CSE 523 ADVANCED PROJECT REPORT, SPRING 2022: QUANTIFYING DANCE MOVEMENTS USING HUMAN POSE ESTIMATION AND TRACKING

Vanshika Sharma

SBU ID: 114365071

ABSTRACT

This project is about taking one of those first steps towards quantifying dance movements based on the space covered by a specific body over a period of time. In this implementation, two things have been measured: first, how much space i.e. the distance a person covers and the kind of movement he or she performs, which could also represent the change in direction or orientation of that person over a period of time. This is executed using a depth camera: ZED 2. These movements are mainly classified into three types and further into their subcategories: Detecting whether a person is moving from side to side and also measuring the distance covered during this, secondly, detecting front and back movements and finally, up and down movements. The subcategories of these movements are discussed in the later sections. These movements are captured and quantified in a 3D space formed in the ZED's field of view.

1. BACKGROUND

1.1. Elements of Dance: Why do we care about space?

Space is the area we move through as we dance. Secondly, shape is considered which is the design of the body, can be open or closed, symmetrical or asymmetrical, angular or curved (individual and group shapes) Third is the focus of the audience (where viewer's eye is drawn) where dancer is the single focus, people look in the direction of movement; Then last but not the least comes: Size - use of size in given space or range of motion, Level - the vertical distance from the floor (high, medium or low), Direction - forward, backwards, up, down, diagonal, Pathways - patterns we make as we move across the floor: straight, curved

Space is the way the dancer occupies the physical world. Art forms like architecture and sculpture take up physical space in three dimensions, but this concept is a little different in dance because the dancer is in motion. So for viewer, the compositional element, space, is the way the dancer moves through and interacts with the physical world.

Humans can't fly or levitate, so how many ways can someone really move through physical space? Quite a few, actually. Within this compositional element, several aspects are considered: the actual direction of the movement (sideways, forwards, backwards, diagonally), as well as implied movement in gestures or placement. The viewer must consider facing, how the front of the body is positioned in relation to the audience. The pattern the dancer moves in as he or she travels through space has to be envisioned. Because this is a three-dimensional art form, we must also consider levels, the relationship of the body to the floor. Do dancers always stay with two feet on the floor? Can they lie on the floor, or leap high above it? All of these movements define the use of space in dance.

1.2. Parameters of Space

Our physical world is defined in a 3 Dimensional system, this is exactly what the ZED 2 camera does for us. It not only defines the space it sees into 3D coordinate system, but also quantifies each coordinate into one unit defining 1 milimeter along X, Y and Z axes. Based on the model we are using, it gives us the exact location of the object/body in this space, with its position as (x,y,z) coordinates. These coordinate information can further be used by various models such as object detection, human tracking, positional tracking, image segmentation, depth sensing, spatial mapping and image capturing. For the task that is being implemented in this project, object detection, human pose estimation and positional tracking components of ZED have been used. Based on 3D space coordinates, we get the specifications of all these three aspects to be integrated and used to get optimal results.

1.3. Experiment Setup

The ZED camera was setup in the lab, and the SDK for this was installed from the Developer's website [1]. The SDK came with an API which integrates ZED with python, and then all the data captured by ZED can be used within python to generate models and get the desired results. The python version used here, is Python 3.6.7, this API currently only supports python 3.6 and 3.7. The hardware I used was my own system: Lenovo Legion 5Pi with NVIDIA GEFORCE RTX 2060 6GB Graphics Card, 16 GB RAM and 1 TB SSD. The ZED Camera, ZED SDK and ZED Python API were then integrated with the python version on my system and was ready to use. The ZED Python API also provides its own models for object detection, human pose estimation and positional tracking, which is used in the program and the information gathered from these is then further used to detect movements and space. Coordinate system of ZED is shown in figure1. This camera has a 120 degree Field of View.

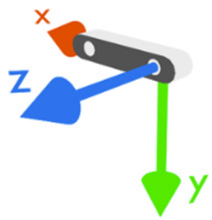


Fig. 1: **ZED 2 Camera Coordinate System**

2. IMPLEMENTATION

This section explains how the quantification of space of movements was carried out and what were the concepts used to implement these two aspects.

