Scientific Computing Department
Faculty of Computers & Information Sciences
Ain Shams University

Real -Time Tracking for Intelligent Surveillance Systems

Thesis submitted to the Department of Scientific Computing, Faculty of Computers and Information Sciences, Ain Shams University, in the partial fulfillment of the requirements for the Ph. D. degree in Computers and Information Sciences (Scientific Computing)

By:

Maryam Nabil Zakaria Al-Berry
M. Sc. In Scientific Computing (2006)
Lecturer Assistant at the Scientific Computing Department
Faculty of Computers & Information Sciences
Ain Shams University

Under Supervision of:

Prof. Dr. Mohammed Fahmy Tolba
Professor at the Scientific Computing Department
Faculty of Computers & Information Sciences
Ain Shams University

Prof. Dr. Ashraf Saad Hussein
Dean of the Faculty of Computing Studies
Arab Open University
Kuwait

Dr. Mohammed Abdel-Megeed Salem
Lecturer at the Scientific Computing Department
Faculty of Computers & Information Sciences
Ain Shams University

Cairo 2015

Acknowledgement

First, thanks to Allah, the most beneficent and the most merciful.

I sincerely acknowledge and express my appreciation to Prof. Dr. Mohamed FahmyTolba for his support, valuable comments and periodic careful evaluation of my work.

I am deeply indebted to Prof. Dr. Ashraf Saad Hussein, for his great effort in revising my work and putting it into a suitable form.

Special thanks to Prof. Dr. Howaida Shedid, head of the Scientific Computing department, for her kind support in the most difficult times.

I would like to thank Ass. Prof. Dr. Hala Mousher Ebeid for the very valuable support and review of my work.

Special thanks to Dr. Mohammed Abdel-Megeed Salem who inspired me with the main contribution and continued to support every detail in this work.

I cannot express my feelings towards my family who supported me unconditionally and backed me up during my studies, wishing me good luck throughout the preparation of this thesis.

Table of Contents

List of Publications

1. M. N. Al-Berry, M. A.-M. Salem, A. S. Hussein and M. F. Tolba, "Motion Detection using Wavelet-enhanced Accumulative Frame Differencing," in*Proc. 8th Intl. Conf. Computer Engineering and Systems*, 2013, pp. 255–261. doi: 10.1109/ICCES.2013.6707215
2. M. N. Al-Berry, M. A.-M. Salem, A. S. Hussein, M. F. Tolba, “Spatio-temporal Motion Detection for Intelligent Surveillance Applications,” *International Journal of Computational Methods,* vol. 12, no. 1, 2015. *doi:* 10.1142/S0219876213500977.
3. M. N. Al-Berry, M. A.-M. Salem, Ebeid, H. M., A. S. Hussein, M. F. Tolba, “Wavelet-Enhanced Detection of Small Slow Objects Moving in Complex Scenes”, submitted to the Frontiers of Computer Science Journal.
4. M. N. Al-Berry, M. A.-M. Salem, H. M. Ebeid, A. S. Hussein, M. F. Tolba, “Action Recognition using Stationary Wavelet-based Motion Images,”in *Proc. IEEE Conf. Intelligent Systems 2014 (IS’14)*, Warsaw, Poland, 2014, pp. 743–753. doi: 10.1007/978-3-319-11310-4_65
5. M. N. Al-Berry, M. A.-M. Salem, H. M. Ebeid, A. S. Hussein, M. F. Tolba, “Directional Stationary Wavelet-based Representation for Human Action Classification,” in *Proc. Intl. Conf. Advanced Machine Learning Technologies and Applications (AMLTA2014),* Cairo, Egypt, 2014, pp. 309–320. doi: 10.1007/978-3-319-13461-1_30.
6. M. N. Al-Berry, H. M. Ebeid, A. S. Hussein, M. F.Tolba, “Human Action Recognition via Multi-scale 3D Stationary Wavelet Analysis,”*in Proc. 14th Intl. Conf. Hybrid Intelligent Systems (HIS2014)*, Kuwait, Kuwait, 2014.
7. M. N. Al-Berry, M. A.-M. Salem, H. M. Ebeid, A. S. Hussein, M. F. Tolba,“Directional Multi-scale Stationary Wavelet-based Representation for Human Action Classification,”in *Handbook of Research on Machine Learning Innovations and Trends*, IGI Global, in press.
8. M. N. Al-Berry, M. A.-M. Salem, H. M. Ebeid, A. S. Hussein, M. F. Tolba,“Fusing Directional Wavelet Local Binary Pattern and Moments for Human Action Recognition,”Provisionally accepted in*IET Computer Vision Journal*.
9. M. N. Al-Berry, M. A.-M. Salem, H. M. Ebeid, A. S. Hussein,

M. F. Tolba, "Action Classification using Weighted Directional Wavelet LBP Histograms,"accepted for publication in *1st Intl. Conf. Advanced Intelligent Systems and Informatics*, BeniSuef, Egypt, 2015.

