Medical Image Analysis 18 (2014) 542–554

Contents lists available at ScienceDirect

Medical Image Analysis

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

Individualized statistical learning from medical image databases: Application to identification of brain lesions

Guray Erus [a,*], Evangelia I. Zacharaki [b], Christos Davatzikos [a]

[a] Center for Biomedical Image Computing and Analytics (CBICA), University of Pennsylvania, Philadelphia, PA, USA
[b] Department of Medical Physics, School of Medicine, University of Patras, Patras, Greece

ARTICLE INFO

Article history:
Received 16 August 2012
Received in revised form 27 November 2013
Accepted 8 February 2014
Available online 17 February 2014

Keywords:
Abnormality segmentation
Statistical learning
PCA
Brain MRI

ABSTRACT

This paper presents a method for capturing statistical variation of normal imaging phenotypes, with emphasis on brain structure. The method aims to estimate the statistical variation of a normative set of images from healthy individuals, and identify abnormalities as deviations from normality. A direct estimation of the statistical variation of the entire volumetric image is challenged by the high-dimensionality of images relative to smaller sample sizes. To overcome this limitation, we iteratively sample a large number of lower dimensional subspaces that capture image characteristics ranging from fine and localized to coarser and more global. Within each subspace, a "target-specific" feature selection strategy is applied to further reduce the dimensionality, by considering only imaging characteristics present in a test subject's images. Marginal probability density functions of selected features are estimated through PCA models, in conjunction with an "estimability" criterion that limits the dimensionality of estimated probability densities according to available sample size and underlying anatomy variation. A test sample is iteratively projected to the subspaces of these marginals as determined by PCA models, and its trajectory delineates potential abnormalities. The method is applied to segmentation of various brain lesion types, and to simulated data on which superiority of the iterative method over straight PCA is demonstrated.

1. Introduction

Voxel-based morphometry (VBM) type of analyses are being increasingly adopted for characterizing neuroanatomical differences between brain images (Ashburner and Friston, 2000; Mechelli et al., 2005). VBM has emerged as an alternative to region of interest (ROI) based approaches. Different brains are compared on voxel-by-voxel basis after being spatially normalized to a common template space. Voxel-wise statistical tests are performed on each individual voxel for a normal population, and voxels that differ from the normal population are flagged and grouped into clusters reflecting pathology, over the entire group, albeit not necessarily at the individual level. A fundamental limitation of this kind of approaches is that it relies on voxel-by-voxel comparisons, and cannot capture more complex imaging patterns.

Voxel-based approaches have also been used in supervised frameworks for segmenting brain lesions. These methods use manually-segmented images annotated by experts, and learn a predictive model from positive and negative training samples obtained from individual voxels with known class labels. Voxel intensities (or values derived from intensities) from single or multiple image channels are used as features that characterize each individual voxel, and a classifier (e.g. kNN, SVM (Burges, 1998; Cortes and Vapnik, 1995)) is trained on these feature vectors (Anbeek et al., 2004; Lao et al., 2008).

Atlas-based methods have been proposed as a way of including spatial context information in lesion segmentation (Prastawa et al., 2004; Shiee et al., 2010; Van Leemput et al., 2001; Wu et al., 2006). A statistical atlas provides the prior probability of each voxel to belong to a particular healthy tissue type. In Prastawa et al. (2004) the atlas is used for sampling voxels from each healthy tissue type. A robust estimator is then used for estimating the probability density function (*pdf*) of the healthy brain tissue intensity. Brain tumors are segmented as outliers of the estimated *pdf*. Similarly, in Van Leemput et al. (2001) the parameters of a stochastic tissue intensity model for normal brain are estimated, while simultaneously detecting lesions as voxels that are not well explained by the model. An atlas is used for prior classification of image voxels into tissue types. In both approaches, tissue specific intensity models are used for estimating the normal variation, and the spatial context information is independently used as a prior.

* Corresponding author. Address: 3600 Market St., Suite 380 Philadelphia, PA 19104, USA. Tel.: +1 215 662 7362.
E-mail address: guray.erus@uphs.upenn.edu (G. Erus).

http://dx.doi.org/10.1016/j.media.2014.02.003
1361-8415/

In this paper, we propose a multi-variate pattern analysis method for segmenting abnormalities. Instead of analyzing voxels individually or estimating intensity distributions of healthy tissue types, we basically aim to learn the statistical variation of the healthy brain anatomy in the high dimensional space, i.e. to estimate the *pdf* of all image voxels, from a training database of normal brain scans.[1] Rather than classifying image voxels independently as normal or abnormal, the proposed method detects patterns of abnormality by projecting an image into the hypervolume of normals as defined by statistical models learned from the training samples. We do not make any assumption about the characteristics of the abnormality, however we assume that normal anatomy follows specific patterns and certain aspects of its statistics can be estimated from a number of representative examples. The method is not specific to a given type of pathology, and thus, it does not require training for many possible types of abnormalities, which is often not practical or even possible, particularly when the pathology is variable and/or unknown in advance.

