DOI: 10.22146/jnteti.v13i1.8145

Identification of Fencing Athletes Based on Anthropometric Measurements Using MediaPipe Pose

Bagas Alif Fimaskoro[1], Suci Aulia[1], Dery Rimasa[2]

[1] School of Applied Science, Telkom University, Bandung, Indonesia
[2] Sports Physical Training Study Program, Universitas Pendidikan Indonesia, Bandung, Indonesia

[Received: 13 June 2023, Revised: 22 August 2023, Accepted: 8 December 2023]
Corresponding Author: Suci Aulia (email: suciaulia@telkomuniversity.ac.id)

ABSTRACT — Over time, numerous developments in digital technology have benefited people, including anthropometric measurements that provide information on an athlete's ability in sports. The use of digital technology in sports must continue, particularly in the National Sports Committee of Indonesia (Komite Olahraga Nasional Indonesia, KONI) of Bandung City. This study proposed a technique for classifying and identifying fencing athletes' talents. This work developed a methodology for evaluating sports talent based on anthropometric measurements of athletes' bodies using the posture detection approach. Fencing and nonfencing athletes in KONI Bandung City were categorized using this talent identification. This study used 36 datasets of body posture images from various skills of the sport. These images were in JPEG or JPG format with a resolution of 3,024 × 4,032 and were acquired using a Canon EOS 1300D camera. This study utilized four points landmarks, which are usually used as measurement components in KONI, to categorize fencing athletes. The four points are shoulder (S), elbow (E), index (I), and hip (H) landmarks. The testing was done using three different dataset settings. According to the test results of all scenarios, scenario 2 had the highest accuracy. This scenario was able to categorize fencing and nonfencing athletes with an accuracy rate of 89% and an average processing time of less than 3 s per image.

KEYWORDS — Anthropometry, Image Processing, Pose Detection, Fencing, Talent Identification.

I. INTRODUCTION

Sport is a movement carried out precisely according to the type of movement with various goals and directions. It is essential for everyone's life as it provides many benefits such as affecting the development of physical growth [1]. The manual system to measure the athletes' body anthropometry is still utilized today, especially in the National Sports Committee of Indonesia (Komite Olahraga Nasional Indonesia, KONI) of Bandung City. Therefore, innovation within KONI, particularly Bandung City KONI, should be made so that KONI is able to fulfill its responsibility on the athlete development. Taking advantage of advancements in technology throughout the digital age, several tools utilizing image processing have been developed to facilitate body anthropometric measurements in athletes and sport talent assistance. Anthropometry is a science related to measuring the dimensions of the human body [2]. Anthropometric measurements are useful for designing equipment and facilities for daily activities [3]. Previous studies have conducted anthropometric measurements of the body. Aligning to [4], measurements in this study were only applied to multiple specific areas. This study measured height and weight using the morphological image processing method. Another study has utilized MATLAB software to create an anthropometric measurement for all athletes [2].

