ISSN No:-2456-2165

[AI-Machine Learning] Optimized Sensorless Human Pose Estimation for a Kpop Dance Application

G. Jeong[1] ; N. Freitas[2] ; Y. Cho[3] ; C. Han[4]
[1]Gyumin Jeong, Hansung University, Devunlimit
[2]Núria Freitas, Catholic University of Korea, Devunlimit
[3]Yoonhwan Cho, Devunlimit
[4]Alan, University of Auckland, Devunlimit
Seoul Business Agency, Seoul Innovation Challenge 2019

Abstract:- There has been a great effort to use technology to make exercise more interactive, measurable and gamified. However, in order to optimize the detection accuracy, these efforts have always translated themselves into motion detection with multiple sensors including purpose specific hardware, which results in extra expenses on both the content production and consumption and induces limitations on the final mobility of the user. In this paper we aim to improve the accuracy, learning speed and detail range of Posenet's AI sensorless human pose detection by using an artificial neural network to optimize its extraction and comparison algorithms, changing the current model that uses a ResNet convolutional neural network (CNN) to a model using DenseNet and developing a new algorithm for detailed corrections using relevant artificial neural networks. The findings here will be applied on a posture correction system for a dance and fitness application.

Keywords:- *Sensorless human pose estimation – Artificial Intelligence (A.I.) – Machine learning – Posenet – DenseNet – Posture correction methods – Human motricity – Motion capture methods – Dance – K-pop – E-sports – South Korea.*

I. INTRODUCTION

A meaningful motion capture at both the teaching and learning ends of an interactive motion education platform is critical for an efficient service. Some of the more demanding motion caption industries have been using commercial options such as marked suits, movement sensors and specialized camera settings such as temperature, infrared or relative distance.

However, differently from the movie and character animation industry, which only requires motion capture once, or the medical industry, that works usually in controlled environments and is more flexible to requirements for a high accuracy, the consumer interactive market requires a very low-cost motion capture at least on one of the service flow ends, which is not controlled by the service provider.

Recently, research has shown to be possible to recognize motion through single cameras, without using extra hardware, using artificial intelligence's machine learning methods.

PoseNet is a machine learning model that allows for Real-time Human Pose Estimation it can be used to estimate either a single pose or multiple poses.

In this paper, we aim to raise the accuracy for movement detection, comparison and feedback for a single RGB machine learning pose estimation. Reminding that these procedures can be made leaving space for further uses on vital data estimations too.

II. OBSERVATION METHODS AND INDICATORS

The observations on this paper have been done using a real-time video taken with a single regular RGB camera.

For further calculations, we measured the number of frames that have skeleton data outputs in a second to set the actual service frame per second (FPS) rate.

In order to avoid large problems when extracting skeleton data through an artificial neural network due to exposure issues and to verify that it works properly in various settings, we also set a standard range of luminosity (Lux).

Lastly, to verify whether skeleton data, and later on vital data, is extracted accurately in various motions, we measured the mean per joint position error (MPJLE) of the neural network machine.

< Overview of Major Performance Indicators >					
Performance Indicator	Unit	Final objectives	World's highest level (Holding company / holding country)	Percentage (%)	Measuring organization
1. Data output speed	fps	More than 24fps	15~30fps (Microsoft, US)	30	TTA accredited certification test
2. Illuminance	lux	More than 5,000lux	Less than 5,000lux (Microsoft, US)	30	TTA accredited certification test
4. Observation position error	mm (mdjpe)	Less than 90mm	More than 110 (Saarland Univ, Germany)	20	TTA accredited certification test

<Sample Definition and Measurement Method>			
Performance Indicator	Sample Definition	Number of samples (n ≥ 5)	Measurement method (standard, environment, result calculation, etc.)
1. Data output speed	1 per System	5	Measured frames per second (fps) in the result data value image taken through the deep learning machine when shooting in real time
2. Illuminance	1 per System	5	Measures illuminance (lux) in the driving environment, and executes in the environment above the standard illuminance value.
4. Observation position error	1 per System	5	Measure the accuracy of the skeleton value obtained by the deep learning machine when shooting in real time

Table 1:- Performance indicators, objectives and observation methods.

III. PROCESS AND SCHEDULES

A. Pre-test procedures:

➢ *Research and development of motion capture technology*

- In order to provide feedback to the user, we needed to extract 3D skeleton data instead of the existing 2D technology.
- Develop an artificial neural network that extracts three-dimensional skeleton data using a single RGB camera.
- In order to utilize the technology as a global service, developments should be made in Python so that can be used without licence costs in Linux OS and it can be deployed later as an Infrastructure as a Service (IaaS) as Platform as a Service (PaaS).

➢ *Contactless Service Control Development*

- Due to the characteristics of the service, prior studies are deemed necessary for improved usability.
- Developed a remote contactless control that can be used in Web, Mobile / PC applications using motion capture technology and applied for international PCT patent.

➢ *Web application prototype development*

- Develop of a web application prototype because it is deemed necessary to verify service utilization and development potential.
- Java Script has the disadvantage that the execution result may be different in each browser, but to solve this problem, we developed in compliance with the web standard of W3C, the international web standards organization.

B. Test procedures:

➢ *Build training data set and test data set*

- It takes three months to build the appropriate training and test datasets by collecting and processing three-dimensional skeleton datasets collected using a single RGB camera.

➢ *Development of Improved Artificial Neural Network Algorithm for Posture Correction*

- It takes 4 months to design and develop an artificial neural network algorithm with improved accuracy and speed compared to the existing neural network algorithm for posture correction.

