Multiple Sclerosis Detection in Multispectral Magnetic Resonance Images with Principal Components Analysis.

Dirk-Jan Kroon, Erik van Oort, and Kees Slump

University of Twente,
Institute for Biomedical Technology & Technical Medicine,
Signals and Systems
Drienerlolaan 5, 7522 NB Enschede, Netherlands
http://www.sas.el.utwente.nl

Abstract. *This paper presents a local feature vector based method for automated Multiple Sclerosis (MS) lesion segmentation of multi spectral MRI data. Twenty datasets from MS patients with FLAIR, T1,T2, MD and FA data with expert annotations are available as training set from the MICCAI 2008 challenge on MS, and 24 test datasets. Our local feature vector method contains neighbourhood voxel intensities, histogram and MS probability atlas information. Principal Component Analysis(PCA) [8] with log-likelihood ratio is used to classify each voxel. MRI suffers from intensity inhomogenities. We try to correct this "bias field" with 3 methods: a genetic algorithm, edge preserving filtering and atlas based correction. A large observer variability exist between expert classifications, but the similarity scores between model and expert classifications are often lower. Our model gives the best classification results with raw data, because bias correction gives artifacts at the edges and flatten large MS lesions.*

Key words: Multiple Sclerosis, Lesions, PCA, MR, Bias Field, Non-rigid Registration, FLAIR, Feature Vector

1 Introduction

Multiple Sclerosis (MS) is thought to be a disease in which the patient immune system damages the isolating layer of myelin around the nerve fibers. This nerve damage is visible in Magnetic Resonance (MR) scans of the brain. For instance, in MRI FLuid Attenuation Inversion Recovery (FLAIR) damage is visible as small bright spots called lesions or plaques. In both clinical trials and every day diagnostic use, the MR scans are manually read and marked by a human expert. This manual segmentation is extremely time consuming because of the large number of MR slices of each patient. Therefore, fully automated MS detection methods are being developed which can classify large amounts of MR data, and do not suffer from inter observer variability [12] [5] [7]. Currently automated

classification results show less agreement with manually classified scans, than there is between independent experts classifications.

Human experts have learned from a large number of scans to look in certain regions of a MR scan for certain abnormalities based on local parameters such as intensity, texture and size. In this paper we present an automated method for MR lesion detection which takes a similar approach. Our model describes every point in a scan by a feature vector, containing voxel intensities of 5 MR modalities, neighbourhood histogram and information from a MS registered probability atlas. A training data set consisting of 20 datasets with T1, T2, FLAIR MD and FA scans with expert classified lesions data, and a test set of 24 datasets without lesion data was provided by the Children's Hospital Boston (CHB)and University of North Carolina (UNC) as part of the MICCAI 2008 MS Lesion Segmentation challenge. The training dataset is used in our model to learn the difference between and correlations in the feature vectors from MS voxels and non-MS voxels.

This paper is organized as follows. In the next section we introduce our method, which uses a PCA model, feature vectors and atlas registration. Bias field correction algorithms are described in the third section. The fourth section describes our results, followed by the last section with our conclusions.

2 Method

Our method consist of three parts which are describe in seperate subsections below. The first subsection is about Principal Component Analysis(PCA) [8], which is the basis of our Model and used to find the correlations in feature vectors of an MS and non-MS class. The second subsection describes the features we used for our model. The third subsection describes an non-rigid registration method for development of the MS probability atlas.

2.1 The PCA Model

Detection and quantization of MS in MR images can be seen as a two class problem. A class of voxels which displays MS properties and class of voxels which do not. Each voxel can be described by a local feature vector, containing features such as voxel and neighbor voxel intensities. A training set of known MS voxels and non-MS voxels can be used to construct the two classes. A distance measure between the feature vector of a certain voxel and the class feature means and variances can be used to classify a voxel as MS or non-MS. We used the PCA classification method introduced in [10] as basis for our classification algorithm. It consists of the following steps: Constructing a matrix L_x with the feature vectors $\mathbf{x}$ of MS voxels from training data, and a matrix G_x from features vectors $\acute{\mathbf{x}}$ of non-MS voxels in the training data.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \ , \hat{\mathbf{x}} = \begin{bmatrix} \acute{x_1} \\ \acute{x_2} \\ \vdots \\ \acute{x_m} \end{bmatrix} \tag{1}$$

$$\mathbf{L_x} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \ , \ \mathbf{G_x} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_n] \tag{2}$$

