NEW DESCRIPTOR FOR OBJECT DETECTION USING AN IMPROVED ENSEMBLE-BASED TECHNIQUE

AMIN MOHAMED AHSAN

UNIVERSITI TEKNOLOGI MALAYSIA

NEW DESCRIPTOR FOR OBJECT DETECTION USING AN IMPROVED ENSEMBLE-BASED TECHNIQUE

AMIN MOHAMED AHSAN

A thesis submitted in fulfilment of the requirements for the award of the degree of Doctor of Philosophy (Computer Science)

> Faculty of Computing Universiti Teknologi Malaysia

> > MAY 2015

To my beloved parents, my wife, children, brothers and sisters

ACKNOWLEDGMENTS

All praise and thanks are due to Allah who "Taught man that which he knew not," and peace and blessings of Allah be upon our prophet, Mohammad and upon all his family and companions. All thanks to Allah for graces and for giving me the strength and endurance to complete this research

During PhD journey, I was in contact with many people, scholars and academicians colleagues. They have contributed in different ways towards my understanding and thoughts. First, I would like to express my high appreciation to my supervisor, Prof. Dr. Dzulkifli Bin Mohamad, for encouragement, guidance, comments and his support which participated actively in the completion of this work.

My sincere appreciation also extends to Dr. Faisal Alsamet for his unlimited help, Dr. Murad Qasem and Dr. Redhwan Shaddad for their support.

ABSTRACT

Object detection is an essential process for further tasks including, but not limited to, object and event detection, object tracking, object recognition, video indexing, motion estimation, image restoration, image registration, image retrieval, and reconstruction of 3D scene. In the recent past, interest point detectors and their descriptors, as local features, have received a great interest in computer vision areas and technologies. These types of features have shown their robustness against different types of deformation due to geometric transformation, photometric transformation and other disturbances. Therefore, they are more accurate and stable than the global ones. Among all interest point detectors and descriptors, the Scale Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) are considered as the most common methods that receive interest from researchers in terms of usage and development; but, getting more accurate, invariant and fast descriptor is still needed. Matching technique is often used to recognize the object based on such features; however, it is not proper for some applications such as searching for an isolated object and it is difficult to be used in object category recognition or to recognize the part-based object. Therefore, learning-based technique, that has been proven to be an effective method in object detection, can be used to overcome the previously mentioned challenges. However, the object required to be detected usually represents a small ratio compared to non-object that causes an imbalanced data problem. The aim of this study is to design and develop an effective model for object detection that is faster, more accurate and it can manage aforementioned challenges. To achieve this goal first, new fast and an accurate descriptor is introduced based on interest points; second, an effective classification method, that mitigates the effect of imbalanced data, is designed based on developed ensemble classifiers; third, an updating scheme of interest point detector is presented to speed up the object detection system. Results show that the proposed features are faster and more invariant than the most common interest-point-based features. The developed technique based on ensemble classifiers produces notable results in terms of accuracy and False Positive rate compared to the traditional one. The speed of object detection system has increased by 30% in average based on the proposed scheme.

ABSTRAK

Pengesanan objek adalah suatu proses yang penting dalam tugasan lanjut termasuk pengesanan objek dan acara, penjejakan objek, pengecaman objek, mengindek video, anggaran pergerakan, restorasi imej, pendaftaran imej, penemuan kembali imej and rekonstruksi adegan 3D. Sebelum ini, pengesan titik minat dan pemarihalnya, iaitu ciri-ciri sedia ada telah menarik minat yang mendalam di dalam bidang visi komputer dan teknologi. Ciri-ciri ini telah mempamerkan kecekalan terhadap beberapa jenis kecacatan yang disebabkan oleh transfomasi geometri, transfomasi fotometri dan gangguan yang lain. Oleh itu, ciri-ciri ini lebih tepat dan stabil daripada ciri-ciri global yang lain. Antara semua pengesan titik minat dan pemarihalnya, Ciri Transform Skala Invarian (SIFT) dan Ciri-Ciri Kelajuan Teguh (SURF) dianggap sebagai dua kaedah sepunya yang paling mendapat perhatian pengkaji dari aspek penggunaan dan pembangunan tetapi penemuan ciri-ciri kategori objek yang lebih cekal dan pantas masih diperlukan. Teknik pemadanan sering digunakan untuk mengenal pasti objek berdasarkan ciri-ciri tersebut. Namun, ia tidak sesuai untuk beberapa aplikasi seperti gelintar objek yang terpencil dan sukar untuk digunakan semasa pengecaman kategori objek atau pengecaman objek yang berdasarkan bahagian-bahagian. Justeru, teknik berdasarkan pembelajaran yang telah dibuktikan sebagai suatu kaedah yang berkesan untuk mengecam objek boleh digunakan untuk mengatasi cabaran-cabaran yang disebut sebelum ini. Walau bagaimanapun, objek yang perlu dikesan lazimnya melambangkan suatu nisbah yang kecil berbanding dengan benda-benda bukan berbentuk objek yang menimbulkan masalah ketidakseimbangan data. Tujuan kajian ini adalah untuk mereka bentuk dan membangunkan satu model yang berkesan untuk mengesan objek dengan lebih pantas, lebih tepat dan dapat mengurus cabaran-cabaran yang disebut tadi. Untuk mencapai matlamat ini, pertama, pengesan yang pantas dan tepat diperkenalkan berdasarkan titik-titik berkepentingan; kedua, satu kaedah klasifikasi yang berkesan dan mampu mengurangkan kesan ketidakseimbangan data yang direka bentuk berdasarkan pengklasifikasi ensembel yang dibangukan; ketiga, dikemukakan suatu skema yang dapat mengemaskini pengesan titik minat untuk mempercepatkan sistem pengecaman objek. Dapatan kajian menunjukkan bahawa ciri-ciri yang dicadangkan adalah lebih pantas dan lebih tepat berbanding ciri-ciri yang berdasarkan titik kepentingan yang sepunya. Teknik yang dibangunkan berdasarkan pengklasifikasi ensembel telah menghasilkan keputusan yang penting berkaitan kejituan dan kadar positif palsu berbanding dengan teknik tradisional. Namun, kelajuan sistem pengesanan objek telah meningkat 30% secara purata berdasarkan skema yang dicadangkan.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGMENTS	iv
	ABSTRACT	V
	ABSTRAK	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	xii
	LIST OF FIGURES	xvi
	LIST OF ABBREVIATIONS	xxiii
	LIST OF APPENDICES	XXV
1	INTRODUCTION	1
	1.1 Introduction	1
	1.2 Problem Background	2
	1.3 Problem Statement	7
	1.4 Research Questions	7
	1.5 Research Aim	8
	1.6 Research Objectives	8
	1.7 Research Scope and Assumptions	9
	1.8 Research Significance	10
	1.9 Summary of Research Contributions	11
	1.10 Terms Definitions	11
	1.11 Thesis Outline	13

vii

LIT	ERAT	URE RE	VIEW	14
2.1	Introd	uction		14
2.2	Objec	t Detectio	n	15
	2.2.1	Features	5	18
		2.2.1.1	Interest Points Detectors	21
		2.2.1.2	Interest Points' Descriptors	25
			2.2.1.2.1 Floating-point-based	
			Descriptors	26
			2.2.1.2.2 Binary-based Descriptors	28
		2.2.1.3	Interest Point Detectors used and their	
			Descriptors	29
			2.2.1.3.1 SIFT Points' Detector and	
			Descriptor	29
			2.2.1.3.2 SURF Points' Detector and	
			Descriptor	33
	2.2.2	Object I	Recognition	37
		2.2.2.1	Supervised Learning Algorithms	39
		2.2.2.2	Technique Used	42
			2.2.2.2.1 Back-propagation Neural	
			Network	42
			2.2.2.2.2 Related Issues	45
	2.2.3	Object I	Localization	47
2.3	Ensen	nble-base	d Recognition	50
2.4	Data s	sets		54
2.5	Perfor	mance Ev	valuation	58
2.6	Relate	ed Works		62
	2.6.1	Object l	Detection Frameworks	62
	2.6.2	Features	S	64
	2.6.3	Recogn	ition	68
2.7	Discu	ssion		72
RES	SEARC	CH METI	HODOLOGY	73
3.1	Introd	uction		73

