Faculty of Electronic and Computer Engineering

WiFi MAC ADDRESS TAGGING ASSISTED FAST SURVEILLANCE VIDEO RETRIEVAL SYSTEM

Tan Kien Leong

Master of Science in Electronic Engineering

2020

WiFi MAC ADDRESS TAGGING ASSISTED FAST SURVEILLANCE VIDEO RETRIEVAL SYSTEM

TAN KIEN LEONG

A thesis submitted
in fulfillment of the requirements for the degree of Master of Science
in Electronic Engineering

Faculty of Electronic and Computer Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2020

DECLARATION

I declare that this thesis entitled “WiFi MAC Address Tagging Assisted Fast Surveillance Video Retrieval System” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : ....

Name : ....

Date : ....

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Master of Science in Electronic Engineering.

Signature : ....

Supervisor Name : ....

Date : ....

DEDICATION

To my beloved father and mother

ABSTRACT

Conventional public safety surveillance video camera systems required 24/7 monitoring of security officers with video wall display installed in the control room. When a crime or incident is reported, all the recorded surveillance video streams nearby the incident area are played back simultaneously on video wall to help locate the target person. The security officers can fast forward the video playback to speed up the video search but it requires massive manpower if there are hundreds of video streams or multiple target persons required to be examined on the video wall. Even today with the Graphics Processing Unit (GPU) that is able to run the person search deep neural network model to automatic search for the target person from a large video database, it can take hours or even days to complete the search. This research aims to determine how to prioritize the surveillance camera video frames that need to be processed by the person search deep neural network model to reduce the time taken for getting the target person in the next camera (the cameras that may recorded the target person according to walkway topology). Thanks to the advancement in artificial intelligence, a person search deep neural network model trained to correctly match thousands of identical person can be used to automate the person search process. The person search matching process required the person in the image to be firstly detected before the matching can be carried out. Eight deep neural network based object detection models are re-trained on 55,272 labelled persons to determine the suitable object detection model that can be used to replace the person detection part of the person search model. As a result, applying Model 3 (Darkflow) for person detection is found to be able to provide reasonable speed/accuracy trade-off (0.62 mAP and 0.04s mean inference time). To further reduce the required time of automated person search without having to scale up the computing hardware, additional metadata (WiFi MAC address of smartphone) collected during the occurrence of the incident can be used to prioritize the retrieving of surveillance video frames for subsequent person search. Three ways of retrieving surveillance video are compared, in term of time taken for getting the target person, with a constructed testbed in UTeM. The developed WiFi sniffer enabled surveillance camera, with 3-stage WiFi frame inspection and the use of collected WiFi signal strength for filtering, is able to tag the collected WiFi MAC addresses to the surveillance video frames according to the time of the MAC address is sniffed. Using the formulated mathematical model, the proposed WiFi MAC address tagging assisted fast surveillance video retrieval method performs 9.6 times better in single person search and 6.2 times better in multiple persons search provided the WiFi MAC address of the target's smartphone is sniffed by the WiFi sniffer of the surveillance camera. Based on these results, the proposed fast video retrieval system with MAC address tagging is proven to take less time to get target person in the next camera as compared to video retrieval system without MAC address tagging. Further research is needed to identify how to prioritize the WiFi MAC address searching when multiple WiFi MAC addresses are sniffed.

