Crowd behavior analysis: A review where physics meets biology

Ven Jyn Kok, Mei Kuan Lim, Chee Seng Chan *

Center of Image and Signal Processing, Faculty of Computer Science & Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

A R T I C L E   I N F O

Article history:
Received 28 May 2015
Received in revised form
17 September 2015
Accepted 2 November 2015
Communicated by Paolo Remagnino
Available online 2 December 2015

Keywords:
Crowd behavior analysis
Biologically inspired
Physics-inspired
Computer vision
Survey

A B S T R A C T

Although the traits emerged in a mass gathering are often non-deliberative, the act of mass impulse may lead to irrevocable crowd disasters. The two-fold increase of carnage in crowd since the past two decades has spurred significant advances in the field of computer vision, towards effective and proactive crowd surveillance. Computer vision studies related to crowd are observed to resonate with the understanding of the emergent behavior in physics (complex systems) and biology (animal swarm). These studies, which are inspired by biology and physics, share surprisingly common insights, and interesting contradictions. However, this aspect of discussion has not been fully explored. Therefore, this survey provides the readers with a review of the state-of-the-art methods in crowd behavior analysis from the physics and biologically inspired perspectives. We provide insights and comprehensive discussions for a broader understanding of the underlying prospect of blending physics and biology studies in computer vision.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The one who follows the crowd will usually go no further than the crowd; the one who walks alone is likely to find herself in places no one has ever been before, Albert Einstein. While this quote is lived by many, this paper is motivated by the contrary. Our work is based on the notion that literally, one follows the crowd will surpass solitary individual, and together with the crowd, ‘venture beyond places’ where no lone individual is capable of venturing to; a phenomenon known as the emergent behavior. Emergent behavior arises in a swarm or crowd with certain class of entities (e.g. insects, human, animals, etc.); whereby, each entity is self-organized and together they portray a complex and coordinated collective behavior. The essence of the emergent behavior is based on a simple rule of thumb, where entities engage with one another using basic interactions. This in turn heightens ones’ sense of responsiveness to the surrounding, and instantaneously brings them closer to their goal. What makes it interesting is that, this resultant phenomenon is not possible to be achieved by solo individuals.

Over the past years, biologists have observed the emergent of collective behaviors in organism, insects and animals and were constantly investigating the underlying mechanism that allows unity in a swarm [1–4]. For example, a school of fish that swims together and yet not colliding with each other, or a flock of starlings steering in the air with the uncanny synchronization. The slime mold that exist as a single-cell organism, congregate to form multicellular when food supplies is scarce, working in tandem to search for the shortest path to food source. Another well-known example is the foraging activity of a colony of ants. Although each ant follows a set of simple rules, the colony as a whole, acts in a sophisticated way that increases its foraging efficiency [5]. Fascinatingly, this similar behavior has been observed in human crowds as well. Amongst the early works that were motivated by the emergent behavior in human crowds was the concept of the 'mind' by Bon in [6] which stated that, when individuals in a crowd gather and coalesce, a new distillation of traits emerged. He referred to the emergent behavior as collective ‘unconsciousness’ that robs every individual member of their opinions, values and beliefs. He put forward that the emergent behavior is very subtle and ignorant to each individual, yet, is capable of forming intriguing collective ‘group mind’ that works wonders. This phenomenon can be seen commonly in a crowded scene. For instance, when two flows of people moving in the reverse directions, a uniform walking lanes for each direction would be formed spontaneously although there is no communication amongst the individuals in the crowd.

In existing literature, the dynamics of human crowd are often studied through analogies with theories in physics and biology. The idea of relating the motion of crowd with fluid, liquid or electrons in aerodynamics, hydrodynamics or continuum mechanics respectively, has generated many research in crowd analysis since the past years [7,8]. Accordingly, physics-inspired studies assume that the individual in a crowd tends to follow the dominant flow of the crowd and
thus, the motion of highly dense crowd resembles fluid. Hence, theories and methods in fluid mechanics are adopted to comprehend the flow of human crowd. In another physics-inspired example, the kinetic theory of gases is applied to model the sparse and random interaction forces amongst individuals in a crowd. On the contrary, from the biology point of view, individuals in a crowd resemble the entities in a swarm. Each individual in the swarm exhibits diverse interaction forces towards achieving the final goal, which is apparently common amongst members in the swarm [9,10]. For example, the motion of individuals in a train station, where everyone is moving with different pace towards the common exit region, or the diverse motion of individuals finding their ways to the boarding area.

Nevertheless, there is no clear distinction between the approaches inspired by the two sciences: physics and biology. Instead, we observe that some terminologies or notions from both approaches share interestingly similar understanding and perspective, while holding onto some minor differences. The study of the human crowd behavior from the perspectives of the two sciences drawn into the field of computer vision is a new and rapidly developing study [11]. It is predominantly deemed as a notion for crowd behavior analysis to enhance and assist the analysis of visual crowd surveillance, which aims to imitate the human visual perception. The capability to emulate human visual perception allows the development of practical systems that provide meaningful and concise description of crowd behavior, to better assist human in crowd surveillance, which is the focal interest of this study.

1.1. Comparisons with previous reviews

Although there have been great interest and a large number of methods have been developed for crowd analysis in general, there are limited comprehensive reviews which focused on crowd behavior understanding [12]. Most existing survey papers [12–17] focus on the computer vision techniques and review the essential features required for application specific crowd analysis. To the best of our knowledge, none of the aforementioned reviews provide in-depth discussion from the perspectives of physics or biologically inspired approaches in the context of crowd behavior analysis.

The closest attempt to bridge the studies between physics and biology in the context of crowd behavior understanding was by Hughes [18]. His work emphasizes on the key distinctions between physics and the actual crowd. Although the discussion was focused only on crowd modeling from the physics perspective, the concept that described the ‘thinking’ component of fluids spurred thought that the interactions between individuals in a crowd is far more complex than particles in fluid. This coincides with the understanding of crowd motion in biology. Another work in [19] categorized the state-of-the-art methods in crowd simulation into three broad approaches which include (i) fluids, (ii) cellular automata and (iii) particles. He suggested the classification of existing work without discussing much on the underlying motives and attributes between these categories. In addition to the 3 broad categories proposed by Leggett [19], Zhan et al. [13] reviewed approaches to infer crowd events by further dividing the ‘particles’ category into agent and nature-based models; leading to 4 categories of crowd models from the non-vision approaches. This includes (i) physics-inspired, (ii) agent-based, (iii) cellular automation and (iv) nature-based. While their work acknowledged the advantages of integrating the non-vision models with computer vision methods for crowd analysis, the in-depth discussion on the different non-vision models from the physics and biology perspectives is lacking. Thida et al. [12] presented a review with systematic comparisons of the state-of-the-art methods in crowd analysis, where the merits and weaknesses of various approaches were discussed comprehensively. Their work is based on the three distinct philosophies for modeling a crowd by Alexiadis et al. [20], where crowd models are categorized as microscopic, mesoscopic and macroscopic. The microscopic model deals with the crowd as discrete individuals while the macroscopic model treats the crowd as a unit. The mesoscopic model combines the properties of the former two models, that is, the microscopic state of pedestrians are maintained with an addition of the general view of crowd. Yet, the gap between the two approaches has not been discussed clearly.