viii

2

3

3.2	Study	Framewo	ork	75
3.3	Object	Object Detection Datasets		78
	3.3.1	Faces D	ataset	78
		3.3.1.1	Training Data	78
		3.3.1.2	Testing Data	81
			3.3.1.2.1 Labeled Data	81
			3.3.1.2.2 Unlabeled Data	83
	3.3.2	Cars Da	taset	84
		3.3.2.1	Training Data	84
		3.3.2.2	Testing Data	86
			3.3.2.2.1 Labeled Data	86
			3.3.2.2.2 Unlabeled Data	88
3.4	Propo	sed Metho	ods	89
3.5	Exper	imental D	esign	91
3.6	Perfor	mance Ev	valuation	92
3.7	Summ	nary		94

4	1
	-

NESTED-WINDOWS OF HISTOGRAM OF

ORIENTED GRADIENTS DESCRIPTOR	95
4.1 Introduction	95
4.2 Methods	96
4.2.1 Motivation	96
4.2.2 SIFT and SURF Descriptors	97
4.2.2.1 SIFT Descriptor	97
4.2.2.2 SURF Descriptor	98
4.2.3 The Proposed Descriptor	99
4.2.4 Single-based Learning Technique	106
4.2.4.1 Learning Phase	106
4.2.4.2 Testing Phase	109
4.3 Experimental Design	110
4.4 Experimental Results	111
4.4.1 Classification Performance	113
4.4.1.1 Faces Classification	114

		4.4.1.2	Cars Classification	116
	4.4.2	Localiza	ation Performance	118
		4.4.2.1	Faces Localization	119
		4.4.2.2	Cars Localization	120
	4.4.3	Invaria	nce	121
		4.4.3.1	Invariance to Rotation	121
		4.4.3.2	Invariance to Scale	124
4.5	Time	Consump	tion	128
4.6	Discu	ssion		130
4.6	Summ	nary		135
AN	IMPRO	OVED EN	NSEMBLE-BASED TECHNIQUE	
FOI	R OBJI	ECT CLA	ASSIFICATION	136
5.1	Introd	uction		136
5.2	Metho	ods		137
	5.2.1	Motivat	ion	137
	5.2.2	An Ense	emble-based Developed Technique	138
		5.2.2.1	Learning Phase	140
		5.2.2.2	Testing Phase	142
5.3	Exper	imental D	Design	144
5.4	Exper	imental R	lesults	144
	5.4.1	Classifi	cation Performance	145
		5.4.1.1	Faces Classification Performance	145
		5.4.1.2	Cars Classification Performance	150
	5.4.2	Localiza	ation Performance	155
		5.4.2.1	Faces Localization Performance	155
		5.4.2.2	Cars Localization Performance	156
	5.4.3	Time Co	onsumption	157
5.5	Discu	ssion		163
5.6	Summ	nary		170

5

6	AN IMPROVED SCHEME OF AN INTEREST POINT				
	DE	ГЕСТО)R	171	
	6.1	Introc	luction	171	
	6.2	Motiv	vation	172	
		6.2.1	Methods	172	
		6.2.2	A Developed Scheme for Speeding up The		
			Object Detection System	172	
	6.3	Expe	imental Results	179	
	6.4	Discu	ssion	185	
	6.5	Sumn	nary	188	
7	CO	NCLU	SIONS	189	
	7.1	Introc	luction	189	
	7.2	Sumn	nary of Research Achievements	190	
	7.3	Resea	rch Contributions	191	
	7.4	Futur	e Works	194	
	7.5	Resea	rch Challenges	195	
REFERENC	ES			196	
Appendices A	-C			210-249	

LIST OF TABLES

TABLE NO.	TITLE	PAGE
2.1	Comparison between the global and local features	21
2.2	Comparison among some existing points' detector regarding some important criterion	25
2.3	Comparison among the most common algorithms (Kotsiantis, et al., 2007)	40
2.4	Summary of some existing studies in regard to localization performance and time consuming	48
2.5	Notes on existing solutions of imbalanced data	53
2.6	INRIA person dataset	56
2.7	Summary of some common datasets used in object detection	
	field	57
2.8	The measurements used in some existing studies	61
2.9	Some proposed frameworks in literature	63
3.1	The relations among phases, study questions, study objectives, methods, chapters, and the evaluation	90
4.1	Comparison among SIFT, SURF, and NHOG regarding some	
	criterion	105
4.2	Summary of Datasets used	112
4.3	Number of nodes of hidden layer for each descriptor	113
4.4	Comparison between SURF and NHOG in terms of classification performance on Faces dataset	114
4.5	Comparison between SIFT and NHOG in terms of classification performance on Faces dataset	114
4.6	Comparison between SURF and NHOG in terms of classification performance on Cars dataset	116
4.7	Comparison between SIFT and NHOG in terms of	

	classification performance on Cars dataset	116
4.8	Comparison between SURF and NHOG in terms of point localization on faces dataset evaluated by Precision measurement	
4.9	Comparison between SIFT and NHOG in terms of point localization on faces dataset evaluated by Precision measurement	119
4.10	Comparison between SURF and NHOG in terms of point localization on cars dataset evaluated by Precision measurement	119
	measurement	120
4.11	Comparison between SIFT and NHOG in terms of point localization on cars dataset evaluated by Precision measurement	
		120
4.12	Comparison between SURF and NHOG regarding to rotation	
	changes	122
4.13	Comparison between SIFT and NHOG regarding to rotation changes	122
4.14	Comparison between SURF and NHOG regarding to Scale	
	changes	125
4.15	Comparison between SIFT and NHOG regarding to Scale changes	126
4.16	Time consumption comparison among SURF, SIFT, and NHOG descriptor based on single classifier	128
1 17	Paired Samples Test on faces dataset single-based classifier	132
4.17	Paired Samples Statistics on faces dataset, single-based	132
4.10	classifier	132
4 10	Daired Samples Test on ears dataset single based elessifier	132
4.19	Paired Samples Test on cars dataset, single-based classifier	155
4.20	classifier	122
5.1	Comparison between SURF and NHOG in terms of classification performance on Faces dataset for E1 and E2	133
	environments	145
5.2	Comparison between SIFT and NHOG in terms of classification performance on Faces dataset for E1 and E2	145
	environments	146
5.3	Comparison between SURF and NHOG in terms of classification performance on faces dataset based on	

	ensemble learning technique	
		148
5.4	Comparison between SIFT and NHOG in terms of classification performance on faces dataset based on ensemble learning technique	
		148
5.5	Comparison between SURF and NHOG in terms of classification performance on cars dataset for E1 and E2 environments	
		150
5.6	Comparison between SIFT and NHOG in terms of classification performance on cars dataset for E1 and E2 environments	
		150
5.7	Comparison between SURF and NHOG in terms of classification performance on cars dataset based on ensemble learning technique	153
5.8	Comparison between SIFT and NHOG in terms of classification performance on cars dataset based on ensemble learning technique	
		153
5.9	Comparison between SURF and NHOG in terms of point localization on faces dataset evaluated by Precision measurement based on ensemble classifiers	
		155
5.10	Comparison between SIFT and NHOG in terms of point localization on faces dataset evaluated by Precision measurement based on ensemble classifiers	
		155
5.11	Comparison between SURF and NHOG in terms of point localization on cars dataset evaluated by Precision measurement based on ensemble classifiers	
	incastrement based on ensemble classifiers	156
5.12	Comparison between SIFT and NHOG in terms of point localization on cars dataset evaluated by Precision	
	incastrement based on ensemble classifiers	156
5.13	Time consuming, SURF vs. NHOG and SIFT vs. NHOG applied on faces dataset	157
5.14	Time consuming, SURF vs. NHOG and SIFT vs. NHOG applied on cars dataset	160
5.15	Paired Samples Test on faces dataset, Single-based vs. Ensemble-based techniques	168
5 16	Paired Samples Statistics on faces dataset for Single based	100
5.10	and Ensemble-based techniques	168