2.1. Space

The space covered was calculated based on two things: displacement of a body and distance covered by the body. To implement this, the first task was to detect the body in the scene, if and when a body is detected, its position is stored in the 3D coordinate system by calculating the mid point of the bounding box (not shown in the output), generated by the object detection model. Once, a body is detected and the region of interest in now defined, the body's landmarks are defined, this is done by the human pose estimation model. When we enable a parameter called enableTracking, it also activates the positional tracking system of ZED, and the position of the entire body as well as body part landmarks is tracked across the duration of the video. This tracking stores all the coordinates for each frame of the video as a list, and this list of coordinated is then used to calculate the displacement and distance of the body. For space coverage, only the 3D coordinates of the body are required. All the 3D positions of the body are looped over and the distance calculated between each consecutive frame using the Euclidean Distance Formula is computed and stored in another python list. The sum of these distances is the total distance (space) covered by the body. For displacement, only the first and last coordinates of the body are used in the Euclidean Distance Formula and computed for it. Euclidean Distance is explained in Figure 2 and 3.

2.2. Movements

The movements performed by the body in space are classified into three categories: Side to side movements, front and back movements and up and down movements.

$$d = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2}$$

Fig. 2: **Euclidean Distance Formula**

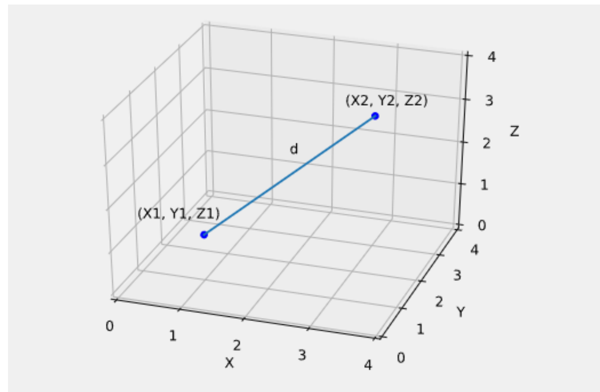


Fig. 3: **Distance in 3D Space**

2.2.1. *Side to Side Movements*

This movement is based on how the body moves in the X-Direction, i.e. its movement along the X-Axis of ZED. ZED uses an image coordinate system which is right-handed with the positive Y-axis pointing down, X-axis pointing right and Z-axis pointing away from the camera. This means that as we go further right (from ZED's perspective) (and left from our perspective), the coordinate value of X increases. Therefore, we can say that when the person facing the camera reaches its maximum value of coordinate while keeping moving along the X-axis until the end of the frame, he/she is moving from Right to Left (from his own perspective), whereas, if he moves from Left to Right, the coordinate value of X would decrease and reach a minimum coordinate in the end of the frame. Based on these concepts, there are four parameters considered while calculating the variations of the side to side movements: Minimum X Coordinate, Maximum X Coordinate, Last X Coordinate, Start X Coordinate. The Variations are explained in figure 4. The distance covered by the person during these movements is calculated by measuring the distance only along X axis.

2.2.2. *Front and Back Movements*

The front and back movements are based on how a person moves along the Z Axis i.e. the axis pointing outwards from the ZED Camera. This axis keeps track of how far or close the person is from the camera. When the person keeps moving towards the camera till the end of the last frame, the Z coordinate reaches a minimum as the coordinates decreases while moving closer and denotes a front movement, whereas

Fig. 4: Some of side to side movement variations

the exact opposite of this denotes a back movement. There is also a third possibility that person moves both backward and forward through the video, this algorithm also records and detects that. These quantification and measurements are based on three parameters: Last Z coordinate, Maximum Z Coordinate and Minimum Z Coordinate. Specific conditions are given in the figure 5. The near and far distances are calculated by calculating the total distance covered along z axis.

2.2.3. Up and Down Movements

These movements happen along the y axis, which is pointed downward in the ZED coordinate system. This means that as we move downward, the coordinate value of y increases, whereas, as we move upward, the coordinate value of y increases. Therefore, when the person reaches his minimum by the end of the frame, this implies that he is moving in the upward direction, whereas, when moving downward, the person reaches his maximum. The variation of these movements and specifications of the conditions are given below. The distance covered during the movement is calculated by measuring the specific distance across y axis. Explained visually in figure 6.

