A Method of Detecting Abnormal Crowd Behavior Events Applied in Air Patrol Robot

Huailin Zhao, Shunzhou Wang, Shifang Xu
School of Electrical and Electronic Engineering, Shanghai Institute of Technology, Shanghai, China

Yani Zhang
School of Computer Science and Information Engineering, Shanghai Institute of Technology, Shanghai, China

Masanori Sugisaka
Alife Robotics Corporation LTD, Oita, Japan
zhao_huailin@yahoo.com

Abstract

When the ground or air patrol robot monitors a certain area, one of the important intelligent functions is to monitoring and alerting unusual event of the monitored area. This paper proposes a method via the number of people in the monitored area to detect the environment is safety or not. We use mature methods to formulate the counting problem and the number of people in each frame can be calculated with Gaussian Process Regression model. We suppose that this application may provides an important technical support for enhancing the patrol robot monitoring effect.

Keywords: Intelligent Surveillance, Gaussian Process Regression, Air Patrol Robot

1. Introduction

It is an important research topic to monitor the emergent events in large public places. With the development of robotics technology in recent years [1,2,3,4], unmanned aerial vehicles (UAVs) play a more and more important role in the field of intelligent monitoring. In this paper, an unmanned aerial vehicle (UAV) is used to monitor the abnormal event of public area. By monitoring the number of people in the specific area, it can alarm in a short time when abnormal number change happens.

A lot of works on the crowd counting algorithm have been studied. The crowd counting algorithm is currently divided into three categories: counting the number of people in the video[5,6,7,8], counting the number of people in a single image[9], and counting the number of people based on the deep learning[10,11]. The mature methods of counting the number of people in the video[5,6,7,12] is generally divided into three steps: 1) foreground segmentation 2) feature extraction 3) crowd regression. The methods of counting people in a single image[9,12] generally split the image into patches, and then extract the feature of each patch, and sum the number of each patch to get the total number of the picture. The crowd counting work based on the deep learning [10,11,12] usually use the population density map of the picture as the supervisory information, and design the convolutional neural network to regression the density map. The prediction of density map can be integrated to get the number of people in the frame.

However, the convolutional neural network performance is sensitive to the quality of the dataset images and training network process is long and hard. Considering the requirements of the performance of real-time and computational resource constraints, we repeat the reference[5,7] experiment and decide to adopt the video-based crowd counting algorithm proposed in the reference[5]. The model prediction results are used as the criteria for UAV monitoring unusual event. The contribution and innovation of crowd counting part in this paper is weak and if you are really interested in the crowd counting research topic we will spare no effort to recommend these works[5,6,7,9,10,11,12] .

2. Design of the Whole System

The flowchart of air patrol robot monitoring abnormal event in crowded public area is shown in Fig.1. Unmanned aerial vehicles monitor the large public places, the collected images are real-time transmit to remote monitoring terminal. The monitoring terminal use the crowd counting algorithm to count the images from the aircraft real-time, and record the number of changes over time. When the crowd number changes dramatically in a short term, the monitoring terminal remind the management that the region may happened abnormal event and should take actions or stay pay much attention to the area immediately.

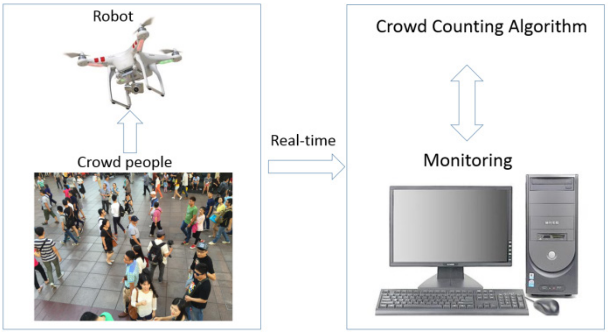


Fig.1. The flowchart of air patrol robot monitoring abnormal event in crowded public area

3. The Crowd Counting Algorithm

Video-based population counting algorithm[5,6,7]is generally divided into three steps: 1) foreground segmentation; 2) feature extraction; 3) regression. In this paper, the method[5] is used, the whole algorithm is shown in Fig.2.

