I wish to express my deepest gratitude to my supervisor, Dr. Hayit Greenspan, for providing me with stimulating guidance during this undertaking in the fascinating field of medical image processing. I am much obliged to Prof. Gomori of Hadassah Medical Center, who with great patience marked the lesions in our data and gave me invaluable clinical insights. I would also like to acknowledge the assistance I received from my fellow students, Arnaldo Mayer and Shiri Gordon, especially in the initial stages of this research. The multiple sclerosis image data used here was generously sent to our lab by Dr. Warfield of Harvard Medical School.

This work would not have reached its conclusion without the continuous encouragement and loving support of my parents, Frida and Aharon, and my grandparents, Frida, Chaia and Michael. I thank them all with the greatest affection.

Love is the divine ingredient ... It is the means by which we accomplish our work

(After David B. Haight)

I lovingly dedicate this thesis to my wife, Lisa, my guiding light and inspiration.
ABSTRACT

A novel statistical video-modeling scheme is applied in this work to sequences of brain MR images taken over time. The objective of the proposed probabilistic framework is to automatically detect, segment and track-in-time regions pertaining to relapsing-remitting multiple sclerosis (MS) lesions. The clinical motivation for providing such an automatic framework is to assist the radiologist in the task of manual spatio-temporal segmentation, which is both time-consuming and non-reproducible. The suggested automatic tools can be of use when MRI acquisitions are performed in regular time intervals, for example, in research that examines the effect of drug therapy over time or in clinical cases where a quantitative analysis of MS progress is required.

Unsupervised clustering via Gaussian mixture modeling is utilized to extract coherent space-time regions (space-time “blobs”) in a four-dimensional feature space (intensity, position (x,y), and time) and corresponding coherent segments in the sequence content. The parameters of the model are determined via the Expectation-Maximization (EM) algorithm according to the maximum likelihood principle. In the detection stage of the framework, the regions corresponding to MS lesions are identified out of the collection of extracted segments by a context-based rule set. Following the detection, lesion segmentation and tracking are performed in a unified manner via global space-time modeling. A key feature of our framework is the analysis of the image-sequence throughout the scheme as a single entity, as opposed to a sequence of separate frames.

Qualitative and quantitative results of the proposed methodology are shown for a sequence of 24 T2-weighted MR images, which was acquired from a relapsing-remitting MS patient over a period of approximately a year. The validation of the framework was performed by comparing our results to an expert radiologist’s manual delineation and by conducting robustness tests using a simulation tool.
Table of Contents

Acronyms, Abbreviations and Symbols .......................................................... v
List of Figures ................................................................................................ vii
List of Tables ................................................................................................ vii

1 Introduction ............................................................................................ 1
  1.1 Clinical background .............................................................................. 1
    1.1.1 Multiple Sclerosis ........................................................................... 1
    1.1.2 Clinical analysis of MS ................................................................. 2
  1.2 Objectives of the proposed framework .................................................. 3

2 Related Works ....................................................................................... 4
  2.1 Segmentation of MS lesions ................................................................. 5
    2.1.1 Van Leemput et al. – lesions as outliers ....................................... 5
    2.1.2 Warfield et al. – separating WM/MSL from GM structures ......... 6
    2.1.3 Overall appreciation ................................................................. 7
  2.2 Tracking of MS Lesions in Time ............................................................ 8
    2.2.1 Kikinis et al. – extending the segmentation problem to 4D ........ 8
    2.2.2 Gerig et al. – analyzing temporal features of voxel intensity ...... 9
    2.2.3 Rey et al. – fitting a model to the intensity vs. time profile ...... 10
    2.2.4 Overall appreciation ............................................................... 11

3 Methodology .......................................................................................... 12
  3.1 Model assumptions and apriori knowledge .......................................... 12
  3.2 Employing GMM for image representation .......................................... 14
  3.3 Employing GMM for image segmentation ........................................... 18

4 The Proposed Framework ...................................................................... 19
  4.1 Preprocessing: extracting regions of interest ...................................... 19
  4.2 Global space-time modeling ................................................................. 22
  4.3 Lesion detection ................................................................................. 23
  4.4 Region merging .................................................................................. 24

5 Experimental Results ........................................................................... 26
  5.1 Data characteristics ........................................................................... 26
  5.2 Preprocessing .................................................................................... 28
5.3 ROI extraction based on an intensity model .................................................... 29
5.4 Global space-time model generation ............................................................... 31
5.5 Lesion detection ............................................................................................... 34
5.6 Lesion merging ............................................................................................... 35

6 Validation ........................................................................................................ 39
6.1 Comparison to an expert’s segmentation......................................................... 39
6.2 Validation via simulation................................................................................. 41

7 Discussion ........................................................................................................ 46
7.1 Modeling multiple sclerosis lesions................................................................. 46
7.2 Determining the model order ........................................................................... 48
7.3 The detection scheme....................................................................................... 49
7.4 The validation problem.................................................................................... 50
7.5 Future work ...................................................................................................... 51

Bibliography ........................................................................................................ 52
Acronyms, Abbreviations and Symbols

ACRONYMS
CC    Connected-Components procedure
CSF   Cerebrospinal Fluid (brain tissue)
DS    Dempster-Shafer Theory
EM    The Expectation-Maximization algorithm
GM    Gray Matter (brain tissue)
GMM   Gaussian Mixture Model
ICC   Intracranial Cavity
KM    The K-means clustering algorithm
MDL   The Minimal Description Length criterion
ML    Maximum Likelihood
MR    Magnetic Resonance
MRF   Markov Random Field
MRI   Magnetic Resonance Imaging
MS    Multiple Sclerosis
MSL   Multiple Sclerosis Lesions
NEX   Number of Excitations (MRI)
PVE   Partial Volume Effect
PVEC  Partial Volume Effect Correction
RG    Region Growing
ROI   Region/s of Interest
T1    Relaxation time in the z-axis (MRI)
T2    Relaxation time in the x-y plane (MRI)
TE    Echo Time (MRI)
TLL   Total Lesion Load
TR    Repetition Time (MRI)
WM    White Matter (brain tissue)

ABBREVIATIONS
E-step  Expectation step (EM algorithm)
M-step  Maximization step (EM algorithm)
SYMBOLS

K  Model order
x, y  Spatial coordinates
d  Dimensionality of the feature-space and feature-vectors
\( R^d \)  d-dimensional real space (the feature-space)
\( \theta \)  The set of model parameters
\( \theta_{ML} \)  Maximum likelihood estimation of \( \theta \)
\( \alpha_j \)  The relative weight of Gaussian cluster \( j \)
\( \mu_j \)  The mean vector of Gaussian cluster \( j \)
\( \Sigma_j \)  The covariance matrix of Gaussian cluster \( j \)
\( \hat{\alpha}_j \)  Updated relative weight of Gaussian cluster \( j \) (EM iterations)
\( \hat{\mu}_j \)  Updated mean vector of Gaussian cluster \( j \) (EM iterations)
\( \hat{\Sigma}_j \)  Updated covariance matrix of Gaussian cluster \( j \) (EM iterations)
\( w_{ij} \)  The probabilistic affiliation value of feature-vector (or pixel) \( t \) to Gaussian cluster \( j \)
x_i  Random variable \( i \); feature-vector \( i \)
p(x)  Prior probability of a random variable \( x \)
f(x/\theta)  Density function for a random variable \( x \) given a parameter set \( \theta \)
f(x/\mu_j, \Sigma_j)  Density function for a random variable \( x \) given the parameters of Gaussian cluster \( j \) (\( \mu_j, \Sigma_j \))
L(\( \theta/x \))  The likelihood of model parameters, \( \theta \), given the input data \( x \)
l_k  Number of free parameters needed for a model of \( k \) mixture Gaussian components
A-D  Labels for cases in the validation results (simulation tool)
I-VIII  Labels for dynamic lesions in the given input sequence
IX-XII  Labels for static lesions in the given input sequence
List of Figures

Figure 1 Example of an MS brain image, and of dynamic and static lesions ... 12
Figure 2 Evolution over time of pixel intensity ................................................. 13
Figure 3 The concept of space-time blobs ......................................................... 17
Figure 4 A schematic description of the proposed framework ......................... 19
Figure 5 The input sequence ........................................................................... 27
Figure 6 Removal of high intensity regions from the boundaries of the brain .. 28
Figure 7 Intensity histogram estimation by the intensity-based GMM .......... 29
Figure 8 Probabilistic segmentation maps for an intensity-based GMM (K=5) 30
Figure 9 ROI extraction .................................................................................. 31
Figure 10 ROI segmentation based on a space-time global model ............... 33
Figure 11 Detection of dynamic lesions & their spatio-temporal segmentation .. 35
Figure 12 Region merging .......................................................... 37
Figure 13 Zoom-in on lesion II in frame 8 ....................................................... 37
Figure 14 Evolution over time of the lesion size .......................................... 38
Figure 15 Manual segmentation of dynamic lesions by an expert radiologist .... 40
Figure 16 Zoom-in on expert and framework segmentation .......................... 40
Figure 17 Comparison of size evolution according to manual segmentation and framework results .......................................................... 41
Figure 18 Simulation of lesions ...................................................................... 43
Figure 19 Validation via the simulation tool – size evolution plots .............. 44
Figure 20 Images relating to cases C and D in Table 2 .................................. 44

List of Tables

Table 1 Summary of phenomena appearing in the segmentation results presented in Figure 10 .......................................................... 32
Table 2 Validation via the simulation tool ..................................................... 43
1 Introduction

1.1 Clinical background

1.1.1 Multiple Sclerosis

Multiple sclerosis (MS) is a neurological disease primarily affecting the central nervous system, typically detected in young adults (ages 20-40). Approximately 400,000 Americans acknowledge having MS, and every week about 200 people are diagnosed. Studies suggest that MS may affect as many as 2.5 million individuals worldwide. Statistically, women seem to be more susceptible to MS than men are and MS occurs more commonly among people with northern European ancestry. A specific cause for the disease has not yet been identified, although viral and autoimmune etiologies are postulated. Genetic and environmental factors are known to contribute to MS, as well [20].

Pathologically, MS is characterized by the presence of areas of demyelination and perivascular inflammation in the brain’s white matter (WM). These damaged areas of scar tissue are also known as multiple sclerosis plaques or lesions (MSL). The disease begins most commonly with the onset of neurological abnormalities. Initial and subsequent symptoms may dramatically vary in their expression and severity over the course of the disease, which usually lasts many years. Typical symptoms range from numbness and double vision to motorial disability (although usually not severe) and mental disorders [4].

There are several variants of MS, which are classified according to the clinical course of the disease. The most common of these are the relapsing-remitting and relapsing-progressive variants, found in approximately 85% of the patients. In such cases (especially in relapsing-remitting MS), the signs of demyelination and inflammation appear and disappear periodically. The patient goes through episodes of acute worsening of neurological function followed by partial or complete recovery periods (remissions) free of disease progression [21].

