IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 8 March 2024, accepted 24 April 2024, date of publication 29 April 2024, date of current version 6 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3395172

APPLIED RESEARCH

Practical Abandoned Object Detection in Real-World Scenarios: Enhancements Using Background Matting With Dense ASPP

MINGU JEONG[1,2], DOHUN KIM[3], AND JOONKI PAIK[1,4], (Senior Member, IEEE)
[1]Department of Image, Chung-Ang University, Seoul 06974, South Korea
[2]Korea Electronics Technology Institute (KETI), Seongnam 13509, South Korea
[3]Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, South Korea
[4]Department of Artificial Intelligence, Chung-Ang University, Seoul 06974, South Korea

Corresponding author: Joonki Paik (paikj@cau.ac.kr)

This work was supported in part by the Korea Government Ministry of Sicence and Information Technology (MSIT) through the Artificial Intelligence Graduate School Program at Chung-Ang University under Grant 2021-0-01341, and in part by the Institute of Civil-Military Technology Cooperation (ICMTC) project 'Research on bio-object recognition and matching algorithm in 3D space' (19CM0035, 20-CM-DD-03).

ABSTRACT The widespread deployment of Closed-Circuit Television (CCTV) systems in public and private spaces has significantly enhanced security measures but also posed unique challenges in accurately interpreting the voluminous data captured, especially in the context of abandoned object detection. This area is critical for identifying potential security threats, including illegal waste disposal, explosives, or lost items, which necessitate sophisticated detection techniques. Traditional methods often struggle with limitations such as false positives/negatives due to dynamic environmental conditions like lighting changes or complex backgrounds. Addressing these challenges, our study proposes a novel abandoned object detection system that integrates background matting and advanced learning algorithms to refine detection accuracy. The system architecture is divided into three key stages: i) preprocessing, to reduce noise and adjust for lighting variations; ii) abandoned object recognition (AOR), employing background matting to distinguish between static and dynamic entities, further enhanced by pedestrian detection to exclude moving objects; and iii) abandoned object decision feature correction (AODFC), which employs feature correlation analysis for precise identification of abandoned objects. The experimental evaluation, conducted across varied real-world settings, demonstrates the method's superior performance over conventional approaches, significantly reducing false identifications while maintaining high detection accuracy. This paper not only presents a comprehensive solution to the challenges of abandoned object detection but also paves the way for future research in enhancing the robustness and applicability of surveillance systems.

INDEX TERMS Object detection, image matting, abandoned object detection, dense ASPP.

I. INTRODUCTION

With the ubiquitous deployment of Closed-Circuit Television (CCTV) in both public and private sectors, there has been a marked increase in research efforts focused on leveraging Computer Vision techniques to analyze and interpret critical scenarios captured in CCTV footage. Abandoned object detection (AOD) has emerged as a focal point of interest within the research community, attributed to its significant implications for public safety and security. This field aims to identify objects that have been intentionally left unattended, posing potential threats such as illegal dumping, explosives, and lost items, thereby necessitating sophisticated detection and analysis techniques [1], [2]. The concept of

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

an 'abandoned object' is defined as any item that has been deliberately discarded or left behind by an individual. The literature on AOD explores a variety of abandonment criteria, which include, but are not limited to, the spatial separation between an object and its owner, and the temporal aspect of how long an object has been left unattended [3]. Furthermore, the research delves into the challenges of accurately identifying and categorizing such objects within the often complex and dynamic environments captured by CCTV systems. Addressing these challenges requires a multifaceted approach, combining advancements in image processing, machine learning, and pattern recognition to develop reliable detection methodologies. In enriching the introduction with additional references, our aim is to provide a comprehensive overview of the AOD landscape, highlighting the critical role this technology plays in enhancing surveillance capabilities and public safety measures. By situating our work within the broader context of existing research, we underscore the contributions our study makes towards advancing the field of AOD, offering novel insights and solutions to the challenges inherent in detecting abandoned objects in diverse and unpredictable environments. This revision aims to more effectively set the stage for your study by detailing the background of AOD with a rich array of references, thereby providing a solid foundation for understanding the importance and complexity of the research area. In revising the categorization of previous studies on abandoned object detection (AOD), we acknowledge the reviewer's insight and have thus reframed the discussion to emphasize the analysis from a background and foreground perspective.Abandoned object detection methodologies can be broadly classified into two intertwined categories based on their primary focus: background analysis and foreground identification. Background analysis methods primarily focus on modeling the environment to detect changes that signify the presence of a newly abandoned object. This category benefits from a profound ability to distinguish between dynamic and static elements within a scene, facilitating both immediate and prolonged detection scenarios. Among the techniques employed, dual background modeling emerges as a particularly prominent method, praised for its effectiveness in long-term abandoned object identification [4], [5]. Foreground identification strategies, on the other hand, leverage advancements in deep learning to directly recognize and classify objects as abandoned based on their features and temporal persistence in the scene. While deep learning approaches offer robustness against common detection challenges such as variable lighting and occlusions, they are complemented by the foundational insights provided by background modeling techniques. Acknowledging the limitations and strengths of both approaches, recent research has increasingly focused on hybrid models that integrate background modeling for dynamic/static distinction with deep learning for sophisticated object recognition. This combined methodology aims to mitigate the drawbacks of each approach such as the sensitivity of background models to environmental noise factors like illumination changes, reflections, and camera movements, and the resource intensive nature of deep learning models thereby offering a more balanced and effective solution for AOD.

Recent advancements in object detection technologies have shown promising results in controlled or indoor settings. However, their applicability and efficiency in outdoor environments remain a significant challenge [6]. The dynamic nature of outdoor settings, characterized by frequent background changes, presents additional hurdles not commonly encountered indoors. This is particularly evident in the domain of abandoned object detection, where the ability to accurately identify objects left unattended in outdoor spaces is critical. Existing methodologies often struggle in such environments, leading to a higher incidence of false positives and false negatives. These systems are typically optimized for static backgrounds and can falter under the influence of variable environmental conditions such as changes in lighting, weather patterns, and ambient noise levels. Given the diversity of potential backgrounds and the complexity inherent in outdoor environments, improving the accuracy and reliability of abandoned object detection systems is paramount. Our motivation stems from this significant gap in the current research landscape, prompting us to explore innovative approaches that enhance detection capabilities in these challenging conditions.

This paper presents a novel Abandoned object detection method that leverages the integration of background matting and learning-based detector techniques to address the critical limitations found in previous research. The proposed approach detects changes in the background, distinguishing between suspected static and dynamic objects, by employing a background matting method. Subsequently, a pedestrian detector is applied to eliminate dynamic objects from the analysis. The proposed method can be described in three main steps as follows: i) Pre-Processing, ii) Abandoned Object Recognition (AOR), and iii) Abandoned Object Decision by Feature Correlation (AODFC).

The proposed abandoned object detection method comprises three key steps, aiming to overcome the limitations of conventional techniques and effectively identify abandoned objects.

In the pre-processing step, two methods are employed to enhance accuracy. Firstly, pixel averaging is applied between sequential frames to minimize pixel variations caused by communication noise in CCTV with RTSP-based communication protocols [7]. Secondly, background information is initialized to handle abrupt changes in illumination, which could affect the background data. By monitoring pixel changes between frames, our system identifies the moment of illumination change, allowing us to reset the background information to that specific point. This process effectively reduces false positives in the background matting method, especially in indoor and outdoor environments where lighting changes may occur.

In the Abandoned Object Recognition (AOR) step, we introduce a system that combines background matting and pedestrian detection techniques to address the drawbacks of traditional background modeling and learning-based methods. Background matting helps detect areas of background variation, revealing regions containing both static and dynamic object noise. Subsequently, a pedestrian detection method is employed to remove dynamic object areas, focusing on analyzing the relationship with the owner to determine abandoned objects. The system stores information from pedestrian detection to make accurate determinations.

Finally, in the Abandoned Object Decision by Feature Correlation (AODFC) step, abandoned objects are identified within the candidate areas detected in previous steps. By performing feature correlation analysis between candidate regions and their counterparts in previous time periods, this step effectively distinguishes abandoned objects from regions generated by noise or dynamic objects. Regions with correlated features against previous frames are not deemed abandoned objects.

The prevalent issue of background subtraction in the context of abandoned object detection presents significant challenges, including sensitivity to lighting changes, the complexity of modeling dynamic backgrounds, and the difficulty in distinguishing objects against complex backgrounds. In addressing these challenges, this paper introduces an innovative method that effectively mitigates the drawbacks associated with traditional background subtraction techniques. By employing a combination of image denoising and light change detection, the proposed approach significantly reduces the impact of lighting variations and improves object identification accuracy. Furthermore, we advance the detection process by utilizing a background matting method to remove complex backgrounds and dynamic object removal to minimize false positives. A crucial step in the proposed method involves the application of a feature correlation technique, which further refines the accuracy of identifying abandoned objects by reducing false positives. the proposed research distinguishes itself through the integration of comprehensive experimental analyses and empirical evidence derived from real-world scenarios. These efforts collectively demonstrate the practical viability and enhanced effectiveness of the proposed method, underscoring its potential to overcome longstanding challenges in abandoned object detection.

The paper is organized as follows: In section II, we delve into the research on existing abandoned object detection systems. The proposed abandoned object detection system is presented in section III. To assess its performance, we evaluate the method using ABODA [8], KISA [9], in-house and VideoMatte [10] datasets in section IV. Lastly, we provide concluding remarks in section V.

II. RELATED WORK

Abandoned object detection (AOD) is a specialized field within object detection, emphasizing the necessity to discern the relationship between an object and its potential owner as well as to identify objects that remain static over time. Research in AOD can be broadly categorized into two methodological approaches: background-aware and foreground-aware methods, each incorporating elements of traditional techniques and deep learning advancements for improved detection accuracy.

