UTD Multi-view Action Dataset

The UTD Multi-view Action Dataset was collected for a study of view-invariance action recognition. This dataset contains data from a Kinect v2 camera and a wearable inertial sensor that is being made available for public use.

The Microsoft Kinect v2 sensor (see Fig. 1(a)) depth images are of size 512 × 424 pixels. The frame rate is approximately 30 frames per second. The Kinect Windows SDK 2.0 software package is used which allows tracking 25 human skeleton joints as shown in Fig. 1(c).

The wearable inertial sensor used for data collection is a small size (1”×1.5”) wireless inertial sensor (see Fig. 1(b)) built in the Embedded Signal Processing (ESP) Laboratory at Texas A&M University. This sensor captures 3-axis acceleration and 3-axis angular velocity which are transmitted wirelessly via a Bluetooth link to a laptop/PC. The sampling rate of the inertial sensor is 50Hz and its measuring range is ±8g for acceleration and ±1000 degrees/second for rotation.

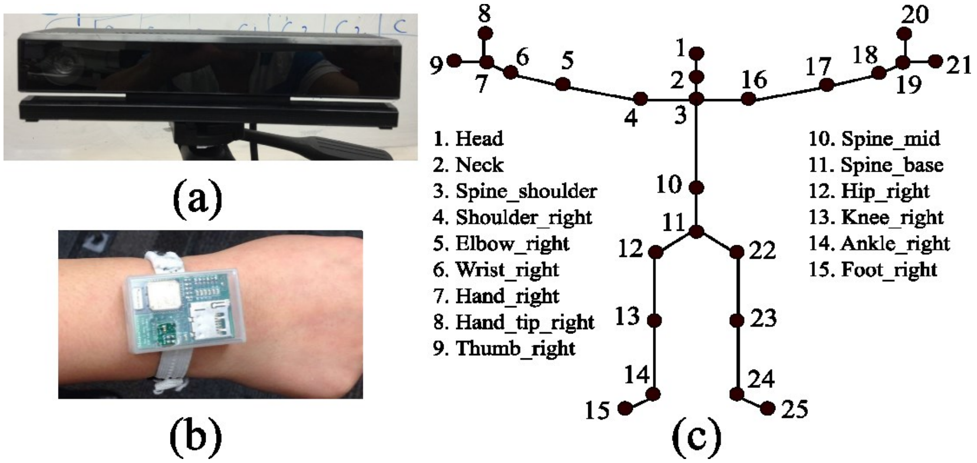

Fig. 1. (a) Microsoft Kinect v2 camera. (b) Wearable inertial sensor. (c) 3D skeleton with 25 tracked skeleton joints.

The dataset was collected by using a Kinect depth camera v2 and the above inertial sensor simultaneously. The inertial sensor was placed on the right wrist of subjects. The dataset included the following six actions: *catch*, *draw circle*, *draw tick*, *draw triangle*, *knock* and *throw.* These six actions were chosen from our UTD-MHAD dataset due to their similarities thus making the action recognition problem more challenging. Five subjects were asked to perform each action with five different subject orientations or views as shown in Fig. 2. For each view, a subject repeated an action 6 times (i.e., 6 trials). Therefore, in total 900 action samples were generated.

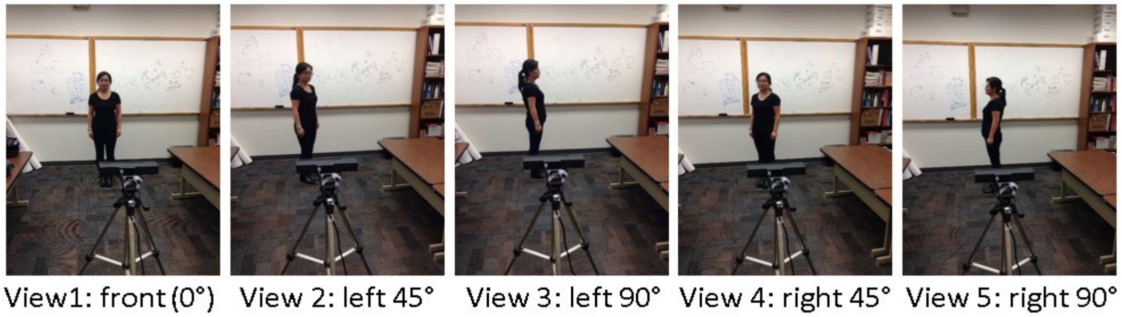

Fig. 2. Five different standing positions of a subject with respect to the Kinect depth camera.

.

Depth images and skeleton joints positions from the Kinect depth camera and accelerations and angular velocities from the inertial sensor were recorded. For the depth data, the size of a depth sequence is 424 x 512 x number_of_frame. For the skeleton data, the 25 joints positions in the 3D coordinates were recorded and the screen coordinates were mapped to the depth images. “skel.world” and “skel.screen” indicate these two types of joint coordinates. The size of “skel.world” is 25 x 3 x number_of_frame, and the size of “skel.screen” is 25 x 2 x num_of_frame. Each row corresponds to the positions of a particular skeleton joint. The order of the joints is as follows:

Row	Joint
1	Base of the spine
2	Middle of the spine
3	Neck
4	Head
5	Left shoulder
6	Left elbow
7	Left wrist
8	Left hand
9	Right shoulder
10	Right elbow
11	Right wrist
12	Right hand
13	Left hip
14	Left knee
15	Left ankle
16	Left foot
17	Right hip
18	Right knee
19	Right ankle
20	Right foot
21	Spine at the shoulder
22	Tip of the left hand
23	Left thumb
24	Tip of the right hand
25	Right thumb

The size of the inertial sensor data is number_of_sample x 6. The six columns (from the first column to the last column) correspond to X-axis acceleration, Y-axis acceleration, Z-axis acceleration, X-axis angular velocity, Y-axis angular velocity, and Z-axis angular velocity.

http://www.utdallas.edu/~kehtar/MultiViewDataset.zip