How are Siblings Similar? How Similar are Siblings? Large-scale Imaging Genetics using Local Image Features

Matthew Toews[1], William Wells[2,3]

[1] The Laboratory for Imagery, Vision and Artificial Intelligence, École de Technologie Supérieure Montréal, Canada (matthew.toews@etsmtl.ca)

[2] Brigham and Women's Hospital, Harvard Medical School

[3] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology Boston, US (sw@bwh.harvard.edu)

ABSTRACT

A novel method is developed for analyzing large sets of 3D medical image data. The image manifold is approximated as a proximity graph between bags-of-features (BoFs), i.e. order-less sets of local invariant features extracted from images. The pairwise geodesic distance between images (or BoFs) is approximated locally by the Jaccard distance metric, which measures BoF overlap based on sets of nearest neighbor (NN) feature correspondences computed across a data set. An efficient computational framework allows the method to scale gracefully with the amount of data. Registration error associated with NN correspondences is shown to decrease with increasing numbers of images *N*. Experiments demonstrate strong links between manifold and genetic proximity, from T1-weighted head MRI of 511 twin and non-twin siblings. Siblings are automatically identified with high probability based on BoF distance. NN correspondences are more numerous in the frontal cortices of twin vs. non-twin siblings. An erroneous duplicate subject is identified based on manifold proximity.

Index Terms— Local invariant image features, Genetics, 3D SIFT-Rank, head MRI, Twin data.

1. INTRODUCTION

Big data-style medical image analysis requires robust, scalable algorithms for processing of large, unordered data sets. A primary algorithmic procedure for many data-intensive medical image analysis applications, including alignment[1], segmentation[2][3] and classification[4], is identifying correspondences between images. How does correspondence vary with the number of images *N*, specifically as $N \to \infty$? To what extent do NN correspondences reflect genetic links between subjects? These questions are not yet fully understood.

We propose analyzing nearest neighbor (NN) image correspondences as a function of the number of images *N*. We show that the accuracy of local feature correspondence is guaranteed to improve as *N* grows large, based on a generative model of image appearance. The image appearance manifold is approximated by a proximity graph across an image set. Each image is represented as a bag-of-features (BoF), i.e. an order-less set of salient local image features, and the Jaccard distance approximates the geodesic distance between BoFs based on NN feature correspondences. Scalability is achieved using an efficient $O(N \log N)$ approximate NN search algorithm based on 3D scale-invariant image features and randomized KD-search trees, and thus avoiding the prohibitive $O(N^2)$ complexity of naïve search algorithms.

Experiments demonstrate the system on T1-weighted MRI data from the Human Connectome Project (HCP) [5], where subjects have extensive genetic links in the form of sibling relationships, including monozygotic (MZ) and dizygotic (DZ) twin pairs. Local feature correspondence accuracy increases with the number of images, as predicted by theory. The appearance manifold, approximated by a NN graph between subject BoFs, is highly consistent with family structure labels, indicating close agreement between the proposed manifold distance and genetic proximity. Unknown duplicate subject instances can be identified by unusually low geodesic distance. Feature geometry (location, scale) is not used for correspondence, but can be used to visualize anatomical regions and tissues most closely linked to genetic proximity. Links between genetics and anatomy have been investigated in group-wise twin studies from MRI[6], however our method is unique in rapidly analyzing genetic-structural links in a data-driven, subject-specific basis. To our knowledge, we report the first results in the literature for identifying sibling relationships in a large set of human head MRI, demonstrating the ability of the method to model genetically-relevant image structure.

2. RELATED WORK

The work in this paper is most closely linked to local feature- or patch-based methods, which leverage local image observations for robust medical image analysis. Dense patch-based methods consider all training data, the computational complexity of NN correspondence is thus generally $O(N^2)$ in the number of images/features, which becomes intractable for large *N*. Approaches for reducing complexity include reducing the number of training images/atlases, e.g. via machine learning[8], constraining correspondence via reduced search spaces[3], compact descriptors[15], etc. Sparse feature-based methods, e.g. using salient key points extracted using the so-called scale

invariant feature transform (SIFT)[9] or 3D SIFT-Rank transform[7], consider small subsets of informative image patches or features. Through the use of distinctive image encodings and efficient approximate NN search routines, e.g. randomized search trees[10], correspondence can be achieved in $O(N \log N)$ computational complexity, thus allowing algorithms to scale gracefully to large sets of general radiological image data.

