Contents lists available at ScienceDirect





Information Fusion

journal homepage: www.elsevier.com/locate/inffus

EFA-BMFM: A multi-criteria framework for the fusion of colour image segmentation



Lazhar Khelifi*, Max Mignotte

DIRO, University of Montreal, Montreal, QC, Canada

ARTICLE INFO

Article history: Received 19 May 2016 Revised 14 February 2017 Accepted 3 March 2017 Available online 18 March 2017

Keywords:

Colour textured image segmentation Combination of multiple segmentations Multi-objective optimization Berkeley image database Entropy

ABSTRACT

Considering the recent progress in the development of practical applications in the field of image processing, it is increasingly important to develop new, efficient and more reliable algorithms to solve an image segmentation problem. To this end, various fusion-based segmentation approaches which use consensus clustering, and which are based on the optimization of a single criterion, have been proposed. One of the greatest challenges with these approaches is to select the best fusion criterion, which gives the best performance for the image segmentation model. In this paper, we propose a new fusion model of image segmentation based on multi-objective optimization, which aims to overcome the limitation and bias caused by a single criterion, and to provide a final improved segmentation. To address the ill-posedness for the search of the best criterion, the proposed fusion model combines two conflicting and complementary criteria for segmentation fusion, namely, the region-based variation of information (VoI) criterion and the contour-based F-measure (precision-recall) criterion using an entropy-based confidence weighting factor. To optimize our energy-based model, we propose an extended local optimization procedure based on superpixels and derived from the iterative conditional mode (ICM) algorithm. This new multiobjective median partition-based approach, which relies on the fusion of inaccurate, quick and spatial clustering results, has emerged as an appealing alternative to the use of traditional segmentation fusion models which exist in the literature. We perform experiments using the Berkeley database with manual ground truth segmentations, and the results clearly show the feasibility and efficiency of the proposed methodology.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The focus of image segmentation is to divide an image into separate regions which have uniform and homogeneous attributes [1]. This step is crucial and important in higher-level tasks such as feature extraction, pattern recognition, and target detection [2]. Several promising methods for segmentation of textured natural images have been recently proposed and reported in the literature. Of those, the ones which are based on the combination of multiple and weak segmentations of the same image to improve the quality of segmentation results are appealing from a theoretical perspective and offer an effective compromise between the complexity of the segmentation model and its efficiency.

* Corresponding author.

E-mail addresses: khelifil@iro.umontreal.ca (L. Khelifi), Mignotte@iro.umontreal.ca (M. Mignotte).

URL: http://www-etud.iro.umontreal.ca/~khelifil/ (L. Khelifi), http://www.iro.umontreal.ca/~mignotte/ (M. Mignotte)

http://dx.doi.org/10.1016/j.inffus.2017.03.001 1566-2535/© 2017 Elsevier B.V. All rights reserved. Most of these approaches, which are used to compute the segmentation fusion result from a set of initial and weak putative segmentation maps, are theoretically based on the notion of median partition. According to a given specific criterion (which can also be expressed as a distance or a similarity index/measure between two segmentation maps), the median partition approach aims to minimize the average of the distances (or to maximize the average of similarity measures), separating the (consensus) solution from the other segmentations to be fused. To date, a large and growing number of fusion-segmentation approaches based on the result of the median partition problem, along with different criteria or different optimization strategies, have been proposed in the literature.

For example, a fusion model of weak segmentations was initially introduced in the evidence accumulation sense in [3] with a co-association matrix, and in [4], it is then based on a minimization of the inertia (or intra-cluster scatter) criterion across cluster instances (represented by the set of local re-quantized label histogram given by each input segmentation to be fused). The fusion of multiple segmentation maps has also been proposed with respect to the Rand Index (RI) criterion (or its probabilistic version), with either a stochastic constrained random walking technique [5] (within a mutual information-based estimator to assess the optimal number of regions), an algebraic optimization method [6], a Bayesian Markov random field model [7], a superpixel-based approach optimized by the expectation maximization procedure [8] or finally, according to a similarity distance function built from the adjusted RI [9] and optimized with a stochastic gradient descent. It should also be noted that the solution of the median partition problem can be determined according to an entropy criterion, either in the variation of information (VoI) sense [10], using a linear complexity and energy-based model optimized by an iterative steepest-local energy descent strategy combined with a spatial connectivity constraint, or in the mutual information sense [11] using expectation maximization (EM) optimization. The fusion of clustering results can also be carried out according to the global consistency criterion (GCE) [12] (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) or based on the precision-recall criterion [13] using a hierarchical relaxation scheme. In this context, Franek et al. [14] proposed a methodology allowing the use of virtually any ensemble clustering method to address the problem of imagesegmentation combination. The strategy is mainly based on a preprocessing step which estimates a superpixel map from the segmentation ensemble in order to reduce the dimensionality of the combinatorial problem. Finally, in remote sensing, there have been reports of the combining model based on the maximum-margin sense (of the hyperplanes between spatial clusters) [15] or the recent Bayesian fusion procedure proposed in [16], in which the class labels obtained from different segmentation maps (obtained from different sensors) are fused by the weights of the evidence model.

In fact, the performance of these energy-based fusion models is related both to the optimization procedure, with its potential ability to find an optimal solution (as quickly as possible), and it also largely depends on the chosen fusion criterion, which defines all the intrinsic properties of the consensus segmentation map to be estimated. However, by assuming that an efficient optimization procedure is designed and implemented (in terms of its ability to quickly find a global optimal and stable solution), it remains unclear whether it can find the most appropriate single criterion allowing both to extract all the useful information contained in the segmentation ensemble and also to model all the complex geometric properties of the final consensus segmentation map. Another way to look at this problem is to understand that if the optimization problem is based on the optimization of a single criterion, the fusion procedure is inherently biased towards searching one particular family of possible solutions; otherwise, some specific regions of the search space contain solutions, which are a priori defined (by the criterion), as acceptable solutions. This may bias and limit the performance of an image segmentation model. To overcome this main disadvantage (the bias caused by a single criterion), we propose an interesting solution to use approaches based on multi-objective optimization in order to design a new fusion-segmentation model which takes advantage of the (potential) complementarity of different objectives (criteria), and enables us to finally obtain a better consensus segmentation result. Following this new strategy, in this work, we introduce a new multi-criteria fusion model weighted by an entropy-based confidence measure (EFA-BMFM). The main goal of this model is to simultaneously combine and optimize two different and complementary segmentation-fusion criteria, namely, the (region-based) Vol criterion and the (contour-based) F-measure (derived from the precision-recall) criterion.

The remainder of the paper is organized as follows. In Section 2, we present basic concepts of multi-objective optimization. In Section 3, we describe the generation of the segmentation ensem-



Fig. 1. The weighted formula approach (WFA).

ble to be fused by our model, while in Section 4, we describe the proposed fusion model, *i.e.*, the used segmentation criteria, the multi-objective function and the optimization strategy of the proposed algorithm for the fusion of image segmentation. We explain the experiments and discussions in Section 5, and in Section 6, we conclude the paper.

2. Multi-objective optimization

The motivation of using multi-objective (MO) optimization comes from all the drawbacks and limitations of using a monoobjective one, as mentioned in our preliminary work [17]. As previously mentioned, the final segmentation solution is inherently biased by the chosen single criterion as well as by the parameters of the model and the possible outliers of the segmentation ensemble. A MO optimization-based segmentation fusion framework enables us to more efficiently extract the useful information contained in the segmentation ensemble according to different criteria or different viewpoints, as well as to model easily all the complex geometric properties of the final consensus segmentation map *a priori* defined as the acceptable solution. To this end, the challenge is to find two different and complementary criteria.

Contrary to the mono-objective optimization case, in the MO optimization case, there are often several conflicting objectives to be simultaneously optimized [18]. Existing approaches which are utilized to solve a MO problem can be distinguished into two classes [19]. The first class is called the Pareto approach (PTA), and aims to provide a set of solutions which are non-dominated with respect to each objective. The second class (adopted in our work) is called the weighted formula approach (WFA), which transforms a MO problem into a problem with a single objective function. This is typically achieved by first assigning a numerical (estimated data-driven) weight to each objective (evaluation criterion), and then combining the values of the weighted criteria. The formula to determine the quality (or cost) *Z* which is related to a given candidate model is written as:

$$Z = w_1 c_1 + w_2 c_2 + \dots + w_n c_n$$
 (1)

with *n* representing the number of evaluation criteria, and w_i are real-valued weights (assigned to criteria c_i) which satisfy the following relations: $0 \le w_i \le 1$ and $\sum_{i=1}^{n} w_i = 1$.

