Segmentation for Medical Image Mining: A Technical Report

Ashraf Elsayed[1], Frans Coenen[1], Marta García-Fiñana[2] and Vanessa Sluming[3]

The University of Liverpool, Liverpool L69 3BX, UK
(1) Department of computer science, {a.el-sayed, coenen}@liv.ac.uk
(2) Centre for Medical Statistics and Health Evaluation, M.Garciafinana@liv.ac.uk
(3) School of Health Sciences, vanessa.sluming@liv.ac.uk

Abstract

Magnetic resonance imaging (MRI) of the brain, followed by automated segmentation of the Corpus Callosum in midsagittal sections has important applications in neurology and neurocognitive research since the size and shape of the Corpus Callosum are shown to be correlated to sex, age, neurodegenerative diseases and various lateralized behaviour in people. The segmenting of the Corpus Callosum is regarded as a critical step in image mining frameworks to classify the brain MR images. In the example presented here a test collection of 76 brain MRI images representing musician and non-musician is used. Two algorithms, based on established work, for segmenting the Corpus Callosum in brain MR images, are presented and evaluated; based on this evaluation a new algorithm is also proposed. The segmentation algorithm operates by first extracting regions satisfying the statistical characteristics (gray level distributions) of the Corpus Callosum that have relatively high intensity values. This is then processed using graph analysis and classification procedures. Test using the musician data set have provided promising results.

1 Introduction

Discovering knowledge from data stored in alphanumeric databases, such as relational databases, has been the focal point of much work in data mining. However, with advances in secondary storage capacity, coupled with a relatively low storage cost, more and more non-standard data is being accumulated. One category of "non-standard" data is image data (others include free text, video, sound, etc). There is currently a very substantial collection of image data that can be mined to discover new and valuable knowledge. The central research issue in image mining is how to preprocess image sets so that they can be represented in a form that supports the application of data mining algorithms. A common representation is that of feature vectors were each image is represented as vector. Typically each vector represents some subset of feature values taken from some global set of features. A trivial example is where images are represented as primitive shape and colour pairs. Thus the global set of tuples might be:

{{blue square}, {red square}, {yellow square}, {blue circle}, {red circle}, {yellow circle}}

which may be used to describe a set of images:

{{blue square}, {red square}, {red circle}}
{{red square}, {yellow square} {blue circle}, {yellow circle}}
{(red box}, {red circle}, {yellow circle}}

However, before this can be done it is first necessary to identify the image objects of interest (i.e. the squares and circles in the above example). A common approach to achieving this is known as "segmentation". Segmentation is the process of finding regions in an image (usually referred to as objects) that share some common attributes (i.e. they are homogenous in some sense).

The process of image segmentation can be helped/enhanced for many applications if there is some application dependent domain knowledge that can be used in the process. In the context of the work described here the author's are interested in MRI "brain scans', and in particularly a specific feature within these scans called the Corpus Callosum. An example image is given in Figure 1. The Corpus Callosum is of interest t to researchers for a number of reasons:

1. The size and shape of the Corpus Callosum are shown to be correlated to sex, age, neurodegenerative diseases (such as epilepsy) and various lateralized behaviour in people.
2. It is conjectured that the size and shape of the Corpus Callosum reflects certain human characteristics (such as a mathematical or musical ability).
3. It is a very distinctive feature in MRI brain scans.
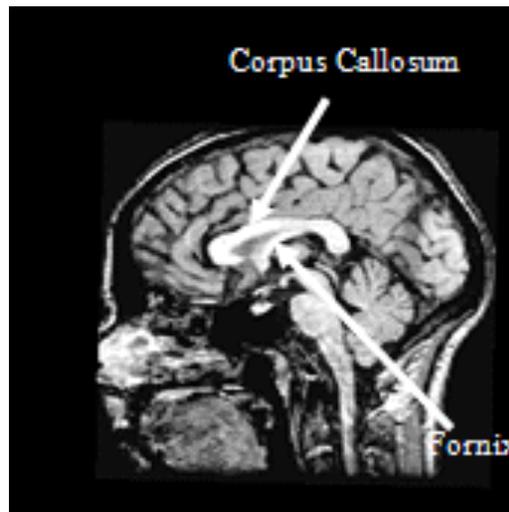
Figure 1: Corpus Callosum in a midsagittal brain MR image.

Several studies indicate that the size and shape of the Corpus Callosum in human brains are correlated to sex [1, 2, 3], age [3, 4], brain growth and degeneration [5, 6], handedness [7] and various types of brain dysfunction [8, 9, 10, 11, 12]. In order to find such correlations in living brains, Magnetic Resonance Imaging (MRI) is regarded as the best method to obtain cross-sectional area and shape information about the Corpus Callosum. In addition, MRI is fast and safe, without any radiation exposure to the subject such as with x-ray CT. Since manual tracing of Corpus Callosum in MRI data is time consuming, operator dependent, and does not directly give quantitative measures of cross-sectional areas or shape, there is a need

for automated and robust methods for localization, delineation and shape description of the Corpus Callosum.

With reference to Figure 1 the authors are interested in building classifiers from existing sets of "training" images that can then be applied to "unseen data", for example to predict epilepsy. The building of classifiers from tabular data is well understood and many examples can be found in the literature. Typical technologies that can be used to build classifiers include Decision Trees (Quinlan 1993) [13], Support Vector Machines (Vapnick 1997) [14], Classification Association Rule Mining (Coenen et al 2007) [15], etc. The process for building classifiers is therefore not an issue. However, in the context of image classification, as noted above, the issue is how to identify the relevant features within the image set to be included in the classifier construction. Thus, with respect to the particular application of interest here, the issue is how to detach the Corpus Callosum region from the rest of the image. From the above one approach is to adopt a form of image segmentation. A large number of segmentation techniques have been reported in the literature [16-23]. This paper reports on several of these in the context of the specific application of interest to the authors: the segmentation of MRI scans to identify the Corpus Callosum region (object) so that it can be further processed and used as input to classification generators.

The paper is organized as follows. Section 2 gives a brief overview of image mining, and Section 3 gives a generic discussion of image segmentation. Two segmentation algorithms are described in details in sections 4 and 5. A new segmentation algorithm is presented in section 6. Experiments and results are illustrated in section 7. This report is concluded in section 8.

2 Image Mining

Advances in image acquisition and storage technology have led to great growth in very large and detailed image databases [24]. A huge amount of image data is generated in our daily life and each field, such as medical image (CT images, ECT images and MR images etc), satellite images and all kinds of digital photographs. These images involve a great number of useful and implicit information that is difficult for users to discover. Image mining can automatically discover this implicit information and patterns from this high volume of images. Image mining is more than just an extension of data mining to image domain; it can be viewed as an interdisciplinary endeavor that draws upon computer vision, image processing, image retrieval, machine learning, artificial intelligence, database and data mining, etc. Research in image mining can be broadly classified into two main directions. The first direction involves domain-specific applications where the focus is to extract the most relevant image features into a form suitable for data mining [25, 26, 27]. The second direction involves general applications where the focus is to generate image patterns that maybe helpful in the understanding of the interaction between high level human perceptions of images and low level image features [28, 29, 30]. Data mining in medical images belongs to the first direction. There are two kinds of frameworks used to characterize image mining systems: function-driven versus information-driven. The first focuses on the functionalities of different component modules to organize image mining systems while the latter is designed as a hierarchical structure with special emphasis on the information needs at various levels in the hierarchy. Zhang et al. [31] proposes an information-driven framework that aims to highlight the role of information at various levels of representation. There are four levels of

information, starting from the lowest Pixel Level, the Object Level, the Semantic Concept Level, and finally to the highest Pattern and Knowledge Level.

