Multiview-Fusion-Based Crowd Density Estimation Method for Dense Crowd

Liu Bai, Cheng Wu, Yiming Wang and Feng Xie
School of Rail Transportation
Soochow University, Suzhou, P. R. China 215139
Email:20174246009@stu.suda.edu.cn, cwu,ymwang,fengxie@suda.edu.cn

***Abstract*—Crowd gathering places are prone to crowd stampede and other public emergencies, resulting in large numbers of casualties and property losses, then, leading to negative social impact. At present, the research on dynamic assessment of crowd gathering safety situation mainly relies on isolated real-time video monitoring, and lacks reliable methods to deal with plenty of video data from different sources, perspectives and granularities. Based on the traffic Internet of things infrastructure, this paper explores the fusion technology of multi-sensor source homogeneous video data. On the basis of the static model of crowd aggregation based on the high-altitude perspective, this paper studies the different source and multi granularity real-time dynamic monitoring video cooperative perception methods in the middle and low altitude and different perspectives. The dynamic scene crowd statistical perception including motion prediction mechanism is used to extract the global coarse-grained motion situation of the crowd from the perspective of high altitude. The multi column convolution depth neural network is used to extract the local fine-grained density features of the crowd with line of sight occlusion in low altitude perspective, thus establishing the holographic model of the temporal and spatial evolution of crowd situation, and proposing a new method of crowd aggregation safety situation assessment. This method is applied to the crowd gathering safety situation assessment of Suzhou city life fountain square, and achieves good results, which provides theoretical support for the safety control of crowd gathering place based on the Internet of things.**

Keywords–Crowd gathering safety situation; Video monitoring; Accident analysis and early warning; Traffic safety.

I. INTRODUCTION

When the flow of people in space is highly concentrated for a long time, the crowd density will rise sharply and the distribution is extremely unreasonable, which increases the potential safety hazards and seriously threatens the personal safety. After the spread of the flow, it will even affect the circulation and control of the surrounding traffic. Typical such events, such as the example occurring in the Shanghai Chen Yi Bund Square on December, 2014. The opposite ow of people formed a hedging caused a crowded stampede accident, resulting in 36 deaths and 49 injuries. In addition, according to incomplete statistics, from 2001 to 2014, there were more than 150 people trampling events around the world, all of which occurred in crowded places. Such incidents are sudden, complex and low-level control, which is extremely lightly to cause large-scale casualties. This also makes the prevention and research against crowded stampede accidents become the urgent needs to developing countries with rapid crowd and relatively backward management.

At present, the monitoring of the crowding degree and trend of population is beneted from the mature use of intelligent surveillance video systems, and its comprehensive perspective coverage provides more data support for crowd density estimation. Some scholars have processed the video frame image, and finally got the number of people and the density of the crowd, with good accuracy [1]. To a certain extent, such detection methods have solved the accuracy problems existing in the current crowd density detection, However, the safety of the current location cannot be determined. On the other hand, in the related research on group disaster dynamics, we are more concerned with pedestrian flow simulation and individual motion models. Helbing et al. analyzed the two phenomena of laminar flow from the laminar flow to the stop flow and turbulent flow after analyzing the video of the Mina/Mecca crowd disaster during the 1426 hours pilgrimage on January 12, 2006 [2]. Insights into the causes of these key population conditions are important for organizing safer group events. Johansson et al. discussed how to study high-density conditions based on appropriate video data on the basis of Helbing, and explained the critical conditions of crowd turbulence, and proposed corresponding measures to improve population safety [3]. Moussaid et al. proposed a cognitive heuristic based cognitive science method that predicts individual trajectories and collective movement patterns [4]. The essence of the pedestrian movement model is to study the spatial and temporal evolution trend of the crowd. The input of the model comes from some simple rules and hypotheses. In practice, these inputs often rely on the help of human experience. In fact, with the continuous development of sensor technology, real-time crowd gathering information can effectively replace artificial experience and become the input of a group disaster model. In view of the above problems, this paper combines the analysis of the overall crowd situation and individual model of the crowd gathering place in the context of the intelligent traffic key technology of multi-layer domain collaborative intelligent sensing and data fusion. The focus is transferred from accurately estimating the number of people on the current picture or signal to the reasonable distribution of crowd density. Considering the distance of the crowd within the space from the attraction point and the psychological state of the crowd and other factors, the individual model and the static model of the crowds gathering are established, and the spatial-temporal evolution of the crowd situation is extracted from the real-time monitoring video, thus proposing the crowds gathering early warning method for multi-domain information. The method is used to analyse the crowd density distribution of Suzhou City Life Squares on a certain day with certain guiding significance.

