An online graph-based anomalous change detection strategy for unsupervised video surveillance

Jongwon KIM and Jeongho CHO*

Abstract

Due to various accidents and crime threats to an unspecified number of people, many surveillance technologies have been studied as an interest in individual security continues to increase throughout society. In particular, intelligent video surveillance technology is one of the most active research areas in the field of surveillance; this popularity has been spurred by recent advances in computer vision/image processing and machine learning. The main goal is to automatically detect, recognize, and analyze objects of interest from collected sensor information and then efficiently extract/utilize this useful information, such as by detecting abnormal events or intruders and recognizing objects. Anomalous event detection is a key component of security, and many existing anomaly detection algorithms rely on a foreground subtraction process to detect changes in the foreground scene. By comparing input image frames with a reference image, changed areas of the image can be efficiently detected. However, this technique can be insensitive to static changes and has difficulties in noisy environments since it depends on a reference image. We propose a new strategy for improved dynamic/static change detection that complements the weak points of existing detection methods, which have low robustness in noisy environments. To achieve this goal, we employed a self-organizing map (SOM) for data clustering and regarded the cluster distribution of neurons, represented by the weight of the optimized SOM, as a directed graph problem. We then applied the shortest path algorithm to recognize anomalous events. The real-time monitoring capability of the proposed change detection system was verified by applying it to self-produced test data and the CDnet-2014 dataset. This system showed robustness against noise that was superior to other surveillance systems in various environments.

Keywords: Video surveillance, Self-organizing maps, Shortest path problem, Anomalous event, Change detection

1 Introduction

Governments and institutions are currently paying a great deal of money to solve and prevent crime and terrorism threats. However, while efforts are being made to develop surveillance technologies and deploy monitoring equipment and personnel, the occurrence of safety-related incidents continues to increase demand for advanced and efficient equipment to improve safety [1]. For this reason, industries and technologies related to anomaly detection, including intruder detection, image change detection, and surveillance systems, are continuously growing. Many of the surveillance systems that are currently available commercially include analog surveillance systems, based on sensor information such as infrared rays and ultrasonic waves, and digital surveillance systems that are centrally managed through servers based on similar sensors and CCTV (closed-circuit television) image information. However, these monitoring systems can have significant consequences if a single misjudgment is made. Therefore, robustness and operational efficiency in various environmental conditions are required. Unfortunately, existing commercial monitoring systems have some disadvantages.

Sensors used in analog surveillance systems are divided into passive sensors and active sensors. Passive sensors detect energy emitted from nature and have the disadvantage of being overly sensitive to heat radiation

* Correspondence: jcho@sch.ac.kr
Department of Electrical Engineering, Soonchunhyang University, Asan
31538, Republic of Korea

Fig. 1 Lattice structure of the SOM
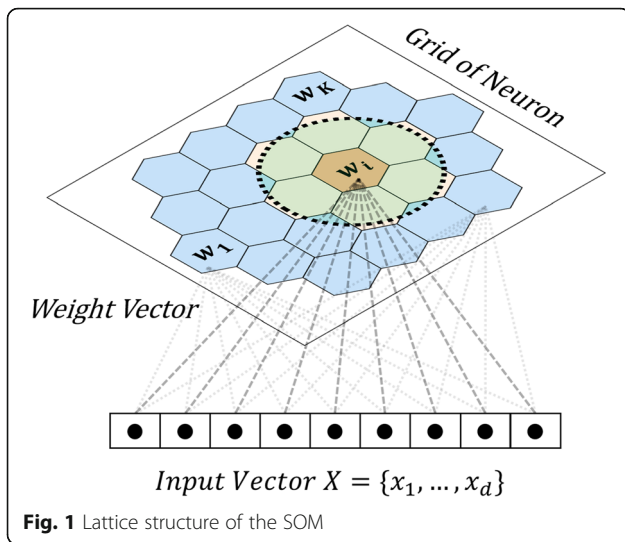
from people, the ambient temperature, and the presence of the sun. Alternatively, active sensors, which have their own energy source for illumination, are used to detect reflected radiation after it is emitted. These have higher reliability but their performance can be deteriorated depending on the terrain of the installation site. In addition, a higher energy source is required compared to passive sensors. Also, as the desired range of illumination increases, astronomical costs are incurred, which leads to operational efficiency problems [2, 3]. A digital surveillance system using an unmanned security system or CCTV to improve the vulnerability of an analog surveillance system is advantageous in that it is highly efficient and free from restrictions related to the installation environment, as compared to existing surveillance systems using a single analog sensor. However, in the case of a centralized supervisory surveillance system consisting of numerous CCTVs, an administrator is still required; thus, this setup is still limited in that it relies heavily on the concentration and judgment of an

administrator; problems related to carelessness or misjudgment can still occur [4, 5].

In addition to the aforementioned surveillance systems, intelligent surveillance systems, which use monitoring algorithms rather than simple surveillance based on video information obtained from CCTVs, have been attracting attention due to recent advancements in computer vision/image processing and the remarkable development of machine learning. The primary role of such a system is to automatically identify and analyze objects of interest and efficiently extract and provide interpretations of scenes, such as abnormal event monitoring, intruder detection, and object recognition, from the collected sensor information. Here, recognition of an abnormal event is made by checking the state difference of an object or phenomenon based on observations over time [6]. Recognition of objects or events requires vast amounts of information that span multiple areas of space and time, and most recognition methods rely on an image subtraction process that detects changes from foreground scenes. Another way to effectively recognize and classify objects is based on machine learning algorithms, such as convolutional neural networks (CNNs) and support vector machines (SVMs) [7–9]. The technologies applied to intelligent surveillance systems allow for very good active judgment and cognitive ability compared to existing surveillance techniques.

