Anisotropic Mean Shift Based Fuzzy C-Means Segmentation of Dermoscopy Images

Huiyu Zhou, Gerald Schaefer, *Member, IEEE*, Abdul H. Sadka, *Senior Member, IEEE*, and M. Emre Celebi, *Member, IEEE*

***Abstract*—Image segmentation is an important task in analysing dermoscopy images as the extraction of the borders of skin lesions provides important cues for accurate diagnosis. One family of segmentation algorithms is based on the idea of clustering pixels with similar characteristics. Fuzzy c-means has been shown to work well for clustering based segmentation, however due to its iterative nature this approach has excessive computational requirements. In this paper, we introduce a new mean shift based fuzzy c-means algorithm that requires less computational time than previous techniques while providing good segmentation results. The proposed segmentation method incorporates a mean field term within the standard fuzzy c-means objective function. Since mean shift can quickly and reliably find cluster centers, the entire strategy is capable of effectively detecting regions within an image. Experimental results on a large dataset of diverse dermoscopy images demonstrate that the presented method accurately and efficiently detects the borders of skin lesions.**

***Index Terms*—Dermoscopy, fuzzy c-means, image segmentation, mean shift, melanoma, skin cancer.**

I. Introduction

MALIGNANT melanoma, the most deadly form of skin cancer, is one of the most rapidly increasing cancers in the world, with an estimated incidence of 62 480 and an estimated total of 8420 deaths in the United States in 2008 alone [1]. Early diagnosis is particularly important since melanoma can be cured with a simple excision if detected early.

Dermoscopy, one of the major tools for the diagnosis of melanoma, is a noninvasive skin imaging technique that involves optical magnification which makes subsurface structures more readily visible compared to conventional clincal images [2]. This in turn reduces screening errors and provides greater differentiation between difficult lesions such as pigmented Spitz nevi and small, clinically equivocal lesions [3]. However, it has also been demonstrated that dermoscopy might lower the diagnostic accuracy in the hands of inexperienced dermatologists [4]. Therefore, in order to minimize diagnostic errors resulting from the difficulty and subjectivity of visual interpretation, the development of computerized image analysis techniques is of paramount importance.

Automatic border detection of lesions is often the first step in the automated or semi-automated analysis of dermoscopy images and is crucial for accurate diagnosis. Image segmentation can be defined as the grouping of similar pixels (i.e., lesion and non-lesion pixels) in a parametric space, where they are associated with each other in the same or different images. Fuzzy c-means (FCM) is a segmentation algorithm that is based on clustering similar pixels in an iterative way where the cluster centers are adjusted during each iteration [5]. Due to its iterative nature the computational cost of the algorithm is relatively high compared to other segmentation techniques. Hence, a number of approaches, e.g., [6] and [7], have been presented that allow for significant speedups while maintaining good segmentation performance.

In this paper we introduce a new mean shift based FCM algorithm that requires less computational time than these established techniques. The proposed method incorporates a mean field term within the standard FCM objective function. Since mean shift can quickly and reliably find cluster centers, the entire strategy is capable of effectively segmenting clusters within an image. We evaluate the proposed algorithm on a large dataset of dermoscopic images. Based on these experiments we show that our approach delivers excellent segmentation of lesions in a computationally efficient manner.

The rest of the paper is organized as follows: In Section II, the original FCM algorithm and its variants are introduced and discussed. Our proposed anisotropic mean shift based FCM approach is described in Section III. Section IV presents extensive comparative results of the proposed scheme and conventional approaches. Finally, conclusions and future directions are given in Section V.

II. Fuzzy C-Means Image Segmentation and Its Variants

A. Classical Fuzzy C-Means

Fuzzy c-means (FCM) is based on the idea of finding cluster centers by iteratively adjusting their positions and evaluation of an objective function similar to the original hard c-means, yet it allows more flexibility by introducing the possibility of partial memberships to clusters. The effect of the general FCM algorithm is illustrated in Fig. 1.

Manuscript received May 03, 2008; revised October 15, 2008. Current version published February 19, 2009. The work of H. Zhou was supported by the European Commission under Grant FP6-045189-STREP (RUSHES). The work of M. E. Celebi was supported by a grant from The Louisiana Board of Regents (LEQSF2008-11-RD-A-12). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jinshan Tang.

H. Zhou and A. H. Sadka are with the School of Engineering and Design, Brunel University, Uxbridge UB8 3PH, U.K. (e-mail: Huiyu.Zhou@brunel.ac.uk; Abdul.Sadka@brunel.ac.uk).

G. Schaefer is with the School of Engineering and Applied Science, Aston University, Birmingham, U.K. (e-mail: g.schaefer@aston.ac.uk).

M. E. Celebi is with the Department of Computer Science, Louisiana State University, Shreveport, LA 71115 USA (e-mail: ecelebi@lsus.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2008.2010631

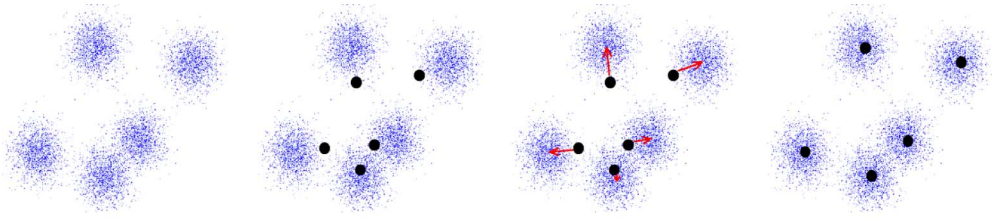

Fig. 1. Illustration of classical FCM that attempts to find appropriate cluster centers. (a) Available clusters. (b) Random centers. (c) Converging and (d) final settlement.

The objective function usually follows the form

$$E = \sum_{j=1}^{C} \sum_{i=1}^{N} \mu_{ij}^{k} \|x_i - c_j\|^2 \tag{1}$$

where μ_{ij}^{k} is the fuzzy membership of sample (or pixel) x_i and the cluster identified by its center c_j, and k is a constant that defines the fuzziness of the resulting partitions.