5.17	Paired Samples Test on cars dataset, Single-based vs.		
	Ensemble-based techniques	169	
5.18	Paired Samples Statistics on cars dataset for Single-based and Ensemble-based techniques	169	
6.1	TPP, FP, TP, and Time consuming regarding the octave and hidden layer settings	176	
6.2	TPP, FP, TP, and Time consuming using NHOG+ with different hidden layer settings	177	
6.3	The average of time (in seconds) and the improvement percentage for NHOG+ over NHOG applied on faces and cars datasets		
		179	
6.4	TPP, FP,TP, and Precision of NHOG+ applied on faces and cars datasets	182	
6.5	Paired Samples Test regarding time consuming, NHOG vs. +NHOG	185	
6.6	Paired Samples Statistics regarding time consuming, NHOG +vs. NHOG	186	
6.7	Paired Samples Test regarding Precision metric, NHOG vs. +NHOG	187	
6.8	Paired Samples Statistics regarding Precision metric, NHOG +vs. NHOG	107	
		10/	

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
1.1	Scenario leading to the study problem	6
2.1	The most common object detection challenges	16
2.2	Taxonomy of the main steps and elements of object detection system	18
2.3	Taxonomy of features in object detection	19
2.4	Most common points' detectors methods	22
2.5	Convolved Gaussian (left), DOG (right) proposed by Lowe (2004)	30
2.6	Maxima and minima of DOG, Lowe (2004)	31
2.7	Interest point descriptor, Lowe (2004)	32
2.8	Computing the upright rectangle area using integral Image Bay et al. (2008)	34
2.9	LOG Approximation (top row), Bay formula (2.13) (bottom (row	35
2.10	Scale-Space: SIFT (left), SURF (right)	35
2.11	Assigning the Orientation, Bay et al. (2008)	36
2.12	Components of SURF Descriptor	37

2.13	Taxonomy of machine learning techniques	39
2.14	Standard architecture of Back-propagation Neural Network	43
2.15	Steps of BP neural network	45
2.16	Taxonomy of imbalanced data solutions	51
2.17	Bi-level filters for Center-Surround (Agrawal, et al., 2008)	65
2.18	Proposed approach for face detection (D. Kim and Dahyot, 2008)	68
2.19	Fusion schemes proposed by (Ludwig, et al., 2009). (a) scheme1, (b) scheme2, (c) scheme3	70
2.20	Cascade classifiers structure (Wang, et al., 2012)	71
3.1	The Study Framework	76
3.2	Some examples of Caltech 101 "faces" dataset, the original images	79
3.3	Some examples of an isolated faces clipped from Caltech 101 "faces" dataset, for training	79
3.4	Indoor examples from SUN397 dataset, for training	80
3.5	Outdoor examples from SUN397 dataset, for training	80
3.6	Some examples of an isolated faces clipped from Caltech 101 "faces" dataset, for classification test	82
3.7	Some examples of non-object from SUN397, for classification test	82
3.8	Some unlabelled images from Caltech 101 "faces", for localization test	83
3.9	Some examples of car object used for training	85

3.10	Some examples of non-object (street) used for cars training	85
3.11	Some examples of non-object (open space) used for cars training	86
3.12	Some examples of car object used for classification test	87
3.13	Some examples of non-object used for classification test	87
3.14	Some unlabelled images from VOC2009, for localization test	88
4.1	SIFT descriptor, Lowe (2004)	97
4.2	Orientation assigning, Bay et al. (2008)	98
4.3	SURF Descriptor Components	99
4.4	Points' detection. (a) original image, (b) detected points	100
4.5	Building up the window around interest point	100
4.6	NHOG features extraction elements: (a) window gradients, (b) nested-windows of gradients, (c) the considered gradients, (d) pyramid representation	103
4.7	The involved process in NHOG	104
4.8	General framework for object detection based on interest points	106
4.9	Training Procedure	108
4.10	Testing Procedure	110
4.11	ROC curve for SURF and NHOG according to hidden layer settings on Faces dataset. (left) H1, (right) H2	115
4.12	ROC curve for SIFT and NHOG according to hidden layer settings on Faces dataset. (left) H1, (right) H2	115
4.13	ROC curve for SURF and NHOG according to hidden layer	117

4.14	ROC curve for SIFT and NHOG according to hidden layer				
	settings on Cars dataset. (left) H1, (right) H2	118			
4.15	Comparison between SURF and NHOG in regards to				
	average of FP along all images at θ	123			
4.16	Comparison between SIFT and NHOG in regards to average				
	of FP along all images at θ	124			
4.17	Comparison between SURF and NHOG in regards to				
	average of FP along all images at different scale value	127			
4.18	Comparison between SIFT and NHOG in regards to average				
	of FP along all images at different scale value	127			
4.19	Time consumption for each image using SURF and NHOG				
	descriptor based on single classifier	129			
4.20	Time consumption for each image using SURF and NHOG				
	descriptor based on single classifier	129			
5.1	Learning and testing phase of the developed technique. (a)				
	learning, (b) testing	139			
5.2	ROC curve for SURF and NHOG according to E1 and E2	–			
	environments on Faces dataset with H1 and H2 setting. (top) H1. (bottom) H2	147			
5.2					
5.5	environments on Faces dataset with H1 and H2 setting. (top)				
	H1, (bottom) H2	147			
5.4	ROC curve for SURF and NHOG according to H1 and H2				
	applied on Faces dataset based on ensemble classifiers. (left)				
	H1, (right) H2	149			
5.5	ROC curve for SIFT and NHOG according to H1 and H2				
	applied on Faces dataset based on ensemble classifiers. (left)	149			

H1, (right) H2

5.6	ROC curve for SURF and NHOG according to E1 and E2 environments on cars dataset with H1 and H2 setting. (top)	
	H1, (bottom) H2	151
5.7	ROC curve for SIFT and NHOG according to E1 and E2 environments on cars dataset with H1 and H2 setting. (top)	
	H1, (bottom) H2	152
5.8	ROC curve for SURF and NHOG according to H1 and H2	
	applied on cars dataset based on ensemble classifiers. (left)	
	H1, (right) H2	154
5.9	ROC curve for SIFT and NHOG according to H1 and H2	
	applied on cars dataset based on ensemble classifiers. (left)	
	H1, (right) H2	154
5.10	Comparison between SURF and NHOG in terms of time	
	consuming for each image using ensemble-based technique	
	applied on faces dataset. (a) H1, (b) H2	158
5.11	Comparison between SIFT and NHOG in terms of time	
	consuming for each image using ensemble-based technique	
	applied on faces dataset. (a) H1, (b) H2	159
5.12	Comparison between SURF and NHOG in terms of time	
	consuming for each image using ensemble-based technique	
	applied on cars dataset.(a) H1, (b) H2	161
5.13	Comparison between SIFT and NHOG in terms of time	
	consuming for each image using ensemble-based technique	
	applied on cars dataset. (a) H1, (b) H2	162
5.14	Comparison between single-based and ensemble-based	
	techniques for SURF and NHOG descriptors on faces	
	dataset in regard to H1 and H2 setting. (top) H1, (bottom)	163
	H2	

