CSCI473/573 Human-Centered Robotics
Project 3: Robot Understanding of Human Behaviors Using Skeleton-Based Representations

Project Assigned: April 10
Multiple due dates (all materials must be submitted to Canvas)
Deliverable 1 (Representation Code) due: April 24, 23:59:59
Deliverable 2 (Complete Code) due: May 3, 23:59:59

In this project, students will implement several skeleton-based representations (Deliverable 1) and use Support Vector Machines (SVMs) (Deliverable 2) to classify human behaviors using a public activity dataset collected from a Kinect V1 sensor. Additionally,students are required to write a report following the format of standard IEEE robotics conferences using LaTeX.

Students are required to program this project using **Python 3**.

Before you start this project, you need to understand the related content in the data featurization lectures (for Deliverable 1) and SVM classification lectures(for Deliverable 2).

I. DATASET

The MSR Daily Activity 3D dataset[1] will be used in this projet, which was one of the most widely applied benchmark dataset in human behavior understanding tasks. This dataset contains 16 human activities, as demonstrated in Figure 2, performed by 10 human subjects. Each subject performs each activity twice, once in a standing position, and the other in a sitting position. Although this dataset contains both color-depth and skeleton data, in this project, we will only explore the skeletal information to construct **skeleton-based human representations** (Deliverable 1).

The skeleton in each frame contains 20 joints, as illustrated by Figure 1. The correspondence between joint names and joint indices is also presented in the figure. For example, Joint #1 is HipCenter and Joint #18 is KeeRight, etc.

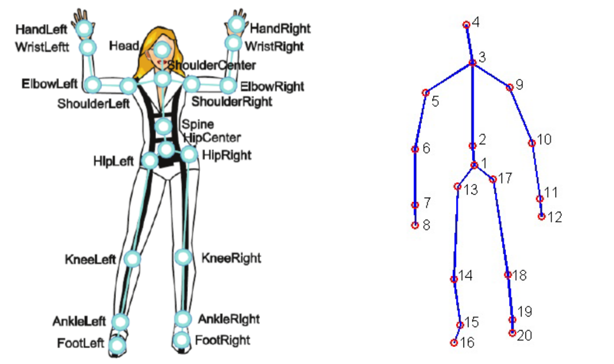
In this project, a pre-formatted skeleton data will be used, which can be downloaded from Canvas

This dataset in this project is a subset of original dataset, which only contains six (6) activity categories:

- CheerUp (`a08`)
- TossPaper (`a10`)
- LieOnSofa (`a12`)
- Walk (`a13`)
- StandUp (`a15`)
- SitDown (`a16`)

This write-up is prepared using LaTeX.

[1]The MSR Daily Activity 3D dataset is removed from the author's website and no longer publicly available after 2017.



(a) Names (b) Indices

Fig. 1: Skeleton joint names and indices from Kinect SDK.

In particular, the dataset contains two folders: `Train` and `Test`. Data instances in the directory `Train` is used for training (and validation). Data instances in `Test` is used for testing the performance. Each instance has a filename like: `a12_s08_e02_skeleton_proj.txt`. This filename means the data instance belongs to activity category 12 (i.e., `a12`, that is "lie down on sofa" as in Figure 2), from human subject 8 (`s08`) at his/her second trial (i.e., `e02`). The dataset contains 16 activity categories, 10 subjects, and 2 trials each subject. Instances from subjects 1–6 are used for training (in the directory `Train`), and instances from subjects 7–10 are used for testing (`Test`).

When you open a data instance file, e.g., `a12_s08_e02_skeleton_proj.txt`, you will see the following:

1	1	0.326	-0.101	2.111
1	2	0.325	-0.058	2.152
1	3	0.319	0.194	2.166

.
.
.

