

UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

**HUMAN GAIT RECOGNITION UNDER
CHANGES OF WALKING CONDITIONS**

by

Lingxiang Yao

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

Sydney, Australia

2021

Certificate of Authorship/Originality

I, Lingxiang Yao, declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Electrical and Data Engineering, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
Signature removed
prior to publication. _____

Date: **Dec/14/2021** _____

Acknowledgements

It is a cherished memory to pursue PhD in UTS. I would like to express my sincere gratitude to all those who help me complete my doctoral study.

Foremost, I would like to express my sincere gratitude to my principal supervisor, A/Prof. Jian Zhang, for his professional guidance and persistent support throughout my PhD study. His enthusiasm, attitude, and devotion towards academy have deeply influenced me, which provides instructions to my future career.

I would also like to appreciate my co-supervisor A/Prof. Qiang Wu for his advice and guidance in both research and lives in Sydney. He always encourages me to focus on the advanced techniques that motivate the research going deeper.

I also deeply appreciate my friend Worapan Kusakunniran in Mahidol University for his significant support. He always spends a lot of time to help me polish my paper, and gives me valuable advice to improve my academic skills.

And, I appreciate the help, support and friendship from my dear colleagues and friends during my doctoral study. Thanks to Jingsong Xu, Yucheng Wang, Xiaoshui Huang, Yazhou Yao, Peng Zhang, Junjie Zhang, Muming Zhao, Zhibin Li, Yongshun Gong, Lu Zhang, Huaxi Huang, Anan Du, Xunxiang Yao, Yan Huang, Guofeng Mei, and all other labmates for their collaboration and discussion. I am also grateful to all my friends in Sydney, Razia Osman, Yumeng Yao, Xin Ba, Jiancheng Li, and Xiao Li for their encouragement and companion.

Finally, I would like to express my sincere thanks to my parents and girlfriend Xia Zhang for their support, trust, encouragement and love throughout my PhD studies..

Lingxiang Yao
August 2021 @ UTS.

List of Publications

Published Papers

- C-1. **Lingxiang Yao**, Worapan Kusakunniran, Qiang Wu, Jian Zhang and Jingsong Xu. “Part-based Collaborative Spatio-temporal Feature Learning for Cloth-changing Gait Recognition.” 2020 25th International Conference on Pattern Recognition (ICPR) (2021): 2057-2064.
- J-1. **Lingxiang Yao**, Worapan Kusakunniran, Qiang Wu and Jian Zhang. “Gait recognition using a few gait frames.” PeerJ Computer Science 7 (2021).
- J-2. **Lingxiang Yao**, Worapan Kusakunniran, Qiang Wu, Jian Zhang, Zhen-min Tang and Wankou Yang. “Robust gait recognition using hybrid descriptors based on Skeleton Gait Energy Image.” Pattern Recognition Letters 150 (2021): 289-296.
- C-2. Junyi Wu, **Lingxiang Yao**, Y. Huang, J. Xu, Qiang Wu and Liqin Huang. “Improving Person Re-Identification Performance Using Body Mask Via Cross-Learning Strategy.” 2019 IEEE Visual Communications and Image Processing (VCIP) (2019): 1-4.
- C-3. Anan Du, Xiaoshui Huang, Jiayuan Zhang, **Lingxiang Yao** and Qiang Wu. “Kpsnet: Keypoint Detection and Feature Extraction for Point Cloud Registration.” 2019 IEEE International Conference on Image Processing (ICIP) (2019): 2576-2580.
- C-4. **Lingxiang Yao**, Worapan Kusakunniran, Qiang Wu, J. Zhang and Zhen-min Tang. “Robust CNN-based Gait Verification and Identification using Skeleton Gait Energy Image.” 2018 Digital Image Computing: Techniques and Applications (DICTA) (2018): 1-7.

