Machine Vision and Applications (2008) 19:345–357
DOI 10.1007/s00138-008-0132-4

SPECIAL ISSUE PAPER

Crowd analysis: a survey

Beibei Zhan · Dorothy N. Monekosso · Paolo Remagnino · Sergio A. Velastin · Li-Qun Xu

Received: 7 May 2007 / Accepted: 21 January 2008 / Published online: 10 April 2008

Abstract In the year 1999 the world population reached 6 billion, doubling the previous census estimate of 1960. Recently, the United States Census Bureau issued a revised forecast for world population showing a projected growth to 9.4 billion by 2050 (US Census Bureau, http://www.census.gov/ipc/www/worldpop.html). Different research disciplines have studied the crowd phenomenon and its dynamics from a social, psychological and computational standpoint respectively. This paper presents a survey on crowd analysis methods employed in computer vision research and discusses perspectives from other research disciplines and how they can contribute to the computer vision approach.

Keywords Crowd studies · Crowd dynamics · Socio-dynamics · Crowd simulations · Computer vision

1 Introduction

The steady population growth, along with the worldwide urbanization, has made the crowd phenomenon more frequent. It is not surprising, therefore, that crowd analysis has received attention from technical and social research disciplines. The crowd phenomenon is of great interest in a large number of applications:

Crowd management: Crowd analysis can be used for developing crowd management strategies, especially for increasingly more frquent and popular events such as sport matches, large concerts, public demonstrations and so on, to avoid crowd related disasters and insure public safety.

Public space design: Crowd analysis can provide guidelines for the design of public spaces, e.g. to make the layout of shopping malls more convenient to costumers or to optimize the space usage of an office.

Virtual environments: Mathematical models of crowds can be employed in virtual environments to enhance the simulation of crowd phenomena, to enrich the human life experience.

Visual surveillance: Crowd analysis can be used for automatic detection of anomalies and alarms. Furthermore, the ability to track individuals in a crowd could help the police to catch suspects.

Intelligent environments: In some intelligent environments which involve large groups of people, crowd analysis is a pre-requisite for assisting the crowd or an individual in the crowd. For example, in a museum deciding how to divert the crowd based on to the patterns of crowd.

Crowd management and public space design are studied by sociologists, psychologists and civil engineers; virtual environments are studied by computer graphic researchers; visual surveillance and intelligent environments are of interest to computer vision researchers. The approach favored by psychology, sociology, civil engineer and computer graphic research is an approach based on human observation and analysis. Sociologists, for instance, study the characters of a crowd as a social phenomenon, exploring human factors. For example, the computational model developed by Seed Projects at Stanford University [1], incorporated human behavior in environments with emergency exits. The Crowd-MAGS Project, which is funded by GEOIDE and the Canadian

B. Zhan · D. N. Monekosso · P. Remagnino (✉) · S. A. Velastin
Digital Imaging Research Centre, Kingston University,
Kingston upon Thames, UK
e-mail: p.remagnino@kingston.ac.uk

L.-Q. Xu
Research and Venturing, BT Group PLC, London, UK

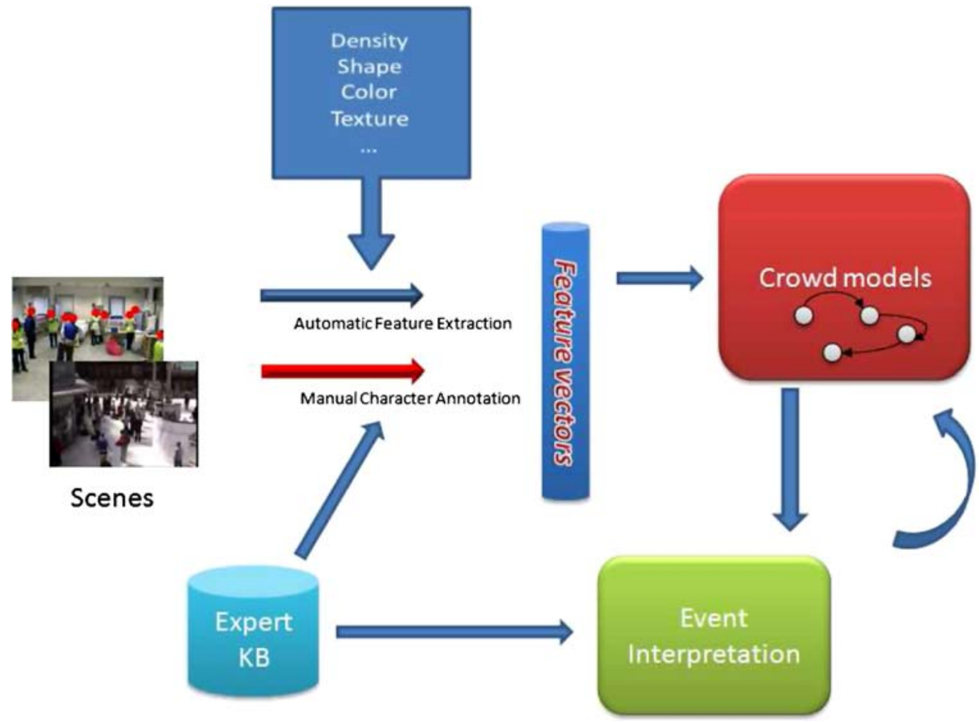

Fig. 1 A framework for Crowd analysis

Network of Centers of Excellence in Geomatics, aims to develop micro-simulations of crowd behaviours and the impact of police or military groups [23]. The Police Academy of the Netherlands and School of Psychology of University of Liverpool are cooperating on a project funded by the UK Home Office: "A European study of the interaction between police and crowds of foreign nationals considered to pose a risk to public order" [2].

On the other hand, computational methods such as those employed in computer graphics and vision methods focus on extracting quantitative features and detecting events in crowds, synthesizing the pehnomenon with mathematical and statistical models. For example, early projects funded by the EPSRC in the UK were concerned with measuring crowd motion and density and hence potentially dangerous situations [26,89,95]. The EU funded project PRISMATICA [75] and ADVISOR [3], completed in 2003, were concerned with the management of public transport networks through CCTV cameras. The UK EPSRC funded project BEHAVE, was concerned with pre-screening of video sequences for the detection of abnormal or crime-oriented behaviour [12]. ISCAPS [44] started in 2005, a consortium of ten European ICT companies and academic organizations, aims to provide automated surveillance of crowded areas. SERKET, a recently started EU project aims to develop methods to prevent terrorism [42].

Figure 1 illustrates the processes involved in crowd analysis. In a crowd scene the attributes of importance are crowd density, location, speed, etc. This information can be extracted either manually or automatically using computer vision techniques. Crowd models can then be built based on the extracted information. Event discovery is achieved using pre-compiled knowledge of the scene or using the computational model, although both approaches can be combined. In both cases the model is updated with newly extracted information.

The paper is organised as follows. Section 2 introduces research in automatic crowd feature extraction. Section 3 discusses existing work on crowd modelling and crowd event inference. Section 4 and 5 provide some examples of how the two complementary approaches can be bridged.

2 Crowd information extraction

The components of crowd analysis from a computer vision perspective are described in Table 1. Essentially, the typology of sensors and their topology influence the scene capture processes; environmental conditions, such as natural and artificial illumination changes often introduce noise; the scene typology affects the type of process one requires to extract the most accurate information of a dyanmic scene.

