Affine Registration of label maps in Label Space

Yogesh Rathi, James Malcolm, Sylvain Bouix, Allen Tannenbaum, Martha E Shenton,

***Abstract*—Two key aspects of coupled multi-object shape analysis and atlas generation are the choice of representation and subsequent registration methods used to align the sample set. For example, a typical brain image can be labeled into three structures: grey matter, white matter and cerebrospinal fluid. Many manipulations such as interpolation, transformation, smoothing, or registration need to be performed on these images before they can be used in further analysis. Current techniques for such analysis tend to trade off performance between the two tasks, performing well for one task but developing problems when used for the other.**

This article proposes to use a representation that is both flexible and well suited for both tasks. We propose to map object labels to vertices of a regular simplex, *e.g.* the unit interval for two labels, a triangle for three labels, a tetrahedron for four labels, etc. This representation, which is routinely used in fuzzy classification, is ideally suited for representing and registering multiple shapes. On closer examination, this representation reveals several desirable properties: algebraic operations may be done directly, label uncertainty is expressed as a weighted mixture of labels (probabilistic interpretation), interpolation is unbiased toward any label or the background, and registration may be performed directly.

We demonstrate these properties by using label space in a gradient descent based registration scheme to obtain a probabilistic atlas. While straightforward, this iterative method is very slow, could get stuck in local minima, and depends heavily on the initial conditions. To address these issues, two fast methods are proposed which serve as coarse registration schemes following which the iterative descent method can be used to refine the results. Further, we derive an analytical formulation for direct computation of the "group mean" from the parameters of pairwise registration of all the images in the sample set. We show results on richly labeled 2D and 3D data sets.

***Index Terms*—Registration, probabilistic atlas, richly labeled images, multi-object shape analysis**

I. Introduction

Multi-object shape analysis is an important task in the medical imaging community. When studying the neuroanatomy of patients, clinical researchers often develop statistical models of important structures which are then useful for population studies or as segmentation priors [18, 21, 23, 29, 28]. The first step for this problem consists in choosing an appropriate shape descriptor capable of representing its statistical variability.

There are two main types of models: explicit and implicit. Splines and medial axis skeletons are two popular examples of explicit models [6, 7, 16, 22, 25, 26]. While providing a reduced parametric representation, explicit models have several drawbacks. For example, they often assume a fixed shape topology, require care in distributing control points, and the representation can get very complicated for complex structures like the white matter or gray matter.

Y. Rathi, S. Bouix, J. Malcolm, and M.E. Shenton are with the Psychiatry Neuroimaging Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston

Y. Rathi is also an adjunct faculty at Georgia Institute of Technology, Atlanta

A. Tannenbaum is a professor at Georgia Institute of Technology, Atlanta

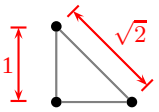


Fig. 1. Tsai et al. [29] proposed mapping each pixel from object label to a point in a space shaped as a non-regular simplex, each vertex corresponding to an object label. Visualized here for the case of two objects and background, the bottom left background (0,0) is a distance of 1 from both labels top (0,1) and right (1,0), while labels are separated from each other by a distance of $\sqrt{2}$.

This work focuses on implicit models [18, 28] which avoid these problems. After mapping the entire volume to another space, the value of each voxel contributes to describe the shape. In this new space, arbitrary topologies may be represented, correspondences are naturally formed between voxels, and there are no control points to redistribute. However, since this shape space is often of higher dimension than the original dataset, one key disadvantage for this type of representational model is that it will usually increase the spatial and computational complexity of the analysis.

A common starting point for shape representation is a simple scalar label map, each pixel indicating the object present at that pixel, *e.g.* a label value of 1 indicating object #1, a value of 2 indicating object #2, etc. Many techniques [28, 3] go on to map this entire volume to another space, the value of each pixel contributing to describe the shape.

The simplest implicit representation is a binary map where each pixel indicates the presence or absence of the object. Signed distance maps (SDM's) are another example of an implicit representation, each pixel having the distance to the nearest object boundary, a negative distance for points inside the object [18, 28].

For the multi-object setting, binary maps may be extended to scalar label maps, each pixel holding a scalar value corresponding to the presence of a particular object; however, this representation is not well suited for algebraic manipulation. For example, if labels are left as scalar values, the arithmetic average of labels with values #1 and #3 would incorrectly indicate the label of value #2, not a mixture of labels #1 and #3.

To address this, mappings of object labels to linear vector spaces were proposed, an approach to which our method is most closely related. The work of Tsai et al. [29] introduced two such representations, each for a particular task. For registration, the authors proposed mapping scalar labels to binary vectors with entries corresponding to labels; a 1 in an entry indicates the presence of the corresponding label at that pixel location. As an example for the case of two labels

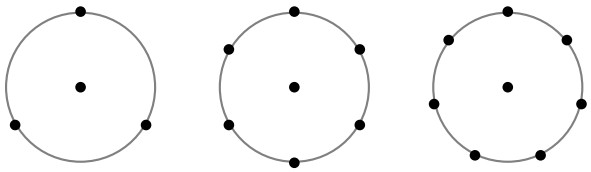


Fig. 2. Example configurations for the S^1 hypersphere representation of [3]: three, six, and seven labels *(left to right)* with background at the center.

and background, Figure 1 visualizes the spatial configuration each pixel is mapped onto. Here the background is at the bottom left origin (0,0) with one label at (1,0) and the other at (0,1). It is also important to note that they go on to perform registration considering each entry of these vectors separately. For shape analysis, Tsai et al. [29] proposed mapping scalar labels to layered SDM's, in this case each layer giving the signed distance to the corresponding object's interface.

Note that in both vector valued representations described in Tsai et al. [29], each label lies on its own axis and so the dimension of the representation grows linearly with the number of labels, *e.g.* two objects require two dimensions, three objects require three dimensions. To address this spatial complexity, Babalola and Cootes [3, 4] propose a lower dimension approximation to replace the binary vectors in registration. By mapping labels to the unit hypersphere S^n, they demonstrate that even configurations involving dozens of labels can be efficiently represented with label locations distributed uniformly on a hypersphere. Figure 2 gives examples for S^1.

Another representation that is widely used in fuzzy classification is that of [5] which uses a positive simplex to represent probabilistic data. It has also been used for representing posterior probability distribution or fuzzy labeling of image segmentation data [13]. In this work, we propose to use it for representation and registration of multiple shapes.

Finally, Pohl et al. [23] indirectly embeds label maps in the logarithm-of-odds space using as intermediate mappings either the binary or SDM representations of [29]. Particularly well suited for probabilistic computations, the logarithm-of-odds space is also a field providing closed operations for addition and scalar multiplication. As with the representations of Tsai et al. [29], the dimensionality of the logarithm-of-odds space increases with each additional object. We should also note that the work of [23] did not address registration, but instead assumed an already registered atlas via [20]. Another related work is that of [13], which performs a mutual-information based non-rigid registration of probabilistic data represented as class labels, similar to the Log-Odds representation [23].

Once the representation is settled upon, registration must be performed to eliminate variation due to differences in pose. A common approach is to register the set to a reference image; however, this then introduces a bias to the shape of the chosen reference. Joshi et al. [15] propose unbiased registration with respect to the mean sample as a template reference. Assuming a general metric space of transformations, they describe registering a sample set with respect to its intrinsic mean and use the L_2 distance for demonstration. A similar approach uses the minimum description length to measure distance from the intrinsic mean [30]. Instead of registering to a mean template, an alternative approach is to minimize per-pixel entropy. Using binary maps Miller et al. [20] demonstrate that this has a similar tendency toward the sample mean. This approach has also been demonstrated on intensity images [31, 32]. Among these energy-based registration techniques, iterative solutions include those that are variational [29, 15] and those that use sampling techniques [32]. Other registration methods can be found in [19, 24].

