A NOVEL FRAMEWORK FOR N-D MULTIMODAL IMAGE SEGMENTATION USING GRAPH CUTS

Asem M. Ali *Aly A. Farag*

Computer Vision and Image Processing Laboratory (CVIP Lab)
University of Louisville, Louisville, KY 40292
{asem,farag}@cvip.uofl.edu, URL: www.cvip.uofl.edu*

ABSTRACT

This work proposes a new MAP-based segmentation framework of multimodal images. In this work a joint MGRF model is used to describe the image. The main focus here is a more accurate model identification. For a known number of classes in the given image, the empirical distributions of this image signals are precisely approximated by a LCG distributions with positive and negative components. Gibbs potential, which is used to identify the spatial interaction between the neighboring pixels, is analytically estimated. Finally, an energy function using the previous models is formulated and is globally minimized using graph cuts. Experiments show that the developed technique gives promising accurate results compared to other known algorithms.

Index Terms— MRF, Graph Cut, LCG

1. INTRODUCTION

Segmentation is a fundamental problem in image processing. This paper addresses the problem of accurately unsupervised segmenting multimodal grayscale images. There are many simple techniques, such as region growing or thresholding, for multimodal image segmentation. Although these techniques are widely known due to their simplicity and speed, but no accurate segmentation can be achieved using these techniques. This is because these techniques depend only on the marginal probability distributions, and in most cases signal ranges for different object overlap. Energy-based segmentation approaches are more robust algorithms. Our proposed framework uses graph cuts as a powerful optimization technique to get the optimal segmentation. Shi and Malik [1] proposed the normalized cut criteria, a measure of both the total dissimilarity between the different image regions as well as the total similarity within the image regions, for graph partitioning. To compute the minimum cut which corresponds to optimum segmentation, they used an eigenvalue system. Boykov and Jolly [2] proposed a framework that used s/t graph cuts to get a globally optimal object extraction method for N-dimensional images. They minimized a cost function that combines region and boundary properties of segments as well as topological constraints. That work illustrated the effectiveness of formulating the object segmentation problem via graph cuts. Since Boykov and Jolly introduced their graph cuts segmentation technique in their seminal paper [2], it became one of the leading approaches in interactive N-D image segmentations, and many publications extended this work in different directions. For more details see [3] and references therein. These works showed the power of graph cuts as a tool for image segmentation, since it optimizes energy functions that can integrate regions, boundary, and shape information. Also, the graph cuts technique offers a reliable and robust globally optimal object segmentation method. Most of these works are interactive segmentation. Although interactive segmentation imposes some topological constraints reflecting certain high-level contextual information about the object, it depends on the user input. The user inputs have to be accurately positioned. Otherwise the segmentation results are changed.

This paper proposes a Maximum-A-Posterior (MAP)-based segmentation approach for multimodal grayscale image segmentation. To model the low level information in the multimodal image, we precisely approximate the empirical distributions of the image signals by a Linear Combination of Gaussian (LCG) distributions with positive and negative components. For accurate model identification, the image is described by a joint Markov Gibbs Random Field (MGRF) model. Then the spatial interaction potential for this MGRF model is analytically estimated. Finally, we use LCG and MGRF models to formulate an energy function, which is globally minimized using graph cuts.

2. PROPOSED FRAMEWORK

The weighted undirected graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is a set of vertices $\mathcal{V}$, and a set of edges $\mathcal{E}$ connecting the vertices. Each edge is assigned a non-negative weight. The set of vertices $\mathcal{V}$ corresponds to the set of image pixels $\mathcal{P}$, and two specially terminal vertices s (source/object), and t (sink/background). The set of edges $\mathcal{E}$ consists of two subsets: *n-links*, the edges connecting the neighboring pixels in the image, and *t-links*, the edges connecting the pixels with the terminals. An s/t cut, which can be computed efficiently in low-order polynomial time, divides the set of image pixels into two subsets, background and object. Consider a neighborhood system$\mathcal{N}$ of all unordered pairs $\{p, q\}$ of neighboring pixels in $\mathcal{P}$. Let $\mathcal{L}$ the set of labels $\{0, 1, .., K\}$, correspond to image modes. Labelling is a mapping from $\mathcal{P}$ to $\mathcal{L}$, and we denote the set of labelling by $\mathbf{f} = \{f_1, \ldots, f_p, \ldots, f_{|\mathcal{P}|}\}$. In other words, the label f_p, which is assigned to the pixel $p \in \mathcal{P}$, classifies it to one of the labels. Now the goal is to find the best labelling $\mathbf{f}$, optimal segmentation, by minimizing the following function