Accepted Papers

- J-1. **Lingxiang Yao**, Worapan Kusakunniran, Qiang Wu, Jian Zhang and Jingsong Xu. "Collaborative Feature Learning for Gait Recognition under Cloth Changes," IEEE Transactions on Circuits and Systems for Video Technology, 2021.
- J-2. **Lingxiang Yao**, Worapan Kusakunniran, Qiang Wu, Jingsong Xu and Jian Zhang. "Recognizing Gaits across Walking and Running Speedes", ACM Transactions on Multimedia Computing Communications and Applications, 2021.

Contents

Certificate	ii
Acknowledgments	iii
List of Publications	iv
List of Figures	xi
List of Tables	xvi
Abstract	xix
1 Introduction	1
1.1 Research Background	1
1.2 Research Problems	3
1.2.1 Gait Recognition under Clothing Changes	4
1.2.2 Gait Recognition across Walking and Running Modes	6
1.2.3 Gait Recognition Using a Few Gait Frames	9
1.3 Thesis Contribution	11
1.4 Thesis Structure	13
2 Literature Review and Related Theories	15
2.1 Gait Recognition under Clothing Changes	15
2.1.1 Appearance-based Methods under Clothing Changes	15
2.1.2 Model-based Methods under Clothing Changes	17
2.2 Cross-mode Gait Recognition	19

2.2.1	Walking Speed-invariant Gait Recognition	19
2.2.2	Cross-Mode Gait Recognition	21
2.3	Low Frame-rate Gait Recognition	22
2.4	Deep Learning-Based Gait Recognition	24
2.4.1	Convolutional Neural Networks	24
2.4.2	Recurrent Neural Networks	26
2.4.3	3D Convolutional Neural Networks	27
2.4.4	Deep AutoEncoders	28
2.4.5	Generative Adversarial Networks	28
2.4.6	Graph Convolutional Networks	29
2.4.7	Capsule Networks	29
2.4.8	Comparison	31
2.5	Gait Datasets	31
2.5.1	The CASIA Gait Dataset B	31
2.5.2	OU-ISIR Treadmill Gait Dataset A	32
2.5.3	OU-ISIR Treadmill Gait Dataset B	33
2.5.4	OU-MVLP Dataset and OUMVLP-Pose Dataset	33
2.6	Related Theories	34
2.6.1	View Transformation Model	34
2.6.2	Gait Recognition Using Bag-of-words Method	36
2.6.3	Self-Attention and Transformers	38
2.7	Summary	39
3	Skeleton Gait Energy Image	40
3.1	Introduction	40

3.2	Proposed Method	42
3.2.1	Skeleton Extraction	42
3.2.2	Skeleton Models	43
3.2.3	Gait Period Analysis	44
3.2.4	Skeleton Gait Energy Image	45
3.3	Experiment	47
3.3.1	Evaluation in a Cloth-changing Environment	48
3.3.2	Evaluation in an Unconstrained Environment with View Changes and Clothes Changes	49
3.4	Conclusion	57
4	Part-based Collaborative Spatio-temporal Feature Learning for Cloth-changing Gait Recognition	58
4.1	Introduction	58
4.2	Proposed Methods	60
4.2.1	Overview	61
4.2.2	Human Body Segmentation	63
4.2.3	Snapshots from View of $T - W$	64
4.2.4	Part-Based Collaborative Feature Learning	66
4.3	Experiments	67
4.3.1	Training and Testing Details	68
4.3.2	Comparison on CASIA Gait Dataset B	69
4.3.3	Comparison on OU-ISIR Treadmill Gait Dataset B	72
4.3.4	Ablation Experiments on CASIA Gait Dataset B	73
4.4	Conclusion	74