Background-Aware Methods: These techniques primarily focus on modeling the environment's background. A seminal approach by Porikli et al. introduced the concept of dual background modeling, utilizing two distinct models operating over different temporal scales to differentiate between dynamic and static objects, thus facilitating effective background subtraction [1]. To address the limitations inherent in traditional background modeling methods, subsequent research has integrated deep learning models to enhance the detection capabilities. For instance, Smeureanu et al. proposed a method that combines background subtraction with a convolutional neural network (CNN) to accurately identify abandoned objects through the analysis of static and dynamic object candidates over temporally adjacent frames [11]. **Foreground-Aware Methods:** In contrast, foreground-aware methods concentrate on the detection and analysis of objects in motion within the scene. Park et al. developed a dual background subtraction technique that leverages long-term foreground information to discern changes in lighting conditions, aiding in the identification of abandoned objects [12]. To further mitigate false positives, integration of object detection networks like the Single Shot Multibox Detector (SSD) has been explored, refining the process of distinguishing abandoned objects from those temporarily stationary [13]. Recent advancements have also explored direct detection strategies that address the challenges posed by complex backgrounds and dynamic environmental conditions. Kim et al. introduced an innovative method that utilizes key-point and object detection networks to identify abandoned luggage by examining its spatial relationship with potential owners, showcasing the application of deep learning to overcome traditional detection challenges [3]. Despite the progress, both traditional and deep learning-based methods encounter challenges, including sensitivity to environmental changes, complex scene dynamics, and the inherent limitations of machine learning models such as occlusion and high rates of false positives. To surmount these obstacles, this paper introduces a novel background matting technique, surpassing traditional background modeling in effectiveness. Furthermore, we propose a comprehensive system for AOD that leverages deep learning for efficient and precise identification of abandoned objects, addressing key shortcomings of previous methodologies. Traditional background and foreground subtraction methods, while foundational to the field of abandoned object detection, encounter significant challenges that undermine their effectiveness and applicability. These methods often fail to adapt to environmental changes, notably lighting variations, making them unsuitable for dynamic outdoor settings. The complexity of accurately modeling backgrounds, coupled with the hurdles in initializing these models in diverse conditions, further diminishes their utility. Such limitations not only restrict their performance but also highlight the fundamental flaws in relying solely on these traditional techniques for detecting abandoned objects. Furthermore, while deep learning models have emerged as powerful tools for enhancing detection capabilities, they introduce a distinct set of challenges that have not been adequately addressed by existing research. Specifically, these models suffer from 'black box' problems, where the lack of transparency and interpretability in their decision-making processes poses significant concerns. The inability of deep learning models to provide insights into their reasoning complicates the task of diagnosing and rectifying errors, such as those caused by occlusions and the misidentification of objects, leading to high rates of false positives. This opacity limits the potential for researchers and practitioners to fully trust and understand the outcomes of these advanced models. To overcome the aforementioned shortcomings, this paper introduces a novel approach that incorporates an innovative background matting method, demonstrating superior efficacy over traditional background modeling techniques. This method significantly enhances the detection of abandoned objects by addressing key challenges such as lighting variations and complex background scenarios. Additionally, we propose a comprehensive system for abandoned object detection that not only leverages the strengths of deep learning for accurate object identification but also addresses its inherent limitations. This system aims to provide a more transparent, interpretable, and reliable solution for detecting abandoned objects, marking a significant advancement in the field. Deep learning has revolutionized the way we approach object detection, extending its utility far beyond traditional applications. In medical imaging, deep learning models have shown exceptional prowess in detecting and diagnosing diseases from complex image data, offering insights that significantly aid in patient care [14]. Similarly, in the agricultural sector, these models have been applied to enhance crop monitoring and yield prediction. A notable example is the GrainSpace dataset, a large-scale resource for fine-grained and domain-adaptive recognition of cereal grains, which exemplifies the application of deep learning in addressing specific challenges within agricultural imagery [15]. Furthermore, deep learning's impact extends to industrial applications, where it is used for anomaly detection and quality control. The MVTec Anomaly Detection (AD) dataset represents a comprehensive real-world resource for unsupervised anomaly detection in industrial images, showcasing the effectiveness of deep learning models in identifying defects and irregularities in manufacturing processes [16].

Building on these diverse applications, our research explores the integration of deep learning techniques within abandoned object detection systems. By drawing upon the principles and methodologies proven effective in domains such as medical imaging, agriculture, and industrial inspection, we aim to enhance the accuracy and reduce false positives in detecting abandoned objects. The adaptability of deep learning models to various detection tasks suggests their potential to significantly improve abandoned object detection when tailored to the unique challenges of this field.

This expanded discussion underscores the wide-reaching implications of deep learning advancements across different sectors and illustrates the potential for these technologies to inform and enhance abandoned object detection systems. By integrating deep learning models that have demonstrated success in fields ranging from medical imaging to agriculture and industrial inspection, we propose a comprehensive approach to abandoned object detection that leverages the full spectrum of deep learning capabilities.

III. PROPOSED METHOD

In this study, we introduce a novel approach to detecting abandoned objects by leveraging image matting techniques, addressing the limitations inherent in traditional background subtraction methods. We call this system Abandoned Object Detection Using Background Matting(AODBM). The proposed method include three distinct stages: i) Pre-Processing, ii) Abandoned Object Recognition (AOR), and iii) Abandoned Object Decision via Feature Correlation (AODFC), as depicted in Fig. 1.

The Pre-Processing stage consists of two subsystems: noise reduction and illumination change detection. The noise reduction subsystem aims to mitigate the pixel variations in videos arising from factors like communication noise, light fluctuations, and other disturbances. Concurrently, the illumination change detection subsystem pinpoints video frames experiencing significant pixel alterations due to lighting changes, recording the respective frame numbers. Such frame number data proves invaluable for initializing the background during background matting.

The AOR subsystem employs the background matting technique alongside the YOLO V7 algorithm to eliminate both backgrounds and dynamic objects from videos. By discarding the background and transient object details, the proposed system is designed to precisely detect only abandoned objects, particularly in CCTV environments.

Meanwhile, the AODFC subsystem leverages feature correlation to minimize potential inaccuracies arising during background matting. We introduce a refined method for abandoned object detection that mitigates false detections. This is achieved by computing the Intersection Over Union (IOU) between the background frame and the subsequent frames derived from the abandoned object detection results secured by the AODFC system. Comprehensive explanations are provided in the subsequent sections.

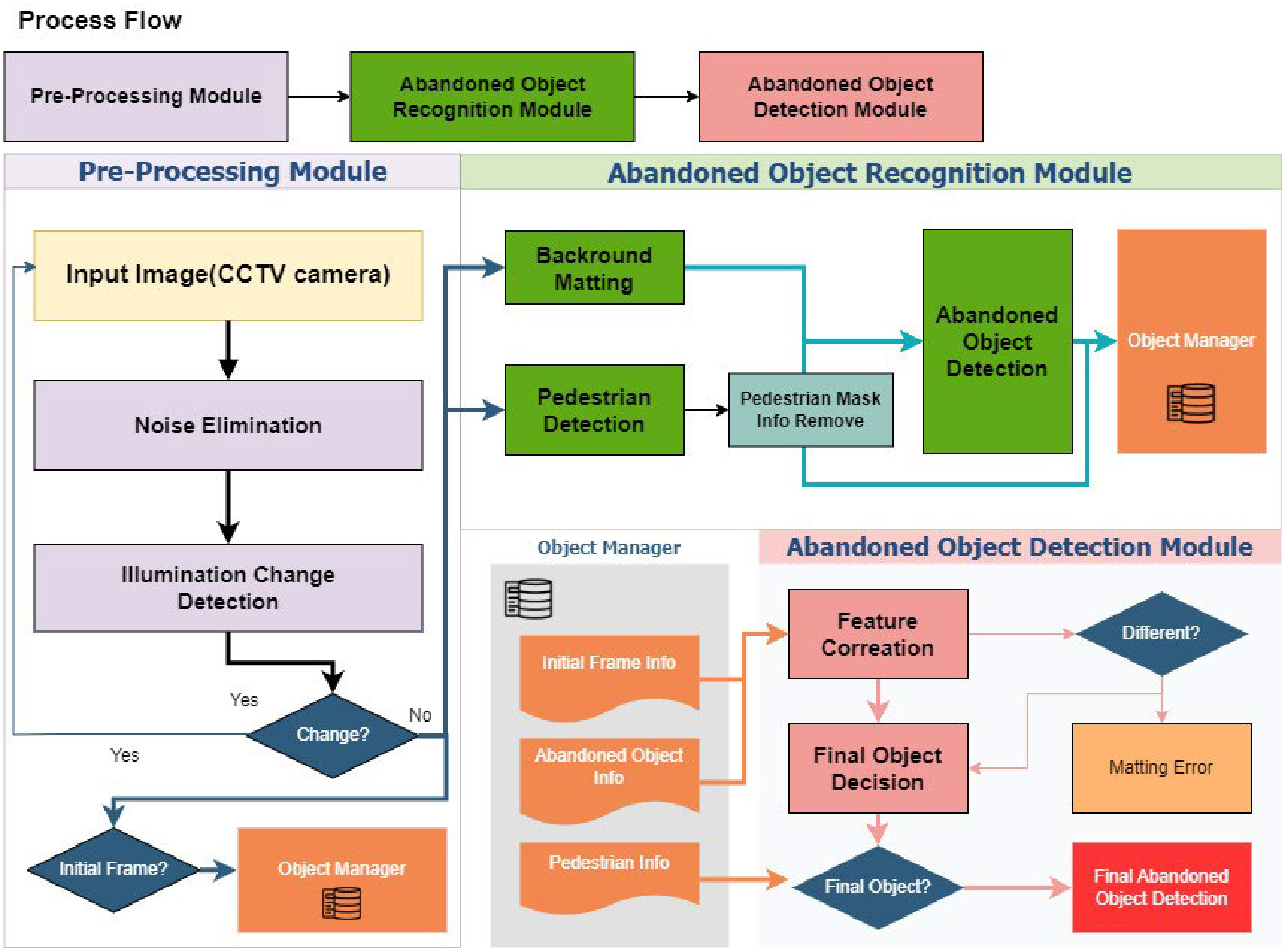


FIGURE 1. Block diagram of the proposed abandoned object detection system.

A. PRE-PROCESSING

1) NOISE REDUCTION

CCTV camera imagery frequently contains noise, which alters image pixels due to several factors, including communication interference, illumination shifts, and thermal noise. Conventional background subtraction or background matting algorithms detect objects based on these pixel fluctuations. In this context, it's evident that the accuracy of background subtraction or matting techniques decreases when applied to images contaminated by noise. Regarding CCTV footage, it's transmitted in real-time using the RTSP communication protocol. While this system offers the benefit of rapid communication speeds due to its real-time image reception, it's concurrently susceptible to communication noise.

