HUMAN RE-IDENTIFICATION USING SIAMESE CONVOLUTIONAL NEURAL NETWORK ON NVIDIA GEFORCE RTX 2060

ELAVARASAN A/L RAJATHURAI

A project report submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Engineering (Electronic and Telecommunication)

School of Electrical Engineering
Faculty of Engineering
Universiti Teknologi Malaysia

FEBRUARY 2021

DEDICATION

Parents who raised me up

Sister who were raised with me

Friends who shared their shoulders with me

Supervisors who lead me

Humanity

ACKNOWLEDGEMENT

Many life lessons are learned throughout the completion of this project. A warm thank you to all the souls and hearts that made me believe that a simple help can mean a lot to someone.

Not forgetting the important person in this process, I would like to express my gratitude and full appreciation to my supervisor Dr. Muhammad Ariff Baharudin for spending his precious time to assist me throughout the entire project. He has been a best guide to me in completing the project. Besides, I would like to thank Dr. Usman Ullah Sheikh for his continuous support. Thank you for his patience and deep knowledge sharing with me

Friends are always the backup plans for all failed attempts. Their level of moral support is incomparable; thus, mere thanks is not sufficient. Your kindness will not be forgotten forever. I would also like to extend my gratitude to my current employer, Panasonic AVC Networks Kuala Lumpur to allow me further my studies without any obstacles.

Finally, the deepest appreciation to my parents, and my sister, that has been supporters of not only my work but for everything I did throughout my life.

ABSTRACT

Human reidentification in multiple cameras with disjoint views is to match a pair of humans appearing in different cameras with non-overlapping views. Human reidentification has been extensively studied in recent years because it plays a significant role in many applications such as human tracking and video retrieval. However, human re-identification is a challenging task due to varying factors such as color, pose, viewpoint, lighting conditions, low resolution and partial occlusion. Most of the existing methods in handling human re-identification task are based on various handcrafted features and metric learning. However, hand-crafted features method requires expert knowledge and requires a lot of time to tune the features and metric learning methods are not powerful enough to exploit the nonlinear relationship of samples. The main objective of this thesis is to implement Siamese Convolutional Neural Network (SCNN) for person re-identification task in multiple cameras on the NVIDIA® GeForce RTX™ 2060 platform, including person detection. This continuous with validation of the applicability of SCNN and compare with existing techniques. In this work, global and local features of human images are extracted from SCNN. The proposed SCNN consists of two identical Convolution Neural Networks with common parameters that can automatically learn hierarchical feature representations from image pixels directly which has advantages than the hand-crafted design and metric learning method. Experiments were conducted with CUHK02 offline database with non-overlapping cameras. The proposed technique demonstrated a person re-identification using SCNN on the NVIDIA® GeForce RTX™ 2060 platform.

ABSTRAK

Pengenalpastian manusia dalam beberapa kamera dengan pandangan yang tidak sama adalah untuk memadankan sepasang manusia yang muncul dalam kamera yang berbeza dengan pandangan yang tidak bertindih. Pengenalpastian manusia telah banyak dikaji dalam beberapa tahun kebelakangan ini kerana memainkan peranan penting dalam banyak aplikasi seperti penjejakan manusia dan pengambilan video. Walau bagaimanapun, identifikasi semula manusia adalah tugas yang mencabar kerana pelbagai faktor seperti warna, pose, sudut pandang, keadaan pencahayaan, resolusi rendah dan oklusi separa. Sebilangan besar kaedah yang ada dalam menangani tugas mengenal pasti semula manusia adalah berdasarkan pelbagai ciri handcraft dan pembelajaran metrik. Walau bagaimanapun, kaedah ciri handcraft memerlukan pengetahuan pakar dan memerlukan banyak masa untuk menyesuaikan ciri dan kaedah pembelajaran metrik tidak cukup kuat untuk mengeksploitasi hubungan sampel yang tidak linear. Objektif utama thesis ini adalah untuk melaksanakan Siamese Convolutional Neural Network (SCNN) untuk tugas pengenalan semula orang dalam beberapa kamera pada platform NVIDIA® GeForce RTX ™ 2060, termasuk pengesanan orang. Ini berterusan dengan pengesahan penerapan SCNN dan bandingkan dengan teknik yang ada. Dalam thesis ini, ciri global dan tempatan dari gambar manusia diekstrak dari SCNN. SCNN yang dicadangkan terdiri daripada dua Convolutional Neural Network yang serupa dengan parameter umum yang secara automatik dapat mempelajari perwakilan ciri hierarki dari piksel gambar secara langsung yang mempunyai kelebihan daripada reka bentuk handcraft dan kaedah pembelajaran metrik. Eksperimen dilakukan dengan data CUHK02 secara offline dengan kamera yang tidak bertindih. Teknik yang dicadangkan menunjukkan pengenalan semula seseorang menggunakan SCNN pada platform NVIDIA® GeForce RTX ™ 2060.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

