The LV Dataset: a Realistic Surveillance Video Dataset for Abnormal Event Detection

Roberto Leyva, Victor Sanchez *Member IEEE* and Chang-Tsun Li *Senior Memeber IEEE*

***Abstract*—In recent years, designing and testing video anomaly detection methods have focused on synthetic or unrealistic sequences. This has mainly four drawbacks: 1) events are controlled and predictable because they are usually performed by actors; 2) environmental conditions, e.g. camera motion and illumination, are usually ideal thus realistic conditions are not well reflected; 3) events are usually short and repetitive; and 4) the material is captured from scenarios that do not necessarily match the testing scenarios. This leads us to propose a new rich collection of realistic videos captured by surveillance cameras in challenging environmental conditions, the Live Videos (LV) dataset. We explore the performance of a number of state-of-the-art video anomaly detection methods on the LV dataset. Our results confirm the need to design methods that are capable of handling realistic videos captured by surveillance cameras with acceptable processing times. The proposed LV dataset, thus, will facilitate the design and testing of such new methods.**

***Index Terms*—Video surveillance, video anomaly detection, online processing.**

I. Introduction

Video anomaly detection has recently attracted great interest in the computer vision community as a tool to automatically detect unusual events without specific *a priori* knowledge about these abnormal events [1, 2]. In order to develop and test video anomaly detection methods, different datasets have been created. These datasets usually contain simulated scenes with actors behaving abnormally, e.g. [3–5]; or more realistic scenes with a very limited number of abnormal events, e.g. [6–8]. In general, existing datasets present four main drawbacks. First, many sequences are recorded using predefined scripts which poorly represent realistic scenarios captured by surveillance cameras. Second, the available training and test data are often acquired by different cameras or from different scenes. Third, many sequences are recorded under ideal environmental conditions which do not represent realistic scenarios encountered by surveillance cameras. And fourth, many sequences usually contain a limited number of abnormal events or the abnormal events are very repetitive. Based on these facts, we propose the Live Videos (LV) dataset, which is a new rich collection of video sequences captured in challenging conditions by surveillance cameras depicting realistic abnormal events; e.g., car accidents, robberies, kidnappings, and other dangerous situations. We evaluate a number of state-of-the-art video anomaly detection methods using the new LV dataset and compare their performance with that attained on commonly used datasets. Based on the evaluation results, we discuss the importance of developing anomaly detection methods that can handle realistic events captured by surveillance cameras in challenging environmental conditions.

Roberto Leyva [1] and Victor Sanchez [2] are with the Computer Science Department at University of Warwick, Coventry, United Kingdom. Chang-Tsun Li [3] is with the School of Computing and Mathematics, Charles Sturt University This research has been funded by the Mexican Ministry of Education, SEP-CONACYT, and the EU H2020 Project - IDENTITY. The corresponding contact addresses of the authors are:
[1] M.R.Leyva-Fernandez@warwick.ac.uk
[2] V.F.Sanchez-Silva@warwick.ac.uk.
[3] chli@csu.edu.au

Fig. 1. Sample frames of the LV dataset. Sequences are captured by surveillance cameras from different view angles at different resolutions and frame rates. Illumination changes and camera motion are also present.

The rest of the paper is organized as follows. Section § II presents a summary of datasets commonly used to develop and evaluate video anomaly detection methods. Section § III presents the proposed LV dataset. Section § IV details the evaluation setup and presents the results of the evaluated methods. We discuss these results in Section § V. We finish this paper with conclusions in Section § VI.

II. Previous Work

A. Existing datasets

Early video anomaly detection proposed by Boiman and Irani [4] attempts to identify observed patches (spatio-temporal regions) whose location and appearance cannot be re-constructed using previously seen patches. In their proposed

TABLE I
CHARACTERISTICS OF EXISTING DATASETS FOR DESIGNING AND TESTING VIDEO ANOMALY DETECTION METHODS.

