A Single RGB Camera Based Gait Analysis With A Mobile Tele-Robot For Healthcare

Ziyang Wang[1], Fani Deligianni[2], Irina Voiculescu[1], Guang-Zhong Yang[3]

[1]Department of Computer Science, University of Oxford
[2]School of Computing Science, University of Glasgow
[3]Institute of Medical Robotics, Shanghai Jiao Tong University

ABSTRACT

With the increasing awareness of high-quality life, there is a growing need for health monitoring devices running robust algorithms in home environment. Health monitoring technologies enable real-time analysis of users' health status, offering long-term healthcare support and reducing hospitalization time. The propose of this work is twofold, the software focuses on the analysis of gait, which is widely adopted for joint correction and assessing any lower limb, or spinal problem. On the hardware side, a novel marker-less gait analysis device using a low-cost RGB camera mounted on a mobile tele-robot is designed. As gait analysis with a single camera is much more challenging compared to previous works utilizing multi-cameras, a RGB-D camera or wearable sensors, we propose using vision-based human pose estimation approaches. More specifically, based on the output of state-of-the-art human pose estimation models, we devise measurements for four bespoke gait parameters: inversion/eversion, dorsiflexion/plantarflexion, ankle and foot progression angles. We thereby classify walking patterns into normal, supination, pronation and limp. We also illustrate how to run the proposed machine learning models in low-resource environments such as a single entry-level CPU. Experiments show that our single RGB camera method achieves competitive performance compared to multi-camera motion capture systems, at smaller hardware costs.

Index Terms— Gait Analysis, Healthcare, Mobile Robot

1. INTRODUCTION

New technologies which monitor users in a non-clinical environment have been growing in popularity. They keep users engaged and encourage proactive healthcare, providing long-term medical support. Among health monitoring tasks, gait disorders are very common in elderly, patients after surgery or sports injury, which influences the quality of life. This work focuses on gait analysis, which measures user walking patterns for joint correction and assessment of lower limb or spinal problem. A noteworthy application of gait analysis is home monitoring of the gait of patients with neurological disorders, which allows for early intervention and reduces medical expenses. Previous works apply depth cameras [1] or optical markers [2], both of which imply a fairly high cost and limit real-life applicability. Motivated by the recent advances in computer vision, especially human pose estimation and its wide applications in biomedical research, gaming industry and surveillance [3], our design involves a novel marker-less gait analysis device with a low-cost RGB camera mounted on a mobile tele-robot. A single-camera-based gait analysis has two main challenges: (A) Depth ambiguities, motion blurs, unconstrained motions and self-occlusion are harmful for gait prediction [4]. (B) A single image can usually be mapped to more than two possible human poses [3]. In order to circumvent these challenges, we use two state-of-the-art 2D and 3D human pose estimation algorithms, Openpose [5] and VNect [6]. We devise measurements for four bespoke gait angular parameters of the lower limb: inversion/eversion, dorsiflexion/plantarflexion, ankle and foot progression angles. These are defined, extracted and measured. We thereby classify walking patterns into normal, supination, pronation and limp based on these gait features.

Multi-camera systems are not easy to deploy in real life environments. By contrast, our single camera system has been developed as an autonomous mobile tele-robot to be used in a home-like environment. Machine learning models have been implemented in such a way as to run on a single entry-level CPU, such as iPad. An iOS APP with a dedicated user interface, where the skeleton of users in front of the camera is shown in real time, is also developed. As well as processing human gait data, the APP also serves as a tele-robot controller and the robot functions well in an uncluttered environment.

We evaluated the proposed method using manually labelled 2D and 3D ground truth data. Our method achieves competitive performance compared to a commercial multi-camera motion-capturing system, while still requiring lower hardware costs. The code for the iOS APP has been made public through Github.

