

Multi-view Warped Mixtures for Shape Correspondence Analysis

Hernán F. García and Mauricio A. Álvarez

June 16, 2019

Abstract

Neuroimage correspondence analysis is critical in applications that model neurodegenerative disease progression. Establishing meaningful relations between non-rigid objects such as brain structures poses a challenging topic in the bio-imaging signal processing field. In this preliminary study, we introduce a novel nonlinear probabilistic multiview warped mixture model to infer shape correspondences of brain structures. We experimentally show how the model proposed can accurately establish meaningful relations between any pair of non-rigid shapes such as those brain structures related to neurodevelopmental in neonatal data.

1 Introduction

The correspondence problem in neuroimage analysis is a challenge research topic consisting in establishing meaningful relations between any pair of brain structures (static registration problem) [13], or analyzing temporal changes of a given neurodegenerative disease across time (dynamic analysis of brain structures) [7].

Most of the correspondence methods for medical image problems focus on computing different similarity metrics based on texture descriptors such as the bag-of-words features [2], largest common point-sets [1], and geodesic contours [12]. Though, these approaches only work over objects of the same size, which gives a poor accuracy in non-rigid matching processes [4].

Although similarity metrics could potentially capture shared information between objects, these metrics are not easy to define [5] since brain structures are non-rigid objects that exhibit morphological changes between subjects (brain volumetry over a population) and shape deformations over time in a neurodegenerative disease (i.e., Alzheimer and Parkinson) [6].

Instead of defining similarity metrics, an alternative approach consists in using unsupervised learning for object matching. These methods aim to establish meaningful correspondences in scenarios where a non-rigid object describes a given shape, and the similarity measure between objects cannot be computed [19]. Variational Bayesian matching [10] and Bayesian canonical correlation analysis [?] are some examples of these methods in which a given probabilistic framework is used to model features between objects and establish shape correspondences. Nonetheless, these methods only handle full correspondence frameworks (i.e., point-to-point matching) and linear analysis over the shape descriptors (i.e., appearance descriptors), which makes them unsuitable to model shared information between non-rigid objects, i.e., tissue shapes in MRI data [17] or volumes of brain structures for studying progression of Alzheimer's disease [15, 8]. High variability of these patterns such as curvedness and size makes it necessary to compute the correspondences between objects in a groupwise manner [14].

Probabilistic groupwise methods for unsupervised clustering have the benefit that we can model multiple view data without any correspondence information. Hence, we can compute shared information among domains instead of analyzing full correspondences by establishing linear relations between objects as Iwata et al. present in [9]. However, these relations are impractical in applications where non-linear representations of shape objects are needed (i.e., non-rigid matching tasks).

2 Multiview Warped Mixture Models

We can use the single view problem of the warped mixture model from [16] in which they warp a latent mixture of Gaussians into nonparametric cluster shapes. The low-dimensional latent mixture model summarizes the properties of the high-dimensional density manifolds describing the data.

Our idea is to introduce a model which warps a multiview latent mixture of Gaussians (possibly MRD) to produce nonparametric cluster shapes.¹

2.1 The model

Let us define a multi-view data set as $\mathcal{Y} = \{\mathbf{Y}^v\}_{v=1}^V$, where each view is defined as $\mathbf{Y}^v \in \mathbb{R}^{N_v \times D_v}$. This leads to the likelihood. Given mixture assignments, likelihood has only two parts: GP-LVM and GMM

$$p(\mathbf{Y}^\mathcal{V} | \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) = \prod_{v=1}^V p(\mathbf{Y}^v | \mathbf{X}, \boldsymbol{\theta}) \times \prod_i \sum_{c=1}^{\infty} \lambda_c \mathcal{N}(\mathbf{x}_i | \boldsymbol{\mu}_c, \mathbf{R}_c^{-1}), \quad \mathbf{x}_i \in \mathbf{Z}_c \quad (1)$$

Based on the iWMM (see [16]) our generative model generates multiple observations $\mathbf{Y}^\mathcal{V}$ according to the following generative process:

1. Draw mixture weights $\boldsymbol{\lambda} \sim \text{GEM}(\eta)$
2. For each cluster $c = 1, \dots, \infty$
 - (a) Draw precision $\mathbf{R}_c \sim \mathcal{W}(\mathbf{S}^{-1}, v)$
 - (b) Draw mean $\boldsymbol{\mu}_c \sim \mathcal{N}(\mathbf{u}, (r\mathbf{R}_c)^{-1})$
3. For each view $v = 1, \dots, \mathcal{V}$
 - (a) For each observation $n = 1, \dots, N_v$
 - i. Draw latent assignment $z_{nv} \sim \text{Mult}(\boldsymbol{\lambda})$
 - ii. Draw latent coordinates $\mathbf{x}_{nv} \sim \mathcal{N}(\boldsymbol{\mu}_{z_{nv}}, \mathbf{R}_{z_{nv}}^{-1})$
4. For each view $v = 1, \dots, \mathcal{V}$
 - (a) For each observed dimension $d = 1, \dots, D_v$
 - i. Draw function $\mathbf{f}_d^v \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}^v)$
5. For each view $v = 1, \dots, \mathcal{V}$
 - (a) For each observed dimension $d = 1, \dots, D_v$
 - i. Draw projection variable $w_d^v \sim \mathcal{N}(0, \rho_d^v)$
 - ii. For each observation $n = 1, \dots, N_v$
 - A. Draw feature $y_{nd}^v \sim \mathcal{N}(w_d^v f_d^v(\mathbf{x}_{nv}), \beta^{-1})$

We define the dimensionalities of our variables as:

- K : real number of clusters
- Q : dimensionality of the Latent Space
- D_v : dimensionality of the input data in the v -th view
- $\boldsymbol{\lambda} \in \mathbb{R}^{K \times 1}$
- $\mathbf{R}_c \in \mathbb{R}^{Q \times Q}$
- $\boldsymbol{\mu}_c \in \mathbb{R}^{Q \times 1}$

¹The possibly low-dimensional latent mixture model allows us to summarize the properties of the high-dimensional clusters (or density manifolds) describing the data. The number of manifolds, as well as the shape and dimension of each manifold is automatically inferred.

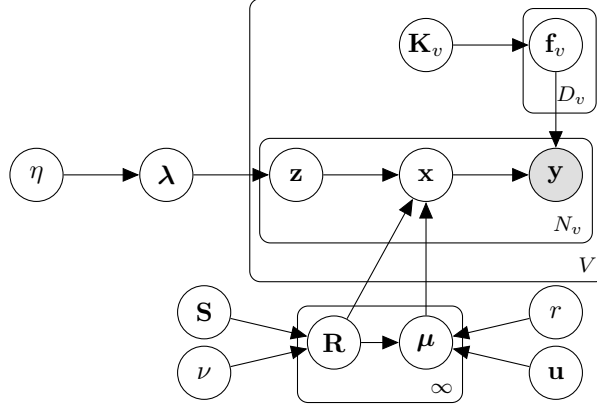


Figure 1: A graphical model representation of the multiview warped mixture model, where the shaded and unshaded nodes indicate observed and latent variables, respectively, and plates indicate repetition.

2.2 Latent Multi-view Warped Mixture Model

Our model is set as a multi-view Gaussian Process Latent Variable model as [11]. First we assume that observations are generated by mapping the latent coordinates through a set of smooth functions, over which Gaussian process priors are placed. Under the GPLVM, the probability of observations given the latent coordinates, integrating out the mapping functions, is defined as

$$p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) = \prod_{v=1}^V p(\mathbf{Y}^v|\mathbf{X}^v, \boldsymbol{\theta}^v) \quad (2)$$

$$= \prod_{v=1}^V \prod_{d=1}^{D_v} p(\mathbf{y}_d^v|\mathbf{X}^v, \boldsymbol{\theta}^v), \quad (3)$$

where \mathbf{y}_d^v represents the d th column of \mathbf{Y}^v and

$$p(\mathbf{y}_d^v|\mathbf{X}^v, \boldsymbol{\theta}^v) = \mathcal{N}(\mathbf{y}_d^v|\mathbf{0}, \beta^{-1}\mathbf{I} + w_d^2\mathbf{K}^v). \quad (4)$$

