Technology Overview: Gait Analysis using 2D and 3D Camera

***Abstract*—In the domains of sports science, biomechanics, osteopathic medicine, and medical diagnostics, gait analysis plays a crucial role. Traditional systems like VICON, which utilize passive markers and sophisticated video cameras, need multiple cameras to produce 3D images, making them costly and intrusive. These systems can be challenging for patients with conditions such as stroke or spinal cord injuries. While markerless systems have been introduced to mitigate these challenges, they often struggle with accuracy, particularly when subjects wear long clothing that hides their gait kinematics. This study seeks to create an affordable, user-friendly, and precise gait analysis system using markerless computer vision techniques. Our method leverages advanced deep learning models for converting 2D images (from smartphone RGB cameras) to 3D reconstructions and pose estimations, thereby improving the accuracy of key joint points and angle measurements. We have worked on a algorithm that accurately identifies the gait cycle and its phases, providing detailed insights into a patient's condition and recovery progress. Through extensive testing and comparison with existing methods, our algorithm demonstrated an overall accuracy of 98.89% for key joint point angles, surpassing the YOLO model. These findings have significant implications for medical and rehabilitation fields, enhancing rehabilitation strategies, optimizing prosthetic designs, and improving patient outcomes. Our system effectively measures the kinematic values of the ankle, knee, and hip, outperforming models like YOLO, which struggle with varying lighting conditions and subjects wearing long clothing.**

I. Introduction

Human gait, the complex pattern of an individual's walking, is defined by a gait cycle that includes a series of movements resulting in locomotion. This cycle starts with the heel strike and continues until the same foot strikes the ground again. Gait analysis, a thorough and systematic approach to evaluating walking, concentrates on three primary components: kinematics (the study of joint movements), kinetics (the analysis of forces), and electromyography (the measurement of muscle activity). Although passive reflective marker-based motion capture systems have become popular for their non-invasiveness compared to the gold standard invasive methods, they face accuracy challenges under different lighting conditions and due to skin movement artifacts. This highlights the need for further advancements in markerless gait analysis techniques. Our primary method involves calculating the gait cycle, identifying its various phases, and measuring the angles associated with these phases with enhanced accuracy. This enables physiotherapists to detect abnormalities, monitor the progress of injured patients, and assess the fitness of sports players, among other applications. We use advanced computer tracking systems like ViTPose and markerless motion capture technologies such as Kinovia to obtain our ground truth values. Traditional gait analysis methods, which often relied on physical markers and simplistic models, encountered challenges such as occlusion issues and data inaccuracies. Our approach aims to overcome these limitations by employing sophisticated, non-intrusive techniques that provide greater precision and flexibility. Initially, we will underscore the importance of gait analysis in rehabilitation and recover

II. Proposed Solutions

The methodology entails a multi-stage process starting with video input framing and person detection. It then proceeds to calculate the complete gait cycle based on the aspect ratio of a person's shape, depth estimation using monodepth2, and 3D reconstruction using STAF on 2D images. Angle calculation is conducted by selecting optimal 2D joint points obtained from ViTPose-H. Temporal analysis and occlusion handling are integrated into the pose estimation process using ViTPose. The gait cycle is further refined by identifying the optimal frame for each of the eight phases, leading to improvements in the overall gait cycle analysis, VIBE algorithm, 3D image reconstruction, and the CSRT algorithm, and the Lucas-Kanade model trained on the latest COCO-Pose dataset. The integration of temporal analysis and occlusion handling has been demonstrated to enhance the accuracy of gait cycle determination. This is further enhanced by incorporating YOLOv8 into the VIBE framework. Improvements have also been made to the ViTPose-H model, especially in foot point detection using the latest COCO-Pose dataset. These advancements are significant for gait analysis, contributing to the development of more accurate and reliable methods for determining the gait cycle. We employed advanced techniques to thoroughly analyze human gait. The input video requirements include a minimum resolution of 800x600 pixels, single-person footage, a 4:3 aspect ratio, and an average frame rate of 30 FPS. Using the OpenCV library, we will frame the video according to the FPS. Person detection will be conducted using the YOLOv8 model trained on the COCO-Pose dataset, followed by image cropping. Pose estimation will involve detecting keypoints and calculating angles using YOLOv8 and geometric angle calculations. The complete gait cycle calculation will include signal smoothing and determining the initial contact of the right leg. Depth estimation will be performed using the Monodepth2 technique, while 3D reconstruction will utilize the STAF model. Temporal analysis will involve motion analysis and shape checking of the person. Occlusion handling will resize 3D mesh files, and pose estimation using ViTPose-H will

detect poses from 3D mesh files and calculate angles. Finally, the gait cycle calculation will identify the right leg, measure the aspect ratio, and calculate gait cycle phases, with results presented cycle-wise in eight phases. Our approach includes the following steps:

- **Video Input Framing**

We start by recording video footage of subjects walking, which is essential for the following analysis steps. The input video must meet certain requirements, such as a minimum resolution of 800x600 pixels, featuring only one person, an aspect ratio of at least 4:3, a frame rate above 25 FPS, showing the full body, having a single background, and being recorded with a single camera. We frame each frame of the video at a rate of 1 frame per second (FPS) using the OpenCV library

- **Person Detection and Cropping**

To separate the subjects from the background, we utilize YOLOv8 for person detection. This enables us to concentrate on the relevant parts of the video, specifically the subjects' movements. We identify a single person in the video using YOLOv8 and crop the images accordingly. The YOLOv8 model is trained using the COCO-Pose dataset.

- **Depth Estimation**

We employ the monodepth2 depth estimation technique to measure individuals' depth in gait analysis. This method is based on a self-supervised learning approach that uses a robust reprojection loss to construct a depth model. To minimize visual artifacts, monodepth2 incorporates an automasking loss to disregard training pixels that deviate from assumptions about camera motion and a full-resolution multi-scale sampling technique (Godard et al., 2019). The challenging task of monocular depth estimation involves determining the depth value of each pixel, or its distance from the camera, based on a single RGB image. It is a crucial need for figuring out scene comprehension in applications like augmented reality, driverless vehicles, and 3D scene reconstruction. The most advanced techniques for monocular depth estimation generally fall into one of two categories: either developing a sophisticated network capable of directly regressing the depth map or dividing the input into windows or bins to reduce computational complexity. Monodepth2 techniques have demonstrated high-quality depth estimation results in various scenarios, including walking, running, and jumping. This method has been trained on the KITTI dataset , which is commonly used for monocular depth estimation tasks. By estimating the depth of subjects, we can accurately reconstruct their 3D movements, offering valuable insights into their biomechanics and kinematics. This information is essential for improving rehabilitation strategies, optimizing prosthetic designs, and enhancing our understanding of human movement patterns.The ViTPose 31 model is a powerful tool that can detect pose from 3D mesh images, calculate the angle between the keypoints using geometric angle calculation, and use the coco_25 dataset .

The ViTPose model is a state-of-the-art pose estimation algorithm developed by the ViTAE-Transformer team. It utilizes a transformer architecture and has demonstrated strong performance in various applications, including human pose estimation. The model is trained on the coco_25 dataset, which contains a vast number of annotated images of humans in various poses. By employing the ViTPose model, we can evaluate the performance of different pose estimation algorithms and select the most suitable one for our specific application.

Using the ViTPose model, we first detected poses from 3D mesh images. This process involves extracting relevant information from the 3D mesh files and converting them into a format that the model can process. We calculate the angles between keypoints using geometric angle calculations. This step is crucial for understanding the subjects' movements and biomechanics. The ViTPose model is a valuable tool for gait analysis, as it enables us to perform accurate pose estimation and make comparative assessments. Utilizing this model allows us to gain a more profound understanding of the subjects' movements and biomechanics. This insight is essential for improving rehabilitation strategies, optimizing prosthetic designs, and advancing our knowledge of human movement patterns.

- **Calculate Complete Gait Cycle**

To calculate the complete gait cycle, we first identify the right leg by comparing the right key points with the left key points. We then remove the initial gait cycle from the video to ensure it includes more than three cycles. Since a complete gait cycle consists of two steps, one for each leg, analyzing multiple cycles helps ensure the accuracy of our results.
To analyze the gait cycle, we determine the relative positions of the legs by subtracting the position of the right leg from that of the left leg at a given time, using YOLO pose estimation techniques. This allows us to calculate the distance between the legs through a geometric distance calculation method.
Let:
-xL and yL be the x and y coordinates of the left ankle, respectively.
-xR and yR be the x and y coordinates of the right ankle, respectively.
Then, the equation can be written as:
$$\textit{distance} = (xR + yR)(xL + yL)$$

The accuracy of gait cycle calculation depends greatly on correctly identifying the right leg. Any errors in this identification can significantly affect the accuracy of the analysis. Therefore, it is crucial to ensure accurate identification of the right leg before proceeding with the gait cycle calculation.

