

DYNAMIC VISION

From Images to Face Recognition

Shaogang Gong

Queen Mary and Westfield College

Stephen J McKenna

University of Dundee

Alexandra Psarrou

University of Westminster



Imperial College Press

DYNAMIC VISION

From Images to Face Recognition

Shaogang Gong

Queen Mary and Westfield College

Stephen J. McKenna

University of Dundee

Alexandra Psarrou

University of Westminster



ICP

Imperial College Press

Published by

Imperial College Press
57 Shelton Street
Covent Garden
London WC2H 9HE

Distributed by

World Scientific Publishing Co. Pte. Ltd.
P O Box 128, Farrer Road, Singapore 912805
USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

DYNAMIC VISION

From Images to Face Recognition

Copyright © 2000 by Imperial College Press

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 1-86094-181-8

This book is printed on acid and chlorine free paper.

Printed in Singapore by Fulsland Offset Printing

To my Parents and Aunt Mae

Shaogang Gong

To Collette

Stephen McKenna

To my Parents

Alexandra Psarrou

*As you set out for Ithaka
hope your road is a long one,
full of adventure, full of discovery.*

— *C.P. Cavafy, Ithaka*

Preface

Face recognition is a task that the human vision system seems to perform almost effortlessly, yet the goal of building computer-based systems with comparable capabilities has proven to be difficult. The task implicitly requires the ability to locate and track faces in scenes that are often complex and dynamic. Recognition is difficult because of variations in factors such as lighting conditions, viewpoint, body movement and facial expression. Although evidence from psychophysical and neurobiological experiments provides intriguing insights into how we might code and recognise faces, its bearings on computational and engineering solutions are far from clear. In this book, we describe models and algorithms that are capable of performing face recognition in a dynamic setting. The key question is how to design computer vision and machine learning algorithms that can operate robustly and quickly under poorly controlled and changing conditions.

The study of face recognition has had an almost unique impact on computer vision and machine learning research at large. It raises many challenging issues and provides a good vehicle for examining some difficult problems in vision and learning. Many of the issues raised are relevant to object recognition in general. In particular, face recognition is not merely a problem of pattern recognition of static pictures; it implicitly but crucially invokes many more general computational tasks concerning the perception of moving objects in dynamic and noisy scenes. Consideration of face recognition as a problem in dynamic vision is perhaps both novel and important. The algorithms described in this book have numerous potential applications in areas such as visual surveillance, multimedia and visually mediated interaction.

There have been several books and edited collections about face recognition written over the years, primarily for studies in cognitive psychology or related topics [38, 39, 41, 42, 43, 46, 378]. In more recent years, there has been an explosion of computer vision conferences and special work-

shops dedicated to the recognition of human faces and gestures [162, 163, 164, 165, 166, 167, 168, 365]. Surprisingly, however, there has been no book that provides a coherent and unified treatment of the issue from a computational and systems perspective. We hope that this book succeeds in providing such a treatment of the subject useful for both academic and industrial research communities.

This book has been written with an emphasis on computationally viable approaches that can be readily adopted for the design and development of real-time, integrated machine vision systems for dynamic object recognition. We present what is fundamentally an algorithmic approach, although this is founded upon recent theories of visual perception and learning and has also drawn from psychophysical and neurobiological data.

We address the range of visual tasks needed to perform recognition in dynamic scenes. In particular, visual attention is focused using motion and colour cues. Face recognition is attempted by a set of co-operating processes that perform face detection, tracking and identification using view-based, 2D face models with spatio-temporal context. The models are obtained by learning and are computationally efficient for recognition. We address recognition in realistic and therefore poorly constrained conditions. Computations are essentially based on a statistical decision making framework realised by the implementation of various statistical learning models and neural networks. The systems described are robust to factors such as changing illumination, poor resolution and large head rotations in depth. We also describe how the visual processes can co-operate in an integrated learning system.

Overall, the book explores the use of visual motion detection and estimation, adaptable colour models, active and animate vision principles, statistical learning in high-dimensional feature spaces, vector space dimensionality reduction, temporal prediction models (e.g. Kalman filters, hidden Markov models and the Condensation algorithm), spatio-temporal context, image filtering, linear modelling techniques (e.g. principal components analysis (PCA) and linear discriminants), non-linear models (e.g. mixture models, support vector machines, nonlinear PCA, hybrid neural networks), spatio-temporal models (e.g. recurrent neural networks), perceptual integration, Bayesian inference, on-line learning, view-based representation and databases for learning.

