Preface

The human visual system is able to visually recognise and interpret object behaviours under different conditions. Yet, the goal of building computer vision based recognition systems with comparable capabilities has proven to be very difficult to achieve. Computational modelling and analysis of object behaviours through visual observation is inherently ill-posed. Many would argue that our cognitive understanding remains unclear about why we associate certain semantic meanings with specific object behaviours and activities. This is because meaningful interpretation of a behaviour is subject to the observer's a priori knowledge, which is at times rather ambiguous. The same behaviour may have different semantic meanings depending upon the context within which it is observed. This ambiguity is exacerbated when many objects are present in a scene. Can a computer based model be constructed that is able to extract all necessary information for describing a behaviour from visual observation alone? Do people behave differently in the presence of others, and if so, how can a model be built to differentiate the expected normal behaviours from those of abnormality? Actions and activities associated with the same behavioural interpretation may be performed differently according to the intended meaning, and different behaviours may be acted in a subtly similar way. The question arises as to whether these differences can be accurately measured visually and robustly computed consistently for meaningful interpretation of behaviour.

Visual analysis of behaviour requires not only to solve the problems of object detection, segmentation, tracking, motion trajectory analysis, but also the modelling of context information and utilisation of non-sensory knowledge when available, such as human annotation of input data or relevance feedback to output signals. Visual analysis of behaviour faces two fundamental challenges in computational complexity and uncertainty. Object behaviours in general exhibit complex spatio-temporal dynamics in a highly dynamical and uncertain environment, for instance, human activities in a crowded public space. Segmenting and modelling human actions and activities in a visual environment is inherently ill-posed, as information processing in visual analysis of behaviour is subject to noise, incompleteness and uncertainty in sensory data. Whilst these visual phenomena are difficult to model analytically, they can be probabilistically modelled much more effectively through statistical machine learning.

Despite these difficulties, it is compelling that one of the most significant developments in computer vision research over the last 20 years has been the rapidly growing interest in automatic visual analysis of behaviour in video data captured from closed-circuit television (CCTV) systems installed in private and public spaces. The study of visual analysis of behaviour has had an almost unique impact on computer vision and machine learning research at large. It raises many challenges and provides a testing platform for examining some difficult problems in computational modelling and algorithm design. Many of the issues raised are relevant to dynamic scene understanding in general, multivariate time series analysis and statistical learning in particular.

Much progress has been made since the early 1990s. Most noticeably, statistical machine learning has become central to computer vision in general, and to visual analysis of behaviour in particular. This is strongly reflected throughout this book as one of the underlying themes. In this book, we study plausible computational models and tractable algorithms that are capable of automatic visual analysis of behaviour in complex and uncertain visual environments, ranging from well-controlled private spaces to highly crowded public scenes. The book aims to reflect the current trends, progress and challenges on visual analysis of behaviour. We hope this book will not only serve as a sampling of recent progress but also highlight some of the challenges and open questions in automatic visual analysis of object behaviour.

There is a growing demand by both governments and commerce worldwide for advanced imaging and computer vision technologies capable of automatically selecting and identifying behaviours of objects in imagery data captured in both public and private spaces for crime prevention and detection, public transport management, personalised healthcare, information management and market studies, asset and facility management. A key question we ask throughout this book is how to design automatic visual learning systems and devices capable of extracting and mining salient information from vast quantity of data. The algorithm design characteristics of such systems aim to provide, with minimum human intervention, machine capabilities for extracting relevant and meaningful semantic descriptions of salient objects and their behaviours for aiding decision-making and situation assessment.