Dataset	Scenes	Subjects	Abnormal Events
Weizmann[4]	1	$[1, 2]$	Crawling, falling, robberies.
QMLU[8]	1	> 100	Jaywalking, prohibited u-turns, improper lane use.
UCSD[6]	2	$[5, \sim 50]$	Non-pedestrian entities.
U-Turn [10]	4	$[1, \sim 30]$	Prohibited u-turns, wrong-way, abandoned baggage.
UMN[3]	3	$[15, \sim 20]$	Panic.
York [5]	6	$[0, \sim 20]$	Wrong-way, payment evasion, running, unusual movements.
Mall[9]	8	$[3, > 100]$	Running, wrong-way direction.
UCF[7]	20	$[\sim 30, > 100]$	Panic, people clashing, fights.

dataset, two actors perform two basic atomic actions, i.e. walking and running. A set of unusual behaviors, such as crawling, a simulated robbery and falling, are then identified as abnormal. In the UMN [3] dataset, people walking in random directions suddenly start to run simulating panic. This dataset shows that abnormal behavior can be generated not only individually but also collectively by a crowd acting as a single entity. In the UCF dataset, Ali and Shah [7] propose to use realistic data from crowd and panic scenes. These scenes are then used to detect people fighting, clashing or in panic. An important aspect of their work is the detection of abnormal behavior when the training scenarios differ from the test scenarios. In the Mall dataset [9], people running is considered as abnormal behavior, while people walking is considered to be normal behavior. In the York dataset, Zaharescu *et al.* [5] propose two video sequences acquired at the entrance and exit of a subway station. Abnormal behavior consists of people entering the station without paying, exiting the station through the entrance, and entering through the exit. Mahadevan *et al.* [6] propose the UCSD dataset, which contains sequences with different crowd densities where the abnormal events are the presence of non-pedestrians, e.g., cyclists, skaters, small carts, and people in wheelchairs. In the QMLU dataset [8], a sequence depicting traffic and pedestrians in a street junction is provided . Normal activity is interrupted when an ambulance and a fire truck enter the scene causing other vehicles to slow down or move in unusual directions.

The main characteristics of the aforementioned datasets are summarized in Table I. These datasets, which comprise not only abnormal events but also abnormal changes in behavior, have successfully facilitated the development of various video anomaly detection methods. However, in order for anomaly detection methods to be useful in practical situations, e.g., with surveillance cameras, datasets should have no semantic definition of the events, since abnormal events in realistic videos are usually not rigidly defined [2].

B. *Video anomaly detection methods*

Detecting abnormal events in practical settings; e.g., surveillance cameras, usually require online processing in order to process frames as they are acquired. In this section we thus review methods that achieve the minimum processing times and are suitable for online processing. For a detailed review of other existing methods, the reader is referred to [1, 11].

Javan and Levine [12] propose a framework based on densely constructed spatio-temporal video volumes, which are organized into large contextual graphs. A hierarchical codebook model for the dominant behaviors is constructed. The framework is capable of simultaneously modeling high-level behaviors and low-level spatial, temporal and spatio-temporal pixel level changes. These spatio-temporal compositions are later improved in [13], in terms of processing times, by considering the temporal differences of video volumes as a descriptor. In [14], Lu *et al.* propose a fast sparse reconstruction where a low rank projection captures intrinsic volumetric compositions of small regions in the video. To enhance this approach, these regions are analyzed in a coarse-to-fine fashion. Biswas *et al.* [15] propose to analyze the motion vectors directly from the compressed motion blocks using a statistical model. To that end, a kernel density estimator is employed to link the vectors' displacement in a probabilistic model. Reddy *et al.* [16] analyze the local responses of Gabor filters in a cell structure via Gaussian models in local spatio-temporal neighborhoods. The method also incorporates foregound information in a two-way inference mechanism.

III. PROPOSED LV DATASET

The proposed LV dataset (see Fig. 1 for examples) comprises video sequences characterized by the following aspects:

- Realistic events without actors performing predefined scripts with a diverse subject interaction.
- Highly unpredictable abnormal events in different scenes, some of them of very short duration.
- Scenario correspondence, where the training and test data are captured from the same scene.
- Challenging environmental conditions; sequences are acquired under changing illumination and camera motion.

The LV dataset consists of 30 sequences with 14 different abnormal events. The main characteristics of this dataset are given in Table II and sample sequences are shown in Fig. 4.

As proposed in [4, 5], we include in the LV dataset sequences where the abnormal events are those that rarely occur in a scene. In this case, the activity in the scene may not

TABLE II
MAIN CHARACTERISTICS OF THE PROPOSED LV DATASET.

