A Precision-Recall Criterion Based Consensus Model For Fusing Multiple Segmentations

Max Mignotte and Charles Hélou*

** Département d'Informatique et de Recherche Opérationnelle (DIRO),
Université de Montréal, Faculté des Arts et des Sciences, Montréal H3C 3J7 QC,
Canada
mignotte@iro.umontreal.ca*

Abstract

This paper presents a general framework for seamlessly combining multiple low cost and inaccurate estimated segmentation maps (with an arbitrary number of regions) of the same scene to achieve a final improved segmentation. The proposed fusion model is derived from the well-known precision-recall criterion, specially dedicated to the specific clustering problem of any spatially indexed data and which is also efficient and widely used in the vision community for evaluating both a region-based segmentation and the quality of contours produced by this segmentation map compared to one or multiple ground-truth segmentations of the same image. The proposed combination framework is here specifically designed to be robust with respect to outlier segmentations (that appear to be inconsistent with the remainder of the segmentation ensemble) and includes an explicit internal regularization factor reflecting the inherent ill-posed nature of the segmentation problem. We propose also a hierarchical and efficient way to optimize the consensus energy function related to this fusion model that exploits a simple and deterministic iterative relaxation strategy combining the different segments or individual regions belonging to the segmentation ensemble in the final solution. The experimental results on the Berkeley database with manual ground truth segmentations show the effectiveness of our combination model.

Keywords: *cluster ensemble algorithm, combination of multiple segmentations, F-measure, precision-recall, segmentation ensemble*

1. Introduction

Image segmentation is a low-level vision task which is often the preliminary step in the development of many high-level image understanding algorithms and computer vision systems such as reconstruction problems [1] or 3D object localization/recognition [2, 3].

A plethora of region-based segmentation methods have been proposed so far to solve the difficult unsupervised segmentation problem of textured natural images. Most of these methods exploit first a texture feature extraction step (whose goal is to characterize each meaningful textured region to be segmented) followed by a clustering technique, attempting to group (with different criteria or strategies) spatially coherent regions sharing similar attributes. Years of research in segmentation have thus focused on finding more sophisticated image features and/or more elaborate clustering techniques and significant improvements in the final segmentation results have been achieved, generally at the cost of an increase in model complexity and/or in computational complexity. These methods include segmentation models exploiting directly clustering schemes [4, 5, 6, 7] using Gaussian mixture modeling,

fuzzy clustering approach [8, 9] or fuzzy sets [10] or after a possibly de-texturing approach [7, 11, 12], mean-shift or more generally mode seeking based procedures [13-15], watershed or [16] region growing strategies [17], lossy coding and compression models [18, 16], wavelet transform [19], MRF [20-26], Bayesian [27] texton-based approach [28] or graph-based models [29-31], variational or level set methods [32, 33, 34, 28, 35], deformable surfaces [36], active contour model [37] (with graph partitioning based approach [38] or curve-based techniques, iterative unsupervised thresholding technique [39, 40], genetic algorithm [41], self-organizing map, manifold learning technique, topology, symbolic object based approach [42] and spectral clustering [43] to name a few.

A recent and effective alternative to these segmentation approaches consists in combining several quickly and coarsely estimated segmentation maps of the same scene associated with simpler segmentation model[1] to achieve a final improved segmentation. In this strategy, instead of looking for the best segmentation algorithm (along with its optimal internal parameters) which is hardly possible considering the different types of existing images, one prefers to look for the best fusion model of segmentations, or more precisely for the most efficient criterion for fusing multiple segmentations.

Combining multiple segmentations can be viewed as a special case of the so called *cluster ensemble problem*, *i.e.,* the concept of combining multiple data clusterings for the improvement of the final clustering result, initially explored in the machine learning field [44-46]. Indeed, a distinctive aspect of image data is its spatial ordering and image segmentation is a process of clustering spatially indexed data. Consequently the grouping of pixels into clusters must take into account not only the similarity in the feature space but also the requirement of their spatial coherence. This distinction allows defining the notion of spatial boundaries between different regions which does not exist in a simple clustering process. This characteristic, inherent to the spatially indexed data of any images, allows defining a very efficient criterion of good segmentation called the global F-measure [47] which will be used as criterion of our fusion model. It is worth mentioning that this ensemble segmentation problem can also be viewed as a special type of denoising problem in which each segmentation (to be fused) is in fact a noisy solution or observation and the final objective is to find a denoised segmentation solution which would be a consensus or a compromise (in terms of level of details, contour accuracy, *etc.,*) exhibited by each input segmentations. In some sense, the final fused segmentation is the average of all the individual segmentations to be combined according a defined criterion.

Despite of decades of intensive research to find a universal region-based segmentation algorithm (and/or selected features) that can successfully segment all images, up to now; few works have been proposed on how to efficiently combine multiple (region-based) segmentations or label fields of the same scene. However, we can cite the fusion model proposed in [4] which merges the individual input segmentations in the within-cluster variance (or inertia) sense (for the set of local label histogram values given by each input segmentations) since, the final segmentation result is optimized by applying a K-means algorithm based fusion scheme. In the same vein, we can also cite the fusion scheme proposed in [48] which uses the same strategy but for the set of local soft labels (computed with a multilevel thresholding scheme) and for which the fusion procedure is thus achieved in the (somewhat) sense of the weighted within-cluster inertia. This fusion of (region-based) segmentation maps can also be achieved in the probabilistic Rand index [49] (PRI) sense,

---

[1] These initial segmentations to be fused can be given either by different (and ideally complementary) algorithms or by the same algorithm with different values of the internal parameters or seeds (for stochastic methods), or by using different features and applied to an input image possibly expressed in different color spaces or transformations (*e.g.*, scale, skew, *etc.,*) or by other means

with a consensus function encoding the set of constraints, in terms of pairs of pixel labels (identical or not), provided by each of the segmentations to be combined. This PRI criterion can be optimized either with an algebraic optimization method [50] or with a random walking approach [51] (and combined with a mutual information based estimator for estimating the optimal number of regions in the final segmentation result), or with an Expectation Maximization (EM) algorithm [52] (combined with integer linear programming and applied on superpixels, preliminary obtained by an over-segmentation) or finally in the penalized PRI sense including a global constraint on the fusion process [53] (restricting the number and the size of the regions) with a Markovian approach and an analytical optimization method. Let us also cite the fusion model proposed in [45] in the evidence accumulation sense (and using a hierarchical agglomerative clustering model) and the one proposed in [54] in the variation of information sense (and using an energy-based model optimized by exploiting an iterative steepest local energy descent strategy combined with a connectivity constraint).

The fusion model, proposed in this paper is an hierarchical energy-based model with a consensus (fusion) function derived from the well-known harmonic mean of precision-recall measure (or global F-measure) [47] widespread used for evaluating a soft (or possibly probabilistic) boundary map or the quality of contours produced by a region-based segmentation map (comparatively to a ground-truth segmentation obtained from an expert). In this new framework, we will see that the proposed resulting consensus energy-based fusion model of segmentation can be efficiently optimized by simply applying a deterministic relaxation scheme on each region (or superpixel) given by each individual segmentations to be combined. In addition, we will see how this model can be efficiently designed to be robust with respect to outlier segmentations.

The remainder of this paper is organized as follows: Section 2 describes the combination model and the optimization strategy used to minimize the consensus energy function related to this model. Section 3 describes the generation of the segmentation ensemble to be fused by our model. Finally, Section 4 presents a set of experimental results and comparisons with existing segmentation techniques on the Berkeley natural image database (including, for quantitative evaluations, ground truth segmentations obtained from human subjects).

2. Proposed Fusion Model

2.1. The F Measure

The global F measure (or harmonic mean of precision-recall measure) [47] provides a performance score, evaluating the agreement between region boundaries of a machine segmentation and its ground-truth segmentation[2]. This latter measure is, in fact, deduced from the well-known precision/recall values that characterize, in the image segmentation case, respectively the fraction of detections that are true boundaries and the fraction of true boundaries detected.

