Review

Comparison of Main Approaches for Extracting Behavior Features from Crowd Flow Analysis

Zeinab Ebrahimpour [1,2,*] **, Wanggen Wan** [1,2]**, Ofelia Cervantes** [3]**, Tianhang Luo** [2,4] **and Hidayat Ullah** [1,2]

1   School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China
2   Institute of Smart City, Shanghai University, Shanghai 200444, China
3   Department of Computing, Electronics and Mechatronics, Universidad de las Américas Puebla 72810, Mexico
4   Sydney Institute of Language and Commerce, Shanghai University, Shanghai 200444, China
*   Correspondence: z_ebrahimpour@shu.edu.cn; Tel.: +86-151-210-40290

check for **updates**

Abstract: Extracting features from crowd flow analysis has become an important research challenge due to its social cost and the impact of inadequate planning of high-quality services and security monitoring on the lives of citizens. This paper descriptively reviews and compares existing crowd analysis approaches based on different data sources. This survey provides the fundamentals of crowd analysis and considers three main approaches: crowd video analysis, crowd spatio-temporal analysis, and crowd social media analysis. The key research contributions in each approach are presented, and the most significant techniques and algorithms used to improve the precision of results that could be integrated into solutions to enhance the quality of services in a smart city are analyzed.

Keywords: urban crowd flow analysis; feature extraction; spatio-temporal data; big data; social media

1. Introduction

We are living in the era of big cities, where the physical world is connected to the virtual world, and, thus concepts such as the crowd become more relevant. A group of people coinciding in the same location together is a crowd. Crowds are divided into two categories: structured and unstructured. In the former category, there is no great difference in the motion direction between members of a crowd, which is mostly in the same direction. Thus, a group contains only one main crowd behavior over time [1]. It is generally assumed that all individuals of the crowd are moving in one direction to track multiple people based on floor fields in a structured crowded scene. In unstructured crowds, participants travel in diverse directions in different spatio-temporal aspects [2]. For instance, crowds at fairs or exhibitions, stadiums, and airports are unstructured crowded scenes. Some models [3] have been proposed to employ a correlated topic model to predict the tracking of individuals in an unstructured crowded scene. Crowd behavior has become a significant concern that has motivated scholars to do research in the field of crowd behavior analysis, to exploit useful patterns that affect everyday life in a city. For instance, in September 2015, thousands of pilgrims were crushed to death in the town of Mina, Mecca, Saudi Arabia, as they were performing their prayers [4]. The celebration of the 2014 New Year's Eve in Shanghai, China, ended in tragedy on the Bund, a waterfront area that is one of the biggest tourist attractions in the city, when massive crowds of people took part in the celebrations, a planned New Year's light show. The event was ultimately cancelled, so the crowd movement of the people was out of the police's control [5]. A similar event happened at the Love Parade music festival in Germany in 2010. On July 24th, a crowd disaster, after a panic, left at least 18 revelers dead and about 500 injured seriously when thousands of people had to push through others

to pass through a weedy path and were crushed. The organizer of the festival reported that no further Love Parades would be held [6].

As a result, crowd analysis is of high importance, particularly for such fields as urban planning, smart city management, public safety, risk assessment, virtual environments, marketing, etc. In addition, urban crowd flow prediction is considered important for both traffic management and public safety, and has become a fundamental urban computing problem. One of the main devices in this area is video surveillance, which collects information from monitored activities. On the other hand, crowd flow prediction as a part of urban computing can help both policy-makers and companies like Uber and Didi create successful business synergy marketing strategies. They can take advantage of these data to balance driver supply and passenger demand, which is useful for traffic congestion and gas consumption, as well as air pollution in cities. Furthermore, if researchers can find better solutions to develop a better way to forecast crowds of people or vehicles in a city, the results will be beneficial for citizens in various aspects of their daily lives (e.g., launching emergency mechanisms, conducting traffic controls, and helping the evolution of security strategies). Undoubtedly, crowd analysis is one of the main tools for crowd flow prediction. This is a very interesting and challenging research field that has been examined via two main approaches: crowd video analysis and crowd analysis based on historical big data. The goals are to extract the abnormal behaviors that occurred in the scene, to define different kinds of events, to obtain the trajectories of motion, and to determine the features of the crowd (i.e., density and location of patterns in crowd, the movement speed of crowd, etc.). The majority of analyses have focused on computer vision applications. Regarding the application of computer vision and graphics on crowd analysis, the review shows that the computer-vision-based algorithms for the estimation of crowd density are divided into three classes: pixel-based analysis, texture analysis, and object-level analysis. Summarily, pixel-based approaches and the methods employing texture analysis explore lower-level properties in an image and do not aim at the identification of individuals in a scene. Thus, they are less precise for counting individuals. Object-level analysis is known to be a suitable way to count individuals and localization in a scene because it is mainly oriented towards person identification. Typically, this kind of analysis suits relatively denser crowds because occlusions become significant in packed crowds [7].

In this paper, three main crowd analysis approaches are considered. Section 2 considers crowd video analysis. Section 3 presents crowd analysis based on spatio-temporal data; crowd analysis based on big social media data is tackled in Section 4. For each approach, the main research contributions existing in the literature are used to underline the most significant trends in the field. A discussion comparing the existing approaches is presented in Section 5.

2. Crowd Video Analysis

In the past decade, one of the most important applications to attract significant research attention in automated crowd analysis has been video surveillance. Video surveillance plays an important role in managing and planning a city to increase the security of citizens. Managing a crowd has become a hot topic in this area to develop an effective system to control the abrupt events in public society. This requires a good understanding and application of crowd behavior analysis and crowd action recognition algorithms. The authors in [8] used the CROSS framework, which focuses more exclusively on a specific kind of behavior in crowded areas using simulation as a social scientific tool for abnormal crowd behavior prediction.

2.1. Crowd Video Behavior Analysis

A variety of algorithms have been proposed to produce reliable (valid) trajectories that can be categorized into two broad groups:

- Microscopic Modeling

This model is presented in order to analyze the behavior of each pedestrian in a crowded scene based on collective information via a holistic approach. This approach can be complemented with macroscopic modeling that simulates more realistic patterns, such as movement orientation, the location of an individual, and the interactions of people with each other.

- Macroscopic Modeling

In this model, the analysis is based on the behavior or movement of a crowd of people without considering the movement of any individual to model a typical (representative, generic) motion pattern. Macroscopic modeling has been applied to track and analyze the behaviors of people in both sparse and dense crowds using such specific properties as velocity, density, and flow.

In addition, the authors in [9] suggested a framework for video surveillance which generally classifies the model into different phases, as shown in Figure 1.
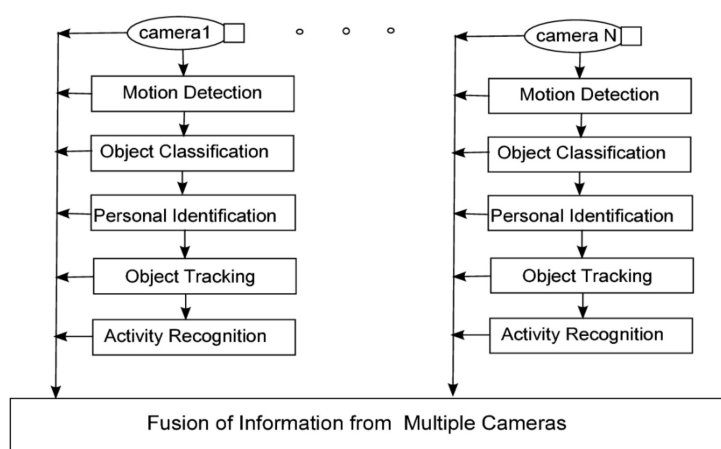

Figure 1. Weiming's proposed framework for video surveillance [9].

Their proposed model can be applied to data collected from different cameras. Each process and task is presented in detail in Table 1.

Table 1. Weiming's proposed model, including processes and tasks [9].

Process	Task
Environmental modeling	Constructing and updating background images from a dynamic sequence.
Motion segmentation	Detecting region corresponding moving objects.
Object classification	Classifying moving regions corresponding to different moving target scenes.
Object tracking	Tracking moving objects from one scene to another in the image sequence.
Behavior understanding and description	Treating a special behavior understanding problem.
Fusion of information from multiple cameras	Viewing information that can overcome occlusion.

Later, Antonakki [10] proposed a bottom up approach for motion detection, object classification, tracking, motion analysis, behavior understanding, and behavior description to classify normal and abnormal behavior using different criteria. Table 2 describes the components for the video surveillance in this model.

Table 2. Video surveillance aspects in Antonakki's model [10].

