

AuraPose: Accurate Human Pose Detection and Behavior Recognition via Enhanced OpenPose with Angular Measurement

Haichen Liu^{a,†}, Jianqiang Mei^{a,*†}, Fan Jia^b, Dandan Zheng^c, Jiamin Yuan^{d,a}, Zheling Wang^{e,a}, Yiran Zhao^{f,a}

^aSchool of Electronic Engineering, Tianjin University of Technology and Education, Tianjin, China

^bRaysov Instrument Co. Ltd., Dandong, China

^cSchool of Electrical and Information Engineering, Tianjin University, Tianjin, China

^dDepartment of Electronic and Electrical Engineering, University of Sheffield, Sheffield, United Kingdom

^eSchool of Engineering, University of New South Wales, Sydney, Australia

^fSchool of Engineering, University of Birmingham, Birmingham, United Kingdom

E-mail: meijianqiang@tute.edu.cn

Abstract—The rapid progress in computer vision and deep-learning technology has positioned human posture detection and estimation as a hot spot research area with sustained development. Nevertheless, prevalent detection technologies frequently encounter issues related to accuracy and robustness which hinders their further practical adoption. In this paper, an enhanced OpenPose algorithm incorporates angular measurement is proposed to tackle above challenges. Firstly, OpenPose is utilized to detect and locate key points of human body. Subsequently, the connections between these points according to the specified posture are optimization. Finally, ultimate angles are measured to deliver the accurate pose detection and behavior recognition. Qualitative analysis on recorded videos illustrate the applicability and robustness of proposed method. Furthermore, quantitative results on COCO2017 dataset compared with the state-of-the-art methods also show competitive performance of proposed method, in which our AR, AP and mAP reached 89.2%, 79.8% and 81.8%, respectively. Through meticulous optimization, our method has the potential to further enhance vision-based human posture detection and measurement technologies, particularly within the domains of medical, biomedical and healthcare systems.

Index Terms—Pose detection, behavior recognition, enhanced OpenPose, angular measurement

I. INTRODUCTION

The enhancement of living standards has propelled heightened awareness of physical fitness. However, individuals often find it challenging to accurately adhere to exercise guidance, leading unsatisfactory workout outcomes and increased risk of physical injury. Consequently, human pose detection and behavior recognition have garnered increasing attention.

Some methods utilize procedural models to optimize motion data with time markers [1] or perform comprehensive analysis of human motion through specific devices combined with data fusion algorithms [2], [3]. However, they all face several challenges, including high model complexity, poor stability, and difficulties in controlling accuracy. [4]–[6].

*Corresponding author, †Equal contribution.

With the advancements in artificial intelligence, the OpenPose framework, which utilizes deep learning technology to enable cost-effective human pose estimation, has been introduced. Extensive validation confirmed that it effectively offers a dependable approach for human pose estimation. Nevertheless, complex scenes where occlusion or contact between human bodies may lead OpenPose to inaccurate motion analysis and difficult key points tracking issue.

In order to address aforementioned challenges, a human pose detection and behavior recognition method is proposed in this paper. Initially, the OpenPose is utilized to identify key points while the connections between them are subsequently optimized based on specified postures. Finally, ultimate angles are measured to provide precise pose detection and behavior recognition.

Layout of the paper is given below. Relative literature is previewed in Section II while proposed method is exhaustively outlined in Section III. Qualitative and quantitative comparisons are conducted in Section IV. The paper summarized in Section V.

II. RELATED WORK

In the field of human pose detection and behavior recognition, classical methods usually rely on computer vision and machine learning. A non-parametric representation, called Part Affinity Fields (PAFs), was proposed by Cao et al. [7]. This method is not affected by the number of people and has excellent real-time accuracy. A single network for pose estimation in natural scenes was also proposed by Hidalgo et al. [8] to solve the scale difference between different key points. Furthermore, Samet et al. [9] invented a novel single-stage pose estimation method, called Hierarchical Point Regression (HPRNet), which achieved the most advanced results on COCO dataset. In addition, Boswell et al. [10] introduced a self-guided five-repeat sit-to-stand testing for extensive quantitative motion analysis. Huang et al. [11] proposed

a system that effectively detects compensatory exercise to evaluate the performance of rehabilitation exercises.

