



---

# 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

<b>TABLE OF CONTENTS .....</b>	<b>2</b>
<b>LIST OF PUBLICATIONS.....</b>	<b>4</b>
<b>LIST OF FIGURES.....</b>	<b>8</b>
<b>LIST OF TABLES.....</b>	<b>11</b>
<b>1 INTRODUCTION .....</b>	<b>13</b>
1.1 PROBLEM DEFINITION .....	13
1.2 OBJECTIVES .....	14
1.3 THESIS ORGANIZATION .....	16
<b>2 INTELLIGENT SURVEILLANCE .....</b>	<b>21</b>
2.1 VISUAL SURVEILLANCE SYSTEM COMPONENTS AND APPLICATION DOMAINS .....	21
2.2 MOVING OBJECT DETECTION TECHNIQUES .....	24
2.2.1 <i>Object Detection Techniques</i> .....	24
2.2.2 <i>Motion-based Object Detection</i> .....	29
2.3 CLASSIFICATION AND IDENTIFICATION OF OBJECTS .....	34
2.4 MOVING OBJECT TRACKING .....	35
2.5 PRACTICAL ISSUES IN REAL WORLD SCENARIOS .....	38
<b>3 HUMAN ACTION AND ACTIVITY RECOGNITION .....</b>	<b>42</b>
3.1 DEFINITIONS AND LEVELS OF ABSTRACTION .....	43
3.2 TAXONOMIES OF ACTION AND ACTIVITY RECOGNITION TECHNIQUES .....	44
3.3 SPATIO-TEMPORAL FOR ACTION RECOGNITION TECHNIQUES .....	48
3.3.1 <i>Global Spatio-temporal Representation and Description</i> .....	49
3.3.2 <i>Local Spatio-temporal Representation and Description</i> .....	56
3.4 SEMANTIC INFORMATION EXTRACTION FOR BEHAVIOR ANALYSIS .....	63
3.5 SUMMARY .....	65
<b>4 WAVELET-ENHANCED DETECTION OF SMALL/SLOW OBJECTS MOVING IN COMPLEX SCENES .....</b>	<b>68</b>
4.1 TRADITIONAL MEMORY-BASED MOVING OBJECT DETECTION .....	68
4.2 PROPOSED WAVELET-ENHANCED MEMORY-BASED MOVING OBJECT DETECTION ..	69
4.3 EXPERIMENTAL RESULTS .....	76
4.3.1 <i>CAVIAR test case scenarios</i> .....	77
4.3.2 <i>Performance Evaluation criteria</i> .....	78
4.3.3 <i>Results and discussion</i> .....	80
4.4 SUMMARY .....	83
<b>5 SPATIO-TEMPORAL MOTION DETECTION FOR INTELLIGENT SURVEILLANCE .....</b>	<b>86</b>

---

5.1 THE PROPOSED 3D STATIONARY WAVELET-BASED MOTION DETECTION TECHNIQUE	87
5.2 PROPOSED SEPARATE SPATIO-TEMPORAL TECHNIQUE .....	91
5.3 COMPLEXITY ANALYSIS .....	93
5.4 EXPERIMENTAL RESULTS AND DISCUSSION .....	96
5.5 SUMMARY .....	104
<b>6 DIRECTIONAL MULTI-SCALE STATIONARY WAVELET-BASED REPRESENTATION FOR HUMAN ACTION CLASSIFICATION .....</b>	<b>106</b>
6.1 WAVELET-BASED MOTION IMAGES: .....	106
6.1.1 <i>Wavelet-based Energy Images</i> .....	106
6.1.2 <i>Wavelet-based History Images</i> .....	107
6.2 DIRECTIONAL WAVELET-BASED MOTION IMAGES: .....	108
6.3 FEATURE EXTRACTION .....	111
6.4 RESULTS AND DISCUSSION .....	112
6.4.1 <i>Description of Datasets</i> .....	112
6.4.2 <i>Experimental Setup</i> .....	114
6.4.3 <i>Experiments and Results</i> .....	115
6.5 SUMMARY .....	124
<b>7 DIRECTIONAL WAVELET LOCAL BINARY PATTERNS .....</b>	<b>127</b>
7.1 LOCAL BINARY PATTERNS .....	127
7.2 PROPOSED DIRECTIONAL WAVELET LOCAL BINARY PATTERN .....	129
7.2.1 <i>Proposed Directional Wavelet Local Binary Pattern Histogram</i>	130
7.2.2 <i>Proposed Weighted Directional Action Representation and Description</i>	133
7.2.3 <i>Global Description using Invariant Moments</i> .....	134
7.3 EXPERIMENTAL RESULTS .....	135
7.3.1 <i>Directional Wavelet Local Binary Pattern Histogram</i> .....	135
7.3.2 <i>Weighted Directional Local Binary Pattern Histograms</i> .....	141
7.4 SUMMARY .....	144
<b>8 CONCLUSIONS AND FUTURE WORK .....</b>	<b>147</b>
<b>REFERENCES .....</b>	<b>151</b>

---

## 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. 8<sup>th</sup> 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. 14<sup>th</sup> 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 *1<sup>st</sup> 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