The structure of this paper is as follows: Section II introduces the work related to crowd density detection and

 ISBN: 978-1-61208-771-9

individual motion models; Section III establishes the static model of crowd gathering based on the theory of personal space, and analyses the crowd situation from the low-altitude local perspective and the high-altitude global perspective respectively, then, establishing the dynamic model of the spatial-temporal evolution of the crowd situation. Section IV applies the static and dynamic model to Suzhou City Life Squares, which guides crowd monitoring and evacuation, and achieves good results. Section V discusses the results.

II. Related Work

This section introduces the crowd density detection methods in the field of machine vision in recent years, and also summarizes the pedestrian model in crowd disaster dynamics. This has inspired the work of this paper.

A. Crowd Density Detection

The core of crowd density model is to calculate and estimate the crowd density. So many methods have been used to estimate the crowd density, and abundant results have been achieved. From the perspective of computer vision research, the crowd density estimation and counting methods in visual surveillance can be divided two classes. That is crowd density detection based on model labeling and crowd density detection based on feature extraction [5].

The methods using model labeling directly can label and count the human model in the image. Luo et al. mapped the crowd image directly to its crowd density map, then, obtained the total number of people by integral [6]. Zhao et al. divided the human body into multiple objects and used the ellipsoidal model for global tracking to calculate the crowd density [7]. Ge et al. proposed a bayesian method for estimating the number and location of individuals in video frames, which combines a spatial stochastic process that controls the number and location of individuals with a conditional marking process for selecting body shape, shape and direction, and nally gives the number of individuals [8]. Rao et al. proposed a method of estimating crowd density by motion hints and hierarchical clustering, which uses optical ow for motion estimation, contour analysis for crowd contour detection, and gets crowd density by clustering [9]. Although this method retains the features of detection targets to the greatest extent, it is easy to cause inaccurate detection results and difcult to meet the requirements due to the blurred individual contour and inaccurate positioning for dense crowds.

The methods using feature extraction can estimate the crowd density by extracting human features or using other parameters instead of human behavior, then, using normalized method. Koki et al. used the rotational angular velocity of human body as test data, and used continuous wavelet transform and machine learning methods to measure crowd density [10]. Ven et al. learned to distinguish crowd characteristics from granules and tted the contours between crowd and background (i.e., non-crowd) regions for density estimation [11]. Oliver et al. compared the application of two texture classification methods of bow and Gabor filters on aeronautical image plaque datasets to distinguish different crowd densities [12]. Zhang et al. proposed a simple and effective Multi-column Convolution Neural Network (MCNN) structure to map the image to its population density map [13]. By using filters with different size of receiving fields, the features of each column of Convolution Neural Network (CNN) can adapt to the changes of head size caused by perspective effect or image resolution and the effect is remarkable. Although the method of calculating individual density by extracting individual characteristics improves the accuracy of population density detection, the uneven distribution of population density leads to the identication of regional safety hazards not only by estimating the accuracy of population density.

B. Pedestrian Behavior Estimation

As mentioned above, on the basis of high-precision crowd density, we also need to pay attention to the position and state information of each individual. On the micro level, we divide the pedestrian motion model into cellular automata model, social force model, agent-based model and so on.

Based on the individual movement analysis of the cellular automata model, Claudio et al. proposed an improved version of the cellular automaton floor field model, using a sub-grid system to increase the maximum density allowed during the simulation and to reproduce the observed phenomena in dense crowd [14]. Ji et al. proposed a new triangular mesh cellular automaton model for the characteristics of high-density crowd evacuation, and accurately simulated the evacuation process of high-density crowd [15]. The advantage of the kinds of model is relatively simple and suitable for pedestrian behavior simulation in large-scale scenes. But its disadvantage is still obvious. That is, the algorithm itself is a heuristic algorithm, whose results with statistical signicance is unpredictable. And it can not be explained rationally due to divergent rule setting.

