Detection of Abandoned and Stolen Objects Based on Dual Background Model and Mask R-CNN

HYESEUNG PARK[ID], SEUNGCHUL PARK, AND YOUNGBOK JOO
Korea University of Technology and Education, Cheonan-si, South Korea

Corresponding author: Seungchul Park (scpark@koreatech.ac.kr)

ABSTRACT Dual background models have been widely used for detecting stationary objects in video surveillance systems. However, there is a problem that both abandoned and stolen objects are equally detected as stationary objects, making it difficult to distinguish them. Another problem is the ghost region created by shadow shift or light changes, which makes the discrimination issue more complicated. In this paper, we present an efficient method to distinguish abandoned objects, stolen objects, and ghost regions in the surveillance video. This method contains two main strategies: the first one is the dual background model for extracting candidate stationary objects, the second one is object segmentation based on mask regions with CNN features (Mask R-CNN) for providing the object mask information. The basic idea is: given a candidate stationary object from the background model, it is checked whether a corresponding segmented object exists in the current video frame or the previous background frame to take into account the current and past situations. And the final state of the candidate stationary object is determined by considering various situations through the comparative analysis technique presented in this paper. The proposed algorithm has qualitatively experimented with our own dataset focusing on the discrimination issue, which generated satisfactory results. Therefore, it is expected to be widely applied to automatic detection of stolen objects as well as abandoned objects in open environments such as exhibition halls and public parks where existing intrusion detection-based security services are difficult to be deployed.

INDEX TERMS Abandoned object detection, stolen object detection, ghost region, dual background model, mask R-CNN.

I. INTRODUCTION

In recent years, researches on intelligent video surveillance system analyzing video automatically without continuous observation by humans have been actively conducted to provide methods for detecting and notifying the occurrence of specific events such as intrusion, loitering, abandonment, crime, and fire detection. Intelligent surveillance systems can reduce human errors by lowering the dependency on humans. They can also improve the response time by generating alarms as soon as events occur [1], [2]. Although systems based only on traditional computer vision technology have suffered limitations in accuracy or performance, recent advances in artificial intelligence have opened practical ways for improvement in various applications. But neither technique is perfect on its own, how to use them properly for specific applications.

Therefore, we propose a novel approach for detecting abandoned and stolen objects based on the conventional background subtraction and an artificial intelligence technology, i.e., Mask R-CNN. The background subtraction technique creates dynamic backgrounds and foregrounds in real-time [3]. The primary way to detect abandoned and stolen objects is to analyze the foreground and to select stationary objects. In this paper, an abandoned object is the object newly placed on the background and a stolen object represents an uncovered background from which an existing object is removed. The problem is that both are technically similar in that they equally appear as stationary objects in the foreground. Besides, the ghost region issue caused by such as shadow shift or light changes is another problem to concern. Ghost regions or ghosts are meaningless areas

The associate editor coordinating the review of this manuscript and approving it for publication was Gangyi Jiang.

that also appear as stationary objects in the foreground [4] even though they are not abandoned or stolen. Therefore, it is difficult to distinguish abandoned objects, stolen objects, and ghost regions by simply analyzing the foreground. For this reason, we intend to solve this problem using artificial intelligence technology.

The methodology we propose is as follows. First, we use a dual background model to detect stationary objects [5]–[11]. The dual background model consists of two sub-models with different learning rates. The short-term model, which has a high learning rate, quickly absorbs moving pixels into the background when they stop. On the other hand, the long-term model updates the background slowly. So stationary pixels can exist for a long time in the foreground. Therefore, the pixels existing as foregrounds in the long-term model but not in the short-term model are extracted as stationary pixels. Adjacent pixels among the stationary pixels are grouped into a blob and considered as an object. The blobs below a certain size are regarded as noises and removed. Simply determining whether an object is stationary based on a single frame could be premature when we use the dual background model. The object may stop temporarily and move immediately. Therefore, temporal transition information is necessary to identify the stationary foreground based on the sequence pattern of each blob [8]. We track the object to check if it is consistently stationary more than a certain amount of time at the same spot. If so, it is considered 'stable' and designated as a 'candidate stationary object'.

The candidate stationary object can be finally determined as an abandoned object, a stolen object, or a ghost region when certain conditions are satisfied. In this paper, Mask R-CNN [12] which provides object mask data is used to verify whether there is a segmented object in the area of the candidate stationary object. If a segmented object is detected in the current frame, its pixels and the pixels corresponding to the candidate stationary object are compared respectively to determine the equality. If the match score is greater than a certain threshold, it is determined as an abandoned object. Then, how do you explain when a segmented object exists in the current frame but doesn't match the candidate? And what if there is no segmented object in the area of the candidate? To answer these questions, we need to take into account the past situation as well as the present. We trace back to a previous frame (precisely, the long-term background frame) to see if any object existed at that time. There could be a segmented object in the frame. However, it is also possible that the segmented object doesn't match the candidate or it doesn't exist at all. Therefore, we propose a method to determine the final state of the candidate stationary object by considering these various clues.

In summary, we use the dual background model based on the traditional background subtraction technique to extract the pixel region estimated to be a candidate stationary object. And we also use Mask R-CNN, one of the widely used artificial intelligence technologies in the field of object detection, to check whether the area corresponds to a trained object.

Furthermore, we present a method for determining whether the candidate stationary object is abandoned, stolen, or a ghost region, taking into account various situations that may occur in the present and the past.

This paper is organized as follows. In Chapter 2, we analyze several related researches and their problems with the detection of abandoned and stolen objects. In Chapter 3, we explain the proposed framework and algorithm for detecting abandoned, stolen objects and ghost regions in detail. In Chapter 4, we prove our algorithm through various experiments using our own dataset. Finally, we conclude in Chapter 5.

II. RELATED WORKS

As mentioned above, abandoned, stolen objects, and ghost regions are equally detected as stationary objects based on the foreground analysis. Most of the relevant researches have mainly focused on detecting abandoned objects. Most of them did not consider stolen objects and ghost regions or left them as challenges.

Lin *et al.* [8], [9] suggested ways to increase the accuracy of abandoned object detection using the dual background model. They also use a pixel-based finite state machine (PFSM) to track the state of the object and the back-tracing verification to find its owner. However, their research did not solve the problem of tracking long-term abandonment and illumination change issue due to the limitation of the background subtraction. Park *et al.* [10], [11] identifies the position and area of the candidate stationary object through the dual background model. They register the template of the candidate, which is used for comparison in the presence authentication process determining the final state. By reducing the dependency on foreground information, they solved long-term abandoned object tracking, occlusion, and illumination change. Wahyono *et al.* [13] and Filonenko *et al.* [14] detect stationary objects using the difference of a reference background and the current background. The triple background model presented by Cuevas *et al.* [15] solves the problem of long-term abandoned object detection and occlusion by adding a long-term model with no background absorption. However, all the studies based on multiple background models focus on the detection of abandoned objects and do not address the problem of distinguishing whether a candidate stationary object is an abandoned object or a stolen object.