$$E(\mathbf{f}) = \sum_{p \in \mathcal{P}} D_p(f_p) + \sum_{\{p,q\} \in \mathcal{N}} V(f_p, f_q), \qquad (1)$$

where $D_p(f_p)$, measures how much assigning a label f_p to pixel p disagrees with the pixel intensity, I_p. A good example for $D_p(f_p)$ represents the regional properties of segments

$$D_p(f_p) = -ln\ P(I_p \mid f_p). \qquad (2)$$

We estimate the empirical distribution $P(.)$ of each class as shown in Sec. 2.1. The second term is the pairwise interaction model which

*This research has been supported by US National Science Foundation Grant IIS-0513974.

represents the penalty for the discontinuity between pixels p and q. The potential function, $V(.,.)$, is defined as follows:

$$V(f_p, f_q) = \begin{cases} \gamma & \text{if } f_p \neq f_q; \\ 0 & \text{if } f_p = f_q \end{cases} . \quad (3)$$

In this work, we uses our analytical approach [4] to estimate the spatial interaction γ between the neighboring pixels in MGRF model, which is used to describe the image as shown in Sec.2.2. For image segmentation, we minimize the energy function Eq.(1) using the graph cut technique [5].

Our proposed framework can segment images have more than one objects. So the energy function Eq.(1) will be defined over a finite set of labels. To solve this problem we use the α-expansion move algorithm [6].

2.1. Gray Level Distribution Probabilistic Model

To accurately model the multimodal image, we approximate the gray level marginal density of each class using a LCG with $C_{\mathrm{p},k}$ positive and $C_{\mathrm{n},k}$ negative components as follows:

$$P(g|k) = \sum_{r=1}^{C_{\mathrm{p},k}} w_{\mathrm{p},k,r}\varphi(g|\theta_{\mathrm{p},k,r}) - \sum_{l=1}^{C_{\mathrm{n},k}} w_{\mathrm{n},k,l}\varphi(g|\theta_{\mathrm{n},k,l}). \quad (4)$$

Where, g is the gray level, k is the class number, $\varphi(g|\theta)$ is a Gaussian density with parameter $\theta \equiv (\mu, \sigma^2)$ with mean μ and variance σ^2, and $w_{\mathrm{p},k,r}$ means the r^{th} positive weight in class k. The model components are both positive and negative and have a restriction that the summation of the weights in Eq.(4) is one: $\sum_{r=1}^{C_{\mathrm{p},k}} w_{\mathrm{p},k,r} - \sum_{l=1}^{C_{\mathrm{n},k}} w_{\mathrm{n},k,l} = 1$. For a known number of dominant modes K, we estimate the parameters of the LCG model. We use the modified EM algorithm [7] to deal with the positive and negative components. Fig.1 shows a summary of the LCG model of a bimodal image.

2.2. Spatial Interaction Model

The simplest model of spatial interaction is the MGRF with the nearest 4 or 6-neighborhood, in 2D or 3D cases, respectively. Therefore, for this specific model the Gibbs potential can be obtained analytically using the maximum likelihood estimator (MLE) for the MGRF [4]. So, The potential γ is given by the following equation:

$$\gamma = \frac{K^2}{K-1}\left(\frac{K-1}{K} - f_{\mathrm{neq}}(\mathbf{f})\right), \quad (5)$$

where K is the number of classes in the volume and $f_{\mathrm{neq}}(\mathbf{f})$ denotes the relative frequency of the not equal labels in the pixel pairs and it is defined as follows.