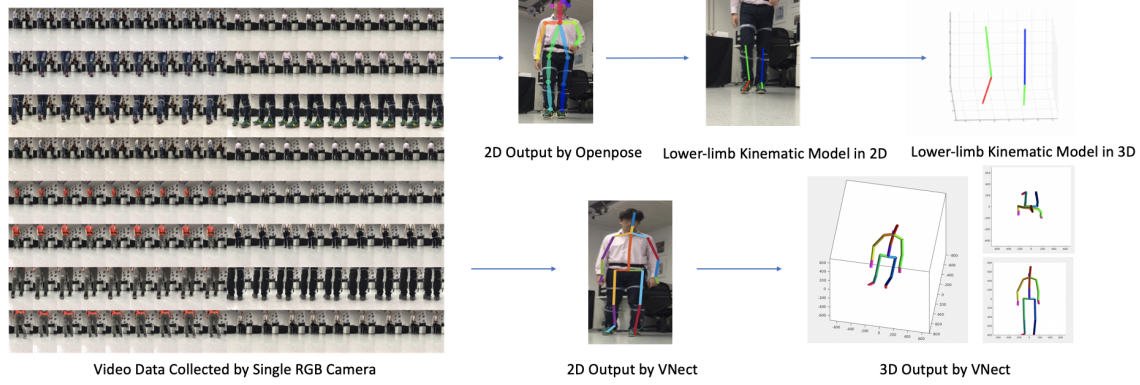Corresponding Author ziyang.wang@cs.ox.ac.uk

Fig. 1: The Example Output for Gait Analysis Based on VNect And Openpose

2. METHODS

2.1. From Human Pose Estimation to Gait Analysis

Both VNect [6] and Openpose [5] are real-time methods to capture skeletal points of human with 2D images.

Openpose applies a bottom-up method of person detection and pose estimation that detects and tracks 17 skeletal feature points based on the Part Affinity Field (PAF). The model can track multiple persons in real time. Figure 1 shows an example of Openpose used for gait analysis. To detect the key points of toes and get the orientation of feet, Gaussian mixture model-based image segmentation is applied [7, 8]. The 2D key points of the joints are mapped to 3D space using active shape modelling together with the dictionary learning [7].

VNect estimates human pose through tracking 21 global skeletal points in 3D, including the foot area. To speed up training and inference, it uses a 50-layer convolutional network. Kinematic skeleton constraints are introduced to improve stability and temporal consistency. Taking the temporal history of the sequence as input, the 2D locations of each joint from the 2D heatmap are transferred to a few links. Then, the length of each link, which is the most important reference in kinematic skeleton constraints, are calculated. Finally, the predictions from the 2D heatmap are mapped to 3D global poses. Only a single person can be tracked with VNect in real-time. Figure 1 shows an example output of VNect, where the 2D skeletal image is drawn based on the heatmap, and the skeletal image in 3D global space.

Several gait analysis parameters are specially extracted and processed. As normal and pathological gaits exhibit varied patterns, six key joints including left and right knees, ankles, toes and their corresponding angles are studied. Inversion/eversion angle, dorsiflexion/plantarflexion angle, and ankle angle are used to evaluate the proposed method's performance. The foot progression angle is studied as a gait angular feature to recognise gait patterns. For example in the i_{th} frame, the left knee, ankle and toe as three joints are repre-

sented $\alpha_{Li} = (xk_{Li}, yk_{Li}, zk_{Li})$, $\beta_{Li} = (xa_{Li}, ya_{Li}, za_{Li})$, $\gamma_{Li} = (xt_{Li}, yt_{Li}, zt_{Li})$. The left shank and foot as two links are represented $\vec{S}_{Li} = \beta_{Li} - \alpha_{Li}$, $\vec{F}_{Li} = \beta_{Li} - \gamma_{Li}$, Inversion/eversion angle is the angle between the foot and horizontal plane. Dorsiflexion/plantarflexion angle is the angle between foot and vertical plane. Ankle angle is the angle between foot and shank. The angular e.g. A_{Li} is illustrated in equation 1.

