TẠP CHÍ KHOA HỌC ĐẠI HỌC TÂN TRÀO

ISSN: 2354 - 1431

http://tckh.daihoctantrao.edu.vn/

3D HUMAN POSE ESTIMATION FROM SPORT VIDEOS USING MEDIAPIPE FRAMEWORK

Thai Son Dinh[1], Thi Thanh Minh Luong [2],Huu Son Do[3], Van Nam Phan[3],Trung Hieu Te[3], Van Hung Le[3]*

[1]Hung Vuong University, Phu Tho, Vietnam

[2]ICTU, Thai Nguyen, Vietnam

[3]Tan Trao University, Tuyen Quang, Vietnam

[*]Email address: dosonhytq@gmail.com

DOI: 10.51453/2354-1431/2023/975

Article info

Received: 16/12/2022

Revised: 06/03/2023

Accepted: 16/5/2023

Keywords:

3D human skeleton, Convolutional, Neural, Networks, Hand action recognition.

Abstract:

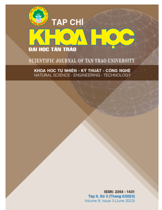
Human posture estimation is important research applied in many fields such as human-machine interaction, surveillance, sports anal-ysis, etc. From there, it is possible to build intuitive and practical applications with science, technology and life. Therefore, fast and accurate estimation of human posture is a pre-processing step but very important in the process of building applications. In this paper, we propose to use Mediapipe, which is a Microsoft built-in frame-work for 3D human pose estimation. The test was evaluated against the MADS (Martial Arts, Dancing, and Sports Dataset) database, in which we focused on sports videos such as: basketball, volleyball, football, rugby, tennis and badminton. The average estimate error is between 100-200mm. The 3D human posture estimation results are a good result in supporting sports analysis.

Vol 9. No 3_May 2023

TẠP CHÍ KHOA HỌC ĐẠI HỌC TÂN TRÀO

ISSN: 2354 - 1431

http://tckh.daihoctantrao.edu.vn/



ƯỚC LƯỢNG TƯ THẾ NGƯỜI 3D TRONG VIDEO THỂ THAO SỬ DỤNG MEDIAPIPE

Đinh Thái Sơn[1], Lương Thị Thanh Minh [2], Đỗ Hữu Sơn[3], Phan Văn Nam[3],Tề Trung Hiếu[3], Lê Văn Hùng[3]*

[1]Trường Đại học Hùng Vương, Phú Thọ, Việt Nam

[2]Trường ĐH Công nghệ thông tin và Truyền thông Thái Nguyên, Việt Nam

[3]Trường Đại học Tân Trào, Việt Nam

[*]Địa chỉ email: dosonhytq@gmail.com

DOI: 10.51453/2354-1431/2023/975

Thông tin bài viết

Ngày nhận bài: 16/12/2022

Ngày sửa bài:06/03/2023

Ngày duyệt đăng: 16/5/2023

Từ khóa:

Bộ xương người 3D, Tích chập, Thần kinh, Mạng, Nhận dạng hành động tay.

Tóm tắt

Ước lượng tư thế người là nghiên cứu quan trọng được áp dụng trong nhiều lĩnh vực như tương tác người máy, giám sát, phân tích thể thao, v.v. Từ đó có thể xây dựng được các ứng dụng trực quan và thiết thực với khoa học công nghệ và đời sống. Do đó việc ước lượng nhanh và chính xác tư thế người là một bước tiền xử lý nhưng rất quan trọng trong quá trình xây dựng các ứng dụng. Trong bài báo này chúng tôi đề xuất sử dụng Mediapipe, là một khung có sẵn của Microsoft cho việc ước lượng tư thế người 3D. Thử nghiệm được đánh giá trên cơ sở dữ liệu MADS (Martial Arts, Dancing, and Sports Dataset), trong đó chúng tôi tập trung vào các video thể thao như: basketball, volleyball, football, rugby, tennis and badminton. Sai số ước lượng trung bình là từ 100-200mm. Các kết ước lượng tư thế người 3D là một kết quả tốt trong hỗ trợ phân tích thể thao.

1. Introduction

In the era of Industry 4.0, the applications of Artificial Intelligence (AI) and computer vision have opened up many potentials in the field of sports. Among them, estimating 3D human pose in sports videos is an area of particular interest, helping to accurately and reliably evaluate and analyze the performance of athletes. To meet this demand, Me- diaPipe - a powerful tool developed by Google, has been widely used in sports research, bringing significant and superior benefits compared to some other 3D human pose estimation methods.

