Provided for non-commercial research and education use. Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/authorsrights

Information Fusion 20 (2014) 7-20

Contents lists available at ScienceDirect

Information Fusion

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

A label field fusion model with a variation of information estimator for image segmentation



INFORMATION FUSION



Max Mignotte*

DIRO, Univ. de Montréal, C.P. 6128, Succ. Centre-ville, H3C 3J7 Montréal, Canada

ARTICLE INFO

Article history: Received 3 December 2012 Received in revised form 10 October 2013 Accepted 14 October 2013 Available online 4 January 2014

Keywords: Cluster ensemble algorithm Combination of multiple segmentations Label field fusion Segmentation ensemble Variation of information

ABSTRACT

This paper proposes a new and reliable segmentation approach based on a fusion framework for combining multiple region-based segmentation maps (with any number of regions) to provide a final improved (i.e., accurate and consistent) segmentation result. The core of this new combination model is based on a consensus (cost) function derived from the recent information Theory based variation of information criterion, proposed by Meila, and allowing to quantify the amount of information that is lost or gained in changing from one clustering to another. In this case, the resulting consensus energy-based segmentation fusion model can be efficiently optimized by exploiting an iterative steepest local energy descent strategy combined with a connectivity constraint. This new framework of segmentation combination, relying on the fusion of inaccurate, quickly and roughly calculated, spatial clustering results, emerges as an appealing alternative to the use of complex segmentation models existing nowadays. Experiments on the Berkeley Segmentation Dataset show that the proposed fusion framework compares favorably to previous techniques in terms of reliability scores.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Segmenting an image is a useful and important pretreatment whose purpose is to represent the image content into different sub-parts or regions of coherent properties (also called *segments* or *superpixels*) with homogeneous characteristics (and corresponding to distinct regions of the image foreground or background). This pixel-level process simplifies and summarizes the image content in order to make it simpler to analyze and useful for the development of any high-level vision tasks [1].

Most of the *region-based* segmentation methods exploit first a texture feature extraction step whose goal is to represent each meaningful individual texture class with statistical (or fractal, morphological, structural, geometrical, etc.) image features and a subsequent clustering technique. In this context, decades of effort in image partitioning have shown that further enhancements (on the result) could be obtained either by exploiting more complex features and/or more advanced clustering techniques attempting to group (with different criteria or strategies) spatially homogeneous regions with coherent properties. These methods include segmentation algorithms exploiting directly clustering schemes [2–5] (with fuzzy sets [6] and/or after a first de-texturing step [5,7,8]), mode seeking algorithms [9–11], watershed transforma-

URL: http://www.iro.umontreal.ca/~mignotte/

tion [12] or region growing techniques [13], compression-based methods [14,12], statistical [15–21], Bayesian [22] or graph-based representations [23–25], variational frameworks [26–28], deformable surfaces [29] and spectral clustering [30] to name a few. In general, these enhancements have led to complex and time-consuming procedures and, for minimization-based segmentation schemes, to highly nonlinear and large-scale optimization problems.

To provide an alternative to these computationally costly and complex segmentation models, a possible and effective approach consists in combining (or seamlessly fusing) multiple low-cost and rough image segmentation results of the same scene associated with simpler segmentation model¹ to provide a final improved segmentation. In this strategy, instead of searching for the best segmentation algorithm (along with its optimal internal parameters) which is hardly possible for the set of all natural images, it is more interesting to find the most appropriate fusion model of segmentations. This strategy has initially been introduced in [31,32] with the restriction that all input segmentation results (to be combined) should contain the same number of regions and then a little later without this restriction, with any number of regions, in [2,33].

The key idea behind this fusion approach is that there exists no universal segmentation algorithm and/or selected features which

^{*} Tel.: +1 514 343 5747; fax: +1 514 343 5834. *E-mail address:* mignotte@iro.umontreal.ca

^{1566-2535/\$ -} see front matter @ 2013 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.inffus.2013.10.012

¹ Other strategies include different algorithms, with possibly different values of their intrinsic parameters (or seeds for stochastic algorithms) and/or different feature modalities for a given input scene possibly converted into different color spaces or spatially transformed (e.g., scale, skew, etc.) or by other means.

could successfully segment all images (and/or it is not easy to know the optimal algorithm or set of features for one particular image). On the other hand, it is logical to think that a gain could be expected from combining the strengths and features of several segmenters or multiple segmentation maps which, individually, might produce some poor segments (i.e., poor segmentation result for some sub-parts of the image) but for which there also often exist good segments. A clever merging of these segmentation results, with poor segments considered as noise (and good segments as reliable information), could produce a superior consensus segmentation than any of the individual input segmentations. This remark applies to all imagery problems dealing with label fields such as motion detection, 3D segmentation, 3D reconstruction, and depth field estimation.

Formally, the approach of fusing few classifiers to achieve better classification performance (better than a single classifier) is called, in statistical learning theory, a mixture of experts or classifier ensembles [34,35]. In this context, Dietterich [35] has given a simple and clear explanation, from representational, computational and statistical points of view, of why the ensembles or predictions of multiple classifiers can improve results. Motivated by these promising results, cluster ensembles that combine multiple data clusterings (instead of multiple classifiers) have then started to gain an increasing interest [36,37] in machine learning (see [38] for a good review of these methods).

The remainder of the document is divided into the following sections: First, relevant literature on fusion models of segmentations is reviewed and discussed in Section 2. Section 3 presents the proposed fusion framework and the minimization technique of the proposed consensus cost function. Section 4 presents the generation of the segmentation ensemble to be fused by the proposed model. Finally, Section 5 shows a variety of experimentation and comparisons with other leading segmentation algorithms.

2. Related work

The problem of combining multiple segmentations can be viewed as a cluster ensemble problem even if, these two problems are mainly different in two points. Indeed, it is worth noting that, in a fusion model of natural image segmentation maps, there are well defined classes (the texture class "tree" or "sky" are different and well defined even if the image is noisy). In addition, in the case of the combination of weak image segmentations, the important characteristics of a (textured) natural image, e.g., the inherent connection and the spatial dependencies between spatial neighboring objects (i.e., pixels) can also be exploited.

Despite of decades of intensive research to find a universal segmentation algorithm (and/or selected features) that can successfully segment all images, few works have been proposed on how to efficiently combine multiple segmentations. Nevertheless, it is important to cite the fusion model proposed in [2] which merges the individual input segmentations in the within-cluster variance (or intra-class inertia) sense (for the set of local label histogram values given by each input segmentations) since, the final segmentation result is optimized by applying a K-means algorithm based fusion scheme. This fusion of segmentation maps can also be achieved in the probabilistic Rand index [39] (PRI) sense, either with an algebraic optimization method [40] or with a random walking approach (and combined with a mutual information based estimator for estimating the optimal number of seed regions) [33] or finally in the penalized Rand index sense including a global constraint on the fusion model (restricting the size and the number of the regions) with a Markovian approach and an analytical optimization method [41]. Let us also cite the fusion model proposed in [37] in the evidence accumulation sense and using a hierarchical agglomerative clustering model.

