ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 39 (2008) 987–996

Texture-based segmentation of diffuse lesions of the brain's white matter

Frithjof Kruggel,[a,*] Joseph Suresh Paul,[a] and Hermann-Josef Gertz[b]

[a]*204 Rockwell Engineering Center, University of California, Irvine, CA 92697-2755, USA*
[b]*University of Leipzig, Leipzig, Germany*

Received 20 June 2007; revised 18 September 2007; accepted 22 September 2007
Available online 11 October 2007

Diffuse lesions of the white matter of the human brain are common pathological findings in magnetic resonance images of elderly subjects. These lesions are typically caused by small vessel diseases (e.g., due to hypertension, diabetes), and related to cognitive decline. Because these lesions are inhomogeneous, unsharp, and faint, but show an intensity pattern that is different from the adjacent healthy tissue, a segmentation based on texture properties is proposed here. This method was successfully applied to a set of 116 image data sets of elderly subjects. Quantitative measures for the lesion load are derived that compare well with results from experts that visually rated lesions on a semiquantitative scale. Texture-based segmentation can be considered as a general method for lesion segmentation, and an outline for adapting this method to similar problems is presented.

Keywords: White matter; Lesion segmentation; MRI; Texture; Aging

Introduction

In elderly patients suffering from cognitive impairment, an important differential diagnosis with therapeutic implications is drawn between the presence of a neurodegenerative disease (e.g., Alzheimer's disease, AD), a cerebrovascular disease (e.g., a multiinfarct cognitive disorder), or a combination thereof. In magnetic resonance imaging (MRI) of the brain, cortical and hippocampal atrophy are found as well-known signs for AD (Wolf et al., 2003), while focal or diffuse lesions are signs of cerebrovascular disease (Englund, 2002; Erkinjuntti et al., 1996; Tullberg et al., 2002). Especially, "periventricular lesions" (PVLs) and "diffuse white matter hyperintensities" (DWMHs) are often found in the subcortical white matter (WM) (Bowen et al., 1990; Deary et al., 2003; Fazekas et al., 1987; Guttmann et al., 1998). The amount of these abnormalities is correlated with age and risk factors for diseases affecting small brain vessels such as hypertension and diabetes (Bowen et al., 1990; Englund, 2002; Fazekas et al., 1998; Tullberg et al., 2002), and increased in the presence of memory disorders or dementia (Bowen et al., 1990; de Groot et al., 2002; Deary et al., 2003; Desmond, 2002; Starr et al., 2003; Tullberg et al., 2002). PVLs are typically adjacent to the corner of the lateral ventricle, as "caps" on the frontal horns, or "bands" along the trigone and occipital horn (see Fig. 1). DWMHs are located in the deep white matter, and appear as faint, nodular or confluent, patchy lesions with fuzzy borders. For an excellent discussion of the neuropathological features of the lesions refer to (Braffman et al., 1988; Englund et al., 1988; Erkinjuntti et al., 1996; Fazekas et al., 1998).

Visual assessment of T_2-weighted MR images is still the most widely used practice for evaluating these lesions: Fazekas et al. (1987) proposed a semiquantitative rating 4-point scale (0: absence; 1: mild; 2: moderate; 3: severe). Intensity is the most important image feature employed in segmentation approaches, either by using MR contrast agents (e.g., in the case of multiple sclerosis lesions, Parodi et al., 2002), by acquiring special MR imaging protocols (e.g., a fast-fluid-attenuated inversion recovery (FLAIR) sequence, Gootjes et al., 2004), or by combining information from multiple protocols (e.g., T_1, T_2, and PD-weighting, Zijdenbos et al., 1994). Lesions are segmented as outliers of the intensity distribution in monomodal images (Jack et al., 2001), or by intensity-based classification in multimodal images (Anbeek et al., 2004; Admiraal-Behloul et al., 2005; Tullberg et al., 2002; Zijdenbos et al., 1994). Due to the partial volume effect, white matter lesions may have similar intensity values as grey matter, so the use of a white matter template (an atlas) was suggested to define the search space for lesions (Admiraal-Behloul et al., 2005; DeCarli et al., 2005). Besides these voxel-based approaches, the spatial homogeneity of lesion may be used as an additional segmentation criterion. These "region-growing" procedures typically require a seed point specified by an expert (Payne et al., 2002; Parodi et al., 2002; Gootjes et al., 2004). Finally, DeCarli et al. (2005) proposed a classification procedure that is based on the intensity distribution of the WM and the distance of a voxel to the next cerebro-spinal fluid (CSF) compartment. For a comprehensive comparison of current ap-

* Corresponding author. Fax: +1 949 824 1727.
E-mail address: fkruggel@uci.edu (F. Kruggel).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2007.09.058

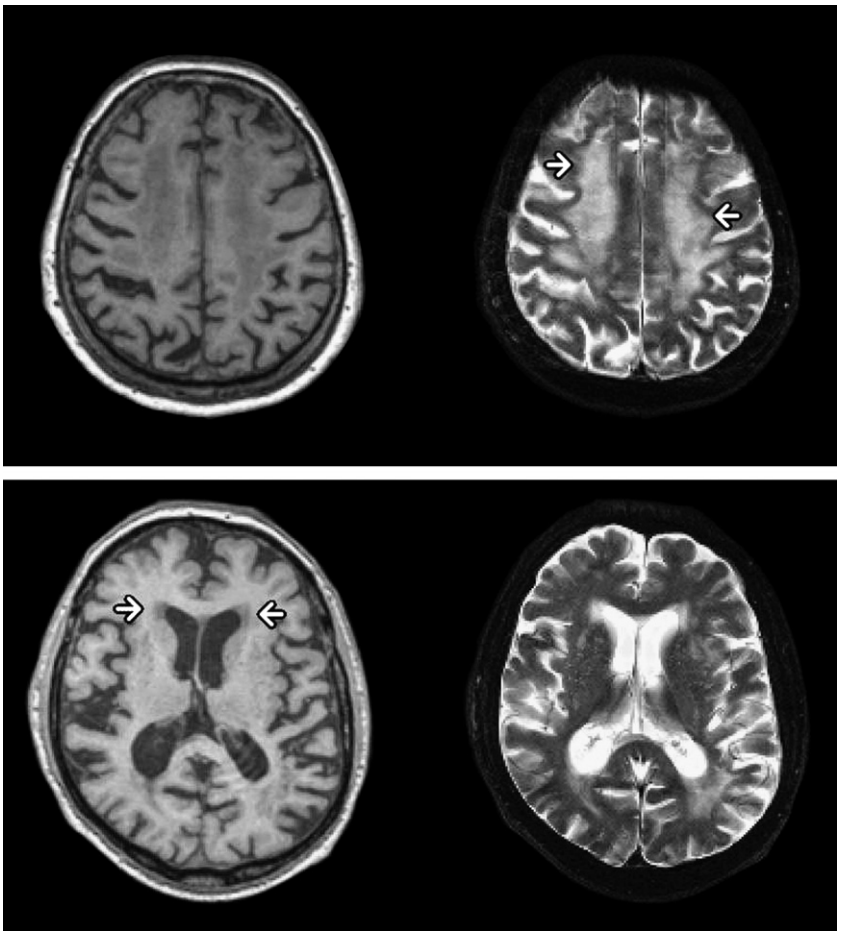

Fig. 1. T_1- (left) and T_2-weighted images (right) of diffuse white matter hyperintensities (DWMH, top) and periventricular lesions (PVL, right). Arrows point to example lesion areas.

proaches for segmenting WM hyperintensities, refer to Yoshita et al. (2005).

