Probabilistic Correspondence Analysis for Neuroimaging Problems

Hernan Felipe Garcia Arias

Universidad Tecnológica de Pereira

This dissertation is submitted for the degree of

Doctor of Engineering

Research in Automatics

Advisor
Dr. Mauricio A. Álvarez López.

Committee
Dr. Yuri Poveda
Dr. Augusto E. Salazar
Dr. Jorge A. Jaramillo

Date
October, 2019.

I would like to dedicate this thesis to my family ...

Acknowledgements

I would firstly like to thank my supervisor Dr. Mauricio A. Álvarez, for his support, inspiration, and guidance during these years. Mauricio was always kind and available to discuss every little detail of my research work. Additionally, he trusted me with exciting ideas and encouraged me to pursue my own ideas.

Thanks to all friends and colleagues at the Automatics research group. It was a privilege to get to know each one of them. Especially, I would like to thank Cristian D. Guarnizo for his support, feedback, and advice. I also thank Cristian A. Torres, Juan José Giraldo, Carlos D. Zuluaga, Hernan D. Vargas, Julian Gil, Jhouben J. Cuesta, Andres F. López, Juan F. López, Pablo A. Alvarado and Diego A. Agudelo, for their friendship and support during this process. Besides, I am very grateful to Dr. Neil D. Lawrence, Dr. Eleni Vasilaki, and Dr. Norberto Malpica for allowing me to have the chance to spend highly valuable time at their research labs.

I would like to acknowledge my source of funding given by Convocatoria 617 - 2013 from Administrative Department of Science, Technology and Innovation of Colombia (COLCIENCIAS).

Finally, I want to thank my family for their love and support. Especially, thanks to mom and dad for their love and patience. Also to the memory of my loving grandmother, "Maria Lilia Garcia de Ríos", who always inspired me to be a better person.

Abstract

Establishing meaningful shape correspondences between objects as in neuroimaging problems is crucial for improving matching processes. For instance, the correspondence problem consists of finding meaningful relations between any pair of brain structures as in the static registration problem, or analyzing temporal changes of a given neurodegenerative disease across time for dynamic brain shape analysis. Unfortunately, in many correspondence frameworks, similarity metrics are required to establish shape relations which are difficult to defined and limit the ability of modeling complex shape variations (i.e., large deformations in multiview shape analysis). Besides, these approaches only work over objects of the same size, and prealigned views, which gives poor performances in non-rigid matching problems. This thesis is concerned with solving these problems through developing unsupervised models for shape correspondence analysis.

In this thesis, we develop two methods for probabilistic correspondence analysis useful in neuroimaging problems. These approaches based on probabilistic nonlinear latent variable models can handle complex shape variations such as high curvedness regions, occlusions, and topological noise. Thus, these shape variations can be encoded through nonlinear mapping functions, which have correspondence assignments in the latent space. We show how to perform this task, creating a system that explores non-rigid object shapes by learning shape correspondences in an unsupervised manner.

The introductory chapters contain a review of the existing approximations to the correspondence problem. A special focus is given to a set of correspondence approaches in the context of shape analysis for image processing.

Secondly, we present 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. Finally, this thesis develops a multi-view warped mixture model having nonparametric cluster shapes, and a corresponding latent space in which each correspondence assignment has an interpretable parametric form. This probabilistic warped mixture representation allows for strong model learn-

ing, resulting in rich latent space representations of the shape variations. The developed methods are demonstrated in experiments with synthetic and real 3D shape datasets. The results show that the developed correspondence approaches perform accurately on shape matching processes, even when the shape structure has complex variations.

Contents

List of Figures

List of Tables

Notation

Mathematical notation

Generalities

V	number of views
N_v	Number of objects (observations) in the vth view
D_v	dimensionality of the v-th view
L_v	Dimensionality of the observed features in the vth view
K	Dimensionality of the latent feature vector
J	Number of correspondences (latent vectors) to which objects are assigned

Operators

$\mathbb{E}[\cdot]$	expected value
$\mathrm{tr}(\cdot)$	trace of a matrix

Functions

$k(\cdot,\cdot)$	covariance function for a Gaussian process of $\mathbf{x}_{vn}$
$f_d(t)$	d-th output or response function evaluated at t
$\boldsymbol{\phi}(\cdot)$	nonlinear mapping function

Vectors and matrices

$\mathbf{x}_{vn}$	Observation of the nth object in the vth view, $\mathbf{x}_{vn} \in \mathbb{R}^{D_v}$
$\boldsymbol{\phi}(\mathbf{x}_{vn})$	Observation of the nth feature object in the vth view, $\boldsymbol{\phi}(\mathbf{x}_{vn}) \in \mathbb{R}^{L_d}$
$\boldsymbol{\zeta}_j$	Latent feature vector for the jth correspondence, $\boldsymbol{\zeta}_j \in \mathbb{R}^K$
$\boldsymbol{B}_v$	Projection matrix for the vth view, $\boldsymbol{B}_v \in \mathbb{R}^{L_d \times K}$
θ_j	Mixture weight for the jth cluster, $\theta_j \geq 0$, $\sum_{j=1}^{\infty} \theta_j = 1$
$\mathbf{K}_v$	covariance matrix with entries $k(\mathbf{x}_{vn}, \mathbf{x}'_{vn})$
$\mathbf{f}_d$	$f_d(t)$ evaluated at $\mathbf{f}_d = [f_d(t_{d,1}), \ldots, f_d(t_{d,N_d})]^\top$
$\mathbf{f}$	vectors $\{\mathbf{f}_d\}_{d=1}^D$, stacked in one column vector
$\mathbf{I}_N$	identity matrix of size N

Abbreviations

LVM	Latent Variable Model
UCM	Unsupervised Clustering Matching
GP	Gaussian Process
GP-LVM	Gaussian Process Latent Variable Model
iGMM	Infinite Gaussian Mixture Model
DP	Dirichlet Process
EM	Expectation Maximization

Chapter 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) (Lin et al., 2014), or analyzing temporal changes of a given neurodegenerative disease across time (dynamic analysis of brain structures) (Durrleman et al., 2014). For instance, modeling brain volumes in neurodevelopmental outcomes is a challenge research topic in applications to asses primary adverse outcome in perinatal asphyxia.[1] Here, brain structures have significant morphology variation as part of brain development, which makes the analysis process a complex task (Weeke et al., 2018).

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 (Bronstein et al., 2011), largest common point-sets (Aiger et al., 2008; Cosa et al., 2013), and geodesic contours (Liang et al., 2015). Though, these approaches only work over objects of the same size, which gives a poor accuracy in non-rigid matching processes (Brunton et al., 2014).

Although similarity metrics could potentially capture shared information between objects, these metrics are not easy to define (Cortés and Serratosa, 2015) 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) (Cosa et al., 2013). Figure 1.1, shows an example of the shape correspondences from matched clusters, we can notice a large amount of geometric variations between the brain shapes which make the computation of such correspondences difficult.

[1]Perinatal asphyxia is a lack of blood flow or gas exchange to or from the fetus in the period immediately before, during, or after the birth process.

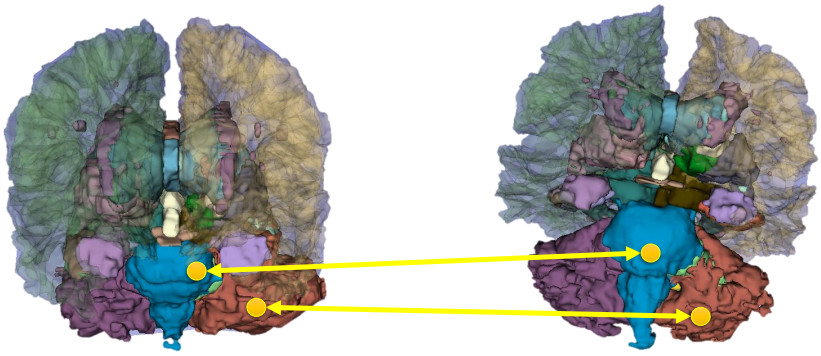


Figure 1.1: Shape correspondences from matched clusters. The goal is to establish meaningful relations between a pair of brain structures.

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 (Yang et al., 2013). Variational Bayesian matching (Klami, 2012) and Bayesian canonical correlation analysis (Klami et al., 2013) 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 (van Kaick et al., 2011) or volumes of brain structures for studying progression of Alzheimer's disease (Hill, 2010; Thompson et al., 2003). High variability of these patterns such as curvedness and size makes it necessary to compute the correspondences between objects in a groupwise manner (Sidorov et al., 2011).

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 (Iwata et al., 2016). However, these relations are impractical in applications where non-linear representations of shape objects are needed (i.e., non-rigid matching tasks).

1.1 Aims

1.1.1 General aims

To develop a probabilistic approach based on latent variable models for shape correspondences analysis, that allows temporal and structure learning of non-rigid shapes relevant in medical imaging problems.

1.1.2 Specific aims

1. To propose a probabilistic model that extends the latent variable model framework for unsupervised object matching, allowing to learn temporal and structure information of shape correspondences between non-rigid objects.

2. To design an inference method for learning the shape structure, temporal information and parameters of the probabilistic latent variable model.

3. To validate the performance of the proposed method over synthetic and medical imaging data, comparing its performance against commonly used shape correspondence methods in the state-of-the-art.

1.2 Outline of the Thesis

From the *modeling perspective*, we introduced nonlinear versions of the models proposed by Iwata et al. (2016) and Damianou et al. (2012). In particular, our contribution is to develop methods for shape correspondence analysis based on nonlinear unsupervised clustering of groupwise 3D shape descriptors. The clustering process was carried out by nonlinear probabilistic latent variable models, with the aim to model the multiview data. In other words, we extend the many-to-many object matching proposed by Iwata et al. (2016) exploring Hilbert space embeddings of the input data as in Cutajar et al. (2017); Rahimi and Recht (2007) and Gaussian Processes to obtain more representative latent descriptions of the input data (Damianou et al., 2012). Once the models are defined, we provide the mathematical framework and algorithms that allow computing the necessary posterior distributions of the probabilistic models.

From the *neuroimage perspective*, our contribution was based on the probabilistic formulation of the correspondence problem in which we explore the process of establishing

meaningful relations between 3D image data through unsupervised clustering of shape descriptors. That is, compute shape correspondences without any similarity measure.

In detail, the rest of the thesis is structured as follows:

- Chapter 2 constitutes a brief introduction to the correspondence problem. A special focus and a unifying view is given for a set of correspondence approaches in the context of shape analysis.

- Chapter 3 defines Nonlinear Unsupervised Clustering Matching (NL-UCM), a non linear latent variable model which incorporates observations from several views. 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 (Rahimi and Recht, 2007). In other words, we extend the many-to-many object matching proposed by Iwata et al. (2016) using Hilbert space embeddings of the input data (Rahimi and Recht, 2007).

- Chapter 4 presents a Multiview Bayesian clustering model in which the correspondences have nonparametric shapes, called the multiview infinite warped mixture model. The density manifolds learned by this model follow the contours of the data density, and have interpretable, parametric forms in the latent space. The marginal likelihood lets us infer the effective dimension and shape of each cluster separately, as well as the number of clusters.

  Chapter 5 summarizes the key contributions of the thesis and discusses ideas for future work.

1.3 Associated Publications and Software

The work presented in Chapter 3 is based on two published papers. First paper (García et al., 2015) states the motivation of the probabilistic analysis of the correspondence problem. Then, paper (García et al., 2018) extend the linear latent variable model proposed by Iwata et al. (2016).

The work presented in Chapter 4 is based on a paper by Garcia, H.F. and Álvarez M. A., which is in preparation at the time of writing this thesis. Finally, four publications

(García et al., 2014, 2016a,b, 2017) co-authored by myself are not discussed in this thesis since, although relevant to the general research area, they explore probabilistic models such as Gaussian processes for affective computing systems, and Bayesian optimization for fitting brain structures.

The **software** developed to accompany the methods described in chapters 3 and 4 is publicly available under a unified repository at `https://github.com/fepo68/probCorr`

Chapter 2

Correspondence Analysis

This chapter describes how to tackle the correspondence problem from different perspectives: similarity-based correspondence, rigid alignment, non-rigid alignment, time-varying registration, deep learning, and probabilistic correspondence approaches. We also describe relevant shape descriptors that allow us to establish the similarity between shapes indirectly.

The original contribution of this chapter is to offer a tutorial showing the implications of different correspondence problems, and some of the frameworks which can be obtained by combining them.

2.1 Overview of correspondence problem

Currently, a variety of problems can be classified in the field of shape correspondences analysis such as shape isomerty-invariant similarity (Bronstein et al., 2009), partial similarity, and non-rigid correspondence (Bronstein et al., 2008). However, since all these different problems address the same fundamental task, we can group them into a unified analysis by considering the following problem: given input shapes $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_N$ (e.g., brain structures), the main task is to establish meaningful relations $\mathcal{R}$ between their elements.[1] Besides, when two shapes are related to each other (i.e., $(s, z) \in \mathcal{R}$) for elements $s \in \mathcal{S}_i$ y $z \in \mathcal{S}_j$ with $i \neq j$), it is commonly said that these shapes are in correspondence or they match each other. The relation can be constrained in different manners, such as one-to-one (e.g., analyze two MRI volumes) (Cosa et al., 2013), one-to-many (e.g., match a given brain atlas to a MRI dataset) (Cabezas et al., 2011), or many-to-many correspondence (e.g., temporal behavior of a MRI dataset) (Zheng et al., 2010).

[1] Meaningful relation means which correspondence should be selected or is closer to our objective.

The correspondence problem can then be divided into the following categories:

- Similarity-based correspondence
- Rigid alignment
- Non-rigid alignment
- Time-varying registration
- Probabilistic correspondence analysis

In addition, for medical imaging analysis, we can compute correspondences for a given pair of brain images by the similarity of their elements or applying some alignment process between shapes and then derive a correspondence from the proximity of the aligned elements (e.g., brain atlas reconstruction (Cabezas et al., 2011)). It is worth noting that the alignment between the shapes is a side product of the computation which is useful and sometimes essential to the underlying application (Salama et al., 2016). We will address the issue of correspondence analysis by considering the case in which no alignment is required, and then we discuss the scenarios where the relation $\mathcal{R}$ is derived from aligning the shapes in a rigid or a non-rigid manner. Figure 2.1, shows an illustration of the shape correspondence scheme on which the problem can be view as a quality measure or a structure similarity.

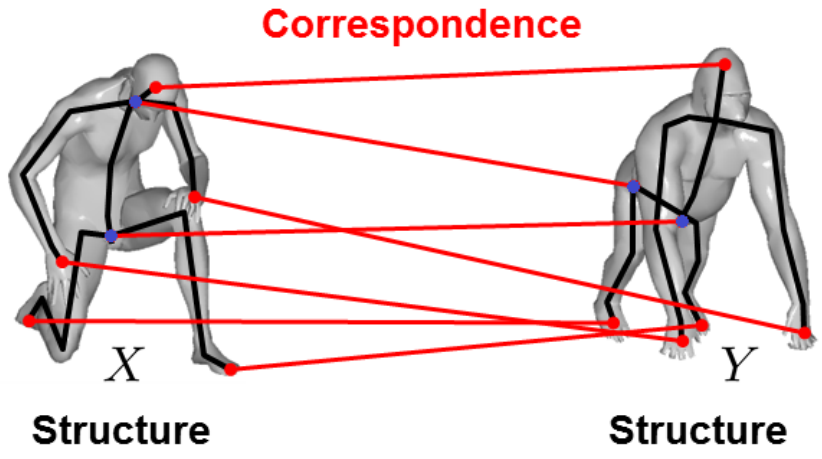


Figure 2.1: Scheme of Shape Correspondence Problem. "Correspondence quality = structure similarity" Bronstein et al. (2008). Blue and red circles relate bad and good matches, respectively.

2.1.1 Similarity-based correspondence

The most common way of computing correspondences is to analyze pairs of shape elements or feature points to estimate the similarity between these features (commonly called the feature matching approach in neuroimaging). These elements are commonly characterized by shape descriptors (e.g., texture properties of brain tissues) (Kanavati et al., 2017). We can obtain the correspondence by optimizing an objective function that allows us to select assignments between pairs of elements. This objective function is composed of two terms, one term seeks to maximize the similarity between the descriptors of corresponding elements, while the other one seeks to minimize the distortion that would be introduced in the shapes if they were deformed to align their corresponding elements (e.g., brain atlas matching) (Cabezas et al., 2011; Castellani et al., 2008).

In the medical imaging field, this is the most common method for neuroimaging analysis (Hill, 2010). Here, texture descriptors are computed from the MRI volume to establish similarity metrics between volumes (e.g., Euclidean, cross-correlation and super-voxels) (Sjoberg and Ahnesjo, 2013). The approach can be global, where the whole MRI volume is parametrized by statistical texture descriptors (Salama et al., 2016; Sjoberg and Ahnesjo, 2013), or local where the texture descriptors are computed in local regions of the MRI volume (Cosa et al., 2013; Hill, 2010; Kanavati et al., 2017).

In principle, we can obtain a solution that is geometrically meaningful by satisfying this objectives function (similarity based metrics). Besides, this solution is typically obtained with a standard optimization method such as quadratic programming (Salama et al., 2016). That is why we can apply feature matching in any context where can be possible to compute a set of descriptors for these elements (MRI volumes) (Castellani et al., 2008; Cortés and Serratosa, 2015).

2.1.2 Rigid alignment

Another way to view the correspondence problem in neuroimaging problems is to present it as a searching process for a geometric transformation that aligns the MRI volumes (input shapes). The most common method for this purpose is the rigid alignment of geometry scans, used for shape acquisition and medical image registration (Aiger et al., 2008; Panda et al., 2017). One important feature of the rigid alignment problem is that the objects do not change from one scanning pass to another (MRI volumes remains their structure after the interaction) and the shapes are assumed to be aligned by a Euclidean transformation involving a rotation and translation (Oliveira and Tavares, 2014; Tam

et al., 2013).

Furthermore, most of the common applications derived from the rigid alignment are focused on capturing a given transformation that can match one MRI volume into another (Panda et al., 2017), initialize a given atlas transformation (e.g., adapt a given brain atlas into a new MRI volume) (Cosa et al., 2013; Sjoberg and Ahnesjo, 2013) and analyze rigid properties of the brain structures such as size, translation and rotation (Ngo et al., 2017; Oliveira and Tavares, 2014).

Shapes alignment is just one example of many medical applications that rely on the assumption of rigidity in the MRI datasets (Hill, 2010; Podolak et al., 2006). For two shapes given as 3D point sets $\mathcal{S}$ and $\mathcal{Z}$ (e.g., brain structure analysis), the problem of rigid alignment can be posed as: find the rigid transformation that, when applied to $\mathcal{S}$, maximizes the number of points in $\mathcal{S}$ that align to points in $\mathcal{Z}$. This goal is usually dependent on a threshold ε that indicates when two points are close enough and can be considered as matching to each other (Cortés and Serratosa, 2015).

Rigid alignment is a challenging problem because the data itself poses many difficulties, which may include noise, outliers, and limited amounts of overlap (Oliveira and Tavares, 2014). Outliers are unwanted points far from the surface, which can seriously affect the alignment process if not discarded (Oliveira and Tavares, 2014). Also, noise information may take the form of unwanted points close to a 3D shape (Salama et al., 2016). This method might be the less robust to compute a correspondence between brain shapes, but it is commonly used to initialize some other methods (e.g., non-rigid methods) due to fast convergence (Aiger et al., 2008).

2.1.3 Non-rigid alignment

Now let us consider the case in which we need to analyze the shape deformation process between a set of brain structures. Unlike the rigid case, where a few correspondences are sufficient to define one candidate rigid transformation, we need to compute both deformation and alignment in the non-rigid case, without prior assumptions (Sotiras et al., 2013). Thus, this method often requires a lot more reliable correspondences to define the matching process (Tam et al., 2013). Some examples of non-rigid alignment processes include brain shape correspondence for objects with different geometries with parts that are semantically related (Cabezas et al., 2011; Chang and Zwicker, 2009; Zhang et al., 2008), and the correspondence of anatomical shapes (i.e., brain structures, organs) (Lin et al., 2014), which can be deformed elastically by introducing stretching to localize regions of the shape (Weingarten et al., 2015).

Due to the local stretching and bending presented between brain structures (e.g., local variations related to shapes in Parkinson disease) in the non-rigid case, we can see the need to analyze how these shapes can be brought into correspondence (Weingarten et al., 2015). This can be achieved by taking into consideration non-rigid transformations over the analyzed MRI volumes (i.e, B-splines and thin-plates) (Wang and Pan, 2014). Moreover, these transformations can be applied separately to local portions of the shapes. Here, this transformation applied to a given brain structure can be represented as a set of displacement vectors with the aim to match accurately the target shape (Sidorov et al., 2011; Xiao et al., 2010).

However, although heuristic solution methods are available in the state-of-the-art, such as diffeomorphic registration (Tang et al., 2013), elastic methods (Soon and Qiu, 2015), and fluid registration methods (Shi et al., 2013), the quality of the results will typically depend on the complexity of the medical problem instance and the level of approximation introduced by these methods (Cai et al., 2015; Wang and Pan, 2014).

2.1.4 Time-varying registration

Space-time correspondences analysis has an increasing attention in the last few years (Weingarten et al., 2015). Since neurodegenerative diseases are difficult to monitoring (e.g., analyze brain evolution over time in Alzheimer's disease), the reconstruction of 3D brain structures acquired over time while moving and deforming (e.g., as the disease evolves on time) is one particular topic in which both rigid and non-rigid alignment have been studied (Cosa et al., 2013; Lin et al., 2014; Weingarten et al., 2015; Zheng et al., 2010). In this framework, a fixed number of scans is acquired per time step (i.e MRI volumes, 3D point cloud of brain shapes), and these scans have to be aligned to perform the reconstruction of both the shape object and the motion sequence (Li et al., 2009; Sharf et al., 2008). However, recovering the temporal transformation is a difficult task, since the dimensionality of MRI data is higher than the data instances that neuroimaging problems usually present (e.g., few MRI volumes to be analyzed for one patient over time). As a consequence, most of the temporal approaches to analyze high-dimensional datasets are focused on capture shape primitives to describe the entire dynamic dataset (i.e., shape skeletons in temporal registration frameworks) (Hill, 2010; Zheng et al., 2010).

In addition, it is noticed that for the classic registration problem, all shapes can be aligned to compose a single and coherent object, while the time-varying problem introduces additional difficulties. The main reason for this assumption is that the brain

shapes may change along time (deformed significantly from one frame to the other in a given neurodegenerative disease) (Hyun et al., 2016; Zheng et al., 2010). However, a large amount of missing data (due to the evolution of the neurodegenerative diseases) that can be present in each frame (Pekelny et al., 2008; Stylianou et al., 2016), and datasets that were captured over sparse time (e.g., MRI data for monitoring brain volumetry) (Chang and Zwicker, 2009; Hyun et al., 2016; Wolk et al., 2017) remain a challenging topic in neuroimaging problems. Furthermore, adding temporal constraints to the correspondence model can help in reducing the size of the search space, and can bring more information about the dynamics of a given brain structure over time (Stylianou et al., 2016; Sússmuth et al., 2008; Weingarten et al., 2015).

2.1.5 Deep learning approaches for shape correspondence

Deep learning has become in a popular learning framework were complex datasets are modeled through combinations of simpler models in a hierarchical, or multilayer manner (i.e., artificial neural networks). Besides, model-based approaches based on deep learning concepts have achieved accurate performances in several fields such as image processing, speech recognition, natural language processing, and computer vision (Tygert et al., 2016).

In particular, convolutional neural network (CNN) architectures currently present state-of-the-art performance on a variety of computer vision tasks when dealing with 1D, 2D, or 3D Euclidean-structured data such as acoustic signals, images, or videos (Lundervold and Lundervold, 2019). Hence, we can consider images as functions on the Euclidean space (plane), or sampled on a grid (e.g., MRI data used for medical image analysis) (Cosa et al., 2013; Wu et al., 2015).

However non-Euclidean geometric data arises in numerous applications that range from modeling users attributes in social networks (i.e., signals of the vertices), to graph models in neuroscience (Tan et al., 2019; Wee et al., 2019). Thus, anatomical and functional structures of the brain can be modeled as Riemannian manifolds (i.e., mesh surfaces) endowed with properties such as texture, shape structure, and curvedness (Su et al., 2015).

Besides in computer vision applications, finding similarity and correspondence between shapes are examples of manifold learning problems (Litany et al., 2016). Convolutional neural networks on non-Euclidean domains (surfaces) were first presented by Masci et al. (2015) with the introduction of the geodesic CNN model. Here, the clas-

sical convolution operation is replaced by a local geodesic[2] system of polar coordinates to extract patches (i.e., a small part of a given brain structure). Then, from these patches, surface descriptors can be learned to compute the correspondences. Moreover, Boscaini et al. (2016) used anisotropic heat kernels (Bronstein and Kokkinos, 2010) as an alternative way of extracting intrinsic patches to represent the shape descriptors.

Therefore, these methods are instances of a broader recent trend of geometric deep learning attempting to generalize successful deep learning paradigms to data with a non-Euclidean underlying structure such as manifolds or graphs (Bronstein et al., 2017). Finally, since this is a newly emerging field on computer vision and our research path is focused on probabilistic models, we are concern about the uncertainty of the shape geometry that will be captured in a latent space, rather than exploring surface descriptors from geometric learning (i.e., compute correspondences through similarity metrics).

2.1.6 Probabilistic correspondence analysis

Although similarity metrics could potentially capture shared information between objects, these metrics are not easy to define (Cortés and Serratosa, 2015) 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 (e.g., Alzheimer and Parkinson) (Cosa et al., 2013).

Instead of defining similarity metrics, an alternative approach consists of using probabilistic methods 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 (Yang et al., 2013). Variational Bayesian matching (Klami, 2012) and Bayesian canonical correlation analysis (Klami et al., 2013) 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 (e.g., point-to-point matching) and linear analysis over the shape descriptors (e.g., appearance descriptors), which makes them unsuitable for modeling shared information between non-rigid objects such as tissue shapes in MRI data (van Kaick et al., 2011), and volumes of brain structures for studying the progression of Alzheimer's disease (Hill, 2010; Thompson et al., 2003). Thus, the high variability of these patterns such as curvedness and size makes it necessary to compute the correspondences between objects in a groupwise

[2] A geodesic is a curve representing in some sense the shortest path between two points in a surface (i.e., Riemannian manifold)

manner (Sidorov et al., 2011).

Recently, approaches formulated as Gaussian Processes Latent Variable Models (GP-LVMs) have been successful in applications derived from a low-dimensional representation that arises the shape structure as a manifold of feature descriptors (i.e., variables that are not directly observed such as latent variables)(Ek et al., 2008b; Shon et al., 2005). Nevertheless, these models consider a single latent variable to represent the multiview data (i.e., data that comes from different sources), assuming that the modalities can be fully aligned (e.g., multimodal image registration between computed tomography (CT) and MRI).

To improve the latent data representation, the idea of a segmented latent space was explored in works such as Ek et al. (2008a), Salzmann et al. (2010) and Damianou et al. (2012), where each view is handled as independent data with an additional private space (to represent the variance which cannot be aligned). 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 (Damianou et al., 2012; Iwata et al., 2016; Tomoharu Iwata, 2013).

Finally, table 2.1 shows a summary of the relevant approaches for shape correspondence analysis. Besides, the table describes the criteria of the different correspondence frameworks.

2.2 Magnetic resonance imaging

In this project, we will use Magnetic resonance imaging (MRI) as the input data where all the analysis will be derived. The motivation is that MRI is the preferred medical technique used in radiology to capture brain images of healthy and pathological tissues. As it has better resolution than CT, MRI offers better visualization of the posterior fossa. The contrast provided by the gray and white matter of the MRI data makes it the best choice for many conditions of the central nervous system, including demyelinating diseases, dementia, cerebrovascular disease, infectious diseases and epilepsy. Since many images are taken milliseconds apart, it shows how the brain responds to different stimuli; researchers can then study both the functional and structural brain abnormalities in psychological disorders (e.g., brain volumetry analysis in Alzheimer's disease) (Salama et al., 2016). MRI is also used in MRI-guided stereotactic surgery and radiosurgery for treatment of intracranial tumors, arteriovenous malformations and other surgically

Table 2.1: Summary of the different frameworks of the correspondence problem

Input		**Geometry representation**		Points, skeletons, feature points, surfaces
		Dimensionality of the data		2D, 2D+Time, 3D, 3D+Time
Output	Correspondence representation		Correspondence and transformation	Translation, rigid, similarity, affine, non-rigid
			Correspondence only	Bijective, injective, crisp, probabilistic
			Full vs. Partial	
			Dense vs. sparse	
Objetive Function	**Similarity-based correspondence**		(Cosa et al., 2013) (Kanavati et al., 2017)	Similarity only, similarity and distortion
	Rigid alignment		(Sjoberg and Ahnesjo, 2013) (Cortés and Serratosa, 2015)	Largest common pointset, geometric distance
	Non-Rigid alignment		(Bronstein et al., 2009) (Bronstein et al., 2011)	Geometric distance and regularization
Approach	**Solution paradigm**	Transformation search	(Soon and Qiu, 2015)	Alignment, point cluster, nonrigid alignment
		Correspondence search	(Cortés and Serratosa, 2015)	Continuous optimization, combinational search
		Hybrid search	(Liu et al., 2015)	ICP, prealignment+ICP, embedding+ICP
		Geometric Deep Learning	(Bronstein et al., 2017)	Volumetric CNNs, Intrinsic
	Global vs. local search		(Boscaini et al., 2016)	CNNs, Geodesic CNNs
	Pairwise vs. groupwise	Probabilistic search	(Iwata et al., 2016) (Damianou et al., 2012)	Linear latent variable Models, Gaussian Processes

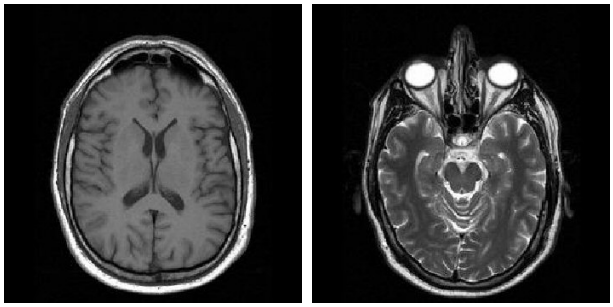
treatable conditions using a device known as the N-localizer (Cabezas et al., 2011; Cosa et al., 2013; Mercier et al., 2012).

MRI of the nervous system uses magnetic fields and radio waves to produce high quality $2D/3D$ images of nervous system structures without the use of ionizing radiation (X-rays) or radioactive tracers (Salama et al., 2016). Image contrast may be weighted to demonstrate different anatomical structures or pathologies. Each tissue returns to its equilibrium state after excitation by the independent processes of T1 (spin-lattice) and T2 (spin-spin) relaxation.

Figure 2.2 shows an example of MRI $T1$ and $T2$ images related to the brain.

2.3 Shape descriptors

Most of the matching methods in the state-of-art use surfaces to represent the shape of a given object. These methods are based on isometric surfaces (Zhang et al., 2008), deformation (Huang et al., 2008), template matching (Li et al., 2009) and articulated shapes (Chang and Zwicker, 2009). Motion reconstruction or dynamic volume registra-



(a) MRI $T1$ sample (b) MRI $T2$ sample

Figure 2.2: Samples of MRI $T1$ and $T2$ images. The above figures show that for different schemes of acquisitions, MRI images exhibit changes in their texture descriptors related to each tissue (e.g., brain structures such as gray and white matter).

tion methods use Time-varying surfaces to analyze the shape variability over time (Sharf et al., 2008). Finally, skeleton shapes is the more general framework of shape representations such as the medial axis and curve skeletons (Gall et al., 2009; Kin-Chung Au et al., 2010).

Extracting representative points from a given shape can give us another set of information about the shape (computing descriptors for these points). These descriptors will be scalar or vectors of scalars that capture some properties about the shape (Lin et al., 2014). From these descriptors, we can establish a given similarity measure between the shapes to be analyzed (Kalogerakis et al., 2010). Furthermore, if two or more descriptors are similar, we can say that their corresponding points should also be similar (Bronstein et al., 2011). Since surfaces can give relevant information about the shape structure of a given brain structure, in this work we will use shape descriptors as input features for the probabilistic correspondence method.

2.3.1 Scale-invariant Heat Kernel Signature (SI-HKS)

Using a shape descriptor that maintains invariance under a wide class of transformations is convenient when it comes to shape analysis. The SI-HKS is a scale-invariant version of the heat kernel descriptor where an intrinsic local shape descriptor based on diffusion scale-space analysis is performed in order to describe a given shape X(Bronstein et al., 2011). Here, brain structures are modeled as Riemannian manifolds in order to compute the shape descriptors from the heat conduction properties (Bronstein and Kokkinos,

2010). Heat propagation on non-Euclidean domains is governed by the heat diffusion equation,

$$\left(\Delta_X + \frac{\partial}{\partial t}\right) u\left(x,t\right) = 0, \tag{2.1}$$

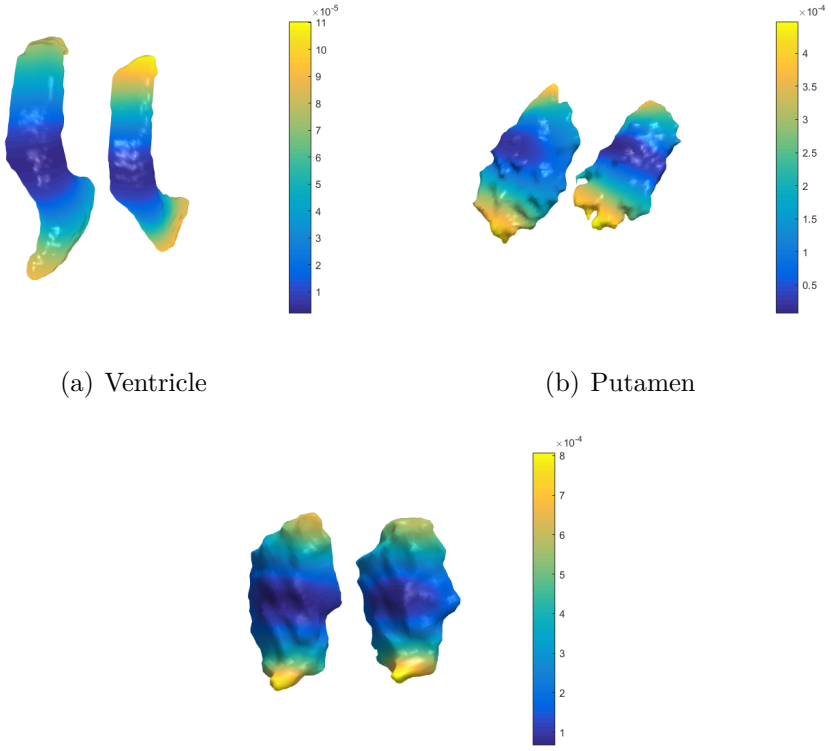
where Δ_X is the Laplace Beltrami operator and $u\left(x,t\right)$ is the heat distribution at a point x at time t. The solution of the equation (2.1) with point distribution $u_0(x) = \delta\left(x - x'\right)$ as initial condition, describes the amount of heat on a given surface at point x in time t. This solution is called the *heat kernel* and is denoted by $K_{X,t}\left(x,x'\right)$. Thus, the heat kernel signature can be described as

$$h\left(x,t\right) = K_{X,t}\left(x,x\right) = \sum_{i=0}^{\infty} e^{-\lambda_i t}\phi_i(x)^2, \tag{2.2}$$

where $\lambda_0, \lambda_1, \ldots \geq 0$ are eigenvalues, and $\phi_0, \phi_1, \ldots$ are the corresponding eigenfunctions of the Laplace-Beltrami operator, satisfying $\Delta_X \cdot \phi_i = \lambda_i \phi_i$. One disadvantage of the heat kernel signatures is their sensitivity to scale. Given a shape X and its scaled version $X' = \beta X$, the new eigenvalues and eigenfunctions will satisfy $\lambda' = \beta^2 \lambda$¸ and $\phi' = \beta\phi$.

