

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

FEATURE-BASED RAPID OBJECT DETECTION:  
FROM FEATURE EXTRACTION TO PARALLELISATION

A thesis presented in partial  
fulfilment of the requirements

for the degree of

Doctor of Philosophy in Computer Sciences

at Massey University, Auckland, New Zealand

Andre Luis Chautard Barczak

2007



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Research Scope . . . . .	2
1.3	Original Contributions . . . . .	2
1.4	Publications related to this work . . . . .	3
1.5	Outline of the Thesis . . . . .	4
<b>2</b>	<b>Literature Review</b>	<b>7</b>
2.1	Computer Vision and Rapid Object Detection . . . . .	7
2.1.1	Real-time Object Detection, Recognition and Tracking . . . . .	8
2.1.2	Face Detection . . . . .	11
2.1.3	Invariant Features and Non-invariant Features . . . . .	13
2.2	Rapid Object Detection: a review of the Viola and Jones Method . . . . .	15
2.2.1	Haar-like Features . . . . .	15
2.2.2	Summed-area Tables . . . . .	16
2.2.3	Adaboost . . . . .	18
2.2.4	Cascades of Classifiers . . . . .	22
2.2.5	Multiresolution Analysis in Feature-based Systems . . . . .	23
2.2.6	Lighting Conditions . . . . .	26
2.2.7	Articulated Objects and Motion Features . . . . .	26
2.3	Parallel Computing . . . . .	27
2.3.1	Types of parallel machines . . . . .	28
2.3.2	Beowulf Clusters . . . . .	29
2.3.3	Cluster Topology . . . . .	29
2.3.4	Bandwidth and Latency . . . . .	30
2.3.5	Speedup in parallel computers . . . . .	31
2.3.6	Message passing paradigm . . . . .	32
2.3.7	Parallel Image Processing . . . . .	32
2.4	Conclusions of Literature Review . . . . .	35

<b>3</b>	<b>Methods and Preliminary Experiments</b>	<b>37</b>
3.1	Implementing the Rapid Object Recognition Training algorithms . . . . .	38
3.1.1	Adaboost: a simple modification for convergence and speed . . . . .	38
3.1.2	Building Classifiers with a set of Convex Hulls . . . . .	42
3.1.3	Marking the positive examples . . . . .	43
3.1.4	Accuracy and Performance Measurements . . . . .	44
3.2	Preliminary Experiments . . . . .	45
3.3	Building Beowulf Clusters . . . . .	46
3.3.1	A brief description of the clusters . . . . .	49
3.4	Performance of Clusters . . . . .	51
3.4.1	Linpack performance results . . . . .	51
3.4.2	An analysis of performance using mpptests . . . . .	52
3.4.3	Ring test using jumpshot . . . . .	54
3.5	Summary . . . . .	56
<b>4</b>	<b>Partial Occlusions in Face Detection</b>	<b>57</b>
4.1	Introduction . . . . .	57
4.2	The Training Process . . . . .	58
4.3	Testing the classifiers . . . . .	60
4.4	Experimental results and discussion . . . . .	61
4.5	Summary . . . . .	63
<b>5</b>	<b>Hand Detection</b>	<b>65</b>
5.1	Related work . . . . .	65
5.2	Experiments with Haar-like features to recognise human hands . . . . .	67
5.2.1	Preparing the Positive Set . . . . .	67
5.2.2	Training . . . . .	68
5.2.3	Detection . . . . .	69
5.3	Results and Discussion . . . . .	70
5.3.1	Accuracy . . . . .	70
5.3.2	Performance . . . . .	71
5.4	Summary . . . . .	73
<b>6</b>	<b>Rotational-invariant Approach</b>	<b>77</b>
6.1	Background . . . . .	78
6.2	Error sources in the computation of Haar-like features . . . . .	80
6.2.1	Computing Haar-like features at $45^\circ$ . . . . .	81
6.2.2	Computing Haar-like features at $26.5^\circ$ and $63.5^\circ$ . . . . .	82
6.3	Converting features to generic angles . . . . .	84