5.15	Comparison between single-based and ensemble-based				
	techniques for SIFT and NHOG descriptors on faces dataset				
	in regards to H1 and H2 setting. (top) H1, (bottom) H2	164			
5.16	Comparison between single-based and ensemble-based				
	techniques for SURF and NHOG descriptors on cars dataset				
	in regards to H1 and H2 setting. (top) H1, (bottom) H2	165			
5.17	Comparison between single-based and ensemble-based				
	techniques for SIFT and NHOG descriptors on cars dataset				
	in regards to H1 and H2 setting. (top) H1, (bottom) H2	166			
6.1	Filters sizes for each octave used in SURF	174			
6.2	SURF point detector, the relation between scale and the				
	detected interest points	174			
6.3	Initial proposed schemes to speed up point's detector, a) 4				
	octaves, b) 3 octaves	175			
6.4	Proposed scheme to speed up point's detector, NHOG+	177			
6.5	Some examples of object localization regarding the octave				
	settings and hidden layer settings, left half H1, right half H2.				
	(a) 4 octaves, (b) 3 octaves, (c) odd octaves.	178			
6.6	Comparison between NHOG and NHOG+ in terms of time				
	consumption regarding hidden layer setting applied on faces				
	dataset. (a) H1, (b) H2	180			
6.7	Comparison between NHOG and NHOG+ in terms of time				
	consumption regarding hidden layer setting applied on cars				
	dataset. (a) H1, (b) H2	181			
6.8	Comparison between NHOG and NHOG+ in terms of FP				
	error regarding hidden layer setting applied on faces dataset.				
	(a) H1, (b) H2	183			

Comparison between NHOG and NHOG+ in terms of FP error regarding hidden layer setting applied on cars dataset. (a) H1, (b) H2 184

LIST OF ABBREVIATIONS

ANN	-	Artificial Neural Network
B-SIFT	-	Binary SIFT
BRIEF	-	Binary Robust Independent Elementary Features
BRISK	-	Binary Robust Invariant Scalable Key-points
CCR	-	Correct Classification Rate
CenSurE	-	Center Surround Extermas
CV	-	Computer Vision
DOC	-	Difference of Gaussian
DR	-	Detection Rate
FAST	-	Features from Accelerated Segment Test
FLDA	-	Fishier Linear Discriminated Analysis
FN	-	False Negative
FNR	-	False Negative Rate
FP	-	False Positive
FPR	-	False Positive Rate
GLOH	-	gradient location-orientation histogram
HOG	-	Histogram of Oriented Gradient
IM-SURF	-	Improved SURF
LOG	-	Laplacian of Gaussian
NHOG	-	Nested-windows of Histogram of Oriented Gradients

	٠	
Y Y	1	٩7
AA		v

- ORD Oriented FAST and Rotated BRIEF
- PCA-SIFT Principal Component Analysis SIFT
- ROC Receiver Operating Characteristic
- SIFT Scale Invariant Feature Transform
- SURF Speeded Up Robust Feature
- SVM Support Vector Machine
- TCB Totally Corrective Boosting
- TN True Negative
- TNR True Negative Rate
- TP True Positive
- TPR True Positive Rate

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	False Positive and True Positive over all images applied on faces and cars datasets using single-based classifier	210
В	False Positive and True Positive over all images applied on faces and cars datasets using ensemble-based classifier	226
С	False Positive and True Positive over all images applied on faces and cars datasets using ensemble-based classifier and NHOG+ scheme	242

CHAPTER 1

INTRODUCTION

1.1 Introduction

Computer vision (CV) is an area that includes approaches and techniques to acquire, process, analyze, and understand an image. Moreover, it produces a new form of symbols information as high-dimensional data from the natural world for further decisions (Morris, 2004). The idea of developing this field is to inspire the abilities of human vision to understand an image by machine (Sonka, 2007). Therefore, an image can be understood by seeing it in another form (i.e. numerical or symbolic information), this information can be extracted by using specific models such as geometry, statistic, learning technique or by combining two or more methods (Forsyth and Ponce, 2002). Additionally, CV is a relatively new area and has great relationship with many other fields such as Machine Learning, Physics, Control Systems, and Medical Imaging. Therefore, it is reasonable that CV has been rapidly developed into many kinds of sub-fields (Calonder, 2010a). Sub-fields of computer vision include, but not limited to, reconstruction of scene, object and events detection, object tracking, recognition of an object, video indexing, motion estimation, and image restoration.

One of the most important processes in CV is the object detection because it is essential process for many other applications such as object tracking, object recognition, object categorization, event detection, and searching for object in sequence of images or video. In addition, detecting and finding a specific object in sequence images or video has become an increasingly important problem with numerous applications, especially with the popularity of smart phones and other devices. That includes searching for people, buildings, landmarks and other interest objects for further analyzing and decisions (Aly, 2011).

1.2 Problem Background

Visual object detection is an extremely troublesome computational issue. The main problem is that each object in the world can cast an unbound number of diverse 2-D images onto the retina as the object's position, pose, lighting, and background changes in respect to the viewer (DiCarlo and Cox, 2007). The variations occurs due to the deformations of non-rigid, and intra-class variability because of shape and other visual properties, also increase the difficulty of tackling this problem (Felzenszwalb et al., 2010). Furthermore, the quality of object detection (detection rate, precision, error) and time consumption add more difficulties to object detection system in which they are trade-off problems.

Full object detection system normally consists of three main elements; features, recognition, and localization. Features can be divided into two types, according to the area (spatial) that they are taken from, and the methods of feature's acquisition. Features based on spatiality can be global or local (Aly, 2011; Li B. et al., 2011; Pang et al., 2012); while the ways of getting features can be categorized into three types; extraction, selection, composition features (Lillywhite et al., 2012b).

Recognizing an object can be done based on either matching or learning technique. Matching technique recognize the object based on the distance between the features of the target and the template stored. On the other hand, the learning technique is a way to build a system that mimic the human ability of learning from examples and it can recognize unknown instance based on prior knowledge and experience. Machine learning techniques can be classified into two types; singlebased and ensemble-based techniques (Xu et al., 2011; Zhang X. and Cheng, 2012). In single-based method, only one algorithm is used to learn and classify the instance (Xu et al., 2011). On the other hand, the ensemble-based approach uses a several algorithms in a certain arrangement to get the final decision of instance recognition (Ceamanos et al., 2010; Galar et al., 2012; Liu and Wang, 2010; Rahman and Verma, 2011; Verikas et al., 2010; Xu et al., 2011).

Features play an important role in object detection and affect directly on the quality of detection (Yilmaz et al., 2006). For any object detection within an image based on features, object needs to be represented by features as a first process by which can be used later in the classification phase (Aly, 2011). Features can be divided into two types; global and local.

Global features: where an image can be represented by extracting features as one dimension which describe the whole image. Information extracted might be histogram of color (Forsyth and Ponce, 2002), or histogram of orientation as in (Dalal and Triggs, 2005). global features are fast and easy in computation which is considered advantage over local features, but they perform worse than local features in term of accuracy (Douze et al., 2009).

Local features: where the features that represent image are extracted from a set of regions by specific descriptor. Generally, there are two steps to extract the local features which are feature detection and feature description (Forsyth and Ponce, 2002; Mikolajczyk and Schmid, 2004a). Feature detection involves detecting interest locations in the image (i.e. corners and edges). Feature description is to describe the image patch that surrounds the interest point; this can be done by histograms of gradients or orientation. There are different types of feature detectors; however among the most well-known ones are Difference of Gaussians (DOG) which was

presented by Lowe (2004), Hessian-Affine and Harris-Affine which were introduced by Mikolajczyk and Schmid (2004), and SURF which brought by Bay et al. (2008) and it is relatively new. Similarly, there are many feature descriptors, SIFT and SURF are the most widely used because of their performance. The local features have superior performance over the global ones; but with large memory usage (Douze et al., 2009).