The mean μ is subtracted in order to center the feature vector matrices.

$$\mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{3}$$

$$\mathbf{L_x} = [\mathbf{x}_1 - \mu_x, \mathbf{x}_2 - \mu_x, \ldots, \mathbf{x}_n - \mu_x] \ , \ \mathbf{G_x} = [\mathbf{x}_1 - \mu_x, \mathbf{x}_2 - \mu_x, \ldots, \mathbf{x}_n - \mu_x] \tag{4}$$

PCA with singular value decomposition (SVD) is used to find the eigenvectors $\mathbf{U_G}$ and eigenvalues λ in the mean centered non-MS feature data $\mathbf{G_m}$.

$$\mathbf{G_m} = \mathbf{U_G} \mathbf{\Sigma_G} \mathbf{V_G^T} \sqrt{n} \tag{5}$$

$$\lambda_i = \mathbf{\Sigma}^2_{(i,i)} \ , \ \lambda = [\lambda_1, \lambda_2, \ldots, \lambda_m] \tag{6}$$

After solving the SVD the 99% largest eigenvalues which describe the main variance in the data are kept $\mathbf{\acute{U}_G}$ and the smallest (1%) are discarded. Then the eigenvectors are multiplied with both the data sets, now the non-MS data is orthogonal.

$$\mathbf{L_{pca1}} = \mathbf{\acute{U}_G}^T \mathbf{L_m} \ , \ \mathbf{G_{pca1}} = \mathbf{\acute{U}_G}^T \mathbf{G_m} \tag{7}$$

PCA is used to find the eigenvectors and values of the MS feature data, the 99% largest eigenvalues which describe the main variance in the data are kept $\mathbf{\acute{U}_L}$ and the 1% smallest are discarded.

$$\mathbf{L_{pca1}} = \mathbf{U_L} \mathbf{\Sigma_L} \mathbf{V_L}^T \sqrt{n} \tag{8}$$

Then the variance in the eigenvectors is normalized to one, and eigenvectors are multiplied with both the data sets. Now the MS data is orthogonal.

$$\mathbf{U_{LN}} = \frac{\mathbf{\acute{U}_L}}{diag(\mathbf{\acute{\Sigma}_L})} \tag{9}$$

$$\mathbf{L_{pca2}} = \mathbf{U_{LN}}^T \mathbf{L_{pca1}} \ , \ \mathbf{G_{pca2}} = \mathbf{U_{LN}}^T \mathbf{G_{pca1}} \tag{10}$$

PCA is used to find the eigenvectors and values in the non-MS feature data. Then the eigenvectors are multiplied with both the data sets. Now both the datasets have orthogonal features.

$$\mathbf{G_{pca2}} = \mathbf{U_{G2}} \mathbf{\Sigma_{G2}} \mathbf{V_{G2}}^T \sqrt{n} \tag{11}$$

$$\mathbf{L_{pca3}} = \mathbf{U_{G2}}^T \mathbf{L_{pca2}} \ , \ \mathbf{G_{pca3}} = \mathbf{U_{G2}}^T \mathbf{G_{pca2}} \tag{12}$$

All the PCA rotations can be merged to one rotation matrix $\mathbf{U_{tot}}$, and after calculating the covariance matrix $\mathbf{C_{G_{pca3}}}$ from the $\mathbf{G_{pca3}}$ data the model is finished and can be used to classify test data.

$$\mathbf{U_{tot}} = \mathbf{U_G U_{LN} U_{G2}} \tag{13}$$

The Mahalanobis distance is used to calculate the distance of the feature vector of a certain test voxel on coordinates x,y,z to both the MS and non-MS classes. A threshold is applied to the log-Likelihood ratio $\tilde{\Lambda}(u, v)$ of the distances to get the binary MS classification.