6.3.1	Error Analysis for Pair of Equivalent Features (PEF) . . . . .	85
6.4	Experimental Results and Discussion . . . . .	86
6.4.1	Experiment 1: error analysis for PEFs using $0^\circ$ and $45^\circ$ . . . . .	86
6.4.2	Experiment 2: error analysis for PEFs using $0^\circ$ and $26.5^\circ$ . . . . .	89
6.4.3	Experiment 3: using PEFs to compute features in classifiers . . . . .	93
6.4.4	Experiment 4: converting the original face classifier using tilted features . . . . .	96
6.5	Summary . . . . .	98
<b>7</b>	<b>Real-time Moment Invariants</b>	<b>101</b>
7.1	Related work . . . . .	102
7.2	Geometric Moment Invariants . . . . .	103
7.2.1	Hu's equations . . . . .	104
7.2.2	Computing Moment Invariants from SATs . . . . .	107
7.2.3	Preliminary Experiments . . . . .	109
7.2.4	Limitations . . . . .	111
7.2.5	Lighting Contrast Stretching . . . . .	111
7.3	Concentric Discs . . . . .	114
7.3.1	Circular Area (Discs) . . . . .	115
7.3.2	Concentric Discs computation . . . . .	116
7.4	Face Detection . . . . .	122
7.4.1	Estimating angles using Hu's moments . . . . .	122
7.4.2	Classifiers for Face Detection using Moments . . . . .	123
7.5	Summary . . . . .	126
<b>8</b>	<b>Digits Recognition and Invariant Features</b>	<b>127</b>
8.1	Related Work . . . . .	128
8.2	Feature Extraction . . . . .	130
8.2.1	The Concentric Discs Method (CDMI) . . . . .	130
8.2.2	An Extended Feature Set for Normalised Central Moments . . . . .	130
8.3	Training Methods . . . . .	131
8.3.1	Convex Hull Classifiers . . . . .	131
8.3.2	AdaBoost . . . . .	132
8.3.3	Confidence Value for Multiple Classifiers . . . . .	132
8.4	Experimental results . . . . .	133
8.4.1	experiment 1: four types of classifiers . . . . .	134
8.4.2	experiment 2: Ada-88- $\psi$ with low training error . . . . .	135
8.4.3	Experiment 3: $\eta$ features with AdaBoost . . . . .	137
8.5	Discussion . . . . .	139

8.6	Summary . . . . .	144
<b>9</b>	<b>A Model for Parallel Classifiers</b>	<b>145</b>
9.1	A Generic Platform for a Mobile Cluster . . . . .	146
9.2	A Performance Model . . . . .	146
9.2.1	Single Cascade . . . . .	146
9.2.2	Concurrent Cascades . . . . .	149
9.2.3	Communication . . . . .	153
9.3	Implementing a mobile cluster using mini-ITX boards . . . . .	154
9.3.1	Speedup . . . . .	156
9.4	Summary . . . . .	157
<b>10</b>	<b>Final Conclusions</b>	<b>159</b>
10.1	Summary . . . . .	159
10.2	Limitations . . . . .	160
10.3	Future Work . . . . .	161
	<b>Appendices</b>	<b>165</b>
<b>A</b>	<b>Implementation of SATs at 26.5°</b>	<b>165</b>
<b>B</b>	<b>Moments with SATs</b>	<b>171</b>
B.1	Deriving The Equations For $\mu_{pq}$ From SATs ( $m_{pq}$ ) . . . . .	171
B.2	4 <sup>th</sup> Order Moment Invariants Derivation . . . . .	174
<b>C</b>	<b>Additional Results for Digits Recognition</b>	<b>179</b>
C.1	Experiment 1: additional Tables . . . . .	179
C.2	Experiment 2: additional Tables . . . . .	181
C.3	Experiment 3: additional Tables . . . . .	182
	<b>Bibliography</b>	<b>184</b>
	<b>Glossary</b>	<b>194</b>

