Article

Social Science Computer Review
1-22

DOI: 10.1177/0894439317726510
journals.sagepub.com/home/ssc
SAGE

Automated Solutions for Crowd Size Estimation

Muhammad Waqar Aziz[1,2], Farhan Naeem[3], Muhammad Hamad Alizai[4], and Khan Bahadar Khan[2]

Abstract

The crowd phenomenon frequently occurs in dense urban living environments. Crowd counting or estimation helps to develop management strategies such as designing safe public places and evacuation plan for emergencies. These strategies are different depending upon the type of event such as political and public demonstrations, sports, and religious events. However, estimating the number of people in crowds at closed or open environments is quite challenging because of the dynamics involved in the process. In addition, crowd estimation itself poses challenges due to randomness in crowd behavior, motion, and an area's geometric specifications. Crowd behavior as well as the area parameters is studied before suggesting any possible technological solution for managing a crowd. This article presents a theoretical understanding of the major crowd size estimation approaches that cannot be achieved through the study of existing survey papers in this area, because the existing survey papers focus on particular technologies/specific areas with no or brief description of the involved steps. Besides, this article also highlights the strength and weakness of crowd size estimation solutions and their possible applications. It is, therefore, believed that the provided information would assist in developing an intelligent system for crowd management.

Keywords

crowd size estimation, crowd density, received signal strength indication, crowd management

Introduction

The steady increase in population in urban living environments has made the occurrence of the crowd a common phenomenon. Crowd gathering can instantly turn to be disastrous, which may lead to deadly accidents (Yogameena & Nagananthini, 2017). Thus, efficient crowd management and control is pivotal to human safety (Dickie, 1995). Administering crowd is not trivial (Chen, Liang, Lee, & Xu, 2007), since all aspects of the occurrence of crowd have to be considered, for example,

[1] Umm Al-Qura University, Makkah, Saudi Arabia
[2] CECOS University of IT and Emerging Sciences, Peshawar, Pakistan
[3] University of Engineering and Technology, Peshawar, Pakistan
[4] Lahore University of Management Sciences (LUMS), Lahore, Pakistan

Corresponding Author:
Muhammad Waqar Aziz, Science and Technology Unit, Umm Al-Qura University, Makkah 21995, Saudi Arabia.
Email: mwabdulaziz@uqu.edu.sa

Table 1. Summary of Existing Survey Papers in Recent Years.

Authors (Year)	Paper Title	Aspects Covered
Yogameena and Nagananthini (2017)	Computer vision–based crowd disaster avoidance system: A survey	Crowd disaster analysis
Zitouni, Bhaskar, Dias, and Al-Mualla (2016)	Advances and trends in visual crowd analysis: A systematic survey and evaluation of crowd modeling techniques	Crowd modeling techniques
Henke (2016)	Estimating crowd size: A multidisciplinary review and framework for analysis	Multidisciplinary crowd estimation techniques
Saleh, Suandi, and Ibrahim (2015)	Recent survey on crowd density estimation and counting for visual surveillance	Computer vision–based crowd density estimation and counting methods for visual surveillance
Li, Chang, Wang, Hong, and Yan (2015)	Crowded scene analysis: A survey	Crowded scene analysis in automated video surveillance
Ryan, Denman, Sridharan, and Fookes (2015)	An evaluation of crowd counting methods, features and regression models	Various image features and regression models; evaluation across multiple data sets

crowd type (static, dynamic), *event type* (sports, theatrical, rally, concert, and parade), *crowd behavior*, and *environment type* (open, closed). Conventional methods of crowd management mostly involved human interventions which have been often unsuccessful (or at least unpleasant to the crowd). Automated systems have been developed to avoid direct human intervention. Based on the analysis of the crowd behavior and the area where crowd occurrence is frequent, these systems can build smart environments to estimate crowd size automatically. This information can further be used to plan and administer a crowd and to ensure a safe environment.

A key factor in crowd management technology is crowd behavior. The most important aspects of crowd behavior such as movement, location, and density are extracted using automated techniques and decisions are then based on it. These aspects of crowd behavior are further analyzed for designing public usage spaces, automatically detecting disturbances, and activating proper strategy for handling the situation. Beside essential for human safety, these analyses can also be used in the development of smart cities (van Waart, Mulder, & de Bont, 2016). For safety control, automated estimation of crowd density and counting is receiving much attention (Saleh, Suandi, & Ibrahim, 2015). Automated solutions such as the ones based on closed-circuit television (CCTV; Regazzoni & Tesei, 1996) exist for real-time crowd density estimation. These systems employ several image processing techniques to estimate the crowd size, detect behaviors, and track and recognize individuals. For example, foreground/background separation and edge detection are used to estimate the area occupied by individuals. This article reviews the major crowd size or density estimation (simply refer to as *crowd estimation* throughout the article) methods.

Additionally, intelligent crowd management system (ICMS) can be developed for managing crowds, which could automatically estimate crowd size and then provide assistance to the crowd/managers. The requirements for setting up such a system must be carefully drawn by considering all aspects of a crowd. Hence, selecting the right solution for a particular set of requirements is pivotal in developing an efficient ICMS. It is therefore necessary to thoroughly understand the available automated crowd estimation solutions, before selecting a particular solution for developing an ICMS. Unfortunately, the existing survey papers in the area of crowd behavior analysis (see Table 1) are unable to provide such an understanding. Although these attempts provide reviews of the technical advancements in the field, they cover different perspectives. These reviews are either based on a

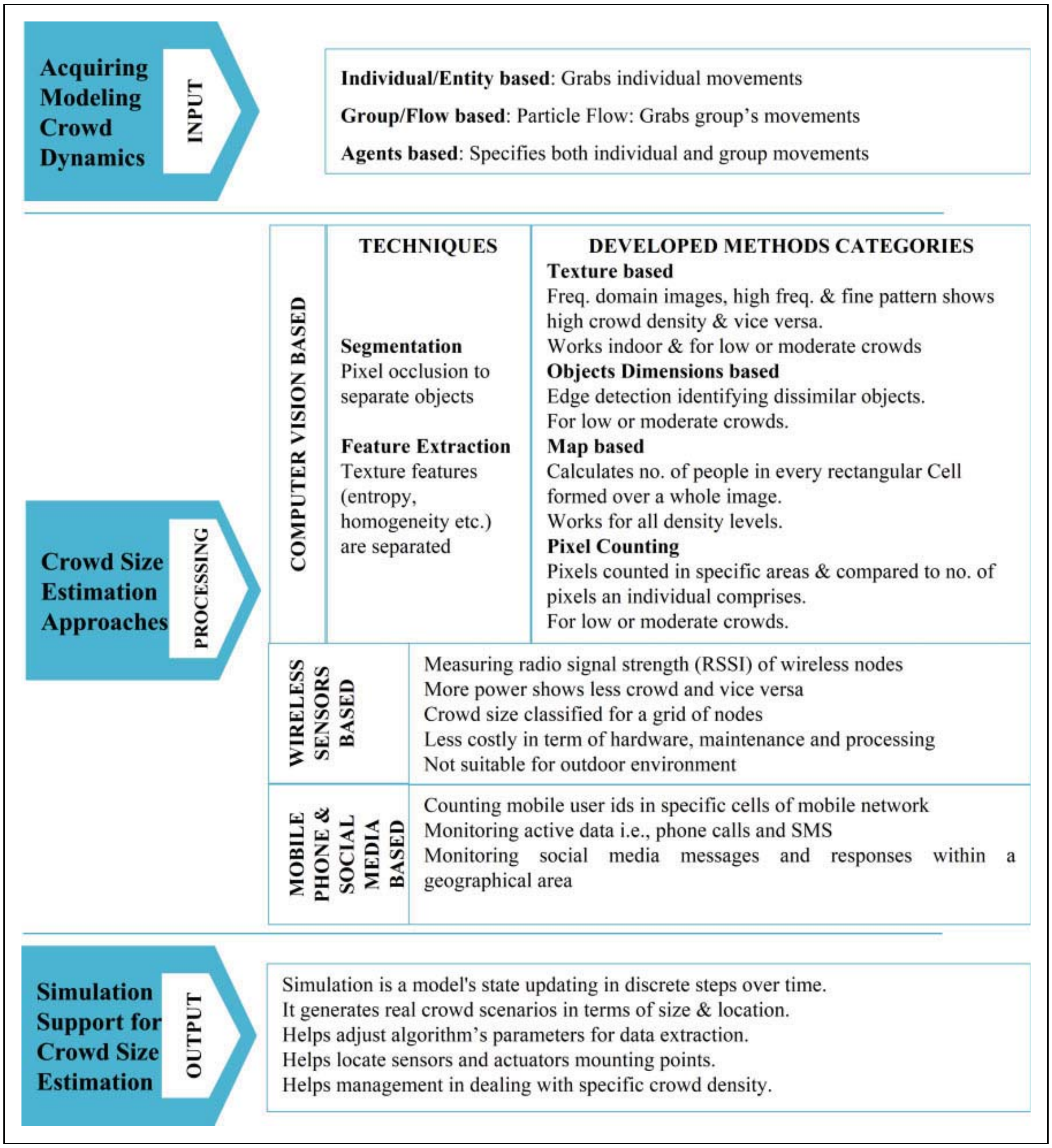

Figure 1. Schematic representation of the topics covered in this review.

particular technology, such as computer vision (Saleh et al., 2015; Yogameena & Nagananthini, 2017), or cover a specifically focused area within that technology, for example, crowd modeling techniques (Zitouni, Bhaskar, Dias, & Al-Mualla, 2016). Furthermore, the description of the methods/techniques used and the steps involved in them is rather brief.