E can reach the global minimum when pixels nearby the centroid of corresponding clusters are assigned higher membership values, while lower membership values are assigned to pixels far from the centroid [8]. Here, the membership is proportional to the probability that a pixel belongs to a specific cluster where the probability is only dependent on the distance between the image pixel and each independent cluster center. The membership functions and the cluster centers are updated by

$$\mu_{ij} = \frac{1}{\sum_{m=1}^{C} \left(\frac{\|x_j - c_i\|}{\|x_j - c_m\|^{2/(k-1)}} \right)} \tag{2}$$

and

$$c_i = \frac{\sum_{j=1}^{N} u_{ij}^{k} x_j}{\sum_{j=1}^{N} u_{ij}^{k}}. \tag{3}$$

The steps involved in fuzzy c-means image segmentation are [5] as follows.

1) Initialize the cluster centers c_i and let $t = 0$.
2) Initialize the fuzzy partition memberships functions μ_{ij} according to (2).
3) Let $t = t+1$ and compute new cluster centers c_i using (3).
4) Repeat Steps 2 to 3 until convergence.

An initial setting for each cluster center is required and FCM converges to a local minimum. The efficiency of FCM has been comprehensively investigated in [9]. To effectively address the inefficiency of the original FCM algorithm several variants of the FCM algorithm have been introduced which we cover briefly in the following subsections.

B. Fast FCM With Random Sampling (RSFCM)

To reduce the computational requirements of FCM, Cheng *et al.* [6] proposed a multistage random sampling strategy. This method has a lower number of feature vectors and also needs fewer iterations to converge. The basic idea is to randomly sample and obtain a small subset of the dataset in order to approximate the cluster centers of the full dataset. This approximation is then used to reduce the number of iterations. The random sampling FCM algorithm consists of two phases. First, a multistage iterative process of a modified FCM is performed. Phase 2 is then a standard FCM with the cluster centers approximated by the final cluster centers from Phase 1.

Phase 1: Randomly initialize the cluster centers c_i

Let $X_{(\Delta\%)}$ be a subset whose number of subsamples is $\Delta\%$ of the N samples contained in the full dataset X and denote the number of stages as n_s. ϵ_1 and ϵ_2 are parameters used as stopping criteria. After the following steps the dataset (denoted as $X_{(n_s * \Delta\%)}$) will include $N * \Delta\%$ samples.

1) Select $X_{(\Delta\%)}$ from the set of the original feature vectors matrix ($z = 1$).
2) Initialize the fuzzy memberships functions μ_{ij} using (2) with $X_{(z*\Delta\%)}$.
3) Compute the stopping condition $\epsilon = \epsilon_1 - z*((\epsilon_1 - \epsilon_2)/n_s)$ and let $j = 0$.
4) Set $j = j + 1$.
5) Compute the cluster centers $c_{(z*\Delta\%)}$ using (3).
6) Compute $\mu_{(z*\Delta\%)}$ using (2).
7) If $\|\mu^{j}_{(z*\Delta\%)} - \mu^{j-1}_{(z*\Delta\%)}\| \geq \epsilon$, then go to Step 4.
8) If $z \leq n_s$ then select another $X_{(\Delta\%)}$ and merge it with the current $X_{(z*\Delta\%)}$ and set $z = z + 1$, otherwise move to Phase 2 of the algorithm.

Phase 2: FCM clustering

1) Initialize μ_{ij} using the results from Phase 1, i.e., $c_{(n_s*\Delta\%)}$ with (3) for the full data set.
2) Go to Step 3 of the conventional FCM algorithm and iterate the algorithm until stopping criterion ϵ_2 is met.

Evidence has shown that this improved FCM with random sampling is able to reduce the computation requested in the classical FCM method [10]. Other variants of this multistage random sampling FCM framework have also been developed and can be found, e.g., in [11] and [12].

C. Enhanced FCM (EnFCM) and Variants

Ahmed *et al.* [13] introduced an alternative to the classical FCM by adding a term that enables the labelling of a pixel to be associated with its neighborhood. As a regulator, the neighborhood term can change the solution towards piecewise homogeneous labelling. As a further extension of this work, Szilágyi *et al.* [7] introduced their EnFCM algorithm where, in order to reduce the computational complexity, a linearly weighted sum image g is formed from the original image, and the local neighbor average image evaluated as

$$g_m = \frac{1}{1+\alpha} \left(x_m + \frac{\alpha}{N_R} \sum_{j \in N_r} x_j \right) \tag{4}$$

where g_m denotes the gray value of the m-th pixel of the image g, x_j represents the neighbors of x_m, N_R is the cardinality of a cluster, N_r represents the set of neighbors inside a window around x_m.

The objective function used for segmenting image g is defined as

$$J = \sum_{i=1}^{C} \sum_{l=1}^{q_c} \gamma_l \mu_{il}^m (g_l - c_i)^2 \tag{5}$$

where q_c denotes the number of the gray levels in the image, and γ_l is the number of the pixels having an intensity equal to l, which refers to intensity levels with $l = 1, 2, \ldots, q_c$. Thus, $\sum_{l=1}^{q_c} \gamma_l = N$ under the constraint that $\sum_{i=1}^{C} \mu_{il} = 1$ for any l.

Finally, we can obtain the following expressions for membership functions and cluster centers [14]:.

$$\mu_{il} = \frac{(g_l - s_i)^{-2/m-1}}{\sum_{j=1}^{C} (g_l - s_j)^{-2/m-1}} \tag{6}$$

and

$$s_i = \frac{\sum_{l=1}^{q_c} \gamma_l \mu_{il}^m g_l}{\sum_{l=1}^{q_c} \gamma_l \mu_{il}^m}. \tag{7}$$

EnFCM considers a number of pixels with similar intensities as a weight. Thus, this process may accelerate the convergence of searching for global similarity. On the other hand, to avoid image blur during the segmentation, which may lead to inaccurate segmentation, Cai *et al.* [14] utilizes a measure S_{ij} in a fast generalized FCM algorithm (FGFCM), which incorporates the local spatial relationship S_{ij}^s and the local gray-level relationship S_{ij}^g, and is defined as

$$S_{ij} = \begin{cases} S_{ij}^s \times S_{ij}^g, & j \neq i \\ 0, & j = i \end{cases} \tag{8}$$

with

$$S_{ij}^s = \exp\left(\frac{-\max(|p_{cj} - p_{ci}|, |q_{cj} - q_{ci}|)}{\lambda_s} \right) \tag{9}$$

and

$$S_{ij}^g = \exp\left(\frac{-||x_i - x_j||^2}{\lambda_g \times \sigma_g^2} \right) \tag{10}$$

where (p_{ci}, q_{ci}) describe the coordinates of the i-th pixel, σ_g is a global scale factor of the spread of S_{ij}^s, and λ_s and λ_g represent scaling factors. S_{ij} replaces α in (4).