Smeureanu and Ionescu [16] perform stationary object detection based on background subtraction and motion estimation. And they tried to improve the detection accuracy of abandoned objects by applying a cascade of convolutional neural networks (CNN). However, they have the overfitting problem because they trained the sub-areas of the experimental video background as negative classes. Shyam *et al.* [17] construct a dual background model using the sViBe [18] modeling method. They also use the PFSM to keep tracking states of foreground pixels. Besides, they classify stationary objects into suspected objects using Single Shot Multibox Detector (SSD) [19]. The problem to consider is that they

unconditionally regard an object to be abandoned if it exists in the candidate area. However, when an object disappears, it can create a foreground area. If there is another object behind it, the system can wrongly determine that it is an abandoned object, not a stolen object.

There have also been attempts to distinguish between abandoned and stolen objects. Connell *et al.* [20] and Venetianer *et al.* [21] store the background just before a candidate stationary object is generated. If the edge energy of the corresponding area in the current frame is higher than the edge energy of the stored background, it is classified into an abandoned object. Unless, it is determined as a stolen object. This approach is applicable in general situations, but does not take into account the case where the edge energy of the background is higher than the debris. Ghost areas, caused by shadow shifts or light changes, are difficult to distinguish because it is difficult to predict how much they affect the edge energy of the background. Tian *et al.* [22], [23] proposed a region growing method that examines color similarity by extending the area from the inside of the candidate stationary to the outside. The problem with this technique is that detection can fail if the color of the object is similar to the background.

III. PROPOSED ALGORITHM FOR DETECTING ABANDONED OBJECT AND STOLEN OBJECT

A. DUAL BACKGROUND MODEL FOR EXTRACTING STATIONARY OBJECTS

In this paper, the proposed algorithm extracts a stationary object that can be finally determined as an abandoned or a stolen object using a dual background model as shown in Figure 1. The dual background model is composed of two sub-models (short-term model and long-term model) with different learning rates. In the short-term model, pixels of an abandoned bag are quickly absorbed into the background (SB) and disappear in the foreground (SF) at the same time. On the other hand, they remain for a relatively long time in the long-term foreground (LF). Therefore, there is no bag in the long-term background (LB). In summary, the SF contains only moving objects and the LF includes moving objects as well as stationary objects. (To be precise, that moment exists.) We calculate the approximate position, size, and shape of a stationary object using these characteristics. The difference foreground (DF) is the result of subtracting the SF from LF. Active pixels adjacent to each other are grouped into a blob and considered as a stationary object in the DF frame. If its size is smaller than a certain threshold, it is considered noise and filtered out. As a result, only stationary objects remain in the DF frame. By the way, simply determining whether an object is stationary based on a single frame could be premature. Because the object may stop temporarily and move immediately. We must use the temporal transition information to identify a stationary object based on the sequence pattern of each blob in the consecutive frames [8]. So we track the object to check if it is consistently stationary more than
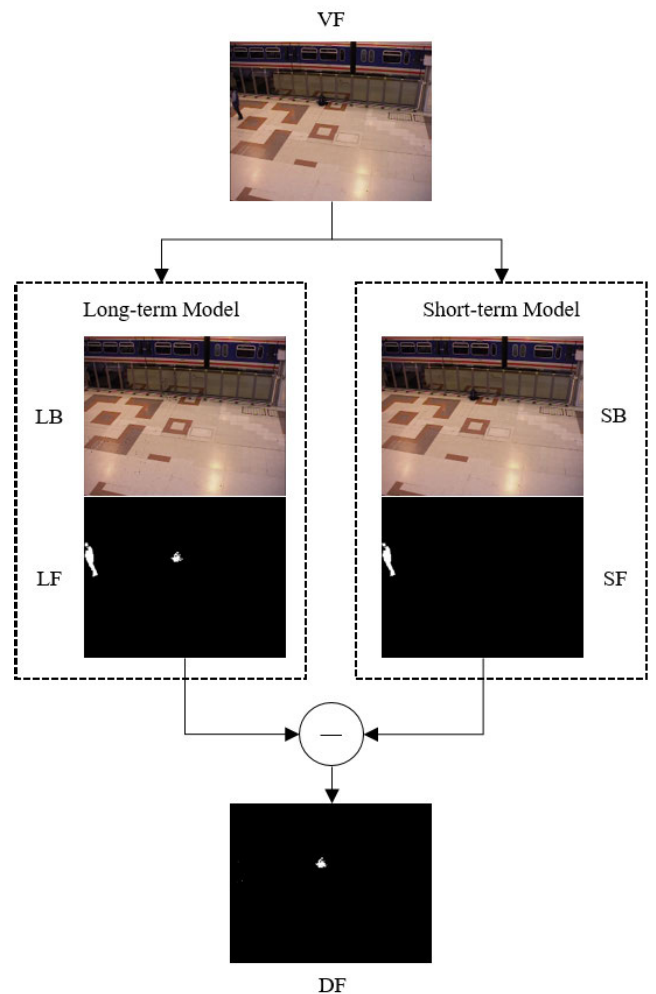

FIGURE 1. Dual background model for extracting stationary objects. The dual background model takes a VF (Someone abandoned a bag) as an input and produces a DF frame as the final result. There are only stationary objects in the DF frame. (VF: Video Frame, SB: Short-term Background, SF: Short-term Foreground, LB: Long-term Background, LF: Long-term Foreground, DF: Difference Foreground).

a certain amount of time at the same spot. If so, it is considered "stable" and designated as a "candidate stationary object". The candidate stationary object has the possibility to be finally determined as an abandoned object, a stolen object, or a ghost region when certain conditions are satisfied.

B. ISSUE OF DISCRIMINATION BETWEEN ABANDONED AND STOLEN OBJECT

Although the dual background model makes it easy to extract stationary objects, it is difficult to distinguish between an abandoned object and a stolen object by simply analyzing the foreground. Figure 2 shows the results of the dual background model in the situation after someone stole a suitcase that has existed as a part of the background. While it has already disappeared from the SB, it is still present in the LB because it has not been absorbed in the background yet. Using this feature, the back-tracing verification method presented in this
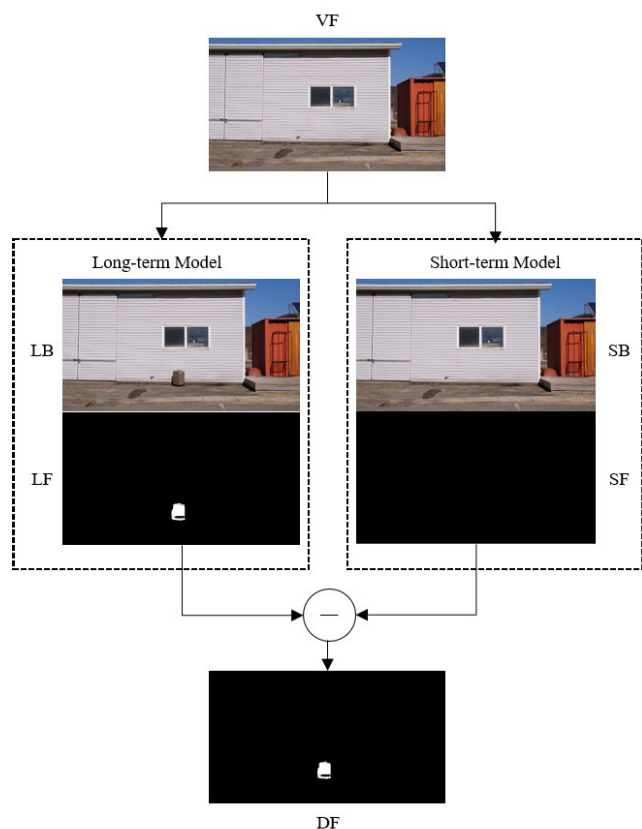
FIGURE 2. When a suitcase that has existed as part of the background is stolen, a stationary object is generated in the DF frame. But it does not exist actually. Therefore, there should be a way to distinguish between abandoned and stolen objects.

