

Geodesic Information Flows

M. Jorge Cardoso¹, Robin Wolz², Marc Modat¹, Nick C. Fox³,
Daniel Rueckert², and Sebastien Ourselin^{1,3}

¹ Centre for Medical Image Computing (CMIC), University College London, UK

² Visual Information Processing Group, Imperial College London, UK

³ Dementia Research Centre (DRC), University College London, UK

Abstract. Homogenising the availability of manually generated information in large databases has been a key challenge of medical imaging for many years. Due to the time consuming nature of manually segmenting, parcellating and localising landmarks in medical images, these sources of information tend to be scarce and limited to small, and sometimes morphologically similar, subsets of data. In this work we explore a new framework where these sources of information can be propagated to morphologically dissimilar images by diffusing and mapping the information through intermediate steps. The spatially variant data embedding uses the local morphology and intensity similarity between images to diffuse the information only between locally similar images. This framework can thus be used to propagate any information from any group of subject to every other subject in a database with great accuracy. Comparison to state-of-the-art propagation methods showed highly statistically significant ($p < 10^{-4}$) improvements in accuracy when propagating both structural parcelations and brain segmentations geodesically.

1 Introduction

Since the advent of open imaging databases, researchers have struggled with the fact that extra sources of clinical, structural and anatomical information are only available on a small subset of the data. These sources of information (e.g. manual parcelations, anatomical landmarks, tissue priors, pathological classification) are usually scarce since they require large amounts of human interaction. Ideally, one would like to be able to automatically extrapolate and propagate this information to morphologically dissimilar datasets in a coherent manner.

In neuroimage analysis, the best example of information propagation is multi-atlas segmentation. Many researchers have shown that propagating structural parcelations from multiple sources, by mapping them to new unseen data using image registration and then fusing the candidate parcelations, provides a good estimation of the true underlying parcelation [1,2]. However, in the case of limited and morphologically clustered source of information, like the 30 young control subjects with an associated parcelation of 83 key brain areas provided by Hammers et al. [3], structural parcelation propagation can be problematic. As these parcelations are defined only on young controls with normal anatomy, it is non trivial to directly map this information to morphologically dissimilar

and pathological subjects [4,5] without introducing large errors. More recently, Wolz *et al.* [6] introduced the LEAP approach (learning embeddings for atlas propagation) for brain segmentation. In LEAP, a low dimensional representation of the data is used to find morphologically similar datasets. This morphological similarity is then used to gradually diffuse the segmentation of the brain from the 30 Hammers atlases [3] to pathological Alzheimer’s diseased patients via morphologically similar intermediate datasets, greatly increasing the segmentation accuracy. However, as the region-of-interest size increases, the morphological embedding becomes less localised, potentially resulting in a decrease in performance. A similar framework, but for geodesic image registration, was also introduced by Hamm *et al.* [7] with the GRAM (geodesic registration on anatomical manifolds) method. This family of step-wise propagation algorithms will become increasingly relevant with the availability of larger and larger databases. Ideally, one would like to slowly diffuse any information from its source to all the other images in a database in an unbiased manner.

This work presents an algorithm where information is propagated along the geodesic path of the local data embeddings. These local embeddings are represented as a voxel-wise implicit undirected graph with a heat kernel based information reconstruction weight that uses both the local image similarity and local image morphology as a metric. The proposed algorithm not only allows information (structural parcelations, tissue atlases, *etc.*) to be transferred between anatomically disparate images with great accuracy, but also allows the estimation of geodesic distance maps that represent the local degree of confidence and extrapolation of the propagated information. Overall, as the reconstruction weights are both symmetric and inverse consistent, the framework can be used to propagate any information from any subject (or group of subjects) to every other subject in a dataset using an unbiased information flow.

2 Methods

This section will first introduce the mathematical framework and the undirected graph for geodesic information flow, followed by the introduction of the distance metric between images. Finally, the geodesic information propagation step is presented for two types of information: numerical and label attributes.

2.1 The Implicit Local Data Embedding

Let a set Y with R images be the full set of observed *grey-valued* anatomical data with the a -th image of this set denoted by Y_a . Each image Y_a is going to be a vector of size N_a , with its i -th voxel denoted by $Y_{a,i}$.