Given scalable correspondence via sparse local features, instance-based learning approaches such as fast approximate K-nearest neighbor (K-NN) inference or kernel density methods may be used[11]. Such methods are highly attractive for big data sets, for several reasons. They are computationally simple and free of explicit parameterization, all feature data is maintained in memory. Computationally expensive training algorithms are thus unnecessary, even as new data arrive, as all inference is performed on-the-fly via so-called lazy-learning methods. Most importantly, the performance of inference is guaranteed to converge to within a small factor of optimal Bayes error as the number of data $N \to \infty$ [12].

3. METHOD

3.1 Local Feature Correspondence

NN correspondence is analyzed using a generative model for local feature data. Let $\{T_i\}$ be a set of transforms mapping each image *i* in a dataset to a reference coordinate system, and let $\{a_{ij}, g_{ij}\}$ be a set of invariant features extracted from images, i.e. geometry g_{ij} (feature location and scale) and appearance a_{ij} descriptors of feature *j* from image *i*. The posterior probability of a transform set $\{T_i\}$ given conditionally independent feature data $\{a_{ij}, g_{ij}\}$ is

$$p(\{T_i\}|\{a_{ij}, g_{ij}\}) \propto p(\{a_{ij}, g_{ij}\}|\{T_i\})p(\{T_i\}),$$
$$= \prod_{ij} p(a_{ij}, g_{ij}|T_i)p(\{T_i\}), \qquad (1)$$

where in Equation (1), the proportionality results from Bayes rule and the second equality from the assumption of conditionally independent features. Factor $p(\{T_i\})$ is a prior probability over the set of transforms, and $p(a_{ij}, g_{ij}|T_i)$ is a density over the geometry and appearance of feature *j* in image *i*, conditioned on transform T_i, also referred to as the likelihood function of T_i.

Dropping subscripts for simplicity, the likelihood function $p(a, g|T)$ linking a single feature to a transform can be expressed in terms of a latent variable *f* taking on a set of *K* discrete values or modes $\{f^1, \dots, f^k, \dots, f^K\}$:

$$p(a, g|T) = \sum_k p(a, g, f^k|T),$$
$$= \sum_k p(a|f^k)p(g|f^k, T)p(f^k). \qquad (2)$$

Equation (2) is a generative mixture model[13][7]. Each mixture component $p(a, g, f^k|T)$ factors into conditional densities over feature appearance $p(a|f^k)$ and geometry $p(g|f^k, T)$ and a mixing probability $p(f^k)$ associated with latent mode f^k. Intuitively, a mode f^k can be thought of as indicating a distinctive semantic image pattern, e.g. a ventricle or cortical fold in neurological MRI, whose occurrence probability is independent of the transform $p(f^k|T) = p(f^k)$. Conditional independence of feature appearance and geometry is a key assumption in Equation (2) specific to invariant feature data, i.e. $p(a|g, f^k, T) = p(a|f^k)$, as feature appearance descriptors are computed in a manner invariant to geometrical deformations (i.e. global similarity transforms + minor local non-linear deformations). This permits efficient inference algorithms based on probable NN appearance descriptor correspondences and obviates the need for search algorithms over geometry or transform parameters.

We analyze the model in Equation (2) in terms of the accuracy of correspondences for increasing numbers of images *N*. As illustrated in Figure 1, NN correspondences can be either correct matches between different features sampled from the same latent conditional density $p(a, g, f^k|T)$, or incorrect matches between different latent modes with similar appearance but different geometrical distributions. In both cases, the appearance distances are similarly low, whereas the geometrical distance (or registration error) is low for correct correspondences but high for incorrect correspondences.

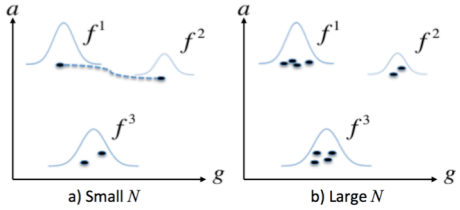

Figure 1: Illustrating the distribution of feature data (dots) sampled from $p(a, g|T)$ in the cases of a) small and b) large numbers of data *N*. In (a) the case of small *N*, incorrect NN correspondences arise between different latent modes (here f^1 and f^2) with similar appearances but different geometry. In (b) the case of large *N*, sufficient sampling permits nominally correct correspondence between features arising from the same modes.

