Enhanced Unsupervised Change Detector for Industrial Surveillance Systems

Ajmal Shahbaz, *Member*, IEEE and Kang-Hyun Jo, *Senior member*, IEEE

***Abstract*—Detecting an intruder that is trespassing a prohibited area is a critical task of intelligent surveillance systems. This task requires a change detector to segment an intruder (foreground object) from the background. The task suffers the inherent drawbacks of change detectors due to the dual-camera sensor (color/IR), illumination changes, night time, static, and camouflaged foreground objects. This paper proposes an enhanced unsupervised change detector (EUCD) to compensate for the aforementioned challenges for industrial sterile zone monitoring. The camera switch detection based on skewness patterns detects a switch between the dual camera sensors (color/IR). The optimal color space selection based on the mean squared error will select tolerant color space (RGB/YCbCr) to illumination changes for modeling the background. Also, the IR camera frames are contrast-enhanced to tackle the camouflaged intruders during the night. The incoming frames are split into respective channels before modeling the background. The background is modeled by Gaussian Mixture Models (GMM). The adaptive background model update scheme is proposed to tackle the various challenges posed by environment and object such as a static foreground object. The comparison is performed on three databases with top-ranked unsupervised change detection algorithms.**

***Index Terms*—Intelligent surveillance systems, camouflaged intruder, dual-camera sensors, IR camera, security system, night time.**

I. Introduction

INTELLIGENT surveillance systems (ISSs) play an important role in protecting sensitive areas. They are proving helpful not only for detecting anomalies but also for tracking subjects of interest. ISSs powered by computer vision algorithms are steadily taking over conventional surveillance systems. They require minimal human intervention and allow the automatic detection of an anomaly.

Sterile zone monitoring is a crucial task of ISS to detect intruders/trespassers in a prohibited area. The definition of a sterile zone depends on the application. It could be a border between countries, a fence of a prison, or a rooftop of a skyscraper. Thus, sterile zone monitoring can be utilized in a wide range of prohibited areas.

Sterile zone monitoring employs the change detection algorithm to segment a foreground object (intruder) from the background. The definition of a foreground object depends on the application. It could be a human walking in a corridor [1], a car parked illegally on a road [2], a bag abandoned at a bus station [3], smoke, or a fire in a forest [4], etc. ISSs are challenged due to inherent and practical drawbacks faced by the change detectors by dual camera sensors (color/IR), illumination changes, dynamic backgrounds, bad weather, static, and camouflaged foreground objects.

The change detection algorithms are classified into two categories based on the training process: unsupervised and supervised algorithms [5]. Unsupervised algorithms are trained online with incoming frames to construct a concrete background model using pixel intensity. The supervised algorithms based on Convolutional Neural Networks (CNN) are trained offline on GPUs with background and foreground information [6].

Gaussian Mixture Models (GMM) [7] is considered as the most popular unsupervised change detection algorithm. GMM models background using mean and variance. In recent years, there have been remarkable improvements in GMM based algorithms such as [8]–[11]. They are cost-efficient and have better accuracy in coping with illumination changes. However, they fail to cope with the camouflage effect due to pixel intensity based background modeling [5].

Self-Balanced SENsitivity SEgmenter (SuBSENSE) [12] exploits texture information around the pixel using local binary patterns (LBP) to model the background. Pixel-based Adaptive Word Consensus Segmenter (PAWCS) [13] improved SuBSENSE using an adaptive threshold strategy. Weight Sample Background Extractor (WeSamBE) [14] improved SuBSENSE by integrating the weighted reward/penalty strategy for pixel-labeling. These algorithms showed promising results at high computational costs [6].

Flux Tensor and Split Gaussian (FTSG) [15] models the background using Gaussian and flux tensors. In Unity There Is Strength (IUTIS) [16] proposed a genetic algorithm to choose subsets of the best performing unsupervised models to construct the final foreground mask. These algorithms are computationally inefficient for low-cost real-time systems [5].

Subspace/low-rank models provide a superior framework to segment a foreground object [17]–[19]. The main theme of such algorithms is to exploit the principal component analysis or the subspace transformation to build the background

Manuscript received January 17, 2020; revised Month xx, xxxx; accepted Month x, xxxx. This work has been supported by the National Research Foundation of Korea (NRF) grant funded by the government (MSIT) (No. 2020R1A2C2008972). (Corresponding author: Kang-Hyun Jo)

Ajmal Shahbaz is with the School of Electrical and Computer Engineering, University of Ulsan, Ulsan, 44610 South Korea (e-mail: ajmal@islab.ulsan.ac.kr). Kang-Hyun Jo is with the Department of Electrical and Computer Engineering, University of Ulsan, Ulsan, 44610 South Korea (e-mail: acejo@ulsan.ac.kr).

model. Multi-Layer Robust Principal Component Analysis (ML-RPCA) [17] exploits low-rank recovery and extracts the background information using multi-dimensional arrays. These methods are based on batch optimization processing. The frames are stored in memory to train the model. Thus, they require high computational costs [19].

Supervised algorithms based on CNN are showing astonishing results [20]–[24]. CNN is trained offline using background and foreground labels for a video. Braham *et al.* [21] trained a model on 50% training frames from a video and tested the remaining frames using the said model. DeepBS [22] trained only one model with 5% training frames from all video sequences of the change detection dataset (CDNet). CascadeCNN [23] trained multiple CNNs with multi-scale input images. The CNN based algorithms outperformed unsupervised algorithms, but requires high-end hardware and lacked real-time performance.

This paper proposes an enhanced unsupervised change detector (EUCD) for the task of industrial sterile zone monitoring. The word enhanced refers to the several improvements implemented over GMM [9] to tackle its inherent drawbacks of dual camera sensors, illumination changes, static, and camouflaged foreground objects. The contributions are as follows:

- A novel camera switch detection (CSD) scheme detects the switch between color and IR sensors (II.A.1).
- A novel optimal color space selection strategy to select illumination change-tolerant color space (RGB/YCbCr) for modeling the background (II.A.2).
- An efficient contrast enhancement scheme for enhancing IR camera frames to tackle camouflaged intruders at night (II.A.3).
- A novel adaptive background model update scheme for updating the background model to tackle the challenges of illumination changes, dynamic backgrounds, moving, and static foreground. (II.B.3).
- There is no public benchmark for an ISS with IR camera-based video sequences. The dataset of HD videos is created and would be publicly available.