Although multiple sclerosis was among the first diseases to be described scientifically, there is no curative treatment available for MS. However, a number of medications can be used to treat the disease symptomatically and reduce the frequency and severity of relapses. In the future, medications targeting specific autoimmune responses, as well as cures designed to assist in remyelination, will possibly help improve the quality of life of MS patients [20].
1.1.2 Clinical analysis of MS

Magnetic resonance imaging (MRI) has long been considered the leading modality for brain scanning because of its high resolution and superior soft tissue contrast. Although it cannot be used as a stand-alone diagnostic tool for cases of MS, MRI is increasingly being used to assess the progression of MS and to evaluate the effect of drug therapy [28]. The analysis of MR images may be done qualitatively and/or quantitatively. Qualitative analysis is performed by visually estimating the disease’s progress, classifying lesions in the scan as either new lesions, changing lesions or static lesions [23]. The most common quantitative parameter is the burden (or load) of the disease expressed in terms of the volume of the lesions.

Researchers have shown that a correlation exists between patient symptoms and the size and anatomical location of lesions ([1], [17]). However, since many lesions seen on MR scans may be in the so-called “silent” areas of the brain, it is not always possible to specifically correlate the scan-based analysis and the patient’s clinical signs and symptoms. In addition, people over the age of 50 often have small findings on MRI that resemble MS, but are actually related to the aging process and have no clinical significance [15]. Other works suggest that areas of abnormal white matter, which are not identified as distinct lesions, should be included as part of the disease burden in order to achieve a better correlation to clinical symptoms [28]. In any case, according to physicians with whom this work was discussed, the detection of lesions in MR scans is still the common clinical procedure. The MSL load is of interest when analyzing a specific scan, as well as when examining temporal pathological processes, especially in the context of drug research [1].

To make a quantitative analysis of the brain scans of the patient, the clinician is required to identify the multiple sclerosis lesions that are present in those scans, i.e., he/she needs to perform manual segmentation. Manual segmentation has several shortcomings. Because of the vast amount of data presented by the MRI modality, manual analysis is very time-consuming. High inter- and intra-observer variability has been demonstrated in several studies, where segmentations, performed by a group of experts, at different time points, were compared ([22], [28]). When dealing with small lesions, the variability may be crucial as the segmentation errors are often in the range of the segmented structure’s volume. The reproducibility of manual segmentation results is further decreased when multi-spectral MRI is used, due to inconsistencies in the manner, in which data from different channels (for example, from T1-weighted
and T2-weighted MR scans) is used by the physician. These shortcomings were the motivation for introducing automated techniques into the process of MSL segmentation and tracking in time. The suggested schemes are either semi-automated, involving a user at specific stages; or fully automated, based on a priori knowledge, such as an anatomical atlas [30].

1.2 Objectives of the proposed framework

The present work focuses on automatically performing two main tasks:

1. Detection and segmentation. The first task is to detect the regions pertaining to multiple sclerosis lesions in a sequence of two-dimensional MR scans of the brain. Each scan in the sequence was acquired at a different time point. The intervals between acquisitions are typically several weeks or several months. \(^1\) The framework is tuned to process T2-weighted MR scans. In the detection stage, an emphasis is placed on the detection of dynamic lesions, i.e., lesions that demonstrate a significant change in size (relapsing-remitting lesions). The task of lesion segmentation is closely related to detection. Following the detection of dynamic lesions, regions pertaining to these lesions are delineated.

2. Tracking in time. The final task is to track the segmented regions through the sequence, thus producing the temporal profile of the size of each dynamic lesion (a separate size vs. time plot for each lesion).

In Chapter 2 of this thesis an overview of related works in the field of segmentation and tracking in time of MS lesions is given. Chapter 3, Methodology, presents the mathematical tools, which are the basis of this work. The proposed framework with its different stages is introduced in Chapter 4. In Chapter 5, experimental results of each of the stages are presented. The manner in which the framework was validated is described in Chapter 6. In Chapter 7, several issues concerning the chosen methodology are discussed and suggestions for future work are given.

---

\(^1\) Registration and bias field correction are not part of the current framework; we assume that these operations were performed on the input sequence prior to the application of our framework and that the scans are, therefore, spatially aligned and have a similar dynamic range of intensities.
2 Related Works

Over the years, quite a number of algorithms for the segmentation of multiple sclerosis lesions in MR images have been described. Various mathematical methods are at the core of these algorithms, for instance, neural networks [10] and fuzzy connectivity principles [29]. In this chapter, several algorithms representing recent research in the field of statistical segmentation and tracking in time of multiple sclerosis lesions are discussed. Works dealing with segmentation include van Leemput et al. [30] and Warfield et al. [32]. The two frameworks model brain images using the voxels’ intensity feature. In [30] intensity-based classification is used with a stochastic model for normal brain images. Regions that are found to be model outliers are labeled as potential MS lesions. Warfield et al. segment the lesions in the WM regions after the gray matter (GM) structures have been identified and removed.

The tracking task is usually approached in one of two manners. The first option, represented here by the work of Kikinis et al., is to consider tracking as an extension of the three-dimensional segmentation problem to 4D [18]. Thus, the lesions are first segmented in 3D and then tracked through the serial data. The other approach attempts to model the temporal features of MSL voxels (typically, intensity) in the sequence. Gerig et al. examine the intensity vs. time profile of each voxel [8]. Voxels that show dynamic change of intensity are labeled as pertaining to MSL. Rey et al. have previously published ([26], later expanded in [24]) a method for detecting the evolution of lesions by comparing two consecutive images at a time based on Thirion’s rigid registration [27]. Methods of this kind are appropriate for short-term analysis, but are limited tools for retrospective diagnosis. The method reviewed here operates on a whole set, and considers both spatial and temporal coherences [25].

Sections 2.1 & 2.2 give an overview of segmentation and tracking algorithms, respectively. Model assumptions, chosen mathematical model and methods, implementation issues (initialization, materials), and evaluation methods are discussed. At the end of each section, the overview is summarized and the algorithms are comparatively appreciated.
2.1 Segmentation of MS lesions

2.1.1 Van Leemput et al. – lesions as outliers

The basic assumption made by van Leemput et al. in their work is that MSL voxels’ intensity could not be described well by a distribution function related to one of the normal brain tissues [30]. Furthermore, explicit MS lesions’ modeling was assumed difficult due to the lesions’ widely varying appearance and to the small amount of MSL voxels in some scans. The authors claimed, therefore, that MS lesions should be regarded as model outliers with respect to a statistical model of a normal brain.

Van Leemput et al. modeled the intensity of normal brain tissues by a statistical model of a mixture of Gaussian probability distribution functions (Gaussian Mixture Model – GMM) [31]. The design of the model assumed three normal brain tissues: WM, gray matter (GM), and cerebrospinal fluid (CSF). The authors added a fourth component to the model, relating to non-brain tissues and background noise. Model parameters are derived according to the maximum likelihood principle via the Expectation-Maximization (EM) algorithm. Van Leemput et al. noticed that the existence of outliers in the data, such as lesion-voxels, might hinder the EM algorithm, since it assumes that every voxel belongs to one of the model components. To overcome this, elements of robust statistics were introduced. A typicality-weight is computed for every voxel and updated every EM iteration. The weight reflects the typicality of each voxel with respect to the model. The EM algorithm was corrected, so that the contribution of each voxel to the calculation of the model’s likelihood depends on its typicality weight. Because the typicality weight of outliers is close to zero, they do not influence the EM algorithm. Furthermore, assessing the belief that a certain voxel is an outlier could be based on the typicality weight. The initial values for preliminary classification and the tissue type priors were derived from a brain atlas.

When the convergence of the EM algorithm is reached, the voxels are assigned to one of the tissues, except for the voxels with low typicality values. These voxels are labeled as outliers. Outliers do not necessarily pertain to MS lesions. Partial volume voxels, where a mixture of tissues is present, may appear as an outlier. Outliers exist in the CSF tissue, as well, comprising dura matter and blood vessel voxels. To avoid false detection of MS lesions two constraints are exercised. The first constraint, the intensity constraint, demands that an MSL voxel be brighter than the average GM
voxel. The other constraint, the contextual one, requires that most of the 3-D neighbors of the voxel belong to MS lesions or to the WM tissue. The contextual information is derived from the data itself, using a Markov random field (MRF). The MRF parameters reflect the cost of class transition between two neighboring voxels (within the image plane and over two adjoining scans). The outlier voxels that fulfill the two constraining conditions are labeled MS lesion voxels.

The algorithm was tested on 10 pairs of MR brain images, each taken from a different MS patient in an interval of approximately a month. Each scanning set contained 24 axial slices. The automatically computed total lesion load ($TLL_a$) was compared with a burden value derived from an expert’s manual segmentation ($TLL_e$). The comparison between the graphs of $TLL_a$ and $TLL_e$ showed that the change in lesion burden detected by the automatic framework follows the change observed by the expert. For example, where the expert indicated a continuously growth of lesion size, the burden computed by the algorithm increased in the same manner. Another validation criterion (the similarity index) showed, however, that the spatial overlap of the expert and framework segmentation results was relatively low. The authors attributed this to a limited incorporation of spatial considerations in their framework.

### 2.1.2 Warfield et al. – separating WM/MSL from GM structures

Warfield et al. approached the MSL segmentation problem from a different angle [32]. Unlike van Leemput et al., they assumed that there is an overlap between the intensity distribution functions of gray matter and lesions. Inter-scan and intra-scan intensity inhomogeneities further enhance the overlapping phenomenon. As a result, a pure statistical approach for the classification of brain tissues (like the one this group presented in [34]) is likely to mix GM voxels with MSL voxels. The authors claimed that MSL are detected by a human expert, although they look like GM, because they are located in WM regions. They, therefore, proposed to identify the other (non-WM) brain structures, and to perform lesion segmentation only on the remaining regions.

The algorithm begins with an EM-based statistical classification, which assigns the voxels to one of the clusters of a GMM. MRI scans of a normal volunteer were also classified and manually labeled to produce a reference anatomical atlas. The atlas was then elastically matched to the patient’s scans. Elastic matching of an atlas can directly segment brain structures that have a regular and consistent shape across many people, e.g., deep GM structures, and the cerebellum. However, the cortex structure is
not well recovered because of its shape variations. Statistical classification also fails to perform cortex segmentation, because it cannot differentiate between the cortex and other GM structures.