Duration	3.93 hours
Frame Rate	7.5 - 30 FPS
Resolution	minimum: QCIF (176×144) maximum: HDTV 720 (1280×720)
Format	MP4 video in H.264
No. of videos/scenes	30/30
URL	https://cvrleyva.wordpress.com/
Anomalous Frames	68989
Events of Interest	34
Abnormal events	Fighting, people clashing, arm robberies, thefts, car accidents, hit and runs, fires, panic, vandalism, kidnapping, homicide, cars in the wrong-way, people falling, loitering, prohibited u-turns and trespassing.
Scenarios	Outdoors/indoors, streets, highways, traffic intersections and public areas.
Crowd density	No subjects to very crowded scenes.

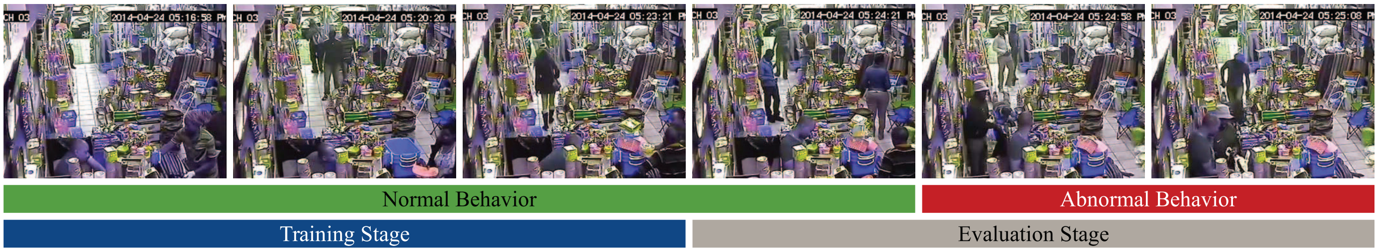

Fig. 2. Example of normal and abnormal frames from the Robbery sequence of the LV dataset. During the training and evaluation stages, normal behavior is present in the scene.

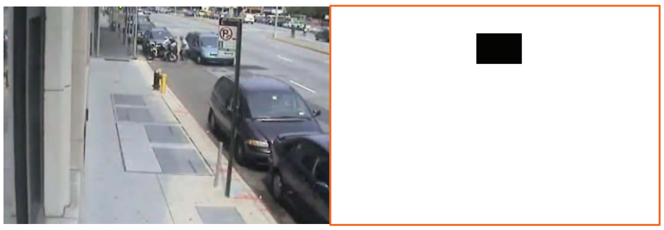

Fig. 3. Sample frame and provided ROI to detect abnormal events. The two burglars putting a motorcycle inside a SUV is the abnormal event.

necessarily increase during the abnormal event; for instance robberies, car accidents and loitering. As proposed in [3, 7, 9], we also include sequences where the activity in the scene may significantly change during the abnormal event; such as fights, people clashing, fire and panic scenes. An example is given in Fig. 2, where the cashier next to the point of sale and customers are deemed to be normal. Although the cashier and the customers perform a relatively large number of actions, when the burglars break into the scene the activity significantly changes. The cashier is beaten by the burglars and some customers are forced to lie on the floor.

It is important to mention that some datasets, e.g., the dataset in [6], provide the ground truth for the specific abnormal objects. Providing such pixel-level ground truth may be very challenging. As an example, consider a video sequence depicting an armed robbery in a connivence store. Are only the burglar and cashier to be considered as anomalous for, respectively, showing a weapon and rising their hands? Should the surrounding people witnessing the robbery and trying to escape (or not to interfere) be considered as anomalous too? In this example, the beginning and end of the chain of individual actions must then be defined in order to correctly determine which objects are abnormal, which may be a difficult task. In the proposed LV dataset, we bound the anomalous event to the spatio-temporal region in the sequence where the main action of interest occurs.

All sequences in the LV dataset contain both abnormal and normal behavior, as in the York dataset [5]. Thus, the LV dataset is not divided into training and test sequences, as in [6]. The labeled frames are provided along with the dataset. As mentioned early, we do not provide the ground truth at pixel-level segmentation of abnormal objects. Instead, the regions of interest (ROIs) are provided in a separate sequence of the same length as that of the training/testing sequence (see Fig. 3). We provide in a separate file the time stamps to determine the frames to be used for training and those for testing in each sequence.

