Geodesic Finite Mixture Models

Institut de Robòtica i Informàtica Industrial

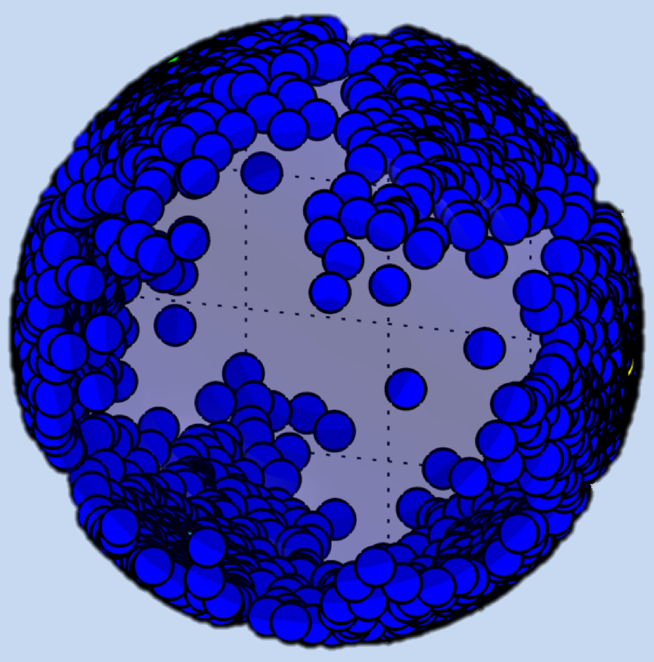
Edgar Simo-Serra, Carme Torras, Francesc Moreno-Noguer

IRI (CSIC-UPC)

Barcelona, Spain

PROBLEM:

- Estimation of the Probability Density Function of data that lies on a known Riemannian Manifold
- Clustering of data on a known Riemannian Manifold

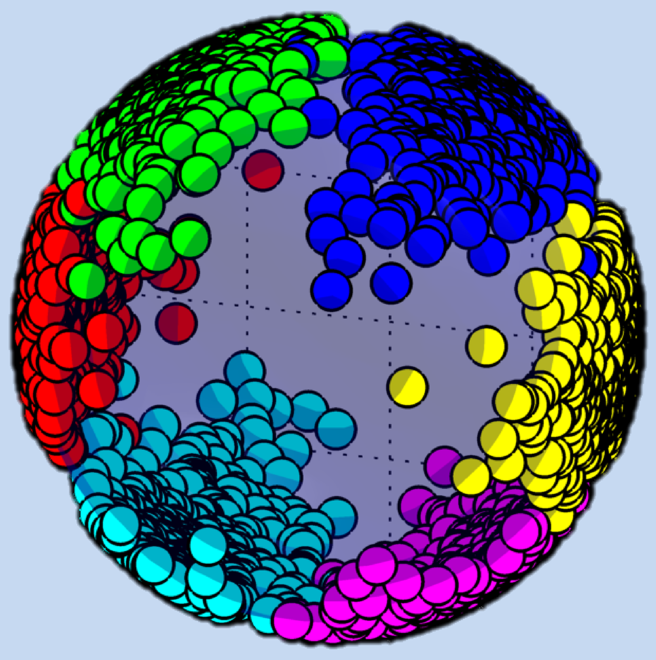


GIVEN:

- ✓ Known Manifold
- ✓ Points on Manifold

WE WANT TO ESTIMATE:

- ✓ Number of Clusters
- ✓ Cluster Parameters

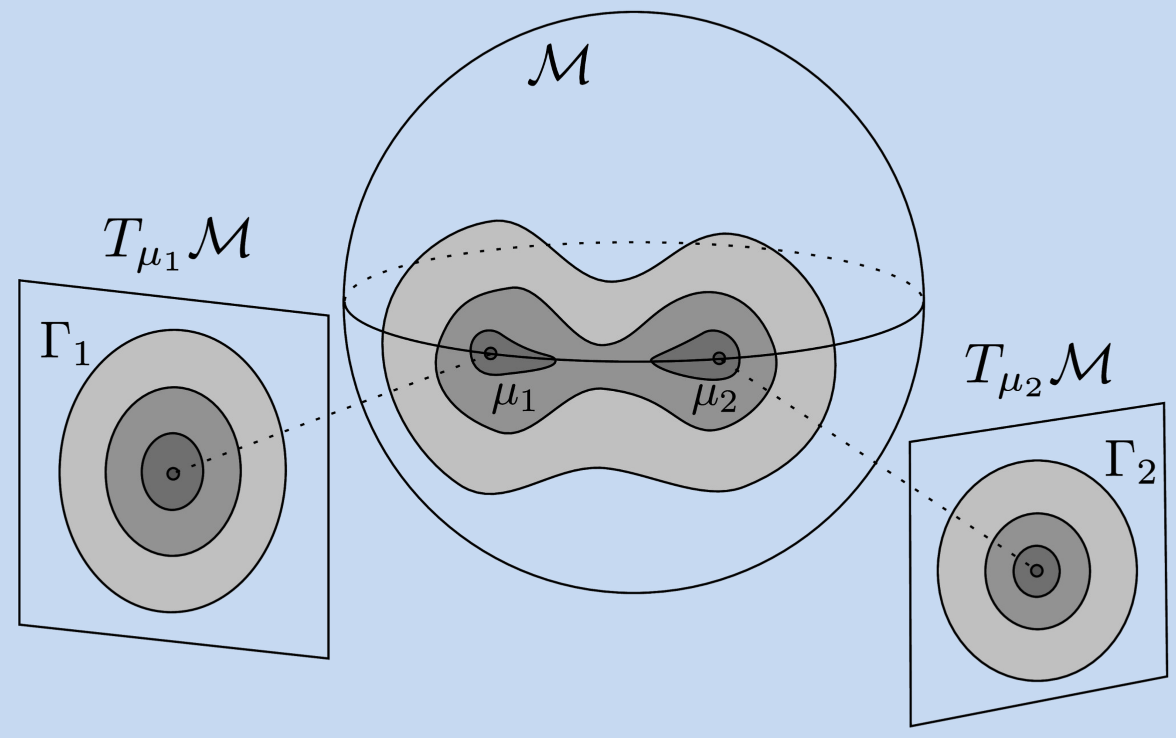


CONTRIBUTIONS:

- Unsupervised clustering algorithm for manifolds
- One tangent space per cluster
- Efficient implementation

KEY FEATURES:

- Generative Model
- Fully unsupervised
- Scales well to large datasets

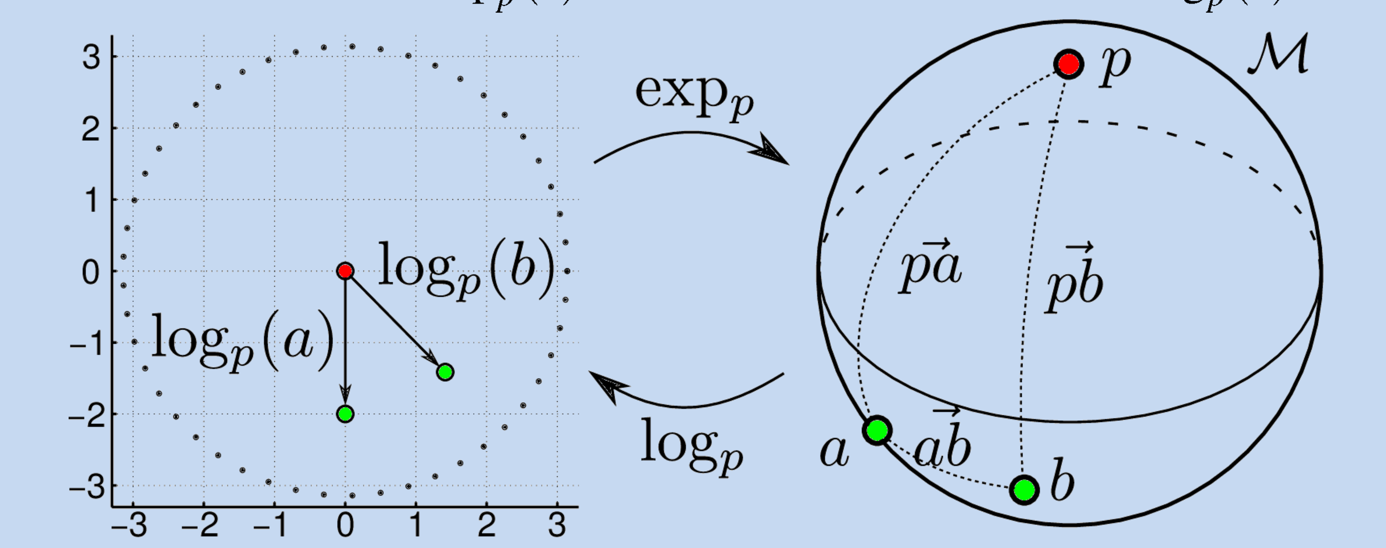


MANIFOLDS, GEODESICS, AND TANGENT SPACES

- **Geodesic distance** between two points on a manifold is the shortest distance between the two points on the manifold
- **Tangent space** is a local approximation of a manifold that is a **Euclidean space**
- Geodesic distances on the tangent space are exact between any point and the origin, but approximate between any pair of arbitrary points

The mapping to and from the tangent space is defined by two operators:

$$\exp_p : \begin{array}{ccc} T_p\mathcal{M} & \longrightarrow & \mathcal{M} \\ v & \longmapsto & \exp_p(v) = x \end{array}, \qquad \log_p : \begin{array}{ccc} \mathcal{M} & \longrightarrow & T_p\mathcal{M} \\ x & \longmapsto & \log_p(x) = v \end{array}$$



STATISTICS ON TANGENT SPACES

- Mean estimated on manifold using **geodesic mean**

Iterate until convergence to obtain geodesic mean

$$\mu = \arg\min_p \sum_{i=1}^{N} d(x_i, p)^2 \quad \Rightarrow \quad \mu(t+1) = \exp_{\mu(t)}\left(\frac{\delta}{N}\sum_{i=1}^{N}\log_{\mu(t)}(x_i)\right)$$

- Covariance estimated on **tangent space** in closed form

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}\log_\mu(x_i)\log_\mu(x_i)^\top$$

- Given the mean and covariance we define a normal distribution on the tangent space as:

$$\mathcal{N}_\mu(\nu, \Sigma^{-1}) = \lambda \exp\left(-\frac{\log_\mu(x)^\top \Sigma^{-1}\log_\mu(x)}{2}\right)$$

UNSUPERVISED FINITE MIXTURE MODELLING

- Extension of unsupervised learning of finite mixture models [1]
- Center each cluster on a tangent space to minimize geodesic error
- **Minimum Message Length (MML)** used to determine number of clusters
- **Expectation-Maximization (EM)** algorithm

E-step

$$w_k^{(i)} = \frac{\alpha_k(t-1)p(x_i|\theta_k(t-1))}{\sum_{k=1}^{K}\alpha_k(t-1)p(x_i|\theta_k(t-1))}$$

M-step

$$\alpha_k(t) = \frac{1}{N}\sum_i^N w_k^{(i)} = \frac{w_k}{N}, \qquad \mu_k(t) = \arg\min_p \sum_{i=1}^{N} d\left(\frac{N}{w_k}w_k^{(i)}x^{(i)}, p\right)^2$$

$$\Sigma_k(t) = \frac{1}{w_k}\sum_{i=1}^{N}\left(\log_{\mu_k(t)}(x^{(i)})\right)\left(\log_{\mu_k(t)}(x^{(i)})\right)^\top w_k^{(i)}$$