# List of Figures

2.1	A generic computer vision Model . . . . .	9
2.2	Histograms as features for computer vision systems . . . . .	14
2.3	Examples of Haar-like Features. . . . .	15
2.4	The Summed-Area Table . . . . .	17
2.5	Examples of tilted Haar-like Features. . . . .	18
2.6	AdaBoost applied in a colour classification problem. . . . .	20
2.7	A cascade of classifiers . . . . .	23
2.8	Data decomposition . . . . .	30
2.9	Four schemes for Parallel Computer Vision. . . . .	34
3.1	Classification using thresholds does not converge. . . . .	38
3.2	This set converges if the the classifier has four stages. . . . .	40
3.3	Example of convex hull used as a classifier. . . . .	43
3.4	Marking a face from the CMU-MIT database . . . . .	44
3.5	ROC curve BASIC versus ALL features, 100 positive examples . . . . .	46
3.6	ROC curve BASIC versus ALL features, 250 positive examples . . . . .	47
3.7	ROC curve BASIC versus ALL features, 500 positive examples . . . . .	47
3.8	ROC curve BASIC versus ALL features, 750 positive examples . . . . .	48
3.9	ROC curve BASIC versus ALL features, 1000 positive examples . . . . .	48
3.10	The Helix Beowulf-type cluster's topology . . . . .	50
3.11	Linpack Rmax rating versus number of processors. . . . .	51
3.12	Point to point messages in Helix (short, blocking) . . . . .	52
3.13	Point to point messages in Sisters (short, blocking) . . . . .	53
3.14	Point to point messages in Helix (long, blocking) . . . . .	53
3.15	Point to point messages in Sisters (long, blocking) . . . . .	54
3.16	Goptest (Broadcast operation) running in 128 processors. . . . .	55
3.17	A ring test using 127KB message and 128 processors . . . . .	55
4.1	Examples of partial occluded faces found in the three sets. . . . .	58
4.2	Example of a false positive object within the negative set. . . . .	59



4.3	The occlusion adding positive examples . . . . .	60
4.4	Hits in several occluded faces. . . . .	62
4.5	ROC curve comparing Classifiers 2 and 3. . . . .	63
5.1	The basic hand images used on the training process. . . . .	67
5.2	Creating the positive set. . . . .	68
5.3	Using several cascades to recognise and determine rotation of hands . . .	69
5.4	The Haar-like features chosen for the first step of each cascade. . . . .	71
5.5	ROC curves comparing the classifiers trained at different angles for test set A. . . . .	73
5.6	Examples of detection at different angles (from test set A). . . . .	74
5.7	Examples of images in test set B . . . . .	75
5.8	Rates for cascades detecting hands in various angles . . . . .	75
6.1	Normal features conversion to other types . . . . .	80
6.2	Comparison between a normal and a tilted feature . . . . .	81
6.3	Comparison between a $0^\circ$ , a $26.5^\circ$ and a $63.5^\circ$ Haar-like feature. . . . .	82
6.4	Computing the SAT for $26.56^\circ$ recursively . . . . .	83
6.5	Corrections for SATs at $26.5^\circ$ . . . . .	84
6.6	Pair of Equivalent Features (PEF) . . . . .	85
6.7	An example of PEF . . . . .	86
6.8	Case where the value of the $22.5^\circ$ feature should be MAX. . . . .	87
6.9	Case where the value of the $22.5^\circ$ feature should be zero. . . . .	87
6.10	Scanning images for PEF comparison . . . . .	88
6.11	Measuring errors for PEFs . . . . .	88
6.12	First frame of Akiyo sequence and the chessboard images. . . . .	88
6.13	Maximum error vs angle for Akiyo (normal+ $45^\circ$ PEFs). . . . .	89
6.14	Average error vs angle for Akiyo (normal+ $45^\circ$ PEFs). . . . .	90
6.15	Maximum error vs angle for Chessboard (normal+ $45^\circ$ PEFs). . . . .	90
6.16	Average error vs angle for Chessboard (normal+ $45^\circ$ PEFs). . . . .	91
6.17	Maximum error vs angle for Akiyo (normal+ $26.5^\circ$ PEFs). . . . .	91
6.18	Maximum error vs angle for Chessboard (normal+ $26.5^\circ$ PEFs). . . . .	92
6.19	Average error vs angle for Akiyo (normal+ $26.5^\circ$ PEFs). . . . .	92
6.20	Average error vs angle for Chessboard (normal+ $26.5^\circ$ PEFs). . . . .	93
6.21	Hit ratio vs angle using 16 layers. . . . .	94
6.22	Hit ratio vs angle, kernel $36 \times 36$ . . . . .	94
6.23	False detection rate vs angle, kernel $36 \times 36$ . . . . .	95
6.24	The classifier $36 \times 36$ with various correction factors. . . . .	96
6.25	The hit rate for classifier $36 \times 36$ with limited range of features. . . . .	97

