A NEW MULTI-CRITERIA FUSION MODEL FOR COLOR TEXTURED IMAGE SEGMENTATION

Lazhar Khelifi, Max Mignotte

Image Processing Laboratory, DIRO, University of Montreal, Montreal, Canada Email: khelifil@iro.umontreal.ca, mignotte@iro.umontreal.ca

ABSTRACT

Fusion of image segmentations using consensus clustering and based on the optimization of a single criterion (commonly called the median partition based approach) may bias and limit the performance of an image segmentation model. To address this issue, we propose, in this paper, a new fusion model of image segmentation based on multi-objective optimization which aims to avoid the bias caused by a single criterion and to achieve a final improved segmentation. The proposed fusion model combines two conflicting and complementary segmentation criteria, namely; the region-based variation of information (VoI) criterion and the contour-based F-Measure (precision-recall) criterion with an entropy-based confidence weighting factor. To optimize our energy-based model we use an optimization procedure derived from the iterative conditional modes (ICM) algorithm. The experimental results on the Berkeley database with manual ground truth segmentations clearly show the effectiveness and the robustness of our multi-objective median partition based approach.

1. INTRODUCTION

Image segmentation aims to divide an image into several disjoint regions with uniform and homogeneous attributes. Combining multiple, weak and quickly estimated segmentation maps of the same image to obtain a final improved segmentation has become an important since competitive approach, over the last few years, to efficiently solve the non-trivial problem of unsupervised segmentation of textured natural images.

Many different fusion segmentation approaches based on various criteria have been proposed until now. Among them, we can mention the fusion model which combines the individual putative segmentations (or clustering results) in the evidence accumulation sense [1], in the within-cluster inertia (or variance) criterion [2] (for the set of neighboring pixel labels), or in the probabilistic version of the Rand index criterion [3]. The fusion of weak segmentations can also be carried out according to the optimal or maximum-margin sense (of the hyperplanes between classes) [4] or recently, in the weights of evidence sense [5] (for satellite image segmentation) to name a few.

The most common way to compute the consensus segmentation from the set of weak segmentation maps is based on the so-called median partition based approach which in fact, consists of finding the consensus solution which minimizes, according a given criterion (which is also expressed as a distance between two segmentations), the sum of the distances separating the (consensus) solution from the other segmentations to be fused. This above-mentioned criterion defines both the most appropriate fusion model of segmentations but also (to a lesser extent) all the intrinsic properties of the consensus segmentation map to be estimated. When this optimization problem is based on the optimization of a single criterion, the fusion procedure is inherently biased towards one particular objective, i.e., some specific regions of the search space containing solutions a priori defined (by the criterion) as acceptable solutions. This may bias and limit the performance of an image segmentation model since it is difficult to find a generic single criterion that both defines the most appropriate fusion model of segmentations and all the complex intrinsic geometric properties of a consensus segmentation. To avoid the bias caused by a single criterion, an alternative consists in using approaches based on multi-objective optimization in order to design a new fusion segmentation model that takes advantage of the complementary of different objectives (criteria) allowing to achieve a final better consensus segmentation.

In this paper, we present a new multi-criteria fusion model weighted by an entropy-based confidence measure (EFA-BMFM). This model aims to combine and optimize simultaneously two different and complementary segmentation fusion criteria; namely the (region-based) variation of information (VoI) criterion and the (contour-based) F-Measure (derived from the precision-recall) criterion.

2. MULTI-OBJECTIVE OPTIMIZATION

Unlike mono-objective optimization, which consists in minimizing a single objective function with respect to a set of parameters, in the multi-objective optimization (MO) case, there are several often conflicting objectives to be simultaneously optimized [6]. Broadly, there are two general approaches of MO [7]. The first, is called the Pareto approach (PTA) whose basic idea is to determine a set of solutions that are non-dominated with respect to each objective. The second (adopted in our work), is called the weighted formula approach (WFA). The main goal of the WFA is to transform a MO problem into a problem with a single objective function. This is typically done by assigning a numerical weight to each objective (evaluation criterion) and then combining the values of the weighted criteria into a single value by either adding all the weighted criteria. Mathematically, the quality Z of a given candidate model is typically given by the following formula:

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

where *n* is the number of evaluation criteria, and w_i denotes the weights (assigned to criterion c_i). The estimation of the weights (also known as importance factors) is crucial and should depend both of the degree of information or confidence one has in the ensemble of segmentations (to be fused), provided by each criterion and also should take into account the difference of scaling between these two criteria (expressed as a distance between a pair of segmentations). This rescaling is important in order to avoid giving too much importance to one of the two criteria (thus making the fusion of the two criteria, totally ineffective). To address this problem we propose a new entropy-based confidence measure (see Section 4.3).