Other papers are more specific towards understanding crowd behavior, disregarding the point of whether the different methods of analysis are inspired by the studies from physics or biology. Each of the works provides critical outlook of existing literature pertaining to the different aspects of crowd analysis and serves as a reference point to all computer vision practitioners in the domain. However, we observed that a great deal of them are focused on physics-inspired approaches. Helbing et al. [21] discussed their analysis on using density and pressure attributes to infer two new phenomena in crowd: the stop-and-go and turbulent flows. Their discussions are highly influenced by physics and provide readers with insights to where and when accidents tend to occur in crowded scenes, and on how the proper management of crowd can ensure prevention of crowd disasters. In another review that is based on the notion that individuals in crowds behave in ways like particles in the fluid is by Moore et al. [22]. Their work adopted the concept of scale in hydrodynamics (the study of liquid in motion) as opposed to the common adaptation of aerodynamics (the study of gaseous or air in motion). The main difference between the two is that in the former, the interaction forces between individuals in the crowd tend to dominate the motion of the individuals, while in the latter, the interactions between individuals are few and random motion is most likely to dominate the crowd behavior. In a more recent review, Jo et al. [23] briefly highlighted the difference between physics-based and physics-inspired methods. Accordingly, physics-based methods are rooted in fundamental physic ideas whereas the latter are inspired by the laws of physics. In [10,24], the limitations of existing physics-inspired models to describe pedestrian behaviors and crowd disasters are discussed comprehensively. This includes the difficulty to capture the complexity of crowd behaviors using a single model and the insufficiency of current models in understanding the interactions between individuals and their environment. Thus, they introduced the integration of cognitive science and physics for a more holistic solution. Some examples of the heuristic rules which are derived from the natural cognitive of human include the assumptions that an individual tends to move towards a possible entry or exit, and that an individual is very likely to move its motion according to his or her gaze angle. Interestingly, the introduction of such simple rules adheres to the concept of emergent behavior, where the collective dynamics of a social system with many interacting individuals can be modeled through simple rules. A more comprehensive review of physics-inspired crowd models covering the 3 main aspects of crowd motion pattern segmentation, crowd behavior recognition and anomaly detection can be found in [17]. While this review provide broad discussion on existing models, algorithms and evaluation protocols of research in crowd, the outlook of computer vision approaches from the perspectives of physics and biology remains unstated. Other relevant researches include the study on crowd dynamics and how the different dynamics of crowd can lead to the various issues in crowd safety by Johansson et al. [25], the modeling of crowd dynamics from the viewpoint of mathematics [26], the analysis of human behaviors from the perspectives of social signal processing [27], the study of crowd dynamics from the psychology perspective by Reicher [28], the underlying rules that lead to collective behaviors for group intelligence problem-solving by Fisher [29] and the comprehensive review on the basic laws of physics and mathematics that describe collective motion which leads to the emergent behavior in groups of animals or humans [30].
<table>
<thead>
<tr>
<th>Paper</th>
<th>Author</th>
<th>Title</th>
<th>Description</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>[18]</td>
<td>R.L. Hughes</td>
<td>The flow of human crowds</td>
<td>Presented the notion that crowd is a continuum with the capability to think. The crowd motion can be represented as ‘thinking fluids’</td>
<td>2003</td>
</tr>
<tr>
<td>[19]</td>
<td>R. Leggett</td>
<td>Real-time crowd simulation: A review</td>
<td>Reviewed the research on real-time crowd simulation, focusing on the approaches that modeled pedestrian as fluid, cellular automata or particles. Also, describe the implementation of CrowdSim model to simulate crowd motion</td>
<td>2004</td>
</tr>
<tr>
<td>[13]</td>
<td>B. Zhan, D.N. Monekosso, P. Remagnino, S.A. Velastin, and LQ. Xu</td>
<td>Crowd analysis: A survey</td>
<td>Surveys crowd analysis methods in computer vision, covering approaches of crowd density estimation, crowd tracking as well as pedestrian and crowd recognition. Discusses crowd model from the perspective of other research discipline and the potential to integrate with computer vision for crowd modeling and events inferencing.</td>
<td>2008</td>
</tr>
<tr>
<td>[29]</td>
<td>L. Fisher</td>
<td>The perfect swarm: The science of complexity in everyday life</td>
<td>Explores the collective behavior emerged from a set of very simple rules of interaction between neighboring entities (specifically insects such as locusts, bees and ants). Focus on the development of group intelligence in human crowd to solve complex problems</td>
<td>2009</td>
</tr>
<tr>
<td>[14]</td>
<td>J.C.S. Jacques Junior, S. Raupp Musse, and C.R. Jung</td>
<td>Crowd analysis using computer vision techniques</td>
<td>Overview of computer vision techniques for crowd analysis, specifically on people tracking, crowd density estimation, event detection, validation and simulation. Also, describe the correlation between crowd simulation and analysis to deal with challenges in crowd analysis</td>
<td>2010</td>
</tr>
<tr>
<td>[22]</td>
<td>B.E. Moore, S. Ali, R. Mehran, and M. Shah</td>
<td>Visual crowd surveillance through a hydrodynamics lens</td>
<td>Reviewed hydrodynamics-based techniques to model the interaction forces between individuals in crowd for crowd analysis in visual surveillance of high-density crowd</td>
<td>2011</td>
</tr>
<tr>
<td>[30]</td>
<td>T. Vicsek and A. Zafeiris</td>
<td>Collective motion</td>
<td>Reviewed the observation and basic laws of collective motion which is on the borderline of several scientific disciplines</td>
<td>2012</td>
</tr>
<tr>
<td>[12]</td>
<td>M. Thida, Y.L. Yong, P. Climent-Perez, H.L. Eng, and P. Remagnino</td>
<td>A literature review on video analytics of crowded scenes</td>
<td>Reviewed on the state-of-the-art approaches in automatic crowd video analysis, by emphasizing on the macroscopic modeling, microscopic modeling and crowd event detection</td>
<td>2013</td>
</tr>
<tr>
<td>[16]</td>
<td>C.C. Loy, K. Chen, S. Gong, and T. Xiang</td>
<td>Crowd counting and profiling: Methodology and evaluation</td>
<td>Reviewed the state-of-the-art approach for video imagery based crowd counting with emphasis on the methodologies and systematic evaluation of different techniques</td>
<td>2013</td>
</tr>
</tbody>
</table>
To the best of the authors’ knowledge, there is no review on biologically inspired algorithms for crowd analysis in computer vision. This is rather surprising, given the plethora of methods that apply biological concepts for crowd analysis today [31–34]. A summary of the existing surveys are presented in Tables 1 and 2.

1.2. Motivation and contributions

A thorough exploration on the current literature provides a broader outlook on the potentials of computer vision to cut across disciplines, especially in the two areas of sciences: physics and biology, to further enhance the efficiency of video surveillance of crowded scenes. Therefore, in this paper, we focus primarily on the attributes of crowd behavior from the study of physics and biology in computer vision. We discuss existing computer vision solutions that integrate these attributes for crowd behavior analysis. Specifically, this paper provides comprehensive review of the state-of-the-art computer vision methods in crowd behavior analysis from the physics and biologically inspired perspectives. We put forward the underlying prospects of leveraging the studies from the two sciences and bringing together different disciplines in the hope of, and envisaging computer vision to the next level. As shown in Fig. 1, to ease understanding and improve readability, we begin with comprehensive introduction of the key attributes: common and conflicting attributes of crowd behavior from the perspectives of the two sciences. This is then followed by detailed discussion on the state-of-the-art crowd analysis applications and the commonly used benchmark dataset in each respective area. Here, we discuss the applications of computer vision in the perspective of their key attributes. The applications are divided into three common tasks in computer vision generally: (i) crowd segmentation, (ii) crowd dynamic analysis and (iii) crowd density estimation. It is important to note that this study seeks to initiate the outlook of computer vision solutions from the viewpoint of physics and biology. This is as opposed to previous surveys, especially computer vision related reviews such as in [17], that focuses on computer vision techniques (e.g. feature representation and model learning) in crowd analysis. Ideally, this study hopes to spark interest in integrating multiple disciplines for the advancement of crowd analysis in computer vision.

