Published in final edited form as:

Artif Intell Med. 2009 September ; 47(1): 75–86. doi:10.1016/j.artmed.2009.05.007.

Morphometric analysis of brain images with reduced number of statistical tests: a study on the gender-related differentiation of the corpus callosum

Despina Kontos^{1,3}, Vasileios Megalooikonomou¹, and James C. Gee²

¹ Data Engineering Laboratory (DEnLab), Center for Information Science and Technology, Temple University, 1805 N. Broad St., Philadelphia, PA, 19122, USA, phone: +1 215 204 5774, fax: +1 215 204 5082, email: vasilis@temple.edu

² University of Pennsylvania, Department of Radiology, 3600 Market Street, Suite 370, Philadelphia, PA 19104-2644, phone: +1 215 662 7109, fax: +1 215 615 3681, email: gee@mail.med.upenn.edu

Summary

Objective—We evaluate the feasibility of applying dynamic recursive partitioning (DRP), an image analysis technique, to perform morphometric analysis. We apply DRP to detect and characterize discriminative morphometric characteristics between anatomical brain structures from different groups of subjects. Our method reduces the number of statistical tests, commonly required by pixel-wise statistics, alleviating the effect of the multiple comparison problem.

Methods and Materials—The main idea of DRP is to partition the two-dimensional (2D) image adaptively into progressively smaller sub-regions until statistically significant discriminative regions are detected. The partitioning process is guided by statistical tests applied on groups of pixels. By performing statistical tests on groups of pixels rather than on individual pixels, the number of statistical tests is effectively reduced. This reduction of statistical tests restricts the effect of the multiple comparison problem (*i.e.* type-I error). We demonstrate an application of DRP for detecting gender-related morphometric differentiation of the corpus callosum. DRP was applied to template deformation fields computed from registered magnetic resonance images of the corpus callosum in order to detect regions of significant expansion or contraction between female and male subjects.

Results—DRP was able to detect regions comparable to those of pixel-wise analysis, while reducing the number of required statistical tests up to almost 50%. The detected regions were in agreement with findings previously reported in the literature. Statistically significant discriminative morphological variability was detected in the posterior corpus callosum region, the isthmus and the anterior corpus callosum. In addition, by operating on groups of pixels, DRP appears to be less prone to detecting spatially diffused and isolated outlier pixels as significant.

Conclusion—DRP can be a viable approach for detecting discriminative morphometric characteristics among groups of subjects, having the potential to alleviate the multiple comparisons' effect by significantly reducing the number of required statistical tests.

3Present address, Corresponding address: Department of Radiology, Hospital of the University, of Pennsylvania, 1 Silverstein Building, 3400 Spruce Street Philadelphia, PA 19104-4206, phone: +1 215 746 8759, fax: +1 215 746 8764, email: despina.kontos@uphs.upenn.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

image analysis; feature extraction; dynamic recursive partitioning; multiple comparison problem; pixel-wise statistics; template deformation morphometry; corpus callosum; magnetic resonance imaging

1. Introduction

Visualizing human physiology with modern imaging technologies has offered valuable insight to understanding anatomy, function and the development of several diseases. Considering in particular the advances in brain imaging [1], great progress has been achieved during the recent years in understanding how anatomical structures are associated to function [2,3] and how the cognitive process is generated [4]. Fascinating insight has been gained by interpreting the process of development [5], and identifying the effects of pathology and aging [6]. To date, most of the research in brain imaging has been based on computerized analysis of brain imaging modalities [2], such as magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), and diffusion tensor imaging (DTI); the goal of many studies has been to identify anatomical or functional differences between different populations, such as healthy individuals and patients [7,8].

Statistical parametric mapping (SPM) analysis is one of the most common approaches that has been used for brain image analysis [8–12]. SPM analysis can be performed to detect differences between images of separate groups of subjects by analyzing each pixel's changes independently and building a corresponding map of statistical values. To ascertain the discriminatory significance of each pixel, a statistical test such as the t-test, ranksum test, or the F-test is applied. A p-value is obtained for each pixel that indicates how likely it is that the pixel's variability across classes is observed by chance. Clustering is often employed in the process to construct highly informative regions with respect to classification.

One of the drawbacks of SPM analysis is that pixel-wise analysis usually requires a large number of statistical tests to be performed. The increased number of statistical tests increases the probability that certain tests will appear positive simply due to chance (*i.e.* type-I error); this effect is usually referred to as the multiple comparison problem. Consider, for example, a 256×190 two-dimensional (2D) MRI image acquired with approximately $0.9 \times 0.9 \text{ mm}^2$ pixel resolution: with a standard 0.01 type-I error, the required 48640 pixel-wise statistical tests could result to approximately 486 false positives; depending on the spatial distribution of these falsely detected pixels, the spatial extent of the effect could be considerable.

Several approaches have been proposed in the literature for controlling the false positive rate [13]. Perhaps the most commonly used approach is the Bonferroni correction [14], which is a rather conservative procedure: the nominal significance level α is replaced with the level α/n for each test, where n is equal to the total number of statistical tests performed; it can be shown that the Bonferroni correction has strong control of type-I error. In practice, when applied to brain image analysis, Bonferroni correction tends to eliminate several true positives along with the false positives. For this reason heuristic modifications such as the sequential Bonferroni correction have been proposed [15]. Clustering is usually applied to detect and discard outlier pixels when constructing discriminative regions. In addition to the multiple comparison problem, pixel-wise statistics appear to be significantly biased towards distribution differences that are highly localized in space [16].

