Crowd Monitoring and Classification: A Survey

Sonu Lamba and Neeta Nain

Abstract Crowd monitoring on public places is very demanding endeavor to accomplish. Huge population and assortment of human actions enforces the crowded scenes to be more continual. Enormous challenges occur into crowd management including proper crowd analysis, identification, monitoring and anomalous activity detection. Due to severe clutter and occlusions, conventional methods for dealing with crowd are not very effective. This paper highlights the various issues involved in analyzing crowd behavior and its dynamics along with classification of crowd analysis techniques. This review summarizes the shortcomings, strength and applicability of existing methods in different environmental scenarios. Furthermore, it overlays the path to device a proficient method of crowd monitoring and classification which can deal with most of the challenges related to this area.

Keywords Crowd monitoring · Behaviour analysis · Crowd classification

1 Introduction

Crowd phenomenon has been an important research issue in the era of computer vision since past few years. Due to growing population, understanding and monitoring crowd behavior has now become an essential exercise and challenge for the security agencies across the world. Thus, various groups of researchers and intellectuals set their mind to find proper solution in such field so as to control and manage crowd in their specific manner. To monitor a large area of crowd there has been an exponential increase in surveillance cameras installed around the world, limited number of human resources is not sufficient in analyzing these large number

S. Lamba (✉) · N. Nain
Department of Computer Science & Engineering, MNIT, Jaipur, India
e-mail: lamba.sonu5@gmail.com

N. Nain
e-mail: nnain.cse@mnit.ac.in

S.K. Bhatia et al. (eds.), *Advances in Computer and Computational Sciences*,
Advances in Intelligent Systems and Computing 553,
DOI 10.1007/978-981-10-3770-2_3

of video frames simultaneously. So there must be an automated way for crowd monitoring and classification. Intelligent surveillance system is one of the extremely important applications of crowd analysis.

Researchers have proposed several methods and techniques to understand crowd behaviors for developing a safe and secure environment in order to avoid crowd congestion, public riots and terror attacks etc. In crowded scenes, standard computer vision techniques are not applicable in first hand manner due to severe occlusion and complex background scenarios. Many computer vision algorithms exist in literature for tracking, detecting and in analyzing behavior of crowded scene. Although, they provide a good result in a low to medium density of population, but it is still a challenge to deal with a dense crowd. There is an adequacy to present a review of crowd monitoring and classification. Most of the recent surveys focused on the activity analysis of a single person or small group of people, rather than focusing on a crowded scenario. The survey papers by Zhan et al. [1] and Teng et al. [2] are the only two focusing on crowd video analysis as per best of our knowledge. Zhan et al. [1] focused on pedestrian detection and tracking in severely occluded scene but crowd behavior understanding and abnormality detection are not at all covered by them. Although, Teng et al. [2] focused on crowd behavior recognition, motion pattern segmentation and anomaly detection but did not provide generally accepted solution for an unseen crowd scene. This survey suggests many open issues for further research in crowd monitoring and classification.

2 Crowd Scene Monitoring Applications

Crowd monitoring has wide range of applications in the real-world scenarios.

1. **Intelligent Surveillance**: For security point of view, the very crowded places should be under camera surveillance such as railway stations, stadiums, subways and shopping malls etc. The normal surveillance system should be replaced by intelligent surveillance which can perform crowd behavior analysis and control the crowd by alarming.
2. **Crowd Management**: Crowded scene analysis helps to develop crowd management strategies. In mass gathering situations, these strategies control the crowd motion in order to avoid the overcrowded situations and public stampede.
3. **Public Space Design**: Monitoring and classification of crowd and their relevant dynamics can provide prior instruction in public space designing by ensuring safety measures and comfort level in the construction of railway stations, airport and terminals etc.
4. **Entertainment**: Computer games could be designed to simulate crowd analysis techniques to derive a correct mathematical model of crowds.

2.1 Motivations to Crowd Monitoring and Classification

In the video scene analysis and understanding, the focus is on object detection, tracking and behavior recognition [3, 4]. The conventional methods are not appropriate or sometimes fail for densely crowded scenes which have severe occlusions, ambiguities and are extremely cluttered, where undetected anomalous activities might lead to adverse situations which are terrible. Some incidents of crowd disaster at mass events are illustrated in Fig. 1. A crowd has both dynamics and psychological characteristics so analysis of behavior is a very complex task. Human crowds are often goal oriented. It is a very difficult to model dynamics of a crowd at an appropriate level. There is a need to detect, count and classify the behavior of crowd in most surveillance scenario. The rest of this paper is structured as follows. In Sect. 2, we introduce the features which are generally used in the literature of crowd scene analysis. Section 3, categorizes crowd analysis and the relevant approaches are detailed. The commonly used database for crowded scene analysis is summarized in Sect. 4. In Sect. 6 we conclude this paper by furnishing some encouraging future directions.