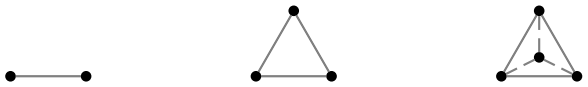


Fig. 3. The first three label space $\mathcal{L}$ configurations: a unit interval in 1D for two labels, a triangle in 2D for three labels, and a tetrahedron in 3D for four labels *(left to right)*.

A. Our contributions

This paper proposes to use a multi-object implicit representation scheme that is widely used to represent probabilistic data [17]. In this representation, object labels are mapped to the vertices of a regular simplex, going from a scalar label value to a coordinate position in a high dimensional space which we term *label space* and denote by $\mathcal{L}$. Visualized in Figure 3, a regular simplex is an n-dimensional analogue of an equilateral triangle. Under convex combinations, label space has several desirable properties: all labels are equally separated in space, addition and scalar multiplication are natural, label uncertainty is expressed as a weighted combination of label vertices, and interpolation is unbiased toward any label including the background.

This representation addresses several problems with current implicit mappings. For example, while the binary vector representation of Tsai et al. [29] was proposed for registration, we will demonstrate that it induces a bias sometimes leading to misalignment, and since our label space representation equally spaces labels, there is no such bias. Additionally, compared to the SDM representation, the proposed method introduces no inherent per-pixel variation across equally labeled regions making it more robust for statistical analysis. Hence, the proposed method better encapsulates the functionality of both representations. We will also demonstrate that, while lowering the spatial demands of the mapping, the hypersphere representation of Babalola and Cootes [3] biases interpolation and can easily lead to erroneous results. The arrangement of label space incurs no such bias allowing convex combinations of arbitrary labels.

Using the *label space* representation, we propose two fast techniques that coarsely register the labeled images to their intrinsic mean. These coarsely registered images then serve as the starting point for a gradient descent based approach which typically converges within 10-20 iterations to a better minimum.

Further, we derive a formulation to analytically compute the parameters that take every image in a given sample set to the group mean. This is done by pairwise registration of each image to every other image in the group, and the affine parameters so obtained can be used to compute the group mean.

The rest of this paper is organized as follows. Section II

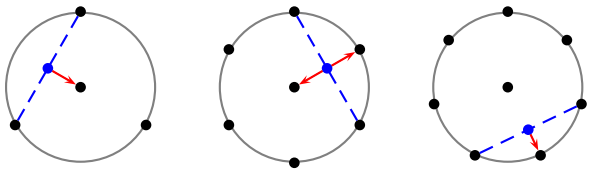
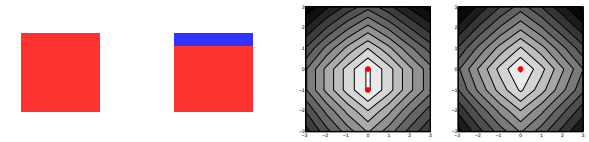


Fig. 4. For the S^1 hypersphere configurations of [3], cases such as these yield erroneous results during interpolation. Judged by nearest neighbor, interpolating between two labels resolves to background, ambiguously either background or another label, and finally another label *(left to right)*.



(a) Reference (b) Image (c) Energy landscape using [29] (d) Energy landscape using label space

Fig. 5. Alignment of an image with a reference template using the representation of [29] results in two possible alignments, the shifted one misaligning along both the top and bottom with respect to the reference *(red dots indicate minima)*. For just x- and y-translation, isocontours of the energy landscape show the non-unique energy minima in (c).

explores several problems that can develop with the implicit representations described above [3, 23, 29]. Section III then describes the proposed label space representation $\mathcal{L}$ documenting several of its properties. Section IV demonstrates variational registration directly within this representation, and Section V describes two methods for performing coarse registration on labeled images and Section VI shows some experimental results on 3D data sets. Finally in Section VII we summarize our work and discuss future research directions.

II. Related representations

In this section, we describe problems that may develop in the representations this present work seeks to extend.

A. Shape representation

The signed distance map (SDM) has been used as a representation in several studies [1, 18, 23, 29, 28]; however, it may produce artifacts during statistical analysis [10]. This is due to the fact that SDM's lie on a nonlinear manifold where linear operations like addition/subtration introduce artifacts [10, 14].

Label maps have inherently little per-pixel variation, pixels far from the interface having the same label as those just off the interface. For statistical analysis in the case of one object, Dambreville et al. [9] demonstrated that binary label maps have higher fidelity compared to SDM's. However, for the multi-object setting, the question then becomes one of how to represent multiple shapes using binary maps? What is needed is a richer feature space suitable for a uniform pairwise separation of labels.

An example of such a richer feature space is that of [3] where labels are mapped to points on the surface of a unit hypersphere S^nplacing the background at the center. This is similar to the binary vector representation described by [29] to spread labels out; however, [3] argue that lower dimensional approximations can be made. They demonstrate that configurations involving dozens of labels can be efficiently represented by distributing label locations uniformly on the unit hypersphere using as few as three dimensions. Since any label may neighbor the background, the background must be placed at the hypersphere center, equally spaced from all other labels. The fundamental assumption is that pixels only vary between labels that are located near to each other on the hypersphere, so the placement of labels is crucial to avoid erroneous label mixtures. For example, Figure 4 demonstrates that if two labels far from each other are mixed, the result may be attributed erroneously to other labels. Notice in particular that the central placement of the background gets in the way when interpolating across the sphere. Smoothing in Figure 7 also demonstrates these inherent effects of the lower dimensional approximation, effects that cannot be avoided unless the dimension approaches label cardinality.

The logarithm-of-odds representation of Pohl et al. [23] provides the third and final shape representation we compare against. Aside from the normalization requirement for closed algebraic manipulation, the main concern when using this representation is the choice of intermediate mapping, a choice that directly impacts the resulting probabilities. The authors explore the use of both representations from [29]; however, both choices have inherent drawbacks.

For the layered SDM intermediate mapping, Pohl et al. [23] notes that SDM's are a subspace of the logarithm-of-odds space. This means that, while the layered SDM's are exactly the logarithm-of-odds representation, results after algebraic manipulation in the logarithm-of-odds space often yield invalid SDM's (but still valid logarithm-of-odds representations). Using such results, computing probabilities as described in [23] may yield erroneous likelihoods. Notice also, that the generalized logistic function is used to compute probabilities. This introduces additional problems as the use of the exponential ensures that these probabilities will always have nonzero character across the entire domain, even in areas never indicated by the sample set.

Using smoothed binary maps as intermediates also leads to problems. To begin, using binary maps directly would mean probabilities of either zero or one, which in the log domain produce singularities. Smoothing lessens such effects yet results in a loss of fine detail along the interface. Also, Pohl et al. [23] shows examples where after normalization the logarithm-of-odds representation develops artifacts at the interface between objects, an effect which is magnified in the logarithm domain.

B. Registration

Tsai et al. [29] propose a binary vector representation specifically for registration. As Figure 1 shows, this representation places labels at the corners of a right-triangular simplex; however, unlike this present work, it is not a regular simplex but has a bias with respect to the background. The background, located at the origin, is a unit distance from any other label, while any two labels, located along a positive axis, are separated by a distance of $\sqrt{2}$. The effect may be seen in registration where there is a bias to misalign labels over the background (penalty 1) rather than over other labels (penalty $\sqrt{2}$).

To demonstrate the effect of this induced bias, consider the

Fig. 6. Proposed $\mathcal{L}$ [3] label space for the case of three labels: a point indicating the equal presence of all three labels *(left)*, and a point indicating the unequal mixed presence of just the left and top labels *(right)*.

example in Figure 5. Using the representation and registration energy of Tsai et al. [29], there are two global minima: the image overlapping and the image shifted up. In the first case, label #1 is misaligned over label #2, while in the second case, a strip of pixels at both the top and bottom are misaligned over the background; that is, because of this bias, there can be twice as many pixels misaligned in the shifted case than in the unshifted. These minima (indicated by red dots in the energy landscapes) are shown only for translation; considering additional pose parameters further increases the number of local minima in the energy landscape representing misalignments. Also, this is not inherent in the energy, as the same phenomenon is observed using the energy in (1).

