NeuroImage **17,** 1504–1511 (2002)
doi:10.1006/nimg.2002.1130

Sexual Dimorphism in the Corpus Callosum: A Characterization of Local Size Variations and a Classification Driven Approach to Morphometry

David J. Pettey* and James C. Gee*,†

**GRASP Laboratory Department of Computer and Information Science and †Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104*

Received August 29, 2001

We present two complementary quantitative approaches to the problem of characterizing morphometric variations between two distinct populations. The case presented focuses solely on local size variations, but the general method can easily be applied to other scalar morphometric quantities. The first method uses a statistical parametric map (SPM) to ascertain a *P* value, which indicates whether any statistically significant differences exist between the populations. The second method focuses on finding the best single measurement which can be used for classifying the two populations. For our case study midsagittal cross sections of the corpora callosa from a population of normal males and females are nonrigidly registered (spatially normalized) to an atlas. The resulting deformations are then used to ascertain (i) whether there are any statistically significant differences between the populations and (ii) whether these differences allow one to perform classification. We make use of the Jacobian of the deformation field and normalize it to account for overall volume changes allowing us to focus on differences which are more related to morphometry than scale. From the (SPM) approach to the problem we find evidence of statistically significant differences in the morphology between the populations. Using a linear discriminant function we find that these differences do not appear to be useful for classification. Thus, this dataset provides an example of how statistically significant effects may not be of much diagnostic value. They may be of interest to the research community, but of little value to the clinician.

INTRODUCTION

The goal of this paper is the presentation of a classification driven approach to morphometry, the main result is the methodology since there are obviously simpler ways to identify the sex of a subject. Our choice to study sex-linked local size variations within the corpus callosum (CC) is mainly driven by the three points: (i) the admittance of a reasonable two-dimensional approach to the problem, (ii) a large existing literature on the problem, and (iii) the ready availability of a sufficiently large dataset. Though our main task is to explain what we feel is an important complementary methodology to already existent morphometric analyses, it is imperative that we present a thorough description of the background to the problem of sex-linked size variations in the corpus callosum. Following this we shall proceed to describe in detail how the data consisting of two-dimensional midsagittal slices of the corpus callosum is transformed to a set of scalar fields defined on a common atlas on which the final statistical analyses are performed.

Callosal morphology as a potential diagnostic criteria has attracted much attention in the research community. Recently, many researchers have focused on shape and size variations of the corpus callosum between distinct populations. There appears to be evidence confirming possible morphological differences between normal subjects and those subjects afflicted with schizophrenia (18, 6, 9), dyslexia (17) or Alzheimer's disease (16). If these effects are found to be distinctive enough we can hope to be able to use these differences in a diagnostic manner. Even barring such a diagnostic success, understanding the pathologies of a disease, in this case morphometric differences, may aid in uncovering the cause.

Many researchers have also turned to looking for sex related differences in the morphology of the CC. Thompson (23) and Bermudez and Zatorre (1) both provide a good overview of previous studies. Thompson correctly questions the use of the Witelson partition or similar partitions used in most studies of the CC searching for group differences. The Witelson partition consists of partitioning the structure using cuts perpendicular to the longest line segment that can be formed by connecting two points on the midsagittal slice of the CC. The first third is called the anterior

1053-8119/02 $35.00

third; the next sixth, the anterior midbody; the following sixth, the posterior midbody; the next two-fifteenths the isthmus; and the remaining fifth the splenium. Other partitions are discussed in (23). Studies typically report statistical differences in the overall or relative size of the whole CC or one of these parts of the CC. This of course may fail to reveal statistically relevant differences which either bridge boundaries between partitions or perhaps occur on scales smaller than the partition and are cancelled out by other competing effects within the same partition. Under the assumption that one can find a correspondence between any two callosa we perform an elastic registration to an atlas, this is sometimes referred to as spatially normalizing the data. Using the resulting set of transformations taking the atlas callosum to each subject's callosum we can search for differences in any subregion, potentially even distributed effects which are not confined to a single contiguous region.