$$A_{Li} = \arccos \frac{\vec{C}_{Li} \cdot \vec{F}_{Li}}{||\vec{C}_{Li}|| \times ||\vec{F}_{Li}||} \tag{1}$$

Foot progression angle is the angle between the line e.g. $\vec{F}_{Li}$ from the calcaneus to the second metatarsal and the line P as $\frac{x-x_0}{m} = \frac{y-y_0}{n} = \frac{z}{1}$ of progression averaged from heel strike to toe off during the stance phase of walking for each step. P is illustrated in equation 2, and calculated by equation 3.

$$\begin{bmatrix} m & x_0 \\ n & y_0 \end{bmatrix} \begin{bmatrix} z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} m & x_0 \\ n & y_0 \end{bmatrix} = \begin{bmatrix} \sum xa_i za_i & \sum xa_i \\ \sum ya_i za_i & \sum ya_i \end{bmatrix} \begin{bmatrix} \sum za_i^2 & \sum za_i \\ \sum za_i & n \end{bmatrix}^{-1} \tag{3}$$

The evaluation experiment was carried out in a lab of 6*6 meters, equipped with a multi-camera acquisition system (SMART-DX) where markers are attached on human bodies. The multi-camera system was used in order to generate ground truth data. Simultaneously, one of the corners of the lab hosted a low-cost single RGB camera system, which collected raw walking data for validation. Four volunteers were asked first to walk naturally and then to imitate three abnormal walking styles: limping, pronating and supinating. In order to synchronize the data between the single camera and the multi-camera system, the volunteers were asked to jump before and after walking. Both 2D and 3D ground truth data were collected by manually annotating the stream from the multi-camera system. A validation pipeline for each approach

is shown in Figure 2, the differences between the estimated results and the ground truth results are calculated as angular errors of the evaluation metric. As a result of the classification network, four separate walking patterns are accurately classified based on different distributions of gait features.
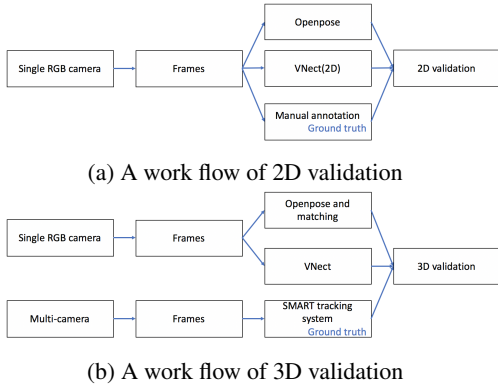

(a) A work flow of 2D validation

(b) A work flow of 3D validation

Fig. 2: Framework of validation

2.2. System Design and Deployment

Having tested and validated the software component of our system, we were keen also to realize its real-life application. Tele-robots are semi-autonomous robots which can be controlled remotely using wireless networks such as WiFi or Bluetooth. Figure 3 illustrates our setup using a Double Robot, an off-the-shelf telepresence robot for iPad tablets, with a single RGB camera. It supports remote control, real-time video chat, auto-answering and self-balancing.
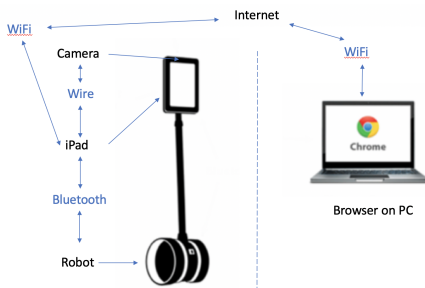