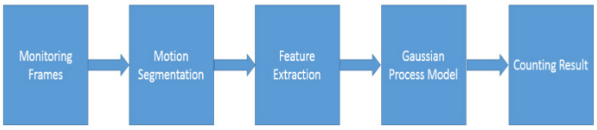
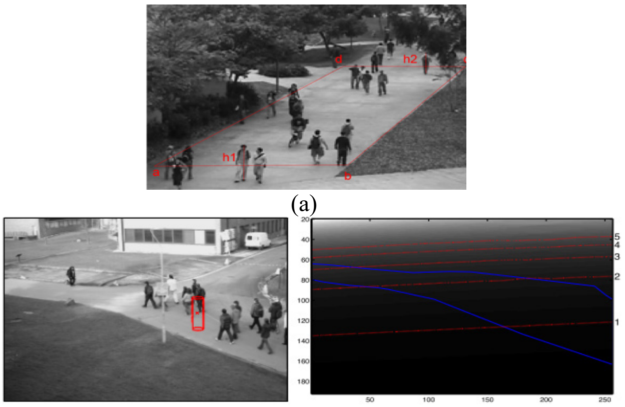


Fig.2. Monitoring video crowd counting system[5]

Due to the perspective of view, people who are close to the camera take more pixels in the image than those who are away from the camera. We use two different perspective normalization method to two different datasets. The perspective normalization is shown in Fig.3.



(a)

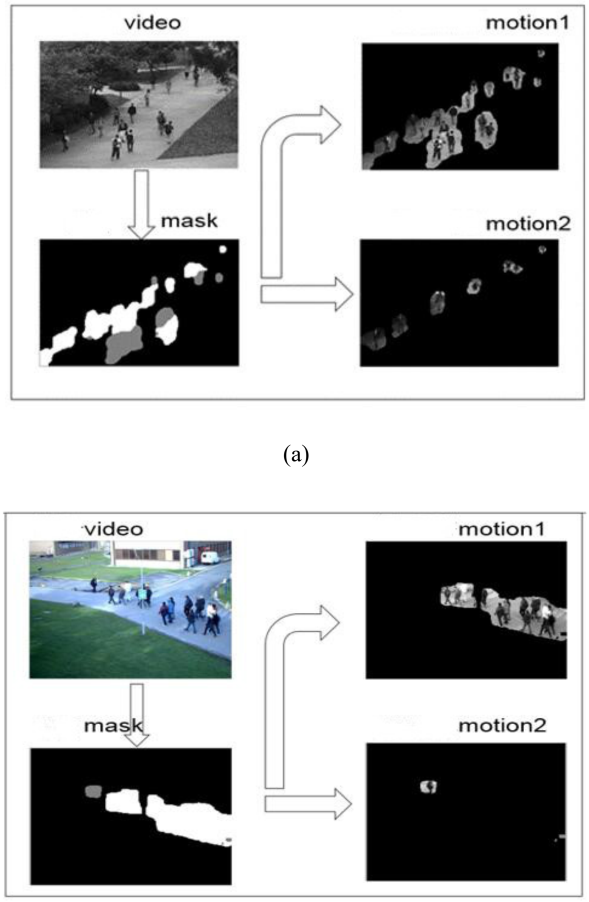
(b)

Fig.3. Perspective normalization (a)UCSD dataset (b)PETS2009 dataset

On the UCSD dataset, we make a ground plane[5], which is scaled to measure the height h_1 of the person at ab and the height h_2 of the person on the cd .We can see the ground plane as in Figure 3(a). The weight of the middle pixel is obtained by multiplying the pixels on ab and cd by the weights 1 and $\frac{h_1|\overline{ab}|}{h_2|\overline{cd}|}$ respectively, and the middle pixel weight is obtained by the linear interpolation between the two lines.

