AUTOMATED ANALYSIS OF SURVEILLANCE VIDEOS

AYESHA CHOUDHARY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY, DELHI
INDIA

JULY 2010

Automated Analysis of Surveillance Videos

by

Ayesha Choudhary

Department of Computer Science and Engineering

Submitted

in fulfillment of the requirements of the degree of

Doctor of Philosophy

to the

Indian Institute of Technology, Delhi

India

July 2010

Certificate

This is to certify that the thesis titled **“Automated Analysis of Surveillance Videos”** being submitted by Ayesha Choudhary to the Indian Institute of Technology, Delhi for the award of the degree of Doctor of Philosophy is a record of bona fide research work carried out by her under our supervision. In our opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The work presented in this thesis has not been submitted elsewhere, either in part or full, for the award of any degree or diploma.

Dr. Subhashis Banerjee
Professor
Department of
Computer Science & Engineering
Indian Institute of Technology Delhi
New Delhi-110016

Dr. Santanu Chaudhury
Professor
Department of
Electrical Engineering
Indian Institute of Technology Delhi
New Delhi-110016

Acknowledgments

It is with a deep sense of gratitude that I thank my supervisors Prof. Subhashis Banerjee and Prof. Santanu Chaudhury for their constant guidance, support and encouragement all through my Ph.D. My sincere thanks to them for giving me the opportunity to do my research work under their superior guidance and for their confidence in me, which has motivated me to learn and explore many things. They have always provided me with the right directions whenever I was lost, and were always forthcoming with words of encouragement whenever I needed them. Without their constant and unwavering support, it would have been impossible for me to reach this stage. I also thank them for the precious time that they generously gave me, in spite of their very many responsibilities in the institute.

I wish to express my heartiest gratitude towards Prof. J. B. Srivastava from the Department of Mathematics who introduced me to the fascinating subject of computer vision. Not only did it change the course of my career, his constant advice and guidance in many aspects of my life has helped me in numerous ways, for which I am always thankful to him. I also thank Prof. Suresh Chandra from the Department of Mathematics who instilled in me a deep interest for research and I am also grateful to him for his support and encouragement at all times.

I thank Prof. Prem Kalra for his encouragement, support and concern and I consider it my good luck to have been his teaching assistant for many semesters during my PhD. I would also like to extend my gratefulness to the other members of my SRC, Prof. Saroj Kaushik and Prof. Arun Kumar for their support and encouragement. I would also like to express my gratitude towards Prof. Sandeep Sen, the Head of the Computer Science Department for his administrative support and words of advice and especially for giving me an opportunity to visit Japan as a delegate in the JENESYS program. I also express my thanks to my supervisors for encouraging me to be a part of the same.

I am grateful to Prof. K. Gupta and Prof. S. Ishtiaque, holding the offices of Dean, Post Graduates and Deputy Director, Administration respectively, for their help and promptness in facilitating me with appropriate funds, which enabled me to present my paper in the Workshop on Camera Networks, and attend CVPR, 2010. I am especially grateful to both of them to enable me to receive an advance on my travel grant as a special case. I am deeply indebted to Prof. Shashi Mathur, then Associate Dean of

Students, who was extremely helpful in providing me with campus accommodation as a special case when I had suffered from a ligament rupture in the ankle at the end of the fifth year of my Ph.D.

I also thank all the faculty members from the Departments of Computer Science, Electrical Engineering and Mathematics, who have taught me and with whom I have interacted during my Ph.D years. It has been a enriching experience to learn from each one of them.

I am extremely thankful to Dr. S. Pathak, homoeopath at the hospital in IIT Delhi, for his magical potions and for always being there to help me cope with my health during the entire period of my research work, specially during the past two years. His calm personality, support and kind words along with his medicines helped in alleviating many problems.

I would like to express my heartfelt gratitude towards Mr. K. R. Kaushik, the technician in-charge of the Vision and Graphics lab in my department, for providing me with all the support to facilitate my research work, be it laboratory space, cameras or work-stations. He has always been ready to help, in every possible way within his capacity and has gone out of his way many a time to make things available at the earliest possible. I am also thankful to him for his concern and support and for his pieces of advice on life. I would like to thank Mrs. Vandana, Mr. Prasad, Mr. S. D. Sharma, Mr. Negi, and Mr. Jaswant, technicians in-charge of the other labs in my department, for their help whenever required. I would specially like to thank Mrs. Renu, the administrative staff in my department, for her prompt help whenever required. I also thank Mrs. Usha Bhola, technician in-charge of the Multimedia lab in Department of Electrical Engineering, for all her prompt help and support and especially for providing me with equipment there also. I thank Mrs. Sushila from the same lab for her help.

