Computerized Medical Imaging and Graphics 38 (2014) 337–347

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Evaluation of supervised methods for the classification of major tissues and subcortical structures in multispectral brain magnetic resonance images

Loredana Murino[a,*], Donatella Granata[a], Maria Francesca Carfora[a], S. Easter Selvan[c], Bruno Alfano[b], Umberto Amato[a], Michele Larobina[b]

[a] Istituto per le Applicazioni del Calcolo 'Mauro Picone' CNR, Napoli, Italy
[b] Istituto di Biostrutture e Bioimmagini CNR, Napoli, Italy
[c] Université catholique de Louvain, Department of Mathematical Engineering, Louvain-la-Neuve, Belgium

ARTICLE INFO

Article history:
Received 30 April 2013
Received in revised form 6 March 2014
Accepted 11 March 2014

Keywords:
Brain
Subcortical structures
Segmentation
Discriminant Analysis
Atlas-based
Denoising

ABSTRACT

This work investigates the capability of supervised classification methods in detecting both major tissues and subcortical structures using multispectral brain magnetic resonance images. First, by means of a realistic digital brain phantom, we investigated the classification performance of various Discriminant Analysis methods, K-Nearest Neighbor and Support Vector Machine. Then, using phantom and real data, we quantitatively assessed the benefits of integrating anatomical information in the classification, in the form of voxels coordinates as additional features to the intensities or tissue probabilistic atlases as priors. In addition we tested the effect of spatial correlations between neighboring voxels and image denoising. For each brain tissue we measured the classification performance in terms of global agreement percentage, false positive and false negative rates and kappa coefficient. The effectiveness of integrating spatial information or a tissue probabilistic atlas has been demonstrated for the aim of accurately classifying brain magnetic resonance images.

1. Introduction

Tissue classification is an important and challenging task in the processing and analysis of brain magnetic resonance (MR) images. In the last two decades, a large number of studies have been oriented in finding volume changes of the major brain tissues – gray matter, white matter and cerebrospinal fluid – in the aging process or in the occurrence and progression of specific diseases [1]. Currently, there is a growing interest in understanding the involvement and the possible morphological changes of minor brain structures both in normal aging processes and in the presence of neurodegenerative disorders [2–4].

Techniques for tissue classification can be broadly divided into supervised and unsupervised depending on the use or not of a training data set. Moreover, some techniques are based on one MR parameter only, commonly a high resolution T1-weighted image, others are multispectral with the aim of combining T1, T2 and PD-weighted information or other image types such as inversion recovery and fluid attenuation inversion recovery [5].

The classification of brain minor structures presents a higher degree of difficulty due to a weaker contrast of these structures with adjacent tissues, which limits intensity-based recognition. Various approaches have been followed to segment these structures. Among the most widely used and freely available methods we mention FreeSurfer and FSL-First. FreeSurfer is a suite of tool for the analysis of neuroimaging data, and one of the first software to provide the segmentation of subcortical structures. The algorithm is essentially based on Markov Random Fields, a statistical methodology used to model the local spatial relationships between voxels as described in detail in [6–8]. FSL-First is a tool for subcortical segmentation, part of the FSL software library [9], that uses a statistical model of object shape in combination with intensity information [10].

Discriminant Analysis is probably the most consolidated statistical methodology for classification, based on an approximation of the probability density function of the signal and on the Bayes decision rule [11]. Among nonstatistical methodologies, the K-Nearest

* Corresponding author at: Istituto per le Applicazioni del Calcolo CNR, Via Pietro Castellino, 111, 80131 Napoli, Italy. Tel.: +39 0816132381; fax: +39 0816132597.
E-mail address: l.murino@na.iac.cnr.it (L. Murino).

http://dx.doi.org/10.1016/j.compmedimag.2014.03.003
0895-6111/

Neighbor (KNN), the Artificial Neural Network [12,13] and the Support Vector Machine (SVM) [14] have been effectively used for segmentation purposes.

Atlas-based approaches exploit anatomical information by means of atlases to be matched to the study one wishes to segment. For a thorough update review, the readers may refer to Cabezas et al. [15] who elaborately discuss a probabilistic atlas-based segmentation that shares some similarities with our scheme. A segmentation framework based on the combined use of a statistical and a topological brain atlas has been proposed by Bazin and Pham [16].

Other methodologies aimed to exploit the spatial voxel correlation have been proposed to be used in conjunction with pattern recognition techniques. For instance, Anbeek et al. [17] added voxel coordinates as an extra feature to the intensity model, which in turn was applied to a KNN classifier. Cutillo and Amato [18] proposed some techniques akin to the Discriminant Analysis to account for the local spatial correlation among voxels.

