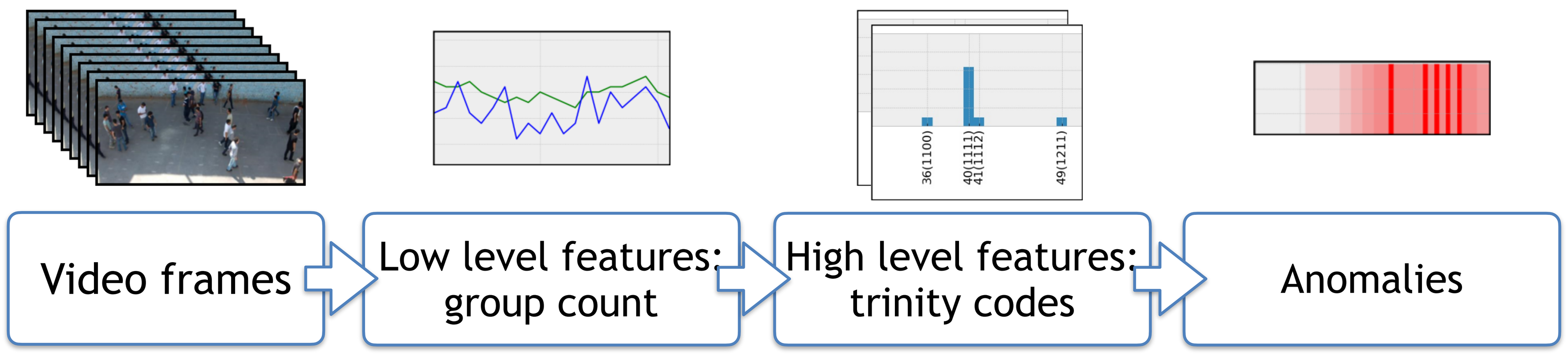
Detecting Anomalies from Video-Sequences: a Novel Descriptor

Giulia Orrù, Davide Ghiani, Maura Pintor, Gian Luca Marcialis, Fabio Roli
Department of Electrical and Electronic Engineering, University of Cagliari, Italy
pralab.diee.unica.it
Email:{giulia.orru, maura.pintor, marcialis, roli}@unica.it, davideghiani@gmail.com

ABSTRACT

We present a novel descriptor, inspired by the one-dimensional local binary pattern (1D-LBP)[1], for crowd anomaly detection. The goal is to measure by appropriate patterns, based on the number of group observed in a time window, the speed of formation and disintegration of groups in the crowd. Our hypothesis is that abrupt variations of the groups' number may be due to an anomalous event that can be accordingly detected, by translating these variations on temporal sequence of strings which are significantly different from the one describing the "no-anomaly" one.

Anomalous Events Descriptor



Our detector is based on the hypothesis that anomalous events happen when multiple group formation events and group breaking-up events suddenly appear in the scenario.
By a time-window sliding over a certain video-sequence's set of frames, and centred on a certain temporal instant, we compare the number of groups computed at that instant with the groups count after and before it. Three cases may happen: the number of groups increases, decreases, or remains unaltered. This leads to have, for each time-window, a set of "trinary" codes.

We compare 4 methods for counting the groups:
- Manual counting as ground truth (MC);
- Clustering of optical flow (COF)[3];
- OpenCV Cascade detector (CD)[4];
- Blob detector (BD).

We then evaluated the histograms of the trinary codes occurrences. A threshold is applied to the histogram central bin, characterizing the state of quiet, and acts as a trigger for detecting the anomaly.

Experimental Protocol

- Motion Emotion data set [2]: 31 video sequences of around 44000 frames in total.
- The videos contain both normal and abnormal behavior, labeled frame-by-frame as 5 classes (panic, fight, congestion, obstacle and neutral).

Plots: anomaly labels are green vertical lines and the corresponding alarms are represented by red vertical lines. The 27-seconds window centred on the anomaly is highlighted in light green.

Metrics:

$$precision = \frac{TP}{TP+FP} \qquad recall = \frac{TP}{TP+FN}$$

$$F1_{score} = 2 * \frac{precision * recall}{precision + recall}$$

Where TP is the number of correctly detected anomalies, FP is the number of false alarms, and FN is the number of undetected anomalies.

Experiment parameters have been set using a grid search, which maximize the $F1_{score}$. The grid search was done both in a supervised way ($F1_{score}$ maximised on all videos) and with a Leave-one-out cross validation ($F1_{score}$ maximised on N-1 videos) and test on the video left out.

