

Fast Shape-Based Nearest-Neighbor Search for Brain MRIs Using Hierarchical Feature Matching

Peihong Zhu, Suyash P. Awate, Samuel Gerber, and Ross Whitaker

Scientific Computing and Imaging Institute, University of Utah, USA

Abstract. This paper presents a fast method for quantifying shape differences/similarities between pairs of magnetic resonance (MR) brain images. Most shape comparisons in the literature require some kind of deformable registration or identification of exact correspondences. The proposed approach relies on an optimal matching of a large collection of features, using a very fast, hierarchical method from the literature, called *spatial pyramid matching* (SPM). This paper shows that edge-based image features in combination with SPM results in a fast similarity measure that captures relevant anatomical information in brain MRI. We present extensive comparisons against known methods for shape-based, k -nearest-neighbor lookup to evaluate the performance of the proposed method. Finally, we show that the method compares favorably with more computation-intensive methods in the construction of local atlases for use in brain MR image segmentation.

1 Introduction

Large collections of medical images are becoming ubiquitous as public resources, and within specific clinical practices. Currently, large studies consist of thousands of images, but in the coming years databases of images of various types will grow to tens of thousands. The availability of such data demands new techniques for image analysis and associated algorithms that are able to efficiently take advantage of these large collections. We begin with a brief discussion of the kinds of algorithms that utilize these large sets of images and how these algorithms demand new technologies for fast image lookup, and end this section with a discussion of how the proposed method addresses this challenge.

One use for a large collection of medical images is to aid in segmentation or tissue classification. *Atlases*, comprising voxel-wise tissue probabilities, incorporate information about spatial location of biological structures. Atlases with hard/soft tissue/object assignments can be used alone, or as “priors”, for segmentation and combined with voxel measurements of a specific image to generate label maps in previously unseen images. Atlases are typically constructed by summarizing information concerning image intensities and anatomical shapes from a training set that includes manual segmentations. The information is summarized in (i) an average image (template) and (ii) a tissue probability map. To segment a test image, the template is warped to the test image and the tissue probabilities in the atlas are then transferred to the test image.

Recent work [1,11,6] shows that there can be significant loss of information in the averaging/summarizing process underlying conventional atlas construction schemes. Alternative strategies rely on *multiatlases* or *nonparametric atlases*. Such schemes consider every image (and associated manual segmentation) in the training set as an atlas. They transfer information from the training set to the test subject by independently warping atlas templates to the latter and, subsequently, warping atlas segmentations to the latter. An important aspect of multi or nonparametric atlases is that one typically selects a small subset of atlases from the training set, comprising the atlas templates that are *most similar* to the test image. Lotjonen et al. [11] use a weighted average of these images, where weights reduce monotonically with *distance*, thus forming a nonparametric estimate of segmentations in the space of images. A small collection of similar images typically registers quite well and results in a *crisper* average, thus providing better quality segmentations compared with those obtained using a summary/average atlas from the entire training set. This has been demonstrated on brain magnetic resonance (MR) images [1] and cardiac MR angiography images [6]. Despite these impressive developments in nonparametric atlases, finding images representing similar anatomy requires registration between the test image and (potentially) every image in the training set. Thus, the problem of efficiently applying these methods to very large training sets remains open.

There are other types of analyses demanding some kind of fast nearest-neighbor (NN) lookup on sets of anatomical images. For instance, Gerber et al. [7] propose a method to learn the underlying manifold structure on sets of brain images, and rely on comparisons (deformable registrations) of each image against every other, resulting in many days of computation time—but ultimately they use only NN relationships to construct the manifold. Wolz et al. [13] similarly use NN relationships to extract biomarkers from a low-dimensional representation under the hypothesis that the low-dimensional coordinates capture information about shape and appearance. Other approaches for learning representations on sets of images exist [2,5], but they share the same need for quickly identifying and quantifying shape-based relationships between similar images.