FIGURE 3. The edge energy of an object can be higher or lower than the background. Also, the color of the object and background can be similar. (1): There is a suitcase on the road, (2): Suitcase image and its edges, (3): Background image and its edges. (4): There is a suitcase on the parking lot, (5): Suitcase image and its edges, (6): Background image and its edges. We used the Canny edge technique [24] to calculate the edges.

paper searches the previous LB frame to look into the past. Details are described in Section D. Analyzing only the DF, as suggested, can draw a conclusion that a stationary object exists now. But it is wrong. Since this issue increases the false detection rate of the surveillance system, there should be a way to distinguish them.

To solve the problem, Venetianer *et al.* [21] presumed that the edge energy of the current frame is higher for abandoned objects and lower for stolen objects. As shown in Figure 3-(1), (2), and (3), the edge energy of an abandoned object is higher than the background like their expectation. What if the edge energy of the background is higher than an abandoned object as shown in Figure 3-(4), (5), and (6)? It will be detected as a stolen object, which is not. Therefore, this method is error-prone.

Meanwhile, Tian *et al.* [22] noted that there is a difference between the color of an abandoned object and its background. They applied the region growing method from the inside of the foreground area to the outside. If the colors of an edge pixel and its adjacent pixel is comparable to each other, it is included in the same region. If the final region is similar to the stationary object of the DF, they regard that the object is abandoned. Unless it is considered stolen. However, the color of an object and its background can be similar as we can see
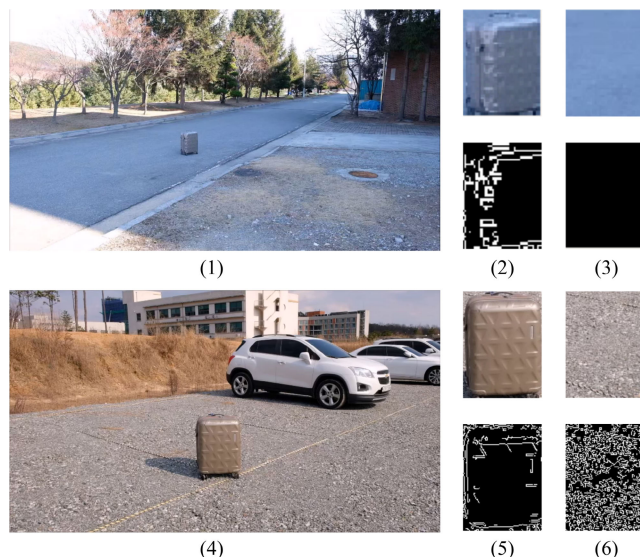
in Figure 3. In this case, the division can be difficult. Even the authors mentioned in their paper that the region growing method failed if the colors of objects and backgrounds were similar.

C. GHOST REGION ISSUE

Figure 4 shows the situation where a ghost region is created. As the shadow of a building moves to the right, a parking line is revealed and falsely determined as a stationary object in the DF frame. Ghost regions can be generated by such as shadow movement and light changes. Since these also increase the false detection rate of the surveillance system, they must be clearly distinguished from both abandoned and stolen objects. As shown in the (4) of Figure 4, Our algorithm identifies the object as a ghost region (the red bounding box). The next section describes the method to solve the discrimination issue including ghost regions.

D. THE PROPOSED ALGORITHM

The purpose of this paper is to accurately detect abandoned and stolen objects. Given that there is a stationary object in the DF, we must be able to distinguish whether it is an abandoned object, a stolen object, or a ghost region. Figure 5 shows the framework of the proposed algorithm in this paper. The dual background model extracts the binary mask of a candidate stationary object (CSO_{DF}) after a specific filtering process. At the same time, object segmentation using Mask R-CNN is performed on the current frame to obtain object masks. The comparison module receives the masks generated by the upper modules. As a result, a segmented object (SO_{VF}) corresponding to the CSO_{DF} is determined. Given no
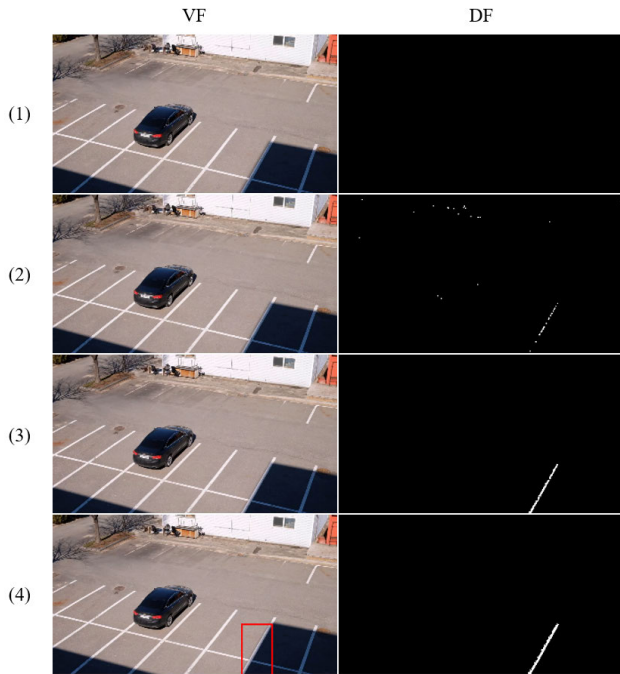
VF   DF

(1)

(2)

(3)

(4)

FIGURE 4. A ghost region is generated. As the shadow of a building moves to the right little by little, a stationary object is created in the DF frame.
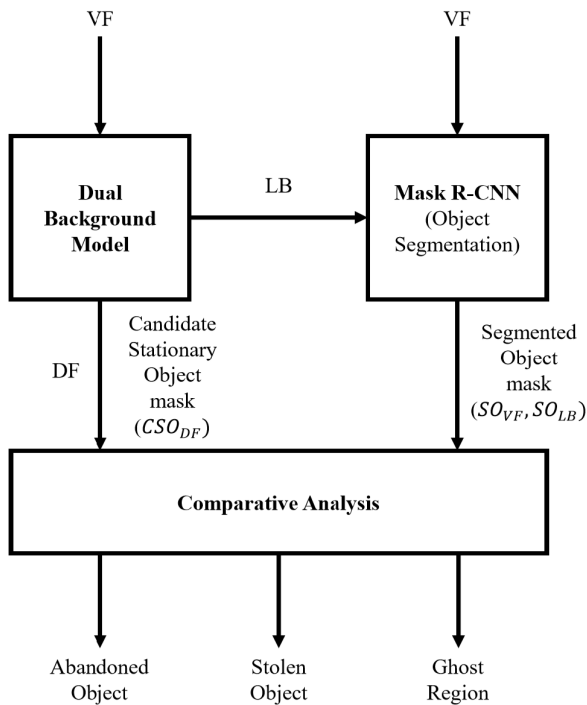

FIGURE 5. Framework of the proposed algorithm.

