Supplementary Material:
Forecasting Human Dynamics from Static Images

Yu-Wei Chao[1], Jimei Yang[2], Brian Price[2], Scott Cohen[2], and Jia Deng[1]

[1]University of Michigan, Ann Arbor
{ywchao,jiadeng}@umich.edu

[2]Adobe Research
{jimyang,bprice,scohen}@adobe.com

1. Human Character Rendering

We demonstrate one potential application of 3D pose forecasting by rendering human characters from 3D skeletal poses. We use the public code provided by Chen *et al.* [1]: We first produce a 3D human shape model from each 3D skeletal pose using SCAPE. We then transfer skin and clothing textures to the 3D human model. Finally, the 3D model is rendered and overlaid on the person's projected bounding box in the input image. Fig. 1 shows the rendered human characters, both textureless and textured, for the qualitative results shown in the Fig. 8 of the paper. We believe the capability of pose forecasting with 3D human rendering may trigger further applications in augmented reality.

2. Performance on Individual Action Classes

Tab. 1 shows the PCK and the number of training videos of each action class. We see that actions with holistic joint motions (e.g. baseball pitch) are more challenging for pose forecasting and thus have lower PCK values even with more training samples, while actions with only partial joint motions (e.g. jump rope) are the opposite.

3. Additional Qualitative Results

We show additional qualitative examples of the forecasted poses in Fig. 2, 3, and 4. Note that the rendered human model also improves the interpretability of the output 3D poses over skeletons. The second example in Fig. 4 shows a failure case of 3D pose recovery. While the forecasted motion of the tennis serve looks plausible in 2D (row 2), the recovered 3D poses are unrealistic in their body configurations (row 4 and 5), which may be difficult to perceive in the visualizations of 3D skeletons (row 3). All qualitative results can also be viewed as videos at http://www.umich.edu/~ywchao/image-play/.

References

[1] W. Chen, H. Wang, Y. Li, H. Su, Z. Wang, C. Tu, D. Lischinski, D. Cohen-Or, and B. Chen. Synthesizing training images for boosting human 3d pose estimation. In *3DV*, 2016. 1, 2

Timestep #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	# Tr
Baseball pitch	79.7	51.2	37.4	30.3	26.3	23.6	22.2	21.5	20.8	20.6	20.5	20.7	20.8	20.7	20.6	20.5	94
Baseball swing	81.2	69.0	54.9	46.7	42.3	40.2	39.1	38.7	38.8	38.9	38.7	38.9	39.0	38.8	38.8	38.7	104
Bench press	69.1	60.6	52.6	50.1	48.8	48.7	48.9	49.3	49.9	50.5	51.3	52.1	52.9	53.6	54.1	54.3	63
Bowl	68.8	53.1	41.1	34.9	31.7	30.0	28.9	28.4	27.7	27.3	27.0	26.9	26.9	27.0	26.9	27.0	123
Clean and jerk	87.5	60.1	52.7	47.9	44.6	41.6	39.9	38.5	38.0	37.5	37.1	36.8	36.9	37.0	37.1	37.1	39
Golf swing	82.1	68.7	59.4	54.2	51.6	50.3	49.8	49.3	48.6	47.5	47.3	47.6	48.0	48.0	47.8	47.6	81
Jump rope	83.6	69.4	60.6	61.1	65.4	69.2	65.6	61.9	62.2	64.9	66.1	64.6	64.2	65.6	67.2	67.6	36
Jumping jacks	85.0	63.9	47.1	41.3	40.7	42.9	46.7	50.0	52.6	53.9	55.4	57.9	60.5	62.9	64.9	65.5	51
Pullup	81.4	65.7	50.9	44.3	42.1	42.3	43.4	44.8	46.7	48.8	50.8	52.5	54.4	55.7	56.4	56.5	89
Pushup	73.3	65.5	57.5	53.1	51.4	51.3	51.9	53.2	54.9	56.6	58.4	60.1	61.6	62.7	63.2	63.2	94
Situp	67.1	48.0	41.6	38.9	37.6	37.1	37.4	38.0	39.0	39.6	40.4	41.2	41.8	42.3	42.6	42.8	45
Squat	81.3	58.4	46.1	42.3	40.8	41.1	42.3	43.7	45.5	47.4	49.3	51.2	53.0	54.8	56.0	56.0	104
Strum guitar	62.4	61.6	61.5	61.2	61.1	61.6	61.1	60.7	60.3	60.2	59.7	59.2	58.6	58.5	58.4	58.3	42
Tennis forehand	80.9	59.3	40.8	31.7	27.4	24.7	22.9	22.0	21.0	20.5	20.1	19.9	19.8	19.7	19.7	19.6	73
Tennis serve	78.8	56.4	41.3	34.1	29.5	26.4	24.3	22.8	21.6	20.7	20.3	20.0	20.0	20.3	20.3	20.2	104

Table 1: PCK@0.05 of 3D-PFNet on individual action classes.