In order to embed the observed data within a manifold, one normally starts by finding a distance between each pair of images. This distance provides insights about the global similarity between the images, and subsequently, about the manifold structure of the data. Theoretically, this global embedding assumes that one can represent the space spanned by the full data in a low dimensional

space. However, due to the complexity of the data, the dimensions of the manifold can lack interpretability and usefulness. For example, Gerber *et al.*[8] explored the manifold structure of the space of brain images and concluded that the first dimension of the manifold represents global ventricular expansion due to disease and ageing, while the second dimension meaning is described as "less obvious". Ideally, one would like to be able to capture the local manifold structure of the brain on a spatially constrained neighbourhood and not the global brain morphology. With one manifold representation per voxel, one would be able to describe the local brain morphology and similarity as a measure of distance between two mapped locations, and respective neighbourhoods, in two different images. However, due to computational and memory requirements, this problem is untractable. As an example, only to store a pairwise distance matrix for one single image at every voxel and assuming a set of 120 neighbouring images with average size 200^3 , one would need approximately 400GB of memory. Furthermore, the memory requirements will grow proportionally to R^2 , where R is the number of datasets. Thus, one cannot have an explicit representation of the manifold at the voxel level. In this work, instead of constructing an explicit representation of the manifold, we implicitly represent the manifold through the local neighbourhood graph of each data point. This greatly reduces both computational complexity and memory requirements, making the problem tractable and linearly scalable with the number of datasets R in the database.

Let D be a set of distance matrices, with the a -th matrix of this set denoted by D_a . Here, D_a will be an $N_a \times (R - 1)$ matrix describing the distance between the image a and each one of the remaining $(R - 1)$ images at every sample position i . More specifically, $D_{a \rightarrow b}(i)$ will contain the distance between the i -th sample of image Y_a and its corresponding sample in image Y_b . We now introduce a threshold d_t over these distances. By doing so, one can now build an undirected graph where the neighbourhood of each data-point is restricted only to the data-points with $D_{a \rightarrow b}(i) < d_t$. Note that the graph is undirected only if the distances are a semi-metric (subadditivity is not required).

In this work, one does not need to explicitly represent the full graph. In order to solve the information diffusion problem at a given location, one only needs to keep track of the graph neighbourhood at that specific location, visually shown in Fig. 1 - right. Here, the realm of observations (the blue connections) from the data point in bold is limited by its direct neighbouring nodes with distances below d_t . Under this undirected graph assumption, a heat kernel decay function $W_{a \rightarrow b}(i)$ is then used to diffuse the information [9]. This kernel is defined as

$$W_{a \rightarrow b}(i) = \begin{cases} e^{-\frac{D_{a \rightarrow b}(i)}{t}} & D_{a \rightarrow b}(i) < d_t \\ 0 & \textit{else} \end{cases} \quad (1)$$

with t being a heat kernel temperature that will determine the speed and the distance the information can diffuse. In this work, we set $d_t = t$, meaning that the choice of t will determine both the maximum distance traveled by the information and the amount of information diffusion that occurs at each iteration.

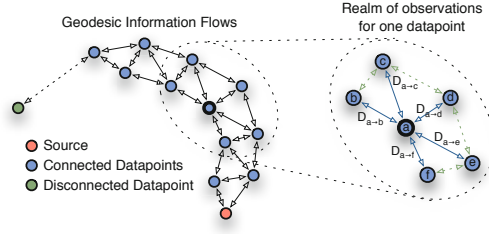


Fig. 1. Left) Implicit manifold with the neighbourhood defined as all the data points within a certain distance. Note that if the manifold is sparse, some data points might be disconnected. Right) Diagram representing the observed and unobserved connections (in blue and green respectively) and distances from the standpoint of the data point a .

2.2 The Distance Metric

The heat kernel decay function is based on the assumption that one can calculate a distance between two nodes in the graph. This distance should be at least a semi-metric, respecting both the coincidence and separation axioms and symmetry. In a medical imaging framework the local distance between images should take into account both local morphology and local image similarity. Gerber et al. [8] proposed to use coordinate transformations as a distance metric. These coordinate transformations map an image Y_a to an image Y_b by finding the optimal transformation $T_{a \rightarrow b}$ that maximises some cost function. In order to be a semi-metric, this coordinate transformation has to be symmetric, inverse consistent and diffeomorphic. In our work, we use a symmetric variant of a non-rigid free-form registration algorithm as described in [10]. Under the symmetry and diffeomorphism constraints, the transformation $T_{a \rightarrow b} = T_{b \rightarrow a}^{-1}$ and $T_{a \rightarrow b} \circ T_{b \rightarrow a} = \text{Id}$, with T^{-1} being the inverse of the transformation, \circ being the composition operator and Id the identity transformation. In order to remove the smoothly varying local affine component of the transformation that characterises the global anatomical shape differences, the low frequency component of the transformation is removed. From the high-frequency version of the transformation, one can then find the displacement field $F_{a \rightarrow b}$ that describes how much a point i in Y_a had to move to match the corresponding point j in Y_b .

