

Shaogang Gong and Tao Xiang

VISUAL ANALYSIS OF BEHAVIOUR

From Pixels to Semantics

February 2011

Springer

To Aleka, Philip and Alexander

Shaogang Gong

To Ning and Rachel

Tao Xiang

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 the 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 Be-*

haviour 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 the 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 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.

Shaogang Gong
Tao Xiang

London
February 2011

Acknowledgements

We shall express our deep gratitude to the many 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 and postdoctoral research assistants at Queen Mary University of London. In particular, we want to thank Chen Change Loy, Tim Hospedales, Jian Li, Wei-Shi Zheng, Jianguo Zhang, Caifeng Shan, Yogesh Raja, Bryan Prosser, Matteo Bregonzio, Parthipan Siva, Lukasz Zalewski, Eng-Jon Ong, Jamie Sherrah, Jeffrey Ng, and Michael Walter for their contributions to this work. We are indebted to Alexandra Psarrou who read a draft carefully and gave us many helpful comments and suggestions.

We shall thank Simon Rees and Wayne Wheeler at Springer for their kind help and patience during the preparation of this book. The book was typeset using \LaTeX .

We gratefully acknowledge the financial support that we have received over the years from UK EPSRC, UK DSTL, UK MOD, UK Home Office, UK TSB, US Army Labs, EU FP7, the Royal Society, BAA, and QinetiQ. Finally, we shall thank our families and friends for all their support.

Contents

Part I INTRODUCTION

1	About Behaviour	3
1.1	Understanding Behaviour	4
1.1.1	Representation and Modelling	5
1.1.2	Detection and Classification	6
1.1.3	Prediction and Association	6
1.2	Opportunities	7
1.2.1	Visual Surveillance	8
1.2.2	Video Indexing and Search	8
1.2.3	Robotics and Healthcare	9
1.2.4	Interaction, Animation and Computer Games	9
1.3	Challenges	10
1.3.1	Complexity	10
1.3.2	Uncertainty	10
1.4	The Approach	11
	References	13
2	Behaviour in Context	15
2.1	Facial Expression	15
2.2	Body Gesture	17
2.3	Human Action	19
2.4	Human Intent	21
2.5	Group Activity	23
2.6	Crowd Behaviour	24
2.7	Distributed Behaviour	26
2.8	Holistic Awareness: Connecting the Dots	29
	References	31

3	Towards Modelling Behaviour	43
3.1	Behaviour Representation	43
3.1.1	Object-based Representation	43
3.1.2	Part-based Representation	47
3.1.3	Pixel-based Representation	48
3.1.4	Event-based Representation	50
3.2	Probabilistic Graphical Models	51
3.2.1	Static Bayesian Networks	54
3.2.2	Dynamic Bayesian Networks	55
3.2.3	Probabilistic Topic Models	56
3.3	Learning Strategies	57
3.3.1	Supervised Learning	58
3.3.2	Unsupervised Learning	58
3.3.3	Semi-Supervised Learning	60
3.3.4	Weakly-Supervised Learning	61
3.3.5	Active Learning	61
	References	63
Part II SINGLE-OBJECT BEHAVIOUR		
4	Understanding Facial Expression	75
4.1	Classification of Images	75
4.1.1	Local Binary Patterns	76
4.1.2	Designing Classifiers	79
4.1.3	Feature Selection by Boosting	83
4.2	Manifold and Temporal Modelling	84
4.2.1	Locality Preserving Projections	84
4.2.2	Bayesian Temporal Models	92
4.3	Discussion	96
	References	97
5	Modelling Gesture	101
5.1	Tracking Gesture	102
5.1.1	Motion Moment Trajectory	102
5.1.2	2D Colour-based Tracking	103
5.1.3	Bayesian Association	106
5.1.4	3D Model-based Tracking	116
5.2	Segmentation and Atomic Action	122
5.2.1	Temporal Segmentation	124
5.2.2	Atomic Actions	125
5.3	Markov Processes	127
5.4	Affective State Analysis	131
5.4.1	Space-Time Interest Points	132
5.4.2	Expression and Gesture Correlation	133

5.5	Discussion	135
References		137
6	Action Recognition	141
6.1	Human Silhouette	142
6.2	Hidden Conditional Random Fields	143
6.2.1	HCRF Potential Function	146
6.2.2	Observable HCRF	146
6.3	Space-Time Clouds	149
6.3.1	Clouds of Space-Time Interest Points	150
6.3.2	Joint Local and Global Feature Representation	157
6.4	Localisation and Detection	158
6.4.1	Tracking Salient Points	161
6.4.2	Automated Annotation	162
6.5	Discussion	166
References		167
Part III GROUP BEHAVIOUR		
7	Supervised Learning of Group Activity	173
7.1	Contextual Events	174
7.1.1	Seeding Event: Measuring Pixel-Change-History	174
7.1.2	Classification of Contextual Events	177
7.2	Activity Segmentation	180
7.2.1	Semantic Content Extraction	181
7.2.2	Semantic Video Segmentation	183
7.3	Dynamic Bayesian Networks	191
7.3.1	Correlations of Temporal Processes	191
7.3.2	Behavioural Interpretation of Activities	196
7.4	Discussion	200
References		202
8	Unsupervised Behaviour Profiling	205
8.1	Off-Line Behaviour Profile Discovery	206
8.1.1	Behaviour Pattern	206
8.1.2	Behaviour Profiling by Data Mining	207
8.1.3	Behaviour Affinity Matrix	208
8.1.4	Eigendecomposition	209
8.1.5	Model Order Selection	209
8.1.6	Quantifying Eigenvector Relevance	210
8.2	On-Line Anomaly Detection	213
8.2.1	A Composite Behaviour Model	213
8.2.2	Run-Time Anomaly Measure	216

8.2.3	On-Line Likelihood Ratio Test	216
8.3	On-Line Incremental Behaviour Modelling	218
8.3.1	Model Bootstrapping	219
8.3.2	Incremental Parameter Update	220
8.3.3	Model Structure Adaptation	223
8.4	Discussion	224
References		226
9	Hierarchical Behaviour Discovery	229
9.1	Local Motion Events	230
9.2	Markov Clustering Topic Model	231
9.2.1	Off-Line Model Learning by Gibbs Sampling	234
9.2.2	On-Line Video Saliency Inference	236
9.3	On-Line Video Screening	237
9.4	Model Complexity Control	240
9.5	Semi-Supervised Learning of Behavioural Saliency	242
9.6	Discussion	243
References		246
10	Learning Behavioural Context	247
10.1	Spatial Context	248
10.1.1	Behaviour-Footprint	250
10.1.2	Semantic Scene Decomposition	250
10.2	Correlational and Temporal Context	252
10.2.1	Learning Regional Context	253
10.2.2	Learning Global Context	256
10.3	Context-Aware Anomaly Detection	258
10.4	Discussion	261
References		263
11	Modelling Rare and Subtle Behaviours	265
11.1	Weakly-Supervised Joint Topic Model	267
11.1.1	Model Structure	267
11.1.2	Model Parameters	270
11.2	On-Line Behaviour Classification	275
11.3	Localisation of Rare Behaviour	278
11.4	Discussion	279
References		281

