2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids)
Nancy, France. November 22-24, 2024

Guidelines For Optimal Human Mesh Generation Using Deep Learning-Driven Avatar Reconstruction For Gait Analysis

Odysseas I. Stavrakakis[1], Athanasios Mastrogeorgiou[1],
Aikaterini Smyrli[1], *Student members, IEEE,* and Evangelos Papadopoulos[1], *Life Fellow, IEEE*

***Abstract*— Gait analysis is essential in many scientific fields; to study it marker-based or markerless motion capture (MoCap) techniques are used. The latter have significantly benefited from the recent rise of research in deep learning (DL) and its applications on human mesh generation. However, insufficient and suboptimal camera viewpoint selections often lead to low-grade human mesh geometries. This paper presents an approach to consistently obtain accurate human meshes using DL-based avatar reconstruction algorithms (ARAs). Our framework provides a systematic approach, utilizing a simulated environment to inform decisions on the number of cameras and their spatial configuration to achieve optimal reconstruction results. These results are enhanced through mesh evaluation, mesh alignment, and surface reconstruction to remove poorly formed geometries and artifacts. Additionally, we present a gait analysis tool, tested in simulation and reality (Fig. 1), that detects gait phase changes, extracts the significant human body joint angles and recreates the animation of the gait cycle in 3D space. The proposed approach is open-source, adjustable, and applicable to various research contexts where gait analysis is essential.**

I. Introduction

Gait analysis is an integral tool in many scientific fields, from clinical diagnostics to biometric security and sports performance optimization [1]. It acts as a diagnostic instrument for analyzing musculoskeletal and neurological conditions, providing insights valuable for the development of treatment plans [2]. The metrics used by gait analysis are objective and quantifiable, providing researchers with a deeper understanding of human biomechanics, with applications in technique refinement and injury prevention [3]. Lastly, biometric characteristics such as the gait cycle are unique and crucial in identification and verification tasks, with high accuracy and resistance to impersonations [4].

The process of gait analysis involves several key steps, crucial in accurately capturing, analyzing, and interpreting motion data. The data are acquired through MoCap systems, preprocessed to remove any noise or artifacts, and analyzed using dedicated software [5], to extract gait parameters such as the stride length, the step width, the ground reaction forces, etc. [6].

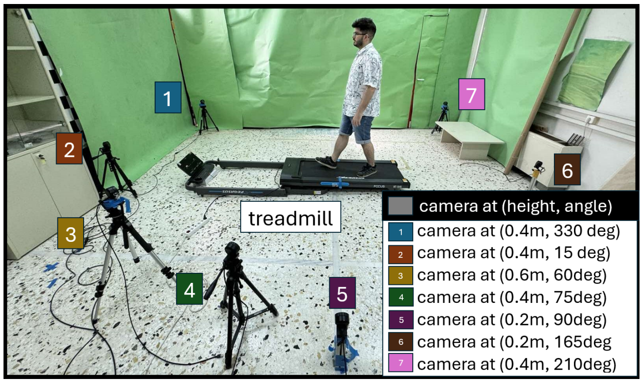

1
2
3
4
5
6
7
treadmill
camera at (height, angle)
1 camera at (0.4m, 330 deg)
2 camera at (0.4m, 15 deg)
3 camera at (0.6m, 60deg)
4 camera at (0.4m, 75deg)
5 camera at (0.2m, 90deg)
6 camera at (0.2m, 165deg
7 camera at (0.4m, 210deg)

Fig. 1: Experimental Setup at CSL NTUA.

Traditionally, gait analysis has relied on marker-based MoCap systems, force plates, and video analysis, with systems like Optitrack, Qualisys, and Vicon at the forefront [7]. While accurate, these systems are expensive and require specialized equipment as well as trained experts to operate them. Conversely, markerless MoCap approaches might be easier to operate but lack specialized capabilities such as gait phase detection and recognition and rely on software restricted by proprietary barriers [8][9].

Recent advancements in computer vision and DL have opened new avenues for more accessible and efficient techniques to be used as gait analysis tools [10]. Several open-source projects have created human avatars directly from RGB images [11][12][13], with implementations mainly directed to game development, augmented reality (AR) and virtual reality (VR). Yet, as we presented in [14], their results are highly dependent on the position and orientation of the camera concerning the human target. Challenging and suboptimal camera angles lead to poor stance detection and badly formed meshes, as demonstrated in [14]. Numerous projects have implemented DL methods to obtain the gait energy image [15], with applications to human recognition [16][17]. However, applying DL-based ARAs for gait analysis remains unexplored.

This paper presents a novel, open-source approach for gait analysis utilizing DL-based ARAs to produce robust results that are significant to the motion analysis community. Specifically, our framework provides: (a) an innovative procedure that optimizes the number and positioning of the cameras based on the selected ARA to mitigate problematic video

[1] O. Stavrakakis is with the School of Mechanical Engineering, National Technical University of Athens. All other authors are with the Athena RC Robotics Institute, Marousi, 15125, Greece, and with the School of Mechanical Engineering, National Technical University of Athens. E-mails: `mc18050@mail.ntua.gr`, {`a.mastrogeorgiou`, `katerina.smyrli`, `egpapado`}`@athenarc.gr`.

This research work was supported by the project “Applied Research for Autonomous Robotic Systems” (MIS5200632) which is implemented within the framework of the National Recovery and Resilience Plan (NNRP) “Greece 2.0” (Measure: 16618- Basic and Applied Research) and is funded by the European Union-NextGenerationEU.