III. Label space

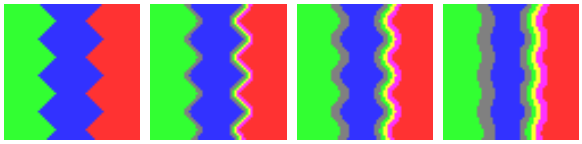
Our goal is to create a robust representation where algebraic operations are natural, label uncertainty is captured, and interpolation is unbiased toward any label. To this end we propose mapping each label to a vertex of a regular simplex; given n labels, including the background, we use a regular simplex which lies in $n-1$ dimensions and denote this by $\mathcal{L}$ (see Figure 3). A regular simplex is an n-dimensional analogue of an equilateral triangle.

A. Probabilistic Interpretation

Label Space has natural probabilistic interpretation as described below and algebraic operations are as natural as vector addition and scalar multiplication in $\mathcal{R}^n$ (under convex combination). Label uncertainty is realized as the weighted mixture of vertices. For example, a pixel representing labels #1, #2, and #3 with equal characteristic would simply be the point $l = \frac{1}{3}l_1 + \frac{1}{3}l_2 + \frac{1}{3}l_3$, a point equidistant from those three vertices (see Figure 6). Using this fact, one can map any point in a probabilistic atlas to a unique point in the label space $\mathcal{L}$, *i.e.* given the probabilities $p_1, .., p_n$ with $\sum p_i = 1$, the corresponding representation in $\mathcal{L}$ is given by $l = \sum p_i l_i \in \mathcal{L}$, where $l_i \in \mathcal{L}$ is a vertex of the simplex. We should however note that only tri-linear interpolation can be used to obtain points that lie within label space. Other interpolation schemes such as sinc or b-spline which do not form a convex combination are not guaranteed to give a mixture label that lies within the simplex.

Alternatively, the probability for $x \in \mathcal{L}$ being label $l \in \mathcal{L}$ is, $P(x = l) = (1 - \|x - l\|)/Z$, where $Z = n - \sum_i \|x - l_i\|$ acts as a normalization constant. We can thus navigate naturally between the space of probabilities and $\mathcal{L}$ using simple algebraic operations. Further, given a point l in $\mathcal{L}$, the most likely label can be obtained by finding the vertex closest to l.

Thus the proposed framework has parallels to the logarithm-of-odds representation, with the advantage that it does not use logarithms, thus removing singularities in the definition of the logit function when any of the probabilities p_i is zero.



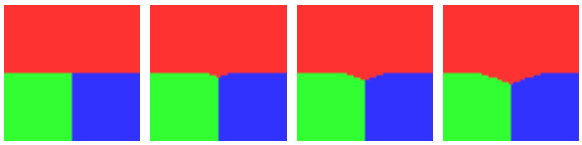
(a) Scalar label map

(b) S^n hypersphere of Babalola and Cootes [3]

(c) $\mathcal{L}$ label space

Fig. 7. Progressive smoothing directly on scalar label maps, the hypersphere representation of Babalola and Cootes [3], and $\mathcal{L}$ label space. Both the scalar label maps and hypersphere representations develop intervening strips of erroneous labels. Only label space is able to correctly capture the label mixtures during smoothing. The rightmost hypersphere in Figure 4 depicts the S^1 configuration used here in (b).

(a) Binary vector representation of Tsai et al. [29]



(b) $\mathcal{L}$ label space

Fig. 8. Progressive smoothing directly on binary vector representation of Tsai et al. [29] and $\mathcal{L}$ label space. Smoothing among several labels in the binary vector representation yields points closer to background *(black)* than any of the original labels. Label space is able to correctly begin to smooth out the sharp corners of the bottom two regions without erroneous introduction of the black background label.

B. Other Properties

To demonstrate some of the advantages of label space, we performed progressive smoothing using the various representations described: scalar label values, the binary vector representation of Tsai et al. [29], the S^n representation of Babalola and Cootes [3], and $\mathcal{L}$ label space. In our first experiment, there are seven labels numbering #1 through #7. In Figure 7, we show an example with jagged stripes of labels #5, #7, and #3, respectively. Scalar label values show the appearance of intervening labels #4, #5, and #6 as the original labels blend, and the hypersphere representation shows the appearance of labels #2, #6, and #4 as interpolation is performed across the hypersphere (the hypersphere configuration used here is the rightmost depicted in Figure 4). In Figure 8, the second experiment shows that the smoothing among multiple labels using binary vectors produces points closest to the background

(black). In both experiments, only label space correctly preserves the interfaces. We should however note that the binary map representation of Tsai et al. [29] can be modified so as to remove the bias towards the background. This can be done by considering the background as a separate label and mapping it to one of the corners of the right triangular simplex (not the origin). Thus, by not mapping any label to the origin results in a representation which is similar to label space albeit with higher dimensionality, i.e., for n labels, the dimensionality is n.

The following table summarizes the comparison between different representations discussed above:

Table 1. Comparison of shape representations

Representation	P.I.*	Bias	Dimensionality
Label Space	Yes	No	n-1
Binary Map	No	Yes	n-1
S^n	No	Yes	3
LogOdds	Yes	not known	n-1

where, P.I.* = Probabilistic Interpretation and dimensionality is given for n labels.

IV. Variational Registration

We demonstrate here the variational registration of a set of maps to their intrinsic mean map, thereby respecting the first order statistics of the sample set. The proposed representation has the advantage of supporting registration directly on the representation. By directly we mean that differentiable vector norms may be used to compare labels.

Common approaches to registration begin by fixing one of the maps as a reference and registering the remaining maps to this fixed map. This is done in both [3, 29]; however, as Joshi et al. [15] describes, this initial choice biases the spatial statistics of the aligned maps. To avoid this bias, Joshi et al. [15] describe registration with respect to a reference that best represents the sample set, i.e., the mean. In addition to avoiding bias, the resulting gradient descent involves far less computation than that proposed in [29] where each map is compared against each other map.

Before presenting the energy used, we first describe the problem borrowing notation from [29]. First, each label map is mapped into the label space (simplex representation) to produce a representation m_i for subject i. For the set of label maps $M = \{m_i\}_{i=1}^N$, our goal is to estimate the set of corresponding pose parameters $P = \{\mathbf{p}_i\}_{i=1}^N$ for optimal alignment. We denote as $\tilde{m}$ the label map m transformed by its pose parameters. An advantage of implicit representations over explicit ones is that, once the label maps have undergone this transformation, we can assume direct per-pixel correspondence between maps and use a vector norm to perform comparison. We model pose using an affine model, and so for 2D, the pose parameter is the vector $\mathbf{p} = [x\ y\ s_x\ s_y\ \theta\ k]^T$ corresponding to x-,y- translation, x-,y-scale, in-plane rotation, and shear. Note that this is a fully affine model as compared to the rigid transformation model used in [29]. The transformed map is defined as $\tilde{m}(\tilde{x}, \tilde{y}) = m(x, y)$ where coordinates are mapped according to $\begin{bmatrix}\tilde{x} & \tilde{y} & 1\end{bmatrix}^T = T(\mathbf{p}) \begin{bmatrix}x & y & 1\end{bmatrix}^T$, where $T(\mathbf{p})$ is the decomposable transformation matrix

$$T(\mathbf{p}) = \underbrace{\begin{bmatrix}1 & 0 & x\\ 0 & 1 & y\\ 0 & 0 & 1\end{bmatrix}}_{M(x,y)} \underbrace{\begin{bmatrix}\cos(\theta) & -\sin(\theta) & 0\\ \sin(\theta) & \cos(\theta) & 0\\ 0 & 0 & 1\end{bmatrix}}_{R(\theta)} \underbrace{\begin{bmatrix}s_x & 0 & 0\\ 0 & s_y & 0\\ 0 & 0 & 1\end{bmatrix}}_{H(s_x,s_y)} \times \underbrace{\begin{bmatrix}1 & k & 0\\ k & 1 & 0\\ 0 & 0 & 1\end{bmatrix}}_{K(k)}$$

for a translation matrix $M(x, y)$, rotation matrix $R(\theta)$, anisotropic scale matrix $H(s_x, s_y)$, and shear matrix $K(k)$, all for the parameters taken from $\mathbf{p}$.