Term	Description
Motion detection	Focused on static or adaptive background subtraction or temporal differencing algorithms to separate the foreground pixels that participate in any kind of motion observed in a given scene.
Object classification	Classifying detected objects into different classes, such as humans or vehicles, that appear in a given scene.
Object detection	Locating the time and extracting the trajectories.
Motion analysis	Using motion information from a low level to identify the types of moving objects. Classifying the activities and calculating the features of motion itself. Classifying into primitive actions.
Behavior understanding	Performing the recognition of behaviors based on these feature values.
Behavior description	Recognizing behavior through nearest neighbor classification.

In real-world crowd scenes, precise segmentation is necessary, and it is a challenge to achieve this aim, especially in the field of violence recognition in video surveillance. Recently, substantial research has been conducted to achieve this aim. The authors in [11] presented Improved Fisher Vectors (IFVs) employing local features and spatio-temporal positions. This method calculates the details of features regardless of spatio-temporal locations. The authors focused on normalizing the center of each trajectory in order to stabilize the size of the feature position vector. Then, the authors input those normalized features into their model. Although the base of the IFV is in the temporal sliding window, the authors sped up the detection of violence for a range of frames by using a summed area table. Thus, their method often does not calculate the temporal segments. These results demonstrate more accurate and faster IFV performance compared to similar approaches. Although many algorithms have been proposed to better determine crowd dynamics in the field of computer vision, progress still faces different challenges including monitoring an immeasurable number of people and their activities, automated camera switching, data fusion, and complex tracking algorithms, which make a significant difference between theory and the monitoring of crowds in real life.

2.2. Crowd Video Action Recognition

Recently, crowd video analysis has been used in various applications, such as video surveillance, single or group activity and action recognition, tracking people or objects, sports video analysis, and human–computer interfaces. The process focuses first on detecting objects or individuals and then tracking them over time, to recognize crowd video action.

- Single Person Action Recognition

Many researchers have studied single person action recognition. Wang et al. [12] proposed a Latent Hierarchical Model (LHM), which is a tree-structure model with different sub-activities. Each video segment has a certain temporal scale, with start and end points that work with the temporal duration of each video segment. In order to speed up classification, they used a latent kernelized Support Vector Machine (SVM) framework. Their method has been used to recognize complex actions. In order to achieve the aim of recognizing complex actions, the authors implemented their method in two different datasets: the Hollywood2 action dataset, including a variety of actions such as driving, eating, greeting, moving, etc.; and the Olympic Sports dataset, including numerous sport actions of volleyball, basketball, bowling, etc. Laptev et al. [13] used movie scripts for the automatic annotation of human action recognition in movies. The authors used the retrieved action samples for visual learning and classified the videos. The model was built by a bag-of-features technique and features were classified using a non-linear SVM. To overcome the limitation of previous works on video datasets, they produced their own dataset of videos with scripts from movies and clustered them by using the k-means algorithm. Islam Shujah et al. [14] detected junction points by subtracting the background.

Key frames were extracted according to distinct poses and a distance-based classifier to distinguish the geometric patterns (GPs) into 8-directional classes in order to obtain the geometric pattern classes. They employed the Lucas–Kanade optical flow algorithm [15] to obtain optical flow between the video frames and achieved satisfactory average performance by evaluating their model based on famous single actions such as running, clapping, sitting down, etc., using the Weizmann dataset. Chen et al. [16] estimated optical flows with the Lucas–Kanade (LK) algorithm [17] to obtain a set of reliable points in the current frame *(ft)* and in the next frame *(ft+1)*.

In order to produce a better prediction, Chen et al. [16] compared the set of the feature points in the current and next frames to remove the set located in the same coordinates. By measuring the weighting factors for each individual and the distance between the feature points, they obtained different clusters. These clusters include similar individual patterns based on factors such as positions and orientations. Then, in the next step, the authors used the individual features to make clusters for the further analysis of their method. They proposed to use an adjacency matrix-based clustering (AMC) algorithm to get a larger cluster because it helps to better detect the abnormal behavior of crowds. Finally, the authors applied a prediction on the force field model to detect abnormal behavior based on four different dominant directions of the people that they clustered. Following the obtained clusters, if a human crowd orientation suddenly changes, it is considered as abnormal behavior (limitation: in order to improve the detection rate it is better to work more on local events). Figure 2 shows the main steps of the approach of Chen et al. [16].
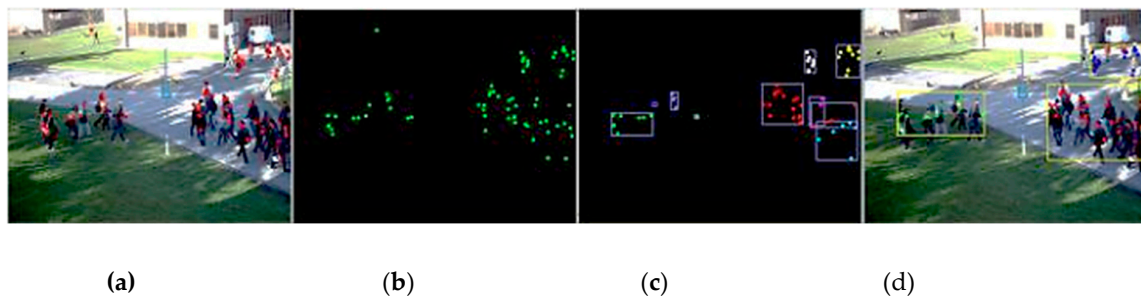

(a)   **(b)**   **(c)**   **(d)**

Figure 2. The proposed model of two-stage clustering presented in [16]: (**a**) the detected feature points; (**b**) the feature points shown without the background; (**c**) clusters obtained after applying first-stage clustering; (**d**) clusters obtained using the adjacency matrix-based clustering (AMC) algorithm.

New innovative methods of simulating individual/pedestrian movement can help to achieve better performance in a broad range of applications, but model evaluation strongly depends on a proper data source. The results of an individual simulation can be useful not only for city planners to manage better services for citizens, but also for developing commercial software such as Legion, Steps, and SimWalk. Regarding pedestrian simulation, Vizzari et al. [18] classified three main approaches to evaluate the behavior of pedestrians in the environment: 1) pedestrians as particles subject to forces; 2) pedestrians as particular states of cells; and 3) pedestrians as autonomous agents.

- Group Activity Recognition

Recently, group activity recognition has been used in a variety of fields (e.g., robotics and human interaction). Shu et al. [19] extended a two-level model toward improving the brittleness of the direct cascading in previous work with an additional energy layer (EL). By end-to-end training of the EL on top of all long short-term memory (LSTM), they captured the dependencies between all LSTM predictions. Existing datasets are too small to train LSTMs, because the feeding-forward of predictions is too brittle. In order to overcome this challenge, Shu et al. [19] took two steps to minimize the energy of all their predictions and maximize their reliability, as shown in Figure 3. Hence, their proposed model is called a confidence-energy recurrent network (CERN).
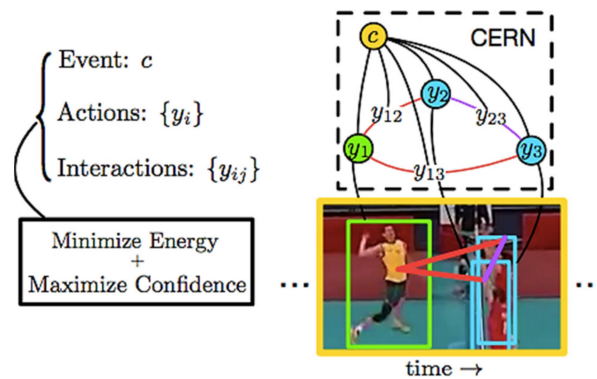
Figure 3. Two-level model by Shu et al. [19]. CERN: confidence-energy recurrent network.

