

©2008 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE

# Classification of Disease Subgroup and Correlation With Disease Severity Using Magnetic Resonance Imaging Whole-Brain Histograms: Application to Magnetization Transfer Ratios and Multiple Sclerosis

J. Dehmeshki\*, G. J. Barker, and P. S. Tofts

**Abstract**—This paper presents a new approach to characterize subtle diffuse changes in multiple sclerosis (MS) using histograms derived from magnetization transfer ratio (MTR) images. Two major parts dominate our histogram analysis; 1) Classification of MTR histograms into control and MS subgroups; 2) Correlation with current disability, as measured by the EDSS scale (a measure of disease severity). Two data reduction schemes are used to reduce the complexity of the analysis: linear discriminant analysis (LDA) and principal component analysis (PCA). LDA is better for the classification of MTR histograms as it takes into account the between-class variation. By using LDA, the space of MTR histograms is transformed to the optimal discriminant space for a nearest mean classifier. In contrast, PCA is useful for correlation with current disability as it takes into account the variation within each subgroup in its process. A multiple regression analysis is used to evaluate the multiple correlation of those principal components with the degree of disability in MS. This is the first application of such classification and correlation techniques to magnetic resonance imaging histogram data. Our MTR histogram analysis approach give improved classification success and improved correlation compared with methods that use traditional histogram features such as peak height and peak location.

**Index Terms**—Histogram analysis, linear discriminant analysis, multiple sclerosis, principal component analysis.

## I. INTRODUCTION

MAGNETIC resonance imaging (MRI) techniques are a valuable way to depict the pathology of multiple sclerosis (MS) *in vivo*. By characterising the extent and nature of pathological change, they offer opportunities to obtain new insights into the underlying pathogenic mechanisms and a sensitive, objective tool with which to monitor treatment effects.

Amongst an increasing number of methods applied there has been a major interest, in recent years, in the application of magnetization transfer (MT) imaging. This approach adds an additional radiofrequency pulse to a traditional MR sequence;

Manuscript received March 14, 2001; revised February 14, 2002. This work was supported by the Multiple Sclerosis Society of Great Britain and Northern Ireland. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was T. Taxt. *Asterisk indicates corresponding author.*

\*J. Dehmeshki is with the Institute of Neurology, University College London, Queen Square, WC1N 3BG, London, U.K (e-mail: j.dehmeshki@ion.ac.uk). He is currently with Medicsight PLC, 46 Berkeley Square, Mayfair, London W1J 5AT (e-mail: jamshid.dehmeshki@medicsight.com).

G. J. Barker and P. S. Tofts are with the Institute of Neurology, University College London, WC1N 3BG, London, U.K.

Publisher Item Identifier S 0278-0062(02)04694-3.

this causes a reduction in the magnetization of the pool of water protons bound to macromolecules. The effect is to reduce the nuclear magnetic resonance (NMR) signal arising from the mobile proton pool, by an amount determined by the size of the bound pool. Measurement of the magnetization transfer ratio (MTR), thus, provides a measure of the structural integrity of tissue. Major reductions in MTR are likely to indicate destructive changes including demyelination and axonal loss, the pathological substrates of symptoms and disability in MS. Less marked reduction in MTR might also be expected with other pathological changes known to occur in MS including oedema, inflammation and gliosis. MTR, thus, provides a reproducible, quantitative measure of brain tissue structures and their modification by pathological change, which in MS occurs not only in visible lesions but also in the normal appearing tissues. There have been numerous recent studies which have used histogram analysis to study the global MTR characteristics of brain tissue [1]–[8].

Various descriptive measures have previously been used to describe the MTR histogram and to measure change; these include histogram peak height, peak location (i.e., mode), average MTR value (i.e., mean), 25th percentile, 50th percentile, and 75th percentile values [1]–[6]. These measures have been used for two purposes: to separate controls from MS subgroups (or to separate subgroups) using simple students t-tests, and to correlate with [expanded disability status scale (EDSS); a measure of disability in MS] [9]. The success of any new disease descriptor at these two tasks is influential in the process of deciding whether the descriptor is likely to be useful for characterising disease progress and its response to treatment. However the features listed above cannot be optimum for these two distinct tasks. The former task (separation of groups) requires features that minimize intra-group variation, while maximizing inter-group variation. The latter task (correlation) requires features that maximally represent variation within the patient dataset (whether a subgroup or all the patients). For example if two subgroups have the same mean EDSS, one cannot expect a feature such as peak height to simultaneously differentiate between the two subgroups and to correlate with EDSS. In addition, such measures are essentially local descriptors of histogram characteristics (i.e., only one part of the histogram is used); because they are local and are chosen arbitrarily (by trial and error), they are unlikely to be optimum as much potential information is ignored. We have recently developed an alternative

TABLE I  
CLINICAL CHARACTERISTICS OF THE PATIENTS AMONG THE DIFFERENT MS SUBGROUPS

	Number of patients	Mean Age	Mean Duration of disease	Mean EDSS
CO	39	35.18 ± 10.64	--	--
BE	11	45.00 ± 5.30	15.82 ± 6.55	2.41 ± 0.76
SP	16	41.94 ± 7.98	16.94 ± 10.91	6.97 ± 0.95
RR	10	35.40 ± 6.27	8.60 ± 6.00	3.80 ± 1.73
PP	46	45.43 ± 9.26	8.06 ± 4.71	5.08 ± 1.23
MS	83	43.50 ± 8.80	10.86 ± 7.65	4.93 ± 1.80

CO = control; BE = benign; SP = secondary progressive; RR = relapsing remitting; PP = primary progressive. Values are given ± standard deviations (SDs).

way of analyzing MTR histograms [7], [8] that has two major advantages over the existing method. First, it takes into account the entire shape of the histograms, and not only just a few arbitrary markers (whatever this is for correlation or classification). Second, instead of performing t-tests to compare groups of patients, patients are classified based on their individual MTR histograms using a pattern recognition scheme. Using this approach, MS patients can be classified correctly into clinical subgroups of MS.

We use linear discriminant analysis (LDA) and principal component analysis (PCA) as data reduction schemes to reduce the complexity of the histograms. We show that the former is more robust for classifying MTR histograms into different subgroups. The latter is useful for correlation analysis as it takes into account the variation within each subgroup. By using these two data reduction techniques, we provide a new approach which uses the entire shape of the histogram for classification and correlation. Classification is more appropriate than a simple t-test, since it works on individual patient data, and a measure of success is produced. We believe this is the first application of such classification and correlation techniques to MRI histogram data.

Our method involves a number of steps which are described in more detail in the remainder of this paper. First, a morphological technique is described to reduce the partial volume effect of tissue adjacent to cerebro-spinal fluid (CSF), in order to minimize the effects of atrophy. The histograms of MTR images are then normalized to the residual brain tissue volume within the scan slices. Using LDA, the space of the MTR histograms is transformed to the optimal discriminant space. In this space, a nearest-mean classifier is used to recognize pairwise the differences between normal controls and the four different subgroups of MS disease. Finally, a multiple regression analysis is used to evaluate the multiple correlation of principal components (PCs) of MS subgroup patients with the degree of disability in MS.

## II. MTR IMAGES

### A. Subjects

The analysis was applied to MTR image data for 83 patients with MS [11 with benign (two male, nine female), ten with relapsing-remitting (four male, six female), 16 with secondary progressive (three male, 13 female), and 46 with primary

progressive disease (30 male, 16 female)] and 39 healthy controls (19 male, 20 female).

Patient and control demographics are provided in Table I.

### B. MRI Acquisition

MRI was carried out using a 1.5-tesla Signa Echo speed Horizon system (General Electric, Milwaukee, WI). A dual spin echo sequence (28 contiguous 5-mm axial slices, TE 30/80 ms, TR 1720 ms, 0.75 NEX, 256 × 256 matrix, field of view 24 × 24 cm) was performed with and without presaturation pulses (total acquisition time 20 min). The presaturation pulse was a Hamming-apodised three lobe sinc pulse with a duration of 16 ms and a peak amplitude of 23.2 μT giving a nominal bandwidth of 250 Hz, applied 1 kHz from the water resonance. Scans with and without presaturation were interleaved for each TR period providing precise co-registration.