The fusion model, presented in this paper, is different and is in fact, an energy-based model derived from the recently introduced Variation of Information (VoI) metric [42,43]. The proposed resulting consensus energy-based fusion model of segmentation can be efficiently optimized by exploiting an iterative steepest local energy descent strategy combined with a connectivity constraint. This fusion of possibly inaccurate, quickly and roughly estimated segmentation results may appear as an attractive and efficient alternative to the use of complex, expensive computational algorithms existing in the segmentation field.

This latter assumption comes from our own effort and past experience in searching the best unsupervised segmentation model (without fusion) of natural images [18–20,5,7,8]. To this end, it is important to understand that this objective is frequently conflicting between the increase in the segmentation model complexity (i.e., its ability to model efficiently different textures or complex interactions with a great number of tunable parameters in the model) and the optimization complexity required to find the solution of the model. It is worth mentioning that fusion models may be also faced with this kind of conflicting choices. A simple and less relevant fusion model is often easy to optimize (such as the model previously proposed, in the inertia sense, in [2]) and conversely, a more reliable and complex fusion model is often more critical to optimize (such as the one, in the PRI sense, proposed by the author in [41]).

3. Proposed fusion framework

3.1. The variation of information

The variation of information (VoI) metric [42,43] is a recent information Theory based measure for comparing two segmentations (partitions) or clusterings, of the same data set. This metric quantifies the information shared between two partitions by measuring, more precisely, the amount of information that is lost or gained in changing from one clustering to another. Equivalently (and conceptually), it also represents roughly the amount of randomness in one segmentation which cannot be explained by the other [42]. The VoI is a true metric on the space of clusterings which is positive, symmetric and obeys the triangle inequality [43].

This Vol metric can also be exploited as a clustering metric that measures the similarity between a possible clustering solution and the underlying ground truth. It was also exploited in the segmentation (or spatial clustering) field [4,14,41], for objectively evaluating the efficiency of different automatic segmentation algorithms.

Let $S^{t} = \{C_{1}^{t}, C_{2}^{t}, \dots, C_{R^{t}}^{t}\}$ and $S^{g} = \{C_{3}^{g}, C_{2}^{g}, \dots, C_{R^{g}}^{g}\}$ be respectively the segmentation test result and the ideal segmentation (or "ground truth"). Let also R^{t} and R^{g} , denote respectively, the cluster number (or number of regions²) in S^{t} and in S^{g} . The Vol between S^{t} and S^{g} is defined as:

$$\operatorname{Vol}(S^{\mathsf{t}}, S^{\mathsf{g}}) = H(S^{\mathsf{t}}) + H(S^{\mathsf{g}}) - 2 \cdot I(S^{\mathsf{t}}, S^{\mathsf{g}})$$
(1)

where $H(S^t)$ and $H(S^g)$ represent respectively the classical entropy associated with the segmentation S^t and S^g and $I(S^t, S^g)$ the mutual information between these two partitions. Let *n* be the number of pixels within the image, n_i^t the number of pixels in the *i*th cluster of the segmentation S^t , n_j^g the number of pixels in the *j*th cluster of the segmentation S^g and finally n_{ij} the number of pixels which are together in the *i*th cluster (or region) of the segmentation S^t and in the *j*th cluster of the segmentation S^g . The entropy is always

² A region is a group of connected pixels sharing a common class label and a class, a set of pixels possessing similar textural characteristics.

positive (it is null only when there is no uncertainty, namely when there is only one cluster) and is defined as:

$$H(S^{t}) = -\sum_{i=1}^{R^{t}} P(i) \log P(i) = -\sum_{i=1}^{R^{t}} \frac{n_{i}^{t}}{n} \log \frac{n_{i}^{t}}{n} H(S^{g}) = -\sum_{j=1}^{R^{g}} P(j) \log P(j) = -\sum_{j=1}^{R^{g}} \frac{n_{j}^{g}}{n} \log \frac{n_{j}^{g}}{n}$$
(2)

with $P(i) = n_i^t/n$ being the probability that a pixel belongs to cluster S^t (respectively $P(j) = n_j/n$ being the probability that a pixel belongs to cluster S^g) in the case where i and j represent two discrete random variables taking respectively R^t and R^g values and uniquely associated to the partition S^t and S^g . Let now $P(i,j) = n_{ij}/n$ represents the probability that a pixel belongs to C_i^t and to C_j^g , the mutual information $I(\cdot)$ between the partitions S^t and S^g is equal to the mutual information between the random variables i and j and is expressed in the following way:

$$I(S^{\mathsf{t}}, S^{\mathsf{g}}) = \sum_{i}^{R^{\mathsf{t}}} \sum_{j}^{R^{\mathsf{g}}} P(i, j) \log \frac{P(i, j)}{P(i) P(j)}$$
(3)

The Vol is a true metric across the lattice of possible clusterings (taking a value of 0 when two clusterings are identical and positive otherwise) and is bounded by $\log n$. However, if S^t and S^g have at most R^{max} clusters (i.e., regions), it is bounded by $2 \log R^{max}$ [43].

This VoI metric was also lately exploited in image partitioning [4,41] as an objective measure for the evaluation of a machine segmentation compared to a set of ground truth segmentations. This set of ground truth segmentations captures, in fact, the inherent variability of each possible (perceptually consistent) interpretation of an input image, segmented at different detail levels by each human segmenter [44,41,45]. Statistically speaking, this variability also reflects the ill-posed nature of the segmentation problem for which, additional constraints need to be imposed in order to provide the desired result, such as the number of classes or regions, a priori chosen by the user. In this case, and as proposed in [4], this concern can be easily addressed by computing the mean VoI metric as a simple and empirical way to take this diversity into account in the set of ground truths. More precisely, let $\{S_k^g\}_{k \leq L}$ be a finite ensemble of *L* manually or ground truth segmented images (of the same scene): $\{S_k^g\}_{k \leq L} = \{S_1^g, S_2^g, \dots, S_L^g\}$. Let S^t be the spatial clustering result to be evaluated by comparison with the manually labeled set $\{S_k^g\}_{k \le l}$, the mean VoI metric is simply defined by:

$$\overline{\operatorname{Vol}}\left(S^{\mathsf{t}}, \left\{S^{\mathsf{g}}_{k}\right\}_{k \leqslant L}\right) = \frac{1}{L} \sum_{k=1}^{L} \operatorname{Vol}\left(S^{\mathsf{t}}, S^{\mathsf{g}}_{k}\right) \tag{4}$$

As a consequence, the mean Vol distance function will give a low value to a segmentation result S^t that is all the more in accordance with the set of the segmentation maps obtained from human experts (i.e., a segmentation map resulting in a consensus or a compromise, in terms of contour accuracy or detail level displayed by each ground-truth segmentation).