Extracting motion and pose from a video in action recognition is connected with the spatio-temporal relations among people. Ibrahim et al. [20] proposed a hierarchical deep-learning-based model to determine what a group of people are doing based on a video scene. In this two-stage approach, they used three cues to recognize the activities performed by the group of people. In the first step, they recognized a single person's action, and then they used long short-term memory to present the temporal dynamics behind how a person's actions change over time. Finally, combining all of those individual representations, the authors discovered group activities in the second step. The authors implemented their model using Caffe. In order to extract the complex features from the bounding box around the person, they used the convolutional neural network (CNN) and the LSTM approaches, in addition to person trajectories. This model was evaluated not only with 1525 frames from 15 videos that were handpicked from YouTube, but also with the Collective Activity Dataset. By implementing this method on the datasets, they determined pedestrian movements such as waiting, queuing, and talking, and more complicated group activities based on volleyball videos (i.e., right/left set, spike, etc.). The results showed almost perfect performance in the dynamic properties of group activities, but due to a lack of consideration of the spatial relations between people in the group, the model became confused and boosted performance in some activities, such as crossing, waiting, and walking. However, person pooling is not able to model a group-to-group context. To solve this problem, M. Wang et al. [21] proposed a hierarchical group-to-group interaction framework implemented in the same dataset. First, the authors generated a sequence of tracked human bounding boxes. In order to partition all human tracklets into spatio-temporal groups, the authors applied clustering and segmentation methods. Next, the authors trained the model to learn inter-group interaction and intra-group human interactions. In previous works, the methods attempted to encode the high-order relationships among people in the scene by inferring the latent graphical structures. However, in this work, the authors deployed a contextual binary encoder. Previous methods were based on pairwise features to model the interactions that are difficult to generalize. Based on the two sub-actions of "move" and "pose" for each person, the authors encoded the person level features and order by the x or y coordinate of the center of the person in the image and output them in another LSTM network. Motion CNN and a pooling-SVM structure were implemented to identify actions in the scene. Based on cellular automata (CA) approaches, Bandini et al. [22] introduced the group-aware pedestrian (GA-Ped) model to simulate pedestrian crowd behavior. This simulation was implemented in an environment containing a lattice of cells. Each cell contained at most one individual who could perform a single action between that cell and the neighboring one. In order to navigate the person's point of interest, each cell was provided with specific floor fields. The authors analyzed the features related to each single person and grouped the similar movements of the people in order to build a crowd movement. They analyzed the different performances of the people as single persons or as groups of different numbers of people. They reduced the congestion of the pilgrims by simulating their model in a real-world case study of Makkah. In a video scene, the background usually changes quickly, and it is difficult to determine the temporal dynamics of the foreground. The main challenge in detecting multiple activities or

individuals is that a specific representation for one task might not be efficient for any others. In order to solve this problem, Bagautdinov et al. [23] introduced multi-scale features by concatenating multiple intermediate activation maps. Since temporal information is very important in action recognition, by using TensorFlow, the authors implemented recurrent neural networks (RNNs) to merge with the information in the temporal domain. They used the standard gated recurrent units (GRUs) for each person in the sequence, with minor modifications. Recurrent neural networks and convolutional neural networks (CNNs) are used frequently to recognize group activities, but Azar et al. [24] used multi-stream convolutional networks to capture different aspects of individual regions, as well as group activities. Spatial and temporal inputs work on different types of frames to classify the actions in RGB format in addition to a pose-map, to determine the location of the body parts of each individual. The authors implemented a linear SVM with TensorFlow for the fusion method and chose a combination of person- and scene-level predictions in RGB, optical flow, and warped optical flow, combined with the person-level predictions of the pose-map to achieve the highest performance. They compared their proposed model with previous works like CERN, and their results showed that they achieved better accuracy in the well-known Volleyball dataset [25].

3. Crowd Spatio-Temporal Analysis

Another significant source of data for crowd analysis is generated by transport, such as public buses or shared bikes that are monitored with GPS. These data allow one to analyze crowd flow in two dimensions: space and time. Big data analytics apply large amounts of data to obtain useful knowledge from hidden patterns. With current technology, we create data continuously in our daily lives. The amount of available digital information has grown rapidly since we started engaging in online shopping, communicating with friends through social networks, and using GPS or Wi-Fi on our smartphones. In this survey, we focused on data with spatio-temporal (ST) intervals based on data aspects, including spatial observations (e.g., shape, direction, or distance). Parking violation data are a typical example of ST data whose locations continuously change [26]. In this case, the traditional locations of data can be defined as objects that are collected during time intervals. A variety of domains, such as climate science, transportation, and Earth sciences deal simultaneously with the two attributes of data—space and time.

Since the dynamic ST properties constantly change, managing such properties is a challenge for aspects like clustering, predictive learning, change detection, frequent pattern mining, anomaly detection, and relationship mining. Accordingly, spatio-temporal-interval data are defined as a tuple $ST = (x; y; t_s; t_e; d)$, where x and y are the spatial information such as longitude and latitude, t_s is the start time of the event, t_e is the end time, and d is the data vector. For each pair of points $\{p;q\}$, a distance $Dist\{p,q\}$ is defined. All the points are in Euclidean space. Therefore, the direct path between two points is the shortest [27]. Spatio-temporal data have been used in various application fields, such as climate science, in which spatial dependencies with similar climatic phenomenon are defined as instances in a specific group over time. In the temporal dimension, we face different uneven time segmentations, while measuring the distance between them is an important challenge. For each piece of spatio-temporal-interval-based data, the temporal domain has a time window $\{t_s; t_e\}$ that shows the duration of an event, and there are n time windows for each point p:

$$\left\{ \left[\lambda_1{}^p . t_{s,1}{}^p . t_e \right] \left[\lambda_2{}^p . t_s, \xi_2{}^p . t_e{}^p \right], \dots \left[\lambda_n{}^p . t_s, . t_e \right] \right\} \tag{1}$$

where t_s and t_e are the start and end times of an event, respectively. As can be seen, the spatial and temporal domains comprise the two main dimensions in different spaces. A wide range of approaches have been studied, such as trajectory crowd prediction mining [27], spatio-temporal clustering [28], time series prediction [29], remote sensing, and social media analysis based on big data [30,31]. The analysis of a typical spatio-temporal database such as one related to a public transportation system like the subway can produce several benefits for citizens in different aspects of their safety and travel

route choice [32,33]. In order to provide better transportation services, especially in subway systems, different methods have been proposed to predict crowd flow [34]. Most of these strategies are based on regression strategies, like auto-regressive integrated moving averages (ARIMAs) [35] or Gaussian processes [36]. Some other methods have been applied, such as neural networks [37], probability trees [38], and wavelet-SVM [39] (e.g., for predicting the population movement in a city, the increasing likelihood of finding new passengers, or predicting traffic congestion for city designers; policy makers can also establish adequate new policies for a clean/green and invulnerable city).

Fan et al. [40] argued that the prediction of human movements is very difficult, particularly when rare behaviors that deviate from normal daily routines are taken into account. Humans' daily routines can be modeled by monitoring behavior over a long period of time. Moreover, a robust human mobility predictor must accurately handle both regular and rare crowd behaviors. As can be seen in Figure 4, there are no major differences between regular days. Accordingly, such behaviors can be readily predicted, as can be seen in Figure 5.
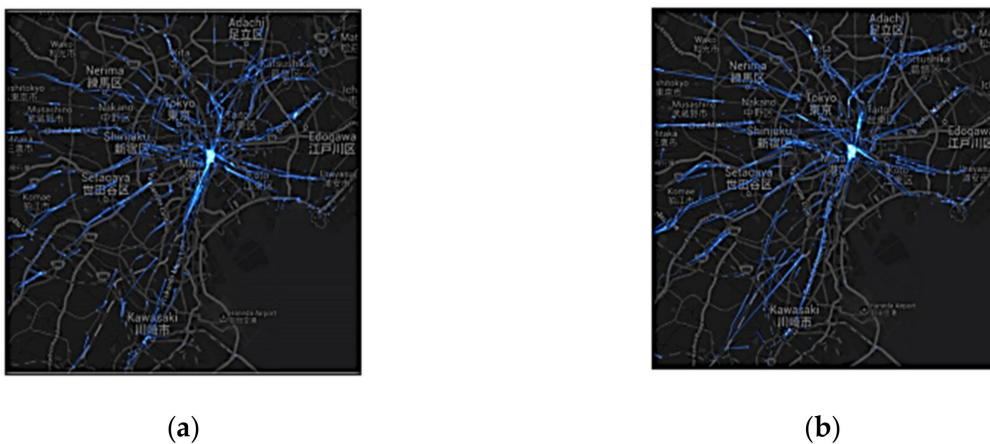

(**a**)   (**b**)

Figure 4. Regular crowd behavior: (**a**) 01 August 2011, 7–9 AM; (**b**) 02 August 2011, 7–9 AM [40].
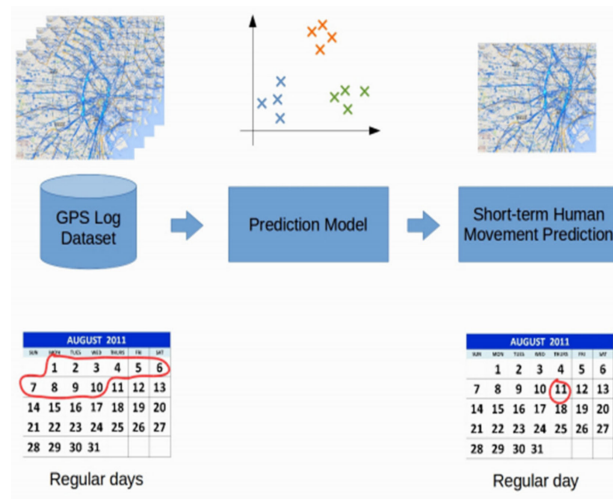

Figure 5. Prediction of regular crowd behavior.

