# A Novel Fusion Approach Based on the Global Consistency Criterion to Fusing Multiple Segmentations

Lazhar Khelifi and Max Mignotte

Abstract-In this paper, we introduce a new fusion model whose objective is to fuse multiple region-based segmentation maps to get a final better segmentation result. The suggested new fusion model is based on an energy function originated from the global consistency error (GCE), a perceptual measure which takes into account the inherent multiscale nature of an image segmentation by measuring the level of refinement existing between two spatial partitions. Combined with a region merging/splitting prior, this new energy-based fusion model of label fields allows to define an interesting penalized likelihood estimation procedure based on the GCE criterion with which the fusion of basic, rapidly-computed segmentation results appears as a relevant alternative compared with other (possibly complex) segmentation techniques proposed in the image segmentation field. The performance of our fusion model was evaluated on the Berkeley dataset including various segmentations given by humans (manual ground truth segmentations). The obtained results clearly demonstrate the efficiency of this fusion model.

*Index Terms*—Berkeley image dataset, cluster ensemble algorithm, color textured image segmentation, combination of multiple segmentations, energy-based model, global consistency error (GCE), label field fusion, penalized likelihood model, segmentation ensemble.

#### I. INTRODUCTION

**C** OMBINING multiple, quickly estimated (and eventually poor or weak) segmentation maps of the same image to obtain a final refined segmentation has become a promising approach, over the last few years, to efficiently solve the difficult problem of unsupervised segmentation [1] of textured natural images.

This strategy is considered as a particular case of the cluster ensemble problem. Originally investigated in machine learning,<sup>1</sup> this approach is also known as the concept of

Manuscript received May 26, 2015; revised September 2, 2015 and October 27, 2015; accepted December 10, 2015. Date of publication March 7, 2016; date of current version August 17, 2017. This paper was recommended by Associate Editor M. Celenk.

The authors are with the Department of Computer Science and Operations Research, Faculty of Arts and Sciences, University of Montreal, Montreal, QC H3C 3J7, Canada (e-mail: khelifil@iro.umontreal.ca; 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/TSMC.2016.2531645

<sup>1</sup>The cluster ensemble problem, itself, is derived from the theory of merging classifiers to improve the performance of individual classifier and also known under the name of classifier ensemble problem or ensemble of predictors, committee machine or mixture of expert classifier [6]–[8].

fusing multiple data clusterings for the amelioration of the final clustering result [2]-[5]. Indeed, an inherent feature of images is the spatial ordering of the data and thus, image segmentation is a clustering procedure for grid-indexed data. In this context, the partitioning into regions must consider both closeness in the feature vector space and spatial coherence property of the image pixels. This approach can also be considered as a special case of restoration/denoising procedure in which each rough segmentation (to be combined) is, in fact, assumed to be a noisy observation or solution and the final goal of a fusion model is to obtain a denoised segmentation solution which could be a compromise or a consensus (in terms of contour accuracy, clusters, number of regions, etc.) provided by each input segmentations. Somehow, the final combined segmentation is the average of all the putative segmentations to be fused with respect to a specific criterion. This approach has first been proposed in [9] and [10] with a constraint specifying that all input segmentations (to be fused) must be composed of the same region number. Shortly after, other fusion approaches have been proposed with an arbitrary number of regions in [11] and [12]. Since these pioneering works, this fusion of multiple segmentations<sup>2</sup> of the same scene in order to get a more accurate and reliable result of segmentation (which would be, in some criterion sense, the average of all the individual segmentation) is now implemented according to several strategies or well-defined criteria.

Following this strategy, we can mention the combination model introduced in [11] which fuses the individual putative segmentations according to the within-point scatter of the cluster instances (described in terms of the set of local requantized label histogram produced by each input segmentations), by simply running a K-means-based fusion procedure. By doing so, the author implicitly assumes, in fact, a finite distribution mixture-based fusion model in [13] which the labels assigned to the different regions (given by each input segmentations to be fused), are modeled as random variables distributed according K spherical clusters with an equal volume (or Gaussian distribution [14] with identical covariance matrix) which can be efficiently clustered with

2168-2216 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications\_standards/publications/rights/index.html for more information.

<sup>&</sup>lt;sup>2</sup>This strategy can also be efficiently exploited, more generally, for various other problems involving label maps other than spatial segmentations (e.g., depth field estimation, motion detection or estimation, 3-D reconstruction/segmentation, etc.).

a K-means algorithm. In a similar way, we can also mention the combination model performed in [15] which follows the same idea but for the set of local soft labels (estimated with a multiscale thresholding technique) and for which the fusion operation is thus performed in the sense of the weighted within class/cluster inertia. This fusion of segmentations can also be carried out according to the probabilistic version of the well-known Rand index [13] (PRI) criterion with an energy-based fusion model in order to estimate the segmentation solution with the maximum number of pairs of pixels having a compatible label relationship with the ensemble of segmentations to be fused. This PRI criterion can be minimized either with a stochastic random walking technique [12] (along with an estimator based on mutual information to estimate the optimal region number), or with an algebraic optimization method [16], or with an expectation maximization procedure [17] (combined with integer linear programming and performed on superpixels, initially estimated by a simple over-segmentation) or also in the penalized PRI sense in conjunction with a global constraint on the combination process [18] (constraining the size and the number of segments) with a Bayesian approach relying on a Markovian energy function to be minimized. Combination of segmentation maps can also be performed according to the variation of information (VoI) criterion [19] (by exploiting an energy-based model minimized by applying a pixel-wise gradient descent method strategy under a spatial coherence constraint). Fusion of segmentations can also be achieved in the evidence accumulation sense [4] (and via a hierarchical agglomerative partitioning strategy), or in the F-measure (or precision-recall criterion) sense [20] (and via a hierarchical relaxation scheme fusing the different segments generated in the segmentation ensemble in the final combined segmentation). Finally, we can also mention the fusion scheme proposed in [21] in the optimal or maximum-margin hyperplane (between classes) sense and in which the hyperspectral image is segmented based on the decision fusion of multiple and individual support vector machine classifiers that are trained in different feature subspaces emerging from a single hyperspectral data set or the recent Bayesian [13] fusion procedure for satellite image segmentation proposed in [22]. In addition, we can cite the image segmentation fusion model using general ensemble clustering methods proposed in [23] or the approach presented in [24] based on a consensus clustering algorithm, called filtered stochastic best one element move minimizing a distance function (called symmetric distance function) with a stochastic gradient descent.

The fusion model, introduced in this paper, is based on the global consistency error (GCE) measure. This graph theory-based measure has been designed to directly take into account the following interesting observation: segmentations produced by experts are generally used as a reference or ground truths for benchmarking segmentations performed by various algorithms (especially for natural images). Even though different people propose different segmentations for the same image, the proposed segmentations differ, essentially, only in the local refinement of regions. In spite of these variabilities, these different segmentations should be interpreted as being consistent, considering that they can express the same image segmented at different levels of detail and, to a certain extend, the GCE measure [13] is designed to take into account this inherent multiscale property of any segmentations made by humans. In our fusion model, this GCE measure, which has thus a perceptual and physical meaning, is herein adopted and tested as a new consensus-based likelihood energy function of a fusion model of multiple weak segmentations.

In the remainder of this paper, we first describe the proposed fusion model and the optimization strategy used to minimize the consensus energy function related to this new fusion model in Section II. In Section III, we present the generation of the segmentation ensemble to be combined with our model. Finally, an ensemble of experimental tests and comparisons with existing segmentation approaches is described in Section IV. In this section, our model of segmentation is tested and benchmarked in the Berkeley color image dataset.

### **II. PROPOSED FUSION MODEL**

# A. GCE Measure

There are a lot of (similarity) metrics in the statistic and vision literature for measuring the agreement between two clusterings or segmentation maps. Among others, we can cite [25], [26]; the Jacquard coefficient [27], a variant of the counting pairs also called the Rand index [13] (whose the probabilistic version is the PRI), the Mirkin distance [28], the set matching measures (including the *F*-measure [20], [29], and the purity and inverse purity [30]), and the information theory-based metrics; namely the VoI [19], *V*-measure [31] or kernel-based metrics (graph kernel or subset significance [33]-based measures [32]) or finally the popular Cohen's kappa [34], [35] measure.

