A Modified Fuzzy C-Means Algorithm for Document Image Segmentation

Mohamed Nooman Ahmed, Ph.D.
Applied Software Research, Lexmark International Inc., Lexington, Kentucky
Email: mahmed@lexmark.com

Abstract

In this paper, we present a new system to segment and label document images into text, images, and background using a modified Fuzzy C-means Algorithm. Each pixel is assigned a feature pattern extracted from the gray level distribution and computed at different scales. The invariant feature pattern is then assigned to a specific region using Fuzzy logic. Our algorithm is formulated by modifying the objective function of the standard fuzzy c-means (FCM) algorithm to allow the labeling of a pixel to be influenced by the labels in its immediate neighborhood. The neighborhood effect acts as a regularizer and biases the solution towards piecewise-homogeneous labelings. Such a regularization is useful in segmenting scans corrupted by scanner noise.

1 Introduction

Several approaches for document segmentation have been proposed.[1-6] These techniques can be broadly classified as bottom-up and top-down. Bottom-up methods start from pixel level and merge regions together into larger and larger components. Top-down techniques apply a priori knowledge about the page to hypothesize and split the page into blocks which are subsequently identified and further subdivided. Top-down approaches work well with pre-specified layouts such as technical papers. However. the performance of these techniques degrades significantly when different components are touching or overlapping.

Among bottom-up approaches, texture-based schemes have attracted much attention.[4-6] These methods treat different components of a document image as different textures. The scanned document images are convolved with a set of masks to generate feature vectors. Each feature vector is then classified into different classes using a pre-trained classifier.

The results of these techniques, however, are highly sensitive to noise. Another problem associated with these approaches is the mask size for extracting local features. If the mask size is too small, it is difficult to detect large scale textures such as large fonts. On the contrary, if a large mask is chosen, the computational complexity will increase dramatically.[6]

To solve the problem of noise sensitivity and to exploit both local features and image structure, we present in this paper a different approach for fuzzy segmentation of document images. Each pixel is assigned a feature pattern consisting of gray level distribution computed at different scales. The invariant feature pattern is then assigned to a specific region using Fuzzy logic.[8-12]

Our algorithm is formulated by modifying the objective function of the standard fuzzy c-means (FCM) algorithm to allow the labeling of a pixel to be influenced by the labels in its immediate neighborhood. The neighborhood effect acts as a regularizer and biases the solution towards piecewise-homogeneous labelings. Such a regularization is useful in segmenting scans corrupted by scanner noise.

2 Feature Extraction

The first step in our approach is to extract a feature pattern for each pixel in a gray level image. A feature vector X is a set of measurements $\{x_1,x_2,..,x_d\}$ which condenses the description of relevant properties of the image into a small, Euclidean feature space of *N* dimension. The number of needed features depends on the complexity of the image. The components of the feature vector may include gray values, gray values through different filters, texture measures, Markov random field features, fractal dimension measures, and gradient magnitudes and directions.[5]

In this paper, we will construct our feature vector from measurements obtained from gray level distribution in a window, *W*, of dimension *w* x *w* centered around each pixel. These features describe the first order gray level distribution without considering the spatial interdepen-dence. Two features were selected:

1. The mean gray level, μ,

$$\mu = \frac{1}{w^2} \sum_{(x,y)\in\mathcal{W}} I(x,y) \tag{1}$$

where *I*(*x*,*y*) is the gray level at location (*x*,*y*).

2. The varianct σ^2 of the gray level variation

$$\sigma^2 = \frac{1}{w^2-1} \sum_{(x,y)\in\mathcal{W}} (I(x,y)-\mu)^2. \tag{2}$$

3 Standard Fuzzy-C-Means

The standard FCM objective function for partitioning $\{x_k\}_{k=1}^N$ into c clusters is given by

$$J = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^p ||x_k - v_i||^2 \tag{3}$$

where $\{x_k\}_{k=1}^N$ are the feature vectors for each pixel, $\{v_i\}_{i=1}^c$ are the prototypes of the clusters and the array $[u_{ik}] = U$ represents a partition matrix, $U \in \mathcal{U}$, namely

$$\mathcal{U}\{\ u_{ik} \in [0,1] \ \ | \ \sum_{i=1}^{c} u_{ik} = 1 \quad \forall k \quad and \quad 0 < \sum_{k=1}^{N} u_{ik} < N \quad \forall i\} \tag{4}$$

The parameter p is a weighting exponent on each fuzzy membership and determines the amount of fuzziness of the resulting classification. The FCM objective function is minimized when high membership values are assigned to voxels whose intensities are close to the centroid of its particular class, and low membership values are assigned when the pixel data is far from the centroid.