In contrast, rare crowd behaviors show that human mobility during such events is different from that on regular days (Figure 6).

Figure 6. Rare crowd behaviors: (**a**) 2 August 2011, 7–9 AM; (**b**) 12 August 2011, 7–9 AM [40].

However, having access to a sufficient data source at the citywide level helps to better predict behavior. The researchers in [40] proposed a novel model called CityMomentum, which is a clustering-based framework. They applied Markov chain to predict the movement of each subject in each cluster using the GPS dataset. A sample of the predictions based on the most recent trajectory clustering can be seen in Figure 7.
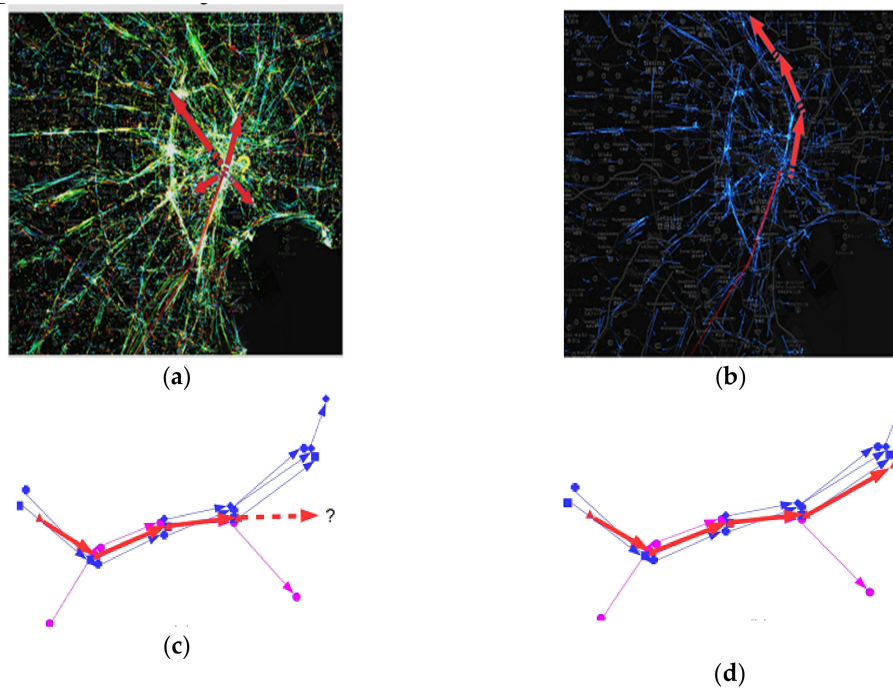

Figure 7. (**a**,**c**) Partitioning subjects' movement into clusters; (**b**,**d**) Prediction of future movements based on the best-matching cluster [40].

In another study, Zhang et al. [41] proposed the Deep Neural Network (DNN-based) prediction model for spatial-temporal data (DeepST). DeepST consists of two sections: spatio-temporal and global. The former takes advantage of a convolutional neural networks to model spatial and temporal dependencies. The latter is applied to capture global factors, such as the day of the week. The architecture of DeepST is illustrated in Figure 8.
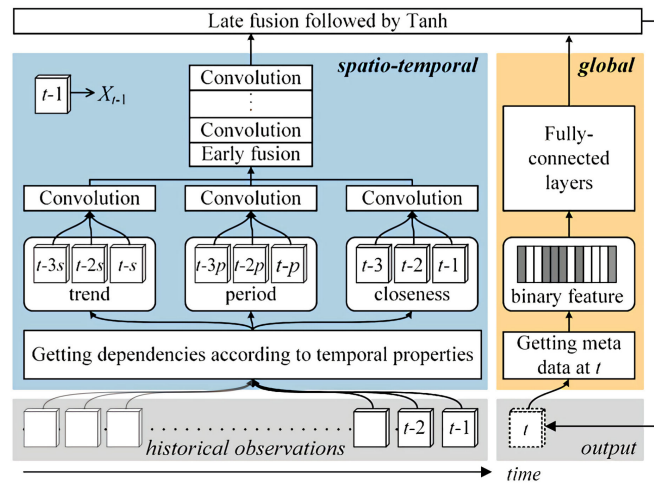
Figure 8. Architecture of DNN-based prediction model for spatial-temporal data (DeepST) [41].

Abadi et al. [42] took advantage of available historical data collected from sensors to adjust the origin-to-destination matrices. Then, they predicted the traffic flow up to 30 min in advance by using real-time traffic data. Their simulations revealed the accuracy of their proposed approach in their case studies on the city of San Francisco. The traffic flow prediction errors were in the range of 2% (5-min forecasting) to 12% (for 30-min forecasting), even with unpredictable events. On the other hand, Alahi et al. [43] quantitatively investigated pedestrians' destinations in train stations. In this proposal, a new descriptor called social affinity maps (SAMs) was employed to link the unobserved trajectories of individuals in the crowd. Their experimental findings revealed an improvement in performance upon employment of SAM features and origin–destination (OD) prior.

Silva et al. [44] proposed a new approach to analyze massive transportation systems and traffic information about individual travelers (Figure 9).
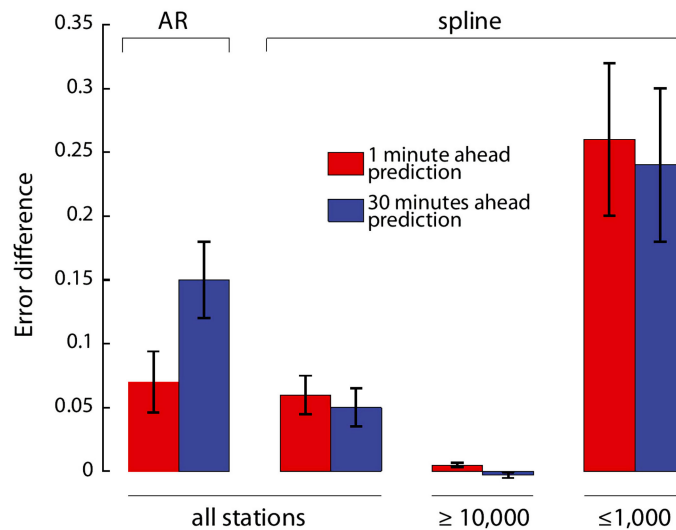

Figure 9. Fivefold cross-validation averages for 1-min-ahead (red) and 30-min-ahead (blue) forecasts.

Li et al. [45] offered a hierarchical prediction model to forecast the quantity of bikes that will be rented from/returned to each station cluster. They initially proposed a bipartite clustering algorithm to cluster the bike stations into groups and count the number of bikes rented from/returned to each group via a gradient boosting regression tree (GBRT). Another proposal for crowd flow prediction in a real-time framework was presented by Toto et al. [46], called PULSE (Prediction Framework for Usage Load on Subway SystEms). The authors extracted the profile features of subway stations, such

as the time-variant and historical traffic of peak hours. Based on the extracted features, the model was designed to select the optimal route for the passengers' target station. In a comparative study on two pedestrian monitoring techniques to predict crowd flow, Martani et al. [47] compared an array of infrared depth sensors and a visible light (RGB) camera. Their findings revealed that the developed RGB-based system performed reliably across a wide range of conditions, while the former approach was demonstrated to be a useful supplement in conditions without significant ambient sunlight, such as underground passageways. Zhengfeng et al. [48] considered multi-modal data for traffic status prediction via taxicab operating data using stacked autoencoders (SAEs). With 91% test accuracy compared to other models, such as the linear regression model and back propagation (BP) neural network, these data depicted superior performance. Some works on crowd flow prediction have mainly focused on entrance and exit passengers rather than on entire city networks, because of the complexity, unpredictability, and difficulty in dealing with real-time data. Non-negative matrix factorization (NMF) is a popular solution for network-wide issues, and online NMF showed better performance in capturing temporal changes [49]. Gong et al. [50] proposed a model taking advantage of the NMF model to capture the dynamic mobility in Sydney train stations. Although their ONMF-OA model could predict the stable flow of people, it could not capture sudden changes in flow. Thus, the authors introduced another model called ONMF-MR to develop a way to predict drastic changes in the flows. Taking advantage of strengths of both models, the authors proposed a hybrid model called ONMF-H for use in real-world applications. They proposed to measure the algorithms using the mean absolute error (MAE) and mean relative error (MRE) [51] in a real-world Opal Card dataset. They used the following expressions:

$$\text{MAE} = \frac{\sum_{i=1}^{m}|v_i - \hat{v}_i|}{m}, \qquad \text{MRE} = \frac{\sum_{i=1}^{m}|v_i - \hat{v}_i|}{\sum_{i=1}^{m}v_i} \qquad (2)$$

where $\hat{v}_i$ is a prediction, v_i is the ground truth, and m is the number of prediction flows.