979-8-3503-7357-8/24/$31.00 ©2024 IEEE

capturing that leads to poorly formed meshes as we presented in [14], (b) an automated process that combines multiple video inputs to create human meshes with improved surface accuracy in the leg and foot regions, and (c) a versatile Blender [18] add-on tool that reconstructs human motion in 3D space and extracts gait parameters. Fig. 1 presents the experimental setup that employs a multiple-input video-capturing process.

This paper is organized as follows: Section II elaborates on the most significant attributes of the camera selection process, Section III presents the custom Blender add-on that assists in gait analysis, Section IV presents the developed pipeline, and Section V contains the experimental results. Section VI concludes the paper and discusses future research directions.

II. Camera Selection Framework

A. *Simulated Environment*

The task of camera selection is foundational for every computer vision framework. In this work, a simulation was developed to obtain unrestricted oversight of the environmental parameters and maintain precision and consistency across different scenarios. The simulation employs elements from Adobe MIXAMO library [19], which provides several combinations of characters and animations, and Blender, which successfully integrates these animations into a 3D scene with realistic environmental conditions. These tools are utilized to produce a naturally-moving *digital actor*, who is recorded by the simulated MoCap system. The videos are used to produce 3D meshes, whose quality is maximized via the optimization of the camera setup, see Fig 1.

Fig. 2 depicts the digital actor in the studied gait cycle, which is comprised of 6 phases: Loading Response (LR), Mid-Stance (MS), Terminal Stance (TS), Pre-Swing (PS), Initial Swing (IS) & Terminal Swing (SW), separated by six events: Initial Contact (IC), Foot Flat (FF), Heel Rise (HR), Contralateral IC (CIC), Toe Off (TO) & Feet Adjacent (FA) [20].

The animation cycle is composed of 30 keyframes, augmented by linear interpolations. This approach guarantees a high degree of granularity in the motion while maintaining sufficient motion resolution. As a result, the gait phases and their transitions are depicted with precision, allowing for a nuanced and accurate representation of the gait.

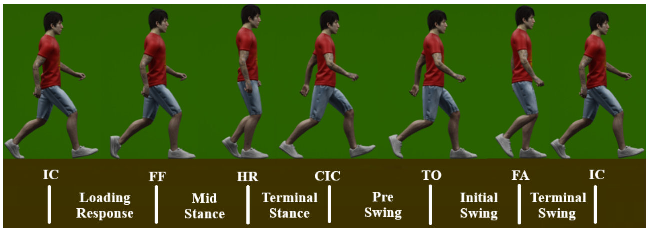


Fig. 2: Gait Cycle with 6 Phases.

B. *Mesh Creation and Evaluation*

The chosen ARA is the Explicit Clothed humans Optimized via Normal integration (ECON) algorithm [11] due to its capability to merge the characteristics of implicit and explicit representations, thereby inferring high-fidelity human meshes from a single RGB image. Compared to other state-of-the-art ARAs, it demonstrates significantly greater consistency and precision in leg reconstruction, which is essential for the study of the gait.

The selected ARA processes the digital actor's simulated recordings. This can be compared to the digital actor's known mesh, which serves as the ground truth (GT). The mean cloud-to-cloud distance metric is employed to evaluate the extracted meshes and identify the optimal combination of cameras that provides an effective view of the target. The most suitable method for obtaining the mean cloud-to-cloud distance is the quadratic regression [21]. A quadratic regression accurately captures 3D geometries and smooth surface variations by utilizing the quadratic function

$$f = \alpha x^2 + \beta x + \gamma xy + \delta y + \epsilon y^2 + g \qquad (1)$$

where $\alpha, \beta, \gamma, \delta, \epsilon, g$ are the curvature coefficients, x and y are the surface coordinates and f is the surface height. The quadratic function facilitates surface comparisons by approximating the reconstructed and GT surfaces using an algebraic expression. The final cloud-to-cloud distance is the spatial difference between the surface heights of the quadratic approximations. Thus, the distance between the GT and the generated meshes is quantified and used as an evaluation score.

Fig. 3 compares full-body and subischial [22] meshes to their corresponding evaluation results using a color grading scale ranging from blue for minimum deviation from the GT to red for the maximum deviation from the GT. Note that the inclusion of the hands and upper body can impact the evaluation score due to mesh inaccuracies in these areas (purple circle in Fig. 3); however, the hands and upper body are not significant for gait analysis. At the same time, excluding the upper body from the evaluation process while maintaining the same color grading scale allows for minor mesh deviations in the lower body (red ellipses in Fig. 3) to be emphasized (light blue ellipses in Fig. 3).

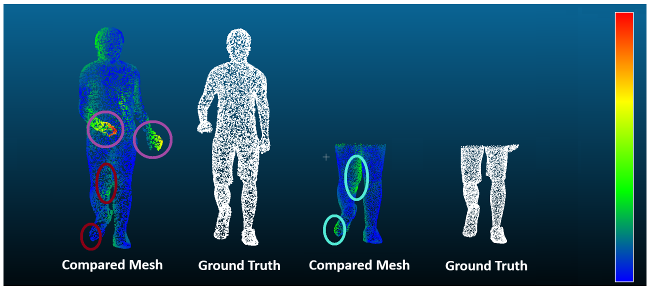


Fig. 3: Full Body versus Subischial Mesh Comparison. The inclusion of the upper body results in a lower evaluation score due to mesh inaccuracies.