ABSTRAK

Sistem kamera pengawasan konvensional memerlukan pemantauan oleh pegawai keselamatan sepanjang hari di dalam bilik kawalan. Apabila jenayah atau insiden dilaporkan, semua rakaman video pengawasan yang dirakam berhampiran kawasan kejadian akan diulang semula secara serentak untuk mencari sasaran orang. Pegawai keselamatan boleh mempercepatkan masa ulangan video untuk mengenalpasti orang yang perlu disasarkan tetapi ia memerlukan tenaga kerja yang banyak jika terdapat ratusan rakaman video atau beberapa sasaran yang perlu dipantau. Walaupun Graphics Processing Unit (GPU) dapat mengaplikasi model rangkaian neural untuk melaksanakan proses pencarian orang secara automatik, tetapi ia masih mengambil masa yang panjang untuk proses pencarian orang jika terdapat ratusan video yang perlu diproses. Kajian ini bertujuan untuk menentukan bagaimana mengutamakan rangka video yang perlu diproses oleh model rangkaian neural bagi mengurangkan masa yang diperlu untuk mencari orang sasaran dalam kamera yang berikut (kamera yang mungkin dikunjung oleh sasaran berdasarkan topologi laluan). Seiring dengan pencapaian dan kemajuan dalam kecerdasan buatan, model rangkaian neural mampu melaksanakan proses pencarian orang secara automatik. Proses pemadanan orang memerlukan imej orang dikesan terlebih dahulu. Lapan model pengesanan objek berasaskan rangkaian neural yang mendalam dilatih semula pada 55,272 orang berlabel untuk menentukan model pengesanan objek yang sesuai seterusnya menggantikan bahagian pengesanan model carian tersebut. Keputusannya, Model 3 (Darkflow) diaplikasikan untuk pengesanan orang mampu memberikan keseimbangan antara kelajuan dan ketepatan dalam melaksanakan tugas (0.62 mAP dan 0.04s purata masa inferensi). Metadata tambahan (alamat MAC WiFi telefon pintar) yang dikumpulkan semasa kejadian boleh digunakan untuk mengutamakan pencarian rangka video yang berkaitan untuk pemprosesan imej berikutnya. Melalui cara ini, masa yang diperlukan untuk pencarian orang dapat dikurangkan tanpa perlu menaikkan taraf perkakasan komputer yang tersedia ada. Tiga cara telah dibandingkan dari segi masa untuk mendapatkan sasaran telah dijalankan pada kawasan ujian yang bertempat di UTeM. Snifer WiFi yang terletak dalam kamera pengawasan mempunyai fungsi tiga peringkat pemeriksaan rangka WiFi dan penggunaan kekuatan isyarat WiFi yang dikumpul untuk penapisan rangka WiFi. Ia digunakan untuk menanda alamat MAC WiFi yang telah dikumpul ke rangka video mengikut masa alamat MAC yang dikesan. Dengan penggunaan model matematik yang dicadangkan, kaedah pencarian video yang dibantu dengan alamat MAC WiFi dapat memberikan 9.6 kali lebih pantas dalam carian orang tunggal dan 6.2 kali lebih pantas dalam mencari orang untuk jumlah yang banyak, dengan andaian alamat MAC WiFi yang dimiliki oleh sasaran orang dapat dikesan oleh snifer WiFi. Kesimpulannya, kaedah pencarian video yang dibantu dengan alamat MAC WiFi terbukti mengambil masa yang lebih pendek berbanding dengan kaedah pencarian video yang tidak dibantu dengan alamat MAC WiFi. Kajian lanjutan diperlukan untuk mengenal pasti bagaimana untuk mengutamakan alamat MAC WiFi apabila terdapat banyak alamat MAC WiFi yang dikumpul.

ACKNOWLEDGEMENTS

I would first like to express my sincere gratitude to my supervisor, Associate Professor Dr. Lim Kim Chuan, for the continuous support of my master study and related research, for his patience, motivation, and immense knowledge.

I would also like to thank my co-supervisor, Associate Professor Dr. Soo Yew Guan, for his guidance in WebApp video surveillance dashboard interface development.

I would also like to thank Collaborative Research in Engineering, Science and Technology (CREST) and Recogine Technology Sdn Bhd for awarded me the Graduate Research Assistant Scholarship Program (GRASP) and provide me the experiment equipment used to build the testbed.

I would also like to thank my colleagues and lab technician in Research Lab 3, Faculty of Electronics and Computer Engineering of UTeM who have provided assistance when I was doing my research.

Finally, I must express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AI	-	Artificial Intelligence
AP	-	Average Precision
API	-	Application Programming Interface
cm	-	centimetre
CNN	-	Convolutional Neural Network
Conv2D	-	Two-dimensional convolution
CUDA	-	Compute Unified Device Architecture
cuDNN	-	CUDA Deep Neural Network
dBm	-	decibels (dB) with reference to one milliwatt
DS	-	Distribution System
FLOPs	-	Floating point operations
FN	-	False Negative
FP	-	False Positive
fps	-	frame per second
GB	-	Gigabyte
GHz	-	Gigahertz
GPU	-	Graphics Processing Unit
IANA	-	Internet Assigned Numbers Authority
ID	-	Identity
IEEE	-	Institute of Electrical and Electronics Engineers