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 the true boundaries detected; a low recall value is typically the result of under segmentation and indicates failure to capture the salient image structure. Thus, precision

---

[2] Historically, the F-measure has its origin in the text mining literature for the purpose of document clustering. In this specific context, this measure is used to quantify the accuracy of a clustering solution, *i.e.,* how close a clustering solution, given by a search engine (in response to a user's query), is compared to a human-defined categorization [55].

quantifies the amount of noise in the output of a detector, while recall quantifies the amount of ground-truth detected. In statistical terms, precision and recall are respectively related to the percentage of false positives (or false alarms) and miss detection rate. 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 (or possibly probabilistic) boundary representation, 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.

Image segmentation is not a well-defined task, since the level to which an image is subdivided is determined by the application at hand. In this context, the precision and recall measures are particularly meaningful since it is reasonable to characterize the higher level processing in terms of how much true signal is required to succeed (recall), and how much noise or false alarms can be tolerated (precision). A particular application can define a relative cost α between these two quantities, which focuses attention at a specific point on the precision-recall curve [47]. The F measure, defined as,

$$F_\alpha = \frac{PR}{\alpha R + (1 - \alpha)P} \tag{1}$$

captures this trade off as the weighted harmonic mean of the precision and recall measures. The best F measure, for a given α (*i.e.,* reflecting the optimal compromise between how much true signal is required and how much false alarm can be tolerated), thus allows to find the best segmentation method required (as a pre-processing step) in the development of a given high-level computer vision system[3]. In our fusion model of multiple segmentations (of the same scene), we will see that α can efficiently act as a regularization term for the final fused segmentation result, favoring over segmentation for values close to zero and merging for values close to one (this will be explicit in Section 2.2). In addition, when a ground-truth segmentation is available, α is, by default, set to 0.5 (*i.e., $F = 2PR/(P+R)$*) in order to objectively evaluate the agreement between region boundaries of a machine segmentation and its ground-truth segmentation.

This F measure based performance score was also recently used in image segmentation [47] as a quantitative measure to compare automatic segmentation of an image to a set of ground truth segmentations. This family of ground truth segmentations is, in fact, the multiple acceptable ground truth segmentations associated with each natural image and reflects the inherent variability of possible interpretations between each human observer of an image. This variability between observers, recently highlighted by the Berkeley segmentation dataset [56] 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 regions or segments not known *a priori*.

Hence, in the absence of a unique ground-truth segmentation, this F measure based 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

---

[3] In this case, the maximal F measure, for a given application, (*i.e.,* for a given α) on a precision-recall curve is generally used as a summary statistic for the performance of the detector on a set of images. 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).

image [57]. The authors [56] address this concern by computing the mean F measure as a means of accounting for this variation in the ground truth set.

More formally, let us consider a set of L manually segmented (ground truth) images $\{S_k^g\}_{k \leq L} = \{S_1^g, S_2^g, \cdots, S_L^g\}$ corresponding to an image. Let S^t be the segmentation to be compared with the manually labeled set, the mean F-measure is simply defined by:

$$\overline{F}_\alpha(S^t, \{S_k^g\}_{k \leq L}) = \frac{1}{L} \sum_{k=1}^{L} F_\alpha(S^t, S_k^g) \qquad (2)$$

As a result, the $\overline{F}_\alpha$-measure will give a high score to a machine segmentation S^t which is consistent (in this criterion sense) with most of the segmentation results given by human experts.

2.2. Consensus Energy-Based Fusion Model

Let us consider now that we have at our disposal, a set of L segmentations $\{S_k\}_{k \leq L} = \{S_1, S_2, \cdots, S_L\}$ (associated with a same scene) to be fused in order to obtain a final improved segmentation result relatively to each member of $\{S_k\}_{k \leq L}$. In this context, the previously defined mean F measure on $\{S_k\}_{k \leq L}$ can be straightforwardly used as the consensus or cost function in this energy-based fusion model and in this framework, the consensus segmentation is simply obtained by the solution of the following optimization problem:

$$\hat{S}_{\overline{F}_\alpha} = \arg \max_{S \in S_n} \overline{F}_\alpha(S, \{S_k\}_{k \leq L}) \qquad (3)$$

With S_n the set of all possible segmentations using n pixels. In this way, this fusion model is a generative model of correct segmentation, which can also be considered as a likelihood of $\hat{S}$, in the maximum $\overline{F}_\alpha$ sense or equivalently in the Maximum Likelihood (ML) sense for this $\overline{F}_\alpha$ criterion, for the given segmentation ensemble $\{S_k\}_{k \leq L}$ (considered as a set of observations). Let us note that this optimization approach is also called the median partition [46] with respect to both the mean F measure criterion (used in this application) and the segmentation ensemble $\{S_k\}_{k \leq L}$.

In our application, in order to increase the robustness of our estimator $\hat{S}_{\overline{F}_\alpha}$ relatively to the outliers, *i.e.,* relatively to the possible segmentation maps belonging to $\{S_k\}_{k \leq L}$ far away (according to our F measure-based criterion) from the averaged or fused segmentation result, (*i.e.,* statistically speaking, relatively to an observation or subset of observations which appears to be inconsistent with the remainder of the segmentation ensemble), we have decided to weight the importance (or equivalently the confidence) of each segmentation of $\{S_k\}_{k \leq L}$ by a coefficient w_k proportional to its mean F measure $\overline{F}_{\alpha=0.5}(S_k, \{S_k\}_{k \leq L})$.

In this context, our final fusion model, generative of averaged segmentation, in the mean F measure sense, is expressed by:

$$\hat{S}_{\overline{F_\alpha}} = \arg\max_{S \in S_n} \left\{ \frac{1}{L} \sum_{k=1}^{L} w_k F_\alpha\left(S, \{S_k\}_{k \leq L}\right) \right\}$$

$$\hat{S}_{\overline{F_\alpha}} = \arg\max_{S \in S_n} \left\{ \overline{F_\alpha}\left(S, \{w_k\}_{k \leq L}, \{S_k\}_{k \leq L}\right) \right\}$$

$$\text{With, } w_k = \frac{1}{Z} \exp\left(\frac{\overline{F_{\alpha=0.5}}\left(S_k, \{S_k\}_{k \leq L}\right)}{\tau} \right) \tag{4}$$

in which Z is a normalizing constant ensuring $\sum_k w_k = L$, and τ is a parameter controlling the decay of the weights, (*e.g.,* for high values of τ, $w_k \approx 1 \; \forall \; k$ and for small values of τ, a large subset of the weights is close to 0, thus removing the outliers which appears to be inconsistent with the remainder of the segmentation ensemble).
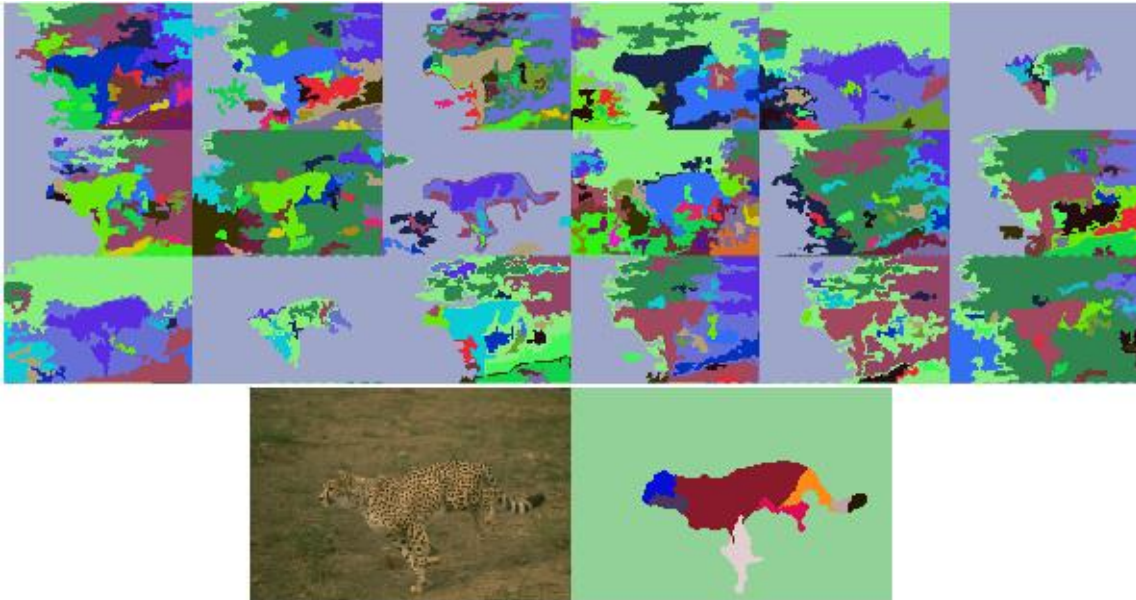

Figure 1. Examples of Segmentation Ensemble and our Fusion Result (Algorithm FMBFM)

From top to bottom; three first rows; K-means clustering results for the segmentation model described in Section 3. Input natural image from the Berkeley image database and final segmentation map resulting of our fusion model.

