AUTOMATED HUMAN RECOGNITION AND TRACKING FOR VIDEO SURVEILLANCE SYSTEM

BY

FADHLAN HAFIZHELMI BIN
KAMARU ZAMAN

A dissertation submitted in partial fulfilment of the requirements for the degree of Master of Science (Communication Engineering)

Kulliyyah of Engineering
International Islamic University
Malaysia

NOVEMBER 2010

ABSTRACT

Recent research in video surveillance system has shown an increasing focus on creating reliable systems utilizing non-computationally expensive technique for observing humans' appearance, movements and activities, thus providing analytical information for advanced human behavior analysis and realistic human modeling. In order for the system to function, it requires robust method for detecting and tracking human from a given input of video streams. In this thesis, a human detection technique suitable for video surveillance is presented which requires fast computations in addition of accurate results. The techniques proposed include adaptive frame differencing for background subtraction, contrast adjustment for shadow removal, and shape based approach for human detection. The tracking technique on the other hand uses correspondence approach. Event Based Video Retrieval (EBVR) system is also proposed for efficient surveillance data management and automated human recognition with unique ID assignment. Proposed human detection and tracking are integrated with EBVR and motion detection into a complete automated surveillance system called Active Vis Video Surveillance Analysis System (AVSAS) which produces good result and real-time performance especially in non-crowded scene. The EBVR system also proves to be able to handle automated human recognition with unique ID assignment accurately.

ملخص البحث

أظهرت الأبحاث الحديثة في مجال أنظمة مراقبة الفيديو زيادة في التركيز على إنشاء أنظمة موثوقة مستفيدة من التقنيات الحسابية غير المكلفة لمراقبة ظهور الأشخاص وحركاتهم وأنشطتهم. وبالتالي إعطاء معلومات تحليلية لنظام تحليل متقدم لسلوك الأشخاص ونمذجة واقعية لهم. لكي يعمل النظام، فإنه يتطلب طريقة قوية لكشف وتتبع الأشخاص من ناحية محددة من بث الفيديو. في هذا البحث، تم تقديم تقنية لكشف الأشخاص مناسبة لمراقبة الفيديو، وهي تتطلب حسابات سريعة بالإضافة إلى نتائج دقيقة. التقنيات المقدمة تتضمن تمييزاً ذا إطار مهيئ لإزالة خلفية الصورة، وتعديل التباين لإزالة الظل، وكذلك طريقة تعتمد على الشكل لكشف الأشخاص. في المقابل، فإن تقنية التتبع تستخدم طريقة التطابق. كذلك تم تقديم نظام استرداد الفيديو المعتمد على الحدث (EBVR) من أجل إدارة فعالة لمعلومات المراقبة ونظام تمييز آلي للأشخاص مع تعيين فريد للشخصية. نظام تمميز الأشخاص وتتبعهم و نظام استرداد الفيديو المعتمد على الحدث (EBVR) وكذلك تمييز الحركة تم جمعهم في نظام مراقبة آلي كامل يسمى نظام التحليل الفعال لمراقبة الفيديو (AVSAS) الذي بدوره يقدم نتائج جيدة وأداءً فورياً خاصة في المشهد غير المزدحم بالأشخاص. نظام استرداد الفيديو المعتمد على الحدث (EBVR) أيضاً يبرهن أنه قادر على التعامل مع تمييز الأشخاص الآلي مع تعيين فريد للشخصية بدقة.

APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Science (Communication Engineering).

.....
Amir Akramin Shafie
Supervisor

.....
Othman O. Khalifa
Co-supervisor

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Science (Communication Engineering).

.....
Teddy Surya Gunawan
Examiner

This dissertation was submitted to the Department of Electrical and Computer Engineering and is accepted as a partial fulfilment of the requirements for the degree of Master of Science (Communication Engineering).