In our fusion model, we use the GCE [36] criterion which (is the only one, to our knowledge that) measures the extent to which one segmentation map can be viewed as a refinement of another segmentation. In this metric sense, a perfect correspondence is obtained if each region in one of the segmentation is a subset (i.e., a refinement) or geometrically similar to a region in the other segmentation. Segmentations with similar GCE can be interpreted as being consistent, inasmuch as they could express the same natural image segmented at a different degree of detail, as it is the case of the segmented images generated by different human observers for which a finer level of detail will be (possibly) merged by another observer in order to give the larger regions of a segmentation thus estimated at a coarser level.

This GCE distance can be exploited as a segmentation measure to evaluate the correspondence of a segmentation machine with a ground truth segmentation. To this end, it was recently proposed in image segmentation [37], [38] as a quantitative and perceptually interesting metric to compare machine segmentations of an image dataset to their respective manually segmented images given by human experts (i.e., a ground truth segmentations) and/or to objectively measure and rank (based on this GCE criterion) the efficiency of different automatic segmentation algorithms.<sup>3</sup>

Let  $S^t = \{C_1^t, C_2^t, \dots, C_{R^t}^t\}$ ,  $S^g = \{C_1^g, C_2^g, \dots, C_{R^g}^g\}$ ,  $R^t$ , and  $R^g$  be, respectively, the segmentation result, the manually segmented image, and the number of regions<sup>4</sup> in  $S^t$  and in  $S^g$ . We consider, for a particular pixel  $p_i$ , the segments in  $S^t$  and  $S^g$  including this pixel. We denote these segments by  $C_{< p_i>}^t$ and  $C_{< p_i>}^g$ , respectively. If one segment is a subset of the other, so the pixel is practically included in the refinement area, and the local error should be equal to zero. If there is no subset relationship, then the two regions overlap in an inconsistent way and the local error ought be different from zero [36]. The local refinement error (LRE) is therefore denoted at pixel  $p_i$  as

$$LRE(S^{t}, S^{g}, p_{i}) = \frac{\left|C^{t}_{\langle p_{i} \rangle} \setminus C^{g}_{\langle p_{i} \rangle}\right|}{|C^{t}_{\langle p_{i} \rangle}|}$$
(1)

where  $\setminus$  represents the set differencing operator and |C| the cardinality of the set of pixels *C*. As noticed in [36], this clustering (or segmentation) error measure is not symmetric and encodes a measure of refinement in only one sense. LRE( $S^t$ ,  $S^g$ ,  $p_i$ ) is equal to 0 specifically if  $S^t$  is a refinement of  $S^g$  at pixel  $p_i$ , but not vice-versa. A possible and natural way to combine the LRE at each pixel into a measure for the whole image is the so-called GCE which constraints all local refinement to be in the same sense in the following way:

$$GCE(S^{t}, S^{g}) = \frac{1}{n} \min\left\{\sum_{i=1}^{n} LRE(S^{t}, S^{g}, p_{i}), \sum_{i=1}^{n} LRE(S^{g}, S^{t}, p_{i})\right\}$$
(2)

where *n* is the pixels number  $p_i$  within the image. This segmentation error, based on the GCE, is a metric whose values belong to the interval [0, 1]. A measure of 0 expressed that there is a perfect match between the two segmentations (identical segmentations) and an error of 1 represents a maximum difference between the two segmentations to be compared.

Although a fundamental problem with the GCE measure is that there are two bad, unrealistic segmentation types (i.e., degenerate segmentations) giving an unusually high score value (i.e., a zero error for GCE) [36]. These two degenerative segmentations are the two following trivial cases; one pixel per region (or segment) and one region per the whole image. The former is, in fact, a detailed improvement (i.e., refinement) of any segmentation, and any segmentation is a refined improvement of the latter. This illustrates why, the GCE measure is useful only when comparing two segmentation maps with an equal number of regions.

In our application, in order to be able to define an energybased fusion model, avoiding the two above-mentioned degenerate segmentation cases, and for which a reliable consensus

<sup>3</sup>In addition, as the semantic gap is generally considered as a difference between low-level segmentation (i.e., labeling decision based on a machine by using pixel information) and high-level segmentation (i.e., based on the human expert's labeling decision), the use of the GCE-based perceptually metric also leads to objectively measure and rank the semantic gap width as well.

<sup>4</sup>A region is a set of connected pixels grouped into the same class and a class, a set of pixels possessing similar textural characteristics.

or compromise resulting segmentation map would be solution, via an optimization scheme (see Section II-B), we have replaced the minimum operator in the GCE by the average operator

$$GCE^{\star}(S^{t}, S^{g}) = \frac{1}{2n} \left\{ \sum_{i=1}^{n} LRE(S^{t}, S^{g}, p_{i}) + \sum_{i=1}^{n} LRE(S^{g}, S^{t}, p_{i}) \right\}.$$
 (3)

This new measure is slightly different, while being a tougher measure than the usual and classical GCE measure since GCE<sup>\*</sup> is always greater than GCE for any automatic segmentation relatively to a given ground truth  $S^{g,5}$ 

The performance score, based on the GCE measure, was also lately used in the segmentation of natural image [40] as a score to compare an unsupervised image segmentation given by an algorithm to an ensemble of ground truth segmentations provided by human experts. This ensemble of slightly different ground truth partitions, given by experts, represents, in essence, the multiple acceptable ground truth segmentations related to each natural image and reflecting the inherent variation of possible (detailed) interpretations (of an image) between each human segmenter. Recently, this variation among human observers, modeled by the Berkeley segmentation database (BSD) [36], comes from the fact that each human generates a segmentation (of a given image) at different levels of detail. These variations highlight also the fact that the image segmentation is inherently an ill-posed problem in which there are different values of the number of classes for the set of more or less detailed segmentations of a given image. Let us finally mention that, as already said, the GCE metric is a measure tolerant to this intrinsic variability between possible interpretations of an image by different human observers. Indeed, this variability is often due to the refinement between human segmentations represented at different levels of image detail, abstraction or resolution. Thus, in the presence of a set of various human segmentations (showing, in fact, a small fraction of all possible perceptually consistent spatial partitions of an image content [41]), this measure of segmentation quality, based on GCE criterion, has to quantify the degree of similarity between an automatic image segmentation (i.e., performed by an algorithm) and this set of possible ground truths. As proposed in [37], this variability can simply be taken into account by estimating the mean GCE value. More precisely, let us assume a set of L manually segmented images  $\{S_k^g\}_{k \leq L} = \{S_1^g, S_2^g, \dots, S_L^g\}$  related to a same scene. Let  $S^t$  be the segmentation to be compared to the manually labeled set,

$$BCE(S^t, S^g) = \frac{1}{n} \sum_{i} \max \{ LRE(S^t, S^g, p_i), LRE(S^g, S^t, p_i) \}$$

in which the problem of degenerate segmentations "cheating" a benchmark also disappears. Nevertheless, this measure does not tolerate refinement at all (more precisely, BCE is a measure that penalizes dissimilarity between segmentations proportional to the degree of region overlap) contrary to our GCE\* measure which tolerates, to a certain extent, a refinement between two segmentations (i.e., which considers, as consistent, two segmentations with a certain different degree of detail).

<sup>&</sup>lt;sup>5</sup>An alternative to avoid the above-mentioned degenerate segmentation cases was also proposed in [39] with the so-called bidirectional consistency error (BCE)

the mean GCE measure is thus given by

$$\overline{\text{GCE}}\left(S^{t}, \left\{S_{k}^{g}\right\}_{k \le L}\right\}\right) = \frac{1}{L} \sum_{k=1}^{L} \text{GCE}\left(S^{t}, S_{k}^{g}\right)$$
(4)

and equivalently, we can define

$$\overline{\operatorname{GCE}^{\star}}\left(S^{t},\left\{S^{g}_{k}\right\}_{k\leq L}\right) = \frac{1}{L}\sum_{k=1}^{L}\operatorname{GCE}^{\star}\left(S^{t},S^{g}_{k}\right).$$
(5)

For example, this  $\overline{\text{GCE}}$  measure will return a high score (i.e., a low value) for an automatic segmentation  $S^t$  which is homogeneous, in the sense of this criterion, with most of the ground truth segmentations provided by human segmenters.

