Video-Based Crowd Density Estimation and Prediction System for Wide-Area Surveillance

CAO Lijun, HUANG Kaiqi

National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing 100190, China

Abstract: Crowd density estimation in wide areas is a challenging problem for visual surveillance. Because of the high risk of degeneration, the safety of public events involving large crowds has always been a major concern. In this paper, we propose a video-based crowd density analysis and prediction system for wide-area surveillance applications. In monocular image sequences, the Accumulated Mosaic Image Difference (AMID) method is applied to extract crowd areas having irregular motion. The specific number of persons and velocity of a crowd can be adequately estimated by our system from the density of crowded areas. Using a multi-camera network, we can obtain predictions of a crowd's density several minutes in advance. The system has been used in real applications, and numerous experiments conducted in real scenes (station, park, plaza) demonstrate the effectiveness and robustness of the proposed method.

Key words: crowd density estimation; prediction system; AMID; visual surveillance

I. INTRODUCTION

Video analysis techniques are becoming increasingly popular in the visual surveillance of public areas because of their great efficiency in gathering information and low cost in human resource. A central topic is the automatic analysis and detection of abnormal events. One particular abnormal event is crowding which may occur wherever a large number of people gather together at public assemblies, sport competitions, or demonstrations (e.g., strikes, protests), etc. Because of the high level of risk, crowding has always been of high concern to relevant authorities. In recent years, a number of security agencies specialized in crowd management have emerged, and the visual surveillance research has studied the automated monitoring crowd movements [1].

Foreground based methods: in Refs. [2-3], the foreground is extracted firstly by background removal using a reference image, then crowd density is computed as a function of the number of foreground pixels; the function itself is obtained by curve fitting. However, these methods may fail when the background changes gradually over time. In Ref. [4], Optical Flow and Background Model (OFBM), which is based on LK optical flow and GEM methods, is computed for the whole image and used for crowd density estimation. This approach overcomes the shortages of optical flow and background subtract, such as sensitiveness of light changing and producing accumulate errors. However, the modeling is time-consuming. In Ref. [5], the foreground of moving crowds is detected by a Bayes decision rule for classification between background and foreground. The number of people in a crowd is computed as a linear function of foreground pixels. In Ref. [6], a Markov Random Fields (MRF) based approach is used to model changes in pixel value, and the optimal foreground is obtained by minimizing a MRF-

Received: 2013-02-06
Revised: 2013-03-27
Editor: WANG Qiao

A novel method is proposed to estimate the crowd density through AMID feature. Based on the method, a prediction algorithm is designed which can estimate the specific number of people and velocity of crowd well. The applications with multiple-camera network demonstrate the system is robust and effective on wide-area surveillance.

based objective function. This method gives good results in subway scenes, but the minimization is very difficult and time-consuming.

Feature based methods: Haar feature based head detection [7] and integral channel features [8] based head detection [9] are adopted to detect human heads in crowds. The total number of people in a crowd is estimated by analyzing the sizes and positions of detected heads, but the method may fail if the observed area is so crowded that few heads can be detected. In Ref. [10], texture feature vectors are extracted from input images and a Support Vector Machine (SVM) is used to solve the regression problem of calculating crowd density. This method is inconvenient for real applications. In the training stage, the ground truth of crowd densities is highly related to specific scenes and needs to be estimated by human experts. Recently, many new approaches for crowd segmentation and density estimation have been proposed. In Refs. [11-16], individuals in moving crowds are detected by tracking and clustering local features. These methods are good ways to estimate both the number and the location of the individuals in moving crowds.

Group based methods: in Ref. [17], the authors propose a group-based method to accurately estimate the number of people in moving groups and track each group reliably. This method deals with the entire area occupied by a group as a whole, rather than trying to detect individuals separately. In Ref. [18], the authors propose a framework in which Lagrangian particle dynamics is used to segment high density crowd flows and detect flow instabilities. In this method, moving crowds are treated as periodic dynamical systems manifested by a time dependent flow field. Ref. [19] uses Size of extracted crowd region as density measurement. Some papers [20-21] pay attention to the crowd behavior modeling for abnormal event detection, so far in constrained environments.