Among all local features, SIFT and SURF are the dominant features over all others in this field in terms of robustness and invariance (Li H. et al., 2013; Pang et al., 2012). However, the high dimensionality and computation complexity of such features consume a large amount of space (memory) and take longer time in recognition, especially against large-scale dataset e.g. SIFT. Besides, SURF descriptor might not make use the full information about the interest point (Li H. et al., 2013); further, SURF has limitation regarding scale changes in case of zoom-out variant (Bay et al., 2008). So, these factors are challenges in some applications especially in real-time and mobile-based applications (Heinly et al., 2012; Li H. et al., 2013; Rister et al., 2013). These challenges give a motivation to introduce new features by converting the original features of SIFT or SURF into a new binary form. Examples of such binary-based features include, Binary Robust Independent Elementary Features (BRIEF) (Calonder et al., 2010b), Oriented FAST and Rotated BRIEF (ORD) (Rublee et al., 2011), and Binary Robust Invariant Scalable Keypoints (BRISK) (Leutenegger et al., 2011). Binary-based methods introduced features that are faster and easy to be computed, however, they still inaccurate, instable, and less descriptive than the original ones (Heinly et al., 2012; Ni, 2012).

Regarding the object recognition; matching technique is used to find the corresponding points between two images depending on the distances between the accompanying features. It is common technique used in object recognition and image registration (Brown et al., 2011). But, computing the descriptor and using the matching against large database adds more complexity to the recognition system (Gauglitz et al., 2011). It is also difficult to be used in object category recognition or

to recognize the part-based object. Moreover, it is not proper for some applications such as searching for an isolated object (Gauglitz et al., 2011).

As alternative to the matching technique, the learning-based technique, it has proven to be an effective method in object detection (Qiu et al., 2012). It is used successfully in category classification especially discriminative-based learning (Gu and Ren, 2010b), also researchers are paid more attention to design method that learn the interest point features (Li B. et al., 2011). Examples of such techniques are, Artificial Neural Network (ANN), boosting, decision trees, and Support Vector Machines (SVM) (Porikli and Yilmaz, 2012).

Since the data in object detection is an imbalance, this is a big challenge facing the learning-based methods as a single learning because the single-based method expect that all classes have the same distribution (He and Garcia, 2009; Sun et al., 2009), which leads to poorly accuracy of classification (Ceamanos et al., 2010; Galar et al., 2012; He and Garcia, 2009; Sun et al., 2009). Therefore, most conventional (standard or single-based) methods fail to be an effective to fix such problem (Ceamanos et al., 2010; Galar et al., 2012; He and Garcia, 2009).

To tackle this challenge, ensemble-based techniques are proposed to improve the accuracy of the classification (Ceamanos et al., 2010; Galar et al., 2012), and to reduce the effects of imbalanced data on the classification quality (Galar et al., 2012; Zhang X. and Cheng, 2012; Zhang Y. et al., 2012). however, this type of learning has structure complexity and need longer time than single-based learner (Xu et al., 2011).

Based on the afore-stated issues; Figure 1.1 summarizes the advantages, limitation of each parts of object detection system and the scenario that led to the study problem.



Adv.=Advantages, Disadv.=Disadvantages

Figure 1.1 : Scenario that led to the study problem

From Figure 1.1, the required solutions of object detection for both features and recognition can be concluded. Regarding the features; Features should be: Accurate, Fast, Invariant, and Easy to be computed. On the other hand; recognition technique should be: Accurate, Fast, and Manage and reduce the effect of imbalance data on recognition quality.

1.3 Problem Statement

Full object detection system consists of three parts, features extraction, object recognition, and object localization. Most researches in literature focus only on one part of the whole system; each part affects the performance of the object detection system to some degree. Moreover, other challenges such as changes in rotation, scale, and imbalanced data effect are also affecting and complicating the system structure. Therefore, the performance of object detection system could be improved by introducing a new descriptor that produces fast, distinctive and invariant features, reducing the effect of imbalanced data by improving ensemble-based learning technique, and modifying the original interest points' detector by which the system can be carried out faster.

1.4 Research Questions

Research questions are:

- How to overcome the shortages of existing descriptors in terms of accuracy and time consumption.
- (ii) How to tackle the imbalance data and gain a high quality of recognition.

- (iii) How to further reduce the detection errors; False Positive Rate (FPR) and False Positive (FP), and keep enough interest points for object localization.
- (iv) How to speed up the whole system of object detection without impact on precision of detection.

1.5 Research Aim

The aim of this research is to introduce a general model for object detection in which it can overcome the limitation of existing methods regarding the descriptor, classification, and speed. The proposed model consists of three parts; *first*, new descriptor based on interest point detectors which is fast and accurate in the same time; *second*, an ensemble-based improved learner based on ANN technique applied on different environments; *third*, an improved scheme to speed up the object detection system; all are integrated in one model.

1.6 Research Objectives

The objectives of this research can be summarized as follows:

(i) To propose a new local descriptor that is faster and more accurate than the most common existing descriptors taken into account the stability under different conditions such as changes in scale and rotation.

- (ii) To develop an effective supervised learning technique to learn the features of the proposed descriptor in (i) based on an ensemble learning in which it can increase the *accuracy*, reduce the *error*, and tackle the imbalance data challenge.
- (iii) To propose a developed scheme that speed up the whole system resulting of the complexity in (ii) with no cost in quality of recognition.

1.7 Research Scope and Assumptions

Study scope and assumptions are as follows:

- (i) Objects used in this study are; human faces and cars.
- (ii) Scenes and environments in which the object is detected are:
 - a. Indoor and outdoor for *faces* detection.
 - b. Street and open space for *cars* detection.
- (iii) Datasets used are:
 - a. Caltech 101 "faces"; for learning and testing.
 - b. PASCAL VOC 2009 "cars"; for testing.
 - c. SUN397; for learning.
 - d. Data for cars learning was collected by the researcher.
- (iv) Measurements used in this research are: Correct classification Rate
 (CCR), True Positive Rate (TPR) or Detection Rate (DR), False Positive
 Rate (FPR), Precision, and False Positive (FP) or 1-Precision. The detail
 in chapter 3.

- (v) Comparisons done as follows:
 - a. Comparing the proposed descriptor against SIFT and SURF descriptors based on measurements mentioned in (iv); besides, comparing the invariance to rotation and scale changes.
 - b. Comparing the ensemble-based classifiers against single-based classifier.
 - c. Time consumption comparisons; the proposed descriptor against SIFT and SURF descriptor, and the proposed scheme (that speed up the object detection system) against ensemble-based classifiers.
- (vi) T-test was used to validate the obtained results regarding the FP error.

1.8 Research Significance

Object detection and its application play an important role in different fields as mentioned earlier; thus, this research takes the same importance beside the additional characteristics as follows:

- Usability: system can be used to detect any object by only providing positive and negative examples into the system; no extra changes are needed.
- (ii) Development: different types of points' detectors can be used if there, the structure of the model remains unchanged. In addition, easy to change the classifier used by another or adding another one beside the existing one, model structure remains also unchanged.

Contributions of this study can be summarized as follows:

- A new local descriptor that is fast, accurate and overcome the existing descriptors in terms of speed, accuracy, and invariance.
- (ii) A developed learning technique based on ensemble learning that produced higher accuracy than single-based learning technique and fixed the imbalance data problem.
- (iii) A developed scheme that speeded up the object detection system by modifying the original SURF point's detector.
- (iv) Tow labelled datasets prepared from datasets mentioned in section 1.7 for training and testing phases.

1.10 Terms Definition

Due to using some terms in different domains, misconception or confusion may occurs; thus, this section provides the definition of terms used along this study, these definitions were taken from literature detailed in chapter 2.