IoU	-	Intersection over Union
IP	-	Internet Protocol
LTS	-	Long Term Support
m	-	Meter
MAC	-	Media Access Control
mAP	-	Mean Average Precision
Max	-	Maximum
Min	-	Minimum
mm	-	millimetre
no.	-	number
OpenCV	-	Open source Computer Vision
OS	-	Operating System
OUI	-	Organizational Unique Identifier
PIL	-	Python Imaging Library
PVC	-	Polymerizing Vinyl Chloride
RAM	-	Random Access Memory
RSSI	-	Receive Signal Strength Indication
s	-	seconds
SSID	-	Service Set Identifier
TP	-	True Positive
WEP	-	Wired Equivalent Privacy
WPA	-	WiFi Protected Access
WPA 2	-	WiFi Protected Access II

LIST OF PUBLICATIONS

Journal

1. Tan, K.L. and Lim, K.C., 2019. Fast Surveillance Video Indexing & Retrieval with WiFi MAC Address Tagging, *Indonesian Journal of Electrical Engineering and Computer Science*, 16(1), pp. 473-481.

Technical Report

1. Tan, K.L., Lim, K.C. and Tan, X.Y., Alvin, S., 2018. MAC Address tagged Fast Video Retrieval System across Multiple Cameras for Person Re-identification, *Recogine Technology Sdn Bhd*.

CHAPTER 1

INTRODUCTION

The WiFi MAC address tagging assisted fast surveillance video retrieval system with deep learning based person search will be firstly explained at the beginning of this chapter. The problem statement, research question, hypothesis, objectives and research scope of this research are presented subsequently. The contribution of this research is presented at the end of this sectionp.

1.1 Background

Surveillance camera systems have been widely used in public places in recent years for the purpose of public safety. Traditional type surveillance camera systems are monitored by security officers 24 hours per day. When an incident has taken place, all the recorded surveillance video streams nearby the incident area are playback simultaneously to help locate the target person. The security officer usually fast forward the video playback to speed up the video search.

Deep learning is one of the greatest achievements in the field of computer vision over the last decade. It outperforms the handcrafted methods in image classification (Krizhevsky et. al., 2012), object detection (Ren et. al., 2015), object segmentation (He et. al., 2017), and others. An automatic approach to search for a target person from the surveillance camera video database is by applying deep neural network model. A person search deep neural network model consists of two parts; the first part is person detector while the second part is person re-identification. The person detector is used to automate locate the persons inside an image. Red bounding boxes are drawn on the image to indicate where the persons are located (see Figure 1.1). The person feature inside the bounding box is extracted and passed to person re-identification model. It compares the similarity of person features of the target person with the person features of the persons appeared in recorded surveillance video frames. For example, given an input image which contains five target persons, the target persons are automatically labelled with a box and a unique ID as shown in Figure 1.2(a). The person re-identification model will search through the surveillance camera video database to find the images that contain either one of the target persons. An example of the result of person re-identification is as shown in Figure 1.2(b). Same ID will be assigned to the person detected in the recorded surveillance camera video frame if that person is detected and recognized as the same person in the input image. The percentage value displayed beside the ID is the percentage of similarity (in between the range of 0%-100%).

