DTU Library

Sparse Decomposition and Modeling of Anatomical Shape Variation

Sjöstrand, Karl; Rostrup, Egill; Ryberg, Charlotte; Larsen, Rasmus; Studholme, Colin; Baezner, Hansjoerg; Ferro, Jose; Fazekas, Franz; Pantoni, Leonardo; Inzitari, Domenico

Total number of authors:
11

Sparse Decomposition and Modeling of Anatomical Shape Variation

Karl Sjöstrand*, Egill Rostrup, Charlotte Ryberg, Rasmus Larsen, Colin Studholme, *Member, IEEE*, Hansjoerg Baezner, Jose Ferro, Franz Fazekas, Leonardo Pantoni, Domenico Inzitari, Gunhild Waldemar, and on behalf of the LADIS study group

Abstract—Recent advances in statistics have spawned powerful methods for regression and data decomposition that promote sparsity, a property that facilitates interpretation of the results. Sparse models use a small subset of the available variables and may perform as well or better than their full counterparts if constructed carefully. In most medical applications, models are required to have both good statistical performance and a relevant clinical interpretation to be of value. Morphometry of the corpus callosum is one illustrative example. This paper presents a method for relating spatial features to clinical outcome data. A set of parsimonious variables is extracted using sparse principal component analysis, producing simple yet characteristic features. The relation of these variables with clinical data is then established using a regression model. The result may be visualized as patterns of anatomical variation related to clinical outcome. In the present application, landmark-based shape data of the corpus callosum is analyzed in relation to age, gender, and clinical tests of walking speed and verbal fluency. To put the data-driven sparse principal component method into perspective, we consider two alternative techniques, one where features are derived using a model-based wavelet approach, and one where the original variables are regressed directly on the outcome.

Index Terms—Corpus callosum (CC), decomposition, Leukoaraiosis And DISability in the elderly (LADIS), principal component analysis (PCA), shape analysis, sparse.

*K. Sjöstrand is with the Department of Informatics and Mathematical Modelling, the Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark (e-mail: kas@imm.dtu.dk).

E. Rostrup is with the Danish Research Center for Magnetic Resonance, Copenhagen University Hospital, DK-2650 Hvidovre, Denmark.

C. Ryberg is with the Memory Disorders Research Group, Department of Neurology, Copenhagen University Hospital, DK-2100 Copenhagen, Denmark and with the Danish Research Center for Magnetic Resonance, Copenhagen University Hospital, DK-2650 Hvidovre, Denmark.

R. Larsen is with the Department of Informatics and Mathematical Modelling, the Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.

C. Studholme is with the Department of Radiology, University of California, San Francisco, CA 94143 USA.

H. Baezner is with the Department of Neurology, University of Heidelberg, Klinikum Mannheim, Mannheim 68167, Germany.

J. Ferro is with the Servico de Neurologia, Centro de Estudos Egas Moniz, Hospital de Santa Maria, 1649-035 Lisboa, Portugal.

F. Fazekas is with the Department of Neurology, Medical University, A-8036 Graz, Austria.

L. Pantoni and D. Inzitari are with the Department of Neurological and Psychiatric Science, University of Florence, 50139 Florence, Italy.

G. Waldemar is with the Memory Disorders Research Group, Department of Neurology, Copenhagen University Hospital, DK-2200 Copenhagen, Denmark.

Digital Object Identifier 10.1109/TMI.2007.898808

I. INTRODUCTION

TRADITIONAL morphometric investigations in medicine make use of simple metrics such as volume, area, length, and various ratios to evaluate relations between structure and function. The outcomes of such studies provide the examiner with an indication of the characteristic anatomy of a clinical population, or spatial features related to pathology, for example. More intricate features provide more information for interpretation, but require a more detailed hypothesis of the process under study. For a clinical investigation that is exploratory in nature, it makes sense to use an exploratory method to extract features. Such variables should ideally have a clear relation to the relevant morphology while imposing as few assumptions on the data as possible. During the last two decades, methods for extracting more complex representations of anatomy from image data of increasingly high resolution have evolved. This has led to the development of methods that allow for the computation of more abstract features such as the mean shape and typical deformation patterns according to the latent shape distribution. Derived variables may be concretized as examples of anatomy, which allows for more detailed investigation and interpretation. Furthermore, the relationship between structural and clinical variables can be analyzed in a formal statistical framework, making the investigation of certain clinical hypotheses possible.

The challenge posed by increasingly complex anatomical representations is to extract physically intuitive parameterizations of spatial variation. Conventional statistical techniques tend to extract global decompositions of spatial data. However, the effects of many biological processes of interest are expected to be anatomically localized, even if the particular location, extent, and frequency are usually unknown.

This paper presents a methodology in which a statistically defined, spatially localized representation of anatomy is automatically extracted. The approach is built on a generic statistical method known as sparse principal component analysis. The paper further describes a way of relating these spatial variables to a clinical outcome variable, producing a characteristic deformation of the present anatomy and indicating its statistical relevance.

Advanced techniques for analyzing the shape of anatomical structures have emerged during the last two decades [1]. A suitable choice of shape parameterization is crucial to ensure correct and efficient analysis, and several techniques have been developed to describe the variability of human anatomy. These techniques include corresponding landmarks [2]–[4], representations in the frequency domain in two [5] and three [6] dimensions, skeleton-based techniques [7], [8], distance transforms

[9], [10], and deformation fields resulting from the registration of a set of images to a common [11], [12].

Most of these methods produce a large number of spatial features. To devise a more manageable model, the features are often arranged into groups according to a spatial or statistical criterion. Cootes *et al.* [3] pioneered the use of principal component analysis (PCA) to decompose sets of landmarks. This provides compact and powerful models for shape-driven segmentation and registration. A more recent example is Davies *et al.* [13], who decomposed sets of landmarks with optimized correspondences using PCA and used the resulting shape features in a classification study of the hippocampus. PCA has also been used to decompose other shape descriptors. For instance, Kelemen *et al.* [14] presented a framework similar to that of Cootes *et al.* [3] for frequency domain descriptors applied to the segmentation of the hippocampus, and Le Briquer and Gee [15] applied PCA to deformation fields extending throughout the entire brain.

The use of PCA as an explanatory basis for interpretation in clinical applications has been limited (Peterson *et al.* [16] is one exception). While PCA is an excellent tool for efficient data representation, the global nature of the derived variables makes interpretation difficult. This motivates the use of an extension to PCA known as sparse PCA (SPCA). While the variables derived by PCA consist of linear combinations of *all* original variables, SPCA forces the weights on some variables towards zero, while others are adjusted to uphold the variance-maximizing properties of PCA. The idea in studies of anatomy is that each variable describes a spatial pattern of variation with a simple structure and a clinically relevant interpretation [17]. Although conceptually simple, calculation of SPCA has proved difficult and several algorithms have been proposed [18]–[24]. The approach advocated in this paper was developed by Zou *et al.* [25] and formulates PCA as a regression problem, using a recent variable selection algorithm [26] to achieve sparsity. Selection of important variables is achieved by penalization of the weights on each variable using the ℓ_1 norm, a methodology introduced with the LASSO regression framework [27], along with a method for its efficient computation [28].

Examples of other statistical decomposition techniques used in shape analysis are factor analysis [29], varimax rotated principal components [30], and independent component analysis [31]. The latter two typically produce approximately sparse representations but lack the flexibility of most SPCA implementations.

In medical image analysis, the use of variable selection algorithms to aid interpretation is gaining momentum. Yushkevich *et al.* [32] employed a support vector machine classification algorithm that incorporates variable selection to select subregions of the hippocampus separating schizophrenic patients from normal controls. A similar algorithm was used by Stoeckel and Fung [33] on SPECT imagery to find regions of the brain that differentiate between healthy subjects and patients with Alzheimer's disease. Fan *et al.* [34] used variable selection on deformation field data in a study of schizophrenia.

The methodology introduced in this paper is applied to a data set of 569 outlines of the corpus callosum (CC) brain structure, obtained from a study on atrophy in an elderly population [35]. The CC provides an illustrative example of a structure that may benefit from a localized analysis. The white-matter fibers defining the CC are organized according to an anterior–posterior topographical organization; tissue loss and discrepancies can therefore be expected to be constrained to specific regions [36]. The CC is perhaps the most popular single nervous structure for morphometric analysis and a wide range of applications in shape analysis exist. Bookstein [2] characterized deformations of the CC using partial thin-plate spline warps. Davatzikos *et al.* [37], Machado and Gee [38], and Dubb *et al.* [39] used deformation field features to find gender differences in the CC. Golland *et al.* [40], [41] takes a classification approach to finding anatomical discrepancies between populations where group differences are characterized by the gradient of the classifier function and applies the method to a study of the CC in affective disorder. Joshi *et al.* [42] extract predefined global and local shape features of the CC using a multiscale medial shape representation. The features are used for classification of schizophrenic and normal subjects.

