Multimedia Tools and Applications (2020) 79:24851–24871
https://doi.org/10.1007/s11042-020-09024-z

Real-time frequency-based detection of a panic behavior in human crowds

Bahya Aldissi[1] · **Heyfa Ammar**[2]

Received: 18 June 2018 / Revised: 22 March 2020 / Accepted: 5 May 2020 /
Published online: 26 June 2020

Abstract

The real-time detection of a panic behavior in a human crowd is of a high interest as it helps alleviating crowd disasters and ensures that timely appropriate action will be taken. However, the fast analysis of video sequences to detect abnormal behaviours is one of the most challenging tasks for computer vision experts. While many research works propose off-line solutions, few studies investigate the real-time analysis of crowded scenes. This may be due to the fact that detecting a panic behaviour is closely related to the analysis of the crowd dynamics, which commonly necessitates heavy computations. In order to alleviate this problem, we propose a real-time panic detection technique that analyzes the crowd movements based on a simple and efficient solution. The key idea of the proposed approach consists of analyzing the interactions between moving edges along the video in the frequency domain. Our contribution is threefold. First, moving edges are considered for analysis along the video. Second, when a panic situation occurs within a human crowd, it leads to interactions between people that are different from those that occur during a normal situation. Therefore, to reveal such a behavior, a new frequency based-feature is proposed. To select the most appropriate frequency domain, the fast fourier transform, the discrete cosine transform and the discrete wavelet transform are investigated. Third, two different formulations of the problem of detecting a panic are explored. The experimental evaluation of the proposed technique shows its outperforming compared to the state-of-the-art approaches in terms of detection rates and execution time.

Keywords Real-time detection · Crowded scenes · Abnormal behaviors · Frequency domain · Clustering data

✉ Heyfa Ammar
heyfa.amar@gmail.com

Bahya Aldissi
behaldissi@kau.edu.sa; bahyaaldissi@gmail.com

[1] FCIT, King Abdulaziz University, KSA, Jeddah, Saudi Arabia

[2] Laboratory of Robotics, Informatics and Complex Systems (RISC-Lab), National Engineering School of Tunis, University of Tunis El Manar, Tunis, Tunisia

1 Introduction

The study of crowd scenes is becoming a field of considerable interest to researchers, mainly due to the rising number of popular events and public places that facilitate the mass gathering of people. Such occasions and spaces include markets, subways, religious festivals, sporting events and public demonstrations [32, 43]. Often, a crowd may induce a disastrous event due to fight, congestion, mass panic or various other reasons [18]. Many crowd disasters occurred recently [1, 6, 14].

In an attempt to prevent such deadly disasters from occurring, most public areas including holy places, campuses, residential areas and airports are now equipped with closed-circuit television (CCTV) surveillance cameras. The incoming video can be automatically analysed to facilitate the early detection of a possible abnormal event. The automatic detection of panic behaviours is the interest of the present study. Panic manifests as a sudden change in the crowd dynamics. It appears in the video feed as an atypical behaviour: moving in different directions, speed increase, collective running, grouping in one region and so on.

Many research projects have been conducted to automatically detect panic behaviours [7, 12, 13, 15, 21, 22, 24, 26, 29, 31, 34, 35, 40–42]. Despite their good detection performances, the majority of them propose off-line solutions. Although they are useful in many situations such as police investigation, it is important to detect a panic situation as soon as it occurs using a real-time detection approach.

To the best of our knowledge, few real-time techniques are proposed in the literature [13, 15, 22, 24, 26, 31, 34, 35, 42]. The common scheme of the panic detection approaches is mainly composed of three steps. First, the motion field is estimated since motion is a crucial characteristic of the crowd dynamics. Second, a feature that characterizes the crowd behavior is extracted. It is based on the way a panic behavior is defined. Third, panic is detected as a deviation of the values, taken by the selected feature, from those obtained during a non panic situation. For instance, in [35], a panic is identified by the presence of *atypical* motion patterns in the scene. Motion patterns of a non panic situation are learned thanks to the computation of motion representative subspaces on videos of normal behaviors. During the testing phase, the motion field of the considered video is estimated. Then, an approximation using the representative subspaces determined through the training stage is carried out. If the error between the estimated motion and its approximation exceeds a threshold adjusted by the user, then the presence of an abnormal behavior is concluded. However, the performances of this technique depend on the amount of trained data and the diversity of the human behavior in crowded scenes makes it difficult the enumeration of all possible normal behaviors.

More recently, a real-time detection technique based on texture modelling is proposed in [34]. It associates the occurrence of a panic behavior to a temporal change of the texture. According to the authors [34], this technique achieves an execution time of 18 ms per image. However, its performances may degrade when the spatial-temporal texture patterns are highly heterogeneous during a non-panic situation.

Detecting grouping and running behaviors are the main interest of the work described in [42]. First, the motion is estimated using an optical flow (OF) estimation approach. This estimation is carried out only on the Harris corners in order to alleviate the computations. Second, two parameters are chosen as features to characterize the two behaviors. The first parameter is a crowd distribution index that reflects the gathering level of people in a local area. The second parameter is the velocity. The combination of the two parameters forms the kinetic energy of the crowd. Running or gathering behaviors are detected if the energy

or the crowd distribution index are greater than a threshold. This technique achieves a near real time execution of 20 images per second.

In [24], panic is defined as an unexpected change in the spatial occupancy of moving objects. An abnormal event is recognized when a high temporal variation of the space occupancy occurs during a given time interval. However, the space occupancy is computed in terms of number of pixels. Thus, it does not take into consideration the way by which the space is occupied along the video. In other terms, the space can be occupied by approximately the same number of moving pixels, but not with the same spatial distribution. The spatial positions of moving blobs are more likely to change during a panic situation as a consequence of the reaction of pedestrians when a dangerous event occurs.

By examining the reported real time techniques, two main limitations can be pointed out. First, although the motion information is important in characterizing the crowd dynamics, the OF estimation yields heavy computations. Also, tracking moving objects or restricting the motion estimation to some points of interest may fail in the presence of occlusions or a highly dense crowd. To alleviate this problem, we suggest to detect moving pixels by computing the absolute differences between pairs of successive images. Contrary to what one might expect, this step does not affect the robustness of the whole proposed system, to noise, occlusions and illumination variations as demonstrated later in Section 4.

The second limitation is that considering a panic as a temporal change in the texture may limit the performances of the system when the scene is highly textured. In the same way, defining a panic as a temporal change in the number of moving pixels within the crowded area may degrade the detection accuracy when people cannot move outside the area in presence of panic. As a solution, and motivated by the fact that the occurrence of a panic situation changes the way people behave with respect to each other, panic is viewed, in the present study, as a sudden change in the interactions between people. The same definition of panic has been considered recently in [29] and led to high detection performances in crowded scenes of any density level. However, the approach described in [29] cannot be investigated for real-time detection due to the computational complexity of the OF applied to all pixels of each image. In the present work, we restrict the analysis of the interactions between moving objects to the interactions between moving edges. Thus, the problem of analysing the spatial interactions between pedestrians is formulated as the problem of analyzing how the spatial distribution of moving edges varies in time. A sudden and remarkable temporal variation is associated to the occurrence of a panic situation. Our rationale is to conduct the analysis of the spatial distribution of moving edges, in the frequency domain. This is motivated by the fact that the spatial distribution of edges is easily perceived in the frequency domain by coefficients of high values. Furthermore, the transformation into the frequency domain allows to get a sparse representation of the spatial discontinuities. Any transformation to the frequency domain could be applied. The fast fourier transform (FFT), the discrete cosine transform (DCT) and the discrete wavelet transform (DWT) are explored in the present work.

To analyze the crowd behavior, a new feature is proposed based on the coefficients obtained through the transformation. A sudden increase in the value of the feature reveals the presence of a panic. In order to temporally locate the panic behavior, the values taken by the proposed feature are classified into two subsets: the first one is related to the normal instances of the video while the second one corresponds to the panic instances. We perform this classification using two different techniques as detailed in Section 3. An experimental comparison between both is presented in Section 4.

