

Estimation Of Number Of People In Crowded Scenes Using Amid And Pdc

**C. Jerlin Sheela Hani, **Dr.S.Sumathi M.E.,Ph.D.*

**PGscholar Department of electronics and communication Adhiyamaan college of engineering, Hosur
**Head of the Department Department of electronics and communication Adhiyamaan College of Engineering, Hosur*

Abstract: Crowd density estimation in wide areas is a challenging problem for visual surveillance. So the safety of people in a crowds has been a major problem. Hence proposed a system to estimate the number of people in crowded scenes using AMID and PDC for wide-area surveillance. 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. By using multi-camera network, the camera can predict the crowd 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. An 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. 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.

Foreground based methods: 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. 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. The number of people in a crowd is computed as a linear function of foreground pixels. Markov Random Fields (MRF) based approach is used to model changes in pixel value, and the optimal foreground is obtained by minimizing a MRF 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 and integral channel features based head detection 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. 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.

Group based methods: This method used 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. Accumulated Mosaic Image Difference (AMID) feature, 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. It has three contributions. First, the AMID feature 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, a novel prediction method for wide-area crowd analysis is done. Lastly, we develop a wide-area crowd surveillance system and the practical applications show the effectiveness of this system.

II. System Framework For Wide-Area Surveillance

2.1 AMID based crowd density estimation

As 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. It develops AMID feature to describe

these intra-crowd motions and estimate the crowd density. Local individual perturbations and movements are the main reasons causing intra-crowd.

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.

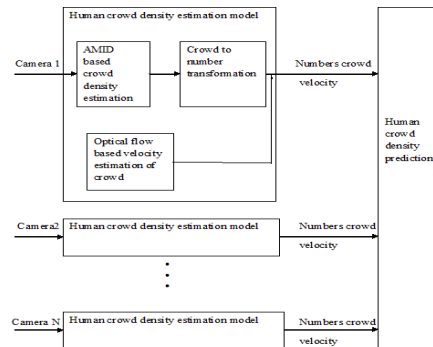


Fig.1:systemframework block diagram

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:

$$MID_t(m, n) = \begin{cases} 1, & \text{if } \|M_t(m, n) - M_{t-1}(m, n)\| > T_t \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where t is the current frame number; (m, n) is the middle location of the corresponding Mo82 saic 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.

A component of a vector; T_t is an adaptive threshold; and $M_t(m, n)$ is a representation value of each local area.

$$M_t(m, n) = \frac{1}{L_M^2} \sum_{M(i,j) \in MB(m,n)} I_t(i, j) \quad (3)$$

Here, $I_t(i, j)$ is the RGB vector of pixel (i, j) at frame $\#t$. LM 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 :

$$AMID_l(m, n, t) = \sum_{K \in W_l(t)} MID_K(m, n) \quad (4)$$

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

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

where $N'_{mid} = \lfloor \frac{N_{mid}}{N_{sw}} \rfloor N_{sw}$ is the valid length of observing time window after division, $t' = t - N'_{mid}$. The

AMID series is updated at every frame and provides amotions, 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:

$$U_t(m, n) = \begin{cases} 1, & \text{if } s_{mt}(m, n, t) - s_{mt}^0 < s_{mt}^{th} \\ & s_{mv}(m, n, t) > s_{mv}^{th} \text{ and } s_{ns}(m, n, t) > s_{ns}^{th} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

Where s_{mt}^{th}, s_{mv}^{th} and s_{ns}^{th} are the 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 intracrowd 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.

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

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

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

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

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: $\{(d_i, n_i)\} i = 1 \dots p$, 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 $[[d_s, d_{s+1}], (0 \leq s \leq p)$, then the number of people is

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

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

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

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. 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

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

by:
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 $[\theta_c - \theta_{offset}, \theta_c + \theta_{offset}]$ are used to get the speed and direction of crowd flow $\theta_p(k), \rho_p(k)$ with the average of k frame.

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

2.4 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 Figure2, 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%.

2.5 Crowd velocity estimation based on optical flow

In this set of experiments, we focus on the accuracy of the crowd velocity estimated by our proposed method. Here we apply our method on three video sequences shown in Figure 3. 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.

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

$$\theta_{real} = \frac{1}{n\delta} \left\| \arctan \frac{\Delta y_i}{\Delta x_i} \right\|$$

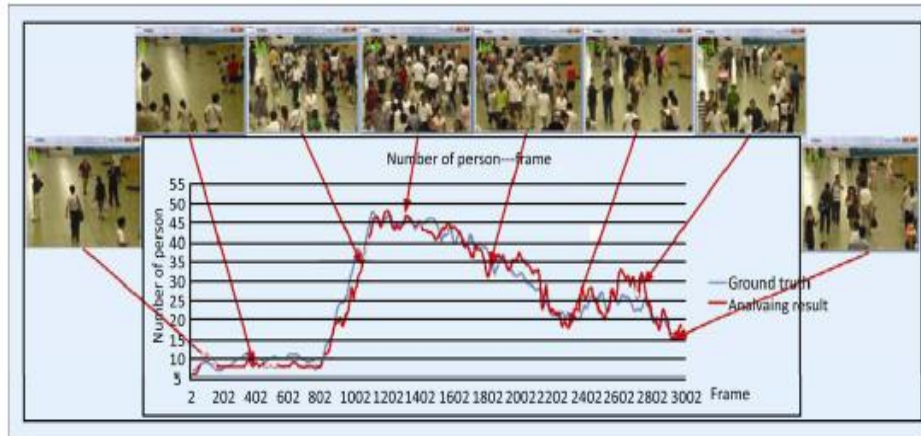


Fig.2 The curve of number of person vs. time for the “subway station” video and some selected frames

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 $f_c(y) = ((y_r - y_v)/(y - y_v))^2$. 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.

2.6 Comparisons and discussion

The crowd information extraction is an important step in crowd density estimation. Here the comparison of AMID method with the GMM method. 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.



Fig.3 Video sequences and the bounding boxes used for real data by hand.

Table I

	θ_{est}	θ_{real}	v_{est}	v_{real}
Video 1	64.2	71.3	0.271	0.301
Video 2	322.4	314.7	1.022	1.089
Video 3	267.5	265.2	0.163	0.221

	GMM based method	AMID without PDC	AMID with PDC
“bus station”	failed	60%	87%
“subway station”	80%	87%	91%
“plaza”	failed	75%	88%

In table I it gives the number of person result comparison between GMM based, AMID based without Perspective Distortion 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 for crowd information extraction. 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 known 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.

III. Conclusion

This paper discussed about 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.

References

- [1] 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.
- [2] 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.
- [3] 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.
- [4] CONTE D, FOGGIA P, PERCANNELLA G. 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. USA, 2010:
- [5] GUO Jinnian, WU Xinyu, CAO Tian 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.
- [6] MA Ruihua, LI Liyuan, HUANG Weimin, 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.
- [7] ZHAN Beibei, MONEKOSSO D N, REMAGNINO P. Crowd Analysis: A Survey[J]. Machine Vision and Applications, 2008, 19(5-6): 345-357.
- [8] 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.
- [9] KILAMBI P, RIBNICK E, JOSHI A J. Estimating Pedestrian Counts in Groups[J]. Computer Vision and Image Understanding, 2008, 110(1): 43-59.
- [10] WU Xinyu, LIANG Guoyuan, LEE K K., 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.