Bayesian Region Growing and MRF-based Minimization for Texture and Colour Segmentation

Ilias Grinias, Nikolaos Komodakis and Georgios Tziritas
Multimedia Informatics Laboratory of Computer Science Department, University of Crete
Heraklion, P.O. Box 2208, 71409, Greece
phone: + (30) 2810 393517, fax: + (30) 2810 393501
e-mail: {grinias, komod, tziritas}@csd.uoc.gr

***Abstract*— We propose a generic, unsupervised feature classification and image segmentation framework, where only the number of classes is assumed as known. Image segmentation is treated as an optimization problem. The framework involves block-based unsupervised clustering using *k-means*, followed by region growing in spatial domain. High confidence statistical criteria are used to compute a map of initial labelled pixels. A new region growing algorithm is introduced, which is named *Independent Flooding Algorithm* and computes a height per label for each one of the unlabeled pixels, using Bayesian dissimilarity criteria. Finally, a *MRF* model is used to incorporate the local pixel interactions of label heights and a graph cuts algorithm performs the final labelling by minimizing the underlying energy. Segmentation results using texture, intensity and color features are presented.**

I. Introduction

Image segmentation is a key step in many computer vision analysis and interpretation tasks. Segmentation of color textured images has eventually become a necessity for many multimedia applications, such as content based image retrieval (CBIR) and object recognition purposes, especially after the development of international standard MPEG-7 [1].

Despite the plethora of methodologies for image segmentation we note the lack of a single, generic paradigm that addresses the whole range of segmentation problems and applications. This is due to the frequent complexity and ill-posedness of segmentation problems and the absence of an unambiguous ground truth. In light of these considerations, interactive segmentation techniques are also frequently employed [2].

Considering image segmentation as an optimization problem we should introduce four kinds of constraints: boundary, shape, region and topology. In edge detection [3] only boundary constraints are taken into account, while such constraints can be integrated in region growing techniques [4]. The last category of techniques incorporates soft topology constraints. Boundary constraints combined with geometric shape constraints lead to geodesic active contours [5], where a global optimization method is applied. This approach is generalized in [6] giving a powerful method because it introduces and deals with boundary, shape and region constraints. The natural counter part of topology flexibility is the difficult incorporation of topological constraints. However, topology constraints could guide the segmentation process and this is the case in our work. Nevertheless, we focus on automatically determining topological constraints based on accurate region features.

We now consider the different cues for image segmentation. Multi-channel filtering approaches for texture analysis have been proposed, using filter-banks of Gabor filters [7], [8] or wavelet packet frame decomposition [9]. In [10], many different multi-channel filtering approaches have been compared. Among the best filters were the Discrete Wavelet Frames (DWF) filter bank [11], which is used for texture modelling in our work. Wavelet frames representation decomposes the image into orthogonal texture components in different scales and orientations and it is translation invariant, a necessary property, when quite precise boundary localization is required. Methods that combine texture and color information for segmentation have also been proposed in the literature [12].

In the proposed *unsupervised* pattern classification and segmentation framework, only the number L of classes is assumed to be known. The proposed framework may be roughly separated in two main components, namely, *feature extraction* and *classification* in the feature space, which is constructed by image data information and *energy minimization* in spatial domain based on the computed features of classes. Referring to Fig. 1 and given as input a) the number of classes, b) the selection of segmentation features and c) the input image, the derived pixel features of chromaticity, texture and intensity are computed, if they have been selected as segmentation features. The next step consists of the block-based classification of features. The feature description of (possibly overlapping) blocks is derived, followed by the optional rejection of heterogeneous blocks, as it is described in detail in [13], to exclude from clustering the blocks which are not similar to their neighboring ones. Homogeneous blocks are given as input to a *k-means* initialization algorithm proposed by Kauffman and Rousseeuw [14], which successively selects a prototype block for each one of the clusters. Clustering of homogeneous blocks is then performed by *k-means* in order to extract a feature vector per class. Among several known distance measures between probability distributions, the Bhattacharyya distance is used herein to measure the distance between block instances as well as between a block and a class in *k-means*. Then, probabilistic distances are used to determine and label pixels that belong to one of the classes with high confidence. Having available the data modelling