3 Feature Categorization in Crowded Scene

A proper feature categorization can benefitted the subsequent tasks. In crowd scene analysis, motion features play a vital role. Representation point of view, motion features can be classified into three levels as: flow-based features, local spatio-temporal features and trajectory/tracklet. In flow based features, each and every pixel is analyzed. In local spatio-temporal features, local information is extracted from 2-D patches or 3-D cubes. On a next stage, being a basic feature of motion representations, trajectory/tracklet evaluates the individual tracks. These feature representations have been used in crowd analysis and perform various tasks as crowd counting, crowd behavior recognition and crowd anomaly detection.

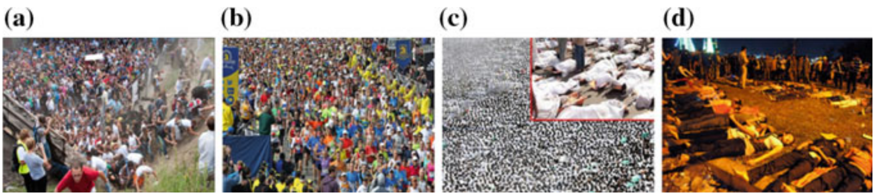

Fig. 1 Examples of crowd calamities incidents at mass gathering events: **a** Love Parade disaster-Duisburg, Germany (2010). **b** Boston marathon bombing Massachusetts, United States (2013). **c** Mina, Mecca-Saudi Arabia (2015). **d** Khmer water festival-Penh, Cambodia (2010)

3.1 Flow-Based Features

When we look at the crowd, our concern is to detect what actions are performed not who is performing it, where a certain set of activities of an individual may visible proportionately random, but to analyze whole crowd scene they have some pattern in their actions [5]. Various flow based features have been presented in recent years [6, 7]. Motion feature can be computed by the conventional optical flow method if the brightness of an image at a time t is represented by I(x, y, t) then

$$\frac{dI}{dt} = \frac{\partial \mathrm{I}}{\partial x}\frac{dx}{dt} + \frac{\partial \mathrm{I}}{\partial y}\frac{dy}{dt} + \frac{\partial \mathrm{I}}{\partial t} = \frac{\partial \mathrm{I}}{\partial x}\mathrm{u} + \frac{\partial \mathrm{I}}{\partial y}\mathrm{v} + \frac{\partial \mathrm{I}}{\partial t}$$

We calculate motion vector (u, v) for all points in an image. The equation is solved by Optical Flow Constraint (OFC) [6], where image brightness is supposed to be constant with respect to time. Flow-based features are further divided into three categories.

- **Optical Flow**: Optical flow is used for motion detection in video sequences by using flow vector of moving objects. However, long range dependencies are not captured by optical flow.
- **Particle Flow**: The computation of particle flow is done by moving grid with the help of optical flow. The trajectories provided by particle flow are related with an initial and later position of a particle. It has shown very excellent results on abnormal behavior detection and crowd segmentation.
- **Streak Flow**: Mehran et al. [8] introduced streak flow which helps in computation of an accurate motion field in crowded video. It also provides temporal evolution of moving object in a period of time and measures flow in visualization and fluid mechanics.

3.2 Local Spatio-Temporal Features

When an optical flow method could not provide sufficient motion information due to an unstructured crowded scene with high density. In such situations, local spatio temporal features are one of the solutions to gain motion flow of crowd. In the estimation of pedestrian movement, the nonuniform motion is produced by any number of objects in each local area. In these kinds of scenarios, only dense local pattern is utilized. To provide the spatio-temporal relationship, two methods are existing such as spatio-temporal gradients [9, 10] and histogram functions [11]. They extract whole motion of crowd video and specify its spatio-temporal distributions based on local 2D patches or 3D cubes. The spatio-temporal gradient, for each pixel i in each patch is calculated as:

$$\left(I_{ix}, I_{iy}, I_{iz} \right)^T = \left(\frac{\partial I}{\partial x} \frac{\partial I}{\partial y} \frac{\partial I}{\partial z} \right)^T$$

where x, y and t are the video's horizontal, vertical, and temporal dimensions, respectively. Motion histogram plots the motion information defined in local region. In fact, it is not suitable for crowd motion analysis because computing motion is not only time consuming but also erroneous. In contrast of motion histogram, a novel feature descriptor called multi-scale histogram of optical flow (MHOF) is proposed by Cong et al. [11]. Spatio-temporal features are widely used in crowd anomaly detection due to their strong descriptive power.