5 Collaborative Feature Learning for Gait Recognition under Clothes Changes	79
5.1 Introduction	79
5.2 Proposed Methods	81
5.2.1 Overview	81
5.2.2 Extracting Features from the Non/less Affected Body Parts	82
5.2.3 Extracting Features from the Skeleton Key-point Regions	85
5.2.4 Assembling Features from Two Different Perspectives	90
5.3 Experiments	91
5.3.1 Training and Testing Details	91
5.3.2 Comparison on CASIA Gait Dataset B	92
5.3.3 Ablation Experiments on CASIA Gait Dataset B	95
5.3.4 Studies of Features from the Non/less Affected Body Parts	98
5.3.5 Complexity Analysis	100
5.4 Conclusion	100
6 Recognizing Gaits across Walking and Running Modes	104
6.1 Introduction	104
6.2 Proposed Methods	105
6.2.1 Generating Spatial Features	106
6.2.2 Generating Temporal Features	110
6.2.3 Assembling Spatial and Temporal Features	116
6.3 Experiments	117
6.3.1 Dataset, Training and Testing Details	117
6.3.2 Comparison on OU-ISIR Treadmill Dataset A	119

6.3.3	Ablation Experiments	121
6.4	Conclusion	125
7	Gait Recognition using a Few Gait Frames	129
7.1	Introduction	129
7.2	Proposed Methods	131
7.2.1	Augmenting More Input Frames Using Different Strategies	131
7.2.2	Generating Gait Features Using Proposed network	134
7.3	Experiments	137
7.3.1	Experiments on CASIA Gait Dataset B	137
7.3.2	Experiments on OU-MVLP Dataset	142
7.4	Discussions	144
7.5	Conclusions	145
8	Conclusions and Future Work	147
8.1	Conclusions	147
8.2	Future Work	150
8.2.1	Improving the Proposed Methods with Deep Learning Tricks	150
8.2.2	Gait Recognition Using a Few Frames	151
8.2.3	Disentanglement Learning	151
8.2.4	Self-supervised Learning	152
8.2.5	Multi-biometric Recognition	153

List of Figures

1.1	Samples from CASIA Gait Dataset B (Yu et al. 2006).	4
1.2	Shape boundaries from OU-ISIR Treadmill Dataset A. Boundaries at walking speeds are shown in red, and boundaries at running speeds are shown in blue.	7
1.3	Average rank-1 accuracies with constraints of silhouette volume on CASIA-B in the setting LT. Accuracies are averaged from all views excluding identical-view cases across 10-times experiments.	10
1.4	Illustration of thesis framework.	14
2.1	Four different networks of CNNs in deep learning-based gait recognition: (a) GEINet in Shiraga et al. (2016); (b) LBNet in Wu et al. (2017); (c) MTNet in Wu et al. (2017); (d) GaitSet in Chao et al. (2019).	25
2.2	Three different architectures for using RNNs in deep learning-based gait recognition systems: (a) RNNs directly learn from the movement of joint positions; (b) RNNs are combined with CNNs; and (c) RNNs recurrently learn the relationships between partial representations in gait templates.	26

2.3	Silhouette sequences from OU-ISIR Treadmill Dataset A. Frame interval is adjusted for demonstration. The left sequences are from the same subject, and the right belong to another. Sequences on the same line mean that subjects are walking/running at the same speed. From top to bottom, their speeds are 2km/h, 5km/h, 7km/h, 8km/h and 10km/h, respectively.	32
2.4	Various clothing combinations in OU-ISIR Treadmill Dataset B. . . .	33
2.5	Sample images from the OU-ISIR Treadmill Dataset B.	34
2.6	Architecture of transformers.	37
3.1	Process of generating Skeleton Gait Energy Image.	41
3.2	Samples of skeleton detection.	42
3.3	Example of a pendulum model. θ is the angle between human thigh and the vertical direction.	44
3.4	Samples of waveforms before (red) and after (blue) the normalization and auto-correlation.	45
3.5	SGEIs (the first two rows) and GEIs (the last two rows) for the same person in two different dressing styles and under five different viewing angles.	46
3.6	Networks used for gait verification.	50
3.7	Framework of empirically combining GEI and SGEI.	52
4.1	Visualization of two silhouette sequences from three views. H/W are the height/width of each individual frame, and T is the timeline along each sequence. For each sequence, top left is from the $H - W$ view, top right is from the $T - H$ view, and bottom is from the $T - W$ view. Different from the $T - W$ snapshots, large differences can be found between the visualized $T - H$ snapshots. . .	59