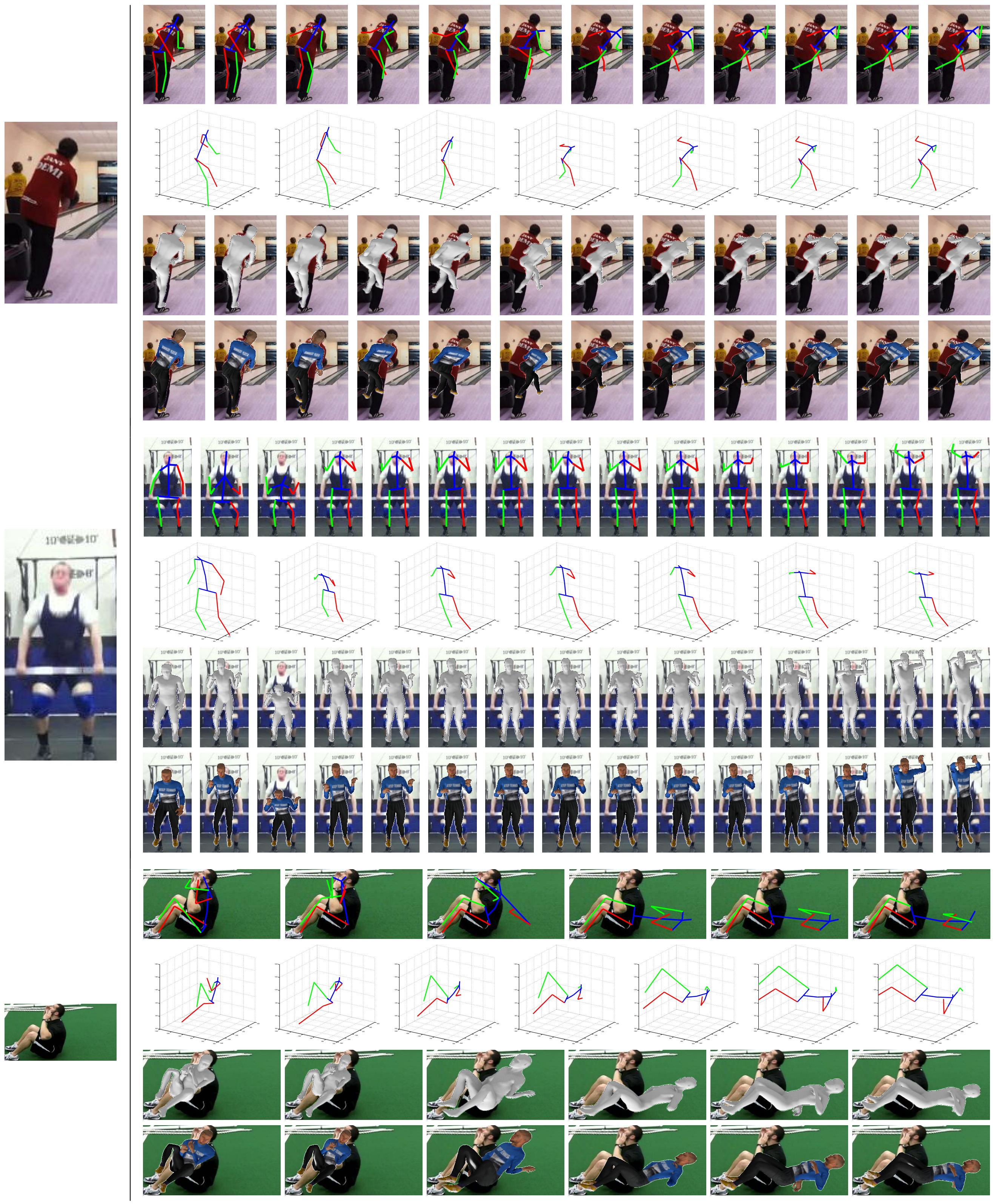

Figure 1: Rendering human characters from the forecasted 3D skeletons. The left column shows the input images. For each input image, we show in the right column our forecasted pose sequence in 2D (row 1) and 3D (row 2), and the rendered human body without texture (row 3) and with skin and cloth textures (row 4). We use the rendering code provided by [1].