C. Camera Grid Initialization

The simulation is initialized by the generation of a camera grid in a cylindrical formation around the target (Fig. 4). To pinpoint the optimal angle increments of the camera grid, we calculate the difference in mesh score for two cameras at different angles, see Fig. 5a. For this study, we consider that a $10mm$ difference between consecutive grid viewpoints is significant. As a result, angular increments of 15°are produced. The height-score correlation is estimated respectively; see Fig. 5 b, and the height increment is set at $0.2m$ to maintain the same score difference ($10mm$). This produces 7 height intervals, starting at $0.0m$, where the camera axis is aligned with the foot, and ending at $1.2m$, i.e., the hip height for 95% of the human population [23]. Respectively, the angle intervals are 24, starting at 0°and ending at 360°, to surround the subject. The definitive camera grid is made up of 168 cameras.

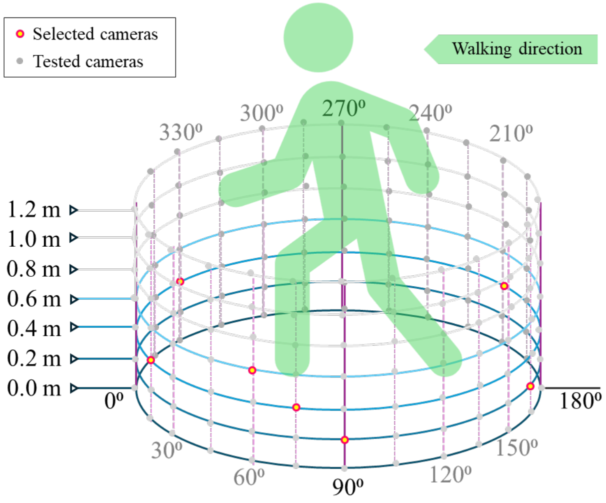


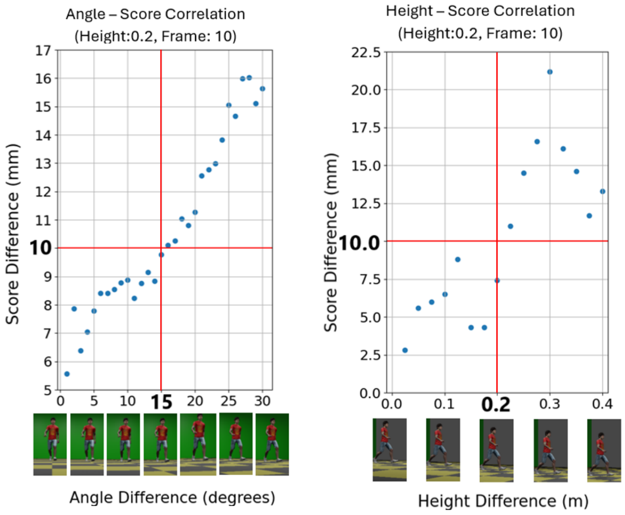
Fig. 4: Cylindrical camera grid of candidate cameras around the human target.

D. MSE Score Calculation

The camera with the optimal viewpoint for every animation keyframe must be identified to generate high-quality meshes. This task is accomplished by determining the camera grid point whose mesh has the minimum difference from the corresponding GT. This process is performed by Algorithm 1.

Initially, the evaluation score matrix $a_{i,j}$ is determined, where $i = 1, ..., F$ is the camera number, $j = 1, ..., F$ is the number of keyframes, and a are the evaluation scores. Then, by identifying the $min(a_{i,j})$ for every row, each camera is assigned the frames that enable the ARA to generate a high quality mesh. Finally, to quantify the performance for this set of cameras, the Mean Square Error (MSE) [24] is calculated using the $min(a_{i,j})$ of each frame:

$$MSE = \frac{1}{F}\sum_{i=1}^{T} min(a_{i,j}) \qquad (2)$$



(a) (b)

Fig. 5: Camera Grid Increments.

Algorithm 1 MSEScoreCalculation

```
Input:
  • $a_{T,F}$, the evaluation scores for $T$ cameras and $F$ keyframes
Output:
  • $MSE$, the mean square errort
  • $CID_F$, the camera ID assigned to each frame
$CID_F = array[]$
for $i$ in $range(1, T)$ do
    $CID_F[i] = a_{i,0}$
    for $j$ in $range(1, F)$ do
        if $a_{i,j} < CID_F[i]$ then
            $CID_F[i] = a_{i,j}$
        end if
    end for
end for
$MSE = 0$
for $i$ in $min_a$ do
    $MSE = MSE + min_a[i]^2)$
end for
$MSE = MSE/T$
return $MSE$ , $CID_F$
```

Tables I and II depict the ideal number of cameras to capture the six gait cycle events and the ideal cameras to capture each of the 30 keyframes of the animation sequence. The minimum number of cameras required is equivalent to the number of cameras that best detect all events presented in Fig. 2. Respectively, the camera set that includes the optimal camera for each keyframe defines the maximum camera count. These values are 4 (min) and 18 (max), respectively, since some cameras are assigned as optimal in more than one instance. It should be noted that a higher motion resolution can be achieved simply by using an animation with more keyframes. Nevertheless, the same guidelines apply, and the maximum camera count directly correlates with the number of keyframes.

TABLE I: Minimum Camera Count.

Event Left Foot	Event Right Foot	Normalized Evaluation Score (NES)	Camera Height (CH)	Camera Angle (CA)
IC	CIC	118	0.2	90
FF	TO	112	0.4	75
HR	FA	94	0.2	90
CIC	IC	120	0.6	60
TO	FF	128	0.2	90
FA	HR	84	0.2	165

TABLE II: Maximum Camera Count.