Bermudez and Zatorre (1), using the Witelson partition, direct their attention to deficiencies in the methodology used in many studies of sexual dimorphism of the CC. First, there are problems associated with small sample sizes or, somewhat equivalently, poorly matched samples with too many confounding factors, especially age and handedness. Second is the problem of what to measure. Should one look for differences in the raw size of the CC, or should one somehow normalize the CC area (when we speak of CC area we are referring to the area of the midsagittal slice) by some measure of brain size, since brain size is already known to be different between males and females? And if we choose to do so, then how should it be done? Bermudez and Zatorre present a clear discussion of several different normalization procedures, and difficulties associated with them. For example, they find that CC area does not appear to scale in the manner one might expect with measures of brain volume. Additionally it appears to have a different scaling behavior in males and females. As a result the correlation of callosal area with measures of brain volume will be different across populations.

For this reason, we have chosen to look at a slightly different quantity. Rather than normalizing CC area with a measure of brain volume or total midsagittal area we have chosen to normalize with respect to total CC area. This gives a very different measure of differences between populations. What we will be examining is the difference in the size of regions of the CC relative to the overall size of the CC. Davatzikos (5) and Machado (19, 20) have previously examined smaller data sets in the same manner and have found apparent differences, though without quantifying the statistical significance of their observations. Bermudez and Zatorre will undoubtedly be nonplussed by the application of yet another normalization measure, but we can only counter that there do appear to be statistical differences between males and females using this measure, and the methodology is quite clear.

We find that in our data set there are no statistically significant localized absolute size differences between the male and female CC. However, we do find significant differences in the relative (normalized by total CC area) sizes of localized regions, most notably an effect in the splenium ($P < 0.002$) and a mild effect in the anterior midbody ($P < 0.3$), which we report only for completeness.

Finally, we wish to reiterate the difference between statistically significant differences and differences which can be used for classification. As mentioned at the opening we sometimes hope that we will be able to make use of statistically significant differences in populations in a diagnostic setting. However, this may not always be possible. In our complementary analysis we will be searching for the best possible scalar diagnostic function (or linear discriminant function) in the hopes of finding some collective measure of a sample which could be used as a classifier. We find in this case that the best such classifier fails to perform well, and as such we should not expect the statistically significant differences observed between our populations to be useful in performing classification. Thus, despite the clear presence of a statistically significant difference in the relative size of a localized region, this difference is not substantial enough to be used for classification purposes. It may nevertheless be of significant value to understanding differences between two populations.

MATERIALS AND METHODS

Subjects

All subjects were right-handed volunteers free from any known neurological disorders. The population consisted of 29 males and 45 females. The male ages fell in the range 19–39 years ($\mu = 25.79$, $\sigma = 5.59$). For the females the age range was 18–37 years ($\mu = 23.56$, $\sigma = 4.19$).

Image Acquisition

Magnetic resonance imaging scans were acquired on a 1.5-T scanner (Signa; General Electric Co, Milwaukee, WI) with a spoiled gradient-recalled pulse sequence using the following parameters: flip angle of 35°, repetition time of 35 ms, echo time of 6 ms, field of view of 24 cm, I repetition, 1-mm slice thickness, and no interslice gaps. Transaxial images were in planes parallel to the orbitomeatal line, with resolution of 0.9375×0.9375 mm^2. Images were resliced along the anterior to posterior commissural (AC–PC) axis to standardize for head tilt. Sagittal images are rotated so that the AC–PC axes are oriented to straight horizontal positions. No parenchymal lesions or skull abnor-

malities were evident neuroradiologically. The midsagittal slice is selected as the sagittal image with only traces of cortex visible. Additional cues include a clear separation of the tectum and tegmentum by the cerebral aqueduct, and good visibility of the fourth ventricle and vermis of the cerebellum.

Generalized k-means clustering was performed to segment the white matter (19). The corpus callosum was then isolated by manual segmentation. The k-means algorithm groups image voxels into k clusters in such a way that the intensity variance within each cluster is minimized. The generalized version additionally imposes a smoothness constraint to favor groupings that are spatially contiguous.

Registration

One additional male subject's segmented callosum was used as an atlas to perform registration. As remarked in (5), the statistics of the variables we are examining should be independent of the shape of the atlas chosen if the registration procedure is physiologically meaningful (*i.e.*, registering anatomically corresponding regions). A few preliminary investigations on our own dataset suggests this is true in practice.