Algorithm 1
F measure-Based Fusion Model

$\{S_k\}_{k \leq L}$	Set of L segmentations to be fused
$\{w_k\}_{k \leq L}$	Set of weights
$\{b_k\}$	Set of superpixels $\in \{S_k\}_{k \leq L}$
$\mathcal{E}$	Set of region labels $\in \{S_k\}_{k \leq L}$
α	Compromise parameter of the Fmeas.
$T_{\max}$	Maximal number of iterations (=12)

1. Initialization

- $\hat{S}_{F_\alpha}^{[0]} = \arg\max_{S \in \{S_k\}_{k \leq L}} \overline{F}_\alpha(S, \{S_k\}_{k \leq L})$
- Compute $\overline{F}_\alpha(S, \{w_k\}_{k \leq L}, \{S_k\}_{k \leq L})$ on $\hat{S}_{F_\alpha}^{[0]}$

2. Steepest Local Energy Ascent

while $p < T_{\max}$ **do**

  for each b_k superpixel $\in \{S_k\}_{k \leq L}$ **do**

  - Draw a new label x according to the uniform distribution in the set $\mathcal{E}$
  - Let $\hat{S}_{F_\alpha}^{[p],\text{new}}$ the new segmentation map including b_k with the region label x
  - Compute $\overline{F}_\alpha^{\text{new}}(S, \{w_k\}, \{S_k\}_{k \leq L})$ on $\hat{S}_{F_\alpha}^{[p],\text{new}}$

  if $\overline{F}_\alpha^{\text{new}} > \overline{F}_\alpha$ **then**

  - $\blacktriangleright \quad \overline{F}_\alpha = \overline{F}_\alpha^{\text{new}}$
  - $\blacktriangleright \quad \hat{S}_{F_\alpha}^{[p]} = \hat{S}_{F_\alpha}^{[p],\text{new}}$

  $p \leftarrow p + 1$

2.1. Fusion Model Optimization

Our fusion model of several label fields, in the $\overline{F}_\alpha$ criterion sense, thus ends up as an optimization problem of a complex non-convex cost function with several local extrema across the lattice of possible clusterings S_n.

The difficulty (and more precisely the non-convexity) of this optimization problem, lies in the fact that many different region-based segmentation images can give the same optimal binary boundary representation, (*i.e.,* the one optimizing our criterion or ensuring the maximal mean F measure). Indeed, it is important to recall that the estimation of the F measure (see Eq. (1)) exploits a squared window search (or a window search with a ball shape) of a few pixels, (*e.g.,* 3 pixels) in order to estimate both the percentage of false positives (or false alarms) and the miss detection rate between the binary boundary representation of the machine segmentation and the ground-truth segmentation. This window search based procedure is essential in order to take into account the fact that each contour of the ground truth

segmentations are not perfect and are hand drawn with a certain degree of accuracy. Due to this level of accuracy of a few pixels (2 or 3 pixel size), it is important to understand that each segmentation map whose binary boundary representation are within this aforementioned accuracy would be considered as similar. In this context, the boundary representation of the optimal region-based segmentation (solution of our fusion model) and the degenerate solution given by the boundary representation of this latter boundary representation (exhibiting a segmentation solution without spatial homogeneity, with one-pixel wide region corresponding to the contours) will give the optimal and maximum F measure.

In order to find a reliable estimation of $\hat{S}_{\overline{F_\alpha}}$ that efficiently maximizes this complex energy function and that leads to a region-based segmentation solution (and not a degenerate solution without spatial homogeneity), we must constrain the solution space in order to avoid region-based segmentation solutions with possibly one-pixel wide regions. In this context, the strategy used in [54], exploiting the local expression of the decrease (or increase) in the consensus function for each pixel update of this consensus segmentation to be estimated (thus avoiding the prohibitive calculus of the global consensus measure for the entire segmentation map thanks to a relaxation scheme based on a pixel-wise optimization strategy) is useless for our consensus measure based on the $\overline{F}_\alpha$ criterion. In order to avoid a prohibitive computational complexity, requiring the calculus of the global consensus criterion for the entire segmentation map, we have decided to apply an optimization procedure based on the set of superpixels[4,5] existing in $\{S_k\}_{k \leq L}$, $i.e.$, the set of regions or segments given by each individual segmentations to be fused. This strategy has a second important advantage. Indeed, it is logical to think that we could efficiently and spatially combine the strengths of multiple segmentation maps which, individually, might produce some poor segments or regions, ($i.e.,$ poor segmentation result for some sub-parts of the image) but for which there also often exist good segments (in other sub-parts of the image). The set of superpixels existing in $\{S_k\}_{k \leq L}$ are likely to contain the different right segments or individual regions of the optimal segmentation solution.

For the optimization procedure, we have chosen the simple Iterative Conditional Modes (ICM) introduced by Besag [60], $i.e.,$ a Gauss-Seidel relaxation, where superpixels (in our hierarchical approach) are updated one at a time. This iterative search technique 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 bad local minima associated with our complex energy function. In order to solve this problem, we can take, as initialization (first iteration), the segmentation map $\hat{S}_{\overline{F_\alpha}}^{[0]}$ such as:

$$\hat{S}_{\overline{F_\alpha}}^{[0]} = \arg \max_{S \in \{S_k\}_{k \leq L}} \overline{F}_\alpha \left(S, \{S_k\}_{k \leq L} \right) \tag{5}$$

---

[4] Let us note that the use of superpixels in an energy-based fusion procedure has been initially proposed in [58] with a different goal, namely the one 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, $etc.,$) and in [59] for restoration application.

[5] Another reliable strategy consists of initializing the ICM with the first N_I optimal input segmentations (in the mean F measure sense) and to finally retain, after convergence of the ICM procedure, the segmentation result ensuring the highest mean F_α measure. This procedure allows improving very slightly the reliability of our fusion model but at a prohibitive computational cost.

i.e., in choosing for the first iteration of the ICM procedure, amongst the L segmentation to be fused, the one ensuring the maximal consensus energy (in the mean F measure sense) of our fusion model (Eq. (4))[5]. In our case, where our optimization problem is a maximization problem, ICM is an iterative steepest local energy ascent algorithm which searches to obtain, for each (super)-pixel to be labeled, the maximum energy label assignment. Starting with $\widehat{S}\,^{[0]}_{\overline{F}_\alpha}$, *i.e.,* a solution not too far from the optimal solution (see Eq. (5)), ICM chooses, at each iteration and (sequentially) for each (super)-pixel, the label (of the final segmentation result), yielding the largest increase of the energy function, conditioned on the labels assigned to its neighbors.

It is also worth mentioning that algorithmically, the computation of the largest increase of the energy function $\overline{F}_\alpha^{\,new}$ (see Algorithm 1) may be efficiently and quickly answered, in C++ by exploiting the BITSET class (which is is very similar to a regular array, but optimizing for space allocation, each element occupies only one bit) and appropriate Boolean operations (such as the logical AND bit-wise operator in order to estimate the precision and the recall values, between two binary contours-based segmentation maps, required to compute the F - measure $\overline{F}_\alpha^{\,new}$). Similar bit-wise operator and bit-set operations are also available in Matlab.

Finally, the overall F measure-Based Fusion Model (FMBFM) algorithm with the iterative steepest local energy ascent strategy and the maximal energy label assignment of each superpixel belonging to $\{S_k\}_{k \leq L}$ is outlined in pseudo-code in Algorithm 1.