With the rapid development of deep learning technology, human pose estimation based on deep neural network has gradually become a research hot-spot. Particularly, Kreiss et al. [12] proposed a new bottom-up multi-person 2D human pose estimation method suitable for urban mobility, in which Part Intensity Field (PIF) was utilized to locate body parts while PAFs was employed for association. Cao et al. [13] also unveiled a new convolutional neural network architecture in which the key points of the human body were estimated through a dual attention mechanism and multi-scale feature fusion. As the proposal of OpenPose, a multi-person pose estimation framework based on deep learning, human pose detection in complex scenes has become feasible. Shen et al. [14] explored a learning assistant system combining OpenPose and fuzzy technology, in which OpenPose is used to extract the human body skeleton. Furthermore, Kim et al. [15] proposed a new method for measuring joint angles and evaluating RULA / REBA ergonomic postures based on three-dimensional human skeleton data obtained by OpenPose. Saiki et al. [16] proposed the hip-knee-ankle (HKA) angle which was measured by OpenPose.

III. METHOD

A. Overview of proposed method

The ultimate goal of our method is to explore accurate human pose detection and behavior recognition from real scenarios. To achieve this, a novel method composed of OpenPose algorithm with angular measurement named AuraPose is proposed in this paper. As depicted in Fig.1, OpenPose is employed to detect and extract the key points of human body from the preprocessed video images. Subsequently, the connections between these points are optimized according to specified posture. Then, the corresponding angle is measured and compared with the standard posture's. We next detail the OpenPose network and angular measurement functions of proposed method in following sections.

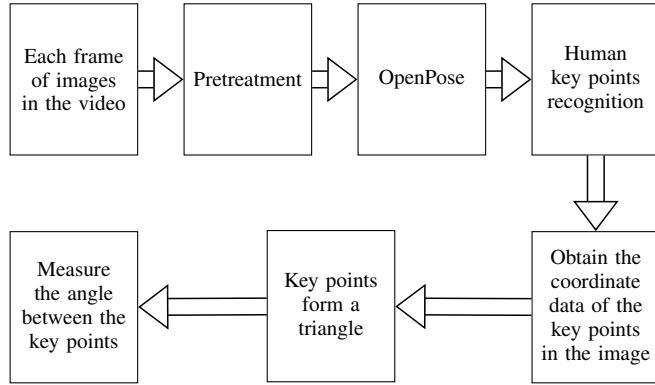


Fig. 1. Flow chart of proposed method.

B. OpenPose network

The OpenPose attitude model with the specific locations of 18 key points are illustrated in Fig.2.

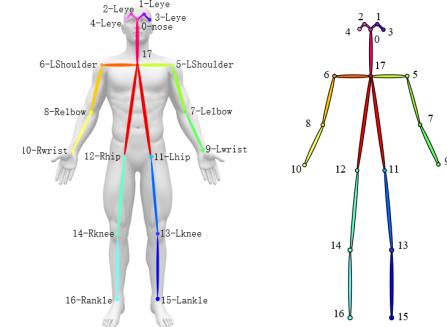


Fig. 2. OpenPose attitude mode with 18 key points (from 0 to 17).

As depicted in Fig.3, the feature maps F , which derived from the 2D image by VGG-19 model, are passed into two branches of Stage 1. Explicitly, branch 1 stands for the Part Confidence Maps (PCMs), in which ρ^1 represents the convolution process predicts the key points. Meanwhile, branch 2 denotes the PAFs, where Φ^1 indicates the predictive convolution. Both branches contain three 3×3 and two 1×1 2D convolutions. Five 7×7 2D convolutions are utilized in enhancement phase to increase the visual perception of image. In addition, at the end of each $t \in \{1, 2, 3, \dots, T\}$ level cascades, two tributaries are finally merged with F together.

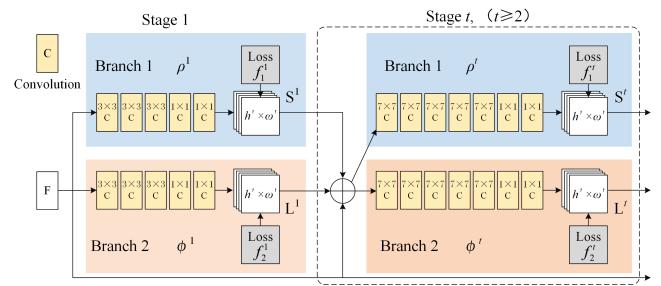


Fig. 3. Schematic structure of OpenPose network [7].

