Published in final edited form as:

Inf Process Med Imaging. 2009 ; 21: 300–313.

A Unified Framework for MR Based Disease Classification

Kilian M. Pohl^{1,2} and Mert R. Sabuncu²

¹ Healthcare Informatics, IBM Almaden Research Center, San Jose, CA, USA pohl@us.ibm.com

² Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. msabuncu@csail.mit.edu

Abstract

In this paper, we employ an anatomical parameterization of spatial warps to reveal structural differences between medical images of healthy control subjects and disease patients. The warps are represented as structure-specific 9-parameter affine transformations, which constitute a global, non-rigid mapping between the atlas and image coordinates. Our method estimates the structure-specific transformation parameters directly from medical scans by minimizing a Kullback-Leibler divergence measure. The resulting parameters are then input to a linear Support Vector Machine classifier, which assigns individual scans to a specific clinical group. The classifier also enables us to interpret the anatomical differences between groups, as we can visualize the discriminative warp that best differentiates the two groups. We test the accuracy of our approach on a data set consisting of Magnetic Resonance scans from 16 first episode schizophrenics and 17 age-matched healthy control subjects. The data set also contains manual labels for four regions of interest in both hemispheres: superior temporal gyrus, amygdala, hippocampus, and para-hippocampal gyrus. On this small size data set, our approach, which performs classification based on the MR images directly, yields a leave-one-out cross-validation accuracy of up to 90%. This compares favorably with the accuracy achieved by state-of-the-art techniques in schizophrenia MRI research.

1 Introduction

Thanks to in-vivo imaging technologies like magnetic resonance images (MRI), the last few decades have witnessed a rapid growth in the amount of research that explores structural and functional abnormalities due to certain diseases (see e.g. [1]). Studies that investigate such inter-group differences rely heavily on the notion of cross-subject correspondence. They ensure cross-subject correspondence, for example, by extracting structure-specific features from labor-intensive manual segmentations [2,3]. An alternative strategy is to investigate thousands of local features, such as in [4–6], but determining an unbiased dense spatial correspondence across a heterogeneous population is challenging. In this paper, we propose an alternative methodology that employs local features, yet takes advantage of well-established structure-based correspondence across individuals.

In schizophrenia research, structural imaging studies have traditionally focused on showing differences between the sizes of certain brain regions [7]. Other studies have shown that various boundary and/or skeleton shape representations can also be employed for discriminating corresponding structures in two different groups [2,3]. These structure-based morphology methods typically require a priori knowledge about abnormal regions, which generally is extracted from manual segmentations [8]. Generating manual segmentations is labor intensive and therefore limits the size of datasets used by structure-based morphological methods. This shortcoming can be addressed by automatic segmentations but any bias or error in the segmentations will negatively impact the analysis.

As an alternative strategy, recent research has proposed automatic comparative analysis techniques that typically investigate thousands of features from each image. There are two flavors to this approach: voxel-based and deformation-based morphology. The first approach attempts to discover statistical differences in, for example, tissue density at various voxel locations between the groups [9]. This assumes a global spatial normalization of the images. Once voxel-wise correspondence is established, various local measurements are compared across subjects. The second approach is to employ a highly nonlinear registration algorithm to align the subjects with a template. A discriminative analysis then can be conducted based on the deformation fields [10] or both the deformation fields and residual images [11]. Rather than performing voxel-by-voxel univariate comparisons, other studies [4–6] have employed a large number of features to improve the statistical separation between two groups. These approaches provide the potential to design accurate classification algorithms that, in the future, can be used for diagnostic purposes. In these techniques, inter-subject spatial correspondence is often achieved using a non-linear intensity-based image registration algorithm, where the quality of alignment is quantified by the similarity between intensity values or local attribute vectors [12], and the warps that capture inter-subject anatomical variability are represented via arbitrary mechanical or probabilistic models. The high-dimensional nature of the problem and modeling assumptions made in the formulations can potentially bias inferences based on these registration results.