3 Segmentation Algorithms

Image segmentation is, arguably, the most important component in the medical image mining process; it is certainly the start point. Segmentation is concerned with the automated division of images into non-overlapping regions. There are characteristic artifacts of medical images called partial-volume effects which make the segmentation task complicated, examples area shadows and signal dropout due to the orientation dependence of acquisition that can result in missing boundaries [32]. Partial-volume effects are artifacts that occur where multiple tissue types contribute to a single pixel, resulting in a blurring of intensity across boundaries. Partial-volume effects are common in medical images, particularly for 3-D CT and MRI data, in which the resolution is not isotropic and, in many cases, is quite poor along one axis of the image. Further complications arise as the contrast between areas of interest is often low. A major difficulty that is specific to the segmentation of MR images is the intensity inhomogeneity artifact [32, 33, 34], which causes a shading effect to appear over the image. The artifact can significantly degrade the performance of methods that assume that the intensity value of a tissue class is constant over the image.

Three image segmentation algorithms suitable for application to brain MR images to extract the Corpus Callosum segment form these images are described in the following subsections:

1. Normalized Cuts algorithm
2. Spectral Segmentation with multiscale graph decomposition
3. Modified Spectral Segmentation

The first two are established approaches reported in the literature, the third is a new approach. Each is discussed in the three following sections.

4. Normalized Cuts Algorithm

Shi and Malik (1997) [35] proposed the Normalized Cuts algorithm to address the image segmentation problem. Broadly the algorithm is founded on Graph Theory. This algorithm is described in details in this section.

The normalized cuts algorithm treats each image pixel as a node in a graph, and addresses the segmentation problem in terms of graph partitioning [36]. More formally the set of points in an arbitrary feature space are represented as a weighted undirected graph $G = (V, E)$, where the nodes or vertices (V) of the graph are the points in the feature space; and the edges (E) represent the connections between each node to each other node. The weight of each edge linking nodes i and j, $w(i, j)$, is a function of the similarity between nodes i and j. For the segmentation (grouping) process we seek to partition the set of vertices into disjoint subsets $V_1; V_2; \ldots ; Vm$, where by some measure the similarity among the vertices in a set V_i is high and, across different sets V_i, V_j the similarity is low. This means that the Normalized Cuts algorithm measures both the total dissimilarity between the different groups as well as the total similarity within the groups.
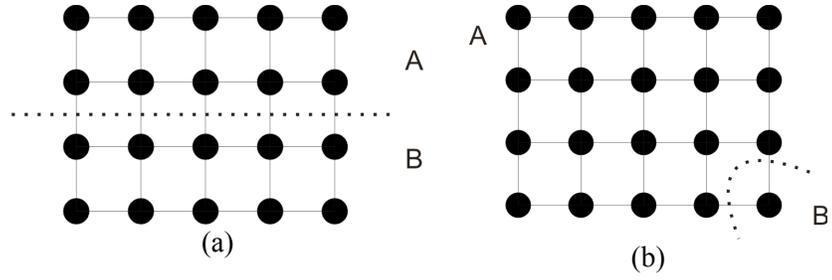
Figure 2: *Example Graph representation of a 5x4 pixel image, (a) two equally sized segments (partitions) A and B, (b) two unequally size segments (partitions) A and B.*

4.1. Fundamental Concepts

Any graph $G = (V, E)$ can be partitioned into two disjoint sets, A, B, provided |V| is greater than 1. The degree of dissimilarity between any two sets A and B (the *cut* value) can be computed as the sum of all the weightings between every node in A and every node in B, thus:

$$cut(A, B) = \sum_{u \in A, v \in B} w(u, v) \tag{1}$$

where *w(u, v)* is the similarity between node u and v. The optimal bipartitioning of a graph is the one that minimizes this cut value. To find the *minimum cut* is a complex problem whose solution requires the comparison of every possible partition. Given N nodes the potential number of partitions is:

$$\sum_{r = 1 \, to \, \left\lfloor \frac{N}{2} \right\rfloor} {}^{N}C_r$$

(The notation ${}^{N}C_r$ is taken from the field of combinatronics and means the number of different combinations of size r that exist in a set of N items.)

Finding the minimum cut is a well-studied problem and there exist efficient algorithms for solving it. Wu and Leahy [35], for example, proposed a clustering method based on the minimum cut criterion. However, the minimum cut criteria favors cutting small sets of isolated nodes in the graph, and can give a bad segmentation. This is because, using equation (1), the cut value will tend to increases with the number of edges crossing between the two partitioned segments. If the two segments are the same size (have the same number of nodes) they will be related by more edges than if they are unequally sized. This is illustrated in Figures 2(a) and (b). In Figure 2(a) partitions A and B have 10 vertices each and consequently there will be 100 (10x10) edges crossing between the partitions; however, in Figure 2(b) partition A has 19 vertices while partition B has only one vertex, consequently there are only 19 edges crossing between the partitions. Assuming the edge weights are inversely proportional to the distance between nodes any cut that partitions out individual nodes will have a smaller cut value than cuts that partition larger groupings of nodes.

To avoid this unnatural bias for partitioning out small sets of points, Shi and Malik (1997) proposed a new measure of disassociation, the normalized cut (Ncut). Given a partition of nodes of a graph, V, into two sets A and B, let x be an $N = |V|$ dimensional indicator vector, $x_i = 1$ if node i is in A and -1, otherwise. Let $d(i) = \sum_j w(i, j)$ be the total connection from node i to all other nodes. With the definitions x and d, we can rewrite $Ncut(A,B)$ as:

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)}$$

$$= \frac{\sum_{(x_i>0,x_j<0)} - w_{ij}x_i x_j}{\sum_{x_i>0} d_i} + \frac{\sum_{(x_i<0,x_j>0)} - w_{ij}x_i x_j}{\sum_{x_i<0} d_i} \qquad (2)$$

where $assoc(A,V) = \sum_{u\in A, t\in V} w(u,t)$ is the total connection from nodes in A to all nodes in the graph and $assoc(B, V)$ is similarly defined.