#### B. Penalized Likelihood-Based Fusion Model

Let us assume now that we have an ensemble of L (different) segmentations  $\{S_k\}_{k \le L} = \{S_1, S_2, \dots, S_L\}$  (of the same scene) to be combined in the goal of providing a final improved segmentation result  $\hat{S}$  (i.e., more accurate than the individual member of  $\{S_k\}_{k \le L}$ ). To this end, a classic strategy for finding a segmentation result  $\hat{S}$ , which would be a consensus or compromise of  $\{S_k\}_{k < L}$ , or equivalently, a strategy for combining/fusing these L individual segmentations, consists in designing an energy-based model generating a segmentation solution which is as close as possible (with the GCE<sup>\*</sup> considered distance) to all the other segmentations or, equivalently, a likelihood estimation model of  $\hat{S}$ , in the minimum  $\overline{\text{GCE}}$ distance sense [or according to the maximum likelihood (ML) principle for this  $\overline{\text{GCE}}^*$  criterion], since this measure, contrary to the GCE measure is not degenerate. This optimization-based approach is sometimes referred to as the median partition [5] with respect to both segmentation ensemble  $\{S_k\}_{k \le L}$  and GCE criterion. In this framework, if  $S_n$  designates the set of all possible segmentations using n pixels, the consensus segmentation (to be estimated in the  $\overline{\text{GCE}}^{\star}$  criterion sense) is then straightforwardly defined as the minimizer of the GCE' function

$$\hat{S}_{\overline{\text{GCE}}^{\star}} = \arg\min_{S \in \mathcal{S}_n} \overline{\text{GCE}}^{\star} (S, \{S_k\}_{k \le L}).$$
(6)

However, the problem of image segmentation remains an ill-posed problem providing different solutions for multiple possible values of regions number (of the final fused segmentation and/or of each segmentation to be fused) and which is a priori unknown. To make this problem a well-posed problem characterized by a unique solution, it is essential to add some constraints on the segmentation process, favoring merging regions or conversely, an over-segmentation. From the probabilistic standpoint, these regularization constraints could be defined via a prior distribution on the segmentation solution  $\hat{S}_{\overline{\text{GCE}}^{\star}}$ . Analytically, this requires to recast our likelihood estimation problem of the consensus segmentation in the penalized likelihood framework by adding, to the simple ML fusion model [see (6)], a regularization term, allowing to integrate knowledge about the types of resulting fused segmentation, a priori considered as acceptable solutions. In our case, we search to estimate a resulting segmentation map providing a reasonable number of segments or regions. In our framework, this property, regarding the types of segmentation maps that we would like to favor, can be efficiently modeled and controlled via a region merging or splitting regularization term related to the different (connected) region area of the resulting consensus segmentation map. In this optic, an interesting global prior, derived from the information theory, is the following region-based regularization term:

$$E_{\operatorname{Reg}}\left(S = \{C_k\}_{k \le R}\right) = \left|-\sum_{k=1}^{R} \left[\frac{|C_k|}{n} \log \frac{|C_k|}{n}\right] - \overline{\mathcal{R}}\right| \quad (7)$$

where we remind that R denotes the region number (or segments) in the segmentation map S, n, and  $|C_k|$  are, respectively, the pixel number within the image and the pixel number in the kth region  $C_k$  of the segmentation map S (i.e., the area, in terms of pixel number, of the region  $C_k$ ).  $\mathcal{R}$  is an internal parameter of our regularization term that defines the mean entropy of the *a priori* defined acceptable segmentation solutions. This penalty term favors merging (i.e., leads to a decrease of the penalty energy term) if the current segmentation solution has an entropy greater than  $\mathcal{R}$  (i.e., in the case of an oversegmentation) and favors splitting in the contrary case. Contrary to the regularization term defined in [18], this one takes into account both region number of the resulting segmentation solution, but also the proportion of these regions. In image segmentation, this information theoretic regularization term (without the absolute value and with  $\overline{\mathcal{R}} = 0$ ) has been used first to restrict the number of clusters of the classical objective function of the fuzzy K-means clustering procedure [42] (i.e., the class number of the segmentation problem) in [43] and later, to efficiently restrict the number of regions of an objective function in a level set segmentation framework [44]. Finally, with this regularization term, a penalized likelihood solution of our fusion model is thus given by

$$\hat{S}_{\overline{\text{GCE}}^{\star}_{\beta}} = \arg\min_{S\in\mathcal{S}_{n}} \left\{ \overline{\text{GCE}}^{\star} (S, \{S_{k}\}_{k\leq L}) + \beta E_{\text{Reg}}(S) \right\}$$
$$= \arg\min_{S\in\mathcal{S}_{n}} \overline{\text{GCE}}^{\star}_{\beta} (S, \{S_{k}\}_{k\leq L})$$
(8)

with  $\beta$  allowing to weight the related contribution of the region splitting/merging argument in our energy-based fusion model.

It is also noteworthy to mention that the region splitting/merging regularization term remains essential in some relatively rare cases in which the segmentation solution may lead to a  $\overline{\text{GCE}}^*$  measure which is minimal in the trivial one region segmentation case. The penalized likelihood approach allows to avoid these (relatively rare) situations. In addition and consequently, this penalized likelihood approach allows also to exploit the original  $\overline{\text{GCE}}$  measure with the minimum operator [see (2)]. A comparison of efficiency between these two error metrics, in our fusion-based segmentation application, will be discussed later, in the experimental results section.

## C. Optimization of the Fusion Model

Our fusion model of multiple label fields, based on the penalized  $\overline{\text{GCE}}^{\star}$  criterion, is therefore formulated as a global optimization problem involving a nonlinear objective function characterized by a huge number of local optima across



Fig. 1. Examples of initial segmentation ensemble and fusion results (Algorithm 1). First three rows: results of *K*-means clustering for the segmentation model presented in Section III. Fourth row: input image chosen from the Berkeley image dataset and final segmentation given by our fusion framework.

the lattice of possible clusterings  $S_n$ . In our case, this optimization problem is difficult to solve, mainly because (among other things) we are not able to express (for this  $\overline{\text{GCE}}^{\star}$  criterion) the local decrease in the energy function for a new label assignment at pixel  $p_i$ , and consequently, we cannot adopt the pixel-wise optimization strategy described in [19] in which a simple Gauss-Seidel type algorithm is exploited. This aforementioned Gauss-Seidel type algorithm is, in fact, a deterministic relaxation scheme or an approximate gradient descent where any pixel of the consensus segmentation to be classified are updated one at a time (by searching the minimum local energy label assignment also called the mode). Nevertheless, in our case, we can adopt the general optimization strategy proposed in [20], in which the strategy of optimization is based on the ensemble of superpixels belonging in  $\{S_k\}_{k \le L}$ , i.e., the segments ensemble or regions provided by each individual segmentations to be fused. This approach has other crucial advantages. First, by considering this set of superpixels as the atomic elements to be segmented in the consensus segmentation (instead of the set of pixels), we considerably decrease the computational complexity of the consensus segmentation process. Second, it is also quite reasonable to think that, if individually, each segmentation (to be fused) might give some poor results of segmentation for some subparts of the image (i.e., bad regions or superpixels) and also conversely good segmented regions (or superpixels) for other subparts of the image, the superpixel ensemble created from  $\{S_k\}_{k \le L}$  is likely to contain the different individual pieces of regions or right segments belonging to the optimal consensus segmentation solution. In this semi-local optimization strategy, the relaxation scheme is based on a variant of the iterative conditional modes (ICMs) [45], i.e., a Gauss–Seidel type process (see Algorithm 1 for more details) which iteratively optimizes

only one superpixel (in our strategy) at a time without considering the effect on other superpixels (until convergence is achieved). On the one hand, this iterative search algorithm is simple and deterministic, however, on the other hand, the main drawback of this technique is to strongly depend on the initialization step, which should be not too far from the ideal solution (in order to prevent the ICM from getting stuck in a local minima far from the global one). To this end, we can take, as initialization, the segmentation map  $\hat{S}_{\overline{\text{GCE}}}^{\star[0]}$  defined as follows:

$$\hat{S}_{\overline{\text{GCE}}_{\beta}^{\star}}^{[0]} = \arg\min_{S \in \{S_k\}_{k \le L}} \overline{\text{GCE}}_{\beta}^{\star} \left(S, \{S_k\}_{k \le L}\right)$$
(9)

i.e., from the *L* segmentation to be combined, we can select the one ensuring the minimal consensus energy (in the  $\overline{\text{GCE}}_{\beta}^{\star}$ sense) of our fusion model. This segmentation will be considered as the first iteration of our penalized likelihood model (8).<sup>6</sup> This iterative algorithm attempts to obtain, for each superpixel to be classified, the minimum energy label assignment. More precisely, it begins with an initialization  $\overline{\text{GCE}}_{\beta}^{\star}$  not far to the optimal segmentation [see (9)], and for each iteration and each atomic region (superpixel), ICMs assigns the label giving the largest decrease of the energy function (to be minimized). We summarize in Algorithm 1, the overall penalized GCE-based fusion model (GCEBFM) algorithm based on the ICM procedure and superpixel set.

<sup>&</sup>lt;sup>6</sup>Another efficient approach consists in running the ICM procedure, independently, with the first  $N_I$  optimal input segmentations extracted from the segmentation ensemble (in the  $\overline{\text{GCE}}_{\beta}^{\star}$  sense) as initialization, and to select, once convergence is achieved, the result of segmentation associated with the lowest  $\overline{\text{GCE}}_{\beta}^{\star}$  energy. This strategy will improve slightly the performance of our combination model, but will increase the computational cost.

#### Algorithm 1: Penalized GCE-Based Fusion Algorithm

$ \begin{array}{c} \overline{\operatorname{GCE}}^{\star}_{\beta} \\ \{S_k\}_{k \leq L} \\ \{b_k\} \\ \{\mathcal{E}_k\} \end{array} $	Penalized mean GCE [see (8)] Set of $L$ segmentations to be fused Set of superpixels $\in \{S_k\}_{k \leq L}$ Set of region labels in $\{S_k\}_{k \leq L}$
$\begin{array}{c} T_{\max} \\ \beta \end{array}$	Maximal number of iterations $(=8)$ Regularization parameter

1. Initialization

• 
$$\hat{S}_{\overline{\text{GCE}}_{\beta}^{\star}}^{[0]} = \arg\min_{S \in \{S_k\}_{k \leq L}} \overline{\text{GCE}}_{\beta}^{\star} \left(S, \{S_k\}_{k \leq L}\right)$$

#### 2. Steepest Local Energy Descent

while  $p < T_{max}$  do

for each 
$$b_k$$
 superpixel  $\in \{S_k\}_{k \le L}$  do  
• Draw a new label  $x$  according to the  
uniform distribution in the set  $\{\mathcal{E}_k\}$   
• Let  $\hat{S}_{\text{GCE}_{\beta}^{*}}^{[p],\text{new}}$  the new segmentation map  
including  $b_k$  with the region label  $x$   
• Compute  $\overline{\text{GCE}}_{\beta}^{\star}(S, \{S_k\}_{k \le L})$  on  $\hat{S}_{\text{GCE}_{\beta}^{*}}^{[p],\text{new}}$   
if  $\overline{GCE}_{\beta}^{\star}(\hat{S}_{\overline{GCE}_{\beta}^{*}}^{[p],\text{new}}) < \overline{GCE}_{\beta}^{\star}(\hat{S}_{\overline{GCE}_{\beta}^{*}}^{[p]})$  then  
 $\downarrow \quad \overline{\text{GCE}}_{\beta}^{\star} = \overline{\text{GCE}}_{\beta}^{\star,\text{new}}$   
 $\downarrow \quad \hat{S}_{\overline{\text{GCE}}_{\beta}^{\star}}^{[p],\text{new}} = \hat{S}_{\overline{\text{GCE}}_{\beta}^{\star}}^{[p],\text{new}}$   
 $p \leftarrow p + 1$ 

### **III. GENERATION OF THE SEGMENTATION ENSEMBLE**

The initial ensemble of segmentations, which will be combined via our fusion model, is rapidly generated, in our case, through the standard K-means method [46] associated with 12 different color spaces in order to ensure variability in the segmentation ensemble, those are, YCbCr, TSL, YIQ, XYZ, h123, P1P2, HSL, LAB, RGB, HSV, i123, and LUV (in [18] more explanations are given on the choice of these color spaces). Also, for the class number K of the K-means, we resort to a metric measuring the complexity relative to each input image, in terms of number of the different texture type present in the natural color image. This metric, presented in [47], is in fact the measure of the absolute deviation  $(L_1 \text{ norm})$  of the ensemble of normalized histograms obtained for each overlapping squared fixed-size  $(N_w)$  neighborhood included within the image. This measure ranges in [0, 1] and an image with different textured regions will provide a complexity value close to 1 (and conversely, a value close to 0 when the image is characterized by few texture types). In our framework

$$K = \text{floor}\left(\frac{1}{2} + \left[K^{\max} \times \text{complexity value}\right]\right) \quad (10)$$

where floor(*x*) is a function that gives the largest integer less than or equal to *x* and  $K^{\max}$  is an upper-bound of the number of classes for a very complex natural image. It is noteworthy to mention that, in our application, we use three different values of  $K^{\max}$  ( $K_1^{\max} = 11$ ,  $K_2^{\max} = 9$ , and  $K_3^{\max} = 3$ ) once again, in order to ensure variability in the segmentation ensemble. In addition, as input multidimensional descriptor of feature, we exploited the ensemble of values (estimated around the pixel to be labeled) of the requantized histogram (with equal bins in each color channel). In our framework, this local histogram is requantized, for each color channels, in a  $N_b = q_b^3$ bin descriptor, estimated on an overlapping, fixed-size squared  $(N_w = 7)$  neighborhood centered around the pixel to be classified with three different seeds for the *K*-means algorithm and with two different values of  $q_b$ , namely  $q_b = 5$  and  $q_b = 4$ . In all, the number of input segmentations, to be combined, is  $60 = 12 \times (3 + 2)$ .<sup>7</sup>

#### **IV. EXPERIMENTAL RESULTS**

#### A. Initial Tests Setup

In all the tests, the evaluation of our fusion scheme [see (8)] is presented for an ensemble of L = 60 segmentations  $\{S_k\}_{k \le L}$  with spatial partitions generated with the simple *K*-means-based segmentation technique introduced in Section III (see Fig. 1). Moreover, for these initial experiments, we have fixed,  $\overline{\mathcal{R}} = 4$ , 2 and  $\beta = 0$ , 01 [see (7) and (8)]. The justification of these internal parameter values (for the fusion algorithm) will be detailed in Section IV-B.

First of all, we have tested the convergence properties of our iterative optimization procedure based on superpixel by choosing, as initialization of our iterative local gradient descent algorithm, various initializations (extracted from our segmentation ensemble  $\{S_k\}_{k \le L}$  and one noninformative (or blind) initialization by creating an image exhibiting K horizontal and identical rectangular regions, thus with K various region labels (see Figs. 2 and 3). Before all, we can notice that our proposed optimization procedure shows good convergence properties in its ability to achieve the optimization of our consensus function of energy. Indeed, the consensus energy function is perhaps not purely convex (three somewhat different solutions are obtained), nevertheless, the obtained final solutions (after 8 iterations) remain very similar. In addition, the final  $GCE_{\beta}$ score along with the resulting final segmentation map, is on average, all the better than the initial segmentation solution is associated to a good initial  $\overline{\text{GCE}}_{\beta}^{\star}$  score (while remaining robust when the initialization is not reliable). Consequently, the combination of the use the superpixels of  $\{S_k\}_{k \leq L}$  along with a good initialization strategy [see (9)] definitely gives good convergence properties to our fusion model. Second, we have tested the influence of parameter  $\mathcal{R}$  [see (7)] on the generated solutions of segmentation. Fig. 4 indicates unambiguously that  $\mathcal{R}$  can be clearly interpreted as a regularization

<sup>&</sup>lt;sup>7</sup>This process aims to ensure the diversity needed to achieve a reliable (i.e., good) set of putative segmentation maps on which the final result will depend. This diversity is crucial to guarantee the availability of more (reliable) information for the consensus function (on which the model of fusion is defined) [5], [18]. The use of different segmentations associated with the same scene, expressed in diverse spaces of color, is (somewhat) equivalent to observing the scene with several sensors or cameras with different characteristics [22], [48] and also a necessary condition for which the fusion model can be efficiently carried out. On the other hand, it is easy to understand that the fusion of similar solutions of segmentation. The time of execution, related to each segmentation achieved by this simple K-means technique is rapid (less than 1 s) for a nonoptimized sequential program in C++.