Each *row* contains five values, representing:

1) `frame_id`,
2) `joint_id`,
3) `joint_position_x`,
4) `joint_position_y`,

Algorithm 1: RAD representation using star skeletons

Input : Training set `Train` or testing set `Test`
Output : `rad_d1` or `rad_d1.t`

for *each instance in* `Train` *or* `Test` **do**
  for *frame t = 1, ..., T* **do**
    Select joints that form a star skeleton (Figure 3);
    Compute and store distances between body extremities to body center ($d_1^t, ..., d_5^t$);
    Compute and store angles between two adjacent body extremities ($\theta_1^t, ..., \theta_5^t$);
  end
  Compute a histogram of N bins for each $\boldsymbol{d}_i = \{d_i^t\}_{t=1}^T$, $i = 1, ..., 5$;
  Compute a histogram of M bins for each $\boldsymbol{\theta}_i = \{\theta_i^t\}_{t=1}^T$, $i = 1, ..., 5$;
  Normalize the histograms by dividing T to compensate for different number of frames in a data instance;
  Concatenate all normalized histograms into a one-dimensional vector of length $5(M + N)$;
  Convert the feature vector as a single line in the `rad_d1_train.txt` or `rad_d1_test.txt` file.
end
return *`rad_d1_train.txt` or `rad_d1_test.txt`*

5) `joint_position_z`.

Each frame contains 20 rows that contain information of all joints in the frame.

II. CSCI 473: Deliverable 1 (Representation Construction)

Students in CSCI 473 must implement two skeleton-based representations during the Deliverable 1.

A. Relative Distances and Angles of Star Skeleton

Students in CSCI 473 are required to implement the human representation based on the **Relative Angles and Distances (RAD) of star skeleton**, as described by Algorithm 1. The objective is to implement the RAD representation to convert all data instances in the folder `Train` into a single training file `rad_d1_train.txt`, each line corresponding the RAD representation of a data instance. Similarly, all instances in the folder `Test` needs to be converted into a single testing file `rad_d1_test.txt`.

B. Customized Representations

Implement a customized skeleton-based representation by choosing different joints other than the joints selected in the star skeleton. For example, you can change reference joints, select other joints other than body extremities, or compute distances of all joints but ignore the orientation information. Your code is required to output a single training file `cust_d1_train` for all training instances, with each row containing the customized representation of an instance, and a single testing file `cust_d1_test`, similar to the task in Section II-A.

C. What to Submit

For Deliverable 1, CSCI 473 students are required to submit a **single tarball** named *D1_firstname_lastname.tar* (or .tar.gz) to the portal named "P3-D1" in Canvas, which must contain the following items:

- A README that provides sufficient instructions needed to compile and execute your code. Your README also needs to document your implementation information, including which joints are used in the RAD representation, how the histograms are computed, and how many bins are used.
- Your code to construct the RAD and customized representations.
- The generated representation data, including `rad_d1_train.txt`, `rad_d1_test.txt`, `cust_d1_train.txt`, and `cust_d1_test.txt`.
- A 1-page report that includes:
  - Section introducing the problem
  - Section discussing the RAD featurization
  - Section discussing the custom featurization you used
  - Section discussing histogram creation

Students are allowed to include a local copy of the training and testing sets within the code directory to make your code self-contained.

III. CSCI 573: Deliverable 1 (Representation Construction)

CSCI 573 students are required to implement three specific skeleton-based representations for Deliverable 1, including RAD, HJPD, and HOD.

A. Relative Distances and Angles of Star Skeleton

Students in CSCI 573 are required to implement the human representation based on the **Relative Angles and Distances (RAD) of star skeleton**, as described by Algorithm 1. The objective is to implement the RAD representation to convert all data instances in the folder `Train` into a single training file `rad_d1_train.txt`, each line corresponding the RAD representation of a data instance. Similarly, all instances in the folder `Test` needs to be converted into a single testing file `rad_d1_test.txt`. This required representation is the same as Section II-A.

