Determining Correspondence in 3D MR Brain Images Using Attribute Vectors as Morphological Signatures of Voxels

Zhong Xue, Member, IEEE, Dinggang Shen, Member, IEEE, Christos Davatzikos, Member, IEEE

Abstract—Finding point correspondence in anatomical images is a key step in shape analysis and deformable registration. This paper proposes an automatic correspondence detection algorithm for intramodality MR brain images of different subjects using wavelet-based attribute vectors defined on every image voxel. The attribute vector is extracted from the wavelet subimages and reflects the image structure in a large neighborhood around the respective voxel in a multi-scale fashion. It plays the role of a morphological signature for each voxel, and our goal is therefore to make it distinctive of the respective voxel. Correspondence is then determined from similarities of attribute vectors. By incorporating the prior knowledge of the spatial relationship among voxels, the ability of the proposed algorithm to find anatomical correspondence is further improved. Experiments with MR images of human brains show that the algorithm performs similarly to experts, even for complex cortical structures.

Index Terms—Correspondence, deformable registration, image matching, wavelet transformations, computational anatomy.

I. INTRODUCTION

DEFORMABLE registration of MR brain images is the process of finding a 3D transformation that maps one individual brain image to another [1-45]. It is used frequently for anatomical segmentation and labeling, for morphological analysis using shape transformations, and for spatial normalization of structural and functional data. Although a large portion of the related literature focuses on models of template deformation based either on statistical models or on physical principles, image intensities or other low-level attributes (e.g., edge information) drive the registration most frequently. In this paper, we devote our effort to design richer attributes for determining voxel correspondences between same modality MR brain images of different subjects, which could be subsequently combined with various deformation models for deformable registration.

According to the features used, deformable registration can be classified into two groups: image similarity-based methods and feature-based methods. In image similarity-based methods [7-29], the deformation from a subject image to a template image is determined by evaluating the degree of similarity between the template and the warped subject images, under constraints imposed by the model of the transformation. Various methods have been used to model the 3D transformation, including elastic, biomechanical, fluid and parametric approaches. Besides image intensity-based methods, other image similarity approaches include maximizing mutual information [9], [10], [22], [20], or matching local frequency representations [28].

Image similarity-based methods use local information, e.g., intensity of each voxel, as image features, which are not rich enough to distinguish a voxel from others. They can therefore be sensitive to initialization and particularly vulnerable to local minima. On the other hand, feature-based methods are less susceptible to these shortcomings. In feature-based methods [30-43], anatomical features such as surfaces, landmark points, or ridges are first detected in two images. Then a 3D spatial transformation is determined so that corresponding features are mapped to each other. Correspondence elsewhere is interpolated using the model of the spatial transformation [30], [32], [33], [35], [37], [41].

Although feature-based methods can rely on manual definition of point or feature correspondence, routine clinical use requires effective automated methods for finding correspondences that can subsequently guide a 3D transformation. Even if fully automated feature segmentation and correspondence detection has been achieved [36], [38], [46], it is typically confined to only a small subset of voxels, namely the feature points. In this paper, an automatic correspondence detection algorithm is proposed for intramodality MR brain images of different subjects using the Wavelet-based Attribute Vectors (WAV) defined on every image voxel. We take the notion of a feature to its extreme: every voxel is regarded as a feature point. An Attribute Vector (AV) that is defined for each voxel serves as its morphological signature [44], [47], [48], and is used to achieve accurate and automated feature identification and matching.

In the literature, methods for voxel correspondence detection often rely on extracting image features such as intensities and gradients [49], moments [44], [50], Gabor features [51], and local frequency representations [28], [52]. In the moment-based feature extraction methods, different kinds of moments of the image data can be used. These moment-based features are generally invariant under rigid transformations. In order to capture the anatomical features of a voxel, the moments of the subimage data within a sliding spherical shell centered on the voxel are calculated and used to form the moment-based AV. However, the moment-based AVs are not compact...
and efficient enough to capture all anatomical subtleties of the image structures around voxels. The reason is that the moments reflect only global features of the subimages within the neighborhoods of voxels, and they lose the spatial information contained in those subimages. To incorporate more anatomical information and make moment-based AV more distinctive, in [44], brain images were first segmented and labeled into white matter, gray matter and CSF, then the AV of a voxel was defined by calculating the rotationally invariant moment features on different tissues [50], [53]. Although experimental results indicated that this algorithm performed well for segmented images, the requirement of pre-segmentation limits its generality. Unlike the moment-based AV, a new wavelet-based AV is designed in this paper to efficiently capture the anatomical features of each voxel from the original MR brain images.

In this study, we focus on design and implementation of a distinctive and robust AV. Generally, good AVs should possess three properties: i) Translation and rotation-invariance, i.e., an AV should not be sensitive to the position and orientation of objects. ii) Uniqueness, i.e., the AV of a voxel should contain enough information to be different from all the other AVs within the image. iii) Robustness to noise, intensity distortion, morphological variability, and partial volume effects, i.e., ideally, corresponding voxels in images of different individuals should have AVs similar to each other and different from the AVs of non-corresponding voxels. Under these ideal conditions, voxel correspondences among different images can be determined directly by examining AV similarities. In practice, we seek representations that make AVs as distinctive as possible.

In our method, the voxel-wise WAVs are designed for correspondence detection in a multi-resolution framework. To construct the WAVs, the Discrete Wavelet Transformation (DWT) of the image data within a sliding window centered on the voxel of interest is performed using the length-2 Daubechies wavelet basis, yielding a series of wavelet subimages that represent different aspects of the original MR brain image in a scale-space fashion. In order to achieve rotation-invariance, we then construct the Wavelet-based AV using a method that maintains spatial information along the radial direction in the wavelet subimages and extracts only statistical image information in the angular directions. The information extracted in WAV contains not only the global information but also the progressively higher frequency and spatial localization information around a voxel, which reflects the geometric information about the underlying anatomical structures. The WAV gives detailed representation of the anatomical features in the vicinity of that voxel, and acts as the voxel’s morphological signature. It is also intended to be rich enough to distinguish among different parts of the anatomy. Therefore, compared with the moment-based AVs, the wavelet-based AVs are more compact and efficient in capturing the anatomical features of the image structures around voxels.

An automatic correspondence detection algorithm for 3D MR brain images is achieved based on WAV matching. Using this algorithm, given a set of voxels in one image, their corresponding voxels are found automatically in another image, and these images are of the same modalities and are obtained using similar acquisition protocols. Three WAV matching methods are proposed and examined in this paper: i) Matching based on similarities of WAVs, i.e., given a voxel from one image, its corresponding voxel in another image is determined by finding the voxel that has the most similar WAV. ii) Matching based on statistical models of WAVs of corresponding voxels. This strategy is based on the Principal Component Analysis (PCA) models of the WAVs of corresponding voxels, and the corresponding voxel in an input image is determined by finding the voxel, whose WAV has the highest pdf value. iii) Matching by incorporating shape prior constraints. This method applies prior shape constraints among a number of voxels/landmarks in order to determine their correspondences among images.

We validate our matching algorithms on human brain data by comparing the correspondences with manually derived ones provided by experts. The similarities of the WAVs of the voxels selected from different sulci are evaluated using a statistical analysis method, and the results reveal a very high specificity of the proposed WAV. Moreover, the three AV matching methods are compared by finding correspondences of a number of landmarks and evaluating the correspondence difference between the detection results of our algorithm and those manually marked by experts. Experimental results indicate that the proposed algorithm performs similarly to experienced raters in defining pairs of anatomically corresponding voxels, while being fully automated.

