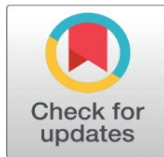


REAL-TIME MOTION IMAGE POSE DECOMPOSITION, CLASSIFICATION AND ANALYSIS

Chien-Hsing Huang ¹  

¹ Department of Information Engineering, I-Shou University, Kaohsiung City, Taiwan, R.O.C., China



Received 03 April 2024
Accepted 06 May 2024
Published 07 June 2024

Corresponding Author

Chien-Hsing Huang, raylan@isu.edu.tw

DOI

[10.29121/ijetmr.v11.i6.2024.1464](https://doi.org/10.29121/ijetmr.v11.i6.2024.1464)

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Copyright: © 2024 The Author(s). This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

With the license CC-BY, authors retain the copyright, allowing anyone to download, reuse, re-print, modify, distribute, and/or copy their contribution. The work must be properly attributed to its author.



ABSTRACT

Exercise can boost metabolism and make the body healthier. It also increases metabolic rate, helping to consume more calories and burn fat. Regular exercise can stimulate the brain to secrete endorphins, making people feel relaxed and happy, and improves self-confidence, and has been shown to reduce symptoms in people with depression and anxiety. However, incorrect exercise posture may cause harm to the body, such as torn ligaments or muscle strains, so good exercise posture is needed to improve sports performance. This article proposes to use artificial intelligence image analysis method to decompose fitness exercise posture images and establish exercise posture cycle samples to assist in completing fitness exercises.

Keywords: Computer Vision, Mediapipe Pose, HNN, Image Classification

1. INTRODUCTION

There is good evidence that both running and football improve aerobic and cardiovascular function at rest, and football can reduce obesity. Conditional evidence suggests that running benefits metabolic health, adiposity, and postural balance, while football improves metabolic health, muscle performance, postural balance, and cardiac function. Evidence for the health benefits of other forms of exercise is either inconclusive or tenuous. Only the correct way of exercise can help your health, otherwise it will not help your health or even harm your health. [Oja et al. \(2015\)](#)

Research into using computer vision to analyze human posture dates back to the 1960s. At the time, researchers used manual feature extraction and matching

methods to estimate human pose. This method is less accurate and sensitive to noise and occlusions in the image. [Andriluka et al. \(2014\)](#), [Dang et al. \(2019\)](#), [Wang et al. \(2021\)](#)

In the 1980s, researchers began using human body models to estimate human posture. This method is more accurate and more robust to noise and occlusions in the image. However, the modeling of human body models is more complex and requires a large amount of data for training.

In the 1990s, researchers began using machine learning algorithms to estimate human posture. The accuracy and robustness of this method are significantly improved. However, machine learning algorithms require large amounts of data for training and are sensitive to noise and occlusions in images.

Since the 21st century, with the development of computer software and hardware, research on human posture analysis has made significant progress. Currently, the accuracy and robustness of human posture analysis have been significantly improved and applied in many fields.

Milestones in the development of human posture analysis:

- **1960s:** Human pose estimation using artificial feature extraction and pairing methods.
- **1980s:** Estimating human pose using human body models.
- **1990s:** Use of machine learning algorithms to estimate human posture.
- **21st century:** The accuracy and robustness of human posture analysis have improved significantly and are applied in many fields.

Future development trends of human posture analysis:

- **Higher accuracy and robustness:** As machine learning algorithms continue to be improved, the accuracy and robustness of human posture analysis will be further improved.
- **Wider application range:** Human posture analysis will be used in more fields, such as virtual reality, augmented reality, robots and other fields.
- **A more humane interaction method:** Through human posture analysis, a more humane interaction method is provided, such as controlling the device through gestures or controlling the device through eye tracking.

According to different sports attributes, sports skills can be divided into the following five categories: [Guthrie \(1952\)](#), [Schmidt \(1991\)](#), [Tereshchenko et al. \(2015\)](#)

- Divided into "open skills" or "closed skills" based on environmental stability.
- According to the continuity of actions, they are divided into "discrete skills", "continuous skills" or "sequential skills".
- Divided into "motor skills" or "cognitive" skills according to the range of activities.
- Classified as "gross motor skills" or "fine motor skills" based on the muscles used.
- Divided into "simple skills" or "complex skills" based on action combinations.