MTR was calculated for each pixel by the formula  $MTR = ([M_0 - M_s]/M_0) \times 100$  percent units (pu) where  $M_s$  and  $M_0$  represent signal intensities with and without presaturation respectively. MTR is measured in units of pu. Typical values are 38 pu for white matter, 33 pu for grey matter and 0 pu for CSF (which has no bound protons).

## III. PREPROCESSING

### A. Partial Volume Reduction

Partial volume problems arise because some voxels (size about 1×1×5 mm) contain more than one tissue. We are interested particularly in those that contain a fraction of CSF. *Partial volume voxels* have signal intensity depending on the relative proportions of the voxel occupied by different tissues. The distribution of MTR values shown in histograms, therefore, contains an unknown error component due to partial volume effects. A two-stage partial volume reduction technique (described below) was used to remove some partial volume voxels from around the sulci and ventricles, which contain some CSF, prior to PC and LD analysis. This gave about a 5% reduction in total number of voxels which could be used in the rest of the analysis.

Two stages are sequentially used to reduce the partial volume effects at the brain /CSF boundaries from the MTR images. These are thresholding and mathematical morphology.

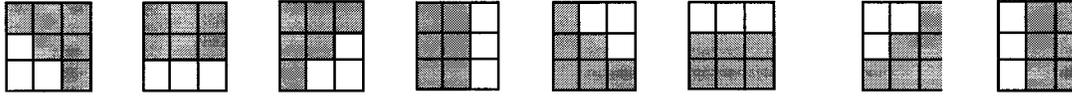


Fig. 1. Morphological structures that match the edges of the brain slices.

1) *Thresholding*: Because partial voxels containing CSF have “partial” intensities dependent on the amount of CSF contained in their pixels, their intensities are likely to be lower than other tissue (since CSF is 0 pu). Thresholding these images by setting all intensities less than a specified threshold (in our case, 10 pu) to zero would remove partial volume voxels with those intensities. MS lesions also have lower MTR values (typically 20 pu in this image dataset) and, therefore, the threshold selected was such that voxels in visible lesions were not removed by this process. Other workers [1]–[5] have used thresholds in a similar manner; we chose this after experimentation, selecting an aggressive value to reduce partial volume effects with CSF.

2) *Mathematical Morphology*: In order not to remove internal voxels, morphological operations [10] were applied on the thresholded images. The application of these morphological operations can be expressed as

$$M(I) = IoL \quad (1)$$

for MTR image  $I$ , morphological operation  $o$ , morphed image  $M$  and structure  $L$ . Of the four operations; erosion, dilation, opening and closing; opening is the most suitable as it removes additional outlying pixels/voxels along the slice boundaries and reinforces the boundary edges, i.e., any removed pixels/voxels which should have been part of the boundary are replaced. The structures used for the morphological operations are shown in Fig. 1 and match the eight possible edges of the brain slices boundaries. The morphological application of these structures to the images in sequence can be expressed as

$$M(I) = (\dots (((IoL^1) oL^2) oL^3) \dots oL^8). \quad (2)$$

The effects of this operation are twofold: internal “missing voxels” (i.e., holes in the mask) caused by thresholding are filled in; and a one-pixel wide strip of partial volume voxels associated with boundaries between CSF and brain tissues is removed.

### B. Histogram Normalization and Noise Reduction

A MTR histogram (i.e., MTR voxel intensity frequency) of whole brain was calculated for each subject. Each MTR histogram was normalized to the residual brain tissue volume by dividing the number of counts in each sampling bin by the total number of voxels. Thus, the total area under the histogram is fixed at unity and the normalized histogram is, therefore, a frequency distribution.

In order to evaluate the effect of noise on our histogram analysis, a smoothing filter (moving average) with width of 0.5 pu was applied to the histograms for noise reduction.

## IV. DATA REDUCTION

There is a large number of bins (depending on the resolution to which the MTR values are stored, in our case 500 bins each of width 0.1 pu) in each MTR histogram compared with the relatively small number of training data sets (i.e., the number of patients used to train the classifier). Any data analysis can, therefore, erroneously focus on meaningless characteristics of individual training cases instead of seeing the broad picture that is essential for general data analysis on the training sets problem.

LDA and PCA are two data reduction schemes designed to reduce the complexity of analysis. The purpose of these procedures is to compress all of the information contained in a 500-bin histogram into a few values. We show that PCA is useful for correlation analysis (see Section VI) as it take into account the variation within each subgroup in its process. In contrast, we expect LDA to be better for classifying MTR histograms into different subgroups. We have measured the success of a nearest mean-group classification applied on LDA spaces.

## V. CLASSIFICATION OF MTR HISTOGRAM

The aim of this section is to show that there are significant differences in MTR histograms between patients with different types of MS disease and healthy controls. There are many available classifiers that could be used for distinguishing between different groups, and for developing rules for classification of a MTR histogram of a subject whose group is unknown. These methods can be divided into two categories; the first category (often called “parametric”) assumes knowledge of the underlying class-conditional probability density function (in our case, probability density function of the MTR histograms for a given clinical group). In many applications (including our own) these would have to be estimated from training sets (a set of correctly classified samples). The second category (often called “non-parametric”) develops a set of decision rules that uses the data to estimate the decision boundaries directly without explicit calculation of the probabilistic density functions. This study uses the second “nonparametric” category; the histogram space is transformed to the optimal discriminant space (which maximizes the separation between the classes in the training set). This discriminant space can be divided into as many regions as there are classes. The boundary between them, the decision boundary, can be used to assign an unknown MTR histogram to a clinical group. The decision boundaries may be linear or nonlinear. Here, we use a linear LDA since this is more appropriate as the number of our samples is small. In general, a LDA is discriminant analysis with linear class boundaries using a feature vector of specified dimension. The feature vector extracted from the histogram using a multiple discriminant transformation, which is defined using training data set. A nearest mean classifier, applied on transformed data spaces, is used for classification.

### A. Linear Discriminant Analysis (LDA)

The aim of LDA is to maximize the ratio of the between-group variance to the within-group variance. LDA is used to create a nearest-mean-group classifier. A brief summary of the techniques is given here, but for more detail see [11] and [12]. The aim of LDA is to maximize

$$\frac{|\tilde{W}_b|}{|\tilde{W}_w|} = \frac{|\phi^t W_b \phi|}{|\phi^t W_w \phi|}. \quad (3)$$

In (3),  $W_b$  is the between-class scatter matrix,  $W_w$  is the within-class scatter matrix, and  $\phi$  is the transformation we are searching for in order to form the optimal discriminant space. We can define the following, with  $\underline{h}^{p,i}$  being the histogram of subject  $p$  in subgroup  $i$ , and  $n_i$  being the number of subjects in subgroup  $i$

$$\bar{\underline{h}}^i = \frac{1}{n_i} \sum_p \underline{h}^{p,i} \quad (4)$$

$$\bar{\underline{h}} = \frac{1}{n} \sum_i n_i \bar{\underline{h}}^i \quad (5)$$

$$W^i = \sum_p (\underline{h}^{p,i} - \bar{\underline{h}}^i) (\underline{h}^{p,i} - \bar{\underline{h}}^i)^t \quad (6)$$

$$W_w = \sum_i W^i \quad (7)$$

$$W_b = \sum_i n_i (\bar{\underline{h}}^i - \bar{\underline{h}}) (\bar{\underline{h}}^i - \bar{\underline{h}})^t. \quad (8)$$

Equation (4) computes subgroup mean of  $n_i$  subjects  $p$  in subgroup  $i$ . Both the within-class scatter  $W_w$  and the between-class scatter  $W_b$  are analogous to their respective covariance matrices.

