GCE-BASED MODEL FOR THE FUSION OF MULTIPLES COLOR IMAGE SEGMENTATIONS

Lazhar Khelifi, Max Mignotte

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

ABSTRACT

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

Index Terms— Color textured image segmentation, combination of multiple segmentations, global consistency error.

1. INTRODUCTION

Combining multiple, quickly estimated (and eventually weak) segmentation maps of the same image to obtain a final refined segmentation has become a promising approach, over the last few years, to efficiently solve the non-trivial problem of unsupervised segmentation of textured natural images.

This approach has been first proposed in [1] with a constraint specifying that all input segmentations (to be fused) must be composed of the same regions number and soon after, without this specific constraint, with an arbitrary number of regions, in [2] and combining the individual putative segmentations which minimizes the within-point scatter of each cluster members. This fusion of segmentations can also be carried out according to the probabilistic version of the well-known Rand [3] index (PRI) criterion [4] or according to the evidence accumulation sense [5], Finally, we can also mention the fusion scheme proposed in [6] in the optimal or maximum-margin hyperplane (between classes) sense or the recent Bayesian fusion procedure for satellite image segmentation proposed in [7] in which the class labels obtained from different segmentation maps are fused by the weights of evidence model.

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

2. PROPOSED FUSION MODEL

2.1. The GCE Measure

In our fusion model we use the global consistency error (GCE) [8] criterion which measures the extent to which one segmentation map can be viewed as a refinement of another segmentation. It was recently proposed in image segmentation [9] as a quantitative and interesting metric to compare machine segmentations of an image database to their respective manually segmented images given by human experts.

Formally, let $S^{\mathrm{t}} = \{C^{\mathrm{t}}_1, C^{\mathrm{t}}_2, \ldots, C^{\mathrm{t}}_{R^{\mathrm{t}}}\}$ & $S^{\mathrm{g}} = \{C^{\mathrm{g}}_1, C^{\mathrm{g}}_2, \ldots, C^{\mathrm{g}}_{R^{\mathrm{g}}}\}$ be respectively the segmentation test result to be measured and the manually segmented image and R^{t} being the number of segments or regions (C) in S^{t} and R^{g} the number of regions in S^{g}. We consider, for a particular pixel p_i, the segments in S^{t} and S^{g} including this pixel. We denote these segments by $C^{\mathrm{t}}_{<p_i>}$ and $C^{\mathrm{g}}_{<p_i>}$ respectively. If one segment is a subset of the other, so the pixel is practically included in the refinement area, and the local error should be equal to zero. If there is no subset relationship, then the two regions overlap in an inconsistent way and the local error ought be different from zero [8]. The local refinement error (LRE) is therefore denoted at pixel p_i as:

$$\mathrm{LRE}(S^{t}, S^{g}, p_i) = \frac{|C^{t}_{<p_i>} \backslash C^{g}_{<p_i>}|}{|C^{t}_{<p_i>}|} \tag{1}$$

where $\backslash$ represents the set differencing operator and $|C|$ the cardinality of the set of pixels C. As noticed in [8], this segmentation error measure is not symmetric and encodes a measure of refinement in only one sense. A possible and natural way to combine the LRE at each pixel into a measure for the whole image is the so-called global consistency error (GCE) which constraints all local refinement to be in the same sense in the following way:

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

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

Although a fundamental problem with the GCE measure is that there are two degenerate segmentation cases that give an aberrantly high score value [8]. These two degenerative segmentations are the two following trivial cases; one pixel per region and one region covering the whole image. The former is, in fact, a detailed refinement of any segmentation, and any segmentation is a refined improvement of the latter.

In our application, in order to be able to define an energy-based fusion model, avoiding the two above-mentioned degenerate segmentation cases, and for which a reliable consensus or compromise resulting segmentation map would be solution, *via* an optimization scheme (see Section 2.2), we have replaced the minimum operator in the GCE by the average operator:

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

Let us finally mention that, as already said, the GCE metric is a measure tolerant to the intrinsic variability existing between possible interpretations of an image by different human observers. As proposed in [9], this variability can simply be taken into account by estimating the mean GCE value, or, in our case, by estimating the mean GCE* value. More precisely, let us assume a set of L manually segmented images $\{S^{g}_{k}\}_{k \leq L} = \{S^{g}_{1}, S^{g}_{2}, \ldots, S^{g}_{L}\}$ related to a same scene. Let S^{t} be the segmentation to be compared with the manually labeled set, the mean GCE* measure is thus given by:

$$\overline{\mathrm{GCE}^{\star}}\big(S^{t}, \{S^{g}_{k}\}_{k \leq L}\big) = \frac{1}{L} \sum_{k=1}^{L} \mathrm{GCE}^{\star}\,(S^{t}, S^{g}_{k}) \tag{4}$$

Fig. 1. From left to right; Input image, chosen from the Berkeley image dataset, Results of K-means clustering for the segmentation model (presented in Section 3.1) and final segmentation given by our fusion model.