In order to achieve scale invariance, Bronstein proposes to remove the dependence of h from the scale factor β (Bronstein and Kokkinos, 2010). First, the HKS is sampled logarithmically in time ($t = \alpha^\tau$) at each shape point x to obtain a discrete function of h as h_τ. Finally, they use the discrete-time Fourier transform of h_τ to shift in time into a complex phase $H'\left(w\right) = H\left(w\right) e^{2\pi w s}$.

Figure 2.3 shows an example of computing SI-HKS descriptors over three kind of brain structures (ventricle, putamen and thalamus). The figure shows that spatial regions related to the same descriptor (i.e., same color), share common information which can be exploited to model the correspondence between brain structures.



(a) Ventricle

(b) Putamen

(c) Thalamus

Figure 2.3: Samples of SI-HKS descriptors for three brain structures: Ventricle, Putamen, and Thalamus, respectively. The figure shows the early and advanced stage of the Alzheimer disease for each brain structure (left and right shape per brain structure). The color bar indicates high and low salience regions, respectively.

Chapter 3

Nonlinear Unsupervised Clustering Matching

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 views instead of analyzing full correspondences by establishing linear relations between objects as Iwata et al. present in (Iwata et al., 2016). However, these relations are impractical in applications where nonlinear representations of shape objects are needed (i.e., non-rigid matching tasks). In this chapter, we introduce a nonlinear version of the model proposed by Iwata et al. (Iwata et al., 2016). 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 (Rahimi and Recht, 2007). In other words, we extend the many-to-many object matching proposed by Iwata et al. (2016) using Hilbert space embeddings of the input data (Rahimi and Recht, 2007). Once the model is defined, we provide an stochastic EM algorithm for computing the necessary posterior distributions of the probabilistic model.

3.1 Unsupervised Cluster Matching via Linear Latent Variable Models

Unsupervised object matching is a task that involves finding relations between observations in different views without any correspondence information. Methods for un-

supervised object matching have been proposed in the state-of-the-art to model shape structures where there are no full correspondences between objects, or it is challenging to compute some similarity measures between them. Yang et al. (2013). Some works such as least squares object matching Yamada and Sugiyama (2011), variational Bayesian matching Klami (2012) and kernelized sorting Quadrianto et al. (2009), use probabilistic methods to modeling features between objects to establish matching processes.

Let us first introduce the classical linear latent variable models for unsupervised object matching. Thus, we describe the model proposed by Iwata et al. (2016) in which a given task for multiview object matching is performed. Besides, this model assumes that there is an infinite number of latent vectors that are shared by all views, and each object is generated from one of the latent vectors and a view-specific projection matrix. Figure 3.1 shows the relationship between latent vectors and objects in two different views, where arrows that indicate the corresponding latent vectors for each observation are hidden.

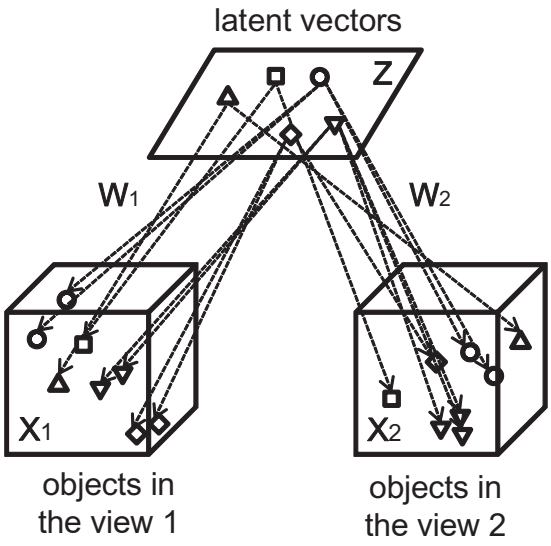


Figure 3.1: Relationship between latent vectors and input observations in two views (Iwata et al., 2016).

3.1.1 The linear model

The linear model assumes that we are given objects (i.e., observations in the input space) for V views $\{\mathbf{X}_v\}_{v=1}^{V}$ where $\mathbf{X}_v = \{\mathbf{x}_{vn}\}_{n=1}^{N_v}$ is a set of observations in the vth view, and $\mathbf{x}_{vn} \in \mathbb{R}^{D_v}$ is the input vector of the nth observation in the vth view. Besides, since we

are unaware of any correspondences between objects in different views, the number of observations N_v and the dimensionality D_v for each object (e.g., brain structure) can be different from those of other structures. Therefore, our task is to match clusters of objects (i.e., groupwise correspondences) across multiple views in an unsupervised manner (Iwata et al., 2016).

As we describe above, this model assumes that there is a potentially infinite number of clusters (i.e., groups of correspondences), and each cluster j has a latent vector $\mathbf{z}_j \in \mathbb{R}^K$, in a K-dimensional latent space. Specifically, the proposed model is an infinite Gaussian mixture model (iGMM), where the probability of object in the input space $\mathbf{x}_{vn}$ is given by

$$p\left(\boldsymbol{x}_{vn}|\boldsymbol{Z},\boldsymbol{W},\boldsymbol{\theta}\right)=\sum_{j=1}^{\infty}\theta_j\mathcal{N}\left(\boldsymbol{x}_{vn}|\boldsymbol{W}_v\boldsymbol{z}_j,\alpha^{-1}\boldsymbol{I}\right) \tag{3.1}$$

where $\mathbf{W}=\{\mathbf{W}_v\}_{v=1}^{V}$ 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. Besides,

Since each object $\mathbf{x}_{vn}$ in the v-th view is generated depending on a view-specific projection matrix $\mathbf{W}_v \in \mathbb{R}^{M_v\times K}$ and a latent vector $\mathbf{z}_{s_{vn}}$ that is selected from a set of latent vectors $\mathbf{Z}=\{\mathbf{z}_j\}_{j=1}^{\infty}$. Here, $s_{dn}=\{1,\ldots,\infty\}$ is the latent cluster assignment of object $\mathbf{x}_{vn}$. Observations that use the same latent vector, or that have the same cluster assignment s_{vn}, are considered to match.

3.2 Probabilistic Nonlinear Latent Variable Model for Groupwise Correspondence

Regarding the multiview dataset in the observed space $\mathbf{X}=\{\mathbf{X}_v\}_{v=1}^{V}$, where $\mathbf{X}_v=\{\mathbf{x}_{vn}\}_{n=1}^{N_v}$ is a set of objects (i.e., shape descriptros) in the vth view, and $\mathbf{x}_{vn}\in\mathcal{R}^{D_v}$ is the input vector of the nth object in the vth view, we aim to map the input observations to the feature space. Thus, by introducing a function $k:\mathcal{X}\times\mathcal{X}\mapsto\mathbb{R}$ called the *kernel*, that performs a given nonlinear mapping over the objects, $\phi:\mathcal{X}\to\mathcal{H}$ such that $\forall x,x'\in\mathcal{X}$,

$$k\left(x,x'\right):=\left\langle\phi\left(x\right),\phi\left(x'\right)\right\rangle_{\mathcal{H}}, \tag{3.2}$$

we can cluster groups of correspondences by using nonlinear mapping functions that represents the shape descriptors in the Hilbert space. 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 (Schölkopf et al., 1999).

As we are unaware of any correspondence between feature sets $\boldsymbol{\Phi} = \{\boldsymbol{\Phi}_v\}_{v=1}^{V}$ in different views, we set different number of feature vectors $\boldsymbol{\Phi}_v = \{\boldsymbol{\phi}(\mathbf{x}_{vn})\}_{n=1}^{N_v}$, and different dimensionalities L_v such that $\boldsymbol{\phi}(\mathbf{x}_{vn}) : \mathcal{R}^{D_v} \rightarrow \mathcal{R}^{L_v}$. 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 $\mathbf{s}_{vn}$, or are related by the same latent feature vector, establish a meaningful correspondence.

Each feature vector in $\boldsymbol{\phi}(\mathbf{x}_{vn}) \in \mathcal{H}$ in the vth view is generated depending on the view-specific projection matrix $\boldsymbol{B}_v \in \mathcal{R}^{L_v \times K}$ and the latent feature vector $\boldsymbol{\zeta}_{s_{vn}}$ 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 $\boldsymbol{\phi}(\mathbf{x}_{vn})$. Then, by using a latent space representation of an infinite Gaussian mixture model, we define the probability of a feature vector $\boldsymbol{\phi}(\mathbf{x}_{vn})$ as

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

where $\boldsymbol{\mathcal{W}} = \{\boldsymbol{B}_v\}_{v=1}^{V}$ 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 a Hilbert space for each feature vector (view-specific), we can handle multiple feature sets with nonlinear properties and different dimensionalities (i.e., size of the brain structures). Figure 3.2 shows the scheme of the proposed model, in which we describe the relationship between feature vectors and latent feature vectors in a Hilbert space.

3.2.1 Stick-breaking representation

A stick-breaking process is an approach to generate a random vector with a Dirichlet distribution $\mathrm{Dir}(\cdot)$, that involves iteratively breaking a stick of length one into k pieces in such a way that the lengths of the k pieces θ_k follow a $\mathrm{Dir}(\gamma)$ distribution (Sethuraman, 1994). First, choose a beta random variable $\beta_1 \sim \mathrm{Beta}(1, \gamma)$ and break of β_1 of the stick. For each remaining segment, choose another beta distributed random variable,

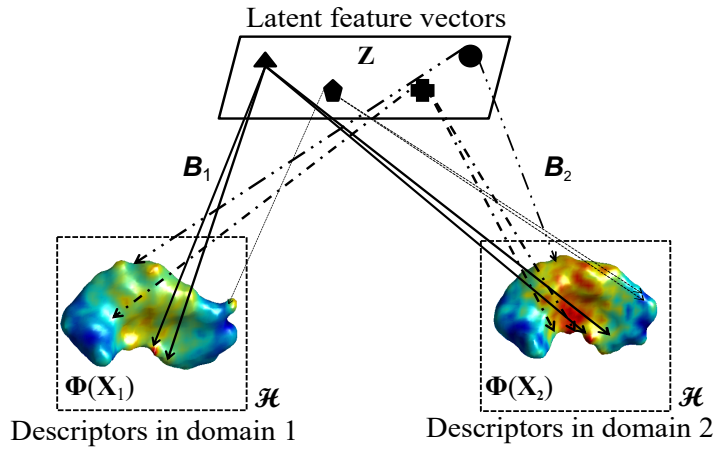


Figure 3.2: Scheme for the unsupervised nonlinear clustering method for groupwise correspondence analysis. The figure shows an example of establishing correspondences in a Hilbert space for two brain structures (left putamen).

and break off that proportion of the remainder of the stick. This gives us an infinite collection of weights θ_k,

$$\beta_k \sim \text{Beta}(1, \gamma), \tag{3.3}$$

$$\theta_k = \beta_k \prod_{j=1}^{k-1} (1 - \beta_j) \quad k = 1, 2, 3, \ldots \tag{3.4}$$

This representation of the Dirichlet process, and its corresponding use in a Dirichlet process mixture allows us to compute a variety of functions of posterior DPs (Gelfand et al., 2005). As in Iwata et al. (2016), we use a stick-breaking process to set the mixture weights $\boldsymbol{\theta}$ for a Dirichlet process with concentration parameter γ.

3.2.2 The model

The proposed model generates feature sets in multiple views $\boldsymbol{\Phi}$ according to the following process,

1. Draw mixture weights $\boldsymbol{\theta} \sim \text{Stick}(\gamma)$
2. Draw a precision parameter $\alpha \sim \mathcal{G}(a, b)$
3. For each cluster: $j = 1, \ldots, \infty$

(a) Draw a latent vector $\boldsymbol{\zeta}_j \sim \mathcal{N}(\mathbf{0}, (\alpha r)^{-1}\mathbf{I})$

4. For each view: $v = 1, \ldots, V$

   (a) For each object: $n = 1, \ldots, N_v$

      i. Draw a cluster assignment $s_{vn} \sim \text{Categorical}(\grave{})$

      ii. Draw a feature vector $\boldsymbol{\phi}(\mathbf{x}_{vn}) \sim \mathcal{N}(\boldsymbol{B}_v \boldsymbol{\zeta}_{s_{vn}}, \alpha^{-1}\mathbf{I})$

Here, r is a precision parameter of the latent vectors, Stick(γ) is the stick-braking process that generates the mixture wights for the DP with concentration parameter γ (Sethuraman, 1994), and $\mathcal{G}(a, b)$ is a Gamma distribution with parameters a and b.

As for the priors of the latent vectors $\mathbf{Z}$ and precision parameter α, we use a Gauss-Gamma distribution, because it is a conjugate-prior for a Gaussian

$$p(\boldsymbol{\phi}(\mathbf{x}_{vn}) | \boldsymbol{\zeta}_j, \alpha, \boldsymbol{\mathcal{W}}) = \mathcal{N}(\boldsymbol{\phi}(\mathbf{x}_{vn}) | \boldsymbol{B}_v \boldsymbol{\zeta}_j, \alpha^{-1}\mathbf{I}), \tag{3.5}$$

and it enables us to analytically integrate out the latent vectors as show in Appendix B.1. The joint probability of the feature vectors $\boldsymbol{\Phi}$, and the cluster assignments $\mathbf{S} = \left\{\{\mathbf{s}_{vn}\}_{n=1}^{N_v}\right\}_{v=1}^{V}$ is given by

$$p(\boldsymbol{\Phi}, \mathbf{S} | \boldsymbol{\mathcal{W}}, a, b, r, \gamma) = p(\mathbf{S}|\gamma)\, p(\boldsymbol{\Phi}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r), \tag{3.6}$$

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 $N = \sum_{v=1}^{V} N_v$ 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 $N_{\cdot j} > 0$. Figure 3.3, shows a graphical model representation of the proposed model, where shaded and unshaded nodes indicate observed and latent variables, respectively.

For our non-linear model, we give the derivation of the likelihood in (3.3), in which latent feature vectors $\mathbf{Z}$ and precision parameter α are analytically integrated out. The

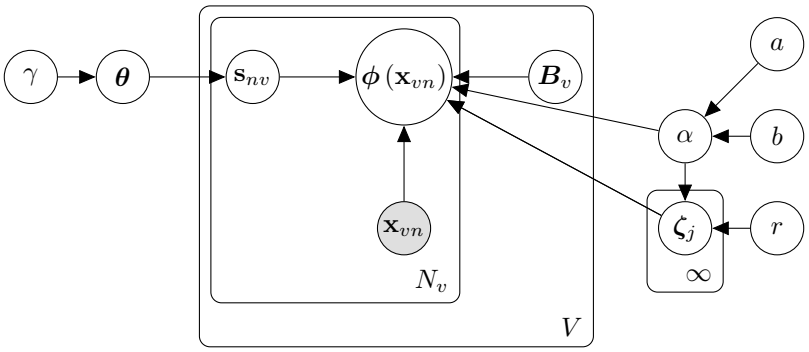


Figure 3.3: A graphical model representation of the Non-Linear Latent Variable Model for Groupwise Correspondence, where the shaded and unshaded nodes indicate observed and latent variables, respectively, and plates indicate repetition.

resulting expression is defined as

$$p\left(\boldsymbol{\Phi}|\mathbf{S},\boldsymbol{\mathcal{W}},a,b,r\right)=(2\pi)^{-\frac{\sum_v L_v N_v}{2}}\, r^{\frac{KJ}{2}}\,\frac{b^a}{b'^{a'}}\,\frac{\Gamma\left(a'\right)}{\Gamma\left(a\right)}\prod_{j=1}^{J}|\boldsymbol{\Lambda}_j|^{1/2}. \tag{3.7}$$

Here, $a'=a+\frac{\sum_v L_v N_v}{2}$,

$$b'=b+\frac{1}{2}\sum_{v=1}^{V}\sum_{n=1}^{N_v}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^{\top}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)-\frac{1}{2}\sum_{j=1}^{J}\boldsymbol{\mu}_j^{\top}\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j, \tag{3.8}$$

and

$$\begin{aligned}\boldsymbol{\mu}_j&=\boldsymbol{\Lambda}_j\sum_{v=1}^{V}\boldsymbol{B}_v^{\top}\sum_{n:\mathbf{s}_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right),\\ |\boldsymbol{\Lambda}_j^{-1}|&=\sum_{v=1}^{V}N_{vj}\boldsymbol{B}_v^{\top}\boldsymbol{B}_v+r\mathbf{I},\end{aligned} \tag{3.9}$$

where r is a parameter for controlling the precision of the latent feature vectors $\mathbf{Z}$, and N_{vj} is the number of feature vectors assigned to cluster j in the v view (see Appendix B.1 for further details). Finally, the posterior for the precision parameter α is given by

$$p\left(\alpha|\boldsymbol{\Phi},\mathbf{S},\boldsymbol{\mathcal{W}},a,b\right)=\mathcal{G}\left(\alpha|a',b'\right), \tag{3.10}$$

and the posterior for the latent vector $\boldsymbol{\zeta}_j$ is given by

$$p\left(\boldsymbol{\zeta}_j|\alpha,\boldsymbol{\Phi},\mathbf{S},\boldsymbol{\mathcal{W}},r\right)=\mathcal{N}\left(\boldsymbol{\zeta}_j|\boldsymbol{\mu}_j,\alpha^{-1}\boldsymbol{\Lambda}_j\right). \tag{3.11}$$

See Appendix B.2 for the derivation.

3.3 Inference

Stochastic expectation maximization (EM) algorithm comes as an attractive alternative to EM. The main idea of Stochastic EM is to impute for the missing data with plausible values given the observed data and a current estimate of the model parameters (McLachlan and Krishnan, 2008). Moreover, Stochastic EM is particularly useful in problems where EM is intractable. For instance, in problems where the latent variables are integrate out to obtain the marginal likelihoods.

To marginalize out the latent feature vectors $\mathbf{Z}$, and the precision parameter α, we use an stochastic EM algorithm (Iwata et al., 2016). Hence, we alternatively iterate collapsed Gibbs sampling for the cluster assignments $\mathbf{S}$, and maximum joint likelihood estimation of the projection matrices $\boldsymbol{\mathcal{W}}$. By collapsing the latent variables (i.e., marginalizing out), the time-consuming step of drawing these variables is skipped, and the sample auto-correlations are usually reduced (Liu, 1994).

In the E-step, a new value for $\mathbf{s}_{dn}$ is sampled from

$$p\left(s_{vn}=j|\boldsymbol{\Phi},\mathbf{S}_{\backslash vn},\boldsymbol{\mathcal{W}},a,b,r,\gamma\right)\propto\frac{p\left(s_{vn}=j,\mathbf{S}_{\backslash vn}|\gamma\right)}{p\left(\mathbf{S}_{\backslash vn}|\gamma\right)}\frac{p\left(\boldsymbol{\Phi}|s_{vn}=j,\mathbf{S}_{\backslash vn},\boldsymbol{\mathcal{W}},a,b,r\right)}{p\left(\boldsymbol{\Phi}_{\backslash vn}|\mathbf{S}_{\backslash vn},\boldsymbol{\mathcal{W}},a,b,r\right)}, \tag{3.12}$$

where $\backslash vn$ represents a value excluding the nth feature vector in the vth view. The first factor in the expression above is given by

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

using (3.7). Then, using (3.7) the second factor becomes

$$\frac{p\left(\boldsymbol{\Phi}|s_{vn}=j,\mathbf{S}_{\backslash vn},\boldsymbol{\mathcal{W}},a,b,r\right)}{p\left(\boldsymbol{\Phi}_{\backslash vn}|\mathbf{S}_{\backslash vn},\boldsymbol{\mathcal{W}},a,b,r\right)}=(2\pi)^{-\frac{D_v}{2}}r^{\frac{1}{2}I(j>J\backslash vn)}\frac{b_{\backslash vn}'^{a'_{vn}}\Gamma\left(a'_{s_{vn}=j}\right)\left|\boldsymbol{\Lambda}_{j,s_{vn}=j}\right|^{\frac{1}{2}}}{b_{s_{vn}=j}'^{a'_{s_{vn}=j}}\Gamma\left(a'_{\backslash vn}\right)\quad\left|\boldsymbol{\Lambda}_{j\backslash vn}\right|^{\frac{1}{2}}}, \tag{3.13}$$

where $s_{vn}=j$ indicates the value when the feature vector $\mathbf{x}_{vn}$ is assigned to cluster j (i.e. correspondence group) as follows (see Appendix B.3 for further details of collapsed

Gibbs sampling stage at E-step),

$$a'_{s_{dn}=j} = a', \tag{3.14}$$

$$b'_{s_{vn}=j} = b'_{\backslash vn} + \frac{1}{2}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^{\top}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right) + \frac{1}{2}\boldsymbol{\mu}_{j\backslash vn}^{\top}\boldsymbol{\Lambda}_{j\backslash vn}^{-1}\boldsymbol{\mu}_{j\backslash vn} - \frac{1}{2}\boldsymbol{\mu}_{j,s_{vn}=j}^{\top}\boldsymbol{\Lambda}_{j,s_{vn}=j}^{-1}\boldsymbol{\mu}_{j,s_{vn}=j}, \tag{3.15}$$

$$\mu_{j,s_{vn}=j} = \boldsymbol{\Lambda}_{j,s_{vn}=j}\left(\boldsymbol{B}_v^{\top}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right) + \boldsymbol{\Lambda}_{j\backslash vn}^{-1}\boldsymbol{\mu}_{j\backslash vn}\right), \tag{3.16}$$

$$\boldsymbol{\Lambda}_{j,s_{vn}=j}^{-1} = \boldsymbol{B}_v^{\top}\boldsymbol{B}_v + \boldsymbol{\Lambda}_{j\backslash vn}^{-1}, \tag{3.17}$$

and $I(\cdot)$ is used to denote the indicator function, i.e. $I(A) = 1$ if A is true, $I(A) = 0$ otherwise.

In the M-step, the projection matrices $\boldsymbol{\mathcal{W}}$ are estimated by maximizing the logarithm of the joint likelihood (3.6). The gradient of the joint likelihood is computed by

$$\begin{aligned}\frac{\partial \log p\left(\mathbf{X}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \boldsymbol{B}_v} &= -\frac{a'}{b'}\frac{\partial b'}{\partial \boldsymbol{B}_v} + \frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\Lambda}_j^{-1}\frac{\partial \boldsymbol{\Lambda}_j}{\partial \boldsymbol{B}_v}\right) \\ &= -\frac{a'}{b'}\left[\sum_{j=1}^{J}\left\{N_{vj}\boldsymbol{B}_v\boldsymbol{\mu}_j\boldsymbol{\mu}_j^{\top} - \sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\boldsymbol{\mu}_j^{\top}\right\}\right] \\ &\quad - \sum_{j=1}^{J}N_{vj}\boldsymbol{B}_v\boldsymbol{\Lambda}_j.\end{aligned} \tag{3.18}$$

We can obtain the projection matrices that maximize the joint likelihood analytically as follows,

$$\boldsymbol{B}_v = -\frac{a'}{b'}\left(\sum_{j=1}^{J}\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\boldsymbol{\mu}_j^{\top}\right)\left(\sum_{j=1}^{J}N_{vj}\boldsymbol{\Lambda}_j + \frac{a'}{b'}N_{vj}\boldsymbol{\mu}_j\boldsymbol{\mu}_j^{\top}\right)^{-1}. \tag{3.19}$$

See Appendix B.3 for the complete derivation of the projection matrices.

Finally, Algorithm 1 shows the procedure for inferring our approach based on the stochastic EM algorithm. For the input, we initialize the cluster assignments $\mathbf{S}$ by randomly selecting an integer from $\{1, \ldots, J\}$. The projection matrices $\boldsymbol{\mathcal{W}}$ are initialized by Gaussian with zero mean and small variance. T is the maximum number of iteration for the learning algorithm.

Algorithm 1 Learning procedure for the nonlinear unsupervised clustering matching.

Input: Multiview dataset projected into the feature space $\boldsymbol{\Phi}$, initial number of clusters J, maximum iteration number T, and hyperparameters a, b, r, γ

Initialize $\mathbf{S}$ and $\boldsymbol{\mathcal{W}}$.
for $t = 1, \ldots, T$ **do**
  // E-Step
  for $v = 1, \ldots, V$ **do**
    for $n = 1, \ldots, N_v$ **do**
      sample s_{vn} using the $p\left(j|\boldsymbol{\Phi}, \mathbf{S}_{\setminus vn}, \boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)$ (3.12) from $j = 1, \ldots, J+1$
      if $s_{vn} = J+1$ **then**
        update the number of clusters $J \leftarrow J+1$
      end if
    end for
  end for
  // M-Step
  for $v = 1, \ldots, V$ **do**
    update each projection matrix $\boldsymbol{B}_v$ using (3.19)
  end for
end for

Output: projection matrices $\boldsymbol{\mathcal{W}}$, and cluster assignments $\mathbf{S}$ (i.e., groupwise correspondences)

3.4 Random Fourier Features

Since the model parameters depicted above depend on the feature vectors $\boldsymbol{\phi}(\mathbf{x}_{vn})$, the expression for $\boldsymbol{\mu}_j$ in the equation (3.9) becomes intractable due to the non-linear relationships between variables. Here, we can notice that a given kernel induces an inner product between infinite dimensional feature vectors. As in Rahimi and Recht (2007), we propose to approximate the mapping functions $\boldsymbol{\phi}(\mathbf{x}_{vn})$ by computing a randomized feature map $\boldsymbol{\varphi}(\mathbf{x}_{vn}) : \mathcal{R}^{D_v} \rightarrow \mathcal{R}^{L_v}$ so that the inner product in equation (3.8) ensures that we can approximate the kernel, $k(\mathbf{x}, \mathbf{x}') = \langle \boldsymbol{\phi}(\mathbf{x}), \boldsymbol{\phi}(\mathbf{x}') \rangle \approx \boldsymbol{\varphi}(\mathbf{x}_{vn})^{\top} \boldsymbol{\varphi}(\mathbf{x}_{vn})$ (Rahimi and Recht, 2007). Thus, we consider the spherical Gaussian kernel as $k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma||\mathbf{x} - \mathbf{x}'||^2)$. Consequently, we compute these feature vectors by using random Fourier bases as

$$\boldsymbol{\varphi}(\mathbf{x}_{vn}) \equiv \sqrt{\frac{2}{L_v}} \begin{bmatrix} \cos\left(\boldsymbol{\omega}_1^{\top} \mathbf{x}_{vn} + \upsilon_1\right) \\ \vdots \\ \cos\left(\boldsymbol{\omega}_{L_d}^{\top} \mathbf{x}_{vn} + \upsilon_{L_d}\right) \end{bmatrix}, \tag{3.20}$$

where $\{\boldsymbol{\omega}_m \sim \mathcal{N}(\mathbf{0}, \beta^{-1}\mathbf{I})\}_{m=1}^{L_v}$ and υ_m is drawn from the uniform distribution as

$$\{\upsilon_m \sim \mathcal{U}(0, 2\pi)\}_{m=1}^{L_d}. \tag{3.21}$$

3.5 Experimental results

3.5.1 Synthetic datasets

The table 3.1, shows four different metrics for comparing partitions (Hubert and Arabie, 1985). Results for adjusted Rand index, unadjusted Rand index, Hubert's index, and Mirkin's index are reported for comparison (See appendix A for details of the evaluation metrics for cluster analysis). Every index was computed for the three synthetic datasets (latent dimensionality $K = 3, \quad 5, \quad 10$), which were averaged over ten experiments. The results show that even for high latent dimensionalities, the proposed method is able to group data successfully ($AR = 0.9096$ for $Synth10$). Thus, the model can infer matching clusters by assuming a shared latent space.

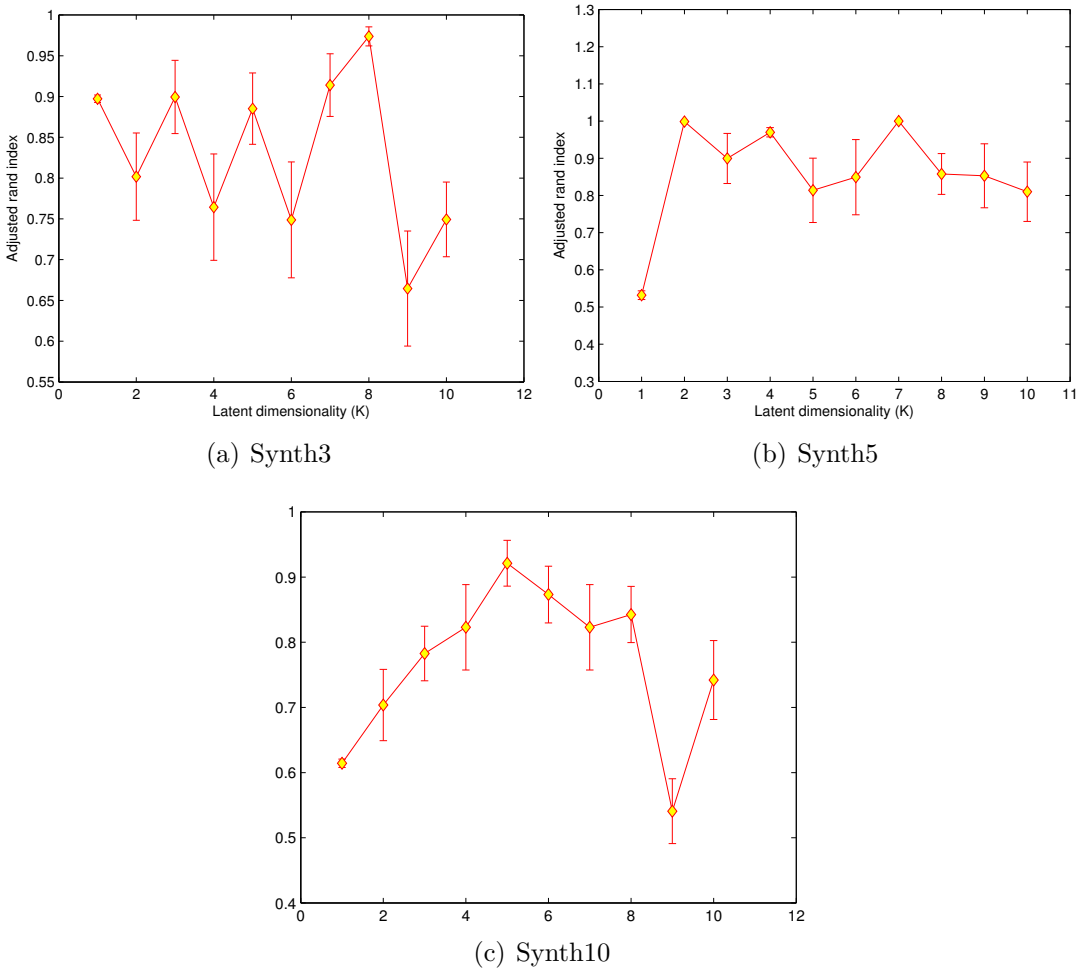
Table 3.1: Average of the Rand indexes and its standard deviation ($mean \pm std$), for the three synthetic datasets. The table shows the adjusted rand intex (AR), the unadjusted rand index (RI), Hubert's index and the Mirkin's index.

Dataset	**AR**	**RI**	**Hubert**	**Mirkin**
Synth3	0.9757 ± 0.0559	0.9897 ± 0.0235	0.9794 ± 0.0469	0.2653 ± 0.0444
Synth5	0.9763 ± 0.0502	0.9898 ± 0.0216	0.9796 ± 0.0433	0.2316 ± 0.0887
Synth10	0.9096 ± 0.1903	0.9607 ± 0.0831	0.9214 ± 0.1661	0.2968 ± 0.1435

Figure 3.4, shows the adjusted Rand index for all three synthetic datasets achieved by the proposed model with different latent dimensionalities. The value was highest when the latent dimensionality of the model was the same (or near) as the true latent dimensionality. The proposed model with $K \neq K*$ also performed accurate matching. This result indicates that the proposed model is robust to the latent dimensionality settings.

In addition, figure 3.5 shows a few examples of the resulting likelihood for the three synthetic datasets (latent space dimensionality $K = 3, 5, 10$). The figure shows that the likelihoods reach their maximums values among the iterations properly.

