Journal of Electronic Imaging

SPIEDigitalLibrary.org/jei

Fusion of thermal- and visible-band video for abandoned object detection

Cigdem Beyan
Ahmet Yigit
Alptekin Temizel

Beyan C, Yigit A, Temizel A; Fusion of thermal- and visible-band video for abandoned object detection.
J. Electron. Imaging. ;20(3):033001 (2011) doi:10.1117/1.3602204.

http://dx.doi.org/10.1117/1.3602204

Fusion of thermal- and visible-band video for abandoned object detection

Cigdem Beyan
Ahmet Yigit
Alptekin Temizel
Middle East Technical University
Graduate School of Informatics
Ankara, 06531, Turkey
E-mail: atemizel@ii.metu.edu.tr

Abstract. *Timely detection of packages that are left unattended in public spaces is a security concern, and rapid detection is important for prevention of potential threats. Because constant surveillance of such places is challenging and labor intensive, automated abandoned-object–detection systems aiding operators have started to be widely used. In many studies, stationary objects, such as people sitting on a bench, are also detected as suspicious objects due to abandoned items being defined as items newly added to the scene and remained stationary for a predefined time. Therefore, any stationary object results in an alarm causing a high number of false alarms. These false alarms could be prevented by classifying suspicious items as living and nonliving objects. In this study, a system for abandoned object detection that aids operators surveilling indoor environments such as airports, railway or metro stations, is proposed. By analysis of information from a thermal- and visible-band camera, people and the objects left behind can be detected and discriminated as living and nonliving, reducing the false-alarm rate. Experiments demonstrate that using data obtained from a thermal camera in addition to a visible-band camera also increases the true detection rate of abandoned objects.* © *2011 SPIE and IS&T.* [DOI: 10.1117/1.3602204]

1 Introduction

In public spaces, such as shopping malls and airports, surveillance system operators often watch a high number of cameras simultaneously to control the security of the environment and these systems are left unattended occasionally. This results in security lapses because criminals might leave suspicious items in the environment that may cause catastrophic events. Therefore, detecting suspicious items on time is crucial to providing the security of such places.

In recent years, several studies have been done to detect abandoned items automatically by using computer-assisted systems. In such systems, the most important issue is attaining low false-alarm rates while not missing the real alarms, because false alarms might render the system ineffective by causing the operators to ignore these alarms.

An abandoned-object–detection method using information coming from multiple cameras to generate alerts when objects move away from each other is given in Ref. 1. In this study, the motion is detected using background subtraction and the ground-truth homography is utilized to handle occlusions. Although the system is simple, it suffers to track fast-moving objects, which may cause false alarms if that object is the owner of the abandoned item. A study combining moving- object tracking with drop-off–event–detection and stationary-object–detection methods in crowded environments is proposed in Ref. 2. For reliable detection, the system requires setting many parameters to detect drop-off events, stationary objects, and objects left unattended. An abandoned-object–detection approach with owner tracking, which can be used in indoor environments, is proposed in Ref. 3. In this study, the owner of the object is determined when the object and owner split, which results in a false owner association and cause a false alarm in the case of occlusion of the owner or the abandoned object. Besides, two people who are entering the scene together and then splitting may cause a false alarm because one of them will be detected as an abandoned object. Additionally, discrimination of an object using its size information is not always reliable because a piece of luggage and, for instance, a child's size could be very similar. A study for detecting and recognizing abandoned objects is presented in Ref. 4. In this study, background change detection based on wavelet coefficients is used to detect objects and histogram gradients are utilized to recognize the objects, while support vector machines are employed for supervised learning. An unattended and stolen object detection algorithm, which includes fusion of color and shape information of static foreground objects, is presented in Ref. 5. In Ref. 6, a study that uses long- and short-term background models to detect abandoned objects is described. This method is a pixel-based method, and each pixel is classified as part of a moving object, an abandoned object, an uncovered background, or a scene background according to changes in the short- and long-term foreground images. Apart from these studies, there are several commercial products aiming to detect abandoned objects.[7,8] Even though all these methods are useful for detecting stationary objects, false detections could still occur. Although the object interaction-based analysis might falsely detect abandoned items when two people move away from each other, stationary-object-detection–based algorithms detect all the stationary objects as abandoned items. For example, people standing steadily or sitting on a bench are also potentially detected as abandoned objects. In such cases, by performing

Paper 10136R received Aug. 4, 2010; revised manuscript received May 6, 2011; accepted for publication Jun. 1, 2011; published online Jul. 8, 2011.

1017-9909/2011/20(3)/033001/12/$25.00 © 2011 SPIE and IS&T

living/nonliving object discrimination, the number of false alarms could be reduced.

Thermal- and visible-band cameras are separately widely used for surveillance. However, both types of cameras have some shortcomings. For instance, lighting changes, shadows, or darkness may cause problems in visible-band camera and may bring false-positive detections. Thermal cameras, on the other hand, are not affected by lighting changes and illumination, but they have a lower signal-to-noise ratio than visible cameras and they sometimes cannot detect objects when an object's thermal properties are similar to the environment's thermal properties. The "halo effect," which appears around very hot or dark objects, is another disadvantage of thermal video.[9] Additionally, thermal reflection is a different type of problem. Wet surfaces, glass, and metals reflect infrared radiation and may cause false alarms while detecting or tracking objects when only thermal technology is used. Although visible video and thermal video have their own limitations, the fusion of these two kinds of data in surveillance systems is encouraging to overcome the drawbacks and to obtain more robust systems. As an example, Ref. 10 proposes a study on fusion of thermal- and visible-band images for detecting and searching objects. A surveillance system that fuses thermal video with visible spectrum video for pedestrian detection and tracking by using rule-based decisions and heuristics is presented in Ref. 11. Another approach uses fuzzy logic and Kalman filtering in the fusion step in order to detect moving objects.[12] Studies focus on a method that uses closed-circuit televison (CCTV) and thermal image fusion for object segmentation and tracking is presented in Refs. 13 and 14. In these studies, objects are tracked separately in the thermal and visible domains and this information is then fused by using transferable belief model.