Let $d(i) = \sum_j w(i, j)$ be the edge weightings connecting node i to all other nodes j. Also let D be an N×N diagonal matrix with d on its diagonal, where d = [d(1), d(2), …, d(N)] and W be an N × N symmetric matrix with W(i, j) = w(i, j) (i.e. each entry in the matrix holds the weighting for the edge linking the nodes i and j).

$$k = \frac{\sum_{x_i>0} d_i}{\sum_i d_i},$$

And **1** be an N × 1 vector of all ones. Using the fact $\frac{1+x}{2}$ and $\frac{1-x}{2}$ are indicator vectors for $x_i > 0$ and $x_i < 0$, respectively, we can rewrite 4[$Ncut$(x)] as:

$$= \frac{(1+x)^T (D-W)(1+x)}{k 1^T D 1} + \frac{(1-x)^T (D-W)(1-x)}{(1-k)1^T D 1}$$

$$= \frac{(x^T (D-W)x + 1^T (D-W)1)}{k(1-k)1^T D 1} + \frac{2(1-2k)1^T (D-W)x}{k(1-k)1^T D 1}.$$

Let

$$\alpha(x) = x^T (D-W)x,$$
$$\beta(x) = 1^T (D-W)x,$$
$$\gamma = 1^T (D-W)1,$$

and

$$M = 1^T D 1,$$

we can then further expand the above equation as:

$$= \frac{(\alpha(x) + \gamma) + 2(1 - 2k)\beta(x)}{k(1 - k)M}$$

$$= \frac{(\alpha(x) + \gamma) + 2(1 - 2k)\beta(x)}{k(1 - k)M} - \frac{2(\alpha(x) + \gamma)}{M} + \frac{2(\alpha(x) + \gamma)}{M}$$

$$= \frac{(\alpha(x) + \gamma) + 2(1 - 2k)\beta(x)}{k(1 - k)M} - \frac{2(\alpha(x) + \gamma)}{M} + \frac{2\alpha(x)}{M} + \frac{2\gamma}{M}.$$

Dropping the last constant term, which in this case equals 0, we get:

$$= \frac{(1 - 2k + 2k^2)(\alpha(x) + \gamma) + 2(1 - 2k)\beta(x)}{k(1 - k)M} + \frac{2\alpha(x)}{M}$$

$$= \frac{\frac{(1 - 2k + 2k^2)}{(1 - k)^2}(\alpha(x) + \gamma) + \frac{2(1 - 2k)}{(1 - k)^2}\beta(x)}{\frac{k}{(1 - k)}M} + \frac{2\alpha(x)}{M}.$$

Letting $b = \dfrac{k}{1 - k}$, and since $\gamma = 0$, the equation becomes:

$$= \frac{(1 + b^2)(\alpha(x) + \gamma) + 2(1 - b^2)\beta(x)}{bM} + \frac{2b\alpha(x)}{bM}$$

$$= \frac{(1 + b^2)(\alpha(x) + \gamma)}{bM} + \frac{2(1 - b^2)\beta(x)}{bM} + \frac{2b\alpha(x)}{bM}$$

$$= \frac{(1 + b^2)(\alpha(x) + \gamma)}{bM} + \frac{2(1 - b^2)\beta(x)}{bM} + \frac{2b\alpha(x)}{bM} - \frac{2b\gamma}{bM}$$

where $\dfrac{2b\gamma}{bM} = 0$:

$$= \frac{(1 + b^2)(x^T(D - W)x + 1^T(D - W)1)}{b1^T D 1} + \frac{2(1 - b^2)1^T(D - W)x}{b1^T D 1}$$

$$+ \frac{2bx^T(D - W)x}{b1^T D 1} - \frac{2b1^T(D - W)1}{b1^T D 1}.$$

$$= \frac{(1 + x)^T(D - W)(1 + x)}{b1^T D 1} + \frac{b^2(1 - x)^T(D - W)(1 - x)}{b1^T D 1} - \frac{2b(1 - x)^T(D - W)(1 + x)}{b1^T D 1}$$

$$= \frac{[(1 + x) - b(1 - x)]^T(D - W)[(1 + x) - b(1 - x)]}{b1^T D 1}.$$

Setting $y = (1 + x) - b(1 - x)$, it is easy to see that:

$$y^T D \mathbf{1} = \sum_{x_i > 0} d_i - b \sum_{x_i < 0} d_i = 0$$

Putting every thing together we have:

$$\min_{A,B} N cut(A, B) = \min_y \frac{y^T (D - W) y}{y^T D y} \tag{3}$$

The above expression is the Rayleigh quotient [25]. If y is relaxed to take real values, the above equation can be minimized by solving the generalized eigenvalue system,

$$(D - W) y = \lambda D y \tag{4}$$

Interestingly the second smallest eigenvector y gives the solution of the normalized cut problem.

4.2 Algorithm

Given an image sequence I. Construct a weighted graph $G = (V,E)$ whose each node is each pixel of the image I. Let N be the number of nodes (pixels), i.e., |V|.

Step 1

Construct an N × N symmetric similarity matrix W as:

$$W_{ij} = \exp\frac{-\|F(i) - F(j)\|_2^2}{\sigma_I^2} * \begin{cases} \exp\dfrac{-\|X(i) - X(j)\|_2^2}{\sigma_X^2} & if \ \|X(i) - X(j)\|_2^2 < r \\ 0 & otherwise \end{cases} \tag{5}$$

where X(i) is the spatial location of node i, i.e., the coordinates in the original image I, and F(i) is a feature vector defined as:

• F(i) = 1 for segmenting point sets,
• F(i) = I(i), the intensity value, for segmenting brightness (gray scale) images,

Let $d(i) = \sum_j w(i, j)$ be the total connection from node i to all other nodes.

Construct an N × N diagonal matrix D with **d** on its diagonal.

Step 2

Solve a generalized eigensystem:

$$(D - W)y = \lambda D y \tag{6}$$

and get an eigenvector with the second smallest eigenvalue.

Step 3

Use the eigenvector to bipartition the graph. In the ideal case, the eigenvector should only take on two discrete values, and the signs tell us exactly how to partition the graph $\left(A = \{V_i | y_i > 0\}, B = \{V_i | y_i \leq 0\}\right)$.

However, **y** is relaxed to take real values; therefore, we need to choose a splitting point. There are several ways such as

- Take 0
- Take median
- Search a splitting point which results in that *Ncut(A, B)* is minimized.

The splitting point which minimizes Ncut value also minimizes

$$\frac{y^{T}(D - W)\, y}{y^{T} D\, y} \tag{7}$$

where $\mathbf{y} = (1 + x) - b(1 - x)$, $b = k/(1 - k)$, $k = \dfrac{\sum_{x_i > 0} d_i}{\sum_i d_i}$

where x is an N dimensional indicator vector, $x_i = 1$ if node i is in A and -1, otherwise.

To find the minimal Ncut, we need to try different values of splitting points. The optimal splitting point is generally around the mean value of the obtained eigenvector.

Step 4

Repeat bipartition recursively. Stop if Ncut value is larger than a pre-specified threshold value (Large Ncut value means that there is no clear partition point any more). Furthermore, stop if the total number of nodes in the partition (Area) is smaller than a pre-specified threshold value.

5. Spectral Segmentation with Multiscale Graph Decomposition

Cour and Shi (2005) [37] proposed a multiscale spectral image segmentation algorithm. This algorithm works on multiple scales of the image in parallel, without iteration, to capture both coarse and fine level details. The Normalized Cut graph partitioning framework of image segmentation is used (see above). They demonstrate that large image graphs can be

compressed into multiple scales capturing image structure at increasingly large neighborhoods. This segmentation algorithm works simultaneously across the graph scales, with an inter-scale constraint to ensure communication and consistency between the segmentations at each scale. This algorithm is described in details in this section.

5.1. Fundamental Concepts

Given an image I, a graph $G = (V, E, W)$ is constructed with the pixels represented by the graph nodes V, and the pixels within a distance $V \leq G_r$ are connected by a graph edge E. The weight $W(i, j)$ measures the likelihood of pixel i and j being in the same cluster. Partitioning of this graph represents the image segmentation [35,37].