We take N_A to be the number of pixels occupied by the corpus callosum chosen as the atlas. We further define (i_k^A, j_k^A) to be the coordinates of the k^{th} pixel in our atlas. Now we need to find the correspondences between the atlas and every other CC. There are many proposed methods, differing mainly in their choice of how to regularize the transformation. Christensen (3) and Bro-Nielsen (2) use a fluid model letting one image flow into the other. Thirion (22) employs a different model using optical flow techniques and Dawant *et al.* (7) have demonstrated the repeatability and agreement with human identified correspondences using this method. We have chosen to employ an elastic-membrane model (described in detail in (15, 14)) though make no claims as to the superiority of this model over other methods.

Whichever method is chosen we obtain for each subject a vector field or deformation field $\mathbf{u}_{\cdot\cdot}(s)$, where

$$\mathbf{u}_{i_k^A, j_k^A}(s) \equiv (u^1_{i_k^A, j_k^A}(s), u^2_{i_k^A, j_k^A}(s)) \tag{1}$$

tells us the displacement needed to place the k^{th} atlas pixel into the correct correspondence with the corpus callosum of subject s. So now instead of a collection of images as our data set we have a collection of deformation fields over a common atlas. For example, if all of the samples were identical to the atlas then there would be no deformations needed to register the atlas to the subject: the vector describing the deformation that atlas pixel k undergoes to put it into the correct position for subject s would be zero for each subject and for each atlas pixel, hence $\mathbf{u}_{i_k^A, j_k^A}(s) = 0$. If each callosa were the same up to a translation then we would have $\mathbf{u}_{i_k^A, j_k^A}(s) = C(s)$, that is the deformation would be a constant for each subject (it would not vary with k), but the constant would depend upon which subject we were registering. Clearly in a morphometric analysis we are not interested in overall translational differences, or overall rotational differences. Thus we will find it more economical to look at functions of $\mathbf{u}$ which are invariant with respect to rigid transformations.

The Determinant of the Jacobian

In general, performing statistical analysis of functions of the deformation fields has been termed deformation-based morphometry (DBM) (4). We have chosen to focus our attention on local size differences between corresponding portions of the corpus callosum. This is by no means the only measure that one could consider. But, as it has been examined previously and the presence or absence of size differences is in dispute (8, 5, 20, 19, 23), we have chosen this to be the quantity of interest in our study. As mentioned at the outset, looking at this scalar field affords us a better means to explore for size changes than would an *a priori* partitioning of the CC.

Recall $\mathbf{u}_{i,j}(s)$ is a displacement field for subject s. In order to calculate the Jacobian of the transformation which takes the atlas into the corpus callosum of subject s it helps to consider $\mathbf{u}_{i,j}(s)$, in a slight abuse of notation, to be a vector field $\mathbf{u}_s(\mathbf{x})$. That is, we consider the atlas to be a region in the plane rather than merely a collection of discrete points. Then we need to recall that the transformation, T_s, which takes the atlas corpus callosum into subject s's is given by,

$$T_s(\mathbf{x}) = \mathbf{x} + \mathbf{u}_s(\mathbf{x}). \tag{2}$$

Finally the quantity we wish to examine is the determinant of the Jacobian of this transformation,

$$\left| \frac{\partial T_s(\mathbf{x})}{\partial \mathbf{x}} \right|. \tag{3}$$

Subsequently, we prefer to go back to our discretized space and denote by $J_k(s)$ the value of the determinant of the Jacobian of the transformation at pixel (i_k^A, j_k^A) in the atlas.

We have now reduced our original image data into a collection of k numbers related to the expansion or contraction required by the atlas at pixel (i_k^A, j_k^A) in order to achieve correspondence. The most important aspect of the new random variables J_k is that they relate to what we consider to be an important physical characteristic (local size change) as well as having the

feature that $J_k(s_1)$ and $J_k(s_2)$ refer to the corresponding physically meaningful measures.

