



University of HUDDERSFIELD

University of Huddersfield Repository

Hao, Yu, Xu, Zhijie, Wang, Jing, Liu, Ying and Fan, Jiulun

An Approach to Detect Crowd Panic Behavior using Flow-based Feature

Original Citation

Hao, Yu, Xu, Zhijie, Wang, Jing, Liu, Ying and Fan, Jiulun (2016) An Approach to Detect Crowd Panic Behavior using Flow-based Feature. In: Proceedings 22nd International Conference on Automation and Computing. IEEE. ISBN 9781862181328

This version is available at <http://eprints.hud.ac.uk/id/eprint/29079/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

<http://eprints.hud.ac.uk/>

An Approach to Detect Crowd Panic Behavior using Flow-based Feature

Yu Hao¹, Zhijie Xu¹, Jing Wang¹, Ying Liu², Jiulun Fan²

¹ School of Computing and Engineering, University of Huddersfield, Huddersfield, UK

¹{y.hao; z.xu; j.wang2}@hud.ac.uk

² School of Telecommunication Engineering, Xi'an University of Posts & Telecommunication, Xi'an, China

²ly_yolanda@sina.com,²jiulunf@xupt.edu.cn

Abstract—With the purpose of achieving automated detection of crowd abnormal behavior in public, this paper discusses the category of typical crowd and individual behaviors and their patterns. Popular image features for abnormal behavior detection are also introduced, including global flow based features such as optical flow, and local spatio-temporal based features such as Spatio-temporal Volume (STV). After reviewing some relative abnormal behavior detection algorithms, a brand-new approach to detect crowd panic behavior has been proposed based on optical flow features in this paper. During the experiments, all panic behaviors are successfully detected. In the end, the future work to improve current approach has been discussed.

Keywords—video processing; behavior detection; optical flow

I. INTRODUCTION

Sudden hazardous behaviors such as brawl or stampede in public area bring significant threatens to the public security. If these behaviors can be predicted, or an alarm can be automatically sent to security department while the event happened, more casualties could be avoided. Since most of urban public areas have surveillance CCTV cameras installed, if the incoming video streams could be analyzed automatically in real-time or close (so-called online) to detect abnormality, human operators might be alerted immediately to verify the event. In this case, more efficient actions could have taken place to reduce the scale of the tragedy.

Behaviors in public area could be categorized into crowd behaviors and individual behaviors. Crowd behaviors are recognized as a group of individual behaviors sharing certain motion connections and impacts. This type of behavior usually occurs when there is a large amount of people in the scene, and each individual is represented by limited number of pixels, and also, actions of the majority parts of the crowd are showing certain patterns. For example, in a panic situation, the crowd is attempting to avoid from something hazardous, thus the dominant moving flow is escaping from certain points and forming a circular shape, or rushing off with high velocity to the same direction. Figure 1 shows some possible flow patterns of crowd behaviors.



Fig. 1. Some possible crowd behavioral patterns. The first one could be the pattern from a cross road; the second one could be at a pedestrian lane; the third and fifth one could be formed at a panic scene; and the fourth one could be a pattern on a Roundabout. This figure is provided by Solmaz [1]

Individual behaviors, on the other hand, usually exist in a relative small part of the scene, which may be surrounded by dominant crowd's behaviors, or exists in a sparse scene with low crowd density. For example, pocket picking in crowd and trespassing. Different from crowd behaviors, detection of individual behaviors requires rich information from the individual objects and local environment, which is a great challenge to the most of the CCTV system because the CCTV cameras are not usually designed to focus on local areas and the resolution of those cameras are usually very low.

This research has been concentrated on the detection of crowd behaviors which have the potential to cause public security issues such as panic and congestion. For example, a serious congestion could be a premise of stampede, which may result heavy casualties. When looking into crowd behavior, it can be further categorized into structured and unstructured crowd behaviors. Li [2] defines the structured pattern as the crowd moves coherently in a common direction, the motion direction does not vary frequently, and each spatial location of the scene contains only one unified crowd behavior over the time. Unstructured crowd patterns represent the scenes with chaotic or random motions, where participants move in different directions at different times, and each spatial location contains multiple crowd behaviors. These scenes possess different dynamic and visual characteristics. Thus distinguishing structured and unstructured crowd patterns could be a feasible way to detect abnormality.

