CERTIFICATE

It is certified that the work contained in the thesis titled ***"Person Re-Identification for Intelligent Surveillance Using Deep Learning"*** by ***Nirbhay Kumar Tagore*** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all requirements of Comprehensive Examination, Candidacy, and SOTA for the award of Ph.D. Degree.

Supervisor Pratik Chattopadhyay 01/11/2021

Dr. Pratik Chattopadhyay
Assistant Professor,
Department of Computer Science and Engineering,
Indian Institute of Technology (BHU), Varanasi,
Uttar Pradesh, INDIA 221005

DECLARATION BY THE CANDIDATE

I, ***Nirbhay Kumar Tagore***, certify that the work embodied in this Ph.D. thesis is my own bonafide work carried out by me under the supervision of ***Dr. Pratik Chattopadhyay*** from ***December 2017*** to ***October 2021*** at ***Department of Computer Science and Engineering***, Indian Institute of Technology (BHU) Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, *etc.* reported in journals, books, magazines, reports, dissertations, theses, *etc.*, or available at websites and have not included them in this thesis and have not cited as my own work.

Date: 01/11/2021

Place: Varanasi

Nirbhay Kumar Tagore

(Nirbhay Kumar Tagore)

CERTIFICATE BY THE SUPERVISOR

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Pratik Chattopadhyay 01/11/2021

(Dr. Pratik Chattopadhyay)
Assistant Professor,
Dept. of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi.

पर्यवेक्षक/Supervisor
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg.
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(काशी हिन्दू विश्वविद्यालय)
(Banaras Hindu University)
वाराणसी/Varanasi-221005

Signature of Head of Department

आचार्य व विभागाध्यक्ष
Professor & Head
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg.
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(बनारस हिन्दू यूनिवर्सिटी)
(Banaras Hindu University)
वाराणसी–२२१००५ / Varanasi-221005

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Person Re-Identification for Intelligent Surveillance Using Deep Learning

Name of the Student: Nirbhay Kumar Tagore

Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University), Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the *Doctor of Philosophy.*

Date: 01/11/21

Place: Varanasi

Nirbhay Kumar Tagore

(Nirbhay Kumar Tagore)

Dedicated to my parents,

Mr. B. S. Tagore

and

Mrs. Roopa Tagore

ACKNOWLEDGEMENT

First and foremost, I would like to thank my supervisor, Dr. Pratik Chattopadhyay, for his invaluable support and assistance. I feel immense pleasure in expressing my profound sense of gratitude and sincere regard for his constant feedback and expertise during all these years. I am eternally grateful to have had the opportunity to work on my thesis under his supervision.

My cordial thanks to all the members of the Department of Computer Science and Engineering for creating an excellent working atmosphere. I would also like to thank the other members of my Doctoral committee, Dr. Sandeep Ghosh, Department of Electrical Engineering, and Dr. Bhaskar Biswas, Department of Computer Science and Engineering, for their help and support throughout the tenure of my studies. Special thanks to Prof. Rajeev Srivastava, Department of Computer Science and Engineering, for his valuable suggestions. I would also like to convey my sincere gratitude to Prof. Sanjay Kumar Singh, Head of the CSE Department, and all the RPEC and DPGC members for their suggestions and endorsement of this work.

I am grateful to my colleagues and friends, Ankit Jaiswal, Naina Yadav, Amit Biswas, Shashank Kumar Singh, Amit Kumar, and Ramakant Kumar, for the long discussions and their brilliant insights that have helped me to overcome the challenges I have faced in the development of this work. Finally, I express my heartfelt gratitude to my parents Mrs. Roopa Tagore and Mr. Brehma Shanker Tagore, and my life partner Swati for their constant support, love, encouragement, and sacrifices. Their affectionate love and care cannot be expressed in words.

With limitless humility, I would like to praise and thank the **"Sankat Mochan Hanuman Ji"** and **"Baba Kashi Vishwanath Ji"**. The almighty, the Merciful compassionate who bestowed me with all the favourable circumstances to achieve the desired goal of life through this crucial juncture.