Background subtraction, which helps to detect changes in environment and to track the moving objects in the environment, is an important component to providing a robust system. There are several background-subtraction techniques in the literature. Running Gaussian average,[15] temporal median filter,[16] mixture of Gaussians,[17] kernel density estimation[18] and eigenbackgrounds[19] are just a few examples. The Codebook approach[20] represents pixel samples as a set of code words. This approach is reported to work well on moving backgrounds and under illumination changes. A background-subtraction technique by using contour-based fusion of thermal and visible data for object detection is proposed in Ref. 21. In this study, background subtraction is applied independently in thermal and visible videos. For thermal video, a statistical method is applied while color and intensity information are used in the visible domain. Then, the input and the background gradient information are combined to obtain the object silhouette. Although this study has a higher performance over visible- or thermal-only studies, it is computationally expensive.

In the literature, studies using the thermal- and visible-band fusion do not aim to detect abandoned objects. Recently, to overcome the limitations of visible-band video, a method to detect abandoned or removed object by using video and passive infrared (PIR) sensor fusion has been proposed.[22]

In this study, we propose a method that could be used to support surveillance system operators in indoor environments, such as airports, railway stations, or metro stations, by providing an alarm when abandoned objects are detected. Different from the other studies in the literature, this method is also capable of classifying objects using their heat signatures as living and nonliving. In this way, false alarms caused by humans remaining stationary or sitting for a while are prevented. In Sec. 2, detailed information about the proposed method is given. In Sec. 3, information about the test environment and results are detailed. Conclusions and future improvements are presented in Sec. 4.

2 Proposed Method

The main steps of the proposed method are illustrated in Fig. 1. As shown in Fig. 1, background-subtraction, abandoned-object detection, and post processing are applied to both thermal and visible data to extract abandoned objects. Then, the abandoned-object information obtained from both modalities is fused. Living-object extraction is done by processing the thermal images. Finally, the abandoned-object results are used in combination with the living-object results to eliminate any false alarms due to stationary people. These steps are described in more detail below in Secs. 2.1–2.5.

2.1 *Background-Subtraction and Abandoned Object Detection*

The improved adaptive Gaussian model proposed in Ref. 23 is a technique that has been reported to produce reliable back-

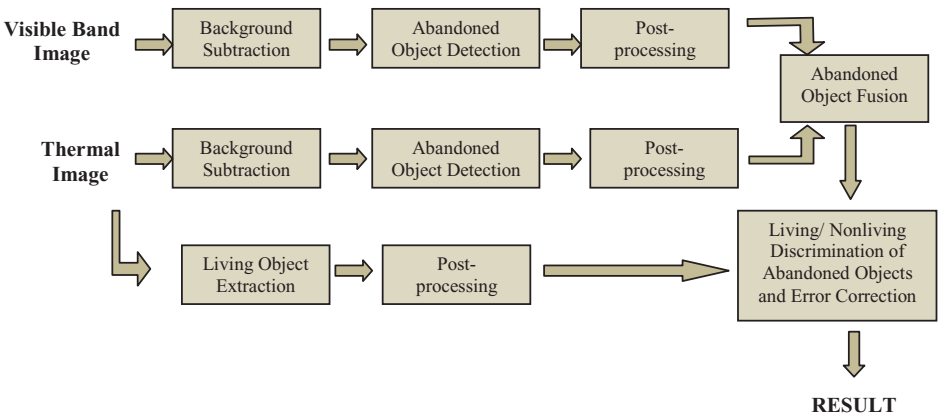

Fig. 1 Block diagram of the proposed system.

ground information while being computationally not very complex. This is a pixel-based method, and each pixel is defined as a mixture of Gaussians with M components as follows:

$$\hat{p}\left(\vec{x}|X_T, \mathrm{BG}+\mathrm{FG}\right)=\sum_{m=1}^{M}\hat{\pi}_m N\left(\vec{x};\hat{\vec{\mu}}_m,\hat{\sigma}_m^2 I\right), \tag{1}$$

where $x^{(t)}$ is the value of pixel at time t, $X_T=\{x^{(t)},\ldots,x^{(\mathrm{t-T})}\}$ is the training set at time t while T is the time period, BG is the background, FG is the foreground, μ_1, μ_2, ..., μ_M and σ_1, σ_2,...., σ_M are the estimates of mean and variance for the Gaussian components respectively. π_1, π_2, ..., π_M are the weight values that are non-negative and summation is equal to 1. The parameters of the model should be updated with new samples to adapt to the changes in background. Equations (2)–(4) show how Gaussian model parameters are being updated.[23]

$$\hat{\pi}_m \leftarrow \hat{\pi}_m+\alpha\left(o_m^{(t)}-\hat{\pi}_m\right), \tag{2}$$

$$\hat{\vec{\mu}}_m \leftarrow \hat{\vec{\mu}}_m+o_m^{(t)}\left(\alpha/\hat{\pi}_m\right)\vec{\delta}_m, \tag{3}$$

$$\hat{\sigma}_m^2 \leftarrow \hat{\sigma}_m^2+o_m^{(t)}(\alpha/\hat{\pi}_m)\left(\vec{\delta}_m^T\vec{\delta}_m-\hat{\sigma}_m^2\right), \tag{4}$$

where $\delta_m=x^{(t)}-\mu_m$, $o^{(t)}$ is an ownership, and α is learning parameter, approximately $\alpha=1/T$, T is the time period. For each component, m is set to 1 if its close component to largest π_m and the others are set to 0. New sample is close to the component if the Mahalanobis distance between them is $<$ st dev. Square distance from the m'th component can be calculated by using.[23]

$$D_m^2\left(\vec{x}^{(t)}\right)=\vec{\delta}_m^T\vec{\delta}_m/\hat{\sigma}_m^2. \tag{5}$$

If the new sample is close to the component, then the new sample belongs to the 99% confidence level and can be determined as a part of the foreground. Because this method does not use a fixed number of components, it is more adaptive and robust when compared to the mixture of Gaussian methods and could automatically select the proper number of components per pixel and update the parameters. In this study, we use this technique because of these advantages.