Because these lesions have an irregular shape and inhomogeneous structure (see Fig. 1), segmentation approaches based on intensity or shape alone may yield imprecise results. Thus, characterizing and classifying these unsharp, faint, inhomogeneous lesions by their textural characteristics appear promising. We briefly indicated this idea in Kovalev et al. (2001), and elaborate here on a detailed procedure for texture-based segmentation of WM lesions. Indeed, a wide variety of textures are encountered in biomedical images, and recent 3D CT and MRI images show rich correlates of the natural texture of organs: examples for oriented textures are muscles fibers or white matter tracts. The advantage over current intensity- or region-based approaches discussed above is that the intensity *pattern* of a lesion is classified here—and different lesion types may be discriminated by their different intensity pattern. The approach is applicable to mono- and multimodal images: using multiple weightings generally increases the sensitivity and specificity of lesion detection.

Intensity properties of textures may be described by a grey-level co-occurrence matrix (COM, refer to Rangayyan, 2005 for a detailed introduction). An element in this two-dimensional matrix represents the probability of occurrence of a pair of intensity levels i_1,i_2 of neighboring voxels v_1,v_2. Thus, the COM describes the joint intensity distribution of neighboring voxels. Other useful characteristics that describe textures include the gradient magnitude g_1,g_2 or the angle a_{12} between gradients at v_1,v_2. These multi-dimensional co-occurrence matrices were introduced by Kovalev and Petrou (1996), and applied to MRI data analysis in Kovalev et al. (2001). We denote an element of the COM as texture feature, and aim at segmenting lesions from normal tissue by discriminating their texture features.

Although we focus on the detection and segmentation of PVL and DWMH here, we emphasize that our approach is applicable for solving similar problems (e.g., the detection of WM lesions in multiple sclerosis, the segmentation of edema around tumor lesions and infarct zones), and is not limited to brain image data, or even MRI as imaging modality. Therefore, we describe a general procedure for computing texture properties from small subvolumes of an image, and develop a strategy for optimizing the discrimination between two or more texture descriptors (e.g., corresponding to WM, PVL, and DWMH). Vectors of texture features may be classified in high-dimensional space, and properties of lesions understood in terms of their distinctive features. We consider as specific strengths of this approach: (1) it is conceptually simple and easy to implement, (2) it is computationally efficient (e.g., it takes less than 60s typical computation time for a brain volume), (3) taking

advantage of multiple imaging modalities is straightforward, and (4) detection and segmentation quality is based on statistical quantities.

In the next section, we formally outline our approach for texture-based segmentation and classification of feature vectors. Three experiments describe the application to the problem of segmenting PVL and DWMH in MR brain images of elderly subjects. A final discussion critically reviews the concept. With our example application, we suggest to replace the semi-quantitative ratings with objective, texture-based measures for quantifying signs of cerebro-vascular disease as revealed by MR imaging.

Algorithms

Our texture-based approach for segmenting template (pathological) structures is based on the idea that prototypical, distinctive properties of this structure are captured in a multi-sort co-occurrence matrix c_p (COM, Kovalev et al., 2001) computed from a (small) subvolume of the whole image domain. Matrices c_i are computed in subvolumes of the same dimensions at all positions i in the image domain Ω, and the similarity $p_i = sim(c_i, c_p)$ is determined. Results are compiled in the form of a "Lesion Probability Map" (LPM), where values close to 1 correspond to a high confidence that the region in the vicinity contains the template structure, and values close to 0 correspond to its absence (e.g., healthy tissue). Note that multiple template structures may be sought for. In this case, results at each position are represented as a vector of conditional probabilities, or converted into a hard classification drawn from the highest conditional probability. In the following sections, we discuss issues of (1) image pre-processing, (2) details of computing multi-sort COMs, and (3) their statistical analysis.

Pre-processing

To compare texture properties between different image data sets, it is obvious that data must have the same spatial resolution and a similar intensity distribution. Thus, pre-processing involves within-scan correction of intensity inhomogeneities and within-group normalization of image intensity and contrast (Schad, 2002). Refer to Kovalev et al. (2001) for a further discussion.

To increase the specificity of detecting a pathological structure based on prior knowledge, and to delimit computation, it is useful (but not required) to define a search space, e.g., for lesions of the white matter, a segmentation of the white matter can be employed as the search domain. Because lesions typically have an intensity distribution that is different from the surrounding tissue, providing an intensity-based segmentation is often insufficient. Whether it is necessary to provide a search domain and how this is accomplished depends on the application. Information specific to our problem is given in the experimental section.

Computation of a co-occurrence matrix

Consider a 3D digital image *im* and its corresponding gradient image *gr*, obtained by convolving *im* with a kernel based on the first derivative of a Gaussian function ∇G with standard deviation σ. The gradient magnitude image *gm* is computed at each location i by: $gm_i = \|gr_i\|$, $\forall\ i \in \Omega_{gr}$, and the gradient direction image *gd* by: $gd_i = gr_i / \|gr_i\|$, $\forall\ i \in \Omega_{gr}$. Denote the binary image *dom* as the search domain for the pathological finding, e.g., the brain or a subcompartment thereof. We briefly describe the computation of a multi-sort COM. Consider a small 3D subvolume of width centered around coordinate x_0, y_0, z_0, that is scanned for all voxel pairs. For each pair of voxels at locations (x_i, y_i, z_i) and (x_j, y_j, z_j) in this subvolume, find the intensities im_i, im_j, gradient magnitudes gm_i, gm_j, and angle ang_{ij} between gradient directions gd_i, gd_j. Note that we only study immediate neighbors, i.e., voxels connected by faces, edges, or corners. Due to the reflection symmetry of the matrices, only 7 of the 26 immediate neighbors must be examined, as indicated by the direction arrays *dx, dy, dz*. Note that we have to ensure that both voxels belong to the search domain (e.g., the white matter). Intensities, gradient magnitudes, and angles are discretized by binning functions into *NI, NG*, and *NA* levels, respectively. Suppose intensity levels of a voxel pair are $l_i = ibin(im_i)$ and $l_j = ibin(im_j)$, then the element $II[l_i][l_j]$ of the intensity co-occurrence matrix has to be incremented. Similarly, the gradient array *GG* and angle array *A* are updated. Valid voxel pairs are counted in *n*.