Finally, figure 3.6 shows the experimental results when the number of views, V, increases for the three synthetic datasets. The result shows that our model improves the



(a) Synth3

(b) Synth5

(c) Synth10

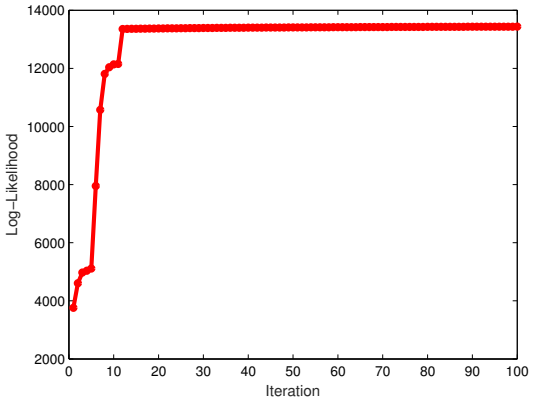
Figure 3.4: Adjusted Rand index achieved by the proposed model with different latent dimensionalities K for the all datasets whose true latent dimensionalities are $K* = 3, \quad 5, \quad 10$.

performance of the Iwata's approach.

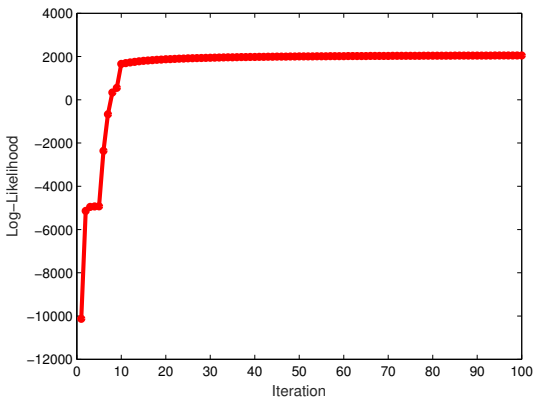
To explore the accuracy of our approach, we show the performance of our approach on real-world non-rigid datasets. Then, we present a comparative analysis of unsupervised clustering methods over well-known machine learning databases. Finally, we discuss the benefits of performing probabilistic correspondence analysis over neuroimaging data.

3.5.2 Non-rigid real-world datasets

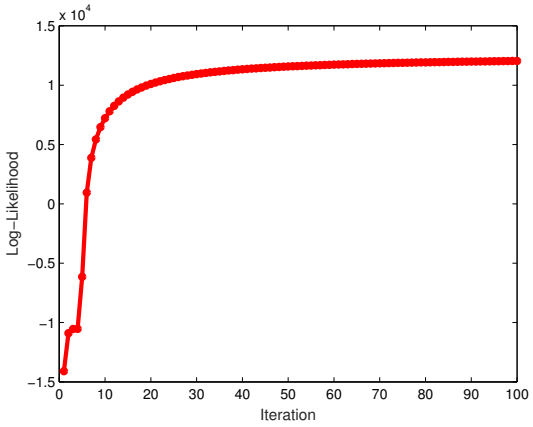
First, we used the Gorilla 3D shapes of the TOSCA dataset to show more comprehensively the correspondence problem over non-rigid shapes. This dataset contains three-



(a) Synth3



(b) Synth5



(c) Synth10

Figure 3.5: Samples of the log-likelihood for the proposed model with different latent dimensionalities K for the all datasets whose true latent dimensionalities are $K* = 3, \quad 5, \quad 10$.

dimensional non-rigid shapes in a variety of poses for non-rigid shape similarity and correspondence experiments which is challenging and can give us an important evaluation of our correspondence method. We used the triangular faces and a list of vertex XYZ coordinates as our input data (Bronstein et al., 2007b).

3.5.3 Real-world datasets

Second, we test our method with three well-known machine learning datasets such as Iris, Glass, and MNIST.[1] We set up our experiments by randomly splitting the input data

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

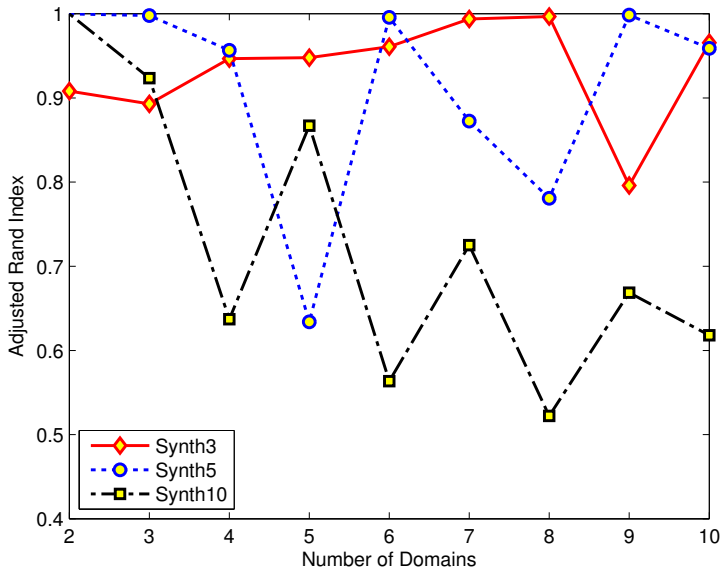


Figure 3.6: Adjusted Rand index for the three synthetic datasets ($K* = 3, \quad 5, \quad 10$) with different numbers of views, V.

(i.e., features of the datasets) into two views as Iwata et al. did for their experiments in (Iwata et al., 2016).

3.5.4 Brain structures dataset

For the neuroimage analysis, 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 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.

3.5.5 TOSCA Non-rigid shapes

To show more appropriately the correspondence results on 3D shapes, we test our method on the TOSCA dataset. We compute 3D shape descriptors based on scale-invariant

Heat Kernel Signatures (SI-HKS) as in (Bronstein and Kokkinos, 2010). We used these descriptors as input data to compute the correspondences. We test our method against KM and UCM in correspondence tasks. Figure 3.7 shows the experimental results of the non-rigid correspondence analysis. The results show that our method establishes more meaningful correspondences between two real-world samples than KM and UCM. We can notice from these results that our approach finds accurate correspondences between shapes (Standing and sitting gorilla respectively). For instance, we can see in figure 3.7(c) that regions related to arms, head, and body match entirely in both shapes (that is not the case for the KM, and UCM approaches, where hands are matched wrongly, see figures 3.7(a) and 3.7(b)).

3.5.6 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 (Iwata et al., 2016). For comparison, we use unsupervised clustering matching (UCM) (Iwata et al., 2016), k-means (KM), and convex kernelized sorting (CKS) (Djuric et al., 2012). Table 3.2 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).

Table 3.2: Adjusted Rand index of the proposed method against the state-of-the-art methods for unsupervised clustering.

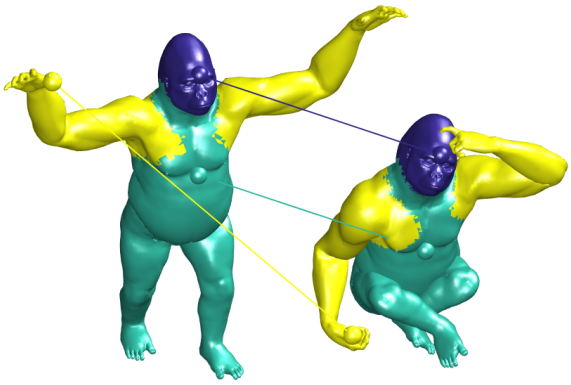
	Approach			
Database	UCM	KM	KM-CKS	**Ours**
Iris	0.383 ± 0.189	0.224 ± 0.0910	0.254 ± 0.154	$\mathbf{0.546 \pm 0.080}$
Glass	0.160 ± 0.020	0.050 ± 0.008	0.052 ± 0.011	$\mathbf{0.378 \pm 0.045}$
MNIST	0.085 ± 0.016	0.030 ± 0.007	0.037 ± 0.008	$\mathbf{0.167 \pm 0.013}$

3.5.7 Groupwise shape correspondences

To establish groupwise correspondences between brain structures, we used again the SI-HKS as shape descriptors. Then, we perform a random feature expansion to compute the features vectors. We set each view as a 3D shape descriptor for a given brain structure.

(a) KM

(b) UCM



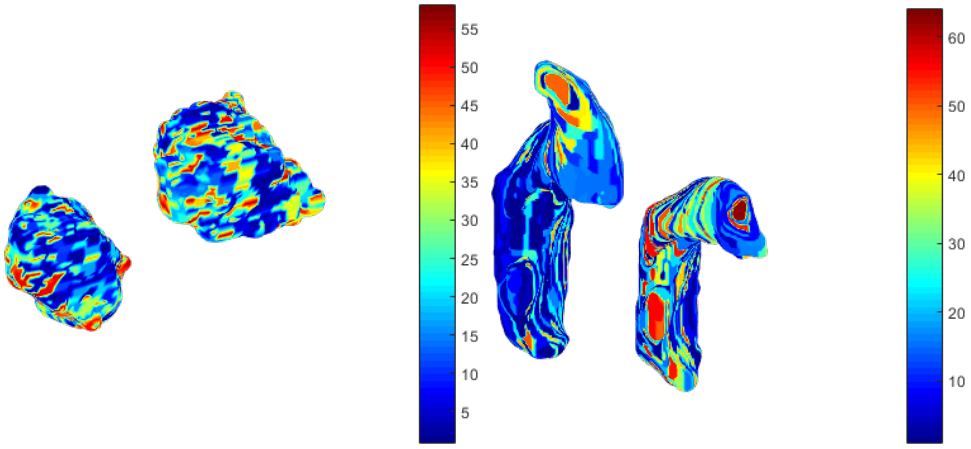
(c) Ours

Figure 3.7: Correspondence estimates for different clustering methods. The figure shows a comparison of correspondence experiments on different Gorillas shapes from TOSCA dataset.

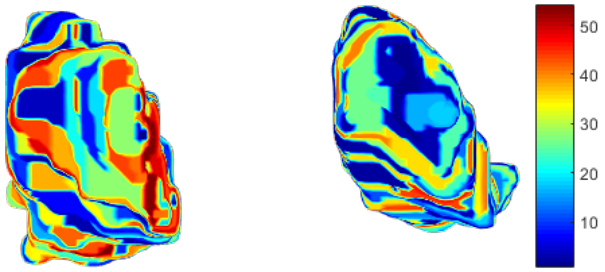
We evaluate our model by using three relevant brain structures in the Alzheimer's disease such as the ventricle, thalamus, and putamen.

Figure 3.8 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 3.8(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

Figure 3.8: 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). Same colors are considered as candidate matching regions.

In addition, table 3.3 shows the comparison for different surface descriptors, SI-HKS, the Classic HKS and the classic Laplace-Beltrami operator (LBO) (Bronstein and Kokkinos, 2010). The results show that when we used SI-HKS, the correspondence is more accurate than HKS and LBO (i.e. adjusted rand index of 0.287 ± 0.035, 0.269 ± 0.030 and 0.221 ± 0.053 for ventricle brain structures). These results are evidence of the invariant properties of SI-HKS, in which we can model locally-elastic deformations that stretch or shrink the brain structure in a given neurodegenerative process.

Finally, the table 3.4 shows both mean and standard deviations computed from

Table 3.3: Adjusted Rand index for different surface descriptors on brain structures.

Brain Structure	Ventricle	Putamen	Thalamus
SI-HKS	$\mathbf{0.287 \pm 0.035}$	$\mathbf{0.312 \pm 0.023}$	$\mathbf{0.332 \pm 0.017}$
HKS	0.269 ± 0.030	0.305 ± 0.042	0.311 ± 0.047
LBO	0.221 ± 0.053	0.253 ± 0.048	0.266 ± 0.042

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 3.4: 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	$\mathbf{0.332 \pm 0.017}$

As for the correspondence evaluation, we use an optimal correspondence metric based on the structure preservation criterion shown in (Kim et al., 2011), which is the Princeton benchmark protocol for correspondence quality. Hence, one can obtain a criterion of shape similarity as the amount of structure distortion. The results for the TOSCA dataset, show that our method computes point-to-point relations for geodesic errors lowers than 0.18, which is a good result in comparison with the correspondences approaches in the state-of-the-art. Additionally, the results for the Brain dataset show geodesic errors lower than 0.29, which is a good result for partial shape matching approaches in the state-of-the-art. These results evidence that for large deformations over a given brain structure, the distortion increases as a result of the neurodegenerative disease (partial correspondence problem).

3.6 Conclusions

In this chapter, 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 views 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 and we can match non-rigid real world shapes accurately. In addition, since the inferred correspondences fit 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.

Chapter 4

Multiview Warped Mixture Model

In the previous chapter, we showed how nonlinear latent variable models allow the automatic determination of the proper relation between nonrigid shapes when building unsupervised matching models for shape correspondence analysis. However, an important issue that arose in the previous chapter is the need to define a view-specific underlying representation for each shape. Hence, modeling shape correspondences with shared latent vectors sometimes lead to mismatches when the number of views grows (Zhao et al., 2017). Moreover, when shape descriptors from different views are related to a common latent feature vector, the information about the private latent structure of a given data-view is limited (i.e., poor representation of complex shape variations within views). Besides, since the feature map for the model proposed in Chapter 3 is built using samples drawn from a spectral measure (i.e., through random Fourier features) (Rahimi and Recht, 2007), the complete kernelized version of the multiview approach remains to be reached.

In consequence, this chapter presents a multiview learning approach to model nonlinear groupwise correspondences of complex shapes by using a multiview warped mixture model. In particular, we assume that each shape descriptor has its correspondence in the latent space, and is generated by warping the latent correspondence via nonlinear mapping functions from the latent space to the input space. Thus, we use Gaussian processes (GP) (Rasmussen and Williams, 2005), which enable us to infer the nonlinear warping function from the input data (i.e., shape descriptors). Gaussian processes have also been used in unsupervised learning and dimensionality reduction scenarios where the output dimensionality of the input data can even be much larger than the instances (i.e., number of shape descriptors per view N_v)(Lawrence and Quiñonero Candela, 2006).

However, the main challenge in the unsupervised GP problems is that the input

data set $\mathbf{X}_v$ is not directly observed. Therefore, Gaussian process latent variable models (GP-LVM) (Lawrence, 2005) offer an elegant solution to this problem by treating the unobserved inputs as latent variables, while employing a product of D_v independent GPs as prior for the latent mapping.

In fact, the presence of multiple views motivate Gaussian Process latent variable models, which relates different views (i.e., pair of 3D shapes) by assuming that the data variance is shared among views (Damianou et al., 2012). Thus, the remaining variance is explained with latent spaces that are private to each view (i.e., by preserving the non-related shape information among views). As a result, this property will derive in a more meaningful correspondence descriptor.

In addition, we assume an infinite Gaussian mixture (iGMM) model in the latent space, which allows us to infer the shape correspondences automatically (i.e., by clustering candidate objects) (Iwata et al., 2016). As in Tomoharu Iwata (2013), we use Gaussian processes to set the priors on the nonlinear mapping functions. Thus, this prior will enable us to infer the nonlinear warping function from the shape descriptors flexibly.

Besides, the model structure allows inference when partial shape correspondence per view is available (i.e., by handling observations of different size per view) and, because the observation spaces have been warped, it is possible to transfer shape information between views by conditioning the model through the underlying concept.

4.1 Gaussian Process Latent Variable Models

Gaussian process latent variable models (GP-LVM) are probabilistic, non-linear, latent variable models that generalize principal component analysis (Lawrence, 2005). Here, a GP-LVM brings an efficient probabilistic mapping and also allows the uncertainties of the low dimensional embedding data to be estimated. By additionally learning a latent dynamical model, the GP-LVM gives a closed-form expression for the joint distribution of the observed sequences and their latent space representations (Ek et al., 2007). Therefore, we can regularize the latent space for modeling shape sequences by incorporating dynamics to the model.

4.1.1 Model definition

Let us define a set of observations $\mathbf{X}_v = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_v}]^\top$ for a single v-th view, with $\mathbf{x}_i \in \mathfrak{R}^{D_v}$. Similarly, let $\mathbf{Z}_v = [\mathbf{z}_1, \ldots, \mathbf{z}_{N_v}]^\top$ denote the matrix whose rows represent corresponding positions in the latent space, $\mathbf{z}_i \in \mathfrak{R}^K$. The Gaussian Process Latent Variable Model relates a high-dimensional data set, $\mathbf{X}_v$, and a low dimensional latent space, $\mathbf{Z}_v$, using a Gaussian process mapping from the latent space to the input space (Lawrence, 2005). Given a covariance function for the Gaussian process, $k(\mathbf{z}, \mathbf{z}')$, the likelihood of the data given the latent positions is defined as

$$p(\mathbf{X}_v|\mathbf{Z}_v, \boldsymbol{\theta}^v) = \frac{1}{\sqrt{(2\pi)^{N_v D_v}|\mathbf{K}_v|^{D_v}}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\mathbf{K}_v^{-1}\mathbf{X}_v\mathbf{X}_v{}^\top\right)\right), \tag{4.1}$$

where elements of the kernel matrix $\mathbf{K}_v$ are defined by the covariance function $(\mathbf{K}_v)_{i,j} = k(\mathbf{z}_i, \mathbf{z}_j)$, and the kernel hyperparameters by $\boldsymbol{\theta}^v$. We use a radial basis function (RBF) kernel with an additive noise term,

$$k(\mathbf{z}, \mathbf{z}') = \alpha \exp\left(-\frac{1}{2l^2}||\mathbf{z} - \mathbf{z}'||^2\right) + \delta_{\mathbf{z},\mathbf{z}'}\beta^{-1},$$

where α is the output variance, l^2 is a scale parameter and $\delta_{\mathbf{z},\mathbf{z}'}$ denotes the Kronecker delta.

The corresponding log-likelihood is then

$$L = -\frac{D_v N_v}{2}\ln(2\pi) - \frac{D_v}{2}\ln|\mathbf{K}_v| - \frac{1}{2}\mathrm{tr}\left(\mathbf{K}_v^{-1}\mathbf{X}_v\mathbf{X}_v{}^\top\right). \tag{4.2}$$

Typically, the GP-LVM is used for dimensionality reduction or high dimensional data visualization, and the latent coordinates are set by maximizing equation (4.1). Currently, the primary methodology for training the GP-LVM model is to find the MAP estimate of $\mathbf{Z}$ while jointly optimize the hyperparameters (Lawrence, 2005). In that way, the Gaussian prior density on $\mathbf{z}_i$ is a regularizer which keeps the latent coordinates close in the feature space (i.e., Local Distance Preservation in the GP-LVM through Back Constraints) (Lawrence and Quiñonero Candela, 2006). As in Tomoharu Iwata (2013), we instead integrate out the latent coordinates as well as the warping function.

4.2 Warped Mixtures for Unsupervised Clustering

Let us first define the single view Warped Mixture Model (WMM) described in Tomoharu Iwata (2013), which is a nonparametric Bayesian model that finds nonlinearly separable clusters. The model aims to compute cluster assignments by warping a latent mixtures of Gaussians. Thus, the simple generative model of non-Gaussian density manifolds allows us to infer shape correspondences nonlinearly. Here, we first introduce the key components of this single view model.

As in the GP-LVM, the infinite WMM assumes a smooth nonlinear mapping from a latent density to an observed density. Here, the only difference is that the iWMM assumes that the latent density is an infinite Gaussian mixture model (iGMM)

$$p(\mathbf{z}) = \sum_{c=1}^{\infty} \lambda_c \mathcal{N}\left(\mathbf{z}|\boldsymbol{\mu}_c, \mathbf{R}_c^{-1}\right), \tag{4.3}$$

where λ_c, $\boldsymbol{\mu}_c$ and $\mathbf{R}_c$, denote the mixture weight, mean, and precision matrix of the cth mixture component. The iWMM can be seen as a generalization of either the GP-LVM or the iGMM: The iWMM with a single fixed spherical Gaussian density on the latent coordinates $p(\mathbf{z})$ corresponds to the GP-LVM, while the iWMM with fixed mapping $\mathbf{x} = \mathbf{z}$ and $K = D$ corresponds to the iGMM (Tomoharu Iwata, 2013).

4.3 Latent Multiview Warped Mixture Model

We can use the single view problem of the infinite warped mixture model (iWMM) from Tomoharu Iwata (2013) in which they warp a latent mixture of Gaussians into nonparametric cluster shapes. Our idea is to extend the single view model by warping a multiview latent mixture of Gaussians to produce nonparametric groupwise correspondences. In addition, the possibly low-dimensional latent mixture model allows us to summarize the properties of the high-dimensional shape correspondences (or density manifolds) describing the shape descriptors. The number of manifolds, as well as the shape and dimension of each manifold is automatically inferred.

4.3.1 The model

Let us define a multiview data set as $\mathbf{X} = \{\mathbf{X}^v\}_{v=1}^V$, where each view is defined as $\mathbf{X}^v \in \mathbb{R}^{N_v \times D_v}$. Our model can be seen as an extension of iGMM, where mixtures are

warped (i.e., by using a multiview GPLVM). Thus, given the mixture assignments (e.g., shape correspondences) the likelihood is set as

$$p\left(\mathbf{X}, \mathbf{Z}|\mathbf{S}, \boldsymbol{\theta}\right) = \prod_{v=1}^{V} p(\mathbf{X}^v|\mathbf{Z}^v, \boldsymbol{\theta}^v) \times \prod_{i \in v} \sum_{c=1}^{\infty} \lambda_c \mathcal{N}\left(\mathbf{z}_i|\boldsymbol{\mu}_c, \mathbf{R}_c^{-1}\right), \quad \mathbf{x}_i \in \mathbf{S}_c \tag{4.4}$$

We define the dimensionalities of our variables as:

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

Based on the iWMM (see Tomoharu Iwata (2013)) our generative model generates multiview observations $\mathbf{X}_v$ according to the following generative process:

1. Draw mixture weights $\boldsymbol{\lambda} \sim \text{Stick}(\eta)$
2. For each cluster $c = 1, \cdots, \infty$
   (a) Draw precision $\mathbf{R}_c \sim \mathcal{W}(\boldsymbol{\Lambda}^{-1}, v)$
   (b) Draw mean $\boldsymbol{\mu}_c \sim \mathcal{N}(\mathbf{u}, (r\mathbf{R}_c)^{-1})$
3. For each view $v = 1, \cdots, V$
   (a) For each observation $n = 1, \cdots, N_v$
      i. Draw latent assignment $s_{nv} \sim \text{Mult}(\boldsymbol{\lambda})$
      ii. Draw latent coordinates $\mathbf{z}_{nv} \sim \mathcal{N}(\boldsymbol{\mu}_{s_{nv}}, \mathbf{R}_{s_{nv}}^{-1})$
4. For each view $v = 1, \cdots, V$
   (a) For each observed dimension $d = 1, \cdots, D_v$
      i. Draw function $\mathbf{f}_d^v \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}^v)$

5. For each view $v = 1, \cdots, V$

   (a) For each observed dimension $d = 1, \cdots, D_v$

      i. Draw projection variable $w_d^v \sim \mathcal{N}(0, \rho_d^v)$

      ii. For each observation $n = 1, \cdots, N_v$

         A. Draw feature $x_{nd}^v \sim \mathcal{N}(w_d^v f_d^v(\mathbf{z}_{nv}), \beta^{-1})$

Here, r is the relative precision of $\boldsymbol{\mu}_c$, Stick(η) is the stick-braking process with parameter η (see 3.2.1), Mult($\boldsymbol{\lambda}$) represents a multinomial distribution with parameter $\boldsymbol{\lambda}$, and $\mathcal{W}(\cdot)$ is the Wishart distribution defined as

$$\mathcal{W}\left(\mathbf{R}|\boldsymbol{\Lambda}^{-1}, \nu\right) = \frac{1}{G}|\mathbf{R}|^{\frac{\nu-K-1}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}\mathbf{R})\right) \tag{4.5}$$

where ν is called the number of degrees of freedom of the distribution, and G is the normalizing constant.

Our model is set as a multiview Gaussian Process Latent Variable model as in Lawrence (2005). First we assume that observations for each view 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

$$p\left(\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta}\right) = \prod_{v=1}^{V} p\left(\mathbf{X}_v|\mathbf{Z}_v, \boldsymbol{\theta}^v\right) = \prod_{v=1}^{V}\prod_{d=1}^{D_v} p\left(\mathbf{x}_d^v|\mathbf{Z}_v, \boldsymbol{\theta}^v\right), \tag{4.6}$$

where $\mathbf{x}_d^v$ represents the dth column of $\mathbf{X}_v$ and

$$p\left(\mathbf{x}_d^v|\mathbf{Z}_v, \boldsymbol{\theta}^v\right) = \mathcal{N}\left(\mathbf{x}_d^v|\mathbf{0}, \beta^{-1}\mathbf{I} + w_d^2\mathbf{K}_v\right). \tag{4.7}$$

Our multiview 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,

$$p(\mathbf{z}_v|\lambda, \boldsymbol{\mu}, \mathbf{R}) = \sum_{c=1}^{\infty} \lambda_c \mathcal{N}\left(\mathbf{z}_v|\boldsymbol{\mu}_c, \mathbf{R}_c^{-1}\right), \tag{4.8}$$

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

As in the iWMM (Tomoharu Iwata, 2013), we place Gaussian-Wishart priors on the Gaussian parameters $\{\boldsymbol{\mu}_c, \mathbf{R}_c\}$.

$$p\left(\boldsymbol{\mu}_c, \mathbf{R}_c\right) = \mathcal{N}\left(\boldsymbol{\mu}_c | \mathbf{u}, (r\mathbf{R}_c)^{-1}\right) \mathcal{W}\left(\mathbf{R}_c | \boldsymbol{\Lambda}^{-1}, \nu\right), \tag{4.9}$$

where $\mathbf{u}$ is the mean of $\boldsymbol{\mu}_c$, r is the relative precision of $\boldsymbol{\mu}_c$, $\boldsymbol{\Lambda}^{-1}$ 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 the input observations to the components. Let s_{nv} be the latent assignment of the nth object in the vth view. The probability of latent coordinates $\mathbf{Z}_v$ given latent assignments $\mathbf{S}_v = (s_1, \ldots, s_{N_v})$ is obtained by integrating out the Gaussian parameters $\{\boldsymbol{\mu}_c, \mathbf{R}_c\}$ as follows:

$$p\left(\mathbf{Z}_v | \mathbf{S}_v, \boldsymbol{\Lambda}, \nu, r\right) = \prod_{c=1}^{\infty} \pi^{-\frac{\sum_v N_{vc} K}{2}} \frac{r^{K/2} |\boldsymbol{\Lambda}|^{\nu/2}}{r_c^{K/2} |\boldsymbol{\Lambda}_c|^{\nu_c/2}} \prod_{q=1}^{K} \frac{\Gamma\left(\frac{\nu_c + 1 - q}{2}\right)}{\Gamma\left(\frac{\nu + 1 - q}{2}\right)}, \tag{4.10}$$

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

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

are the posterior Gaussian-Wishart parameters of the cth component. We use a Dirichlet process with concentration parameter η for infinite mixture modeling in the latent space.

The probability of $\mathbf{S}$ is given as follows

$$p\left(\mathbf{S}|\eta\right) = \prod_{v=1}^{V} \frac{\eta^{C} \prod_{c=1}^{C} \left(N_{vc} - 1\right)!}{\eta\left(\eta + 1\right) \cdots \left(\eta + N_{v} - 1\right)}, \tag{4.11}$$

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

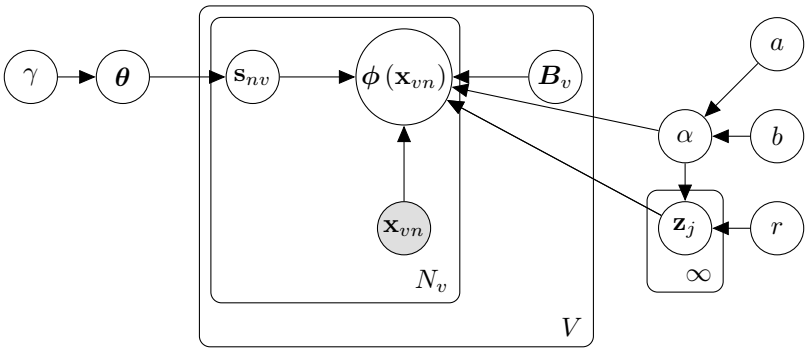
$$p\left(\mathbf{X}, \mathbf{Z}, \mathbf{S}|\boldsymbol{\theta}, \nu, \mathbf{u}, r, \eta, \right) = \prod_{v=1}^{V} p\left(\mathbf{X}_{v}|\mathbf{Z}_{v}, \boldsymbol{\theta}^{v}\right) p\left(\mathbf{Z}_{v}|\mathbf{S}_{v}, \boldsymbol{\Lambda}, \nu, \mathbf{u}, r\right) p\left(\mathbf{S}_{v}|\eta\right). \tag{4.12}$$

Figures 4.1(a) and 4.1(b) show the graphical representation of the proposed models (figure 4.1(a) refers to the model proposed in chapter 3), where the shaded and unshaded nodes indicate observed and latent variables respectively, and plates indicate repetition. The probabilistic nature of the multiview WMM let us automatically infer the number, dimension, and shape of a set of nonlinear multiview manifolds, and represent the observed multiview data in a low-dimensional latent space (see figure 4.1(b)). We assume for this model that each shape descriptor has its correspondence in the latent space, whereas the model in chapter 3) assumes that latent vectors $\mathbf{z}_j$ are shared among views.

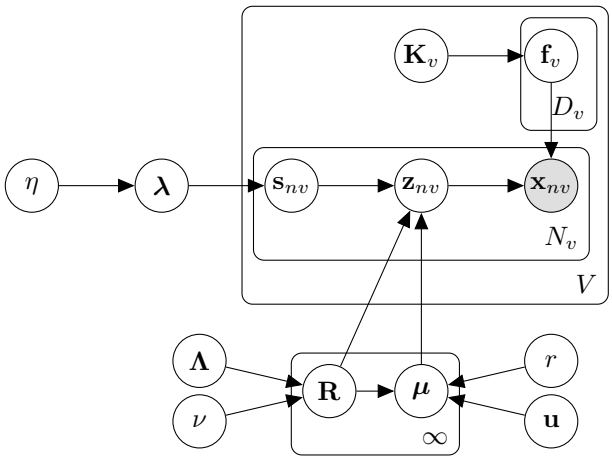
4.3.2 Inference

We infer the posterior distribution of the latent coordinates $\mathbf{Z} = \{\mathbf{Z}_v\}_v^V$ and cluster assignments $\mathbf{S}_v$ using Markov chain Monte Carlo (MCMC). In particular, we alternate collapsed Gibbs sampling of $\mathbf{S}$ (Liu, 1994), and Hamiltonian Monte Carlo (HMC) sampling of $\mathbf{Z}$ (see appendix C). Given $\mathbf{Z}_v$, we can efficiently sample $\mathbf{S}_v$ using collapsed Gibbs sampling, integrating out the mixture parameters. Then from $\mathbf{S}_v$, we can calculate the gradient of the unnormalized posterior distribution of $\mathbf{Z}_v$, integrating over warping functions. This gradient allows us to sample $\mathbf{Z}$ using hybrid Monte Carlo (see appendix C.1.4).

First, we derive the collapsed Gibbs sampling for $\mathbf{S}$. Here, given a sample of $\mathbf{Z}$, $p\left(\mathbf{S}|\mathbf{Z}, \boldsymbol{\Lambda}, v, \mathbf{u}, r, \eta\right)$ does not depend on $\mathbf{X}_v$. This lets resampling cluster assignments, integrating out the iGMM likelihood in closed form (equation (4.10)). Given the current state of all but one latent component s_{nv}, a new value for s_{nv} is sampled from the following probability distributions:



(a) NL-UCM



(b) Multiview WMM

Figure 4.1: Evolution of the structure of unsupervised multiview model variants. 4.1(a) NL-UCM, where latent vectors $\mathbf{z}_j$ are shared among views. Figure 4.1(b), shows the graphical model we propose in this chapter. In this model, we assume that each shape descriptor has its correspondence in the latent space $\mathbf{z}_{nv}$, and is generated by warping the latent correspondence via nonlinear mapping functions from the latent space to the multiview input space. Differences with the single-view model as in Tomoharu Iwata (2013) can be seen in the multiview formulation of the graphical model.

$$p\left(s_{nv}=c|\mathbf{Z},\mathbf{S}_{\backslash nv},\boldsymbol{\Lambda},v,\mathbf{u},r,\eta\right)\propto\begin{cases}N_{c\backslash nv}\cdot p\left(\mathbf{z}_{nv}|\mathbf{Z}_{c\backslash nv},\mathbf{S},v,\mathbf{u},r\right) & \text{existing components}\\ \eta\cdot p\left(\mathbf{z}_{nv}|\mathbf{S},v,\mathbf{u},r\right) & \text{a new cluster,}\end{cases}\tag{4.13}$$

where $\mathbf{Z}_c=\{\mathbf{z}_{nv}|s_{nv}=c\}$ is the set of latent coordinates assigned to the c^{th} component, and $\backslash nv$ represents the value or set when excluding the n-th observation in the v-th view. We can analytically calculate $p\left(\mathbf{z}_{nv}|\mathbf{Z}_{c\backslash nv},\boldsymbol{\Lambda},v,\mathbf{u},r\right)$ leading to (Fink, 1997):

$$p\left(\mathbf{z}_{nv}|\mathbf{Z}_{c\backslash nv},\mathbf{\Lambda},v,\mathbf{u},r\right)=\pi^{-\frac{N_{c\backslash nv}K}{2}}\frac{r_{c\backslash nv}^{K/2}|\mathbf{\Lambda}_{c\backslash nv}|^{\nu_{c\backslash nv}/2}}{r'^{K/2}_{c\backslash nv}|\mathbf{\Lambda'}_{c\backslash nv}|^{\nu'_{c\backslash nv}/2}}\prod_{q=1}^{K}\frac{\Gamma\left(\frac{\nu'_{c\backslash nv}+1-q}{2}\right)}{\Gamma\left(\frac{\nu_{c\backslash nv}+1-q}{2}\right)},$$

where ν'_c, r'_c, and $\mathbf{\Lambda}'_c$ are the posterior parameters for the Gaussian-Wishart distribution of the c-th component (i.e., correspondence group), when observation n-th of the v-th view has been assigned to it. Finally, we compute the determinant of $|\mathbf{\Lambda'}_{c\backslash nv}|$ using the rank-one Cholesky update.[1]

To sample the latent vectors $\mathbf{Z}_v$ through HMC, we need to compute the gradient of the log-unnormalized posterior distribution $\log p(\mathbf{X}_v|\mathbf{Z}_v,\boldsymbol{\theta}^v)+\log p(\mathbf{Z}_v|\mathbf{S}_v,\mathbf{\Lambda},\nu,\mathbf{u},r)$. Thus, the first term of the gradient is computed as

$$\frac{\partial\log p(\mathbf{X}_v|\mathbf{Z}_v,\boldsymbol{\theta}^v)}{\partial\mathbf{K}_v}=-\frac{1}{2}D_v\mathbf{K}_v^{-1}+\frac{1}{2}\mathbf{K}_v^{-1}\mathbf{X}_v\mathbf{X}_v{}^T\mathbf{K}_v^{-1},\tag{4.14}$$

with a Gaussian Kernel per view with additive noise as

$$\frac{\partial k\left(\mathbf{z}_{nv},\mathbf{z}_{mv}\right)}{\partial\mathbf{z}_{nv}}=-\frac{\sigma_{vf}^2}{\ell_v^2}\exp\left(-\frac{1}{2\ell_v^2}\left(\mathbf{z}_{nv}-\mathbf{z}_{mv}\right)^\top\left(\mathbf{z}_{nv}-\mathbf{z}_{mv}\right)\right)\left(\mathbf{z}_{nv}-\mathbf{z}_{mv}\right).\tag{4.15}$$

Finally te second term of the gradient is set as

$$\frac{\partial\log p(\mathbf{Z}_v|\mathbf{S}_v,\mathbf{\Lambda},\nu,\mathbf{u},r)}{\partial\mathbf{z}_{nv}}=-\nu_{s_{nv}}\mathbf{\Lambda}_{s_{nv}}^{-1}\left(\mathbf{z}_{nv}-\mathbf{u}_{s_{nv}}\right).\tag{4.16}$$

We also infer kernel parameters $\boldsymbol{\theta}^v$ via HMC, using the gradient of the log unnormalized posterior with respect to the kernel parameters (see appendix C.2 for more details). Finally, Algorithm 2 shows the process to obtain samples from the posterior $p(\mathbf{Z},\mathbf{S}|\mathbf{X},\boldsymbol{\theta},\mathbf{\Lambda},\nu,\mathbf{u},r,\eta)$ by iterating the following procedure.

