

Synthetic Human Model Dataset for Skeleton Driven Non-rigid Motion Tracking and 3D Reconstruction

Shafeeq Elanattil, Peyman Moghadam

Robotics and Autonomous Systems, CSIRO Data61, Brisbane, Australia
Queensland University of Technology, Brisbane, Australia
{shafeeq.elanattil, peyman.moghadam}@data61.csiro.au

Abstract

We introduce a synthetic dataset for evaluating non-rigid 3D human reconstruction based on conventional RGB-D cameras. The dataset consist of seven motion sequences of a single human model. For each motion sequence per-frame ground truth geometry and ground truth skeleton are given. The dataset also contains skinning weights of the human model. More information about the dataset can be found at: <https://research.csiro.au/robotics/our-work/databases/synthetic-human-model-dataset/>

1 Introduction

Volumetric 3D reconstruction for rigid scenes and objects is a well studied problem in computer vision and robotics [1, 2, 3]. Often reconstructed 3D maps are fused with other complementary modalities such as RGB information [4], non-visible imaging information such as thermal-infrared [5, 6, 7] or sound [8] for application such as medical imaging [9], disaster response [10] and energy auditing [11].

The more general scenario, where the objects or scenes are dynamic and undergo non-rigid deformation, is still a challenge to be solved [12]. There are only a few publicly datasets available for evaluating RGB-D based

non-rigid 3D reconstruction. Those datasets [13] are for general non-rigid subjects and not specific to human bodies. Even though the dataset published with [12] has the frame-to-frame live ground truth geometry and camera trajectory [14], they do not have ground truth skeleton joints and have very small non-rigid motion. We found that accurate skeleton joints play an important role in human performance capture algorithms. Motivated by this we developed a synthetic dataset which contains ground truth geometry and skeleton joints. The dataset contains human motion sequences which possess high frame-to-frame non-rigid motion.

2 Differences with previously published dataset

Our previous published dataset [14] has frame-to-frame live ground truth geometry and camera trajectory as the corresponding work [12] targets for usage of camera pose for non-rigid reconstruction. Frame-to-frame non-rigid motion in that sequences is very small. In the current publication [15] we are more focused on reconstructing non-rigid movements of human subjects like boxing, jumping etc.

These sequence possess high frame-to-frame non-rigid movements. This dataset also provides ground truth skeleton joints along with per frame ground truth geometry. In addition we are using a human model for motion tracking. The dataset also contains our human model with skinning weight information.

3 The Dataset

Our dataset [16] consists of

1. Ground truth of human 3D geometry at each frame in world coordinate frame.
2. Ground truth of skeleton points at each frame in world coordinate frame.
3. Extrinsic parameters of RGB and depth cameras
4. RGB and depth images.

Name	N	Mean	Min	Max	Std
Jump Balance	60	0.988	0.263	2.605	0.629
Punch Strike	250	0.444	0.084	0.938	0.201
Boxing	245	0.650	0.015	1.589	0.312
Sword Play	248	0.521	0.082	1.165	0.252
Exercise	248	0.733	0.068	1.919	0.456
Kick Ball	161	0.536	0.030	2.752	0.607
Direct Traffic	251	0.578	0.126	1.912	0.260

Table 1: Details of the synthetic data. Each row have sequence name, number of frames in sequence (N), and statistics of joint motion are given.

The dataset consist of seven motion sequences of varying motions characteristics. Table 1 shows motion statistics of the corresponding data sequences. The motion is estimated as the sum of joint movement in each frame. We assign the same name as used in the CMU Mocap dataset for each sequence. The first two columns in Table 1 show name and number of frames in the sequence. The remaining columns shows the motion statistics for each data sequence.

Elanattil *et al.* [12] outlines the detail of the design and production of this synthetic dataset.

3.1 Data Description

The dataset consists of eight folders in which seven of them contains motion sequence data and remaining one contains our human model data. Each motion sequence folders named as the corresponding sequence name as shown in Table 1. Each sequence folder is structured as follows.

```

| color
|_ depth
|_ gt
|_ skeleton
|_ transformation.txt

```

Each motion sequence folder contains color, depth, gt, skeleton sub-folders and a transformation text file. The color and depth folders contains RGB and depth images respectively. The file names have the following form:

- **frame_XXXXX.png**: the RGB image of the scene;
- **depth_XXXXX.png**: the depth image;

The gt and skeleton folders contains ground truth mesh and ground truth skeleton joints correspond to each frame. The file name has the form:

- **mesh_XXXXX.ply**: mesh file of human in world co-ordinates;
- **skeletonCSV_XXXXX.csv**: skeleton joints of the human in camera co-ordinates;

where **XXXXX** is an integer number representing the frame number within the data sequence. In each skeleton file, joint locations of each skeleton bones are given. The transformation.txt file contains the world to camera transformation matrix.