During the training process, relaxation supervision is performed. The calculation of S^t and L^t is shown as follow:

$$S^t = \rho^t(F, S^{t-1}, L^{t-1}), \forall t \geq 2 \quad (1)$$

$$L^t = \Phi^t(F, S^{t-1}, L^{t-1}), \forall t \geq 2 \quad (2)$$

Moreover, to alleviate the gradient vanishing effect for deeper network layers during iterations, following loss function is employed at the end of each stage:

$$f = \sum_{t=1}^T (f_S^t + f_L^t) \quad (3)$$

in which f_S^t and f_L^t are further described as:

$$f_S^{t_k} = \sum_{j=1}^J \sum_P W(P) \cdot \|S_j^{t_k}(P) - S_j^*(P)\|_2^2 \quad (4)$$

$$f_L^{t_i} = \sum_{c=1}^C \sum_P W(P) \cdot \|L_c^{t_i}(P) - L_c^*(P)\|_2^2 \quad (5)$$

As shown in Equation (4) and (5), t_k represents the t_k -th stage of the predicted PCMs while t_i represents the t_i -th stage of the predicted PAFs. $S_j^*(P)$ and $L_c^*(P)$ are the ground truth of PCMs and PAFs, respectively. Furthermore, W is a binary mask. $W(P) = 0$ represents the position P is not annotated. The existence of $W(P) = 1$ can solve the problem of additional loss caused by data loss [7].

C. Action angle measurement

Aforementioned, the angles between the joint points for specified postures are measured and compared with standard ground truths to deliver the accurate pose detection and behavior recognition of complex scenes. It should be noted that the angular measurement is invariant to body shape, height, skin color, age and other characteristics. Our angular measurement scheme is depicted in Fig.4 while the measurement algorithm is present in Algorithm 1.

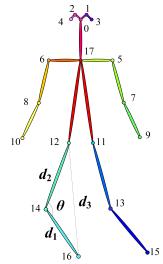


Fig. 4. Angle measurement schematic.

Algorithm 1: Human Action Angle Measurement

Require: Three 2D key points of selected human body: p_0 , p_1 and p_2 . Three sides of the triangle: d_1 , d_2 and d_3 .

- 1 **if** $(p_1[0] = 0) \vee (p_1[1] = 0) \vee (p_0[0] = 0) \vee (p_0[1] = 0) \vee (p_2[0] = 0) \vee (p_2[1] = 0)$ **then**
- 2 **return** -2.0 // Key point missing or abnormal
- 3 **if** $a \cdot b = 0$ **then**
- 4 **return** -1.0 // Key points are collinear, no valid triangle
- 5 $d_1 \leftarrow \sqrt{(p_1[0] - p_0[0])^2 + (p_1[1] - p_0[1])^2}$
- 6 $d_2 \leftarrow \sqrt{(p_1[0] - p_2[0])^2 + (p_1[1] - p_2[1])^2}$
- 7 $d_3 \leftarrow \sqrt{(p_2[0] - p_0[0])^2 + (p_2[1] - p_0[1])^2}$
- 8 $\theta \leftarrow \arccos \frac{d_1^2 + d_2^2 - d_3^2}{2d_1d_2}$

As shown in Algorithm 1, the Euclidean distances between p_0 , p_1 and p_2 are calculated as d_1 , d_2 and d_3 (line 5 to 7 in Algorithm 1) after dealing with some invalid situation, such as key point missing, abnormal, collinear and so on. The value of the specified angle is obtained at line 8 in Algorithm 1.

IV. RESULTS AND DISCUSSION

A. Experimental environment and comparison criterion

All experiments of proposed method leverage the TensorFlow deep learning framework on Windows 10 with an Intel Core i5-8250U and 8GB memory. It should be noted that we trained our method on a NVIDIA GeForce MX150 using random samples of COCO2017 dataset. Some samples used in the experiment are shown in Fig.5. The human key points extraction results are shown in Fig.6.



Fig. 5. Image samples of the COCO2017 dataset utilized in our experiment.



Fig. 6. Extraction results on samples from COCO2017 dataset by OpenPose.