As summarized in Figure 1, this article provides insight into the following major crowd size estimation solutions: (a) computer vision–based methods, (b) wireless sensor networks (WSNs)-based methods, and (c) mobile phones/Internet data-based methods. These solutions are selected as per the study of existing literature, based on the fundamental technologies available for crowd estimation. Although the multidisciplinary review of crowd estimation methods (Henke, 2016) can be considered close to this work, it doesn't provide theoretical understanding of the reviewed methods. On the other hand, the aim of this article is to provide a one-stop reference point that offers a theoretical understanding of the above-mentioned crowd estimation technologies. To this end, this article provides the detail of the techniques used in these solutions and the steps involved in them. Moreover, the article also discusses the strengths and weaknesses of these solutions in different kinds of environments and scenarios. The article also provides a review of the available simulation support for crowd estimation. It is anticipated that this knowledge would assist the developers in selecting appropriate solutions for developing an ICMS.

This article is organized into six sections (including the Introduction section): The second section explains crowd size estimation, key requirements of size estimation, and the workflow of automated system for crowd size estimation. The third section presents detail description of different crowd estimation approaches that include computer vision–based approaches (Computer Vision–Based Approach subsection), WSNs-based approaches (WSNs-Based Approach subsection), and approaches based on mobile phones/Internet data (Mobile Phones and Internet Data subsection). The fourth section discusses simulation support for crowd estimation, whereas the discussion is provided in the fifth section. Finally, the sixth section concludes the article.

Crowd Size Estimation

Crowd estimation refers to a technique used to count number of people in a crowd. This counting has become a key requirement in managing urban environment. Counting becomes difficult as the crowd becomes bigger and denser because of overlapping among people, which makes only part of their bodies appear. Moreover, people are not always stationary, categorized into lines, or have the same behavior. This section provides a review of the requirements and challenges of a crowd estimation system.

Crowd Behavior and Characteristics

Crowd types and characteristics are defined by its behavior which is a key factor that distinguishes different crowds depending upon the purpose of gathering, for example, protests, sports, or religious events have specific behavior. Many of the behaviors (Chen et al., 2007; Cheng, Chen, & Fang, 2013) observed so far are clear, but some still remain unclassified (Thalmann, 2007). Crowd analysis focuses on extracting information from the crowd such as *angry crowd*, *level of crowd comfort*, and *crowd density*, which could be used to take appropriate actions in managing crowds, for example, giving automated directions for emergency exit in case of fire threats. The crowd is diverse in nature: it can be static, dynamic, structured, and unstructured. A number of computational algorithms have been developed to analyze different types of crowds efficiently.

Assisting and handling a crowd is based on the amount of the information we have about the crowd and the environment. Analyzing crowds to extract information is achieved automatically through methods such as computer vision and WSNs. Computer vision uses video footage/still images, whereas WSNs employ radio signal strength and link quality indication for estimating crowd. Meaningful features that help in estimating individuals are extracted using gray-level dependence matrix (GLDM) method (Haralick, 1979). Consequently, crowds are classified using intelligent classifiers based on neural network and other probabilistic algorithms. A computational model is then developed (Zhan, Monekosso, Remagnino, Velastin, & Xu, 2008) which analyzes the extracted information and shows the status of the crowd. The most common crowd characteristics are summarized in Table 2.

Crowd Density

Crowds are categorized based on the density of the people that could be low, medium, dense, and jammed (Polus, Schofer, & Ushpiz, 1983) in a specified area. Estimating crowd density by counting

Table 2. Common Crowd Characteristics.

Characteristics	Categories
Density	Low, medium, dense, and jammed
Behavior (associated with event)	Normal, angry, and violent
Sequence	Structured and unstructured
Mobility	Static and dynamic

heads in a crowd is not always feasible (Watson & Yip, 2011). Crowd size categories are based on density metric defined as the number of people per unit area. The area under study is divided into smaller subareas in order to know how far apart the people are standing. At arm's length, one person occupies an area of .93 m^2; at close distance, it is .42 m^2; and in tightly packed situations, it is .23 m^2.

$$\text{Number of people} = \text{Area} \times \text{Density}. \tag{1}$$

An acceptable estimate of the actual number of people is obtained, as well as standard error. Area and density are assumed to be approximately independent; hence, delta rule (Hevia, 2008) is used for representing *Relative Standard Error*:

$$\frac{se(\hat{N})}{\hat{N}} = \sqrt{\frac{se^2(\hat{A})}{A^2} + \frac{se^2(\hat{D})}{D^2}}, \tag{2}$$

where se is the standard error, A is area, D is crowd density, and N is the number of people. Although the tools for estimating crowd density have improved, this main principle remains the same.

Modeling Crowd Dynamics

Public areas, whether closed or open, have some repeatedly occurring dynamics for a particular type of event. Crowd information can be extracted efficiently using such dynamics that represent the status of the event. The crowd size estimation system requires these crowd dynamics to be modeled. Crowd models are designed based on this status to develop crowd management mechanisms. Figure 2 shows the workflow of a modeled system. A model's computations describe and predict the collective effects of mass behavior by identifying the relationship between the characteristics of crowds (Mehran, Oyama, & Shah, 2009; Sjarif, Shamsuddin, Hashim, & Yuhaniz, 2011). These models are designed based on the following approaches:

1. *Individual* or *entity-based* model describes individuals in a crowd. The movement area of individuals is divided into smaller areas known as cells, thus making a grid. Movements in these cells are automated through cellular automation method. This model has proven to be good for low and medium scale crowds but performs poorly in dense crowd situations (Mehran et al., 2009). Furthermore, it lacks the property of producing real human animations in simulations (Pelechano, Allbeck, & Badler, 2008). Helbing and Molnar (1995) use this method to model a system for pedestrian dynamics. This modeling mechanism follows microscopic methodology (Antonini, Bierlaire, & Weber, 2006; Papadimitriou, Yannis, & Golias, 2009; Robin, Antonini, Bierlaire, & Cruz, 2009)—*describing individuals in the crowd*.
2. *Group* or *flow-based* approach models a crowd as a continuous flow of a group. Although designed to simulate large and dense gathering, it lacks the ability to grab individual movement (Pelechano et al., 2008). Particle flow is the common method used to model flow-based crowd movements. This type of modeling is mainly used in emergency evacuations and

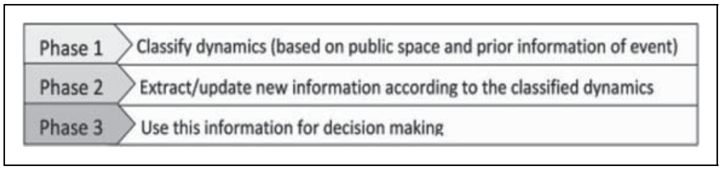

Figure 2. Workflow of a modeled system.

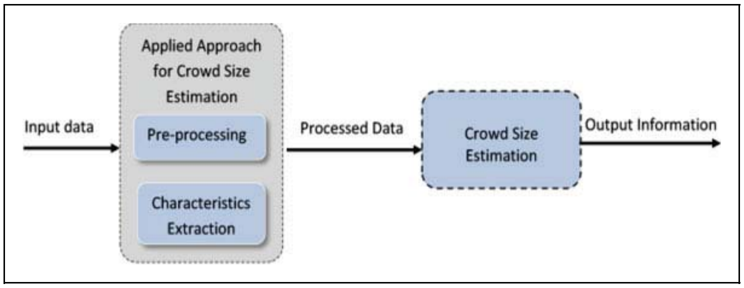

Figure 3. Crowd size estimation: Automated system.

follows macroscopic, *dealing with a whole crowd* (Hughes, 2002), and mesoscopic, *combining microscopic and macroscopic* (Huang, Wong, Zhang, Shu, & Lam, 2009), methodologies.

