NONLINEAR PROBABILISTIC LATENT VARIABLE MODELS FOR GROUPWISE CORRESPONDENCE ANALYSIS IN BRAIN STRUCTURES

Hernán F. García[⋆]*, Álvaro A. Orozco*[⋆]*, and Mauricio A. Álvarez*[†]

[⋆] Automatics Research Group, Universidad Tecnológica de Pereira, Pereira, Colombia
[†] Department of Computer Science, The University of Sheffield, Sheffield, UK

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 bioimaging signal processing field. In this paper, we introduce a novel nonlinear probabilistic latent variable model approach to infer shape correspondences of brain structures. To this end, we perform an unsupervised clustering process that is automatically carried out by a nonlinear kernelized probabilistic latent variable model. The kernel embeddings are accomplished by using random Fourier features as nonlinear mappings of 3D shape descriptors. We experimentally show how the model proposed can accurately establish meaningful relations between any pair of nonrigid shapes such as those brain structures related to the Alzheimer's disease.

Index Terms— Correspondence problem, Probabilistic latent variable models, Random Fourier features, Neuroimage analysis.

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) [1], or analyzing temporal changes of a given neurodegenerative disease across time (dynamic analysis of brain structures) [2].

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 [3], largest common point-sets [4], and geodesic contours [5]. Though, most of these approaches only work over objects of the same size, which gives a poor accuracy in nonrigid matching processes [6].

Although similarity metrics could potentially capture shared information between objects, these metrics are not easy to define [7] since brain structures are nonrigid 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) [8].

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 nonrigid object describes a given shape, and the similarity measure between objects cannot be computed [9]. Variational Bayesian matching [10] and Bayesian canonical correlation analysis [11] 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 [12] or volumes of brain structures for studying progression of Alzheimer's disease [13, 14]. High variability of these patterns such as curvedness and size makes it necessary to compute the correspondences between objects in a groupwise manner [15].

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 point-to-point relations [16]. In this paper, we introduce a nonlinear version of the model proposed by Iwata et al. [16]. In particular, we provide a method for shape correspondence analysis based on nonlinear unsupervised clustering of groupwise 3D shape descriptors. The clustering process is carried out by a nonlinear probabilistic latent variable model, in which we use random Fourier features of the input data observations [17]. In other words, we extend the many-to-many object matching proposed by [16] using Hilbert space embeddings of the input data [17]. Once the model is defined, we provide an stochastic EM algorithm for computing the necessary posterior distributions of the probabilistic model.

The rest of paper proceeds as follows. Section 2 presents the nonlinear unsupervised clustering approach. In section 3, we report a comparison of our method with commonly used approaches for unsupervised clustering. Then, we show the

experimental results for the brain structures dataset. The paper concludes in section 4 with some conclusions and future works of the proposed method.

2. MATERIALS AND METHODS

2.1. Probabilistic Non-Linear Latent Variable Model for Groupwise Correspondence

Let us define the input data set in D domains as $\mathbf{X} = \{\mathbf{X}_d\}_{d=1}^{D}$, where $\mathbf{X}_d = \{\mathbf{x}_{dn}\}_{n=1}^{N_d}$ is a set of objects in the dth domain, and $\mathbf{x}_{dn} \in \mathcal{R}^{M_d}$ is the input vector of the nth object in the dth domain. As in nonlinear latent variable models, we want to map the input data to a feature space through a nonlinear map $\boldsymbol{\phi}(\cdot)$, so we can compute clusters of feature vectors by using nonlinear functions [17].