We anticipate that this book will be of special interest to researchers and academics interested in computer vision, visual recognition and machine learning. It should also be of interest to industrial research scientists and managers keen to exploit this emerging technology and develop automated face and human recognition systems for a host of commercial applications including visual surveillance, verification, access control and video-conferencing. Finally, this book should be of use to post-graduate students of computer science, electronic and systems engineering and perhaps also of cognitive psychology.

The topics in this book cover a wide range of multi-disciplinary issues and draw on several fields of study without requiring too deep an understanding of any one area in particular. Nevertheless, some basic knowledge of applied mathematics would be useful for the reader. In particular, it would be convenient if one were familiar with vectors and matrices, eigenvectors and eigenvalues, some linear algebra, multivariate analysis, probability, statistics and elementary calculus at the level of 1st or 2nd year undergraduate mathematics. However, the non-mathematically inclined reader should be able to skip over many of the equations and still understand much of the content.

Shaogang Gong
Stephen McKenna
Alexandra Psarrou

October 1999. London and Dundee

Nomenclature

Vectors are column vectors, i.e. $\mathbf{x}^T \equiv [x_1 \ x_2 \ \dots \ x_N]$. A list is written as (x_1, x_2, \dots, x_N) while a set or sequence is denoted by $\{x_1, x_2, \dots, x_N\}$. Furthermore, $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ denotes a set or sequence of lists. Other commonly used symbols in the book are:

N, n	Input space dimensionality, index
M, m	Number of examples, index
C, c	Number of classes, index
K, k	Number of basis functions or discrete states, index
T, t	Number of time-steps (frames), time variable or index
i, j	Indices
x, \bar{y}	Coordinates in the image plane
s, r, a	Scale, orientation and aspect ratio in the image plane
θ, ϕ	Tilt and yaw (rotation out of the image plane)
κ, ξ	A constant, an error variable
λ	Eigenvalue, wavelength or a hidden Markov model
σ, μ, Σ	Standard deviation, mean vector and covariance matrix
\mathbf{u}, α	Eigenvector, a representation vector
\mathbf{a}	Parameter vector
$\mathbf{x}, \mathbf{q}, \mathbf{y}$	Observation, state and interpretation label vectors
$\tilde{\mathbf{x}}, \mathbf{x}^*$	An approximation to \mathbf{x} , a prediction of \mathbf{x}
γ, ξ	Rotation in depth (both yaw and tilt)
$\mathcal{X}, \mathcal{Q}, \mathcal{Y}$	Sets or sequences of observations, states, interpretations
\mathcal{O}, \mathcal{S}	Object, scene background
$ A , A^T, I$	Determinant and transpose of matrix A , identity matrix
\mathbb{R}^N	N -dimensional space of reals
$f(\cdot), d(\cdot), h(\cdot)$	A function, a distance function and a similarity function
$P(\cdot), E(\cdot), L(\cdot)$	Probability, expectation and likelihood
$p(\cdot)$	Probability density function (PDF)
$G(\cdot), \phi(\cdot)$	Gaussian function (normal distribution), kernel function
$I(\cdot)$	Intensity function (monochrome image)
$\mathcal{E}(\cdot), \mathcal{L}(\cdot), \mathcal{R}(\cdot)$	Error function, loss function and risk functional
$\ \mathbf{x}\ $	L_2 norm of \mathbf{x} (Euclidean length)
\ln	Logarithm to base e
\otimes	Convolution

Acknowledgements

We should like to express our considerable gratitude to the many fine people who have helped us in the process of writing this book. The experiments described herein would not have been possible without the work of PhD students, research assistants and post-docs at QMW and the University of Westminster. In particular, we want to thank J.J. Collins, Yongmin Li, Peter Loft, Jeffrey Ng, Eng-Jon Ong, Yogesh Raja, Sami Romdhani, Jamie Sherrah, Jonathan Tanner, Fernando de la Torre, Paul Verity and Michael Walter for their contributions to this work. We are indebted to Hilary Buxton, Robert Koger and Dennis Parkinson, all of whom read a draft carefully and gave us many insightful comments and suggestions.

We should also like to thank the anonymous reviewers for their constructive feedback as well as John Navas at Imperial College Press and R. Sankaran at World Scientific Publishing for the kind help we have received in the final stages of preparing this book. The book was typeset using L^AT_EX.

Our thanks are also due to Wolfgang Konen and Elke Sassmannshausen at ZN GmbH in Bochum, Johannes Steffens at Eyematic Interfaces Inc., James Kottas at Miros Inc. and Martin Levine at VisionSphere Technologies Inc. for kindly responding to our enquiries concerning their companies' products which are summarised in Appendix B.