3. *Agents-based* crowd model is also seen as a multi-agent system (Macal & North, 2009), which uses its multiple intelligent agents (IAs; i.e., sensors and actuators) to perceive and react to an environment change (Šalamon, 2011). Goals are assigned to IAs by the rules placed in the system based on observations and prior knowledge of the scenario for which they are designed. Learning models are more efficient due to the capability of IAs to learn and use knowledge to achieve their assigned goals. These models are then used for extracting new information. This type of model works well for low and medium crowds but is inefficient for dense and large crowds. However, agent-based models have recently been improved using systems like high-density autonomous crowd (HiDAC; Pelechano, Allbeck, & Badler, 2007) and multi-agent communication for evacuation simulation (MACES; Hung, Tam, Lam, & Aberer, 2013).

Automated Systems for Crowd Size Estimation

An automated system reduces or removes the need of human interaction in a process. It speeds up a process, reduce human labor costs, and eliminate human error. Automated systems are developed to handle jobs that are difficult for humans. Automated systems have some limitation as well such as the high initial cost of the system, limited level of intelligence for certain scenarios, and research and development cost from time to time. An overview of the automated crowd size estimation system is depicted in Figure 3. It consists of two main functional parts, that is, applied technique to extract required information from the received data and classification of crowd size based on the processed data. Different testing units and their functionalities include:

- *Applied approach* is a functional unit where a particular technique such as computer vision or WSNs for crowd size estimation is applied. It consists of:
  1. *Preprocessing* removes noise and prepares data for processing. In vision-based approach, the unwanted pixel values that distort the image or video frame is removed, as explained

in Preprocessing subsection. While in WSNs-based approach, the received radio signal may also have unwanted signal values added to it due to interference from the environment, which is filtered out.

2. *Characteristics extraction* is used to extract meaningful features to be used for counting individuals. In vision-based approach, it can be separating background from individuals or recognizing individuals in multiple video frames, as explained in Characteristics Extraction subsection. In WSNs-based approach, it is to extract signal strength and wireless link quality of the received signal to use it for estimation.

- *Crowd size estimation* is used to classify the number of individuals into categories and density levels. It separates sets of different objects into groups. Processed data of certain selected features for crowd estimation are given to the next level of the system. Classifier's algorithms count individuals, assign proper density levels, and activate corresponding responses to the classified situations. The density levels and the responses are set in the classifier's memory during modeling of the system.

Crowd Size Estimation Approaches

Vision is the first choice that comes up naturally when counting people, so most of the development is done in vision-based crowd size estimation. CCTV is the most popular and most developed method seen and known. However, wireless sensor-based techniques have also been employed that use wireless signal strength measurements to estimate the crowd sizes. Recently, mobile phone and social media data have been used to estimate the number of people in a specific area. In the following subsection, all these approaches are revisited.

Computer Vision–Based Approach

Computer vision has been the first choice for scalable projects to analyze dynamic crowd activity using statistical methods and mathematical algorithms applied to the video feed. Research dates back to early 90s since developments in image processing on understanding and managing crowd using modern technologies (Davies, Yin, & Velastin, 1995). The research has mainly been focused on segmenting, tracking, and analyzing individual human such as detecting head, arms, and legs to understand gestures and human behavior. However, dealing with crowd poses different sets of challenges. Although a straightforward extension of these techniques seems suitable for crowded scenes, it is difficult to segment individuals in a crowd using computational algorithms due to occlusions of indefinite kinds. Traditional mathematical methods to detect objects or humans lose their accuracy when applied to crowded footage. Therefore, it makes much more sense to use the crowd as a single entity.

Crowd monitoring and management in real time through CCTV is a typical system employed worldwide. However, it requires human observers for continuous monitoring of connected array of television monitors. It is a tedious job and observers tend to lose concentration. Therefore, solutions have been proposed to monitor crowds automatically. Crowd models are developed representing behavior and scenarios through image processing methods by analyzing video data. The generalized view of a computer vision–based crowd size estimation techniques is presented in Figure 4. These techniques are explained below.

Preprocessing. This stage is used to prepare the data into such a form that is ready to be used by the characteristic extraction algorithms. The involved steps are:

1. *Noise removal*: Digital images are vulnerable to noise, that is, unwanted captured information which is the result of errors in the image capturing process. This noise is usually

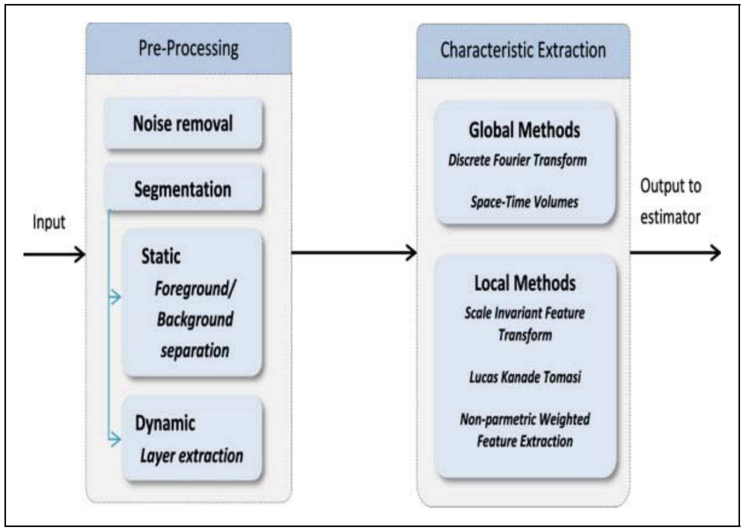

Figure 4. Summary of vision-based techniques.

unwanted pixel values that do not reflect the actual pixel values of the image. Filtering noise is the solution to the problem. Different filters based on the noise type are used to filter out unwanted pixel values.

2. *Segmentation*: Segmenting an image or video frames is the first step to categorize objects. Foreground detection is used to separate areas of interest, such as people, objects, vehicles, text, from the background using a reference background image or model. In surveillance videos, frame of interest is subtracted from reference frame to detect objects. It is the most common procedure to detect moving objects in videos. For example, text on the vehicle identification plate is identified and entered into the database of a car park using video footage. One weakness of this method is to have a reference image for the complete area of interest which may not be possible in a single image or video frame. Segmentation is further categorized as follows:
   (a) Static
   A crowd that is relatively stationary from a single vantage point (preferably from a camera perspective) is considered static. Sporting events, musical concerts, political protests, and celebrations of religious events are all examples of static crowd. In some pre-organized events in proper locations such as sports in stadiums or concerts, it is easy to estimate or count number of people accurately due to controlled entrance of the crowd through marked entrances.

   Foreground/background separation: A still image of the uncrowded area is used as the *reference image* to be compared with the crowded area to estimate crowd density. Ranging from simple to complex, a number of background subtraction techniques are discussed in Piccardi (2004). In Davies, Yin, and Velastin (1995), the input image is subtracted from reference image separating people from the background. Afterward, elements in the new image are computed by applying edge detection. This clearly outlines different objects that are separated by masking, that is, *separating different objects identified by pixel occlusion*. A more advanced solution (Hu, Zheng, Wang, & Li, 2013) uses the same basic principle but add an adaptive classifier. This classifier is trained by input samples in the training phase, thus improved background models are used for segmenting crowd in testing phase. Characteristics are extracted using the weighted area concept by giving importance to certain areas of interest. Individuals are separated and then counted by a linear regression method. Background separation is a simple mechanism and thus is less expensive in terms of computational and storage requirement. However, its simplicity causes incorrect categorization of pixels requiring frequent background updates.

(b) Dynamic

A crowd that is in motion from a reference point is known as dynamic. This motion may be organized and regular, such as entry into an event area, and unorganized, such as protests heading toward a protest area or disorderly crowd that has intent toward trouble and violence known as a *mob*. In all such impromptu celebrations, religious events, and political protests, it becomes very difficult to count since there are no marked ingress and egress points. Such crowd is dealt with using layer extraction techniques.

Layer extraction: Segmentation of scenes into regions of similar structure from video frames is performed by a compelling approach of layer extraction. Foreground is detected and separated using layer extraction. This mechanism of extracting layers from a sequence of images is widely used by many researchers (Jojic & Frey, 2001; Kumar, Torr, & Zisserman, 2008; Šalamon, 2011; Torr, Szeliski, & Anandan, 2001; Wang & Adelson, 1993). Layer extraction is commonly used for motion segmentation, that is, *separating the static background from moving objects*. Two steps are followed in motion segmentation:

i. Description of layers—number of layers and their motion parameters.
ii. Pixel assignment—discarded and assigned pixels of an image sequence to described layers.

From multiple images, the invariant values of pixels are considered as a background layer. A moving object will have changed pixel values from frame to frame which defines another layer. Both the invariant and variant values are used to segment out the background and moving objects. The moving objects are also categorized into layers.