The community has suggested various dimensionality reduction methods (e.g. Principal Component Analysis [13]) and feature selection techniques (e.g. Recursive Feature Elimination [8]) to tackle the problem of learning in very high dimensions with a small training set. The outcome of these methods tend to be a complex mixture and/or a subset of the original feature space. Alternatively, researchers have proposed image features encoding local statistics or texture[6]. However, the complex nature of these features can obscure classification and make clinical interpretation difficult.

In this paper, we propose a population analysis methodology that takes advantage of the well-established structure-based correspondence across individuals. This approach is different from a majority of today's morphology methods that employ local descriptors and assume correspondence between these. In the proposed framework, we encode the inter-subject spatial mapping via structure-specific affine registration parameters. These parameters are global to each structure so that we do not introduce any further ambiguity to the correspondence problem.

Similar to [14,15], we compute the transformation parameters directly from the images by employing a Bayesian framework that models the relationship between the segmentation labels, transformation parameters and image intensities. Each structure-specific transformation is encoded by a global, 9-parameter affine model. Our method then uses a linear Support Vector Machine (SVM) [16] classifier to jointly examine subtle differences in the affine registration parameters across all structures while characterizing natural variability within the population. This analysis reveals the anatomical separation between two groups. The separation can be analyzed by generating warps based on the weights of the linear classifier [2]. We use these warps to visualize the most significant anatomical differences between the groups under study.

We test the accuracy of the proposed method on a data set consisting of 16 first episode schizophrenics and 17 age-matched healthy subjects [17]. The original morphometry study that used this data [17] studied volumetric measurements of eight temporal lobe structures, obtained from manual segmentations of the MR images. The goal of that study was to show statistical differences in independent volumetric measurements between first-order schizophrenics and controls. In the present paper, we show that a linear SVM trained on all eight volumetric measurements achieves a leave-oneout cross-validation accuracy of 69%. In comparison, the linear SVM based on our structure specific 9-parameter affine parameters

achieves a cross-validation accuracy of up to 90%. This accuracy also compares favorably with state-of-the-art methods for schizophrenia classification based on MRI data [4].

The contributions of this paper are multi-fold: we present a new population analysis approach that takes advantage of unambiguous, well-established correspondences between anatomical structures to reduce the dimensionality of the problem. In the proposed framework, the discriminative features can be interpreted in anatomical terms. Furthermore, the integrated automatic segmentation procedure will allow us to deploy our analysis on large collections of image data, for which we may not have manual segmentations.

2 Classifying Images Based on Registration Parameters

Our population analysis methodology has two stages: the extraction of structure-specific image features and the discriminative statistical analysis of these feature samples. In the remainder of this section, we first derive a formulation for extracting structure-specific registration parameters from images. We then use a linear classifier to conduct a statistical analysis on these feature samples.

2.1 An Iterative Approach for Computing Warps

Let $\mathcal{T} \triangleq (\mathcal{T}_1, \dots, \mathcal{T}_N)$ be the parameter set of structure-specific transformations, where N is the number of structures of interest (including a background label). In the following, we will assume the existence of an atlas and corresponding coordinate system, which we define in more detail in Section 2.2. For each structure $i \in \{1, \dots, N\}$, $\mathcal{T}_i$ defines a mapping the image coordinate frame to the atlas coordinate frame. We are interested in estimating $\mathcal{T}$ directly from the image I using a Maximum A Posteriori (MAP) formulation:

$$\widehat{\mathcal{T}} = \arg \max_{\mathcal{T}} \log P(\mathcal{T}|I) = \arg \max_{\mathcal{T}} \log P(I, \mathcal{T}). \quad (1)$$

At this point, we introduce another random variable $\mathcal{L}$, which encodes the anatomical label at each voxel. As commonly done in the literature, we will factorize the joint probability on $P(I, \mathcal{T}, \mathcal{L})$ using three distributions:

- a prior distribution on the transformation parameters: $P(\mathcal{T})$,
- a conditional label prior distribution: $P(\mathcal{L}|\mathcal{T})$, and
- the image likelihood, which we assume to be independent of $\mathcal{T}$: $P(I|\mathcal{L}, \mathcal{T}) = P(I|\mathcal{L})$.

