A Statistical Parts-Based Model of Anatomical Variability

Matthew Toews* and Tal Arbel

***Abstract*—In this paper, we present a statistical parts-based model (PBM) of appearance, applied to the problem of modeling intersubject anatomical variability in magnetic resonance (MR) brain images. In contrast to global image models such as the active appearance model (AAM), the PBM consists of a collection of localized image regions, referred to as *parts*, whose appearance, geometry and occurrence frequency are quantified statistically. The parts-based approach explicitly addresses the case where one-to-one correspondence does not exist between all subjects in a population due to anatomical differences, as model parts are not required to appear in all subjects. The model is constructed through a fully automatic machine learning algorithm, identifying image patterns that appear with statistical regularity in a large collection of subject images. Parts are represented by generic scale-invariant features, and the model can, therefore, be applied to a wide variety of image domains. Experimentation based on 2-D MR slices shows that a PBM learned from a set of 102 subjects can be robustly fit to 50 new subjects with accuracy comparable to 3 human raters. Additionally, it is shown that unlike global models such as the AAM, PBM fitting is stable in the presence of unexpected, local perturbation.**

***Index Terms*—Intersubject variability, invariant feature, parts-based model, statistical appearance model.**

I. Introduction

In magnetic resonance (MR) images of the human brain, no two subjects are identical. Anatomical structure or tissue may exhibit significant appearance variation from one subject to the next, or may simply not exist in all subjects, in the case of pathology for example. This phenomenon is referred to as intersubject appearance variability, or the manner in which images of different subjects vary within a population. Effectively quantifying intersubject variability is of great importance to the medical imaging community, as it lies at the heart of understanding how anatomical structure varies within a population. This leads to various open research questions: What image structures are common within a population, what structures are rare, and how do they vary in appearance and geometry? In what ways is a (possibly pathological or abnormal) subject similar to or different from the population?

There is general agreement that intersubject variability within a population should be quantified statistically, via a common reference frame or model with respect to which subjects can be compared. The particular manner in which variability is quantified depends on the purpose of the model. Certain models are intended to quantify appearance and geometry for intersubject registration [12] or segmentation [30]. Others quantify localized intensity or segmentation variation once subjects have been aligned into a common reference frame [1]. Yet others quantify the deformation fields that bring subjects into alignment [31], [37]. In order to quantify variability over a set of different subjects in a meaningful manner, they must first be aligned to ensure that quantification is based on the same underlying brain structure. This requires addressing the task of intersubject (or similarly model-to-subject) registration.

A large number of techniques have been proposed for the task of intersubject registration, the majority of which either implicitly or explicitly formulate the task as one of determining one-to-one correspondence between subjects or between a model and a subject [19], [31]. Given the degree of variability present in brain imagery of different subjects however, particularly in highly variable regions such as the cortex or in case of abnormal or pathological subjects, it is reasonable to expect that one-to-one correspondence may not exist. Furthermore, it is reasonable to expect that models based on one-to-one correspondence perform poorly in locations where such correspondence does not exist. Although the case of diffeomorphic relationships between subjects has received a large amount of attention [19], [23], [39], the case where one-to-one correspondence does not exist is rarely modeled, with the exception of outlier detection techniques [4].

The contribution of this paper is a new parts-based statistical model of appearance, designed specifically to quantify intersubject variability within a population in the case where *one-to-one correspondence between subjects does not exist*. The model represents a population such as MR brain images as a collection of "parts," which we define as spatially localized image regions. Each model part consists of an appearance, a geometrical description and an occurrence frequency, all of which are quantified statistically. Our model explicitly accounts for occlusion and intersubject variation on a local scale, as model parts are not expected to (and typically do not) occur in all images. Model parts are based on generic scale-invariant image features, oriented image regions that can be automatically extracted in a wide variety of images and robustly aligned between images over a range of intensity and geometrical deformations. As such, model parts do not represent obvious anatomical structures that a neuroanatomist might identify per se, but rather image patterns that occur with statistical regularity within a population.

The parts-based model (PBM) we present represents several important advancements with respect to current statistical

Manuscript received September 15, 2006; revised November 1, 2006. *Asterisk indicates corresponding author.*

*M. Toews is with the Centre for Intelligent Machines, McGill University, Montreal, QC H3A 2A7, Canada (e-mail: mtoews@cim.mcgill.ca).

T. Arbel is with the Centre for Intelligent Machines, McGill University, Montreal, QC H3A 2A7, Canada (e-mail: arbel@cim.mcgill.ca).

Digital Object Identifier 10.1109/TMI.2007.892510

appearance models. The model can be constructed over a large set of training images via a fully automatic machine learning procedure with no manual interaction, automatically discovering a set of image parts that occur with regularity in a population in the presence of significant intersubject variation. All subjects of a population can be modeled simultaneously without making *a priori* classifications as to which subjects are "normal," as image structure common to multiple subjects is automatically recognized, allowing irregular structure due to subject-specific or abnormal characteristics such as pathology to be disregarded. The model can be robustly fit to new images in the presence of intersubject variation and abnormality, in addition to global image translation, rotation and scale changes, without relying on iterative search techniques prone to getting trapped in suboptimal local minima when initialized outside of a "capture range." Parts-based model fitting is stable in the presence of local image perturbations, unlike the fitting of global models such as the active appearance model (AAM) [12], in the sense that a local perturbation results in a local change to the fitting solution. Finally, the set of localized model parts serves as a natural and intuitive basis for describing anatomical structure, in comparison with other representations such as modes of global covariance.

The remainder of the paper is organized as follows: we begin with an overview of related work in Section II. In Section III, we present the PBM, including algorithms for automatic model learning and fitting. In Section IV, we present experimentation based on 2-D slices of T1-weighted MR brain images from the ICBM (International Consortium for Brain Mapping) 152 data set [9], consisting of 152 volumes of normal subjects, 88 male and 66 female, aged 24.6 ± 4.8 years. We demonstrate that a set of model parts can be automatically learnt from a large set of training images, and sorted according to their statistics in order to identify the image patterns most representative of the training set. Furthermore, quantitative evaluation of model-to-subject registration reveals that the PBM can be automatically fit to new images with accuracy comparable to human raters, and comparison with the AAM demonstrates the superior stability of PBM fitting in the presence of local perturbations. Finally, we conclude with a discussion in Section V.

II. Related Work

In this section, we outline the main bodies of research relevant to our approach, namely intersubject registration, statistical appearance modeling, and parts-based modeling.

A. Intersubject Registration

In order to quantify variability between subjects in a population, correspondence must first be established between subjects, via intersubject registration. In general, most registration techniques attempt to determine a one-to-one diffeomorphic mapping from one subject to the next [19], [31], driven by a measure of image similarity. Low-parameter linear registration techniques are simple and capable of determining initial coarse alignment between subjects, but do not properly account for intersubject variation on a local scale. Deformable registration techniques involving the estimation of highly parameterized deformation fields from one image to the next have been proposed to account for variation on a local scale. Different deformable registration formulations can generally be distinguished by the manner in which the deformation field is constrained or regularized [22], examples include elastic registration [2], fluid registration [7], finite element models [18], thin plate splines [6], etc.

In the case of intersubject registration, however, it is not clear that the estimated deformation fields represent meaningful correspondence between underlying tissues [31]. Different regularization approaches will generally result in different deformation field solutions for a given pair of subjects, particularly in regions where correspondence is ambiguous, such as areas of homogenous image intensity. In general, the higher the number of deformation parameters, the more ways a deformation field can be constructed between any two subjects to obtain a near-perfect mapping in terms of pixel intensity error. This could explain the proliferation of novel highly-parameterized registration formulations in the medical imaging literature that report low error in terms pixel intensity. Recently, quantitative comparison of 6 different intersubject registration methods showed no significant difference between high and low parameter registration approaches [22], in which case the principle of Occam's razor [16] would suggest that the simpler, reduced parameter model would be more plausible.

In general, these difficulties illustrate the challenge of coping with intersubject variability in registration, particularly in the absence of a gold standard for verification of intersubject registration. Given these difficulties, the *a priori* assumption of the existence of a one-to-one mapping between different subjects is often difficult to justify, considering highly variable regions such as the cortex, the cases of pathology and tissue resection, etc. We propose that intersubject registration relax the requirement for a one-to-one mapping, and focus instead on identifying distinct patterns of local image structure that occur with a certain probability within a population.

