DEVELOPMENT OF A MARKERLESS MOTION CAPTURE SYSTEM BASED ON OPENPOSE

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Engineering from the Institut Teknologi Bandung

By
Alvin
NIM: 23121017
Faculty of Mechanical and Aerospace Engineering

INSTITUT TEKNOLOGI BANDUNG
December 2022

Chapter I Introduction

This research was proposed as continuation of the final project "3D Running Ground Reaction Forces Prediction using LSTMs based on kinematics parameters". There are several considerations made during decisions making and will be explained in this chapter.

I.1 Research Background

In the final project, LSTM (Long Short-Term Memory) based models had been created and successfully predicted the 3D Ground Reaction Forces of running with decent accuracy [1]. These models take kinematics data, position, velocity, and acceleration, as their input, which can be obtained by using motion capture on a subject running on a treadmill. The accuracy was evaluated using the test dataset which is presented in Table 1.1. The Ground Reaction Forces (GRF) prediction was motivated by how overpriced an instrumented treadmill are. But to be able to pass inputs to the models, a robust method on capturing running motion was needed.

Table I.1 Model's performance [1]

Speed	GRFX (Anteroposterior)		GRFY (Vertical)		GRFZ (Mediolateral)	
	Mean RMSE (± S.D.)	Mean Coeff. Correlation	Mean RMSE (± S.D.)	Mean Coeff. Correlation	Mean RMSE (± S.D.)	Mean Coeff. Correlation
2,5 m/s	1,87E-02 ± 7,71E-03	0,990	4,66E-02 ± 4,12E-02	0,998	7,00E-03 ± 2,50E-03	0,987
3,5 m/s	1,92E-02 ± 6,33E-03	0,994	4,38E-02 ± 2,35E-02	0,999	7,50E-03 ± 4,60E-03	0,983
4,5 m/s	1,93E-02 ± 4,90E-03	0,996	4,30E-02 ± 1,34E-02	0,999	8,00E-03 ± 2,40E-03	0,987
Average	1,91E-02 ± 6,31E-03	0,993	4,45E-02 ± 2,60E-02	0,999	7,50E-03 ± 3,17E-03	0,985

There have been several motion capture methods that were developed in ITB Biomechanics Laboratory. The first one is the optical marker-based motion capture which use LED (Light Emitting Diode) lights as markers [2]. This method was used most often because of its simplicity in the image segmentation method. Like many other active marker-based motion capture methods, this method should be conducted in a dark room to contrast markers against the background so that it will be easy to segment the markers from the background. But this method has several shortcomings like the markers tend to shake while the subject is moving, it takes time to attach all the markers, and lastly the darkroom itself.

The other method is the markerless motion capture that was developed in 2020 [3]. This method allows us to do motion capture in a bright room without any marker which overcomes most of the problems stated before. However, the method created at that time was not able to analyze 3D motion as it depends on the sagittal view of the subject. Moreover, it is computationally costly because it used Particle Swamp Optimization to adjust the position of the human model to the silhouette at every single frame.

To overcome the disadvantages of the developed method by Lab Biomekanika, the author looks for another alternative that can analyze human pose on every motion and every is angle view free. In the year around 2010 there was a popular method called human pose estimation. This method relies on Artificial Intelligence especially Deep Learning as its base model for predicting the joint position of human pose. It started from tree structured model [4] to CNNs model based [5] for reliable local observations on the image. Moreover, the method also evolved from single person pose estimation to multiple persons pose estimation by changing from top-down strategy to bottom-up approach. The bottom-up approach was proved to have much less computational cost as it did not need to predict every person at first which solve the multiple person complexity in a single frame [6]. As an addition, PAFs (Part Affinity Field) was also introduced so the model can differentiate between a person and another accurately. This was proposed by research called the OpenPose model [7]. This model was using the PAFs as vector to encode unstructured pairwise relationships between body parts and heatmaps as the joint position probability itself. It could even handle multi-person prediction in real-time. Although it is still 2D, this can be used to substitute the early stage of marker-based motion capture up to the segmentation method. But because it was trained by using a random and general human pose dataset, it was not accurate enough for motion capture purposes compared to the conventional marker-based motion capture or even inertial sensors [8, 9].

To overcome OpenPose shortcomings, in this research, the original OpenPose will be fine-tuned using motion capture data. By fine-tuning, the parameters inside the original OpenPose model will be adjusted to fit the motion capture task and thus have better accuracy. In addition to that, the motion capture data should consist of both images of the subject in a bright room that will clearly show the subject and the marker positions. This kind of data can be obtained by using color-based motion capture data that will also be developed in this research.

I.2 Research Background

The problems identified for this research consist of:

1. There has not been any developed motion capture method in ITB Biomechanical Laboratory that can capture subject in bright room using marker.
2. OpenPose's accuracy is not satisfactory for motion capture task compared to marker-based or inertial sensors-based motion capture.

I.3 Research Purposes

This research's purposes can be divided into two main purposes, which are:

1. To create a color-based motion capture method for collecting the data required for fine-tuning.
2. To fine-tune OpenPose using motion capture data so that it will have better accuracy in motion capture tasks.

I.4 Problem Limitation

In this research, several limitations are applied to simplify the problem. These limitations consist of:

1. The motions that will be analyzed are only walking and running.

2. There will be only one subject in the frame.

I.5 Methodology

This research pipeline can be divided into two main parts. The first one is to create color-based motion capture and the other is to fine-tune the OpenPose using the data collected from color-based motion capture. These processes can be represented using a flowchart as shown in Figure I.1.