12 Man in the Loop	283
12.1 Active Behaviour Learning Strategy	285
12.2 Local Block-based Behaviour	287
12.3 Bayesian Classification	289
12.4 Query Criteria	291
12.4.1 Likelihood Criterion	291
12.4.2 Uncertainty Criterion	292
12.5 Adaptive Query Selection	294
12.6 Discussion	297
References	299
Part IV DISTRIBUTED BEHAVIOUR	
13 Multi-Camera Behaviour Correlation	303
13.1 Multi-View Activity Representation	306
13.1.1 Local Bivariate Time-Series Events	306
13.1.2 Activity-based Scene Decomposition	307
13.2 Learning Pair-Wise Correlation	310
13.2.1 Cross Canonical Correlation Analysis	311
13.2.2 Time-Delayed Mutual Information Analysis	313
13.3 Multi-Camera Topology Inference	314
13.4 Discussion	316
References	317
14 Person Re-Identification	319
14.1 Re-Identification by Ranking	321
14.1.1 Support Vector Ranking	321
14.1.2 Scalability and Complexity	323
14.1.3 Ensemble RankSVM	324
14.2 Context-Aware Search	326
14.3 Discussion	328
References	331
15 Connecting the Dots	333
15.1 Global Behaviour Segmentation	333
15.2 Bayesian Behaviour Graphs	337
15.2.1 A Time-Delayed Probabilistic Graphical Model	337
15.2.2 Bayesian Graph Structure Learning	339
15.2.3 Bayesian Graph Parameter Learning	344
15.2.4 Cumulative Anomaly Score	345
15.2.5 Incremental Model Structure Learning	348
15.3 Global Awareness	353
15.3.1 Time-Ordered Latent Dirichlet Allocation	353

15.3.2 On-Line Prediction and Anomaly Detection	355
15.4 Discussion	358
References	360
Epilogue	363
Index	365

Acronyms

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
BIC	Bayesian information criterion
CCA	canonical correlation analysis
CCTV	closed-circuit television
CONDENSATION	conditional density propagation
CRF	conditional random field
EM	expectation-maximisation
DBN	dynamic Bayesian network
HCI	human computer interaction
HCRF	hidden conditional random field
HOG	histogram of oriented gradients
FACS	facial action coding system
FOV	field of view
FPS	frame per second
HMM	hidden Markov model
KL	Kullback-Leibler
LBP	local binary patterns
LDA	latent Dirichlet allocation
LPP	locality preserving projection
MAP	maximum a posteriori
MCMC	Markov chain Monte Carlo
MLE	maximum likelihood estimation
MRF	Markov random field
PCA	principal component analysis
PGM	probabilistic graphical model
PTM	probabilistic topic model
PTZ	pan-tilt-zoom
SIFT	scale-invariant feature transform
SLPP	supervised locality preserving projection

SVM	support vector machine
xCCA	cross canonical correlation analysis

Part I
INTRODUCTION

Chapter 1

About Behaviour

Understanding and interpreting behaviours of objects, and in particular those of humans, is central to social interaction and communication. Commonly, one considers that behaviours are the actions and reactions of a person or animal in response to external or internal stimuli. There is, however, a plethora of wider considerations of what behaviour is, ranging from economical (Simon, 1955), organisational (Rollinson, 2004), social (Sherman and Sherman, 1930), to sensory attentional such as *visual behaviour* (Ingle et al., 1982). Visual behaviour refers to the actions or reactions of a sensory mechanism in response to a visual stimulus, for example, the navigation mechanism of nocturnal bees in dim light (Warrant, 2008), visual search by eye movement of infants (Gough, 1962) or drivers in response to their surrounding environment (Harbluk and Noy, 2002). If visual behaviour as a search mechanism is a perceptual function that scans actively a visual environment in order to focus attention and seek an object of interest among distracters (Ltti et al., 1998), *visual analysis of behaviour* is a perceptual task that interprets actions and reactions of objects, such as people, interacting or co-existing with other objects in a visual environment (Buxton and Gong, 1995; Gong et al., 2002; Xiang and Gong, 2006). The study of visual analysis of behaviour, and in particularly of human behaviour, is the focus of this book.

Recognising objects visually by behaviour and activity rather than shape and size plays an important role in a primate visual system (Barbur et al., 1980; Schiller and Koerner, 1971; Weiskrantz, 1972). In a visual environment of multiple objects co-existing and interacting, it becomes necessary to identify objects not only by their appearance but also by what they do. The latter provides richer information about objects especially when visual data is spatially ambiguous and incomplete. For instance, some animals, such as most snakes, have a very poor visual sensing system, which is unable to capture sufficient visual appearance of objects but very sensitive to movements for detecting preys and predators. The human visual system is highly efficient for scanning through large quantity of low-level imagery data and selecting salient information for a high-level semantic interpretation and gaining situational awareness.

1.1 Understanding Behaviour

Since 1970s, the computer vision community has endeavoured to bring about intelligent perceptual capabilities to artificial visual sensors. Computer vision aims to build artificial mechanisms and devices capable of mimicking the sensing capabilities of biological vision systems (Marr, 1982). This endeavour is intensified in recent years by the need for understanding massive quantity of video data, with the aim to not only comprehend objects spatially in a snapshot but also their spatio-temporal relations over time in a sequence of images. For understanding a dynamically changing social environment, a computer vision system can be designed to interpret behaviours from object actions and interactions captured visually in that environment. A significant driver for visual analysis of behaviour is automated visual surveillance, which aims to automatically interpret human activities and detect unusual events that could pose a threat to public security and safety.

If a behaviour is considered the way how an object acts, often in relation to other objects in the same visual environment, the focus of this book is on visual analysis of human behaviour and behaviours of object that are manipulated by humans, for example, vehicles driven by people. There are many interchangeable terms used in the literature concerning behaviour, including activities, actions, events, and movements. They correspond to different spatial and temporal context within which a behaviour can be defined. One may consider a behaviour hierarchy of three layers:

1. Atomic actions correspond to instantaneous atomic entities upon which an action is formed. For example, in a running action, the atomic action could be 'left leg moving in front of the right leg'. In a returning action in tennis, it could be 'swing right hand' followed by 'rotating the upper body'.
2. Actions correspond to a sequence of atomic actions that fulfil a function or purpose. For instance, walking, running, or serving a tennis ball.
3. Activities are composed of sequences of actions over space and time. For example, 'a person walking from a living room to a kitchen to fetch a cup of water', or 'two people playing tennis'. Whilst actions are likely associated with a single object in isolation, activities are almost inevitably concerned with either interactions between objects, or an object engaging with the surrounding environment.

In general, visual analysis of behaviour is about constructing models and developing devices for automatic analysis and interpretation of object actions and activities captured in a visual environment. To that end, visual analysis of behaviour focuses on three essential functions:

1. Representation and modelling: To extract and encode visual information from imagery data in a more concise form that also captures intrinsic characteristics of objects of interest;
2. Detection and classification: To discover and search for salient, perhaps also unique, characteristics of certain object behaviour patterns from large quantity of visual observations, and to discriminate them against known categories of semantic and meaningful interpretation;

3. Prediction and association: To forecast future events based on the past and current interpretation of behaviour patterns, and to forge object identification through behavioural expectation and trend.

We consider that automated visual analysis of behaviour is information processing of visual data, capable of not only modelling previously observed object behaviours, but also detecting, recognising and predicting unseen behavioural patterns and associations.

1.1.1 Representation and Modelling

A human observer can recognise behaviours of interest directly from visual observation. This suggests that imagery data embed useful information for semantic interpretation of object behaviour. Behaviour representation addresses the question of what information must be extracted from images and in what form, so that object behaviour can be understood and recognised visually. The human visual system utilises various visual cues and contextual information for recognising objects and their behaviour (Humphreys and Bruce, 1989). For instance, the specific stripe pattern and its colour is a useful cue for human to recognise a tiger and distinguish it from other cats such as lions. Similarly, the movement as well as posture of a tiger can reveal its intended action: running, walking, or about to strike. It is clear that different sources and types of visual information need be utilised for modelling and understanding object behaviour.