As in [15, 32], we assume the intrinsic mean map $\tilde{\mu}$ of the sample set to best represent the population. We then attempt to minimize the energy defined as the squared distance between each transformed label map $\tilde{m}$ and this mean map $\tilde{\mu}$ of the set $\tilde{M}$ as it converges:

$$d^2 = \frac{\sum_{i=1}^N \|\tilde{m}_i - \tilde{\mu}\|^2}{\sum_{i=1}^N \|\tilde{m}_i + \tilde{\mu}\|^2}, \tag{1}$$

where $\tilde{\mu} = \frac{1}{N}\sum_{i=1}^N \tilde{m}_i$, and while $\|\cdot\|$ may be any differentiable norm, we take it to be the elemental L_2 inner product $\|x\| = \langle x, x\rangle^{1/2} = \sqrt{\int x^2 dx}$. The denominator ensures that all of the images do not shrink in size to decrease the energy during gradient descent.

In this section, we describe a variational approach to registration. Specifically we perform gradient descent to solve for the pose parameters minimizing this distance. We find the gradient of this distance, taken with respect to the pose $\mathbf{p}_j$, to be:

$$\begin{aligned}\nabla_{\mathbf{p}_j} d^2 = & 2\frac{\langle \nabla_{\mathbf{p}_j}\tilde{m}_j,\ \tilde{m}_j - \tilde{\mu}\rangle}{\sum_{i=1}^N \|\tilde{m}_i + \tilde{\mu}\|^2} \\ & + 2\frac{\sum_{i=1}^N \|\tilde{m}_i - \tilde{\mu}\|^2 \langle \nabla_{\mathbf{p}_j}\tilde{m}_j,\ \tilde{m}_j + \tilde{\mu}\rangle}{\left(\sum_{i=1}^N \|\tilde{m}_i + \tilde{\mu}\|^2\right)^2}\end{aligned} \tag{2}$$

Notice that terms involving other label maps ($\tilde{m}_i$ for $i \neq j$) fall out and that the gradient of the mean contributes nothing. It remains to define $\nabla_{\mathbf{p}_j}\tilde{m}_j$. For the k^{th} element of the pose parameter vector $\mathbf{p}_j$, using the chain rule produces

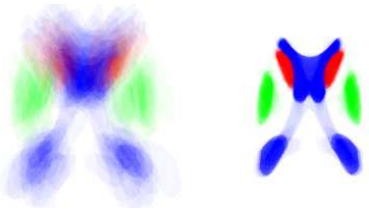
$$\nabla_{\mathbf{p}_j^k}\tilde{m}_j = \begin{bmatrix}\frac{\partial \tilde{m}_j}{\partial \tilde{x}} & \frac{\partial \tilde{m}_j}{\partial \tilde{y}} & 0\end{bmatrix} \frac{\partial T(\mathbf{p}_j)}{\partial \mathbf{p}_j^k} \begin{bmatrix}x\\ y\\ 1\end{bmatrix},$$

where $\frac{\partial T(\mathbf{p}_j)}{\partial \mathbf{p}_j^k}$ is computed for each pose parameter, for example,

$$\frac{\partial T(\mathbf{p}_j)}{\partial \mathbf{p}_j^1} = \frac{\partial T(\mathbf{p}_j)}{\partial x} = \frac{\partial M(x,y)}{\partial x}\ R(\theta)\ H(s_x, s_y)\ K(k).$$

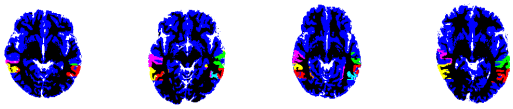
Matrix derivatives are taken componentwise, *e.g.*

$$\frac{\partial M(x,y)}{\partial x} = \begin{bmatrix}0 & 0 & 1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}.$$

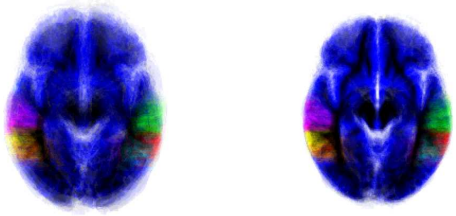


(a) Original (b) Aligned

Fig. 9. Alignment of a set of 30 maps used in the study by Tsai et al. [29]. The original and aligned sets are superimposed for visualization.



(a) Example maps from unaligned set



(b) Original (c) Aligned

Fig. 10. Alignment of a set of 33 maps with eight labels and background obtained from manual MRI segmentations. The original and aligned sets are superimposed for visualization.

Using a forward Euler scheme for gradient descent, in terms of $\nabla_{\mathbf{p}_j} d^2$, we have the update equation for pose parameter $\mathbf{p}_j$

$$\mathbf{p}_j^{t+1} = \mathbf{p}_j^t - \Delta t_{\mathbf{p}} \nabla_{\mathbf{p}_j} d^2,$$

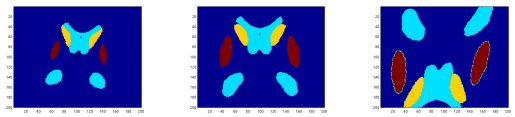
where t denotes the iteration number and $\Delta t_{\mathbf{p}}$ is the step size for updating $\mathbf{p}_j$. This jointly aligns the set of maps M while jointly aligning all labeled regions among the maps. Finally, gradient descent proceeds by repeated calculation of $\nabla_{\mathbf{p}_j} d^2$ and adjustment of $\mathbf{p}_j$ for each map in the set until convergence.

We then turned to verifying our method using the 2D data from the study by Tsai et al. [29], where a synthetic data set was generated using four labels. Figure 9 shows alignment on the entire data set.

Lastly, we performed registration using 2D maps obtained from expert manual segmentation of 33 patient MRI scans involving eight labels and background. Figure 10 shows examples from the original unaligned set as well as the superimposed maps after alignment.

V. Coarse Registration

While gradient descent is one of the most popular techniques for registration, it has three main drawbacks that we intend to address in this section: susceptibility to local minima, dependence on the starting point, and the difficulty in choosing the time step $\Delta t_{\mathbf{p}}$ (Equation IV). To address these problems, we propose a coarse registration strategy following which the gradient descent based algorithm typically converges in 10-20 iterations. In the proposed algorithm, we utilize information



(a) Image1 (b) Image2 (c) LS solution

Fig. 11. Aligning brain images with 4 labels. Right: Result of aligning using least squares solution.

already available in the labeled images. As a first step, compute the area centroid of each of the n labels in the two images we intend to register. Now we have n sets of corresponding points. It is very tempting to think that we can now directly recover the affine transformation relating the two images; however, we will now discuss why this is not the case.

A. What doesn't work

For a 2D image, if $n < 3$ (or $n < 4$ for 3D images), then we have an under-determined system with fewer variables (centroids) than unknowns (the parameters of the affine transformation). Hence, there are infinitely many satisfying solutions, requiring a least squares solution which often yields poor results. Assuming enough labels, if we use the least squares approach as detailed in [12], we still do not obtain satisfactory results as shown in Figure 11. This is clear when we register the centroids of the labels in Figure 11.

This result gives the minimum error in the least squares sense for aligning 4 points (centroids) in the images shown in Figure 11 instead of the shapes represented by the labels. Hence, we obtain an un-realistic estimate of the transformation matrix.