$$\mathbf{u} = \mathbf{U_{tot}}^T(\mathbf{x} - \mu_\mathbf{G}) \ , \ \mathbf{v} = \mathbf{U_{tot}}^T(\mathbf{x} - \mu_\mathbf{L}) \tag{14}$$

$$\mathbf{\Lambda}\ (x, y, z) = \frac{1}{2}(\mathbf{u}^T \mathbf{C_{G_{pca3}}}^{-1}\mathbf{u} - ||\mathbf{v}||^2) + ln(|\mathbf{C_{G_{pca3}}}^{-1}|) \tag{15}$$

$$\mathbf{MS}(x, y, z) = \tilde{\mathbf{\Lambda}}(x, y, z) \geq t \tag{16}$$

2.2 Features

The feature vector used to classify each voxel contains MR neighbourhood intensities because they will describe the texture of the MS area. The derivatives of the voxel intensities are added to the feature vector because MS lesions are often lines or blobs in the image. The histogram information is added to provide the model with low pass intensity information of a certain region. To exclude false MS voxel detection in areas were MS is less probable, the location and distance to the brain center, and the MS atlas probability of a voxel are used as features. The table below shows the whole feature vector.

Table 1. Feature Vector

Number	Feature Description
165	Intensities of 33 voxel neighbourhood of T1, T2, FLAIR, MD and FA.
30	Intensities derivates of T1, T2, FLAIR, MD and FA in the xyz directions.
19	Local 19 bins histogram of $20x20x20$ voxel area of the FLAIR dataset.
6	From the T1, T2 and FLAIR dataset the location and height of the highest peak in a local 30 bin histogram.
3	Normalized location in xyz coordinates of the voxel.
1	Distance to approximate center of brain
1	Atlas based MS probability of the voxel
225 (total)	

2.3 Non-Rigid Atlas Registration

Our method uses a MS atlas as feature vector. This atlas is constructed from the expert annotated lesions in the training data. First the widely used NMI T1 atlas

from SPM [4] is affinely fitted (rotation, translation and resize) to all T1 training datasets. Next a non-rigid B-spline based algorithm and white matter atlas is used to get a better alignment in the white matter area. The inverse transform of the atlas registration is used to align the expert annotated lesion datasets of all patients, which are used to generate a MS probability atlas. This MS probability atlas is Gaussian smoothed, to "compensate" for the small number of datasets used.

The rigid alignment of the T1 atlas with the T1 patient datasets is based on a nonlinear least-squares optimizer from Matlab. This algorithm is a subspace trust-region method and is based on the interior-reflective Newton method described in [3] and [2]. Our system input variables are 3D rotation, size and translation. The error solved is the squared distance between the intensities of the voxels in the affine transformed atlas and the voxels on the same location in the patient T1 data.

The non-rigid alignment with the white matter probability atlas is based on the method introduced by Ruecker et al. [9]. This method is based on a Free Form Deformation (FFD) with a B-spline grid, and error calculation is done with normalized mutual information. To define a spline based FFD, we denote the domain of the image volume as $\mathbf{\Omega} = (x, y, z)|0 \leq x < X, 0 \leq y < Y, 0 \leq z < Z$. Let $\mathbf{\Phi}$ denote a n_x x n_y x n_z mesh of control points $\phi_{i,j,k}$ with uniform spacing d_x,d_y and d_z. Then, the FFD with transformation T can be written as the 3-D tensor product of the familiar 1D cubic B-splines

$$\mathbf{T} = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} \mathbf{B}_l(v)\mathbf{B}_m(u)\mathbf{B}_n(w)\mathbf{\Phi}_(i+l, j+m, k+n) \tag{17}$$

where

$$i = \lfloor x/n_x \rfloor - 1 \ , \ j = \lfloor y/n_y \rfloor - 1 \ , \ k = \lfloor z/n_z \rfloor - 1 \tag{18}$$

$$u = x/d_x - \lfloor x/d_x \rfloor \ , \ v = y/d_y - \lfloor y/d_y \rfloor \ , \ w = z/d_z - \lfloor z/d_z \rfloor \tag{19}$$

Note: equation 19 is our corrected version of the equation in the paper of Ruecker et al. [9] which stated that $u = x/n_x - \lfloor x/n_x \rfloor$.