The remainder of this paper is organized as follows: Section 2 introduces the terms swarm, crowd and their relation to computer vision. In Sections 3 and 4, we outline the various attributes of crowd and discuss their similarities and differences with modeled crowd behavior. Section 5 presents a summary of the state-of-the-art computer vision applications, particularly in the branch of crowd behavior analysis. In addition, we discuss these applications with regards to the shared attributes between the two sciences to ease understanding of the concept of emergent behavior. Section 6 provides the opinions of the authors with regards to the forthcoming of a multidisciplinary crowd behavior analysis. Finally, in Section 7, we would conclude with our insights on the potential of spanning the distance between the physics and biologically inspired approaches in computer vision, for crowd behavior analysis and understanding in particular.

Table 2
The disciplines and criteria emphasized in the review papers on crowd behavior analysis. Note that item without check-mark indicates that the topic is not discussed comprehensively in the respective review paper, but may have been mentioned intrinsically in the context.

<table>
<thead>
<tr>
<th>Year</th>
<th>Paper</th>
<th>Discipline of discussion</th>
<th>Aspect of discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Physic</td>
<td>Biology</td>
</tr>
<tr>
<td>2003</td>
<td>[18]</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>2004</td>
<td>[19]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2008</td>
<td>[13]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2009</td>
<td>[29]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2010</td>
<td>[14]</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2011</td>
<td>[22]</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>2011</td>
<td>[15]</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2012</td>
<td>[30]</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2013</td>
<td>[12]</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2013</td>
<td>[23]</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>2013</td>
<td>[16]</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2014</td>
<td>[17]</td>
<td>✓</td>
<td>–</td>
</tr>
</tbody>
</table>

Fig. 1. The overall organization of this review.
2. Swarm, crowd and computer vision – the beginnings

2.1. What is a swarm or crowd?

A swarm or crowd is generally referred to as a collection of spatially proximate class of entities. Most commonly, the entities are associated to insects such as bees and ants, thus elucidating the enormous research efforts in swarm intelligence that are motivated by biology. For example, the well-known Ant Colony Optimization algorithm that was inspired by the foraging behavior of a colony of ants and the Particle Swarm Optimization algorithm which simulates the synchronized movement of a flock of birds. Beekman et al. [35] advocate that swarm intelligence is biology. On the other hand, the physicists have for a long time used the term swarm to refer to particles or electrons to describe transport equations and fluid flows [36–38]. Examples of swarm are as shown in Fig. 2. For decades, the biologist and physicists have been studying the behaviors of social animals to investigate the underlying mechanism that allows unity in a swarm. Only in the late-1980s that computer scientists begin to discover the potentials of the swarm intelligence and proposed scientific insights of these algorithms into the different applications in the field of computer vision such as robotics [39,40], optimization [41,42] object tracking [43,44]. In fact, the great leap forward was made by Reynolds [45], where he created a computer model known as Boids, comprising a large group of virtual agents that mimicked the coordinated movement of a flock of birds. This simulation applied simple rules to control the steering behaviors of its agent: separation (keeping some distance from other agents), alignment (move at a velocity that matches with local flock mates) and cohesion (move towards the average position of local flock mates). The graphical representation of the rules applied by Boids is as shown in Fig. 3. Since then, great strides have been taken to investigate if the emergent of complex behaviors is indeed caused by simple rules and interactions amongst individuals in a large group, mainly through simulations [9]. In the same way, the simulated model by Helbing and Molnar [9], and the Boids has sparked a wide interest in the adaptations of the emergent behaviors to solve complex computer vision problems.

2.2. What is the role of computer vision in crowd analysis?

One particular field of interest in public security involves visual surveillance of mass gathering of people, such as during public assemblies (e.g. music festivals and religious events) and demonstrations (e.g. strikes and protests) as illustrated in Fig. 4. The security of mass crowd during public events has always been of high concern to relevant authorities due to the dynamics and degeneration risk. Any crowded environment has high tendency to plunge into a panic atmosphere given physical stress (i.e. overcrowding) or sudden external pressure (i.e. shootings and fire), where the consequences are often devastating. Various examples from historical incidents have shown how things can easily get out of control when mass of people come together during big events. Some examples of disasters that have happened in the past are as shown in Fig. 5.

One must understand that in crowded scenes, where crowds of hundreds or even thousands gather, video monitoring is a daunting task. Often, incidents within a crowd went unnoticed due to the inherent limitations from depending solely on manual monitoring by CCTV operators. The limitations are commonly due to (i) sheer number of screens to be monitored, (ii) boredom and human fatigue, (iii) distractions and interferences, and (iv) the complexity and uncertainty of human behavior. In most scenarios, the consequences of not being alert of overcrowding and fail detection of suspicious activities may ultimately lead to
unfavorable incidents which are irreversible and catastrophic. Table 3 lists some cases of crowd disasters at mass gathering events. Still [46] and Soomaroo and Murray [47] provide comprehensive summary of crowd disaster.

Carnage in crowd happens for a variety of reasons and have seen a two-fold increase in the past two decades [48,49]. The aftermath investigations surrounding most of the crowd disasters conclude that there were missed opportunities to use technology for crowd behavior understanding to achieve proactive surveillance [50]. Hence, the recent years have seen significant advances in using computers and technologies, specifically in the field of computer vision, to assist humans in the task of video monitoring for a more efficient and proactive crowd surveillance (as shown in Fig. 6). This survey draws a bead on physics and biology inspired methods for crowd analysis, which is a challenging research topic in computer vision.

3. Common attributes of crowd

In this paper, the primary focus is behavior analysis of human crowds. From the literature, it is observed that studies from the two sciences, biological and physics inspired approaches for crowd analysis in computer vision are generally based on the common attributes: decentralization, collective motion and emergent behavior. The underlying idea of crowd behavior exploited in the respective approaches is that the emergent behavior is caused by basic interactions between individuals without any force of leadership. Complex collective motions are formed, and this in turn, resulted in emergent behavior.

3.1. Decentralized

Decentralized decision making is the concept of having no leaders, where there is no centralized control structure to dictate how individual agents should behave. The decision arising from a process of decentralized decision making are often related to the functional result of group intelligence and crowd wisdom in the domain of biology, or moving particles in physics. Generally, the decentralized mechanism requires positive and negative feedback, amplification and multiple interactions between agents to establish a collective unconscious that allows emergent behavior [68]. Table 4 shows examples of the overlapping criteria from the perspectives of the two sciences: biology and physics.

The decentralized behavior or decision making mechanism is indeed a distinct attribute in most of the physics and biologically inspired approaches in computer vision that give rise to emergent behavior. Simulations and observations of insect, animal and human behaviors in biology and particles in physics have supported this notion. One of the common scenarios to depict the decentralized aspect in crowd behavior is the formation of lanes (unidirectional or bidirectional) in crowded areas such as malls.

Unrelated individuals in crowd are able to create smooth traffic flows without collision by having minimal interactions according to simple rules. Individuals within a crowd use repulsive forces to reach their destination. They stay close to the shortest route to their destination. They stay close to the shortest route to their destination. They stay close to the shortest route to their destination. They stay close to the shortest route to their destination.

Similar models which simulate this decentralized behavior have been adapted to achieve effective evacuation planning [69,70]. Nonetheless, there are several researches that contradict the decentralized attributes. Crowd intelligence is disregarded by introducing leaders to include subgroup behavior [71,72] or to allow analysis of evacuation efficiency [73–75]. Fig. 6 illustrates the graphical deinition of centralized and decentralized decision making. We reckon that there is still a gap between centralize/ decentralize crowd behavior model in computer vision and real world crowd scenarios.

3.2. Collective motion

A comprehensive review on the observations and basic laws to describe collective motion from mathematics and physics is as presented in [30]. Here, the collective motion is considered as a phenomenon occurred due to ordered macroscopic behavior of constituent
Table 3
Examples of crowd disasters at mass events.