GMM is employed as a change detector due to its adaptive nature, good accuracy, and low computational cost [5]. However, other unsupervised change detectors can be integrated with the proposed algorithm. Thus, the proposed algorithm is also integrated with SuBSENSE, PAWCS, WeSamBE, and ML-RPCA to demonstrate its effectiveness. The proposed algorithm is tested on three databases namely the Imagery Library for Intelligent Detection Systems (i-LIDS) dataset [25], ISLab-Industrial Sterile Zone Monitoring (ISL-ISZM) dataset, and the Change Detection (CDNet) dataset [26].

The remainder of the paper is organized as follows: Section II explains the proposed algorithm in detail. Section III provides quantitative, qualitative, and computational evidence in support of the proposed algorithm. Section IV provides conclusions.

II. Proposed Algorithm

The proposed algorithm consists of two modules, as shown in Fig. 1:

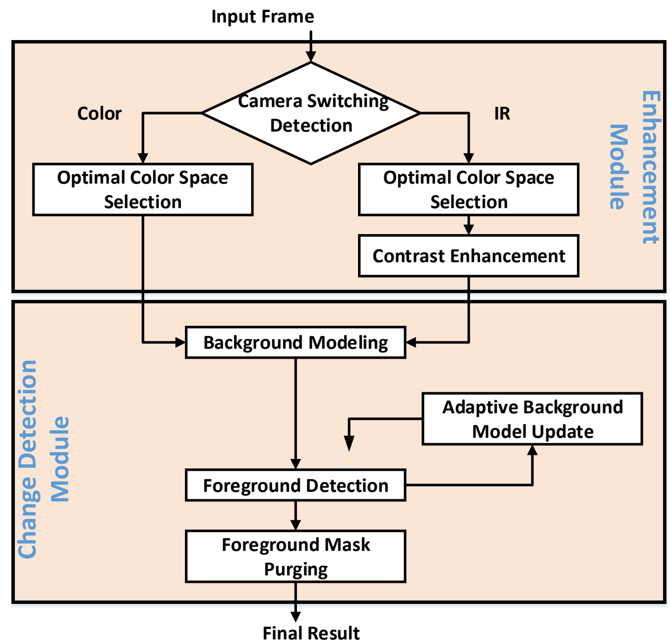


Fig. 1. System diagram of the proposed algorithm.

1) Enhancement Module: The type of input frames (Color/IR) is detected using camera switching detection. Later, the input is pass through optimal color space selection to select the optimal color space (RGB/YCbCr) to tackle illumination changes. Also, the IR input is contrast-enhanced to distinguish camouflaged intruders from the background.
2) Change Detection Module: The enhanced input is then used to model the background and detect the foreground. The background model is updated automatically during the whole process. The foreground mask is purged to get the final result.

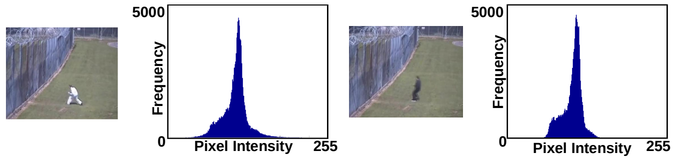
A. Enhancement Module

1) Camera Switch Detection (CSD): Current ISSs employ dual camera sensors for the day (color) and night (IR). Changes in sunlight intensity causes a switch between the camera sensors signaling the time of day. Such a scheme while economical comes with severe drawbacks. The switch between sensors may distort the unsupervised change detector which exploits the background model to segment an intruder from a scene. Such distortion results in false positives. Also, the IR sensor may pose a strong camouflage effect. Due to pixel intensity based background modeling, it could lead to false negatives resulting in ISSs failure.

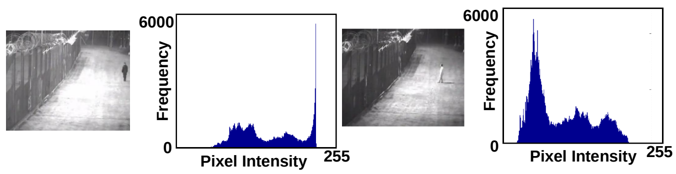
A novel camera switch detection (CSD) scheme has been proposed to tackle the aforementioned challenges. The premise is derived from the skewness patterns of the color and IR camera. The color camera gives balanced information about a scene and a varied range of intensity. While the IR camera provides information in shades of gray and a congested intensity range. Thus, the skewness patterns of both camera sensors differ remarkably and follow these three patterns:

1) If $\mu = m = M$, it is classified as symmetry.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3013747, IEEE Transactions on Industrial Electronics



(a) The color camera frames with their skewness patterns. 1st frame (μ = 122, m = 123, M =129, $|M-m|$ = 6, $|M-\mu|$ = 7) and 2nd frame(μ = 123, m = 123, M =129, $|M-m|$ = 6, $|M-\mu|$ = 6)



(b) The IR camera frames with their skewness patterns. 1st frame (μ = 166, m = 156, M =254, $|M-m|$ = 98, $|M-\mu|$ = 88) and 2nd frame(μ = 120, m = 102, M =64, $|M-m|$ = 38, $|M-\mu|$ = 56).

Fig. 2. Skewness patterns exhibited by the color and IR camera, where x- axis and y-axis shows pixel intensity and frequency respectively.

2) If $\mu > m > M$, it is classified as left-skewed.
3) If $\mu < m < M$, it is classified as right-skewed.

where μ, m, and M are mean, median, and mode of an image respectively. Fig. 2 visualizes the skewness patterns in the day (color) and night (IR) frames. The day frames follow a nearly symmetrical pattern (Fig. 2a), whereas the night frames exhibit either left or right skewness (Fig. 2b). Following the symmetrical pattern, the mean μ, median m, and mode M of the day frames were approximately equal (Fig. 2a). However, the night frames showed that the mean μ, median m, and mode M were far apart and followed either a left or a right skewed pattern (Fig. 2b).

The CSD criterion is formulated from three skewness patterns to detect the switch from a color to an IR camera sensor. The criterion can be written as:

$$CSD = \begin{cases} IR, & |M-m| \geq T \vee |M-\mu| \geq T \\ Color, & otherwise \end{cases} \tag{1}$$

where T is the CSD threshold selected heuristically (Section III.B). The mean μ, median m, and mode M are scalar entities averaged over the three image channels. For example, the mean μ is the sum of all the pixels divided by the total number of pixels in an image averaged over three channels. If there is a switch between the camera sensors then the CSD signals to initialize the background modeling again. Also, if the IR camera is detected, the incoming frames are contrast-enhanced before modeling the background. The criterion is simple yet powerful to detect the left and right skewed incoming IR images.

