Malicious Attacks Detection in Crowded Areas Using Deep Learning-Based Approach

Fouzi Harrou, Mohamad Mazen Hittawe, Ying Sun and Ouadi Beya

Abnormal Event Detection

With the increasing need to ensure people's safety in crowded areas, the development of a systematic anomaly detection mechanism is becoming indispensable. Here are a few examples of recent malicious attacks targeting crowded areas in big cities: in 2016, a truck driver attacked and killed 84 persons walking in the promenade in Nice, France; and on 19 December, 2016, a truck was deliberately driven into the Christmas market, in Berlin, Germany, killing 12 people and injuring 56 others. These attacks demonstrate the need for efficient monitoring systems to avoid such devastating attacks. To do so, early detection and prevention abilities are vital. Detecting and localizing abnormal events in crowded scenes is important, and has significant implications in video surveillance applications. Video surveillance can be challenging, as abnormal events can be unpredictable and changing, based on the context. Accurately detecting and localizing anomalies in videos is a powerful tool that can help to improve security and understand the behavior of anomalies. In this paper, we present an automated vision-based monitoring scheme specifically designed for atypical event-detection and localization in crowded areas.

To achieve the reliable detection of abnormal events based on videos, a substantial amount of research effort has been invested over the last two decades [1]. Several methods have been designed using trajectories analysis to uncover atypical events [1], [2]. However, these techniques usually need precise tracking solutions and are highly sensitive to occlusion [3]. Alternatively, other techniques use spatio-temporal features for representing the events in the video and do not require trajectories analysis [4]. Techniques falling into this category use low-level local visual features (e.g., motion or texture) for background modeling and construction of template behavior [4]. In [5], a multi-scale and non-parametric approach is proposed to detect and locate occurred atypical events in real-time. In [6], at first the optical flow is applied to select video volumes of interest; and then a convolutional neural network (CNN) is used to extract relevant features. Many studies have been done to extract spatiotemporal features based on the idea of Bag of Video words (BOV). Essentially, the backbone idea of this concept to detect abnormal events consists of using local video volumes based on dense sampling or by selecting points of interest. Unfortunately, this approach ignores the relationship between video

volumes. Recently, alternative solutions have been proposed using the contextual information, but high computational capacity is needed for the implementation of these methods, which makes them unsuited in real-time purposes. In [7], the authors present fully CNNs, combining a pre-trained CNN and another convolutional layer trained using sparse auto-encoder. In [8], Bouindour *et al.* propose to use CNNs and a 1-Class support vector machine algorithm to detect anomalies in video datasets. Note that the major challenge is to extract relevant descriptors and design detection schemes able to uncover unusual and atypical behaviors with a high detection rate and a minimum of false alarms [9]–[11].

Recently, with the rapid advancements in deep learning and computational technologies, representations of data are accomplished in a sophisticated way and learned by end-to-end neural networks. This study was motivated by the strong capacity of the deep learning CNN to extract relevant features from images. Particularly, this paper proposes an efficient vision-based approach for attack detection and localization in crowded areas. This approach merges the desirable properties of a CNN to learn relevant features from videos, the flexibility of the k-Nearest Neighbor (kNN) algorithm, and the sensitivity of double Exponentially Weighted Moving Average (DEWMA) to sense small changes in time series data.

In this paper, we treat the problem of malicious attacks in crowded areas as an anomaly detection problem based on features extracted from the CNN model. The design of a CNN-based detection approach is performed in two phases. The first phase consists of constructing the reference CNN model that mimics the normal situations based on the data without abnormal events (attacks). In addition, we computed the detection threshold of the kNN-DEWMA approach based on the CNN features. In the second phase, we apply the constructed CNN model to generated features based on testing, and we apply the kNN-DEWMA approach to the new features using the detection threshold previously computed to detect abnormal events.

Specifically, the CNN features are examined using the proposed kNN-DEWMA approach for the purpose of atypical events detection. Importantly, this approach takes advantage of the great ability of the kNN algorithm to quanify the similarity between normal and abnormal CNN's features to better uncover atypical events. This is mainly due to its capacity to handle nonlinear features without making hypotheses on the data distribution.