3.2. Energy-based fusion model

Let now $\{S_k\}_{k \in L}$ be a finite ensemble of *L* segmentations, $\{S_k\}_{k \in L} = \{S_1, S_2, \dots, S_L\}$ (related to a same image), to be reconciled and combined with the aim of achieving a final improved (accurate and consistent) segmentation solution, better than the segmentation result of each member of $\{S_k\}_{k \in L}$. As indicated in Section 3.1, the mean Vol metric can be used directly as the consensus or cost function in this energy-based fusion model. In this framework, the consensus segmentation is simply obtained from the solution of the following optimization problem:

$$\widehat{S}_{\overline{\text{Vol}}} = \arg\min_{S \in S_n} \overline{\text{Vol}}(S, \{S_k\}_{k \leqslant L})$$
(5)

with S_n the set of all possible segmentations using *n* pixels. In this way, this fusion model is a generative model of *likely* (segmentation) solutions, which is able to give an estimate of \hat{S} , in the minimum $\overline{\text{Vol}}$ error sense. Let us note that this optimization approach is also called the *median partition* [38] with respect to both the mean Vol criterion (used in this application) and the segmentation ensemble $\{S_k\}_{k \leq L}$.

Let us note that the reason why minimizing Eq. (5) would result in a lower VoI value with respect to the (family) of ground-truth segmentations, stems directly from the theory of consensus or aggregation of clusterings (see Section 1). This theory explains all the interest of combining different and complementary data clusterings for a particular dataset. As already said in the Introduction Section, this consensus strategy can also efficiently be viewed and understood as a (special type of) denoising problem in which each segmentation (to be fused) is in fact a noisy segmentation solution (or a noisy observation), given by a clustering algorithm sensitive to the initial settings (i.e., number of classes, feature descriptors, color-space used, etc.). In estimation theory, an estimator based on the average (or the weighted average) operation generally yields to an optimal denoised solution (when the noise is uncorrelated). In the consensus theory, an interesting denoised segmentation solution would also be the average (or a consensus or compromise) of all the individual segmentations to be combined, according a defined criterion [38]. Concretely speaking, the noise affecting each segmentation solution (to be removed) would correspond to poor segments (i.e., poor segmentation result for some sub-parts of the input image in the segmentation ensemble) and the information (to be kept) would correspond to existing good segments in the segmentation ensemble. A median partition approach allows to estimate this average segmentation solution according a defined criterion.

Let us finally note that it is important to generate an appropriate segmentation ensemble, because the final result is necessarily conditioned by the initial clusterings to be fused. A given segmentation ensemble should ideally include, a lot of complementary information. The complementarity and the diversity are recommended, since the more varied the set of initial segmentations is, the more information for the consensus cost function is available [38].

3.3. Fusion model for image segmentation

The minimization of Eq. (5) allows to find the best compromise solution, which is also optimal according to the $\overline{\text{Vol}}$ criterion. This fusion procedure of multiple label fields or segmentations thus herein appears as a complex minimization problem for a $\overline{\text{Vol}}$ -based consensus function exhibiting several local minima across the lattice of possible clusterings S_n .

This complex consensus function can be efficiently minimized by the stochastic (and thus computationally costly) simulated annealing [15] algorithm which is both, insensitive to initialization and is guaranteed to find the global minima. Another choice is an iterative Gauss–Seidel-type relaxation scheme (also called Iterative Conditional Modes or ICM [46]), where pixels are updated one at a time. This iterative search technique is fast (since deterministic) and easy to implement, but has also the drawback of requiring a good initialization, not too far from the global minima (corresponding to the desired solution). In order to set the initialization M. Mignotte / Information Fusion 20 (2014) 7-20



Fig. 1. At iteration p, the pixel at site s initially belongs to the mth region (or segment) and is 4-connected with the xth region. If there is a (positive) decrease in \overline{Vol} energy when the new label x is assigned to this pixel; i.e., if $\Delta \overline{\text{Vol}}\left(\widehat{S}_{\overline{\text{Vol}}}^{[p]}, \{S_k\}_{k \leq L}\right)_{s:m \to x} > 0, \text{ then the label } m \text{ is replaced by the label } x \text{ at site } s.$

of the ICM properly to obtain a good solution, the ICM procedure is initialized with the $\widehat{S}_{\overline{\text{Vol}}}^{[0]}$ such as:

$$\widehat{S}_{\overline{\text{Vol}}}^{[0]} = \arg\min_{S \in \{S_k\}_{k \leq L}} \overline{\text{Vol}}(S, \{S_k\}_{k \leq L})$$
(6)

i.e., in initializing the ICM procedure, by the segmentation map (among the *L* segmentation results to be reconciled), ensuring the minimal consensus energy (in the mean VoI sense) of the proposed fusion model (Eq. (5)).³

Technically, ICM is an iterative steepest local energy descent algorithm which searches to obtain, for each pixel to be labeled, the minimum energy label assignment. Starting with $\widehat{S}_{\overline{Vol}}^{[0]}$, i.e., a solution not too far from the optimal solution (see Eq. (6)), ICM chooses, at each iteration and (sequentially) for each pixel, the label (of the final segmentation result), yielding the largest decrease of the cost function, conditioned on the labels assigned to its neighbors. In addition, to speed up the convergence of the ICM procedure, the important features of a natural image, especially the notion of homogeneity of the segmentation result (i.e., the spatial dependencies between spatial neighboring labels) and also the fact that a region, or cluster, in this application, necessarily form a set or group of connected pixels can also be efficiently exploited.

To this end, let the pixel at site *s* of $\widehat{S}_{Vol}^{[p]}$ (at iteration *p*) assumed to (initially) belong to region or cluster *m*. A new label is assigned to this pixel (e.g., the xth label) if three conditions are fulfilled:

- 1. $x \leq R$ where *R* is the number of regions in $\widehat{S}_{\overline{Vol}}^{[p]}$. 2. This pixel (at site *s*) is connected with the *x*-th cluster or region in $\widehat{S}_{\overline{\text{Vol}}}^{[p]}$.
- There is a decrease of the energy function $\overline{\text{Vol}}(.)$ (Eq. (5)), when 3. the pixel at site s (belonging to the *m*-th label or region) is assigned to the *x*th label (see Fig. 1).