Overall Results

	Supervised			Leave-one-out		
	Precision	Recall	F1	Precision	Recall	F1
MC	88.89%	94.12%	91.43%	79.31%	71.87%	75.41%
COF	71.11%	88.89%	79.01%	52.50%	60.00%	56.00%
CD	75.00%	91.67%	82.50%	73.17%	83.33%	77.92%
BD	70.45%	86.11%	77.50%	56.52%	74.29%	64.20%

The group counting method affects the performance of the descriptor. The most reliable method is the Cascade Detector which on all videos achieves better detection performance than even the Manual Counting in the Leave-one-out protocol.

The difference in performance between the supervised and the leave-one-out protocol suggests that a more accurate setting of the parameters calculated by grid search would allow for more reliable detection.

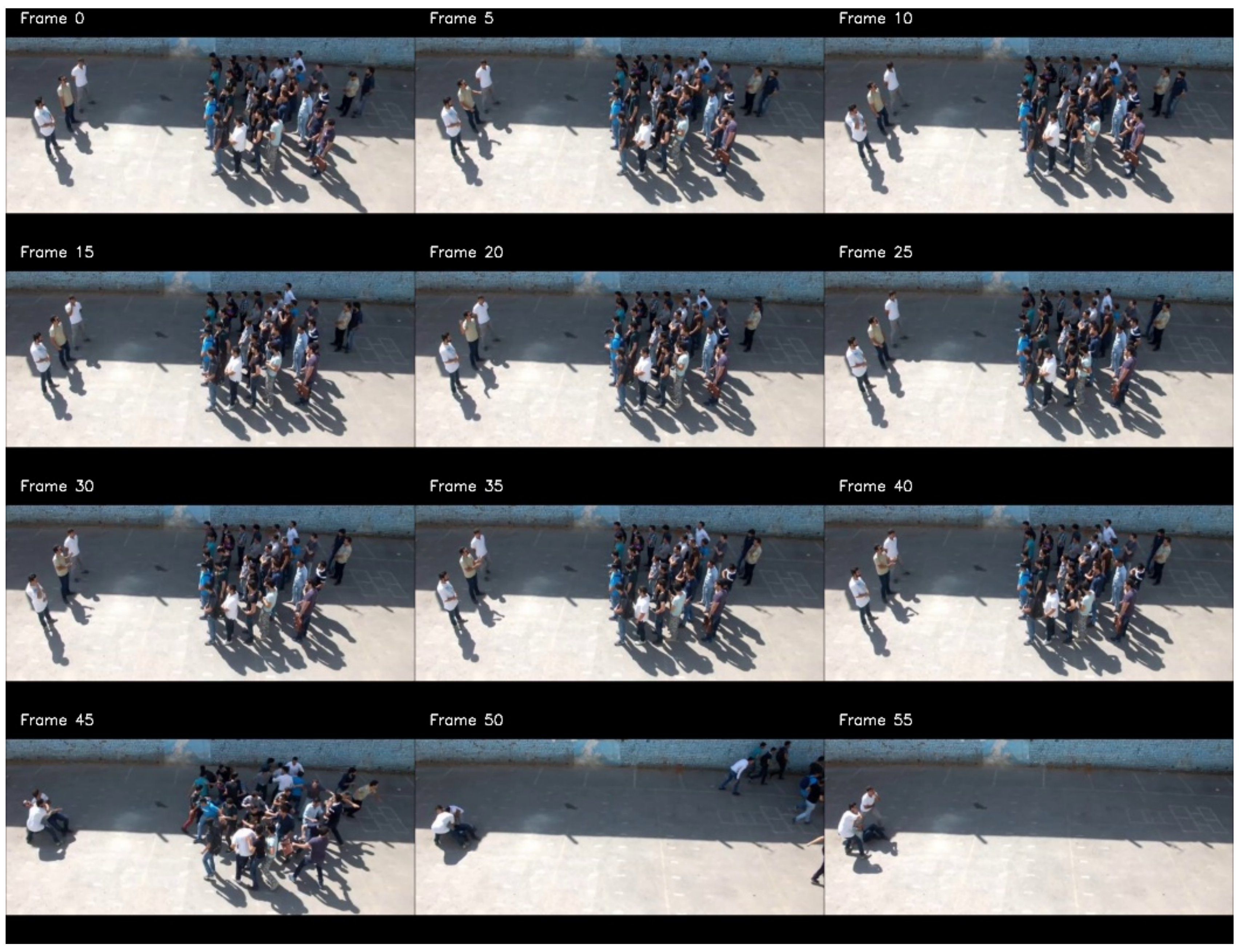
Moreover, a further analysis that distinguishes lateral videos from frontal ones, highlighted that the different position of the camera leads to very different performances. Lateral views appear to have a deeper observation surface than frontals and this might suggest that the descriptor has more "time" to observe changes in the scene, especially in case of anomalous events.