[1]The Cholesky decomposition is a decomposition of a Hermitian, positive-definite matrix into the product of a lower triangular matrix and its conjugate transpose, which is useful for efficient numerical solutions. Here, rank-one Cholesky update uses only the diagonal and upper triangle of the original Cholesky decomposition of $\mathbf{\Lambda}$.

Algorithm 2 Sampling procedure for the MV-WMM.

Input: latent assignment $\mathbf{S}$, scale matrix $\mathbf{\Lambda}$, model parameters ν, $\mathbf{u}$, r

1: **for** $v = 1, \ldots, V$ **do**
2: **for** $n = 1, \ldots, N_v$ **do**
3: Sample the correspondence assignment s_{nv} by collapsed Gibbs sampling of $p\left(\mathbf{s}_{nv} = c|\mathbf{X}, \mathbf{S}_{\backslash nv}, \mathbf{\Lambda}, v, \mathbf{u}, r, \eta\right)$ as in (4.13)
4: **end for**
5: Sample latent coordinates $\mathbf{Z}_v$ and kernel hyperparameters $\boldsymbol{\theta}^v$ using hybrid Monte Carlo
6: **end for**

Output: latent cluster assignments $\mathbf{S}$ (i.e., groupwise correspondences) and latent feature vectors $\mathbf{Z}$

4.4 Multiview Constraints for Dynamic Analysis

As for the multiview learning approach, we are facing two types of shape analysis. Intra shapes will be related to those topological changes within the same object (i.e., brain structure analysis over time). Also, inter shapes will be referred to those changes that same objects exhibit but comes from different sources (i.e., brain structure analysis between patients) (Cosa et al., 2013). Since we want to model both intra and inter shapes, by using our multiview learning approach, it is important to constrain our model to avoid overtraining (i.e., maintain correspondence accuracy as the complexity of the shapes increases).

For instance, assessing the neurodevelopmental progress for a given patient with Alzheimer disease, is highly related to the temporal shape variability that brain structures exhibit among views (Hill, 2010). Besides, as for the large variations presented in infants with hypoxic-ischemic encephalopathy (HIE)[2], assessing the neurodevelopmental outcome is crucial when few observations are available (i.e., small sets of MRI studies to monitoring the encephalopathy) (Weeke et al., 2018).

Regarding the study of the temporal component through latent variable models, the prior knowledge is incorporated into the inference of the dynamical model where latent factors determine the observable structure (Hallac et al., 2017). In fact, temporal information modeled through latent variables is considered to derive in a single latent space which represents the global variations of the dynamical system (Anandkumar et al., 2013; Jalali and Sanghavi, 2012). Hence, we choose to constrain the projection vector $\mathbf{w}_v$ and then add a regularizer term to prevent that latent vectors within views start too

[2]Hypoxic-ischemic encephalopathy, or HIE, is the brain injury caused by oxygen deprivation to the brain, also commonly known as intrapartum asphyxia.

far apart. Thus, shape variations in the observed space within views will maintain close.

$$\log p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta}) = \sum_{v=1}^{V} \log p\left(\mathbf{X}_v|\mathbf{Z}_v, \boldsymbol{\theta}^v\right) + \frac{\lambda}{2} \sum_{v=1}^{V} ||\mathbf{w}_v||^2, \tag{4.17}$$

where λ is the regularization coefficient, and $\mathbf{w}_v$ is the view-specific projection vector.

4.5 Results

4.5.1 Clustering performance on real datasets

We first test our model on common clustering ML datasets shown in table 4.1 (see appendix C.3 for details of the ML datasets). 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).

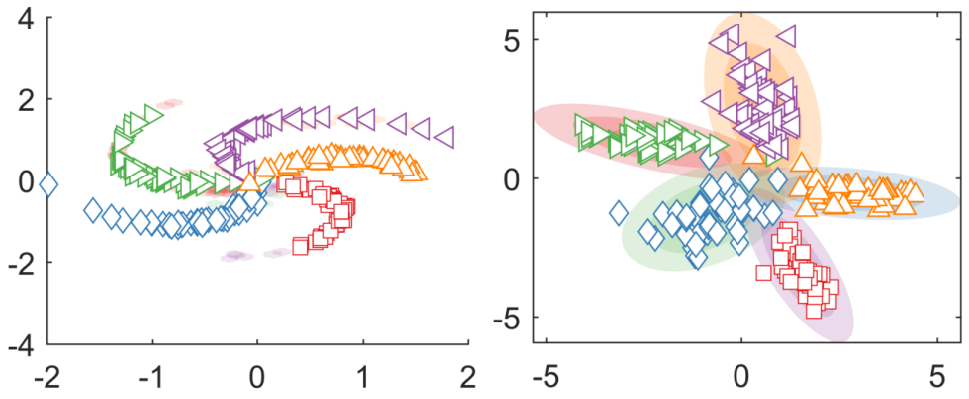
Table 4.1: Average Rand index for evaluating clustering performance.

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

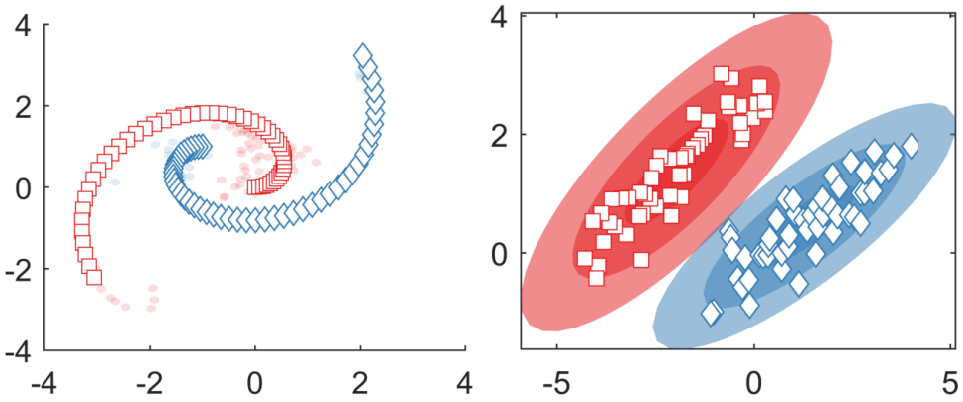
Figure 4.2 shows an example of the clustering matching on synthetic datasets (pinwheel 4.2(a), 2-curve 4.2(b), and 2-circle 4.2(c)). The results show that our model separates the two non-Gaussian clusters in the observed space, representing them using two Gaussian-shaped clusters in the latent space (i.e., clustering performance on the 2-curve and 2-circle dataset). Finally, figure 4.2(a) shows that our approach can recover an analogous latent structure from more complex datasets.

4.5.2 Comparison with linear approaches

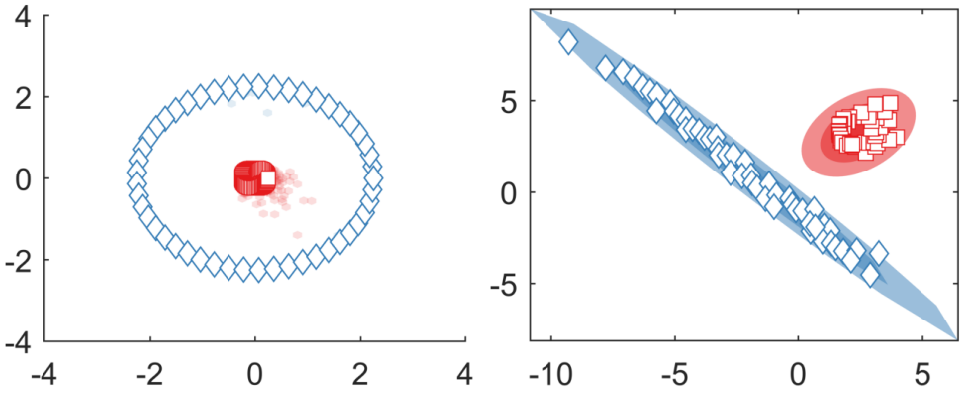
Secondly, we test the performance of our approaches (both NL-UCM and MV-WMM) regarding the adjusted Rand index (we report both average and standard deviation), to quantify the similarity between the inferred clusters (Iwata et al., 2016) and the true labels. For comparison, we use unsupervised clustering matching (UCM) (Iwata et al., 2016), k-means (KM), and convex kernelized sorting (CKS) (Djuric et al., 2012). Table



(a) Pinwheel



(b) 2-curve



(c) 2-circle

Figure 4.2: Experimental results for the Pinwheel, 2-curve and circles dataset. Left column: Observed input data (white markers), and cluster densities inferred by the model (colors). Right column: Latent and Gaussian components from a single sample from the posterior. Each marker plotted in the latent space ($K = 2$) corresponds to an input data in the observed space.

4.2 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 0.33 for the MNIST dataset against 0.085 obtained from the UCM method).

Table 4.2: Adjusted Rand index of the proposed method against the state-of-the-art methods for unsupervised clustering.

	Approach				
Database	UCM	KM	KM-CKS	**NL-UCM**	**MV-WMM**
Iris	0.383 ± 0.189	0.224 ± 0.091	0.254 ± 0.154	0.546 ± 0.080	$\mathbf{0.553 \pm 0.010}$
Glass	0.160 ± 0.020	0.050 ± 0.008	0.052 ± 0.011	0.378 ± 0.045	$\mathbf{0.384 \pm 0.003}$
MNIST	0.085 ± 0.016	0.030 ± 0.007	0.037 ± 0.008	0.167 ± 0.013	$\mathbf{0.334 \pm 0.011}$

4.5.3 Non-rigid 3D shape datasets

To show more appropriately the correspondence results on 3D shapes, we test our method again on the TOSCA dataset (Bronstein et al., 2007a). We compute 3D shape descriptors based on scale-invariant Heat Kernel Signatures (SI-HKS) as in Bronstein and Kokkinos (2010). We used these descriptors as input data to compute the correspondences. Thus, we test our model with nonrigid shapes in a variety of poses for non-rigid correspondence experiments. We use 3D objects exhibiting two different poses, such as gorillas, dogs, lions, centaurs, among others. Figures 4.3 to 4.7 show the predicted correspondences between two shapes. Here, shape regions exhibiting the same color are considered to match (i.e., blue color on the left knee for the gorilla shapes). Besides, the zoomed areas relate more fine correspondences (i.e., green and blue colors for head in the gorilla shapes). Also, figure 4.4 illustrates the case where some regions (cheeks and nose of the Dog) are matched wrongly. Particularly, latent descriptors for this part of the object are assigned to different clusters. Hence, the model seeks to relate this area for a specific view (i.e., Dog figure in the left colored in goldenrod).

Also, the results show that our model establishes meaningful relations between objects (i.e., see figures 4.3 to 4.5, for the Gorillas, dogs, centaurs, and cat shapes respectively) with Rand indexes above 0.6. This gives us a quantitative evaluation of the shape matching where the analyzed shapes can be matched accurately. Moreover, we show additional experiments for non-rigid 3D objects in appendix C.4.1.

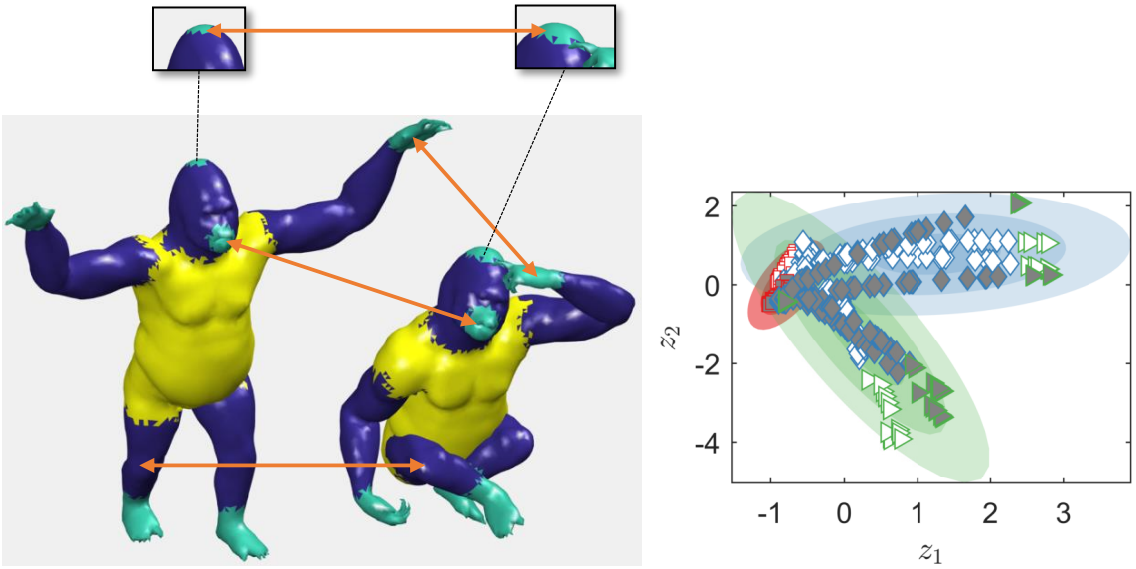


Figure 4.3: Experimental results for the TOSCA dataset. Current mixture parameters (input space), along with the latent positions ($K = 2$) for two shapes exhibiting different poses. Average Rand Index 0.667. Orange arrows describe candidate matches between shapes. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

In addition, table 4.3 shows the clustering performance of the proposed method against the state-of-the-art approaches for unsupervised object matching. The results show that our approach ($K = 2$) efficiently establishes meaningful relations by exhibiting better average Rand indexes concerning the compared methods (i.e., 0.69 for the MV-WMM).

Table 4.3: Comparison of the unsupervised clustering performance on the TOSCA dataset. Average Rand index and standard deviation are reported for the state-of-the-art methods for unsupervised learning.

	Approach				
Database	UCM	KM	KM-CKS	**NL-UCM**	**MV-WMM**
TOSCA	0.454 ± 0.140	$0.0.326 \pm 0.012$	0.355 ± 0.115	$\mathbf{0.604 \pm 0.080}$	$\mathbf{0.698 \pm 0.050}$

Furthermore, figure 4.8 shows a quantitative comparison of our shape correspondence method for the dataset in terms of the normalized geodesic error[3] (see appendix A.3

[3]The normalized geodesics error measures the distance between the ground-truth and predicted correspondence for a given pair of shapes.

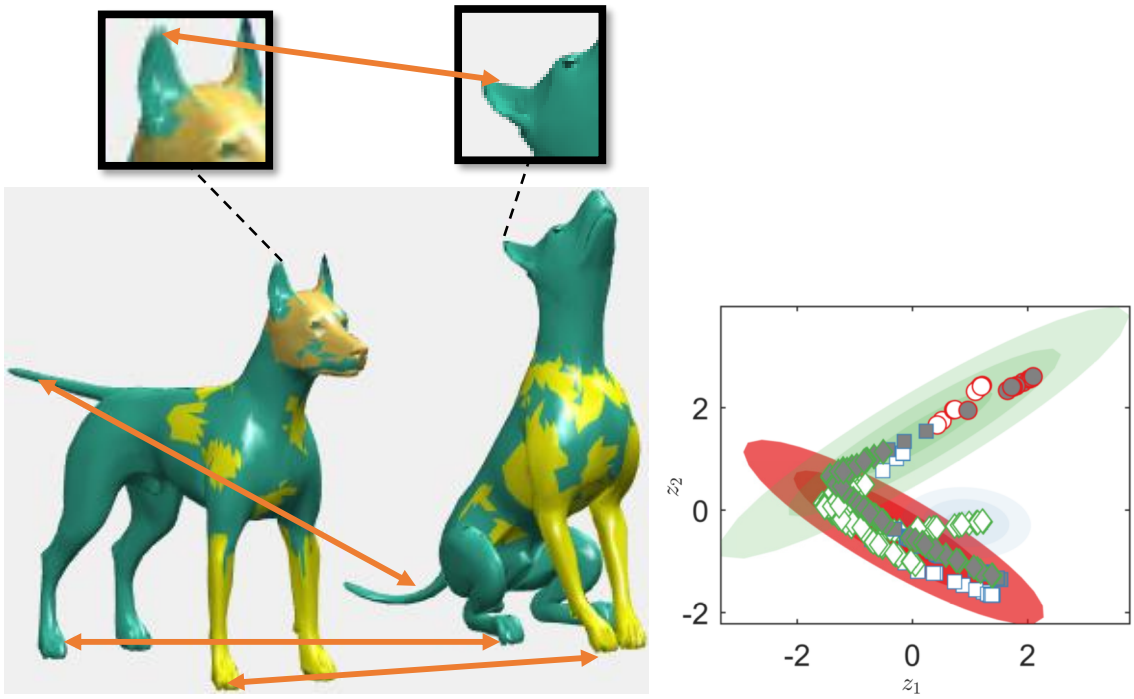


Figure 4.4: Experimental results for the TOSCA dataset. Current mixture parameters (input space), along with the latent positions ($K = 2$) for two shapes exhibiting different poses. Average Rand Index 0.6040. Orange arrows describe candidate matches between shapes. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

for details of the evaluation metrics). We measure the total percentage of matched points within a variable amount of normalized geodesic error according to the Princeton benchmark protocol (Kim et al., 2011). The results show that our approach performs an accurate matching since the 100% of the correspondences are reached for geodesic errors lesser than 0.15 which give us a relevant estimate of shape correspondences.

Finally, table 4.4 shows a comparison with relevant state-of-the-art approaches for this dataset. We compare our model against a) Unsupervised Learning of Dense Shape Correspondence (UL-DSC) (Halimi et al., 2019), b) product manifold filter (PMF) (Vestner et al., 2017), c) Blended intrinsic maps (BIM) (Kim et al., 2011), d) spectral generalized multidimensional scaling (SGMDS) (Aflalo et al., 2016), e) functional maps (FM) (Ovsjanikov et al., 2012), and f) Random Forest (RF) (Rodolà et al., 2014). We show our model to estimate accurate correspondences in comparison with the reported works, even for recent graph convolutional neural network approaches as in Halimi et al. (2019)(i.e., both approaches reach 90% of correspondences for geodesic errors lowers than 0.01).

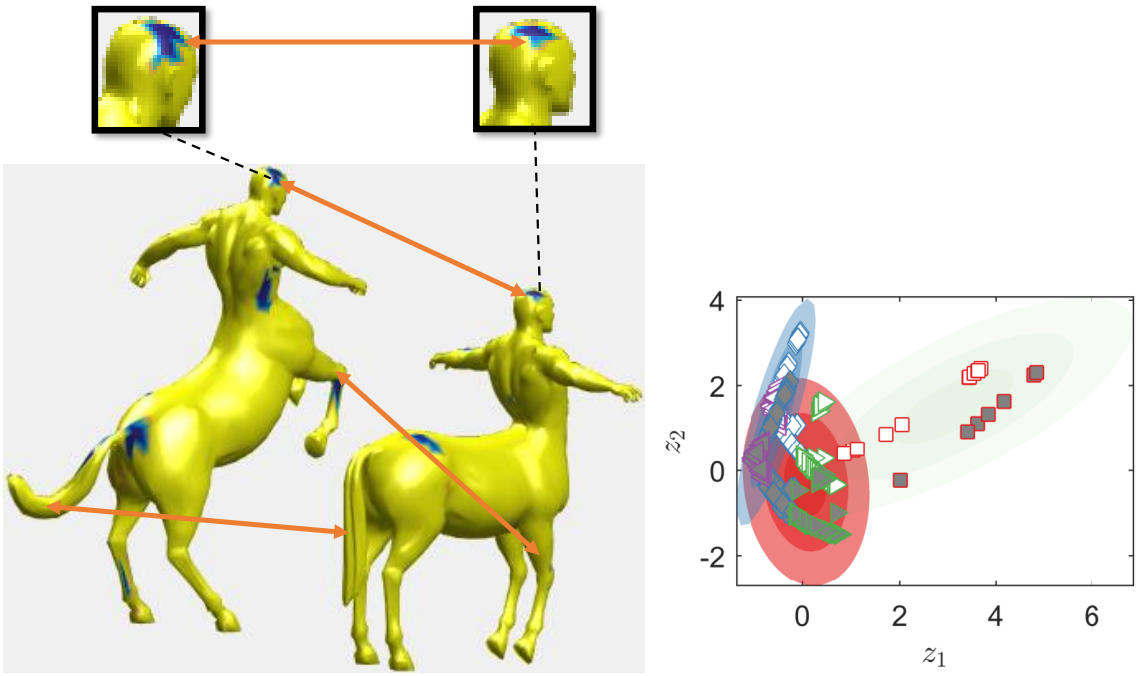


Figure 4.5: Experimental results for the TOSCA dataset. Current mixture parameters (input space), along with the latent positions ($K = 2$) for two shapes exhibiting different poses. Average Rand Index 0.6232. Orange arrows describe candidate matches between shapes. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

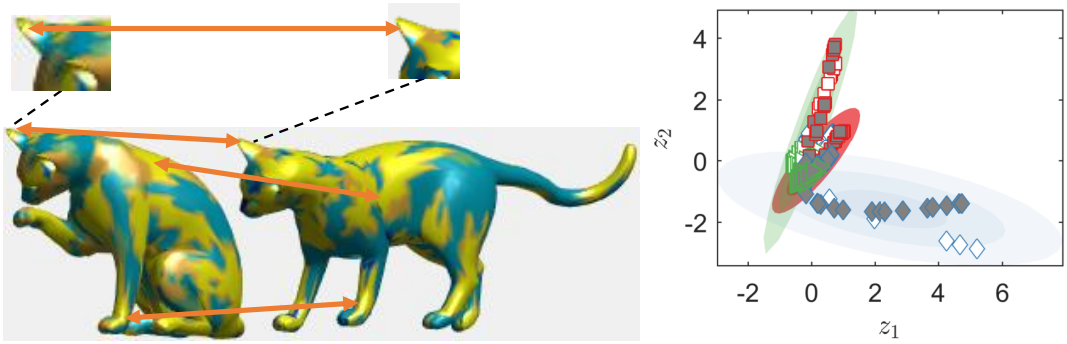


Figure 4.6: Experimental results for the TOSCA dataset. Current mixture parameters (input space), along with the latent positions ($K = 2$) for two shapes exhibiting different poses. Average Rand Index 0.8016. Orange arrows describe candidate matches between shapes. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

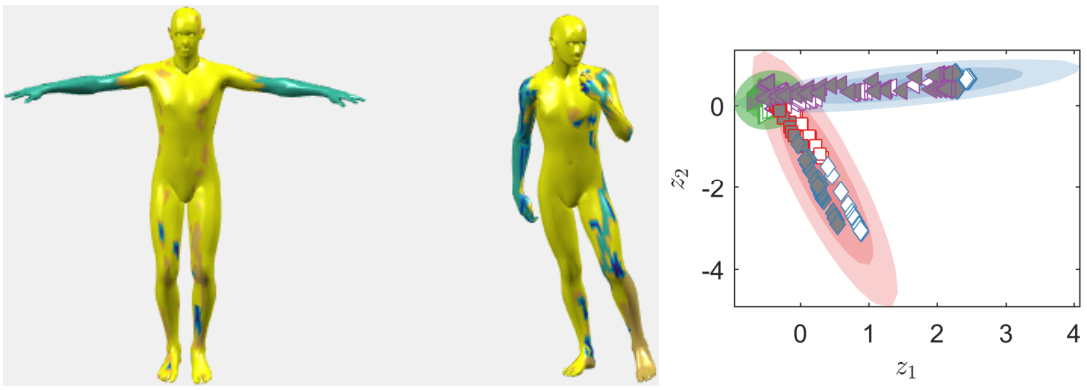


Figure 4.7: Experimental results for the TOSCA dataset. Current mixture parameters (input space), along with the latent positions ($K = 2$) for two shapes exhibiting different poses. Average Rand Index 0.7049. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

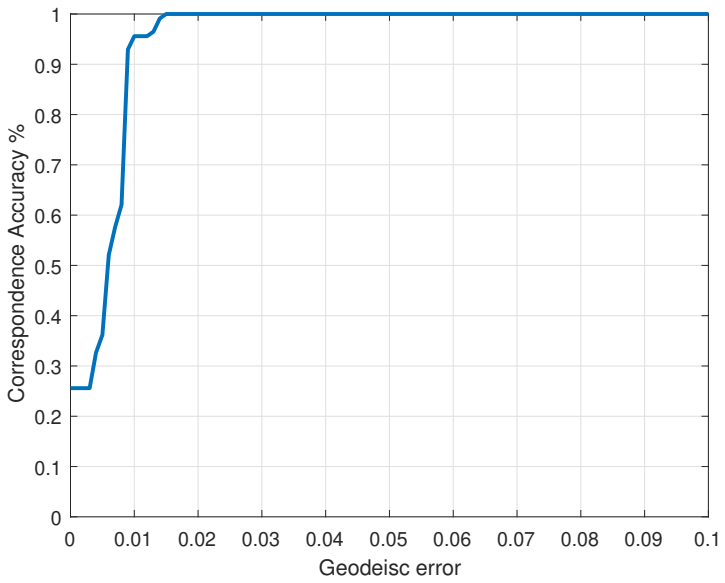


Figure 4.8: Performance of the proposed correspondence method on the TOSCA nonrigid dataset. The performances were evaluated using the Princeton benchmark (Kim et al., 2011). Each curve is averaged over all shapes in all classes. We plot a cumulative curve showing the percentage of matches that are at most geodesically distant from the ground-truth correspondence on the reference shape (i.e., one of the shapes is set as the reference one).

Table 4.4: Comparison of the matching performances on the TOSCA dataset. We report the geodesic error for which 90% of correspondences are reached, which give us an accurate estimate of the matching performance. Also, we report the standard deviation (std), for which the percentage of correspondences are reached.

Approach	Geodesic error	std
UL-DSC (Halimi et al., 2019)	0.010	n/a
PMF (Vestner et al., 2017)	0.030	n/a
BIM (Kim et al., 2011)	0.060	n/a
SGMDS (Aflalo et al., 2016)	0.040	n/a
FM (Ovsjanikov et al., 2012)	0.065	n/a
RF (Rodolà et al., 2014)	0.045	n/a
ours	**0.010**	0.002

4.5.4 Deformable 3D Shapes with Topological Noise

In this section, we test our model with the KIDS dataset (Lähner et al., 2016). This data consists of a collection of 3D shapes undergoing within-class deformations that include topological noise. This noise simulate coalescence of spatially close surface regions, a scenario that frequently occurs when dealing with real data under sub-optimal acquisition conditions (i.e., surface models for medical image analysis). The dataset is based on the fat kid from the KIDS dataset with additional poses. Figures 4.9 to 4.11 show the predicted correspondences between two different shapes exhibiting different poses with topological noise. The results show that even when the pair of shapes are contaminated with this type of noise, the model behaves appropriately as a consequence of the probabilistic nonlinear latent representation. Besides, average Rand indexes for these experiments remains above 0.65, which means that our approach can efficiently establish shape correspondences even when the shape structure is not well defined.

Finally, figure 4.12 shows a quantitative comparison of our shape correspondence method for shapes with topological noise in terms of the normalized geodesic error. The results show that our approach performs an acceptable matching since the 80% of the correspondences are reached for geodesic errors lesser than 0.2 which give us a relevant estimate of shape correspondences.

4.5.5 Partial matching

In this section, we aim to test our model in a particularly challenging and widely studied dataset for partial matching. Here, we examine the scenario of partial correspondence,

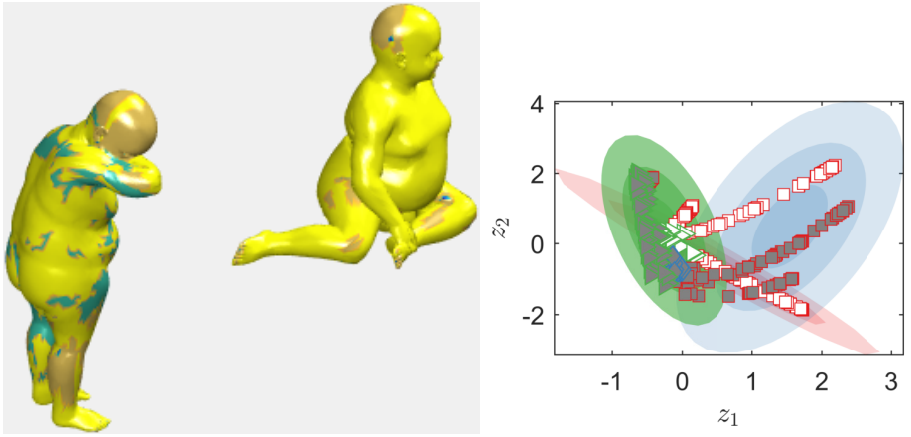


Figure 4.9: Experimental results for the KIDS dataset. Current mixture parameters (input space), along with the latent positions ($K = 2$) for two shapes exhibiting different poses. Average Rand Index 0.7432. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

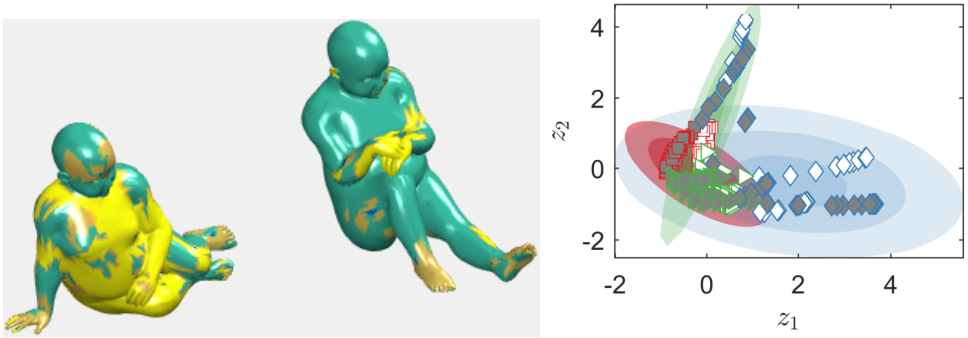


Figure 4.10: Experimental results for the KIDS dataset. Current mixture parameters (input space), along with the latent positions ($K = 2$) for two shapes exhibiting different poses. Average Rand Index 0.7179. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

where one shape exhibits a portion of the reference 3D object and has to match it with a deformable version.

Partial correspondence problems arise in numerous applications that involve real data acquisition by 3D sensors, which leads to missing parts occlusions and partial views (i.e., MRI acquisition problems in medical image analysis). Thus, we use the SHREC'16 benchmark to evaluate the performance of our method for establishing correspondences between a full shape and its deformed versions (Cosmo et al., 2016).

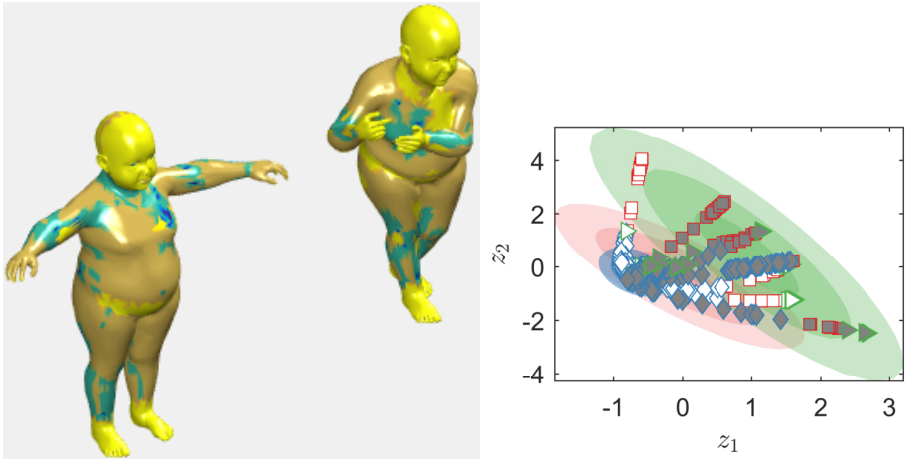


Figure 4.11: Experimental results for the KIDS dataset. Current mixture parameters (input space), along with the latent positions ($K = 2$) for two shapes exhibiting different poses. Average Rand Index 0.7288. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

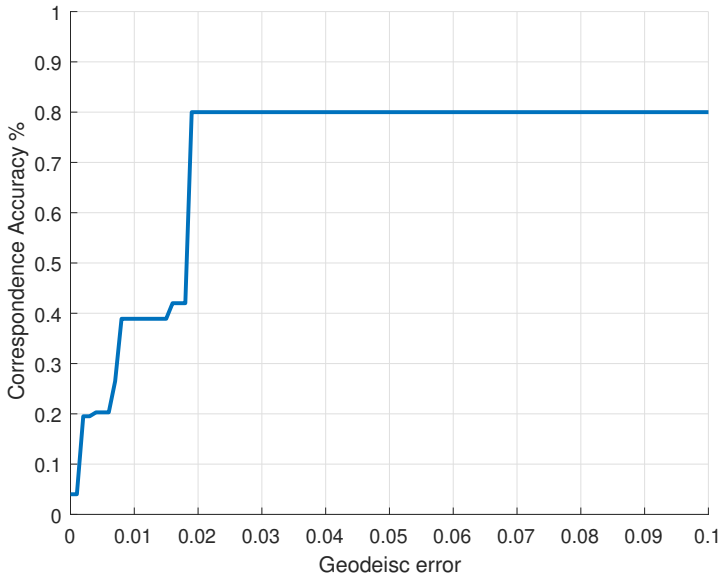


Figure 4.12: Performance of the proposed method on the KIDS dataset. The performances were evaluated using the Princeton benchmark (as in the previous section). Each curve is averaged over all shapes in all classes. We again plot a cumulative curve showing the percentage of matches that are at most geodesically distant from the ground-truth correspondence on the reference shape.