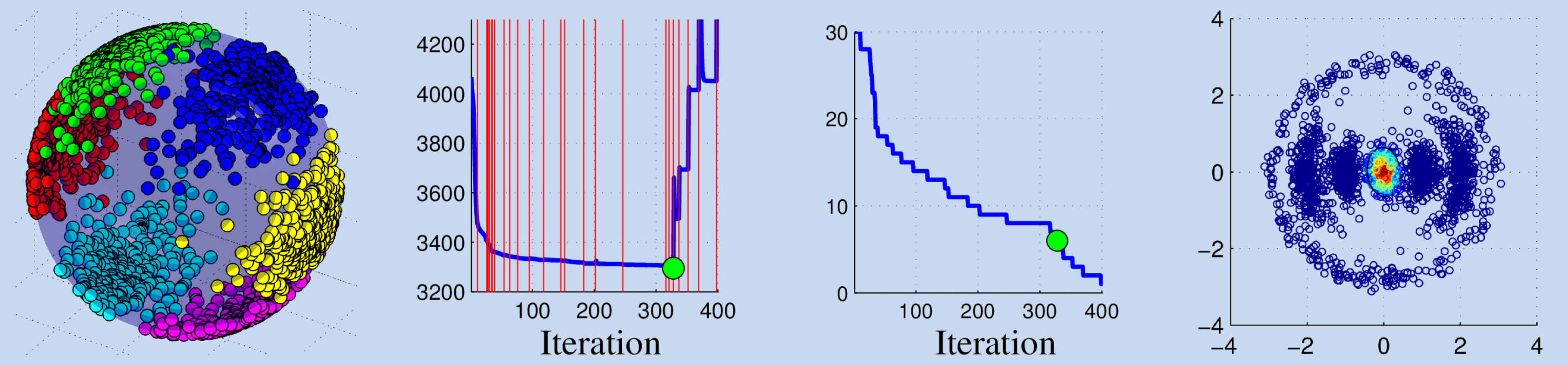
Final Distribution $p(x|\theta) = \sum_{k=1}^{K}\alpha_k p(x|\theta_k)$ with $p(x|\theta_k) \approx \mathcal{N}_{\mu_k}\left(0, \Sigma_k^{-1}\right)$

Regression $p(x_A|x_B,\theta) = \frac{p(x_A, x_B|\theta)}{p(x_B|\theta_B)} = \frac{\sum_{k=1}^{K}\alpha_k p(x_B|\theta_{k,B})p(x_A|x_B,\theta_k)}{\sum_{k=1}^{K}\alpha_k p(x_B|\theta_{k,B})}$

EXPERIMENTAL RESULTS:

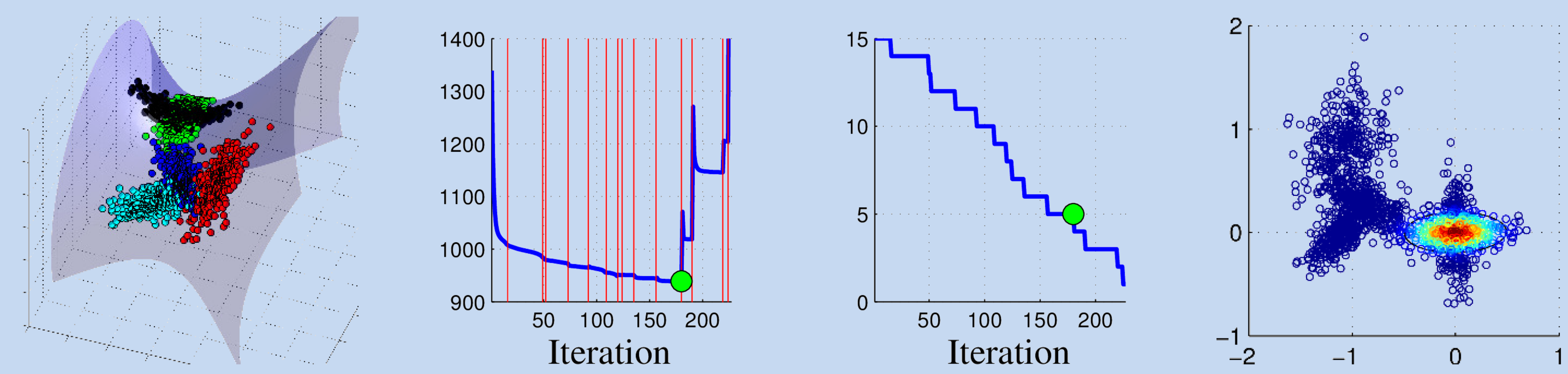
- Evaluate on **synthetic** and **real datasets**