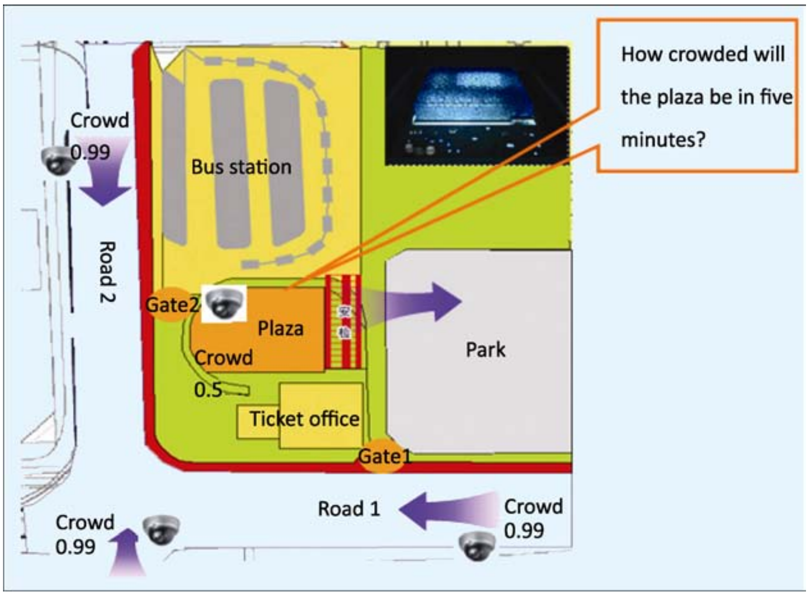

Fig.1 *The plaza with two nearby roads*

Most of these approaches mentioned focus on single region crowd analysis, which can be divided into crowd information extraction and crowd density modeling. Texture based methods are often used to extract the crowd information, which include speed, direction and location of a crowd in a video sequence and so on. These methods cannot work for high density crowds. First, individual detection is nearly impossible due to heavy occlusions or poor view angles. Another problem is that the gathering people may stay motionless for quite a while, thus foreground detection by background modeling or feature point tracking is very difficult. The previous work of crowd density estimation of our group is the Accumulated Mosaic Image Difference (AMID) feature [16], which is suitable for some real scenes, but recently, the real application of crowd monitoring always asks for collaborative work of multiple cameras in wide areas, which is also a challenge for most of surveillance systems.

In this paper, we propose a crowd analysis system which can be explored in public areas such as bus stations, subways and plazas. The degree of crowding is estimated in monocular image sequences. And future crowd densities and velocities are predicted using the information obtained from a number of cameras. Figure 1 shows a plaza with an entry from two roads and an exit to a park. Predicting the crowd density in the plaza can help decisionmakers present overcrowding.

This paper has three contributions. First, the AMID feature [16] is used to estimate the number of people transformation in crowds, which can be used for estimating crowd densities and velocities. Secondly, with the information from multi-cameras, we propose a novel prediction method for wide-area crowd analysis through. Lastly, we develop a wide-area crowd surveillance system and the practical applications show the effectiveness of this system. The remainder of the paper is organized as follows. In Section II, we describe the system framework and every component, including the AMID based crowd information extraction algorithm, crowd-number of people transformation, speed and direction computation and prediction. Experimental results and analysis are presented in Section III. Finally, we draw a conclusion in Section IV.

II. System Framework for Wide-Area Surveillance

In this section, we give the details of the human crowd analysis and prediction system. As Figure 2 shows, this system includes crowd density analysis in images taken by a single camera and crowd density prediction in images taken by multi-cameras. When the crowd density is obtained from images taken by many cameras, we can predict crowd levels at specific places a few minutes into the future. Next the details of the system will be given.

2.1 AMID based crowd density estimation

As we know, high-density crowds often contains subtle meaningless motions, and these tiny motions happen in the whole time in crowds, like people's turning around and raising heads. We can refer to such crowds as stable crowds and to the tiny motions happening in stable crowds as intra-crowd motions. The previous work of our group [16] develops AMID feature to describe these intra-crowd motions and estimate the crowd density. Local individual perturbations and movements are

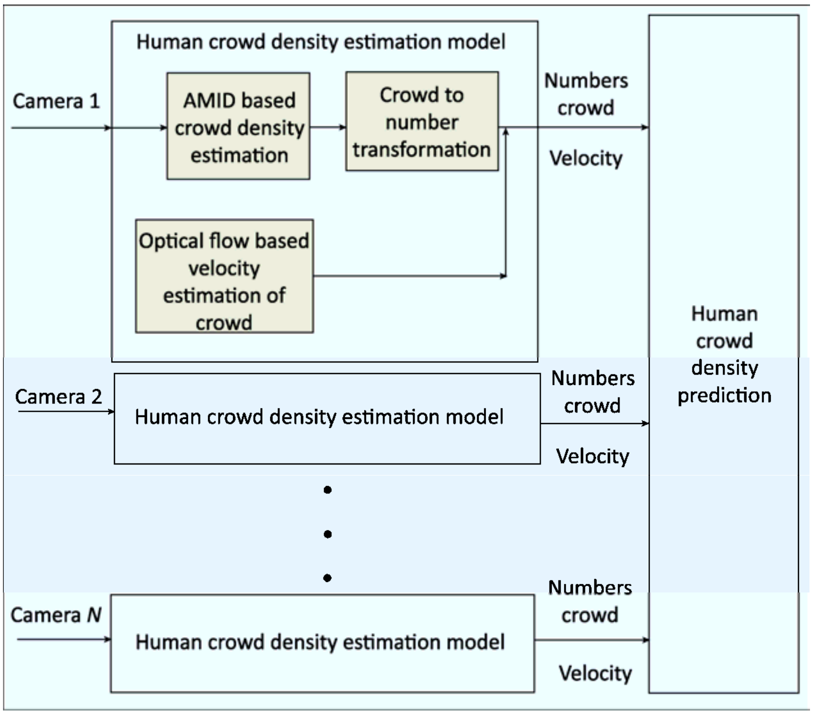
Human crowd density estimation model
Camera 1
AMID based crowd density estimation
Crowd to number transformation
Numbers crowd
Velocity
Optical flow based velocity estimation of crowd
Camera 2
Human crowd density estimation model
Numbers crowd
Velocity
Camera N
Human crowd density estimation model
Numbers crowd
Velocity
Human crowd density prediction

Fig.2 *System framework for wide-area surveillance*

the main reasons causing intra-crowd. It is indeed nearly impossible for all people who are waiting in a local crowded area keep absolute still all the time, so intra-crowd motions happen almost everywhere in crowded areas and their distribution areas reveal the size of crowded areas. AMID feature uses this incessant characteristic of high crowd to achieve the estimation.