.....
Othman Khalifa
Head, Department of Electrical
and Computer Engineering

This dissertation was submitted to the Kulliyyah of Engineering and is accepted as a partial fulfilment of the requirements for the degree of Master of Science (Communication Engineering).

.....
Amir Akramin Shafie
Dean, Kulliyyah of Engineering

DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Fadhlan HafizHelmi Bin Kamaru Zaman

Signature …… Date …….....

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

AUTOMATED HUMAN RECOGNITION AND TRACKING FOR VIDEO SURVEILLANCE SYSTEM

Affirmed by Fadhlan Hafiz Helmi Bin Kamaru Zaman

…… ……

Signature Date

ACKNOWLEDGEMENTS

First of all, I would like to express my utmost heartfelt gratitude to almighty Allah swt. for granting me the ability to finish my master dissertation. My foremost appreciation goes to Dr. Amir Akramin Shafie, my supervisor, for his academic guidance and many other supports throughout my master program. This dissertation could not have been successfully completed without his guidance, insightful instructions, and constructive criticism throughout my research. He is a model of teacher and a supervisor besides a perfect academic advisor. Over the years, I have learned from him not only effective research approaches and good study habits, but also academic writing and an optimistic, positive attitude. I am also thankful to another supervisor of my master research who is also the Head of Electrical and Computer Engineering department, Prof. Dr. Othman O. Khalifa for his invaluable guidance to my thesis.

Special gratitude goes to respected Deputy Dean, Postgraduate and Research Prof. Momoh-Jimoh E. Salami, and Postgraduate coordinator, Assoc. Prof. Dr. Zahirul Alam for their comments and guidance from time to time which made me able to finish my thesis on time. Special thanks also go to Universiti Teknologi MARA (UiTM) especially to the Fakulti Kejuruteraan Elektrik (FKE) for awarding me the scholarship throughout my period of study. My master study will not be completed without their invaluable assistance. This research also has been funded by the MSC Malaysia and I am grateful towards MSC Malaysia for their great support to this project.

I am also heavily indebted to my parents, Kamaru Zaman Bin Asli and Robiah Binti Embong, and my beloved wife, Wan Noraini Binti Wan Razab, for their continuous support and concern. My graduate studies would not have been possible without the support of my parents, my wife and other members of my family, and to my colleague Md. Hazrat Ali, I am thankful for his great co-operation to complete this dissertation and helping me during my study period. I am also very grateful to my friends and colleagues who helped me to accomplish my research successfully.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
AFD	Adaptive Frame Differencing
ATM	Automated Teller Machine
AVSAS	ActiveVis Video Surveillance Analysis System
BMP	Bitmap
C#	C Sharp
CAM	Contrast Adjustment Method
CAVIAR	Context Aware Vision
CBS	Camera Blind Spot
CBVR	Content Based Video Retrieval
CCD	Charge-Coupled Device
CCTV	Close Circuit Television
DWT	Discrete Wavelet Transform
EBVR	Event Based Video Retrieval
EDA	Effective Detection Area
FPS	Frame per Second
GIS	Geographical Information System
GMM	Gaussian Mixture Model
GUI	Graphical User Interface
HCSD	Human Color Structure Descriptor
HES	Human Entering Scene
HLS	Human Leaving Scene
HOG	Histogram of Oriented Gradients
HRS	Human Roaming in Scene
HSV	Hue, Saturation, and Value
ID	Identification Number
JPEG	Joint Photographic Experts Group
LPP	Locality Preserving Projection
MAP	Maximum A Posteriori
MPIH	Most Probable Initial Human
NDA	Non-effective Detection Area
PCNN	Pulse Coupled Neural Network
PFINDER	Person Finder
PNG	Portable Network Graphics
RBF	Radial Basis Function
RGB	Red, Green and Blue
ROC	Receiver Operating Characteristics
SCS	Shadow Confidence Score
SVM	Support Vector Machine
TV	Television
W^4	Who? When? Where? What?
YCbCr	Luminance, Blue-difference, Red-difference
YUV	Luminance, Chrominance, Value