$$f_{\mathrm{neq}}(\mathbf{f}) = \frac{1}{|\mathbf{T_N}|}\sum_{\{p,q\}\in\mathbf{T_N}} \delta(f_p \neq f_q), \quad (6)$$

where $\mathbf{T_N} = \{\{p,q\} : p,q \in \mathcal{P}; \{p,q\} \in \mathcal{N}\}$ is the family of the neighboring pixel pairs supporting the Gibbs potentials. The indicator function, $\delta(\boldsymbol{A})$ equals one when the condition $\boldsymbol{A}$ is true, and zero otherwise.

3. EXPERIMENTS AND DISCUSSION

To assess the performance of the proposed segmentation approach, we tested it on several N-D multimodal images. To validate the segmentation of proposed approach, we compare it with results of both the mean shift algorithm [8] and normalized cuts algorithm [1] (We used the authors' codes EDISON[1] and NCUTS [2], respectively). Note that when using these codes, we conducted several trials in order to select the tuning parameters that give the best results. Ground truth experiments (on synthetic 1D, 2D and 3D) are done to statistically measure the proposed algorithm performance. Fig.2 shows a 1D image (signal) with 2 classes (1,-1) distorted by a White Gaussian Noise (WGN) with standard deviation $\sigma_n = 1$. To illustrate that the standard MLE, which only uses the proposed LCG model's threshold without spatial relations, doesn't give satisfactory results, we present on Fig.2 the MLE result as well as the proposed approach result. To get some statistical analysis, we distorted the original signal with different WGNs (with different σ_n). For each σ_n we generated 50 noisy signals, and computed average relative errors of proposed approach outputs in comparison to the original signal. Means μ_ε and standard deviations σ_ε of the relative errors are shown in Table1. Results in this table illustrate the robustness to noise of the proposed algorithm. It is worth mentioning that a "similar" experiment was carried out in [9]. However, in that work, the authors used constant $\sigma_n = 1$ and performed their analysis w.r.t. spatial interaction parameter (named λ) instead of the noise as was done here. In comparison to results reported in [9], note that for the specific case of $\sigma_n = 1$, our analytical estimate for γ leads to an output with almost the same average relative error, in comparison to the original signal($\approx 3\%$), as the error reported in [9] (see Fig.7 [9]). This illustrates the accuracy of the proposed analytical approach that estimate γ.

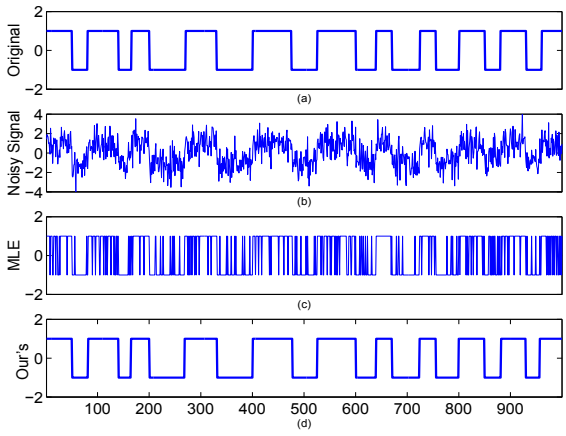
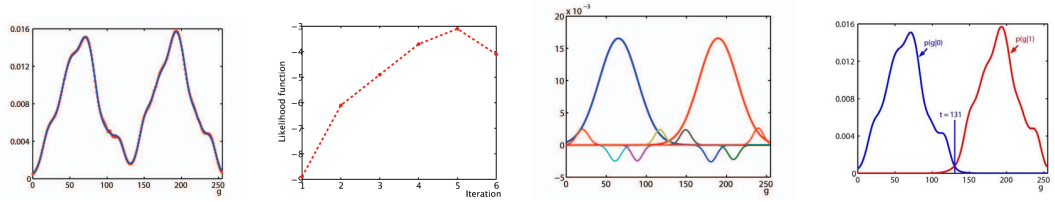


Fig. 2. 1D image labeling: (a)The original signal is distorted by WGN (b) Distorted signal (c) MLE outputs (d) Proposed algorithm output.