Assigning weights to graph edges
The pair-wise pixel affinity graph determines the segmentation accuracy. Therefore, as recommended in [37] two simple local grouping cues are used which are the intensity and contours.

Grey Level Intensity: neighboring pixels with close intensity are most likely to be in the same region.

$$W_I(i, j) = e^{-\frac{\|X_i - X_j\|^2}{\sigma_x} - \frac{\|I_i - I_j\|^2}{\sigma_I}} \tag{8}$$

Where X_i and I_i represent pixel location and intensity.

Connecting pixels considering only intensity and location usually gives bad segmentation due to the texture that is present in the brain MR images. Therefore the principal image contours (edges) are also considered for the segmentation of brain MR images.

Dominant Contours: the image edges are considered useful when the neighboring regions have the same cutter. The affinity between two pixels is calculated by measuring the image edges between them.

$$W_C(i, j) = e^{-\max_{x \in line(i,j)} \|Edge(x)\|^2 / \sigma_C} \tag{9}$$

Where $line(i, j)$ is a straight line joining pixels i and j and $Edge(x)$ is the edge strength at location(x).

The two cues are combined in this work in the form:

$$\sqrt{W_I(i, j) \times W_C(i, j)} + \alpha W_C(i, j) \tag{10}$$

With larger G_r, the objects with faint contours pop out more clearly, but the graph affinity matrix becomes denser. A larger graph radius G_r generally makes segmentation better. Long range graph connections facilitate propagation of local grouping cues across larger image regions. This effect allows us to better detect objects with faint contours in a cluttered background.

5.2. Decomposition of graph into multiple scales

The graph links can be separated into different scales according to their underlying spatial separation:

$$W = W_1 + W_2 + ... + W_S,$$ (11)

where W_s contains affinity between pixels with certain spatial separation range:

$$W_S(i,j) \neq 0$$

only if :

$$G_{r,S-1} < r_{ij} \leq G_{r,S} \text{ where } r_{ij} = \|X_i - X_j\|^2.$$

The representative pixels can be determined at graph scale W_s as follows. For the first graph scale W_1 every pixel as graph node is taken, and connect pixels within r distance apart by a graph edge. For the second graph scale W_2, there are no short graph connections; pixels at distance *2r + 1* apart in the original image grid are sampled as representative nodes. Applying this procedure recursively, at scale s, representative pixels at *(2r + 1)^{s-1}* distance apart on the original image grid are sampled , as shown in Fig. 3.The representative pixels in each scale will be denoted by I_s, and denote W_C^S as a compressed affinity matrix with connections between the representative pixels in I_s. The different scales of the graph are defined on different layers of the image pyramid, each a sub-sample of the original image.

As Fig. 4 illustrates, such a small number of connections can have virtually the same effect as a large fully connected graph. With a maximal graph connection radius G_r, the affinity matrix W_{Full} probably doesn't fit in memory. It can be decomposed into short-range and long range connections: $W_{Full} = W_1 + W_2$, and compress W_2 with a low-rank approximation: $W_2 \approx C_{1,2}^T W_2^C C_{1,2}$. W_2^C can be computed either directly on a sub-sampled image, or by sampling values from W_1. The interpolation matrix $C_{1,2}$ from scale 2 to scale 1 will be introduced later on to couple segmentations at each scale.
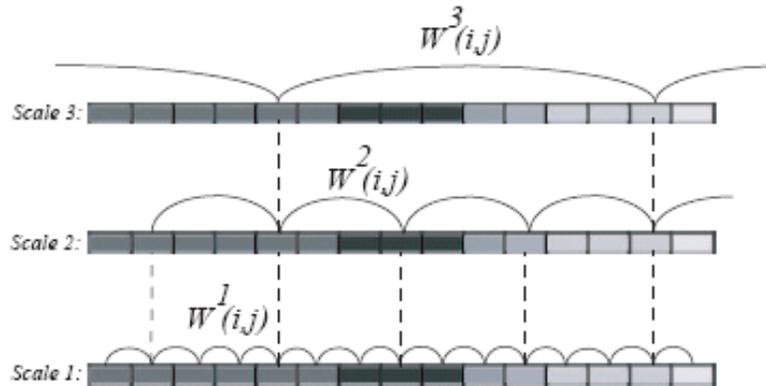
Figure 3: *1D view of multiple-scale graph decomposition with r = 1. Large radius graphs can be decomposed into different scales, each containing connections with specific range of*

spatial separation: $W=W_1+W_2+:::+W_S$. At larger scales, the graph weights vary slowly in a neighborhood, we can sample them using representative pixels at $(2r + 1)^{s-1}$ distance apart.

$$W_{\text{Full}} = W_1 + W_2 \approx W_1 + C_{1,2}{}^T W_2^c C_{1,2} = W_{\text{reconstruction}}$$
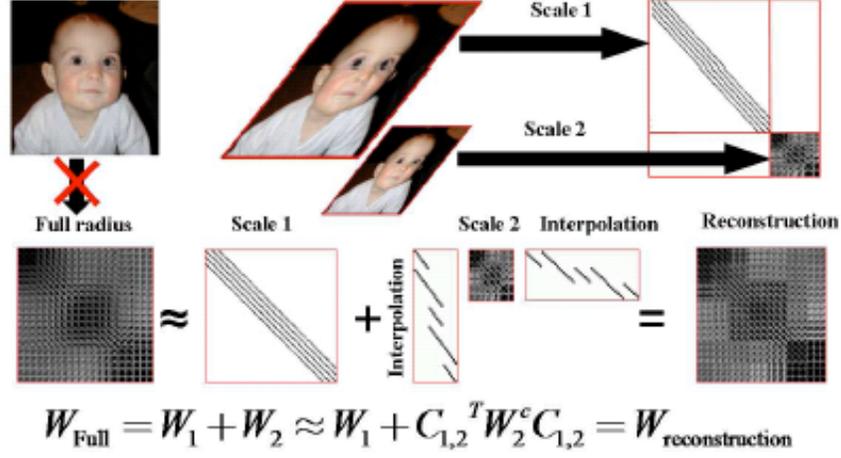
Figure 4: *Example of multiscale graph compression.*

5.3. Parallel segmentation across scales

Let $X_s \in \{0,1\}^{N_S \times K}$ be the partitioning matrix at scale s, $X_s(i,k) = 1$ iff graph node $i \in I_s$ belongs to partition k. The multiscale partition matrix X and the block diagonal multiscale affinity matrix W are formed as follows:

$$X = \begin{pmatrix} X_1 \\ . \\ . \\ X_s \end{pmatrix} , \quad W = \begin{pmatrix} W_1^C & & 0 \\ & . & \\ & & . \\ 0 & & W_s^C \end{pmatrix} \tag{12}$$

Direct partitioning of graph W gives the trivial segmentation, grouping all the nodes in a given scale as one segment. For multiscale segmentation, the segmentation costs should ensure propagation across the different scales to reach a consistent segmentation at all scales. At the finest graph scale, the segmentation should take into account graph links at all coarser levels. The cross-scale consistency is simple: the coarse-scale segmentation (X_{s+1}) should be locally an average of the fine-scale segmentation (X_s). This is done by constraining the multiscale partitioning vector X to verify: for all node i in layer I_{s+1}, $X_{s+1}(i) = \frac{1}{|N_i|} \sum_{j \in N_i} X_s(j)$.