Hence, the newly generated image g is updated as

$$g_i = \frac{\sum_{j \in N_i} S_{ij} x_j}{S_{ij}} \tag{11}$$

and is restricted to [0, 255] due to the denominator.

Given a predefined number of clusters C and a threshold value $\epsilon > 0$, the reported FGFCM algorithm [14] proceeds in the following steps.

1) Initialize the clusters c_j.
2) Compute the local similarity measures S_{ij} using (8) for all neighbours and windows over the image.
3) Compute linearly-weighted summed image g using (11).
4) Update the membership partitions using (6).
5) Update the cluster centers c_i using (7).
6) If $\sum_{i=1}^{C} ||c_{i(\text{old})} - c_{i(\text{new}}||^2 > \epsilon$ go to Step 4.

Similar efforts to improve the computational efficiency and robustness have also been reported in [15] and [16].

D. Other FCM Variants

Other variants of the classical FCM algorithm can be classified into two groups: those with added spatial constraints, and those with optimization of termination conditions or objective functions.

1) FCM With Spatial Constraints: There are certain similarities when considering spatial contents of an image. For example, a number of regions in the image can be very similar to each other in intensity or colour. These similarities can be labelled before any segmenting process starts. During the actual segmenting process, one of the areas from the similar groups will be utilized for segmentation, while the others may be directly assigned to the same clusters as the former with little computational effort. This is the basis of one strategy to applying spatial constraints to the standard FCM.

Pham [17] proposed an improved FCM objective function with an added spatial penalty term in the membership functions. This technique needs some extented computational efforts to search for an appropriate penalty term. However, the entire FCM scheme is of lower computational complexity upon determination of the penalty term.

psFCM, as proposed by Hung and Yang [18], is a two-stage scheme. A smaller data set is extracted from the entire image using the classical $k - d$ tree method, followed by a standard FCM segmentation which uses the cluster centers previously generated. This strategy reduces the computational requirements of the FCM segmentation significantly. Eschrich *et al.* [11] presented the brFCM algorithm, which can reduce the number of distinct patterns by aggregating similar examples and then using a weighted exemplar in the FCM process.

2) FCM With Optimisation of Functionals: Modifications or adjustment of membership functions can also be used to reduce the number of iterations required in the FCM scheme. The motivation behind this strategy is the possibility of simplifying the original membership functions or modifying the classical convergence criterion so as to accelerate the segmentation procedure.

Höppner [10] re-organized the original data sets as a tree before segmentation starts, leading to fast convergence of the later process. Unfortunately, this re-organisation is not an ideal model in the presence of large data sets or increasing number of clusters [19]. Cannon *et al.* [20] reported a speed-up factor of 6 for an improved FCM scheme by look-up tables for exponential and distance function. Frequent updating of the standard FCM can be used to reduce the iterations and hence improves the computational efficiency [21].

A similarity-driven cluster merging method was proposed by Xiong *et al.* [22]. This method takes into account the similarity between clusters by a fuzzy cluster similarity matrix, and an adaptive threshold is used for merging. De Gruijter and McBratney [23] modified the objective function to account for

outliers (extragrades) and hence improve the performance of the FCM in noisy environments.

III. Anisotropic Mean Shift Based FCM

In this subsection, we will present a new combinatorial approach to fuzzy c-means segmentation that utilizes an anisotropic mean shift algorithm coupled with fuzzy segmentation.

Mean shift based techniques have been shown to be capable of estimating the local density gradients of similar pixels. These gradient estimates are iteratively performed so that all pixels can find similar pixels in corresponding images [24], [25]. A standard mean shift approach method uses radially symmetric kernels. Unfortunately, the temporal coherence will be reduced in the presence of irregular structures and noise in the image. This reduced coherence may not be properly detected by radially symmetric kernels and thus, an improved mean shift approach, namely anisotropic kernel mean shift [26], provides better performance.

A. Proposed Algorithm

In mean shift algorithms the image clusters are continuously moved along the gradient of the density function before they become stationary. Those points gathering in an outlined area are treated as the members of the same segment. To determine the membership of an image point a density estimate at the point needs to be conducted. In other words, similarity computation must be achieved between this point and the center of the segment. Furthermore, the coherence between this point and its surrounding image points needs to be discovered (e.g., colour or intensity consistency), as this coherence can be used to remove any inconsistency such as image artifacts or noise. In this subsection, we mainly discuss about the estimation of the density function of an image point (this kernel density estimation is also known as the Parzen window technique).

The motivation of introducing the density estimation based segmentation is that the image space can be represented by empirical probability density functions (PDF) of certain parameters (e.g., color or intensity). Dense or sparse regions of similar image points correspond to local maxima or minima of the PDF (or the modes of the unknown density) [25]. After the modes have been located in the image, the membership of an image point to a particular segment will be determined.

A kernel density estimate on an image point is defined by

$$\tilde{f}(x) = \frac{1}{N}\sum_{i=1}^{N} K(x - x_i) \tag{12}$$

with

$$K(x) = |H|^{-1/2} K(H^{-1/2}x) \tag{13}$$

where N is the number of samples, and x_i stands for a sample from an unknown density function f. $K(\cdot)$ is the d-variate kernel function with compact support satisfying the regularity constraints, and H is a symmetric positive definite $d \times d$ bandwidth matrix. Usually, we have $K(x) = k_e(x)$, where $k_e(x)$ is a convex decreasing function, e.g., for a Gaussian kernel

$$k_e(x) = c_t e^{-x/2} \tag{14}$$

or for an Epanechnikov kernel

$$k_e(x) = c_t \max(1 - x, 0) \tag{15}$$

where c_t is a normalising constant.

