

# Multi-view Warped Mixtures for Shape Correspondence Analysis

Hernán F. García and Mauricio A. Álvarez

June 11, 2019

## Abstract

Probabilistic approach for the correspondence problem based on the Single View WMM[4].

## 1 Multiview Warped Mixture Models

We can use the single view problem of the warped mixture model from [4] in which they warp a latent mixture of Gaussians into nonparametric cluster shapes. The low-dimensional latent mixture model summarizes the properties of the high-dimensional density manifolds describing the data.

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

### 1.1 The model

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

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

Based on the iWMM (see [4]) our generative model generates multiple observations  $\mathbf{Y}^V$  according to the following generative process:

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

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

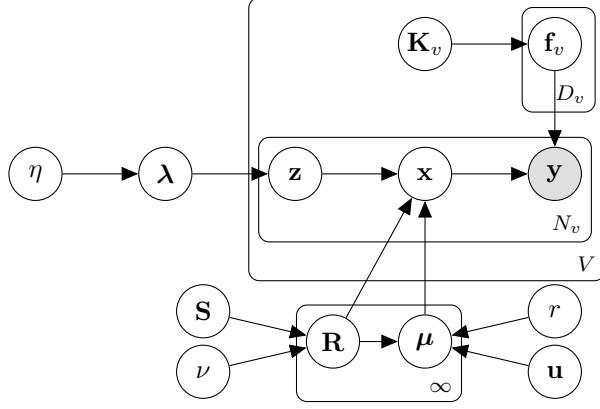


Figure 1: A graphical model representation of the multiview warped mixture model, where the shaded and unshaded nodes indicate observed and latent variables, respectively, and plates indicate repetition.

- ii. For each observation  $n = 1, \dots, N_v$
- A. Draw feature  $y_{nd}^v \sim \mathcal{N}(\mathbf{w}_d^v f_d^v(\mathbf{x}_{nv}), \beta^{-1})$

We define the dimensionalities of our variables as:

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

## 1.2 Latent Multi-view Warped Mixture Model

Our model is set as a multi-view Gaussian Process Latent Variable model as [3]. First we assume that observations are generated by mapping the latent coordinates through a set of smooth functions, over which Gaussian process priors are placed. Under the GPLVM, the probability of observations given the latent coordinates, integrating out the mapping functions, is defined as

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

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

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

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

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

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

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

As in the iWMM [4], we place Gaussian-Wishart priors on the Gaussian parameters  $\{\boldsymbol{\mu}_c, \mathbf{R}_c\}$ .

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

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

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

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

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

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

are the posterior Gaussian-Wishart parameters of the  $c$ th component. As in [4], we use a Dirichlet process with concentration parameter  $\eta$  for infinite mixture modeling in the latent space.

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

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

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

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

### 1.3 Inference

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

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

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

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

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

## 1.4 Regularized WMM

Since we want to constrain the  $\mathbf{w}^v$

## 2 Results

### 2.1 Clustering performance on real datasets

Table 1: Average Rand index for evaluating clustering performance.

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

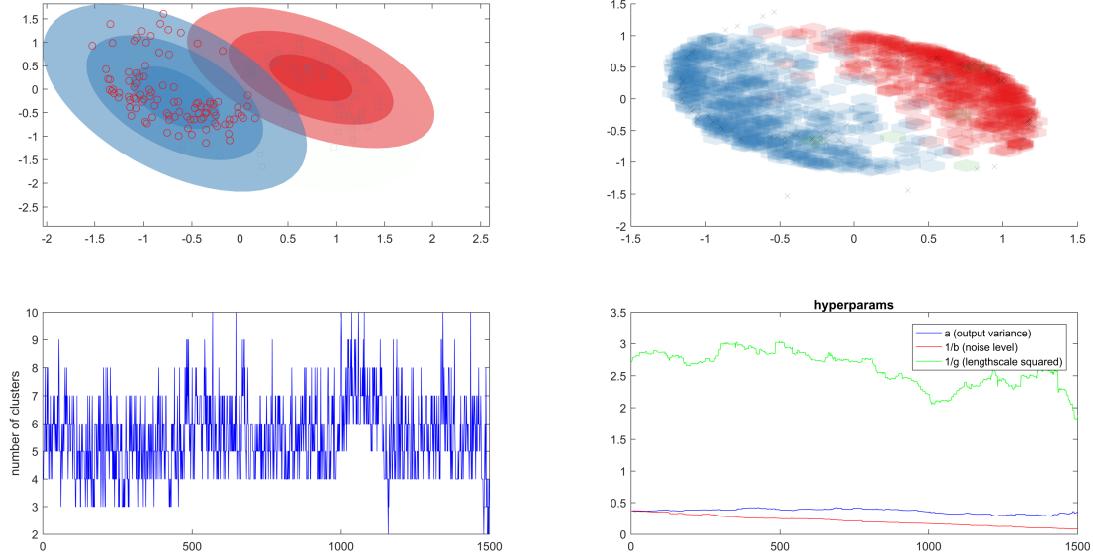


Figure 2: Experimental results for the Pinwheel dataset. Current mixture parameters, along with the latent positions (Top left). Original data and predicted assignments (Top right).

### 2.2 Comparison with linear approaches

First, 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 [1] and the true labels. For comparison, we use unsupervised clustering matching (UCM) [1], k-means (KM), and convex kernelized sorting (CKS) [?]. Table 2 shows that our approaches outperforms the state-of-the-art methods for unsupervised clustering for the three databases. The results also show that by mapping the observed data through random feature expansions, the model can handle real-world datasets with better performance than linear approaches as in the case of NL-UCM (i.e., 0.17 for the MNIST dataset against 0.085 obtained from the UCM method).