The major reason for double exponentially smoothing kNN measurements (kNN-DEWMA) is to include all of the information from past and actual samples in the decision rule, which make it sensitive to small anomalies. Moreover, we applied the DEWMA scheme to kNN

measurements to incorporate all of the information from past and current measurements in the decision process, which offers more sensitivity to small changes. Furthermore, to obtain a flexible and assumption-free approach, the non-parametric kernel density estimation is used to compute the detection threshold. Lastly, for the spatial localization of abnormal events in each frame of the video, the Gaussian Distribution of Mahalanobis Distance (GDMD) is applied only on the frames flagged by the kNN-DEWMA detector. Results on benchmark data indicate that the developed method has a higher detection efficiency than the traditional competitors by a large margin.

Proposed CNN-KNN for Abnormal Attacks Detection

This section presents a vision-based method for attack detection and localization in crowded areas. The implementation of this method is performed in two steps: (1) extracting robust features from normal video data, based on the first two convolution layers of a pre-trained CNN; and (2) applying the proposed KNN-DEWMA scheme to check the CNN features to detect anomalies in the video (Fig. 1).

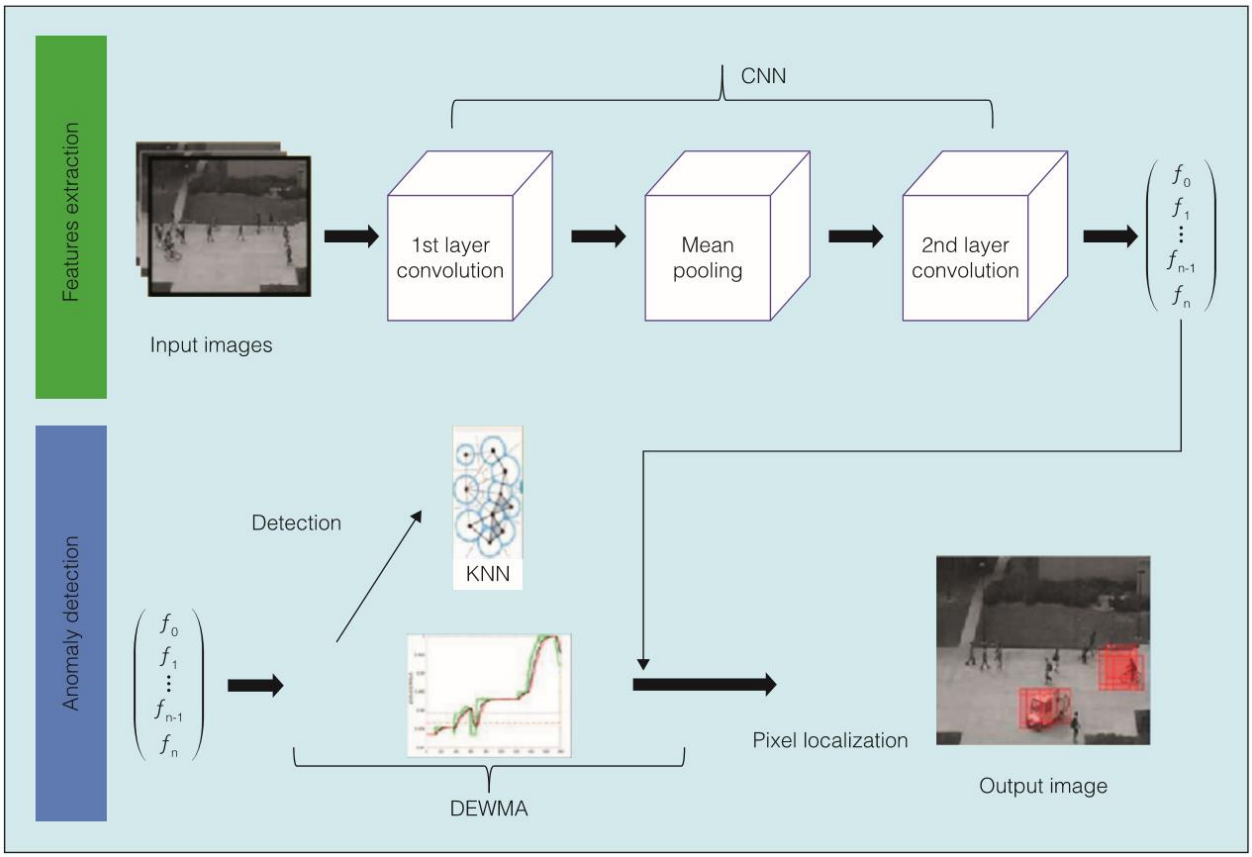

Fig. 1. Conceptual representation of the proposed monitoring method.