If a single global spherical bandwidth is applied, $H = h^2\mathbf{I}$ (where $\mathbf{I}$ is the identity matrix), then we have the classical form as

$$\tilde{f}(x) = \frac{1}{Nh^d}\sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right). \tag{16}$$

Since the kernel can be divided into two different radially symmetric kernels, we have the kernel density estimate as

$$\begin{aligned}\tilde{f}(x;\mathbf{c}) = \sum_{i=1}^{N} & \frac{1}{N(h^\alpha)^p(h^\beta)^q} k^\alpha \\ & \times \left(\left\|\left(\mathbf{c}^\alpha - x_i^\alpha\right)/h^\alpha\right\|^2\right) \\ & \times k^\beta\left(\left\|\left(\mathbf{c}^\beta - x_i^\beta\right)/h^\beta\right\|^2\right)\end{aligned} \tag{17}$$

where $\mathbf{c}$ represents a vector of cluster centers, p and q are two ratios, and α and β denote the spatial and temporal components respectively [26]. Classical mean shift utilizes symmetric kernels that may experience a lack of temporal coherence in the regions where the intensity gradients exist with a slope relative to the evolving segment. In contrast, anisotropic kernel mean shift links with every data point by an anisotropic kernel. This kernel associated with a pixel can update its shape, scale and orientation. The density estimator is represented by

$$\begin{aligned}\tilde{f}(x;\mathbf{c}) = \frac{1}{N}\sum_{i=1}^{N} & \frac{1}{h^\beta\left(H_i^\alpha\right)^q} k^\alpha \\ & \times \left(d\left(\mathbf{c}^\alpha, x_i^\alpha, H_i^\alpha\right)\right) \\ & \times k^\beta\left(\left\|\left(\mathbf{c}^\beta - x_i^\beta\right)/\left(h^\beta H_i^\alpha\right)\right\|^2\right)\end{aligned} \tag{18}$$

where $d\left(c_i^\alpha, x_i^\alpha, H_i^\alpha\right)$ is the Mahalanobis distance

$$d\left(\mathbf{c}^\alpha, x_i^\alpha, H_i^\alpha\right) = \left(x_i^\alpha - \mathbf{c}^\alpha\right)^T H_i^{\alpha-1}\left(x_i^\alpha - \mathbf{c}^\alpha\right). \tag{19}$$

Anisotropic mean shift is intended to modulate the kernels during the mean shift procedure. The objective is to keep reducing the Mahalanobis distance so as to group similar samples as much as possible. First, the anisotropic bandwidth matrix H_i^α is estimated using a standard radially symmetric diagonal H_i^α and h^β. The neighborhood of pixels around $\mathbf{c}$ has the following constraints:

$$\begin{cases} k_e^\alpha\left(d\left(\mathbf{c}, x_i, H_i^\alpha\right)\right) < 1 \\ k_e^\beta\left(\left\|\left(\mathbf{c} - x_i\right)/\left(h^\beta H_i^\alpha\right)\right\|^2\right) < 1. \end{cases} \tag{20}$$

Fig. 2. Examples of AMSFCM iterative segmentation. (a) Original image. (b) Three iterations. (c) Four iterations. (d) Six iterations.

A new full matrix $\bar{H}_i^{\alpha}$ will use the variance of $(\mathbf{c} - x_i)$ as its components. To show how the modulation of $\bar{H}_i^{\alpha}$ happens, we first decompose the required bandwidth matrix to

$$\bar{H}_i^{\alpha} = \lambda V\, AV^T \tag{21}$$

where λ is a scalar, V is a matrix of normalized eigenvectors, and A is a diagonal matrix of eigenvalues whose diagonal elements a_i satisfy [26]

$$\prod_{i=1}^{p} a_i = 1. \tag{22}$$

The bandwidth matrix is updated by adding more and more points to the computational list: the more image points with similar colour or intensity gather in the same segments, the less total Mahalanobis distance between the image points and the centers of individual segments will be obtained [refer to (19)–(22)].

In the proposed algorithm we combine fuzzy c-means and anisotropic mean shift segmentation. A significant difference between our approach and other similar methods is that our algorithm continuously inherits and updates the states, based on the interaction of FCM and mean shift. Stemming from the algorithm reported in [26], the proposed anisotropic mean shift based FCM (AMSFCM) proceeds in the following steps.

1) Initialize the cluster centers c_j. Let the iteration count $t = 0$.
2) Initialize the fuzzy partitions μ_{ij} using (2).
3) Increment $t = t + 1$ and compute c_j using (3) for all clusters.
4) Update μ_{ij} using (2). This is an FCM process.
5) For each pixel x_i one needs to estimate the density with anisotropic kernels and related color radius using (18)–(21). For simplicity, $\bar{H}_i^{\alpha}$ can just apply variances at the diagonal items with other zero components. Note that mean shift is employed after the FCM stage.
6) Calculate the mean shift vector and then iterate until the mean shift, $M^{+}(x_i) - M^{-}(x_i)$, is less than 0.01 considering the previous position and a normalized position change—see the equation shown at the bottom of the page. with $\nu = 0.5$.
7) Merge pixels that possess less Mahalanobis distances than the predefined thresholds.
8) Repeat Steps 3) to 7) until $|\mu_{ij}^{t} - \mu_{ij}^{t-1}| < \epsilon_0$ (ϵ_0 is a pre-set threshold).

Fig. 2 illustrates how the segmentation evolves using the proposed AMSFCM algorithm. In this example, the segmentation optimally converges after six iterations.

B. Convergence Behaviors

1) Classical FCM: Classical FCM is one of the sub-optimal segmentation algorithms, which sacrifices global optimality to the improved numerical efficiency and flexibility of the segmentation process. The computational cost of FCM heavily depends on the number of image points that need to be processed in each iteration.

To obtain a global minimal solution, we differentiate both sides of (1) with respect to c_k and then set them to zero

$$\frac{\partial E}{\partial c_k} = \frac{\partial}{\partial c_k} \sum_{j=1}^{C} \sum_{i=1}^{N} \mu_{ij}^{k} \|x_i - c_j\|^2 = 0. \tag{23}$$

The right-hand side of (1) has an upper bound that leads to

$$\sum_{j=1}^{C} \sum_{i=1}^{N} \mu_{ij}^{k} \left(\|x_i - c_m\|^2 + \|c_m - c_j\|^2 \right) \geq \sum_{j=1}^{C} \sum_{i=1}^{N} \mu_{ij}^{k} (x_i - c_j)^2 \tag{24}$$

where c_m stands for the mean value of c_j. Introducing (24) into (23), one can observe that the derivative of E with respect to c_k will be dominated by the sum of distance between c_m and c_j. The faster this distance is reduced, the better asymptotic performance the entire FCM holds.