IV. Experiments and Results

We conduct experiments to evaluate the LV dataset on four of the fastest performance methods reported in the literature, namely the method prosed by Javan in [13], by Reddy in [16], by Biswas and Babu in [15], and that proposed by Lu in [14]. We also evaluate these methods, plus other state-of-the art methods, on two of the most popular existing datasets, i.e., the UMN [3] and UCSD [17] datasets, in order to compare their performance attained with more realistic videos. For the LV dataset, after training the corresponding inference model using the normal frames of each sequence, we process the remaining frames in the sequence and classify them to rank the framework. During evaluation, if a normal frame is labeled as abnormal, we count the associated event as a false negative. An abnormal frame is considered to be a true positive if at least one of the ROIs is correctly detected; otherwise the frame is considered as a false positive. A ROI is considered to be correctly detected if at least 20% of its spatial extent is labeled as abnormal. This specific criterion is used because a ROI may contain both foreground and background pixels, and it is a well known fact that background subtraction in complex scenes is very challenging. Unlike other existing datasets, we do not evaluate the detection at the frame level in the LV dataset. This is because in such frame level evaluations, a method can be *lucky* [6] in detecting something that is not the main event of interest. In such cases, the method may detect an abnormal frame correctly but not the specific events that make it so.

To evaluate Reddy's method, we employ the code available in [16] keeping the default parameter setting. To evaluate Lu's method, we employ the implementation available in [18]. To evaluate the method of Biswas and Babu, we use the code available in [19] to estimate the interpolation steps; a maximum of $N = 5$ GMM model components are used. The other parameters are kept as proposed by the authors. Because some of the sequences in the LV dataset contain very large frames (e.g. 1280×720), we re-scale them to a fixed frame size of 160×240. We run all experiments without tuning parameters according to the environmental conditions. To rank the approaches, we compute the Receiver Operating Characteristic (ROC) of each method and report the Equal Error Rate (EER) and Area Under Curve (AUC). We also record the computation time to verify that the methods achieve online performance for all the videos in the tested dataset. A method is said to meet online requirements if the total frame processing time is less or equal to the Frame Per Second (FPS) rate; for example, for a 30 FPS the maximum frame processing time is 33ms. Results for the LV dataset, in terms of EER/AUC