Figure 2 demonstrates the process and results of applying the image enhancement and background matting techniques to images degraded by communication noise. Figure 2 (a) displays an image compromised by communication interference, highlighting the challenges such noise poses to image clarity and object detection. Figure 2 (b) shows the initial attempt at applying the background matting algorithm to the noise-affected image, illustrating the limitations of conventional matting methods in the presence of significant noise.

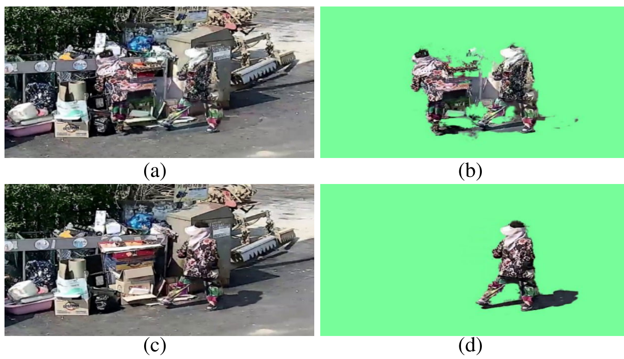


FIGURE 2. (a) Image affected by communication noise, (b) result of applying the background matting algorithm to the noise-affected image, (c) image after noise reduction, and (d) result of background matting post-noise reduction.

Recognizing the need for a robust preprocessing step to address this issue, we introduce a video normalization technique aimed at mitigating the effects of communication noise, as depicted in Panel (c). This technique not only reduces noise but also preserves the integrity of the image, making it a crucial preliminary step before further processing. Panel (d) presents the result of applying the background matting algorithm post-noise reduction, showcasing a marked improvement in matting quality and object detection accuracy. Furthermore, this technique preserves the pixels from each video frame, presenting an advantage in discerning illumination variations by monitoring these pixel alterations. In addressing the challenges posed by communication interference in video frames, our study introduces an advanced noise reduction technique through video normalization, a critical component of the proposed abandoned object detection system. This technique is meticulously designed to mitigate the detrimental effects of noise on image quality, thereby significantly enhancing the detection accuracy of abandoned objects. The primary contribution of the video normalization process lies in its ability to preserve essential pixel information across video frames. This preservation is crucial for accurately discerning illumination variations and subtle movements, which are pivotal in abandoned object detection scenarios. By closely monitoring and analyzing these pixel alterations, the proposed method adeptly identifies and compensates for noise-induced distortions without compromising the integrity of the original data. However, a potential challenge with normalization is the risk of distorting or omitting critical patterns present in the video frames. Such alterations could inadvertently normalize characteristics vital for accurate object detection. To counteract this, we have implemented strategic adjustments within the normalization algorithm. These adjustments are meticulously calibrated to ensure that crucial information, particularly related to object characteristics and background details, is retained and accurately represented during the normalization process. This nuanced approach sets the proposed method apart from conventional video normalization techniques, which may not adequately address the preservation of essential data characteristics. Our modifications to the standard video normalization process include adaptive thresholding and selective normalization strategies. These enhancements are specifically tailored to maintain the fidelity of critical visual features, such as object contours and texture details, which are essential for the subsequent steps of the abandoned object detection methodology. By integrating these tailored adjustments, the video normalization technique not only reduces noise more effectively but also enhances the overall performance of the background matting and object detection stages. This innovative approach represents a significant advancement over previous studies, which may not fully account for the intricate balance required between noise reduction and the preservation of vital image features in the context of abandoned object detection.

(a)

(b)

FIGURE 3. Examples of illumination variations: (a) indoor shift and (b) outdoor shift.

The mechanisms for pixel storage and illumination change detection within the image are elaborated upon in the subsequent section.

2) INITIALIZATION ALGORITHM BASED ON ILLUMINATION CHANGE DETECTION

Camera environments frequently display regions within images that experience significant pixel variations as a result of illumination changes. Outdoors, these illumination shifts might arise from events such as sunset, while indoors, changes in artificial lighting can cause similar effects. Fig. 3 illustrates instances of illumination variation within images. Fig. 3(a) captures the illumination shift in an indoor setting, whereas Fig. 3(b) depicts it in an outdoor context. Object detection systems that rely on pixel analysis often encounter challenges with increased false positives triggered by such illumination alterations.

The background subtraction technique, also known as the background matting algorithm, can extract dynamic objects based on pixel alterations. However, when a pixel shift is caused by an illumination change, the system misinterprets it as a dynamic object, leading to inaccurate outcomes. In this section, we propose an efficient illumination change detection algorithm designed to promptly detect lighting shifts and set the initial frame for the background matting algorithm. The proposed approach hinges on analyzing the pixel data from the frame stored during the noise reduction phase. For both outdoor and indoor scenarios, widespread pixel shifts in the image signify illumination changes. Through this paper, we demonstrate that pixel variations are minimal in images that have undergone denoising via pixel normalization. Nonetheless, it remains feasible to identify substantial pixel shifts across large portions of the image due to abrupt illumination changes.

To detect changes in illumination, the proposed method starts by converting the continuous stream of RGB frames into the YCbCr color space, which typically comprises the Y channel containing brightness information, and Cb and Cr channels that hold color details. In this approach, we introduce an algorithm that detects illumination shifts by analyzing the Y-values within the transformed YCbCr space, focusing particularly on brightness variations.

It's essential to examine illumination shifts across the entire image. When evaluating such shifts in a video, we compare an initial reference frame with successive incoming frames. This reveals the progression of illumination change. A binary analysis is conducted between the initial frame and the following frames directed into the image matting procedure, which can be mathematically expressed as:

Algorithm 1 Illumination Change Detection Algorithm

if day situation to night Situation **then**
  if $Y_{img}(i,j) < y_{mean} - y_{std}$ **then**
    $binary\ mask[i,j] = [255, 255, 255]$
  else if $Y_{img}(i,j) > y_{mean} - y_{std}$ **then**
    $binary_{m}ask[i,j] = [0, 0, 0]$
  end if
else if night Situation to day Situation **then**
  if $Y_{img}(i,j) > y_{mean} - y_{std}$ **then**
    $binary\ mask[i,j] = [255, 255, 255]$
  els if $Y_{img}(i,j) < y_{mean} - y_{std}$ **then**
    $binary_{m}ask[i,j] = [0, 0, 0]$
  end if
end if

where i is the x-coordinate and j is the y-coordinate of the image. Y_{img} is a Y-value with brightness information converted from an RGB image to the YCbCr color space. y_{mean} is the average value of the Y_{img} values for the entire image. y_{std} is the standard deviation of the Y_{img} values for the entire image. If the difference between y_{mean} and y_{std} values is greater than the Y_{img} value, the pixel value is converted to the [255,255,255] value, otherwise it is converted to [0,0,0] value to calculate the binary mask. Conversely, In the night to day situation, binary mask is calculated by converting pixels to the [255,255,255] value if the difference between y_{mean} and y_{std} less than the Y_{img} value, and [0,0,0] value if it is greater.

The binary mask is derived by calculating both the mean and standard deviation of the Y-channel values within the input image. Based on algorithm 1, the proposed method formulates a binary mask from both the initial frame fed into the image matting and the current frame, subsequently computing the difference between the two. The difference is evaluated in absolute terms, highlighting the effectiveness of the proposed method in detecting illumination variations. Fig. 4 illustrates the results of binary mask subtraction using the proposed method. 'A' denotes the starting point, 'B' represents a midpoint, and 'C' signifies when the illumination change takes place. Fig. 4a displays the subtraction outcomes between the binary masks of 'A' and 'B'. Meanwhile, Fig. 4b presents the differential result between the binary masks of 'B' and 'C'.

Based on the results, it is evident that the subtracted values of the binary mask differ based on illumination shifts. Fig. 5 shows a graph that evaluates the number of pixels in each region of the subtracted binary mask. Points A, B, and C in Fig. 5 correspond to the same reference points in Fig. 4.

(a)

(b)

FIGURE 4. Detected regions of illumination change: (a) Binary image without any illumination shift and (b) binary image indicating a detected illumination variation.

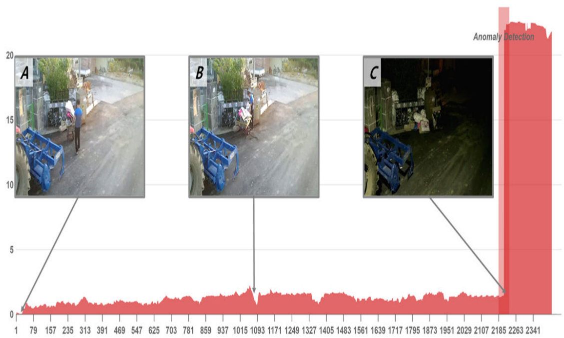


FIGURE 5. Graphical representation of detected illumination changes.

The illumination change between the initial frame at time A and the mid-point at time B is marginal. On the other hand, the contrast in illumination between times C and A—post the change in lighting—is prominently depicted in the graph.

The proposed approach thoroghly initializes the starting frame for background matting based on the identified illumination shifts. The background matting technique employed in this study identifies dynamic objects by examining pixel changes between the inaugural frame and those that follow. Given this framework, the proposed strategy ensures background matting remains resilient to illumination alterations, provided the initial frame input is updated to the frame succeeding the illumination shift.

B. ABANDONED OBJECT RECOGNITION (AoR)

1) BACKGROUND MATTING WITH DENSE ATROUS SPATIAL PYRAMID POOLING (ASPP)

In conventional approaches, abandoned object detection largely depends on methods rooted in background modeling to detect static objects. Such methods deploy background subtraction to perceive pixel alterations, thereby adaptively identifying static objects. Although these background modeling techniques demonstrate expertise in detecting static objects, they also come with their set of challenges. A major concern is the increased occurrence of false positives during significant pixel changes.

In this study, we introduce an improved approach to detect static objects by using the capabilities of the background matting technique, addressing the limitations associated with traditional background modeling. Essentially, background matting distinguishes the foreground from the background in an image. However, its distinction from background modeling lies in its diminished susceptibility to pixel changes. Because of their reduced sensitivity to pixel alterations, background matting methods, such as the one presented by Xu et al. [17], tend to outperform their background modeling counterparts, especially when paired with suitable pre-processing measures.

Traditionally, image matting has been employed in image compositing, assisting in the removal of backgrounds and the incorporation of new ones. This makes image matting algorithms stand out for their adeptness at removing backgrounds while preserving sharp foreground boundaries. With this perspective, this paper focuses on the use of an image-matting approach, optimized for accurately detecting static objects.