On the PETS2009 dataset, the perspective map is approximating a person moving in a 3-D scene to a cylinder with a height of 1.75 m and a radius of 0.25 m.[7] For each pixel (x, y) in the 2-D camera view, the cylinder is positioned in the 3D scene so that the center of the cylinder is projected to (x, y) in the 2-D view as shown in Figure 3(b), which is shown on the left. The total number of pixels used to fill the cylinder is expressed as c (x, y). The perspective is then calculated as $M(x,y) = c(230,123)/c(x,y)$, where the coordinates (230,123) correspond to the reference person on the right side of the sidewalk. Figure 3(b) is a perspective view with the contour line (red) which is indicated that the pixel weight at that location is {1, ..., 5}.

The purpose of segmentation is to segment the crowd people from the image to facilitate the subsequent step feature extraction. The performance of the segmentation is directly related to the final count precision, so it is an important factor limiting the performance of crowd counting algorithms. Commonly used segmentation algorithms[12] are Optical Flow, Mixture of Dynamic Textures[6], Wavelets and so on. The disadvantage of this motion-based foreground segmentation algorithm is obvious. If the person does not move in the video, the stationary person will be divided into the background, which affects the performance of the crowd counting[12]. In this paper, we use the Mixture of Dynamic Textures[6] to process the foreground segmentation on the UCSD dataset and PETS2009 dataset respectively. The concrete process is shown in Fig.4.



(a)

(b)

Fig. 4. Foreground segmentation based on mixture of dynamic textures (a)UCSD dataset (b)PETS2009 dataset

The mean absolute error (MAE) and the mean square error (MSE) are commonly used to measure the performance of the algorithm.

$$MAE = \frac{1}{N}\sum_1^N |z_i - \hat{z}_i| \qquad (1)$$

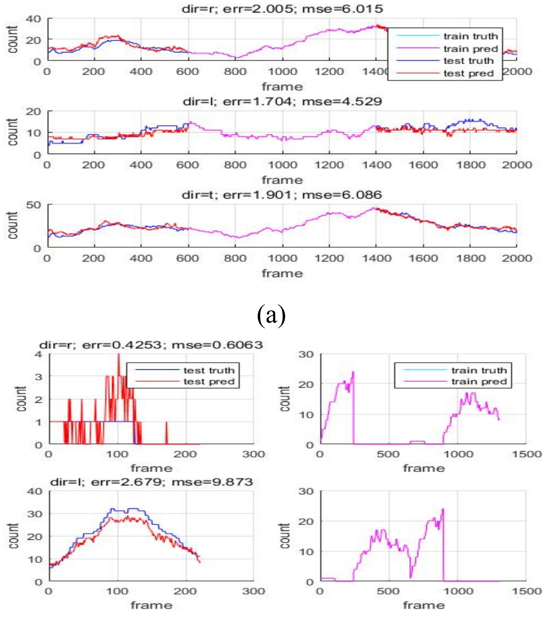
$$MSE = \sqrt{\frac{1}{N}\sum_1^N (z_i - \hat{z}_i)^2} \qquad (2)$$

Where N is the number of pictures to be tested (number of video frames). z_i is the number of the i-th frame. $\hat{z}_i$ is the estimated number of the algorithm.

After completion of the foreground segmentation, a variety of low-level features are extracted from the foreground (population) obtained from the segmentation. There are some common features[12]: Area and Perimeter of Crowd Mask, Edge Count, Edge Orientation, Texture Features, Minkowski Dimension, and so on. In this project, two datasets are tested respectively, and the feature of each image is saved as a feature vector. There are 29 features in each feature vector of the UCSD dataset[5]. Each feature vector of the PETS2009 dataset has 30 features[7].

4. Experiments

In this part, we repeat the reference papers[5,7] experiments. We use the regression model to regress the feature extracted in the previous step to the number of people in the image. The regression can be a simple linear regression, or a complex nonlinear regression. Commonly regression methods are Linear Regression, Piecewise Linear Regression, Ridge Regression, Gaussian Process Regression, and so on. We use Gaussian process regression to predict the number of people in the image of UCSD dataset and PETS2009 dataset. The prediction result is shown as in Fig.5.



