Automatic Segmentation of Abdominal Fat from CT Data

Amol Pednekar1  Alok N. Bandekar1  Ioannis A. Kakadiaris1  Morteza Naghavi2
1Visual Computing Lab, Dept. of Computer Science, Univ. of Houston, Houston, TX, USA
2Association for Eradication of Heart Attack, Houston, TX, USA

Abstract

Abdominal visceral fat accumulation is one of the most important cardiovascular risk factors. Currently, Computed Tomography and Magnetic Resonance images are manually segmented to quantify abdominal fat distribution. The manual delineation of subcutaneous and visceral fat is labor intensive, time consuming, and subject to inter- and intra-observer variability. An automatic segmentation method would eliminate intra- and inter-observer variability and provide more consistent results. In this paper, we present a hierarchical, multi-class, multi-feature, fuzzy affinity-based computational framework for tissue segmentation in medical images. We have applied this framework for automatic segmentation of abdominal fat. An evaluation of the accuracy of our method indicates bias and limits of agreement comparable to the inter-observer variability inherent in manual segmentation.

1. Introduction

A recent survey reveals that over 64% of Americans are overweight or obese [3]. The presence of excess fat in the abdomen (out of proportion to the total body fat) is an independent predictor of morbidity. Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are often used to quantify abdominal fat distribution. CT is preferred over MRI because it provides good contrast between fat and non-fat tissues and it is less expensive. In current clinical practice, quantification of the fat tissue is performed manually by an expert who has a good knowledge of the anatomy. It is a very time consuming and cumbersome process subject to inter and intra-observer variations. Thus, the development of a computer-assisted method which provides unbiased and consistent results with minimal human intervention is highly desirable.

The normal CT attenuation ranges for the fat tissues is defined as the interval within (−190 < Hu < −30) [12]. However, the average Hounsfield units (Hu) for the fat tissue varies across subjects and it also depends on the CT scanner [12]. Thus, assessment of fat distribution requires a case-specific flexible attenuation range. Recent techniques estimate the abdominal fat distribution using a threshold within the mean plus-minus two standard deviations [12]. This method, called flexible threshold method (FTM), is independent of the relative spatial location of the pixel with the neighborhood pixels. As a result, this segmentation using an intensity-based threshold alone is not accurate. The Fuzzy Connectedness-Based Image Segmentation framework developed by Udupa and his collaborators [14] effectively captures the fuzzy “hanging togetherness” (perception of an object region in spite of heterogeneity of scene intensity) of image elements specified by their strength of scene connectivity. This framework was further extended to take into consideration object scale [11], relative strength of connectedness [9], and multiple objects [10]. A hybrid segmentation method [13] that combines fuzzy connectedness segmentation, Voronoi diagram classification, and deformable model based smoothing algorithms was applied for the segmentation of adipose tissue from whole body MRI scans. In our previous work, we have presented a composite fuzzy affinity using dynamic weights for the affinity components [6].

In this paper, we present an automatic tissue segmentation method which combines the intensity and texture information with local “hanging togetherness” within a tissue class. The fuzzy connectedness constraint for the object extraction in the frameworks presented in [6, 7, 14] is relaxed to allow global segmentation. Thus, instead of applying a space-invariant global threshold value, our computational framework adapts the threshold value locally to account for the local “hanging togetherness” of the tissue to be segmented. Specifically, we present a hierarchical, multi-class, multi-feature, fuzzy affinity-based framework for tissue segmentation. The global class affinity is computed based on discrepancy measures of the given image element (spatial element - spel) pair with respect to the learned distributions of the prominent and neighboring objects in the image. We use the Mahalanobis metric as the discrepancy measure. The local fuzzy affinity between two spels is computed based on their spatial nearness as well as the similarity of their intensity and texture-based features. The Mahalanobis metric is used to compute the similarity of spels in the intensity and texture-based feature space. The most discriminant combination of texture features for specific object regions and for a specific modality are determined in the
training phase of our framework. Specifically, during the training phase we compute the first order and second order statistics at various scales covering the entire scale spectrum of the image. The Fisher’s criterion is used to quantize the discriminating power of combinations of texture features for the tissue classes of interest. The most discriminating feature combinations are determined based of the cumulative discriminating power of each feature.

The remainder of the paper is organized as follows. Section 2 details the formulation of our method. In Section 3, we describe abdominal fat segmentation in CT images using our framework. In Section 4, we present results from our method and a comparison with previous methods.