Three main contributions are proposed in this work:

1. The first contribution aims to alleviate the heavy computations resulting from applying a motion estimation technique, by considering the absolute differences between pairs of successive images of the video. This solution allows a fast analysis reaching 406 frames per second (fps) as shown in Table 4.
2. Considering a panic situation as a sudden change in the interactions between people, our second contribution consists of associating the spatial distribution of moving edges to people interactions. Thus, the problem of analyzing people interactions is formulated as the problem of analyzing the temporal variation of moving edges distribution.
3. As a third contribution, we propose a new feature that characterizes the interactions between pedestrians. Our rationale is to sparsely represent moving edges in the frequency domain where they are expressed by coefficients of high values. When panic starts, the spatial distribution of moving edges suddenly changes implying a remarkable change in the values of the coefficients. Hence, the feature we propose is the sum of the coefficient absolute values at each instant.

The rest of the paper is organized as the following. The datasets used in this study are described in Section 2. Then, the proposed system is detailed in Section 3 and experimentally evaluated in Section 4. Next, the results are discussed in Section 5. Finally, some conclusions are drawn in Section 6.

2 Datasets

A variety of datasets are used in order to deal with various scenes. As depicted in Table 1, videos including artificial and real behaviors with different density levels, and different image sizes are analyzed.

A brief overview of each data set is given in what follows.

University of Minnesota (UMN) dataset It is a public dataset produced by the University of Minnesota, USA [23]. It is composed of 11 video sequences representing escape events, and captured in various contexts: Lawn, Indoor and Plaza. People in these videos walk around normally until an abnormal event occurs which makes them run away. A ground truth (GT) of this dataset is available in [12].

Motion Estimation Dataset (MED) is a public dataset that includes 11 videos of panic behaviors [25]. Typical scenarios are: putting down a suspicious backpack, earthquake, hoodlum attack and sniper attack. The GT of this dataset is annotated and made publicly available by authors of the work [25].

Table 1 Characteristics of the datasets

Name	*Scene*	*Behavior*	*Number of frames*	*Frame size*	*Frame rate*	*Density level*
MED	Real	Artificial	45000	480 × 854	30	Medium/High
UMN	Real	Artificial	7739	320 × 240	30	Low
PETS 2009	Real	Artificial	1944	768 × 576	7	Low
Festival crowd	Real	Real	713	640 × 360	25	High
Bull running festival	Real	Real	596	640 × 360	25	High

Performance Evaluation of Tracking and Surveillance 2009 (PETS2009) dataset is recorded at University of Reading, UK [9]. It includes many scenarios, where each scenario is captured from four different views. Two scenarios are analyzed in the present study. In the first scenario of 107 images, people start walking from the left side until an abnormal event occurs which makes them run away. The second scenario is composed of 378 images. People in this scenario start gathering to the middle until an abnormal event occurs which makes them run away in different directions. The videos of this dataset are challenging as they contain frequent illumination variations.

Festival crowd [38] This video is a real scene of high people density. It records a festival event and shows people who are initially gathered until an abnormal event occurs. This video is challenging as it includes frequent people interactions, obstacles and occlusions.

Bull-running festival [37] This video records a bull-running festival in Spain. In the beginning, it shows people walking, then they start freeing space for the coming bulls. After that, some of the bulls enter in the scene which causes people running. Critical occlusions appear in this video.

3 Method

The proposed approach is composed of four main stages as shown by the block-diagram of Fig. 1. Given the streaming video transmitted by the CCTV camera, the K images of the video are transformed into a grayscaling level. The first step of the proposed method consists of computing the absolute differences $\{D^{(k)}\}_k$ between pairs of successive images $I^{(k)}$ and $I^{(k+1)}$ ($\forall k = 1, \ldots, K-1$). This phase allows to locate the moving edges at each instant. Second, the resulting maps $\{D^{(k)}\}_k$ are transformed in a frequency domain. The obtained coefficients of high absolute values in the frequency domain correspond to the spatial discontinuities within the map $D^{(k)}$. These coefficients allow to reveal the way the moving edges are distributed within a local area. In the same way, local homogeneous regions, such as non moving areas, are represented by coefficients of low absolute values. The absolute values of the coefficients of $D^{(k)}$ are summed, giving rise to $S^{(k)}$. Third, the variation of $S^{(k)}$ along time ($\forall k = 1, \ldots, K-1$) allows to identify whether a remarkable increase, which may be associated to a panic situation, exists. To detect a panic behavior within the set $\{S^{(k)}\}_k$, two alternatives are explored in this study: a clustering based approach and a statistical based approach. Finally, the detection performances are refined by removing false alarms through a postprocessing phase.

3.1 Absolute image differences computation

This step aims to detect moving edges of the objects present in the video by a minimum number of computations. Hence, the absolute difference $D^{(k)}$ between two successive images $I^{(k)}$ and $I^{(k+1)}$ of the video is carried out as:

$$D^{(k)} = |I^{(k+1)} - I^{(k)}|, \quad \forall k = 1, \ldots, K-1 \tag{1}$$

where K is the number of images in the video. The resulting matrices $\{D^{(k)}\}_{k=1,\ldots,K-1}$ locate the moving edges between successive instants. Furthermore, they reveal the spatial distribution of the moving pixels at each instant. This distribution undergoes variations in a certain range during a non panic situation. When a panic occurs, it largely varies due to