2) Proposed AMSFCM: To find cluster centers we can utilize the gradient of the density estimator. Let

$$\tilde{k}\left(\left\| \frac{\mathbf{c} - x_i}{\tilde{h}} \right\|^2 \right) = k^{\alpha} \left(d\left(\mathbf{c}^{\alpha}, x_i^{\alpha}, H_i^{\alpha} \right) \right) k^{\beta} \times \left(\left\| \frac{\mathbf{c}^{\beta} - x_i^{\beta}}{h^{\beta} H_i^{\alpha}} \right\|^2 \right). \tag{25}$$

$$M^{+}(x_i) = \nu M^{-}(x_i) + (1 - \nu) \frac{\sum_{j=1}^{N} (x_j - M^{-}(x_i)) \left\| \left(M^{-}(x_i^{\beta}) - x_j^{\beta} \right) / \left(h^{\beta} H_j^{\alpha} \right) \right\|^2}{\sum_{j=1}^{N} \left\| \left(M^{-}\left(x_i^{\beta} \right) - x_j^{\beta} \right) / \left(h^{\beta} H_j^{\alpha} \right) \right\|^2}$$

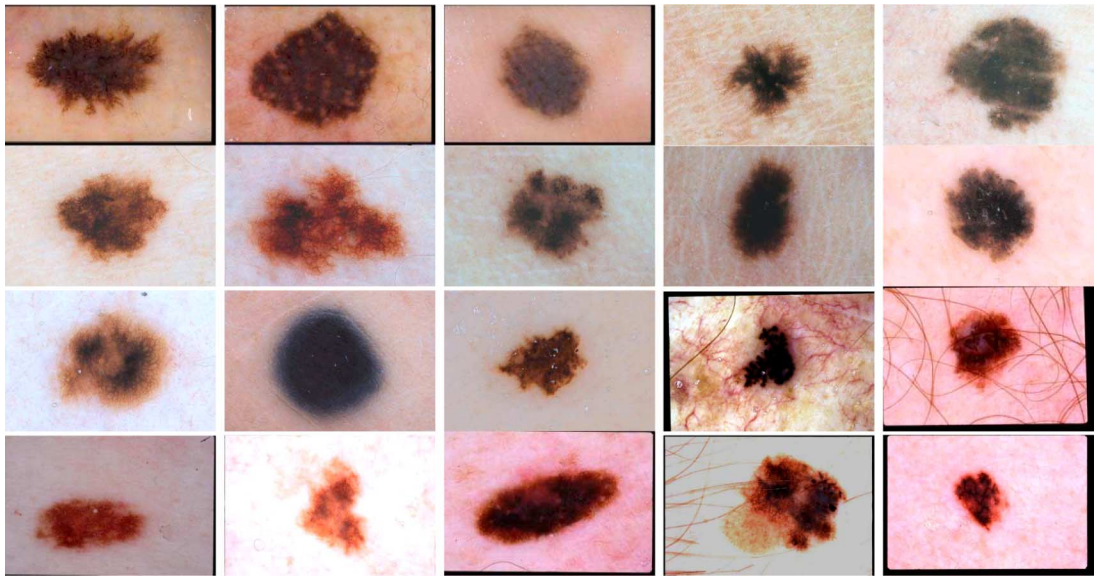

Fig. 3. Subset of the dermoscopic image set used in the evaluation.

Then we have

$$\frac{\partial \tilde{f}(x;\mathbf{c})}{\partial \mathbf{c}} = C_s \left[\sum_{i=1}^{N} g\left(\left\|\frac{\mathbf{c}-\mathbf{x_i}}{h}\right\|^2\right)\right] \times \left[\frac{\sum_{i=1}^{N} x_i g\left(\left\|\frac{\mathbf{c}-\mathbf{x_i}}{h}\right\|^2\right)}{\sum_{i=1}^{N} g\left(\left\|\frac{\mathbf{c}-\mathbf{x_i}}{h}\right\|^2\right)} - \mathbf{c}\right] \quad (26)$$

in which the constant C_s is expressed as

$$C_s = \frac{2c_d}{Nh^{d+2}} \quad (27)$$

and c_d is the corresponding normalization constant [25]

$$\begin{aligned} & g\left(\left\|\frac{\mathbf{c}-\mathbf{x_i}}{h}\right\|^2\right) \\ &= \frac{\partial}{\partial \mathbf{c}} \tilde{k}\left(\left\|\frac{\mathbf{c}-x_i}{\tilde{h}}\right\|^2\right) \\ &= \left(d\left(\mathbf{c}^{\alpha}, x_i^{\alpha}, H_i^{\alpha}\right)\right)' k_1' k_2 \\ &\quad + k_1 k_2' \frac{2\mathbf{c}^{\beta-1}\left(\mathbf{c}^{\beta}-x_i^{\beta}\right)}{h^{2\beta}\left(H_i^{\alpha}\right)^2} \end{aligned} \quad (28)$$

and

$$\times \begin{cases} k_1 = k^{\alpha}\left(d\left(\mathbf{c}^{\alpha}, x_i^{\alpha}, H_i^{\alpha}\right)\right), \\ k_2 = k^{\beta}\left(\left\|\frac{\mathbf{c}^{\beta}-x_i^{\beta}}{h^{\beta}H_i^{\alpha}}\right\|^2\right). \end{cases} \quad (29)$$

The regulation term is

$$\tilde{f}_r(x;\mathbf{c}) = C_s \left[\sum_{i=1}^{N} g\left(\left\|\frac{\mathbf{c}-\mathbf{x_i}}{h}\right\|^2\right)\right] \quad (30)$$

and the mean shift term is

$$\mathbf{m}_s(x;\mathbf{c}) = \left[\frac{\sum_{i=1}^{N} x_i g\left(\left\|\frac{\mathbf{c}-\mathbf{x_i}}{h}\right\|^2\right)}{\sum_{i=1}^{N} g\left(\left\|\frac{\mathbf{c}-\mathbf{x_i}}{h}\right\|^2\right)} - \mathbf{c}\right]. \quad (31)$$

Referring to [25], we can obtain the following expression:

$$\mathbf{m}_s(x;\mathbf{c}) = C_v \frac{\frac{\partial \tilde{f}(x;\mathbf{c})}{\partial \mathbf{c}}}{\tilde{f}_r(x;\mathbf{c})} \quad (32)$$

where C_v is a constant. According to the Capture Theorem [27], the trajectories of the gradient method introduced here are attracted by local maxima if they are unique stationary points within a small neighborhood. In other words

$$\frac{\partial \tilde{f}(x;\mathbf{c})}{\partial \mathbf{c}} = 0. \quad (33)$$

In due convergence, we will have an optimal c_m from (33), where the magnitude of the mean shift vector approaches 0.