Next, we can rewrite Equation (1) by marginalizing $P(I, \mathcal{T}, \mathcal{L})$ over all possible label maps $\mathcal{L}$:

$$\begin{aligned} \widehat{\mathcal{T}} &= \arg \max_{\mathcal{T}} \log \sum_{\mathcal{L}} P(I, \mathcal{L}, \mathcal{T}) = \arg \max_{\mathcal{T}} \log \sum_{\mathcal{L}} P(I, \mathcal{L}|\mathcal{T}) + \log P(\mathcal{T}) \\ &= \arg \max_{\mathcal{T}} \log \sum_{\mathcal{L}} P(I|\mathcal{L}) \cdot P(\mathcal{L}|\mathcal{T}) + \log P(\mathcal{T}) \end{aligned} \quad (2)$$

where Σ_A represents the sum over all possible values of A .

Equation (2) can be optimized using an Expectation Maximization (EM) strategy. Given a current estimate $\mathcal{T}'$ of $\widehat{\mathcal{T}}$, we can expand Equation (2) with the posterior label probability $P(\mathcal{L}|\mathcal{T}', I)$

$$\widehat{\mathcal{T}} = \arg \max_{\mathcal{T}} \log \sum_{\mathcal{L}} \frac{P(\mathcal{L}|\mathcal{T}', I) \cdot P(I|\mathcal{L}) \cdot P(\mathcal{L}|\mathcal{T})}{P(\mathcal{L}|\mathcal{T}', I)} + \log P(\mathcal{T})$$

and then apply Jensen's inequality to obtain an easier to optimize, tight lower bound:

$$\log \sum_{\mathcal{L}} P(\mathcal{L}|\mathcal{T}', I) \cdot \frac{P(I|\mathcal{L}) \cdot P(\mathcal{L}|\mathcal{T})}{P(\mathcal{L}|\mathcal{T}', I)} \geq \sum_{\mathcal{L}} P(\mathcal{L}|\mathcal{T}', I) \cdot \log \left(\frac{P(I|\mathcal{L}) \cdot P(\mathcal{L}|\mathcal{T})}{P(\mathcal{L}|\mathcal{T}', I)} \right),$$

which achieves equality for $\mathcal{T} = \mathcal{T}'$. We can then update $T^{[prime]}$ by maximizing the lower bound:

$$\mathcal{T}' \leftarrow \arg \max_{\mathcal{T}} \sum_{\mathcal{L}} P(\mathcal{L}|\mathcal{T}', I) \cdot \log \left(\frac{P(\mathcal{L}|\mathcal{T})}{P(\mathcal{L}|\mathcal{T}', I)} \right) + \log P(\mathcal{T}) = \arg \min_{\mathcal{T}} D_{KL} \left(P(\mathcal{L}|\mathcal{T}', I) \parallel P(\mathcal{L}|\mathcal{T}) \right) - \log P(\mathcal{T}), \quad (3)$$

where we have dropped terms that are independent of $\mathcal{T}$ and

$$D_{KL}(P \parallel Q) = \sum_{\mathcal{L}} P(\mathcal{L}) \log \frac{P(\mathcal{L})}{Q(\mathcal{L})}$$

is the Kullback-Leibler divergence [18], a non-commutative measure of the difference between two probability distributions P and Q .

We note that this iterative scheme is an instance of the EM algorithm with the image I being the observed data, the label map $\mathcal{L}$ the unobserved data, and $\mathcal{T}$ the parameter. Similar to [14, 15], the Expectation Step (E-Step) then computes the weights

$$\mathcal{W}(\mathcal{L}) \triangleq P(\mathcal{L}|\mathcal{T}', I) \quad (4)$$

and the Maximization Step (M-Step) updates the registration parameters by solving Equation (3).