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

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

### 2.3 Non-rigid 3D shape datasets

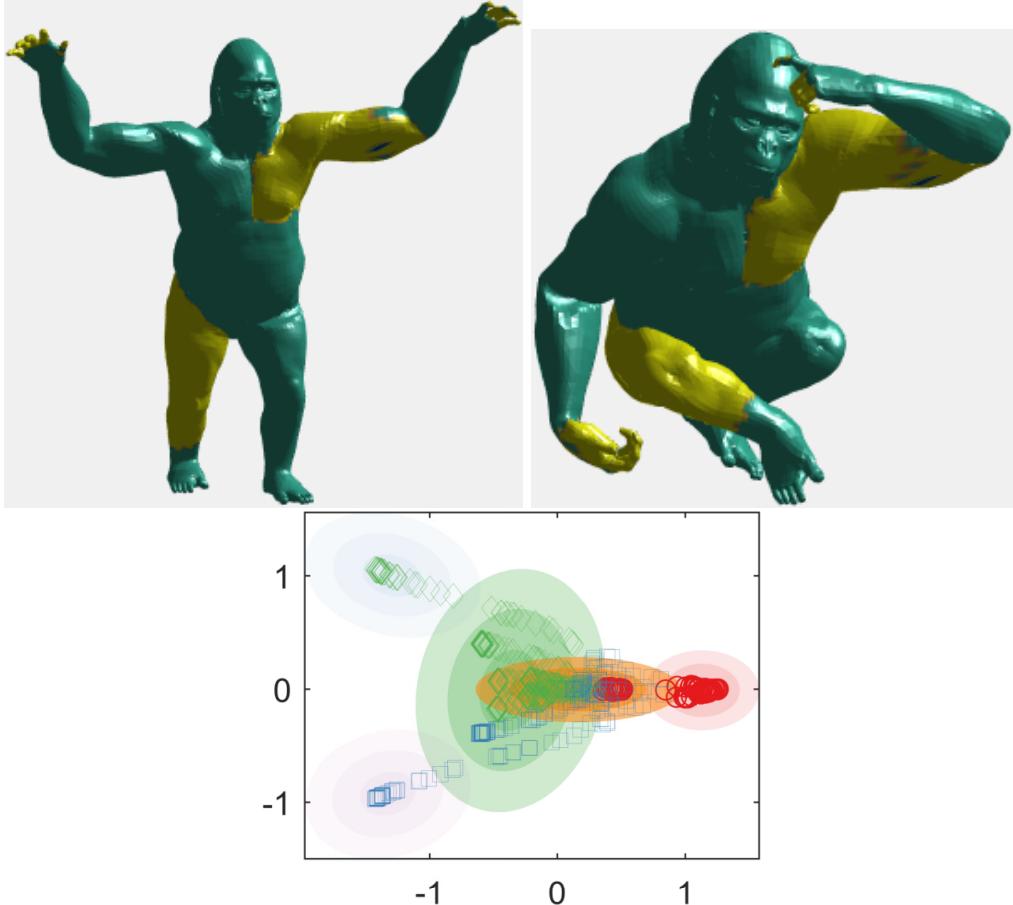


Figure 3: Experimental results for the TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.6487

### 2.4 Deformable 3D Shapes with Topological Noise

In this section, we test our model with KIDS dataset [2]. This data consists of a collection of 3D shapes undergoing within-class deformations that include topological changes. The changes simulate coalescence of spatially close surface regions – a scenario that frequently occurs when dealing with real data under suboptimal acquisition conditions. The dataset is based on the fat kid from the KIDS dataset with additional poses.

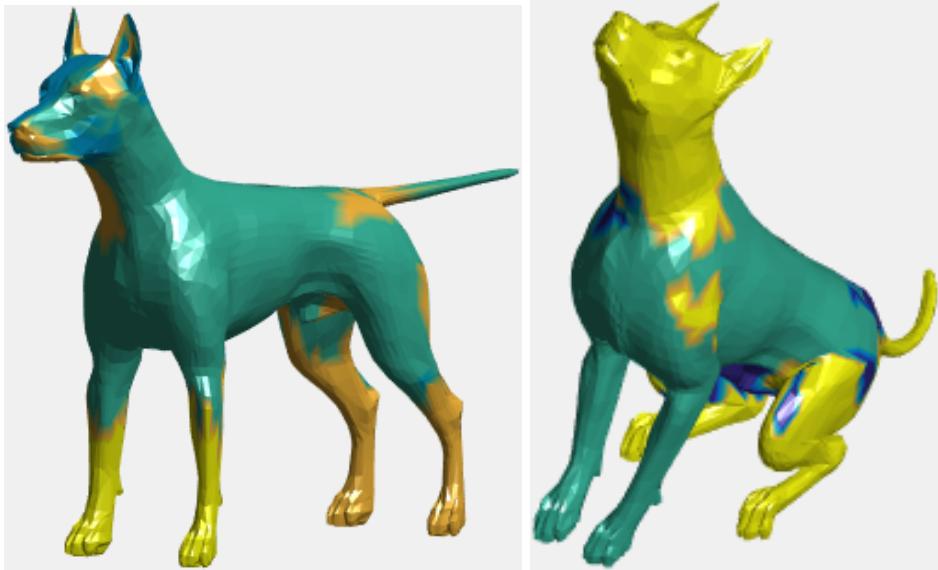


Figure 4: Experimental results for the TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.5894