In order to predict the future crowd flow in a target city via the transmission of the crow flow dynamic patterns learned through a source city at the regional level, Wang et al. [31] proposed a deep transfer learning framework called RegionTrans. Their main idea was to find inter-city region pairs sharing similar crowd-flow dynamic patterns and then use these region pairs as proxies to efficiently transfer knowledge from the source city to the target city. For this purpose, they tackled two challenges: 1. Since few crowd flow data existed in the target city, it was impossible to directly compute a reliable crowd flow similarity between a region in the source city and a region in the target city. Then, how is it possible to find strongly similar inter-city region pairs? 2. Since the available deep-learning methods are often entirely designed to predict citywide crowd flow, it is difficult to take advantage of region-level knowledge. Then, how is it possible to incorporate inter-city regional similarity information for effective deep transfer learning? The mentioned framework, RegionTrans, is believed to be novel for three reasons:

- It employs auxiliary data to obtain the inter-city region similarities associated with crowd flow dynamics.
- It represents the design of a deep spatio-temporal model with a hidden layer especially for storing region latent representations.
- It suggests a learning algorithm to transmit knowledge from a source city to a target city according to the latent representations of the inter-city similar-region pairs.

Given their evaluation results, when the recorded history of the target city was shorter, the improvement of RegionTrans was more considerable, indicating that the introduced intercity similar-region pairs are valuable for transfer learning, especially when the target data are extremely scarce. Compared to the fine-tuned DeepST and ST-Resnet (which, in the source city, provides a good starting point for optimizing network parameters), RegionTrans outperformed the mentioned baselines by further considering inter-city regional similarity information in transfer learning.

To forecast the crowd flows in a city using big data, Hoang et al. [52] proposed a scalable prediction framework that exploits multiple complex factors that are effective on crowds. Due to their limitations, a macro-level view of the crowd was investigated. This investigation was done via the prediction of two types of new-flows and end-flows in every region. In doing so, they had three major challenges: (1) the multiple complex factors that are effective on crowd flows (e.g., the effects of weather on daily routines); (2) the flow dependencies between different types of flows in intra/inter-region dependencies; (3) city-scale prediction—a prediction needs to be prepared instantly. In this way, city-scale forecasting is computationally intensive. As a result, an efficient predictive model was proposed to obtain forecasting citywide crowd flow (FCCF), as follows:

1. In order to tackle data scarcity and provide a practical citywide solution, a city was initially divided into small regions, and those specific regions with similar crowd flows were divided into different clusters.
2. Relying on intrinsic Gaussian Markov random field (IGMRF), a seasonal model to predict the periodic flow was proposed.
3. Using the region neighbor and weather information, the authors suggested a spatio-temporal model to predict the deviations.

Finally, as the main contribution, investigations into three real-world datasets (taxis and bikes) revealed that FCCF performed better in terms of accuracy compared to the baseline.

4. Crowd Social Media Analysis

In recent years, a variety of mobile apps have come to market. These apps allow users to share their location and temporal data (called check-in data). Location-based social networks (LBSNs) and geo-tagged social microblogs such as Twitter, LinkedIn, and Sina Weibo have become popular among people within the past few years. Although ST data have been used widely for different purposes in urban planning, social media data have the advantage of containing the user's interests and purposes, such as where/when/why they are going to a specific location. This massive source of data has yielded new prospective applications in crowd analysis for researchers in different areas, such as marketing and trend detections. Social media data allow researchers to investigate new insights at a higher level of analysis, for example computing individual tracks, the purpose of travel, or connection dependencies in a crowd [53]. On the other hand, analyzing the social connection between users is important to find the impact of a message to create a crowd. A large description and analysis of the literature available on how social media data is used is presented by Stieglitz et al. [54]. This process of analysis is classified into four different steps: data discovery, collection, preparation, and analysis. The same steps are necessary for crowd mobility analysis, and existing research results show that, similar to other applications, the main challenge lies in the discovery of meaningful patterns. Authors have argued that in order to overcome the challenges in analyzing social media data, the use of computer science techniques is essential. In crowd analysis, this is confirmed by the possibly random behavior of the crowd due to very complex social factors.

Liu et al. [55] proposed a model that extracts the inter-urban mobility of check-in data to explore the hidden features of a citizen's footprint when and where the check-in happened. They used a gravity model to discover the spatial interactions, formulated as

$$I_{ij} = \frac{K P_i P_j}{f(ij)} \tag{3}$$

where I_{ij} and f_{ij} denote the interaction from i to j and the distance between two places, and P_i and P_j are the repulsion of place i and the attraction of place j, respectively; particle swarm optimization (PSO) method was used to find the best fit. The authors studied their trips to China and built a spatial network where the edge weights represent the interaction strengths.

Li et al. [56] presented a paradigm to utilize check-in data with some adopted weighted standard deviational ellipse (WSDE) methods to disclose some features of the movement from the suburbs to central urban areas in different periods of time. Based on one-year social media data, the authors investigated the extraction of population features in three main railway stations in Wuhan, China. They analyzed the density of people in different time buckets (P1: 00:00–02:00, P2: 02:00–04:00). In Figure 10, the colors represent the population level; for example, red parts have a density of over 5000 and green parts have a density of less than 5000 people.
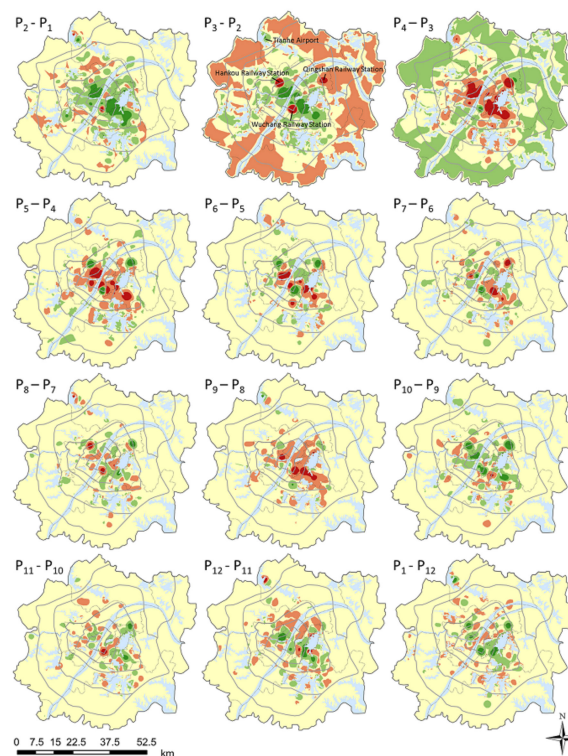

Figure 10. Distribution of the population in Wuhan [56].

Wu et al. [57] built a mechanism of an agent-based model that represents the transition of trip demands. The model integrates human activities as well as movement approaches in two different categories. The authors divided human activities into locationally mandatory activities (LMAs) and locationally stochastic activities (LSAs) to determine the exact date that people travel or move between locations. With the help of demand tags, the authors identified the aim of the trips—for example, dining, entertainment, or the temporal aspects of other groups. Since analyzing crowd mobility from social media is entirely dependent on data, the analysis and patterns cannot be used for other urban areas, which means that the specific patterns of a city are not suitable to fit into another city in order to extract features unless there will be similar patterns in the source and target data of the two different cities. Osorio et al. [58] claimed that Twitter data provide useful insights, because it is easy to use online applications. Origin–destination (OD) analysis is an efficient tool for obtaining useful information from social media. Since traditional ways of obtaining data from surveys are static, time-consuming, and expensive, a new generation of applications using data from online sources was introduced to access larger sources of data free of charge and analyze the spatio-temporal aspects of those data [59]. Compared to previous works [60,61], Osorio et al. [58] employed extra sources of data to both the origin and destination of the travel metrics in order to obtain better precision. Land Registry (cadaster) maps and population residential data from official sources (i.e., a census) from the city of Madrid were used to evaluate the origin travel metrics. On the other hand, information about workplaces was taken from the National Institute of Social Security records. After storing the data, they eliminated and

filtered the data based on different aspects, including: 1) bot tweets; 2) a selection of the weekend's tweets, filtering data from Fridays after 2 pm to the end of the holidays; 3) a selection of tweets from very similar locations; and 4) the selection of users with fewer temporal changes and a number of tweets less than five. In addition, to identify the home and workplace, they differentiated between day and night, so tweets sent during the night were considered as being sent from home. Compared to other works, an effort was made to detect the median point and calculate the closest location of a user aggregating tweet locations with the Land Registry data. The same method was employed to find workplaces based on time of day, with a time interval of 15 min. Finally, a relationship matrix was found between the home and the workplace, with the number of tweets and the number of residents by each district.