B. Statistical Appearance Models

Given that exact correspondence between subjects of a population is generally difficult and ill-posed, research has focused on statistical appearance models to quantify intersubject variability. Statistical models consist of parameters which can be estimated from a set of data in order to quantify variability in a meaningful manner. Parameters of image-based models are typically based on variables of image intensity, image space, and mappings between different images. As images consist of a large number of data samples (intensities), directly modeling all samples and their interdependencies is generally intractable. Statistical modeling must, thus, resort to simplifying assumptions regarding parameters and their dependencies, which can be broadly classified according to the image scale at which modeling takes place, i.e. global versus local models.

Global models refer to those considering the entire spatial extent of the image simultaneously [12], [32]. Variation in global models is typically quantified via a linear Gaussian model, consisting of additive "modes" of covariance about a mean. As all image data are related linearly, the modeling assumption is one of statistical dependence between spatially separated image regions, and model simplification arises from the fact that the majority of data variance can be accounted for by a small number principle components [38]. Global models can be applied in a variety of ways, e.g. over combined intensity and shape [12], over deformation fields between subjects [32], etc. Local models refer to those that consider variation on a spatially localized scale, typically once images have been brought into alignment, such as morphometry methods [1]. The modeling assumption is that once aligned within a common reference frame, spatially separated image regions can be treated as statistically independent, and model simplification arises from the fact that only local interdependencies are modeled. Other models such as Markov models [30] can be considered a hybrid of global and local characteristics, where modeling is based on spatially local interactions which can propagate globally.

A major difficulty in attempting to statistically quantify intersubject variability is that models based either implicitly or explicitly on the assumption of one-to-one correspondence tend to break down when the assumption is invalid. Global models assuming statistical dependency between spatially distinct regions are generally unable to properly account for deformation on a local scale. While this may be less important when considering the shape of simple isolated structures such as the hypothalamus [30], it is problematic when modeling the brain as a whole where localized variations are commonplace. While local models are able to account for spatially localized phenomena, they are prone to confounding observations of different underlying brain structure when the assumption of one-to-one correspondence is invalid, and susceptible to error introduced by the method used to obtain alignment, such as bias [23]. To minimize these problems, modeling is often based on "normal" brains free of abnormality [19], [37]. It has been observed, however, that even brains classified as clinically "normal" can vary significantly [32].

C. *Parts-Based Modeling*

The parts-based approach to modeling generally advocates representing a complex image pattern, such as a brain image, in terms of a collection of simpler, independently observable parts [5]. A practical PBM requires a reliable automated means of identifying or detecting model parts in images. The development of robust generic scale-invariant feature detectors [26], [29] has recently led to automated approaches for learning PBMs from large, noisy data sets [14], [17]. Our approach builds on a recent model capable of learning a rich multi-modal model of hundreds or thousands of features [35], [36]. This approach is well suited to describing the complex nature of brain anatomy, particularly in the case where one-to-one correspondence is ambiguous or nonexistent, as model parts are not required to occur in all images.

III. A Statistical Parts-Based Appearance Model

In this section, we describe our PBM, including scale-invariant features, the model components, the probabilistic model formulation, and the automated procedures to learn the model from a set of subject images and to fit the model to new subjects.

A. *Generic Scale-Invariant Features*

The statistical parts-based appearance model presented in this paper is designed to be general and automatically applicable in a wide variety of imaging contexts. The model parts are, therefore, based on generic image features that can be automatically detected and matched in robust manner. This represents a significant advantage over approaches based on special-purpose detectors for specific image structures (i.e. sulci [27]) which do not generalize easily to new contexts, or manual landmark selection which is tedious for large data sets and prone to interrater variation, as a human must decide which landmarks are optimal and how many to use.

Automatic generic feature detection for image correspondence is based on the notion that a small set of distinct, localizable image patterns can be reliably extracted and matched from one image to the next. In the computer vision literature, early detection approaches focused on identifying the image location of features such as corners, e.g. the Harris detector based on analysis of image derivatives within a local window [21]. Scale-invariant features improved upon such approaches by localizing image patterns in scale in addition to image location, following the observation that pattern distinctiveness was intimately tied to the scale or size of the local image window around the pattern [33]. The general invariant-feature approach is based on detecting image patterns in a manner such that they can be automatically normalized with respect to intensity and geometric variation, at which point they can be efficiently matched between images without requiring an explicit search over deformation parameters.

Recent feature detectors are invariant to linear intensity variation, in addition to geometrical deformations such as translation, rotation, scale and affine transformations [26], [29], and have been shown to be effective in a wide range of image contexts. Fast feature detection algorithms exist based on image pyramids. Feature geometrical parameters such as location x, orientation θ and scale σ are recovered in the extraction process and can be used to formulate multiple independent hypotheses as to the geometrical transform between images, leading to occlusion and outlier-resistant correspondence. In addition, scale-invariant features can be extracted from a variety of different image properties such as blobs [26], edges [29], phase [8] and entropy [24], all of which can be used in conjunction.

Despite their attractive properties, scale-invariant features are not widely used for intersubject registration. This is due primarily to intersubject variability, where the variation of features extracted in images of different subjects is generally too significant to reliably determine correspondence. As seen in Fig. 1, many features result from ambiguous or subject-specific image structure, and as such are not useful for determining correspondence between subjects. Although this can be considered a shortcoming of generic feature correspondence, we argue that it merely reflects the difficulty of the intersubject registration task, and supports the notion that one-to-one correspondence between different subjects may not generally exist. The PBM presented in this section is based on statistically quantifying the occurrence frequency, appearance and geometry of features over a large set of subjects, thereby learning a set of parts that

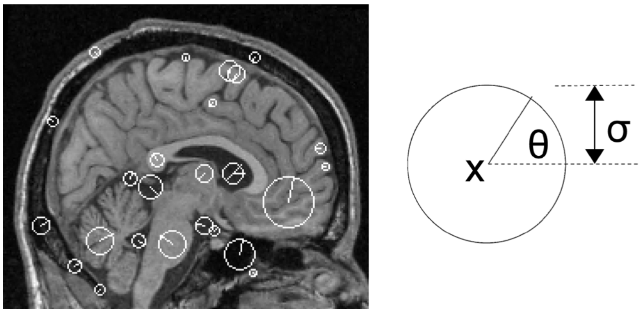


Fig. 1. A subset of scale-invariant features extracted in a T1-weighted MR brain image. Scale-invariant features, illustrated as white circles inset by radial lines, are oriented regions characterized geometrically by their location x, orientation θ and scale σ within an image. The geometric characterization results from the detection process and the image content. The same features can be detected in the presence of image translation, scale and orientation change, in addition to local intensity change. Note that many features result from subject-specific or ambiguous image structure, such as the skull.

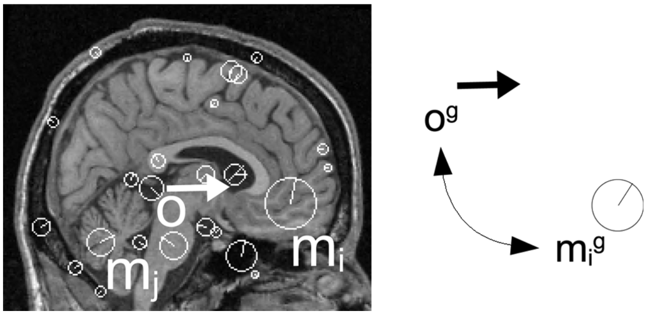


Fig. 2. Illustrating model parts and a reference frame in a sagittal slice of a T1-weighted MR brain image. In the left image, reference frame o is illustrated as a white arrow, and represents the projection of the Talairach AC-PC line onto the slice. Reference frame and model part geometry o^g and m_i^g are related via an invertible linear transform $t_i : m_i^g \rightarrow o^g$, as illustrated by the diagram on the right and, thus, a single observed part is sufficient to infer the reference frame geometry.

can be used to determine statistically meaningful correspondence between images of different subjects. In the following section, we present the components and the probabilistic formulation of the model, and describe how such a model can be learned from training data, and fit to a new image instances.