Even though this displacement field will describe the morphological differences between the brains, we also combine it with an intensity similarity metric in order to assess the local similarity between the images after transformation [11]. This similarity term is necessary to characterise both the local differences in tissue appearance due to pathology (e.g. damaged WM in dementia) and also some possible local registration errors. The local similarity between an image Y_a and an image Y_b transformed by $T_{b \rightarrow a}$, denoted by $L_{a \rightarrow b}$, can be calculated as the local sum of squared differences (LSSD) between the intensity in these images, using a cubic B-spline kernel as a local smoothing function. We combine the two metrics together by setting $D_a(i, b) = \alpha L_{a \rightarrow b}(i) + (1 - \alpha) F_{a \rightarrow b}(i)$, with α being a relative weight, meaning that both a low displacement and a low LSSD are

necessary to obtain a low distance $D_a(i, b)$ between images. In this work $\alpha = 0.5$. Optimisation of α will be part of future work.

2.3 Geodesic Information Flows

The two previous sections have defined the neighbourhood graph and the distance metric. This section will make use of the graph structure to introduce the concept of propagating information between neighbouring nodes of the graph.

Let I be a set of vectors, with the a -th vector of this set denoted by I_a . Each vector I_a has its i -th element denoted by $I_{a,i}$, representing the associated information at location i for image Y_a . Assuming that the information to be propagated is only available in a subset of images within the full database, only some of the vectors I_a will be defined. Now, let an indicator vector K_a , indexed by $K_{a,i}$, characterise the status of the information at location $I_{a,i}$. Here, $K_{a,i}$ is set to 2 if $I_{a,i}$ is a source of information, set to 1 if the information has diffused and reached the location $I_{a,i}$ or 0 if the information does not exist at location $I_{a,i}$. As the realm of observations at each spatial location is limited by its closest neighbours, the best approximation for data flow is given by a normalised weighted sum of the information available within the neighbourhood. Thus,

$$I_{a,i} = \frac{\sum_{\forall b \in K_{b,j} > 0} W_{a \rightarrow b}(j) I_{b,j}}{\sum_{\forall b \in K_{b,j} > 0} W_{a \rightarrow b}(j)} \quad (2)$$

with j being the spatially transformed coordinate i into the space of image I_b , mapped using the previously described transformation. The information flow is thus governed by the heat kernel-derived weights $W_{a \rightarrow b}(i)$. If the set $b \in (K_{b,j} > 0)$ is not empty, the information in $I_{b,j}$ can then be propagated to $I_{a,i}$. In this case, $K_{a,i}$ is set to 1, else $K_{a,i}$ is set to 0. Note that Eq. 2 is only valid for floating point data propagation like a probabilistic atlas. The same equation can be re-formulated in a weighted label fusion scheme, by making $I_{a,i}$ equal to $p(I_{a,i}, l)$, representing the probability that location i in image a has label l , and by making $I_{b,j}$ equal to $p(I_{b,j}, l)$, representing the probability that location j has label l in image b . Eq 2 is solved iteratively for all samples where $K_{a,i} < 2$, until all are marked as solved. The number of times $I_{a,i}$ could not be solved because $K_{b,j} > 0 \forall b$ is an empty set, represents the number of steps through the graph's geodesic path necessary to transport the information from its source. The number of steps represent the amount of extrapolation of information, where larger extrapolation should result in lower accuracy.

Finally, because the temperature t will determine the neighbourhood size and consequently the existence of a connection from every information source to all the targets, the geodesic information flow is solved multiple times for several values of t . This temperature t is varied between 5 and 1 with decrements of 0.5. In a similar fashion to an annealing process, the information at the lowest possible temperature is kept as the answer, as the distance traveled by each extrapolation step is the lowest.

3 Validation

The data used in this work, depicted in Fig. 2, is compromised of two sets: 30 young controls with associated structural parcelation of 83 key structures [3] (<http://www.brain-development.org>); 90 subjects from the ADNI database (<http://adni.loni.ucla.edu>), subdivided into 30 controls, 30 Mild Cognitive Impairment (MCI) and 30 Alzheimer’s diseased (AD) patients with associated manual segmentations of the brain. The current work aims at homogenising databases under the assumption that extra information is only available on a subset of the data. From these sources of information, measuring the information extrapolation accuracy will always be limited by the anatomical and pathological variability within the full dataset and by the range of available segmentations. Furthermore, the most complex sources of information, like the 30 young controls with full brain parcelations, are simply not available in pathological subjects. This makes the validation anecdotal for untested morphologies. The proposed validation will thus have two components. First, the overlap accuracy of multi-label information propagation will be estimated and compared to MAPER [4] using a leave one out approach on the 30 young controls. Then, the accuracy of information extrapolation accuracy will be characterised by propagating the brain segmentations from the elderly control group to the MCI and AD patients.