Our multi-view iWMM assumes that the latent coordinates (per View) are generated from a Dirichlet process mixture model. In particular, we use the following infinite Gaussian mixture model,

$$p(\mathbf{x}^v|\lambda_c, \boldsymbol{\mu}_c, \mathbf{R}_c) = \sum_{c=1}^{\infty} \lambda_c \mathcal{N}(\mathbf{x}^v|\boldsymbol{\mu}_c, \mathbf{R}_c^{-1}), \quad (5)$$

where λ_c , $\boldsymbol{\mu}_c$ and \mathbf{R}_c are the mixture weight, mean, and precision matrix of the c th mixture component.

As in the iWMM [16], we place Gaussian-Wishart priors on the Gaussian parameters $\{\boldsymbol{\mu}_c, \mathbf{R}_c\}$.

$$p(\boldsymbol{\mu}_c, \mathbf{R}_c) = \mathcal{N}(\boldsymbol{\mu}_c|\mathbf{u}, (r\mathbf{R}_c)^{-1}) \mathcal{W}(\mathbf{R}_c|\mathbf{S}^{-1}, \nu) \quad (6)$$

where \mathbf{u} is the mean of $\boldsymbol{\mu}_c$, r is the relative precision of $\boldsymbol{\mu}_c$, \mathbf{S}^{-1} is the scale matrix for \mathbf{R}_c , and ν is the number of degrees of freedom for \mathbf{R}_c .

By using conjugate Gaussian-Wishart priors for the parameters of the Gaussian mixture components, we can analytically integrate out those parameters, given the assignments of points to components. Let z_n^v be the latent assignment of the n th object in the v th view. The probability of latent coordinates \mathbf{X} given latent assignments $\mathbf{Z}^v = (z_1, \dots, z_{N_v})$ is obtained by integrating out the Gaussian parameters $\{\boldsymbol{\mu}_c, \mathbf{R}_c\}$ as follows:

$$p(\mathbf{X}|\mathbf{Z}, \mathbf{S}, \nu, r) = \prod_{c=1}^{\infty} \pi^{-\frac{\sum_v N_{vc}Q}{2}} \frac{r^{Q/2} |\mathbf{S}|^{\nu/2}}{r_c^{Q/2} |\mathbf{S}_c|^{\nu_c/2}} \prod_{q=1}^Q \frac{\Gamma(\frac{\nu_c+1-q}{2})}{\Gamma(\frac{\nu+1-q}{2})}, \quad (7)$$

where N_{vc} is the number of objects in the v th view assigned to the c th cluster, $\Gamma(\cdot)$ is the Gamma function and

$$\begin{aligned} r_c &= r + \sum_v N_{vc}, \quad \nu_c = \nu + \sum_v N_{vc}, \\ \mathbf{u}_c &= \frac{r\mathbf{u} + \sum_{v=1}^V \sum_{n:z_{nv}=c} \mathbf{x}_{nv}}{r + \sum_v N_{vc}}, \\ \mathbf{S}_c &= \mathbf{S} + \sum_{v=1}^V \sum_{n:z_{nv}=c} \mathbf{x}_{nv} \mathbf{x}_{nv}^\top + r\mathbf{u}\mathbf{u}^\top - r_c \mathbf{u}_c \mathbf{u}_c^\top, \end{aligned}$$

are the posterior Gaussian-Wishart parameters of the c th component. As in [16], we use a Dirichlet process with concentration parameter η for infinite mixture modeling in the latent space.