SO_{VF} matching the CSO_{DF}, another object segmentation process is performed once again on one of the previous LB frames where the CSO_{DF} first appeared in the DF frame. In this paper, this is called back-tracing verification, and the segmented object obtained from the LB frame is

named SO_{LB}. After that, we compare the states of the SO_{VF} and SO_{LB}. Depending on the combination of comparison results, the final state of the CSO_{DF} is determined as one of three categories: abandoned, stolen, or ghost.

To see how equal the two given masks are, we compare the pixels of each region one by one. The method for calculating the match score is as follows. The top, left, width, and height constituting the bounding box of a CSO_{DF} are defined as 't', 'l', 'w', and 'h', respectively. The pixel configuration in the bounding box of the CSO_{DF} is expressed by w × h matrix A:

$$A = [a_{ij}] \quad (l \leq i \leq l + w, t \leq j \leq t + h) \quad (1)$$

$$a_{ij} = \begin{cases} 1, & a_{ij} \in CSO\,mask \\ 0, & otherwise \end{cases} \quad (2)$$

We define the total number of a_{ij} with the value of 1 as 'n', which is used to calculate the final match score. A w × h matrix B represents the segmented object within the same bounding box:

$$B = [b_{ij}] \quad (l \leq i \leq l + w, t \leq j \leq t + h) \quad (3)$$

$$b_{ij} = \begin{cases} 1, & b_{ij} \in Segmented\,Objectmask \\ 0, & otherwise \end{cases} \quad (4)$$

We compare all the corresponding elements of the two matrices to compute matched pixels. To express this result, we define a w × h matrix R and the element r_{ij} has the following values:

$$R = [r_{ij}] \quad (l \leq i \leq l + w, t \leq j \leq t + h) \quad (5)$$

$$r_{ij} = \begin{cases} 1, & a_{ij} = b_{ij} \\ 0, & otherwise \end{cases} \quad (6)$$

If r_{ij} has a value of 1, it represents a matched pixel. The total number of matched pixels is defined as 'm'. Finally, a match score is defined as a ratio of the number of matched pixels (m) to the number of pixels of the CSO_{DF} (n):

$$matchscore(\%) = \frac{m}{n} \times 100 \quad (7)$$

The match result depends on whether the match score is over a certain threshold. This method allows us to compare the CSO_{DF} and SO_{VF}. The comparison of the SO_{VF} and SO_{LB} is performed in the same way.

If two masks match each other, it means that they are identical. Otherwise, there are two cases. First, their sizes are very different. This is particularly relevant to the CSO_{DF}. Because the background subtraction can't represent only the object itself as the foreground. An additional region such as shadows of an object may appear in the foreground, which may create a CSO_{DF} larger than the object. Second, there is no object. Since there can be various states in the past and the present, the match results must be combined to determine the final state of the CSO_{DF}.

Figure 6 is the flowchart of the proposed algorithm. First, we check if there are stationary blobs in the DF frame. Among them, blobs smaller than a certain size are considered
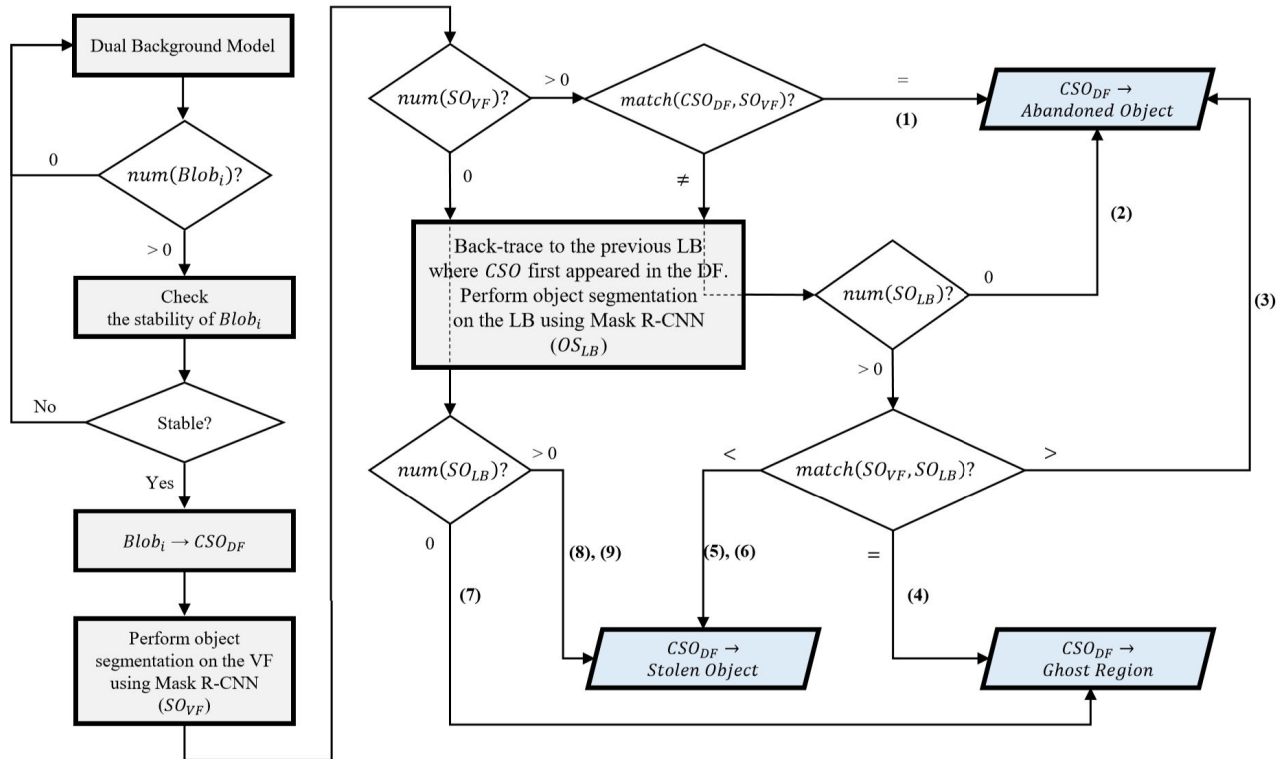
FIGURE 6. Flowchart of the proposed algorithm for detecting abandoned, stolen objects and ghost regions. Each number from (1) to (9) corresponds to those of Figure 7, respectively.

noises and discarded. The remaining ones can be temporarily stationary objects. Therefore, we need to make sure that the blobs appear repeatedly at the same location in consecutive frames. In this paper, this process is called 'checking the stability of a blob'. If the number of appearance of a stationary blob satisfies a condition, it is regarded as stable.