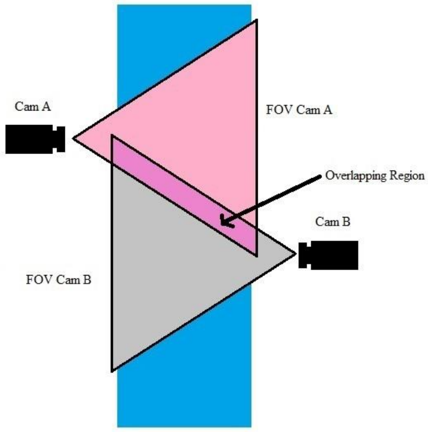
1.1 Overview of Human Re-identification

The demand for the installation of closed-circuit television (or CCTV) camera networks has increased recently to address variety of security issues. CCTV camera networks are being installed at home, office, shopping centers, sport centers and airports. However, it is not an easy task for human operators to continually observe CCTV over multiple cameras especially when tracking human of interest. Hence, a computer vision system is required to assist human operators in recognizing individual humans throughout an entire camera network. The problem of observing a human of interest across multiple camera networks is known as a human reidentification problem [1–6].

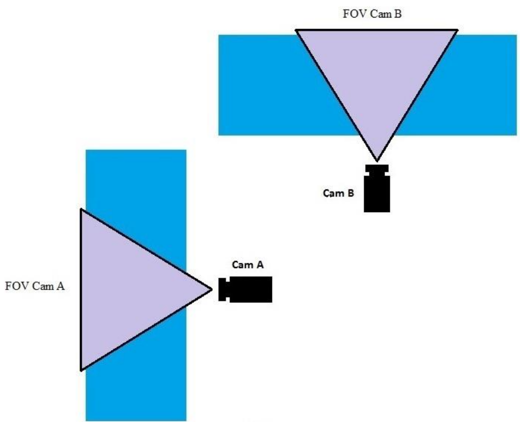
Human reidentification divided into two categories which are appearance-based approach and biometric [7]. Example of biometric approaches for reidentification are face [8], gait [9], iris [10] and fingerprint recognition [11]. However, iris and fingerprint recognition are not suitable for reidentification at wide area video surveillance field of view because recognition of iris and fingerprint requires human cooperation in the monitored environment or high-resolution images, which are not available in common surveillance systems [12]. Compared to iris and fingerprint recognition, gait and face recognition do not require human cooperation and can operate without interrupting or interfering with the human's activity [13]. However, face and gait recognition will only achieve good performance of recognition when some conditions and constraints are achieved. Unfortunately, some of these constraints are not satisfied by most deployed surveillance systems [7].

Biometric approaches are mainly dependent on the camera view and orientation of the human with the camera. Based on the reasons above, biometric approaches are not very suitable for human reidentification in surveillance systems.

Appearance based approaches for human reidentification are more suitable for wide area video surveillance systems because it is less constrained than biometric approaches and more adapted to video surveillance requirements such as does not require human cooperation, low resolution images and no specific conditions and constraint are required [7]. Human reidentification with appearance-based approaches is a central task in surveillance system which is used to match a pair of humans appearing in different cameras with non-overlapping views [14]. The difference between general camera setup with overlapping views and non-overlapping views are shown in Figure 1.1. In most surveillance systems, cameras with nonoverlapping views are applied because it is impossible to cover all the area of interest by using multiple overlapping cameras due to economic and computational reasons. Surveillance over wide-areas such as area of law enforcement, airport and office buildings requires a network of cameras that are sparsely distributed without overlapping field of views. Human reidentification has been extensively studied in recent years due to its various applications such as in surveillance systems with nonoverlapping views.



(a)



(b)

Figure 1.1 Camera network with (a) overlapping views and (b) non-overlapping views.

1.2 Problem Statement

Human reidentification problem is a challenging task and received a great attention of researchers in recent years. In most practical scenarios, the gap between camera views in a surveillance system is quite large due to economic and computational reason. Since images obtained from surveillance cameras have low resolution region of interest (centering humans) which is around 128x48 pixels because taken from long distances, human biometric information such as face and gait are not suitable to be used for reidentification purpose. Therefore, appearance of human becomes an important feature to solve reidentification task. Moreover, appearance of a human varies across multiple cameras due to difference in viewpoint, pose and illumination. Moreover, low resolution image has fewer useful details for classification and especially in non-overlapping views [15–21]. Thus, a better approach is needed for handling the low-resolution issue to increase the accuracy and speed of human reidentification task.