3. Segmentation Ensemble Generation

The initial segmentation maps, which will be fused by our fusion framework are simply given, in our application, by a *K*-means [61] clustering technique, with respectively; different features, several values of K and expressed in 12 different color spaces, namely; RGB, HSV, YIQ, XYZ, LAB, LUV, i123, h123, YCbCr, TSL, HSL, P1P2 (see [53] for a justification of these color spaces and for references), *i.e.,*.

1. As the number of classes *K* of the *K*-means algorithm, we use for each image, a metric measuring the complexity, in terms of the number of different texture types, of a natural color image. This metric, introduced in [62] is herein defined as the measure of the absolute deviation (L_1 norm) of the set of normalized histograms obtained for each overlapping squared fixed-size (N_w) neighborhood contained within the input image. This measure ranges in [0, 1] and an image with several different texture types will result in value of complexity close to 1. In our application,

    $$K = 1 + ceil\ (K^{\,max} \times complexity \quad value\ ) \qquad (6)$$

    where ceil(*x*) is a function that round *x* up to the nearest integer and $K^{\,max}$ is an upperbound of the number of classes for a very complex natural image. In our application, we use three different values of $K^{\,max}$, namely $K_1^{\,max} = 8$, $K_2^{\,max} = K_1^{\,max} + 1$ and $K_3^{\,max} = K_1^{\,max} / 2$.

2. As input multidimensional feature descriptor, we used the set of values of the requantized color histogram, with equidistant binning, estimated around the pixel to be classified. In our application, this local histogram is equally re-quantized, 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 segmented with two different values of q_b, namely $q_b = 5$ and $q_b = 4$.

For a total of $12 \times (3 + 2) = 60$ input segmentations to be fused. This generation process allows us to ensure the diversity required to obtain a good, (*i.e.,* reliable) segmentation ensemble on which the final result will be conditioned. It is worth mentioning that the more varied the set of segmentations is, the more information for the consensus function (on which the fusion model is based) is available [53, 46] (conversely, it is logic to think that a combination of similar segmentation solutions could not give an improved segmentation that outperforms the individual ensemble members).
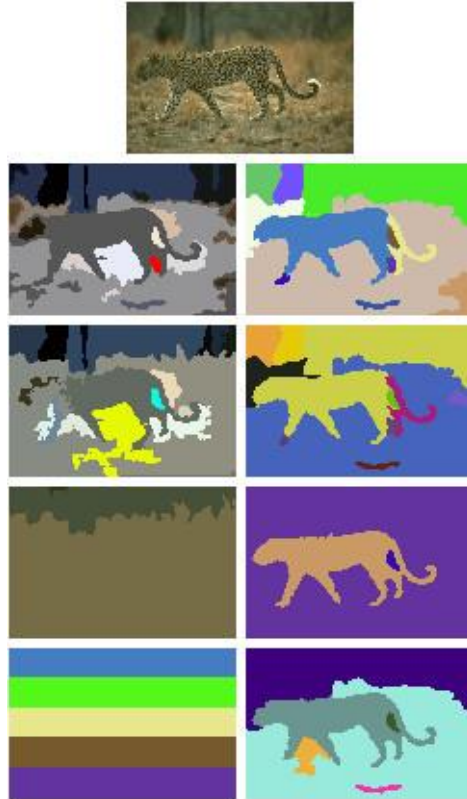

Figure 2. Example of Fusion Convergence Result on Four Different Initializations for the Berkeley Image (n^0 134052)

Left: initialization and right: result at the convergence of our FMBFM model (12 iterations). From top to bottom, the original image, the input segmentations (from the segmentation ensemble $\{S_k\}_{k \leq L}$ which have the best and the L/2 = 30-th best $\overline{F}_\alpha$ score and the input segmentation which have the worst $\overline{F}_\alpha$ score and one blind (or non-informative) initialization.

4. Experimental Results

4.1. Setup and Initial Tests

In all the experiments, we have considered our fusion model (see Eq. (4)) based on a segmentation ensemble $\{S_k\}_{k \leq L}$ generated as indicated in Section 3 (see Figure 1 for an example of segmentation ensemble generated by our K-means based procedure). In addition, for these initial tests, we have set $K_1^{max} = 8$, $\tau = 16$ and $\alpha = 0.86$ (these values of the internal parameters of our fusion model will be explained in Section 4.2).

First, we have tested the convergence of our iterative optimization procedure by taking, as initialization of our ICM-based iterative steepest local energy ascent algorithm, respectively, the input segmentations (of our segmentation ensemble $\{S_k\}_{k \leq L}$ which have the best, (*i.e.,* maximal) $\overline{F}_\alpha$ score, the $L/2 = 30$-th best score, the worst (*i.e.*, minimal) $\overline{F}_\alpha$ score and one blind (or non-informative) initialization by considering an image spatially divided by $K = 5$ horizontal rectangles with K different labels (see Figure 2 and 3). We can notice that our strategy, consisting in relaxing the set of superpixels belonging to the segmentation ensemble, remains robust to the initialization. We can also notice than our strategy, consisting in choosing for the first iteration of the ICM procedure, the segmentation (amongst the L segmentation to be fused) closest to the optimal solution of the consensus energy function of our fusion model (Eq. (4)), allows to improve somewhat the final segmentation result.

Second, we have tested the role of the parameter α (see Eq. (4) and Algorithm 1.) on the obtained segmentation solutions. Figure 4 shows clearly that α efficiently acts as a regularization parameter of our fusion model favoring oversegmentation (for value close to 0) and merging (for value close to 1).
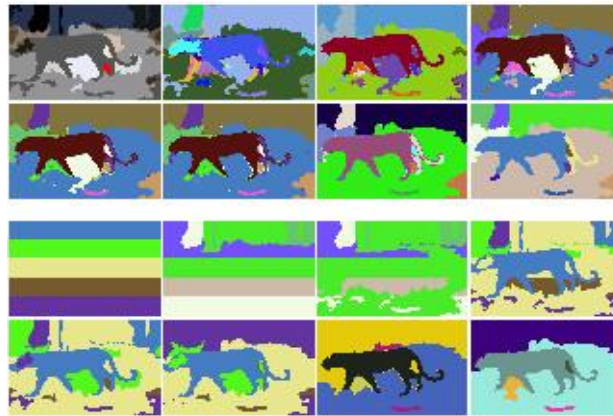

Figure 3. From Lexicographic Order, Evolution of the Resulting Segmentation Map along the Iterations of the Relaxation Process for 1-] the Initial Segmentation which have the Best $\overline{F}_\alpha$ Score and 2-] for One Non Informative Initialization
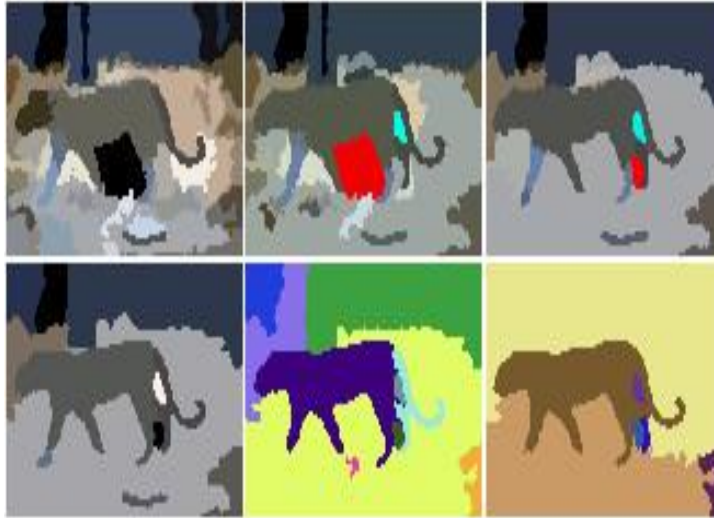
Figure 4. Example of Segmentation Solutions Obtained for Different Values of α, from Top to Bottom and Left to Right, α = {0.5, 0.7, 0.8, 0.86, 0.9, 0.99}

4.2. Performance Measures & Comparison With State-Of-The-Art Methods

In these experiments, we have tested our fusion model as segmentation algorithm on the Berkeley segmentation database (BSD300) [56] for which the color images are normalized to have the longest side equal 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.