B. A Solution

We thus propose to use the second moments of each of the labels along with the centroids. While the centroids can be easily used to obtain the translation parameters, the covariance matrix for each of the labels can be used to obtain the scale, rotation and skew parameters. For the rest of the discussion, we will consider registering two 3D images. Thus, the covariance matrix essentially represents the corresponding label by an ellipsoid. For a given label l, let C_1 and C_2 represent the covariance matrices obtained from two images I_1 and I_2. The rotation matrix R that aligns these two ellipsoids can be obtained as follows:

$$[U\ S\ V] = \text{SVD}(C_1^T C_2);\ R = UV^T,$$

where SVD is the singular value decomposition of a matrix. Given angles θ_{xl}, θ_{yl} and θ_{zl}, the corresponding rotation matrix is given by

$$K = \begin{bmatrix} c_y c_z & -c_y s_z & s_y \\ c_z s_x s_y + c_x s_z & -s_z s_x s_y + c_x c_z & -s_x c_y \\ -c_z s_y c_x + s_x s_z & s_z s_y c_x + s_x c_z & c_x c_y \end{bmatrix},$$

where $c_i = \cos(\theta_{il}), s_i = \sin(\theta_{il}),\ i = \{x, y, z\}$. Equating R and K, we obtain the following formulae for recovering the

rotation angles:

$$\theta_{xl} = \arctan\{\frac{-R_{23}}{R_{33}}\},\ \theta_{yl} = \arctan\{\frac{R_{13}c_1}{R_{33}}\}$$
$$\theta_{zl} = \arctan\{\frac{-R_{12}}{R_{11}}\}, \quad (3)$$

where R_{ij} is the i^{th} row and j^{th} column of R. The above equation can be used to obtain the rotation angles for each of the labels in the two images about their corresponding centroids. The final rotation for the image about its centroid is calculated as a weighted sum of the individual rotations:

$$\tilde{\theta}_i = \sum_{l=1}^{n} w_j \theta_{il},\ i \in \{x, y, z\}$$

where the weights w_j can be specified by the user. A byproduct of this scheme is the flexibility to assign higher weights to certain labels (giving them more importance) than others. This is useful, for example, in labeled brain images where more importance can be given to aligning subcortical structures like caudate nucleus, hippocampus etc versus cortical structures with lots of cortical folds and inherently complex variability.

The scaling parameters can be obtained from the length of the axes of the ellipsoids. Let A_{i1} and A_{i2} be the length of the axes of the ellipsoid represented by covariance matrices C_1 and C_2. Then the scaling parameters in each of x, y and z-direction for label l can be calculated using

$$s_{il} = \sqrt{\frac{A_{i1}}{A_{i2}}},\ i \in \{x, y, z\}.$$

Once again, the overall scaling for the image can be calculated as a weighted sum of all labels:

$$\tilde{s}_i = \sum_{l=1}^{n} w_j s_{il},\ i \in \{x, y, z\}.$$

Using the co-ordinates of the centroids $c_{1l} = [x_1\ y_1\ z_1]$ and $c_{2l} = [x_2\ y_2\ z_2]$ for label l, one can obtain the following expression to calculate the skew parameters:

$$k_{xl} = \frac{x_2 - x_1}{y_1 + z_1},\ k_{yl} = \frac{y_2 - y_1}{x_1 + z_1},\ k_{zl} = \frac{z_2 - z_1}{x_1 + y_1}.$$

The skew parameters for the entire image can then be calculated using weighted sums:

$$\tilde{k}_i = \sum_{l=1}^{n} w_l k_{il},\ i \in \{x, y, z\}.$$

Using the above analysis, one can easily recover the affine transformation parameters that relate two richly labeled images I_1 and I_2. The question still remains unanswered as to how to register a group of labeled images to their intrinsic mean, when the mean itself is not known. We address this issue in the next section.

C. Aligning to the Mean

Our goal in this section is to build a probabilistic atlas from a group of labeled images $\{I_1, I_2, ..., I_N\}$. One way to build an unbiased atlas is to align all the images to their intrinsic mean. Most of the existing algorithms explicitly compute the mean image and iteratively refine the estimate at each iteration. In this work, we depart from this methodology and directly estimate the affine transformation T_i that takes image I_i to the group mean $\tilde{I}$ without explicitly computing the mean.

We demonstrate the entire algorithm by a simple example where we align three images $\{I_1, I_2, I_3\}$ to their mean. The technique however works for any N. As a first step, we coarsely align each image to every other image in the group using the method described earlier in section V-B. We thus have an affine transformation that relates I_1 to I_2 and I_3. Let x_{12},x_{13}, x_{23} be x-translation that aligns $I_1 \leftrightarrow I_2$, $I_1 \leftrightarrow I_3$ and $I_2 \leftrightarrow I_3$ respectively. Then, at iteration 1, we calculate the average x-translation $\tilde{x}_{im}$ required to move image i to the group mean μ:

Step 1. x-translation

	I_1	I_2	I_3	$\tilde{x}_{im}^{(1)}$
I_1	0	x_{12}	x_{13}	$\tilde{x}_{1m}^{(1)} = 0.5(x_{12} + x_{13})$
I_2	$-x_{12}$	0	x_{23}	$\tilde{x}_{2m}^{(1)} = 0.5(-x_{12} + x_{23})$
I_3	$-x_{13}$	$-x_{23}$	0	$\tilde{x}_{3m}^{(1)} = 0.5(-x_{13} - x_{23})$

In the second step, we assume that this average x-translation has been applied to each image, which allows us to update the table as follows:

Step 2. x-translation

	I_1	I_2	I_3
I_1	0	$x_{12}^{(1)} = x_{12} - \tilde{x}_{1m}^{(1)} - \tilde{x}_{2m}^{(1)}$	$x_{13}^{(1)} = x_{13} - \tilde{x}_{1m}^{(1)} - \tilde{x}_{3m}^{(1)}$
I_2	$-x_{12}^{(1)}$	0	$x_{23}^{(1)} = x_{23} - \tilde{x}_{2m}^{(1)} - \tilde{x}_{3m}^{(1)}$
I_3	$-x_{13}^{(1)}$	$-x_{23}^{(1)}$	0

The final x-translation that takes any image I_i to the mean image μ is then given by $\tilde{x}_{im}^{(1)} + \tilde{x}_{im}^{(2)} + ... + \tilde{x}_{im}^{(t)}$, for which a closed form expression can be computed as shown in Appendix A. The main intuition behind the procedure outlined above is to find the average x-translation required to align all the images by explicitly manipulating the x-translation parameters without actually applying the transformation to the images. The same technique is used to find all the other translation, rotation, skew and scale parameters. An important point to note is that, instead of using the scale parameters s_i directly, $\log(s_i)$ is used making operations of addition and subtraction valid on the space of scale parameters. An inverse exponential mapping can be used to obtain the actual scale values.

The final output of this procedure is a set of affine parameters that take every image I_i to the intrinsic group mean μ. This method of finding the intrinsic mean is not limited to labeled images, but can be used for gray-scale images if the affine transformation parameters that relate them are obtained using any suitable method. This is one of main contributions of this work.

D. A Faster Alternative

The method outlined above requires registering every image I_i to every other image I_j, $j \neq i$ of the group. This means, for a group of N images, we need to find $\frac{N(N-1)}{2}$

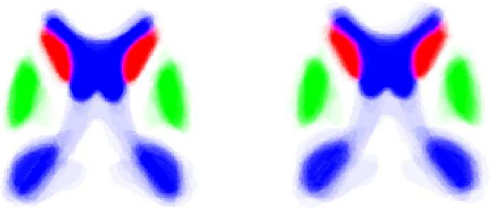


Fig. 12. Coarse Registration using full coarse registration (left) and using composition (right).

affine transformations. To reduce this computational load, a suboptimal strategy is to find only $N-1$ affine transformations A_{ij} that relate any image I_i to all the other images $I_j, j \in \{1,2,i-1,i+1,..N\}$, where we have arbitrarily fixed the image I_i. As described earlier, to compute the mean, we need affine transformations that relate every image to every other image. This can be achieved by using composition of transformations as follows: Let $i=1$ and assume we have computed A_{1j} for all j. Now, we can compute A_{23} using $A_{23} = A_{12}^{-1}A_{13}$, $A_{34} = A_{13}^{-1}A_{14}$ and so on. This strategy provides a very fast alternative to the method presented in the previous section, but it is suboptimal since the errors get accumulated while computing all the other transformations using compositions (see Table 2 for comparison). In our experiments, this faster method still gave a good starting point for the gradient descent based algorithm.