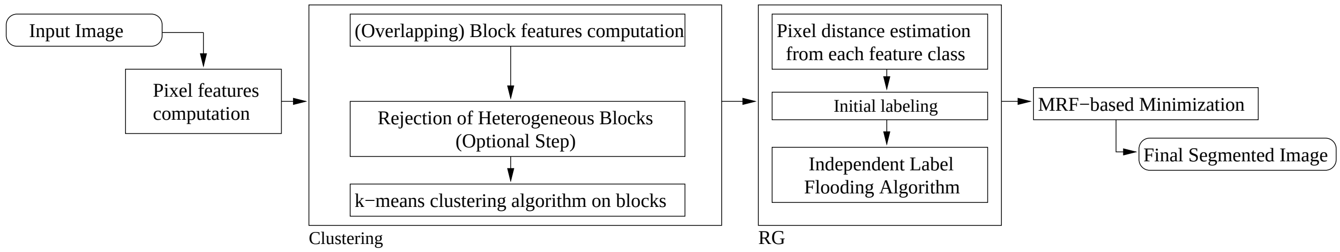


Fig. 1. Flowchart of the proposed segmentation framework.

and the initial map of correctly labelled pixels, we propose a new *Region Growing* (RG) algorithm in order to compute a topographic surface per label by assigning a height to each unlabeled pixel, using Bayesian dissimilarity criteria. Finally, the local pixel interactions of label heights are modeled by a *Markov random field (MRF)* model and the underlying energy is minimized by a novel *primal-dual* algorithm.

II. Flooding Process for Label Propagation

A. A min-max criterion for labelling

In what follows, it is assumed that a method such as that of Subsection II-B, which assigns pixel regions to classes with high confidence has been performed. Let $S = \bigcup_{l=0}^{L-1} S_l$ be the set of those initially labelled pixels. For any unlabeled pixel s we can consider all the paths linking it to a labelled set or region. A path $\mathcal{C}_l(s)$ is a sequence of adjacent pixels $\{s_0, \ldots, s_n\}$, where $s_n = s$, while all pixels of the sequence are unlabeled, except s_0 which has label l. The cost of a particular path is defined as being equal to the maximum cost of a pixel classification according to the Bayesian rule and along the path

$$\max_{i=1,\ldots,n} d_l^B(s_i),$$

with

$$d_l^B(s) = -\ln \Pr\{l|\boldsymbol{\xi}(s)\} = -\ln \frac{P_l p_l(\boldsymbol{\xi}(s))}{\sum_{k=0}^{L-1} P_k p_k(\boldsymbol{\xi}(s))},$$

where P_k is the *a priori* probability of class k.

Therefore, for each l a topographic surface on a discrete grid is defined, considering 4-connected pixels. The initially labelled pixels are defined to be at the zero level, while the height of the unlabeled pixels is given by the Bayesian rule. Indeed, $d_l^B(s)$ are always non-negative.

Finally the labelling problem becomes equivalent to search for the shortest path under the above cost, as we can define the distance of any unlabeled pixel from the different classes as being the lowest height to climb for reaching site s,

$$\delta_l(s) = \min_{\mathcal{C}_l(s)} \max_{s_i \in \mathcal{C}_l(s)} d_l^B(s_i). \tag{1}$$

Therefore the decisions are topology constrained.

If we consider the graph of unlabeled sites with 4-connections and the labelled connected components, we can define an edge weight as follows

$$w(s_{i-1}, s_i) = \max(d_l^B(s_{i-1}), d_l^B(s_i)).$$

The weight thus defined is an ultra-metric distance measure. Paths considered previously belong to the *minimum spanning tree* of this graph. Therefore the computation of $\delta_l(s)$ necessitates the construction of the *minimum spanning tree*. Prim's algorithm could be used adequately.

On the other hand, it is very interesting to remark that the labelling problem, as posed here, consists of constructing a topographic surface, as that for finding watershed lines [15]. Hence, we can use a region growing procedure, like the immersion (flooding) algorithm [16] for computing the above defined heights and distances and for classifying pixels, taking into account region features and topology constraints. We present now a new label initialization method which is followed by a novel flooding algorithm in order to determine the optimal label for each initially unlabeled site.

