A single camera based automatic movement assessment system to rehabilitation exercise therapy

Shuoheng YANG
Department of Orthopaedics and Traumatology
The University of Hong Kong
Hong Kong, China
shuoheng@connect.hku.hk

Xiaodong LI
Orthopedics Center
The University of Hong Kong - Shenzhen Hospital
Shenzhen, China
lixd@hku-szh.org

Zhuoman XU
Department of Physiotherapy
The University of Hong Kong - Shenzhen Hospital
Shenzhen, China
xuzm@hku-szh.org

Yunli FAN
Department of Physiotherapy
The University of Hong Kong - Shenzhen Hospital
Shenzhen, China
fanyl@hku-szh.org

Yong HU
Department of Orthopaedics and Traumatology
The University of Hong Kong
Hong Kong, China
yhud@hku.hk

2024 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA) | 979-8-3503-2299-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/CIVEMSA58715.2024.10586470

***Abstract*—Rehabilitation exercise therapy movement automatic assessment is a promising field, the potential applications including reducing hospital service costs and assisting patients in home exercises. In recent years, AI based movement assessment methods have been widely proposed. However, existing research has the following issues: 1) The devices used to obtain human skeletal coordinates are complex and difficult to deploy,which has hindered its applications in large scale. 2) The scoring methods used for assessment are often lack clear feedback, making it difficult for patients to receive movement corrective advice. To address above problems, this study has proposed an analysis system named the SC-AMAS(single-camera based automatic movement assessment system). The system utilizes human skeletal coordinates estimation technology to obtain estimated 3D skeleton keypoints from a single RGB-camera,which have significantly reduce the usage complexity. And the scoring method in this system is constructed by computational rules from medical expert knowledge, which can provide patients with clear text type feedback, as well as the movement performance score. Additionally, our system applies a privacy protection mechanism to the video stream, preventing the leakage of patients' physical information when third parties view the video. We hope that this intelligent system can promote the feasibility of related research and development.**

***Index Terms*—physiotherapy, movement assessment**

I. Introduction

Rehabilitation exercise therapy requires long-term commitment from patients, but the high cost of medical treatment and shortage of professional physical therapists often limit patient compliance, thus affecting the effectiveness of rehabilitation. In recent years, with the development of artificial intelligence technology, there's a strong interest in rehabilitation movement assessment systems based on automatic evaluation methods has raised. The potential applications of these methods including assisting physical therapists in movement assessment, which can enable them to serve more patients simultaneously, and supporting patients in performing rehabilitation exercises at home, thereby reducing transportation costs to hospitals. However, recent research has demonstrated the following characteristics:

(i)Existing literature extensively uses complex and expensive devices such as inertial measurement units (IMUs) [11] [12], RGB-D cameras [17] [16] [18] [19], and movement capture devices [14] [15]. These devices can accurately capture the precise position of the patient's body, providing high-accuracy assessment results. However, due to the high deployment cost, it is challenging to apply them in the patient's daily training environment. Therefore, existing research lacks feasible large-scale device application solutions.

(ii)When pure machine learning model was used for movement performance scoring, although more accurately match the scoring results of human experts. Due to the limited interpretation ability of machine learning models, existing scoring systems struggle to identify the exact time when the wrong movement occurred during the movement process and provide corrective advice. Therefore, patients find it difficult to obtain guidance from these systems [1] [13].

To address these issues, recent literature has proposed potential solutions. In Reference [6], the authors proposed the possibility of using a single RGB camera for rehabilitation therapy assessment systems. However, this system only utilizes 2D

* Yong HU is the corresponding author.

human skeleton coordinates, which has limited the effectiveness of rehabilitation movement assessment [6]. In Reference [7], the authors proposed the possibility of using a single RGB camera to obtain 3D skeletal data of the human body. However, since a single RGB camera cannot fully reproduce the real 3D environment through 3D reconstruction method(resulting in multiple computational results), the authors assume the acquisition of 3D skeleton keypoints have been acquired but do not provide specific technical details. In Reference [2], the authors have used AI model to estimate 3D skeleton keypoints from 2D skeleton of the human body, but this study only provides a pose estimation solution and does not specifically apply it to movement assessment.