In AMID, the directions of intra-crowd motions will not be taken into consideration because of random directions in intra-crowd motions. It only measures their locations and temporal characteristics in local areas. The locations of intra-crowd motions can be obtained by local image change detection. The AMID feature comes from Mosaic Image Difference (MID), which is gained by dividing the image into local areas. The MID series are:

$$\{\mathrm{MID}_k(m,n) \mid t - N_{\mathrm{mid}} < k \leqslant t\} \tag{1}$$

$$\mathrm{MID}_t(m,n) = \begin{cases} 1, \text{if } \ \|M_t(m,n) - M_{t-1}(m,n)\|_\infty > T_t \\ 0, \text{otherwise} \end{cases} \tag{2}$$

where t is the current frame number; (m, n) is the middle location of the corresponding Mosaic Block (MB) (m, n); N_{mid} is the width of the observing time window and t means this series is updated at every new frame. In Eq. (2), $\|\cdot\|_\infty$ denotes the maximum absolute component of a vector; T_t is an adaptive threshold; and $M_t(m,n)$ is a representation value of each local area.

$$M_t(m,n) = \frac{1}{L_M^2} \sum_{(i,j)\in MB(m,n)} \boldsymbol{I}_t(i,j) \qquad (3)$$

Here, $\boldsymbol{I}_t(i,j)$ is the RGB vector of pixel (i, j) at frame #t. L_M is the size of the MB.

The AMID series is obtained by evenly dividing the observing time window into a number of sub time windows and accumulating the MID features of MB (m, n) in each sub time window separately. The value of the l-th element in AMID series of MB (m, n) at frame #t is given by:

$$\text{AMID}_l(m,n,t) = \sum_{k\in W_l(t)} \text{MID}_k(m,n) \qquad (4)$$

where $l = 1\ldots N_{sw}$. N_{sw} is the number of sub observing time windows and $W_l(t)$ denotes the l-th sub observing time window at frame #t:

$$W_l(t) = \left\{ k \mid t' + \frac{N'_{\text{mid}}}{N_{sw}}(l-1)+1 \leqslant k \leqslant t' + \frac{N'_{\text{mid}}}{N_{sw}} l \right\} \qquad (5)$$

where $N'_{\text{mid}} = [N_{\text{mid}}/N_{sw}]N_{sw}$ is the valid length of observing time window after division, $t' = t - N'_{\text{mid}}$. The AMID series is updated at every frame and provides a good description of the local intra-crowd motions, because it depicts the change of a local area in each time piece of the observing time window.

The indicator function that whether MB (m, n) should be labeled as a foreground area (crowded area) or a background area at frame #t is determined below:

$$U_t(m,n) = \begin{cases} 1, & \begin{array}{c} \text{if } \left|S_{mt}(m,n,t) - S^0{}_{mt}\right| < S_{mt}^{th}, \\ S_{mv}(m,n,t) > S_{mv}^{th}, \\ \text{and } S_{ns}(m,n,t) > S_{ns}^{th} \end{array} \\ 0, & \text{otherwise} \end{cases} \qquad (6)$$

where S_{mt}^{th}, S_{mv}^{th} and S_{ns}^{th} are three important statistics of local intro-crowd motions which can be defined: 1) $S_{mt}(m, n, t)$ — the mean time when motions happened; 2) $S_{vt}(m, n, t)$ — the variance of time when motions happened; 3) $S_{ns}(m, n, t)$ — the number of sub observing time windows in which motions happened. The indicator function $U_t(m, n)$ describes the temporal scattering degree of local intra-crowd motions according to the given assumption of uniform distribution. When $U_t(m, n) = 1$, it means the local motions scatter extensively enough that they are most likely caused by stable crowds.

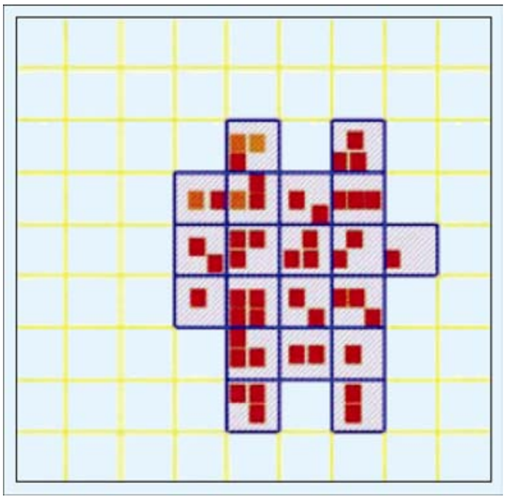

Fig.3 *Illustration of the gridding method* (*red blocks: crowded areas detected by temporal statistical analysis of local intra-crowd motions, blue shading grids: the foreground area obtained*)

Thus, the entire crowded area can be easily obtained by the gridding method. The foreground area at frame #t can be denoted as:

$$R_{fg}(t) = \{(i,j) \mid (i,j) \in (GB(p,q) \cap R_{\text{roi}}), G_t(p,q) = 1\} \qquad (7)$$

where (i, j) denotes a pixel's position.