To detect abandoned objects, we adopted the method proposed in Ref. 6 where a dual foreground approach utilizing two background models: one long term and one short term. In this method, these two background models are initialized with the same parameters except the learning parameter. The short-term background model should have a higher learning parameter so that it could update the background faster. On the other hand, the long-term background model should have a lower learning parameter in order to update the background more slowly.

Abandoned objects are temporally static objects in the background. Therefore, pixels of abandoned objects are detected as background for the higher learning rate and detected as foreground in the shorter learning rate.[6]

$$(x,y)\,\text{is a pixel of}=\begin{cases}\text{moving object}, & F_{\mathrm{L}}(x,y)=1\ \Lambda\ F_{\mathrm{s}}(x,y)=1\\ \text{abandoned object}, & F_{\mathrm{L}}(x,y)=1\ \Lambda\ F_{\mathrm{s}}(x,y)=0\\ \text{uncovered backgorund object}, & F_{\mathrm{L}}(x,y)=0\ \Lambda\ F_{\mathrm{s}}(x,y)=1\\ \text{scene background object}, & F_{\mathrm{L}}(x,y)=0\ \Lambda\ F_{\mathrm{s}}(x,y)=0\end{cases}, \tag{6}$$

where F_{L} is the long-term foreground model and F_{S} is the short-term foreground model.

At every frame, we update two backgrounds B_{L} and B_{S} and estimate two foregrounds F_{L} and F_{S}. F_{L} includes moving objects and abandoned objects, while F_{S} includes moving objects as well as some noise. To extract abandoned objects from the background, the rules given in Eq. (6) are used.

By using two binary foreground images, which are obtained from the previous operation, the evidence image (which is used to decide whether a pixel belongs to an abandoned object or not) are obtained. Evidence images are created using[6]

$$E(x,y)=\begin{cases}E(x,y)+1 & F_{\mathrm{L}}(x,y)=1\wedge F_{\mathrm{s}}(x,y)=0\\ E(x,y)-k & F_{\mathrm{L}}(x,y)\neq 1\vee F_{\mathrm{s}}(x,y)\neq 0\\ \max_{\mathrm{e}} & E(x,y)>\max_{\mathrm{e}}\\ 0 & E(x,y)<0\end{cases}, \tag{7}$$

where $E(x,y)$ is the pixel value at (x,y), k is the decay constant, and $\max_{\mathrm{e}}$ is the maximum value that a pixel value can have in the evidence image. If the pixel belongs to foreground object in the long-term background and background in the short-term background, then the corresponding pixel value of the evidence image is increased. Otherwise, the corresponding pixel value is decreased with the decay constant. If the corresponding pixel value reaches to $\max_{\mathrm{e}}$, then it is decided that this pixel belongs to an abandoned object. Example evidence images for thermal and visible data are shown in Fig. 2.

This method does not require any object-initialization or object-tracking facility and it only contains pixelwise op-

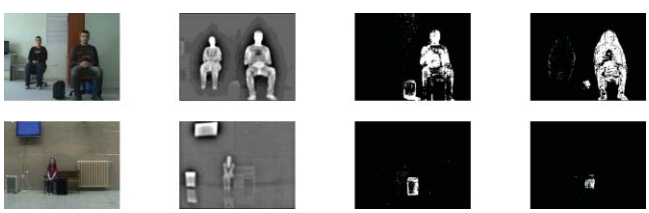

Fig. 2 (a) Visible band image, (b) corresponding thermal image, (c) visible evidence image, and (d) thermal evidence image.

Fig. 3 (a) Thermal images and (b) result of the living-object detection.

erations. Moreover, it could accurately find the boundary of items although they are fully or partially occluded, and is also successful in crowded scenes in which many abandoned-object–detection algorithms fail.[6] As is demonstrated in Ref. 6, this method is successful when the abandoned item is placed under a shadowed area under the table, to detect very small abandoned items and when there is significant motion in the scene.

2.2 *Living Object Extraction*

After detection of abandoned objects, the objects are labeled as living or nonliving. Because the feature that is used to discriminate objects as living or not cannot be obtained from visible data heat signature information from thermal images is used.

Thermal domain images are constructed from energy emitted by objects, and living objects emit more energy compared to nonliving objects. Hence, pixels of living objects appear brighter than pixels of nonliving objects (in a white-hot setting). The study in Ref. 24 used this fact to extract brighter pixels from gray-scale images. We also utilized local intensity operation (LIO) in a similar fashion that brightens the bright pixels and darkens the dark pixels.

According to this method, $I(x,y)$ is given as a pixel in a thermal image written as z_0, and neighbors of it $I(x-1,y-1)$, $I(x-1,y)$, $I(x-1,y+1)$, $I(x,y-1)$, $I(x,y+1)$, $I(x+1,y-1)$, $I(x+1,y)$, $I(x+1,y+1)$ are written as z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8, respectively. Then, Z will be the product of the neighboring pixels,

$$Z = \prod_{k=0}^{8} Z_k. \tag{8}$$

A new image is created according to Z for each pixel in the thermal image by defining intensity brightness operation by using

$$g(x, y) = Z, \tag{9}$$

where $g(x, y)$ is the pixel value at (x, y) of the new image.

After word these image pixels are normalized to the gray-scale range. The normalization process was done by dividing these pixels into the maximum pixel value within this image.

Although, this operation increases the brightness of bright pixels and the darkness of the dark pixels to get better results, we segmented this new image by using entropy thresholding.[25] In this thresholding method, background entropy and foreground entropy are calculated individually as given in following equation for each gray level:

$$H_b(T) = -\sum_{g=0}^{T} \frac{p(g)}{P(T)} \log \frac{p(g)}{P(T)},$$
$$H_f(T) = -\sum_{g=T+1}^{G} \frac{p(g)}{P(T)} \log \frac{p(g)}{P(T)}, \tag{10}$$

where $H_b(T)$ represents the entropy of the background, and $H_f(T)$ represents the entropy of the foreground. G is the maximum gray-level value in the image, T is the threshold value, $p(g)$ is the probability mass function, and $P(T)$ are the background and foreground probabilities.