2.2 An Example Implementation

Let us now instantiate the probabilistic model of the previous section. We will denote the stacked up image voxels as $I \triangleq (I_1, \dots, I_K)$, where K is the number of voxels. Similarly, the (possibly unknown) label map $\mathcal{L} \triangleq (\mathcal{L}_1, \dots, \mathcal{L}_K)$ is a vector with K entries, where $\mathcal{L}_j \in \{1, \dots, N\}$ denotes the label at voxel j (including a label for background). For simplicity, we will assume a suitable interpolator and treat the atlas $\mathcal{A} \triangleq (\mathcal{A}_1, \dots, \mathcal{A}_N)$ as a collection of N continuous label probability images defined in the atlas coordinate frame, where at each location $k \in \{1, \dots, K\}$, $\mathcal{A}_i(k)$ encodes the prior probability of structure $i \in 1, \dots, N$. The structure-specific registration parameters $\{\mathcal{T}_i\}$ are length-9 vectors (which include 3 translational, 3 rotational and 3 scaling parameters) that define an affine mapping from the image coordinates to the atlas $\mathcal{A}_i$. Abusing notation, we will use $\mathcal{T}_i(k)$ to denote the mapped point of the image voxel k .

We now make several assumptions that allow us to simplify Equation (3):

- The registration parameters $\{\mathcal{T}_i\}$ have independent uniform priors:

$$P(\mathcal{T}) = \prod_i P(\mathcal{T}_i), P(\mathcal{T}_i) = \text{const.}$$

- The image likelihood is spatially independent $P(I|\mathcal{L}) = \prod_k P(I_k|\mathcal{L}_k)$ and $P(I_k|\mathcal{L}_k)$ depends on the imaging modality.
- The conditional label distribution is spatially independent, $P(\mathcal{L}|\mathcal{T}) = \prod_k P(\mathcal{L}_k|\mathcal{T})$, and defined as $P(\mathcal{L}_k=i|\mathcal{T}) = \frac{A_i(\mathcal{T}_i(k))}{\sum_l A_l(\mathcal{T}_l(k))}$.

Based on these assumptions, we can rewrite the E-step of Equation (4) as

$$\mathcal{W}(\mathcal{L}) = \prod_k \mathcal{W}_k(\mathcal{L}_k) \quad \text{with} \quad \mathcal{W}_k(\mathcal{L}_k) \propto P(I_k|\mathcal{L}_k) \cdot A_{\mathcal{L}_k}(\mathcal{T}'_{\mathcal{L}_k}(k)). \quad (5)$$

Using the uniform prior on $\mathcal{T}$ and inserting Equation (5) into Equation (3) yields:

$$\begin{aligned} \mathcal{T}' &\leftarrow \arg \min_{\mathcal{T}} \sum_k D_{KL}(\mathcal{W}_k(\mathcal{L}_k) \| P(\mathcal{L}_k|\mathcal{T})) \\ &= \arg \min_{\mathcal{T}} \sum_k D_{KL} \left(\mathcal{W}_k(\mathcal{L}_k) \left\| \frac{A_{\mathcal{L}_k}(\mathcal{T}_{\mathcal{L}_k}(k))}{\sum_l A_l(\mathcal{T}_l(k))} \right\| \right) \\ &= \arg \min_{\mathcal{T}} \sum_k \left(D_{KL} \left(P(I_k|\mathcal{L}_k) A_{\mathcal{L}_k}(\mathcal{T}'_{\mathcal{L}_k}(k)) \left\| A_{\mathcal{L}_k}(\mathcal{T}_{\mathcal{L}_k}(k)) \right\| \right) - \log \left(\sum_l A_l(\mathcal{T}_l(k)) \right) \right), \end{aligned} \quad (6)$$

where we abuse notation and only require that the second argument of the KL divergence normalizes to some constant, not necessarily 1. Note that in our implementation, we assume the image is globally aligned with the atlas coordinate frame, and thus fix the transformation associated with the background label to this global transformation.

In summary, the algorithm estimates the transformation parameters $\mathcal{T}$ by repeatedly solving Equation (6) until convergence. We note that the update equations do not specify the label likelihood $P(I_k|\mathcal{L}_k)$, as its definition depends on the type of medical scan, which in our case is MR data. Many methods in the literature that deal with MRI, such as [14, 15, 19, 20], model $P(I_k|\mathcal{L}_k)$ using an independent Gaussian distribution with a mean modulated by a spatially varying multiplicative factor that captures MR inhomogeneity. For convenience, we use the same model and sequentially estimate its parameters within the EM algorithm (see also [21]).