B. Label Initialization

The output of *label initialization* is a set of spatially connected regions of pixels, which are classified to class l with high confidence, using statistical tests. For each pixel s and class l , the distances in a window Π_W of dimension $(2W+1)^2$ are averaged, resulting to the metric:

$$d_l^{SB}(s) = \sum_{z \in \Pi_W} d_l^B(s+z).$$

Then, image pixels are sorted in ascending order according to that metric and a user-given percentage of the sites with minimum average distance are retained and get labeled.

This method may also be considered as an algorithm to determine initial regions of high confidence for the construction of minimum spanning tree, for each label of the image. Indeed, metric $d_l^{SB}(s)$ could be interpreted as the weight of the spanning sub-tree which is constructed using all the pixels of window Π_W. The pixels s of minimum $d_l^{SB}(s)$ are placed on topographic valleys of minimum height, thus constituting the better initialization option for label flooding.

C. Flooding or Region Growing Algorithm

The **Independent Label Flooding Algorithm (ILFA)** determines the optimum label as that minimizing $\delta_l(s)$. L independent flooding procedures are needed and the best label is selected. The algorithm follows the principle of *Region Growing* [4], [17]. In RG, the initially labelled pixels, composing spatially connected regions, are grown by iteratively considering neighboring pixels. For ILFA the main objective of the growing procedure is to compute the distances, and not to directly label pixels. Among all neighboring pixels to the set S_l, that are unlabeled and of unknown distance from label l, the nearest pixel is found, according to Equation (1). Growing proceeds until no more pixels can be added to the expanding regions, because their propagating contour reaches only pixels with different initial labels.

III. MRF-Based Minimization

Given the region growing measurements derived in Equation (1), we then propose to optimize a discrete *MRF* in order to decide what the final labels should be. In this manner, we aim at capturing the local interactions between pixels, which will help us to refine and correct the labels that were assigned during the previous stage of our algorithm. In general, the problem of optimizing a 1^{st}-order discrete MRF can be formulated as follows: we are given a weighted graph $\mathcal{G}$ (with nodes $\mathcal{V}$, edges $\mathcal{E}$ and weights w_{sz}), and we seek to assign a label l_s (from a discrete set of labels $\mathcal{L}$) to each node $s \in \mathcal{V}$, so that the following cost is minimized:

$$\sum\nolimits_{s\in\mathcal{V}} \mathfrak{c}(l_s) + \sum\nolimits_{(s,z)\in\mathcal{E}} w_{sz} d^P(l_s, l_z). \tag{2}$$

Here, $\mathfrak{c}(\cdot)$, $d^P(\cdot,\cdot)$ determine the singleton and pairwise MRF potential functions respectively.

In our case, the singleton potentials will be set according to the region growing measurements derived in Equation (1), i.e. $\mathfrak{c}(l_s) = \delta_{l_s}(s)$, while the pairwise potentials will be set according to the Potts function, i.e.:

$$d^P(l_s, l_z) = \begin{cases} 1, & l_s \neq l_z \\ 0, & l_s = l_z \end{cases} \tag{3}$$

Furthermore, all weights w_{sz} will be set equal to a user-specified constant.

For minimizing the MRF energy in (2), we will make use of the recently proposed *primal-dual* method in [18], which casts the MRF optimization problem as an integer program and then makes use of the duality theory of linear programming in order to derive solutions that are provably almost optimal. Furthermore, that algorithm proves to be faster than α-expansion and also applies to a much wider class of MRFs.

IV. Experimental Results

We give experimental results to evaluate the performance of our method. Texture analysis is based on a Discrete Wavelet Frames (DWF) filter bank [11], resulting in a set of $n = 3K+1$ components for each pixel s, $\boldsymbol{\zeta}(s) = \{\zeta_i(s), 1 \leq i \leq n\}$, where K is the number of analysis scales. The first $3K$ descriptors are the details in orientation for each scale, while the n-th component is the approximation. Alternatively, the intensity $\mathcal{I}(s)$ of each pixel may be used instead of the approximation as the n-th component or this component may not be used at all. The chromaticity coordinates $\mathbf{c}(s) = (a(s), b(s))$ of the *Lab* color space are used for color representation, when color information is taken under consideration. Texture details of each class l, $0 \leq l < L$, are assumed to be zero-mean, Gaussian distributed and uncorrelated and they are represented by the variance of the texture components of pixels belonging to label l, while approximation/intensity and color of the class are represented by 1D and 2D histograms of intensity and chromaticity of classes pixels, respectively.