A behaviour representation needs to accommodate both cumulative and temporal information about an object. In order to recognise an object and its behaviour, the human visual system relates any visual stimuli falling on to the retina to a set of knowledge and expectation about the object under observation: how it *should* look like and how it is *supposed* to behave (Gregory, 1970; Helmholtz, 1962). Behaviour representation should address the need for extracting visual information that can facilitate the association of visual observation with semantic interpretation. In other words, representation of visual data is part of a computational mechanism that contributes towards constructing and accumulating knowledge about behaviour. For example, modelling the action of a person walking can be considered as to learn the prototypical and generic knowledge of walking based on limited observations of walking examples, so that when an unseen instance of walking is observed, it can be recognised by utilising the accumulated *a priori* knowledge. An important difference between object modelling and behaviour modelling is that a behaviour model should benefit more from capturing temporal information about behaviour. Object recognition in large only considers spatial information. For instance, a behaviour model is built based on visual observation of a person's daily routine in an office which consists of meetings, tea breaks, paper works and a lunch at certain times of every day. What has been done so far can then have a significant influence on the correct interpretation of what this person is about to do (Agre, 1989).

A computational model of behaviour performs both representation and matching. For representing object behaviour, one considers a model capable of capturing distinctive characteristics of an object in action and activity. A good behaviour representation aims to describe an object sufficiently well for both generalisation and discrimination in model matching. Model matching is a computational process to either explain away new instances of observation against known object behaviours, considered as its generalisation capacity, or discriminate one type of object behaviour from the others, regarded as its discrimination ability. For effective model matching, a representation needs to separate visual observation of different object behaviour types or classes in a representational space, and to maintain such separations given noisy and incomplete visual observations.

1.1.2 Detection and Classification

Generally speaking, visual classification is a process of categorising selected visual observations of interest into known classes. Classification is based on an assumption that segmentation and selection of interesting observations have already been taken place. On the other hand, visual detection aims to discover and locate patterns of interest, regardless of class interpretation, from a vast quantify of visual observations. For instance, for action recognition, a model is required to detect and segment instances of actions from a continuous observation of a visual scene. Detection in crowded scenes, such as detecting people fighting or falling in crowd, becomes challenging as objects of interest can be swamped by distracters and background clutters. To spot and recognise actions from a sea of background activities, the task of detection often poses a greater challenge than classification.

The problem of behaviour detection is further compounded when the behaviour to be detected is unknown *a priori*. A common aim of visual analysis of behaviour is to learn a model that is capable of detecting unseen abnormal behaviour patterns whilst recognising novel instances of known normal behaviour patterns. To that end, an anomaly is defined as an atypical and un-random behaviour pattern not represented by sufficient observations. However, in order to differentiate anomaly from trivial unseen instances or outright statistical outliers, one should consider that an anomaly satisfies a specificity constraint to known normal behaviours, i.e. true anomalies lie in the vicinity of known normal behaviours without being recognised as any.

1.1.3 Prediction and Association

An activity is usually formed by a series of object actions executed following certain temporal order at certain durations. Moreover, the ordering and durations of constituent actions can be highly variable and complex. To model such visual observa-

tions, a behaviour can be considered as a temporal process, or a time series function. An important feature of a model of temporal processes is to make prediction. To that end, behaviour prediction is concerned with detecting a future occurrence of a known behaviour based on visual observations so far. For instance, if the daily routine of a person's activities in a kitchen during breakfast time is well understood and modelled, the model can facilitate prediction of this person's next action when certain actions have been observed: the person could be expected to make coffee after finishing frying an egg. Behaviour prediction is particularly useful for explaining away partial observations, for instance, in a crowded scene when visual observation is discontinuous and heavily polluted, or for detecting and preventing likely harmful events before they take place.

Visual analysis of behaviour can assist object identification by providing contextual knowledge on how objects of interest should behave in addition to how they look. For instance, human gait describes the way people walk and can be a useful means to identify different individuals. Similarly, the way people perform different gestures may also reveal their identities. Behaviour analysis can help to determine when and where a visual identification match is most likely to be valid and relevant. For instance, in a crowded public place such as an airport arrival hall, it is infeasible to consider facial imagery identification for all the people all the time. A key to successful visual identification in such an environment is effective visual search. Behaviour analysis can assist in determining when and where objects of interest should be sought and matched against. Moreover, behaviour analysis can provide focus of attention for visual identification. Detecting people acting out of norm can activate identification with improved effectiveness and efficiency. Conversely, in order to derive a semantic interpretation of an object's behaviour, knowing what and who the object is can help. For instance, a train station staff's behaviour can be distinctively different from that of a normal passenger. Recognising a person as a member of staff in a public space can assist in interpreting correctly the behaviour of the person in question.

1.2 Opportunities

Automated visual analysis of behaviour provides some key building blocks towards an artificial intelligent vision system. To experiment with computational models of behaviour by constructing automatic recognition devices may help us with better understanding of how the human visual system bridges sensory mechanisms and semantic understanding. Behaviour analysis offers a great deal of attractive opportunities for application, despite that deploying automated visual analysis of behaviour to a realistic environment is still at its infancy. Here we outline some of the emerging applications for automated visual analysis of behaviour.

1.2.1 Visual Surveillance

There has been an accelerated expansion of closed-circuit television (CCTV) surveillance in recent years, largely in response to rising anxieties about crime and its threat to security and safety. Visual surveillance is to monitor the behaviour of people or other objects using visual sensors, typically CCTV cameras. Substantial amount of surveillance cameras have been deployed in public spaces, ranging from transport infrastructures, such as airports and underground stations, to shopping centres, sport arenas and residential streets, serving as a tool for crime reduction and risk management. Conventional video surveillance systems rely heavily on human operators to monitor activities and determine the actions to be taken upon occurrence of an incident, for example, tracking suspicious target from one camera to another camera, or alerting relevant agencies to areas of concern. Unfortunately, many actionable incidents are simply miss-detected in such a manual system due to inherent limitations from deploying solely human operators eyeballing CCTV screens. These limitations include: (1) excessive number of video screens to monitor, (2) boredom and tiredness due to prolonged monitoring, (3) lack of *a priori* and readily accessible knowledge on what to look for, and (4) distraction by additional operational responsibilities. As a result, surveillance footages are often used merely as passive records for post-event investigation. Miss-detection of important events can be perilous in critical surveillance tasks such as border control or airport surveillance. It has become an operational burden to screen and search exhaustively colossal amount of video data generated from growing number of cameras in public spaces. Automated computer vision systems for visual analysis of behaviour provide the potential for deploying never-tiring computers to perform routine video analysis and screening tasks, whilst assisting human operators to focus attention on more relevant threats, thus improving the efficiency and effectiveness of a surveillance system.

1.2.2 Video Indexing and Search

We are living in a digital age with huge amount of digital media, especially videos, being generated at every single moment in the forms of surveillance videos, on-line news footages, home videos, mobile videos, and broadcasting videos. However, once generated, very rarely they are watched. For instance, most visual data collected by surveillance systems are never watched. The only time when they are examined is when a certain incident or crime has occurred and a law enforcement organisation needs to perform a post-event analysis. Unless specific time of the incident is known, it is extremely difficult to search for an event such as someone throws a punch in front of a nightclub. For a person with a large home video collection, it is a tedious and time consuming task to indexing the videos so that they can be searched efficiently for footages of a certain type of actions or activities from years gone by. For film or TV video archive, it is also a very challenging task to search for a specific footage without text meta information, specific knowledge about the

name of a subject, or the time of an event. What is missing and increasingly desired is the ability to visually search archives by what has happened, that is, automated visual search of object behaviours with categorisation.