B. Model Components

The PBM consists of a set of parts $\{m_i\}$ modeled with respect to a common reference frame o. By reference frame, we mean a geometrical structure with respect to which the geometric variability of m_i can be normalized and quantified. Although the particular reference frame used is application specific and can be chosen in a variety of ways, a well-known example in the context of MR brain imagery is the midplane line defining the Talairach stereotaxic space [34], which passes from the superior aspect of the anterior commissure to the inferior aspect of the posterior commissure, as illustrated in Fig. 2. The significance of the reference frame to statistical modeling is that different parts m_i and m_j can be considered as statistically independent given o. Specifically, knowing o, parts m_i can be automatically normalized with respect to reference frame scale, rotation and translation, at which point their appearance and remaining geometrical variation can be statistically quantified locally. This is tied closely to the Bookstein definition of shape as the geometric variability remaining in a pattern once global scale, rotation and translation are removed [15].

In this paper, a model part is denoted as $m_i : \{m_i^b, m_i^g, m_i^a\}$ representing the occurrence, geometry and appearance of a scale-invariant feature within an image, respectively. Part occurrence m_i^b is a binary random variable representing the probability of part presence (or absence) in an image. Part geometry $m_i^g : \{x_i, \theta_i, \sigma_i\}$ is an oriented region in $\mathbf{R}^N$ image space, represented by N-parameter location x_i, an $N-1$ parameter orientation θ_i, and a scale σ_i. Part appearance m_i^a describes the image content at region m_i^g, and can generally be parameterized in a number of ways, such as principle components [38].

The reference frame is denoted as $o : \{o^b, o^g\}$ representing the occurrence and geometry of a common structure with respect to which the geometric variability of parts can be quantified. Note that o is identical to a model part with the exception of the appearance component. This is because o cannot be observed directly, but must be inferred via model parts m_i. Reference frame occurrence o^b is a binary random variable indicating the presence or absence of the reference frame, whose significance will be made clear in the discussion of model learning and fitting. Reference frame geometry o^g is parameterized in the same manner as model part geometry m^g, implying that a single observed part m_i is sufficient to infer o^g in a new image, via a learned linear relationship. In the following section, we describe the formulation of the probabilistic model underpinning quantification of part appearance, geometry and occurrence frequency.

C. Probabilistic Model Formulation

Our model consists of a set of M model parts $\{m_i\}$, which when observed in a new image can be used infer the reference frame o. Assuming that parts m_i are conditionally independent given o, the posterior probability of o given $\{m_i\}$ can be expressed using Bayes rule as

$$p(o|\{m_i\}) = \frac{p(o)p(\{m_i\}|o)}{p(\{m_i\})} = \frac{p(o)\prod_i^M p(m_i|o)}{p(\{m_i\})} \tag{1}$$

where $p(o)$ is a prior over reference frame geometry and occurrence and $p(m_i|o)$ is the likelihood of part m_i given o. The assumption of conditional independence of localized parts generally states that knowing reference frame o, all remaining part variation m_i can be described locally. This assumption represents a significant departure from global modeling approaches, such as the multi-variate Gaussian based on principle components, where all image intensities are correlated and statistically dependent. Such global models cannot account for deformation on a local scale, however, where the appearance of one part of the image violates the learned linear model with respect to other parts of the image. We demonstrate this later in experimentation.

Our model focuses principally on the likelihood term $p(m_i|o)$, which can be expressed as

$$\begin{aligned} p(m_i|o) &= p\left(m_i^a, m_i^b|o\right) p\left(m_i^g|o\right) \\ &= p\left(m_i^a|m_i^b\right) p\left(m_i^b|o^b\right) p\left(m_i^g|o^b, o^g\right) \end{aligned} \tag{2}$$

TABLE I
OCCURRENCE PROBABILITY $p(m_i^b|o^b)$

	o^b	m_i^b	**Interpretation**
π_i^0	0	0	Not applicable
π_i^1	0	1	False part occurrence
π_i^2	1	0	Occluded part
π_i^3	1	1	True part occurrence

under the assumptions that m^a and m^b are statistically independent of m^g given o, and that m^a and o are statistically independent given m^b. The first of these assumptions generally states that knowing o, the appearance and occurrence of a part offer no further information regarding part geometry. The second assumption generally states that given knowledge of part occurrence, knowing o offers no additional information regarding part appearance. We describe the three terms of (2) below.

1) Appearance Likelihood $p(m_i^a|m_i^b)$*:* Appearance likelihood $p(m_i^a|m_i^b)$ represents the appearance of a part after geometry normalization, and can generally be modeled as a multivariate Gaussian distribution in an appearance space with mean and covariance parameters μ_i^a, Σ_i^a. There are two distributions that must be estimated, corresponding to the cases $m_i^{b=1}$ and $m_i^{b=0}$. In the later case, we take $p(m_i^a|m_i^b)$ to be constant or uniform. We discuss the specific appearance space used in Section IV.

2) Occurrence Probability $p(m_i^b|o^b)$*:* Occurrence probability $p(m_i^b|o^b)$ represents part occurrence given reference frame occurrence, and is modeled as a discrete multinomial distribution with event count parameters $\pi_i = \{\pi_i^0, \ldots, \pi_i^3\}$ for the 4 possible combinations of binary events. Table I lists the events and their interpretations. Event π_i^0 is not of interest, as neither the part nor the reference frame are present. Event π_i^1 represents the case where part m_i is observed in the absence of reference frame o, this can be viewed as a false part occurrence. Event π_i^2 represents the case where reference frame o is present but part m_i is absent, this is the case where the part is occluded or unobservable due to intersubject variability. Finally, event π_i^3 represents the case that both part m_i and reference frame o are present.

An aspect of the occurrence probability bearing mention at this point is the likelihood ratio of true versus false part occurrences

$$\frac{p\left(m_i^{b=1}|o^{b=1}\right)}{p\left(m_i^{b=1}|o^{b=0}\right)} = \frac{\pi_i^3}{\pi_i^1}. \tag{3}$$

We refer to this ratio as the *distinctiveness* of a part within this paper, as it provides a measure of how reliably a part can be identified in the presence of noise and false matches [14]. The distinctiveness plays an important role both in automatic model learning and in model fitting, described later.

3) Geometrical Likelihood $p(m_i^g|o^b, o^g)$*:* Geometrical likelihood $p(m_i^g|o^b, o^g)$ models the residual error of a linear transform from part to reference frame geometry $t_i : m_i^g \rightarrow o^g$, and is represented as a Gaussian distribution with mean and covariance parameters μ_i^g, Σ_i^g. As with the appearance likelihood, there are two distributions corresponding to cases $o^{b=1}$ and $o^{b=0}$. In the later case, o^g and, thus, $m_i^g \rightarrow o^g$ are undefined, and we treat the geometrical likelihood as uniform or constant. In order to characterize geometrical error in a scale-invariant manner, the scale dimension is transformed logarithmically, and translation magnitude is normalized by reference frame scale.

D. Model Learning

Learning involves automatically identifying a set of informative model parts $\{m_i\}$ and estimating the parameters of their appearance, geometry and occurrence frequency distributions, based on a set of subject images. Prior to learning, each training image is processed by labeling a reference frame o^g and automatically extracting features $\{m_i^a, m_i^g\}$. For the purpose of this paper, we adopt the AC-PC line defining the Talairach stereotactic reference frame as o^g, as illustrated in Fig. 2. Other definitions could be used depending on the image context, the sole constraint being that the reference frame represent a stable scale-invariant structure shared by all subjects being modeled. Labeling can be done manually by defining a single line segment corresponding to o^g in each subject image, or in an approximate manner via linear registration of MR volumes into the same stereotactic space. We adopt the latter approach, using MR volumes preregistered into the MNI stereotactic space [10].

Parameter estimation is based on a set of data vectors of the form $\{m_i^a, m_i^g, o_i^g\}$, where o_i signifies the reference frame instance associated with feature m_i. A model part corresponds to a cluster of data vectors that are similar in geometry and appearance, and parameter estimation requires identifying these clusters. A variety of clustering techniques could be used for this purpose, but as the data vector space contains a large, unknown number of clusters or modes, approaches such as EM (expectation-maximization) or K-means are ineffective [16], as they require prior knowledge as to the number of parts present, and tend to either group vectors arising from different clusters or separate vectors arising from the same cluster. We, thus, adopt a different approach, identifying all isolated clusters in a robust manner similar to the mean shift technique [11]. We do this by treating each feature m_i as a potential cluster center or model part, grouping features into a reduced set of parts, and finally estimating part statistics.