The necessity of 3D human pose estimation in the context of Industry 4.0:

In the era of Industry 4.0, the application of information technology and artificial intelligence in sports is making significant progress. Professional sports tournaments are increasingly being closely monitored and evaluated, from statistical data to live sports data analysis. In particular, estimating 3D human pose in sports videos is becoming an important tool for measuring athletes' performance, thereby improving sports training, skill development, and effectiveness. This requires an accurate, fast, and reliable method to estimate 3D human pose from sports video data, and that is where MediaPipe proves to be an ideal choice.

The importance of MediaPipe in 3D human pose estimation:

MediaPipe is a powerful tool developed by Google, providing significant features for estimating 3D human pose from sports video data. First, MediaPipe supports high-precision detection of head, shoulder, knee, and other points on the body, helping to accurately determine the position of human frames in sports videos. In addition, MediaPipe also provides the ability to estimate 3D human pose, allowing measurement of parameters such as rotation angle, direction of movement, and distance between points on the body. This helps to accurately and fully evaluate athletes' postures during sports competitions, thereby providing important information for making accurate decisions in training.

A prominent advantage of MediaPipe is its consistency and reliability in estimating 3D human pose. MediaPipe has been trained on a large amount of data and optimized to achieve high accuracy, while ensuring stability and reliability in all situations. In addition, MediaPipe also supports parallel computing, allowing for quick processing of sports videos with high frame rates.

Compared to other 3D human pose estimation methods, MediaPipe has several advantages. Firstly, MediaPipe is an open-source tool, which means it is publicly available and can be customized and expanded to meet the specific needs of each project. This provides flexibility and high customization for MediaPipe users in research and practical applications. Secondly, MediaPipe provides 3D human pose estimation features on mobile devices, simplifying deployment and integration into mobile sports applications, such as tracking individual sports activities or detecting poses in sports games. Finally, the consistency and reliability of MediaPipe in 3D human pose estimation is a strength of this tool. MediaPipe achieves high accuracy and reliability in estimating 3D human poses in different conditions, from low light to fast motion, meeting all requirements of real-time sports applications.

With the strong development of digital technology and artificial intelligence applications in the context of the Fourth Industrial Revolution, 3D human pose estimation in sports videos is becoming a promising research and application field. MediaPipe has demonstrated its important role in providing efficient and consistent solutions for 3D human pose estimation in sports videos. With accurate and reliable 3D human pose estimation features, the ability to integrate on mobile devices, and the consistency and stability of results, MediaPipe is becoming a useful and potential tool in research and practical applications in the sports field.

In the current 4.0 context, 3D human pose estimation in sports videos is a promising and potential field. MediaPipe has provided an efficient, consistent, and reliable solution for 3D human pose estimation in sports videos, with the ability to integrate on mobile devices and flexibility in customization according to the specific needs of each project. The importance of MediaPipe lies in its ability to provide detailed information on 3D human poses in sports, improving training, education, and accurate decision-making processes. With superior advantages compared to other 3D human pose estimation methods, MediaPipe is becoming a useful tool in developing research and practical applications in the sports field. The combination of accuracy, consistency, and flexibility makes MediaPipe a promising tool for the future of sports.

2. Related studies

In the article by authors Lin and colleagues [1] They found that Mediapipe's 3D human body recognition library has the highest accuracy in this field. detecting 3D human postures. In this paper, they address this issue by improving the z-depth from the first-person camera because it cannot accurately detect the z-value in human postures when tilted. Adjusting the z-value of the human posture in different body postures is done by normalizing the simulated proportions of each body. Finally, to solve the lag and periodic noise problems in many frames due to the body's movement speed, this paper verifies that the accuracy of 3D human posture detection has improved based on Mediapipe to over 90% through testing multiple postures for people of different heights, weights, ages, and genders.. Fig **??** Indicating the complete inaccuracy of Mediapipe in human pose recognition and indicates that Mediapipe, after calibration, is capable of more accurate recognition.

Fig 1: Mediapipe does not accurately detect human postures, and Mediapipe recognizes human postures after calibration.

In the article by authors Radhakrishna and colleagues [2]. This article presents a new algorithm for extracting a quaternion from a 2D camera, which is used to estimate the pose. The estimation problem is often solved using stereo cameras and gravity sensors to obtain depth (z), but using these devices comes with significant delay and economic cost. Using Mediapipe, the article proposes extracting a quaternion from a 2D human image frame with a delay of less than 50 milliseconds and using low computational resources. The purpose of this algorithm is to detect and react at the last minute for autonomous robots. The algorithm aims to overcome financial barriers and improve availability for control system-related robot researchers.