Fig. 2. Example of fusion convergence result on three various initializations for the Berkeley image (n<sup>0</sup>187039). Left: initialization. Right: segmentation result after eight iterations of our GCEBFM fusion model. From top to bottom: the original image, the two input segmentations (from the segmentation set) which have the best and the worst  $\overline{\text{GCE}}_{\beta}^{\star}$  value, and one noninformative (or blind) initialization.



Fig. 3. Progression of the segmentation result (from lexicographic order) during the iterations of the relaxation process beginning with a noninformative (blind) initialization.

parameter of the final number of regions of our combination scheme; favoring under-segmentation, for low values of  $\overline{\mathcal{R}}$  (and consequently penalizing small regions) or splitting, for great values of  $\overline{\mathcal{R}}$ . To further test the regularization role of  $\overline{\mathcal{R}}$  in our fusion model, we have also plotted in Fig. 6, the average regions number for each image of the BSD300 as a function of the value of  $\overline{\mathcal{R}}$ . In our case, the value for  $\overline{\mathcal{R}} = 4, 2$  (see Section IV-B) allows to obtain 23 regions, on average, on the BSD300. It is worth recalling that the





Fig. 4. Example of segmentation solutions generated for different values of  $\overline{\mathcal{R}}$  ( $\beta = 0.01$ ), from top to bottom and left to right,  $\overline{\mathcal{R}} = \{1.2, 2.2, 3.2, 4.2\}$ , respectively, segmentation map results with 4, 12, 20, and 22 regions.



Fig. 5. Example of fusion result using, respectively, L = 5, 10, 30, and 60 input segmentations (i.e., 1, 2, 6, and 12 color spaces). We can also compare the segmentation results with the segmentation maps given by a simple *K*-means algorithm (see examples of segmentation maps in the segmentation ensemble in Fig. 1).



Fig. 6. Plot of the average number of different regions obtained for each segmentation (of the BSD300) as a function of the value of  $\overline{\mathcal{R}}$ .

average regions number belonging to the set of human segmentation ensemble of the BSD300 is around this value (see [37]).

#### TABLE I

AVERAGE PERFORMANCE, RELATED TO THE PRI METRIC, OF SEVERAL REGION-BASED SEGMENTATION ALGORITHMS (WITH OR WITHOUT A FUSION MODEL STRATEGY) ON THE BSD300, RANKED IN THE DESCENDING ORDER OF THEIR PRI SCORE (THE HIGHER VALUE IS THE BETTER) AND CONSIDERING ONLY THE (PUBLISHED) SEGMENTATION METHODS WITH A PRI SCORE ABOVE 0.75 [11], [16]–[20], [37], [38], [47], [50]–[68]

	ALGORITHMS	PRI [49]
	-HUMANS- (in [37])	0,87
el	-GCEBFM-	0,80
ision Mod	(2014) -VOIBFM- [19]	0,81
	(2014) -FMBFM- [20]	0,80
	(2010) -PRIF- [18]	0,80
ı Fu	(2012) -SFSBM- [47]	0,79
With	(2008) -FCR- [11]	0,79
	(2009) -Consensus- [16]	0,78
	(2007) -CTM- [37],[38]	0,76
	(2012) -MDSCCT- [50]	0,81
	(2011) -gPb-owt-ucm- $[51]$	0,81
	(2012) -AMUS [17]	0,80
	(2009) -MIS- $[53]$	0,80
	(2011) -SCKM- [54]	0,80
del	(2008) -CTex- [52]	0,80
Mo	(2004) -FH- $[55]$ (in [37])	0,78
on	(2011) -MD2S- $[56]$	0,78
lusi	(2009) -HMC- $[57]$	0,78
ıt F	(2009) -Total Var- [58]	0,78
chot	(2009) -A-IFS HRI- [59]	0,77
Wit	(2001) -JSEG- $[60]$ (in [52])	0,77
	(2011) -KM- $[61]$	0,76
	(2006) -Av. Diss- $[62]$ (in [51])	0,76
	(2011) -SCL- $[63]$	0,76
	(2005) -Mscuts- $[64]$ (in [58])	0,76
	(2003) -Mean-Shift- $[65]$ (in [37])	0,75
	(2008) -NTP- [66]	0,75
	(2010) -iHMRF- [67]	0,75
	(2005) -NCuts- $[64]$ (in [51])	0,75
	(2006) -SWA- $[68]$ (in [51])	0,75

#### B. Performances and Comparison

In this section, we have benchmarked our model of fusion as algorithm of segmentation on the BSD300 [36] (with images normalized to have the longest side equal to 320 pixels). The segmentation results are then super-sampled in order to obtain segmentation images with the original resolution ( $481 \times 321$ ) before the estimation of the performance metrics.

To this end, several performance measures computed on the full image dataset will be indicated for a fair comparison with the other state-of-the-art segmenters proposed in the literature. These measures of performance include first and foremost the PRI [49] score, which seems to be among the most correlated (in term of visual perception) with manual segmentations [37] and which is generally exploited for segmentations based on region. This PRI score computes the percentage of pairs of pixel labels perfectly labeled in the result of segmentation and a value equal to PRI = 0.75 means that, on average, 75% of pairs of pixel labels are correctly labeled (on average) in the results of segmentation on the BSD300.

#### TABLE II

Average Performance of Diverse Region-Based Segmentation Algorithms (With or Without a Fusion Model Strategy) for Three Different Performances (Distance) Measures (the Lower Value Is the Better) on the BSD300 [11], [17]–[19], [37], [38], [50], [54]–[56], [65], [69]

	ALGORITHMS	VoI	GCE	BDE
-	-HUMANS-	1,10	0,08	4,99
With Fusion Model	-GCEBFM-	2,10	0,19	8,73
	-VOIBFM- [19]	1,88	$0,\!20$	9,30
	-FCR- [11]	$^{2,30}$	$0,\!21$	8,99
	-CTM- [37,38]	2,02	$0,\!19$	9,90
	-PRIF- [18]	1,97	$0,\!21$	8,45
out Fusion Model	-MDSCCT- [50]	2,00	$0,\!20$	$7,\!95$
	-SCKM- [54]	$2,\!11$	$0,\!23$	10,09
	-MD2S- [56]	2.36	0.23	$10,\!37$
	-Mean-Shift- [65] (in [37])	2,48	$0,\!26$	9,70
	-NCuts- [69] (in [37])	2,93	$0,\!22$	9,60
	-FH- [55] (in [37])	2,66	$0,\!19$	9,95
Witl	-AMUS- [17]	1,68	$0,\!17$	-

To guarantee the integrity of the benchmark results, the two control parameters of our algorithm of segmentation [i.e.,  $\overline{\mathcal{R}}$ and  $\beta$ , see (7) and (8)] are optimized on the ensemble of training images by using a local search procedure (with a fixed step-size) on a discrete grid, on the (hyper)parameter space and in the feasible ranges of parameter values ( $\beta \in [10^{-3} : 10^{-1}]$ [step-size =  $10^{-3}$ ] and  $\overline{\mathcal{R}} \in [3:6]$  [step-size = 0.2]. We have found that  $\overline{\mathcal{R}} = 4, 2$  and  $\beta = 10^{-2}$  are reliable hyperparameters for the model yielding interesting 0, 80 PRI value (see Table I).