(Nirbhay Kumar Tagore)

Contents

List of Figures **xi**

List of Tables **xvi**

List of Symbols **xvii**

List of Abbreviations **xix**

Abstract **xxi**

1 Introduction **1**

1.1 Person Re-Identification 1

1.1.1 Possible Scenarios and Application Areas for Person Re-Identification 2

1.1.2 Need for Development of Automated Re-id Technique 3

1.1.3 Data Acquisition Devices to Carry-Out Re-id in Surveillance Applications 4

1.1.4 Challenges in Computer Vision-Based Person Re-id Applied to Surveillance Sites 6

1.2 Motivation of the Work 7

1.3 Contributions of the Thesis 9

1.3.1 Developing Improved Person Re-identification Approaches from Still Images 9

1.3.2 Effective Handling of Motion Features in Video-Based Person Re-identification 11

1.3.3 Handling Occlusion in Images 13

1.3.4 Handling Occlusion in Videos 15

1.3.5 Constructing A New Data Set and Making the Pre-Trained Models Publicly Available 17

1.4 Organization of the Thesis 17

2 Related work **19**
2.1 Traditional Person Re-Identification Methods 19
2.1.1 Contextual Methods 20
2.1.2 Non-Contextual Methods 21
2.2 Modern Approaches 24
2.2.1 Deep Learning-based Approaches 24
2.2.2 Siamese Network-based Approaches 26
2.2.3 GAN-based Approaches 28
2.3 Scopes for Further Research 35
2.4 Data Set Description and Evaluation Metrics 37
2.4.1 Image-based Re-Identification Data Set 37
2.4.2 Video-based Re-Identification Data Set 40
2.4.3 Evaluation Metrics 41
2.5 Tools and Frameworks 42
2.5.1 Keras 43
2.5.2 Tensorflow 43
2.5.3 Pytorch 43
2.6 Summary 44

3 Person Re-Identification from Still Images **47**
3.1 Multi-Scale Feature Extraction for Person Re-Identification 49
3.1.1 Multi-Scale Siamese *(SMSNet)* Architecture 50
3.1.2 *SMSNet* Training 51
3.1.3 Experimental Evaluation 53
3.1.4 Limitations of the Approach 59
3.2 Hierarchical Classification for Person Re-Identification 60
3.2.1 *SCB* Training 61
3.2.2 Hierarchical Approach to Re-Identification 63
3.2.3 Experimental Evaluation 65
3.3 Summary 80

4 Temporal Attention Features for Video-Based Person Re-Identification 83
4.1 Sub-Network Architecture Details 86
4.1.1 Full-Body Pose Attention Network 86
4.1.2 Motion Pooling Network 87
4.1.3 Long-Short Term Memory Network 88
4.2 Experiments and Results 91

4.3 Summary 99

5 Occlusion Handling in Image-based Person Re-Identification 101

5.1 Synthetic Occlusion Generation for Training Deep Neural Network Models 103

5.2 Occlusion Handling in Image Frames 104

5.2.1 *OHGAN*-based Reconstruction 104

5.2.2 *Autoencoder*-based Reconstruction 107

5.2.3 Reconstruction Results Using *OHGAN* and *Autoencoder* 108

5.2.4 Fine-Tuning Reconstruction Results with *DCGAN* 110

5.2.5 Re-Identification Results Using Baseline Networks 113

5.3 Summary 119

6 Occlusion Handling in Video-Based Person Re-Identification 121

6.1 *Conv-LSTM*-based Occlusion Reconstruction 122

6.2 Reconstruction Results 125

6.3 Person Re-identification and Experimental Evaluation 128

6.4 Summary 137

7 Conclusions and Future Work 139

References 148

List of Publications 170

List of Figures

1.1 Images of same person captured using two different cameras at different times . . . 1
1.2 Scenario where spatial features extracted from non-sequential frames are used for re-identification . . . 10
1.3 Scenario where spatio-temporal features extracted from sequential frames are used for re-identification . . . 12
1.4 Re-identification scenario with occlusion present in non-sequential image frames . . . 13
1.5 Re-identification framework for occlusion handling in sequential frames 15