In looking for  $\phi$ , we can define

$$\underline{y} = \phi^t \underline{h} \quad (9)$$

$$\psi^i \equiv \{ \underline{y}^i | \underline{h}^i \in \text{ith group}, \underline{y}^i = \phi^t \underline{h}^i \}$$

$$\bar{\underline{y}}^i = \frac{1}{n_i} \sum_{\underline{y} \in \psi^i} \underline{y} \quad (10)$$

$$\bar{\underline{y}} = \frac{1}{n} \sum_i n_i \bar{\underline{y}}^i \quad (11)$$

$$\tilde{W}_w = \sum_i \sum_{\underline{y} \in \psi^i} (\underline{y} - \bar{\underline{y}}^i) (\underline{y} - \bar{\underline{y}}^i)^t \quad (12)$$

$$\tilde{W}_b = \sum_i n_i (\bar{\underline{y}}^i - \bar{\underline{y}}) (\bar{\underline{y}}^i - \bar{\underline{y}})^t. \quad (13)$$

It follows from this that

$$\tilde{W}_w = \phi^t W_w \phi \quad (14)$$

$$\tilde{W}_b = \phi^t W_b \phi. \quad (15)$$

Taking the determinant of a scatter matrix is equivalent to finding the product of the eigenvalues, which corresponds to the product of the variance. As may be seen with reference to (3), by maximizing this ratio, we are looking for a transform  $\phi$  that maximizes the between-class variance with respect to the within-class variance. The solution of (3) can be shown to correspond to the generalized eigenvectors of

$$W_b \underline{\phi}_j = \lambda_j W_w \underline{\phi}_j$$

where the vectors  $\underline{\phi}_j$  then form the columns of the matrix  $\phi$ .

In addition, the individual dimensions of the discriminant space created by each eigenvector  $\underline{\phi}_j$  are now ordered. The between-class variances in dimension  $j$  is proportional to the eigenvalue  $\lambda_j$ . Assuming a constant within-class variance, the higher the between class variance of a dimension, the better the discriminant capacity of that dimension.

One additional step can be taken to scale all the within-class variances to uniform size in the discriminant space.

The variance in dimension  $j$  can be computed as  $\underline{\phi}_j^t W_w \underline{\phi}_j$  and each dimension can be scaled by replacing  $\underline{\phi}_j$  with

$$\hat{\underline{\phi}}_j = \frac{\underline{\phi}_j}{\sqrt{\underline{\phi}_j^t W_w \underline{\phi}_j}} \quad (17)$$

giving each new dimension uniform variance. This allows us to use the nearest-mean group classifier, in discriminant space, since this classifier does not account for differing variances between dimensions. The decision as to whether the particular MTR histogram is allocated to one subgroup or another is based on measuring the (Euclidean) distance between its transform scores (created by LDA) and the centroids of all the subgroups in discriminant space.

### B. Identifying Regions of Histogram and Images With Significant Between-Group Variation. (LDA-Based Eigenvectors and Eigenimages)

The largest elements of each eigenvector derived from (16) show which regions of a histogram contribute most to the variation *between group*. These regions of significant variation can also be highlighted in the brain slice using the eigenvectors. This involves mapping the MTR values (intensity values) in the brain slice images through the eigenvectors and displaying the result. This is described by

$$E_j[x, y] = |\underline{v}_j[I[x, y]]| \quad (18)$$

for eigenimage  $E_j$  and image  $I$  (of the individual subject) with points  $(x, y)$  and using eigenvector  $\underline{v}_j$ . Thus,  $I(x, y)$  is the MTR value at a particular voxel in the image (location  $(x, y)$ );  $\underline{v}_j[I(x, y)]$  is the value of the eigenvector  $\underline{v}_j$  at this MTR value (i.e.,  $\underline{v}_j$  indexed by the MTR value). The positive and negative extrema of the eigenvector as discussed earlier will give the significant regions. The absolute value of these give high intensity values for these significant regions and, therefore, eigenimages formed from mapping through the absolute eigenvector highlight significant regions as being bright areas.

## VI. CORRELATION WITH CURRENT DISEASE SEVERITY (EDSS)

This section demonstrates that MTR histograms provide a global measure of tissue structural change that corresponds closely to the degree of disability (EDSS) in MS. The EDSS is used by clinicians as a ‘‘gold standard’’ measure of disease severity in MS, although it has several shortcomings. It has only discrete values (0.5–10 in steps of 0.5), with an inter-observer agreement of about 1.0, it is nonlinear, and is biased toward locomotor disability, largely ignoring other symptoms such as neuropsychiatric ones. Nonetheless, the acceptance by the clinical community of new MRI surrogate markers of

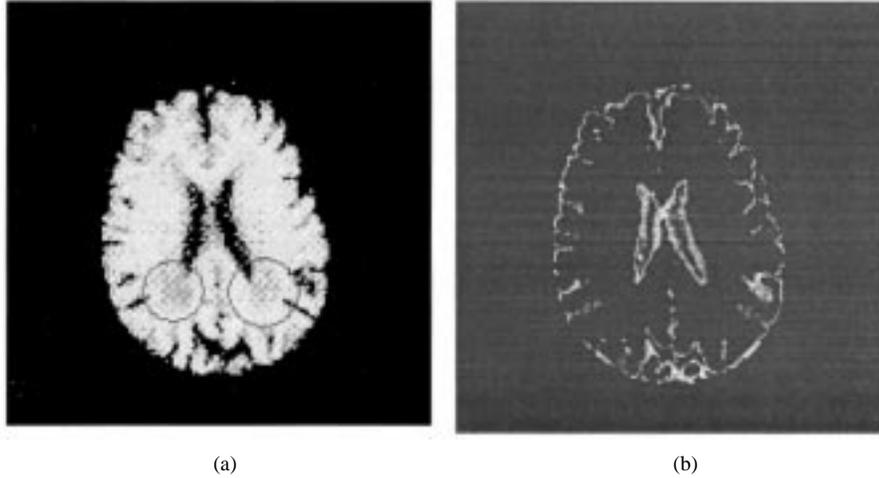


Fig. 2. The removal of partial volume voxels using thresholding and mathematical morphology. (a) Brain slice with highlighted lesions. (b) Extracted partial volume voxels.

disease is largely dependent on being able to show correlation of these markers with EDSS [13]. We have used a multiple linear regression analysis to evaluate the correlation between the first few PCs and EDSS. PC analysis was used to reduce the dimensionality of the MTR histogram, while retaining as much of the variation present in the histogram set as possible. Therefore, the first few PCs were introduced as global features, which represent the most useful information within the MTR histogram. This multiple correlation analysis is appropriate, as the individual PCs were essentially uncorrelated. The F test was used to test for the significance of the multiple correlation coefficient [14].

#### A. Principal Component Analysis (PCA)

Data reduction can be achieved by extracting the PCs [15] of the covariance matrix of MTR histograms, computed by treating the histograms as replicates of a  $N$ -variate observation ( $N$  is the number of bins in each MTR histogram). The PCs are uncorrelated with each other and are ordered in decreasing proportion of variation present in all of the original MTR histograms.

PCs capture the characteristic significant variations for each of the patient groups. These characteristic variations are in the form of eigenvalues and eigenvectors evaluated from the MTR histogram data [15]. The process of evaluating the eigenvalues and eigenvectors is described below. First, covariance matrix of MTR histograms,  $\underline{h}^p = \{h_1^p, \dots, h_N^p\}$  is evaluated

$$w_{k,l} = w_{l,k} = \frac{1}{n} \sum_p (h_k^p - \bar{h}_k) (h_l^p - \bar{h}_l). \quad (19)$$

Where

$$\bar{h}_k = \frac{1}{n} \sum_p h_k^p$$

$$\bar{h}_l = \frac{1}{n} \sum_p h_l^p$$

and  $n$  stands for number of patients  $p$  that comprises the training set,  $k$  and  $l$  are elements in the index  $N$  where  $N$  is the number of bins in histogram.

Then, eigenvalues  $\lambda$  and eigenvectors  $\underline{v} = (v_1, v_2, \dots, v_N)$  of this matrix are evaluated

$$W \underline{v} = \lambda \underline{v}. \quad (20)$$

The  $m$  most significant (largest) eigenvalues and their corresponding eigenvectors are selected to characterize the variation of each patient group. The percentage variation ( $PV(m)$ ) covered by these  $m$  significant components is calculated as follow.

$$PV(m) = \frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^N \lambda_j} \times 100\% \quad (21)$$

where  $N$  is number of bins in MTR histogram, and  $\lambda_j$  are the eigenvalues.