Anomalous change detection for the implementation of intelligent surveillance systems is one of the most challenging and long-standing tasks in computer vision [10, 11]. Various simple methods for the implementation of intelligent surveillance systems have been proposed to detect changes in an image. These include using global illumination based on a single grayscale/color image without a moving object and using a median filter based on a temporal image filter [12, 13]. The use of Gaussian mixture models (e.g., WrenGA and Grimson GMM [14, 15]) has been proposed to describe the background of animated
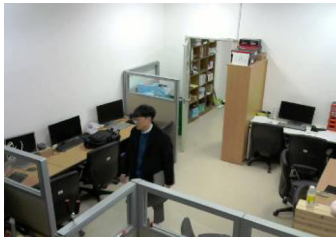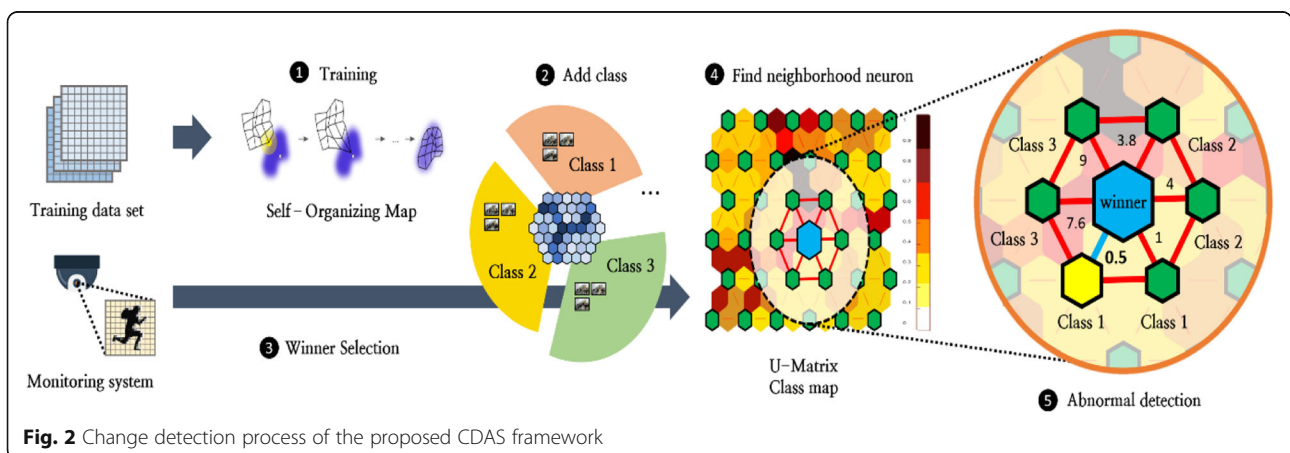

Fig. 2 Change detection process of the proposed CDAS framework

Table 1 Sample images classified into three levels

Training dataset (scenario)			
Level	Class 1 (normal)	Class 2 (static abnormal)	Class 3 (dynamic abnormal)
	Negative	Positive	Positive
	No people	Little movement of objects or people	Objects or people moving constantly

textures, and a high-performance background subtraction algorithm using a deep convolution network (e.g., FeSeg-Net and DeepBS [16, 17]) has also been proposed recently. These intelligent surveillance methods provide better performance than analog/digital surveillance systems, but there are still some drawbacks. Methods based on image subtraction are relatively simple to implement but very susceptible to abnormal situations, such as static/dynamic changes, image brightness changes, and noise, because they depend on a reference image [18–20]. Alternatively, for machine learning-based detection algorithms, high-performance hardware is required due to the demand for high levels of computation ability in the learning process, and a vast amount and wide variety of data learning processes are also involved [21–23]. In addition, it is difficult to detect static abnormalities when tracking a moving object based on pixel changes, and this technique may be vulnerable when there is noise or a change in shading. An intelligent surveillance system guarantees superior surveillance performance; however, because it requires a lot of

time and cost for operation, the availability and efficiency in various fields are limited and commercialization has progressed slowly.

In this paper, we propose a new intelligent anomaly detection system to complement the low robustness of existing surveillance systems for detecting static anomalies in various environments. Our proposed system also addresses cost-efficiency problems. To achieve this, we consider a clustering method that can learn the topology and distribution of the input, exhibit robustness to noise, and classify the entire scene as an attribute. To implement the proposed change detection architecture, we employ a representative data-clustering technique, i.e., a self-organizing map (SOM); SOMs have been applied in various fields, such as data visualization, process monitoring, and data analysis [24, 25]. After training the SOM, we classify abnormalities by analyzing the cluster distribution of the neurons, which are represented by the connection weights, of the optimized SOM. The weights associated with neurons are regarded as
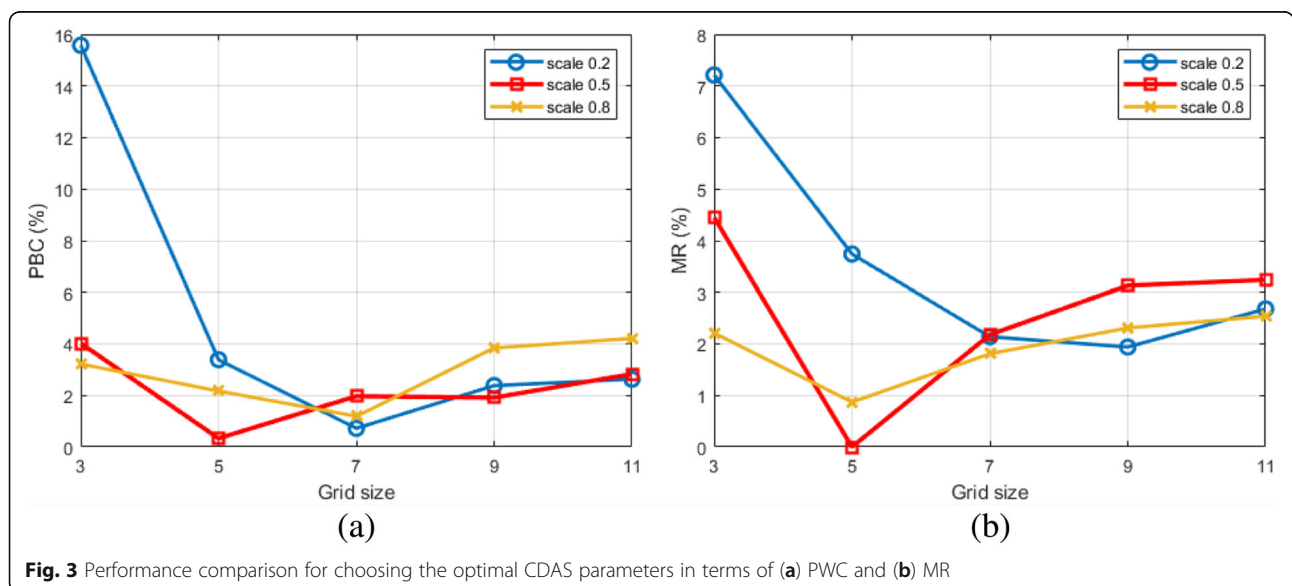

