Extracting Spatio-temporal Texture Signatures for Crowd Abnormality Detection

Yu Hao[1,2], Jing Wang[2], Ying Liu[1], Zhijie Xu[2], Jiulun Fan[1]
[1]School of Communication and Information Engineering, Xi'an University of Posts & Telecommunication, Xi'an, China
[1]y.hao@hud.ac.uk, [1]ly_yolanda@sina.com, [1]jiulunf@xupt.edu.cn
[2]School of Computing and Engineering, University of Huddersfield, Huddersfield, UK
{[2]j.wang2, [2]z.xu}@hud.ac.uk

***Abstract*—In order to achieve automatic prediction and warning of hazardous crowd behaviors, a Spatio-Temporal Volume (STV) analysis method is proposed in this research to detect crowd abnormality recorded in CCTV streams. The method starts from building STV models using video data. STV slices – called Spatio-Temporal Textures (STT) - can then be analyzed to detect crowded regions. After calculating the Gray Level Co-occurrence Matrix (GLCM) among those regions, abnormal crowd behavior can be identified, including panic behaviors and other behavioral patterns. In this research, the proposed STT signatures have been defined and experimented on benchmarking video databases. The proposed algorithm has shown a promising accuracy and efficiency for detecting crowd-based abnormal behaviors. It has been proved that the STT signatures are suitable descriptors for detecting certain crowd events, which provide an encouraging direction for real-time surveillance and video retrieval applications.**

Keywords*—Crowd abnormality, Spatio-Temporal Volume, STT Signature

I. BACKGROUND

The data obtained from CCTV cameras are 2D image sequences called frames. Each frame of these video data contains spatial information of various visual patterns. However, a single frame doesn't contain useful temporal information such as motions and trajectories of those patterns. In order to explore the temporal features, two or more consecutive frames need to be analyzed together. Horn and Schunck [1] first introduced Optical Flow Field to describe instant motion between two consecutive frames, fellow researchers [2] have conducted experiments using optical flow features to realize crowd abnormality detection. However, the computational complexity of optical flow is high, which limits its real-time performance. Other approaches such as motion field have been used to extract spatio-temporal information, yet suffering from object tracking problems caused by occlusion [3]. In this research, low-level and 2D texture features are utilized to avoid excessive computation because they only require pixel-wise calculation. This approach has ensured the computational time is independent from the actual crowds' density and event types. Following sections will explain how to reconstruct and utilize texture features for abnormal crowd behavior detection in details.

II. SPATIO-TEMPORAL VOLUME CONSTRUCTION

Spatio-Temporal Volume (STV) is first proposed by Aldelson and Bergen [4]. Figure 1 illustrates the STV construction process. In Figure 1(a), by stacking selected video frames in time sequence, a three dimensional RGB data block is obtained as shown in Figure 1(b).

To further process obtained STV models, slices of a STV model called Spatio-Temporal Textures (STT) can be extracted to learn patterns from the texture such as human or vehicle movement. For example Niyogi [5] has used STT to analyze the gait of individual pedestrian. In Figure 1(c) STV is cut either horizontally or vertically at certain position along time axis, to obtain STTs, and Figure 1(d) shows an example of extracted STTs describing pedestrians' motion through time. In this paper, various STT patterns - STT signatures - are studied to enabling the differentiation of ordinary crowd patterns from the abnormal ones.