2) Optimal Color Space Selection (OCSS): Sterile zone monitoring is an outdoor task. There will be a time at which the ISS faces sudden or variable illumination changes. This may result in false positives. The optimal color space selection (OCSS) aims at tackling the illumination changes by selecting tolerant color space (RGB/YCbCr) to model the background. The effectiveness of both color spaces for illumination changes has been documented in the literature [6]. Several works have proposed the application of multiple color spaces to tackle illumination changes [11]. Such algorithms are cost ineffective as they maintain multiple background models. The OCSS selects the optimal color space to model the background which is a cost-efficient solution.

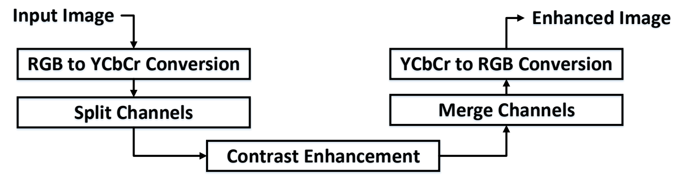


Fig. 3. Cost-efficient contrast enhancement (CE) scheme for the IR frames.

The premise of OCSS is derived from the working principle of the human eye. The human eye has two different cells called rods and cones. They supplement each other according to illumination changes. Rods are effective in general conditions while cones are designed to work in a variable or sudden illumination changes. The color spaces RGB and YCbCr are analogous to rods and cones respectively [11]. Following this premise, the OCSS was proposed which aids in deciding the optimal color space tolerant to the illumination changes.

The OCSS exploits mean squared error μ_{se} to select the optimal color space. It is a measure of image similarity between consecutive frames and sensitive to illumination changes. The initial frames (say 100) without foreground information were used to calculate μ_{se} for both color spaces is given as:

$$\mu_{se} = \frac{1}{mn} \sum_{i=0}^{m} \sum_{j=0}^{n} [I(i,j) - G(i,j)]^2, \tag{2}$$

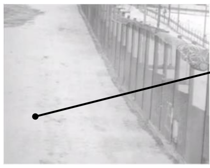
where $I(i,j)$ and $G(i,j)$ are input image and ground-truth image. m and n are the number of pixels in respective frames. The first frame of the input sequence without foreground information is selected as the ground truth $G(i,j)$. It is possible to get such a frame as ISS has the liberty to record input sequences without foreground information. Also, initial frames (100-200 frames) of ISS benchmark (e.g., i-LIDS datasets) are recorded without foreground information [25].

The optimal color space is selected as the one satisfying the following criterion:

$$\mu_{se}^{avg} \leq 5\mu_{se}^{1}, \tag{3}$$

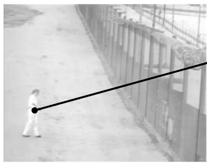
where μ_{se}^{avg} is the average mean squared error of consecutive frames. μ_{se}^{1} is the mean squared error between the first frame and ground-truth. Here the first frame is the one after the selected ground-truth. Such criterion is inferred from the foreground detection rule for unsupervised change detectors, which allows a deviation of pixel intensities from $\pm$ 5 incorporated as the background (Eq. 5). If both color spaces satisfy the condition, RGB color space is selected.

3) Contrast Enhancement (CE): The IR input frames may pose a strong camouflage effect, i.e., the foreground object and background have similar pixel intensity. The cost-efficient contrast enhancement (CE) schema is proposed to tackle the camouflaged intruder at night as shown in Fig. 3. The



231	231	230	230	230
229	230	230	231	230
230	231	232	231	230
230	230	230	231	231
231	230	230	230	231

(a) IR frame without foreground object.



232	235	236	232	227
231	234	235	231	227
226	229	233	232	226
229	235	237	226	224
233	242	243	232	221

(b) IR frame with camouflaged foreground object.

247	248	254	247	250
254	252	251	249	251
253	252	251	250	251
251	250	252	251	250
251	250	250	250	250

138	139	141	136	141
139	140	146	154	144
145	146	152	153	143
148	150	154	143	143
146	146	147	149	144

(c) Contrast-enhanced IR frame of (b).

Fig. 4. Visualizing the contrast enhancement (CE). There is significant pixel intensity difference between background and camouflaged intruder after CE.

incoming IR frames are converted to YCbCr color space if required. If the OCSS selects RGB as optimal color space, it will be converted to YCbCr for applying CE and then converted back to RGB for further processing. As CE is an intensity stretching operation, the ideal color space would be the one showing intensity values instead of color, i.e., YCbCr.

If CE is directly applied to RGB color space, it may cause color imbalance leading to false positives due to a noisy video. The input frames are split into their respective channels. The probability mass function (PMF) and cumulative density function (CDF) are computed and mapped to the intensity range. Later, the channels are merged and color space is converted back to RGB, if necessary.

Fig. 4 shows the effectiveness of the CE in differentiating the camouflaged intruder in the IR input frames. Fig. 4a shows the IR input without camouflaged intruder (only background) with pixel intensities in a 5×5 region. Similarly, Fig. 4b shows the IR input with the camouflaged intruder. The pixel intensities of both images (4a and 4b) in the specified 5×5 regions are similar. Such small differences are hard to detect by change detectors due to pixel intensity based background modeling [6]. Fig. 4c shows the contrast-enhanced version of Fig. 4b. It can be seen that the contrast has been increased between the background and the camouflaged intruder. This helps to detect the intruder effectively by the change detection module.

B. Change Detection Module

1) Background Modeling: The background is modeled from the initial frames (say 100) without foreground information. Each frame is split into their respective color channels (e.g, R, G, B). Each pixel in its respective channel is modeled using GMM [9]. The probability P of a pixel X at time t being background is formulated as:

$$P_{X_t} = \sum_{i=1}^{G} \omega_{i,t} \eta(X_t; \mu_{i,t}, \sigma_{i,t}^2), \tag{4}$$

where G, $\omega_{i,t}$, $\mu_{i,t}$, and $\sigma_{i,t}^2$ are number of Gaussian, estimate of weight, mean, and variance of the ith Gaussian in the mixture at time t. Since only Y channel carries information while Cb and Cr are useless without actual colors. Thus, the Y channel is used to model the background in the YCbCr-based IR camera frame.

