2010 International Conference on Pattern Recognition

# Multiple atlas inference and population analysis with spectral clustering

Giorgos Sfikas[1,2], Christian Heinrich[1], Christophoros Nikou[2]

[1]LSIIT-University of Strasbourg, 67412 Illkirch cedex, France
[2]Department of Computer Science, University of Ioannina, 45110 Ioannina, Greece

***Abstract*—In medical imaging, constructing an atlas and bringing an image set in a single common reference frame may easily lead the analysis to erroneous conclusions, especially when the population under study is heterogeneous. In this paper, we propose a framework based on spectral clustering that is capable of partitioning an image population into sets that require a separate atlas, and identifying the most suitable templates to be used as coordinate reference frames. The spectral analysis step relies on pairwise distances that express anatomical differences between subjects as a function of the diffeomorphic warp required to match the one subject onto the other, plus residual information. The methodology is validated numerically on artificial and medical imaging data.**

## I. Introduction

In medical image analysis an *atlas*, in its simplest form, is the image of the average subject out of a given training population. Atlas construction dates back to early attempts which were based on one, or at best a few, medical images [1], [2]. In the more elaborate sense of the term, *atlas* may refer to models of the image population, aiming to describe intensity, structural, vascular and functional variability. Information conveyed by the atlas can be used for various purposes, including normalization to the atlas coordinate frame using the atlas as a registration template, segmentation of certain tissues of interest, studying and identifying pathologies and structural anomalies in new patients [3], [4], [5].

In order to infer useful information out of a set of images, the first step is to bring them into correspondence in a common coordinate frame. This is typically performed by choosing a single template from the training population, and have the rest of the data registered one-by-one onto the frame of this subject. However, if the choice of the template subject is done poorly, the resulting atlas will be biased toward the template. This problem has been addressed by searching for the least-biased template [3] or by groupwise registration to a common coordinate frame [4], [6], that does not require any template choice. However, these approaches work under the hypothesis, explicitly stated or not, that the examined population is unimodal. Such a hypothesis may be damaging for the population analysis if in fact there exist multiple modes in the data, like for example a group of control subjects mixed with neurodegenerative disease-suffering patients.

Giorgos Sfikas was supported by a grant from *Région Alsace* (France)

Identification of multiple modes has been explored in [7], where the population is clustered using the mean shift algorithm. The subjects are brought pairwise into correspondence using a registration algorithm then inter-subject distances - using data representations in image space- are calculated to infer the data neighborhoods required by mean shift. The registration step has to be repeated for every iteration of mean shift, resulting in a heavy computational load for large data sets. In [8], a generative model was proposed that corresponds each population mode to a kernel of a Gaussian mixture model (GMM). The model is then solved using an EM-based algorithm, simultaneously registering the training subjects to their respective calculated templates. While this model is clearly more sophisticated than a simple GMM, say over subject intensities, it is well-known that the obtained result in mixture models is highly sensitive to initialization, especially in high-dimensionality problems such as the problem at hand. Thus, the analysis could easily be biased towards false modes.

In the current work, we propose a method to infer multiple atlases out of a population of images, that does not require choosing any initial template estimates and does not assume any single common reference frame prior to the final result. The presented algorithm is based on spectral clustering [9]. Spectral clustering treats data as an undirected weighted graph, constructed by means of their intersubject distances; it has been used successfully in a wide variety of applications [10], [11]. In [12], spectral clustering was used to classify MR images using feature vectors based on atlas-based constructed segmentations of various brain structures. In the proposed method, subject distances are calculated using directly the result of trigonometric-kernel based non-rigid pairwise registration [13]. We argue that such distances can be used in a meaningful manner and finally evaluate the entire framework numerically.

The proposed scheme was tested on an artificial image set as well as on real brain MR images with simulated deformations, yielding promising results.

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.612

## II. Generalities and spectral clustering

Let us assume a training set $X$ of $N$ images $\{x_1, x_2, ..., x_N\}$. Our goal is to effectively cluster the set into $K$ groups, assigning a separate representative image - the *atlases* - to each of these clusters. Given the number $K$ of required atlases / clusters and a notion of similarity between the training data, spectral clustering can give us a partition of the data that maximizes intra-cluster while minimizing inter-cluster image similarity. This is done by exploiting some useful properties of the *graph Laplacian* matrix $L$ [9]. Using the eigenstructure of $L$, every datum is represented as a $K$-variate vector, and *k*-means is used to finally cluster the resulting $K$-dimensional set. Spectral clustering can solve complex setups of the training data, even when the data clusters do not form strictly convex sets.