1.2.3 Robotics and Healthcare

A key area for robotics research in recent years is to develop autonomous robots that can see and interact with people and objects, known as social robots (Breazeal, 2002). Such a robot may provide a useful device in serving an aging society, as a companion and tireless personal assistant to elderly people or people with a disability. In order to interact with people, a robot must be able to understand the behaviour of the person who is interacting with. This ability can be based on gesture recognition, such as recognising waving and initialising a hand-shake, interpreting facial expression, and inferring intent by body posture. Earlier robotics research had focused more on static object recognition, manipulation and navigation through a stationary environment. More recently, there has been a shift towards developing robots capable of mimicking human behaviour and interacting with people. To that end, enabling a robot to perform automated visual analysis of human behaviour becomes essential. Related to the development of a social robot, personalised healthcare in an aging society has gained increasing prominence in recent years. To be able to collect, disseminate and make sense of sensory information from and to an elderly person in a timely fashion is the key. To that end, automated visual analysis of human behaviour can provide quantitative and routine assessment of a person's behavioural status needed for personalised illness detection and incident detection, e.g. a fall. Such sensor based systems can reduce the cost of providing personalised health care, enabling elderly people to lead a more healthy and socially inclusive life style (Yang, 2006).

1.2.4 Interaction, Animation and Computer Games

Increasingly more intelligent and user friendly human computer interaction (HCI) are needed for applications such as a game console that can recognise a player's gesture and intention using visual sensors, and a teleconferencing system that can control cameras according to the behaviour of participants. In such automated HCI systems using sensors, effective visual analysis of human behaviour is central to meaningful interaction and communication. Not surprisingly, animation for film production and gaming industries are also relying more on automated visual analysis of human behaviour for creating special effects and building visual avatars that can interact with players. By modelling human behaviour including gesture and facial expression, animations can be generated to create virtual characters, known as

avatars, in films and for computer games. With players' behaviour recognised automatically, these avatars can also interact with players in real-time gaming.

1.3 Challenges

Understanding object behaviour from visual observation alone is challenging because it is intrinsically an ill-posed problem. This is equally true for both humans and computers. Visual interpretation of behaviour can be ambiguous and is subject to changing context. Visually identical behaviours may have different meanings depending on the environment in which activities are taken place. For instance, when a person is seen waving on a beach, is he greeting somebody? swatting an insect? or calling for help as his friend is drowning? In general, visual analysis of behaviour faces two fundamental challenges.

1.3.1 Complexity

Compared to object recognition in static images, an extra dimension of time needs to be considered in modelling and explaining object behaviour. This makes the problem more complex. Let us consider human behaviour as an example. Human has an articulated body and the same category of body behaviours can be acted in different ways largely due to temporal variations, for example, waving fast versus slowly. This results in behaviours of identical semantics look visually different, known as large intra-class variation. On the other hand, behaviours of different semantic classes, such as jogging versus running, can be visually similar, known as small inter-class variation. Beyond single object behaviour, a behaviour can be of multiple interacting objects characterised by their temporal ordering. In a more extreme case, a behaviour is defined in the context of a crowd where many people co-exist both spatially and temporally. In general, behaviours are defined in different spatial and temporal context.

1.3.2 Uncertainty

Based on visual information alone to describe object behaviour is inherently partial and incomplete. Unlike a human observer, when a computer is asked to interpret behaviour without other sources of information except imagery data, the problem is compounded by visual information only available in two-dimensional images of a three-dimensional space, lack of contextual knowledge, and in the presence of imaging noise.

Two-dimensional visual data give rise to visual occlusion on objects under observation. This renders not all behavioural information can be observed visually. For instance, for two people interacting with each other, depending on the camera angle, almost inevitably part of or the full body of a person is self-occluded. As a result, semantic interpretation of behaviour is made considerably harder when only partial information is available.

Behaviour interpretation is highly context dependent. However, contextual information is not always directly observable, nor necessarily always visual. For instance, on a motorway when there is a congestion, a driver often wishes to find out the cause of the congestion in order to estimate likely time delay, whether the congestion is due to an accident or road work ahead. However, that information is often unavailable in the driver's field of view, as it is likely to be located miles away. Taking another example, on a train platform, passengers start to leave the platform due to an announcement by the station staff that the train line is closed due to signal failure. This information is in audio form therefore not captured by visual observation on passenger behaviours. To interpret behaviour by visual information alone introduces additional uncertainty due to a lack of access to non-visual contextual knowledge.

Visual data are noisy, either due to sensor limitations or because of operational constraints. This problem is particularly acute for video based behaviour analysis when video resolution is often very low both spatially and temporally. For instance, a typical 24 hours 7 days video surveillance system in use today generates video footages with a frame-rate of less than three frames per second, and with heavy compression for saving storage space. Imaging noise degrades visual details available for analysis. This can further cause visual information processing to introduce additional error. For instance, if object trajectories are used for behaviour analysis, object tracking errors can increase significantly in low frame-rate and highly compressed video data.

1.4 The Approach

We set out the scope of this book by introducing the problem of visual analysis of behaviour. We have considered the core functions of behaviour analysis from a computational perspective, and outlined the opportunities and challenges for visual analysis of behaviour. In the remaining chapters of Part I, we first give an overview on different domains of visual analysis of behaviour to highlight the importance and relevance of understanding behaviour in context. This is followed by an introduction to some of the core computational and machine learning concepts used throughout the book. Following Part I, the book is organised into further three parts according to the type of behaviour and the level of complexity involved, ranging from facial expression, human gesture, single object action, multiple object activity, crowd behaviour analysis, to distributed behaviour analysis.

Part II describes methods for modelling single-object behaviours including facial expression, gesture, and action. Different representations and modelling tools are considered and their strengths and weaknesses are discussed.

Part III is dedicated to group behaviour understanding. We consider models for exploring context to fulfil the task of behaviour profiling and abnormal behaviour detection. Different learning strategies are investigated, including supervised learning, unsupervised learning, semi-supervised learning, incremental and adaptive learning, weakly-supervised learning, and active learning. These learning strategies are designed to address different aspects of a model learning problem in different observation scenarios according to the availability of visual data and human feedback.

Whilst Part II and Part III consider behaviours observed from a single camera view, Part IV addresses the problem of understanding distributed behaviours from multiple observational viewpoints. An emphasis is specially placed on non-overlapping multi-camera views. In particular, we investigate the problems of behaviour correlation across camera views for camera topology estimation and global anomaly detection, and the association of people across non-overlapping camera views, known as re-identification.

References

- P.E. Agre. *The dynamic structure of everyday life*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1989.
- J.L. Barbur, K.H. Ruddock, and V.A. Waterfield. Human visual responses in the absence of the geniculo-calcarine projection. *Brain*, 103(4):905–928, 1980.
- C.L. Breazeal, editor. *Designing Sociable Robots*. The MIT Press, 2002.
- H. Buxton and S. Gong. Visual surveillance in a dynamic and uncertain world. *Artificial Intelligence*, 78(1-2):431–459, 1995.
- S. Gong, J. Ng, and J. Sherrah. On the semantics of visual behaviour, structured events and trajectories of human action. *Image and Vision Computing*, 20(12): 873–888, October 2002.
- D. Gough. The visual behaviour of infants in the first few weeks of life. *Proceedings of the Royal Society of Medicine*, 55(4):308–310, April 1962.
- R.L. Gregory. *The Intelligent Eye*. Weidenfeld and Nicolson, London, 1970.
- J.L. Harbluk and Y.I. Noy. The impact of cognitive distraction on driver visual behaviour and vehicle control. Technical Report TP 13889 E, Road Safety Directorate and Motor Vehicle Regulation Directorate, Canadian Minister of Transport, 2002.
- H. von Helmholtz. *Popular Scientific Lectures*, chapter “The Recent Progress of the Theory of Vision”. Dover Publications, 1962.
- G.W. Humphreys and V. Bruce. *Visual Cognition: Computational, Experimental and Neuropsychological Perspectives*. Lawrence Erlbaum Associates, Hove, East Sussex, 1989.
- D.J. Ingle, M.A. Goodale, and R.J.W. Mansfield, editors. *Analysis of Visual Behaviour*. The MIT Press, January 1982.
- L. Lti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, November 1998.
- D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman & Co., 1982.
- D. Rollinson. *Organisational Behaviour and Analysis: An Integrated Approach*. Prentice Hall, December 2004.
- P.H. Schiller and F. Koerner. Discharge characteristics of single units in superior colliculus of the alert rhesus monkey. *Journal of Neurophysiology*, 34(5):920–935, September 1971.
- M.D. Sherman and I.C. Sherman. The process of human behaviour. *Journal of Mental Science*, 76:337–338, 1930.
- H.A. Simon. A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99–118, February 1955.
- E.J. Warrant. Seeing in the dark: Vision and visual behaviour in nocturnal bees and wasps. *Journal of Experimental Biology*, 211:1737–1746, May 2008.
- L. Weiskrantz. Review lecture: Behavioural analysis of the monkey’s visual nervous system. *Proceedings of the Royal Society*, 182:427–455, 1972.