6.26	Detecting rotated faces using PEFs . . . . .	97
6.27	Converting Haar-like feature maintaining the proportional area . . . . .	98
6.28	Comparative ROC curve for upright and tilted cascades for the CMU-MIT database. . . . .	99
6.29	Some results for the tilted cascade converted from OpenCV. . . . .	99
7.1	Method overview . . . . .	103
7.2	Grey scale samples of digits 9. . . . .	110
7.3	A sample of the working algorithm . . . . .	111
7.4	Merging positive hits with Moments Invariants . . . . .	112
7.5	An example of contrast stretching . . . . .	113
7.6	Computing the moments for an approximation of a circular area . . . . .	117
7.7	The Concentric Discs Moment Invariants approach (CDMI) . . . . .	117
7.8	Test image from Gonzalez and Woods (2002). . . . .	118
7.9	Concentric disc area approximation . . . . .	121
7.10	Angles estimated for faces . . . . .	123
7.11	Face detection using CDMI. . . . .	124
7.12	The false detection rate versus the number of weak classifiers (1000 FERET samples) . . . . .	125
7.13	The false detection rate versus the number of weak classifiers (2000 FERET samples) . . . . .	126
8.1	The CMDI for handwritten digit recognition . . . . .	130
8.2	Computing 12 $\eta$ s per square area. . . . .	131
8.3	AdaBoost trained with sub-sets of the negative set . . . . .	132
8.4	ROC curves for each classifier group, Ada-88- $\psi$ -N. . . . .	137
8.5	ROC curves for each classifier group, $\eta$ -84-N. . . . .	139
8.6	Number of errors caused by wrong choices between classifiers $\eta$ -84-N. . . . .	140
8.7	Compared misclassifications for the three $\eta$ and $\psi$ feature types. . . . .	141
8.8	Distribution of the feature type usage per class for Ada-88- $\psi$ -N. . . . .	142
8.9	Distribution of the feature type usage per class for $\eta$ -84-N. . . . .	142
8.10	Examples of misclassifications for $\eta$ with 84 features . . . . .	143
8.11	Intersection of two competing classifiers for an unknown character. . . . .	143
8.12	Hypothetical handwritten digits showing a gradual transition between certain digits. . . . .	144
9.1	The proposed hardware platform. . . . .	147
9.2	Percentage of sub-windows per layer. . . . .	148
9.3	Features per layer for $Sub = 1000000$ , $f_0 = 3$ and $C = 10$ . . . . .	149
9.4	% of the features computed up to layer $x$ . . . . .	150

9.5	The detector using multiple cascades in one slave node. . . . .	150
9.6	Time distribution for concurrent cascades running on one processor. . . . .	151
9.7	Frame rate for concurrent classifiers running on one processor. . . . .	151
9.8	Percentage of the total weak classifiers computed up to layer n. . . . .	152
9.9	Frame rate for concurrent classifiers running on one processor. . . . .	152
9.10	Time to broadcast a frame. . . . .	154
9.11	A mobile cluster. . . . .	155
9.12	Parallelisation scenario: runtime and communication times. . . . .	156
9.13	Speedups with different number of cascades and processors . . . . .	158
A.1	SAT for $26.5^\circ$ . . . . .	165