The probability of \mathbf{Z} is given as follows

$$p(\mathbf{Z}|\eta) = \prod_{v=1}^V \frac{\eta^C \prod_{c=1}^C (N_{vc} - 1)!}{\eta(\eta + 1) \cdots (\eta + N_v - 1)}, \quad (8)$$

where C is the number of components for which $N_{vc} > 0$. The joint distribution is given by

$$p(\mathbf{Y}, \mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}, \nu, \mathbf{u}, r\eta) = \prod_{v=1}^V p(\mathbf{Y}^v|\mathbf{X}^v, \boldsymbol{\theta}^v) p(\mathbf{X}^v|\mathbf{Z}_v, \mathbf{S}_v, \nu, \mathbf{u}, r) p(\mathbf{Z}_v|\eta). \quad (9)$$

2.3 Inference

We infer the posterior distribution of the latent coordinates $\mathbf{X} = \{\mathbf{X}^v\}_v^V$ and cluster assignments \mathbf{Z}^v using Markov chain Monte Carlo (MCMC). In particular, we alternate collapsed Gibbs sampling of \mathbf{Z} , and hybrid Monte Carlo sampling of \mathbf{X} . Given \mathbf{X} , we can efficiently sample \mathbf{Z}^v using collapsed Gibbs sampling, integrating out the mixture parameters. Given \mathbf{Z}^v , we can calculate the gradient of the unnormalized posterior distribution of \mathbf{X} , integrating over warping functions. This gradient allows us to sample \mathbf{X} using hybrid Monte Carlo.

First, we explain collapsed Gibbs sampling for \mathbf{Z} . Given a sample of \mathbf{X} , $p(\mathbf{Z}|\mathbf{X}, \mathbf{S}, v, \mathbf{u}, r, \eta)$ does not depend on \mathbf{Y} . This lets resample cluster assignments, integrating out the iGMM likelihood in close form. Given the current state of all but one latent component z_n , a new value for z_n is sampled from the following probability:

$$p(z_{nv} = c|\mathbf{X}, \mathbf{Z}_{\setminus nv}, \mathbf{S}, v, \mathbf{u}, r, \eta) \propto \begin{cases} N_{vc \setminus nv} \cdot p(\mathbf{x}_{nv}|\mathbf{X}_c, \mathbf{S}, v, \mathbf{u}, r) & \text{existing components} \\ \eta \cdot p(\mathbf{x}_{nv}|\mathbf{S}, v, \mathbf{u}, r) & \text{a new cluster} \end{cases}$$

where $\mathbf{X}_c = \{\mathbf{x}_{nv} | z_{nv} = c\}$ is the set of latent coordinates assigned to the c^{th} component, and $\setminus nv$ represents the value or set when excluding the n -th observation in the v -th view. We can analytically calculate $p(\mathbf{x}_{nv}|\mathbf{X}_c, \mathbf{S}, v, \mathbf{u}, r)$ as follows:

$$p(\mathbf{x}_{nv}|\mathbf{X}_c, \mathbf{S}, v, \mathbf{u}, r) = \pi^{-\frac{N_{vc \setminus n} Q}{2}} \frac{r^{Q/2} |\mathbf{S}|^{\nu/2}}{r_c^{Q/2} |\mathbf{S}_c|^{\nu_c/2}} \prod_{q=1}^Q \frac{\Gamma(\frac{\nu_c + 1 - q}{2})}{\Gamma(\frac{\nu + 1 - q}{2})}$$

3 Results

3.1 Clustering performance on real datasets

We first test our model on common clustering ML datasets shown in table 1. Results show that by using nonlinear models to perform the clustering task the cluster assignment becomes more accurate (None of these datasets can be appropriately clustered through linear approaches such as GMMs).

Table 1: Average Rand index for evaluating clustering performance.

Approach	Database					
	Wine	2-curve	3-semi	2-circle	Pinwheel	Vowel
MV-WMM($Q = 2$)	0.68 ± 0.03	0.83 ± 0.02	0.83 ± 0.01	0.88 ± 0.02	0.87 ± 0.02	0.65 ± 0.01
MV-WMM($Q = D$)	0.85 ± 0.02	0.83 ± 0.02	0.83 ± 0.01	0.88 ± 0.02	0.87 ± 0.02	0.73 ± 0.02