Figure 3 illustrates the principle of the Gridding method. From Eq. (4), we can see that the computing time of extracting AMID feature for estimation is just linearly on the size of interested areas and the computational complexity is

$$o\left(mn \sum_{(i,j)\in MB(m,n)} \boldsymbol{I}_t(i,j) \right) = o(\boldsymbol{I}) \qquad (8)$$

2.2 Crowd density to number of people transformation

In order to predict crowd densities, the transformation from crowd density to number of people should be known and vice versa. Crowd density is a value between 0 and 1, which cannot be used for the prediction directly. Here the linear fitting method is used to give the estimation of the number of people. Based on some testing measurements in advance, we can know *p* groups mapping relation between crowd density and the number of people: $\{(d_i, n_i)\}_{i=1\ldots p}$, where d_j is the crowd density of *i*-th group; n_i is the number of people. With the crowd density d_i, the relationship can be decided by prior measurement. If one new region with the crowd *d* belongs to the *s* group $[d_s, d_{s+1}], (0 \leqslant s \leqslant p)]$, then the number of people is

$$n(d) = \frac{n_{s+1} - n_s}{d_{s+1} - d_s}(d - d_s) + n_s \qquad (9)$$

and vice versa. If we get the number of people, then the crowd density can also be estimated:

$$d(n) = \frac{d_{r+1} - d_r}{n_{r+1} - n_r}(n - n_r) + d_r \qquad (10)$$

where *r* is the *r*-th region $[n_r, n_{r+1}]$, and *d*(*n*) is the crowd density to be decided.

2.3 Crowd velocity estimation based on optical flow

The velocity of the crowds is also required for the prediction. Here we get the direction and speed based on optical flow [22]. We can get the optical flow as $OF_u(k)$ and $OF_v(k)$, where *u* and *v* stands for the horizontal and vertical components, and *k* is the frame. Then the Speed $OF_\rho(k)$ and direction $OF_\theta(k)$ of the optical flow are obtained by:

$$OF_\rho(k) = \sqrt{OF_u^{\,2}(k) + OF_v^{\,2}(k)} \qquad (11)$$

To obtain $OF_\theta(k)$, we sample it with *M* bins histogram (*M* is 4 or 8). The $M_{\max}$ is chosen from the histogram and we get the center of crowd flow $\theta_c = \frac{(2M_{\max} - 1)\pi}{M}$. The pixels in $[\theta_c - \theta_{\text{offset}}, \theta_c + \theta_{\text{offset}}]$ are used to get the speed and direction of crowd flow $\theta_p(k), \rho_p(k)$ with the average of *k* frame.

$$OF_\theta(k) = \begin{cases} \arctan\frac{u}{v}, & u > 0, v > 0 \\ \arctan\frac{u}{v} + \pi, & u < 0 \\ \arctan\frac{u}{v} + 2\pi, & u > 0, v < 0 \end{cases} \qquad (12)$$

2.4 Crowd density prediction based on directed structural diagram

It is useful to estimate the crowd density at a govern location in few minutes in the future. One directed structural diagram is designed according to the distance and direction of a crowd. As Figure 4 shows, region A is the major node (for example, plaza), which is surveyed by cameras C0, B, D, E, and F are sub nodes under the surveillance of cameras C1, C2, C3 and C4. S1, S2, S3 and S4 are the distance away from place A respectively. The predicted number of people in region A is W_A', which can be computed with the present number of people W_A. Input people W_B, W_D, W_E from B, D, E and output people W_F. These inputs and outputs can be computed by the method purposed in Section 2.2.

$$W_A' = W_A + W_B + W_D + W_E - W_F \qquad (13)$$

where W_A is known. Here we just give the computation step for W_B as follows:

1) Computing the time to main node (A) for every unit $T_i + (i-1)\Delta\ (i = 1,2,3,4......30)$;

2) Computing the sum of people N$_i$ in Ω units time interval (Ω > Δ, Ω =IΔ, I is integer);

3) Computing the max time $T_{i\max}$ where the people from sub node to main node (such