In image matting research, Fully Convolutional Networks (FCN) [18] and U-Net architectures [19] are notable developments as seen in studies such as [20] and [21]. However, they face challenges when dealing with high resolution and occlusion scenarios. A recent advancement tries to mitigate these issues by introducing an image matting technique that employs Atrous Spatial Pyramid Pooling (ASPP) [10], [22]. Originated from semantic segmentation, this approach uses ASPP configurations for image matting tasks. Moving away from conventional rule-based background modeling, this technique enhances the accuracy of background separation by adjusting dilation rates, thereby extracting richer details from the feature map. To enhance computational efficiency and streamline the process, parallel layers are incorporated, as depicted in Fig. 6.

The ASPP structure employs various dilation rates to gather data. However, as the dilation rate escalates, there is noticeable loss in feature information. When implemented in typical CCTV scenarios using this ASPP configuration, substantial matting errors emerge, as depicted in Fig. 6.

In this study, we introduce a network that incorporates Dense ASPP [23] to address the limitations observed in ASPP-based networks when conducting background matting. The revised architecture, which replaces the conventional ASPP section (often regarded as the ‘‘neck’’ of the background matting algorithm) with Dense ASPP, is illustrated in Fig. 8. While retaining the Backbone, Decoder, and Refiner from the ASPP-based background matting, we substituted the ASPP with the Dense ASPP. This novel ASPP variant serves as a remedy to the issue of information loss in traditional ASPP by ensuring that features are continually added during dilation. Figure 8 presents the revised network architecture, which incorporates a Dense ASPP for improved feature extraction and object detection accuracy. The figure highlights the sequential flow of data through the network, starting with the input layer where images are fed into the system. Each stage of the network is annotated to indicate the transformation and shaping of data as it progresses through the network layers. The detailed breakdown of the Dense ASPP structure is a key feature of this revision.

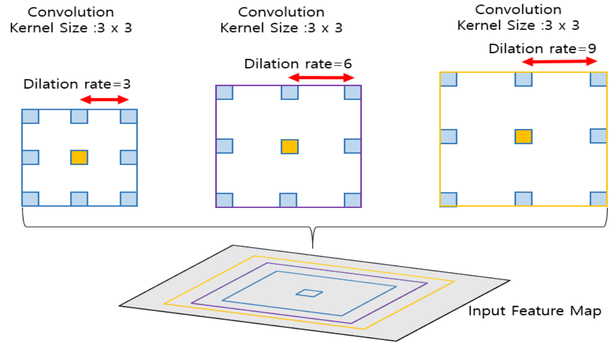


FIGURE 6. Schematic representation of the Atrous Spatial Pyramid Pooling (ASPP) structure. The design features varying dilation rates to comprehensively extract spatial information from different scales of the feature map. The integration of parallel layers in the structure emphasizes our commitment to balancing computational efficacy with the richness of information capture, streamlining the process without compromising on detail.

FIGURE 7. Visualization of outcomes obtained from employing the matting algorithm integrated with ASPP.

It outlines the specific configurations of atrous convolutions used to capture multi-scale contextual information, which is crucial for accurately identifying abandoned objects in varied scenes. The annotations include the dimensions of feature maps at various points in the network, providing insight into how the input image is processed and transformed into a feature-rich representation. Additionally, the figure captions have been expanded to describe the role of each component within the network, including how video normalization techniques are integrated to enhance the quality of input images and the specific contributions of each layer to the overall detection process. This comprehensive visualization and description aim to make the network's operations transparent and understandable, highlighting the innovations and modifications made to adapt the Dense ASPP structure for the specific challenges of abandoned object detection. This enhanced figure, with its added captions and detailed breakdown, provides a comprehensive overview of the network architecture and its components, ensuring a deeper understanding of the methodological advancements our research contributes to the field of abandoned object detection.

Test results for the proposed network are shown in Fig. 9. We used the same dilation rate as ASPP in our test. When changing to Dense ASPP with the same settings, there was a noticeable change. In contrast to Fig. 7, we see a significant reduction in the portion that corresponds to matting errors. In the dynamic object part, the separation between the object and the background is more subtle than with ASPP. Also, in the abandoned object portion, the objects were more precisely isolated to include less background than Fig. 7. For the same dilatation rate, the application of Dense ASPP results in less semantic information loss than the network with existing ASPP.

2) ABANDONED OBJECT DETECTION

In the previous section, Background Matting with Dense ASPP, we proposed an efficient way to detect dynamic and static objects that are foreground. The dual background modeling method was divided into long-term and short-term, which allowed us to efficiently separate static and dynamic objects. However, background-matting-based methods do not distinguish between dynamic and static objects, which requires additional systems. Also, an abandoned object is an object that has been discarded by a person. Therefore, many abandoned object detection algorithms use pedestrian detection to determine if a detected object is abandoned based on the relationship between the detected object and the pedestrian. For this reason, pedestrian detection is essential for abandoned object detection. In this paper, pedestrian detection is performed to remove the area of a dynamic object, a person, and to calculate the relationship for determining abandoned objects. The pedestrian detection algorithm used was YOLO V7 [24]. YOLO-based methods are one-stage detection algorithms that have been widely used in real-time object detection systems due to their fast speed and robust performance. In addition, YOLO V7 can be used by applying various modules such as pose estimation and instance segmentation, so it is suitable for the algorithm for removing pedestrian areas in this paper. Instance segmentation is a pixel-by-pixel classification of the boundaries of a trained object, excluding the background, in a bounding box region.

In the proposed abandoned object detection system, the process of abandoned object determination and dynamic object region removal is performed through pedestrian detection methods. Dynamic object regions detected by the background matting method have detailed border detection. However, the results obtained using background matting make it difficult to distinguish between abandoned objects and luggage. In this paper, we propose a method for detecting dynamic objects using the background matting method. The object that the pedestrian was holding is a category of dynamic object because it moves with the person as part of the person before being abandoned. Therefore, the proposed method uses pedestrian detection to remove pedestrian regions in order to get only objects that correspond to abandoned objects as dynamic objects. Fig. 10 shows the results of the dynamic object detection method with pedestrian region removal. Fig. 10(a) shows the original image, Fig. 10(b) shows the result of background matting, and Fig. 10(c) shows the result of removing the pedestrian area. As shown in Fig. 10(c), the results of the background matting effectively removed only the instance region of the pedestrian to detect the luggage.

However, there is noise in the non-intersecting parts of the region, depending on the results of background matching and pedestrian detection. To remove subtle noise, we used Gassian Blur and morphology methods in post-processing to denoise the image. We propose a method for detecting presumed abandoned objects by specifying the area remaining in the image after removing noise.

C. ABANDONED OBJECT DECISION BY FEATURE CORRELATION (AODFC) SYSTEM

In background matting, the first frame of a video is typically employed for matting. Yet, in diverse CCTV settings, there are instances where this initial frame contains dynamic objects. When such dynamic objects populate the initial image in the background matting approach, this leads to false positives in those regions occupied by the dynamic entities, as illustrated in Fig. 11. However, in the method we put forth, such artifacts aren't mistakenly identified as abandoned objects since they lack any association with pedestrians.

However, if a false positive occurs in an area near a pedestrian, a problem occurs where a meaningless region is detected as an abandoned object. In this paper, we propose a final abandoned object determination method using feature correlation method to solve the mentioned false detection problem. The proposed method first specifies the region of abandoned objects through the abandoned object detection performed in Section B. Fig. 12 shows a cropped image of

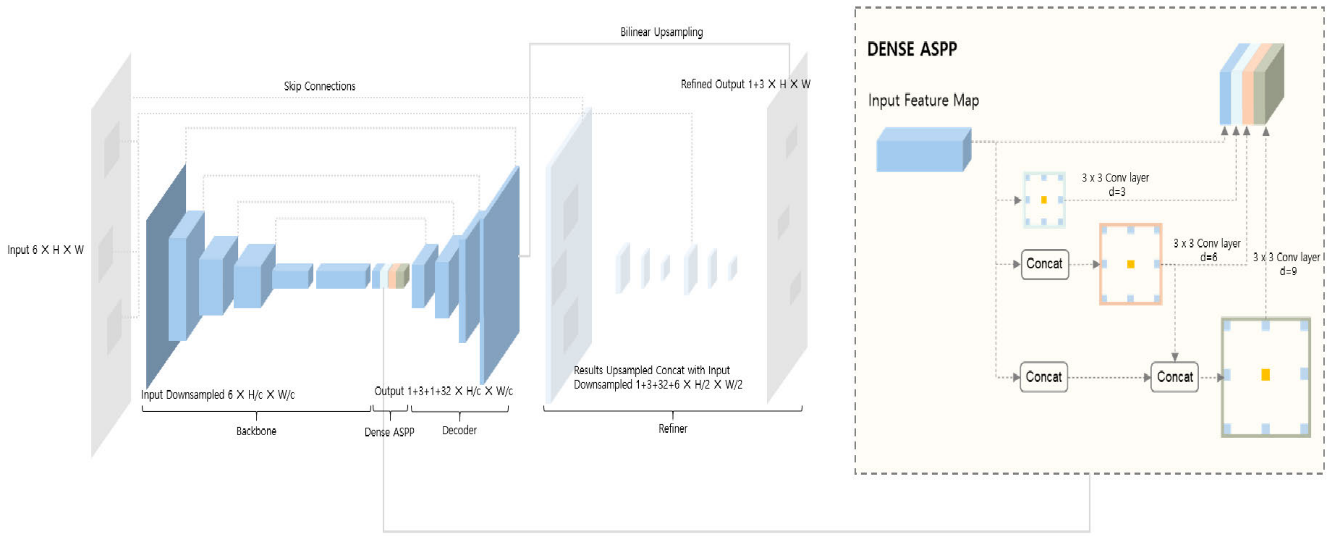


FIGURE 8. Revised network architecture featuring Dense ASPP and detailed breakdown of the Dense ASPP structure.