SPM analysis has been used in combination with template deformation morphometry (TDM) for quantifying structural variations in anatomical structures of the human brain. In TDM,

registration of a template brain image is performed to match each individual subject in the dataset [17–19]; pixel-wise information is extracted about the degree of expansion or contraction of each pixel during the registration process. Essentially, each image is mapped to a deformation vector field, which can be quantified to a scalar measurement by computing the corresponding Jacobians in the template space. By applying pixel-wise statistics on the Jacobians, an SPM can be generated and a significance threshold can be applied to detect discriminative morphometric variations.

One of the structures in the human brain that has attracted a lot of research in the past decades is the corpus callosum. The corpus callosum can be easily identified as a white matter structure in the mid-sagittal section of the brain [20]. It is a structure that facilitates primarily the communication between the two cerebral hemispheres of the human brain, being of critical importance when interpreting the neurological process of cognitive tasks. Studies have supported that the corpus callosum is critically engaged in the development of disorders such as schizophrenia [21] and Alzheimer's disease [22]. Research on the gender-based dimorphism of the corpus callosum has also raised significant discussion during the past decades [23]; being critical to inter-hemispheric communication, the corpus callosum has often been accounted for differences in cognition between males and females. Many investigators have sought to identify variation on the overall size of the corpus callosum when examining male and female populations [24]. More recent studies have investigated local morphological gender-related differentiation using template deformation morphometry and pixel-based analysis [18,25,26].

In this paper we evaluate the feasibility of applying dynamic recursive partitioning (DRP), an image analysis technique suitable for detecting discriminative image regions between groups of subjects, to perform morphometric analysis. The main idea of DRP is to partition the image adaptively into progressively smaller sub-regions until statistically significant discriminative regions are detected. The partitioning process is guided by statistical tests that are applied to groups of pixels rather than to individual pixels; for this reason the number of statistical tests is effectively reduced compared to SPM analysis. Depending on the required degree of control for type-I error, p-value correction methods can also be incorporated to further restrict the effect of the multiple comparison problem.

The algorithmic outline of DRP was initially introduced and evaluated primarily with synthetic and realistic images [27,28]. Preliminary evaluation of DRP with functional brain images showed its potential to be used effectively for medical image analysis; DRP was able to reduce the number of statistical tests by two orders of magnitude compared to pixel-wise statistics, while also improving classification accuracy by 15% [8,27]. Here, we evaluate the feasibility of DRP for anatomical imaging, and particularly for performing morphometric analysis. Morphometry introduces different challenges than functional imaging. While patterns of functional activity are usually analyzed within the entire image region in functional imaging, in morphometry, template deformation statistics are usually computed from well-defined anatomical structures as a necessary first level of analysis; DRP can then be applied as a postprocessing second level of analysis. A preliminary report on this study was previously presented by Kontos *et al.* [29].

2. Materials and Methods

2.1 Data and Preprocessing

Our dataset included 2D MRI midsagittal slices acquired from 93 healthy female and 93 healthy male right-handed individuals. The images were obtained from the Schizophrenia Center database of the University of Pennsylvania [26]. The MRI images were acquired on a 1.5T GE scanner (TR=35, TE=6, flip=35, slices= 1 × 1mm, FOV=24cm). Transaxial images were in planes parallel to the orbitomeatal line, with resolution of $0.9375 \times 0.9375 \text{ mm}^2$. A k-means

clustering algorithm which groups image voxels into k clusters by minimizing the intensity variance within each cluster was employed to segment the corpus callosa [30]. Because the performance of this algorithm was highly dependent on the brightness of the input image and number of clusters, we opted to perform cluster segmentation on the same data set using $k = 3, 4$, and 5 and scaled at multiple brightness levels. This process yielded 18 sets of segmentation images (3 segmentation levels $\times 6$ brightness settings). For each mid-sagittal image, the best segmentation image from the set of 18 was chosen and manually edited so that the callosa was completely separated from surrounding structures. This segmentation image was then used to mask and extract the corpus callosum from the mid-sagittal section. An additional male's subject callosum was used as a template to perform registration of our dataset. Note that since the shape of the callosa are highly stereotypical and well within the capacity of the registration method's ability to establish robust correspondences between them, the template anatomy does not bias the registration. This robustness to bias has been documented in several previous studies where this particular method has been applied [18,26,30,31]. The template callosum was mapped through pixel-wise transformation to each subject's callosum using an elastic-membrane model [32], and the template deformation vector fields were computed for each subject. All of the registration results were carefully inspected for anatomic accuracy.

In order to quantify the degree of contraction or expansion required at each pixel of the atlas for each subject, the Jacobians of the template deformation vectors were computed; this is a procedure that is commonly used in TDM [18,19,26]. To further compensate for the overall size variation between female and male callosum, we applied a normalization with respect to the total cross-sectional callosal area. We rescaled the Jacobian value at each pixel of every subject by dividing with the sum of the Jacobians in the entire callosal area, so that the overall callosal area is consistent; Figure 1 shows an example of a normalized Jacobian image. This normalization allows the analysis applied further by DRP to focus on local morphometric differentiation among the male and female subjects that are relatively compared to the overall callosal size; differences that are attributed to overall callosum size variation and scaling are not considered. This type of analysis has been previously shown to reveal statistically significant regions when applying TDM to a smaller subset of our dataset [18]; the details of the registration process and the Jacobian computations have been described previously by Pettey and Gee [18].

2.2 Dynamic Recursive Partitioning

We applied DRP to the callosum images of the Jacobian determinants computed from the template deformation fields. To ensure a clear description of the methodology, we elaborate DRP for the two-class problem of our corpus callosum dataset using two-sample statistical tests; an extension to more than two classes can also be implemented using appropriate statistics, such as one-way analysis of variance (ANOVA) [33]. We will refer to one class as M (male) and the other class as F (female), having $N_M=93$ and $N_F=93$ number of images respectively.

