Using Deep Learning Models Combined with Crowd Emotion Models to Identify Abnormal Behaviors in Crowds

Xiao Li¹,a, Yu Yang¹,* Linyang Li² and Yiming Xu¹,b
¹College of Information Engineering, Engineering University of PAP, Xi’an Shaanxi China
²Department of Basic Course, Engineering University of PAP, Xi’an Shaanxi China
Email: aannytiny@163.com; banny119988@163.com

Abstract. Aiming at the problem that the definition of crowd abnormal behavior detection is ambiguous and difficult to combine with context semantics, an algorithm using OCC human emotion model combined with crowd entropy is proposed. First calculate the crowd entropy for the crowd, and determine whether the entropy value is abnormal, if it is abnormal, further extract the optical flow OF and HOG. Then project it into two-dimensional vector data, send it to CNN for local feature extraction and combine with OCC model to achieve the description of crowd emotions. Finally, predict whether the abnormality occurs according to the judgment factor. Verified on the data set, this method shows a high accuracy.

1. Introduction
The detection of crowd abnormal behavior is divided into crowd density estimation and crowd behavior recognition, and the latter is the focus and difficulty of current research. Researchers have proposed various methods for directly extracting the trajectory state of the crowd in the video from the underlying image information features to the high-level structural modeling. For low-density people, the methods of individual appearance abnormality detection and individual behavior recognition[1] are mostly used, such as analysis methods based on behavior feature extraction, behavior analysis methods based on wavelet transform algorithm and analysis methods based on rigid body kinematics, of which the most commonly used. It is a behavior analysis method based on behavior feature extraction. Commonly used features are Histogram of Oriented Gradient (HOG), Haar-like, Edgelet, and color features. For medium and high-density people, the method of overall analysis is mostly used, such as judging the change of the entropy value of the people[2,3], mutation of the crowd’s exercise energy[4, 5], abnormal trajectory[6, 7], etc. The definition of crowd behavior abnormality has always been a puzzle. For the abnormality of individuals, after the training of the appearance characteristics of normal individuals, when the behavior of the test set is different from the trained model, it can be judged as abnormal. And for the movement characteristics of the crowd, we usually use the parameters deviate from the threshold, such as entropy value, CMI peak value[8] and motion energy, etc.

2. Proposed Algorithm
First, calculate the entropy of the crowd and determine whether it exceeds the threshold. If it exceeds the threshold, the next calculation is performed. Otherwise, it is determined as a normal crowd; for the crowd with abnormal entropy value, the optical flow (OF) and gradient histogram from the tension-based motion descriptor are combined, and then projects it into two-dimensional vector data, through CNN further feature extraction, combined with the OCC model to describe the crowd’s emotions.
Finally achieve accurate description of crowd characteristics and crowd abnormal behavior of detection.

**Contribution:** 1) Propose to combine motion features and emotional psychological features to fully describe different crowd attributes; 2) Use wavelet transform to project features into a two-dimensional view, and then input into a simple deep learning framework, without affecting the speed of calculation, to realize accurate detection and early warning of crowd behaviors; 3) It can achieve rapid and accurate detection of abnormal behaviors of medium and high-density crowds.

![Flow chart of this article](image)

### 3. Feature Extraction

#### 3.1. Image Segmentation and Judgment of Entropy

In the video, useful information is partial, so the region of interest needs to be extracted. This paper uses the frame difference method to extract the foreground target area. In the middle and high-density crowd, because the crowd is relatively crowded, the feature extraction of the entire motion area will have a large computational complexity, so we segment the motion area.

When initially judging the abnormality of the crowd, the crowd entropy is used. The calculation formula of entropy value is as follows:

\[
S(W) = -\sum_{i=1}^{||A||} p(Q_i) \ln p(Q_i)
\]  

Where W is the state set composed of all microscopic states of the crowd at a time, \(W = \{Q_i; i = 1,2,...,||A||\}\), and \(||A||\) is the total number of microscopic state of the crowd at a certain time. \(p(Q_i)\) is the probability of the crowd appearing under \(Q_i\). \(S\) represents the statistical average value of the probability \(p(Q_i)\) in each microscopic state \(Q_i\), and \(\ln p(Q_i)\) is the weight. Theoretically, the maximum value of entropy is \(\ln ||A||\), and half of this value is used as the threshold. When it exceeds 3/2 of the threshold, anomaly occurs.