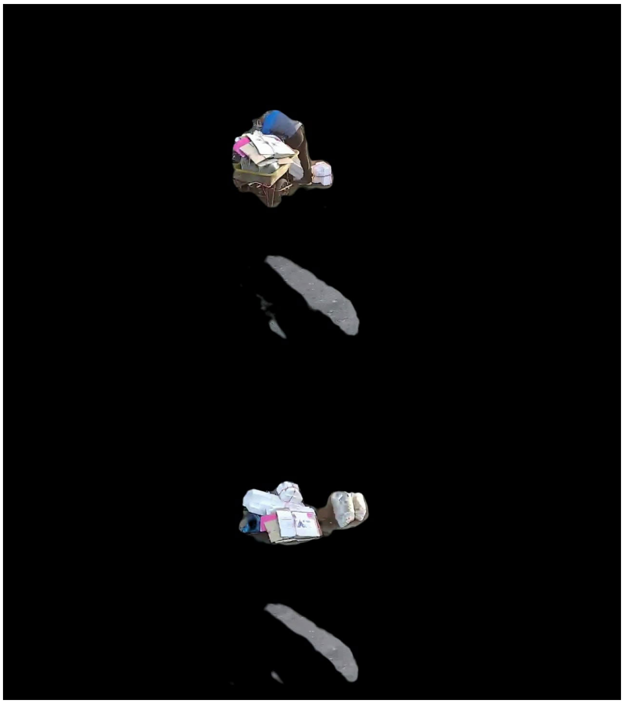


FIGURE 9. Matting outcomes derived from the implementation of the Dense ASPP algorithm.

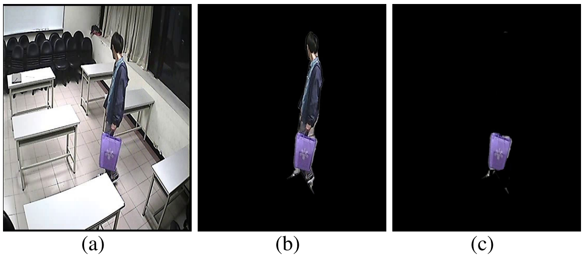


FIGURE 10. Experimental results for luggage detection: (a) original scene, (b) background matting output, and (c) post-pedestrian area exclusion.

FIGURE 11. Illustration of matting inaccuracies when the initial background frame contains dynamic entities.

the areas corresponding to the coordinates of the abandoned objects in the video sequence. In the proposed method, the final abandoned object is determined by feature correlation in the time period when the abandoned object was not present and the time period when the abandoned object was present. In Figs. 12(a), (b), and (c) show images extracted from the region before the abandoned object occurred, and Figs. 12(d), (e), and (f) show images extracted from the region after the abandoned object occurred. In this paper, in order to determine abandoned objects, it is necessary to show a low correlation value in comparison to the initial frame.

The feature correlation method proposed in this paper to determine abandoned objects is to extract the boundary of the object and compare the correlation between the edges. To compare the correlation between edges, we use the Intersection of Union (IOU) method, which is a popular evaluation method for tracking algorithms and semantic segmentation. Fig. 13 shows the result of extracting edges using the Canny algorithm [25].

The edges derived from Fig. 13 are evaluated using the IOU metric to ascertain if they correspond to an object. The

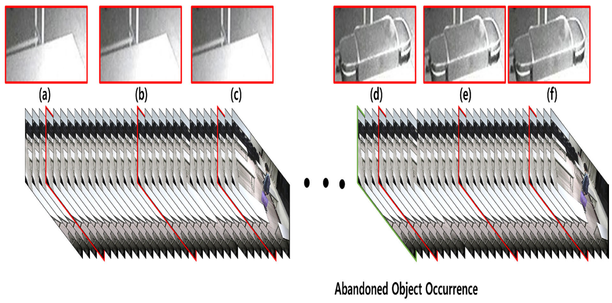


FIGURE 12. (a-c): Images from the region prior to the occurrence of the abandoned object and (d-f): Images from the region following the occurrence of the abandoned object.

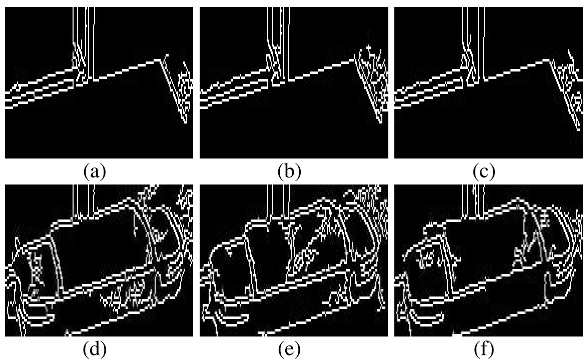


FIGURE 13. Edge detection results using Canny algorithm.

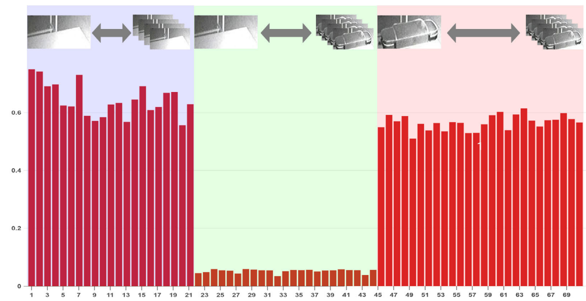


FIGURE 14. Graphical representation of iou scores in feature correlation. The blue segment in the left indicates the correlation scores between backgrounds, the green section in the center represents scores between backgrounds and abandoned objects, and the red portion in the right illustrates scores solely between abandoned objects.

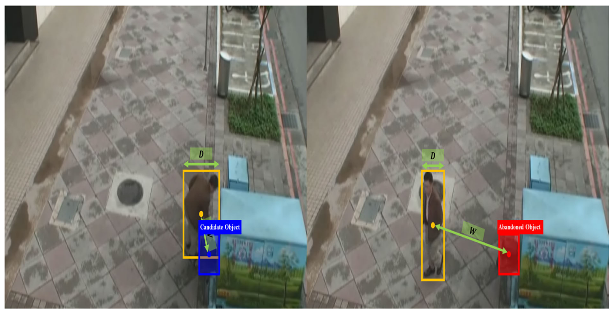


FIGURE 15. Visualization of the decision-making process for identifying abandoned objects.

formula for computing the IOU is defined as:

$$\text{IOU} = \frac{\text{area of overlap}}{\text{area of union}}. \tag{1}$$

Initially, edges are identified utilizing the Canny algorithm. Subsequently, the morphology technique is employed to emphasize the contours. Using these emphasized contours, the union and intersection of pixels at equivalent coordinates are calculated, leading to the determination of the IOU score, as given in Equation 1. In the proposed approach, this IOU score plays a pivotal role in identifying the ultimate abandoned object.

Table 1 presents the IOU scores derived from the edge extractions illustrated in Fig. 13. When comparing the boundaries of the abandoned objects against the background, the IOU scores tend to become lower. On the other hand, the IOU scores are notably higher when comparing within abandoned objects or within backgrounds.

Fig. 14 shows a graphical representation of the IOU scores. The blue segment represents the IOU scores between backgrounds, the green highlights the scores between backgrounds and abandoned objects, and the red illustrates the scores between abandoned objects. The graph clearly describes a significant difference between the scores for identical scenarios and those involving abandoned objects. In this study, we introduced a method that leverages object boundary-based feature correlation to curtail false positives in detecting abandoned objects. The effectiveness of the proposed technique in recognizing the final abandoned object is validated by the IOU scores, as shown in Fig.14.

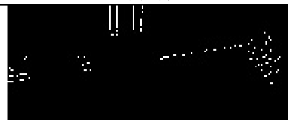
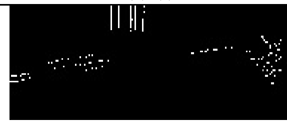
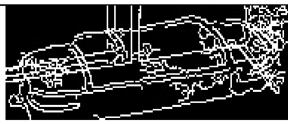
Similar to existing methods for abandoned object detection, this study employs a strategy that determines the final abandoned object by leveraging the width of the pedestrian and the distance between them and the abandoned items. As depicted in Figure 15, w represents the pedestrian's width, while d denotes the distance. In the study by Kim et al. [3], a distance d that exceeded twice the value of w was considered significant. An object was finally classified as 'abandoned' if it maintained that separation for a period exceeding 5 seconds. This research adopts the same criteria as proposed by Kim et al. to finalize the classification of an abandoned object.

In this study, we introduce a comprehensive approach to detect abandoned objects, incorporating several key innovations to address the challenges inherent in this task. In the preprocessing stage, the proposed method includes advanced noise reduction and light change detection techniques, which are crucial for preparing the input data for effective processing. We enhance background matting by incorporating features into the atrous spatial pyramid pooling (ASPP) module, significantly improving the system's ability to differentiate between background and foreground objects. Additionally, by identifying and removing dynamic objects from the analysis, we further increase the precision in detecting abandoned objects. A crucial aspect of the proposed

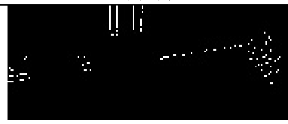
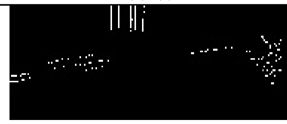
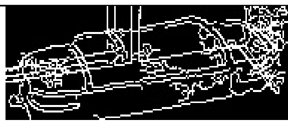
TABLE 1. Comparison of correlation values for different features.

	(a)-(b)	(a)-(c)	(b)-(c)
Area of Overlap			
Area of Union			
IOU	0.384	0.365	0.411

Correlation values with background regions

	(d)-(e)	(d)-(f)	(e)-(f)
Area of Overlap			
Area of Union			
IOU	0.416	0.354	0.404

Correlation values with abandoned object regions

	(a)-(d)	(b)-(e)	(c)-(f)
Area of Overlap			
Area of Union			
IOU	0.082	0.078	0.065

Correlation value between the abandoned object region and the background region

approach is the implementation of feature collocation analysis and the assessment of the spatial relationship between abandoned objects and people. This step is instrumental in dramatically reducing false positives, a common issue in abandoned object detection. The effectiveness of each component, along with the overall method, is rigorously validated through extensive experimentation. We conducted tests in a variety of real-world settings, encompassing diverse environmental conditions and scenarios to ensure the robustness and generalizability of the proposed approach. The experimental results demonstrate not only the superior detection accuracy of the proposed method compared to existing techniques but also validate its practical applicability and reliability in real-world contexts. Thus, the assurance of the proposed scheme's effectiveness is grounded in the methodological innovations that we have introduced and the comprehensive experimental evidence that supports their efficacy in improving abandoned object detection.

