Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group

Raviteja Vemulapalli, Felipe Arrate and Rama Chellappa

Department of Electrical and Computer Engineering, University of Maryland, College Park.

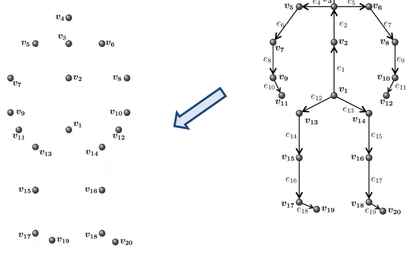
Skeleton-based Action Recognition

- Humans can recognize many actions directly from skeletal sequences.
- Skeletal data can be generated in real time using the algorithm of [Shotton *et. al.*]

Skeletal data → Skeletal Representation → Temporal Modeling & Classification → Action label

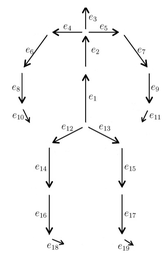
Human Skeleton

Set of points



Representation: Joint coordinates

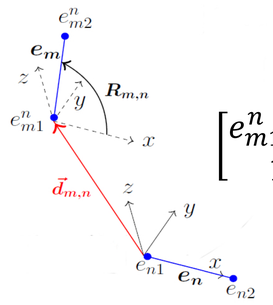
Set of rigid rods



Representation: Joint angles

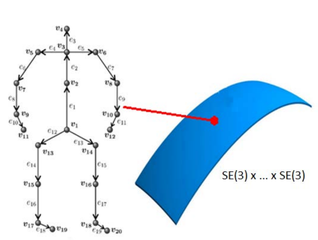
Proposed Skeletal Representation

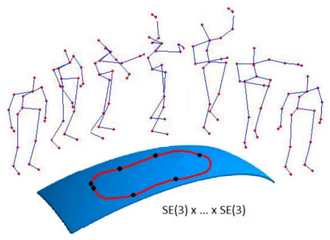
- For action recognition, we need a skeletal representation whose temporal evolution directly describes the relative motion between various body parts.
- We represent a skeleton using the relative 3D geometry between different body parts.



Element of special Euclidean group SE(3)

$$\begin{bmatrix} e^n_{m1}(t) & e^n_{m2}(t) \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} R_{m,n}(t) & \vec{d}_{m,n}(t) \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} e_{n1}(t) & s_{m,n} * e_{n2}(t) \\ 1 & 1 \end{bmatrix}$$





- $SE(3) \times \cdots \times SE(3)$ is a curved, smooth manifold.
- We approximate the action curves in the Lie group $SE(3) \times \cdots \times SE(3)$ by mapping them to the Lie algebra $\mathfrak{se}(3) \times \cdots \times \mathfrak{se}(3)$, which is the tangent space at identity.

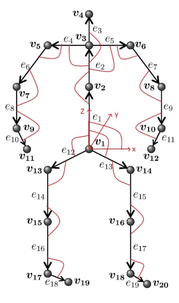
Temporal Modeling & Classification

Curve in $\mathfrak{se}(3) \times \cdots \times \mathfrak{se}(3)$ → Dynamic Time Warping → Fourier Temporal Pyramid Representation → Linear SVM → Action label

Experimental Results

Joint positions (JP): Concatenation of the joint coordinates.

Relative joint positions (RJP): Concatenation of the 3D vectors $\overrightarrow{v_i v_j},\ 1 \leq i < j \leq 20$.



Joint angles (JA): Concatenation of the quaternions corresponding to the joint angles.

Individual body part locations(BPL): Each body part is represented as a point in $SE(3)$ using its relative 3D geometry with respect to the global x-axis.

- MSR-Action3D: 20 actions, 10 subjects

JP	RJP	JA	BPL	Proposed
87.22	88.23	81.83	83.54	**92.46**

- UTKinect-Action: 10 actions, 10 subjects

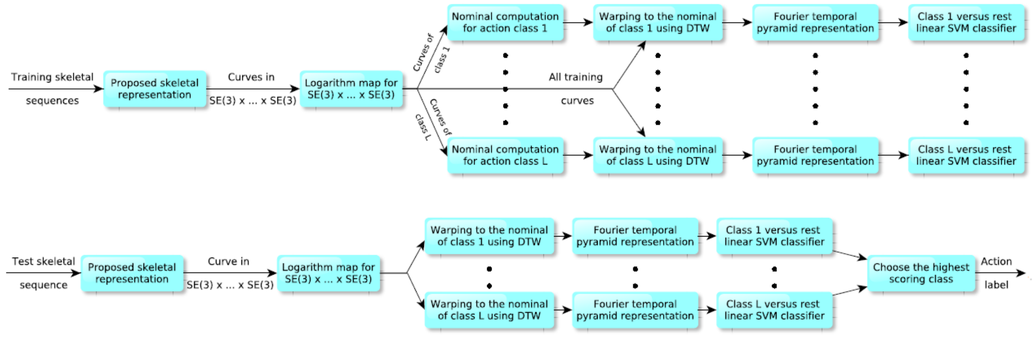
JP	RJP	JA	BPL	Proposed
94.68	95.58	94.07	94.57	**97.08**

- Florence3D-Action: 9 actions, 10 subjects

JP	RJP	JA	BPL	Proposed
85.26	85.20	81.36	80.80	**90.88**

State-of-the-art (skeleton based)	MSR-A3D	UTKA	F3DA
	90.90	90.92	82.00

Proposed Recognition System



References

- J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman and A. Blake, "Real-time Human Pose Recognition in Parts From a Single Depth Image", CVPR, 2011.
- A. Veeraraghavan, A. Srivastava, A. K. Roy-Chowdhury and R.Chellappa, "Rate-invariant Recognition of Humans and Their Activities", IEEE Trans. on Image Processing, 18(6):1326–1339, 2009.
- J. Wang, Z. Liu, Y. Wu and J. Yuan, "Mining Actionlet Ensemble for Action Recognition with Depth Cameras", CVPR, 2012.