Then, the sum of the background entropy and foreground entropy is calculated. Once this step is done for all gray levels, the maximum entropy is used as the threshold value and the image is converted to binary using this threshold.

Example results are shown in Fig. 3. This algorithm may not find the object precisely, and some gaps may observed due to the clothing. These problems are rectified with post-processing explained in Sec. 2.3.

It must be noted that, besides the living objects, hot objects such as heating systems, radiators, or any nonliving objects that are hotter than the environment, are also captured brighter than the other objects [such as TV and radiator in Fig. 3(a)]. However, because such objects belong to the background, in this method they are not detected as living or nonliving abandoned objects and false alarms are prevented.

2.3 *Postprocessing*

It is expected that there are some holes in binary images obtained in Secs. 2.1 and 2.2. To make these objects a single piece, we must to complete and close these holes in binary images by using some morphological operations. First, objects in binary images were completed by hole filling. Then, these binary objects were closed. Different from binary images of visible data, binary images of thermal data were eroded to reduce the halo effect around the objects.

2.4 *Fusion of Abandoned Object Results from Different Modalities*

The same methods—background subtraction, abandoned-object detection, and post processing—are applied with the same parameters to both visible and thermal images. In the background-subtraction step, by using two background models (long- and short-term background) two foreground models for each modality are obtained separately. Then, evidence images of thermal- and visible-band data are produced to detect abandoned objects. Next is the postprocessing step,

which is mentioned in Sec. 2.3. After the postprocessing step, binary evidence images of thermal and visible band are fused according to

$$R(x,y)=\begin{cases}0, & I_T(x,y)=0 \wedge I_V(x,y)=0\\ 1, & \text{otherwise}\end{cases}, \tag{11}$$

where $R(x,y)$ is the fusion result, $I_T(x, y)$ is pixel value at (x, y) of binary evidence thermal image and $I_V(x, y)$ is pixel value at (x, y) of binary evidence visible image.

2.5 *Living/Nonliving Discrimination of Abandoned Objects and Error Correction*

As a result of Sec. 2.4, binary mask images for the stationary objects are obtained (including both abandoned and living objects). The next step is discrimination of abandoned objects as living or not using the heat signatures. An abandoned object can be determined as alive or not by fusing the abandoned-object mask, which is obtained in Sec. 2.4 and the living-object mask, which is obtained in Sec. 2.2. To combine features coming from both the thermal and visible domains and classify an object as a nonliving or living abandoned object, Eq. (12) is used,

$$G(x,y)=\begin{cases}\text{nonliving abandoned object}, & E_s(x,y)\neq 0 \wedge F_s(x,y)=0\\ \text{living abandoned object}, & E_s(x,y)\neq 0 \wedge F_s(x,y)\neq 0\end{cases}, \tag{12}$$

where $G(x, y)$ is the fusion result, $E_s(x, y)$ is the pixel value of evidence image and $F_s(x, y)$ is the pixel value of living object mask image. An example fusion result is shown in Fig. 4, in which abandoned objects and other stationary objects are shown in different shades.

As is seen in Fig. 4, there are discrimination errors around the person. These errors may occur due to inaccuracies in registration of the thermal- and visible-band images. Because it is known that all pixels of objects are connected to each other with a neighboring relationship, a connected-component labeling with eight-connectivity can be used. As a result of connected-component labeling, objects and also errors are separated. Then, bounding boxes that surrounds the objects are found. To eliminate errors around a person, the density of objects in its rectangle is calculated using

$$D = N/A_{\text{rect}}, \tag{13}$$

where D is the density of object, N is the total number of pixels that the object has, and A_{rect} is the area of the bounding rectangle. The object was only accepted if its density was greater than the density threshold. The corrected result is shown in Fig. 4(d).

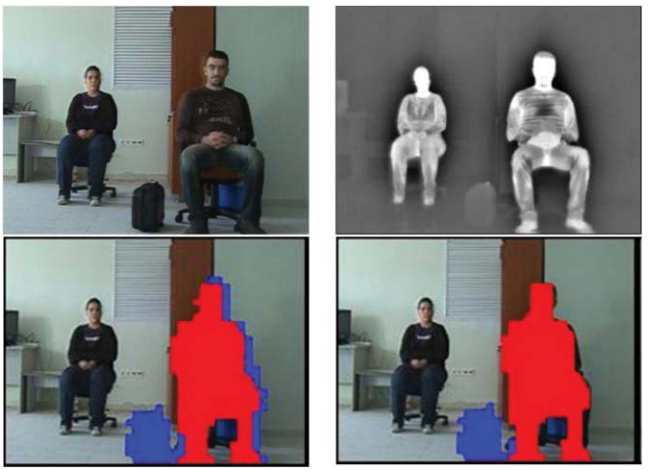

Fig. 4 (a) Visible-band image, (b) corresponding thermal image, (c) result of living/nonliving discrimination with error, and (d) corrected result of living/nonliving discrimination.

3 Test Results

3.1 *Data Sets*

Some video sequences included in the i-LIDS bag-detection dataset[26] (AB-Easy, AB-Medium, and AB-Hard) have been used to test the abandoned-object detection algorithm in Ref. 6. Moreover, the i-LIDS bag-detection data set,[26] which contains crowded, real-time metro video sequences, has been used to test the proposed method by simulating the thermal images, because this data set does not contain thermal-band sequences. In the literature, there is no available data set covering both modalities to detect unattended objects. Furthermore, 11 videos that reflect different scenarios in an indoor environment have been captured simultaneously in both the thermal and visible domains to test the proposed method. The proposed method has been tested with various sizes and colors of bags, such as black luggage; white, black, and red handbags, and dark blue backpack, and the distance between the field of view and cameras has been varied. To test the performance of the proposed method while thermal reflection exists, videos have been captured in different places having high and low reflectance floors. In addition, the algorithm has been tested in crowded scenes and when multiple occlusions exist. To prove the proposed method is not affected from hot objects in the scene, such as heating system, radiators, etc., the method has also been tested in such environments. Table 1 illustrates the details of the scenarios used in the evaluation. Video sequences are publicly available at http://ii.metu.edu.tr/content/abandoned-object-detection.