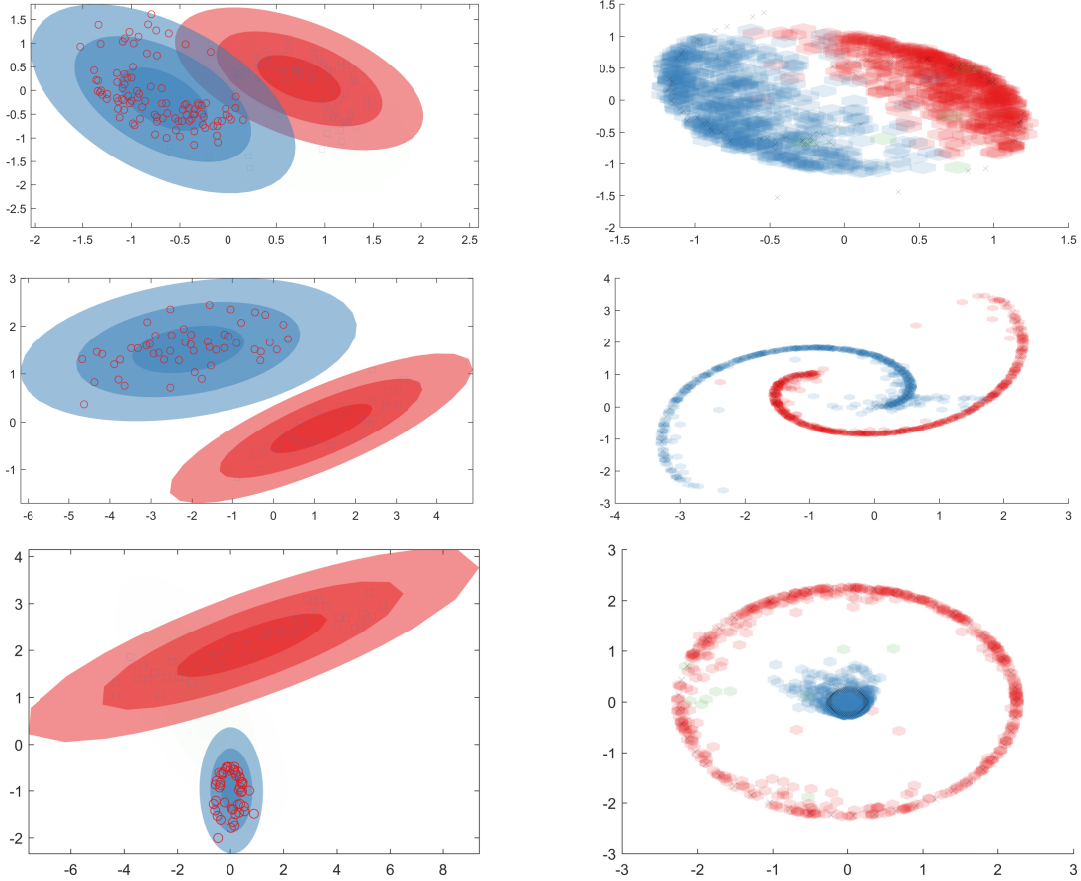


Figure 2: Experimental results for the Pinwheel, spiral and circles dataset. Current mixture parameters, along with the latent positions (left). Original data and predicted assignments (right).

3.2 Comparison with linear approaches

Secondly, we test the performance of our approaches (both NL-UCM and MV-WMM) regarding the adjusted Rand index (we report both average and standard deviation), to quantify the similarity between the inferred clusters [9] and the true labels. For comparison, we use unsupervised clustering matching (UCM) [9], k-means (KM), and convex kernelized sorting (CKS) [?]. Table 2 shows that our approaches outperforms the state-of-the-art methods for unsupervised clustering for the three databases. The results also show that by mapping the observed data through non-linear mappings, the models can handle real-world datasets with better performance than linear approaches as in the case of NL-UCM and MV-WMM (i.e., 0.17 and 0.13 for the MNIST dataset against 0.085 obtained from the UCM method).

Table 2: Adjusted Rand index of the proposed method against the state-of-the-art methods for unsupervised clustering.