The advantage of the method presented in this paper over previous work is the extraction of interpretable localized features governed by few and weak assumptions. The central assumption is on the extent of the deformations; however, we propose to alleviate this assumption by extracting features on several scales.

To put the SPCA method into perspective, we provide a comparison with two alternative analysis methods. The first analyzes the original shape features (landmarks) directly to provide a sparse representation of anatomy. The second method challenges a potential shortcoming of a data-driven process such as PCA or SPCA in that a minor but clinically relevant variation may be omitted. We therefore include a model-based method for decomposition based on the wavelet transform. Multiscale representation of curves using the wavelet transform has found applications in both computer graphics [43] and image analysis [44]. The wavelet transform decomposes the anatomy into coefficients of both scale and localization [45] and offers a sparse orthogonal shape basis with acceptable interpretability.

Characteristic deformation patterns of the CC are derived for four different clinical variables. Focus is on shape differences of the CC due to gender [37]–[39], [46]–[49], but results are also given for age effects, verbal fluency, and walking speed. Using the same data set, atrophy of the CC has previously been shown to correlate with general cognitive and physical decline [36], [50].

II. METHODS

To understand and quantify a complex process such as the variability of anatomy, one has to compromise between a general model and a compact model. The first property means that it should be possible to model any conceivable deformation pattern, while the second property ensures that the number of variables used to do this is kept small, allowing more power for subsequent statistical analysis. If the intended use of the model goes beyond prediction, interpretability adds to this list of requirements. Many anatomical processes are expected to be localized, leading to high correlations between neighboring features. This property can be used to derive variables where a single variable may describe deformations across several features in an anatomically plausible fashion. Furthermore, restricting the analysis to relevant variation only, the number of variables can be reduced.

In the following, we will review two methods for deriving such variables.

A. Principal Component Analysis

The first method is perhaps the most well known and widely used method for data decomposition in general, PCA. To introduce the method, as well as the notation and terminology used throughout the rest of this paper, a brief explanation will be given here.

PCA takes a mean-centered $(n \times p)$ data matrix $\mathbf{X}$, with n being the number of observations and p being the number of variables, and transforms it by $\mathbf{Z} = \mathbf{XB}$ such that the derived variables (the columns of $\mathbf{Z}$) are uncorrelated and correspond to directions of maximal variance in the data. The derived coordinate axes are the columns of $\mathbf{B}$, called *loading vectors*, with individual elements known as *loadings*. These are at right angles with each other; PCA is simply a rotation of the original coordinate system, and the $(p \times p)$ loading matrix $\mathbf{B}$ is the rotation matrix. The new variables (the columns of $\mathbf{Z}$) are known as principal components (PCs). Usually, only the first k components, $k < p$, are retained since these explain the majority of the sample set variance. This makes $\mathbf{Z}$ $(n \times k)$ and $\mathbf{B}$ $(p \times k)$. The loading matrix can be calculated using singular value decomposition of the data matrix $\mathbf{X}$ or by eigenanalysis of the corresponding covariance or correlation matrix.

B. Sparse Principal Component Analysis

SPCA can be described as an extension of PCA, where a constraint of the number of nonzero loadings is added. The recent development in statistical methods for variable selection in regression has resulted in an SPCA approach described by Zou and Hastie [25]. This method is used throughout this paper and the idea will be described here in brief. For a complete description, consult [25] and the preliminary papers [26]–[28]. Refer to [17] for an introduction on using SPCA to decompose shape data.

The regression methods used in the calculation of SPCA all originate from ordinary least squares (OLS) approximations. The independent variable $\mathbf{y}$ is approximated by a linear combination of the dependent variables in $\mathbf{X}$. The coefficients for each variable (column) of $\mathbf{X}$ are contained in $\mathbf{b}$

$$\mathbf{b}_{\text{OLS}} = \arg \min_{\mathbf{b}} \|\mathbf{y} - \mathbf{Xb}\|^2 \qquad (1)$$

where $\|\cdot\|$ represents the ℓ_2 norm. This is the best linear unbiased estimator given a number of assumptions, such as independent and identically distributed (i.i.d.) residuals. However, if some bias is allowed, estimators can be found with lower mean square error than OLS when tested on an unseen set of observations. A common way of implementing this is by introducing some constraint on the coefficients in $\mathbf{b}$. The methods described here use constraints on either the ℓ_1 norm or the ℓ_2 norm of $\mathbf{b}$, or both. Adding the ℓ_2 constraint gives

$$\mathbf{b}_{\text{ridge}} = \arg \min_{\mathbf{b}} \|\mathbf{y} - \mathbf{Xb}\|^2 + \lambda \|\mathbf{b}\|^2. \qquad (2)$$

This is known as ridge regression [51]. Sufficiently large values of λ will shrink the coefficients of $\mathbf{b}$. The shrinkage introduces bias but lowers the variance of the estimates. Careful selection of λ may lead to improved prediction accuracy, but of more interest here are the improved numerical properties, making estimation in cases where $p > n$ is feasible [52]. Replacing the ℓ_2 norm in the constraint with the ℓ_1 norm gives

$$\mathbf{b}_{\text{LASSO}} = \arg \min_{\mathbf{b}} \|\mathbf{y} - \mathbf{Xb}\|^2 + \delta \|\mathbf{b}\|_1 \qquad (3)$$

where $\|\mathbf{b}\|_1 = \sum_{i=1}^{p} |b_i|$. This is the LASSO method [27]. Using the ℓ_1 norm not only shrinks the coefficients, but also drives them one by one to exactly zero as δ increases. This implements a form of variable selection, as minor coefficients will be set to zero in a controllable fashion, while the remaining coefficients will be used to minimize the size of the regression residuals.

A third possibility is to use a combination of the constraints from ridge regression and the LASSO. This approach is known as the elastic net [26] and has the form

$$\mathbf{b}_{\text{EN}} = \arg \min_{\mathbf{b}} \|\mathbf{y} - \mathbf{Xb}\|^2 + \lambda \|\mathbf{b}\|^2 + \delta \|\mathbf{b}\|_1. \qquad (4)$$

The main benefit of the elastic net is that it better handles cases where $p > n$. The elastic net can be formulated as a LASSO problem on augmented variables and is solved using the same algorithm, outlined as follows.

OLS and ridge regression have closed-form solutions, that is, $\mathbf{b}_{\text{OLS}}$ and $\mathbf{b}_{\text{ridge}}$ can be expressed as functions of the random variable $\mathbf{y}$ $\mathbf{b}_{\text{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ and $\mathbf{b}_{\text{ridge}} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$. This is not true for the LASSO and elastic net methods. For many years, LASSO solutions were found using standard optimization techniques, which made for long computation times. In 2002, Efron *et al.* [28] published a report on a new regression method which they called least angle regression (LARS). Although conceptually different, the method is shown to be very similar to LASSO, and through a small modification, the exact LASSO solution can be computed. The method is built on a powerful geometric framework, through which a computationally thrifty algorithm is conceived. The paper shows that the coefficients $\mathbf{b}$ are piecewise linear with respect to the regularization parameter δ, with breakpoints as variables enter or leave the model. The breakpoints can be established using standard linear algebra. Using this property, the entire regularization path can be computed. Starting with the empty model ($\mathbf{b} = \mathbf{0}$), variables are added and occasionally subtracted as δ grows until all variables are nonzero and the full least squares solution is reached. Hereby, the LARS path algorithm returns the solutions for all possible values of δ. The computational cost of obtaining the entire LASSO regularization path is the same as for a single least squares fit.