10. M. N. Al-Berry, M. A.-M. Salem, H. M. Ebeid, A. S. Hussein, M. F. Tolba, "Weighted Directional 3D Stationary Wavelet-based Action Classification,"*Egyptian Computer Science Journal*, vol. 39, no. 2, 2015.
11. M. N. Al-Berry, M. A.-M. Salem, H. M. Ebeid, A. S. Hussein, M. F. Tolba, "Recent Challenges and Advances in Spatio-Temporal Action Recognition," accepted in *Applications of Intelligent Optimization in Biology and Medicine,* Springer-Verlag.

Abstract

Intelligent surveillance is very important for security-sensitive fields. Generally, surveillance can be defined as the observation of changing information, activities or behaviors for some purpose. The framework of visual surveillance systems includes environment modeling, motion detection, object classification, tracking as well as behavior understanding and description. Environment modeling is the module responsible for creating and updating dynamic models for the environment. Motion detection is the module responsible for segmenting moving objects from static or irrelevant background. This module is the base for any subsequent processing; thus it must be accurate, robust and fast. A surveillance scenario may contain different types of moving objects, therefore the system must classify the detected objects. The tracking module tracks the classified moving objects from one frame to another and then the behavior of the tracked objects is analyzed and a description of actions/activities is provided.

The first contribution of this thesis is mainly concerned with the problem of motion detection. Two innovative spatio-temporal wavelet-based motion detection techniques are proposed, combining the advantages of wavelets, multi-resolution analysis and data fusion to enhance the performance without raising the complexity. The first proposed technique is based on 3D Stationary Wavelet Transform (SWT), which combines spatial and temporal analysis into a single 3D transform by applying 1D analysis in the x-, y- and t- domains. The second proposed technique is

implementing the 3D transform as two separate spatio-temporal analyses. Both of the proposed techniques are compared to the recent techniques using a benchmark dataset. In addition, the proposed 3D technique is compared to another 3D wavelet-based technique using a traffic monitoring dataset. Both of the proposed techniques outperform traditional techniques, especially in the cases of low contrast scenes and those having non-uniform illumination and they succeeded to detect moving objects in bad and time-varying illumination conditions.

The second contribution is in the field of human action classification. A new method is proposed by using the 3D Stationary Wavelet Transform (SWT) and combining it with a Local Binary Pattern (LBP) histogram to represent and describe the human actions in video sequences. The directional and multi-scale information encoded in the wavelet coefficients is utilized to obtain robust global and local descriptions in a unified feature vector. This unified vector is used to train standard classifiers. The performance of the new method was examined in two different ways. One way is by fusing the global and local features, generated from directional sub-bands, in one feature vector and using the fused feature vector for training the classifiers. The second way uses the features of different directional bands separately to train multiple classifiers with a voting scheme to vote for the best match. The performance of the proposed descriptors is verified using two standard datasets. The proposed method achieved high accuracy in comparison with the existing methods.

List of Figures

List of Tables

1

Introduction

1.1 Problem Definition
1.2 Objectives
1.3 Thesis Organization

1 Introduction

1.1 Problem Definition

Intelligent surveillance [1]is crucial for security sensitive fields such as airports, parking lots andbanks. Generally, surveillance may be defined as the observation of changing information, activities or behaviors for some purpose. Perception of changing information and activities in surveillance systems usually relies on cameras which capture various details about objects in the scene. A general framework of visual surveillance systems includes environment modeling, motion segmentation, object classification, tracking as well as behavior understanding and description [2]. Environment modeling is the module responsible for creating and updating dynamic models for the environment. Motion detection is the module of segmenting moving objects from static or irrelevant backgrounds. This module is the base for any subsequent processing; thus it must be accurate, robust and fast.

A surveillance scenario may contain different types of moving objects such as; humans[3] and vehicles [4] and therefore the system must classify the detected objects. The tracking module tracks the classified moving objects from one frame to another and then the behavior of the tracked objects is analyzed and a description of actions/activities is provided. If the system is equipped with multiple cameras[5], a data fusion module is needed to fuse data captured by the different cameras. The data fusion module may also fuse spatial, temporal and color information to solve some problems like occlusions.For the aforesaid framework, it is interesting to point out that the motion detection module concerns with classification, tracking and behavior understanding and considering the efficiency of these processes.

A huge amount of surveillance applicationsneed human action and activity recognition [6, 7, 8] as a basic module. Intelligent Surveillance applications [9, 10, 11] include systems that are used to detect abnormal behavior [12, 13] in security sensitive areas [7], crowd behavior surveillance [14, 15], group activity recognition [16] and human identification using behavioral biometrics [17, 18]. This makes the area of human action and activity recognition still an open research area despite being relatively old. Another reason is the existence of various challenges that face the task of action and activity recognition. Challenges include variations in the environment, dynamic backgrounds, variations in the performer posture and clothes, variation of the performance of the actions or variations in the rate of execution of the action [19, 20, 21, 22]. More complex activity and behavior understanding, encounter some other challenges including, the number of modalities to be used, how to fuse these modalities, how much of the context affects the process of learning and recognition among others [23]. These challenges cannot be completely avoided and thus the need for robust action representation and description increases.

1.2 Objectives

The main objective of this thesis is to propose a framework for joint detection and recognition of human actions in a surveillance