2.2.4. Cross Dimensional Movements

These movements are recorded when a person is detected to be moving along two or more dimensions simultaneously. For example, if a person moves away from the camera and also keeps going to the left. In this case, both X and Z coordinates will change significantly and record high values of distances covered. To classify these movements, there is a minimum threshold value set to each axis movement based on the multiple observations. If the distance covered along an axis is detected to go above this threshold, only then a person is defined to be moving along that axis, otherwise no movement is detected along that axis. This is considered to take care of the sensitivity and error scope of the ZED camera. Therefore, if the threshold values for two axes is crossed, then it qualifies for a cross-dimensional movement. Conditions are mentioned in the code.

3. OBSERVATIONS AND RESULTS

There are three kinds of movements which are considered. The variations in each of them as well as their subcategories are plotted in the graphs. The x-axis is the number of frames passed, for this particular instance it was taken to be 150 with a rate of 10 fps. The y axis shows the distance covered by the person along the three axes and their comparisons with one another. When the person is performing side to side movements, the maximum distance covered is along the x-axis, which makes the x-axis line on top, whereas for front and back movements, the highest values along z-axis, similarly for up and down movements, y-axis distance is the greatest. Plots shown in figure 9.

4. FUTURE WORK

Based on the functionalities implemented in this project, this work can be further used for a lot of extensions. This provides a way to map space with movements, and based on the basic movements implemented here, they can be a gateway to analyze, quantify and predict even more complex movements which are performed by dancers. The way the 3D coordinates have been analyzed for the positioning of the entire body, specific body part landmarks movements can also be tracked for their positioning and velocity in space. The space measured in this work can further be used to map the whole surface area covered by a person moving across the scene and measure how much area he covered along with the distance. These movements quantified and predicted for their behaviour in space can be used to train models and generate a supervised

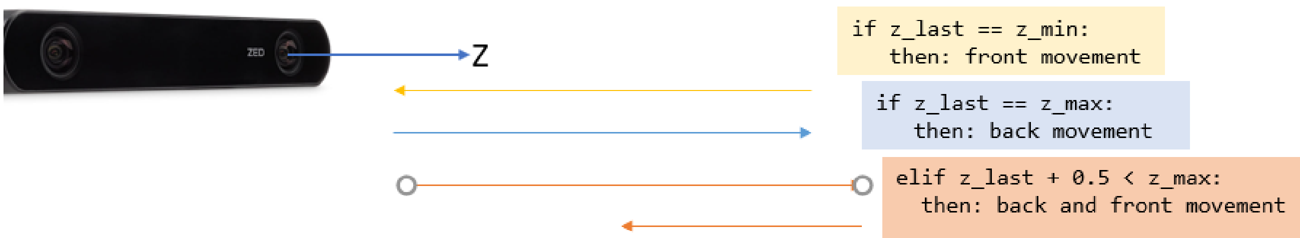


Fig. 5: Front and Back movement variations

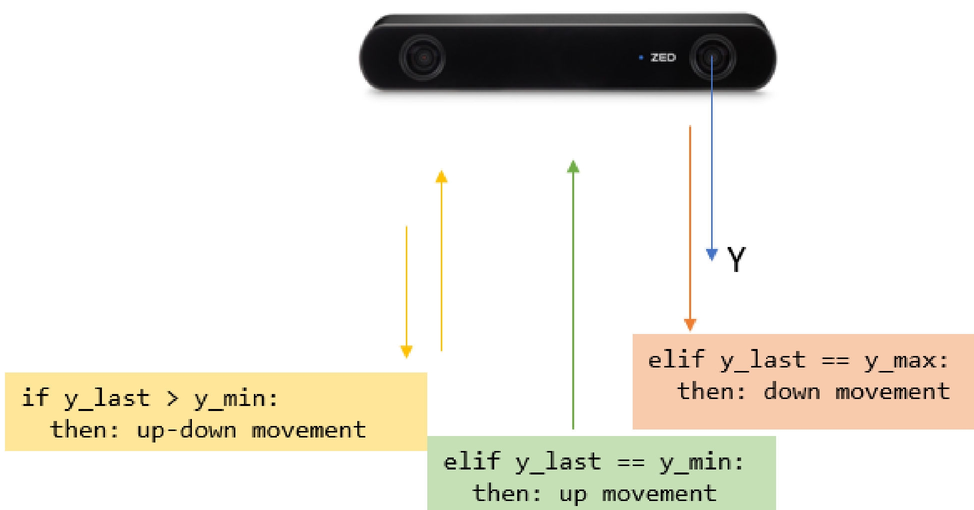
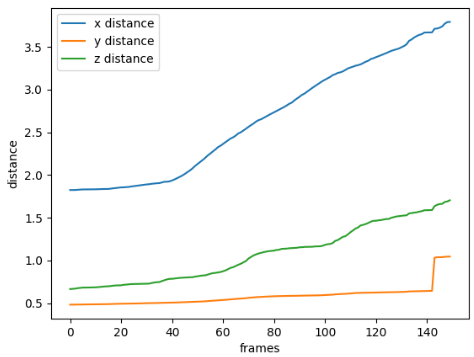
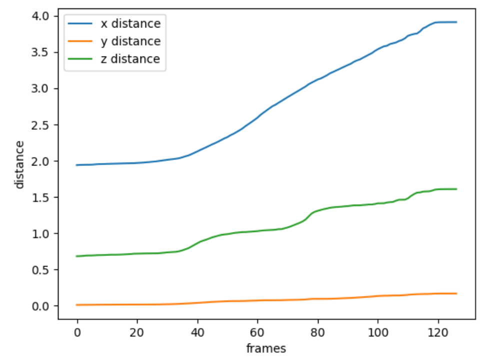