In this context, for the simultaneous classification of major tissues and subcortical structures, we focused our attention on supervised methods suitable to work with multiple input channels represented by MR images acquired with different contrasts. In the first part of this work we attempted to answer the following question: "Is it possible to classify brain tissues and subcortical brain structures on the basis of image contrast using a trained statistically based methodology and multispectral data?" By employing a gold-standard phantom data, we compared several methods belonging to the family of Discriminant Analysis techniques apart from two nonstatistical methods, i.e., KNN and SVM. In the second part of the paper, the performance of the classification methods has been quantitatively assessed when the classification scheme incorporates spatial information about tissues: x, y, z coordinates of image voxels have been considered as spatial features in addition to the intensity features for Discriminant Analysis methods and KNN. As an alternative, the use of a probabilistic atlas for all the intracranial brain tissues has been considered as prior spatial information for Discriminant Analysis methods. Results on synthetic data, and a comparison with FSL-First are reported and discussed. For the quantitative evaluation of the various classification methods the present work made use of a digital brain phantom as our gold standard. We considered a recently developed phantom, claimed to simulate the actual brain with a larger number of tissues and an increased realism compared to currently available alternatives [19]. To test robustness of the considered approaches on real data, a set of ten real studies has been included into the evaluation.

The paper is organized as follows. Section 2 describes the synthetic and real data used in this study, and the proposed methods for the MRI data classification. Section 3 presents the experiments conducted to evaluate the performance of the classification methods with and without the inclusion of spatial information. Section 4 contains a discussion of the results and Section 5 the conclusions.

2. Materials and methods

2.1. MRI data

A realistic digital brain phantom [19], available at http://lab.ibb.cnr.it/, is used to evaluate and measure the performance of the classification methods. The following 17 classes are defined in the phantom as listed in Table 1: 11 intracranial tissues (Gray Matter, White Matter, Cerebro Spinal Fluid, Pallidus, Putamen, Thalamus, Caudate Nucleus, Red Nucleus, Dentate Nucleus, Substantia Nigra and Intracranial Connective), and five extracranial ones (Fat, Muscle, Vitreous Humor, Extracranial Connective and Extracranial Fluid); a further class (LowPD) comprising intra- and extra-cranial voxels characterized by a low proton density value and not otherwise assigned to any other class is included. The model contains $256 \times 256 \times 150$ near-isotropic 0.9375 mm $\times$ 0.9375 mm $\times$ 1 mm voxels; each voxel of the model is labeled according to its assignment to a particular class. The corresponding simulated signals are provided in the form of T_{1w} (510/15ms TR/TE) and P_{Dw}–T_{2w} (1867/15-90ms TR/TE) axial slices with a slice thickness selectable between 2 and 5 mm for a conventional spin-echo sequence, and in 1 mm thick axial slices (TR = 9.9 ms, flip angle = 10°, TE = 3.5 ms) for a 3D T_{1w} FFE sequence.

Table 1
List of tissues included in the digital brain phantom. Their abbreviation and the 1 mm resolution volumes are also shown.

Tissue	Abbreviation	Volume (cc)
Gray Matter	GM	854.86
White Matter	WM	572.62
Cerebro Spinal Fluid	CSF	182.45
Pallidus	PAL	3.99
Putamen	PUT	9.73
Thalamus	THA	13.95
Caudate Nucleus	CN	10.35
Substantia Nigra	SN	1.19
Red Nucleus	RN	0.66
Dentate Nucleus	DN	1.67
Intracranial Connective	ICC	20.04
Fat	FAT	409.18
Muscle	MUS	552.09
Vitreous Humor	VH	14.25
Extracranial Connective	ECC	54.20
Extracranial Fluid	ECF	16.54
LowPD	LPD	825.68
All		3543.45

For the first experiment two multispectral datasets with 4 mm and 2 mm slice thickness have been generated from the phantom; 2 mm is the minimum slice thickness for the 2D multislice spin-echo simulation. While the original model is composed of 150 slices, the MR signals include 37 slices of thickness 4 mm and 75 slices of thickness 2 mm. Then, starting from the original model, we reassigned voxel labels by grouping the slices 2 by 2 and 4 by 4, respectively. In the case of non-pure 2 and 4 mm resolution voxels (i.e., voxels composed of 1 mm resolution voxels corresponding to different tissues), their labels have been assigned by a voting strategy based on the highest label occurrence for the group of selected slices. In case of a tie among two or more tissues, the voting strategy is applied to a neighborhood of the voxel sized 3×3 in the same axial slice. The voxels and their tissue labels form the ground-truth.

The second experiment considers 3D T_{1w} FFE and P_{Dw}–T_{2w} spin-echo signals. As the FFE volume includes 150 slices, the P_{Dw}–T_{2w} volume has been resliced to obtain 150 slices using a linear interpolation.

The digital brain phantom model serves as a reference to evaluate the performance of the classification methods as well as to derive the training dataset needed by supervised methods. To avoid the estimation bias, which may arise when the training and test datasets coincide, N voxels were randomly selected from the digital model as a training dataset and the actual classification was performed on the remaining voxels. We chose N = 100,000, which is 10% of the full volume at 4 mm resolution. Each tissue was guaranteed to have the same percentage of its voxels in the training set. The value of N was chosen through some experiments based on the classification of the digital phantom: note that no significant variations were observed for higher values of N. In order to avoid the influence of a particular realization of noise, several experiments with the same N were conducted with different realizations of noise; however, the results did not vary significantly. Incidentally, using a fraction of the total volume as the training dataset speeds up the training process of the classification methods such as the SVM, where the computational cost of the training step becomes