IV. EXPERIMENT

Development of technology to extract skeleton data using artificial neural network

A. Select Data Set

Dataset Name	Size	Obs
Densepose	2.1 GB	Train(0.6GB) eval(1.3GB)
Mpi-inf-3dhp	11GB	2 set 4 sequence total
MPII-HUMAN-POSE	12GB	

Table 2

B. Data Set Pre-processing

Requires video for testing skeleton data, converts resolution to 480x360 using FFMPEG, and saves it as a png file.

C. Training

- Improved accuracy by changing the ResNet network used to extract skeleton data into DenseNet network to improve the learning effect and speed of machine learning.
- Execution Environment (Hardware / software)
  CPU Intel i7-7700k
  Memory DDR4 32GB
  Graphic Card Nvidia GTX1660Ti 6GB
  SSD 256 GB

- Execution Count / Time

Based on the GPU memory capacity, it is set to batch size 2.

400 times of study, total time required one week.

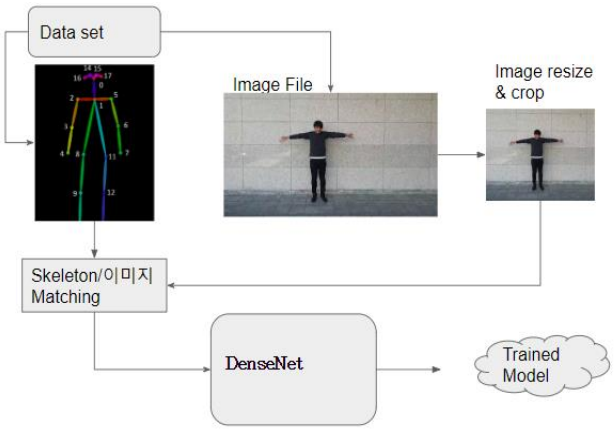

Fig 1:- Cloud software flow chart extracting skeleton data using artificial neural network

D. Verification

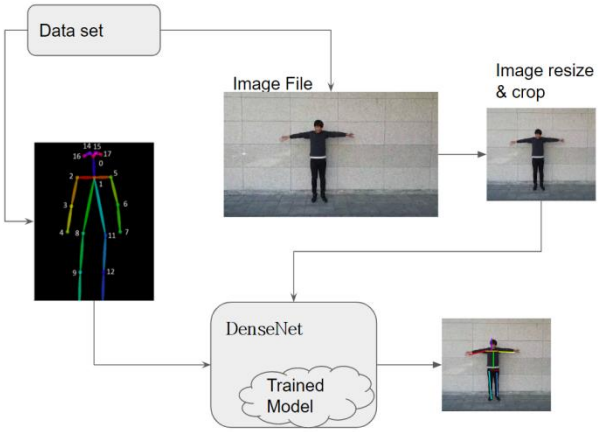

Fig 2:- Cloud software verification extracting skeleton data using artificial neural network

V. PROCEDURES

The test procedure for extracting skeleton data is as follows. Install Linux and install the utilities associated with the source you developed, just as you measured your heart rate. And similarly, we will install a GPU (including cuda) on Linux. This is also installed in the same way as the heart rate was measured. When the installation is complete, download and install the anaconda distribution package from https://docs.anaconda.com/anaconda/install/. Unzip the downloaded file to configure the environment and install the required library.

VI. RESULTS AND DISCUSSION

The experiment has tested the joint position error for neural network using DenseNet from real-time video data taken with a single regular RGB camera.

The test was conducted five times and it was confirmed the tests met the final development goals presented in the previous section 2, ‘Observation methods and indicators’.

The joint position errors were all 10.224 mm, meeting the target value of 90 mm or less.

The images used for the joint position error test were above the target of 5000 lux when the illumination intensity was measured at the time of image shooting, as shown in Figure 3, and were used to verify that the product was operating normally.

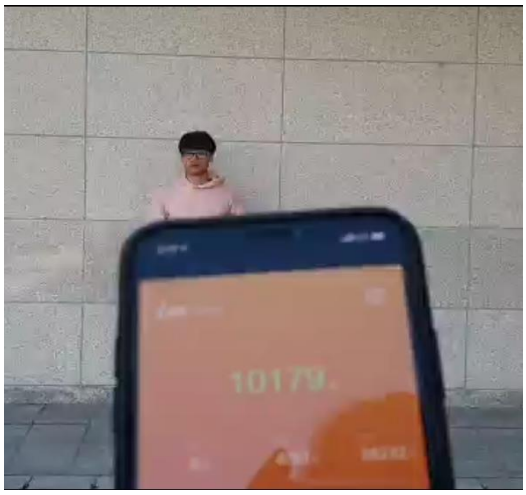

Fig 3:- Illuminance value when taking test images

Index	Analyzed joint position images	Mean per joint position error
1	1000	10.224 mm
2	1000	10.224 mm
3	1000	10.224 mm
4	1000	10.224 mm
5	1000	10.224 mm
Average		10.224 mm

Table 3:- Measurement of MPJLE of images taken in 9000LUX or higher environment.

REFERENCES

[1]. https://github.com/tensorflow/tfjs-models/tree/master/posenet
[2]. http://gvv.mpi-inf.mpg.de/projects/VNect/
[3]. https://medium.com/tensorflow/real-time-human-pose-estimation-in-the-browser-with-tensorflow-js-7dd0bc881cd5