The subvolume size *W* is determined from textural properties of the pathological finding, and the spatial resolution of the desired probability map. With larger *W*, textural properties are more robustly described because more voxel pairs are examined; however, the spatial resolution of the resulting map decreases. For small values of *W*, bins of the co-occurrence matrices may not receive enough counts to represent the texture properly. Intensity, gradient magnitude, and gradient angle are discretized into *NI, NG*, and *NA* levels, respectively. The choice of this parameter is critical, so that sufficient counts are found when sampling over a (small) subvolume. Note that the discretization can be limited to a subrange of the possible values in the domain, e.g., for intensity:

$$ibin(im_i) = NI * \frac{im_i - imin}{imax - imin} \tag{1}$$

where *imin, imax* denote the bounds of the intensity interval. A similar mapping is defined for the gradients with limits *gmin, gmax*. The full range of gradient angles was used (0–180°). The determination of these bounds and suitable values for the parameters *NI, NG, NA*, and *W* are discussed in the experimental section. To avoid unnecessary computation, images *im* and *gm* can be replaced by their binned versions in a pre-processing step, so repeated calls to *ibin(·), gbin(·)* are not required here.

Since the search domain in dom is highly irregular, the number *n* of voxel pairs examined in a specific subvolume depends on the local structure in the subvolume, and is typically less than the maximum: $n <= 7 * W^3$. To allow a comparison between matrices obtained at different positions, bin-wise normalization by *n* is necessary. It is useful to impose a lower bound on *n*, so that only those matrices are considered that are based on a sufficient number of samples.

For further statistical analysis, arrays *II* of dimensions $NI \times NI$, *GG* of dimensions $NG \times NG$, and *A* of dimension *NA* are reshaped into a feature vector *f* with dimension $NF = NI^2 + NG^2 + NA$. Due to the chosen normalization, elements of each array *II, GG, A* sum up to 1, so elements f_1 range in [0,1], and $\sum_i f_1 = 3$ for $n > 0$. Remember that this feature vector describes the texture properties in a small subregion around location x_0, y_0, z_0, so this operation has to be repeated for all locations in the search domain dom.

Note this feature vector can be "stacked up" from different COMs by including different imaging modalities (e.g., T_1-, T_2-weighted images) or timepoints (e.g., for serial examinations). In this former case, taking advantage of a different imaging modality

may result in an increased specificity for detecting a template structure. In the latter case, a higher sensitivity for detecting changes with time may be achieved. Of course, data must be spatially aligned (e.g., by linear registration) during pre-processing.

Statistical analysis

A feature vector f can be understood as pointing towards a specific location in a texture feature space of NF dimensions. Generally, the number of observed feature vectors is much larger than NF, so a classification of feature vectors can be dealt with in this high-dimensional space. We assume that points corresponding to a specific tissue type or template structure form a distinct cluster in this space, i.e., follow a multi-dimensional Gaussian distribution. For dimensions related to intensities and gradient angles, this assumption is reasonable. The gradient magnitude shows a leftward-skewed distribution (see experimental section), that is, however, still approximately Gaussian distributed. Several options are available to study properties of feature vectors in this space:

- *Generation of a prototypical descriptor*: An expert marks positions in a single or multiple images where a template structure is present. If the points in feature space corresponding to COMs computed from these annotations form a cluster, then the center of this cluster can be regarded as a prototypical descriptor for this template. This prototypical feature vector can be computed as an element-wise (weighted) average of all expert-marked instances.
- *Similarity metric*: The similarity (f_i, f_j) between two feature vectors can be defined as their L_1-, L_2-, or L_∞-distance. The empirical probability density function of this distance can be used to derive a voxel-wise probability for the presence of a template structure.
- *Optimization of a discrimination problem*: Consider two sets of feature vectors derived from two different sample tissues (e.g., white matter and lesion), and perform an element-wise t-test between feature vectors of both samples. Sign and magnitude of the test scores offer important insight which COM bins (i.e., which texture features) are most discriminative. The absolute sum of the element-wise test scores is a measure for the discriminative ability for the current setting of parameters W, NI, NG, NA, the intensity ($imin$, $imax$), and gradient limits ($gmin$, $gmax$). This sum can be used as a measure for an automatical optimization of these 8 parameters.
- *Clustering*: Any suitable clustering approach may employed to develop more complex discrimination functions, e.g., increasing the discrimination problem from 2 to k classes. More specifically, analyzing the properties of pathological tissue is interesting: Is it possible to discriminate different regions, types, or degree of pathology within a region-of-interest?

Note that it is not necessary to perform the classification or discrimination task in high-dimensional space. In our experience, a principal component analysis (PCA) reveals that the first two components typically represent 80% of the total variance, and the first 8–10 components up to 95% for the example problem studied here. Understanding and optimizing the discrimination problem are much easier in low-dimensional space. Once a discrimination problem is optimized, it is rather a computational disadvantage to map each feature vector into a lower-dimensional space and perform the classification in this space.

Experiments

Subjects and MR imaging

Subjects were recruited from the Leipzig Longitudinal Study of Aging (LEILA 75+) (Riedel-Heller et al., 2000). Subjects were collected to represent a characteristic sample of the elderly with a cognitive continuum from normality to mild/moderate dementia within a narrow age range of 75–85 years. All subjects and/or their legal caregivers gave informed written consent to participate in this study that was approved by the local ethics committee. None of the subjects suffered from temporal lobe epilepsy, brain tumors, or major vessel infarcts. A history or presence of other cancer, diabetes, heart disease, Parkinson-like features, and mild depressive symptoms did not lead to exclusion from the study. A subset of 116 subjects was selected for this study (75 female,

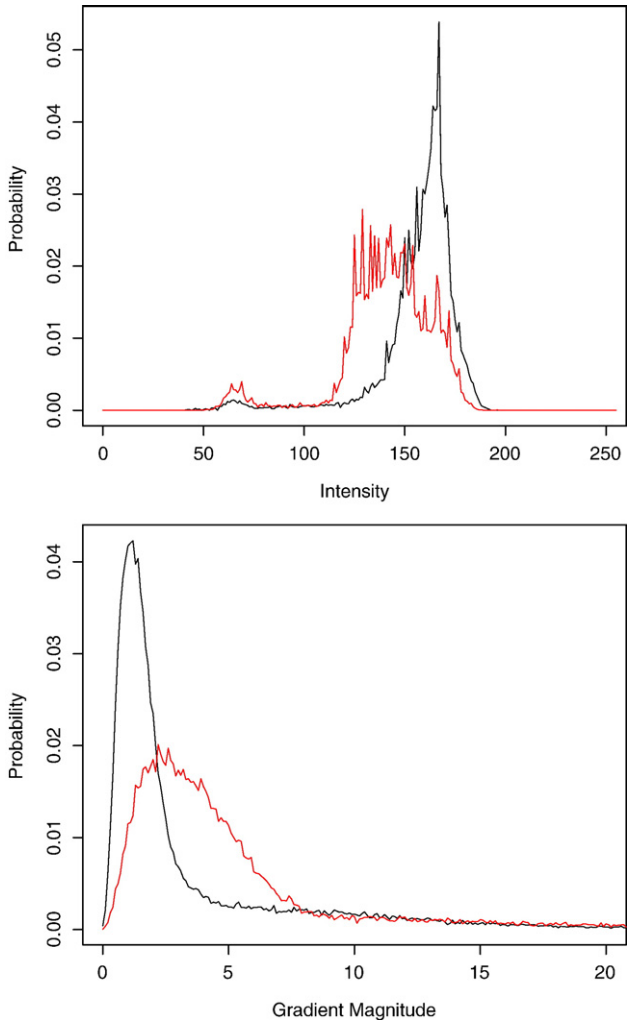

Fig. 2. Top: Intensity histogram from T_1-weighted MRI data of subvolumes sampled in healthy WM (black) and lesions (red). Below: Corresponding gradient magnitude for healthy WM (black) and lesions (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mean age: 78.56 years, range 60–91 years; 41 male, mean age: 78.27 years, range 66–88 years, age difference between gender nonsignificant).

