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

ABSTRAK

PENGEMBANGAN SISTEM PENANGKAP GERAK TANPA PENANDA BERBASIS OPENPOSE

Oleh

Nama Alvin

NIM: 23121017

Fakultas Teknik Mesin dan Dirgantara

Dalam penelitian sebelumnya, telah dibuat beberapa model untuk memprediksi gaya reaksi tanah secara 3D dengan metode LSTM (*Long Short-Term Memory*). Model tersebut bekerja dengan menerima data kinematika gerak yang didapatkan dari penangkap gerak sebagai inputnya. Akan tetapi, terdapat beberapa kelemahan pada sistem penangkap gerak yang telah dikembangkan oleh Lab Biomekanika ITB yang diantaranya ialah gerak harus dilakukan dalam ruangan gelap dan penanda cenderung bergoyang saat subjek bergerak. Untuk mengatasi kelemahan diatas, metode estimasi pose manusia yang berdasar pada *Deep Learning* dan *Computer Vision* dapat diterapkan sebagai sistem penangkap gerak tanpa penanda. Pada tahun 2020 terdapat kemajuan yang pesat saat model OpenPose dirilis karena dapat memprediksi pose manusia secara instan meski terdapat banyak subjek dalam gambar. Namun, telah dibuktikan bahwa OpenPose tidak memiliki akurasi yang cukup untuk menggantikan sistem penangkap gerak tanpa penanda. Oleh sebab itu, pada penelitian ini dilakukan fine-tuning terhadap OpenPose agar dapat memprediksi dengan lebih akurat untuk tugas penangkapan gerak berjalan atau berlari manusia. Namun untuk dapat melakukan *fine-tuning* tersebut, diperlukan data gambar yang dapat menunjukkan subjek dan penanda dengan jelas. Lalu dilakukan juga *inpainting* untuk menghilangkan penanda dari gambar subjek sehingga didapatkan gambar subjek tanpa penanda dan posisi penanda tersebut sebelum dihilangkan. Setelah dilakukan pelatihan ulang, ditemukan bahwa akurasi model dinilai cukup dengan kesalahan rata-rata 17-*pixel*.

Kata kunci: Penanda, *fine-tuning*, OpenPose, penangkap gerak

ABSTRACT

DEVELOPMENT OF A MARKERLESS MOTION CAPTURE SYSTEM BASED ON OPENPOSE

By
Alvin
NIM: 23121017
Faculty of Mechanical and Aerospace Engineering

In previous research, several models to predict the 3D Ground Reaction Forces of running had been created using LSTM (Long Short-Term Memory). The models take kinematic data obtained through motion capture as input. However, there were disadvantages to the currently developed method in ITB Biomechanics Lab such as the need for a dark room, or the tendency of markers to shake while the subject is moving. To overcome this, human pose estimation based on deep learning and computer vision can be applied as a markerless motion capture system. There was a huge leap in this field back in 2020, by a bottom-up approach model OpenPose. This model can predict human pose in real-time even for multi person problem. But recently, it was reported that OpenPose is not as accurate as marker-based or sensor-based motion capture. Hence, to alleviate this inaccuracy, it is proposed to fine-tune the original OpenPose with motion capture data so that it will learn and perform better for motion capture task especially for running or gait. But to get the fine-tuning done, the subject and marker should be clearly visible. Later, the frames will be inpainted to hide the marker shown in the image leaving only the subject for the input data. Lastly, after training it was discovered that the model has good accuracy with the average of 17-pixel error.

Keywords*:* marker, fine-tuning, OpenPose, motion capture

APPROVAL SHEET

DEVELOPMENT OF A MARKERLESS MOTION CAPTURE SYSTEM BASED ON OPENPOSE

By

Alvin

NIM: 23121017

Faculty of Mechanical and Aerospace Engineering

Institut Teknologi Bandung

Approval on

Date: 12th December 2022

by

Supervisor

Prof.Ir. Andi Isra Mahyuddin Ph.D.

Co-Supervisor 1	Co-Supervisor 2
Ferryanto, ST., MT.	Pramudita Satria Palar, S.T, M.T, Ph.D.

ACKNOWLEDGEMENT

First, I would like to express my gratitude and sincere thanks to my respected supervisors Prof.Ir. Andi Isra Mahyuddin Ph.D., Ferryanto, ST., MT., and Pramudita Satria Palar, S.T, M.T, Ph.D. for the guidance, insight, and positive criticism they gave throughout the process of this work. Next, I would also like to thank my friends Nardo Rizaldy and Nico Jackson that helped me during the data acquisition process because without them it would be impossible to finish in a month. Lastly, I would also like to express my gratitude for all the subjects that has been willing to come to ITB during Covid-19 pandemic and without them there would be no data for this work to be done.

CONTENT

LIST OF FIGURES

LIST OF TABLES

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