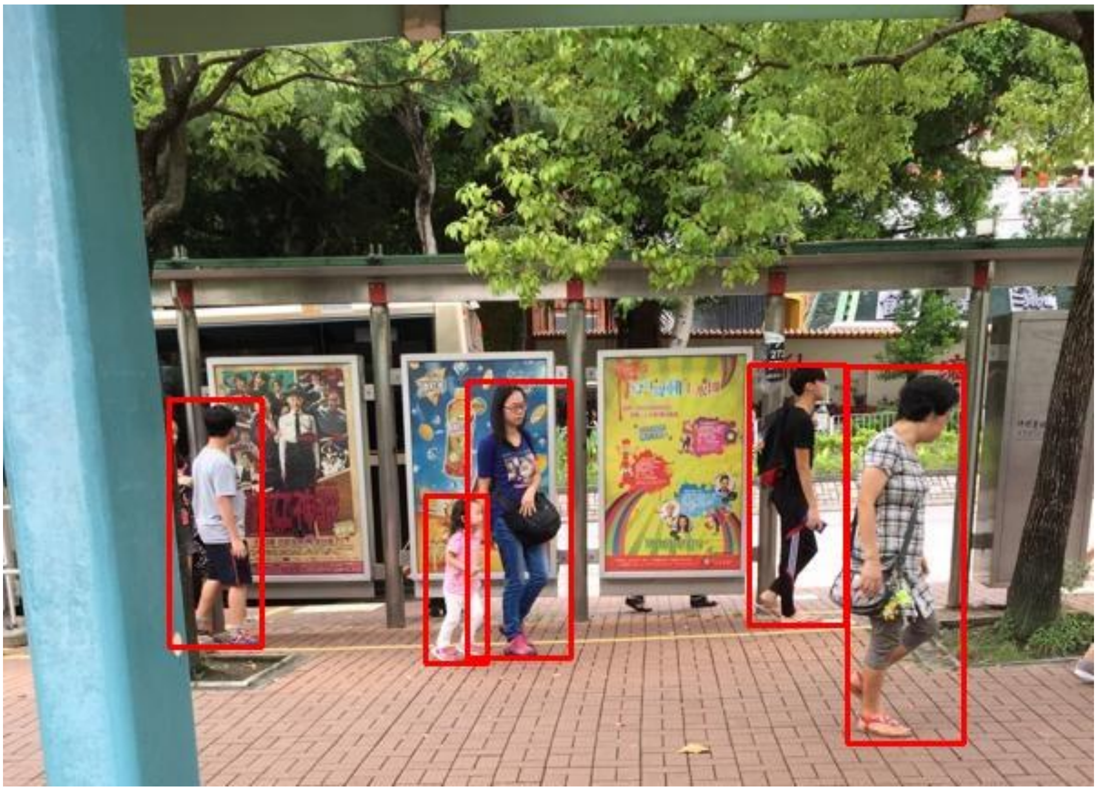

Figure 1.1: Example of automated person detection (CUHK-SYSU person search dataset, 2016)

