A Dual-Source Approach for 3D Pose Estimation from a Single Image Supplementary Material

Hashim Yasin*[1], Umar Iqbal*[2], Björn Krüger[3], Andreas Weber[1], and Juergen Gall[2]

[1]Multimedia, Simulation, Virtual Reality Group, University of Bonn, Germany
[2]Computer Vision Group, University of Bonn, Germany
[3]Gokhale Method Institute, Stanford, USA

1. Qualitative Results

We present some qualitative results for the Human3.6M dataset [2] as well as Leeds Sports pose dataset [3]. Human3.6M dataset contains images captured in an indoor environment while Leeds Sports pose dataset consists of realistic images taken from the internet. For experiments on Leeds Sports pose dataset we train our regression forests and pictorial structure model using 1000 training images provided with the dataset, and use CMU motion capture dataset to develop our motion capture database. A few examples of resulting 3D pose estimates for both datasets are shown in Figure 1 and Figure 2, respectively. As evident in Figure 1 and Figure 2, our approach shows very good performance even for highly articulated poses, and also for images captured in unconstrained environments.

2. CMU Motion Capture Dataset

In Table 1 we provide details about the sequences taken from CMU motion capture dataset in order to develop our motion capture database used in this paper. We downsample these sequences from 120 Hz to 30 Hz that results in 360K poses for our CMU motion capture database.

References

[1] CMU. Carnegie mellon university graphics lab: Motion capture database, 2014. mocap.cs.cmu.edu. 4

[2] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *TPAMI*, 2014. 1, 2

[3] S. Johnson and M. Everingham. Clustered pose and nonlinear appearance models for human pose estimation. In *BMVC*, 2010. 1, 3

*authors contributed equally

Figure 1: A few qualitative results from Human3.6M dataset [2]: **(a)** represents input images, **(b)** shows refined 2D poses while **(c)** and **(d)** correspond to estimated 3D poses from two different views.