Based on the individual movement analysis of the social force model, Helbing et al. suggested that pedestrian movement be described as ”social forces”, which are not directly imposed by the pedestrian’s personal environment, but the measurement of the intrinsic motivation of the individual to perform the task [16]. Yang et al. proposed a pedestrian dynamics correction method based on social force model. By comparing the density-velocity and density-flow maps with the basic maps, it was verified that the guided crowd model can better reflect the pedestrian behavior characteristics in emergency situations [17]. However, the social force model lacks a clear and effective mechanism to ensure that pedestrians do not excessively contact (also known as overlap), so anti-overlap mechanism needs to be introduced.

Based on the individual motion analysis of the agent model, Tak et al. proposed an Agent-based Redestrian Cell Transmission Model (A-PCTM), which shows the flexibility of switching destinations and selecting driving directions according to the situation ahead [18]. Ben et al. presented an agent-based Cellular Automata (CA) environment modeling method that simulated four different evacuation scenarios and effectively guided crowd evacuation [19]. Was et al. proposed a proxy-based non-homogeneous cellular automaton model and an asynchronous cellular automaton model, enabling people to simulate pedestrian complex decision processes in complex environments [20].

III. Model Establishment

In this section, a static early warning model of crowd aggregation is established, and the crowd gathering state is derived through formulas. In addition, a dynamic spatio-temporal evolution model of the crowd situation is also established, and

 ISBN: 978-1-61208-771-9

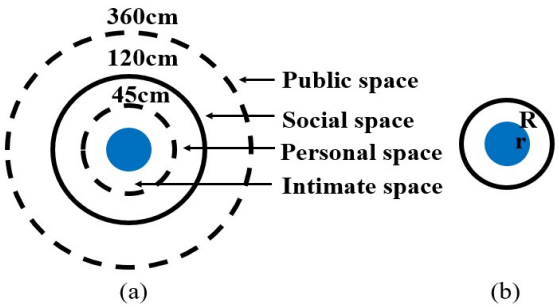

Figure 1. Individual model based on personal space theory.

the calculation methods of the crowd density are given from two perspectives: high altitude and low altitude.

A. Static Early Warning Model of Crowd Aggregation

In 1966, Edward Hall proposed the theory of personal space, which distinguishes four personal spatial distances, intimate distance, personal distance, social distance and public distance [21], as shown in Figure 1(a). Personal space theory is a kind of intimacy interpretation that serves the public relations of the public, which only divides the distance of the individual space and generally just applies to individual motion research. However, personal space theory does not directly quantify the relationship between population density and distance, and has great limitations for the application of large crowded places. On the basis of personal space theory, we defined the individual model as a solid circle with radius r, and the personal space is defined as a hollow circle with radius R, as shown in Figure 1(b). In the crowded place, we fitted the relationship between density and distance to analyze the distribution law of population density in large crowds. Firstly, we define the aggregation state T when a crowd gathers in a site, assuming that the crowd is in absolute static state with the most reasonable and safe distribution state, whose optimality depends on the characteristics of the site and the nature of the event. We assume that the determinant is expressed as the attraction of the site, therefore, the site is divided into n attraction points $O_j(j = 1, 2, ..., n)$. Each attraction point will cause the crowd to distribute according to some rules in the range of itself, so the position and size of each individual are different. These aggregates of individuals with different positions and sizes of individual space are called the crowd distribution at the attraction point, which is recorded as $U(O_j)$, and the aggregation state of the site is as follows:

$$T = \sum_{j=1}^{n} U(O_j), \frac{\partial T}{\partial t} = 0 \tag{1}$$

According to the actual situation, we set a plurality of attraction points for the place, and select one attraction point O_1. Then, we set two individual activity ranges closest to and farthest from the attraction point O_1. R_{min} is the radius of the nearest individual activity range from O_1, and R_{max} is the radius of the farthest individual activity range from O_1, as follows the formula calculating the change trend of the personal space radius R:

$$R = \tan\theta * x + A = \frac{R_{max} - R_{min}}{L} * x + A \tag{2}$$

Where θ is the angle between the center line of the two individuals range of activity and the horizontal plane. L is the straight line distance between the center of the farthest individual range of activity and the attraction point. A is the constant of $0.22 \sim 0.25$, representing the radius of the first individual space closest to O_1. x is a variable between 0 and L that changing along with L, as shown in Figure 2.

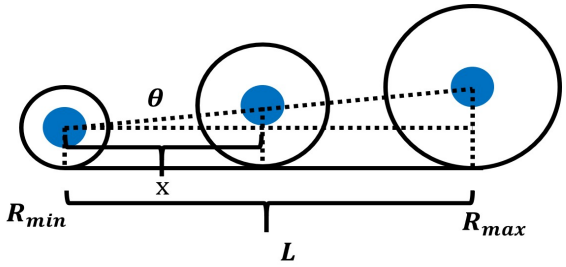

Figure 2. Relationship between personal activity space and distance.