In this case, the local decrease in the energy function $\overline{VoI}(.)$ when the above-mentioned two first conditions are fulfilled and for a new label assignment at site s, from label m to label x, can be written as follows (directly derived from Eqs. 4, 1, 2 and 3):

$$\begin{split} \Delta \overline{\mathrm{Vol}} \left(\widehat{S}_{\overline{\mathrm{Vol}}}^{[p]}, \{S_k\}_{k \leqslant L} \right)_{s:m \to x} &= L \cdot \left\{ -\frac{n_m}{n} \log\left(\frac{n_m}{n}\right) - \frac{n_x}{n} \log\left(\frac{n_x}{n}\right) \\ &+ \frac{(n_m - 1)}{n} \log\left(\frac{n_m - 1}{n}\right) + \frac{(n_x + 1)}{n} \log\left(\frac{n_x + 1}{n}\right) \right\} \\ &- 2 \cdot \sum_{l=1}^{L} \left\{ \frac{n_{m,\mathcal{L}_s^l}}{n} \log\left(\frac{n_{m\mathcal{L}_s^l}}{n} \cdot \frac{n}{n_m} \cdot \frac{n}{n_{\mathcal{L}_s^l}}\right) \\ &+ \frac{n_{\chi\mathcal{L}_s^l}}{n} \log\left(\frac{n_{\chi\mathcal{L}_s^l}}{n} \cdot \frac{n}{n_{\chi\mathcal{L}_s^l}}\right) \\ &- \frac{(n_{m\mathcal{L}_s^l} - 1)}{n} \log\left(\frac{(n_{m\mathcal{L}_s^l} - 1)}{n} \cdot \frac{n}{(n_m - 1)} \cdot \frac{n}{n_{\mathcal{L}_s^l}}\right) \\ &- \frac{(n_{\chi\mathcal{L}_s^l} + 1)}{n} \log\left(\frac{(n_{\chi\mathcal{L}_s^l} + 1)}{n} \cdot \frac{n}{(n_x + 1)} \cdot \frac{n}{n_{\mathcal{L}_s^l}}\right) \right\} \end{split}$$
(7)

where \mathcal{L}_{s}^{l} denotes the label at site *s* of the *l*-th segmentations ($l \leq L$) of the segmentation ensemble $\{S_k\}_{k \leq L}$ and it is important to recall that $n_{m\mathcal{L}_s^l}$ designates the number of pixels which are together in the *m*th cluster (or region) of the segmentation *S* and in the \mathcal{L}_{s}^{l} th cluster of the segmentation $S_l \in \{S_k\}_{k \leq L}$. Finally, the overall Vol-Based Fusion Model (VOIBFM) algorithm with the iterative steepest local energy descent strategy ensuring the connectivity constraint above-expressed and the minimum energy label assignment is outlined in pseudo-code in Algorithm 1.

Algorithm 1 VoI-Based Fusion Model VoI Mean VoI (See Equation (5)) $\{S_k\}_{k < L}$ Set of L segmentations to be fused $T_{\rm max}$ Maximal number of iterations (=40) 1. Initialization $\hat{S}^{[0]}_{\overline{\text{Vol}}} = \arg\min_{S \in \{S_k\}_{k \leq L}} \overline{\text{Vol}} \left(S, \{S_k\}_{k \leq L}\right)$ 2. Steepest Local Energy Descent while $p < T_{max}$ do for each pixel with label l_s at site s do • Let \mathcal{E} the set of labels $\neq l_s$ contained in the local (squared) fixed-size $(N_w = 7)$ neighborhood of s

• Draw a new label x according to the uniform distribution in the set \mathcal{E}

• if $x = \emptyset$ then continue;

• if pixel s with label x is not 4-connected with the x-th region in $\hat{S}_{\overline{vol}}^{[p]}$ then continue;

• Compute $\Delta \overline{\text{VoI}}\left(\hat{S}_{\overline{\text{VoI}}}^{[p]}, \{S_k\}_{k \leq L}\right)_{s:l_s \to x}$ (See Equation (7))

• If $\Delta \overline{\text{VoI}}\left(\hat{S}_{\overline{\text{VoI}}}^{[p]}, \{S_k\}_{k \leq L}\right)_{s:l_s \to x} > 0$ Then replace label l_s by label x at site s

 $p \leftarrow p + 1$

It is worth mentioning that, for the same final result, the global Vol distance (Eq. (4)) could be calculated instead of its local expression derived in Eq. (7), for each pixel of the consensus

³ Another reliable strategy consists of initializing the ICM with the first N_1 optimal input segmentations (in the mean Vol sense) and to finally retain, after convergence of the ICM procedure, the segmentation result ensuring the lowest mean Vol energy.

M. Mignotte/Information Fusion 20 (2014) 7–20



Fig. 2. Examples of segmentation ensemble and fusion result (algorithm VOIBFM). From lexicographic order. Three first rows of the image; *K*-means classification results (21 out of 60 *K*-means segmentations with different colors used to reflect different regions) for the segmentation model described in Section 4. Natural image (number n⁰134052) from the BSDS300 and final fusion result. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



Fig. 3. Examples of complexity values on some images of the BSDS300. From left to right, complexity value = 0.235, 0.364, 0.515, 0.661, 0.803 setting the number of classes (*K*) of the *K*-means clustering algorithm respectively to 2, 3, 4, 5 and 6 in the *K*-means segmentation model.



Fig. 4. First row and from left to right; a natural image from the BSDS300 ($n^{0}229036$) and the resulting fusion map estimated from its set of 7 input hand-labeled ground-truth segmentations [44] with their number of regions and their consensus energy (the two first outliers related to the two first maximal \overline{Vol} energy are indicated in bold).

segmentation, and then, used the same classification strategy consisting of choosing the label leading to the minimum global VoI distance, but this latter strategy would be done at a prohibitive computational cost, increased (approximately) by two orders of magnitude compared to the pixel-wise local optimization strategy using Eq. (7).

4. Segmentation ensemble generation

The initial segmentations, that this fusion framework searches to combine, are herein obtained by a *K*-means [47] clustering method, with 10 different color spaces (see [41] for a justification of these color spaces, references and more details) and 6 different feature modalities, for a total of 60 input segmentation results to be merged (see Fig. 2).

This generation process allows us to ensure the diversity required to obtain a good (i.e., reliable) segmentation ensemble on which the final result will be conditioned. It is worth mentioning that the more varied the set of segmentations is, the more information for the consensus function (on which the fusion model is based) is available [41,38]. Conversely, it is logic to think that a combination of similar segmentation solutions could not give an improved segmentation that outperforms the individual ensemble members.

Besides, the class number of the *K*-means (*K*) within this framework, is adaptively estimated, for each input image of the BSDS300, by using a distance quantifying its complexity and variability in terms of its total number of distinct and differentiable texture classes within the image. This metric, introduced in [48] ranges in [0, 1] (with a value of 1 for an input image with a large number of texture patterns) (see Fig. 3). More precisely,

$$K = 1 + \operatorname{ceil}(K_{\max} \times \operatorname{complexity value})$$
(8)

where ceil(*x*) denotes a ceiling function which rounds *x* up to the closest integer value and $K_{\text{max}} = 7$ herein somewhat represents an estimation of the maximum number of distinct texture classes present in a highly-textured input scene. We recall that the complexity value of an image is simply (as defined in [48]) the measure of the mean absolute deviation (with the L_1 norm) of the (coarsely quantized) color histogram of each overlapping squared (fixed-size 7×7) neighborhood contained within the image.

