OBJECT-ORIENTED ANOMALY DETECTION IN SURVEILLANCE VIDEOS

Xiaodan Li, Weihai Li, Bin Liu, Qiankun Liu, Nenghai Yu
School of Information Science and Technology, University of Science and Technology of China
Key Laboratory of Electromagnetic Space Information, the Chinese Academy of Sciences
lxd1030@mail.ustc.edu.cn; whli@ustc.edu.cn; flowice@ustc.edu.cn; liuqk3@mail.ustc.edu.cn; ynh@ustc.edu.cn

1. Motivation

Most existing methods are patch or trajectory-based, which either lack semantic understanding of scenes or didn't perform well in crowded scenes. So we proposes a novel and effective algorithm by incorporating object detection and tracking with full utilization of spatial and temporal information.

2. Dynamic Image Object Detection

$$\begin{cases} H = Angle\ of\ optical\ flow \\ S = Magnitude\ of\ optical\ flow \\ I = Image\ intensity \end{cases}$$

3. Detect Anomaly

Appearance: As a by-product of object detection, we can obtain the object categories and the corresponding confidence scores. We check whether each detected object belongs to normal classes. If not and the confidence score is beyond 0.9, it will be regarded as anomaly.

Location: To deal with location anomaly like walking on the grass, a background model is firstly established with principle component analysis which considers background and foreground as a low-rank matrix and a sparse error matrix respectively.

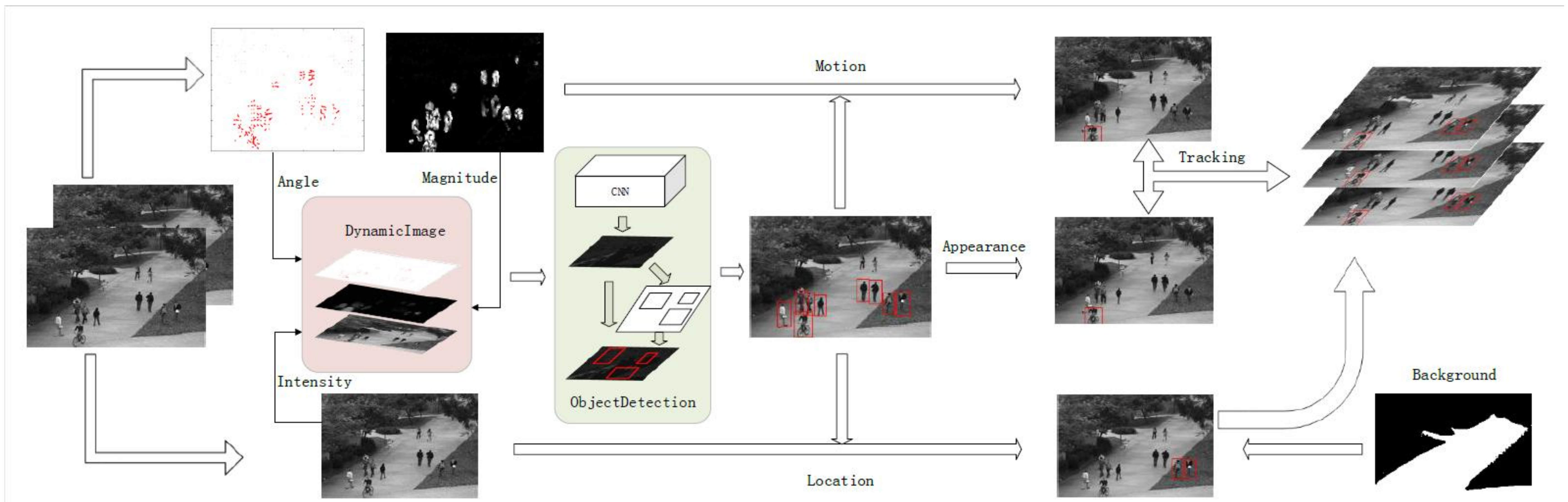


Fig. 1. Framework of the proposed algorithm.

Motion: We use motion energy E_m to reflect the speed of object motion to detect velocity anomalies.

$$E_m = \sum_{i=1}^{N} v_i^2 / N$$

where N is the number of pixels of an object, v_i is the magnitude of optical flow at each pixel.

Besides, a new effective and scale-insensitive feature called histogram variance of optical flow angle (**HVOFA**) is also developed to detect rigid objects like cars in the crowds, which counts the frequency of different directions just by angles of optical flow.

$$HOFA = [f_1, f_2, \dots, f_B], \qquad \sum_{i=1}^{B} f_i = N$$

$$HVOFA = \sum_{i=1}^{B} (f_i - \bar{f})^2 \le \left(\sum_{i=1}^{B} f_i\right)^2 - N^2/B$$

where B is the number of directions in HOFA, f_i is the number of pixels of a direction. $\bar{f} = N/B$.

Only when all pixels follow the same direction, the inequality gets to be equal. The higher is the HVOFA, the more rigid is the object.