Assume that for a given domain, e.g. brain MRI, the majority of feature data are sampled from conditional densities $p(a, g|f^k, T)$ arising from a finite (possibly very large) set of latent feature *modes* f^k. Define this latent feature set as all f^k for which the occurrence probability $p(f^k)$ is greater than some small value ϵ, i.e. $\{f^k: p(f^k) > \epsilon\}$. Correct NN correspondences between anatomically homologous features are achieved when the least probable mode is sampled at least twice. The probability of this event occurring is lower bounded by the following expression:

$$p\left(\left|\{a_{ij}: a_{ij} \sim p(a, g|f^k, T)\}\right| \geq 2|N\right) =$$
$$1 - \sum_{i=0,1} \binom{i}{N} \epsilon^i (1-\epsilon)^{N-i}, \qquad (3)$$

which increases monotonically with *N*. To show this, it is sufficient to show that the binomial expression representing the probability of 0 or 1 samples $\sum_{i=0,1} \binom{i}{N} \epsilon^i (1-\epsilon)^{N-i} = (1-\epsilon)^N + N\epsilon^1(1-\epsilon)^{N-1}$ decreases monotonically as $N \to \infty$. The first term $(1-\epsilon)^N$ decreases monotonically since $0 < 1-\epsilon < 1$. For the second term $g(N) =$

$N\epsilon^1(1-\epsilon)^{N-1}$, we show that the ratio $g(N+1)/g(N)$ converges to $1-\epsilon<1$, and thus that $g(N+1)<g(N)$:

$$\lim_{N\to\infty}\frac{(N+1)\epsilon^1(1-\epsilon)^N}{(N)\epsilon^1(1-\epsilon)^{N-1}}=\lim_{N\to\infty}\frac{N+1}{N}(1-\epsilon)=1-\epsilon<1. \quad (4)$$

Thus the probability of sampling two features from the same mode as in Equation (3), and thus achieving low-error NN correspondence, increases with the number of images *N*.

3.2 Image Manifold, Bag-of-Features Correspondences

The previous section demonstrated that as the number of images *N* increases, correspondence based purely on local appearance information becomes increasingly accurate. This motivates modeling images as orderless bags-of-features (BoF), which requires no explicit information regarding feature geometry. The BoF space is difficult to characterize, as the number of features varies from one image to another. However the intrinsic manifold of subject appearance can be approximated by a proximity graph in which nodes are image BoFs and edge lengths are defined by a pairwise metric, locally approximating the geodesic distance.

For clarity, NN feature-to-feature correspondences are defined across large set of BoFs $\{a_{ij}\}$ based on an appearance descriptor distance metric $d(a_{ij},a'_{ij})$, and manifold construction requires a pairwise BoF-to-BoF distance metric. Let $A=\{a_{Aj}\}$ and $B=\{a_{Bj}\}$ be sets of local appearance descriptors, i.e. BoFs extracted from two different images. Furthermore, let NN correspondences define equivalence relations between descriptor elements in sets *A* and *B*. We propose using the Jaccard distance between sets[14] as a natural pairwise metric:

$$J(A,B)=1-|A\cap B|/|A\cup B|, \quad (5)$$

where $|A\cap B|$ is number of common features between sets *A* and *B* induced by the equivalence of NN descriptors, and $|A\cup B|=|A|+|B|-|A\cap B|$.