To capture the scenarios, for the thermal-video sequence, an OPGAL Eye-R640 uncooled infrared camera, which captures 25 fps at 320 × 240 resolution, and, for the visible-band sequence, a Sony HDR-HC1 camera that also captures 25 fps at 320 × 240 images, were used.

Table 1 Details of the test videos used in the evaluation. All video sequences have resolution of 320 × 240 pixels and captured at 25 fps.

Scenario	No. of frames	Approx. Distance between the objects of interest and the cameras (m)	No. of living objects	No. of nonliving objects	General scenario description
Set 1	1500	1.5	2	1	Handbag left unattended
					Low reflectance floor
Set 2	2300	1.5	2	0	No alarm case
					Low reflectance floor
Set 3	2340	1.5	2	1	Occlusion
					Handbag left unattended
					Low reflectance floor
Set 4	2151	2.5	2	1	Handbag left unattended
					Contrast between the background and the handbag is low
					Low reflectance floor
Set 5	2051	2.5	2	1	Handbag left unattended
					Contrast between the background and the handbag is low
					Low reflectance floor
Set 6	1714	2.5	1	1	Occlusion
					Handbag left unattended when contrast between the background and the handbag is low
					High reflectance floor
Set 7	1576	4	>4	1	Multiple occlusion
					Luggage left unattended
					High reflectance floor
Set 8	1508	4	>3	1	Multiple occlusion
					Backpack left unattended
					High reflectance floor
Set 9	1755	4	>3	1	Multiple occlusion
					Handbag left unattended
					High reflectance floor
Set 10	1268	4	1	2	Heating system on
					Luggage left unattended
					High reflectance floor
Set 11	1763	4	>3	2	Heating system on
					Luggage left unattended
					High reflectance floor

3.2 *Results*

In this study, detection of abandoned objects is the goal; hence, true alarms are defined as detection of nonliving abandoned objects at correct positions. The correctness and completeness of the abandoned living or nonliving object detections were not critical because the system aims to support surveillance system operators by immediately providing an alarm and marking the area when a suspicious object is detected.

For the tests, first, images captured from thermal and visible cameras were registered. Both thermal- and visible-band cameras are adjusted properly to capture a similar field of view (FOV). However, it is not practically possible to capture exactly the same field of view for both thermal- and visible-band cameras, because these cameras have different parameters (such as different sensor types and lenses). Therefore, a crop operation was performed for both thermal- and visible-band frames to set almost the same FOV for both thermal and visible images. Then, homography was performed manually by selecting reference points in both the thermal and visible domains for image registration. To find the corresponding pixels of each pixel, an homography matrix was constructed. To obtain the homography matrix, Eqs. (14) and (15) and reference points selected from both thermal and visible images were used,

$$V_{\text{ref}} = H \times T_{\text{ref}} \tag{14}$$

$$H = V_{\text{ref}} \times T_{\text{ref}}^{-1} \tag{15}$$

where V_{ref} is the reference point matrix for visible domain, T_{ref} is the reference point matrix for thermal domain, and H is the homography matrix for registration. In this study, 20 reference points were selected for each data set. Once the capture and homography parameters are obtained, these parameters can be used as long as the camera positions are not changed. Therefore, in real-life systems, when the camera positions are fixed and registration parameters are set, the proposed method works without requiring any user intervention.

While performing abandoned-object detection (Sec. 2.1), for the short-term background model, learning rate α was taken as 0.02; on the other hand, for the long-term background model, α was equal to 0.0002. The threshold on the squared Mahalanobis distance was taken as 16, which means 4 standard deviation in order to provide 99% confidence, the Gaussian number was taken as 4, and the initial standard deviation was taken as 11 for both kinds of background models. The density and threshold (in terms of number of pixels) while applying living and nonliving object discrimination and error correction (Sec. 2.5) was selected as 0.4 and 1000, respectively.

For the i-LIDS bag-detection data set,[26] it is assumed that an abandoned object cannot be left in the way that a subway train passes. Hence, this area was masked out as a nondetection area, and false alarms that might be raised when subway trains stop were eliminated. Figure 5 shows the abandoned-object detection algorithm's[6] test results for AB-Easy, AB-Medium, and AB-Hard included in the i-LIDS bag-detection data set.[26] In Fig. 5, objects detected as abandoned are shown. For AB-Easy data, the luggage was detected but two people were also detected as abandoned objects. For AB-Medium data, besides the luggage, the algorithm[6] also found persons sitting on a bench as abandoned objects, and a total of 10 people were detected as abandoned objects. For AB-Hard data, luggage and a total of 11 people who were standing were detected as abandoned items. To sum up, this method[6] is not successful to discriminate the different types of objects; it causes many false alarms because it detects a person who is stationary for a long time as abandoned object.

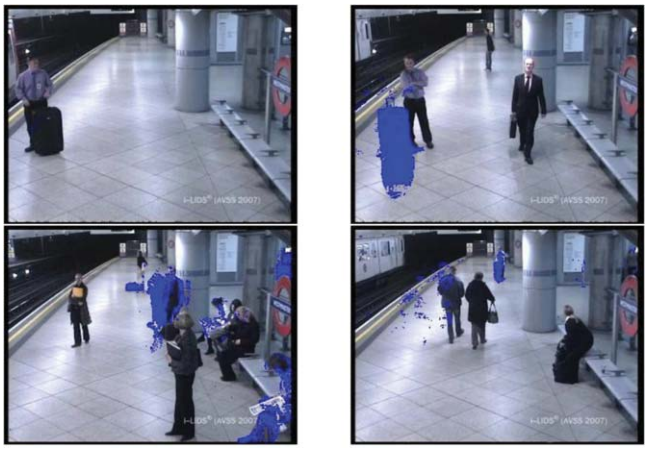

Fig. 5 (a) AB-Easy without detection of abandoned-object, (b) result of abandoned-object detection for AB-Easy, (c) result of abandoned-object detection for AB-Medium, and (d) result of abandoned-object detection for AB-Hard.