3.1 Surveillance setup for re-identification . . . 49
3.2 Insight view of the proposed Siamese Multi-scale Network (*SMSNet*) architecture. The first layer of convolution is unaffected by dilation parameters. All other layers are dilated with rates 1, 2, and 3, and feature aggregation has been done after each convolution layer in form of concatenation. The feature difference is computed after the fourth convolution layer. . . . 50
3.3 Overall framework of the re-identification approach . . . 52
3.4 Range of three-fold cross-validation accuracy for various combinations of parameters η and γ corresponding to different the data sets by setting different initial weights of the network . . . 55
3.5 Box and whiskers plot showing the performance of the proposed approach after five different times of run on four data sets i.e., *VIPeR*, *CUHK_01*, *CUHK_03*, and *Market1501* . . . 56
3.6 Cumulative matching characteristic curves showing improvement in re-identification accuracy with rank for the different approaches corresponding to: (a) *CUHK_01*, (b) *CUHK_03*, (c) *Market1501*, and (d) *VIPeR* data sets . . . 58

3.7 A block diagram of the proposed hierarchical approach to person re-identification 60
3.8 Re-Identification framework applied on the reduced gallery set 62
3.9 Siamese Convolution Box (*SCB*) 62
3.10 Range of five-fold cross-validation accuracy for various combinations of parameters η and γ corresponding to different the data sets by setting different initial weights of the network 66
3.11 Elbow Curves for (a) *VIPeR* and (b) *CUHK_03* data 67
3.12 Variation of the cluster mapping accuracy with increment in the number of clusters for *CUHK_03* data 69
3.13 Accuracy and standard deviation of accuracy values on the four test sets constructed from (a) *CUHK_01*, (b) *CUHK_03*, (c) *Market1501*, (d) *VIPeR*, and (e) *IIT(BHU) Re-Identification* Data 71
3.14 Comparative performance analysis of the different re-identification approaches by means of Cumulative Matching Characteristic curves on (a) *CUHK_01*, (b) *CUHK_03*, (c) *Market1501*, (d) *VIPeR*, and (e)*IIT(BHU) Re-Identification* Data 73
3.15 Cumulative Matching Characteristic curves showing improvement in re-identification accuracy with the rank of the proposed algorithm with and without the clustering phase for the (a) *CUHK_01*, (b) *CUHK_03*, (c) *Market1501* data sets 75
3.16 Percentage accuracy and the standard deviation of the accuracy values on 100 test samples of the *CUHK_01*, *CUHK_03*, and *Market1501* data sets obtained by training our model with different initialization parameters five times 76
3.17 Cumulative Matching Characteristic curves showing improvement of re-identification accuracy with rank on a test set with similar clothing conditions with and without considering the cluster determination and mapping phases 78

4.1 A block diagram of the re-identification approach 85
4.2 Feature maps generated at the intermediate layers of (a) *FPAN*, (b) *MPN*, (c) *RNN*, (d) *T-MAN* 90
4.3 Rank 1 accuracy for different combinations of RNN Models (i.e., *Simple RNN*, *LSTM*, and *GRU*) 91
4.4 Rank 1 accuracy obtained by executing our ensemble *T-MAN* model three times along with the standard deviation 96

5.1 Overall framework of the re-identification approach: (a) Occlusion reconstruction, and (b) Re-identification 102
5.2 An original frame from the *CUHK_03* data and the corresponding occluded frames generated by adding varying degrees of synthetic occlusion . . . 104
5.3 Occlusion Handling GAN (*OHGAN*) architecture 105
5.4 An insight view of an *Autoencoder* 107
5.5 Images in the first row represent the synthetically occluded samples while the respective images in the second and third row show the generated images from the *OHGAN* and *Autoencoder* 110
5.6 Sample ground-truth unoccluded non-sequential frames (1^{st} row), frames with synthetic occlusion (2^{nd} row), reconstruction using *Autoencoder* (3^{rd} row), and fine-tuning using *DCGAN* (4^{th} row) 112
5.7 Comparison of Rank 1 accuracy of (i) *SCB*, (ii) *Autoencoder+DCGAN+SCB*, (iii) *Autoencoder+DCGAN+SCB (part-based)* 119