5. Experimental results

5.1. Initial tests

First, as in [41], the proposed fusion model has been tested on a segmentation ensemble of hand-segmented images associated to a Berkeley image. To keep this experiment interesting and instructive, the same Berkeley image proposed in [41] has been considered, i.e., a natural image showing wide variation across the set of hand-segmented images. The result (see Fig. 4) displays a consensus segmentation in term of level of details (or mean number of segments) compared to the ground-truth set, while being different to the resulting fusion map given by the fusion model proposed in [41], in the penalized probabilistic Rand index sense. It is also worth pointing out that the \overline{VoI} energy given by the considered combination framework could be afterward exploited to estimate the similarity and consistency of each ground-truth human segmentation comparatively to the other ground-truth segmentations. This Vol energy could thus be useful to identify possible outliers (i.e., segmentations that appear to be inconsistent with the remainder of the segmentation ensemble, according to a



Fig. 5. Example of fusion convergence result on six different initializations for the Berkeley image (n⁰134052). Left: initialization and Right: result at convergence of the proposed VOIBFM model. From top to bottom, the two input segmentations which have the best and the 5th (i.e., minimum $\overline{\text{Vol}}$ energy) $\overline{\text{Vol}}$ score (50 iterations for the gradient descent), the input segmentation which has the L/2 = 30th best $\overline{\text{Vol}}$ score and the input segmentation which has the worst $\overline{\text{Vol}}$ score (50 iterations) and two blind (or non informative) initializations (150 iterations).

M. Mignotte / Information Fusion 20 (2014) 7-20



Fig. 6. Evolution of the segmentation map (0th, 1st, 2nd, 5th, 9th, 13th, 22nd 50th, 80th, 85th) iterations and evolution of the consensus energy function along the number of iterations of the proposed VOIBFM fusion model starting from a blind (or non informative) initialization.

40

Iterations

50

60

70

80

30

defined criterion), in order either to remove them or to use this measure of confidence as a confidence prior, to assign higher fusion weight to certain input segmentations.

2.8

2.6

2.4 0

10

20

Now, for the tests, the proposed fusion method has been experimented, based on a segmentation ensemble $\{S_k\}_{k \leq L}$ with L = 60(see Section 4 and Fig. 2 for an example of segmentation ensemble). For the estimation of the number of classes K of this clustering procedure, it is important to recall that Eq. (8) (relying on the estimation of the complexity measure associated with each image) is herein used.

Second, the convergence of the iterative optimization procedure has been tested by taking, as initialization of the ICM-based iterative steepest local energy descent algorithm, respectively, the two input segmentations (of the segmentation ensemble $\{S_k\}_{k \leq L}$) which have the best (i.e., minimal) $\overline{\text{Vol}}$ score, the L/2 = 30th best score, the worst (i.e. maximal) Vol score and two blind (or non informative) initializations by considering an image spatially divided by K horizontal and vertical rectangles with K different labels (see Figs. 5 and 6). It can be seen that the consensus cost function is clearly non-convex and complex with many local minima. It can also be noted that the strategy, consisting in initializing the ICM procedure by the segmentation map that is closest to the solution (i.e., with the best $\overline{\text{Vol}}$ score) (Eq. (6)) seems a good initialization strategy even considering the fact that the segmentation map associated to the L/2 = 30th best Vol score (or equivalently the L/2 = 30th worst Vol score) gives a final resulting segmentation map close to the one given by the best \overline{VoI} score.

5.2. Evaluation

Throughout these experiments, the proposed fusion algorithm has been evaluated on the BSDS300 [49,44,50] in which the natural color images were downsampled to a size whose longest side is set to 320 pixels.

For a fair comparison, Figs. 7 and 8 present the segmentation results, provided by the proposed method, on the same subset of images (of the BSDS300) used (and shown) in [41] (based on a PRI-based segmentation combination model). The results for the whole dataset are freely accessible at the author's website (see Section 5.4).

To compare the efficiency of the proposed model to recent leading segmentation methods, four different and complementary performance measures have been computed. First, a reliable performance metric, very efficient for measuring the agreement between a region-based segmentation and multiple (hand-generated) ground-truths segmented from experts; namely the Probabilistic Rand Index⁴ (PRI) [51] (see Table 1). In fact, the PRI score simply measures the percentage (on average, by considering the set of ground truth images) of correctly classified pixel pairs. Additional comparisons have been performed with the VoI metric (see Section 3.1) and two other performance metrics, namely; the GCE [44] and the BDE [52] (see Table 2), demonstrating that the considered approach achieves sound results for several different and complementary clustering quality measures compared to the other previously published segmentation methods.

5.3. Comparison with existing methods and discussion

The proposed segmentation procedure gives the best PRI-based score amongst the leading algorithms proposed in the segmentation field (PRI=0.81 means that the proposed method ensures 81 percent of correctly classified pixel pairs, on average, in the BSDS300). Fig. 9 shows respectively the repartition of the local PRI score and some interesting spatial statistics about the size of

⁴ The Matlab source code, implemented by Yang (and publicly accessible on-line at address http://www.eecs.berkeley.edu/~yang/software/lossy_segmentation/) has been herein exploited.

M. Mignotte/Information Fusion 20 (2014) 7–20



Fig. 7. Example of segmentations obtained by the VOIBFM algorithm on several images of the BSDS300 (see also Tables 1 and 2 for quantitative performance measures and the author's website for the segmentation results on the entire database).

M. Mignotte/Information Fusion 20 (2014) 7–20



Fig. 8. Example of segmentations obtained by the VOIBFM algorithm on several images of the BSDS300 (see also Tables 1 and 2 for quantitative performance measures and the author's website for the segmentation results on the entire database).

segments (or regions) given by the VOIBFM algorithm over the BSDS300.

This segmentation method based on the fusion of multiple *imperfect* segmentations, produced from different features and color spaces, is also the best to date in term of VoI score, compared to

the best competing segmentation methods. To a certain extent, this is reassuring, since this performance score is also the criterion of this fusion model. Thus, as it was initially presupposed, a reliable fusion model, in the VoI sense, for combining multiple weak (and low-cost) segmentations $\{S_k\}_{k \in L}$ gives also a good VoI score

M. Mignotte/Information Fusion 20 (2014) 7-20

Table 1

Average PRI-based performance measure, for different segmentation (into regions) algorithms [53,41,3,29,5,2,48,24,7,20,40,27,6,13,28,4,26,54,30,9,25,21,55,14,23], (only those above PRI = 0.75 are herein referenced and are ranked in descending order of PRI-based performance value across the BSDS300). The bold values indicate the best score.

ALGORITHMS	PRI
HUMANS (in [4])	0.87
VOIBFM	0.81
-2011- gPb-owt-ucm [53]	0.81
-2010- PRIF [41]	0.80
-2008- CTex [3]	0.80
-2009- MIS [29]	0.80
-2011- SCKM [5]	0.80
-2008- FCR [2]	0.79
-2012- SFSBM [48]	0.79
-2004- FH [24] (in 4)	0.78
-2011- MD2S [7]	0.78
-2009- HMC [20]	0.78
-2009- Consensus [40]	0.78
-2009- Total Var. [27]	0.78
-2009- A-IFS HRI [6]	0.77
-2001- JSEG [13] (in [3])	0.77
-2011- KM [28]	0.76
-2007- CTM [4,14]	0.76
-2008- Av. Diss. [26] (in [53])	0.76
-2011- SCL [54]	0.76
-2005- Mscuts [30] (in [27])	0.76
-2002- Mean-Shift [9] (in [4])	0.75
-2008- NTP [25]	0.75
-2010- iHMRF [21]	0.75
-2005- NCuts [30] (in [53])	0.75
-2006- SWA [55] (in [53])	0.75

Table 2

Average performance measures, for different segmentation (into regions) algorithms [53,41,29,5,2,48,24,7,20,27,13,28,4,26,54,30,9,25,55,14,23], in terms of several complementary rate of good segmentation using a distance measure (smaller distance based scores are thus preferred) over the BSDS300. The bold values indicate the best score.