A geometric representation of the WFA approach is given in Fig 1. In fact, the minimization of Z can be analysed by searching

the value of the tangency point *A* for which the line *T* with slope $-w_1/w_2$ (associated with $c_2 = -\frac{W_1}{W_2}c_1 + \frac{Z}{W_2}$ in the case of two objectives) just touches the boundary of the set of feasible solutions (*FS*) (related to the couple $[c_1, c_2]$). Note that the estimation of the weights (also known as the *importance factors*) is an essential step, and should be based on the degree of information or the confidence levels regarding the ensemble of segmentations (to be fused) provided by each criterion, along with the difference of the scaling between these two criteria. This re-scaling is essential to prevent either of the two criteria from being assigned too much significance; otherwise, it would make the fusion of the two criteria ineffective. To address this issue, we propose an entropy-based confidence measure (see Section 4.3).

3. Generation of the initial segmentations

In our application, it is simple to acquire the initial segmentations (see Fig. 2) used by our fusion Framework. To do this, we employ a *K*-means [20] clustering technique, with an image expressed in 12 different colour spaces,¹ namely: RGB, HSV, YIQ, XYZ, LAB, LUV, i123, h123, YCbCr, TSL, HSL and P1P2. For each input image of the BSDS300, we predict the cluster number of the *K*-means algorithm (*K*) using a metric which measures the complexity in terms of the number of distinct texture classes within the image. This metric, which is defined in [21], has a range of [0, 1], where a value close to 0 means that the image has few texture patterns, and a value close to 1 means that the image has several different types of texture. Mathematically, the value of *K* is written as:

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

where floor(*x*) is a function which 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. In our application, we used three different values of K^{max} , namely $K_1^{\text{max}} = 11$, $K_2^{\text{max}} = 9$ and $K_3^{\text{max}} = 3$. Additional details about the complexity value of an image are given in [10]. Note that in our case, the complexity is a measure of the absolute deviation (L_1 norm) of the set of normalized histograms or feature vectors for each overlapping squared fixed-size (N_w) neighbourhood contained within the input image.

Moreover, we used a set of values of the re-quantized colour histogram (as a feature vector for the *K*-means) with equidistant binning, which is estimated around the pixel to be classified. In our framework, this local histogram is equally re-quantized for each of the three-colour channels in a $N_b = q_b^3$ bin descriptor. This descriptor is computed on an overlapping squared fixed-size ($N_w = 7$) neighbourhood, which is centered around the pixel to be segmented using three different values of K^{max} for the *K*-means algorithm, and using 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.

It should be noted that different weak segmentations (resulting from a simple *K*-means expressed in different colour spaces) used in our fusion model can be easily viewed as different and complementary image channels, as provided by various sensors. In this context, our fusion model has the same goal of a multi-sensor data fusion scheme [22–24], which aims to take advantage of the complementarity in the data in order to improve the final result. In addition, different values of K^{max} (which is related to the cluster number) and q_b (related to the level of resolution of the texture model used in the *K*-means) enable us to generate a consistent variability in the segmentation ensemble, and considers the inherently ill-posed nature of the segmentation problem, which is due to the large number of possible partitions for a single image, and which can also be segmented at different levels of resolution or detail by different human observers.

4. Proposed fusion method

4.1. Region-based Vol criterion

The Vol [25] is an information theoretic criterion used for comparing two segmentations (partitions) or clusterings. By measuring the amount of information which is lost or gained while switching from one clustering to another, this metric aims to quantify the information shared between two partitions. In particular, the Vol takes a value of 0 when two clusterings are identical, but ≤ 1 otherwise. Similarly, it also expresses roughly the amount of randomness in one segmentation which cannot be explained by the other [26].

Let us assume that there is a machine segmentation to be computed (or compared) $S^a = \{C_1^a, C_2^a, \dots, C_{R^a}^a\}$ relative to a (ideal) manually segmented image $S^b = \{C_1^b, C_2^b, \dots, C_{R^b}^b\}$, where R^a represents the number of segments or regions (*C*) in S^a and R^b denotes the number of regions in S^b . The Vol distance between S^a and S^b can be written as follows:

$$Vol(S^{a}, S^{b}) = H(S^{a}) + H(S^{b}) - 2I(S^{a}, S^{b})$$
(3)

where $H(S^a)$ and $H(S^b)$ denote the entropy associated with the segmentation S^a , and S^b and $I(S^a, S^b)$ represent the mutual information between these two spatial partitions. Let n be the number of pixels within the image, let n_i^a be the number of pixels in the *i*th cluster i of the segmentation S^a , n_j^b the number of pixels in the *j*th cluster j of the segmentation S^b and finally, n_j^i the number of pixels which are together in the *i*th cluster (or region) of the segmentation S^a and in the *j*th cluster of the segmentation S^b . Note that the entropy is always positive or zero in the case where there is no uncertainty (when there is only one cluster), and is given by:

$$H(S^{a}) = -\sum_{i=1}^{R^{a}} P(i) \log P(i) = -\sum_{i=1}^{R^{a}} \left(\frac{n_{i}^{a}}{n}\right) \log \left(\frac{n_{i}^{a}}{n}\right)$$
(4)

$$H(S^{b}) = -\sum_{j=1}^{R^{b}} P(j) \log P(j) = -\sum_{j=1}^{R^{b}} \left(\frac{n_{j}^{b}}{n}\right) \log \left(\frac{n_{j}^{b}}{n}\right)$$
(5)

where $P(i) = n_i^a/n$ represents the probability that a pixel belongs to cluster S^a (respectively $P(j) = n_j^b/n$ being the probability that a pixel belongs to cluster S^b) in the case where *i* and *j* represent two discrete random variables with values of R^a and R^b , respectively, and uniquely related to the partition S^a and S^b . Now, let us assume that $P(i,j) = n_{ij}/n$ represents the probability when a pixel belongs to C_i^a and to C_j^b , which is the mutual information between the partitions S^a , and S^b is equal to the mutual information between the

¹ It should be noted that each colour space has an interesting specific property which is efficiently taken into account in our application in order to better diversify the segmentation ensemble (to be fused), and thus making a more reliable final fusion procedure. For example, RGB is an additive colour system based on trichromatic theory, and is nonlinear with visual perception. This space colour appears to be optimal for tracking applications [27]. The LAB colour system approximates human vision, and its component closely matches the human perception of lightness [28]. The LUV components provide an Euclidean colour space yielding a perceptually uniform spacing of colour approximating a Riemannian space [29]. The HSV is interesting in order to decouple chromatic information from the shading effect [30]. The YIQ colour channels have the property of being able to code the luminance and chrominance information, which are useful in compression applications. Besides, this system is intended to take advantage of human colour characteristics. XYZ has the advantage of being more psycho-visually linear, although they are nonlinear in terms of linear-component colour mixing. Each of these properties will be efficiently combined by our fusion technique.



Fig. 2. Examples of initial segmentation set and combination result (output of Algorithm 1). (a) Results of K-means clustering. (b) Input image ID 198054 selected from the Berkeley image dataset. (c) Final segmentation given by our fusion framework. (d) Contour superimposed on the colour image.



Fig. 3. Two images from the BSDS300 (a) and its ground truth boundaries (b). Segmentation results obtained by our EFA-BMFM are shown in (c).

random variables *i* and *j*, and is expressed as:

$$I(S^{a}, S^{b}) = \sum_{i=1}^{R^{a}} \sum_{j=1}^{R^{b}} P(i,j) \log\left(\frac{P(i,j)}{P(i)P(j)}\right).$$
(6)

4.2. Contour-based F-measure criterion

In the field of statistical analysis, the F-measure score (also called the F-score or F1 score) is defined as a measure of a test's accuracy. We obtained the results of the F-measure from a combination of two complementary measures, i.e. precision (*Pr*) and recall (*Re*). In the (contour-based) image segmentation domain, these two scores respectively represent the fraction of detections which are true boundaries and the fraction of the true boundaries detected [13]. In particular, a low precision value is typically the result of significant over-segmentation, and highlights the fact that a large number of boundary pixels have poor localization. On the contrary, the recall is low when there is significant undersegmentation or when there is a failure to capture the salient image structure (in terms of contours). In other words, precision and recall can be understood in terms of the rate of *false positives* and *missed detection*.

Mathematically, let us assume that a segmentation result $S^a = \{C_1^a, C_2^a, \ldots, C_{R^a}^a\}$ has to be compared with a manually segmented image $S^b = \{C_1^b, C_2^b, \ldots, C_{R^b}^b\}$ (considered as ground truth), where R^a represents the number of regions (*C*) in S^a and R^b denotes the number of regions in S^b . Now, let B_{C^a} be the set of pixels which belong to the boundary of the segment C^a in the segmentation S^a (B_{C^b} is the set of pixels belonging to the boundary of the segment C^b in the ground truth segmentation S^b). The precision (*Pr*) and recall (*Re*) are then respectively defined as:

$$Pr = \frac{|B_{C^{a}} \cap B_{C^{b}}|}{|B_{C^{a}}|}, \quad Re = \frac{|B_{C^{a}} \cap B_{C^{b}}|}{|B_{C^{b}}|}$$
(7)

where \cap denotes the intersection operator and |X| represents the cardinality of the set of pixel *X*.