Mean Corrected J_k

In addition to examining the statistics of J_k we will be looking at relative size variations by rescaling J_k so that the overall callosal area is constant. We will denote the rescaled variable by,

$$j_k(s) = \frac{J_k(s)}{\sum_k J_k(s)}. \tag{4}$$

Looking at j_k rather than J_k is our way of trying to normalize for gross (global) variations. Bermudez and Zatorre (1) provide an excellent overview of the various methods other authors have employed, including normalizations involving frontal brain volume or total brain mass through a rescaling or through their use as a covariate in a statistical model. The mean corrected Jacobian appeared to us an obvious choice. This clearly gives us some measure of shape, since overall size variations have already been taken into account. However, we emphasize that this is by no means the only scalar field of **u** that one could use to measure shape.

The Logarithm of J_k

In the following sections we will be examining statistical properties of J_k in a pointwise and a collective fashion. Most of the standard statistical tests employed rely on the assumption of a Gaussian distribution for the underlying random variables. Naively, one would expect that a change in the size by a factor of two in expansion or contraction are apt to be more equally likely than two constant shifts in J_k. After all with a base assumption that $J_k = 1$ a positive shift by 1 implies a doubling in size, whereas a negative shift by 1 would imply an infinite contraction! Accordingly, the distribution of J_k is more likely log-normal than normal. Thus, we introduce,

$$J_k^{log}(s) = \log(J_k(s)) \tag{5}$$

as a measure to examine size differences. In all that follows, the statistical tests may be applied to either J_k or $J_k^{\log}$ and in some cases we will present both. Historically, most other authors have used J_k and we feel obliged to note that the differences are not extreme, yet we do believe that $J_k^{\log}$ is the more appropriate measure. Additionally, we will denote by $j_k^{\log}$ the logarithm of the rescaled J_k.

Statistical Analysis

So far we have condensed the morphological differences between our subjects from a vector field $\mathbf{u}_{i,j}(s)$ characterizing the full transformation from the atlas to the subject, down to a scalar field $j_k(s)$ characterizing relative local size variations at the k^{th} atlas pixel. Now we wish to employ statistical tests on j_k (J_k) to search for statistically significant differences in the relative (absolute) size variations between the two populations.

Pointwise Statistics

At this point Gee, Machado, and Davatzikos (19, 5) decided to examine each j_k individually and compute an effect size for the difference between the two populations for each pixel. Take μ_k^f and μ_k^m to be the average of j_k over the females and males respectively and σ_k^f and σ_k^m to be the respective standard deviations. Then, defining,

$$\sigma_k = \sqrt{\frac{(n_f - 1)(\sigma_k^f)^2 + (n_m - 1)(\sigma_k^m)^2}{n_m + n_f - 2}}, \tag{6}$$

to be the pooled variance (where n_m and n_f are the number of male and female subjects, respectively) we compute the effect size for the k^{th} pixel,

$$e_k^j = \frac{\mu_k^f - \mu_k^m}{\sigma_k}. \tag{7}$$

This is a measure of how different j_k is between the two populations. We can now determine which regions of the atlas are deformed in characteristically different ways for males and females by examining which pixels have a large effect size associated with them. Typically, one looks for an effect size greater than 1 as an indication that the distributions are significantly different. We can set a specified threshold and then shade all of the pixels whose effect size is greater than that threshold to see which regions of the atlas are relevant for discriminating between males and females.

The effect size is basically telling us how well we might be able to perform classification based upon the value of j_k. However, it does not tell us how certain we are that the difference observed is real and not just an artifact of noise. To do this, we perform a simple means test between the two populations under the null hypothesis that the means are equal and with the assumption that the variances of j_k within the two populations are equal. For each pixel we calculate a t score,

$$t_k^j = \frac{\mu_k^f - \mu_k^m}{\sigma_k \sqrt{\frac{1}{n_m} + \frac{1}{n_f}}}. \tag{8}$$

Given the large number of degrees of freedom (86 − 2 = 84) provided by our large sample size, the t scores are very close to being actual z scores and can be interpreted as such with little error. In any case, these t scores allow us to calculate a P value which we use to determine whether to accept or reject the null hypothesis. The collection of t scores across all pixels forms what is commonly referred to as a statistical parametric map (SPM) (11, 10, 13, 12). Indeed, performing an SPM99 (21) analysis of the data using a PET model, two-sample t test will yield the same raw t score map described above.