The percentage variation allows us to choose the number of PCs ( $m$ ) to be used for our further MTR histogram analysis. Note that the PCs,  $p$ , of each histogram are linear combination (dot product) of the eigenvectors and the histogram

$$p = \underline{v} \underline{h} = v_1 h_1 + v_2 h_2 + \dots + v_N h_N. \quad (22)$$

#### B. Identifying Regions of Histogram and Images With Significant Within-Group Variation. (PC-Based Shading Histogram and Eigenimages)

The largest elements of each eigenvector show which regions of a histogram contribute most to the variation *within* a group (Note that this is distinct from the *between-group* differences highlighted by LDA). These regions can be highlighted in the brain slice using the PCA eigenvectors in a similar manner to (18).

## VII. RESULTS

#### A. Partial Volume Reduction

Images illustrating the reduction of partial volume effects using thresholding and mathematical morphology can be seen in Fig. 2.

An example of a MT image with circled lesions is shown in Fig. 2(a) along with partial volume voxels extracted [Fig. 2(b)]

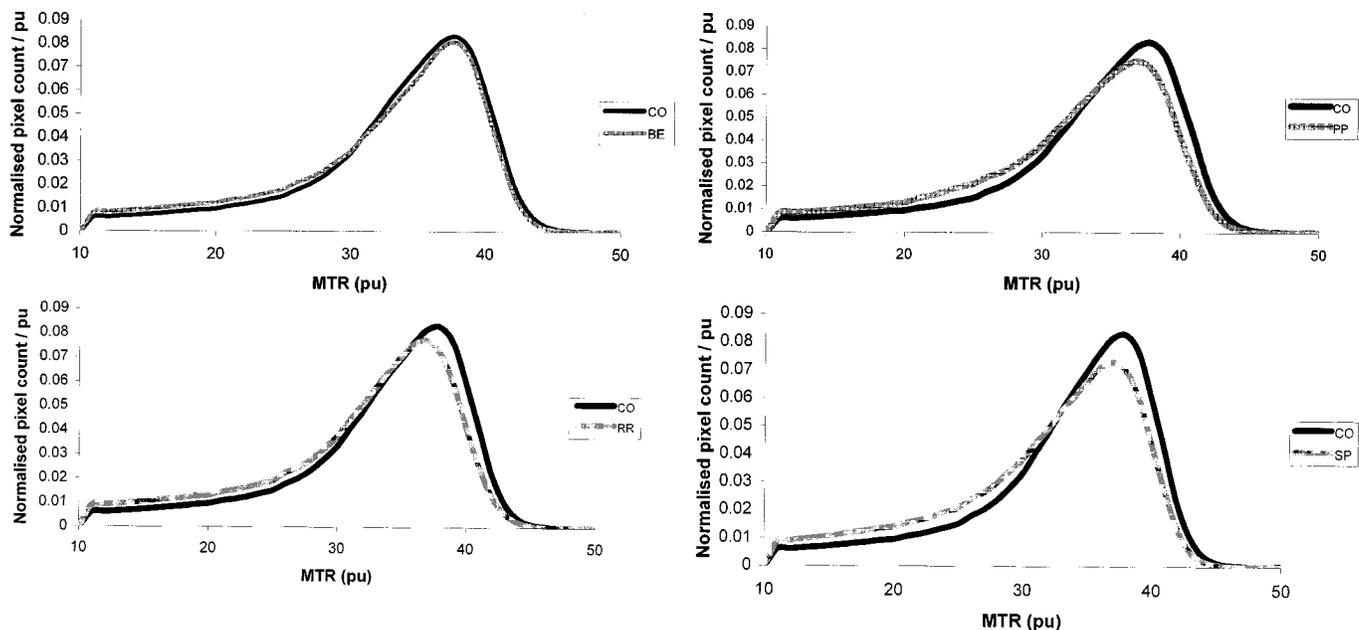


Fig. 3. MTR histograms (after CSF removal and normalization) among the five groups; CO = control; PP = Primary progressive; RR = Relapsing Remitting; Be = Benign; SP = Secondary Progressive.

using the technique described in Section III-A. The final brain slice image retains all of its internal tissue including lesions, but has a “cleaner” boundary as a result of the partial volume removal.

### B. Group Mean Histograms

Fig. 3 displays the mean MTR histograms for the MS subgroups and controls, showing clear differences in peak height and location, indicating that there are differences between control and MS subgroups. This shifting of the peak position may be significantly influenced by pixels that are not in visible lesions but in the normal appearing brain. This shifting of the peaks downward and to the left indicates that the MS subgroups in general have a lower fraction of high density tissue (healthy tissue) or, in other words, a higher fraction of low density tissue (lesion and deteriorated tissue) than the control group does.

### C. Classification Results

Ideally, the classifier should compare a clinically unknown MTR histogram with a pre-existing database of the MTR histograms and classify it to one of several possible groups, with an indication of the degree of certainty of the classification. However, total validation of this system requires large databases of preclassified MTR histograms. It will take some time to construct these databases and consequently obtain a complete validation of our automatic system. In the meantime, because we currently have only a small database, we restrict our classifier to a binary distinguishing between pairs of clinical MTR groups, in order to provide a preliminary test of our system. However a binary or three-way classification may often be sufficient, since, to use the system optimally, the clinical question being asked must guide a potential restriction on the number of possible classes (e.g., there may be prior information that the possibility of a particular subgroup MS has already been excluded, in which case

TABLE II  
SUCCESS RATES FOR BINARY CLASSIFICATION (NO. OF SUCCESSES/TOTAL NO. OF SUBJECTS IN THE TWO GROUPS) USING LDA CLASSIFICATION

Training		Classification result
Set 1	Set 2	Leave-one-out
CO	BE	43/50
CO	SP	50/55
CO	RR	45/49
CO	PP	75/86
CO	MS	100/122
BE	SP	21/27
BE	RR	17/21
BE	PP	41/57
SP	RR	18/26
SP	PP	40/62
RR	PP	37/56

CO = control; BE = benign; SP = secondary progressive; RR = relapsing remitting; PP = primary progressive.

the classifier should not attempt to answer whether the subject has that particular form of MS).

1) *Binary Classification Results:* Table II shows the result of a binary classification of patients into the different groups using LDA classifier applied on the MTR histograms. To avoid the “selection bias” problem, due to the fact that the number of patients is not large enough, the leave-one-out method is used for validation [16], [17]. The number of correctly classified samples and the total number of samples involved in each binary comparison are given in the Table II. Using LDA, the success rate of binary classification was 86%–92%, depending on which MS groups were being compared with control.

2) *Comparison of LDA Scores With Conventional Histogram Parameters:* In order to make a convincing comparison of our LDA score method with conventional histogram parameters,

