UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

LONG-TERM PERSON RE-IDENTIFICATION IN THE WILD

by

Peng Zhang

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Sydney, Australia

2020

Certificate of Authorship/Originality

I, Peng Zhang, declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the 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.

Production Note:

Signature: Signature removed prior to publication.

Date: 16/06/2020

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.

I would like to express my deepest appreciation, first and foremost, to my supervisor, A/Prof. Qiang Wu, for his professional guidance and warm encouragement. I am deeply impressed by his insight understanding on computer vision and his rich skills on research writing and presentation. He can always inspire me with fancy ideas and rigorous logic. His enthusiasm, attitude and devotion towards academy have deeply influenced me, which provides instructions to my future career. It is the luckiest thing that has him as my supervisor.

I also want to express my sincere attitude to my co-supervisor Dr. Jingsong Xu and A/Prof. Jian Zhang. Jingsong gives me a lot of constructive suggestions for my research and helps me polish papers. Besides, he always encourages me to focus on the advanced techniques that motivate the research going deeper. Jian organized many interesting seminars that provide us opportunities to share and communicate our research progress.

Then, I wish to give thanks to A/Prof. Xianye Ben, who was my supervisor during the master study in Shandong University. She guided me into the field of computer vision and provided me endless support to pursue PhD study. Without her recommendation, I might not have opportunity to study in UTS and work with A/Prof. Qiang Wu.

And, I appreciate the help, support and friendship from my dear colleagues and friends during my doctoral study. Thanks to Yifan Zuo, Zongjian Zhang, Qian Li, Yan Huang, Xunxiang Yao, Lingxiang Yao, Muming Zhao, Xiaoshui Huang, Junjie Zhang, Lina Li, Lu Zhang, Yongshun Gong, Zhibin Li and all other labmates for their collaboration and discussion. Without their help, I cannot collect my experimental data. I am also grateful to all my friends in Sydney: Lei Sang, Lin Zhu, Tao Shen, Mengyao Li, Xiaolin Zhang, Wentao Li, Mingjie Li, Zhuo Tang, Xin Ba and Shuo Yang for their encouragement and companion. Thank you guys that bring me to try delicious food and visit beautiful landmarks in Sydney.

Finally, I would like to express my sincere thanks to my parents and girlfriend Weiyu for their endless support, trust, encouragement and love throughout my studies these years.

Peng Zhang

February 2020 @ UTS.

List of Publications

Journal Papers

J-1. **P. Zhang**, J. Xu, Q. Wu, Y. Huang and J. Zhang, "Top-Push Constrained Modality-Adaptive Dictionary Learning for Cross-Modality Person Re-Identification," *IEEE Transactions on Circuits and Systems for Video Technology*, Early Access, 2019.

J-2. X. Ben, **P. Zhang**, Z. Lai, R. Yan, X. Zhai and W. Meng, "A general tensor representation framework for cross-view gait recognition," *Pattern Recognition*, vol. 90, pp. 87-98, 2019.

J-3. X. Yao, Q. Wu, **P. Zhang** and F. Bao, "Adaptive rational fractal interpolation function for image super-resolution via local fractal analysis," *Image and Vision Computing*, vol. 82, pp. 39-49, 2019.

J-4. X. Ben, C. Gong, **P. Zhang**, X. Jia, Q. Wu and W. Meng, "Coupled patch alignment for matching cross-view gaits," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 3142-3157, 2019.

J-5. X. Ben, C. Gong, **P. Zhang**, R. Yan, Q. Wu and W. Meng, "Coupled bilinear discriminant projection for cross-view gait recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, pp. 734-747, 2020.

J-6. Y. Huang, J. Xu, Q. Wu, Y. Zhong, **P. Zhang** and Z. Zhang, "Beyond Scalar Neuron: Adopting Vector-Neuron Capsules for Long-Term Person Re-Identification," *IEEE Transactions on Circuits and Systems for Video Technology*, Early Access, 2019.

J-7. X. Yao, Q. Wu, **P. Zhang** and F. Bao, "Weighted Adaptive Image Super-Resolution Scheme based on Local Fractal Feature and Image Roughness", *IEEE Transactions on Multimedia*, Accepted, 2020.

Conference Papers

C-1. **P. Zhang**, Q. Wu and J. Xu, ”VT-GAN: View Transformation GAN for Gait RecognitionAcross Views,” *The International Joint Conference on Neural Network (IJCNN)*, Budapest, 14-19 July, 2019.

C-2. **P. Zhang**, Q. Wu and J. Xu, ”VN-GAN: Identity-preserved Variation Normalizing GAN for Gait Recognition,” *The International Joint Conference on Neural Network (IJCNN)*, Budapest, 14-19 July, 2019.