In this work, we calculate all inter-subject distances and thus construct first a *distance* matrix. This matrix is subsequently transformed to give the similarity matrix required to compute the Laplacian. The choice of this transformation is not a trivial task [9]; we choose to use a *k*-nearest neighbor graph, which produces a sparse similarity matrix and depends on a practically intuitive parameter. Graph-theoretical results [9] suggest a value of around $k = log(N)$ as a good first approximation, which worked well in our experiments.

## III. Deformable registration model and proposed method

In order to construct the distance matrix and subsequently the graph Laplacian matrix, we quantify anatomical differences between each image pair using deformable non-rigid registration. In this work we have used the non-rigid diffeomorphic registration model proposed in [13], which employs grid-based trigonometric kernel interpolating warps. In this model, by constraining the offset of the control points to lie within a certain bound, the calculated deformation is easily guaranteed to be diffeomorphic. Diffeomorphism is a desirable trait in the medical imaging and computational anatomy context [6], as it guarantees invertibility and no folding or tearing.

The registration can be easily described by a finite set of parameters. Matching image $x_i$ to image $x_j$ will yield a warp parameter vector $d_{ij}$, of dimensionality equal to the number of the warp degrees of freedom, which are analogous to the number of control point displacements. This representation offers us the advantage that any convex combination of such vectors corresponds to a valid diffeomorphic deformation, given that the original vectors correspond to known diffeomorphisms themselves [13]. Any interpolating straight line in the parameter space between known diffeomorphisms comprises of valid diffeomorphisms, describing a continuous displacement of the control points. Under the hypothesis that traveling along this interpolating line may also have an intuitive meaning, in this work we consider the Euclidean norm of the warp vector $d_{ij}$ as the anatomical distance between two images $x_i$ and $x_j$.

However, warping $x_i$ according to $d_{ij}$ is not guaranteed to give back $x_j$ exactly. Large differences between images are hard to be well represented by a diffeomorphism, and in practice we will have a *residual* between the warped and target image not captured by the transformation [14]. We account for this by concatenating the residual norm as a separate variate to the warp vector $d_{ij}$[1], weighted so that it is commensurate with the control point displacements. In order to calculate the proper weight, we compute the residual between a source random-intensity image and a warp on the same random image. The ratio between the magnitude of the computed residual and the weight of the used warp (if not unitary) gives the required weight. This weighting is similar-in-spirit with the weighting between warps of different multiresolution levels used [13]; also the computed weight value empirically has shown to be relatively robust to the realization of the random image used, as well as the warp choice.

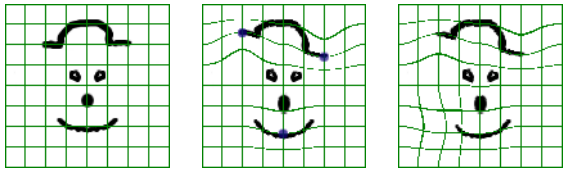

Figure 1. Template and synthesized sketches from the toy set. Left: The original drawing, with coordinate grids shown. Center: Deformation is applied as described in text; a blue circle marks the displaced control points. Right: Registration result of the original to the deformed image.

Note that we have not defined any vector representation in some common reference space for all the training data. Vectors $d_{ij}$ are only meaningful as describing the distance between only the given subjects $x_i, x_j$. Having a common reference space for the deformations would be equivalent to finding and using a single template subject, which is in itself not a straightforward task and would bias the analysis towards the used atlas [3], [6]. On the other hand, this would refute our primary hypothesis that multiple atlases exist and represent the data. Avoiding a common feature vector representation altogether fits well with spectral clustering, where all that is required is a way to define the inter-datum distances.

Once all inter-subject distances $\|d_{ij}\|$ are computed, we calculate[2] the similarity matrix and cluster $X$ as described in the previous section. Finally, we choose as representative atlas of each cluster their respective medoids, as they min-

[1]From this point onward, $d_{ij}$ will refer to this augmented vector that comprises the residual.

[2]Since in general the registration result $x_i$ to $x_j$ will not be the same as $x_j$ to $x_i$, $\|d_{ij}\| \neq \|d_{ji}\|$. Hence, we work with the symmetric component of the similarity matrix.

imize inter-subject distance in their respective clusters and can be easily computed using the distance matrix.