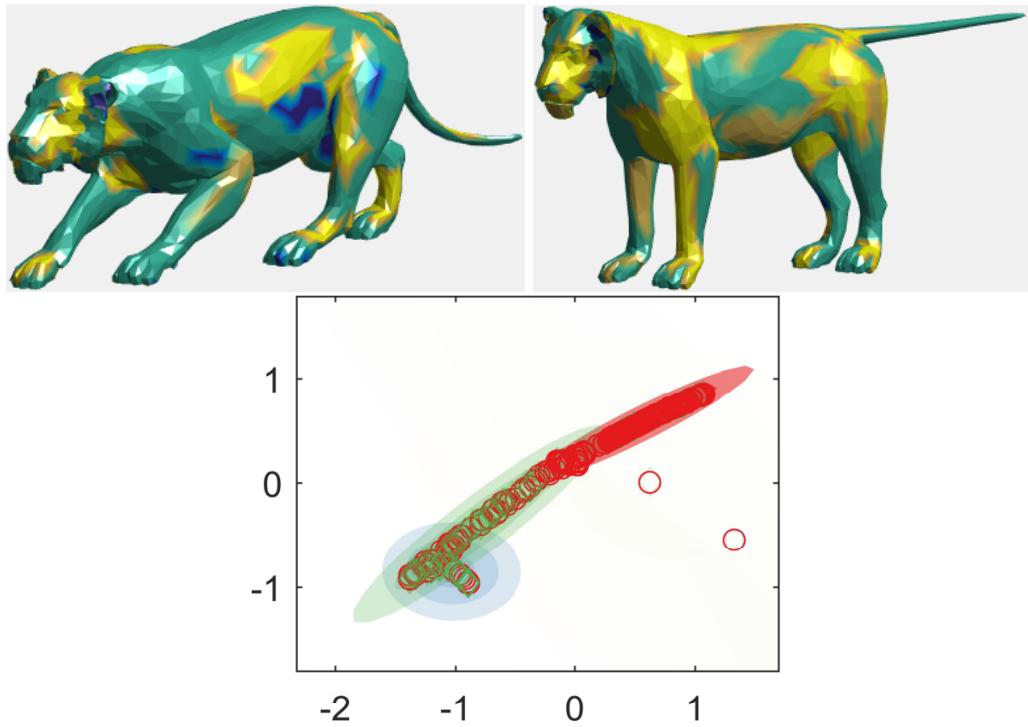


Figure 5: Experimental results for the TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.6317

## 2.5 Partial matching

### 2.5.1 Shapes undergoing a single cut

### 2.5.2 Shapes with irregular holes

Finally we test our method on *holes dataset*. This dataset contains irregular holes and multiple cuts for different shapes.

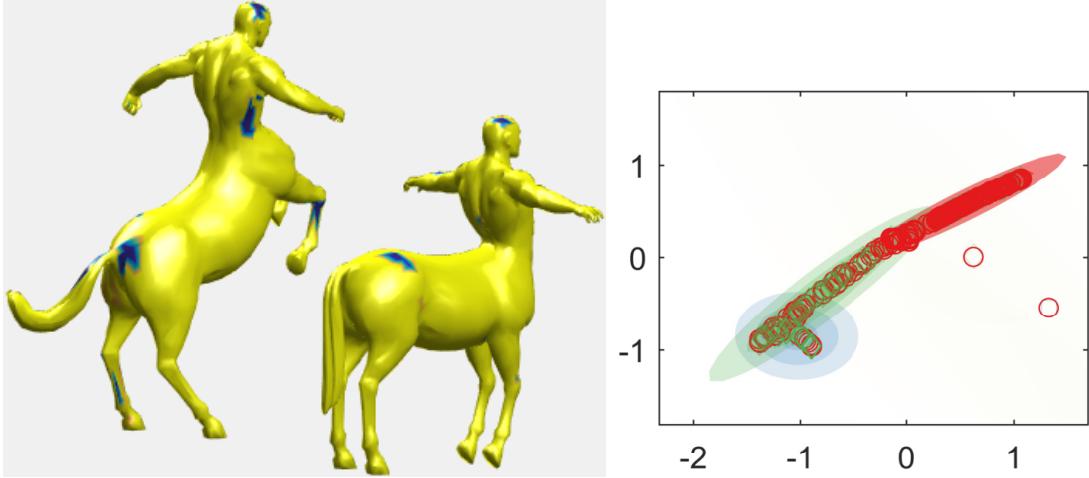


Figure 6: Experimental results for the TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.6232

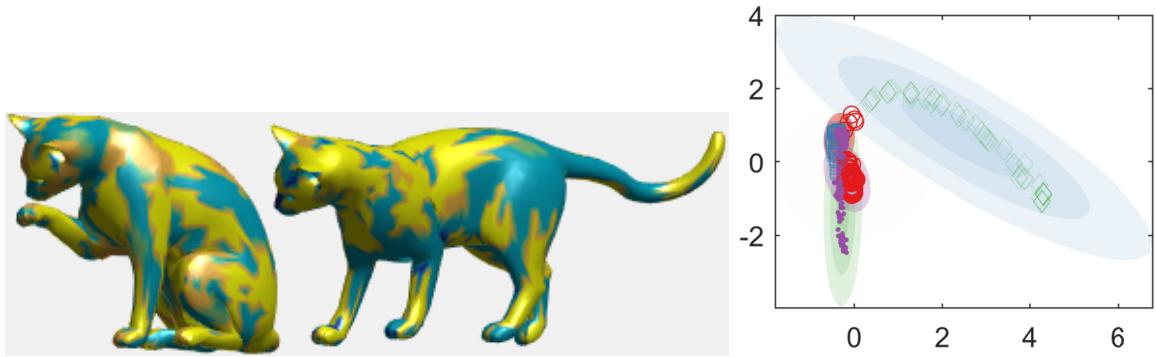


Figure 7: Experimental results for the TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.6921

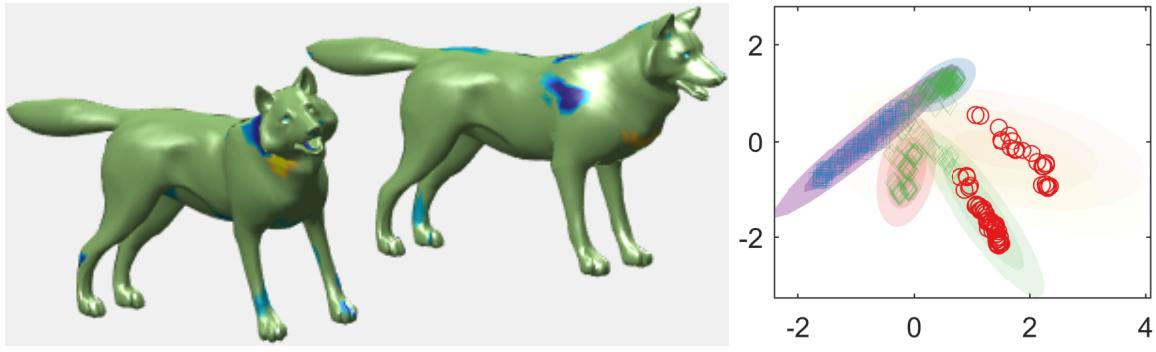


Figure 8: Experimental results for the TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.8073

### 3 Brain structure dataset

### References

- [1] 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.
- [2] Z. L'Shrec'16: Matching of deformable shapes with topological noise.

