Occlusion-Aware Human Mesh Model-Based Gait Recognition

Dr Kiran Macwan[1] , Dr. Ajay Kumar[2]

[1] LINCOLN University [2] LINCOLN University

pdf.Kirankumari@lincoln.edu.my, ajay.phdcse@gmail.com

1. Problem Background and Motivation

Earlier approaches to occluded gait recognition can broadly be divided into reconstruction-based and reconstruction-free methods. Reconstruction-based techniques attempt to recover an unobstructed silhouette or gait representation before performing feature extraction and matching. Reconstruction-free techniques instead learn features directly from incomplete observations or compare only regions that remain visible in both samples. Although these strategies can be useful, many depend on a full-body bounding box for cropping, scale normalization, and body-center registration. This assumption becomes unrealistic when the complete person cannot be observed.

A model-based strategy offers a more natural alternative. By fitting a human model to the visible evidence, the system can estimate body pose, body shape, and the location of the complete body even when some regions are hidden. The proposed work extends this idea from individual images to continuous gait sequences. Temporal information is important because the visible region can change from frame to frame, and a sequence provides complementary evidence about the person's movement.

Keywords : Occluded Gait Recognition, Reconstruction-Free Methods, Human Body Modeling, Temporal Gait Analysis, Pose and Shape Estimation

2. Proposed Framework

For every frame, the visible portion of the walking person is cropped using a square bounding box and resized while preserving its aspect ratio. A sequence encoder then estimates human-mesh parameters together with global rotation and camera parameters. Instead of treating each frame independently, the encoder incorporates temporal information through a bidirectional gated recurrent unit (BiGRU). This allows information from both earlier and later frames to contribute to the estimation of the current body pose and shape.

The SMPL representation describes body shape using ten parameters and body pose using sixty-nine parameters corresponding to the rotations of 23 joints. The model produces a triangulated human mesh with 6,980 vertices, from which three-dimensional joint locations can be calculated. A weak-perspective camera model is used to project the estimated three-dimensional body into the image plane. The sequence encoder extracts frame-level features with ResNet-50, processes them with a BiGRU, and uses a regressor with iterative feedback to estimate the model parameters.

3. Phase Synchronization and Occlusion Attenuation

A major challenge is that different occlusion patterns can cause different model-fitting errors even for the same person. To reduce this intra-person variation, the framework first estimates the gait phase of every frame. The phase is represented as a cyclic two-dimensional sine-cosine vector. A convolutional network followed by temporal processing predicts the phase sequence, while estimation, smoothness, and ordering constraints encourage the predicted phases to remain accurate and temporally consistent.

After phase estimation, the model parameters are synchronized to a canonical gait cycle using interpolation. This places corresponding walking stances into comparable temporal positions. The synchronized shape and pose parameters are then passed to an occlusion attenuation transformation. Importantly, the camera scale and translation parameters provide clues about the location and extent of the hidden body region. An attenuation network uses these camera cues together with the initial shape and pose parameters to learn updates that make the resulting representation less dependent on the particular occlusion pattern.

A paired similarity loss encourages representations from different occlusion conditions of the same individual to become closer. A sequential constraint also promotes temporal consistency and a stable body shape throughout a gait sequence.

4. Recognition Strategy and Training

After attenuation, the framework constructs identity features from both body shape and body pose. The unified shape parameters are used directly as a shape representation. Three-dimensional joint trajectories derived from the estimated mesh are passed through a convolutional network to obtain discriminative pose features. The two feature types can be evaluated separately or combined through score-level fusion.

The complete system is optimized end-to-end using a unified objective. The loss combines gait-phase estimation, paired similarity, sequential consistency, SMPL and joint estimation, and recognition objectives. Triplet-based recognition learning encourages genuine samples to have smaller feature distances than samples from different identities. This joint optimization seeks a balance between accurate human-model estimation and effective identity discrimination.

5. Experimental Methodology

Because there is no publicly available gait dataset designed specifically for systematic RGB occlusion experiments, the study creates simulated occlusion samples from the OU-MVLP dataset. OU-MVLP contains 10,307 subjects captured from 14 views. Four views—0°, 30°, 60°, and 90°—are selected for the main experiments. The training and testing populations are disjoint, with 5,153 subjects used for training and 5,154 for testing.

The experiments concentrate on vertical occlusion, which can make the person's full height unavailable. Four patterns are created: fixed top occlusion (FT), fixed bottom occlusion (FB), changing top occlusion (CT), and changing bottom occlusion (CB). Fixed cases use occlusion ratios of 20%, 40%, and 60%. Changing cases simulate occlusion that varies throughout a gait cycle. The visible regions are cropped as they would be after pedestrian detection in a real scene.

6. Implementation and Evaluation

The cropped inputs are resized to 224 × 224 pixels, and 25 consecutive frames are used as one sequence, approximately covering one gait cycle for most subjects. The feature extractor and regressor are initialized using a pretrained partial-HMR model. The system is trained with Adam using a batch structure of eight subjects and eight samples per subject. Recognition is evaluated using rank-1 identification accuracy and equal error rate (EER) for identification and verification tasks.