2.2. Penalized Likelihood Based Fusion Model

Let us assume now that we have an ensemble of L (different) segmentations $\{S_k\}_{k \leq L} = \{S_1, S_2, \ldots, S_L\}$ of the same scene to be combined with the goal of providing a final improved segmentation result $\hat{S}$. A classic strategy for finding a segmentation result $\hat{S}$, which would be a consensus or compromise of $\{S_k\}_{k \leq L}$ consists in designing an energy-based model generating a segmentation solution which is as close as possible (with the GCE* considered distance) to all the other segmentations, since this measure, contrary to the $\overline{\mathrm{GCE}}$ measure, is not degenerate. This optimization-based approach is sometimes referred to as the *median partition* [10] approach with respect to both the segmentation ensemble $\{S_k\}_{k \leq L}$ and the ($\overline{\mathrm{GCE}}^{\star}$ in our case) criterion. In this framework, if $\mathcal{S}_n$ designates the set of all possible segmentations using n pixels, the consensus segmentation $\hat{S}_{\overline{\mathrm{GCE}}^{\star}}$ (optimal in the GCE* sense) is then straightforwardly defined as the minimizer of the $\overline{\mathrm{GCE}}^{\star}$ function:

$$\hat{S}_{\overline{\mathrm{GCE}}^{\star}} = \arg \min_{S \in \mathcal{S}_n} \overline{\mathrm{GCE}}^{\star}\big(S, \{S_k\}_{k \leq L}\big) \tag{5}$$

However, the problem of image segmentation remains an ill-posed problem, providing different solutions regarding to the multiple possible values of the number of regions, which is *a priori* unknown.[1] To make this problem a well-posed problem, characterized by a unique solution, it is essential to add some constraints on the segmentation process, favoring merging regions or conversely, splitting existing regions. Analytically, this requires to recast our likelihood estimation

[1]Different human experts can segment the same image at different levels of detail.

problem of the consensus segmentation in the penalized likelihood framework by adding, to the simple (likelihood) fusion model [see (5)], a regularization term, allowing to integrate knowledge about the types of resulting fused segmentations, *a priori* considered as acceptable solutions. In our case, we search to estimate a resulting segmentation map providing a reasonable number of segments or regions. In this optic, an interesting global prior, derived from the information theory, is the following region-based regularization term:

$$E_{\text{Reg}}\Big(S = \{C_k\}_{k\leq R}\Big) = \left| - \sum_{k=1}^{R} \left[\frac{|C_k|}{n} \log \frac{|C_k|}{n} \right] - \overline{\mathcal{R}} \right| \quad (6)$$

where we remind that R denotes the number of regions in the segmentation map S, n and $|C_k|$ are respectively the number of pixels within the image and the number of pixels in the k-th region C_k of the segmentation map S. $\overline{\mathcal{R}}$ is an internal parameter of our regularization term that defines the mean entropy of the *a priori* defined acceptable segmentation solutions. This penalty term favors merging if the current segmentation solution has an entropy greater than $\overline{\mathcal{R}}$ (i.e., in the case of an over-segmentation) and favors splitting in the contrary case. Finally, with this regularization term, a penalized likelihood solution of our fusion model is thus given by:

$$\begin{aligned} \hat{S}_{\overline{\text{GCE}}^{\star}_{\beta}} &= \arg \min_{S\in\mathcal{S}_n} \Big\{ \overline{\text{GCE}}^{\star}\big(S, \{S_k\}_{k\leq L}\big) + \beta\, E_{\text{Reg}}(S) \Big\} \\ &= \arg \min_{S\in\mathcal{S}_n} \overline{\text{GCE}}^{\star}_{\beta}\big(S, \{S_k\}_{k\leq L}\big) \end{aligned} \quad (7)$$