Let us revisit (24), (32), and (33). It has been proven that the mean shift with a form as in (26) converges if the kernel $K(\cdot)$ has a convex and monotonically decreasing profile [25]. While the kernel function is approaching to its convergence, the mean value c_m of the cluster centers is available and can be a prediction for next iteration of FCM segmentation. This helps reduce computational efforts addressed in the FCM segmentation procedure afterwards. The convergence speed of the mean shift relies on the $(\partial \tilde{f}(x;\mathbf{c}))/(\partial \mathbf{c})$, which normally is very fast due to fast mean calculation.

IV. Experimental Evaluation

The proposed segmentation algorithm was evaluated on a set of 100 dermoscopy images (30 invasive malignant melanoma and 70 benign) obtained from the EDRA Interactive Atlas of Dermoscopy [2] and the dermatology practises of Dr. A. Marghoob (New York, NY), Dr. H. Rabinovitz (Plantation, FL) and Dr. S. Meznies (Sydney, Australia). The benign lesions included nevocellular nevi and dysplastic nevi. A subset of the images is shown in Fig. 3. Manual borders were obtained by selecting a number of points on the lesion border, connecting these with a 2nd-order B-spline and finally filling the resulting closed curve. Three sets of manual borders were determined

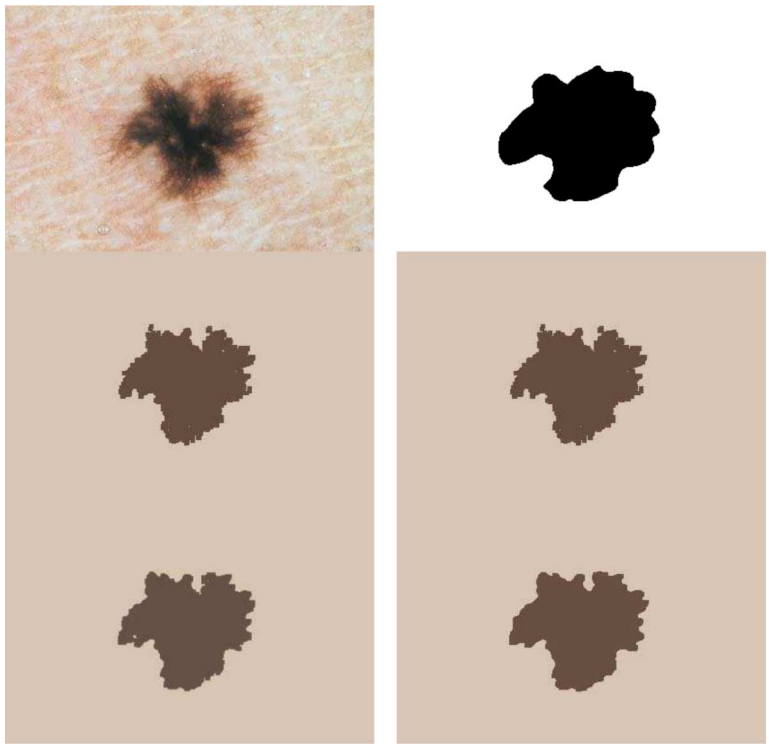

Fig. 4. Segmentation comparison of original image (upper left), ground truth (upper right), FCM (middle left), RSFCM (middle right), EnFCM (bottom left) and AMSFCM (bottom right) for image 15.

by dermatologists Dr. W. Stoeckker, Dr. J. Malters, and Dr. J. Grichnik using this method and serve as a ground truth for the experiments.

For our experimental evaluation, we used a PC with Intel(R) Core(TM)2 CPU (2.66 GHz) and 2 GB RAM. The algorithms that we compared are conventional FCM [5], EnFCM [7], RSFCM [6], and the proposed AMSFCM. In a final stage, morphological processing is employed for smoothing the segmentation outcomes, especially the image borders and removing small isolated areas.

An example of the segmentations obtained by the various algorithms is given in Fig. 4 which shows one of the ground truth segmentations together with the results by all four methods. It can be observed that the segmentations produced by classical FCM and RSFCM are less smooth than those by EnFCM and AMSFCM. This is due to: 1) RSFCM uses FCM in the second phase so they both have approximate convergence characteristics and 2) EnFCM and AMSFCM take into account weighted image pixels so their outcomes are smoothed in the FCM stage. Clearly, smoother borders are more realistic and also conform better to the manual segmentations derived by the dermatologists. The second observation is also reflected in Fig. 5, where original images are segmented using different FCM algorithms and the lesion borders are then extracted. It is also noticed that different algorithms generate similar results forigure 5, while the proposed AMSFCM algorithm has clearly the best border result for the third example.

For each image segmentation we record the number of True Positives TP (the number of pixels that were classified both by the algorithm and the expert as lesion pixels), True Negatives TN (the number of pixels that were classified both by the algorithm and the experts as non-lesion pixels), False Positives FP (the number of instances where a non-lesion pixel was falsely classified as part of a lesion by an algorithm) and False Negatives FN (the number of instances where an lesion pixels was falsely classified as non-lesion by an algorithm). From this we can then calculate the sensitivity SE (or true positive rate) as

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{34}$$

and the specificity SP (or true negative rate) as

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}. \tag{35}$$

In Table I we list the sensitivity and specificity obtained by all algorithms over the entire database and compared to all three ground truth segmentations (average SE and SP based on all three manual segmentations are reported). It can be seen that the proposed AMSFCM performs significantly better with a median sensitivity of about 78% while the other algorithms achieve only a sensitivity of about 74%. In addition, our algorithm provides more consistent results as indicated by the lower variance of SE. As specificity is fairly similar for all algorithms, we can conclude that AMSFCM provides the best segmentation on the given dataset.