The remainder of the paper is organized as follows: Section 2 introduces the proposed methods in detail, including WAV design and WAV matching for correspondence detection. Experimental results are presented in Section 3, and Section 4 concludes the results from this study.

II. METHOD

We approach automatic correspondence detection as a procedure that requires the following three steps: i) design of voxel-wise attribute vectors that capture the image structure around each voxel, ii) calculation of the similarities between a voxel’s AV and the AVs of candidate matching voxels, and iii) selection of the optimal matching voxel based on AV similarity, and in certain cases, also based on prior statistical knowledge about shape variability. A new automatic correspondence detection algorithm is proposed, which consists of the above three steps. It emphasizes the design and implementation of the wavelet-based attribute vector so that each voxel has a distinctive WAV serving as its morphological signature.

A. Design of Wavelet-Based Attribute Vectors

Although AV design may depend on the problem at hand, a good AV possesses the property of rotation-invariance, and should be made as distinctive as possible so that one voxel can be distinguished from other voxels in the same or different images. In this section, following the introduction of a general approach to extract the rotation-invariant attribute vector, the procedure for constructing WAV is described in detail. Moreover, compared with the moment-based correspondence detection, the advantages of the wavelet-based AV are demonstrated experimentally.
1) Extracting Rotation-Invariant Attributes — Radial Profiling: To construct the translation and rotation-invariant attribute vector of a voxel, we would ideally need to project the image within the sliding window centered on that voxel onto a multi-resolution and rotationally invariant basis. Since no rotationally invariant complete basis has been constructed in 3D, we achieve rotation-invariance by first calculating the multi-resolution feature subimage, and then maintaining only the statistical image information in the angular directions, along with the full spatial information in the radial direction. In this way, the attribute vector obtained is rotation-invariant to the feature subimage. Moreover, it is rotation-invariant to the input image, provided the feature subimage always follows the rotation of that input image.

Specifically, for an input image \( I \), a general approach to construct the rotation-invariant attribute vector of a given voxel \( x_0 \) includes the following three steps:

\[
\begin{align*}
\text{image } I, \text{ voxel } x_0 & \rightarrow \text{ subimage } I_{x_0}(x) \\
& \rightarrow \text{ feature subimage } L_{x_0}(x) \\
& \rightarrow \text{ pdf } h_{\Delta r,n}(l).
\end{align*}
\]

\( I_{x_0}(x) \) is the subimage within a sliding window centered on \( x_0 \), and \( L_{x_0}(x) \) is the feature subimage calculated from \( I_{x_0}(x) \), which includes image features organized in a multi-resolution way as described in the next section. \( h_{\Delta r,n}(l) \) is the probability density function (pdf) of the feature subimage’s intensity \( f \) within a spherical shell with inner radius \( r = n \Delta r \) and outer radius \( r = (n+1)\Delta r \), where \( n \) is the index and \( \Delta r \) is the thickness of a spherical shell. \( h_{\Delta r,n}(l) \) is calculated from the feature subimage \( L_{x_0}(x) \) as follows: First, using the spherical coordinate system, \( L_{x_0}(x) \) can be denoted as \( L_{\text{spherical}}^h(r, \theta, \phi) \) (with \( x_0 \) as the origin), where \( r \) is the radius, \( \theta \) is the azimuthal angle and \( \phi \) is the polar angle respectively; Then, if we treat the intensities of \( L_{\text{spherical}}^h(r, \theta, \phi) \) as realizations of a random variable \( l \), the respective pdf \( h_{\Delta r,n}(l) \) for \( \theta \in [0, 2\pi), \phi \in [0, \pi] \), and \( r \in [n \Delta r, (n+1)\Delta r) \) can act as a rotation-invariant attribute vector of voxel \( x_0 \). \( N_r \) is the number of the spherical shells. In fact, \( h_{\Delta r,n}(l) \) can be estimated by calculating the histograms of \( L_{\text{spherical}}^h(r, \theta, \phi) \) within the \( N_r \) spherical shells respectively. In summary, step 1 makes the attribute vector invariant to translations; step 2 calculates the multi-resolution feature subimage; and step 3 extracts the rotationally invariant features, by discarding spatial information along the angular directions.

In order to derive a computationally tractable attribute vector, we represent \( h_{\Delta r,n}(l) \) by its moments, which reflect its most important characteristics. Specifically, in this paper, we propose a method called Radial Profiling (RP), which only calculates the low-order moments of \( h_{\Delta r,n}(l) \), i.e. the mean and variance:

\[
\begin{align*}
u(n) &= \int_{L_{\min}}^{L_{\max}} l \cdot h_{\Delta r,n}(l) \, dl, \\
w(n) &= \int_{L_{\min}}^{L_{\max}} (l - \mu(n))^2 \cdot h_{\Delta r,n}(l) \, dl.
\end{align*}
\]

where \([L_{\min}, L_{\max}]\) is the range of the intensity values of the feature subimage \( L_{x_0} \). \( \mu(n) \) and \( w(n) \) are actually the mean and variance of all the feature values within the \( n \)th spherical shell, with the internal radius being \( n \Delta r \). Then, a rotation-invariant attribute vector at voxel \( x_0 \) is given by

\[
\mathbf{v} = [\mu(0), \mu(1), w(1), \ldots, \mu(N_r - 1), w(N_r - 1)]^T.
\]

\( n = 0 \) refers to the innermost shell and \( n = N_r - 1 \) refers to the outermost shell (see Fig.1). Thus \( N_r \Delta r \) is the radius of the spherical neighborhood. This RP method will be used to construct the wavelet-based attribute vector in the following section.

2) Wavelet-Based Attribute Vector: In this section, we describe how to design and implement the WAV.

Design of WAV: Since the AV of a voxel is designed to reflect the image structure in the vicinity of that voxel, it should be calculated from the image data within a neighborhood of the voxel. Moreover, the size of the neighborhood should be large enough so that adequate anatomical context around the voxel of interest is used in the calculation of the AV. For example, if the neighborhood were one voxel big, the AV would only include the image intensity of the respective voxel, and it would not reflect any anatomical information other than the tissue type at that voxel. Based on experience, we set the volume of the neighborhood to about 10-15% of the brain volume [44]. For a given neighborhood, the main challenge in designing the AV at a voxel is to find those parameters that represent the corresponding anatomy in a compact way.