# List of Tables

4.1	The training parameters for the classifiers. . . . .	59
4.2	Testing images: number of occluded frames and non-occluded frames. . .	61
4.3	Hit rates (%) for the sets of images using different classifiers. . . . .	61
5.1	The kernel sizes for each cascade. . . . .	69
5.2	Hit ratios and false positives for the first test set (test set A). . . . .	72
5.3	Comparing the cascades for different angles using a rotated test set (test set B). . . . .	72
6.1	Performance of tilted cascades. . . . .	98
7.1	Runtimes for moment invariants computation. . . . .	110
7.2	Variances for contrast stretching. . . . .	114
7.3	11 moments computed over a square area for image in figure 7.8. . . . .	119
7.4	11 moments computed over an approximation of a circular area. . . . .	120
7.5	Variances for the approximation of concentric disc features from 1 to 0.5 in diameter. . . . .	120
7.6	Variances for the actual concentric disc areas. . . . .	121
8.1	Results reported on the MNIST database. . . . .	129
8.2	Training errors for four feature sets. . . . .	135
8.3	Overall Confusion Matrix for Ada-88- $\psi$ -N. . . . .	136
8.4	Test errors for feature type $\psi_n$ (90 concurrent classifiers) for the 10000 digits MNIST test set. . . . .	136
8.5	Overall Confusion Matrix for $\eta$ with 84 features ( $\eta$ -84-N). . . . .	138
8.6	Test errors for feature type $\eta_{pq}$ (90 concurrent classifiers) for the 10000 digits MNIST test set. . . . .	138
9.1	Results for the broadcasting times and SAT runtimes. . . . .	155
C.1	CH-11- $\psi$ : false detection rate (%) per classifier. . . . .	179
C.2	CH-88- $\psi$ : false detection rate (%) per classifier. . . . .	180

C.3	Overall Confusion Matrix for Ada-11- $\psi$ -50. . . . .	180
C.4	Overall Confusion Matrix for Ada-88- $\psi$ -50. . . . .	180
C.5	$\psi$ -Ada-88-N: hit rate (%) per classifier. . . . .	181
C.6	$\psi$ -Ada-88-N: false detection rate (%). . . . .	181
C.7	$\eta - 84 - N$ : hit rate (%) per classifier. . . . .	182
C.8	$\eta - 84 - N$ : false detection rate (%). . . . .	182
C.9	Number of errors caused by wrong choices between classifiers for $\eta$ feature type, 84 features, N stages . . . . .	183

## List of Notations

$A_{disc}$	Sum of the pixels within a disc .....	116
$A_{square}$	Sum of the pixels within a square or rectangular area ..	115
$\alpha_t$	Factor associated to the error of $h_t$ .....	19
$\beta_t$	Factor associated to the error of $h_t$ .....	20
$C_n$	Positional factor for moment invariants .....	113
$D_t(i)$	Weight of sample $i$ at round $t$ .....	19
$\eta_{pq}$	Normalised central moment .....	104
$\varepsilon$	Efficiency .....	153
$f$	Scaling factor for a kernel .....	25
$f(x)$	number of features computed up to layer $x$ .....	148
$f_s$	Serial fraction of an algorithm .....	31
$F$	False detection rate .....	148
$\gamma$	Exponential factor for normalised central moments .....	104
$H$	Height of an image .....	24
$h_t(x_i)$	Weak classifier function .....	19
$i(x, y)$	Image pixel at $(x, y)$ .....	16
$\bar{i}(x, y)$	Image resulting from a contrast stretching operation .....	26
$I(x, y)$	Summed-area Table element at $(x, y)$ .....	17
$I_r(x, y)$	Rotated SAT element .....	18
$m_{pq}$	2D geometric moment of order $pq$ .....	104
$m_{pq}(x, y)$	SAT element for order $pq$ .....	107
$\bar{m}_{pq}$	Moment invariant with contrast stretching .....	113
$\mu$	Mean (statistics) .....	26
$\mu_{pq}$	Central moment .....	104
$M$	Width of a kernel .....	24
$N$	Height of a kernel .....	24
$pt_n$	Element of a SAT .....	115
$P(A B)$	Conditional probability of A, given B .....	133
$\phi_n$	Hu's moment invariants .....	105
$\psi_n$	Flesser's moment invariants .....	107