For fixed sample sizes the information content of the effect size and the t score is identical—the values differ only by a scale factor. The t score tells us how sure we are that the means differ between the two populations and the effect size tells us how well this difference can be used for classification.

Cluster Statistics

Above we simply looked at each pixel independently and have found that there are no significant differences between the populations. However, this does not exhaust the possible means for finding a difference. The inherent spatial nature of our data suggests that we should also look at the statistics of clusters of pixels, clusters of neighboring pixels and possibly collections of distant clusters. If we observe above threshold effect sizes for some pixels we expect them to be in clusters. We expect that if one pixel reveals a large effect size then neighboring pixels will be more likely to have a large effect size. Furthermore, if j_k is larger in the males than in the females we anticipate that $j_{k'}$ for neighboring pixels will also be larger, not smaller. Succinctly, we anticipate there are correlations between the different j_k's. Additionally, there may be correlations between pixels widely separated in space. Looking at the pointwise statistics ignores all these correlations.

Because of the nature of the registration algorithm we expect t_k^j to approximate a unit Gaussian field with some finite spatial correlation σ (not related to the σ_k's above) which can be estimated using SPM99. The complications associated with calculating corrected P values are discussed in (11, 10, 13, 12). One can look at the likelihood of having any one connected region exceed a given threshold, or the likelihood of having a connected region above a given size for a given threshold. We choose to look at the latter in order to identify significant regions of size change. That is, we will look at a thresholded map of the t scores and locate connected regions where the t scores are above the chosen threshold. For each region we can then calculate a P value for a region of this size appearing at random in a Gaussian field with the same σ. Because of the complications associated with irregular boundaries and such, rather than calculate the P values analytically we have performed simulations to obtain the relevant P values for clusters above a certain threshold and size. Furthermore, because the t score map may not be adequately described by a Gaussian field with a uniform correlative structure, we have also employed a permutation test on the data to calculate corrected P values.

Collective Effects and the Linear Discriminant Function

Looking at SPM-style cluster statistics allows us to identify contiguous regions which might be of significance. Another approach which tries to capitalize on the information hidden in the correlations between pixels is to look at linear combinations of the j_k's. This not only allows the correlations between neighboring pixels to be used, but also any correlations between disparate regions. Here, we are not inherently assuming that neighboring pixels are any more or less correlated than any other pixels. The plan now is to look at all linear combinations of the j_k's (up to some overall scale factor) and see which one provides the best discrimination between the two populations.

In practice, because the number of samples is small compared to the number of random variables (there are 3325 pixels in our atlas) our results would be dominated by noise artifacts if we were to look at all linear combinations of the j_k's. So instead, we first thin down the number of random variables of relevance. We chose to keep only the first few principal components of the j_k's, where

$$p_i(s) = \sum_k \hat{\mathrm{e}}_k^i j_k(s) = \hat{\mathrm{e}}_k^i j_k(s) \tag{9}$$

is the i^{th} principal component and $\hat{\mathrm{e}}_k^i$ are the components of $\hat{\mathrm{e}}^i$, a unit vector in the direction of the i^{th} principal axis (in the last line we have employed the Einstein summation convention of summing over repeated indices). Recall that $p_1(s)$ is the linear combination of the j_k's with the largest variance, $p_2(s)$ has the same property for all linear combinations with a principal axis orthogonal to $\hat{\mathrm{e}}^1$ and so on.

At this point Machado and Gee (20) perform a type of factor analysis employing the principal components of the J_k's scaled so as to have unit variance. They then find a collection of factors or simply new random variables which are linear combinations of the J_k's. Some of these factors appear to be localized to particular regions of the corpus callosum though none seem to give rise to random variables which can be used to classify the populations. A word of caution is in order; even though the resulting random variables or factors may have a large effect size they still may be of little use for classification because of the small size of the data sets used. One really must perform a cross-validation or jack-knife analysis to be confident that the results are not simply due to noise.