2D Case: To assess the robustness of the proposed approach, we tested it on 600 synthetic 2D multimodal images and compared the results with ground truths. Examples of these images and proposed approach results are shown in Fig.3. Relative errors ε of our results in comparison to ground truths and computation times τ are also given. Fig.4 shows EDISON outputs , and NCUTS outputs for input images shown in Fig.3. To get more statistics, we generated 10 3-modal data sets, each of which consists of 30 images, with different Signal to Noise Ratios (SNR). We run the proposed algorithm on the data sets and computed the average relative error for each data set. Fig. 5 shows the SNR and the corresponding average relative segmentation error. The error at SNR -5 dB is dramatically large and the proposed algorithm missed one of the objects due to the great amount of noise.

[1] Available at www.caip.rutgers.edu/riul/research/code.html
[2] Available at www.cis.upenn.edu/ jshi/software/

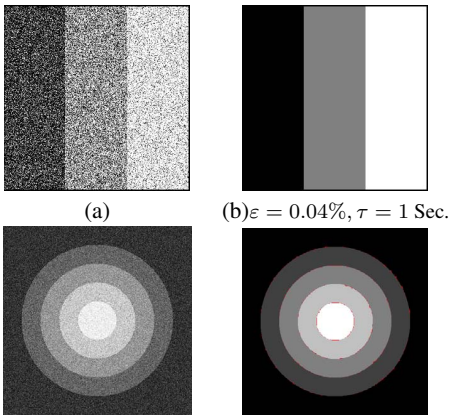


(a) (b) (c) (d)

Fig. 1. Summary of LCG model of a bimodal image. (a) LCG approximation of the bimodal image, (b) Changes of the log likelihood with the iterations (c) The componets of the LCG model, and (d) The LCG models of each class with the best threshold

Table 1. 1D Statistical results.

σ_n	0.5	0.7	1.0	1.2	1.5
μ_ε %	0.25	1.0	3.3	6.7	11.05
σ_ε %	0.2	0.65	1.9	8.0	13.1



(a) (b)$\varepsilon = 0.04\%, \tau = 1$ Sec.

(c) (d)$\varepsilon = 0.41\%, \tau = 7$ Sec.

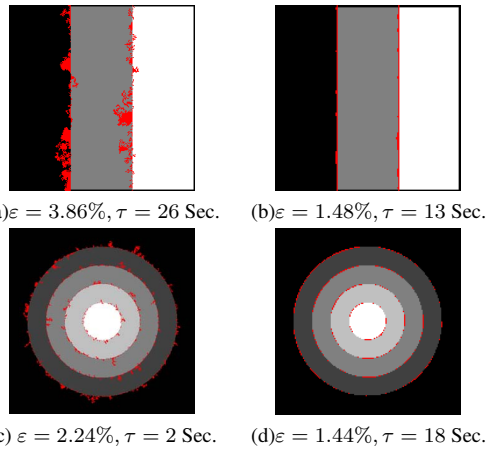
Fig. 3. Segmentation results (a) The gray level image with SNR=1 dB, (b) Proposed approach segmentation.(c) The gray level image with SNR=5 dB, (d) Proposed approach segmentation. (Error shown in red)

We repeated the same scenario for some 5-modal images (see Fig.3), and the corresponding error versus SNR is plotted on Fig. 6. For a comparison issue, the segmentation error of Iterative Conditional Modes (ICM) [10], EDISON, and NCUTS techniques are also illustrated.

3D Case: To test the algorithm on 3D case, we generated a bimodal volume phantom (256x256x50). The relative error ε=0.53% between our segmentation and the ground truth confirms the high accuracy of the proposed segmentation framework in 3D cases. To emphasize the importance of segmenting the 3D object using a 3D graph instead of independently segmenting each 2D slice of the volume (e.g, see Fig. 7 (a)), we present the two corresponding segmentation results on Fig. 7 (b) and (c), respectively.