$s$	Scale factor .....	16
$Sub$	Number of sub-windows in an image given $s$ and $t$ . ....	147
$S(x)$	Number of sub-windows assessed by $x$ layers of a cascade	148
$\sigma$	Variance (statistics) .....	26
$t$	Translation factor .....	24
$T_m$	Runtime on a multiple processor system .....	31
$T_s$	Runtime on a single processor system .....	31
$\theta$	Direction of an object given by $2_{nd}$ order moments ....	122
$V$	Feature value .....	16
$V_{normal}, V_{tilted}$	Haar-like feature values .....	84
$w_n$	Constants that define a Haar-like feature .....	16
$W$	Width of an image .....	24
$x_i$	Feature array .....	19
$\bar{x}$	Ratio between first $p$ order and zeroth order moments .	104
$y_i$	Class array .....	19
$\bar{y}$	Ratio between first $q$ order and zeroth order moments ..	104

## **Abstract**

This thesis studies rapid object detection, focusing on feature-based methods. Firstly, modifications of training and detection of the Viola-Jones method are made to improve performance and overcome some of the current limitations such as rotation, occlusion and articulation. New classifiers produced by training and by converting existing classifiers are tested in face detection and hand detection.

Secondly, the nature of invariant features in terms of the computational complexity, discrimination power and invariance to rotation and scaling are discussed. A new feature extraction method called Concentric Discs Moment Invariants (CDMI) is developed based on moment invariants and summed-area tables. The dimensionality of this set of features can be increased by using additional concentric discs, rather than using higher order moments. The CDMI set has useful properties, such as speed, rotation invariance, scaling invariance, and rapid contrast stretching can be easily implemented. The results of experiments with face detection shows a clear improvement in accuracy and performance of the CDMI method compared to the standard moment invariants method. Both the CDMI and its variant, using central moments from concentric squares, are used to assess the strength of the method applied to hand-written digits recognition.

Finally, the parallelisation of the detection algorithm is discussed. A new model for the specific case of the Viola-Jones method is proposed and tested experimentally. This model takes advantage of the structure of classifiers and of the multi-resolution approach associated with the detection method. The model shows that high speedups can be achieved by broadcasting frames and carrying out the computation of one or more cascades in each node.





## Acknowledgements

I would like to thank my supervisors Dr. Martin Johnson and Dr. Chris Messom for their support and for providing an interesting and motivating environment for research.

I have also to thank my PhD colleagues who in the last few years shared the challenging activities related to research, in particular Farhad Dadgostar (with whom I share the authorship of two papers), and Chao Fan. Other colleagues in the institute that also have been actively discussing research issues and sharing ideas include Samuel Alexander and Tony Meyer. Thanks to Ravi Chemudugunta (with whom I share the authorship of one paper) for his collaboration in the mini-cluster project and for his UDP protocol for broadcasting frames.

This research would not be possible without the full support of the institute's head of department. I would like to thank Prof. Robert Mckibbin and Massey University for the opportunity to carry out the PhD project while working at IIMS. I also would like to thank the support and encouragement from Dr. Napoleon Reyes, Dr. Houssein Sarrafzadeh, Prof. Ken Hawick, Dr. Heath James, Dr. Chris Scoggings and all my colleagues at IIMS.

Thanks to my wife and my three children. The finalisation of my thesis was only possible with their full support and understanding.