2.3 Classification of Structure-specific Transformation Parameters

In the second stage of our analysis, the estimated transformation parameters $\widehat{\mathcal{T}}$ are fed into a linear Support Vector Machine (SVM) classifier [16]. In the following, we use $\mathcal{T}$ to denote a vector obtained by concatenating the structure-specific transformation parameters $\mathcal{T}_i$ (excluding the background label). Given a collection of transformation parameter values from two groups, the linear classifier searches for a weight vector λ that maximizes the “spread” between the projected scalar values $t \triangleq \lambda^T \cdot \mathcal{T}$ of the two groups. One advantage of using a linear classifier is that its weights can be interpreted as change along the normal of the

separating hyperplane [2]: changing the value of a sample in the direction of λ will bring it closer to the positive class, while a change in the opposite direction will have the reverse effect. In our context, this allows us to illustrate the discriminating warp between the two groups of interest.

3 Classifying MR images into Schizophrenic vs. Healthy

3.1 Data

We test the accuracy of our classifier on a data set that contains MR images of 33 subjects. Sixteen of the patients were diagnosed as first episode schizophrenia and the remaining seventeen are age-matched healthy control subjects. Because first episode patients are relatively free of confounds such as the long-term effects of medication, any difference between these two groups may be considered as a more direct indication of disease-specific abnormalities compared to, say, chronic cases. Each scan was acquired on 1.5 Tesla General Electric Scanner using a SPoiled Gradient-Recalled (SPGR) sequence with a voxel dimension of $0.9375 \times 0.9375 \times 1.5 \text{ mm}^3$. Each scan was manually segmented into the (right and left) Superior Temporal Gyrus (STG) and three medial temporal lobe structures: Hippocampus (HIP), Amygdala (AMY) and Parahippocampal Gyrus (PHG). Further details about the data set can be found in [17].

3.2 Atlas Construction

Our framework requires the construction of an atlas from a collection of manually labeled training data. We first perform a global co-registration of all the 33 MRI volumes: each image is aligned with a dynamically evolving mean template [22,23]. The similarity measure is the mutual information between pixel intensities of the image and current estimate of the template, and the transformations are parameterized using a global 9 parameter affine model. We choose this approach based on our comparison of different atlas generation methods on the same data set [24].

Once all the images are spatially normalized, we compute the atlas $\mathcal{A}$. For each test subject, we construct a separate *leave-one-out* atlas: at each atlas voxel, we compute the frequency of each anatomical label, excluding the test subject. Figure 1 visualizes such an atlas, where we have rendered the 50% iso-probability surfaces for each one of the labels of interest and overlaid these on slices of the average MRI volume.

3.3 Transformation Parameters

The structure-specific transformations are parameterized using a 9 parameter affine model. The parameters include 3 transformations (t_x, t_y, t_z) along the three axes, 3 rotation angles ($\theta_x, \theta_y, \theta_z$) around the x, y, z axes, and 3 scale coefficients (s_x, s_y, s_z) along the three axes. In our convention, the x coordinate represents the medial-lateral axis with increasing values indicating the left, increasing y denotes more superior and increasing z denotes more anterior. Formally, the transformation that maps the image coordinate $k \triangleq (x, y, z)$ to the atlas coordinate

$\tilde{k} \triangleq (\tilde{x}, \tilde{y}, \tilde{z})$ is:

$$\tilde{k}^T = S \cdot R \cdot (x, y, z)^T + (t_x, t_y, t_z)^T, \quad (7)$$

where S is a diagonal matrix with diagonal elements defined by (s_x, s_y, s_z) and

$R \triangleq R_x(\theta_x) \cdot R_y(\theta_y) \cdot R_z(\theta_z)$ is the rotation matrix with $R_u(\theta_u)$ being the rotation matrix around

axis ' u '. We note that we use for the SVM the logarithm of the scale coefficients instead of the scale coefficient directly.