CHAPTER ONE

INTRODUCTION

1.1 INTRODUCTION

Video surveillance is an important tool to enhance public safety and privacy protection. According to Hosik et al (2009), surveillance cameras as in CCTV systems for instance are widely deployed in strategic places such as airports, banks, and public transportation facilities, as well as in public places such as stores, elevators, and hallways. Hosik et al (2009) added that the surveillance cameras in London provided key photos of the men who bombed the underground system in July 2005 while the latest terrorist attack in London was foiled in 2007, partly thanks to the millions of surveillance cameras that London authorities have installed across the city. In order to fight crimes, video surveillance is also used in commercial locations such as banks, automated teller machines (ATMs), supermarkets, and parking areas to prevent and track criminal activities whereas consumer adoptions of video surveillance also have soared in recent years due to the increasing concern on privacy protection (Limin et al, 2009).

According to Li-Qun (2007), the reduction in unit cost, seemingly easy to use and more versatile functionalities of the emerging video surveillance infrastructure have played a key role for the now widespread deployment of surveillance systems by various vertical industry sectors like government, transport, retail, banking and finance concerning safety, security and business intelligence analysis. This is further prompted by the currently acute political and socioeconomic drive in view of increasing level

and complexity of criminal acts and frequency and severity of other disastrous incidents / events.

The widespread use of video surveillance has then demanded new technologies and functionalities leading to smarter surveillance systems. These concern mainly research on visual scene understanding algorithms, enabling a variety of generic, scene adaptable, or domain specific functions such as multi-object detection, tracking and occlusion handling, event detection and visualization, visual scene characterization and segmentation, abnormal behaviours analysis, crowd scene analysis, multi-camera cooperative situation awareness, the structuring and search of large archived video databases, and so on (Li-Qun, 2007).

1.2 BACKGROUND OF THE STUDY

This research generally aims to integrate computer vision into video surveillance system in order to increase its accuracy, robustness and efficient data management. Computer vision broadly refers to the discipline where extraction of useful 2D and/or 3D information from one or more images is of interest (Rama Chellappa et al, 2005). Since the human visual system works by extracting information from the images formed on the retina of the eye, developments in computer vision are inevitably compared to the abilities of the human vision system. One of the basic tasks of the human visual system is to recognize humans and objects and spatial relationships among them. Similarly, one of the main goals of computer vision researchers is to develop methods for localization and recognition of objects in a scene.

Rama Chellappa et al (2005) also added that traditionally, problems in computer vision have been grouped into three areas that have vaguely defined boundaries. At low level, the goal is to extract features such as edges, corners, lines,

segmented regions, track features over a sequence of frames or to compute optical flow. At the intermediate level, using the output of the low-level modules, one is interested in grouping of features, in estimation of depth using stereopsis, and in motion and structure estimation. At the high level, the intermediate-level outputs are combined with available knowledge about the scene, objects and tasks so that descriptions of objects can be derived.

Recent research in computer vision has increasingly focused on building systems for observing humans and understanding their appearance, movements, and activities, providing advanced interfaces for interacting with humans, and creating realistic models of humans for various purposes (Ogale, 2006). For the last decades, video analysis and understanding has been one of the main active fields in computer vision and image analysis where applications relying on this field are various, like video surveillance, object tracking and traffic monitoring. (Allili et al, 2007). The ability of computer vision to recognize human from the image and works similar to human eye has made computer vision to be used widely in various applications especially in video analysis.