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
<th>Place</th>
<th>Description</th>
<th>Casualties</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ibrox disaster (football match)</td>
<td>January 1971</td>
<td>Glasgow, UK</td>
<td>Fire started deliberately in the alcove</td>
<td>66 deaths, 140 injured, 48 injured</td>
<td>[51]</td>
</tr>
<tr>
<td>Hillsborough disaster (football match)</td>
<td>April 1989</td>
<td>Sheffield, UK</td>
<td>Crush due to overrunning surge against barrier</td>
<td>142 deaths, no data available for injured</td>
<td>[52]</td>
</tr>
<tr>
<td>Orkney stadium disaster</td>
<td>January 1991</td>
<td>Shetland Islands, UK</td>
<td>Progressive crowd collapse caused by the sheer number of pilgrims</td>
<td>286 deaths, 28 injured</td>
<td>[53]</td>
</tr>
<tr>
<td>New Year’s eve stampede Lan Kwai Fong, Hong Kong</td>
<td>January 1993</td>
<td>Hong Kong, China</td>
<td>Crush due to sudden panic during fireworks display</td>
<td>11 deaths, 207 injured</td>
<td>[54]</td>
</tr>
<tr>
<td>Hajj disaster Mecca, Saudi Arabia</td>
<td>May 1994</td>
<td>Mecca, Saudi Arabia</td>
<td>Progressive crowd collapse caused by the sheer number of pilgrims</td>
<td>266 deaths, 98 injured</td>
<td>[55]</td>
</tr>
<tr>
<td>Akashi pedestrian bridge accident</td>
<td>July 2001</td>
<td>Akashi, Japan</td>
<td>Crush due to sudden panic during fireworks display</td>
<td>74 deaths, 627 injured</td>
<td>[56]</td>
</tr>
<tr>
<td>Miyun lantern festival disaster</td>
<td>July 2004</td>
<td>Beijing, China</td>
<td>Crush due to sudden panic during fireworks display</td>
<td>74 deaths, 627 injured</td>
<td>[57]</td>
</tr>
<tr>
<td>Philsports stadium - Manilla Philippines</td>
<td>February 2006</td>
<td>Manila, Philippines</td>
<td>Crash due to sudden panic during fireworks display</td>
<td>74 deaths, 627 injured</td>
<td>[58]</td>
</tr>
<tr>
<td>Love Parade disaster Duisburg, Germany</td>
<td>July 2010</td>
<td>Duisburg, Germany</td>
<td>Crash due to sudden panic during fireworks display</td>
<td>74 deaths, 627 injured</td>
<td>[59]</td>
</tr>
<tr>
<td>Khmer water festival Phnom Penh, Cambodia</td>
<td>November 2010</td>
<td>Phnom Penh, Cambodia</td>
<td>Crash due to sudden panic during fireworks display</td>
<td>74 deaths, 627 injured</td>
<td>[60]</td>
</tr>
<tr>
<td>Boston marathon bombing</td>
<td>November 2011</td>
<td>Boston, USA</td>
<td>Two pressure cooker bombs exploded near the finishing line, where the crowd of spectators gathered</td>
<td>37 deaths, 347 injured</td>
<td>[61]</td>
</tr>
<tr>
<td>Love Parade disaster Duisburg, Germany</td>
<td>July 2010</td>
<td>Duisburg, Germany</td>
<td>Crush due to sudden panic during fireworks display</td>
<td>37 deaths, 347 injured</td>
<td>[62]</td>
</tr>
<tr>
<td>Khmer water festival Phnom Penh, Cambodia</td>
<td>November 2010</td>
<td>Phnom Penh, Cambodia</td>
<td>Crash due to sudden panic during fireworks display</td>
<td>37 deaths, 347 injured</td>
<td>[63]</td>
</tr>
<tr>
<td>Boston marathon bombing</td>
<td>November 2011</td>
<td>Boston, USA</td>
<td>Two pressure cooker bombs exploded near the finishing line, where the crowd of spectators gathered</td>
<td>37 deaths, 347 injured</td>
<td>[64]</td>
</tr>
</tbody>
</table>

Moussaid et al. [101] proposed two heuristics that are based on cognitive information to model the desired walking directions, speeds of individuals and physical contacts between individuals in a crowd. They propose using combination of these behavioral heuristics with contact forces for a large set of complex collective dynamics. What makes this work interesting is the proposal of integrating a cognitive science approach into the commonly physics inspired attributes for a more realistic modeling of collective social behaviors, in particular of human crowds. In another variation by Helbing and Molnar [9], the motion of individuals is described as if they are subjected to a set of rules known as the ‘social force’. The ‘social force’ is not directly exerted by the individuals’ personal environment. Instead, it is a measure for the unique motivation of each individual in performing certain actions. For example, each individual motion is driven by the nearest entries or exits in a train station, and the tendency of keeping a distance or gap with other individuals to avoid collision. This method is influenced by the Newtonian mechanics in physics. Promising results from the various studies provide evidence that the interactions between individuals in a group are indeed simple, although their resulting collective patterns are highly complex.

Subsequently, a number of higher-level applications of crowd analysis such as detection of abnormalities within dense crowd scenes are developed by exploiting the notion of collective motion [65,88–94] as illustrated in Fig. 8. Their works assume that the motion of individuals tends to follow the collective flow of the crowd, and any deviations from the observed collective patterns are deemed as abnormal. Despite the promising results in estimating abnormality as well as in reproducing the observed features of collective crowd, behaviors of individuals constituting a crowd are unique in nature and thus, affect the scalability and robustness of crowd models. Their dynamics, which comprise the local and global interactions amongst the group and environment further complicate analysis. A summary of the understanding of collective motion from the perspectives of biology and physics is described in Table 5.

3.3. Emergent behavior

Emergence is the process of complex pattern formation from simple rules. Thus an emergent behavior or emergent property can arise when a number of individuals operate collectively in an environment, forming complex behaviors. A critical distinction between the emergent behavior and collective motion is that the former is a phenomenon resulting from collective motion, while the later describes the self-organization pattern of crowd. In both nature and engineering, complex behaviors can emerge as a result of distributed collective processes or collective group behavior. The physicists, biologists, philosophers and computational scientists have for a long time study the emergent properties in their respective domains to shed light into the problem of understanding emergent behavior [95]. From the physics point of view, the emergence of complex behaviors has been observed in photons and electrons in quantum systems and fluid dynamics [37,96–99]. Here, the collective phenomenon observed in macroscopic systems is brought about by the microscopic constituents with one another and with their environment. A biological example of the
emergent behavior is an ant colony. An ant as a single entity has limited memory and is capable to perform only simple actions. However, an ant colony expresses a complex, collective behavior which provides intelligent solutions to problems such as finding the shortest route from the nest to a food source [5,100–102]. Fig. 9 illustrates an example scenario of the emergent behavior in human crowd. Meanwhile, Table 6 describes examples of emergent behavior in physics and biology.

The underlying theory that we observe to be consistent across the wide range of domains is that, the emergent of the highly complex behavior in a collection of individuals (i.e. animal, electron, neuron, and particle) is often a result of individuals following a set of simple rules. Furthermore, these interactions are established to be decentralized, without any external coordination or the active role of a leader. Another characteristic that is consistent across experts from different areas is that emergent behavior is often unpredictable and unprecedented. Work that focuses on the emergent behavior of crowd includes [88], where a linear approximation of the dynamic system is used to categorize varying emergent crowd behaviors based on eigenvalues over an interval of time. Their method classifies crowd behaviors into bottleneck, lane, arch, fountainhead and blocking (as illustrated in Fig. 10). Likewise, Allain et al. [103] focuses on seven key crowd behaviors generated using Lagrangian forces as shown in Fig. 11. Despite the promising results, questions may arise on the adequacy on imposing selected predefined ‘patterns’ and set of ‘rules’ in existing models to capture the complexity of real-world scenarios.
4. Conflict attributes of crowd

In the research of crowd behavior analysis, abstractions are made regarding motions of individuals in crowd. The two main aspects discussed in this section is the tendency and capability of individuals to think and have preferences in the crowd.