The authors suggested a cortex segmentation method based on region growing. In their framework, the cortex is modeled as a 3D-connected thick “blanket” crumpled over itself. The region growing seed is chosen at the outer edge of the cortex. The growing constraints include an anatomical constraint (based on the atlas), an intensity constraint (based on the EM classification), and a gradient constraint. The gradient constraint ensures that no regions that presented a significant rate of intensity change were included in the growing region, thus the constraint eliminated the inclusion of lesions. After all the GM structures in the brain had been segmented, the remaining regions were labeled as WM. The classification of voxels to normal WM and to lesions was done by a two-class minimum distance classifier.

The algorithm was applied to scans of 16 patients. Warfield et al. compared their results to results manually obtained by a group of five experts to validate the segmentation of GM structures. MSL segmentation results were qualitatively compared to the results of the pure statistical classification method of [34]. The improvement of segmentation results was also demonstrated by using a simulated scan. In a later work published by the same group (Wei et al., [33]), the authors added partial volume effect correction (PVEC) measures, taken from their tracking framework, which is discussed in the next section. Wei et al. reported that the addition of PVEC further improved the results.

2.1.3 Overall appreciation

The van Leemput et al. and Warfield et al. algorithms have several common features, including the basic modeling scheme and the use of an atlas. Both van Leemput et al. and Warfield et al. model the intensity feature of a brain MR image by a GMM. However, the two groups disagree on the proper approach to statistically modeling MS lesions. To summarize, it seems that Warfield et al. presented a more anatomically valid method, better incorporating case-specific data. Nevertheless, the automation, the self-training capabilities, and the modularity of van Leemput et al.’s scheme give an impression of an elegant engineering design. It would be of interest to examine the results of the van Leemput et al. algorithm, with an incorporation of spatial consideration at the classification stage.
2.2 Tracking of MS Lesions in Time

2.2.1 Kikinis et al. – extending the segmentation problem to 4D

Kikinis et al. approached the task of tracking as an extension of the three-dimensional segmentation problem to 4D [18]. They added to the model assumptions of the segmentation schemes² (see Section 2.1) the supposition that the location of lesions within the brain does not shift as time goes by. This last assumption allows a correspondence to be made between segmented regions in different scans, based on 4D connectivity procedures, such as the connected-components algorithm.

The intensities in the sequence are modeled in Kikinis et al.’s framework by a GMM of four clusters, corresponding to GM, WM, CSF and MSL. The model parameters are learned separately for each scan using the EM algorithm and, consequently, each scan is segmented individually. In order to provide tracking in time of lesions, a four-dimensional connected-component labeling is applied to the segmentation maps. Lesions that are too small or appear in only a few frames are eliminated from the list of detected lesions.

In order to initialize their modeling procedure, the authors used interactive calibration, in which the user chose typical voxels for each tissue in the scan. The researchers claimed that such a training procedure could be done on a limited number of scans and, from then on, the initial parameters could be constant, thus allowing a completely automatic procedure. Aside from model initialization, the interactive segmentation was the input for intracranial cavity (ICC) masking. This part of the framework included binarization based on the knowledge of which tissues comprise the ICC, 3D connected-component labeling and erosion/dilation.

The partial volume effect (PVE) was identified by Kikinis et al. as one of the main causes of classification errors in the output of their framework. The errors were particularly common for pixels on the brain surface that contain both brain tissue (skull bone) and CSF. Such pixels often have intensity similar to that of WM lesions. In order to deal with this problem, extensive postprocessing, based on morphologic operators and connectivity principles, was added.

² Their starting point was the pure statistical segmentation scheme in [34] that was also adopted by the same group for a segmentation framework (presented here as the Warfield et al. algorithm).
The framework was validated by showing that the estimated volumes of normal tissues were consistent through the sequence. The lesion detection results were not compared to an expert’s tracking.

2.2.2 Gerig et al. – analyzing temporal features of voxel intensity

The basic concept of Gerig et al.’s algorithm [8] is that the temporal change of voxel features has sufficient discrimination power for the identification and segmentation of MSL. Therefore, no segmentation procedures are applied to separate scans. For every voxel, a feature-vector is calculated based on its intensity temporal evolution, described by the intensity vs. time graph (the voxel’s temporal profile). A constant time interval between the scans is assumed. The zero-level of the intensity vs. time profile is the mean value of the profile. The feature-vectors consisted of:

1. The statistical variance of the voxel intensity over time: the intensity variability of normal tissues, attributed to white noise, is assumed considerably smaller than the statistical variance of lesion intensity.

2. The number of sign changes: a noisy but otherwise static signal will introduce frequent sign changes (relative to the mean intensity of the voxel). Voxels affected by multiple sclerosis will show large fluctuations, i.e., fewer sign-changes.

3. The maximum and minimum time gradient: MSL voxels demonstrate more abrupt changes of intensity.

The statistical scheme for establishing the likelihood of a sampled voxel to be an MSL voxel was based on the Dempster-Shafer (DS) theory [11]. The DS theory makes inferences from incomplete and uncertain knowledge, provided by different independent knowledge sources. In this case, the required inferences were "belongs to a MS lesion" or "does not belong”, and the knowledge sources were the feature-vectors. The output of the DS scheme is a likelihood parameter ranging from 0 to 1. To achieve a binary segmentation, an empirically chosen threshold was applied to the DS results.

As in the case of 3D segmentations, results of the algorithms are presented in the form of segmented images for a qualitative evaluation, and in the form of the total lesion load for quantitative analysis. The Gerig et al. algorithm was applied to 11 sets of serial MR images of multiple sclerosis patients: ten from the BIOMORPH project
and one from Brigham and Women’s Hospital, Boston. The results corresponding to real MR images were evaluated only qualitatively. Quantitative results were given for the segmentation of simulated images. The researchers replicated a 3-D set of a healthy volunteer to produce a serial MR data set. Eighteen lesions of varying size, contrast and lifetime (temporal extent) were generated using a non-linear diffusion process, and inserted in both white matter and gray matter regions. The choice of the lesion simulation model was done based on the observation that lesions were always diffuse, and most often showed a sharp increase and a smooth decrease of size and brightness. Gerig et al. reported detection of all the simulated lesions. The ratio between detected and implanted voxels, calculated in the scan in which the lesion reaches its peak size, was about 80% for most of the lesions. However, in other scans the rate was considerably worse, especially for very small or low-contrast lesions and at the borders of the lesion's lifetime. The estimate of the lifetime of the detected lesion voxels was often too short.

2.2.3 Rey et al. – fitting a model to the intensity vs. time profile
At the core of Rey et al.’s algorithm [25] is a parametric model for the intensity vs. time profile of lesion-voxels. The authors observed that the expanding stage is usually more rapid, while the shrinking stage is smoother. Based on this assumption they suggested an asymmetric “Gaussian” intensity curve as a model representing the intensity vs. time behavior of relapsing-remitting MSL. In this work only lesions that both appear and disappear, within the time covered by the sequence, are considered.

The fitting of the parametric model to the temporal intensity profile of the voxels in the sequence is calculated (in the least mean squares sense) in the first stage of the algorithm. Not all the model parameters could have been derived from the data itself, since the temporal resolution of the data sets (typically 1-6 weeks) is usually low compared with the pathological process time extent (~ 8 weeks). The authors chose to predefined the time extent of the intensity increase and decrease, based on an analysis of temporal profiles of many lesions. The voxels, the intensity profile of which fits well the parametric model, are labeled as potential lesion-voxels.

The aim of the second stage of Rey et al.’s algorithm is to extract statistically significant clusters of hyper-intense voxels based on the results of the first stage. The statistical significance depends on the cluster size. The null hypothesis of the model is
that no lesions exist in the image. Significant clusters are those with a size probability smaller than a threshold, under the null hypothesis.³

Rey et al. validated their algorithm based on visual observations only, as they claimed that no "ground truth" exists for their data set. The algorithm was applied to two T2-weighted serial data sets taken over a year: the first consisted of 24 scans and the other consisted of 10 scans. Visual inspection showed detection of all evolving lesions.

2.2.4 Overall appreciation
Comparing the tracking algorithms in terms of results is difficult, due to the different methods of result presentation and validation selected by the authors. The Rey et al. algorithm [25] has the advantage of incorporating spatial connectivity considerations that are absent in the work of Gerig et al [8].⁴ Gerig et al.’s method, on the other hand, seems to be more generic and robust in treating a variety of forms of MS lesion evolution (for example lesions appearing multiple times at the same location).

The most mature of the tracking methods appears to be the work presented by Kikinis et al. [18]. A distinct advantage of this method is the straightforward transition from segmentation maps to regions (via connected components). Another feature of interest in this work is the outlining of potential sources for detection errors (mainly false positives). A limitation of Kikinis et al.’s framework is its dependence on intensity-based segmentation, which may fail to delineate the full spatial extent of lesion-regions, because of the lesion-center/lesion-surround effect discussed in the present work (Section 3.1).

³ The second stage of the algorithm was not sufficiently explained in the article, and therefore, the description given here is limited.

⁴ In fact, this research group has recently published a spatio-temporal extension to their algorithm [35]. The algorithm is not reviewed here, because of limited mathematical elaboration in the article.
3 Methodology

The mathematical tools at the core of our framework are presented in this chapter. We start by outlining the assumptions and apriori knowledge that motivate the design of the framework. In the second section the Gaussian mixture model, which is employed here to model the image-sequence in a chosen feature-space, is described. The section includes also descriptions of the EM and K-means algorithms. The utilization of a GMM for statistical segmentation of an image-sequence is presented in the third section.

3.1 Model assumptions and apriori knowledge

In order to achieve an improved performance of a segmentation algorithm some assumptions regarding the input data are usually made. In the case of MS lesions, suppositions concerning a combination of lesion characteristics are often made ([18], [30], [32]). In this work, we use apriori knowledge concerning the intensity, shape and location characteristics of lesions, in the design of our framework.

![Figure 1](a) An example of a T2-weighted image of an MS patient. One of the lesions is circled. (b) Size evolution (in pixels) over time of dynamic (solid line) and static (dash line) lesions.

An example of a T2-weighted image with MSL is given in Figure 1(a) (one of the lesions is circled). The most distinct of the lesions’ visual features is their relatively high intensity. In addition, lesion regions are usually convex. Lesions are known to appear in the brain WM [4] and therefore it can be assumed that lesions are unlikely to be found along the brain’s vertical centerline (the line that goes through the corpus
callosum). Furthermore, high-intensity regions connected to the brain’s boundaries are usually not lesions [18].

The size of lesion regions can change considerably during the time interval between consecutive frames in the sequence. In such a case, we refer to the lesion as a *dynamic* lesion. Other lesions that undergo only small changes in size are *static* lesions. Figure 1(b) shows size vs. time plots typical of dynamic and static lesions. The proposed framework is aimed at detecting lesions showing dynamic behavior.