As we are unaware of any correspondence between feature sets $\boldsymbol{\Phi} = \{\boldsymbol{\Phi}_d\}_{d=1}^{D}$ in different domains, we set different number of feature vectors $\boldsymbol{\Phi}_d = \{\boldsymbol{\phi}(\mathbf{x}_{dn})\}_{n=1}^{N_d}$, and different dimensionalities L_d such that $\boldsymbol{\phi}(\mathbf{x}_{dn}) : \mathcal{R}^{M_d} \rightarrow \mathcal{R}^{L_d}$. Our approach assumes that we can find an infinite number of correspondences between feature vectors, and each correspondence j has a latent feature vector $\boldsymbol{\zeta}_j \in \mathcal{R}^K$ in a latent space of dimension K. Thus, feature vectors that have the same cluster assignments $\mathbf{s}_{dn}$, or are related by the same latent feature vector, establish a meaningful correspondence.

Each feature vector in $\boldsymbol{\phi}(\mathbf{x}_{dn}) \in \mathcal{H}$ in the dth domain is generated depending on the domain-specific projection matrix $\boldsymbol{B}_d \in \mathcal{R}^{L_d \times K}$ and the latent feature vector $\boldsymbol{\zeta}_{s_{dn}}$ that is selected from a set of latent feature vectors $\mathbf{Z} = \{\boldsymbol{\zeta}_j\}_{j=1}^{\infty}$. Here, $s_{dn} = \{1, \ldots, \infty\}$ is the cluster assignment of feature vector $\boldsymbol{\phi}(\mathbf{x}_{dn})$. Then, by using a latent space representation of an infinite Gaussian mixture model, we define the probability of a feature vector $\boldsymbol{\phi}(\mathbf{x}_{dn})$ as

$$p(\boldsymbol{\phi}(\mathbf{x}_{dn})|\mathbf{Z}, \boldsymbol{\mathcal{W}}, \boldsymbol{\theta}) = \sum_{j=1}^{\infty} \theta_j \mathcal{N}\left(\boldsymbol{\phi}(\mathbf{x}_{dn}) | \boldsymbol{B}_d \boldsymbol{\zeta}_j, \alpha^{-1}\mathbf{I}\right),$$

where $\boldsymbol{\mathcal{W}} = \{\boldsymbol{B}_d\}_{d=1}^{D}$ is a set of projections matrices, $\boldsymbol{\theta} = (\theta_j)_{j=1}^{\infty}$ are the mixture weights, θ_j represents the probability that the jth cluster is chosen and α is a precision parameter. By employing different projection matrices in Hilbert space for each feature vector (domain-specific), we can handle multiple feature sets with nonlinear properties and different dimensionalities (i.e., size of the brain structures). Figure 1 shows the scheme of the proposed model, in which we describe the relationship between feature vectors and latent feature vectors in Hilbert space.

As in [16], we use a stick-breaking process to set the mixture weights $\boldsymbol{\theta}$ for a Dirichlet process with concentration parameter γ. The joint probability of the feature vectors $\boldsymbol{\Phi}$, and the cluster assignments $\mathbf{S} = \left\{\{\mathbf{s}_{dn}\}_{n=1}^{N_d}\right\}_{d=1}^{D}$ is given by

$$p(\boldsymbol{\Phi}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma) = p(\mathbf{S}|\gamma)\, p(\boldsymbol{\Phi}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r), \quad (1)$$

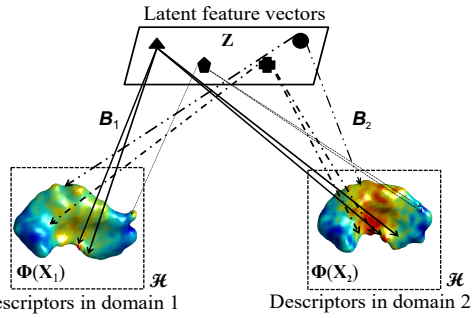


Fig. 1. Scheme for the unsupervised nonlinear clustering method for groupwise correspondence analysis. The figure shows an example of establishing correspondences in Hilbert space for two brain structures (left putamen).

where a, b and r are the hyperparameters.