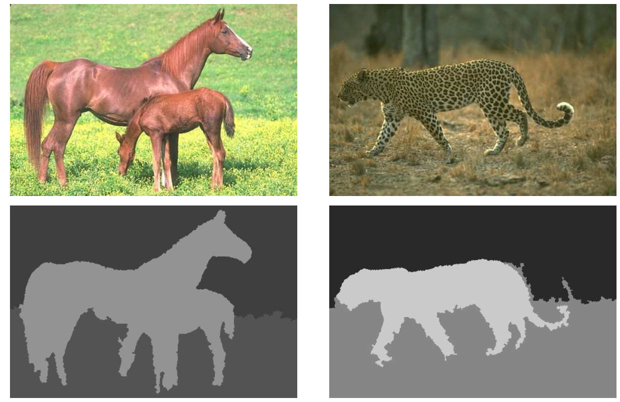


Fig. 4. Segmentation results for two images of the *Berkley Segmentation Dataset*.

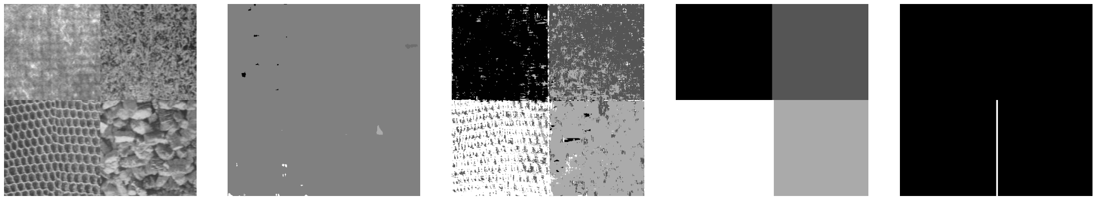
An illustrative example of the method is given using the original *four regions* image of Fig. 2(a). Histogram of intensity values has been used together with high frequency texture analysis in this experiment. After the extraction and block-based clustering of features by *k-means*, the initial labelled regions of high confidence are depicted in the map of Fig. 2(b). In Fig. 2(c) we see what would be the result if ILFA was used as labelling algorithm, while in Fig. 2(d) the result of the minimization of the metric computed by ILFA (Eq. (1)), is depicted. Finally, in Fig. 2(e) we see the 0.45% erroneously labelled pixels of the optimization process.

In [19] are reported results of texture classification on nine synthetic texture mosaics using various techniques mainly based on frame representations. Our results are summarized in Table I and compared to those of [13] only for the *five regions* images of [19], due to limited space. The original images and the corresponding segmentation results are given in Figure 3.

We also present results for images of the *Berkley Segmentation Dataset*. Although these images are mainly used as benchmark for gradient-based segmentation algorithms, the combination of high frequency texture components, intensity and *Lab* chromaticity as segmentation features under our general Bayesian framework, leads to very good segmentation results, as it is shown in Fig. 4.

V. Conclusion

A generic, unsupervised feature classification and image segmentation framework, has been proposed. Image segmentation has been treated as an optimization problem. The framework involves block-based unsupervised clustering, followed by a new region growing algorithm which is named *Independent Flooding Algorithm* and computes a topographic surface per label, using Bayesian dissimilarity criteria. Finally,



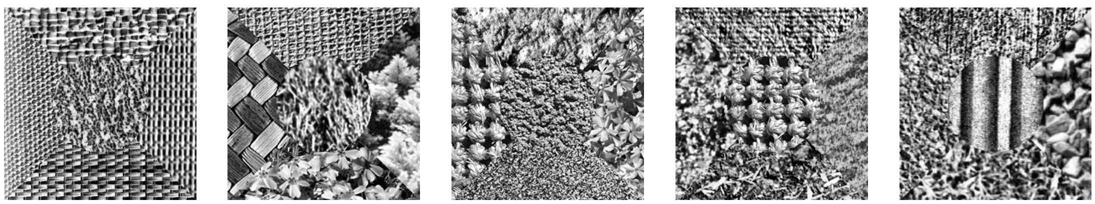
(a) (b) (c) (d) (e)

Fig. 2. Texture segmentation for the *four regions* image.