Treating each extracted feature m_i as a potential model part, feature grouping proceeds as follows. The first step is to generate a set G_i of data vectors m_j similar in geometry to m_i, such that the error in predicting the reference frame geometry is less than an empirically determined threshold Thres^g

$$G_i = \left\{\forall\, m_j : \left|t_i\left(m_j^g\right) - o_j^g\right| < Thres^g\right\}. \tag{4}$$

Recall that $t_i : m_i^g \rightarrow o_i^g$ is a linear transform between feature and reference frame geometries. $t_i(m_j^g)$, thus, represents the geometry of reference frame o_j^g as predicted by m_i^g, o_i^g and m_j^g. $\mathrm{Thres}^g = \{T_x, T_\theta, T_\sigma\}$ represents a scale-invariant threshold on the maximum acceptable error permitted in the location, orientation and scale of the predicted reference frame geometry. As Thres^g is not related to features themselves but rather to their ability to predict the reference frame geometry, a single threshold is applicable to all features. We adopt a threshold of

$T_x = (\sigma/2)$ voxels (where σ is the scale of o_j^g), $T_\theta = 15$ degrees and $T_\sigma = log(1.5)$ (a scaling factor range of 0.66 to 1.5 times).

The next step is to generate a set A_i of data vectors m_j similar in appearance to m_i, such that the difference between feature appearances m_i^a and m_j^a is less than a threshold Thres_i^a

$$A_i = \left\{\forall\, m_j : dist\left(m_i^a, m_j^a\right) < Thres_i^a\right\}. \qquad (5)$$

Thres_i^a represents a threshold on the Euclidean distance between m_i^a and m_j^a, which is consistent with the Mahalanobis distance in a feature space of independent and identically distributed features [16]. Here, Thres_i^a is automatically set to maximize the ratio of features that are similar in appearance and geometry versus features that are similar in appearance but not in geometry

$$\mathrm{Thres}_i^a = \underset{Thres_i^a}{\mathrm{argmax}} \left\{ \frac{|G_i \cap A_i|}{|\bar{G}_i \cap A_i|} \right\}. \qquad (6)$$

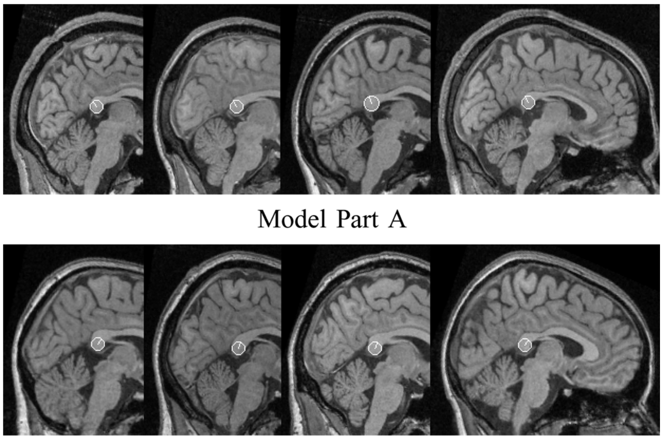
Note that the ratio in (6) is equivalent to the distinctiveness in (3), and Thres_i^a is, thus, determined to maximize the likelihood of a correct match versus an incorrect match.

Once G_i and A_i have been determined for each feature m_i, the set of features can be reduced into a small number of representative model parts. There are several mechanisms by which this is possible. Features with arbitrarily low distinctiveness can be removed [35], as they are uninformative. Features with high mutual information with other features can also be removed [3], [35], as they are redundant. We generate a set R of features m_j to remove, where m_j are similar to, but occur less frequently than, some other feature m_i

$$R = \{\forall\, m_j : m_j \in G_i \cap A_i, |G_j \cap A_j| < |G_i \cap A_i|\}. \qquad (7)$$

This approach has the effect of discarding redundant features, while maintaining those that are most representative as determined by their occurrence frequency. Feature removal generally reduces the entire feature set by an order of magnitude into a set of model parts. Estimation of part parameters then proceeds as follows. Events o^b and m_i^b are determined by membership in sets G_i and A_i, respectively, allowing the estimation of event count parameters π_i. Geometry and appearance parameters $\{\mu_i^g, \Sigma_i^g\}$ and $\{\mu_i^a, \Sigma_i^a\}$ are determined from the geometries and appearances of features in set $G_i \cap A_i$.

It is important to note that this learning process is effective at determining a set of distinct model parts useful for the task of model-to-subject registration, as the ratio of correct versus incorrect matches is maximized. Model parts do not necessarily correspond one-to-one with anatomical structures. Features resulting from a single anatomical structure or region, for instance, generally exhibit several stable modes of geometry and appearance, due to the interaction of anatomical variability and the particular feature detector used. In such a cases, model learning generally represents each mode as a different model part. This is illustrated in Fig. 3, where the same section of the corpus callosum results in two parts with distinct modes of orientation. Relaxing the threshold Thres^g on the permitted geometrical error results in fewer parts, each capable of describing a wider range of variability of the underlying image content, at the cost of decreased part distinctiveness.



Model Part A

Model Part B

Fig. 3. Illustrating instances of two different model parts, labeled A and B, arising from the same anatomical structure. In general, the same underlying anatomical structure can give rise to multiple model parts in the learning process, due to the interaction between the feature detector used and the image characteristics of the anatomical structure. Here, a section of the corpus callosum results in features with two significant orientation modes, which are grouped into two distinct model parts by the learning process.

E. Model Fitting

Once the model has been learned, it can be fit to a new image by inferring the geometry of the reference frame o based on features extracted in the new image. Unlike other registration or fitting techniques based on iterative algorithms which tend to go awry when started outside a "capture range" of the optimal solution, the PBM can be fit globally in a robust manner. Fitting begins by first matching features extracted in the new image to the learned model parts. An image feature m is considered to match a model part m_i if $dist(m_i^a, m^a) < Thres_i^a$, as in (5). While the reference frame geometry is initially unknown, each matching image feature/model part fixes m_i^a and m_i^g as evidence in likelihood (2), and infers a hypothesis as to the reference frame geometry o^g in the new image via the learned linear relationship. Dense clusters of similar hypotheses indicate the presence of a reference frame instance, and their corresponding features indicate model-to-subject matches. We are interested in evaluating whether a hypothesis cluster is the result of a true model instance or a random noisy match, i.e. $o = \{o^g, o^{b=1}\}$ or $\overline{o} = \{o^g, o^{b=0}\}$. These two hypotheses can be compared via a Bayes decision ratio

$$\gamma(o) = \frac{p\left(o|\{m_i\}\right)}{p\left(\overline{o}|\{m_i\}\right)} = \frac{p(o)}{p(\overline{o})} \prod_{i=1}^{M} \frac{p(m_i|o)}{p(m_i|\overline{o})} \qquad (8)$$

where high $\gamma(o)$ indicates the presence of a model, and $p(o)/p(\overline{o})$ is a constant representing the expected ratio of true to false model instances. Note that the value of $\gamma(o)$ is in large part determined by the part distinctiveness defined in (3), as highly distinctive parts will carry greater weight in determining the model fit. Fig. 4 illustrates the result of fitting a model to a new sagittal slice, including the initial set of hypotheses o^g, the hypothesis maximizing $\gamma(o)$, and the features which contributed to the maximal hypothesis. Note that due to the nature of invariant features, the model can be automatically

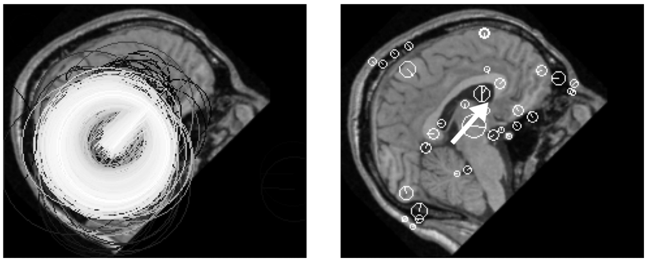


Fig. 4. Illustrating fitting of the PBM to a new image, based on sagittal MR slices. The new image has been scaled, rotated and translated in order to illustrate the scale-invariant nature of model fitting. The left image illustrates all reference frame hypotheses o^g based on matches from image features to model parts m_i, where the intensity of each o^g is proportional to the distinctiveness (3) of the model part that produced it. Note the dense clustering of distinct o^g hypotheses surrounding the true value, defined as the projection of the AC-PC line in the sagittal plane, see Fig. 2. The right image illustrates the hypothesis o^g maximizing $\gamma(o)$ as a white arrow, along with features in the new image that were matched to the model, contributing to the fitting hypothesis.

fit in the presence of global image translation, orientation and scale changes, in addition to intersubject variability.