The problem of fast searching in image spaces has received a lot of attention in computer vision for content-based image retrieval and recognition. In this context, Grauman and Darrell [8] proposed an efficient hierarchical approximation for measuring similarity between histograms of features using a *pyramid match kernel* (PMK). While [8] applies PMK as bags of image features in which spatial location is typically ignored, Lazebnik et al. [10] proposed a variation on the PMK approach, called *spatial pyramid matching* (SPM), which computes approximate geometric correspondences based on histograms of feature locations. They apply SPM primarily to content-based image retrieval, where relatively coarse approximations to feature localization and matching are justified. However, these methods *do* offer bounded-error approximations of L1, bipartite feature matching—raising the possibility that they would have utility in evaluating anatomical shape similarity. This observation is supported, in part, by medical image registration algorithms that rely on hierarchical feature matching [12].

The purpose of this paper is to examine the application of a fast, hierarchical approach for matching image features to the problem of shape-based lookups into large databases of medical images. The result is a novel method for fast, shape-based, approximate search of medical images. The proposed method is several orders of magnitude faster than typical 3D registration methods, thereby making NN searches practical for large databases of medical images. We describe a mechanism by which MR brain images can be reduced to feature maps that, combined with the matching algorithms above, offer acceptable approximations to deformation distance. This paper also validates the accuracy of the fast searches relative to accuracies produced by registration methods. Furthermore, the paper presents results on the application of the proposed method for brain-tissue segmentation, where the results demonstrate segmentation performance as good or better than with registration-based metrics.

2 Methods

The proposed method for measuring the similarity between two brain MR images relies on hierarchical feature matching using SPM [10]. This section reviews the formulation for SPM and describes its application to shape comparison on brain MR images. The use of SPM on image features entails a pipeline of four steps: (i) image preprocessing—intensity and spatial normalization and edge-preserving filtering, (ii) feature extraction—Canny edge detection, (iii) feature labeling—data-driven codebook generation and label assignment, and (iv) SPM comparisons—construction and comparison of a collection of labeled, multilevel feature histograms. Steps (i)–(ii) rely on public domain implementations (i.e. ITK) for affine registration, intensity normalization, filtering, and edge detection that are well known from the literature. We describe steps (iii) and (iv), next.

Feature Extraction and Codebook Construction: The proposed method relies on edges to capture anatomical shape. SPM matches a set of feature maps in a way that approximates L1, bipartite matching, but does not enforce any kind of smoothness on the matching, which is typically important for medical image registration. In order to avoid mismatching nearby edges associated with different anatomies, we classify edges into different types based on local shape. For edge classification, we use the orientation and curvature of the level sets of the input image at each edge point, which are computed by first- and second-order derivatives of Gaussians. Then each edge is classified by (i) a clustering process on the orientation-curvature features extracted at all edge voxels followed by (ii) a quantization/coding process that assigns a label to every edge voxel based on the cluster center to which it is nearest. We use k -means for clustering and refer to cluster centers as the *codebook* of possible edge patterns $\mathcal{C} = \{c_1, \dots, c_k\}$ in brain MR images. Figure 1 shows examples of the hard-coded edge maps. To alleviate errors/artifacts in the hard quantization/coding process, we assign “soft” labels as in fuzzy-c-means, where each edge point has a set of membership values of belonging to every cluster.

Shape Similarity using Spatial Pyramid Matching: Each edge type in the codebook is considered as a different feature map, and each feature map is represented as a multilevel spatial histogram, or spatial *pyramid*. For each edge type SPM approximates the distance between nearby edges by the *overlap* of the spatial histograms of the edge images at each level. Features in two different pyramids, are “matched” if they lie in the same bin, at a specific level in the pyramid. The algorithm assigns matches starting from the finest level and proceeds progressively through coarser levels—to a final single bin, where any remaining features are matched. Once labels get matched at a specific level of the pyramid, they are not considered for matches at coarser levels. This pyramid matching process is essentially an approximate, greedy algorithm for bipartite matching between labeled point sets. For each feature match at each level there is an associated weight that assigns a cost (or reward) for that match. To approximate L1 matching, one would use a cost proportional to the bin size at each level in the hierarchy, and sum over each bin at all levels. This is, however, sensitive to mismatches, outliers, and inherent inaccuracies in binning (worse at coarser levels). As an alternative, Grauman and Darrell [8] assign weights that are the reciprocal of the bin size (i.e. rewarding finer-level matches) and show that the resulting similarity measure is a Mercer kernel.