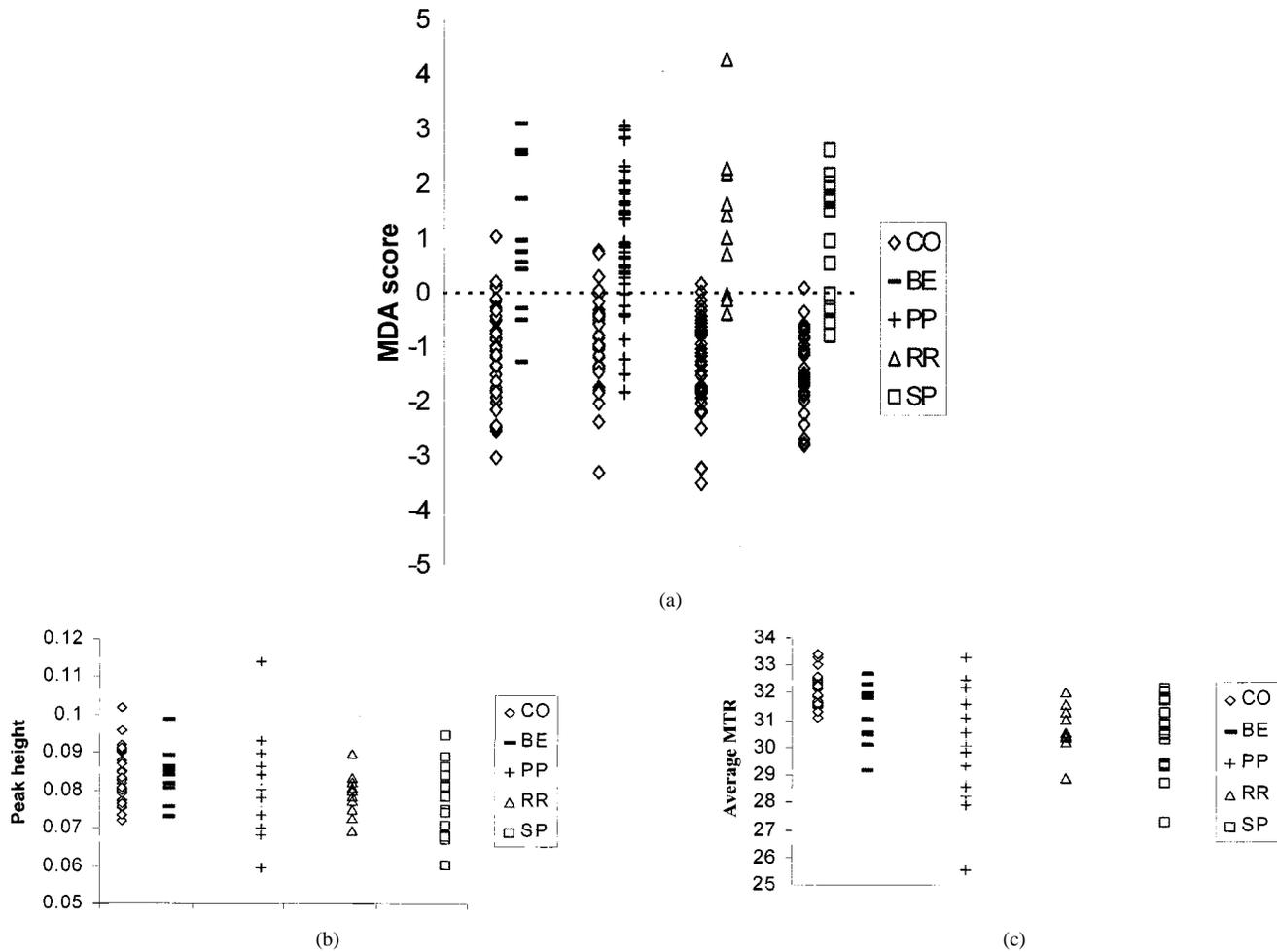


Fig. 4. Sample scatter plots for (a) the distribution of discriminant scores (created by MDA analysis) and two established features [(b) peak height and (c) average MTR] Note that the MDA scores relate to binary comparisons, so that the each control has a score which differs according to what comparison is being made.

TABLE III

PERFORMANCE OF LDA SCORE AND CONVENTIONAL HISTOGRAM PARAMETERS IN SEPARATING MS SUBGROUPS AND CONTROL.  $p$  VALUES FOR STUDENT  $t$ -TEST ARE SHOWN. NOTE THAT THE LDA  $p$ -VALUES ARE ALWAYS LOWER THAN THOSE ACHIEVED BY ANY OTHER PARAMETERS, USUALLY BY A LARGE FACTOR

	Conventional features						LDA
	AVMTR	PH	PL	MTR25%	MTR50%	MTR75%	score
CO-BE	0.02	NS	NS	0.03	NS	NS	<b>0.0003</b>
CO-SP	0.0001	0.013	NS	0.0001	0.0009	0.004	<b>5e-6</b>
CO-RR	0.0001	0.027	0.008	0.002	0.0001	0.007	<b>0.00002</b>
CO-PP	0.0001	NS	0.002	0.0001	0.0001	0.0001	<b>1e-12</b>
BE-SP	NS	0.061	NS	NS	0.06	0.07	<b>0.001</b>
BE-RR	NS	NS	0.04	NS	NS	NS	<b>0.008</b>
BE-PP	NS	NS	0.018	NS	0.03	0.02	<b>0.002</b>
SP-RR	NS	NS	NS	NS	NS	NS	<b>0.002</b>
SP-PP	NS	NS	NS	NS	NS	NS	<b>0.002</b>
RR-PP	NS	NS	NS	NS	NS	NS	<b>0.004</b>

AVMTR = Average brain MTR; PH = peak height, PL = peak location; MTR25%, MTR50%, MTR75%: = MTR AT THE 25TH, 50TH, 75TH PERCENTAGE; CO = control; BE = benign; SP = secondary progressive; RR = relapsing remitting; PP = primary progressive. NS = not significant.

Fig. 4 shows some sample scatter plots for the distribution of discriminant scores (created by LDA analysis in binary classification) and two established features (peak height and average MTR). This figure clearly shows the power of discriminant anal-

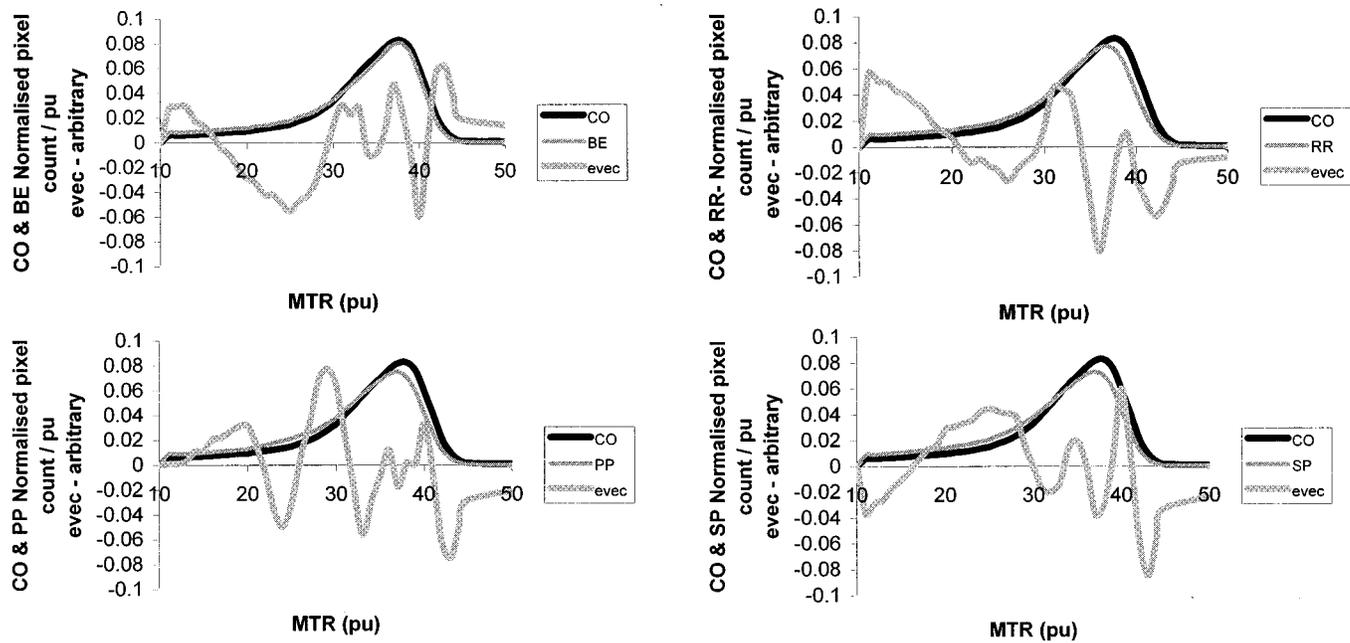


Fig. 5. MDA eigenvector for binary comparisons between each MS subgroup and control. Note that the number of eigenvectors needed to describe the comparison is at most one less than the number of groups being compared, so for these binary comparison only a single eigenvector is generated.