An important challenge of multi-variate analysis on image data is the problem of high-dimensionality and small sample size. For example, a typical MRI scan of the brain includes several millions of measurements on respective voxels. Moreover, the structure and function of many organs, particularly of the brain and the heart, are very complex and difficult to summarize with a small number of variables extracted from such images. Finally, pathology can cause even more complex and subtle changes in the imaging characteristics, thereby rendering it extremely difficult to train an algorithm to find such changes, without knowing in advance what features to look for. Even though the complexity of the problem would call for thousands of scans to be used as training examples, typical imaging studies often offer just one or two hundred scans for training, or even less. Learning algorithms therefore are limited with respect to the complexity of image patterns they can learn, and the precision with which they can learn them and identify them in new images. We focus on addressing the challenges of such direct estimation of statistical variation.

A common approach to learning from an image database involves some form of dimensionality reduction, such as principal component analysis (PCA) or some sparse decomposition of the training examples, so that a relatively small and manageable set of "features" is extracted from the training data. These features then define the feature space within which each image is projected, prior to building statistical models that estimate its characteristics. If a number of normal brain MRI scans is available, PCA can be used to approximate the statistical distribution of these images, under Gaussianity assumptions. When a new image is presented, its location within the PCA subspace is determined via projection to the principal eigenvectors. Normality or abnormality of the test image can be evaluated by calculating its respective likelihood, given the PCA model of the training set. Similarly, the pattern of abnormality can be estimated by projecting the test image to the subspace of normal brain images, as determined by the PCA. However, the dimensionality of the PCA space is bounded by the number of training samples. Consequently, a PCA model of the whole brain obtained from a limited number of training samples can only capture relatively global patterns of variation in the data.

In order to overcome the small sample size limitation, building upon prior work (Erus et al., 2010), we first propose to sample a large number of lower dimensional subspaces, each of which is sufficiently small relative to the underlying image variations and the available sample size. In the experiments herein, each of these subspaces represents image patches, albeit it does not necessarily have to. The subspaces are derived in a multi-scale fashion, and capture image characteristics ranging from fine and localized to coarser and relatively more global. We impose conditions that allow the statistical variations of these subspaces to be reliably estimated from the available training data. The main premise here is that, an imaging pattern that is consistent with a large number of marginal *pdf*s, is likely to be consistent with the overall *pdf*, which has not been explicitly estimated due to the high image dimensionality.

We further reduce the dimensionality within each image patch by applying an individualized feature selection strategy. Learning methods generally try to find the features that best represent the entire set of training examples and their variability. We propose an alternative approach based on a "target specific" feature selection strategy. When the goal is to analyze a specific test image, called the "target", not belonging to the training image database, the optimal feature extraction and dimensionality reduction for classification of the target image is not necessarily the one that is optimal for the entire training set, but one that best learns the features that are relevant for the specific target image. For example, instead of attempting to learn a statistical model or classifier that reflects all possible variations of normal brain anatomy, we only need to learn the variations of the specific anatomical features encountered in the target brain. Such target-specific learning improves our ability to learn from a database of examples, as it focuses on the features that are present in the specific target being analyzed. These features, of course, are different for different target images. In our experiments, a smaller set of target-specific features are selected within each image patch through an approach based on wavelet thresholding (Donoho, 1995), since it is a method that has been very successful in dimensionality reduction of images and signals. However other sparse decompositions can be readily used instead (Sjostrand et al., 2007). Typically, a small number of features, or expansion coefficients, is necessary to construct the main characteristics of a target image.

The validation experiments are applied on FLAIR (FLuid Attenuation Inversion Recovery Magnetic Resonance Imaging)-MRI images, for the segmentation of two types of abnormalities with different signal and spatial characteristics: white matter lesions and cortical infarcts. We evaluate the segmentation accuracy on both simulated and real abnormalities. WMLs show up as hyperintensities with respect to surrounding healthy white matter (WM) tissues on FLAIR images. An infarct is generally the result of a stroke that occurs when the blood supply to the brain is interrupted, due to cerebrovascular disease (CVD). It consists of necrosis, i.e. a region of dead brain tissue, typically surrounded by a rim of tissue that is not dead but not entirely healthy, either. Many clinical studies investigating CVD require segmentation of the necrosis (Barkhof, 2003). The accurate segmentation of these regions is difficult as their intensity patterns are similar to the adjoining cerebrospinal fluid (CSF).

The paper is organized as follows: Section 2 describes the proposed method. The experimental results are given and discussed in Section 3. Section 4 summarizes and concludes with additional discussions and future perspectives.

2. Method

Consider medical images of a normative population coregistered to a common domain Ω as realizations of a d-dimensional random vector I, consisting of d scalar random variables $[x_1, x_2, \ldots, x_d]$ corresponding to intensities of the image voxels. The joint *pdf* of I,

[1] Through the text the term "normal scan" refers to a brain image from a healthy subject without abnormalities, unless otherwise stated.