As a continuation, this research proposed a design of an automatic anthropometric measurement system based on image processing to be implemented at the Bandung City KONI, mainly for fencing athletes. This identification can be utilized to differentiate body measurements between fencers and nonfencing athletes with focus on the upper body, namely arm length, shoulder width, and waist width. Human pose estimation technologies are currently being extensively researched in several fields such as sports, home training, surveillance, job monitoring, cultural activities, gesture control, home elder care, and virtual world avatars. Human pose estimation is usually separated into 2D and 3D model estimation methods, single-object and multiple-object methods based on the number of target subjects, monocular and multiview image methods based on the quantity of shooting cameras, and single image and video methods based on the input type [5]–[9]. Notably, human pose estimation is divided into single-stage and two-stage methods based on the structure of the deep learning process. Two groups of single-stage algorithms immediately convert input images into 3D body joint coordinates: regression-based approaches [10], [11] and detection-based methods [12], [13]. The regression-based methods directly estimate the location of connections to the location of the root joint [10], [11] or the joint angles utilizing a kinematic model with numerous joints and bones [12], [13]. On the contrary, the detection-based methods predict a likelihood heatmap for each joint, whose location is determined by using the heatmap's maximum likelihood. Lifting 2D to 3D, also known as leveraging 2D pose estimation discoveries for 3D human pose estimation, is an active area of research. It is because 2D pose estimation has a larger number of datasets in real-world scenarios that provide accurate joint coordinates compared to 3D pose estimate [14]. Long short-term memory (LSTM) has been used to derive advantage of the interactions between joints [15]. Meanwhile, generative adversarial networks (GANs) are frequently employed to create a more realistic 3D human pose [16]. The deep learning methods of 3D pose estimation from 2D images should overcome challenging issues such as a shortage of in-the-wild datasets, high demand for different posture data, depth ambiguities, and vast search state space for each joint despite ongoing technological advancements. A powerful personal computer (PC) with numerous graphic processing units (GPUs) is also required to run deep learning software. Due to the significant coupling between processes, changing an application of perception to include additional processing steps or inference models can be challenging. It also takes time to create the same application for several platforms because these processes typically need to be optimized for a particular device to function correctly and effectively. By interpreting and integrating several perception models into consistent pipelines, MediaPipe solves these problems. The pipeline architecture contains the instructions required to draw conclusions from the sensory data and produce the perceived outcomes. Since MediaPipe components have a similar interface focused on time-series data, reusing them in other pipelines throughout subsequent projects is simple [17]. Subsequently, every pipeline can function uniformly across multiple platforms, allowing the practitioner to create the application on computers and later deploy it on mobile devices.

This study employed a two-stage pose estimation method using MediaPipe Pose (MPP) to execute a human pose estimate package on a single-board computer. This study used the MPP, Google-provided open source and cross-platform framework, to estimate the 2D human joint coordinates in each image database. The MPP uses machine learning to construct pipelines and interpret cognitive data presented as images [17]–[19]. This fencing athlete's identification is expected to help the Bandung City KONI to take anthropometric body measurements quickly and precisely. The rest of this paper is organized as follows. Sections 2 and 3 present the proposed framework for pose estimation and talent identification. Section 4 presents the experiment's results. Finally, Section 5 presents our conclusions.

II. ANTHROPOMETRIC MEASUREMENT METHODS

A. ANTHROPOMETRY

Anthropometric measurements are used to determine a suitable body for several sports [20]. Anthropometry is the measurement of the human body, including the determination of length, width, diameter, and circumference; it also involves the calculation of ratios and proportions based on two or more measurements to identify body shape, size, and topography [21].

Anthropometry is a scientific field that focuses on the measurement and analysis of human body dimensions [22]. Anthropometry is widely applied in sports. One of the essential aspects in identifying sports talent and achieving good sports performance is strengthening the ideal body size at peak performance. Anthropometry utilizes information obtained from body measurements to distinguish people, groups, and other entities. An anthropomorphic system of proportions based on human body dimensions can be used to differentiate between adults and children. One method is static anthropometry, which involves measuring the body in conventional positions without doing any movements. Meanwhile, the other method is dynamic anthropometry, which is carried out when a person performs specific movements related to certain tasks.

B. RED, GREEN, BLUE (RGB) COLOR

Each color in a color image is represented by 4, 8, 16, or 24 bits of information for each pixel, with the number of colors ranging from 16; 256; 65,536; or 16 million. Red, green, and blue (RGB) make up the main hues of the actual color. RGB color refers to the combination of the three primary color-forming elements, which will eventually create a color image [23]. Red is a matrix that states the degree of brightness for the red color (for example, for a grayscale of 0–255, a value of 0 represents black and 255 represents red). Green is a matrix that states the degree of brightness for green. Blue is a matrix that states the degree of brightness for blue [24]. A 3 × 3 matrix transformation can convert the RGB data found in a pixel into the CIE XYZ color space. Tristimulus values, a configuration of three light-linear components that perform the CIE color-matching function, were used in this process. Specific colors are represented as always positive values in the XYZ color space [25].

