Orthographic Projection Linear Regression for Single Image 3D Human Pose Estimation

Yahui Zhang, Shaodi You and Theo Gevers
Computer Vision Lab, Infomatics Institute, University of Amsterdam
y.zhang5@uva.nl

UNIVERSITEIT VAN AMSTERDAM

1. Overview

Goal: Predict 3D huamn joint locations in the camera coordinate from a single image.
Challenge: In-the-wild images are extremely complex and do not have corresponding 3D ground truth.
Contribution:

- We propose a novel orthographic projection and linear regression to constrain the 3D and 2D poses.
- A network is proposed which is adaptive to various in-the-wild images without retraining the 3D pose.
- Our network achieves state-of-the-art performance on the Human3.6m dataset and generalizes well to in-the-wild datasets.

2. Limitation

The perspective projection from 3D to 2D poses may cause problems. To minimize the error between 2D projections and 2D ground truth, the value of depth (i.e., z axis) will deviate to find the optimal solution. This may lead to overfitting. Therefore, an orthographic projection is employed in this work to solve this issue.

6. Qualitative Results

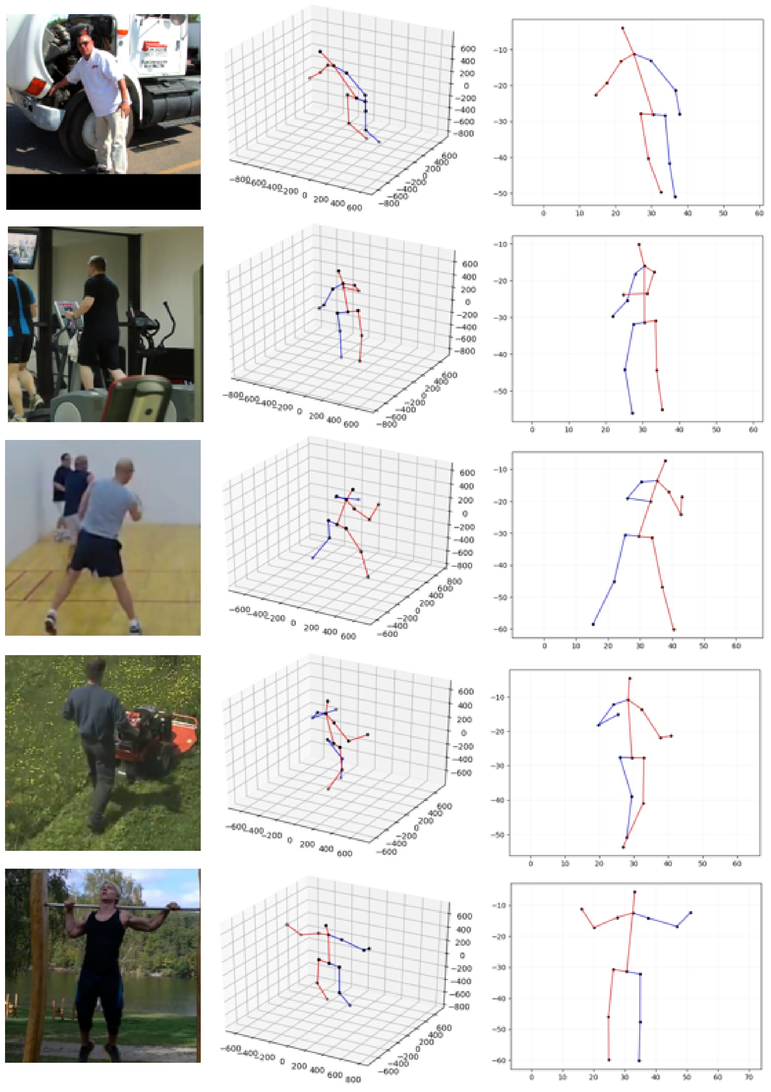

Figure 3. Qualitative results generated by our method on MPII [3] and LSP [4] datasets.