Frame	NES	CH	CA	Frame	NES	CH	CA
1	118	0.2	90	16	120	0.6	90
2	121	0.4	90	17	132	1.0	90
3	112	0.4	90	18	128	0.2	90
4	115	0.4	45	19	131	0.2	90
5	129	0.4	60	20	135	0.4	270
6	110	0.4	345	21	117	0.4	210
7	94	0.2	165	22	84	0.2	165
8	110	0.2	330	23	90	0.4	15
9	125	0.8	75	24	96	0.2	15
10	128	0.8	270	25	116	0.4	75
11	124	0.2	90	26	116	0.4	75
12	123	0.2	90	27	135	0.6	255
13	125	0.6	90	28	116	0.8	90
14	109	0.6	90	29	103	0.2	90
15	108	0.4	90	30	110	0.6	90

E. Camera Addition - Forward Pass

A forward pass is implemented to initialize the camera selection framework. It begins with the minimum camera count and iteratively incorporates the next optimal camera until the maximum camera count is reached. Algorithm 2 presents this process and identifies the camera, whose addition minimizes the MSE.

The Attainable Improvement Index (AII) is introduced to quantify the MSE scores produced from each iteration. AII is defined as

$$AII = \frac{MSE_{max} - MSE_{current}}{MSE_{max} - MSE_{min}} \tag{2}$$

where MSE_{max} is the maximum attainable score, MSE_{min} is the minimum attainable score, and $MSE_{current}$ is the score of the current iteration. MSE_{max} occurs for the lowest camera count and MSE_{min} for the highest camera count. Thus, AII illustrates the percentage of the maximum attainable score achieved with each iteration, as shown in Fig. 7. The camera addition process is terminated when the addition of another camera leads to a less than 5% improvement in the quality of the resulting meshes. This produces an AII close to 90%, corresponding to 10 cameras.

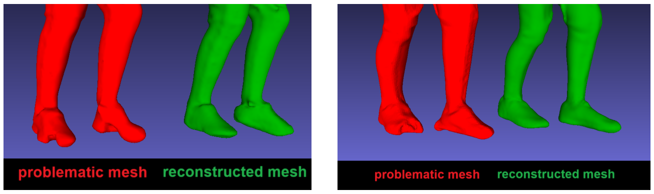
Algorithm 2 OptimalCameraAddition

```
Input:
  • $MSE_B$, the score for base camera set
  • $a_{B,F}$ base camera values
  • $a_{C,F}$ contender camera values, where $C = Total - B$
Output:
  • $MSE_{B+1,min}$, the new $MSE$ score
  • $camera_{added}$, the added camera
$base_a = matrix[]$
for $i$ in $range(1, B)$ do
    for $j$ in $range(1, F)$ do
        $base_B[i, j] = a_{i,j}$
    end for
end for
$MSE_{B+1,min} = MSE_B$
$camera_{added} = 0$
for $k$ in $range(1, C)$ do
    for $j$ in $range(1, F)$ do
        $cont_a[j] = a_{k,j}$
    end for
    $base_{B+1} = base_B + cont_a$
    $MSE_{B+1}$ = ModelScoreCalculation $(base_{B+1})$
    if $MSE_{B+1} < MSE_{B+1,min}$ then
        $MSE_{B+1,min} = MSE_{B+1}$
        $camera_{added} = k$
    end if
end for
return $MSE_{B+1,min}, camera_{added}$
```

F. Surface Alignment and Reconstruction

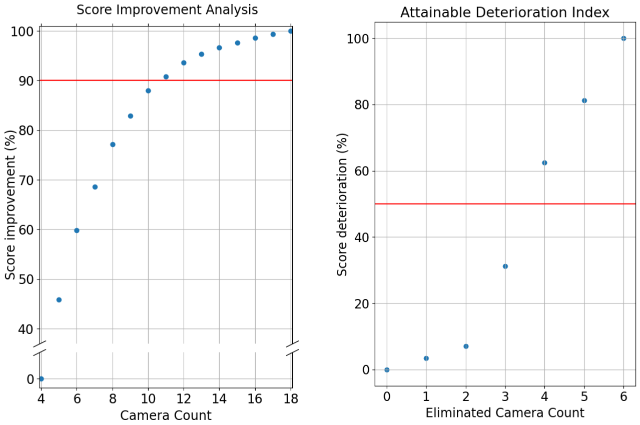
A crucial part of the camera selection is the prioritization of meshes with accurate leg positioning (ALP). However, ALP meshes sometimes have surface irregularities, especially in the sub-ankle/foot region. However, some meshes exhibit robust foot geometries (RFGs) but imprecise leg positioning. To produce optimized meshes, corresponding ALP and RFG meshes are combined through the process of Poisson surface reconstruction [25] with Iterative Closest Point (ICP) [26] alignment.

This procedure is initialized by performing a new evaluation round exclusively for the foot region of the meshes. Thus, the RFG meshes for each frame are derived. The ALP meshes are produced by Algorithm 2. Subsequently, the ICP algorithm is performed between the ALP and SFG meshes of the corresponding frames to align them. Finally, utilizing the Poisson surface reconstruction , the leg surfaces of the ALP mesh are combined with the foot surfaces of the RFG mesh, producing improved meshes like the ones in Fig. 6.



(a) Frame 3 (b) Frame 24

Fig. 6: Surface reconstruction examples.



(a) AII (b) ADI

Fig. 7: Score metrics for forward and reverse pass.

G. Camera Elimination - Reverse Pass

The surface reconstruction process can be extended from mesh optimization to camera elimination. Following the elimination of a camera, the frames initially assigned to it are redistributed to the cameras that offer the next optimal mesh of the leg region. This redistribution may result in quality degradation due to the sub-optimal viewpoint of the newly assigned cameras. This degradation is mitigated by performing surface reconstruction, enabling the outcomes to attain mesh quality identical to that of the version before the camera elimination. Thus, the eliminated camera can be rendered redundant.