$\mathbf{B}_l$ represents the lth basis function of the B-spline

$$\begin{aligned} \mathbf{B}_0(u) &= (1-u)^3/6 \\ \mathbf{B}_1(u) &= (3u^3 - 6u^2 + 4)/6 \\ \mathbf{B}_2(u) &= (-3u^3 - 3u^2 + 3u + 1)/6 \\ \mathbf{B}_3(u) &= u^3/6 \end{aligned} \tag{20}$$

The alignment of the transformed atlas $\mathbf{A}(\mathbf{\Phi})$ to the T1 dataset $\mathbf{B}$ with a certain control grid $\mathbf{\Phi}$, can be described by the normalized mutual information $\mathbf{C}(\mathbf{\Phi})$, which can be written as an error measure $\mathbf{E}(\mathbf{\Phi})$.

$$\mathbf{C}(\mathbf{\Phi}) = \frac{H(\mathbf{A}(\mathbf{\Phi})) + H(\mathbf{B})}{H(\mathbf{A}(\mathbf{\Phi}), \mathbf{B})} \ , \ \mathbf{E}(\mathbf{\Phi}) = 2 - \mathbf{C}(\mathbf{\Phi}) \tag{21}$$

The iterative alignment is done with a steepest gradient decent of the error function $\mathbf{E}$. The gradient of a certain control point $\mathbf{\Phi}(i)$ is calculated by moving a control point a Δ step up, down, right, left, forwards and backwards, and calculation of the error ϵ in the affected area for all those 6 atlas transformations, after which the gradient $\nabla\mathbf{E}(\mathbf{\Phi}(i))$ on the point is known.

$$\mathbf{\Phi}(t+1) = \mathbf{\Phi}(t) + \mu\frac{\nabla\mathbf{E}}{||\nabla\mathbf{E}||} \tag{22}$$

$$\nabla\mathbf{E} = \frac{\partial\mathbf{E}(\mathbf{\Phi})}{\partial\mathbf{\Phi}} \tag{23}$$

Our final implementation of this algorithm contains control mesh refinement and other improvements, and is available as "Matlab C-code" on Mathworks.com [1].

3 Bias Field Correction

MR scans suffer from slowly varying intensity inhomogeneities which is called "the Bias Field". This low frequency artifact is caused by imperfections in the RF coils, inhomogenities in the magnetic field and some problems associated with the acquisition sequences. Our PCA method will also be influenced by the bias field, because the varying intensity influences the linear correlation between the used features. Therefore we will test different bias field correction methods: a genetic algorithm, atlas based correction, edge preserving filtering and histogram correction.

3.1 Genetic Bias Field Correction

The Genetic bias field method introduced here can be used to correct MR modalities with a single histogram peak, such as FLAIR, FA and MD. The RF inhomogenities will broaden the histogram peaks. This characteristic will be used to correct the datasets. As basis of this method, a genetic algorithm was used to select, cross and mutate a pool of intensity correction models. The bias field is slowly varying in a single slice and can be described by a small intensity grid with size $9x9$. Initially, a pool of 100 random grids is generated with values between 0.5 and 2. With every iteration, a three model tournament system [1] is used to randomly select, cross and mutate the population in the pool. In a tournament system, a fixed number of models are randomly selected from the population. The ones with the lowest fitness score are deleted and replaced by crossing and mutating the remaining models. Those models are then returned to the pool and the process is repeated.

To determine the fitness of a model, this grid model is resized with bicubic interpolation to fit a MRI slice. This matrix is then multiplied with this slice to apply the correction. The fitness t of this corrected slice is calculated with the following function. Note that h represents the histogram of the slice and

[1] url: http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=20057

that the highest (256) and lowest bin (1) are ignored to prevent the model from converging on an all black or all white solution.

$$t = \sum_{i=2}^{255} h(i)^2 \tag{24}$$

The final population of 100 models is used to correct every slice of the data by processing every slice and applying the model with the highest fitness for that slice.

3.2 Atlas Based Correction

Atlas based bias field correction can be done in two ways. The first method is to non-rigidly fit a T1 or T2 atlas to a T1 respectively T2 patient scan, and improve the low-pass correspondence between the patient scan and atlas. The second method is to fit a white, gray and CSF probability atlas to the patient scan, and combine intensity based probabilities with the probabilities from the atlas to determine the bias field.