Fig. 3 Performance comparison for choosing the optimal CDAS parameters in terms of (**a**) PWC and (**b**) MR

Table 2 Comparison results with the hand-labeled dataset

Method	Frame Diff	Global Ilu	MF	Grimson GMM	Zivkovic GMM	SC-SOBS	CDAS
FNR	0.024	0.000	0.031	0.074	0.086	0.013	0.010
FPR	0.813	0.034	0.706	0.812	0.719	0.000	0.000
PWC (%)	33.751	0.316	19.409	17.693	16.848	0.422	0.331
FAR (%)	10.000	0.000	22.913	64.214	68.823	2.432	3.534
MR (%)	31.274	0.344	19.108	14.243	13.753	0.010	0.000

directional graph problems, and the shortest path algorithm is applied to determine the abnormality. In the test phase, the proposed algorithm finds the winner neuron in the image that has been newly input to the SOM and searches for the neuron class located in the shortest path to determine whether anomalies are detected based on dynamic environmental changes.

We confirmed the real-time monitoring capability of the proposed change detection system by using self-produced test data in an indoor environment. We also used the CDnet-2014 dataset, which is a change detection benchmark dataset. This system showed superior robustness in various environments, compared with other surveillance systems. Therefore, we expect that the proposed system can be used for practical applications [26].

2 Methods

2.1 Training of the SOM for image clustering

Self-organizing maps (SOMs) are unsupervised data clustering algorithms inspired by human cognitive processes and neurological conditions. They have been applied in various fields along with the development of data clustering techniques [27–29]. The SOM is characterized by carrying out dimension reduction of high-dimensional input vectors and data clustering simultaneously. This is done through the winner-take-all learning mechanism and visualization in a two-dimensional form by extracting feature points of complex and non-linear data [30]. Generally, the SOM is composed of a two-dimensional lattice structure, and each neuron on the grid consists of a weight vector. The Euclidean distance between neurons i and $i^{'}$, which are the positions of neurons (r_1^i, r_2^i) and $(r_1^{i^{'}}, r_2^{i^{'}})$, respectively, is given by Eq. (1).

$$\text{dist}\left(i, i^{'}\right) = \left\| \left(r_1^i, r_2^i\right) - \left(r_1^{i^{'}}, r_2^{i^{'}}\right) \right\| \tag{1}$$

Each neuron i is connected to a prototype weight vector $W_i = \{w_{i1}, ..., w_{id}\}$, which represents a cluster of input vectors. Here, d is the dimension of the input vectors, and the number of neurons on the lattice can be structurally extended by several hundred (or more) depending on the complexity of the data [31, 32]. The lattice structure of the SOM is shown in Fig. 1. Here, the input vector is rearranged as a one-dimensional vector $X \in \mathbb{R}^{d \times 1}$, compressed into a single gray point, and normalized from the RGB image data ($N_{\text{row}} \times N_{\text{col}}$ pixels) collected from the surveillance camera.

The learning of the SOM proceeds in an iterative way in the direction of optimizing the connection strength with the neurons. The topological distance between the input vector X and the weight vector W_i of each neuron is obtained in each epoch t. At this instance, the neuron closest to the input vector is the winning neuron, which is also defined as the best matching unit (BMU). This is selected through the competitive learning process, as described in Eq. (2) [33].

$$\text{BMU(X)} = \underset{i \in \{1, ..., K\}}{\text{argmin}} \|X - W_i(t)\| \tag{2}$$

Then, all the weight vectors around the BMU are adjusted for $i \in \{1, ..., K\}$ as follows:

$$\begin{aligned} W_i(t+1) = {} & W_i(t) \\ & + \tau(t)\delta_{\text{BMU}(X),i}\left(X - W_{\text{BMU}(X)}(t)\right) \end{aligned} \tag{3}$$

Here, $\tau(t)$ is a decaying learning rate that determines the learning speed. Additionally, $\delta_{(\text{BMU}(X), i)}$, which is defined as a neighborhood function, adjusts the connection

Table 3 Comparison results with the hand-labeled dataset in a noisy environment

Method	Frame Diff	Global Ilu	MF	Grimson GMM	Zivkovic GMM	SC-SOBS	CDAS
FNR	0.043	0.056	0.064	0.036	0.034	0.024	0.000
FPR	0.823	0.047	0.672	0.796	0.794	0.831	0.163
PWC (%)	42.891	5.023	13.520	30.894	30.184	38.540	7.995
FAR (%)	28.530	51.478	56.184	21.766	22.214	31.253	0.000
MR (%)	33.351	0.186	9.076	31.94	31.85	10.150	8.834

Table 4 Comparison results with the hand-labeled dataset when the brightness changes

Method	Frame Diff	Global Ilu	MF	Grimson GMM	Zivkovic GMM	SC-SOBS	CDAS
FNR	0.084	0.043	0.087	0.063	0.068	0.043	0.000
FPR	0.895	0.037	0.788	0.724	0.716	0.823	0.289
PWC (%)	44.723	4.164	15.790	20.186	21.013	29.531	3.694
FAR (%)	48.244	42.659	74.126	17.689	16.871	24.584	2.065
MR (%)	44.366	0.158	9.723	28.847	27.980	7.244	3.861

strength with neighborhood neurons around the BMU into a concave Gaussian filter-type scalar, as described in Eq. (4).

$$\delta_{\text{BMU}(X),i} = \exp\left(-\frac{\text{dist}(i, \text{BMU})^2}{2\sigma^2(t)}\right) \tag{4}$$

Here, $\sigma^2(t)$ is the decaying variance representing the radius of the neighborhood function. As a result, the SOM is optimized by adjusting the connection strength around the BMU and its neighbors through the competitive learning between the input vector and the neuron.