Characteristics extraction. Characteristics detection, extraction, and representation play a vital role in an efficient recognition system. A characteristic or feature is a part of total information which is related to a certain computational function for doing a particular job (Kumari & Mitra, 2011). Avoiding the processing of foreground separation, characteristics of image or video are directly calculated based on observations of the crowded areas that highlight certain textures—information related to the spatial arrangement of intensities and color in an image (Scovanner, Ali, & Shah, 2007). Texture characteristics (e.g., light intensity, contrast, entropy, homogeneity; Piccardi, 2004) are high in dense crowds and vice versa. Earlier solutions (Pelechano et al., 2007, 2008) for detecting required features of a crowd performed poorly due to perspective distortion. It occurs due to the distances between near and far objects in an image. Far objects appear smaller and their pixels are not distinguishable by texture methods. Piccardi (2004) solves this problem by using geometric correction applied to all foreground pixels equally with some derived scales in analysis of geometric correction.

More recently, Mariano, Tran, Hung, and Amouroux (2017) have used horizontal run-length distributions feature to estimate the dimensions of objects in several locations of an image. The horizontal run-length features are computed from the binary images, accumulated over a certain learning period and maintained for each location in the image. The object dimensions are then estimated from these feature distributions. While estimating the person size well, this approach also deals with partial occlusion and perspective distortion. Characteristics are extracted from the whole image or video sequence (i.e., globally) or from targeted areas (i.e., locally). Whereas global characteristics are prone to noise, pixel occlusion, and dependency on viewpoint, local features are not. The subcategories are discussed below:

1. *Global methods*: These methods extract characteristics by considering the whole image. It is used for extraction of important characteristics of image frames. These methods are sensitive to noise, occlusion, and variant to viewpoint changes when extracting local characteristics of the image.
   (a) Discrete Fourier transform (DFT)
   DFT is a global characteristic that is used to remove noise caused due to luminosity effects, blurring, and false contouring. The frequency domain information is extracted from video sequence or multiple images in blocks by average DFT that provides information (*intensity values*) about the structure of foreground in the spatial domain (Kumari & Mitra, 2011). There is a difference between the values of foreground and background. The characteristics are then classified by *k*-nearest neighbor algorithm.
   (b) Space–time volumes (STVs)
   STVs (*x*–*y* axis spatial coordinates and time) show behavior similarity by measuring space–time intensity values to show if different video segments have similar motion pattern, that is, catching the continuity of a movement. It is a global characteristic of showing space–time relationship. Actions, shapes, or structure and its orientation are extracted from categorized video segments (Blank, Gorelick, Shechtman, Irani, & Basri, 2005; Shechtman & Irani, 2005). The measure of similarity pattern enables to identify people/group of people performing the same type of activity. This method is usually used for individual behavior and action recognition, but due to its broad structure of extracting similar patterns, it is suitable for detecting crowd behavior pattern as well. It allows freedom from foreground/background separation, learning activities, and tracking making it computationally inexpensive.
2. *Local methods*: Layer extraction methods are made more immune to errors by processing targeted areas in an image or video sequence instead of the whole image. Scale-invariant feature transform (SIFT; Lowe, 1999, 2004; Scovanner et al., 2007) is an invariant method for characteristics extraction designed to avoid scaling and rotation by using only grayscale information. SIFT descriptors are used with Shi–Tomasi detector in pedestrian counting using thermal images (Kristoffersen, Dueholm, Gade, & Moeslund, 2016). The use of stereo thermal camera method shows promising results in handling changing lighting conditions and heavy occlusions. Lucas–Kanade–Tomasi method (Lucas & Kanade, 1981; Shi, 1994) solves appearance and shape disorientation by differential optical flow of local neighborhood of considered pixel. Nonparametric weighted feature extraction (Kuo & Landgrebe, 2004; Lin, Hsu, & Lin, 2010) uses histogram vectors which efficiently solve multiclass high dimensionality patterns. Using nonparametric matrices (Kuo & Landgrebe, 2004) reduces outliers effects.

Overview of developed methods in vision. Several crowd estimation methods are developed using the aforementioned techniques and algorithms. These methods can be divided into the following main categories:

1. Texture-based estimation
   The texture provides information related to the spatial arrangement of intensities and color in an image (Shapiro & Stockman, 2001). For this purpose, images are converted to the frequency domain. Statistical and spectral methods such as Fourier spectrum, GLDM, and histogram generation based on the matrix are applied. Images with high frequency show fine patterns and smooth histogram which correspond to high crowd density, while low crowd density images show coarse patterns and dispersed histograms. Based on this information, ranges of crowd densities are defined into the classifier to categorize crowd density

into very low, low, medium, high, very high, and jammed. Marana, Velastin, Costa, and Lotufo (1998) have used this method and identified its pros and cons.

2. Objects dimensions identification
   Objects are identified based on different dimensions. The image is converted to a binary image, that is, black and white color variation by using edge detection. Edge detection captures discontinuities in brightness. Different structures are identified and separated. The dimensions that correspond to people are already defined in the classifier which estimates the number and classifies it into a density category.
3. Map-based estimation
   An image is divided into cells using perspective projection—the mapping or drawing of 3-D object onto a 2-D plane—forming a map of multiresolution cells. Features are extracted from each cell separating the background from people. The number of people is calculated in each cell and then added to get the total (Qian, Wu, & Xu, 2011). For uniformity and ease of calculations, different cell sizes are normalized to a certain value. Since in dense crowded situations, the total body is not visible, only heads are counted in this method. This is why it is suitable for most crowd scenarios. Accurate density map calculations lead to accurate counting.
4. Pixel counting
   Pixel counting can be applied to areas of interest in a separated foreground image from the background. The foreground image is divided into rectangular areas using perspective projection to specify areas of interest. Objects besides people are occluded. A relationship between the numbers of pixels comprising an individual is useful to detect the number of individuals in an image regardless of pixel position in the scene (Ma, Li, Huang, & Tian, 2004). This feature leads to crowd density estimation.

WSN-Based Approach

Monitoring, guiding, and surveillance applications are totally dependent on video monitoring and camera sensors. Dependency of these applications on visual feed is logical. Video monitoring and still images provide instant snapshots of crowd movements. Computer vision tools and algorithms have evolved over time from exhaustive research to run on the video feed to analyze the crowd flow, crowd distribution, crowd density, estimate the number of people, and instantly update databases for guiding or controlling the crowd. It has become a classical solution for various kinds of monitoring applications. Granted its advantages, there is a downside to it: In order to get an accurate and precise result, computer vision tools rely on images or video feeds from multiple cameras mounted on predetermined locations. The number of cameras required for accurate results is directly proportional to the crowd size. The overall cost of the entire network of cameras and their connectivity and the complexity it introduces scores unimaginably high. Moreover, automatically analyzing large image database in real time is hard due to large computational overhead. Scalable covered areas tend to have a different lighting environment in different areas introducing further challenges. All these problems are secondary to the invasion of privacy of people being monitored.

For the purpose of crowd density estimation, an alternative relatively inexpensive and sufficiently accurate approach is to utilize signal strength characteristics in WSNs. Instead of using a host of cameras, a grid of WSN nodes can be deployed. Using received signal strength indicator (RSSI), WSN nodes are able to accurately (with tweaking and calibration) estimate crowd distribution inside the grid of nodes. Yuan and Zhao (Yuan, Zhao, Qiu, & Xi, 2013) comprehensively studied WSNs-based crowd estimation in dynamic environments. The iterative approach taken by them consists of training, monitoring, and calibrating phases. RSSI values are collected from all nodes to construct a training database. Then, various density levels of crowd are clustered by using K-means algorithm.

Finally, spatial–temporal stability method is used to eliminate noise and deviation. The result obtained by the iterative method is accurate because the deviation is convergent.

Training phase. This phase builds fingerprint database by measuring the threshold of sensor nodes to estimate crowd density. First off, RSSI values are calculated from a grid of nodes in static environments (when the environment is empty, i.e., no people). Detection threshold r_s is measured by the equation:

$$r_s = \frac{\sum_{i=1}^{t} s_i}{t}, \tag{3}$$

where s_i ($s_1, s_2, \ldots, s_t$) denote signal strength for each node and t is user defined time.

$$\sigma_s = \sqrt{\frac{\sum (s_i - r_s)^2}{t}}, \tag{4}$$

where σ_s is the standard deviation of RSSI for each node. In case of dynamic environment (people are present), a node's RSSI is used to detect interference caused by people in order to calculate crowd distribution. In this case, a number θ is used as a threshold to detect if interference of a node is likely to estimate crowd density. This value can be obtained by combining r and σ. If s is the received RSSI of a node and Equation 5 holds true, then the detection threshold is set by Equation 6. Here, k is rectify factor and δ is the signal strength change value that is initialized by training experience.

$$|s - r| \geq \delta, \tag{5}$$

$$\delta = k\sigma. \tag{6}$$

The detection threshold changes continuously with iterations of the algorithm. In fact, multiple hours are required for a training period depending on environment. The fingerprint database is updated along with iterations. The FingerPrint method used by Indoor global system for mobile communication localization (Otsason, Varshavsky, LaMarca, & De Lara, 2005) is adopted to judge if the change in cell is worth focusing. If the change in cell exceeds the threshold, its status changes from "quiet" (change is too little) to "active" (change should be further analyzed).