The neighborhood N_i specifies the projection of $i \in I_{s+1}$ on the finer layer I_s, and is simply defined on a regularly spaced grid of size $\rho = 2r + 1$, the sampling factor. Then, the cross-scale interpolation matrix $C_{s,s+1}$ between nodes in layer I_s and those in coarser layer I_{s+1} can be defined as:

$$C_{s,s+1}(i,j) = \begin{cases} \dfrac{1}{|N_i|} & if \ j \in N_i \\ 0 & otherwise \end{cases} \qquad (13)$$

and the cross-scale segmentation constraint matrix C can be defined as:

$$C = \begin{pmatrix} C_{1,2} & -I_2 & & & 0 \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & \cdot \\ 0 & & & C_{1s-1,s} & -I_S \end{pmatrix} \qquad (14)$$

Where the cross-scale segmentation constraint equation:

$$CX = 0 \qquad (15)$$

With this constraint, the segmentation is forced to propagate across the scales to reach a consistent segmentation at all scales.

The segmentation criterion can be defined as the constrained multiscale Normalized Cut:

$$\max imize \, \varepsilon(X) = \frac{1}{K} \sum_{l=1}^{K} \frac{X_l^T W X_l}{X_l^T D X_l} \qquad (16)$$

Subject to $CX = 0$, $X \in \{0,1\}^{N^* \times K}$ where $N^* = \sum_s N_s$

Finding the optimal Ncut graph partitioning is NP hard. A spectral graph partitioning technique allows us to solve this problem using a continuous space solution by computing the K eigenvectors corresponding to the K largest eigenvalues in:

$$WV = \lambda DV \qquad (17)$$

Using Lagrange multipliers [37], the constrained Ncut optimization amounts to find eigenvectors of

$$\overline{W}\,\overline{V} = \lambda \overline{V} \qquad (18)$$

$$where \ \overline{W} = QD^{-1/2}WD^{-1/2}Q \qquad (19)$$

$$and \ Q = I - D^{-1/2}C^T \left(C D^{-1} C^T \right)^1 C D^{-1/2} \qquad (20)$$

Interestingly the first K eigenvectors give the solution of the constrained multiscale Normalized Cut.

5.4. Algorithm

Step 1

Given $p \times q$ image I, for s=1..S (S = #scales):

   a) sample $\dfrac{p}{\rho} \times \dfrac{q}{\rho}$ pixels $i \in I_s$ from I_{s-1} on a regular grid, where is the sampling factor.

   b) compute constraint $C_{s-1,s}(i,j) = \dfrac{1}{|N_i|}$ $\forall j \in N_i$ sampling neighborhood of i.

   c) compute affinity W_s^c on I_s with small radius r, using image edges at scale s.

Step 2

Compute W, C from $\left(W_s^c, C_{s,s+1} \right)$ as in (12), (14)

Step 3

Compute Q using (20), compute $\overline{V}$, the first K eigenvectors of $\overline{W} = Q\, D^{-1/2}\, W\, D^{-1/2}\, Q$.

Step 4

Compute $V = D^{-\frac{1}{2}}\, \overline{V}$ and discretize to compute the closest possible to a binary vector X.

6. Modified Spectral Segmentation Algorithm

The main contribution of this report is a modification of the spectral segmentation with multiscale graph decomposition method to improve its performance when applied to specifically to brain MR images. The modified algorithm consists of a series of operations. The Corpus Callosum, which is located at the center of the brain, is considered as white matter tissue. Although one can visually recognize the outline of the Corpus Callosum in Fig. 1, portions of its boundary are indistinct, which can make it difficult to develop an automated segmentation algorithm based on edge information alone. A problem with applying segmentation is that, quite often, variation within the Corpus Callosum can be comparable or exceed the difference between the Corpus Callosum and the surrounding tissues. In addition, parts of the boundary between the Corpus Callosum and surrounding tissue are indistinct and thus very difficult to define due to similar gray levels, particularly on the top portion of the Corpus Callosum and between the Corpus Callosum and the fornix which is not part of the Corpus Callosum.

6.1. Fundamental Concepts

Fig. 5 shows pixel values histogram of the Corpus Callosum derived from 30 MR images. The number of gray levels of all the images used in this report is 256. It can be seen that the Corpus Callosum has relatively high intensity values. In addition, the distribution of intensity values of the Corpus Callosum follows the normal distribution. All global regions and neighborhood regions follow Normal distribution. This is because for most medical images, the noise can be assumed to follow Normal distribution.
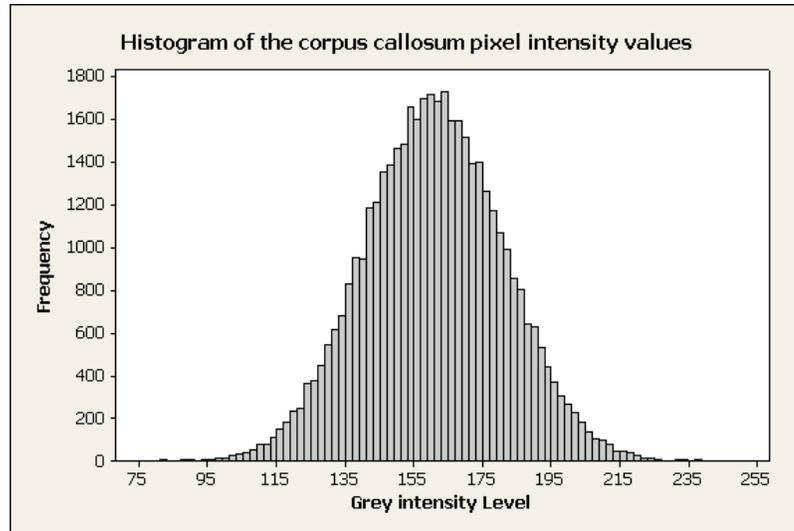
Figure 5: *Histogram of the pixel grayscale values of the Corpus Callosum.*

Fig. 6 shows that the Corpus Callosum pixel values follow the normal distribution with mean $\overline{X}$ = 160.2 and standard deviation S = 20.25.
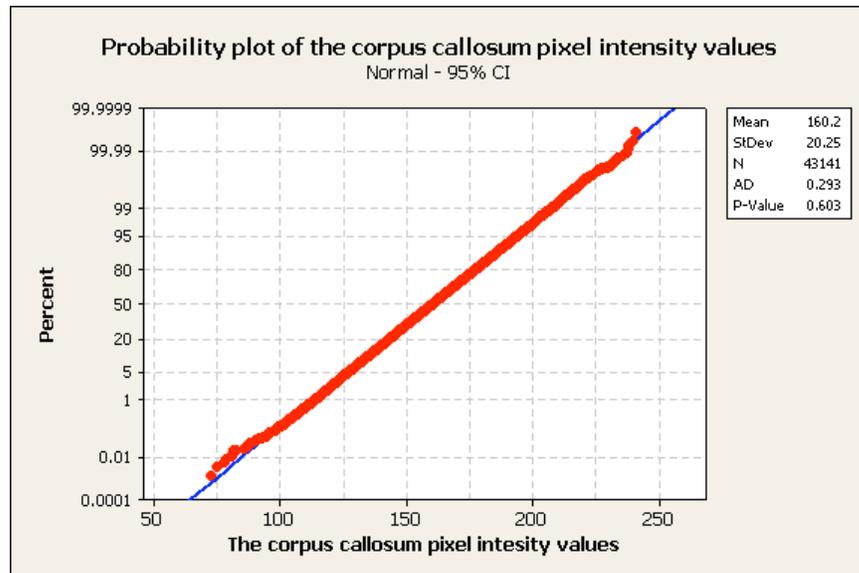
Figure 6: *Probability plot of the Corpus Callosum pixel values.*