In order to validate our segmentation model, several performance metrics will be estimated (for the entire image database) for an objective comparison with the other segmenters. These performance measures include the PRI [63] score which is highly correlated with human hand-segmentations [6] and widely used in the region-based segmentation field. This PRI score quantifies the percentage of pairs of pixel labels correctly classified in the segmentation results and a score equal to PRI=0.80 means that, on average, 80% of pairs of pixel labels are correctly classified in the segmentation results on the BSD300.

In order to ensure the integrity of the evaluation, the internal parameters of our segmentation algorithm, namely K_1^{max} required for the segmentation ensemble generation (see Section 3) and α (and to a lesser measure, the parameter τ) for the fusion model (see Eq. (4)) 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 $K_1^{max} \in [5-10]$ [step-size = 1], $\alpha \in [0.5-1]$ [step-size = 0.02] and $\tau \in [0.125-16]$ [step-size = a power of 2.0]. We have found that $K_1^{max} = 8$, $\alpha = 0.86$ and $\tau = 16^6$ is a good set of internal parameters leading to a very good PRI score of 0.80 (see Table 1).

Consequently, a good strategy (in the case of a fusion model based on the *F* measure criterion) consists in proposing several over-segmentations in the segmentation ensemble and a high value of regularization for α, thus favoring a final segmentation result with contours predominantly found in this set of segmentations.

For comparison, we now illustrate the results of our segmentation algorithm by showing the same segmented images (see Figures 7 and 8) as those shown in the fusion model proposed in [53, 54] and in the segmentation algorithms proposed in [12, 11]. The results for

the entire database will be available on the website of the author. It may be noted that our segmentation procedure gives a very competitive PRI score among the state-of-the-art segmentation methods recently proposed in the literature. Figure 5 shows respectively the distribution of the PRI measure and the number and size of regions obtained by our FMBFM algorithm over the BSD300.
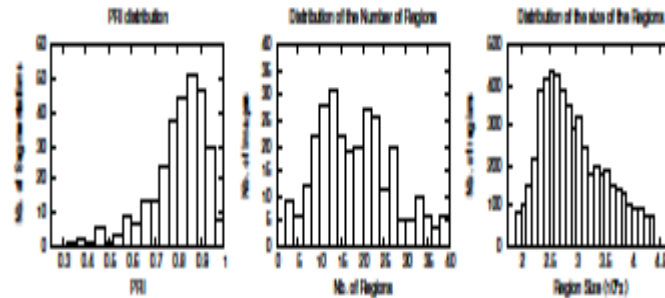

Figure 5. From Left to Right, Distribution of the -1- PRI measure -2- Number and -3- Size of Regions Over the 300 Segmented Images of the Berkeley Image Database

We have also compared our segmentation method with the VoI metric [68], the GCE[56] and the BDE [69] (see Table 2) (for which a lower distance is better), showing that our method gives competitive results for some other metrics based on different criteria and compared to state-of-the arts.

We have also tested the performance of our fusion method as segmentation method, in term of F measure (see Table 3). First, we should remember that this measure is best appropriate for contour detection methods giving a "soft" boundary representation since this benchmark measure also finds (from a soft edge map) the optimal threshold value ensuring the best F measure [70] over the BSD300. In this spirit, we will let this benchmark measure in choosing the optimal threshold on a soft contour map provided by averaging, 6 times, the set of "hard", (*i.e.,* binary) boundary representations of our segmentation method with K^{max}, the number of classes of the segmentation step, varying in an interval containing an upper and lower bound of the number of classes, *e.g.,* $K^{max} \in [K^{max} : K^{max} + 6]$. We have obtained $F = 0.62@(R = 0.65, P = 0.59)$ for the BSD300, which remains competitive compared to the state-of-the art existing segmentation methods and a significant improvement, compared to $F = 0.53@(R = 0.70, P = 0.42)$, the overall score obtained by the segmentation result achieved by a single K-means based segmentation map (with the features proposed in Section 3) for each image, without fusion method.

ALGORITHMS	PRI [63]
HUMANS on [73]	0.87
FMBFM	**0.80**
-2014- VOIBFM [54]	0.81
-2012- MDSCCT [12]	0.81
-2011- gPb-owt-ucm [64]	0.81
-2012- AMUS [52]	0.80
-2010- PRIF [53]	0.80
-2008- CTex [5]	0.80
-2009- MIS [36]	0.80
-2011- SCKM [7]	0.80
-2008- FCR [4]	0.79
-2012- SPSBM [62]	0.79
-2004- FH [30] on [s]	0.78
-2011- MD2S [11]	0.78
-2009- HMC [25]	0.78
-2009- Consensus [50]	0.78
-2009- Total Var. [35]	0.78
-2009- A-IFS HRI [10]	0.77
-2001- JSEG [17] on [s]	0.77
-2011- KM [34]	0.76
-2007- CTM [6,18]	0.76
-2008- Av. Diss. [52] on [s4]	0.76
-2011- SCL [65]	0.76
-2005- Mscuts [43] on [73]	0.76
-2003- Mean-Shift [13] on [s]	0.75
-2008- NTP [31]	0.75
-2010- iHMRF [26]	0.75
-2005- NCuts [43] on [s4]	0.75
-2006- SWA [66] on [s4]	0.75

Table 1. Average Performance, in Term of PRI Measure, of Several (Region-Based) Segmentation Algorithms on the BSD300, Ranked According to their PRI Score and Considering only the (Published) Segmentation Methods with a PRI Score above 0.75

ALGORITHMS	VoI	GCE	BDE
HUMANS	1.10	0.08	4.99
FMBFM	**2.01**	**0.20**	**8.49**
VOIBFM [54]	1.88	0.20	9.30
MDSCCT [12]	2.00	0.20	7.95
PRIF [53]	1.97	0.21	8.45
SCKM [7]	2.11	0.23	10.09
MD2S [11]	2.36	0.23	10.37
FCR [4]	2.30	0.21	8.99
CTM [6,18]	2.02	0.19	9.90
Mean-Shift [13] on [s]	2.48	0.26	9.70
NCuts [29] on [s]	2.93	0.22	9.60
FH [30] on [s]	2.66	0.19	9.95
AMUS [52]	1.68	0.17	-

Table 2. Average Performance of Several Algorithms for Different Performance Measures (Lower is Better) on the BSD300

ALGORITHMS	F Measure [56]
HUMANS on jsp	0.79
FMBFM	0.62
-2011- gPb-owt-ucm [64]	0.71
-2010- PRIF [53]	0.64
-2012- MDSCCT [12]	0.63
-2003- Mean-Shift [13] on jstp	0.63
-2000- NCuts [43] on jstp	0.62
-2007- CTM [6] on jstp	0.58
-2004- FH [30] on jsp	0.58
-1986- (Canny [67]) on jsp	0.58
-2006- SWA [66] on jstp	0.56
-2008- PCR [4]	0.56
K-means (with features proposed in Sect. 3)	0.53
Quad-Tree on jstp	0.37

Table 3. Average Performance, in Term of F Measure, of Several Segmentation (Into Contours) Algorithms and Contour Detectors (in Parentheses) on the BSD300
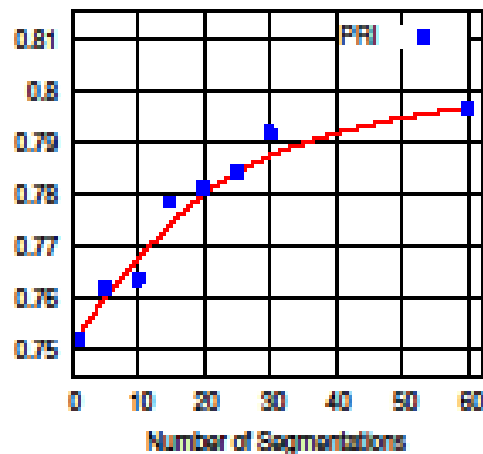