II. LITERATURE REVIEW

To detect crowd behaviors, the very first step is to extract useful feature data from the video stream. The most frequently used features for crowd abnormal behavior detection include global flow-based features and local spatio-temporal based features.

By setting a grid of points on two consecutive frames, motion information of each point is stored as global flow-base feature. The motion information could be optical flow [1], force flow [5], and tracklet [14]. Global flow-based feature describes overall motion patterns in scene, dominant motion magnitude and tendency can be obtained from it. Li [2] gave the definition of optical flow: optical flow is to compute pixel wise instantaneous motion between consecutive frames. Optical flow is robust to multiple and simultaneous camera and object motions, and it is widely used in crowd motion detection and segmentation. However, optical flow does not capture long-range temporal dependencies, and cannot represent spatial and temporal properties of a flow. Several researches are conducted to enhance the performance of typical optical flow. Mehran [3] introduced a so-called streak flow technique by using the notion of streak line to mark-up the motion field for crowd analysis. The study has also provided critical comparisons of optical flow, particle flow, and streak flow methods. Impressive results employing particle flow have been demonstrated on crowd segmentation and abnormal crowd behavior detection. However, in particle flow the spatial changes are still ignored, and time consuming is significant. Dirk and Peter [4] proposed a model to describe the interaction forces between individuals in a crowd - named Social Force Model (SFM). According to the SFM concept, Mehran [5] obtained desired velocity of each particle using the extracted actual particle velocity but modified by the SFM equation.

Flow based global scale features can be utilized to detect dominant events, however for the unstructured high density crowd scenes, even a fine-grain representation such as optical flow would not provide enough motion information for processing. Thus spatio-temporal features are used to detect abnormality compensate the deficiencies. The related methods generally consider the motion as a whole, and characterize its spatio-temporal distributions based on local 2D patches and 3D cubic regions. Local spatio-temporal features have good performance in motion understanding due to their strong descriptive power, and unlike global flow based features, the temporal information is preserved. Adelson and Bergen [6] first introduced Spatio-Temporal Volume (STV). STV shows promising global feature representation and pattern recognition potentials inherent from its nature. An STV model is capable of encapsulating static and dynamic video content features, hence simplifying an event recognition task into corresponding 3D geometric feature extraction and matching operations. A framework is proposed to detect abnormal behavior using STV by Wang and Xu [7]. They introduced an innovative pattern recognition algorithm developed to harness the promising characteristics of the STV event models. A region intersection (RI) based 3D shape-matching method is proposed to compare the STV shapes extracted from video inputs to the predefined 3D event templates. Chan [8] proposed a new generative model named as Mixture of Dynamic Textures (MDT), in which a collection of video sequences are modeled as samples from a set of underlying dynamic textures. Mahadevan [9] introduced dynamic texture based models of normalcy over both space and time in his abnormal behavior detection framework.

Once features are extracted, different approaches can be implemented to detect abnormal behaviors using these

information. As previously explained, crowd behaviors can be categorized into structured and unstructured behaviors. If the studied scene is structured, a crowd behavior can be detected by simply matching predefined templates to the live-feed. However to the unstructured scene, a dominant motion flow does not exist, hence further processes will be required, such as using relevant histogram-base operations to identify certain patterns.

