Out-of-Clinic Measurement of Time-Domain Gait Parameters From Low-Frame-Rate Videos

Yuliang Fu
North Carolina State University
Raleigh, NC, USA
yfu34@ncsu.edu

Yanxin Luo
Independent Researcher
Durham, NC, USA
luoyx28@hotmail.com

Sidy Ndiaye
North Carolina State University
Raleigh, NC, USA
sndiaye@ncsu.edu

Chenhan Xu
North Carolina State University
Raleigh, NC, USA
cxu34@ncsu.edu

Abstract

Time-domain gait parameters provide important indicators for clinical assessments, but their measurement often requires specialized in-clinic devices. Video provides an accessible alternative for out-of-clinic gait monitoring, but its low frame rate limits its ability of precise gait-event timing. This paper presents a video-based method for measuring time-domain gait parameters by performing sub-frame gait-event localization from human pose sequences. To effectively model sub-frame gait transitions, we employ velocity augmentation for poses and a spatial-temporal encoder with a multi-head decoder. Evaluation on the UnderPressure dataset shows that the system maintains a high ten-ms-level accuracy across stride, step, stance, and swing time with 30 fps input, and is robust against reduced frame rates. These results suggest that our method overcomes the limitation of video frame rate in time-domain gait parameter measurements.

CCS Concepts

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Keywords

Gait parameters, Biomedical sensing, Video processing

ACM Reference Format:
Yuliang Fu, Yanxin Luo, Sidy Ndiaye, and Chenhan Xu. 2026. Out-of-Clinic Measurement of Time-Domain Gait Parameters From Low-Frame-Rate Videos. In *The 4th International Workshop on Human-Centered Sensing, Modeling, and Intelligent Systems (HumanSys '26), October 26–30, 2026, Austin, TX, USA*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3842436.3843790



This work is licensed under a Creative Commons Attribution 4.0 International License.
HumanSys '26, Austin, TX, USA
© 2026 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/10.1145/3842436.3843790

1 INTRODUCTION

Human gait provides a clinically informative window into mobility, neurological function, rehabilitation progress, and disease progression. Spatiotemporal gait parameters are widely used in clinical assessment, including post-stroke rehabilitation, Parkinson's disease assessment, fall-risk evaluation, and general mobility monitoring [2].

Despite their clinical value, quantitative gait measurements are difficult to obtain in everyday settings. Standard gait analysis relies on instrumented walkways or pressure mats, which typically require dedicated equipment, trained operators, and in-person clinic visits [6]. As a result, gait assessment is often limited to sparse snapshots collected during scheduled appointments, bringing challenges for continuous symptom monitoring. Enabling accurate out-of-clinic gait measurement could make gait monitoring more accessible and longitudinal, allowing functional changes to be observed in everyday environments.

Recent advances in sensing and machine learning have motivated alternative approaches for more accessible gait assessment. In particular, video is attractive for being contactless, inexpensive, and increasingly available through smart devices. Prior video-based gait analysis has demonstrated encouraging performance for parameters such as step length, gait speed, pressure distribution, and other spatial measures [3, 7]. However, temporal parameters that depend on precise gait-event timing remain more challenging. A key practical barrier is the temporal resolution of commodity video. Clinical reference systems such as pressure mats commonly sample at 100 Hz or higher. In contrast, most everyday videos are recorded at around 30 frames per second (FPS). This sampling rate gap is consequential for parameters such as swing time and stride time, which require accurate identification of foot contact and foot-off events. Prior work has noted that

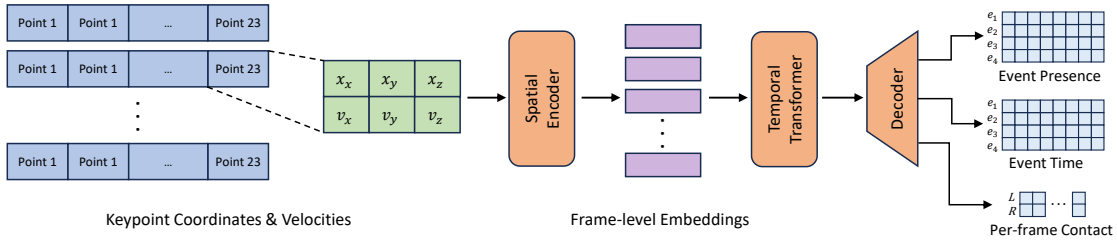


Figure 1: Sub-frame level event detection model.

low frame rates make time-domain parameters particularly difficult to estimate from pose-based video analysis [1]. Thus, while video-based gait analysis is promising for scalable out-of-clinic monitoring, its use for fine-grained time-domain gait measurement remains underexplored.