Real images and Medical Volumes: Fig. 8 shows more segmentations of the proposed approach, EDISON, and NCUTS of a starfish image. In each output, we draw the object boundary (shown in red), and give the computation time.

Medical images are good examples of multimodal images. In such a case, errors are evaluated with respect to ground truths produced by a radiologist. Due to the closeness of the gray levels be-



(a)$\varepsilon = 3.86\%, \tau = 26$ Sec. (b)$\varepsilon = 1.48\%, \tau = 13$ Sec.

(c) $\varepsilon = 2.24\%, \tau = 2$ Sec. (d)$\varepsilon = 1.44\%, \tau = 18$ Sec.

Fig. 4. Segmentation results: for the 3-modal image(a) EDISON output, (b) NCUTS output . For the 5-modal image(c) EDISON output, (d) NCUTS output. (Error shown in red)

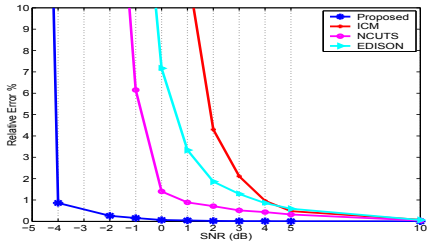


Fig. 5. The changes of the misclassification error with the SNR for 3-modal images

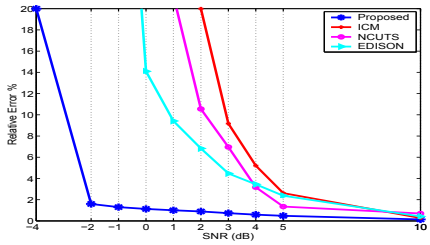
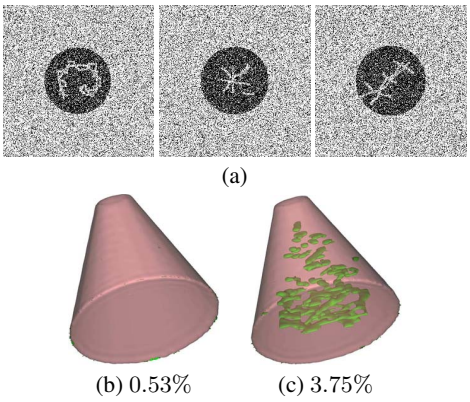


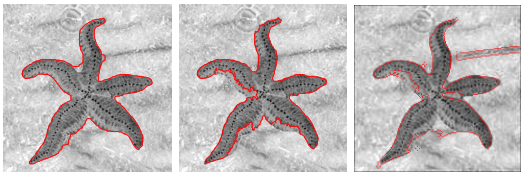
Fig. 6. The changes of the misclassification error with the SNR for 5-modal images



(a)

(b) 0.53% (c) 3.75%

Fig. 7. Synthetic volume segmentation results: (a)Slices from the synthetic volume, (b) 3D segmentation (c) 2D segmentation (Error shown in green)

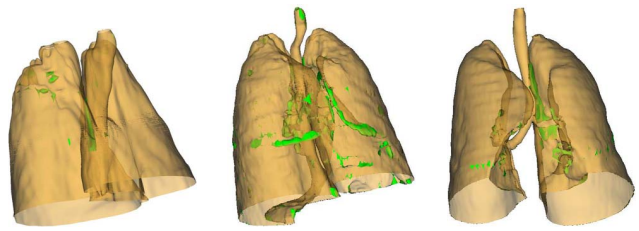


(g)$\tau = 1$ Sec. (h) $\tau = 6$ Sec. (i) $\tau = 8$ Sec.

Fig. 8. Real image segmentation results Left: Proposed Algorithm, middle: EDISON , and right: NCUT.