ALGORITHMS	VoI	GCE	BDE
HUMANS	1.10	0.08	4.99
VOIBFM	1.88	0.20	9.30
gPb-owt-ucm [53]	1.65	-	-
PRIF [41]	1.97	0.21	8.45
MIS [29]	1.93	0.19	7.83
SCKM [5]	2.11	0.23	10.09
MD2S [7]	2.36	0.23	10.37
FCR [2]	2.30	0.21	8.99
HMC [20]	3.87	0.30	8.93
Total Var. [27]	1.82	0.18	16.24
JSEG[13] (in [27])	2,32	0.20	14.40
SFSBM [48]	2.21	0.21	8.87
KM [28]	2.41	-	-
CTM [4],[14]	2.02	0.19	9.90
Av. Diss. [26] (in [53])	2.62	-	-
SCL [54]	2.14	0.18	9.38
MsCUtS [30] (in [27])	2.47	0.19	15.10
Mean-Shift [9] (in [4])	2.48	0.26	9.70
NTP[25](in[27])	2.49	0.24	16.30
NCIItS [23] (in [4])	2.93	0.22	9.60
FH [24] (in [4])	2.66	0.19	9.95
SWA [55] (in [53])	2.06	-	-

compared to the set of ground-truths. The proposed segmentation method has been evaluated and compared in a precision-recall framework (with the so-called F-measure) in order to assess its efficiency. First, it is important to note that the proposed algorithm is inherently disadvantaged, in this framework, for two reasons. First, it has been trained to optimize another criterion (i.e., the PRI instead of the F-measure criterion). Second, it is also important to remember that this precision-recall framework will inherently





Fig. 9. From top to bottom, distribution of the (1) PRI measure, (2) number and (3) size of regions over the 300 segmented images of the BSDS300.

Table 3

Average F-measure-based performance measure for different segmentation algorithms [9,30,55,56] on the Weizmann dataset (single and two objects database).

Algorithms	F-measure score		
	Single object database	Two objects database	
VOIBFM (<i>K</i> ≤ 4)	0.82 ± 0.016	0.77 ± 0.027	
VOIBFM (K with Eq. (8))	0.78 ± 0.015	0.76 ± 0.025	
AggProc [56]	0.86 ± 0.012	0.68 ± 0.005	
Mean-Shift [9] (in [56])	0.57±0.023	0.61±0.023	
NCuts [30] (in [56])	0.72±0.012	0.58 ± 0.059	
SWA V1 [55] (in [56])	0.83±0.016	$0.66 {\pm} 0.065$	
SWA V2 [55] (in [56])	$0.76 {\pm} 0.018$	-	

favor contour-based segmentation procedures with a "soft" contour map as output over region-based segmentation techniques giving a binary edge representation (as the proposed segmentation method). Indeed, this benchmark procedure will automatically estimate the optimal binary labeling corresponding to the best

M. Mignotte/Information Fusion 20 (2014) 7-20



Fig. 10. Segmentation results obtained by the proposed VOIBFM algorithm on the five first images of the Weizmann dataset (first and second row: single object dataset and third and fourth row: two-object dataset).



Fig. 11. Evolution of the PRI (higher is better) and Vol measures (lower is better) as a function of the number of segmentations (*L*) to be fused for the VOIBFM algorithm. More precisely for L = 1, 6, 12, 18, 24, ..., 60 segmentations (by considering first, one *K*-means segmentation and then by considering six segmentations for each color space and 1, 2, 3, ..., 10 color spaces). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

F-measure score [49] for the entire BSDS300 dataset. To overcome this latter disadvantage, it is more interesting that this benchmark measure itself chooses the optimal threshold on a soft boundary image provided by averaging several (6, in this test), binary contour maps as those given by the proposed scheme with 6 different seeds for the *K*-means based segmentation ensemble. In this case, the obtained score is *F*-measure = 0.61@(Recall = 0.59)

, Precision = 0.63) for the test set of the BSDS300, which is quite good in comparison to the other leading segmentation algorithms.

Experiments have also been conducted on the Weizmann Data-Set (WDS), which is composed of images that clearly depict one or two object/s in the foreground that differ from its surroundings [56]. It is important to note that the WDS is thus especially suited for segmentation algorithms *a priori* specifying that the number of

M. Mignotte/Information Fusion 20 (2014) 7-20



Fig. 12. From left to right, distribution of the (1) complexity measure for each image of the BSDS300 and (2) the distribution of the value of *K*, number of classes of the *K*-means clustering algorithm, for each segmentation of $\{S_k\}_{k \in L}$ (see (8) and Footnote 2).



Fig. 13. Evolution of the PRI (higher is better) as a function of the value of K_{max} for the proposed VOIBFM algorithm (L = 60).



Fig. 14. Evolution of the PRI score (higher is better) as a function of the value of the number of outliers (i.e., bad segmentations in the Vol sense) excluded from the segmentation ensemble (including initially 60 input segmentations).

classes is set to two (i.e., object and background) or set to three (i.e., two objects and background) whereas the proposed VOIBFM algorithm has been optimized (especially through Eq. (8)) on the BSDS300 for which, on average, there are 20 regions per image (and per human segmenter) [4]. Knowing this, the proposed VOI-BFM algorithm reports an F-measure score of 0.78 ± 0.015 for the segment who fits the best the foreground [56] on the WDS (single object dataset) and 0.82 ± 0.016 if the constraint $K \leq 4$ is added (*K* being the number of classes of the *K*-means, see Eq. (8)) which is competitive compared to the scores listed in [56]. For the two-object database, the final score is 0.76 ± 0.025 (and 0.77 ± 0.027 with the constraint $K \leq 4$) which is the highest averaged F-measure score (on the two-object WDS) presented in the website's author [56] (see Table 3 and Fig. 10).

The results of the experiment illustrated in Fig. 11 show that our fusion model has good asymptotic properties. Indeed, the PRI or Vol performance measures are even better than the size of the segmentation ensemble L is high. This test demonstrates the validity of the proposed fusion procedure and shows also that the performance scores are perfectible if the segmentation ensemble is completed by other (and different) segmentation maps (of the same scene).

Statistics on the distribution of the complexity measure for each image of the BSDS300 and the distribution of the value of K (number of classes of the K-means) for each segmentation of $\{S_k\}_{k \leq L}$ (see Eq. (8) and 2) are given in Fig. 12. Evolution of the PRI measure as a function of the value of K_{max} (see (8)) is also given in Fig. 13. It is worth mentioning that the proposed fusion model is a bit sensitive to this parameter K_{max} and this one has to be in the range [6,9] in order to keep a PRI score greater than or equal to 0.80.