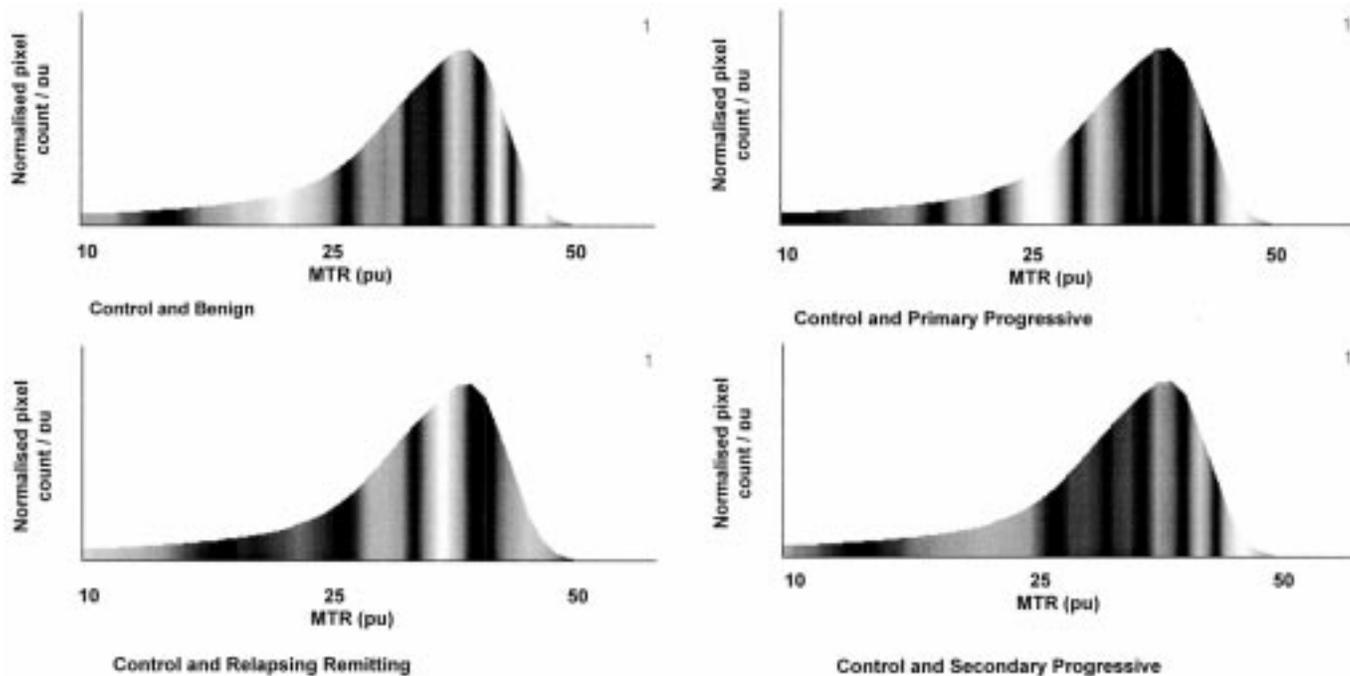


Fig. 6. MDA-Shaded histograms for binary comparisons between each MS subgroup and controls.

ysis in differentiating subgroups using their MTR characteristics. It is apparent also that conventional measures provide a far less clear differentiation. Student t-tests were also carried out between some subgroups to compare our LDA score method with conventional histogram parameters (peak height and average MTR value) (Table III). We investigated whether the LDA score would provide better separation between the groups (i.e., lower p-values). In separating the different clinical subgroups, LDA was always better than conventional MTR histogram parameters.

3) *LDA-Based Eigenvector and Eigenimages*: The eigenvectors created by LDA analysis contain information from the whole histogram, in a way that has been optimized to place most emphasis on those parts that contribute most to separating the groups. Fig. 5 shows the LDA eigenvectors derived from binary comparison between MS group and control group, along with the histogram of a typical patient from that group. The positive and negative extrema of the eigenvectors are the region of that group’s histograms that contribute the most to group separation (i.e., between each MS subgroup and control). Fig. 6

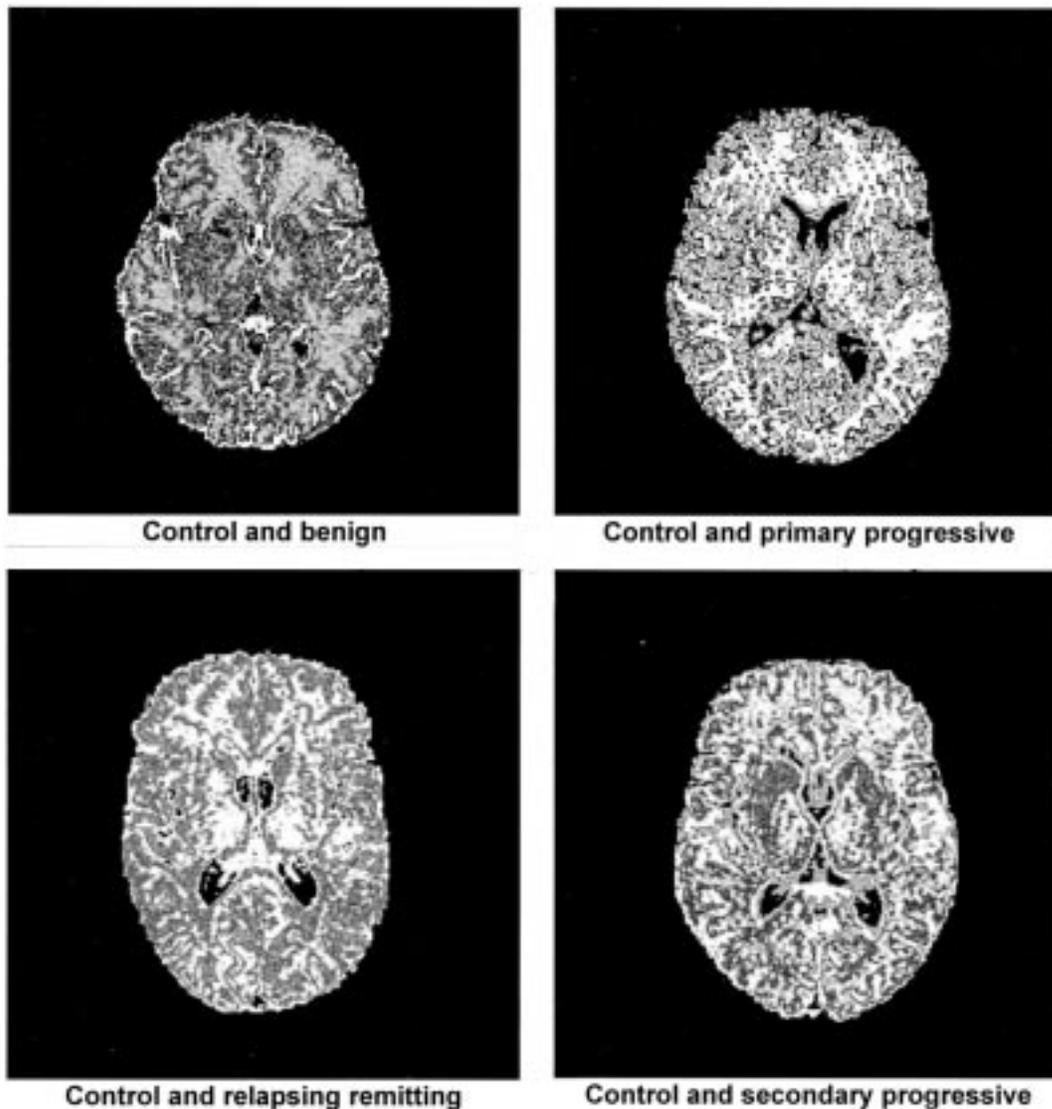


Fig. 7. MDA Eigen images relating to binary comparisons between each MS subgroup and control. Bright areas correspond to tissues whose intensities in the original MTR images fall within the ranges which best discriminate the groups. A typical image from each subgroup is mapped through the eigenvector (Fig. 5) to produce an intensity dependent on the amplitude of the eigenvector at that value of MTR. The gray scale is arbitrary.

shows shaded histograms which qualitatively illustrate these significant regions with dark regions contributing the least to group separation (between each MS subgroup and control) and bright regions contributing the most.