The main idea of DRP is to partition 2D image adaptively into progressively smaller subregions until regions with statistically significant morphological variability are detected. The entire image is initially treated as one region (*i.e.* rectangle). An adaptive quad-tree [34] splitting of the space into smaller regions is performed; a region is partitioned by dividing each spatial dimension equally into two. The selectivity of the partitioning is guided by statistical tests; a region is partitioned if the statistical test indicates that the region does not have enough morphological differentiation between the two classes (*i.e.*, males and females).

In this particular application, we consider as a representative feature of every region the median V_{median}^J of the corresponding Jacobian values; pixels belonging to image background are not considered in this calculation. Because the Jacobian values quantify the degree of expansion

or shrinkage at each pixel, the median of the Jacobian determinants for a group of pixels provides a measure of the central tendency of the pixel-wise template shrinkage or expansion within the particular region. In general, other descriptive statistics, such as the mean, sum and mode, could potentially be used as representative features. However, in the specific application examined here, outlier pixel-wise Jacobian values could introduce significant bias in the characterization of a region's morphological variation when using, for example, the region's mean value as a representative feature; this could be particularly important especially in the first levels of the DRP, where the regions under consideration contain a large number of pixels. For this reason, the median was chosen as a more reliable measure, since it has been shown to be a more robust statistic than the mean [35].

DRP proceeds by partitioning each region if the V_{median}^J feature does not have sufficient discriminative power to distinguish between the two classes. A statistically significant difference among the V_{median}^J features of two classes indicates significant overall morphometric variation for a particular region. The discriminative power of a region is determined by applying a statistical test (*i.e.*, t-test) on the corresponding V_{median}^J features of the two classes. A p-value threshold is assigned in order to define the desired level of discriminatory significance and guide the selectivity of the splitting. DRP detects discriminative regions in a multiresolution manner by operating in a coarse-to-fine grain basis; given the resolution of the image, the partitioning procedure progresses recursively until all remaining rectangles are discriminative or a rectangle becomes so small that it cannot be further partitioned. Figure 2 illustrates the main idea of the adaptive partitioning process; the image regions corresponding to background are excluded from the analysis.

One of the underlying assumptions made for applying DRP is that any patterns of morphological variation in the images are generated by independent (unrelated) distributions. This assumption is realistic considering that the two classes contain images of unpaired subjects generated under different conditions (*i.e.*, male and female subjects in our application). Under this assumption, the following hypothesis testing scenario is examined at each step of the recursive partitioning process for the V_{median}^J features:

H_0 (null hypothesis): The distribution of the V_{median}^J features is the same for the two classes; the region under consideration is not sufficiently discriminative between males and females.

H_1 (alternative hypothesis): The distribution of the V_{median}^J features is not the same for the two classes; the region under consideration is sufficiently discriminative between males and females.

Depending on the properties of the dataset, either parametric or non-parametric statistical tests can be employed for guiding the selectivity of the image partitioning. The unpaired t-test [33] can be used to evaluate the null hypothesis H_0 for the computed V_{median}^J features, in the case that the normality assumption holds for the pixel-wise value distribution across the images of each class, and equal within population variance is observed. Alternatively, in the case that the normality assumption and the equal within population variance cannot be validated for the data, the non-parametric Wilcoxon rank-sum test [33] can be used to test the null hypothesis H_0 for the V_{median}^J features. The Wilcoxon rank-sum test makes no distributional assumption and is based on the sum of ranks of the V_{median}^J features in each of the two classes [33]. Table 1 summarizes all the preconditioned assumptions made for applying DRP.

Compared to pixel-wise analysis, DRP effectively reduces the number of statistical tests; this reduction is due to the fact that the tests are applied selectively to groups of pixels (*i.e.*, rectangles), rather than to individual pixels, focusing only on certain potentially discriminative sub-regions. The DRP reduction of statistical tests is essentially dependent on the dimensionality and the resolution of the original images as well as the desired level of resolution

for detecting discriminative regions, as guided by the particular application at hand. To illustrate this, let us consider a 2D image of $m \times n$ pixels. Let us also denote with L the maximum number of levels that the image partitioning is allowed to proceed; in DRP, this parameter is defined by the user and can be optimally assessed by considering the dimensions of the data and the desired maximum resolution of the detected discriminative ROIs. Through the adaptive partitioning process, we define as l the current level of the DRP partitioning (*i.e.*, $l=0, \dots, L$). We note that $l=0$ corresponds to the root node of the partitioning quad-tree which represents the entire original 2D image.

At each level l of DRP, the maximum number of rectangles that could be considered for further splitting is equal to:

$$R_{l\max} \leq (2^D)^l, \quad (1)$$

where $D=2$ for a 2D image. At each level l , the number of statistical tests $Tests_l$ applied by DRP has an upper bound that is equal to the maximum number of rectangles:

$$Tests_l \leq R_{l\max} = (2^D)^l \quad (2)$$

Hence, the upper bound estimate for the total number of statistical tests that can be applied by DRP is equal to:

$$Tests_{DRP} = \sum_{l=0}^L Tests_l \leq \sum_{l=0}^L (2^D)^l \quad (3)$$

Considering that the minimum desired resolution of DRP partitioning can be defined according to the original image resolution as

$$x_L = \frac{m}{2^L}, \text{ and } y_L = \frac{n}{2^L}, \quad (4)$$

Eq. 3 can provide an estimate for the maximum number of statistical tests in terms of the $m \times n$ dimensions of the original image:

$$Tests_{DRP} \leq \sum_{l=0}^L (2^D)^l = \sum_{l=0}^L (2^D)^{\log_2(m/x_L)} = \sum_{l=0}^L (2^D)^{\log_2(n/y_L)} \quad (5)$$