4.2	Flowchart of the method proposed in Chapter 4.	61
4.3	Structure of the network proposed in Chapter 4.	62
4.4	Segmentation of human bodies.	63
4.5	Comparison of signals generated using wearable devices and snapshots generated from the $T - W$ view in our method. For each sample, the horizontal axis represents the timeline T . (a) is cited from Gafurov et al. (2006).	64
4.6	Samples of snapshots from the $T - W$ view for two persons with different dressing styles. Samples in the 2rd and 3st columns are obtained from the head part, and samples in the 4-8th columns are obtained from the crus part.	65
4.7	Relationship between the accuracy and the input frame number.	73
5.1	Overview framework of method proposed in this chapter.	81
5.2	The sub-network for extracting features from the non/less affected body parts.	82
5.3	The sub-network for extracting features from the estimated skeleton key-point regions.	85
6.1	Overview framework of our proposed method.	106

- 6.2 Network to formulate spatial features. For each sequence, five frames are randomly sampled as input. Beginning with a convolutional unit as the first convolutional stage, more convolutional stages are created by appending more convolutional units. After each stage a sub-network is generated, and all sub-networks are arranged in parallel. An information exchanging scheme is also modeled using the upsampling and downsampling operations between sub-networks (See Figure 6.3). For the first and second sub-networks, a pooling operation and HPM are used to project features of each frame into a more discriminative sequence feature for distinguishing different IDs. Moreover, at the top a global pipeline is proposed to extract global features for distinguishing walking and running modes. 107
- 6.3 Information exchanging between sub-networks. For indication only. . 108
- 6.4 Network to generate temporal features. For each input sequence, a series of 25 temporal templates are first averaged as input. The R(2+1)D convolutions are utilized to integrate temporal features from the fragmented temporal templates, and HPM is utilized to map the integrated features into a more discriminative subspace. The 2D-CNN convolutions are also utilized to extract spatial features about walking and running modes from each individual temporal templates and guide this network to be more concerned about the differences across these two modes. 111
- 6.5 Temporal templates sampled for the same person. From top to bottom, the speeds are 2km/h, 5km/h, 7km/h, 8km/h and 10km/h, respectively. 113

7.1	Samples for three different augmentation strategies. The first row reveals the original silhouettes in three walking conditions. The left three rows show sampled samples generated using skeleton modeling, horizontal flipping, or cycleGAN.	133
7.2	The structure of our proposed two-branch network.	135

List of Tables

2.1	Averaged rank-1 accuracies (%) on CASIA-B under different settings, excluding identical-view cases.	30
3.1	Rank-1 accuracy (%) when the probe is in a long coat and the gallery is in the normal dressing.	48
3.2	Rank-1 accuracy (%) when the probe is in the normal dressing and the gallery is in a long coat.	48
3.3	Rank-1 accuracy (%) when the gallery viewing angle is 72°	51
3.4	Rank-1 accuracy (%) when the gallery viewing angle is 90°	51
3.5	Rank-1 accuracy (%) when the gallery viewing angle is 108°	51
3.6	Rank-1 accuracy (%) when the gallery viewing angle is 54°	53
3.7	Rank-1 accuracy (%) when the gallery viewing angle is 72°	53
3.8	Rank-1 accuracy (%) when the gallery viewing angle is 90°	53
3.9	Rank-1 accuracy (%) when the gallery viewing angle is 108°	54
3.10	Rank-1 accuracy (%) when the gallery viewing angle is 126°	54
3.11	Rank-1 accuracy when the probe viewing angle is 72°	55
3.12	Rank-1 accuracy when the probe viewing angle is 90°	55
3.13	Rank-1 accuracy when the probe viewing angle is 108°	56
4.1	Rank-1 accuracy (%) under the same viewing angle on CASIA-B.	68
4.2	Rank-1 accuracy (%) under the same viewing angle on CASIA-B.	69