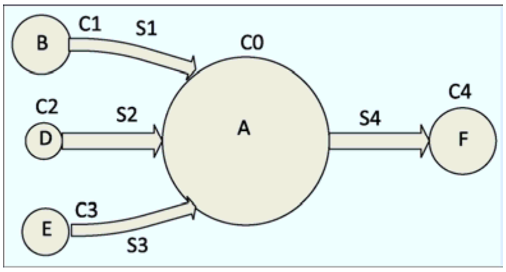

Fig.4 *Directed structural diagram for prediction*

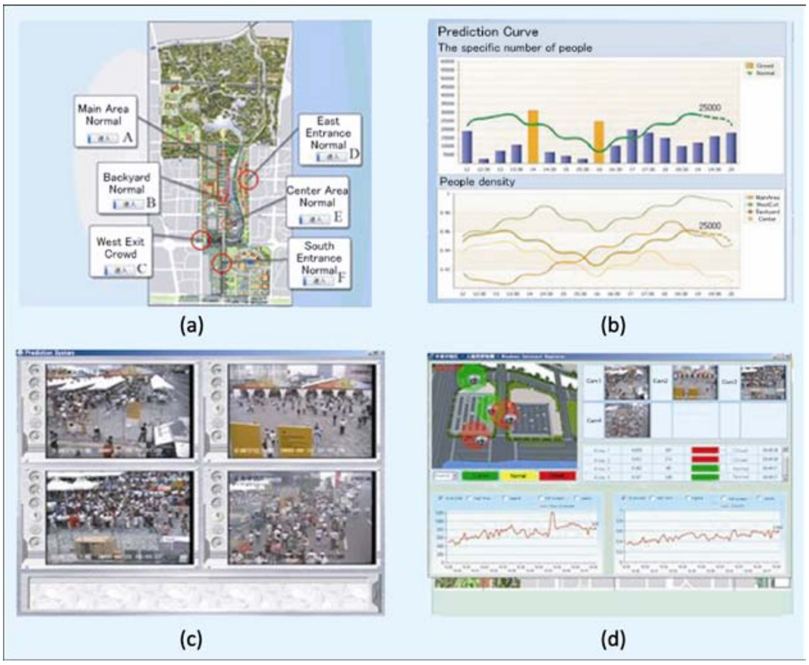

Fig.5 *Map of the ceremony and system interfaces* **(a)** *map of the ceremony;* **(b)** *analyzing curves;* **(c)** *analyzing interface;* **(d)** *alarming interface*

as region B to A);

4) If $T_0 < T_{i\max}$, then $W_B = \sum_{T_0} N_i$, $T_i < T_0$;

5) Else $W_B = \sum_{T_0} N_i + N_0 \times (T_0 - T_{i\max})$;

where W is the number of people in region B at present time; L is the length of region B. V is the average speed; Δ is unit (the frame rate of camera). Output number in basic time interval is $N_i = \rho L_i$, where $\rho = \frac{W}{L}$ is the crowd of unit; $L_i = V \times \Delta$ is the distance of basic time interval. t_0 is the start time; T_0 is the during time.

III. Experimental Results

The system has been used in a large ceremony. Figure 5 (a) is a map of the ceremony. Six regions labeled A, B, C, D, E, F (including bus station, entrance 1, entrance 2, …) are analyzed. The predicted results of crowd density and the number of people in plaza are shown in Figure 5 (b)[1]. The top half shows the predicted number of people and the lower half shows the real-time curves of crowd density. The total deviation of the predicted number of people does not exceed 10% of the actual value in the final statistics according to the ground truth from the sale of tickets. Next we will give more results about the crowd density analysis algorithm.

Our algorithm is implemented on a PC with a P4 3.0 GHz CPU and 512 MB memory in C++ programming language. With real-time processing of a 320 × 240 video, the CPU usage is less than 50%. In our real application system, one PC can analyze four cameras. To test the adaptability of our proposed method in different real scenes, the experiments are conducted in a bus station scene, a subway station scene and a plaza respectively.

3.1 Crowd density and number of people estimation

In this part, the experiments of crowd density estimation and number of people are given in “bus station” video and “subway station” video.

In Figure 6, the curve of number of people vs. time for the “subway station” video is given as well as some example frames. The blue line is the ground truth which is given manually. The red line is the analyzing results. The green number in the left-up corner of the frame is the estimated number of persons. In the beginning, there are few persons in this scene (about 8 at Frame#202); later, more and more persons come into the scene (about 43 at Frame#1402); then the number of persons decreases with the time. From the figure, we can see that the analyzing results consist well with the ground truth in most time. The estimated accurate rate is over 90%.

3.2 Crowd velocity estimation based on optical flow

In this set of experiments, we focus on the accuracy of the crowd velocity estimated by our purposed method. Here we apply our method on three video sequences shown in Figure 7. To get the real data, we select several distinctive people as our targets and figure out their bounding box by hand.

We calculate the mean values of their real velocities as the real velocities, which is used to be compared with our estimated velocities.

$$v_{\text{real}} = \frac{1}{n\delta}\sum_{i=1}^{n} f_c(y_i)\sqrt{\Delta x_i^2 + \Delta y_i^2} \qquad (14)$$

[1] The figure is a sketch map to show how the system works, however, the details of the system is restricted and we will give the crowd density estimation results.