This quantity can be compared to the number of statistical tests performed by pixel-wise analysis for the same image:

$$Tests_{pixel-wise} = m \times n \quad (6)$$

Control of Type-I Error—While DRP reduces, compared to pixel-wise analysis, the overall number of statistical tests, p-value correction methods can additionally be incorporated when a more stringent control of the type-I error (*i.e.*, false positives) is required. For DRP, we

propose combining the use of Bonferroni correction and false discovery rate (FDR) [36,37] to achieve a more stringent false positive control. DRP typically performs very few statistical tests during the first levels of the partitioning (see Eq. 2). For example DRP has an upper bound of 1, 4, 16, and 64 statistical tests for $l=0, \dots, 3$ respectively; this is a relatively small number of tests compared to pixel-wise statistics. For each of these levels of DRP the standard Bonferroni correction can be applied since the dimensions x_l and y_l of the sub-regions (*i.e.* rectangles) are large enough to safely assume that the feature V_{median}^J of each region is statistically independent and spatially uncorrelated with the features extracted from the other regions [13, 14]. As the number of statistical tests increases exponentially with the levels of partitioning (see Eq. 2), we propose using FDR to explicitly control the type-I error.

FDR is a recently proposed statistical method that controls the number of false positives when multiple hypothesis tests are performed [36,37]. Unlike the commonly used random field approaches that are widely used in pixel-based analysis for controlling false positives [9,38], FDR makes no assumptions for the distribution of the pixel values. The FDR procedure selects an optimal p-value threshold for rejecting the null hypothesis H_0 , which adapts to the properties of the image data. FDR refers to the ratio of the false positive tests among the tests which reject the null hypothesis. In the case of DRP analysis, the FDR ratio can be defined as

$$q = \frac{R_{l \text{ false positive}}}{R_{l \text{ positive}}}, \text{ where } R_{l \text{ false positive}} \text{ are the false positive rectangles and } R_{l \text{ positive}} \text{ the total number of rectangles for which the null hypothesis } H_0 \text{ is rejected at level } l.$$

Assuming that $Tests_l$ statistical tests are being performed at level l , the optimal p-value threshold can be obtained with the following procedure [36,37]:

1. Define an acceptable rate q for FDR ($0 \leq q \leq 1$).
2. Order the p-values obtained from applying statistical tests to the rectangles of level l ; $p(1) \leq p(2) \leq \dots \leq p(Tests_l)$.
3. Find the largest i , $r = \max(i)$ such that $p(i) \leq \frac{i}{Tests_l} \times \frac{q}{c(Tests_l)}$, where $c(Tests_l)$ is a constant depending on the distribution of the p-values obtained for level l . For template deformation Jacobian images it is reasonable to assume positive dependence of pixels. Particularly for DRP, the rectangles during the first partitioning levels $l=0 \dots 4$ are large enough to assume statistical independence. For both of these assumptions it has been proven that $c(Tests_l) = 1$ [36,37].
4. Select the p-value threshold $P_{thres} = p(r)$ and reject the null hypothesis H_0 for the corresponding rectangles having p-values $p(1), \dots, p(r)$.

2.3 Classification

In certain applications, there is a need to perform classification using spatially discriminative properties. Such applications include, but are not limited to, computer aided diagnosis (CAD), similarity retrieval and clustering of similar images [7,18,19,26]. The V_{median}^J feature computed by DRP from the final discriminative regions can be used to construct a characterization vector $f = [V_{median1}^J, V_{median2}^J, \dots, V_{medianN}^J]$, where N equals to the total number of the detected discriminative regions; note that these regions can be of different sizes depending on the level of DRP that they were identified as statistically significant. The characterization vector f can be thought of as a signature representing the discriminative morphological characteristics of each subject. These signatures can be used as inputs to train any established classifier, such as neural networks, decision trees, and Bayes normal classifiers. Classification in this type of analysis can be used to validate the significance of the detected discriminative regions.

3. Results

We applied DPR to the Jacobian determinants computed from the template deformation fields of the male and female callosum images. Due to the nature of our dataset, it was reasonable to assume that the two populations (*i.e.*, male and female images) are most likely to have been generated by normal distributions with equal within class variance; under this assumption, the unpaired t-test could safely be applied. To validate this assumption, we examined the cross-subject pixel-wise value distribution and tested for normality. Note that during the execution of DPR, statistics are applied on aggregate pixel values (*i.e.*, the median) computed from the same (*i.e.*, corresponding) spatial regions across each subject's image within each class. These aggregate values are then compared between the two classes using a statistical test (*i.e.* t-test or rank-sum). Therefore, we are actually interested on whether the normality assumption is valid on a pixel-wise basis across the subjects of each class, rather than for the overall distribution of pixel-values in each class. We applied the Lilliefors test of normality [39] at $\alpha=0.05$ significance level. For the 'male' class of our dataset 82% of the pixels appeared to follow a pixel-wise normal distribution throughout the class' subjects. For the 'female' class of our dataset, 85% of the pixels appeared to follow a pixel-wise normal distribution. Figures 3.a and 3.b show the corresponding pixels in each class that follow a normal distribution (in white). Considering that the majority of the pixels followed a normal distribution, we applied the t-test statistic (which is based on the normality assumption). In addition, in order to study how the selection of statistical test affects the detected discriminative regions, we also applied the non-parametric ranksum test for comparison, which does not rely on normality assumption. The results demonstrated that the detected significant regions are indeed comparable for the two statistical tests (Figures 4–5). We attribute this observation to the relatively large sample of our dataset ($N=93$) which allows for obtaining robust parametric statistics when applying the t-test even when the normality criteria are not as stringent for the minority of pixels appearing as not normal.