2. Methods

2.1. Hierarchical Multi-Class Multi-Feature Fuzzy Affinity

In this section, we review our formulation of hierarchical, multi-class, multi-feature, fuzzy affinity using a required set of terminology, definitions, and framework introduced in [14]. An image is considered as a two-dimensional Euclidean space \( R^2 \) subdivided into spatial elements (spels) called pixels. A digital space \( Z^2 \), where the coordinates of pixel correspond to a point, is the set of all pixels of \( R^2 \). For any fuzzy relationship \( \rho \), the strength of \( \rho \) between \( c \) and \( d \) is represented by a membership function \( \mu_{\rho}(c, d) \). If a fuzzy relation \( \alpha = \{(c, d), \mu_{\alpha}(c, d)\} \in Z^2 \) is reflexive and symmetric, it is said to be a fuzzy spel adjacency, which describes the spatial relationship between the two spels. For any pixels \( c, d \in Z^2 \), \( \mu_{\alpha}(c, d) \) is assumed to be a hard adjacency relation, such that:

\[
\mu_{\alpha}(c, d) = \begin{cases} 
1 & \text{if } ||c - d|| \leq 1, \\
0 & \text{otherwise},
\end{cases}
\]

where \( ||c - d|| \) represents the Euclidean distance between \( c \) and \( d \). The pair \( (Z^2, \alpha) \), where \( \alpha \) is a fuzzy spel adjacency is called a fuzzy digital space. The concept of fuzzy digital space characterizes the underlying digital grid system independent of any image related concepts. A scene over a fuzzy digital space \( (Z^2, \alpha) \) is a pair \( \mathcal{C} = (C, f) \) where \( C = \{c - b_j \leq c_j \leq b_j\} \) for some \( b \in Z^2 \) is a finite two-dimensional rectangular array of pixels, \( f \) is a scene intensity function whose domain is \( C \), called the scene domain, and the range is a set of integers \([L, H]\). If \( C \) is a scene over \( Z^2 \) in which the range of \( f \) is \([0,1]\), then \( C \) is called a binary scene over \( (Z^2, \alpha) \). In an object class identification process the aim is to capture the local “hanging togetherness” of pixels.

2.1.1 Local Fuzzy Spel Affinity

Any fuzzy relation \( \kappa \) in \( C \) is said to be a fuzzy spel affinity in \( C \) if it is reflexive and symmetric. We define the local fuzzy spel affinity \( (\mu_{\kappa}) \) to consist of three primary components:

1) an object feature intensity component \( (\mu_{\phi}) \), 2) an intensity homogeneity component \( (\mu_{\psi}) \), and 3) a texture feature component \( (\mu_{\varphi}) \), as follows:

\[
\mu_{\kappa}(c, d) = \mu_{\alpha}(c, d)g(\mu_{\phi}(c, d), \mu_{\psi}(c, d), \mu_{\varphi}(c, d)).
\]

We combine the affinity components to form a composite local fuzzy spel affinity. Thus, the fuzzy relation \( \kappa \) in \( Z^2 \) indicates the degree of local “hanging togetherness” of pixels \( c \) and \( d \) in the vector space of feature vectors:

\[
x = \frac{1}{2}(f(c) + f(d)), f(c) - f(d), \frac{1}{2}(t(c) + t(d))\]

where \( f(c) \) and \( f(d) \) are the image intensities, and \( t(c) \) and \( t(d) \) are the texture features at pixels \( c \) and \( d \). The similarity of the pixels’ feature vectors is computed using the Mahalanobis metric:

\[
m_d^2 = (x_{(c-d)} - \bar{x}_{(c-d)})^T S_{x_{(c-d)}}^{-1} (x_{(c-d)} - \bar{x}_{(c-d)}),
\]

where \( x_{(c-d)} \), \( \bar{x}_{(c-d)} \), \( S_{x_{(c-d)}} \) are the feature vector, the mean feature vector, and the covariance matrix in the direction from \( c \) to \( d \). The bias in intensity in a specific direction is thus accounted for by allowing different levels and signs of intensity homogeneities in different directions of adjacency. Thus, this formulation accounts for different levels of the change in intensity values in the horizontal (east, west) or vertical (north, south) directions. The advantage of using the Mahalanobis metric is that it weights the differences in various feature dimensions by the range of variability. Another advantage of using the Mahalanobis metric for discrimination is that the distances are computed in units of standard deviation from the group mean. This allows us to assign a statistical probability to that measurement. The local fuzzy spel affinity is computed as:

\[
\mu_{\kappa}(c, d) = \frac{1}{1 + m_d}
\]

to ensure \( \mu_{\kappa}(c, d) \in Z^2 \rightarrow [0,1] \) and it is reflexive and symmetric. Thus, \( \mu_{\kappa}(c, d) \) defines the probability of the pixel pair belonging to the target object class. The threshold for the class identifier can be set based on the probability distribution of a specific feature space for a particular application.