3.4 Automatic Segmentation of New Subject

Given a new subject with no manual labels, we run the EM-algorithm described in the previous section that jointly estimates the structure-specific registration parameters and segmentation labels. For each test subject we use its corresponding leave-one-out atlas. In practice, we have also found that treating the background label differently from the other labels improves segmentation accuracy: we fix the background label's transformation to a global transformation obtained through a pairwise image registration procedure, where the new subject is aligned with the atlas template using a B-spline transformation model [25,26]. We initialize our EM algorithm using segmentations obtained from an atlas-based method [21] that solves registration (with a B-spline model) and segmentation sequentially. As we report in [24], this method, that we use for initialization, achieves accurate segmentation for all ROIs except PHG.

3.5 Statistical Analysis

The structure-specific transformation parameters of each new subject are then considered as samples for our statistical analysis. Figure 5 plots five features that demonstrate the highest individual separation (measured using an unpaired t-test) between schizophrenics and controls. Note that, even the feature with the lowest p-value (medial-lateral translation of the left PHG) does not demonstrate complete separation between the groups. We first train a linear SVM on all 33 subjects to analyze the separation between the groups. We achieve 100% training accuracy, which can be observed in Figure 5(f) that plots the transformation parameters linearly projected on to the normal of the SVM's separating hyperplane. Since the training data was linearly separable, the linear SVM classifier was parameter-free.

As we previously discussed, the weights of the trained linear classifier can be used to analyze the discrimination between the groups. Let λ denote the weights of the classifier, such that $\lambda^T \mathcal{T}$ is positive for age-matched healthy control subjects. If we apply a warp with $\mathcal{T} = \alpha \lambda$, for some $\alpha > 0$, to the *total* atlas $\mathcal{A}$ computed on all subjects, this warped atlas will represent the "healthier" population. Similarly a warp in the opposite direction $-\alpha \lambda$ will represent a more "schizophrenic" population. We can therefore view λ as the discriminating transformation between schizophrenics and controls. Figure 2 overlays these warped atlases on top of each other to illustrate the discriminating transformation. Note that, in this figure we have chosen $\alpha = 5$, exaggerating the discrimination between the two groups for illustrative purposes.

Figure 3 shows the weights of the linear SVM classifier. Here, we performed 33 leave-one-out training procedures to obtain 33 different sets of weights. The plot shows the average over these cases and the error bars indicate the standard error, to illustrate the variability of the weights. In this analysis, the translational and scaling components are easier to interpret: e.g. a translation along the x-axis tells us whether a structure is more lateral or medial in one group and a scale along the y-axis indicates a length difference in inferior-superior direction. On the other hand, the rotational parameters are trickier to understand: a z-rotation represents a tilt of the sagittal plane. Some preliminary interpretations suggested by the analysis of Figure 3 are that in normal controls: the left STG is longer in the posterior-anterior direction, the left AMY is more posterior, and the left PHG is more lateral. Figure 2 visualizes these differences all together. We leave a detailed analysis of these results to future work.

To understand how generalization accuracy changes as a function of training data size we conducted the following jackknife experiment. We held out each subject and performed a repeated random sampling cross-validation on the remaining 32 subjects, while varying the training set size between 24 and 31. Thus, we obtained a cross-validation accuracy estimate

for each hold-out subject and training set size. Figure 4 plots the average and standard deviation of these 33 values as a function of training data size. We note that the standard deviation, indicated by the error bars, reflects our uncertainty in the generalization performance but is an underestimate of the underlying variance [27]. We observe that generalization accuracy continues to increase with the training data size, suggesting that more data would be helpful in building a better classifier for discriminating between schizophrenics versus controls. The final generalization accuracy reaches 90%.

As a further analysis, we trained separate linear SVMs on the 9 affine parameters of each structure. Only the parameters associated with left AMY, left and right HIP, and left PHG yielded a significant discrimination of the two populations, with 61%, 70%, 54% and 60% as the respective leave-one-out (LOO) cross-validation accuracy values. We also trained linear SVMs on the parameters corresponding to each hemisphere independently. According to our LOO cross-validation, the two hemispheres manifested the same discriminative power, with 67% accuracy. To measure the discriminative power of volumetric measurements only and provide a benchmark against our results, we trained a linear SVM on the volumes of the manual segmentations. Thus each sample was a length-8 vector of the absolute sizes of the manually delineated regions of interest. LOO cross-validation accuracy for this feature set was 69%. This result suggests that the 9 parameter characterization of each structure obtained from our framework yields an improved discriminative power than size measurements of manual segmentations.