3.3 Trajectory/Tracklet

Comparing with other features, trajectory is more attractive and semantic than flow based and spatio temporal representation. We can observe crowd scene activities by motion features due to its repetitive pattern. The features such as acceleration, motion energy and relative distance between objects etc. are extracted from trajectories of crowd video. However, as previously mentioned, object detection, feature extraction and tracking are not performed accurately due to severe occlusion and clutter. To overcome these difficulties and obtaining complete trajectory, a motion feature is introduced in [12]. They are a piece of trajectory acquired from the tracker within a short interval termed as tracklet.

When severe occlusion and clutter scene are aroused, tracklets are terminated because they are more stable and rarely change with scenes. In [12–14], they used track-lets to obtain complete trajectories for tracking or activity recognition. In [12] various tracklet based methodologies are proposed to analyzing and clustering semantic regions in unstructured areas. In their work, they extract tracklets from dense crowd video then apply a defined model and it executes the spatial and temporal coherence between tracklets and eventually obtains a behavior pattern to analyze densely crowded scenes. An approach to segment video in the form of trajectories that helps in further analysis of activities present in a video is proposed in [15].

4 Crowd Analysis and Monitoring

The research on crowd analysis is comprehensively divided into three parts: People counting/density estimation, people tracking and behavior understanding or anomaly detection. A taxonomy of crowd analysis and monitoring is systematically represented in Fig. 2 with subcategories so the readers can easily understand the

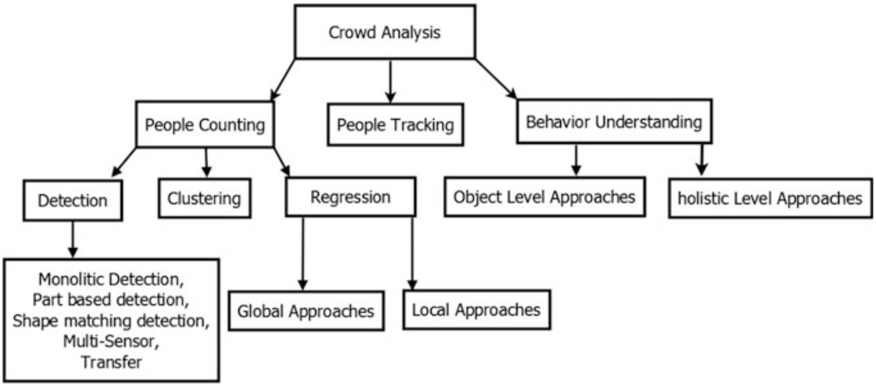

Fig. 2 Systematic representation of crowd monitoring and classification

crowd phenomenon. Brief explanations of these categories of crowd classification are provided in the following subsections.

4.1 People Counting or Density Estimation

People counting are mainly focused on overcrowded areas for both security and safety purposes [16]. People counting or density estimation is a dominant problem to define the level of a crowd as dense or sparse. People counting can be applied on static images and video sequences in both outdoor and indoor scenarios. In recent years, people counting can be arranged as: counting by detection, counting by clustering and counting by regression [17]. The various crowd counting strategies are illustrated in Table 1 along with its advantages and shortcomings.

4.2 People Tracking

In people tracking, we need to locate the position of an individual in successive frames. The problems of people counting and tracking are correlated, as both have the target of identifying people in crowded scene. However, the problem of counting generally needs to approximate the number of participants present in crowd, instead of their position. On the other side, the tracking problem involves to locate each individual in the frames as a function of time. Mikel et al. [29] combine crowd density and tracking of individual people by optimizing joint energy function. A ground truth density Fo(p) is proposed by them as a kernel density estimation based on the positions of annotated points as

Table 1 Comparative study of people counting literature with advantages and disadvantages