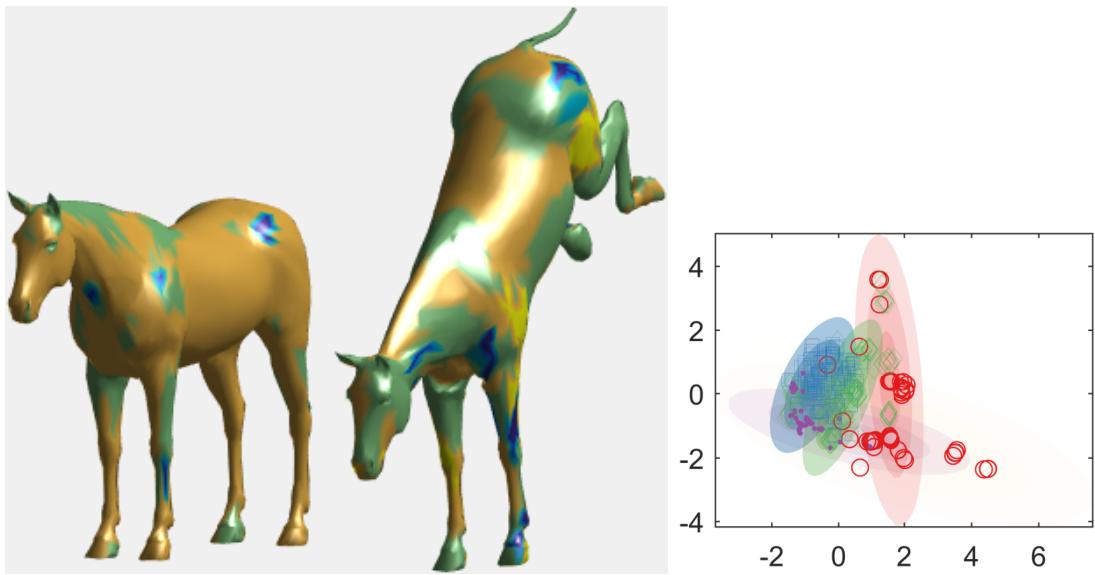


Figure 9: Experimental results for the TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.6576

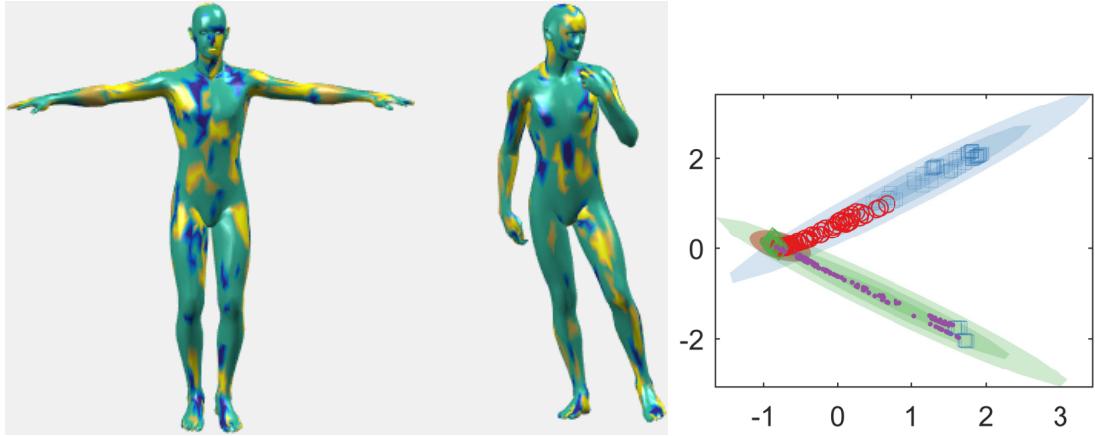


Figure 10: Experimental results for the TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.7049

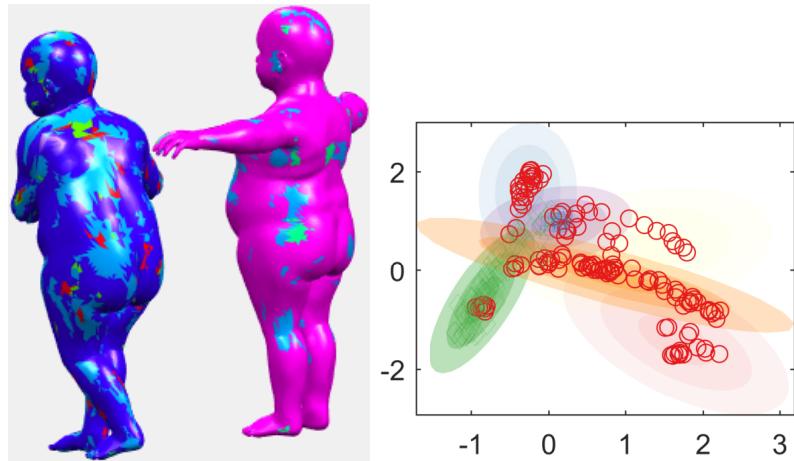


Figure 11: Experimental results for the KIDS dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.7003

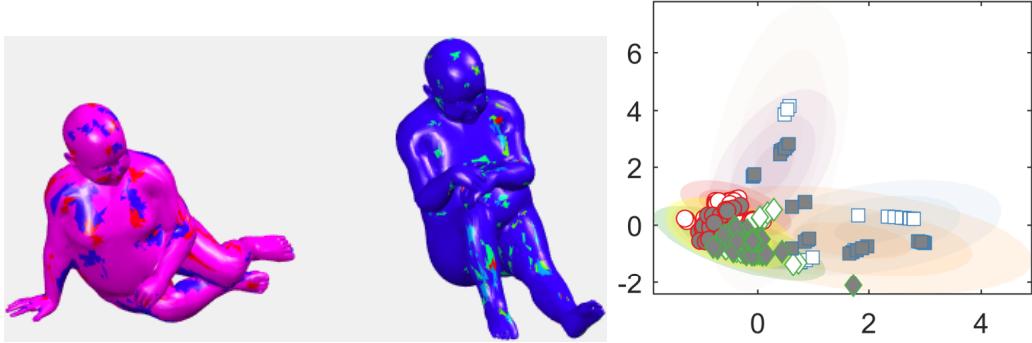


Figure 12: Experimental results for the KIDS dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.7432