2) Foreground Detection: The background model is compared with an incoming frame with the foreground information. The foreground detection rule to mark particular pixels at time t as the foreground is:

$$|X_t - \mu_{i,t}| > \lambda \sigma_{i,t}, \tag{5}$$

where λ = 2.5 is the foreground detection threshold inferred from the 68-95-99.7 standard deviation σ rule in statistics [5]. 1 σ, 2 σ, and 3 σ covers 68%, 95%, and 99.7% of pixel values within a Gaussian. Thus, a pixel value located at more than 2.5 σ (99%) away from the estimated mean component of a Gaussian is labeled as foreground.

3) Adaptive Background Model Update: The new background and foreground values need to be updated in the background model after foreground detection. The general scheme [2]-[5] to update the current pixel value in the new background model is as weighted sum of the pixel value in the current frame and pixel value in the previous background model:

$$B_t = \alpha I_t + (1 - \alpha) B_{t-1}, \tag{6}$$

where B_t, I_t, B_{t-1}, and α is the new background model, current pixel value, previous background model, and learning rate respectively.

Learning rate α is a crucial parameter to decides how long a certain pixel classified as foreground, will stay as a foreground. A fixed α value ranging between 0 to 1 is usually utilized to update a background model [6]-[10]. However, a fast-changing scene needs a high α value such as illumination changes, dynamic backgrounds, and moving foregrounds. For example, leaves moving on a tree (dynamic backgrounds) may be labeled as foreground and should promptly be labeled as background.

A slowly-changing scene requires a low α. For example, a static foreground object (SFO) is a challenge when a foreground object enters a scene and stays static at a certain position for a long time. SFO would be diffused into the background over time due to the background model update. Hence, an adaptive background model update scheme is required for an ISS to tackle the aforementioned challenges.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3013747, IEEE Transactions on Industrial Electronics

Several works addressed the fixed α problem by modeling and updating the background model with multiple learning rates [2, 3, 27]. Wahyono $et\ al.$ [2, 3] proposed a dual-learning rate scheme to model and update the background models separately. The scheme is only focused on extracting SFO by subtracting two foreground masks. Lin $et\ al.$ [27] proposed four learning rates to deal with the illumination changes, dynamic backgrounds, and moving foreground objects. The extracted foreground masks were aggregated to obtain the final foreground mask [27]. Such schemes are computationally inefficient due to multiple background models maintenance.

A novel adaptive background model update scheme is proposed based on the measure of change of foreground pixels f_r in the scene. The innovation lies in its ability to track the changes in a scene based on the f_r using a single background model. Depending on f_r, four optimal learning rates are automatically switched in the background model update process. The rate of change of foreground pixels [11] is written as:

$$f_r = \frac{f_n^t \text{ - } f_{avg}^t}{f_{avg}^t}, \tag{7}$$

where f_n^t and f_{avg}^t is the number of foreground pixels at time t and average of foreground pixels at time t in a scene. The background model update process is initialized with a minimum value of f_r and is translated into four optimal learning rates, defined in the literature [2]-[9]. The criterion to assign different learning rate $\alpha \rightarrow L$ is defined as:

$$L = \begin{cases} \alpha = 0.1, & f_r \geq 1 \\ \alpha/10, & 1.0 > f_r \geq 0.5 \\ \alpha/100, & 0.5 > f_r \geq 0.1 \\ \alpha/1000, & 0.1 > f_r \geq 0.01 \end{cases} \tag{8}$$

where L, α, and f_r is the adaptive learning rate, optimal learning rates, and rate of change of foreground pixels. A high f_r corresponds to a fast-changing scene or a moving foreground object. As a fast-changing scene requires a high α. Thus, the background model is updated with α = 0.1. As the foreground object stays in a particular position for a long time, f_r will start decreasing to a minimum where a foreground object becomes static (SFO). Thus, the background model is updated with a low α according to the f_r. Such correspondence between f_r and four widely adopted learning rates α helps to tackle the challenges of illumination changes, dynamic backgrounds, moving, and static foreground.

4) Aggregating and Purging Foreground Mask: Later the foreground masks obtained by the respective color spaces are aggregated as follows:

$$f = \sum_{c=1}^{C} f_c \geq 2, \tag{9}$$

whereas Y channel is the final foreground mask in the YCbCr based IR frame. The aggregated foreground mask might have some isolated noise and cavities in the foreground object. Morphological opening and closing are applied to eliminate isolated noise and fill the cavities in the foreground object. The kernel size for opening (3×3) and closing (5×5) were kept as small as possible to keep the foreground object intact.

TABLE I
DATASETS DESCRIPTION.

Video name	Duration	Time	Scenario
		i-LIDS Dataset	
1	0:30	Day	Walking
2	0:30	Day	Running
3	0:30	Day	Crawling
4	0:30	Day	Walking slowly
5	0:30	Day	Walking fast
6	0:30	Night	Walking away from camera
7	0:30	Night	Walking away from camera slowly
8	0:30	Night	Far from camera
9	0:30	Night	Camouflage intruder
10	0:30	Night	Camouflage intruder
		ISL-ISZM Dataset	
11	1:21	Day	Walking
12	0:54	Day	Running
13	1:35	Day	Multiple intruders
14	0:50	Day	Dynamic background
15	0:58	Day	Dynamic background
16	0:16	Night	Walking fast
17	0:14	Night	Camouflage intruder
18	0:31	Night	Camouflage intruder
19	0:25	Night	Camouflage intruder
20	0:35	Night	Multiple intruders

III. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed algorithm is compared with the top-ranked unsupervised change detection algorithms such as GMM [7]-[9], SuBSENSE [12], PAWCS [13], WeSamBE [14], and ML-RPCA [17]. Supervised change detection algorithms are not included in the comparison as they are trained offline with foreground and background information. This consensus has been reached by the wider change detection research community (http://www.changedetection.net/).

A. Datasets Description

1) i-LIDS and ISL-ISZM dataset: Table I shows a detailed description of datasets with the challenges. The i-LIDS dataset is the UK government benchmark for video surveillance systems. The i-LIDS and ISL-ISZM dataset consist of 10 videos each, with 5 videos each for day and night. Each video in the i-LIDS dataset consists of 1,000 frames. While each video in ISL-ISZM dataset varies from 1,000-2,300 frames. The i-LIDS dataset presents the scenario of fence monitoring. ISL-ISZM dataset was filmed mimicking the challenges of the i-LIDS dataset in industrial settings. The dataset can be accessed at https://drive.google.com/file/d/1qIRUPgAQeY42zeRlTg2oTqvCC9ImYxMG/view?usp=sharing.