Barbara Krausz [10] used a video set of crowded parade to compute two-dimensional histograms of motion magnitude and motion direction of the flow vectors of the entire frame. All of the resulting two-dimensional histograms are clustered using the k-means algorithm. Non-Negative Matrix Factorization (NMF) is applied to decompose complex histograms of magnitude and motion direction into a set of five basis histograms. A symmetry factor is then calculated to detect congestion in current scene. Solmaz [11] used Jacobian matrix of global optical flow, and introduced two eigenvalues to determine the dynamic stability of points in the optical flow. Dominant crowd behaviors are divided into five categories, which are Bottlenecks, Fountainheads, Lane Formation, Ring/Arch Formation and Blocking. Unknown crowd events are transformed into eigenvalues-based patterns to match those templates and to detect current crowd state. Mehran [5] calculated social force flow with extracted optical flow, then K-means clustering is applied on non-zero flow area to obtain several clusters. With a corpus of clusters, Latent Dirichlet Allocation (LDA) [12] is implemented to discover the distribution of L topics for the normal crowd behavior. Using the modified Expectation Maximization (EM) algorithm in [12], Bag of Words (BoW) model is used to maximize the likelihood of corpus. Based on a fixed threshold on the estimated likelihood, frames can be labeled normal or abnormal.

This paper's approach proposed a solution utilizing global flow based features to fast detect crowd's panic event, with less time consumption comparing to previously introduced approaches. Some disadvantages such as adaptive ability and accuracy are also discussed in the following section.

III. A CROWD BEHAVIOR OF PANIC DETECTION APPROACH

In this research, an approach is proposed to detect crowd abnormal behavior of panic. The typical pipeline of an abnormal behavior detection system often include phases and modules for data extraction, model training, anomaly detection and behavior matching. The framework of different phases are shown in Figure 2.

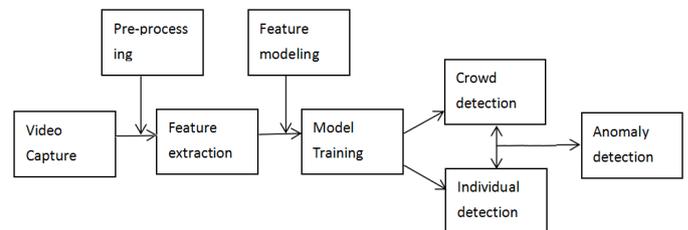


Fig. 2. Detailed crowd abnormal behavior detection framework

The very first step is video capturing that is usually obtained from standard video sources such as online CCTV or off-line video databases. In the following section, UMN dataset [13] will be used extensively for experiment design and test purposes. Before extracting features such as calculating the optical flow, some pre-processing operations such as background subtraction can be implemented, in order to get smoother feature and save computational time. Once features are extracted, the feature values sometimes need to be modeled, because it is not sufficient to detect the abnormal behavior with low level raw features in most of the time. Next with the modeled feature data, histograms or probability models are built up, then the crowd or individual behaviors can be analyzed.

The operational flows of the proposed system are explained in the pseudo code as listed as Figure 3.

<p>Initialization For each frames k</p> <p>Feature Extraction Calculate Optical flow between k and $k+1$ Apply a neighborhood average to smooth the obtained optical flow features</p> <p>Feature Modeling Calculate the summation S of absolute value of optical flow vectors of current frame k and $k+1$</p> <p>Model Training Calculate the average value A of S in the first F frames Assume a threshold T</p> <p>Anomaly Detection After training phase, IF $S - A > T$ THEN return abnormal ELSE return normal</p>

Fig. 3. Pseudo Code of the proposed crowd abnormal behavior detection model

A. Initialization

In this step, video footages are loaded and parameters are set up, such as the F value of training length, the threshold T etc. When the value F is set, the algorithm will use first F frames of video footage to complete the training stage. The F value should be adjusted to adapt different circumstances. The threshold value T will be used to judge if current state is abnormal in anomaly detection phase. Instead of using real time video streams, UMN was used for its popularity in system benchmarking and cross referencing.

Then for each frames k , the following procedure is implemented repetitively.

B. Feature Extraction

Firstly, taking frames k and $k+1$ from loaded video data, then apply Horn Shunck Optical flow method to obtain optical flow feature data, noted as u_k . Before further processing, a neighborhood average to u_k will be performed to reduce noise, then processed u_k is noted as v_k . The comparison between u_k and v_k is shown in the Figure 4, the left figure represents u_k , the middle figure represents v_k and the image to the right is frame

k . It is noticed that the left figure correctly show the shapes of individuals in frame k . However the shapes are not clear enough to observe. After apply neighborhood average, in the middle figure the shapes of individuals look clearer, which make the patterns easier to utilize. In addition, a cluster algorithm can be applied to the processed optical flow for a better result.