#### 3.2. Extraction of Emotion Information

##### 3.2.1. Extraction of OF

In the polynomial-based model, we use the linear combination of orthogonal polynomials proposed to approximate the optical flow vector field.

Where \(V^1(x_1,x_2)\) and \(V^2(x_1,x_2)\) are the horizontal and vertical displacements of point x. The optical flow \(V^1\) and \(V^2\) are approximated by each projection function \(V\) on each polynomial \(P_{ij}\).

Definition F:

\[
F = (V^1(x_1,x_2),V^2(x_1,x_2)).
\]

\[
V^1(x_1,x_2) = \sum_{i=0}^{g} \sum_{g=0}^{g-1} v^1_{ij} P_{ij} \quad \text{and} \quad V^2(x_1,x_2) = \sum_{i=0}^{g} \sum_{g=0}^{g-1} v^2_{ij} P_{ij}.
\]
\[ T^{OF} = \sum_b T_f. \] (4)

3.2.2. Histogram of Gradient Tensor

Here we need to calculate the gradient of all \( n \) points in the image \( I_i \), through the histograms of gradients \( h_j = h_{k,l} \), where \( k \in [1,nb_{\theta}] \) and \( l \in [1,nb_{\psi}] \), where \( nb_{\theta} \) and \( nb_{\psi} \) are respectively represents the number of \( \theta \) and \( \psi \). We used equally-divided intervals to obtain \( nb_{\theta} \) and \( nb_{\psi} \) bins.

The HOG tensor of frame \( j \) with \( m \) bins is:
\[ T^{HOG} = \sum_j h_j h_j^T. \] (5)

Calculate the extracted histogram of the entire image, and each cell of the image is represented by a vector. We use a grid of \( n_x \) and \( n_y \) blocks to divide the \( f \) frame into \( x \) and \( y \) directions. Now calculate the tensor of all bins in frame \( j \):
\[ T_j = \sum h^{CP}_j h^{CP}_j^T. \] (6)

The global descriptor of the segmented image and gradient histogram is calculated as:
\[ T^{HOG} = \sum_{j=1}^n T_j. \] (7)

Finally, the histograms of the approximation of the tensor, the optical flow approximation and the gradient approximation are obtained from the video data:
\[ T = T^{OF} + T^{HOG} \]

3.2.3. Two-dimensional Projection of Features

For the extracted OF and HOG feature vectors, converted to a two-dimensional map, we use wavelet transform, by calculating the vertical and horizontal projections of some wavelet sub-band quantization coefficients, and then vertically stitching to obtain a two-dimensional map of features.

3.3. Emotional Analysis

The OCC model is based on the concept of evaluation and attributes emotional arousal to a subjective interpretation of the individual environment. The judgment of the emotions of the existing crowd can be used for behavior prediction, and the detection of the emotions of the existing crowd needs to be combined with the PAD (pleasure-arousal-dominance) model, which determines the average emotional state of the representative life situation samples described by Mehrabian[10]. OCC emotion is always related to PAD status[11]. The PAD space uses such a map and its three orthogonal scales to evaluate emotional tendencies.

When the crowd's emotions exceed a certain threshold, the individual is likely to have the ability to communicate emotions. The threshold of expression ability of individual \( j \) comes from the normal distribution with mean value \( 0.5 - 0.5\psi^E \) and standard deviation \( (0.5 - 0.5\psi^E)/10 \). We classify this group of people as a triggerable group, which can be used to predict the next behavior of this group based on the type of emotion affected. We use the above two models to preliminarily judge the emotional state of the crowd, and use it as an input of CNN, and the projected two-dimensional data as another input, to further extract and classify the characteristics of the crowd.

4. Experiments and Analysis

On 4.0 GHz CPU, 64-bit Windows 10 operating system, MATLAB 2018a, Open CV and python programming languages are used as development tools for simulation experiments. In order to verify the effectiveness of the proposed algorithm, this paper selects the UCSD dataset and MED[12] as the experimental objects.