Figure 12 shows registration of a set 30 2D labeled images using both the coarse strategies outlined above.

VI. Experiments with 3D images

A. Registration

To test the label space representation and the proposed coarse registration techniques, we aligned a set of 3D labeled brain images. In our first experiment, we aligned a set of 20 images each containing 4 labels (grey matter, white matter, cerebrospinal fluid and skull of brain) plus background, a few slices of which are shown in Figure 13. The surface corresponding to the mean of the registered images is shown in Figure 14. In this experiment, the input images contained very noisy labeling of the skull (see Figure 13). To minimize their effect on the overall registration, we assigned a very low weight to this label in our coarse registration strategy which is quite a useful feature in such scenarios. We computed the total number of misaligned voxels with the group mean as a quantitative measure to assess the efficacy of the proposed algorithms. The mathematical formulation of this error metric is given below:

$$Er = \frac{\sum_{i=1}^{N} \mathcal{H}(I_i - \mu)}{\sum_{i=1}^{N} V(I_i)}, \tag{4}$$

where, $\mathcal{H}$ is a function which measures the total number of voxels that do not have the same label, $V(I_i)$ is the volume in voxels of the image I_i and μ is the calculated group mean of the label maps. Other error metrics used in the literature can be found in [8].

In our second experiment, we registered 10 images each

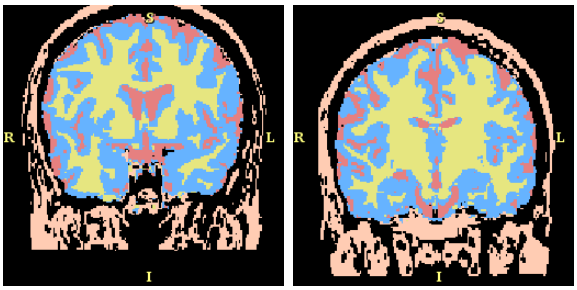


Fig. 13. Experiment 1: Slices from the brain images to be aligned

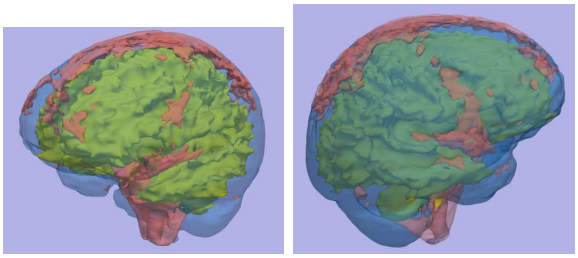


Fig. 14. Different view of the rendering of labels of the intrinsic mean image.

containing 14 labels (see Figure 15). In order to reduce the computational load, we implemented a scale-space based registration algorithm, where the coarse registration was performed on low resolution images while gradient descent was done on the original image resolution. While computing the lower scale images, the important property of label space of not being biased to a particular label as described earlier comes in quite handy by not introducing spurious labels in random places.

Figure 16 shows the 3D rendering of the mean image registered after doing the coarse registration followed by gradient descent. The following table gives the total number of misaligned voxels from the mean image (as a percentage of the total volume) for each of the two data sets.

Table 2. Misaligned voxels Er for all methods.

Data Set	GD alone	CR+GD	CRC + GD
Experiment 1	11.71%	7.95%	8.27%
Experiment 2	7.62%	6.52%	6.60%

where, GD = gradient descent, CR + GD = coarse registration followed by gradient descent, CRC + GD = coarse registration with composition followed by gradient descent.

For the data set in experiment 1, it took 78 iterations for GD algorithm (4 days of computational time on a typical fast machine), 7 iterations for CR+GD (48 hours of computational time) and 15 iterations for CRC+GD (24 hours of computational time) for convergence. In Experiment 2, it took only 8 iterations for CR+GD, 16 iterations for convergence with CRC+GD and 70 iterations with GD alone. From both these experiments (see Table 2), it is clear that using the coarse registration methods described in this work allows for better registration of the label maps than a pure gradient descent based approach.

Further, to test the dependence of CRC+GD on the choice of the initial reference image, we used 10 different reference images and computed the Misaligned voxels Er from

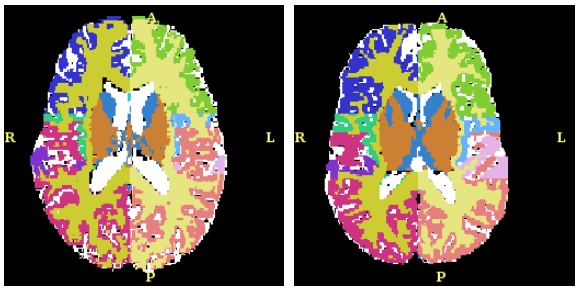


Fig. 15. Experiment 2:Slices from the brain images to be aligned

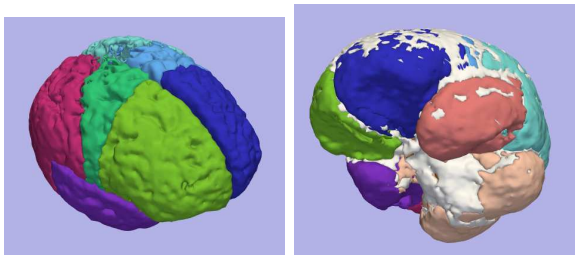


Fig. 16. Rendering of labels of the intrinsic mean image. Left: The different lobes of the brain. Right: All the 14 labels of the mean image.

the group mean. The following table gives the average and standard deviation (taken over 10 initializations) Er for each of the experiments:

Table 3. Misaligned voxels Er for CRC+GD.

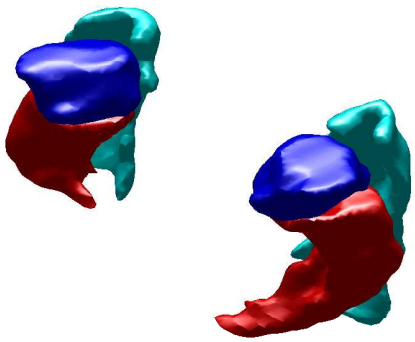
Data Set	CRC + GD
Experiment 1	8.31 ± 0.12 %
Experiment 2	6.57 ± 0.04 %

From the table, it is clear that the choice of initial reference image has a very limited impact on the final outcome (the group mean is very similar in all cases).

B. Multi-label Shape Learning using PCA

In this experiment, we examine a set of 31 patient brains with manually segmented amygdala, hippocampus, and parahippocampus.[1] Figure 17 shows these subcortical structures. Using label space as the underlying representation to eliminate bias, we performed affine registration [29].

[1]Data obtained from the NAMIC data repository of the Brigham and Women's Hospital, Boston.



(a) Subcortical structures

Fig. 17. Manually segmented amygdala, hippocampus, and parahippocampus *(blue, red, cyan)*.

To demonstrate the detail maintained in label space probabilistic calculations, Figure 18 shows the conditional probabilities for the amygdala and hippocampus of the first three patients, *i.e*. the probability of the amygdala given the probability of the amygdala in the other patients. For visualization, one slice is taken that passes through both regions. These probabilities were computed via the formula in Section III. Conditional probabilities like this may be used to judge inter-rater segmentation. Notice the sharp tail of the hippocampus, a feature which is easily lost in the intermediate representation used in the logarithm-of-odds approach.

We then compared label space against both the SDMs and binary vectors for use in linear principal component analysis (PCA); see [29, 18] and the references therein for a description of this technique. We chose to separate out a test patient, compute the PCA basis on the remaining patients, and examine the projection of the test patient onto that learned basis (leave-one-out test). This is standard in PCA modeling and is also referred to as the generalization metric [11, 27]. As a side note, we used the eigenvectors representing 99% of the variation in the data.

