

Abnormal Behavior Detection in Crowded Scenes Based on Optical Flow Connected Components

Oscar E. Rojas^(✉) and Clesio Luis Tozzi

School of Electrical and Computer Engineering, UNICAMP, Av. Albert Einstein,
400, Campinas, SP, Brazil

oscar.rojas87@gmail.com, clesio@dca.fee.unicamp.br

Abstract. This paper presents a new approach for automatic abnormal behavior detection in crowded scenes. Background subtraction algorithm, optical flow and connected component analysis are used to define the optical flow connected components (OFCC). An unsupervised normal behavior model is computed using the main magnitude and direction of each OFCC. Experimental results on the standards UCSD and UMN anomaly detection and localization benchmarks demonstrate the method performance compared to other approaches considering detection rate and processing time.

Keywords: Anomaly behavior detection · Optical flow · Crowded scenes analysis · Anomaly localization

1 Introduction

Abnormal behavior analysis on crowded scenes is an important and growing research field. Video cameras, given their ease installation and low cost, have been widely used for monitoring internal and external areas such as buildings, parks, stadiums etc. With the world's population increasing, the presence of people in common areas has been increasing too. Algorithms for pose detection and action recognition for single or, in some cases, very low density groups of people are extensively treated in the pattern recognition community. Nevertheless, abnormal behavior detection and localization in crowded scenes remains an open problem due to high levels of occlusion and the impractical approach of individual segmentation.

The concept of abnormal behavior is always associated with the scene context, a behavior considered as normal in a scene may be considered abnormal in other. These specific conditions increase the difficulties for automatic analysis and require specific modeling of the abnormal behavior for each particular scene.

In order to build such models many algorithms have been proposed. In [6] optical flow is used to compute interaction forces between adjacent pixels and a model, known as Social Force Model, is created based in a bag of words approach

for classify frames either normal or abnormal. In [5] dynamic textures (DT) are used to model the appearance and dynamics of normal behavior, samples with a low probabilistic values in the model are labeled as abnormal. In [7] entropy and energetic concepts are used as features to model the probability of finding abnormal behavior in the scene. Natural language processing is used in [10] as classification algorithm for features based on viscous fluid field concepts.

Many algorithms employs machine learning techniques as classification tool. Support Vector Machine (SVM) are used in [8, 11] for classify histograms of the orientation of optical flow. Multilayer Perceptron Neural Networks is used in [13]. k-Nearest Neighbors is used in [1] for classify outlier observed trajectories as abnormal behavior. Finally, Fuzzy C-Means are used in [2, 3] to derive an unsupervised model for the crowd’s trajectory patters.

In general, to construct the feature vector used in many of the algorithms described above, a several set of parameters must be correctly set in order to achieve the performance reported by the authors. Some of the state-of-the-art methods are based in complex probabilistic models which leads high processing time. Despite the processing time per frame is reported only for a very few papers, it is in general high. For example, in [5] the authors reported a test time of 25 s per frame for 160×240 pixel images and in [12] the reported test time per frame is 5 s in videos with 320×240 pixel resolution.

The main contribution for this paper is present a simple but efficient method that reduce the processing time per frames in near real time allowing practical use.

The rest of this paper is organized as follows. Section 2 describes the proposed approach. Section 3 presents the experimental results. Section 4 presents the conclusions.

2 Proposed Method

The general pipeline for the proposed approach is shown in Fig. 1. The five initial modules (1 to 5 in Fig. 1) aims to compute the model features and are the same for the training and test phases. These initial modules are described in Sect. 2.1. In the training phase, represented by module 6, frames with normal