Fig. 7 shows that with threshold interval ($\overline{X} \pm S$), we can barely recognize the Corpus Callosum. With threshold interval ($\overline{X} \pm 2S$), a relatively distinct callosal shape is evident with few other non-adjacent structures. With threshold interval ($\overline{X} \pm 3S$), the Corpus Callosum is clearly defined, although more other non-adjacent structure are visible. With threshold interval wider than ($\overline{X} \pm 3S$), the Corpus Callosum starts to be connected to surrounding tissues. Although the threshold values may be different depending on individual images, this property of high intensity values of the Corpus Callosum can be exploited to yield a segmentation algorithm that is efficient across images. Therefore, we choose the $\overline{X} \pm 3S$ interval for applying to MR images as a threshold interval to extract the Corpus Callosum and the objects with the same intensity values.

(a)

(b)

(c)

(d)
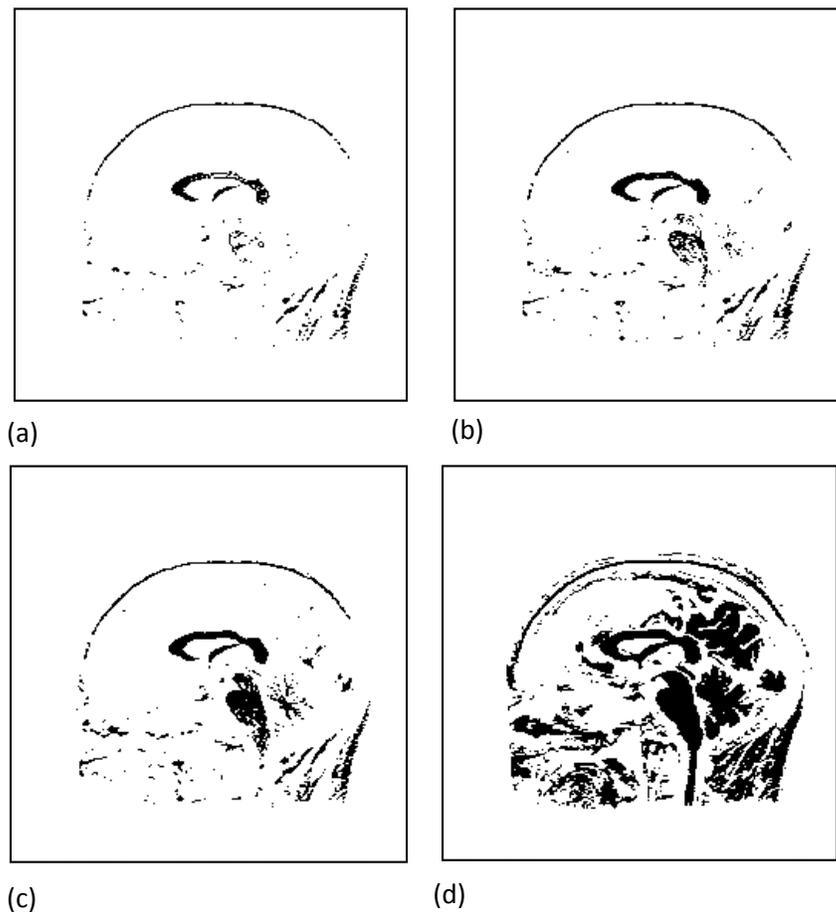
Figure 7: *Thresholding with various threshold intervals. (a) the image obtained by applying a $\overline{X} \pm S$ threshold interval, (b) the image obtained by applying a $\overline{X} \pm 2S$ threshold interval, (c) the image obtained by applying a $\overline{X} \pm 3S$ threshold interval, . (d) the image obtained by applying a wider threshold interval than $\overline{X} \pm 3S$ interval.*

6.2. Algorithm

Step 1

Given $p \times q$ image I, Apply $\overline{X} \pm 3S$ threshold interval for image I to exclude the pixel values which are outside this interval.

Step 2
Apply Spectral Segmentation with Multiscale Graph Decomposition algorithm described in section 5 to the output image from step 1.