2.2 Image change detection based on the SOM

By using this learning mechanism and assigning a class to the two-dimensional lattice structure based on the clusters of neurons, classification and prediction problems can be solved by the competing process of new inputs [34–36]. Considering this, the proposed change detection architecture based on the SOM (referred to as CDAS) learns the image data collected from the surveillance camera based on the SOM and assigns classes to the cluster characteristics of optimized neurons. Next, the BMU is obtained through the input of the image propagated in real-time, and the neuron closest to the BMU is defined as the nearest neighbor (NN). To account for the cluster distribution characteristics of the optimized grid structure, the mutual distance to the BMU is regarded as the shortest path search problem. This uses the weighted directed graph based on the connection strength with neighbors, as defined

in the unified distance matrix (U-Matrix), which is a visualization technique used with SOMs [37].

Consider a directed graph $G = (A, E)$ consisting of a finite set of nodes (the position of neurons) $A \in \{1, ..., K\}$ and E, denoting a set of m_{ij} directed edges, which represent the connection strength of a neighboring neuron pair (i, j) based on the U-Matrix. Additionally, two nodes, i.e., an origin node (BMU) $s \in A$ and a destination node (target) $t \in A$, which represents neurons containing class information, are specified. Then, the shortest path problem $SP_{s \to t}$ is applied to find the path with the minimum total distance from s to t, as described in Eq. (5) [38, 39].

$$SP_{s \to t} = \min \sum_{(i,j) \in A} m_{ij}(\text{dist}(i, j)) \tag{5}$$

The distance between the finally obtained BMU and the neuron cluster is obtained by calculating $SP_{s \to t}$, and the detection of the image change can be found through the class information of the SOM. This makes it possible to reduce the error probability of anomaly detection that occurs in existing anomaly detection schemes. We can construct a real-time anomaly detection system with improved monitoring robustness and efficiency in a noisy environment. Figure 2 illustrates the change detection process of the proposed CDAS. In summary, the proposed CDAS reduces the detection error probability in a real operating environment. It can be operated at a relatively low cost and improve the robustness and efficiency compared to conventional surveillance systems.

Table 5 Sample images classified into two levels

Training dataset (Scenario)		
Level	Class 1 (normal)	Class 3 (dynamic abnormal)
	Negative	Positive
	No Change	Objects or people moving constantly

Table 6 Comparisons of FNR of the competing methods over tested scenarios

Scenario	Frame Diff	Global Ilu	MF	Grimson GMM	Zivkovic GMM	SC-SOBS	CDAS
Office	0.071	0.013	0.038	0.136	0.109	0.014	0.023
Fall	0.878	0.858	0.903	0.866	0.897	0.658	0.650
Traffic	0.711	0.353	0.409	0.401	0.200	0.198	0.137
Over pass	0.613	0.606	0.679	0.508	0.497	0.181	0.412

3 Results and discussion

In order to evaluate the capability of the proposed CDAS, anomaly detection tests were conducted based on two datasets: hand-labeled data for an indoor environment and the public CDnet-2014 (change detection) benchmark dataset. It has been confirmed that the static abnormal condition can be detected by the CDAS. It also operates normally, even in the presence of noise. Qualitative results and future directions are discussed below.

3.1 Performance evaluation with the hand-labeled dataset

First, data in an indoor office environment, which are generally applicable to various indoor and outdoor environments requiring surveillance, were produced and tested to evaluate the proposed surveillance system. A total of 3600 images were captured at a resolution of 1280 × 720 and 30 frame/s using a Logitech HD webcam in an indoor space (8 m × 6 m). The collected video images were classified as one of three image change detection levels depending on static changes, such as a door opening/closing, the number of subjects present in the image, and the degree of motion. The levels include the normal state with no motion (negative/class 1), a static abnormal state with minimal static environment change (positive/class 2), and a dynamic abnormal state with a large dynamic change (positive/class 3). Examples of images from the indoor office environment, classified into these three levels are illustrated in Table 1.

For training the SOM, we randomly extracted 120 images from the 3600 single images that were collected. Performance evaluation was carried out using four video scenarios, which contained all classes with various static and dynamic changes, in order to avoid the possibility of duplication with learning data. We also evaluated the robustness of surveillance on continuous images. The detection of change, which can be interpreted as a binary classification, is regarded as a classification problem by determining a normal state (negative = 0) in which there is no change and an abnormal state (positive = 1) in which a change in the static or dynamic environment is detected. For this binary classifier problem, we have considered the binary classification performance metrics, listed in Eq. (6)–(10), as recommended in the CDnet-2014 motion detection benchmark; they are false negative rate (FNR), false positive rate (FPR), percentage of wrong classification (PWC), false alarm rate (FAR), and missing rate (MR).

$$FNR = \frac{FN}{TP + FN} \tag{6}$$

$$FPR = \frac{FP}{FP + TN} \tag{7}$$

$$PWC = \frac{FN + FP}{TP + FP + FN + TN} \times 100 \tag{8}$$

$$FAR = \frac{FN}{FN + TN} \times 100 \tag{9}$$

$$MR = \frac{FP}{TP + FP} \times 100 \tag{10}$$

Here, FN is the number of false negatives, TN is the number of true negatives, FP is the number of false positives, and TP is the number of true positives. Note that because the goal of our task is global anomaly detection, the performance metrics listed above are calculated at the frame level, not at each pixel level so that it is considered a frame with abnormal changes if the percentage of changed pixels is higher than 10% of the total.