Three-dimensional (3D) T_1-weighted high resolution MRI brain data sets were obtained on a Siemens Vision 1.5 T scanner using a 3D sequence (MPRAGE, TR 11.4 ms, TE 4.4 ms, 128 slices, matrix 256×256, voxel size $0.9 \times 0.9 \times 1.5$ mm). In addition, T_2-weighted MR images were acquired (TR 5016 ms, TE 132 ms, matrix 357×512, 19 slices, 5-mm slices, gap 1.5 mm, field of view 255×255 mm, transversal).

Imaging data were rated by a neuroradiologist for the presence of PVH and DWMH on a 4-point rating scale (0: absence; 1: mild; 2: moderate; 3: severe) (Fazekas et al., 1987).

Pre-processing

The 3D T_1-weighted data were aligned with the stereotaxic coordinate system using a 9-parameter transformation (translation, rotation and scaling) while interpolating to an isotropic voxel size of 1 mm. A fourth-order *b*-spline interpolation was used in the spatial transformation. Data were corrected for intensity inhomogeneities by a fuzzy segmentation approach using 5 classes (Pham and Prince, 1999), yielding an intensity-corrected T_1-weighted image. Within-group normalization of image intensity and contrast was performed on the distribution information obtained in the segmentation step. Data sets were finally cropped to a minimum box enclosing the head of $200 \times 256 \times 200$ mm extent. T_2-weighted data were registered with the T_1-weighted data using a 9-parameter transformation.

Experiment 1: optimizing the discrimination problem

The first experiment demonstrates the influence of the parameter setting on the discrimination problem, and is meant as an example of the optimization approach for similar detection problems. In 12 T_1-weighted data sets, a neuroradiologist marked 42 centers of subvolumes, of which 18 are attributed as "normal WM" and 24 as "DWMH". Gradient images were computed from all data sets ($\sigma = 1.0$), and regions of 15^3 voxels centered around the annotation points cut out from intensity and gradient volumes. The PDFs of intensity and gradient magnitude are shown in Fig. 2.

There is a clear distinction of the intensity distribution in both sets: the distribution in the lesion is broader, and the mean shifted to lower values. It is sufficient to focus on an interval [120, 180] for computing the *II* matrices. There is also a clear distinction of the gradient magnitude between groups: the distribution in the lesion is broader, and the mean shifted to higher values, corresponding to a more inhomogeneous pattern. The interval [0, 9] was chosen for the *GG* matrices. The bin count was set to $NI = NG = NA = 6$ in all experiments. COMs were computed for all subvolumes and widths $W \in \{7, 9, 11, 13, 15\}$. To examine which

Table 2
Discrimination power of different matrix combinations at different window widths

Descriptor	7	9	11
(A)			
II	3.112	2.293	1.626
GG	7.022	8.897	9.880
A	3.812	3.969	4.102
II+GG	7.713	7.156	6.599
II+GG+A	7.730	7.103	6.463
(B)			
II	0.804	0.723	0.652
GG	0.991	1.000	1.000
A	0.853	0.907	0.906
II+GG	0.986	0.981	0.975
II+GG+A	0.997	0.983	0.976

(A) Results for *t*-tests of L_1-distances between co-occurrence matrices vs. average of normals, expressed as *z*-scores. (B) AUC for detection. The best discrimination power at the smallest window width is found for the *II+GG+A* combination at width 7.

Table 1
Bin-wise *z*-scores for window size $W = 7$ for matrices *II* (top), *GG* (middle), and *A* (below)

Intensity	120–129	130–139	140–149	150–159	160–169	170–179
120–129	0.000	4.130	2.586	1.388	0.000	0.000
130–139	3.986	3.368	1.474	1.112	−0.060	0.000
140–149	1.766	2.005	0.526	−0.659	−1.267	0.000
150–159	0.477	0.698	−0.979	−2.991	−3.213	−1.405
160–169	0.000	−0.774	−0.227	−3.735	−3.915	−2.212
170–179	0.000	0.000	0.000	−2.088	−2.733	−2.028
Grad. mag.	0.00–1.49	1.50–2.99	3.00–4.49	4.50–5.99	6.00–7.49	7.50–8.99
0.00–1.49	−5.714	−6.819	−2.670	−0.215	−0.419	−0.849
1.50–2.99	−5.854	−5.926	−2.261	1.155	1.709	0.006
3.00–4.49	−2.278	−1.739	0.109	4.616	4.331	3.332
4.50–5.99	1.935	2.288	3.788	6.012	5.012	4.341
6.00–7.49	1.137	2.522	4.322	5.322	4.842	4.779
7.50–8.99	1.010	1.153	3.054	3.425	4.337	3.985
Grad. ang.	0–29	30–59	60–89	90–119	120–149	150–179
	4.147	−1.815	−4.069	−4.353	−2.852	−3.788

Positive *z*-scores correspond to (significantly) higher counts in pathological regions. Compare with the intensity and gradient magnitude distribution in Fig. 2. For pathological regions, the *II* scores show a shift from high-intensity to low-intensity bins along the main diagonal of *II*, along with a spread-out to off-diagonal bins. Likewise, higher gradient magnitudes are found in pathological regions.

matrix bins are most discriminative, *t*-tests were computed for each bin, comparing controls vs. lesions, and results expressed as *z*-scores for $W=7$ in Table 1.

Positive *z*-scores correspond to (significantly) higher counts in pathological regions. The average intensity decrease in the intensity histograms from normal to pathological regions corresponds here to a shift from high-intensity to low-intensity bins along the main diagonal of *II*, along with a spread-out to off-diagonal bins. Note that 13/36 bins show significant differences ($|z|>2$), and 7 bins are empty. For *GG*, a shift to higher gradient magnitudes is found, corresponding to the higher gradient magnitudes in the histogram. Here, *z*-scores are much higher, 23/36 bins show significant differences, and all bins are populated. Choosing larger windows results in marginally higher *z*-scores, presumably because COMs were computed on a larger data basis and, thus, are more robust.

The average co-occurrence descriptor was determined by element-wise averaging the corresponding feature vectors for the normal subvolumes. For different combinations of *II, GG*, and *A*, and different window widths, L_1-distances were computed to this average descriptor. *t*-tests were computed for the L_1-distances for the normal and pathological descriptors, and test scores converted to *z*-scores. Results are compiled in Table 2, left. Higher *z*-scores denote a better discrimination between normal and pathological descriptors. Alternatively, "receiver-operating-characteristic" (ROC) curves were computed while varying the L_1-distance threshold. Results are compiled as the "area under curve" (AUC) in Table 2, right.

GG matrices show the highest discrimination power which is in agreement with the large z-score differences found for *GG* bins above. The power increases with window size. Descriptors *II* and *A* are less discriminative, and the power decreases slightly with window size. Among distance measures, L_1, L_2 and L_∞, L_1-distances always achieve the highest discrimination. The best performance at the smallest window size is found for $II+GG+A$ and size 7.