3. SEGMENTATION TO BE FUSED

The initial segmentation maps which will be fused together are simply generated, in our application, as in [3], by a simple *K*-means clustering technique (with the Euclidean similarity distance and a stochastic initial seed) represented in 12 different color spaces (i.e., RGB, HSV, LAB, YCbCr, TSL, YIQ, XYZ, h123, P1P2, HSL, i123, LUV) with three different values of the number of classes K^1 , and two different values of the number of bins ($N_b \in \{4^3, 5^3\}$) for each local requantized color histogram (used as a feature vector for the *K*-means) computed on an overlapping squared fixed-size ($N_w = 7$) neighborhood centered around the pixel to be classified for a total of $12 \times (3 + 2) = 60$ input segmentations (see [3] for a justification of these color spaces and for reference).

4. FUSION OF SEGMENTATION

4.1. VoI criterion

The variation of information (VoI) metric (or criterion) [8] 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 (taking a value of 0 when two clusterings)

are identical and positive but ≤ 1 otherwise). Equivalently, it also represents roughly the amount of randomness in one segmentation which cannot be explained by the other [9].

Formally, let $S^a = \{C_1^a, C_2^a, \ldots, C_{R^a}^a\}$ & $S^b = \{C_1^b, C_2^b, \ldots, \ldots, C_{R^b}^b\}$ be respectively the machine segmentation to be measured between the manually segmented image and R^a being the number of segments or regions (C) in S^a and R^b the number of regions in S^b . The VoI distance between S^a and S^b is defined as follows:

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

where $H(S^a)$ and $H(S^b)$ represent respectively the entropy associated with the segmentation S^a and S^b and $I(S^a, S^b)$ the mutual information between these two spatial partitions. Let n be the number of pixels within the image, n_i^a 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 . The entropy is always positive (it is null only when there is no uncertainty, namely when there is only one cluster) and is defined as follows:

$$H(S^{a}) = -\sum_{i=1}^{R^{a}} P(i) \log P(i) = -\sum_{i=1}^{R^{a}} \left(\frac{n_{i}^{a}}{n}\right) \log \left(\frac{n_{i}^{a}}{n}\right)$$
(3)

$$H(S^{b}) = -\sum_{j=1}^{R^{b}} P(j) \log P(j) = -\sum_{j=1}^{R^{b}} (\frac{n_{j}^{b}}{n}) \log (\frac{n_{j}^{b}}{n})$$
(4)

with $P(i) = n_i^a/n$ being 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 taking respectively R^a and R^b values and uniquely associated to the partition S^a and S^b . Let now $P(i,j) = n_{ij}/n$ represents the probability that a pixel belongs to C_i^a and to C_j^b , the mutual information between the partitions S^a and S^b is equal to the mutual information between the random variables *i* and *j* and is expressed in the following way:

$$I(S^{a}, S^{b}) = \sum_{i=1}^{R^{a}} \sum_{j=1}^{R^{b}} P(i,j) \log\left(\frac{P(i,j)}{P(i).P(j)}\right)$$
(5)

4.2. F-measure criterion

The F-measure is in fact a combination of two complementary measures; precision and recall, which are commonly used by information retrieval theorists and practitioners [10]. In the (contour-based) image segmentation case, these two scores represent, respectively, the fraction of detections that are true boundaries and the fraction of the true boundaries detected [11]. In the one hand, the precision measure is low when there is significant over-segmentation, or when a large number of

¹To estimate the number of classes K (for each input image), we use an interesting metric which measures the image complexity (based on the number of its different texture classes) (see [3]).

boundary pixels have poor localization. On the other hand, a low recall value is typically the result of under-segmentation and indicates failure to capture the salient image structure. A particular application can define a relative cost α between these two quantities, which focuses attention on a specific point on the precision-recall curve [12]. In this case, the F-Measure is then defined as:

$$F_{\alpha} = \frac{PR}{\alpha R + (1 - \alpha)P} \tag{6}$$

and is within the range [0, 1] where a score equals to 1 indicates that two segmentations are identical (i.e., with identical contours).