Features Extraction

Various model Structures of Deep Neural networks were shown to be effective for many tasks because of their structures utilize hidden features. One of the effective tools for features extraction in deep neural networks is CNN, and in particular, Alexnet [12] which has been trained and learned on a huge number of images to extract robust features from these images. In our proposed scheme, the first step is to use many frames: Ft, Ft-1, Ft-2, D = [Ft, Ft-1, Ft -2] as input to the first two convolutional layers of Alexnet to generate the features maps [13]. In the second step, we use one pooling layer between the two convolutional layers of the network to minimize the number of parameters and control the overfitting in the network. Thus, a matrix of dimensions 529×256 is our resultant feature map; each row of this matrix represents the feature vector of one each patch in the input image, as shown in Fig. 1.

Proposed Monitoring Approach

In this strategy, the KNN-DEWMA is used to evaluate the features produced by the CNN model (Alexnet). This approach combines the desirable features of the KNN and the detection capacity of the DEWMA scheme to detect very small changes [14]. The KNN is a very efficient nonparametric algorithm for separating different features; this feature is essential, in particular when data are non-Gaussian- distributed or not linearly separable [15], [16]. Overall, KNN separates normal data from abnormal data by measuring the distance between the actual observation and the k-nearest neighbors of anomaly-free data. In this study, KNN is applied to the features extracted from the CNN model (Alexnet) to check the presence of abnormal events in the supervised area. Large KNN distances are used as an indicator to detect anomalies. Here, the DEWMA monitoring scheme is employed to check the KNN distances for detecting atypical events. This approach allows the detection of anomalies without using any data labeling. The KNN-DEWMA statistic W_t is computed as:

$$\begin{aligned} W_0 = S_0 = \mu_0\,, \\ W_t = v s_t + (1-v) w_{t-1} \\ S_t = v d_t + (1-v) s_{t-1}, \qquad t = 1,2,\ldots\ldots,n. \end{aligned} \tag{1}$$

where d_t represents the KNN distance of the actual CNN features and the training features. Note that the DEWMA scheme, with two different smoothing parameters to compute the statistics s_t and W_t in (1), offers similar performance that the DEWMA with equal parameters as demonstrated in [17]. Here, we use DEWMA with an equal smoothing constant, as recommended in [17]. The variance of the DEWMA statistic is given as:

$$var(w_t) = v^4{\sigma_0}^2 \frac{1+(1-v)^2-(1-v)^{2t}((t+1)^2-(2t^2+2t-1)(1-v)^2+t^2(1-v)^4)}{(1-(1-v)^2)^3} \quad (2)$$

For large t, the asymptotic variance can be computed as:

$$Var_{asymtotic}(w_t) = {\sigma_0}^2 \frac{v(2-2v+v^2)}{(2-v)^3} \quad (3)$$

The anomaly detection can be performed by plotting the KNN-based DEWMA statistic with its upper and lower detection thresholds U and L:

$$U, L = \mu_0 \pm k\sigma_0 \sqrt{\frac{v(2-2v+v^2)}{(2-v)^3}}, \quad (4)$$

where μ_0 and σ_0 represent the mean and the standard deviation for the training data. In the design DEWMA scheme, we need to specify the values of the parameters v and k. k is usually called the width of the detection thresholds. It is frequently selected in practice to be 3 that corresponds to a false alarm rate of 0.27%, implying that 99.73% of the observations should be contained within the detection limits in normal conditions. On other hand, v is usually called the forgetting or smoothing parameter ($0 < v \leq 1$) that defines the temporal memory of the DEWMA statistic. From (1), we can see that the selection of smaller values to v gives more weight to the past measurements and thus the detection schemes will be more sensitive to uncover small changes. When a large value of v is chosen, then more importance is given to the actual measurement and less importance is given to the previous measurements. In practice, to uncover small changes in process mean, the value of v is generally taken within 0.2 and 0.3 [17], [18]. More details about DEWMA can be found in [17]. An anomaly is flagged when the KNN-DEWMA statistic exceeds the control limits. It should be noted that the detection thresholds in (4) are derived based on the normality assumption of the data.