Fig. 3: The framework of how Double Robot works

The software framework is shown in Figure 4. Since the Openpose model is computationally expensive, we modified it using MobileNet [9] to accelerate the model processing. The modified model in the Caffe format is then converted to the CoreML format, which is applicable in an iOS platform. The frame rate achieves 3.5 in the 5th iPad whereas the original model running on Macbook Pro is only 0.6. Since the iPad was connected to the Internet through WIFI and controlled robot with Bluetooth, a remote control system was developed using the UDP protocol.
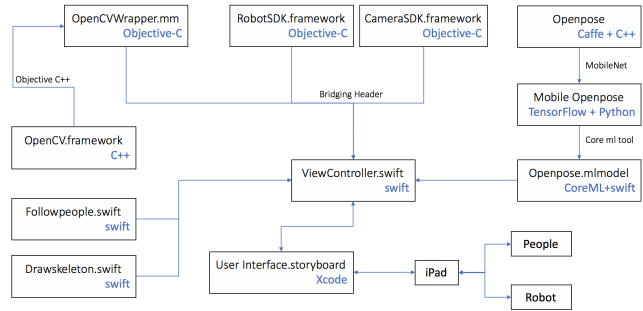

Fig. 4: The framework of software development for iOS APP

3. RESULTS

As the machine learning model has shown the best performance, the evaluation work in this paper focus on whether it could be utilised for gait analysis and comparison between a low-cost single RGB camera and commercial multi-camera with markers. Each of the dorsiflexion/plantarflexion and inversion/eversion angles measured by our system were compared against the values obtained from manual annotations and multi-camera. Any angular difference from the ground truth was considered to be an error. A total of 21,210 frames were studied. The comparisons are shown in Figure 5 and 6.

For 2D validation, we use histograms where the X-axis indicates angular error in degrees and the Y-axis is the percentage of frames where such error occurs. For 3D validation, we use line charts where the X-axis shows the percentage of the gait cycle and the Y-axis is average angular error in degree. The estimation results are shown in box charts which illustrate clearly the distribution of angular features. Figure 5a and 6a show the total 2D angular errors of inversion/eversion and dorsiflexion/plantarflexion. Figure 5b and 6b show the total 2D angular errors for each condition. Figure 5c, 6c , 5d and 6d show the 2D angular errors for inversion/eversion angle and dorsiflexion/plantarflexion angle for four walking patterns, respectively. Figure 5e and 6e present the average 3D foot-leg angular errors in a gait cycle. Figure 5f and 6f present the foot progression angle distribution across different walking conditions.

For 2D angular features, VNect is slightly better than Openpose. The reason is that the toe point is estimated in different ways. VNect model uses toe points during training, but we have to apply ellipse fitting and image segmentation methods to estimate toe points for OpenPose, which results in error propagation. The 3D accuracy of VNect is substantially lower than modified Openpose-based approach. We conjecture that the experimental conditions (camera angle/walking route) do not meet optimal requirements for the 3D kinematic model of VNect. Furthermore, the model does not perform well when the upper body is occluded, which occurs when subjects come very close to the camera. Openpose perform

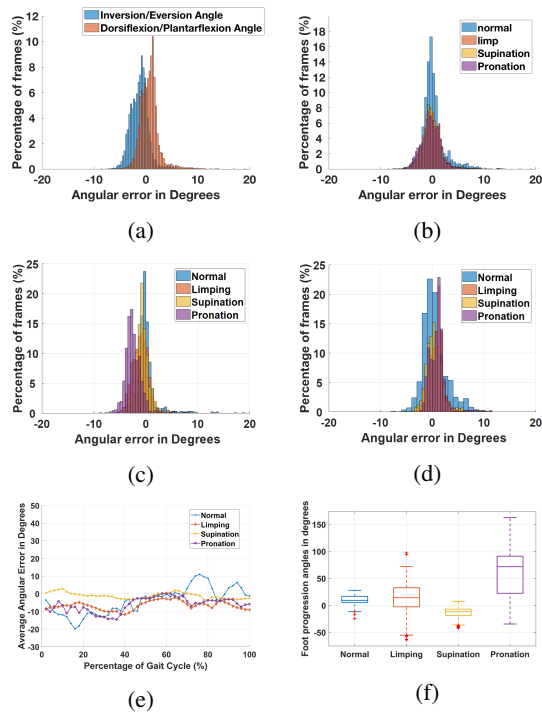on par with VNect on the foot progression angular features.

