Lie Algebra-Based Kinematic Prior for 3D Human Pose Tracking

Edgar Simo-Serra, Carme Torras, Francesc Moreno-Noguer
Institut de Robòtica i Informàtica Industrial (CSIC-UPC)
Barcelona, Spain

PROBLEM:

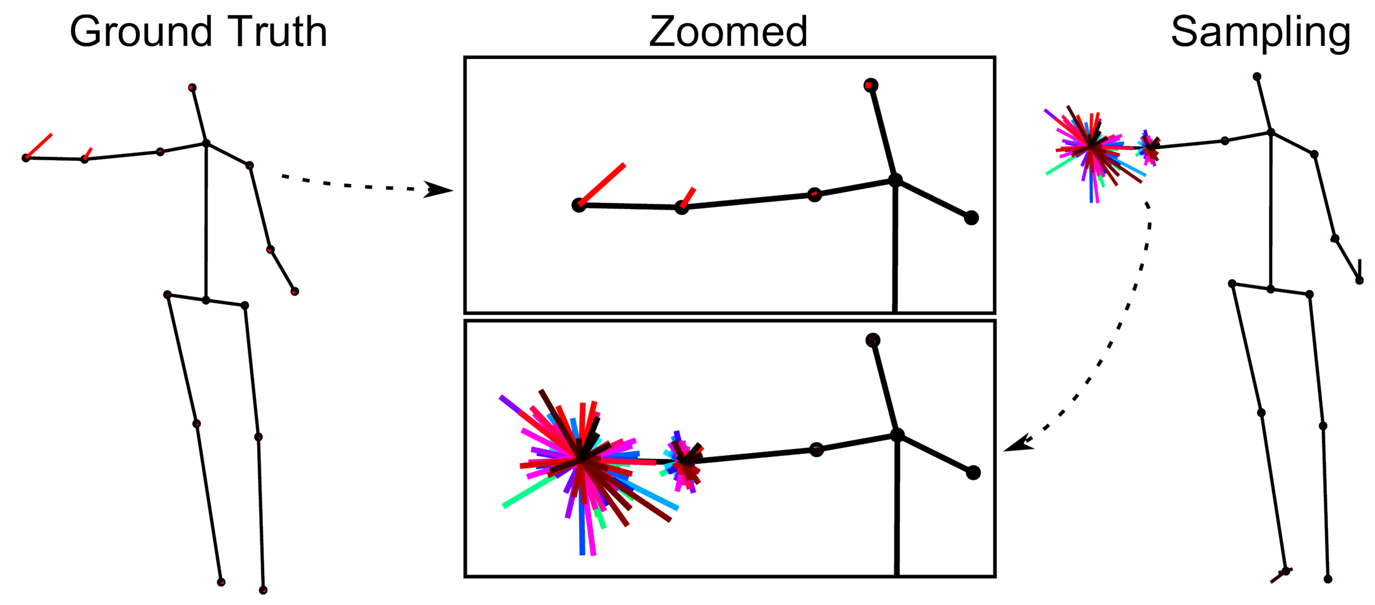
- Predicting motion hypothesis from poses
- Only generating feasible pose hypothesis

CONTRIBUTIONS:

- Joint Pose and Kinematic Manifold
- Efficient approach for sampling
- Outperforms widely used Gaussian diffusion models

KEY FEATURES:

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



GIVEN:

✓ 3D positions of joints

WE WANT TO ESTIMATE:

✓ Distribution of velocities

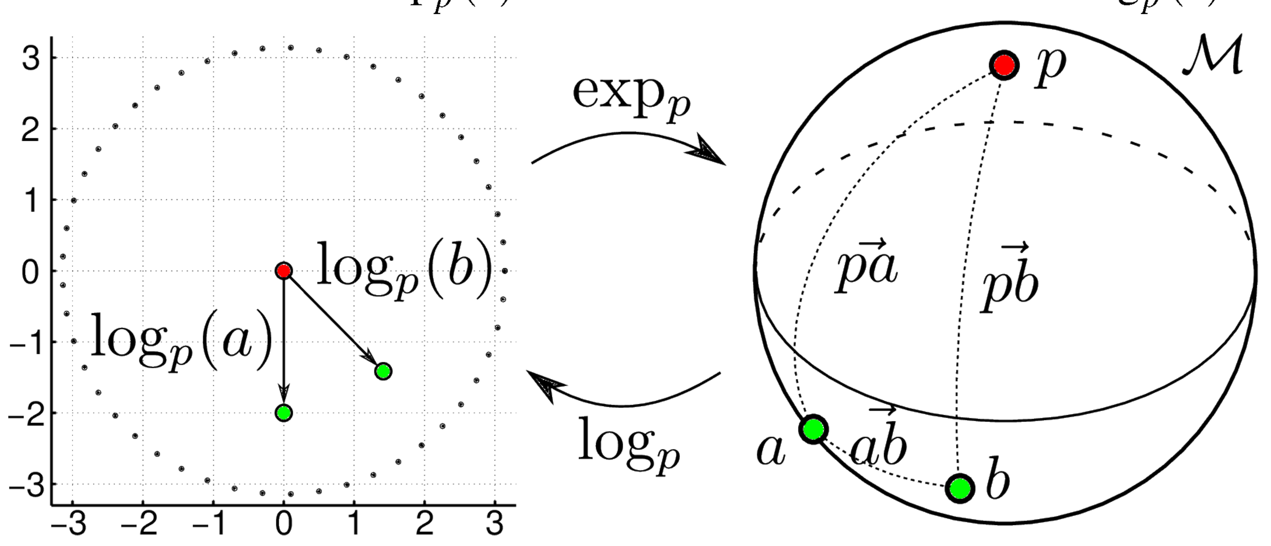
Prior	Complexity	Scales	Consistent
Gaussian diffusion	Low	Yes	No
GPLVM [1]	Low	No	No
GPDM [2]	Medium	No	No
hGPLVM [3]	Medium	No	No
CRBM [4]	High	Yes	No
GCMFA [5]	High	No	No
GFMM (Ours)	Low	Yes	Yes