At this time, the personal activity space has a corresponding relationship with the distance. Assuming that the activity radius of the individual i is R_i, the area occupied by the individual at that place is S_i, and the density ρ_i at that place is $\frac{1}{S_i}$:

$$\rho_i = \frac{1}{S_i} = \frac{1}{\pi {R_i}^2} = \frac{1}{\pi(\tan\theta * x + A)^2} \tag{3}$$

So, we get the corresponding relationship between different size of activity space and the density of the individual's position at this time. Then, we use the least square method to fit the discrete points of different density in different size of activity space, so as to determine the relationship between density and distance. The main idea of the least square method is to solve unknown parameters so that the sum of squares of residual errors can be minimized:

$$E = \sum_{i=1}^{n} (\rho_i - \hat{\rho}_i)^2 \tag{4}$$

The observed value ρ_i is the density of the position of the individual obtained after we calculate the trend of density change. The theoretical value $\hat{\rho}_i$ is the value of the polynomial after we obtain the specific coefficient under the set order. The objective function is also the loss function that is often said in machine learning. Our goal is to obtain the parameters when the objective function is minimized. The final fitting result is the relationship between the distance and the density at the attraction point O_1. After calculating the density at different distances, there is a microscopic individual combination N, which can be regarded as a population aggregation state $U(O_1)$ from a macro perspective:

$$U(O_1) = N = \sum_{i=1}^{n} P_i \tag{5}$$

Among them, P_i is the information set of the location of the individual i and the current personal space size. In the same way, the aggregation state $U(O_2)$ at the second attraction point O_2 is calculated.By analogy, a crowded static early warning model $T = \sum_{j=1}^{n} U(O_j)$ can be obtained.

B. Dynamic Evolution Model of Crowd Situation

The perception of crowd density is divided into high-altitude overall perspective and low-altitude local perspective. Low-altitude camera equipment tends to capture rich human characteristics and perceive local population density more accurately. When it is necessary to perceive the overall crowd situation as a whole, highlighting individual characteristics is not conducive to observing the overall movement trend, so the

 ISBN: 978-1-61208-771-9

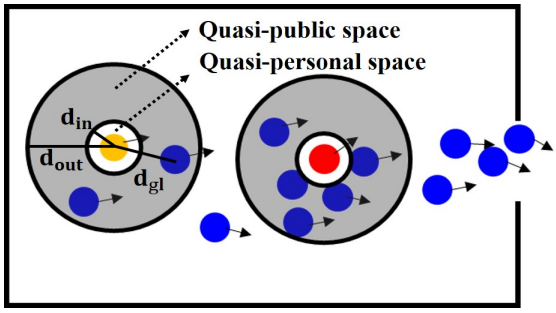

Figure 3. Density judgment of key moving points.

position and movement trend of the crowd at different times can be perceived by using high-altitude camera equipment.

The crowd gathering static early warning model describes the gathering state T of the site. If the crowd distribution V of all the moments t in the activity is described, the dynamic evolution of the crowd situation can be obtained. The static model of the crowd is a certain moment of the dynamic model:$V_t = T$. Since the perception of the crowd situation pays more attention to the group behavior, the characteristics of the person itself are no longer concerned. Therefore, what we need to know is the position evolution and the movement trend of the key moving point Q in the video image at different moments. The point Q can be obtained by extracting the foreground of the moving target from the Gaussian mixture model. Assuming that the mixed gaussian model consists of K Gaussian models, the probability density function is as follows:

$$p(w) = \sum_{k=1}^{K} p(k)p(w|k) = \sum_{k=1}^{K} \pi_k N(w|\pi_k, \sum k) \quad (6)$$

Where $p(w|k) = N(w|\pi_k, \sum k)$ is the probability density function of the gaussian model k, that is, the probability of generating w after the model k selected, $p(k) = \pi_k$ is the weight of the gaussian model k, that is, the prior probability of the model k is selected, and $\sum_{k=1}^{K} \pi_k = 1$. Then, the open operation and morphological denoising are performed on the model results. Finally, the foreground image consist of key moving points is obtain.