(i) Interest Point

The most important point that can repeatable over scale-spaces, it can be edge or corner; sometimes it called *key-point*.

(ii) Detector

It is the method that detects the interest point.

(iii) Descriptor

Descriptor is the method that describe the detected interest point; describing the interest point means extract the substantial information about that point. Sometimes the method called *extractor* and the extracted information called *feature*.

(iv) Object Recognition

Object recognition is this study means, distinguishing the patterns that belonging to the object class from that they are not; it answer the question "Is there object within scene or not?" This includes the matching technique and learning methods.

(v) Object Localization

Objection localization means if there an object exist within scene, determine the location of that object; so, it answer the question "Where the object location within scene?"

(vi) Object detection

The meaning of object detection embodies both object recognition and localization altogether.

(vii) Single-based classifier

Normally, single-based classifier means only one classifier used for learning and classification of patterns; sometimes it called traditional or conventional method.

(viii) Ensemble-based classifiers

Ensemble-based methods often refer to a collection of classifiers, either by combining different classification techniques or variants of the same technique.

(ix) An imbalanced data

Data that has minority and majority among its classes is called imbalanced data; that's mean the distribution of instances for classes is not equal.

1.11 Thesis Outline

This thesis consists of 7 chapters organized as follows: Chapter 1 introduces an overview on object detection system and its applications; besides, still problems and challenges, problem statement, study questions, study objectives, scope of the study, the aim and significance of the study and study contributions, all were explained. Chapter 2 illustrates the literature review and related studies to object detection, this includes the challenges with respect to features, recognition, localization, datasets, and evaluation measurements used in this area; in addition, the proposed solution in literature for each and still challenges. The research methodology of this study is illustrated in Chapter 3. Chapter 4 describes the proposed descriptor in detail which then integrated with a traditional classification technique as one model. Chapter 5 presents the solution for imbalanced data problem by developing an effective technique based on ensemble classifiers. A developed scheme for speeding up the object detection system is introduced in Chapter 6. Research summary, contributions, and the suggestions for the future work are detailed in Chapter 7.

REFERENCES

- Aanæs, H., Dahl, A. L., and Pedersen, K. S. (2012). Interesting Interest Points. Int J Comput Vis. 97, 18-35.
- Agrawal, M., Konolige, K., and Blas, M. (2008). Censure: Center surround extremas for realtime feature detection and matching. *Computer Vision–ECCV 2008*, 102-115.
- Alwaili, S. (2007). Image-based object detection and identification. Doctor Philosophy, Universiti Teknologi Malaysia, Faculty Geoinformation Science and Engineering.
- Aly, M. A.-E.-D. (2011). Searching Large-Scale Image Collections. Doctor Philosophy, California Institute of Technology, Pasadena, California.
- Barczak, A. L. (2007). Feature Based Rapid Object Detection: From Feature Extraction to Parallelisation. Doctor of Philosophy. Massey University, Auckland, New Zealand.
- Batista, G. E., Prati, R. C., and Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. ACM Sigkdd Explorations Newsletter, 6(1), 20-29.
- Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (SURF). Computer Vision and Image Understanding. 110(3), 346-359.
- Brown, M., Hua, G., and Winder, S. (2011). Discriminative learning of local image descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 33(1), 43-57.

- Calonder, M. (2010a). Robust, High-Speed Interest Point Matching for Real-Time Applications. Doctor Philosophy, Swiss Federal Institute of Technology, Lausanne, Switzerland.
- Calonder, M., Lepetit, V., Strecha, C., and Fua, P. (2010b). Brief: Binary robust independent elementary features *Computer Vision–ECCV 2010*. 778-792: Springer.
- Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., and Fua, P. (2012c). BRIEF: Computing a local binary descriptor very fast. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 34(7), 1281-1298.
- Caruana, R., and Niculescu-Mizil, A. (2006a). An empirical comparison of supervised learning algorithms. Proceedings of the 23rd international conference on Machine learning: ACM, New York, NY, USA, 96-103.
- Caruana, R., Karampatziakis, N., and Yessenalina, A. (2008b). An empirical evaluation of supervised learning in high dimensions. Proceedings of the 25th international conference on Machine learning: ACM, New York, NY, USA, 161-168.
- Ceamanos, X., Waske, B., Benediktsson, J. A., Chanussot, J., Fauvel, M., and Sveinsson, J. R. (2010). A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. *International Journal of Image and Data Fusion*, 1(4), 293-307.
- Chapelle, O., Schölkopf, B., and Zien, A. (2006). *Semi-supervised learning*. Vol. 2: MIT press Cambridge.
- Chawla, N. V., Japkowicz, N., and Kotcz, A. (2004a). Editorial: special issue on learning from imbalanced data sets. ACM Sigkdd Explorations Newsletter, 6(1), 1-6.
- Chawla, N. V., Cieslak, D. A., Hall, L. O., and Joshi, A. (2008b). Automatically countering imbalance and its empirical relationship to cost. *Data Mining and Knowledge Discovery*, *17*(2), 225-252.

- Cigizoglu, H. K., and Alp, M. (2006). Generalized regression neural network in modelling river sediment yield. Advances in Engineering Software, 37(2), 63-68.
- Dai, Q., and Hoiem, D. (2012). Learning to localize detected objects. Proceedings
- of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, Providence, RI: IEE, 3322-3329.
- Dalal, N., and Triggs, B. (2005). *Histograms of oriented gradients for human detection*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, San Diego, CA, USA: IEE, 886-893.
- De Bock, K. W., Coussement, K., and Van den Poel, D. (2010). Ensemble classification based on generalized additive models. *Computational Statistics* and Data Analysis, 54(6), 1535-1546.
- DiCarlo, J. J., and Cox, D. D. (2007). Untangling invariant object recognition. *Trends in cognitive sciences*, 11(8), 333-341.
- Dogan, E., Sengorur, B., and Koklu, R. (2009). Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *Journal of Environmental Management*, 90(2), 1229-1235.
- Douze, M., Jégou, H., Sandhawalia, H., Amsaleg, L., and Schmid, C. (2009). Evaluation of gist descriptors for web-scale image search. Proceedings of the ACM International Conference on Image and Video Retrieval CIVR. New York, NY, USA: ACM,19
- Ebrahimi, M., and Mayol-Cuevas, W. W. (2009). SUSurE: Speeded up surround extrema feature detector and descriptor for realtime applications. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, Miami, FL:IEEE, 9-14.
- Egrioglu, E., Aladag, C. H., and Gunay, S. (2008). A new model selection strategy in artificial neural networks. *Applied mathematics and computation*. 195(2), 591-597.

- Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International journal of computer vision*. 88(2), 303-338.
- Favelle, S. K., Palmisano, S., and Avery, G. (2011). Face viewpoint effects about three axes: The role of configural and featural processing. *Perception-London.* 40(7), 761.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*. 27(8), 861-874.
- Fei-Fei, L., Fergus, R., and Perona, P. (2006a). One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4), 594-611.
- Fei-Fei, L., Fergus, R., and Perona, P. (2007b). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*. 106(1), 59-70.
- Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2010). Object Detection with Discriminatively Trained Part-Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9), 1627-1645.
- Forsyth, D. A., and Ponce, J. (2002). *Computer Vision: A modern approach*. (Frist ed.): Prentice Hall.
- Freitas, A., Costa-Pereira, A., and Brazdil, P. (2007). Cost-sensitive decision trees applied to medical data. *Data Warehousing and Knowledge Discovery*. Springer, 303-312.
- Galar, M., Fernández, A., Barrenechea, E., Bustince, H., and Herrera, F. (2012). A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 42*(4), 463-484.