4.3	Rank-1 accuracy (%) under different walking conditions on CASIA-B.	70
4.4	Averaged rank-1 accuracy (%) on CASIA-B under three different experimental settings, excluding identical-view cases.	76
4.5	Comparison on OU-ISIR Treadmill Dataset B by accuracies (%).	77
4.6	Influence of feature components.	78
5.1	Comparison on CASIA-B under the same normal viewing angle by accuracies (%).	92
5.2	Rank-1 accuracy (%) on CASIA-B under different walking conditions.	93
5.3	Averaged rank-1 accuracy (%) on CASIA-B under three different settings, excluding identical-view cases.	101
5.4	Ablation experiment results (%) on CASIA-B using setting LT, excluding identical-view cases.	102
5.5	Accuracies (%) of different body parts.	103
5.6	Complexity Analysis.	103
6.1	Averaged rank-1 accuracy (%) on OU-ISIR Treadmill Dataset A.	119
6.2	Rank-1 accuracy (%) of our proposed method on OU-ISIR Treadmill Dataset A. G/P denotes Gallery/Probe.	126
6.3	Ablation experiments of spatial features.	127
6.4	Ablation experiments of temporal features.	128
6.5	Ablation experiments of assembling features.	128
7.1	Averaged rank-1 accuracies (%) on CASIA-B, excluding identical-view cases, using the same input frame number.	139
7.2	Averaged rank-1 accuracies (%) on CASIA-B, excluding identical-view cases, using different input frame numbers.	141

7.3	Averaged rank-1 accuracies (%) on OU-MVLP, excluding identical-view cases, using the same input frame number.	143
-----	---	-----

ABSTRACT

HUMAN GAIT RECOGNITION UNDER CHANGES OF WALKING CONDITIONS

by

Lingxiang Yao

Gait has been gathering extensive research interest for its non-fungible position in applications, *e.g.*, security surveillance and forensic identification. First, it is difficult to disguise one's gait, since walking is necessary for human mobility. Second, it works well in an unconstrained condition and can be attained at a distance without physical contact or proximal sensing. However, although recently different methods have been proposed for gait recognition, gait analysis is still in its infancy. Most methods enable to garner a remarkable recognition performance when the gallery and the probe are in a similar situation. However, when exterior factors affect a person's gait and changes occur in human appearances, a significant performance degradation happens.

Among these exterior factors, clothing variations and mode changes can be treated as the most influential factors for gait recognition. It is advisable to identify a person using gait, since each person exhibits his/her walking patterns in a sufficiently unique and fairly characteristic way. However, clothing variations can significantly influence available features to be used in the future recognition process, while walking/running modes can change human motions made by limbs and thus dramatically influence the instinct walking patterns of each person. Hence, in this thesis different methods have been proposed for gait recognition to handle the difficulties of clothing variations and walking/running mode changes.

First, given that model-based methods are less vulnerable to clothing variances, a more robust model-based gait feature, Skeleton Gait Energy Image (SGEI), is formed to handle this cloth-changing gait recognition problem. Then, since clothing changes

can cause different impacts to different body parts, a part-based collaborative spatio-temporal feature learning method is also proposed for cloth-changing gait recognition by concatenating features from the non/less affected body parts under the correlative $H-W$ and $T-W$ views. Based on the aforementioned two methods, another efficient network is proposed for cloth-changing gait recognition. This network consists of two sub-networks, aiming to produce part-based features from the non/less affected body parts and the estimated skeleton key-point regions. Moreover, in order to address the walking-vs-running problem in a cross-mode manner, a feasible hybrid method is also proposed in this thesis. Distinct from most cross-mode gait recognition methods, this method focuses on learning mode-invariant features for each person from their innate patterns between walking and running modes. Multi-task learning strategies are also used to enhance the efficiency of these learned features. Finally, given that the above-mentioned methods are all proposed based on sufficient input data, a complementary solution is given when only a few gait frames can be offered.

To sum up, the main objective of this thesis is to address the problems of clothing variations and walking/running mode changes for gait recognition, thus four different methods have been proposed in this thesis. Besides, related experiments have proved that these proposed methods can obtain a remarkable performance when tackling the cloth-changing and walking-vs-running gait recognition problems.