Figure 6. Evolution of the PRI (Higher is Better) as a Function of the Number of Segmentations (L) to be Fused for our FMBFM Algorithm. More Precisely for *L* = 1, 5, 10, 15, 20, 30, 60 Segmentations (by Considering First, One *K*-means Segmentation and then by Considering 5 Segmentation for each Color Space and 1, 2, 3, 4, 6, 12 Color Spaces)
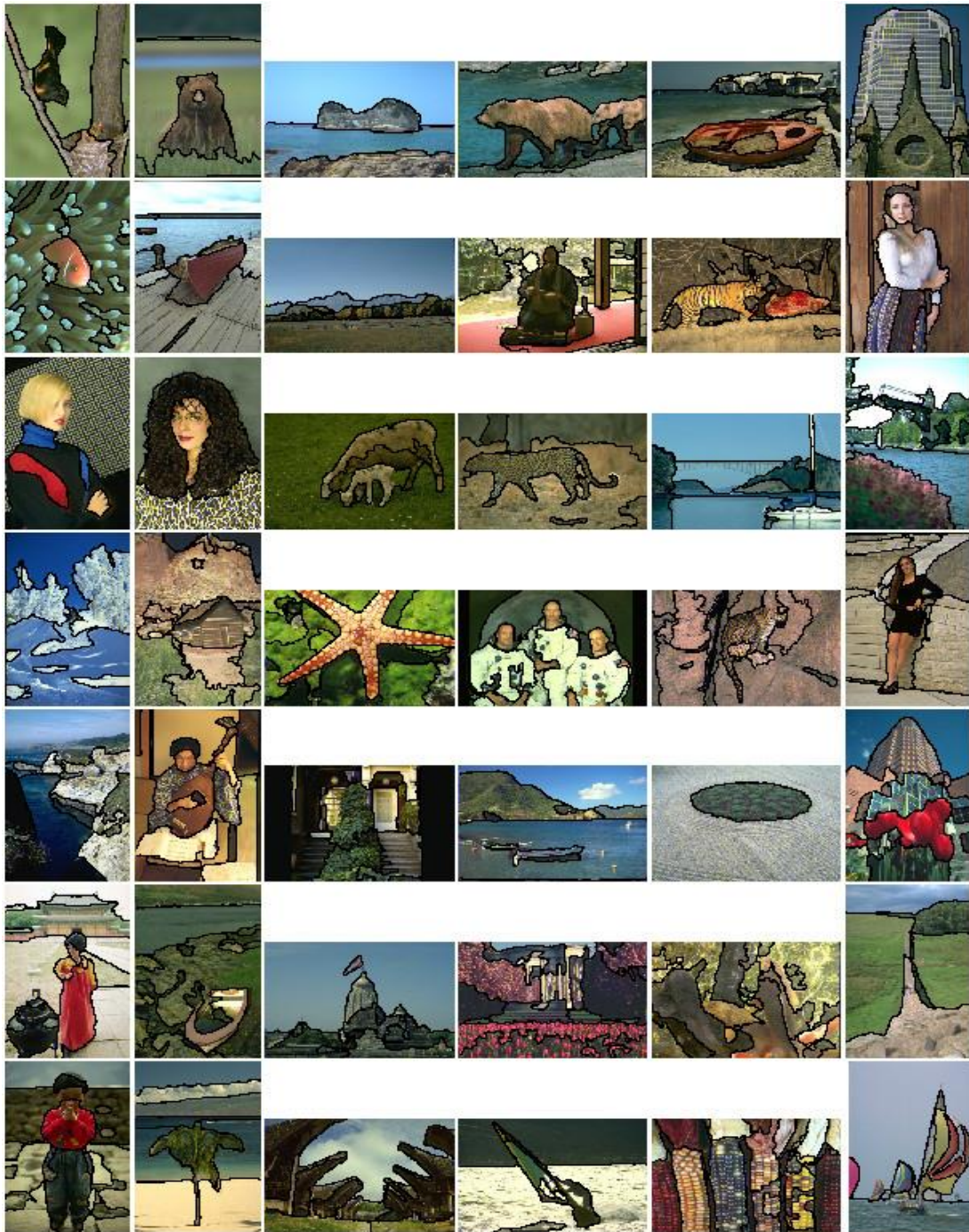
Figure 7. Example of Segmentations Obtained by our Algorithm FMBFM on Several Images of the Berkeley Image Database (see also Tables 1, 2 and 3 for Quantitative Performance Measures and on the Website of the Author for the Segmentation Results on the Entire Database)
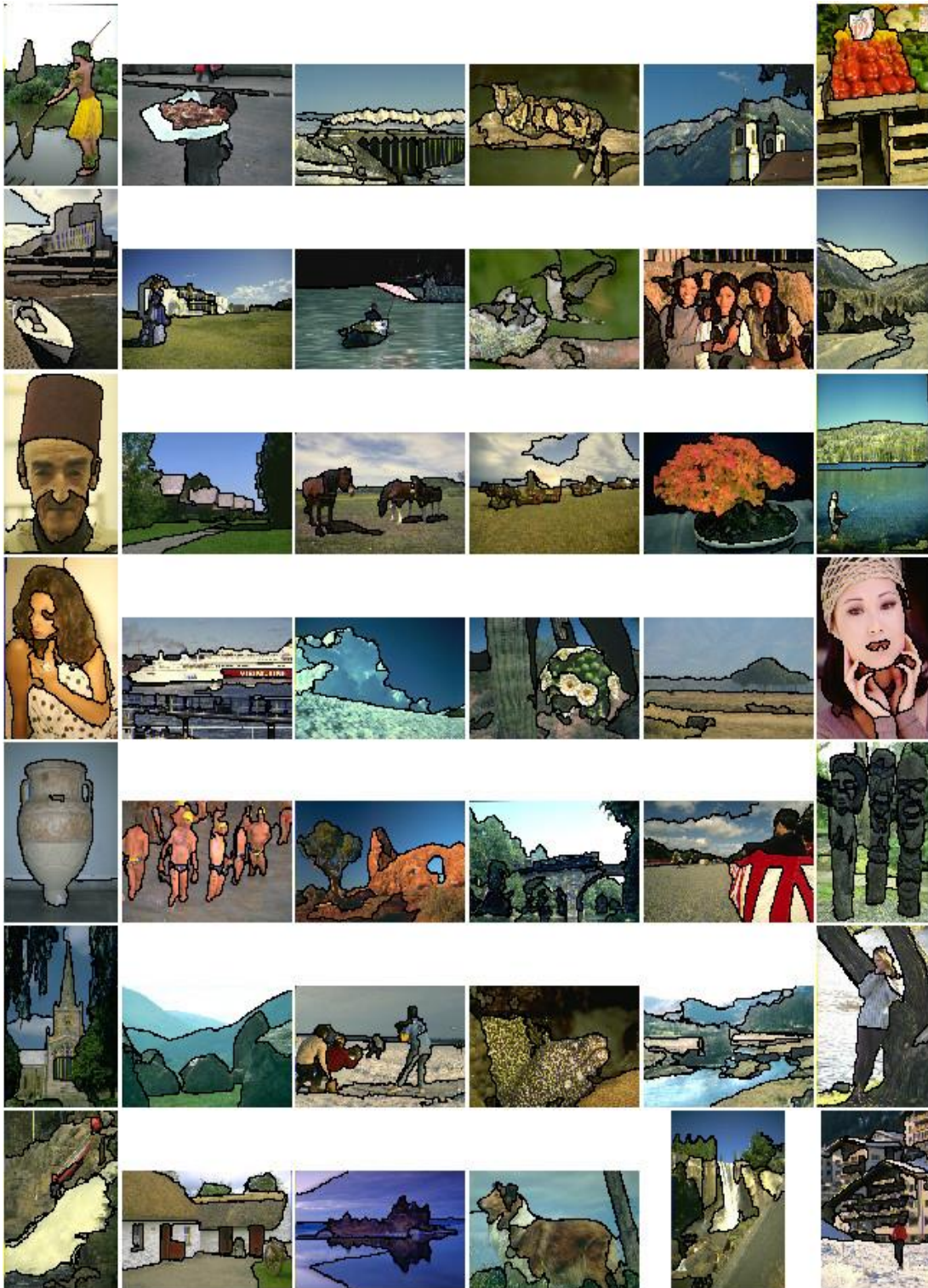
Figure 8. Example of Segmentations Obtained by our Algorithm FMBFM on Several Images of the Berkeley Image Database (see also Tables 1, 2 and 3 for Quantitative Performance Measures and on the Website of the Author for the Segmentation Results on the Entire Database)

4.3. Discussion

As we can notice, our method of fusion of simple, quickly estimated segmentation results appears to be very competitive for different kinds of performance metrics and thus appears as an interesting alternative to complex, computationally demanding segmentation models existing in the literature. We can also observe (see Figure 6) that the PRI performance measures are better when L (number of segmentation to be fused) is high. This experiment shows the validity of our fusion procedure and shows also that our performance scores are perfectible if the segmentation ensemble is completed by other (and different and/or ideally complementary) segmentation maps (of the same scene).