Generally, 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 a probabilistic boundary representation, a precision-recall curve displays the trade-off between the absence of noise and the fidelity to the ground truth, considering that the threshold parameter of the boundary detector varies. A specific application² can characterise the relative cost α between these two amounts, which highlights a particular point on the precision-recall curve [31]. In this case, the new expression of the F-measure is given as follows:

$$F_{\alpha} = \frac{Pr \times Re}{\alpha \times Re + (1 - \alpha) \times Pr}$$
(8)

which is within the range [0, 1] where a score equal to 1 indicates that two segmentations are identical (*i.e.* they have identical contours).

4.3. Multi-objective function

The VoI and F-measures, which are described in Sections 4.1 and 4.2, are in fact frequently used to validate a new segmentation method [7,31,32] as two complementary comparison measures which enable the assessment of an automatic segmentation (i.e. given by an algorithm) relative to a set of ground truth segmentations (provided by a set of human experts). This summarizes the possible (and consistent) interpretation of an input image segmented at different levels of detail or resolution levels (see Fig. 3). Let $\{S_k^b\}_{k \le L} = \{S_1^b, S_2^b, \dots, S_L^b\}$ be a finite ensemble of *L* manually obtained ground truth segmented images of the same scene (segmented by L different human experts at different levels of detail), and S^a be the spatial clustering result to be estimated by making a comparison with the manually labeled set $\{S_{k}^{b}\}_{k < I}$. The mean F-measure and the mean VoI metrics are simply the two metrics which consider this set of possible ground truth segmentations, i.e.:

$$\overline{\mathsf{C}}\left(S^{\mathsf{a}}, \{S^{b}_{k}\}_{k \le L}\right) = \frac{1}{L} \sum_{k=1}^{L} \mathsf{C}\left(S^{\mathsf{a}}, S^{b}_{k}\right) \tag{9}$$

with $C \in \{\text{Vol, } F_{\alpha}\}$. In particular, the $\overline{\text{Vol}}$ distance function will give a low value (on the contrary, the \overline{F}_{α} measure function will give a high value) to a segmentation result S^a which is in good agreement with the set of segmentation maps obtained from human experts.

In our case, we aim to obtain a final improved segmentation result \hat{S} by the fusion of a family of *L* segmentations $\{S_k\}_{k \leq L} = \{S_1, S_2, \ldots, S_L\}$ (associated with the same scene or image), with the hope that the result is more accurate than that of each individual member of $\{S_k\}_{k \leq L}$. To this end, these two complementary criteria, namely the contour-based F-measure and the region-based Vol measure, can be used directly as an MO cost function in an energy-based model. From this point of view, the consensus segmentation $\hat{S}_{\overline{\text{MO}}}$ is simply obtained as the result of the following bi-criteria optimization problem:

$$\hat{S}_{\overline{\text{MO}}} = \arg\min_{S \in S_n} \overline{\text{MO}} \left(S, \{S_k\}_{k \le L} \right) \quad \text{with} :$$
 (10)

² In the case of an algorithm performing a search task, it is usually preferable to have a lower rate of false positives (higher precision) than a low rate of missed detections (high recall).

$$\overline{\text{MO}}(S, \{S_k\}_{k \le L}) = \mathsf{w}_{\text{Vol}} \overline{\text{Vol}}(S, \{S_k\}_{k \le L}) + \frac{\mathsf{w}_{\mathsf{F}_{\alpha}}}{\overline{\mathsf{F}}_{\alpha}(S, \{S_k\}_{k \le L})}$$
(11)

where S is a segmentation map belonging to the set of possible segmentations ($S \in S_n$). The importance (or weighting) factors w_{VoI} and $w_{F_{\alpha}}$ must be data-driven and estimated based on the concept of the informational importance of the segmentation ensemble given a criterion, or according to the traditional multiplecriteria analysis decision making (MCDM) problem under uncertainty [33] based on the intrinsic information generated by the segmentation ensemble through each criterion.

In our model, we can use the entropy value to measure the amount of decision information contained in the segmentation ensemble and related to each criterion as follows (for the VoI criterion and similarly for the F-measure):

$$e_{\text{Vol}} = -D \sum_{i=1}^{L} \left\{ \frac{\overline{\text{Vol}}(S_i, \{S_k\}_{k \le L})}{S_{\overline{\text{Vol}}}} \log \frac{\overline{\text{Vol}}(S_i, \{S_k\}_{k \le L})}{S_{\overline{\text{Vol}}}} \right\}$$
(12)

where:
$$S_{\overline{\text{Vol}}} = \sum_{j=1}^{L} \overline{\text{Vol}}(S_j, \{S_k\}_{k \le L})$$
 (13)

where $D=1/\log(L)$ is a constant which guarantees $0 \leq e_{Vol} \leq 1.$ In this context, the degree of divergence of the intrinsic information (or the contrast intensity) of the VoI and the F_{α} criterion can be measured as follows:

$$\mathbf{d}_{\mathrm{Vol}} = 1 - \mathbf{e}_{\mathrm{Vol}} \tag{14}$$

$$\mathbf{d}_{F_{\alpha}} = 1 - \mathbf{e}_{F_{\alpha}} \tag{15}$$

and finally, the objective weight for each criterion (VoI and F_{α}) is thus defined by:

$$W_{Vol} = \frac{d_{Vol}}{d_{Vol} + d_{F_{\alpha}}} \quad \in [0, 1]$$
(16)

$$W_{F_{\alpha}} = \frac{d_{F_{\alpha}}}{d_{Vol} + d_{F_{\alpha}}} \quad \in [0, 1].$$
(17)

In this manner, the entropy generated by the set of mean pairwise VoI distances of each weak segmentation (i.e. the set of rough segmentations to be fused) is first computed to obtain e_{VOI} (in addition, the entropy generated by the set of mean pairwise F_{α} distances of each weak segmentation allows us to obtain $e_{E_{\alpha}}$). Then, $e_{\rm VOI}$ and $e_{F_{\rm ev}}$ enable us to estimate the degrees of divergence of the intrinsic information related to each criterion, *i.e.* d_{Vol} or $d_{F_{u}}$ (also referred to as the inherent contrast intensity [33]), and are finally both used to compute the weight *W* associated with each criterion.

Conceptually, the entropy e_{VOI} or $e_{F_{\alpha}}$ defines the uncertainty of distribution of mean pairwise distances (related to each criterion). For example, if the set of weak segmentation maps to be fused have similar pairwise mean distances relative to the VoI criterion, this VoI criterion transmits too little information (relative to the other F_{α} criterion) to the fusion (decision maker) model [34]. As a result, the weight W_{Vol} of this Vol criterion is less because this criterion becomes less important for our fusion model.

4.4. Optimization of the fusion model

To enable us to solve this consensus function, in the bi-criteria sense, we resort to a deterministic search technique, which is called the iterative conditional mode (ICM), proposed by Besag [35] (i.e. a Gauss-Seidel relaxation), where pixels are updated one at a time. In this work, we used a much more effective enhancement of the ICM algorithm, which involves utilizing a superpixel

Algorithm 1. EFA-based fusion model algorithm.