In this paper, we use the wavelet decomposition to extract the multi-resolution features of the image data within that neighborhood. The wavelet decomposition is a tool for multi-resolution and scale-space analysis, and it is widely used in the area of signal/image processing, compression and recognition. In 1989, Mallat et. al. proposed the DWT for extracting features from images [54], which represents not only global (low resolution) but also local (high resolution) image information. In this paper the length-2 Daubechies wavelet is applied, which has orthogonal and symmetric basis functions. It has compact support and at least one vanishing moment. The low-pass and high-pass filters are \([1/\sqrt{2}, 1/\sqrt{2}]\) and \([1/\sqrt{2}, -1/\sqrt{2}]\) respectively.
To calculate the WAV of a voxel $x_0$ under consideration, the DWT decomposition of the image data within a sliding window centered on $x_0$, $I_{x_0}(x)$, is performed. The DWT generates 8 component subimages, including one low-pass subimage $I_{LL}^{(1)}$, and seven high-pass subimages $I_{LH}^{(1)}, I_{HL}^{(1)}, I_{HH}^{(1)}, I_{LLH}^{(1)}, I_{LHL}^{(1)}, I_{HLL}^{(1)}$, etc. The three-letter subscript represents which filter is used along the three coordinate axes $x$, $y$, and $z$ respectively. $L$ represents that the low-pass wavelet filter is used in the corresponding coordinate axis, and $H$ indicates that the high-pass wavelet filter is used in that axis. The low-pass subimage, $I_{LL}^{(1)}$, can be further decomposed, resulting in a second-level DWT decomposition. This procedure continues until the $J$th level DWT decomposition is performed ($J$ is a predefined number). To construct the WAV of a voxel at each DWT level $j$, selected high-pass subimages are combined to form a feature subimage $L_{x_0}^{(j)}(x)$ as will be detailed below. Then the RP method described in Section 2.1.1 is applied to $L_{x_0}^{(j)}(x)$, ($j = 1, 2, \ldots, J$) respectively, resulting in $J$ wavelet feature vectors each formed according to Eq.(3). In addition, the feature vector of the low-pass subimage $I_{LL}^{(J)}$ at DWT level $J$ is also calculated using RP. The WAV is then constructed by concatenating all these $J + 1$ feature vectors.

Since the wavelet decomposition is performed on the image data inside the sliding window centered on the voxel of interest, the DWT subimages are invariant to translations that are integral numbers of voxels. However, the individual DWT subimage is not rotationally invariant, i.e.

$$DWT(R(I_{x_0})) \neq R(DWT(I_{x_0})).$$

where $DWT(I_{x_0})$ represents one of the DWT component subimages of the input subimage $I_{x_0}$, and $R(\cdot)$ is a rotation operator.

Therefore, if a WAV were to be extracted from the individual wavelet high-pass subimage using the RP method, it would not be rotation-invariant: in fact it is very sensitive to rotation. To make the WAV insensitive to rotation, we combine three high-pass subimages into one feature subimage at each level (similar to that in [55]), and calculate WAV from this combined feature subimage. The feature subimage $L_{x_0}^{(j)}(x)$ at level $j$ is calculated as

$$L_{x_0}^{(j)}(x) = \sqrt{|I_{LH}^{(j)}(x)|^2 + |I_{HL}^{(j)}(x)|^2 + |I_{HH}^{(j)}(x)|^2},$$

where $I_{LH}^{(j)}(x)$, $I_{HL}^{(j)}(x)$ and $I_{HH}^{(j)}(x)$ are the selected high-pass subimages at DWT level $j$. These three subimages reflect most high-pass information of the object along the three different coordinates of the 3D space, and a combination of them reflects the high-frequency information of the original image structure well. Therefore, although theoretically $L_{x_0}^{(j)}(x)$ is not rotation-invariant, the experimental results show that the method is effective as long as the rotation is relatively small (less than 25 degrees) since it combines information about high-pass subimages in all three coordinate axes.

Using the RP method, the attribute vector $v_j$ at the $j$th DWT level is calculated from $L_{x_0}^{(j)}(x)$, where $x_0$ is the center of $L_{x_0}^{(j)}(x)$. Combining all the $J$-level attribute vectors $v_1, v_2, \ldots, v_J$, the WAV of voxel $x_0$, denoted by $v_w$, is expressed as

$$v_w = [\lambda_1 v_1^T, \lambda_2 v_2^T, \ldots, \lambda_J v_J^T, \lambda_D v_D^T]^T,$$

where $v_D$ is calculated from $I_{LL}^{(J)}$ using the RP method, $\lambda_1, \ldots, \lambda_J$ and $\lambda_D$ are weighting coefficients to be discussed later. This procedure for constructing the WAV is illustrated in the two-dimensional case in Fig.2.

**Implementation of WAV — A Computationally Efficient Approximation:** Unfortunately, the computational load necessary to calculate the ideal WAV presented in the previous section for every voxel of an image is very heavy: for each voxel, the $J$-level DWT decompositions should be applied to the image data within the sliding window, where the volume of the window is set to about 10-15% of the brain volume. To solve this problem, a multi-resolution framework is proposed to calculate WAV, which not only acts as an approximation of the original algorithm, but also has the effect of feature reduction and selection.

Based on two facts, this procedure is a computationally efficient calculation of certain wavelet expansion parameters that we need to include in the AV: i) Relatively global anatomical characteristics must be determined from a larger window size...
around the voxel of interest. Because these characteristics are global, in their estimation one does not need to consider all voxels in this large neighborhood, but only a small fraction of them. Our premise is that global characteristics calculated from a small set of downsampled voxels in a large neighborhood will be very similar to those calculated from all voxels in the same neighborhood. Therefore, computational efficiency does not compromise the calculation of the WAV. ii) High-frequency local characteristics only need to be calculated from a small neighborhood around the voxel: high frequency parameters in a large neighborhood would be less relevant to the voxel of interest, since they pertain to local shape information to be included in other voxels’ WAVs.

In this framework, the multi-resolution images are generated by first using Gaussian filtering and down-sampling. In our experiments, the standard deviation of Gaussian function is set to 0.4mm, and 1/2 down-sampling is used. This standard deviation of Gaussian function is chosen so that the low-resolution images are not so blurry and still retain much high frequency information. From experimental results we found that 0.4mm is the best compromise over a variety of our brain images. Then multi-level DWTs are performed at each resolution level, and the feature vectors are calculated using the RP method. Finally, the WAV of a voxel is constructed by combining all the feature vectors. This procedure is illustrated in Fig.3 and is described as follows:

- Re-sample the input image so that it has isotropic voxel size. Spline-based interpolation methods that introduce minimal blurring to the images are used to calculate the new intensities of isotropic voxels. Then, normalize the intensities of the input image (with maximal value as 255 and minimal value as 0) and calculate the multi-resolution images using Gaussian filtering and down-sampling:

\[ I_k(\mathbf{x}), \quad 0 \leq k \leq K, \]

where \( I_0 \) represents the input image after normalizing the intensities \( (I_0 = I) \), and \( I_K \) is the image at the lowest resolution. The down-sampling rate, \( R \), is 2. If the current voxel in the finest resolution image is \( I_0 = x_0 \), then this voxel in image \( I_k \) is \( P_k = x_0/R^k \).

- For each image \( I_k \) at resolution level \( k \), set the center of the sliding window to the current voxel \( P_k \). Then perform \( J_k \)-level DWT decompositions of the image data inside the sliding window.

- Calculate \( \mathbf{v}_{\mathbf{w}}^{(k)} \), the WAV of the current voxel \( P_k \) at resolution level \( k \), from the \( J_k \)-level DWT subimages using Eq.(6). Fig.3 describes this procedure, where 1-level DWT decomposition of the image data inside the sliding window is utilized.