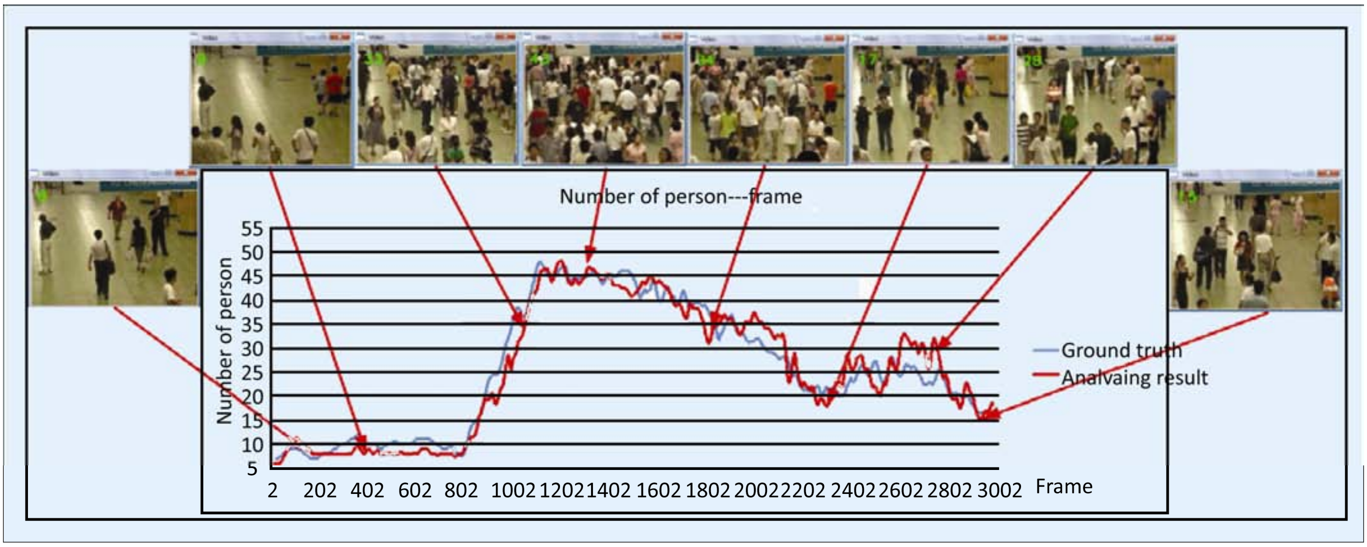

Fig.6 *The curve of number of person vs. time for the "subway station" video and some selected frames*

$$\theta_{\text{real}} = \frac{1}{n\delta}\left\| \arctan\frac{\Delta y_i}{\Delta x_i} \right\| \qquad (15)$$

where i is the frame index we extract for bounding boxes; Δx_i and Δy_i are distances of the targets moving on the images. $f_c(y_i)$ is a weight set to indicate the practical length per pixel and is calculated through $f_c(y) = ((y_r - y_v)/(y - y_v))^2$ [16]. δ is the parameter of down sampling, and here we set it to 25. $\|\cdot\|$ is used to get the directions ranging between 0 and 360.

The comparisons are listed in Table I. From Table I, we can figure out that the estimation from our purposed method is very near to the real data.

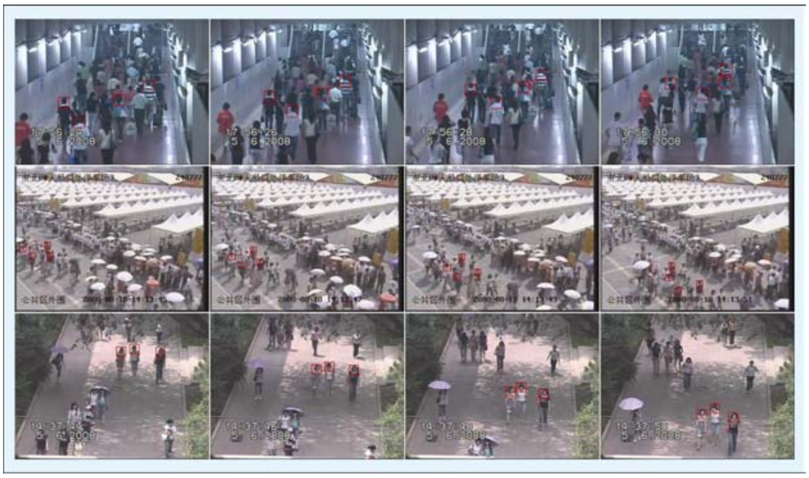

Fig.7 *Video sequences and the bounding boxes used for real data by hand*

Table I *The comparisons between estimation and handed-data*

	θ_{est}	θ_{real}	v_{est}	v_{real}
Video1	64.2	71.3	0.271	0.301
Video2	322.4	314.7	1.022	1.089
Video3	267.5	265.2	0.163	0.221

3.3 Comparisons and discussion

In this section, we will give the comparison of our method with some present methods. As we have mentioned, the crowd information extraction is an important step in crowd density estimation. Here we compare our AMID method with the GMM method [23]. GMM is a well-known background modeling method, which can work in clutter background in real time for motion detection and can handle swaying trees, ocean waves, etc., while GMM cannot be used for high crowd density estimation. Figure 8 gives some comparison results between GMM and AMID method for crowd information extraction. Figure 8 (a) is the original image; Figure 8 (b) is the results based on GMM and Figure 8 (c) is the results based on AMID. From the figure, we can see that GMM cannot extract most of the crowd information because of the high density, which causes the slow motion. Our AMID method can extract the crowd information effectively.