Monitoring phase. K-means (MacQueen, 1967) algorithm or Lloyd's (1982) algorithm is a clustering technique widely used in signal processing for cluster analysis in data mining. It is applied on fingerprints obtained in the training phase to generate different levels and types of density of the crowd. K-means algorithm has many variants and techniques. Common among them is an iterative refinement technique which basically iterates between two steps until convergence is achieved. To start the iteration process, the Forgy (1965) partition method is used as an initialization technique. This method randomly chooses k observations (from fingerprint database) and then use them as initial means $m_1, m_2 \ldots m_k$.

Assignment step: Each observation (signal strength) is assigned to the cluster with the closest mean to the means chosen by Forgy method. This partition the observations according to Voronoi diagram. Voronoi diagram is a method of dividing space into regions. A seed (mean in this case) is specified in the start and all other points closer to the seed are clustered together.

$$S_i^t = X_j - m_i^t \leq X_j - m_j^t, \tag{7}$$

where X is a region of divided space.

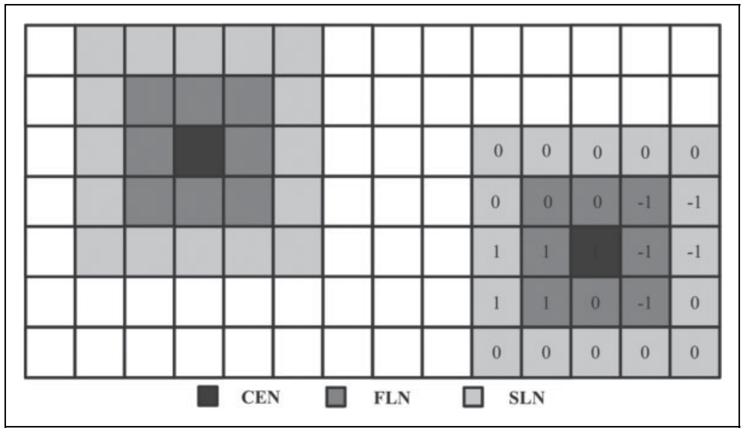

Figure 5. Calibration of FSA using FLN and SLN: Courtesy of Yuan and Zhao (Yuan, Zhao, Qiu, & Xi, 2013). FSA = the areas which contain FLN and SLN; FLN = first layer neighbors; SLN = second layer neighbors.

Update step: New means are calculated to be centerfolds of observations in the cluster.

$$m_i^{t+1} = \frac{1}{|S_i^t|} \sum_{x_j \in S_i^{(t)}} X_j. \tag{8}$$

This algorithm is considered to have converged when assignments seize to change. Calculations can be made to change the clusters for different types of crowd. To apply temporal–spatial analysis, density levels are assigned as DEH, DEM, and DEL for high density, medium density, and low density areas, respectively.

Calibrating phase. Due to the unstable nature of RSSI and other sensitive characteristics of wireless channels, sensor data are prone to errors and deviations. Furthermore, multipath phenomenon, node placement, antenna orientation, hardware discrepancies, and environmental factors such as humidity and temperature have a greater influence on RSSI. To rectify these errors and deviations, it is necessary to set some rules to calibrate sensor readings. This is achieved by neighbor field correlation and adjacent slots stability.

Two important aspects of calibration process are temporal and spatial correlations. Variations of crowd density and RSSI change in a short period of time are little or almost zero. On the other hand, changes in density in one cell are closely related to neighboring cells if the crowd is moving. This is analyzed in spatial correlation part of the calibration process. This entire process works on two basic assumptions: (i) in particular user-defined period, crowd density of certain subareas is stable (temporal characteristics) and (ii) closer subareas have influence on crowd in the study of a particular subarea (spatial characteristics). User-defined time period plays pivotal role in the process. The shorter time period will require high computations and longer time periods will fade the difference between spatial and temporal correlation.

To understand this process, Yuan, Zhao, Qiu, and Xi (2013) illustrated an example, as depicted in Figure 5. The cell under study is marked as CEN, whereas the immediate neighbors are marked first layer neighbors (FLNs) and outer layer neighbors are marked second layer neighbors (SLNs). All other neighbors are omitted for the purpose of clarity and simplicity. α and β are two coefficients of the FLN and SLN, respectively, to measure the influence of these layers. The difference in RSSI (ΔRSSI) between two timestamps (t_m, t_{m-1}) is calculated by:

$$\Delta\text{RSSI} = \left| |s - r|_{t_m} - |s - r|_{t_{m-1}} \right|, \tag{9}$$

where s is signal strength and r is the threshold of signal strength.

Now the RSSI variance of one cell can be denoted by:

$$\gamma = 0 : \ \Delta\mathrm{RSSI} < \delta, \tag{10}$$

$$\gamma = 1 : \ \Delta\mathrm{RSSI} > \delta \text{ and } \mathrm{RSSI}_{t_m} > \ \mathrm{RSSI}_{t_{m-1}}, \tag{11}$$

$$\gamma = \ -1 : \ \Delta\mathrm{RSSI} > \delta \text{ and } \mathrm{RSSI}_{t_m} < \ \mathrm{RSSI}_{t_{m-1}}, \tag{12}$$

where γ is signal strength value based on RSSI. For multiple cells, the sum of RSSI variance is calculated which is denoted by $\mathrm{RV}_{n1,\ n2,\ldots,\ nk}$ ($n1$, $n2, \ldots, nk$ are the cells).

$$\mathrm{RV}_{n1,n2,\ldots,nk} = \sum_{ni} \gamma. \tag{13}$$

For example, the value of $RV = 0$ represents very little change in RSSI, whereas the $RV = 6$ shows a sharp increase in crowd density and $RV = -6$ denotes sharp decrease. The fluctuations of RSSI on different layers can be presented as $\Delta\mathrm{RSSI}_{\mathrm{CEN}}$, $\Delta\mathrm{RSSI}_{\mathrm{FLN}}$, and $\Delta\mathrm{RSSI}_{\mathrm{SLN}}$.

$$|\Delta\mathrm{RSSI}_{\mathrm{CEN}} + \Delta\mathrm{RSSI}_{\mathrm{FLN}} \times \alpha + \Delta\mathrm{RSSI}_{\mathrm{SLN}} \times \beta| \approx 0, \tag{14}$$

$$\text{then}, -\Delta \approx \mathrm{RSSI}_{\mathrm{CEN}} + \Delta\mathrm{RSSI}_{\mathrm{FLN}} \times \alpha + \Delta\mathrm{RSSI}_{\mathrm{SLN}} \times \beta. \tag{15}$$

If deviation between $-\Delta\mathrm{RSSI}_{\mathrm{CEN}}$ and $(\Delta\mathrm{RSSI}_{\mathrm{FLN}} \times \alpha + \Delta\mathrm{RSSI}_{\mathrm{SLN}} \times \beta) > \varepsilon$, where ε is a threshold, it demonstrates that $\Delta\mathrm{RSSI}_{\mathrm{CEN}}$ is incorrect. Hence, in order to guarantee accuracy, $\Delta\mathrm{RSSI}_{\mathrm{CEN}}$ and $-(\Delta\mathrm{RSSI}_{\mathrm{FLN}} \times \alpha + \Delta\mathrm{RSSI}_{\mathrm{SLN}} \times \beta)$ need to be rectified. With further iterations, deviation and error gradually decrease and the final result converges to within the desired accuracy after few rounds.

Extensive experimentation is performed in Yuan et al. (2013) to show the effectiveness and feasibility of WSNs to estimate crowd density. The experimental setup uses TelosB nodes deployed in a 4 m $\times$ 4 m grid with a height of 1.4 m above ground. This distance between each node can be extended to 8 m by trading-off the estimation accuracy. However, even that distance is not convenient for deployment except if the space is specifically designed for deployment of such WSN systems.

Mobile Phones and Internet Data

The new trend of increased usage of smartphones opens a new gateway to estimate and control crowds in different situations. In addition, the use of social media data has been explored in different areas such as for electoral prediction (Gayo-Avello, 2013). Usage of both mobile service and social media to interact presents two new options to be used for crowd estimation. Both can be used in conjunction with each other to accurately estimate crowd size. This method can surpass other methods in quick estimation as it does not involve extra algorithms to be applied on the available data, crucial in emergency situations. Studies have shown (Choi & Varian, 2012; Quercia, Lathia, Calabrese, Di Lorenzo, & Crowcroft, 2010) that cellular network and Internet data can be efficiently used in economic trends, attendance to an event, traffic management, urban area planning and crowd size estimation, and so on.