The basic condition to be satisfied for detecting an abandoned object is as follows. First, there is a stationary object. Second, it did not exist before. When a stable blob is selected, it is designated as a CSO_{DF}. To check whether an object exists in the CSO_{DF} area, object segmentation is performed on the current frame using Mask R-CNN. If there is a SO_{VF} in the CSO_{DF} area, a match score is calculated to verify their equality. If they match each other, we consider that it is an abandoned object. Figure 7-(1) corresponds to this case.

Given that a SO_{VF} exists but doesn't match the CSO_{DF}, we need to see the previous situation through the back-tracing verification which has two steps: First, looking for the previous LB frame which corresponds to where the CSO_{DF} is first created. The reason for selecting a specific LB frame is that an object existing as a foreground in the DF frame is not yet absorbed into the background. Second, performing an object segmentation on the LB frame to get the mask of an object. If there is no SO_{LB} in the LB frame, it is determined to be an abandoned object. The reason is that there is now

a new stationary object in the place where there was nothing. Then, why doesn't the SO_{VF} match the CSO_{DF} in the first place? This is because a shadow has been created together as shown in Figure 7-(2), i.e., there is a SO_{VF} smaller than the CSO_{DF} in the absence of the SO_{LB}.

The problem to consider occurs when a SO_{LB} is present. In this case, it should be compared to the SO_{VF}. If the SO_{LB} is equal to the SO_{VF} as shown in Figure 7-(4), it is assumed that it has existed from the past and only a large ghost region, e.g., shadow, covering its area has been created as a CSO_{DF}. Therefore, it is determined as a ghost region. What if the SO_{LB} is smaller than the SO_{VF} like Figure 7-(3)? This indicates that the current object SO_{VF} obscures another small object SO_{LB} and creates an additional shadow equal to the size of the CSO_{DF} at the same time. Therefore, the CSO_{DF} is determined as an abandoned object. If the SO_{LB} is greater than the SO_{VF} as shown in Figure 7-(5) and (6) (regardless of the match score of the CSO_{DF} and SO_{LB}), it means that an uncovered background appears as a foreground after the object has disappeared. And at the same time, another hidden object has revealed in the uncovered background. Therefore, it is determined as a stolen object.

In the absence of both the SO_{VF} and SO_{LB} as shown in Figure 7-(7), the object has never existed from the beginning. Then why is the CSO_{DF} created? The reason is that a ghost region is created. If a SO_{LB} is present while there is no
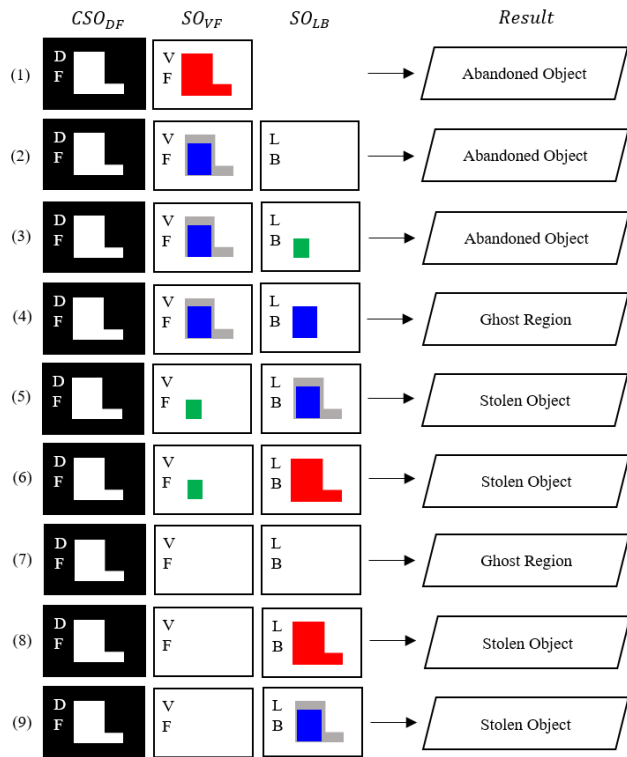
FIGURE 7. The final state of an object according to the states of the CSO_{DF}, SO_{VF}, and SO_{LB}. Objects are colored in red, blue, and green. Shadows are colored in gray.

SO_{VF}, the object is considered stolen, regardless of its size, as shown in Figure 7-(8) and (9).

In summary, in this section, we first covered how to select a candidate stationary object from several foreground blobs generated from a dual background model. To distinguish the status of a given candidate stationary object, we were assisted by Mask R-CNN, one of the famous deep learning-based object segmentation techniques. Object segmentation process is performed on the current frame and the previous background frame to consider not only the present but also the past. The candidate stationary object is compared and analyzed with corresponding segmented objects in each frame. Depending on the result, its final state is determined.

The processes after the object state has been determined are as follows: If it is abandoned, additional processes such as searching for the owner, template registration, etc. can be performed for further tracking. The final state is determined through the presence authentication process at the end of the timer. If the object is stolen, a warning alarm can be triggered immediately. However, if it is a ghost region, nothing happens.

IV. EXPERIMENTS
A. EXPERIMENTAL ENVIRONMENT
We experimented with the PETS2006 [25], ABODA [26], and our own video dataset to demonstrate that the algorithm proposed in this paper can accurately identify and detect abandoned objects, stolen objects, and ghost regions. Since the PETS2006 and ABODA datasets are specialized for detecting abandoned objects, this chapter discusses experiments with our database. The system's core equipment used in the experiments are the Intel® Xeon® W-2123 CPU 3.60GHz, 16.0GB RAM, and NVIDIA TITAN RTX GPU. The ratio of the learning rate of the short-term and long-term model is set to 50:500. For each model, KNN background subtraction technique [3] is used. The threshold values for the match score are set to 90%. For object segmentation, we used the PyTorch based Mask R-CNN [27] distributed by Facebook Research on github.
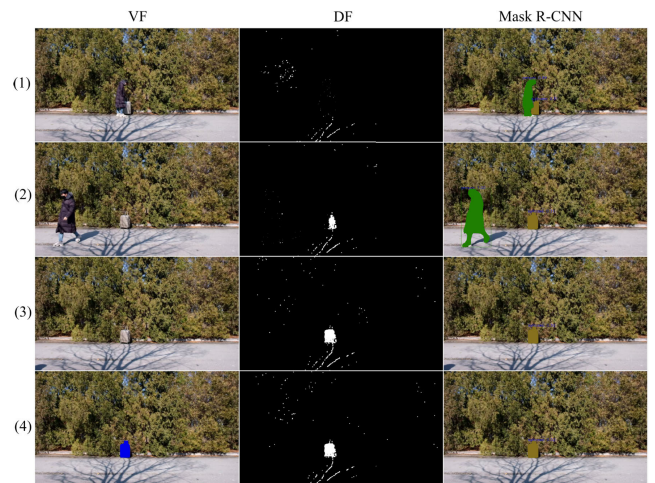

FIGURE 8. Abandoned object detection, which corresponds to the case of Figure 7-(1).