By marginalizing out the mixture weights $\boldsymbol{\theta}$, $p(\mathbf{S}|\gamma)$ becomes

$$p(\mathbf{S}|\gamma) = \frac{\gamma^J \prod_{j=1}^{J} (N_{\cdot j} - 1)!}{\gamma(\gamma+1)\cdots(\gamma+N-1)},$$

where $N = \sum_{d=1}^{D} N_d$ is the total number of feature vectors, $N_{\cdot j}$ represents the number of feature vectors assigned to the cluster j, and J is the number of clusters that satisfies $N_{\cdot j} > 0$.

For our non-linear model, we give the derivation of the likelihood in (1), in which latent feature vectors $\mathbf{Z}$ and precision parameter α are analytically integrated out. The resulting expression is defined as

$$p(\boldsymbol{\Phi}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r) = (2\pi)^{-\frac{\sum_d L_d N_d}{2}} r^{\frac{KJ}{2}} \frac{b^a}{b'^{a'}} \times \frac{\Gamma(a')}{\Gamma(a)} \prod_{j=1}^{J} |\boldsymbol{\Lambda}_j|^{1/2}. \quad (2)$$

Here, $a' = a + \frac{\sum_d L_d N_d}{2}$,

$$b' = b + \frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d} \boldsymbol{\phi}(\mathbf{x}_{dn})^{\top} \boldsymbol{\phi}(\mathbf{x}_{dn}) - \frac{1}{2}\sum_{j=1}^{J} \boldsymbol{\mu}_j^{\top} \boldsymbol{\Lambda}_j^{-1} \boldsymbol{\mu}_j, \quad (3)$$

and

$$\boldsymbol{\mu}_j = \mathbf{C}_j \sum_{d=1}^{D} \boldsymbol{B}_d^{\top} \sum_{n:\mathbf{s}_{dn}=j} \boldsymbol{\phi}(\mathbf{x}_{dn}),$$
$$\boldsymbol{\Lambda}_j^{-1} = \sum_{d=1}^{D} N_{dj} \boldsymbol{B}_d^{\top} \boldsymbol{B}_d + r\mathbf{I}, \quad (4)$$

where r is a parameter for controlling the precision of the latent feature vectors $\mathbf{Z}$, and N_{dj} is the number of feature vectors assigned to cluster j in the d domain.

2.2. Inference

To marginalize out the latent feature vectors $\mathbf{Z}$, and the precision parameter α, we use stochastic EM algorithm [16]. Hence, we alternatively iterate collapsed Gibbs sampling for the cluster assignments $\mathbf{S}$, and maximum joint likelihood estimation of the projection matrices $\boldsymbol{\mathcal{W}}$. In the E-step, a new value for $\mathbf{s}_{dn}$ is sampled from

$$p\left(s_{dn}=j|\boldsymbol{\Phi},\mathbf{S}_{\backslash dn},\boldsymbol{\mathcal{W}},a,b,r,\gamma\right)\propto\frac{p\left(s_{dn}=j,\mathbf{S}_{\backslash dn}|\gamma\right)}{p\left(\mathbf{S}_{\backslash dn}|\gamma\right)}$$
$$\times\frac{p\left(\boldsymbol{\Phi}|s_{dn}=j,\mathbf{S}_{\backslash dn},\boldsymbol{\mathcal{W}},a,b,r\right)}{p\left(\boldsymbol{\Phi}_{\backslash dn}|\mathbf{S}_{\backslash dn},\boldsymbol{\mathcal{W}},a,b,r\right)},$$

where $\backslash dn$ represents a value excluding the nth feature vector in the dth domain. The first factor in the expression above is given by

$$\frac{p\left(s_{dn}=j,\mathbf{S}_{\backslash dn}|\gamma\right)}{p\left(\mathbf{S}_{\backslash dn}|\gamma\right)}=\begin{cases}\frac{N_{.j\backslash dn}}{N-1+\gamma} & \text{for an existing cluster}\\ \frac{\gamma}{N-1+\gamma} & \text{for a new cluster.}\end{cases}$$