An example result for the proposed method is shown in Fig. 4. In this video, the person on the left-hand side was not detected as a suspicious object because she always was in the scene from the beginning and stayed stationary in the environment. Both the person on the right-hand side and his bag were detected as suspicious objects as expected. However, different from Ref. 6, the person was labeled as a living object and, hence, the false alarm due to the sitting person is eliminated. The bag was labeled as nonliving abandoned object.

Figure 6 presents another test result of the proposed method. In this video, the proposed algorithm detected the person as an abandoned object but did not raise an alarm because it is a living object. After a while, when another person comes to the scene and sits with his bag on the side, the algorithm detected both the person and his bag; the bag was labeled as an abandoned nonliving object, and the system raised an alarm. The region between the legs of the person and the bag are detected together as a single abandoned object because they are connected. This inaccurate boundary detection can be corrected by utilizing a shadow-removal algorithm. The sitting person was successfully detected as a living object. Hence, a correct detection of living and abandoned objects is achieved.

Figures 7 and 8 shows the test results when multiple occlusions occur in a high reflectance floor. In Fig. 7, a piece of luggage is left unattended, while in Fig. 8 a smaller backpack is left unattended. In both sequences, all the nonliving abandoned objects were detected correctly and alarms were raised as expected. In Fig. 7, the person was not detected as a living abandoned object because he was not stationary while he was sitting on the bench. On the other hand, in Fig. 8, legs of the sitting person were detected as an abandoned living

Fig. 6 Result of the proposed method when a handbag is left unattended and the contrast between the background and the handbag is low (low-reflectance floor).

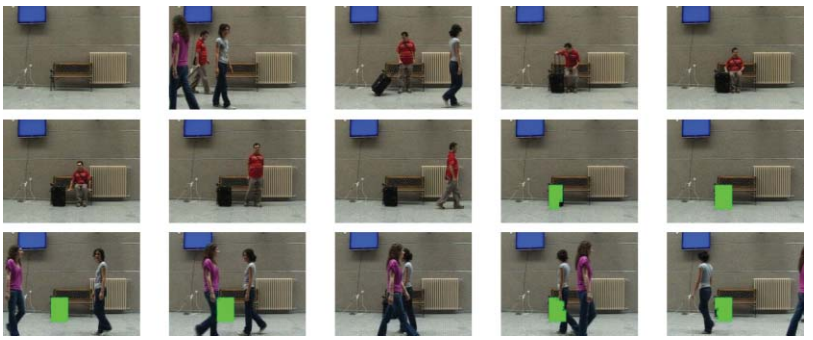

Fig. 7 Result of the proposed method when there are multiple occlusions and luggage is left unattended on the high-reflectance floor.

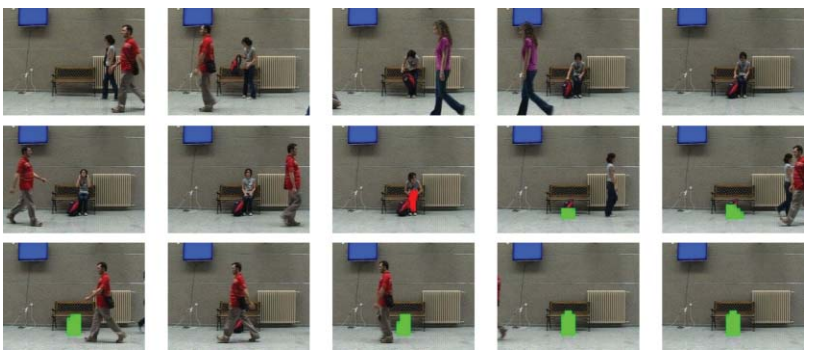

Fig. 8 Result of the proposed method when there is occlusion and a backpack is left unattended on the high-reflectance floor.

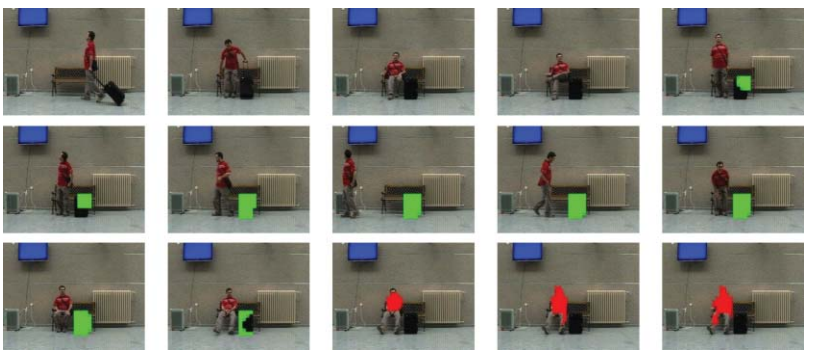

Fig. 9 Result of the proposed method when the heating radiator is on and luggage is left unattended on the high-reflectance floor.

Fig. 10 A handbag is left unattended, when contrast between the background and the handbag is low (floor has low reflectance): (a) Result of the method that does not implement abandoned object detection in thermal data and (b) result of the proposed method.

object because they were stationary. The rest of the body was not detected as abandoned because of the small movements while sitting. The proposed method gives successful results with objects of different sizes. Additionally, it is suitable to detect abandoned objects when the reflectance of the floor is high which causes thermal reflection. The method also handles multiple occlusions due to people walking successfully.

Figure 9 illustrates the proposed method's result with a different data set when there is a hot radiator in the field of view of the cameras. In this video sequence, luggage was detected as a suspicious abandoned object as expected and person was labeled as living object. Although the radiator (see Fig. 3) is segmented as a hot object with the LIO operation (Sec. 2.2), because it is part of the background, both in the visible and thermal domains it does not get detected as an abandoned object. Hence, it can be said that the proposed method does not produce any false alarms due to hot objects in the environment.