IV. EXPERIMENTAL RESULTS

In this paper, we detect abandoned objects using a network of systems proposed in the Proposed Method section. This section presents the performance evaluation and detection performance of the proposed system using ABODA [8], KISA [9],In-house and VideoMatte [10] dataset. ABODA, KISA, and In-house dataset were used to determine whether the proposed algorithm can detect abandoned objects in general situations. To evaluate the quantitative performance of background matting with Dense ASPP, we trained the VideoMatte dataset and evaluated using In-house dataset. VideoMatte dataset consisted of a total of 210,000 images, split 479:5 to form the training and validation sets. The equipment of the proposed system is TITAN XP GPU 12GB, RAM 32G, CPU E5-2640 V4 2.40GHz for training and experimentation of the proposed system. The experimental results are categorized into two parts: Evaluation of the background matting model combined with Dense ASPP, and Evaluation of the abandoned object detection performance.

A. EVALUATION OF THE BACKGROUND MATTING MODEL COMBINED WITH DENSE ASPP

In this paper, we proposed an Image Matting model that combines Dense ASPP and a series of systems for detecting Abandoned object using this model. This section describes the performance evaluation of the proposed image matting model. The evaluation [26] of replacing background matting with ASPP with Dense ASPP background mattingis performed using Mean Squared Error(MSE), Sum of Absolute Difference(SAD), and Gradient(Grad) for the alpha map.

The MSE is the average of the squared errors. Error is the difference between the predicted value by the algorithm and the ground truth value. (2) shows the expression for the MSE.

$$E = \frac{1}{2}\sum_{i=0}^{n}(y_i - \tilde{y}_i)^2 \tag{2}$$

where n represents the data amount, y_i denote the correct result for i-th training data and $\tilde{y}$ denote the predicted value with i-th training data.

MSE is characterized by a rapidly increasing loss function value as the error increases. The loss function E can be characterized as varying proportionally to the square of the error. The derivative is not constant, and as the error grows, the derivative grows as well.

SAD Error sets a window for all pixels within the disparity search range of the predicted alpha map and ground truth alpha map. For each region, the absolute value difference is calculated as shown in (3) after setting the disparity search range for the prediction alpha map and ground truth alpha map.

$$E = \sum_{i=0}^{n} |y_i - \tilde{y}_i| \tag{3}$$

where n indicates the quantity of data, y_i represents the correct value for the i-th training data, and $\tilde{y}$ represents the predicted value with the i-th training data.

After computing the absolute value difference between windows, the absolute value difference computed within a window is summed and used as the matching cost. If the difference between the pixel values in the window is large, it means that the prediction and ground truth values are not similar.

The calculation method for Grad Error measures a variety of gradients, including commonly used angular errors between gradient vectors. This is expressed as (4).

$$E = \sum_{i=0}^{n}(\nabla a_i - \nabla a'_i)^2 \tag{4}$$

where difference between the slope of a in the computed alpha map and a' in the ground truth alpha map is defined by (4). ∇a_i and $\nabla a'_i$ represent the normalized gradient values at image pixel i computed by convolving the alpha map with a first-order Gaussian derivative filter and n is the number of pixels in the image.

The performance of the proposed method was evaluated by training the background matting with ASPP and background matting with Dense ASPP on the VideoMatte dataset and comparing the errors using In-house dataset. Evaluation uses MSE, SAD, and Grad among the image matting evaluation methods of Rhemannet al. [26]. Table 2 shows the quantitative evaluation performance of the existing method and the proposed method. The results of using the background matting algorithm with Dense ASPP showed lower errors than the results of using ASPP. When comparing MSE, the model with Desne ASPP shows less error than ASPP by about 0.7 and SAD by about 0.3. Grad error is different by about 0.12. Therefore, we demonstrate that the proposed background matting model with Dense ASPP is more accurate in background matting than the existing methods in general CCTV situations. Therefore, we demonstrate that the proposed background matting model with Dense ASPP is more accurate in background matting than the existing methods in general CCTV situations.

TABLE 2. Comparison of performance between Dense ASPP and ASPP.

	MSE	SAD	Grad
Background Matting(Ours)	**3.242**	**1.441**	**1.260**
Background Matting(ASPP)	3.965	1.766	1.372

B. EVALUATION OF THE ABANDONED OBJECT DETECTION PERFORMANCE

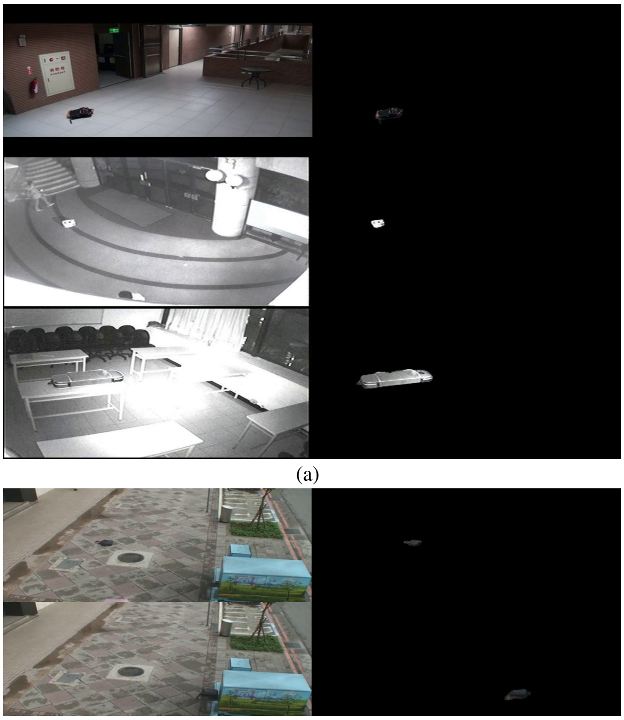
In this paper, we perform a performance evaluation of abandonment detection using the ABODA dataset, a public dataset. In addition, since the amount of ABODA datasets is not much, we evaluated the performance of the proposed method by using KISA dataset used by HLDnet and In-house CCTV dataset. The ABODA dataset consisted of a total of 11 video sets, with various scenarios for different situations. Videos 1, 9, and 10 show abandoned objects in indoor situations and Videos 2 to 4 show abandoned objects in outdoor situations. Videos 5 to 8 show indoor situations to determine whether the judgment of abandoned object detection is accurate due to light changes. Video 11 attempts to determine whether abandoned object detection in crowd situations is accurate, but it is unclear whether GT is present. Table 3 shows the comparison between the proposed abandoned object detection system and the existing abandoned object detection results. In the comparative analysis, we introduce the Precision metric to evaluate the performance of the proposed system against existing learning-based abandoned object detection methods, such as HLDnet and Shyam's method. The Precision metric, defined in Equation 5, serves as a key indicator of detection accuracy by measuring the proportion of true positive detections out of all positive detections (both true positives and false positives).

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

This metric is crucial for understanding the effectiveness of the detection system in accurately identifying abandoned objects while minimizing the occurrence of false positives. The inclusion of Precision, alongside the IoU method, provides a more nuanced and detailed assessment of detection

TABLE 3. Comparative evaluation of abandoned object detection methods on the avoda dataset.

Video	Scenario	GT	AODBM(Ours)			HLDNet [3]			Park *et al.* [27]			Lin [28]			Wahyono [29]			Patrick [30]			Shyam [31]		
			TP	FP	Precision	TP	FP	Precision	TP	FP	Precision	TP	FP	Precision	TP	FP	Precision	TP	FP	Precision	TP	FP	Precision
V1	Indoor	1	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0
V2	Outdoor	1	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0	0	1	0.0	1	0	1.0	1	0	1.0
V3	Outdoor	1	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0	0	1	0.0	1	0	1.0	1	0	1.0
V4	Outdoor	1	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0	0	1	0.0	1	0	1.0	1	0	1.0
V5	Night	1	1	0	1.0	1	0	1.0	1	0	1.0	1	1	0.5	1	0	1.0	1	0	1.0	1	0	1.0
V6	Illumination	2	2	0	1.0	2	0	1.0	2	0	1.0	2	0	1.0	0	0	0.0	2	0	1.0	2	0	1.0
V7	Illumination	1	1	0	1.0	1	0	1.0	1	0	1.0	1	1	0.5	0	0	0.0	1	1	0.5	1	0	1.0
V8	Illumination	1	1	0	1.0	1	0	1.0	1	0	1.0	1	1	0.5	0	0	0.0	1	0	1.0	1	0	1.0
V9	Indoor	1	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0
V10	Indoor	1	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0	1	0	1.0
V11	Crowded	-	-	-	-	-	-	-	-	-	-	1	3	0.25	-	-	-	1	1	0.5	-	-	-



(a)

(b)

FIGURE 16. Abandoned object detection using the AVODA dataset.

performance, facilitating a direct comparison with other state-of-the-art methods.

By incorporating these metrics into Tables 3 and 5, the revised manuscript offers a more thorough and quantifiable evaluation of the abandoned object detection system's performance. This enhancement addresses the reviewer's request for the inclusion of accuracy-related metrics and ensures that the proposed system's capabilities are assessed using established and relevant evaluation criteria. Through this approach, we aim to demonstrate the efficacy of the proposed method and its comparability, if not superiority, to existing techniques in the field.

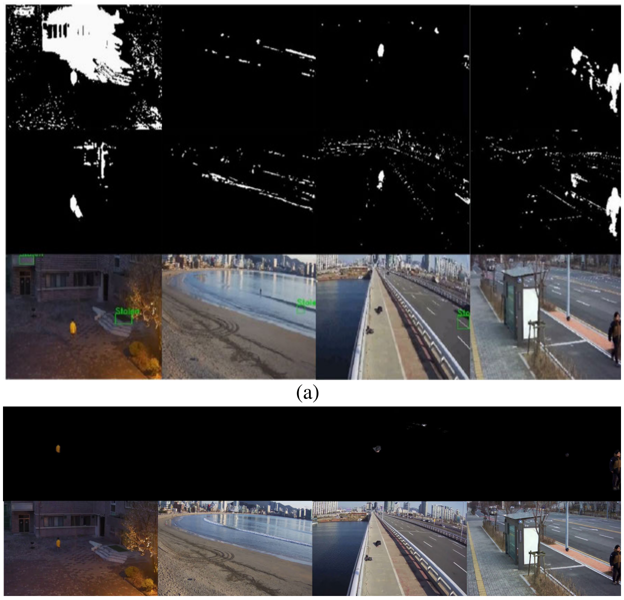
Fig. 16 shows the detection results for indoor and outdoor situations for the AVODA dataset. In Fig. 16(a), it is seen that the abandoned object is robustly separated from the background and detected in both brightly and darkly light conditions in an indoor situation. Fig. 16(b) shows that the abandoned object is accurately separated from the background even in outdoor situations, resulting in high detection performance.