For a fair comparison, we now present the results of our fusion model by displaying the same segmented images (see Figs. 10 and 11) as those presented in the model of fusion introduced in [18] and [19]. The results concerning the whole dataset are accessible on-line via this link: "http://www.etud.iro.umontreal.ca/~khelifil/ResearchMaterial/gcebfm.html."

In order to ensure an effective comparison with other segmentation methods we have also used the VoI measure [70], the GCE [36], and the boundary displacement error (BDE) [71] (this metric measures the average displacement error of boundary pixels between two segmented images, especially, it defines the error of one boundary pixel as the Euclidean distance between the pixel and the closest pixel in the other boundary image) (see Table II, the lower distance is better). The results show that our method provides a competitive result for some other metrics based on different criteria and comparatively to state-of-the-arts.

In addition, and as it has been proposed in Section II-B, we have used our penalized likelihood approach with the original  $\overline{\text{GCE}}$  consensus energy function, with the minimum operator, [i.e., by using (2) instead of (3)] and tuned the internal parameters of our segmentation model, noted  $\beta$  and  $\overline{\mathcal{R}}$  on the ensemble of training images via a local search approach on a



Fig. 7. Distribution of the PRI metric, the number and the size of regions over the 300 segmented images of the Berkeley image dataset.

discrete grid. We have found that  $\overline{\mathcal{R}} = 4.2$  and  $\beta = 0.0375$  are optimal hyper-parameters giving the following performance measures; PRI = 0.78, VoI = 2.22, GCE = 0.20, and BDE = 10.43, significantly less better than our GCE\*-based fusion model.

#### C. Discussion

As we can notice, our fusion model of simple, rapidly estimated segmentation results is very competitive for different kinds of performance measures and can be regarded as a robust alternative to complex, computationally demanding segmentation models existing in the literature.

We have compared our segmentation algorithm (called GCEBFM) against several unsupervised algorithms. From Table II, we can conclude that our method performs overall better than the others for different and complementary performance measures and especially for the PRI measure (which is important because this measure is highly correlated with human hand segmentations) and with the GCE measure which is closely related to the classification error via



Fig. 8. From lexicographic order, progression of the PRI (the higher value is better) and VoI, GCE, and BDE metrics (the lower value is better) according to the segmentations number (*L*) to be fused for our GCEBFM algorithm. Precisely, for L = 1, 5, ..., 60 segmentations [by considering first, one *K*-means segmentation (according to the RGB color space) and then by considering five segmentation for each color space and 1, 2, ..., 12 color spaces].



Fig. 9. First row: three natural images from the BSD300. Second row: the result of segmentation provided by the MDSCCT algorithm. Third row: the result of segmentation obtained by our algorithm GCEBFM.

the computation of the overlap degree between two segmentations (and this good performance is also due to our fusion model which is based on this specific criterion). Statistics on the segmentation results of our method (e.g., the distribution of the PRI, the distribution of the number of regions and size



Fig. 10. Example of segmentations obtained by our algorithm GCEBFM on several images of the Berkeley image dataset (see also Tables I and II for quantitative performance measures and http://www.etud.iro.umontreal.ca/~khelifil/ResearchMaterial/gcebfm.html for the segmentation results on the entire dataset).

of the regions of the segmented Berkeley database images), for our algorithm are given in Fig. 7. These statistics show us that the average number of regions, estimated by our algorithm, is close to the average value given by humans (24 regions) and the PRI distribution shows us that few segmentation exhibits a bad PRI score even for the most difficult segmentation cases.



Fig. 11. Example of segmentations obtained by our algorithm GCEBFM on several images of the Berkeley image dataset (see also Tables I and II for quantitative performance measures and http://www.etud.iro.umontreal.ca/~khelifil/ResearchMaterial/gcebfm.html for the segmentation results on the entire dataset).

Moreover, we can observe (see Figs. 5 and 8) that the PRI, VoI, BDE, and GCE performance scores are better when L (the segmentation number to be merged) is high. This test

shows the validity and the potentiality of our fusion procedure and demonstrates also that our performance scores are perfectible if the segmentation ensemble is completed by other

TABLE III Comparison of Scores Between the GCEBFM and the MDSCCT Algorithms on the 300 Images of the BSDS300. Each Value Points Out the Number of Images of the BSDS300 That Obtain the Best Score

ALGORITHMS MEASURES	GCEBFM	MDSCCT
-GCE-	121	179
-VOI-	166	134
-BDE-	120	180
-PRI-	132	168

(and complementary or different) segmentation maps (of the same image).

The experimental results (see Table II) show that our fusion model outperforms all other fusion approaches in term of GCE measure. On the one hand, this result is driven by the fact that our model is based on an energy function originating from the GCE. On the other hand, with the addition of an efficient entropy-based regularization term (see Section II-B), our model can accurately (and adaptively) estimate a resulting segmentation map with the optimal number of regions, thus yielding a good similarity score between our segmentation results and the ground truth segmentations (given by humane expert). Also, in terms of BDE score, our model outperforms all other fusion approaches excepted the PRIF model. Let us finally add that our model gives a good compromise (comparing to other methods) between all complementary performance measures mentioned in Tables I and II (this last point is important, since this good compromise between relevant complementary performance indicators, is also a clear indication of the quality of segmentations produced by our algorithm).

The PRI, VoI, BDE, and GCE measures are quite different for a given image compared to the measures obtained by other approaches like the multidimensional scaling-based segmentation model for the fusion of contour and texture cues (MDSCCT) algorithm (see Table III and Fig. 9). It means that these two methods perform differently and well for different images. This is not surprising since these two methods are, by nature, very different from each other (the MDSCCT is a purely algorithmic approach, on the contrary, our GCEBFM algorithm is a fusion model whose objective is to combine different region-based segmentation maps). This fact may suggest that these two methods extract complementary region information and, consequently, could be paired up or combined together to achieve better results.

Also, it is important to note that the GCEBFM method's performance strongly depends on the level of diversity existing in the initial ensemble of segmentations. This means that a better strategy for the generation of the segmentation ensemble could ensure better performance results for our fusion model.

#### D. Computational Complexity

Due to our optimization strategy based on the ensemble of superpixels (see Algorithm 1), the time complexity of our fusion algorithm is  $O(nLN_sN_o)$ , where  $n, L, N_s$ , and  $N_o$  are, respectively, the pixel number within the image, the number of segmentations to be fused, the number of superpixels existing in the set of segmentations (to be fused) and  $N_o < T_{\text{max}}$ , the

 TABLE IV

 Average CPU Time for Different Segmentation Algorithms

	ALGORITHMS	PRI	CPU time (s)	On [image size]
ion Model	-GCEBFM-	0,80	$\simeq 180$	$[320 \times 214]$
	-VOIBFM- [19]	0,81	$\simeq 60$	$[320 \times 214]$
	-FMBFM- [20]	0,80	$\simeq 90$	$[320 \times 214]$
ı Fus	-CTM- [37,38]	0,76	$\simeq 180$	$[320 \times 200]$
Witł	-PRIF- [18]	0,80	$\simeq 20$	$[320 \times 214]$
	-FCR- [11]	0,79	$\simeq 60$	$[320 \times 200]$
nout Fusion Model	-MDSCCT- [50]	0,81	$\simeq 60$	$[320 \times 214]$
	-CTex- [52]	0,80	$\simeq 85$	$[184 \times 184]$
	-FH- [55]	0,78	$\simeq 1$	$[320 \times 200]$
	-HMC- [57]	0,78	$\simeq 80$	$[320 \times 200]$
With	-JSEG- [60]	0,77	$\simeq 6$	$[184 \times 184]$

number of iterations of the steepest local energy descent (since our iterative optimizer can stop before the maximum number of iterations  $T_{\text{max}}$ , when convergence is reached).