2.1.2 Global Class Affinity

In our hierarchical fuzzy affinity, the local pixel affinities are assigned only if the probability of \( c \) and \( d \) belonging to the neighboring objects’ classes is much less than 0.01. The neighboring objects are defined as the objects with common boundaries in Euclidean space. For a given pixel pair \((c, d)\), we compute the discrepancy measure with respect to
the learned distributions of neighboring classes. We compute the discrepancy measure of a pixel pair from a known class in terms of its Mahalanobis distance. Then, the minimum discrepancy measure which provides the probability of pixel pair belonging to a certain class is given by:

$$J(c, d) = \min_{1 \leq i \leq b} m_d(c, d),$$

where $b$ is the number of neighboring classes to the target object. If $J(c, d) < 3$ for any neighboring class distribution other than the target object class then the local pixel affinity $\mu_{\kappa(c,d)}$ is set to zero. Otherwise, its local pixel affinity is computed as described in Section 2.1.1.

In summary, our image segmentation framework consists of the following steps:

I. Training Phase

A. Estimate object-specific feature distributions using a training data set.
   
   **Step 1:** Compute relevant features for domain specific objects.
   
   **Step 2:** Compute the most discriminant features.
   
   **Step 3:** Construct a template (mean shape) using the landmark points in the training images for the seed region (Not applicable in all domains).

II. Deployment Phase

B. Initialize the target object seed region.
   
   **Step 4:** Compute the target object seed pixel using domain specific knowledge.
   
   C. Compute the fuzzy connectedness-based object.
   
   **Step 5:** Compute global class affinity for a given spel.
   
   **Step 6:** If the spel is not determined to be a member of non-target objects, then compute local fuzzy affinity.
   
   **Step 7:** Compute the global object affinity.
   
   **Step 8:** Compute the fuzzy extent of the target object.

3. Abdominal fat segmentation in CT images

We apply our hierarchical, multi-class, multi-feature, fuzzy affinity framework for the automatic segmentation of the abdominal fat from CT images. The abdominal fat consists of visceral, retroperitoneal, and subcutaneous fat. Figure 1 depicts a 3D visualization of a subject’s CT data with two CT slices depicted as orthoslices in a volume.

Figure 1: Visualization of a subject’s CT data.

3.1. Experimental Data

The study population consisted of 80 randomly chosen subjects with 5 CT scan images per patient. The data were obtained by a CT abdominal scan between the fourth and fifth lumbar vertebrae (L4-L5) (i.e., at the level of umbilicus). Scanning was performed at 130 kV and 200 mA and the field of view ranged from 30 to 50 cm. Slice thickness was 6.0 mm in all subjects. We divided the data into 40 training and 40 test data sets.

3.2. Steps 1-2: Training Phase

The average CT value for the fat tissue is well separated from the rest of the tissue. However, the variance of the average CT value for fat for a specific subject and for a specific scanner is unknown. In addition, the intensity-based discrimination of fat from non-fat tissue does not capture the local spatial “hanging togetherness” of the fat. We use the texture features of the fat and its neighboring non-fat tissue types to increase the discriminating power. Specifically, first we computed Laws’ features [4] and Gabor’s texture features [5] for the abdominal CT images. In the training phase, we used 25 Laws’ features and a Gabor filter bank consisting of 4 scales and 6 orientations. We manually labelled in the 200 slices the regions of fat, nonfat, and background. We extracted pixel features for the classes fat and non-fat from the feature images. The most discriminating feature combinations were determined according to their cumulative discriminating power. We found that spot-spot (ss) and spot-level (sl) Laws’ features are the most discriminating. The convolution of spot with spot, and spot with level kernels provides the ss and sl filters. Furthermore, the combination of intensity and ss features provided the highest discrimination between fat and non-fat tissues. Thus,
our feature vector consists of the pixel pair intensity and the ss Laws’ feature in a neighborhood of 5x5 pixels. Since the average CT values for the fat tissue are well separated from other tissues we can separate the non-fat distribution into two distributions \((-190 < H_u < -30)\) [12]. These distributions were used to find the probability of pixel pair belonging to the non-fat class.

3.3. Steps 3-4: Automatic seed initialization

During the deployment phase of our framework, we obtain the sample statistics of a tissue by choosing a seed point and computing sample statistics around it. Selection of the seed point is very critical. We obviate the need of a manual seed selection by automatic seed initialization using Active Shape Models (ASM) [2].