Because of the complexity and diversity related to the dimensions of the input data, it is necessary to derive the optimal value of the parameters of the proposed detection system based on strategic learning in order to obtain the optimal performance. Factors influencing the proposed system performance are related to preprocessing, such as size adjustment and normalization of the

Table 7 Comparisons of FPR of the competing methods over tested scenarios

Scenario	Frame Diff	Global Ilu	MF	Grimson GMM	Zivkovic GMM	SC-SOBS	CDAS
Office	0.033	0.000	0.000	0.028	0.014	0.000	0.000
Fall	0.097	0.066	0.081	0.042	0.025	0.008	0.115
Traffic	0.083	0.135	0.174	0.055	0.060	0.071	0.070
Over pass	0.002	0.010	0.035	0.029	0.033	0.016	0.003

Table 8 Comparisons of PWC of the competing methods over tested scenarios

Scenario	Frame Diff	Global Illu	MF	Grimson GMM	Zivkovic GMM	SC-SOBS	CDAS
Office	3.851	0.084	0.936	2.368	1.987	0.134	0.530
Fall	31.075	28.272	30.575	26.625	26.321	24.458	37.063
Traffic	64.677	49.032	45.645	30.161	28.709	29.244	24.171
Over pass	13.200	13.633	17.241	13.233	13.133	8.338	11.284

input image, as well as the density of the two-dimensional lattice structure composed of neurons, i.e., the size of the SOM. To optimize the proposed CDAS, we compared the change detection performance in terms of the PWC and MR values while varying the size of the SOM and the size of the input image; these results are shown in Fig. 3. The same data and parameters of $\tau(t)$ and $\sigma^2(t)$ were used for SOM learning. When changing the size of the image data, the PWC and MR values remained relatively low, regardless of the grid size of the SOM when using a half-compressed image. The highest performance was achieved with a grid size of 5×5; thus, this was selected for future tests.

To evaluate the CDAS, after optimization via learning, we adopted detection techniques that have been widely used for change detection and require relatively small computation. We employed the frame difference (Frame Diff) scheme, which tracks movement by looking at the difference between the current image and the previous one [40]; a median filter (MF), which detects objects through filter-based background removal [32]; a global illumination (Global Ilu) algorithm, which removes backgrounds from pre-trained images [41]; the Gaussian mixture model that uses two mixture of Gaussians, Grimson GMM [15]; and the Gaussian mixture model containing a non-fixed number of Gaussian models, Zivkovic GMM [42]. Additionally, an artificial network-based change detection methodology named spatially coherent self-organizing background subtraction (SC-SOBS) [43] was also considered. All of the detection techniques used in this comparative study are methods for removing the background and tracking an object. If the number of detected pixels is above a certain detection threshold, the result is determined as being normal or abnormal; Table 2 shows these results. In the table, the proposed change detection scheme showed good performance over the entire test image regardless of the degree of abnormality. Although the FNR

values were low for all of the detection techniques used in the performance comparison tests, the Frame Diff, MF, and GMM-based methods showed inferior performance for all of the other measures. The high FNR values indicate relatively low detection performance in the classes 1 and 2 test images in terms of the dynamic variation. Additionally, although the PWC and FAR values of the proposed CDAS are higher than the results by the Global Ilu, this difference is negligible.

Additional experiments were conducted to evaluate the robustness of the proposed system in a noisy environment. Additive white Gaussian noise (AWGN), which can occur during various processes (e.g., image signal processing, data transmission, and storage), with zero mean and unit variance scaled by 0.2 is applied to the image to evaluate performance. In addition, we assessed the robustness of the proposed system to light noise by also adding a value, which is generated randomly in the range from − 75 to 75, for the entire pixel of the original image, taking into consideration of the change in brightness of the surroundings due to the monitoring environment and weather conditions. These results are shown in Tables 3 and 4. For the conventional detection methods we considered, including Frame Diff, MF, GMM-based methods, and SC-SOBS, the degree of change in the image due to noise was too frequent, and the calculated measures were over the upper limits, causing the PWC, FAL(R), and MR values to become large. These existing methods are inappropriate for the detection of image changes in noisy environments. The CDAS showed the PWC and MR values that were somewhat higher than the Global Ilu method, which was due to the degradation of noise immunity in some test intervals. However, the CDAS showed a FAR value of 0%, while Global Ilu and SC-SOBS exceeded 40% and 20%, respectively, proving the relative noise robustness of the CDAS.

Overall, these results demonstrate the outstanding scalability of the CDAS and its overwhelmingly positive

Table 9 Comparisons of FAR of the competing methods over tested scenarios

Scenario	Frame Diff	Global Illu	MF	Grimson GMM	Zivkovic GMM	SC-SOBS	CDAS
Office	6.841	0.000	0.000	3.160	2.284	8.331	0.000
Fall	26.765	25.664	27.014	25.433	25.688	27.516	34.651
Traffic	77.641	84.408	70.817	66.671	56.250	31.412	47.260
Over pass	14.223	14.138	15.960	12.742	12.178	11.207	9.124

Table 10 Comparisons of MR of the competing methods over tested scenarios

Scenario	Frame Diff	Global Ilu	MF	Grimson GMM	Zivkovic GMM	SC-SOBS	CDAS
Office	2.660	0.157	1.328	2.481	2.144	0.000	0.751
Fall	68.105	55.587	69.616	44.578	39.459	3.866	7.880
Traffic	39.906	33.871	27.823	28.282	27.980	18.271	17.552
Over pass	3.992	9.025	28.919	18.836	19.395	2.798	1.247

performance in a noisy environment. The immunity to noise of a surveillance system is one of the most important factors because a single error can lead to failure of the system.