Based on the above circumstances, this study proposes the SC-AMAS system (Single RGB-camera based automatic movement assessment system) to address the following issues: 1) Using the GHUM 3D Pose Landmark Model to estimate 3D keypoints in movement videos captured by a single RGB camera, thus reducing the need of complicated devices. 2) Utilizing a rule-based scoring system that incorporates domain knowledge from the medical field, providing not only traditional scoring but also corrective advice. 3) Additionally, the system employing video processing techniques to the original video, which increases patient privacy protection when external individuals who assess the video scores manually.

II. Method

The overall workflow of this system consists of four parts: video data upload, motion feature extraction, movement performance scoring with feedback, and privacy protection for the video stream. Fig1 has shown the core workflow of the system. The following parts would provide a detailed description about the functionality and principles of this system.

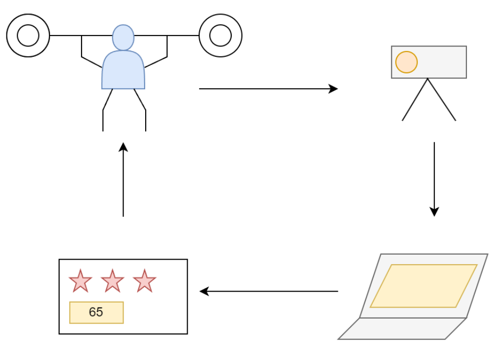

Fig. 1. The basic workflow of movement assessment system.

A. System Environment

This system is designed to be used on the Windows operating system and primarily relies on software packages including Python 3.7, Mediapipe Python version, and OpenPose Python version. The hardware environment includes a 720p RGB camera or a smart phone to record 4K video (with video resolution reduced to 1920x1080 for furtering analyzing), Intel i7 8th generation processor, and 16GB of RAM. Additionally, we do not utilize a GPU for video feature extraction, this selection try to make sure that the system could be deployed on lightweight hardware platforms in further work.

B. Video Data record & Upload

This function includes video recording and saving, as well as uploading existing videos. The function is supported by the open-source OpenPose package to drive a single RGB camera and provide other video stream processing methods such as resolution reduction. Users can freely record videos or upload videos by clicking the start button. The recorded videos are saved in the default folder but are not automatically uploaded for analysis, giving users the necessary authority to decide which videos to upload.

C. Motion Feature Extraction

1. Pose Detection Model: The pose detection model is used to convert motion video data into human skeleton data, where the common form of human skeletal data is in 2D [8]. In this study, the BlazePose algorithm based on Mediapipe is used to extract 2D human skeleton motion data. BlazePose is a bottom-up pose detection algorithm that proposed by Google. In its current release versions,it contains three branches: lite version, full version, and heavy version. The lite version is suitable for ultra-lightweight devices, while the full version is designed for devices with medium computing capabilities, and the heavy version is for devices with abundant computational resources, respectively.

2. 3D Human Skeletal Data Extraction: Generally, current pose detection algorithm provides 2D human skeletal data from video stream, but predicting real-world human movement requires using 3D skeleton data. In this study, the GHUM model is used for realize this function [20]. GHUM is a statistical model that can predict estimated 3D skeletal data based on 2D human skeleton inputs. The training dataset of GHUM is based on a large human body corpus, which is developed by Google.

D. scoring and feed back

These two parts involves analyzing the results of motion feature extraction and presenting visual feedback to the user. Typically, the input for movement scoring is the human skeleton keypoints model (a common result of motion feature extraction), and the common scoring methods include rule-based methods and template-based methods [5] .

Rule-based methods are the most common type of scoring method, where rules are formulated by experts and translated into measurable features. For example, most squat depth can be formulated as maximum knee bending angle and arm extension can be formulated as elbow bending angle. So this method provide movements corrective feedback more explicitly. On the other hand, template-based methods use predefined standard movements and their corresponding skeleton keypoints as labels, and the user's movement score is calculated by measuring the distance to the labels. This method have considered the subject's whole body pose in movement