In the static model, the individual model represents a human body. In the crowd situation model in this section, the individual model is described to the simulation of the key moving point Q. At this time, the personal space is correspondingly transformed into the motion space between the key moving points, and the personal space distance is redefined as a point. The personal space distance is also redefined as the distance between the set of points $\sum_{i=1}^{n} Q_i$:

$$d_{gl} = \sqrt{(m_g - m_l)^2 + (n_g - n_l)^2}, g \neq l, l \in [1, n-1] \quad (7)$$

d_{gl} denotes the distance between any moving point Q_g and other moving points Q_l, (m_g, n_g), (m_l, n_l) are the position coordinates of the moving point Q_g and the moving point Q_l, respectively.The distance and direction of motion between the points change with time. We specify a personal space for each moving point Q_i, shown as the inner circle in Figure 3. And the public space shown as the outer circle in Figure 3. The number of moving points contained in the space between them is used as the basis for dividing the density:

$$d_{in} < d_{gl} < d_{out} \quad (8)$$

Where d_{in} represents the distance from the center of the circle to the inner circle, and d_{out} represents the distance from the center of the circle to the outer circle.

The perception of local crowd situation depends on the video and image acquired by low-altitude camera with obvious individual characteristics. We use Multi-column Convolution Neural Network (MCNN) model to extract human head features of different sizes [13]. The original image obtains different size of human head features through parallel networks with different sizes of three-column filters. Finally, the obtained features are weighted linearly to obtain the crowd density map. The model uses the maximum pooling layer of 2*2 and the activation function of the linear rectifier function, and integrates the three-column feature map. The loss function uses the optimized Euclidean loss function, which can standardize the density map of the network output:

$$L(\beta) = \frac{1}{2N} \sum_{i=1}^{N} ||F(X_i, \beta) - F_i||_2^2. \quad (9)$$

Here β is the network parameter to be optimized, N is the number of training images, X_i is the input image, F_i is the ground truth density map corresponding to X_i, $F(X_i, \beta)$ is the density map generated by MCNN. Figure 4 shows the structure of our MCNN. The three-column network structure has the same number of convolution layers and functions for each column except that the size of the filter. The purpose is to capture the head features of different sizes. Therefore, the first column is taken as an example. Enter an image of unlimited size, the first layer whose filters with 7*7 size to capture the local human head features. Then, max pooling is applied for each 22 region to reduce the resolution of the upper layer image to $\frac{1}{4}$ of the original image, the number of parameters is reduced, and more useful features are extracted. At the end, the features are weighted and stacked by a 1*1 filter, so that the output results are averaged for density grading processing. The density normalization process mainly depends on the Gaussian kernel function. This paper proposes that the adaptive Gaussian kernel function is slightly different from Zhang[14]:

$$F(x) = \sum_{i=1}^{M} \delta_i * G(x, \sigma_i) \quad (10)$$

Where δ_i represents the impulse function of each head, M is the number of heads in the image. And σ_i denotes the maximum head distance of the adaptation within a certain range (Using the maximum is to make the crowds more dense), $\sigma_i = \alpha max(d_{ij})$. α is the weight value of the adaptive range. In our experiment, it shows that when $\alpha = 0.5$ the crowds intensity is the most consistent with the actual situation.

IV. Model Application

Located near Jinji Lake in Suzhou City in China, the city life square covers an area of 4300 square meters and periodically carries out large-scale fountain projection activities, with a maximum of 35,000 people. Our crowds gathering model was been applied in this square. The data come from the video camera equipment covering the inside and the exits of the square and the construction drawings of the construction of the site. The crowd distribution of the square was been investigated on the spot. The concrete implementation is as follows.

A. Application of Static Early Warning Model

We select the fountain where it is an attraction point $O_1(n = 1)$, then, establish a spatial coordinate system choos-

