**Project Report – Age Estimation from Posture**

Lotem Nadir, October 2021

**Goal**

The main purpose of this project is to develop algorithms and a pipeline for age estimation from RGB-D images, using the posture of the subject. The project consists of two tasks: (1) posture estimation from RGB-D images and (2) age prediction from posture.

**Dataset**

A dataset of 15 recording sessions was collected. In each recording session, a person was recorded using 3 RealSense cameras (RGB-D) and Vicon sensors (3D coordinates). The RealSense cameras record 3 different angles: front, back and side.

A picture containing floor, indoor

Description automatically generatedA picture containing floor, indoor

Description automatically generatedA picture containing floor, ground, sport

Description automatically generatedA picture containing scatter chart

Description automatically generated

Working with multiple sensors raises the following challenges:

* The frames from the different RealSense cameras and the Vicon system are not aligned in time, and synchronizing is required.
* The RealSense cameras and the Vicon sensors are not calibrated. Calibration is required in order to find the transformation between the Vicon coordinate system and the RealSense camera coordinate system.
* The FPS of the RealSense cameras is 30, and the FPS of the Vicon system is 120. In some recordings, the FPS of the RealSense cameras is 15.
* Since two of the RealSense cameras in each recording session are connected to the same laptop, there is a frame drop in the output of these cameras.

**Methods**

In this section I will be referring to each of the tasks separately.

1. Posture estimation from RGB-D images – My main contribution was by cleaning the dataset and finding a method to calibrate RealSense cameras and the Vicon sensors. For the dataset cleaning, automatic method was first examined, using the OpenPose model in order to detect the T-pose at the beginning of each recording. OpenPose performed poorly of the “side” shooting angles. Due to the frame-drop in the RealSense cameras, manual fixes on the “front” and “back” angles were required frequently. Since this method was not effective nor accurate enough, the T-pose was manually detected in all recordings. In order to deal with the different FPS of the sensors, every 4th frame was taken from the Vicon recordings. Another method of averaging every 4 frames in the Vicon recordings was considered. In order to check which method is better, an angle in the neck was calculated in both methods. The difference between the two methods was negligible. In order to deal with the frame-drop in the RealSense cameras, the differences in the frames numbers were extracted, and the correlated Vicon frames were trimmed.

Chart

Description automatically generated

Recordings were cleaned and trimmed for all 15 sessions. Validation of the synchronizing was done manually.  
For the calibration process, Kabsch’s algorithm was used in order to find the transformation between the Vicon coordinate system and the RealSense camera coordinate system: given 2 sets of N paired points in D dimensions, Kabsch’s algorithm calculates the rotation matrix that minimizes the RMSD between the two sets, using singular value decomposition (SVD). The calibration is done using a single frame from the RealSense camera and the corresponding frame from the Vicon system. Kabsch’s algorithm alone performed poorly. Several improvements were made to the data before applying Kabsch’s algorithm on it: Removing points with noisy depth value, averaging the depth value of each point with nei ghboring pixels, sampling sub-group of the points with lowest projection error. After applying the improvements, the current projection error rate is 60mm (RMSE).

A picture containing indoor

Description automatically generatedA picture containing indoor

Description automatically generated

For future recordings, a static calibration device was built.

1. Age prediction from posture – Two methods were examined for this purpose:A picture containing indoor

   Description automatically generated
   1. Predicting age from the Vicon points – PointNet was trained as a binary classifier (“old” or “young”). The network resulted in high overfitting on the trainset. This might be caused by the fact that the data has low variance due to the Vicon high FPS. In order to increase the variance in the data, the dataset was re-generated, this time only frames that have a difference of at least 80mm were kept to the dataset.

A picture containing text, clock

Description automatically generated

The overfitting problem was not solved by this process. More data is required in order to reduce the overfitting.

* 1. Predicting age from angles – Four angles represent the human posture were chosen, each one was defined by 3 3D points from the Vicon data. For each frame in the dataset, the four angles were extracted, and the age was converted into a binary label (“old” or “young”) for that sample. Dimensionality reduction algorithms were applied on the data in order to visualize it.

Chart, scatter chart

Description automatically generated

Several classification algorithms such as KNN, SVM, random forest, were trained on the data. Results on the test set were poor, indicating more data is required for this problem.

My code is maintained [here](https://github.com/Lotemn102/skeleton-RGBDto3D). There are documentation comments throughout the code. I’ve added in the README file instructions for reproducing my steps.

**Future Work**

1. For the 3D skeleton extraction from RGB-D images, there is a need to develop an automatic pipeline for synchronizing the videos in time, and calibrating them. For the time synchronization, we can start the recordings by constant times (10sec, 15sec...) or develop an automatic first frame recognizer, based on some pose the subject does (which can be detected in all shooting angles). For the calibration, the board Omer created can be used.
2. For the age prediction from the 3D skeleton, more data is required. Maybe free datasets of 3D human body can be used:
   1. [3DPeople](https://cv.iri.upc-csic.es/) (80 different subjects, mostly young, RGB-D, 3D skeleton)
   2. [ScanDB](http://gvvperfcapeva.mpi-inf.mpg.de/public/ScanDB/) (114 different subjects, meshes)
   3. [People-Snapshot](https://graphics.tu-bs.de/people-snapshot) (24 different young subjects, meshes)
   4. [Human3.6M](http://vision.imar.ro/human3.6m/description.php) (11 different young subjects, meshes)
   5. [Buff](http://buff.is.tue.mpg.de/) (6 different young subjects, meshes)
   6. [USCS](https://graphics.soe.ucsc.edu/data/BodyModels/index.html) (3000 different people, old & young, meshes)
   7. [MPII](http://humanshape.mpi-inf.mpg.de/) (4300 different people, old & young, meshes) – MIGHT BE OVERLAPPING WITH USCS