4.4. Algorithm

The segmentation procedure takes, on average, between 70 and 90 seconds for a Core i7 Intel© , 3.2 GHz, 6403 bogomips and non-optimized code running on Linux. More precisely, the two steps, (*i.e.,* 1] estimations of the L = 60 weak segmentations to be fused and 2] the minimization step of our fusion procedure) takes respectively, on average, one minute for the segmentation ensemble generation and less than 30 seconds for the fusion step and for a 320 $\times$ 214 image. Let us add that the initial segmentations to be fused and the proposed energy-based fusion method can be easily computed in parallel. It is straightforward for the generation of the segmentation ensemble but also truth for our fusion model by considering a Jacobi-type version of the Gauss-Seidel based ICM procedure [71]. The final energy-based 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 (up to a factor of 200) as indicated in [71].

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

5. Conclusion

In this paper, we have presented a new and efficient fusion model of segmentation based on a consensus energy-based fusion procedure whose goal is to combine, in the precision-recall sense, multiple (simple) segmentation maps to achieve a final improved segmentation result. This framework of segmentation combination performs well compared to the best existing state-of-the-art segmentation methods and thus appears as an interesting alternative to complex segmentation models existing in the literature. It remains simple to implement, perfectible (by increasing the number of segmentation to be fused), robust to outliers and is easily parallelizable (and thus especially well-suited for the next generation massively parallel computers, embedded graphics or multi-core processors). In addition, the proposed hierarchical optimization approach based on a deterministic relaxation scheme combining the set of superpixels belonging to the segmentation ensemble is simple, efficient and also general enough to be applied to other fusion models of label fields (in the sense of other criteria). Furthermore, this fusion model includes an explicit internal regularization factor reflecting the optimal compromise between how much true signal is required and how much false alarms can be tolerated in the resulting segmentation map, thus allowing to design the appropriate fusion model allowing to find the best segmentation map required, as a pre-processing step in the development of a given high-level computer vision system.

References

[1] S. Benameur, M. Mignotte, F. Destrempes and J. D. Guise, "Three-dimensional biplanar reconstruction of scoliotic rib cage using the estimation of a mixture of probabilical prior models", IEEE trans. on Biomedical Engineering, 3D reconstruction, vol. 52, no. 10, **(2005)**, pp. 2041–2057.

[2] M. Mignotte, C. Collet, P. P´erez and P. Bouthemy, "Hybrid genetic optimization and statistical model-based approach for the classification of shadow shapes in sonar imagery", IEEE Trans. Pattern Anal. Machine Intell., vol. 22, no. 2, **(2000)**, pp. 129–141,

[3] F. Destrempes and M. Mignotte, "Localization of shapes using statistical models and stochastic optimization", IEEE Trans. Pattern Anal. Machine Intell., vol. 29, no. 9, **(2007)**, pp. 1603–1615.

[4] M. Mignotte, "Segmentation by fusion of histogram-based K-means clusters in different color spaces", IEEE Trans. Image Processing, vol. 17, **(2008)**, pp. 780–787.

[5] D. E. Ilea and P. F. Whelan, "Ctex- an adaptive unsupervised segmentation algorithm on color-texture coherence", IEEE Trans. Image Processing, vol. 17, no. 10, **(2008)**, pp. 1926–1939.

[6] A. Y. Yang, J. Wright, S. Sastry and Y. Ma, "Unsupervised segmentation of natural images via lossy data compression", Computer Vision and Image Understanding, vol. 110, no. 2, **(2008)**, May, pp. 212–225,

[7] M. Mignotte, "A de-texturing and spatially constrained K-means approach for image segmentation", Pattern Recognition letter, vol. 32, no. 2, **(2011)** January, pp. 359–367.

[8] D. M´ujica-Vargas, F. J. Gallegos-Funes, A. J. Rosales-Silva and J. Rubio, "Robust c-prototypes algorithms for color image segmentation", EURASIP Journal on Image and Video Processing, vol. 63, **(2013).**

[9] S. Xu, L. Hu, X. Yang and X. Liu, "A cluster number adaptive fuzzy c-means algorithm for image segmentation", International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 6, no. 5, **(2013).**

[10] M. M. Mushrif and A. K. Ray, "A-IFS histon based multithresholding algorithm for color image segmentation", IEEE Signal Processing Lett., vol. 16, no. 3, pp. 168–171, **(2009).**

[11] M. Mignotte, "MDS-based multiresolution nonlinear dimensionality reduction model for color image segmentation", IEEE Trans. Neural Networks, vol. 22, no. 3, **(2011)** March**,** pp. 447–460.

[12] ——, "Mds-based segmentation model for the fusion of contour and texture cues in natural images", Computer Vision and Image Understanding, vol. 116, **(2012)** September, pp. 981–990.

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

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

[15] M. A. Carreira-Perpinan, "Fast nonparametric clustering with Gaussian blurring mean-shift", in Proc.of the International Conference on Machine Learning (ICML'06), **(2006)**, pp. 153–160.

[16] I. Mecimore and C. D. Creusere, "Unsupervised bitstream based segmentation of images", in Proc. of the Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop 2009, **(2009)** January, pp. 643–647.

[17] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video", IEEE Trans. Pattern Anal. Machine Intell., vol. 23, no. 8, **(2001)**, pp. 800–810.

[18] Y. Ma, H. Derksen, W. Hong and J. Wright, "Segmentation of multivariate mixed data via lossy coding and compression", IEEE Trans. Pattern Anal. Machine Intell., vol. 29, no. 9, **(2007)**, pp. 1546–1562.

[19] H. Wu, M. Li, M. Zhang, J. Zheng and J. Shen, "Texture segmentation via scattering transform", International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 6, no. 2, **(2013).**

[20] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images", IEEE Trans. Pattern Anal. Machine Intell., vol. 6, no. 6, **(1984)**, pp. 721–741.

[21] M. Mignotte, C. Collet, P. P´erez and P. Bouthemy, "Three-class Markovian segmentation of high resolution sonar images", Computer Vision and Image Understanding, vol. 76, no. 3, **(1999)**, pp. 191–204.

[22] M. Mignotte, C. Collet, P. P´erez and P. Bouthemy, "Sonar image segmentation using a hierarchical MRF model," IEEE Trans. Image Processing, vol. 9, no. 7, **(2000)**, pp. 1216–1231.

[23] 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 Processing, vol. 14, no. 8, **(2005)** August, pp. 1096–1108.

[24] F. Destrempes, J.-F. Angers and M. Mignotte, "Fusion of hidden Markov Random Field models and its Bayesian estimation", IEEE Trans. Image Processing, vol. 15, no. 10, **(2006)** October, pp. 2920–2935.

[25] H. Rachid and M. Mignotte, "A hierarchical graph-based Markovian clustering approach for the unsupervised segmentation of textured color images", in Proc. of the IEEE International Conference on Image Processing (ICIP'09), Cairo, Egypt, **(2009)** November, pp. 1365–1368.

[26] S. Chatzis and G. Tsechpenakis, "The infinite hidden Markov random field model", IEEE Trans. Neural Networks, vol. 21, no. 6, **(2010)**, pp. 1004–1014.

[27] S. Chen, L. Cao and Y. Wang, "Image segmentation by ML-MAP estimations", IEEE Trans. Image Processing, vol. 19, no. 9, **(2010)**, pp. 2254 – 2264.

[28] X. He, Z. Song and J. Fan, "A novel level set image segmentation approach with autonomous initialization contour", International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 6, no. 4, **(2013).**

[29] J. Shi and J. Malik, "Normalized cuts and image segmentation", IEEE Trans. Pattern Anal. Machine Intell., vol. 22, no. 8, **(2000)**, pp. 888–905.

[30] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation", International Journal on Computer Vision, vol. 59, **(2004)**, pp. 167–181.