In the M-step, the projection matrices $\boldsymbol{\mathcal{W}}$ are estimated by maximizing the logarithm of the joint likelihood (1). The gradient of the joint likelihood is computed by

$$\frac{\partial\log p\left(\boldsymbol{\Phi},\mathbf{S}|\boldsymbol{\mathcal{W}},a,b,r,\gamma\right)}{\partial\boldsymbol{B}_d}=-\frac{a'}{b'}\left[\sum_{j=1}^{J}\left\{N_{dj}\boldsymbol{B}_d\boldsymbol{\mu}_j\boldsymbol{\mu}_j^{\top}\right.\right.$$
$$\left.\left.-\sum_{n:s_{dn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)\boldsymbol{\mu}_j^{\top}\right\}\right]-\sum_{j=1}^{J}N_{dj}\boldsymbol{B}_d\boldsymbol{\Lambda}_j.$$

2.3. Random Fourier Features

Since the model parameters depicted above depend on the feature vectors $\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)$, the expression for $\boldsymbol{\mu}_j$ in the equation (4) becomes intractable for kernelized methods[1]. As in [18], we propose to approximate the mapping functions $\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)$ by computing a randomized feature map $\boldsymbol{\varphi}\left(\mathbf{x}_{dn}\right):\mathcal{R}^{M_d}\rightarrow\mathcal{R}^{L_d}$ so that the inner product in equation (3) ensures that we can approximate the kernel, $k\left(\mathbf{x},\mathbf{x}'\right)=\left\langle\boldsymbol{\phi}\left(\mathbf{x}\right),\boldsymbol{\phi}\left(\mathbf{x}'\right)\right\rangle\approx\boldsymbol{\varphi}\left(\mathbf{x}_{dn}\right)^{\top}\boldsymbol{\varphi}\left(\mathbf{x}_{dn}\right)$ [17]. Consequently, we compute these feature vectors by using random Fourier bases as

$$\boldsymbol{\varphi}\left(\mathbf{x}_{dn}\right)\equiv\sqrt{\frac{2}{L_d}}\begin{bmatrix}\cos\left(\boldsymbol{\omega}_1^{\top}\mathbf{x}_{dn}+\upsilon_1\right)\\ \vdots\\ \cos\left(\boldsymbol{\omega}_{L_d}^{\top}\mathbf{x}_{dn}+\upsilon_{L_d}\right)\end{bmatrix},\qquad(5)$$

where $\left\{\boldsymbol{\omega}_m\sim\mathcal{N}\left(\mathbf{0},\beta^{-1}\mathbf{I}\right)\right\}_{m=1}^{L_d}$ and υ_m is draw from the uniform distribution $\left\{\upsilon_m\sim\mathcal{U}\left(0,2\pi\right)\right\}_{m=1}^{L_d}$.

[1] A given kernel induces an inner product between infinite dimensional feature vectors.

2.4. Databases

2.4.1. Real-world datasets

First, we test our method with three well-known machine learning datasets such as Iris, Glass, and MNIST[2]. We set up our experiments by randomly splitting the input data (i.e., features of the datasets) into two domains as Iwata et al. did for their experiments in [16].

2.4.2. Brain structures dataset

For the neuroimage analysis, we used the MRI *DB-UTP* database from the Technological University of Pereira. 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 $1mm\times1mm\times1mm$ voxel size and slices of 512x512 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.

3. RESULTS

To explore the accuracy of our approach, we first present a comparative analysis of unsupervised clustering methods over well-known machine learning databases. Then, we discuss the benefits of performing probabilistic correspondence analysis over neuroimaging data.

3.1. Comparison with linear approaches

First, we test the performance of our approach regarding the adjusted Rand index (we report both average and standard deviation), to quantify the similarity between the inferred clusters [16]. For comparison, we use unsupervised clustering matching (UCM)[16], k-means (KM), and convex kernelized sorting (CKS) [19]. Table 1 shows that our approach 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 random feature expansions, the model can handle real-world datasets with better performance than linear approaches (i.e., 0.17 for the MNIST dataset against 0.085 obtained from the UCM method).