4.1. Bias or non-bias

Biologically, entities in crowd have natural tendency or pre-dilection. A fairly good example is the schooling species of golden shiners (Notemigonus crysoleucas) where they have pre-existing bias towards yellow targets. Similarly, every individual is exclusive...
where each has their own unique behavior and inclination. Individuals in crowd are ‘bonded’ by one common focal points [105], in which each individual has low relatedness with varying self-interest [106]. Hence, the task to understand and model the varying predilection and behavior of each entity is complex.

Nevertheless, when individuals are integrated as a crowd, they unconsciously alter their behavior in line with the responses of neighboring entities. As exhibited in [106], despite the presence of powerful minority with strong inclination, with sufficient amount of uninformed entities, the crowd will come to a majority decision.

In another work by Klucharev et al. [107], they found that when individuals are made aware of the ‘opinions’ of the crowd, individuals will adjust their judgments to align with the opinions of the general norm of crowd, neuronal responses are triggered and manifested in the rostral cingulated zone and ventral striatum in the brain [107], which leads to tendency to adapt.

The natural responses of a human crowd to conform to social norm and the general crowd makes it tolerable to simplify and relax the ‘biasness’ at individual level when modeling and understanding crowd behavior in both biological and physic inspired approaches. Crowd behavior models emphasize more on the collective effect of entities – emergent behavior [109]. In [110], an algorithm to track the path maneuvered by individuals in a dense crowd based on the observation that locomotive pattern of an individual is in line with the collective patterns of the crowd and the layout of the environment is proposed. Alternatively, in [85], the idea of conformity of individual to crowd is used to infer the past and future behavior of individual. Then again, to analyze the crowd behaviors in a real-world scenario, questions might
The negative impact of the emergent behaviors such as herding and clogging, a traditional two exits may lead to herding behavior and clogging at one of the exits. By contrast, preliminary results that are worthy of future investigations in order to understand the contrast between the individualistic and collective behavior. Their findings provide preliminary results that are worthy of future investigations in order to understand the ‘thinking’ and ‘non-thinking’ components as well as the ‘bias’ and ‘non-bias’ aspects of the crowd; with the aim to provide better solutions that are able to mitigate the negative impact of the emergent behaviors such as herding and clogging, towards creating a safer environment.

arise as up to which extent an individual is conforming to the norm of general crowd.

4.2. Thinking or non-thinking

Originally, works on crowd behavior are frequently associated with the laws of physics, presuming that it is sensible to view the crowd as a homogeneous mass of bodies [111,112]. Each individual within the crowd is characterized as a non-thinking particle which has no tendency and capability to make decisions. The motions and directions of each particle are dictated by the external forces (i.e., boundaries, neighboring particles, etc.) [22]. For instance, motion of mass of individual is equated with the flow of fluid steered by the pathway with the rate of motions computed analog with the law of fluid dynamic.

In 1995, Sime [111] questioned the practicality of representing human in crowd as non-thinking particles (or ball bearings); forsaking the rules of behaviors. Often, the minute behaviors and reactions of individuals within the crowd to surrounding are the vital interactions that affects the crowd motion as a whole. For example, a person who fell can become an obstacle to a smooth flow of a crowd, that may deteriorate and lead to the occurrence of stampede and death. The lack of behavioral complexity in crowd models makes it an imprecise description of ‘real life’ crowd flow.

With that being said, many new approaches integrated physic and psychological aspects in crowd behavior analysis; exploring crowd behavior from both macroscopic (crowd as a whole) and microscopic (interactions between individuals) perspective. Among the new approaches is the model proposed by Helbing et al. [32]. The authors modeled crowd behavior by simulating the tendency of individuals to conform to social norm and reactions (i.e. panic) evoked as response to the layout and atmosphere of the environment.

Nevertheless, computational complexity is directly proportional to the density of a crowd. There is a dire need to achieve equilibrium between the understanding that individuals within a crowd are capable of making decision and the incorporation of the amount of influence an individual behavior has to the others in the crowd. Fig. 12 demonstrates an example scenario where the behaviors of crowd in the real-world differ from the common assumption of physics inspired models. This stir thoughts and interest on the need to understand further the varying perspectives on the attributes of crowd (bias or non-bias, thinking or non-thinking).

5. Applications and dataset

Crowd behavior analysis and understanding is a subject of great scientific interest that involve multidisciplinary field, including computer vision, psychology, sociology, physics and civil engineering to understand the multifaceted of the study of crowd. Despite the possibility of being an inexhaustible source of research, it is motivated by the need for an enhanced public safety in the society [113].

In the following section, we will discuss the public datasets corresponding to the three branches of applications of crowd behavior analysis: (i) crowd segmentation, (ii) crowd dynamic analysis, and (iii) crowd density estimation. In particular, the discussion is driven by the methods to model the collective motion and the contextual understanding of the emergent behavior for various applications.

5.1. Dataset

As crowd behavior analysis in the field of computer vision prosper, public datasets start to gain importance in the vision community to meet different research challenges. Refer to Tables 7 and 8 for a complete list of the currently available datasets.

Although different researches from varying scientific fields are trying to analyze the same physical entity (i.e. crowds of individuals), the approaches have advanced independently [114]. As such, different techniques, comparative comparisons and benchmark datasets developed are characteristically of its own; resulting in the difficulty to summarize the evaluation protocol and performance comparisons in this area. In an effort to make the next great leap forward in crowd behavior analysis, we strongly believe that there is a need of a common platform to evaluate the various analytical aspects of crowd.

5.2. Crowd segmentation

Generally, works in crowd segmentation assume that crowd is an agglomeration of pedestrians [85]. Even though each individual has their own goal destination and motion tendency, they appear to share common motion dynamics when observed over time in a crowded scene. This is due in part to the tendency of individuals to follow the dominant flow owing to the physical structure of the scene, and the social conventions of the crowd dynamics [90]. Therefore, crowd segmentation is commonly the basis of many abnormal event detection systems, whereby finding the abnormal regions in a given scene is accomplished by discovering the deviation from the regular flow, also known as coherent motion, as stored in the crowd motion model. In some scenarios, crowd segmentation is applied prior to estimating the density of crowd [126].

The main focus of crowd segmentation methods is on grouping regions with similar motion dynamics or coherency [127]. Most often, rather than computing the trajectories of individuals (microscopic), holistic approaches (macroscopic) in crowd segmentation methods build a crowd motion model using instantaneous motions of the entire scene such as the flow field [65,128,129]. There are, however some work which is based on
tracking individuals and accumulating their trajectories over a period of time to obtain coherent motion [85]. Tracking approaches, regardless of whether they are using distance or model-based representations are very challenging in crowd scenes [130]. This is because the trajectories are highly fragmented with many missing observations due to the complex interactions, occlusions.
between individuals in the crowd and background clutters. Therefore, tracking in crowded scenes often incorporate scene or contextual information for accurate trajectory estimation [131]. Although coherent motion is generally the macroscopic and microscopic observations of collective movements of individuals, recent studies show that it can also be characterized by mesoscopic models. Mesoscopic approaches such as in [85,132,130] model crowd activities using the interactions between an individual and its local neighborhood. In [133], fragments of trajectories over a short period of time, known as tracklets are used to analyze motion coherency. The general view of crowd is nevertheless retained, by incorporating the information on the sources and sinks. Their model encourages the tracklets to have the same sources and sinks as the segmented regions. Despite the fact that