Our method is based on the probabilities atlases. First step of our method is to non-rigidly fit the white, gray and CSF probability atlases. Second step is to calculate a weighted histogram for each tissue type. The equation below is used looping through all voxel locations x,y,z.

$$\mathbf{H}(\lfloor \mathbf{I}(x) N_{bins} \rfloor) = \mathbf{H}(\lfloor \mathbf{I}(x,y,z) N_{bins} \rfloor) + \mathbf{P}_{white}(x,y,z)\ ,\ with\ \mathbf{I} \in [0,1] \tag{25}$$

Third step is to calculate the white, gray, CSF and background intensity probabilities by simply dividing each histogram by the sum of all bins. Fourth step is combining the atlas $\mathbf{P_a}$ and intensity $\mathbf{P_i}$ probabilities for each class.

$$\mathbf{P}(x,y,z) = \mathbf{P_a}(x,y,z)\mathbf{P_i}(\mathbf{I}(x,y,z)) \tag{26}$$

Fifth step is take only voxels with high probability for a certain matter (1%), and calculate the bias field on this pixels by dividing through the mean intensity of the class.

$$\mathbf{I_{bias}}(\acute{x},\acute{y},\acute{z}) = \mathbf{I_{white}}(\acute{x},\acute{y},\acute{z})/\mu_{white} \tag{27}$$

This sparse bias field is interpolated with Delaunay triangulation interpolation to fit the whole volume. Last step is dividing by the bias field to correct it, after which the total method is repeated a couple of times.

3.3 Edge Preserving Filtering

The most straightforward method to correct the slowly varying bias field, is to low pass filter the MR scans, and divide the MR scan by the low pass filtered scan. Simple Gaussian filtering will filter over region edges which causes under

and over compensation of the bias field. A filter which smooth regions while preserving the edges is diffusion filtering. We use gradient anisotropic diffusion filtering (FGAD) which is available in ITK [6]. In this implementation the conductance term $\mathbf{C}(x)$ is chosen as a function of the intensity gradient magnitude $||\nabla I(x)||$ reducing the strength of diffusion at edge pixels. We set the conductance constant k to 3.

$$\mathbf{C}(x) = e^{-(\frac{||\nabla \mathbf{I}(x)||}{k})^2} \tag{28}$$

4 Results

We have split our training set with expert annotations in two groups, 5 datasets from UNC (1,3,5,7 and 9) are used as test set and the other 15 datasets with annotations of a CHB expert are used as training set. The training sets and test sets are compared, and because there are large differences in appearance, only the 6 datasets with the closest features are used to train the model for a certain test set. The Sørensen [11] similarity index of pixel overlap is used to compare the annotations of the CHB expert with the annotations from a UNC expert, with raw data and bias corrected data classification results from our model. The threshold values for binary model segmentations from the obtained likelihood segmentations are values chosen for maximal pixel overlap with the CHB expert.

The classification similarity scores are shown in figure 1. The similarity between the CHB expert and UNC expert is higher than our model classifications on datasets 1,3,5 and 7, but also shows high observer variability. Our model classification with raw data scored 3.7 times better than the expert on dataset 9. The reason that bias corrected data scored often lower than raw MR data, is probably due to artifacts at the edges and large lesion areas which are intensity flattened. In FLAIR data MS shows as bright spots but also those bias correction edge artifacts, see figure 2. Figure 2 also shows a typical classification result with RAW data, which seems to match the bright spots in FLAIR.

The results for the test data are shown in tables 2,3 and 4. Note that the STAPLE results are discarded with regards to the overall score. The results show that our method scores about the same on the UNC and CHB rated data. Only the volume difference is slightly better with the CHB rater. It is also noteworthy that the scores vary by a large amount. With some cases the score is higher than the human expert (e.g. CHB case 03), but with others it is very poor (e.g. CHB case 01). It is also obvious that in some cases the gap between the CHB and the UNC rater is quite large. For example, the score for the volume difference for UNC case 02 is 0 with the UNC rater and 95 with the CHB rater. In general, the scores show that the number of true and false positives is not yet good enough.