C. GRAYSCALE

Greyscale refers to an image that possesses a grayscale level [26]. The grayscale image consists of the RGB colors, all of which possess an identical intensity [27]. In contrast to color images, where each pixel requires three intensities, grayscale images only require one intensity value. The degree of gray is represented by the 8-bit integer intensity of the grayscale image, which ranges from 0 to 255 with 0 indicating black and 255 indicating white; the value between them is the degree of gray [28].

D. ARM LENGTH CALCULATION

The measurement of the arm's length can be determined by calculating the number of pixels measured from the fingertips of the right hand to those of the left hand and multiplying them by the ratio of the hand's length, which is commonly denoted as $hand_l$. This study calculated the pixel length value of black objects represented by a plus sign (+) on a background and compared the actual length size of the object in centimeters to determine the $hand_l$ ratio. Meanwhile, $Object_{pl}$ denotes the number of object pixel lengths. The following is the equation for calculating arm length [29]:

$$Arm\ length = Object_{pl} \times hand_l. \qquad (1)$$

E. WAIST CIRCUMFERENCE CALCULATION

The waist circumference can be determined by counting the number of pixels of the object's waist in the front view and the number of pixels of the object's waist in the side view. Determining the diameter 1 (d_1) and diameter 2 (d_2) waist on the object will yield the pixel values. Subsequently, the d_1 and d_2 values in pixels are converted into centimeters. Equation (2) shows the formula for calculating waist circumference [30]:

$$Waist\ circumference = \frac{1}{2} \times \pi \times (d_1 + d_2). \qquad (2)$$

F. HUMAN POSE ESTIMATION

1) THE 2D POSE ESTIMATION

Human pose estimation has been thoroughly researched over the past few years, from pictorial structures [31], [32] to more current convolutional neural network (CNN) techniques [33]. Based on prior study, there are two groups of pose estimation algorithms: regression-based and detection-based methods. In detection-based approaches for posture estimation, each pixel in a heatmap represents a joint's detection score, which consider pose estimation as a heatmap prediction problem [34]. Prior study investigated the idea of stacked architectures, residual connections, and multiscale processing [35]. They subsequently developed the stacked hourglass network, which significantly improved the performance of 2D posture estimation issues. Since then, state-of-the-art