More specifically, the experiments were performed using a p-value threshold of $p < 0.01$ and also the most commonly used p-value threshold of $p < 0.05$. The maximum level of partitioning was set equal to $L=5$; this is equivalent to a minimum partitioning resolution of approximately $2 \times 5 \text{ mm}^2$, which can be considered as reasonable for our particular application. Significant morphological shape differentiation was detected in the posterior corpus callosum region also known as the splenium. This specific region was identified as significant in all the different experimental settings (*i.e.*, combination of statistical test and p-value threshold). Besides the splenium, structural variability was also identified for specific experimental settings in parts of the isthmus and the anterior corpus callosum. Figure 4 shows discriminative regions of morphological dimorphism that were identified by DPR, overlaid on the anatomical template. These regions were consistent with previously reported findings in the literature [18,25,26].

We compare these regions with the ones obtained when applying pixel-wise statistical analysis on the same dataset using t-test and ranksum test to estimate significance of each pixel. Figure 5 illustrates the corresponding regions identified by pixel-wise analysis for the same experimental settings; the splenium was detected in all experimental settings, while the anterior part of the callosum as also identified as significant for certain parameters of the analysis. Table 2 shows the number of statistical tests performed by DPR compared to pixel-wise analysis. DPR was able to detect regions comparable to those of pixel-wise analysis, while reducing the number of required statistical tests by more than 50%. In addition, as shown in figures 4 and 5, by operating on a group of pixels, DPR is less prone to detecting spatially diffused and isolated outlier pixels as significant.

We compared the areas detected by DPR when applying the conventional p-value thresholds (*i.e.*, $p < 0.01$ or $p < 0.05$) to the areas detected when using more stringent p-value criteria; we

applied DRP including Bonferroni correction and FDR [36,37] to achieve a more stringent control of type-I error. Bonferroni correction was applied during $l=0, \dots, 3$ levels of partitioning; in the last level of partitioning ($L=4$) FDR was applied with $q=0.54$; the p-value was assigned by FDR to $p=0.0008$ (see section 2.2 for the definition of FDR). As shown in figure 6.a, the lower p-value threshold assigned by FDR resulted in focusing the detected areas only within the area of the splenium; this area has been consistently reported throughout the literature [18,25,26]. To compare, we applied pixel-wise analysis with FDR; parameter q was set at $q=0.53$, and the p-value was assigned to $p=0.0042$. Bonferroni correction was not considered for the pixel-wise statistics, because the total number of statistical tests applied simultaneously is too large (*i.e.* 3219 pixels within the callosum area); Bonferroni correction could result in a p-value adjustment equal to $p\text{-value}=0.01/3219=0.3 \times 10^{-5}$, which has the potential to eliminate several true positives along with the false positives. As shown in figure 6.b the area detected by pixel-wise analysis with FDR is comparable to the area detected by DRP with FDR; the splenium appears to have the most significant morphological differentiation between the male and female groups.

Classification

For the purpose of fully evaluating the detected discriminative areas, we performed classification experiments. We used the V^J_{median} features extracted from the detected discriminative regions in figure 4 as input attributes to train linear and quadratic Bayes normal classifiers; the class labels were set to *male*=1 and *female*=2. We experimented with training set sizes ranging from 65% to 75% of the available data. The classification performance reached up to 60%–65% accuracy. These classification results are in agreement with previous attempts to classify corpus callosum images based on features extracted from sexually dimorphic regions [18]. They also validate the general belief that despite being able to detect regions of gender-based morphologic variability in the corpus callosum, the differentiation is not prevalent enough to rely on these regions for gender classification.

4. Discussion

The results demonstrate that DRP could potentially provide an alternative viable technique for morphometric analysis of anatomical brain structures, particularly when the reduction of statistical tests is a desirable parameter in the application at hand. A potential concern for the applicability of DRP is the rectangular shape of the regions on which the adaptive partitioning process operates. However, the rectangular shape of the DRP regions is not as restrictive as one could suspect. For morphometric analysis where one single and well-defined anatomical structure is of interest (which is usually previously segmented), the detection of rectangular regions is not affected by the potential bias of including several neighboring structures without being able to explicitly attribute corresponding partial effects. In addition, since DRP regions follow the structure of the pixel, when DRP proceeds to a larger depth it has the ability to consider very small regions of a very few pixels or even single pixels. Discriminative regions of different shapes can also be formed by several neighboring discriminative rectangles of possibly different sizes. In the case that the significant sub-regions are small and heterogeneous, (*i.e.*, overlapping different anatomical structures), DRP's p-value threshold could be further reduced, by using the p-value correction techniques, to find the particular anatomic sub-regions that contributes more to the observed group difference. One could also consider combining techniques and stopping DRP when highly heterogeneous subregions are detected to perform pixel-wise analysis only within these regions; a significant overall reduction in the number of statistical tests should also be expected even under this particular scenario.

The advantages and disadvantages of DRP compared to other commonly used approaches can further be examined considering the particular application at hand. For example, approaches

for determining spatially localized differences of a certain shape between groups of subjects could also be based on statistical comparisons between features extracted from the shape's boundary. The problem is that the dimensionality of the feature space is usually so large that the statistical tests' significance threshold must be greatly reduced to protect against detection of spurious group differences resulting in the reduction of sensitivity of these techniques in detecting small shape differences [40]. To address this problem, one needs shape parameterization techniques that are concise, (*i.e.*, capture the most prominent shape characteristics using a small number of parameters), and spatially localized, accounting for the shape of a spatially restricted sub-region in each parameter. These properties are very attractive for parameterization techniques since the statistical power of tests on those parameters is reduced as little as possible by corrections for the multiple comparison problem [41]. Techniques based on principal components analysis (PCA) try to represent shape in a number of spatially-localized sub-regions using the eigenvectors of the covariance matrix of the vector of coordinates.