- **Calculate Phase within Gait Cycle**

Determining the phase within the gait cycle is another critical step in comprehending the overall structure of the subjects' movements. By integrating angle calculations with temporal analysis, we can identify the gait cycle and its associated phases. To identify the right leg, we compare the right key points with the left key points and measure the aspect ratio of the video. This process ensures the video contains more than three cycles and that the right leg is accurately identified.
After identifying the right leg, we remove the initial gait

cycle from the video to ensure that the remaining cycles are consistent and representative of the subject's typical gait. We then calculate the gait cycle by subtracting the position of the right leg from the left leg at each time point. To determine the phase within the gait cycle, we analyze the changes in joint angles over time and compare them to the gait cycle. Using the positive values of the left and right leg subtraction, we can identify the various phases of the gait cycle, including initial contact, loading response, mid-stance, terminal stance, pre-swing, initial swing, mid-swing, and terminal swing. Accurately calculating these phases provides a deeper understanding of the subjects' movements and biomechanics.

Environmental Setup

To ensure the success of our work, we relied on a comprehensive set of tools encompassing hardware, software, and development components. Regarding hardware, we used an Windows setup with an NVIDIA 3060, 16GB of RAM, and a 1TB SSD, providing the necessary computational power and storage capacity. For software, our toolkit included Google Colab, GPU, Kinovea, and MS PowerPoint for data analysis, visualization, and presentation tasks. Additionally, we employed various development tools such as iOS, Python 3.10.12, OpenCV, NumPy, Seaborn, Pandas, Matplotlib, Ultralytics, Torch, MXNet, GluonCV, Pytube, Hugging Face Hub, JSON, PIL, SciPy, and VitInference. Each tool played a critical role at different stages of the project, from data preprocessing to model development and evaluation. The successful completion of our work was made possible by the collective utilization of these tools.

Dataset Evaluation

In this section, we assess the performance of our proposed gait analysis system using the recorded video dataset as the primary dataset. This dataset was captured in a controlled environment with optimal lighting conditions, minimal background distractions, and a camera setup that provided a clear view of the participant's gait.

The environment was set up to minimize distractions and ensure optimal lighting conditions. The room was dimly lit with a single light source positioned at a 45-degree angle to the participant's body. The background consisted of a plain white wall, free from any objects or distractions that could interfere with the participant's gait.

The camera was placed at a height of 2 meters and angled at 45 degrees relative to the participant's body. It was calibrated to capture the participant's entire body, from head to toe, at a resolution of 800x600 pixels and a frame rate of 30 frames per second. The camera settings were optimized for image quality, with an exposure of 1/100th of a second and an ISO of 100. Additionally, the camera was set to capture images in RAW format to ensure the highest image quality.

Participants were asked to wear casual clothing that did not restrict their movement, chosen to minimize potential distractions or interference with their gait.

Alongside the recorded video dataset, we used the COCO-Pose dataset to train the YOLOv8 model. The COCO-Pose dataset is a large-scale collection of over 200,000 images and annotations used for training and evaluating pose estimation models, with each image annotated for 25 keypoints. Additionally, we utilized the coco_25 dataset to train the ViTPose-H model. The coco_25 dataset is a subset of the COCO-Pose dataset, comprising 25,000 images, each with annotations for 25 keypoints.

Mathematical Consideration

Initial Contact (IC): This phase marks the beginning of the gait cycle, representing the moment when the foot first touches the ground. **IC = mn_min[i]**

Loading Response (LR): This phase indicates the moment when the heel of the foot strikes the ground. It also refers to the phase where body weight is transferred to the stance leg following initial contact. **LR = IC +mn_max[i]/ 2**

Mid-Stance (MS): This phase occurs when the body weight is directly over the stance leg, and the leg is supporting the load. **MS = mn_max[i]4**

Terminal Stance (TS): This phase represents the final portion of the gait cycle, where the foot is in contact with the ground and the body's weight is supported by the foot. Marks the end of the stance phase, just before the swing phase begins. **TS = mnmin[i+1]**

Pre-Swing (PS): This is the phase of transition between stance and swing, preparing the leg to lift off the ground. **PS = TS+mn_max[i+1]/2**

Initial Swing (IS): The leg begins to swing forward, lifting off the ground. **IS = mn_max[i+1]4**

Mid-Swing (MS): This phase represents the middle portion of the gait cycle, where the foot is lifted off the ground and swung forward. The midpoint of the swing phase, is where the leg is at its highest point. **MSw = IS+mn_min[i+2]/2**

Terminal Swing (TS): This phase marks the concluding part of the gait cycle, where the foot is lifted off the ground and swings forward in preparation for the next step. It is the final segment of the swing phase, just before the next cycle starts. **TSw = mn _min [i+2]**

Conclusions

In this study, we introduce a novel computer vision-based method for human gait analysis, providing a robust, cost-effective, markerless, and user-friendly system. Utilizing advanced technologies such as YOLO and ViTPose, we demonstrate significant potential for enhancing robustness and accuracy. However, a major limitation of this approach is managing motion in 3D frames. Although improvements in gait analysis using markerless computer vision techniques have been shown, the accuracy of results may be impacted by 3D frame motion. We are still in the process of developing and finalizing the application.