KNN-DEWMA Scheme

The conventional parametric DEWMA control procedure is suitable when the Gaussian hypothesis is verified. However, this assumption cannot be guaranteed here, as the kNN distances are used as input for the DEWMA scheme. Basically, the kNN is used to compute the dissimilarity between the normal and abnormal CNN features. Thus, in that case, the Gaussianity assumptions of the kNN distances can be violated. Then, the performance of the

DEWMA-kNN scheme using a detection threshold computed based on the Gaussian distribution could be significantly degraded. To alleviate this limitation, an assumption-free approach is introduced by using kernel density estimation [19] for estimating the distribution of kNN distances and setting up the detection threshold.

- Step: 1 Compute kNN-DEWMA statistic, Di, for each CNN features data,
- Step 2: From the distribution of Di, a nonparametric threshold of the kNN-DEWMA procedure is computed as (1 - α)-th quantile of the distribution of the KNN-DEWMA statistic, W_t, obtained using the kernel density estimator.
- Step 3: Declare atypical events if the kNN-DEWMA statistic exceeds the control limit.

Abnormal Attacks Localization Using GDMD

We convert a feature space into a distance space by calculating the Mahalanobis distance between the features F_i extracted (in features extraction step) from the normal data (without anomaly). We then create the model of these extracted features by finding the Gaussian distribution of Mahalanobis distances (GDMD) of the training data.

Each new test frame f_{test} is assigned to the first two convolutional layers to extract the features Ft. Then, we find the Mahalanobis distance dm between each row of F_t and the GDMD model of the normal data, if the distance d_m is greater than threshold s (selected using the experimentations). We classify this patch in the frame test as abnormal attack. Because the convolution and pooling of a fully connected network are approximately invertible, we can localize the abnormal attack in the frame test f_{test}.

Results and Discussion

We tested the performance of our proposed method on a Pedestrian (Ped2 UCSD) dataset (http://www.svcl.ucsd.edu/projects/anomaly). This dataset contained different outdoors scenes with complex abnormal behaviors in some parts of its videos. The UCSD Ped2 dataset consisted of 16 training videos to build the model of normal data and 12 videos for validation. This dataset is complex because the video contains many occlusions and low-resolution images obtained from multiple crowded areas.

An atypical event is considered as abnormal behavior if there is a non-pedestrian action in the video, e.g. a skateboarder, bicycle and/or car among walking pedestrians. The detection quality of the two KNN-DEWMA methods was quantified using a false positive rate (FPR), a true positive rate (TPR), Precision, Accuracy, F-measure, Recall, equal error rate (EER) and the area under the curve (AUC). In our method, a False Negative (FN) was counted for each frame with abnormal events that were not detected by our method, and a False Positive (FP) was

counted for each number of frames without any abnormality flagged as an abnormal frame by our method.

The kNN-DEWMA scheme with nonparametric threshold was applied on the features generated by CNN, for the 12 scenarios in the USCD Ped2 dataset. In our study, the smoothing parameter 'ν' was set to 0.25. The monitoring results are summarized in Table 1 and highlight the superiority of the DEWMA scheme with a nonparametric threshold, by the fact it achieved an AUC of 0.94 and had a lower EER of 0.05 (Frame level).

Table 1 – Quantitative comparison of KNN-DEWMA methods for abnormal attack detection

Average	TPR	FPR	Accuracy	Precision	F-Measure	AUC	Recall	EER
KNN-DEWMA (PA)	0.99	0.24	0.91	0.90	0.94	0.87	0.99	0.09
KNN-DEWMA (NP)	0.96	0.08	0.95	0.97	0.96	0.94	0.96	0.05

Fig. 2 illustrates the detection of abnormal events (red color) in the UCSD Ped2 dataset, with "INPUT" representing the input videos, "Detection" representing the detection of anomalies detection in video frames, and "Localization" representing the localization of anomalies in each frame.