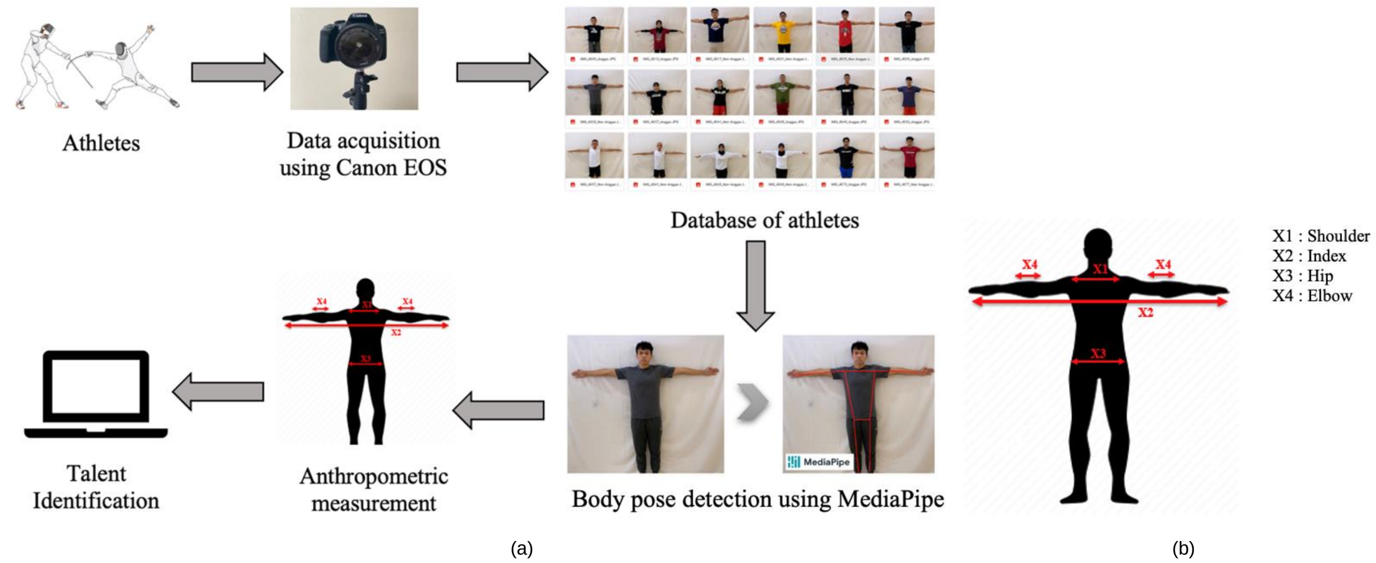

Figure 2. Images of (a) system identification block diagrams, (b) landmark in this research.

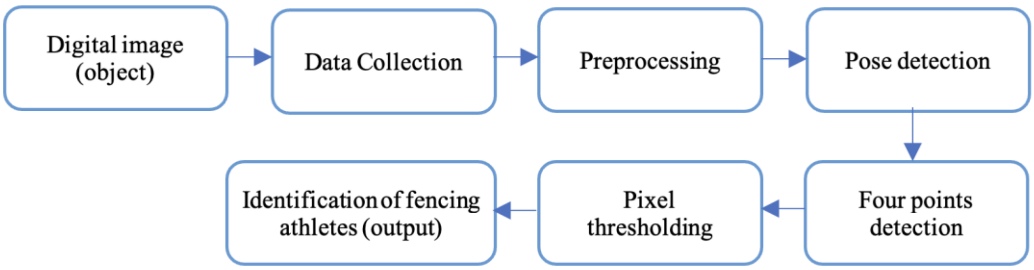

Figure 3. Block diagram of image processing.

nonfencing athletes. The participants, both fencing and nonfencing athletes, were students of the Faculty of Sport and Health Education, Universitas Pendidikan Indonesia, and consisted of men and women aged 20–22 years. Like most sports, fencing has a diverse range of participants. Physical traits, such as age, weight, height, power, speed, and general athleticism, are varied. In the dataset, fencing participants have a minimum height of 175 cm for men and 160 cm for women.

2) PREPROCESSING

Preprocessing was one of the first methods employed in image processing. Preprocessing seeks to minimize or increase data size, remove any existing noise, clarify image features, and change the original data into expected data that will later form a required object [16]. This process was carried out in this study for scenarios 2 and 3. In scenario 2, the contrast of the dataset was improved. In scenario 3, on the other hand, the original dataset was turned into black-and-white images. Preprocessing can also be done to improve an image from existing noise. It was done to remove parts of the input image that were not needed for the subsequent process.

3) POSE DETECTION

Pose detection using MediaPipe resulted in 33 points [17]–[19], as seen in Figure 1. Nevertheless, as previously mentioned, this study only utilized four points landmark as variables: shoulder (X_1), index (X_2), hip (X_3), and elbow (X_4) as seen in Figure 2.

4) PIXEL THRESHOLDING

Thresholding aims to determine the minimum and maximum pixel values in an image [20]. Pixel thresholding is one of the main techniques in clustering or classifying athletes into fencing and nonfencing categories. In this study, pixel thresholding segmented the measurement results (pixel) for each landmark (X_1), (X_2), (X_3), and (X_4) based on the actual size (pixel) of fencing athletes that had been simulated in the training process.