For fair comparison, following similarity measurements were calculated for quality evaluation:

- 1) **AP** (Average Precision). For the bottom-up method of multi-person pose estimation, the calculation of **AP** is shown as following:

$$AP = \frac{\sum_p \delta(OKS_p > T)}{\sum_p 1} \quad (6)$$

in which **OKS** (Object Key Point Similarity) is the similarity of joint points. The calculation of **OKS** is shown as following:

$$OKS = \frac{\sum_i \left[\exp\left(-\frac{d_i^2}{2s^2 k_i^2}\right) \delta(\nu_i > 0) \right]}{\sum_i [\delta(\nu_i > 0)]} \quad (7)$$

- 2) **mAP** (Mean Average Precision). The **mAP** is the mean of the average accuracy. By setting different **OKS** thresholds T , different **AP** are obtained, and then the average of all **AP** is taken. The value range of **OKS** threshold T is S , and **mAP** is the mean value of

AP obtained under different thresholds. The calculation of mAP is shown as following:

$$mAP = \text{mean}\{\text{AP}@S\} \quad (8)$$

In addition, there are a series of OKS -based performance indicators are widely utilized, namely AP^{50} , AP^{75} , AP^M , AP^L and average recall (AR). In general, AP^{50} and AP^{75} represent the accuracy of more than 50% and 75% similarity to joint points while AP^M and AP^L represent the accuracy of identifying small-scale and large-scale populations, respectively. Furthermore, AR represents the average value of the proportion of the size of predicted correct part to the size of annotation result.

B. Model training

First of all, we pre-trained the model following the method in COCO dataset [17], which serves as the initial weight for our training. Then, 80% of the COCO2017 dataset were used for training while the remaining were employed for evaluating. In particular, the model iterates more than 39,000 times for training. The learning rate is gradually reduced from $4e^{-4}$ to 0 while the value of batch size is fixed to 1.

As shown in Fig.7, the purple solid curve represents the PAFs loss with respect to the iterations during the training process. Meanwhile, the orange dots curve indicates the key point heat map loss. The two loss curves are both rapidly decreasing and tending towards stability as the iteration increased. It should be noted that the final PAFs loss is lower than the key point heat map loss.

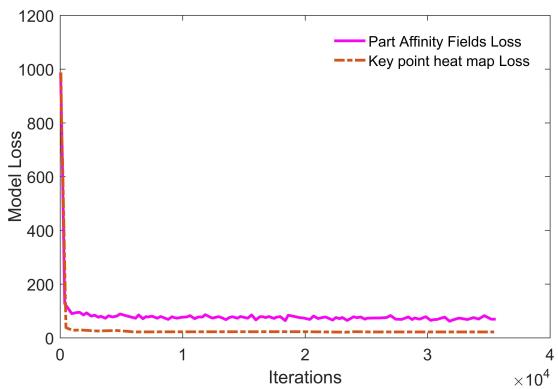


Fig. 7. Changing trend of two losses at different training iterations.

C. Key points detection and angular measurement

In this section, as shown in Fig.8, the key points detection and angles measurement results through proposed method are presented for corresponding high knee and calf stretching actions in real scenes. In particular, the detection and measurement results of the left and right high knees are presented in Fig.8(a) and 8(b), respectively. Moreover, from left to right within Fig.8(a) and 8(b), detection results of lifting the left and right legs are also shown, respectively. The yellow color

texts represent the detection results of key points and the key points are connected by lines. The red color texts represent the right leg angles while the green color texts represent the detection results of the left leg angles. Meanwhile, as depicted within Fig.8(c), the different leg angles of calf stretching can be accurately measured.

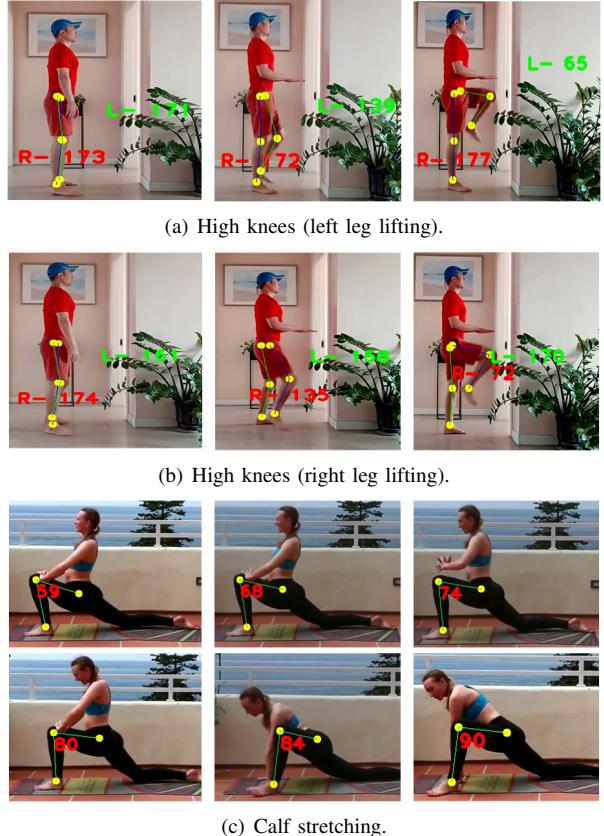


Fig. 8. Key point detection and angular measurement results.