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**

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

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



STATISTICS ON TANGENT SPACES

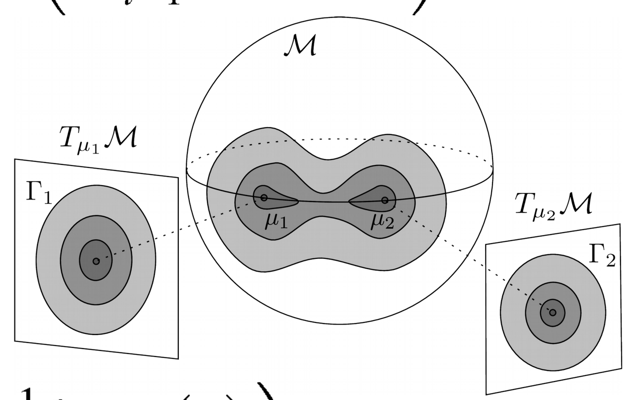
- Mean estimated on manifold using **geodesic mean**

Iterate until convergence to obtain geodesic mean

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

- 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:

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

UNSUPERVISED FINITE MIXTURE MODELLING

- Extension of unsupervised learning of finite mixture models [6]
- 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

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

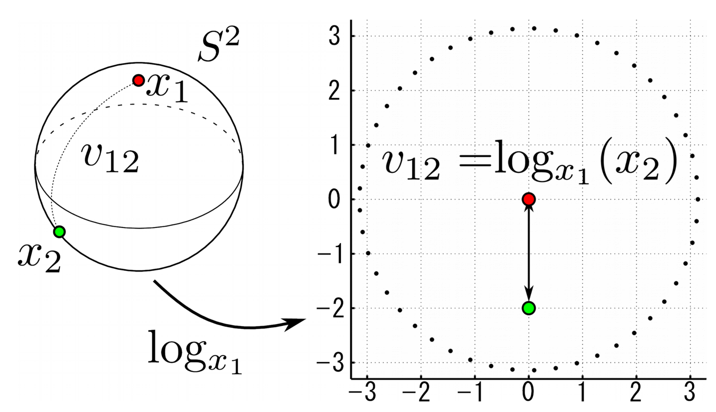
JOINT POSE AND KINEMATIC MANIFOLD

- Rotation between consecutive joints modeled as $SO(3/2)^n$ manifold
- Kinematics modeled with associated quotient of Lie algebras $\mathfrak{so}(3/2)^n$

$$v_{12} = \log_{x_1}(x_2), \quad x_1, x_2 \in SO(3/2)^n, \quad v_{12} \in \mathfrak{so}(3/2)^n$$

$$\log_{v_1}(v_2) = v_2 - v_1$$

$$\exp_{v_1}(v_2) = v_2 + v_1$$



- Joint distribution of Pose and Kinematic is learnt

KINEMATIC MODEL

- Kinematics conditioned on pose

$$p(v|x,\theta) = \frac{p(x,v|\theta)}{p(x|\theta_x)} = \frac{\sum_{k=1}^{K}\alpha_k p(x|\theta_{k,x})p(v|x,\theta_k)}{\sum_{k=1}^{K}\alpha_k p(x|\theta_{k,x})}.$$