Database	Approach				
	UCM	KM	KM-CKS	NL-UCM	MV-WMM
Iris	0.383 ± 0.189	0.224 ± 0.0910	0.254 ± 0.154	0.546 ± 0.080	0.498 ± 0.001
Glass	0.160 ± 0.020	0.050 ± 0.008	0.052 ± 0.011	0.378 ± 0.045	0.384 ± 0.003
MNIST	0.085 ± 0.016	0.030 ± 0.007	0.037 ± 0.008	0.167 ± 0.013	0.133 ± 0.004

3.3 Neurodegenerative brain dataset

With the aim to model shape variability in Brain structures, we test our model on real medical image data. Here we used the MRI *DB-UTP* database from the Universidad Tecnológica de Pereira, COL. This database contains volumetric MRI data from four patients with Parkinson’s disease (at earlier and advanced stage of the disease). The database was labeled by neurosurgeons from *NEUROCENTRO*: The Institute of Parkinson and Epilepsy, located in Pereira-Colombia. The database contains *T1* sequences with $1mm \times 1mm \times 1mm$ voxel size and slices of 512×512 pixels. The atlas was derived from a volumetric *T1*-weighted MR-scans, using semi-automated image segmentation, and three-dimensional reconstruction techniques. The current version of this dataset consists of 1) the original volumetric whole brain MRI of the volunteers; 2) a set of detailed label maps and 3) the three-dimensional models of the labeled anatomical brain structures.

To establish groupwise correspondences between brain structures, we used SI-HKS as shape descriptors [3]. We evaluate our model by using three relevant brain structures in the Alzheimer’s disease such as the ventricle, thalamus, and putamen. Figures 3, 4 and 5 show the experimental results of the brain correspondence analysis. These experiments test our framework working on three brain structures at different times of the disease (early and advanced stage). From the results, it can be noticed that even when the brain volumetry of a given shape (i.e., see Ventricle results in figure 3) has lost part of their mass as consequence of the neurodegenerative process, our model is capable of establishing relevant correspondences between brain structures.

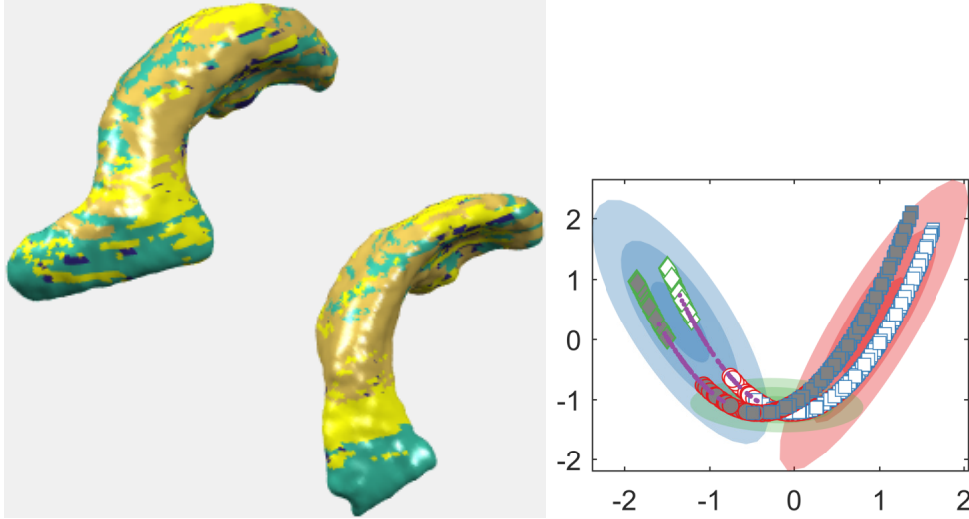


Figure 3: Experimental results for the brain structures dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the ventricle. Average Rand Index 0.6071