Mathematical n	otation:
VoI	Mean VoI
\overline{F}_{lpha}	Mean F-measure
MO	Multi-objective function
$\{S_j\}_{j\leq J}$	Set of J segmentations to be fused
$\{z_j\}_{j\leq J}$	Set of weights
$\{b_j\}$	Set of superpixels $\in \{S_j\}_{j \leq J}$
ε	Set of region labels in $\{S_j\}_{j \leq J}$
T_{\max}	Maximal number of iterations (=10)
$S_{I_{best}}$	Fusion segmentation result
α	F-measure compromise parameter
e _{voI}	Entropy of the VoI criterion
$e_{F_{lpha}}$	Entropy of the F-measure criterion

Input: $\{S_j\}_{j < J}$

Output: $S_{I_{best}}$

- A. Initialization:
- 1: Compute e_{Vol} (see (12))

2: Compute $e_{F_{\alpha}}$ 3: if $e_{Vol} < e_{F}$ then

$$\begin{array}{ll} S: \ \mathbf{n} \ \mathbf{c}_{Vol} < \mathbf{c}_{F_{\alpha}} \ \mathbf{n} \ \mathbf{end} \\ 4: \qquad \mathbf{S}^{[0]} = \operatorname{promin} \\ \end{array}$$

4:
$$S_I^{[0]} = \arg \min_{S \in \{S_j\}_{j \le J}} \overline{\text{VoI}}(S, \{S_j\}_{j \le J})$$

5: else

$$S_I^{[0]} = \arg\min_{S \in \{S_j\}_{j \le J}} \overline{\mathbf{F}_{\alpha}}(S, \{S_j\}_{j \le J})$$

6

B. Steepest Local Energy Descent:

8: while $p < T_{\text{max}}$ do

- for each b_j superpixel $\in \{S_j\}_{j \leq J}$ do 9:
- 10: Draw a new label x according to the uniform distribution in the set \mathcal{E}
- Let $S_I^{[p],\text{new}}$ the new segmentation map including b_j with 11: the region label x
- Compute $\overline{MO}(S_I, \{S_j\}_{j \le J})$ on $S_I^{[p],\text{new}}$ (see (10)) if $\overline{MO}(S_{\overline{MO}}^{[p],\text{new}}) < \overline{MO}(S_{\overline{MO}}^{[p]})$ then $\overline{MO} = \overline{MO}^{\text{new}}$ $S_I^{[p]} = S_I^{[p],\text{new}}$ $S_{I_{best}} = S_I^{[p]}$ end if 12: 13. 14: 15: 16: 17: end if end for 18: 19: $p \leftarrow p + 1$ 20: end while

(i.e. the regions or segments given by each individual segmentation S_k generated by the K-means algorithm) concept instead of pixels. This superpixel-based strategy makes our consensus energy function nearly convex by adding several region-based constraints (among other advantages over the pixel-based fusion method [36]). However, with the lack of proper initialization, this algorithm will converge towards a bad local minima(i.e. a local minima which is far away from the global minimum, and which gives a poor segmentation result).

Again, to solve this problem, we resort to the entropy values of each criteria (see (12)). Thus, we select the criteria which gives the minimal entropy (i.e. the most informative criterion; see Section 4.3), and for the first iteration of the ICM, of the L segmentations to be fused, we then choose the one which ensures the minimal consensus energy (in this selected criterion sense) of our fusion model. Because this iterative algorithm amounts to achieving simultaneously, for each superpixel to be labeled, the minimum value of (11), we call this segmentation algorithm a multicriteria fusion model based on the entropy-weighted formula approach (EFA-BMFM). The pseudo-code of EFA-BMFM is shown in Algorithm 1.



Fig. 4. Example of fusion convergence result for three various initializations. (a) Berkeley image ID 229036 and its ground-truth segmentations. (b) A non informative (or blind) initialization. (c) The worst input segmentation. (d) The best input segmentation (from the segmentation set) selected by the entropy method (see Section 4.4). (e), (f) and (g) segmentation results after 10 iterations of our EFA-BMFM fusion model (resulting from (b), (c) and (d), respectively).

5. Experimental tests and results

5.1. Data set and benchmarks

In order to measure the performance of the proposed fusion model, we validate our approach using the famous Berkeley segmentation database (BSDS300) [37]. Recently, this dataset has been enriched to BSDS500³ [38] with 200 additional test colour images of size 481×321 . In order to quantify the efficacy of the proposed segmentation algorithm, for each colour image, the BSDS300 and the BSDS500 offer a set of benchmark segmentation results (*i.e.* ground truth), given by human observers (between 4 and 7). In addition, we used the Matlab source code proposed in [32] with the aim of estimating the different quantitative performance measures (*i.e.* the four image segmentation indices

presented in Section 5.3). This code is available online at: http: //www.eecs.berkeley.edu/~yang/software/lossysegmentation. In addition, to test the effectiveness for other types of images, we tested our proposed method on the aerial image segmentation dataset (ASD)⁴ [39], and we performed a quantitative evaluation using two medical images (a brain magnetic resonance imaging (MRI) and a cornea image) recently used in [40] and [41].

5.2. Initial tests

Our initial tests can be divided into two main stages. First, we tested the convergence properties of our ICM procedure based on superpixels by choosing as the initialization of our iterative local gradient-descent algorithm various initializations extracted from our segmentation ensemble $\{S_k\}_{k \leq L}$ (these convergence proper-

³ The BSDS300 [37] and the BSDS500 [38] are available online at: https://www2. eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html.

⁴ The ASD [39] is available online at: http://web.ornl.gov/~jiy/ASD/Aerial Image Segmentation Dataset.html.



Average error of different initialization methods (using PRI)

Fig. 5. Average error of different initialization methods (for the probabilistic Rand index (PRI) performance measure) on the BSDS300.



Fig. 6. Progression of the segmentation result as a function of the number of segmentations (L) to be fused for the EFA-BMFM algorithm. More precisely, for L = 12, 24, 36, 48 and 60 segmentations.



Fig. 7. Progression of the Vol, (lower is better) and the PRI (higher is better) according to the segmentation number (L) to be fused for our proposed EFA-BMFM algorithm (on the BSDS500). Precisely, for L = 1, 12, 24, 36, 48 and 60 segmentations.

ties have been discussed in Section 5.7). From our results, the final energy value, along with the resulting final segmentation map, is on average better when the initial segmentation solution is associated with an initialization chosen by our proposed entropybased method, while it remains robust to other initializations (see Section 4.4 and Fig. 4). We also found that the average error for the PRI performance measure (on the BSDS300) is lower when the initial segmentation solution is associated with an initialization chosen by our entropy-based method (Init - best in Fig. 5).

Secondly, we tested the effect of the number of initial segmentations on the accuracy of the final segmentation result. Qualitatively, Fig. 6 shows that the final consensus result is even better than the size of the segmentation ensemble L is high. Quantitatively, we observed that the different performance measures (see Section 5.3) are improved when we increase the number of initial segmentations. This test demonstrates the validity of the proposed fusion procedure, and shows that the segmentation results can be



Fig. 8. A sample of results obtained by applying our proposed algorithm to images from the Berkeley dataset compared to other algorithms. From left to right: original images, FCR [4], SCKM [51], MD2S [50], GCEBFM [44], MDSCCT [48] and our method (EFA-BMFM).

enhanced if the segmentation ensemble is completed by other segmentation maps of the same scene.

5.3. Performance measures and results

In an attempt to test and evaluate our fusion segmentation model, we employed four performance metrics which are most popular in the literature. These well-known performance measures⁵ are:

- 1. The Probabilistic Rand index (PRI) [42] counts the fraction of pairs of pixels whose labels are consistent between the computed segmentation and the human segmentation, averaging through all of the ground-truth segmentation of a given image.
- 2. The boundary displacement error (BDE) [43] measures the average displacement error of boundary pixels between two segmented images. In particular, it defines the error of one boundary pixel as the distance between the pixel and the closest pixel in the other boundary image.
- 3. The variation of information (VoI) [25] defines the distance between two segmentations as the average conditional entropy of one segmentation given the other; it measures the amount of

⁵ The GCE metric is in the range [0;1], where a score of 0 indicates that there is a perfect match between two segmentations and an error of 1 represents a maximum difference between the two segmentations to be compared [44]. Also, the PRI metric is in the range [0;1], where higher values indicating greater similarity between two segmentations [7]. For the BDE measure, a value near-zero indicates high quality of the image segmentation, and its maximum value can be the length of the image segmentation [43]. The VOI metric taking a value of 0 when two segment

tations are identical and positive otherwise. This metric is in the range [0;log(n)], where *n* denotes the number of pixels within the image [10].

Table 1

Performance of several segmentation algorithms (with or without a fusion model strategy) for three different performance measures: Vol, GCE and BDE (lower is better), on the BSDS300.

ALGORITHMS	VoI	GCE	BDE
HUMANS	1.10	0.08	4.99
Algorithms With Fusion Model			
EFA-BMFM	1.870	0.198	8.284
-2016- GCEBFM [44]	2.10	0.19	8.73
-2014- FMBFM [13]	2.01	0.20	8.49
-2014- VOIBFM [10]	1.88	0.20	9.30
-2014- SFSBM [21]	2.21	0.21	8.87
-2010- PRIF [7]	1.97	0.21	8.45
-2008- FCR [4]	2.30	0.21	8.99
-2007- CTM _{γ=20} [32]	2.02	0.19	9.90
Algorithms Without Fusion Model			
-2016- DGA-AMS [46]	2.03	-	-
-2014- CRKM [47]	2.35	-	-
-2012- MDSCCT [48]	2.00	0.20	7.95
-2012- AMUS [8]	1.68	0.17	-
-2011- KM [49]	2.41	-	-
-2011- MD2S [50]	2.36	0.23	10.37
-2010- SCKM [51]	2.11	0.23	10.09
-2009- MIS [52]	1.93	0.19	7.83
-2009- HMC [53]	3.87	0.30	8.93
-2008- NTP [54]	2.49	0.24	16.30
-2008- Av. Diss [55]	2.62	-	-
-2005- NCuts _{$K=20$} [56] (in [32])	2.93	0.22	9.60
-2004- $FH_{\sum=0.5,k=500}$ [57] (in [32])	2.66	0.19	9.95
-2002- Mean-Shift [58]	2.48	0.26	9.70

Table 2

Performance of several segmentation algorithms (with or without a fusion model strategy) for the PRI performance measure (higher is better) on the BSDS300.