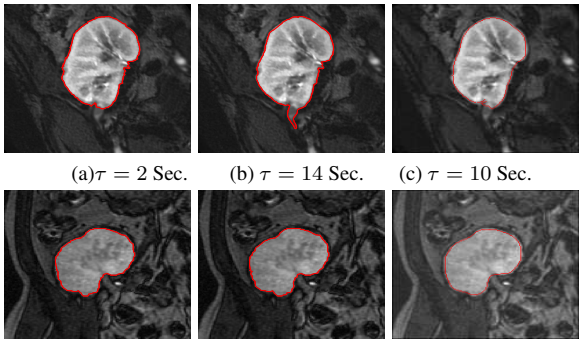
tween the abnormal tissues in the lung and the chest tissues, interactive segmentation for the Computed Tomography (CT) lung images is improper. We run the proposed approach on seven human chest CT scans, three of which are presented on Fig. 9. All misclassified pixels in our results are located at the boundary. Therefore; our segmentation did not miss any abnormal tissues, which are important if lung segmentation is a pre-step in a detection of lung nodules system. Segmentations of Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) of human kidneys are shown in Fig. 10 segmentation errors w.r.t. ground truth are [(a)$\varepsilon = 1.7\%$,(b)$\varepsilon = 1.1\%$,(c)$\varepsilon = 1.1\%$, (d)$\varepsilon = 0.1\%$, (e)$\varepsilon = 0.01\%$, (f)$\varepsilon = 0.6\%$].

In this paper, we have presented a novel approach for automatic multimodal grayscale image segmentation using the graph cuts algorithm. A joint MGRF model is used to describe the input image and its desired map with more accurate model identification. The image gray levels distribution is precisely approximated by a LCG distribu-



(a) ε= 1.95% (b) ε= 2.08% (c) ε= 2.17%

Fig. 9. Proposed approach segmentation results of different lung volumes



(a)$\tau = 2$ Sec. (b) $\tau = 14$ Sec. (c) $\tau = 10$ Sec.

(d)$\tau = 1$ Sec. (e) $\tau = 10$ Sec. (f) $\tau = 19$ Sec.

Fig. 10. Kidney segmentation results. Left: Proposed Algorithm, middle: EDISON, and right: NCUT.

tions with positive and negative components. The image is initially segmented using the thresholds of this LCG model, therefore, no user interaction is needed. From all previous segmentation results, of synthetic and real grayscale multimodal images, we can say that without optimizing any tuning parameters, the proposed approach is fast, robust to noise, and gives accurate results compared to the state-of-the-art algorithms (e.g., in 1D case [9] , and in 2D case [8] and [1]). Moreover, our approach is easily extended to segment 3D volumes.

4. REFERENCES

[1] Jianbo Shi and Jitendra Malik, "Normalized cuts and image segmentation," *IEEE TPAMI*, vol. 22, no. 8, pp. 888–905, 2000.

[2] Y. Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proc. of ICCV*, 2001, vol. 1, pp. 105–112.

[3] Yuri Boykov and Gareth Funka-Lea, "Graph cuts and efficient N-D image segmentation," *Int. J. of Comp. Vis.*, vol. 70, no. 2, pp. 109–131, 2006.

[4] Asem M. Ali, Ayman S. El-Baz, and Aly A. Farag, "A novel framework for accurate lung segmentation using graph cuts," in *Proce. of IEEE ISBI*, 2007, pp. 908–911.

[5] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?," *IEEE TPAMI*, vol. 26, no. 2, pp. 147–159, 2004.

[6] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximation energy minimization via garph cuts," *IEEE TPAMI*, vol. 23, no. 11, pp. 1222–1239, 2001.

[7] A.A. Farag, A. El-Baz, and G. Gimel'farb, "Density estimation using modified expectation maximization for a linear combination of gaussians," in *Proc. of ICIP*, 2004, vol. 3, pp. 1871 – 1874.

[8] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE TPAMI*, vol. 24, no. 5, pp. 603–619, 2002.

[9] J. Keuchel, C. Schnorr, C.and Schellewald, and D. Cremers, "Binary partitioning, perceptual grouping, and restoration with semidefinite programming," *IEEE TPAMI*, vol. 25, no. 11, pp. 1364–1379, 2003.

[10] J. E. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc. B*, vol. 48, pp. 259–302, 1986.