 ISBN: 978-1-61208-771-9

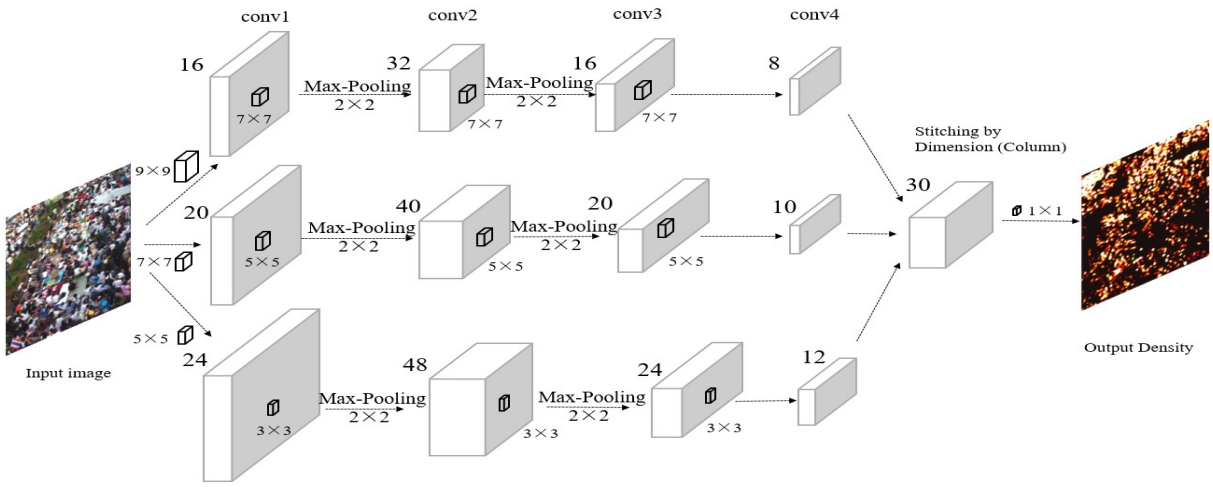

Figure 4. MCNN Network Structure [13].

ing O_1 as the coordinate origin. The radius r of the individual model is determined by the shoulder width of the person standing. according to [22] [23], a solid circle with radius $r = 22cm$ is used as an individual model in this paper. Through field survey and CAD drawing measurement, we set $L = 200m$ in the static early warning model, and take R_{max} and R_{min} as $0.62m$ and $0.25m$ respectively according to the queued waiting state pedestrian service level table [24]. When the least squares method is fitted, we calculate it repeatedly. When the loss E reaches 0.1 at the first time, the fitting effect is the best, and the relationship between the crowd density ρ and the distance d from the attraction point is calculated as:

$$\rho = -1.8782e^{-12}d^5 + 2.3706e^{-9}d^4 - 1.187e^{-6}d^3 + 0.0030768d^2 - 0.045949d + 3.9659 \quad (11)$$

Taking $N = 50000$ different distance corresponding to different density of points cyclic calculation, these point sets N is the crowd aggregation state $U(O_1)$ at the point O_1:

$$U(O_1) = \sum_{i=1}^{n} P_i = \sum_{i=1}^{n} P(d_i, \rho_i) \quad (12)$$

thus establishing a static early warning model of crowd aggregation taking Fountain Square as an example.

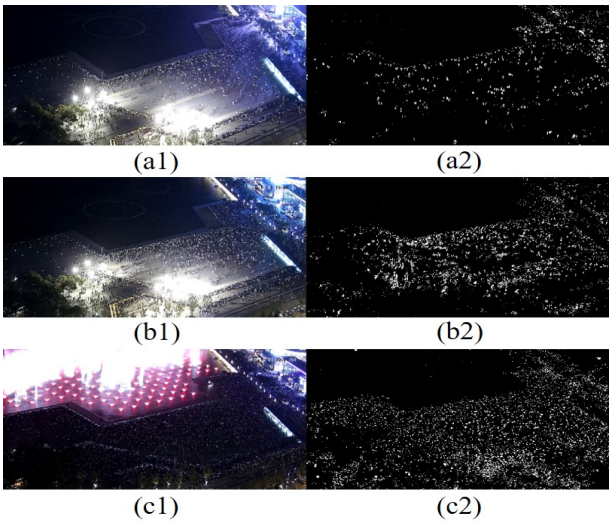

Figure 5. Key moving points map of fountain square in different time period(a1) Low crowd density real-time video (a2) Low crowd density key moving points set (b1) Medium crowd density real-time video (b2) Medium crowd density key moving points set (c1) High crowd density real-time video (c2) High crowd density key moving points set.