ALGORITHMS	PRI
HUMANS	0.87
Algorithms With Fusion Model	
EFA-BMFM	0.806
-2016- GCEBFM [44]	0.80
-2014- FMBFM [13]	0.80
-2014- VOIBFM [10]	0.81
-2014- SFSBM [21]	0.79
-2010- PRIF [7]	0.80
-2009- Consensus [6]	0.78
-2008- FCR [4]	0.79
-2007- CTM $_{\gamma=20}$ [32]	0.76
Algorithms Without Fusion Model	
-2016- LSI [59]	0.80
-2014- CRKM [47]	0.75
-2011- SCKM [51]	0.80
-2011- MD2S [50]	0.78
-2011- KM [49]	0.76
-2009- MIS [52]	0.80
-2009- HMC [53]	0.78
-2009- Total Var [60]	0.78
-2009- A-IFS HRI [61]	0.77
-2008- CTex [62]	0.80
-2004 - FH _{$\Sigma=0.5,k=500$} [57] (in [32])	0.78
-2005- NCuts _{$K=20$} [56] (in [32])	0.72
-2002- Mean-Shift [58]	0.75
$-2001 - JSEG_{c=255,s=1.0,m=0.4}$ [63] (in [62])	0.77

information which is lost or gained while switching from one region to another.

4. The global consistency error (GCE) [37] determines the extent to which one segmentation map can be viewed as a refinement of another segmentation map. In this way, for a perfect match, every region in one of the segmentations must be a refinement (*i.e.*, a subset) of a region in the other segmentation.

Table 3

Performance of several segmentation algorithms (with or without a fusion model strategy) for three different performance measures: Vol, GCE and BDE (lower is better), on the BSDS500.

ALGORITHMS	VoI	GCE	BDE
HUMANS	1.10	0.08	4.99
Algorithms With Fusion Model			
EFA-BMFM	1.97	0.21	7.90
-2016- GCEBFM [44]	2.18	0.20	8.61
-2014- FMBFM [13]	2.00	0.21	8.19
-2014- VOIBFM [10]	1.95	0.21	9.00
-2010- PRIF [7]	2.10	0.21	8.88
-2008- FCR [4]	2.40	0.22	8.77
-2007- CTM [32] (in [64])	1.97	-	-
Algorithms Without Fusion Model			
-2011- WMS _{$d_{1}=20$} [65] (in [64])	2.10	-	-
-2004- $FH_{\Sigma=0.8}$ [57] (in [64])	2.18	-	-
-2002- Mean-Shift [58] (in [64])	2.00	-	-

Table 4

Performance of several segmentation algorithms (with or without a fusion model strategy) for the PRI performance measure (higher is better) on the BSDS500.

ALGORITHMS	PRI
HUMANS	0.87
Algorithms With Fusion Model	
EFA-BMFM	0.81
-2016- GCEBFM [44]	0.80
-2014- FMBFM [13]	0.80
-2014- VOIBFM [10]	0.80
-2010- PRIF [7]	0.79
-2008- FCR [4]	0.79
-2007- CTM [32] (in [64])	0.73
Algorithms Without Fusion Model	
-2004- $FH_{\Sigma=0.8}$ [57] (in [64])	0.77
-2011- $WMS_{d_{\Lambda}=20}$ [65] (in [64])	0.75
-2002- Mean-Shift [58] (in [64])	0.77

As can be seen from the results given in Tables 1 and 2, for the BSDS300, our method generally outperforms the state-of-the art algorithms in terms of the different distance measures with: BDE = 8.284, VoI = 1.870, GCE = 0.198 (a lower value is better) and PRI = 0.806 (a higher value is better). From the tables, we also see that if we compare our results to a mono-objective approach (FMBFM and VOIBFM) based on the same single criterion, we obtain significantly better results. This shows clearly that our strategy of combining two complementary (contour and region-based) criteria of segmentation (the VoI and the F-measure) is effective. In addition, from the data in Tables 3 and 4, we observe that for the BSDS500, our method gives comparable performance results compared to different algorithms with or without the fusion model when: BDE = 7.90, VoI = 1.97, GCE = 0.21 (a lower value is better) and PRI = 0.81. Moreover, Fig. 7, we observe that the PRI and VoI performance scores are better when L (the segmentation number to be fused) is high. This test shows that our performance scores can be further improved if we increase the number of segmentations to be fused. In addition, for better comparison, in Fig. 8, we present a sample of results obtained by applying our algorithm to some images from the Berkeley dataset compared to other stateof-the-art algorithms. In addition, Fig. 9 displays a small number of segmented images which are similar to those shown in the monocriterion fusion model (FMBFM and VOIBFM) proposed in [10] and [13], respectively. Fig. 10 shows the best and worst segmentation results (in the PRI sense) from the BSDS300. The results for the entire database will be available on the website of the author. Fig. 11 shows the distribution of the PRI, BDE, VoI and GCE measures. From this figure, we can conclude that few segmentations



Fig. 9. Additional segmentation results obtained from the BSDS300.





Fig. 10. Best and worst segmentation results (in the PRI sense) obtained from the BSDS300. First column: (a) image ID 167062 and (b) its segmentation result (PRI=0.99). Second column: (c) image ID 175043 and (d) its segmentation result (PRI = 0.37).



Fig. 11. Distribution of the BDE, GCE, PRI and Vol measures over the 300 segmented images of the BSDS300.



Fig. 12. Distribution of the number and size of regions over the 300 segmented images of the BSDS300.

exhibit poor PRI and BDE scores even for the most difficult segmentation cases. Moreover, Fig. 12 shows the distribution of the number and size of regions obtained by our EFA-BMFM algorithm over the BSDS300.

5.4. Comparison of medical image segmentation

Medical image segmentation is an important part of medical analysis, and is also a process which is clearly different from the segmentation of natural (textured colour) images because input medical images are generally in grey levels, have low contrast and are noisy. We performed two experiments on medical images to demonstrate the effectiveness and flexibility of our segmentation approach. In the first experiment, we used a brain magnetic resonance imaging (MRI), as shown in Fig. 13. The results, which were obtained by using the region-based model via local similarity factor (RLSF), the global active contour model (these two models which are based on active contour were recently proposed in [40]) and our EFA-BMFM model, are shown in Fig. 13(b)-(d), respectively. As can be seen, our method outperforms the global active contour model and gives an interesting result compared to the segmentation achieved by the RLSF model. In the second experiment, we tested our model on a real cornea image, and we compared the segmentation result provided by our EFA-BMFM model with the results given by the fast global minimization (FGM) [45] and the double fitting terms of multiplicative and difference (DMD) [41] models (see Fig. 14). We observe that the quality of the segmentation obtained by the FGM model for this cornea image is not as good as those of the DMD and EFA-BMFM. The reason for this is (as mentioned in [41]) that the image with intensity inhomogeneity is too challenging for the FGM.

5.5. Comparison of segmentation methods for aerial image segmentation

We also benchmarked our fusion model as a segmentation method using the aerial image segmentation dataset (ASD) [39]. This new image dataset contains 80 high-resolution aerial images, with spatial resolutions ranging from 0.3 to 1.0 m, including different scenes as schools, residential areas, cities, warehouses and power plants. The images were normalized to realize a resolution



(a)

(b)



Fig. 13. Comparison of two region-based active contour models on a brain MRI. (a) original image. (b) segmentation of the RLSF model [40]. (c) segmentation of the global active contour model [40]. (d) segmentation achieved by our EFA-BMFM model.

Table 5

Boundary benchmarks on the aerial image segmentation dataset (ASD). Results obtained for different segmentation methods (with or without the fusion model strategy). The figure shows the F-measures (higher is better) when choosing an optimal scale for the entire dataset (ODS) or per image (OIS).

ALGORITHMS	ODS	OIS			
HUMANS	0.68	0.69			
Algorithms Without Fusion	on Model				
FH [57]	0.59	0.62			
SRM [66]	0.58	0.60			
Mean shift [58]	0.56	0.58			
JSEG [63]	0.54	0.56			
FSEG [67]	0.58	0.61			
MSEG [68]	0.53	0.57			
Algorithms With Fusion Model					
EFA-BMFM	0.50	0.50			
VOIBFM [10]	0.36	0.36			
FMBFM [13]	0.53	0.53			

of 312 \times 312 pixels, and the segmentation results were then supersampled in order to obtain segmentation images with the original resolution (512 \times 512 pixels).