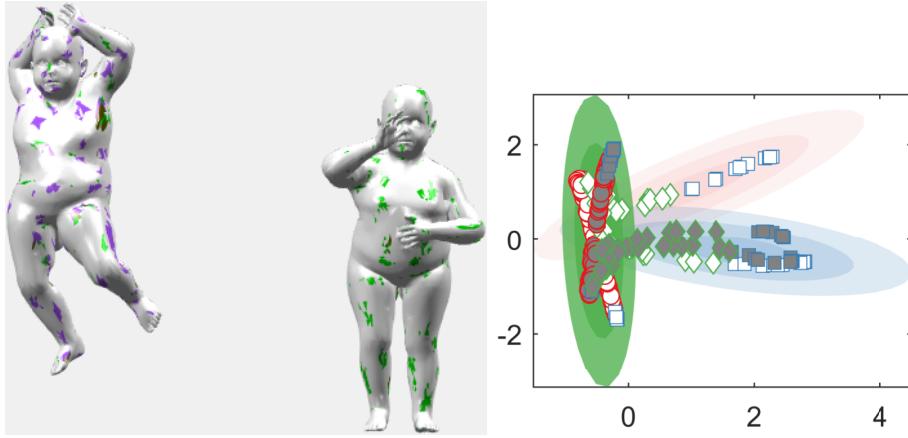


Figure 13: Experimental results for the KIDS dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.7179

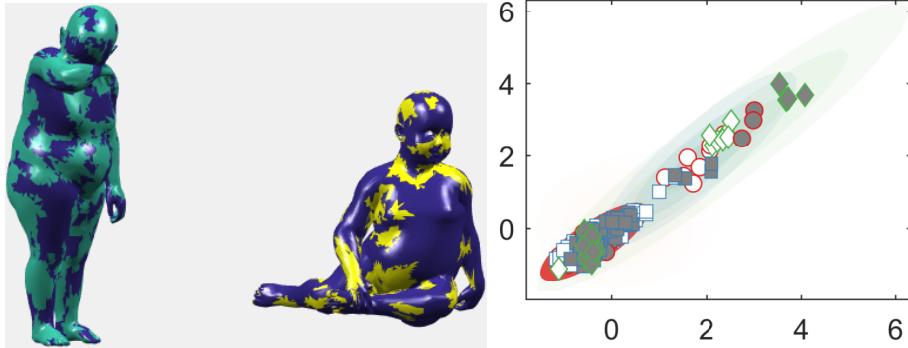


Figure 14: Experimental results for the KIDS dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different poses. Average Rand Index 0.6557

- [3] Neil Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In *In NIPS*, page 2004, 2003.
- [4] Zoubin Ghahramani Tomoharu Iwata, David Duvenaud. Warped mixtures for nonparametric cluster shapes. In *29th Conference on Uncertainty in Artificial Intelligence*, pages 311–319, 2013.

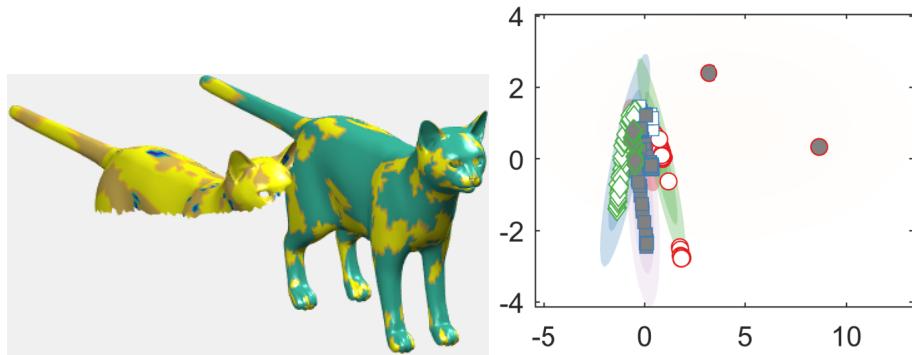


Figure 15: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the shape. Average Rand Index 0.8016

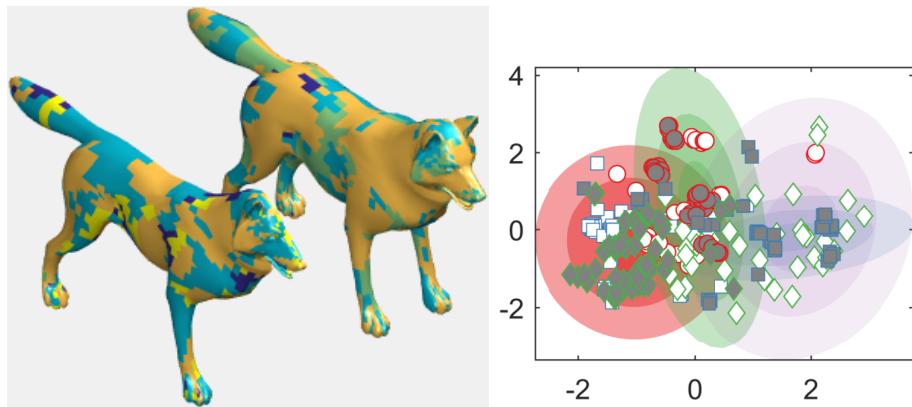


Figure 16: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the shape. Average Rand Index 0.6365

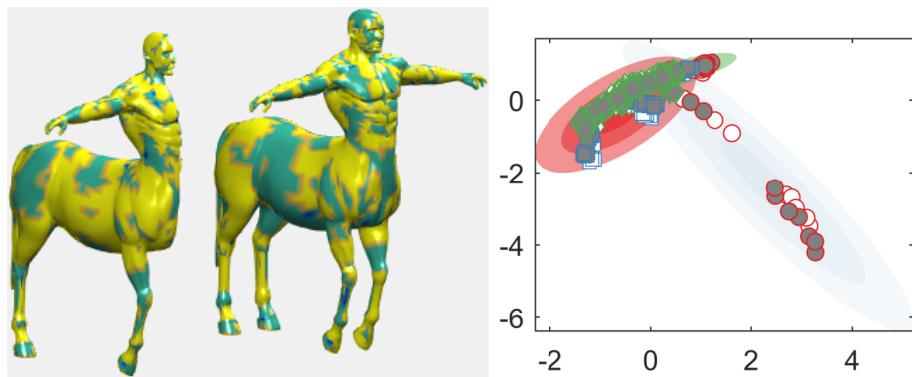


Figure 17: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the shape. Average Rand Index 0.5838