6.1 Overall framework of the proposed video-based re-identification approach in the presence of occlusion 122
6.2 Insight of *Conv-LSTM* Model 124
6.3 First and second rows show unoccluded and synthetically occluded frames from the *IIT (BHU)* Data set, while third and fourth rows present the reconstructed frames from the *Conv-LSTM* model and the corresponding fine-tuned frames through *DCGAN* 126
6.4 Box and whiskers plot showing the stability of the reconstruction model *Conv-LSTM+DCGAN* on the synthetically occluded test sets of *IIT (BHU) Re-identification* data and *PRID2011* data 127
6.5 Comparative study in terms of rank-based accuracy of *T-MAN* with *SCB* and the approaches [1–3] using synthetically occluded (a) *IIT (BHU) Re-identification* data and (b) *PRID2011* data 130
6.6 Comparison of robustness of all four proposed methods *M1*, *M2*, *M3*, and *M4* on various test sets constructed from the (i) video-based *PRID2011* data, and (ii) image-based *Market-1501* data corrupted with 20% synthetic occlusion 135
6.7 Comparison of stability of the Deep Learning models used in the four proposed methods *M1*, *M2*, *M3*, and *M4* by training the models multiple times from scratch on the same data and observing the overall re-identification accuracy of the approaches on (a) video-based *PRID2011* data, and (ii) image-based *Market-1501* data corrupted with 20% synthetic occlusion 136

List of Tables

2.1 Summary of some recent approaches to person re-identification 29
2.2 Image-based re-identification data sets 39
2.3 Video-based re-identification data sets 41

3.1 Layer specification of each Siamese Multi-scale Network (*SMSNet*) . . . 51
3.2 Data set split information 54
3.3 Comparison of Rank 1 accuracy (in %) for 100 test_ids of our proposed approach with state-of-the-art techniques 57
3.4 Layer specification of the *SCB* network. Both the starting convolution layers are pooled with size 2×2 and the last two layers are exempted from pooling. ⋆ represents the concatenation of fully connected layers . 61
3.5 Average cross-validation accuracy of the proposed approach on different data sets and average response time (in milli-secs) 68
3.6 Comparative study of Rank 1 accuracy of different re-identification approaches on a test set of subjects with similar clothing conditions 79
3.7 Comparison of Rank 1 accuracy (in %) for 100 test_ids 80

4.1 Comparative results on *PRID-2011* data set for Ranks 1, 5 and 10 . . 92
4.2 Comparison results on *iLIDS-VID* data set for Ranks 1, 5 and 10 . . . 93
4.3 Comparison results on *MARS* data set for Ranks 1, 5 and 20 with Mean Average Precision (map) 94
4.4 Comparative analysis of different combinations of proposed models (*FPAN*, *MPN*, and *T-MAN*) on the *MARS* data set 95
4.5 Comparison of Rank 1 accuracy for Siamese Convolution Box *(SCB)* and Temporal Motion Aware Network *(T-MAN)* 97
4.6 Rank 1 accuracy of Siamese Convolution Box *(SCB)* (with and without clustering) and Temporal Motion Aware Network *(T-MAN)* on a data set with similar clothing conditions 98

4.7 Comparison of Rank 1 accuracy for simple *T-MAN* and *Hierarchical T-MAN* 99

5.1 Layer specification of the *OHGAN* architecture 106
5.2 Layer-wise configuration of the *Autoencoder* 108
5.3 Dice scores obtained from the two reconstruction models at different margin values for different data sets 109
5.4 Layer-wise specification of the *DCGAN* 111
5.5 Dice similarity coefficient (DSC) values for the reconstruction models . 112
5.6 Comparative results on *CUHK_01* data set for Ranks 1, 5, and 10 . . . 114
5.7 Comparative results on *CUHK_03* data set for Ranks 1, 5, and 10 . . . 115
5.8 Comparative results on *Market1501* data set for Ranks 1, 5, and 10 along with mean average precision (*map*) 116
5.9 Comparison of Rank 1 accuracy of *OHGAN+PCB*, *Autoencoder+DCGAN+PCB*, and *Autoencoder+DCGAN+SCB* on synthetically occluded samples generated from *Partial ReID*, *Partial iLIDS*, and *IIT(BHU) Re-identification* sets 118