Given two images A and B of a particular edge type $c \in \mathcal{C}$, consider their spatial pyramid representations h_A and h_B , with L levels in order of increasing resolution. Each level l is a spatial histogram with 2^{l-1} bin along each image dimension d , with a total of $M_l = 2^{d(l-1)}$ bins. Let h_A^l and h_B^l be the representations of A and B , respectively, at level l with $h_A^l(i)$ and $h_B^l(i)$ the *soft* count of edges of type c in the i th bin. Then, the number of matches between h_A^l and h_B^l is given by the *histogram intersection*:

$$I(h_A^l, h_B^l) = \sum_{i=1}^{M_l} \min(h_A^l(i), h_B^l(i)). \quad (1)$$

Given the matches that have already occurred at levels $[l+1, L]$ finer than l , the *new matches* occurring at level $l < L$ are: $N_l = I(h_A^l, h_B^l) - I(h_A^{l+1}, h_B^{l+1})$. At the finest level L , all the matches $I(h_A^L, h_B^L)$ are new: $N_L = I(h_A^L, h_B^L)$. If h_A and h_B are the same, then all (new) matches occur at the finest level L .

Similarity between point-sets A and B is measured using the PMK,

$$\kappa(A, B) = \sum_{l=1}^L w_l N_l = I(h_A^L, h_B^L) + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}} (I(h_A^l, h_B^l) - I(h_A^{l+1}, h_B^{l+1})), \quad (2)$$

which is a weighted combination of the new matches at each level. The matches at finer levels are given higher weights because matches at finer levels indicate lower separation distance between the matched points. Specifically, the weight w_l , for matches at level l , is $w_l = 1/(2^{L-l})$ that decreases exponentially with level coarseness; this rate is consistent with the exponential increase of the bin size with level coarseness. For the finest level, $w_L = 1$.

To ensure a maximum PMK similarity $\kappa(\cdot, \cdot)$ of 1, $\kappa(\cdot, \cdot)$ is normalized as $\tilde{\kappa}(A, B) = \kappa(A, B) / \sqrt{\kappa(A, A)\kappa(B, B)}$. Then, **SPM similarity** between two

images is the sum of normalized PMK similarities for each edge type: $S(A, B) = \sum_{c=1}^k \tilde{\kappa}(A_c, B_c)$, where $\tilde{\kappa}(A_c, B_c)$ is the normalized PMK similarity for the c th code. SPM's algorithmic complexity is linear in the point-set cardinality, number of pyramid levels, and number of codes.

3 Evaluation Methodology

The SPM approach is compared to diffeomorphic registration (LDDMM) [4], which captures geodesic distances of deformation fields, and an elastic registration approach (not strictly a metric) as employed in [7]. This is not an exhaustive list of metrics for registration-based shape comparison, but is representative of the state of the art in terms of compute time and performance. Our hypothesis is that SPM is comparable to the various options for deformation metrics (different algorithms and parameters) in the evaluation of shape differences and k -NNs.

Consider a training set of brain images $\mathcal{B} = \{B_1, \dots, B_M\}$ and a test set $\mathcal{A} = \{A_1, \dots, A_N\}$. To compare k -NN performance, we find the k -NNs $B_j \in \mathcal{B}$ of all $A_k \in \mathcal{A}$ using (i) SPM with 2 different codebook sizes, (ii) elastic registration with 2 values of the standard parameter, say λ , which balances the weight between the image-match and deformation smoothness, and (iii) diffeomorphic registration with two λ values. Our evaluation metrics are:

1. *Accuracy* (π): For image A_i , let the k -NNs given by two different methods (among SPM, diffeomorphic registration, and elastic registration) be $\eta_S(A_i, k) \subset \mathcal{B}$ and $\eta_R(A_i, k) \subset \mathcal{B}$. Then, π for S relative to R is:

$$\pi = (1/N) \sum_{i=1}^N |\eta_R(A_i, k) \cap \eta_S(A_i, k)| / |\eta_R(A_i, k)|.$$
2. ϵ -ball radius ratio (γ): If images in \mathcal{B} are very similar to each other, there are many neighbors within a relatively small distance. Thus, we propose an ϵ -ball metric, from topology theory, to quantify the extra distance introduced by differences in k -NNs. Given the k -NNs $\eta_R(A_i, k)$, we find the smallest set $\eta_S(A_i, k^*)$ such that $\eta_R(A_i, k) \subset \eta_S(A_i, k^*)$. Note: $k^* \geq k$. Then,

$$\gamma = (1/N) \sum_{i=1}^N [\max_{B \in \eta_S(A_i, k^*)} d_R(A_i, B)] / [\max_{B \in \eta_R(A_i, k)} d_R(A_i, B)],$$
where $d_R(A_i, B)$ is the registration-based distance between A_i and B ; $\gamma \geq 1$.
3. To compare performance for atlas-based segmentation, we use the *Dice* overlap measure between an atlas-based tissue segmentation A and the ground-truth tissue segmentation G . The Dice overlap is: $(2|A \cap G|) / (|A| + |G|)$.

The evaluation employs: (i) a database \mathcal{B} of 259 skull stripped, T1-weighted brain MR images (normal humans; image size $160 \times 192 \times 176$ voxels) with probabilistic tissue segmentations and (ii) a database \mathcal{A} of 20 BrainWeb [3] images.

4 Results and Discussion

The main motivation of using SPM similarity is to enable *fast* searches for NNs. Thus, we begin by discussing the computational cost and times for matching tasks between two 3D brain images SPM, elastic registration, and diffeomorphic registration. Diffeomorphic registration between two brain MR images with 200^3

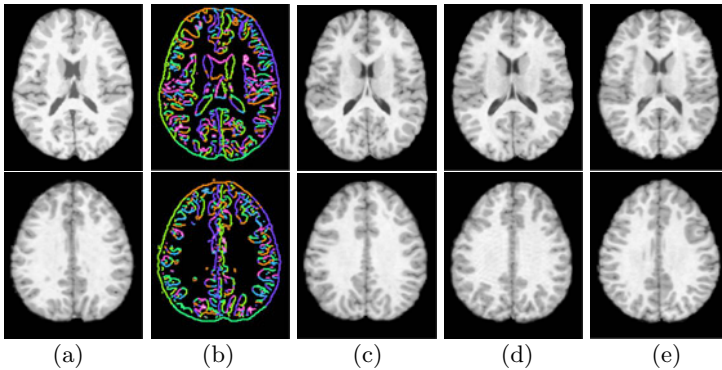


Fig. 1. (a) Axial slices from 2 examples of 3D T1w brain MR images in the 259-brain database (see text). (b) Coded edge maps associated with (a). (c)–(e) 3 nearest neighbors (NNs), for the images in (a), found by the proposed method. The NNs clearly reflect the overall shape of the skull, ventricles, etc. SPM similarities between the top and bottom row images are very small, reflecting the large differences in skull shape.

voxels, using LDDMM [4] with 100 iterations and 5 intermediate timepoints, requires roughly 8×10^{13} flops. An elastic registration requires approximately 50-times fewer, i.e. 2×10^{12} , flops. For the same example, SPM requires less than 10^8 flops. Typical registration methods have a high demand for trilinear interpolations, which can result in poor memory-access patterns. SPM, on the other hand, eliminates interpolations entirely. Codebook construction and edge labeling are computationally light and need to be performed just once, during preprocessing, and require computation of the same order as in preprocessing for registration. Run times on conventional, serial CPUs confirm, approximately, these calculations. A careful C++ CPU implementation of LDDMM, registration between two brain images requires around 4-6 hours, while a similar implementation of elastic registration requires roughly 10 minutes. A Matlab implementation of SPM takes roughly 40 seconds. The experiments in this paper used a GPU implementation [9] of LDDMM (roughly 10 minutes required). In real systems to be deployed one would implement SPM on a parallel architecture such as a GPU (a $100\times$ speedup over Matlab is expected). Such experiments becomes architecture specific and are beyond the scope of this paper.