- Galleguillos, C., Babenko, B., Rabinovich, A., and Belongie, S. (2008). Weakly supervised object localization with stable segmentations *Computer Vision– ECCV 2008*: Springer,193-207
- Ganganwar, V. (2012). An overview of classification algorithms for imbalanced datasets. International Journal of Emerging Technology and Advanced Engineering. 2(4), 42-47.
- Gauglitz, S., Höllerer, T., and Turk, M. (2011). Evaluation of interest point detectors and feature descriptors for visual tracking. *International journal of computer vision.* 94(3), 335-360.
- Gil, A., Mozos, O. M., Ballesta, M., and Reinoso, O. (2010). A comparative evaluation of interest point detectors and local descriptors for visual SLAM. *Machine Vision and Applications*. 21(6), 905-920.
- Griffin, G., Holub, A., and Perona, P. (2007). Caltech-256 object category dataset.
- Gu, C., Lim, J. J., Arbeláez, P., and Malik, J. (2009a). *Recognition using regions*. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, Miami, FL:IEEE, 1030-1037.
- Gu, C., and Ren, X. (2010b). Discriminative mixture-of-templates for viewpoint classification. *Computer Vision–ECCV 2010*, 408-421.
- Harris, C., and Stephens, M. (1988). A combined corner and edge detector.
- Harzallah, H., Jurie, F., and Schmid, C. (2009). *Combining efficient object localization and image classification*. ICCV, Kyoto, 237-244.
- He, H., and Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284.
- Heinly, J., Dunn, E., and Frahm, J.-M. (2012). Comparative evaluation of binary features *Computer Vision–ECCV 2012*: Springer, 759-773.
- Heitz, G., Elidan, G., Packer, B., and Koller, D. (2009). Shape-based object localization for descriptive classification. *International journal of computer vision*. 84(1), 40-62.

- Hu, X., and Weng, Q. (2009). Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks. *Remote Sensing of Environment.* 113(10), 2089-2102.
- Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006a). Extreme learning machine: theory and applications. *Neurocomputing*. 70(1), 489-501.
- Huang, G.-B., Wang, D. H., and Lan, Y. (2011b). Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics*. 2(2), 107-122.
- Juan, L., and Gwun, O. (2009). A comparison of sift, pca-sift and surf. *International Journal of Image Processing*. 3(4), 143-152.
- Ke, Y., and Sukthankar, R. (2004). PCA-SIFT: A more distinctive representation for local image descriptors. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition CVPR, 506-513.
- Khan, N. Y., McCane, B., and Wyvill, G. (2011). Sift and surf performance evaluation against various image deformations on benchmark dataset. International Conference on Digital Image Computing Techniques and Applications DICTA, Noosa, QLD:IEEE, 501-506.
- Kim, D., and Dahyot, R. (2008). Face components detection using SURF descriptors and SVMs. International Conference on Machine Vision and Image IMVIP, Portrush, 51-56.
- Kim, S., and Lee, J. (2012). Scale Invariant Small Target Detection by Optimizing Signal to Clutter Ratio in Heterogeneous Bacground for Infrared Search and Track. *Pattern Recognition.* 45, 393.
- KISHORE, G. N. (2011). *STUDY OF OBJECT DETECTION AND READING*. Doctor Philosophy, National Institute of Technology Rourkela.
- Kittler, J., Hatef, M., Duin, R. P., and Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3), 226-239.

- Kotsiantis, S. B., Zaharakis, I., and Pintelas, P. (2007). Supervised machine learning: A review of classification techniques.
- Lampert, C. H., Blaschko, M. B., and Hofmann, T. (2008). Beyond sliding windows: Object localization by efficient subwindow search. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition CVPR, Anchorage, AK:IEEE, 1-8.
- Laptev, I. (2009). Improving object detection with boosted histograms. *Image and Vision Computing*. 27(5), 535-544.
- Lei, Z., Fang, T., and Li, D. (2010). Histogram of oriented gradient detector with color-invariant gradients in Gaussian color space. *Optical Engineering*. 49(10), 1-7.
- Leutenegger, S., Chli, M., and Siegwart, R. Y. (2011). *BRISK: Binary robust invariant scalable keypoints*. IEEE International Conference on Computer Vision ICCV, Barcelona:IEEE, 2548-2555.
- Li, B., Xiao, R., Li, Z., Cai, R., Lu, B.-L., and Zhang, L. (2011). *Rank-SIFT: Learning to rank repeatable local interest points.* IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI:IEEE, 1737-1744.
- Li, C. (2007). Classifying imbalanced data using a bagging ensemble variation (BEV). Proceedings of the 45th annual southeast regional conference: ACM, New York, NY, USA,203-208.
- Li, H., Xu, T., Li, J., and Zhang, L. (2013). Face Recognition based on Improved SURF. Third International Conference on Intelligent System Design and Engineering Applications (ISDEA), Hong Kong, 755-758.
- Liang, Z., Wang, M., Zhou, X., Lin, L., and Li, W. (2012). Salient object detection based on regions. *Multimedia Tools and Applications*. 1-28.
- Lillywhite, K., Lee, D. J., and Zhang, D. (2009a). *Real-time human detection using histograms of oriented gradients on a GPU*. WACV, Snowbird, UT, 1-6.

- Lillywhite, K., Tippetts, B., and Lee, D. J. (2012b). Self-tuned Evolution-COnstructed features for general object recognition. *Pattern Recognition*. *45*(1), 241-251.
- Lindeberg, T. (1994a). Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*. 21(1-2), 225-270.
- Lindeberg, T. (1998b). Feature detection with automatic scale selection. International journal of computer vision. 30(2), 79-116.
- Liu, N., and Wang, H. (2010). Ensemble based extreme learning machine. *Signal Processing Letters, IEEE. 17*(8), 754-757.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision. 60(2), 91-110.
- Ludwig, O., Delgado, D., Gonçalves, V., and Nunes, U. (2009). Trainable classifierfusion schemes: an application to pedestrian detection. 12th International IEEE Conference on Intelligent Transportation Systems ITSC, St. Louis, MO:IEEE, 1-6.
- Mahiddine, A., Seinturier, J., Boï, J.-M., and Merad, P. D. D. (2012). Performances Analysis of Underwater Image Preprocessing Techniques on the Repeatability of SIFT and SURF Descriptors. 20th International Conference on Computer Graphics, Visualization and Computer Vision WSCG, 275-282.
- Mair, E., Hager, G. D., Burschka, D., Suppa, M., and Hirzinger, G. (2010). Adaptive and generic corner detection based on the accelerated segment test *Computer Vision–ECCV* : Springer, 183-196.
- McAuley, J. J., and Caetano, T. S. (2012). Fast matching of large point sets under occlusions. *Pattern Recognition*. 45(1), 563-569.
- Mikolajczyk, K., and Schmid, C. (2004a). Scale and affine invariant interest point detectors. *International journal of computer vision*. 60(1), 63-86.
- Mikolajczyk, K., and Schmid, C. (2005b). A performance evaluation of local descriptors. *PAMI*. 27(10), 1615–1630.

- Miller, I., Campbell, M., Huttenlocher, D., Kline, F. R., Nathan, A., Lupashin, S., et al. (2008). Team Cornell's Skynet: Robust perception and planning in an urban environment. *Journal of Field Robotics*. 25(8), 493-527.
- Mohan, A., Papageorgiou, C., and Poggio, T. (2001). Example-based object detection in images by components. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(4), 349-361.
- Morel, J.-M., and Yu, G. (2009). ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*. 2(2), 438-469.
- Morris, T. (2004). Computer Vision and Image Processing: Palgrave Macmillan.
- Nagarajan, B., and Balasubramanie, P. (2008). Object classification in static images with cluttered background using statistical feature based neural classifier. *Asian J. Inform. Technol.* 7, 162-167.
- Neubeck, A., and Van Gool, L. (2006). Efficient non-maximum suppression. IEEE International Cnoference on Pattern Recognition ICPR, Hong Kong, 850-855.
- Ni, Z.-S. (2012). B-SIFT: A Binary SIFT Based Local Image Feature Descriptor. Fourth International Conference on Digital Home (ICDH), Guangzhou, China,117-121.
- Opelt, A., Fussenegger, M., Pinz, A., and Auer, P. (2004). Weak hypotheses and boosting for generic object detection and recognition. *Computer Vision*-*ECCV 2004*. 71-84.
- Osadchy, M., and Keren, D. (2004). Efficient detection under varying illumination conditions and image plane rotations. *Computer Vision and Image Understanding*. 93(3), 245-259.
- Paliwal, M., and Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*. *36*(1), 2-17.
- Pang, Y., Li, W., Yuan, Y., and Pan, J. (2012). Fully affine invariant SURF for image matching. *Neurocomputing*. 85, 6-10.