In this work, we aim to accurately measure time-domain gait parameters from low-frame-rate videos. We focus on five clinically meaningful gait parameters: stride, step, stance, swing, and double support time. To address the mismatch between commodity video frame rates and the timing precision required for gait-event detection, *we propose a video-based gait-event localization model that augments pose sequences with velocity cues.* The model predicts foot-contact and foot-off events at sub-frame resolution, from which we further compute the temporal gait parameters. Evaluation on the UnderPressure dataset shows ten-ms-level accuracy across diverse activities and robustness to reduced frame rates.

2 METHODS

2.1 Task Formulation

We formulate out-of-clinic time-domain gait measurement as a gait-event timing problem from short video-derived pose sequences. Given a sliding one-second window of 3D human pose, our goal is to estimate the sub-frame timing of four fundamental gait events: left-foot contact, right-foot contact, left-foot departure, and right-foot departure. These events are then used to compute temporal gait parameters.

Unlike frame-level classification, accurate gait-parameter measurement requires estimating event times at a finer temporal resolution than the input video frame rate. Therefore, our method predicts normalized event times within each input window rather than only assigning events to discrete video frames. A practical challenge is that multiple events of the same type may occur within a single window, especially during faster movements. Therefore, instead of predicting only one timestamp for each event class, we use a set-based output formulation. For each of the four event classes, the model predicts multiple candidate event slots, each of which contains a normalized event-time estimate. This formulation allows the model to represent a variable number of gait events within a fixed-length input window.

2.2 Velocity-Based Pose Augmentation

The input to the model is a sequence of body poses represented by 23 3D joint coordinates derived from recorded videos. To expose motion cues that are critical for identifying foot contact transitions, we augment the pose sequence with finite-difference velocity features. For each joint, the velocity at time t is computed as $v_t = x_t - x_{t-1}$ and concatenated with the original joint features. This gives the model direct access to short-term kinematic changes around foot contact and foot-off transitions, supporting sub-frame event localization.

2.3 Model Architecture

The overall structure of our model is shown in Fig. 1. To model body topology embedded in the keypoint input, we first employ a pose encoder consisting of a graph convolutional network (GCN) embedding module to map each velocity-augmented pose frame into a latent representation. Such a representation facilitates following temporal processing, while preserving motion information from the original joints. Following the GCN, a temporal encoder then processes the embedded pose sequence. Consistent with our post-hoc processing nature, we allow the encoder to use bidirectional temporal context and apply attention to the full input window. This enables the model to reason about gait-event transitions based on the surrounding motion trajectory.

Finally, the encoded sequence is decoded by a set prediction head based on MLPs. The head outputs event-presence flags and normalized event times for all event classes, and allows eight events to be detected for each class. In parallel, an auxiliary contact detection head predicts the left and right foot contact state at each frame from the shared temporal representation. The auxiliary head encourages the encoder to capture contact-related dynamics, which directly underlie the target contact and foot-off events. Moreover, while event-time labels are sparse and occur only at contact or departure instants, contact-state labels are available for every frame and both feet. Therefore, this dense supervision helps anchor the temporal representation to foot contact dynamics and complements the sparse event-localization objective.

2.4 Gait Event-informed Loss Function

We train the model with a weighted multi-task loss combining event-time regression, event-presence classification, and auxiliary contact-state prediction:

$$\mathcal{L} = \lambda_t \mathcal{L}_t + \lambda_p \mathcal{L}_p + \lambda_c \mathcal{L}_c.$$

The event-time loss $\mathcal{L}_t$ is the mean absolute error of the event timings. The event-presence loss $\mathcal{L}_p$ is a binary cross-entropy loss over all event slots. The auxiliary contact loss $\mathcal{L}_c$ is per-frame contact-state cross-entropy for both feet. Because our set-output formulation does not assign fixed semantic identities to event slots, predictions and ground-truth events are matched by temporal order within each event class. $\mathcal{L}_t$ is applied only to matched events in both ground truth and the output, with extra or empty slots supervised only by the presence loss.

2.5 Gait Parameter Computation

After event prediction, temporal gait parameters are computed from the estimated contact and departure times of the two feet. Stance time is measured as the duration between foot contact and the subsequent departure of the same foot, while swing time is measured as the duration between foot departure and the next contact of the same foot. Step time is computed between alternating foot-contact events, and stride time is computed between consecutive contacts of the same foot. Double support time is derived from the overlap structure of the left and right foot contact. In this way, the model converts low-frame-rate video pose sequences into clinically meaningful time-domain gait parameters.