- For each voxel \( x_0 \) in the input image \( I \), its WAV can be formulated as a concatenation of all the WAVs in different resolutions, i.e., \( \mathbf{v}_{\mathbf{w}}^{(0)}, \mathbf{v}_{\mathbf{w}}^{(1)}, \mathbf{v}_{\mathbf{w}}^{(2)}, \ldots, \mathbf{v}_{\mathbf{w}}^{(K)} \). Before the concatenation, the WAVs in the lowest resolution level need to be interpolated or upsampled into the highest resolution level. In our work, we use linear interpolation of WAVs to eliminate blockiness in the similarity map. Therefore, the WAVs \( \mathbf{v}_{\mathbf{w}}^{(k)} (k = 1, \ldots, K) \) are first interpolated into the highest resolution level using linear interpolation, where they are denoted by \( \mathbf{v}_{\mathbf{w}}^{(k)} \). Then the WAV of voxel \( x_0 \), \( \mathbf{v}_{\mathbf{w}} \), can be written as

\[ \mathbf{v}_{\mathbf{w}} = [\lambda_0^{(0)} \mathbf{v}_{\mathbf{w}}^{(0)T}, \lambda_1^{(1)} \mathbf{v}_{\mathbf{w}}^{(1)T}, \ldots, \lambda_K^{(K)} \mathbf{v}_{\mathbf{w}}^{(K)T}]^T. \]  

where \( \lambda_0, \ldots, \lambda_K \) are the weighting coefficients. In Section 3.2, we will describe how to select these parameters.

3) Comparison Between Wavelet-based AV and Moment-based AV: Compared to the moment-based attribute vector that we have previously examined in [44], the WAV is more compact and efficient to capture the image structure around a voxel since it is extracted from the multi-resolution wavelet
Fig. 4. Comparison results of finding voxel correspondences using wavelet-based AVs and moment-based AVs. (a) $P$ is the voxel selected from image one; (b) the correspondence detection result using similarity of WAV; $P'$ is the corresponding voxel found from image two by selecting the voxel with the most similar attribute vector; (c) correspondence detection result using similarity of moment-based AV; (d) the similarity map of WAVs, which shows the similarities between the WAV of $P$ in image one and the WAVs of all the voxels in image two; (e) the similarity map of moment-based AVs. The colorbar illustrates that red (close to 1) means there is high similarity between two AVs, while blue (close to 0) represents a lower similarity. This figure is representative of the relatively more distinctive nature of the WAV.

For comparison purposes, we use the same three resolutions, four attributes for each resolution, and the same radius for spherical neighborhoods for the two kinds of AVs. i.e., the radii are 7.5mm for the high resolution, 15mm for the middle resolution, and 30mm for the low resolution images. For WAV, to calculate the four attributes at each resolution, we first perform 1-level DWT on the image data within the neighborhood of the voxel, then use one spherical shell for calculating RP on the wavelet subimages, while for moment-based AV, we calculate one zero-order and three second-order rotation-invariant moment features [53] of the image data within the spherical neighborhoods.

After the AVs are calculated, the similarity of two attribute vectors is calculated using Eq.(11). Given a voxel in one image, its corresponding voxel in another image is then determined by finding the voxel that has the most similar AV. Fig.4 presents the results of correspondence detection using wavelet and moment-based AVs. Fig.4(a) shows the voxel $P$ selected from image one. Fig.4(b) and Fig.4(c) show the corresponding voxels of $P$ in image two, $P'$, determined using similarity of wavelet and moment attribute vectors, respectively. Note $P'$ in Fig.4(b) and Fig.4(c) are different since different AVs are used.

The similarity maps of wavelet AVs and moment AVs are shown in Fig.4(d) and Fig.4(e), which give the similarities between the AV of $P$ and the AVs of all the voxels in image two. In the similarity maps, blue means a low similarity between two AVs (close to 0), while red or brown represents a high similarity (close to 1). The white crosses indicate the positions of $P'$ in the similarity maps. It can be seen clearly from Fig.4(d) that there is only one brown peak. However, the peak in Fig.4(e) is not obvious. This experiment indicates that wavelet-based AVs are more efficient and distinctive than
moment-based ones under similar conditions.

B. Correspondence Detection Using Statistical Models of WAVs

Having designed the attribute vectors, in this section we describe how to determine the voxel correspondence among MRI images using attribute vectors.

1) The Forward Matching Problem: Because of their convenient properties and easiness for mathematical analysis, Gaussian distributions are frequently adopted to model random variables for practical purposes. In this paper, we treat the variability of the WAVs of corresponding voxels, which may be influenced by many uncertain small effects, as normally distributed. Specifically, the set \( \Omega_v(x) \) of the AVs of the corresponding voxels at a particular location \( x \) in a standard space is approximated by a single Gaussian distribution with mean \( \bar{v} \) and covariance matrix \( \Sigma_v \).

Therefore, the pdf value of an AV, \( v(y) \), at location \( y \) in an individual image belonging to \( \Omega_v(x) \) is given by

\[
f(v(y)|\Omega_v(x)) = \eta \cdot \exp\left\{ -\frac{1}{2} E_v \right\}, \tag{9}\]

where \( \eta \) is a normalization coefficient, and \( E_v = (v(y) - \bar{v})^T \Sigma_v^{-1} (v(y) - \bar{v}) \) is the Mahalanobis distance between \( v(y) \) and \( \bar{v} \). Then, in the input image, the corresponding voxel \( y^* \) of \( \Omega_v(x) \) is determined by maximizing Eq.(9),

\[
y^*(x) = \arg \max_{y \in \Omega_v(x)} \left\{ f(v(y)|\Omega_v(x)) \right\}, \tag{10}\]

where \( N_s(x) \) denotes the searching area in the subject image.

When there are no training samples available, correspondence can be determined by maximizing the similarity between two attribute vectors:

\[
f(v(y)|v_m(x)) = \exp\left\{ -\frac{||v(y) - v_m(x)||^2}{\epsilon} \right\}, \tag{11}\]

where \( v_m(x) \) is the attribute vector of the model (template) image at location \( x \), and \( \epsilon \) is a parameter that would ideally depend on the expected range of \( ||v - v_m||\).

When training samples are available, the mean \( \bar{v} \) and covariance matrix \( \Sigma_v \) can be estimated from those samples, and Eq.(9) is used to calculate the pdf value. However, since generally the number of the training samples is small, \( \Sigma_v \) becomes a highly singular matrix, thus not allowing us to determine \( f(v|\Omega_v) \). To solve this problem, the Karhunen-Loeve transform [56] is first used to decouple the calculation of \( E_v \) into a weighted sum of uncorrelated components through calculating the eigenvectors and eigenvalues of \( \Sigma_v \). Then, the PCA [57] is used to estimate an approximation of \( E_v \) using only the major components of \( \Sigma_v \). Therefore, the Mahalanobis distance \( E_v \) is approximated by \( \hat{E}_v \), which captures the major degrees of statistical variability of AVs,

\[
\hat{E}_v = \sum_{i=1}^{Q_v} \frac{b_i^2}{\gamma_i}, \tag{12}\]

where \( b_i \) is the \( i \)th element of vector \( b \), \( Q_v \) is the number of the principal eigenvectors used, and \( \gamma_i, i = 1, 2, ..., Q_v \) are the \( Q_v \) largest eigenvalues of \( \Sigma_v \). \( b \) is calculated by

\[
b = \Phi^T (v - \bar{v}), \tag{13}\]

where \( \Phi \) is the matrix formed by the \( Q_v \) eigenvectors of \( \Sigma_v \), which correspond to the \( Q_v \) largest eigenvalues \( (\gamma_i, 1 \leq i \leq Q_v ) \) of \( \Sigma_v \).