PCA and SPCA are strongly related to these regression algorithms. One way of describing PCA using regression is by treating each principal component as a response vector and regressing this on the p variables using ridge regression

$$\hat{\mathbf{b}}_i = \arg \min_{\mathbf{b}_i} \|\mathbf{z}_i - \mathbf{Xb}_i\|^2 + \lambda \|\mathbf{b}_i\|^2. \qquad (5)$$

The minimizing coefficient vector $\hat{\mathbf{b}}_i$ normalized to unit length is exactly the ith principal loading vector, independent of the choice of λ [25]. A direct approach to sparse PCA is obtained by adding the ℓ_1 (LASSO) constraint

$$\hat{\mathbf{b}}_i = \arg\min_{\mathbf{b}_i} \|\mathbf{z}_i - \mathbf{X}\mathbf{b}_i\|^2 + \lambda\|\mathbf{b}_i\|^2 + \delta\|\mathbf{b}_i\|_1. \quad (6)$$

The regression procedure will calculate a loading vector $\hat{\mathbf{b}}_i$ such that the resulting PC is close to $\mathbf{z}_i$ while being sparse. The weakness of this approach is that all solutions are constrained to the immediate vicinity of a regular PCA. A better approach would be to approximate the *properties* of PCA, rather than its exact results. Specifically, the columns of the loading matrix $\mathbf{B}$ should be near orthogonal and describe directions of high variance in the data set. Zou and Hastie propose a problem formulation called the *SPCA criterion* [25] to address this

$$(\hat{\mathbf{A}}, \hat{\mathbf{B}}) = \arg\min_{\mathbf{A},\mathbf{B}} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{A}\mathbf{B}^T\mathbf{x}_i\|^2$$
$$+ \lambda \sum_{j=1}^{k} \|\mathbf{b}_j\|^2 + \sum_{j=1}^{k} \delta_j\|\mathbf{b}_j\|_1$$
$$\text{subject to } \mathbf{A}^T\mathbf{A} = \mathbf{I}_k. \quad (7)$$

To clarify this expression, it will be broken down into components. First, $\mathbf{B}^T\mathbf{x}_i$ takes the variables of observation i and projects them onto the principal axes (loading vectors) of $\mathbf{B}$. Note that $\mathbf{x}_i$ denotes the ith column of $\mathbf{X}^T$. Only k PCs are retained, meaning that some information is lost in this transformation. Next, $\mathbf{A}\mathbf{B}^T\mathbf{x}_i$ takes the scores of $\mathbf{B}^T\mathbf{x}_i$ and transforms them back into the original space. The orthogonality constraint on $\mathbf{A}$ makes sure $\mathbf{B}$ is near orthogonal. The whole term $\sum_{i=1}^{n}\|\mathbf{x}_i - \mathbf{A}\mathbf{B}^T\mathbf{x}_i\|^2$ measures the reconstruction error. The remaining constraints are the same as for elastic net regression, driving the columns of $\mathbf{B}$ towards sparsity and ensuring good numerical properties in cases where $p > n$. Some further insight into this criterion is given by considering the loss function alone, with the additional constraint $\mathbf{B} = \mathbf{A}$

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{A}\mathbf{A}^T\mathbf{x}_i\|^2 \text{ subject to } \mathbf{A}^T\mathbf{A} = \mathbf{I}_k. \quad (8)$$

The minimizer of this function is given by the first k loading vectors of a standard PCA; this equation is, in fact, the basis for a derivation of PCA [52] other that the standard variance-maximization approach. One of the key results of the SPCA paper [25] is that the constraint $\mathbf{B} = \mathbf{A}$ can be omitted given the addition of an ℓ_2 penalty term

$$(\hat{\mathbf{A}}, \hat{\mathbf{B}}) = \arg\min_{\mathbf{A},\mathbf{B}} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{A}\mathbf{B}^T\mathbf{x}_i\|^2$$
$$+ \lambda \sum_{j=1}^{k} \|\mathbf{b}_j\|^2$$
$$\text{subject to } \mathbf{A}^T\mathbf{A} = \mathbf{I}_k. \quad (9)$$

The columns of $\mathbf{B}$ (normalized to unit length) will still give the exact PCA solution. The SPCA criterion then augments this formulation by the addition of the ℓ_1 term, making it possible to

estimate loading vectors that range from the results of a standard PCA to various sparse approximations.

The constraint weight λ must be chosen beforehand and has the same value for all PCs, while δ may be set to different values for each PC, offering good flexibility. The level of sparsity can also be defined by specifying a target number of active variables. This is done by terminating the elastic net estimation when a suitable number of variables has entered the model. This stopping criterion is very useful in practice.

Equation (7) resembles the elastic net formulation, but there is a significant difference. Instead of estimating a single coefficient vector, this problem has two matrices of unknown coefficients, $\mathbf{A}$ and $\mathbf{B}$. A reasonably efficient optimization method for minimizing the SPCA criterion is presented in [25]. First, assume $\mathbf{A}$ is known. Expanding and rearranging (7) shows that $\mathbf{B}$ can be estimated by solving k independent elastic net problems, one for each column of $\mathbf{B}$ (loading vector). Referring to the elastic net formulation in (4), the predictor matrix is $\mathbf{X}$ as usual, while $\mathbf{y} = \mathbf{X}\mathbf{a}_i$, where $\mathbf{a}_i$ is the ith column of $\mathbf{A}$. On the other hand, if $\mathbf{B}$ is known, $\mathbf{A}$ can be calculated using a singular value decomposition; if $\mathbf{X}^T\mathbf{X}\mathbf{B} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, then $\mathbf{A} = \mathbf{U}\mathbf{V}^T$. Since both matrices are unknown, an initial guess is made and $\mathbf{A}$ and $\mathbf{B}$ are estimated alternately until convergence. The standard option is to initialize $\mathbf{A}$ to the loadings of the k first ordinary principal components.

C. Statistical Analysis

The goal of the analysis is to determine the relationship between the derived variables (loading vectors) and a clinical outcome variable. Clinical variables here are assumed to consist of a single score for each patient (e.g., age) and are therefore n dimensional. However, methods such as PCA and SPCA derive new variables that are p dimensional, that is, each variable can be interpreted as a perturbation of the mean observation. As a preliminary step, the presence of each PCA/SPCA variable in each subject must be measured. We propose to do this via univariate regression. The following model formulates the idea, where the presence z of deformation mode j is determined for the shape corresponding to subject i (row vector) by

$$\mathbf{x}_i^T = z\mathbf{b}_j + \boldsymbol{\varepsilon}. \quad (10)$$

The loading vectors $\mathbf{b}_j$ have unit length for both PCA and SPCA, yielding the least squares estimate $z = \mathbf{x}_i\mathbf{b}_j$. This is simply the (i, j)th entry of the $n \times k$ scores matrix $\mathbf{Z}$, which, as described in Section II-A, is estimated by $\mathbf{Z} = \mathbf{X}\mathbf{B}$, also for SPCA. The presence z can be interpreted as a measure of correlation between shape i and deformation j.

The scores matrix provides k n-dimensional variables that can be related to clinical outcome. In this paper, we propose to establish this relation via a series of univariate tests. This approach is similar to those used in, e.g., analysis of functional images and deformation/tensor-based analysis [39], where separate tests are performed at each voxel of an image volume. The statistical properties of the scores vectors are often better suited to a regression analysis than clinical variables, which may be categorical or ordinal (ordered categorical). We therefore assign

the scores vector as the outcome variable. The test for a relationship between spatial variable i and the clinical outcome $\mathbf{y}$ becomes

$$\mathbf{z}_i = \beta_i \mathbf{y} + \boldsymbol{\varepsilon}. \tag{11}$$

Confounding variables enter the model on the right-hand side as covariates. This simple regression model is solved using the least squares criterion, providing access to a range of statistical properties, most notably t-scores with corresponding p-values, measuring the probability that a significant relation is declared when the variables are in fact unrelated.

Using the above analysis, the relationship between the outcome and each spatial variable is established. A complication with this approach is that significance levels should be adjusted for the number of comparisons performed. Bonferroni correction provides one simple procedure, where any test probabilities (p-values) are multiplied by the number of tests performed. This provides strong control over the family-wise (type-I) error rate—the probability that one or more tests are falsely rejected is less than the nominal significance level α. However, this procedure is generally too conservative, leading to unnecessarily high p-values. A more powerful alternative, also with strong control over type-I errors, is provided by nonparametric permutation testing procedures. The specific method used here is described in detail, in e.g., [53], and is based on finding the empirical distribution of a *maximal statistic*. First, we will review the basics of permutation testing and then briefly explain how this may be used to adjust a set of p-values for multiple comparisons.

The idea of permutation testing is that if two variables are in fact unrelated, then the results (for instance from a correlation or regression analysis) should not change notably even though the elements of one of the variables have been randomly shuffled around [54]. By permuting the dependent variable in the regression analysis in (11) R times, where R is some large integer number ($R > 999$), an estimate of the empirical distribution function (EDF) under the null hypothesis is obtained as the histogram of the corresponding t-statistics of the independent variable of interest. Calculating the proportion of t-values exceeding the t-value obtained from the original (non-permuted) regression analysis provides a nonparametric estimate of the p-value of the independent variable. Providing the standard assumptions of the regression analysis in (11) hold, these p-values will be in close agreement with those obtained from a classical parametric analysis.