3 EVALUATION

3.1 Dataset and Evaluation Setup

We evaluate our method on the public UnderPressure dataset, which provides motion data synchronized with foot pressure measurements for foot contact analysis [5]. The dataset consists of data from 10 subjects, with human poses captured with MoCap cameras at 240 FPS, and foot pressure distributions sampled with pressure insoles at 100 Hz. To align with our gait event detection objective, we convert the synchronized contact signals into gait-event annotations by detecting the zero points and rising/falling edges of the pressure readings. To represent videos collected from common commercial devices, the human poses are down-sampled to 30 FPS before being fed into our processing pipeline.

We use six categories of locomotion activities from the dataset: walking front and backward, jogging front and backward, stair climbing, and jogging sidesteps. These activities cover common gait activities with proven clinical significance while also introducing variation in motion speed and foot-ground interaction patterns [4, 8]. It's notable that double support time is not applied to jogging since it is the defining feature for walking. Therefore, this activity set provides a comprehensive test of whether the model can recover reliable gait parameters under diverse temporal dynamics.

To find proper weights λ for our loss function, we perform a grid search on a four-subject subset with one subject left out for validation. A combination of $\lambda_t = 1, \lambda_p = 0.25, \lambda_c = 0.1$ reaches the lowest event time MAE and is used for all evaluation trainings. To reflect practical usage with unseen users, we perform all evaluations with a leave-one-subject-out setup and report the metrics on a 10-subject average.

3.2 Baselines

We compare our method with two representative baselines, UnderPressure [5] and FootFormer [3]. The two methods reflect state-of-the-art approaches for estimating human gait parameters from visual input, including foot pressure and contact flags at the frame level. To compare them with our event-timing formulation, we convert their predicted contact sequences into gait events using transition detection, and the same event-to-parameter conversion procedure is applied to ensure a fair comparison.

3.3 Results

Overall Accuracy Across Gait Parameters. We compare the five time-domain gait parameters derived from our method with the baselines in Fig. 2. Across all evaluated parameters, our method achieves lower error than both baselines. Specifically, our method reaches MAE around 10 ms for stride, step, stance, and swing times, with step time reaching the lowest MAE of 6.7 ms. Double support time shows a larger MAE, primarily because its calculation relies on all four gait events, instead of two for other indicators. With the input having approximately 33 ms frame interval, these results suggest that the proposed event-timing formulation can recover fine-grained temporal gait information beyond the nominal video frame resolution.

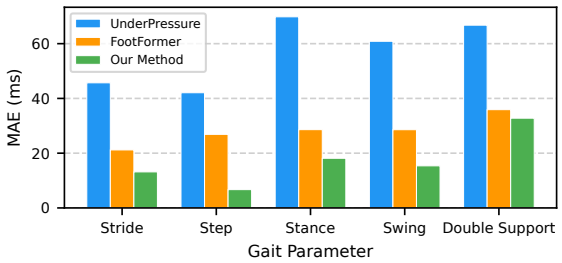


Figure 2: Gait parameter MAEs with different methods.

To see how our method perform on different gait activities, we compare its performance in Fig. 3. Among all activities, stair climbing brings the biggest challenges, showing

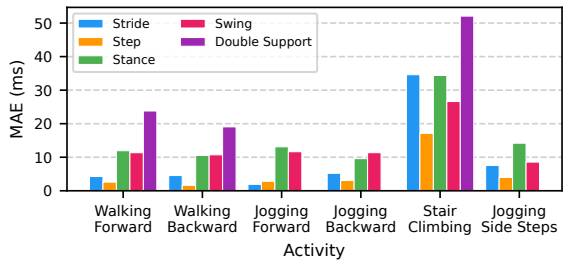


Figure 3: System error across different activities.

slightly higher errors across gait parameters. This is expected because foot-ground contact during stair motion is more variable: the two feet interact with surfaces at different heights, and contact/departure events can involve more complex transitions than level-ground locomotion. In all other activities, our approach maintains low MAE regardless of motion speed and directions, showing that the learned temporal representation remains effective on diverse foot-contact dynamics.

Robustness to Input Frame Rate. To investigate how input frame rate affects our gait-parameter estimation system, we vary the input pose frame rate among 20, 25, and 30 FPS and evaluate the resulting gait-parameter errors using our approach. As shown in Fig. 4, the performance of gait parameter prediction remains relatively consistent across the tested frame rates, even though reducing the input frame rate increases the time interval between observed frames. This result indicates that the model successfully learns temporal motion patterns within the pose window and estimates event timing at a sub-frame resolution.