3.2 Performance evaluation with the CDnet-2014 dataset

We used the ChangeDetection.net (CDnet-2014) dataset to validate the performance of the CDAS, which was designed as a binary classifier, in a wider variety of environments (e.g., office, environment in dynamic background, camera jitter), as shown in Table 5. The dataset consists of 2050 image frames (360 × 240 resolution) collected at a speed of 17 frames/s. The ground truth label was regarded as a binary classification problem consisting only of class 1 (normal) and class 3 (dynamic abnormal) conditions. After labeling the images corresponding to each sequence, 10% of the total data was used for learning and 90% was used for evaluation.

The optimal parameter values of the CDAS were derived as 0.5 and 5 × 5 for the image scale and the size of the SOM through learning. Performance testing of the CDAS was carried out in various environments and was compared with conventional change detection algorithms. The same performance metrics used in the previous section were used. As shown in Tables 6, 7, 8, 9, and 10, the CDAS showed good performance, similar to the results observed in the previous section, for all of the evaluation metrics. In particular, it demonstrated superior performance compared to existing Frame Diff, Global Ilu, and GMM-based methods, and is similar to SC-SOBS in scenarios of "fall," "traffic," and "over pass" with dynamic background or camera jitter noise. It was very similar to the Global Ilu and MF results in the case of "office" scenario, with very small performance differences. In addition, since the data used for the test did not include the class 2 level, both the Frame Diff and

GMM-based methods, which were relatively weak in static abnormal conditions, showed low PWC and MR values of higher than 30%.

As seen in Tables 11 and 12, which show the results for the robustness evaluation in environments where white noise is added or a brightness change exists, excellent change detection performance of the proposed technique is confirmed, as compared with the conventional methods. Regardless of the type of noise, the PWC, FAR, and MR values of the conventional methods were greatly increased, and the overall detection performance was significantly degraded. The Global Ilu, Frame Diff, SC-SOBS, and GMM methods, which showed excellent performance in a noise-free environment, had FAR values exceeding 90% in a noisy environment and were unable to detect a normal condition. In an additional test in a noisy environment, which was made by varying the irregular brightness, the conventional detection techniques once again showed large FAR and MR values, confirming their weakness to noise. In the case of the proposed CDAS, the PWC, FAR, and MR values in all noisy environments are similar to the results in a noise-free environment, indicating excellent robustness to noise and confirming that the change detection capability is maintained despite the presence of interference.

Figure 4 compares the performance of the proposed method with those of conventional detection techniques depending on the degree of image change using the CDnet-2014 "office" dataset. It also depicts some of the change detection results in the noisy environment and includes the original image. The CDAS demonstrated excellent change detection performance and can distinguish any of the classes existing in the video. Unlike the other algorithms, which cannot detect a variation due to their low noise robustness when white noise is added or

Table 11 Comparison results with the CDnet-2014 "office" dataset in a noisy environment

	Frame Diff	Global Ilu	MF	Grimson GMM	Zivkovic GMM	SC-SOBS	CDAS
FNR	0.293	0.221	0.018	0.223	0.241	0.101	0.008
FPR	0.025	0.000	0.036	0.374	0.365	0.667	0.020
PWC (%)	28.531	20.198	1.361	20.384	20.176	10.954	0.580
FAR (%)	97.000	70.700	2.205	71.745	72.923	95.384	0.000
MR (%)	0.784	0.000	1.023	5.584	5.264	1.079	0.795

Table 12 Comparison results with the CDnet-2014 "office" dataset when the brightness changes

	Frame Diff	Global Ilu	MF	Grimson GMM	Zivkovic GMM	SC-SOBS	CDAS
FNR	0.248	0.105	0.134	0.013	0.013	0.005	0.000
FPR	0.635	0.019	0.040	0.0442	0.048	0.988	0.002
PWC (%)	38.295	7.513	11.518	22.768	23.225	39.293	0.577
FAR (%)	50.945	26.357	38.121	2.986	2.718	36.363	0.000
MR (%)	33.241	0.015	0.908	34.131	33.416	39.315	0.774

brightness changes exist, the CDAS maintained its high-detection performance without performance degradation over the entire range. This confirms the outstanding ability of the proposed CDAS and its very high robustness in noisy environments.

4 Conclusion

In this paper, we proposed a real-time intelligent surveillance system, i.e., a SOM-based image change detection technique, to overcome problems related to low noise robustness and operational cost-efficiency, which are
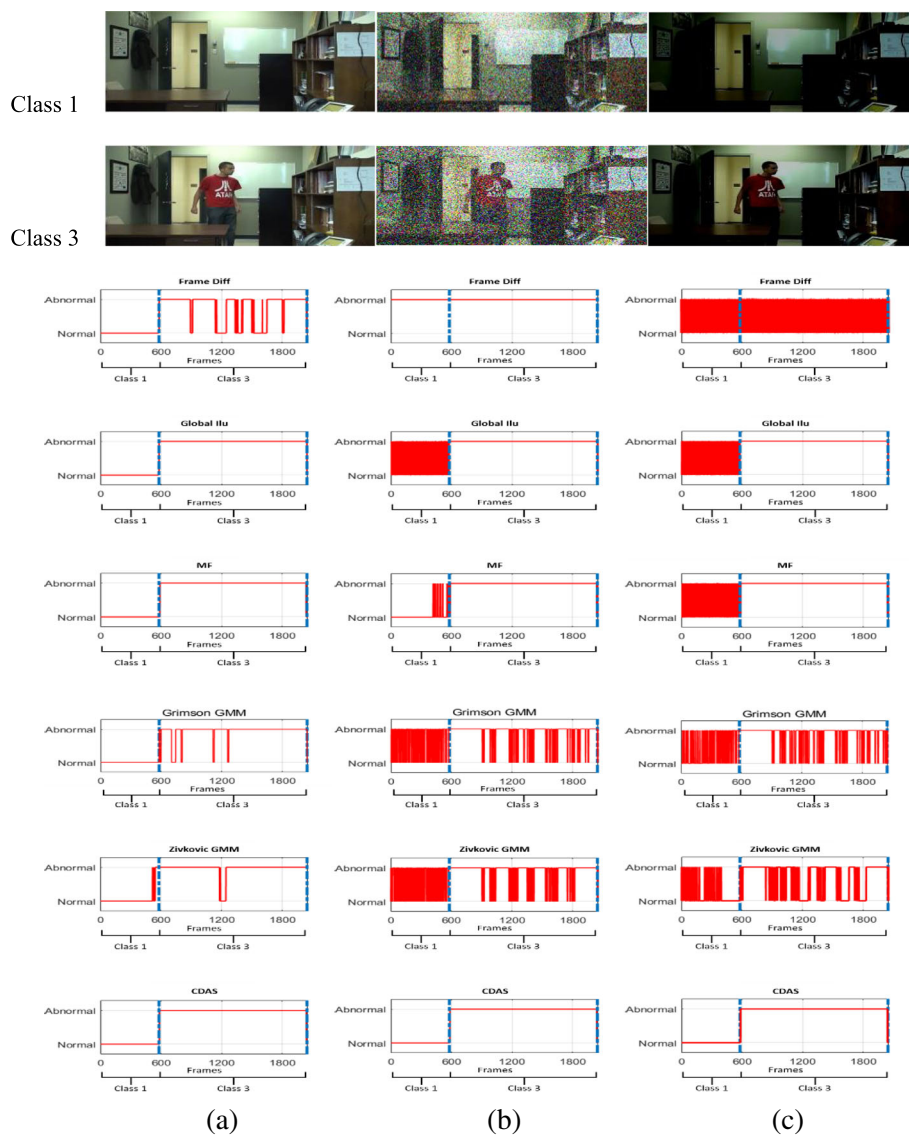