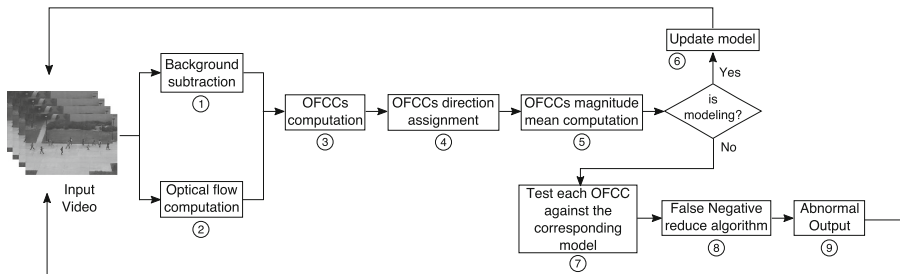


Fig. 1. General pipeline for the proposed method.

behavior are used to update the model as described in Sect. 2.2. In the test phase, represented by module 7, each new sample is compared with the model and classified as normal or abnormal as described in Sect. 2.3. A false positive reduction methodology, represented by module 8, is also described in Sect. 2.3.

2.1 OFCCs Computation

In the training phase a sequences of N frames are used to build a normal behavior model. The algorithm presented in [9] is used to compute the background model and a foreground mask I_{fm} of each frame is obtained.

In order to reduce the noise and de computational load a connected components labeling algorithm is used to obtain the blobs (b_1, b_2, \dots, b_n) where n is the total number of blobs in the foreground mask I_{fm} . In parallel to foreground extraction the dense optical flow of each frame is computed using [14]. The optical flow vectors are used to obtain the magnitude $m(x, y)$ and direction $\theta(x, y)$ values of each (x, y) point in the input image.

An Optical Flow Connected Component $OFCC_i$ can be defined as the set of values $[m(x, y), \theta(x, y)]$ for all (x, y) points belonging to the i -th blob, as expressed in Eq. 1.

$$OFCC_i = [m(x, y), \theta(x, y)] \forall (x, y) \in b_i. \quad (1)$$

The main direction $\bar{\theta}_i$ of the i -th OFCC is computed as follows. A histogram of the direction values of $OFCC_i$ is obtained with a fixed bin width of $\Delta\theta = 45^\circ$. The angle associated with the highest bin is used as the main direction $\bar{\theta}_i$ of $OFCC_i$.

The main magnitude \bar{m}_i of $OFCC_i$ is obtained as the statistical mean of the magnitudes values in $OFCC_i$ as shown in Eq. 2.

$$\bar{m}_i = \frac{1}{S} \sum_{k=1}^S m(x, y) \mid m(x, y) \in OFCC_i \quad (2)$$

where S is the total number of magnitude values in $OFCC_i$. Finally, the main direction $\bar{\theta}_i$ and the main magnitude \bar{m}_i values of each $OFCC_i$ are used to construct the normal behavior model.

2.2 Normal Behavior Model

In this algorithm the behavioral model is composed of m matrices (A_1, A_2, \dots, A_m) where m is computed as

$$m = \frac{360}{\Delta\theta} \quad (3)$$

and represents the number of possible values that $\bar{\theta}_i$ can adopt. For instance, if $\Delta\theta = 45^\circ$ then $m = 8$ matrices will be defined.

For each frame in the training video a set of n OFCCs are obtained as described in the previous section. After compute the $\bar{\theta}_i$ and \bar{m}_i values of each $OFCC_i$ the corresponding A matrix number η is obtained as,

$$\eta = \frac{\bar{\theta}_i}{\Delta\theta} \tag{4}$$

Then, the values of the A_η matrix are updated using the next condition

$$A_\eta(x, y) = \begin{cases} \bar{m}_i, & \text{if } \bar{m}_i > A_\eta(x, y), \forall (x, y) \in b_i. \\ A_\eta(x, y), & \text{otherwise} \end{cases} \tag{5}$$

At the end of the training phase each matrix A_η will store the maximum principal magnitude \bar{m}_i in the full training video at each point (x, y) at the direction $\eta * \Delta\theta$.

Figure 2 shows an example of a normal behavior model with $\eta = (1, 2, \dots, 8)$ matrices for $\Delta\theta = 45^\circ$. A color map was applied to each matrix A_η for better visualization.