One advantage of this nonparametric approach is that it provides additional information that can be used to adjust the p-values obtained for multiple comparisons. This information comes in the form of the distribution of the maximal statistic. This statistic consists of the maximal absolute t-value over all tests for each permutation. For the ith repetition, we denote this value as $t_i^{\max}$. After R repetitions, an approximation of the EDF for the maximal statistic is obtained. The *critical value* is defined as the $\lfloor \alpha R \rfloor + 1$ largest member of this distribution. Any t-values exceeding this value are deemed significant at the α level. In practice, we do not need to compute the critical value. An adjusted p-value can be obtained directly from the EDF

of the maximal statistic as the proportion of values exceeding the t-value t from the original regression analysis. Formally, this corresponds to $p_{\text{adjusted}} = (1 + \#\{t_i^{\max} > t\})/(R + 1)$, where $\#$ denotes the number of elements in a set [54].

D. Application to Shape Analysis

In this section, we will describe more specifically how the methods outlined previously are applied to landmark-based shape analysis. We adopt the definition of shape by Kendall [55], stating that shape information is what remains in a data set, when translational, rotational, and scaling effects have been filtered out. The shapes are therefore aligned using a general Procrustes analysis [4]. The removal of scale differences deserves some attention in this application. Many anatomical discrepancies, age-related changes is one example, are likely to include a component of pure scale. Obviously, a sparse decomposition is not suitable for describing global properties with preserved interpretability, which is why we recommend removing such differences. In the subsequent analysis of the results, this fact must be taken into consideration. A separate analysis of area/volume differences may be used to complement the study of local shape variability.

1) PCA Application: Global patterns of shape variability are obtained through a principal component analysis, performed by a singular value decomposition of the centered data matrix $\mathbf{X}$. This matrix consists of the Procrustes-aligned shapes, where the mean shape has been subtracted from each row. Typically, all landmarks contribute to the variance of the data set, meaning that each new variable $\mathbf{b}_i$ (column of $\mathbf{B}$) will affect the entire outline. The usual practice is to truncate the set of variables to account for, e.g., 95% of the total variation. This can be done easily, since the variance explained by $\mathbf{b}_i$ is given directly by the ith eigenvalue of $\text{cov}(\mathbf{X})$. This reduction has a number of advantages. For instance, it excludes noisy (wiggly) deformation modes and it simplifies and strengthens the subsequent statistical analysis.

2) SPCA Application: As for PCA, SPCA is applied to the aligned and centered shapes contained in the data matrix $\mathbf{X}$. A number of parameters govern the results. Also akin to PCA, a choice must be made on the number of variables to retain. Unlike PCA, this must be done in advance. A rough number is provided by the number of variables deemed significant in the PCA analysis, since when estimating an excess of variables, the SPCA algorithm tends to produce highly correlated variables. The next parameter to set is λ in relation to the ℓ_2 constraint. Empirical evidence [17] supported by some theoretical results [25] suggests that the results are largely independent on the specific choice of this parameter. Typically, it is set at a small positive value to ensure good numerical properties. Finally, the parameters δ_j must be set, governing the amount of sparsity of the decomposition. This choice is dependent on the anatomical scale of interest and must be carefully chosen for each application. For many purposes, δ_j will be equal for all j, resulting in the same deformation size for each $\mathbf{b}_i$.

3) Tabulation and Visualization: The most thorough way of presenting the results is a table showing each deformation mode and the significance level for each of these associated with each tested clinical outcome variable. Such a presentation minimizes

the risk of misleading the reader but may also become time consuming and complex to draw conclusions from. In order to construct a sample anatomy related to a specific outcome variable, we suggest creating a compound deformation of a template shape (for most purposes, the mean shape). Each deformation mode exceeding the nominal significance level α contributes to this deformation with strength proportional to its corresponding β (regression coefficient) value. If both the spatial and clinical variables are standardized (zero mean, unit variance) prior to the regression analysis, the coefficients can be interpreted as the change (in standard deviations) in the spatial (response) variable introduced by a unit change in the clinical variable. For interpretational purposes, the use of the β values directly as weights on the various deformations may not produce an anatomically meaningful pattern. Therefore, we instead choose to normalize the β values within the group of spatial variables being tested such that the maximal point-to-point distance is set to an appropriate value. The relative sizes of the deformations will still be correct using this method, but the absolute strengths of the relationships are lost, a fact that must be taken into consideration when analyzing the results. This approach is used in the display of deformations in this paper.

E. Alternative Methods

This section provides a brief explanation of two alternative methods for relating clinical outcome to localized representations of anatomy. One represents a simple and direct analysis, while the other provides a model-based alternative to the data-driven decomposition of PCA/SPCA.

1) Direct Analysis of Original Variables: PCA derives variables that capture global properties of the relevant anatomy, while SPCA provides a more localized alternative. If the analysis is made increasingly localized, the derived variables will ultimately consist of a single component (x or y coordinate in the case of 2-D shape analysis). This results in an immediate and simple approach where the original spatial variables enter (11) one by one on the left-hand side, and their individual relation to the clinical outcome is established.

2) Decomposition Using Wavelet Transform: The pitfall of using subspace techniques such as PCA is that subtle but interesting information may be lost. A minor deformation may be strongly related to a clinical variable, but since the contribution to the sample variance is low, the effect may not be modeled or simply discarded. It is therefore of interest to find a basis where each variable is clinically relevant and all the variance of the original data set is preserved. The wavelet transform may provide one such basis.

A wavelet is a waveform of limited duration. The wavelet transform breaks the original signal into scaled and translated versions of a predefined *mother wavelet* [45]. The original signal is first divided into two parts of low and high scale. These representations are known as the approximation (coarse scale) and the detail (fine scale). The approximation is then further divided in an equivalent fashion, and the process is repeated a suitable number of times. This yields a hierarchy of coefficients organized in a tree structure according to scale and location depicted in Fig. 1. Each wavelet coefficient represents a deformation across several landmarks that is localized in both scale
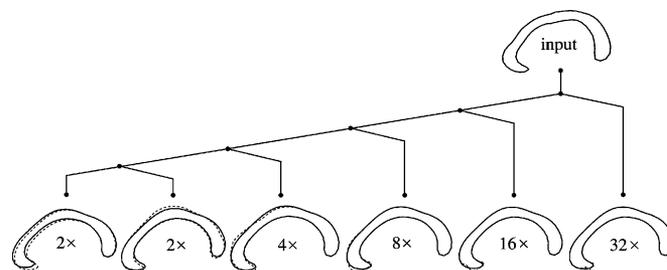

Fig. 1. Hierarchical representation of a shape in wavelet domain. Numbers represent number of wavelet coefficients on each level. Leftmost branch represents approximation, while other branches correspond to detail at different scales. At each branch, one example of resulting shape deformation is shown (dashed line) with mean shape (solid line) as a reference.
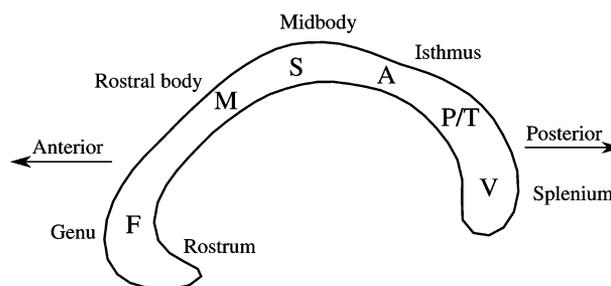

Fig. 2. Subregions and approximate fiber connectivity of CC. Connectivity labels are F (frontal), M (motor), S (somatosensory), A (auditory), P/T (parietotemporal), and V (visual). Image is adapted from [58] and is based on a postmortem study [49].

(spatial extent) and position along the outline. The first order *coiflet* wavelet is used here, as this was deemed suitable for describing local shape changes because of its low complexity and high symmetry. This particular wavelet is orthogonal, meaning that the variance and structure of the original shape data are preserved. In the present analysis, x and y coordinates are treated separately as 1-D periodic functions. The two resulting wavelet coordinate vectors are concatenated into a single observation. This process is repeated for all shapes in the data set, producing a set of variables of the same size as the original data.