In the article by authors Radhakrishna and colleagues [3]. They use Mediapipe to develop a new algorithm for limiting traffic accidents for selfdriving cars, as accidents often occur when obstacles (such as pedestrians) suddenly appear in the path of self-driving cars, giving the robots less time to react. They use Mediapipe to build a new algorithm to estimate the intention of any pedestrian selected in an image into a logical state. This allows them to bypass the previously necessary use of a deep learning algorithm with relatively high latency. The model achieved an average test accuracy of 83.56% with a reliable variance of 0.0042 while operating at an average latency of 48 milliseconds, showing significant advantages over the current standard of using convolutional networks for this perception task. Fig 2 represents the algorithm for evaluating the intention of pedestrians.

The next article by authors V.A.R.Barao and colleagues [4]. They evaluated Google Mediapipe Pose Estimation to be able to predict and return the keypoints of a human pose based on a single viewpoint with high accuracy. Compared to other human pose estimation algorithms such as YOLO, Google Mediapipe can estimate poses in all low- latency platforms. In this paper, they designed functions to extract useful parameters from Google Mediapipe and designed useful applications based on the extracted information from Mediapipe Pose. With these parameters, they designed practical applications on all camera-related platforms and fields.

In the article by authors Kim Jong Wook and colleagues [5] They used Mediapipe to build a model for detecting and recognizing human motion in order to detect falls and injuries in elderly people living alone. Although deep learning methods are actively developing in this field, they have limitations in estimating missing or occluded poses in the training dataset. Therefore, they propose a lighter approach combining an available 2D pose estimation method, a more complex human model, and a fast optimization method to estimate joint angles for 3D pose estimation. This is a new idea, and the issue of deep ambiguity in 3D pose estimation is resolved by adding a loss function based on deviations of the center of mass from the centers of the two supporting feet and functions related to the appropriate range of joint angles. To verify the accuracy of this model, they performed six daily pose returns for estimation. With an average joint coordinate error of 0.097m and an average angle error per joint of 10.017 degrees. Additionally, to confirm the practicality, videos of exercise activities and a person falling were recorded, and joint angle trajectories were generated as a result of 3D pose estimation. The realtime execution time for optimization on each frame was measured to be 0.033 seconds on a non-GPU computer, showing the feasibility of the proposed method similar to a real-time system. Fig 3 illustrates the overall diagram of the proposed model.

3. Using Mediapipe to estimate 3D human poses from sports videos.

3.1. Mediapipe Pose

MediaPipe Pose (MPP) is a cross-platform opensource framework provided by Google, used to estimate the 2D human joint coordinates in each image frame. MediaPipe Pose builds pipelines and processes perception data as video using machine learning (ML). MPP uses BlazePose [6] to extract 32 2D features on the human body as shown in fig 5. BlazePose is a lightweight ML architecture that achieves real-time performance on mobile phones and personal computers with CPU inference capability. When using normalized coordinates to estimate poses, the inverse ratio must be multiplied by the y-axis pixel value. Among the estimated MPP features, we use 12 features to estimate arbitrary poses and motions, which are 11, 12, 13, 14, 15, 16, 23, 24, 25, 26, 27, and 28, as shown in Fig 4.

3.2. Humanoid Robot Model

The human body must be represented by a humanoid robot model with joints and links similar to those of humans to recreate 3D human postures from 2D joint data collected from the MPP joint capture system. Therefore, arbitrary 3D postures can be reconstructed from 2D images taken at different angles and distances from the camera by measuring the length of the links in pixels and estimating the joint angles of the humanoid robot model using optimization methods. In general, the humanoid robot is described by links and joints based on the Denavit-Hartenberg (DH) method [7], where a reference coordinate frame is placed on the support base. As the goal of the current method is to generate and estimate postures of the humanoid robot that are as close to human-like as possible, we have improved our previous humanoid robot model [8] to create arbitrary postures as follows:

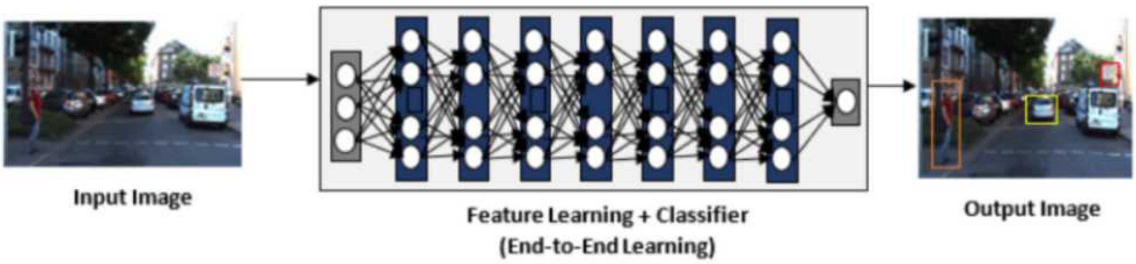


Fig 2: Represents the algorithm for evaluating the intention of pedes

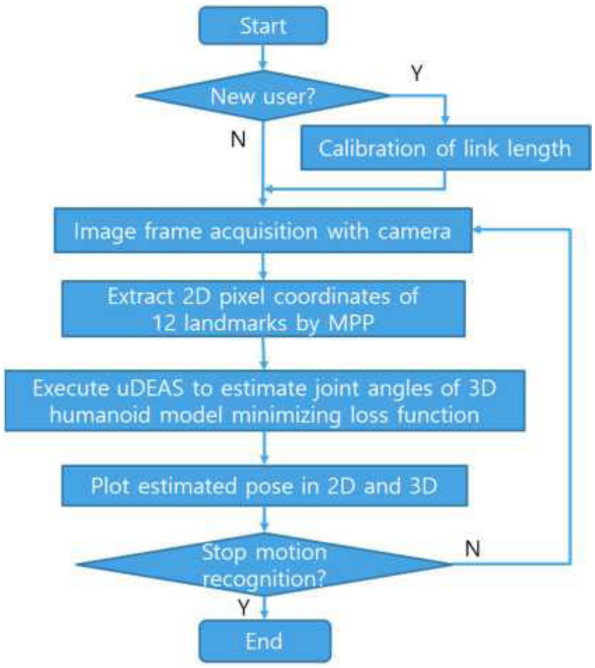


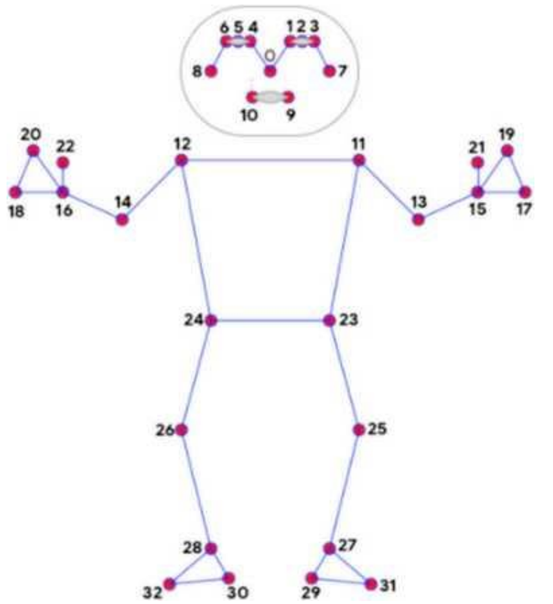
Fig 3: The overall diagram of the proposed model.

• Set the origin of the reference coordinate system at the center of the body joint coordinates, in order to create arbitrary postures.

• Add three lumbar spine joints with 3 DoF at the center of the pelvis bone to create postures with upper body movement separately.

• The redefinition of the rotation direction for all joint variables is necessary to align with the Vicon joint collection system for easy exchange of joint measurement data within the system.

As shown in the fig 5, The proposed humanoid robot model includes a total of 23 joint variables, comprising 12 joint angles with axes perpendicular to the sagittal plane and 7 joint angles with axes perpendicular to the frontal plane. (**0***hd*, **0***lhh*, **0***lr*, **0***tr*, **0***hp*, **0***ln*, **0***an*) There are 7 joints with rotation axes perpendicular to the frontal plane of the body. (**O***sh*, **o***tr*, **o***hp*, **o***an*) and 4 joint angles have axes perpendicular to the frontal plane of the body. ($\wedge_{hd}$, $\wedge_{tr}$, ^*hp*) Here, the indices under hd, sh, el, tr, hp, kn, and an refer to the joints of the head, shoulders, elbows, upper body, hips, knees, and ankles, respectively, and the indices on l and r denote the left and right sides, respectively.



0. nose
1. Ieft_eye_inner
2. Ieft_eye
3. Ieft_eye_outer
4. right_eye_inner
5. right_eye
6. right_eye_outer
7. Ieft_ear
8. right.ear
9. mouthjeft
10. mouth_right
11. Ieft_shoulder
12. right.shoulder
13. Ieft_elbow
14. right_elbow
15. Ieft_wrist
16. right_wrist
17. Ieft_pinky
18. right_pinky
19. leftjndex
20. rightjndex
21. Ieft_thumb
22. right_thumb
23. Ieft_hip
24. right_hip
25. Ieft_knee
26. right_knee
27. Ieft_ankle
28. right_ankle
29. Ieft_heel
30. right.heel
31. Ieft_foot_index
32. right_footjndex

Fig 4: Definition of features in Mediapipe Pose.