Visual surveillance methods have been devloped to estimate motion of objects and people in the scene, in isolation or in groups; a review can be found in [38]. When video is analysed for very crowded scenes, conventional computer vision methods are not appropriate, in these cases methods must be designed to cope with extreme clutter. Features from conventional image processing are still employed, such as colour, shape and texture etc. However, sophisticated methods have been developed to retrieve crowd information. In the following sections we will review the existing state of the art.

Table 1 Features in crowd analysis by computer vision methods

Sensor typology and topology	Environmental conditions	Scene typology	
		Individual characters	Collective
Moving or Static platform Number of cameras Type of video sequence: color or gray scale, etc.	Indoor/outdoor Level of clutter Light condition, etc.	location/velocity/etc. Appearance, etc.	Crowd density Average speed, etc.

2.1 Crowd density measurement

An important crowd feature is crowd density and it is natural to think that crowd of different density should receive a different level of attention. Polus et al. [74] provide a clear idea of the problem of *level of services* for a pedestrian flow defined as: free flow, restricted flow, dense flow, and jammed flow according to a density metric defined as the number of pedestrians per unit area. Here we review some research either estimating the crowd density directly or counting number of pedestrians which provide information for density estimation.

Research methods have been proposed for crowd analysis which employ background removal techniques. In [95] a reference image with only background is used to classify image pixels as belonging to either pedestrians or background. A functional relationship between the number of pedestrian-classified pixels and number of people is then established manually for the measurement of crowd density. Another example is proposed by Ma et al. [61] using background removal. A mathematical relation for geometric correction for the ground plane is derived. The authors proved that it can be directly applied to all foreground pixels. A linear relation between the number of pixels and number of persons was derived by applying the geometric correction. These works have a typical assumption that the number of foreground pixels are proportional to the number of people, which is only true when there are not serious occlutions between people Dong et al. [27] makes use of examples to map the global shape feature to configurations of humans directly. This training based algorithm is a quite novel approach but the problem of how to decide the size of the training dataset remains unclear.

Image processing and pattern recognition techniques are also used for the analysis of the scene to estimate the crowd density. Marana et al. [64] assume that images of low-density crowds tend to present coarse texture, while images of dense crowds tend to present fine textures. Self-organising neural maps [65] combined with Minkowski fractal dimensions [63] are employed to deduce the crowd density from the texture of the image. The work by Marana is compared in [76] with another method that uses Chebyshev moments. An optimization of performance under different illumination conditions is discussed. Lin et al. [60] present a system that estimates the crowd size through the recognition of the head contour using Haar wavelet transform (HWT) and support vector machines (SVM).

Approach of information fusion has also be applied, e.g. Yang et al. [94] estimate the number of people directly from groups of image sensors. For each sensor, foreground objects are segmented from the background, and the resulting silhouettes are aggregated over the sensor network. A geometric algorithm is then introduced to limit the number and possible locations of people using silhouettes extracted by each sensor. Alternative methods combine several techniques, to achieve more accurate and reliable measurements. For example, in [89], an edge-based technique is integrated with background removal using a Kalman filter. Marana et al. [62] use different methods including Fourier and Fractal analysis and classifiers to estimate the crowd density level. Kong et al. [54,55] employ background subtraction and edge detection; the work defined the extracted edge orientation and blob size histograms as features. The relationship between the feature histograms and the number of pedestrian is learned from labelled training data. Obvious more cues may indicate a more accurate solution.

2.2 Recognition

Conventional visual surveillance focuses on object detection and tracking. In essence, image processing techniques are employed to extract the chromatic and shape information of the moving objects and the background for detecting and tracking purposes.

For crowd dynamics modeling, detecting and tracking are also important as they provide the location and velocity features of the dynamics. Crowded scenes add a degree of complexity to the conventional detection and tracking problem of single individuals. In the following sections we concentrate on methodologies for crowded situations.

2.2.1 Face and Head Recognition

Face is the most discriminating feature of the human body and many researchers try to detect pedestrian through face detection. Majority of the existing research employs supervised learning methods. Here we review a few attempts to detect the faces in complex scenes.

Early works like [87] in which a technique using genetic algorithms is employed for face localization in a complex scene. The system proceeds with a training phase to generate a simple object mean image using a single object image, and a test phase using arbitrary images.

However the previous work highly depends on the training set and if the faces appear at different sizes and orientations, it may require a very large training set and long processing time. Hence different techniques have been developed to address the problem of multi-view face detection. Li et al. [59] proposes a pyramid structure that adopts coarse-to-fine strategy to handle pose variance. Another approach is by Jone et al. [45], in this work different detectors are for different views of the face, and a decision tree is trained to determine the viewpoint class. Huang et al. [39] uses Width-First Search tree structure to improve the performance in both speed and accuracy. This kind of work is quite likely to be adopt into crowd analysis, espeically from a single camera view, as the problem of human pose and the perspective are both compensated here.

Methodologies for stereo face detections in crowd have also been developed. For example Huang et al. [40] propose a three steps technique: first extracting the likelihood evidence of heads from the stereo image by scale-adaptive filtering; then spurious clues are suppressed from the extracted points according to the average human height; finally the human heads are located by applying a mean-shift algorithm to the likelihood map.

2.2.2 Pedestrian and Crowd recognition

Pedestrian detection and tracking is a well studied problem in computer vision. Many methods have been proposed, such as using the afore mentioned background removal technique, or combining chromatic and shape information of the tracked pedestrians. The following sections discuss the methods that try to provide a solution for pedestrian detection in crowded scenes.

- **Occlusion handling.** Occlusion caused by the high clutter of the pedestrian in crowd scene is the major challenge for crowd detection problem. Some research addresses the problem by using human body parts. Wu et al. [92] propose a method to detect multiple-partially occluded human in a single image. Edgelet features are introduced in their work. Part detectors based on edgelet features are learned by a boosting method. Responses of part detectors are combined to form a joint likelihood model that includes cases of multiple, possibly inter-occluded humans. The human detection problem is then formulated as one of maximum a posteriori (MAP) estimation. The models of group of people in [29] are initialised based on segmenting the body into regions by modelling their appearance and spatial distribution. A framework uses maximum likelihood estimation to estimate the best arrangement of people in term of a 2D translation that yields segmentation for the foreground region. Occlusion reasoning is then conducted to recover relative depth information.

  Leibe et al. [58] present a different algorithm that integrates evidence in multiple iterations and from different sources. Local cue is based on a scale-invariant extension of Implicit Shape Model (ISM), and global consistency is enforced by adding the information from global shape cues. Local and global cues are combined via a probabilistic top-down segmentation to detect the pedestrian.
- **Moving Views.** Special solutions are required for moving platforms for some of the applications e.g. for on-board vision system to assist a driver. Some of the implementations make assumptions of human appearance.In Broggi et al.'s work [16] a coarse detection of pedestrian is computed through the processing of a single image based on shape of human body assumption of symmetry, size and ratio. Heisele et al. [33] apply spatio-temporal methodologies by recognizing walking pedestrian based on the characteristic motion of the legs of a pedestrian walking parallel to the image plane. Each image is segmented into region-like image parts by clustering pixels in a combined color/postion feature space. A classfier is then used to extract the clusters which are mostly like to be the pedestrian's legs.