Selecting 3500 random pairs, we compute SPM similarities and registration-based distances. Linear regression between: (i) registration-based distances provided by different methods gives slopes around 0.83, R^2 around 0.75; (ii) SPM similarity and registration-based distance gives slopes around -0.73 , R^2 around 0.6. This indicates that SPM similarity correlates very well with registration-based distances.

For accuracy we consider a subset of 100 from the 259 brain MR images. We compute SPM similarities and registration-based distances for all pairs in the dataset. Subsequently, we compute π for k -NN searches with $k = 10$ for every image in the dataset and average the values over the dataset (the standard deviation is around 0.13 for all results). Figure 2 indicates that (i) π for SPM, relative

		k -NN Accuracy						
	Diff ₁	Diff ₂	Elas ₁	Elas ₂	SPM ₁	SPM ₆	SPM ₁₈	
Diff ₁	1	0.39	0.22	0.35	0.25	0.32	0.32	
Diff ₂		1	0.51	0.69	0.45	0.53	0.53	
Elas ₁			1	0.45	0.36	0.36	0.36	
Elas ₂				1	0.42	0.52	0.53	
SPM ₁					1	0.56	0.52	
SPM ₆						1	0.86	
SPM ₁₈							1	

		Average ϵ -Ball Radius Ratio						
	Diff ₁	Diff ₂	Elas ₁	Elas ₂	SPM ₁	SPM ₆	SPM ₁₈	
Diff ₁	1	1.24	1.30	1.25	1.38	1.33	1.32	
Diff ₂	1.26	1	1.20	1.16	1.33	1.29	1.26	
Elas ₁	1.29	1.19	1	1.23	1.29	1.27	1.27	
Elast ₂	1.13	1.07	1.10	1	1.12	1.09	1.09	

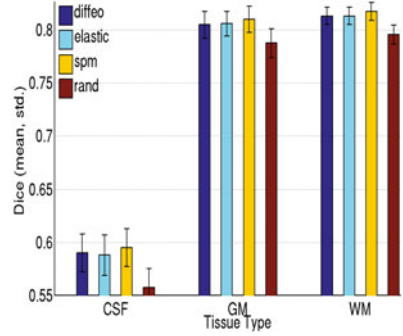


Fig. 2. [Top-Left] k -NN accuracy π for different parameter values (smoothness of warp) of diffeomorphic and elastic registration as well as different numbers of edge codes (subscripts) for SPM. [Bottom-Left] Average ϵ -ball radius ratio γ for different methods compared to reference distances listed in the first column. The first two rows use diffeomorphic registration to give the reference $d_R(\cdot, \cdot)$, while the last two use elastic registration. [Right] Dice overlaps for atlas-based segmentation of 20 BrainWeb images [3]. Atlases are constructed using 10 nearest neighbors selected by diffeomorphic registration, elastic registration, SPM, and random selection.

to each registration method, is very close to π measured for one registration method relative to another; and (ii) using codebooks, rather than simply edges, improves SPM’s performance. Figure 2 also shows that SPM’s performance is also robust respect to the size of the codebook. The second table in Figure 2 shows ϵ -ball radius ratios, indicating that SPM and registration-based methods misidentify k NNs in a way that introduces similar errors in image distances.