3.3. Camera reflection effect

When deciding on full-body posing, the position and angle of the camera with respect to the subject are also important factors. The camera position determines the overall size of the body, which can be reflected in a human figure model by multiplying the factory dimensions with all the displayed limb lengths in fig 5. It means that as the camera moves further away, the size of the body decreases, that is Y <1 And vice versa, which means Y > 1. Moreover, the relative camera viewpoint causes the same standing posture to look different, as shown in fig 6. Fig 6a illustrates the scenario where the camera captures an object from top to bottom. Fig 6b illustrates a situation in which the camera is rotated 90 degrees clockwise. Fig 6c shows the camera viewing the object from the front left. The differences in posture due to this camera angle can be mathematically described by the relationship between the body coordinate frame and the camera-based coordinate frame, as illustrated in fig 7. Body vertical angle $Ó_{bd}$, Body oblique angle 0_{bd},and the horizontal angle of the body $\wedge_{bd}$ corresponding to three posture changes in fig 6. The polarities of these body angle parameters are determined to conform to the Vicon sign convention for joint angles, for example $Ó_{bd} > 0$ to tilt forward, $0_{bd} > 0$ to tilt towards the left, and $\wedge_{bd} > 0$ to rotate left [9] And vice versa.

3.4. Fast global optimization method

To estimate joint angles in a humanoid robot model that corresponds to the MPP bone model for the current frame, an inverse kinematics mechanism is necessary. However, solving inverse kinematics based on the formula of a structured humanoid robot like the one in fig 5 is time-consuming. uDEAS was developed to address nonlinear and multimodal technical issues in the field of information technology. uDEAS verifies the globally optimal performance fastest and most reliably on seven low-dimensional standard benchmark functions (from two to six dimensions), three highdimensional benchmark functions (up to 30 di-mensions) [10], optimal designs of Gabor filters [11], and generates common paths for the up and down stairs walking of a humanoid robot [12]. In addition, a modified version of uDEAS can also search for integer variables, called cDEAS (Combinatorial DEAS), which has recently been developed and applied to optimize the hybrid energy systems [13]. uDEAS is a global optimization method that combines local and global search methods. For the local search method in uDEAS, all optimization variables are represented by binary strings, similar to the genetic algorithm (GA) [14]. The basic unit of local search is a session formed by a single binary split search (BSS) and multiple one-dimensional search operations (UDS) with a binary string for each variable. BSS adds 0 and 1 to the end of the selected string as the least significant bit (LSB), for example: 010_2 ^ 012 ^ 0112, where inserting 0 (1) into the new LSB position of the binary string corresponds to decreasing (increasing) its decoded real value compared to the parent string [15]. For example, suppose the binary string 010_2 is decoded by a decoding function into a real value of 0.3 and a cost value of 0.7, i.e., J(d(0102)) = J(0.3) = 0.7. Meanwhile, the binary string 0112 is decoded into 0.1 and its cost value is 0.3, i.e., $J(d(011_2)) = J(0.1) = 0.3$. Since J(0.1) < J(0.3), increasing the current variable will become a good search plan thanks to BSS. Meanwhile, UDS adds or subtracts 1 from the binary string in an optimally directional manner found in the previously discovered BSS. With the example of BSS mentioned above, UDS will generate a better-directed binary string than BSS, such as 1002 (the first UDS) ^ 1012 (the second UDS) ^ until no further cost reduction can be achieved. This BSS-UDS pair plays a crucial role in balancing exploitation and exploration in local search.

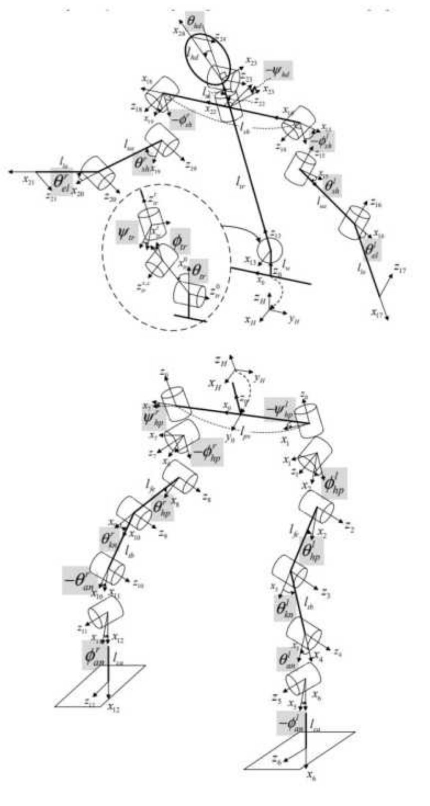


Fig 5: 3D humanoid Robot Model.