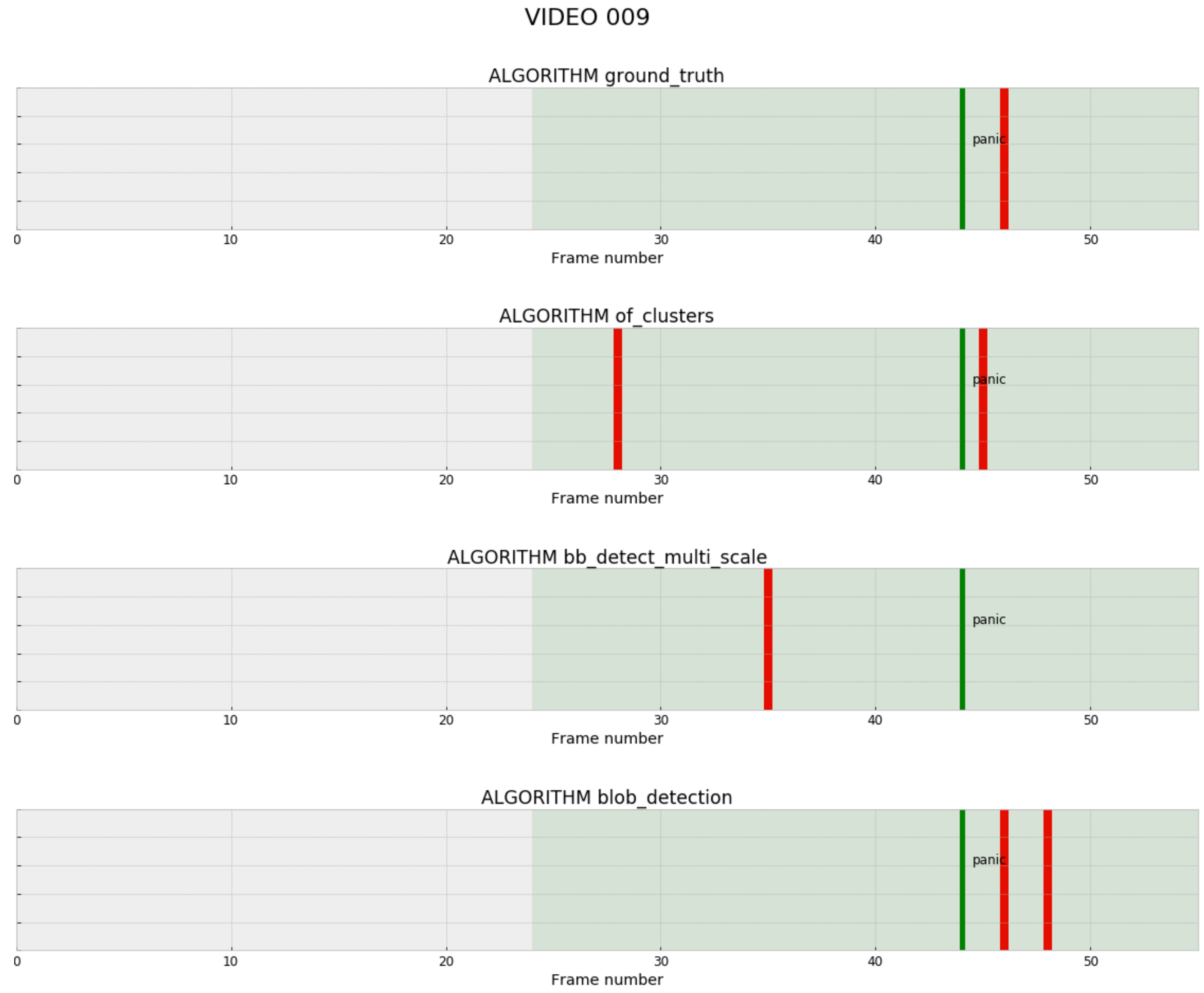
Lateral view slope $> 5°$:

Frontal view slope $<= 5°$:

Results

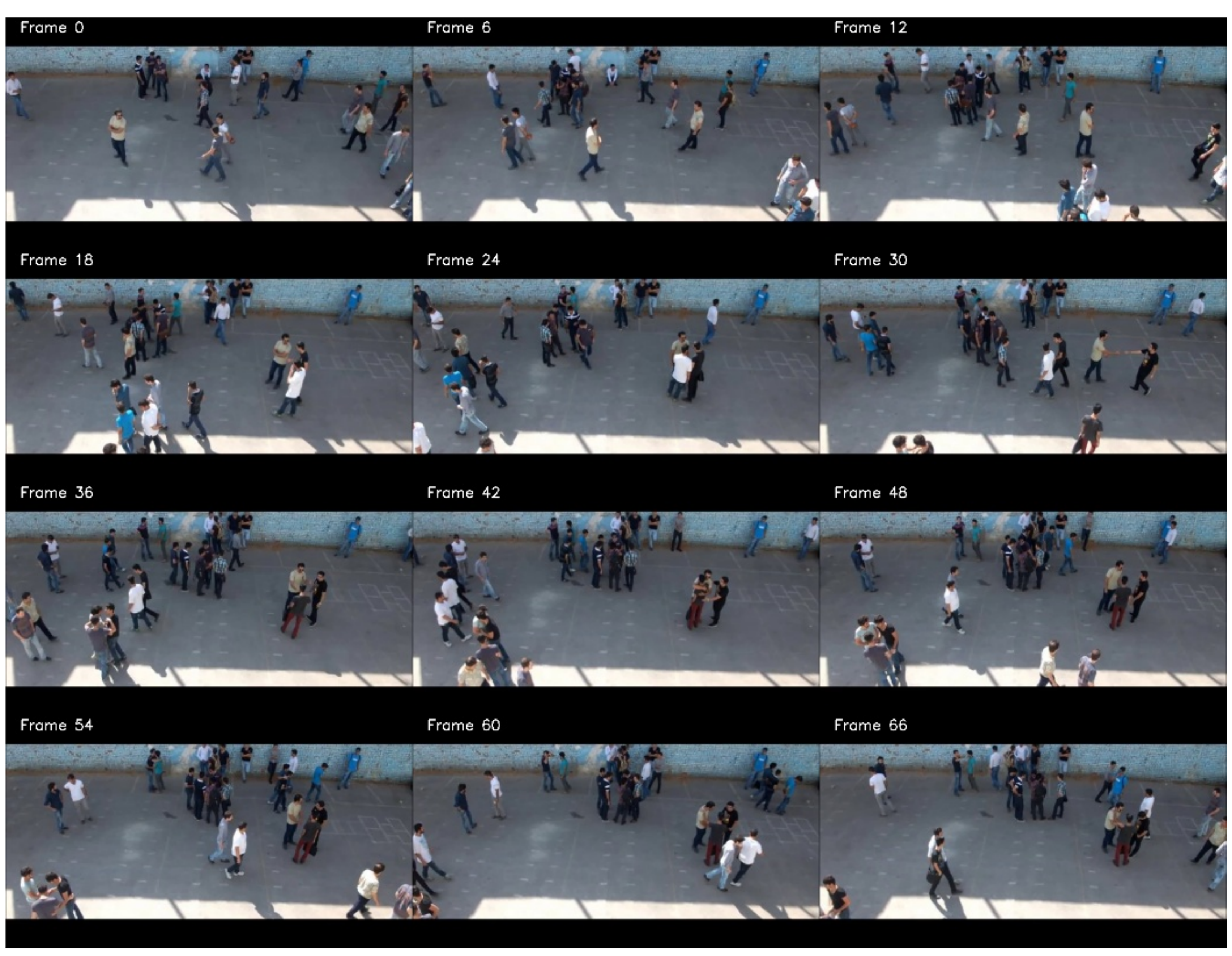
Video 009, panic event:

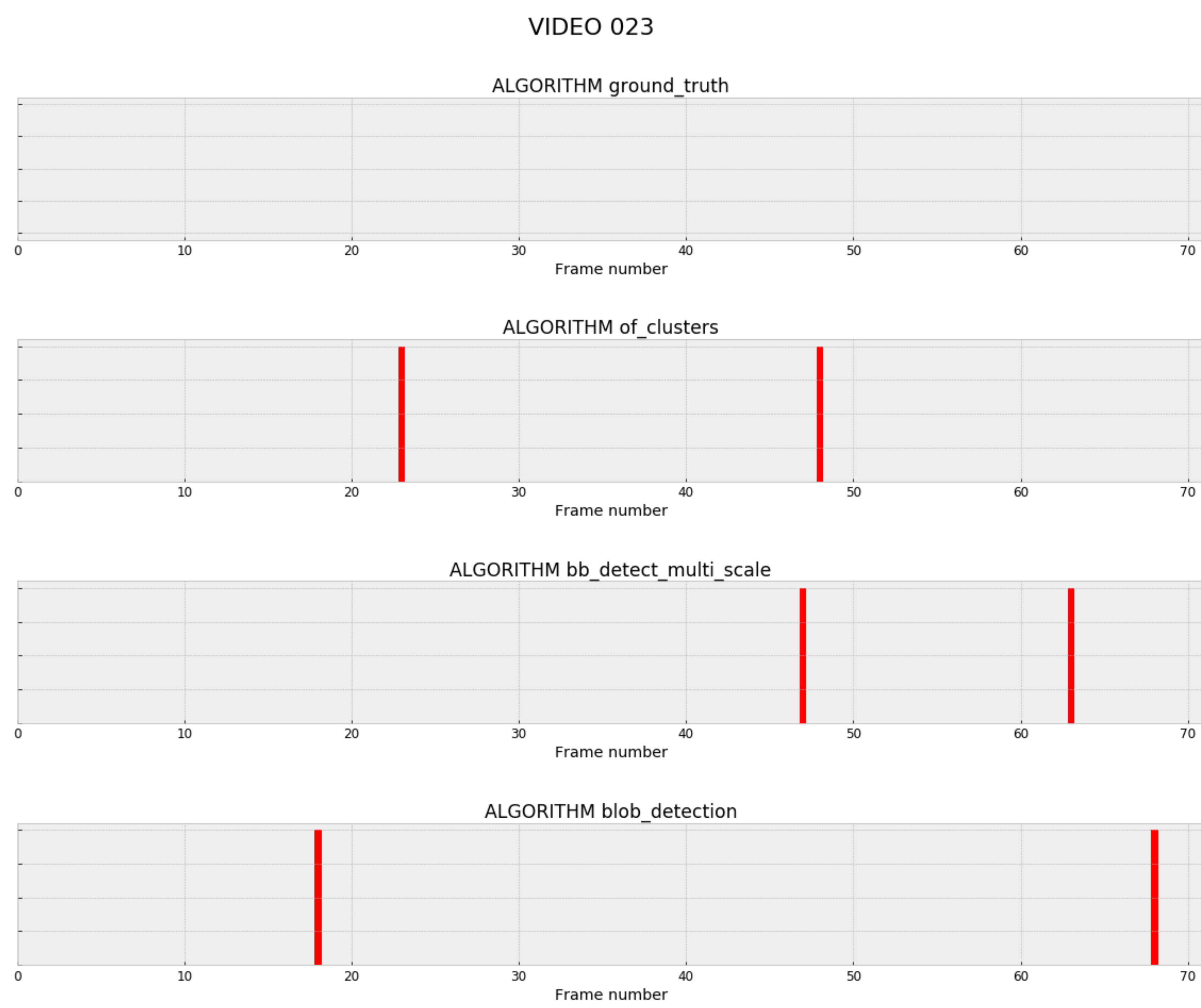




In situation fully in agreement with the hypothesis behind our work, as the video 009, the detectors are able to correctly detect anomalies. This video is characterised by a structured crowd, and contains a constant number of groups. The panic event generates movement whereby the detectors are able to notice the anomaly.

Video 023, normal behavior:





Instead, they make mistakes in the case of unstructured and sparse crowd whit very small changes in the number of groups, as in the video 023.

References

[1] Chatlani, N., & Soraghan, J. J. (2010). Local binary patterns for 1-D signal processing. 95-99.18th European Signal Processing Conference (EUSIPCO-2010), Aalborg, Denmark.
[2] H. Rabiee, J. Haddadnia, H. Mousavi, M. Kalantarzadeh, M. Nabi, and V. Murino, “Novel dataset for fine-grained abnormal behavior understanding in crowd,” in 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2016, pp. 95-101.
[3] B. K. Horn and B. G. Schunck, “Determining optical flow,” Artificial intelligence, vol. 17, no. 1-3, pp. 185–203,1981.
[4] G. Bradski, “The OpenCV Library,” Dr. Dobb’s Journal of Software Tools, 2000.

Conclusions

- White box descriptor easily adaptable depending on the real context and the type of anomaly to be detected.
- The descriptor works well with events that clearly match our hypothesis of quickly number groups changes and presents errors if the disruptions and aggregations are slow and controlled.
- Possibility to build a pipeline without delay for all the methods analyzed.

Acknowledgment

Supported by the Italian Ministry of Education, University and Research (MIUR) within the PRIN2017 - BullyBuster - A framework for bullying and cyberbullying action detection by computer vision and artificial intelligence methods and algorithms

Bully Buster