  Different from above, Shashua et al. [81] describe a functional and architectural breakdown pedestrian detection system. Single classification is based on a scheme of breaking down the class variability by repeatedly training a set of relatively simple classification performance results. The path from single-frame to system level performance includes the integration of additional cues measure over time, situation specific features and via building up additional object categories consisting of vehicles and stationary background structures.
- **Spatial-temporal methods.** Besides conventional cues of pedestrian appearance, space-temporal cues are used for detection. Brostow et al. [17] tackle the problem by tracking simple image features and probabilistically grouping them into clusters representing independently moving entities. Space-time proximity and trajectory coherence through image space are used as the only probabilistic criteria for clustering. Moreover, this motion-based detection could be easily extended to tracking of individuals in dense crowds by merging the outcomes.

  In extremely cluttered scenes, individual pedestrian cannot be properly segmented in the image. However, sometimes the *crowd* within which the pedestrians share a similar purpose can be recognizad. Reisman et al. [79] propose a scheme that uses slices in the spatio-temporal domain to detect inward motion as well as intersections between multiple moving objects. The system calculates a probability distribution function for left and right inward motion and uses these probability distribution functions to infer a decision for crowd detection.

2.3 Tracking

Tracking has been proposed to localize the interested object in time-space. Also the velocity feature can be derived afterwards. Though as a natural extension of detection, tracking has its own problem to recognize and identify pedestrians in the consecutive frames. Tracking could be regarded as the most popular topic in visual surveillance, however, currently for crowd analysis, most of the techniques are validated only for multiple (e.g. up to ten) people.

As discussed in the last subsection, occlusions could occur very frequently when there are many objects and people in the scene. Tracking techniques have to overcome the problem in order to continuously track before, during and after the occurrence of occlusions. A comprehensive review on occlusion handling can be found in [31]. A formulation of the occlusion problem is provided, and the techniques are divided in two groups: merge-split approach, which addresses the problem to re-establish object identities following a split, and straight-through approaches, which maintains object identities at all times.

The following text covers three aspects: the techniques which are developed to track multiple people(objects) without any assumptions of the dependence of their motion, e.g. interactions etc.; the techniques which try to explain the interactions between the pedestrians; and also some practical analysis of handling the problem of occlusion in the crowd situation.

2.3.1 Tracking methodologies

Crowd scenes increase the complexity of tracking because there are multiple moving objects in the scene. Quite a few techniques are developed based on the colour, geometry and other features for tracking.

- **Likelihood.** Color, edge etc. are the most popular features in tracking. In crowd salient traceable image features are particular interested for tracking. As one of the good candidates, interest points (IPs) are employed in [31,67]. In both works the IPs are obtained by a popular colour Harris detector. Gabriel characterized IPs by their position relative to the estimated centre of the object and Mathes built a point distribution model between ASM and AAM. Both of the methods require a pre-defined region (or object) of interest. There salient features are benefit from their robustness under different light conditions. The tracking inference using these features can work better under occlusions than using the entire contour. Therefore the usage of those features could be more appliable to large amount of people in the scene.
- **Human body model.** Methods using models of human bodies or human body parts have been developed for tracking in complex crowded scenes, which are usually completed with probabilistic frameworks. Zhao et al. [99, 100] have been working on the former approach, using explicit 3D human shape models. The problem of detection and tracking are formulated as one of Bayesian inference to find the best interpretation given the image observations. The latter one as the work from Wu et al. [93] extend the previous detection work in [92] (which has been discussed) using edgelet features to human body part detectors. Tracking is implemented by probabilistic data association, i.e. matching the object hypotheses with the detected response.
- **Tracking inference strategies.** Tracking inference strategies have been developed for the problem of tracking multiple objects. For non-linear and non-Gaussian dynamic models, particle filter technique, also known as CONDENSATION [43], is one of the most popular among those. Particle filters are sequential Monte Carlo methods based upon a point mass (or 'particle') representations of probability densities [28]. Large portion of multiple object tracking work have employed this technique. For example, Venegas et al. [90] use particle filter to track the moving objects by generating hypotheses on the top-view reconstruction of the scene. Okuma et al. [72] combine mixture particle filters and Adaboost algorithm. Sidenbladh et al. [82] extend the particle filter formulation according to finite set statistics (FISST) for tracking. Cai et al. [18] tackle the problem by embedding the meanshift algorithm into the particle filter framework. Koller-Meier et al. [53] introduce an extension of the CONDENSATION algorithm that relies on a single probability distribution of describe the likely states of multiple objects. Kang et al. [46] propose the discrete shape model and the competition rule to improve the performance of the condensation tracker for real time tracking.

To address data association problem, there are Multiple Hyphotheses Tracker (MHT) and Joint Probabilistic Data Association Filter (JPDAF). MHT tries to keep the track of all the possible hyphotheses over time [78]. A details summary and a discussion of MHT for multiple target tracking is included in [13]. MHT suffers from the storage of the redundant track, hence some of the work propose extensions and modifications to the algorithm to get better performances, e.g. [32]. JPDAF computes a Bayesian estimation of correspondence between the different features and the different objects, e.g. Rasmussen and Hager [77] apply this technique with color region and snake-based tracker. Another approach has been introduced by Karlsson [47], which uses Monte Carlo method.

The fusion of the different cues from a number of detection and tracking algorithms are also used to produce a more robust tracker. Siebel et al. [83] propose a tracking system containing three co-operating parts: an Active Shape Tracker, a Region Tracker, and a Head detector. Spengler and Schiele [85] proposes an approach based on the principles of self-organization of the integration mechanism and self-adaptation of the cue models during the tracking. Cues from different sensors and models can increase dimension of information, which is preferable in the multiple objects situations. However, the goodness of integration scheme is very crucial in these algorithms.

2.3.2 Tracking interacting people

In certain cases, interaction happens frequently in crowded scene. Researchers have shown great interest in studying these interactions to get the new perspectives on tracking techniques.

Some of the work formulate the interaction to enhance the tracking scheme. For example both Smith et al. [84] and Khan et al. [49] propose to use Markov Chain Monte Carlo (MCMC) and the particle filter. Smith used a joint multi-object state-space formulation and a trans-dimensional MCMC particle filter to recursively estimate the multi-object configuration and search efficiently the state-space. Khan developed a joint tracker that included a motion model to maintain the identity of targets throughout and interaction, thus to reduce tracker failure. Pre-defined motion models are used in this approach, with the trade-off between improving the tracking performance in crowd with known interations and the adaption of the motion model to arbitrary crowd.

Some researchers interpret interactions as relationships between pedestrians and a group (pedestrian merging/splitting into groups). Marques et al. [66] propose a two-layer solution to overcome the problem. The first layer produces a set of spatio temporal strokes based on low level operations to track the active regions. The second layer performs a consistent labelling of detected segments using a statistical model based on Bayesian networks which is recursively computed during the tracking operation. Mckenna et al. [69] perform tracking at three levels: regions, people and groups. Background subtraction is used to cope with shadows and unreliable colour cues. Colour information is used to disambiguate occlusions and to provide qualitative estimates of depth ordering and position. Pedestrian merging and group splitting are frequent phenomena in the crowded scene, however, the major challenge for this kind of methods is to recover the object lable after splitting from the group.

