Multiview-Consistent Semi-Supervised Learning for 3D Human Pose Estimation

Rahul Mitra[1*], Nitesh Gundavarapu[2*], Abhishek Sharma[3], Arjun Jain[3,4]

IIT Bombay[1] and UC San Diego[2] and Axogyan AI[3] and IISc Bangalore[4]

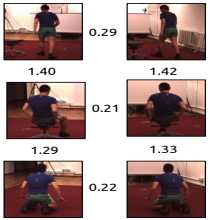
CVPR 2020, Seattle

* - Equal Contribution

Overview

- Motivation:
  - Available annotated 3D pose datasets are biased towards indoor environments do not generalise well for real-world scenarios.
  - Multi-view time-synchronized human motion videos are easier to obtain than 3D annotated images.

- Contribution:
  - Novel metric learning based semi-supervised framework using multi-view videos (**MCSS**) and limited 3D annotations.
  - Jointly learning a view-invariant embedding capturing pose similarity(shown in Fig.) and regressing 3D pose from it.



Embedding distances. Rows: same pose, different viewpoint. Cols: differen pose, same viewpoint.

Proposed Approach

Proposed MCSS approach consists of two modules trained jointly,

- Multiview-consistent metric-learning from time synchronised videos
- 3D-pose regression with limited 3D supervision

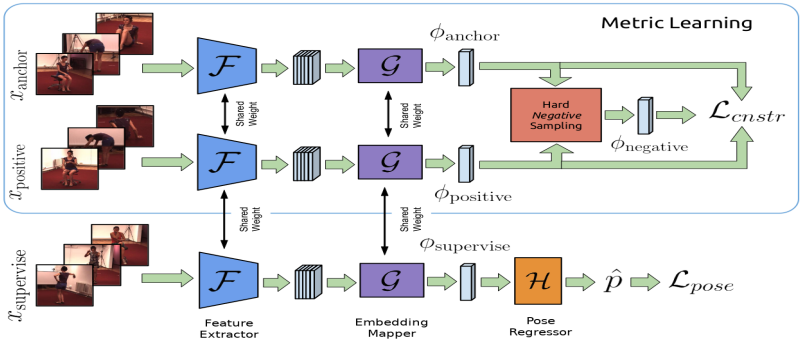
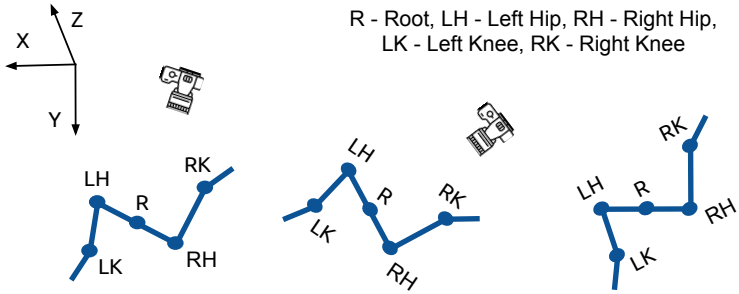


Figure: (x_{anchor}, x_{positive}) - same pose, different views. The *Hard Negative Sampling* module performs in-batch hard mining.

Canonical Pose Representation

- Regressing pose in local camera co-ordinates from our embedding is a *one-to-many* function. In MoCap system's *Global* co-ordinates poses are not invariant when they are rigid transformations of one another.
- Above problems are eliminated in our *Canonical* pose, where the bone connecting 'pelvis' and 'left-hip' joint is always kept parallel to the XZ plane.



Global Pose (Time T1, View V1)

Global Pose (Time T2, View V2)

Canonical Pose (Time T1, T2)

Pose Estimation Results

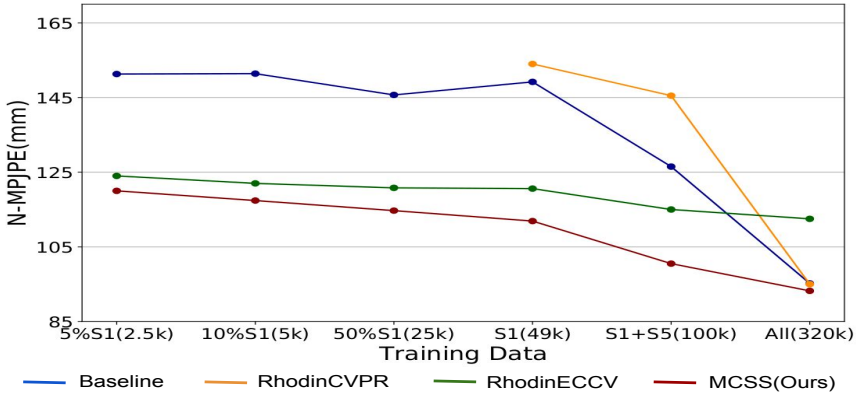


Figure: N-MPJPE of MCSS against other methods on Human3.6M(Ionescu rt al. 2013) dataset with progressively limited 3D supervision. Our proposed method outperforms RhodinECCV(Rhodin et al. ECCV 2018) and RhodinCVPR(Rhodin et al. CVPR 2018).

Qualitative Pose Estimation Results

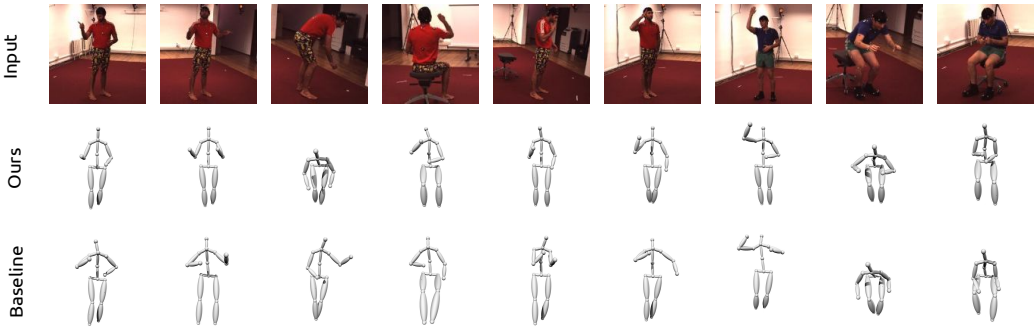


Figure: Qualitative results on *canonical* pose estimation by our (**MCSS**) against our **Baseline** on Human 3.6M test split trained with limited 3D supervision. Our method produces more accurate estimates.