III. RESULTS

The proposed method was applied to a large data set of 2-D outlines of the CC brain structure. The CC is the band of fibers connecting the hemispheres of the brain. These fibers are organized in the approximate anterior to posterior topographical organization depicted in Fig. 2. The data set is part of the longitudinal LADIS (Leukoaraiosis And DISability in the elderly) study, involving 12 European countries and more than 700 patients. Refer to [35] for a complete description of this study and the project protocol. This paper presents a cross-sectional study based on baseline data with 569 (312 female) subjects. The shape data was extracted from the baseline MR images (3-D sagittal or coronal T1-weighted MPRAGE, voxel size $1 \times 1 \times 1$ mm). In the mid-sagittal plane, the CC was registered using a learning-based active appearance model [56], [57], trained on 62 CC examples, each manually annotated with 78 corresponding landmarks. The automatic registration
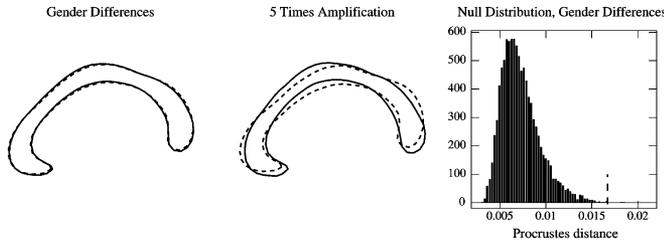
Fig. 3. Mean female CC shape (dashed) versus male (solid). Also shown is empirical null distribution function of Procrustes distance between two shapes. Observed distance is represented by dashed vertical line, corresponding to $p = 0.0015$.

Fig. 4. Example deformation modes. Each group of deformations represents one scale. Notation $\mathrm{SPCA}(k)$ denotes a sparse decomposition with k nonzero components. Mean shape is shown in between representations of deformations in positive (solid line) and negative (dashed line) directions, respectively. Deformations have been appropriately scaled for visualization.

was followed by manual inspection and correction by an expert reviewer, unaware of any clinical status [50].

Initially, a simple test was performed to see whether the shape of the male and female CC differed significantly. The full Procrustes distance was used to measure the discrepancy between two shapes. This measure is a normalized sum of point-to-point distances between the aligned shapes $\mathbf{w}$ and $\mathbf{y}$ in complex notation [4]

$$d_F(\mathbf{y}, \mathbf{w}) = \sqrt{1 - \frac{\mathbf{y}^*\mathbf{w}\mathbf{w}^*\mathbf{y}}{\mathbf{w}^*\mathbf{w}\mathbf{y}^*\mathbf{y}}} \qquad (12)$$

where $\mathbf{w}^*$ is the transpose complex conjugate of $\mathbf{w}$. The Procrustes distance between the male and female mean shapes was found to be $d_F(\bar{\mathbf{x}}_{\mathrm{male}}, \bar{\mathbf{x}}_{\mathrm{female}}) = 0.0167$. Placing this value on the null distribution estimated by calculation of the Procrustes distance based on a large number of permutations of the data set (cf., [53], [54], and Section II-C), the shapes were found to differ significantly ($p = 0.0015$, $R = 9999$ repetitions). Fig. 3 shows the female versus the male mean CC shapes and the corresponding null distribution. The dashed line indicates the nominal Procrustes distance $d_F(\bar{\mathbf{x}}_{\mathrm{male}}, \bar{\mathbf{x}}_{\mathrm{female}})$.

The described algorithm for sparse principal component decomposition was applied to the Procrustes-aligned shape data. The anatomical scale of any deformations related to the clinical outcome variables of interest is unknown. Three decompositions on three different scales were therefore calculated. The extent of the deformations was set to 5, 20, and 50 nonzero components, corresponding to 3%, 13%, and 32% of the total number of components ($2 \cdot 78 = 156$). This choice of scales provides a relatively large span of deformations while maintaining interpretability. A standard PCA was also applied, obviously corresponding to 100% nonzero components. Fig. 4 shows the resulting deformations. Note the coherence of the sparse deformation patterns. This property is in no way enforced by the algorithm and neither are such assumptions desired from a fully exploratory method. Instead, the coherence is a result of the high correlations between adjacent landmarks. In theory, however, there is nothing to keep the deformations from breaking up into an arbitrary number of separate effects, and this is seen to occur to some extent for SPCA(20) and SPCA(50).

The deformations for each SPCA scale and for PCA were related to four clinical outcome variables using the univariate
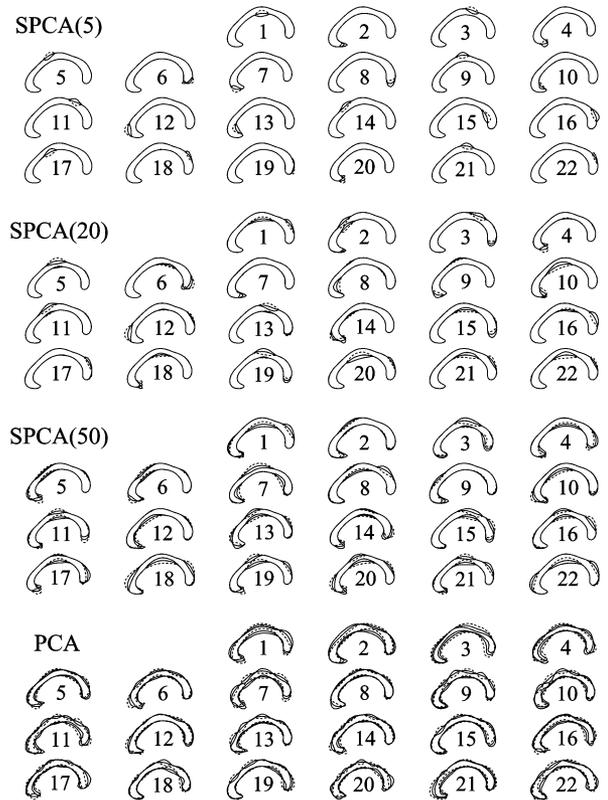
regression scheme outlined in Section II-C. The variables are gender (male/female), age (years), walking speed (meters/second), and verbal fluency (words/minute). In the tests for gender and age, no confounding variables were identified. For walking speed and verbal fluency, the results were adjusted for age, gender, level of education, and the logarithm of the volume of white-matter hyperintensities, as suggested by previous studies on the same data set [36], [50].

The results for each clinical variable are given in Fig. 5. As described in Section II-D, the deformations shown for each scale and variable are the compounded results for each deformation mode corresponding to an adjusted p-value below $\alpha = 0.05$. To provide more specific results in the case of gender differences, Table I lists the resulting coefficient values for each deformation mode and scale with corresponding significance levels.

To put the data-driven SPCA method into perspective, tests for each clinical outcome variable were also investigated through a direct analysis of the original variables and by using the model-based wavelet approach. Fig. 6 shows the results from these tests.

IV. DISCUSSION

This paper has introduced a method for relating localized, anatomically meaningful patterns of variation to clinical outcome using a method for the estimation of sparse principal components.
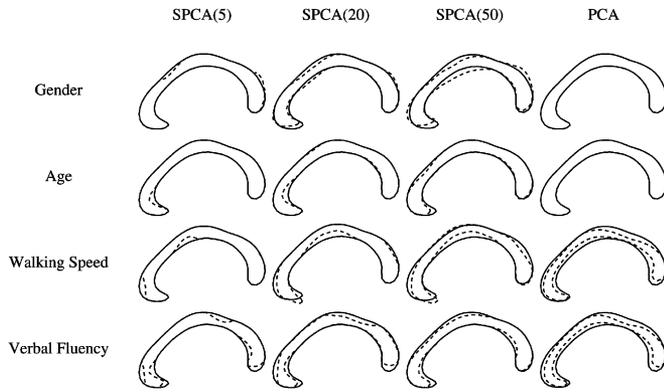
Fig. 5. Results for each clinical outcome variable and scale of decomposition. Mean shape is denoted by solid lines, while dashed lines represent a more female CC, old age, and lower scores for walking speed and verbal fluency. Results for verbal fluency have not been corrected for multiple comparisons. Deformations show a high degree of consistency over different scales and are sufficiently coherent and regular for clinical interpretation.