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

2.3. Optimization of the Fusion Model

In our case, this optimization problem is difficult to solve, mainly because (among other things) we are not able to express (for this $\overline{\text{GCE}}^{\star}$ criterion) the local decrease in the energy function for a new label assignment at pixel p_i. Nevertheless, we can adopt the general optimization strategy proposed in [11], in which the optimization is based on the ensemble of superpixels belonging in $\{S_k\}_{k\leq L}$, i.e., the ensemble of segments or regions provided by each individual segmentations to be fused. This approach has also other crucial advantages. First, by considering this set of superpixels, as being the atomic elements that make up the consensus segmentation (instead of the set of pixels), we considerably decrease the computational complexity of the consensus segmentation process. Second, it is also quite reasonable to think that, if individually, each segmentation (to be fused) might give some poor results of segmentation for some sub-parts of the image (i.e., bad regions or superpixels) and also conversely good segmented regions (or superpixels) for other sub-parts of the image, the superpixel ensemble created from $\{S_k\}_{k\leq L}$ is likely to contain the different individual pieces of regions or right segments belonging to the optimal consensus segmentation solution. In this semi-local optimization strategy, the relaxation scheme is based on a variant of the iterative conditional modes (ICM) [12] i.e., a Gauss-Seidel type process which iteratively optimizes only one superpixel (in our strategy) at a time without considering the effect on other superpixels (until convergence is achieved).

3. EXPERIMENTAL RESULTS

3.1. Initial Tests Setup

In all the tests, the evaluation of our fusion scheme [see (7)] is presented for an ensemble of $L = 60$ segmentations $\{S_k\}_{k\leq L}$ with spatial partitions generated with the simple K-means based segmentation technique associated with 12 different color spaces in order to ensure variability in the segmentation ensemble. Those are; RGB, LAB, HSV YCbCr, TSL, YIQ, XYZ, h123, P1P2, HSL, i123, LUV, and using as features the set of values of the requantized color histogram around the pixel to be classified.[2]

As initial test, we have evaluated the influence of parameter $\overline{\mathcal{R}}$ [see (6)] on the generated solutions of segmentation. Fig. 2 indicates unambiguously that $\overline{\mathcal{R}}$ can be clearly interpreted as a regularization parameter of the final number of regions of our combination scheme; favoring under-segmentation, for low values of $\overline{\mathcal{R}}$ (and consequently penalizing small regions) or splitting, for great values of $\overline{\mathcal{R}}$. To further test the regularization role of $\overline{\mathcal{R}}$ in our fusion model, we have also calculated the average regions number in each image of the BSD300 as a function of the value of $\overline{\mathcal{R}}$. In our case, the value for $\overline{\mathcal{R}} = 4.2$ (see Section 3.2) allows to obtain 23 regions, on average, on the BSD300. This parameter is interesting since the average regions number belonging to the (training) set of human segmentation ensemble of the BSD300 is around this value (see [9]).

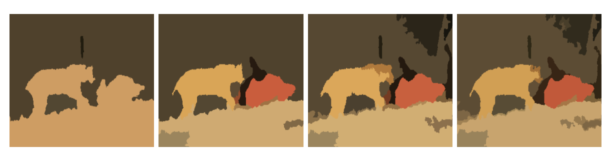

Fig. 2. Example of segmentation solutions generated for different values of $\overline{\mathcal{R}}$ ($\beta = 0.01$), from left to right, $\overline{\mathcal{R}} = \{1.2, 2.2, 3.2, 4.2\}$, respectively segmentation map results with $4, 12, 20, 22$ regions.

[2] See [4] for a justification of these color spaces and for the histogram-based features.

Table 1. Performance of several region-based segmentation algorithms (with or without a fusion model strategy) for four different performance (or distance) 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
GCEBFM	**2.10**	**0.19**	**8.73**	**0.80**
FMBFM [11]	1.88	0.20	9.30	0.80
PRIF [4]	1.97	0.21	8.45	0.80
FCR [2]	2.30	0.21	8.99	0.79
CTM [9]	2.02	0.19	9.90	0.76
FH [13] (in [9])	2.66	0.19	9.95	0.78
SCKM [14]	2.11	0.23	10.09	0.80
Mean-Shift (in [9])	2.48	0.26	9.70	0.75
NCuts (in [9])	2.93	0.22	9.60	0.72
gPb-owt-ucm [15]	1.65	-	-	0.81

