

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.

Signature: Production Note:  
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 Recognition Across 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 grounding grids 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

Certificate	ii
Acknowledgments	iii
List of Publications	v
List of Figures	xii
List of Tables	xviii
Abstract	xx
Abbreviation	xxii
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.1.1 Conventional Short-term Person re-ID . . . . .	2
1.1.2 Challenging Long-term Person Re-ID . . . . .	8
1.2 Research Problems . . . . .	11
1.2.1 Pure Motion Estimation from Dense Trajectories . . . . .	12
1.2.2 View Bias Mitigation via GAN . . . . .	13
1.2.3 Camera Modality Bias Mitigation . . . . .	13
1.2.4 Learning Hybrid Representation via Neural Networks . . . . .	14
1.2.5 Dataset Collection . . . . .	14
1.3 Thesis Contribution . . . . .	15
1.4 Thesis Structure . . . . .	16

<b>2 Literature Review and Related Theories</b>	<b>18</b>
2.1 Review on Current Works of Person Re-ID . . . . .	18
2.1.1 Conventional Short-term Person Re-ID . . . . .	18
2.1.2 Contemporary Long-term Person Re-ID . . . . .	25
2.1.3 Person Re-ID after Long-time Gap . . . . .	26
2.1.4 Benchmark Datasets for Person Re-ID . . . . .	27
2.2 Motion Attribute for Person Re-ID . . . . .	30
2.2.1 Dense Trajectory . . . . .	30
2.2.2 Fisher Vector . . . . .	32
2.2.3 Gait Energy Image . . . . .	33
2.3 Key Techniques of Heterogeneous Person Re-ID . . . . .	34
2.3.1 Cross-modality Dictionary Learning . . . . .	34
2.3.2 Cross-view Person Re-ID Using Gaits . . . . .	35
2.4 Summary . . . . .	36
<b>3 Long-term Person Re-identification using True Motion from Videos</b>	<b>37</b>
3.1 Introduction . . . . .	37
3.1.1 Problem Formulation . . . . .	37
3.1.2 Motivation . . . . .	39
3.2 Fine Motion Encoding . . . . .	39
3.2.1 Body-action Pyramid Model . . . . .	40
3.2.2 Motion Trajectories for Re-ID . . . . .	42
3.2.3 Trajectory-aligned Motion Statistics . . . . .	43
3.2.4 Fisher Vector Encoding of Motions . . . . .	44



3.2.5	Feature Fusion . . . . .	45
3.3	Dataset . . . . .	46
3.3.1	Benchmark Datasets . . . . .	46
3.3.2	New Motion-ReID Dataset . . . . .	47
3.4	Experiments . . . . .	49
3.4.1	Experiment Setting . . . . .	50
3.4.2	Experiment Analysis . . . . .	51
3.5	Summary . . . . .	56
<b>4</b>	<b>GANs for CVGLT-reID</b>	<b>58</b>
4.1	Problem Formulation . . . . .	58
4.2	Identity-preserved Variation Normalizing GAN . . . . .	59
4.2.1	Motivations . . . . .	59
4.2.2	Proposed Method . . . . .	62
4.2.3	Network Architecture Implementation . . . . .	67
4.2.4	Experiments . . . . .	68
4.2.5	Conclusions . . . . .	77
4.3	View Transformation GAN . . . . .	78
4.3.1	Motivations . . . . .	78
4.3.2	Proposed Framework . . . . .	82
4.3.3	Experiments . . . . .	86
4.3.4	Conclusions . . . . .	96
4.4	Summary . . . . .	96
<b>5</b>	<b>Top-push Constrained Modality-adaptive Dictionary Learning for Cross-modality Person Re-ID</b>	<b>97</b>