2. EXPERIMENTS AND RESULTS

Experiments use T1-weighted head MRI of 511 subjects, including 51 MZ and 42 DZ twin pairs, and 240 non-twin siblings[5], acquired at 3 Tesla and processed at isotropic 0.7mm resolution. All data are approximately aligned to a template subject via a robust linear transform prior to processing for the purpose of evaluating error, note however that alignment is not required for feature correspondence. Salient features are extracted from MRI volumes according to the 3D SIFT-Rank method using the implementation described in[7]. Briefly, salient feature geometry g_{ij} (location and scale) is defined by extrema in a difference-of-Gaussian image scale-space, and local feature appearance a_{ij} is encoded via the SIFT-Rank descriptor, a 64-element vector of rank-ordered histograms of local image gradient orientations. NN descriptor correspondence uses the Euclidean distance metric and the randomized KD-tree search algorithm[10]. Approximately 2000 features are extracted in each image for a total of 1,000,000 features. Feature extraction requires approximately 20 seconds per volume, after which all feature-to-feature and BoF-to-BoF computation requires on the order of several minutes on a single core processor. The memory footprint of feature data is approximately 100X smaller than original image data. Note the BrainPrint approach [15] also provides a compact descriptor, but requires pre-segmented neuroanatomical structures which may not be readily available, e.g. due to pathology, inter-subject variability, non-brain images, etc.

Increasing Accuracy of NN Feature Correspondence: Trials are performed for increasing numbers of subject BoFs *N*, where for each trial, NN correspondence is computed between all features extracted in different subjects, and registration error is measured as the mean and median Euclidean distance (mm) between the locations of corresponding features in normalized reference space.

The graph of error vs. image number *N* in Figure 2 exhibits a monotonically decreasing trend, as predicted in Section 3.1 and Equation (3). This demonstrates a '*big data*' effect, where as more data arrive, modes f^k from the latent set are sampled multiple times, resulting in nominally correct NN descriptor correspondences and reduced registration error.

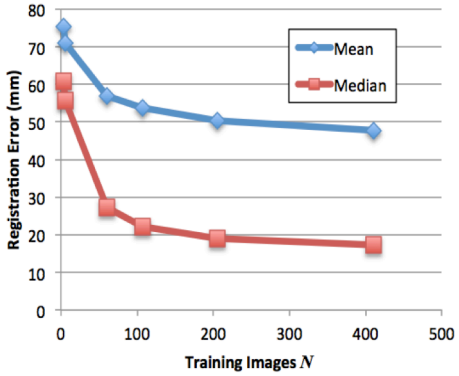

Figure 2: The graph shows registration error of NN correspondences as a function of increasing image set size *N*, mean and median errors decrease monotonically. Note that the median is noticeably lower than the mean, which is skewed by a percentage of spurious outliers. Note also that nominal error is non-zero here due to inter-subject variability within reference space, this could be reduced, e.g. via feature-based alignment [7].

Head MRI Manifold and Familial Links: The head MRI manifold is approximated by the NN graph across all HCP subject images, by first 1) identifying feature-to-feature NN correspondences via Euclidean distance between SIFT-Rank descriptors, then 2) identifying image-level NNs based on Jaccard distance between BoFs. The NN graph is highly consistent with family labels, as shown in Table 1. NN images according to Jaccard distance include 100% of MZ twins, 50% of DZ twins and 23% of non-twin siblings. Differences between features implicated correspondences between twins vs. non-twin correspondences are visualized in Figure 3, note correspondences in the frontal cortex.

Unexpected Result: An unrelated subject pair exhibits an extremely low Jaccard distance, e.g. 0.6 vs. 0.94 for the closest MZ twin pair. Upon visual inspection, it is evident

that the two images in fact represent the same subject. This illustrates an interesting practical application of the method: maintaining the integrity of large medical image databases.

Table 1: Rates for correctly pairing of twins (MZ, DZ) and non-twin siblings (SIB) based on NN Jaccard distance between bags-of-features. Note 100% of MZ twin pairs are automatically identified. Note also that not all 511 subjects have siblings in the HCP data set.

Sibling (counts)	Genetic Overlap	Environmental Factors	Jaccard Distance $J(A,B)$	
			Mean ± Stdev	Correct NN
MZ (102)	1.0	Similar	0.966 ± 0.007	**100%**
DZ (94)	0.5	Similar	0.972 ± 0.010	**50%**
SIB (240)	0.5	Different	0.981 ± 0.004	**23%**