When using mobile phone data, first of all the cells of the network in the event's location are identified. Then, the number of people can be obtained through the user service ID while the active data, that is, phone calls and short message service can be used to identify active users. For instance, Mamei and Colonna (2016) have used call detail records to estimate the number of people to an event. They have conducted a study to estimate crowd size based on mobile phone data for 43 different events. For the monitored cells of the network, which corresponded to the geographical area of the event, the data are saved using a time-based saving activity. Their results have shown less than 15% of median error of the actual number of people.

Social networks also provide the ability to analyze the geospatial dimensions of a particular area. For example, Sinnott and Chen (2016) have collected the coordinate properties information for crowd estimation using the Twitter's application program interfaces. It is also possible to use both mobile phone and social media data to estimate crowd size (Botta, Moat, & Preis, 2015). Their study is based on long-term data collection of a specific area. Both active and passive data of mobile phone user and social media data sent and received in the geographical location are considered to estimate crowd size in a specific time and specific event in a location.

Simulation Support for Crowd Size Estimation

Crowd synthesis (or simulation) is an important and challenging phenomenon that can be used to improve crowd analysis. Crowd simulation is related to reproduction of realistic crowds using computer graphic techniques (Junior, Musse, & Jung, 2010). To this end, several crowd simulation models have been built (Thalmann, 2007) in recent years to represent crowd status using crowd information. Modeling of crowd dynamics (discussed in Modeling Crowd Dynamics subsection) makes the developing strategies efficient and possible to validate. In addition, automated systems developed (Helbing, Farkas, & Vicsek, 2000) based on a modeled system can be further fine-tuned to react to particular crowded situations using virtual environment evaluations. Thus, the performance of a system can be evaluated through simulations based on real scenarios.

Simulation Models

Crowd simulation models are realistic if based on real data captured from the actual environment, for example, spatial distribution of a crowd in a real scene can be used to initialize a crowd simulator and track trajectories to guide the motion (Musse, Jung, Jacques, & Braun, 2007). In addition, validation requires ground truth which in the case of crowd control is nearly impossible to obtain beforehand. A model that can simulate all possible crowd behaviors has not yet been achieved due to lack of real knowledge of the crowd and large number of behaviors that are still unobserved and unexplained (Junior et al., 2010). Yet, in the absence of real data from the crowded scenes, simulations can be used to a greater extent to validate and train a system. Applications of crowd simulation are phenomenal. Simulating real-world crowds, such as at airports and football stadiums, can help understand the flow of the crowd in a specific time frame such as entering or leaving the stadium or airport (Narain, Golas, Curtis, & Lin, 2009). For an already built area, crowd simulations provide key information such as normal capacity limit, emergency route planning, obstacles that can be avoided, and most commonly used walkways. This information can be used effectively for deploying sensors and actuators of crowd surveillance and estimation system. Moreover, virtual crowd's simulation has made its mark in areas such as urban area planning, emergency training and response, administrative policy-making, architectural structures designs, and large gatherings control.

Despite providing these advantages, the crowd simulation algorithms still have several challenges. For example, based on mathematical models, the existing crowd simulators try to reproduce the average behavior of human beings and thus are unable to simulate unpredictable behaviors. To replicate crowd movement and behavior, observed knowledge of a particular type of crowd and the environment plays a vital role. Moreover, the visual quality of the simulation is another important issue that needs to consider a number of parameters related to rendering, environment, and virtual humans (Junior et al., 2010). A variety of features that characterize the crowd simulation system are shown in Table 3.

Table 3. Main Features of Crowd Simulation System.

Crowd Simulation Features	Description
Crowd size	Shows number of individuals a system can simulate in real time for designing and planning of building
Environment type	Simulations are based on appropriate type, that is, indoor or outdoor
Model type	Description of decisions that are to be taken based on rules implemented
System structure	Space represented by coordinates, that is, continuous or divided into areas, that is, grid
Individuality/groups behaviors	Set of rules specifying normal and abnormal behaviors related with particular actions
Hazards	Rules defining hazards related to specific environments which should be avoided
Signaling and alarms	Interaction between crowd and environment, for example, signs and alarms for safe evacuation

Simulators

The main goals of many commercial applications available for crowd simulations are usually crowd densities, flow rate, identifying congestion areas, and evacuation procedure and time required for it. The majority of the simulators is usually designed for animation purpose rather than simulation to visualize structure design and provide support to enhance it. They usually do not perform well when applied to 3-D virtual environments because these methods consider particle flow and cannot identify human movements and behaviors (Pelechano et al., 2008).

Agents-based models and simulation systems discussed in Modeling Crowd Dynamics subsection are the advanced mechanisms used for 3-D simulation. Most agents-based methods work on the free flow of autonomous and interacting agents and hence applied for low or medium crowd environments. Every individual in the crowd is treated as an agent (Zhou et al., 2010). However, HiDAC (Pelechano et al., 2007) and MACES (Hung et al., 2013) and Crowd with Aleatoric, Reactive, Opportunistic, and Scheduled Actions (CAROSA) are developed for simulating dense crowds. The above three systems are combined in a single framework for handling dense crowd using multi-agent simulations (Pelechano et al., 2008). Additional properties such as personality, communication with the environment, and other agents are added. Geometric properties of area for navigation in a virtual structure and communication for way finding are key properties for making it a better simulation framework for any scenario. This method is developed to consider real-world scenarios of dense crowd.

Software development companies provide general solutions for crowd management, including aspects such as size estimation and abnormal behavior detection. Size estimation poses different challenges in different environments and scenarios for which these available software not only model and simulate a particular environment but also deploy the best possible solution based on simulation results. Some of the commercially available software is highlighted in Table 4.

Discussion

Despite having several real-life applications, accurately estimating the crowd size is really difficult. This section presents the analysis of the crowd estimation solutions reviewed in this article (as detailed in Crowd Size Estimation Approaches section). The comparison of these solutions is provided in Table 5, which includes the accuracy of these methods and the data sets used by them. Among the computer vision–based methods, texture-based estimation is good for low and medium crowd estimation but not for dense crowds. Because in texture-based estimation, exact head count is

Table 4. Crowd Modeling, Simulation, and Analysis Software.

Software Name	Properties/Uses
Legion (www.legion.com/why-legion)	Simulating crowd flow to help design, test, and validate open and closed venues
MASSIVE (www.massivesoftware.com/engineering.html)	Modeling the idiosyncrasies of complex, real-life behaviors into agents who use visual and auditory cues, as we do in real life
MassMotion (www.oasyssoftware.com/products/engineering/massmotion.html)	Pedestrian simulation and crowd analysis tool that provides designers and operators with clear information about crowding, usage patterns, and occupant safety in a facility
STEPS (www.steps.mottmac.com)	An agent-based microsimulation tool developed for the simulation of pedestrian movement
NetLogo (ccl.northwestern.edu/netlogo)	A multi-agent programmable modeling environment for research, freely available
Pedestrian Dynamics (www.pedestrian-dynamics.com/pedestrian-dynamics/pedestrian-dynamics.html)	For the creation and execution of large pedestrian simulation models in complex infrastructures for testing performance and safety
XFlow (www.nextlimit.com/products/name/xflow)	Its a next generation Computational Fluid Dynamics software system that uses a proprietary, particle-based, mesh less approach
RealFlow (www.nextlimit.com/products/name/realflow)	For the simulation of fluids (particle flow) and body dynamics
VISuite (www.ipsotek.com/en/products)	Hybrid video analytics engine
Video Turnstile (videoturnstile.com)	Use CCTV to count people in both indoors or outdoors
Iomniscient iQ-Series	Crowd management

Note. CCTV = closed-circuit television.

Table 5. Comparison of the Major Crowd Estimation Methods.

Method	Associated with	Accuracy	Data Set
Computer vision–based methods			
Texture-based estimation	Marana, Velastin, Costa, and Lotufo (1998)	82% (statistical GLDM); 80% (spectral method)	Liverpool Street Railway Station (London, UK)
Object dimension specification	Marana, Costa, Lotufo, and Velastin (1999)	near 85% (GLDM); around 75% (Minkowski fractal dimension)	Liverpool Street Railway Station (London, UK)
Map-based estimation	Qian, Wu, and Xu (2011)	No information	No information
Pixel counting	Ma, Li, Huang, and Tian (2004)	No information	Various surveillance scenes
Wireless sensor network	Yuan, Zhao, Qiu, and Xi (2013)	94% (static); 86% (moving crowd)	Experiments in lab
Mobile phones and Internet data	Botta, Moat, Preis (2015)	Within 95%	San Siro Football Stadium Linate Airport (Milan, Italy)

Note. GLDM = gray-level dependence matrix.

not possible due to limited data. Although object dimensions identification is an efficient method that identifies dimensions, that is, $individuals/m^2$, it is not suitable for dense crowd situations, as the limited data don't show exact head count. While map-based estimation works for both indoor and outdoor environments, the outdoor environments should have shadow elimination filters and light intensity adjustments. However, map-based estimation is only limited to very low or medium crowds where the whole body of an individual is visible. Moreover, it is also computationally intensive task and thus affects throughput. The computer vision–based methods in general face many problems such as occlusion, illumination variation, extensive computation, and so on.