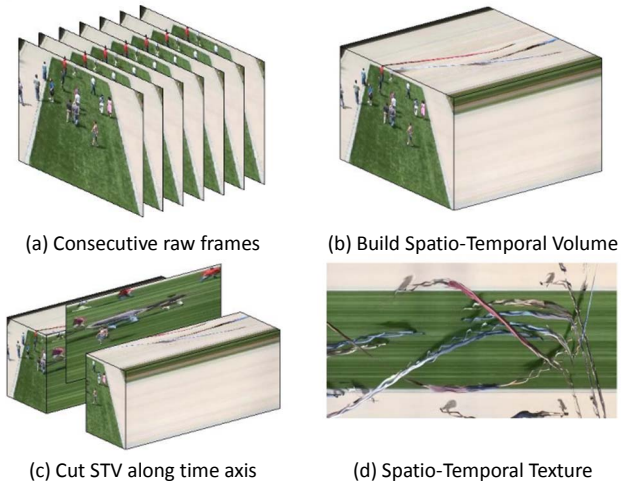


Fig. 1. Procedures to obtain STV and STT from raw video data

III. STT Analysis and Signature definition

A. STT Feature Classification

Depending on different constructing patterns, STT features can be roughly classified into Statistical Texture Features, Model Type Texture Features and Signal Domain Texture Features.

1) Statistical Texture Features

This type of feature is obtained by collecting gray scale value related between target pixel and neighbor pixel, and then calculating first-, second- and higher- degree statistical features such as contrast, variance etc. The most frequently used statistical texture features is Grey Level Co-occurrence Matrix (GLCM) [6], which will be probed in next section.

2) Model Type Texture Features

This type assumes texture can be described by certain parameter controlled distribution model. How to find the most accurate parameter value is the core issue of this approach. Benezeth [7] proposed an algorithm using Hidden Markov Model (HMM) associated with a Spatio-Temporal neighborhood co-occurrence matrix to describe the texture feature.

3) Signal Domain Texture Features

In this approach, textures are defined in a transform domain by certain transformation or filters such as wavelet [8]. It is based on the assumption that the energy distribution of frequency domain can be used to classify textures.

In this research, GLCM feature has been explored to test its performance in STT signature identification. Since GLCM belongs to statistical domain, sophisticated algorithms for detecting semantically high level information can be avoided to save computation time. In the meantime, detailed information such as individual movement isn't necessary, because the focus of this research is the detection of abnormal crowd behavior. Therefore, the main strategy of this approach is to extract raw GLCM texture feature from relevant STTs. Once acquired these features, histogram distributions along time will be observed to study crowd patterns. A five-stage process flow of this approach is shown in Figure 2.

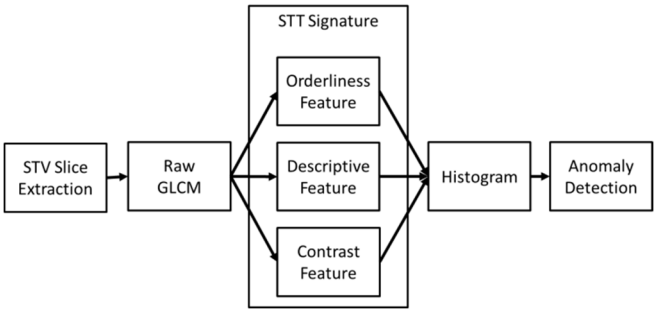


Fig.2. Structure of proposed approach

The Grey Level Co-occurrence Matrix (GLCM), known as Grey Tone Spatial Dependency Matrix, is first proposed by Robert M Haralick [6]. By definition the GLCM is a statistic tabulation of the probability of different pixel grey scale values occurs in an image. In brief, assuming the gray scale of current image is divided to three levels, GLCM will store the number of neighboring pairs of these three levels. A sample GLCM is shown in Figure 3.

	1	2	3
1	3	0	2
2	0	4	0
3	3	5	1

Fig.3. A sample GLCM, gray scale level is set to 3.

Details of the GLCM algorithm can be found in [9]. In most cases, STTs are irregular, thus the obtained results of G are very likely to be asymmetric. According to the GLCM definition, G represents the gray-scale pair relations along one direction, the transposed matrix is then calculated to represent the relation matrix along opposite direction, and then the symmetric matrix S is obtained by adding G' and G, to represent the complete relations a direction. Next step is the normalization, the probability matrix P is obtained from S by using the following equation (1).

$$P_{i,j} = \frac{S_{i,j}}{\sum_{i,j=0}^{N-1} S_{i,j}} \tag{1}$$

Next, texture features can be calculated from the probability matrix P. The resulting low level texture patterns are named here as STT signatures namely contrast signatures, orderliness signatures, and descriptive statistical signatures.