Table 5 shows the overall F-measure of different segmentation algorithms under two different scale settings. The first is the score under the optimal data set scale (ODS), and the average F-measure of 80 images at each scale is calculated and the best measure across scales is reported. The second is the score under the optimal image scale (OIS), which uses the best F-measure across scales for each image, and the average measure over images is reported.⁶ As can be seen from the data on Table 5, our method outperforms the VOIBFM fusion model in terms of both the ODS and OIS, and it re-

Table 6

Fusion segmentation models and complexity.

	EFA-BMFM	GCEBFM [44]	VOIBFM [10]
K-means step (generation of initial segmentations)	$O(N \times K \times I \times d)$	$O(N \times K \times I \times d)$	$O(N \times K \times I \times d)$
Fusion step	$O(Nbsup \times n)$	$O(Nbsup \times n)$	O(n)

mains generally competitive compared to segmentation algorithms without a fusion strategy. In addition, and for better comparison, samples of the results obtained by applying our algorithm to some images from the ASD dataset compared to other state-of-the art algorithms are given in Fig. 15.

5.6. Algorithm complexity

With respect to the time complexity, the first step of our algorithm (the generation of the initial ensemble of segmentations) has a complexity equal to $O(N \cdot K \cdot I \cdot d)$, where N, K, I and d are the number of points of each cluster, the number of clusters, the number of iterations and the dimension of each point to be clustered, respectively. Moreover, the second step (fusion algorithm) is characterized by a complexity time of $O(Nbsup \cdot n)$, where n is the pixel number within the image and Nbsup represents the number of superpixels existing in the set of segmentations to be fused (see Table 6 for a comparison with other methods).

As another important aspect, in terms of the execution time, the segmentation operation takes on average about 240 s for an Intel 64 Processor core i7-4800M Q, 2.7 GHz with 8GB of RAM memory and non-optimized code running on Linux; on average, it takes 60 s to generate the segmentation ensemble and approximately 180 s for the fusion step and for a 320×214 image (Table 7 compares the average computational time for an image segmentation and for different segmentation algorithms whose PRI is greater than 0.76). Further, it is important to note that the algorithm can easily be parallelized (using the parallel capabilities of a graphic

⁶ The soft contour map is provided by averaging, 6 times, the set of hard (*i.e.* binary) boundary representations of our segmentation method with different values of K_{max} (the number of classes of the segmentation).



(a)

(b)



(c)

(d)

Fig. 14. Comparison of two segmentation methods on segmenting a real cornea image. (a) original image of size 256×256 . (b) detection using the FGM method [45] (5000 iterations). (c) detection using the DMD method [41] (5 iterations). (d) detection resulting from our EFA-BMFM model (10 iterations).

Table 7Average CPU time for different segmentation algorithms for
the BSDS300.

ALGORITHMS	PRI	CPU time (s)	[image size]
-EFA-BMFM-	0.80	≃240	[320 × 214]
-GCEBFM- [44]	0.80	≃180	[320 × 214]
-VOIBFM- [10]	0.81	≃60	[320 × 214]
-FMBFM- [13]	0.80	≃90	[320 × 214]
-CTM- [32]	0.76	$\simeq 180$	$[320 \times 200]$
-PRIF- [7]	0.80	≃20	[320 × 214]
-FCR- [4]	0.79	≃ 6 0	$[320 \times 200]$
-MDSCCT- [48]	0.81	≃ 6 0	[320 × 214]
-CTex- [62]	0.80	≃85	[184 × 184]
-HMC- [53]	0.78	≃80	$[320 \times 200]$
-LSI- [59]	0.80	≃60	[481 × 321]

processor unit) because its two steps (described above) are purely independent. Finally, to enable comparisons with future segmentation methods, the source code (in C++ language) of our model

and the ensemble of segmented images are publicly accessible here: http://www-etud.iro.umontreal.ca/~khelifil/ResearchMaterial/ efa-bmfm.html.

5.7. Discussion

The most obvious finding to emerge from the above analysis is that the use of the MO optimization concept enables us to design a new fusion model that takes advantages of the complementarity of different segmentation criteria.

This interesting model appears to be very competitive for different kinds of performance measures, and it therefore appears as an alternative to complex and computationally demanding segmentation models which exist in the literature. Moreover, another possible alternative analysis is given in Table 8. In fact, from this table, we can confirm that the performance measures are quite different for a given image compared to the values obtained by other approaches. Thus, our model outperforms the VOIBFM [10] fusion



Fig. 15. A sample of results obtained by applying our algorithm to images from the aerial image dataset [39] compared to other popular segmentation algorithms (gPb-owtucm [38], Felz-Hutt (FH) [57], SRM [66], Mean shift [58], JSEG [63], FSEG [67] and MSEG [68]). The first row shows six example images. The second row overlays segment boundaries generated by four subjects, where the darker pixels correspond to the boundaries marked by more subjects. The last row shows the results obtained by our method (EFA-BMFM).

model and the MDSCCT [48] algorithm (a purely algorithmic approach), in terms of the number of images of the BSDS300 which obtain the best GCE, BDE and PRI scores. These results provide further support for the hypothesis that our model appears to be very competitive against other methods with or without a fusion strategy. Compared to the mono-objective approach, the combination of

two objectives makes our fusion algorithm slower, confirming the hypothesis in [69], and indicating that a high number of objectives cause additional challenges. However, it appears that the choice of using super-pixels with the ICM (as an optimization algorithm) limits this problem as the execution time remains close to those of other algorithms. In this context, we present a convergence analy-





Fig. 16. Convergence analysis. (a) input image ID 187039 selected from the BSDS300. (b) change of the segmentation map of our EFA-BMFM fusion model starting from a blind (or non informative) initialization. (c) evolution of the consensus energy function along the number of iterations of the EFA-BMFM.

sis of a Berkeley colour image, shown in Fig. 16. Fig. 16 shows (a) the original Berkeley image ID 187039 selected from the BSDS300, (b) the evolution of the segmentation map of our EFA-BMFM fusion model starting from a blind (or noninformative) initialization and (c) the evolution of the consensus energy function along the number of iterations of the EFA-BMFM. In Fig. 16 (c), we observe

that our EFA-BMFM model converged to a minimum energy value after 5 iterations. It should be noted that this faster convergence speed of our model resulted from the use of superpixels.

As mentioned in Section 1, to date, there have been no reports of the application of current knowledge of MO optimization to the field of the fusion of colour image segmentation. These interest-

Table 8

Comparison of scores between the EFA-BMFM and other segmentation algorithms for the 300 images of the BSDS300. Each value indicates the number of images of the BSDS300 which obtain the best score.

MEASURES	EFA-BMFM Vs GCEBFM [44]		EFA-BMFM Vs MDSCCT [48]		EFA-BMFM Vs VOIBFM [10]	
	EFA-BMFM	GCEBFM	EFA-BMFM	MDSCCT	EFA-BMFM	VOIBFM
GCE	216	84	261	39	167	133
VOI	143	157	122	178	134	166
BDE	151	149	175	125	201	99
PRI	147	153	167	133	160	140