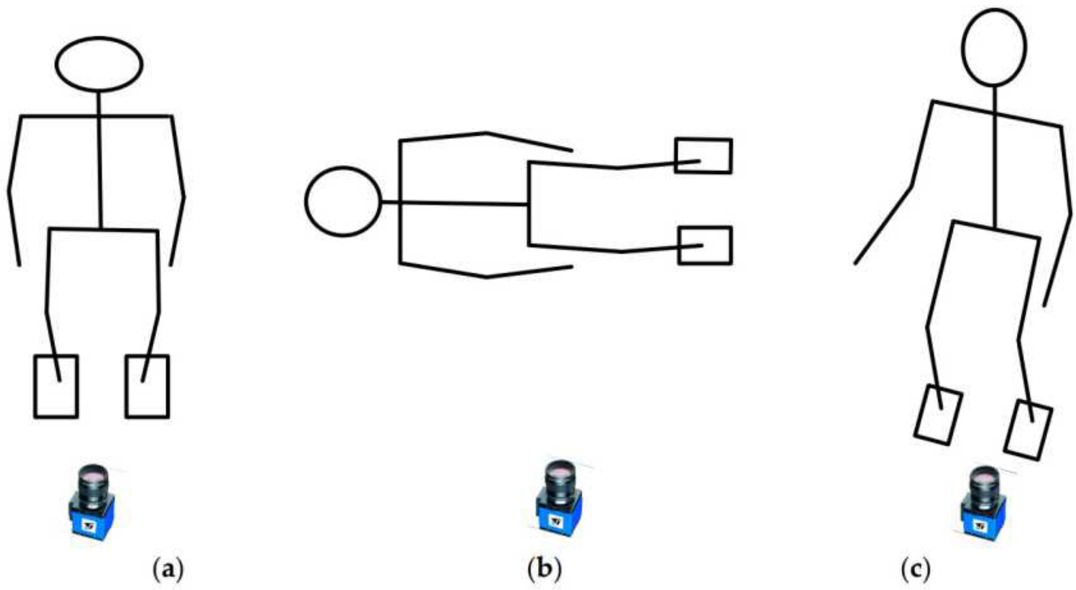


Fig 6: The mesmerizing poses look different depending on the camera's perspective: (a) from top to bottom, (b) rotated 90 degrees clockwise, (c) from the front left.

For the n-dimensional problem, uDEAS stacks n chains together to form a binary matrix with n rows.

• Step 1. Initialization: Create an n X m binary matrix with elements being randomly chosen binary digits. The length of the row is set to m0. The optimization variable vector is defined as v = [v1v2..vn]T.

• Step 2. Start the first working session with i =1.

• Step 3. BSS: From the current best matrix M, the binary vector of the jth row (= J(i)) is chosen:

• j(M) = [amaj(m-1)..aj 1],ajk = [0,1], 1 < k < m (1)

Adding 0 or 1 to the LSB (least significant bit) of the row vector will yield:

rj = [a jm a j(m-1)..a j1 0], rj = [ajmaj(m-1)..aj1 1] (2)

Then, these sequences are decoded into real values and replaced by the j-th variable of the current optimal variable vector v* as follows:

v- = v+ = v*,v-(j) = d(rj-),v+(j) = d(r+j) (3)

For the n-dimensional problem, uDEAS stacks n chains together to form a binary matrix with n rows.

Step 4. UDS: Depending on the direction of u(j), perform addition or subtraction on row j, described as follows:

rj(M) = r*j(M) + u(j) (4)

Check if the new item rj contributes to further loss reduction. If yes, update the binary sequence and current variables as the new optimum, and continue to Step 4.

On the contrary, proceed to Step 5.

• Step 5. Save the UDS sequence of the best result, r*(M), into the j-th row of the current best matrix.

• Step 6. Ifi < n, set i = i + 1. Continue to Step 3. Otherwise, if the current string length m is shorter than the maximum allowed length mf, set i = 1, increase the string length index to m = m + 1, and continue to Step 2. In the case where m = mf , proceed to Step 7.

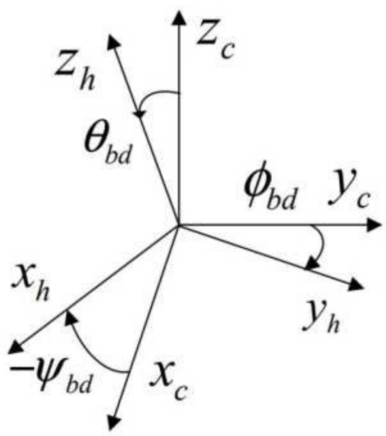


Fig 7: The relationship between the $x_h y_h z_h$ reputation coordination of the body and the camera-based coordinates of frame $x_c y_c z_c$

• Step 7. If the number of restarts is less than the specified value, continue to Step 1. Otherwise, end the current local search process and select the global minimum value with the smallest cost value among the local values found so far.