Reference and year	Methods	Advantage	Disadvantage	Datasets
2015 [18]	DPM + SIFT + GLCM + Fourier	Combination of multiple features	Perform count only on still images	UCF dataset
2011 [19]	BP-neural network	High accuracy in low density	False detection in dense crowd	Masjid al-Haram
2013 [20]	Kinect + HOG + SVM	Overcome occlusion and overlaps	Constraint to kinect camera	Real time video
2013 [21]	HOG + SIFT + MRF	Count in extremely dense crowd	No explicit density function, count in still images only	50 crowd images with 64 K annotated Humans
2010 [22]	SURF + SVR	In several cases more accurate than Albiols [23]	Problems due to partial occlusions and perspective	PETS2009
2014 [24]	Channel state information (CSI)	Device free crowd counting	Cannot work well in a dim or dark environment	Own data set
2015 [25]	Multi-view head shoulder detection	Both static and moving crowd count	Limited to sparse crowd, poor performance on people with strange clothes	Indoor videos: classroom, meeting room
2015 [26]	Part based detection	Detect partially visible humans	Difficult in low resolution, camera position	Internet source: flickers
2013 [27]	CNN + Deep learning	Introduction of a novel loss function	Time consumes for training	Millions of images for a very deep network
2015 [28]	Deep convolution neural network (CNN)	Solve the cross-scene crowd counting problem	Pre-trained CNN model required	UCF CC 50 dataset, UCSD dataset

$$Fo(p) = \frac{1}{2\pi\vartheta^2}\sum_i \exp\left(\frac{\|\epsilon_i - \mathrm{p}\|}{2\vartheta^2}\right)$$

where ϑ is size of feature in feature map, ϵ is ground-truth annotations of feature positions and p is pixel in an image. In [1], object tracking algorithms are comprehensively covered along with taxonomy of approaches with some references on crowd tracking.

4.3 Crowd Behavior Analysis

To understand and monitor the crowd behavior is still a challenge despite the various advances in human behavior analysis. Abnormal behavior can be defined in various ways due to its personalized essence. It has been seeding much confusion in the literature. Some researchers describe the abnormality in terms of frequency. The event which occurs infrequently is called as abnormal or which happens rarely. Crowd can be classified as structured or unstructured as shown in Fig. 3. It is easy to analyze structured crowd but an unstructured crowd is very dangerous due to random motion. In behavior understanding we mainly focus on velocities, direction of flow and abnormal events like fighting, running etc.

Validation of crowd behavior is again a challenging problem because ground truth video footage containing specific abnormal behaviors in the typical crowd are not easily accessible and available. To resolve the problem of validation of ground-truth video, Andrade et al. [30], [8] achieved controlled situation with known ground truth data set to test their algorithm by using crowd simulation algorithm. However, [8] discriminate normal and abnormal behavior of crowd by exploring some concepts which are related to crowd simulation. They divided crowd behavior analysis approaches in object-based and holistic based. Object based approach defines crowd as a collection of individual person whereas in a holistic approach, focus on individual difference is ignored. This approach considers the entire individual in a crowd to have similar motion characteristics.

Fig. 3 Illustration of crowded scenarios: **a** structured event, **b** unstructured event

Table 2 Crowd video dataset

Dataset	Description	Size	Label	Accessibility
UCF	Videos of crowds, vehicle flows and other high density moving objects	38 video	Partial	Yes
UMN	Scenarios of 3 different indoor and outdoor events	11 videos	All	Yes
UCSD	34 training video clips and 36 testing videos	98 video clips	Partial	Yes
CUHK	1 traffic video sequence and 1 crowd video sequence	2 videos	Partial	Yes
QMUL	3 traffic video sequence and 1 pedestrian's video sequence	4 videos	Partial	Yes
PETS2009	8 video sequences of different crowd activities with calibration data	8 videos	All	Yes
Rodriguezs	Large collection of crowd videos with 100 labeled object trajectories	520 videos	Partial	No
UCF	Behavior image sequences from the web videos and PETS2009	61 sequences	All	Yes
Violent-flows	Real-world video sequences of crowd violence	246 videos	Partial	Yes

5 Crowd Video Dataset

When we deal with crowd video analysis, we need to validate our results with some real-world database for which some datasets are publically available and accessible. Table 2, illustrates some set of real-data bases with its brief descriptions which includes size, label and accessibility.

6 Summary and Future Scope

This paper explores the various aspects related to crowd modeling and crowd analysis which uses various techniques for various applications in the real word. A detailed comparison of the state-of-the-art methods related to people counting has been summarized along with its advantages and shortcomings. This survey presents a futuristic view of the crowd monitoring and classification. An integrated framework is required for crowd management which can deal with any type of crowd analysis ranging from the panic situations to the large scale misbehaved crowds. It is done in order to identify the common lacunae of the existing techniques and to overlay a path for the further research in this area. Though, various researches have been concluded but some issues are still untouched, which demand further research as:

- Multi-sensor Information Fusion.
- Tracking-Learning Detection Framework.
- Deep Learning with neural network for crowd analysis.
- Wireless Sensor Networks for density estimation.
- Real-Time Processing and Generalization.