- T. Xiang and S. Gong. Beyond tracking: Modelling activity and understanding behaviour. *International Journal of Computer Vision*, 67(1):21–51, May 2006.
- G.Z. Yang, editor. *Body Sensor Networks*. Springer, May 2006.

Index

Symbols

1-norm 82
1D *see* one-dimensional
24/7 11, 189, 224
2D *see* two-dimensional
2D view 116
3D *see* three-dimensional
3D model 116
3D model based tracking 116
3D skeleton 116

A

absolute difference vector 322
accumulation factor 175
action 4
action cuboid 162
action detection 21, 158
action localisation 158
action model 143, 149
action recognition 19, 141
action representation 143, 149
 bag-of-words 48
action unit 17
active learning 61, 62, 279, 285
 adaptive weighting 287
 critical example 286
 expected error reduction 285
 expected model change 285
 likelihood criterion 285, 291
 pool-based 285
 query criterion 285
 stream-based 285
 uncertainty criterion 285, 292
activity 4, 23, 196, 230
activity classification 197

activity graph 196
activity model 191
activity recognition 173
activity representation
 bivariate time-series 306
 hierarchical 196
activity segmentation 180
activity transition matrix 196
acyclicity check 344
AdaBoost 83
affective body gesture 131
affective state 131, 135
affinity 208
affinity feature vector 208
affinity matrix 208, 250, 289, 307
 eigenvector 208
 normalised 209
affinity metric 207
aging society 9
alignment 76, 87
alternative hypothesis 216
ambient space 84
animation 9
annotation 162
 automated 159, 162, 165
 learning 162
anomalous region 260
anomaly 6, 205, 258, 261
 specificity 6
anomaly detection 59, 213
 cumulative factor 216
 on-line likelihood ratio test 216
 probe pattern 213
 sensitivity 216, 261
 specificity 216, 261
 threshold 216
appearance feature 28

- association 5, 7
 - atomic action 4, 122, 125
 - atomic component *see* gesture
 - atypical co-occurrence 240
 - automated visual surveillance 4
 - automatic model order selection 209
 - avatar 9, 109
 - average Kullback-Leibler divergence 293
- B**
- background subtraction 48, 306
 - bag-of-words 57, 133, 149, 157, 231
 - document size 238
 - batch EM 221
 - batch learning 236, *see also* model learning
 - batch mode *see* model learning
 - Baum-Welch algorithm 193, 334
 - Bayes factor 197
 - Bayes net *see* Bayesian network
 - Bayes' rule 128, 290
 - Bayesian adaptation 60
 - Bayesian classification 289
 - Bayesian data association 106, 107
 - Bayesian filtering 236
 - Bayesian graph 337, 342, *see also* static Bayesian network
 - Bayesian information criterion 178, 186, 192, 334, 342
 - model under-fitting 210
 - Bayesian learning 344, 350
 - conjugate prior 290
 - Bayesian model selection 197
 - Bayesian network 110
 - belief revision 111
 - belief update 111
 - clique 113
 - double count evidence 114
 - dynamic 53
 - explain away evidence 114
 - explanation 111
 - extension 111
 - join tree 113
 - most probable explanation 111
 - singly connected 113
 - static 53
 - Bayesian network learning
 - batch mode 348
 - incremental 349
 - mutual information 339
 - prior structure 340
 - scoring function 342
 - Bayesian Occam's razor 277
 - Bayesian parameter inference 233
 - Bayesian saliency 236
 - Bayesian temporal model 92
 - behaviour 3
 - abnormal 205
 - computational model 6
 - correlational context 247, 252
 - global interpretation 303
 - latent intent 303
 - modelling 5
 - normal 213
 - profile discovery 206
 - semantics 54
 - spatial context 247, 248, 306
 - temporal context 248, 252
 - temporal segmentation 334
 - behaviour affinity matrix 208
 - behaviour class distribution 283
 - behaviour classification
 - on-line 275
 - behaviour correlation 26
 - behaviour detection 6
 - behaviour hierarchy 4
 - behaviour interpretation
 - multivariate space 51
 - behaviour localisation 278
 - behaviour model
 - abnormality 219
 - abnormality update 222
 - adaptive 218
 - approximate abnormal 220
 - bootstrapping 219
 - complexity 229
 - composite 213
 - hierachical structure 229
 - incremental learning 218
 - incremental update 220
 - initialisation 219
 - normality 218
 - normality update 221
 - rarity 219
 - rejection threshold 220
 - uncertainty 229
 - behaviour pattern 23, 205, 206
 - behaviour posterior 233
 - behaviour prediction 7
 - behaviour profile 278
 - behaviour profiling 59, 201, 205
 - automatic 201, 205
 - unsupervised 205, 207
 - behaviour representation 5, 43, 54, 287
 - behaviour-footprint 250
 - behavioural context 247, 349
 - abrupt change 349
 - gradual change 349

- behavioural interpretation 51, 196
 - semantics 54
 - statistical learning 51
 - behavioural saliency 236, 242
 - irregularity 242
 - known 243
 - rarity 242
 - unknown 243
 - behavioural surprise 238, 240
 - between-class scatter matrix 81
 - between-sets covariance matrix 312, *see also*
 - canonical correlation analysis
 - Bhattacharyya distance 319
 - BIC *see* Bayesian information criterion
 - BIC score 334, 342
 - big picture 333, 335
 - binary classification 205, 283
 - false positive 205
 - sensitivity 205
 - specificity 205
 - true positive 205
 - binary classifier 81, 156
 - binary decision 81
 - biometric feature 28
 - biometrics 319
 - body gesture 17
 - body language 18
 - Boolean function 325
 - boosting 83
 - boosting learning 321
 - bottom-up 51, 196
 - bottom-up event detection 200
 - bottom-up segmentation 188
 - bounding box 44, 105, *see also* object
- C**
- camera connectivity matrix 315
 - camera topology inference 28
 - camera view
 - spatial gap 27
 - temporal gap 27
 - canonical correlation 311
 - canonical correlation analysis 134
 - between-sets covariance matrix 134
 - canonical factor 134
 - canonical variate 134, 311
 - optimal basis vector 311
 - principal angle 134
 - within-set covariance matrix 134
 - canonical direction 231
 - cardinal direction *see* canonical direction
 - cascaded LDA 253
 - cascaded learning 83
 - cascaded topic model 248, 252
 - Catch-22 problem 180, 327
 - CCA *see* canonical correlation analysis
 - CCTV *see* closed-circuit television
 - chain code 142
 - Chi square statistic 80
 - Chow-Liu tree 340
 - class relevance measure 155
 - class separability 284
 - class-conditional marginal likelihood 277
 - classification 4, 6, 53, 75
 - classification sensitivity 205
 - classifier 79
 - closed-circuit television 8, 303
 - closed-world 304
 - cloud of interest points *see* cloud-of-points
 - cloud-of-points 149, 157
 - class relevance measure 155
 - global scalar feature 155
 - clustering 58, 105, 126, 177, 256
 - co-occurrence 29, 57, 230, 231
 - co-occurring
 - event 230, 235
 - topic 235
 - coarse labelling 262
 - codebook 289
 - colour space
 - hue-saturation 103
 - complexity 10
 - computational complexity 56, 342
 - computational tractability 230, 343
 - computer games 9, 18
 - computer vision 4
 - CONDENSATION algorithm 106, 119, 129
 - conditional density propagation 106, 129
 - conditional independence 53, 290, 339
 - test 340
 - conditional probability 54, 234
 - conditional probability distribution 51, 290, 339
 - conditional random field 143, *see also*
 - Markov random field
 - conjugate prior 344
 - BDeu prior 345
 - connected component 105, 177
 - connecting-the-dots 29, 333
 - consensus probability *see* query-by-committee
 - context 11
 - global 248, 252, 256
 - local 248
 - multiple scale 248
 - regional 252
 - context dependent 11