Algorithm 3 describes the process of identifying the cameras with the lowest contribution to the MSE by finding the ones whose elimination results in the minimum MSE increase. The Attainable Deterioration Index (ADI) is introduced, defined as

$$ADI = \frac{MSE_{current} - MSE_{min}}{MSE_{max} - MSE_{min}} \tag{3}$$

In the reverse pass, the MSE_{min} corresponds to a 10 camera set, and the MSE_{max} corresponds to a 4 camera set. Fig. 7 shows that by selecting a 50% threshold for ADI, the eliminated cameras are 3, resulting in a setup with 7 cameras.

III. Blender Add-on

Following the simulation-informed design of the camera setup, the meshes are combined in a 3D animation of the digital actor, which can be used to analyze the actor's gait. For this purpose, the Gait Analysis Module for Avatar Reconstruction in 3D (GAMA 3D) was developed as an add-on for the Blender toolbox. Users can import the generated meshes, apply spatiotemporal transformations, and reconstruct the original animation in a virtual 3D environment. Applying a Boolean modifier extracts the intersections between the floor-plane and the feet of each mesh, thereby establishing the virtual footprint. The area of the virtual footprint is utilized to measure the accuracy of the mesh reconstruction and detect phase-changing events. Fig. 8 demonstrates the virtual footprint results at frames with such events.

Algorithm 3 Optimal Camera Elimination

Input:

- MSE_B, the score for base camera set
- $a_{B,F}$ base camera values

Output:

- $MSE_{B-1,min}$, the new MSE score
- $camera_{elim}$, the eliminated camera

```
$base_a = matrix[]$
for $i$ in $range(1, B)$ do
    for $j$ in $range(1, F)$ do
        $base_B[i, j] = a_{i,j}$
    end for
end for
$MSE_{B-1,min} = 2MSE_B$
$camera_{elim} = 0$
for $i$ in $range(1, B)$ do
    $base_{B-1} = base_B - base_B[i]$
    $MSE_{B-1}$ = MSEScoreCalculation ($base_{B-1}$)
    if $MSE_{B-1} < MSE_{B-1,min}$ then
        $MSE_{B-1,min} = MSE_{B-1}$
        $camera_{elim} = i$
    end if
end for
return $MSE_{B-1,min}, camera_{elim}$
```

IV. Experimental Validation

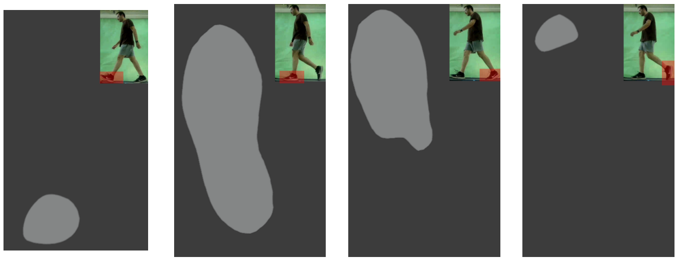
A. Real World Setup

The developed framework has facilitated the identification of the optimal camera configuration via a simulated setup of up to 168 cameras and the contribution of an avid digital actor. The significant number of cameras and the large number of simulated gaits that were tested would not have been possible without the simulation-based approach. In our laboratory, a physical implementation of the optimal 7-camera layout has been designed to validate the results.

The experimental setup consists of a treadmill placed in the center of the action space and 7 Razer Kiyo Pro [27] cameras that capture full HD video at $60FPS$. These cameras are positioned around the target in a circular formation of $2.5m$ radius, as shown in Fig. 1.

B. Pipeline Summary

The framework used in our experiments employs the pipeline presented in Fig. 9. Specifically:



(a) IC (b) FF (c) HR (b) TO

Fig. 8: Left Footprint Event Examples.

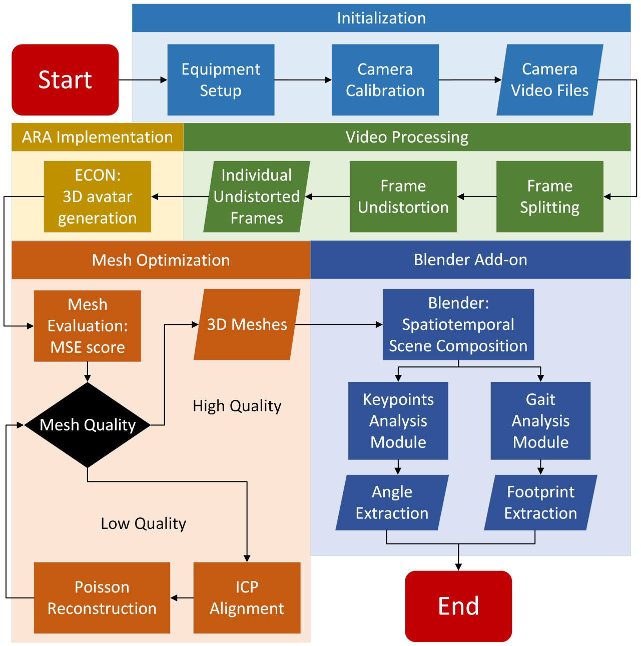


Fig. 9: The steps of the experimental pipeline.

Initialization: This involves the equipment setup, the calculation of the intrinsic and extrinsic camera parameters [28], and the recording of the footage from all cameras.

Video Processing: The footage is split into monocular inputs, which in turn are broken down into individual frames. Then each frame is undistorted using the camera matrix [29].

ARA Integration: ECON [11] is applied to obtain 3D reconstructed human meshes. With minor adaptations, ECON can be replaced with any ARA based on Skinned Multi-Person Linear Model (SMPL) [30].