TABLE 4. IOU comparison between HLDnet and our system with AVODA dataset.

Video	Scenario	HLDnet	Ours
V1	Indoor	0.8186	**0.9713**
V2	Outdoor	0.8903	**0.9104**
V3	Outdoor	0.8788	**0.9057**
V4	Outdoor	0.9398	**0.9357**
V5	Night	0.8210	**0.9039**
V6	Illumination	0.8315	**0.8716**
V7	Illumination	0.8940	**0.9240**
V8	Illumination	0.8251	**0.8872**
V9	Indoor	0.9372	**0.9429**
V10	Indoor	0.9529	**0.9763**
V11	Crowded	-	-

Performance evaluation on the Avoda dataset also demonstrates the accuracy of abandoned object detection with IOU values as shown in Table 4. In general, IOU values represent detection performance in object detection models, but in abandoned object detection, IOU values represent the number of successful detections in frames with abandoned objects. In videos 1 to 10, the proposed method performs better than HLDnet's method in detecting abandoned objects.

Existing abandoned object detection methods based on background subtraction perform well for detecting stationary objects, but are still challenging in outdoor situations. Fig. 17(a) shows the results using background subtraction with the KISA dataset. In Fig. 17(b), the results of background matting using the KISA dataset are shown. From these results, the background subtraction method performs poorly in general CCTV situations. In outdoor situations, there is a lot of pixel variation, which can cause false positives in the presence of waves or moving vehicles. However, in our proposed system, we handle pixel changes and use a background matting method that is less sensitive to pixel changes than the background subtraction method. In Fig. 17(b), the results show that the object is separated with less noise than the background subtraction method.



(a)

(b)

FIGURE 17. Comparison of our proposed method and background subtraction using the KISA dataset.

TABLE 5. Comparison of detection results using the KISA dataset, illustrating the performance of dual background, HLDNet, and the proposed method in terms of IOU metrics.

Scene	GT	Dual Background [5]	HLDNet	AODBM(ours)	Dual Background IOU	HLDNet IOU	AODBM(ours) IOU
Illumination	0	4	0	0	0	0	0
Wave	0	3	0	0	0	0	0
Road	0	2	0	0	0	0	0
Road	1	0	1	1	0	0.778	**0.891**

The dataset in Fig. 17 is used to verify the detection results. The background subtraction method causes many false positives in outdoor situations, as shown in Table 5. The abandoned object detection system, which incorporates background matting, achieves detection performance comparable to that of HLDNet. We employed the Intersection over Union (IOU) calculation method, as detailed in Table 4, to assess the precision of detected objects in comparison to ground truth annotations for both HLDNet and the proposed method. The resulting IOUs are presented in the table below, illustrating the effectiveness of the proposed approach.

HLDnet proceeds with abandoned object detection by prioritizing abandoned objects through a Gaussian filtered network. However, the process of detecting abandoned objects is affected by the coordinate information of the hand, which causes false positives near the hand position. Fig. 18 shows the results of false positives when using HLDnet.

Unlike HLDnet, the proposed system detects abandoned objects with pedestrian objects removed, so pedestrian objects are not misidentified as abandoned objects. Fig. 19 shows the result of removing pedestrian objects and detecting only abandoned objects as a binary image. It can be seen that abandoned objects are detected correctly with less false positives because objects are removed which are not abandoned objects.

As with the KISA dataset, the In-house dataset was used to determine that our system accurately detects abandoned objects in outdoor situations. Unlike the AVODA dataset, the in-house dataset contains objects in the video that look like abandoned objects. Table 6(a) and (b) show the day time video and (c) and (d) show the night time video. In the case of the background subtraction method, Table 6(a), (c) shows that the information about the abandoned object is not well discriminated. In addition, the noise in the background subtraction, as shown in Table 6(b), (d), makes it too difficult to make accurate estimates of abandoned objects. However, in our system, we use Background Matting, which we replaced with Dense ASPP, to detect abandoned objects. Even if there are many similar objects, such as abandoned objects, the background matting results in Table 6 show that it accurately separates the abandoned objects from the background. It solves the problems of existing background subtraction detection methods and detects abandoned objects more robustly.

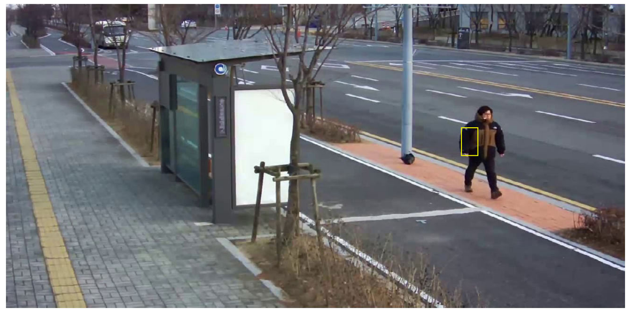


FIGURE 18. False positives problems with HLDnet.

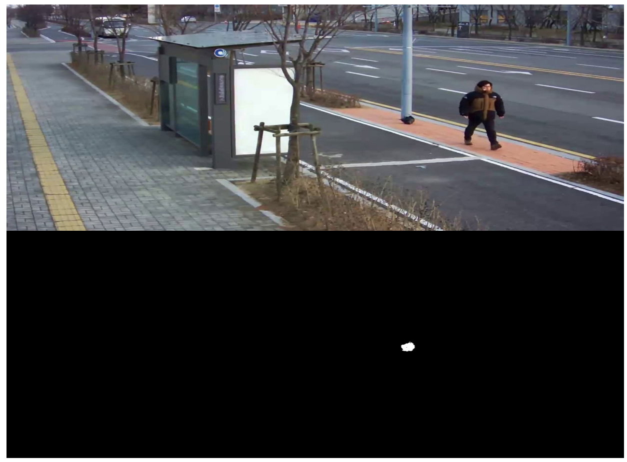


FIGURE 19. Abandoned object detection using the proposed system.

Table 7 shows a comparison of the detection results of the background subtraction method and the proposed system with the in-house dataset in Table 6. The background subtraction method fails to detect GT or produces false positives, similar to the KISA dataset detection results. However, the proposed system can be seen to accurately detect GT.

C. ERROR CASE ANALYSIS

Despite achieving a generally high detection rate for abandoned objects in CCTV footage, the proposed system

TABLE 6. Comparison of background subtraction and the proposed system using in-house dataset result.

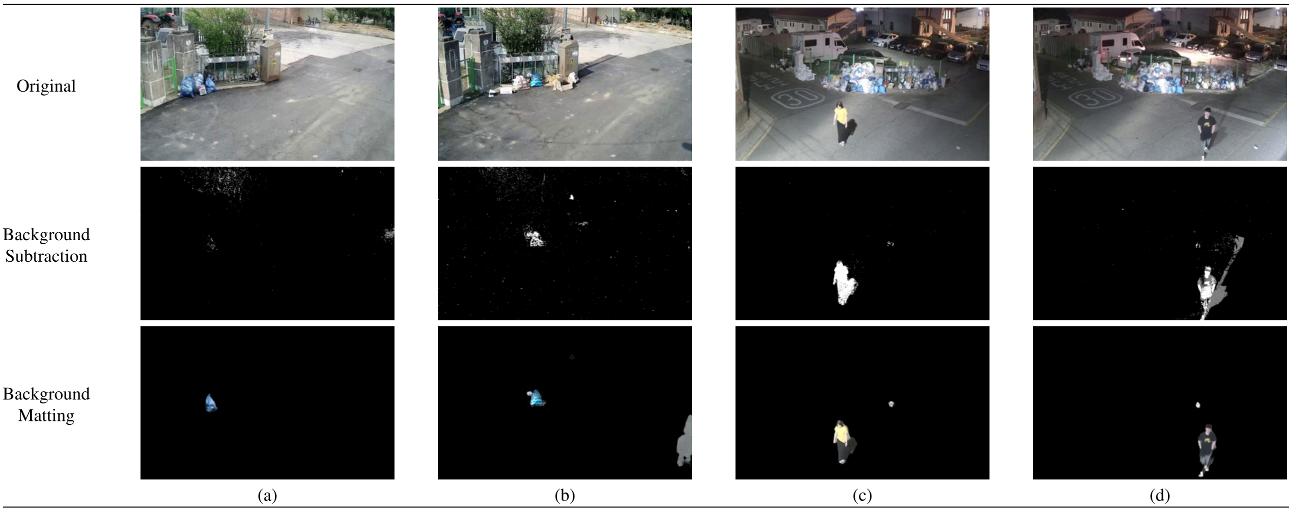


TABLE 7. Comparison of detections using in-house dataset.

Scene	GT	Background substraction [28]	AODBM(ours)
(a) Day	1	0	**1**
(b) Day	1	2	**1**
(c) Night	1	0	**1**
(d) Night	1	3	**1**

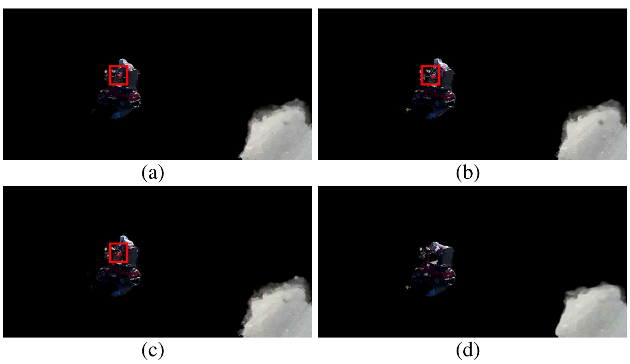


FIGURE 20. (a) to (d) illustrate error cases where the detection system fails to identify objects correctly being discarded into a container.

encounters specific error scenarios, particularly when an object is discarded into a stationary container like a box or bin. These instances often result in the non-detection of the object, as demonstrated in Figure 20, where the system fails to recognize an object being placed into a box. To enhance detection in such cases, we propose acquiring data from multiple angles and implementing exception handling mechanisms tailored to these specific scenarios. These adjustments aim to bolster the model's detection capabilities and ensure more consistent and reliable identification of abandoned objects. In this paper, we proposed an abandoned object detection method using Dense ASPP, and the experimental results were divided into two parts for performance evaluation. In Part 1, we demonstrated that the proposed background matching algorithm with DenseASPP shows lower error values than the model with ASPP, resulting in improved performance compared to the existing method. Part 2 demonstrated that the abandoned object detection method with background matting provides more accurate abandoned object detection than the background subtraction method using AVODA, KISA, and in-house datasets.In this context, we show that the abandoned object detection method proposed in this paper is suitable for utilization in real-life CCTV and surveillance systems.