[31] J. Wang, Y. Jia, X.-S. Hua, C. Zhang and L. Quan, "Normalized tree partitionning for image segmentation", in IEEE Computer Society Conference on computer vision and pattern recognition (CVPR'08), Anchorage, AK (USA), **(2008)**June, pp. 1–8.

[32] L. Bertelli, B. Sumengen, B. Manjunath and F. Gibou, "A variational framework for multi-region pairwise similarity-based image segmentation", IEEE Trans. Pattern Anal. Machine Intell., vol. 30, no. 8, **(2008)**, pp. 1400–1414.

[33] M. Donoser, M. Urschler, M. Hirzer and H. Bishof, "Saliency driven total variational segmentation", in Proc. of the IEEE Int'l Conf. Computer Vision (ICCV'09), **(2009)**.

[34] M. B. Salah, A. Mitiche and I. B. Ayed, "Multiregion image segmentation by parametric kernel graph cuts", IEEE Trans. Image Processing, vol. 20, no. 2, **(2011)**, pp. 545–557.

[35] Q. Chen and C. He, "Integrating clustering with level set method for piecewise constant mumford-shah model", EURASIP Journal on Image and Video Processing, vol. 1, **(2014)**.

[36] M. Krninidis and I. Pitas, "Color texture segmentation based on the modal energy of deformable surfaces," IEEE Trans. Image Processing, vol. 7, no. 18, **(2009)**, pp. 1613–1622.

[37] Y. Wang and C. He, "Image segmentation algorithm by piecewise smooth approximation," EURASIP Journal on Image and Video Processing, vol. 16, **(2012)**.

[38] S. Nath and K. Palaniappan, "Fast graph partitioning active contours for image segmentation using histograms," EURASIP Journal on Image and Video Processing, **(2009)**.

[39] Y. Chen and O.-C. Chen, "Image segmentation method using thresholds automatically determined from picture contents," EURASIP Journal on Image and Video Processing, **(2009)**.

[40] F. Nie, J. Li, T. Tu and P. Zhang, "Image segmentation using two-dimensional extension of minimum within-class variance criterion," International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 6, no. 5, **(2013)**.

[41] S. Chabrier, C. Rosenberger, B. Emile and H. Laurent, "Optimization-based image segmentation by genetic algorithms", EURASIP Journal on Image and Video Processing, vol. 842029, **(2008)**.

[42] G. U. Maheswari, K. Ramar, D. Manimegalai, V. Gomathi and G. Gowrision, "An adaptive color texture segmentation using similarity measure of symbolic object approach", International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 4, no. 4, **(2011)**.

[43] T. Cour, F. Benezit and J. Shi, "Spectral segmentation with multiscale graph decomposition", in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), **(2005)**.

[44] A. Strehl and J. Ghosh, "Cluster ensembles - a knowledge reuse framework for combining multiple partitions", Journal on Machine Learning Research, JMLR, vol. 3, **(2001)**, pp. 583–617.

[45] A. Fred and A. Jain, "Data clustering using evidence accumulation," in In Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02), August 2002, pp. 276–280.

[46] S. Vega-Pons and J. Ruiz-Shulcloper, "A survey of clustering ensemble algorithms," International Journal of Pattern Recognition and Artificial Intelligence, IJPRAI, vol. 25, no. 3, pp. 337–372, 2011.

[47] D. Martin, C. Fowlkes and J. Malik, "Learning to detect natural image boundaries using local bright-ness, color and texture cues", IEEE Trans. Pattern Anal. Machine Intell., vol. 26, no. 5, **(2004)** May, pp. 530–549.

[48] R. Harrabi and E. B. Braiek, "Color image segmentation using multi-level thresholding approach and data fusion techniques: application in the breast cancer cells images", EURASIP Journal on Image and Video Processing, vol. 11, **(2012)**.

[49] W. M. Rand, "Objective criteria for the evaluation of clustering methods", Journal of the American Statistical Association, vol. 66, no. 336, **(1971)**, pp. 846–850.

[50] S. Ghosh, J. Pfeiffer and J. Mulligan, "A general framework for reconciling multiple weak segmentations of an image", in Proc of the Workshop on Applications of Computer Vision, (WACV'09), Snowbird, Utah, USA, **(2009)** December, pp. 1–8.

[51] P. Wattuya, K. Rothaus, J.-S. Praßni and X. Jiang, "A random walker based approach to combining multiple segmentations," in Proc of the 19th International Conference on Pattern Recognition (ICPR'08), Tampa, Florida, USA, **(2008)** December, pp. 1–4.

[52] A. Alush and J. Goldberger, "Ensemble segmentation using efficient integer linear programming", IEEE Trans. Pattern Anal. Machine Intell., vol. 34, no. 10, **(2012)**, pp. 1966–1977.

[53] M. Mignotte, "A label field fusion Bayesian model and its penalized maximum Rand estimator for image segmentation", IEEE Trans. Image Processing, vol. 19, no. 6, **(2010)**, pp. 1610–1624.

[54] "A label field fusion model with a variation of information estimator for image segmentation," Fusion Information, **(2014)**.

[55] B. Larsen and C. Aone, "Fast and effective text mining using linear time document clustering", Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, **(1999)**, pp. 16–29.

[56] 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'l Conf. Computer Vision, vol. 2, **(2001)** July, pp. 416–423.

[57] R. Unnikrishnan, C. Pantofaru and M. Hebert, "Toward objective evaluation of image segmentation algorithms," IEEE Trans. Pattern Anal. Machine Intell., vol. 29, **(2007)**, pp. 929–944.

[58] P.-M. Jodoin, M. Mignotte and C. Rosenberger, "Segmentation framework based on label field fusion", IEEE Trans. on Image Processing, vol. 16, no. 10, **(2007)** October, pp. 2535–2550,

[59] M. Mignotte, "A segmentation-based regularization term for image deconvolution", IEEE Trans. On Image Processing, vol. 15, no. 7, **(2006)**, pp. 1973–1984.

[60] J. Besag, "On the statistical analysis of dirty pictures", Journal of the Royal Statistical Society, vol. B-48, **(1986)**, pp. 259–302.

[61] S. P. Lloyd, "Least squares quantization in PCM", IEEE Trans. Inform. Theory, vol. 28, no. 2, **(1982)**, pp. 129–136.

[62] M. Mignotte, "A non-stationary MRF model for image segmentation from a soft boundary map", Pattern Analysis and Applications, **(2012)** April.

[63] R. Unnikrishnan, C. Pantofaru and M. Hebert, "A measure for objective evaluation of image segmentation algorithms", in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Workshop on Empirical Evaluation Methods in Computer Vision, vol. 3, **(2005)** June, pp.34–41.

[64] P. Arbelaez, M. Maire, C. Fowlkes and J. Malik, "Contour detection and hierarchical image segmentation", IEEE Trans. Pattern Anal. Machine Intell., vol. 33, no. 5, **(2011)** May, pp. 898–916,

[65] R. Huanga, N. Sangb, D.Luoc and Q. Tangd, "Image segmentation via coherent clustering in l*a*b* color space", Pattern Recognition Letters, vol. 32, no. 7, **(2011)**, pp. 891–902.

[66] E. Sharon, M. Galun, D. Sharon, R. Basri and A. Brandt, "Hierarchy and adaptivity in segmenting visual scenes," Nature, vol. 442, **(2006)**, pp. 810–813.

[67] J. Canny, "A computational approach to edge detection", IEEE Trans. Pattern Anal. Machine Intell., vol. 8, no. 6, **(1986)**, pp. 679–698.

[68] M. Meila, "Comparing clusterings–an information based distance", Journal of Multivariate Analysis, vol. 98, no. 5, **(2007)**, pp. 873–895.

[69] J. Freixenet, X. Munoz, D. Raba, J. Marti and X. Cufi, "Yet another survey on image segmentation: Region and boundary information integration", in Proc. 7th European Conference on Computer Vision ECCV02, **(2002)**, pp. III: 408 ff.

[70] D. Martin and C. Fowlkes, "The Berkeley segmentation database and benchmark", image database and source code publicly available at address http//www.cs.berkeley.edu/projects/vision/grouping/segbench/.

[71] P.-M. Jodoin and M. Mignotte, "Markovian segmentation and parameter estimation on graphics hardware", Journal of Electronic Imaging, vol. 15, no. 3, **(2006)** July-September, pp. 033 015–1–15.