7. Experiments and Results

We applied the proposed procedure to find the Corpus Callosum in medsagittal brain MR images (256 by 256, 256 gray levels). The Normalized Cuts algorithm generates poor segmentation with almost all 76 MR images. The proposed algorithm was tested on 76 subjects and we obtained generally satisfactory results. In some cases, the fornix is connected to the Corpus Callosum. The fornix appears as a tail-like protrusion descending from the mid of bottom boundary of the Corpus Callosum. Quite often, it has almost the same gray level as the Corpus Callosum and is very difficult to separate based solely on gray levels.
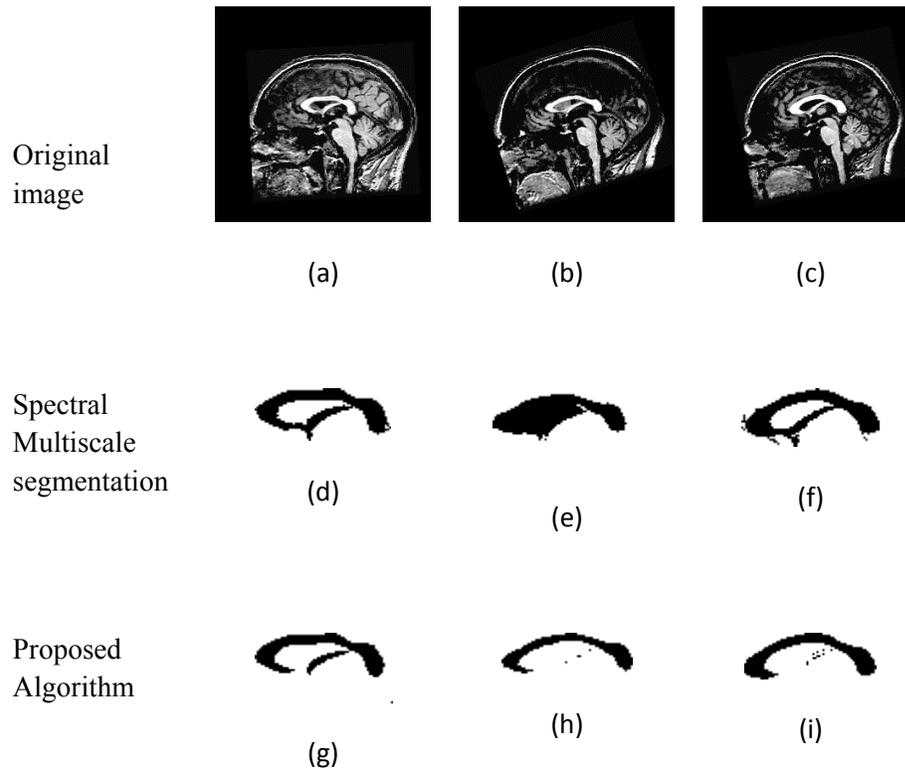
Figure 8: The results obtained by the spectral multiscale segmentation algorithm and the proposed algorithm.
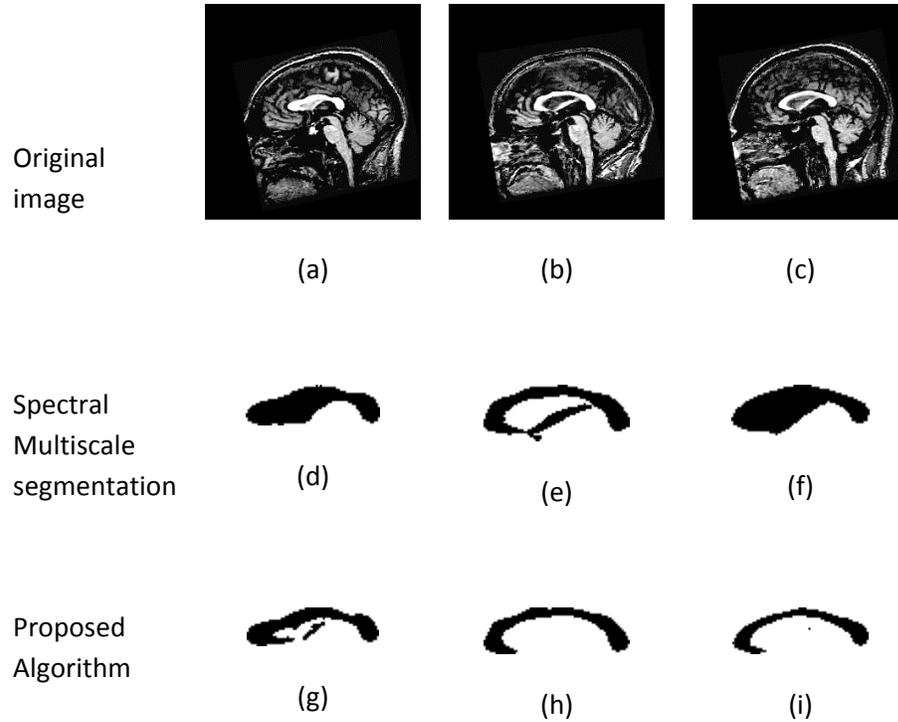
Figure 9: The results obtained by the spectral multiscale segmentation algorithm and the proposed algorithm.

Figs. 8-10 show some of the results with the Corpus Callosum found by the spectral segmentation with multiscale graph decomposition algorithm and the proposed algorithm highlighted. As can be seen, the proposed algorithm gives better results than spectral multiscale algorithm. Out of 76 images, the proposed algorithm was able to find the Corpus Callosum reasonably accurately as in (Figs. 10h, 10i, 11g, and 11h). Out of 76 MR, 38 images where the Corpus Callosum was reasonably accurately segmented, images also included the fornix as in (Figs. 9g, 9h, 9i, 10g, 11i).

Original image			
	(a)	(b)	(c)
Spectral Multiscale segmentation			
	(d)	(e)	(f)
Proposed Algorithm			
	(g)	(h)	(i)
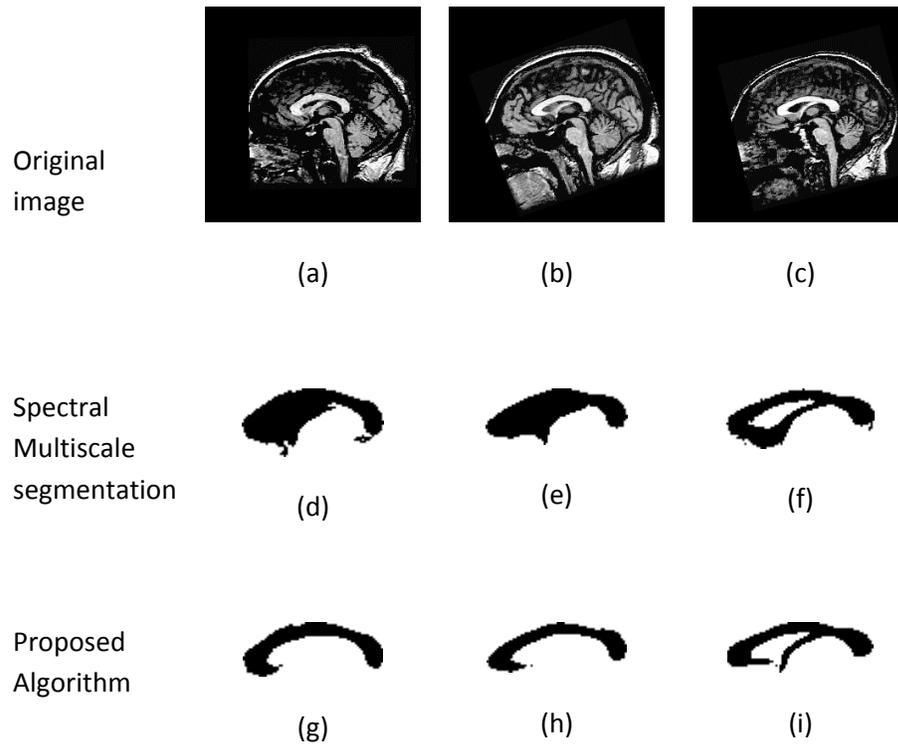
Figure 10: The results obtained by the spectral multiscale segmentation algorithm and the proposed algorithm.

8. Conclusion

We propose an automatic algorithm that segments the Corpus Callosum from midsagittal brain MR images. The algorithm utilizes thresholding and spectral segmentation with multiscale graph decomposition. The boundary between the Corpus Callosum and surrounding tissues can be difficult to detect and, in some cases, artifacts are included in the segmented Corpus Callosum. To remove such artifacts, a thresholding technique is proposed. Experiments showed that the technique can be successfully applied to a wide range of MR images. Once the Corpus Callosum is segmented from surrounding tissues, it can be used as a reliable landmark to find other brain structures, help segment the brain from the surrounding tissues, and also can be used in classification of images across individuals according to its size and shape within image mining framework. Results obtained so far look promising but we need to improve several aspects in our research effort. We need to make post processing procedure applied to some resulted CC segments to remove the fornix which is not part of the Corpus Callosum. We also try to find another ultimate segmentation algorithm which extracts efficiently the Corpus Callosum.

References

[1] L. S. Allen, M. F. Richey, Y. M. Chain and R. A. Gorski (1991). Sex differences in the Corpus Callosum of the living human being. Journal of Neuroscience, Vol. 11, pp. 933-942.

[2] C. Davatzikos, M. Vaillant, S. M. Resnick, J. L. Prince, S. Letovsky, and R. N. Bryan (1996). A computerized approach for morphological analysis of the Corpus Callosum. Journal of Computer Assisted Tomography, Vol. 20, pp: 88-97.

[3] D. Salat, A. Ward, J. A. Kaye and J. S. Janowsky (1997). Sex differences in the Corpus Callosum with aging. Neurobiology of Aging, Vol. 18, No. 2, pp. 191-197.