The segmentation operation takes, on average, about 2 and 3 min for an Athlon-AMD 64-Proc-3500+, 2.2 GHz, 4422.40 bogomips and nonoptimized code running on Linux; namely, the two steps (i.e., the estimations of the L = 60 weak segmentations to be combined and the minimization step of our fusion algorithm) takes, respectively, on average, 1 min to generate the segmentation ensemble and approximately 2 or 3 min for the fusion step and for a  $320 \times 214$  image (Table IV compares the average computational time for an image segmentation and for different segmentation algorithms whose PRI is greater than 0.76). Also, it is important to mention that the initial segmentations to be combined and the proposed energy-based fusion algorithm could easily be processed in parallel or could efficiently use multicore processors. It is straightforward for the generation of the set of segmentations but also truth for our fusion model by an application of a Jacobi-type version of the Gauss-Seidel-based ICM procedure [72]. The final energy-based minimization can be efficiently performed via the use of the parallel abilities of a graphic processor unit (integrated on most computers) which could significantly speed up the algorithm.

Finally, the source code (in C++ language) of our model and the ensemble of segmented images are publicly accessible via this link: http://www.etud.iro.umontreal.ca/~khelifil/ ResearchMaterial/gcebfm.html in the goal to make possible eventual comparisons with different performance measures and future segmentation methods.

## V. CONCLUSION

In this paper, we have introduced a novel and efficient fusion model whose objective is to fuse multiple segmentation maps to provide a final improved segmentation result, in the GCE sense. This new fusion criterion has the appealing property to be perceptual and specifically well suited to the inherent multiscale nature of any image segmentations (which could be possibly viewed as a refinement of another segmentation). More generally, this new fusion scheme can be exploited for any clustering problems using spatially indexed data (e.g., motion detection or estimation, 3-D reconstruction, depth field estimation, 3-D segmentation, etc.). In order to include an explicit regularization hyper parameter overcoming the inherent ill-posed nature of the segmentation problem, we have recasted our likelihood estimation problem of the consensus segmentation (or the so-called median partition) in the penalized likelihood framework by adding, to the simple ML fusion model a merging regularization term allowing to integrate knowledge about the types of resulting fused segmentation, a priori considered as acceptable solutions. This penalized likelihood estimation procedure remains simple to implement, perfectible, by incrementing the number of segmentation to be fused, adapted to lower outliers, general enough to be applied to different other problems dealing with label fields and is suitable to be implemented in parallel or to fully take advantage of multicore (or multi-CPU) systems.

### REFERENCES

- W. Tao, H. Jin, and Y. Zhang, "Color image segmentation based on mean shift and normalized cuts," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1382–1389, Oct. 2007.
- [2] A. Strehl and J. Ghosh, "Cluster ensembles—A knowledge reuse framework for combining multiple partitions," J. Mach. Learn. Res., vol. 3, pp. 583–617, Mar. 2003.
- [3] D. Parikh and R. Polikar, "An ensemble-based incremental learning approach to data fusion," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 2, pp. 437–450, Apr. 2007.
- [4] A. L. N. Fred and A. K. Jain, "Data clustering using evidence accumulation," in *Proc. 16th Int. Conf. Pattern Recognit. (ICPR)*, Quebec City, QC, Canada, Aug. 2002, pp. 276–280.
- [5] S. Vega-Pons and J. Ruiz-Shulcloper, "A survey of clustering ensemble algorithms," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 25, no. 3, pp. 337–372, 2011.
- [6] L. I. Kuncheva, "Switching between selection and fusion in combining classifiers: An experiment," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 32, no. 2, pp. 146–156, Apr. 2002.
- [7] A. J. Sharkey, Combining Artificial Neural Nets Ensemble and Modular Multi-Net Systems, 1st ed. Secaucus, NJ, USA: Springer-Verlag New York, 1999.
- [8] T. G. Dietterich, "Ensemble methods in machine learning," in Proc. 1st Int. Workshop Multiple Classifier Syst., Cagliari, Italy, 2000, pp. 1–15.
- [9] Y. Jiang and Z.-H. Zhou, "SOM ensemble-based image segmentation," *Neural Process. Lett.*, vol. 20, no. 3, pp. 171–178, 2004.
- [10] J. Keuchel and D. Küttel, "Efficient combination of probabilistic sampling approximations for robust image segmentation," in *Proc. DAGM Symp.*, Berlin, Germany, 2006, pp. 41–50.
- [11] 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.
- [12] P. Wattuya, K. Rothaus, J.-S. Prassni, and X. Jiang, "A random walker based approach to combining multiple segmentations," in *Proc. 19th Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2008, Tampa, FL, USA, pp. 1–4.
- [13] W. M. Rand, "Objective criteria for the evaluation of clustering methods," J. Amer. Stat. Assoc., vol. 66, no. 336, pp. 846–850, 1971.
- [14] T. M. Nguyen and Q. M. J. Wu, "Gaussian-mixture-model-based spatial neighborhood relationships for pixel labeling problem," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 1, pp. 193–202, Feb. 2012.
- [15] R. Harrabi and E. B. Braiek, "Color image segmentation using multilevel thresholding approach and data fusion techniques: Application in the breast cancer cells images," *EURASIP J. Image Video Process.*, vol. 2012, no. 1, pp. 1–11, 2012.
- [16] S. Ghosh, J. J. Pfeiffer, and J. Mulligan, "A general framework for reconciling multiple weak segmentations of an image," in *Proc. Workshop Appl. Comput. Vis. (WACV)*, Dec. 2009, pp. 1–8.
- [17] A. Alush and J. Goldberger, "Ensemble segmentation using efficient integer linear programming," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1966–1977, Oct. 2012.
- [18] M. Mignotte, "A label field fusion Bayesian model and its penalized maximum Rand estimator for image segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1610–1624, Jun. 2010.

- [19] M. Mignotte, "A label field fusion model with a variation of information estimator for image segmentation," *Inf. Fusion*, vol. 20, pp. 7–20, Nov. 2014.
- [20] C. Hélou and M. Mignotte, "A precision-recall criterion based consensus model for fusing multiple segmentations," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 7, no. 3, pp. 61–82, 2014.
- [21] X. Ceamanos *et al.*, "A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data," *Int. J. Image Data Fusion*, vol. 1, no. 4, pp. 293–307, 2010.
- [22] B. Song and P. Li, "A novel decision fusion method based on weights of evidence model," *Int. J. Image Data Fusion*, vol. 5, no. 2, pp. 123–137, 2014.
- [23] L. Franek, D. D. Abdala, S. Vega-Pons, and X. Jiang, "Image segmentation fusion using general ensemble clustering methods," in *Proc. 10th Asian Conf. Comput. Vis. Vol. Part IV (ACCV)*, Queenstown, New Zealand, 2011, pp. 373–384.
- [24] M. Ozay, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Fusion of image segmentation algorithms using consensus clustering," in *Proc. 20th IEEE Int. Conf. Image Process. (ICIP)*, Melbourne, VIC, Australia, Sep. 2013, pp. 4049–4053.
- [25] S. Vega-Pons and J. Ruiz-Shulcloper, "A survey of clustering ensemble algorithms," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 25, no. 3, pp. 337–372, 2011.
- [26] R. Unnikrishnan and M. Hebert, "Measures of similarity," in *Proc.* 7th IEEE Workshops Appl. Comput. Vis. (WACV/MOTIONS), vol. 1. Breckenridge, CO, USA, Jan. 2005, p. 394.
- [27] A. Ben-Hur, A. Elisseeff, and I. Guyon, "A stability based method for discovering structure in clustered data," in *Proc. Pac. Symp. Biocomput.*, 2002, pp. 6–17.
- [28] B. G. Mirkin, *Mathematical Classification and Clustering* (Nonconvex Optimization and its Applications). Dordrecht, The Netherlands: Kluwer, 1996.
- [29] S. V. Dongen, "Performance criteria for graph clustering and Markov cluster experiments," Nat. Res. Inst. Math. Comput. Sci., Amsterdam, The Netherlands, Tech. Rep. INS-R0012, 2000.
- [30] Y. Zhao and G. Karypis, "Criterion functions for document clustering: Experiments and analysis," Dept. Comput. Sci., Univ. Minnesota, Minneapolis, MN, USA, Tech. Rep. 01-40, 2002.
- [31] A. Rosenberg and J. B. Hirschberg, "V-measure: A conditional entropybased external cluster evaluation measure," in *Proc. Joint Conf. Empirical Methods Nat. Lang. Process. Comput. Nat. Lang. Learn.* (*EMNLP-CONLL*), Prague, Czech Republic, 2007, pp. 410–420.
- [32] S. Vega-Pons, J. Correa-Morris, and J. Ruiz-Shulcloper, "Weighted cluster ensemble using a kernel consensus function," in *Proc. 13th Iberoamerican Congr. Pattern Recognit. (CIARP)*, Havana, Cuba, 2008, pp. 195–202.
- [33] S. Vega-Pons, J. Correa-Morris, and J. Ruiz-Shulcloper, "Weighted partition consensus via kernels," *Pattern Recognit.*, vol. 43, no. 8, pp. 2712–2724, 2010.
- [34] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, p. 37, 1960.
- [35] M. Banerjee, M. Capozzoli, L. McSweeney, and D. Sinha, "Beyond kappa: A review of interrater agreement measures," *Can. J. Stat.*, vol. 27, no. 1, pp. 3–23, 1999.
- [36] 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. Vis. (ICCV)*, vol. 2. Vancouver, BC, Canada, Jul. 2001, pp. 416–423.
- [37] A. Y. Yang, J. Wright, S. S. Sastry, and Y. Ma, "Unsupervised segmentation of natural images via lossy data compression," *Comput. Vis. Image Understand.*, vol. 110, no. 2, pp. 212–225, May 2008.
- [38] Y. Ma, H. Derksen, W. Hong, and J. Wright, "Segmentation of multivariate mixed data via lossy data coding and compression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1546–1562, Sep. 2007.
- [39] D. R. Martin, "An empirical approach to grouping and segmentation," Ph.D. dissertation, Dept. EECS, Univ. California, Berkeley, CA, USA, Aug. 2003.
- [40] D. R. Martin, C. 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.
- [41] 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.