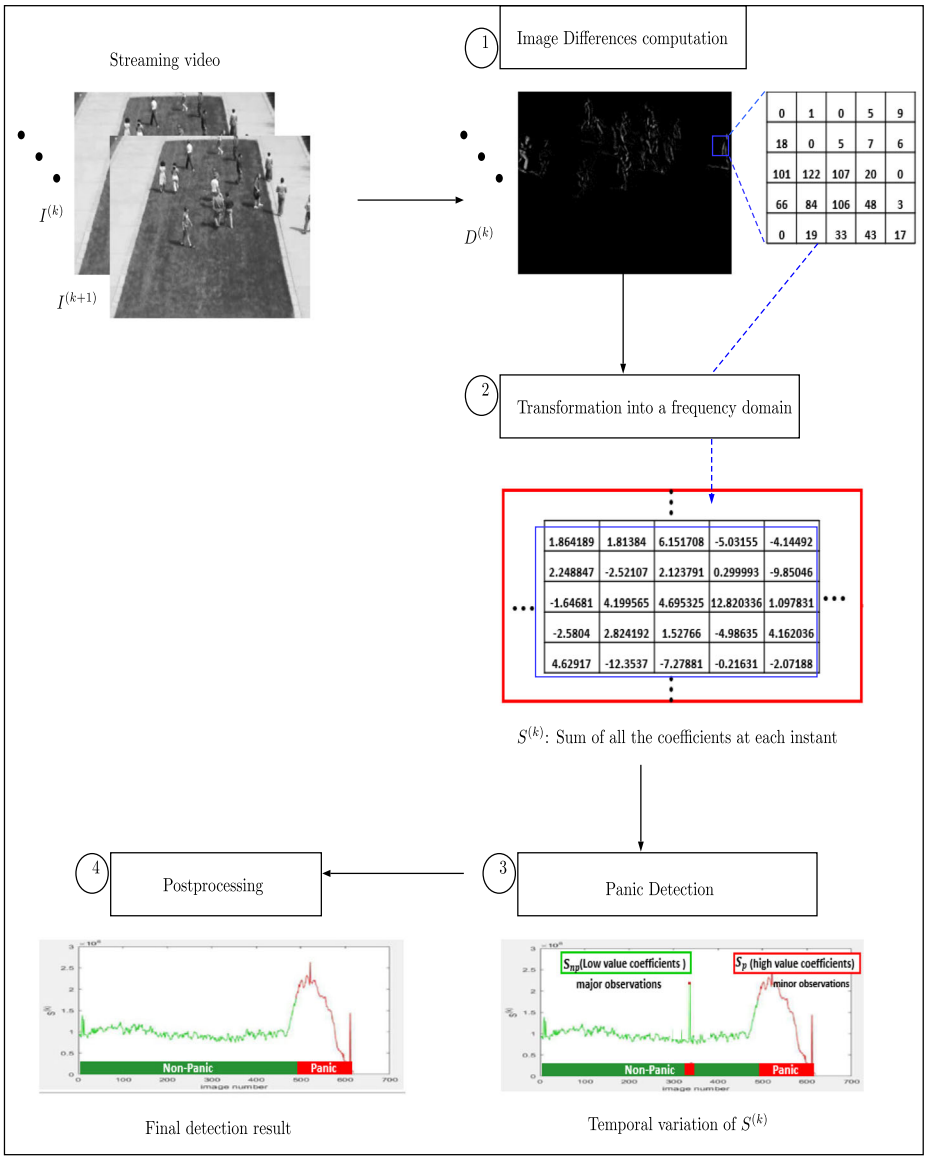


Fig. 1 Block-diagram of the proposed approach

a sudden and a remarkable change in people behavior. An illustration is given in Fig. 2 where the first column depicts an image extracted during a non panic situation along with the corresponding map $D^{(k)}$. The second column shows an image extracted during a panic situation and its related map $D^{(k)}$. The bright pixels in $D^{(k)}$ indicate a high intensity differences between the successive images. It is well noticeable that small variations exist during a non panic situation; whereas in a panic situation, the number of moving pixels increases, and the absolute pixel intensity differences between successive images are higher(shown

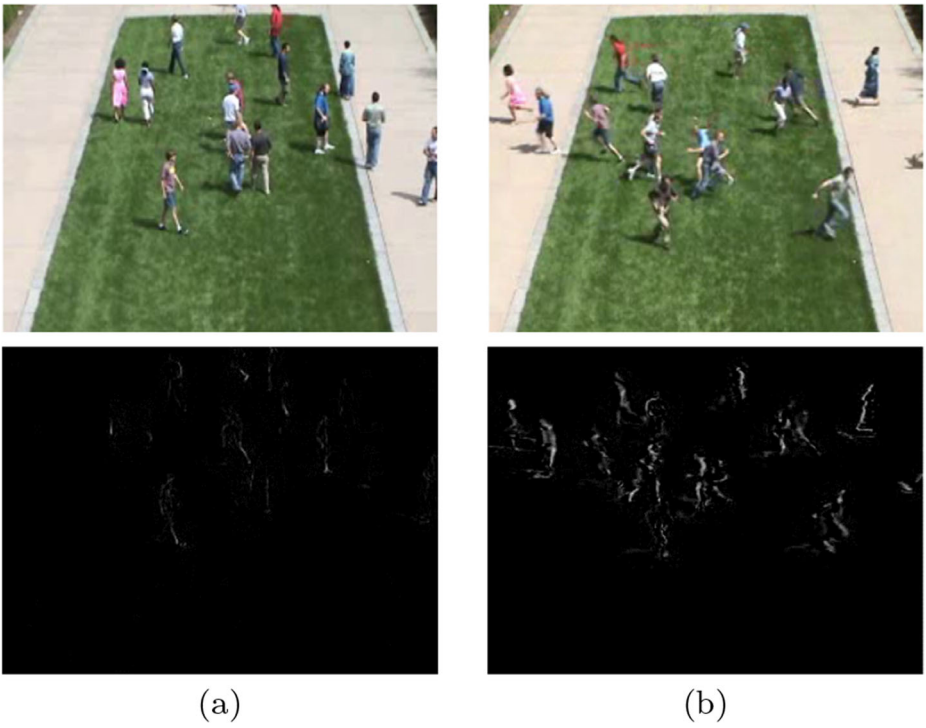


Fig. 2 Distributions of moving pixels during **a** non panic **b** a panic situation

by bright pixels in $D^{(k)}$) as a consequence of a faster change in the characteristics of the pedestrian movements during panic.

In order to detect this remarkable spatial-temporal change, we resort to the analysis of the distribution of moving edges in the frequency domain as explained in the next paragraph.

3.2 Proposed frequency-based feature for the characterization of the crowd dynamics

The aim of this step is to analyze the behavior of the pedestrians at each instant through analyzing the spatial distribution of the corresponding moving edges. For this, the transformation of $D^{(k)}, \forall k = 1, \ldots, K-1$ in a frequency domain is retained for its efficiency to locate discontinuities in one hand, and the sparse representation it offers on the other hand. The FFT [3], the DCT [2] and the DWT [19] are explored in the present work. The transformation $\mathcal{T}_{F_d}(D^{(k)})$ of $D^{(k)}, \forall k = 1, \ldots, K-1$ in a frequency domain $F_d \in \{FFT, DCT, DWT\}$ yields a set of coefficients $\{c^{(k)}\}$ stored in a matrix $C^{(k)}$. The spatial discontinuities in $D^{(k)}$ are transformed into high-magnitude coefficients in the frequency domain, which represent a minority among the whole set $\{c^{(k)}\}$. On the contrary, a majority of low-magnitude coefficients correspond to the local spatial homogeneities as illustrated by the histograms of Fig. 3. Furthermore, the differences in people interactions between the non panic and the panic situation are well highlighted. For instance, in this excerpt, with the same pedestrians present in both situations, the magnitude's range of the set of coefficients $\{c^{(k)}\}$ is larger during a panic than in case of normal behaviors. In addition, the number of

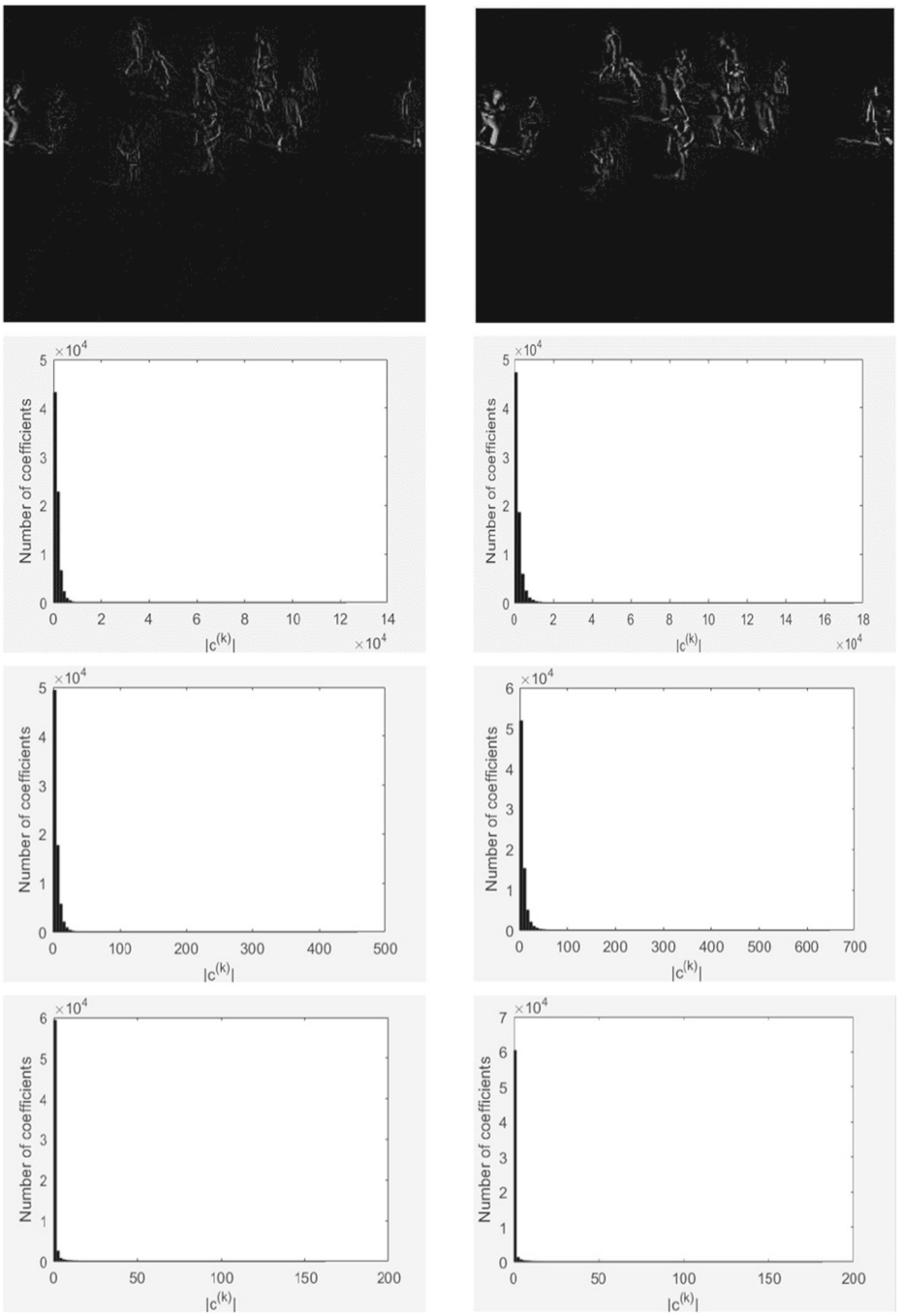


Fig. 3 Distribution of the coefficient'magnitudes during a non-panic (first column) and a panic (second column) situation, using FFT (second row), DCT (third row) and DWT (fourth row)

high-magnitude coefficients during a panic situation is greater than their number during a non panic situation. It is to be noted that the same behavior is observed regardless the chosen frequency domain.