4. Experiment and Results

4.1. Database

The purpose of collecting data on martial arts, dance, and sports (abbreviated as MADS) is to provide challenging action sequences for estimating posture from multi-view or depth data. The MADS dataset includes 5 types of actions, including TaiChi, Karate, Jazz, Hip-Hop, and Sport. All actions are performed by professionals, and they provide video data on actual action postures in the real world, not from existing datasets. For depth video, they chose stereo cameras as the data collection method because there are no datasets that collect human actions using this method - although the noise factor is high, stereo cameras are not affected by infrared noise and can operate outdoors.

4.2. Dataset

Their dataset includes actions from martial arts, dance, and sports, which are common actions but not everyday actions in life. Tai-chi is a traditional Chinese martial art consisting of smooth movements. Karate is a Japanese martial art with many striking actions such as punching, kicking, elbow strikes, and bare-hand techniques. The selected martial arts actions are very different from everyday actions and require more strength, speed, and balance.

Jazz dance evolved from American vernacular dance and includes many body spins and arm movements with a wide range of turning motion. Hip-hop dance involves street dance moves with hip-hop music and includes hip thrusts as well as many shoulder and body rolls. For dance actions, there are more body spins than other types of actions, which makes tracking based on dynamics more difficult.

The sports videos include badminton, basketball, soccer, tennis, and volleyball, which are popular sports worldwide. For convenience of data collection, they allowed actors to perform actions without a ball or racket. The actions for basketball, soccer, and volleyball include shooting, passing, and blocking. For badminton and tennis, the actions recorded are hitting the ball. For football, the actions include throwing and blocking. The actions in MADS are more complex and challenging than normal actions. First, they have a large range of motion, while some postures will not appear in normal actions. Second, there is more occlusion and many interactions between limbs. Third, some actions will occur very quickly, compared to the frame rate used to record video (10 frames per second for Tai-chi and Karate, and 20 frames per second for jazz, hip-hop, and sports), leading to motion blur.

A total of 5 actors were used to collect data, with each actor performing a specific type of action. Two martial arts experts performed prearranged sequences of movements, called “forms” in Tai-chi or “katas” in Karate, two professional dancers to perform jazz and hip-hop dance routines, and one athlete to perform sports movements. The subjects signed a consent form allowing the distribution of data for academic purposes.

The subjects wore natural clothing. Multiangle and 3D depth videos were recorded separately. As the 3D depth video was recorded from only one viewpoint, the subjects were instructed to adjust their movement sequences to face the camera as much as possible. In particular, the subjects were asked to avoid turning their back to the camera during any movements. For example, if the movement involved a 180-degree turn, and the subject ended up with their back facing the camera, we requested that the subject perform a complete 360-degree turn. In another example, if the subject was facing the camera on the right and the movement was a turn to the left, leaving their back facing the camera, the subject would turn to the right to face the camera. Therefore, the action sequences in the multi-angle and 3D depth videos are different for some movements.

5. Conclusions and Future Works

Estimating 3D human posture from color image data is a challenging research problem in computer vision. Despite being the subject of much research attention, the results obtained still suffer from significant errors. In this paper, we conducted an experiment on using the Mediapipe Framework to estimate 3D human posture from RGB sports video. The estimation results still had a large error. In the future, we will continue to improve the estimation models using convolutional neural networks to achieve higher estimation accuracy.

4.3. Results

The results of estimating 3D human posture on sports videos (basketball, volleyball, football, rugby, tennis and badminton) from the MADS database are presented in table 1.

Fig 8 illustrates the results of estimating the posture of a badminton player from the MADS database, with the input being a color image and the output being the estimated human pose in 3D space. The estimated human joints are represented as red points, and the segments connecting the joints are the parts of the human body.

Tab 1: The results of 3D human pose estimation on sports videos from the MADS dataset.

Deep learning model	Videos	MPJPE (mm)
	basketball	124.8
Mediapipe Pose [6]	volleyball football	120.7 118.5
	rugby	115.4
	tennis	112.1
	badminton	111.2