2) Change Detection Dataset (CDNet): The CDNet is a comprehensive dataset with 150,000 video frames and manually labeled ground truths. There are 11 categories with 4-6 videos in each category. There are 54 video sequences with each video consisting of 900-8,000 video frames. The proposed algorithm is tested on 5 categories relevant to ISS such as baseline, bad weather, dynamic backgrounds, thermal, and shadows. The categories accumulate to 25 videos and almost 80,000 video frames.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3013747, IEEE Transactions on Industrial Electronics

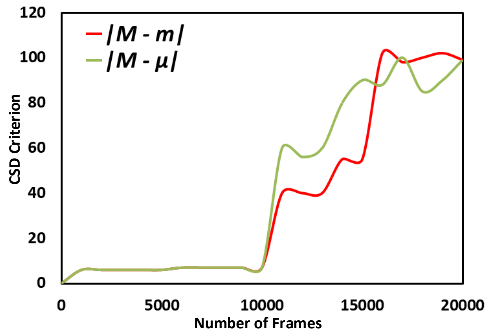


Fig. 5. The CSD criterion, where x-axis and y-axis shows the number of frames and CSD criterion (Eq. 1).

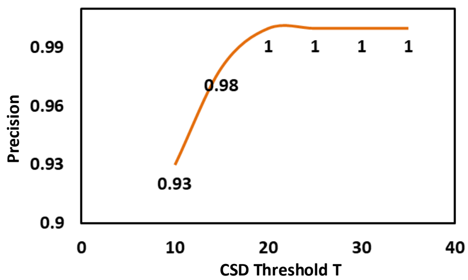


Fig. 6. The ablation study of CSD threshold T, where x-axis and y-axis shows the CSD threshold T (Eq. 1) and precision.

TABLE II
PARAMETER SETTING.

Parameter Name	Symbol	Value
CSD Threshold	T	20
Number of Gaussian	G	3
Foreground Detection Threshold	λ	2.5
Aggregated Foreground Mask	F_c	≥ 2

B. Parameter Setting

Table II shows the parameters setting used in the proposed algorithm along with the definition. All the video sequences were tested using the same parameter setting. The optimal values were chosen through extensive experiments. Also, optimal parameters were kept for GMM and its improvements. Similarly, SuBSENSE, PAWCS, WeSamBE, and ML-RPCA were applied with the original setting. The proposed algorithm employs 4 parameters only, fewer than the comparative algorithms. For example, SuBSENSE and its improvements have more than 10 parameters to tweak.

Fig. 5 shows the variation of CSD criterion, $|M\text{-}m|$ (green line) and$|M\text{-}\ \mu|$ (red line), in the day (color) and night frames (IR). The day and night frames (10,000 each) from the i-LIDS and ISL-ISZM dataset were used to get the optimal value of the CSD threshold T. The frames come from six different background settings.

The frames were arranged as day sequences followed by night sequences. The difference of $|M\text{-}m|$ and $|M\text{-}\ \mu|$ for day sequences was small, i.e., 4-9. This difference jumped above 40 and fluctuates between 40-90 for the night sequences. The variation of $|M\text{-}m|$ and $|M\text{-}\ \mu|$ between day (color) and night (IR) frames helps in deciding the CSD threshold T, as shown in Table II. CSD is powerful to detect either a left or right skewed IR frames due to its dual condition, i.e., $|M\text{-}m|$ and $|M\text{-}\ \mu|$.

The ablation study of the CSD threshold T is shown in Fig. 6. Six values of T=$\{10, 15, 20, 25, 30, 35\}$ were evaluated. 10,000 frames from the change detection dataset (4,000), i-LIDS dataset (3,000), and ISL-ISZM dataset (3,000) were employed with 5,000 frames from each day (color) and night (IR). The frames were different from the ones evaluated in Fig. 5 and comes with eight different background settings. The threshold value (T=10) close to the CSD color camera range (4-9) gives a precision of 0.93. However, T=$\{20, 25, 30, 35\}$ achieved 100% precision. This is due to the big gap of $|M\text{-}m|$ or $|M\text{-}\ \mu|$ for color and IR frames which helps to decide the optimal T (Fig. 5).

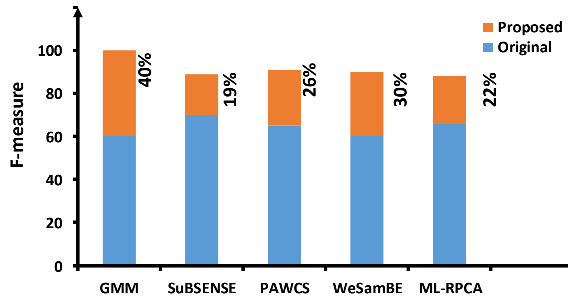


Fig. 7. Quantitative comparison of the proposed algorithm with the comparative algorithms. The blue column shows comparative algorithms. The orange column shows the proposed algorithm integrated with comparative algorithms.

C. Quantitative Analysis

The quantitative analysis on i-LIDS and ISL-ISZM dataset using F- measure F are shown in Fig 7. The blue column shows a comparative algorithm while the orange column shows the proposed algorithm integrated with the corresponding algorithm. The success criterion is defined by the i-LIDS dataset for the ISS evaluation. The intruder (true positive) should be detected for at least 75% of a particular video sequence. The analysis is shown in Fig. 7 is the average value over the 20 videos from both datasets. Each video contributes 5% of the overall F-measure.

GMM and its improvements were successful in 12 sequences of both datasets. Hence, it has a 60% F-measure (Fig. 7). SuBSENSE was able to detect and track an intruder in 14 sequences, PAWCS in 13, and WeSamBE in 12. ML-RPCA was able to detect an intruder in all sequences. However, these algorithms gave false positives. These false positives resulted in a decrease in their overall F-measures (Fig. 7).

The proposed algorithm with GMM showed impressive performance by tackling the camouflaged intruder in the night. It was able to detect and track intruders in all sequences without false positives. The proposed algorithm was also integrated with other comparative algorithms to show its generalization and effectiveness. It improved the performance of the comparative algorithms from 19-40%.