In summary, given the set of AVs, \( \Omega_v(x) \), determined from a particular location \( x \) in a number of training samples, the criterion for finding the corresponding voxel in an input image is: find the voxel \( y^* \), whose AV \( v(y^*) \) has the maximal pdf value in \( \Omega_v(x) \), by searching within a neighborhood \( N_s(x) \).

2) The Backward Matching Problem — Consistency: In the previous section, we described the forward matching problem, \( i.e. \) given a particular location \( x \) in the model image, how to determine the corresponding voxel \( y^* \) in an individual’s image. In this section, the backward matching problem is discussed, with the objective of investigating how consistent a detected correspondence is. Inverse consistency has been discussed in [14], [27], [44], [48]. In particular, \( y^* \) is a consistent match for \( x \) if \( x \) is also a match for \( y^* \), when one considers the inverse matching problem. The backward matching procedure maximizes the likelihood of the AV \( v(y^*) \) belonging to an AV set \( \Omega_v(z) \) within \( N_m(y^*) \), a neighborhood of \( y^* \) in the model image domain:

\[
x^* = \arg \max_{\zeta \in N_m(y^*)} \left\{ f(v(y^*)|\Omega_v(z)) \right\}. \tag{14}\]

Fig. 5 shows the forward and backward matching schematically. It can be seen that if \( x = x^* \), or the distance between them,

\[
d(x, x^*) = ||x - x^*||, \tag{15}\]

is small, the mapping \( x \rightarrow y \) and \( y \rightarrow y^* \) is consistent. Therefore, given two images, by calculating the forward and backward correspondences of all the voxels in one image, consistent voxel correspondences can be determined by thresholding the consistency distances. For example, given a threshold \( C > 0 \), if \( d(x, x^*) \leq C \), \( (x, x^*) \) is considered to be a consistently corresponding voxel pair. Thus a smaller threshold \( C \) will yield a smaller number of consistent pairs.

Ideally, we would like to find unique correspondence, \( i.e. \) \( y^* \) is the only good match of \( x \), and \( x \) is the only good match of \( y^* \). Since it would be computationally infeasible to test uniqueness of correspondence, we use inverse consistency as a reasonable alternative.

Fig.6(a) and Fig.6(b) give some examples of the voxels that yield different consistency distances, and Fig.6(c) shows the histogram of all the voxels vs. \( d(x, x^*) \). It can be seen from
the histogram that almost half of the voxels of the image have distances less than 5mm. Generally, the voxels located at boundaries and interior areas (such as the boundary of the ventricle, the roots of sulci, etc) yield more consistent correspondence results, while those located within the regions of uniform intensity and the cortical areas may yield less accurate correspondences. Because the local variability around cortical areas across different subjects is relatively higher than that of the interior areas, there is less consistency in these areas. Other factors may also affect the inverse-consistency of correspondence. For example, the similarity between the AVs of two voxels might be high, but it just happens that other voxels located in that vicinity also have similar AVs.

C. Correspondence Detection by Incorporating Shape Constraints

In the previous matching schemes, AV similarities or the statistics of AVs of corresponding voxels were examined independently for each voxel, and the spatial constraints among candidate matches were not taken into account. For example, in Fig.7, voxels $A'$ and $B'$ are the best matches of voxels $A$ and $B$ respectively. Suppose $C'$ and $C''$ are two best candidate matches for voxel $C$, after considering the spatial relationship among the voxels, $C'$ would be chosen as the corresponding voxel of voxel $C$. Voxel $C''$ is possible to be selected only if AV similarity is examined independently and the voxel’s spatial location with respect to other voxels is not taken into consideration. As a result, more robust correspondence detection results can be obtained by incorporating the shape constraints.

Therefore, the other matching scheme used in our experiments considers the collection of voxels, for which matches are sought as a deformable shape, and it finds matches based on AV similarity as well as constraints imposed by the shape’s expected range of variation. The standard approach of Active Shape Models [58] is used here to build the statistical model of the shapes.

Suppose a shape $S = \{s_1, s_2, \ldots, s_M\}$ in the standard shape domain is formed by $M$ voxels $s_i$, ($i = 1, \ldots, M$), the spatial coordinates of all the voxels can be concatenated into a vector $s$. Denote the set of all training samples as $\Omega_s$, which we assume follows a single high-dimensional Gaussian distribution, with mean $\bar{s}$ and covariance $\Sigma_s$, the likelihood of a new shape $s$ belonging to $\Omega_s$ is defined as

$$f(s|\Omega_s) = \xi \cdot \exp\left\{ \frac{-1}{2} E_s \right\} ,$$

where $\xi$ is the normalization coefficient, and $E_s = (s - \bar{s})^T \Sigma_s^{-1}(s - \bar{s})$ is the Mahalanobis distance between $s$ and $\bar{s}$.

This statistical shape model can be efficiently and robustly estimated from a number of training samples using PCA. First, in order to train the model, the landmarks of $N$ sample shapes are manually marked by experienced raters so as to acquire the point correspondences among those images. Then, by selecting one sample shape to define the standard shape domain, all the other $N - 1$ sample shapes are transformed into this standard shape domain using affine transformations. The parameters of the affine transformations are determined using the least squares error method. Finally, the Mahalanobis distance $E_s$ is
estimated by \( \hat{E}_s \) through performing PCA on the normalized sample shapes:

\[
\hat{E}_s = \sum_{i=1}^{Q_s} b_{s,i}^2 \gamma_{s,i} \tag{17}
\]

where \( \gamma_{s,i} \) (\( i = 1, \ldots, Q_s \)) are the \( Q_s \) largest eigenvalues of the covariance matrix \( \Sigma_s \) of the sample shapes, \( b_{s,i} \) is the \( i \)th element of vector \( b_s \), and \( b_s = \Phi_s^T (s - \bar{s}) \). \( \Phi_s \) is the matrix formed by the corresponding \( Q_s \) eigenvectors of \( \Sigma_s \). On the other hand, given a new vector \( b_s \), its corresponding shape can be reconstructed by using

\[
s = \bar{s} + \Phi_s b_s. \tag{18}
\]

Having calculated the likelihood \( f(s|\Omega_s) \) of a new shape \( s \) belonging to the statistical model, we use the following optimization procedure to constrain a new shape formed by the corresponding voxels found in the image domain, \( s^I \):

Constrain \( s^I \) using the statistical shape model, until

\[
f(s|\Omega_s) \geq \tau, \tag{19}
\]

where \( \tau \) is a threshold and \( s \) is the new shape in the standard shape domain, \( s^I = T^{-1}(s^I) \). \( T(\cdot) \) is the affine transformation from the standard shape domain to the image domain, and \( T^{-1}(\cdot) \) is its inverse transformation.

This iterative mechanism is described as follows:

- Perform correspondence detection of each voxel of the constrained shape \( s^I_n \) based on AV similarity. By searching within the neighborhood of each voxel of \( s_n^I \) and selecting its best candidate match that yields the highest AV similarity, the new corresponding shape \( s^I \) is determined. Note \( s^I_n \) is not available for the first iteration, thus it is set as \( \bar{s} \).

- Calculate the affine transformation \( T(\cdot) \) between the mean shape \( \bar{s} \) and the new corresponding shape \( s^I \), using the least squares error method.

Thus, the new shape in the standard shape domain can be obtained by,

\[
s = T^{-1}(s^I). \tag{20}
\]

- Apply shape constraint to \( s \) in the standard shape domain and update the result in the image domain accordingly.