5.1	Introduction . . . . .	97
5.1.1	Problem Formulation . . . . .	97
5.1.2	Motivation . . . . .	98
5.2	Top-push Constrained Modality-Adaptative Dictionary Learning . . .	101
5.2.1	Framework Overview . . . . .	101
5.2.2	Objectives . . . . .	102
5.2.3	Optimization . . . . .	106
5.2.4	Complexity Analysis . . . . .	110
5.2.5	Matching for Cross-modality Person Re-ID . . . . .	110
5.3	Experiments . . . . .	110
5.3.1	Experiments on SYSU-MM01 . . . . .	112
5.3.2	Experiments on BIWI RGBD-ID dataset . . . . .	119
5.3.3	Parameter Analysis . . . . .	124
5.3.4	Effects of Energy Terms . . . . .	126
5.4	Summary . . . . .	127
<b>6</b>	<b>Learning Hybrid Representations over Tracklets for Person Re-ID in the Wild</b>	<b>128</b>
6.1	Introduction . . . . .	129
6.1.1	Problem Formulation . . . . .	129
6.1.2	Motivation . . . . .	129
6.2	The Proposed Method . . . . .	131
6.2.1	Structure Overview . . . . .	131
6.2.2	Set-based Subtle Identity Net . . . . .	132
6.2.3	Skeleton-based Motion Identity Net . . . . .	136

6.2.4	Two-stream Networks . . . . .	141
6.3	CVID-reID Dataset . . . . .	142
6.4	Experiments . . . . .	144
6.4.1	Network Structure and Implementation Details . . . . .	145
6.4.2	Datasets and Experiment Setting . . . . .	146
6.4.3	Challenges of Long-term Person Re-ID . . . . .	146
6.4.4	Effectiveness of SpTSkM for Long-term Person Re-ID . . . . .	147
6.4.5	Robustness for CST-reID . . . . .	150
6.4.6	Ablation Study . . . . .	151
6.4.7	Parameter Analysis . . . . .	155
6.5	Summary . . . . .	157
<b>7</b>	<b>Conclusions and Future Work</b>	<b>158</b>
7.1	Conclusions . . . . .	158
7.2	Future Work . . . . .	160

## List of Figures

1.1	An example of the CST-reID setting . . . . .	2
1.2	An example of person Re-ID as the retrieval application . . . . .	3
1.3	Pipeline of a complete person Re-ID system . . . . .	3
1.4	A comparison of variations in four benchmark datasets: (a) VIPeR, (b) CUHK03, (c) MARS, (d) DukeMTMC-reID . . . . .	5
1.5	An example of three different types of sensors. . . . .	6
1.6	An example of long-term person re-ID setting which includes two cameras deployed at distinct gates in one building. Pedestrians collected in the same day are put in boxes with the same colour. . . . .	8
1.7	Illustration of challenges in long-term person re-ID. Images in the same row is taken from the same TSI . . . . .	9
1.8	An comparison of CST-reID and CCM-reID. Modality of query sample and gallery samples are same for CST-reID, but it is different for CCM-reID. . . . .	11
1.9	Relationship structure of our research problems . . . . .	12
1.10	Illustration of thesis structure . . . . .	17
3.1	Illustration of LTG-reID challenges. Each line of images is collected from same camera, and each column of images belongs to the same subject. Images from last two rows are captured by a long-time interval with the top row. . . . .	38

3.2	Framework of the proposed FITD model. It consists of two phases: model training and feature extracting. . . . .	40
3.3	Top: our proposed body-action pyramid model consists of eight body-action units, which is labelled from 1 to 8. Middle: dense trajectories in each body-action unit. Bottom: Fisher vectors correspond to the eight units. . . . .	41
3.4	Illustration of samples in the proposed Motion-ReID dataset. . . . .	48
3.5	Video recording timeline of a subject. . . . .	49
3.6	An example of the eight evaluation subsets in Motion-ReID. . . . .	50
4.1	Overview of Siamese structure of the proposed VN-GAN. It includes two-stream networks where each stream consists of the proposed coarse-to-fine design in Figure 4.2. Noting that the variation discriminator $\mathbf{D}_v$ and variation classifier $\mathbf{P}$ in Stage-I are not shown in the figure, because the figure only aims to show the Siamese design. Details of each branch is illustrated in Figure 4.2. . . . .	60
4.2	Overview of the coarse-to-fine structure. It is sub-branch of the proposed VN-GAN which consists of two stages: coarse gait image generation and refinement. In the framework, $\odot$ denotes channel concatenation and $\ominus$ denotes image difference, $\oplus$ denotes image summation. . . . .	62
4.3	An example of GEIs from 11 views, i.e., $0^\circ, 18^\circ, \dots, 180^\circ$ from left to right, and images at each column stand for GEIs from the same view angle. The first three rows compares GEIs from three different people in normal walking pattern. The last two rows includes the same subject with the first row but different walking conditions, i.e., the subject carries a bag in the forth row and wear a coat in the last row. . . . .	70
4.4	Average accuracies except the identical view on three probe sets. . . . .	73