TABLE I
REGRESSION COEFFICIENTS β_i [CF. (11)] FROM INVESTIGATION OF CC GENDER DIFFERENCES. SIGNIFICANCE LEVELS ARE INDICATED BY $*(p < 0.05)$, $**(p < 0.01)$, AND $***(p < 0.001)$, CORRECTED FOR MULTIPLE COMPARISONS USING PERMUTATION TESTING. ROW NUMBERS REFER TO DEFORMATION MODES SHOWN IN FIG. 4

|	SPCA(5)	SPCA(20)	SPCA(50)	PCA
1	-0.0034	-0.0075	-0.0092	-0.0131
2	0.0043	0.0074 **	0.0100 **	0.0077
3	-0.0017	0.0043	0.0067	-0.0007
4	0.0025	-0.0036	-0.0032	-0.0041
5	-0.0038	-0.0044	-0.0099 *	0.0018
6	-0.0032	-0.0022	-0.0086 **	0.0027
7	-0.0030	0.0058 *	-0.0083 **	-0.0020
8	0.0040	-0.0039 *	-0.0081 *	-0.0012
9	0.0024	0.0048 *	0.0068 *	0.0001
10	0.0010	0.0056	0.0105 **	0.0004
11	0.0009	0.0061 *	-0.0089 *	-0.0010
12	0.0032	0.0061	0.0077 *	0.0008
13	0.0028	-0.0034	0.0107 *	-0.0009
14	0.0041 *	0.0069 *	0.0016	-0.0000
15	0.0011	0.0070	0.0080	0.0007
16	-0.0011	-0.0059	0.0096 *	-0.0005
17	0.0041	-0.0064 *	-0.0053	-0.0005
18	-0.0048	0.0041	0.0085	-0.0007
19	-0.0042 *	0.0053	-0.0110 *	0.0004
20	-0.0020	-0.0074	0.0099 *	0.0009
21	0.0017	0.0069	-0.0085	-0.0006
22	-0.0047 *	0.0071	-0.0092	0.0004
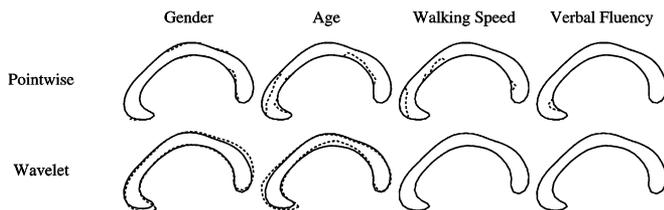
Fig. 6. Results for all four clinical outcome variables using direct component-wise approach (top row) and wavelet coefficient approach (bottom row), showing mean shape (solid) versus a more female shape (dashed). Methods seem inferior to proposed method in terms of statistical power, specificity, and interpretability.

A. Method

The results presented in Fig. 4 suggest that the SPCA method is a useful method for deriving localized and interpretable patterns of variability. The computational complexity is reasonable in the present case of relatively many observations, but limited dimensionality. Computation times varied from seconds, for low scale deformations, to minutes, for more complex cases. Convergence also seems to vary considerably, with almost immediate convergence in some cases and slower and more irregular convergence in others. Alternative or approximate optimization schemes for the SPCA criterion in (7) should be a focus of future work. For application to higher dimensional data, we supply a discussion as follows.

Splitting the testing procedure performed to relate spatial deformations to clinical outcome data into a series of univariate tests has both benefits and drawbacks. Most importantly, it provides a strong form of regularization. Each model contains a low number of variables (one plus any covariates), making the analysis more stable in cases with few observations. The main disadvantage is that this analysis disregards the correlation structure between variables. However, PCA scores are uncorrelated and are therefore unaffected by this property. SPCA scores generally show stronger patterns of correlation and the SPCA analysis may be more notably influenced by this limitation. Estimation methods that take the correlation structure between spatial variables into consideration are also relevant for further investigation.

Two alternative methods for a localized analysis of anatomy were outlined. Arguably, the results obtained using these methods (cf., Fig. 6) were inferior to those of the proposed method. The point-based method suffers from two apparent disadvantages. The high number of degrees of freedom makes the method prone to overfitting. Disparate results may be obtained for adjacent points, leading to variational patterns that are scattered or irregular, and therefore difficult to interpret. The SPCA method circumvents this problem by making sure that each variable represents an anatomically meaningful pattern over several data points. The second problem is the high number of variables. Procedures for adjustment for multiple comparisons such as Bonferroni correction or the permutation method outlined in Section II-C tend to adjust more for more high-dimensional models, effectively resulting in lower levels of significance. The discouraging results obtained using the wavelet representation seem to be due to the spatial appearance of the derived variables, which look implausible from an anatomical viewpoint (cf., Fig. 1). The poor results may therefore be due to an improper choice of mother wavelet. The first order coiflet was used here, because of its low complexity and high degree of symmetry. Reissell presents a type of wavelet called pseudocoiflet [43], which is custom-designed for curve and surface representation and may be a more suitable choice. Further, the wavelet representation also suffers from multiple testing problems, as the number of variables involved is equal to the number of variables in the original data. To alleviate this, the wavelet representation can either be truncated, or separate analyses can be performed at each wavelet scale. Preliminary tests using the latter approach did not point to an improvement in the results.

There exist a few interesting alternatives to SPCA to construct sparse representations of anatomy, most notably independent component analysis (ICA) [31] and varimax rotated principal components [30]. Some experiments using these bases

have been carried out, with results similar to those of SPCA. One disadvantage shared by both ICA and factor rotation is that the patterns produced are only approximately sparse. The residual variation makes the results more difficult to interpret.

1) Extension to 3-D and Higher Dimensions: The CC outlines used here to validate the method are represented by planar shape data. However, the outline of the method, from the extraction of spatially sparse and meaningful features to the subsequent analysis of the relation of these to clinical data, is applicable to data of any dimension, modality, and topology, given that its distribution is suitable for linear modeling. With an increasing number of variables, such as for shape data in three dimensions, comes an increase in computational burden and memory requirements. The core problem for most SPCA algorithms is the need to calculate and store the $p \times p$ covariance matrix of the variables involved. The algorithm presented here uses sequential up- and down-dating of the Cholesky factorization of the covariance matrix [59], such that only currently active variables are considered. With k active variables, this limits the storage requirement to a $k \times k$ matrix. The complexity of the algorithm is therefore more due to the number of nonzero components than to the total number variables involved.

In cases where a very large number of variables must be considered, such as for complex shape representations in three dimensions or for functional MRI analyses, the optimization problem in (7) becomes too complex and the alternating estimation algorithm will not converge. It turns out that the criterion in (7) is valid for any positive value of λ and that the solutions are not particularly dependent on the choice of this parameter [17], [25]. Specifically, a computationally efficient algorithm emerges for $\lambda = +\infty$. In this case, the complex elastic net process to estimate $\mathbf{B}$ can be replaced by a simpler soft-thresholding rule

$$\mathbf{b}_j = \left(\left| \mathbf{a}_j^T \mathbf{X}^T \mathbf{X} \right| - \frac{\delta_j}{2} \right)_+ \mathrm{Sign}\left(\mathbf{a}_j^T \mathbf{X}^T \mathbf{X} \right) \qquad (13)$$

where $(\cdot)_+ = \max(0, \cdot)$ and $\mathbf{a}_j$ is the jth column of $\mathbf{A}$. Note that the $p \times p$ matrix $\mathbf{X}^T \mathbf{X}$ does not need to be explicitly calculated and stored if the matrix operations are properly ordered. Some preliminary results on using this method for exploratory analyses of fMRI data can be found in [60].

SPCA and its related methods for regression are available as add-on packages for the statistical environment R.[1]

B. Clinical Application

We will now comment on the results for the application of the method on the CC data. These comments are provided to support the method only, a more thorough clinical investigation with subsequent interpretation is deferred to a separate paper. The sexual dimorphism of the CC is a closely investigated subject that has yielded disparate results. However, several authors [37], [39], [46], [48] report on a more bulbous splenium for

[1]Similar implementations for the MATLAB platform are available from www.imm.dtu.dk/~kas/software/spca.

females. The present results clearly agree with this finding. The results also agree with the male/female mean shape differences depicted in Fig. 3. The advantage of using the proposed method is the additional information on localization. In a number of limited regions along the boundary, the method quantifies the strength of the relevant discrepancies, giving more detailed anatomical information. Moreover, any global method such as measures of callosal area or the Procrustes distance measure used in this paper may not prove to be significant if the differences are small and highly localized. Using sparse decomposition, such differences can be identified and quantified correctly.

The deformation of the CC corresponding to the measure of walking speed provides an example that nicely demonstrates the potential of the method. In the third row of Fig. 5, some thinning can be seen in the genu area, but more interestingly, a clear deformation is also present in the rostral body, corresponding well to the area of the CC containing fibers related to the motor cortex (cf., Fig. 2). All SPCA scales show this effect to some extent.

The results for verbal fluency did not reach significant levels when corrected for multiple comparisons. In Fig. 5, the corresponding unadjusted deformations for $p < 0.05$ are shown. Although not highly significant, the results again make anatomical sense. On scales SPCA(5) and SPCA(20), a thinning of the isthmus subregion occurs. Referring to Fig. 2, this seems to correspond to atrophy of the fiber tissue connecting to brain regions involved in auditory tasks. This result is also in accordance with previous results based on the same data set [36], where verbal fluency was found to correlate exclusively with the rostrum and isthmus regions. The latter paper used measures of callosal area based on a partitioning the CC into subregions and declared significance at level $\alpha = 0.01$, not corrected for multiple comparisons.

The deformation modes extracted using PCA did not provide much interpretational value in this application. For gender and age, no deformations correlated significantly with the outcome. For walking speed and verbal fluency, PCA yielded some significant results, but the limited interpretational power becomes apparent in the results. Effects are present throughout the entire boundary, and inference of structure-function relationships become difficult.