Closed skills such as gymnastics, individual skills such as throwing, athletic skills such as high jumping, and simple skills such as running are all simple movements performed repeatedly in a stable environment. Performing or training such sports requires stable monitoring of the movement process, so it is suitable to use sports posture analysis assistance to reduce the burden on coaches and provide timely feedback to athletes.

2. MEDIAPIPE POSE

Artificial intelligence's image analysis of sports postures has evolved from the past image classification to distinguish different postures, to analyzing the spatial positions of human joints and branches in images, and establishing human movement models. However, subsequent feature classification processing is required to distinguish different movement postures.

This article uses MediaPipe Pose developed by Google Research to analyze the human posture in the image, obtain the spatial coordinates of the human body feature points, calculate the curvature of each joint of the human body, normalize it into a posture feature vector, and then use unsupervised machine learning to analyze the posture movement. Automatic classification. , find the cyclic samples of gesture motion from the time series of moving images. Samples can be compared with real-time sports images to help athletes complete cycle movement postures correctly.

The MediaPipe framework focuses on (a) selecting and developing appropriate machine learning algorithms and models, (b) building a series of prototypes and demonstrations, (c) balancing resource consumption with solution quality, and finally (d) identifying and mitigating problem cases. Developers can use MediaPipe to build prototypes by combining existing awareness components, improve them into complete cross-platform applications, and measure system performance and resource consumption on the target platform. [Lugaresi et al. \(2019\)](#)

The MediaPipe pose landmark task detects human body landmarks in images or videos. Can be used to identify key body positions, analyze postures and classify movements. This task uses a machine learning (ML) model that processes a single image or movie. This task outputs body pose landmarks in image coordinates and 3D world coordinates. [Google MediaPipe \(n.d.\)](#)

Figure 1

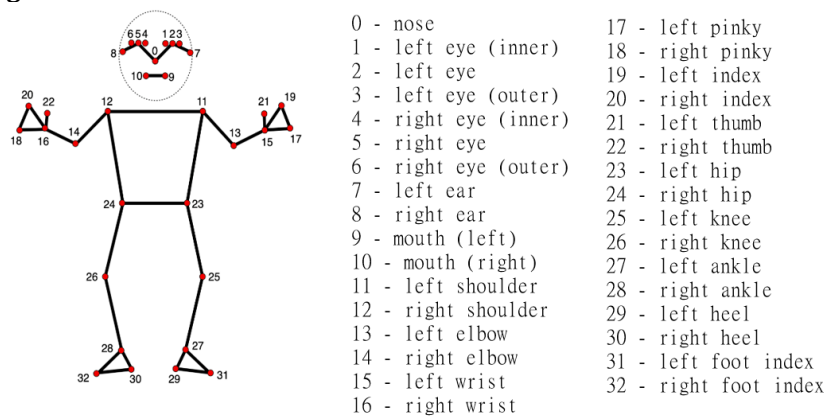


Figure 1 33 Body landmark locations, representing the approximate location of the body parts

From https://developers.google.com/mediapipe/solutions/vision/pose_landmarker Schmidt (1991)

After obtaining the spatial coordinates of each joint point of the body through MediaPipe, the bending degree of each joint is used as a feature of human posture classification, and the left shoulder, left elbow, right shoulder, right elbow, left hip, left knee, right hip and the angle between the right knees constitutes the feature vector P^f representing the posture:

$$P^f = [A_{23,11,13} \quad A_{11,13,15} \quad A_{24,12,14} \quad A_{12,14,16} \quad A_{11,23,25} \quad A_{23,25,27} \quad A_{12,24,26} \quad A_{24,26,28}] \quad (1)$$

$$A_{i,j,k} = \cos^{-1}\left(\frac{\vec{j} \cdot \vec{k}}{|\vec{j}| |\vec{k}|}\right) \quad (2)$$

3. HOPFIELD NEURAL NETWORK (HNN)

Next, you need to build a machine learning module to learn to distinguish posture changes at different stages during exercise. Considering the sample differences in image sampling and the reasonable variation range of human movement, this article uses the associative learning network to learn the training samples. When the input state data is incomplete and complete data needs to be inferred, apply the example's internal storage rules. [Pajares \(2006\)](#)

Hopfield Neural Network (HNN) is an associative learning network proposed by J. Hopfield in 1982 and consists of feature vectors. By learning the associative memory rules, remember the feature vectors of the training samples. Incomplete or noisy feature vectors can approximate the most similar training samples in iterative operations through associative memory rules to complete sample classification.