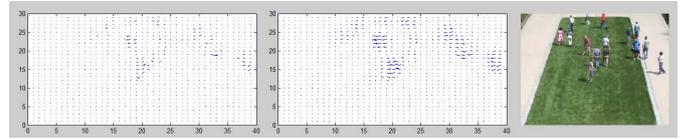


Fig. 4. A comparison of optical flows before and after the neighborhood mean procedure

C. Feature Modeling

For panic behavior, it can be expected that motion velocity of people in the scene could vary drastically before and after the abnormal event triggered. Figure 5 shows the changing of optical flow motion magnitude between normal and abnormal state, it's easy to observe that in panic state, the motion magnitude is greater than itself in normal state. It may not easy to analyze the motion change of each person, because the change isn't large enough compared to noise, the detection result could be inaccurate. However the global changing magnitude could be significant large to use, by calculating the summation of every motion magnitude in the scene.

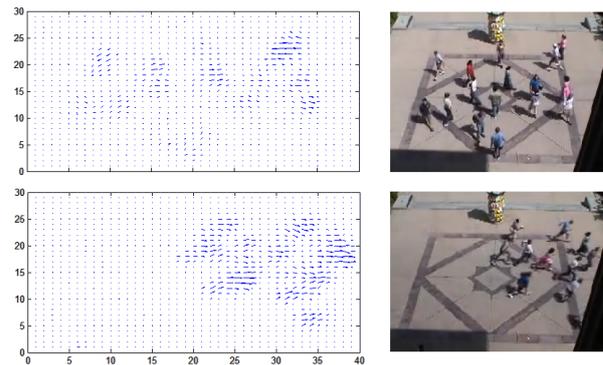


Fig. 5. Changing of Magnitude in Panic event

To utilize the extracted and processed optical flow features, the summation S is calculated based on the absolute value of v_k , as in (1)

$$S = \sum_{w=1}^W \sum_{h=1}^H |v_k^{w,h}| \quad (1)$$

in which w is the width of matrix v_k and h is the height of matrix v_k .

D. Model Training

Once examined the pattern variations of the summation value, a reasonable threshold can be set up to inspect if the incoming frames are going through changes from "normal" to "abnormal". There are two approaches to decide a threshold, (a) building up and a training model through learning the normal states, (b) defining the proper threshold value based on users' experience in an empirical manner. Here both

approaches are utilized to represent the normal state. The summation value of several initial frames of normal state are used to obtain an average value, this value shows normal state. In this experiment the threshold value to detect abnormality is fixed, however in future work it's value can be automatically set depends on current scene. In this phase, the average value of S of first F frames, named A , is calculated.

$A = AVG(S_1, \dots, S_F)$. Then based on obtained A , a reasonable threshold T is chosen for abnormality detection. The following Figure 6 shows the difference between S values and A values. Because the overall velocity of the crowd is low and constant, the shape of curve remains stable in first one hundred frames. By observing the curve, we set the threshold value T to 20 in this case. Noticed that in training phase, there is no need distinctively marking or detecting any abnormality.

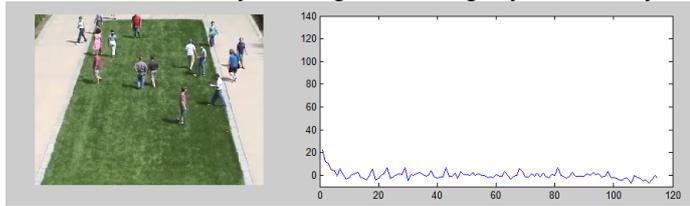


Fig. 6. Difference values s between S and A value for each frame

E. Anomaly Detection

Once the model training phase is completed, $k > F$, the system is ready to detect crowd anomaly. For each frame k , if the absolute difference of S and A is greater than T , the current frame is considered abnormal, otherwise it is considered normal. It could be expressed as $|S_k - A| > T$, if this condition is true, return abnormal, and vice versa. The following Figure 7 shows the detected abnormality, the abnormal behavior begins after 450 frames. When the abnormality is detected, the word 'anomaly' is shown at left top corner of the figure. It is observed when the pedestrians starts to running in fear, the overall S value increase significantly, the maximum of $|S-A|$ reaches 120, which is much larger than T . In this case the threshold value can be set larger to decrease the FP rate.