Template construction: We select the seed using a subcutaneous fat template. To that end, we construct a Point Distribution Model for selected landmark points around the subcutaneous fat area. We first threshold the original images within the threshold range of \((-190 < H_u < -30)\) to obtain the fat region. Then, we label \((n = 36)\) landmark points - \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\) - in each binary image. We create a subcutaneous fat template by selecting 36 landmark points based on the anatomy of the abdominal region. Landmarks points are selected based on the curvature of the boundary of the subcutaneous fat area. Each template is represented as:

\[
x = \{x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n\}^T.
\]

Corresponding points are then aligned to remove affine differences in the images using Procrustes analysis [8], before obtaining the mean shape \(\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i\). Figure 2 depicts the mean shape of the subcutaneous fat template. To obtain the typical modes of variation of the shape, Principal Component Analysis is applied and only the first \(t\) eigenvectors corresponding to the largest \(t\) eigenvalues accounting for a specified percentage of the variation (95%) are retained. Any shape in the class can thus be represented as:

\[
x \approx \bar{x} + \Phi b,
\]

where \(b\) is a vector of weights for each variation mode and \(\Phi\) is the matrix consisting of \(t\) principal modes of variation. To construct a gray-level appearance model, we obtain a gray level profile of intensities along the normal to each landmark. Then, the mean normalized derivative as well as the covariance of each profile is computed across all images.

Seed initialization: During the deployment phase of our framework, we select the seed region by fitting the subcutaneous fat template to the thresholded CT image. Specifically, the seed point is chosen as the centroid of the region enclosed by the landmarks 14, 15, 16, 30, 31, and 32. Figure 3 depicts the seed point and the location of the 36 landmarks after ASM fitting. Next, the subcutaneous and visceral fat is segmented by employing Steps 5-8 in our framework.

4 Results and Discussion

We evaluated the results of our method (UHAFS: University of Houston Automatic Fat Segmentation) against expert manual segmentations of subcutaneous and visceral fat for 20 subjects. Experienced physicians/radiologists delineated the fat region in the CT slices which were used as gold standard. The qualitative results of our algorithm for Subjects 1-5 are depicted in Figs. 5 and 7. We evaluated the results of our algorithm by computing the three measures of accuracy recommended by [13]. Specifically, we compute the false negatives (FN), false positives (FP), true negatives (TN), and true positives (TP) by computing the number of pixels that were classified as the background and the region of interest (ROI), both correctly and incorrectly. The true positive rate is defined as the percentage of correctly classified object (i.e., fat tissue) pixels to the total number of object pixels. The true negative rate is defined as the percentage of
correctly identified non-object (i.e., non-fat tissue and background) to the total number of non-object pixels. Accuracy is defined as the percentage of correctly classified pixels (in object and non-object) with respect to the total number of pixels in the images. Accuracy denotes the degree to which the segmentation agrees with the ground truth.

Figure 8 depicts the accuracy, the true positive rate, and the true negative rate obtained using FTM and UHAFS. Note that the true positive rates of the two methods are comparable, but UHAFS has a higher true negative rate and it is more accurate. Figure 4 depicts the overlap ratio within the 95% confidence intervals obtained by applying the two algorithms 10 times in each image.

We define the false negative fraction (FNF) to indicate the fraction of tissue that is included in the ground truth but missed by the method: FNF = 100*FN / (TP+FN). The false positive fraction (FPF) indicates the amount of tissue falsely identified by the method as a fraction of the total amount of tissue in the ground truth: FPF = 100*FP / (TP+FN). Figure 6 reveals a significant decrease in the FNF and a slight increase in the FPF obtained by UHAFS as compared to FTM pointing to the reliability of our method.

The agreement between the expert manual segmentations and UHAFS for the fat area computation as well as the inter-observer variability were assessed according to the method described by Bland and Altman [1]. Let FR1 denote the fat area labelled by the first reader, FR2 denote the fat area labelled by the second reader, and FR3 denote the fat area labelled by the third reader. The inter-observer and UHAFS mean biases and variabilities for the first reader, the second reader, and the third reader, for the fat area estimation are depicted in Figs. 9 and 10. Finally, our method exhibits a considerable reduction in time as compared to manual tracing (from 5-10 mins to less than 10 seconds).

5 Conclusions
In this paper, we have presented an automatic and robust segmentation method for quantifying abdominal fat in CT images across subjects and across scanners. We have also successfully applied our framework for automatic extraction of endocardial and epicardial contours of the left ventricle in cine bFFE MR images [7].

References
Figure 6: (a) False negative fraction and (b) false positive fraction for UHAFS and FTM.


Figure 7: (a, e, i) Original CT images from subject-3, subject-4, and subject-5, respectively. (b, f, j) Manually segmented images (white pixels denote fat tissue). (c, g, k) Segmentation results using FTM. (d, h, l) Segmentation results using UHAFS.

Figure 8: Performance evaluation of UHAFS and FTM: (a) accuracy (%), (b) true positive rate (%), and (c) true negative rate (%).
Figure 9: Bland-Altman analysis pertaining to the manual segmentation.

Figure 10: Bland-Altman analysis pertaining to the results obtained using UHAFS.