B. Signature Definitions

Contrast signatures describe how drastically the gray scale value of current image changes in terms of contrast, dissimilarity, homogeneity and similarity.

The farther pixel pairs from diagonal in P represents higher difference in gray scale values (contrast), thus the contrast can be obtained by (2).

$$CON = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2 \tag{2}$$

Similar to contrast, dissimilarity also represents difference in gray scale values, except it increases linearly instead of exponentially, dissimilarity can be obtained by (3).

$$DIS = \sum_{i,j=0}^{N-1} P_{i,j}|i-j| \tag{3}$$

Homogeneity is also called Inverse Different Moment (IDM), on the contrary, homogeneity represents how less the contrast is, when the contrast of image is low, value of homogeneity could be large. Equation (4) shows how to get homogeneity.

$$HOM = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \tag{4}$$

Similar to dissimilarity, linear version of homogeneity can be obtained by using (5).

$$SIM = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+|i-j|} \quad (5)$$

The Table 1 gives a comparison of contrast related patterns for different image. GLCM calculation window size is set to 50 by 50 pixels large, GLCM direction is set to horizontal, calculation step is fixed to 1 pixel, and number of gray scale level is set to 8. Figure in Table 1.(a) is less contrastive than image Table 1.(d), thus the result shows that image Table 1.(a) has less GLCM contrast and dissimilarity values, and larger homogeneity and similarity values.

Orderliness related patterns describe how orderly or regular of gray scale value distribution, including Angular Second Moment, Energy and Entropy. Angular Second Moment (ASM) comes from physics [11], used to measure rotational acceleration. ASM could be obtained using (6), its value increases while the orderliness of gray scale value distribution is high.

$$ASM = \sum_{i,j=0}^{N-1} {P_{i,j}}^2 \quad (6)$$

The Energy equals to the square root of ASM, as (7).

$$ENR = \sqrt{ASM_{i,j}} \quad (7)$$

On contrary to energy, entropy describes how irregular current gray scale distribution is, value of entropy decreases when distribution is less orderly. Entropy can be expressed as (8).

$$ENT = \sum_{i,j=0}^{N-1} P_{i,j}(-lnP_{i,j}) \quad (8)$$

In Table 1 the orderliness signatures are also compared for six different images. Image Table 1.(a) clearly shows more regular patterns than Image Table 1.(d), so it can be expected that the entropy of Image Table 1.(a) is less than Image Table 1. (d).

Descriptive Statistics related patterns consist of statistics derived from GLC matrix, including GLCM Mean, GLCM Variance and GLCM Correlation. It needs to be emphasized that these patterns describes the statistic of pixel pair gray scale relation, but not typical gray scale value. Two GLCM Mean values can be obtained by using (9), note that the probability matrix P is symmetric, two mean values are identical.

$$\mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j}) \qquad \mu_j = \sum_{i,j=0}^{N-1} j(P_{i,j}) \quad (9)$$

GLCM Variance and Standard Deviation can be obtained through (10) and (11).

$$\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j}(i-\mu_i)^2 \qquad \sigma_j^2 = \sum_{i,j=0}^{N-1} P_{i,j}(j-\mu_j)^2 \quad (10)$$

$$\sigma_i = \sqrt{\sigma_i^2} \qquad \sigma_j = \sqrt{\sigma_j^2} \quad (11)$$

Finally according to the calculated GLCM Mean and GLCM Variance, the GLCM Correlation is obtained by using (12).

$$COR = \sum_{i,j=0}^{N-1} P_{i,j}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right] \quad (12)$$