Image	(a)	(b)	(c)	(d)	(e)	Mean
Error of method	2.3%	3.72%	3.74%	2.50%	3.94%	3.24%
Error (SMLFM)	5.8%	5.4%	8.8%	8.3%	4.9%	6.64%

TABLE I

ERROR PERCENTAGE RESULTS ON IMAGES FROM [10] (P. 300)



(a) (b) (c) (d) (e)

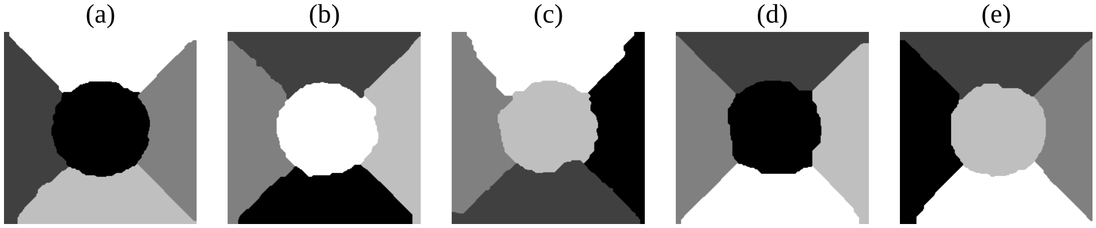


Fig. 3. Texture segmentation for 5 natural textures.

a *MRF* model is used to incorporate the local pixel interactions of label surface heights and a graph cuts algorithm gives the final labelling by minimizing the underlying energy.

REFERENCES

[1] P. Salembier and F. Marques, "Region-based representations of image and video: segmentation tools for multimedia services," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 1147–1169, Dec. 1999.

[2] Y. Boykov and G.Funka-Lea, "Graph cuts and efficient n-d image segmentation," *Intern. Journal of Computer Vision*, vol. 70, pp. 109–131, 2006.

[3] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 679–698, 1986.

[4] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 641–647, June 1994.

[5] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Intern. Journal of Computer Vision*, vol. 22, pp. 61–79, 1997.

[6] N. Paragios, *Geodesic Active Regions and Level Set Methods: Contributions and Applications in Artificial Vision*. Ecole Superiore en Science Informatique, Universite de Nice-Sophia Antipolis/I.N.R.I.A., 2000.

[7] A. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition*, vol. 24, pp. 1167–1186, Dec. 1991.

[8] P. P. Rangu and B. Yegnanarayana, "Segmentation of Gabor filtered textures using deterministic relaxation," *IEEE Trans. Image Processing*, vol. 5, pp. 1625–1636, Dec. 1996.

[9] A. Laine and J. Fan, "Frame representation for texture segmentation," *IEEE Trans. Image Processing*, vol. 5, pp. 771–780, May 1996.

[10] T. Randen and J. H. Husoy, "Filtering for texture classification: A comparative study," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 571–582, Apr. 1999.

[11] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Trans. Image Processing*, vol. 4, pp. 1549–1560, Nov. 1995.

[12] S. Belongie, C. Carson, H. Greenspan, and J. Malik, "Color- and texture-based image segmentation using EM and its application to content based image retrieval," in *Proc. IEEE International Conference of Computer Vision (ICCV98)*, 1998.

[13] S. Liapis, E. Sifakis, and G. Tziritas, "Colour and texture segmentation using wavelet frame analysis, deterministic relaxation and fast marching algorithms," *Journal of Visual Communication and Image Representation*, vol. 15, no. 1, pp. 1–26, Mar. 2004.

[14] L. Kauffman and P. Rousseeuw, *Finding Groups in Data. An Introduction to Cluster Analysis*. John Willey and Sons, 1990.

[15] F. Meyer, "Topographic distance and watershed lines," *Signal Processing*, vol. 38, pp. 113–125, 1994.

[16] L. Vincent and P.Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 583–598, 1991.

[17] S. A. Hojjattolesami and J. Kittler, "Region growing - a new approach," *IEEE Trans. Image Processing*, vol. 7, no. 7, pp. 1079–1084, July 1998.

[18] N. Komodakis and G. Tziritas, "A new framework for approximate labeling via graph-cuts," in *ICCV 2005*.

[19] K. Skretting and J. H. Husoy, "Texture classification using sparse frame-based representations," *EURASIP Journal on Applied Signal Processing*, 2006.