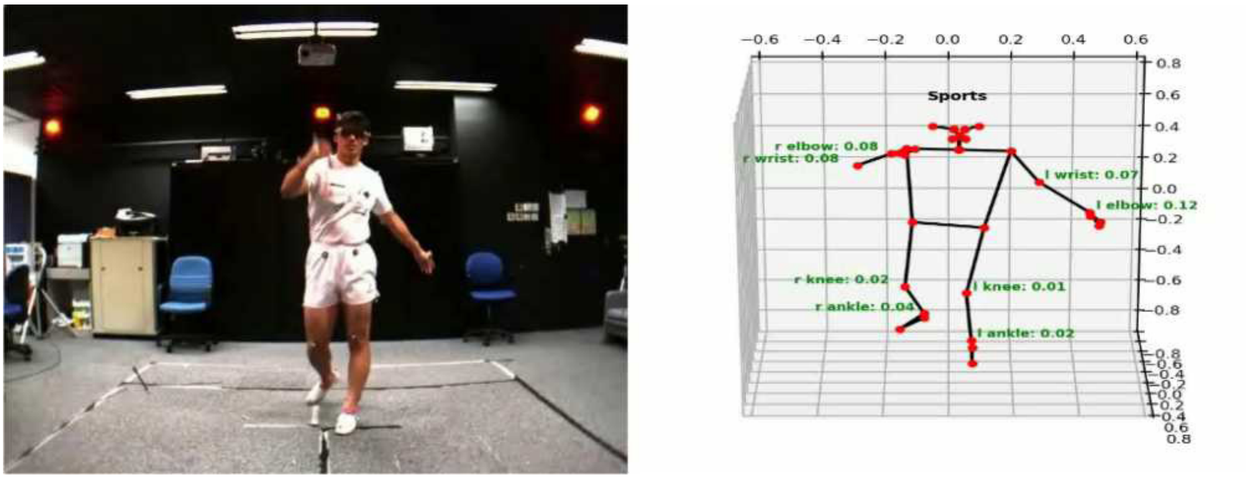


Fig 8: Illustration of 3D human pose estimation results. On the left is the input color image data, and on the right is the estimated 3D skeleton frame of the human.

References

[1] Y. Lin, X. Jiao, L. Zhao, "Detection of 3D Human Posture Based on Improved Mediapipe," Journal of Computer and Communications, **11**(02), 102-121, 2023, doi:10.4236/jcc.2023.112008.

[2] S. Radhakrishna, A. Balasubramanyam, "Economical Quaternion Extraction from a Human Skeletal Pose Estimate using 2-D Cameras," 2023.

[3] S. Radhakrishna, A. Balasubramanyam, "Pedestrian Intention Classifier using ID3 Modelled Decision Trees for IoT Edge Devices," https://arxiv.org/abs/2304.00206, 2023.

[4] V.A.R.Barao, R.C.Coata, J.A.Shibli, M.Bertolini, J.G.S.Souza, "APPLICATIONS OF GOOGLE MEDIAPIPE POSE ESTIMATION USING A SINGLE CAMERA," Braz Dent J., **33**(1), 1-12, 2022.

[5] J. W. Kim, J. Y. Choi, E. J. Ha, J. H. Choi, "Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model," Applied Sciences (Switzerland), **13**(4), 2023, doi:10.3390/app13042700.

[6] Bazarevsky, V., Grishchenko, I. On-Device, "Real-Time Body Pose Tracking with MediaPipe BlazePose," Google Research, 2020.

[7] Denavit, J., Hartenberg, R.S, "A kinematic notation for lower-pair mechanisms based on matrices," Appl. Mech. 1955, 77, 215-221, 1955.

[8] Kim, Tran, Dang, Kang, "Motion and walking stabilization of humanoids using sensory reflex control," Int. J. Adv. Robot. Syst, 2016.

[9] V. Goldberg, Available online, https://www.vicon.com/, 2021.

[10] J. W. Kim, T. Kim, J. Y. Choi, S. W. Kim, "On the global convergence of univariate dynamic encoding

algorithm for searches (uDEAS)," Int. J. Control Autom. Syst, 2008.

[11] J. Yun, Choi, J.-W. Kim, S. Kim, " Automatic detection of cracks in raw steel block using Gabor filter optimized by univariate dynamic encoding algorithm for searches (uDEAS)," NDT E Int, 2009.

[12] E. Kim, M. Kim, S. Kim, J. Kim, "Trajectory generation schemes for bipedal ascending and descending stairs using univariate dynamic encoding algorithm for searches (uDEAS)," Int. J. Control Autom. Syst, 1061-1071,2010.

[13] J. Kim, H. Ahn, H. Seo, S. Lee, "Optimization of So- lar/Fuel Cell Hybrid Energy System Using the Combinatorial Dynamic Encoding Algorithm for Searches (cDEAS)," Energies 2022, 2779, 2022.

[14] D. Goldberg, "Genetic Algorithm in Search, Optimization and Machine Learning," Addison Wesley: Berkeley, CA, USA, 1999.

[15] J.-W. Kim, T. Kim, Y. Park, S. Kim, " On load motor parameter identification using univariate dynamic encoding algorithm for searches (uDEAS)," IEEE Trans. Energy Convers, 804—-813, 2008.