The experiments compare the proposed approach with GaitSet, GaitGL, and ModelGait. The results show that ModelGait, although effective on complete images, loses performance when the input is heavily occluded. The proposed framework performs substantially better because it explicitly models temporal information and reduces the dependency of model fitting on the occlusion pattern. Combining shape and pose features provides the strongest overall recognition performance.

7. Main Findings

The reported results indicate an average improvement of approximately 15 percentage points in rank-1 identification and about 2 percentage points in EER compared with the state-of-the-art comparison methods in the evaluated occlusion setting. The analysis also shows that changing-occlusion sequences can be easier than fixed severe occlusion because they contain frames with smaller hidden regions. Bottom-body occlusion generally causes a larger performance reduction than top-body occlusion, emphasizing the importance of lower-body information for gait recognition.

Cross-occlusion experiments further demonstrate that the proposed representation remains more effective when the probe and gallery samples have different occlusion patterns. Recognition is strongest when the patterns are similar, while substantial differences in occlusion ratio and camera view reduce performance. Nevertheless, the model consistently outperforms the comparison methods across the evaluated cross-occlusion cases.

8. Ablation, Robustness, and Limitations

Ablation experiments confirm that the major components contribute to the final system. Removing the occlusion attenuation module reduces the average rank-1 identification rate from 41.2% to 32.5% in the reported pose-feature experiment. The GRU, phase synchronization, and attenuation components therefore play important roles in maintaining useful representations under occlusion.

The study also examines color and texture effects. When clothing colors are manually changed, the estimated human meshes remain broadly similar, showing that the model representation is comparatively insensitive to color and texture. Grad-CAM analysis indicates that the network can use visible legs, joints, stride information, shoulders, torso structure, and other available body cues to infer hidden regions.

The method is not without limitations. Severe fixed occlusion, especially around 60%, can still produce errors in estimated body pose and stride. The authors report examples where lower-body stride is underestimated or the upper body is predicted as too straight. Horizontal occlusion can also depend on reliable tracking, and phase synchronization may be difficult when substantially less than half of a gait cycle is visible.

9. Conclusion

The paper presents a model-based solution for gait recognition when the human body is only partially visible. Rather than depending on an unavailable full-body bounding box, the framework estimates an SMPL human mesh directly from visible regions and exploits temporal information across the gait sequence. Phase estimation and synchronization align comparable walking stances, while the occlusion attenuation module uses camera information to reduce representation changes caused by different hidden regions.

Overall, the study demonstrates that combining human shape, pose, temporal modeling, and occlusion-aware attenuation can make gait recognition considerably more robust to partial visibility. The reported experiments on simulated OU-MVLP occlusions show clear advantages over established appearance-based and model-based baselines. The approach is therefore relevant to practical surveillance and forensic scenarios in which complete observations of a walking person cannot be guaranteed.

References

[1] C. Fan, Y. Peng, C. Cao, X. Liu, S. Wang, Y. Huang, and J. Wang, “GaitSet: Regarding gait as a set for cross-view gait recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 8, pp. 2781-2797, Aug. 2021, doi: 10.1109/TPAMI.2020.2983804.

[2] Y. Fan, X. Liu, C. Cao, Y. Huang, and J. Wang, "GaitPart: Temporal part-based model for gait recognition," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 14225-14234, doi: 10.1109/CVPR42600.2020.01424.

[3] Z. Lin, Z. Wang, Z. Liu, and Y. Huang, "GaitGL: Global and local representation learning for gait recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 6, pp. 3579-3591, Jun. 2022, doi: 10.1109/TCSVT.2021.3113918.

[4] X. Li, C. Xu, X. Zhu, and J. Sun, "Context-aware spatiotemporal learning for gait recognition," IEEE Transactions on Image Processing, vol. 31, pp. 5643-5656, 2022, doi: 10.1109/TIP.2022.3199345.

[5] H. Xu, Y. Wang, and W. Ouyang, "Gait recognition via disentangled representation learning," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 10034-10043, doi: 10.1109/CVPR42600.2020.01005.

[6] M. Z. Uddin, "A survey on gait recognition under occlusion," Journal of Visual Communication and Image Representation, vol. 75, p. 103036, Mar. 2021, doi: 10.1016/j.jvcir.2021.103036.

[7] Z. Wang, Z. Liu, Z. Lin, and Y. Huang, "GaitAdapt: A continual learning framework for evolving gait recognition," IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 5, no. 1, pp. 119-131, Jan. 2023, doi: 10.1109/TBIOM.2022.3218501.

[8] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view and pace on gait recognition," in Proc. 18th Int. Conf. Pattern Recognit. (ICPR), 2006, vol. 4, pp. 441-444, doi: 10.1109/ICPR.2006.102.

Y. Makihara, R. Sagawa, Y. Miki, T. Echigo, and Y. Yagi, "The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition, IEEE Transactions on Information Forensics and Security, vol. 7, no. 5, pp. 1587-1598, Oct. 2012, doi: 10.1109/TIFS.2012.2204321.