Sullivan et al. [86] label tracking targets by exploring the trajectories. Trajectories of when a target is isolated are found and it is claimed that these trajectories end when targets interact. A graph structure has been formed by the interactions of these trajectories. This method could be very useful for off-line crowd analyzing but for online processing it may have a bottleneck in the storage of the trajectories.

2.3.3 Tracking from multiple views

For large public areas the use of a multi-camera system is required to cover most of the monitored areas.

For the multi-camera system arrangement, Mittal et al. [70] present a system named M2Tracker using multiple synchronized cameras located far from each other for segmenting, detecting and tracking multiple people in a cluttered scene. First, a region-based stereo algorithm is introduced for finding 3D points inside an object. Then, a scheme is developed dynamically assigning priors for different objects at each pixel. Finally, the evidences gather from different camera pairs are combined using occlusion analysis to obtain a globally optimum detection and tracking of objects. A different arrangement of cameras is used in [20]. The method uses both static and Pan-Tilt-Zoom (PTZ) cameras. The static cameras are used to locate people in the scene, while the PTZ cameras *lock-on* to the individuals and provide visual attention. The underlying visual processes rely on colour segmentation, movement tracking and shape information to locate target candidates and colour indexing methods to register these candidates with the PTZ cameras.

Meanwhile special techniques have been developed for the tracking from multiview, normally a planar homography constraint would be included. For example in [48], feet regions of the people are located by the constraint. The contiguous spatio-temporal region formed by the feet regoins belonging to the same person are clustered as the track of the person. In [50] people's ground points are located and a multi-hypothesis framework using particle filter is developed for tracking.

3 Crowd modelling and events inference

Dynamics in public spaces can indeed be recurrent. Crowd information can be better exploited to indicate the status of the crowd so that crowd events can be inferred. Crowd models have been built to represent these status, either implicitly or explicitly. On the other hand, some research makes direct use of crowd information instead of building models. In such cases, the events are usually inferred based on some prior knowledge of the properties of the particular scene and the crowd.In this section, crowd models and events inference in computer vision will be presented as well as some crowd models from non vision areas.

3.1 Crowd models and crowd events inference in computer vision

In computer vision crowd modelling is achieved based on the extracted information from visual data and normally can be employed in crowd events inference. Meanwhile there are also some approaches attempt to infer events without construction of models. Here examples are given for both of the cases.

– In computer vision approach crowd models are built as representations of recurrent behaviours by analysing video data of the crowd through vision methods.
  Zhan et al. [96–98] propose a crowd model using accumulated motion and foreground (moving objects) information of a crowded scene. This is implemented by two probability density functions (PDFs): Occurrence PDF and Orientation PDF associated with every non-overlapped block ($n \times n$ pixels) of the image. The Occurrence model indicates the frequency of the block covered by the foreground features, and the Orientation PDF indicates the probability of each orientation of the foreground feature on that block could take. A preliminary data mining of the PDF models is given to find the major (most frequent) path of the crowd.
  Andrade et al. [6–8] characterize crowd behaviour by observing the crowd optical flow associated with the crowd and use unsupervised feature extraction to encode normal crowd behaviour. The unsupervised feature extraction applies spectral clustering to find the optimal number of models to represent normal motion patterns. The motion models are HMMs to cope with the variable number of motion samples that might be present in each observation window. The objective of this model is to detect abnormal event in crowd scenes.
– Apart from building models, in crowd monitoring systems of computer vision, the extracted information is used to recognize the event, usually under some assumptions of involved crowds. Early work on crowd monitoring using image processing is reviewed by Davies et al. [26].
  More recent work like in Boghossian et al. [14], a system is presented using computer vision techniques to estimate the paths and directions of crowd flows in CCTV images and improve the perception of scene dynamics by offering on-line illustrations.
  Maurin et al. [68] propose a system to detect, track, and monitor both pedestrians (crowds) and vehicles. The system contains a detection scheme based on optical flow that can locate vehicles, individual pedestrians and crowd. The detection phase is followed by the tracking phase that tracks all the detected entities. Traffic objects are tracked and a rich set of descriptors are computed for each object including a wealth of information (position, velocity, acceleration/deceleration, bounding box, and shape).
  Cupillard et al. carry out event recognition by means of *behaviour*, in [24,25] an approach using multiple cameras is presented. The algorithm relies on both low level motion detection and tracking, and a high level module which recognizes pedefined scenarios corresponding to specific behaviours.
  Michael et al. [19] present a method jointly performing recognition of complex events and linking fragmented tracks. The recognition work is implemented by combing appearance and kinematic constraints from tracking and constraints from a hypothesized event model.
  In these methods specially assumptions of crowd are usually involved, indicating that some prior knowledge are required for events inference. These methods may be very efficient and computational unexpensive for some particular systems that the interested events are simple and clear, though this is not always the case in general situations.

3.2 Crowd models from non vision approach

Computational models aim at describing and predicting the collective effects of crowd behaviour by identifying the relationship between crowd features. There are three distinct philosophies for modelling a crowd; traffic analysis [30] proposes a categorisation, where crowd models can be defined as microscopic, mesoscopic and macroscopic. The microscopic model deals with pedestrians as discrete individuals; the macroscopic model deals with a crowd as a whole and the mesoscopic model combines the properties of the previous two, either keeping a crowd as a homogeneous mass but considering an internal *force* or keeping the characters of the individuals while maintaining a general view of the entire crowd (Fig. 2). In the following some typical techniques of crowd modelling will be introduced and some examples will be given.

– **Physics inspired models.** Several quantitative factors of crowds and pedestrians are measurable. This fact encourages researchers to look for the mathematical models of crowd dynamics.
  Helbing has a series of work upon this topic. His first experiment is in [34], with a stochastic formulation at microscopic level, a gas kinetic formulation at the

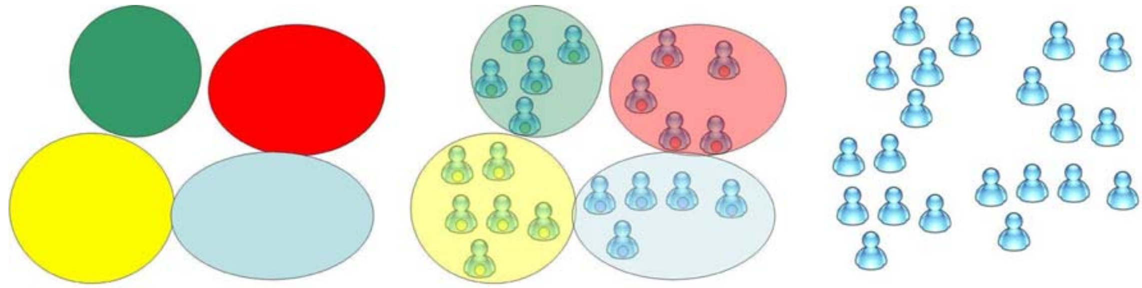

Fig. 2 *Left* Macroscopic, *centre* Mesoscopic, *right* Microscopic

mesoscopic level, and fluid dynamic equations at the macroscopic level for the crowd model. Later he [36] proposes another more popular microscopic model: social force model based on the social field theory. The social force represents the effect of the environment; it is a quantity that describes the concrete motivation to act. In [37] the model is used to reproduce the emergence of several empirically observed collective patterns of motion. Moreover, simulations of crowd dynamics based on a generalized force model for the escape panic phenomenon are presented in [35]. Also quite a few works have been developed upon this work, for example in [21] additional pattern is introduced by considering the unequal information distribution in a crowd.