- [42] L. Chen, C. L. P. Chen, and M. Lu, "A multiple-kernel fuzzy c-means algorithm for image segmentation," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 5, pp. 1263–1274, Oct. 2011.
- [43] A. Lorette, X. Descombes, and J. Zerubia, "Fully unsupervised fuzzy clustering with entropy criterion," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, Barcelona, Spain, Sep. 2000, pp. 986–989.
- [44] I. B. Ayed and A. Mitiche, "A region merging prior for variational level set image segmentation," *IEEE Trans. Image Process.*, vol. 17, no. 12, pp. 2301–2311, Dec. 2008.
- [45] J. Besag, "On the statistical analysis of dirty pictures," J. Roy. Stat. Soc., vol. B-48, no. 3, pp. 259–302, 1986.
- [46] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.
- [47] M. Mignotte, "A non-stationary MRF model for image segmentation from a soft boundary map," *Pattern Anal. Appl.*, vol. 17, no. 1, pp. 129–139, Apr. 2014.
- [48] S. Chitroub, "Classifier combination and score level fusion: Concepts and practical aspects," *Int. J. Image Data Fusion*, vol. 1, no. 2, pp. 113–135, 2010.
- [49] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "A measure for objective evaluation of image segmentation algorithms," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR) Meas. Objective Eval. Image Segmentation Algorithms*, vol. 3. San Diego, CA, USA, Jun. 2005, pp. 34–41.
- [50] M. Mignotte, "MDS-based segmentation model for the fusion of contour and texture cues in natural images," *Comput. Vis. Image Understand.*, vol. 116, no. 9, pp. 981–990, Sep. 2012.
- [51] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.
- [52] D. E. Ilea and P. F. Whelan, "CTex—An adaptive unsupervised segmentation algorithm based on color-texture coherence," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1926–1939, Oct. 2008.
- [53] M. Krninidis and I. Pitas, "Color texture segmentation based on the modal energy of deformable surfaces," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1613–1622, Jul. 2009.
- [54] M. Mignotte, "A de-texturing and spatially constrained K-means approach for image segmentation," *Pattern Recognit. Lett.*, vol. 32, no. 2, pp. 359–367, Jan. 2011.
- [55] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," Int. J. Comput. Vis., vol. 59, no. 2, pp. 167–181, 2004.
- [56] M. Mignotte, "MDS-based multiresolution nonlinear dimensionality reduction model for color image segmentation," *IEEE Trans. Neural Netw.*, vol. 22, no. 3, pp. 447–460, Mar. 2011.
- [57] R. Hedjam and M. Mignotte, "A hierarchical graph-based Markovian clustering approach for the unsupervised segmentation of textured color images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Cairo, Egypt, Nov. 2009, pp. 1365–1368.
- [58] M. Donoser, M. Urschler, M. Hirzer, and H. Bishof, "Saliency driven total variation segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Kyoto, Japan, 2009, pp. 817–824.
- [59] M. M. Mushrif and A. K. Ray, "A-IFS histon based multithresholding algorithm for color image segmentation," *IEEE Signal Process. Lett.*, vol. 16, no. 3, pp. 168–171, Apr. 2009.
- [60] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of colortexture regions in images and video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 8, pp. 800–810, Aug. 2001.
- [61] M. B. Salah, A. Mitiche, and I. B. Ayed, "Multiregion image segmentation by parametric kernel graph cuts," *IEEE Trans. Image Process.*, vol. 20, no. 2, pp. 545–557, Feb. 2011.
- [62] L. Bertelli, B. Sumengen, B. S. Manjunath, and F. Gibou, "A variational framework for multiregion pairwise-similarity-based image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 8, pp. 1400–1414, Aug. 2008.
- [63] R. Huang, N. Sang, D. Luo, and Q. Tang, "Image segmentation via coherent clustering in l\*a\*b\* color space," *Pattern Recognit. Lett.*, vol. 32, no. 7, pp. 891–902, 2011.
- [64] T. Cour, F. Benezit, and J. Shi, "Spectral segmentation with multiscale graph decomposition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2005, pp. 1124–1131.
- [65] 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.
- [66] J. Wang, Y. Jia, X.-S. Hua, C. Zhang, and L. Quan, "Normalized tree partitioning for image segmentation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

- [67] S. P. Chatzis and G. Tsechpenakis, "The infinite hidden Markov random field model," *IEEE Trans. Neural Netw.*, vol. 21, no. 6, pp. 1004–1014, Jun. 2010.
- [68] E. Sharon, M. Galun, D. Sharon, R. Basri, and A. Brandt, "Hierarchy and adaptivity in segmenting visual scenes," *Nature*, vol. 442, pp. 810–813, Aug. 2006.
- [69] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [70] M. Meilă, "Comparing clusterings—An information based distance," J. Multivar. Anal., vol. 98, no. 5, pp. 873–895, 2007.
- [71] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí, "Yet another survey on image segmentation: Region and boundary information integration," in *Proc. 7th Eur. Conf. Comput. Vis. (ECCV)*, Copenhagen, Denmark, 2002, pp. 408–422.
- [72] P.-M. Jodoin and M. Mignotte, "Markovian segmentation and parameter estimation on graphics hardware," *J. Electron. Imag.*, vol. 15, no. 3, Jul./Sep. 2006, Art. no. 033005.



Lazhar Khelifi received the bachelor's degree in computer science from Gafsa University, Gafsa, Tunisia, in 2010, and the master's degree in computer science from University of Tunis El Manar, Tunis, Tunisia, in 2012. He is currently pursuing the Ph.D. degree in computer science with the Department of Computer Science and Operations Research, University of Montreal, Montreal, QC, Canada.

His research interests include image segmentation, fusion, and multiobjective optimization.



Max Mignotte received the D.E.A. degree in digital signal, image, and speech processing from the Grenoble Institute of Technology, Grenoble, France, in 1993, and the Ph.D. degree in electronics and computer engineering from the University of Bretagne Occidental, Brest, France, and the Digital Signal Laboratory, French Naval Academy, Brest, in 1998.

He was a French Institute for Research in Computer Science and Automation Post-Doctoral Fellow with the Department of Computer Science

and Operations Research, University of Montreal, Montreal, QC, Canada, from 1998 to 1999, where he is currently an Associate Professor with the Computer Vision and Geometric Modeling Laboratory. He is also a member of the Laboratorire de Recherche en Imagerie et Orthopédie (LIO) at the Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CHUM), Montreal, Hôpital Notre Dame, Montreal and a Researcher with 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.