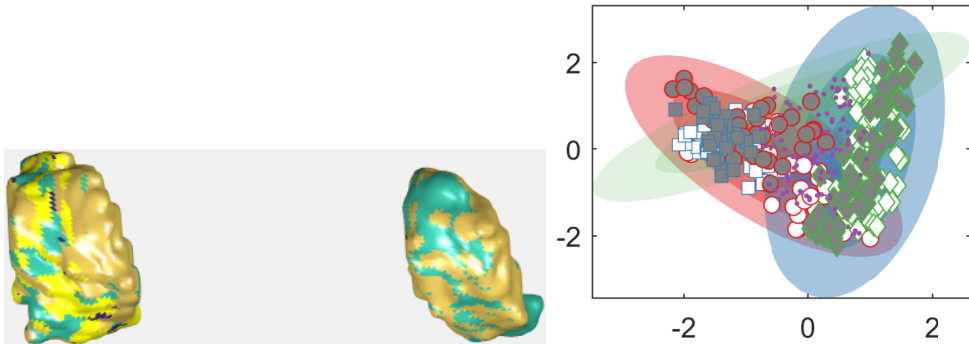


Figure 4: Experimental results for the brain structures dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the thalamus. Average Rand Index 0.6467

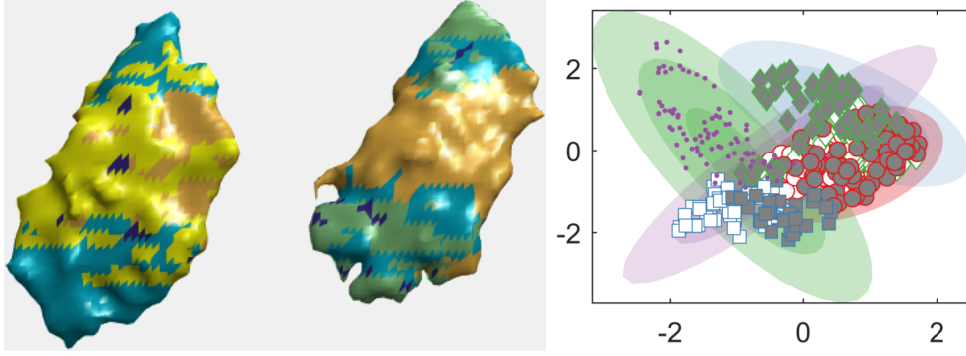


Figure 5: Experimental results for the brain structures dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the putamen. Average Rand Index 0.6467

3.4 Neurodevelopmental dataset: In-utero brain model

Finally, we test our model on a couple of in-utero MRI scans of a fetus acquired at two-time points, 25 and 31 weeks. Figure 6 shows groupwise correspondences between the neonatal brain volumes. The results show that matched clusters between shapes are related to slight changes as part of the neurodevelopmental outcome. Besides, different clustered regions are related to significant changes over the brain volume as part of normal development. Finally, these preliminary results can lead to a new type of scores to predict Neurodevelopmental Outcome that uses unsupervised multiview learning to establish meaningful relationships between a set of MRI scans [18].

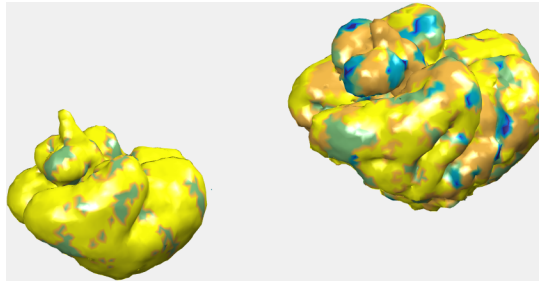


Figure 6: Experimental results for the brain Neurodevelopmental dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the brain volume. Average Rand Index 0.6948

References

- [1] D. Aiger, N. J. Mitra, and D. Cohen-Or. 4-points congruent sets for robust surface registration. *ACM Transactions on Graphics*, 27(3):#85, 1–10, 2008.
- [2] Alexander M. Bronstein, Michael M. Bronstein, Leonidas J. Guibas, and Maks Ovsjanikov. Shape google: Geometric words and expressions for invariant shape retrieval. *ACM Trans. Graph.*, 30(1):1:1–1:20, February 2011.
- [3] Michael M. Bronstein and Iasonas Kokkinos. Scale-invariant kernel signatures for non-rigid shape recognition. In *In Proc. CVPR*, 2010.
- [4] Alan Brunton, Augusto Salazar, Timo Bolkart, and Stefanie Wuhler. Review of statistical shape spaces for 3d data with comparative analysis for human faces. *Computer Vision and Image Understanding*, 128(0):1 – 17, 2014.
- [5] Xavier Cortés and Francesc Serratos. An interactive method for the image alignment problem based on partially supervised correspondence. *Expert Syst. Appl.*, 42(1):179–192, 2015.