Multi-label Propagation Accuracy: The accuracy of propagating information through a geodesic path was compared to MAPER [4], a direct information fusion method based on majority voting. As the amount of parcelations available for validation is limited, a leave-one-out cross validation was performed only on the 30 young controls that have manual brain parcelations. One should note that the limited availability of segmentations restricts the range of morphological variability in the propagation, thus not representing the real performance when segmenting morphologically dissimilar subjects.

In this paper, the Dice score was used as a measure of accuracy. The mean Dice scores per structure for the the leave-one-out cross validation are shown in Table 1. Out of 83 structures, 15 structures had a significantly higher Dice score using the Geodesic information Flow when compared to MAPER, while only two structures (lingual gyrus and superior parietal gyrus) were better segmented in MAPER. The mean Dice score over all structures and all patients for the

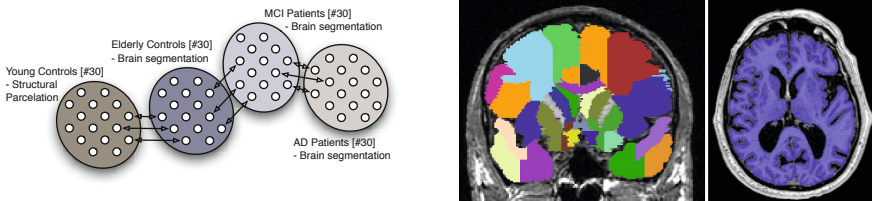


Fig. 2. Left) All the sets of data used in this work and their associated information. Right) One dataset from the Hammers atlas overlaid with its associated structural parcelation; an ADNI AD subject with the associated brain segmentation

Table 1. Mean Dice coefficient for a set of key structures, comparing the proposed method (GIF) with MAPER [4]. Statistically higher mean Dice is shown in bold font. Only a limited number of structures are shown due to restrictive space availability.

Structure	Unilateral Structures					
	GIF	MAPER	p-value			
All Structures	0.8179	0.8089	$< 10^{-4}$			
Corp. callos.	0.8802	0.8674	$< 10^{-4}$			
Brainstem	0.9534	0.9377	$< 10^{-4}$			
Structure	Left Side			Right Side		
	GIF	MAPER	p-value	GIF	MAPER	p-value
Hippocampus	0.8439	0.8335	0.0048	0.8251	0.8211	0.2036
Amygdala	0.8263	0.7922	$< 10^{-4}$	0.8241	0.7830	$< 10^{-4}$
Caudate nucl.	0.8983	0.8923	0.0380	0.9004	0.8955	0.0478
Nucleus acc.	0.7581	0.6834	$< 10^{-4}$	0.7317	0.6707	$< 10^{-4}$
Putamen	0.9069	0.8916	$< 10^{-4}$	0.9107	0.8959	$< 10^{-4}$
Thalamus	0.9210	0.8879	$< 10^{-4}$	0.9205	0.8852	$< 10^{-4}$
Pallidum	0.8562	0.7661	$< 10^{-4}$	0.8554	0.7672	$< 10^{-4}$

proposed method (0.8197) was significantly higher ($p < 10^{-4}$) than in MAPER (0.8089). An example of the propagation to a highly atrophied subject from the ADNI database is shown in Fig. 3 (right).

Information Extrapolation Accuracy: In the previous subsection, the accuracy of propagating information through a geodesic path was limited to a morphologically similar set of subjects. Thus, the previous validation does not capture the ability to extrapolate information to anatomically disparate subjects. The information extrapolation accuracy is thus assessed by using only a subset (the elderly control group) of all the manual brain segmentations. This morphologically clustered set of data is then used to segment both the MCI and AD groups. The proposed geodesic propagation algorithm is compared to a direct propagation algorithm based on the locally weighted majority voting algorithm [12]. The results are presented in Fig. 3 (left). The mean (std) Dice score for the proposed geodesic method was 0.940(0.009) and 0.947(0.008) for the AD an MCI groups respectively while for the direct method, the mean (std) Dice score was 0.934(0.009) and 0.942(0.008) for the AD an MCI groups respectively.