B. Histogram of Joint Position Differences (HJPD)

Given the 3D location of a joint (x, y, z) and a reference joint (x_c, y_c, z_c) in the world coordinate, the joint displacement is defined as:

$$(\Delta x, \Delta y, \Delta z) = (x, y, z) - (x_c, y_c, z_c) \qquad (1)$$

The reference joint can be the skeleton centroid or a fixed joint. For each temporal sequence of human skeletons (in a data instance), a histogram is computed for the displacement along each dimension, i.e., $\Delta x, \Delta y, \Delta z$. Then, the computed histograms are concatenated into a single vector as a feature.

This HJPD representation is similar to the RAD representation, except that it uses all joints and ignores the pairwise

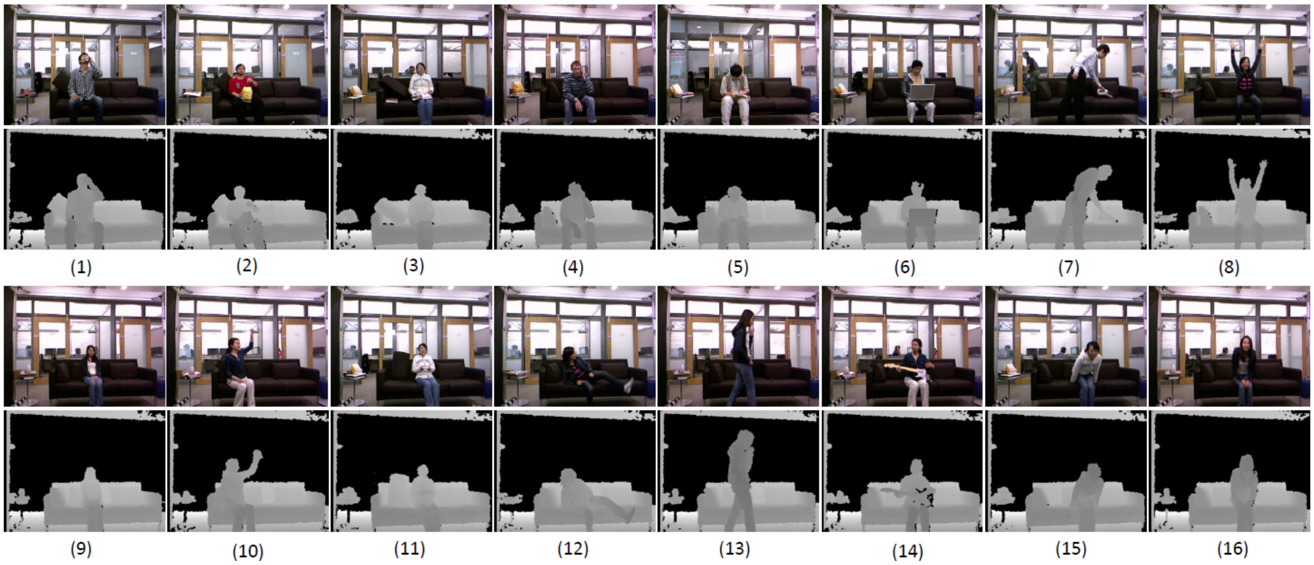


Fig. 2: The full MSR Daily Activity 3D dataset contains sixteen human activities: (1) drink, (2) eat, (3) read book, (4) call cellphone, (5) write on a paper, (6) use laptop, (7) use vacuum cleaner, (8) cheer up, (9) sit still, (10) toss paper, (11) play game, (12) lie down on sofa, (13) walk, (14) play guitar, (15) stand up, (16) sit down.