Fig. 7 shows eigenimages [see (18)] which highlight the pixels in the original MTR image whose intensity fall within regions of the histogram corresponding to significant differences between the groups; regions that contribute most to the separation between the two groups are very bright while those with little or no contribution are dark. It is interesting to note the distribution pattern of the bright regions in the primary progressive eigenimages; the variations seen over the whole images may reflect the more diffuse nature of the disease in these patients [18]. Note that the eigenvector has been calculated globally, and shows areas of *potential* abnormality; thus, it is not surprising that some tissue that is highlighted, such as the choroids plexus, could be normal and still caught in the range of MTR values that have most variation.

4) *Effects of Age and Atrophy*: In a previous study [19] we measured the effect of age on the mean MTR of normal white matter. The reduction is extremely small (about 0.23 pu/decade); this is too small to explain the significant reductions we saw for MS. In addition we divided the normal group into the youngest and oldest halves (mean ages 25.78 [ $\pm 5.0$  SD] and 44.10 [ $\pm 5.5$  SD] years respectively). LDA was applied in an attempt to classify the subjects from their histograms. The success rate was 23/39 (i.e., 59%); if the groups overlapped completely we would expect a success rate of 50% from chance. Thus, we conclude that LDA is unlikely to be significantly influenced by age in this study.

#### D. Correlation Result: Multiple Regression Analysis (EDSS) Results

Other workers [1]–[6] have previously reported good correlation of histogram parameters such as the peak height with EDSS.

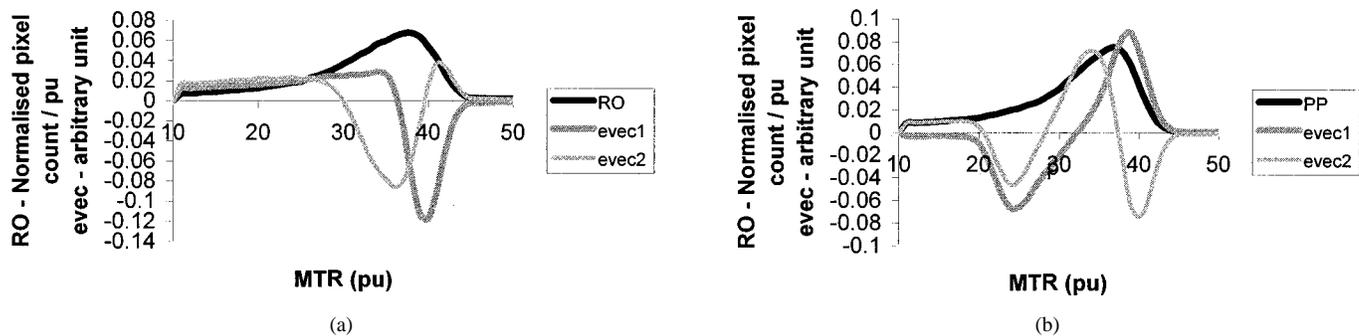


Fig. 8. PCA eigenvector. (a) Relapsing onset (RO) Histogram along with first two PC eigenvector. (b) Primary progressive group histogram along with first two PC eigenvectors. Note that (unlike the MDA example above, which relates to a binary comparison of groups) we are here extracting information from all subjects within a particular group, so multiple eigenvectors can be created (limited only by the number of bins in the histogram). Here, we display the first two (most significant) eigenvectors.

TABLE IV  
RESULTS OF CORRELATION ANALYSIS BETWEEN EACH SUBGROUPS AND EDSS

	PP	RO	MS
AVMTR*	NS	NS	-0.25(p=0.01)
PH*	NS	-0.41(p=0.01)	-0.25(p=0.01)
PL*	NS	NS	NS
MTR25%*	NS	-0.46(p=0.004)	-0.25(p=0.02)
MTR50%*	NS	-0.34(p=0.03)	-0.21(p=0.04)
MTR75%*	NS	-0.35(p=0.03)	-0.25(p=0.02)
PCs (Multi) <sup>+</sup>	0.40 (p<0.025)	0.51 (p<0.005)	0.30 (p<0.01)

AVMTR = average brain MTR; PH = peak height, PL = peak location; MTR25%, MTR50%, MTR75% = MTR AT THE 25TH, 50TH, 75TH PERCENTAGE. PP = primary progressive; RR = relapsing remitting; BE = benign; SP = secondary progressive; RO = RR + BE + Sp. (\*) traditional MTR parameters; the spearman's rank correlation analysis was used; (+) PCs: the f test is used for the significance of the multiple correlation coefficient. NS = not significant.

It is to be expected that any results obtained using previous parameterizations may be less optimal than results obtained using PC analysis [using the first three PCs as these cover 90% of the variability based on (21)], which is more general and not restricted to this subset of parameterizations. However, it would be interesting to compare such measures with those provided by PCA. Therefore, a Spearman's rank correlation analysis was used to evaluate the relationship between each of the traditional MTR parameters (peak height, peak location, average MTR, MTR at the 25th, 50th, and 75th percentile) and EDSS. Table IV shows the correlation between EDSS and traditional MTR histogram parameters for all patients and for the PP subgroups and relapsing onset (RO) (because of small numbers within the BE, RR, and SP groups it was considered appropriate to combine them into the single larger RO group for disability correlation). In the PP group, there was no correlation between individual conventional MTR parameters and EDSS, but using multiple correlation analysis there was a significant correlation with the PCs of the histogram ( $r = 0.40$ ). In the RO group, moderate associations were seen between a number of conventional MTR parameters and EDSS, the strongest being with the 25th percentile MTR ( $r = -0.46$ ). However a still stronger correlation was seen with the PCs ( $r = 0.51$ ). It can be seen from Table IV that multiple PC analysis revealed more robust correlation be-

tween the MTR histogram and disability in the two subgroups (PP, RO).

1) *PCA-Based Eigenvector and Eigenimages*: Fig. 8 shows the first two eigenvectors for each group, along with the histogram of a typical patient from that group. The positive and negative extrema of the eigenvectors are the regions of that group's histograms that contribute the most to its variation. Fig. 8 shows that low regions of the histograms in general did not contribute significantly to the variation in the MTR histogram data within an MS subgroup. Shaded histograms qualitatively illustrate these significant regions with dark regions contributing the least to within group variation and bright regions contributing the most (Fig. 9).

## VIII. CONCLUSION

This study explores and demonstrates the application of the Linear Discriminant transform and PCA to MTR images in MS, and has shown that these techniques may be used to characterize various subtypes of MS. This alternative way of analyzing MTR histograms has two major advantages over the existing methods. First, it takes into account the entire shape of histograms, and not only just a few arbitrary features (e.g., peak location and height). Second, instead of performing t-tests to compare groups of patients, patients are classified based on their individual MTR histograms using LDA. Using this method, MS patients could be classified reasonably accurately into clinical subgroups of MS. MRI measurements are increasingly being used as surrogate markers in drug trials; this work implies that LDA and PCA-derived features will be more sensitive and specific at predicting biological and clinical change than existing features are.

LDA of MTR histograms has been shown to provide effective classification of disease subgroups and controls. Even a perfect classifier might not achieve 100% success, since we do not have a perfect 'Gold Standard' description of the class of each subject; the clinical subgroups may not have been assigned completely correctly, and some of the patients may even be in transition between groups.