3.2 Camera Parameters

In our synthetic environment the RGB and depth cameras are placed in same location and orientation. Therefor the extrinsic matrix between RGB and Depth cameras is 4×4 identity matrix. The intrinsic parameter matrix for both RGB and depth images is

$$\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1050 & 0 & 480 \\ 0 & 1050 & 480 \\ 0 & 0 & 1 \end{bmatrix}.$$

3.3 Human Model

Our human model is a mesh with 17,021 vertices and 31,492 faces. Skinning bone indexes and weights of each vertices are given in **labels.csv** and **weights.csv** files of the model_information folder. Our model is based on 14 bone skeleton. The skeleton bone indexes and corresponding bone names are shown in Figure 1. The Figure 2 shows our human model which is color coded based on the highest skinning weight for the corresponding skeleton bone. Note that the color coding used for skeleton (Figure 1) and human model (Figure 2) is different.

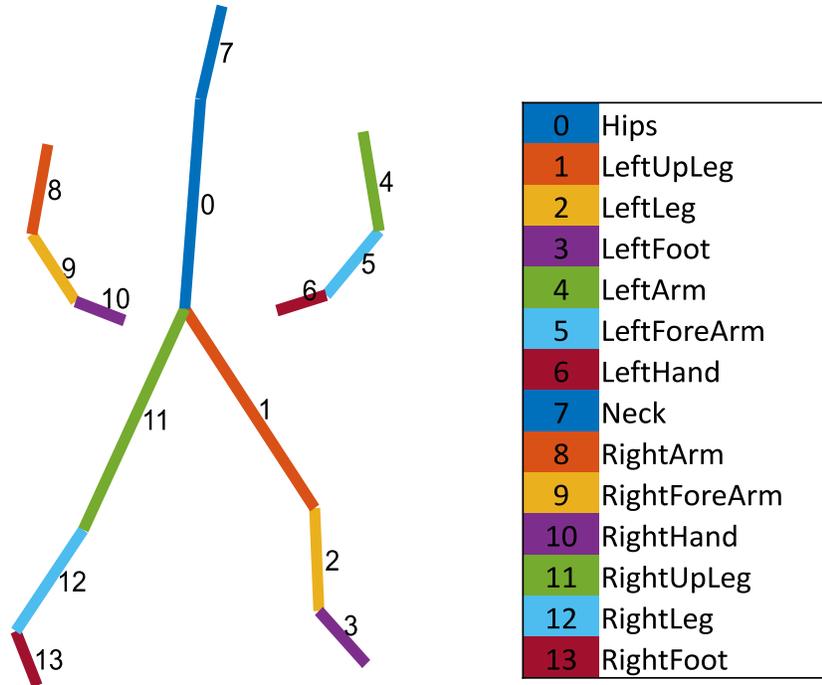


Figure 1: The skeleton used in our human model is shown. The skeleton bone indexes and corresponding bone names are shown.

3.4 Obtaining The Data

This dataset can be downloaded from <https://research.csiro.au/robotics/our-work/databases/synthetic-human-model-dataset/>.

When making use of this data we ask that [15] [16] are cited.

3.5 Example Data Visualizations

Sample data from our dataset is shown in Figures 3 and 4 respectively. Figure 3 shows RGB image, depth image, ground truth mesh and 3D skeleton of a particular frame in 'Boxing' motion sequence. Figure 4 shows RGB images of 'Boxing', 'Exercise' and 'Jumping' motion sequences in the dataset.



Figure 2: The color coded human model is shown. Each point in the mesh is color coded based on the highest skinning weight for the corresponding skeleton bone.