4.3. Multi-objective function

The VoI and F-measure metrics (presented in Section 4.1 and Section 4.2), are now commonly used in segmentation [3, 12, 13] as two (complementary) objective measures for the evaluation of an automatic segmentation (i.e., given by an algorithm) compared to a set of ground truth segmentations given by different human experts and capturing, in fact, the inherent variability of each possible (perceptually consistent) interpretation of an input image, segmented at different detail levels by each human segmenter.

More precisely, let $\{S_k^b\}_{k \leq L} = \{S_1^b, S_2^b, \ldots, S_L^b\}$ be a finite ensemble of L manually ground truth segmented images of the same scene (segmented by L different human experts at possibly different levels of details) and S^a be the spatial clustering result to be evaluated by comparison with the manually labeled set $\{S_k^b\}_{k \leq L}$. The mean VoI and the mean F-measure metrics are simply the two metrics which take into account this set of possible ground truth segmentations, i.e.:

$$\overline{\mathbf{C}}\left(S^{\mathbf{a}}, \{S^{b}_{k}\}_{k \leq L}\right) = \frac{1}{L} \sum_{k=1}^{L} \mathbf{C}\left(S^{\mathbf{a}}, S^{b}_{k}\right) \tag{7}$$

with $C = \{\text{VoI}, F_{\alpha}\}$. Concretely, the $\overline{\text{VoI}}$ distance function will give a low value (conversely, the \overline{F}_{α} measure function will give a high value) to a segmentation result S^{a} that is all the more in accordance with the set of the segmentation maps obtained from human experts.

Let us consider now that we have a family of L segmentations $\{S_k\}_{k \leq L} = \{S_1, S_2, \ldots, S_L\}$ (associated with a same scene) to be fused in order to obtain a final improved segmentation result \hat{S} (more accurate than each individual member of $\{S_k\}_{k \leq L}$). These two complementary criteria; namely the contour-based F-measure and the region-based VoI measure can be used directly as a multi-objective cost function in an energy based model. In this context, the consensus segmentation \hat{S}_{MO} is simply obtained as the result of the following

bi-criteria optimization problem:

$$\hat{S}_{\overline{MO}} = \arg\min_{S \in \mathcal{S}_n} \overline{MO}\left(S, \{S_k\}_{k \le L}\right) \quad \text{with}: \qquad (8)$$

$$\overline{\mathrm{MO}}(.) = \mathrm{w}_{\mathrm{Vol}} \overline{\mathrm{Vol}}(S, \{S_k\}_{k \le L}) + \frac{\mathrm{w}_{\mathrm{F}}}{\overline{\mathrm{F}}_{\alpha}(S, \{S_k\}_{k \le L})} \quad (9)$$

where the importance (or weighting) factors w_{vol} and w_F must be estimated in order to take into account firstly, the possible difference of scaling existing between these two metrics and secondly, the confidence one has in the ensemble of segmentations (to be fused) (see Section 2). In our case, the first requirement is fulfilled because our two criteria are inherently normalized within the range [0, 1]. For the second requirement, let us note that the confidence one has in the ensemble of segmentations (or the confidence provided by each criterion) is directly related to the degree of variation (measure of disorder or uncertainty) of each (distance or) criterion in the ensemble of *L* segmentation to be fused. This degree of uncertainty can be quantified by computing the entropy weight of each criterion in the following way (for the VoI criterion)²:

$$\mathbf{e}_{\text{vol}} = -D \sum_{i=1}^{L} \left\{ P(\overline{\text{VoI}}(.)) \log P(\overline{\text{VoI}}(.)) \right\}$$
(10)

with $D = 1/\log(L)$. It is important to note that, the smaller the entropy is, the greater the amount of information (or confidence) provided by the criteria we have, and consequently, the weight of the criteria should be larger. To this end, the weights are then normalized in the following way:

$$W_{\text{vol}} = \frac{1 - e_{\text{vol}}}{(1 - e_{\text{vol}}) + (1 - e_{F_{\alpha}})} \in [0, 1] \quad (11)$$

$$\mathbf{W}_{\mathbf{F}_{\alpha}} = \frac{1 - \mathbf{e}_{F_{\alpha}}}{(1 - \mathbf{e}_{\text{vol}}) + (1 - \mathbf{e}_{\mathbf{F}_{\alpha}})} \in [0, 1] \quad (12)$$