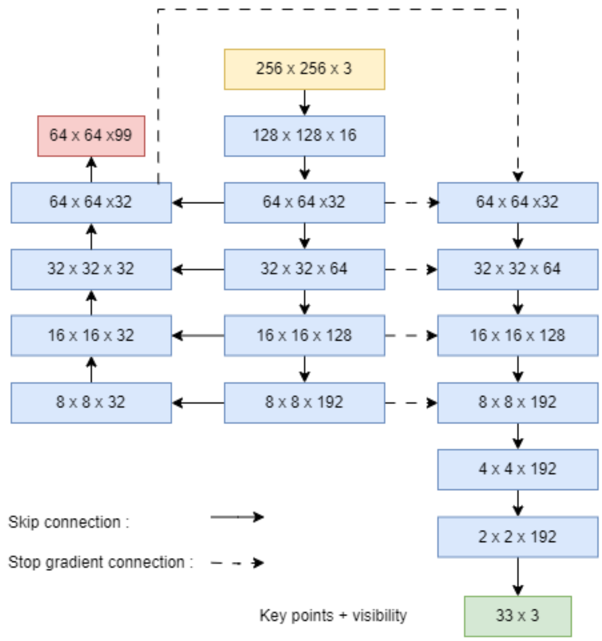

Fig. 2. Architecture of BlazePose. [8]

scoring, such as balance and coordination of different body parts, so its result in a model that can better fits human expert scoring behavior. In this study, a rule-based method is chosen based on its a successful application on KIMORE dataset [5]. The reasons of score deductions can be explained by specific rule, which can meet the strict requirements of movement therapy standards in practical environment and provide movement correction suggestions to patients.

As for the specific rule definition, this method defined the movement description by two parts. The first part is called Primary Objective(PO) , which is a kind of objective that subject try to achieve, for example the arm extension when doing lifting. The second part is called Control Factor(CF), which represents the rules that the subject should maintain their posture during the movement implementation process, for example,when subject doing the lifting, the requirement about keep the subject's feet don't move. The score calculation was demonstrated in below:

$$FS = W_{PO} * POS + W_{CF} * CFS \tag{1}$$

$$POS = W_1 * PO_{r1} + W_2 * PO_{r2} + ... + W_i * PO_{ri} \tag{2}$$

$$CFS = W_1 * CF_{r1} + W_2 * CF_{r2} + ... + W_i * CF_{ri} \tag{3}$$

In this scoring method, the movement assessment score is named as Final Score(FS), and it is calculated by the weight sum of scores in PO and CF part. In PO or CF part,a basic score is given to each rules. correspondingly, then the total score in PO(POS) or CF(CFS) part is calculated the weight sum of each rules. Noted that all the weights(e.g. W_1) and rules(e.g. CF_{r1}) are predefined by rehabilitation exercise expert.

TABLE I
RULES EXAMPLE OF ONE MOVEMENT

Rule ID	***Rule description***	***measurable features***
PO1	The maximum angle of knees should be less than 90 degrees	double knees' maximum angle
CF1	keep the feet has same distance with shoulder	knees' distance shoulder distance
CF2	keep the forearm parallel	elbows' distance wrists' distance

E. Privacy feature protection

Generally speaking, the privacy features in motion videos usually are patients' body information and environmental information, some of which are related to the movement assessment process and we call relevant information in this paper. The potential relevant information including body shape, limb positions, etc. Hiding these relevant information may affect human experts' judgment when watching in the videos, making the movement assessment process impractical. On the other hand, the information that unrelated with movement assessment include clothing, home environment, etc, these kind of information should be protected. However, the unrelated information cannot be completely separated from the relevant information in the video. We employed 2 color space transformation solutions to systematically replace the original video colors, thus protecting the patient's information. This method has illustrates in below:

$$\begin{aligned} &Image\ channel(R) -> Image\ channel(B) \\ &Image\ channel(G) -> Image\ channel(G) \\ &Image\ channel(B) -> Image\ channel(R) \end{aligned} \tag{4}$$

or

$$Color\ space(RGB) -> Color\ space(HSV) \tag{5}$$

This method can easily the preserve the related information while protecting patient privacy, the two color transformation methods can let user choose the method that can fit the different color environments.

III. RESULT

From the perspective of AI application, whether a system is successful depends on its user experience. However, since there're very limited number of users with experience using different systems, we compare similar products based on the following aspects we have selected: 1. Scene factors related to system usage. 2. System features: In this part, the system's own features include pose detection accuracy, movement scoring definition, and privacy protection. In addition, although there are no works in the existing literature that can be directly compared with our system, we have selected works that can achieve similar goals.