These observations motivated us to propose a new feature $S^{(k)}$ defined for each $D^{(k)}$ by :

$$S^{(k)} = \begin{cases} \sum\limits_{(r,s)\in C^{(k)}} |c^{(k)}(r,s)|, & \text{if } F_d \in \{FFT, DCT\} \\ \sum\limits_{j=1}^{J} \sum\limits_{o=1}^{\mathcal{O}} \sum\limits_{(r,s)\in c^{(k)}_{(j,o)}} |c^{(k)}_{(j,o)}(r,s)|, & \text{if } F_d = DWT. \end{cases} \tag{2}$$

where J is the number of wavelet decomposition levels, $\mathcal{O}$ is the number of orientations at each level and $c^{(k)}_{(j,o)}$ is the wavelet subband at the resolution level $j = 1, \ldots, J$ and the orientation $o = 1, \ldots, \mathcal{O}$. For a dyadic wavelet, $\mathcal{O} = 3$ and $c^{(k)}_{(j,1)}$ denotes the horizontal subband ($o = 1$) at the resolution level j, $c^{(k)}_{(j,2)}$ denotes the vertical subband ($o = 2$) and $c^{(k)}_{(j,3)}$ is the diagonal subband ($o = 3$). As explained later in this paper (Section 4), several dyadic wavelet transforms are experimented with different decomposition levels ranging from $J = 1$ to $J = 3$. It is found that a 1-level decomposition ($J = 1$) yields the best detection performances.

The feature $S^{(k)}$ allows to quantify the discontinuities between moving pixels at each instant. Furthermore, it facilitates the distinction between a non panic and a panic behavior.

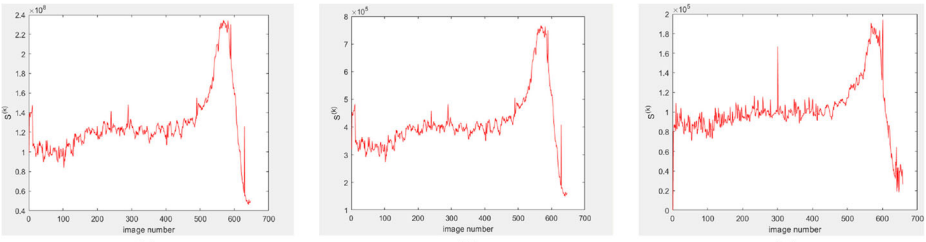
The examination of the temporal variation of $S^{(k)}$ reveals a sudden change in its values when a panic occurs. An illustration of this behavior is depicted in Fig. 4 where the temporal variation of $S^{(k)}$ along the video 9 of the UMN dataset is displayed.

As can be noticed, the values of $S^{(k)}$ vary within the same range until the 551^{st} image where a remarkable jump occurs due to the occurrence of a panic behavior, and lasts for about 120 images. Then, the curve drops when people leave the scene. Another jump is also noticed within the images 301 and 302, when the DWT is applied. However, this peak does not correspond to a panic given its very short duration and is automatically eliminated according to the processing described in Section 3.4.

The next step consists of automatically detecting the high values of $S^{(k)}$ as they reveal the presence of a panic situation.

3.3 Panic detection

Two relevant and distinguishable behaviors are present in a video containing a panic situation : non panic and panic related behaviors. They are reflected by the presence of two



(a) (b) (c)

Fig. 4 Temporal variation of the proposed feature along the video 9 of the UMN dataset by using **a** FFT. **b** DCT. **c** DWT

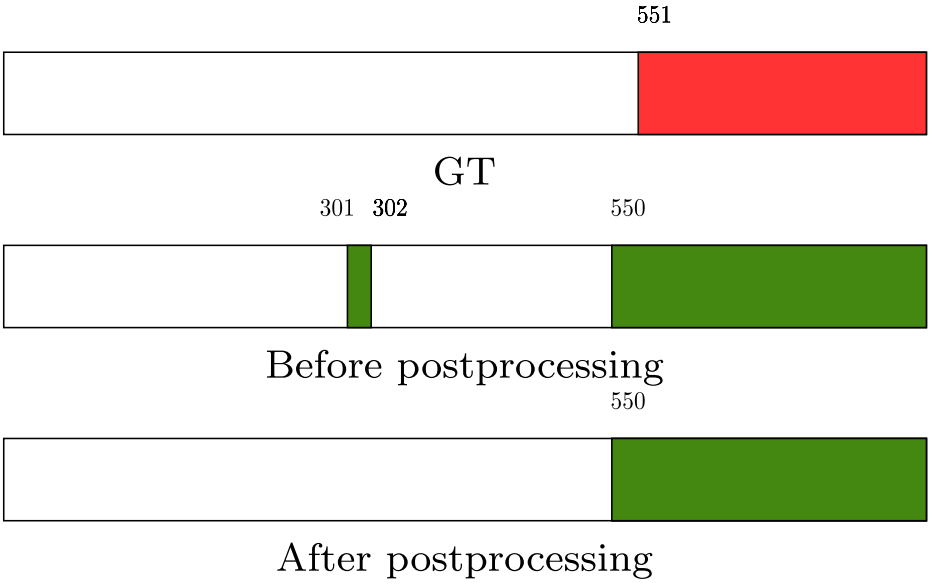


Fig. 5 Detection using MCD before and after postprocessing

classes of values in the set $\mathcal{S} = \{S^{(k)}\}_k$ respectively : low and high values. In this study, we propose to formulate the problem of detecting the high values by using two different formulations. The first one considers classifying the set $\mathcal{S}$ into 2 classes using a clustering technique [16]. The second formulation, proposed in [29], considers the high values as atypical observations that statistically deviate from the distribution followed by the low values. Besides, without loss of generality, the high values are assumed to be a minority within the set $\mathcal{S}$ and hence are considered as outliers, detected thanks to the use of a statistical test for outlier detection [27]. We investigate the two formulations and we compare them in terms of detection performances and execution time.

3.3.1 Clustering based detection

The objective of this step is to differentiate between the data observations in $\mathcal{S}$ that correspond to a panic behavior, from those related to a normal behavior, by using a clustering technique. The idea is to build clusters of data by grouping in each cluster the data points that are as close as possible to each other with respect to a given distance, in one hand. On the other hand, the distance between clusters is required to be as large as possible. To detect the values of $\mathcal{S}$ that correspond to a panic, two clusters have to be identified. The first cluster S_{np} corresponds to the values obtained during a non-panic situation; while the second cluster S_{p} includes high values that are related to a panic situation.

Several clustering techniques are proposed in the literature [10, 11, 20, 28]. The comparison of their performances in detecting panic is conducted in Section 4.