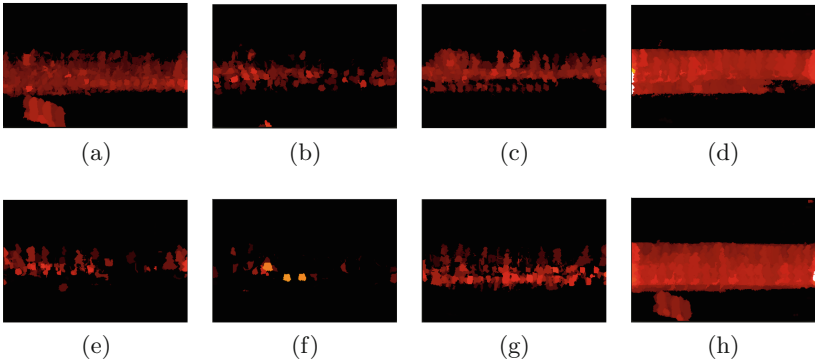


Fig. 2. The magnitude model for eight directions ($\eta = 8$). Each image shows the pixel highest magnitude for direction angles between (a) $(0^\circ-45^\circ]$, (b) $(45^\circ-90^\circ]$, (c) $(90^\circ-135^\circ]$, (d) $(135^\circ-180^\circ]$, (e) $(180^\circ-225^\circ]$, (f) $(225^\circ-270^\circ]$, (g) $(270^\circ-315^\circ]$ and (h) $(315^\circ-360^\circ]$. (Color figure online)

2.3 Abnormality Detection

After all the training frames have been processed and the model is completed, test videos with both normal and abnormal behaviors can be analyzed.

The set of OFCCs and their main directions $\bar{\theta}_i$ s are obtained as described in Sect. 2.1 for each video frame. To determine if the $OFCC_i$ is abnormal or not its main direction $\bar{\theta}_i$ is used to find the corresponding A_η matrix with η computed using Eq. 4. Next the maximum value \hat{a}_η in A_η within the same region defined by the blob b_i is founded according to

$$\hat{a}_\eta = \max(A_\eta(x, y) \mid (x, y) \in b_i). \quad (6)$$

Then, the comparison between each $m(x, y)$ value in $OFCC_i$ and \hat{a}_η is done as follows. If $m(x, y)$ is greater than \hat{a}_η then the pixel (x, y) is marked as abnormal, otherwise its marked as normal.

After compare all the magnitude values in $OFCC_i$ an abnormal binary mask image $I_{ab}(x, y)$ with the same size as the input frames, can be use to store the abnormal marked pixels as $I_{ab}(x, y) = 1$ and the normal ones as $I_{ab}(x, y) = 0$.

In order to improve the algorithm performance a FIFO type list with fixed size M is defined and filled up with the latest M binary images $I_{ab}(x, y)$. To consider an OFCC as abnormal it must appear at least a number W of times in the list. The list size M and the number W are user controlled parameters and can be used for sensitivity adjustment, since a higher value of W means a higher alarm delay time.

3 Results and Comparisons

The proposed algorithm was implemented in Qt/C++ using OpenCV on a 2.7 GHz Intel Core i7 PC with 16 GB of RAM. The method was tested in two popular datasets: UMN¹ and UCSD². Figure 3 shows a frame for each of the scenarios in the UMN dataset and the abnormality detected by the proposed approach. The frame size in all UMN videos is 320×240 pixels. The frame size in the UCSDPed1 videos is 238×158 and for UCSDPed2 is 360×240 pixels.



Fig. 3. Examples of normal (*top*) and abnormal (*bottom*) situations in UMN dataset.

Figure 4 shows three examples frames with abnormal behavior for each of the two scenarios in the UCSD dataset.

The proposed method was compared with similar state-of-the-art algorithms including Mixture Dynamic Texture (MDT) [5], Mixture of Optical Flow

¹ http://mha.cs.umn.edu/proj_events.shtml.