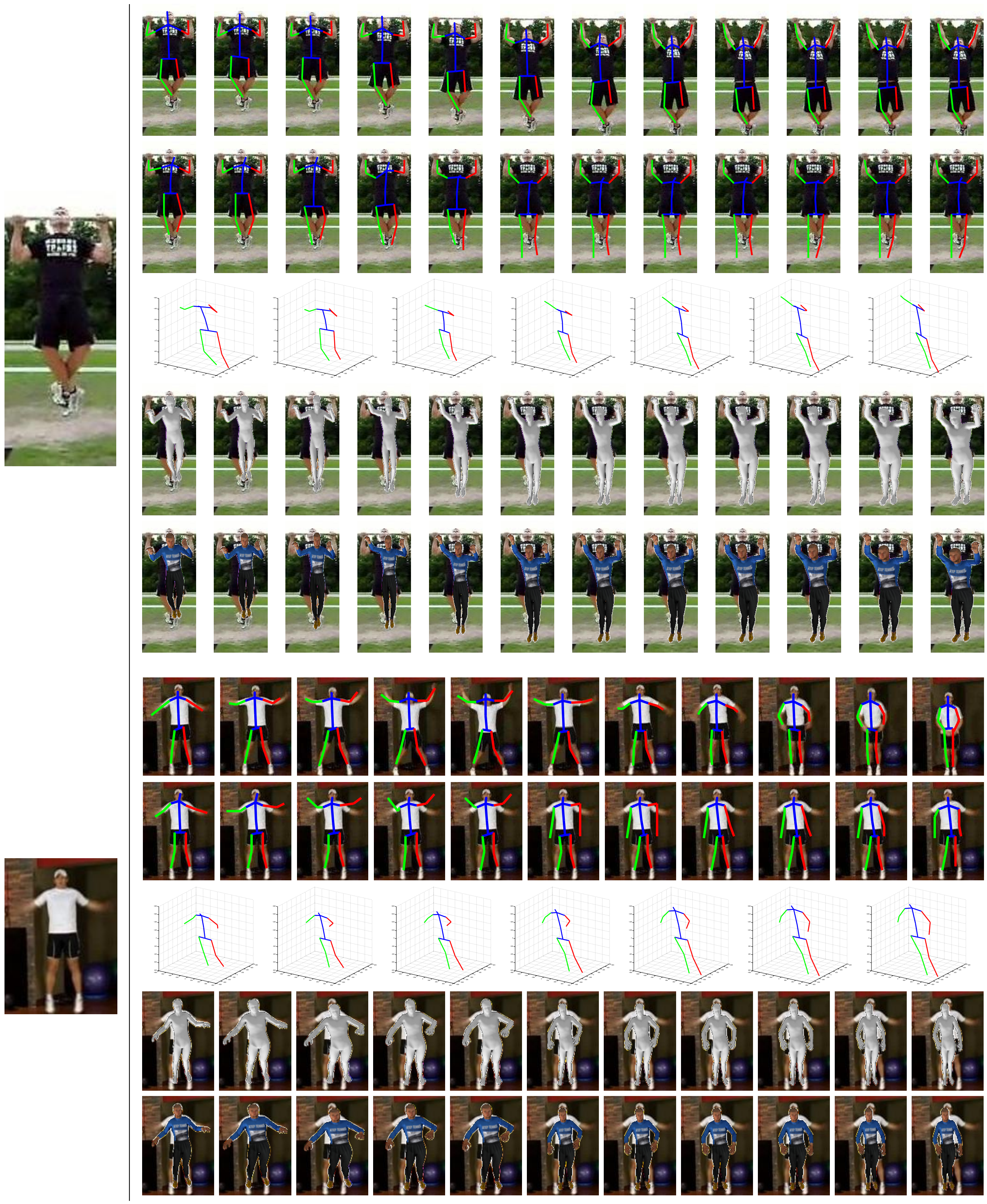

Figure 2: Additional qualitative results of pose forecasting. The left column shows the input images. For each input image, we show in the right column the sequence of ground-truth frame and pose (row 1), our forecasted pose sequence in 2D (row 2) and 3D (row 3), and the rendered human body without texture (row 4) and with skin and cloth textures (row 5).