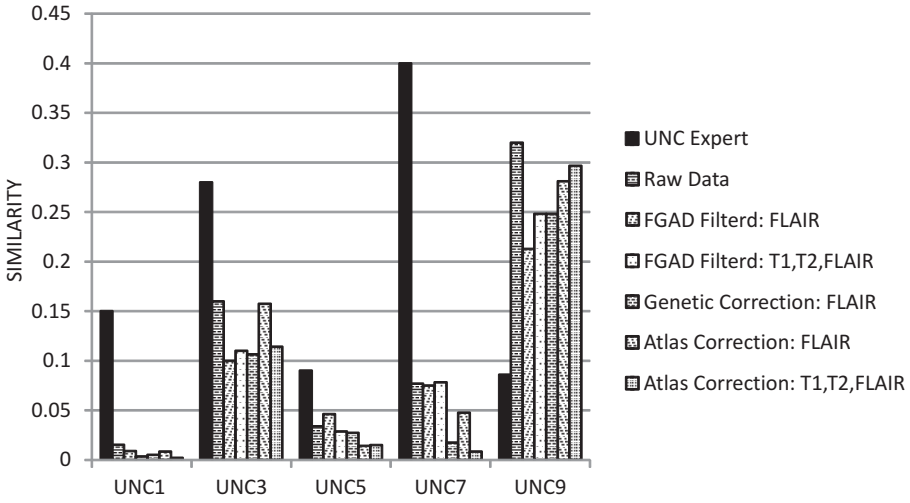

Fig. 1. Similarity between model results and CHB expert annotations

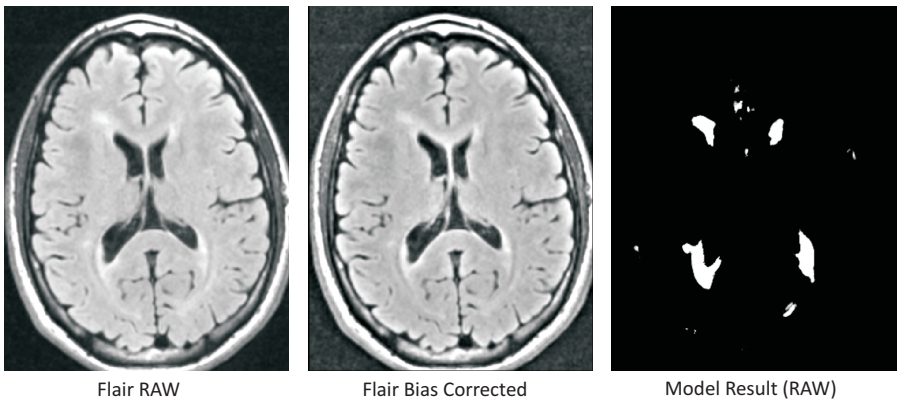

Fig. 2. Raw FLAIR data, FGAD bias corrected FLAIR and classification result. *Bright artifacts visible at the image edges after filtering and large MS areas less bright. Classification result using raw data and threshold value of zero.*

Table 2. Results part 1

Ground Truth	UNC rater							
All Datasets	Volume Diff. [%]	Score	Avg. Dist [mm]	Score	True Pos. [%]	Score	False Pos. [%]	Score
UNCtest1Case01	178.9	74	6.4	87	32.6	70	92.4	53
UNCtest1Case02	908.2	0	6.5	87	64.7	88	95.4	52
UNCtest1Case03	103.1	85	7.0	86	16.9	61	85.3	58
UNCtest1Case04	149.7	78	15.3	68	44.7	77	85.8	57
UNCtest1Case05	359.0	47	5.8	88	59.5	85	96.1	51
UNCtest1Case06	55.4	92	7.2	85	41.4	75	92.9	53
UNCtest1Case07	212.9	69	6.5	87	27.9	67	92.5	53
UNCtest1Case08	247.9	64	7.4	85	34.0	71	82.5	59
UNCtest1Case09	573.6	16	38.3	21	0.0	51	100.0	49
UNCtest1Case10	237.4	65	18.5	62	15.0	60	96.9	51
CHBtest1Case01	1768.8	0	14.6	70	40.0	74	98.7	50
CHBtest1Case02	615.9	10	10.9	78	50.0	80	97.5	50
CHBtest1Case03	24.9	96	7.7	84	35.7	72	93.2	53
CHBtest1Case04	282.3	59	10.2	79	72.7	93	95.7	51
CHBtest1Case05	2519.2	0	9.8	80	51.9	81	98.8	49
CHBtest1Case06	59.8	91	4.4	91	66.7	89	89.0	55
CHBtest1Case07	401.5	41	7.4	85	31.7	69	94.3	52
CHBtest1Case08	371.4	46	5.9	88	74.1	94	92.0	54
CHBtest1Case09	159.2	77	4.7	90	28.9	68	67.2	69
CHBtest1Case10	773.3	0	10.4	79	47.4	78	97.6	50
CHBtest1Case11	727.2	0	11.1	77	36.4	72	96.5	51
CHBtest1Case12	20.4	97	4.2	91	18.1	62	74.5	64
CHBtest1Case13	465.6	32	14.5	70	60.0	86	91.5	54
CHBtest1Case15	35.6	95	5.0	90	32.9	70	84.0	58
All Average	469.6	51	10.0	79	41.0	75	91.3	54
All UNC	302.6	59	11.9	75	33.7	71	92.0	54
All CHB	588.8	46	8.6	82	46.2	78	90.8	54