The underlying assumption of the suggested methodology is that the image intensities and their space-time distribution can be modeled by a GMM. The model is based on the following observations:

1. Gaussian behavior of the intensity feature – regarding this issue, we follow previous works that suggest modeling brain intensities by a GMM ([18], [31], [32], [34]).
2. Spatial convexity of the lesion regions
3. Distribution of the intensity feature in time – Figure 2 shows the intensity vs. time plots of several pixels taken from the center of the circled lesion in Figure 1(a), from its surround and from the background. The lesion-pixels present a clear relapsing-remitting pattern, while background-pixels demonstrate noisy fluctuations around their mean.

![Evolution over time of pixel intensity for pixels sampled from different regions of the image](image_url)

**Figure 2** Evolution over time of pixel intensity for pixels sampled from different regions of the image: lesion-center (*solid line*), lesion-surround (*dash line*), and non-lesion tissue (*dot line*).
Figure 2 conveys an important distinction between lesion-center and lesion-surround. The lesion-surround pixels present a pattern of behavior similar to that of the central pixels. However, their peak intensity is lower and the time interval in which they show a relapsing-remitting pattern is shorter. The conclusion that the central part of the lesion has different features from the lesion-surround corresponds to the fact that the lesion-surround often includes a mixture of lesion and edema or lesion and normal WM ([4], [18]). We employ this knowledge to provide a more accurate delineation of the lesion region, as will be described in Section 4.4.

### 3.2 Employing GMM for image representation

In this work, we utilize Gaussian mixture models to represent the input image-sequence in a feature-space of our choice. In order to construct the GMM for a given sequence of images, a transition is first made from the image domain to a selected feature-space. Following the feature extraction, each pixel is represented by a feature-vector and the image-sequence as a whole is represented by a collection of feature-vectors in the feature-space. Note that the dimensionality of the feature-vectors and feature-space is dependent on the chosen features and may be reduced or augmented in a modular way as needed.

The distribution of a random variable $X \in \mathbb{R}^d$ is a mixture of $K$ Gaussians, if its density function is:

$$f(x|\theta) = \sum_{j=1}^{K} \alpha_j \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} \exp\left\{ -\frac{1}{2} (x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j) \right\},$$

such that the parameter set $\theta = \{ \alpha_j, \mu_j, \Sigma_j \}_{j=1}^{K}$ consists of:

- $\alpha_j > 0$; $\sum_{j=1}^{K} \alpha_j = 1$;
- $\mu_j \in \mathbb{R}^d$; $\Sigma_j$ is a $d \times d$ positive definite matrix;

where $\mu_j$ and $\Sigma_j$ are the mean vector and covariance matrix of the $j$-th Gaussian of the model, respectively; $\alpha_j$ is the relative weight of Gaussian $j$; $d$ is the dimension of the chosen feature-space.
Learning a Gaussian mixture model from the feature data is, in essence, an unsupervised clustering task. The EM algorithm is used to determine the GMM parameters according to the maximum likelihood principle [6]. Given a set of feature-vectors $x_1, \ldots, x_n$, the maximum likelihood estimation of $\theta$ is:

$$\theta_{ML} = \arg \max_{\theta} f(x_1, \ldots, x_n \mid \theta).$$

(2)

The EM algorithm obtains $\theta_{ML}$ by iterating two steps: the Expectation step (E-step) and the Maximization step (M-step). Based on the current estimation of the parameter set, the E-step provides the probabilistic classification values, $w_{tj}$, which indicate the probabilistic affiliation of pixel $x_t$ to Gaussian $j$:

$$w_{tj} = \frac{\alpha_j f \left( x_t \mid \mu_j, \Sigma_j \right)}{\sum_{j=1}^{K} \alpha_j f \left( x_t \mid \mu_j, \Sigma_j \right)}.$$  

(3)

In the M-step, model parameters are re-estimated using E-step classification values:

\[
\begin{align*}
\hat{\alpha}_j &\leftarrow \frac{1}{n} \sum_{t=1}^{n} w_{tj} \\
\hat{\mu}_j &\leftarrow \frac{\sum_{t=1}^{n} w_{tj} x_t}{\sum_{t=1}^{n} w_{tj}} \\
\hat{\Sigma}_j &\leftarrow \frac{\sum_{t=1}^{n} w_{tj} (x_t - \hat{\mu}_j)(x_t - \hat{\mu}_j)^T}{\sum_{t=1}^{n} w_{tj}}.
\end{align*}
\]

(4)

The iterative updating process is repeated until a predefined condition is fulfilled. Possible conditions include a threshold applied to the increase of the log-likelihood and a limit on the number of iterations.

Initial values for the EM algorithm can be obtained by incorporating a priori knowledge (e.g., via an anatomical atlas [30]) or by extracting the values from the data itself. We utilize the K-means algorithm to perform the data-driven initialization [7]. The purpose of the K-means algorithm is to provide initial classification of the data into $K$ clusters in the chosen feature-space. The algorithm includes two iterating steps – classifying the data and determining the clusters’ means. Each pixel is assigned to the nearest cluster mean, in terms of the Euclidean distance in the feature-space. Based on the classification results, a new mean value is calculated for each cluster.
The K-means algorithm is initialized by assigning all the pixels to a single cluster. In order to proceed to the next iteration, two new mean values are generated based on the mean of the largest cluster. It should be noted that the generation of the two new values is done in a deterministic manner so as not to affect the reproducibility of the framework results. New mean values are generated until the required number of clusters (K) is reached. The algorithm terminates when the change in the K means becomes smaller than a predefined threshold. The parameters of the clusters are used to initialize the EM algorithm. Our implementation ensures that these initial clusters are statistically stable, i.e., that their covariance matrices are not singular.

It is common knowledge that the number of mixture components (or number of means) may be of great importance for the accurate representation of a given input. Ideally, K should have the value that best suits the natural number of clusters present in the input. However, in practice, it is difficult to determine this “natural” number when attempting to model a sequence in a multi-dimensional feature-space. Furthermore, even for a 1D intensity-based model K is usually larger than the number of tissues present in the images, due to the partial volume effect. An optimization criterion for K may be established based on a tradeoff between performance and the number of parameters used for describing the mixture distribution. The Minimum Description Length (MDL; [5]) is just such a criterion, commonly used for the selection of a model’s order in still-image processing [14]. It is possible to use the MDL criterion in video processing, as well [13].

The MDL suggests choosing a K that maximizes the expression:

$$\log L(\theta | x) - \frac{l_k}{2} \log n,$$

where $L(\theta | x)$ is the likelihood of the parameter set given the input data and $l_k$ is the number of free parameters needed for a model with K mixture components. In the case of a Gaussian mixture model with full covariance matrices, we have:

$$l_k = (k - 1) + kd + k\left(\frac{d(d + 1)}{2}\right).$$

When two models of different K values fit the data equally well, the simpler model, i.e., the lower K, is chosen. The MDL criterion is useful for choosing the model order where a reasonable range for the possible values of K can be established. In other cases, K has to be predefined heuristically or with the aid of a preprocessing stage.
Figure 3 Demonstration of the concept of space-time blobs. A 2sigma projection of the Gaussian in the x-y-t space is shown. The left blob corresponds to a region that is present mainly in a certain frame, while the right blob corresponds to a region that has a temporal extent over several frames and demonstrates a movement along the x-axis.

One can find in the relevant literature the use of a one-dimensional GMM (Equation 1, d=1) to model the intensity feature of brain images ([31], [34]). Typically, each Gaussian in such a model corresponds to a different brain tissue. In the current work, the GMM is utilized for intensity-based modeling in the 1D intensity space, as well as for global space-time representation in a 4D feature-space of intensity, spatial position (x,y) and time. In the global model, each Gaussian corresponds to what is from hereon termed a “space-time blob” [3]. Each blob has a spatial extent in x-y, which represents a certain region in an image, and a temporal extent, which represents the presence of the region across several frames. The blob represents, therefore, a space-time continuity of a region with certain intensity characteristics. Figure 3 demonstrates the concept of a space-time blob. The left image describes a blob that is mainly present in a certain frame, while the right image shows a blob that is present in several frames.
3.3 Employing GMM for image segmentation

Given the model parameters, a correspondence can be made between the coherent regions in the feature-space and homogeneous regions in the image plane. Denote

\[ f(x|\mu_j, \Sigma_j) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} \exp\left\{ -\frac{1}{2} (x-\mu_j)^T \Sigma_j^{-1} (x-\mu_j) \right\} \]  

(7)

as the aposteriori probability function of input sample \( x \), given the parameters of the \( j \)-th Gaussian component of the learned model. The probabilistic affiliation of pixel \( x \) with cluster (Gaussian) \( j \) is given by:

\[ p(\text{Label}(x) = j) = \frac{\alpha_j f(x|\mu_j, \Sigma_j)}{f(x|\theta)}. \]  

(8)

Segmentation maps are generated by assigning each pixel in the sequence to the most probable Gaussian cluster, i.e., to the component \( j \) of the model that maximizes the aposteriori probability:

\[ \text{Label}(x) = \arg\max_j \alpha_j f(x|\mu_j, \Sigma_j). \]  

(9)

The use of the GMM in image-sequence statistical representation and segmentation was described in this chapter in its generic form. The next chapter details the employment of the GMM for these purposes in the specific context of lesion detection and tracking-in-time.
4 The Proposed Framework

The proposed framework for MS lesions detection and tracking includes four main stages, as shown in Figure 4.\(^5\) The first stage is a preprocessing stage, in which the framework’s regions of interest (ROI) are extracted. In the second stage, global space-time modeling by a GMM is performed and space-time clusters are extracted. The dynamic lesions, i.e., the lesions that grow and shrink in size over time, are detected and segmented in the third stage. The final stage is aimed at producing a more complete delineation of the lesions via region merging.

![Figure 4 A schematic description of the framework proposed for the lesion detection & tracking (spatio-temporal segmentation) application.](image)

4.1 Preprocessing: extracting regions of interest

One of the most distinct visual features of MSL is their high intensity. We start therefore by focusing on the intensity feature and model the entire sequence by an intensity-based GMM (Equation 1; \(d=1\)). The model parameters are determined via the EM algorithm (Equations 3-4) initialized by the K-means algorithm. The model order, \(K\), is chosen within the range of 3 to 6, so as to reflect the number of different tissues present ([30], [32]). Each component of the model represents coherent intensity clusters in the feature domain. Following the model generation, a correspondence is made between the coherent regions in the feature space and homogeneous regions in the image plane. Each pixel of each frame is assigned to the most probable Gaussian cluster, i.e., to the component of the model that maximizes the aposteriori probability (Equation 9).