A fully automatical parameter optimization is possible, e.g., using AUC as an optimization criterion while varying *imin, imax, gmin, gmax*, *W*, and perhaps, the bin counts *NI, NG*, and *NA*. However, the discrimination power derived here is optimistic: subvolumes were known to completely encompass either lesion or white matter, and fully belonged to the search domain. When processing 3D data sets, subwindows under study are rather expected to contain different tissues and/or lesions, and generally will not completely lie within the search domain.

Experiment 2: sensitivity and specificity of detection

Next, the discrimination power was tested in a full 3D data set. A neuroradiologist provided a manual segmentation of lesion areas in a data set with a high prevalence of WM abnormalities. The union of the WM segmentation (as obtained from the fuzzy classification) and the lesion annotation was used as the search space. We took advantage of using both weightings for the classification problem. The distribution of the intensity and gradient magnitude in the annotated WM and lesion voxels was studied similar as described above, yielding an intensity range for T_1: [130, 190], T_2: [80, 160], and a gradient range for T_1: [0, 12], T_2: [0, 24].

COMs were computed for both weightings and all voxels in the search space. A random subset of 1920 feature vectors (lesion: 2%, WM: 0.25%) was used as a training sample. PCA was applied to retain only 2 components, representing 83% of the cumulative variance. Because we assume that each cluster may be modeled by a multi-dimensional Gaussian distribution, it is natural to model the whole data set as a mixture of multi-dimensional Gaussian distributions. No constraints were imposed on the structure of the covariance matrix here, and parameters were estimated using the expectation- maximization (EM) algorithm (Fraley and Raftery, 2002). The optimal number of clusters may be determined from the Bayesian information criterion (BIC). The simplest model led to a clear distinction of 2 classes (Fig. 3), denoted as WM and lesion voxels, respectively.

While the feature vectors corresponding to the WM voxels show an almost symmetric distribution w.r.t. both components in the center of the plot, the lesion descriptors are clearly separated as a distinct cluster. To construct a discrimination function, 8 PCA components were retained that represented 92.5% of the total

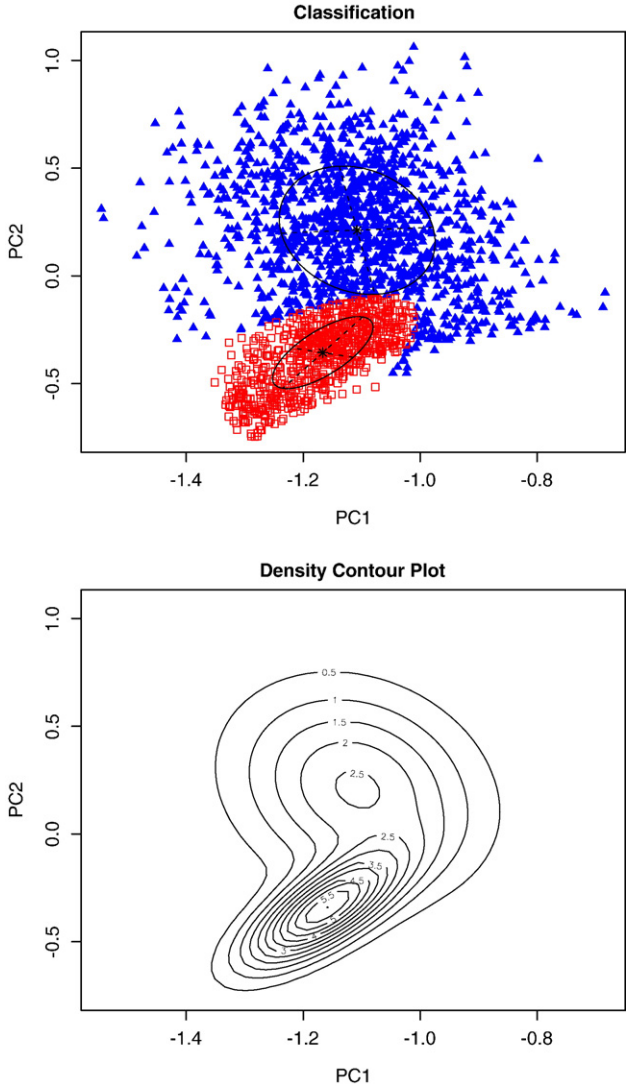

Fig. 3. A random subset of 1920 feature vectors from an MRI data set was used as a training sample. PCA was applied to retain only 2 components, representing 83% of the cumulative variance. Top: Clustering of feature vectors into two clusters, denoted as WM (blue) and lesion (red). Below: Corresponding contour map of the clustering.

variance. Classes WM and lesion were described by the mean m and variance-covariance matrix c in this 8-dimensional space.

Using this information from the training sample, feature vectors f_i for subvolumes around all voxels in the search space were computed and mapped into the 8-dimensional space. The distance to both class centers was determined, e.g.,

$$d_{i,\mathrm{WM}} = (f_i - m_{\mathrm{WM}})^{\top} c_{\mathrm{WM}}^{-1} (f_i - m_{\mathrm{WM}})$$

and converted into a "lesion probability":

$$p = 1 - \frac{d_{\mathrm{LES}}}{d_{\mathrm{WM}} + d_{\mathrm{LES}}}$$

Using the expert annotation as reference, the ROC curve indicated an optimal threshold of $p=0.13$ with a sensitivity $se=0.901$ and specificity $sp=0.913$. The example segmentation is shown in Fig. 4.

False positive detections were typically found in the white matter stalks of the gyri. They are small, and their peak probability is rather low ($p<0.5$). Post-processing involves thresholding the probability map, labeling connected components, and removing small and low-probability components. Thus, the specificity can be raised to $sp=0.994$. False negative detections were typically found on the border between WM and lesion, and account for about 30% of the border voxels of the lesion. Considering the faint, fuzzy, and irregular aspect of the lesion border in the images, this rate is arguably related to the precision of the manual annotation.

This scheme is easily extended to deal with multiple different lesion types, either by increasing the number of classes, or by