(a)

(b)

Fig.5. Crowd counting method employed in (a)UCSD dataset and (b)PETS2009 dataset

For the UCSD dataset, 800 pictures are used as the training set, and the remaining 1200 pictures are used as test set. The output of the Gaussian process regression is rounded to the nearest integer to generate a population count and record the mean square error (MSE) and the mean absolute error between the estimate and the true value. The results of the crowd counting algorithm for counting the UCSD dataset are shown in Figure 5(a)

The MSE of the two groups were 6.015 and 4.529, respectively. The accuracy of the algorithm basically meet the design requirements.

The whole area of scene S1.L1 of PETS2009 dataset was tested by the crowd counting algorithm, and 1308 images are used as training set and the video 13-57, 13-59F, 13-59F, 14-03,14-03F, 221 images are used as a test set. The output of the Gaussian process regression is rounded to the nearest integer to generate a population count and record the mean square error (MSE) and the absolute error between the estimate and the true value. The population count algorithm for the PETS2009 data set scene S1.L1 count results shown in Figure 5(b), the MSE of the two groups of people counting are 0.6063 and 9.873, the accuracy of the crowd counting algorithm basically meet the design requirements.

5. Conclusion

In this paper, we design a monitoring function of patrol robot to detect the abnormal crowd behavior events applied in the monitoring area with the changes of crowd counting results. We choose the typically machine learning method rather than convolution neural network to apply in this specific filed due to the real-time performance and the computation of the algorithm. We repeat the reference papers experiment and the prediction result of the crowd counting algorithm in our experiment is not the state of the art, but can basically meet the design purposes. In the future, we will use the frames we collected from real world rather than use the existing datasets to test our whole system performance and we will design a new crowd counting algorithm to increase the accuracy and attempt to apply deep learning in the this whole patrol robot monitoring system.

References

1. Shunzhou Wang, Huailin Zhao, Xuyao Hao. Design of An Intelligent Housekeeping Robot Based on IOT. *Proceedings of 2015 International Conference Intelligent Informatics and BioMedical Sciences* (ICIIBMS2015), 2015.11, p197-200.
2. Huailin Zhao, Lin Wang, Bei Wang, Masanori Sugisaka, “System development of an artificial assistant suit”, *Artificial Life and Robotics* (ISSN1433-5298), Vol.17, 2013, No.3-4, p331-335.
3. Huailin Zhao, Bei Wang, “Configuration of the Mckibben Muscles and Action Intention Detection for an Artificial Assistant Suit”, *International Journal of Advanced Robotic Systems* (ISSN 1729-8806), Vol. 9, 2012, p1-7.
4. Huailin Zhao, Masanori Sugisaka, “Simulation study of CMAC control for the robot joint actuated by McKibben muscles”, *Applied Mathematics and Computation* (ISSN0096-3003), Vol.203, 2008, p457-462
5. Chan A B, Liang Z S J, Vasconcelos N. Privacy preserving crowd monitoring: Counting people without people models or tracking. *Computer Vision and Pattern Recognition, 2008. CVPR 2008.* IEEE Conference on. IEEE, 2008: 1-7.
6. Chan A B, Vasconcelos N. Modeling, clustering, and segmenting video with mixtures of dynamic textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(5): 909-926.
7. Chan A B, Morrow M, Vasconcelos N. Analysis of crowded scenes using holistic properties. *Performance Evaluation of Tracking and Surveillance workshop at CVPR*. 2009: 101-108.
8. Fengzhi Dai, Huailin Zhao, Resonance algorithm for image segmentation, *Computer Engineering and Design* (ISSN1000-7024), Vol.28 (No.23), 2007, p5657-5660.
9. Idrees H, Saleemi I, Seibert C, et al. Multi-source Multi-scale Counting in Extremely Dense Crowd Images. *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2013:2547-2554.
10. Zhang C, Li H, Wang X, et al. Cross-scene crowd counting via deep convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015: 833-841.
11. Zhang Y, Zhou D, Chen S, et al. Single-image crowd counting via multi-column convolutional neural network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016: 589-597.
12. Zhang Y. Research on Crowd Counting Estimation. http://mp.weixin.qq.com/s?__biz=MzI1NTE4NTUwOQ==&mid=2650325105&idx=1&sn=939f46fd57a80f86b9eb050dd2cf1652&scene=4#wechat_redirect. 2016.