IV. Experimentation

When dealing with MR brain images, there are several general questions that can be asked of a statistical model of appearance. What is the extent of the variability within a population (i.e. what does the model tell us)? How does a particular subject relate to the population (i.e. model-to-subject registration)? How does the model compare with others in the literature? In this section, we show how the statistical PBM can be used to answer these questions. In Section IV-A, we detail the result of model learning based on 102 subjects from the ICBM 152 data set [9]. In Section IV-B, we evaluate model-to-subject registration of the learned PBM, where the model is fit to a set of 50 new test subjects not used in learning. Finally in Section IV-C, we compare model fitting of the PBM and the AAM, showing that the PBM can be fit in a manner that is robust and stable in the presence of local deformations.

A. Learning

For the purpose of experimentation, we learn a PBM from 102 sagittal slices of the ICBM 152 data set using the fully automatic procedure described in Section III-D. As mentioned in Section III-A, a variety of techniques can be used to identify scale-invariant features from on a number of image characteristics. For the purpose of this study, we use the so-called SIFT (scale-invariant feature transform) technique [26].[1] We chose the SIFT feature detector and appearance representation, as they have been shown to be superior in terms of detection repeatability in comparisons with other approaches [28], and are currently widely used in the computer vision literature. Briefly, SIFT features are extracted as maxima/minima in a DoG (difference-of-Gaussian) scale space image pyramid. The pyramid location at which features are extracted determines the location and scale components x_i and σ_i of the feature geometry m_i^g. The orientation component θ_i is then determined from peaks in an orientation histogram generated from local image derivatives. Features extracted in the DoG scale space tend to correspond to blob-like image structures. The SIFT appearance representation m_i^a is a 128-value vector, corresponding to bins of a histogram of image first derivatives quantized into $8 \times 4 \times 4 = 128$ bins over orientation and 2-D spatial location, respectively.

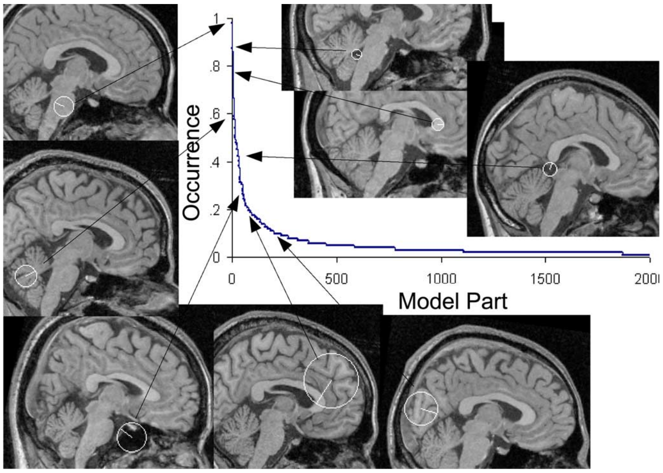


Fig. 5. A graph of learned model parts sorted by descending occurrence frequency $\pi_i^3 = p(m_i^{b=1}|o^{b=1})$. The images illustrate parts which occur at the indicated frequency within the population. Note that part occurrence drops off sharply, indicating that a relatively small number of model parts are common to all brains, whereas a large number are specific to a small number of brains.

The result of learning is a set of spatially localized parts, including their occurrence frequency, geometrical variability and appearance variability, a natural and intuitive representation for describing anatomical variability. Fig. 5 illustrates a graph of the likelihood of model part occurrence in the population, i.e. $p(m_i^{b=1}|o^{b=1})$, sorted in order of descending likelihood. Note that a small number of model parts occur in a relatively high percentage of subjects, indicating structure that is stable and repeatably identifiable in many members of the population. Conversely, a relatively large number of model parts are related to subject-specific characteristics or noise, occurring in only a small number of subjects. Such parts are of limited use in describing the anatomy of the population, and may be disregarded. Note that in general, no model parts are detected in all 102 subject brains. This is not to say that common anatomical structures such as the pons or corpus callosum are not present in all brains, but that scale-invariant features arising from such structures tend to cluster into several distinct modes of appearance and geometry and are identified as distinct model parts, as illustrated in Fig. 3 of Section III-D. This is a characteristic of our learning approach, which seeks a set of maximally distinctive image patterns that arise from regularity in underlying brain anatomy, for the task of intersubject registration.

Model parts can be potentially useful in a number of ways. In the following sections, we demonstrate that model parts serve as a basis for robust, stable model-to-subject registration, and that a subset of common parts can be used to quantify model fitting accuracy. Model part statistics can be used in order to interpret the result of model fitting in a meaningful, quantitative manner. For example, the distinctiveness in (3) represents the certainty with which a model part can be identified in a new image, and the geometrical likelihood in Section III-C-3 represents the variability that can be expected in localizing a model part in space, orientation and scale. For the purpose of anatom-

[1] An efficient SIFT implementation is publicly available [25].

ical study, model parts could be grouped and assigned semantic labels according to their underlying anatomical structures and tissues, after which point model-to-subject registration could be used to propagate these labels in an automatic, robust manner to hundreds of new subjects. It is possible that distributions of part appearance and geometry could serve as useful indicators of disease or abnormality. For example the geometry of features lying on the corpus callosum could potentially serve as robust indicators of schizophrenia as in [37]. Part geometry and occurrence variables could potentially be used improve morphometric analysis [1], by improving intersubject alignment and indicating where intersubject alignment may not be valid. Although many of these possibilities fall outside the scope of this paper, they represent interesting directions for future investigation.

B. Model-to-Subject Registration

The goal of model-to-subject registration is to determine how a new subject relates to its population. The PBM does this by identifying the features in the subject image that are representative of the population, as determined through learning. In this section, we describe model-to-subject registration trials where the model learned from 102 subjects described in Section IV-A is automatically fit to the remaining 50 new test subjects not used in learning, via the algorithm described in Section III-E. No gold standard exists for evaluating the ground truth accuracy of intersubject registration [31] and we, thus, compare the automatic part registration with respect to manual part registration, as established by 3 different human raters. Since there are many model parts, none of which are generally identified in all subjects, we focus on a set of 25 test parts that occur frequently and throughout the brain during model learning. Fitting trials identify a subset of these 25 parts in each test subject, which serves as the basis for fitting evaluation. The number of test parts identified per subject image is 10 on average and ranges from 4 to 16. The number of instances of each test part identified over all 50 trials is 20 on average and ranges from 4 to 47. In total, 516 part instances are considered. Fig. 6 illustrates a set of 4 test subject images which together contain at least one instance of each test model part.

Since model parts are defined via a fully automatic learning procedure, and may not necessarily match obvious anatomical structures, human raters themselves must first be taught the appearances and geometries of the parts before attempting to localize them in new images. This is done by showing the raters images of model part instances identified in different subjects during the model learning process. Specifically, ten images such as those in Fig. 3 of a single model part are shown in a looping video sequence. The rater is asked to watch the videos, and then determine the model part locations in all test subject images within which the part was identified during model fitting. Note that model parts contain a much richer description than simple spatial location, including orientation and scale information in addition to a measure of distinctiveness. These aspects could also be established and verified by human raters, however to do so is difficult and labor-intensive and we, thus, restrict our evaluation to part location in this study. As a measure of fitting quality, we calculate the target registration error (TRE) [20] for each part, between locations identified by the model and human raters (model-to-rater) and between raters (interrater). The TRE generally measures the discrepancy between different locations of a "target" identified by different methods in an image.

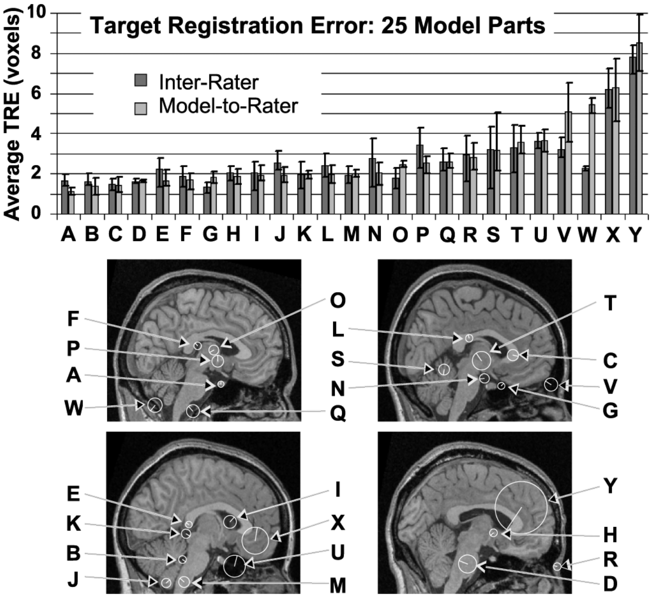


Fig. 6. The interrater and model-to-rater TRE for 25 test model parts, averaged over 50 test subject images and sorted in order of ascending model-to-rater TRE. The height of the bars represents the mean TRE, and error bars indicate the minimum and maximum TRE. The 4 images below illustrate instances of the indicated test part in images. Note that agreement between interrater and model-to-rater TRE indicates that individual model parts can be automatically localized with similar precision to human raters, validating model fitting on a part-by-part basis. Note also that the TRE varies from one part to the next, primarily due to the scale of the part in question, where the larger the part scale, the greater error associated with its localization.