Table 3. Results part 2

Ground Truth	CHB rater							
All Datasets	Volume Diff. [%]	Score	Avg. Dist [mm]	Score	True Pos. [%]	Score	False Pos. [%]	Score
UNCtest1Case01	307.8	55	8.9	82	34.4	71	94.1	52
UNCtest1Case02	33.1	95	2.0	96	52.3	81	80.3	61
UNCtest1Case03	162.5	76	6.8	86	19.9	63	85.0	58
UNCtest1Case04	287.5	58	18.3	62	59.3	85	89.0	55
UNCtest1Case05	937.0	0	8.2	83	78.3	96	97.7	50
UNCtest1Case06	99.2	85	16.6	66	62.5	87	96.6	51
UNCtest1Case07	626.0	8	8.6	82	40.0	74	93.3	53
UNCtest1Case08	468.6	31	9.2	81	66.7	89	80.7	61
UNCtest1Case09	850.3	0	45.8	6	0.0	51	100.0	49
UNCtest1Case10	1119.7	0	23.5	52	33.3	70	97.4	50
CHBtest1Case01	2595.8	0	19.5	60	61.3	86	99.2	49
CHBtest1Case02	205.1	70	5.6	88	52.6	81	93.0	53
CHBtest1Case03	39.6	94	6.9	86	40.0	74	88.3	56
CHBtest1Case04	83.9	88	5.2	89	61.1	86	91.4	54
CHBtest1Case05	396.8	42	3.2	93	65.2	89	93.2	53
CHBtest1Case06	66.9	90	4.5	91	40.9	75	93.2	53
CHBtest1Case07	205.1	70	2.7	95	28.9	68	89.5	55
CHBtest1Case08	215.5	68	4.5	91	61.8	87	93.5	53
CHBtest1Case09	118.6	83	3.3	93	22.2	64	62.5	72
CHBtest1Case10	327.0	52	3.7	92	48.3	79	90.0	55
CHBtest1Case11	167.5	75	6.1	87	34.5	71	89.1	55
CHBtest1Case12	20.8	97	4.1	92	17.9	62	72.3	66
CHBtest1Case13	246.6	64	4.9	90	38.1	73	86.6	57
CHBtest1Case15	78.8	88	5.4	89	29.8	68	85.0	58
All Average	402.5	58	9.5	80	43.7	76	89.2	55
All UNC	489.2	41	14.8	70	44.7	77	91.4	54
All CHB	340.6	70	5.7	88	43.0	76	87.6	56