3. Orthographic Projection Linear Regression

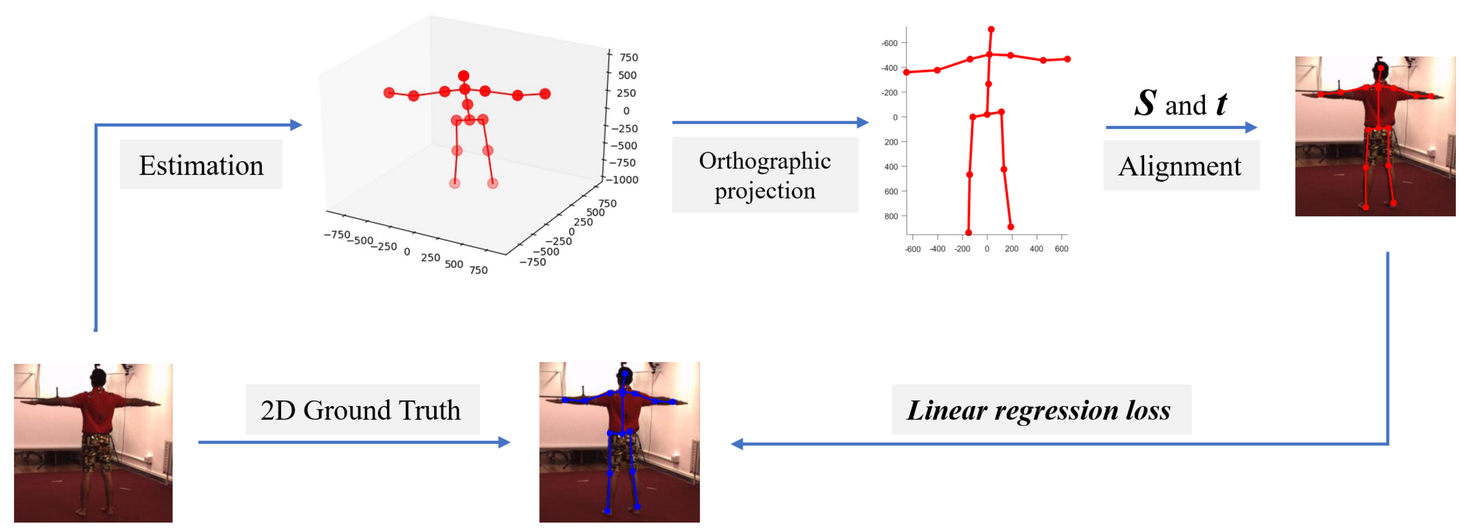

Figure 1. The general idea of matching 3D with 2D poses by the orthographic projection linear regression method.

4. Architecture

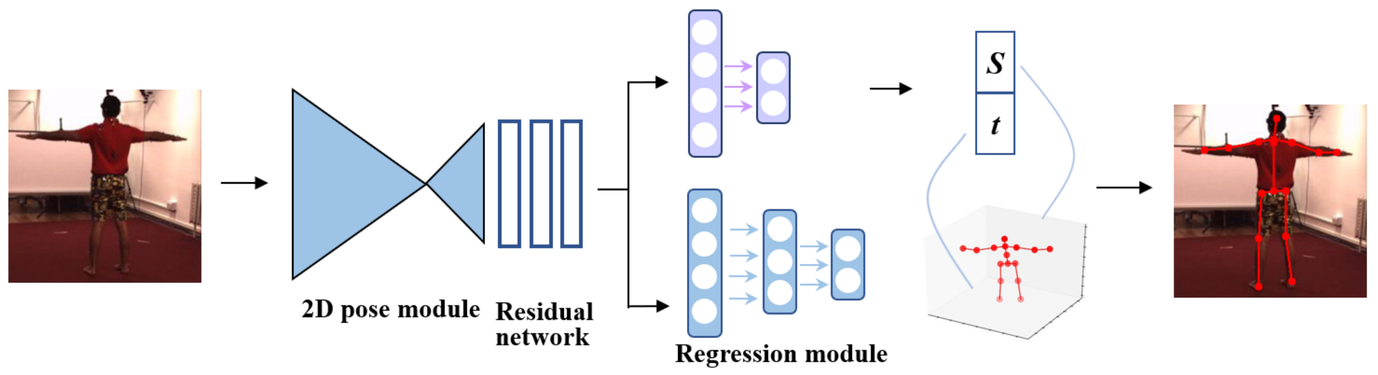

Figure 2. The overview of the proposed framework.

5. Experiments

Datasets and Metric

- MPI-INF-3DHP [1]
  Percentage of Correct Keypoints (PCK).
  Aera under the Curve (AUC)
  * *The larger, the better.*
- Human3.6m [2]
  Protocol #1: Mean Per Joint Position Error (MPJPE).
  Protocol #2: MPJPE after a rigid transformation (PA MPJPE).
  * *The smaller, the better.*

Table 1. Experimental results on MPI-INF-3DHP dataset [1].

Methods	Extra information	PCK	AUC
Mehta *et al.* (3DV'17)		64.7	31.7
Zhou *et al.* (ICCV'17)	Post-processing	68.2	32.5
Yang *et al.* (CVPR'18)		69.0	32.0
Habibie *et al.* (CVPR'19)	Extra training set	70.4	36.0
Want *et al.* (CVPR'19)	Extra training set	81.8	54.8
Ci *et al.* (ICCV'19)	2D Pose	74.0	34.7
Ours (w/o $\mathcal{L}_{OPLR}$)		23.9	8.9
Ours (full)		66.8	31.9

Methods (Using rigid transformation)	Extra information	PCK	AUC
Habibie *et al.* (CVPR'19)	Extra training set	82.9	45.4
Ours		**84.4**	**46.9**

Table 2. Experimental results on Human3.6m dataset [2].