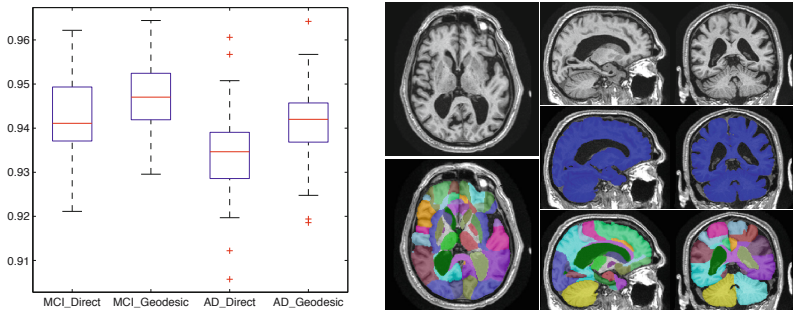


Fig. 3. Left) Dice scores for direct and geodesic propagation of brain mask. Right) An example of the propagation of both the structural parcellation and brain segmentation to an highly atrophied AD subjects (ID:1281) from the ADNI database. Note the correct ventricle segmentation and the smooth deep grey matter parcellation.

This represents a statistically significant ($p < 10^{-4}$) increase in segmentation accuracy when using a two-tailed paired t-test for statistical comparison. Note that, due to the lack of post-processing and the limited size of the training set, one should not compare these results with other brain segmentation methods.

4 Conclusion

This work presents an algorithm where information is geodesically propagated through a local implicit neighbourhood graph. Application to structural parcellation and brain segmentation propagation has demonstrated the significant ($p < 10^{-4}$) advantages of the proposed framework when compared to state of the art methods. Overall, the proposed framework can be used to propagate any information from a group of subjects to every other subject in a dataset. All the software and results used for this work is available at <http://niftyseg.sf.org>.

Acknowledgements. This study was supported by the EPSRC (EP/H046410/1), the CBRC Strategic Investment Award (Ref. 168) and the 7th Framework Programme by the European Commission (<http://cordis.europa.eu/ist/>)

References

1. Rohlfing, T., Russakoff, D.B., Maurer, C.R.: Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE Transactions on Medical Imaging* 23(8), 983–994 (2004)
2. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *NeuroImage* 46(3), 726–738 (2009)
3. Hammers, A., Chen, C., Lemieux, L.: Statistical neuroanatomy of the human inferior frontal gyrus and probabilistic atlas in a stereotaxic space. *HBM* (2007)
4. Heckemann, R.A., Keihaninejad, S., Aljabar, P.: Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. *NeuroImage* (51), 221–227 (2010)
5. Cardoso, M.J., Modat, M., Keihaninejad, S., Cash, D., Ourselin, S.: Multi-STEPS: Multi-label Similarity and Truth Estimation for Propagated Segmentations. In: *MMBIA*, pp. 153–158 (2012)
6. Wolz, R., Aljabar, P., Hajnal, J.V., Hammers, A., Rueckert, D.: Alzheimer’s Disease Neuroimaging Initiative: LEAP: learning embeddings for atlas propagation. *NeuroImage* 49(2), 1316–1325 (2010)
7. Hamm, J., Ye, D.H., Verma, R., Davatzikos, C.: GRAM: A framework for geodesic registration on anatomical manifolds. *MedIA* 14(5), 633–642 (2010)
8. Gerber, S., Tasdizen, T., Joshi, S., Whitaker, R.: On the Manifold Structure of the Space of Brain Images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part I. LNCS*, vol. 5761, pp. 305–312. Springer, Heidelberg (2009)
9. Kondor, R.I., Lafferty, J.: Diffusion kernels on graphs and other discrete input spaces. In: *ICML (January 2002)*

10. Modat, M., Cardoso, M.J., Daga, P., Cash, D., Fox, N.C., Ourselin, S.: Inverse-Consistent Symmetric Free Form Deformation. In: Dawant, B.M., Christensen, G.E., Fitzpatrick, J.M., Rueckert, D. (eds.) WBIR 2012. LNCS, vol. 7359, pp. 79–88. Springer, Heidelberg (2012)
11. Souvenir, R., Pless, R.: Image distance functions for manifold learning. *Image and Vision Computing* 25(3), 365–373 (2007)
12. Yushkevich, P.A., Wang, H., Pluta, J., Das, S.R., Craige, C., Avants, B.B., Weiner, M.W., Mueller, S.: Nearly automatic segmentation of hippocampal subfields in in vivo focal T2-weighted MRI. *NeuroImage* 53(4), 1208–1224 (2010)