Fig. 4 Performance comparisons with the CDnet-2014 dataset with the CDAS and other surveillance systems. (**a**) In a noise-free environment. (**b**) In the presence of noise. (**c**) When the brightness changes, the pictures on the top are sample images of an "office" scenario

inherent in existing video surveillance systems. The proposed detection method was optimized via clustering through the competitive learning process of the SOM. Then, classes were assigned according to the cluster characteristics of neurons in a two-dimensional lattice structure. The similarity between image data (input in real-time) and the optimized neuron was checked. We classified and predicted the class of the nearest neighbor neurons based on a weighted directed graph, and finally determined the change of the image based on the class information of the classified winner neuron, i.e., BMU. In order to verify the superiority of the proposed system, we conducted comparative tests with other detection methods using both a hand-labeled dataset with various environmental changes and the CDnet-2014 dataset. We successfully demonstrated that our method is suitable for image change detection. In particular, the ability to detect static anomalies (object movement or on-screen placement changes) and maintain monitoring in noisy environments, which are critical to surveillance systems, was proven to be better than other change detection systems. The proposed system can be applied in various industrial environments, including indoor and outdoor monitoring. Additionally, it can become a more generalized anomaly monitoring system through the development of learning reinforcement techniques to further improve robustness in the future.

Abbreviations
CCTV: Closed-circuit television; CDAS: Change Detection Architecture based on SOM; CNN: Convolution neural network; FAR: False alarm rate; FNR: False negative rate; FPR: False positive rate; GMM: Gaussian mixture model; MF: Median filter; MR: Missing rate; NN: Nearest neighbor; PWC: Percentage of wrong classification; SOM: Self-organizing map; SVM: Support vector machine; U-Matrix: Unified distance matrix

Author's contributions
All authors took part in the discussion of the work described in this paper. All authors read and approved the final manuscript.

Availability of data and materials
Please contact the corresponding author for data requests.

Competing interests
The authors declare that they have no competing interests.

References
1. Wang, M. L., Huang, C. C., & Lin, H. Y. (2006, June). An intelligent surveillance system based on an omnidirectional vision sensor. In 2006 IEEE Conference on Cybernetics and Intelligent Systems (pp. 1-6). IEEE.
2. M. Valera, S.A. Velastin, Intelligent distributed surveillance systems: a review. IEE Proceedings-Vision, Image and Signal Processing **152**(2), 192–204 (2005)
3. F. Ortega-Zamorano, M.A. Molina-Cabello, E. López-Rubio, E.J. Palomo, Smart motion detection sensor based on video processing using self-organizing maps. Expert Systems with Applications **64**, 476–489 (2016)
4. B. Sun, S. Velastin, Fusing visual and audio information in a distributed intelligent surveillance system for public transport systems. Acta Autom. Sin **20**(3), 393–407 (2003)
5. Tao, J., Turjo, M., Wong, M. F., Wang, M., & Tan, Y. P. (2005, December). Fall incidents detection for intelligent video surveillance. In 2005 5th International Conference on Information Communications & Signal Processing (pp. 1590-1594). IEEE.
6. A. Singh, Digital change detection techniques using remotely sensed data. International Journal of Remote Sensing. **10**(6), 898–1003 (1988)
7. Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A. K., & Davis, L. S. (2016). Learning temporal regularity in video sequences. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 733-742).
8. Xu, D., Ricci, E., Yan, Y., Song, J., & Sebe, N. (2015). Learning deep representations of appearance and motion for anomalous event detection. arXiv preprint arXiv:1510.01553.
9. Malisiewicz, T., Gupta, A., & Efros, A. A. (2011, November). Ensemble of exemplar-SVMs for object detection and beyond. In Iccv (Vol. 1, No. 2, p. 6).
10. Cui, X., Liu, Q., Gao, M., & Metaxas, D. N. (2011, June). Abnormal detection using interaction energy potentials. In CVPR 2011 (pp. 3161-3167). IEEE.
11. Leo M., Furnari A., Medioni G.G., Trivedi M., Farinella G.M. (2019) Deep learning for assistive computer vision. In: Leal-Taixé L., Roth S. (eds) Computer Vision – ECCV 2018 Workshops. ECCV 2018. Lecture Notes in Computer Science, vol 11134. Springer, Cham
12. P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, et al., Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proceedings of the National Academy of Sciences **96**(6), 2907–2912 (1999)
13. S.V. Verdú, M.O. Garcia, C. Senabre, A.G. Marin, F.G. Franco, Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps. IEEE Transactions on Power Systems **21**(4), 1672–1682 (2006)
14. C.R. Wren, A. Azarbayejani, T. Darrell, A.P. Pentland, Pfinder: real-time tracking of the human body. IEEE Transactions on pattern analysis and machine intelligence **19**(7), 780–785 (1997)
15. Stauffer, C., & Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149) (Vol. 2, pp. 246-252). IEEE.
16. M. Babaee, D.T. Dinh, G. Rigoll, A deep convolutional neural network for video sequence background subtraction. Pattern Recognition **76**, 635–649 (2018)
17. L.A. Lim, H.Y. Keles, Foreground segmentation using convolutional neural networks for multiscale feature encoding. Pattern Recognition Letters **112**, 256–262 (2018)
18. D. Vallejo, F.J. Villanueva, J.A. Albusac, C. Glez-Morcillo, J.J. Castro-Schez, Intelligent surveillance for understanding events in urban traffic environments. International Journal of Distributed Sensor Networks **10**(8), 723819 (2014)
19. M. Al-Nawashi, O.M. Al-Hazaimeh, M. Saraee, A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments. Neural Computing and Applications **28**(1), 565–572 (2017)
20. D. Murray, A. Basu, Motion tracking with an active camera. IEEE transactions on pattern analysis and machine intelligence **16**(5), 449–459 (1994)
21. M.J. Roshtkhari, M.D. Levine, An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions. Computer vision and image understanding **117**(10), 1436–1452 (2013)
22. Sultani, W., Chen, C., & Shah, M. (2018). Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6479-6488).
23. Gordo, A., Almazán, J., Revaud, J., & Larlus, D. (2016, October). Deep image retrieval: Learning global representations for image search. In European conference on computer vision (pp. 241-257). Springer, Cham.
24. L. Peeters, F. Bacao, V. Lobo, A. Dassargues, Exploratory data analysis and clustering of multivariate spatial hydrogeological data by means of GEO3DSOM, a variant of Kohonen's Self-Organizing Map. Hydrology and Earth System Sciences **11**, 1309-1321 (2007)