Protocol #1	Dir.	Dis.	Eat	Gre.	Phon.	Pose	Pur.	Sit	SitD.	Smo.	Phot.	Wait	Walk	WalkD.	WalkP.	Avg
Zhou *et al.* (CVPR'16)	87.4	109.3	87.1	103.2	116.2	143.3	106.9	99.8	124.5	199.2	107.4	118.1	114.2	79.4	97.7	113.0
Chen *et al.* (CVPR'17)	89.9	97.6	90.0	107.9	107.3	93.6	136.1	133.1	240.1	106.7	139.2	106.2	87.0	114.1	90.6	114.2
Pavlakos *et al.* (CVPR'17)	67.4	71.9	66.7	69.1	72.0	77.0	65.0	68.3	83.7	96.5	71.7	65.8	74.9	59.1	63.2	71.9
Mehta *et al.* (3DV'17)	57.5	68.6	59.6	67.3	78.1	56.9	69.1	98.0	117.5	69.5	82.4	68.0	55.3	76.5	61.4	72.9
Zhou *et al.* (ICCV'17)	54.8	60.7	58.2	71.4	62.0	65.5	53.8	55.6	75.2	111.6	64.1	66.0	51.4	63.2	55.3	64.9
Sun *et al.* (ICCV'17)	52.8	54.8	54.2	54.3	61.8	67.2	53.1	53.6	71.7	86.7	61.5	53.4	61.6	47.1	53.4	59.1
Luo *et al.* (BMVC'18)	53.5	60.9	56.3	59.1	64.3	74.4	55.4	63.4	74.8	98.0	61.1	58.2	70.6	49.1	55.7	63.7
Yang *et al.* (CVPR'18)	51.5	58.9	50.4	57.0	62.1	65.4	49.8	52.7	69.2	85.2	57.4	58.4	43.6	60.1	47.7	58.6
Zhao *et al.* (CVPR'19)	47.3	60.7	51.4	60.5	61.1	49.9	47.3	68.1	86.2	55.0	67.8	61.0	42.1	60.6	45.3	57.6
Ours	**46.0**	55.3	50.6	**53.5**	**57.5**	**46.3**	49.4	71.7	87.9	56.6	68.4	53.5	**41.4**	57.9	46.6	**56.2**

Protocol #2	Dir.	Dis.	Eat	Gre.	Phon.	Pose	Pur.	Sit	SitD.	Smo.	Phot.	Wait	Walk	WalkD.	WalkP.	Avg
Moreno-Noguer (CVPR'17)	66.1	61.7	84.5	73.7	65.2	67.2	60.9	67.3	103.5	74.6	92.6	69.6	71.5	78.0	73.2	74.0
Sun *et al.* (ICCV'17)	42.1	44.3	45.0	45.4	51.5	53.0	43.2	41.3	59.3	73.3	51.0	44.0	48.0	38.3	44.8	48.3
Luo *et al.* (BMVC'18)	40.8	44.6	42.1	45.1	48.3	54.6	41.2	42.9	55.5	69.9	46.7	42.5	48.0	36.0	41.4	46.6
Yang *et al.* (CVPR'18)	26.9	30.9	36.3	39.9	43.9	47.4	28.8	29.4	36.9	58.4	41.5	30.5	29.5	42.5	32.2	**37.7**
Zhou *et al.* (TPAMI'18)	47.9	48.8	52.7	55.0	56.8	65.5	49.0	45.5	60.8	81.1	53.7	51.6	54.8	50.4	55.9	55.3
Ours	35.8	41.0	42.3	42.0	**43.4**	**36.3**	36.7	55.1	66.5	**45.0**	49.6	41.2	32.9	43.9	39.0	43.4

7. Conclusions

In this paper, we proposed an orthographic projection linear regression module to construct a relation between the 3D human pose, 2D human pose projection and 2D image appearance. Experiments on several datasets validated the effectiveness and generalization ability of the proposed method qualitatively and quantitatively.

8. References

[1] Dushyant Mehta, Helge Rhodin, Dan Casas, Pascal Fua, Oleksandr Sotnychenko, Weipeng Xu, and Christian Theobalt. Monocular 3d human pose estimation in the wild using improved cnn supervision. In *3DV*, pages 506–516. IEEE, 2017.

[2] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *TPAMI*, 36(7):1325–1339, 2013.

[3] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *CVPR*, pages 3686–3693, 2014.

[4] Sam Johnson and Mark Everingham. Clustered pose and nonlinear appearance models for human pose estimation. In *bmvc*, volume 2, page 5. Citeseer, 2010.