I am extremely grateful to my friends and colleagues Smruti and Tanveer, for going out of their way to help me with the arduous experiments, in spite of being completely busy with their own research work. I also express my heartfelt thanks for their encouragement, support and well-timed help which they always extended with eagerness. I am thankful to Gaurav Sharma and Manish Pal, with whom I have collaborated for my research work, for their help, support and team spirit. I am specially thankful to Gaurav for coming to my place to work when I was bedridden with the ligament rupture. It was because of him that I could continue my work during that period. I would also like to thank Nitin, Shubham, Mayank and Dipankar, students from the Department of Computer Science and Engineering, with whom I have worked during the course of my research. It was a pleasure to work with all of them.

I am extremely thankful to my friends Vaishali, Uma and Muralidhara for their help, support and words of encouragement, when we were colleagues in IIT, and also later, when they moved on with their lives. I would like to express my gratitude towards my friends and colleagues Subhajit, Parag, Taposh, Gaurav Harit, Aryabartta, Anant, Vikram, Jamshid, Ajay Aggarwal, Anup Gangwar and Rohit Khandekar, for their help and support. It has been a memorable journey and I am glad to have found some close friends along the way.

I would like to thank my current colleagues from the Visualization lab Rudra, Shibhashis, Chinmay, Nisha, Manoj, Symantak, Yamuna Prasad, Arindam and Swati for their help and for creating an atmosphere of camaraderie in the lab and making it a wonderful place to work. I thank my friends and colleagues from the Multimedia Lab, especially Anupama, Ritu, Indu, Hassan, Tanmoy and Subarna for their help and for making me feel a part of the Multimedia lab also.

Last but not the least, I would like to express my heartiest gratitude towards my husband Devabrat, my mother, my father, my sister Anita, my nephew Rian , brother-in-law Rahul, my parents-in-law, sisters-in-law and their families for their unfettered support, unwavering encouragement and undeterred faith in me. It would not have been possible to complete my research work without their love, encouragement and support. I especially thank Devabrat for all the many ways he helped me in making my Ph.D. a possibility. I thank him for taking on many responsibilities along with his own office and M.B.A. studies so that I could spend my time working in the lab. I thank my parents for the many ways they did their best to make my life easy and comfortable and most of all, for supporting me in all my decisions in life. I thank my sister and brother-in-law for taking on most of the family responsibilities and also thank my sisters-in-law and their husbands for doing the same on the other side. I especially thank my parents-in-law for being extremely understanding, loving, and encouraging me to pursue my career. I thank my nephews and nieces for making my life interesting and for their affection. I especially thank Rian, my sister's son for being there in my life and for helping me learn to maintain a balance in my life. I also thank my childhood friends Shachi, Aditi, Neha and Surabhi for their encouragement and faith in me.

Ayesha Choudhary

Abstract

Automated analysis of surveillance videos is an important area of research. This is mostly due to the growing need for surveillance of wide areas, such as airports, military bases and public places. As the areas under surveillance are growing wider and more complex, the difficulties of manual analysis of surveillance videos are increasing. At the same time, with the increase in size and complexity of the areas under surveillance, automated analysis of surveillance videos has become a challenging area of research. This has led to an increase in the various aspects of automated analysis of surveillance videos that need to be addressed.

In this thesis, we address some of the aspects of automated analysis of surveillance videos, such as event discovery, summarization and mining in single camera surveillance systems and real-time distributed composite event recognition in a calibrated pan-tilt camera network. We also address the problem of distributed self-calibration of a pan-tilt camera network. In the following paragraphs, we briefly outline our contributions.