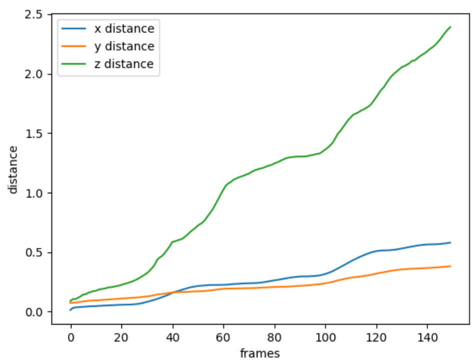
Fig. 6: Up and Down movement variations



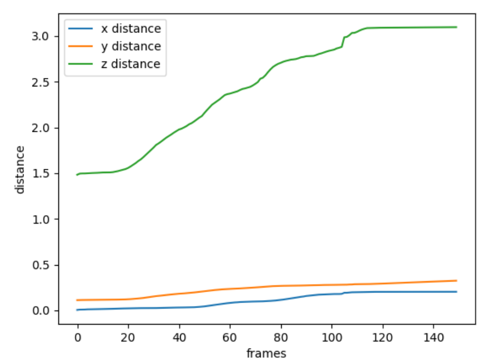
(a) L2R



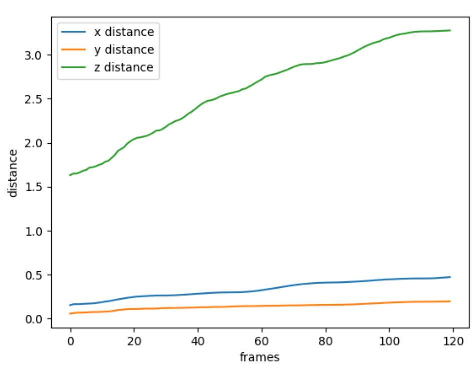
(b) R2L



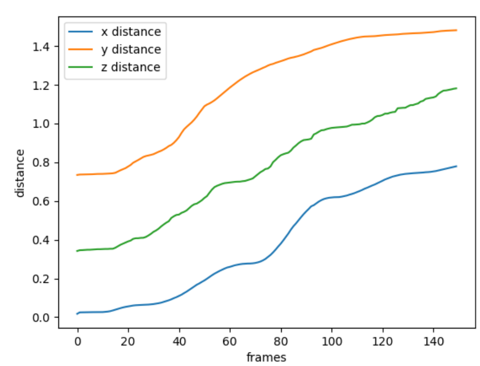
(c) F-B



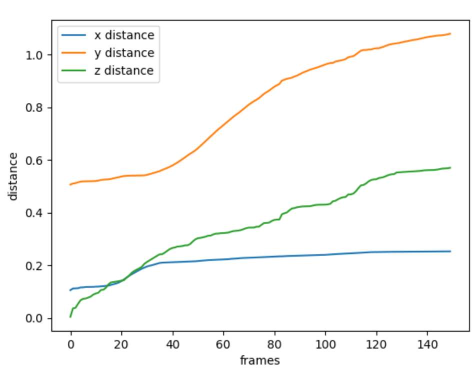
(d) FRONT



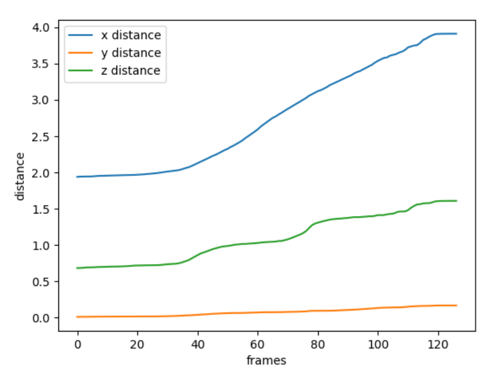
(e) BACK



(f) UP

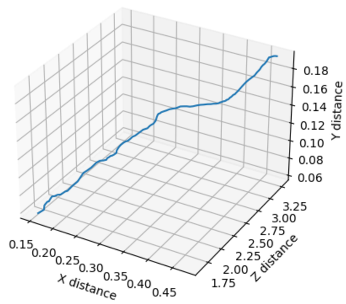


(g) DOWN

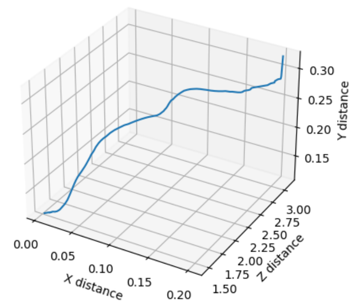


(h) R2LR

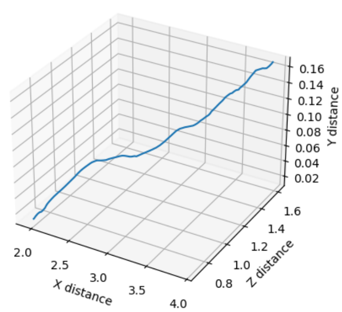
Fig. 7: Plots of Various Movement Variations



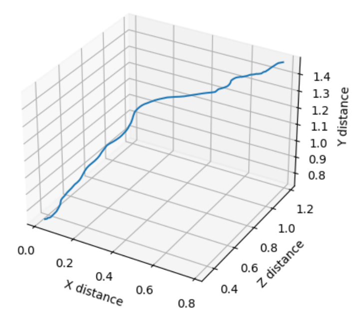
(a) BACK3D



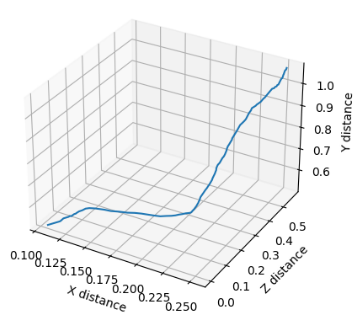
(b) FRONT3D



(c) R2L3D



(d) UP3D



(e) DOWN3D

Fig. 8: Plots of Various Axis Movements wrt Each Other

learning dataset to do this, the coordinate variations and distances mapped for their corresponding movements can act as the training dataset and can be tested upon for unknown coordinate variations and distances. The similar thing can further be done by using the coordinate variation of body part landmarks. These movmements then can also either be used in Laban Movement Analysis or generate a similar model. This model can then be enhanced for its accuracy and precision considering the sensitivity of the ZED camera and the scope of error. The ZED camera's data can be further analysed and the distance tracking and movement recognition can be optimized based on this metrics.

5. REFERENCES

1. StereoLabs reference: https://www.stereolabs.com/docs/positional-tracking/coordinate-frames/
2. ZED: https://www.stereolabs.com/zed-2/
3. ZED SDK: https://www.stereolabs.com/developers/release/
4. Distance Formula: https://en.wikipedia.org/wiki/Euclideandistance
5. Dance Concepts: https://www.nelson.kyschools.us/userfiles/-4/
6. Images: https://www.omnicalculator.com/math/3d-distance
7. PyMC Tutorial: https://docs.pymc.io/en/v3/nbtutorials/index.html
8. Bayesian Network Reference: https://www.bayesfusion.com/bayesian-networks/
9. Movements in Dance: https://kcimc.medium.com/discrete-figures-7d9e9c275c47
10. Dance:https://us.humankinetics.com/blogs/excerpt/elements-of-dance