  - Calculate \( \hat{E}_s \) using Eq.(17), and get \( f(s|\Omega_s) \) by substituting \( \hat{E}_s \) to \( E_s \) using Eq.(16).

  - If Eq.(19) is not satisfied, scale \( b_s \) by \( \eta \) (\( 0 < \eta < 1 \)) until Eq.(19) holds: \( b_s = \eta b_s \). Note the scaling of \( b_s \) will make \( \hat{E}_s \) in Eq.(17) smaller, and hence \( f(s|\Omega_s) \) in Eq.(19) will be larger.

  - Calculate a new shape in the standard shape domain, \( s_n^I \), by substituting \( b_s \) in Eq.(18), thus the constrained new shape in the image domain is

\[
s^I_n = T(s_n^I). \tag{21}
\]

- Iteratively perform the above procedure until the corresponding shape \( s^I_n \) converges.

### III. EXPERIMENTAL RESULTS

In this section, experiments are carried out to evaluate the proposed automatic correspondence detection algorithm. In the first set of experiments, a number of voxels are selected from one image, and their corresponding voxels in other images are determined. To evaluate the performance of the algorithm quantitatively, in the second set of experiments, the WAV is evaluated by analyzing the statistics of the WAVs of corresponding voxels from various sulci and showing that AV is indeed highly distinctive and characteristic of the underlying anatomy. Moreover, three AV matching schemes are compared by evaluating the correspondence difference between the detection results of our algorithm and those manually marked by experts.

#### A. Correspondence Detection Results Using Similarity of WAVs

In this experiment, we show some representative examples of correspondence detection using the proposed wavelet-based attribute vector. The images used in all the experiments are 3D T1-weighted gradient echo MR images of the normal human brains of individuals aging from 59 to 90. We used 20 subjects, and there is one image for each subject. The intensities of all the images have been normalized to the range [0,255]. The images have been interpolated using spline-based method and re-sampled so that they have isotropic voxel sizes. The original voxel size is 0.9375mm, 0.9375mm, and 1.5mm in x, y, and z directions. In the experiments, the voxel size after re-sampling is 0.9375mm in all directions, and the new image size is 256x256x144. Prior to the experiments, the global difference between the MR brain images have been corrected based on point correspondences determined by using the Gradient-based Attribute Vectors (GAV), which are described in Appendix.

The intensity normalization was performed using a histogram transformation method [59]. First, a number of landmark configurations are automatically selected from the histogram of an input image and that of the model image respectively. Then, a continuous monotonic 1-D transformation function is interpolated from these landmark pairs. At last, the intensities of the input image are transformed using this transformation function. In our experiments, we used the landmark configuration \( L_A \) for each histogram according to Eq.(1) of [59], which includes 11 landmarks. Please refer to [59] for details. This normalization procedure makes the intensities and contrasts of the normalized images similar to those of the model image.

The WAVs used in all the following experiments were calculated as follows: Three resolution levels of the images are used to calculate the WAV of each voxel. The sliding window size is set to 16x16x16 for the lowest level, 8x8x8 for the middle and highest resolutions. Two-level DWT decompositions are performed to calculate the wavelet subimages at each resolution level, where \( \lambda_1, \lambda_2 \) and \( \lambda_D \) are set to 1. In addition, in order to reduce the size of the WAV vectors, \( N_r \) is set to 2 for the lowest resolution and 1 for other resolutions, and \( \Delta r \) is 4 for all the resolutions. Also, the weighting coefficients to concatenate the WAVs at different resolutions are set as
Fig. 8. Examples of the results of correspondence detection using similarity of WAVs. The corresponding voxels in image two were determined by finding the voxels with the most similar WAV compared to the selected voxels in image one. (a) Selected voxels in image one; all the voxels are from the three slices shown; (b) the corresponding voxels found in image two. The size of the crosses shows whether a detected corresponding voxel is in a neighboring slice above (smaller crosses) or below (larger crosses) the one displayed.

Fig. 9. 3D images of the matching results of the right column of Fig. 8. (a) Selected voxels in image one; (b) corresponding voxels in image two.
\[ \lambda_0' = 0.15, \lambda_1' = 0.2 \text{ and } \lambda_2' = 1.0 \] (The selection of these weighting coefficients is discussed in Section 3.2). The length-2 Daubechies wavelets are used to calculate the DWT subimages.

The correspondence detection is performed using the similarity of WAVs defined in Eq.(11). Given a voxel from one image, its corresponding voxel in another image is determined by finding the voxel, whose WAV is the most similar to the WAV of the voxel given.

We tested correspondence detection on hundreds of voxels from the images of the 20 subjects. The voxels were selected randomly throughout the entire brain, including those voxels inside and at the boundary of ventricles, and voxels of sulci and gyri, whose correspondences are of interests to us. Each time we manually selected one voxel from one image, and then automatically found and marked its corresponding voxels in other images. We then justified the goodness of detection by visually comparing these two voxels. The correspondence results of our approach were generally satisfactory. Note that our goal is not to use our approach for defining correspondence everywhere in the brain, but rather to use it to determine reliable correspondences of a large number of voxels between images, so that the correspondence detection algorithm can be used in the future in conjunction with deformable registration methods. Fig.8 gives representative correspondence detection results. In order to be able to display these results, we selected voxels in one image from only 3 levels in the brain. Fig.8(a) shows these levels and the selected voxels. The corresponding voxels found from the second image are shown in Fig.8(b). Since voxels from the same slice of image one can correspond to voxels in image two that do not lie on the same slice, the crosses ‘+’ with different sizes are used to represent the positions of the resultant matching voxels. The size of the cross indicates whether the corresponding voxel is above (smaller crosses) or below (larger crosses) the displayed slice.

Since different brains have different shapes, it is often difficult to appreciate correspondences (e.g. right column of Fig.8(a)). Therefore we display volumetric renderings of those images in Fig.9 and Fig.10, with corresponding voxels overlaid. Matching performance is typically better for the interior voxels than for surface voxels. This may be due to the fact that shape variation on the surface is larger than that in deep structures. Moreover, for surface voxels, a large part of the sliding window used for determining the AV includes the background. i.e. there is less anatomical context to guide the matching process.

The computation time for all the WAVs of a typical 3D image (with size 256x256x144) is about 1 hour in an SGI workstation (500MHz), where three resolutions are used, and their sizes are 256x256x144, 128x128x72, 64x64x36 respectively.

B. Statistical Analysis of WAVs

In order to evaluate the ability of the WAVs to distinctly characterize different anatomical locations, in this section we examine the WAVs of the corresponding voxels located on roots of different sulci. To facilitate cortical landmark selection, which would be difficult based on cross-sectional images, we utilized 3D renderings of the sulcal curves obtained from cortical surface reconstructions using the method described in [46], [60], [61].

As shown in Fig.11(b), five sulci from the left hemisphere are used, and they are the central sulcus, the superior temporal sulcus, the superior frontal sulcus, the cingulate sulcus and the calcarine sulcus. In order to be able to display the results, we selected 2 voxels from each sulcal curves, i.e., 10 voxels from all the sulci. These five sulci are chosen because they are the major and most representative sulci of the brain, which can be manually traced on the cortical surface reliably. We are therefore interested in evaluating the ability of the proposed WAV to distinguish the sulcal voxels of them. It is desired that these representative voxels have very different WAVs.