3.3.2 Statistical detection

The aim is to partition $\mathcal{S}$ into two homogeneous subsets: a subset S_{np} of the majority of observations related to a non panic situation and another subset S_{p} containing a minority of

observations of remarkably higher values that are related to a panic situation. Motivated by the characteristics of each subset and the differences between them, we emphasize the possibility of identifying them through a hypothesis testing. More precisely, the observations in S_{p} are considered to be deviating from the statistical distribution followed by the ones in S_{np}. Their detection can therefore be performed following two phases. The first phase aims to estimate the mean and variance of S_{np} by analyzing $\mathcal{S}$ robustly to the presence of the other category of observations (those being part of S_{p}). To this aim, the Minimum Covariance Determinant (MCD) estimator is retained for its efficiency and relatively low computational complexity [27]. The second phase aims to deduce the set S_{p}, given the estimated parameters of the distribution of S_{np}.

1. Parameters estimation: The key idea of MCD to estimate the mean and the variance of $\mathcal{S}$ robustly to the presence of the observations of S_p, is to look for the most concentrated subset in $\mathcal{S}$ of size $h = (1-\alpha)(K-1)$ among h-subsets, given a confidence level $0 < \alpha < 1$. Hence, the observations $s_i \in \mathcal{S}$ are firstly ordered in an increasing order. Then, contiguous h-subsets H_i are built as $H_i = \{s_{(i)}, \ldots, s_{(i+h-1)}\}$. For each subset, the mean and the variance are computed. Then, the most concentrated subset is the one whose variance σ_c^2 is the minimum among the variances of all the subsets H_i. Its mean is denoted by μ_c.
2. Detection of panic related observations: As outlined before, panic related observations have distinguishable values compared to the non panic related ones, and hence are considered as outliers. An observation s_i of $\mathcal{S}$ is considered as an outlier if its distance $d(s_i, \mu_c, \sigma_c)$ from the mean μ_c relatively to σ_c exceeds a tabulated threshold T derived with respect to a confidence level α. This distance is defined by:

$$d(s_i, \mu_c, \sigma_c) = \frac{|s_i - \mu_c|}{\sigma_c}, \forall i \in 1, \ldots, K-1. \tag{3}$$

Hence, the two subsets S_{np} and S_{p} of $\mathcal{S}$ related respectively to non panic and panic situations are deduced by:

$$\begin{aligned} S_{\text{np}} &= \{s_i \in \mathcal{S}; d(s_i, \mu_c, \sigma_c) < T\}, \\ S_{\text{p}} &= \{s_i \in \mathcal{S}; d(s_i, \mu_c, \sigma_c) \geq T\}. \end{aligned} \tag{4}$$

The MCD source code is part of the LIBRA package which is available at https://wis.kuleuven.be/stat/robust/LIBRA/LIBRA-home.

Figure 5 shows the detection result of the statistical test for outlier detection, when applied to the set $\mathcal{S}$ of Fig. 4c. As expected, the images 301 and 302 are considered as being part of the panic images. Other than these images, the panic event is detected earlier by just one image. To improve the detection performances, we propose a postprocessing step that aims to reduce the false detections.

3.4 Postprocessing

The proposed detection technique yields some false detections that should be reduced. To this aim and without loss of generality, the following assumptions are stated:

– A panic behavior cannot happen over less than one second.
– A panic behavior occurs once within a processed video.

The first assumption means that if N successive images are detected as containing a panic behavior and that N is less than the number N^* of images per second (equivalently, the

sampling rate of the video), then, those images are considered as false detections and are discarded from the set of detections. According to the second assumption, it is then possible to identify the sequential number of the image when panic starts. It is the one whose all subsequent images were also identified as anomalous. As depicted in Fig. 5, the result of applying this processing shows the effective elimination of the false detections that are located separately to the sequence of panic images.

4 Results

To evaluate the performances of the proposed technique, four rounds of tests are conducted. The aim of the first round is to select the wavelet parameters that yield the best performances. In the second round, we seek to retain the panic detection method that yields the most accurate results. For this, common clustering techniques as well as the MCD test are confronted. The selection of the most suitable frequency domain is carried out in the third round. Finally, after retaining the appropriate parameters of the system, the detection performances are evaluated with respect to some highly-accurate offline techniques of the literature, and some real-time techniques.
In order to quantify the performances of the tested techniques, the correct detection rate P_c (which is the same as the accuracy), the false detection rate P_f, the precision and the recall are computed. They are respectively defined in terms of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) by:

$$P_c = \frac{TP + TN}{K}, \quad P_f = \frac{FP + FN}{K} \tag{5}$$

$$Recall = \frac{TP}{TP + FN}, \quad Precision = \frac{TP}{TP + FP} \tag{6}$$

The proposed technique is also evaluated in terms of execution time (in number of frames per second (fps)) when it runs on a PC with a 64 bit Core(TM) i7 2.80 GHz CPU, 16 GB RAM and Windows 10. MATLAB 2017 and the WAVELAB library [33] are used for the implementation.

4.1 Wavelet parameters

In this category of tests, we aim to select the most suitable wavelet function and the optimal number of decomposition levels.

Selection of the wavelet function Several wavelet functions exist in the literature [5, 8, 17, 36, 39] and [30]. The performances of the proposed approach are evaluated with respect to some wavelets in order to retain the most suitable one: Haar [36], Beylkin [4], Vaidyanathan [39], Coiflet [8] order 1, Daubechies [17] order 20, Symmlet [5] order 10 and Battle [30] order 5. Figure 6 shows the detection rates when the MCD method is applied. Each result represented in a curve reflects the P_f value (along the x-axis) and the P_c value (along the y-axis) of a specific video in the UMN dataset. The performances are almost the same for all the wavelets except for video 3, where they are degraded when using the Symmlet, the Beylkin and the Battle wavelets. According to these results, the Coiflet wavelet yields the best performances and it is retained for the remaining tests.

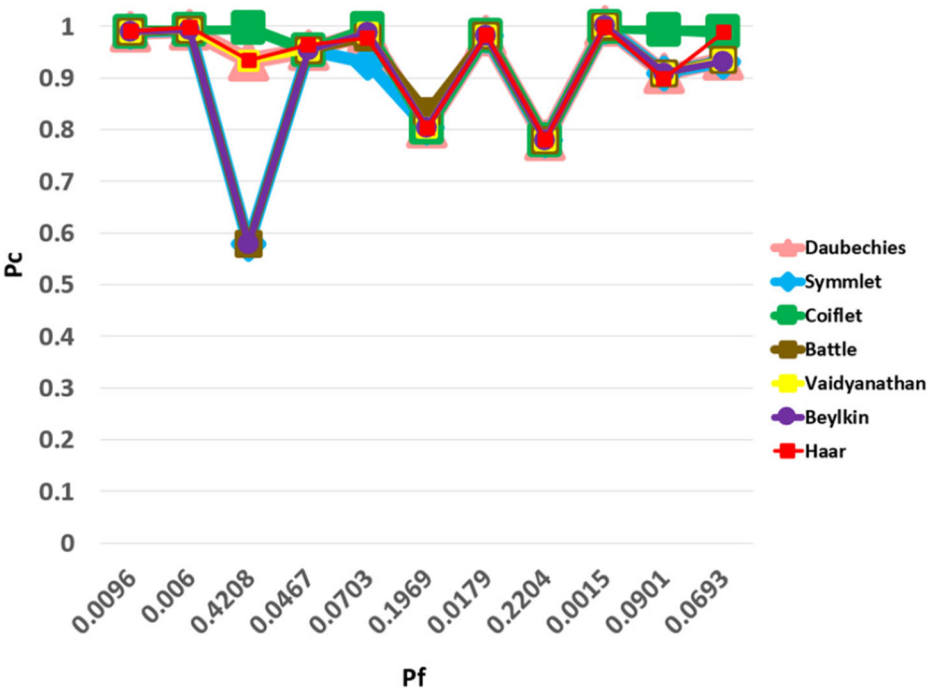
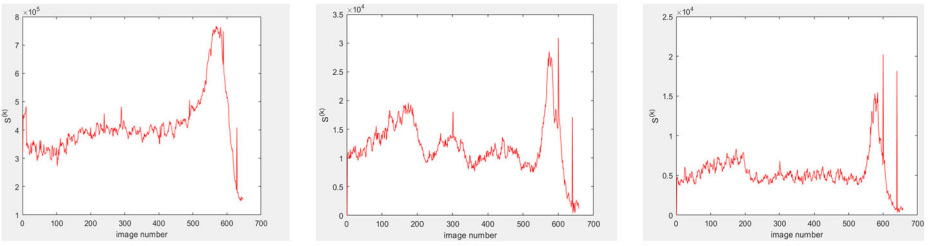


Fig. 6 Detection performances using different wavelet functions and the MCD test

Selection of the number of resolution levels Tests are conducted in order to decide about the optimal number of resolution levels that ensures high detection performances. 1-level, 2-level and 3-level wavelet decompositions are investigated on all the videos of the UMN dataset. The temporal variations of $S^{(k)}$ along the video 9 of the UMN dataset, when $J = 1$, $J = 2$ and $J = 3$ are respectively illustrated in Fig. 7a, b and c. The panic starting at the 551^{st} image according to the ground truth (GT) is visible when $J = 1$ and $J = 3$ and two classes of values can be clearly identified in the set $\mathcal{S}$, unlike the one obtained when $J = 2$.