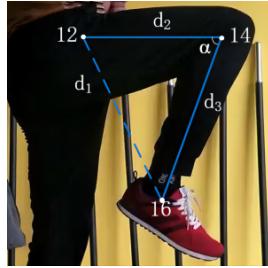
D. Behavior recognition analysis

1) *High knees*: The main reasons why many people struggle to perform high knees with standard form is their inability to maintain a 90° angle between the thigh and calf during high-frequency leg movements. Following this issue, we can take the right hip, right knee, and right ankle as key points, labeling them as points 12, 14, and 16 respectively. These points form a triangle with vertex 14 and side lengths d_1 , d_2 , and d_3 (see Fig.9(b)). The value of angle α is measured on-the-fly and compared with the 90° . The difference between angle α and angle of 90° can be utilized as the condition to determine whether the behavior is standard. By quantitatively comparing the angle α with the standard 90° angle.

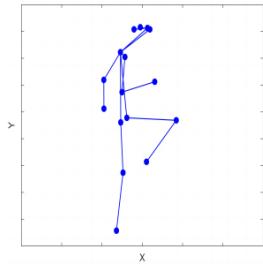
Furthermore, Fig.9(c) presents a schematic diagram illustrating the outcomes of key points extraction. As shown in Fig.9(d), it is the angular changing of angle α in each frame. Each trough in the waveform represents the completion of an action. The more condensed the waveform, the higher



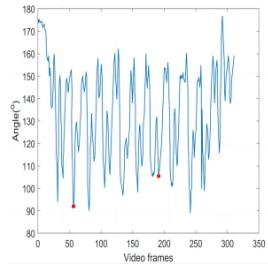
(a) Extraction effect.



(b) Zoom in effect.



(c) Extract key points.



(d) Changing trend of α .

Fig. 9. Demonstration of high knees situation.

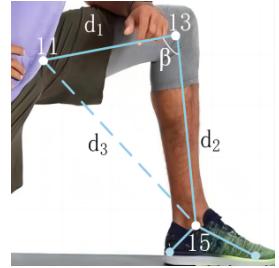
the user's high leg running frequency. Two trough points are randomly selected for analysis. As shown in Fig.9(d), the first red spot from left to right represents the completion of the fourth high leg lifting movement. At this point, the frame number of the input video is the 56-th frame. The value of angle α under this frame is 91.84° . Compared with the standard angle 90° , the angular difference is small which indicates that the action is considered to be standard. The second red spot from left to right corresponds to the 15-th high knees lifting. At this point, the frame number of the input video is the 191-th frame, and the value of angle α is 105.48° . Compared with the standard angle 90° , the large angular difference indicates the action is not standard.

2) *Calf stretching*: The specific action of the calf stretching is to separate the legs into lunges before and after, the left leg is in front, the right knee is on the ground, the calf is attached to the ground and the backside is moved forward, maintaining 10-15 seconds or uninterrupted repetition. The difficulty of this behavior is that the angle formed between the thigh and the calf reaches 90° . As shown in Fig.10(a), the calf stretching demonstration is arbitrarily selected. The key points corresponding to the left hip, left knee and left ankle are key points 11,13 and 15 respectively, forming a triangle with 13 as the vertex and the side lengths of d_1 , d_2 and d_3 (see Fig.10(b)). The angle β of the vertex 13 in the calf stretching is measured in real time, and the difference between the standard angle 90° is compared which can be used as a condition to determine whether the action is standard.

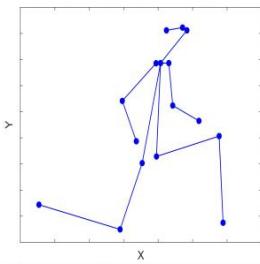
Fig.10(c) illustrates a schematic diagram presenting the results of all extracted key points. As shown in Fig.10(d), the degree change of angle β in the number of frames of the input user's calf stretching action video is analyzed. Each trough in



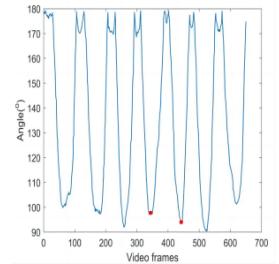
(a) Extraction effect.