B. ACQUISITION DESCRIPTION

In this study, testing was carried out with 36 athletes of all types of sports in the Bandung City KONI, all of whom were of different body size. This testing was carried out with other measuring point parameters: shoulder (S), elbow (E), index (I), and hip (H). These images were taken using a camera lens with sufficient lighting so that there are no disturbing shadows when taking the picture. The camera used was Canon EOS 1300D with a height of ± 1 m from the floor surface, 2 m from the camera to the object, and angle of 90° between the camera and the object.

IV. RESULTS

In this research, testing was performed based on three dataset scenarios, each using 36 datasets taken directly at the Bandung City KONI. In scenario 1, the testing was conducted using original datasets. In scenario 2, the testing employed different datasets from the original ones because the contrast was changed. In scenario 3, the testing used datasets that were converted from their original RGB color into black and white.

The accuracy was calculated by comparing the measurement by the system proposed in this this research with the actual size by the Bandung City KONI team. The measure of a fencing athlete consisted of four points components: elbow length, shoulder width, index length, and hip width as seen in Figure 4.

[36] W. Yang *et al*., “Learning feature pyramids for human pose estimation,” *2017 IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 1290–1299, doi: 10.1109/ICCV.2017.144.

[37] X. Chu *et al*., “Multi-context attention for human pose estimation,” *2017 IEEE Conf. Comput. Vis., Pattern Recognition (CVPR)*, 2017, pp. 5669–5678, doi: 10.1109/CVPR.2017.601.

[38] Y. Chen *et al*., “Adversarial PoseNet: A structure-aware convolutional network for human pose estimation,” *2017 IEEE Int. Conf. Comput. Vis. (ICCV),* 2017, pp. 1221–1230, doi: 10.1109/ICCV.2017.137.

[39] C.J. Chou, J.T. Chien, and H.T. Chen, “Self adversarial training for human pose estimation,” *2018 Asia-Pac. Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, 2018, pp. 17–30, doi: 10.23919/APSIPA.2018.8659538.

[40] A. Toshev and C. Szegedy, “DeepPose: Human pose estimation via deep neural networks,” *2014 IEEE Conf. Comput. Vis., Pattern Recognit. (CVPR)*, 2014, pp. 1653–1660, doi: 10.1109/CVPR.2014.214.

[41] J. Carreira, P. Agrawal, K. Fragkiadaki, and J. Malik, “Human pose estimation with iterative error feedback,” *2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 4733–4742, doi: 10.1109/CVPR.2016.512.

[42] D.C. Luvizon, H. Tabia, and D. Picard, “Human pose regression by combining indirect part detection and contextual information,” *Comput. Graph*., vol. 85, pp. 15–22, Dec. 2019, doi: 10.1016/j.cag.2019.09.002.

[43] D. Mehta *et al.,* “VNect: Real-time 3D human pose estimation with a single RGB camera,” *ACM Trans. Graph*., vol. 36, no. 4, pp. 1–14, Aug. 2017, doi: 10.1145/3072959.3073596.

[44] J. Martinez, R. Hossain, J. Romero, and J.J. Little, “A simple yet effective baseline for 3D human pose estimation,” *2017 IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2659–2668, doi: 10.1109/ICCV.2017.288.

[45] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, “Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments,” *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 36, no. 7, pp. 1325–1339, Jul. 2014, doi: 10.1109/TPAMI.2013.248.

[46] C.H. Chen and D. Ramanan, “3D human pose estimation = 2D pose estimation + matching,” *2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 5759–5767, doi: 10.1109/CVPR.2017.610.

[47] S. Liang, X. Sun, and Y. Wei, “Compositional human pose regression,” *Comput. Vis., Image Underst.*, vol. 176–177, pp. 1–8, Nov./Dec. 2018, doi: 10.1016/j.cviu.2018.10.006.