C-3. **P. Zhang**, Q. Wu, J. Xu and J. Zhang, ”Long-term person re-identification using true motion from videos,” *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, 12-15 March, 2018, pp. 494-502.

C-4. H. Song, H. Dong, and **P. Zhang**, ”A virtual instrument for diagnosis to substation groundinggrids in harsh electromagnetic environment,” *IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Turin, 22-25 May, 2017.

Submitted Papers

J-1. **P. Zhang**, Q. Wu, J. Xu and Y. Huang, ”Learning Hybrid Representations over Walking Tracklet for Long-term Person Re-Identification in The Wild,” *IEEE Transactions on Multimedia*, Under review, 2020.

J-2. Y. Huang, Q. Wu, J. Xu, Y. Zhong, **P. Zhang** and Z. Zhang, ”Learning from Decoupled Semantic Cue for Infrared-Visible Person Re-identification”, *IEEE Transactions on Information Forensics and Security*, Under review, 2020.

J-3. Y. Huang, Q. Wu, J. Xu, Y. Zhong, **P. Zhang** and Z. Zhang, ”Alleviating Modality Bias Training for Infrared-Visible Person Re-identification”, *IEEE Transactions on Multimedia*, Under review, 2020.

Contents

List of Figures

List of Tables

ABSTRACT

LONG-TERM PERSON RE-IDENTIFICATION IN THE WILD

by

Peng Zhang

Person re-identification (re-ID) has been attracting extensive research interest because of its non-fungible position in applications such as surveillance security, criminal investigation and forensic reasoning. Existing works assume that pedestrians keep their clothes unchanged while passing across disjoint cameras in a short period. It narrows person re-ID to a short-term problem and incurs solutions using appearance-based similarity measurement. However, this assumption is not always true in practice. For example, pedestrians are high likely to re-appear after a long-time period, such as several days. This emerging problem is termed as long-term person re-ID (LT-reID).

Regarding different types of sensors deployed, LT-reID is divided into two subtasks: person re-ID after a long-time gap (LTG-reID) and cross-camera-modality person re-ID (CCM-reID). LTG-reID utilizes only RGB cameras, while CCM-reID employs different types of sensors. Besides challenges in classical person re-ID, CCM-reID faces additional data distribution discrepancy caused by modality difference, and LTG-reID suffers severe within-person appearance inconsistency caused by clothing changes. These variations seriously degrade the performance of existing re-ID methods.

To address the aforementioned problems, this thesis investigates LT-reID from four aspects: motion pattern mining, view bias mitigation, cross-modality matching and hybrid representation learning. Motion pattern mining aims to address LTG-reID by crafting true motion information. To this point, a fine motion encoding method is proposed, which extracts motion patterns hierarchically by encod-

ing trajectory-aligned descriptors with Fisher vectors in a spatial-aligned pyramid. View bias mitigation targets on narrowing discrepancy caused by viewpoint difference. This thesis proposes two solutions: VN-GAN normalizes gaits from various views into a unified one, and VT-GAN achieves view transformation between gaits from any two views. Cross-modality matching aims to learn modality-invariant representations. To this end, this thesis proposes to asymmetrically project heterogeneous features across modalities onto a modality-agnostic space and simultaneously reconstruct the projected data using a shared dictionary on the space. Hybrid representation learning explores both subtle identity properties and motion patterns. Regarding that, a two-stream network is proposed: the space-time stream performs on image sequences to learn identity-related patterns, e.g., body geometric structure and movement, and skeleton motion stream operates on normalized 3D skeleton sequences to learn motion patterns.

Moreover, two datasets particular for LTG-reID are presented: Motion-reID is collected by two real-world surveillance cameras, and CVID-reID involves tracklets clipped from street-shot videos of celebrities on the Internet. Both datasets include abundant within-person cloth variations, highly dynamic background and diverse camera viewpoints, which promote the development of LT-reID research.

Abbreviation

CCM-reID - cross camera modality person re-identification

CLT-reID -contemporary long-term person re-identification

CMC - cumulative matching characteristic

CNN -convolutional neural network

CST-reID - conventional short-term person re-identification

CVGLT-reID - cross-view gait-based long-term person re-ID

DT -dense trajectory

FITD - fine motion encoding

GAN - generative adversarial network

GCN - graph convolutional network

GEI - gait energy image

GMM - Gaussian mixture model

LT-reID - log-term person re-ID

LTG-reID -person re-ID after long-time gap

mAP - mean average precision

PCA - principle component analysis

re-ID - re-identification

SILTP - Scale Invariant Ternary Pattern

SOTA - state-of-the-art

TCMDL - top-push constrained modality-adaptive dictionary learning

TSI - Target subject of interest

VN-GAN - variational normalizing generative adversarial network

VT-GAN - view transformation generative adversarial network