4.5	Comparison with the SOTA approaches on ProbeNM subset at three distinct probe views $54^\circ$ , $90^\circ$ and $126^\circ$ . . . . .	74
4.6	An example of normalized gaits of six subjects. The three rows stand for input gaits, generated gaits and reference gaits, respectively. Each column represents one subject. . . . .	76
4.7	Comparison between gaitGAN and the proposed VT-GAN. (a) The former normalizes GEIs from arbitrary views to a reference one, and (b) the latter directly transform gait images between any pair of views.	79
4.8	The architecture of the proposed VT-GAN. It consists of three modules, i.e., generator $\mathbf{G}$ , discriminator $\mathbf{D}$ and similarity preserver $\Phi$ . Generator $\mathbf{G}$ inputs the concatenation of source/condition gait image from arbitrary view and target view indicator, and synthesizes gait image from target view. Discriminator $\mathbf{D}$ learns to distinguish between synthesised gait image and real gait image and classify real gait image to its corresponding view. Similarity preserver $\Phi$ learns to pull positive gait pairs together and push negative gait pairs away.	81
4.9	An example of GEIs from 11 views. Each row includes GEIs of the same subject from $0^\circ$ to $180^\circ$ with $18^\circ$ interval. (a) shows GEIs of three different subjects in terms of three rows. (b) exhibits GEIs in three different conditions of the same subject, i.e., NN, BG and CL. .	86
4.10	Comparison of accuracies on the three probe subsets at the 11 probe views. For each probe view, accuracies of the rest views are excluded.	88
4.11	Comparison with previous methods at three representative probe views on ProbeNM. . . . .	89
4.12	Visualization of synthesised gaits between any two views. Images in blue box are input gaits, and the ones in red box are reference gaits. The rest $11 \times 11$ images are synthesised gaits conditioned on the corresponding input image and view indicator. . . . .	93

4.13	Visualization of synthesised gaits of three different subject. Each row includes gaits from the same subject. And, the first column is reference gaits in $90^\circ$ , the rest rows are synthesised gaits from input gaits in $0^\circ$ to $180^\circ$ . . . . .	93
4.14	Visualization of synthesised gaits of the same subject in three distinct conditions, i.e., normal walking, carrying a bag and wearing a coat corresponding to three rows, respectively. The first column includes reference gaits of the same subject in $90^\circ$ and the rest columns consist of synthesised gaits from various views. . . . .	94
4.15	An example of some bad samples and their corresponding translated gaits. For each pair, the left lists input gaits and the right shows synthesised gaits. . . . .	94
5.1	Illustration of how our asymmetric mapping briataaedges the data gaps. We performed PCA on samples from BIWI RGBD-ID dataset Munaro et al. (2014b) for visualization. Each shape (circle or triangle) represents samples from one modality. (a) original data distribution, (b) distribution in the shared space learned by asymmetric mapping. . . . .	99
5.2	Illustration of the proposed TCMDL model. Objects with the same shape (hallow/solid) represent the same person. Input images from cross modalities are mapped into a subspace in which a shared dictionary is learned. Meanwhile, the encoding coefficients are regularized by a top-push ranking constraint embedded Laplacian-like graph. . . . .	100

5.3	Examples of samples in SYSU-MM01 dataset. Images from cameras 1-3 in the blue box are captured on indoor scenes while images from camera 4-6 in the green box are captured on outdoor scenes. Cameras 1, 2, 4, 5 are visual light sensors and cameras 3, 6 are near-infrared sensors. Every column represents images from the same person. . . . .	112
5.4	Examples of images in the BIWI RGBD-ID dataset. Images in the top row are RGB images and in the bottom row are depth images (shown by pseudo-colour) as well as skeletons. . . . .	119
5.5	Examples of point clouds. Images in top row is the RGB images in BIWI RGBD-ID dataset and images in bottom row are their corresponding visualization of point clouds. . . . .	120
5.6	Parameter analysis using DZP for all-search mode. Rank-1 and mAP accuracy with different parameters (a) $\lambda_1$ , (b) $\lambda_2$ , (c) $\beta$ and dictionary size (d) $K$ are reported. . . . .	124
5.7	Parameter analysis on subset #1 of BIWI RGBD-ID dataset. Rank-1 and mAP accuracy in terms of different parameters (a) $\lambda_1$ , (b) $\beta$ , and dictionary size (c) $K$ are reported. . . . .	125
6.1	An overview of the proposed SpTskM framework. " $\oplus$ " denotes element-wise addition, "SP" denotes set pooling, "GAP" denotes global average pooling. Different color represents different processing flow. . . . .	131
6.2	The proposed 2D cross-attention module (2DCAM) for an image. The boxes denote 2D convolution operations with $1 \times 1$ kernels. . . .	133
6.3	An illustration of 3D pose estimation. Joints are highlighted by circle. Origin of coordinate is located on the middle hip of the estimated skeleton. . . . .	137