B. DETERMINING THE FINAL STATE OF A STATIONARY OBJECT (ABANDONED, STOLEN OBJECT, AND GHOST REGION) USING THE DUAL BACKGROUND MODEL AND MASK R-CNN
Figure 8 handles a situation that a person abandons a suitcase and disappears, which corresponds to Figure 7-(1). The events that occur over time in the scenario are as follows. Figure 8-(1) illustrates a person putting down a suitcase. The foreground of the stationary object has not yet been created in the DF frame. And then, the person abandoned the suitcase and disappears. As a result, the foreground blob of the object begins to be created in the DF frame. The complete blob is created in the DF frame as shown in Figure 8-(3). Then, the stability of the blob is verified and it is designated as a CSO_{DF}. There is a SO_{VF} matching the the CSO_{DF}. Therefore, the CSO_{DF} is finally considered an abandoned object (colored in blue). This is the basic case for detecting an abandoned object, because the match score of the SO_{VF} and CSO_{DF} exceeds the threshold.

Figure 9 handles a basic scenario for detecting a stolen object, which corresponds to the case of Figure 7-(8). It shows a situation where someone takes an object that has been a part of the background. A suitcase exists as a part of
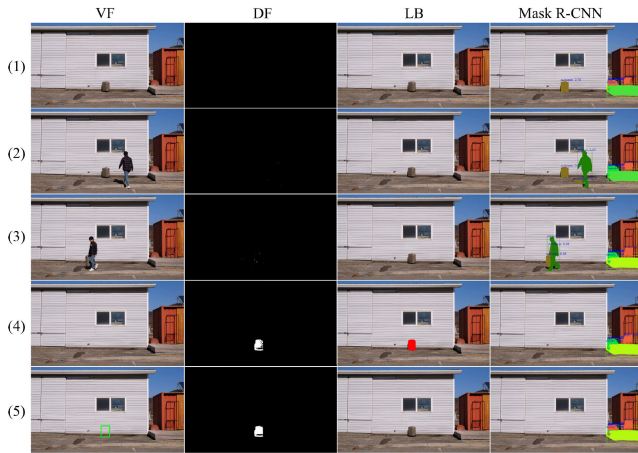
FIGURE 9. Stolen object detection when a CSO_{DF} is identical to a SO_{LB}, which corresponds to the case of Figure 7-(8).

the background in Figure 9-(1). Someone approaches the suitcase. And then, he takes the suitcase and disappears. In Figure 9-(4), the foreground blob of the stationary object is created in the DF frame. Figure 9-(5) is showing that the stability of the blob is verified and it is designated as a CSO_{DF}. The surveillance system back-traces to the LB frame corresponding to where the complete stationary blob is created because there is no SO_{VF}. And a SO_{LB} is detected in the LB of Figure 9-(4) (colored in red). Therefore, the CSO_{DF} is considered as a stolen object and it is marked with a green bounding box.
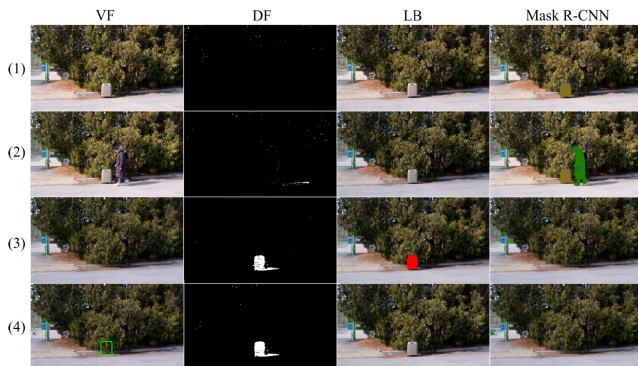

FIGURE 10. Stolen object detection when a CSO_{DF} is larger than a SO_{LB}, which corresponds to the case of Figure 7-(9).

Figure 10 shows a similar situation where a CSO_{DF} is larger than a SO_{LB} because of the shadow of a missing object. A suitcase exists as a part of the background in Figure 10-(1). Figure 10-(2) shows a person taking the suitcase. At the time of Figure 10-(3), the foreground blob of the stationary object is generated in the DF frame. Note that the shadow of the suitcase is created as the foreground. Even if the CSO_{DF} does not match the SO_{LB}, it is eventually determined as a stolen object in Figure 10-(4) because there is no SO_{VF} as shown in Figure 7-(8) and (9).
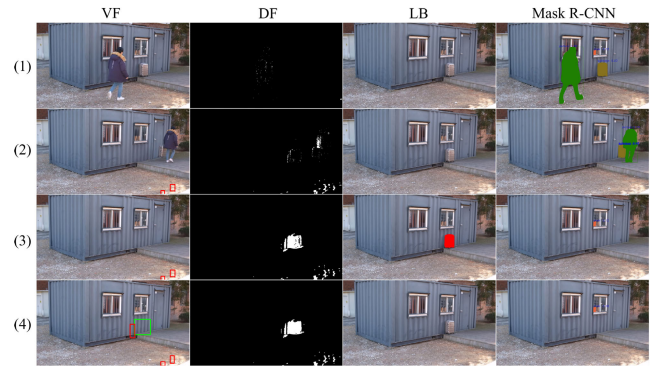

FIGURE 11. Another example of stolen object detection when a CSO_{DF} is larger than a SO_{LB}, which corresponds to the case of Figure 7-(9). And this figure also shows a ghost region detection in the absence of both the SO_{VF} and SO_{LB}, which corresponds to the case of Figure 7-(7).

Figure 11 deals with the almost identical situation to Figure 10, but shows that the surveillance system can detect stolen objects and ghost regions simultaneously using the proposed algorithm. In Figure 11-(1), a suitcase exists as a part of the background and someone is approaching the suitcase. The person takes the suitcase and leaves the area. As we can see, some ghost regions are created because of the light coming through the branches in Figure 11-(2). Figure 11-(3) is showing that the foreground blob of the stationary object is generated in the DF frame. Note that the shadow of the suitcase is also created as the foreground. The object shadow is divided by the exterior pattern of the building so that a part of the shadow is detected as a ghost region. Although the CSO_{DF} does not match the SO_{LB}, it is simply determined to be a stolen object in Figure 11-(4) because there is no SO_{VF}.
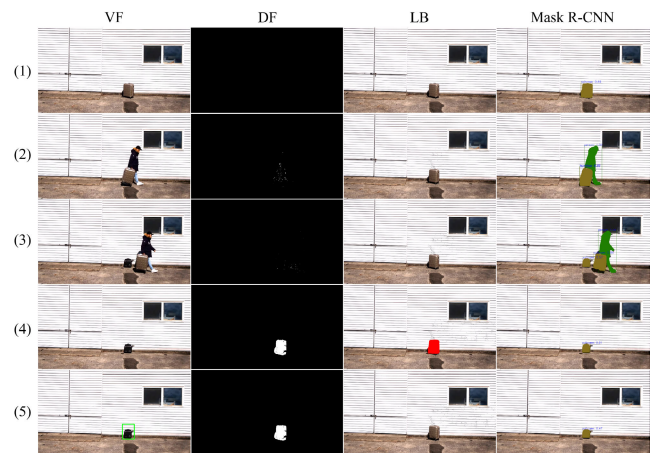

FIGURE 12. Stolen object detection when a CSO_{DF} does not match a SO_{VF} and the SO_{VF} is smaller than a SO_{LB}, which corresponds to the case of Figure 7-(6).