V. CONCLUSION

SPCA is introduced as an attractive method for extracting strictly sparse and anatomically meaningful variables from a data set. While the results may be interesting for direct analysis, this paper shows how to relate these spatial variables to clinical outcome data, making it possible to derive typical deformation patterns related to e.g., pathology. As an illustrative example, results are presented on the basis of a large data set of CC outlines for several clinical target variables, demonstrating the capabilities of the method. The method has been compared to both a simple point-based alternative as well as decomposition using a wavelet transform. The results suggest that these methods are either less precise, or offer inferior interpretability compared to the sparse principal component analysis approach.

APPENDIX
LIST OF PARTICIPANTS IN THE LADIS STUDY

Helsinki, Finland (Memory Research Unit, Department of Neurology, Helsinki University): Timo Erkinjuntti, M.D., Ph.D., Tarja Pohjasvaara, M.D., Ph.D., Pia Pihanen, M.D., Raija Ylikoski, Ph.D., Hanna Jokinen, Lic.Psych., Meija-Marjut Somerkoski, M.Psych., Riitta Mäntylä, M.D., Ph.D., Oili Salonen, M.D., Ph.D.; *Graz, Austria* (Department of Neurology and MRI Institute, Medical University): Franz Fazekas, M.D., Reinhold Schmidt, M.D., Stefan Ropele, Ph.D., Alexandra Seewann, M.D., Katja Petrovic, MagPsychol, Ulrike Garmehi; *Lisbon, Portugal* (Serviço de Neurologia, Centro de Estudos Egas Moniz, Hospital de Santa Maria): José M. Ferro, M.D., Ph.D., Ana Verdelho, M.D., Sofia Madureira, PsyD; *Amsterdam, The Netherlands* (Department of Neurology, VU Medical Center): Philip Scheltens, M.D., Ph.D., Ilse van Straaten, M.D., Alida Gouw, M.D., Wiesje van der Flier, Ph.D., Frederik Barkhof, M.D., Ph.D.; *Gothenburg, Sweden* (Institute of Clinical Neuroscience, Goteborg University): Anders Wallin, M.D., Ph.D., Michael Jonsson, M.D., Karin Lind, M.D., Arto Nordlund, PsyD, Sindre Rolstad, PsyD, Kerstin Gustavsson, RN; *Huddinge, Sweden* (Neurotec Department, Section of Clinical Geriatrics, Karolinska Universitetssjukhuset): Lars-Olof Wahlund, M.D., Ph.D., Militta Crisby, M.D., Ph.D., Anna Pettersson, physiotherapist, Kaarina Amberla, PsyD; *Paris, France* (Department of Neurology, Hopital Lariboisiere): Hugues Chabriat, M.D., Ph.D., Ludovic Benoit, M.D., Karen Hernandez, Solene Pointeau, Annie Kurtz, Daniel Reizine, M.D.; *Mannheim, Germany* (Department of Neurology, University of Heidelberg, Klinikum Mannheim): Michael Hennerici, M.D., Christian Blahak, M.D., Hansjoerg Baezner, M.D., Martin Wiarda, PsyD, Susanne Seip, RN; *Copenhagen, Denmark* (Copenhagen University Hospital: Memory Disorders Research Unit, Department of Neurology, Rigshospitalet, and the Danish Magnetic Resonance Research Center, Hvidovre Hospital): Gunhild Waldemar, M.D., DM.Sc., Egill Rostrup, M.D., M.Sc., Charlotte Ryberg, M.Sc., Tim Dyrby, M.Sc., Olaf B. Paulson, M.D., DM.Sc.; *Newcastle-upon-Tyne, UK* (Institute for Ageing and Health, University of Newcastle): John O'Brien, DM, Sanjeet Pakrasi, MRCPsych, Thais Minnet, Ph.D., Michael Firbank, Ph.D., Jenny Dean, Ph.D., Pascale Harrison, BSc, Philip English, DCR. The coordinating center is in *Florence, Italy* (Department of Neurological and Psychiatric Sciences, University of Florence): Domenico Inzitari, M.D. (Study Coordinator); Leonardo Pantoni, M.D., Ph.D., Anna Maria Basile, M.D., Giovanna Carlucci, M.D., Michela Simoni, M.D., Giovanni Pracucci, M.D., Monica Martini, M.D., Eliana Magnani, M.D., Anna Poggesi, M.D., Luciano Bartolini, Ph.D., Emilia Salvadori, Ph.D., Marco Moretti, M.D., Mario Mascalchi, M.D., Ph.D. The LADIS Steering Committee is formed by Domenico Inzitari, M.D., Timo Erkinjuntti, M.D., Ph.D., Philip Scheltens, M.D., Ph.D., Marieke Visser, M.D., Ph.D., and Peter Langhorne, Ph.D.

ACKNOWLEDGMENT

REFERENCES

[1] F. Bookstein, "Shape and the information in medical images: A decade of the morphometric synthesis," in *Proc. Workshop Math. Methods Biomed. Image Anal.* , 1996, pp. 2–12.

[2] F. Bookstein, "Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape," *Med. Image Anal.*, vol. 1, no. 3, pp. 225–243, 1997.

[3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—Their training and application," *Comput. Vision Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.

[4] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*. New York: Wiley, 1999.

[5] L. H. Staib and J. S. Duncan, "Boundary finding with parametrically deformable models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 11, pp. 1061–1075, Nov. 1992.

[6] C. Brechbuhler, G. Gerig, and O. Kubler, "Parametrization of closed surfaces for 3-D shape description," *Comput. Vision Image Understand.*, vol. 61, no. 2, pp. 154–170, 1995.

[7] F. Bookstein, "The line-skeleton," *Comput. Graphics Image Process.*, vol. 11, no. 2, pp. 123–137, 1979.

[8] P. Golland, W. Eric, and L. Grimson, "Fixed topology skeletons," in *Proc. Comput. Vision Pattern Recognition Conf.*, 2000, vol. 1, pp. 10–17.

[9] G. Borgefors, "Distance transformations in digital images," *Comput. Vision, Graphics, Image Process.*, vol. 34, no. 3, pp. 344–371, Jun. 1986.

[10] M. Leventon, W. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2000, vol. 1, pp. 316–323.

[11] J. Ashburner and K. J. Friston, "Nonlinear spatial normalization using basis functions," *Human Brain Mapp.*, vol. 7, no. 4, pp. 254–266, 1999.

[12] C. Studholme, V. Cardenas, R. Blumenfeld, N. Schuff, H. Rosen, B. Miller, and M. Weiner, "Deformation tensor morphometry of semantic dementia with quantitative validation," *NeuroImage*, vol. 21, no. 4, pp. 1387–1398, 2004.

[13] R. Davies, C. Twining, P. Allen, T. Cootes, and C. Taylor, "Shape discrimination in the hippocampus using an MDL model," in *Proc. Inf. Process. Med. Imag., IPMI 2003*, vol. 18, pp. 38–50.

[14] A. Kelemen, G. Szekely, and G. Gerig, "Three-dimensional model-based segmentation of brain MRI," in *Proc. Biomed. Image Anal.*, 1998, pp. 4–13.

[15] L. Le Briquer and J. Gee, "Design of a statistical model of brain shape," in *Proc. 15th Int. Conf. Inf. Process. Med. Imag.*, 1997, pp. 477–82.

[16] B. S. Peterson, P. A. Feineigle, L. H. Staib, and J. C. Gore, "Automated measurement of latent morphological features in the human corpus callosum published online 21 Feb. 2001," *Human Brain Mapp.*, vol. 12, no. 4, pp. 232–245, 2001.

[17] K. Sjöstrand, M. Stegmann, and R. Larsen, "Sparse principal component analysis in medical shape modeling," in *Proc. Int. Symp. Med. Imag.*, San Diego, CA, Feb. 2006, vol. 6144.

[18] C. Chennubhotla and A. Jepson, "Sparse PCA extracting multi-scale structure from data," in *Proc. IEEE Int. Conf. Computer Vision*, 2001, vol. 1, pp. 641–647.

[19] A. d'Aspremont, L. E. Ghaoui, M. I. Jordan, and G. R. G. Lanckriet, "A direct formulation for sparse PCA using semidefinite programming," in *A Direct Formulation for Sparse PCA Using Semidefinite Programming 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds. Cambridge, MA: MIT Press, 2005, pp. 41–48.

[20] R. Hausman, "Constrained multivariate analysis," in *Optimization in Statistics*, ser. Studies in the management sciences, S. Zanakis and J. Rustagi, Eds. Amsterdam, The Netherlands: North-Holland, 1982, pp. 137–151.