4 Conclusion

In this paper, we developed a deformation-based classifier that directly assigns a medical scan to a healthy or disease patient group. The assignment is based on an anatomical parameterization of the warp that leverages a structure-based correspondence across individuals. In our implementation, the parameters are a set of structure-specific affine registration parameters. These parameters are computed by minimizing the Kullback-Leibler divergence between the posterior label probability of the subject and structure prior probability of the atlas. The affine registration parameters are then fed into a linear support vector machine for classifying the image scan. The weights of this linear classifier can then be interpreted as a warp that discriminates the two groups. On an MR data set with 16 first episode schizophrenics and 17 age-matched healthy subjects, our approach achieves a promising cross-validation accuracy of up to 90%.

Acknowledgments

The authors would like to thank William Wells, Koen Van Leemput and Wanmei Ou for their feedback on this work, and Martha Shenton for providing the data. The research was partially supported by the following grants: NIH NIBIB N01-MH-60270, NIH NCRR P41-RR13218, NIH NINDS R01-NS051826.

References

1. Shenton M, Dickey C, Frumin M, McCarley R. A review of MRI findings in schizophrenia. *Schizophrenia Research* 2001;49(1-2):1–52. [PubMed: 11343862]
2. Golland P, Grimson W, Kikinis R. Statistical shape analysis using fixed topology skeletons: Corpus callosum study. *Information Processing in Medical Imaging* 1999:382–387.
3. Styner M, Lieberman JA, Pantazis D, Gerig G. Boundary and medial shape analysis of the hippocampus in schizophrenia. *Medical Image Analysis* 2004;8(3):197–203. [PubMed: 15450215]
4. Davatzikos C, Shen D, Gur RC, Wu X, Liu D, Fan Y, Hughett P, Turetsky BI, Gur RE. Whole-brain morphometric study of schizophrenia reveals a spatially complex set of focal abnormalities. *Archives of General Psychiatry* 2005;62:1218–1227. [PubMed: 16275809]