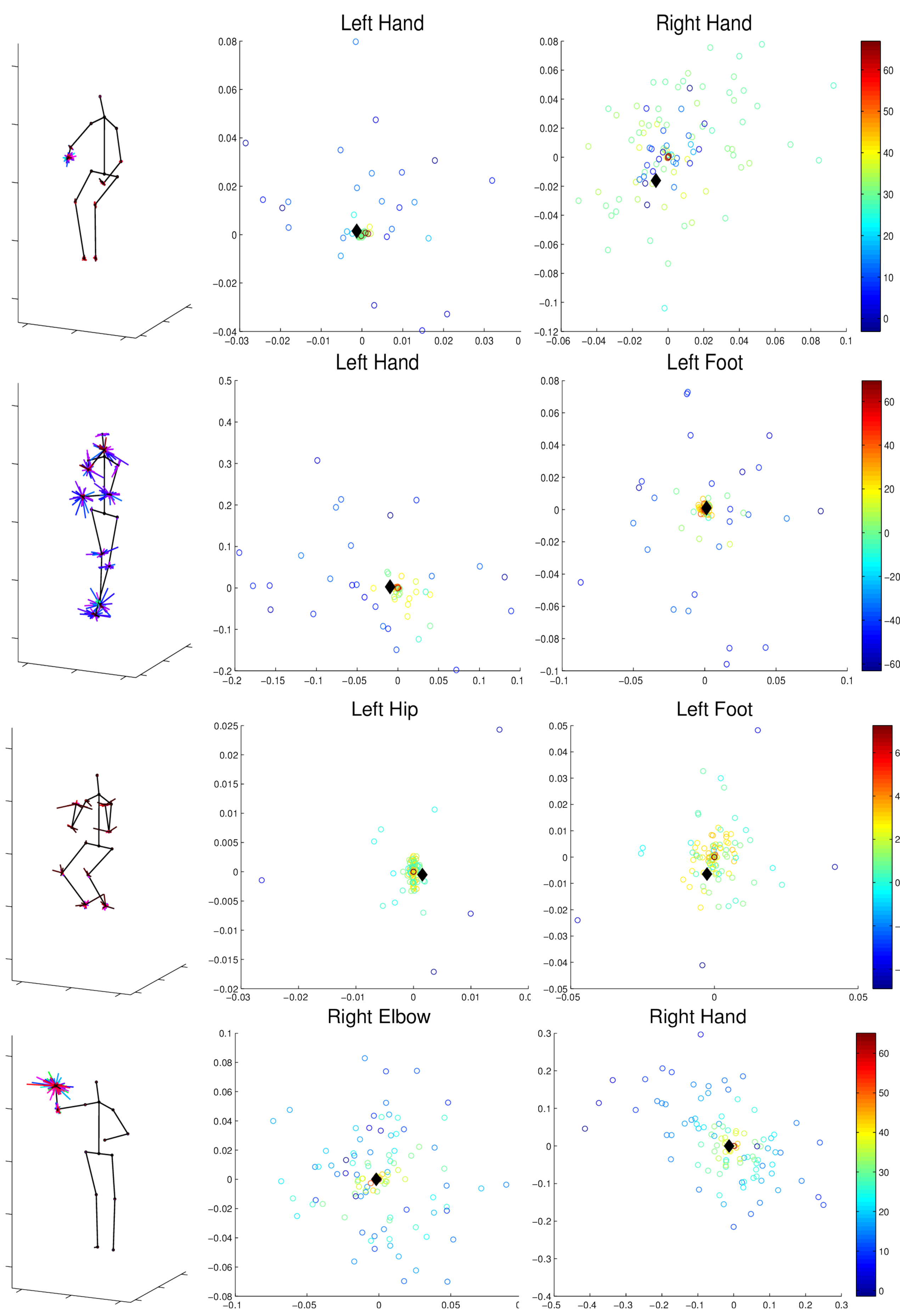
- Regression gives another mixture model $\Sigma_k^{-1} = \Gamma_k = \begin{bmatrix}\Gamma_{k,x} & \Gamma_{k,vx} \\ \Gamma_{k,vx} & \Gamma_{k,v}\end{bmatrix}$

$$p(v|x,\theta_k) = \mathcal{N}_{\mu_v}\left(\Gamma_{k,vx}\Gamma_{k,x}^{-1}\log_{\mu_{k,x}}(x_x),\, \Gamma_{k,v} - \Gamma_{k,vx}\Gamma_{k,x}^{-1}\Gamma_{k,vx}\right)$$

EXPERIMENTAL RESULTS

- Evaluation on Human3.6M dataset
- 15 categories of actions
- 6 actors used for training, 1 actor used for testing

Method	Log-likelihood Train	Log-likelihood Test
Samples	465,325	62,064
Gaussian diffusion	5.4325	5.4349
local Gaussian diffusion	6.4193	6.4206
Ours (30%, 211 clusters)	9.3382	11.7874
Ours (15%, 147 clusters)	8.9544	11.8714



- **GFMM code available:**
http://www.iri.upc.edu/people/esimo/code/gfmm/

REFERENCES

[1] N. D. Lawrence. Probabilistic Non-linear Principal Component Analysis with Gaussian Process Latent Variable Models. JMLR, 6:1783-1816, 2005.
[2] J. Wang, D. Fleet, and A. Hertzmann. Gaussian process dynamical models. In NIPS, 2005.
[3] M. Andriluka, S. Roth, and B. Schiele. Monocular 3D Pose Estimation and Tracking by Detection. In CVPR, 2010.
[4] G. Taylor, L. Sigal, D. Fleet, and G. Hinton. Dynamical binary latent variable models for 3d human pose tracking. In CVPR, 2010.
[5] R. Li, T.-P. Tian, S. Sclaroff, and M.-H. Yang. 3d human motion tracking with a coordinated mixture of factor analyzers. IJCV, 87(1-2):170-190, 2010.
[6] E. Simo-Serra, C. Torras, and F. Moreno-Noguer. Geodesic Finite Mixture Models. In BMVC, 2014.