In contrast to the former works, macroscopic models often draw an analogy between the crowd a continuum responding to local influence. Hughes [41] is more interested in modelling rational, goal-directed pedestrians. His theory does not govern the behaviours of any individual pedestrians, as it is a macroscopic model; instead the crowd is divided into (approximate) pedestrian types where pedestrians in each type have the same walking habits.

Physics inspired models are widely used to study crowds from different perspectives, e.g. to study the effects of introducing autonomous robots into crowds [52], or to model a historic scene [80]. The interrelations of the factors and equations (e.g. employing the same factors in different level equations) imply the possibility of having a model encompassing all the levels. Also the quantitative analysis of crowd dynamics can be relatively easy to be adapted into computer-based algorithms.

– **Agent based models.** These are qualitative models include employing fuzzy methods to describe the relations of factors and crowd motion instead of pure mathematical methods. Agent-based models use agents to represent the pedestrian or the crowd. Many examples are from the former, e.g. in [71] crowd, crowd individuals have their own emotional parameter to govern behaviour while they belong to a collection of goal-directed groups on mesoscopic level. In [73] the agents are modelled following the concept of non-adaptive behaviours. Non-adaptive crowd behaviours refer to the destructive actions that a crowd may experience in emergency situations. The human and social models are categorised into the individual, the interactions among individuals, and the group and the environment three non-independent levels (Fig. 3). Brenner et al. [15] provide an example model by assuming that people at the same location experience the same psychological and environmental influences.

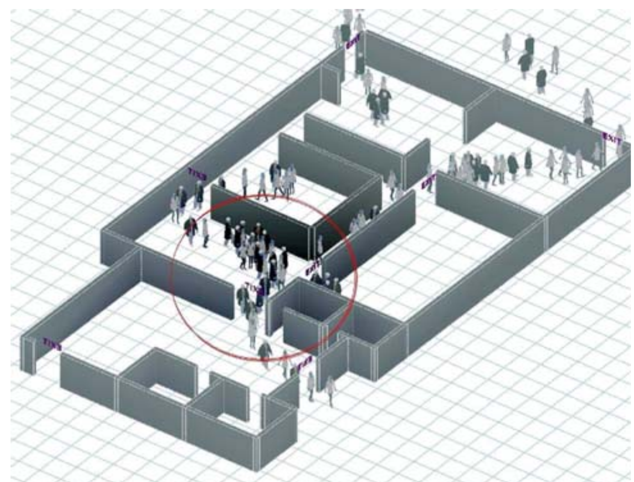

Fig. 3 A screenshot of XiaoShan Pan's work: human agents try to self-organise into exiting lines

Some work of the agent-based models have already been commercialised, such as the work of Keith Still at Crowd Dynamics Ltd [22] and LEGION international LTD [57], both provide pedestrian simulations for space design an planning, based on agent technology. For example the model developed by Crowd Dynamics Ltd aims to simulate how people react to their environment in a variety of conditions (Fig. 4).

Usually, these examples employ agent to act as individual pedestrians and only concern the microscopic level.

– **Cellular automation models.** Another research approach employs the construction of local models, where active area has been virtually devided into cells. An example is

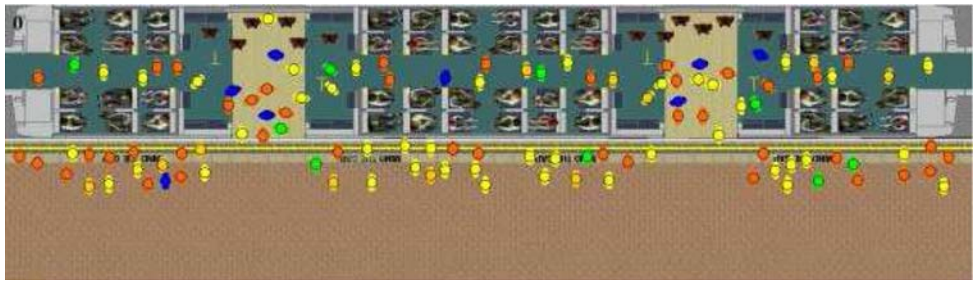

Fig. 4 Dwell analysis by Crowd Dynamics Ltd, using agent to assess the throughput of specific geometric designs

a commercialized tool EGRESS of AEA Technology Plc [4]. In EGRESS the floor area of an environment is covered with cells equivalent to the minimum occupancy area of a person. The used cells can represent free floor area, a wall or a blockage, a cell with a person, or a region with some other attributes. Pedestrians move between cells following predefined rules. Krez et al. [56] present a model of pedestrian motion using both floor field and agents. The model consists of three floor fields: *Static floor field* for each cell contains the information of the distance to the exit; *Dynamic floor field* changes by the motion of the pedestrains and the third floor fields saves the distance of a cell to the next wall.

– **Nature based models.** Some of the models take their inspiration from nature. The emotional ant model [11] extends the psychological information using biologically inspired ant agent as a crowd. Four different cognitive behaviours of crowd have been modelled and transition behaviour is modelled using fuzzy logic.
  Kirchner et al. [51] apply a bionics approach to the cellular automation model by describing the interaction between the pedestrians using ideas from chemotaxis. The simulation of the evacuation from a large room is also presented to show the ability of the model to represent different types of behaviours.

4 Examples of bridging the research

Computer simulation can be used to evaluate the developed system's performance. Considering that real visual evidences for abnormal scenarios are rare or unsafe to reproduce in a controllable way, Andrade et al. [5] have developed an approach generating simulations to allow training and validation of computer vision systems applied to crowd monitoring. The simulation is generated by a pedestrian path model and a pedestrian body model. Vu et al. [91] conceive a test framework that generats 3D animations corresponding to behaviours recognised by an interpretation system. In other words, this is a test system for a given interpretation system by generating test animations.

Non-vision models can be borrowed for computer vision analysis. Anotonini et al. [9,10] propose a framework using discrete choice model, which is widely used in traffic simulations, for pedestrian dynamics modelling. The framework models short-term behaviours of individuals as a response to the presence of other pedestrians. The model is calibrated using data from actual pedestrian movements, manually taken from video sequences. The work is applied to the problem of the target detection in the particular case of pedestrian tracking.

5 Conclusions and discussion

This paper provides a review on current crowd analysis work in computer vision. Perspectives from sociology, psychology and computer graphics are presented, as these research fields also have contributed to an in-depth study on crowd analysis and modelling. Sociological and psychological studies on the crowd phenomenon make use of human observations. Their studies indicate various ways to represent and model people relationships in isolation and as part of a more or less large group of people. The microscopic, mesoscopic and macroscopic levels are defined to characterise people as individuals part of crowd. The computer vision approach tackles the problem of automatically extracting information sufficient to characterise some special crowd events.