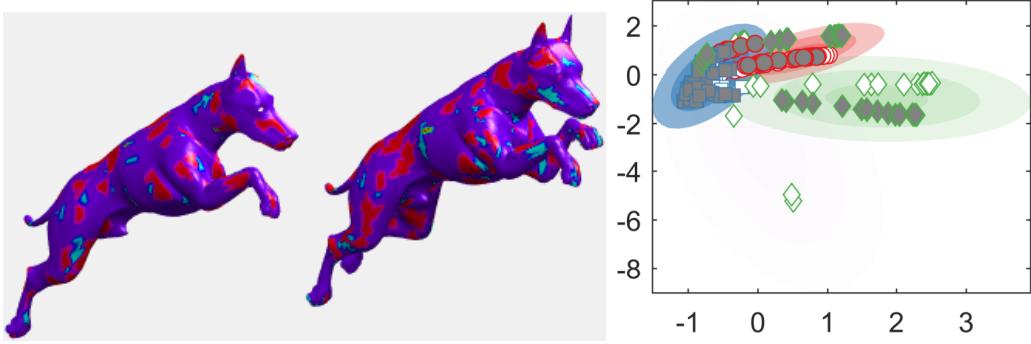


Figure 18: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the shape. Average Rand Index 0.7744

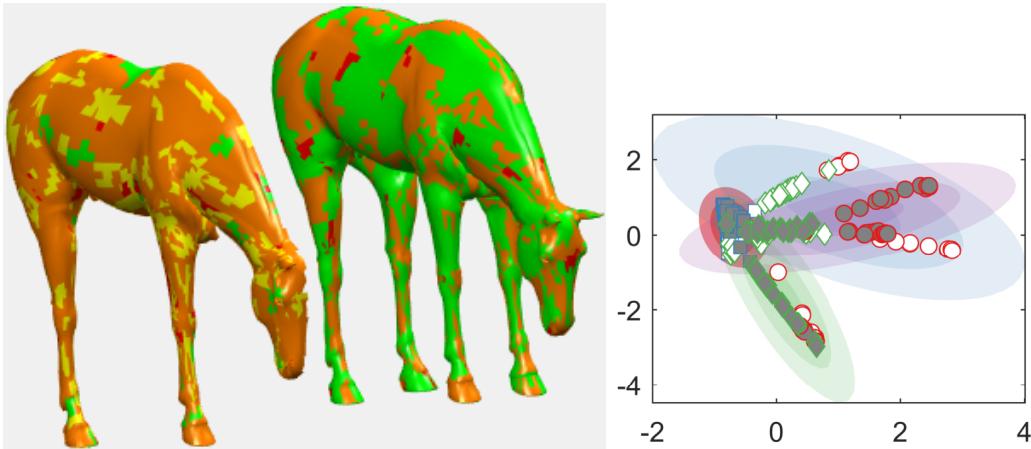


Figure 19: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the shape. Average Rand Index 0.7931

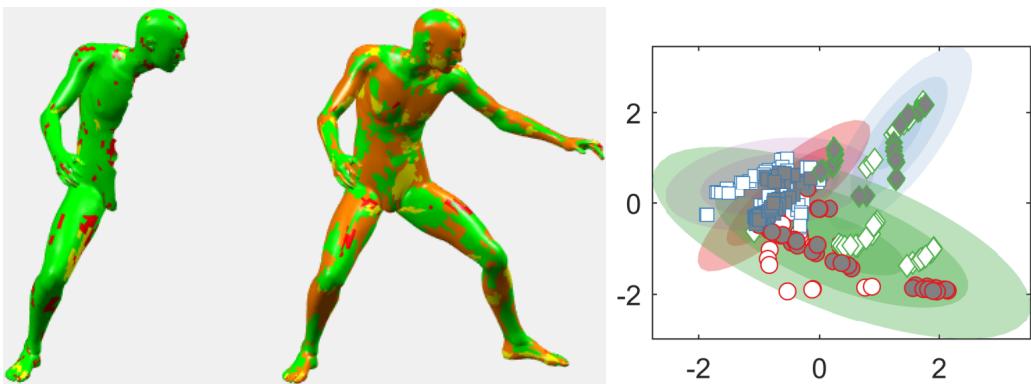


Figure 20: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the shape. Average Rand Index 0.6566

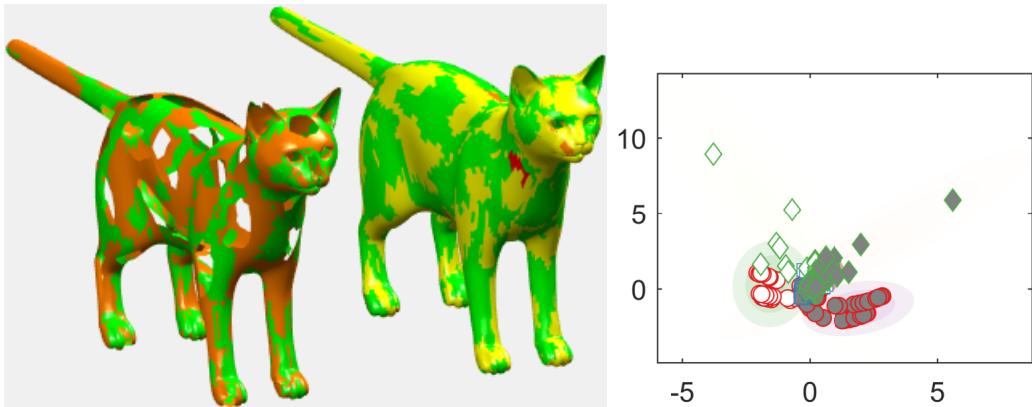


Figure 21: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the shape. Average Rand Index 0.6568