We first propose a framework for multi-perspective summarization and mining of video data using unsupervised and incremental learning. We propose a tupular representation of the analytics data of a video that takes into consideration the fact that video data consists of information in multiple attributes or components and in multiple layers of abstraction. We introduce the concept of component based clustering to summarize the video data in the various attributes, chosen by the user. This allows the user to discover and learn the patterns in each of the attribute of the video data independently. We propose an incremental clustering algorithm that can be used with any type of data, symbolic or numeric, and is independent of pre-defining the number of clusters and the cluster radii. We apply our incremental clustering algorithm to cluster the data in each of the attribute spaces on-the-fly. This allows us to deal with the large amounts of video data easily as well as get a summary of the past as the video is analyzed. We also extend our incremental clustering algorithm for semi-supervised incremental clustering to deal with situations when labeled data is present *a priori*. We also extend

our incremental clustering algorithm for temporal incremental clustering to be able to find temporal patterns in the video data. This is essential for discovering events that are temporally related as well as for discovering events that are unusual on the basis of their time of occurrence. We develop a cluster algebra that defines operations on clusters from various attribute spaces and give a rough set [1] based analysis that shows that the algebraic operations on the component clusters leads to forming clusters in the higher dimensional spaces that are close approximation of the clusters formed by directly clustering in the higher dimensional spaces. We also define a *usualness* measure of a cluster that defines how usual a cluster is and use it to discover unusual patterns as well as learn usual ones. Our cluster algebra together with the *usualness* measure gives the user a flexible and powerful tool to explore the space of all clusters and their various combinations for multi-perspective summarization and mining of the video data patterns based on the requirement of the system.

Next, we propose an event discovery and anomaly detection framework for learning the usual and detecting the unusual events in the area under observation, given a sufficient amount of training data. We develop a generic segmentation procedure to segment the objects in the form of space-time patches. Our segmentation scheme uses video epitomes [2] to be able to segment a moving object in the form of space-time patches, even when the video has low resolution and lost frames. We use probabilistic Latent Semantic Analysis (pLSA) [3], to discover and learn a pre-defined number of event classes from the training data. We extend pLSA to detect unusual events in novel video clips of that area. We further extend pLSA to *time pLSA* to learn the temporal distribution of the discovered event classes. We show that this framework directly extends to a static multi-camera scenario for learning the correlations among events related in time.

If the area under surveillance is wide, a single camera becomes inadequate to observe the entire area and capture all the events that occur at the various locations of this area. Therefore, multi-camera networks are required to observe and analyze events that occur across space and time in a wide area. Pan-tilt camera networks are better suited for wide area surveillance, as compared to static camera networks as a fewer number of cameras are capable of observing wide areas for surveillance. However, a multi-camera network requires to be calibrated with respect to a global coordinate system for real-time surveillance and activity analysis. In this thesis, we propose a method for distributed self-calibration of a pan-tilt camera network using multi-layered belief propagation. We also propose a distributed self-calibration algorithm for a large static camera network using multi-layered belief propagation that requires automatic computation of corresponding points between at most three views at a time. We also develop an algorithm for automatic computation of corresponding points among multiple views of the same scene. We use this algorithm to discover the set of cameras that have a common view of the area. Since the cameras can pan and tilt, it is essential to discover the cameras that view a common region every time a camera pans and/or tilts. We show that by applying multi-layered belief propagation, we are able to obtain globally consistent and accurate camera parameters for each camera in the pan-tilt camera network in a distributed manner.

For a calibrated pan-tilt camera network, we propose a real-time distributed framework for composite event recognition. A composite event in a pan-tilt camera network comprises of events that are observed by smaller sub-networks of cameras, either simultaneously or sequentially in different locations across space and time. Distributed recognition of composite events requires distributed multi-camera, multi-object tracking and distributed recognition of events that comprise the composite event. We propose a multi-layered belief propagation based approach for reaching a consensus, within and across sub-networks, on the global identities of the objects that move around in the pan-tilt camera network. This allows us to perform multi-camera, multi-object tracking in a pan-tilt camera network in a distributed manner. For distributed recognition of an event by multiple cameras simultaneously viewing the event, we propose the use of belief propagation to reach a consensus on the event recognized by each camera independently. We propose a hidden Markov model (HMM) [4] based approach for composite event recognition as they are best suited for recognition of a temporal sequence of observations when the state of the system is dynamic and unknown. We also present an application of our framework, where an event implies interaction between a pair of objects. In this case, the composite event is composed of various different interactions between pairs of objects as they move in the complete area under observation of the pan-tilt camera network. We also propose a novel pLSA [3] based algorithm for discovery and learning of pair-wise interactions as well as to detect unusual pair-wise interactions. Our online, real-time composite event recognition framework is capable of recognizing usual and detecting unusual events and composite events in a distributed manner in a pan-tilt camera network.

Contents