The TRE for each test model part averaged over all 50 test images is illustrated in Fig. 6. Overall, the interrater and model-to-rater TREs are similar, indicating that individual model parts can be automatically fit with similar accuracy to human raters on a part-by-part basis. Localization is more precise for certain parts than for others, for both interrater and model-to-rater cases. This is primarily due to part scale, as large-scale parts such as those arising from cerebral lobes are intrinsically more difficult to localize with precision than small-scale features. For part W, the interrater TRE is somewhat lower than the model-to-rater TRE, indicating some disagreement between human raters and automatic model fitting for this particular model part. Subsequent investigation revealed that model part W had a relatively high intrinsic geometrical variability, as reflected in term $p(m_i^g|o^b, o^g)$, whereas human raters tended to agree as to where they felt the part should occur. Note that model-to-rater agreement could be forced by tightening the geometrical consistency constraint in model learning as mentioned in Section III-D, but as is, the high geometrical uncertainty has already been quantified by $p(m_i^g|o^b, o^g)$ and accounted for in model fitting.

It is also of interest to know the error with which the model can be registered to individual test images. The TRE for each of the 50 test images averaged over identified test model parts is illustrated in Fig. 7. Here, agreement between interrater and model-to-rater TRE indicates that automatic model fitting is similar in accuracy to human raters on a per-image basis. The average interrater and model-to-rater TREs over all images are

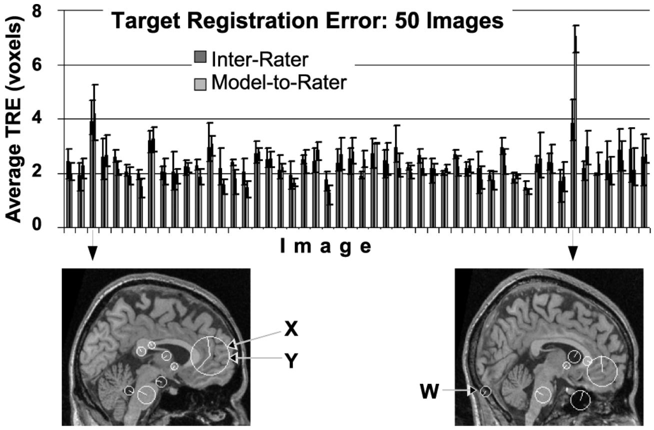


Fig. 7. The interrater and model-to-rater TRE for each test subject image averaged over identified test model parts. The height of the bars represents the mean TRE, and error bars indicate the minimum and maximum TRE. Note that general agreement between interrater and model-to-rater TRE indicates that the model can be automatically fit with similar precision to human raters, validating PBM fitting on an image-by-image basis. The images below illustrate two test subjects for which the TRE is noticeably higher than average, for both interrater and model-to-rater cases. On the left, this is due to the fact that the feature detector has fused parts X and Y into a single region with two dominant orientations. On the right, the part W has been mismatched to a similar-looking yet incorrect feature, in the absence of a feature detected at its correct location just beneath the cerebellum. Note that part W is known to have relatively high intrinsic geometrical uncertainty as quantified by model term $p(m_i^g|o^b, o^g)$, although this uncertainty is not accounted for by the average TRE measure.

similar, 2.3 and 2.4 voxels, respectively. The two images shown below illustrate two test subjects for which the TRE is somewhat higher than average for both interrater and model-to-rater cases, indicating increased fitting difficulty for both man and machine. On the left, this is due to the fact that the feature detector has fused model parts X and Y into a single region with two dominate orientations. On the right, model part W has been mismatched to a similar-looking yet incorrect feature in the absence of a feature detected at the correct location just beneath the cerebellum. As previously mentioned, the high geometrical uncertainty associated with part W is quantified in term $p(m_i^g|o^b, o^g)$ and its influence on fitting accuracy is already discounted by the model. The TRE measure in Fig. 7 does not reflect this uncertainty, however, as the errors of all identified model parts are weighted equally.

C. Parts-Based Modeling Versus Global Modeling

In this section, we compare parts-based modeling to the global modeling approach common in the literature. The PBM describes a set of brain images as a collection of spatially localized, conditionally independent parts. In contrast, global models such as the AAM [12] assume one-to-one correspondence between all subject images, and represent the population terms of global modes of covariance about a mean, in which spatially distinct regions are coupled in linear relationship and are, thus, statistically dependent. Forcing such a global model to fit in locations where correspondence is invalid can adversely affect the entire fit, including locations where valid correspondence exists. We hypothesize that the PBM fitting is, therefore, more robust to unexpected intersubject variability on a local scale than the AAM, as the PBM specifically accounts for such variability and avoids forcing the assumption of one-to-one correspondence in locations where it is invalid. To test this hypothesis, we compare AAM and PBM fitting, where an AAM[2] and a PBM are trained on same set of 102 subjects, and fit to the same independent test set of 50 subjects.

The AAM and the PBM differ significantly in both training and fitting. AAM training requires manually determining a set of point correspondences in all 102 training images, after which a linear Gaussian model over image intensity and point location is estimated. Establishing manual point correspondence is tedious, requires a human to decide which and how many points to use, and is subject to interrater variability. PBM learning is fully automatic and requires no manual intervention, as features are determined by an automatic detector. AAM fitting is an iterative process, starting from an initial guess and occasionally falling into suboptimal local maxima when outside of a particular "capture range." During experimentation, we performed multiple restarts in order to obtain the best possible AAM fitting solutions. In contrast, the PBM fitting produces a robust, globally optimal solution, even in the presence of image translation, rotation and scale change.

Directly comparing AAM and PBM fitting to a new subjects is difficult for several reasons. First, the fitting solution output of the two approaches is fundamentally different: the AAM produces a smooth mapping from model to subject, whereas the PBM identifies the presence and local geometries of a discrete set of modeled parts. Second, there is no gold standard for evaluating the ground truth accuracy of intersubject registration [31]. Third, little guidance exists in selecting a set of manual AAM landmarks leading on an optimal model. We, thus, compare the two models in terms of their stability in the presence of artificial perturbation. To this end, we perform two different sets of fitting trials: the first is based on 50 normal test subjects. The second is based the same subjects with the addition of a localized, artificial intensity perturbation. Model fitting stability can then be evaluated in terms of the per-image TRE averaged over all model locations identified before and after perturbation. For the PBM, locations are based on identified test part locations as described in the previous section, and for the AAM, locations are based on the landmarks defining the model.

The first step in comparing the two models is to establish a baseline in terms of AAM model fitting accuracy. To do this, we construct an AAM based on the set of 102 training subject images, which we then fit automatically to the remaining 50 test subjects. The AAM is defined by 6 manually chosen landmark points as illustrated in Fig. 8. The results of automatic AAM fitting trials are compared to manually-labeled solutions established by a single human rater in terms of the TRE, and illustrated in Fig. 8. Note that in 2 of 50 trials, the TRE is exceptionally high, indicating that the AAM has converged to suboptimal, incorrect solutions. We determine an average TRE threshold of 10 voxels as our definition of a successful/unsuccessful model fitting trial, and exclude all subjects for which the TRE of model fitting is greater than this threshold. Note that as illustrated in Fig. 7, all PBM fitting trials are successful by this definition. The average per-image TRE for successful fitting trials before

[2] The AAM implementation used is publicly available [13].