(b) Zoom in effect.



(c) Extract key points.



(d) Changing trend of β .

Fig. 10. Demonstration of calf stretching situation.

the waveform indicates the completion of an action. We select the fourth and fifth trough point. The first red spot from left to right corresponds to the fourth trough point, representing the completion of the fourth set of actions and corresponding to the 342-th frame of the input video. In this frame, the angle of angle β is 97.38° . Compared with the standard angle of 90° , this angle is significantly different, suggesting that the action needs to be corrected as it is not up to the standard. The second red point from left to right corresponds to the fifth trough point, indicating the completion of the fifth set of actions and corresponding to the 445-th frame of the input video. The angle of angle β in this frame is 93.91° . Compared with the standard angle 90° , this angle is not significantly different, suggesting standard action.

E. Discussion

Detailed quantitative comparison with the state-of-the-art methods based on COCO2017 dataset are presented in Table I, in which the red and blue color represent the optimal and second ranked results, respectively. In addition, the performance metrics of proposed method are detailed as follows: AR attains 89.2%, AP achieves 79.8%, AP^{50} attains 90.1%, AP^{75} reaches 83.5%, AP^M achieves 73.1%, AP^L reaches 81.2% and mAP attains 81.8%.

It should be pointed out that although proposed method cannot achieve optimal results on all metrics, it still shows competitive advantages in specified situation. To this end, it is clear that proposed method outperforms than the other state-of-the-art methods in AP, AP^{75} , AP^M and mAP metrics, which validates an competitive pose detection and behavior recognition results.

TABLE I

QUANTITATIVE COMPARISON BETWEEN PROPOSED METHOD AND OTHER STATE-OF-THE-ART METHODS BASED ON THE COCO2017 DATASET.

Model	AR(%)	AP(%)	AP ⁵⁰ (%)	AP ⁷⁵ (%)	AP ^M (%)	AP ^L (%)	mAP(%)
HRNet-W32 [18]	78.9	73.4	89.5	80.7	70.2	80.1	—
HRNet-W48 [18]	90.4	75.1	90.6	82.2	71.5	81.8	—
Small HRNet [19]	61.5	55.2	85.8	61.4	51.7	61.2	—
Dite-HRNet [20]	72.1	65.9	87.3	74.0	63.2	71.6	—
AlphaPose [21]	—	—	89.2	79.1	68.0	78.6	72.3
Proposed	89.2	79.8	90.1	83.5	73.1	81.2	81.8

V. CONCLUSIONS

A method named AuraPose is proposed in the paper for accurate human pose detection and behavior recognition in complex scenarios. Specifically, improved OpenPose is utilized to detect and locate the key points of human body while angles of joint points are measured to deliver specified posture. Qualitative and quantitative comparisons against the state-of-the-art methods demonstrate applicability and robustness of our method, which improves AP by 4.7%, AP⁷⁵ by 1.3%, AP^M by 1.6%, and mAP by 9.5%. Furthermore, AuraPose emerges as an effective tool for fitness enthusiasts, coaches, and medical professionals to monitor the postures and movements of themselves, their trainees, and patients.

ACKNOWLEDGEMENTS

This work is supported by Science and Technology Planning Project of Liaoning Province(Grant No.1654770952010), Teaching Reform and Quality Construction Research Project of Tianjin University of Technology and Education(Grant No.JGY2022-03).