Denoting the selected voxels as \( P_1, P_2, \ldots, P_{10} \), the distributions of their WAVs are estimated from 57 training samples using PCA. The the pdf values of different voxels belonging to different classes of WAVs are calculated using Eq.(9). Table I lists the results. Each row corresponds to a class of the WAVs at the respective location. For example, the 5th entry of the 3rd row represents the pdf value of \( P_5 \) belonging to the class of \( P_3 \). (This value is averaged over 57 different brain images). It can be seen from the table that the diagonal values are
Fig. 11. 10 voxels are selected from different sulcal curves to evaluate the distinction of the designed WAVs. (a) the 3D view of a brain image; (b) the 10 voxels selected from the 5 sulcal curves of the brain.

**TABLE I**

<table>
<thead>
<tr>
<th>Class of $P_1$</th>
<th>$P_2$</th>
<th>$P_3$</th>
<th>$P_4$</th>
<th>$P_5$</th>
<th>$P_6$</th>
<th>$P_7$</th>
<th>$P_8$</th>
<th>$P_9$</th>
<th>$P_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class of $P_1$</td>
<td>0.192</td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.016</td>
<td>0.006</td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>Class of $P_2$</td>
<td>0.000</td>
<td>0.167</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Class of $P_3$</td>
<td>0.001</td>
<td>0.000</td>
<td>0.278</td>
<td>0.003</td>
<td>0.007</td>
<td>0.028</td>
<td>0.033</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Class of $P_4$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.334</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Class of $P_5$</td>
<td>0.015</td>
<td>0.001</td>
<td>0.014</td>
<td>0.000</td>
<td>0.408</td>
<td>0.064</td>
<td>0.000</td>
<td>0.018</td>
<td>0.000</td>
</tr>
<tr>
<td>Class of $P_6$</td>
<td>0.029</td>
<td>0.021</td>
<td>0.002</td>
<td>0.000</td>
<td>0.004</td>
<td>0.216</td>
<td>0.003</td>
<td>0.014</td>
<td>0.000</td>
</tr>
<tr>
<td>Class of $P_7$</td>
<td>0.002</td>
<td>0.005</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.170</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>Class of $P_8$</td>
<td>0.004</td>
<td>0.003</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.012</td>
<td>0.000</td>
<td>0.175</td>
<td>0.000</td>
</tr>
<tr>
<td>Class of $P_9$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.006</td>
<td>0.000</td>
<td>0.181</td>
</tr>
<tr>
<td>Class of $P_{10}$</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.020</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
</tr>
</tbody>
</table>

much larger than non-diagonal values, which implies a higher probability that the WAV of a voxel belongs to the class of its corresponding voxels in other brains. Moreover, the low pdf values of the WAVs of noncorresponding voxels indicate that the designed WAV is distinctive enough, and therefore it is specific in determining correspondence of these voxels across these 57 individuals.

Further, we explain how to select the weighting coefficients $\lambda_0, \lambda_1, ..., \lambda_K$ used in the experiments (See Eq.(8)). Given a number of corresponding voxels from different parts of sample images, we can estimate the distribution of WAVs for each corresponding voxel location and obtain the table of pdf values like Table 1. Since different configurations of the weighting coefficients yield different element values in the table, the weighting coefficients are determined by performing a global search within the range $[0, 1]$, so that the ones that yield the highest pdf values for the diagonal elements and the lowest values for other elements are chosen. In our experiments, we have used the 10 sulcal voxels described above, along with the 18 voxels used in the next experiment (see Fig.12) as our corresponding voxel sets for training the coefficients. The coefficients are fixed in all experiments after training. Therefore, the matching results might be biased toward the structures around these voxels. In future work, we would like to create a broader selection of training voxels. Moreover, if the correspondences among a number of sample images are available, spatially-adaptive coefficients can be obtained by evaluating the goodness of correspondence detection at different parts/locations of the brain.

**C. Comparison of the AV Matching Methods**

In this section we present a quantitative evaluation of the correspondence automatically determined using different AV matching methods, by comparing it with the correspondence defined by human experts. 20 MR brain images from different normal subjects were selected from our database. For each image, 18 landmarks were manually marked by two experts separately. Fig.12 shows one of the sample images with the marked landmark voxels.

Three sets of experiments were carried out and their matching performances are compared with that of the experts.

- **Experiment 1: Use WAV similarity** (see Eq.(11) in Section 2.2.1). In this experiment, given voxels from one image, their corresponding voxels in other images are determined by finding the voxels that have the most similar WAVs.

- **Experiment 2: Use the statistical models of WAVs** (see Eq.(10) in Section 2.2.1). In this experiment, half of the images marked by each expert are selected as the training samples, and the rest are used for testing.
Fig. 12. One of the sample images marked by the medical experts (all the 18 marked voxels are shown). Automatic correspondence detection was found to be similar to expert definitions for these 18 voxels.
**Experiment 3: Use shape constraints (see Section 2.3).**

In this experiment, the prior shape constraints among the landmarks are used for correspondence detection. The 10 training images and 10 test images are the same as those of Experiment 2.

The method to quantitatively analyze the accuracy of the matching results is described as follows.

- Calculate the mean of the detection disagreement of each experiment. If we denote the voxel \( j \) marked by expert \( k, (k = \{1, 2\}) \) in image \( i \) as \( \mathbf{P}_{i,j,k} \), and the corresponding voxel determined by the proposed algorithm as \( \mathbf{P}_{i,j,k}' \), then the correspondence disagreement is

\[
e_{i,j,k} = \| \mathbf{P}_{i,j,k} - \mathbf{P}_{i,j,k}' \|.
\]

For voxel \( j \), the mean of \( e_{i,j,k} \) across all the images and experts provides the measurements of the correspondence differences.

- Compare the correspondence differences of the three experiments with the disagreement between the experts, which is defined in a way analogous to Eq.(22).

Fig. 13 shows the means and the standard deviations of \( e_{i,j,k} \) for the 18 manually marked landmarks between two experts across 20 images. It can be seen that the differences are large, particularly in cortical landmarks. For example, the mean of the absolute differences for landmark 7 is 12.5mm. The mean of the absolute difference between the corresponding landmarks found by our algorithm with different matching schemes and the voxels marked by the experts for the three experiments are given in Fig. 14. The mean, standard deviation, and root mean square (RMS) values of the differences in Fig. 14 are shown in Table II.

We also performed paired two-tail hypothesis tests on the experimental results. The results show that at significance level 0.05, there is no significant difference between the correspondence differences of Experiment 3 (using shape constraints) and the disagreements of the experts (\( p=0.14 \)). On the other hand, the difference between the correspondence differences of Experiment 1 (using WAV similarity) and the disagreements of experts (\( p=0.01 \)), and between those of Experiment 2 (using statistical models of WAVs) and those of the experts (\( p=0.047 \)), reached statistical significance.

The accuracy of the three matching schemes can be analyzed by using the mean values shown in Table II. It can be seen that for the 18 landmarks, the mean of correspondence differences is 9.7mm for Experiment 1. This largest value indicates that the method only using WAV similarity performs the worst among all the methods. On the other hand, the mean of the correspondence differences of Experiment 3 is 4.6mm, and it is obvious that this method using the shape constraints among the landmarks outperforms the other two methods. In summary, the correspondence detection algorithm using the statistical models of the WAVs of corresponding landmarks (Experiment 2) performs better than that only using WAV similarity. This is because the variability of WAVs of corresponding landmarks is characterized by the statistical model, which can tolerate the variability of corresponding anatomical structures of different subjects better. Moreover, the correspondence detection results can be further improved by incorporating the prior shape constraints among the landmarks. Comparing these mean values, we can see that the spatial shape constraints play the most important role in improving the performance of correspondence detection.