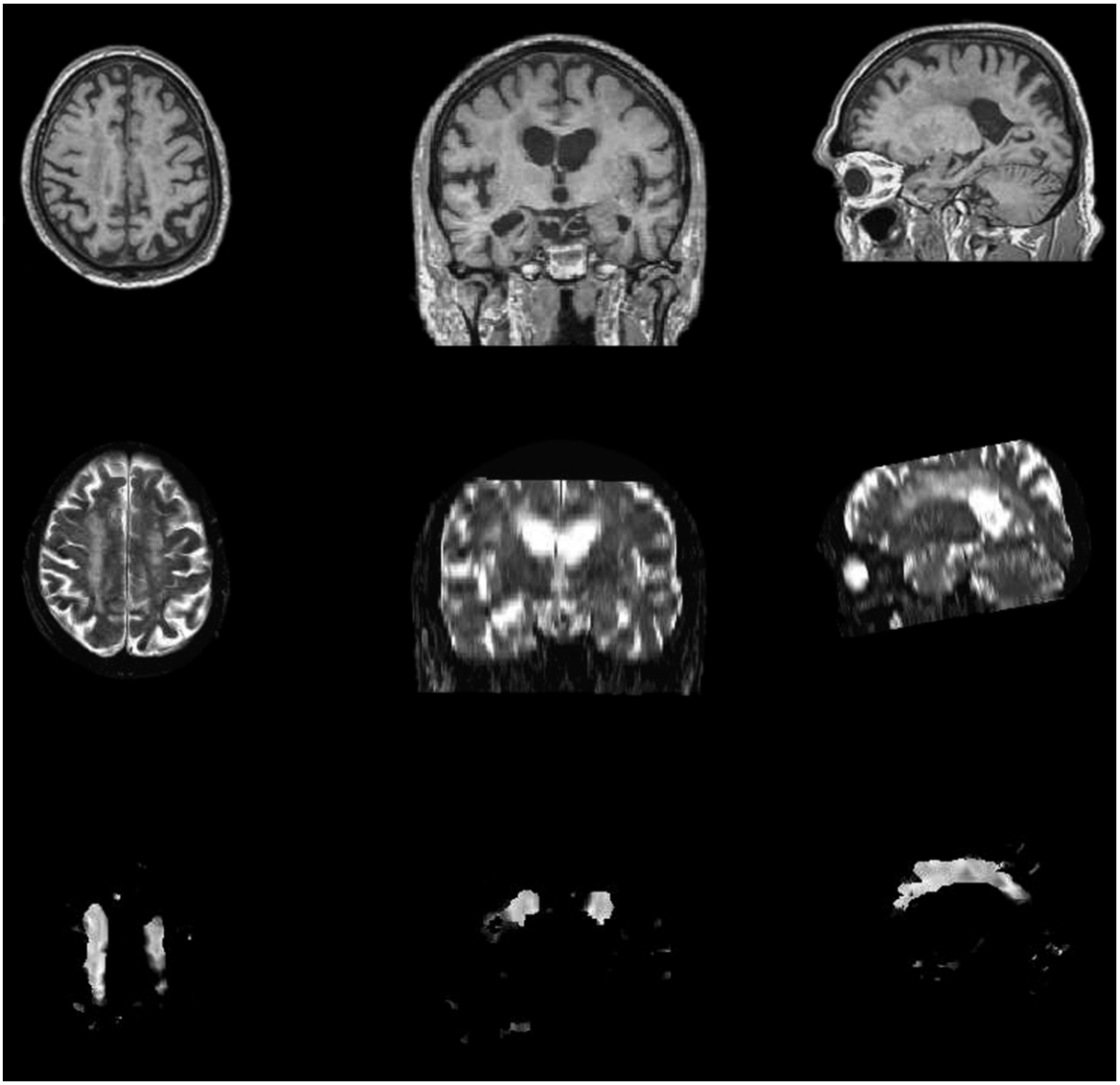

Fig. 4. Top row: Orthogonal views of T_1-weighted image. Middle row: T_2-weighted image, registered with the T_1-weighted image. Bottom row: Segmentation result, shown as a lesion probability map. Intensities range between 0 (healthy, in black) and 1 (lesion, in white). The cumulative lesion probability may be interpreted as the lesion load.

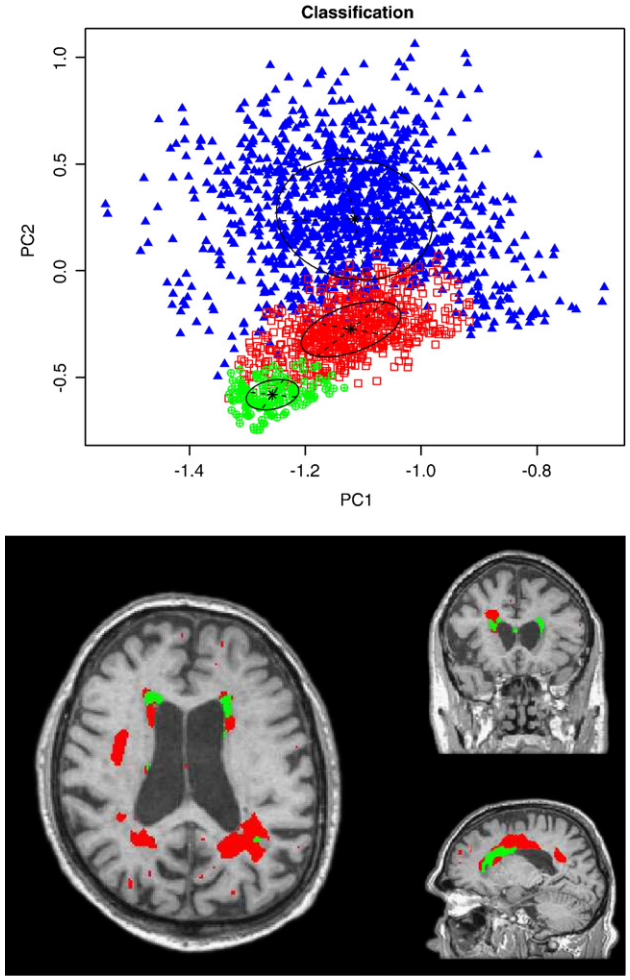

Fig. 5. Top: Discrimination of feature vectors corresponding to WM (blue), diffuse white matter hyperintensities (DWMH, red), and periventricular lesions (PVL, green). Below: Overlay of classification result onto the T_1-weighted image with DWMH (red) and PVL (green).

subclassification of the lesion voxels. The model with the best BIC value in the classification above was found with 3 classes (see Fig. 5), where (roughly) the former class of lesion voxels is now discriminated into two classes. The "green class" has a higher intensity difference from WM and higher gradient magnitude than the "red class". We consistently find voxels of the "green class" adjacent to the ventricles, and address them as "periventricular lesions" (PVL) in accordance with the description by Payne et al. (2002). Likewise "red voxels" are rather located deep in the white matter, and thus addressed as "deep white matter hyperintensities" (DWMH).

Experiment 3: analyzing single cases and groups

Finally, all data sets in the patient sample were analyzed for the presence of DWMH and PVL using texture-based segmentation. We were interested in relating a segmentation-based lesion measure with data of the expert rating scales. All 116 data sets were successfully segmented using the texture-based classifier. We recorded the number of lesion voxels n_{LES} and the lesion load l_{LES}, computed as the integral lesion probability. Both variables showed a highly significant correlation with the expert ratings ($p<10e{-}7$), but not with age and gender. However, the explained variance was rather low ($R^2=0.2433$). After visual assessment of several cases, we regard inconsistency in the expert ratings as the major source of the unexplained variance. In our view, the lesion load is the most reliable measure for quantifying WM lesions.

A subset of 28 subjects with a high prevalence of WM lesions was chosen, and another subject without a significant amount of lesions was selected as reference. A nonlinear registration between the T_1-weighted data sets of the reference and each of the 28 subjects was computed. The resulting deformation field was used to warp all lesion probability maps onto the reference. The averaged probability volume corresponds to a map of the prevalence of these WM lesions. Example axial slices at different levels above the AC-PC plane are compiled in Fig. 6. As stated in the introduction, these WM lesions are typically found in the deep supra- and periventricular WM. The processing time for a single data set, including the computation of the gradient volumes, the co-occurrence matrices, their classification, and post-processing takes about 210 s on a standard workstation (AMD 2.4 GHz processor).