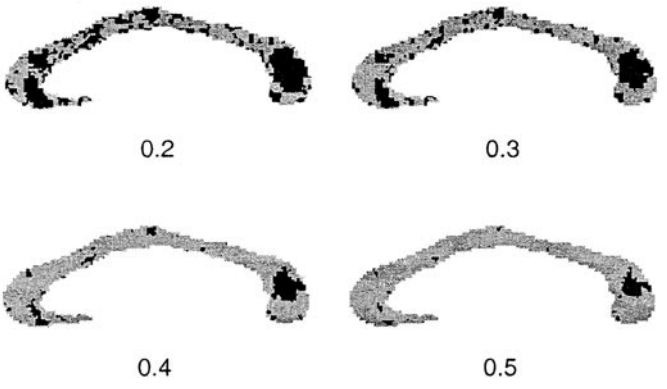

FIG. 1. Thresholded images of the effect size of j_k performed pointwise where the black regions indicate pixels where the effect size is above the threshold. The threshold values used were 0.2, 0.3, 0.4, and 0.5, as indicated in the images above. At the 0.5 level, we see very few pixels that have a corresponding effect size above this threshold.

Factor analysis has its own criteria for the choice of linear combinations of random variables which are of interest, and these criteria do not presuppose that classification is the goal. Our goal is to find the linear combination which best discriminates between the populations and to this end we apply linear discriminant analysis to the p_i's to find the linear discriminant function $f(s)$. Since f is a linear combination of the p_i's we can write

$$f(s) = d_i p_i(s). \tag{10}$$

This is the linear function of the p_i's which best discriminates between the two populations. Here **d** is again a unit vector. Since each p_i is a linear combination of the j_k's we can also write $f(s)$ as a linear combination of the j_k's:

$$f(s) = \hat{e}_k^{(LDF)} j_k(s). \tag{11}$$

Now by examining which pixels influence $f(s)$ the most (which values of k have large "weights", $\hat{e}_k^{(LDF)}$) we can see which regions of the CC are most associated with differences between the two populations and thus which regions will be useful for classification.

The whole procedure can be repeated with varying numbers of principal components and using a jackknife procedure, leaving one subject out and then seeing how the classifier classifies that subject and then repeating for each subject. We did precisely this keeping from 3–30 principal components.

As presented, our linear discriminant analysis is complementary to the SPM-like analysis. In the latter test looking at the size of connected regions above the threshold, we can identify those regions which we are most confident are statistically significant. However, examining the linear discriminant function would appear to be a better way to identify regions which are most important for classification purposes and determining differences between the populations. In general, these regions need not be the same as the regions identified through the SPM analysis. It is possible to be certain that a difference between two populations exists even if this difference is so small as to be not particularly useful for performing classification (e.g. two Gaussian populations with large variances and slightly different means).

RESULTS

Effect Size

With our samples, we find that there are no points where the effect size of j_k is greater than 1 and even for a relatively small threshold there do not appear to be very many relevant pixels (see Fig. 1). This is in contrast to earlier results on smaller sample sizes which found that the region of the splenium appeared to be significantly different between the male and female populations (5, 20, 19). We find similar results for J_k and the logarithms.

SPM and t Scores

Figure 2a contains thresholded images of the t scores for J_k using a threshold for a pixel-wise (or raw) P value of 0.1 (meaning that ignoring correlations we would expect 10% of the pixels to be above threshold). There is one rather large region in the image in the anterior midbody. A standard SPM99 analysis yielded the FWHM (full-width half-max) of the correlative structure in the Gaussian field corresponding to a $\sigma = 2.4$. Using this σ and the atlas corpus callosum for a mask we simulated 25000 Gaussian fields to find estimates of corrected P values for connected regions above the threshold. The cluster yields $P < 0.3$.

For j_k we find a more significant cluster. Figure 2b contains the corresponding thresholded images for j_k. Here we again see a cluster in the anterior midbody, but now we also see a much larger cluster in the posterior region of the splenium. A cluster-level analysis again reveals that the anterior cluster has $P < 0.3$, but that the cluster near the splenium has a much higher significance with $P < 0.002$. Again, we also find similar results for the logarithms of both J_k and j_k. Additionally, we performed a permutation test to calculate a P

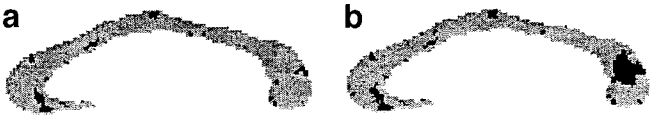

FIG. 2. Thresholded images of the t scores of (a) J_k and (b) j_k performed pointwise, with a raw P value threshold value of 0.1.