1.3 Objective

Based on the current issues surrounding human reidentification across multiple cameras, the two main objectives of this research can be expressed as follows:

1. To implement SCNN for human reidentification task in multiple cameras, including human detection.

2. To validate the applicability of SCNN in NVIDIA® GeForce RTX™ 2060 and compare with existing techniques.

1.4 Scope

This research focuses on developing a human reidentification system. Hence in this research:

1. The process of human detection is in the scope of this work.

2. The common challenges such as illumination and viewpoint are considered in the proposed human reidentification system.

3. The human reidentification system is prototyped on a NVIDIA® GeForce RTX™ 2060.

4. The proposed work is based on Siamese Convolution Neural Network.

REFERENCES

1. Metzler, J. Two-stage appearance-based re-identification of humans in low-resolution videos. 2012 IEEE International Workshop on Information Forensics and Security (WIFS). 2012. 19–24.
2. Metzler, J. Appearance-Based Re-identification of Humans in LowResolution Videos Using Means of Covariance Descriptors. 2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance. 2012. 191–196.
3. Zheng, W.-S., Gong, S. and Xiang, T. Person Re-identification by Probabilistic Relative Distance Comparison. Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society. 2011, CVPR ’11. 649–656.
4. Zheng, W. S., Gong, S. and Xiang, T. Reidentification by Relative Distance Comparison. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013. 35(3): 653–668.
5. Liu, H., Qin, L., Cheng, Z. and Huang, Q. Set-based classification for person re-identification utilizing mutual-information. 2013 IEEE International Conference on Image Processing. 2013. 3078–3082.
6. Yuan, G., Zhang, Z. and Wang, Y. Enhancing Person Re-identification by Robust Structural Metric Learning. 2013 Seventh International Conference on Image and Graphics. 2013. 453–458.
7. Souded, M. People detection, tracking and re-identification through a video camera network. Theses. Université Nice Sophia Antipolis. 2013. URL https://tel.archives-ouvertes.fr/tel-00913072.
8. Zhao, W., Chellappa, R., Phillips, P. J. and Rosenfeld, A. Face recognition: A literature survey. ACM computing surveys (CSUR), 2003. 35(4): 399–458.

9. Lee, L. and Grimson, W. E. L. Gait analysis for recognition and classification. Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on. IEEE. 2002. 155–162.
10. Ma, L., Wang, Y. and Tan, T. Iris recognition based on multichannel Gabor filtering. Proc. Fifth Asian Conf. Computer Vision. 2002, vol. 1. 279–283.
11. Jea, T.-Y. and Govindaraju, V. A minutia-based partial fingerprint recognition system. Pattern Recognition, 2005. 38(10): 1672–1684.
12. Yang, J., Shi, Z. and Vela, P. A. Person Reidentification by Kernel PCA Based Appearance Learning. 2011 Canadian Conference on Computer and Robot Vision. 2011. 227–233.
13. Bashir, K., Xiang, T. and Gong, S. Gait recognition without subject cooperation. Pattern Recognition Letters, 2010. 31(13): 2052 – 2060. Metaheuristic Intelligence Based Image Processing.
14. Hirzer, M., Beleznai, C., Roth, P. M. and Bischof, H. Image Analysis: 17th Scandinavian Conference, SCIA 2011, Ystad, Sweden, May 2011. Proceedings, Berlin, Heidelberg: Springer Berlin Heidelberg, chap. Person Re-identification by Descriptive and Discriminative Classification. 2011, 91–102.
15. Zheng, W.-S., Gong, S. and Xiang, T. Associating Groups of People. Proceedings of the British Machine Vision Conference. BMVA Press. 2009. 23.1–23.11.
16. Figueira, D., Bazzani, L., Minh, H. Q., Cristani, M., Bernardino, A. and Murino, V. Semi-supervised Multi-feature Learning for Person Reidentification. Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on. 2013. 111–116.
17. Sivic, J., Zitnick, C. L. and Szeliski, R. Finding People in Repeated Shots of the Same Scene. Proceedings of the British Machine Vision Conference. BMVA Press. 2006. 93.1–93.10.
18. Bird, N. D., Masoud, O., Papanikolopoulos, N. P. and Isaacs, A. Detection of Loitering Individuals in Public Transportation Areas. IEEE Transactions on Intelligent Transportation Systems, 2005. 6(2): 167–177.