By applying the same test to all the videos, the detection rates depicted in Fig. 8 show that a 1-level decomposition and a 3-level decomposition yield close performances. Therefore,



(a) (b) (c)

Fig. 7 Temporal variation of the proposed feature according to the number of resolution levels **a** $J = 1$ **b** $J = 2$ **c** $J = 3$. Application to video 9 of the UMN dataset

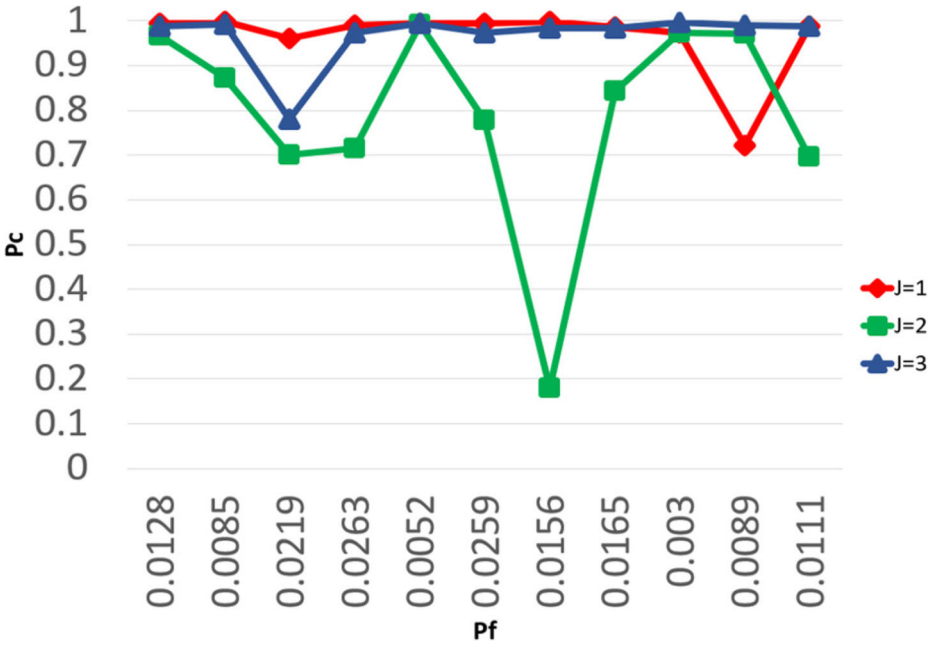


Fig. 8 Comparison of the detection performances when $J = 1$, $J = 2$ and $J = 3$. Application to all the videos of the UMN dataset

a 1-level wavelet decomposition is retained as it requires less computations compared to the 3-level decomposition.

4.2 Selection of the detection technique

In order to select the most appropriate detection technique, different clustering methods such as k-means [11], the Partitioning Around Medoids (PAM) method [16] and skinny-dip [20] are investigated and compared to the MCD statistical test for outlier detection [27]. Furthermore, different values of the confidence level α (0.01, 0.05, 0.1) are considered to evaluate the performances of the system when the MCD test is used. Figure 9 as well as Table 2 show that the performances of the MCD test with $\alpha = 0.01$ outperforms the other methods with an average detection rate of 0.98.

4.3 Comparison between the frequency domains

Using the retained parameters of the system, namely the MCD test with $\alpha = 0.01$, this round of tests aims to select the most appropriate frequency domain in terms of detection performances and execution time. Therefore, FFT, DCT and the DWT using the coiflet wavelet function with one level of decomposition, are explored.

Regarding the execution time, Table 3 shows that the FFT transform yields the fastest execution, followed by the DCT transform and then the DWT. However, for all the data sets, the detection rates obtained using each of the three frequency domains are very close, except for the PETS2009 and MED datasets where the DWT outperforms DCT and FFT. The execution time indicates that the technique operates in real-time although the image

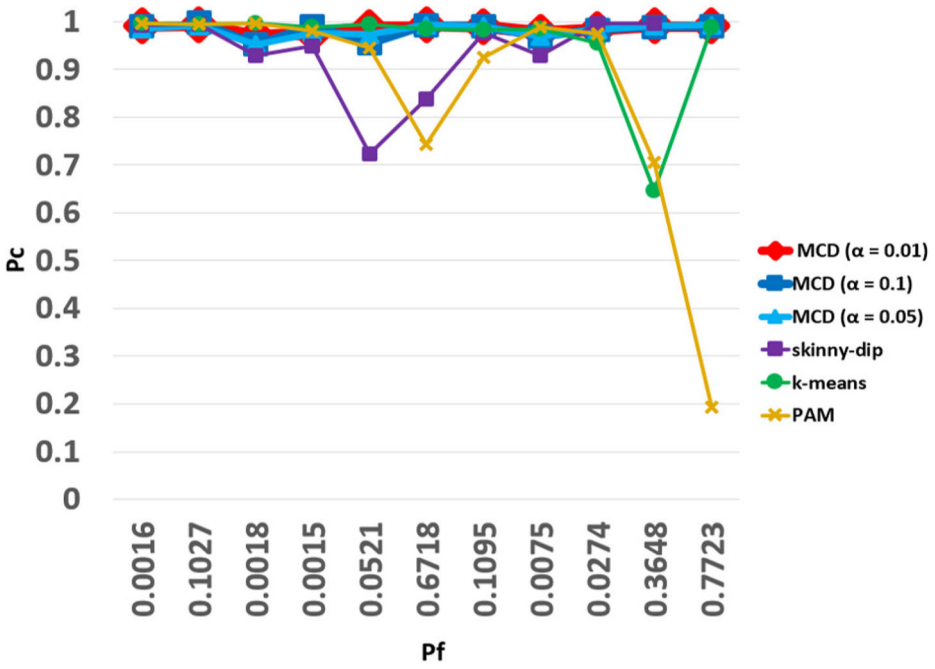


Fig. 9 Comparison between the detection performances obtained with PAM, skinny-dip, k-mean clustering algorithms and the MCD test. Application to all the videos of the UMN dataset

dimensions of the video are larger than in the other data sets. It is worth pointing out that the use of the DWT requires that the dimensions of the images be of the form 2^n where $n \in \mathbb{N}^*$. That is, if this condition is not satisfied, the images are zero-padded. This explains in part the less fast computation of the technique when the DWT is used compared to the use of DCT and FFT.