References

- [1] C. Park, P. Moghadam, S. Kim, A. Elfes, C. Fookes, and S. Sridharan, “Elastic LiDAR Fusion: Dense map-centric continuous-time slam,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, May 2018, pp. 1206–1213.
- [2] C. Park, S. Kim, P. Moghadam, J. Guo, S. Sridharan, and C. Fookes, “Robust photogeometric localization over time for map-centric loop closure,” *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1768–

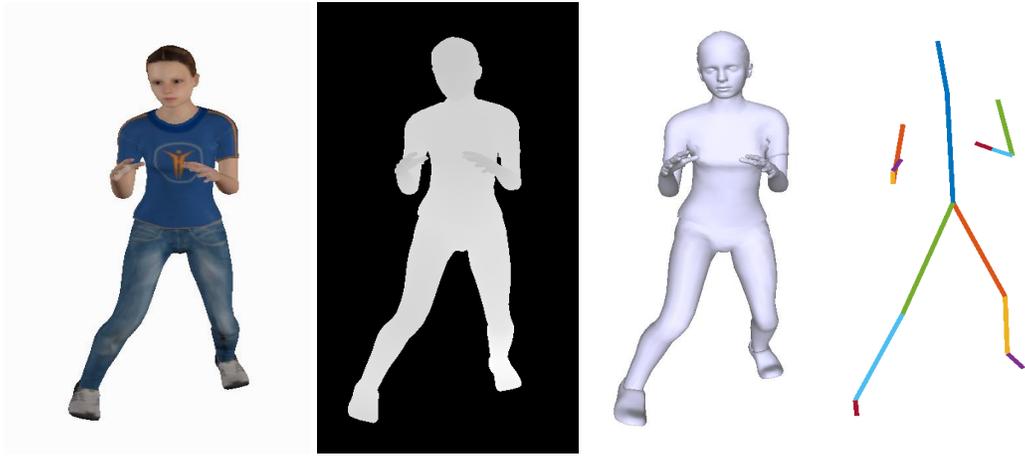


Figure 3: One sample data from 'boxing' sequence is shown. Each frame in the sequence consists of RGB image, depth image, ground truth geometry and 3D skeleton (left to right).

1775, April 2019.

- [3] C. Park, S. Kim, P. Moghadam, C. Fookes, and S. Sridharan, "Probabilistic Surfel Fusion for Dense LiDAR Mapping," *International Conference on Computer Vision Workshops (ICCVW)*, pp. 2418–2426, 2017.
- [4] P. Moghadam, M. Bosse, and R. Zlot, "Line-based extrinsic calibration of range and image sensors," in *International Conference on Robotics and Automation (ICRA)*, May 2013, pp. 3685–3691.
- [5] P. Moghadam and S. Vidas, "HeatWave: the next generation of thermography devices," in *Thermosense: Thermal Infrared Applications XXXVI*, vol. 9105, 2014.
- [6] S. Vidas, P. Moghadam, and M. Bosse, "3D thermal mapping of building interiors using an RGB-D and thermal camera," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2013, pp. 2311–2318.
- [7] S. Vidas, P. Moghadam, and S. Sridharan, "Real-time mobile 3D temperature mapping," *IEEE Sensors Journal*, vol. 15, no. 2, pp. 1145–1152, Feb 2015.