4.4. Optimization of the Fusion Model

In order to solve this consensus function, in the bi-criteria sense, we use the set of superpixels [15] (the segments ensemble or regions provided by each individual segmentations to be fused) along with a fast search technique, called the iterative conditional modes (ICM), proposed by Besag [16], i.e.; a Gauss-Seidel relaxation where (in our case super-) pixels are updated one at a time.

This algorithm requires a proper initialization (otherwise, it will converge towards a bad local minima). To this end, we resort once again to the entropy values of each criteria [see (10)] by first selecting the criteria which gives the minimal entropy (i.e., the most informative criterion, see Section 4.3) and then by choosing, for the first iteration of the ICM, among the *L* segmentation to be fused, the one ensuring the minimal

²Note that $P(\overline{\text{VoI}}(.)) = \overline{\text{VoI}}(.,.) / \sum_{i=1}^{L} \overline{\text{VoI}}(.,.)$, this normalization step makes sure that each criterion value is limited between 0 and 1 [14].



Fig. 1. Example of fusion convergence result on three various initializations for the Berkeley image (n0 229036). Top: initialization and Bottom: segmentation result after 10 iterations of our EFA-BMFM fusion model. From left to right, a non informative (or blind) initialization, the worst and the best input segmentation (from the segmentation set) selected by the entropy method (see Section 4.4).

consensus energy (in this selected criterion sense) of our fusion model. This iterative algorithm amounts to obtaining simultaneously, for each superpixel to be labeled, the minimum value of 9.

5. EXPERIMENTS

To evaluate the efficiency of our fusion model we validate our approach on the famous Berkeley segmentation database (BSD300) [17].

First of all, we have tested the convergence properties of our ICM procedure based on superpixels by choosing, as initialization of our iterative local gradient descent algorithm, various initializations (extracted from our segmentation ensemble $\{S_k\}_{k \leq L}$). We have noticed that, the final energy value along with the resulting final segmentation map, is on average, the better when the initial segmentation solution is associated to an initialization chosen by our entropy-based method (see Section 4.4) (while remaining robust with other initializations). Consequently, the combination of the use the superpixels of $\{S_k\}_{k \leq L}$ along with a good initialization strategy definitely gives good convergence properties for our fusion model (see Fig. 1).

In order to test and compare our segmentation model, we adopt four performance metrics which are most popular in the literature. These performance measures include the VoI [8], the GCE [17] and the BDE [18] distance (for which a lower distance is better) and the PRI [19] (for which a higher measure is better) showing that our method gives competitive results compared to state-of-the art (see Table 1). Also, if we

Table 1. Performance of several segmentation algorithms (with or without a fusion model strategy) for four different performance measures: VoI, GCE, BDE (lower is better), and PRI (higher is better) on the BSD300.

ALGORITHMS	VoI	GCE	BDE	PRI
HUMANS	1.10	0.08	4.99	0.87
EFA-BMFM	1.87	0.19	8.16	0.80
-2014-FMBFM [11]	2.01	0.20	8.49	0.80
-2014-VOIBFM [20]	1.88	0.20	9.30	0.81
-2011-gPb-owt-ucm [21]	1.65	-	-	0.81
-2010-PRIF [3]	1.97	0.21	8.45	0.80
-2008-FCR [2]	2.30	0.21	8.99	0.79
-2007-CTM [13]	2.02	0.19	9.90	0.76
-2014-CRKM [22]	2.35	-	-	0.75
-2002-Mean-Shift [23]	2.48	0.26	9.70	0.75
-2005-NCuts (in [13])	2.93	0.22	9.60	0.72



Fig. 2. Segmentation results obtained by our algorithm EFA-BMFM on three different images of the Berkeley image dataset.

compare our results to a mono-objective approach (FMBFM and VOIBFM) based in the same single criterion, we obtain significantly better results (see Table 1). 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 (see Fig. 2).