A. Scene factors related to system usage

In this part, We comparing the following perspective of each proposed systems: ease of deployment(system collaboration requirement), professional requirements for operators, potential price, potential time. Noted that the estimated data was based on common experience.

TABLE II
SYSTEM COMPARISON

System name	SC-AMAS	[16]	[11]
collaboration requirement free	✓	✓	✓
professional operator requirement free	✓	✓	
estimated deployment time(mins)	**10**	30	20
estimated hardware price($)	**1000**	3000	8000

B. System features

- Pose Detection speed & accuracy: Pose detection accuracy and speed can directly affect the user experience, for our selected method, the corresponding result was tested and was originally demonstrated in [8], includes the FPS speed and accuracy on two public dataset. Noted that FPS result can just represent the detection speed for specific video on specific hardware, and it subject to change based on different runtime environment and video resolution.
- Movement Scoring feedback: One of the key feature of this system is to provide visible feedback to patients, and the text was selected as visible and was built by the corresponding rules that predefined by sports expert. In this system, We have selected one movement of public exercise as an example, the corresponding rule definition has shown in Tab. 1. Since we used the rules with the form applied on the KIMORE dataset, the rules has been listed a special form in the table, noted that to increase the number of movement rule database, further research may need professional advice from sports expert.
- Privacy Protection: Since Privacy Protection is trying to reducing the unrelated information when expert watching the video, to obtain the feedback of of this method as practicable as possible, in this part, we simulate the real usage environment to let the user acquire enough experience for our systems. Specifically, an example video for exercise have been recorded as a experimental data. The video before and after processing have shown in Fig.3, the part A is the original recorded video part B is the original pose estimation result. In part C and part D, we have used first and second used color space transition method respectively.

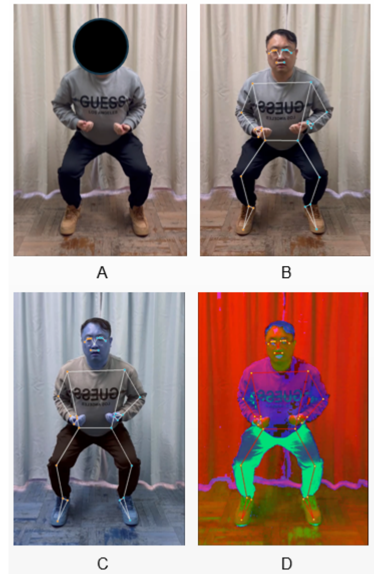

Fig. 3. privacy protection result

IV. DISCUSSION

This section discusses the potential improvement directions for the system from the perspectives of scene factors and the system's own function based on the aforementioned results. We provide three potential directions that we believe will be contributed to the development of remote movement assessment systems.

1. Lightweight platform deployment: Although the system can achieve remote movement assessment with lower system resources, migrating to lightweight platforms can still improve cost-effectiveness. For example, common lightweight hardware platforms include mobile devices and lightweight computers (such as Raspberry Pi). However, the current system runs on the Windows 10 operating system, which cannot be deployed on lightweight platforms. Therefore, there is still space to reduce the system's development cost by adding platform selection and migration.

2. movement scoring & feedback: Currently, the system applied a rule-based method. Although this method can provide accurate action correction feedback, it heavily relies on the summarization and transformation of domain knowledge by professional sports medicine expert. Since rules are difficult to describe certain motion characteristics, such as the overall coordination of the body, this method lacks the flexibility comparing with template-based methods. A more reasonable approach would be to combine the two methods to leverage the advantages of different approaches.

3. Privacy protection: Currently, the system uses color transformation method to protect the privacy of users in the videos by disturbing the video color. Although this method saves computational resources, it still can't fully protected information that unrelated to movement assessment, such as the user's face appearance. Potential solutions include using human body part recognition technology to detect and conceal the user's unrelated body while avoiding mask the body parts that may affect the assessment process.

V. Conclusion

This paper presents a motion assessment system based on a monocular RGB camera, which utilizes advanced computer vision techniques to obtain estimated 3D coordinates of the human body. This design significantly reduces the deployment costs for users, including space and equipment factors, and this designed has contributed to the development of remote motion assessment research. By extracting domain knowledge in the field of motion, this paper establishes rule-based evaluations for therapeutic movement, which may be applicable in rehabilitation medicine-related movements, and providing potential directions for the application of AI methods in this field. Regarding the unresolved issues and potential improvement directions of this system, we believe they are equally applicable to similar remote assessment scenario, and we hope our work can promote the development of related research.