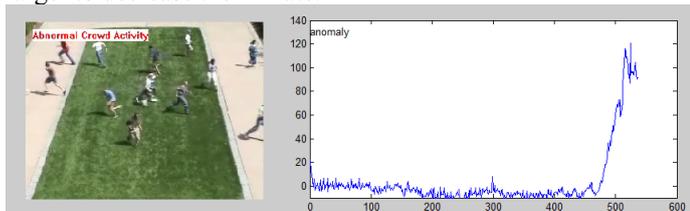


Fig. 7. Detection results using the proposed framework

IV. EXPERIMENT ENVIRONMENT AND RESULT

The experimental requirements for the research are relatively simple: a PC computer with any main stream operating system and MATLAB installed.

UMN [13] dataset is used for experiment. The UMN is a publicly available dataset maintained by researchers in the University of Minnesota. It includes 11 independent videos, and each video consists of normal crowd behaviors and

sequences of the panic behavior. UMN data set is widely used in lab-based experiments (the criticism is it is based on simplified/idealized setting against the often much more challenging actual field trial scenarios).

The algorithm introduced in previous section is applied on all 11 video clips of UMN data set, the result is shown as Figure 8. Blue line represents detected result, red line represents ground truth. Value of line reaches 0 means the current state is normal, when it reaches 1 means the panic happens or detected. It can be observed that panic behaviors in all video clips are successfully detected, however a common pattern of the results is that the detected abnormality vanished very fast and didn't match ground truth, two possible reasons are responsible for this issue. The first reason is the threshold T is set to static value for all the videos, thus when most of the panic people moves out of scene, the overall magnitude decrease fast and soon become smaller than the threshold. To solve this problem, a self-adjusting threshold mechanism can be introduced to set proper threshold value. The second reason is that the ground truth is manually labeled, which may be not accurate enough.