The capability of computer vision to also perceive image and shape in various physical conditions and constraints which might not be suitable for human eye can in fact enhance visual surveillance's accuracy and performance. Video surveillance systems are indeed becoming increasingly intelligent; thanks to the deployment of computer-vision an algorithm for detecting suspicious movement and identifying people and objects (Hosik et al, 2009). These intelligent applications help monitoring public area, counting interested objects passing through, reporting any suspicious behaviour, and so on. As more and more intelligence is sought by video surveillance

applications, there is an increasing demand for more robust real-time human recognition and tracking systems (Yuk, 2006)

Beleznai and Bischof (2009) suggested that reliable human detection is a key algorithmic component of many application-oriented computer vision systems, for instance in automated visual surveillance, automotive safety, human-computer interaction and multimedia processing. To realize more robust and secure video surveillance system, an automated human recognition and tracking system is needed which can analyze video streams fed by several surveillance cameras in real time, by utilizing fast-computation techniques without compromising the accuracy and performance of that particular surveillance system. The system also needs to be able to analyze past recorded events, store the video data, and manage the search for archived video data effectively. Dalal and Triggs (2005), Qiang (2006) and Qing Jun (2008) suggested however, detecting humans have been proven to be a challenging task because of the wide variability in appearance due to clothing, articulation and illumination conditions that are common in outdoor scenes

Another important aspect of video surveillance is the data management. Thi-Lan et al (2008) suggested that the increasing number of cameras provides a huge amount of video data and therefore, video data retrieval facilities become very useful for many purposes. He added while some approaches have been proposed for video retrieval in meetings, movies, broadcast news, and sports, very few work has been done for surveillance video retrieval. Current achievements on automatic video understanding such as object detection, object tracking and event recognition, though not perfect, are reliable enough to build efficient surveillance video indexing and retrieval systems.

Although a tremendous amount of work has been done in computer vision to enhance existing visual surveillance system, there are many issues still open and deserved further research. Among those issues as discussed earlier, are the precision and performance of moving human detection and tracking especially in real-time system currently available in the market, as well as the surveillance video and event data management which are being tackled in this research.

1.3 PROBLEMS STATEMENT

According to Li-Qun (2007), in the UK, the surveillance cameras are a familiar phenomenon (street sight), which appears in almost every town centre, public space, shopping mall, and event venue. There are, however, various practical problems associated with this rapid market expansion of CCTV systems.

According to Yoo and Park (2008), there has been a sudden increase in the number of cameras being deployed lately, and it gives rise to the problem of insufficient monitoring staff. Moreover, the monitoring of a wall of displays by human operators for a prolonged time is an impractical and improbable task, and a security guard might find himself unable to monitor large numbers of surveillance monitors without slipping any unwanted activities past him (Fang et al, 2008). McLeod (1996) also suggested that the reliability of human operators is a question of major significance. He added even the best and most conscientious operators will suffer from boredom and lose of concentration. A guard or human operator may be unaware of minute details, develop fatigue, and has sight limitations.

Another associated video surveillance problem, is the very existence and still on the increase of the overwhelming volume of raw video data that needs to be monitored lives on display, captured and transmitted through a network, stored in

medium and re-examined whenever a certain request arises. Yan et al (2009) claimed that large volume of information makes the manual observation and evidence extraction from video surveillance systems are extremely difficult. Hence, the manual search for a piece of evidence or an incident from terabytes of video data collected from multiple sources is painfully slow, laborious and expensive. This is true since an effective video surveillance system usually operates 24 hours a day and 7 days a week, which in turn would accumulate gigabytes if not terabytes of data in just one month.

There also regulations for large buildings or business surfaces that requires at any moment the number of people present in different areas or enclosures must be known. This way the authorities can properly act in case of fire or any other reason approaches that requires the complete evacuation of the building. People counter systems are intended to ensure that organization involved in the hospitality and leisure industries is capable to comply with modem fire regulations. (Gardel et al, 2007)

Having gone through all the associated problems and requirements of video surveillance system and the concept of computer vision, it is clear that the solution for various problems aforementioned is an automated, advanced and efficient computer scene analyzer integrated with video surveillance system. Thus, this will serve to greatly increase the effectiveness and performance of visual surveillance system, reduce the human operators' workload and the need for human capital, manage the surveillance data efficiently, and more importantly overcome human visual limitations thus producing tighter security.