<table>
<thead>
<tr>
<th>Dataset name</th>
<th>Established year</th>
<th>Dataset reference</th>
<th>Crowd segmentation</th>
<th>Crowd dynamic analysis</th>
<th>Crowd density estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCF dataset – crowd segmentation</td>
<td>2007</td>
<td>[65]</td>
<td>(Qualitative evaluation)</td>
<td>(Qualitative evaluation)</td>
<td>-</td>
</tr>
<tr>
<td>UCF dataset – crowd tracking</td>
<td>2008</td>
<td>[110]</td>
<td>Person tracking accuracy</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>UCSD Pedestrian Dataset</td>
<td>2008</td>
<td>[115]</td>
<td>Marathon-1: 74.9% (143/199 individuals)</td>
<td>–</td>
<td>Mean squared error</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Marathon-2: 97.5% (117/120 individuals)</td>
<td></td>
<td>Away: 4.181 Towards: 1.291</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Marathon-3: 76% (38/50 individuals)</td>
<td>False positive rate: 0.936</td>
<td>Absolute error Away: 1.621</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>True positive rate: 0.936</td>
<td>Towards: 0.869</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Area under ROC: 0.9727</td>
<td></td>
</tr>
<tr>
<td>QMUL Junction dataset</td>
<td>2008</td>
<td>[116]</td>
<td>–</td>
<td>Anomaly detection</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Area under ROC: 0.9765</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UMN Crowd dataset</td>
<td>2009</td>
<td>[119]</td>
<td>–</td>
<td>Anomaly detection</td>
<td>–</td>
</tr>
<tr>
<td>UCSD Anomaly Detection dataset</td>
<td>2010</td>
<td>[120]</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Equal error rate: 25%</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Crowd and group data</td>
<td>2011</td>
<td>[121]</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Global anomaly detection (Area under ROC)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>UMN dataset: 0.9961</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Prison riot dataset: 0.8903</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>UCF dataset: 0.986</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PETS 2009: 0.9414 (scene 1), 0.9914</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(scene 2)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Data-Driven Crowd dataset</td>
<td>2011</td>
<td>[122]</td>
<td>Tracking typical crowd behavior</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean tracking error: 47.47 ± 1.27 pixels</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tracking rare/abrupt events</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean tracking error: 46.88 pixels</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Train Station dataset</td>
<td>2012</td>
<td>[85]</td>
<td>(Qualitative evaluation)</td>
<td>–</td>
<td>Mean absolute error: 3.15</td>
</tr>
<tr>
<td>Mall dataset</td>
<td>2012</td>
<td>[123]</td>
<td>–</td>
<td>–</td>
<td>Mean squared error: 15.7</td>
</tr>
<tr>
<td>AGORASET Violent-Flows dataset</td>
<td>2012</td>
<td>[103]</td>
<td>–</td>
<td>–</td>
<td>Mean deviation error: 0.0586</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>[124]</td>
<td>Crowd violence video classification: 81.3 ± 0.21%</td>
<td>–</td>
<td>Absolute Difference/image: 419.5 ± 541.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Crowd violence detection: 88.23%</td>
<td>–</td>
<td>Normalized Absolute Difference/image: 31.3 ± 27.1</td>
</tr>
<tr>
<td>UCF dataset – crowd counting</td>
<td>2013</td>
<td>[66]</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Group state analysis</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normalized mutual information: 0.48</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>CUHK Crowd dataset</td>
<td>2014</td>
<td>[125]</td>
<td>Purity: 0.78</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rand index: 0.83</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Average accuracy: 60%</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Crowd video classification</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Average accuracy: 70%</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Crowd saliency detection</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(No. of detection/Labeled region)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Crowding: 12/13 (1 missed detection)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sources &amp; sinks: 14/19 (5 missed detection)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Local irregularity: 47/43 (2 missed detection, 6 false detection)</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

Table 8
Publicly available datasets introduced for specific crowd behavior analysis with the reported quantitative results from the corresponding reference.
Input sequence: Crowd in the pilgrimage or Hajj scene.

Example of the segmented regions, where regions with similar behavior in the Lyapunov sense are merged as proposed in [65]. The different colors in the results represent different flow segments.

Input sequence: Crowd in the marathon scene.

Example of the segmented regions, where the region growing scheme is used to segment crowd flow based on its flow field as proposed in [129]. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Fig. 13. Examples of the crowd segmentation results from state-of-the-art methods. (a) Input sequence: crowd in the pilgrimage or Hajj scene. (b) Example of the segmented regions, where regions with similar behavior in the Lyapunov sense are merged as proposed in [65]. The different colors in the results represent different flow segments. (c) Input sequence: crowd in the marathon scene. (d) Example of the segmented regions, where the region growing scheme is used to segment crowd flow based on its flow field as proposed in [129].

A similar crowd detection algorithm proposed by Reisman et al. [134] works in the spatio-temporal domain, the system relies on the inward and intersection motions of opposite moving individuals to infer presence of crowd. Examples of the segmentation results from related works are as illustrated in Fig. 13.

While the earlier discussed works are fixated on segmenting coherent motions as a cue of crowd on videos or image sequences, another variant perform crowd detection on still images as proposed in [135]. Their work allows discrimination between crowd and non-crowd regions by utilizing low-level local feature from single crowd image. Responses for each pixel are defined using pyramid pixel-grid approach to exploit the properties of crowd (at narrow scale, the basic element should resemble a human; whereas at large scale, crowd regions exhibit repetitive features of individuals in crowd). Sample results are shown in Fig. 14. Similarly, Fagette et al. [136] proposed an unsupervised method using multi-scale texture-based features to detect and localize dense crowds in images. The unsupervised method allows detection on images without the need to have prior knowledge of the scene or context. In another variation, Idrees et al. [137] proposed a new direction to localize crowd segments by detecting humans in dense crowds. Their detection method uses the collective crowd attribute based on the observation that scale of individuals in local neighborhood is similar.

Regardless of the low-level features used to segment regions of crowd, crowd segmentation methods are indirectly based on the assumptions that there is indeed collective motion between individuals in the crowd. The collective motion between individuals that move together with consistent speed and motion direction describes the self-organization pattern of crowd and can be observed commonly in public spaces such as mall and underground station. In crowd segmentation, the phenomenon arising from the collective homogeneous and coherent motion such as the formation of unidirectional lanes and shortest path is the emergent behavior. In real world applications, the emergent behavior changes according to time (e.g., traffic flow and crowd density increases during peak hours) and thus, segmentation methods must be flexible enough to cope with the differing motion activity.

5.3. Crowd dynamic analysis

The formation of crowd and mass gathering often poses challenges to public safety if it is not handled effectively, particularly when panic arises among surging individuals [138]. Therefore, amongst the major goal of computer vision systems is to detect and analyze the motion dynamics of crowded scenes, in the hope towards profiling and identifying salient motion behaviors which could lead to potential unfavorable events. In the literature, there has been a variety of terms used to refer to salient motion behavior including interesting, irregular, suspicious, anomaly, uncommon, unusual, rare, atypical and outliers. The definition of salient motion behavior has been causing much debate and confusion in the literature due to the subjective nature and complexity of human behaviors. In particular, they can be categorized into 2 broad understanding, where an event is considered salient if: (i) there is deviation from the ordinary observed or learned events (i.e., the event having low occurrence or statistical representation in the learned model) and (ii) the event is not known or it is outstanding.