\(^5\) A preliminary three-stage version of the framework was introduced in [12].
Our goal in this stage is to extract high-intensity regions in the sequence. We define these regions as our ROI, since they include the pixels that belong to MS lesions. The segmented regions that correspond to the two highest-mean-intensity Gaussian clusters are extracted and are used as the input data for the following stages. We choose to extract two clusters and not just one in order to ensure that pixels pertaining to lesion-surround (typically of lower intensity, as shown in Section 2.1) are included in the ROI. Results of utilizing the intensity model are presented in Section 5.3.

A common preprocessing procedure, which is performed by all the algorithms that segment brain tissues, is the extraction of the intracranial cavity ([9], [19]). ICC extraction algorithms may be quite complex, if intended to deal with a large variety of brain scans taken from different angles and all possible values of the z-axis. In our case, however, we can assume the scans are axially oriented, and that they were taken at a z-axis point that shows the brain as a single connected region. These assumptions allow the simplification of the ICC extraction. The sequence undergoes binarization using a relatively low threshold (0.2), followed by a binary region growing procedure with a seed at the center point of the scan. The selected region is used as a mask, which is applied to the original sequence.

In addition to ICC extraction, two procedures are performed in this stage prior to the intensity model generation and following it, in order to eliminate potential false positives and unnecessary “noise” from the data. First, hyper-intense regions on the boundaries of the brain are removed. The pixels in these regions may be attributed to the CSF, the skull, or to a CSF-skull partial volume effect. These PVE artifacts are a well-known source of classification errors produced by an intensity-based segmentation algorithm, especially when the input consists of MS images [18]. The removal of boundary pixels is based on a region growing procedure (RG) applied to each frame separately (the design of this procedure draws from the cortex segmentation in [32] and the PVE correction in [18]):
1. The brain boundaries are detected by the Canny operator [2], applied to a binary image (all pixels with intensity significantly above 0 are set to 1).
2. A binary mask is created by thresholding. The threshold value is 75% of the maximal intensity value.
3. The RG seed is chosen as one of the pixels that is joint to the boundary detected in step 1 and the mask of step 2.
4. RG is performed on the binary map. The output of RG is the new binary mask.
5. If a pixel appears in at least one of the frames, all the pixels at the same position throughout the sequence are added to the binary mask. This rule ensures the consistency of the boundary removal, thus avoiding the boundary regions that are cancelled out only in part of the frames being erroneously regarded as dynamic lesions.
6. The binary mask is applied to the sequence (removal of boundaries).

The second preprocessing measure is the removal of small connected-regions from the sequence following the extraction of the ROI. Additional processing measures aimed at removing these potential false positives are not essential in our scheme. As long as there are no false negatives, the preliminary estimation of the ROI can be used as the basis for the identification of dynamic lesions out of the entire set of MS lesions. However, since a considerable number of regions that are definitely not lesions can be removed by applying simple criteria, several additional rules are applied to the extracted ROI. Regions that are smaller than a predefined threshold are assumed to be artifacts. Since lesions are typically not present along the vertical centerline of the brain, regions that cross the centerline are removed as well. The centerline is calculated as the average between the two extreme lines, on which there are non-zero pixels. Finally, if a region of significant size has a perimeter vs. size ratio higher than a predefined threshold, it is attributed to sulci (lace-like peripheral brain structures) and is removed from the set of ROI. The results of the preprocessing procedures are shown in Sections 5.2 and 5.3.

6 For our purpose, an accurate determination of the centerline (for example, according to symmetry consideration) is not required, since small deviations in its estimation do not effect the selection of regions for removal. The centerline is therefore calculated in a simplistic manner.
4.2 Global space-time modeling

A global space-time model is generated next. The extracted ROI are modeled by a GMM in a four-dimensional feature-space (d=4 in Equation 1), which includes intensity, spatial position (x,y) and time (the index of the frame within the sequence) as features. The feature space can also be three-dimensional, excluding time from the set of features. The decision whether or not to include time as a feature depends on the emphasis one wishes to put on the temporal separation of events in the sequence. When a 3D feature space is used, a lesion region, which appears, disappears and then appears again at the same spatial location, may be assigned to the same blob and not to two separate blobs, as one might expect. Including the time feature may solve this problem, but may also have the side effect of assigning a static region to several blobs, each covering a different temporal extent. The appearance of such a phenomenon is more likely to occur as the time interval between frames increases.

The EM algorithm is utilized to determine the GMM parameters. Two methods are suggested for initializing the EM algorithm, based on the preliminary intensity model. One method, from hereon referred to as the “connected-components version” (“CC version”), applies a 3D (space-time) connected-components procedure to the segmentation maps of the intensity-based model (this CC procedure is similar to the one in [18]). Thus, each space-time connected-component consists only of pixels that are assigned to a certain intensity-based cluster. The number of connected-components is the global model’s order. Each connected component is used to determine the initial parameters for a corresponding space-time Gaussian of the global GMM.

The second method, henceforth termed “K-means version” (“KM version”), initializes the EM algorithm via a data-driven K-means procedure. The input to the KM algorithm is a collection of feature-vectors extracted from the ROI. For this version, the model order is chosen empirically for a given input data, since criteria such as the MDL are not practical in this case (see Section 3.2).

Once generated, using one of the initialization methods, the global space-time model can next be utilized for the segmentation of the image sequence (Equations 7-9). Attention should be given to the fact that in the proposed methodology, a unique set of blobs is used for modeling the entire frame-sequence. The same blobs are also used in the segmentation of the sequence, where each pixel is probabilistically affiliated to one of the blobs in the set. Pixels in a specific frame, which are labeled as
pertaining to a certain blob, correspond to a segmented region. Moreover, because a space-time model is used and the sequence is processed as a single entity, each blob’s region of support is known from the segmentation maps in each frame. As a result, a certain region is marked by the same label in all the frames in which it is present. A by-product of the segmentation process is, therefore, the temporal tracking of regions. The unification of segmentation and tracking in our scheme is unique when compared to other works in the field ([8], [18], [24], [25]). The results of segmentation using a global model are given in Section 5.4.

The blob characteristics of each tracked region are stored for the lesion identification stage that follows.

4.3 Lesion detection

An important component of our framework is the detection of blobs pertaining to MSL (lesion-blobs), out of the larger collection of blobs. The global modeling stage provides a collection of space-time blobs, each parameterized by a mean feature-vector and a covariance matrix. An additional output of the modeling stage is the segmented sequence. In this stage, as well as in the merging stage that follows, we use context-dependent region-level rules, incorporating knowledge derived from clinical insights (Section 3.1) in order to detect lesion blobs:

1. A threshold is applied to the mean intensity of the blob. The mean intensity is used as a discriminating feature due to the hyper-intense appearance of MS lesions in T2-weighted MR images. We suggest deriving an adaptive threshold from the parameters of the intensity-based model: the average of the mean intensity value of the two most hyper-intense clusters.

2. We focus on the identification of relapsing-remitting lesions; an emphasis is thus placed on the detection of the dynamic behavior of blobs within the time-interval of the sequence. Blobs that demonstrate a significant change in size along the sequence are likely to pertain to MSL, since the structure of normal tissue remains relatively constant during the period that is typical to our case.7 We therefore apply a blob-level constraint on the

---

7 One should also bear in mind that MS lesions, as opposed to tumors, are not known to cause deformation of surrounding tissues [4].
blob’s size variability over time. The size variability is normalized by the mean size of the blob and is calculated along the entire sequence, that is, frames in which the blob does not exist are also taken into account in order to further emphasize the dynamic nature of blobs. An adaptive threshold is used, taken as the average of the size variability values of all the blobs. In order to enable the detection of lesions that are present in a very small number of frames, blobs of significant size with a short temporal extent are considered as well.

4.4 Region merging

Lesion-center and lesion-surround regions have different characteristics (Section 3.1). In an unsupervised modeling scheme, the entire region of the lesion is likely to be assigned to more than one blob. In order to produce an accurate delineation of the lesion region, we suggest a merging process to combine the lesion-center pixels with those of the lesion-surround. In this stage, it is assumed that all the lesion-center blobs are detected in the lesion-detection stage (Section 4.3). The merging procedure requires identification of lesion-surround blobs. For each lesion-center blob, a lesion-surround blob is detected according to the following criteria:

1. Mean spatial features (x,y) close, in terms of Euclidean distance, to the mean spatial features of the lesion-center. This criterion ensures the merging of spatially close regions.
2. Mean-intensity above a threshold, which ensures that no regions pertaining to background-blobs are merged. The intensity threshold is defined as the 2-sigma point below the mean intensity value of the second-highest cluster of the intensity-based model.
3. Temporal overlap with the lesion-center blob. This criterion ensures that lesion-surround blobs are affiliated only with their corresponding lesion-center blob, and not with another blob that appears in the same place, but at a different time-point. Temporally sequential lesion-center and lesion-surround blobs (there is an interval of one frame between the disappearance of one blob and the appearance of the other) are also considered overlapping.
Following the extraction of lesion-center and lesion-surround blob pairs, the segmented regions corresponding to these blobs are merged to produce an updated lesion region. An optional postprocessing stage can be applied to the results of the merging stage, in order to ensure that the marked lesions are connected regions of clinically significant size. If a lesion blob corresponds to a region that consists of more than one large connected sub-region, each sub-region will be labeled as a different region (splitting of segmented regions). Results of lesion detection and tracking in time before the region merging and following it are presented in Sections 5.5 and 5.6, respectively.

***

To summarize, we set two main objectives for our framework: detection of lesions and their tracking in time. Both objectives are achieved by the scheme described in this chapter. Lesion-blobs are detected out of the collection of blobs, which are provided by the global modeling of the ROI, using a set of context-based criteria. Once the lesion-blobs are detected, each of their corresponding lesion-regions is delineated in each frame (segmentation) and marked by the same label throughout the sequence (tracking in time).
5 Experimental Results

5.1 Data characteristics

The framework was applied to a T2-weighted serial data set received from Harvard Medical School and Brigham and Women’s Hospital in Boston. The images were acquired using a General Electrics Signa 1.5T with a double echo spin-echo pulse sequence (TR 3000 ms, TE 30/80 ms) and half-Fourier sampling (0.5 NEX). The acquisitions provided 54 slices with a 3-mm distance between slices, 3-mm slice thickness and a 256×256 resolution\(^8\), thus resulting in voxel dimensions of 0.48×0.48×3 mm\(^3\). In our experiment, a single slice from each set of scans was considered.

All 24 frames of the series, which were acquired over a period of approximately a year, are presented in Figure 5. Twelve of the most distinct lesions in the sequence are manually labeled (the labeling is used throughout this thesis to refer to the lesions in this sequence). The lesions have similar intensity features and are of a convex shape. Their visual appearance along the sequence differs greatly, however, due mainly to variability in their size vs. time profile. In Figure 5, lesions I-VIII are dynamic lesions, while lesions IX-XII are static lesions (see Section 3.1).