Shapes undergoing a single cut

In this scenario, we analyze the shapes belonging to the cuts dataset. These objects have a whole part of the shape missing in correspondence of a clean-cut over the surface (i.e.,

shape that is represented by half of the entire object). Figures 4.13 to 4.16 show the experimental results for the partial matching scenario. The results show that by exploiting the nonlinear probabilistic latent space, our model is able to establishes meaningful relations even when only a part of the 3D shape is observed. For instance, figure 4.16 shows that different forms of partiality can be tackled if adequately represented latent space (i.e., corresponding regions are shown in similar color).

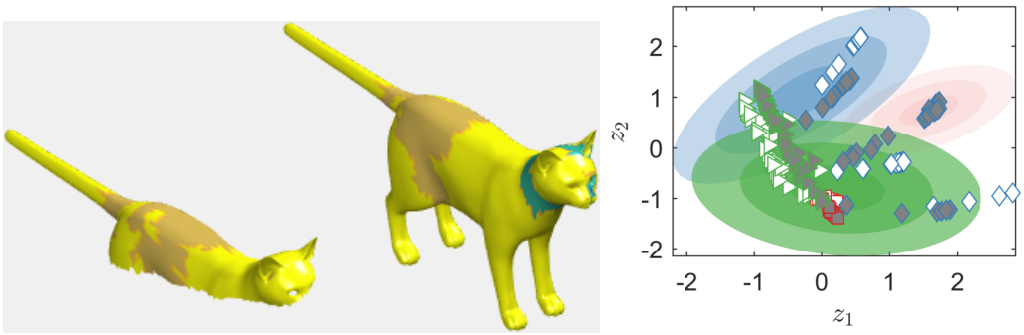


Figure 4.13: Experimental results for the partial matching SHREC'16 dataset. Current mixture parameters, along with the latent positions ($K = 2$) for two shapes exhibiting part-to-whole matching. Average Rand Index 0.6921. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

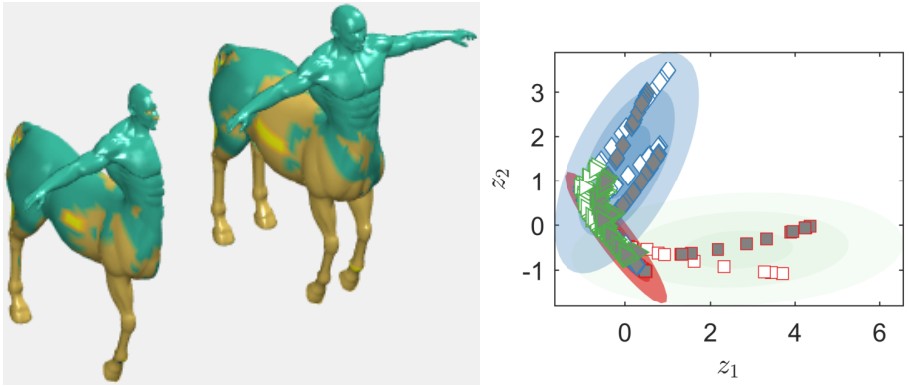


Figure 4.14: Experimental results for the partial matching SHREC'16 dataset. Current mixture parameters, along with the latent positions ($K = 2$) for two shapes exhibiting part-to-whole matching. Average Rand Index 0.6838. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

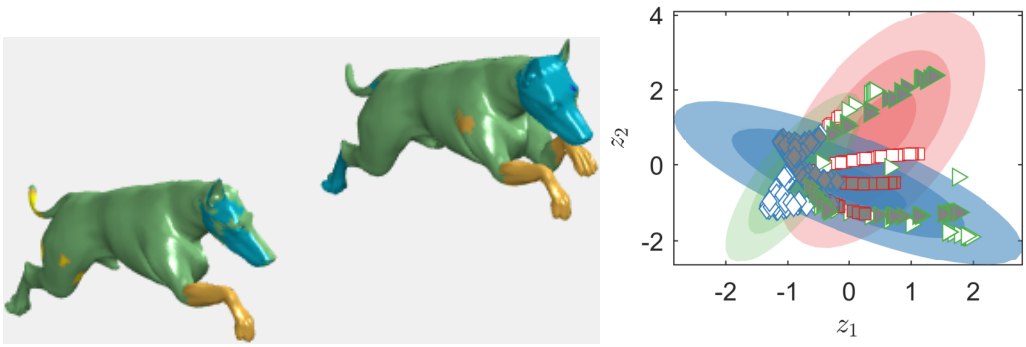


Figure 4.15: Experimental results for the partial matching SHREC'16 dataset. Current mixture parameters, along with the latent positions ($K = 2$) for two shapes exhibiting part-to-whole matching. Average Rand Index 0.7744. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

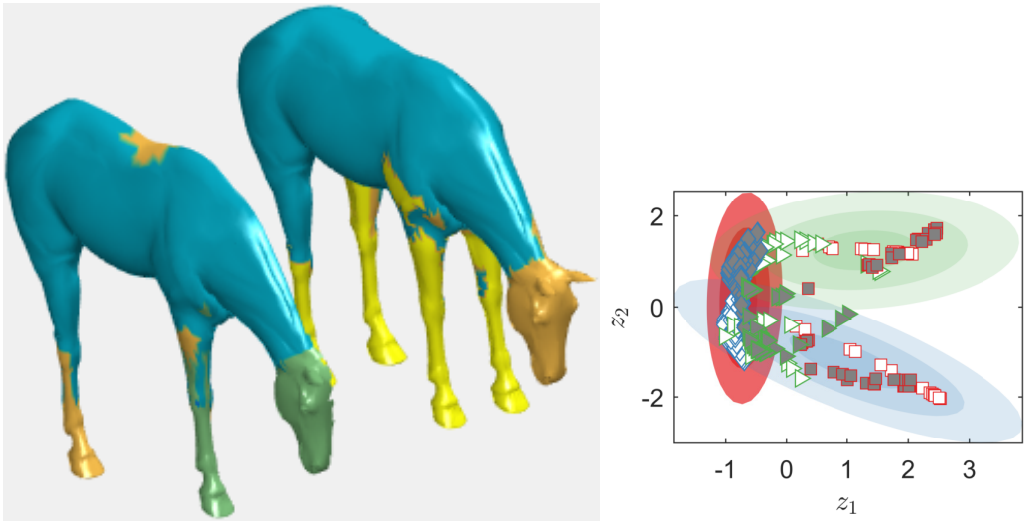


Figure 4.16: Experimental results for the partial matching SHREC'16 dataset. Current mixture parameters, along with the latent positions ($K = 2$) for two shapes exhibiting part-to-whole matching. Average Rand Index 0.7931. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

Shapes with irregular holes

Another interesting dataset that can evidence the performance of our method in shape correspondence problems is the holes dataset. The holes dataset a more challenging set because it contains shapes whose surface has been eroded starting from some random seeds over the surface, causing them to have holes and irregular cuts. Thus, 3D objects

inside each dataset present different amounts of missing surface, ranging approximately from 10% to 60% of the missing area. Figures 4.17 to 4.20 show the correspondence performance in which we can notice that the matching process is performed with average Rand indexes from 0.56 to 0.69. Thus, our model still establishes meaningful relations when irregular holes are presented in the analyzed shapes. As a consequence, our model can handle shape occlusions with acceptable performances (i.e., average Rand indexes above 0.56).

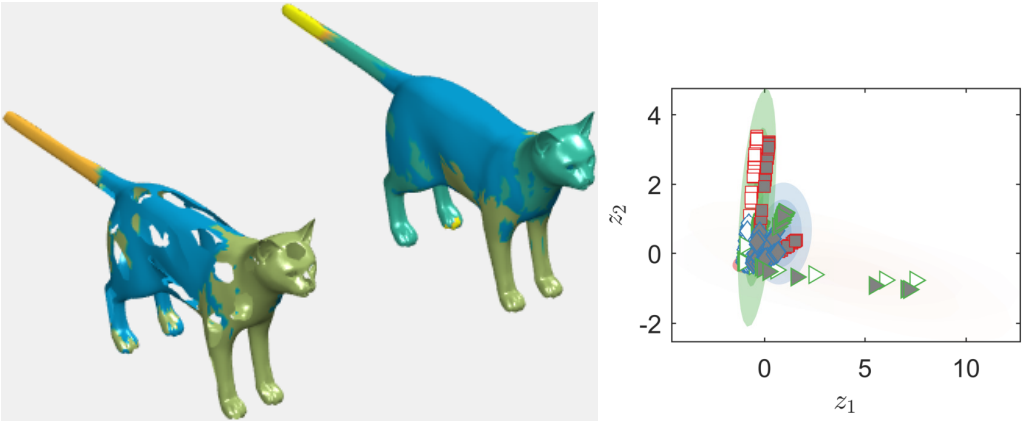


Figure 4.17: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions ($K = 2$) for two shapes exhibiting part-to-whole matching. Average Rand Index 0.6568. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

Finally, figure 4.21 shows a quantitative comparison of our shape correspondence method for both cuts and holes in terms of the normalized geodesic error. The results show that our approach performs an accurate matching since the 90% of the correspondences are reached for geodesic errors lesser than 0.2 which outperforms relevant shape correspondences approaches in the state-of-the-art (see (Cosmo et al., 2016)).

Finally, table 4.5 shows a qualitative evaluation of the matching performance by our approach in comparison with the SHREC'16 framework (Lähner et al., 2016). Average percentage of matches are reported in Table 4.5. Here, we compare against Partial functional maps (PFM) (Rodolí et al., 2017), Random forest (RF) (Rodolà et al., 2014), Scale-invariant isometric matching (IM) (Sahillioglu and Yemez, 2012), Game-theoretic matching (GT) (Torsello, 2012), and anisotropic convolutional neural networks (ACNN) (Boscaini et al., 2016). The results show that our approach performs partial matching tasks accurately, in comparison with state-of-the-art approaches (i.e., Average percentage

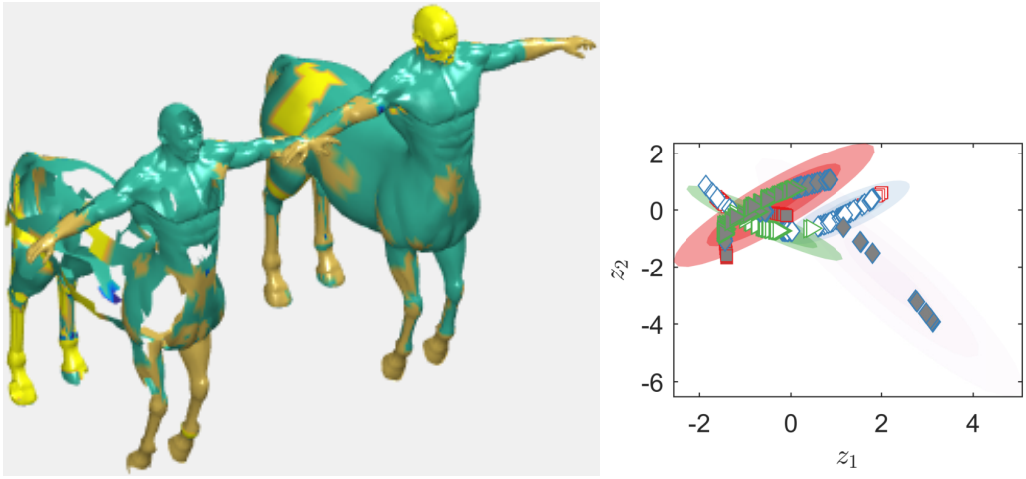


Figure 4.18: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions ($K = 2$) for two shapes exhibiting part-to-whole matching. Average Rand Index 0.6849. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

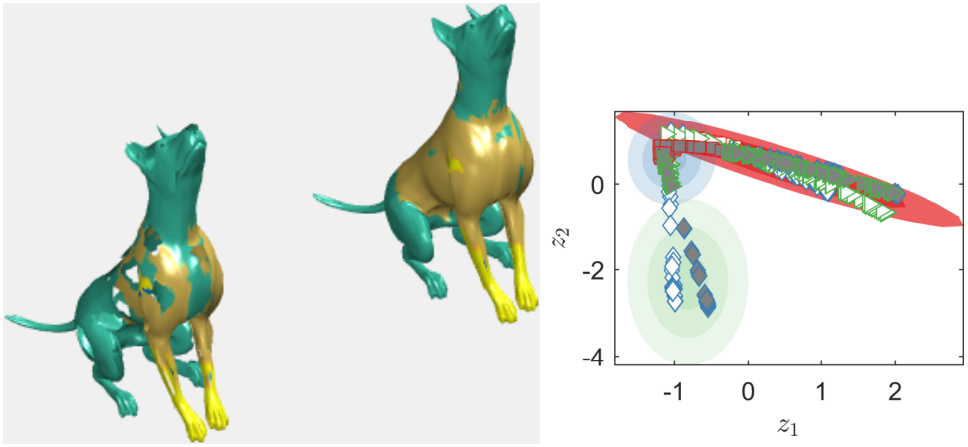


Figure 4.19: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions ($K = 2$) for two shapes exhibiting part-to-whole matching. Average Rand Index 0.5991. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

of matches of 95% and 90% for both datasets).

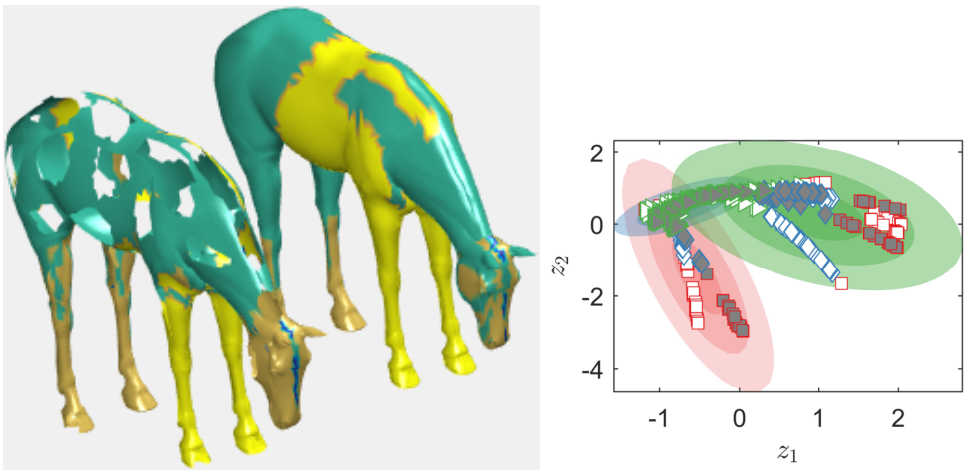


Figure 4.20: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions ($K = 2$) for two shapes exhibiting part-to-whole matching. Average Rand Index 0.6569. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

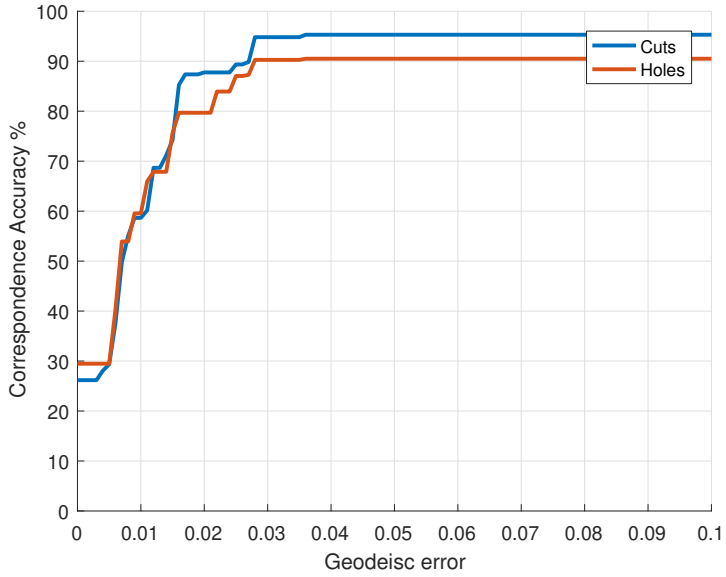


Figure 4.21: Performance of the proposed correspondence method on the SHREC'16 - Partial dataset (cut and holes shapes were tested to obtain the accuracy). The performances were evaluated using the Princeton benchmark. Each curve is averaged over all shapes in all classes. We plot a cumulative curve showing the percentage of matches that are at most geodesically distant from the ground-truth correspondence on the reference shape for the cuts and holes datasets.

Table 4.5: Average percentage of matches obtained by each method on the two datasets (Holes and Cuts datasets in SHREC'16).

	Approach					
Database	PFM	RF	IM	GT	ACNN	**MV-WMM**
cuts	89.6	87.4	61.3	51.0	**96.55**	95.3
holes	83.2	45.1	78.2	76.4	87.6	**90.5**

4.5.6 Neurodegenerative brain dataset

With the aim to model shape variability among time in Brain structures (see section 4.4 for details), 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 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.

To establish groupwise correspondences between brain structures, we used SI-HKS as shape descriptors (Bronstein and Kokkinos, 2010). We evaluate our model by using three relevant brain structures in the Parkinson's and Alzheimer's diseases such as the ventricle, thalamus, and putamen. Figures 4.22, 4.23 and 4.24 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). We set the regularization coefficient λ to be 0.01 in our experiments. Thus, at this value, the regularization parameter reduces overfitting, so that latent positions will maintain close within views. Therefore, shape correspondences among views will change softly (i.e., by avoid creating new shape relations abruptly). In our experiments, we found that varying the regularization parameter influences the correspondence estimate. For instance, increasing lambda results in less overfitting (i.e., deriving in meaningful shape relations) but also greater bias, which causes that correspondences for minor brain shape variations, will be skipped (i.e., lower Rand indexes).

From the results, it can be noticed that even when the brain volumetry of a given

shape (i.e., see Ventricle results in figure 4.22) has lost part of their mass as consequence of the neurodegenerative process, our model is capable of establishing relevant correspondences between brain structures. This result evidences the benefit of modeling shape variability among time (i.e., dynamic analysis) with multiview constraint (see section 4.4). As a result, our model can efficiently capture neurodevelopmental changes while preserving topological changes within the same brain structure.

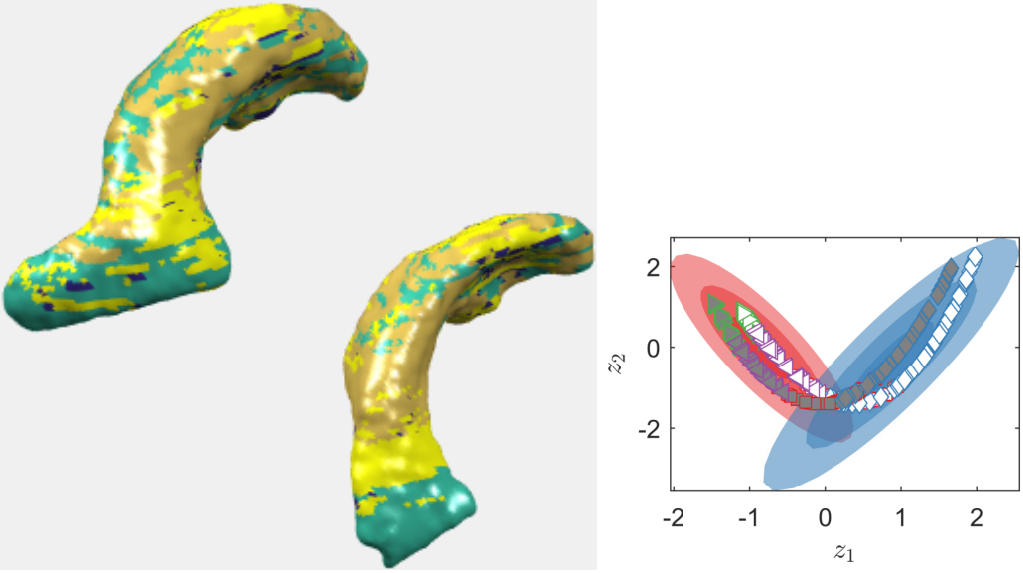


Figure 4.22: Experimental results for the brain structures dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the ventricle. Average Rand Index 0.6071. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

Finally, figure 4.25 shows a quantitative evaluation of our shape correspondence method in terms of the normalized geodesic error. The results show that our approach reaches the 85% of the correspondences for geodesic errors lesser than 0.03, which is a relevant indicator for shape matching approaches in the state-of-the-art. That is, we can perform an appropriate matching task for neuroimaging problems, if at least 80% of the correspondences are reached for geodesic errors below 0.1 (Lähner et al., 2016).

4.5.7 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 4.26 shows groupwise correspondences between

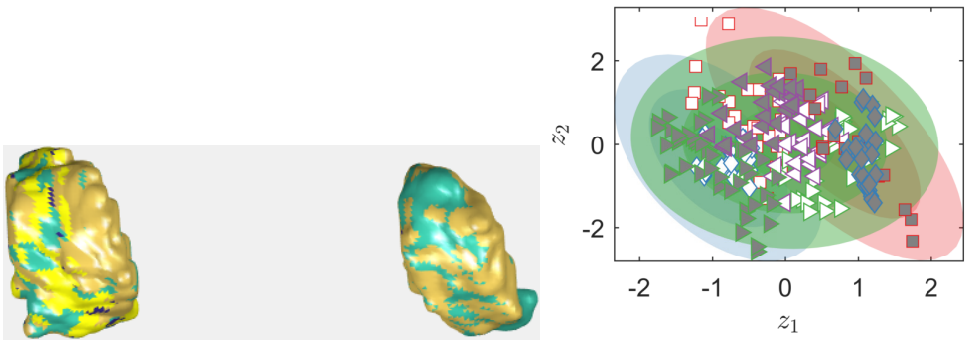


Figure 4.23: Experimental results for the brain structures dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the thalamus. Average Rand Index 0.6467. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

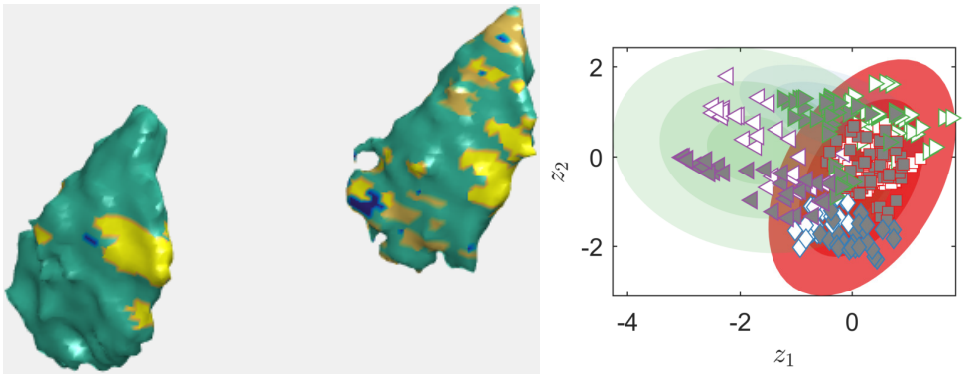


Figure 4.24: Experimental results for the brain structures dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the putamen. Average Rand Index 0.6467. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

the neonatal brain volumes. In this experiment, we set the regularization coefficient of λ to be 0.25. Thus, at this value, the regularization parameter will allow creating new shape relations properly. As a result, new brain regions for the second time-point (i.e., 31 weeks) will be created as part of the neurodevelopmental. 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 (see zoomed areas in black plates from figure 4.26). Preliminary results can lead to a new types of scores to predict neurodevelopmental outcome that uses unsupervised multiview learning to establish

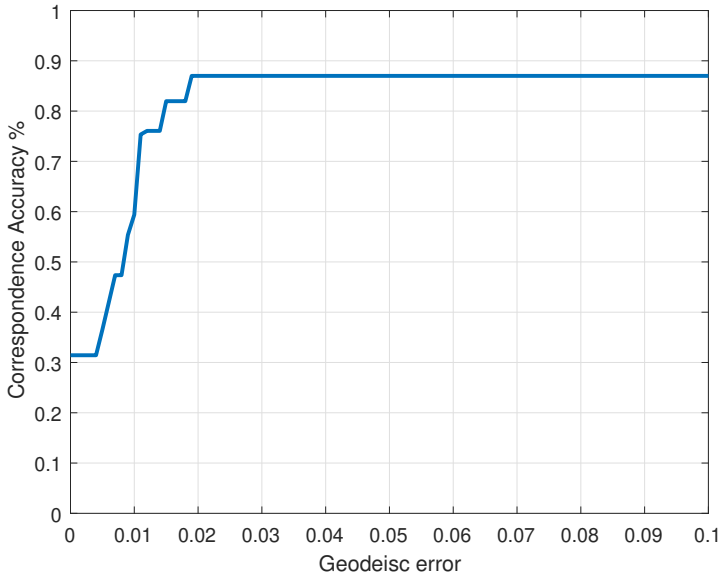


Figure 4.25: Performance of the proposed correspondence method on the Neurodegenerative dataset. The performances were evaluated using the Princeton benchmark. Each curve is averaged over all brain sequences in all classes (i.e., brain structure analysis). We again plot a cumulative curve showing the percentage of matches that are at most geodesically distant from the ground-truth correspondence on the reference shape.

meaningful relationships between a set of MRI scans (Weeke et al., 2018).

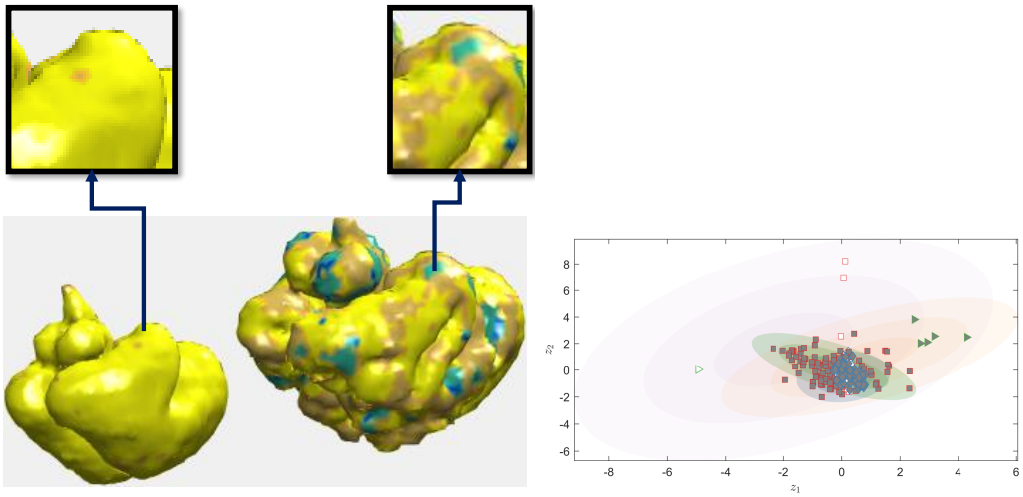


Figure 4.26: Experimental results for the brain Neurodevelopmental dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the brain volume. Average Rand Index 0.6948. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

Figure 4.27 shows an example of the matching process using the established correspondences. The results show that meaningful relations are established as part of the unsupervised matching on which the latent space is exploited to find groupwise correspondences. In addition, the results show that the brain regions assigned to the same group show similar neurodevelopment (i.e. groups that are candidates to match).

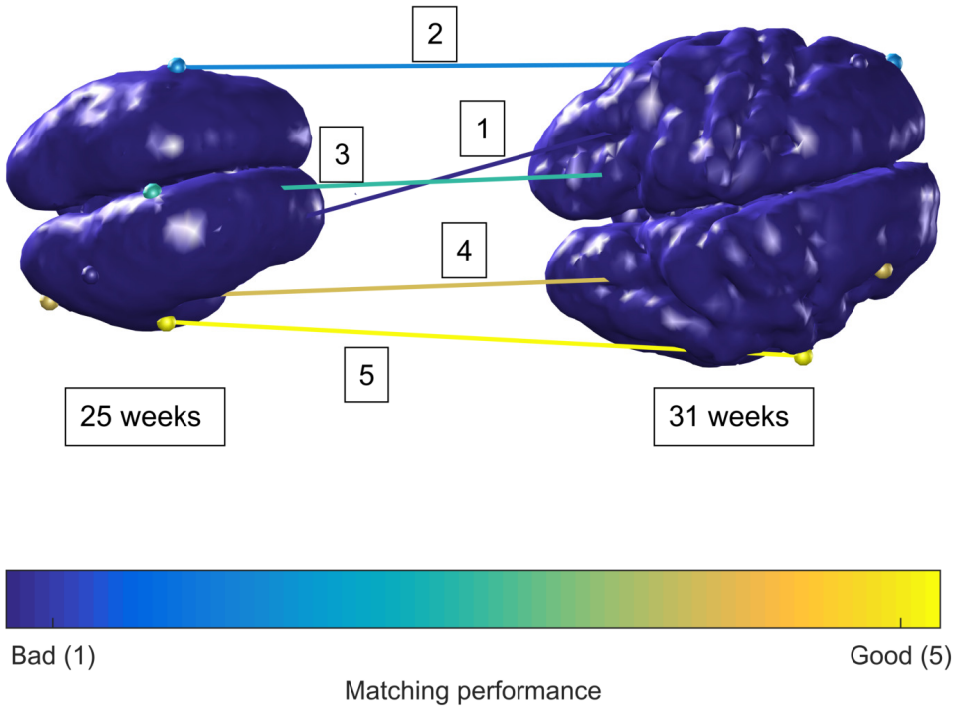


Figure 4.27: Matching result for the brain Neurodevelopmental dataset. The figure shows the established correspondences, starting from good to bad matches (yellow to dark blue in the colormap). For instance, matching number 5 exhibits a good correspondence estimate on which the neurodevelopmental corresponds to a plausible match (right temporal lobe). Also, matching number 1 shows a weak estimate on which the brain regions are related wrongly (right occipital lobe in the 25 weeks brain to the left frontal lobe in the 31 weeks brain).

4.6 Conclusions

This chapter presented a multiview learning approach to model nonlinear groupwise correspondences of complex shapes by using a multiview warped mixture model. In particular, since each shape descriptor has its correspondence in the latent space, and it is generated by warping the latent correspondence via nonlinear mapping functions,

our model is able to establish meaningful relations between different set of 3D shapes.

The approach is very effective in modeling ambiguous views such as those that present irregular cuts or topological noise. Thus, by incorporating multiview constraints, our model is able to match both intra and inter shapes efficiently. Besides, the resulting normalized geodesic errors evidence that by using multiview warped mixture representations of the input data, a given matching process can be performed accurately.

A current limitation of the model is that the model complexity grows proportional to the shape dimensionality. Thus, as future work, we plan to perform stochastic variational inference to allow modeling large datasets (i.e., high-resolution scans) (De Sousa Ribeiro et al., 2019).

Chapter 5

Conclusions and Future Work

This thesis considered probabilistic nonlinear latent variable models for shape correspondence analysis. The objective of learning the shape structure in an unsupervised manner is to learn the object structure without any similarity measure and to exploit the latent representation that leads to meaningful matches between shapes. However, unsupervised clustering methods for multiview learning are associated with many degrees of freedom and, hence, are challenging to learn. Therefore, this thesis was concerned with developing the mathematical framework and algorithms that allow for relevant latent representation on which an unsupervised clustering task was performed.

This chapter summarizes the contributions and research work done in the thesis, besides some future research lines are presented.

5.1 Conclusions

- Chapter 2 provided a unifying view of existing approximation to the correspondence problem. A special focus is given for a set of correspondence approaches in the context of shape analysis.

- Chapter 3 presented a nonlinear latent variable model for unsupervised groupwise correspondences which incorporates observations from several views. We provided a method for shape correspondence analysis based on nonlinear unsupervised clustering of groupwise 3D shape descriptors. The clustering process was carried out by a nonlinear probabilistic latent variable model, in which we used random Fourier features of the input data observations (Rahimi and Recht, 2007). In other words, we extended the many-to-many object matching proposed by Iwata et al. (2016)

using Hilbert space embeddings of the input data (Rahimi and Recht, 2007).

- Chapter 4 presented a Multiview Bayesian clustering model in which the correspondences have nonparametric shapes, called the multiview infinite warped mixture model. The density manifolds learned by this model follow the contours of the data density, and have interpretable, parametric forms in the latent space. The marginal likelihood lets us infer the effective dimension and shape of each cluster separately, as well as the number of groups of correspondences. Besides, by using a multiview GPLVM, the model can handle more relevant latent representations which derived in meaningful correspondences.

- In general, the developed models can establish relevant correspondences for neuroimaging problems by exploiting the latent representation captured through the probabilistic multiview approaches. Besides, in chapter 4, it was shown that even when large shape variations are modeled, probabilistic multiview models are still able to learn the shape structure between views (i.e., temporal shape variability in the neurodevelopmental experiments). Besides, by constraining the multiview model the dynamic analysis of brain volumes becomes more accurate. Thus, meaningful relations between brain shapes are established as part of the neurodevelopmental.

5.2 Future Work

Some interesting paths for future work involve approaches to solving current limitations of the presented methods and further extensions of the developed methodologies. In particular, a common limitation of the unsupervised shape correspondence methods is difficulty in scaling them up. This is a key research line because of the dimensionalities of the 3D shapes are a challenging topic.