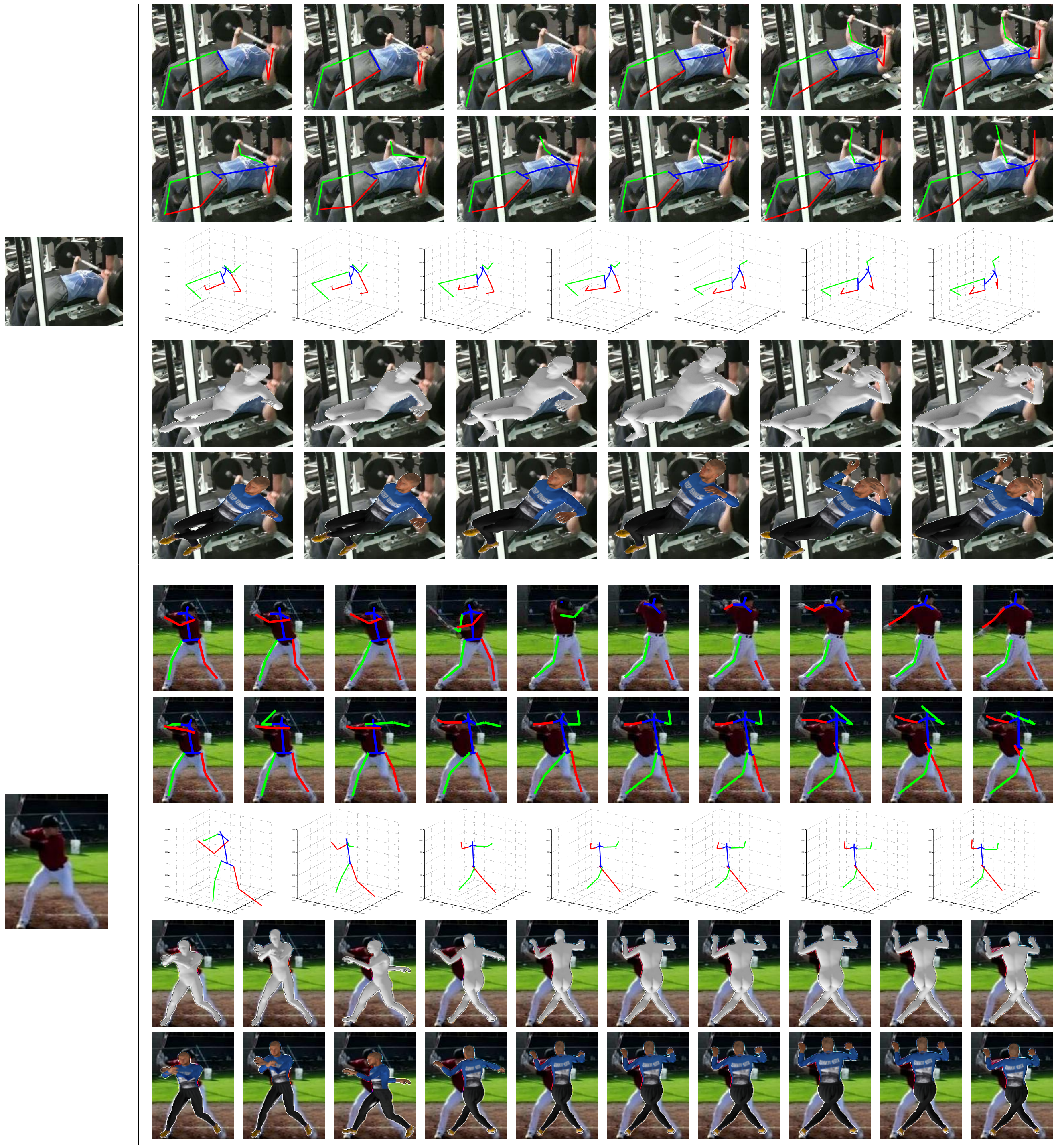

Figure 3: Additional qualitative results of pose forecasting. The left column shows the input images. For each input image, we show in the right column the sequence of ground-truth frame and pose (row 1), our forecasted pose sequence in 2D (row 2) and 3D (row 3), and the rendered human body without texture (row 4) and with skin and cloth textures (row 5).