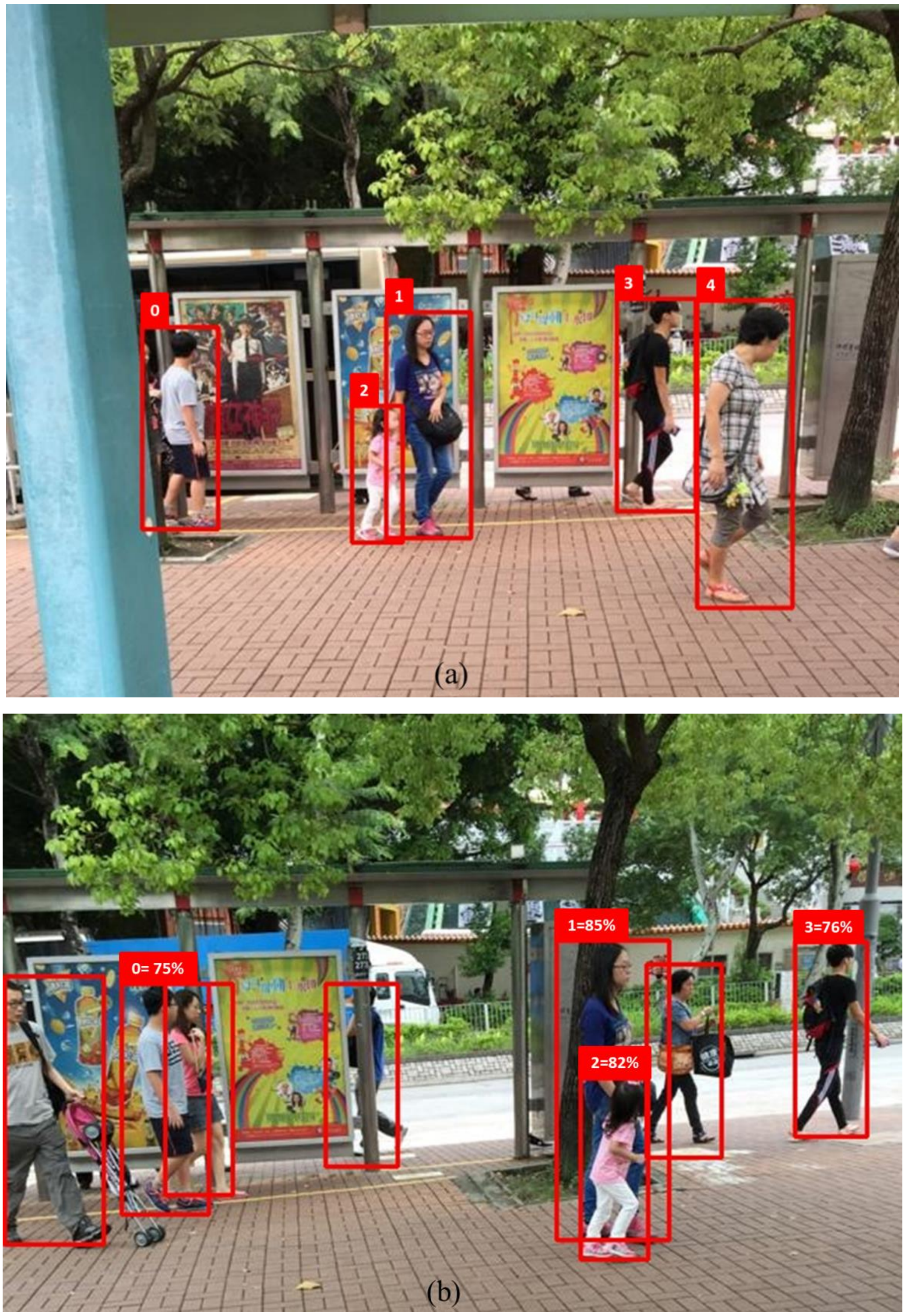

Figure 1.2: (a) The persons of input image are automatic labelled with a unique ID (b) The result of person re-identification (CUHK-SYSU person search dataset, 2016)

Nowadays, people carry their smartphone with WiFi turned on wherever they go in order for their smartphone to automatically connect to Internet when there is WiFi service available. When the WiFi of the smartphone is turned on, the smartphone will broadcast management frames known as probe request to discover all nearby access point. The probe request contains WiFi media access control (MAC) address which is a unique identifier assigned to the network interface card of the smartphone. The length of WiFi MAC address is six bytes and separated by colons. The MAC address is generated by using OUI (Organizationally Unique Identifier) number provided by IANA (Internet Assigned Numbers Authority). The MAC address is never encrypted even though the WiFi devices are connected to a WiFi network with security encryption enabled (Wired Equivalent Privacy (WEP), WiFi Protected Access (WPA), and WiFi Protected Access II (WPA2)). Besides probe request, there are other WiFi frames broadcast by smartphone that contain WiFi MAC address for example: disassociation, authentication, deauthentication, and so on. Since each smartphone has a unique WiFi MAC address, it can be used to track a person who carries a smartphone with WiFi turned on. Research work has been carried out using WiFi MAC address for WiFi tracking (Petre et. al., 2017) (Julien, 2015) (Xu et. al., 2013) (Musa et.al., 2012).

1.2 Problem statement

With rapid urbanization, IP cameras are almost everywhere in our daily life such as pedestrian walkways, road junctions, schools, markets and public transport which provide a greater level of public safety. However, these IP cameras generate large volume of video data across time. This becomes an issue when there is an incident that happens for example a lost child or crime suspect, and the surveillance camera video frames that contain the target person need to be retrieved in the shortest time to minimize further damage. Every second counts to avoid the situation from becoming more serious.

In the old days, human beings needed extensive time and intensive man-power to retrieve surveillance camera video frames that contained target person from a large video database. When an incident occurred, all the recorded surveillance video streams nearby the incident area were playback simultaneously to help locate the target person. The security officers would fast forward the video playback to speed up the video search but it required massive manpower if there were hundreds of video streams that needed to be monitored. The situation became more challenging to the security officers if there were multiple targets being tracked at the same time. Hence, an automatic approach to search for target person by applying deep learning model is applicable.

The speed and accuracy of a person search deep neural network model is mainly affected by the person detection part. A good person detector can provide more useful person features for the person re-identification model to perform person re-identification process (useful person features provide better accuracy in person re-identification). A deeper neural network model can extract more useful person features (higher accuracy) but at the same time it will consume more processing time (slower speed) (Huang et. al. 2016). Therefore, an appropriate person detector which provides reasonable speed and accuracy trade off need to be selected for person search deep neural network model.

Even today with the Graphics Processing Unit (GPU) that is able to run the person search deep neural network model to automatic search for the target person from a large video database, it can take hours or even days to complete the search. The video processing time can be reduced by using multiple desktop GPUs or a more powerful server grade GPU to run the person search deep neural network model, but this requires more investment on GPU and it is not cost-effective to allocate the GPU for every available camera. This has