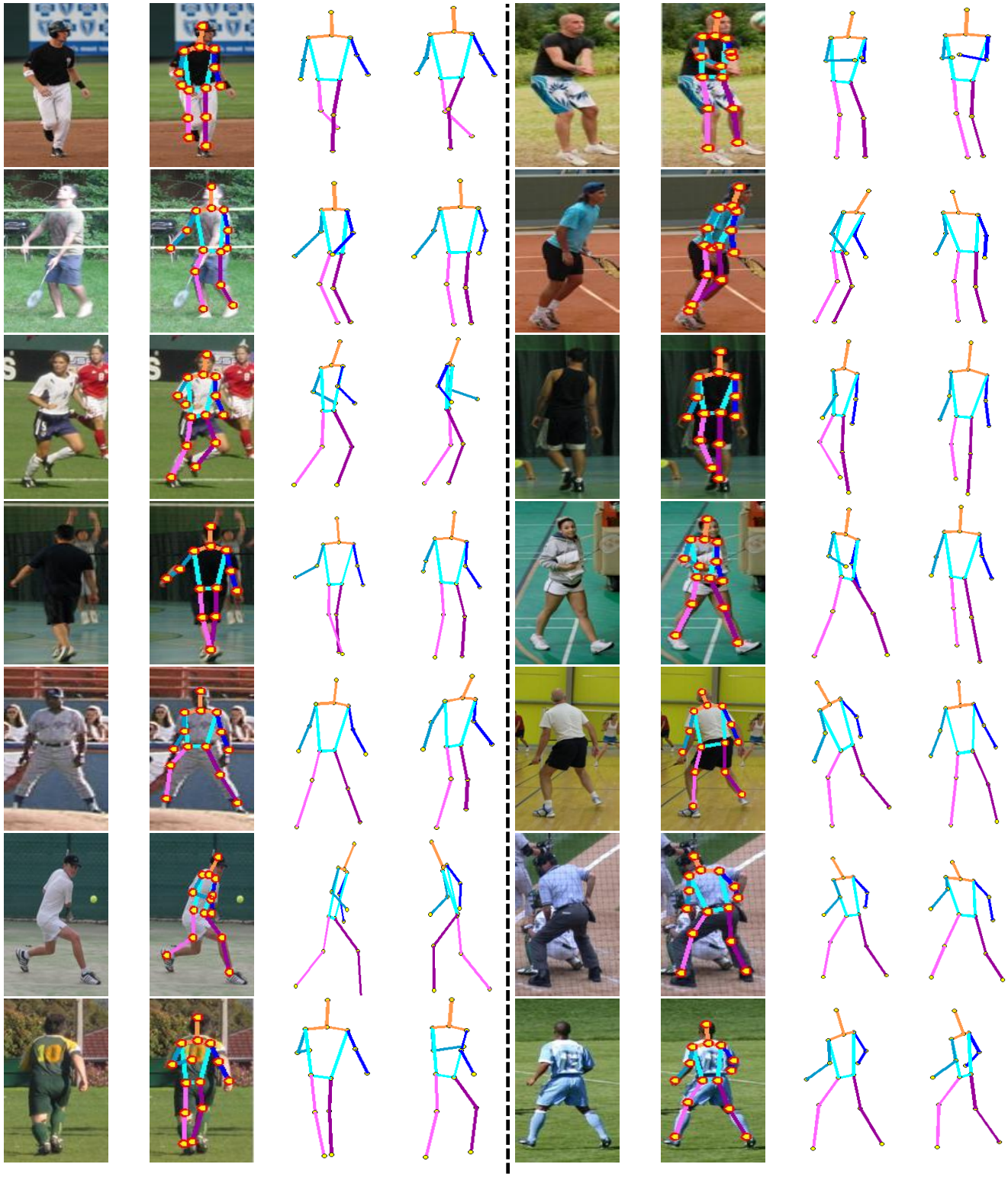

Figure 2: A few qualitative results from Leeds Sports pose dataset [3]: **(a)** represents input images, **(b)** shows refined 2D poses while **(c)** and **(d)** correspond to estimated 3D poses from two different views.

Subjects	Activities	Number of Trails
2	Various expressions and human behaviors (walk, punch, bend over, rise, lift arm *etc*.)	10
3	Walk on uneven terrain	4
7	Walk	12
8	Walk	11
9	Run	12
13	Various everyday behaviors (sit, stand up, boxing, climb ladder, sweep floor *etc*.)	42
14	Various everyday behaviors (direct traffic, wave, wash windows, jump up to grab *etc*.)	37
15	Various everyday behaviors, dance moves	14
16	Run, jump, walk	58
17	Different walking styles	10
18	Human interaction and communication (2 subjects - subject A)	15
19	Human interaction and communication (2 subjects - subject B)	15
20	Human interaction - at play, formations (2 subjects - subject A)	13
21	Human interaction - at play, formations (2 subjects - subject B)	13
33	Throw and catch football (2 subjects - subject A)	2
34	Throw and catch football (2 subjects - subject B)	2
35	Walk, run	34
36	Walk on uneven terrain	37
37	Walk	1
38	Walk, run around in a circle	4
45	Walk	1
46	Walk	1
47	Walk	1
69	Walking (walk sideways and turn, walk backwards and turn, walk and turn *etc*.)	75
81	Pushing a box; jumping off a ledge; walks	18
82	Jumping; pushing; emotional walks	18
91	Walks and turns	62
104	Motion (jog, excited walk, slow walk, run through *etc*.)	57
105	Motion (traffic walk, depressed walk, jump forward, jump turn, small jump *etc*.)	62
107	Walking with obstacles 1	14
108	Walking with obstacles 2	28
115	Bending over	10
123	Carry Suitcase with Varying Weights	13
132	Varying Weird Walks	56
138	Marching, Walking and Talking	55
139	Action Walks, sneaking, wounded, looking around	34
140	Getting Up From Ground	9
141	General Subject Capture	34
142	Stylized Walks	22
143	General Subject Capture	42

Table 1: List of sequences taken from CMU motion capture dataset [1] used in our experiments.