5. Lao Z, Shena D, Xuea Z, Karacalia B, Resnickb SM, Davatzikos C. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *NeuroImage* 2004;21:46–57. [PubMed: 14741641]
6. Liu Y, Teverovskiy L, Carmichael O, Kikinis R, Shenton M, Carter C, Stenger VA, Davis S, Aizenstein H, Becker J, Lopez O, Meltzer C. Discriminative mr image feature analysis for automatic schizophrenia and alzheimers disease classification. *Medical Image Computing and Computer-Assisted Intervention* 2004:393–401.
7. Pruessner J, Li L, Serles W, Pruessner M, Collins D, Kabani N, Lupien S, Evans A. Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: Minimizing the discrepancies between laboratories. *Cerebral Cortex* 2000;10:433–442. [PubMed: 10769253]
8. Fan Y, Shen D, Davatzikos C. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. *Medical Image Computing and Computer-Assisted Intervention* 2005:1–8. [PubMed: 16685822]
9. Ashburner J, Friston K. Voxel-based morphometry - the methods. *NeuroImage* 2000;11:805–821. [PubMed: 10860804]
10. Volz H, Gaser C, Sauer H. Supporting evidence for the model of cognitive dysmetria in schizophrenia: a structural magnetic resonance imaging study using deformation-based morphometry. *Schizophrenia Research* 2000;46:45–56. [PubMed: 11099885]
11. Davatzikos C, Genc A, Xu D, Resnick S. Voxel-based morphometry using the ravens maps: methods and validation using simulated longitudinal atrophy. *NeuroImage* 2001;14:1361–1369. [PubMed: 11707092]
12. Shen D, Davatzikos C. Hammer: Hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging* 2002;21:1421–1439. [PubMed: 12575879]
13. Narr KL, Bilder RM, Toga AW, Woods RP, Rex DE, Szeszkó PR, Robinson D, Sevy S, Gunduz-Bruce H, Wang YP, DeLuca H, Thompson PM. Mapping cortical thickness and gray matter concentration in first episode schizophrenia. *Cerebral Cortex* 2005;15(6):708–719. [PubMed: 15371291]
14. Ashburner J, Friston K. Unified segmentation. *NeuroImage* 2005;26(3):839–851. [PubMed: 15955494]
15. Pohl KM, Fisher J, Grimson W, Kikinis R, Wells W. A Bayesian model for joint segmentation and registration. *NeuroImage* 2006;31(1):228–239. [PubMed: 16466677]
16. Vapnik, V. *The Nature of Statistical Learning Theory*. Springer-Verlag; 1999.
17. Hirayasu Y, Shenton ME, Salisbury D, Dickey C, Fischer IA, Mazzoni P, Kisler T, Arakaki H, Kwon JS, Anderson JE, Yurgelun-Todd D, Tohen M, McCarley RW. Lower left temporal lobe MRI volumes in patients with first-episode schizophrenia compared with psychotic patients with first-episode affective disorder and normal subjects. *The American Journal of Psychiatry* 1998;155(10):1384–1391. [PubMed: 9766770]
18. Kullback S, Leibler R. On information and sufficiency. *The Annals of Mathematical Statistics* 1951;22:79–86.
19. Van Leemput K, Maes F, Vandermeulen D, Suetens P. Automated model-based bias field correction of MR images of the brain. *IEEE Transactions on Medical Imaging* 1999;18(10):885–895. [PubMed: 10628948]
20. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging* 2001;20(1):45–57. [PubMed: 11293691]
21. Pohl K, Bouix S, Nakamura M, Rohlfing T, McCarley R, Kikinis R, Grimson W, Shenton M, Wells W. A hierarchical algorithm for MR brain image parcellation. *IEEE Transactions on Medical Imaging* 2007;26(9):1201–1212. [PubMed: 17896593]
22. Guimond A, Meunier J, Thirion JP. Average brain models: A convergence study. *Computer Vision and Image Understanding* 1999;77(2):192–210.
23. Lorenzen P, Prastawa M, Davis B, Gerig G, Bullitt E, Joshi S. Multi-modal image set registration and atlas formation. *Medical Image Analysis* 2006;10(3):440–451. [PubMed: 15919231]

24. Zöllei L, Shenton M, Wells W, Pohl K. The impact of atlas formation methods on atlas-guided brain segmentation, statistical registration. Pair-wise and Group-wise Alignment and Atlas Formation Workshop at MICCAI 2007: Medical Image Computing and Computer-Assisted Intervention 2007:39–46.
25. Rueckert D, Sonoda L, Hayes C, Hill D, Leach M, Hawkes D. Non-rigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging* 1999;18(8):712–721. [PubMed: 10534053]
26. Rohlfing T, Maurer CR Jr. Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. *IEEE Transactions on Information Technology in Biomedicine* 2003;7(1):16–25. [PubMed: 12670015]
27. Bengio Y, Grandvalet Y. No unbiased estimator of the variance of k-fold cross-validation. *Journal of Machine Learning Research* 2004;(5):1089–1105.

Fig. 1.

Three dimensional renderings of the 50% iso-probability surfaces for each structure over-laid on the average MRI scan: STG is shown in brown and blue; HIP in cyan and yellow; AMY in red and green; and PHG in pink and purple.

Fig. 2.

Discriminating transformation between the two groups: Axial and coronal views of (a-b) the atlas warped towards the control group (in color) overlaid on the atlas warped towards the schizophrenic group (Schiz.) in semi-transparent gray; (c-d) the atlas warped towards the schizophrenic group (in color) overlaid on the atlas warped towards the control group (semi-transparent gray).

Fig. 3.
The SVM weights for the affine parameters of the right and left anatomical structures of interest

Fig. 4.
Cross-validation accuracy versus training size.

Fig. 5.

(a-e): Boxplots of the top five structure-specific transformation parameters that demonstrate the highest separation between the two groups. p-values for unpaired t-tests are included in the title of each plot. (f) The transformation parameters when linearly projected on to the normal of the separating hyperplane of the SVM.