Within the group of dynamic lesions, there are additional differences in temporal behavior, which demonstrate the various cases tackled by MSL detection and tracking frameworks. Lesion II is present in the first frame, expands in later frames and then shrinks until it completely disappears by frame 13. To the contrary, lesion IV appears at the end of the first third of the sequence and from then on demonstrates only small size fluctuations. Lesion VIII (frame 24) is a minuscule plaque, probably corresponding to the very onset of a relapsing-remitting lesion. In contrast with lesion VIII, lesion V represents an abrupt appearance of a large lesion at the end of the sequence. Lesion V is interesting in another respect, since it appears in a location that is very close to that of lesion III, which is barely visible after frame 12.

In addition to differences in temporal behavior, it should be noted that the lesions also differ in their location within the brain. Although most of the lesions are deep in the WM tissue, several lesions (e.g., lesion II) are very close to the brain’s surface.

\(^8\) All the values of size and distance thresholds given in this thesis are tuned for this resolution.
Figure 5  The input sequence of a relapsing-remitting multiple sclerosis patient’s MR brain images. The index of each frame is given in parenthesis. Twelve lesions are labeled: lesions I-VIII are dynamic lesions; lesions IX-XII are static lesions.
5.2 Preprocessing

The data sequence we received underwent several preprocessing procedures [18], including sequence registration with respect to the first frame of the series and the correction of intensity inhomogeneities. Since the intracranial cavity was already extracted, the stage of ICC extraction in our framework did not affect the sequence and therefore its output is not given here.

We performed an additional preprocessing procedure, which removed high intensity regions on the boundaries of the brain area (Section 4.1). For the given input, the intensity threshold in step 2 of the scheme presented in Section 4.1 was 0.75. Figure 6 shows the result of the process for the first image of the sequence. As described in Section 4.1, the framework ensures that the locations of removed pixels are the same for all the images in the sequence, and, therefore the result for a certain frame represents well the output of this stage for the entire sequence. By comparing the original frame to the result, it is evident that a considerable amount of the hyper-intense pixels on the boundaries was removed, thus eliminating potential false positives from the sequence. The boundary elimination algorithm is, nevertheless, sensitive enough so as not to remove pixels pertaining to lesions that are close to the boundaries of the ICC (for example, lesion II).

Figure 6 Removal of high intensity regions from the boundaries of the brain. These regions pertain to the CSF and skull tissues and may cause false positives in the lesion detection stage. The first frame of the sequence is shown: original image (left) and image after the removal of boundaries (right).
5.3 ROI extraction based on an intensity model

An intensity-based model was learned via an EM learning process. We examined possible values of the model order, K, by applying the MDL criterion (Equation 5) to models of three to six Gaussians. The MDL reached its highest value for K=5, which was chosen as the model order. Figure 7 displays the intensity-based GMM with five components, plotted against the intensity distribution of the input data. The relatively small size of the lesion-regions is evident in the fact that the Gaussian of the highest mean is barely visible due to its low weight.

![Intensity histogram estimation by the intensity-based GMM. The intensity-based GMM (dashed line) set against the intensity distribution of the input sequence (solid line). The data is modeled by a GMM of five weighted Gaussians, shown by the dash-dot lines.](image)

Figure 7 Intensity histogram estimation by the intensity-based GMM. The intensity-based GMM (dashed line) set against the intensity distribution of the input sequence (solid line). The data is modeled by a GMM of five weighted Gaussians, shown by the dash-dot lines.

Figure 8 shows a set of probability maps (Equation 7) extracted for the five clusters of the intensity-based GMM. The separation into different tissue clusters is evident. For example, Gaussian 1 corresponds to the WM region, while Gaussian 5 represents the centers of the lesions and edge regions. The regions of interest consist, in this case, of the pixels affiliated with clusters 4 and 5 (the two highest-intensity clusters, which are the magenta and cyan regions in the segmented frame in Figure 8).
Figure 8 Probabilistic segmentation maps for five different clusters of an intensity-based GMM (K=5). The figure refers to the first frame of the sequence. Brightness indicates a higher probability of affiliation. The sixth image is the frame following probabilistic segmentation based on the segmentation maps. Each pixel was labeled by a different pseudo-color according to the cluster to which it was assigned.

Figure 9 presents the ROI extracted in frames 1, 8 and 21 of the sequence. When comparing the images in Figure 5 to those in Figure 9(a), it is noticeable that all clearly visible lesions are marked, i.e., the results hold no false negatives. The largest false positives are spatially concentrated on the central vertical axis. The following procedures were undertaken to remove false positives (Section 4.1):

1. Noise removal - an average size threshold of twenty pixels per frame or an overall size threshold of 150 pixels was applied.
2. Removal of regions that cross the vertical centerline.
3. Removal of sulci regions - connected components of an overall size of 150 pixels or more, the perimeter/size ratio of which exceeds 0.85%, were removed.

Figure 9(b) shows the improvement achieved by applying these criteria. The regions marked in Figure 9(b) are the ROI that were passed on as input to the space-time global modeling. The overall size of these ROI ranges from 500 to 1000 pixels per frame, which is 2.5%-5% of the total brain area. The ROI extraction stage achieved therefore a considerable reduction in the size of data to be modeled by the space-time model. Lesions were searched in the subsequent stages only within these regions.
Figure 9  ROI extraction. The regions, in which potential lesions are to be searched, are marked by a dark color. (a) Initial ROI (b) ROI following cleaning.

5.4 Global space-time model generation

The sequence was modeled in a 4D feature-space by a GMM, the parameters of which were determined by the EM algorithm (Section 4.2). The convergence of the EM algorithm is based on the log likelihood measure. A threshold of 1% was used. We have found experimentally that the above convergence methodology works well for our purposes.

Two methods are utilized for the transition from an intensity-based model to initial parameters for a space-time model. The CC version uses a 26-connectivity three-dimensional connected-components process. CC is applied separately to each of the clusters extracted in the preprocessing stage. In this case, CC was first applied to the pixels of the sequence that were assigned to cluster 4 (see Figure 8) and then to the pixels of cluster 5. A collection of twenty components was extracted. This collection was used to initialize a GMM of twenty Gaussians.
The KM version obtains the initial parameters for the EM algorithm from the data itself. Several values in the range of 20 to 100 were tested for the model order, K. For the given input sequence, we finally set K to be 75, since this value was found to provide good segmentation results in terms of the spatial separation of regions. The results of the KM version shown in this chapter correspond therefore to space-time modeling by a GMM of seventy-five Gaussians.

Figure 10(a) shows the ROI in frames [1 3 4 8 12 15 20 21] following segmentation using the space-time GMM (KM version). It should be noted that the input for this stage consists only of the ROI extracted in the preprocessing stage (Figure 9). The results are overlaid on the corresponding brain images for visualization purposes. Each region that corresponds to a different blob is marked by a different pseudo-color. The results of the CC version are presented in Figure 10(b).

In the context of our framework, in order for a segmentation to be considered successful, each region pertaining to an MS lesion has to be marked by a unique color, different from the labeling of its surround or other lesions. Overall, the separation of the lesions is satisfactory. Several interesting observations are summarized in Table 1.

Table 1 Summary of phenomena appearing in the segmentation results presented in Figure 10.

<table>
<thead>
<tr>
<th>Phenomenon</th>
<th>KM version</th>
<th>CC Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marking of spatially close lesions</td>
<td>Spatially close, but not necessarily touching, lesions may be marked by the same color (for example, lesions I and X in frame 1).</td>
<td>Two spatially overlapping lesion-regions are likely to be marked by the same color (lesions VI and IX in frame 21).</td>
</tr>
</tbody>
</table>
| The effect of time as a modeling feature| The effect discussed in Section 3.2 is evident here - Lesion II’s center is marked by one color in frame 1, by a different one in frames 3 and 4, and by yet another color in frame 8. | The spatial features are more dominant:
  - Advantage: lesion II is marked by the same color in all frames.
  - Disadvantage: similar regions appearing at different time points but at the same spatial location may be attributed to the same blob (notice lesion III in frame 8 and lesion V in frame 21). |
| Lesion-center and lesion-surround (see Sections 2.1 and 3.4) | The region of lesion II in frame 4 is marked by several colors, indicating that it corresponds to several blobs and not just to one. | Lesions III and V demonstrate the center/surround phenomenon in these results. |
Figure 10 ROI segmentation based on a space-time global model (4D feature-space): (a) KM version (b) CC version. Each pseudo-color corresponds to a different segmented region and to a different space-time blob. The segmented regions referred to in Table 1 are labeled. The numbering is based on manual correspondence to the originally defined lesion-regions (Figure 5).
5.5 Lesion detection

The detection scheme (Section 4.3), which is based on the mean intensity and size variance of the blob, was next applied to the results of the modeling stage. Nine of the 75 extracted blobs (KM version) were identified as potential dynamic lesion-blobs. Detection results for the KM version are shown in Figure 11. Labels correspond to the labels of Figure 5. In Figure 11, only detected lesions are labeled. Each detected lesion is marked by a different color, unique throughout the sequence. The marking of lesions by the same color both in a specific frame and through the sequence demonstrates our framework’s ability to unite segmentation and tracking in time, thus providing a single-stage spatio-temporal segmentation. Portions of the ROI that were not detected as pertaining to dynamic lesions are in black. For example, dynamic lesions were not detected in frame 15 and therefore all the ROI remain black.

When examining Figure 11, it should be noted that:

1. All the lesions that show a significant size change within the observed temporal interval (lesions I-VII) are marked in at least one frame.
2. Static lesions, such as the one labeled XI in Figure 5, are rightfully undetected and unmarked.
3. Lesions IX and XII, which we defined as static, are nevertheless identified as dynamic lesions in some frames. The reason for these few errors is probably the splitting of the lesions’ regions into several blobs along the time axis. One should keep in mind, however, that these regions are nonetheless lesions and these false positives are, therefore, errors only in terms of our definitions of static/dynamic. In other words, in the results shown in Figure 11 no region that is not sclerotic was detected as pertaining to a lesion.
4. Spatially close lesions, such as lesions I and X, are united, as was expected in light of the observation in Table 1.
5. Several visually detectable lesions were not identified in some of the frames (e.g., lesion IV). Furthermore, the delineation of some of the lesions does not seem accurate as the identified blobs failed to cover the full extent of the lesion (compare Figure 13(a) and (b)). These deficiencies are due to the lesion-center/lesion-blob effect and are treated in the following region merging stage.
Figure 11 Detection of dynamic lesions and their spatio-temporal segmentation. Detected lesion regions are color marked and labeled. Portions of the ROI that were not detected as pertaining to dynamic lesions keep their black marking from Figure 9.