In another work, Mariano et al. [59] predicted human mobility using a hybrid dataset with a gravity model and machine learning methods. They combined Flickr data with the Airline origin and Destination Survey made by the US Bureau of Transport Statistics. In order to get two different types of mobility—flight travels and daily commutes—a collection of human movements was obtained. A pair connection was proposed between point *i*, which is the time when the user took the photo containing the geo-tagged information, and the subsequent photo taken at point *j*. The total number of users who travelled between point *i* and point *j* indicates the weight. The authors made a dataset of pictures, and each user was calculated by the total number of users who travelled between *i* and *j*. They made a dataset of pictures from Flickr to build a flow matrix by country borders defined by the Survey using the gravity model with linear regression. The radiation model proposed to obtain the total number of travelers between nodes *i* and *j* is as follows:

$$Z_{ij} = T_i \frac{m_i n_j}{\left(m_i + s_{ij}\right)\left(m_i + n_j + s_{ij}\right)} \tag{4}$$

where m_i and n_j represent the populations of the nodes, and S_{ij} is the total population.

There are a number of parameters that influence the intra-urban mobility of people, including transportation networks, economic status, different types of urban environments, and transportation. Thus, it has been revealed that each type of case study requires independent exploration to find valuable patterns for future decision-making purposes. By combining the check-in data with other sources of demographic data, more insights could be achieved.

Using a similar approach, Rashisi et al. [62] underline that obtaining data from social media is simple, but analyzing huge amounts of data to extract meaningful crowd mobility patterns remains a challenging task. To improve the quality of the results, new data-mining and semantic techniques must be applied. However, the real potential of extracting features that could be used for predicting crowd mobility behavior is yet to be explored. Some of the pros and cons of social media data sources have been underlined, such as availability, cost, preparation, land use, socio-demographics, and planning. In particular, the kind of analysis that can be applied and the knowledge that can be obtained from each data source is developed. LinkedIn is considered to be a business-oriented social networking platform and a good data source to extract useful patterns about the social and economic statuses of its users. However, social media platforms like Instagram and Foursquare provide geo-tagged data that can be used to analyze travelling behavior. In addition, by applying text mining techniques on tweets and text, more knowledge can be extracted and used for transport analysis.

An interesting approach that combines video analysis, spatio-temporal data, and social network analysis is presented by Chaker et al. [63], who produced several spatio-temporal cuboids by partitioning video scenes and detected crowd behavior using a local social network (LSN). Since each cuboid evolved with time and was updated by an LSN, detecting abnormal behavior became possible. Social media data can be also used to incorporate crowdsourcing and guide the crowd to move to safe locations in emergency situations. The research project described by Rogstadious et al. [64] integrates real-time crowdsourcing to identify the appropriate tasks and strategies in the case of a disaster. Another interesting topic in the analysis of social media data is to determine the impact of social connections on

people's mobility. For instance, Cho et al. [65] investigated the relationship between human mobility and social relationships using location-based check-in data. They found that social relationships can explain up to 30% of all human movements, while periodic behavior explains 50%–70% of movements. Social media data can be used for trend detection, marketing purposes, and route determination. With this goal in mind, Seranatne et al. [66] analyzed 26,000 tweets, specifically to determine the route to Lady Gaga's concert. Since this artist has 41.2 million followers, it is pertinent to analyze the route taken by the crowds of people going to the concert. The authors applied a Kernel Density Estimation (KDE) algorithm to the data in order to determine the hot spot of clusters and then used sentiment and drift analysis to characterize the trajectories. In another study, Hu et al. [67] focused on detecting commercial areas for business development. After cleaning the noise from the data, the authors applied exploratory spatial data analysis (ESDA) testing to determine the hot spots and used the standard deviational ellipses to determine their scope. In order to achieve more accurate results, they used building boundaries. One of the limitations of LBSN data is its position uncertainty, which requires data cleaning in a pre-processing step before further feature extraction. User authorization is a critical issue when obtaining data from an LBSN with crawler tools from various sources. Although social media data facilitate new kinds of crowd analyses, they provide reliable results if they are merged with other types of data.

5. Discussion

In this paper, we categorized existing research results of crowd flow analyses based on three types of datasets: video, spatio-temporal data, and social media data. We did not focus on traditional survey records based on questions and answers because this source of data is static, requires human resources/significant time to obtain, and was used in past decades before the emergence of new technologies. In all the analyzed works, the main challenge was the availability and quality of the dataset. Figure 11 shows a general overview of crowd analysis.
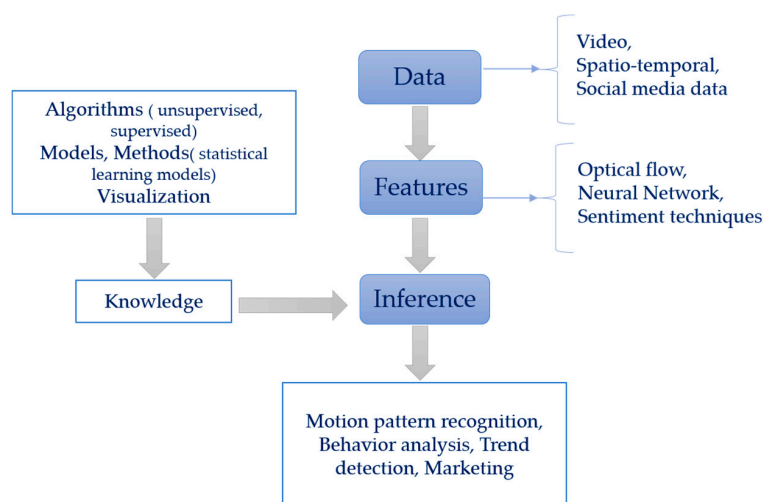

Figure 11. Structure of crowd analysis.

A significant source for visual analysis is taken from video surveillance cameras to analyze crowds in a city, thereby helping policy makers make better decisions to provide better safety and security services, greater quality in transportation systems, etc. Since these data are frequently confidential, it is a challenge for researchers to gain access to reliable datasets. In many cases, studies are accomplished to predict crowd behavior based on other kinds of datasets (due to the limitations in getting access to a dataset of video surveillance, their high cost, their confidentiality, and the small size of the dataset). This expensive traditional type of data is used to study human social behavior and interactions as single or group activities. Concerning the methodologies and techniques used for

crowd flow video analysis, different methods have been presented, including graphical models and pooling methods, to predict people's behaviors or actions at two levels: single individuals and groups of people. Energy-based models have also been used to minimize the energy in activity recognition. Recently, some researchers have applied convolutional neural networks and recurrent neural networks to accomplish a better temporal analysis. The most recent method for modeling temporal information is the multi-stream convolutional network, which takes advantage of both the CNN and RNN models to achieve better accuracy. The most common application in this field is video surveillance to detect abnormal behavior or actions, in order to provide a safer environment in stadiums, concerts, festivals, exhibitions, etc.

In the second type of source data (spatio-temporal), authors aim to study the trajectory patterns using GPS or Wi-Fi data to trace the different movements of people, thus producing a network of people who are moving in different places, trajectories, roads, etc. In trajectory-based data, the goal is to study traffic flow, in order to minimize the congestion and maximize the flow of traffic on the road. In these cases, analyzing real-time data and computation costs remains a concern.

Most of the works in this field have focused on major places in a city, such as important metro stations, well-known districts, etc., rather than on the entire city. In general, the challenge to overcome when managing spatio-temporal mobile data is the difficulty and complexity of analyzing and predicting future crowd trajectories based on real-time big data at the entire level. Concerning the methods used, RNNs have been used repeatedly to capture spatial or temporal dependencies. In order to model both dependencies, convolutional LSTMs have been introduced, but this technique has not been proven to be efficient for very long ranges of temporal data. The most recent trend in this area is to explore deep-learning methods to predict the flow of people at the city level.

Finally, at present, social media as a new source of data does not suffer from the limitations of cost and space (millions of data records can be continuously obtained and/or stored, including different types of text, photos, voice, and spatial information). In this field, most authors have tried to illustrate the differences between the human mobility behaviors of specific population groups. For example, Rizwan et al. [68] showed that in Shanghai, China, females and males do different activities and their behaviors are different during the day than during the night. The real challenge is the extraction of significant patterns from these data, to open new avenues for future research. Thus, social media data are used to analyze travel demand, the purpose of travel, and to determine the influence of social connections on human mobility or crowd movement. This could be an interesting topic not only for tourism applications but also for business development, to advertise new products that attract crowds of people to a specific location. Table 3 indicates a summary of comparison between different data types.