Table 4. Results part 3

Final Score		STAPLE		
All Datasets	Total Score	Specificity	Sensitivity	PPV
UNC test1 Case01	68	0.8822	0.4505	0.1523
UNC test1 Case02	70	0.7635	0.7344	0.3158
UNC test1 Case03	72	0.8664	0.4108	0.1451
UNC test1 Case04	68	0.7793	0.3512	0.1158
UNC test1 Case05	63	0.7805	0.5510	0.0934
UNC test1 Case06	74	0.9730	0.1715	0.2918
UNC test1 Case07	62	0.8157	0.1419	0.0294
UNC test1 Case08	68	0.8757	0.3792	0.0803
UNC test1 Case09	30	0.9362	0.3213	0.1607
UNC test1 Case10	51	0.9389	0.3213	0.1607
CHB test1 Case01	49	0.6243	0.7091	0.0455
CHB test1 Case02	64	0.8284	0.5202	0.1343
CHB test1 Case03	77	0.9853	0.2813	0.2298
CHB test1 Case04	75	0.9581	0.6480	0.4469
CHB test1 Case05	61	0.7655	0.5986	0.1192
CHB test1 Case06	79	0.9257	0.4546	0.3023
CHB test1 Case07	67	0.8238	0.6850	0.2279
CHB test1 Case08	72	0.8500	0.7620	0.2139
CHB test1 Case09	77	0.8856	0.4703	0.2268
CHB test1 Case10	61	0.8710	0.6355	0.1943
CHB test1 Case11	61	0.8657	0.6425	0.2087
CHB test1 Case12	79	0.9515	0.4085	0.4104
CHB test1 Case13	66	0.7631	0.7073	0.1517
CHB test1 Case15	77	0.8979	0.5156	0.3628
All Average	66	0.8586	0.5039	0.2009
All UNC	63	0.8611	0.4054	0.1548
All CHB	69	0.8568	0.5742	0.2339

5 Conclusions

We have described a local feature vector based method for automated MS lesion segmentation of multi spectral MR data. All the described bias correction methods causes artifacts at region edges, which is the main reasons that our model performs better with RAW than bias corrected MR data. The two expert MS classifications have large inter-observer variability with similarities scores in the neighbourhood of 0.20. On most of the data sets, our model has a significant lower similarity with a expert classification than the classification similarity between experts. Future improvements in this model can be made by research for better local features and using separate PCA reductions for MS and non-MS data. A larger training set may also help to improve our method. Main conclusion, our method does it's intended job but needs improving before it can compete with a human expert. Also, be careful with enabling or using bias correction on MR data. The resulting artifacts can give falls MS positives/negatives.

6 Acknowledgment

This research is supported within the Virtual Patient Project by a grant from the Province of Overijssel and is part of the Twente innovation route.

References

1. T. Bäck. *Evolutionary Algorithms in Theory and Practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford University Press US, 1996.
2. T.F. Coleman and Y. Li. On the convergence of reflective newton methods for large-scale nonlinear minimization subject to bounds". Technical Report TR92-1314, Cornell University, November 1992.
3. T.F. Coleman and Y. Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, 6(2):418–445, 1996.
4. K.J. Friston, J.T. Ashburner, S. Kiebel, T.E. Nichols, and W.D. Penny. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, 2006.
5. A. Hadjiprocopis and P. Tofts. An automatic lesion segmentation method for fast spin echo magnetic resonance images using an ensemble of neural networks. *Neural Networks for Signal Processing, 2003. NNSP'03. 2003 IEEE 13th Workshop on*, pages 709–718, Sept. 2003.
6. L. Ibez, W. Schroeder, L. Ng, J. Cates, and the Insight Software Consortium. The itk software guide.
7. R. Khayati, M. Vafadust, F. Towhidkhah, and M. Nabavi. Fully automatic segmentation of multiple sclerosis lesions in brain mr flair images using adaptive mixtures method and markov random field model. *Comput. Biol. Med.*, 38(3):379–390, 2008.
8. K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.
9. D. Rueckert, L. Sonoda, C Hayes, D.L.G. Hill, M Leach, and D.J. Hawkes. Non-rigid registration using free-form deformations: Application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

10. X. Shang and R.N.J. Veldhuis. Registration of hand-grip pattern in smart gun. In *17th Annual Workshop on Circuits, Veldhoven, The Netherlands*, pages 192–195, Utrecht, November 2006. STW.
11. T Sorensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons, biol. skrifter 5, 1948.
12. Y Wu, SK Warfield, IL Tan, WM Wells, DS Meier, RA van Schijndel, F Barkhof, and CR. Guttmann. Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. *NeuroImage*, 33(3):1205–15, 2006.