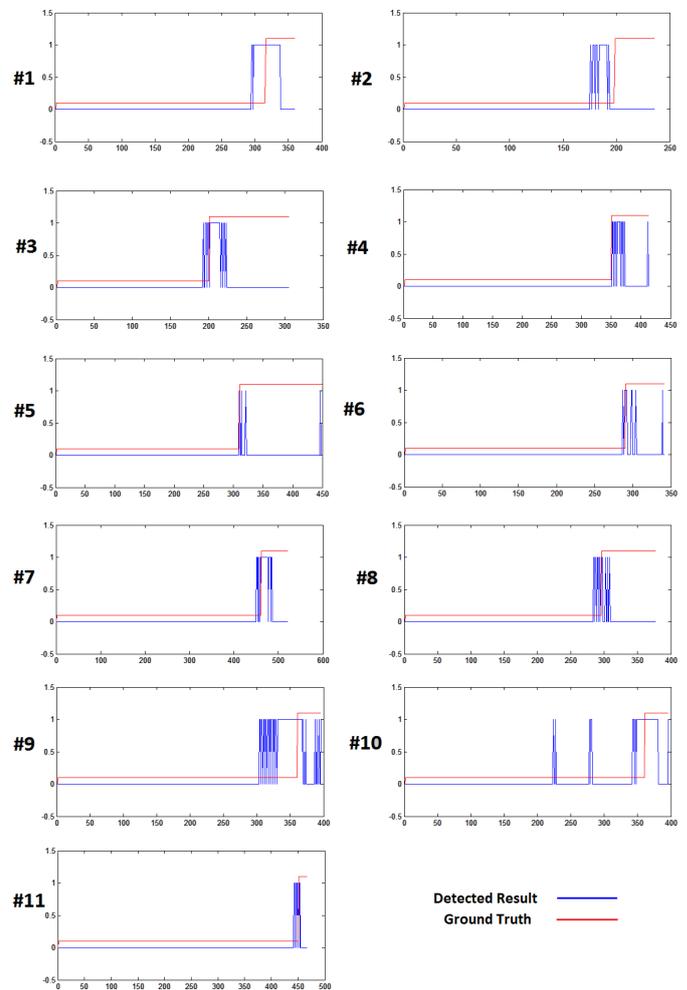


Fig. 8. Detection result of the proposed approach

REFERENCES

V. CONCLUSION AND FUTURE WORK

In this paper, we focused on the crowd behaviors detection algorithms. The paper reviewed many relative researches in this field and proposed an approach to detect panic abnormal behavior in crowd with relatively high density. The proposed approach successfully detected all panic behaviors in UMN dataset, however there are still issues need to be further worked on.

In the next phase of this research, further studies of global scale crowd abnormal behaviors will be carried out, their patterns explored, and categorized in the perspective of computer vision applications similar to Solmaz [1] approach. The crowd scene's dominant flows can be further analyzed in order to adapt more specific situations.

Utilizing global scale flow based feature patterns to build an innovative descriptor is one of the main goals for this research. In previous section, despite a panic crowd abnormality can be successfully detected, the nature of the abnormality is still vague and needs further analysis and appropriate interpretations. As in Krausz's work [10], a symmetry value is proposed to represent the congestion situation of the crowd, exploring a more innovative and convincing descriptor would be another strengthen of our detecting system.

For scenes with high crowd density, because of the computational complexity and the occlusion issues, especially the difficulty in tracking, segmentation of individuals are proven challenging. Thus other techniques, for example, to split image into smaller regions then investigate their probability and likelihood of containing interested features need to be explored. Furthermore, if the occlusion issue need to be tackled to enable effective tracking on individual basis.

- [1] B. Solmaz, B. E. Moore, and M. Shah, "Identifying behaviors in crowd scenes using stability analysis for dynamical systems", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2064-2070, Oct. 2012
- [2] Teng Li, "Crowded scene analysis: a survey". *IEEE transactions on circuits and systems for video technology*, vol. 25, no. 3, march 2015
- [3] Ramin Mehran, B. E. Moore, and M. Shah, "A streakline representation of flow in crowded scenes", in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 439-452.
- [4] Hsin-Chun Tsai, "The optical flow-based analysis of human behavior-specific system". *Orange Technologies (ICOT)*, 2013 International Conference on Year: 2013
- [5] Ramin Mehran. "Abnormal crowd behavior detection using social force model". 2009 IEEE
- [6] E.H Adelson, J.R. Bergen, "Spatio-temporal energy models for the perception of motion", *Journal of the Optical Society of America A* 2(1985) 284-299.
- [7] Wang J and Xu Z, "STV-based video feature processing for action recognition" *[J] // Signal Processing*, 2012, 93(8): 2151-2168.
- [8] A. B. Chan and N. Vasconcelos. "Modeling, clustering, and segmenting video with mixtures of dynamic textures". *PAMI*, 30(5):909-926, May 2008.
- [9] Vijay Mahadevan, "Anomaly detection in crowded scenes", 2010 IEEE.
- [10] Barbara Krausz and Christian Bauckhage. "Loveparade 2010: automatic video analysis of a crowd disaster". In *CVIU* 116(2012) 307-319
- [11] B. Solmaz, B. E. Moore, and M. Shah, "Identifying behaviors in crowd scenes using stability analysis for dynamical systems", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2064-2070, Oct. 2012.
- [12] A. Y. N. David M. Blei and M. I. Jordan. "Latent dirichlet allocation". *Journal of Machine Learning Research*, 3:993- 1022, 2003.
- [13] UMN Crowd Dataset. [Online]. Available: http://mha.cs.umn.edu/proj_events.shtml#crowd
- [14] Seung-Hwan Bae, and Kuk-Jin Yoon, "Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning", 2014 IEEE Conference on Computer Vision and Pattern Recognition. Pages: 1218 - 1225