SPHERE EXAMPLE

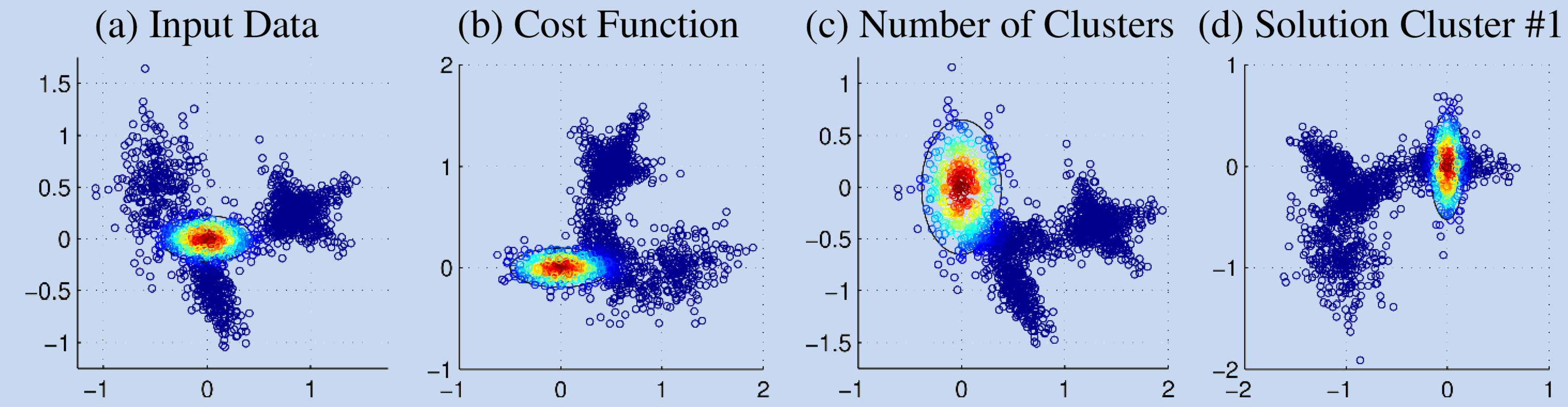


(a) Input Data (b) Cost Function (c) Number of Clusters (d) Solution Cluster #1

QUADRATIC SURFACE EXAMPLE



(a) Input Data (b) Cost Function (c) Number of Clusters (d) Solution Cluster #1



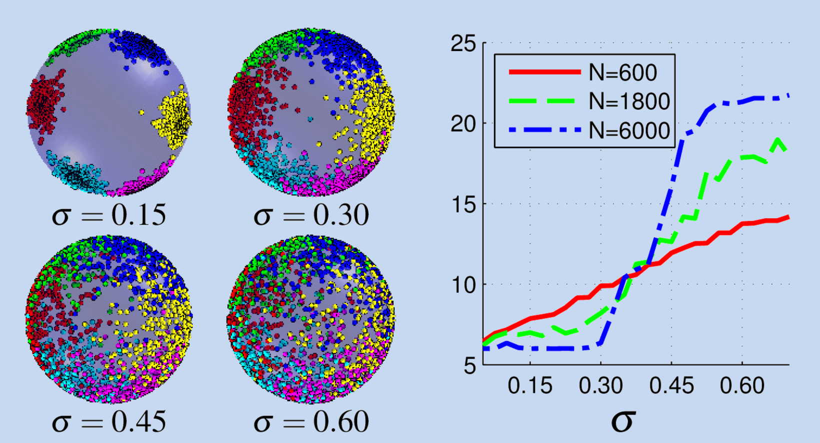
(e) Solution Cluster #2 (f) Solution Cluster #3 (g) Solution Cluster #4 (h) Solution Cluster #5