In Table I, we also give the number of person result comparison between GMM based, AMID based without Perspective Distortion

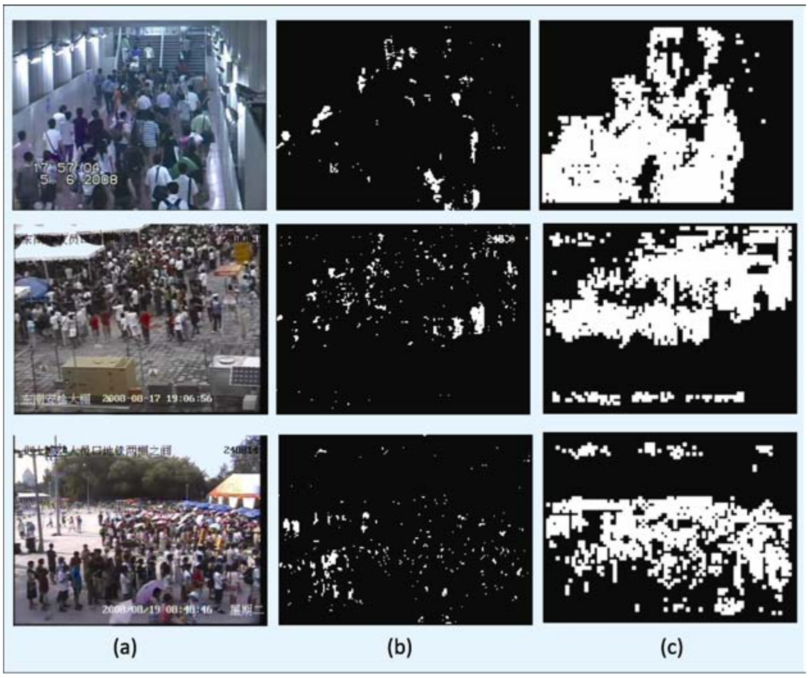

Fig.8 *Crowd information extraction comparison between GMM and AMID method.* **(a)** *original image* **(b)** *motion extracted based on GMM* [20] **(c)** *motion extracted based on AMID*

Table II *Results comparison between GMM based, AMID without PDC and AMID with PDC for number of person*

	GMM based method [15]	AMID without PDC	AMID with PDC
"bus station"	Failed[2]	60%	87%
"subway station"	80%	87%	91%
"plaza"	failed	75%	88%

Correctness (PDC) and AMID based with PDC on three videos. The ground truth is given manually. The GMM background modeling method is used in method [15] for crowd information extraction. From the Table II, we can see that traditional method as Ref. [15] cannot work in "bus station" and "plaza" video, which are occupied with slow motion and high density. As for "subway station" video, which is easier for motion extraction, the GMM based method can get an accuracy of 80%. Our AMID based method can work in three scenes, which is better than GMM based method in "subway station" video. It should be noticed that PDC is useful for the number of person estimation especially for "bus station" and "plaza" videos. "Bus station" video seems flat and low and "plaza" video has good depth of field, which cause great perspective distortion, so the accuracy rate can be improved over 20% by the step of PDC. For "subway station" video, the distortion is little and the improvement is not much.

IV. Conclusion

In this study, we have proposed a crowd density estimation and prediction system for wide-area security. AMID based approach is applied to detect crowded areas and a geometry module is included to correct perspective distortion. The number of people in a crowd is estimated by the liner fitting method and the velocity is also obtained by the optical flow method. After crowd density and velocity are estimated, the prediction module is used to estimate the crowd density at designated points at a later time. Compared to existing methods, the proposed method is a real time system for applications and the crowd density analysis algorithm can work properly in both low and high crowd density scenes. Experiments and real applications demonstrate the effectiveness and robustness of our method in real scenes although there are some aspects to be improved in the system. In the future, we will consider how to choose the parameter (duration time) adaptation for different scenes.

Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grant No. 61175007; the National Key Technologies R&D Program under Grant No. 2012BAH07B01; and the National Key Basic Research Program of China (973 Program) under Grant No. 2012CB316302.

References

[1] ZHAN Beibei, MONEKOSSO D N, REMAGNINO P, *et al.* Crowd Analysis: A Survey[J]. Machine Vision and Applications, 2008, 19(5-6): 345-357.

[2] XU Liqun, ANJULAN A. Crowd Behaviors Analysis in Dynamic Visual Scenes of Complex Environment[C]// Proceedings of the 15th IEEE International Conference on Image Proces-

[2] The "failed" indicates that the algorithm provides a meaningless results on the video; the accuracy of its performance is lower than 20%.

sing, 2008 (ICIP 2008): October 12-15, 2008. San Diego, CA, USA, 2008: 9-12.

[3] GUO Jinnian, WU Xinyu, CAO Tian, *et al.* Crowd Density Estimation Via Markov Random Field (MRF)[C]// Proceedings of 2010 8th World Congress on Intelligent Control and Automation (WCICA): July 7-9, 2010. Jinan, China, 2010: 258-263.

[4] MA Ruihua, LI Liyuan, HUANG Weimin, *et al.* On Pixel Count Based Crowd Density Estimation for Visual Surveillance[C]// Proceedings of 2004 IEEE Conference on Cybernetics and Intelligent Systems: December 1-3, 2004. Singapore, 2004: 170-173.

[5] PARAGIOS N, RAMESH V. A MRF Based Approach for Real-Time Subway Monitoring[C]// Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition: December 8-14, 2001. Kauai, HI, USA, 2001: 1034-1040.