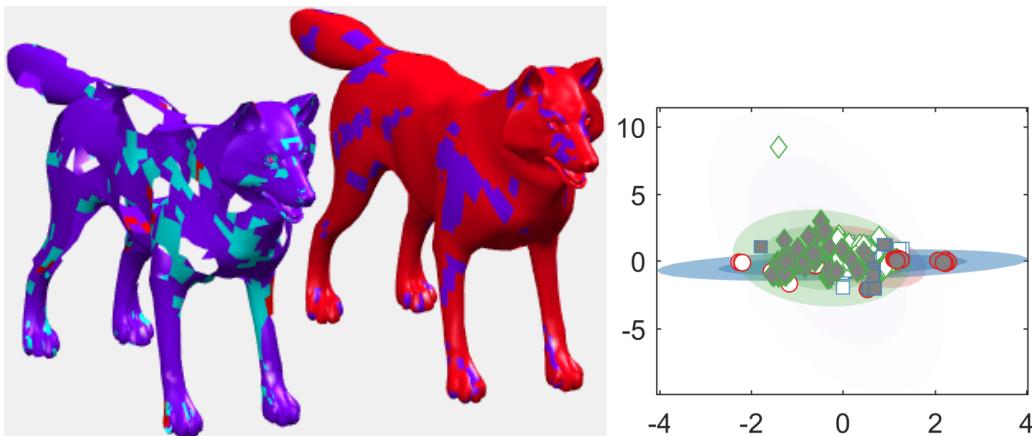


Figure 22: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the shape. Average Rand Index 0.5505

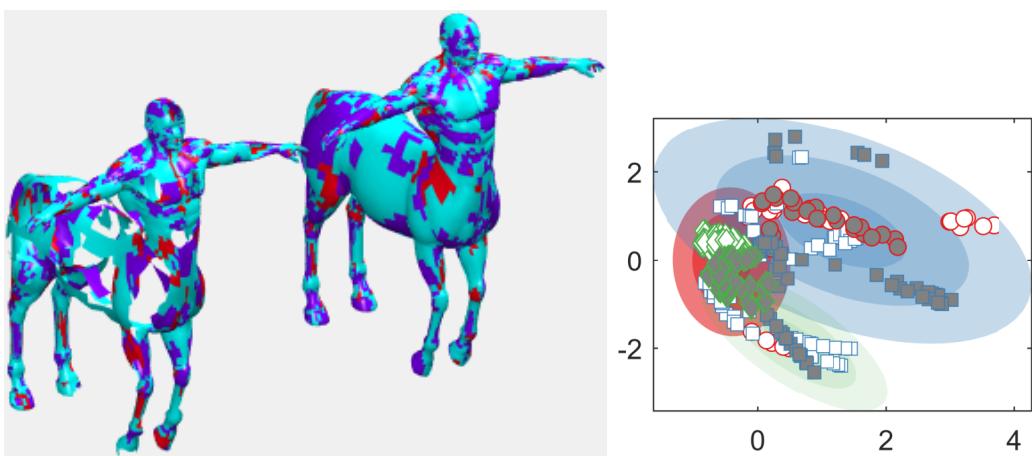


Figure 23: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the shape. Average Rand Index 0.6849

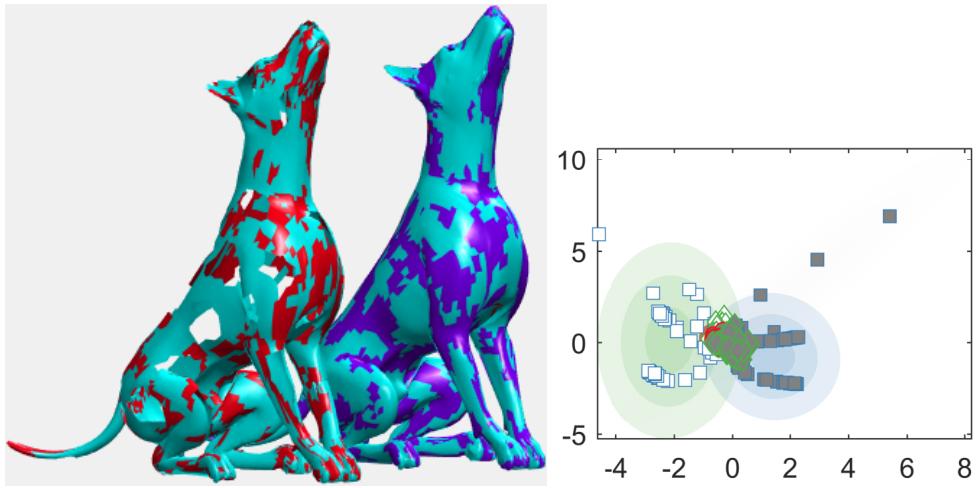


Figure 24: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the shape. Average Rand Index 0.5991

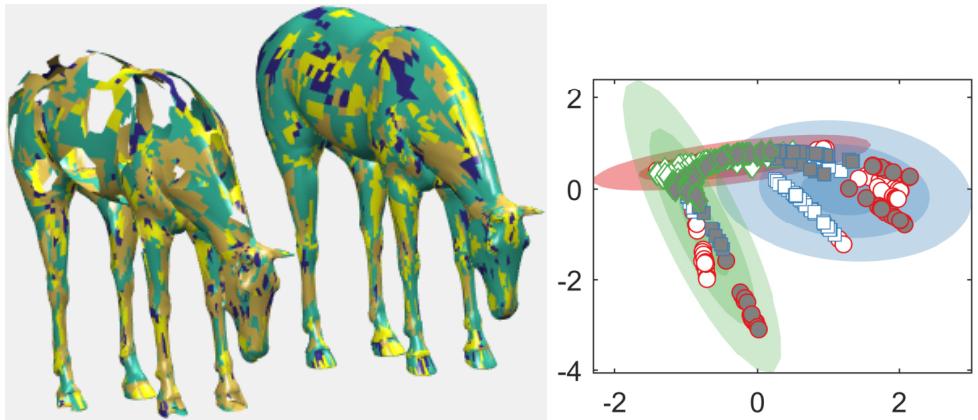


Figure 25: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the shape. Average Rand Index 0.6569