4 Modified Fuzzy C-Means Objective Function

We propose a modification to Eq.(3) by introducing a term that allow the labeling of a pixel to be influenced by the labels in its immediate neighborhood. As mentioned before, the neighborhood effect acts as a regularizer and biases the solution towards piecewise-homogeneous labeling. Such a regularization is useful in segmenting scans corrupted by salt and pepper noise. The modified objective function is given by

$$J_m = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^p ||x_k - v_i||^2 + \frac{\alpha}{N_R}\sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^p \left(\sum_{x_r\in\mathcal{N}_k} ||x_r - v_i||^2\right) \tag{5}$$

where $\mathcal{N}_k$ stands for the set of neighbors that exist in a window around x_k and N_R is the cardinality of $\mathcal{N}_k$. The effect of the neighbors term is controlled by the parameter α. The relative importance of the regularizing term is inversely proportional to the signal to noise ratio (SNR) of the MRI signal. Lower SNR would require a higher value of the parameter α.

Formally, the optimization problem comes in the form

$$\min_{U,\ \{v_i\}_{i=1}^c} J_m \quad subject\ to \quad U \in \mathcal{U} \tag{6}$$

5 Parameter Estimation

The objective function J_m can be minimized in a fashion similar to the standard FCM algorithm. Taking the first derivatives of J_m with respect to u_{ik}, v_i, and setting them to zero results in two necessary but not sufficient conditions for J_m to be at a local extrema. In the following subsections, we will derive these three conditions.

5.1 Membership Evaluation

The constrained optimization in equation (6) will be solved using one Lagrange multiplier

$$F_m = \sum_{i=1}^{c}\sum_{k=1}^{N}\left(u_{ik}^p D_{ik} + \frac{\alpha}{N_R}u_{ik}^p\gamma_i\right) + \lambda(1-\sum_{i=1}^{c}u_{ik}) \tag{7}$$

where $D_{ik} = \|x_k - v_i\|^2$ and $\gamma_i = \left(\sum_{x_r\in N_k}\|x_r - v_i\|^2\right)$

Taking the derivative of F_m w.r.t u_{ik} and setting the result to zero, we have, for $p > 1$

$$\left[\frac{\delta F_m}{\delta u_{ik}} = p u_{ik}^{p-1} D_{ik} + \frac{\alpha p}{N_R} u_{ik}^p \gamma_i - \lambda\right]_{u_{ik}=u_{ik}^*} = 0 \tag{8}$$

Solving for u^*_{ik} we have

$$u_{ik}^* = \left(\frac{\lambda}{p(D_{ik} + \frac{\alpha}{N_R}\gamma_i)}\right)^{\frac{1}{p-1}} \tag{9}$$

Since $\sum_{j-1}^{c} u_{jk} = 1 \ \ \forall k,$ then

$$\sum_{j=1}^{c}\left(\frac{\lambda}{p(D_{jk} + \frac{\alpha}{N_R}\gamma_j)}\right)^{\frac{1}{p-1}} = 1 \tag{10}$$

or

$$\lambda = \frac{p}{\left(\sum_{j=1}^{c}\left(\frac{1}{(D_{jk}+\frac{\alpha}{N_R}\gamma_j)}\right)^{\frac{1}{p-1}}\right)^{p-1}} \tag{11}$$

Substituting into equation (9), the zero-gradient condition for the membership estimator can be rewritten as,

$$u_{ik}^{*} = \frac{1}{\sum_{j=1}^{c} \left(\frac{D_{ik} + \frac{\alpha}{N_R}\gamma_i}{D_{jk} + \frac{\alpha}{N_R}\gamma_j} \right)^{\frac{1}{p-1}}} \quad (12)$$

5.2 Cluster Prototype Updating

Using the standard Eucledian distance and taking the derivative of F_m w.r.t v_i and setting the result to zero we have;

$$\left[\begin{array}{l} \sum_{k=1}^{N} u_{ik}^{p}(x_k - v_i) + \\ \quad \sum_{k=1}^{N} u_{ik}^{p} \frac{\alpha}{N_R} \sum_{y_r \in \mathcal{N}_k} (x_r - v_i) \end{array} \right]_{v_i = v_i^*} = 0. \quad (13)$$

Solving for v_i we have

$$v_i^* = \frac{\sum_{k=1}^{N} u_{ik}^{p} \left((x_k) + \frac{\alpha}{N_R} \sum_{x_r \in \mathcal{N}_k} (x_r) \right)}{(1+\alpha) \sum_{k=1}^{N} u_{ik}^{p}} \quad (14)$$

6 Results

We tested our algorithm with several images scanned at 600 dpi. Six different features were extracted for every pixel. These features included first order statistics evaluated at three different scales. The images were segmented into text, images, and background. Results are shown in Figures 1 and 2.