[2] We use the database of handwritten digits developed by LeCun *et. al.* available at `http://yann.lecun.com/exdb/mnist/`

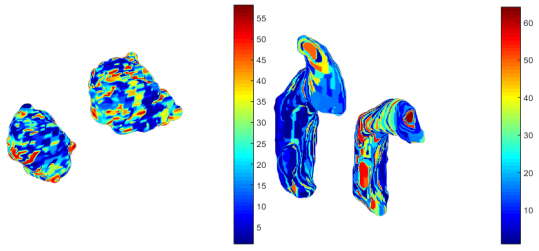
Table 1. Adjusted Rand index of the proposed method against the state-of-the-art methods for unsupervised clustering.

	Approach			
Database	UCM	KM	KM-CKS	**Ours**
Iris	0.383 ± 0.189	0.224 ± 0.0910	0.254 ± 0.154	$\mathbf{0.546 \pm 0.080}$
Glass	0.160 ± 0.020	0.050 ± 0.008	0.052 ± 0.011	$\mathbf{0.378 \pm 0.045}$
MNIST	0.085 ± 0.016	0.030 ± 0.007	0.037 ± 0.008	$\mathbf{0.167 \pm 0.013}$

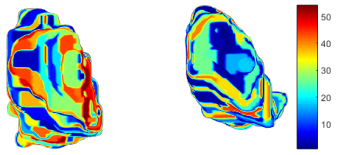
3.2. Groupwise shape correspondences

To establish groupwise correspondences between brain structures, we compute 3D shape descriptors based on scale-invariant Heat Kernel Signatures as in [20]. From these shape descriptors, we perform a random feature expansion to compute the features vectors. We set each domain as a 3D shape descriptor for a given brain structure. We evaluate our model by using three relevant brain structures in the Alzheimer's disease such as the ventricle, thalamus, and putamen.

Figure 2 shows the experimental results of the brain correspondence analysis. These experiments show our framework working with two 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 Putamen results in figure 2(a)) has lost part of their mass as consequence of the neurodegenerative process, our model is capable of establishing relevant correspondences between brain structures.



(a) Putamen (b) Ventricle



(c) Thalamus

Fig. 2. Experimental results of brain correspondences analysis using the proposed method. The figure shows a comparison between brain structures at different stages of the disease (left and right depicts early and advance stage of the disease).

Finally, the table 2 shows both mean and standard deviations computed from ground-truth correspondences established through Voronoi tessellation. Here, the results show that our model has better performance than the unsupervised linear approach. The results prove that by modeling nonlinear mapping functions of the shape descriptors, the model can establish meaningful correspondences between brain structures.

Table 2. Adjusted Rand index for the groupwise correspondence analysis on brain structures.

Brain Structure	UCM	**Ours**
Ventricle	0.092 ± 0.015	$\mathbf{0.287 \pm 0.035}$
Putamen	0.098 ± 0.013	$\mathbf{0.312 \pm 0.023}$
Thalamus	0.157 ± 0.003	$\mathbf{0.332 \pm 0.017}$

4. CONCLUSIONS

In this paper, we have presented an unsupervised clustering method for brain correspondence analysis through random Features expansion. We demonstrated that by using random Fourier features, the clustering process becomes more accurate in comparison with common state-of-the-art methods. Besides, the latent feature space shared among domains holds more relevant information about the nonlinear mapping of the random feature expansion. Moreover, the experimental results showed that our approach establishes meaningful correspondences between 3D brain structures. In addition, since the inferred correspondences fits a ground-truth Voronoi tessellation accurately, our method proved to be useful in applications derived from matching processes. As future works, we plan to analyze other inference methods based on variational inference to make our model fully Bayesian.