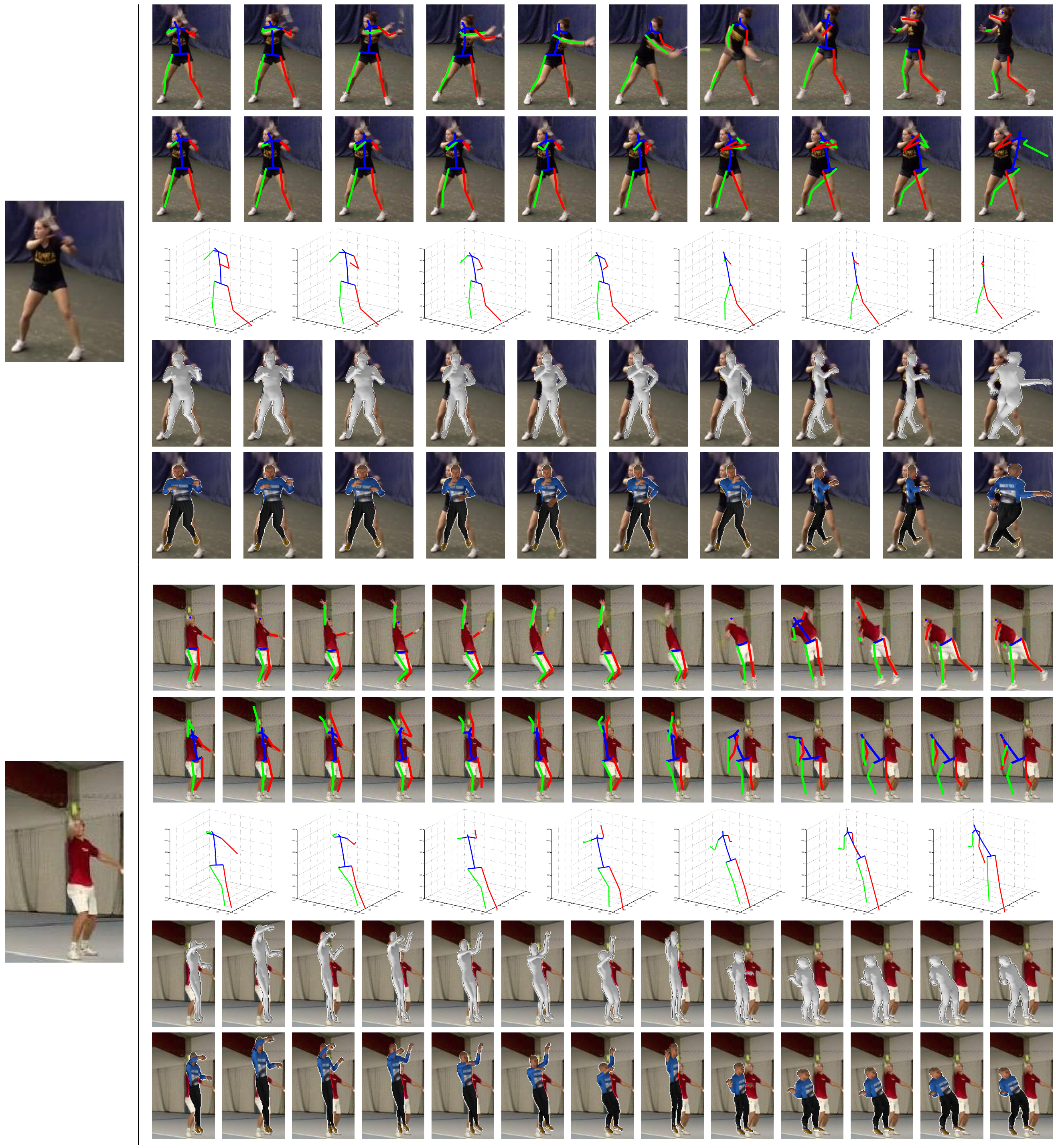

Figure 4: Additional qualitative results of pose forecasting. The left column shows the input images. For each input image, we show in the right column the sequence of ground-truth frame and pose (row 1), our forecasted pose sequence in 2D (row 2) and 3D (row 3), and the rendered human body without texture (row 4) and with skin and cloth textures (row 5).