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

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

ISBN 1-86094-181-8

This book is printed on acid and chlorine free paper.

Printed in Singapore by FuIsland 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 workshops 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. $\mathbf{x}^{\mathrm{T}} \equiv [x_1\, x_2\, \ldots\, x_N]$. A list is written as $(x_1, x_2, \ldots x_N)$ while a set or sequence is denoted by $\{x_1, x_2, \ldots x_N\}$. Furthermore, $\{(x_1, y_1), (x_2, y_2), \ldots (x_N, y_N)\}$ 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, 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
$\sigma, \boldsymbol{\mu}, \boldsymbol{\Sigma}$	Standard deviation, mean vector and covariance matrix
$\mathbf{u}, \boldsymbol{\alpha}$	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 $\mathbf{x}$
$\gamma, \boldsymbol{\xi}$	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
$\lvert \boldsymbol{A} \rvert, \boldsymbol{A}^{\mathrm{T}}, \boldsymbol{I}$	Determinant and transpose of matrix $\boldsymbol{A}$, identity matrix
$\Re^N$	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
$\lvert\lvert \mathbf{x} \rvert\rvert$	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 LaTeX.

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

PART III MODELS OF IDENTITY 16