However, PCA generally leads to global components [42,43]. Independent components analysis (ICA) and principal factor analysis (PFA) do not directly optimize a locality-related objective function when estimating the eigenvectors, and they usually generate spatially-localized components [44,45]. Predefined spatially located regions have also been integrated into PCA [46]. Recently proposed localized component analysis techniques [40] use a linear subspace to reduce the number of variables required for localized shape comparisons boosting the power of statistical tests and also reducing the multiple comparison problem. They simultaneously optimize the eigenvectors for spatial locality and conciseness. DRP is expected to be less computationally expensive than the component or localized component analysis techniques due to the simple calculations involved in the space partitioning and the reduced number of statistical tests applied to the partitions. Finally, by operating on aggregate pixel values (*i.e.* the mean), DRP is less likely to be affected by image registration errors, which usually tend to affect more the approaches that rely on pixel-wise gray level values for analysis.

In terms of the computational complexity and execution time of DRP one should note that the reduction of statistical tests does not necessarily result in a reduction of computational time. The main motivation behind our proposed approach is to reduce the number of statistical tests primarily for alleviating the effect of the multiple comparison problem, and as a secondary objective, to potentially reduce the actual computational time, which will depend on the particular algorithmic implementation. While simple pixel-wise analysis is a much more straightforward approach to implement computationally, DRP, particularly when considering Bonferroni and FDR correction for multiple comparisons, includes the algorithmic implementation of advanced data structures such as the quad-tree [34,47]. Therefore, depending also on the software implementation platform (*i.e.* programming language), the computational resources needed for handling such a data structure might vary. For example, in the application presented in our experiments, where MATLAB® (Mathworks Inc.) was used for implementation, prototyping, and proof-of-concept evaluation, the computational demands could be significant, since MATLAB is an interpretable language. Using a compiled language (*i.e.*, C or C++) could provide faster implementations. Nevertheless, for the dataset used in our evaluations both approaches were relatively fast, in the order of few seconds each. Representative time requirements are 17sec for DRP with Bonferroni and FDR correction for multiple comparisons and 3sec for pixel-wise analysis for the same initialization parameters (*i.e.* same statistical test and significance level). However, as our results demonstrate, while the pixel-wise analysis appears to run faster in our MATLAB environment, it potentially suffers more from the multiple comparison problem, as illustrated by comparing Figures 4 and 5, where scattered isolated pixels are indicated as significant within the corpus callosum area.

DRP could potentially have a particular advantage in the case of three-dimensional (3D) image analysis, where the number of voxel-wise statistics could increase exponentially, having the potential to result in even greater reduction of statistical tests. Preliminary results have indeed demonstrated this advantage of DRP. Evaluation with a pilot dataset of 3D functional brain images showed that DRP can reduce the number of statistical tests by two orders of magnitude, while also outperforming other commonly used medical image classification techniques [8]; DRP outperformed Maximum-Likelihood classification by 20%, Kullback-Leibler classification by 24%, and a static partitioning classification approach by 24% [8,27]. Further work is underway to fully evaluate the applicability and the potential advantages of DRP for 3D medical image analysis.

5. Conclusions

We evaluated the feasibility of applying DRP for detecting discriminative morphometric characteristics between anatomical brain structures of different groups of subjects. In DRP, the image is adaptively partitioned into progressively smaller sub-regions, until regions of significant morphological differentiation are detected. The partitioning process is guided by statistical tests applied on groups of pixels, resulting in significant reduction of statistical tests, compared to the commonly used pixel-wise approaches. Therefore, DRP has the potential to alleviate the effect of the multiple comparison problem (*i.e.* type-I error) in morphometric analysis applications. Here, we applied DRP to the Jacobians of 2D template deformation fields of MRI corpus callosum images acquired from a group of male and female subjects, for the purpose of detecting gender-based morphological variability. DRP detected statistically significant discriminative morphological variability in the posterior corpus callosum region, the isthmus and the anterior corpus callosum; these findings are supported by previous reports in the literature. We compared DRP to pixel-wise morphometric analysis. DRP was able to detect regions comparable to those of pixel-wise analysis, while reducing the number of required statistical tests up to almost 50%.

Acknowledgments

This work was supported in part by NIH Research Grant #1 R01 MH68066-04 funded by NIMH, NINDS and NIA, and by NSF Research Grant IIS-0237921 and Infrastructure Grant ANI-0124390. The funding agencies specifically disclaim responsibility for any analyses, interpretations and conclusions. The authors would also like to thank the anonymous reviewers for their very constructive comments and suggestions.

References

1. Koslow, SH.; Huerta, MF., editors. Neuroinformatics: An Overview of the Human Brain Project. Erlbaum; Mahway, NJ: 1997.
2. Megalooikonomou V, Ford J, Shen L, Makedon F, Saykin A. Data mining in brain imaging. Statistical Methods in Medical Research 2000;9:359–394. [PubMed: 11084713]
3. Letovsky SI, Whitehead SH, Paik CH, Miller GA, Gerber J, Herskovits EH, et al. A brain image database for structure-function analysis. American Journal of Neuroradiology 1998;19:1869–1877. [PubMed: 9874539]
4. Grossman M, Koenig P, DeVita C, Glosser G, Alsop D, Detre J, et al. Neural representation of verb meaning: an fMRI study. Humman Brain Mapping 2002;15:124–134.
5. Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, et al. Brain development during childhood and adolescence: a longitudinal MRI study. Nature Neuroscience 1999;2:861–863.
6. Resnick SM, Pham DL, Kraut MA, Zonderman AB, Davatzikos C. Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. Journal of Neuroscience 2003;23:3295–3301. [PubMed: 12716936]