Discussion

An approach for segmenting WM lesion based on their texture features was presented. Multi-sort co-occurrence matrices capture intensity, gradient, and gradient-angle distributions of neighboring voxel pairs in a small subvolume. This matrix can be recast as a vector that directs to a point in a high-dimensional texture feature space. Because lesions have different textural properties compared to the surrounding healthy tissue, it is possible to discriminate lesion-containing subvolumes in the high-dimensional texture space. We described an outline for the computation of the co-occurrence matrices and for the optimization of the discrimination problem. As an example, we provided a solution for the difficult segmentation of diffuse white matter lesions. Our approach is conceptually simple, computationally efficient, and based on statistical decisions. It is straightforward to include multiple imaging modalities—beyond our example of combining T_1- and T_2-weighted MR images, the integration of CCT-MRI or PET-MRI might be of interest.

With our segmentation problem, we typically find a high detection sensitivity, but achieving a high specificity can be an issue. False positive detections are found in subvolumes that have similar characteristics as WM lesions. Such detections are typically found at the WM/GM interface, especially in the hindbrain, and at the border of the thalamus and the third ventricle. In our case, it is easy to distinguish these false positive based on location in a post-processing step.

Some approaches for the automatical detection of WM lesions use a nonlinear spatial normalization to compare a subjects WM density with that of a group (DeCarli et al., 2005; Yoshita et al., 2005). We rather describe textural properties in a small neighborhood and classify them according to a discriminant function derived from a group. Thus, a spatial normalization is not required. By retaining the individual data space our approach allows quantifying the individual lesion load.

Although there is some neurobiological controversy about the validity of discriminating different types of white matter lesions (Chimowitz et al., 1989; Fazekas et al., 1987; Fernando et al., 2004; Payne et al., 2002; Yoshita et al., 2005), our results provide

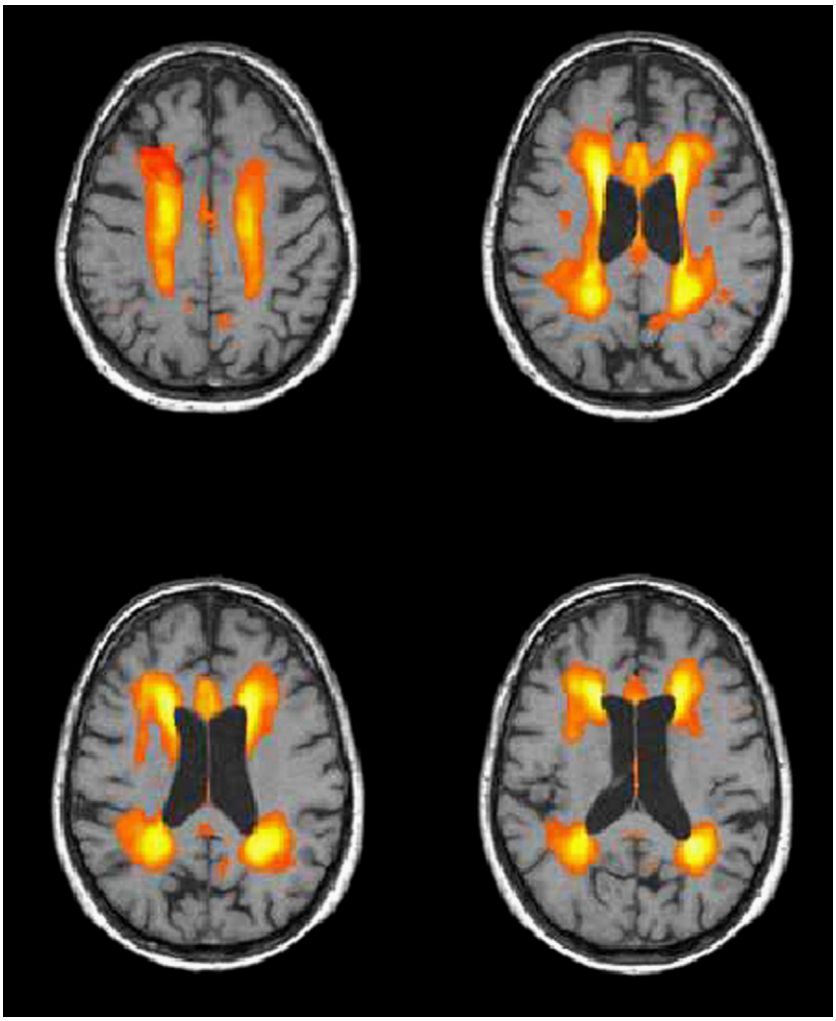

Fig. 6. Prevalence of WM lesions in axial slices 38 mm (top left), 28 mm (top right), 23 mm (bottom left), and 18 mm (bottom right) above the plane defined by the anterior and posterior commissures.

some support for separating periventricular lesions and deep white matter hyperintensities based on their textural features and location. The latter regions have rather faint borders, a "cloudlike" appearance and are located deep in the white matter (Fig. 1, top). PVL regions are characterized by their relatively sharp borders, a low intensity in T_1-weighted images, and their localization adjacent to the ventricles, especially at the anterior tip of the lateral ventricle (Fig. 1, bottom). Although the discrimination of both lesion types is in line with a post-mortem MRI-histology study (Fernando et al., 2004), we cannot provide a histological proof for our classification that is only drawn from in-vivo MRI data.

The application of this approach for segmenting lesions with similar appearance is straightforward, most notably, in multiple sclerosis (e.g., Leemput et al., 2001), in tumors and edema, CNS infections (e.g., toxoplasmosis), and hereditary WM diseases (e.g., leukodystrophy). But not only lesion size and location are important. The acuity of a disease is often related to a specific appearance, i.e., the distance of feature vectors from the normal tissue distribution may be used in the staging of tumors. "Old" and "fresh" lesions in multiple sclerosis may be distinguished.

Compared with manual segmentation and rating, advantages of an automatical approach for lesion segmentation include increased reliability, consistency, and reproducibility. Quantitative information about lesion size and characteristics may lead to a deeper understanding of disease patterns and their timecourse, and provide assistance in diagnostic and therapeutic decisions. With the increasing spatial resolution of the new generation scanners, the rich texture of tissues and lesions in organs is revealed, and we can take advantage of textural features for segmentation.

Acknowledgment

This research was partly supported by Interdisziplinäres Zentrum für Klinische Forschung (IZKF) at the University of Leipzig (Projects C8 and C15).

References

Admiraal-Behloul, F., van den Heuvel, D.M.J., Olofsen, H., van Osch, M.J.P., van der Grond, J., van Buchem, M.A., Reiber, J.H.C., 2005. Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. NeuroImage 28, 607–615.

Anbeek, P., Vincken, K.L., van Osch, M.J.P., Bisschops, R.H.C., van der

Grond, J., 2004. Probabilistic segmentation of white matter lesions in MR imaging. NeuroImage 21, 1037–1044.

Bowen, B.C., Barker, W.W., Loewenstein, D.A., Sheldon, J., Duara, R., 1990. MR signal abnormalities in memory disorder and dementia. Am. J. Radiol. 11, 1285–1292.