The main ideas for the probabilistic correspondence problem are summarized

NL-UCLM: Concerning the nonlinear unsupervised clustering matching, some possible future research lines could be defined as

1. We can relax the assumption that the observations are linear with respect to their latent vectors by using nonlinear matrix factorization techniques (see Lawrence and Urtasun (2009)). From this work, we point out that by

marginalizing out the mapping matrix $\mathbf{W}$, we can set our model as a Bayesian multi-output regression model.

Our model can be formalized as

$$p\left(\mathbf{x}_{vn}|\mathbf{Z},\mathbf{W},\boldsymbol{\theta}\right)=\sum_{j=1}^{\infty}\theta_j\prod_{m=1}^{M_d}\mathcal{N}\left(\mathbf{x}_{dnm}|f_{dm}\left(\mathbf{z}_j\right),\alpha^{-1}\right), \tag{5.1}$$

where f_{dm} can handled from two perspectives

(a) Set the same kernel matrix for all views

$$f_{dm}\sim\mathcal{GP}\left(\mu_{dm}\left(\mathbf{z}_j\right),k\left(\mathbf{z}_j,\mathbf{z}'_j\right)\right) \tag{5.2}$$

(b) One covariance matrix for each domain

$$f_{dm}\sim\mathcal{GP}\left(\mu_{dm}\left(\mathbf{z}_j\right),k_d\left(\mathbf{z}_j,\mathbf{z}'_j\right)\right) \tag{5.3}$$

2. As in the in the linear model of coregionalization (LMC), the outputs are expressed as linear combinations of independent random functions (Álvarez et al., 2012). Consider a set of D outputs $\{\mathbf{f}_d(\mathbf{z}_j)\}_{d=1}^{D}$ with $\mathbf{f}_d(\mathbf{z}_j)\in\mathbb{R}^{M_d}$. By adopting this framework, our model can be formulated as

$$p\left(\mathbf{x}_{vn}|\mathbf{Z},\mathbf{W},\boldsymbol{\theta}\right)=\sum_{j=1}^{\infty}\theta_j\mathcal{N}\left(\mathbf{x}_{vn}|\mathbf{f}_d(\mathbf{z}_j),\alpha^{-1}\mathbf{I}\right), \tag{5.4}$$

where each component can be expressed as:

(a) By using the linear model of coregionalization

$$\mathbf{f}_d(\mathbf{z}_j)\sim\mathcal{GP}\left(\mathbf{0},\sum_{q=1}^{Q}\mathbf{B}_q k_d\left(\mathbf{z}_j,\mathbf{z}'_j\right)\right), \tag{5.5}$$

where $\mathbf{B}_q=\mathbf{L}_q^{\top}\mathbf{L}_q\in\mathbb{R}^{M_d\times M_d}$ is the coregionalization matrix (computed from the Cholesky decomposition)

(b) By using simplified version of the LMC, known as the intrinsic coregionalization model (ICM) (see Álvarez et al. (2012)), assumes that the elements of the coregionalization matrix $\mathbf{B}_q$ can be written as a scaled

version of the elements b_q, which do not depend on the particular output functions $f_d(\mathbf{z}_j)$.

$$\mathbf{f}_d(\mathbf{z}_j) \sim \mathcal{GP}\left(\mathbf{0}, \mathbf{B}k\left(\mathbf{z}_j, \mathbf{z}'_j\right)\right), \tag{5.6}$$

MV-WMM: Concerning the multiview warped mixture model, some possible future research lines could be defined as

- First, by analyzing other prior on the latent densities, our mode can exploit underlying complex representations of geometric shapes (i.e., structure modeling in medical image analysis when a considerable part of the view is missing). Hence, we can adopt hierarchical clustering to efficiently model parts of the shape within a hierarchical framework (Johnson et al., 2016).
- Finally, to exploit the model uncertainty, one final future line would be based on building a Deep Gaussian Process that can handle unaligned multiview data to establish robust shape correspondences (Damianou and Lawrence, 2013).

Appendix A

Performance metrics

A.1 Rand Index

The Rand index is a measure of the similarity between two data clusterings (Hubert and Arabie, 1985). Given a set of n elements $S = \{o_1, \ldots, o_n\}$ and two partitions of S to compare, $X = \{X_1, \ldots, X_r\}$, a partition of S into r subsets, and $Y = \{Y_1, \ldots, Y_s\}$, a partition of S into s subsets, define the following:

- a, the number of pairs of elements in S that are in the same subset in X and in the same subset in Y.
- b, the number of pairs of elements in S that are in different subsets in X and in different subsets in Y.
- c, the number of pairs of elements in S that are in the same subset in X and in different subsets in Y. d, the number of pairs of elements in S that are in different subsets in X and in the same subset in Y.

The Rand index is defined as:

$$RI = \frac{a+b}{a+b+c+d} \tag{A.1}$$

Intuitively, $a + b$ can be considered as the number of agreements between X and Y and $c + d$ as the number of disagreements between X and Y.

A.2 Adjusted Rand Index

For the evaluation of the unsupervised clustering task, we used the adjusted Rand index (Hubert and Arabie, 1985), which quantifies the similarity between inferred clusters and true clusters. It takes a value from -1.0 to 1.0 represents random clustering. A higher value indicates better clustering performance. The adjusted Rand index becomes high when object pairs that belong to one true cluster are assigned to one inferred cluster simultaneously, and when object pairs that belong to different true clusters are assigned to different inferred clusters.

The adjusted Rand index is the corrected-for-chance version of the Rand index defined as

$$\text{AdjustedIndex} = \frac{\text{Index-ExpectedIndex}}{\text{MaxIndex-Expected Index}}. \tag{A.2}$$

A.3 Normalized geodesic error

The evaluation of the correspondence quality was based on the normalized geodesic error of the given matching process from the ground-truth sub-vertex correspondence (i.e., partial to the whole shape).

In general, the following performance indices were considered accordingly to the Princeton correspondence benchmark (Kim et al., 2011):

- Total percentage of matched points within a variable amount of normalized geodesic error is computed for a given pair of 3D shapes.
- Mean geodesic error over all the intra and inter pair of 3D shapes (i.e., objects that change between views and those who are similar but comes from different sources).

A.3.1 Error measure

For the evaluation of the correspondence quality, we refer to the Princeton benchmark protocol (Kim et al., 2011) for point-wise maps. Let be the full model shape in a canonical pose $\mathcal{S}_{\mathcal{N}}$ (i.e., reference shape to compare with) and one of its corresponding partial version. Assume that a correspondence algorithm produces a pair of points $(x, y) \in \mathcal{S}_{\mathcal{N}} \times \mathcal{S}_{\mathcal{M}}$ observations, whereas the ground-truth correspondence is (x, y^*). Then, the inaccuracy of the correspondence is measured as

$$\epsilon(x) = \frac{d_{\mathcal{S}_\mathcal{M}(y,y^*)}}{\text{area}(\mathcal{S}_\mathcal{M})^{1/2}} \tag{A.3}$$

and has units of normalized length on $\mathcal{S}_\mathcal{M}$ (ideally, zero). Here $d_{\mathcal{S}_\mathcal{M}}$ is the geodesic distance on $\mathcal{S}_\mathcal{M}$.

Appendix B

Nonlinear Probabilistic Latent Variable Models

This appendix concerns the Nonlinear Unsupervised Clustering Matching framework developed in Chapter 3 and which is used as a backbone for the methodology developed in the rest of the chapters. The details of this appendix aim at providing a useful reference for the complete derivation of the method.

B.1 Likelihood for the nonlinear model

The joint probability of the feature set $\mathbf{\Phi}$, and the cluster assignments $\mathbf{S} = \left\{ \{s_{vn}\}_{n=1}^{N_V} \right\}_{v=1}^{V}$ are given by

$$p\left(\mathbf{\Phi}, \mathbf{S} | \boldsymbol{\mathcal{W}}, a, b, r, \gamma\right) = p\left(\mathbf{S}|\gamma\right) p\left(\mathbf{\Phi}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r\right). \tag{B.1}$$

By marginalizing out latent vectors $\mathbf{Z}$ and the precision parameter α, the second factor of (B.1) is computed by

$$
\begin{aligned}
p\left(\boldsymbol{\Phi}|\mathbf{S},\boldsymbol{\mathcal{W}},a,b,r\right) &= \int\int \prod_{v=1}^{V}\prod_{n=1}^{N_v} \mathcal{N}\left(\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)|\boldsymbol{B}_v\boldsymbol{\zeta}_{s_{vn}},\alpha^{-1}\mathbf{I}\right)\mathcal{G}\left(\alpha|a,b\right)\times \\
&\prod_{j=1}^{J}\mathcal{N}\left(\boldsymbol{\zeta}_j|\mathbf{0},\left(\alpha r\right)^{-1}\mathbf{I}\right)d\mathbf{Z}d\alpha \\
&= \int\int \prod_{v=1}^{V}\prod_{n=1}^{N_v}\left(\frac{\alpha}{2\pi}\right)^{L_v/2}\exp\left(-\frac{\alpha}{2}||\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)-\boldsymbol{B}_v\boldsymbol{\zeta}_{s_{vn}}||^2\right)\prod_{j=1}^{J}\left(\frac{\alpha r}{2\pi}\right)^{K/2} \\
&\times\exp\left(-\frac{\alpha r}{2}||\boldsymbol{\zeta}_j||^2\right)\frac{b^a\alpha^{a-1}}{\Gamma\left(a\right)}\exp\left(-b\alpha\right)d\mathbf{Z}d\alpha \\
&= \frac{b^a}{\Gamma\left(a\right)}\int\int\left(\frac{\alpha}{2\pi}\right)^{\sum_v L_vN_v/2}\exp\left(-\frac{\alpha}{2}\sum_{v=1}^{V}\sum_{n=1}^{N_v}||\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)-\boldsymbol{B}_v\boldsymbol{\zeta}_{s_{vn}}||^2\right) \\
&\times\left(\frac{\alpha r}{2\pi}\right)^{KJ/2}\exp\left(-\frac{\alpha r}{2}\sum_{j=1}^{J}||\boldsymbol{\zeta}_j||^2\right)\exp\left(-b\alpha\right)\alpha^{a-1}d\mathbf{Z}d\alpha
\end{aligned}
\tag{B.2}
$$

Solving for the first exponential term

$$
\begin{aligned}
\exp\left(-\frac{\alpha}{2}\sum_{v=1}^{V}\sum_{n=1}^{N_v}||\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)-\boldsymbol{B}_v\boldsymbol{\zeta}_{s_{vn}}||^2\right) &= \exp\Big(-\frac{\alpha}{2}\sum_{v=1}^{V}\sum_{n=1}^{N_v}\left[\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^\top\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\right. \\
&\left.-2\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)+\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\zeta}_{s_{vn}}\right]\Big)
\end{aligned}
\tag{B.3}
$$

The equation in (B.2) becomes

$$
\begin{aligned}
p\left(\boldsymbol{\Phi}|\mathbf{S},\boldsymbol{\mathcal{W}},a,b,r\right) &= \frac{b^a}{\Gamma\left(a\right)}\int\int\left(\frac{\alpha}{2\pi}\right)^{\sum_v L_vN_v/2}\exp\Big(-\frac{\alpha}{2}\sum_{v=1}^{V}\sum_{n=1}^{N_v}\left[\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^\top\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\right. \\
&\left.-2\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)+\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\zeta}_{s_{vn}}\right]\Big)\left(\frac{\alpha r}{2\pi}\right)^{KJ/2} \\
&\times\exp\left(-\frac{\alpha r}{2}\sum_{j=1}^{J}\boldsymbol{\zeta}_j^\top\boldsymbol{\zeta}_j\right)\exp\left(-b\alpha\right)\alpha^{a-1}d\mathbf{Z}d\alpha.
\end{aligned}
\tag{B.4}
$$

The exponential terms in (B.4) becomes

$$\exp\Bigg(-\frac{\alpha}{2}\sum_{v=1}^{V}\sum_{n=1}^{N_v}[\boldsymbol{\phi}(\mathbf{x}_{vn})^\top\boldsymbol{\phi}(\mathbf{x}_{vn})-2\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{\phi}(\mathbf{x}_{vn})+\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\zeta}_{s_{vn}}]$$
$$-\frac{\alpha r}{2}\sum_{j=1}^{J}\boldsymbol{\zeta}_j^\top\boldsymbol{\zeta}_j-b\alpha\Bigg)=\exp\Bigg(-\frac{\alpha}{2}\sum_{v=1}^{V}\sum_{n=1}^{N_v}[\boldsymbol{\phi}(\mathbf{x}_{vn})^\top\boldsymbol{\phi}(\mathbf{x}_{vn})-b\alpha]\Bigg)\times$$
$$\exp\Bigg(-\frac{\alpha}{2}\sum_{v=1}^{V}\sum_{n=1}^{N_v}[-2\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{\phi}(\mathbf{x}_{vn})+\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\zeta}_{s_{vn}}]-\frac{\alpha r}{2}\sum_{j=1}^{J}\boldsymbol{\zeta}_j^\top\boldsymbol{\zeta}_j\Bigg) \tag{B.5}$$

By analyzing the nth objects that has the cluster assignment j ($n : s_{vn} = j$), the second factor in (B.5) becomes

$$\exp\Bigg(-\frac{\alpha}{2}\sum_{v=1}^{V}\sum_{n=1}^{N_v}[-2\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{\phi}(\mathbf{x}_{vn})+\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\zeta}_{s_{vn}}]-\frac{\alpha r}{2}\sum_{j=1}^{J}\boldsymbol{\zeta}_j^\top\boldsymbol{\zeta}_j\Bigg)=$$
$$\exp\Bigg(-\frac{\alpha}{2}\sum_{v=1}^{V}\sum_{n:s_{vn}\neq j}[-2\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{\phi}(\mathbf{x}_{vn})+\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\zeta}_{s_{vn}}]\Bigg)$$
$$\times\underbrace{\exp\Bigg(-\frac{\alpha}{2}\sum_{v=1}^{V}\sum_{n:s_{vn}=j}[-2\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{\phi}(\mathbf{x}_{vn})+\boldsymbol{\zeta}_{s_{vn}}^\top\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\zeta}_{s_{vn}}]-\frac{\alpha r}{2}\sum_{j=1}^{J}\boldsymbol{\zeta}_j^\top\boldsymbol{\zeta}_j\Bigg)}_{C}$$
$$=\exp\Bigg(-\frac{\alpha}{2}\sum_{v=1}^{V}\sum_{j=1}^{J}N_{vj}[-2\boldsymbol{\zeta}_j^\top\boldsymbol{B}_v^\top\boldsymbol{\phi}(\mathbf{x}_{vn})+\boldsymbol{\zeta}_j^\top\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\zeta}_j]-\frac{\alpha r}{2}\sum_{j=1}^{J}\boldsymbol{\zeta}_j^\top\boldsymbol{\zeta}_j\Bigg)$$
$$=\exp\Bigg(-\frac{\alpha}{2}\sum_{v=1}^{V}\sum_{j=1}^{J}[-2\boldsymbol{\zeta}_j^\top\boldsymbol{B}_v^\top\sum_{n:s_{vn}=j}\boldsymbol{\phi}(\mathbf{x}_{vn})+\boldsymbol{\zeta}_j^\top N_{vj}\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\zeta}_j]-\frac{\alpha r}{2}\sum_{j=1}^{J}\boldsymbol{\zeta}_j^\top\boldsymbol{\zeta}_j\Bigg)$$
$$=\exp\Bigg(-\frac{\alpha}{2}\sum_{j=1}^{J}[-2\boldsymbol{\zeta}_j^\top\sum_{v=1}^{V}\boldsymbol{B}_v^\top\sum_{n:s_{vn}=j}\boldsymbol{\phi}(\mathbf{x}_{vn})+\boldsymbol{\zeta}_j^\top\sum_{v=1}^{V}N_{vj}\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\zeta}_j]-\frac{\alpha r}{2}\sum_{j=1}^{J}\boldsymbol{\zeta}_j^\top\boldsymbol{\zeta}_j\Bigg)$$
$$=\exp\Bigg(-\frac{\alpha}{2}\sum_{j=1}^{J}[-2\boldsymbol{\zeta}_j^\top\sum_{v=1}^{V}\boldsymbol{B}_v^\top\sum_{n:s_{vn}=j}\boldsymbol{\phi}(\mathbf{x}_{vn})+\boldsymbol{\zeta}_j^\top\sum_{v=1}^{V}N_{vj}\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\zeta}_j+r\boldsymbol{\zeta}_j^\top\boldsymbol{\zeta}_j]\Bigg)$$
$$=\exp\Bigg(-\frac{\alpha}{2}\sum_{j=1}^{J}[-2\boldsymbol{\zeta}_j^\top\sum_{v=1}^{V}\boldsymbol{B}_v^\top\sum_{n:s_{vn}=j}\boldsymbol{\phi}(\mathbf{x}_{vn})+\boldsymbol{\zeta}_j^\top\left(\sum_{v=1}^{V}N_{vj}\boldsymbol{B}_v^\top\boldsymbol{B}_v+r\mathbf{I}\right)\boldsymbol{\zeta}_j]\Bigg). \tag{B.6}$$

By using the quadratic property

$$-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^\top\boldsymbol{C}^{-1}(\mathbf{z}-\boldsymbol{\mu})=-\frac{1}{2}\left[\mathbf{z}^\top\boldsymbol{C}^{-1}\mathbf{z}-2\mathbf{z}^\top\boldsymbol{C}^{-1}\boldsymbol{\mu}+\boldsymbol{\mu}^\top\boldsymbol{C}^{-1}\boldsymbol{\mu}\right], \tag{B.7}$$

where

$$\boldsymbol{\Lambda}_j^{-1} = \sum_{v=1}^{V} N_{vj} \boldsymbol{B}_v^\top \boldsymbol{B}_v + r\mathbf{I}, \tag{B.8}$$

and

$$\begin{aligned} -2\mathbf{z}^\top \boldsymbol{\Lambda}_j^{-1} \boldsymbol{\mu} &= -2\boldsymbol{\zeta}_j^\top \sum_{v=1}^{V} \boldsymbol{B}_v^\top \sum_{n:s_{vn}=j} \boldsymbol{\phi}\left(\mathbf{x}_{vn}\right) \\ \boldsymbol{\mu}_j &= \boldsymbol{\Lambda}_j \sum_{v=1}^{V} \boldsymbol{B}_v^\top \sum_{n:s_{vn}=j} \boldsymbol{\phi}\left(\mathbf{x}_{vn}\right). \end{aligned} \tag{B.9}$$

By completing the square as: $\arg = \arg + \frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{C}^{-1}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{C}^{-1}\boldsymbol{\mu}$, the argument in (B.6) becomes

$$\begin{aligned} &\exp\left(-\frac{\alpha}{2} \sum_{j=1}^{J} [-2\boldsymbol{\zeta}_j^\top \sum_{v=1}^{V} \boldsymbol{B}_v^\top \sum_{n:s_{vn}=j} \boldsymbol{\phi}\left(\mathbf{x}_{vn}\right) + \boldsymbol{\zeta}_j^\top \left(\sum_{v=1}^{V} N_{vj} \boldsymbol{B}_v^\top \boldsymbol{B}_v + r\mathbf{I} \right) \boldsymbol{\zeta}_j] \right) \\ &= \exp\left(-\frac{\alpha}{2} \Big[\sum_{j=1}^{J} \left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right)^\top \boldsymbol{\Lambda}_j^{-1} \left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right) \Big] \right) \exp\left(-\frac{\alpha}{2} \sum_{j=1}^{J} \boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1} \boldsymbol{\mu}_j \right) \end{aligned} \tag{B.10}$$

Substituting (B.10) in (B.4) give us

$$\begin{aligned} p\left(\boldsymbol{\Phi}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r\right) = \frac{b^a}{\Gamma\left(a\right)} \iint \left(\frac{\alpha}{2\pi}\right)^{\sum_v L_v N_v/2} \left(\frac{\alpha r}{2\pi}\right)^{KJ/2} \times \\ \exp\left(-\frac{\alpha}{2} \Big[\sum_{j=1}^{J} \left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right)^\top \boldsymbol{\Lambda}_j^{-1} \left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right) \Big] \right) d\mathbf{Z} \\ \exp\left(-\alpha \left[\frac{1}{2} \sum_{v=1}^{V} \sum_{n=1}^{N_v} \boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^\top \boldsymbol{\phi}\left(\mathbf{x}_{vn}\right) - \frac{1}{2} \sum_{j=1}^{J} \boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1} \boldsymbol{\mu}_j + b \right] \right) \alpha^{a-1} d\alpha \end{aligned} \tag{B.11}$$

In equation (B.11), factors related to $\mathbf{Z}$ are grouped together. We integrated out $\mathbf{Z}$ using

$$\int \exp\left(-\frac{1}{2}\left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right)^\top \left[\alpha^{-1}\boldsymbol{\Lambda}_j\right]^{-1}\left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right)\right) d\boldsymbol{\zeta}_j = (2\pi)^{K/2}\left|\alpha^{-1}\boldsymbol{\Lambda}_j\right|^{1/2}$$
$$= (2\pi)^{K/2}\,\alpha^{-K/2}|\boldsymbol{\Lambda}_j|^{1/2}, \tag{B.12}$$

which is the normalization constant of P-dimensional Gaussian distribution. Since we have the sum over the number of correspondences (latent vectors), K, the above equation ranges for all of these clusters. The equation (B.11), becomes

$$p\left(\boldsymbol{\Phi}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r\right) = \frac{b^a}{\Gamma\left(a\right)} \int \left(\frac{\alpha}{2\pi}\right)^{\sum_v L_v N_v/2} \left(\frac{\alpha r}{2\pi}\right)^{KJ/2} \prod_{j=1}^{J} \left[\,(2\pi)^{K/2}\,\alpha^{-K/2}|\boldsymbol{\Lambda}_j|^{1/2}\right]$$
$$\exp\left(-\alpha\left[\frac{1}{2}\sum_{v=1}^{V}\sum_{n=1}^{N_v}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^\top\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right) - \frac{1}{2}\sum_{j=1}^{J}\boldsymbol{\mu}_j^\top\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j + b\right]\right)\alpha^{a-1}d\alpha \tag{B.13}$$

$$= \frac{b^a}{\Gamma\left(a\right)} \int \left(\frac{\alpha}{2\pi}\right)^{\sum_v L_v N_v/2} \left(\frac{\alpha r}{2\pi}\right)^{KJ/2} (2\pi)^{KJ/2}\,\alpha^{-KJ/2} \prod_{j=1}^{J}|\boldsymbol{\Lambda}_j|^{1/2}$$
$$\exp\left(-\alpha\left[\frac{1}{2}\sum_{v=1}^{V}\sum_{n=1}^{N_v}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^\top\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right) - \frac{1}{2}\sum_{j=1}^{J}\boldsymbol{\mu}_j^\top\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j + b\right]\right)\alpha^{a-1}d\alpha \tag{B.14}$$

The α parameter is integrated out by using the following normalization constant of a Gamma distribution

$$\int \alpha^{a'-1}\exp\left(-b\alpha\right)d\alpha = \frac{\Gamma\left(a'\right)}{b'^{a'}}. \tag{B.15}$$

Finally the likelihood is given by

$$p\left(\boldsymbol{\Phi}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r\right) = (2\pi)^{-\frac{\sum_v L_v N_v}{2}}\, r^{\frac{KJ}{2}}\,\frac{b^a}{b'^{a'}}\,\frac{\Gamma\left(a'\right)}{\Gamma\left(a\right)}\prod_{j=1}^{J}|\boldsymbol{\Lambda}_j|^{1/2}, \tag{B.16}$$

Here,

$$a' = a + \frac{\sum_v L_v N_v}{2}, \qquad b' = b + \frac{1}{2}\sum_{v=1}^{V}\sum_{n=1}^{N_v}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^\top\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right) - \frac{1}{2}\sum_{j=1}^{J}\boldsymbol{\mu}_j^\top\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j, \tag{B.17}$$

$$\boldsymbol{\mu}_j = \boldsymbol{\Lambda}_j \sum_{v=1}^{V} \boldsymbol{B}_v^\top \sum_{n:s_{vn}=j} \boldsymbol{\phi}\left(\mathbf{x}_{vn}\right), \qquad \boldsymbol{\Lambda}_j^{-1} = \sum_{v=1}^{V} N_{vj} \boldsymbol{B}_v^\top \boldsymbol{B}_v + r\mathbf{I}, \tag{B.18}$$

where N_{vj} is the number of descriptors assigned to cluster j in the shape v (view).

B.2 Posterior

The posterior for the precision parameter α is given by

$$p\left(\alpha|\boldsymbol{\Phi}, \mathbf{S}, \boldsymbol{\mathcal{W}}, a, b\right) = \mathcal{G}\left(\alpha|a', b'\right), \tag{B.19}$$

and the posterior for the latent vector $\boldsymbol{\zeta}_j$ is given by

$$p\left(\boldsymbol{\zeta}_j|\alpha, \boldsymbol{\Phi}, \mathbf{S}, \boldsymbol{\mathcal{W}}, r\right) = \mathcal{N}\left(\boldsymbol{\zeta}_j|\boldsymbol{\mu}_j, \alpha^{-1}\boldsymbol{\Lambda}_j\right) \tag{B.20}$$

The derivation for these posteriors is given by

$$\begin{aligned}
&p\left(\alpha|\boldsymbol{\Phi}, \mathbf{S}, \boldsymbol{\mathcal{W}}, a, b\right) \prod_{j=1}^{J} p\left(\boldsymbol{\zeta}_j|\alpha, \boldsymbol{\Phi}, \mathbf{S}, \boldsymbol{\mathcal{W}}, r\right) \propto p\left(\boldsymbol{\Phi}|\alpha, \mathbf{Z}, \mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r\right) p\left(\alpha|a, b\right) \times \\
&\prod_{j=1}^{J} p\left(\boldsymbol{\zeta}_j|\alpha, r\right) = \prod_{v=1}^{V} \prod_{n=1}^{N_v} \mathcal{N}\left(\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)|\boldsymbol{B}_v\boldsymbol{\zeta}_{s_{vn}}, \alpha^{-1}\mathbf{I}\right) \mathcal{G}\left(\alpha|a, b\right) \times \prod_{j=1}^{J} \mathcal{N}\left(\boldsymbol{\zeta}_j|\mathbf{0}, \left(\alpha r\right)^{-1}\mathbf{I}\right) \\
&= \prod_{v=1}^{V} \prod_{n=1}^{N_v} \left(\frac{\alpha}{2\pi}\right)^{L_d/2} \exp\left(-\frac{\alpha}{2}||\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right) - \boldsymbol{B}_v\boldsymbol{\zeta}_{s_{vn}}||^2\right) \prod_{j=1}^{J} \left(\frac{\alpha r}{2\pi}\right)^{K/2} \\
&\times \exp\left(-\frac{\alpha r}{2}||\boldsymbol{\zeta}_j||^2\right) \frac{b^a \alpha^{a-1}}{\Gamma\left(a\right)} \exp\left(-b\alpha\right) \\
&\propto \alpha^{a'-1} \exp\left(-b'\alpha\right) \prod_{j=1}^{J} |\boldsymbol{\Lambda}_j|^{-1/2} \exp\left(-\frac{\alpha}{2}\left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right)^\top \boldsymbol{\Lambda}_j^{-1}\left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right)\right) \\
&\propto \mathcal{G}\left(\alpha|a', b'\right) \prod_{j=1}^{J} \mathcal{N}\left(\boldsymbol{\zeta}_j|\boldsymbol{\mu}_j, \alpha^{-1}\boldsymbol{\Lambda}_j\right).
\end{aligned} \tag{B.21}$$

B.3 Inference for the nonlinear Model

E-Step

Let us recall the equation (3.7) that represents the marginalized (latent vectors, $\mathbf{Z}$, are integrated out) likelihood probability of the objects

$$p\left(\boldsymbol{\Phi}|\mathbf{S},\boldsymbol{\mathcal{W}},a,b,r\right)=(2\pi)^{-\frac{\sum_v L_v N_v}{2}}\, r^{\frac{KJ}{2}}\,\frac{b^a}{b\prime^{a\prime}}\,\frac{\Gamma\left(a\prime\right)}{\Gamma\left(a\right)}\prod_{j=1}^{J}|\boldsymbol{\Lambda}_j|^{1/2}. \tag{B.22}$$

with

$$a\prime = a+\frac{\sum_v L_v N_v}{2} \tag{B.23}$$

$$b' = b+\frac{1}{2}\sum_{v=1}^{V}\sum_{n=1}^{N_v}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^{\top}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)-\frac{1}{2}\sum_{j=1}^{J}\boldsymbol{\mu}_j^{\top}\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j, \tag{B.24}$$

$$\boldsymbol{\mu}_j = \boldsymbol{\Lambda}_j\sum_{v=1}^{V}\boldsymbol{B}_v^{\top}\sum_{n:\mathbf{s}_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right),$$

$$|\boldsymbol{\Lambda}_j^{-1}| = \sum_{v=1}^{V}N_{vj}\boldsymbol{B}_v^{\top}\boldsymbol{B}_v+r\mathbf{I}. \tag{B.25}$$

Now, let us take a look at the E-step, specifically to equation (3.13), which is given by

$$\frac{p\left(\boldsymbol{\Phi}|s_{vn}=j,\mathbf{S}_{\backslash vn},\boldsymbol{\mathcal{W}},a,b,r\right)}{p\left(\boldsymbol{\Phi}_{\backslash vn}|\mathbf{S}_{\backslash vn},\boldsymbol{\mathcal{W}},a,b,r\right)} = (2\pi)^{-\frac{D_v}{2}}r^{\frac{1}{2}I(j>J\backslash vn)}\frac{b_{\backslash vn}^{\prime a'_{vn}}\Gamma\left(a'_{s_{vn}=j}\right)\left|\boldsymbol{\Lambda}_{j,s_{vn}=j}\right|^{\frac{1}{2}}}{b_{s_{vn}=j}^{\prime a'_{s_{vn}=j}}\Gamma\left(a'_{\backslash vn}\right)\quad\left|\boldsymbol{\Lambda}_{j\backslash vn}\right|^{\frac{1}{2}}}, \tag{B.26}$$

The above equation represents the ratio between the probability of changing the assignment of object $\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)$ to cluster j, wrt the likelihood obtained by removing the same object. Note that the denominator value is always the same across the reassignments of $\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)$ to the different clusters. Additionally, if we remove or re-assign any object, then we require to recalculate all b', $\boldsymbol{\mu}_j$ and $\boldsymbol{\Lambda}_j^{-1}$. Considering the above mentioned, let us take a look at the full form of the denominator and the numerator,

$$p(\boldsymbol{\Phi}|s_{vn}=j,\mathbf{S}_{\backslash vn},\boldsymbol{\mathcal{W}},a,b,r) = (2\pi)^{-\frac{\sum_v M_v N_v}{2}}\gamma^{\frac{KJ}{2}}\frac{b^a}{b'^{a'}_{s_{vn}=j}}\frac{\Gamma(a')}{\Gamma(a)}\prod_{i=1,i\neq j}^{J}|\boldsymbol{\Lambda}_{i,\backslash vn}|^{\frac{1}{2}}|\boldsymbol{\Lambda}_{j,s_{vn}=j}|^{\frac{1}{2}}. \tag{B.27}$$

$$p(\boldsymbol{\Phi}_{\backslash vn}|\mathbf{S}_{\backslash vn},\boldsymbol{\mathcal{W}},a,b,r)=(2\pi)^{-\frac{\sum_{k\neq v}D_kN_k}{2}-\frac{D_v(N_v-1)}{2}}\gamma^{\frac{KJ}{2}}\frac{b^a}{b'^{a'_{\backslash vn}}_{\backslash vn}}\frac{\Gamma(a'_{\backslash vn})}{\Gamma(a)}\prod_{i=1}^{J}|\boldsymbol{\Lambda}_{i\backslash vn}|^{\frac{1}{2}}, \tag{B.28}$$

For completeness, the following expressions are given

$$a'_{\backslash vn}=a+\frac{\sum_{k\neq v}^{V}D_kN_k}{2}+\frac{D_v(N_v-1)}{2},$$

$$b'_{\backslash vn}=b+\frac{1}{2}\sum_{v=1}^{V}\sum_{m=1,m\neq n}^{N_v}\boldsymbol{\phi}\left(\mathbf{x}_{vm}\right)^{\top}\boldsymbol{\phi}\left(\mathbf{x}_{vm}\right)-\frac{1}{2}\sum_{j=1}^{J}\boldsymbol{\mu}_{j\backslash vn}\boldsymbol{\Lambda}^{-1}_{j\backslash vn}\boldsymbol{\mu}_{j\backslash vn},$$

$$\boldsymbol{\mu}_{j\backslash vn}=\boldsymbol{\Lambda}_{j\backslash vn}\sum_{v=1}^{V}\boldsymbol{B}_v^{\top}\sum_{m:s_{vn}=j,m\neq n}\boldsymbol{\phi}\left(\mathbf{x}_{vm}\right),$$

$$\boldsymbol{\Lambda}^{-1}_{j\backslash nv}=\sum_{k\neq v}^{V}N_{kj}\boldsymbol{B}_k^{\top}\boldsymbol{B}_k+(N_{vj}-1)\boldsymbol{B}_v^{\top}\boldsymbol{B}_v+r\mathbf{I}.$$

It is clear how we get the relation given in (B.26) using (B.27) and (B.28). But, what do we do when $j=J+1$? We start by calculating (B.27) when $j=J+1$, which is given in (B.29).

$$p(\boldsymbol{\Phi}|s_{vn}=J{+}1,\mathbf{S}_{\backslash vn},\boldsymbol{\mathcal{W}},a,b,r)=(2\pi)^{-\frac{\sum_v M_vN_v}{2}}\gamma^{\frac{KJ}{2}}\frac{b^a}{b'^{a'}_{s_{vn}=J+1}}\frac{\Gamma(a')}{\Gamma(a)}\prod_{j=1}^{J}|\boldsymbol{\Lambda}_{j,\backslash vn}|^{\frac{1}{2}}|\boldsymbol{\Lambda}_{s_{vn}=J+1}|^{\frac{1}{2}}. \tag{B.29}$$

with

$$b'_{s_{vn}=J+1}=b'_{\backslash vn}+\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^{\top}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)-\boldsymbol{\mu}_{s_{vn}=J+1}\boldsymbol{\Lambda}^{-1}_{s_{vn}=J+1}\boldsymbol{\mu}_{s_{vn}=J+1}$$

$$\boldsymbol{\mu}_{s_{vn}=J+1}=\boldsymbol{\Lambda}_{s_{vn}=J+1}\boldsymbol{B}_v^{\top}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)$$

$$\boldsymbol{\Lambda}^{-1}_{s_{vn}=J+1}=\boldsymbol{B}_v^{\top}\boldsymbol{B}_v+r\mathbf{I}.$$

Next, using (B.28) and (B.29), we are able to calculate (B.26) for $j=J+1$ as

$$\frac{p(\boldsymbol{\Phi}|s_{vn}=J+1,\mathbf{S}_{\backslash vn},\boldsymbol{\mathcal{W}},a,b,r)}{p(\boldsymbol{\Phi}_{\backslash vn}|\mathbf{S}_{\backslash vn},\boldsymbol{\mathcal{W}},a,b,r)}=(2\pi)^{-\frac{D_v}{2}}\gamma^{\frac{1}{2}}\frac{b'^{a'_{\backslash vn}}_{\backslash vn}}{b'^{a'}_{s_{vn}=J+1}}\frac{\Gamma(a')}{\Gamma(a'_{\backslash vn})}|\boldsymbol{\Lambda}_{s_{vn}=J+1}|^{\frac{1}{2}}. \tag{B.30}$$

Another case to consider is when the j-th cluster (between 1 and J) has one object, i.e. $N_{\cdot j}=1$. Clearly, if this object is moved to another cluster or removed, then the number of clusters is decreased by one.