Compared with the method in the literature on the UCSD data set, the results of AUC and equal error rate EER are shown in Table 1. It can be seen that the method we proposed shows good performance on Ped1 and can effectively monitor abnormal frames. But the values of AUC and EER are not optimal on Ped2, which may be related to the shooting angle.
Table 1. AUC and EER(%) of various published method

| Ped1 |       |       | Ped2 |       |       |
|------|-------|-------|------|-------|-------|
| Approach | AUC | EER(%) | Approach | AUC | EER(%) |
| HOOF | 0.514 | 48.9 | HOOF | 0.612 | 40.2 |
| HOFM | 0.714 | 32.9 | HOFM | 0.898 | 19.1 |
| MDT temporal | 0.830 | 22.8 | MDT temporal | 0.766 | 27.8 |
| MDT spatial | 0.610 | 43.8 | MDT spatial | 0.751 | 28.9 |
| Adam(LMH) | 0.634 | 38.9 | Adam(LMH) | 0.582 | 45.8 |
| Proposed | 0.912 | 18.1 | Proposed | 0.751 | 30.1 |

Table 2 is a comparison experiment on the MED dataset. The comparison method is to extract low-level features. It can be seen that our algorithm has an accuracy of up to 90.91, indicating that compared to low-level features, our method using emotion-based combined motion information can obtain a more accurate description of the behavior.

Table 2. performance of our approach compared with other methods in MED

| Methods | Accuracy |
|---------|----------|
| HOG     | 38.90    |
| HOF     | 37.68    |
| MBH     | 38.52    |
| Dense trajectory | 38.72 |
| Emotion attribute | 43.56 |
| Our approach | 90.91 |

5. Conclusions
This paper uses the entropy value to make an initial judgment of anomalies, combined with the crowd emotion models OCC, PAD to describe the crowd's emotional psychological characteristics, and as an input of CNN, combined with the optical flow and gradient tensor to extract the crowd's motion features to achieve abnormal accurate detection of behavior. The main advantage of the tensor-based model is that there are many changes in the situation, such as the emergence of new moving objects, there is no need to add new descriptors, that is, strong scene adaptability. The introduction of emotion models can realize the semantic interpretation of behaviors and increase the accuracy of the judgment of crowd status.

6. Acknowledgments
This work was supported by Foundation Research Fund of Engineering University of PAP (WJY201906) and Scientific Research Project of Equipping the Armed Force in 2018 (WJ20182A020020-2).

7. References
[1] Zitouni M S , Sluzek A , Bhaskar H . Towards understanding socio-cognitive behaviors of crowds from visual surveillance data[J]. 2019
[2] XUGUANG ZHANG , QINAN YU1, HUI YU. Physics Inspired Methods for Crowd Video Surveillance and Analysis: A Survey.2018
[3] Akpulat M , Ekinci M . Detecting interaction/complexity within crowd movements using braid entropy[J]. Frontiers of Information Technology & Electronic Engineering, 2018
[4] Lamba S , Nain N . Detecting anomalous crowd scenes by oriented Tracklets' approach in active contour region[J]. Multimedia Tools and Applications, 2019, 78(22):31101-31120
[5] Zhao C. A Survey on Image Style Transfer Approaches Using Deep Learning[J]. Journal of Physics: Conference Series, 2020, 1453:012129
[6] Halbe M, Vyas V, Vaidya Y M. Abnormal Crowd Behavior Detection Based on Combined Approach of Energy Model and Threshold[C]// International Conference on Pattern Recognition and Machine Intelligence. 2017

[7] Feizi A. Hierarchical detection of abnormal behaviors in video surveillance through modeling normal behaviors based on AUC maximization[J]. Soft Computing, 2020:1-13

[8] Rabiee H, Haddadnia J, Mousavi H, et al. Novel dataset for fine-grained abnormal behavior understanding in crowd[C]// IEEE International Conference on Advanced Video & Signal Based Surveillance. IEEE, 2016

[9] Halbe M, Vyas V, Vaidya Y M. Abnormal Crowd Behavior Detection Based on Combined Approach of Energy Model and Threshold[C]// International Conference on Pattern Recognition and Machine Intelligence. 2017

[10] Mota, V.F., de Almeida Perez, E., Da Maurilio, M.L., Vieira, M.B., Gosselin, P.-H.: A tensor motion descriptor based on histograms of gradients and optical flow 39(1), 85–91 (2014)

[11] A. Mehrabian, “Pleasure-arousal-dominance: A general frame work for describing and measuring individual differences in temperament,” Current Psychol., vol. 14, no. 4, pp. 261–292, 1996

[12] Rabiee, H., Haddadnia, J., Mousavi, H., Kalantarzadeh, M., Nabi, M., Murino, V.: Novel dataset for fine-grained abnormal behavior understanding in crowd. In: 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS, pp. 95–101. IEEE (2016)