In Table 1, signatures from six STTs are listed. These clips are extracted at different parts of the STV model, images (a-c) are from normal part, and images (d-f) are from abnormal part. By comparing pattern values of normal and abnormal clips, following patterns can be identified. Firstly image clip at normal state usually has lower contrast, entropy and variance. Images (a-c) all have lower contrast than images (d-f). Image (b) has higher contrast than Image (a) and (c), it's because most contrast is generated by the lawn. Secondly image clip at normal state usually has higher ASM value than clips at abnormal state. Thirdly, among all other patterns, Contrast, ASM, Entropy and Variance show most significant changes between normal and abnormal states. Thus these four patterns are assumed to be most appropriate for abnormal detection, and are labeled in Table 1.

IV. Abnormal Detection using STT signatures

The gray scale image transformed from the STT in Figure 1(d) is displayed in Figure 4. The test video is from UMN dataset [10]. All videos from this dataset contain a normal crowd scene following with an abnormal event, mostly panic behavior. The ground truth of normal and abnormal behaviors is manually marked on Figure 4, by using a color bar at the bottom of the figure. The brighter bar indicates normal state and the darker bar indicates abnormal state. It can be observed that different visual patterns of this figure match the labeled ground truth. Also same color bars are used for labeling ground truth in following figures. It is expected that the differences of patterns will reflect on the defined STT signatures. According to the definition of STT, the column index represents frame index in original video, thus by summing up each column calculated by GLCM texture features, the change of GLCM feature patterns over time can be observed. As stated in following Section, Contrast, ASM, Entropy and Variance are used for performance evaluation.

TABLE I. COMPARISON BETWEEN TEXTURE PATTERNS OF SPATIO-TEMPORAL TEXTURES

	 (a)	 (b)	 (c)	 (d)	 (e)	 (f)
CON	0.2437	0.3237	0.2669	0.6853	0.5735	0.6473
DIS	0.1922	0.2110	0.1935	0.3947	0.3645	0.4278
HOM	0.9085	0.9049	0.9103	0.8302	0.8379	0.8078
SIM	0.9101	0.9094	0.9132	0.8405	0.8459	0.8174
ASM	0.3538	0.2134	0.4124	0.1853	0.2062	0.1767
ENR	0.5948	0.4619	0.6422	0.4304	0.4541	0.4203
ENT	1.2977	2.0858	1.5294	2.3325	2.1747	2.3599
MEAN	2.4933	4.7598	2.7865	4.0635	2.6410	2.8265
VAR	0.3859	3.7115	0.6728	2.5150	1.0509	2.8265
SDEV	0.6212	1.9265	0.8202	1.5859	1.0251	1.0729
COR	0.6843	0.9564	0.8016	0.8638	0.7272	0.7188
Normal	Yes	Yes	Yes	No	No	No

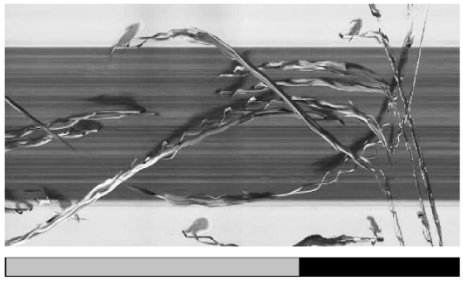


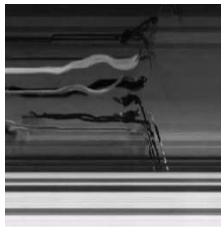
Fig. 4. Normality and Abnormality in STT

Figure 5 exhibits a result comparison between three different videos. All three videos are obtained from UMN dataset. The videos are recorded in different indoor and outdoor settings, which include a lawn, a hallway and a plaza. All of the videos contain a normal state following with a crowd panic event. The Figure 5(a)(c)(e) gives snapshots of each video. The Figure 5(b)(d)(f) shows the extracted STTs of each video, by careful positioning, all panic video slices shows similar patterns. Next the Contrast, ASM, Entropy and Variance features are calculated separately on these videos to examine the performance. The reason of choosing these features is they show the most significant fluctuation in previously conducted experiment, and the most drastic fluctuating feature indicates the highest possibility for detection.