context-aware 30
 contextual event 174, 249
 contextual information 11
 contextual knowledge 7
 contextually incoherent 258, 261
 continuous variable 55
 convolution operator 132
 corpus
 topic model 240
 correlation scaling factor 308
 correspondence problem 28, 44
 cost function 80, 192
 penalty term 192
 tractability 192
 coupled hidden Markov model 56
 covariance matrix 117
 CRF *see* conditional random field
 cross canonical correlation analysis 311
 cross-validation 162, 323
 crowd behaviour 24
 crowd flow analysis 50
 cumulative abnormality score 347
 cumulative scene vector 182
 trajectory 183
 curse of dimensionality 85, 208
 curvature primal sketch 124

D

data
 gallery 237
 informative 285
 probe 237
 test 237
 testing 51
 training 18, 51, 57, 237
 data association 106
 data mining 201, 207, 284
 DBN *see* dynamic Bayesian network
 decay factor 175, 347
 deceptive intent 21
 decision boundary 62, 284
 detection 4, 6, 158
 detection rate 347
 diagonal matrix 86
 dimensionality reduction 85, 117, 154, 208
 eigen-decomposition 209
 feature selection 155
 directed acyclic graph 110, 337, *see also*
 Bayesian network
 directed probabilistic link 192
 Dirichlet distribution 267, 270, 290, 292
 posterior 292
 prior 232

Dirichlet-multinomial conjugate structure
 234
 disagreement measure *see* query-by-
 committee
 discrete variable 55, 111, 337
 discrete-curve-evolution 188
 discrimination 6
 discriminative model 51
 displacement vector 49
 dissimilarity metric 307
 distance metric 58, 207
 distributed behaviour 26
 distributed camera network *see* multi-camera
 system
 distribution overlap 283
 document
 querying 231
 similarity matching 231
 dominant motion direction 289
 dynamic Bayesian network 55, 185, 191,
 208, 230, 333
 topology 185
 dynamic correlation
 parameter 191
 structure 191
 dynamic programming 344
 dynamic scene analysis 191
 dynamic texture 84
 dynamic time warping 207
 dynamic topic model 230, 238
 dynamically-multi-linked HMM 56, 191

E

eigen-decomposition 209
 eigenvalue 86, 117, 210, 312
 eigenvector 117, 312
 relevance learning 210
 selection 210
 EM *see* expectation-maximisation
 EM algorithm *see* expectation-maximisation
 algorithm
 emerging global behaviour 334, 345
 emotion 131
 ensemble learning 324
 ensemble RankSVM 324
 learning weak ranker 324
 entropy ratio 177
 ergodic model 162, 334
 Euclidean 85, 188
 event
 atomic action 51
 classification 177
 clustering 177

- contextual 50
- recognition 198
- event-based representation 50, 174
- exact learning 344
- expectation-maximisation 59, 270
- expectation-maximisation algorithm 104, 186
 - E-step 186, 222
 - M-step 186, 222
- expression classification 75
- expression manifold 84
- extended Viterbi algorithm 199

F

- face recognition 7
- face detection 83, 105
- facial action coding system 17
- facial expression 15, 75
 - classification 75
 - recognition 15, 75
 - representation 75
- FACS *see* facial action coding system
- factored sampling 129
- factorisation 56, 57
- false alarm 61, 283, 347
- false positive 165, 205
- feature extraction 152
- feature fusion 157
- feature representation 157
 - global 157
 - joint 157
 - local 157
- feature selection 83, 150, 154
 - kernel learning 155
 - unsupervised 210
 - variance 154
- feature space 80, 81, 133
 - correlation 133
 - fusion 133
- field of view 303
 - non-overlapping 303
- filtering threshold 352
- fitness value 120
- focus of attention 7
- forward-backward algorithm 192, 193, *see also* Baum-Welch algorithm
- forward-backward procedure *see* forward-backward algorithm
- forward-backward relevance 187
- frame differencing 306
- frame rate 305
- fundamental pattern *see* uniform pattern
- fusion 133

- expression and gesture 133

G

- Gabor filter 109, 132, 151
 - carrier 151
 - envelope 151
- Gabor wavelets 76, *see also* Gabor filter
- gait analysis 23
- gallery image 92, 320
- Gamma distribution 292
 - shape parameter 292
- Gamma function *see* Gamma distribution
- Gaussian mixture model *see* mixture of Gaussian
- Gaussian noise 129
- generalisation 6, 53
- generalised eigenvalue problem 86
- generative model 53, 231, 266
- gesture 101
 - affective 19
 - atomic action 122
 - atomic component 122
 - non-affective 19
 - sign language 19
- gesture component space 125
- gesture recognition 17, 101
 - body language 101
 - emotional state 101
 - expression 101
 - hand gesture 101
- gesture segmentation 122
- gesture tracking 102
- Gibbs sampler 234, 270
 - collapsed 270
- Gibbs sampling 230, 234, 270, 277
- Gibbs-EM algorithm 271, 273
- global abnormal behaviour 333, 347, 357
- global activity 29
- global activity analysis 29, 303
- global activity phase 334
- global anomaly detection 345
 - cumulative anomaly score 345
- global awareness 353
- global Bayesian behaviour graph 337
- global behaviour segmentation 333
- global behaviour time series 334
- global context LDA 256
- global situational awareness 29, 333
- global trail 29
- gradient descent 145
- gradient feature 50
- graph classification problem 141
- graph pruning 339, 351

graph theory 53
 graphical model 54
 directed acyclic 54
 undirected 143, 146
 graphical model complexity 192
 group activity 23, 173, 191
 event-based 174
 modelling 173
 supervised learning 173
 group association context 28

H

hand orientation 109
 histogram 114
 hard negative mining procedure 163
 hard-wired 261
 harmonic mean 241, 277
 HCI *see* human computer interaction
 HCRF *see* hidden conditional random field
 observable 146
 potential function 145, 146
 healthcare 9
 heterogeneous sensor 27, 30, 303
 heuristic search 341, 344
 hidden conditional random field 143, *see also* Markov random field
 graphical model 146
 initialisation 146
 non-convexity 146
 observable 146
 potential function 146
 undirected graph 146
 hidden holistic pattern 333
 hidden Markov model 55, 129, 185, 333
 HMM channelling 129
 HMM pathing 147
 on-line filtering 335
 hidden state variable 55
 hierachical behaviour discovery 229
 hierarchical clustering 235
 hierarchical hidden Markov model 56
 hierarchical model 230
 hierarchical topic 256
 hinge loss function 157, 323
 histogram dissimilarity measure 80
 histogram of oriented gradients 50
 HMM *see* hidden Markov model
 HOG *see* histogram of oriented gradients
 holistic awareness 29
 holistic context 29
 holistic interpretation 29
 human action 19
 human body

2D view ambiguity 116
 2D view model 116
 3D skeleton model 116
 nonlinear space 117
 observation distance 121
 observation subspace 121
 skeleton distance 121
 skeleton subspace 121
 view ambiguity 122
 human computer interaction 9, 16, 18
 human feedback 61, 262, 284
 query criterion 291
 human gait 7
 human gesture 17
 human intent 21
 human silhouette 116
 motion moment 143
 spectrum feature 142
 human-guided data mining 279
 hybrid 2D-3D representation 116
 hyper-parameter 233, 268, 271
 Dirichlet prior 271
 hyperplane 80
 hypothesis test 216