Researchers have found that saliency can be identified and localized by exploiting the motion dynamics in crowded scene [65,89,88,90,139,140] (examples as shown in Fig. 15). In particular,
it has been observed that high motion dynamics and irregularities in the crowd motion are indeed good indicators of anomaly. Here, the high motion dynamics and irregularities constitutes to the emergent behavior. Lim et al. [90], proposed using the global similarity structure, which is a projection of the low-level representation of crowd motion to identify anomaly. Their experiments demonstrated that in dense crowd scenes such as the pilgrimage and marathon scenarios, the motions of individuals tend to follow the regular or dominant flow, resulting in stable motion dynamics. Thus, the possibility of anomalies taking place can be considered when there is high motion dynamics (unstable) and irregularities in a particular region. Some example scenarios of high motion dynamics and irregularities include stop-and-go waves, bottleneck and sources and sinks. This hypothesis again alludes to the aforementioned concept of collective motion, where individuals has high tendency to move with the crowd. In another variation in [141,142], the low-level representation are used to describe the ‘atomic’ activities while the interactions between individuals are modeled for higher level understanding of crowd activity. Under the Bayesian model, their method detects saliency assigning marginal likelihood to the quantized motion. These methods are closer to the mesoscopic understanding of crowd, where the motion of individuals is combined with their interactions within the crowd to infer anomaly. Tackling salient motion behavior of crowd scene from a different aspect, Shao et al. [125] focuses on group profiles in crowd scene based on the notion that groups are the primary entities that constitute a crowd. The proposed framework aims at understanding crowd dynamics on a group level by using a fundamental set of intra and inter-group properties comprising collectiveness, stability, uniformity and conflict. There are also other approaches that adopt learning methods to interpret crowd dynamics for saliency detection. Generally, large amount of data is required to enable good supervised/unsupervised learning for discriminative or generative crowd models. However, a major challenge in the context of crowd analysis in surveillance applications is the lack of abnormal or ground truth events for training. Andrade et al. [143] proposed an unsupervised method to model the degree of similarity between the trained model and new

**Fig. 14.** Sample results of crowd segmentation from another perspective, where regions containing crowd are segmented accordingly as proposed in [135]. The true positives are highlighted in green whereas the false positives are represented by the red areas. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

**Fig. 15.** Comparisons between state-of-the-art anomaly detectors on the corrupted marathon sequence. (a) Input sequence: Crowd in the marathon scene. The abnormal region (enclosed in the red bounding box) is simulated by inserting synthetic instabilities into the original videos. (b) Example abnormal region detected using the global motion saliency detection method based on spectral analysis as proposed in [89]. (c) Example abnormal region detected by exploiting the instability flow information as proposed in [65]. (d) Example output using the global similarity structure as proposed in [90] with the addition of discovering intrinsic structure of the motion dynamics, as illustrated in the colored regions. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)
unseen video data. Their method applied spectral clustering on the flow field information to find the optimal number of models to represent normal motion dynamics, followed by a variation of Hidden Markov Model (HMM) for learning. Another variation in [144] proposed a 3D Gaussian distributions representation of spatio-temporal motion patterns which is then fed into a variant of HMM to discover the relationships between these patterns. Saliency is defined as statistical deviations within the video sequences of the same scene. In the more recent works in [120,145], a joint models of appearance and dynamics is proposed, known as the dynamic textures (DT). Hierarchical mixtures of DT models are then performed, where the spatial and temporal saliency scores are integrated across time, space and scale with a conditional random field (CRF). Here, saliency is defined as events of low probability with respect to a model or normal crowd behavior. In [146], a non-learning method for crowd dynamic analysis is proposed, to mitigate the need of requiring a huge amount of data for accurate learning. Instead, their proposed method detects saliency by observing the deviations of features between a set of points-of-interest (POI) over a time series. In particular, the feature measurements include density, velocity and motion direction. Although the non-learning method provides convenient solution, it is restricted to a particular behavior or event such as detecting collapse flow near escalator exits and may not be ideal in dealing with the complexity of real-world scenarios. Similarly, a non-learning method based on threshold on kinetic energy of the crowd model is proposed to detect specific event such as running [147]. In addition, various unsupervised learning methods which suggests multi-level analysis (i.e. coarse-to-fine, global and local feature extractions) are proposed to deal with the complexity of crowd behavior [148,149].

5.4. Crowd density estimation

Not all events with large gathering of people are conducted in an enclosed venue with turnstiles where crowd density estimation can be administered seamlessly. And for some events such as parades or political protest, employing professionals to conduct human counting is infeasible. Nevertheless, estimating density of crowd is of utmost importance to better administer the well-being of crowd as a whole, development of public space design and accurate documentation of historical events. The Hillsborough disaster [53] is an example of the consequences of overcrowding.

On the contrary to the former two aspects of crowd behavior analysis, crowd density estimation is independent of the ‘thinking’ component of each entity in crowd. Existing work on crowd density estimation depends mainly on collective motion and appearance cues, with respect to the type of inputs (i.e. crowd video sequences or single crowd image). Different techniques are adopted to cope with crowd scene of varying density. The greater density of crowd in a scene, the more complicated the task to estimate crowd density where dynamic occlusions come into picture. It is infeasible to discern different person and ones’ body parts when a person may only be occupying few pixels [66] and further rendered by background clutter. For instance, framework that performs clustering of coherent trajectories to represent a moving entity, and inferring number of individual in the scene by Rabaud and Belongie [150], is limited to crowd scenes of sparse crowd where continuous sets of image frames are accessible. The results presented in their work illustrated that for some crowd scene where individuals are closely positioned with each other, trajectories are incorrectly merged. This is due to the phenomenon of collective motion occurring between moving interacting entities. Using an analogous perception, Li et al. [151] estimate the numbers of people in crowd by implementing foreground segmentation and head–shoulder detection approach. The proposed method was intended to address stationary crowd, where subtle motions of individual is crucial and deeply relied on in defining foreground segments. Nonetheless, the proposed framework is susceptible to inter-occlusion between individuals, particularly prominent in a dense crowd scene. Ge and Collins [152], proposed a Bayesian marked point process to detect individuals in crowd where clear silhouette of individuals is required for accurate projection to a trained set for accurate detection and counting of individuals. In another study, Ge and Collins [153] uses a generative sampling-based approach that leverage on multi-view geometry to achieve estimation of density of individuals in crowd. The work assumes that individuals in a crowd retain a certain space with each other (i.e. separation), which is one of the rules of interaction between entities in the crowd. Hence, individuals should not be occluded from all viewing angle. Example results of the aforementioned methods are shown in Fig. 16.

Alleviating the need to detect each person in a crowd, some works [66,123,154–158] uses low level crowd features (appearance cue) formed based on the collectives of crowd to estimate crowd density. Marana et al. [154] presented a method based on texture analysis to estimate crowd density, where the estimation is given in terms of discrete ranges (i.e. very low, low, moderate, high and very high). Their objective was to challenge scenes of dense crowd where each individual is greatly occluded. They assumed that crowd scene of high density tend to illustrate fine textures, whereas crowd scene of low density are mostly made up of coarse patterns. Crowd density estimation by Davies et al. [155] is one of the earliest works that uses regression approach to learn a linear relationship between low-level raw features (e.g. number of edge pixels) and crowd density. Similarly, works in [115,159] propose to extract dynamic texture from homogeneous motion crowd segments and focus on learning mapping between large set of feature responses and density. A problem commonly encountered in regression based density estimation is perspective distortion, where individuals who are closer to the camera view appear larger than those who are positioned further away from the camera. The problem is exacerbated when single regression function is used for the whole image space. To address this problem, perspective normalization plays a key role by bringing the perceived size of individuals at different depths to the same scale. Another approach is to divide the image space into different cells and each cell is modeled by a regression function to mitigate the influence of perspective distortion. Idrees et al. [66] estimate the number of individuals given single dense crowd image by leveraging the harmonic textures elements of crowd from finer scales and appearances cues to approximate the density of crowd per image patch. The system uses regression approach to infer the count of individuals per patch and multi-scale random fields to refine the counts of individuals per image. Chen et al. [123] proposed a multi-output regression approach to estimate crowd density in sparse crowd images. Low-level features extracted are shared among spatially localized regions to achieve more accurate counts prediction, indicating correlation between local regions of crowd scene is crucial. To compensate for insufficient and imbalanced training data inherent in regression approach, Zhang et al. [160] proposed to utilize a label distribution learning method. Crowd images are annotated with label distributions, and thus can contribute to the learning of its real class and the neighboring classes. Consequently, training data for each class increases significantly. Chen et al. [161] introduced an attribute based crowd density estimation framework to address the data sparsity problem inherent in regression model. Low-level features extracted from image samples are mapped onto a cumulative attribute space using multi-output regression model to exploit the cumulative dependent nature between classes. Another regression model is learned to estimate crowd density using the attributes as input.
In another variations, Lempitsky and Zisserman [162] model the density function over pixel grids, where integral over any region in the image would yield the density of object within. Kong et al. [163] uses feed-forward neural network to map the correlation between feature histogram from low-level features and number of pedestrian. Line-of-interest (LOI) counting approach in [164] regards crowd motions as fluid flow to count number of individuals crossing the detection line within a time frame using flow velocity vector and dynamic mosaics. For a more comprehensive review of works in crowd density estimation, the reader is referred to [16]. A complete taxonomy of the various approaches for crowd counting is provided and the key components of crowd density estimation framework are discussed in detail.