(a) Snapshot of UMN#1

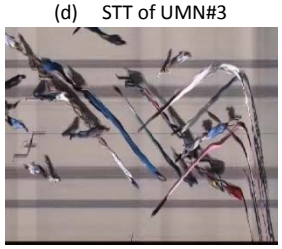
(b) STT of UMN#1

(c) Snapshot of UMN#3

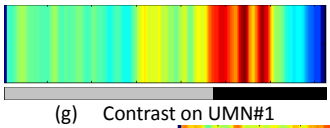


(d) STT of UMN#3

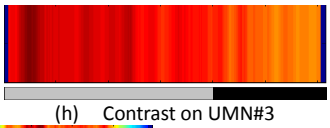
(e) Snapshot of UMN#10



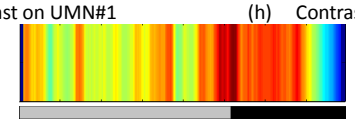
(f) STT of UMN#10



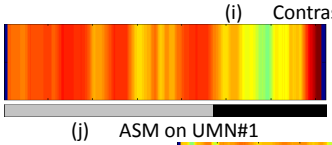
(g) Contrast on UMN#1



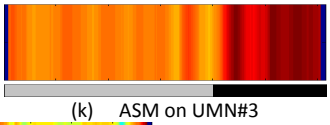
(h) Contrast on UMN#3



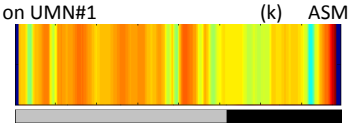
(i) Contrast on UMN#10



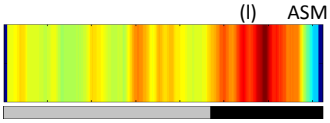
(j) ASM on UMN#1



(k) ASM on UMN#3



(l) ASM on UMN#10



(m) Entropy on UMN#1

(n) Entropy on UMN#3

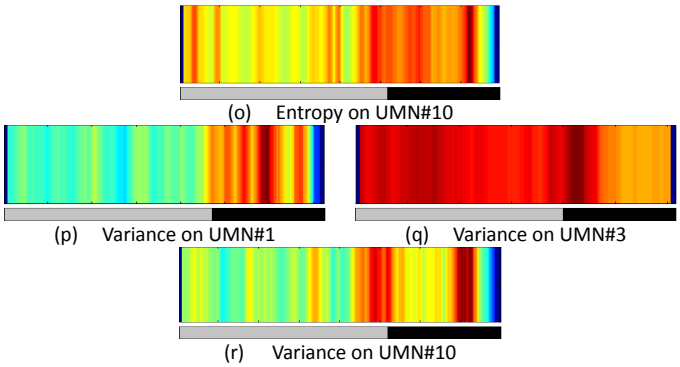


Fig. 5. Result comparison between multiple videos

By examining the results of Contrast, ASM, Entropy and Variance on three videos, it can be concluded that when the crowd abnormal behavior happens, magnitude of Contrast, Entropy and Variance have a significant surge, and on the contrary, the magnitude of ASM reduces. Under this observation, magnitude of Contrast, ASM, Entropy and Variance are combined as a salient descriptor for detecting crowd abnormal behavior.

The performance of proposed signature on the first video footage of UMN dataset is shown in figure 6. The approach is also compared with these based on Optical Flow [12] and Social Force Model [13]. The result shows that the detection accuracy of proposed approach is better than Optical Flow but worse than Social Force Model. However the processing speed of proposed approach is faster than both of the approaches. For example, it takes about 100 seconds to obtain optical flow patterns of UMN's first video, yet it takes only about 20 seconds to obtain proposed signature. This still makes STT GLCM signature a competitive approach for crowd abnormal behavior detection.