In conclusion, among the three WAV matching schemes, the method incorporating shape constraints performs the best. Moreover, the hypothesis tests indicate that its performance is similar to that of the experts (\( p=0.14 \)). We intend to use the wavelet-based attribute vector and the candidate correspondence implied by it in order to guide deformable registration algorithms.

It worth noting that for all the three WAV matching schemes the correspondence detection procedure is fully automated. Of course, in order to build the spatial shape model of the landmarks and the statistical models of the WAVs of corresponding landmarks, prior knowledge of landmark correspondences among a number of sample images is required. In this paper, these landmarks were manually marked by experts.
to train the statistical models. The forward matching method using WAV similarity can be used to match a large number of voxels among images, when the voxel correspondences used for training are not available.

We are exploring automated landmark definition techniques for the training set, as we have previously done in [41]. Since precisely defining correspondences for a large number of landmarks is practically infeasible, we can define a "silver standard" by manually pointing (labeling) a number of brain regions, such as gyri or sulci, and then using an automated deformable registration method. The availability of adequately large number of pointed regions is likely to help the deformable registration method to find the correct transformation, which in turn defines a large number of correspondences. Although not exactly a gold standard, this approach is able to provide reasonably good statistical models to constrain our automated correspondence detection approach.

IV. CONCLUSION

This paper proposed an automatic correspondence detection algorithm for 3D intramodality and intersubject MR brain images using wavelet-based attribute vectors. The wavelet-based AVs are designed to reflect the morphological signatures of voxels, and correspondence detection is achieved by matching WAVs. Three AV matching schemes are presented: i) matching using WAV similarity, ii) matching using statistical models of WAVs, and iii) matching by incorporating shape constraints. Experimental results indicate that the algorithm performs worse than experts for WAV similarity and statistical models of WAVs, but as good (p=0.14) as experts when shape constraints are used in addition to statistical models of WAVs. i.e. the difference between the correspondence differences of the proposed method using shape constraints and the disagreements of the experts is small and not significant (p=0.14). Therefore, it has potential for use in automatic image registration by being incorporated into image warping algorithms.

APPENDIX I

GLOBAL LINEAR TRANSFORMATION OF IMAGES USING GRADIENT-BASED ATTRIBUTE VECTORS (GAVS)

In order to globally correct the rotation between two images, the Gradient-based AV (GAV), with the advantage of being rotation-invariant, is also proposed. Although a standard image similarity-based affine registration method, e.g. [12], can also be used to compensate the global transformation between images, in this paper we propose to use the GAV-based approach, which in contrast to those standard methods, is relatively less affected by image shape/intensity variability, partial volume effects, local minima, etc. Moreover, the GAV-based method can tolerate a large rotation between images, which is not the case for an image similarity-based method that needs to be initialized quite close to the true solution or to be performed with a multi-resolution search strategy.

A. Design of Gradient-based Attribute Vector

The Gradient-based AV (GAV) has the advantage of being rotation-invariant, an important property which guarantees that the similarity of GAVs is independent of the orientation of objects relative to the scanner. Unlike the construction of WAV, Gaussian filters are used to calculate the multi-resolution features in step 2 of the approach presented in Section 2.1.1, i.e. the GAV of a voxel is acquired by directly performing the RP method on the absolute gradient fields of the Gaussian filtered images at different scales, and then combining all those feature vectors together. Because the absolute gradient field of an image follows the rotation of that image, the GAV extracted using RP is also rotation-invariant. The construction of GAV is described as follows:

First, smooth the normalized input image using Gaussian functions. At resolution $j$, the Gaussian filtered image is

$$I_{\sigma_j}(x) = I(x) * G_{\sigma_j}(x), \quad (1 \leq j \leq J)$$  \hfill (23)

where $\sigma_j$ is the standard deviation of the Gaussian function $G_{\sigma_j}(x)$, and $\sigma_1 < \sigma_2 < \ldots < \sigma_J$.

Then, the gradient magnitude $L_{\sigma_j}(x)$ of the filtered image $I_{\sigma_j}(x)$ is calculated by

$$L_{\sigma_j}(x) = \sqrt{\frac{\partial I_{\sigma_j}}{\partial x}^2 + \frac{\partial I_{\sigma_j}}{\partial y}^2 + \frac{\partial I_{\sigma_j}}{\partial z}^2}.$$  \hfill (24)

If $I_{\sigma_j}$ is rotated in a given angle in 3D space with w.r.t. $x_0$, its gradient magnitude $L_{\sigma_j}$ will follow the same rotate.

Finally, the gradient attribute vector, $v^{(j)}_g$, at each resolution level $k$ can be obtained by applying RP to the gradient field $L_{\sigma_j}(x)$, and the GAV, $v_g$, of a voxel $x_0$ is constructed by concatenating all these vectors:

$$v_g = [\lambda_1 v^{(1)}_g]^T, \lambda_2 v^{(2)}_g]^T, \ldots, \lambda_k v^{(j)}_g]^T]^T.$$  \hfill (25)

B. Estimation of Affine Transformation Based on GAV-Based Correspondences

After designed the GAV, given a set of voxels in image $I_1$, their corresponding voxels in image $I_2$ can be determined by

<table>
<thead>
<tr>
<th></th>
<th>Between Experts</th>
<th>Experiment 1: use WAV similarity</th>
<th>Experiment 2: use statistical model of WAVs</th>
<th>Experiment 3: use shape constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>5.9</td>
<td>9.7</td>
<td>7.6</td>
<td>4.6</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>3.6</td>
<td>2.7</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>RMS</td>
<td>6.7</td>
<td>10.1</td>
<td>7.8</td>
<td>4.8</td>
</tr>
</tbody>
</table>
evaluating similarities of GAVs. Then, the affine transformation between $I_1$ and $I_2$ can be estimated by minimizing the following sum of squared errors:

$$
\hat{T} = \arg \min_T \left\{ \sum_{i=1}^{N} \| y_i - T(x_i) \|^2 \right\}
$$  \hspace{1cm} (26)

where $x_i$ (i=1...N) are the selected voxels in $I_1$, $y_i$ (i=1...N) are the corresponding voxels determined in $I_2$, and $T'$ is the affine transformation.

In this way, $I_1$ can be globally transformed by the estimated affine transformation $\hat{T}$, so that $\hat{T}(I_1)$ is similar to $I_2$ in scale and orientation. Fig.15 shows two examples of the results. Fig.15 (a) and (b) are two different input images, and Fig.15 (c) is the model image. The goal is to estimate the global affine transformations between the input images and the model image, so as to correct the global differences. In order to do so, we first selected a number of voxels (500 in the experiments) from the model image, and then determined their corresponding voxels in the input images. Based on these voxel correspondences, one input image can be globally transformed onto the model image domain using the least squares error method described above. The affine transformed images are shown in Fig.15 (d) and (e). It can be seen that after affine transformations, the images are very similar to the model image.

**REFERENCES**


A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Ang, and IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. XX, NO. XX, JUNE 2004 18