The major setback of WSNs-based method is the deployment of the sensor grid in the field. The distance between each node can be extended to 8 m by trading-off the estimation accuracy. However, even that distance is not convenient for deployment except if the space is specifically designed for deployment of such WSN systems. The use of mobile phone and social media data to estimate crowd size is both good for emergency situations and urban planning such as traffic management. However, to estimate correctly, this requires that every person must have a mobile phone and Internet access, he or she is using the mobile phone/social media, and his or her location-based services are activated (in case of estimation via social media data). It is thus concluded that the key to deploying a successful solution for a specific purpose is to know the requirements, the environment, and crowd scenarios. Secondly, the solution needs to be thoroughly evaluated to know its qualities and limitations in terms of detection and estimation mechanisms.

Summary

Recent developments in automated solutions for crowd estimation are quite remarkable keeping in mind the limitations of the current technologies. More advanced ideas and solutions are being discussed yet it is hard to develop a perfect solution suitable for all environments and situations. This article presents a review of major automated crowd estimation approaches to help in their selection for a particular deployment scenario. Instead of reviewing a particular technology-based solutions, as in case of the existing survey papers, theoretical understanding of three different approaches (computer vision, WSNs, and Internet data/mobile phone-based approaches) is provided. In addition, a detailed account of the pros and cons of each approach is provided in terms of accuracy and suitability for different deployment conditions. However, it is worth mentioning that any crowd estimation system itself depends on multiple aspects of technological and environmental factors that need to be considered.

Acknowledgments

The authors would also like to thank Dr. Nasru Minallah and Dr. Laiq Hasan, Department of Computer System Engineering, University of Engineering and Technology (UET), Peshawar, Pakistan, for their valuable suggestions.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is funded by the Transportation and Crowd Management Center of Research Excellence, Umm Al-Qura University, Saudi Arabia.

References

Antonini, G., Bierlaire, M., & Weber, M. (2006). Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological*, *40*, 667–687.

Blank, M., Gorelick, L., Shechtman, E., Irani, M., & Basri, R. (2005, October). Actions as space-time shapes. In *Tenth IEEE International Conference on Computer Vision (ICCV'05)* (Vol. *2*, pp. 1395–1402). Beijing, China: IEEE.

Botta, F., Moat, H. S., & Preis, T. (2015). Quantifying crowd size with mobile phone and Twitter data. *Royal Society Open Science*, *2*, 150162.

Chen, Y., Liang, G., Lee, K. K., & Xu, Y. (2007, July). Abnormal behavior detection by multi-SVM-based Bayesian network. In *2007 IEEE International Conference on Information Acquisition (ICIA'07)* (pp. 298–303). Seogwipo-si, South Korea: IEEE.

Cheng, K. W., Chen, Y. T., & Fang, W. H. (2013, October 21). Abnormal crowd behavior detection and localization using maximum sub-sequence search. In A. Doulamis, M. Bertini, N. D. Doulamis, J. Gonzalez, & A. Voulodimos (Eds.), *Proceedings of the 4th ACM/IEEE International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream ARTEMIS'13* (pp. 49–58). Barcelona, Spain: ACM.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, *88*, 2–9.

Davies, A. C., Yin, J. H., & Velastin, S. A. (1995). Crowd monitoring using image processing. *Electronics & Communication Engineering Journal*, *7*, 37–47.

Dickie, J. F. (1995). Major crowd catastrophes. *Safety Science*, *18*, 309–320.

Forgy, E. W. (1965). Cluster analysis of multivariate data: Efficiency versus interpretability of classifications. *Biometrics*, *21*, 768–769.

Gayo-Avello, D. (2013). A meta-analysis of state-of-the-art electoral prediction from Twitter data. *Social Science Computer Review*, *31*, 649–679.

Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, *67*, 786–804.

Helbing, D., Farkas, I., & Vicsek, T. (2000). Simulating dynamical features of escape panic. *Nature*, *407*, 487–490.

Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical Review E*, *51*, 4282.

Henke, L. L. (2016). Estimating crowd size: A multidisciplinary review and framework for analysis. *Business Studies Journal*, *8*, 27–38.

Hevia, C. (2008). Standard errors using the delta method and GMM. mimeo.

Hu, X., Zheng, H., Wang, W., & Li, X. (2013). A novel approach for crowd video monitoring of subway platforms. *Optik—International Journal for Light and Electron Optics*, *124*, 5301–5306.

Huang, L., Wong, S. C., Zhang, M., Shu, C. W., & Lam, W. H. (2009). Revisiting Hughes' dynamic continuum model for pedestrian flow and the development of an efficient solution algorithm. *Transportation Research Part B: Methodological*, *43*, 127–141.

Hughes, R. L. (2002). A continuum theory for the flow of pedestrians. *Transportation Research Part B: Methodological*, *36*, 507–535.

Hung, N. Q. V., Tam, N. T., Lam, N. T., & Aberer, K. (2013, October). An evaluation of aggregation techniques in crowdsourcing. In X. Lin, Y. Manolopoulos, D. Srivastava, & G. Huang (Eds.), *International Conference on Web Information Systems Engineering: WISE 2013, Part II, LNCS 8181* (pp. 1–15). Berlin, Germany: Springer.

Jojic, N., & Frey, B. J. (2001). Learning flexible sprites in video layers. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001* (Vol. 1, pp. I–I). Kauai, HI, USA: IEEE Computer Society Press.

Junior, J. C. S. J., Musse, S. R., & Jung, C. R. (2010). Crowd analysis using computer vision techniques. *IEEE Signal Processing Magazine*, *27*, 66–77.

Kristoffersen, M. S., Dueholm, J. V., Gade, R., & Moeslund, T. B. (2016). Pedestrian counting with occlusion handling using stereo thermal cameras. *Sensors*, *16*, 62.

Kumar, M. P., Torr, P. H., & Zisserman, A. (2008). Learning layered motion segmentations of video. *International Journal of Computer Vision*, *76*, 301–319.

Kumari, S., & Mitra, S. K. (2011, December). Human action recognition using DFT. In *2011 Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)* (pp. 239–242). Hubli, India: IEEE.

Kuo, B. C., & Landgrebe, D. A. (2004). Nonparametric weighted feature extraction for classification. *IEEE Transactions on Geoscience and Remote Sensing*, *42*, 1096–1105.

Li, T., Chang, H., Wang, M., Ni, B., Hong, R., & Yan, S. (2015). Crowded scene analysis: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, *25*, 367–386.

Lin, C. H., Hsu, F. S., & Lin, W. Y. (2010). Recognizing human actions using NWFE-based histogram vectors. *EURASIP Journal on Advances in Signal Processing*, *2010*, 9.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*, 129–137.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision* (Vol. 2, pp. 1150–1157). Kerkyra, Greece: IEEE.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*, 91–110.

Lucas, B. D., & Kanade, T. (1981, August). An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence, IJCAI'81* (Vol. 2, pp. 674–679), Vancouver, Canada. San Francisco, CA: Morgan Kaufmann Publishers Inc.

Ma, R., Li, L., Huang, W., & Tian, Q. (2004, December). On pixel count based crowd density estimation for visual surveillance. In *2004 IEEE Conference on Cybernetics and Intelligent Systems*, Singapore, Singapore. (Vol. 1, pp. 170–173). IEEE.

Macal, C. M., & North, M. J. (2009, December). Agent-based modeling and simulation. In *Winter Simulation Conference* (WSC'09) (pp. 86–98). Austin, TX: Winter Simulation Conference.

MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). Berkeley, Calif: University of California Press.

Mamei, M., & Colonna, M. (2016). Estimating attendance from cellular network data. *International Journal of Geographical Information Science*, *30*, 1281–1301.

Marana, A. N., Costa, L. D. F., Lotufo, R. A., & Velastin, S. A. (1999, March). Estimating crowd density with Minkowski fractal dimension. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. Proceedings, ICASSP99* (Cat. No. 99CH36258) (Vol. 6, pp. 3521–3524). Phoenix, AZ: IEEE. Washington, DC: IEEE Computer Society.