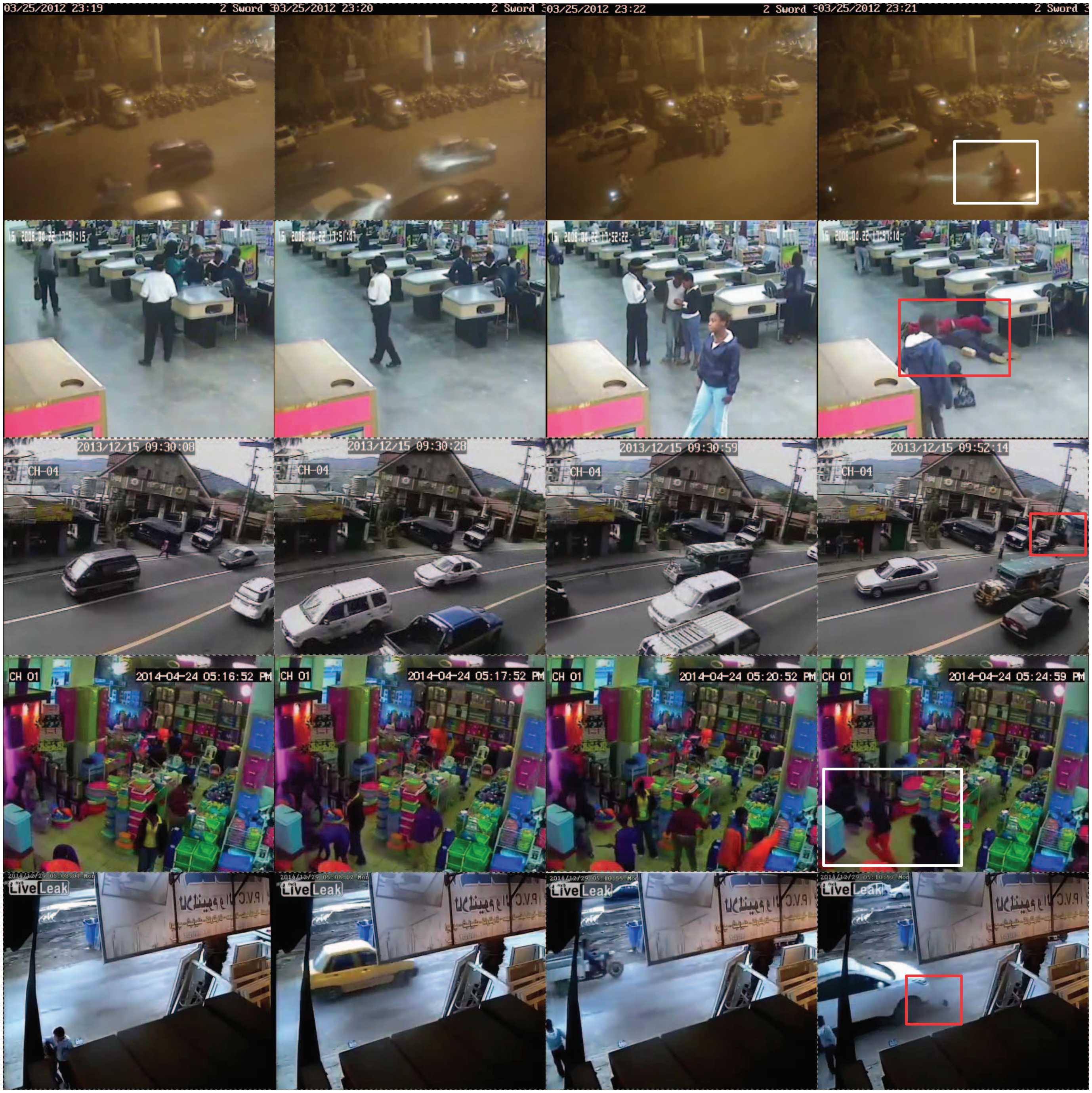

Fig. 4. Sample sequences of the LV dataset. Abnormal events are highlighted by a rectangle. **1st row. Wrong Way Sequence**: traffic goes in a single direction; suddenly some men come out of three cars and people start running; one of them starts shooting (top right corner of frame) and traffic slows down; motorcycles start circulating in the wrong way. **2nd row. Robbery Sequence**: a security guard is at the exit of a supermarket; three armed men suddenly brake into the supermarket; they beat one of the cashiers and force some costumers to lie on the floor. **3th row. Traffic Accident Sequence**: a two-way road with pedestrians on the sidewalks; a truck crashes into a house hitting a car and a light pole. **4th row. Panic Sequence**: costumers in a convenience store look around and pick up items; suddenly (captured by another surveillance camera) four armed men enter the store and costumers start running; some costumers try to hide and escape through the exit. **5th row. Run Over Sequence**: traffic in a two-way road; three children are playing close to road when a car runs over one of them; the driver continues; some adults help the kid. Sequences are captured in poor lighting conditions (e.g., *Wrong Way*); with slight camera motion (e.g. *Traffic Accident*); or the motion of objects is considerably weak making it challenging to detect events (e.g. *Panic*).

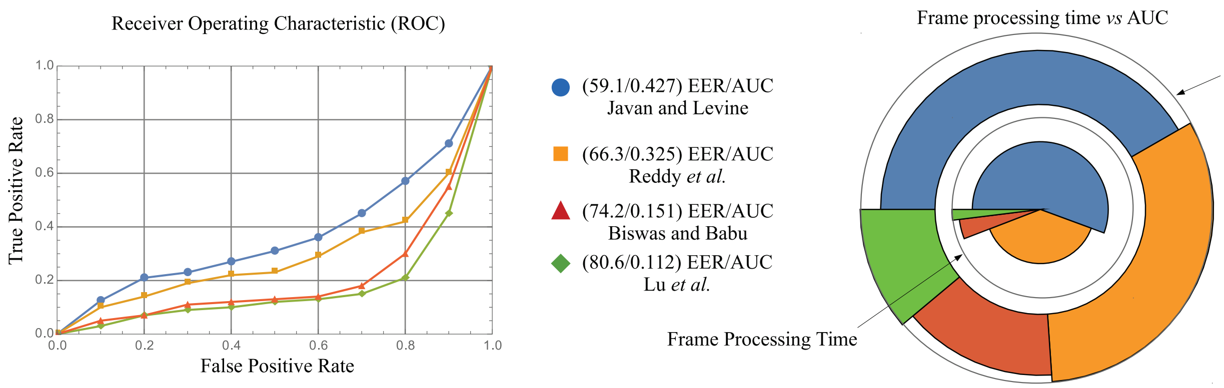

Fig. 5. (Left) ROC for different methods using the LC dataset. (Right) Frame processing time (inner circle) and AUC (outer circle) for different methods.

and ROC, are reported in Table III and Fig 5 (left). Results for the UMN dataset, in terms of the AUC, are tabulated in Table IV; while those for the UMN dataset, in terms of EER, are tabulated in Table V. In all tables, we also indicate which methods achieve online performance. For low FPS values, it is very likely that a method meets the requirements for online performance; we thus only mark a method as online if it achieves online performance for all videos in the dataset.