References

1. Y. Y. Tang, C. Y. Suen, and M. Cheriet, "Document Analysis and Understanding: A brief Survey" *Proc. Int. Conf. on Document Analysis and Recognition,* pp.17-31, 1991.
2. H. Cheng and C. Bouman, "Trainable Context Model for Multiscale Segmentation" *Proc. Int. Conf. on Image Processing (ICIP),* October 1998.
3. L. O'Gorman, "The Document Spectrum for Page Layout Analysis" *IEEE Trans. on Pattern Matching and Machine Intelligence (PAMI),* Vol. 15, No. 11, November 1993.
4. K. Etemad, D. Doermann, and R. Chellapa, "Page segmentation using decision integration and wavelet packets," *Proc. Int. Conf. of Pattern Recognition,* Vol. 2, pp. 345-349, 1994.
5. A. K. Jain and Y. Zhong, "Page segmentation using texture analysis," *Pattern Recognition,* Vol. 29, no 5, pp. 343-770, 1996.
6. H. Cheng, C. Bouman, and J. Allebach, "Multiscale Document Segmentation," *Proc. of IS&T 50th Annual Conference,* pp. 417-425, May 1997.
7. S. Dellepiane et al., "Model generation and model matching of real images by fuzzy approach," *Pattern Recognition,* Vol. 25, No.2, pp. 115-137, 1992.
8. J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.,* Vol. 3, pp. 32-57, 1973.
9. J. Bezdek, "A convergence theorem for the fuzzy ISO-DATA clustering algorithms," *IEEE Trans. on Pattern Ana. Machine Intell. (PAMI),* 1980.
10. C. Xu, D. Pham, and J. Prince, "Finding the brain cortex using fuzzy segmentation, isosurfaces, and deformable surfaces," *Proc. of the XVth Int. Conf. on Information Processing in Medical Imaging (IPMI 97),* pp. 399-404, 1997.
11. D. L. Pham, and J. L. Prince, "Adaptive fuzzy segmentation of magnetic resonance images," *IEEE Trans. Medical Imaging,* Vol. 18, No. 9, pp. 737-752, 1999.
12. M. N. Ahmed, S. M. Yamany, N. A. Mohamed, A. A. Farag, and T. Moriarty "Bias field estimation and adaptive segmentation of MRI data using modified fuzzy C-means algorithm," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recogn.,* Fort Collins, Colorado, Vol. 1, pp. 250-255, June 1999.

Biography

Dr. Mohamed Nooman Ahmed is a research scientist at Lexmark international. He obtained his Ph.D. in Computer Science and Engineering from the University of Louisville, KY, in 1998.

His research interests include image processing, enhancement, neural networks, pattern recognition and medical imaging.

Figure 1: Original and segmented document into text (black), image (gray), and background (white)

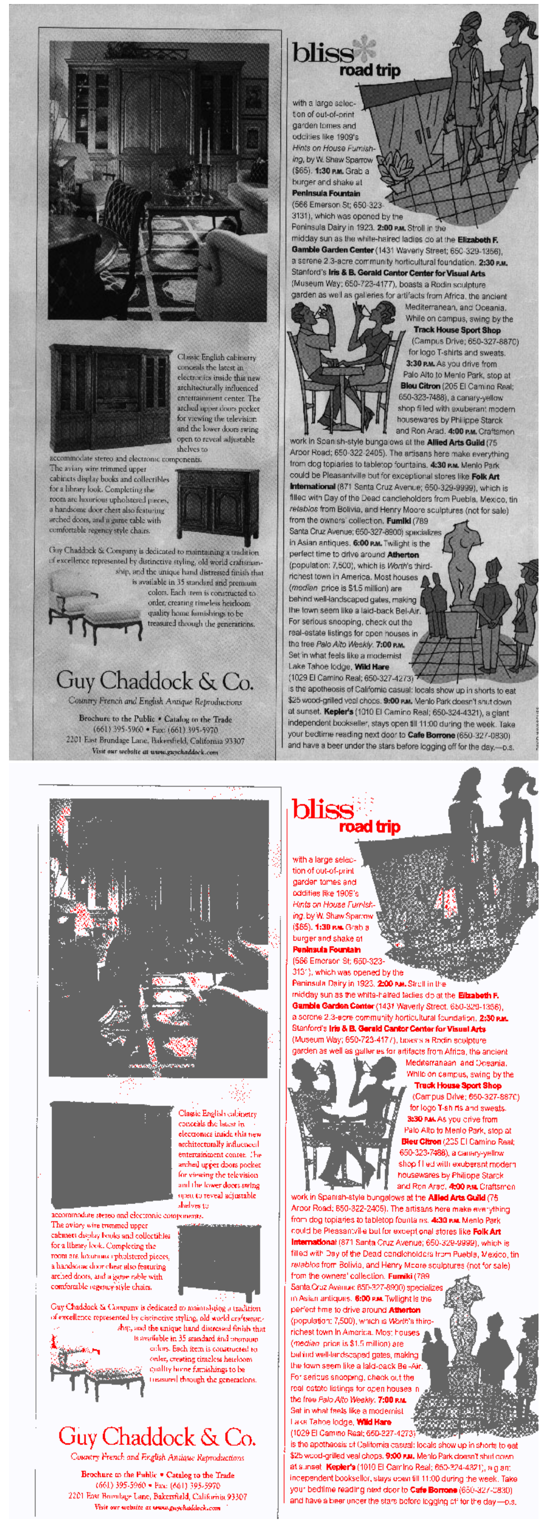

Figure 2: Original and segmented document into text (black), image (gray), and background (white)