PCA has been shown to provide good correlation performance relating MTR histogram features to disability, as quantified using the EDSS. This has been shown by the good correlation between the PCs and EDSS. The EDSS scale is a nonlinear categorical (discontinuous) scale that combines

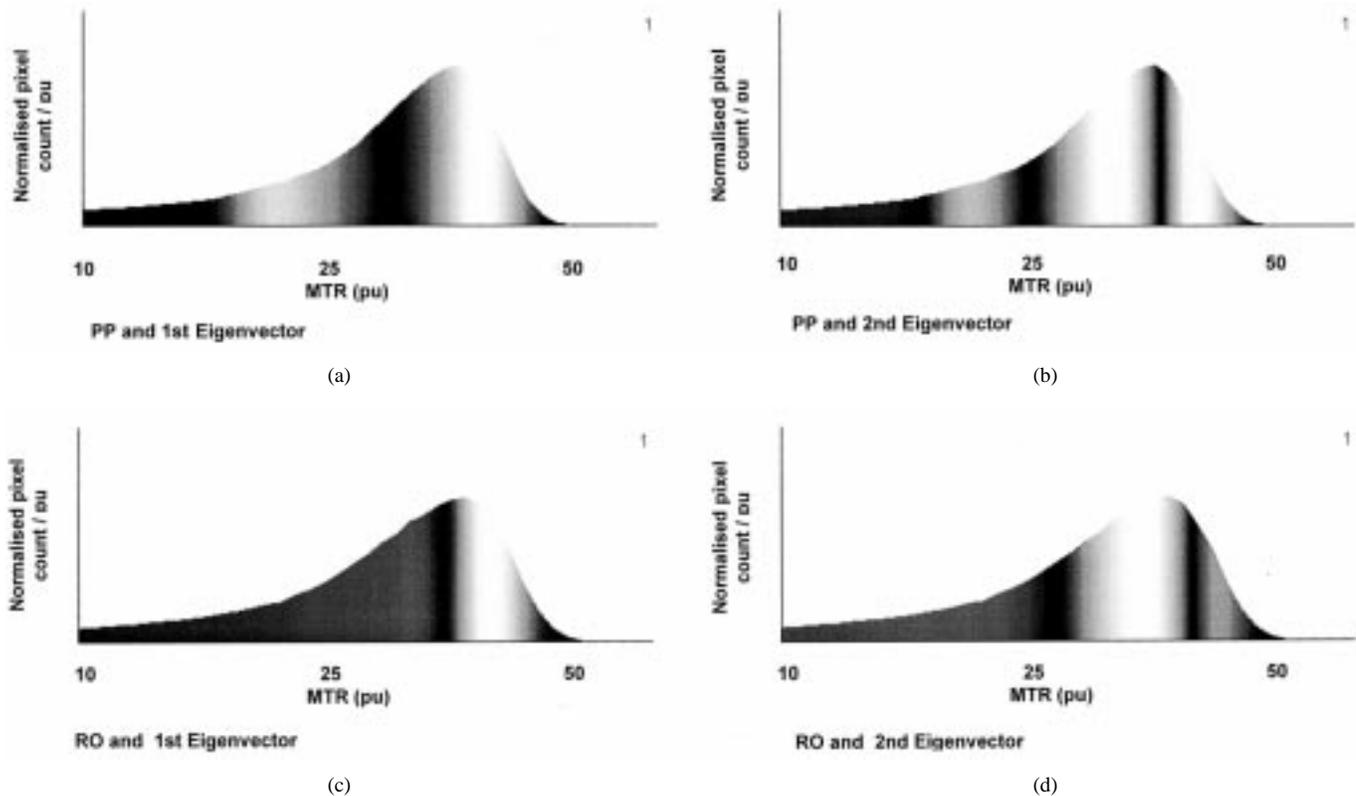


Fig. 9. PCA- shaded histogram corresponding to Fig. 8. (a), (b) Primary Progressive Group Shaded Histograms. (c), (d) Relapsing onset group shaded histograms.

several aspects of disability, thus, it would be very surprising if we achieved 100% correlation [13].

The analysis methods used here have been chosen to be optimal for correlation and classification; they make no attempt to give a biological interpretation of the data. Indeed this is a well-known characteristic of PCA—optimal correlation is obtained but interpretation is not advanced. Nonetheless, the eigenvector, and the consequent images, do give an indication of the spatial location of the biological abnormalities (variation).

The proposed approaches to MTR histogram data appear robust and offer relevant information that should allow monitoring of MS in multicenter studies. For this to be implemented successfully, retraining of the method will be initially required using control and patients subjects studied at the different sites. However, this offers new potential for standardization of quantitative data. As such, this technique deserves further investigation.

#### ACKNOWLEDGMENT

The authors would like to thank Dr N. Silver and Dr S. M. Leary for their MTR data. They would also like to thank M. Sidique and M. Richards for organizing presentation of the paper.

#### REFERENCES

- [1] M. A. Van Buchem, J. C. McGowan, D. L. Kolson, M. Polansky, and R. I. Grossman, "Quantitative volumetric magnetization transfer analysis in multiple sclerosis: estimation of macroscopic and microscopic disease burden," *Magn. Reson. Med.*, vol. 36, pp. 632–636, 1996.
- [2] M. A. Van Buchem, J. K. Udupa, J. C. McGowan, Y. Miki, F. H. Heyning, M. P. Boncoeur-Martel, D. L. Kolson, M. Polansky, and R. I. Grossman, "Global volumetric estimation of disease burden in multiple sclerosis based on magnetization transfer imaging," *Amer. J. Neuroradiol.*, vol. 18, pp. 1287–1290, 1997.
- [3] M. A. Van Buchem, R. I. Grossman, C. Armstrong, M. Polansky, Y. Miki, F. H. Heyning, M. P. Boncoeur-Martel, L. Wei, J. K. Udupa, M. Grossman, D. L. Kolson, and J. C. McGowan, "Correlation of volumetric magnetization transfer imaging with clinical data in MS," *Neurology*, vol. 50, pp. 1609–1617, 1998.
- [4] M. Rovaris, M. Filippi, M. Falautano, L. Minicucci, M. A. Rocca, V. Martinelli, and G. Comi, "Relation between MR abnormalities and patterns of cognitive impairment in multiple sclerosis," *Neurology*, vol. 50, pp. 1601–1608, 1998.
- [5] M. Filippi, G. Ianucci, C. Tortorella, L. Minicucci, M. A. Horsfield, B. Colombo, M. P. Sormani, and G. Comi, "Comparison of clinical MS phenotypes using conventional and magnetization transfer MRI," *Neurology*, vol. 52, pp. 588–594, 1999.
- [6] P. A. M. Hofman, G. J. Kemerink, J. Jolles, and J. T. Wilmink, "Quantitative analysis of magnetization transfer images of the brain: Effect of closed head injury, age and sex on white matter," *Magn. Reson. Med.*, vol. 42, pp. 803–806, 1999.
- [7] J. Dehmeshki, N. C. Silver, S. M. Leary, P. S. Tofts, A. J. Thompson, and D. H. Miller, "Magnetization transfer ratio histogram analysis of multiple sclerosis subgroups," *J. Neurological Sci.*, vol. 185, pp. 11–17, 2001.
- [8] J. Dehmeshki, N. C. Silver, A. C. Ruto, S. Arridge, D. H. Miller, and P. S. Tofts, "Detection and classification of multiple sclerosis disease uses magnetization transfer ratio images," in *Proc. 1st Int. Workshop Image and Signal Processing and Analysis IWISPA*, 2000, pp. 127–132.
- [9] J. F. Kurtzke, "Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS)," *Neurology*, vol. 33, pp. 1444–52, 1983.
- [10] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-44, pp. 532–550, 1987.
- [11] B. D. Ripley, *Pattern Recognition and Neural Networks*. Cambridge, MA: Cambridge Univ. Press, 1997.
- [12] A. Webb, *Statistical Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1999.

- [13] A. J. Thompson and J. C. Hobart, "Multiple sclerosis: assessment of disability and disability scales," *J. Neurol.*, vol. 245, no. 4, pp. 189–96, 1998.
- [14] P. Armitage and G. Berry, *Statistical Methods in Medical Research*. Oxford, U.K.: Blackwell, 1994.
- [15] J. E. Jackson, *A User's Guide to Principal Components*. New York: Wiley, 1991.
- [16] K. Fukunaga and R. Hayes, "Estimation of classifier performance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 1087–1097, 1989.
- [17] A. R. Tate, "Statistical patterns recognition for the analysis of the biomedical magnetic resonance spectra," *J. Magn. Reson. Anal.*, vol. 3, pp. 63–78, 1997.
- [18] A. J. Thompson, C. H. Polman, D. H. Miller, W. I. McDonald, M. M. Flippi, and J. De SA, "Primary progressive multiple sclerosis," *Brain*, vol. 120, pp. 1085–96, 1997.
- [19] N. C. Silver, G. J. Barker, D. G. MacManus, P. S. Tofts, and D. H. Miller, "Magnetization transfer ratio of normal brain white matter: A normative database spanning four decades of life," *J. Neurol. Neurosurg. Psychiatry*, vol. 62, pp. 223–228, 1997.