In order to further confirm the efficiency of the VOIBFM fusion procedure, experiments have been performed with the combination of 60 mean shift based segmentations [9] achieved by randomly varying the parameters h_s and h_r within respectively the interval [2, ..., 15] and [20, ..., 50], in the HSV color space. Once again, the proposed fusion procedure shows a consistent improvement in performance compared to each individual segmentation (see Tables 1 and 2), and despite the fact that the mean shift algorithm does not take into account all the texture information. The final scores are: PRI = 0.795, VoI = 2.383, GCE = 0.153, BDE = 9.461.

Finally, the VOIBFM fusion procedure has also been experimented on a segmentation ensemble in which, iteratively, a certain number of outliers (i.e., bad input K-means based segmentations in the mean Vol sense) were excluded. Fig. 14 illustrates the evolution of the PRI score for a given value of the number of outliers excluded from the segmentation ensemble. It is important to notice that this strategy does not allow to improve the final segmentation result in terms of (average) PRI score. This experiment also shows us two important points. First, all the segmentations are, to some extent, important. Even a bad input segmentation map (in the mean Vol sense) remains informative and allows to improve the final average PRI-based good classification rate. Indeed, a bad segmentation may be globally bad, according to a defined criterion, and nevertheless able to produce some good segments (for some sub-parts of the image) which may be informative for the subsequent fusion procedure. Second, the strategy of excluding, from the fusion procedure, 59 outliers out of 60, i.e., to choose amongst the 60 segmentations to be fused, the one ensuring the minimal consensus energy, in the mean VoI sense (i.e., this is equivalent to just consider the initialization step of the proposed algorithm, see Eq. (6)), without gradient descent-based fusion process

(because, being the single element of the segmentation ensemble), allows to obtain an interesting PRI score (averaged over the BSDS300) equals to 0.79. This result also confirms that the proposed initialization strategy is reliable. The fusion process based on the steepest local energy descent, on this best segmentation (in the mean VoI sense), then allows to further improve this PRI score up to 0.81.

To finally summarize, the herein proposed fusion model has the advantage to be derived from a recent and appealing information Theory based criterion and be both reliable and easy and fast to optimize.

5.4. Algorithm

It takes respectively, on average, one minute for the segmentation ensemble generation and less than 15 s for the fusion step (for a 320×214 image and for a 2.2 GHz, AMD Athlon 64 Processor 3500+). It is important to note that the two aforementioned steps can easily be parallelized. It is straightforward for the generation of the segmentation ensemble but also true for the proposed fusion model by considering a Jacobi-type version of the Gauss–Seidel based ICM procedure [57].

The code, data, and all that is necessary for reproduction of the results are freely accessible on the author's website (directory: http://ResearchMaterial/voibfm.html).

6. Conclusion

Throughout this paper, the ability of a novel and reliable segmentation strategy, based on a general framework for consistently combining, in the VOI sense, multiple (simple) segmentation maps to produce a final improved partition result has been demonstrated. This new deterministic fusion procedure, based on the median partition approach, has a computational complexity linear in the number of pixels (and a fixed number of $T_{max} = 50$ iterations are sufficient to reach a convergence) and thus is scalable to large datasets. In addition, this fusion method can combine segmentation maps with a possibly different number of regions while being able to determining a different number of regions for each consensus segmentation map. This new segmentation method based on the framework of segmentation combination performs very favorably compared to other competing and currently available segmentation schemes. Besides, it remains easy to code, improvable (by considering and adding other spatial clusterings in the segmentation ensemble) and may be applied to many other problems dealing with label fields (such as 3D segmentation, 3D reconstruction, depth field or motion estimation). In addition, this free parameter fusion model is easily parallelizable and thus quite suitable for advanced processor architecture with multi-core technologies. For all these reasons, the proposed fusion method may therefore be seen as an attractive strategy for solving the difficult image segmentation problem.

References

- F. Destrempes, M. Mignotte, Localization of shapes using statistical models and stochastic optimization, IEEE Trans. Patt. Anal. Mach. Intell. 29 (9) (2007) 1603–1615.
- [2] M. Mignotte, Segmentation by fusion of histogram-based K-means clusters in different color spaces, IEEE Trans. Image Process. 17 (5) (2008) 780–787.
- [3] D.E. Ilea, P.F. Whelan, Ctex-an adaptive unsupervised segmentation algorithm based on color-texture coherence, IEEE Trans. Image Process. 17 (10) (2008) 1926–1939.
- [4] A.Y. Yang, J. Wright, S. Sastry, Y. Ma, Unsupervised segmentation of natural images via lossy data compression, Comput. Vision Image Understand. 110 (2) (2008) 212–225.
- [5] M. Mignotte, A de-texturing and spatially constrained K-means approach for image segmentation, Pattern Recog. Lett. 32 (2) (2011) 359–367.