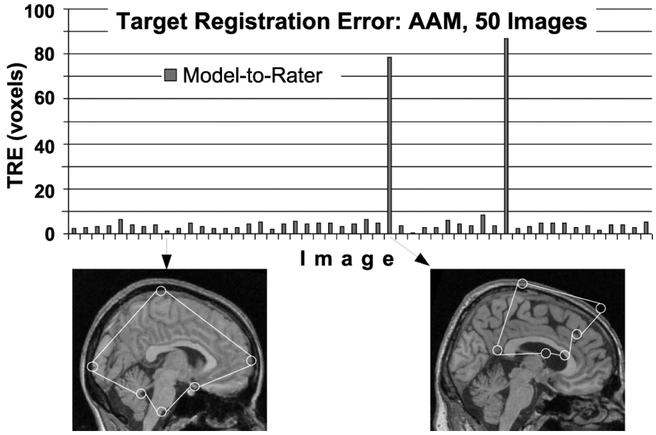


Fig. 8. The model-to-rater TRE of AAM fitting for 50 test subject images, averaged over all landmark locations identified in each image. As illustrated in the lower left image, the AAM created for the purpose of experimentation is defined by 6 manually selected points, 3 lying on distinct subcortical structures and 3 on cortical extremities. The modeled image content lies within the convex hull of the 6 points, and includes both cortical and subcortical structure. The TRE is calculated from the 6 landmark point locations determined both via automatic AAM fitting and manual identification by a single human rater, and the height of the bars represents the average TRE of all 6 points. Note that in 2 of the 50 subjects, the model incorrectly converges to suboptimal incorrect solutions with extraordinarily high TRE. The lower left and right images illustrate examples of successful and unsuccessful AAM fitting trials, respectively, where a successful fitting trial is deemed to be one in which the average TRE is less than 10 voxels. The unsuccessful trials are due to the inability of the AAM to cope with normal, unexpected intersubject variability, such as the large, diagonally-oriented ventricles of the subject in the lower right image.

perturbation is 2.3 voxels for the PBM and 3.8 voxels for the AAM fitting trials.

Having established that both the AAM and the PBM can be successfully fit to a set of 48 subjects, where a successful fit is defined by an average TRE of less than 10 voxels, we can evaluate the stability of fitting in the presence of perturbation. Given that an accurate, stable solution was obtained in fitting to the 48 normal test subjects, how does this solution vary when a localized artificial perturbation is introduced in each of the test subjects? The perturbation we consider consists of a single black circle inserted at random locations in each test subject image, thereby modifying the image appearance locally in a manner reminiscent of a tumor or a resection. The circle is of radius 16 voxels, occupying approximately 2% of the slice of size 217 × 181 voxels. The image intensity at the circle center is 0, with a border blended smoothly into the original image in order to simulate a more natural image boundary. Intuitively, since the perturbation has only affected the image intensity in a local image region, the fitting solution should should not change except perhaps in a neighborhood surrounding the perturbed area. As a measure of model fitting stability, we consider the per-image TRE between fitting solutions obtained before and after the perturbation, which we refer to as the original-to-perturbed TRE.

The original-to-perturbed TRE for both the AAM and the PBM is illustrated in Fig. 9. Note that a large number of AAM fitting trials failed to converge to similar solutions before and after perturbation, as evidenced by exceptionally high original-to-perturbed TRE. As hypothesized, all AAM fitting solutions undergo a global change after perturbation, extending to image regions obviously unaffected by the perturbation, see Fig. 10. This contrasts sharply with the PBM fitting solutions. While the *minimum* original-to-perturbed TRE for the AAM is 2.9 voxels, the *maximum* for the PBM is 0.5 voxels and is barely visible on the graph in Fig. 9, indicating that PBM fitting is stable in the presence of unexpected local variation. As seen in Fig. 10, PBM fitting solutions are virtually identical before and after perturbation, with the exception of fewer matches in a local neighborhood around the perturbation. This is because the perturbation is recognized as new and unmodeled image structure by the PBM, and can be safely ignored. The size of the neighborhood affected by the perturbation is defined both by the scale of the perturbation and the extent of the filter used in feature detection, which in the case of SIFT features is the width of a truncated Gaussian filter.

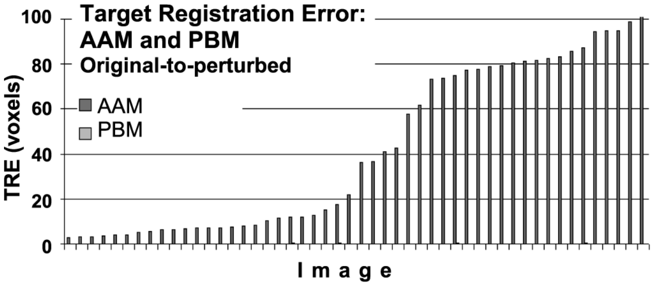
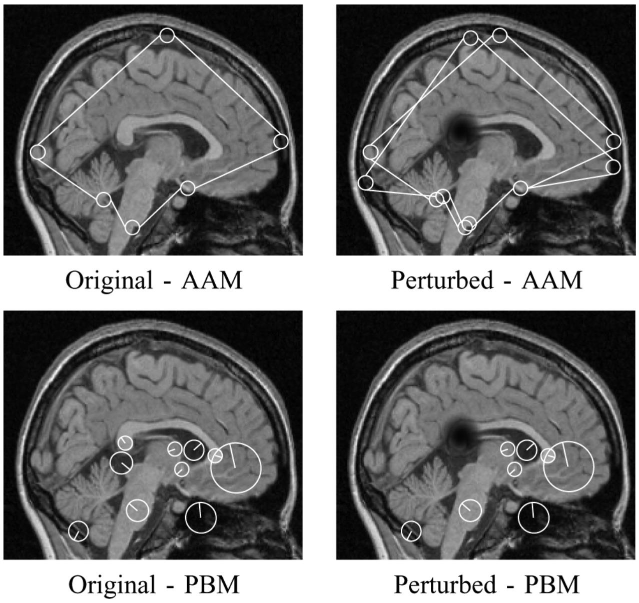


Fig. 9. The original-to-perturbed TRE for AAM and PBM fitting, averaged over AAM landmarks and identified PBM test parts in each of 50 test images, and sorted according to increasing AAM TRE. Note that AAM fitting is generally unstable, as a local image perturbation generally induces a global perturbation in the fitting solution, as evidenced by nonzero AAM TRE. In many cases, AAM solutions become completely invalid, resulting in exceptionally high TRE. While the original-to-perturbed TRE for the AAM ranges from [2.9–110] voxels, the PBM TRE ranges from [0.01–0.5] voxels and is barely visible on the graph.

V. Discussion

In this paper, we presented a statistical parts-based appearance model applied to MR brain imagery. The PBM represents images of a population as a collection of spatially localized image regions, or model parts, each of which consists of an appearance, a geometry and an occurrence frequency. The model specifically addresses the case where one-to-one correspondence between subjects does not exist due to anatomical variability, as model parts are not required to appear in all images. Model parts do not necessarily represent obvious anatomical structures, but rather image patterns which arise from the underlying brain anatomy and occur with statistical regularity within a population. Experimentation shows that the PBM can be automatically learned from a large set of 2-D MR images, in order to identify and statistically quantify common image structure in the form of invariant features. The model can be robustly fit to new images with accuracy comparable to human raters, both on a part-by-part and an image-by-image basis. Additionally, PBM fitting is stable in the presence of unexpected local image perturbation, in contrast to the fitting of global models such as the AAM which is generally unstable as a local perturbation induces a global change in the fitting solution.


Original - AAM

Perturbed - AAM

Original - PBM

Perturbed - PBM

Fig. 10. Illustrating the stability of AAM and PBM fitting in the presence of a local perturbation. The upper left image illustrates the original AAM fit to a new subject. The upper right image illustrates the original and the perturbed AAM fits to the same subject, after the introduction of a local, artificial perturbation in image content akin to the effect of a tumor or a resection. Note that the local perturbation gives rise to a global change in the AAM solution, extending to brain structure unaffected by the change. The two lower images illustrate PBM fitting for the same two images. Note that the PBM fit remains virtually unaffected except in a neighborhood of the perturbation, where two model test parts disappear due to the appearance of the novel, unmodeled image structure. The original-to-perturbed TRE for this test subject is 8.4 voxels for the AAM and 0.03 voxels for the PBM.