The proposed method was also compared to our previous study,[27] where abandoned-object detection in the thermal domain is not implemented. In Ref. 27, abandoned-object detection is done using only the visible band, while the thermal band was only used to discriminate an object as living or nonliving. The results of this work are shown in Fig. 10. When the proposed method and the method in Ref. 27 were compared to each other, in both methods, a person was detected as an abandoned living object and did not cause any false alarm. However, the bag could not be detected as an abandoned nonliving object when the method in Ref. 27 is implemented because the color of the bag is very similar to the background color (the contrast between the object and the background is low), which causes the real alarm to be missed. On the other hand, the proposed approach utilizing abandoned-object detection in both modalities was successful in finding the bag and labeling it as a nonliving object (Fig. 10).

To test the proposed method in real-life scenarios and to compare it to existing studies, such as Refs. 6 and 27, the i-LIDS bag-detection data set,[26] which contains crowded scene video sequences including many stationary living objects and multiple occlusions, has been used. Because this data set does not contain thermal-band sequences and there is no data set covering both modalities for detection of abandoned objects in literature, we simulated the thermal images for the data set. Therefore, background subtraction, abandoned-object detection, and postprocessing have been applied in the visible domain; living-object extraction, which should be performed in the thermal domain, has been applied manually; and these two segmentation results have been used to discriminate living/nonliving abandoned objects and error correction. A sample result is given in Fig. 11.

Similar results were obtained for all the test videos. As a consequence of these results, it can be said that the proposed method decreases the number of false alarms by classifying the objects using their heat signatures while not missing the real alarms. It is successful when there are multiple occlusions and suitable for use in real-life applications. Moreover, the method is not affected from a heating system or hot nonliving objects with temperatures close to or more than the living object in the field of view. It has also been shown that the proposed method increases the detection rate by using the thermal domain background especially when the contrast between the background and the foreground object is low.

To evaluate the performance of the proposed method quantitatively, false/true detection performances and the missed alarms (which may occur when the abandoned-object–detection algorithm could not detect the abandoned object and consequently does not raise an alarm), using data sets of the proposed method listed in Table 1 and the methods of Refs. 6 and 27 are given in Table 2.

As can be seen from Table 2, no false alarm is observed with the proposed method; hence, it provides a lower false alarm rate than the method in Ref. 6. On the other hand, while the method proposed in Ref. 27 misses two real alarms, the proposed method successfully identifies all the alarm cases.

Besides the comparison given in Table 2, the overall precision and recall values using the true, false, and missed alarms have been calculated. While calculating the overall precision and recall values, all video sequences are considered. The following equations are used, where true positive (TP) is the total number of true alarms, false positive (FP) is the total number of false alarms, and false negative (FN) is the total number of missed alarms:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{16}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{17}$$

According to the results the overall precision was found as 1 and overall recall was found as 1 because there are no missed or false alarms; whereas overall precision is calculated as 1 and recall is calculated as 0.48 for study in Ref. 6 due to false alarms, and overall precision is determined as 0.8 and recall is determined as 1 for study in Ref. 27 because it misses some real alarms.

Fig. 11 Simulated result using the i-LIDSAB-Medium data set, which contains a crowded scene including many stationary living objects, multiple occlusions.

Table 2 Performances of the proposed method and methods [6] and [27].

	Detection using dual foregrounds[a]			Detection using thermal and visible band[b]			Proposed method		
Scenario	No. of true alarms	No. of false alarms	No. of missed alarms	No. of true alarms	No. of false alarms	No. of missed alarms	No. of true alarms	No. of false alarms	No. of missed alarms
Set 1	1	0	0	1	0	0	1	0	0
Set 2	0	1	0	0	0	0	0	0	0
Set 3	1	2	0	1	0	0	1	0	0
Set 4	1	2	0	1	0	0	1	0	0
Set 5	1	2	0	0	0	1	1	0	0
Set 6	1	1	0	0	0	1	1	0	0
Set 7	1	0	0	1	0	0	1	0	0
Set 8	1	1	0	1	0	0	1	0	0
Set 9	1	1	0	1	0	0	1	0	0
Set 10	1	0	0	1	0	0	1	0	0
Set 11	1	1	0	1	0	0	1	0	0

[a]Reference 6.
[b]Reference 27.

4 Conclusions and Future Work

In this paper, a system that detects abandoned objects in an indoor environment by using thermal- and visible-band cameras is proposed. The aim of this study is aiding surveillance system operators who are working in indoor environments, such as airports, railway stations, metro stations, and shopping malls by providing an alarm when an abandoned object is detected.

The difference of proposed system from proposed systems in the literature is that it prevents false alarms due to stationary living objects and only generates alarms for abandoned nonliving objects. Using the captured data sets, it has been shown that the algorithm works as expected by filtering out the false alarms due to stationary persons and has high precision and recall values in terms of false, true, and missing alarms. Also, the robustness of the abandoned-object–detection algorithm has been increased by using a thermal domain background in association with the visible domain background. The accuracy and the robustness of the system have been evaluated by using different scenarios having multiple occlusions and different types of objects in the environment while items of different sizes are left unattended.

In future work, the proposed method can be tested in an outdoor environment. On the other hand, a data-fusion step that is currently being applied on the logical level could be improved to increase the completeness of the detected abandoned objects. Aside from this, a tracking utility can be added to the system in order to track living/nonliving objects separately to detect the person who left the object.

Acknowledgments

We thank Fatih Ömrüuzun, Ersin Karaman, Püren Güler, Deniz Emeksiz, and Mustafa Teke for their great help in capturing the test video sequences.