5.6 Lesion merging

Results at the output of the merging algorithm applied to the results of Figure 11 are shown in Figure 12. Lesions in Figure 12 are marked using the same pseudo-colors as in Figure 11. The set of criteria used to determine whether a blob is a lesion-surround blob (Section 4.4) includes:

1. A distance of less than 4 pixels between the blob center and the center of a detected lesion blob;
2. Mean intensity higher than 0.6.
3. Temporal overlap with the closest center-lesion blob.

In comparison with Figure 11, Figure 12 shows an improvement in the detection of the dynamic MSL presence in the sequence. The merging process improved the results mainly in two respects. First, the delineation of lesions is better, as evident by the marking of lesion-surround regions that were black in Figure 11. For example, note lesion number II, shown enlarged in Figure 13. Other examples of better delineations are lesion IV in frame 8 and lesion VII in frame 21. The second contribution of the merging process is the tracking of detected lesions in the frames in which they were missed in Figure 11. For instance, lesion II is now marked in frame 1 and lesion IV is detected also in frames 12, 15 and 20.
An error that the merging process did not overcome is the marking of lesion IV by two different colors along the sequence. The meaning of this result is that the blobs corresponding to the lesion’s early presence were not merged with the blobs belonging to its appearance in the final frames of the sequence. The error may be attributed to the appearance of lesion VI in frame 21. Because of its proximity to lesion IV, lesion VI is unified with its neighbor to create a larger lesion with a spatial center that is different from the center of lesion IV in earlier frames. Thus, the merging algorithm did not identify these two seemingly different lesions as appropriate for merging.

The results following region merging for the CC version are shown in Figure 12(b). A comparison to Figure 12(a) shows that:

1. The strong spatial orientation of the CC method enables the separation of close lesions, such as lesions IV and VI, and the detection of lesion I without the error of identifying lesion X as dynamic.
2. Lesions III and V are still not separated.
3. The emphasis on spatial continuity may cause artifacts. For example, once lesion IX is erroneously merged with lesion VI in frame 21, the false positive detection propagates to all the frames.
4. The CC version’s results are likely to under-estimate the dynamic lesion burden, as detection tends to be more “conservative”, i.e., lesions are less easily detected as showing dynamic behavior (notice, for example, the misdetection of lesion VII).

A quantitative summary of the framework’s output is given Figure 14. A burden evolution plot for each detected lesion is shown. Each colored line in Figure 14 corresponds to an individual lesion candidate marked by the same color in Figure 12(a). The temporal characteristics of the lesions as seen in Figure 14 match the visual observations outlined in Section 5.1. Overall, the results of Figure 12 and Figure 14 demonstrate that the proposed scheme is effective in tackling the task of tracking highly dynamic MS lesions.
Figure 12 Region merging: (a) KM version; (b) CC version.

Figure 13 Zoom-in on lesion II in frame 8. (a) Original image; (b) Detection results; (c) Model adaptation results.
**Figure 14** Evolution over time of lesion size following the region merging stage. Color code follows Figure 12.
6 Validation

The quantification of the accuracy of automated segmentation (and tracking) is difficult in medical images, as “ground truth” segmentation for clinical data is absent. Manual segmentation by experts is the most common approach to addressing the validation problem and yet it suffers from intra-expert and inter-expert variability [32]. Automated algorithms, such as simulation tools [8], were sought in order to remove the variability introduced by experts.

We chose to validate our framework by combining the two aforementioned methods. The results presented in Chapter 5 are compared to the results of an expert’s manual segmentation. In addition, a simulation tool was developed, enabling us to perform robustness tests, in which the framework was given as input a large number of simulated sequences.

6.1 Comparison to an expert’s segmentation

The sequence described in Section 4.1 was given to an expert radiologist. The expert was asked to perform manual delineation of the lesions he identified in each frame. Based on an examination of the sequence as a whole, he also indicated which of the lesions are dynamic lesions. Figure 15 presents the manual segmentation of the lesions that the expert identified as dynamic in the example frames of Figure 12. A comparison of Figure 15 and Figure 12 shows that all six dynamic lesions marked by the expert were detected by our framework (the expert united the lesions that we labeled III and V). A zoom-in on the segmentation of two lesions as performed by the expert and by the framework is given in Figure 16. The delineation is very similar, with the expert marking slightly larger regions.
Figure 15 Manual segmentation of dynamic lesions by an expert radiologist.

Figure 16 Zoom-in on expert segmentation (left image) and framework segmentation (right image, taken from Figure 12). (a) Lesion II in frame 8; (b) lesion III in frame 8.

The expert’s estimation of the size evolution of the dynamic lesions (I-VII) is shown in Figure 17 alongside the framework’s results. The expert’s and framework’s size evolution estimates show the same pattern of disease progress and are therefore clinically compatible. The correlation factor between the plots is 0.93-1.0 (except for lesion I, for which the correlation factor is 0.8). In terms of absolute size estimation, the expert typically marks larger areas. The apparent under-segmentation of lesion-regions by the framework can be attributed to a stricter choice of ROI and more constraining detection rules. It seems that the expert tends to include in his segmentation also additional abnormal WM regions, not only lesions.
6.2 Validation via simulation

The design of our simulation tool follows the concept presented in [8], using a different lesion-model. A simulated image-sequence is generated using the following steps:

1. A normal brain scan (Figure 18(a)) is replicated to create the sequence in which the lesions will be implanted (we assume a constant time interval between frames). White Gaussian noise is added to the sequence in order to differentiate the frames in the sequence.
2. The lesion-center is created. In our case we model the lesion as an asymmetric 3D ellipsoid with the following parameters:

- Center in \([x \ y \ t]\); the center in \(t\) indicates the frame index in which the lesion will reach its peak size. This value can be within the time interval of the simulated sequence or outside it, thus allowing the simulation of lesions that only contract or only expand through the sequence.
- Radii in \(x\) and \(y\), determining the lesion’s spatial extent, i.e., its size.
- Time extent for the lesion’s expansion and for the lesion’s shrinking (these two values are not necessarily identical, hence the asymmetric ellipsoid). There is also a parameter indicating the time extent in which the lesion fluctuates around its peak size before shrinking.
- The lesion’s peak intensity.

3. We take the lesion-center/lesion-surround phenomenon into account by implanting a lesion, which is a superposition of two ellipsoids instead of just one. The lesion-surround ellipsoid is created as described in the previous step. It has a center close to that of the lesion-center and similar radii. Its time extent is typically shorter and its peak intensity is lower.

4. The superposition of the two ellipsoids is implanted in the simulated sequence according to the space-time center parameter. The implanting is done by replacing the pixels of the original images by the lesion pixels.

5. The image-sequence is filtered in order to smooth the boundaries of the implanted regions.

The simulation tool was used to generate 30 image-sequences, 15 frames each (256x256 pixels), containing overall 75 lesions with a large variety of parameters. Examples of images with implanted lesions can be seen in Figure 18(b). For each lesion the size profile (the number of implanted pixels in each frame) was stored. We chose to implant only dynamic lesions. The expected result is therefore the detection and tracking of all the simulated lesions. All 75 implanted lesions were indeed detected and tracked, without any false positives (a correlation coefficient above 0.98 between the implanted and detected size evolution graphs). An important sensitivity condition, which we discovered during the robustness tests, is that the framework would usually require an overall lesion size of approximately 80 pixels in order to detect the lesion.
Figure 18  Simulation of lesions. (a) The normal brain scan on which the simulated sequences is based.
(b) Several examples of simulated lesions, which were implanted into (a).

Table 2  Validation via the simulation tool. Comparison of the implanted number of pixels vs. the size of the lesion as tracked by the framework (KM version) along the sequence. Several example cases are shown: (a) an accurate detection of an expanding-contracting lesion; (b) an accurate detection of a contracting lesion; (c) over-segmentation; (d) detection of two implanted lesions as one.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Imp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>23</td>
<td>63</td>
<td>110</td>
<td>118</td>
<td>117</td>
<td>113</td>
<td>101</td>
<td>72</td>
<td>51</td>
<td>45</td>
<td>17</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Det.</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>15</td>
<td>53</td>
<td>98</td>
<td>100</td>
<td>99</td>
<td>87</td>
<td>60</td>
<td>45</td>
<td>33</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>Imp.</td>
<td>89</td>
<td>68</td>
<td>45</td>
<td>37</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Det.</td>
<td>77</td>
<td>54</td>
<td>37</td>
<td>30</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Imp.</td>
<td>33</td>
<td>73</td>
<td>77</td>
<td>65</td>
<td>49</td>
<td>33</td>
<td>25</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Det.</td>
<td>23</td>
<td>53</td>
<td>85</td>
<td>92</td>
<td>74</td>
<td>64</td>
<td>50</td>
<td>43</td>
<td>35</td>
<td>20</td>
<td>22</td>
<td>19</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>D</td>
<td>Imp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>33</td>
<td>83</td>
<td>120</td>
<td>115</td>
<td>86</td>
<td>65</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Imp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>35</td>
<td>57</td>
<td>73</td>
<td>71</td>
<td>57</td>
<td>57</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Det.</td>
<td>17</td>
<td>19</td>
<td>18</td>
<td>38</td>
<td>81</td>
<td>120</td>
<td>119</td>
<td>90</td>
<td>95</td>
<td>85</td>
<td>74</td>
<td>72</td>
<td>63</td>
<td>60</td>
</tr>
</tbody>
</table>
**Figure 19** Validation via the simulation tool. The lesions’ size evolution according to the framework (dash line) is compared to the “ground truth” (solid line). (a)-(d) correspond to the cases presented in Table 2. In (d) there are two solid lines because, in this case, the framework merged two different lesions into one.

**Figure 20** Images relating to cases C (left) and D (right) in Table 2. Case C – inclusion of non-lesion regions; Case D – frames showing the merging of two lesions into one.
A total of 63% of the detected simulated lesions were segmented and tracked with no artifacts. The segmentation and tracking results for four simulated lesions are shown in Table 2 along with the number of implanted pixels, as recorded by the simulation tool ("ground truth"). Case A (an expanding-contrasting lesion) and case B (a shrinking lesion) are examples of a successful detection. The framework’s results tend to under-estimate the size of the lesion, due to the inclusion of relatively low intensity pixels in the implanted lesions. Nevertheless, the plots of size vs. time for these cases (Figure 19(a-b)) show that the framework’s results are compatible in their pattern to the simulation tool’s “ground truth”.

In 20% of the cases there is a constant bias in the results of the framework, which is caused by over-segmentation. The framework merges a close non-lesion region with the detected lesion. Case C in Table 2 is such a case, as can also be seen in Figure 19(c). Figure 20(a) shows that in this case a small amount of high intensity pixels on the boundaries of the brain were erroneously included as part of the ROI. Consequently, these pixels were merged with a lesion that was implanted close to them.