- [6] A. Cosa, S. Canals, A. Valles-Lluch, and D. Moratal. Unsupervised segmentation of brain regions with similar microstructural properties: Application to alcoholism. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 1053–1056, July 2013.
- [7] S. Durrleman, T Fletcher, G. Gerig, and M Niethammer, editors. *Spatio-temporal Image Analysis for Longitudinal and Time-Series Image Data*, volume 8682. Springer, 2014.
- [8] Derek Hill. Neuroimaging to assess safety and efficacy of ad therapies. *Expert Opinion on Investigational Drugs*, 19(1):23–26, 2010. PMID: 19947893.
- [9] Tomoharu Iwata, Tsutomu Hirao, and Naonori Ueda. Probabilistic latent variable models for unsupervised many-to-many object matching. *Information Processing and Management*, 52(4):682 – 697, 2016.
- [10] Arto Klami. Variational bayesian matching. In *Proceedings of the 4th Asian Conference on Machine Learning, ACML 2012, Singapore, Singapore, November 4-6, 2012*, pages 205–220, 2012.
- [11] Neil Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In *In NIPS*, page 2004, 2003.
- [12] Luming Liang, Andrzej Szymczak, and Mingqiang Wei. Geodesic spin contour for partial near-isometric matching. *Computers & Graphics*, 46:156–171, 2015.
- [13] Dongdong Lin, Vince D. Calhoun, and Yu-Ping Wang. Correspondence between fmri and SNP data by group sparse canonical correlation analysis. *Medical Image Analysis*, 18(6):891–902, 2014.
- [14] K. A. Sidorov, S. Richmond, and D. Marshall. Efficient groupwise non-rigid registration of textured surfaces. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR ’11*, pages 2401–2408, Washington, DC, USA, 2011. IEEE Computer Society.
- [15] Paul M. Thompson, Kiralee M. Hayashi, Greig de Zubicaray, Andrew L. Janke, Stephen E. Rose, James Semple, David Herman, Michael S. Hong, Stephanie S. Dittmer, David M. Doddrell, and Arthur W. Toga. Dynamics of gray matter loss in alzheimer’s disease. *The Journal of Neuroscience*, 23(3):994–1005, 2003.
- [16] Zoubin Ghahramani Tomoharu Iwata, David Duvenaud. Warped mixtures for nonparametric cluster shapes. In *29th Conference on Uncertainty in Artificial Intelligence*, pages 311–319, 2013.
- [17] Oliver van Kaick, Andrea Tagliasacchi, Oana Sidi, Hao Zhang, Daniel Cohen-Or, Lior Wolf, , and Ghassan Hamarneh. Prior knowledge for part correspondence. *Computer Graphics Forum (Proc. Eurographics)*, 30(2):553–562, 2011.
- [18] Lauren C. Weeke, Floris Groenendaal, Kalyani Mudigonda, Mats Blennow, Maarten H. Lequin, Linda C. Meiners, Ingrid C. van Haastert, Manon J. Benders, Boubou Hallberg, and Linda S. de Vries. A novel magnetic resonance imaging score predicts neurodevelopmental outcome after perinatal asphyxia and therapeutic hypothermia. *The Journal of Pediatrics*, 192:33–40, 1 2018.
- [19] Xu Yang, Hong Qiao, and Zhi-Yong Liu. Partial correspondence based on subgraph matching. *Neurocomputing*, 122:193 – 197, 2013. Advances in cognitive and ubiquitous computing Selected papers from the Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2012).