4.4 Performances evaluation compared to the state-of-the-art techniques

The performances evaluation of the proposed system compared to the state-of-the-art techniques is carried out in two stages. As it is important to maintain a high detection accuracy

Table 2 Comparison between MCD, PAM, k-means and skinny-dip methods based on DWT. Application to the UMN dataset

	P_c	P_f	*Precision*	*Recall*
MCD,$\alpha = 0.1$	0.9779	0.0220	0.9552	0.9390
MCD,$\alpha = 0.05$	0.9793	0.0206	0.9671	0.9267
MCD,$\alpha = 0.01$	**0.9824**	**0.0175**	**0.9804**	**0.9260**
PAM	0.8079	0.1920	0.7292	0.9503
k-means	0.9558	0.0441	0.9099	0.9695
Skinny-dip	0.9326	0.0673	0.9696	0.7835

The values in bold are the best values found compared to the other techniques mentioned in this table

Table 3 Panic detection results based on DWT, DCT and FFT using MCD

Dataset	*Transformation*	P_c	P_f	*Precision*	*Recall*	*Execution time(fps)*
MED	DWT	**0.9447**	**0.0552**	**0.9906**	**0.7977**	7
	DCT	0.9383	0.0616	0.9912	0.7627	24
	FFT	0.9379	0.0620	0.9912	0.7591	**57**
UMN	DWT	0.9824	0.01754	0.9804	0.9260	31
	DCT	0.9864	0.0135	0.9743	0.9671	250
	FFT	**0.9864**	**0.0135**	**0.9743**	**0.9671**	**358**
PETS 2009	DWT	**0.8766**	**0.1234**	**0.9744**	**0.7977**	8
	DCT	0.8114	0.1886	0.9786	0.7202	24
	FFT	0.8114	0.1886	0.9786	0.7202	**64**
Bull running	DWT	0.9530	0.047	1	0.9381	5
	DCT	0.9530	0.047	1	0.9381	55
	FFT	**0.9530**	**0.047**	**1**	**0.9381**	**128**
Festival crowd	DWT	0.9369	0.0631	0.9365	1	7
	DCT	0.9453	0.0547	0.9445	1	45
	FFT	**0.9453**	**0.0547**	**0.9445**	**1**	**124**

The values in bold are the best values found compared to the other techniques mentioned in this table

Table 4 Performances comparison between our approach and offline approach in [12] and [29] using the UMN dataset

Dataset		Off-line techniques				Real-time technique		
UMN		[12]		[29]		Proposed technique		
Video	K	P_f	P_c	P_f	P_c	P_f	P_c	Execution time (fps)
1	625	**0.01**	**0.99**	**0.01**	**0.99**	**0.01**	**0.99**	322
2	828	**0**	**1**	0.01	0.99	0.01	0.99	345
3	549	0.03	0.97	**0.02**	**0.98**	0.03	0.97	315
4	685	0.03	0.97	**0.01**	**0.99**	0.02	0.98	318
5	768	0.03	0.97	**0**	**1**	0.01	0.99	406
6	579	0.03	0.97	**0.01**	**0.99**	**0.01**	**0.99**	366
7	895	0.02	0.98	**0.01**	**0.99**	**0.01**	**0.99**	401
8	667	0.03	0.97	**0.02**	**0.98**	**0.02**	**0.98**	368
9	658	**0**	**1**	**0**	**1**	0.01	0.98	355
10	677	**0.01**	**0.99**	**0.01**	**0.99**	**0.01**	**0.99**	363
11	808	**0.01**	**0.99**	**0.01**	**0.99**	**0.01**	**0.99**	377
Average		0.017	0.982	**0.01**	**0.99**	0.013	0.986	358

The values in bold are the best values found compared to the other techniques mentioned in this table

Table 5 Performances comparison between our approach and offline approach. Application to the PETS2009 dataset

		[7]	[40]	[21]	[41]	Proposed technique
First scenario	view 1	0.37	0.56	0.63	**0.92**	0.75
	view 2	0.37	0.83	0.70	0.83	**0.87**
	view 3	0.37	0.81	0.52	**0.89**	0.75
	view 4	0.37	0.52	0.48	**0.90**	0.75
Average accuracy		0.37	0.68	0.58	**0.88**	0.78
Second scenario	view 1	0.94	0.94	0.91	0.96	**0.99**
	view 2	0.63	0.92	0.89	0.94	**0.99**
	view 3	0.95	0.94	0.94	0.95	**0.99**
	view 4	**0.96**	0.89	0.64	0.91	0.92
Average accuracy		0.87	0.92	0.84	0.94	**0.98**

The values in bold are the best values found compared to the other techniques mentioned in this table

while operating in real-time, the objective of the first stage is to compare the accuracy of the system with some offline techniques [7, 12, 21, 29, 40, 41].

In the second stage, comparisons with real-time techniques [13, 22, 24, 26, 31, 34, 35, 42] are performed in terms of accuracy and execution time.

Table 4 shows that in average, the proposed technique outperforms the technique in [12] and is slightly less accurate than the technique in [29] when tests are performed on the UMN dataset, with an average accuracy of 0.986 against 0.99. These results are considered as excellent since the proposed system operates in real-time with an average computational speed of 358 *fps*.

In the same way, Table 5 depicts the performances obtained on PETS2009 dataset. In average, the proposed technique performs better than [7, 40] and [21] for both scenarios, outperforms the technique [41] for the second scenario and is slightly less accurate than [41] for the first scenario.

Besides, Table 6 shows that the technique we propose outperforms the technique in [25] when the MED dataset is experimented, and for any of the frequency transforms.

Real videos are also tested and the performances are depicted in Table 7 in comparison to the technique of [29]. Good performances are obtained by the proposed system even though they are less accurate than those of [29].

In the second stage, the performances of the proposed system are tested on the UMN data set and compared to related real-time techniques. The results are reported in Table 8 and show the outperforming of the proposed system in terms of both accuracy and execution time, for the three frequency domains.

Table 6 Detection performances on the MED dataset. Comparison between the proposed approach and [25]

Research work	*Accuracy*
[25]	0.7482
Proposed (MCD,DWT)	**0.9447**
Proposed (MCD,DCT)	0.9383
Proposed (MCD,FFT)	0.9379

The value in bold is the best value found compared to the other techniques mentioned in the table

Table 7 Performances comparison between our approach and offline approach [29]

Dataset	[29]		Proposed technique	
	P_f	P_c	P_f	P_c
Bull running	**0.0008**	**0.992**	0.047	0.953
Festival crowd	**0.013**	**0.978**	0.0547	0.9453

The value in bold is the best value found compared to the other techniques mentioned in the table

5 Discussion

The present study describes a new real-time approach for the detection of panic behaviors in crowded scenes. Three main contributions are proposed for which, efficiency, accuracy and high speed are experimentally proved. The first contribution aims to alleviate the heavy computations resulting from applying a motion estimation technique, by considering the differences between successive images of the video. This solution allows to locate moving edges with a fast execution. Furthermore, panic is defined as a sudden change in the interactions between people. This is reflected by a change in the spatial distribution of the moving edges in addition to the increase of the number of moving pixels as a consequence of the fast behavior's change of people. In order to characterize the distribution of moving pixels during a panic and a normal situation, our second contribution consists of representing the moving edges in a frequency domain, allowing a sparse representation of the spatial discontinuities. The FFT, the DCT and the DWT are explored in the present study. Tests conducted on several challenging videos, with different density levels of pedestrians, show the high performances and the high speed of the proposed system as depicted in Tables 3, 4, 5, 6, 7 and 8. The experimental comparison between the three frequency domains in terms of performances show that they perform well and that the detection rates are close. In terms of execution time, the FFT based system yields the highest execution speed, followed by the DCT, then the DWT.

Table 8 Comparison in terms of accuracy and execution time between the reported real-time detection techniques and the proposed technique. Application to UMN dataset

Real-time techniques	*Execution time (fps)*	*Accuracy*
[31]	9	0.89
[42]	20	Not mentioned
[35]	20	Not mentioned
[26]	Not mentioned	0.85
[24]	20	0.95
[34]	30	0.85
[13]	5	Not mentioned
[22]	25	0.98
Proposed (MCD,DWT)	31	0.98
Proposed (MCD,DCT)	250	0.99
Proposed (MCD,FFT)	**358**	**0.99**

The value in bold is the best value found compared to the other techniques mentioned in the table

Our third contribution refers to the detection of panic related data by exploring two formulations. The first formulation considers distinguishing between the normal-related data and the panic-related data by using a clustering technique; whereas the second formulation is based on a hypothesis testing, in which panic-related data are considered as aberrant compared to the data resulting from a normal situation. Figure 9 and Table 2 illustrate the good performances of the system for both formulations, with a slight outperforming of the second formulation.

The proposed system is evaluated with regard to offline and real-time detection techniques and the results show its high performances.