7. Pokrajac D, Megalooikonomou V, Lazarevic A, Kontos D, Obradovic Z. Applying Spatial Distribution Analysis Techniques to Classification of 3D Medical Images. *Artificial Intelligence in Medicine* 2005;33:261–280. [PubMed: 15811790]
8. Kontos, D.; Megalooikonomou, V.; Prokrajac, D.; Lazarevic, A.; Obradovic, Z.; Ford, J., et al. Extraction of Discriminative Functional MRI Activation Patterns and an Application to Alzheimer's Disease. In: Barillot, C.; Haynor, DR.; Hellier, P., editors. *Lecture Notes in Computer Science (LNCS)*; 7th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI); Berlin Heidelberg: Springer-Verlag; 2004. p. 727-735.
9. Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A Unified Statistical Approach to Determining Significant Signals in Images of Cerebral Activation. *Human Brain Mapping* 1996;4:58–73.
10. Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ. Statistical Parametric Maps in Functional Imaging: A General Linear Approach. *Human Brain Mapping* 1995;2:189–210.
11. Friston, K. Statistical parametric mapping and other analyses of functional imaging data. In: Toga, A.; Mazziotta, J., editors. *Brain Mapping: The Methods*. Academic Press; San Diego: 1996.
12. Friston K. Statistical parametric mapping: ontology and current issues. *Journal of Cerebral Blood Flow and Metabolism* 1995;15:361–370. [PubMed: 7713993]
13. Nichols T, Hayasaka S. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research* 2003;12:419–46. [PubMed: 14599004]
14. Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* 1936;8:3–62.
15. Andersen, E. *Introduction to the Statistical Analysis of Categorical data*. Springer Verlag; Berlin: 1997.
16. Davatzikos C. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage* 2004;23:17–20. [PubMed: 15325347]
17. Ashburner J, Friston K. Voxel-based morphometry-The methods. *NeuroImage* 2000;11:805–821. [PubMed: 10860804]
18. Pettey DJ, Gee JC. Sexual dimorphism in the corpus callosum: a characterization of local size variations and a classification driven approach to morphometry. *NeuroImage* 2002;17:1504–1511. [PubMed: 12414289]
19. Dubb A, Xie Z, Gur R, Gur R, Gee J. Characterization of brain plasticity in schizophrenia using template deformation. *Academic Radiology* 2005;12:3–9. [PubMed: 15691720]
20. Bloom JS, Hynd GW. The role of the corpus callosum in interhemispheric transfer of information: excitation or inhibition? *Neuropsychology Review* 2005;15:59–71. [PubMed: 16211466]
21. Dubb, A.; Avants, B.; Gur, R.; Gee, JC. Shape Characterization of the Corpus Callosum in Schizophrenia Using Template Deformation. In: Dohi, T.; Kikinis Me, R., editors. *Lecture Notes in Computer Science (LNCS)*; Medical Image Computing and Computer-Assisted Intervention (MICCAI); Berlin Heidelberg: Springer-Verlag; 2002. p. 381-388.
22. Velakoulis D, Pantelis C, McGorry PD, Dudgeon P, Brewer W, Cook M, et al. Hippocampal volume in first-episode psychoses and chronic schizophrenia: a high-resolution magnetic resonance imaging study. *Archives of General Psychiatry* 1999;56:133–41. [PubMed: 10025437]
23. Bishop K, Wahlsten D. Sex differences in the human corpus callosum: Myth or reality? *Neuroscience and Behavioral Reviews* 1997;21:581–601.
24. Allen LN, Richey MF, Chai YM, Gorski RA. Sex differences in the corpus callosum of the living human being. *Journal of Neuroscience* 1991;11:933–942. [PubMed: 2010816]
25. Davatzikos C, Vaillant M, Resnick SM, Prince JL, Letovsky S, Bryan RN. A computerized approach for morphological analysis of the corpus callosum. *Journal of Computer Assisted Tomography* 1996;20:88–97. [PubMed: 8576488]
26. Dubb A, Gur R, Avants B, Gee J. Characterization of sexual dimorphism in the human corpus callosum. *NeuroImage* 2003;20:512–519. [PubMed: 14527611]
27. Megalooikonomou V, Kontos D, Pokrajac D, Lazarevic A, Obradovic Z. An adaptive partitioning approach for mining discriminant regions in 3D image data. *Journal of Intelligent Information Systems* 2008 ;31:217–243.