25. P. Stefanovic, O. Kurasova, Visual analysis of self-organizing maps. Nonlinear Analysis: Modeling and Control **16**(4), 488-504 (2011)

26. Wang, Y., Jodoin, P. M., Porikli, F., Konrad, J., Benezeth, Y., & Ishwar, P. (2014). CDnet 2014: an expanded change detection benchmark dataset. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 387-394).

27. M.L. Shahreza, D. Moazzami, B. Moshiri, M.R. Delavar, Anomaly detection using a self-organizing map and particle swarm optimization. Scientia Iranica **18**(6), 1460–1468 (2011)

28. R. Xiao, R. Cui, M. Lin, L. Chen, Y. Ni, X. Lin, SOMDNCD: image change detection based on self-organizing maps and deep neural networks. IEEE Access **6**, 35915–35925 (2018)

29. Tian, J., Azarian, M. H., & Pecht, M. (2014, July). Anomaly detection using self-organizing maps-based k-nearest neighbor algorithm. In Proceedings of the European Conference of the Prognostics and Health Management Society.

30. Olson, D. L., & Delen, D. (2008). Advanced data mining techniques. Springer Science & Business Media.

31. Cohen, I., & Medioni, G. (1999, June). Detecting and tracking moving objects for video surveillance. In cvpr (p. 2319). IEEE.

32. Z. Zivkovic, F. Van Der Heijden, Efficient adaptive density estimation per image pixel for the task of background subtraction. Pattern recognition letters **27**(7), 773–780 (2006)

33. P. Törönen, M. Kolehmainen, G. Wong, E. Castrén, Analysis of gene expression data using self-organizing maps. FEBS letters **451**(2), 142–146 (1999)

34. R. Dlugosz, T. Talaska, W. Pedrycz, R. Wojtyna, Realization of the conscience mechanism in CMOS implementation of winner-takes-all self-organizing neural networks. IEEE Transactions on Neural Networks **21**(6), 961–971 (2010)

35. T. Russo, P. Carpentieri, F. Fiorentino, E. Arneri, M. Scardi, A. Cioffi, S. Cataudella, Modeling landings profiles of fishing vessels: An application of Self-Organizing Maps to VMS and logbook data. Fisheries Research **181**, 34–47 (2016)

36. A. Ultsch, *Self-organizing neural networks for visualisation and classification. In Information and classification (pp. 307-313)* (Springer, Berlin, Heidelberg, 1993)

37. A. Ultsch, Kohonen's self-organizing feature maps for exploratory data analysis. Proc. INNC90, 305–308 (1990)

38. G. Yu, J. Yang, On the robust shortest path problem. Computers & Operations Research **25**(6s), 457–468 (1998)

39. Broumi, S., Bakal, A., Talea, M., Smarandache, F., & Vladareanu, L. (2016, November). Applying Dijkstra algorithm for solving neutrosophic shortest path problem. In 2016 International Conference on Advanced Mechatronic Systems (ICAMechS) (pp. 412-416). IEEE.

40. Bouwmans, T., Porikli, F., Höferlin, B., & Vacavant, A. (Eds.). (2014). Background modeling and foreground detection for video surveillance. CRC press.

41. Vijverberg, J. A., Loomans, M. J., Koeleman, C. J., & de With, P. H. (2009, September). Global illumination compensation for background subtraction using Gaussian-based background difference modeling. In 2009.

42. Zivkovic, Z. (2004, August). Improved adaptive Gaussian mixture model for background subtraction. In ICPR (2) (pp. 28-31).

43. Maddalena, L., & Petrosino, A. (2012, June). The SOBS algorithm: what are the limits?. In 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (pp. 21-26). IEEE.

Publisher's Note