[21] I. Jolliffe, N. Trendafilov, and M. Uddin, "A modified principal component technique based on the LASSO," *J. Computational Graph. Stat.*, vol. 12, no. 3, pp. 531–547, 2003.

[22] B. Moghaddam, Y. Weiss, and S. Avidan, "Spectral bounds for sparse PCA: Exact and greedy algorithms," in *Advances in Neural Information Processing Systems 18*, Y. Weiss, B. Scholkopf, and J. Platt, Eds. Cambridge, MA: MIT Press, 2006, pp. 915–922.

[23] V. Rousson and T. Gasser, "Simple component analysis," *J. R. Stat. Soc.: Series C (Appl. Stat.)*, vol. 53, no. 4, pp. 539–555, 2004.

[24] S. Vines, "Simple principal components," *J. R. Stat. Soc.: Series C (Appl. Stat.)*, vol. 49, no. 4, pp. 441–451, 2000.

[25] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Computational Graph. Stat.*, vol. 15, no. 2, pp. 265–265, 2006.

[26] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc.: Series B (Stat. Methodol.)*, vol. 67, no. 2, pp. 301–320, 2005.

[27] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. R. Stat. Soc.—Series B Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[28] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–451, 2004.

[29] A. Machado, J. Gee, and M. Campos, "Structural shape characterization via exploratory factor analysis," *Artif. Intell. Med.*, vol. 30, no. 2, pp. 97–118, 2004.

[30] M. Stegmann and K. Sjöstrand, "Sparse modeling of landmark and texture variability using the orthomax criterion," in *Proc. Int. Symp. Med. Imag.*, San Diego, CA, Feb. 2006.

[31] M. Üzümcü, A. Frangi, J. Reiber, and B. Lelieveldt, J. F. M. Sonka, Ed., "The use of independent component analysis in statistical shape models," in *Proc. Int. Symp. Med. Imag.*, San Diego, CA, 2003, vol. 5032.

[32] P. Yushkevich, S. Joshi, S. Pizer, J. Csernansky, and L. Wang, "Feature selection for shape-based classification of biological objects," in *Inf. Process. Med. Imag. (IPMI)*, C. Taylor and J. Noble, Eds. New York: Springer-Verlag, 2003, vol. 2732, Lecture Notes in Computer Science, pp. 114–125.

[33] J. Stoeckel and G. Fung, "SVM feature selection for classification of spect images of Alzheimer's disease using spatial information," in *Proc. 5th IEEE Int. Conf. Data Mining*, Nov. 2005.

[34] Y. Fan, D. Shen, R. C. Gur, R. E. Gur, and C. Davatzikos, "Compare: Classification of morphological patterns using adaptive regional elements," *IEEE Tran. Med. Imag.*, vol. 26, no. 1, pp. 93–105, Jan. 2007.

[35] L. Pantoni, A. M. Basile, G. Pracucci, K. Asplund, J. Bogousslavsky, H. Chabriat, T. Erkinjuntti, F. Fazekas, J. M. Ferro, M. Hennerici, J. O'brien, P. Scheltens, M. C. Visser, L. O. Wahlund, G. Waldemar, A. Wallin, and D. Inzitari, "Impact of age-related cerebral white matter changes on the transition to disability—The LADIS study: Rationale, design and methodology," *Neuroepidemiology*, vol. 24, no. 1-2, pp. 51–62, 2005.

[36] H. Jokinen, C. Ryberg, H. Kalska, R. Ylikoski, E. Rostrup, M. B. Stegmann, G. Waldemar, S. Madureira, J. M. Ferro, E. C. W. v. Straaten, P. Scheltens, F. Barkhof, F. Fazekas, R. Schmidt, L. Pantoni, D. Inzitari, and T. Erkinjuntti, "Corpus callosum atrophy is associated with mental slowing and executive deficits in subjects with age-related white matter hyperintensities. The LADIS study," *J. Neurol., Neurosurgery, Psychiatry*, 2006.

[37] C. Davatzikos, M. Vaillant, S. Resnick, J. Prince, S. Letovsky, and R. Bryan, "A computerized approach for morphological analysis of the corpus callosum," *J. Comput. Assisted Tomogr.*, vol. 20, no. 1, pp. 88–97, 1996.

[38] A. Machado and J. Gee, "Atlas warping for brain morphometry," in *Proc. SPIE—Int. Soc. Opt. Eng.*, 1998, vol. 3338, pp. 642–51.

[39] A. Dubb, R. Gur, B. Avants, and J. Gee, "Characterization of sexual dimorphism in the human corpus callosum," *NeuroImage*, vol. 20, no. 1, pp. 512–519, Sep. 2003.

[40] P. Golland, W. Grimson, M. Shenton, and R. Kikinis, "Deformation analysis for shape based classification," in *Proc. 17th Int. Conf. Inf. Process. Med. Imag.*, 2001, vol. 2082, pp. 517–30.

[41] P. Golland, W. Grimson, M. Shenton, and R. Kikinis, "Detection and analysis of statistical differences in anatomical shape," *Med. Image Anal.*, vol. 9, no. 1, pp. 69–86, 2005.

[42] S. Joshi, S. Pizer, P. Fletcher, P. Yushkevich, A. Thall, and J. Marron, "Multiscale deformable model segmentation and statistical shape analysis using medial descriptions," *IEEE Trans. Med. Imag.*, vol. 21, no. 4, pp. 538–550, Apr. 2002.

[43] L.-M. Reissell, "Wavelet multiresolution representation of curves and surfaces," *Graphical Models Image Process.*, vol. 58, no. 3, pp. 198–217, 1996.

[44] C. Davatzikos, X. Tao, and D. Shen, "Hierarchical active shape models, using the wavelet transform," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 414–423, Mar. 2003.

[45] I. Daubechies and Y. Meyer, "Ten lectures on wavelets," *Bull. Am. Math. Soc.*, vol. 28, no. 2, pp. 350–359, 1993.

[46] L. Allen, M. Richey, Y. Chai, and R. Gorski, "Sex differences in the corpus callosum of the living human being," *Neuroscience*, vol. 11, no. 4, pp. 933–942, Apr. 1991.

[47] P. Bermudez and R. Zatorre, "Sexual dimorphism in the corpus callosum: Methodological considerations in MRI morphometry," *NeuroImage*, vol. 13, no. 6, pp. 1121–1130, 2001.

[48] S. Clarke, H. v. d. L. R. Kraftsik, and G. Innocenti, "Forms and measures of adult and developing human corpus callosum: Is there sexual dimorphism," *J. Comparative Neurol.*, vol. 280, pp. 213–230, 1989.

[49] S. Witelson, "Hand and sex differences in the isthmus and genu of the human corpus callosum: A postmortem morphological study," *Brain*, vol. 112, no. 799-835, 1989.

[50] C. Ryberg, E. Rostrup, M. B. Stegmann, F. Barkhof, P. Scheltens, E. C. W. v. Straaten, F. Fazekas, R. Schmidt, J. M. Ferro, H. Baezner, T. Erkinjuntti, H. Jokinen, L. Wahlund, J. OŠBrien, A. M. Basile, L. Pantoni, D. Inzitari, and G. Waldemar, "Clinical significance of corpus callosum atrophy in a mixed elderly population," *Neurobiol. Aging*, 2006.

[51] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation from nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[52] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag, 2001.

[53] T. Nichols and A. Holmes, "Nonparametric permutation tests for functional neuroimaging: A primer with examples," *Human Brain Mapp.*, vol. 15, no. 1, pp. 1–25, 2002.

[54] A. Davison and D. Hinkley, *Bootstrap Methods and Their Application*, 5th ed. Cambridge, MA: Cambridge Univ. Press, 2003.

[55] D. Kendall, "The diffusion of shape," *Advances Appl. Probability*, vol. 9, pp. 428–430, 1977.

[56] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.

[57] M. Stegmann, B. Ersbøll, and R. Larsen, "FAME—A flexible appearance modelling environment," *IEEE Trans. Med. Imag.*, vol. 22, no. 10, pp. 1319–1331, Oct. 2003.

[58] E. Zaidel and M. Iacoboni, *The Parallel Brain: The Cognitive Neuroscience of the Corpus Callosum*, ser. Issues in Clinical and Cognitive Neuropsychology, M. Iacoboni, Ed. New York: Bradford, 2003.

[59] G. Golub and C. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[60] K. Sjöstrand, T. E. Lund, K. H. Madsen, and R. Larsen, "Sparse PCA, a new method for unsupervised analyses of fMRI data," in *Proc. Int. Soc. Magnetic Resonance In Medicine*, Berkeley, CA, May 2006.