Fig. 5: Evaluation and Angular Features by Openpose

Fig. 6: Evaluation and Angular Features by VNect

4. DISCUSSION

We investigated vision-based gait analysis with a single RGB camera. We evaluated the accuracy of two state-of-the-art human pose estimation methods and validated them compared to a 3D multi-camera system. We also summarize the gait angular features of four different types of gait. Vision-based pose estimation requires low-cost camera but high computing power. Modification for machine learning models and porting work to an embedded system is completed. A demo mobile robot was further developed for real-life applications and be operated smoothly, widely and stably. The future work should be concentrated on the accuracy of human lower limb estimation and improving robot performance.

5. REFERENCES

[1] Yao Guo, Fani Deligianni, and etc, "3-d canonical pose estimation and abnormal gait recognition with a single rgb-d camera," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3617–3624, 2019.

[2] Paavo Vartiainen, Timo Bragge, Jari P. Arokoski, and Pasi A. Karjalainen, "Nonlinear state-space modeling of human motion using 2-d marker observations," *IEEE Trans. Biomed. Engineering*, vol. 61, no. 7, pp. 2167–2178, 2014.
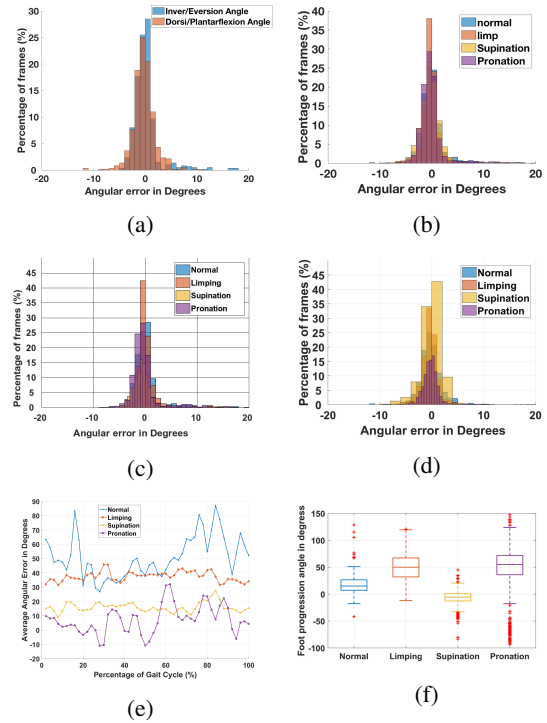
[3] Nikolaos Sarafianos, Bogdan Boteanu, and etc, "3d human pose estimation: A review of the literature and analysis of covariates," *Computer Vision and Image Understanding*, vol. 152, pp. 1–20, 2016.

[4] Fani Deligianni, Charence Wong, Benny Lo, and Guang-Zhong Yang, "A fusion framework to estimate plantar ground force distributions and ankle dynamics," *Information Fusion*, vol. 41, pp. 255–263, 2018.

[5] Zhe Cao, Tomas Simon, and etc, "Realtime multi-person 2d pose estimation using part affinity fields," pp. 1302–1310, 2017.

[6] Dushyant Mehta, Srinath Sridhar, and etc, "Vnect: real-time 3d human pose estimation with a single RGB camera," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 44:1–44:14, 2017.

[7] X. Gu, Fani Deligianni, and etc, "Markerless gait analysis based on a single RGB camera," in *15th IEEE International Conference on Wearable and Implantable Body Sensor Networks, BSN 2018, Las Vegas, NV, USA, March 4-7, 2018*, 2018, pp. 42–45.

[8] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake, ""grabcut": interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004.

[9] Andrew G. Howard, Menglong Zhu, and etc, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *CoRR*, vol. abs/1704.04861, 2017.