References

1. Zhan, Beibei, et al. "Crowd analysis: a survey." Machine Vision and Applications 19.5–6 (2008): 345–357.
2. Li, Teng, et al. "Crowded scene analysis: A survey." Circuits and Systems for Video Technology, IEEE Transactions on 25.3 (2015): 367–386.
3. Hu, Weiming, et al. "A survey on visual surveillance of object motion and behaviors." Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 34.3 (2004): 334–352.
4. Badal, Tapas, et al. "An Adaptive Codebook Model for Change Detection with Dynamic Background." 2015 11th International Conference on Signal-Image Technology and Internet Based Systems (SITIS). IEEE, 2015.
5. Leggett, Richard. Real-time crowd simulation: A review. R. Leggett, 2004.
6. Hu, Min, Saad Ali, and Mubarak Shah. "Detecting global motion patterns in complex videos." Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE, 2008.
7. Wang, Xiaofei, et al. "A high accuracy flow segmentation method in crowded scenes based on streakline." Optik-International Journal for Light and Electron Optics 125.3 (2014): 924–929.
8. Mehran, Ramin, Akira Oyama, and Mubarak Shah. "Abnormal crowd behavior detection using social force model." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.
9. Kratz, Louis, and Ko Nishino. "Tracking pedestrians using local spatio-temporal motion patterns in extremely crowded scenes." Pattern Analysis and Machine Intelligence, IEEE Transactions on 34.5 (2012): 987–1002.
10. Kratz, Louis, and Ko Nishino. "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.
11. Cong, Yang, Junsong Yuan, and Ji Liu. "Abnormal event detection in crowded scenes using sparse representation." Pattern Recognition 46.7 (2013): 1851–1864.
12. Zhou, Bolei, Xiaogang Wang, and Xiaoou Tang. "Random _eld topic model for semantic region analysis in crowded scenes from tracklets." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.
13. Bak, Slawomir, et al. "Multi-target tracking by discriminative analysis on Riemannian manifold." Image Processing (ICIP), 2012 19th IEEE International Conference on. IEEE, 2012.
14. Kuo, Cheng-Hao, Chang Huang, and Ramakant Nevatia. "Multi-target tracking by on-line learned discriminative appearance models." Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010.
15. Badal, Tapas, Neeta Nain, and Mushtaq Ahmed. "Video partitioning by segmenting moving object trajectories." Seventh International Conference on Machine Vision (ICMV 2014). International Society for Optics and Photonics, 2015.
16. Fruin, John J. Pedestrian planning and design. No. 206 pp. 1971.
17. Chen, Ke, et al. "Feature Mining for Localised Crowd Counting." BMVC. Vol. 1. No. 2. 2012.

18. Bansal, Ankan, and K. S. Venkatesh. "People Counting in High Density Crowds from Still Images." arXiv preprint arXiv:1507.08445 (2015).
19. Hussain, Norhaida, et al. "CDES: A pixel-based crowd density estimation system for Masjid al-Haram." Safety Science 49.6 (2011): 824–833.
20. Tian, Qing, et al. "Human detection using HOG features of head and shoulder based on depth map." Journal of Software 8.9 (2013): 2223–2230.
21. Idrees, Haroon, et al. "Multi-source multi-scale counting in extremely dense crowd images." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013.
22. Conte, Donatello, et al. "A method for counting moving people in video surveillance videos." EURASIP Journal on Advances in Signal Processing 2010 (2010).
23. Albiol, A., Silla, M.J., Albiol, A., Mossi, J.M., 2009. Video analysis using corner motion statistics. In: IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, pp. 31–38.
24. Xi, Wei, et al. "Electronic frog eye: Counting crowd using WiFi." INFOCOM, 2014 Proceedings IEEE. IEEE, 2014.
25. Luo, Jun, et al. "Real-time people counting for indoor scenes." Signal Processing (2015).
26. Fradi, Hajer, and Jean-Luc Dugelay. "Towards crowd density-aware video surveillance applications." Information Fusion 24 (2015): 3–15.
27. Lebano_, Logan, and Haroon Idrees. "Counting in Dense Crowds using Deep Learning".
28. Zhang, Cong, et al. "Cross-scene crowd counting via deep convolutional neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
29. Rodriguez, Mikel, et al. "Density-aware person detection and tracking in crowds." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.
30. Andrade, Ernesto L., Scott Blunsden, and Robert B. Fisher. "Modeling crowd scenes for event detection." Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. Vol. 1. IEEE, 2006.