V. CONCLUSION

In this paper, we proposed a system to detect abandoned objects by using background matting with Dense ASPP. The proposed system was able to reduce false positives through Pre-Processing, Abandoned Object Recognition(AOR), and Abandoned Object Decision by Feature Correlation(AODFC). to solve the errors of the existing abandoned object detection methods. In Pre-Processing, image normalization is used to address issues such as communication noise, and illumination changes are identified to eliminate false positives. The AOR system detects abandoned objects by using background matting and removing human object information to overcome the difficulty of detecting abandoned objects due to occlusion and similar objects. Finally, the AODFC System detects the final abandoned object through feature correlation analysis based on the abandoned object coordinates found in the AOR to reduce false positives. In addressing the challenges posed by background subtraction in abandoned object detection, the proposed research introduces a novel approach that significantly mitigates common issues such as sensitivity to lighting variations, the complexity of background modeling, and the handling of shadows and complex backgrounds. Traditional methods often falter when faced with these dynamic environmental factors, leading to decreased accuracy and reliability. the proposed method leverages advanced deep learning techniques to effectively overcome these obstacles, enhancing the system's ability to distinguish abandoned objects from their surroundings with greater precision. By eliminating the aforementioned problems, the proposed system not only improves the detection of abandoned objects but also sets a foundation for future advancements in the field. The use of deep learning offers a flexible and robust framework capable of adapting to the nuanced variations of outdoor environments, thus significantly reducing false positives and negatives that have plagued previous methodologies. Looking forward, we acknowledge the potential for further refinement and optimization of the proposed system. Future research directions include stabilizing and streamlining the deep learning architecture to enhance system performance and efficiency. Additionally, we advocate for the expansion of the proposed method to encompass the detection of hazardous objects such as explosives and drugs. This progression would mark a significant step forward in the development of comprehensive security and surveillance systems capable of addressing a wider range of threats.

REFERENCES

[1] F. Porikli, ''Background subtraction using adaptive kernel density estimation,'' in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Aug. 2006, pp. 1106–1113.
[2] S. Colreavy-Donnelly, F. Caraffini, S. Kuhn, M. Gongora, J. Florez-Lozano, and C. Parra, ''Shallow buried improvised explosive device detection via convolutional neural networks,'' *Integr. Comput.-Aided Eng.*, vol. 27, no. 4, pp. 403–416, Sep. 2020.
[3] D. Kim, H. Kim, Y. Mok, and J. Paik, ''HLDNet: Abandoned object detection using hand luggage detection network,'' *IEEE Consum. Electron. Mag.*, vol. 11, no. 4, pp. 45–56, Jul. 2022.
[4] Y. Xu, J. Dong, B. Zhang, and D. Xu, ''Background modeling methods in video analysis: A review and comparative evaluation,'' *CAAI Trans. Intell. Technol.*, vol. 1, no. 1, pp. 43–60, Jan. 2016.
[5] X. Yang, X. Wang, X. Wang, Y. Song, and Y. Zhang, ''Dual background modeling for robust unattended object detection in video surveillance,'' *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 6, pp. 1853–1867, 2019.
[6] Y. Chen, R. Zhang, and J. Zhang, ''Outdoor abandoned object detection based on deep learning and background subtraction,'' *IEEE Access*, vol. 7, pp. 132395–132403, 2019.
[7] H. Schulzrinne, A. Rao, and R. Lanphier, ''RFC2326: Real time streaming protocol (RTSP),'' Apr. 1998. [Online]. Available: https://www.rfceditor.org/info/rfc2326
[8] Y. Wang, R. Zhang, and J. Zhang, ''Abandoned object detection based on multimodal fusion and domain adaptation,'' in *Proc. IEEE 16th Int. Conf. Inf. Fusion*, Sep. 2019, pp. 1–8.
[9] K. T.-H. Kim and S.-H. Park, ''Abandoned object detection in video surveillance using a dual background model,'' in *Proc. 14th Int. Conf. Inf. Fusion*, 2016, pp. 1–8.
[10] S. Lin, A. Ryabtsev, S. Sengupta, B. Curless, S. Seitz, and I. Kemelmacher-Shlizerman, ''Real-time high-resolution background matting,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8758–8767.
[11] I. Smeureanu, L. Ciobanu, C. Popa, M. Miron, N. Ilie, and A. Ionescu, ''Abnormal object detection for smart cities using video surveillance,'' in *Proc. 17th Int. Conf. Adv. Concepts Intell. Vis. Syst. (ACIVS)*, 2021, pp. 1–6.
[12] S. Park and K. Hong, ''Dual background subtraction based on long-term foreground information for illumination variations,'' in *Proc. 12th Int. Conf. Inf. Fusion*, 2009, pp. 1–8.
[13] W. Y. Liu and J. Zhang, ''Abandoned object detection using pixel-based finite state machine and single shot multibox detector,'' in *IEEE Int. Conf. Image Process. (ICIP)*, Jul. 2021, pp. 4587–4591.
[14] D. Shen, G. Wu, and H.-I. Suk, ''Deep learning in medical image analysis,'' *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.
[15] L. Fan, Y. Ding, D. Fan, D. Di, M. Pagnucco, and Y. Song, ''GrainSpace: A large-scale dataset for fine-grained and domain-adaptive recognition of cereal grains,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 21084–21093.
[16] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, ''MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9584–9592.
[17] N. Xu, B. Price, S. Cohen, and T. Huang, ''Deep image matting,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 311–320.
[18] J. Long, E. Shelhamer, and T. Darrell, ''Fully convolutional networks for semantic segmentation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
[19] O. Ronneberger, P. Fischer, and T. Brox, ''U-Net: Convolutional networks for biomedical image segmentation,'' in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.
[20] Y. Zhang, L. Gong, L. Fan, P. Ren, Q. Huang, H. Bao, and W. Xu, ''A late fusion CNN for digital matting,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7461–7470.
[21] Y. Li and H. Lu, ''Natural image matting via guided contextual attention,'' in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 11450–11457.
[22] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, ''Encoder–decoder with atrous separable convolution for semantic image segmentation,'' in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.
[23] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, ''DenseASPP for semantic segmentation in street scenes,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3684–3692.
[24] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, ''YOLOV7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.
[25] J. Canny, ''A computational approach to edge detection,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vols. PETS-8, no. 6, pp. 679–698, Nov. 1986.
[26] C. Rhemann, C. Rother, J. Wang, M. Gelautz, P. Kohli, and P. Rott, ''A perceptually motivated online benchmark for image matting,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1826–1833.
[27] 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.
[28] M. K. S. Patil and P. G. A. Kulkarni, ''Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance,'' *IARJSET*, vol. 5, no. 9, pp. 7–13, Sep. 2018.
[29] 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.
[30] P. Krusch, E. Bochinski, V. Eiselein, and T. Sikora, ''A consistent two-level metric for evaluation of automated abandoned object detection methods,'' in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 4352–4356.
[31] 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)*, Jul. 2018, pp. 1–6.

MINGU JEONG was born in Gumi, South Korea, in 1992. He received the bachelor's degree in electric engineering from Korea National University of Transportation, South Korea, in 2019. He is currently pursuing the Master of Science degree in digital imaging engineering with Chung-Ang University. His research interests include object detection, object recognition, artificial intelligence, and image matting.

DOHUN KIM was born in Seoul, South Korea, in 1992. He received the Ph.D. degree in imaging engineering from Chung-Ang University, in 2024. From 2019 to 2022, he was a dedicated Researcher with the Mobility Platform Center, Korea Electronics Technology Institute (KETI). Since 2024, he has been the Cooperative Director of the Institute of Electronics and Information Engineers (IEIE), South Korea. Currently, he is with the Intelligent Information Convergence Research Laboratory, Electronics and Telecommunications Research Institute (ETRI), South Korea.

JOONKI PAIK (Senior Member, IEEE) was born in Seoul, South Korea, in 1960. He received the B.S. degree in control and instrumentation engineering from Seoul National University, in 1984, and the M.S. and Ph.D. degrees in electrical engineering and computer science from Northwestern University, in 1987 and 1990, respectively.

He began his career at Samsung Electronics, from 1990 to 1993, where he played a key role in designing image stabilization chipsets for consumer camcorders. In 1993, he joined as a Faculty Member with Chung-Ang University, Seoul, where he is currently a Professor with the Graduate School of Advanced Imaging Science, Multimedia, and Film. From 1999 to 2002, he was a Visiting Professor with the Department of Electrical and Computer Engineering, University of Tennessee, Knoxville. Since 2005, he has been the Director of the National Research Laboratory, South Korea, specializing in image processing and intelligent systems. He was the Dean of the Graduate School of Advanced Imaging Science, Multimedia, and Film, from 2005 to 2007, and the Director of the Seoul Future Contents Convergence Cluster. In 2008, he was a full-time Technical Consultant with the Systems LSI Division, Samsung Electronics, where he developed various computational photographic techniques, including an extended depth of field systems. He has a notable influence in scientific and governmental circles in South Korea. Since 2020, he has been the Vice President of Academic Affairs with Chung-Ang University. In an exceptional move in 2021, he simultaneously assumed the roles of the Vice President of Research and the Dean of the Artificial Intelligence Graduate School, Chung-Ang University, for a one-year term. Expanding his scope of responsibilities, in 2022, he accepted a five-year appointment as the Project Manager of the Military AI Education Program under Korea's Department of Defense. With a career spanning over three decades, he has made significant contributions to the fields of image processing, intelligent systems, and higher education.

Dr. Paik is a member of the Presidential Advisory Board for Scientific/Technical Policy with the Korean Government and serves as a Technical Consultant for computational forensics with the Korean Supreme Prosecutor's Office. His accolades include being a two-time recipient of the Chester-Sall Award from the IEEE Consumer Electronics Society. He also received the Academic Award from the Institute of Electronic Engineers of Korea and the Best Research Professor Award from Chung-Ang University. He has actively participated in various professional societies. He served the Consumer Electronics Society of the IEEE in several capacities, including as a member of the editorial board, Vice President of International Affairs, and Director of the Sister and Related Societies Committee. In 2018, he was the President of the Institute of Electronics and Information Engineers.