28. Megalooikonomou, V.; Pokrajac, D.; Lazarevic, A.; Obradovic, Z. Effective Classification of 3-D Image Data using Partitioning Methods. In: Erbacher, RF., et al., editors. Proc of SPIE 14th Annual Symposium in Electronic Imaging: Conference on Visualization and Data Analysis. SPIE; Bellingham, WA: 2002. p. 62-73.
29. Kontos, D.; Megalooikonomou, V.; Gee, J. Reducing the computational cost for statistical medical image analysis: An MRI study on the sexual morphological differentiation of the corpus callosum. In: Tsymbal, A.; Cunningham, P., editors. Proc of the 18th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2005). IEEE Computer Society Press; Los Alamitos, CA: 2005. p. 282-287.
30. Machado, AMC.; Gee, JC. Atlas Warping for Brain Morphometry. In: Hanson, KM., editor. Proc SPIE Medical Imaging 1998: Image Processing. SPIE; Bellingham, WA: 1998.
31. Machado AM, Simon TJ, Nguyen V, McDonald-McGinn DM, Zackai EH, Gee JC. Corpus callosum morphology and ventricular size in chromosome 22q11.2 deletion syndrome. Brain Research 2007;1131:197–210. [PubMed: 17169351]
32. Gee JC. On Matching Brain Volumes. Pattern Recognition 1999;32:99–111.
33. Petrie, A.; Sabin, C. Medical Statistics at a Glance. Blackwell Publishing; Oxford UK: 2000.
34. Samet H. The quadtree and related hierarchical data structure. ACM Computing Surveys 1984;16:187–260.
35. Devore, JL. Probability and Statistics for Engineering and the Sciences. Thomson Brooks/Cole; Belmont, CA: 2007.
36. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B 1995;57:289–300.
37. Genovese CR, Lazar NA, Nichols TE. Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. NeuroImage 2002;15:870–878. [PubMed: 11906227]
38. Poline, JB.; Holmes, AP.; Worsley, KJ.; Friston, KJ. Making statistical inferences. In: Frackowiak, RSJ., et al., editors. Human Brain Function. Academic Press, Elsevier; San Diego, CA: 1997.
39. Sheskin, D. Handbook of parametric and nonparametric statistical procedures. Chapman & Hall/CRC; Boca Raton, FL: 2004.
40. Alcantara, D.; Carmichael, O.; Delson, E.; Harcourt-Smith, W.; Sterner, K.; Frost, S., et al. Localized Components Analysis. In: Karssemeijer, N.; Lelieveldt, B., editors. Lecture Notes in Computer Science (LNCS); Proc. of Information Processing in Medical Imaging (IPMI 2007); Berlin Heidelberg: Springer-Verlag; 2007. p. 519-531.
41. Curran-Everett D. Multiple comparisons: philosophies and illustrations. Am J Physiol Regul Integr Comp Physiol 2000;279:R1–R8. [PubMed: 10896857]
42. Sjostrand, K.; Stegmann, MB.; Larsen, R. Sparse principal component analysis in medical shape modeling. In: Reinhardt, JM.; Pluim, JPW., editors. Proc of SPIE Medical Imaging: Image Processing. SPIE; Bellingham WA: 2006. p. 61444Xp. 1-12.
43. Stegmann, MB.; Sjostrand, K.; Larsen, R. Sparse modeling of landmark and texture variability using the orthomax criterion. In: Reinhardt, JM.; Pluim, JPW., editors. Proc of SPIE Medical Imaging: Image Processing. SPIE; Bellingham WA: 2006. p. 61441Gp. 1-12.
44. Uzumcu, M.; Frangi, A.; Sonka, M.; Reiber, J.; Lelieveldt, B. ICA vs. PCA active appearance models: Application to cardiac MR segmentation. In: Ellis, RE.; Peters, TM., editors. Lecture Notes in Computer Science (LNCS); Proc. of Medical Image Computing and Computer Assisted Intervention (MICCAI); Berlin Heidelberg: Springer-Verlag; 2003. p. 451-458.
45. Ballester, MAG.; Linguraru, MG.; Aguirre, MR.; Ayache, N. On the adequacy of principal factor analysis for the study of shape variability. In: Fitzpatrick, JM.; Reinhardt, JM., editors. Proc. of SPIE Medical Imaging: Image Processing; Bellingham WA: SPIE; 2005. p. 1392-1399.
46. Vermaak, J.; Perez, P. Constrained subspace modeling. Proc. of Computer Vision and Pattern Recognition (CVPR); Los Alamitos CA: IEEE Computer Society Press; 2003. p. 106-13.
47. Finkel RA, Bentley JL. Quad trees: a data structure for retrieval on composite keys. Acta Informatica 1974;4:1–9.

Figure 1.
An example of a normalized Jacobian corpus callosum image for a female subject.

Figure 2.
The main idea of DRP: applying statistical tests on groups of pixels to adaptively partition the image into progressively smaller sub-regions.

Pixel-Wise Normality Test for the male class

Lilliefors test at 0.05 significance: Normal pixels ($p > 0.05$) are shown in white

(a)

Pixel-Wise Normality Test for the female class

Lilliefors test at 0.05 significance: Normal pixels ($p > 0.05$) are shown in white

(b)

Figure 3.

Pixel-wise test for normality in the 'male' and 'female' classes using Lilliefors test: normal pixels with $p > 0.05$ are shown in white.

Figure 4. Discriminative regions of morphological differentiation detected by DRP with t-test and (a) $p < 0.05$, (b) $p < 0.01$, and ranksum test with (c) $p < 0.05$, (d) $p < 0.01$.

Figure 5. Discriminative regions of morphological differentiation detected by pixel-wise analysis with t-test and (a) $p < 0.05$, (b) $p < 0.01$, and ranksum test with (c) $p < 0.05$, (d) $p < 0.01$.

(a)

(b)

Figure 6. Discriminative region of morphological differentiation detected (a) by DRP using Bonferroni correction and FDR and (b) by pixel-wise analysis using FDR.

Table 1
A summary of all the preconditioned assumptions required for applying DRP.

Preconditioned assumptions
1. Two or more different groups of images generated by independent distributions.
2. Image background and non-relevant regions have been identified and segmented.
3. Application-dependent image pre-processing has been applied.
4. Equal a-priori pixel-wise probabilities for contributing to morphological differentiation.

Table 2

Comparison of the number of statistical tests performed by DRP and pixel-wise analysis.

Method	Experimental settings		Number of statistical tests
	Statistical test	P-Value threshold	
DRP	t-test	0.05	1684
		0.01	1945
	ranksum	0.05	1774
		0.01	1963
Pixel-wise analysis	t-test	0.05/0.01	3219
	ranksum	0.05/0.01	3219