As we can see from Fig. 5 (right), there is an important trade-off between accuracy and processing times. The methods proposed by Reddy and Javan [13, 16] fail to meet online requirements for the proposed LV dataset. Although the methods proposed by Biswas and Lu [14, 15] easily meet the online requirements, they attain a considerably inferior performance compared to the other compared methods.

TABLE III
EER/AUC FOR THE LV DATASET

Authors	EER/AUC	Frame Processing Time	On-line Performance
Javan and Levine [13]	**59.1 / 0.427**	185 ms	
Reddy *et al.* [16]	66.3 / 0.325	128 ms	
Biswas and Babu [15]	74.2 / 0.151	13 ms	✓
Lu *et al.* [14]	80.6 / 0.112	**6.5 ms**	✓

TABLE IV
AUC FOR THE UMN DATASET

Authors	AUC	Frame Processing Time	On-line Performance
Li *et al.* [20]	0.996	1100 ms	
Cong *et al.* [21]	0.973	3800 ms	
Zhu *et al.* [22]	0.997	4600 ms	
Biswas and Babu [15]	0.736	**14 ms**	✓

V. Discussion

Based on Table III, we can confirm that testing realistic scenes is very challenging. For existing datasets, e.g. the UMN and UCSD datasets, it has been recently reported nearly perfect detection (see Table IV, AUC>0.99; and high detection rates in Table V). It is important to recall, however, that the videos of most of the existing datasets do not present the same challenges as those found in realistic surveillance videos. Thus analyzing realistic material is a good opportunity to enhance the performance of existing methods. During the evaluation, we observe that some of the tested methods in Table III consistently failed in poorly illuminated scenes of the LV dataset. The methods in [13–15] rely on frame gradients as a source of motion information, which may not be easy to extract in poorly illuminated scenes. Another problem of some of these methods is their memory demands. For example, the method proposed by Javan in [13] requires to store very large sets of features to generate the models (in the order of millions for some videos). In this case, it is necessary to split and randomly subsample the data to process the videos of the LV dataset. We record memory peaks demands of 23GB just to store the features for some of these videos. The method proposed by Biswas and Babu in [15] and that proposed by Lu in [14] achieve online performance for the LV dataset, but trigger a high number of false alarms even in well illuminated scenes. Therefore, these online methods require significant improvements to achieve acceptable performances on surveillance videos. We also observe that the method proposed by Reddy in [16], despite not being the most accurate one on the LV dataset, is able to detect the areas in which motion takes place in poorly illuminated scenes thanks to the foreground extractor. However, the inference mechanism is not

TABLE V
EER FOR THE PEDS1 SCENE OF THE UCSD DATASET.

Authors	EER Frame Level	EER Pixel Level	Frame Processing Time	On-line Performance
Reddy *et al.* [16]	22.5	32	140 ms	
Javan and Levine [13]	15	27	190 ms	
Hu *et al.* [23] †	18	36	200 ms	
Cheng *et al.* [24]	23.7	37.3	500 ms	
Li *et al.* [20]	17.8	–	1100 ms	
Cheng *et al.* [25]	19.9	38.8	1100 ms	
Cong *et al.* [21]	23	51.2	3800 ms	
Zhu *et al.* [22]	15	–	4600 ms	
Lu *et al.* [14]	**15**	59.1	**6 ms**	✓
Biswas and Babu [15]	24.66	50.95	14 ms	✓

† After optical flow and histogram calculations.

activated because it considers both, optical flow information and foreground information. Optical flow in poorly illuminated scenes tends to be very weak if no significant movement take places. Thus, the method considerably misses detections because of this two-way criterion. Overall, results confirm that designing methods that are suitable to process realistic videos acquired by surveillance cameras, preferably in an online fashion, is an important untackled challenge. The LV dataset introduced in this paper opens a new door to develop and test the next generation of anomaly detection methods for surveillance videos.