Acknowledgments

This research is developed as part of the project financed by COLCIENCIAS-COL with code 111074455860. H.F. García is funded by Colciencias under the program: *Formación de alto nivel para la ciencia, la tecnología y la innovación - Convocatoria 617 de 2013*. H.F García would also thank The University of Sheffield for supporting him during the Postgraduate Research Visit in the Department of Computer Science.

5. REFERENCES

[1] Dongdong Lin, Vince D. Calhoun, and Yu-Ping Wang, "Correspondence between fmri and SNP data by group sparse canonical correlation analysis," *Medical Image Analysis*, vol. 18, no. 6, pp. 891–902, 2014.

[2] S. Durrleman, T Fletcher, G. Gerig, and M Niethammer, Eds., *Spatio-temporal Image Analysis for Longitudinal and Time-Series Image Data*, vol. 8682. Springer, 2014.

[3] 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.*, vol. 30, no. 1, pp. 1:1–1:20, Feb. 2011.

[4] D. Aiger, N. J. Mitra, and D. Cohen-Or, "4-points congruent sets for robust surface registration," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. #85, 1–10, 2008.

[5] Luming Liang, Andrzej Szymczak, and Mingqiang Wei, "Geodesic spin contour for partial near-isometric matching," *Computers & Graphics*, vol. 46, pp. 156–171, 2015.

[6] Alan Brunton, Augusto Salazar, Timo Bolkart, and Stefanie Wuhrer, "Review of statistical shape spaces for 3d data with comparative analysis for human faces," *Computer Vision and Image Understanding*, vol. 128, no. 0, pp. 1 – 17, 2014.

[7] Xavier Cortés and Francesc Serratosa, "An interactive method for the image alignment problem based on partially supervised correspondence," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 179–192, 2015.

[8] 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*, July 2013, pp. 1053–1056.

[9] Xu Yang, Hong Qiao, and Zhi-Yong Liu, "Partial correspondence based on subgraph matching," *Neurocomputing*, vol. 122, pp. 193 – 197, 2013.

[10] Arto Klami, "Variational Bayesian Matching," in *Proceedings of the 4th Asian Conference on Machine Learning, ACML 2012, Singapore, Singapore, November 4-6, 2012*, 2012, pp. 205–220.

[11] Arto Klami, Seppo Virtanen, and Samuel Kaski, "Bayesian canonical correlation analysis," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 965–1003, Apr. 2013.

[12] 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)*, vol. 30, no. 2, pp. 553–562, 2011.

[13] 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*, vol. 23, no. 3, pp. 994–1005, 2003.

[14] Derek Hill, "Neuroimaging to assess safety and efficacy of ad therapies," *Expert Opinion on Investigational Drugs*, vol. 19, no. 1, pp. 23–26, 2010, PMID: 19947893.

[15] 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*, Washington, DC, USA, 2011, CVPR '11, pp. 2401–2408, IEEE Computer Society.

[16] Tomoharu Iwata, Tsutomu Hirao, and Naonori Ueda, "Probabilistic latent variable models for unsupervised many-to-many object matching," *Information Processing and Management*, vol. 52, no. 4, pp. 682 – 697, 2016.

[17] Ali Rahimi and Ben Recht, "Random features for large-scale kernel machines," in *In Neural Infomration Processing Systems*, 2007.

[18] Kurt Cutajar, Edwin V. Bonilla, Pietro Michiardi, and Maurizio Filippone, "Random feature expansions for deep Gaussian processes," in *Proceedings of the 34th International Conference on Machine Learning*. 06–11 Aug 2017, vol. 70, pp. 884–893, PMLR.

[19] Nemanja Djuric, Mihajlo Grbovic, and Slobodan Vucetic, "Convex kernelized sorting," in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012, pp. 893–899.

[20] Michael M. Bronstein and Iasonas Kokkinos, "Scale-invariant kernel signatures for non-rigid shape recognition," in *In Proc. CVPR*, 2010.