6.1 Layer-wise specification of the *Conv-LSTM*. Here, ConvLSTM2d_i represents the i^{th} layer of the model 124
6.2 Dice score for the reconstruction model (*Conv-LSTM*) at different margin values for the *IIT (BHU) Re-identification* and *PRID2011* data set . . 127
6.3 Comparative study of Rank 1 accuracy of the different methods on the occlusion reconstructed images by *ConvLSTM+DCGAN* for the *IIT(BHU) Re-identification* and *PRID2011* data sets 129
6.4 Comparative study of Rank 1 accuracy of *T-MAN* (with and without reconstruction) for the synthetically occluded *IIT(BHU) Re-identification* and *PRID2011* data sets 131
6.5 Comparative study of Rank 1 accuracy of all the proposed approaches on *iLIDS-VID*, *PRID2011*, and *IIT(BHU) Re-identification* data sets . 133
6.6 Comparative study of Rank 1 accuracy of all the proposed approaches on *CUHK_01*, *CUHK_03*, and *Market-1501* data sets 134

List of Symbols

Symbol	Description
Cam_i	Camera 'i'
f_i	Feature map i
$\mathbb{K}$	Cross-input neighborhood distance
$\mathcal{N}[g_i(x, y)]$	Neighborhood around pixel (x, y)
$\mathcal{C}_i$	Cluster centre
S_t	Subject 't'
$sim(A, B)$	Similarity score between 'A' and 'B'
K	Number of clusters
$\mathcal{K}$	Number of top matching clusters
Ci	Hyper Parameter Configuration
η	Learning rate
γ	Weight decay factor
μ	Mean
σ	Standard Deviation
b_c^t	Average of 't' feature vectors
m_c	Attention feature vector of clip 'C'
W_o	Weight matrix
$g_{i,a}$	Ground-Truth for sample (i,a)
$p_{i,a}$	Prediction for sample (i,a)
L_G	Generator loss
$G_e(I)$	Generated reconstructed image

Abbreviations

Abbreviation	**Description**
CNN	Convolutional Neural Network
LSTM	Long-short term memory
GAN	Generative Adversarial Network
NN	Neural Network
RNN	Recurrent Neural Network
GRU	Gated Recurrent Unit
SMSNet	Siamese Multi-scale Network
SCB	Siamese Convolution Box
FPAN	Full-body Pose Attention Network
MPN	Motion Pooling Network
T-MAN	Temporal Motion Aware Network
CMC	Cumulative Matching Characteristic
STD	Standard Deviation
map	Mean Average Precision
DSC	Dice Similarity Co-efficient
$Conv-LSTM$	Convolutional LSTM
DCGAN	Deep Convolutional GAN
ReLU	Rectified Linear Unit

Abstract

Person re-identification refers to the process of finding one-one correspondences among images/videos of individuals captured by different cameras which may have overlapping/ non-overlapping fields of view. It plays a central role in tracking and monitoring crowd movement in public places, and hence it serves as an essential means for providing public security in surveillance sites. In this thesis, we target to come up with plausible approaches to Computer Vision-based person re-identification that can be conveniently deployed in surveillance setups where the movement of multiple persons is monitored by a network of cameras. In Chapter 1 of the thesis, we introduce the problem of re-identification, the challenges involved, along with the motivation of the present work, and main contributions of the thesis with highlights. Next, in Chapter 2, we present a thorough literature survey on person re-identification starting from the traditional contextual and non-contextual approaches to the modern Deep Learning-based approaches. In this chapter, we also present a thorough insight of the trend of research in the domain person re-identification by highlighting the summary and limitations of the recently published work in tabular form, from where we figure out the scopes for further research in this area.

It has been observed from the literature survey that the initial approaches to image-based person re-identification mostly consider color-based appearance descriptors for matching, whereas the modern approaches employ deep features to make the prediction more accurate and robust. While the initial approaches are passive and not so

reliable in the presence of varying lighting conditions or varying scales of the captured images, the modern Deep Learning approaches suffer from the use of large number of parameters that makes the process time-intensive specially if the gallery set is very large. The use of multi-scale features for person re-identification or fusion of the passive methods with Deep Neural Network-based methods is expected to significantly improve the overall effectiveness of re-identification, which we have studied in Chapter 3. Here, we propose two approaches to image-based person re-identification that deal with the extraction of effective spatial features from images through the use of (i) a multi-scale feature generation technique, and (ii) a hierarchical combination of color-based and Deep Siamese network-based features. We make a thorough comparative study among these two proposed techniques and also other state-of-the-art techniques and observe that both these outperform the existing approaches in terms of accuracy. Also, among the two, the second approach has been seen to provide a more consistent performance across different data sets and is less time-intensive due to following a hierarchical classification scheme.