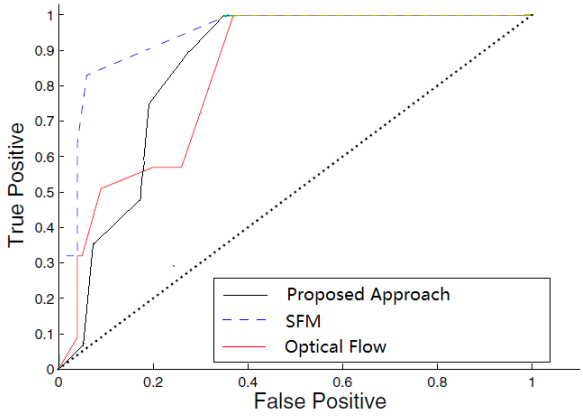


Fig.6. Results of abnormal detection on UMN Dataset

V. Conclusion and Future Works

In conclusion, GLCM texture features in STTs extracted from STV models are studied and utilized to detect abnormal crowd behaviors. Under all tested scenarios, the panic crowd abnormal behaviors are successfully detected. In a fully automated system, the selection of slice positions need to be decided by standardized models and operations. Thus how to extract STTs with most information in the shortest time will be investigated in the future.

Acknowledgement

This research is supported by Shenzhen science and Technology plan project (No.GJHZ20160301164521358), by Science and Technology Department of Shaanxi Province (No.2016GY-123).

References

[1] Horn, B. K. P, and Schunck, B. G., "Determining optical flow", Artificial Intelligence, vol. 17, pp 185-203, 1981

[2] Yu Hao, Zhijie Xu, and Jing Wang, "An approach to detect crowd panic behavior using flow-based feature", 2016 22nd International Conference on Automation and Computing (ICAC), pp 462-466, 2016, DOI: 10.1109 /IConAC.2016. 7604963

[3] Teng Li, Huan Chang, Meng Wang, Bingbing Ni, Richang Hong, and Shuicheng Yan. "Crowded scene analysis: a survey". IEEE Transactions on Circuits and Systems for Video Technology, v 25, n 3, p 367-386, March 1, 2015

[4] Aldelson, E., and Bergen, J.R., "Spatiotemporal energy models for the perception of motion", Journal Optical Society of America, vol. 2, 284-299, 1985

[5] Niyogi, S., and Adelson, E., "Analyzing and recognizing walking figures in XYT", Proceedings of IEEE International Conference of Computer Vision and Pattern Recognition, pp 469-474, 1994

[6] Robert M Haralick, K Shanmugam, and Its'hak Dinstein. "Textural features for image classification". IEEE Transactions on Systems, Man, and Cybernetics. SMC-3 (6): 610–621,1973

[7] Y. Benezeth, P.-M. Jodoin, V. Saligrama, C. and Rosenberger. "Abnormal events detection based on spatio-temporal co-occurences". 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp 2458 – 2465, 2009

[8] Shen Jianbing, Jin Xiaogang, Zhou Chuan, and Zhao Hanli. "Dynamic textures using wavelet analysis". Lecture Notes in Computer Science, v 3942 LNCS, p 1070-1073, 2006, Technologies for E-Learning and Digital Entertainment - First International Conference

[9] "The GLCM Tutorial", http://www.fp.ucalgary.ca/mhallbey/ tutorial.htm

[10] UMN dataset. http://www.cs.ucf.edu/~ramin/projects/

[11] Kress, Joel D. Voter, and Arthur F. "Model description of transition metals using the rotated second moment approximation". Radiation Effects and Defects in Solids, v 129, n 1-2, p 45-53, 1994

[12] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. Proc. CVPR, 2009. 3162, 3166

[13] S. Wu, B. Moore, and M. Shah. Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. Proc. CVPR, 2010. 3162, 3166