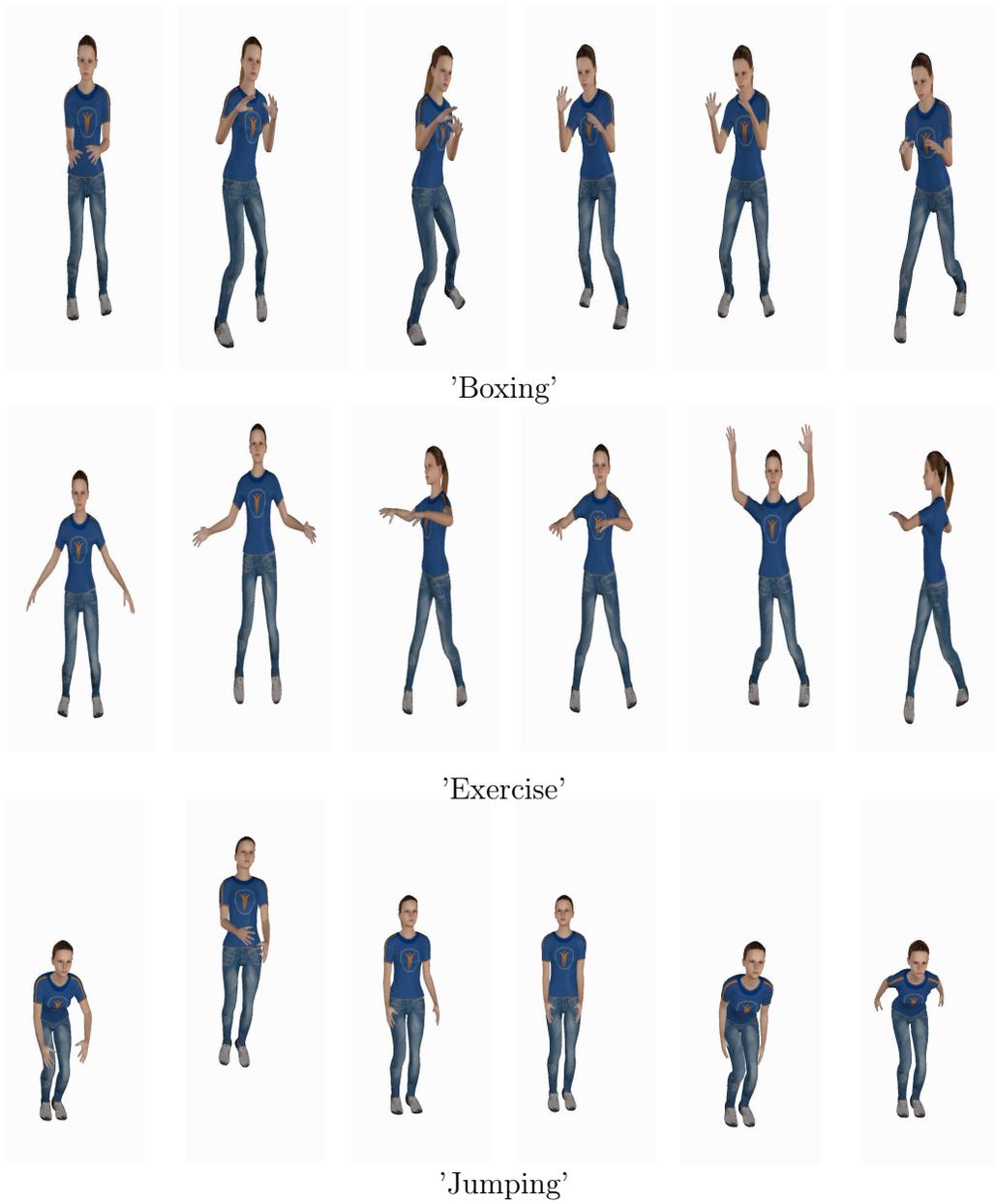


Figure 4: RGB images from the three motion sequences: 'Boxing', 'Exercise' and 'Jumping' are shown from top to bottom.

- [8] P. Moghadam, B. Evans, and E. Duff, “Sage: Semantic annotation of georeferenced environments,” *Journal of Intelligent & Robotic Systems*, vol. 83, no. 3-4, pp. 635–648, 2016.
- [9] P. Moghadam and S. Vidas, “HeatWave: the next generation of thermography devices,” in *Thermosense: Thermal Infrared Applications XXXVI*, vol. 9105. International Society for Optics and Photonics, 2014, p. 91050F.
- [10] K. Nagatani, K. Otake, and K. Yoshida, “Three-dimensional thermography mapping for mobile rescue robots,” in *Field and Service Robotics*. Springer, 2014, pp. 49–63.
- [11] S. Vidas and P. Moghadam, “Heatwave: A handheld 3D thermography system for energy auditing,” *Energy and buildings*, vol. 66, pp. 445–460, 2013.
- [12] S. Elanattil, P. Moghadam, S. Sridharan, C. Fookes, and M. Cox, “Non-rigid reconstruction with a single moving RGB-D camera,” in *International Conference on Pattern Recognition (ICPR)*, 2018.
- [13] M. Slavcheva, M. Baust, D. Cremers, and S. Ilic, “Killingfusion: Non-rigid 3 D reconstruction without correspondences,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 3, no. 4, 2017, p. 7.
- [14] S. Elanattil and P. Moghadam, “Synthetic data for non-rigid 3D reconstruction using a moving RGB-D camera,” in *CSIRO. Data Collection*, 2018. [Online]. Available: <https://doi.org/10.25919/5b7b60176d0cd>
- [15] S. Elanattil, P. Moghadam, S. Denman, S. Sridharan, and C. Fookes, “Skeleton driven non-rigid motion tracking and 3D reconstruction,” in *2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2018, pp. 1–8.
- [16] S. Elanattil and P. Moghadam, “Synthetic human model dataset for skeleton driven non-rigid motion tracking and 3D reconstruction,” in *CSIRO. Data Collection*, 2019. [Online]. Available: <https://doi.org/10.25919/5c495488b0f4e>