- [6] M.M. Mushrif, A.K. Ray, A-IFS histon based multithresholding algorithm for color image segmentation, IEEE Signal Process. Lett. 16 (3) (2009) 168–171.
- [7] M. Mignotte, MDS-based multiresolution nonlinear dimensionality reduction model for color image segmentation, IEEE Trans. Neural Netw. 22 (3) (2011) 447–460.
- [8] M. Mignotte, MDS-based segmentation model for the fusion of contour and texture cues in natural images, Comput. Vision Image Understand. 116 (9) (2012) 981–990.
- [9] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 603–619.
- [10] Q. Luo, T. Khoshgoftaar, Unsupervised multiscale color image segmentation based on MDL principle, IEEE Trans. Image Process. 15 (9) (2006) 2755–2761.
- [11] M.A. Carreira-Perpinan, Fast nonparametric clustering with Gaussian blurring mean-shift, in: Proceedings of the 23rd International Conference on Machine Learning (ICML'06), vol. 148, Pittsburgh, Pennsylvania, USA, 2006, pp. 153– 160.
- [12] I. Mecimore, C.D. Creusere, Unsupervised bitstream based segmentation of images, in: Proceedings of the Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop 2009, Marco Island, FL, USA, 2009, pp. 643–647.
- [13] Y. Deng, B.S. Manjunath, Unsupervised segmentation of color-texture regions in images and video, IEEE Trans. Pattern Anal. Mach. Intell. 23 (8) (2001) 800– 810.
- [14] Y. Ma, H. Derksen, W. Hong, J. Wright, Segmentation of multivariate mixed data via lossy coding and compression, IEEE Trans. Pattern Anal. Mach. Intell. 29 (9) (2007) 1546–1562.
- [15] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Mach. Intell. 6 (6) (1984) 721–741.
- [16] M. Mignotte, C. Collet, P. Pérez, P. Bouthemy, Three-class Markovian segmentation of high resolution sonar images, Comput. Vision Image Understand. 76 (3) (1999) 191–204.
- [17] M. Mignotte, C. Collet, P. Pérez, P. Bouthemy, Sonar image segmentation using an unsupervised hierarchical MRF model, IEEE Trans. Image Process. 9 (7) (2000) 1216–1231.
- [18] F. Destrempes, M. Mignotte, J.-F. Angers, A stochastic method for Bayesian estimation of hidden Markov models with application to a color model, IEEE Trans. Image Process. 14 (8) (2005) 1096–1108.
- [19] F. Destrempes, J.-F. Angers, M. Mignotte, Fusion of hidden Markov random field models and its Bayesian estimation, IEEE Trans. Image Process. 15 (10) (2006) 2920–2935.
- [20] H. Rachid, M. Mignotte, A hierarchical graph-based Markovian clustering approach for the unsupervised segmentation of textured color images, in: Proceedings of the 16th IEEE International Conference on Image Processing (ICIP'09), Cairo, Egypt, 2009, pp. 1365–1368.
- [21] S. Chatzis, G. Tsechpenakis, The infinite hidden Markov random field model, IEEE Trans. Neural Netw. 21 (6) (2010) 1004–1014.
- [22] S. Chen, L. Cao, Y. Wang, Image segmentation by ML-MAP estimations, IEEE Trans. Image Process. 19 (9) (2010) 2254–2264.
- [23] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 888–905.
- [24] P. Felzenszwalb, D. Huttenlocher, Efficient graph-based image segmentation, Int. J. Comp. Vision 59 (2) (2004) 167–181.
- [25] 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 (CVPR'08), Anchorage, Alaska, USA, 2008, pp. 1–8.
- [26] L. Bertelli, B. Sumengen, B. 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.
- [27] M. Donoser, M. Urschler, M. Hirzer, H. Bishof, Saliency driven total variational segmentation, in: Proceedings of the 12th IEEE International Conference on Computer Vision (ICCV'09), Kyoto, Japan, 2009, pp. 817–824.
- [28] 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.
- [29] M. Krinidis, I. Pitas, Color texture segmentation based on the modal energy of deformable surfaces, IEEE Trans. Image Process. 7 (18) (2009) 1613–1622.
- [30] T. Cour, F. Benezit, J. Shi, Spectral segmentation with multiscale graph decomposition, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 2, San-Diego, California, USA, 2005, pp. 1124–1131.
- [31] Y. Jiang, Z.-H. Zhou, SOM ensemble-based image segmentation, Neural Process. Lett. 20 (3) (2004) 171–178.
- [32] J. Keuchel, D. Kuttel, Efficient combination of probabilistic sampling approximations for robust image segmentation, in: 28th Annual Symposium of the German Association for Pattern Recognition, DAGM-Symposium, Lecture Notes in Computer Science, Berlin, Germany, 2006, pp. 41–50.
- [33] P. Wattuya, K. Rothaus, J.-S. Prani, X. Jiang, A random walker based approach to combining multiple segmentations, in: Proceedings of the 19th International Conference on Pattern Recognition (ICPR'08), Tampa, Florida, USA, 2008, pp. 1–4.
- [34] A.J. Sharkey, Combining Artificial Neural Nets Ensemble and Modular Multi-Net Systems, Springer-Verlag, New York, Inc., 1999. ISBN:185233004X.
- [35] T. Dietterich, Ensemble methods in machine learning, in: L.N.I.C. Science (Ed.), Proceedings of the First International Workshop on Multiple Classifier Systems, LNCS, Multiple Classifier Systems, vol. 1857, Springer, 2000, pp. 1–15.

Author's personal copy

M. Mignotte/Information Fusion 20 (2014) 7-20

- [36] A. Strehl, J. Ghosh, C. Cardie, Cluster ensembles a knowledge reuse framework for combining multiple partitions, J. Mach. Learn. Res. 3 (December) (2002) 583–617.
- [37] A. Fred, A. Jain, Data clustering using evidence accumulation, in: Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02), Quebec City, Canada, 2002, pp. 276–280.
- [38] S. Vega-Pons, J. Ruiz-Shulcloper, A survey of clustering ensemble algorithms, Int. J. Pattern Recog. Artif. Intell. 25 (3) (2011) 337–372.
- [39] W.M. Rand, Objective criteria for the evaluation of clustering methods, J. Am. Statist. Assoc. 66 (336) (1971) 846–850.
- [40] S. Ghosh, 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, (WACV'09), Snowbird, Utah, USA, 2009, pp. 1–8.
- [41] 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.
- [42] M. Meila, Comparing clusterings an axiomatic view., in: Proceedings of the 2005 22nd International Conference on Machine Learning (ICML'05), Bonn, Germany, 2005, pp. 577–584.
- [43] M. Meila, Comparing clusterings-an information based distance, J. Multivar. Anal. 98 (5) (2007) 873-895.
- [44] 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 (ICCV'01), vol. 2, Vancouver, British Columbia, Canada, 2001, pp. 416–423.
- [45] R. Unnikrishnan, C. Pantofaru, M. Hebert, Toward objective evaluation of image segmentation algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 29 (6) (2007) 929–944.
- [46] J. Besag, On the statistical analysis of dirty pictures, J. R. Statist. Soc. B-48 (1986) 259–302.

- [47] S.P. Lloyd, Least squares quantization in PCM, IEEE Trans. Inform. Theory 28 (2) (1982) 129–136.
- [48] M. Mignotte, A non-stationary MRF model for image segmentation from a soft boundary map, Pattern Anal. Appl. (2012) 1–11.
- [49] D. Martin, C. Fowlkes, The Berkeley Segmentation Database and Benchmark, Image Database and Source Code Publicly http://www.cs.berkeley.edu/ projects/vision/grouping/segbench/>.
- [50] D. Martin, C. Fowlkes, J. Malik, Learning to detect natural image boundaries using local brightness, color and texture cues, IEEE Trans. Pattern Anal. Mach. Intell. 26 (5) (2004) 530–549.
- [51] 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 (CVPR'05), Workshop on Empirical Evaluation Methods in Computer Vision, vol. 3, San Diego, California, USA, 2005, pp. 34–41.
- [52] J. Freixenet, X. Munoz, 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 (ECCV'02), Copenhagen, Denmark, 2002, p. III: 408 ff.
- [53] 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.
- [54] R. Huanga, N. Sangb, D. Luoc, Q. Tangd, Image segmentation via coherent clustering in L^AB^{*} color space, Pattern Recog, Lett. 32 (7) (2011) 891–902.
- [55] E. Sharon, M. Galun, D. Sharon, R. Basri, A. Brandt, Hierarchy and adaptivity in segmenting visual scenes, Nature 442 (17) (2006) 810–813.
- [56] S. Alpert, M. Galun, R. Basri, A. Brandt, Image segmentation by probabilistic bottom-up aggregation and cue integration, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07), Minneapolis, Minnesota, USA, 2007, pp. 1–8.
- [57] P.-M. Jodoin, M. Mignotte, Markovian segmentation and parameter estimation on graphics hardware, J. Electron. Imag. 15 (3) (2006) 033015-1–033015-15.

20