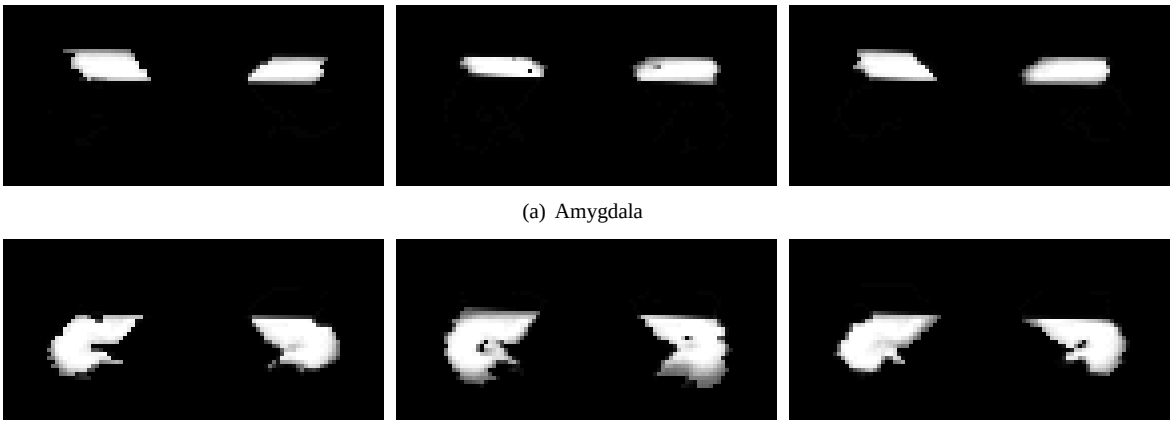
We employed the Dice coefficient as a measure of percent overlap between the test map M and its projection $\hat{M}$. The Dice coefficient is defined as the amount of overlapping volume divided by the average total volume: $D(M, \hat{M}) = 2|M \cap \hat{M}|/(|M| + |\hat{M}|)$, where $|\cdot|$ denotes volume. Table 4 gives the mean and variance of the Dice coefficient for each tissue class[2]. We found that label space is better than SDMs because of the inherent low per pixel variance and also better than binary vectors because it can represent unbiased label mixtures; SDMs and binary vectors require thresholding to determine label ownership. Note that the parahippocampus results improve significantly when label uncertainty is taken into account. Figure 19 shows the "generalization", "compactness" and "specificity" of the PCA modeling space as described in [11, 27]. Generalization is defined as the ability to describe previously unseen shapes (see Table V). Compactness is the ability to use minimal set of parameters to describe the entire shape and specificity is the ability to describe only valid instances of the shape. For all the three shapes discussed in this paper, the label space (black) does better than binary maps for all the three measures.

VII. Discussion

This paper describes label space, a coupled multi-object implicit representation borrowed from the fuzzy classification literature. For this representation, we demonstrated that algebraic operations may be done directly, label uncertainty is expressed equivalently as a weighted mixture of labels or in a probabilistic manner, and interpolation and smoothing is unbiased toward any label or the background. Further, the present representation can be used within the LogOdds framework as well to obtain a different probabilistic interpretation.

In order to avoid some of the inherent drawbacks of gradient descent based techniques, namely, getting stuck in local minima, we proposed two coarse registration methods

[2]Data reported as: *mean* $\pm$ *variance*



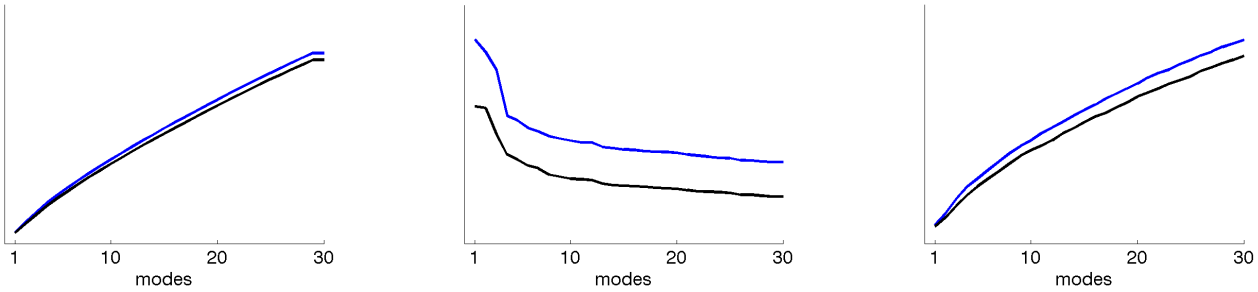
(a) Amygdala

(b) Hippocampus

Fig. 18. For one slice, conditional probability of the presence of amygdala and hippocampus of the first three patients given the other patients *(white indicates high probability)*. Notice the fine detail of the hippocampus, features which may be lost when using smoothed intermediate representations [23].

TABLE V
DICE COEFFICIENT OF PROJECTION USING PCA

	SDMs	Binary vectors	Label space
Amygdala	0.783 ± 0.0031	0.825 ± 0.0014	0.855 ± 0.0004
Hippocampus	0.782 ± 0.0016	0.819 ± 0.0006	0.843 ± 0.0003
Parahippocampus	0.494 ± 0.0033	0.561 ± 0.0017	0.773 ± 0.0003



(a) Compactness (b) Generalization (c) Specificity

Fig. 19. Compactness, Generalization and Specificity of the PCA modeling space. Label Space (black) and binary maps (blue). Label Space does better for all the three measures.

following which the gradient descent method converges to a better minimum implying better registration. We demonstrate this using a few examples on 3D data sets. A new method to compute the intrinsic mean of a set of labeled images is also proposed. This method can be used to coarsely align any type of images (not just labeled) if the affine parameters that relate them are known.

The coarse registration method also allows assigning different weights to each of the labels which is another useful property to have if there is prior knowledge about the robustness of the segmented labels. Further, the method can be easily extended to use the poly-affine model of [2] for non-rigid registration of label maps.

Nevertheless, modeling shapes in label space does have its limitations. One major drawback to label space is the spatial demand. It might be possible to extend the hypersphere representation [3] to avoid interpolation issues (see Figure 4) by taking into consideration the empirical presence of neighbor pairings when determining vertex distribution.

A different limitation of the coarse registration strategy is that it is dependent upon the accuracy of the first two area moments of the labels. Hence, if there are large outliers in the image, it can result in estimating the wrong affine transformation for that image.

VIII. Proof of Convergence

Here, we prove the convergence of the iterative technique described in Section V-C for coarse registration. For the sake of simplicity, let us assume $N = 3$, i.e., we are registering 3 images. At the end of iteration 1, we have computed $x_{12}^{(1)}, x_{13}^{(1)}, x_{23}^{(1)}$. Further, we also know $\tilde{x}_{im}^{(1)}$. Another important property that holds is

$$\tilde{x}_{1m}^{(1)} + \tilde{x}_{2m}^{(1)} + \tilde{x}_{3m}^{(1)} = 0. \tag{5}$$

This can be easily verified from the table in Step 1.