TABLE III
QUANTITATIVE ANALYSIS ON THE CDNET. THE TABLE SHOWS THE AVERAGE VALUE OF 7 PERFORMANCE METRICS NAMELY RECALL R, SPECIFICITY Sp, FALSE POSITIVE RATE FPR, FALSE NEGATIVE RATE FNR, PERCENTAGE OF WRONG CLASSIFICATIONS PWC, F-MEASURE F, AND PRECISION P.

Algorithm	R	Sp	FPR	FNR	PWC	F	P
GMM	0.7334	0.9928	0.0071	0.2660	1.9973	0.7164	0.7663
Proposed+GMM	**0.7897**	0.9946	0.0054	0.2123	1.4748	**0.8028**	**0.8242**
SuBSENSE	0.8616	0.9958	0.0041	0.1383	0.4855	0.8691	0.8895
Proposed+SuBSENSE	**0.8861**	0.9966	0.0033	0.1138	0.7916	**0.8988**	**0.9133**

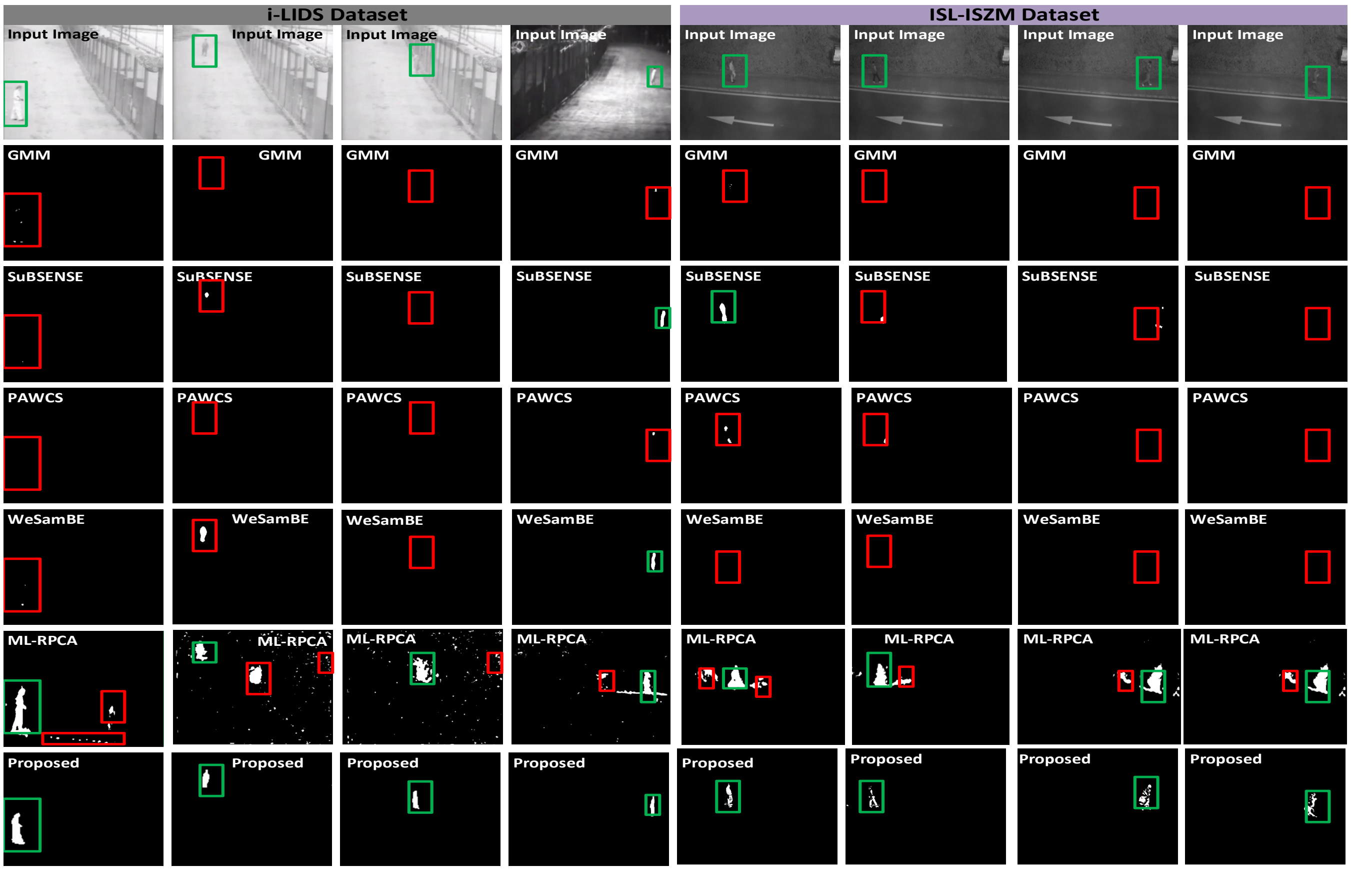


Fig. 8. Qualitative comparison of the proposed algorithm with the comparative algorithms. From top to bottom, rows show the input frame and respective foreground mask obtained from GMM and its improvements, SuBSENSE, PAWCS, WeSamBE, ML-RPCA, and proposed algorithm. Green boxes show true positives (intruder) and red boxes show false positives (illumination changes and shadows) or false negatives (miss-detection of intruder).

Table III shows the quantitative analysis of the 5 categories of CDNet such as baseline, dynamic backgrounds, bad weather, thermal, and shadows. The performance metrics are calculated by pixel-wise comparison between foreground mask and ground-truth using the software provided by the CDNet team. The quantitative results of the comparative algorithms are available on the CDNet website. The proposed algorithm improved GMM 4-6% in performance metrics such as R, P, and F. The proposed algorithm showed better precision, which is crucial for the ISS. Similarly, the proposed algorithm was integrated with SuBSENSE. It also improved the SuBSENSE by the 2-3% in terms of R, P, and F.

D. Qualitative Analysis

Fig. 8 shows the qualitative analysis of the proposed algorithm with the top-ranked change detection algorithms. The general scenario of the video sequences is the intruder entering the prohibited area. The night time sequences are shown to support the superior performance of the proposed algorithm.

The GMM and its improvements (2nd row) failed to detect the intruders properly. They segmented intruders partially. For example, the head and the shoes of the camouflaged intruder were different from the background (1st image). Also, GMM segmented only the head of the intruder which was different from the background (4th image). However, GMM failed all the challenges in ISL-ISZM dataset.