Marana, A. N., Velastin, S. A., Costa, L. D. F., & Lotufo, R. A. (1998). Automatic estimation of crowd density using texture. *Safety Science*, *28*, 165–175.

Mariano, V., Tran, L. D., Hung, T. Q., & Amouroux, E. (2017, April). Person size estimation in image sequences using foreground run-length distributions. In *2017 Seventh International Conference on Information Science and Technology (ICIST)* (pp. 246–250). Da Nang, Vietnam: IEEE.

Mehran, R., Oyama, A., & Shah, M. (2009, June). Abnormal crowd behavior detection using social force model. In *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009* (pp. 935–942). Miami, FL: IEEE.

Musse, S. R., Jung, C. R., Jacques, J., & Braun, A. (2007). Using computer vision to simulate the motion of virtual agents. *Computer Animation and Virtual Worlds*, *18*, 83–93.

Narain, R., Golas, A., Curtis, S., & Lin, M. C. (2009, December). Aggregate dynamics for dense crowd simulation. In *ACM Transactions on Graphics (TOG), Proceeding SIGGRAPH Asia '09 ACM SIGGRAPH Asia* (Vol. 28, p. 122). ACM. , Yokohama, Japan

Otsason, V., Varshavsky, A., LaMarca, A., & De Lara, E. (2005, September). Accurate GSM indoor localization. In M. Beigl, S. Intille, J. Rekimoto, & H. Tokuda (Eds.), *International Conference on Ubiquitous Computing*, *UbiComp 2005, LNCS 3660* (pp. 141–158). Berlin, Germany: Springer.

Papadimitriou, E., Yannis, G., & Golias, J. (2009). A critical assessment of pedestrian behaviour models. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*, 242–255.

Pelechano, N., Allbeck, J. M., & Badler, N. I. (2007, August). Controlling individual agents in high-density crowd simulation. In *Proceedings of the 2007 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (pp. 99–108). San Diego, CA: Eurographics Association.

Pelechano, N., Allbeck, J. M., & Badler, N. I. (2008). Virtual crowds: Methods, simulation, and control. *Synthesis Lectures on Computer Graphics and Animation*, *3*, 1–176.

Piccardi, M. (2004, October). Background subtraction techniques: A review. In *2004 IEEE International Conference on Systems, Man and Cybernetics*, 10–13 October 2004 (IEEE Cat. No.04CH37583) (Vol. 4, pp. 3099–3104). The Hague, Netherlands: IEEE.

Polus, A., Schofer, J. L., & Ushpiz, A. (1983). Pedestrian flow and level of service. *Journal of Transportation Engineering*, *109*, 46–56.

Qian, H., Wu, X., & Xu, Y. (2011). *Intelligent surveillance systems* (Vol. 51). Intelligent Systems, Control and Automation: Science and Engineering. Dordrecht, the Netherland: Springer Science & Business Media.

Quercia, D., Lathia, N., Calabrese, F., Di Lorenzo, G., & Crowcroft, J. (2010, December). Recommending social events from mobile phone location data. In *2010 IEEE 10th International Conference on Data Mining (ICDM)* (pp. 971–976). Sydney, Australia: IEEE.

Regazzoni, C. S., & Tesei, A. (1996). Distributed data fusion for real-time crowding estimation. *Signal Processing*, *53*, 47–63.

Robin, T., Antonini, G., Bierlaire, M., & Cruz, J. (2009). Specification, estimation and validation of a pedestrian walking behavior model. *Transportation Research Part B: Methodological*, *43*, 36–56.

Ryan, D., Denman, S., Sridharan, S., & Fookes, C. (2015). An evaluation of crowd counting methods, features and regression models. *Computer Vision and Image Understanding*, *130*, 1–17.

Šalamon, T. (2011). *Design of agent-based models: Developing computer simulations for a better understanding of social processes*. Repin, Czech Republic: Tomáš Bruckner.

Saleh, S. A. M., Suandi, S. A., & Ibrahim, H. (2015). Recent survey on crowd density estimation and counting for visual surveillance. *Engineering Applications of Artificial Intelligence*, *41*, 103–114.

Scovanner, P., Ali, S., & Shah, M. (2007, September). A 3-dimensional sift descriptor and its application to action recognition. In *Proceedings of the 15th ACM International Conference on Multimedia* (pp. 357–360). Augsburg, Germany: ACM.

Shapiro, L., & Stockman, G. C. (2001). *Computer vision* (1st ed.). Upper Saddle River, NJ: Prentice Hall.

Shechtman, E., & Irani, M. (2005, June). Space-time behavior based correlation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005* (Vol. 1, pp. 405–412). San Diego, CA, USA: IEEE.

Shi, J. (1994, June). Good features to track. In *1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94* (pp. 593–600). Seattle, WA: IEEE.

Sinnott, R. O., & Chen, W. (2016, March). Estimating crowd sizes through social media. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, (pp. 1–6), 14–16 March 2016. Sydney, NSW, Australia: IEEE.

Sjarif, N. N. A., Shamsuddin, S. M., Hashim, S. Z. M., & Yuhaniz, S. S. (2011, June). Crowd analysis and its applications. In J. M. Zain, W. M. bt W. Mohd, & E. El-Qawasmeh (Eds.), *International Conference on Software Engineering and Computer Systems. ICSECS 2011, Part I, CCIS 179* (pp. 687–697). Berlin, Germany: Springer.

Thalmann, D. (2007). *Crowd simulation*. New York, NY: John Wiley.

Torr, P. H., Szeliski, R., & Anandan, P. (2001). An integrated Bayesian approach to layer extraction from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*, 297–303.

van Waart, P., Mulder, I., & de Bont, C. (2016). A participatory approach for envisioning a smart city. *Social Science Computer Review*, *34*, 708–723.

Wang, J. Y., & Adelson, E. H. (1993, June). Layered representation for motion analysis. In *1993 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings CVPR'93* (pp. 361–366), 15–17 June 1993. New York, NY, USA: IEEE.

Watson, R., & Yip, P. (2011). How many were there when it mattered? *Significance*, *8*, 104–107.

Yogameena, B., & Nagananthini, C. (2017). Computer vision based crowd disaster avoidance system: A survey. *International Journal of Disaster Risk Reduction*, *22*, 95–129.

Yuan, Y., Zhao, J., Qiu, C., & Xi, W. (2013). Estimating crowd density in an RF-based dynamic environment. *IEEE Sensors Journal*, *13*, 3837–3845.

Zhan, B., Monekosso, D. N., Remagnino, P., Velastin, S. A., & Xu, L. Q. (2008). Crowd analysis: A survey. *Machine Vision and Applications*, *19*, 345–357.

Zhou, S., Chen, D., Cai, W., Luo, L., Low, M. Y. H., Tian, F., . . . Hamilton, B. D. (2010). Crowd modeling and simulation technologies. *ACM Transactions on Modeling and Computer Simulation*, *20*, 20.

Zitouni, M. S., Bhaskar, H., Dias, J., & Al-Mualla, M. E. (2016). Advances and trends in visual crowd analysis: A systematic survey and evaluation of crowd modeling techniques. *Neurocomputing*, *186*, 139–159.

Author Biographies

Muhammad W. Aziz is an assistant professor in Department of Computer Science, CECOS University, Pakistan. Previously, he served as an assistant professor in Umm Al-Qura University, Saudi Arabia for 3 years. He received his PhD from Universiti Teknologi Malaysia in 2013. He published more than 20 research articles in indexed and well-reputed journals and conference proceedings. The research areas of his interest are embedded real-time software development and smart environment. Email: mwabdulaziz@uqu.edu.sa

Farhan Naeem is a research assistant at the EmNets lab at the UET, Peshawar. He is supervised by Dr. M. H. Alizai. He is currently involved in multiple research projects in the area of wireless sensor networks, delay tolerant networks, and error tolerant communications. He completed his BS and MS in computer systems engineering with distinction from UET, Peshawar. Email: farhan.naeem1@gmail.com

Muhammad H. Alizai is an assistant professor at Lahore University of Management Sciences (LUMS). Previously, he was a research assistant at ComSys Group, RWTH Aachen University, Germany. He received his PhD and MSc degrees from RWTH Aachen University in 2012 and 2007, respectively. His research interests are in mobile applications, Internet of things, sensornets, and delay tolerant networking. He is particularly interested in applications, protocols, architectures, and evaluation tools for future networks. He directs the SysNet lab at LUMS. Email: hamad.alizai@lums.edu.pk

Khan Bahadar Khan received his PhD and MS degrees in Electronic Engineering from International Islamic University Islamabad, Pakistan in 2017 and 2010, respectively. He is also working as an assistant professor in Department of computer science, CECOS University, Pakistan. He has published more than 10 papers in various international and national conferences and journals. His research interests include medical image processing, signal processing, computer vision, and machine learning. Email: kb.khattak@gmail.com