As we have noted before, computational efficiency is a crucial issue when considering FCM based segmentation. We record the number of iteration required in each FCM approach for evaluation, which in turn enables us to make a comparison regarding the relative efficiency of the different approaches. We normalized them so that the classical FCM algorithm is assigned 1.00 while the other ones represent the relative fractions they take

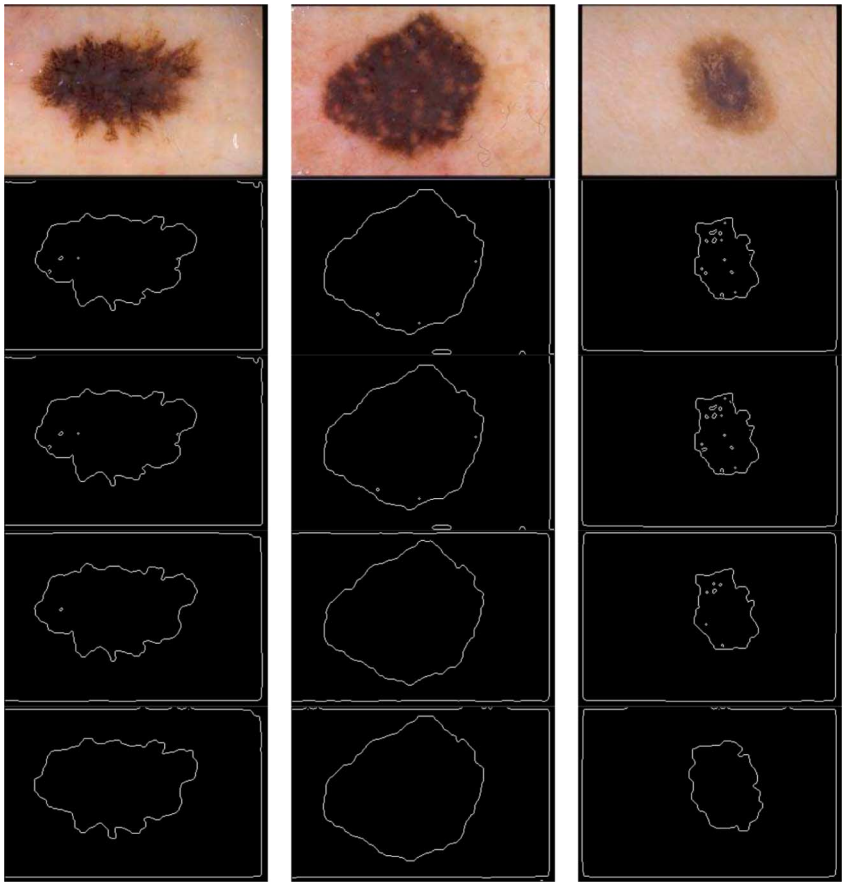

Fig. 5. Border detection of exemplar segmented images (row 1—original images; row 2—FCM results; row 3—RSFCM results; row 4—EnFCM results and row 5—AMSFCM results).

TABLE I
SEGMENTATION PERFORMANCE ON THE COMPLETE DATASET. FOR EACH ALGORITHM THE MEDIAN SENSITIVITY AND SPECIFICITY ARE GIVEN. THE VALUES IN BRACKETS INDICATE THE STANDARD DEVIATIONS OF THE MEASURES

Algorithm	Sensitivity	Specificity
FCM	0.739 (0.120)	0.99 (0.056)
RSFCM	0.738 (0.118)	0.99 (0.052)
EnFCM	0.740 (0.118)	0.99 (0.061)
AMSFCM	0.776 (0.113)	0.99 (0.065)

TABLE II
EFFICIENCY ANALYSIS OF THE DIFFERENT ALGORITHMS. REPORTED IS THE RELATIVE EFFICIENCY COMPARED TO THE CONVENTIONAL FCM ALGORITHM. THE VALUES IN BRACKETS INDICATE THE STANDARD DEVIATIONS OF THE MEASURES

Algorithm	computational cost
FCM	1.00 (0.00)
RSFCM	0.67 (0.11)
EnFCM	0.80 (0.09)
AMSFCM	0.63 (0.09)

compared to this. The results are presented in Table II from which it can be seen that the proposed AMSFCM takes computation efforts of 37%, 4%, and 17% less than compared to FCM, RSFCM and EnFCM respectively.

Overall, it is evident that the proposed approach provides a very useful tool for the analysis of dermoscopic images. Not only does it provide the best segmentation results among the algorithms investigated, it also is the most efficient method.

V. CONCLUSION

Fuzzy c-means based algorithms are frequently used to segment medical images but are also computational intensive. In this paper we have introduced a new mean shift based fuzzy c-means segmentation algorithm. The proposed method incorporates a mean field term within the standard fuzzy c-means objective function. Based on a large set of dermoscopic images, we have shown that the proposed segmentation technique AMSFCM is not only more efficient than other fuzzy c-means approaches but that it is also capable of providing superior segmentation. The developed algorithm hence provides a useful tool as a first stage in the automatic or semi-automatic analysis of skin lesion images.

ACKNOWLEDGMENT

The assistance of W. V. Stoecker, MD, J. M. Malters, MD, and J. M. Grichnik, MD, in obtaining the manual borders is gratefully acknowledged.

REFERENCES

[1] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, and M. Thun, "Cancer statistics, 2008," *CA Cancer J. Clin.*, vol. 58, no. 2, pp. 71–96, 2008.