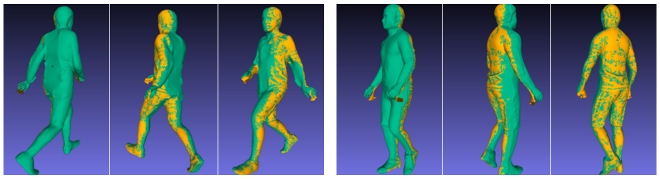
Mesh Optimization: The quality of the mesh is evaluated. The mesh is forwarded to the next step or enhanced using the surface reconstruction with ICP alignment process.

Blender Implementation: The produced meshes are combined into an animation utilizing GAMA 3D. By integrating Blender modifiers with custom-made functionalities, information about the joint angles of the human actor is calculated, the distinct phases of the gait cycle are recognized, and gait parameters are extracted.

V. Results

A. *Mesh Results*

We conducted several experiments in the simulated and real-world environments to validate the findings from the camera selection pipeline. Examples of meshes obtained by applying the pipeline of Fig. 9 are presented in Fig. 10.



(a) frame1, cam at (0.2m, 90°) (b) frame 7, cam at (0.2m, 165°)

Fig. 10: High-quality Meshes produced by the camera selection framework. Areas visible by the cameras are highlighted with orange.

B. *Camera Positioning and Assignment*

The selected cameras produce the optimal mesh quality. However, there are regions where the produced meshes are consistently good. Regions examples in Fig. 11 are LR-B, TS-E, IS-A, and SW-B, highlighted with dark green. Notably, they are symmetrical with respect to the axis of motion, and the chosen angle-height combinations are always within them. For example, camera 5 from Fig. 1 utilizes regions LR-E and TS-E. Although the framework

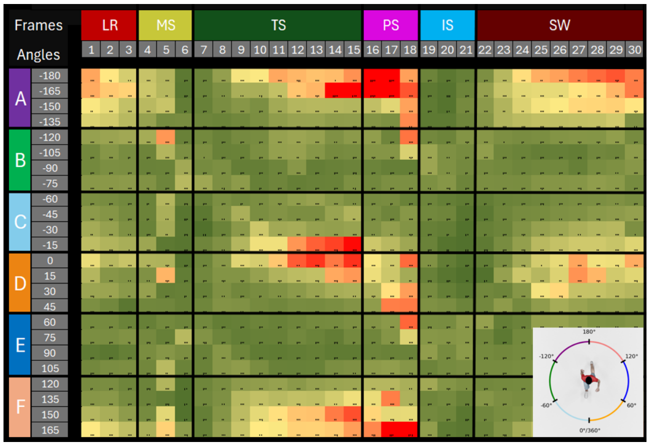
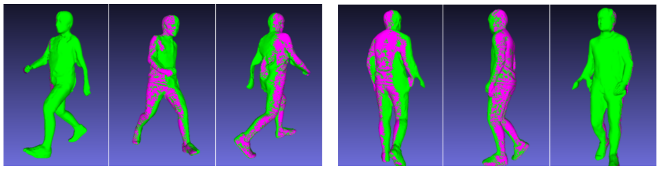


Fig. 11: Evaluation Score Heatmap.



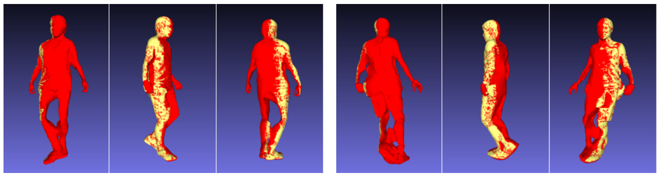
(a) frame1, cam at (0.6m, 270°) (b) frame 7, cam at (0.4m, 135°)

Fig. 12: Meshes from camera positions within acceptable regions. Areas visible by the cams are highlighted with pink. The high quality of the meshes is maintained

yields conclusive results based on quantitative data regarding the number and pose of the cameras needed, the Camera Positioning and Assignment (CPA) process can be broadened and made less restrictive. This can be achieved by extending the accepted CPA area to include the green regions. Fig. 12 presents meshes produced from cameras placed in alternate positions but within these regions and, thus, display similar quality to those of Fig. 10.

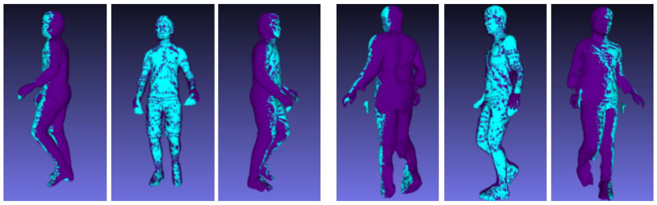
Conversely, some regions produce meshes with consistently low quality and, therefore, should be avoided. Examples in Fig. 11 highlighted with red are LR-A, PS-F, TS-C, and SW-D. Fig. 13 shows mesh examples from cameras placed within these regions that display unnatural joint placement and deformed external geometries.

Contrary to the camera positioning, the camera number displays less flexibility. Decreasing the camera number significantly reduces the versatility of the CPA process, even within the accepted regions, resulting in poor mesh quality for specific frames, as presented in Fig. 14.



(a) frame 4, cam at (0.4m, 255°) (b) frame 7, cam at (0.2m, 75°)

Fig. 13: Meshes from cam positions within non-acceptable regions. Areas visible by the cameras are highlighted with yellow. A clear mesh quality degradation is displayed.



(a) frame 4, cam at (0.4m, 15°) (b) frame 7, cam at (0.2m, 60°)

Fig. 14: Meshes produced by a three-camera set. Areas visible by the cameras are highlighted in blue. A drop in mesh quality is observed.