The proposed algorithm determines that a CSO_{DF} is an stolen object even if it does not match the SO_{VF} and also the SO_{VF} is smaller than the SO_{LB}, as shown in Figure 7-(5), (6). We prove our algorithm can handle this situation in Figure 12. Because the CSO_{DF} is larger than the SO_{VF} in Figure 12-(5),
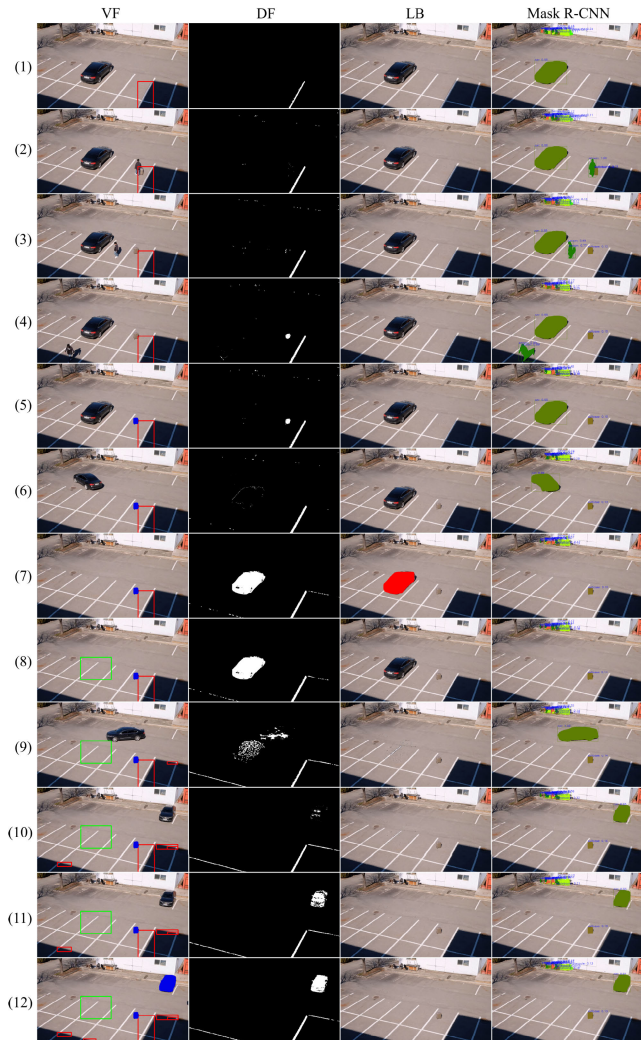
FIGURE 13. Experiment to demonstrate the ability to detect abandoned objects, stolen objects, and ghost regions simultaneously.

they don't match each other. According to the algorithm, the SO_{LB} is detected in Figure 12-(4) through the back-tracing verification, which shows that the SO_{LB} is larger than the SO_{VF}. In other words, the SO_{VF} was placed behind the SO_{LB}. Therefore, the CSO_{DF} is determined to be a stolen object.

Figure 13 shows an experiment to verify that the proposed algorithm distinguishes multiple CSO_{DF}s correctly. As the shadow of a building moves to the right, a CSO_{DF} is generated. It is determined as a ghost region in Figure 13-(1) because there is no SO_{VF} and SO_{LB}. After someone left a suitcase, another CSO_{DF} is created in Figure 13-(4). It is determined to be an abandoned object in Figure 13-(5) because there is a matching SO_{VF}. After the parked car disappears, a CSO_{DF} is created in Figure 13-(7). Since there is no SO_{VF}, the algorithm perform the back-tracing verification and find a SO_{LB} in Figure 13-(7). So the third CSO_{DF} is determined to be a stolen object in Figure 13-(8). The newly parked vehicle are determined as an abandoned object in Figure 13-(12). Of course, parking a car is different from abandoning an

object, but we wanted to show that the proposed algorithm also can be applied to the detection of illegally parked vehicles.

V. CONCLUSION

Researches on existing intelligent video surveillance systems that rely on the foreground analysis generated by the background subtraction have a problem that abandoned objects look like stolen objects and ghost regions. Therefore, it is important to accurately classify these three cases because it can increase the false detection rate of the system. In this paper, we presented a novel algorithm based on traditional image processing techniques and artificial intelligence technology to precisely distinguish abandoned objects, stolen objects, and ghost regions. The proposed algorithm first uses the dual background model to specify the position and area of a candidate stationary object (CSO_{DF}). Then, segmented objects are extracted by performing object segmentation (Mask R-CNN) on the current video frame (SO_{VF}) and the previous LB frame (SO_{LB}) determined via back-tracing verification, which returns the pixel area constituting several types of objects. Depending on the result of comparative analysis on CSO_{DF}, SO_{VF}, and SO_{LB}, the candidate stationary object is accurately classified into an abandoned object, a stolen object, and a ghost region. We showed that the proposed algorithm accurately detects the state of an object in various situations through experiments on our own dataset. Therefore, it is expected to be widely used for real-time detection of not only abandoned, stolen objects in open environments such as exhibition halls and parks but also for illegally parked vehicles, etc. Due to native characteristics of the deep learning model, there is a limitation that only trained objects can be detected. Also, if the video has a low resolution, it can be difficult to identify objects. However, since the release of Mask R-CNN in 2017, more training datasets have been released and also new technologies such as detectron2 [28] and YOLACT [29] have been continuously developed. Therefore, we expect that the methodology presented in this paper will become more robust over time.

REFERENCES

[1] E. Luna, J. San Miguel, D. Ortego, and J. Martínez, "Abandoned object detection in video-surveillance: Survey and comparison," *Sensors*, vol. 18, no. 12, p. 4290, Dec. 2018, doi: 10.3390/s18124290.

[2] N. Bird, S. Atev, N. Caramelli, R. Martin, O. Masoud, and N. Papanikolopoulos, "Real time, online detection of abandoned objects in public areas," in *Proc. IEEE Int. Conf. Robot. Autom. ICRA.*, Orlando, FL, USA, May 2006, pp. 3775–3780, doi: 10.1109/ROBOT.2006.1642279.

[3] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 773–780, May 2006, doi: 10.1016/j.patrec.2005.11.005.

[4] S. Lu, J. Zhang, and D. Feng, "An efficient method for detecting ghost and left objects in surveillance video," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Sep. 2007, pp. 540–545, doi: 10.1109/AVSS.2007.4425368.

[5] J. Martínez-del-Rincon, J. E. Herrero-Jaraba, J. R. Gómez, and C. Orrite-Urunuela, "Automatic left luggage detection and tracking using multi-camera UKF," in *Proc. IEEE 9th IEEE Int. Workshop PETS*, New York, NY, USA, vol. 18, Jun. 2006, pp. 59–66.

[6] F. Porikli, Y. Ivanov, and T. Haga, "Robust abandoned object detection using dual foregrounds," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 1, Dec. 2007, Art. no. 197875, doi: 10.1155/2008/197875.

[7] R. H. Evangelio, T. Senst, and T. Sikora, "Detection of static objects for the task of video surveillance," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2011, pp. 534–540, doi: 10.1109/WACV.2011.5711550.