Next, the utility of SPM is evaluated for *atlas-driven segmentation* by comparing the quality of local k NN based atlases. Notice that SPM does *not provide a coordinate transformation*, thus after finding the k -NNs with each method, we use the LDDMM registration to warp the k -NN images (and segmentations) to the test image and average their tissue probabilities to construct the atlas. In order to isolate k -NN search performance, tissue values are assigned based only on the local atlas and no intensity based segmentation is used. We used 20 test images from the BrainWeb dataset [3] and build the k -NN atlases from the 259-brain image database. To evaluate the efficacy of each atlas we use the Dice overlap of segmentation with respect to ground truth. We average these Dice values over all 20 test brains. Figure 2 shows that Dice overlaps using SPM outperform the others for this dataset. Dice overlaps for elastic and diffeomorphic registration are very close to each other. Random selection of images reduces performance by several percent, which is consistent with findings in [6,1].

Conclusion: This paper demonstrates the effectiveness, in terms of both compute times and accuracy, of an SPM-based k -NN search for brain MR images. Some engineering decisions, such as limiting features to edges and coding based on orientation and curvature, are effective, but warrant further investigation. For instance, one might learn features directly from image patches, especially if

the images in question are not dominated by high-contrast edges. The question of generally applicability is also quite interesting. While the results for brain images are quite promising, successful demonstrations on other types of data, such as cardiac, have to be completed. However, the basic properties of the proposed framework, including a great deal of flexibility in possible features and robustness to feature outliers, promises to be more general.

Acknowledgements. This work is supported by the NIH/NCRR Center for Integrative Biomedical Computing - 2P41 RR0112553-12, and the NIH/NCBC National Alliance for Medical Image Computing - U54-EB005149.

References

1. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage* 46(3), 726–738 (2009)
2. Aljabar, P., Wolz, R., Srinivasan, L., Counsell, S., Boardman, J., Murgasova, M., Doria, V., Rutherford, M., Edwards, A., Hajnal, J., Rueckert, D.: Combining morphological information in a manifold learning framework: Application to neonatal MRI. In: Jiang, T., Navab, N., Pluim, J.P., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6363, pp. 1–8. Springer, Heidelberg (2010)
3. Aubert-Broche, B., Collins, D., Evans, A.: A new improved version of the realistic digital brain phantom. *Neuro Image* 32(1), 138–145 (2006)
4. Beg, F., Miller, M., Troune, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comp. Vis.* 61(2) (2005)
5. Chen, T., Rangarajan, A., Eisenschenk, S.J., Vemuri, B.C.: Construction of neuroanatomical shape complex atlas from 3D brain MRI. In: Jiang, T., Navab, N., Pluim, J.P., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6363, pp. 65–72. Springer, Heidelberg (2010)
6. Depa, M., Sabuncu, M.R., Holmvang, G., Nezafat, R., Schmidt, E.J., Golland, P.: Robust atlas-based segmentation of highly variable anatomy: Left atrium segmentation. In: Camara, O. (ed.) *MICCAI Workshop Stat. Atlases Comp. Models Heart*, pp. 1–8 (2010)
7. Gerber, S., Tasdizen, T., Fletcher, P., Joshi, S., Whitaker, R.: ADNI: Manifold modeling for brain population analysis. *Med. Imag. Analysis* 14(5), 643–653 (2010)
8. Grauman, K., Darrell, T.: The pyramid match kernel: Efficient learning with sets of image features. *J. Mach. Learn. Res.* 2, 725–760 (2007)
9. Ha, L., Kruger, J., Fletcher, T., Joshi, S., Silva, C.T.: Fast parallel unbiased diffeomorphic atlas construction on multi-graphics processing units. In: *Euro. Symp. Parallel Graph. Vis.*, pp. 65–72 (2009)
10. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: *IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 2169–2178 (2006)
11. Lotjonen, J., Wolz, R., Koikkalainen, J., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D.: ADNI: Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage* 49(3), 2352–2365 (2010)
12. Shen, D., Davatzikos, C.: HAMMER: Hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imag.* 21(11), 1421–1439 (2002)
13. Wolz, R., Aljabar, P., Hajnal, J.V., Rueckert, D.: Manifold learning for biomarker discovery in MR imaging. In: Wang, F., Yan, P., Suzuki, K., Shen, D. (eds.) *Conf. Mach. Learn. Med. Imag.*, pp. 116–123 (2010)