I

identification 7
 ill-posed 10, 180, 326
 image cell 267
 image classification 75
 image descriptor 48
 imbalance class distribution 287
 importance sampling 277
 inactivity 181
 inactivity break-up 182
 incremental EM *see* incremental expectation-maximisation
 incremental expectation-maximisation 59, 218, 221
 off-line 221
 on-line 221
 incremental learning 289, 348, 349
 incremental structure learning
 Naïve method 349
 information processing 5
 inter-camera
 appearance variance 303
 inter-camera gap 303
 inverse kinematics 116
 irregularity 237, *see also* behavioural saliency
 Isomap 85
 isotropic distribution 102

J

joined-up reasoning 305
 joint feature vector descriptor 143
 joint probability distribution 53, 314

K

k-means 207, 209, 256, 334
 k-nearest neighbour 82, 94, 164, 320
 greedy 164
 K2 algorithm 341
 re-formulation 341
 kernel function 80, 156
 kernel learning 155, 321
 kernel space 80
 key-frame matching 207
 KL divergence *see* Kullback-Leibler
 divergence
 Kullback-Leibler divergence 255, 277, 287,
 294
 distance metric 294
 non-symmetric 294
 symmetric 294

L

L1 norm 320
 label switching 237, 243
 Lagrange multiplier 80
 Laplace Beltrami operator 85
 Laplacian eigenmap 85
 Laplacian matrix 86
 largest eigenvector 209
 latent behaviour 235
 latent Dirichlet allocation 29, 231, 253, 267,
 353
 time-ordered 353
 latent structure 232
 latent topic 231
 LBP *see* local binary pattern
 LBP histogram 78, 83
 LDA *see* latent Dirichlet allocation
 learning
 model parameter 193
 model structure 191
 procedure 234
 rate 221
 strategy 57
 learning strategy *see* learning
 Levenshtein distance 207
 likelihood function 94
 likelihood score 291
 likelihood test 60

linear combination 311
 linear dimensionality reduction 85
 linear discriminant analysis 81, 231
 linear programming 82
 linearly non-separable 80
 local (greedy) searching 211
 local binary pattern 76
 local block behaviour pattern 287
 local motion event 230
 localisation 158
 locality preserving projection 85
 supervised 88
 locally linear embedding 85
 log-likelihood 104, 145, 213, 254, 255, 345,
 356
 normalised 213
 logistic regression 163
 low-activity region 311
 LPP *see* locality preserving projection

M

man-in-the-loop 279
 manifold 84
 manifold alignment 87
 MAP *see* maximum a posteriori
 marginal likelihood 236, 240, 277
 marginal probability distribution 314
 Markov chain 127, 162, 230
 Markov chain Monte Carlo 233, 234, 278,
 344
 Markov clustering topic model 57, 230, 231
 Markov process 127, 128
 first-order 128, 334
 gesture 127
 second-order 128
 state 128
 state space 128
 Markov property 93, 128
 Markov random field
 conditional random field 143
 hidden conditional random field 143
 Markovian assumption 335, *see also* Markov
 property
 maximum a posteriori 237, 349
 maximum canonical correlation 312
 maximum likelihood 104
 maximum likelihood estimation 104, 145,
 178, 192, 211, 344
 maximum margin 80
 maximum posterior probability 221
 MCMC *see* Markov chain Monte Carlo
 mean field approximation 277
 mean shift 105

- message-passing process 111
 - metadata feature 231
 - minimum description length 126, 147
 - minimum eigenvalue solution 86
 - mixing probability 211, 220
 - mixture component trimming 223
 - mixture model 126
 - mixture of Gaussian 103, 126, 129, 147
 - mixture of probabilistic principal components 59
 - MLE *see* maximum likelihood estimation
 - model exploitation 286
 - model exploration 286
 - model inference 270
 - model input
 - human annotation 284
 - model learning
 - batch mode 325, 339, 348
 - incremental 339, 348
 - shared common basis 266
 - statistical insufficiency 272
 - structure 267
 - model matching 6
 - model order 178
 - model order selection 126, 178, 208, 334
 - automatic 209
 - model output
 - human feedback 284
 - model over-fitting 233, 240, 266
 - model parameter update 220
 - model robustness 229
 - model sensitivity 216
 - model specificity 216
 - model structure 191
 - adaptation 223
 - discovery 191
 - model topology 191
 - model under-fitting 210, 251
 - model-assisted search 279
 - modelling 4
 - behaviour 43
 - gesture 101
 - most probable explanation 192
 - motion
 - descriptor 47, 162
 - feature 177
 - template 47
 - motion direction profile 289
 - motion history volume 46
 - motion moment 102
 - first-order 102
 - second-order 102
 - trajectory 102
 - zeroth-order 102
 - motion-history-image 46
 - multi-camera behaviour correlation 26
 - multi-camera distributed behaviour 26
 - multi-camera system 26, 303
 - blind area 27
 - disjoint views 303
 - surveillance 303
 - multi-camera topology 314
 - connected regions 315
 - region connectivity matrix 315
 - multi-channel kernel 162
 - multi-instance learning 159
 - multi-object behaviour 230
 - multi-observation hidden Markov model 56, 185
 - multi-observation HMM *see* multi-observation hidden Markov model
 - multi-view activity representation 306
 - multimodal 210
 - multinomial distribution 231, 290, 294, 339
 - conjugate 345
 - multiple kernel learning 155
 - mutual information 314
 - analysis 313
- N**
- naïve Bayesian classifier 290
 - natural gesture 122
 - nearest neighbour 60, 86
 - nearest neighbour classifier 82, *see also*
 - k-nearest neighbour
 - negative training video 163
 - network scoring 352
 - neutral expression 86
 - Newton optimisation 323
 - non-visual context 114
 - nonverbal communication 131
 - normal distribution 211
 - normalised probability score 291
 - normality score 260
- O**
- object
 - appearance attribute 44
 - bounding box 44
 - descriptor 44
 - tracking 44, 327
 - tracklet 46
 - trajectory 44
 - object association 27
 - object association by detection 327
 - object representation 47

- bag-of-words 47
- constellation model 47
- object segmentation 47, 48, 180
- object-based representation 43
- objective function 81, 86, 145, 156, 322
- observation
 - distributed viewpoints 303
 - multiple viewpoints 303
 - oversampling 303
 - single viewpoint 303
- observation probability 55
- observation variable 55
- observational space 185
 - factorisation 185
- off-line processing 183
- on-line anomaly detection 206, 213
 - likelihood ratio test 217
 - maximum likelihood estimation 217
- on-line filtered inference 236
- on-line likelihood ratio test 216
- on-line model adaptation 59
- on-line processing 183, 230
- on-line video screening 237
- one class learning 283
- one-dimensional 122, 185
- one-shot learning 266, 273
- one-versus-rest 81, 83, 156, 217
- open-world 310
- optical flow 47, 49, 231, 287
 - Lucas-Kanade method 287
- optimisation problem 156
- ordering constraint 344
- outlier 59
- outlier detection 266, 278, 283
- over-fitting *see* model over-fitting
- oversampling 83

P

- pair-wise correlation 310, 337
- pan-tilt-zoom 27
- parallel hidden Markov model 56
- parameter learning 344
 - maximum likelihood estimation 344
- part-based representation 47
- PCA *see* principal component analysis
- Pearson's correlation coefficient 307
- pedestrian detection 83
- person re-identification 24, 28, 319
 - absolute similarity score 321
 - context-aware search 326
 - contextual information 327
 - matching criterion 320
 - ranking 321