C. Footprint Area Plots

Importing these meshes to the GAMA 3D add-on calculates the area of each foot's footprint. Fig. 15 presents the contact area of the left and right foot during one gait cycle. Changes in the contact area indicate phase-changing events and can, therefore, be utilized for phase detection, an important parameter of the gait.

D. Joint Angle Plots

Among the parameters with exceptional scientific importance to the gait analysis community are the joint angles of (a) the hip, (b) the knee, and (c) the ankle. These joint angles are derived using specific landmarks of the leg and foot [31]. Fig. 16 presents the joint trajectories utilizing the keypoints provided by the SMPL model compared to average joint trajectories provided by [32]. The experimental results are within the context of the population in the database. Note that the plots for the right leg and foot have been offset to align the corresponding IC phase with 0 % of the gait cycle.

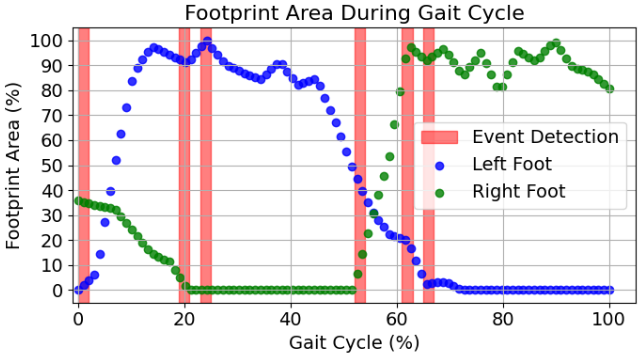


Fig. 15: The percentage of the footprint in contact with the ground. The phase-changing events are highlighted in red.

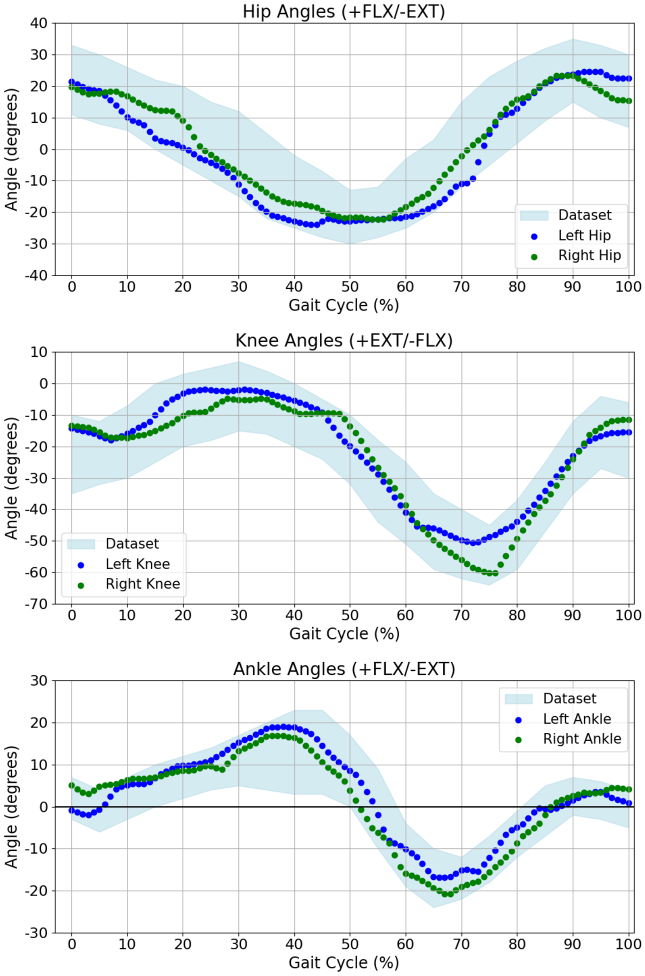


Fig. 16: Joint trajectories of the left and right legs as obtained from our framework, compared to database averages obtained from [32].

VI. Conclusion

This paper presented a novel approach to streamline the camera selection process of gait analysis using DL-driven avatar reconstruction, which can greatly impact the quality of the produced human meshes as presented in our previous work [14]. Demonstrated by the simulated and real-world experiments, our approach: (a) determines the optimal number and position of cameras for (but not limited to) one avatar reconstruction algorithm, (b) evaluates and refines the meshes in cases of poor quality, and (c) combines the results into a virtual animation and extracts gait parameters. Future research directions include the integration of more state-of-the-art ARAs or developing a gait-centric ARA from the ground up. The GAMA 3D add-on could be further enhanced with additional gait analysis functionalities, such as calculating cadence, stride length, step width, etc. Lastly, integrating the work in [33] into the proposed framework may provide a complete and comprehensive tool to better facilitate the analysis of the human gait.

References

[1] Alexander A. Hulleck, Dinesh Menoth Mohan, Nabil Abdallah, Mahmoud El Rich, and Karim Khalaf, "Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies," *Frontiers in Medical Technology*, vol. 4, pp. 901331, Dec. 2022.

[2] Nikita Prajapati, Amandeep Kaur, and Dimple Sethi, ," in *A Review on Clinical Gait Analysis*, 06 2021, pp. 967–974.

[3] Lokesh Arunagiri, "Gait analysis system using smart insoles to improve sports performance," *Karlos Sports science Physical Therapy*, 06 2024.

[4] Maryam Babaee, Linwei Li, and Gerhard Rigoll, "Gait recognition from incomplete gait cycle," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 768–772.

[5] Nathaniel Goldfarb, Alek Lewis, Alex Tacescu, and Gregory S. Fischer, "Open source vicon toolkit for motion capture and gait analysis," *Computer Methods and Programs in Biomedicine*, vol. 212, pp. 106414, 2021.