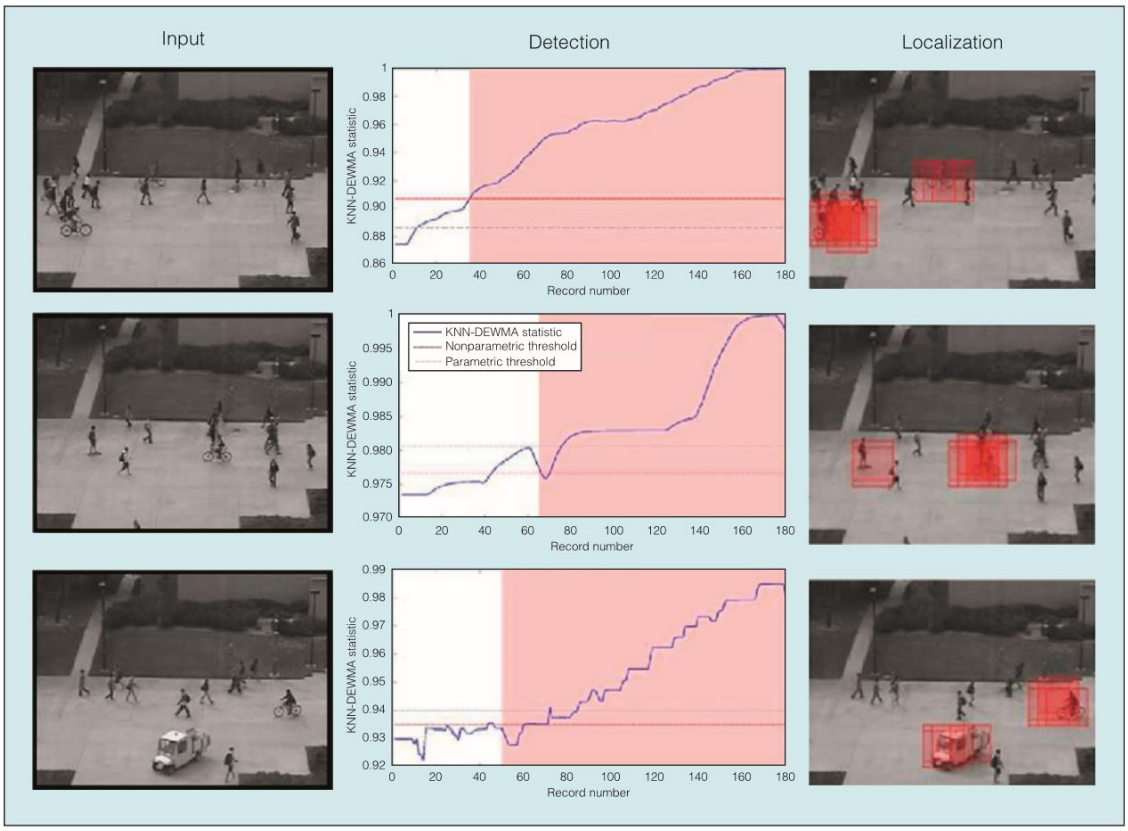

Fig. 2. Detection and localization results of the proposed scheme using UCSD PED2 data: "INPUT" represents the input videos, "Detection" represents the detection of anomalies in video frames, and "Localization" represents the localization of anomalies in each frame.

Fig. 3 shows that better performances are obtained with the CNN-based DEWMA-kNN approach, in comparison with some state-of-the-art machine learning procedures using PED2 dataset. Our method performs better than most of the state-of-the-art methods at the frame level. Additionally, results show a higher efficacy of the nonparametric DEWMA-kNN algorithm, compared with the parametric one. This is mainly due to the fact that the kNN-kNN incorporates all existing information from previous and actual observations in the decision, which extends its detection performance.