[9] Y. Wu, Y. Li, and Y. Fu, "Deep learning for gait recognition: A survey," ACM Computing Surveys, vol. 54, no. 8, pp. 1-37, Oct. 2021, doi: 10.1145⁄3465255.

[10] S. Hou, X. Bai, and Z. Wang, "Gait recognition with flexible temporal resolution," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 12, pp. 9459-9475, Dec. 2022, doi: 10.1109/TPAMI.2021.3130873.

[11] J. Zhang, Y. Huang, C. Cao, and J. Wang, "Learning effective gait features with 3D convolutional networks," IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 9, pp. 3013-3025, Sep. 2020, doi: 10.1109/TCSVT.2019.2931228.

[12] T. He, Z. Liu, Z. Lin, and Y. Huang, "Multi-scale part-based representation for gait recognition," IEEE Transactions on Multimedia, vol. 23, pp. 2386-2397, 2021, doi: 10.1109/TMM.2020.3010531.

[13] H. Zhao, X. Xi, and W. Ouyang, "Appearance-based gait recognition via a deep long short-term memory network," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 2933-2945, 2020, doi: 10.1109/TIFS.2020.2978993.

[14] Z. Liu, Z. Wang, Z. Lin, and Y. Huang, “GaitRef: Refining sequential skeleton-based gait recognition,” IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 5, no. 2, pp. 249-261, Apr. 2023, doi: 10.1109/TBIOM.2023.3242456.

[15] Z. Wang, Z. Liu, Z. Lin, and Y. Huang, “Learning to learn gait: A continual learning perspective,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2021, pp. 12437-12446, doi: 10.1109/ICCV48922.2021.01222.

[16] S. Teepe, T. S. Beenders, and M. T. V. D. Heuvel, “GaitViT: A Vision Transformer for gait recognition,” in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV), 2022, pp. 1534-1543, doi: 10.1109/WACV51458.2022.00159.

[17] Y. Li, Y. Wu, and Y. Fu, “Occlusion-robust gait recognition by adversarial learning,” IEEE Transactions on Image Processing, vol. 29, pp. 6715-6728, 2020, doi: 10.1109/TIP.2020.2996841.

[18] Z. Zhang, L. Wang, and W. Ouyang, “Gait recognition with attentive spatial-temporal feature learning,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 10, pp. 3969-3981, Oct. 2021, doi: 10.1109/TCSVT.2020.3045831.

[19] A. R. J. Al-Kaf, A. A. A. El-Sawy, and M. A. A. El-Sayed, “A survey of gait recognition for healthcare applications,” Journal of Medical Systems, vol. 44, no. 5, p. 96, Apr. 2020, doi: 10.1007/s10916-020-01563-y.

[20] S. G. Larsen, M. De Zee, and J. B. Simonsen, “The effect of aging on gait: A systematic review and meta-analysis,” Journal of Gerontology: Series A, vol. 75, no. 9,
pp. 1754-1763, Sep. 2020, doi: 10.1093/gerona/glaa068.

[21] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, “Continual lifelong learning with neural networks: A review,” Neural Networks, vol. 113, pp. 54-71, May 2019, doi: 10.1016/j.neunet.2019.01.012.

[22] Y. Huang, J. Wang, and C. Cao, “Cross-view gait recognition by feature fusion,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 10, pp. 2544-2557, Oct. 2020, doi: 10.1109/TPAMI.2019.2918485.

[23] X. Zhu, Y. Huang, C. Cao, and J. Wang, “View-invariant gait recognition with attentive multi-biometric fusion,” IEEE Transactions on Information Forensics and Security, vol. 16, pp. 278-291, 2021, doi: 10.1109/TIFS.2020.3009641.

[24] Z. Wu, Y. Huang, C. Cao, and J. Wang, “A comprehensive study on gait biometrics,”
ACM Computing Surveys, vol. 53, no. 6, pp. 1-36, Dec. 2020, doi: 10.1145⁄3417981.

[25] C. Xu, X. Li, X. Zhu, and J. Sun, “Gait recognition with disentangled spatial and temporal representations,” IEEE Transactions on Multimedia, vol. 24, pp. 2565-2577, 2022, doi: 10.1109/TMM.2021.3083248.

[26] J. Liu, Y. Huang, C. Cao, and J. Wang, “Memory-augmented graph neural networks for gait recognition,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021, pp. 11877-11886, doi:

10.1109/CVPR46437.2021.01170.

[27] W. Wang, Y. Huang, C. Cao, and J. Wang, “Learning to estimate 3D human pose and shape from a single image,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 9, pp. 3148-3163, Sep. 2021, doi: 10.1109/TPAMI.2020.2994480.

[28] I. An, B. Kim, and H. J. Kim, “Gait recognition from a single image using a deep siamese network,” IEEE Access, vol. 8, pp. 115338-115349, 2020, doi: 10.1109/ACCESS.2020.3003781.

[29] S. Yu, H. Chen, and E. I. Chong, “GaitGANv2: A generative adversarial network for gait recognition with carrying and clothing variations,” Neurocomputing, vol. 437, pp. 124-135, May 2021, doi: 10.1016/j.neucom.2021.01.036.