## IV. Experiments

We have tested the proposed scheme on a toy data set, as well as on a synthetic data set based on real MR images. The toy data set consists of 40 $128 \times 128$ grey-level images, each being a deformed version of a cartoon-like drawing (fig. 1). Note that these deformations were created using bilinear interpolation to calculate warps for points not in the control point grid, while the registration model as already discussed uses trigonometric kernels. Half of the toy set depicts the cartoon wearing the hat straight (*'hat straight'*), while the other half of the set shows the hat not worn straight, in varying positions (*'hat oblique'*) (fig.2) . In all cases a random facial expression deformation was applied. The control point coordinates deformation on the left side of the hat $\pi_l$ and the deformation on the right side $\pi_r$ are constrained to satisfy $\pi_l^2 + \pi_r^2 = const.$. At the same time $\pi_l, \pi_r$ are dependent on the displacement of the bottom control point (see fig.1), thereby producing a helix-like structure for the 'hat oblique' data.

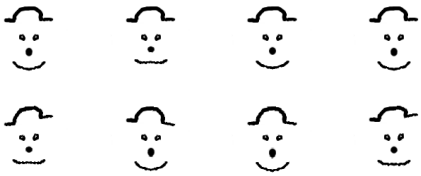

Figure 2. Samples from the toy image set. The cartoons on the top row (class 1) have their hat straight, while the cartoons on the bottom have their hat oblique (class 2).

The toy set images were processed with the proposed scheme, and the two clusters / population modes were identified successfully. For purposes of comparison, we have performed the same test with other well-known clustering algorithms, namely *k*-means, a Gaussian mixture model and mean shift, the latter used in [7] in the same context as in the present work. As the algorithms used for comparison require a data vector representation to work, we have warped all training data to a fixed template and calculated the resulting warp vectors. In practice, an unsuitable template could result in problematic warp vectors, inhibiting performance; this is a strong point for the proposed method, which skips this step.

The results are presented in table I, where we show correct classification ratio (ratio of subjects classified correctly, *CCR*) mean values for 1000 random algorithm initializations. For the proposed method, there is no result variation due to initialization (though it is possible in general that the *k*-means used in the spectral analysis final step will not lead to the same single solution always). For mean shift, we show results for neighborhood bandwidth $h = 0.1$, which

Table I
Correct classification ratio (CCR) results for toy set (hat straight/hat oblique) trials.

| *Method* | *CCR* |
|---|---|
| Proposed | $80.0\% \pm 0.0$ |
| mean shift | $62.6\% \pm 0.4$ |
| *k*-means | $60.2\% \pm 7.0$ |
| GMM | $60.1\% \pm 10.0$ |

calculated the actual number of classes (2). For $h < 0.093$ mean shift clusters the data in more than 2 classes and in general overclusters arbitrarily the data, with the number of classes rapidly increasing as the bandwidth decreases. For $h > 0.12$ on the other hand, the data form a single cluster. Note that using a $k$-nearest neighbor neighborhood definition for mean shift, resulted in collapsing all data to a single mode for any value of $k$.

While it is clear that the scenario may be overly simple and artificial, the point we want to make here is that there may exist clusters of deformations -corresponding to the warp vector distance introduced- with a complex, non-convex structure. At the same time these can correspond to an *intuitively* simple training set partition, like the hat straight/hat oblique scenario presented. Spectral clustering can capture such structures, whereas more conventional clustering methods fail.

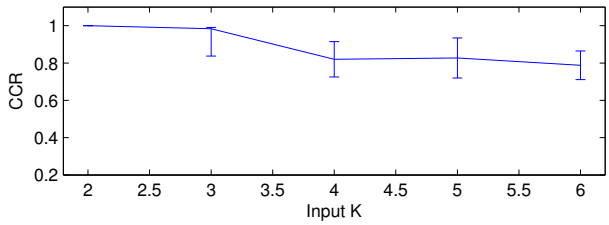


Figure 3. Correct classification ratio (CCR) results for the OASIS brain MRI dataset with respect to the number of clusters K.

In order to test the proposed scheme on medical imaging data, we synthesized a brain image dataset based on a number of $208 \times 176$ MR axial brain slices obtained from the OASIS brain repository [15]. The images were taken from corresponding slices of affinely-registered, gain-field corrected whole brain MRIs. Each image was created by randomly deforming a multiresolution grid of $2\times2, 4\times4$ and $8 \times 8$ control points (see fig. 4). Therefore, the synthesized images for a given real image should ideally belong to the same cluster, i.e. should be assigned to the same atlas. We used 6 original images, and synthesized 4 images for each original image; hence we had 30 subjects in total, corresponding to 6 real templates. The correct classification ratio results are summarized in fig. 3, where we also show corresponding results for subsets of the original set. Each of these subsets comprises $K = 2...5$ original MRIs along with

the corresponding synthesized subjects. In all, we remark that for all values of $K$ the average CCR is close to 1, exhibiting little variance. For the 2-template set in particular, subjects are clustered perfectly. Also, as $K$ increases the problem difficulty also increases, however there is no noticeable decay on the classification result.