² <http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm>.

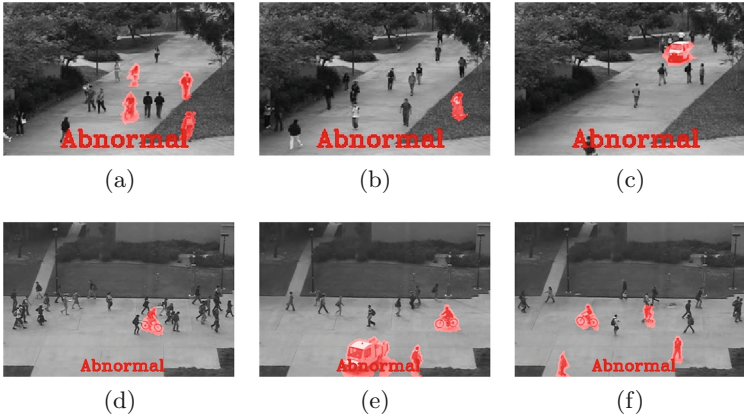


Fig. 4. Example of abnormal behavior detected in the UCSD dataset. UCSDped1 (*top*) and UCSDped2 (*below*).

(MPPCA) [4], Social Force [6], Social Force with MPPCA [4] and the Hierarchical Activity Approach [12]. Figure 5 shows the Receiver Operation Characteristic (ROC) curves for the proposed method and the comparative algorithms, taken from [12]. Table 1 shows the Area Under the ROC curve (AUC) for the five comparative methods and the proposed one. Finally, Fig. 6 shows the processing time per frame for some state-of-the-art algorithms and the proposed in this paper.

The ground truth provided by the UCSD dataset, and used for performance evaluation in all the comparison methods, labels people in wheelchair as abnormal behavior, even when their speed is lower than the speed of walking people. This leads to additional False Negative detected frames because, in the presented algorithm, this situation is not considered as abnormal. A second situation when the output of the presented algorithm differs from the ground truth is when

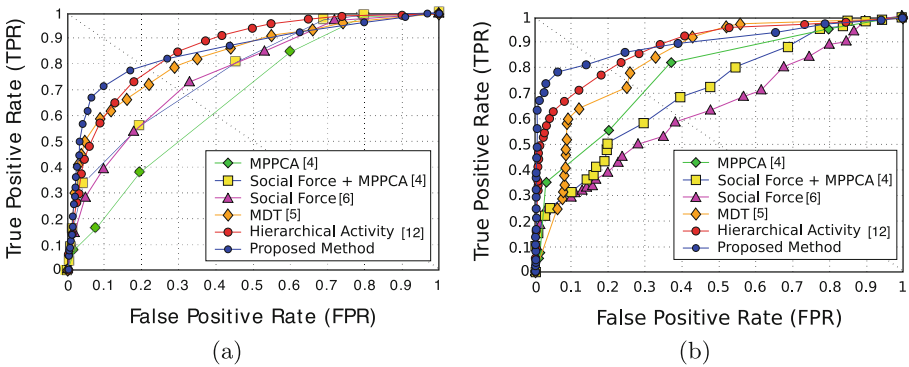


Fig. 5. Quantitative comparison of abnormal behavior detection in (a) UCSDped1 and (b) UCSDped2 against state-of-the-art algorithms.

somebody, in the test phase, walks in a region where no people walked in the training phase. Examples of this type of abnormal detection are shown in Fig. 4(a), (b), (e) and (f). Frames that present only this kind of abnormality are ignored in the comparison results.

Table 1. Comparison of the Area Under Curve of the proposed method compared with the others algorithms.