[6] Footbionics, "The gait cycle," `https://www.footbionics.com/For+Patients/gait_cycle.html`, Acc.:2024-06-12.

[7] "Compare Sports Tech - Compare Motion Capture Systems," `https://www.comparesportstech.com/compare-motion-capture-systems`, Accessed: 2024-06-16.

[8] THEIA Markerless, "Theia 3d markerless motion capture system," `https://www.theiamarkerless.ca`, Accessed: 2024-06-12.

[9] Move AI, "What is markerless motion capture?," `https://www.move.ai/blog-posts/what-is-markerless-motion-capture`, 2024, Accessed: 2024-06-21.

[10] Abdullah S. Alharthi, Syed U. Yunas, and Krikor B. Ozanyan, "Deep learning for monitoring of human gait: A review," *IEEE Sensors Journal*, vol. 19, no. 21, pp. 9575–9591, 2019.

[11] Yuliang Xiu, Jinlong Yang, Xu Cao, Dimitrios Tzionas, and Michael J. Black, "ECON: Explicit Clothed humans Optimized via Normal integration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023.

[12] Chen Guo, Tianjian Jiang, Xu Chen, Jie Song, and Otmar Hilliges, "Vid2avatar: 3d avatar reconstruction from videos in the wild via self-supervised scene decomposition," in *Proceedings of the IEEE/CVF Conference on Computer Vision & Pattern Recognition (CVPR)*, 2023.

[13] Xihe Yang, Xingyu Chen, Daiheng Gao, Shaohui Wang, Xiaoguang Han, and Baoyuan Wang, "Have-fun: Human avatar reconstruction from few-shot unconstrained images," *Proceedings of the IEEE/CVF Conference on Computer Vision & Pattern Recognition (CVPR)*, 2024.

[14] Odysseas Stavrakakis, Athanasios Mastrogeorgiou, Aikaterini Smyrli, and Evangelos Papadopoulos, "Gait analysis with trinocular computer vision using deep learning," in *2023 IEEE International Conference on Image Processing Challenges and Workshops (ICIPCW)*, 2023, pp. 3702–3706.

[15] Chao Fan, Junhao Liang, Chuanfu Shen, Saihui Hou, Yongzhen Huang, and Shiqi Yu, "Opengait: Revisiting gait recognition towards better practicality," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 9707–9716.

[16] Reed D. Gurchiek, Rebecca H. Choquette, Bruce D. Beynnon, James R. Slauterbeck, Timothy W. Tourville, Michael J. Toth, and Ryan S. McGinnis, "Remote gait analysis using wearable sensors detects asymmetric gait patterns in patients recovering from acl reconstruction," in *2019 IEEE 16th International Conference on Wearable & Implantable Body Sensor Networks (BSN)*, 2019, pp. 1–4.

[17] Prof. Jaychand Upadhyay, Prof. Tad Gonsalves, Rohan Paranjpe, Hiralal Purohit, and Rohan Joshi, "Biometric identification using gait analysis by deep learning," in *2020 IEEE International Conference for Innovation in Technology (INOCON)*, 2020, pp. 1–4.

[18] Blender Foundation, "Blender - a 3d modelling and rendering package," 2024.

[19] Adobe Systems Incorporated, "Adobe mixamo," 2024.

[20] Physiopedia, "The gait cycle," `https://www.physio-pedia.com/The_Gait_Cycle#cite_ref-Vaughan_1-0`, Acc.: 2024-06-12.

[21] Ling Guo, Akil Narayan, and Tao Zhou, "Constructing least-squares polynomial approximations," *SIAM Review*, vol. 62, pp. 483–508, May 2020.

[22] Maria Ines Varela-Silva and Barry Bogin, *Leg Length and Anthropometric Applications: Effects on Health and Disease*, pp. 769–783, Handbook of Anthropometry: Physical Measures of Human Form in Health and Disease, 01 2012.

[23] Prabodh Acharya, "Online anthropometry: Measurements for product design - chapter 2.5," 2024, Accessed: 2024-06-30.

[24] Alexei Botchkarev, "A new typology design of performance metrics to measure errors in machine learning regression algorithms," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 14, pp. 045–076, 2019.

[25] Cheng-Hao Lu and Xiu-Hong Chen, "Improved iterative poisson point cloud surface reconstruction," in *2023 3rd International Conference on Digital Society and Intelligent Systems (DSInS)*, 2023, pp. 382–385.

[26] Philipp Glira, Norbert Pfeifer, Christian Briese, and Camillo Ressl, "A correspondence framework for als strip adjustments based on variants of the icp algorithm," *Photogrammetrie-Fernerkundung-Geoinformation*, vol. 2015, no. 4, pp. 275–289, 2015.

[27] Razer Inc., "Razer kiyo pro," 2024, Accessed: 2024-07-15.

[28] Zhengyou Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[29] OpenCV Contributors, "Camera calibration with opencv," Online documentation, 2024, Accessed: 2024-06-30.

[30] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black, "Expressive body capture: 3D hands, face, and body from a single image," in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 10975–10985.

[31] Selin Fandakli, Halil Okumuş, and Mehmet OZTURK, ," in *A Study of Human Walking Biomechanics for Ankle-Foot Prosthesis Design*, 10 2018, pp. 1–5.

[32] Carlos A. Fukuchi et al., "A public dataset of overground and treadmill walking kinematics and kinetics in healthy individuals," *PeerJ*, 2018.

[33] Konstantina Tsintzira, "Kinematic and dynamic study of human gait through image analysis," M.S. thesis, NTUA, 2024.