6 Conclusion and future work

A new panic detection approach is proposed in this study. The aim is to analyze the crowd dynamics and detect a possible panic behavior in real-time and without requiring a prior knowledge about the video under consideration. For this, a new feature is proposed based on the computation of the image differences and the analysis of the moving edges in frequency domains. Then, two formulations of the panic detection problem are explored and compared in terms of accuracy and execution time. The approach is evaluated using several datasets and showed its high performances.

In the future work, we will study the effectiveness of other solutions for the detection of moving edges, such as the foreground extraction, and their impact on the system performances.

Acknowledgments This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant No. (DG-046-612-1140). The authors, therefore, gratefully acknowledge the DSR technical and financial support.

References

1. Agency TOSP (2015) Hajj / civil defense: 150 pilgrims died and 400 others injured in a stampede in Mina. http://www.spa.gov.sa
2. Ahmed N, Natarajan T, Rao KR (1974) Discrete cosine transform. IEEE Trans Comput 100(1):90–93
3. Bergland GD (1969) A guided tour of the fast Fourier transform. IEEE spectrum 6(7):41–52
4. Bleistein N (1987) On the imaging of reflectors in the earth. Geophysics 52(7):931–942
5. Cai C, CaiHarrington PDB (1998) Different discrete wavelet transforms applied to denoising analytical data. Journal of chemical information and computer sciences 38(6):1161–1170
6. Catherine ES, Shoichet E, Botelho G (2016) Footage shows suspects in Brussels attack. http://edition.cnn.com/2016/03/22/europe/brussels-explosions
7. Chen D-Y, Huang P-C (2011) Motion-based unusual event detection in human crowds. J Vis Commun Image Represent 22(2):178–186
8. Daubechies I et al (1991) Ten lectures on wavelets. In: CBMS-NSF regional conference series in applied mathematics, vol 61, no. 4
9. Ferryman JA (2009) Pets2009benchmarkdata. http://www.cvg.reading.ac.uk/PETS2009/a.html
10. Firdaus S, Uddin MA (2015) A survey on clustering algorithms and complexity analysis. Int J Comput Scie Issues (IJCSI) 12(2):62
11. Forgy EW (1965) Cluster analysis of multivariate data: efficiency versus interpretability of classifications. Biometrics 21:768–769
12. Fradi H, Dugelay J-L (2015) Towards crowd density-aware video surveillance applications. Inf Fus 24:3–15
13. Fradi H, Luvison B, Pham Q-C (2017) Crowd behavior analysis using local mid-level visual descriptors. IEEE Trans Circuits Syst Video Technol 27(3):589–602

14. Guardian T (2017) More than a dozen fans killed in stampede at Angolan football match. https://www.theguardian.com/world/2017/feb/10/17-fans-killed-stampede-football-match-angola
15. Gunduz AE, Ongun C, Temizel TT, Temizel A (2016) Density aware anomaly detection in crowded scenes. IET Computer Vision 10(5):376–383
16. Kaufman L, Rousseeuw PJ (2009) Finding groups in data: An introduction to cluster analysis, vol 344. Wiley, New York
17. Lewis AS, Knowles G (1991) VLSI architecture for 2D Daubechies wavelet transform without multipliers. Elect Lett 27(2):171–173
18. Li T, Chang H, Wang M, Ni B, Hong R, Yan S (2015) Crowded scene analysis: A survey. IEEE trans Circ Syst Video Technol 25(3):367–386
19. Mallat S (1999) A wavelet tour of signal processing, Academic Press, San Diego
20. Maurus S, Plant C (2016) Skinny-dip: Clustering in a sea of noise. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 1055–1064
21. Mehran R, Oyama A, Shah M (2009) Abnormal crowd behavior detection using social force model. In: IEEE conference on computer vision and pattern recognition, 2009. CVPR 2009. IEEE, pp 935–942
22. Nady A, Atia A, Abutabl AE (2018) Real-time abnormal event detection in crowded scenes. J Theo Appl Inf Technol 96:6064–6075
23. University of Minnesota (2006) Unusual crowd activity dataset. http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi
24. Pennisi A, Bloisi DD, Iocchi L (2016) Online real-time crowd behavior detection in video sequences. Comput Vis Image Und 144:166–176
25. Rabiee H, Haddadnia J, Mousavi H, Kalantarzadeh M, Nabi M, Murino V (2016) Novel dataset for fine-grained abnormal behavior understanding in crowd. In: 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, pp 95–101
26. Roshtkhari MJ, Levine MD (2013) An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions. Comput Vis Image Und 117(10):1436–1452
27. Rousseeuw PJ, Driessen KV (1999) A fast algorithm for the minimum covariance determinant estimator. Technometrics 41(3):212–223
28. Sajana T, Rani CS, Narayana K (2016) A survey on clustering techniques for big data mining. Ind J Sci Technol 9(3):14721835
29. Shehab D, Ammar H (2018) Statistical detection of a panic behavior in crowded scenes. Mach Vis Appl 30:1–13
30. Sheng Y, Roberge D, Szu HH (1992) Optical wavelet transform. Opt Eng 31(9):1840–1846
31. Shi Y, Gao Y, Wang R (2010) Real-time abnormal event detection in complicated scenes. In: 2010 20th international conference on pattern recognition (ICPR). IEEE, pp 3653–3656
32. Thida M, Yong YL, Climent-Pérez P, Eng H-L, Remagnino P (2013) A literature review on video analytics of crowded scenes. Springer, Berlin, pp 17–36
33. Stanford University (2016) WAVELAB 850. http://statweb.stanford.edu/~wavelab/
34. Wang J, Xu Z (2016) Spatio-temporal texture modelling for real-time crowd anomaly detection. Comput Vis Image Und 144:177–187
35. Wang L, Dong M (2012) Real-time detection of abnormal crowd behavior using a matrix approximation-based approach. In: 2012 19th IEEE international conference on image processing (ICIP), pp 2701–2704
36. Wang Q, Deng X (1999) Damage detection with spatial wavelets. Int J Solids Struct 36(23):3443–3468
37. waze digital (1966) cloud digital asset management platform. http://commerce.wazeedigital.com/license/clip/14121797.do
38. waze digital (2001) cloud digital asset management platform. http://commerce.wazeedigital.com/license/clip/3682865.do
39. Wickerhauser MV (1996) Adapted wavelet analysis: from theory to software. AK Peters/CRC Press
40. Wu S, Moore BE, Shah M (2010) Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. In: 2010 IEEE computer society conference on computer vision and pattern recognition, pp 2054–2060
41. Wu S, Wong H-S, Yu Z (2014) A Bayesian model for crowd escape behavior detection. IEEE Trans Circ Syst Video Technol 24(1):85–98
42. Xiong G, Wu X, Chen Y-L, Ou Y (2011) Abnormal crowd behavior detection based on the energy model. In: 2011 IEEE international conference on information and automation (ICIA), pp 495–500
43. Zhan B, Monekosso DN, Remagnino P, Velastin SA, Xu L-Q (2008) Crowd analysis: A survey. Mach Vis Appl 19(5):345–357. https://doi.org/10.1007/s00138-008-0132-4

Bahya Aldissi received her Master's degree in Information Technology in 2019 from the Faculty of Computing and Information Technology, King Abdulaziz University, Saudi Arabia and her Bachelor's degree in InformationTechnology in 2012 from the Faculty of Computing and Information Technology, King Abdulaziz University, Saudi Arabia. She is currently a Teaching Assistant at the College of Computer Science and Engineering, University of Jeddah. Her research interests include Computer Vision and Video Processing.

Heyfa Ammar is an Associate Professor in Computing and Information Technology. She is also a research member in Robotics-Informatics and Complex Systems (RISCENIT) research laboratory, at the national engineering school of Tunis (University of Tunis ElManar, Tunisia). Her research interests include video and image processing. She received her Ph.D degree in Information Technology and Communications in 2012 from the Higher School of Communications of Tunis (SupCom), her Master's degree in Mathematical engineering from the school of polytechnics of Tunisia in 2005 and her Engineer degree in computer science from the National School of computer sciences (ENSI, Tunisia) in 2002. She also worked as a computer engineer in the R&D department of Alcatel (now Nokia).