[8] K. Lin, S.-C. Chen, C.-S. Chen, D.-T. Lin, and Y.-P. Hung, "Abandoned object detection via temporal consistency modeling and backtracing verification for visual surveillance," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 7, pp. 1359–1370, Jul. 2015, doi: 10.1109/TIFS.2015.2408263.

[9] K. Lin, S.-C. Chen, C.-S. Chen, D.-T. Lin, and Y.-P. Hung, "Left-luggage detection from Finite-State-Machine analysis in static-camera videos," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Stockholm, Sweden, Aug. 2014, pp. 4600–4605, doi: 10.1109/ICPR.2014.787.

[10] H. Park, S. Park, and Y. Joo, "Robust detection of abandoned object for smart video surveillance in illumination changes," *Sensors*, vol. 19, no. 23, p. 5114, Nov. 2019, doi: 10.3390/s19235114.

[11] H. Park, S. Park, and Y. Joo, "Robust real-time detection of abandoned objects using a dual background model," *KSII Trans. Internet Inf. Syst.*, vol. 14, no. 2, Feb. 2020, doi: 10.3837/tiis.2020.02.017.

[12] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Left-luggage detection from finite-state-machine analysis in static-camera videos," in *Proc. ICCV*, Venice, Italy, Oct. 2017, pp. 2961-2969, doi: 10.1109/ICPR.2014.787.

[13] Wahyono, A. Filonenko, and K.-H. Jo, "Unattended object identification for intelligent surveillance systems using sequence of dual background difference," *IEEE Trans. Ind. Informat.*, vol. 12, no. 6, pp. 2247–2255, Dec. 2016, doi: 10.1109/TII.2016.2605582.

[14] Wahyono, A. Filonenko, and K.-H. Jo, "Detecting abandoned objects in crowded scenes of surveillance videos using adaptive dual background model," in *Proc. 8th Int. Conf. Hum. Syst. Interact. (HSI)*, Jun. 2015, pp. 25–27, doi: 10.1109/HSI.2015.7170670.

[15] C. Cuevas, R. Martinez, D. Berjon, and N. Garcia, "Detection of stationary foreground objects using multiple nonparametric background-foreground models on a finite state machine," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1127–1142, Mar. 2017, doi: 10.1109/TIP.2016.2642779.

[16] S. Smeureanu and R. T. Ionescu, "Real-time deep learning method for abandoned luggage detection in video," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 1775–1779, doi: 10.23919/EUSIPCO.2018.8553156.

[17] D. Shyam, A. Kot, and C. Athalye, "Abandoned object detection using pixel-based finite state machine and single shot multibox detector," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, San Diego, CA, USA, Jul. 2018, pp. 23–27, doi: 10.1109/ICME.2018.8486464.

[18] H. Xu and F. Yu, "Improved compressive tracking in surveillance scenes," in *Proc. 7th Int. Conf. Image Graph.*, Qingdao, China, Jul. 2013, pp. 26–28, doi: 10.1109/ICIG.2013.176.

[19] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multiBox detector," in *Proc. ECCV*, Amsterdam, The Netherlands, Oct. 2016, pp. 21–37.

[20] J. Connell, A. W. Senior, A. Hampapur, Y.-L. Tian, L. Brown, and S. Pankanti, "Detection and tracking in the IBM PeopleVision system," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Taipei, Taiwan, vol. 2, Jun. 2004, pp. 1403–1406, doi: 10.1109/ICME.2004.1394495.

[21] P. L. Venetianer, Z. Zhang, W. Yin, and A. J. Lipton, "Stationary target detection using the objectvideo surveillance system," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Sep. 2007, pp. 242–247, doi: 10.1109/AVSS.2007.4425317.

[22] Y. Tian, R. S. Feris, H. Liu, A. Hampapur, and M.-T. Sun, "Robust detection of abandoned and removed objects in complex surveillance videos," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 41, no. 5, pp. 565–576, Sep. 2011, doi: 10.1109/TSMCC.2010.2065803.

[23] Y. Tian, R. Ferris, and A. Hampapur, "Real-time detection of abandoned and removed objects in complex environments," in *Proc. 8th Int. Workshop Visual Surveill. (VS)*, Marseille, France, Oct. 2008, pp. 1–9.

[24] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vols. PAMI–8, no. 6, pp. 679–698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.

[25] *PETS 2006 Benchmark Data*. Accessed: Dec. 1, 2018. [Online]. Available: http://www.cvg.reading.ac.uk/PETS2006/data.html

[26] *ABODA Dataset*. Accessed: Mar. 17, 2020. [Online]. Available: https://github.com/kevinlin311tw/ABODA

[27] F. Massa and R. Girshick. (2019). *Maskrcnn-Benchmark: Fast, Modular Reference Implementation of Instance Segmentation and Object Detection Algorithms in PyTorch*. Accessed: Mar. 19, 2020. [Online]. Available: https://github.com/facebookresearch/maskrcnn-benchmark

[28] Y. Wu, A. Kirillov, F. Massa, W, -Y. Lo, and R. Girshick. (2019). *Detectron2*. Accessed: Mar. 19, 2020. [Online]. Available: https://github.com/facebookresearch/detectron2

[29] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: Real-time instance segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9157–9166.

HYESEUNG PARK received the B.S. degree in computer engineering and science and the M.S. degree in electronics and computer engineering and science from Hanyang University, South Korea, in 2012 and 2014, respectively. He is currently pursuing the Ph.D. degree in computer science and engineering with the Korea University of Technology and Education. His research interests include image processing, intelligent video surveillance systems, object detection, depth prediction, and artificial intelligence.

SEUNGCHUL PARK received the B.S. degree in computer science and statistics from Seoul National University, South Korea, in 1985, the M.S. degree in computer science from the Korea Advanced Institute of Science and Technology, and the Ph.D. degree in computer engineering from Seoul National University, in 1996. Before joining the Korea University of Technology and Education as a Professor, in 2004, he worked for the Electronics and Telecommunications Research Institute, IBM Korea, Hyundai Electronics (current SK Hynix), and Hyundai Networks, as a Software and Network Engineer. His current research interests include blockchain, network security, and multimedia communications.

YOUNGBOK JOO received the B.S. and M.S. degrees in computer science from Yonsei University, Seoul, South Korea, in 1991 and 1993, respectively, and the M.S. and Ph.D. degrees in computer science and engineering from the University of New South Wales, Sydney, Australia, in 1997 and 2000, respectively. From February 1991 to January 1993, he was an Assistant Researcher with Samsung Electronics, Ki-Heung, South Korea. From March 1997 to June 2000, he was a Researcher with Platypus Technology, Sydney. From July 2000 to October 2001, he was a Senior Researcher with Q-vis Ltd., Perth, Australia. From December 2001 to June 2002, he was a Visiting Researcher with The University of Western Australia, Perth. From July 2002 to May 2006, he was a Senior Researcher with Lynx Engineering Consultant, Perth. From June 2006 to 2008, he was a Research Professor with Kyungbook National University, Daegu, South Korea and Yonsei University. Since 2009, he has been an Associate Professor with the Korea University of Technology and Education. His current research interests are pattern recognition, face detection and recognition, and automatic vision inspection.

• • •