VI. Conclusions

Most of the state-of-the-art video anomaly detection methods have been designed and tested on datasets that poorly reflect realistic events commonly found in videos acquired by surveillance cameras, thus hindering their applicability in practical situations. It is therefore important to facilitate the design and testing of inference mechanisms for realistic data. To this end, we proposed a new rich collection of realistic videos captured by surveillance cameras, the LV dataset. This paper also evaluated a number of state-of-the-art methods on the LV dataset and compared their performance with that attained on existing datasets. Results confirm that important improvements are needed to improve detection accuracy and reduce frame processing times for realistic surveillance videos. The proposed LV dataset is therefore aimed at facilitating the development of these improvements.

References

[1] A. Sodemann, M. Ross, and B. Borghetti, "A review of anomaly detection in automated surveillance," *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, 2012*, vol. 42, no. 6, pp. 1257–1272, Nov 2012.

[2] C. Piciarelli and G. Foresti, "Surveillance-oriented event detection in video streams," *IEEE Intelligent Systems, 2011*, vol. 26, no. 3, pp. 32–41, May 2011.

[3] "Univeristy of minnesota dataset," http://mha.cs.umn.edu/proj_events.shtml.

[4] O. Boiman and M. Irani, "Detecting irregularities in images and in video," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 17–31, 2007. [Online]. Available: http://dx.doi.org/10.1007/s11263-006-0009-9

[5] A. Zaharescu and R. Wildes, "Anomalous behaviour detection using spatiotemporal oriented energies, subset inclusion histogram comparison and event-driven processing," in *Computer Vision – ECCV 2010*, ser. Lecture Notes in Computer Science, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Springer Berlin Heidelberg, 2010, vol. 6311, pp. 563–576. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-15549-9_41

[6] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010*, June 2010, pp. 1975–1981.

[7] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007*, June 2007, pp. 1–6.

[8] C. C. Loy, T. Xiang, and S. Gong, "Modelling multi-object activity by gaussian processes." in *BMVC*, 2009, pp. 1–11.

[9] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008*, vol. 30, no. 3, pp. 555–560, March 2008.

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

[11] O. Popoola and K. Wang, "Video-based abnormal human behavior recognition;a review," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews,2012*, vol. 42, no. 6, pp. 865–878, Nov 2012.

[12] M. Roshtkhari and M. Levine, "Online dominant and anomalous behavior detection in videos," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013*, June 2013, pp. 2611–2618.

[13] M. J. Roshtkhari and M. D. Levine, "An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions," *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1436 – 1452, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077314213001239

[14] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 fps in matlab," in *IEEE International Conference on Computer Vision (ICCV), 2013*, Dec 2013, pp. 2720–2727.

[15] S. Biswas and R. Babu, "Real time anomaly detection in h.264 compressed videos," in *National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013*, Dec 2013, pp. 1–4.

[16] V. Reddy, C. Sanderson, and B. Lovell, "Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2011*, June 2011, pp. 55–61.

[17] "Ucsd anomaly detection dataset," http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm.

[18] http://shijianping.me/.

[19] "Kernel density estimator," http://uk.mathworks.com/matlabcentral/fileexchang

[20] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014*, vol. 36, no. 1, pp. 18–32, Jan 2014.

[21] Y. Cong, J. Yuan, and Y. Tang, "Video anomaly search in crowded scenes via spatio-temporal motion context," *IEEE Transactions on Information Forensics and Security, 2013*, vol. 8, no. 10, pp. 1590–1599, Oct 2013.

[22] X. Zhu, J. Liu, J. Wang, C. Li, and H. Lu, "Sparse representation for robust abnormality detection in crowded scenes," *Pattern Recognition*, vol. 47, no. 5, pp. 1791 – 1799, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320313005049

[23] Y. Hu, Y. Zhang, and L. Davis, "Unsupervised abnormal crowd activity detection using semiparametric scan statistic," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2013*, June 2013, pp. 767–774.

[24] K.-W. Cheng, Y.-T. Chen, and W.-H. Fang, "Video anomaly detection and localization using hierarchical feature representation and gaussian process regression," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015*, June 2015, pp. 2909–2917.

[25] K. Cheng, Y. Chen, and W. Fang, "Gaussian process regression-based video anomaly detection and localization with hierarchical feature representation," *IEEE Transactions on Image Processing, 2015*, vol. 24, no. 12, pp. 5288–5301, Dec 2015.