REFERENCES

- [1] Z. Liu, L. Zhou, H. Leung, and H. P. Shum, “Kinect posture reconstruction based on a local mixture of gaussian process models,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 11, pp. 2437–2450, 2016.
- [2] V. Moreno, B. Curto, J. A. García-Estebar, F. Hernández Zaballos, P. Alonso Hernández, and F. J. Serrano, “Husp: A smart haptic probe for reliable training in musculoskeletal evaluation using motion sensors,” *Sensors*, vol. 19, no. 1, 2019. [Online]. Available: <https://www.mdpi.com/1424-8220/19/1/101>
- [3] N. Valencia-Jiménez, A. Leal-Junior, L. Avellar, L. Vargas-Valencia, P. Caicedo-Rodríguez, A. A. Ramírez-Duque, M. Lyra, C. Marques, T. Bastos, and A. Frizera, “A comparative study of markerless systems based on color-depth cameras, polymer optical fiber curvature sensors, and inertial measurement units: Towards increasing the accuracy in joint angle estimation,” *Electronics*, vol. 8, no. 2, 2019. [Online]. Available: <https://www.mdpi.com/2079-9292/8/2/173>
- [4] L. Xu, Y. Liu, W. Cheng, K. Guo, G. Zhou, Q. Dai, and L. Fang, “Flycap: Markerless motion capture using multiple autonomous flying cameras,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 8, pp. 2284–2297, 2018.
- [5] C.-H. Chen and D. Ramanan, “3d human pose estimation = 2d pose estimation + matching,” 2017.
- [6] W.-N. Lie, G.-H. Lin, L.-S. Shih, Y. Hsu, T. H. Nguyen, and Q. Nguyen Quang Nhu, “Fully convolutional network for 3d human skeleton estimation from a single view for action analysis,” in *2019 IEEE International Conference on Multimedia , Expo Workshops (ICMEW)*, 2019, pp. 1–6.
- [7] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, “Openpose: Realtime multi-person 2d pose estimation using part affinity fields,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2021.
- [8] G. Hidalgo, Y. Raaj, H. Idrees, D. Xiang, H. Joo, T. Simon, and Y. Sheikh, “Single-network whole-body pose estimation,” 2019.
- [9] N. Samei and E. Akbas, “Hprnet: Hierarchical point regression for whole-body human pose estimation,” *Image and Vision Computing*, vol. 115, p. 104285, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0262885621001906>
- [10] M. A. Boswell, Kidziński, J. L. Hicks, S. D. Uhlrich, A. Falisse, and S. L. Delp, “Smartphone videos of the sit-to-stand test predict osteoarthritis and health outcomes in a nationwide study,” *NPJ digital medicine*, vol. 6, no. 1, p. 32, March 2023. [Online]. Available: <https://europapmc.org/articles/PMC9985590>
- [11] C.-H. Huang, C.-F. Lin, C.-A. Chen, C.-H. Hwang, and D.-C. Huang, “Real-time rehabilitation exercise performance evaluation system using deep learning and thermal image,” in *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2020, pp. 1–6.
- [12] S. Kreiss, L. Bertoni, and A. Alahi, “Pifpaf: Composite fields for human pose estimation,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 11969–11978.
- [13] D. Cao, W. Liu, W. Xing, and X. Wei, “Human pose estimation based on feature enhancement and multi-scale feature fusion,” *Signal, Image and Video Processing*, vol. 17, no. 3, pp. 643–650, 2022.
- [14] S.-W. Shen, W.-C. Huang, I. T. Anggraini, N. Funabiki, and C.-P. Fan, “Exercise and performance learning assistant system for self-practice dynamic yoga by openpose and fuzzy based design,” in *2022 10th International Conference on Information and Education Technology (ICIET)*, 2022, pp. 16–21.
- [15] W. Kim, J. Sung, D. Saakes, C. Huang, and S. Xiong, “Ergonomic postural assessment using a new open-source human pose estimation technology (openpose),” *International Journal of Industrial Ergonomics*, vol. 84, p. 103164, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169814121000822>
- [16] Y. Saiki, T. Kabata, T. Ojima, Y. Kajino, D. Inoue, T. Ohmori, J. Yoshitani, T. Ueno, Y. Yamamuro, and A. a. Taninaka, “Reliability and validity of openpose for measuring hip-knee-ankle angle in patients with knee osteoarthritis,” *Scientific Reports*, vol. 13, no. 1, 2023.
- [17] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 740–755.
- [18] K. Sun, B. Xiao, D. Liu, and J. Wang, “Deep high-resolution representation learning for human pose estimation,” 2019.
- [19] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, “Deep high-resolution representation learning for visual recognition,” 2020.
- [20] Q. Li, Z. Zhang, F. Xiao, F. Zhang, and B. Bhanu, “Dite-hrnet: Dynamic lightweight high-resolution network for human pose estimation,” 2022.
- [21] W. An, S. Yu, Y. Makihara, X. Wu, C. Xu, Y. Yu, R. Liao, and Y. Yagi, “Performance evaluation of model-based gait on multi-view very large population database with pose sequences,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2, no. 4, pp. 421–430, 2020.