19. Hamdoun, O., Moutarde, F., Stanciulescu, B. and Steux, B. Person Re-identification In Multi-camera System by Signature Based On Interest Point Descriptors Collected on Short Video Sequences. Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on. 2008. 1–6.
20. Cong, D.-N. T., Khoudour, L., Achard, C., Meurie, C. and Lezoray, O. People Re-identification by Spectral Classification of Silhouettes. Signal Processing, 2010. 90(8): 2362 – 2374. Special Section on Processing and Analysis of High-Dimensional Masses of Image and Signal Data.
21. BÄEk, S., Corvee, E., Bremond, F. and Thonnat, M. Boosted Human Re-identification Using Riemannian Manifolds. Image and Vision Computing, 2012. 30(6âA ¸S7): 443 – 452.
22. Gheissari, N., Sebastian, T. B. and Hartley, R. Person Reidentification Using Spatiotemporal Appearance. Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. 2006, vol. 2. 1528–1535.
23. Farenzena, M., Bazzani, L., Perina, A., Murino, V. and Cristani, M. Person Re-identification by Symmetry-driven Accumulation of Local Features. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. 2010. 2360–2367.
24. Forssen, P. E. Maximally Stable Colour Regions for Recognition and Matching. Computer Vision and Pattern Recognition, 2007. CVPR ’07. IEEE Conference on. 2007. 1–8.
25. Wang, X., Doretto, G., Sebastian, T., Rittscher, J. and Tu, P. Shape and Appearance Context Modeling. Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. 2007. 1–8.
26. Ma, B., Su, Y. and Jurie, F. BiCov: A Novel Image Representation for Person Re-identification and Face Verification. Proceedings of the British Machine Vision Conference. BMVA Press. 2012. 57.1–57.11.
27. Hirzer, M., Roth, P. M., Köstinger, M. and Bischof, H. Computer Vision – ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VI, Berlin, Heidelberg: Springer Berlin Heidelberg, chap. Relaxed Pairwise Learned Metric for Person Reidentification. 2012, 780–793.

28. Ma, B., Su, Y. and Jurie, F. Covariance Descriptor based on Bio-inspired Features for Person Re-identification and Face Verification. Image and Vision Computing, 2014. 32(6-7): 379–390.
29. K. B. Low,. Human re-identification based on Siamese convolution neural network with decision fusion of global and local features. 2017.
30. Lianyang Ma, Person Re-Identification Over Camera Networks Using Multi-Task Distance Metric Learning. 2014, IEEE Transactions on Image Processing
31. Vidya Dharan,. A Face Recognition System Accelerated On Embedded Graphics Processing Unit (GPU) Enabled By Compute Unified Device Architecture (CUDA). 2017
32. Understanding Categorical Cross-Entropy Loss, Binary Cross-Entropy Loss, Softmax Loss, Logistic Loss, Focal Loss and all those confusing names. https://gombru.github.io/2018/05/23/cross_entropy_loss/ (Accessed on 07/23/2020)
33. Gentle Introduction to the Adam Optimization Algorithm for Deep Learning. https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/#:~:text=Adam%20is%20a%20replacement%20optimization,sparse%20gradients%20on%20noisy%20problems. (Accessed on 07/23/2020)
34. Li, W. and Wang, X. Locally Aligned Feature Transforms across Views. Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. 2013.
35. ARM Information Center. http://infocenter.arm.com/help/index.jsp?topic=/com.arm.doc.dui0555a/BEIGDEGC.html. (Accessed on 10/15/2016).
36. Building a One-shot Learning Network with PyTorch. https://towardsdatascience.com/building-a-one-shot-learning-network-with-pytorch-d1c3a5fafa4a (Accessed on 07/23/2020).
37. A Comprehensive Guide to Convolutional Neural Networks https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

38. Computer Vision for Human–Machine Interaction, Computer Vision and Pattern Recognition, 2018.
39. What is Rectified Linear Unit (ReLU)? | Introduction to ReLU Activation Function https://www.mygreatlearning.com/blog/relu-activation-function/
40. ReLu, https://deepai.org/machine-learning-glossary-and-terms/relu
41. Introduction to ReLU Activation Function, https://www.mygreatlearning.com/blog/relu-activation-function/
42. A Gentle Introduction to Pooling Layers for Convolutional Neural Networks, https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/
43. Bromley, J., Guyon, I., Lecun, Y., SÃd’ckinger, E. and Shah, R. Signature Verification using a "Siamese" Time Delay Neural Network. In NIPS Proc. 1994.
44. Chopra, S., Hadsell, R. and LeCun, Y. Learning a Similarity Metric Discriminatively, with Application to Face Verification. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05) - Volume 1 - Volume 01. IEEE Computer Society. 2005, CVPR ’05. 539–546.