6. CONCLUSION

This paper presents a novel and efficient multi-criteria fusion model based on the entropy-weighted formula approach (EFA-BMFM), whose goal is to combine 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). Applied on the BSD300, the proposed segmentation model gives competitive results compared to other segmentation models which proves the effectiveness and the robustness of our multi-criteria fusion approach.

7. REFERENCES

- A. Fred and A. Jain, "Data clustering using evidence accumulation," in *In Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02)*, August 2002, pp. 276–280.
- [2] M. Mignotte, "Segmentation by fusion of histogram-based Kmeans clusters in different color spaces," *IEEE Trans. Image Processing*, vol. 17, pp. 780–787, 2008.
- [3] M. Mignotte, "A label field fusion Bayesian model and its penalized maximum Rand estimator for image segmentation," *IEEE Trans. Image Processing*, vol. 19, no. 6, pp. 1610–1624, 2010.
- [4] X. Ceamanos, B. Waske, J. A. Benediktsson, J. Chanussot, M. Fauvel, and J. R. Sveinsson, "A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data," *International Journal of Image and Data Fusion*, vol. 1, no. 4, pp. 293–307, 2010.
- [5] B. Song and P. Li, "A novel decision fusion method based on weights of evidence model," *International Journal of Image and Data Fusion*, vol. 5, no. 2, pp. 123–137, 2014.
- [6] L. Khelifi, I. Zidi, K. Zidi, and K. Ghedira, "A hybrid approach based on multi-objective simulated annealing and tabu search to solve the dynamic dial a ride problem," in *International Conference on Advanced Logistics and Transport (ICALT)*, 2013, May 2013, pp. 227–232.
- [7] B. Wei and R. Mandava, "Multi-objective nature-inspired clustering techniques for image segmentation," in *IEEE Conference* on Cybernetics and Intelligent Systems (CIS), 2010, June 2010, pp. 150–155.
- [8] M. Meila, "Comparing clusterings-an information based distance," *Journal of Multivariate Analysis*, vol. 98, no. 5, pp. 873– 895, 2007.
- [9] M. Meil, "Comparing clusterings: an axiomatic view," in In ICML 05: Proceedings of the 22nd international conference on Machine learning. ACM Press, 2005, pp. 577–584.
- [10] G. Salton and M. J. McGill, Introduction to Modern Information Retrieval. New York, NY, USA: McGraw-Hill, Inc., 1986.
- [11] C. Hélou and M. Mignotte, "A precision-recall criterion based consensus model for fusing multiple segmentations," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 7, no. 3, pp. 61–82, July 2014.
- [12] D. Martin, C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color and texture cues," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 5, pp. 530–549, May 2004.
- [13] A. Y. Yang, J. Wright, S. Sastry, and Y. Ma, "Unsupervised segmentation of natural images via lossy data compression," *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 212– 225, May 2008.
- [14] T. C. Wang and H. D. Lee, "Developing a fuzzy TOPSIS approach based on subjective weights and objective weights," *Expert Systems with Applications*, vol. 36, no. 5, pp. 8980–8985, 2009.

- [15] Z. Li, X.-M. Wu, and S.-F. Chang, "Segmentation using superpixels: A bipartite graph partitioning approach," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 789–796.
- [16] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society*, vol. B-48, pp. 259–302, 1986.
- [17] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int'l Conf. Computer Vision (ICCV'01)*, vol. 2, July 2001, pp. 416–423.
- [18] J. Freixenet, X. Munoz, D. Raba, J. Marti, and X. Cufi, "Yet another survey on image segmentation: Region and boundary information integration," in *Proc. 7th European Conference on Computer Vision (ECCV02)*, 2002, p. III: 408 ff.
- [19] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "A measure for objective evaluation of image segmentation algorithms," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Workshop on Empirical Evaluation Methods in Computer Vision*, vol. 3, June 2005, pp. 34–41.
- [20] M. Mignotte, "A label field fusion model with a variation of information estimator for image segmentation," *Information Fusion*, vol. 20, no. 0, pp. 7 – 20, 2014.
- [21] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.
- [22] M. Ben Salah, I. Ben Ayed, J. Yuan, and H. Zhang, "Convexrelaxed kernel mapping for image segmentation," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1143–1153, March 2014.
- [23] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 5, pp. 603–619, 2002.