6.4	An illustration of 3D skeleton normalization. Origin of coordinate is denoted by the red circle. . . . .	137
6.5	(a) An illustration of constructed spatio-temporal graph; (b) An illustration of partition strategy. $\star$ denotes gravity center. For a neighbor set, it is categorized into three subsets: root joint (in red), centripetal joints (in magenta) and centrifugal joints (in orange). . . . .	140
6.6	The pipeline of data acquisition process for the proposed CVID-reID.	142
6.7	An example of six celebrities in the proposed CVID-reID dataset. It can be seen abundant of intra-person clothing changes exist in the dataset. . . . .	144
6.8	Performance comparison of four baseline methods on MARS and CVID-reID in terms of (a) identification accuracy (Rank-1) and (b) mAP. . . . .	148
6.9	A comparison of intra-person variations between CVID-reID (top row) and MARS (bottom row). . . . .	148
6.10	Analysis of identification performance versus hyper-parameters (a) Number of Frames in a training tracklet, (b) $\alpha_i$ in Eq. 6.7, and (c) $\gamma$ in Eq. 6.9. . . . .	154

## List of Tables

1.1	A comparison between short-term person and long-term person re-ID.	10
2.1	A comparison of existing datasets for person re-ID in both short-term and long-term scenarios.	29
3.1	A comparison of proposed FITD with other popular features on PRID2011 dataset and PRID BK dataset	52
3.2	A comparison of proposed FITD with different encoding methods.	54
3.3	A comparison of proposed FITD with different fusion methods.	55
3.4	A comparison of proposed FITD with other popular features on Motion-ReID dataset.	55
4.1	Identification accuracy on ProbeNM subset.	71
4.2	Identification accuracy on ProbeBG subset.	72
4.3	Identification accuracy on ProbeCL subset.	72
4.4	Comparison of average accuracies of 10 gallery views (The corresponding view is excluded.) on ProbeNM with probe view $54^\circ$ , $90^\circ$ and $126^\circ$ .	75
4.5	Average accuracies of 10 gallery views (The corresponding view is excluded.) on ProbeNM with 5 distinct probe views. ‘s1’ demotes training with Stage-I of the proposed VN-GAN.	77

4.6	Comparison with recent works on CVGLT-reID on ProbeNM. Average identification accuracies except the corresponding view are reported. . . . .	91
4.7	Evaluation of effectiveness of the identity preserver in VT-GAN. . . . .	92
5.1	Results using LOMO (%). '-' means result is not reported. . . . .	115
5.2	Results using DZP (%). '-' means result is not reported. . . . .	117
5.3	Comparison with the state-of-the-art methods for thermal-visible Re-ID on SYSU-MM01 dataset. . . . .	118
5.4	Results on BIWI RGBD-ID subset #1 (%). . . . .	122
5.5	Results on BIWI RGBD-ID subset #2 (%). . . . .	122
5.6	Results on BIWI RGBD-ID subset #3 (%). . . . .	123
6.1	Summarization of potential properties for person re-ID. It is categorized by attribute's time-gap stability. . . . .	130
6.2	Performance comparison between SOTA video-based person re-ID methods and proposed SpTskM on CVID-reID for long-term person re-ID (%). . . . .	149
6.3	Performance comparison between SOTA video-based person re-ID methods and proposed SpTskM on MARS for short-term person re-ID (%). Best and second-best results are highlighted by bold and underline, respectively. . . . .	151
6.4	Ablation study of SSIN stream on CVID-reID dataset (%). . . . .	153
6.5	Ablation Study of Set Aggregation Method (%). . . . .	153
6.6	Ablation Study of SMIN on CVID-reID (%). . . . .	154
6.7	Ablation Study of Dual Stream Networks on CVID-ReID (%). . . . .	155

# 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 sub-tasks: 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