ing results provided by our model are related both to the general-

- [6] S. Ghosh, J.J. Pfeiffer, J. Mulligan, A general framework for reconciling multiple weak segmentations of an image, in: Proceedings of the Workshop on Applications of Computer Vision, 2009, pp. 1-8, doi:10.1109/WACV.2009.5403029
- M. Mignotte, A label field fusion Bayesian model and its penalized maximum Rand estimator for image segmentation, IEEE Trans. Image Process. 19 (6) (2010) 1610-1624, doi:10.1109/TIP.2010.2044965
- [8] A. Alush, J. Goldberger, Ensemble segmentation using efficient integer linear programming, IEEE Trans. Pattern Anal. Mach. Intell. 34 (10) (2012) 1966-1977, doi:10.1109/TPAMI.2011.280.
- [9] M. Ozay, F.T.Y. Vural, S.R. Kulkarni, H.V. Poor, Fusion of image segmentation algorithms using consensus clustering, in: Proceedings of the 20th IEEE International Conference on Image Processing, 2013, pp. 4049-4053, doi:10.1109/ ICIP.2013.6738834
- [10] M. Mignotte, A label field fusion model with a variation of information estimator for image segmentation, Inf. Fusion 20 (2014) 7-20, doi:10.1016/j.inffus. 2013.10.012.
- [11] H. Wang, Y. Zhang, R. Nie, Y. Yang, B. Peng, T. Li, Bayesian image segmentation fusion, Knowl. Based Syst. 71 (2014) 162-168, doi:10.1016/j.knosys.2014.07.021.
- [12] L. Khelifi, M. Mignotte, GCE-based model for the fusion of multiples colour image segmentations, in: Proceedings of the 23th IEEE International Conference on Image Processing, 2016, pp. 2574-2578, doi:10.1109/ICIP.2016.7532824
- [13] C. Helou, M. Mignotte, A precision-recall criterion based consensus model for fusing multiple segmentations, Int. J. Signal Process. Image Process. Pattern Recognit. 7 (3) (2014) 61-82, doi:10.14257/ijsip.2014.7.3.07
- [14] L. Franek, D.D. Abdala, S. Vega-Pons, X. Jiang, Image segmentation fusion using general ensemble clustering methods, in: R. Kimmel, R. Klette, A. Sugimoto (Eds.), Computer Vision ? ACCV 2010. Lecture Notes in Computer Science, Vol 6495, Springer, Berlin, Heidelberg, 2011, pp. 373-384, doi:10.1007/ 978-3-642-19282-1_30.
- [15] X. Ceamanos, B. Waske, J.A. Benediktsson, J. Chanussot, M. Fauvel, J.R. Sveinsson, A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data, Int. J. Image Data Fusion 1 (4) (2010) 293-307, doi:10.1080/19479832.2010.485935
- [16] B. Song, P. Li, A novel decision fusion method based on weights of evidence model, Int. J. Image Data Fusion 5 (2) (2014) 123-137, doi:10.1080/19479832. 2014.894143
- [17] L. Khelifi, M. Mignotte, A new multi-criteria fusion model for colour textured image segmentation, in: Proceedings of the 23th IEEE International Conference on Image Processing, 2016, pp. 2579-2583, doi:10.1109/ICIP.2016.7532825.
- [18] L. Khelifi, I. Zidi, K. Zidi, K. Ghedira, A hybrid approach based on multiobjective simulated annealing and tabu search to solve the dynamic dial a ride problem, in: Proceedings of the International Conference on Advanced Logistics and Transport. 2013. pp. 227-232. doi:10.1109/JCAdLT.2013.6568464.
- [19] B.C. Wei, R. Mandava, Multi-objective nature-inspired clustering techniques for image segmentation, in: Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems (CIS), 2010, pp. 150-155, doi:10.1109/ICCIS.2010. 5518564.
- [20] S. Lloyd, Least squares quantization in pcm, IEEE Trans. Inf. Theory 28 (2) (1982) 129-137. doi:10.1109/TIT.1982.1056489.
- [21] M. Mignotte, A non-stationary MRF model for image segmentation from a soft boundary map, Pattern Anal. Appl. 17 (1) (2014) 129-139, doi:10.1007/ s10044-012-0272-z.
- [22] B. Khaleghi, A. Khamis, F.O. Karray, S.N. Razavi, Multisensor data fusion: a review of the state-of-the-art, Inf. Fusion 14 (1) (2013) 28-44, doi:10.1016/j. inffus 2011 08 001
- [23] A.A. Goshtasby, S. Nikolov, Image fusion: advances in the state of the art, Inf. Fusion 8 (2) (2007) 114-118, doi:10.1016/j.inffus.2006.04.001
- [24] Y. Liu, S. Liu, Z. Wang, Multi-focus image fusion with dense SIFT, Inf. Fusion 23 (2015) 139-155, doi:10.1016/j.inffus.2014.05.004
- [25] M. Meilã, Comparing clusterings an information based distance, J. Multivar. Anal. 98 (5) (2007) 873-895, doi:10.1016/j.jmva.2006.11.013.
- [26] M. Meilã, Comparing clusterings an axiomatic view, in: Proceedings of the 22nd International Conference on Machine Learning, 2005, pp. 577-584, doi:10.1145/1102351.1102424.
- [27] E. Maggio, A. Cavallaro, Multi-part target representation for colour tracking, in: Proceedings of the IEEE International Conference on Image Processing, 2005, pp. 729-732, doi:10.1109/ICIP.2005.1529854.
- [28] M. Millnert, Signal processing, image processing and pattern recognition, S. Banks, Prentice-Hall, Englewood Cliffs, NJ, 1990, ISBN 0-13-812579-1, xiv + 410 pp., 22.95, Int. J. Adapt. Control Signal Process. 6 (5) (1992) 519-520, doi:10.1002/acs.4480060511.

ity and the relative applicability of this MO concept with different segmentation criteria.

6. Conclusion

In this paper, we present a new and efficient multi-criteria fusion model based on the entropy-weighted formula approach (EFA-BMFM). The proposed model combines multiple segmentation maps to achieve a final improved segmentation result. This model is based on two complementary (contour and region-based) criteria of segmentation (the VoI and the F-measure criteria). We applied the proposed segmentation model to BSDS300, BSDS500, ASD and medical images, and the proposed model appears to be comparable to or even outperform other segmentation models, which proves the effectiveness and robustness of our multi-criteria fusion approach. In our model, the fusion process is performed at three different conceptual and hierarchical levels; first, at the criterion level, because the proposed fusion model combines two conflicting and complementary criteria; second, at the (segmentation) decision level by exploiting the combination of different and weak segmentations of the same image (expressed in different colour spaces); third, at the (pixel-)data level, and this is done by considering the set of superpixels as the atomic elements to be segmented in the consensus segmentation (instead of the set of pixels). Although our current multi-criteria fusion model is reasonably efficient and the superpixel strategy makes our energy function nearly convex, it would be interesting to optimize the consensus function with other optimization algorithms such as the exploration/selection/estimation (ESE) [70] or genetic algorithms. Thus, these algorithms are guaranteed to find the optimal solution; however, they have the drawback of a huge computational time. To overcome this problem, we can use the parallel computing capabilities of a graphic processor unit (GPU) (based on its massively parallel architecture consisting of thousands of smaller, which are designed to handle multiple tasks simultaneously). For all these reasons, the proposed fusion method may therefore be seen as an attractive strategy for solving the difficult image segmentation problem.

References

- [1] Y. Wu, P. Zhang, M. Li, Q. Zhang, F. Wang, L. Jia, SAR image multiclass segmentation using a multiscale and multidirection triplet Markov fields model in non-subsampled contourlet transform domain. Inf. Fusion 14 (4) (2013) 441-449, doi:10.1016/j.inffus.2012.12.001.
- [2] Y. Chen, A.B. Cremers, Z. Cao, Interactive colour image segmentation via iterative evidential labeling, Inf. Fusion 20 (2014) 292-304, doi:10.1016/j.inffus. 2014.03.007
- [3] A.L.N. Fred, A.k. Jain, Data clustering using evidence accumulation, in: Proceedings of the 16th International Conference on Pattern Recognition, 2002. pp. 276–280. doi:10.1109/JCPR.2002.1047450.
- [4] M. Mignotte, Segmentation by fusion of histogram-based k-means clusters in different colour spaces, IEEE Trans. Image Process. 17 (5) (2008) 780-787, doi:10.1109/TIP.2008.920761.
- [5] P. Wattuya, K. Rothaus, J.-S. Prassni, X. Jiang, A random walker based approach to combining multiple segmentations, in: Proceedings of the 19th International Conference on Pattern Recognition, 2008, pp. 1-4, doi:10.1109/ICPR. 2008 4761577