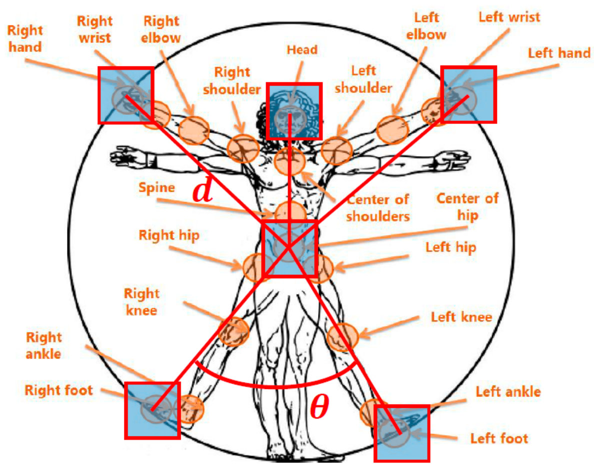


Fig. 3: Illustration of human representation based on relative distance and angles of star skeleton

angles. Refer to Section 3.3 of reference [1] for more details, which is available online at: `https://ieeexplore.ieee.org/document/6836044`.

Your code is required to generate a single training file `hjpd_d1_train.txt` for all training instances, with each row containing the HJPD representation of an instance, and a single testing file `hjpd_d1_test.txt`.

C. Histogram of Oriented Displacements (HOD)

You need to implement the skeleton-based representation of Histogram of Oriented Displacements (HOD), as introduced in Section 3 of reference paper [2], including the technique of Temporal Pyramid. The paper is available publicly at: `https://www.ijcai.org/Proceedings/13/Papers/203.pdf`.

Your code is required to generate a single training file `hod_d1_train.txt` for all training instances, with each row containing the HOD representation of an instance, and a single testing file `hod_d1_test.txt`.

D. What to Submit

For Deliverable 1, students in CSCI 573 are required to submit a **single tarball**, named *D1_firstname_lastname.tar* (or .tar.gz) to the Canvas portal named "P3-D1", which must contain the following items:

- A README that provides sufficient instructions needed to compile and execute your code. Your README also needs to document your implementation information, for example, including which joints are used in the RAD representation, and how the histograms are computed and how many bins are used in your HJPD and HOD representations.
- All your code to construct the RAD, HJPD, and HOD representations.
- All the generated skeleton-based representation data, including `rad_d1_train.txt`, `rad_d1_test.txt`, `hjpd_d1_train.txt`, `hjpd_d1_test.txt`, `hod_d1_train.txt`, and `hod_d1_test.txt`.
- A 1-2 page report that includes:
  - Section introducing the problem
  - Sections discussing each of the featurization methods and their respective histograms

Students are allowed to include a local copy of the training and testing sets within the code directory to make your code-self contained.

IV. Support Vector Machines

(PART OF DELIVERABLE 2)

The second deliverable of this project (in Deliverable 2) is to understand and apply Support Vector Machines (SVM) to enable robot learning in practical applications (i.e., behavior understanding in our project). The write-up of Deliverable 2 will be posted after Deliverable 1 is due.

V. GRADING

The total score of this project 3 deadline 2 is 20 points. Your grade will be based on the quality of your project implementation and the documentation of your findings in the report.

	Excellent (100%)	Needs Work (70%)	Poor(30%)
Code (10 points)	Code runs without issues and is clearly documented	Code contains significant bugs	Code does not run
Write-up (10)	Write-up contains all the necessary sections, is well written and well formatted	Write-up does not contain sufficient information or is not well written	Writeup is not formatted or is missing significant sections of work

TABLE I: Grading rubric for Project 3 deadline 1

Students in CSCI573 will be graded more strictly on the quality of the code implementation and paper presentation. The instructor expects a more through analysis of the experimental results, and a good implementation of skeleton-based representations and robot learning methods for human behavior understanding. The paper should have the “look and feel” of a technical conference paper, with logical flow, good grammar, sound arguments, illustrative figures, etc.

REFERENCES

[1] H. Rahmani, A. Mahmood, D. Q. Huynh, and A. Mian, “Real time action recognition using histograms of depth gradients and random decision forests,” in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2014.

[2] M. A. Gowayyed, M. Torki, M. E. Hussein, and M. El-Saban, “Histogram of oriented displacements (hod): Describing trajectories of human joints for action recognition.,” in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2013.