Anotonini gives a good example of employing non-vision model, however, his work only uses very limited information and only acts as a *clear* tracker. The works of non-vision analysis present in our paper show that all of the factors or information extracted from the real world using computer vision techniques are inter-related. Moreover, they have proposed the probable relationships in their works, which represent the human understanding of crowd dynamics. On the other hand, computer vision techniques have the ability of exploiting the special environmental constraints, which could be applied to calibrate the proposed models. We can claim that it is possible that to develop intelligent systems combining these works with computer vision approaches. The system would be capable of automatically understanding and modelling the crowd behaviours which works at both instantaneous and recurrent level.

Acknowledgments This research was partially funded by the British Telecommuncation Group PLC. The authors would like to thank the Associate Editor Dr. Hai Tao and the two reviewers for valuable comments that helped improve the clarity of presentation of this paper.

References

1. 2004-2005, 2005-2006, CIFE, Seed, Project, Stanford, University: http://eil.stanford.edu/egress/
2. Adang, O.M., Stott, C.: A European study of the interaction between police and crowds of foreign nationals considered to pose a risk to public order. http://policestudies.homestead.com/Euro2004.html
3. Advisor: http://advisor.matrasi-tls.fr/
4. AEA, Techology: a technical summary of the aea egress code. Technical Report **1** (2002)
5. Andrade, E., Fisher, R.: Simulation of crowd problems for computer vision. In: First International Workshop on Crowd Simulation, vol. 3, pp. 71–80 (2005)
6. Andrade, E., Fisher, R.: Hidden Markov models for optical flow analysis in crowds. In: Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), vol. 01, pp. 460–463. IEEE Computer Society Washington, (2006)
7. Andrade, E., Fisher, R.: Modelling crowd scenes for event detection. In: Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), vol. 01, pp. 175–178. IEEE Computer Society Washington, DC (2006)
8. Andrade, E.L., Blunsden, S., Fisher, R.B.: Performance analysis of event detection models in crowded scenes. In: Proceedings of Workshop on Towards Robust Visual Surveillance Techniques and Systems at Visual Information Engineering 2006, pp. 427–432. Bangalore, India (2006)
9. Antonini, G., Bierlaire, M., Weber, M.: Simulation of pedestrian behaviour using a discrete choice model calibrated on actual motion data. In: 4th STRC Swiss Transport Research Conference. Ascona (2004)
10. Antonini, G., Venegas, S., Thiran J.P., Bierlaire, M.: A discrete choice pedestrian behaviour model in visual tracking systems. In: Advanced Concepts for Intelligent Vision Systems, pp. 273–280. Brussels, Belgium (2004)
11. Banarjee, S., Grosan, C., Abarha, A.: Emotional ant based modeling of crowd dynamics. In: Seventh International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC'05), pp. 279–286 (2005)
12. Behave: http://www.homepages.informatics.ed.ac.uk/rbf/BEHAVE/
13. Blackman, S.: Multiple hypothesis tracking for multiple target tracking. Aerospace Electron. Syst. Mag. IEEE **19**(1), 5–18 (2004)
14. Boghossian, B., Velastin, S.: Motion-based machine vision techniques for the management of large crowds. In: The 6th IEEE International Conference on Electronics, Circuits and Systems, vol. 2 (1999)
15. Brenner, M., Wijermans, N., Nussle, T., de Boer, B.: Simulating and controlling civilian crowds in robocup rescue. In: Proceedings of RoboCup 2005: Robot Soccer World Cup IX, Osaka (2005)
16. Broggi, A., Bertozzi, M., Fascioli, A., Sechi, M.: Shape-based pedestrian detection. In: Proceedings of the IEEE Intelligent Vehicles Symposium 2000. Dearbon (MI), USA (2000)
17. Brostow, G., Cipolla, R.: Unsupervised Bayesian detection of independent motion in crowds. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 594–601. IEEE Computer Society Washington, DC, USA (2006)
18. Cai, Y., de Freitas, N., Little, J.J.: Robust visual tracking for multiple targets. In: European Conference on Computer Vision, LNCS, vol. 3954, pp. 107–118. Springer, Heidelberg (2006)
19. Chan, M.T., Hoogs, A., Bhotika, R., Perera, A., Schmiederer, J., Doretto, G.: Joint recognition of complex events and track matching. In: CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1615–1622. IEEE Computer Society, Washington, DC, USA (2006). doi:10.1109/CVPR.2006.160
20. Chang, T., Gong, S., Ong, E.: Tracking multiple people under occlusion using multiple cameras. In: British Machine Vision Conference, pp. 566–575 (2000)
21. Chu, J., Li, J., Xu, M., Zhao, L.: Simulating escape panic based on the mechanism of asymmetric information distribution. In: Complex Systems Summer School Final Project Papers. Santa Fe Institute, Santa Fe (2005).
22. Crowd, Dynamics: http://www.crowddynamics.com/
23. Crowd, MAGS: http://www2.ift.ulaval.ca/~muscamags/Dnd-crowdmags-project.htm
24. Cupillard, F., Bremond, F., Thonnat, M.: Behaviour recognition for individuals, groups of people and crowd. IEE Seminar Digests **7** (2003)
25. Cupillard, F., Bremond, F., Thonnat, M., INRIA, F.: Group behavior recognition with multiple cameras. Applications of Computer Vision, 2002 (WACV 2002). In: Proceedings of Sixth IEEE Workshop, pp. 177–183 (2002)
26. Davies, A., Yin, J., Velastin, S.: Crowd monitoring using image processing. Electron. Commun. Eng. J. **7**(1), 37–47 (1995)
27. Dong, L., Parameswaran, V., Ramesh, V., Zoghlami., I.: Fast Crowd Segmentation Using Shape Indexing. Rio de Janeiro, Brazil (2007)
28. Doucet, A., Godsill, S., Andrieu, C.: On sequential Monte Carlo sampling methods for Bayesian filtering (2000)
29. Elgammal, A., Davis, L.: Probabilistic framework for segmenting people under occlusion. In: Eighth IEEE International Conference on Computer Vision, 2001. Proceedings of ICCV 2001, vol. 2, pp. 145–152 (2001)
30. FHWA.: Traffic analysis tools primer,traffic analysis toolbox (1) (2004). http://ops.fhwa.dot.gov/trafficanalysistools/tat-vol1/index
31. Gabriel, P., Verly, J., Piater, J., Genon, A.: The state of the art in multiple object tracking under occlusion in video sequences. Advanced Concepts for Intell. Vis. Syst., pp. 166–173 (2003)
32. Han, M., Xu, W., Tao, H., Gong, Y.: An algorithm for multiple object trajectory tracking. Computer Vision and Pattern Recognition, 2004. In: CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference, vol. 1 (2004)
33. Heisele B., Woehler C. (1998) Motion-based recognition of pedestrians. In: Proceedings of Fourteenth International Conference on Pattern Recognition, 1998, vol. 2, pp. 1325–1330 (1998)
34. Helbing, D.: Models for pedestrian behavior (1992). http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/9805089
35. Helbing, D., Farkas, I., Vicsek, T.: Simulating dynamical features of escape panic. Lett. Nat. **407**, 487–490 (2000)
36. Helbing, D., Molnár, P.: Social force model for pedestrian dynamics. Phys. Rev. E **51**(5), 4282–4286 (1995)
37. Helbing, D., Molnar, P.: Self-organization phenomena in pedestrian crowds (1997). http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/9806152
38. Hu, W., Tan, T., Wang, L., Maybank, S.: A survey on visual surveillance of object motion and behaviors. IEEE Trans. Syst., Man Cybernet. C Appl. Rev. **34**(3), 334–352 (2004)