In summary, the PBM represents several important advancements with respect to statistical appearance models typically applied to quantifying anatomical brain variability. These are as follows:

1) The model can be constructed via a fully automatic machine learning procedure capable of dealing with a large image set containing significant intersubject variation.
2) The model can be robustly fit to new subjects to obtain a globally optimal solution, in the presence of intersubject variability in addition to global image translation, rotation and scale changes.
3) Model fitting is stable, in the sense that a localized image deformation results in a localized change in the fitting solution.
4) All subjects of a population can be modeled simultaneously without making *a priori* classifications as to which subjects are "normal."
5) The spatially localized model parts identified by the model offer an intuitive means of describing and communicating anatomical variability within a population.

There are many of future directions open to parts-based modeling of brain anatomy. Subjects could be grouped according to their model parts, in order to identify subpopulations sharing similar anatomical characteristics. In addition, parts shared by different subjects could serve as a basis for robust intersubject registration, describing where meaningful intersubject correspondence can be expected, based on the learned model [36]. Different types of invariant features incorporated into the model could prove useful for modeling different aspects of brain anatomy. Although the current line of experimentation revealed the presence of stable parts primarily in the subcortical region, training on larger data sets could potentially reveal modes of appearance within the cortex.

Due to the generality of the model, it should prove useful in a wide variety of other medical imaging domains for the study of anatomical structure within a population, an important area of computational anatomy. The experimentation presented in this paper was based on 2-D imagery, as the primary goals were to validate the PBM and to contrast parts-based and global modeling, which was greatly facilitated using publicly available 2-D implementations of invariant feature detectors and AAMs. The results obtained are encouraging and will naturally lead to the development of robust scale-invariant feature detectors for modeling images in higher dimensions, for example 3-D volumetric and 4-D temporal data. Such feature detectors will employ an image pyramid-based peak detection methodology similar to 2-D implementations in order to determine feature location and scale, and 2 or 3-D orientation histograms to determine feature orientation. We are currently pursuing parts-based modeling on large sets of full 3-D MR volumes, in order to identify commonalities in upper-cortical regions of the brain.

Acknowledgment

The authors would like to thank Dr. D. L. Collins of the McConnell Brain Imaging Centre, Montreal Neurological Institute, for his helpful comments and insight regarding the experimental portion of this paper and for providing the images used throughout.

References

[1] J. Ashburner and K. J. Frison, "Voxel-based morphometry-the methods," *NeuroImage*, vol. 11, no. 23, pp. 805–821, 2000.

[2] R. Bajcsy and S. Kovacic, "Multiresolution elastic matching," *Comput. Vis., Graphic. Image Process.*, vol. 46, pp. 1–21, 1989.

[3] E. Bart, E. Byvatov, and S. Ullman, "View-invariant recognition using corresponding object fragments," in *Proc. ECCV*, 2004, pp. 152–165.

[4] R. Beichel, H. Bischof, F. Leberl, and M. Sonka, "Robust active appearance models and their application to medical image analysis," *IEEE Trans. Med. Imag.*, vol. 24, no. 9, pp. 1151–1169, Sep. 2005.

[5] I. Beiderman, "Recognition-by-components: a theory of human image understanding," *Psychological Rev.*, vol. 94, no. 2, pp. 115–147, 1987.

[6] F. L. Bookstein, "Thin-plate splines and the atlas problem for biomedical images," *Inf. Process. Med. Imag.*, pp. 326–342, 1991.

[7] M. Bro-Nielsen and C. Gramkow, "Fast fluid registration of medical images," *Visualization Biomed. Computing*, pp. 267–276, 1996.

[8] G. Carneiro and A. D. Jepson, "Multi-scale phase-based local features," in *Proc. CVPR*, 2003, vol. 1, pp. 736–743.

[9] D. L. Collins, N. J. Kabani, and A. C. Evans, "Automatic volume estimation of gross cerebral structures," in *Proc. 4th Int. Conf. Functional Mapping of the Human Brain*, A. Evans, Ed., 1998.

[10] D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, "Automatic 3D inter-subject registration of mr volumetric data in standardized talairach space," *J. Comput. Assisted Tomogr.*, vol. 18, no. 2, pp. 192–205, 1994.

[11] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, 2002.

[12] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–684, Jun. 2001.

[13] T. Cootes, AM Tools [Online]. Available: http://www.isbe.man.ac.uk/bim/software

[14] G. Dorko and C. Schmid, "Selection of scale-invariant parts for object class recognition," in *Proc. ICCV*, 2003, pp. 634–640.

[15] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*. New York: Wiley, 1998.

[16] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[17] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proc. CVPR*, 2003, pp. 264–271.

[18] J. C. Gee, D. R. Haynor, M. Reikvich, and R. Bajcsy, "Finite element approach to warping of brain images," *Proc. SPIE Medical Imaging 1994: Image Processing*, vol. 2167, pp. 18–27, 1994.

[19] U. Grenander and M. I. Miller, "Computational anatomy: an emerging discipline," *Quart. Appl. Math.*, vol. LVI, no. 4, pp. 617–693, 1998.

[20] J. V. Hajnal, D. L. G. Hill, and D. J. Hawkes, *Medical Image Registration*. Boca Raton, FL: CRC Press, 2003.

[21] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vision Conf.*, 1988, pp. 147–151.

[22] P. Hellier, C. Barillot, I. Corouge, B. Gibaud, G. Le Goualher, D. L. Collins, A. Evans, G. Malandain, N. Ayache, G. E. Christensen, and H. J. Johnson, "Retrospective evaluation of intersubject brain registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1120–1130, Sep. 2003.

[23] S. Joshi, B. David, M. Jomier, and G. Gerig, "Unbiased diffeomorphic atlas construction for computational anatomy," *NeuroImage*, vol. LVI, no. 23, pp. 151–160, 2004.

[24] T. Kadir and M. Brady, "Saliency, scale and image description," *Int. J. Comput. Vis.*, vol. 45, no. 2, pp. 83–105, 2001.

[25] D. Lowe, Demo Software: Sift Keypoint Detector [Online]. Available: http://www.cs.ubc.ca/lowe/keypoints/

[26] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[27] L. M. Lui, Y. Wang, T. F. Chan, and P. M. Thompson, "Automatic landmark tracking applied to optimize brain conformal mapping," presented at the Int. Symp. Biomedical Imaging, Washington, D.C., 2006.

[28] K. Mikolajczk and C. Schmid, "A performance evaluation of local descriptors," *Comput. Vis. Pattern Recognit.*, vol. 2, pp. 257–263, 2003.

[29] ——, "Scale and affine invariant interest point detectors," *Int. J. Comput. Vis.*, vol. 60, no. 1, pp. 63–86, 2004.

[30] S. M. Pizer, P. T. Fletcher, S. Joshi, A. Thall, J. Z. Chen, Y. Fridman, D. S. Fritsch, A. Graham Gash, J. M. Glotzer, M. Jiroutek, C. Lu, K. E. Muller, G. Tracton, P. Yushkevich, and E. L. Chaney, "Deformable m-reps for 3D medical image segmentation," *Int. J. Comput. Vis.*, vol. 55, no. 2–3, pp. 85–106, 2003.

[31] D. Reuckert, *Nonrigid Registration: Concepts, Algorithms and Applications*. Boca Raton, FL: CRC Press, 2003, ch. 13, pp. 281–301.

[32] D. Rueckert, A. F. Frangi, and J. A. Schnabel, "Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 1014–1025, Aug. 2003.

[33] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 19, no. 5, pp. 530–535, 1997.

[34] J. Talairach and P. Tournoux, *Co-Planar Stereotactic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging*. Stuttgart, Germany: Georg Thieme Verlag, 1988.

[35] M. Toews and T. Arbel, "Detection over viewpoint via the object class invariant," in *Proc. ICPR*, 2006, vol. 1, pp. 765–768.

[36] M. Toews, D. Louis Collins, and T. Arbel, "A statistical parts-based appearance model of inter-subject variability," in *Proc. MICCAI*, 2006, vol. I, pp. 232–240.

[37] A. W. Toga, P. M. Thompson, M. S. Mega, K. L. Narr, and R. E. Blanton, "Probabilistic approaches for atlasing normal and disease-specific brain variability," *Anat. Embryol.*, vol. 204, pp. 267–282, 2001.

[38] M. Turk and A. P. Pentland, "Eigenfaces for recognition," *CogNeuro*, vol. 3, no. 1, pp. 71–96, 1991.

[39] C. J. Twining, T. Cootes, S. Marsland, V. Petrovic, R. Schestowitz, and C. J. Taylor, "A unified information-theoretic approach to groupwise non-rigid registration and model building," in *Proc. IPMI*, 2005, pp. 1–14.