B. Application of Dynamic Evolution Model

We intercept the key frames of the video images captured by the aerial camera. We extract the foreground image of the moving target through the Gauss mixture model, and get the set of the key moving points. At low density, we take 18724 moving points, 43779 moving points at medium density, and 61386 moving points at high density, as shown in Figure 5. According to the distance between the moving points, we can judge the density grade at the point whether it is between d_{in} and d_{out}. Here, we select $d_{in} = 1$ and $d_{out} = 8$ to get the density grade of the key moving points, which reflects the crowd situation at that time, as shown in Figure 6 (a1), (b1), (c1). At low density, the distribution of moving points is scattered and the intensity is relatively light. At medium density, the moving points cover the image area in a large area, and some of the point sets are in a highly concentrated state. While at high density, the moving points basically occupy the image area, showing a trend of global high density, reflecting the high crowding in the square at the moment. The distance between people is very short, which leads to people moving slowly and potential safety hazards.

We performed a curve analysis of the population density at the same angle for the three densities. According to the collected data, a density curve is shown, as shown by the red lines in Figure 6 (a2), (b2) and (c2), the horizontal axis represents the distance from the attraction point, and the vertical axis represents the density at that point. Comparing the crowd density change of static early warning model of crowd aggregation (blue line in Figure 6) with that of low density, medium density and high density.We found that, when the crowd density reaches high density, the density line almost exceeds the warning line, which means potential safety hazards could be in the square, as shown in Figure 6 (c2). Under the high density, pedestrians often have inevitable contact with each other. It is impossible to walk horizontally or reversely, and the flow of people is extremely unstable, which is in accordance with the conclusion of Figure 5 (c1), (c2). We carry out dynamic crowd situational awareness of high-

 ISBN: 978-1-61208-771-9

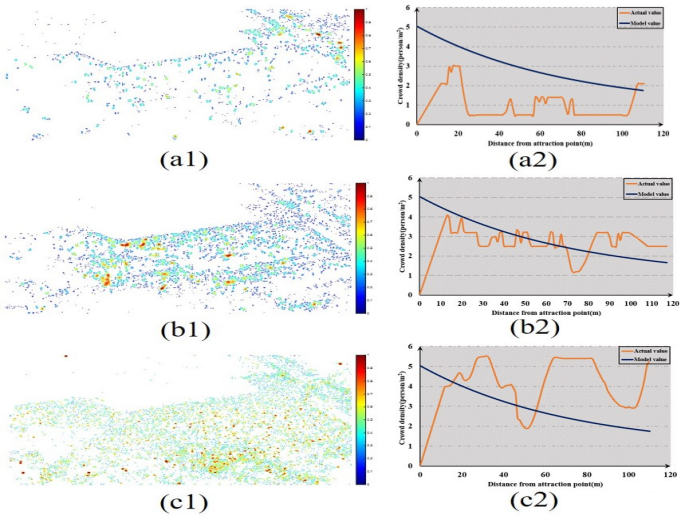

Figure 6. The crowd density curve of the early warning model and the actual population density curve under three states. (a) Early warning curve and low density curve (b) Early warning curve and medium density curve (c) Early warning curve and high density curve.

altitude video, which mainly includes the following three steps: extracting video key frames, extracting foreground images of sports people through Gaussian mixture model, and performing density calculation on all key moving points to obtain density classification of moving points. After the above work is carried out, the perception of the crowd situation can be realized to capture the preference trend of the crowd evacuation. Then, we can formulate the strategy for evacuation of the crowd, which can guide the staff to carry out the evacuation work quickly and effectively. Based on the perception of the crowd situation, we finally get the direction and magnitude of the crowd flow, which provides a reference for our subsequent simulation parameter settings. In addition, we also attempted to analyze the density of low-altitude crowd.

V. Conclusion

In this paper, we discussed the limitations of Hall's personal space theory in the crowded scene. On the basis of this, we explored the quantitative relationship between crowd density and distance, so that the static early warning model for crowd gathering can be established. In contrast, we used the characteristics of overall video images to perceive the temporal and spatial evolution of the crowd situation and established a dynamic model of the crowd situation. The joint analysis of static early warning model and dynamic model can comprehensively perceive the real-time situation of the crowd and improve the public safety of the site. We have successfully applied our models in the Fountain Square of Suzhou City. Through the video images acquired by the high-altitude camera equipment, we perceived the changes in the crowd situation of the site. In future work, we will fuse heterogeneous multi-granularity surveillance videos and supplement the entire situation with local actual number to estimate the number of people in the entire venue.

References

[1] S. Pu, T. Song, Y. Zhang, and D. Xie, "Estimation of crowd density in surveillance scenes based on deep convolutional neural network," Procedia computer science, vol. 111, 2017, pp. 154–159.