We gratefully acknowledge the financial support that we have received over the years from UK EPSRC, EU ESPRIT and HCM networks, the BBC, BT and Safehouse Technologies Inc. Finally, we should like to thank our families and friends for all their support.

*Perchance the best chance of reproducing the ancient
Greek temperament would be to cross the Scots with the
Chinese.*

— *Murray Christopher Grieve (Hugh McDiarmid), Lucky Poet*

Contents

Preface	xv
PART I BACKGROUND	1
1 About Face	3
1.1 The Visual Face	3
1.2 The Changing Face	4
1.3 Computing Faces	6
1.4 Biological Perspectives	11
1.5 The Approach	12
2 Perception and Representation	15
2.1 A Distal Object	16
2.2 Representation by 3D Reconstruction	18
2.3 Two-dimensional View-based Representation	19
2.4 Image Template-based Representation	21
2.5 The Correspondence Problem and Alignment	22
2.6 Biological Perspectives	26
2.7 Discussion	29
3 Learning under Uncertainty	31
3.1 Statistical Learning	32
3.2 Learning as Function Approximation	33
3.3 Bayesian Inference and MAP Classification	36
3.4 Learning as Density Estimation	37
3.4.1 Parametric Models	37
3.4.2 Non-parametric Models	39
3.4.3 Semi-parametric Models	40
3.5 Unsupervised Learning without Density Estimation	41

3.5.1	Dimensionality Reduction	43
3.5.2	Clustering	43
3.6	Linear Classification and Regression	44
3.6.1	Least-squares	45
3.6.2	Linear Support Vector Machines	45
3.7	Non-linear Classification and Regression	48
3.7.1	Multi-layer Networks	48
3.7.2	Support Vector Machines	50
3.8	Adaptation	51
3.9	Biological Perspectives	53
3.10	Discussion	54

PART II FROM SENSORY TO MEANINGFUL PERCEPTION 57

4	Selective Attention: Where to Look	59
4.1	Pre-attentive Visual Cues from Motion	60
4.1.1	Measuring Temporal Change	61
4.1.2	Motion Estimation	62
4.2	Learning Object-based Colour Cues	65
4.2.1	Colour Spaces	66
4.2.2	Colour Density Models	68
4.3	Perceptual Grouping for Selective Attention	71
4.4	Data Fusion for Perceptual Grouping	73
4.5	Temporal Matching and Tracking	76
4.6	Biological Perspectives	77
4.7	Discussion	78
5	A Face Model: What to Look For	81
5.1	Person-independent Face Models for Detection	82
5.1.1	Feature-based Models	82
5.1.2	Holistic Models	83
5.1.3	The Face Class	85
5.2	Modelling the Face Class	86
5.2.1	Principal Components Analysis for a Face Model	87
5.2.2	Density Estimation in Local PCA Spaces	88
5.3	Modelling a Near-face Class	89
5.4	Learning a Decision Boundary	90

5.4.1	Face Detection in Dynamic Scenes	91
5.4.2	Normalisation	93
5.4.3	Face Detection using Multi-layer Perceptrons	95
5.4.4	Face Detection using Support Vector Machines	97
5.5	Perceptual Search	99
5.6	Biological Perspectives	101
5.7	Discussion	102
6	Understanding Pose	103
6.1	Feature and Template-based Correspondence	105
6.2	The Face Space across Views: Pose Manifolds	105
6.3	The Effect of Gabor Wavelet Filters on Pose Manifolds	111
6.4	Template Matching as Affine Transformation	113
6.5	Similarities to Prototypes across Views	118
6.6	Learning View-based Support Vector Machines	121
6.7	Biological Perspectives	123
6.8	Discussion	123
7	Prediction and Adaptation	125
7.1	Temporal Observations	128
7.2	Propagating First-order Markov Processes	129
7.3	Kalman Filters	131
7.4	Propagating Non-Gaussian Conditional Densities	132
	7.4.1 Learning Priors using HMMs and EM	132
	7.4.2 Observation Augmented Density Propagation	133
7.5	Tracking Attended Regions	134
7.6	Adaptive Colour Models	136
7.7	Selective Adaptation	140
7.8	Tracking Faces	145
7.9	Pose Tracking	147
	7.9.1 Person-specific Pose Tracking	149
	7.9.2 Person-independent Pose Tracking	150
7.10	Biological Perspectives	156
7.11	Discussion	157