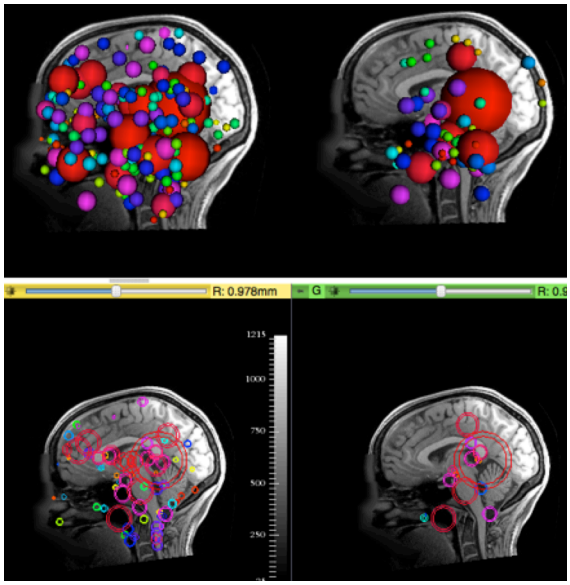

Figure 3: Illustrating bags-of-features (BoFs, spheres/circles) implicated in NN feature matches from a single subject to its MZ twin (left) and non-twin sibling (right), viewed in 3D (above) and in mid-sagittal slices (below). Note the higher concentration of NN matches in the frontal cortex for MZ twin matches (left) vs. non-twin siblings (right), reflecting established neuroanatomical similarities [6].

4. DISCUSSION

We present a method for big data medical image analysis using local invariant features, based solely on local appearance descriptor correspondences in an instance-based learning framework. The geometrical accuracy associated with NN descriptor correspondences is shown to improve with increasing numbers of images *N*. The image manifold across a large set of brain MRIs is approximated using a NN graph and the Jaccard distance, where twin and non-twin siblings are automatically identified as NNs with high probability. Feature correspondences can be visualized to understand how image-to-image similarity varies with spatial location throughout the anatomy of interest. An unintentional duplicate subject is identified and reported to the HPC study. Future directions will include investigating the significance of feature correspondences in the context of heritability studies. The approach is general and is being applied to chest CT, prostate ultrasound and full-body MRI.

Acknowledgement: NIH grant funding: P41EB015902.

5. REFERENCES

[1] V. Potesil, T. Kadir, G. Platsch, and S. M. Brady, "Personalized Graphical Models for Anatomical Landmark Localization in Whole-Body Medical Images," *Int. J. Comput. Vis.*, 111(1), pp. 29–49, 2014.

[2] C. Wachinger, M. Toews, G. Langs, W. Wells, and P. Golland, "Keypoint Transfer Segmentation," in *Inf. Proc. in Medical Imaging*, 2015, pp. 233–245.

[3] P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *NeuroImage*, 54(2), pp. 940–954, 2011.

[4] M. Toews, C. Wachinger, R. S. J. Estepar, and W. M. W. III, "A Feature-Based Approach to Big Data Analysis of Medical Images," in *Inf. Proc. in Medical Imaging*, 2015, pp. 339–350.

[5] D. C. Van Essen et al., "The Human Connectome Project: A data acquisition perspective," *NeuroImage*, vol. 62, no. 4, pp. 2222–2231, 2012.

[6] P. M. Thompson et al., "Genetic influences on brain structure," *Nat. Neurosci.*, 4(12), pp. 1253–1258, 2001.

[7] M. Toews and W. M. Wells III, "Efficient and robust model-to-image alignment using 3D scale-invariant features," *Med. Image Anal.*, 17(3), pp. 271–282, 2013.

[8] A. J. Asman, Y. Huo, A. J. Plassard, and B. A. Landman, "Multi-atlas learner fusion: An efficient segmentation approach for large-scale data," *Med. Image Anal.*, vol. 26, no. 1, pp. 82–91, 2015.

[9] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[10] M. Muja and D. G. Lowe, "Scalable Nearest Neighbor Algorithms for High Dimensional Data," *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(11), pp. 2227–2240, 2014.

[11] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Mach. Learn.*, 6(1), pp. 37–66, 1991.

[12] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, 13(1), pp. 21–27, 1967.

[13] A. Myronenko and X. Song, "Point Set Registration: Coherent Point Drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(12), pp. 2262–2275, 2010.

[14] M. Levandowsky and D. Winter, "Distance between Sets," *Nature*, 234(5323), pp. 34–35, 1971.

[15] C. Wachinger, P. Golland, W. Kremen, B. Fischl, and M. Reuter, "BrainPrint: A discriminative characterization of brain morphology," *NeuroImage*, vol. 109, pp. 232–248, 2015.