[2] D. Helbing, A. Johansson, and H. Z. Al-Abideen, "Dynamics of crowd disasters: An empirical study," Physical review E, vol. 75, no. 4, 2007, p. 046109.

[3] A. Johansson, D. Helbing, H. Z. Al-Abideen, and S. Al-Bosta, "From crowd dynamics to crowd safety: a video-based analysis," Advances in Complex Systems, vol. 11, no. 04, 2008, pp. 497–527.

[4] M. Moussaïd, D. Helbing, and G. Theraulaz, "How simple rules determine pedestrian behavior and crowd disasters," Proceedings of the National Academy of Sciences, vol. 108, no. 17, 2011, pp. 6884–6888.

[5] S. A. M. Saleh, S. A. Suandi, and H. Ibrahim, "Recent survey on crowd density estimation and counting for visual surveillance," Engineering Applications of Artificial Intelligence, vol. 41, 2015, pp. 103–114.

[6] H. Luo et al., "A high-density crowd counting method based on convolutional feature fusion," Applied Sciences, vol. 8, no. 12, 2018, p. 2367.

[7] T. Zhao and R. Nevatia, "Tracking multiple humans in complex situations," IEEE Transactions on Pattern Analysis & Machine Intelligence, no. 9, 2004, pp. 1208–1221.

[8] W. Ge and R. T. Collins, "Crowd density analysis with marked point processes [applications corner]," IEEE Signal Processing Magazine, vol. 27, no. 5, 2010, pp. 107–123.

[9] A. S. Rao, J. Gubbi, S. Marusic, and M. Palaniswami, "Estimation of crowd density by clustering motion cues," The Visual Computer, vol. 31, no. 11, 2015, pp. 1533–1552.

[10] K. Nagao, D. Yanagisawa, and K. Nishinari, "Estimation of crowd density applying wavelet transform and machine learning," Physica A: Statistical Mechanics and its Applications, vol. 510, 2018, pp. 145–163.

[11] V. J. Kok and C. S. Chan, "Granular-based dense crowd density estimation," Multimedia Tools and Applications, vol. 77, no. 15, 2018, pp. 20 227–20 246.

[12] O. Meynberg, S. Cui, and P. Reinartz, "Detection of high-density crowds in aerial images using texture classification," Remote Sensing, vol. 8, no. 6, 2016, p. 470.

[13] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 589–597.

[14] C. Feliciani and K. Nishinari, "An improved cellular automata model to simulate the behavior of high density crowd and validation by experimental data," Physica A: Statistical Mechanics and its Applications, vol. 451, 2016, pp. 135–148.

[15] J. Ji, L. Lu, Z. Jin, S. Wei, and L. Ni, "A cellular automata model for high-density crowd evacuation using triangle grids," Physica A: Statistical Mechanics and its Applications, vol. 509, 2018, pp. 1034–1045.

[16] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," Physical review E, vol. 51, no. 5, 1995, p. 4282.

[17] X. Yang, H. Dong, Q. Wang, Y. Chen, and X. Hu, "Guided crowd dynamics via modified social force model," Physica A: Statistical Mechanics and its Applications, vol. 411, 2014, pp. 63–73.

[18] S. Tak, S. Kim, and H. Yeo, "Agent-based pedestrian cell transmission model for evacuation," Transportmetrica A: transport science, vol. 14, no. 5-6, 2018, pp. 484–502.

[19] X. Ben, X. Huang, Z. Zhuang, R. Yan, and S. Xu, "Agent-based approach for crowded pedestrian evacuation simulation," IET Intelligent Transport Systems, vol. 7, no. 1, 2013, pp. 55–67.

[20] J. Was and R. Lubas, "Towards realistic and effective agent-based models of crowd dynamics," Neurocomputing, vol. 146, 2014, pp. 199–209.

[21] E. T. Hall, The Hidden Dimension. Garden City, NY: Doubleday, 1966, vol. 609.

[22] Y.-C. Lin, M.-J. J. Wang, and E. M. Wang, "The comparisons of anthropometric characteristics among four peoples in east asia," Applied Ergonomics, vol. 35, no. 2, 2004, pp. 173–178.

[23] C. C. Gordon et al., "2010 anthropometric survey of us marine corps personnel: methods and summary statistics," Army Natick Soldier Research Development and Engineering Center Ma, Tech. Rep., 2013.

[24] H. C. Manual, "Highway capacity manual," Washington, DC, vol. 2, 2000.

 ISBN: 978-1-61208-771-9