References

1. E. Auvinet, E. Grossmann, C. Rougier, M. Dahmane, and J. Meunier, "Left-luggage detection using homographies and simple heuristics," in *Proc. of IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance*, pp. 51–58 (2006).
2. S. Guler and M. K. Farrow, "Abandoned object detection in crowded places," in *Proc. of IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance*, pp. 99–106 (2006).
3. S. Ferrando, G. Gera, M. Massa, and C. Regazzoni, "A new method for real time abandoned object detection and owner tracking," in *Proc. of IEEE Int. Conf. on Image Processing*, pp. 3329–3332 (2006).
4. R. Miezianko and D. Pokrajac, "Detecting and recognizing abandoned objects in crowded environments," in *Proc. of Computer Vision Systems*, pp. 241–250 (2008).
5. J. C. SanMiguel and J. M. Martínez, "Robust unattended and stolen object detection by fusing simple algorithms," in *Proc. of IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pp. 18–25 (2008).
6. F. Porikli, Y. Ivanov, and T. Haga, "Robust abandoned object detection using dual foregrounds," *EURASIP J. Adv. Signal Process.* **2008**, 1–10 (2008).
7. Nice Vision Video Analytics, http://www.nice.com/products/video/analytics.php, (10 January 2010).
8. Iomniscient Non-Motion Detector, http://iomniscient.com/, (10 January 2010).
9. P. Kumar, A. Mittal, and P. Kumar, "Study of robust and intelligent surveillance in visible and multimodal framework," *Informatica* **32**, 63–77 (2008).
10. D. A. Fay, A. M. Waxman, M. Aguilar, D. B. Ireland, J. P. Racamato, W. D. Ross, W. W. Streilein, and M. I. Braun, "Fusion of multi-sensor imagery for night vision: color visualization, target learning and search," in *Proc. of Int. Conf. on Information Fusion*, pp. 215–219 (2000).
11. H. Torresan, B. Turgeon, C. Ibarra-Castanedo, P. Hébert, and X. Maldague, "Advanced surveillance systems: combining video and thermal imagery for pedestrian detection," *Proc. SPIE* **5405**, 506–515 (2004).
12. P. Kumar, A. Mittal, and P. Kumar, "Fusion of thermal infrared and visible spectrum video for robust surveillance," in *Proc. Indian Conf. on Computer Vision, Graphics and Image Processing*, pp. 528–539 (2006).
13. C. O. Conaire, N. O. Connor, E. Cooke, and A. Smeaton, "Multispectral object segmentation and retrieval in surveillance video," in *Proc. of IEEE Int. Conf. on Image Processing*, pp. 2381–2384 (2006).
14. C. O. Conaire, N. O. Connor, E. Cooke, and A. Smeaton, "Comparison of fusion methods for thermo-visual surveillance tracking," presented at *Int. Conf. on Information Fusion* (2006).
15. C. Wren, A. Azarhayejani, and T. A. P. Darrell, "Pfinder: real-time tracking of the human," *IEEE Trans. Pattern Anal. Mach. Intel.* **19**(7), 780–785 (1997).
16. M. Piccardi, "Background subtraction techniques: a review," presented at *IEEE Int. Conf. on Systems, Man, and Cybernetics* (2004).
17. C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. of IEEE Comput. Vis. Pattern Recogn.*, pp. 246–252 (1999).
18. A. Elgammal, D. Harwood, and L. S. Davis, "Non parametric model for background subtraction," in *Proc. of IEEE ICCV Frame Rate Workshop*, pp. 751–767 (1999).
19. Z. Xu, P. Shi, and I. G. Yu-Hua, "An eigenbackground subtraction method using recursive error compensation," *Lecture Notes in Computer Science, Advances in Multimedia Information Processing — PCM 2006*, pp. 779–787, Springer, New York (2006).
20. K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Background modeling and subtraction by codebook construction," in *Proc. of IEEE Int. Conf. Image Processing*, pp. 3061–3064 (2004).
21. J. W. Davis and V. Sharma, "Fusion-based background-subtraction using contour saliency," in *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 20–26 (2005).
22. M. Magno, F. Tombari, D. Brunelli, L. D. Stefano, and L. Benini, "Multimodal abandoned/removed object detection for low power video surveillance systems," in *Proc. of IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pp. 188–193 (2009).
23. Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. of Int. Conf. on Pattern Recognition*, pp. 28–33 (2004).
24. R. Heriansyah and S. A. R. Abu-Bakar, "Defect detection in thermal image for nondestructive evaluation of petrochemical equipments," *NDT & E Int.* **42**(8), 729–740 (2009).
25. M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *J. Electron. Imaging* **13**(1), 146–165 (2004).
26. i-LIDS bag detection data set for AVSS 2007, http://www.elec.qmul.ac.uk/staffinfo/andrea/avss2007_d.html, (19 January 2010).
27. A. Yigit and A. Temizel, "Abandoned object detection using thermal and visible band image fusion," in *Proc. of IEEE Signal Processing, Communication and Applications Conf.*, pp. 617–620 (in Turkish) (2010).

Cigdem Beyan received her BSc from the Department of Computer Engineering at Baskent University, Turkey, in 2008. Currently, she is an MSc student in the Department of Information Systems, Graduate School of Informatics, Middle East Technical University (METU), Turkey, and works at Department of Computer Engineering at Baskent University as a research assistant. She is also a researcher in the Bioinference Research Group at Baskent University and a member of Virtual Reality and Computer Vision Research Group in METU, Informatics Institute. Her research interests include computer vision, information systems, bioinformatics, machine learning, pattern recognition, and statistical data analysis.

Ahmet Yigit received his BSc from the Department of Computer Engineering at Middle East Technical University (METU), Turkey, in 2005, and MSc from the Department of Information Systems, Informatics Institute, METU, in 2010. Currently, he is a PhD student in the same department and works for HAVELSAN (Turkish Armed Forces Foundation Company) as a specialist software engineer. His research interests are computer vision, image processing, pattern recognition, information systems, and computer networks.

Alptekin Temizel received his BSc in electrical and electronic engineering from Middle East Technical University (METU), Ankara, Turkey (1999) and his PhD from the Centre for Vision, Speech and Signal Processing, University of Surrey, United Kingdom (2006). Between 1999 and 2001, he worked as a research assistant at the University of Hertfordshire, United Kingdom. He cofounded Visioprime Ltd., UK, a company developing intelligent video systems for security and surveillance applications, and worked as a senior research engineer in the company from 2001 to 2006. Since 2006, he has been a lecturer in the Graduate School of Informatics, at METU, and consultant to several research and development companies. He is the principle investigator of the Virtual Reality and Computer Vision Research Group and authored/coauthored several patents and papers on image-video processing and video surveillance. He received the Overseas Research Student Award (UK) in 2000, has a bronze medal from the 1st National (Turkish) Science Olympics, and received the NVIDIA Professor Partnership Award in 2009.