There have been several books on human modelling and visual surveillance over the years, including *Face Detection and Gesture Recognition for Human-Computer Interaction* by Yang and Ahuja (2001); *Analyzing Video Sequences of Multiple Humans* by Ohya, Utsumi and Yamato (2002); *A Unified Framework for Video Summarization, Browsing and Retrieval: with Applications to Consumer and Surveillance Video* by Xiong, Radhakrishnan, Divakaran, Rui and Huang (2005); *Human Identification based on Gait* by Nixon, Tan and Chellapa (2005); and *Automated Multi-Camera Surveillance* by Javed and Shah (2008). There are also a number of books and edited collections on behaviour studies from cognitive, social and psychological perspectives, including *Analysis of Visual Behaviour* edited by Ingle, Goodale and Mansfield (1982); *Hand and Mind: What Gestures Reveal about Thought* by McNeill (1992); *Measuring Behaviour* by Martin (1993); *Understanding Human Behaviour* by Mynatt and Doherty (2001); and *Understanding Human Behaviour and the Social Environment* by Zastrow and Kirst-Ashman (2003). However, there has

been no book that provides a comprehensive and unified treatment of visual analysis of behaviour from a computational modelling and algorithm design perspective.

This book has been written with an emphasis on computationally viable approaches that can be readily adopted for the design and development of intelligent computer vision systems for automatic visual analysis of behaviour. We present what is fundamentally a computational algorithmic approach, founded on recent advances in visual representation and statistical machine learning theories. This approach should also be attractive to researchers and system developers who would like to both learn established techniques for visual analysis of object behaviour, and gain insight into up-to-date research focus and directions for the coming years. We hope that this book succeeds in providing such a treatment of the subject useful not only for the academic research communities, both also the commerce and industry.

Overall, the book addresses a broad range of behaviour modelling problems, from established areas of human facial expression, body gesture and action analysis to emerging new research topics in learning group activity models, unsupervised behaviour profiling, hierarchical behaviour discovery, learning behavioural context, modelling rare behaviours, 'man-in-the-loop' active learning of behaviours, multi-camera behaviour correlation, person re-identification, and 'connecting-the-dots' for global abnormal behaviour detection. The book also gives in depth treatment to some popular computer vision and statistical machine learning techniques, including the Bayesian information criterion, Bayesian networks, 'bag-of-words' representation, canonical correlation analysis, dynamic Bayesian networks, Gaussian mixtures, Gibbs sampling, hidden conditional random fields, hidden Markov models, human silhouette shapes, latent Dirichlet allocation, local binary patterns, locality preserving projection, Markov processes, probabilistic graphical models, probabilistic topic models, space-time interest points, spectral clustering, and support vector machines.

The computational framework presented in this book can also be applied to modelling behaviours exhibited by many other types of spatio-temporal dynamical systems, either in isolation or in interaction, and therefore can be beneficial to a wider range of fields of studies, including Internet network behaviour analysis and profiling, banking behaviour profiling, financial market analysis and forecasting, bioinformatics, and human cognitive behaviour studies.

We anticipate that this book will be of special interest to researchers and academics interested in computer vision, video analysis and machine learning. It should be of interest to industrial research scientists and commercial developers keen to exploit this emerging technology for commercial applications including visual surveillance for security and safety, information and asset management, public transport and traffic management, personalised healthcare in assisting elderly and disabled, video indexing and search, human computer interaction, robotics, animation and computer games. This book should also be of use to post-graduate students of computer science, mathematics, engineering, physics, behavioural science, and cognitive psychology. Finally, it may provide government policy makers and commercial managers an informed guide on the potentials and limitations in deploying intelligent video analytics systems.

The topics in this book cover a wide range of computational modelling and algorithm design issues. Some knowledge of mathematics would be useful for the reader. In particular, it would be convenient if one were familiar with vectors and matrices, eigenvectors and eigenvalues, linear algebra, optimisation, multivariate analysis, probability, statistics and calculus at the level of post-graduate mathematics. However, the non-mathematically inclined reader should be able to skip over many of the equations and still understand much of the content.

London, UK

Shaogang Gong
Tao Xiang



http://www.springer.com/978-0-85729-669-6

Visual Analysis of Behaviour
From Pixels to Semantics
Gong, S.; Xiang, T.
2011, XIX, 356 p., Hardcover
ISBN: 978-0-85729-669-6