- Papageorgiou, C. P., Oren, M., and Poggio, T. (1998a). A general framework for object detection. Sixth International Conference on Computer Vision, Bombay,555-562.
- Papageorgiou, C. P. (2000b). A trainable system for object detection in images and video sequences. Doctor Philosophy, MIT.
- Parks, D. H. (2006). *Object detection and analysis using coherency filtering*. Doctor Philosophy, McGill University.
- Pedersoli, M., Gonzàlez, J., Bagdanov, A. D., and Villanueva, J. J. (2010). Recursive coarse-to-fine localization for fast object detection *Computer Vision–ECCV*: Springer, 280-293.
- Pinto, N., Cox, D. D., and DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? *PLoS computational biology*. 4(1), e27.
- Ponce, J., Berg, T. L., Everingham, M., Forsyth, D. A., Hebert, M., Lazebnik, S., et al. (2006). Dataset issues in object recognition *Toward category-level object recognition* : Springer, 29-48.
- Porikli, F., and Yilmaz, A. (2012). Object Detection and Tracking. *Video Analytics* for Business Intelligence. 3-41.
- Qiu, X. N., Liu, S. R., and Song, J. T. (2012). Generic Object Detection Based on Boosting Embedded with Bag-of-Words. *Applied Mechanics and Materials*. 109, 285-289.
- Rahman, A., and Verma, B. (2011). Novel layered clustering-based approach for generating ensemble of classifiers. *IEEE Transactions on Neural Networks*, 22(5), 781-792.
- Rister, B., Wang, G., Wu, M., and Cavallaro, J. R. (2013). A fast and efficient SIFT detector using the mobile GPU. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, 2674-2678.
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Revie*. 33(1-2), 1-39.

- Rosten, E., and Drummond, T. (2006). Machine learning for high-speed corner detection. *Computer Vision–ECCV: Springer*, 430-443.
- Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). ORB: an efficient alternative to SIFT or SURF. IEEE International Conference on Computer Vision (ICCV), Barcelona, 2564-2571.
- Sahbi, H., and Geman, D. (2006). A hierarchy of support vector machines for pattern detection. *The Journal of Machine Learning Research.* 7, 2087-2123.
- Scherer, M., Walter, M., and Schreck, T. (2010). *Histograms of oriented gradients* for 3d object retrieval. 18th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision WSCG, 41-48.
- Schmid, C., and Mohr, R. (1997). Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(5), 530-535.
- Singh, K. P., Basant, A., Malik, A., and Jain, G. (2009). Artificial neural network modeling of the river water quality—a case study. *Ecological Modelling*. 220(6), 888-895.
- Singh, Y., Bhatia, P. K., and Sangwan, O. (2007). A review of studies on machine learning techniques. *International Journal of Computer Science and Security*. *1*(1), 70-84.
- Smith, S. M., and Brady, J. M. (1997). SUSAN—A new approach to low level image processing. *International journal of computer vision*. 23(1), 45-78.
- Sonka, M., Hlavac, V., and Boyle, R. (2007). Image processing, analysis, and machine vision. Cengage Learning.
- Su, H., and Chong, K. T. (2007). Induction machine condition monitoring using neural network modeling. IEEE Transactions onIndustrial Electronics, 54(1), 241-249.
- Sun, Y., Wong, A. K., and Kamel, M. S. (2009). Classification of imbalanced data: A review. International Journal of Pattern Recognition and Artificial Intelligence. 23(04), 687-719.

Szeliski, R. (2010). Computer vision: algorithms and applications: Springer.

- Tamminen, T., and Lampinen, J. (2004). A Bayesian occlusion model for sequential object matching. Proceeding of the British Conference on Machine Vision Conference, 547-556.
- Taylor, M. E., and Stone, P. (2009). Transfer learning for reinforcement learning domains: A survey. *The Journal of Machine Learning Research*. 10, 1633-1685.
- Trung, Q. N., Phan, D., Kim, S. H., Na, I. S., and Yang, H. J. (2014). Recursive Coarse-to-Fine Localization for Fast Object Detection. *International Journal* of Control and Automation. 7(1), 235-242.
- Tuytelaars, T., and Mikolajczyk, K. (2008). Local invariant feature detectors: a survey. *Foundations and Trends*® *in Computer Graphics and Vision.* 3(3), 177-280.
- Verikas, A., Kalsyte, Z., Bacauskiene, M., and Gelzinis, A. (2010). Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey. *Soft Computing*. 14(9), 995-1010.
- Viola, P., and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. IEEE conference on Computer Vision and Pattern Recognition CVPR, 511-518.
- Wang, P., Shen, C., Barnes, N., and Zheng, H. (2012). Fast and Robust Object Detection Using Asymmetric Totally Corrective Boosting. *IEEE Transactions on Neural Networks and Learning Systems*, 23(1), 33-46.
- Wu, G., and Chang, E. Y. (2005). KBA: Kernel boundary alignment considering imbalanced data distribution. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 786-795.
- Xiao, B., Ma, J. F., and Cui, J. T. (2012). Combined blur, translation, scale and rotation invariant image recognition by Radon and pseudo Fourier-Mellin transforms. *Pattern Recognition*. 45(1), 314–321.

- Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., and Torralba, A. (2010). Sun database: Large-scale scene recognition from abbey to zoo. IEEE conference on Computer vision and pattern recognition (CVPR), San Francisco, CA, 3485-3492.
- Xu, Y., Cao, X., and Qiao, H. (2011). An efficient tree classifier ensemble-based approach for pedestrian detection. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 41*(1), 107-117.
- Yang, Y., and Newsam, S. (2011). Spatial pyramid co-occurrence for image classification. IEEE International Conference on Computer Vision (ICCV), Barcelona, 1465-1472.
- Yeh, T., Lee, J. J., and Darrell, T. (2009). Fast concurrent object localization and recognition. IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR. Miami, FL, 280-287.
- Yilmaz, A., Javed, O., and Shah, M. (2006). Object tracking: A survey. Acm Computing Surveys (CSUR). 38(4), 13.
- Yu, T.-H., Woodford, O. J., and Cipolla, R. (2011). An evaluation of volumetric interest points. International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), Hangzhou, 282-289.
- Zhang, X., and Cheng, C. (2012). Imbalanced data classification algorithm based on boosting and cascade model. IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, 2861-2866.
- Zhang, Y., Zhang, D., Mi, G., Ma, D., Li, G., Guo, Y., et al. (2012). Using ensemble methods to deal with imbalanced data in predicting protein–protein interactions. *Computational biology and chemistry*. 36, 36-41.
- Zhang, Z., Cao, Y., Salvi, D., Oliver, K., Waggoner, J., and Wang, S. (2010). Freeshape subwindow search for object localization. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, 1086-1093.

Zhu, X., Vondrick, C., Ramanan, D., and Fowlkes, C. (2012). Do We Need More Training Data or Better Models for Object Detection? BMVC, Citeseer, 3, 1-11.