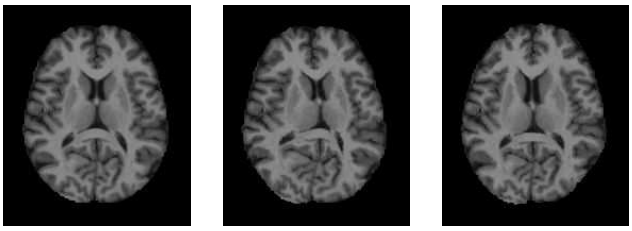

Figure 4. Medical imaging data. Left: Brain MR axial slice taken from the OASIS dataset. Center: Synthesized deformed slice. Right: Derived template, corresponding to the medoid of the respective computed cluster.

## V. Conclusion and future work

In view of the fact that a single common template or reference frame is not suitable for heterogenous image populations, we have proposed a framework capable of identifying such clusters requiring a separate atlas. In using spectral clustering, we avoid a common reference frame representation for our training set; only a one-time inter-subject distances computation is required. Another novelty of our approach is the distance metric proposed, resulting naturally from the diffeomorphism framework of [13] which guarantees convex combinations of valid warps to be also valid. Our distance incorporates residual information, albeit in an *ad hoc* manner; discovering more elaborate distance metrics can be a future research perspective. Another direction of research would be to estimate the number of population classes, which is required beforehand in spectral analysis. There is related work based on Laplacian matrix eigenvalue-based heuristics [9], [10].

Concluding, we showed by numerical evaluation that the proposed method can capture difficult anatomical differences and work well in both the non-medical and medical imaging context. Let us point out that our framework is straightforwardly extensible towards handling three-dimensional images and deformations - this will allow us to have the proposed method tested on $3D$ image training sets fully-comprising of real medical images.

## References


[1] C. V. Economo and G. N. Koskinas, *Die Cytoarchitektonik der Hirnrinde des Erwachsenen menschen*. Springer Verlag, Berlin, 1925.

[2] J. Talairach and P. Tournoux, *Co-planar stereotaxic atlas of the human brain*. Thieme Medical Publishers: New York, 1988.

[3] S. Joshi, B. Davis, M. Jomier, and G. Gerig, "Unbiased diffeomorphic atlas construction for computational anatomy," *NeuroImage*, vol. 23, pp. S151–S160, 2004.

[4] C. Studholme and V. Cardenas, "A template free approach to volumetric spatial normalization of brain anatomy," *Pattern Recognition Letters*, vol. 25, no. 10, pp. 1191–1202, 2004.

[5] P. M. Thompson, M. S. Mega, K. L. Narr, E. R. Sowell, R. E. Blanton, and A. W. Toga, "Brain image analysis and atlas construction," in *Handbook of Medical Image Analysis*, S. M. Fitzpatrick M, Ed. SPIE Press, Bellingham, WA, 2000.

[6] C. Taylor, C. Twining, and R. Davies, *Statistical models of shape: Optimisation and evaluation*. Springer-Verlag London, 2008.

[7] D. Blezek and J. Miller, "Atlas stratification," *Medical Image Analysis*, vol. 11, pp. 443–457, 2007.

[8] M. Sabuncu, S. Balci, M. Shenton, and P. Golland, "Image-driven population analysis through mixture modelling," *IEEE Transactions on Medical Imaging*, vol. 28, no. 9, 2009.

[9] U. von Luxburg, "A tutorial on spectral clustering," *Statistical Computing*, vol. 17, pp. 395–416, 2007.

[10] V. Chasanis, A. Likas, and N. Galatsanos, "Scene detection in videos using shot clustering and sequence alignment," *IEEE Transactions on Multimedia*, vol. 11, no. 1, pp. 89–100, 2009.

[11] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[12] P. Aljabar, D. Rueckert, and W. Crum, "Spectral clustering as a diagnostic tool in cross-sectional MR studies: An application to mild dementia," in *Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2008, pp. 442–449.

[13] T. Cootes, C. Twining, K. Babalola, and C. Taylor, "Diffeomorphic statistical shape models," *Image and Vision Computing*, vol. 26, pp. 326–332, 2008.

[14] S. Baloch and C. Davatzikos, "Morphological appearance manifolds in computational anatomy: Groupwise registration and morphological analysis," *NeuroImage*, vol. 45, pp. S73–S85, 2009.

[15] D. Marcus, T. Wang, J. Parker, J. Csernansky, J. Morris, and R. Buckner, "Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *Journal of Cognitive Neuroscience*, vol. 19, pp. 1498–1507, 2007.