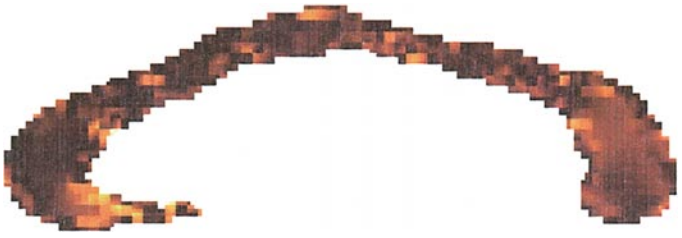

FIG. 3. An image of the absolute value of the weights for the best linear discriminant function for j_k using the first 15 principal components. Generally the splenium appears to be a region of large weights, but the highest weights appear to be in a smaller region in the anterior.

value for finding a cluster at least as large as the one near the splenium. Randomly partitioning the subjects into groups of 29 ("males") and 45 ("females") we perform the same analysis to find a t score map and then count the number of clusters above the threshold. Performing this procedure 20000 times yields a $P < 0.04$, which is still significant but much less so than the 0.002 value obtained from the Gaussian fields method. This should not be too surprising since in the first method we assume that the correlative structure is uniform throughout the callosum, whereas, this is most likely not the case.

Linear Discriminant Function

As mentioned earlier we performed a jack-knife analysis of the classifier using 3–30 principal components in constructing f for j_k. We found that the reliability of the classifier did not vary much, or systematically as we varied the number of principal components used and that we could correctly classify the left out subject only about 60% of the time. In Fig. 3 we present the weights associated with f when using 15 principal components. Note that the posterior portions of the corpus callosum do appear to be associated with large weights in f, though given the poor performance of the classifier we cannot deduce too much from this. We are led to believe that although the pointwise statistics indicate differences exist, they are not relevant enough to allow for classification. Performing the same analysis for J_k or the logarithms yields similar results.

DISCUSSION

We have presented an approach to a classification driven method for morphometry. Additionally, for our present problem of callosal morphologic variations between males and females we have presented a thorough investigation of pointwise statistics of the mean adjusted Jacobian, as well as cluster statistics of this same variable. Our approach to cluster statistics employed a classical SPM style approach as well as a less assuming permutation analysis. The final analysis focused on a complementary linear discriminant analysis of the data.

We have presented evidence that there are statistically significant differences in the relative sizes of regions in the male and female corpus callosum. We have found that while the splenium appears to be significantly different in relative size between the populations, it is not so different in absolute terms.

We have also presented a new complementary statistical method for distinguishing between and characterizing populations using principal components and a linear discriminant function. The linear discriminant function attempts to build the best scalar valued random variable for performing classification. We find that in the current data set the differences between the populations (while statistically significant) are not large enough to allow for a good classification. In connection with the rather unambiguous results from the SPM analysis ($P < 0.04$), which provides us with an assurance that there are statistically significant differences in morphology between the populations, this result tells us that these differences are nevertheless insufficient to yield a good linear classifier. We caution, that there may still be measures other than the Jacobian which could provide us with a good classifier. The main purpose of our current work was to present a coherent methodology which can be used with any scalar measure and we simply chose to use one for which a reasonable amount of prior data were available.

ACKNOWLEDGMENTS

We thank Colin Studholme for pointing out the importance of the log-normal distribution. The authors are also grateful to Veda Maany for assistance in preparing the midsagittal data; to the Schizophrenia Research Center, Neuropsychiatry section, of the University of Pennsylvania for sharing the corpus callosum data; and to the U.S.P.H.S. for supporting this work under grants NS33662, LM03504, MH62100, AG15116, and AG17586.