HNN is a single-layer network. Use the sigmoid function to perform bipolar processing on the feature vector P^f obtained using Mediapipe in the previous section and then input it, so that the variable value processed by each processing unit is bipolar and exists in any two processing units. The interaction relationship is represented by the connection weight W_{ij} , which is the sum of the products of the i -th and j -th features in all training samples. The state value vector $Net^{(k+1)}$ of the processing unit is defined as $W \cdot X^{(k)}$, where W is the interaction relationship between the features of the training sample, and $X^{(k)}$ is the feature of the training sample after k times iteration. After the iterative operation is completed, the state value $Net^{(k+1)}$ is updated to the new sample value $X^{(k+1)}$, and the sample is compared to see whether it converges to an approximately known sample. If convergence occurs, classification is complete. Otherwise, continue iterating until the sample is close to the known sample.

$$S(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

$$W_{ij} = \sum_F P_i^f \times P_j^f \quad (4)$$

$$Net^{(k+1)} = W \cdot X^{(k)} \quad (5)$$

$$X_i^{(k)} = \begin{cases} 1 & Net_i^{(k)} > 0 \\ 0 & Net_i^{(k)} = 0 \\ -1 & Net_i^{(k)} < 0 \end{cases} \quad (6)$$

4. EXPERIMENTAL RESULTS

Divide the squat exercise into four stages: Stage 1: Stand with your feet shoulder-width apart; Stage 2: Bend your knees smoothly and lower your hips; Stage 3: Bend your knees until your thighs are parallel to the ground; Stage 4 : Extend your knees and return to the first phase. A complete set of squats should consist of one to four phases. Calculate posture weights 1, 0, and -1 to represent the three stages of standing, excessive exercise, and squatting respectively. The schematic posture is shown in [Figure 2](#) below.

Figure 2

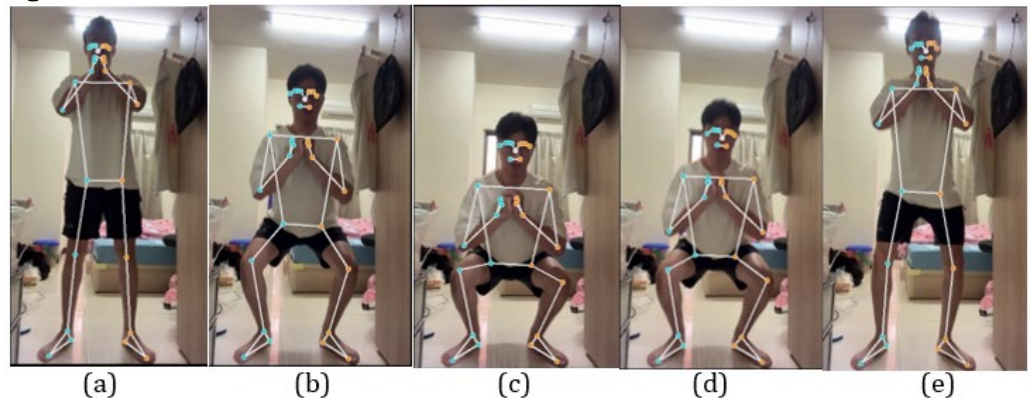


Figure 2 Classification of Squat Postures: (A) Upright; (B) Upright to Squat; (C) Squat; (D) Squat to Upright; (E) Upright.

[Figure 3](#) shows the classification of squat exercise video pictures. The peaks and troughs represent standing and squatting respectively. The athlete's fatigue level can be seen from the duration. It can be found from the cycle of posture changes that after the 125th and 170th frames, the athletes did not complete the squat.