- [29] Z. Kato, T.-C. Pong, A Markov random field image segmentation model for colour textured images, Image Vision Comput. 24 (10) (2006) 1103–1114, doi:10.1016/j.imavis.2006.03.005.
- [30] P. Pérez, C. Hue, J. Vermaak, M. Gangnet, Colour-based probabilistic tracking, in: Proceedings of the European Conference on Computer Vision, 2002, pp. 661–675, doi:10.1007/3-540-47969-4_44.
- [31] D.R. Martin, C.C. Fowlkes, J. Malik, Learning to detect natural image boundaries using local brightness, colour, and texture cues, IEEE Trans. Pattern Anal. Mach. Intell. 26 (5) (2004) 530–549, doi:10.1109/TPAMI.2004.1273918.
- [32] A.Y. Yang, J. Wright, Y. Ma, S.S. Sastry, Unsupervised segmentation of natural images via lossy data compression, Comput. Vision Image Understanding 110 (2) (2008) 212–225, doi:10.1016/j.cviu.2007.07.005.
- [33] H. Deng, C.-H. Yeh, R.J. Willis, Inter-company comparison using modified TOP-SIS with objective weights, Comput. Oper. Res. 27 (10) (2000) 963–973, doi:10. 1016/S0305-0548(99)00069-6.
- [34] T.-C. Wang, H.-D. Lee, Developing a fuzzy TOPSIS approach based on subjective weights and objective weights, Expert Syst. Appl. 36 (5) (2009) 8980–8985, doi:10.1016/j.eswa.2008.11.035.
- [35] J. Besag, On the statistical analysis of dirty pictures, J. R. Stat. Soc. B-48 (1986) 259–302.
- [36] J.J. Lewis, R.J. O'Callaghan, S.G. Nikolov, D.R. Bull, N. Canagarajah, Pixel- and region-based image fusion with complex wavelets, Inf. Fusion 8 (2) (2007) 119–130, doi:10.1016/j.inffus.2005.09.006.
- [37] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings of the 8th International Conference on Computer Vision, volume 2, 2001, pp. 416–423, doi:10.1109/ICCV.2001.937655.
- [38] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (5) (2011) 898– 916, doi:10.1109/TPAMI.2010.161.
- [39] J. Yuan, S.S. Gleason, A.M. Cheriyadat, Systematic benchmarking for aerial image segmentation, IEEE Geosci. Remote Sens. Lett. 10 (6) (2013) 1527–1531, doi:10.1007/10.1109/LGRS.2013.2261453.
- [40] S. Niu, Q. Chen, L. de Sisternes, Z. Ji, Z. Zhou, D.L. Rubin, Robust noise regionbased active contour model via local similarity factor for image segmentation, Pattern Recognit. 61 (2017) 104–119, doi:10.1016/j.patcog.2016.07.022.
- [41] H. Ali, N. Badshah, K. Chen, G.A. Khan, A variational model with hybrid images data fitting energies for segmentation of images with intensity inhomogeneity, Pattern Recognit. 51 (2016) 27–42, doi:10.1016/j.patcog.2015.08.022.
- [42] R. Unnikrishnan, C. Pantofaru, M. Hebert, A measure for objective evaluation of image segmentation algorithms, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Workshop on Empirical Evaluation Methods in Computer Vision, Vol. 3, 2005, pp. 34–41, doi:10.1109/CVPR.2005. 390.
- [43] J. Freixenet, X. Muñoz, D. Raba, J. Marti, X. Cufi, Yet another survey on image segmentation: region and boundary information integration, in: Proceedings of the 7th European Conference on Computer Vision, 2002, pp. 408–422, doi:10. 1007/3-540-47977-5_27.
- [44] L. Khelifi, M. Mignotte, A novel fusion approach based on the global consistency criterion to fusing multiple segmentations, IEEE Trans. Syst. Man Cybern. (2016) 1–14, doi:10.1109/TSMC.2016.2531645.
- [45] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, S. Osher, Fast global minimization of the active contour/snake model, J. Math. Imaging Vision 28 (2) (2007) 151–167, doi:10.1007/s10851-007-0002-0.
- [46] M.A. Jaffar, A dynamic fuzzy genetic algorithm for natural image segmentation using adaptive mean shift, J. Exp. Theor. Artif.Intell. 29 (1) (2017) 149–156, doi:10.1080/0952813X.2015.1132263.
- [47] M.B. Salah, I.B. Ayed, J. Yuan, H. Zhang, Convex-relaxed kernel mapping for image segmentation, IEEE Trans. Image Process. 23 (3) (2014) 1143–1153, doi:10.1109/TIP.2013.2297019.
- [48] M. Mignotte, MDS-based segmentation model for the fusion of contour and texture cues in natural images, Comput. Vision Image Understanding 116 (9) (2012) 981–990, doi:10.1016/j.cviu.2012.05.002.
- [49] M.B. Salah, A. Mitiche, I.B. Ayed, Multiregion image segmentation by parametric kernel graph cuts, IEEE Trans. Image Process. 20 (2) (2011) 545–557, doi:10.1109/TIP.2010.2066982.
- [50] M. Mignotte, MDS-based multiresolution nonlinear dimensionality reduction model for colour image segmentation, IEEE Trans. Neural Netw. 22 (3) (2011) 447–460, doi:10.1109/TNN.2010.2101614.

- [51] M. Mignotte, A de-texturing and spatially constrained k-means approach for image segmentation, Pattern Recognit. Lett. 32 (2) (2011) 359–367, doi:10. 1016/j.patrec.2010.09.016.
- [52] M. Krinidis, I. Pitas, Colour texture segmentation based on the modal energy of deformable surfaces, IEEE Trans. Image Process. 18 (7) (2009) 1613–1622, doi:10.1109/TIP.2009.2018002.
- [53] R. Hedjam, M. Mignotte, A hierarchical graph-based Markovian clustering approach for the unsupervised segmentation of textured colour images, in: Proceedings of the IEEE International Conference on Image Processing (ICIP'09), 2009, pp. 1365–1368, doi:10.1109/ICIP.2009.5413555.
- [54] J. Wang, Y. Jia, X.-S. Hua, C. Zhang, L. Quan, Normalized tree partitionning for image segmentation, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, AK (USA), 2008, pp. 1–8, doi:10.1109/ CVPR.2008.4587454.
- [55] L. Bertelli, B. Sumengen, B.S. Manjunath, F. Gibou, A variational framework for multi-region pairwise similarity-based image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 30 (8) (2008) 1400–1414, doi:10.1109/TPAMI.2007. 70785.
- [56] T. Cour, F. Benezit, J. Shi, Spectral segmentation with multiscale graph decomposition, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 2, 2005, pp. 1124–1131, doi:10.1109/CVPR. 2005.332.
- [57] P.F. Felzenszwalb, D.P. Huttenlocher, Efficient graph-based image segmentation, Int. J. Comput. Vision 59 (2) (2004) 167–181, doi:10.1023/B:VISI.0000022288. 19776.77.
- [58] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 603–619, doi:10. 1109/34.1000236.
- [59] L. Dong, N. Feng, Q. Zhang, LSI: latent semantic inference for natural image segmentation, Pattern Recognit. 59 (2016) 282–291, doi:10.1016/j.patcog.2016. 03.005.
- [60] M. Donoser, M. Urschler, M. Hirzer, H. Bischof, Saliency driven total variation segmentation, in: Proceedings of the 12th IEEE International Conference on Computer Vision, 2009, pp. 817–824, doi:10.1109/ICCV.2009.5459296.
- [61] M.M. Mushrif, A.K. Ray, A-IFS histon based multi-thresholding algorithm for colour image segmentation, IEEE Signal Process. Lett. 16 (3) (2009) 168–171, doi:10.1109/LSP.2008.2010820.
- [62] D.E. Ilea, P.F. Whelan, Ctex- an adaptive unsupervised segmentation algorithm on colour-texture coherence, IEEE Trans. Image Process. 17 (10) (2008) 1926– 1939, doi:10.1109/TIP.2008.2001047.
- [63] Y. Deng, B.S. Manjunath, Unsupervised segmentation of colour-texture regions in images and video, IEEE Trans. Pattern Anal. Mach. Intell. 23 (8) (2001) 800– 810, doi:10.1109/34.946985.
- [64] S. Li, D.O. Wu, Modularity-based image segmentation, IEEE Trans. Circuits Syst. Video Technol. 25 (4) (2015) 570–581, doi:10.1109/TCSVT.2014.2360028.
- [65] A. Browet, P.-A. Absil, P.V. Dooren, Community detection for hierarchical image segmentation, in: Proceedings of the International Workshop Combinatorial Image Analysis, 2011, pp. 358–371, doi:10.1007/978-3-642-21073-0_32.
- [66] R. Nock, F. Nielsen, Statistical region merging, IEEE Trans. Pattern Anal. Mach. Intell. 26 (11) (2004) 1452–1458, doi:10.1109/tpami.2004.110.
- [67] X. Liu, D.L. Wang, A spectral histogram model for texton modeling and texture discrimination, Vision Res. 42 (23) (2002) 2617–2634, doi:10.1109/10. 1016/S0042-6989(02)00297-3.
- [68] U.C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, M. Heynen, Multiresolution, object-oriented fuzzy analysis of remote sensing data for GISready information, J. Photogramm. Remote Sens. 58 (3–4) (2004) 239–258, doi:10.1016/j.isprsjprs.2003.10.002.
- [69] D. Brockhoff, E. Zitzler, Are all objectives necessary? On dimensionality reduction in evolutionary multi-objective optimization, in: T.P. Runarsson, H.G. Beyer, E. Burke, J.J. Merelo-Guervós, L.D. Whitley, X. Yao (Eds.), Parallel Problem Solving from Nature - PPSN IX: 9th International Conference, Reykjavik, Iceland, September 9–13, volume 4193, Proceedings, ser. Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2006, 2006, pp. 533–542, doi:10.1007/11844297_54.
- [70] F. Destrempes, M. Mignotte, J.-F. Angers, A stochastic method for Bayesian estimation of hidden Markov models with application to a colour model, IEEE Trans. Image Process. 14 (8) (2005) 1096–1108, doi:10.1109/TIP.2005.851710.