Table 3. Comparative summary of the different data types in crowd flow analysis.

Data Type	Purpose	Application	Method/Algorithm	Advantages	Disadvantages
Video	Security interests, anomalous behavior recognition	Video surveillance, transportation stations, malls	Graphical models/optical flow	Real-time data	High cost, confidential
Spatio-temporal data (GPS/Wi-Fi)	Crowd management, urban planning	Navigation in festivals, sports events, concerts	Machine learning/recurrent neural networks (RNNs)	More precise (~5 m error)—distinguish between transportation mode/energy usage ~50% GPS	High energy usage, signal problem/lack of large size, privacy issues
Social media	Activity participation, travel demand	Business strategy, tourism	Linguistic and text mining techniques/LDA[1]/DBSCAN[2]	Cheap, easy access, large size	Challenges in extracting useful insights, internet access

[1] Latent Dirichlet Allocation, [2] Density-Based Spatial Clustering of Applications with Noise.

As with video data, spatio-temporal data and social media data all continue to raise open questions about designing, implementing, and testing the most adequate algorithms to automatically discover patterns that will support decision making.

6. Conclusions

This survey offers the fundamentals of crowd analysis and considers three main approaches: crowd video analysis, crowd spatio-temporal analysis, and crowd social media analysis. For each type of data, several papers were considered to illustrate the main challenges and most popular solutions proposed by the authors. The main techniques used were described for each data type, and the best results were discussed. Crowd flow analysis is evolving towards crowd flow prediction. Several works are already applying advanced machine learning techniques to detect behavior patterns and predict the mobility of a crowd in real time. These results will help produce adequate and timely decision making to protect and save lives in smart cities.

References

1. Zheng, Y.; Capra, L.; Wolfson, O.; Yang, H. Urban computing: Concepts, methodologies, and applications. *Acm Trans. Intell. Syst. Technol.* **2014**, *5*, 38. [CrossRef]
2. Xue, Y.; Liu, P.; Tao, Y.; Tang, X. Abnormal Prediction of Dense Crowd Videos by a Purpose–Driven Lattice Boltzmann Model. *Int. J. Appl. Math. Comput. Sci.* **2017**, *27*, 181–194. [CrossRef]
3. Rodriguez, M.; Ali, S.; Kanade, T. Tracking in unstructured crowded scenes. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 1389–1396.
4. Hajj Pilgrimage Stampede: A Visual Guide to the Fatal Rush Near Mecca. Available online: https://www.theguardian.com/world/ng-interactive/2015/sep/24/hajj-pilgrimage-stampede-visual-guide-fatal-crush-mecca (accessed on 24 September 2015).
5. China New Year's Eve Crush: At Least 36 Killed and 47 Injured in Shanghai after 'Fake Money Thrown from Balcony of Nightclub'. Available online: https://www.independent.co.uk/news/world/asia/new-year-2015-at-least-35-killed-and-42-injured-in-shanghai-stampede-9952461.html (accessed on 31 December 2014).
6. Festivalgoers Killed in Stampede at Love Parade in Germany. Available online: https://www.theguardian.com/world/2010/jul/24/love-parade-festival-tunnel-stampede (accessed on 24 July 2010).
7. Junior, J.C.S.J.; Musse, S.R.; Jung, C.R. Crowd analysis using computer vision techniques. *IEEE Signal. Process. Mag.* **2010**, *27*, 66–77.
8. Brenner, M.; Wijermans, N.; Nussle, T.; De Boer, B. Simulating and controlling civilian crowds in robocup rescue. In Proceedings of the RoboCup, Osaka, Japan, 13–19 July 2005; Available online: https://pdfs.semanticscholar.org/891b/8d457218cf6d81b49d328c1a2be5608b123c.pdf (accessed on 7 October 2019).
9. Hu, W.; Tan, T.; Wang, L.; Maybank, S. A survey on visual surveillance of object motion and behaviors. *IEEE Trans. Syst. Man Cybern. Part C* **2004**, *34*, 334–352. [CrossRef]
10. Antonakaki, P.; Kosmopoulos, D.; Perantonis, S.J. Detecting abnormal human behaviour using multiple cameras. *Signal. Process.* **2009**, *89*, 1723–1738. [CrossRef]
11. Bilinski, P.; Bremond, F. Human violence recognition and detection in surveillance videos. In Proceedings of the 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Colorado Springs, CO, USA, 23–26 August 2016; pp. 30–36.

12. Wang, L.; Qiao, Y.; Tang, X. Latent hierarchical model of temporal structure for complex activity classification. *IEEE Trans. Image Process.* **2013**, *23*, 810–822. [CrossRef] [PubMed]

13. Laptev, I.; Marszałek, M.; Schmid, C.; Rozenfeld, B. Learning realistic human actions from movies. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008.

14. Islam, S.; Qasim, T.; Yasir, M.; Bhatti, N.; Mahmood, H.; Zia, M. Single-and two-person action recognition based on silhouette shape and optical point descriptors. 2018, 12, 853–860. *Signal Image Video Proc.* **2018**, *12*, 853–860. [CrossRef]

15. Baumann, F. Action recognition with hog-of features. In Proceedings of the German Conference on Pattern Recognition, Saarbrücken, Germany, 3–6 September 2013; pp. 243–248.

16. Chen, D.-Y.; Huang, P.-C. Dynamic human crowd modeling and its application to anomalous events detcetion. In Proceedings of the IEEE International Conference on Multimedia and Expo, Singapore, 19–23 July 2010; pp. 1582–1587.

17. Lucas, B.D.; Kanade, T. An iterative image registration technique with an application to stereo vision. In Proceedings of the Seventh International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada, 24–28 August 1981; Volume 674, pp. 121–130.

18. Vizzari, G.; Bandini, S. Studying Pedestrian and Crowd Dynamics through Integrated Analysis and Synthesis. *IEEE Intell. Syst.* **2013**, *28*, 56–60. [CrossRef]

19. Shu, T.; Todorovic, S.; Zhu, S.-C. CERN: Confidence-energy recurrent network for group activity recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5523–5531.

20. Ibrahim, M.S.; Muralidharan, S.; Deng, Z.; Vahdat, A.; Mori, G. A hierarchical deep temporal model for group activity recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1971–1980.

21. Wang, M.; Ni, B.; Yang, X. Recurrent modeling of interaction context for collective activity recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3048–3056.

22. Bandini, S.; Vizzari, G. Modeling a Crowd of Groups: Multidisciplinary and Methodological Challenges. In *Modeling, Simulation and Visual Analysis of Crowds*; Springer: New York, NY, USA, 2013; pp. 99–122.

23. Bagautdinov, T.; Alahi, A.; Fleuret, F.; Fua, P.; Savarese, S. Social scene understanding: End-to-end multi-person action localization and collective activity recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4315–4324.

24. Azar, S.M.; Atigh, M.G.; Nickabadi, A. A Multi-Stream Convolutional Neural Network Framework for Group Activity Recognition. *arXiv* **2018**, arXiv:1812.10328.

25. Muralidharan, S. A Hierarchical Deep Temporal Model for Group Activity Recognition. Applied Sciences: School of Computing Science. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.

26. Xu, Y.; Wang, B.; Kong, Q.-j.; Liu, Y.; Wang, F.-Y.; Xu, Y.; Wang, F. Spatio-temporal variable selection based support vector regression for urban traffic flow prediction. In Proceedings of the 93rd Annual Meeting of the Transportation Research Board, Washington, DC, USA, 12–16 January 2014; pp. 14–1994.

27. Ying, J.J.-C.; Lee, W.-C.; Weng, T.-C.; Tseng, V.S. Semantic trajectory mining for location prediction. In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 1–4 November 2011; pp. 34–43.

28. Li, J.; Benediktsson, J.A.; Zhang, B.; Yang, T.; Plaza, A. Spatial technology and social media in remote sensing: Challenges and opportunities [point of view]. *Proc. IEEE* **2017**, *105*, 1583–1585. [CrossRef]

29. Zhang, J.; Zheng, Y.; Qi, D. Deep spatio-temporal residual networks for citywide crowd flows prediction. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–10 February 2017.