Now, let us compute $\tilde{x}_{1m}^{(2)}$:

$$\begin{aligned}\tilde{x}_{1m}^{(2)} &= 0.5\left(x_{12}^{(1)} - \tilde{x}_{1m}^{(1)} - \tilde{x}_{2m}^{(1)} + x_{13}^{(1)} - \tilde{x}_{2m}^{(1)} - \tilde{x}_{3m}^{(1)}\right) \\ &= 0.5\left(\tilde{x}_{2m}^{(1)} + \tilde{x}_{3m}^{(1)}\right) \\ &= -0.5\tilde{x}_{1m}^{(1)},\end{aligned} \tag{6}$$

where the last equality follows from (5). Thus, $\tilde{x}_{1m}^{(3)} = 0.5\tilde{x}_{1m}^{(2)}$ and so on. Indeed, one can directly compute the sum of such a converging series: $\sum_{t=1}^{\infty} x_{1m}^{(t)} = x_{1m}^{(1)} - x_{1m}^{(1)}\frac{1}{1+2}$. Generalizing this for N images, we get the following closed form expression for the final translation for each image:

$$\begin{aligned}\sum_{t=1}^{\infty} x_{im}^{(t)} &= x_{im}^{(1)} + \\ &\quad x_{im}^{(1)} \sum_{t=1}^{\infty}\left(-\frac{1}{(N-1)^{(2t-1)}} + \frac{1}{(N-1)^{(2t)}}\right) \\ &= x_{im}^{(1)} - x_{im}^{(1)}\frac{1}{1+N-1} \\ &= x_{im}^{(1)} - x_{im}^{(1)}\frac{1}{N}\end{aligned} \tag{7}$$

One obtains similar expressions for computing all the translation, rotation and skew parameters. For the scale parameter, one has to take the logarithm, compute the average scale in log space and map back using the exponential.

Acknowledgement

The authors would like to thank Padmapriya Srinivasan for help with the 3D simulations and Marc Neithammer for interesting discussions on this topic. This work was supported in part by a Department of Veteran Affairs Merit Award (Dr. M Shenton, Dr. R McCarley), the VA Schizophrenia Center Grant (RM, MS) and NIH grants:P41 RR13218 (MS), K05 MH 070047 (MS), 1P50MH080272-01 (MS), R01 MH 52807 (RM), R01 MH 50740 (MS) and NA-MIC (NIH) grant U54 GM072977-01 (Dr. Ron Kikinis).

References

[1] H. Abd and A. Farag. Shape representation and registration using vector distance functions. In *IEEE. Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[2] V. Arsigny, X. Pennec, and N. Ayache. Polyrigid and polyaffine transformations: a novel geometrical tool to deal with non-rigid deformations - application to the registration of histological slices. *Medical Image Analysis*, 9:507–523, 2005.

[3] K. Babalola and T. Cootes. Groupwise registration of richly labeled images. In *Medical Image Analysis and Understanding*, 2006.

[4] K. Babalola and T. Cootes. Registering richly labeled 3D images. In *Proc. of the Int. Symp. on Biomedical Images*, pages 868–871, 2006.

[5] JC Bezdek, LO Hall, MC Clark, D.B. Goldgof, and LP Clarke. Medical image analysis with fuzzy models. *Statistical Methods in Medical Research*, 6(3):191, 1997.

[6] C. Brechbuhler, G. Gerig, and O. Kubler. Parametrization of closed surfaces for 3-D shape description. *Computer Vision and Image Understanding*, 61(2):154–170, 1995.

[7] T. Cootes and C. Taylor. Combining point distribution models with shape models based on finite element analysis. *Image and Visual Computing*, 13(5):403–409, 1995.

[8] WR Crum, O. Camara, and DLG Hill. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transactions on Medical Imaging*, 25(11):1451–1461, 2006.

[9] S. Dambreville, Y. Rathi, and A. Tannen. Shape-Based Approach to Robust Image Segmentation using Kernel PCA. In *IEEE. Computer Vision and Pattern Recognition*, pages 977–984, 2006.

[10] S. Dambreville, Y. Rathi, and A. Tannenbaum. A shape-based approach to robust image segmentation. *Lecture Notes in Computer Science*, 4141:173, 2006.

[11] R.H. Davies, C.J. Twining, P.D. Allen, T.F. Cootes, and C.J. Taylor. Shape discrimination in the hippocampus using an MDL model. *Lecture Notes in Computer Science*, pages 38–50, 2003.

[12] I. L. Dryden. General shape and registration analysis. In O. Barndorff-Nielsen, W. S. Kendall, and M. N. M. van Lieshout, editors, *Stochastic Geometry: likelihood and computation*, pages 333–364, London, 1999. Chapman and Hall.

[13] E. DAgostino, F. Maes, D. Vandermeulen, and P. Suetens. An information theoretic approach for non-rigid image registration using voxel class probabilities. *Medical Image Analysis*, 10(3):413–431, 2006.

[14] P. Golland, WE. Grimson, M. Shenton, and R. Kikinis. Detection and analysis of statistical differences in anatomical shape. *Medical Image Analysis*, 9(1):69–86, 2005.

[15] S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:151–160, 2004.

[16] A. Kelemen, G. Székely, and G. Gerig. Three-dimensional model-based segmentation. In *Int. Workshop on Model Based 3D Image Analysis*, pages 4–13, 1998.

[17] JT Kent and KV Mardia. Spatial classification using fuzzy membership models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5):659–671, 1988.

[18] M. Leventon, E. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *IEEE. Computer Vision and Pattern Recognition*, pages 1316–1324, 2000.

[19] P. Lorenzen, M. Prastawa, B. Davis, G. Gerig, E. Bullitt, and S. Joshi. Multi-modal image set registration and atlas formation. *Medical image analysis*, 10(3):440–451, 2006.

[20] E. Miller, N. Matsakis, and P. Viola. Learning from one example through shared densities on transforms. In *IEEE. Computer Vision and Pattern Recognition*, pages 1464–1471, 2000.

[21] D. Nain, S. Haker, A. Bobick, and A. Tannenbaum. Multiscale 3-d shape representation and segmentation using spherical wavelets. *Trans. on Medical Imaging*, 26(4), 2007.

[22] S. Pizer, G. Gerig, S. Joshi, and S. Aylward. Multiscale medial shape-based analysis of image objects. In *Emerging Medical Imaging Technology*, volume 91, 2003.

[23] K. Pohl, J. Fisher, S. Bouix, M. Shenton, R. McCarley, W. Grimson, R. Kikinis, and W. Wells. Using the logarithm of odds to define a vector space on probabilistic atlases. *Medical Image Analysis*, 11(5):465–477, 2007.

[24] T. Rohlfing, D.B. Russakoff, M.J. Murphy, and C.R. Maurer Jr. An intensity-based registration algorithm for probabilistic images and its application for 2-D to 3-D image registration. In *Proc. SPIE*, volume 4683, pages 581–591, 2002.

[25] ME Shenton, R. Kikinis, FA Jolesz, SD Pollak, M. LeMay, CG Wible, H. Hokama, J. Martin, D. Metcalf, M. Coleman, et al. Abnormalities of the left temporal lobe and thought disorder in schizophrenia. A quantitative magnetic resonance imaging study. *New England Journal of Medicine*, 327(9):604–612, 1992.

[26] M. Styner, J. Lieberman, D. Pantazis, and G. Gerig. Boundary and medial shape analysis of the hippocampus in schizophrenia. *Medical Image Analysis*, 8(3), 2004.

[27] M.A. Styner, K.T. Rajamani, L.P. Nolte, G. Zsemlye, G. Szekely, C.J. Taylor, and R.H. Davies. Evaluation of 3D correspondence methods for model building. *Lecture Notes in Computer Science*, pages 63–75, 2003.

[28] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *Trans. on Medical Imaging*, 22(2), 2003.

[29] A. Tsai, W. Wells, C. Tempany, E. Grimson, and A. Willsky. Mutual information in coupled multi-shape model for medical image segmentation. *Medical Image Analysis*, 8(4), 2004.

[30] C.J. Twining, S. Marsland, and CJ Taylor. Groupwise non-rigid registration: The minimum description length approach. In *Proceedings of the British Machine Vision Conference (BMVC*, volume 1, pages 417–426, 2004.

[31] S.K. Warfield, J. Rexilius, P.S. Huppi, T.E. Inder, E.G. Miller, WM Wells, G.P. Zientara, F.A. Jolesz, and R. Kikinis. A binary entropy measure to assess nonrigid registration algorithms. *Lecture Notes in Computer Science*, pages 266–274, 2001.

[32] L. Zollei, E. Learned-Miller, E. Grimson, and W. Wells. Efficient population registration of 3D data. *Lecture Notes in Computer Science*, 3765:291–301, 2005.