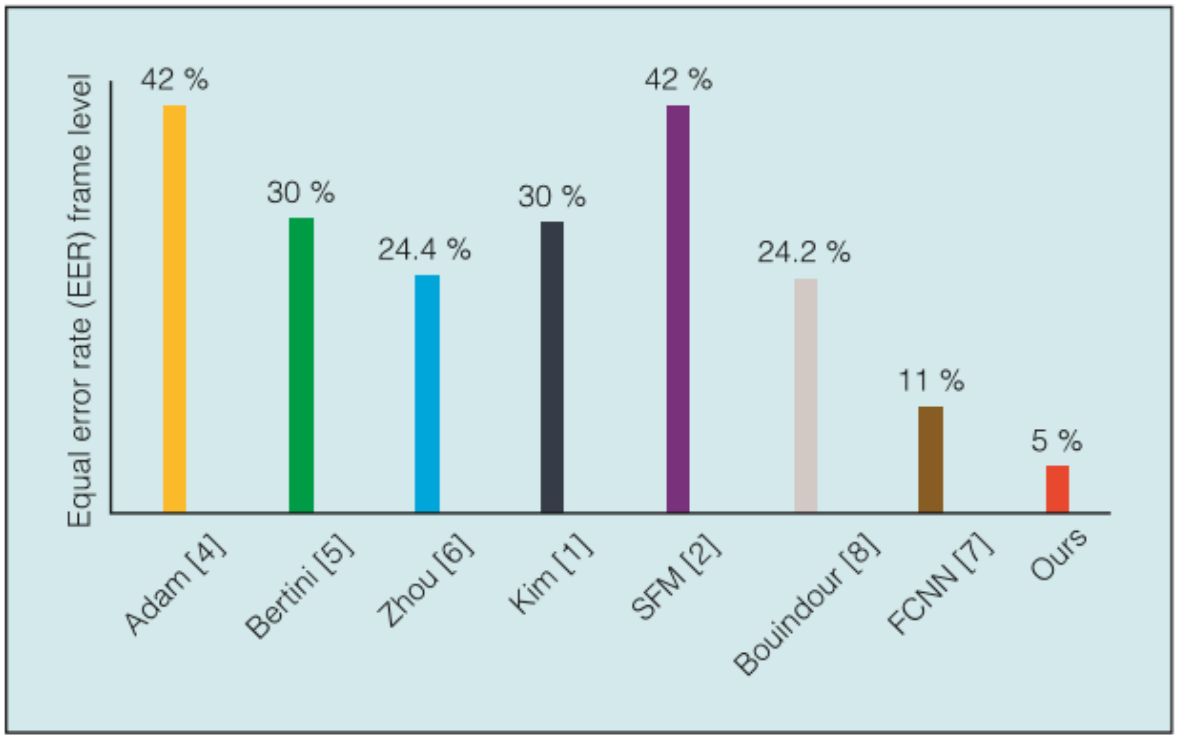

Fig. 3. Comparison of frame-level detection using the proposed approach with traditional methods.

Conclusion

Crowded urban areas are becoming more vulnerable to terrorist threats. Thus, the need for vision-based monitoring systems that can detect and localize abnormal objects in crowded areas is becoming essential. Here, we propose an innovative vision-based detection and localization method for the detection of atypical events in crowded areas. Our method uses an integrated approach merging the benefits of the CNN model, the kNN algorithm and the DEWMA monitoring scheme. CNN is used to extract complex and pertinent features from the captured videos from the supervised area. kNN is then applied to compute the dissimilarity between the

actual CNN features and the CNN features from anomaly-free videos. The DEWMA statistical monitoring scheme is subsequently used for sensing abnormal changes based on kNN distances, followed by the computation of a nonparametric threshold for DEWMA, using KDE to extend its flexibility and sensitivity to small changes. Lastly, the GDMD is applied to localize abnormal events for each frame in the videos. The performance of the proposed method is verified using UCSD datasets, which clearly show a higher efficiency of our method, in comparison with other state-of-the-art methods.

Acknowledgement

This work was supported by King Abdullah University of Science and Technology, Office of Sponsored Research, under Award no: OSR-2019-CRG7-3800.

References

[1] J. Kim and K. Grauman, "Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2009)*, pp. 2921-2928, 2009.

[2] R. Mehran, A. Oyama, and M. Shah. "Abnormal crowd behavior detection using social force model," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2009)*, pp. 935-942, 2009.

[3] X. Ma, H. Wang, B. Xue, M. Zhou, B. Ji, and Y. Li, "Depth-based human fall detection via shape features and improved extreme learning machine," *IEEE J. Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1915-1922, 2014.

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

[5] M. Bertini, A. Del Bimbo, and L. Seidenari, "Multi-scale and real-time non-parametric approach for anomaly detection and localization," *Computer Vision and Image Understanding*, vol. 116, no. 3, pp. 320-329, 2012.

[6] S. Zhou, W. Shen, D. Zeng, M. Fang, Y. Wei, and Z. Zhang, "Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes," *Signal Processing: Image Commun.*, vol. 47, pp. 358-368, 2016.

[7] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayedd, and R. Klette, "Fully convolutional neural network for fast anomaly detection in crowded scenes," 2016.