PART III	MODELS OF IDENTITY	16
8	Single-View Identification	16
8.1	Identification Tasks	16
8.2	Nearest-neighbour Template Matching	16
8.3	Representing Knowledge of Facial Appearance	16
8.4	Statistical Knowledge of Facial Appearance	16
8.4.1	Low Dimensionality: Principal Components Analysis	16
8.4.2	Separability: Linear Discriminant Analysis	17
8.4.3	Sparseness and Topography: Local Feature Analysis	17
8.5	Statistical Knowledge of Identity	17
8.5.1	Identification Tasks Revisited	17
8.5.2	Class-conditional Densities for Modelling Identity	17
8.6	Structural Knowledge: The Role of Correspondence	18
8.6.1	Beyond Alignment: Correspondence at a Single View	18
8.6.2	Combining Statistical and Structural Models	18
8.7	Biological Perspectives	18
8.8	Discussion	18
9	Multi-View Identification	18
9.1	View-based Models	18
9.2	The Role of Prior Knowledge	19
9.3	View Correspondence in Identification	19
9.3.1	Learning Linear Shape Models	19
9.3.2	Nonlinear Shape Models	19
9.4	Generalisation from a Single View	19
9.4.1	Identity by Linear Combination	19
9.4.2	Identity by Similarity to Prototype Views	20
9.5	Generalisation from Multiple Views	20
9.6	Biological Perspectives	20
9.7	Discussion	20
10	Identifying Moving Faces	20
10.1	Biological Perspectives	21
10.1.1	Learning and Identity Constancy	21
10.1.2	The Effect of Temporal Order of Pose on Learning	21
10.1.3	The Effect of Motion on Familiar Face Identification	21
10.2	Computational Theories of Temporal Identification	21

10.2.1	Atemporal Representation	213
10.2.2	Spatio-temporal Signatures	214
10.3	Identification using Holistic Temporal Trajectories	215
10.4	Identification by Continuous View Transformation	218
10.5	An Experimental System	219
10.6	Discussion	225
PART IV	PERCEPTION IN CONTEXT	227
11	Perceptual Integration	229
11.1	Sensory and Model-based Vision	230
11.2	Perceptual Fusion	231
11.3	Perceptual Inference	236
11.3.1	Inference using Hidden Markov Models	240
11.3.2	Closed-loop Perceptual Control	241
11.4	Vision as Cooperating Processes	242
11.4.1	Visual Attention and Grouping	242
11.4.2	Face Detection, Tracking and Identification	245
11.5	Biological Perspectives	247
11.6	Discussion	250
12	Beyond Faces	253
12.1	Multi-modal Identification	254
12.2	Visually Mediated Interaction	255
12.3	Visual Surveillance and Monitoring	258
12.4	Immersive Virtual Reality	260
12.5	Visual Database Screening	261
PART V	APPENDICES	265
A	Databases	267
A.1	Database Acquisition and Design	267
A.1.1	Intrinsic Experimental Variables	268
A.1.2	Extrinsic Experimental Variables	268
A.2	Acquisition of a Pose-labelled Database	269
A.2.1	Using Artificial Markers	269
A.2.2	Using Sensors and a Calibrated Camera	271

A.3	Benchmarking	273
A.4	Commercial Databases	275
A.5	Public Domain Face Databases	277
A.6	Discussion	279
B	Commercial Systems	281
B.1	System Characterisation	282
B.2	A View on the Industry	283
B.2.1	Visionics Corporation	284
B.2.2	Miros Inc.	286
B.2.3	VisionSpheres Technologies	287
B.2.4	Eigenface-based Systems: Viisage Technologies, Intelli- gent Verification Systems and Facia Reco	289
B.2.5	Systems based on Facial Feature Matching: Plettac Elec- tronic Security GmbH, ZN Bochum GmbH and Eyematic Interfaces Inc.	291
B.3	Discussion	295
C	Mathematical Details	297
C.1	Principal Components Analysis	297
C.2	Linear Discriminant Analysis	298
C.3	Gaussian Mixture Estimation	300
C.3.1	Expectation-maximisation	300
C.3.2	Automatic Model Order Selection	301
C.3.3	Adaptive EM for Non-stationary Distributions	302
C.4	Kalman Filters	304
C.4.1	Zero-order Prediction	304
C.4.2	First-order Prediction	305
C.5	Bayesian Belief Networks	306
C.6	Hidden Markov Models	309
C.7	Gabor Wavelets	312
	Bibliography	315
	Index	339