Figure 3

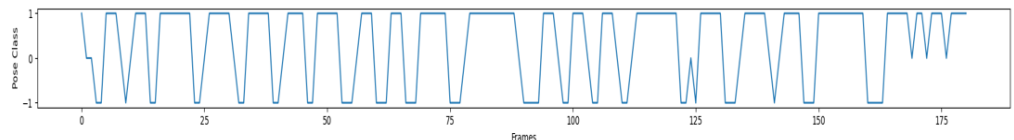


Figure 3 The Posture Changes in the Squat Motion Video, The Horizontal Axis is the Frame Number, and the Vertical Axis is the Athlete's Posture in the Frame: 1 is the Standing Posture; 0 is the Transition Posture Between Standing and Squatting; -1 is the Squatting Posture.

5. CONCLUSIONS

Theoretically, a Hopfield network with n processing units can represent $2n$ possible image samples. Hopfield himself proposed in 1982 that the memory capacity limit of this network was $0.15n$. Other scholars know from theoretical discussions that the memory capacity is $n/(4 \cdot \log_2 n)$. This experiment uses MediaPipe Pose to process images, calculate human skeleton information, and then calculate the bending angles of limb joints as posture classification features, which can effectively compress the image content into feature vectors. Vector information enables HNN to remember the sample feature classification of motion postures. Compared with directly using the spatial distribution of pixels in the image as a processing unit, it can better utilize the memory association mechanism of HNN.

In this experiment, HNN can converge in 10 iterations, that is, the athlete's action phase can be quickly distinguished within one frame. From long-term sports image recordings, changes in frequency cycles can analyze the smoothness or slowness of an athlete's posture. Very helpful for analyzing repetitive motion type movements. People who contributed to the work but do not fit criteria for authorship should be listed in the Acknowledgments, along with their contributions. It is advised that authors ensure that anyone named in the acknowledgments agrees to being so named. Funding sources that have supported the work should also be cited.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

- Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014). 2d Human Pose Estimation: New Benchmark and State of the Art Analysis. In Proceedings of the IEEE Conference on computer Vision and Pattern Recognition, 3686-3693. <https://doi.org/10.1109/CVPR.2014.471>
- Dang, Q., Yin, J., Wang, B., & Zheng, W. (2019). Deep Learning Based 2d Human Pose Estimation: A Survey. *Tsinghua Science and Technology*, 24(6), 663-676. <https://doi.org/10.26599/TST.2018.9010100>
- Google MediaPipe (n.d.). Pose Landmark Detection Guide.
- Guthrie, E. R. (1952). *The Psychology of Learning*. New York: Harper & Row.
- Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., & Grundmann, M. (2019). Mediapipe: A Framework for Building Perception Pipelines. <https://doi.org/10.48550/arXiv.1906.08172>
- Oja, P., Titze, S., Kokko, S., Kujala, U. M., Heinonen, A., Kelly, P., & Foster, C. (2015). Health Benefits of Different Sport Disciplines for Adults: Systematic Review of Observational and Intervention Studies with Meta-Analysis. *British Journal of Sports Medicine*, 49(7), 434-440. <https://doi.org/10.1136/bjsports-2014-093885>
- Pajares, G. (2006). A Hopfield Neural Network for Image Change Detection. *IEEE Transactions on Neural Networks*, 17(5), 1250-1264. <https://doi.org/10.1109/TNN.2006.875978>
- Schmidt, R. A. (1991). *Motor Learning & Performance: From Principles to Practice*. Champaign, IL: Human Kinetics. <https://psycnet.apa.org/record/1993-97677-000>
- Tereshchenko, I., Otsupok, A., Krupenya, S., Liauchuk, T., & Boloban, V. (2015). Coordination Training of Sportsmen, Specializing in Sport Kinds of Gymnastic. *Physical Education of Students*, 19(3), 52-65. <https://doi.org/10.15561/20755279.2015.0307>
- Wang, J., Tan, S., Zhen, X., Xu, S., Zheng, F., He, Z., & Shao, L. (2021). Deep 3D Human Pose Estimation: A Review. *Computer Vision and Image Understanding*, 210. <https://doi.org/10.1016/j.cviu.2021.103225>