[4] S. Weis, M. Kimbacher, E. Wenger, and A. Neuhold (1993). Morphometric analysis of the Corpus Callosum using MRI: Correlation of measurements with aging in healthy individuals. American Journal of Neuroradiology, Vol. 14, pp. 637-645.

[5] H. Hampel, S. J. Teipel, G. E. Alexander, B. Horwitz, D. Teichberg, M. B. Schapiro, and S. I. Rapoport (1998). Corpus Callosum atrophy is a possible indicator of region- and cell type-specific neuronal degeneration in Alzheimer disease. Archives of Neurology, Vol. 55, pp: 193-198.

[6] I. K. Lyoo, A. Satlin, C. K. Lee and P. F. Renshaw (1997). Regional atrophy of the Corpus Callosum in subjects with Alzheimer's disease and multi-infarct dementia. Psychiatry Research, Vol. 74, No. 2, pp. 63-72.

[7] P. E. Cowell, A. Kertesz, and V. H. Denenberg (1993). Multiple dimensions of handedness and the human Corpus Callosum. Neurology, Vol. 43, pp. 2353-2357.

[8] R. Duara, A. Kushch, K. Gross-Glenn, W. Barker, B. Jallad, S. Pascal, D. Loewenstein, J. Sheldon, M. Rabin, B. Levin and H. Lubs (1991). Neuroanatomic differences between dyslexic and normal readers on magnetic resonance imaging scans. Archives of Neurology, Vol. 48, No. 4, pp. 410-416.

[9] J. Duncan, Y.Wang, A. A. Amini, R. Greene, L. Kier, J. Gore, J. Holahan, S. Shaywitz, J. Fletcher, R. Bronen, and B. Shaywitz (1996). An MRI-based study of the Corpus Callosum in dyslexic and normal children. Neurology.

[10] G. W. Hynd, J. Hall, E. S. Novey, D. Eliopulos, K. Black, J. J. Gonzalez, J. E. Edmonds, C. Riccio, and M. Cohen (1995). Dyslexia and Corpus Callosum morphology. Archives of Neurology, Vol. 52, pp. 32-38.

[11] J. M. Rumsey, M. F. Casanova, G. B. Mannheim, N. Patronas, N. DeVaughn, S. D. Hamburger and T. Aquino (1996). Corpus Callosum morphology, as measured withMRI, in dyslexic men. Biological Psychiatry, Vol. 39, No. 9, pp. 769-775.

[12] J. M. Rumsey, M. F. Casanova, G. B. Mannheim, N. Patronas, N. DeVaughn, S. D. Hamburger and T. Aquino (1996). Corpus Callosum morphology, as measured withMRI, in dyslexic men. Biological Psychiatry, Vol. 39, No. 9, pp. 769-775.

[13] Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

[14] Harris Drucker, Chris J.C. Burges, Linda Kaufman, Alex Smola and Vladimir Vapnik (1997). "Support Vector Regression Machines".

[15] Yanbo J. Wang, Qin Xin, Frans Coenen: A Novel Rule Weighting Approach in Classification Association Rule Mining. ICDM Workshops 2007: 271-276

[16] J. Bezdek, L. Hall, and L. Clarke, "Review of MR image segmentation techniques using pattern recognition", Med. Phys.vol. 20, pp. 1033–48, 1993.

[17] R. Schalkoff, Pattern Recognition: Statistical, Structural and Neural Approaches. New York: Wiley & Sons. 364 pp., 1992.

[18] G. Coleman, H. Andrews, "Image segmentation by clustering", Proc. IEEE vol. 5, pp. 773–85, 1979.

[19] T. Lei, W. Sewchand, "Statistical approach to X-ray CT imaging and its applications in image analysis. II. A new stochastic model-based image segmentation technique for X-ray CT image", IEEE Trans. Med. Imaging vol. 11, no.1, pp. 62–69, 1992.

[20] Z. Liang, J. MacFall, and D Harrington, "Parameter estimation and tissue segmentation from multispectral MR images", IEEE Trans. Med. Imaging vol 13, pp. 441–49, 1994.

[21] J. Rajapakse, J. Giedd, and J. Rapoport, "Statistical approach to segmentation of single-channel cerebralMRimages", IEEE Trans. Med. Imaging vol. 16, pp. 176–86, 1997.

[22] T. Pappas, "An adaptive clustering algorithm for image segmentation", IEEE Trans. Signal Process, vol. 40, pp. 901–14, 1992.

[23] K. Held, E. Kops, B Krause, W. Wells, R Kikinis, et al., "Markov random field segmentation of brain MR images", IEEE Trans. Med. Imaging vol. 16, no. 6, pp. 878–86, 1997.

[24] O. R. Zaiane, J. Han, Z.-N. Li, and J. Hou, " Mining multimedia data", In CASCON'98: Meeting of Minds, Toronto, 1998.

[25] W. Hsu, M.L. Lee, and K.G. Goh, " Image Mining in IRIS: Integrated Retinal Information System", ACM SIGMOD, 2000.

[26] U. M. Fayyad, S. G. Djorgovski, and N. Weir, " Automating the analysis and cataloging of sky surveys", Advances in Knowledge Discovery and Data Mining, pp. 471–493, 1996.

[27] A. Kitamoto, " Data Mining for Typhoon Image Collection", In Second International Workshop on Multimedia Data Mining (MDM/KDD'2001), 2001.

[28] W. Hsu, M. L. Lee, J. Zhang, "Mining: Trends and Developments,"Journal of Intelligent Information Systems, vol. 19, no. 1, pp. 7-23, 2002.

[29] C. Ordonez, and E. Omiecinski , "Discovering Association Rules Based on Image Content", In IEEE Advances in Digital Libraries Conference, 1999.

[30] O. R. Zaiane, J. Han, Z. N. Li, J. Y. Chiang, and S. Chee, "MultiMediaMiner: A system prototype for multimedia data mining", In Proc. ACM-SIGMOD, Seattle, 1998.

[31] J. Zhang, W. Hsu, and M. L. Lee, "An Information-Driven Framework for Image Mining", In 12th Int. Conference on Database and Expert Systems Applications, 2001.

[32] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation", Annual Review of Biomedical Engineering, vol. 2, pp. 315–337, 2000.

[33] A. Simmons, P. Tofts, G. Barker, and S. Arridge , "Sources of intensity nonuniformity in spin echo images at 1.5T", Magnetic Resonance in Medicine, Vol. 32, no. 1, pp. 121–28, 1994.

[34] J. Sled, G. Pike, "Standing-wave and RF penetration artifacts caused by elliptic geometry: an electrodynamic analysis of MRI", IEEE Trans. Med. Imaging vol. 17, pp. 653–62, 1998.

[35] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2000.

[36] Z. Wu and R. Leahy, "An Optimal Graph Theoretic Approach to Data Clustering: Theory and Its Application to Image Segmentation", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 15, no. 11, pp. 1,101 - 1,113, 1993.

[37] T. Cour , F. Benezit, and J. Shi, "Spectral Segmentation with Multiscale Graph Decomposition". In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cvpr'05) - Volume 2 - Volume 02 (June 20 - 26, 2005) . CVPR. IEEE Computer Society, Washington, DC, 1124-1131, 2005