Although there exist several re-identification techniques that work with videos/set of sequential frames, these all depend on a single model prediction. However, since video data sets are less extensive compared to image data sets, prediction from a single model may not be reliable. Hence, we propose to employ an ensemble of recurrent network models for the prediction so that the different spatio-temporal aspects of the motion data can be exploited for re-identification. Our proposed ensemble architecture is discussed in depth in Chapter 4, which combines the predictions from a Full-Body Pose Attention Network, a Motion Pooling Network, and a Long-Short Term Memory Network to re-identify an individual for a set of gallery subjects. Through extensive experiments and comparative study, we observe that fusion of the spatio-temporal information extracted by these three sub-networks helps in performing accurate re-identification from video data. We also observe that the use of spatial features alone

is also not so effective in situations where subjects are engaged in some sequential activities like walking, running, etc., and also in situations where subjects have almost similar clothing conditions.

The images/videos captured by the cameras in a surveillance zone are usually corrupted with occlusion caused by other static/dynamic objects present in the scene. To the best of our knowledge, although there exist a few Deep Learning-based occlusion reconstruction strategies in the context of person re-identification, none of these consider occlusion reconstruction and re-identification as two separate modules. Rather, these methods train a single Deep Neural Network to perform re-identification directly from the input occluded frames. It appears that the effectiveness of these approaches can be improved by training two separate dedicated Deep Neural Network architectures for occlusion reconstruction and re-identification and stacking them during deployment as a single end-to-end model. In Chapter 5, we have proposed two such improved techniques that reconstruct the occluded frames by employing Deep Neural Network generators, one of which is based on *UNet+DCGAN* with skip connections between the convolution and the deconvolution layers, while the other is based on *Autoencoder+DCGAN* without any skip connections. Following the reconstruction phase, another Deep Learning classifier is used for re-identification. We make a rigorous comparative study between the two proposed techniques and observe that the network formed by stacking *Autoencoder+DCGAN* performs the best between the two. Classification of the reconstructed images using a Siamese Network-based classifier shows that our proposed method outperforms the existing person re-identification techniques working with occluded sequences.

It may also be noted that, none of the existing techniques that handle occlusion in the person re-identification task are capable of exploiting the available spatio-temporal information if the input is a video sequence rather than a set of non-sequential frames. Due to relying on spatial pixel-based information only, the reconstruction quality of

these existing methods is poor in case a frame of the input sequence is heavily occluded. This limitation of existing techniques can be overcome by effectively utilizing the spatio-temporal information present in the adjacent sequential frames of a video sequence while making prediction about the missing/occluded frames, which we have considered in Chapter 6. Specifically, we propose an algorithm for occlusion reconstruction from videos by employing a *Conv-LSTM*-based generator and a *DCGAN*-based fine-tuner. The reconstruction and re-identification results given by our method on video data sets corrupted with occlusion are quite good and also outperform the related approaches for most experimental settings.

It may be noted that apart from *PRID2011* and *iLIDS-VID* data, most existing re-identification data sets do not consist of frames with a sequence of activities. To test the effectiveness of diverse video-based re-identification data sets, we construct another indoor data in our laboratory with frontal walking videos from 41 subjects and use it to evaluate the performances of the different approaches proposed in the Chapters 3-5 as well as for carrying out detailed comparative studies. This data set has been termed as the *IIT (BHU) Re-identification* data set and it has been made publicly available to the research community for further comparison. In Chapter 6, we also conduct an experiment to present an unified interpretation of results of all the approaches discussed in Chapters 3-6 using both the image-based and video-based occluded re-identification data sets. Our observation is that the approach proposed in Chapter 5 is most suited to carry out re-identification from non-sequential frames, while that proposed in Chapter 6 is suited for dealing with sequential frames captured by surveillance cameras in most real-life surveillance sites. In a more constrained setup, where clean input images/videos of a target subject are available, the approaches discussed in Chapters 3 and 4 can be conveniently used. Each of our trained models has been made publicly available to the research community for further comparative studies. Finally, in Chapter 7, we conclude the thesis and give insights to some future directions of work in the area of

person re-identification.

Keywords: *Image and Video-based Person Re-identification, Siamese Convolution Box, Temporal Motion Aware Network, Generative Modeling, Occlusion Reconstruction, Autoencoder, Convolutional LSTM*