References

[1] Raza, A., Qadri, A. M., Akhtar, I., Samee, N. A., & Alabdulhafith, M. (2023). LogRF: An approach to human pose estimation using skeleton landmarks for physiotherapy fitness exercise correction. IEEE Access.

[2] Shah, Milind, et al. "Computer Vision & Deep Learning based Realtime and Pre-Recorded Human Pose Estimation." 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC). IEEE, 2023.

[3] Garg, B., Postlmayr, A., Cosman, P., & Dey, S. (2023, June). Short: Deep Learning Approach to Skeletal Performance Evaluation of Physical Therapy Exercises. In Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies (pp. 168-172).

[4] Bahukhandi, U., & Gupta, S. (2021). Yoga pose detection and classification using machine learning techniques. Int Res J Mod Eng Technol Sci, 3(12), 13-15.

[5] Capecci, M., Ceravolo, M. G., Ferracuti, F., Iarlori, S., Monteriu, A., Romeo, L., & Verdini, F. (2019). The kimore dataset: Kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27(7), 1436-1448.

[6] Guo, Q., & Khan, S. S. (2021). Exercise-specific feature extraction approach for assessing physical rehabilitation. In 4th IJCAI Workshop on AI for Aging, Rehabilitation and Intelligent Assisted Living. IJCAI.

[7] Garg, B., Postlmayr, A., Cosman, P., & Dey, S. (2023, June). Short: Deep Learning Approach to Skeletal Performance Evaluation of Physical Therapy Exercises. In Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies (pp. 168-172).

[8] Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020). Blazepose: On-device real-time body pose tracking. arXiv preprint arXiv:2006.10204.

[9] Mroz, S., Baddour, N., McGuirk, C., Juneau, P., Tu, A., Cheung, K., & Lemaire, E. (2021, December). Comparing the quality of human pose estimation with blazepose or openpose. In 2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART) (pp. 1-4). IEEE.

[10] Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7291-7299).

[11] Milanko, S., & Jain, S. (2020). LiftRight: Quantifying strength training performance using a wearable sensor. Smart Health, 16, 100115.

[12] Gu, C., Lin, W., He, X., Zhang, L., & Zhang, M. (2023). IMU-based motion capture system for rehabilitation applications: A systematic review. Biomimetic Intelligence and Robotics, 3(2), 100097.

[13] Du, C., Graham, S., Depp, C., & Nguyen, T. (2021, November). Assessing physical rehabilitation exercises using graph convolutional network with self-supervised regularization. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 281-285). IEEE.

[14] Bell, K. M., Onyeukwu, C., McClincy, M. P., Allen, M., Bechard, L., Mukherjee, A., ... & Irrgang, J. J. (2019). Verification of a portable motion tracking system for remote management of physical rehabilitation of the knee. Sensors, 19(5), 1021.

[15] Webster, D., & Celik, O. (2014, February). Experimental evaluation of Microsoft Kinect's accuracy and capture rate for stroke rehabilitation applications. In 2014 IEEE Haptics Symposium (HAPTICS) (pp. 455-460). IEEE.

[16] Liao, Y., Vakanski, A., & Xian, M. (2020). A deep learning framework for assessing physical rehabilitation exercises. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(2), 468-477.

[17] Mousavi Hondori, H., & Khademi, M. (2014). A review on technical and clinical impact of microsoft kinect on physical therapy and rehabilitation. Journal of medical engineering, 2014.

[18] Da Gama, A., Fallavollita, P., Teichrieb, V., & Navab, N. (2015). Motor rehabilitation using Kinect: a systematic review. Games for health journal, 4(2), 123-135.

[19] Webster, D., & Celik, O. (2014). Systematic review of Kinect applications in elderly care and stroke rehabilitation. Journal of neuroengineering and rehabilitation, 11, 1-24.

[20] Xu, H., Bazavan, E. G., Zanfir, A., Freeman, W. T., Sukthankar, R., & Sminchisescu, C. (2020). Ghum & ghuml: Generative 3d human shape and articulated pose models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6184-6193).