In 17% of the cases, two different lesions were merged into one. Case D in Table 2 and its corresponding plot of Figure 19(d) demonstrate that the detected lesion is in fact a superposition of two implanted lesions. The images in Figure 20(b) show that these two lesions are spatially close. The merging of two close lesions was already observed in Section 4.6 and is more typical of the KM version, since in this scheme the fitting of the blobs is less spatially accurate than in the CC version scheme.
7 Discussion

The purpose of the current work was to apply a statistical space-time modeling scheme to the specific medical image processing problem of detecting and tracking-in-time of multiple sclerosis lesions. The generic modeling scheme was adapted so as to better treat the input data of sequences of two-dimensional MR images of the brain. A lesion detection criteria set, incorporating context-based considerations, was implemented and tested. The performance of the proposed framework was demonstrated using a well-known benchmark sequence of 24 2D frames. Strong correlation was achieved between the automatically obtained segmentation results and the expert’s manual segmentation. Robustness tests via a simulation tool further confirmed the validity of the proposed methodology.

In this chapter, the proposed framework is judged in light of other works in the field of MSL segmentation and tracking, and conceptual issues are discussed. At the end of the chapter, topics for future work are suggested.

7.1 Modeling multiple sclerosis lesions

An important assumption of the suggested methodology is that the image intensities and their space-time distribution can be modeled by a GMM. Modeling the intensity feature of a brain MR image by a GMM, where each Gaussian typically relates to one of the brain tissues, is an acceptable approach in the design of normal brain segmentation frameworks ([31], [34]). However, researchers disagree as to which is the proper approach for statistically modeling MS lesions, the visual characteristics of which are more varied than those of normal tissues. Furthermore, lesions are often very small structures, thus providing little data to the modeling scheme. Taking into account these phenomena, van Leemput et al. [30] suggested treating lesion-voxels as outliers to a normal brain model. Our intensity-based model does not follow [30], but rather Warfield et al. [32], who explicitly assigned one of the clusters in their model to fit MSL, contrary to the outliers approach. We believe that treating the sequence as a single entity (and not frame-by-frame like in other works [18]) enables us to gather enough data regarding lesion-regions and to overcome the problem of statistically modeling them. The results of our initial intensity-based modeling show that a satisfactory separation of lesion-regions is possible for the given data (Section 5.3).
In all the above-mentioned segmentation frameworks, the only modeled feature is the intensity. In such models, voxels may be classified according to tissue type independently of their neighbors’ classification, since spatial location within the image is not taken into account. In [30], an attempt to overcome this limitation was made by adding spatial considerations in the form of a contextual constraint at the end of each learning iteration. The use of an atlas at the initialization stage is another method to bolster the spatial accuracy of the segmentation results ([30], [32], [33]). Nevertheless, the best method for achieving improved spatial delineations is, probably, to add voxel location to the set of modeled features, as we do in the global space-time model of our framework.

Working in the context of tracking-in-time, Rey et al. [25] and Gerig et al. [8] both agreed that spatial considerations should be incorporated into the algorithm: in [25] a spatio-temporal extension to the framework was suggested as future work, and in [8] spatial considerations were introduced via post-processing. The spatio-temporal nature of the work of Kikinis et al. relates to the use of a 4D connected-components process, applied to intensity-based segmentation maps. However, this implementation is limited by its dependency on the intensity-based model, which might fail to capture the full spatial extent of lesion-regions. The importance of a truly unified spatio-temporal model in tracking applications has not yet been thoroughly studied and is discussed in this thesis for the first time. Our global model incorporates both space and time features (in addition to intensity). Following the lesion detection stage, segmentation and tracking of lesions are performed simultaneously in a unified stage of spatio-temporal segmentation, and not separately as in other works.

We suggest two methods for initializing the EM algorithm – the CC version and the KM version. The manner in which the time feature is handled is one of the main respects in which these methods differ. Time is unlike the other features in the space-time model in that it does not have a strict range and therefore it is not clear how the time-feature values should be set. We chose one option, which is to give the pixels in each frame a time-value equal to the frame index (normalized by the sequence length). The CC version is less sensitive to time values, since it performs initial clustering based on the intensity model segmentation, considering time only in terms of the order of frames within the sequence. The KM version, on the other hand, learns the initial parameters by clustering the data itself in the chosen feature-space. Since the implementation of our K-means algorithm is based on Euclidean distances,
normalizing one of the features changes the distance between feature-vectors and therefore may change the resulting clusters. The manner, in which this feature of the K-means algorithm could be employed to emphasize different sequence-features, and as a result different characteristics of lesion-pixels, is yet to be thoroughly investigated. Replacing the distance metrics in the K-means algorithm could be another issue worthy of examination.

7.2 Determining the model order

The initialization of probabilistic frameworks that model brain images is often done via the use of an anatomical atlas ([31], [32]). An atlas may be very useful when a model of normal brain tissues is required. However, the data it provides regarding pathological abnormalities is more limited. More importantly, incorporating an anatomical atlas demands addressing the non-trivial task of registering it to the given patient scan. In our framework, no registration of an anatomical atlas is required, since the modeling stages of the framework are data-driven. A disadvantage of data-driven modeling is that its complexity increases with the increase in data size, or in other words, with the increase in the number of pixels in the image-sequence. The main difficulty in generating such a model is establishing the model's order, K. Determining K is a crucial element in the modeling scheme, and may effect the capability to accurately delineate regions of interest and distinguish between static and dynamic regions.

Ideally, K should reflect the number of intensity-homogeneous space-time connected-regions in the input sequence. For large images and long image-sequences, the value of the optimal K, if such a value can be set at all, may be very high. The learning of high order models is computationally expensive and requires extensive stability-guarantee measures in order to avoid the appearance of very small clusters in the model. The initial stage of the framework, in which unwanted data is removed and only the regions of interest are extracted, enables space-time modeling by a GMM of a reasonable order.

However, establishing K for the space-time model is not simple, even after the extraction of ROI. The possible range of K may be very wide, as well as input-dependent, thus hindering the use of the MDL criterion. One option is to predefined the model order based on empirically acquired knowledge, as done here for the KM version, which performs data-driven initialization. Another option is to use a
connected-components procedure for model initialization. When the CC version is used, no predefinition is required, since each connected-component initializes a different Gaussian in the model. It should be noted that the CC version generated a model with half the number of Gaussians than the KM version model’s, which may explain why the KM version results are prone to over-splitting of regions. The KM version, on the other hand, is less dependent on the intensity-based modeling than the CC version and it allows for a more direct incorporation of other features.

7.3 The detection scheme

In the detection stage of our framework, we concentrate on identifying relapsing-remitting lesions (dynamic lesions). While current schemes detect lesion-voxels based on voxel-level rules ([8], [25]), in this work we propose a unique set of criteria that incorporate region-based (blob) features. Region-based features, such as the size-profile, are attainable in a direct manner from our global model, since each blob in the model corresponds to a space-time region in the sequence. This region-level analysis can be regarded as compatible with the definition of a lesion as a spatially connected entity with a unique space-time profile.

We examined several criteria for lesion detection based on the parameters of the space-time blobs. Rules regarding the intensity change of lesion-voxels (like those presented in [8]) cannot be used in our model, since our space-time blobs correspond to intensity-homogeneous regions. We considered adding the intensity delta as one of the modeling features, but ruled out this option because of the feature’s non-Gaussian behavior. Among the parameters provided by the global modeling stage, we found that the most characteristic of lesion-regions is their mean intensity. The covariance matrix does not seem to hold additional parameters effective in distinguishing lesion-regions. This can be attributed to the fact that the lesion’s “movement” does not translate into changes in the spatial center, but rather into expanding and contrasting with relation to a constant center. Since the term dynamic lesion refers by definition to change in size, we took advantage of the data available via the global segmentation maps and added a second criterion relating to size variability. The results of the framework show that criteria applied to mean intensity and size variability are sufficiently discriminating.
The region merging stage is an important contributor to our framework’s ability to tackle the context-specific problems presented by the medical data. The design of our global model facilitates the process of region merging. Adding a merging stage improves the framework’s results, as it enables the detection of lesions at extreme points of their temporal existence and the identification of suspected edema regions around lesions. A possible artifact of this stage may be the labeling of two close lesions as pertaining to a single lesion and the merging of small non-lesion regions. If such artifacts do appear, a relatively simple postprocessing can be applied to the results of the merging stage, in order to ensure that the marked lesions are connected regions of clinically significant size.

An emphasis was placed in the design of the framework on avoiding “hard” predefined thresholds by replacing them instead with adaptive thresholds. However, the validity of these thresholding mechanisms is still in question, since the framework was applied only to a very limited number of real images. It is reasonable to assume that given a larger input data set, the thresholds could be fine-tuned for a specific modeling scheme to further improve results.

7.4 The validation problem

One of the main problems concerning MSL segmentation and tracking-in-time frameworks is the lack of an accuracy metric, known as the “ground truth” problem. This limitation is probably the reason that none of the related works in the field included, as far as we know, a quantitative comparison to other frameworks. In addition, the absence of an accuracy measure raises a difficulty in the stage of framework validation. Different approaches were offered in the relevant literature for solving this problem. A common means for quantitative validation is the comparison to an expert’s manual segmentation [30]. We employed this approach to demonstrate the clinical acceptability of our framework. Through the interaction with the radiologist, the manual process revealed itself, however, as extremely prone to inter-expert and intra-expert variability, thus placing a question mark on the validity of this method for quantitative validation.

In light of the variability in expert segmentation results and the small number of available datasets, we developed a simulation tool that enabled us to test our framework’s robustness. In the context of the current work, we found the simulation tool satisfactory. In future work, this tool may be used to perform additional tests,
examining issues such as sensitivity to changes in framework parameters and effects of input misregistration and noise. However, solving the validation problem via a simulation immediately raises another problem, which is how to validate the simulation tool itself. When considering the limited clinical knowledge of the pathological processes of the MS disease, the task of establishing whether the simulated images truly imitate the real ones seems far from trivial.

7.5 Future work

The completion of two tasks is essential for the clinical validity of the framework: expanding the framework so it would be able to treat volumetric data, and investigating larger datasets for a more complete experimental study and clinical evaluation. The modularity of the GMM will be instrumental in adding another spatial feature to the modeling scheme. Handling three-dimensional regions, instead of two-dimensional ones, will require generalizing the detection and merging criteria to suit a 5D framework.

Obtaining datasets of MS patients taken over time might prove difficult outside the context of a study specifically dedicated to this purpose, since close follow-ups of the disease’s progress are not common procedure in all clinical centers treating MS patients. One should also take into account that available data is unlikely to be preprocessed. This suggests incorporating registration and bias field correction stages to our framework. Once the 5D framework is constructed and validated, it could be adapted to deal with similar medical image processing problems, such as the detection and tracking-in-time of tumors or Alzheimer lesions.
Bibliography