3.2. Performances and Comparison

In order to evaluate the performance of our fusion model (as segmentation algorithm) we have evaluated it on the Berkeley Segmentation Dataset (BSD300) [8] (with images normalized to have the longest side equal to 320 pixels). To guarantee the integrity of the benchmark results, the two control parameters of our segmentation algorithm [i.e., $\overline{\mathcal{R}}$ and β, see (6) (7)] are optimized on the training images of the BSDS300 by using a local search procedure on a discrete grid, on the (hyper)parameter space and within a feasible range of parameter values (i.e., $\beta \in [10^{-3} : 10^{-1}]$ [step-size = 10^{-3}] and $\overline{\mathcal{R}} \in [3 : 6]$ [step-size =0.2]. We have found that $\overline{\mathcal{R}} = 4.2$ and $\beta = 10^{-2}$ are reliable hyper-parameters for our model yielding a competitive PRI [3] score (PRI=0.80) among the state of the art segmentation approaches existing in the literature. This PRI score (which is highly correlated with the human visual system [9]) estimates, in fact, the percentage of pixel pairs that are correctly segmented and a score equal to PRI=0.80 more precisely means that, on average, 80 % of label pairs are correctly segmented in the BSD300.

Also, in order to ensure an effective comparison with other segmentation methods we have used the VoI measure [16], the GCE [8] and the BDE [17] (see Table 1) (the lower distance is better). The results show that our method provides a competitive result for some other metrics based on different criteria and comparatively to state-of-the arts. Moreover, we can observe (see Fig. 3) that the PRI, VoI, BDE, GCE performance scores are better when L (the segmentation number to be merged) is high. This test shows the validity and the potentiality of our fusion procedure and demonstrates also that our performance scores are perfectible if the segmentation ensemble is completed by other (and complementary) segmentation maps of the same image. Also, we present in Fig. 4 four segmentations obtained by our algorithm. Finally, in term of execution time, the overall segmentation scheme takes, on average, 2 minutes for an Athlon-AMD 64-Proc-3500+, 2.2 GHz, 4422.40 bogomips and non-optimized code running on Linux.

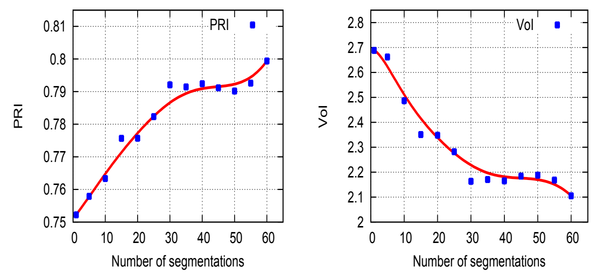

Fig. 3. Progression of the PRI and the VoI metrics according to the segmentations number (L) to be fused for our GCEBFM algorithm. Precisely, for $L = 1, 5, ..., 60$ segmentations (by considering first, one K-means segmentation (according to the RGB color space) and then by considering five segmentation for each color space and $1, 2, \ldots, 12$ color spaces).

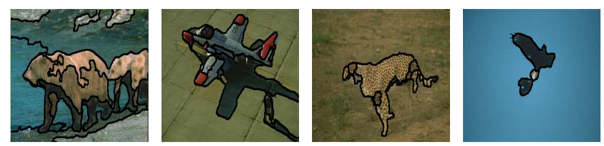

Fig. 4. Example of segmentations obtained by our algorithm GCEBFM on several images of the Berkeley image dataset.

4. CONCLUSION

In this work, we have introduced an efficient fusion model whose objective is to fuse multiple segmentation maps to provide a final improved segmentation result, in the (perceptual) global consistency error sense. In order to overcome the inherent ill-posed nature of the segmentation problem, we have re-cast the estimation problem of the consensus segmentation in the penalized likelihood framework by adding to this fusion model, a merging entropy-based regularization term allowing to incorporate knowledge about the types of resulting fused segmentations, *a priori* considered as acceptable solutions. This penalized likelihood estimation procedure remains simple to implement, efficient in terms of various criteria, perfectible (by incrementing the number of segmentations to be fused), robust to outliers and is also general enough to be applied to different other problems dealing with label fields.

5. REFERENCES

[1] Y. Jiang and Z.-H. Zhou, "SOM ensemble-based image segmentation," *Neural Processing Letters*, vol. 20, no. 3, pp. 171–178, 2004.

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

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

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] C. Hélou and M. Mignotte, "A precision-recall criterion based consensus model for fusing multiple segmentations," *IJSIP International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 7, no. 3, pp. 61–82, 2014.

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

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

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

[15] 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.

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

[17] 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.