M-Step

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

$$\frac{\partial \log p\left(\boldsymbol{\Phi}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \boldsymbol{B}_v} = \frac{\partial \log p\left(\mathbf{S}|\gamma\right)}{\partial \boldsymbol{B}_v} + \frac{\partial \log p\left(\boldsymbol{\Phi}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r\right)}{\partial \boldsymbol{B}_v}, \tag{B.31}$$

Since the derivative of the first term in the above expression is zero, the expression becomes

$$\begin{aligned}
\frac{\partial \log p\left(\mathbf{X}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \boldsymbol{B}_v} &= \frac{\partial \log p\left(\mathbf{X}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r\right)}{\partial \boldsymbol{B}_v} \\
&= \frac{\partial \log \left[(2\pi)^{-\frac{\sum_v L_v N_v}{2}} r^{\frac{JK}{2}} \frac{b^a}{b'^{a'}} \frac{\Gamma(a')}{\Gamma(a)} \prod_{j=1}^{J} |\boldsymbol{\Lambda}_j|^{1/2} \right]}{\partial \boldsymbol{B}_v}, \\
&= \frac{\partial \log \left[\frac{cte}{b'^{a'}} \prod_{j=1}^{J} |\boldsymbol{\Lambda}_j|^{1/2} \right]}{\partial \boldsymbol{B}_v}, \\
&= -\frac{a'}{b'} \frac{\partial b'}{\partial \boldsymbol{B}_v} + \frac{1}{2} \sum_{j=1}^{J} \frac{\partial \log |\boldsymbol{\Lambda}_j|}{\partial \boldsymbol{B}_v} + 0, \\
&= -\frac{a'}{b'} \frac{\partial b'}{\partial \boldsymbol{B}_v} + \frac{1}{2} \sum_{j=1}^{J} \operatorname{tr}\left(\boldsymbol{\Lambda}_j^{-1} \frac{\partial \boldsymbol{\Lambda}_j}{\partial \boldsymbol{B}_v} \right)
\end{aligned} \tag{B.32}$$

where

$$\frac{\partial b'}{\partial \boldsymbol{B}_v} = \frac{\partial}{\partial \boldsymbol{B}_v} \Big[\frac{1}{2} \sum_{v=1}^{V} \sum_{n=1}^{N_v} \boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^\top \boldsymbol{\phi}\left(\mathbf{x}_{vn}\right) - \frac{1}{2} \sum_{j=1}^{J} \boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1} \boldsymbol{\mu}_j + b \Big], \tag{B.33}$$

Here the second factor of the argument is the only which depends on $\boldsymbol{B}_v$.

$$\begin{aligned}
\frac{\partial b'}{\partial \boldsymbol{B}_v} &= \frac{\partial}{\partial \boldsymbol{B}_v}\Big[-\frac{1}{2}\sum_{j=1}^{J}\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\Big] \\
&= -\frac{1}{2}\sum_{j=1}^{J}\frac{\partial}{\partial \boldsymbol{B}_v}\Big[\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\Big] = -\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\frac{\partial}{\partial \boldsymbol{B}_v}\Big[\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\Big]\right) \\
&= -\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\frac{\partial \boldsymbol{\mu}_j^\top}{\partial \boldsymbol{B}_v}\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j + \boldsymbol{\mu}_j^\top\left(\frac{\partial \boldsymbol{\Lambda}_j^{-1}}{\partial \boldsymbol{B}_v}\boldsymbol{\mu}_j + \boldsymbol{\Lambda}_j^{-1}\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{B}_v}\right)\right) \\
&= -\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\partial\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right) + \operatorname{tr}\left(\boldsymbol{\mu}_j^\top \partial\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right) + \operatorname{tr}\left(\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\partial\boldsymbol{\mu}_j\right), \qquad \text{(B.34)}
\end{aligned}$$

by applying trace properties (transpose elements)

$$\begin{aligned}
\frac{\partial b'}{\partial \boldsymbol{B}_v} &= -\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\partial\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right) + \operatorname{tr}\left(\boldsymbol{\mu}_j^\top \partial\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right) + \operatorname{tr}\left(\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\partial\boldsymbol{\mu}_j\right) \\
&= -\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\partial\boldsymbol{\mu}_j\right) + \operatorname{tr}\left(\boldsymbol{\mu}_j^\top \partial\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right) + \operatorname{tr}\left(\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\partial\boldsymbol{\mu}_j\right) \\
&= -\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top \partial\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right) + 2\operatorname{tr}\left(\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\partial\boldsymbol{\mu}_j\right) \\
&= \underbrace{-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\partial\boldsymbol{\mu}_j\right)}_{A}\underbrace{-\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top \partial\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right)}_{B} \qquad \text{(B.35)}
\end{aligned}$$

First, for the B part we have:

$$
\begin{aligned}
-\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^{\top}\partial\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right) &= -\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^{\top}\partial\left[\sum_{v=1}^{V}N_{vj}\boldsymbol{B}_v^{\top}\boldsymbol{B}_v + r\mathbf{I}\right]\boldsymbol{\mu}_j\right) \\
&= -\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^{\top}\left[N_{vj}\left(\partial\boldsymbol{B}_v^{\top}\boldsymbol{B}_v + \boldsymbol{B}_v^{\top}\partial\boldsymbol{B}_v\right)\right]\boldsymbol{\mu}_j\right) \\
&= -\frac{1}{2}\sum_{j=1}^{J}\left[N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j^{\top}\partial\boldsymbol{B}_v^{\top}\boldsymbol{B}_v\boldsymbol{\mu}_j\right) + N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j^{\top}\boldsymbol{B}_v^{\top}\partial\boldsymbol{B}_v\boldsymbol{\mu}_j\right)\right] \\
&= -\frac{1}{2}\sum_{j=1}^{J}\left[N_{vj}\operatorname{tr}\left(\left(\boldsymbol{\mu}_j^{\top}\partial\boldsymbol{B}_v^{\top}\boldsymbol{B}_v\boldsymbol{\mu}_j\right)^{\top}\right) + N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^{\top}\boldsymbol{B}_v^{\top}\partial\boldsymbol{B}_v\right)\right] \\
&= -\frac{1}{2}\sum_{j=1}^{J}\left[N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^{\top}\boldsymbol{B}_v^{\top}\partial\boldsymbol{B}_v\boldsymbol{\mu}_j\right) + N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^{\top}\boldsymbol{B}_v^{\top}\partial\boldsymbol{B}_v\right)\right] \\
&= -\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^{\top}\boldsymbol{B}_v^{\top}\partial\boldsymbol{B}_v\right). \qquad \text{(B.36)}
\end{aligned}
$$

By using derivatives properties for trace forms as

$$
\frac{\partial\operatorname{tr}\left[F\left(\mathbf{X}\right)\right]}{\partial\mathbf{X}} = f\left(\mathbf{X}\right)^{\top}, \qquad \text{(B.37)}
$$

where $f\left(\cdot\right)$ is the scalar derivative of $F\left(\cdot\right)$, the equation (B.36) becomes

$$
-\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^{\top}\partial\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right) = -\sum_{j=1}^{J}N_{vj}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^{\top}\boldsymbol{B}_v^{\top}\right)^{\top} = -\sum_{j=1}^{J}N_{vj}\boldsymbol{B}_v\boldsymbol{\mu}_j\boldsymbol{\mu}_j^{\top}. \qquad \text{(B.38)}
$$

Besides, for the A part we have:

$$
-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^{\top}\boldsymbol{\Lambda}_j^{-1}\partial\boldsymbol{\mu}_j\right) \rightarrow \frac{\partial\boldsymbol{\mu}_j}{\partial\boldsymbol{B}_v} = \frac{\partial}{\partial\boldsymbol{B}_v}\left[\boldsymbol{\Lambda}_j\sum_{v=1}^{V}\boldsymbol{B}_v^{\top}\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\right]. \qquad \text{(B.39)}
$$

The derivative for $\boldsymbol{\mu}_j$ is given by

$$
\frac{\partial\boldsymbol{\mu}_j}{\partial\boldsymbol{B}_v} = \underbrace{\partial\boldsymbol{\Lambda}_j\left(\sum_{v=1}^{V}\boldsymbol{B}_v^{\top}\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\right)}_{C} + \underbrace{\boldsymbol{\Lambda}_j\partial\left(\sum_{v=1}^{V}\boldsymbol{B}_v^{\top}\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\right)}_{V} \qquad \text{(B.40)}
$$

For the C part, we have

$$\begin{aligned}\partial\boldsymbol{\Lambda}_j\left(\sum_{v=1}^{V}\boldsymbol{B}_v^\top\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\right) &= \frac{\partial}{\partial\boldsymbol{B}_v}\left[\left(\sum_{v=1}^{V}N_{vj}\boldsymbol{B}_v^\top\boldsymbol{B}_v+r\mathbf{I}\right)^{-1}\right]\left(\sum_{v=1}^{V}\boldsymbol{B}_v^\top\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\right)\\ &= -\boldsymbol{\Lambda}_j\frac{\partial}{\partial\boldsymbol{B}_v}\left[\sum_{v=1}^{V}N_{vj}\boldsymbol{B}_v^\top\boldsymbol{B}_v+r\mathbf{I}\right]\boldsymbol{\Lambda}_j\sum_{v=1}^{V}\boldsymbol{B}_v^\top\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\\ &= -\boldsymbol{\Lambda}_j\left(N_{vj}\left[\partial\boldsymbol{B}_v^\top\boldsymbol{B}_v+\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right]\right)\boldsymbol{\mu}_j. \qquad \text{(B.41)}\end{aligned}$$

The v part is computed as

$$\boldsymbol{\Lambda}_j\partial\left(\sum_{v=1}^{V}\boldsymbol{B}_v^\top\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\right)=\boldsymbol{\Lambda}_j\partial\boldsymbol{B}_v^\top\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right). \qquad \text{(B.42)}$$

Then the A part becomes,

$$
\begin{aligned}
&-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\boldsymbol{\Lambda}_j^{-1}\partial\boldsymbol{\mu}_j\right)= \\
&-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\boldsymbol{\Lambda}_j^{-1}\left[-\boldsymbol{\Lambda}_j\left(N_{vj}\left[\partial\boldsymbol{B}_v^\top\boldsymbol{B}_v+\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right]\right)\boldsymbol{\mu}_j+\boldsymbol{\Lambda}_j\partial\boldsymbol{B}_v^\top\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\right]\right) \\
&=\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\Lambda}_j\left(\partial\boldsymbol{B}_v^\top\boldsymbol{B}_v\right)\boldsymbol{\mu}_j\right)+\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\Lambda}_j\left(\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right)\boldsymbol{\mu}_j\right) \\
&-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\Lambda}_j\partial\boldsymbol{B}_v^\top\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\right) \\
&=\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\partial\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\mu}_j\right)+\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\boldsymbol{\mu}_j\right) \\
&-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\partial\boldsymbol{B}_v^\top\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\right) \\
&=\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\left(\boldsymbol{\mu}_j^\top\partial\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\mu}_j\right)^\top\right)+\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right) \\
&-\sum_{j=1}^{J}\operatorname{tr}\left(\left(\boldsymbol{\mu}_j^\top\partial\boldsymbol{B}_v^\top\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\right)^\top\right) \\
&=\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\boldsymbol{\mu}_j\right)+\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right) \\
&-\sum_{j=1}^{J}\operatorname{tr}\left(\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^\top\partial\boldsymbol{B}_v\boldsymbol{\mu}_j\right) \\
&=\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right)+\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right) \\
&-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^\top\partial\boldsymbol{B}_v\right) \\
&=2\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right)-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^\top\partial\boldsymbol{B}_v\right). \qquad \text{(B.43)}
\end{aligned}
$$

By using the derivatives properties for trace forms described above, we have

$$
\begin{aligned}
-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\boldsymbol{\Lambda}_j^{-1}\partial\boldsymbol{\mu}_j\right) &= 2\sum_{j=1}^{J}N_{vj}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\boldsymbol{B}_v^\top\right)^\top-\sum_{j=1}^{J}\left(\boldsymbol{\mu}_j\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)^\top\right)^\top\\
&= 2\sum_{j=1}^{J}N_{vj}\boldsymbol{B}_v\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top-\sum_{j=1}^{J}\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\boldsymbol{\mu}_j^\top. \qquad \text{(B.44)}
\end{aligned}
$$

Finally

$$
\begin{aligned}
\frac{\partial b'}{\partial\boldsymbol{B}_v} &= 2\sum_{j=1}^{J}N_{vj}\boldsymbol{B}_v\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top-\sum_{j=1}^{J}\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\boldsymbol{\mu}_j^\top--\sum_{j=1}^{J}N_{vj}\boldsymbol{B}_v\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\\
&= \sum_{j=1}^{J}N_{vj}\boldsymbol{B}_v\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top-\sum_{j=1}^{J}\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\boldsymbol{\mu}_j^\top\\
&= \sum_{j=1}^{J}\left\{N_{vj}\boldsymbol{B}_v\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top-\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\boldsymbol{\mu}_j^\top\right\}. \qquad \text{(B.45)}
\end{aligned}
$$

For the part B,

$$
\begin{aligned}
\frac{1}{2}\sum_{j=1}^{J}\operatorname{Tr}\left(\boldsymbol{\Lambda}_j^{-1}\frac{\partial\boldsymbol{\Lambda}_j}{\partial\boldsymbol{B}_v}\right) &= \frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\Lambda}_j^{-1}\left(-\boldsymbol{\Lambda}_j\left(N_{vj}\left[\partial\boldsymbol{B}_v^\top\boldsymbol{B}_v+\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right]\right)\boldsymbol{\Lambda}_j\right)\right)\\
&= -\frac{1}{2}\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\Lambda}_j\left[\partial\boldsymbol{B}_v^\top\boldsymbol{B}_v+\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right]\boldsymbol{\Lambda}_j\right)\\
&= -\frac{1}{2}\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\Lambda}_j\left[\partial\boldsymbol{B}_v^\top\boldsymbol{B}_v+\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right]\boldsymbol{\Lambda}_j\right)\\
&= -\frac{1}{2}\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\partial\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\Lambda}_j\right)-\frac{1}{2}\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\boldsymbol{\Lambda}_j\right)\\
&= -\frac{1}{2}\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\left(\partial\boldsymbol{B}_v^\top\boldsymbol{B}_v\boldsymbol{\Lambda}_j\right)^\top\right)-\frac{1}{2}\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\Lambda}_j\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right)\\
&= -\frac{1}{2}\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\Lambda}_j\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right)-\frac{1}{2}\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\Lambda}_j\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right)\\
&= -\sum_{j=1}^{J}N_{vj}\operatorname{tr}\left(\boldsymbol{\Lambda}_j\boldsymbol{B}_v^\top\partial\boldsymbol{B}_v\right)=-\sum_{j=1}^{J}N_{vj}\boldsymbol{B}_v\boldsymbol{\Lambda}_j. \qquad \text{(B.46)}
\end{aligned}
$$

Finally the derivative of the log-likelihood is computed as

$$\frac{\partial \log p\left(\mathbf{X}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \boldsymbol{B}_v} = -\frac{a'}{b'}\frac{\partial b'}{\partial \boldsymbol{B}_v} + \frac{1}{2}\sum_{j=1}^{J} \mathrm{tr}\left(\boldsymbol{\Lambda}_j^{-1}\frac{\partial \boldsymbol{\Lambda}_j}{\partial \boldsymbol{B}_v}\right)$$
$$= -\frac{a'}{b'}\left[\sum_{j=1}^{J}\left\{N_{vj}\boldsymbol{B}_v\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top - \sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\boldsymbol{\mu}_j^\top\right\}\right] - \sum_{j=1}^{J} N_{vj}\boldsymbol{B}_v\boldsymbol{\Lambda}_j. \tag{B.47}$$

We can obtain the projection matrices that maximize the joint likelihood analytically as follows,

$$\boldsymbol{B}_v = -\frac{a'}{b'}\left(\sum_{j=1}^{J}\sum_{n:s_{vn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{vn}\right)\boldsymbol{\mu}_j^\top\right)\left(\sum_{j=1}^{J} N_{vj}\boldsymbol{\Lambda}_j + \frac{a'}{b'}N_{vj}\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\right)^{-1}. \tag{B.48}$$

Appendix C

Multiview Warped Mixture Models

This appendix concerns the Multiview Warped Mixture Model framework developed in Chapter 4 and which is used as a backbone for the methodology developed in the rest of the chapter. The details of this appendix aim at providing a useful reference for the complete derivation of the method.

C.1 Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) is a Markov chain Monte Carlo (MCMC) method that uses the derivatives of the density function being sampled to generate efficient transitions spanning the posterior (MacKay, 2002).

The goal of sampling is to draw from a density $p(\boldsymbol{\theta})$ for parameters $\boldsymbol{\theta}$. This is typically a Bayesian posterior $p(\boldsymbol{\theta}|\mathbf{x})$ given data $\mathbf{x}$.

C.1.1 Auxiliary Momentum Variable

HMC introduces auxiliary momentum variables $\boldsymbol{\rho}$ and draws from a joint density

$$p(\boldsymbol{\rho}, \boldsymbol{\theta}) = p(\boldsymbol{\rho}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{C.1}$$

Commonly, the auxiliary density is a multivariate normal that does not depend on the model parameters $\boldsymbol{\theta}$,

$$\boldsymbol{\rho} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{C.2}$$

where the covariance matrix $\boldsymbol{\Sigma}$ acts as an Euclidean metric to rotate and scale the target distribution.

C.1.2 The Hamiltonian

The joint density $p(\boldsymbol{\rho}, \boldsymbol{\theta})$ defines a Hamiltonian

$$\begin{aligned} H(\boldsymbol{\rho}, \boldsymbol{\theta}) &= -\log p(\boldsymbol{\rho}, \boldsymbol{\theta}), && \text{(C.3)} \\ &= -\log p(\boldsymbol{\rho}|\boldsymbol{\theta}) - \log p(\boldsymbol{\theta}), && \text{(C.4)} \\ &= T(\boldsymbol{\rho}|\boldsymbol{\theta}) + V(\boldsymbol{\theta}), && \text{(C.5)} \end{aligned}$$

where the term $T(\boldsymbol{\rho}|\boldsymbol{\theta}) = -\log p(\boldsymbol{\rho}|\boldsymbol{\theta})$, is called the *kinetic energy* and the term $V(\boldsymbol{\theta}) = -\log p(\boldsymbol{\theta})$, is called the *potential energy.*

C.1.3 Transitions for the HMC

Starting from the current value of the parameters $\boldsymbol{\theta}$, a transition to a new state of the model parameters is generated in two stages before being subjected to a Metropolis accept step. Hence, the joint system $(\boldsymbol{\theta}, \boldsymbol{\rho})$ made up of the current parameter values $\boldsymbol{\theta}$ and new momentum $\boldsymbol{\rho}$ is evolved via Hamilton's equations,

$$\begin{aligned} \frac{d\boldsymbol{\theta}}{dt} &= +\frac{\partial T}{\partial \boldsymbol{\rho}}, && \text{(C.6)} \\ \frac{d\boldsymbol{\rho}}{dt} &= -\frac{\partial V}{\partial \boldsymbol{\theta}}. && \text{(C.7)} \end{aligned}$$

The Hamiltonian Monte Carlo algorithm starts at a specified initial set of parameters $\boldsymbol{\theta}$. Then, for a given number of iterations, a new momentum vector is sampled and the current value of the parameter $\boldsymbol{\theta}$ is updated using the leapfrog integrator[1] with discretization time ϵ and number of steps L according to the Hamiltonian dynamics

[1] Leapfrog integrator is a numerical integration algorithm specifically adapted to provide stable results for Hamiltonian systems of equations.

(MacKay, 2002). Then a Metropolis acceptance step is applied, and a decision is made whether to update to the new state $(\boldsymbol{\theta}^*, \boldsymbol{\rho}^*)$ or keep the existing state.

C.1.4 Approximation of $p\left(\mathbf{Z}_v|\mathbf{Y}_v\right)$

We approximate $p(\mathbf{z}_v^*|\mathbf{Z}_v, \mathbf{Y}_v)$ by sampling points from the latent mixture and warping them, using the following procedure

1. Draw latent assignment

$$s_v^\star \sim \text{Mult}\left(\frac{N_{v1}}{N_v+\eta}, \cdots, \frac{N_{vC}}{N_v+\eta}, \frac{\eta}{N_v+\eta}\right) \tag{C.8}$$

2. Draw precision matrix

$$\mathbf{R}^\star \sim \mathcal{W}\left(\boldsymbol{\Lambda}_{s_v^\star}^{-1}, \nu_{s_v^\star}\right) \tag{C.9}$$

3. Draw mean

$$\boldsymbol{\mu}^\star \sim \mathcal{N}\left(\mathbf{u}_{s_v^\star}, \left(r_{s_v^\star}\mathbf{R}^\star\right)^{-1}\right) \tag{C.10}$$

4. Draw latent coordinates

$$\mathbf{z}_v^\star \sim \mathcal{N}\left(\boldsymbol{\mu}^\star, \mathbf{R}^{\star-1}\right) \tag{C.11}$$

C.2 Gradients for Hamiltonian Monte Carlo

To sample latent vectors $\mathbf{Z}_v$ from the posterior distribution $p\left(\mathbf{Z}_v|\mathbf{S}_v, \mathbf{X}_v, \boldsymbol{\theta}^v, \boldsymbol{\Lambda}, \nu, \mathbf{u}, \mathbf{r}\right)$ we need to compute the gradient of the log-unnormalized-posterior w.r.t. $\mathbf{Z}_v$ as

$$\frac{\partial}{\partial \mathbf{Z}_v}[\log p(\mathbf{X}_v|\mathbf{Z}_v, \boldsymbol{\theta}^v) + \log p(\mathbf{Z}_v|\mathbf{S}_v, \boldsymbol{\Lambda}, \nu, \mathbf{u}, r)], \tag{C.12}$$

where the first term of the gradient in (C.12) is computed as

$$\frac{\partial \log p(\mathbf{X}_v|\mathbf{Z}_v,\boldsymbol{\theta})}{\partial \mathbf{Z}_v} \rightarrow \frac{\partial \log p(\mathbf{X}_v|\mathbf{Z}_v,\boldsymbol{\theta})}{\partial \mathbf{K}_v} = -\frac{1}{2}D_v\mathbf{K}_v^{-1} + \frac{1}{2}\mathbf{K}_v^{-1}\mathbf{X}_v\mathbf{X}_v{}^T\mathbf{K}_v^{-1}, \tag{C.13}$$

where we used the chain rule to compute the kernel derivatives as,

$$\frac{\partial k\left(\mathbf{x}_{nv},\mathbf{x}_{mv}\right)}{\partial \mathbf{x}_{nv}} = -\frac{\sigma_{vf}^2}{\ell_v^2}\exp\left(-\frac{1}{2\ell_v^2}\left(\mathbf{x}_{nv}-\mathbf{x}_{mv}\right)^\top\left(\mathbf{x}_{nv}-\mathbf{x}_{mv}\right)\right)\left(\mathbf{x}_{nv}-\mathbf{x}_{mv}\right). \tag{C.14}$$

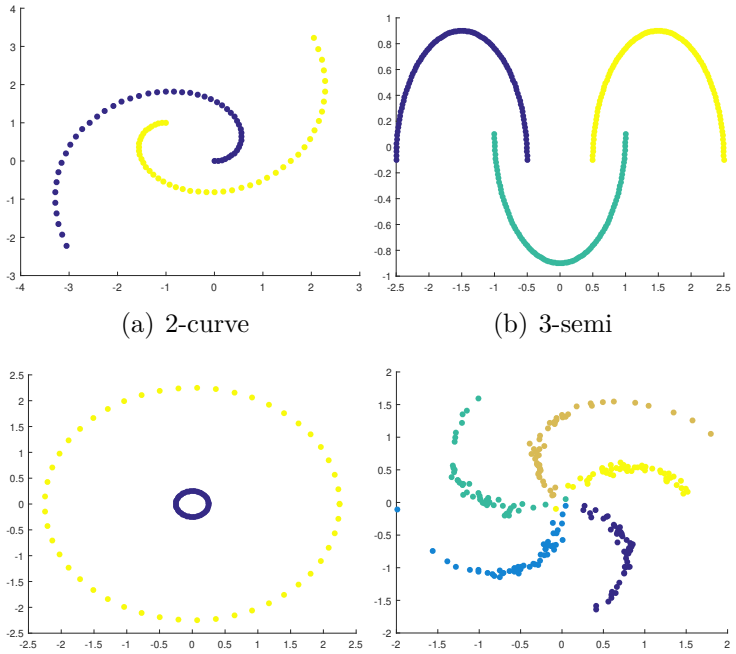
Finally te second term of the equation (4.14) is set as

$$\frac{\partial \log p(\mathbf{X}_v|\mathbf{Z}_v,\boldsymbol{S}^v,\nu,\mathbf{u},r)}{\partial \mathbf{x}_{nv}} = -\nu_{z_m}\boldsymbol{S}_{z_{mv}}^{-1}\left(\mathbf{x}_{nv}-\mathbf{u}_{z_n}\right). \tag{C.15}$$

We also infer kernel parameters $\boldsymbol{\theta}^v$ via HMC, using the gradient of the log unnormalized posterior with respect to the kernel parameters.

C.3 Common clustering ML datasets

We define here the real-world and synthetic datasets used as input data to asses de model performance. None of these datasets can be appropriately clustered by a Gaussian mixture model (GMM) (i.e., computing the cluster assignments in the input space). We use wine, and vowel datasets obtained from the LIBSVM multi-class data (Chang and Lin, 2011). In addition, we use four synthetic datasets: 2-curveC.1(a), 3-semi C.1(b), 2-circle C.1(c), and pinwheel C.1(d). 2-curve and 3-semi are synthetic datasets having 2 and 3 curved lines in which a classical GMM with many components could separate the two lines only by breaking each line into many clusters. 2-circle is an interesting manifold learning dataset consisting of two concentric circles. Finally, Pinwheel is a five-armed variant of the pinwheel dataset of Adams and Ghahramani (2009) generated by warping a mixture of Gaussians into a spiral. Figure C.1, shows an example of the four synthetic datasets.



(a) 2-curve (b) 3-semi

(c) 2-circle (d) Pinwheel

Figure C.1: Examples of the real-world and synthetic datasets used as input data to asses de model performance.

C.4 Additional Results for the MVWMM

We show additional experiments for the non-rigid 3D objects. Results show that our approach can efficiently match different shape structures, even with shape exhibiting partial information.

C.4.1 Non-rigid shape correspondence

Figures C.2 to C.4 show the correspondence results for additional shapes exhibiting different poses. The results also shows that our model can efficiently establish meaningful correspondences for different shape structures (average Rand indexes of 0.81, 0.77, and 0.72 respectively).

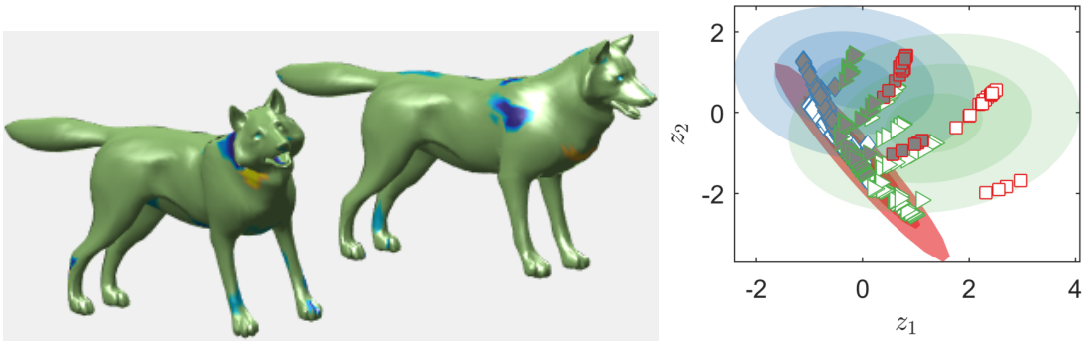


Figure C.2: Experimental results for the TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.8073. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

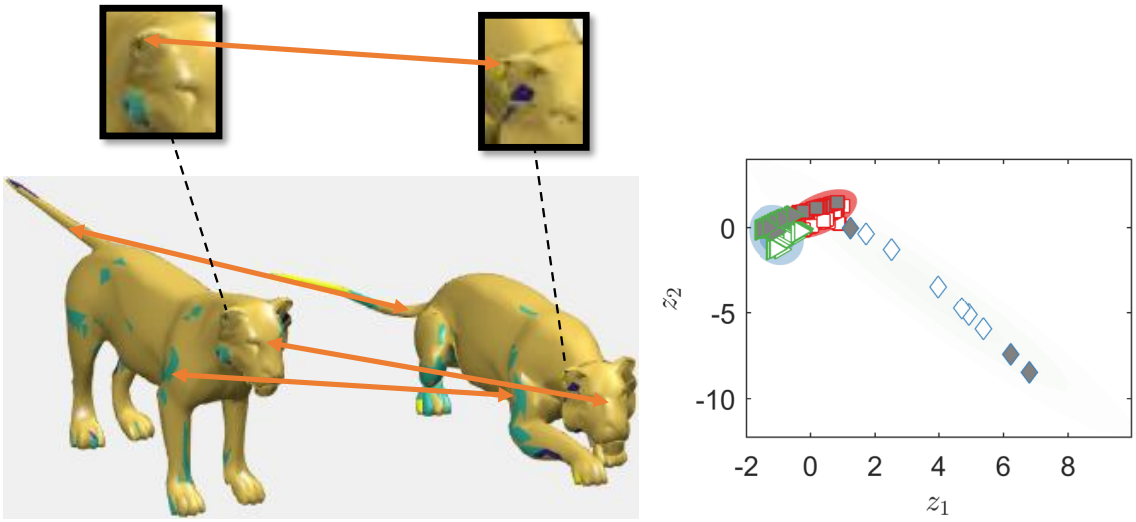


Figure C.3: Experimental results for the TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.7675. Orange arrows describe candidate matches between shapes. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

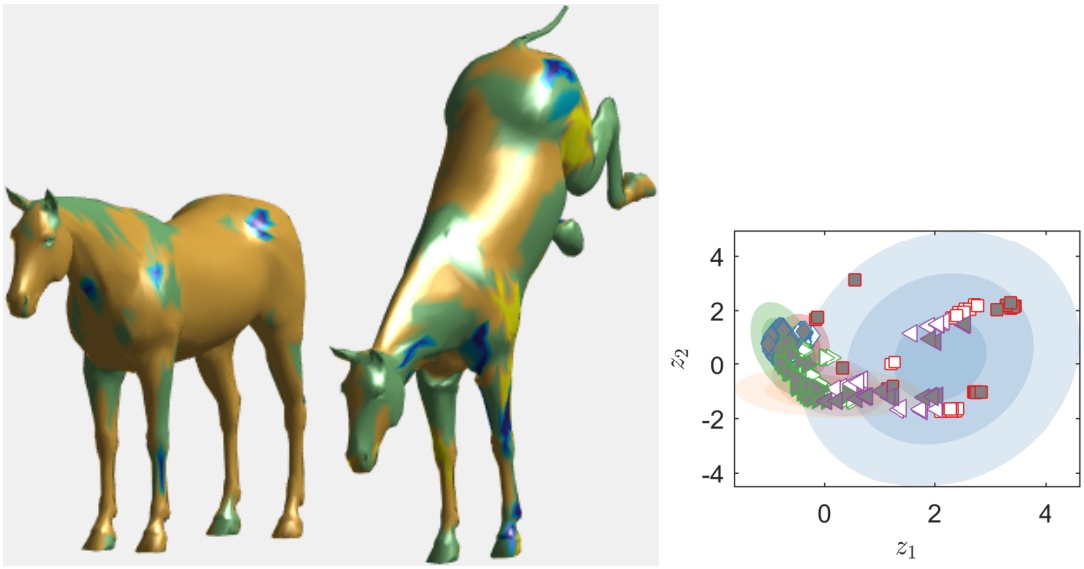


Figure C.4: Experimental results for the TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.7238. Same colors in related regions are considered candidate matches. Gray and white markers relate a specific view (left and right object), so similar markers are candidate matches in the latent space.