Color-based motion capture will be created using Python and OpenCV library in Anaconda3. The process starts with reading the video and dividing it into frames. The OpenCV will read the image in BGR (Blue Gray Red) format by default, so next, it needs to be converted into HSV (Hue Saturation Value) format. After converting the frames, thresholding can be done by selecting the parameters, H, S, and V value to segment the specified color resulting blobs. Lastly, the centroid of each blob will be detected and exported in a CSV (Comma Separated Value) file. The marker used was created by 3D printing and then it will be painted green.

To fine-tune the OpenPose, some data should be collected using the previously created color-based motion capture. But these data should be the images showing only the subject without the marker and the position of the marker. To obtain the images without any marker, preprocessing should be conducted to the original images captured by the camera. This preprocessing step was called inpainting, a process to infer back the part of a picture using deep learning. After the markers were hidden out, the images and corresponding marker positions then will be paired to be a dataset. Next, this dataset was split into the train set and test set. The train set was used to fine-tune all the parameters inside the original OpenPose, and the test set was used to evaluate the fine-tuned OpenPose. The model then was tested on a video without any marker for checking whether the model really worked as expected or not. Moreover, deployment pipeline was also built to deliver this model so that it can be used in the future. In Machine Learning Operations (MLOps), creating a model was usually called research phase and deploying the model for use is usually called the deployment phase. The deployment phase usually included the

Application Programming Interface (API), Docker and many other services. Although in this research, the author has not planned until the monitoring phase but the deployment should be sufficient for further research. Therefore, a project repository was also built along the way by using GitHub.

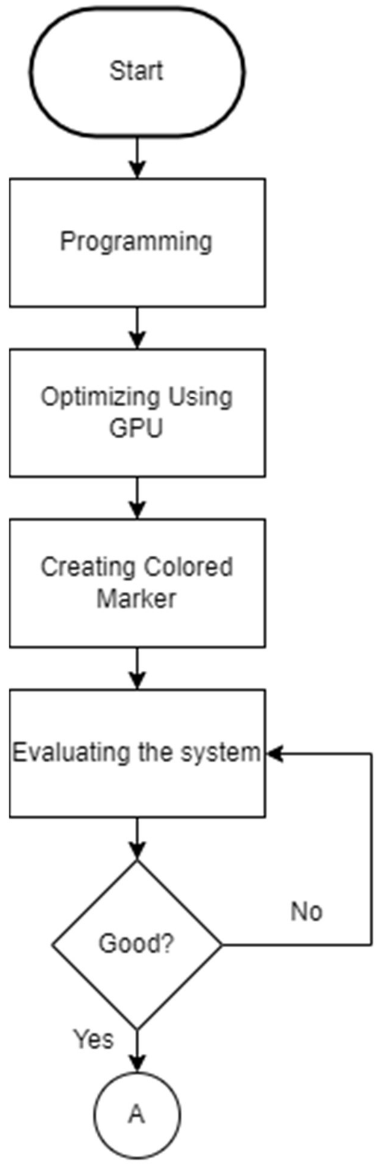

Figure I.1 Research flowchart

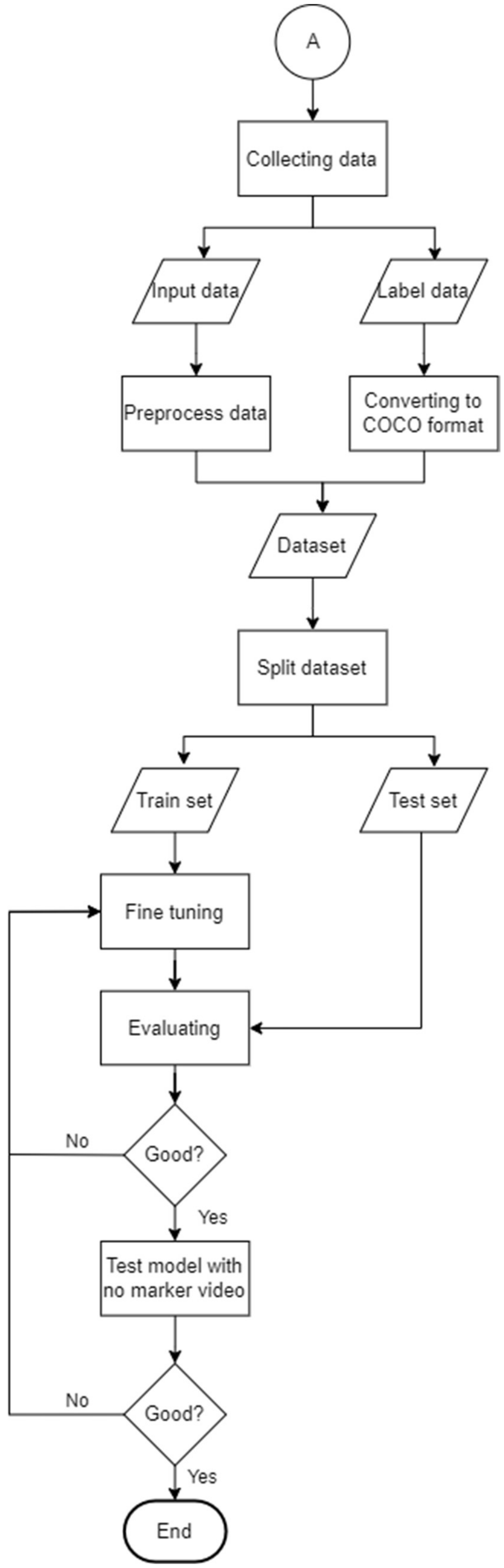

Figure I.2 Research flowchart (continued)