Braffman, B.H., Zimmerman, R.A., Trojanowski, J.Q., Gonatas, N.K., Hickey, W.F., Schlaepfer, W.W., 1988. Brain MR: pathologic correlation with gross and histopathology. 2. Hyperintense white-matter foci in the elderly. Am. J. Roentgenol. 151, 559–566.

Chimowitz, M.I., Awad, I.A., Furlan, A.J., 1989. Periventricular lesions on MRI. Facts and theories. Stroke 20, 963–967.

de Groot, J.C., de Leeuw, F.E., Oudkerk, M., van Gijn, J., Hofman, A., Jolles, J., Breteler, M.M.B., 2002. Periventricular cerebral white matter lesions predict rate of cognitive decline. Ann. Neurol. 52, 335–341.

Deary, I.J., Leaper, S.A., Murray, A.D., Staff, R.T., Whalley, L.J., 2003. Cerebral white matter abnormalities and lifetime cognitive change: a 67-year follow-up of the Scottish Mental Survey of 1932. Psychol. Aging 18, 140–148.

DeCarli, C., Fletcher, E., Ramey, V., Harvey, D., Jagust, W.J., 2005. Anatomical mapping of white matter hyperintensities (WMH): exploring the relationship between periventricular WMH, deep WMH, and total WMH burden. Stroke 36, 50–55.

Desmond, D.W., 2002. Cognition and white matter lesions. Cerebrovasc. Dis. 13, 53–57.

Englund, E., 2002. Neuropathology of white matter lesions in vascular cognitive impairment. Cerebrovasc. Dis. 13, 11–15.

Englund, E., Brun, A., Alling, C., 1988. White matter changes in dementia of Alzheimer's type: biochemical and neuropathological correlates. Brain 111, 1425–1439.

Erkinjuntti, T., Benavente, O., Eliasziw, M., Munoz, D.G., Sulkava, R., Haltia, M., Hachinski, V., 1996. Diffuse vacuolization (spongiosis) and arteriolosclerosis in the frontal white matter occurs in vascular dementia. Arch. Neurol. 53, 325–332.

Fazekas, F., Chawluk, J.B., Alavi, A., Hurtig, H.I., Zimmerman, R.A., 1987. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. Am. J. Neuroradiol. 8, 421–426.

Fazekas, F., Schmidt, R., Scheltens, P., 1998. Pathophysiologic mechanisms in the development of age-related white matter changes of the brain. Dement. Geriatr. Cogn. Disord. 9, 2–5 (suppl).

Fernando, M.S., O'Brien, J.T., Perry, R.H., English, P., Forster, G., McMeekin, W., Slade, J.Y., Golkhar, A., Matthes, F.E., Barbert, R., Kalaria, R.N., Ince, P.G., 2004. Comparison of the pathology of cerebral white matter with post-mortem magnetic resonance imaging (MRI) in the elderly brain. Neuropathol. Appl. Neurobiol. 30, 385–395.

Fraley, C., Raftery, A.E., 2002. Model-based clustering, discriminant analysis, and density estimation. J. Am. Stat. Assoc. 97, 611–631.

Gootjes, L., Teipel, S.J., Zebuhr, Y., Schwarz, R., Leinsinger, G., Scheltens, P., Möller, H.J., Hampel, H., 2004. Regional distribution of white matter hyperintensities in vascular dementia, Alzheimer's disease and healthy aging. Dement. Geriatr. Cogn. Disord. 18, 180–188.

Guttmann, C.R., Jolesz, F.A., Kikinis, R., Killiany, R.J., Moss, M.B., Sandor, T., Albert, M.S., 1998. White matter changes with normal aging. Neurology 50, 972–978.

Jack, C.R., O'Brien, P.C., Rettman, D.W., Shiung, M.M., Xu, Y., Muthupillai, R., Manduca, A., Avula, R., Erickson, B.J., 2001. FLAIR histogram segmentation for measurement of leukoaraiosis volume. J. Magn. Reson. Imaging 14, 668–676.

Kovalev, V.A., Petrou, M., 1996. Multidimensional co-occurrence matrices for object recognition and matching. Graph. Models Image Proc. 58, 187–197.

Kovalev, V.A., Kruggel, F., Gertz, H.J., von Cramon, D.Y., 2001. 3D texture analysis of MRI brain datasets. IEEE Trans. Med. Imag. 20, 424–433.

Leemput, K.V., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P., 2001. Automated segmentation of Multiple Sclerosis lesions by model outlier detection. IEEE Trans. Med. Imag. 20, 677–688.

Parodi, R.C., Sardanelli, F., Renzetti, P., Rosso, E., Losacco, C., Ferrari, A., Levrero, F., Pilot, A., Inglese, M., Mancardi, G.L., 2002. Growing region segmentation software (GRES) for quantitative magnetic resonance imaging of multiple sclerosis: intra- and inter-observer agreement variability: a comparison with manual contouring method. Eur. Radiol. 12, 866–871.

Payne, M.E., Fetzer, D.L., MacFall, J.R., Provenzale, J.M., Byrum, C.E., Krishnan, K.R.R., 2002. Development of a semi-automated method for quantification of MRI gray and white matter lesions in geriatric subjects. Psych. Res. Neuroimag. 115, 63–77.

Pham, D.L., Prince, J.L., 1999. An adaptive fuzzy segmentation algorithm of magnetic resonance images. IEEE Trans. Med. Imag. 18, 737–752.

Rangayyan, R.M., 2005. Biomedical image analysis. CRC Press, London, pp. 583–796.

Riedel-Heller, S.G., Schork, A., Matschinger, H., Angermeyer, M.C., 2000. Recruitment procedures and their impact on the prevalence of dementia. Results from the Leipzig Longitudinal Study of the Aged (LEILA 75+). Neuroepidemiology 19, 130–140.

Schad, L.R., 2002. Problems in texture analysis with magnetic resonance imaging. Dial. Clin. Neurosci. 4, 235–242.

Starr, J.M., Leaper, S.A., Murray, A.D., Lemmon, H.A., Staff, R.T., Deary, I.J., Whalley, L.J., 2003. Brain white matter lesions are detected by magnetic resonance imaging are associated with balance and gait speed. J. Neurol. Neurosurg. Psychiatry 74, 94–98.

Tullberg, M., Hultin, L., Ekholm, S., Mansson, J.E., Fredman, P., Wikkelse, C., 2002. White matter changes in normal pressure hydrocephalus and Binswanger disease: specificity, predictive value and correlations to axonal degeneration and demyelination. Acta Neurol. Scand. 105, 417–426.

Wolf, H., Jelic, V., Gertz, H.J., Nordberg, A., Julin, P., Wahlund, L.O., 2003. A critical discussion of the role of neuroimaging in mild cognitive impairment. Acta Neurol. Scand. 107, 52–76.

Yoshita, M., Fletcher, E., DeCarli, C., 2005. Current concepts of analysis of cerebral white matter hyperintensities on magnetic resonance imaging. Top. Magn. Reson. Imaging 16, 399–407.

Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palmer, A.C., 1994. Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans. Med. Imag. 13, 716–724.