[8] S. Bouindour, M. M. Hittawe, S. Mahfouz, and H. Snoussi, "Abnormal event detection using convolutional neural networks and 1-Class SVM classifier anomaly detection in crowded scenes," *IET*, ICDP 2017.

[9] S. Shirmohammadi and A. Ferrero, "Camera as the instrument: the rising trend of vision based measurement," *IEEE Instrum. and Meas. Mag.*, vol. 17, no. 3, pp. 41-47, 2014.

[10] M. M. Hittawe, D. Sidibé, and F. Mériaudeau, "A machine vision based approach for timber knots detection," in *Proc. 12th Int. Conf. Quality Control by Artificial Vision*, Int. Soc. for Optics and Photonics, Apr. 2015.

[11] O. Beya, M. M. Hittawe, D. Sidibé, D. and F. Meriaudeau, "Automatic detection and tracking of animal sperm cells in microscopy images," in *Proc. 11th IEEE Int. Conf on Signal-Image Technol. and Internet-Based Syst. (SITIS)*, pp. 155-159, Nov. 2015.

[12] F. Gao, J. Lin, H. Liu and S. Lu, "A novel VBM framework of fiber recognition based on image segmentation and DCNN," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 963-973, April 2020.[13] C. Nuzzi, S. Pasinetti, M. Lancini, F. Docchio, and G. Sansoni, "Deep learning-based hand gesture recognition for collaborative robots," *IEEE Instrum. Meas. Mag*., vol. 22, no. 2, pp. 44-51. 2019.

[14] S. E. Shamma and A. K. Shamma, "Development and evaluation of control charts using double exponentially weighted moving averages," *Int. J. Quality and Reliability Manage.*, vol. 9, no. 6, 1992.

[15] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. New York, NY, USA: Elsevier, 2011.

[16] F. Harrou, N. Zerrouki, Y. Sun, and A. Houacine, "Vision-based fall detection system for improving safety of elderly people," *IEEE Instrum. Meas. Mag*., vol. 20, no. 6, pp. 49-55, Dec. 2017.

[17] L. Zhang and G. Chen, "An extended ewma mean chart," *Quality Technol. and Quantitative Manage.*, vol. 2, no. 1, pp. 39-52, 2005.

[18] F. Harrou and M. N. Nounou, "Monitoring linear antenna arrays using an exponentially weighted moving average-based fault detection scheme," *Syst. Science and Control Eng. an Open Access J.*, vol. 2, no. 1, pp. 433-443, 2014.

[19] E. B. Martin and A. J. Morris, "Non-parametric confidence bounds for process performance monitoring charts," *J. Process Control*, vol. 6, no. 6, pp. 349-358, 1996.

Fouzi Harrou (fouzi.harrou@kaust.edu.sa) is a Research scientist at the King Abdullah University of Science and Technology, Saudi Arabia. He received the M.Sc. degree in telecommunications and networking from the University of Paris VI in 2006 and the Ph.D. degree in systems optimization and security in 2010 from the University Technology of Troyes, France. His current research interests include anomaly detection and process monitoring, machine learning and deep learning.

Mohamad Mazen Hittawe (mohamad.hittawe@kaust.edu.sa) is a Post-doctoral Researcher at the King Abdullah University of Science and Technology, Saudi Arabia. He received his M.S. degree in computer science from the Ecole d'ingenieurs Polytechnique of Tours University, France in 2012 and his Ph.D. degree in machine learning and computer vision from the University of Burgundy, France in 2016. His current research interests include machine learning, deep learning and visual analytics.

Ying Sun (ying.sun@kaust.edu.sa) is an Associate Professor of statistics with the Division of Computer, Electrical and Mathematical Sciences and Engineering at King Abdullah University of Science and Technology, Saudi Arabia, where she joined in June, 2014 after one-year service as an Assistant Professor in the Department of Statistics at the Ohio State University, USA. She leads a multidisciplinary research group on environmental statistics, dedicated to developing statistical models and methods for space-time data to solve important environmental problems.

Ouadi Beya (beya.ouadi@u-paris.fr) is a Professor of signal processing, informatics, applied physics and mathematics at Paris University, France. He received his Ph.D. degree in instrumentation, signal and image processing from the University of Burgundy, France. His current research interests include biosignals, image and signal processing, machine learning, and deep learning.