39. Huang, C., Ai, H., Li, Y., Lao, S.: Vector boosting for rotation invariant multi-view face detection. In: Tenth IEEE International Conference on Computer Vision, vol. 1, pp. 446–453 (2005)
40. Huang X., Li L., Sim T. (2004) Stereo-based human head detection from crowd scenes. In: International Conference on Image Processing, 2004. ICIP'04, Vol. 2, pp. 1353–1356
41. Hughes, R.: A continuum theory for the flow of pedestrians. Trans. Res. B Methodol. **36**(6), 507–535 (2002)
42. Inria: http://www.inria.fr/rapportsactivite/RA2005/orion/uid1.html
43. Isard, M., Blake, A.: A mixed-state CONDENSATION tracker with automatic model-switching. In: IEEE International Conference on Computer Vision, pp. 107–112 (1998). http://citeseer.ist.psu.edu/isard98mixedstate.html
44. ISCAPS: http://www.iscaps.reading.ac.uk/home.htm
45. Jones, M., Viola, P.: Fast multi-view face detection. Mitsubishi Electric Research Lab TR-20003-96 (2003)
46. Kang, H., Kim, D., Bang, S.: Real-time multiple people tracking using competitive condensation. Proc. Int. Conf. Pattern Recogn. **1**, 413–416 (2002)
47. Karlsson, R., Gustafsson, F.: Monte Carlo data association for multiple target tracking. Target Tracking: Algorithms and Applications (Ref. No. 2001/174), IEE, vol. 1 (2001)
48. Khan, S.M., Shah, M.: A multiview approach to tracking people in crowded scenes using a planar homography constraint. In: 9th European Conference on Computer Vision, LNCS, vol. 3954, pp. 133–146. Springer, Heidelberg (2006)
49. Khan, Z., Balch, T., Dellaert, F.: MCMC-based particle filtering for tracking a variable number of interacting targets. IEEE Trans. Pattern Anal. Mach. Intell. **27**(11), 1805–1819 (2005)
50. Kim, K., Davis, L.S.: Multi-camera tracking and segmentation of occluded people on ground plane using search-guided particle filtering. In: European Conference on Computer Vision, LNCS, vol. 3953, pp. 98–109. Springer, Heidelberg (2006)
51. Kirchner, A., Schadschneider, A.: Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics. Phys. A Stat. Mech. Appl. **312**(1–2), 260–276 (2002)
52. Kirkland, J., Maciejewski, A.: A simulation of attempts to influence crowd dynamics. IEEE Int. Conf. Syst. Man Cybernet. 4328–4333 (2003)
53. Koller-Meier, E., Ade, F.: Tracking multiple objects using the Condensation algorithm. Robot. Auton. Syst. **34**(2-3), 93–105 (2001)
54. Kong, D., Gray, D., Tao, H.: Counting Pedestrians in crowds using viewpoint invariant training. In: British Machine Vision Conference (2005)
55. Kong, D., Gray, D., Tao, H.: A viewpoint invariant approach for crowd counting. In: Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), vol. 03, pp. 1187–1190 (2006)
56. Kretz, T., Schreckenberg, M.: F.a.s.t. — floor field — and agent-based simulation tool (2006)
57. Legion: http://www.legion.biz/about/index.html
58. Leibe, B., Seemann, E., Schiele, B.: Pedestrian detection in crowded scenes. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. CVPR 2005. **1** (2005)
59. Li, S.Z., Zhu, L., Zhang, Z., Blake, A., Zhang, H., Shum, H.: Statistical learning of multi-view face detection. In: Proceedings of the 7th European Conference on Computer Vision-Part IV, pp. 67–81. Springer, London (2002)
60. Lin, S., Chen, J., Chao, H.: Estimation of number of people in crowded scenes using perspective transformation. IEEE Trans. Syst. Man Cybernet. A, **31**(6), 645–654 (2001)
61. Ma, R., Li, L., Huang, W., Tian, Q.: On pixel count based crowd density estimation for visual surveillance. IEEE Conf. Cybernet. Intell. Syst. **1** (2004)
62. Marana, A., da Costa, L., Lotufo, R., Velastin, S.: On the efficacy of texture analysis for crowd monitoring. In: Proceedings of the International Symposium on Computer Graphics, Image Processing, and Vision, vol. 00, p. 354 (1998)
63. Marana, A., Da Fontoura Costa, L., Lotufo, R., Velastin, S.: Estimating crowd density with Minkowski fractal dimension. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. ICASSP'99. vol. 6, 3521–3524 (1999)
64. Marana, A., Velastin, S., Costa, L., Lotufo, R.: Estimation of crowd density using image processing. In: IEE Colloquium on Image Processing for Security Applications (Digest No: 1997/074), p. 11 (1997)
65. Marana, A., Velastin, S., Costa, L., Lotufo, R.: Automatic estimation of crowd density using texture. Safety Sci. **28**(3), 165–175 (1998)
66. Marques, J., Jorge, P., Abrantes, A., Lemos, J.: Tracking groups of Pedestrians in video sequences. IEEE 2003 Conf. Comput. Vis. Pattern Recogn. Workshop, **9**, 101 (2003)
67. Mathes, T., Piater, J.: Robust non-rigid object tracking using point distribution models. Bri. Mach. Vis. Conf. **2** (2005)
68. Maurin, B., Masoud, O., Papanikolopoulos, N.: Monitoring crowded traffic scenes. In: Proceedings of The IEEE 5th International Conference on Intelligent Transportation Systems, 2002. pp. 19–24 (2002)
69. McKenna, S., Jabri, S., Duric, Z., Rosenfeld, A., Wechsler, H.: Tracking groups of people. Comput. Vis. Image Understanding **80**(1), 42–56 (2000)
70. Mittal, A., Davis, L.: M 2 tracker: a multi-view approach to segmenting and tracking people in a cluttered scene. Int. J. Comput. Vis. **51**(3), 189–203 (2003)
71. Musse, S., Thalmann, D.: A model of human crowd behavior: group inter-relationship and collision detection analysis. Proc. Workshop Comput. Anim. Simul. Eurograph. **97**, 39–51 (1997)
72. Okuma, K., Taleghani, A., de Freitas, N., Little, J., Lowe, D.: A boosted particle filter: Multitarget detection and tracking. Eur. Conf. Comput. Vis. **1**, 28–39 (2004)
73. Pan, X., Han, C., Dauber, K., Law, K.: Human and social behavior in computational modeling and analysis of egress. Automation in Construction **15**(4), 448–461 (2006)
74. Polus, A., Schofer, J., Ushpiz, A.: Pedestrian Flow and Level of Service. J. Transportation Eng. **109**(1), 46–56 (1983)
75. Prismatica: http://prismatica.king.ac.uk/
76. Rahmalan, H., Nixon, M., Carter, J.: On crowd density estimation for surveillance. The Institution of Engineering and Technology Conference on Crime and Security, pp. 540C–545C (2006)
77. Rasmussen, C., Hager, G.: Joint probabilistic techniques for tracking multi-part objects. Computer Vision and Pattern Recognition, 1998. In: Proceedings of 1998 IEEE Computer Society Conference, pp. 16–21 (1998)
78. Reid, D.: An algorithm for tracking multiple targets. Automat. Contr. IEEE Trans. **24**(6), 843–854 (1979)
79. Reisman, P., Mano, O., Avidan, S., Shashua, A., Ltd, M., Jerusalem, I.: Crowd detection in video sequences. Intell. Vehicles Symp. IEEE, 66–71 (2004)
80. R.R., C., Hughes, R.: Mathematical modelling of a mediaeval battle: the battle of agincourt. Math. Compute. Simul. **64(2)**, 259–269 (2004)
81. Shashua, A., Gdalyahu, Y., Hayun, G.: Pedestrian detection for driving assistance systems: single-frame classification and system level performance. Intelligent Vehicles Symposium, 2004 IEEE pp. 1–6 (2004)