[6] LIN S F, CHEN J Y, CHAO H X. Estimation of Number of People in Crowded Scenes Using Perspective Transformation[J]. IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans, 2001, 31(6): 645-654.

[7] SUBBURAMAN V B, DESCAMPS A, CARINCOTTE C. Counting People in the Crowd Using a Generic Head Detector[C]// Proceedings of 2012 IEEE 9th International Conference on Advanced Video and Signal-Based Surveillance (AVSS): September 18-21, 2012. Beijing, China, 2012: 470-475.

[8] WU Xinyu, LIANG Guoyuan, LEE K K, *et al.* Crowd Density Estimation Using Texture Analysis and Learning[C]// Proceedings of IEEE International Conference on Robotics and Biomimetics, 2006 (ROBIO'06): December 17-20, 2006. Kunming, China, 2006: 214-219.

[9] DOLLAR P, TU Zhuowen, PERONA P, *et al.* Integral Channel Features[C]// Proceedings of 2009 British Machine Vision Conference: September 7-10, 2009. Rama Chellappa, UK, 2009: 1-11.

[10] BROSTOW G J, CIPOLLA R. Unsupervised Bayesian Detection of Independent Motion in Crowds[C]// Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06): June 17-22, 2006. New York, NY, USA, 2006: 594-601.

[11] CONTE D, FOGGIA P, PERCANNELLA G, *et al.* A Method for Counting People in Crowded Scenes[C]// Proceedings of 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS): August 29-September 1, 2010. Boston, MA, USA, 2010: 225-232.

[12] YU Haibin, HE Zhiwei, LIU Yuanyuan, *et al.* A Crowd Flow Estimation Method Based on Dynamic Texture and GRNN[C]// Proceedings of 2012 7th IEEE Conference on Industrial Electronics and Applications (ICIEA): July 18-20, 2012. Singapore, 2012: 79-84.

[13] YANG Hua, SU Hang, ZHENG Shibao, *et al.* The Large-Scale Crowd Density Estimation Based on Sparse Spatiotemporal Local Binary Pattern[C] // Proceedings of 2011 IEEE International Conference on Multimedia and Expo (ICME): July 11-15, 2011. Barcelona, Spain, 2011: 1-6.

[14] RABAUD V, BELONGIE S. Counting Crowded Moving Objects[C]// Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition: June 17-22, 2006. New York, NY, USA, 2006: 705-711.

[15] KILAMBI P, RIBNICK E, JOSHI A J, *et al.* Estimating Pedestrian Counts in Groups[J]. Computer Vision and Image Understanding, 2008, 110(1): 43-59.

[16] ZHANG Zhaoxiang, LI Min. Crowd Density Estimation Based on Statistical Analysis of Local Intra-crowd Motions for Public Area Surveillance[J]. Optical Engineering, 2012, 54(1): 047204-047204-13.

[17] ALI S, SHAH M. A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis[C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007 (CVPR'07): June 17-22, 2007. Minneapolis, MN, USA, 2007: 1-6.

[18] SAXENA S, BRÉMOND F, THONNAT M, *et al.* Crowd Behavior Recognition for Video Surveillance[C]// Proceedings of the 10th International Conference on Advanced Concepts for Intelligent Vision Systems (ACVIS'08): October 20-24, 2008. Juan-les-Pins, France, 2008: 970-981.

[19] SIRMACEK, B, REINARTZ, P. Automatic Crowd Density and Motion Analysis in Airborne Image Sequences Based on a Probabilistic Framework[C]// Proceedings of 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops): November 6-13, 2011. Barcelona, Spain, 2011: 898-905.

[20] ANDRADE E L, BLUNSDEN S, FISHER R B. Modeling Crowd Scenes for Event Detection[C]// Proceedings of the 18th International Conference on Pattern Recognition, 2006 (ICPR 2006): August 20-24, 2006. Hong Kong, China, 2006: 175-178.

[21] LI Wei, WU Xiaojuan, MATSUMOTO K, *et al.*

Crowd Foreground Detection and Density Estimation Based on Moment[C]// Proceedings of 2010 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), 2010: July 11-14, 2010. Qingdao, China, 2010: 130-135.

[22] HORN B K P, SCHUNCK B G. Determining Optical Flow[J]. Artificial Intelligence, 1981, 17(1-3): 185-203.

[23] STAUFFER C, GRIMSON W E L. Learning Patterns of Activity Using Real-Time Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 747-757.

Biographies

CAO Lijun, is a Ph.D. candidate in the area of pattern recognition and intelligent system, and an Engineer at the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), China. He received the M.S. degree in photogrammetry and remote sensing from Beijing University of Aeronautics and Astronautics, China in 2010. His current research interests include pedestrian detection and tracking, motion analysis and their application in real scenes. Email: ljcao@nlpr.ia.ac.cn

HUANG Kaiqi, is a Professor at the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), China. He received the M.S. degree in electrical engineering from Nanjing University of Science and Technology, China, and the Ph.D. degree in signal and information processing from Southeast University, China. He is a Senior Member of the IEEE. His current research interests include visual surveillance, digital image processing, pattern recognition and biological based vision. Email: kqhuang@nlpr.ia.ac.cn