30. Chi, M.; Plaza, A.; Benediktsson, J.A.; Sun, Z.; Shen, J.; Zhu, Y. Big data for remote sensing: Challenges and opportunities. *Proc. IEEE* **2016**, *104*, 2207–2219. [CrossRef]

31. Wang, L.; Geng, X.; Ma, X.; Liu, F.; Yang, Q. Crowd flow prediction by deep spatio-temporal transfer learning. *arXiv* **2018**, arXiv:1802.00386.

32. Yao, H.; Tang, X.; Wei, H.; Zheng, G.; Yu, Y.; Li, Z. Modeling spatial-temporal dynamics for traffic prediction. *arXiv* **2018**, arXiv:1803.01254.

33. Ni, M.; He, Q.; Gao, J. Forecasting the subway passenger flow under event occurrences with social media. *IEEE Trans. Intell. Transp. Syst.* **2016**, *18*, 1623–1632. [CrossRef]

34. Zhao, J.; Qu, Q.; Zhang, F.; Xu, C.; Liu, S. Spatio-temporal analysis of passenger travel patterns in massive smart card data. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 3135–3146. [CrossRef]

35. Williams, B.M.; Hoel, L.A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *J. Transp. Eng.* **2003**, *129*, 664–672. [CrossRef]

36. Zhou, J.; Tung, A.K. Smiler: A semi-lazy time series prediction system for sensors. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, VIC, Australia, 31 May–4 June 2015; pp. 1871–1886.

37. Wei, Y.; Chen, M.-C. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transp. Res. Part. C: Emerg. Technol.* **2012**, *21*, 148–162. [CrossRef]

38. Leng, B.; Zeng, J.; Xiong, Z.; Lv, W.; Wan, Y. Probability tree based passenger flow prediction and its application to the Beijing subway system. *Front. Comput. Sci.* **2013**, *7*, 195–203. [CrossRef]

39. Sun, Y.; Leng, B.; Guan, W. A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. *Neurocomputing* **2015**, *166*, 109–121. [CrossRef]

40. Fan, Z.; Song, X.; Shibasaki, R.; Adachi, R. CityMomentum: An online approach for crowd behavior prediction at a citywide level. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 7–11 September 2015; pp. 559–569.

41. Zhang, J.; Zheng, Y.; Qi, D.; Li, R.; Yi, X. DNN-based prediction model for spatio-temporal data. In Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Burlingame, CA, USA, 31 October 31–3 November 2016; p. 92.

42. Abadi, A.; Rajabioun, T.; Ioannou, P.A. Traffic flow prediction for road transportation networks with limited traffic data. *IEEE Trans. Intell. Transp. Syst.* **2014**, *16*, 653–662. [CrossRef]

43. Alahi, A.; Ramanathan, V.; Fei-Fei, L. Socially-aware large-scale crowd forecasting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 2203–2210.

44. Silva, R.; Kang, S.M.; Airoldi, E.M. Predicting traffic volumes and estimating the effects of shocks in massive transportation systems. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5643–5648. [CrossRef] [PubMed]

45. Li, Y.; Zheng, Y.; Zhang, H.; Chen, L. Traffic prediction in a bike-sharing system. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, VA, USA, 3–6 November 2015; p. 33.

46. Toto, E.; Rundensteiner, E.A.; Li, Y.; Jordan, R.; Ishutkina, M.; Claypool, K.; Luo, J.; Zhang, F. Pulse: A real time system for crowd flow prediction at metropolitan subway stations. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Riva del Garda, Italy, 19–23 September 2016; Springer: Cham, Switzerland, 2016; pp. 112–128.

47. Martani, C.; Stent, S.; Acikgoz, S.; Soga, K.; Bain, D.; Jin, Y. Pedestrian monitoring techniques for crowd-flow prediction. *Smart Infrastruct. Constr.* **2017**, *170*, 17–27. [CrossRef]

48. Zhengfeng, H.; Pengjun, Z.; Wenjun, X.; Gang, R. SAE for the prediction of road traffic status from taxicab operating data and bus smart card data. *Int. J. Mod. Phys. C* **2017**, *28*, 1750121. [CrossRef]

49. Pei, Y.; Chakraborty, N.; Sycara, K. Nonnegative matrix tri-factorization with graph regularization for community detection in social networks. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.

50. Gong, Y.; Li, Z.; Zhang, J.; Liu, W.; Zheng, Y.; Kirsch, C. Network-wide Crowd Flow Prediction of Sydney Trains via customized Online Non-negative Matrix Factorization. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; pp. 1243–1252.

51. Yi, X.; Zheng, Y.; Zhang, J.; Li, T. ST-MVL: Filling Missing Values in Geo-Sensory Time Series Data. June 2016. Available online: https://www.microsoft.com/en-us/research/publication/st-mvl-filling-missing-values-in-geo-sensory-time-series-data/ (accessed on 7 October 2019).

52. Hoang, M.X.; Zheng, Y.; Singh, A.K. Forecasting citywide crowd flows based on big data. In Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Burlingame, CA, USA, 31 October–3 November 2016; Available online: https://www.microsoft.com/en-us/research/publication/forecasting-citywide-crowd-flows-based-big-data (accessed on 7 October 2019).

53. Riboni, D.; Bettini, C. Incremental release of differentially-private check-in data. *Pervasive Mob. Comput.* **2015**, *16*, 220–238. [CrossRef]

54. Stieglitz, S.; Mirbabaie, M.; Ross, B.; Neuberger, C. Social media analytics–Challenges in topic discovery, data collection, and data preparation. *Int. J. Inf. Manag.* **2018**, *39*, 156–168. [CrossRef]

55. Liu, Y.; Sui, Z.; Kang, C.; Gao, Y. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *Plos ONE* **2014**, *9*, e86026. [CrossRef]

56. Li, L.; Yang, L.; Zhu, H.; Dai, R. Explorative analysis of Wuhan intra-urban human mobility using social media check-in data. *Plos ONE* **2015**, *10*, e0135286. [CrossRef]

57. Wu, L.; Zhi, Y.; Sui, Z.; Liu, Y. Intra-urban human mobility and activity transition: Evidence from social media check-in data. *Plos ONE* **2014**, *9*, e97010. [CrossRef]

58. Osorio-Arjona, J.; García-Palomares, J.C.J.C. Social media and urban mobility: Using twitter to calculate home-work travel matrices. *Cities* **2019**, *89*, 268–280. [CrossRef]

59. Beiró, M.G.; Panisson, A.; Tizzoni, M.; Cattuto, C. Predicting human mobility through the assimilation of social media traces into mobility models. *Epj Data Sci.* **2016**, *5*, 30. [CrossRef]

60. Picornell, M.; Ruiz, T.; Lenormand, M.; Ramasco, J.J.; Dubernet, T.; Frías-Martínez, E. Exploring the potential of phone call data to characterize the relationship between social network and travel behavior. *Transp. Res. Part. C Emerg. Technol.* **2015**, *42*, 647–668. [CrossRef]

61. Lee, J.H.; Gao, S.; Goulias, K.G. Can Twitter data be used to validate travel demand models. In Proceedings of the Iatbr –Windsor, Washington, DC, USA, 10–14 January 2016.

62. Rashidi, T.H.; Abbasi, A.; Maghrebi, M.; Hasan, S.; Waller, T.S. Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transp. Res. Part. C: Emerg. Technol.* **2017**, *75*, 197–211. [CrossRef]

63. Chaker, R.; Al Aghbari, Z.; Junejo, I.N. Social network model for crowd anomaly detection and localization. *Pattern Recognit.* **2017**, *61*, 266–281. [CrossRef]

64. Rogstadius, J.; Kostakos, V.; Laredo, J.; Vukovic, M. Towards real-time emergency response using crowd supported analysis of social media. In Proceedings of the CHI workshop on crowdsourcing and human computation, systems, studies and platforms, Vancouver, BC, Canada, 7 May 2011.

65. Cho, E.; Myers, S.A.; Leskovec, J. Friendship and mobility: User movement in location-based social networks. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, San Diego, CA, USA, 21–24 August 2011; pp. 1082–1090.

66. Senaratne, H.; Bröring, A.; Schreck, T.; Lehle, D. Moving on Twitter: Using episodic hotspot and drift analysis to detect and characterise spatial trajectories. In Proceedings of the 7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks, Dallas/Fort Worth, TX, USA, 4 November 2014; pp. 23–30.

67. Hu, Q.; Bai, G.; Wang, S.; Ai, M. Extraction and monitoring approach of dynamic urban commercial area using check-in data from Weibo. *Sustain. Cities Soc.* **2019**, *45*, 508–521. [CrossRef]

68. Rizwan, M.; Wan, W.; Cervantes, O.; Gwiazdzinski, L. Using location-based social media data to observe check-in behavior and gender difference: Bringing weibo data into play. *Sprs Int. J. Geo-Inf.* **2018**, *7*, 196. [CrossRef]