82. Sidenbladh, H., Wirkander, S.: Tracking random sets of vehicles in terrain. In: Proceedings of 2003 IEEE Workshop on Multi-Object Tracking vol. 9, 98 (2003)
83. Siebel, N., Maybank, S.: Fusion of multiple tracking algorithms for robust people tracking. In: European Conference on Computer Vision, pp. 373–387 (2002)
84. Smith, K., Gatica-Perez, D., Odobez, J.: Using particles to track varying numbers of interacting people. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1–01, pp. 962–969 (2005)
85. Spengler, M., Schiele, B.: Towards robust multi-cue integration for visual tracking. Mach. Vis. Appl. **14**(1), 50–58 (2003)
86. Sullivan, J., Carlsson, S.: Tracking and labelling of interacting multiple targets. In: European Conference on Computer Vision, LNCS, vol. 3953, pp. 619–632. Springer, Heidelberg (2006)
87. Swets, D., Punch, B.: Genetic algorithms for object localization in a complex scene. In: IEEE International Conference on Image Processing, pp. 595–598 (1995)
88. US Census Bureau: http://www.census.gov/ipc/www/worldpop.html
89. Velastin, S., Yin, J., Davies, A., Vicencio-Silva, M., Allsop, R., Penn, A.: Automated measurement of crowd density and motion using imageprocessing. Road traffic monitoring and control, 1994. In: Seventh International Conference, pp. 127–132 (1994)
90. Venegas, S., Knebel, S., Thiran, J.: Multi-object tracking using particle filter algorithm on the top-view plan. Technical report,LTS-REPORT-2004-003, EPFL (2004). http://infoscience.epfl.ch/getfile.py?mode=best&recid=87041
91. Vu, V., Bremond, F., Thonnat, M.: Human behaviour visualisation and simulation for automatic video understanding. In: Proceedings of the 10th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG–2002), Plzen–Bory, Czech Republic, pp. 485–492 (2002)
92. Wu, B., Nevatia, R.: Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors. In: Tenth IEEE International Conference on Computer Vision, 2005. ICCV 2005, vol. 1, 90–97 (2005)
93. Wu, B., Nevatia, R.: Tracking of multiple, partially occluded humans based on static body part detection. In: CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 951–958 (2006)
94. Yang, D., Gonzalez-Banos, H., Guibas, L.: Counting people in crowds with a real-time network of simple image sensors. In: Proceedings of Ninth IEEE International Conference on Computer Vision, 2003, pp. 122–129 (2003)
95. Yin, J., Velastin, S., Davies, A.: Image Processing Techniques for Crowd Density Estimation Using a Reference Image. Proc. 2nd Asia-Pacific Conf. Comput. Vis. **3**, 6–10 (1995)
96. Zhan, B., Remagnino, P., Velastin, S.: Analysing Crowd Intelligence. Second AIxIA Workshop on Ambient Intelligence (2005)
97. Zhan, B., Remagnino, P., Velastin, S.: Mining paths of complex crowd scenes. Lecture Notes in Computer Science pp. 126–133 (2005). ISBN/ISSN 3-540-30750-8
98. Zhan, B., Remagnino, P., Velastin, S.: Visual analysis of crowded pedestrain scenes. XLIII Congresso Annuale AICA, pp. 549–555 (2005)
99. Zhao, T., Nevatia, R.: Tracking multiple humans in complex situations. Pattern Anal. Mach. Intell. IEEE Trans. **26**(9), 1208–1221 (2004)
100. Zhao, T., Nevatia, R.: Tracking multiple humans in crowded environment. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition vol. 2, II–406–II–413 (2004)

Author biographies

Beibei Zhan received the B.Eng. degree in Computer Science and Technology from Xi'an Jiaotong University, China, in 2003, and M.Sc. degree in Vision Imaging and Virtual Environments from University College London,UK, in 2004. Currently she is a Ph.D. student at the Digital Imaging Research Centre at Kingston University, UK. Her current research interests include vision based behaviour analysis, machine learning, ambient intelligence and visual surveillance.

Dorothy N. Monekosso (PhD, 1999) is a Senior Lecturer in the Faculty of Computing, Information Science and Mathematics at Kingston University. Prior to joining Kingston University, Dr. Monekosso worked for over 10 years in the field of spacecraft engineering and spacecraft autonomy. Her current research interests include machine learning, intelligent systems, and robotics.

Paolo Remagnino (1988, Laurea Diploma and 1993, PhD) is a Reader in the Faulty of Computing, Information Systems and Mathematics at Kingston University. Dr. Remagnino is one of the main promoters of Ambient Intelligence and a research in Video Surveillance. Dr. Remagnino is a member of the editorial board of the Machine Vision and Applications and the Expert Systems journals and an Associate Editor of the International Journal of Robotics and Automation. His research interests include machine vision, learning and intelligence.

Sergio A. Velastin received the B.Sc. degree in Electronics, M.Sc. (Research) degree in Digital Image Processing and the Ph.D. from the University of Manchester, UK, in 1978, 1979 and 1982, respectively. Currently he is a Reader (Associate Professor) at the Digital Imaging Research Centre at Kingston University UK and also its Director. His research interests include computer vision for crowd and pedestrian monitoring and personal security as well as distributed visual surveillance systems. Dr. Velastin is a member of the IEEE and the British Machine Vision Association (BMVA).

Li-Qun Xu is a Principal Researcher in Visual computing and Multimedia Research of BT Group Research and Venturing Division. He joined BT in 1996 and has since managed and initiated a broad range of research projects in BT's drive for excellence and business growth in managed networked ICT services. His technical expertise includes dynamic visual scene understanding, video analytics for security and safety, human-centred ambient intelligence, and multimedia indexing and retrieval. He holds over 16 European and international patents and pending applications. He serves as independent expert for EU Commission Services for the last five years, participating in project review and proposal evaluation in the diverse areas of networked media, intelligent content and security monitoring. He is an executive team member of IET visual information engineering professional network and expert group member of Transport for London (TfL) Realtime Integration Programme. Prior to his joining BT he worked in Chinese Academy of Sciences and several UK Universities as research and faculty member for 8 years.