SuBSENSE (3rd row) segmented the intruder in one sequence of the i-LIDS dataset, as the intruder was significantly different from the background (4th image). It segmented the intruders partially in some sequences of i-LIDS and ISL-ISZM dataset, for which the part of the intruders was significantly different from the background (4th and 5th image).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3013747, IEEE Transactions on Industrial Electronics

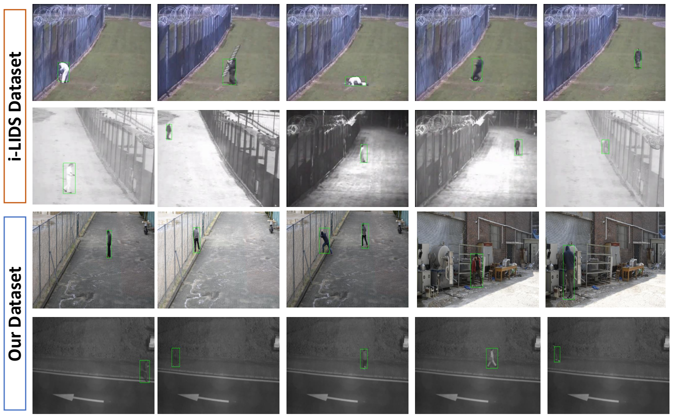


Fig. 9. Final detection results of the proposed algorithm on all the video sequences of i-LIDS and ISL-ISZM dataset.

PAWCS (4th row), similar to GMM and SuBSENSE, also segmented the intruders partially in both datasets. WeSamBE (5th row) had better performance than SuBSENSE and PAWCS on the i-LIDS dataset. It segmented the intruders in two sequences (the 2nd and 4th images). However, it failed all the sequences of ISL-ISZM dataset. It is evident from Fig. 8 that SuBSENSE, PAWCS, and WeSamBE failed to cope with ISL-ISZM dataset.

ML-RPCA (6th row) was able to segment the intruder in all the sequences of both datasets. However, it labeled large portions of the background as foreground (false positives). The proposed algorithm (7th row) was able to detect the precise geometry of the camouflaged intruder in all the video sequences. It was able to cope with strong camouflage effects, illumination changes, and static foreground object.

Fig. 9 shows the final detection result of the proposed algorithm on all the video sequences of i-LIDS and ISL-ISZM dataset. The results are arranged in numerical order as described in Table I. For instance, the 1st result in Fig. 9 refers to the 1st video in Table I. It is evident that night sequences are more challenging (2nd and 4th row).

The i-LIDs dataset is a standard benchmark and the scenes were developed in a controlled environment. ISL-ISZM dataset is more challenging as it has illumination changes, dynamic backgrounds, shadows, and camouflaged intruders. It is hard to distinguish between the camouflaged intruders and the background even to the naked eye (4th row). The performance of the proposed algorithm on three different databases with several challenges proves its generalization and effectiveness. All video results can be accessed via https://drive.google.com/open?id=1FwzwndHDrf3qC0x6lpCG2AOK1bWbO0v4.

E. Computational Analysis

The proposed algorithm was implemented in the OpenCV based C++ environment. It utilized hardware with Intel Core i5-3.80 GHz and 8 GB RAM. The video sequences were resized to 640×480. The comparative algorithms were also implemented on the same machine.

Table IV shows the computational analysis in terms of the average frames per second (fps). GMM and its improvements have good processing speed but failed to detect an intruder overall. SuBSENSE, PAWCS, and WeSamBE have low processing speed as they built the background models using the texture information. Similarly, ML-RPCA builds models using sub-space and requires batch processing which is computation inefficient. Thus, such methods are unsuitable for a real-time system with low-cost hardware. The proposed algorithm integrated with GMM outperformed the comparative algorithms with real-time performance.

TABLE IV
COMPUTATIONAL ANALYSIS

Algorithm	Processing Speed (fps)
GMM	25-35
Proposed+GMM	28-30
SuBSENSE	4
Proposed+SuBSENSE	3.9
PAWCS	2
Proposed+PAWCS	2
WeSamBE	2
Proposed+WeSamBE	2
ML-RPCA	0.5
Proposed+ML-RPCA	0.5

TABLE V
OPERATION-WISE PROCESSING TIME

Operation	Processing Time (ms)
Camera Switching Detection	1.2
Optimal Color Space Selection	2.3
Contrast Enhancement	±3.1
Adaptive Background Update	5.6
Foreground Mask Purging	0.3
Background Modeling	16.6
Foreground Detection	6.6
Total	32.6±3.1

TABLE VI
MINIMUM HARDWARE EVALUATION FOR REAL-TIME PERFORMANCE

CPU	RAM	Image Size	Processing Speed (fps)
Core i5-3.80 GHz	8 GB	640×480	30
Core i3-2.66 GHz	8 GB	640×480	21
Quad core-2.90 GHz	4 GB	640×480	13

Table V shows the operation-wise processing time in milliseconds (ms). The background modeling and foreground detection operations from GMM constitute most of the processing time. The enhancements contribute to 28-34% (9.1±3.1 ms) of the total processing time (32.6±3.1 ms). The camera switch detection scheme helps to apply contrast enhancement only on IR frames. This saves significant processing time (3.1 ms).

Table VI shows the minimum hardware requirement to achieve real-time performance in terms of the average frames per second (fps). The proposed algorithm was further tested on two low-end hardware. It runs at 21 fps and 13 fps on Intel Core i3 and Intel Quad-core CPUs respectively, which is fit for the real-time requirement of the i-LIDS benchmark for video surveillance systems (12 fps) [25].

IV. CONCLUSION

This paper proposed an enhanced unsupervised change detector for industrial sterile zone monitoring. Its ability to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3013747, IEEE Transactions on Industrial Electronics

be integrated with other change detectors show promising prospects. It was tested on three databases, 45 videos, and more than 100,000 video sequences. It outperformed top-ranked change detection algorithms with real-time performance. It improves other change detector's performance to the IR camera. Also, it improves their overall performance on the change detection dataset from 2-5%. The proposed enhancements are light-weight and only contribute to 28-34% of total processing time. The authors wish to integrate the proposed algorithm in high-level video surveillance tasks such as anomaly detection and abandoned object detection.

References

[1] A. Ullah, K. Muhammad, J. Del Ser, S. W. Baik, and V. H. C. de Albuquerque, "Activity recognition using temporal optical flow convolutional features and multilayer lstm," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9692–9702, 2019.

[2] Wahyono and K. Jo, "Cumulative dual foreground differences for illegally parked vehicles detection," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2464–2473, 2017.