REFERENCES

1. Bermudez, P., and Zatorre, R. J. 2001. Sexual dimorphism in the corpus callosum: Methodological considerations in MRI morphometry. *NeuroImage* **13:** 1121–1130.
2. Bro-Nielsen, M., and Gramkow, C. 1996. Fast fluid registration of medical images. In *Proc. Visualization Biomedical Computing,* Vol. 1131, pp. 267–276. Springer-Verlag, Berlin, Germany.
3. Christensen, G. E., Miller, M. I., and Vannier, M. 1994. 3D brain mapping using a deformable neuroanatomy. *Phys. Med. Biol.* **39**(3): 609–618.
4. Chung, M. K., Worsley, K. J., Paus, T., Cherif, C., Collins, D. L., Giedd, J. N., Rapoport, J. L., and Evans, A. C. 2001. A unified statistical approach to deformation-based morphometry. *NeuroImage* **14:** 595–606.
5. Davatzikos, C., Vaillant, M., Resnick, S. M., Prince, J. L., Letovsky, S., and Bryan, R. N. 1996. A computerized approach for morphological analysis of the corpus callosum. *J. Comput. Assist. Tomogr.* **20**(1): 88–97.

6. David, A. S. 1992. Schizophrenia and the corpus callosum: Developmental, structural, and functional relationships. *Behav. Brain Res.* **64:** 203–211.
7. Dawant, B. M., Hartmann, S. L., Thirion, J. P., Maes, F., Vandermeulen, D., and Demaerel, P. 1999. Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations: Part I, Methodology and validation on normal subjects. *IEEE Trans. Med. Imag.* **18**(10): 909–916.
8. de Lacoste-Utamsing, C., and Holloway, R. L. 1982. Sexual dimorphism in the human corpus callosum. *Science* **216:** 1431–1432.
9. Delisi, L. E., Tew, W., Xie, S., and Hoff, A. L. 1995. A prospective follow-up study of brain morphology and cognition in first-episode schizophrenic patients: Preliminary findings. *Biol. Psychiatry* **36:** 349–360.
10. Friston, J. J., Evans, A. C., Marrett, S., and Neelin, P. 1992. A three-dimensional statistical analysis for CBF activation studies in human brain. *J. Cereb. Blood Flow Metab.* **12:** 900–918.
11. Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. J. 1991. Comparing functional (PET) images: The assessment of significant change. *J. Cereb. Blood Flow Metab.* **11:** 690–699.
12. Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., and Frackowiak, R. S. J. 1995. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* **2:** 189–210.
13. Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., and Evans, A. C. 1994. Assessing the significance of focal activations using their spatial extent. *Hum. Brain Mapp.* **1:** 214–220.
14. Gee, J. C. 1999. On matching brain volumes. *Pattern Recogn.* **32:** 99–111.
15. Gee, J. C., and Bajcsy, R. K. 1999. Elastic matching: Continuum mechanical and probabilistic analysis. In *Brain Warping* (A. W. Toga, Ed.), pp. 183–197. Academic Press, San Diego.
16. Hampel, H., Teipel, S. J., Alexander, G. E., Horowitz, B., Teichberg, D., Shapiro, M. B., and Rapoport, S. I. 1998. Corpus callosum atrophy is a possible indicator of region and cell type specific neuronal degeneration in Alzheimer disease. *Arch. Neurol.* **55:** 193–198.
17. Hynd, G. W., Hall, J., Novey, E. S., and Eliopulos, D. 1995. Dyslexia and corpus callosum morphology. *Arch. Neurol.* **52:** 32–38.
18. Lewine, R., Flashman, L., Gulley, L., and Beardsley, S. 1991. Sexual dimorphism in the corpus callosum and schizophrenia. *Schizophren. Res.* **4:** 63–64.
19. Machado, A. M. C., and Gee, J. C. 1998. Atlas warping for brain morphometry. In *Medical Imaging 1998: Image Processing*, pp. 642–651. SPIE, Bellingham, WA.
20. Machado, A. M. C., Gee, J. C., and Campos, M. F. M. 2000. A factor analytic approach to structural characterization. In *Mathematical Methods in Biomedical Image Analysis*, pp. 219–226. IEEE Comp. Soc., Los Alamitos, CA.
21. www.fil.ion.ucl.ac.uk/spm/spm99.html.
22. Thirion, J. P. 1998. Non-rigid matching using demons. *Med. Image Anal.* **2**(3): 243–260.
23. Thompson, P. M., Narr, K. L., Blanton, R. E., and Toga, A. W. Mapping structural alterations of the corpus callosum during brain development and degeneration. In *Proc. of the NATO ASI on the Corpus Callosum*. Kluwer, in press.