Appendix D

Publications

D.1 Published Papers

- Hernán F. García, Mauricio A. Álvarez, and Álvaro Orozco. Groupwise shape correspondences on 3d brain structures using probabilistic latent variable models. In *Advances in Visual Computing - 11th International Symposium, ISVC 2015, Las Vegas, NV, USA, December 14-16, 2015, Proceedings, Part I*, volume 9474 of *Lecture Notes in Computer Science*, pages 491–500. Springer, 2015. doi: 10.1007/978-3-319-27857-5_44. URL https://doi.org/10.1007/978-3-319-27857-5_44

- Hernán F. García, Álvaro Ángel Orozco, and Mauricio A. Álvarez. Nonlinear probabilistic latent variable models for groupwise correspondence analysis in brain structures. In *28th IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2018, Aalborg, Denmark, September 17-20, 2018*, pages 1–6, 2018. doi: 10.1109/MLSP.2018.8516934. URL https://doi.org/10.1109/MLSP.2018.8516934

- Hernán F. García, Mauricio A. Álvarez, and Álvaro Á. Orozco. Dynamic facial landmarking selection for emotion recognition using gaussian processes. *J. Multimodal User Interfaces*, 11(4):327–340, 2017. doi: 10.1007/s12193-017-0256-9. URL https://doi.org/10.1007/s12193-017-0256-9

- Hernán F. García, Mauricio A. Álvarez, and Álvaro Á. Orozco. Bayesian optimization for fitting 3d morphable models of brain structures. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - 21st Iberoamerican Congress, CIARP 2016, Lima, Peru, November 8-11, 2016, Pro-*

ceedings, pages 291–299, 2016a. doi: 10.1007/978-3-319-52277-7_36. URL https://doi.org/10.1007/978-3-319-52277-7_36

- Hernán F. García, Mauricio A. Álvarez, and Álvaro Á. Orozco. Gaussian process dynamical models for multimodal affect recognition. In *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2016, Orlando, FL, USA, August 16-20, 2016*, pages 850–853, 2016b. doi: 10.1109/EMBC.2016.7590834. URL https://doi.org/10.1109/EMBC.2016.7590834

- Hernán F. García, Mauricio A. Álvarez, and Álvaro Orozco. Bayesian shape models with shape priors for MRI brain segmentation. In *Advances in Visual Computing - 10th International Symposium, ISVC 2014, Las Vegas, NV, USA, December 8-10, 2014, Proceedings, Part II*, pages 851–860, 2014. doi: 10.1007/978-3-319-14364-4_82. URL https://doi.org/10.1007/978-3-319-14364-4_82

References

Ryan Prescott Adams and Zoubin Ghahramani. Archipelago: Nonparametric bayesian semi-supervised learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 1–8, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-516-1. doi: 10.1145/1553374.1553375. URL http://doi.acm.org/10.1145/1553374.1553375.

Yonathan Aflalo, Anastasia Dubrovina, and Ron Kimmel. Spectral generalized multi-dimensional scaling. *Int. J. Comput. Vision*, 118(3):380–392, July 2016. ISSN 0920-5691. doi: 10.1007/s11263-016-0883-8. URL http://dx.doi.org/10.1007/s11263-016-0883-8.

Dror Aiger, Niloy J. Mitra, and Daniel Cohen-Or. 4pointss congruent sets for robust pairwise surface registration. In *ACM SIGGRAPH 2008 Papers*, SIGGRAPH '08, pages 85–1 85–10, New York, NY, USA, 2008. ACM. ISBN 978-1-4503-0112-1.

Mauricio A. Álvarez, Lorenzo Rosasco, and Neil D. Lawrence. Kernels for vector-valued functions: A review. *Found. Trends Mach. Learn.*, 4(3):195–266, March 2012. ISSN 1935-8237. doi: 10.1561/2200000036. URL http://dx.doi.org/10.1561/2200000036.

Animashree Anandkumar, Daniel Hsu, Adel Javanmard, and Sham Kakade. Learning linear bayesian networks with latent variables. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 249–257, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

Davide Boscaini, Jonathan Masci, Emanuele Rodoià, and Michael Bronstein. Learning shape correspondence with anisotropic convolutional neural networks. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 3197–3205, USA, 2016. Curran Associates Inc. ISBN 978-1-5108-3881-9. URL http://dl.acm.org/citation.cfm?id=3157382.3157455.

Alexander Bronstein, Michael Bronstein, and Ron Kimmel. *Isometry Invariant Similarity*, pages 205–215. Springer New York, New York, NY, 2009. ISBN 978-0-387-73301-2. doi: 10.1007/978-0-387-73301-2_10. URL https://doi.org/10.1007/978-0-387-73301-2_10.

Alexander M. Bronstein, Michael M. Bronstein, and Ron Kimmel. Calculus of non-rigid surfaces for geometry and texture manipulation. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, pages 902–913, 2007a.

Alexander M. Bronstein, Michael M. Bronstein, and Ron Kimmel. Calculus of nonrigid surfaces for geometry and texture manipulation. *IEEE Trans. Vis. Comput. Graph.*, 13(5):902–913, 2007b. doi: 10.1109/TVCG.2007.1041. URL https://doi.org/10.1109/TVCG.2007.1041.

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. ISSN 0730-0301. doi: 10.1145/1899404.1899405. URL http://doi.acm.org/10.1145/1899404.1899405.

A.M. Bronstein, M.M. Bronstein, and R. Kimmel. *Numerical Geometry of Non-Rigid Shapes*. Monographs in Computer Science. Springer New York, 2008. ISBN 9780387733012. URL https://books.google.com.co/books?id=de2jrRCNpwEC.

M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst. Geometric deep learning: Going beyond euclidean data. *IEEE Signal Processing Magazine*, 34 (4):18–42, July 2017. ISSN 1053-5888. doi: 10.1109/MSP.2017.2693418.

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

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*, 128(0):1 – 17, 2014. ISSN 1077-3142.

Mariano Cabezas, Arnau Oliver, Xavier Lladó, Jordi Freixenet, and Meritxell Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Computer Methods and Programs in Biomedicine*, 104(3):e158 – e177, 2011. ISSN 0169-2607. doi: http://dx.doi.org/10.1016/j.cmpb.2011.07.015.

Yiqi Cai, Xiaohu Guo, Zichun Zhong, and Weihua Mao. Dynamic meshing for deformable image registration. *Computer-Aided Design*, 58(0):141 – 150, 2015. ISSN 0010-4485. doi: http://dx.doi.org/10.1016/j.cad.2014.08.009. URL http://www.sciencedirect.com/science/article/pii/S001044851400178X. Solid and Physical Modeling 2014.

Umberto Castellani, Marco Cristani, Simone Fantoni, and Vittorio Murino. Sparse points matching by combining 3d mesh saliency with statistical descriptors. *Comput. Graph. Forum*, 27(2):643–652, 2008. URL http://dblp.uni-trier.de/db/journals/cgf/cgf27.html.

Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Will Chang and Matthias Zwicker. Range scan registration using reduced deformable models. *Comput. Graph. Forum*, 28(2):447–456, 2009. URL http://dblp.uni-trier.de/db/journals/cgf/cgf28.html.

Xavier Cortés and Francesc Serratosa. An interactive method for the image alignment problem based on partially supervised correspondence. *Expert Syst. Appl.*, 42(1): 179–192, 2015. doi: 10.1016/j.eswa.2014.07.051. URL http://dx.doi.org/10.1016/j.eswa.2014.07.051.

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. doi: 10.1109/EMBC.2013.6609685.

L. Cosmo, E. Rodolà, M. M. Bronstein, A. Torsello, D. Cremers, and Y. Sahillioğlu. Partial matching of deformable shapes. In *Proceedings of the Eurographics 2016 Workshop on 3D Object Retrieval*, 3DOR '16, pages 61–67, Goslar Germany, Germany, 2016. Eurographics Association. ISBN 978-3-03868-004-8. doi: 10.2312/3dor.20161089. URL https://doi.org/10.2312/3dor.20161089.

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*, volume 70, pages 884–893. PMLR, 06–11 Aug 2017.

Andreas Damianou and Neil Lawrence. Deep Gaussian processes. In C. Carvalho and P. Ravikumar, editors, *Proceedings of the Sixteenth International Workshop on Artificial Intelligence and Statistics (AISTATS)*, AISTATS '13, pages 207–215. JMLR W&CP 31, 2013.

Andreas C. Damianou, Carl Henrik Ek, Michalis K. Titsias, and Neil D. Lawrence. Manifold relevance determination. In *ICML*. icml.cc / Omnipress, 2012. URL http://dblp.uni-trier.de/db/conf/icml/icml2012.html.

Fabio De Sousa Ribeiro, Francesco Calivá, Mark Swainson, Kjartan Gudmundsson, Georgios Leontidis, and Stefanos Kollias. Deep bayesian self-training. *Neural Computing and Applications*, Jul 2019. doi: 10.1007/s00521-019-04332-4.

Nemanja Djuric, Mihajlo Grbovic, and Slobodan Vucetic. Convex kernelized sorting. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, pages 893–899, 2012.

S. Durrleman, T Fletcher, G. Gerig, and M Niethammer, editors. *Spatio-temporal Image Analysis for Longitudinal and Time-Series Image Data*, volume 8682, 2014. Springer. URL http://link.springer.com/book/10.1007/978-3-319-14905-9.

Carl Henrik Ek, Philip H. S. Torr, and Neil D. Lawrence. Gaussian process latent variable models for human pose estimation. In *Machine Learning for Multimodal Interaction , 4th International Workshop, MLMI 2007, Brno, Czech Republic, June 28-30, 2007, Revised Selected Papers*, pages 132–143, 2007. doi: 10.1007/978-3-540-78155-4_12. URL https://doi.org/10.1007/978-3-540-78155-4_12.

Carl Henrik Ek, Jonathan Rihan, Philip H. S. Torr, Grégory Rogez, and Neil D. Lawrence. Ambiguity modeling in latent spaces. In *MLMI*, volume 5237 of *Lecture Notes in Computer Science*, pages 62–73. Springer, 2008a.

Carl Henrik Ek, Philip H. S. Torr, and Neil D. Lawrence. Gaussian process latent variable models for human pose estimation. In Andrei Popescu-Belis, Steve Renals, and Hervé Bourlard, editors, *Machine Learning for Multimodal Interaction (MLMI*

2007), volume 4892 of *LNCS*, page 132, Brno, Czech Republic, 00 2008b. Springer-Verlag. doi: 10.1007/978-3-540-78155-4_12. URL http://inverseprobability.com/publications/ek-pose07.html.

Daniel Fink. A compendium of conjugate priors, 1997.

Juergen Gall, Carsten Stoll, Edilson De Aguiar, Christian Theobalt, Bodo Rosenhahn, and Hans peter Seidel. Motion capture using joint skeleton tracking and surface estimation. In *IN IEEE CONF. ON COMPUTER VISION AND PATTERN RECOGNITION*, 2009.

Hernán F. García, Mauricio A. Álvarez, and Álvaro Orozco. Bayesian shape models with shape priors for MRI brain segmentation. In *Advances in Visual Computing - 10th International Symposium, ISVC 2014, Las Vegas, NV, USA, December 8-10, 2014, Proceedings, Part II*, pages 851–860, 2014. doi: 10.1007/978-3-319-14364-4_82. URL https://doi.org/10.1007/978-3-319-14364-4_82.

Hernán F. García, Mauricio A. Álvarez, and Álvaro Orozco. Groupwise shape correspondences on 3d brain structures using probabilistic latent variable models. In *Advances in Visual Computing - 11th International Symposium, ISVC 2015, Las Vegas, NV, USA, December 14-16, 2015, Proceedings, Part I*, volume 9474 of *Lecture Notes in Computer Science*, pages 491–500. Springer, 2015. doi: 10.1007/978-3-319-27857-5_44. URL https://doi.org/10.1007/978-3-319-27857-5_44.

Hernán F. García, Mauricio A. Álvarez, and Álvaro Á. Orozco. Bayesian optimization for fitting 3d morphable models of brain structures. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - 21st Iberoamerican Congress, CIARP 2016, Lima, Peru, November 8-11, 2016, Proceedings*, pages 291–299, 2016a. doi: 10.1007/978-3-319-52277-7_36. URL https://doi.org/10.1007/978-3-319-52277-7_36.

Hernán F. García, Mauricio A. Álvarez, and Álvaro Á. Orozco. Gaussian process dynamical models for multimodal affect recognition. In *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2016, Orlando, FL, USA, August 16-20, 2016*, pages 850–853, 2016b. doi: 10.1109/EMBC.2016.7590834. URL https://doi.org/10.1109/EMBC.2016.7590834.

Hernán F. García, Mauricio A. Álvarez, and Álvaro Á. Orozco. Dynamic facial landmarking selection for emotion recognition using gaussian processes. *J. Multimodal User Interfaces*, 11(4):327–340, 2017. doi: 10.1007/s12193-017-0256-9. URL https://doi.org/10.1007/s12193-017-0256-9.

Hernán F. García, Álvaro Ángel Orozco, and Mauricio A. Álvarez. Nonlinear probabilistic latent variable models for groupwise correspondence analysis in brain structures. In *28th IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2018, Aalborg, Denmark, September 17-20, 2018*, pages 1–6, 2018. doi: 10.1109/MLSP.2018.8516934. URL https://doi.org/10.1109/MLSP.2018.8516934.

Alan E Gelfand, Athanasios Kottas, and Steven N MacEachern. Bayesian nonparametric spatial modeling with dirichlet process mixing. *Journal of the American Statistical Association*, 100(471):1021–1035, 2005. doi: 10.1198/016214504000002078. URL https://doi.org/10.1198/016214504000002078.

Oshri Halimi, Or Litany, Emanuele Rodola, Alex M Bronstein, and Ron Kimmel. Unsupervised learning of dense shape correspondence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4370–4379, 2019.

David Hallac, Youngsuk Park, Stephen Boyd, and Jure Leskovec. Network inference via the time-varying graphical lasso. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pages 205–213, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4887-4. doi: 10.1145/3097983.3098037. URL http://doi.acm.org/10.1145/3097983.3098037.

Derek Hill. Neuroimaging to assess safety and efficacy of ad therapies. *Expert Opinion on Investigational Drugs*, 19(1):23–26, 2010. doi: 10.1517/13543780903381320. URL http://informahealthcare.com/doi/abs/10.1517/13543780903381320. PMID: 19947893.

Qi-Xing Huang, Bart Adams, Martin Wicke, and Leonidas J. Guibas. Non-rigid registration under isometric deformations. In *Proceedings of the Symposium on Geometry Processing*, SGP 08, pages 1449–1457, Aire-la-Ville, Switzerland, Switzerland, 2008. Eurographics Association. URL http://dl.acm.org/citation.cfmid=1731309.1731330.

Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2:193–218, 1985. doi: 10.1007/BF01908075.

Jung Won Hyun, Yimei Li, Chao Huang, Martin Styner, Weili Lin, and Hongtu Zhu. Stgp: Spatio-temporal gaussian process models for longitudinal neuroimaging data. *NeuroImage*, 134:550 – 562, 2016. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2016.04.023. URL http://www.sciencedirect.com/science/article/pii/S105381191630057X.

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. ISSN 0306-4573. doi: http://dx.doi.org/10.1016/j.ipm.2015.12.013.

Ali Jalali and Sujay Sanghavi. Learning the dependence graph of time series with latent factors. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, pages 619–626, USA, 2012. Omnipress. ISBN 978-1-4503-1285-1. URL http://dl.acm.org/citation.cfm?id=3042573.3042655.

Matthew Johnson, David K Duvenaud, Alex Wiltschko, Ryan P Adams, and Sandeep R Datta. Composing graphical models with neural networks for structured representations and fast inference. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 2946–2954. Curran Associates, Inc., 2016.

Evangelos Kalogerakis, Aaron Hertzmann, and Karan Singh. Learning 3d mesh segmentation and labeling. In *ACM SIGGRAPH 2010 Papers*, SIGGRAPH '10, pages 102:1–102:12, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0210-4. doi: 10.1145/1833349.1778839. URL http://doi.acm.org/10.1145/1833349.1778839.

Fahdi Kanavati, Tong Tong, Kazunari Misawa, Michitaka Fujiwara, Kensaku Mori, Daniel Rueckert, and Ben Glocker. Supervoxel classification forests for estimating pairwise image correspondences. *Pattern Recognition*, 63:561–569, 2017. doi: 10.1016/j.patcog.2016.09.026. URL https://doi.org/10.1016/j.patcog.2016.09.026.

Vladimir G. Kim, Yaron Lipman, and Thomas Funkhouser. Blended intrinsic maps. *ACM Trans. Graph.*, 30(4):79:1–79:12, July 2011. ISSN 0730-0301. doi: 10.1145/2010324.1964974. URL http://doi.acm.org/10.1145/2010324.1964974.

Oscar Kin-Chung Au, Chiew-Lan Tai, Daniel Cohen-Or, Youyi Zheng, and Hongbo Fu. Electors voting for fast automatic shape correspondence. *Computer Graphics Forum*, 29(2):645–654, 2010. doi: 10.1111/j.1467-8659.2009.01634.x.

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. URL http://jmlr.csail.mit.edu/proceedings/papers/v25/klami12.html.

Arto Klami, Seppo Virtanen, and Samuel Kaski. Bayesian canonical correlation analysis. *J. Mach. Learn. Res.*, 14(1):965–1003, April 2013. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=2502581.2502612.

Neil Lawrence. Probabilistic non-linear principal component analysis with gaussian process latent variable models. *J. Mach. Learn. Res.*, 6:1783–1816, December 2005. ISSN 1532-4435.

Neil D. Lawrence and Joaquin Quiñonero Candela. Local distance preservation in the gp-lvm through back constraints. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pages 513–520, New York, NY, USA, 2006. ACM. ISBN 1-59593-383-2.

Neil D. Lawrence and Raquel Urtasun. Non-linear matrix factorization with gaussian processes. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 601–608, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-516-1. doi: 10.1145/1553374.1553452. URL http://doi.acm.org/10.1145/1553374.1553452.

Z. Lähner, E. Rodolà, M. M. Bronstein, D. Cremers, O. Burghard, L. Cosmo, A. Dieckmann, R. Klein, and Y. Sahillioglu. Matching of Deformable Shapes with Topological Noise. In A. Ferreira, A. Giachetti, and D. Giorgi, editors, *Eurographics Workshop on 3D Object Retrieval*. The Eurographics Association, 2016. ISBN 978-3-03868-004-8. doi: 10.2312/3dor.20161088.

Hao Li, Bart Adams, Leonidas J. Guibas, and Mark Pauly. Robust single-view geometry and motion reconstruction. *ACM Trans. Graph.*, 28(5):175:1–175:10, December 2009. ISSN 0730-0301. doi: 10.1145/1618452.1618521. URL http://doi.acm.org/10.1145/1618452.1618521.

Luming Liang, Andrzej Szymczak, and Mingqiang Wei. Geodesic spin contour for partial near-isometric matching. *Computers & Graphics*, 46:156–171, 2015. doi: 10.1016/j.cag.2014.09.024. URL http://dx.doi.org/10.1016/j.cag.2014.09.024.

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. doi: 10.1016/j.media.2013.10.010. URL http://dx.doi.org/10.1016/j.media.2013.10.010.

O. Litany, E. Rodolà, A. M. Bronstein, M. M. Bronstein, and D. Cremers. Non-rigid puzzles. *Computer Graphics Forum*, 35(5):135–143, 2016. doi: 10.1111/cgf.12970.

A. Liu, X. Chen, M. J. McKeown, and Z. J. Wang. A sticky weighted regression model for time-varying resting-state brain connectivity estimation. *IEEE Transactions on Biomedical Engineering*, 62(2):501–510, Feb 2015. ISSN 0018-9294. doi: 10.1109/TBME.2014.2359211.

Jun S. Liu. The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89 (427):958–966, 1994. doi: 10.1080/01621459.1994.10476829. URL https://doi.org/10.1080/01621459.1994.10476829.

Alexander Selvikvåg Lundervold and Arvid Lundervold. An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, 29(2): 102 – 127, 2019. ISSN 0939-3889. doi: https://doi.org/10.1016/j.zemedi.2018.11.002. URL http://www.sciencedirect.com/science/article/pii/S0939388918301181. Special Issue: Deep Learning in Medical Physics.

David J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, New York, NY, USA, 2002. ISBN 0521642981.

Jonathan Masci, Davide Boscaini, Michael M. Bronstein, and Pierre Vandergheynst. Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, ICCVW '15, pages 832–840, Washington, DC, USA, 2015. IEEE Computer Society. ISBN 978-1-4673-9711-7. doi: 10.1109/ICCVW.2015.112. URL http://dx.doi.org/10.1109/ICCVW.2015.112.

Geoffrey J. McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions*. Wiley series in probability and statistics. Wiley, Hoboken, NJ, 2. ed edition, 2008. ISBN 978-0-471-20170-0. URL http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+52983362X&sourceid=fbw_bibsonomy.

Laurence Mercier, Rolando F. Del Maestro, Kevin Petrecca, David Araujo, Claire Haegelen, and D. Louis Collins. Online database of clinical mr and ultrasound images of brain tumors. *Medical Physics*, 39(6):3253–3261, 2012. doi: http://dx.doi.org/10.1118/1.4709600. URL http://scitation.aip.org/content/aapm/journal/medphys/39/6/10.1118/1.4709600.

Phuc Ngo, Yukiko Kenmochi, Akihiro Sugimoto, Hugues Talbot, and Nicolas Passat. Discrete rigid registration: A local graph-search approach. *Discrete Applied Mathematics*, 216, Part 2:461 – 481, 2017. ISSN 0166-218X. doi: https://doi.org/10.1016/j.dam.2016.05.005. URL http://www.sciencedirect.com/science/article/pii/S0166218X16302220. Graph-theoretic and Polyhedral Combinatorics Issues and Approaches in Imaging Sciences.

Francisco P.M. Oliveira and João Manuel R.S. Tavares. Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2): 73–93, 2014. doi: 10.1080/10255842.2012.670855. URL http://dx.doi.org/10.1080/10255842.2012.670855. PMID: 22435355.

Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas Guibas. Functional maps: A flexible representation of maps between shapes. *ACM Trans. Graph.*, 31(4):30:1–30:11, July 2012. ISSN 0730-0301. doi: 10.1145/2185520.2185526. URL http://doi.acm.org/10.1145/2185520.2185526.

Rutuparna Panda, Sanjay Agrawal, Madhusmita Sahoo, and Rajashree Nayak. A novel evolutionary rigid body docking algorithm for medical image registration. *Swarm and Evolutionary Computation*, 33:108 – 118, 2017. ISSN 2210-6502. doi: https://doi.org/10.1016/j.swevo.2016.11.002. URL http://www.sciencedirect.com/science/article/pii/S2210650216304011.

Pekelny, Yuri, Gotsman, and Craig. Articulated object reconstruction and markerless motion capture from depth video. *Computer Graphics Forum*, 27(2):399–408, apr 2008. ISSN 0167-7055. doi: 10.1111/j.1467-8659.2008.01137.x.

Joshua Podolak, Philip Shilane, Aleksey Golovinskiy, Szymon Rusinkiewicz, and Thomas Funkhouser. A planar-reflective symmetry transform for 3d shapes. *ACM Trans. Graph.*, 25(3):549–559, July 2006. ISSN 0730-0301.

Novi Quadrianto, Le Song, and Alex J. Smola. Kernelized sorting. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 1289–1296. Curran Associates, Inc., 2009. URL http://papers.nips.cc/paper/3608-kernelized-sorting.pdf.

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

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN 026218253X.

Emanuele Rodolà, Samuel Rota Bulò, Thomas Windheuser, Matthias Vestner, and Daniel Cremers. Dense non-rigid shape correspondence using random forests. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, pages 4177–4184, Washington, DC, USA, 2014. IEEE Computer Society. ISBN 978-1-4799-5118-5. doi: 10.1109/CVPR.2014.532. URL https://doi.org/10.1109/CVPR.2014.532.

E. Rodolí, L. Cosmo, M. M. Bronstein, A. Torsello, and D. Cremers. Partial functional correspondence. *Comput. Graph. Forum*, 36(1):222–236, January 2017. ISSN 0167-7055. doi: 10.1111/cgf.12797. URL https://doi.org/10.1111/cgf.12797.

Y. Sahillioglu and Y. Yemez. Scale normalization for isometric shape matching. *Comput. Graph. Forum*, 31(7pt2):2233–2240, September 2012. ISSN 0167-7055. doi: 10.1111/j.1467-8659.2012.03216.x. URL http://dx.doi.org/10.1111/j.1467-8659.2012.03216.x.

Alsiagy A. Salama, Reda A. Alarabawy, W. El-shehaby, D. El-amrousy, Michael S. Baghdadi, and Mina F. Rizkallah. Brain volumetrics, regional cortical thickness and radiographic findings in children with cyanotic congenital heart disease using quantitative magnetic resonance imaging. *The Egyptian Journal of Radiology and Nuclear Medicine*, pages –, 2016. ISSN 0378-603X. doi: http://dx.doi.org/10.1016/j.ejrnm.2016.09.003. URL http://www.sciencedirect.com/science/article/pii/S0378603X1630170X.

Mathieu Salzmann, Carl Henrik Ek, Raquel Urtasun, and Trevor Darrell. Factorized orthogonal latent spaces. In Yee Whye Teh and Mike Titterington, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 701–708, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. PMLR. URL http://proceedings.mlr.press/v9/salzmann10a.html.

Bernhard Schölkopf, Alexander J. Smola, and Klaus-Robert Müller. Advances in kernel methods. chapter Kernel Principal Component Analysis, pages 327–352. MIT Press, Cambridge, MA, USA, 1999. ISBN 0-262-19416-3. URL http://dl.acm.org/citation.cfm?id=299094.299113.

Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4: 639–650, 1994.

Andrei Sharf, Dan A. Alcantara, Thomas Lewiner, Chen Greif, Alla Sheffer, Nina Amenta, and Daniel Cohen-Or. Space-time surface reconstruction using incompressible flow. *ACM Trans. Graph.*, 27(5):110:1–110:10, December 2008. ISSN 0730-0301. doi: 10.1145/1409060.1409063. URL http://doi.acm.org/10.1145/1409060.1409063.

Jie Shi, Paul M. Thompson, Boris Gutman, and Yalin Wang. Surface fluid registration of conformal representation: Application to detect disease burden and genetic influence on hippocampus. *NeuroImage*, 78:111 – 134, 2013. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2013.04.018. URL http://www.sciencedirect.com/science/article/pii/S1053811913003455.

Aaron P. Shon, Keith Grochow, Aaron Hertzmann, and Rajesh P. N. Rao. Learning shared latent structure for image synthesis and robotic imitation. In *NIPS*, pages 1233–1240, 2005.

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. ISBN 978-1-4577-0394-2. doi: 10.1109/CVPR.2011.5995632. URL http://dx.doi.org/10.1109/CVPR.2011.5995632.

C. Sjoberg and A. Ahnesjo. Multi-atlas based segmentation using probabilistic label fusion with adaptive weighting of image similarity measures. *Computer Methods and Programs in Biomedicine*, 110(3):308 – 319, 2013. ISSN 0169-2607. doi: http://dx.doi.org/10.1016/j.cmpb.2012.12.006. URL http://www.sciencedirect.com/science/article/pii/S0169260713000023.

Hock Wei Soon and Anqi Qiu. Individualized diffeomorphic mapping of brains with large cortical infarcts. *Magnetic Resonance Imaging*, 33(1):110 – 123, 2015. ISSN 0730-725X. doi: http://dx.doi.org/10.1016/j.mri.2014.09.001. URL http://www.sciencedirect.com/science/article/pii/S0730725X14002896.

A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: A survey. *Medical Imaging, IEEE Transactions on*, 32(7):1153–1190, July 2013. ISSN 0278-0062. doi: 10.1109/TMI.2013.2265603.

Petros Stylianou, Chen Hoffmann, Ilan Blat, and Sagi Harnof. Neuroimaging for patient selection for medial temporal lobe epilepsy surgery: Part 1 structural neuroimaging. *Journal of Clinical Neuroscience*, 23:14 – 22, 2016. ISSN 0967-5868. doi: https://doi.org/10.1016/j.jocn.2015.04.019. URL http://www.sciencedirect.com/science/article/pii/S0967586815003525.

Zhengyu Su, Wei Zeng, Yalin Wang, Zhong Lin Lu, and Xianfeng Gu. Shape classification using wasserstein distance for brain morphometry analysis. *Lecture Notes in Computer Science*, 9123:411–423, 2015. ISSN 0302-9743. doi: 10.1007/978-3-319-19992-4_32.

Jochen Süssmuth, Marco Winter, and Günther Greiner. Reconstructing animated meshes from time-varying point clouds. In *Proceedings of the Symposium on Geometry Processing*, SGP '08, pages 1469–1476, Aire-la-Ville, Switzerland, Switzerland, 2008. Eurographics Association. URL http://dl.acm.org/citation.cfm?id=1731309.1731332.

G. K. L. Tam, Zhi-Quan Cheng, Yu-Kun Lai, F. C. Langbein, Yonghuai Liu, D. Marshall, R. R. Martin, Xian-Fang Sun, and P. L. Rosin. Registration of 3d point clouds and meshes: A survey from rigid to nonrigid. *IEEE Transactions on Visualization and Computer Graphics*, 19(7):1199–1217, 2013. ISSN 1077-2626. doi: http://doi.ieeecomputersociety.org/10.1109/TVCG.2012.310.

Qiaoyu Tan, Ninghao Liu, and Xia Hu. Deep representation learning for social network analysis. *Frontiers in Big Data*, 2:2, 2019. ISSN 2624-909X. doi: 10.3389/fdata.2019.00002. URL https://www.frontiersin.org/article/10.3389/fdata.2019.00002.

Wanjie Tang, Bin Li, Xiaoqi Huang, Xiaoyu Jiang, Fei Li, Lijuan Wang, Taolin Chen, Jinhui Wang, Qiyong Gong, and Yanchun Yang. Morphometric brain characterization of refractory obsessive–compulsive disorder: Diffeomorphic anatomic registration using exponentiated lie algebra. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 46:126 – 131, 2013. ISSN 0278-5846. doi: https://doi.org/10.

1016/j.pnpbp.2013.07.011. URL http://www.sciencedirect.com/science/article/pii/S0278584613001528.

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.

Zoubin Ghahramani Tomoharu Iwata, David Duvenaud. Warped mixtures for nonparametric cluster shapes. In *29th Conference on Uncertainty in Artificial Intelligence*, pages 311–319, 2013.

Andrea Torsello. A game-theoretic approach to deformable shape matching. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR '12, pages 182–189, Washington, DC, USA, 2012. IEEE Computer Society. ISBN 978-1-4673-1226-4. URL http://dl.acm.org/citation.cfm?id=2354409.2354702.

Mark Tygert, Joan Bruna, Soumith Chintala, Yann LeCun, Serkan Piantino, and Arthur Szlam. A mathematical motivation for complex-valued convolutional networks. *Neural Comput.*, 28(5):815–825, May 2016. ISSN 0899-7667. doi: 10.1162/NECO_a_00824. URL http://dx.doi.org/10.1162/NECO_a_00824.

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.

M. Vestner, Z. Lähner, A. Boyarski, O. Litany, R. Slossberg, T. Remez, E. Rodola, A. Bronstein, M. Bronstein, R. Kimmel, and D. Cremers. Efficient deformable shape correspondence via kernel matching. In *2017 International Conference on 3D Vision (3DV)*, pages 517–526, Oct 2017. doi: 10.1109/3DV.2017.00065.

Lingfeng Wang and Chunhong Pan. Nonrigid medical image registration with locally linear reconstruction. *Neurocomputing*, 145:303–315, 2014. ISSN 0925-2312. doi: http://dx.doi.org/10.1016/j.neucom.2014.05.030.

Chong-Yaw Wee, Chaoqiang Liu, Annie Lee, Joann S. Poh, Hui Ji, and Anqi Qiu. Cortical graph neural network for ad and mci diagnosis and transfer learning across populations. *NeuroImage: Clinical*, 23:101929, 2019. ISSN 2213-1582. doi: https://doi.org/10.1016/j.nicl.2019.101929. URL http://www.sciencedirect.com/science/article/pii/S2213158219302797.

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. ISSN 0022-3476. doi: 10.1016/j.jpeds.2017.09.043.

Carol P. Weingarten, Mark H. Sundman, Patrick Hickey, and Nan kuei Chen. Neuroimaging of parkinson's disease: Expanding views. *Neuroscience and Biobehavioral Reviews*, 59:16 – 52, 2015. ISSN 0149-7634. doi: https://doi.org/10.1016/j.neubiorev.2015.09.007. URL http://www.sciencedirect.com/science/article/pii/S0149763415002420.

David A. Wolk, Sandhitsu R. Das, Susanne G. Mueller, Michael W. Weiner, and Paul A. Yushkevich. Medial temporal lobe subregional morphometry using high resolution {MRI} in alzheimer's disease. *Neurobiology of Aging*, 49:204 – 213, 2017. ISSN 0197-4580. doi: https://doi.org/10.1016/j.neurobiolaging.2016.09.011. URL http://www.sciencedirect.com/science/article/pii/S019745801630224X.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 1912–1920, 2015. doi: 10.1109/CVPR.2015.7298801. URL https://doi.org/10.1109/CVPR.2015.7298801.

Di Xiao, David Zahra, Pierrick Bourgeat, Paula Berghofer, Oscar Acosta Tamayo, Catriona Wimberley, Marie Claude Gregoire, and Olivier Salvado. An improved 3d shape context based non-rigid registration method and its application to small animal skeletons registration. *Computerized Medical Imaging and Graphics*, 34(4):321 – 332, 2010. ISSN 0895-6111. doi: http://dx.doi.org/10.1016/j.compmedimag.2009.12.003. URL http://www.sciencedirect.com/science/article/pii/S0895611109001414.

Makoto Yamada and Masashi Sugiyama. Cross-domain object matching with model selection. In Geoffrey J. Gordon and David B. Dunson, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS-11)*, volume 15, pages 807–815. Journal of Machine Learning Research - Workshop and Conference Proceedings, 2011. URL http://www.jmlr.org/proceedings/papers/v15/yamada11a/yamada11a.pdf.

Xu Yang, Hong Qiao, and Zhi-Yong Liu. Partial correspondence based on subgraph matching. *Neurocomputing*, 122:193 – 197, 2013. ISSN 0925-2312. Advances in cognitive and ubiquitous computingSelected papers from the Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2012).

Hao Zhang, Alla Sheffer, Daniel Cohen-Or, Qingnan Zhou, Oliver van Kaick, and Andrea Tagliasacchi. Deformation-driven shape correspondence. *Computer Graphics Forum (Special Issue of Symposium on Geometry Processing 2008)*, 27(5), 2008.

Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43 – 54, 2017. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2017.02.007.

Qian Zheng, Andrei Sharf, Andrea Tagliasacchi, Baoquan Chen, Hao Zhang, Alla Sheffer, and Daniel Cohen Or. Consensus skeleton for non-rigid space-time registration. *Computer Graphcis Forum (Special Issue of Eurographics)*, 29(2):635–644, 2010.