6. The forthcoming crowd behavior analysis

There are several aspects of crowd behavior analysis which the authors believe are at their infancy and have the potential to develop further.

Stationary crowd: Crowds may essentially develop into two types, i.e., stationary or dynamic (moving) crowds. Stationary crowds are usually found as spectators or audiences at concerts, rallies, performances and speeches. Dynamic crowds are defined as crowd which is on the move, such as pilgrims that walk around the Kaaba during Hajj.

Most of the existing work on crowd focuses on moving patterns of individuals in the scene to infer their activities. Motion is often detected by using standard approaches such as frame-differencing to more complicated techniques such as dense optical flow. The estimated motion patterns are then analyzed to deduce various suggestions on the crowd activities. On the other hand, stationary crowd analysis has never been sufficiently investigated although the non-motion characteristics can provide rich information. This counter-intuitive approach of stationary crowd analysis is based on the notion that individuals or groups that remain in a particular area for a long time are worthy of attention. An earlier work by a well-known video analytics provider, the iOmniscient in [165], provides a non-motion detection algorithm that has the ability to handle occlusion. The system is able to cope with hundreds of people moving around in a busy scene, to detect abandoned object as long as the object is visible for 50% of the time. In a more advanced and recent work [166,167], a stationary crowd analysis method is proposed to detect four major activities; group gathering, stopping by, relocating and deforming. This work alludes to the findings of [168], where their simulation on groups in crowd shows that stationary groups have greater impact on the dynamics of the scene than moving groups in some cases. This is justified further by simulating individuals forming stationary groups. The formation of stationary groups acts as an obstruction that changes the motion directions and dynamics of other individuals in the scene. Stationary crowd analysis is still at its early stage of research and is definitely worthy of upcoming investigations for a broader degree of scene understanding and traffic pattern analysis, in particular.

Still images: The study of literature in crowd behavior analysis found that most works are focused on video. Only limited areas of crowd behavior analysis are focused on using single image. Generally, the video-based crowd analysis captures motions of crowd over a duration or throughout a sequence of images. Meanwhile, image-based crowd analysis are based on a single image only, and the crowd can be stationary or dynamic. Although using video-based techniques have the advantage of utilizing temporal information, image-based methods allow separation between the
appearance cues of crowd and background clutter. The appearance cues from image-based methods therefore, can be utilized to complement the temporal information from videos for an enhanced crowd behavior understanding.

Large-scale: Most recently, the computer vision field has witnessed a great leap forward through the adoption of deep learning neural networks to solve vision problems, including face recognition, image classification and pedestrian detection [169–171]. One of the interesting features of deep learning is its capability to learn and train models based on large quantity of data, also known as the ‘Big Data’. ‘Big Data’ has earned various definitions across the different domains, and in this context it refers to the exponential growth and wide availability of digital data caused by the proliferation of CCTVs in public spaces nowadays. The explosion of videos on real-time public monitoring creates opportunities to utilize the large learning capacity of deep learning and models in the domain of video analysis. However, it is important to note that the conventional perception of using deep learning models as a black-box that is able to miraculously solve vision problems is reaching its plateau. Instead, there is a paradigm shift in the vision research community where vision problems are solved from the machine learning perspectives by casting them as high-dimensional data transform problems [172,173]. We anticipate to witness the potentials of applying deep learning mechanisms into the context of crowd behavior analysis.

Despite the robustness and flexibility of the deep learning architecture in learning an optimized feature representation automatically from the input data, it is highly dependent on the domain knowledge. This will lead to data-driven or application-driven discovery of more sophisticated deep learning models in the future. The application of deep learning to the field of video-based analysis is fairly new and amongst the earliest attempt of utilizing deep models for human crowds is presented in [174]. In [174], the proposed framework has not fully utilized the motion information in video, as the motion filters are pre-trained before jointly optimized with the appearance filters. It would be interesting to see the evolution of the current deep learning framework to deal with the temporal or motion information in video sequences, in conjunction with the context of this review. Several other attempts at using deep learning architecture for crowd behavior analysis focuses on crowd scenes understanding [175] and crowd density estimation [176]. The potentials of deep learning in transfer learning have yet to be fully realized in this field. The technical challenges imposed with this new way of looking at vision problems is relatively new and much more needed to be investigated in order to realized its potentials.

7. Conclusion

At this end, it is acknowledged that the precise resemblance or distinction between the physics and biological inspired approaches for crowd behavior analysis is rather vague. This is emphasized further by the complex nature of human behavior, especially in a crowd, and the various perceptions of it from the context of computer vision. Nevertheless, this survey aims to provide a platform within which, to address both the well explored and the neglected corners of these two sciences in the aspect of crowd behavior analysis, in particular. It is not the intention of this survey to take a stand on the implicit hierarchy of sciences which has long plagued the research scientists; with physics as the most respectable and biology as the conceptually poor cousin [177], or the recent opinion that applauds biology [178,179]. Instead, this survey observes the paradigm shift that integrates the two sciences, to some degree, and the potentials of exploiting the two opinions to advance further the field of computer vision. This survey believes that most biological processes are governed by the laws of physics, yet it does not deny the progressive use of biological metaphors to understand problems in the various domains, from computer to physical systems.

We suggest the need to bridge the different disciplines in coping with the exceedingly complex nature of human behavior. Understanding the interface between physics and biologically inspired algorithms is crucial towards developing ‘living’ computer vision systems. Recently, a new study of animal swarms has uncovered a new characteristic of their collective behavior when overcrowding sets in [180]. This study was inspired by the condensed matter models, used for example in the study of magnetism in physics. In the reverse scenario that depicts how physics methods are inspired by biology includes introducing the concept of ‘thinking particles’ for a higher level of fluid simulation [111]. In this study, we merely described the direct and indirect connections between two main disciplines: the biology and physics domain. We found that they are indeed complimentary, although the fusion of both has yet to be fully utilized in the area of understanding and analyzing crowd behaviors. We reckon that the understanding of human behavior is far more complex and limitless to be modeled by these two approaches alone, and that there is a considerable potential for multidisciplinary areas involving statistical physics, sociology, philosophy and other branches of life sciences. In this survey, we raised questions for future research, to bridge the gap between current solutions and the real world requirements. From the continuously growing number of exciting new publications on physics and biologically inspired algorithms to solve computer vision, we conclude that this is indeed an emerging field that is worthy of future investigations.

Acknowledgment

This research is supported by the High Impact MoE Grant UM, C/625/1/HIR/MoE/FCST/08, H-22001-00-B00008 from the Ministry of Education Malaysia.

References


K. Kitto, Modelling and generating complex emergent behaviour (Ph.D. thesis), University of South Australia, 2006.


