Skeletons Don't Lie: Can AI Decode Your Moves?

We are given a dataset containing sequences of human movements, each composed of multiple frames (`FrameNumber`). Each frame includes the 3D positions of the 25 body joints, represented by the coordinates: `J1X`, `J1Y`, `J1Z`, ..., `J25X`, `J25Y`, `J25Z`. Each frame is associated with a sequence identifier (`IDSample`) and two labels (`Action` and `Camera`) indicating the performed action and the camera that recorded the sequence.

To understand the spatial distribution of the 25 joint points, you can analyze the following figure:

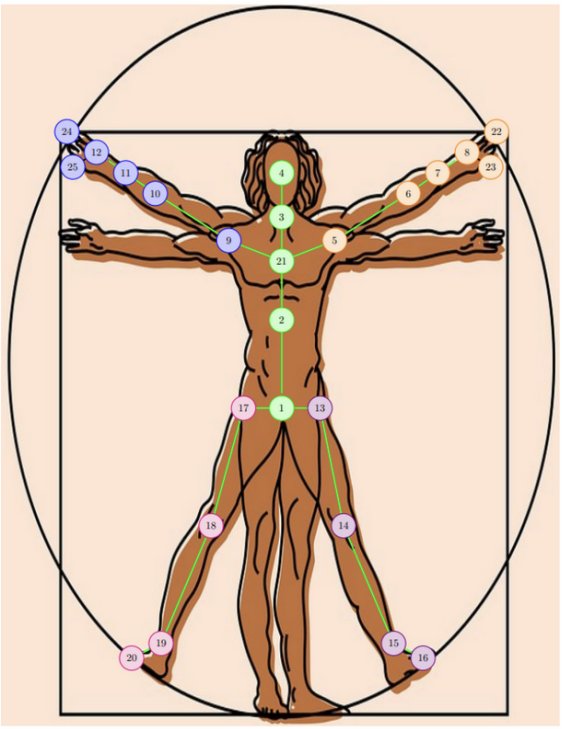

Figure 1: img.png

The possible actions are:

- **0**: wear jacket
- **1**: take off a hat/cap
- **2**: hopping (one foot jumping)
- **3**: jump up
- **4**: falling

The possible cameras are:

- **1**: 45° angle view (left or right)
- **2**: front view
- **3**: side view (left or right)

Subtask 1: Counting frames per sequence (10 points)

For each unique `IDSample`, determine how many frames are available in the test dataset (`test_data.csv`). This analysis helps in understanding the temporal-spatial distribution of the data for each action.

Subtask 2: Sequence-level classification (40 points)

Using the training data, train a classification model that can recognize the action performed in a sequence of frames. Apply the model to the sequences in the test set and specify, for each sequence (set of frames for a unique `IDSample`), which action is predicted by the model as the most probable.

Evaluation metric

The performance of the model will be evaluated using **accuracy**.

This metric expresses the proportion of test sequences for which the model correctly predicted the label (`Action`) associated with the performed action.

To receive full points, the obtained accuracy must be at least 0.965.

Subtask 3: Camera identification (50 points)

Using the training data, train a model capable of predicting, for a sequence of frames, which camera was used to record that sequence. Apply the model to the sequences in the test set and specify, for each `IDSample`, the estimated camera that is most likely.

This subtask aims to assess the ability to understand spatial variations in the data, determined by the camera's position.

Evaluation metric

The performance of the model will be evaluated using **accuracy**.

This metric expresses the proportion of test sequences for which the model correctly predicted the label (`Camera`) associated with the camera that recorded the sequence.

To receive full points, the obtained accuracy must be at least 0.8.

Output format

The final result must be a CSV file named **`output.csv`**, which should contain **exactly 3 columns**:

- `subtaskID` - represents the subtask number (1, 2, 3)
- `datapointID` - refers to the `IDSample` column from `test_data.csv`
- `answer` - the corresponding answer for that datapoint and subtask