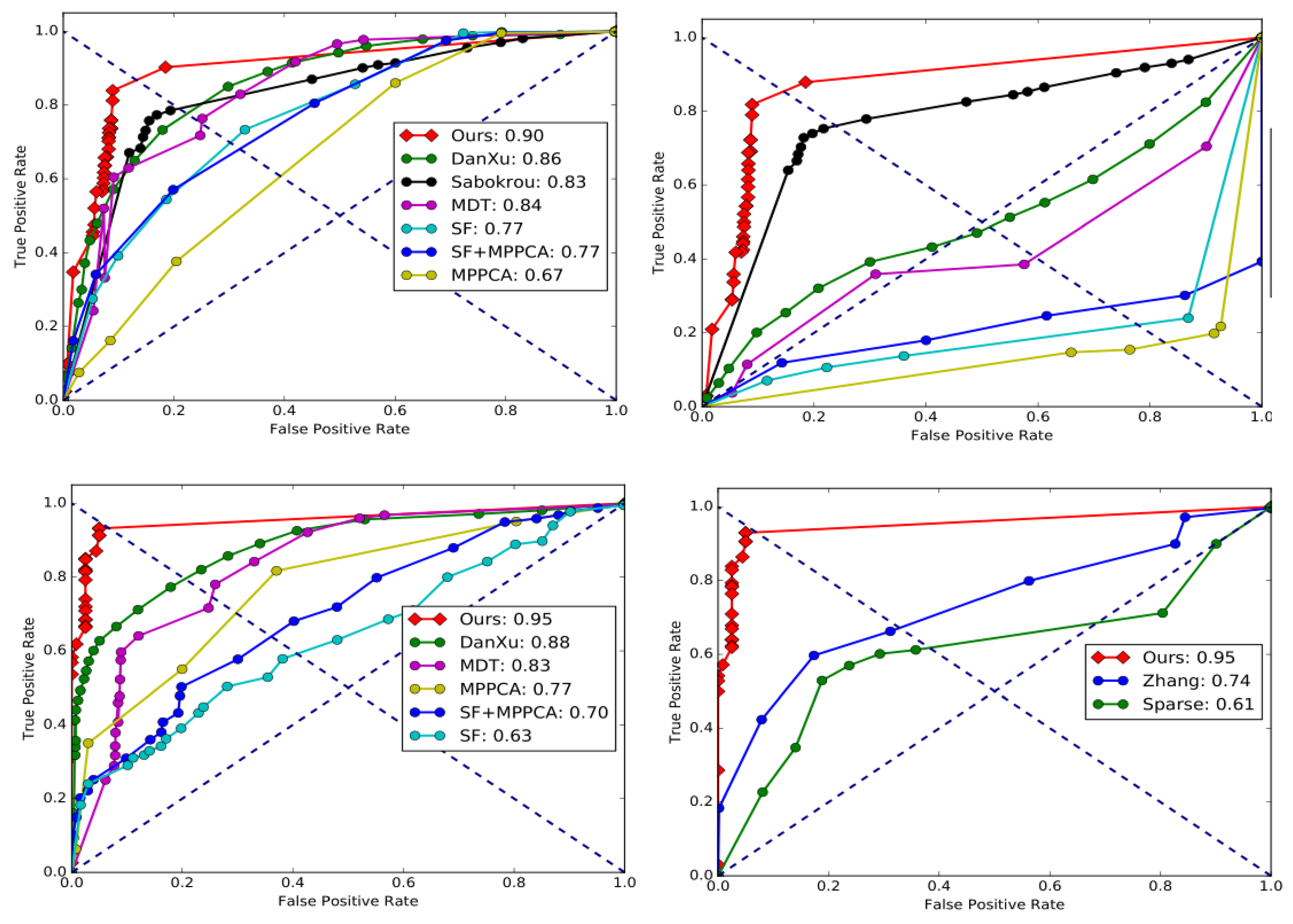
4. Post-processing with Tracking

In video anomaly detection, distant objects are too small to be detected, so we use tracking to pick up missing targets.

5. Our New Dataset: USVD

http://www.ubicom.ustc.edu.cn/source/

6. Experiments



ROC curve: The first, second row: Ped1, Ped2; The first, second column: Frame ,pixel level

Table 1. EER (%) of Frame and Pixel Level on UCSD

Method	Ped1		Ped2	
	Frame	Pixel	Frame	Pixel
SF [24]	31	76	42	80
MPPCA [25]	40	82	30	71
SF+MPPCA [1]	32	71	36	72
Dan Xu [26]	22	42	20	-
Conv-AE [23]	27.9	-	21.7	-
Cascade [13]	**9.1**	15.8	8.2	19
RFCN	46.2	46.8	19.1	21.8
RFCN + T*	33.2	35.3	11.3	14.5
D-RFCN + T*	23.7	26.2	10.5	10.7
D-RFCN + T* + V* + E*	15.5	17.8	6.6	6.9
D-RFCN + T* + V* + E* + B*	13.1	**14.5**	**6.6**	**6.9**

* T: tracking, V: HVOFA, E: energy, B: background model