[3] Wahyono, A. Filonenko, and K. Jo, "Unattended object identification for intelligent surveillance systems using sequence of dual background difference," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 6, pp. 2247–2255, 2016.

[4] A. Filonenko, D. C. Hernandez, and K. Jo, "Fast smoke detection for video surveillance using cuda," *IEEE Transactions on Industrial Informatics*, vol. 14, DOI 10.1109/TII.2017.2757457, no. 2, pp. 725–733, Feb. 2018.

[5] A. Shahbaz, L. Kurnianggoro, Wahyono, and K.-H. Jo, "Recent Advances in the Field of Foreground Detection: An Overview," in *Advanced Topics in Intelligent Information and Database Systems*, pp. 261–269. Springer International Publishing, 2017.

[6] T. Bouwmans, S. Javed, M. Sultana, and S. K. Jung, "Deep neural network concepts for background subtraction:a systematic review and comparative evaluation," *Neural Networks*, vol. 117, pp. 8 – 66, 2019.

[7] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *Computer Vision and Pattern Recognition. IEEE Computer Society Conference on.*, vol. 2, p. 252 Vol. 2, 1999.

[8] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Pattern Recognition. ICPR. Proceedings of the 17th International Conference on*, vol. 2, pp. 28–31 Vol.2, Aug. 2004.

[9] P. KaewTraKulPong and R. Bowden, *An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection*, pp. 135–144. Springer US, 2002.

[10] Y. Chen, J. Wang, and H. Lu, "Learning sharable models for robust background subtraction," in *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, Jun. 2015.

[11] H. Sajid and S. C. S. Cheung, "Background subtraction for static amp; moving camera," in *IEEE International Conference on Image Processing (ICIP)*, pp. 4530–4534, Sep. 2015.

[12] P. L. St-Charles, G. A. Bilodeau, and R. Bergevin, "Subsense: A universal change detection method with local adaptive sensitivity," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 359–373, Jan. 2015.

[13] P. L. St-Charles, G. Bilodeau, and R. Bergevin, "Universal background subtraction using word consensus models," *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4768–4781, Oct. 2016.

[14] S. Jiang and X. Lu, "Wesambe: A weight-sample-based method for background subtraction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2105–2115, Sep. 2018.

[15] R. Wang, F. Bunyak, G. Seetharaman, and K. Palaniappan, "Static and moving object detection using flux tensor with split gaussian models," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 420–424, Jun. 2014.

[16] S. Bianco, G. Ciocca, and R. Schettini, "Combination of video change detection algorithms by genetic programming," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 6, pp. 914–928, 2017.

[17] J. H. G. Zuluaga, A. Gomez, A. Salazar, and A. D. Pulido, "Camera-trap images segmentation using multi-layer robust principal component analysis," *CoRR*, vol. abs/1701.08180, 2017.

[18] G. Liu and S. Yan, "Active subspace: Toward scalable low-rank learning," *Neural Computation*, vol. 24, no. 12, pp. 3371–3394, Dec. 2012.

[19] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis," *J. ACM*, vol. 58, DOI 10.1145/1970392.1970395, no. 3, pp. 11:1–11:37, Jun. 2011.

[20] D. Sakkos, H. Liu, J. Han, and L. Shao, "End-to-end video background subtraction with 3d convolutional neural networks," *Multimedia Tools and Applications*, vol. 77, no. 17, pp. 23 023–23 041, Sep. 2018.

[21] M. Braham and M. V. Droogenbroeck, "Deep background subtraction with scene-specific convolutional neural networks," in *International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 1–4, May. 2016.

[22] M. Babaee, D. T. Dinh, and G. Rigoll, "A deep convolutional neural network for video sequence background subtraction," *Pattern Recogn.*, vol. 76, no. C, pp. 635–649, Apr. 2018.

[23] Y. Wang, Z. Luo, and P.-M. Jodoin, "Interactive deep learning method for segmenting moving objects," *Pattern Recogn. Lett.*, vol. 96, no. C, pp. 66–75, Sep. 2017.

[24] M. Braham, S. Piérard, and M. Van Droogenbroeck, "Semantic background subtraction," in *IEEE International Conference on Image Processing (ICIP)*, DOI 10.1109/ICIP.2017.8297144, pp. 4552–4556, Beijing, China, Sep. 2017. [Online]. Available: http://hdl.handle.net/2268/213419

[25] H. O. S. D. Branch, "Imagery library for intelligent detection systems (i-lids)," in *Crime and Security. The Institution of Engineering and Technology Conference on*, pp. 445–448, Jun. 2006.

[26] Y. Wang, P. M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "Cdnet 2014: An expanded change detection benchmark dataset," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 393–400, Jun. 2014.

[27] H. Lin, J. Chuang, and T. Liu, "Regularized background adaptation: A novel learning rate control scheme for gaussian mixture modeling," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 822–836, 2011.

Ajmal Shahbaz (S'16–M'20) Ajmal Shahbaz received the Bachelors of Electronics Engineering from Islamia University of Bahawalpur, Bahawalpur, Pakistan in 2014. He is working towards the Ph.D degree in Electrical Engineering, University of Ulsan, Ulsan, South Korea. His research interests include computer vision, image processing, and deep learning.

Kang-Hyun Jo (M'96–SM'16) Kang-Hyun Jo received the BS degree from Busan National University, Korea, 1989 and MS. and Ph.D. degrees in Computer Controlled Machinery from Osaka University, Japan, in 1993 and 1997, respectively. After a year of experience at ETRI as a postdoctoral research fellow, he joined the School of Electrical Engineering, University of Ulsan, Ulsan, Korea. He has served as a director or an AdCom member of Institute of Control, Robotics and Systems, The Society of Instrument and Control Engineers, and IEEE IES Technical Committee on Human Factors Chair, AdCom member, and the Secretary until 2019. Currently, he is serving as Faculty Dean of School of Electrical Engineering, University of Ulsan. He has also been involved in organizing many international conferences such as International Workshop on Frontiers of Computer Vision, International Conference on Intelligent Computation, International Conference on Industrial Technology, International Conference on Human System Interactions, and Annual Conference of the IEEE Industrial Electronics Society. At present, he is an Editorial Board Member for the International Journal of Control, Automation, and Systems and a Vice President for Membership of ICROS. His research interests include computer vision, robotics, autonomous vehicle, and ambient intelligence.