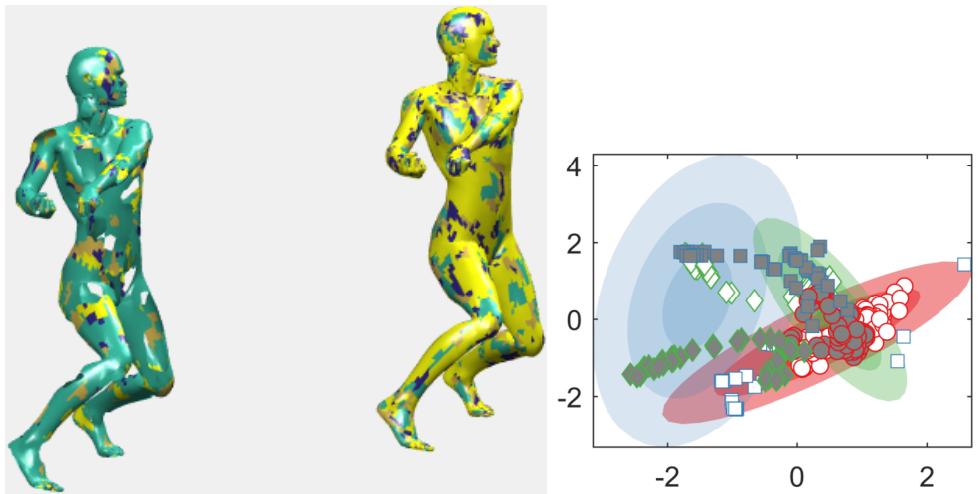


Figure 26: Experimental results for the partial matching TOSCA dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the shape. Average Rand Index 0.6107

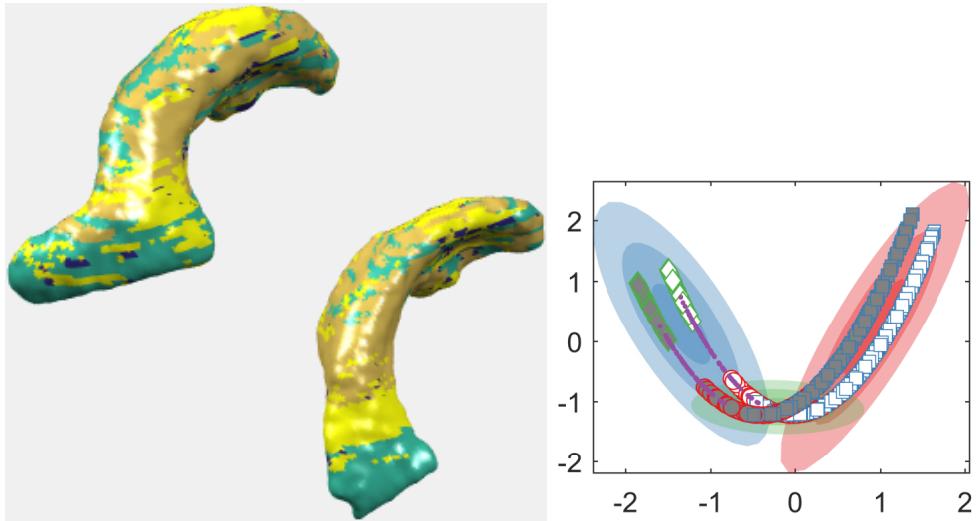


Figure 27: Experimental results for the brain structures dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the ventricle. Average Rand Index 0.6071

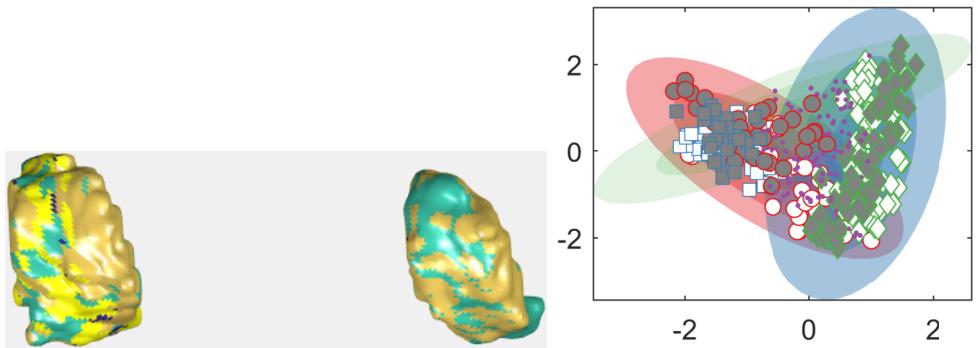


Figure 28: Experimental results for the brain structures dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the thalamus. Average Rand Index 0.6467

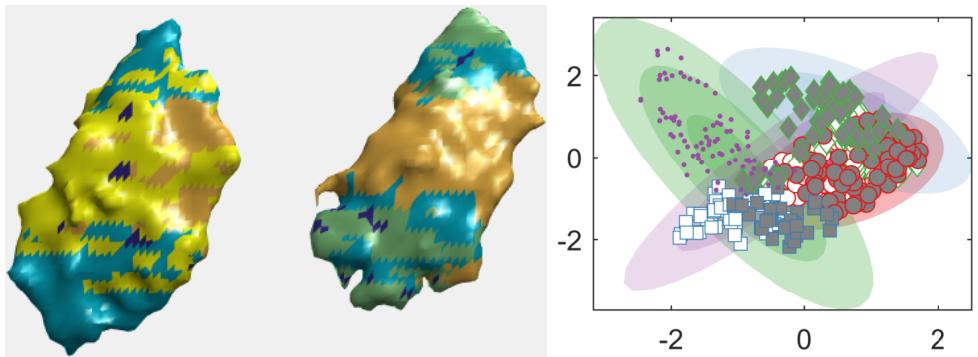


Figure 29: Experimental results for the brain structures dataset. Current mixture parameters, along with the latent positions for two shapes exhibiting different parts of the putamen. Average Rand Index 0.6467