[2] G. Argenziano, H. Soyer, and V. De Giorgi, *Dermoscopy: A Tutorial*. Milan, Italy: EDRE Medical Publishing & New Media, 2002.
[3] K. Steiner, M. Binder, M. Schemper, K. Wolff, and P. H. , "Statistical evaluation of epiluminescence dermoscopy criteria for melanocytic pigmented lesions," *J. Amer. Acad. Dermatol*, vol. 29, no. 4, pp. 581–588, 1993.
[4] M. Binder, M. Schwarz, A. Winkler, A. Steiner, A. Kaider, K. Wolff, and H. Pehamberger, "Epiluminescence microscopy. a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists," *Arch. Dermatol.*, vol. 131, no. 3, 1995.
[5] J. Bezdek, "A convergence theorem for the fuzzy isodata clustering algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 1, pp. 1–8, Jan. 1980.
[6] T. Cheng, D. Goldgof, and L. Hall, "Fast fuzzy clustering," *Fuzzy Sets Syst.*, vol. 93, pp. 49–56, 1998.
[7] L. Szilágyi, Z. Benyo, S. Szilagyi, and H. Adam, "Mr brain image segmentation using an enhanced fuzzy c-means algorithm," in *Proc. 25th Annu. Information Conf. IEEE EMBS*, 2003, pp. 17–21.
[8] K. Chuang, S. Tzeng, H. Chen, J. Wu, and T. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Comput. Med. Imag. Graph.*, vol. 30, pp. 9–15, 2006.
[9] R. Hu and L. Hathaway, "On efficiency of optimization in fuzzy c-means," *Neural, Parallel Sci. Comput.*, vol. 10, pp. 141–156, 2002.
[10] F. Höppner, "Speeding up fuzzy c-means: Using a hierarchical data organisation to control the precision of membership calculation," *Fuzzy Sets Syst.*, vol. 128, no. 3, pp. 365–376, 2002.
[11] S. Eschrich, J. Ke, L. Hall, and D. Goldgof, "Fast accurate fuzzy clustering through data reduction," *IEEE Trans. Fuzzy Syst.*, vol. 11, pp. 262–270, 2003.
[12] J. Kolen and T. Hutcheson, "Reducing the time complexity of the fuzzy c-means algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 263–267, Apr. 2002.
[13] M. Ahmed, S. Yamany, N. Mohamed, A. Farag, and T. Moriaty, "A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 21, no. 3, pp. 193–199, Mar. 2002.
[14] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognit.*, vol. 40, no. 3, pp. 825–838, 2007.
[15] J. Leski, "Toward a robust fuzzy clustering," *Fuzzy Sets Syst.*, vol. 137, pp. 215–233, 2003.
[16] S. Chen and D. Zhang, "Robust image segmentation using fcm with spatial constraints based on new kernel-induced distance measure," *IEEE Trans. Syst., Man, Cybern. B*, vol. 34, no. 4, pp. 1907–1916, Aug. 2004.
[17] D. Pham, "Fuzzy clustering with spatial constraints," in *Proc. IEEE Int. Conf. Image Processing*, 2002, pp. 65–68.
[18] M.-C. Hung and D.-L. Yang, "An efficient fuzzy c-means clustering algorithm," in *Proc. 2001 IEEE Int. Conf. Data Mining*, Washington, DC, 2001, pp. 225–232, IEEE Computer Society.
[19] Y.-S. Chen, B. Chen, and W. Hsu, "Efficient fuzzy c-means clustering for image data," *J. Electron. Imag.*, vol. 14, no. 1, pp. 013017-1–013017-13, 2005.
[20] R. Cannon, J. Dave, and J. Bezdek, "Efficient implementation of the fuzzy c-means clustering algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 2, pp. 248–255, 1986.
[21] M. Kamel and S. Selim, "New algorithms for solving the fuzzy clustering problem," *Pattern Recognit.*, vol. 27, no. 3, pp. 421–428, 1994.
[22] X. Xiong, K. Chan, and K. Tan, "Similarity-driven cluster merging method for unsupervised fuzzy clustering," in *Proc. 20th Conf. Uncertainty in Artificial Intelligence*, Arlington, VA, 2004, pp. 611–618, AUAI Press.
[23] J. De Gruijter and A. McBratney, *Classification and Related Methods of Data Analysis, Elsevier, Amsterdam, The Netherlands*. Amsterdam, The Netherlands: Elsevier, 1988.
[24] D. Comaniciu and P. Meer, "Mean shift analysis and applications," in *Proc. IEEE Conf. Computer Vision*, 1999, pp. 1197–1203.
[25] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
[26] J. Wang, B. Thiesson, Y. Xu, and M. Cohen, "Image and video segmentation by anisotropic kernel mean shift," in *Proc. Eur. Conf. Computer Vision*, 2004, pp. 238–249.
[27] D. Betsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.

Huiyu Zhou received the B.Eng. degree in radio technology from Huangzhong University of Science and Technology, Huangzhong, China, in 1990, the M.Sc. degree in biomedical engineering from the University of Dundee, Dundee, U.K., in 2002 and the Ph.D. degree in computer vision from the Herriot-Watt University, Edinburgh, Scotland in 2006.

Currently, he is a Research Fellow in the School of Engineering and Design, Brunel University, Uxbridge, U.K. His research interests include computer vision, human motion analysis, intelligent systems, and human computer interface. He has published widely in international journals and conferences.

Gerald Schaefer (M'04) received the B.Sc. degree in computing from the University of Derby, Derby, U.K., and the Ph.D. degree in computer vision from the University of East Anglia, Norwich, U.K.

He was with the Colour and Imaging Institute, University of Derby, as a Research Associate (1997–1999), as Senior Research Fellow at the School of Information Systems, University of East Anglia (2000–2001), and as Senior Lecturer in Computing at the School of Computing and Informatics at Nottingham Trent University (2001–2006). In September 2006, he joined the School of Engineering and Applied Science at Aston University, Birmingham, U.K. His research interests include color image analysis, physics-based vision, image retrieval, and image coding.

Abdul H. Sadka (SM'07) is the head of Electronic and Computer Engineering and the Director of the Centre for Media Communications Research at Brunel University, Uxbridge, U.K. He has almost 15 years of experience in academic leadership and excellence and is an internationally renowned expert in visual media processing and communications with an extensive track record of scientific achievements and peer-recognized research excellence. To date, he has attracted over $2M worth of research grant and contracts in his capacity as a Principal Investigator. He has published widely in international journals and conferences and is the author of a highly regarded book, *Compressed Video Communications* (Wiley, 2002). He holds three patents in the video transport and compression area. He acts as a Scientific Advisor and Consultant to several key companies in the international telecommunications sector.

M. Emre Celebi (M'03) received the B.Sc. degree in computer engineering from Middle East Technical University, Ankara, Turkey, in 2002 and the M.Sc. and Ph.D. degrees in computer science and engineering from the University of Texas at Arlington in 2003 and 2006, respectively.

He is currently an Assistant Professor in the Department of Computer Science, Louisiana State University, Shreveport. His research interests include medical image analysis, color image processing, content-based image retrieval, and open-source software development.