Algorithm	Area Under Curve (AUC)	
	UCSDped1	UCSDped2
MPPCA [4]	0.59	0.693
Social force [6]	0.675	0.556
Social force + MPPCA [4]	0.668	0.613
MDT [5]	0.818	0.829
Hierarchical activity [12]	0.854	0.882
Proposed method	0.852	0.958

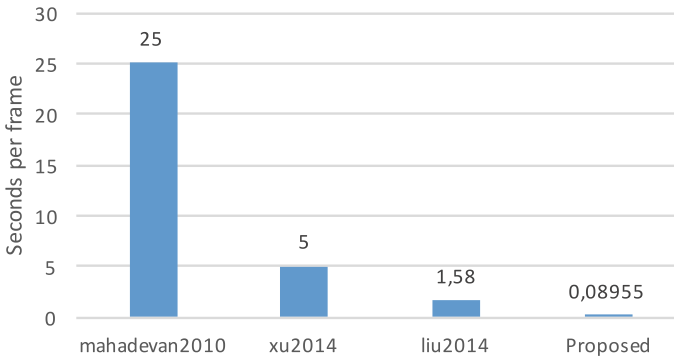


Fig. 6. Comparison of consumed time per frame with others state-of-the-art algorithms. The showed time is for the test phase in UCSDped1.

4 Conclusions

This paper presents a new method for abnormal behavior detection. It's based on optical flow and connected component analysis. From the experimental results it can be concluded that, when compared to other state of the art methods, the proposed method presents better performance in abnormal detection in the UCSDped2 dataset and is very close to the best one in the UCSDped1 but, as shown in Fig. 6 it presents the lowest processing time, near to real-time processing which allows practical use in modern computers.

Acknowledgments. The authors wish to thank Conselho Nacional de Desenvolvimento Científico (CNPq), Brazilian Research Support Foundations, for sponsoring this work.

References

1. Alvar, M., Torsello, A., Sanchez-Miralles, A., Armingol, J.M.: Abnormal behavior detection using dominant sets. *Mach. Vis. Appl.* **25**(5), 1351–1368 (2014)
2. Chen, Z., Tian, Y., Zeng, W., Huang, T.: Detecting abnormal behaviors in surveillance videos based on fuzzy clustering and multiple Auto-Encoders. In: *IEEE International Conference on Multimedia and Expo*, pp. 1–6 (2015)
3. Cui, J., Liu, W., Xing, W.: Crowd behaviors analysis and abnormal detection based on surveillance data. *J. Vis. Lang. Comput.* **25**, 628–636 (2014)
4. Kim, J., Grauman, K.: Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates. In: *Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR)* (2009)
5. Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N.: Anomaly detection in crowded scenes. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1975–1981 (2010)
6. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 935–942 (2009)
7. Ren, W.Y., Li, G.H., Chen, J., Liang, H.Z.: Abnormal crowd behavior detection using behavior entropy model. In: *International Conference on Wavelet Analysis and Pattern Recognition*, pp. 212–221 (2012)
8. Snoussi, H., Wang, T.: Detection of abnormal visual events via global optical flow orientation histogram. *IEEE Trans. Inf. Forensics Secur.* **9**(6), 988–998 (2014)
9. Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. In: *Proceedings in IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2 (1999)
10. Su, H., Yang, H., Zheng, S., Fan, Y., Wei, S.: Crowd event perception based on spatio-temporal viscous fluid field. In: *IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance*, pp. 458–463 (2012)
11. Wang, T., Snoussi, H.: Histograms of optical flow orientation for visual abnormal events detection. In: *IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, pp. 13–18 (2012)
12. Xu, D., Song, R., Wu, X., Li, N., Feng, W., Qian, H.: Video anomaly detection based on a hierarchical activity discovery within spatio-temporal contexts. *Neurocomputing* **143**, 144–152 (2014)
13. Zhang, D., Peng, H., Haibin, Y., Lu, Y.: Crowd abnormal behavior detection based on machine learning. *Inf. Technol. J.* **12**, 1199–1205 (2013)
14. Zivkovic, Z., van der Heijden, F.: Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recogn. Lett.* **27**(7), 773–780 (2006)