- relative ranking 321
- personalised healthcare 9
- phoneme 122
- pipelined approach 229
- pixel-based representation 48
- pixel-change-history 49, 174
- point clouds 150
- point distribution model 116, 117
- point estimation 270
- Polya distribution 269
- polynomial 81
- positive importance weight 323
- posterior distribution 255
- posterior probability 93
- prediction 5, 131
 - blind search 131
 - guided search 131
- predictive likelihood 233, 237, 241
- Prewitt edge detector 152
- Prim's algorithm 340
- principal component analysis 85, 117
 - dimensionality reduction 117
 - hierarchical 117
- prior domain knowledge 243
- probabilistic dynamic graph 191
- probabilistic graphical model 51, 53, 191, 337
- probabilistic latent semantic analysis 53, 258
- probabilistic relative distance comparison 320
- probabilistic topic model 53, 56, 230
 - hierarchical Dirichlet processes 53
 - latent Dirichlet allocation 53
- probability density function 103, 210
- probability theory 54
- probe behaviour pattern 216
- probe image 92, 319, 320
- probe pattern 213
- probe video 259
- profiling 59
- PTM *see* probabilistic topic model
- PTZ *see* pan-tilt-zoom

Q

- Quasi-Newton optimisation 145
- query criterion 291, *see also* human feedback
 - adaptive selection 294
- query score 291
- query-by-committee 286, 292
 - consensus probability 293
 - disagreement measure 293
 - modified 286

vote entropy 292

R

radial basis function 81
 random process 125
 random variable 53, 110, 134
 multivariate 54
 RankBoost 321
 ranking function 322
 ranking problem 321
 RankSVM 321
 complexity 323
 ensemble 324
 scalability 323
 spatial complexity 324
 rare behaviour 262, 266
 benign 284
 classification 265
 model 265
 rarity *see* behavioural saliency
 re-identification 12, *see also* person
 re-identification
 reasoning 109
 regional
 context 253
 correlational context 254
 temporal context 254
 regional activity 316
 regional activity affinity matrix 312
 regional activity pattern 310, 334
 regional temporal phase 256
 relative entropy 287
 relative ranking score 322
 relevance feedback 284
 relevance learning 210
 relevance rank 322
 representation 4
 representation space 6
 robust linear programming 82
 robustness 205, 230, 333
 rule-based 53, 58
 run-time anomaly measure 216

S

salient behaviour 230, 242
 salient pixel group 177
 salient point 161
 sampling bias 210
 scalability 333
 scale-invariant feature transform 48, 161
 scaling factor 251
 spatial 308

scatter matrix
 between-class 82
 within-class 82
 scene decomposition 289, 306
 activity-based 307
 scene layout 305
 scene vector 181
 scoring function optimisation 340
 seeding event 174
 segmentation 47, 48, 122, 180, 250
 self-occlusion 116, 120
 semantic gap 190
 semantic region 247
 semantic scene
 decomposition 250
 segmentation 251
 semantic video indexing 190
 semantic video search 190
 semantics 54
 semi-supervised learning 60, 242, 285
 sensitivity 216, 230, 333, *see also* model
 sensitivity
 sequential minimisation optimization 157
 shape descriptor 142
 chain code 143
 shape template 47
 SIFT *see* scale-invariant feature transform
 SIFT descriptor 161
 silhouette 20, 47, 142
 similarity measure 86
 single-camera system 303
 situational awareness 29, 247
 holistic 303
 skeleton model 116
 skin colour 103, 107
 model 103
 sliding-window 158, 162, 188, 189, 353
 SLPP *see* locality preserving projection
 social robot 9
 space-time cloud 149
 space-time cuboid 162
 space-time descriptor 149
 space-time interest point 48, 132, 149
 bag-of-words 133
 cloud 150
 cuboid 132
 descriptor 132
 detection 132
 histogram 132
 prototype 132
 prototype library 132
 volume 132
 space-time shape template 47
 sparse data 210

sparse example 262
 sparsity problem 272
 spatial topology 316
 spatio-temporal space 17
 specificity 216, *see also* model specificity
 spectral clustering 208, 250, 289, 334
 spectrum feature 142
 state inference 198
 state space 117, 191
 factorisation 191
 state transition 18
 state transition probability 55
 static Bayesian network 54
 statistical behaviour model 51, 57
 statistical insufficiency 266
 stochastic process 18, 128
 string distance 207
 strong classifier 83
 structure learning 339
 constraint-based learning 340
 subspace 81
 nonlinear 117
 subspace analysis 81
 subtle behaviour 266
 model 265
 subtle unusual behaviour 283
 sufficient statistics 221, 350, 351
 supervised behaviour modelling 266
 supervised learning 58, 173
 limitation 266
 support vector 80
 support vector machine 80, 105, 135, 156,
 163, 321
 support vector ranking 321
 synchronised global behaviour 334
 synchronised global space 305

T

template matching 79, 128
 template-based representation 46
 temporal correlation 191
 temporal model 84, 92
 temporal offset 334
 temporal process 7
 temporal segmentation 122, 124
 temporal topology 316
 text document analysis 47, 57, 231
 text mining 57
 texture descriptor 76
 three-dimensional 109
 time delay index 311
 time series function 7
 time varying random process 128

time-delayed mutual information 313
 time-delayed probabilistic graphical model
 337
 top-down 51, 196
 rules 261
 top-down model inference 200
 topic
 inter-regional 256
 profile 258
 regional 256
 topic model 231
 complexity 240
 corpus 240
 document 231, 240
 topics 240
 topic profile 256
 topic simplex 256, 259
 topological sorting 340
 topology inference 27
 tracking 44, 103
 body part 103
 discontinuous motion 106
 holistic body 102
 salient point 161
 tracklet *see* object
 tractability 333
 trajectory *see* object
 trajectory transition descriptor 161
 trajectory-based representation 44
 transformation matrix 81
 two-dimensional 103

U

uncertainty 10
 under-fitting *see* model under-fitting
 uniform pattern 77
 unimodal 210
 unit association 48
 unit formation *see* segmentation
 unlikely
 behavioural 237
 dynamical 237
 intrinsic 237
 unsupervised behaviour modelling 266
 unsupervised behaviour profiling 207
 unsupervised clustering 191
 unsupervised feature selection 210
 unsupervised incremental learning 59
 unsupervised learning 58, 201, 283
 unusual behaviour 283
 unusual behaviour detection 59, 61

V

validation set 163, 324, 352
 variance 154
 variation 75
 between-class 75
 inter-class 10
 intra-class 10
 within-class 75
 variational distribution 255
 variational EM 255
 variational inference 255, 356
 variational parameter 255
 vector descriptor 142
 video 173, 181
 bag-of-words 48, 267
 clip 231, 254
 clip categorisation 236
 content 181
 content trajectory 187
 document 254, 259
 representation 173
 segmentation 173
 semantic interpretation 173
 stream 173
 video analysis
 behaviour-based 279
 video content analysis 190
 video content breakpoint 185
 video corpus 267
 video descriptor 48
 video document 57, 232, 267
 video document category 232
 video indexing 8
 video mining 201
 video polyline 183

video saliency 236
 video screening 236
 video search 8
 video segmentation 180, 183, 185
 on-line 187
 semantic content 183
 video semantic content 181
 extraction 181
 video topic 57, 232
 video word 48, 57, 232, 267
 view selection 120
 visual analysis of behaviour 3, 4
 visual appearance 303
 visual behaviour 3
 visual context 15, 24, 247
 visual cue 5, 107
 colour 107
 motion 107
 texture 50
 visual saliency 161
 visual search 7, 9
 visual surveillance 8, 59
 Viterbi algorithm 198
 vote entropy *see* query-by-committee

W

weak classifier 83
 weakly-supervised joint topic model 267
 weakly-supervised learning 61, 262
 wearable camera 30
 wide-area scene 26, 303, 333
 wide-area space 303
 within-class scatter matrix 81
 within-set covariance matrix 312