ROBUSTNESS ANALYSIS

Comparison with

- 1-TM: One tangent plane
- vMF: von Mises-Fischer distributions [2]
- GMM: Euclidean mixture model

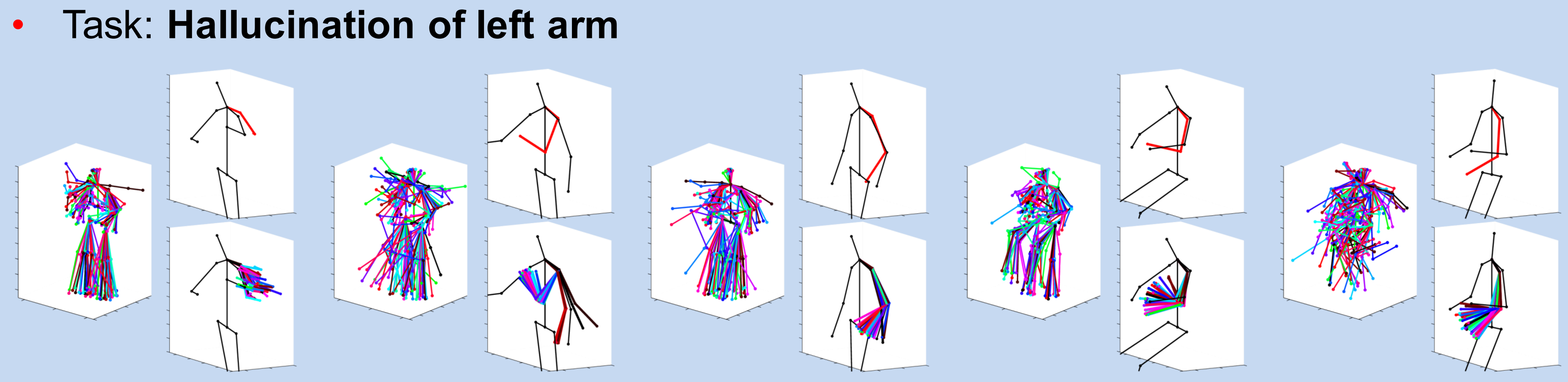
	6 Clusters		8 Clusters	
	Clusters	Correct	Clusters	Correct
Ours	6.09 (0.32)	0.92	8.00 (0.00)	1.00
1-TM	7.05 (1.38)	0.46	15.25 (2.17)	0.00
vMF	16.59 (1.71)	0.00	19.86 (2.35)	0.00



(a) Example Input Data (b) Mean Clusters

REAL REGRESSION EXAMPLE

- Trained on full poses of Human3.6m dataset [3]
- Task: **Hallucination of left arm**



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3 (d) Scenario 4 (e) Scenario 5

	Scenario 1			Scenario 2			Scenario 3			Scenario 4			Scenario 5		
	MGJE	MJE	MLLE	MGJE	MJE	MLLE	MGJE	MJE	MLLE	MGJE	MJE	MLLE	MGJE	MJE	MLLE
Ours	0.446	105.8	0.0	0.468	110.1	0.0	0.349	81.7	0.0	0.458	108.2	0.0	0.597	135.7	0.0
vMF	0.481	114.5	0.0	0.568	134.8	0.0	0.470	110.2	0.0	0.496	118.0	0.0	0.698	162.3	0.0
1-TM	0.522	123.0	0.0	0.640	148.7	0.0	0.535	124.9	0.0	0.548	130.2	0.0	0.765	175.1	0.0
GMM	1.111	103.1	19.0	1.167	106.6	27.5	1.152	77.6	11.3	1.272	101.0	14.2	1.401	127.3	24.8

REFERENCES

[1] M. Figueiredo and A. Jain. Unsupervised Learning of Finite Mixture Models. *PAMI*, 2002.

[2] A. Banerjee, I. S. Dhillon, J. Ghosh, S. Sra, and G. Ridgeway. Clustering on the unit hypersphere using von mises-fisher distributions. *JMLR*, 2005.

[3] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *PAMI*, 2014.