Figure 1.1 The proposed framework for action detection and recognition	16
Figure 2.1 General components of an intelligent surveillance system	23
Figure 2.2 Object representations (a) Centroid, (b) multiple points, (c) rectangular patch, (d) elliptical patch, (e) part-based multiple patches, (f) object skeleton, (g) complete object contour, (h) control points on object contour, (i) object silhouette. [42]	26
Figure 2.3 Interest points detected by applying (a) the Harris, (b) the KLT and (c) SIFT operators.	28
Figure 3.1 Action recognition steps	42
Figure 3.2 Different taxonomies of action levels.	44
Figure 3.3 Summary of different taxonomies of action and activity recognition techniques	48
Figure 3.4 First row: Different exercise types. Second row: Their corresponding MHI's. Derived from the work of Shao et al. [86]	50
Figure 3.5 Walk action from two different view angles. The first row shows sample frames. The second row shows extracted contours. The third row shows corresponding STV. Derived from the work of Yilmaz and Shah [104]	53
Figure 3.6 Space-time shapes of “jack”, “walk” and “run” actions. [109]	54
Figure 3.7 Motion History Volumes (MHV) [89]	55
Figure 3.8 Local space-time interest points extracted by [119]	57
Figure 3.9 Space-time volume and extracted space time interest points (cuboids). From the work of Dollar et al. [123]	58
Figure 3.10 Left hand side: Interest point detected for the leg pattern. Right hand side: Space-time features projected on single frames. Derived from [116]	59
Figure 3.11 From left to right: 2D SIFT descriptor, use of 2D SIFT with a sequence of images, 3D SIFT with its 3D sub-volumes. Derived from [129]	60
Figure 3.12 Representation of Action in (a) XYZ plane, (b) XYT plane. Derived from the work of Sheikh et al. [138].	62
Figure 3.13 Summary of spatio-temporal techniques	63
Figure 4.1 The DWT used in the pre-processing step	72
Figure 4.2 Frame differencing applied on the wavelet coefficient	73
Figure 4.3 Wavelet analysis used in the data validation step	74

---

Figure 4.4 Wavelet Analysis applied on accumulative absolute differences	76
Figure 4.5 Sample frames from some of the selected sequences.	78
Figure 4.6 Comparison of the average performance of all studied methods	82
Figure 4.7 (a) Sample frame from Seq3. (b) The output of the background subtraction method. (c) The output of the proposed method V.1	82
Figure 5.1 (left hand side) Block of the time-varying data formed using two successive frames. (right hand side) A 3D Stationary Wavelet Transform implemented as three successive 1D transforms. $\Phi(\cdot)$ is the scaling function and $\psi(\cdot)$ is the associated wavelet function.	88
Figure 5.2 Block diagram of the proposed 3D stationary wavelet-based technique	91
Figure 5.3 Block diagram of the proposed separate spatio-temporal technique	93
Figure 5.4 Average performance using automatic thresholding	98
Figure 5.5 Left hand column: CD rates using manual thresholding right hand column: ROC curves for different threshold values	100
Figure 5.6 Sample frames from the sequence frankfurt11	102
Figure 5.7 Illustration of different performances of the proposed 3D technique	103
Figure 6.1 Illustration of 3D stationary wavelet-based motion detection	107
Figure 6.2 Forming directional wavelet-based templates using 3D SWT	110
Figure 6.3 Four sample frames of the “jack” action from Weizmann dataset [160]	113
Figure 6.4. Sample of KTH dataset actions and scenarios.	114
Figure 6.5 Confusion matrices obtained in different experiments. The first row shows the results from Wavelet Energy Images. The second row shows the results from Wavelet History Image.	116
Figure 6.6 Confusion matrices for using the 3D stationary wavelet transform. (a) Mahalanobis distance metric, (b) Quadratic discriminant analysis.	117
Figure 6.7 Directional wavelet energy (first row) and directional wavelet history (second row) of the “jack” action	119
Figure 6.8 Confusion matrices obtained using different directional wavelet energy sub-band using Weizmann dataset (a) ADD band, (b) DAD band and (c) DDD band.	120

---

Figure 6.9. Confusion matrices obtained using different directional history sub-bands using Weizmann dataset (a) ADD band, (b) DAD band and (c) DDD band	121
Figure 6.10. Confusion matrices obtained using different directional history sub-bands using KTH dataset. Left hand column: ADD band, center column: DAD band and right hand column DDD band	122
Figure 6.11. Results obtained using the combined feature vector.	123
Figure 7.1 Example of extended LBP ( $R = 2$ , $P = 8$ ). Figure adapted from [28]	128
Figure 7.2 Examples of uniform patterns that describe texture primitives. Figure adapted from [28]	129
Figure 7.3 Illustration of the proposed approach of feature extraction	130
Fig. 7.4. Illustration of the proposed weighted directional LBP method	133
Figure 7.5 Analysis of the effect of changing the neighborhood size of the LBP	137
Figure 7.6 Confusion matrix obtained using the voting scheme	138
Figure 7.7 Confusion matrices obtained for the KTH dataset using the voting scheme.	140

---

### List of Tables

Table 4.1 Performance of the proposed method on CAVIAR dataset .	81
Table 4.2 The Average Match Metric for all studied methods.....	83
Table 5.1 Relative performance of the proposed techniques.....	101
Table 6.1 Accuracy comparison between the proposed method and some reference methods .....	118
Table 6.2. Performance comparison between the proposed method and state-of the art methods.....	124
Table 7.1 Classification results for the Weizmann dataset.....	138
Table 7.2 Classification accuracies for separate sub-bands using KTH dataset .....	138
Table 7.3 classification accuracy using feature fusion vs. voting schemes for the KTH dataset.....	140
Table 7.4: Performance Comparison with Existing Techniques .....	141
Table 7.5 Performance obtained on the Weizmann dataset .....	142
Table 7.6 Accuracy using combined feature vector vs voting between directional feature vectors on the KTH dataset .....	142
Table 7.7 Accuracy using combined feature vector vs voting for gait actions in the KTH dataset.....	143
Table 7.8 Performance Comparison with Existing Techniques .....	144

---

# 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 and banks. 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 applications need 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