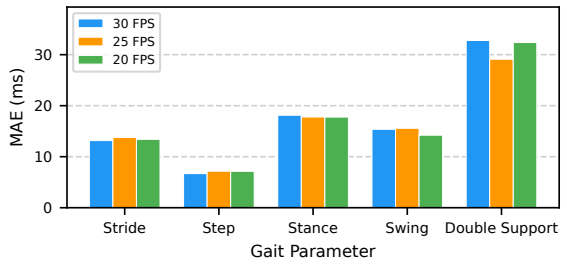


Figure 4: System error under reduced input FPS.

Such compatibility with lower frame rates suggests high practical usability with a wider range of consumer cameras and mobile devices, including settings where video quality or frame rate may be limited. In addition, reliable performance at reduced frame rates can also lower computational and storage requirements, making the approach more suitable for scalable out-of-clinic gait monitoring.

4 DISCUSSION AND CONCLUSION

In this paper, we presented a video-based method for estimating time-domain gait parameters through sub-frame localization of foot contact and foot-off events. On UnderPressure, the method outperforms contact-detection baselines, achieves ten-ms-level accuracy for most parameters, and remains robust at reduced input frame rates. These findings suggest that low-frame-rate video can support accessible out-of-clinic gait assessment.

Despite the system's promising performance, several limitations remain. Stair climbing is harder because foot-ground interactions vary with terrain height; incorporating terrain cues from video may improve robustness. The current system also focuses on contact and foot-off events, while richer gait analysis may require additional events such as heel strike, toe-off, mid-swing, and foot-flat. Extending the event set is a promising direction for future work.

Acknowledgments

This work was supported by the National Science Foundation under grant No. 2625745.

References

[1] Chang Soon Tony Hii, Kok Beng Gan, Nasharuddin Zainal, Norlinah Mohamed Ibrahim, Shahrul Azmin, Siti Hajar Mat Desa, Bart van de Warrenburg, and Huay Woon You. 2023. Automated gait analysis based on a marker-free pose estimation model. *Sensors* 23, 14 (2023), 6489.

[2] Abdul Aziz Hulleck, Dhanya Menoth Mohan, Nada Abdallah, Marwan El Rich, and Kinda Khalaf. 2022. Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies. *Frontiers in medical technology* 4 (2022), 901331.

[3] Keaton Yukio Kraiger, Jingjing Li, Skanda Bharadwaj, Jesse Scott, Robert T. Collins, and Yanxi Liu. 2025. FootFormer: Estimating Stability from Visual Input. In *36th British Machine Vision Conference 2025, BMVC 2025, Sheffield, UK, November 24-27, 2025*. BMVA.

[4] Tracy Milane, Nicolas Vuillerme, Pascal Petit, Elke Warmerdam, Robbin Romijnders, Edoardo Bianchini, Walter Maetzler, and Clint Hansen. 2025. Impact of forward and backward walking on gait parameters across parkinson's disease stages and severity: a prospective observational study. *BMC neurology* 25, 1 (2025), 379.

[5] Lucas Mourot, Ludovic Hoyet, François Le Clerc, and Pierre Hellier. 2022. UnderPressure: Deep Learning for Foot Contact Detection, Ground Reaction Force Estimation and Footskate Cleanup. *Computer Graphics Forum* 41, 8 (Dec. 2022), 195–206. doi:10.1111/cgf.14635

[6] Monica Parati, Emilia Ambrosini, Beatrice De Maria, Matteo Gallotta, Laura A Dalla Vecchia, Giorgio Ferriero, and Simona Ferrante. 2022. The reliability of gait parameters captured via instrumented walkways: A systematic review and meta-analysis. *European Journal of Physical and Rehabilitation Medicine* 58, 3 (2022), 363.

[7] Jan Stenum, Melody M Hsu, Alexander Y Pantelyat, and Ryan T Roemmich. 2024. Clinical gait analysis using video-based pose estimation: Multiple perspectives, clinical populations, and measuring change. *PLOS Digital Health* 3, 3 (2024), e0000467.

[8] Srikant Vallabhajosula, Chi Wei Tan, Mukul Mukherjee, Austin J Davidson, and Nicholas Stergiou. 2015. Biomechanical analyses of stair-climbing while dual-tasking. *Journal of biomechanics* 48, 6 (2015), 921–929.