

Research Article

AI-Based W-Band Suspicious Object Detection System for Moving Persons: Two-Stage Walkthrough Configuration and Recognition Optimization

Zheng Wen ,¹ Keping Yu ,² Xin Qi ,² Toshio Sato ,² San Hlaing Myint ,² Kazuhiko Tamesue ,² Yutaka Katsuyama ,² Hironori Dobashi,³ Yasushi Murakami,³ Ikuo Koyama,³ Kiyohito Tokuda,² Wataru Kameyama,^{1,2} and Takuro Sato²

¹School of Fundamental Science and Engineering, Waseda University, Tokyo 169-8050, Japan

²Global Information and Telecommunication Institute, Waseda University, Shinjuku, Tokyo 169-8050, Japan

³Toshiba Infrastructure Systems & Solutions Corporation, Kawasaki 212-8585, Japan

Correspondence should be addressed to Xin Qi; samqixin@aoni.waseda.jp

Received 10 March 2022; Revised 10 May 2022; Accepted 18 May 2022; Published 10 June 2022

Academic Editor: Kuruva Lakshmanna

Copyright © 2022 Zheng Wen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, terrorist attacks have been spreading worldwide and become a public hazard to human society. The suspicious object detection system is an effective way to prevent terrorist attacks in public places. However, traditional systems face two main challenges: First, they need to conduct security checks at the entrance one by one, which leads to crowding; second, they rely heavily on screeners' ability to understand security images, which can easily lead to misjudgment. To address these issues, we propose an AI-based W-band suspicious object detection system for moving persons that can perform a two-stage walkthrough screening for suspicious objects in an open area to maintain high throughput. The 1st screening uses millimeter wave radar and cameras to automatically screen suspects who may have concealed suspicious objects in an open area. The 2nd screening involves security personnel using a hybrid imager with active and passive imaging capabilities to identify the specific suspicious objects carried by the suspect. Convolutional neural network (CNN) based artificial intelligence (AI) technology will be used to improve the accuracy and speed of suspicious object detection. We performed an experiment to validate the proposed system. The usability and safety of the system are demonstrated by recognition rate (aka accuracy rate) or both recall and precision rate. In addition, in the process of improving the suspicious object recognition rate by AI techniques, we use generative adversarial network to help build a suspicious object database and successfully validate the effectiveness of the method and the factors affecting the suspicious object recognition rate to optimize the system.

1. Introduction

Terrorism is a great damage to the world peace, which can affect the security of the whole country and cause great harm to the social stability. Moreover, terrorist attacks can cause a large number of casualties and property damage, threatening people's lives and property security. So far, terrorism is a disaster that has affected many countries and taken the lives of countless innocent people [1]. Open areas are characterized by a large number of people, complex composition, and dense distribution; plus, the existence of uncertainty factors can easily become the main target of terrorist attacks

[2]. How to prevent terrorist attacks is a global scholarly research focus [3]. Security screening system is an effective way to prevent terrorist attacks in open area. At present, security screening systems (such as metal detectors [4], X-ray scanning detectors [5], and surveillance cameras [6]) have been used in airports, railway stations, bus stations, and other crowded places, playing an important role in the prevention and suppression of crime and protection of people's lives and property [7]. At the same time, the traditional security screening equipment is also facing challenges [8]. On the one hand, the traditional security equipment "blind spot", gun parts, ceramic knives, bottles of alcohol, and other

dangerous goods are easy to ignore, bringing security risks. On the other hand, the security checker uses manual judgment, which depends largely on the ability of the screeners to understand the X-ray fluoroscopic images, requiring long periods of concentration by the screeners, which can easily lead to misjudgments and missed judgments, and the cost of personnel in the process of use far exceeds the cost of equipment. Moreover, the current method is to conduct security checks at each gate entrance one by one (e.g., airport security), which is feasible but inefficient, seriously affecting the throughput of open areas and leading to massive congestion. Therefore, how to automatically perform security detection of moving people in open areas and maintain high throughput has become an urgent problem to be solved. Millimeter wave imaging [9–11] and artificial intelligence [12, 13] technologies, as promising emerging technologies, can be well applied to security screening systems to improve the probability of identifying dangerous objects.

In recent years, as the global counter-terrorism situation has become increasingly severe, security screenings at airports, high-speed railways, large event venues, and key government departments have received widespread attention [14]. In the face of the new situation, security screening technology has received unprecedented development. In the electromagnetic wave band, millimeter wave has a shorter wavelength than the low frequency microwave wave, so for a given antenna aperture, has a higher spatial resolution and accuracy [15]. Compared with the high-frequency terahertz wave and far-infrared light, the millimeter wave can penetrate the surface of clothing and is quasi-all-weather work capable. It is very suitable for coping with the current explosive devices and prohibited items concealed, diversified, miniaturized features, and is an ideal frequency band for human security screening [16]. In addition, compared to X-rays, the millimeter wave has a non-ionizing property and is therefore considered no hazard for human exposure.

With the development and application of artificial intelligence technology [17], dangerous goods intelligent identification technology has been gradually applied to the traditional security industry. This technology can reduce labor costs and improve the efficiency of security checks [18]. It does not require large-scale replacement of security equipment, and does not change the existing security model, only the use of software or hardware upgrades; it can realize the intelligent upgrade of the security machine. In the process of using artificial intelligence technology to identify suspicious objects, suspicious object database is very important because it directly affects the recognition rate of suspicious objects by artificial intelligence technology. The ultimate goal is to obtain a higher rate of suspicious object recognition. For this purpose, high-resolution data from millimeter-wave band imagers and object recognition using artificial intelligence are applied. We use active/passive imagers [19] and simulation to collect and generate object images for suspicious object database to acquire training and evaluation data for object recognition. We also use the generative adversarial network (GAN) [20, 21] to augment object images into more inconsistent images for training data, making the recognition performance better.

Seeking to mitigate these limitations and based on [22], we configure a two-stage walkthrough screening of AI-based W-band suspicious object detection system for moving persons and optimization its recognition rate. Different from [22], this paper uses multiple active/passive images (by hybrid imager) for the 2nd screening to identify the suspect object, while [22] uses only one passive image in the process. The specific contributions are as follows:

- (1) This paper proposes a W-band suspicious object detection system, which is capable of uninterrupted security inspection for moving persons in open areas. The system also applies artificial intelligence techniques to improve the recognition rate of suspicious objects by this solution
- (2) This paper verifies the usability and security of the proposed system through experiments. The factors affecting the recognition rate of suspicious objects are evaluated, and the ways to optimize the system and improve the quality of service are analyzed

The rest of this paper is organized as follows. In Section I, we present the need for AI-based W-band suspicious object detection systems for moving person. Section II presents the recent advances regarding surveillance network system, millimeter wave technology, and artificial intelligence techniques for image detection. In Section III, we detail the purpose of the proposed system, the system architecture, and the system configuration. In Section IV, the system is experimentally verified and optimized. Finally, the paper concludes in Section V.

2. Related Works

2.1. Surveillance Network Systems. In today's society, there is a large population and many security risks. As people's security requirements increase and economic conditions improve, the number of surveillance cameras is growing faster and faster, and the coverage is becoming more and more extensive [23]. Traditional video surveillance only provides simple functions such as video capture, storage, and playback used to record what happened; it is difficult to play the role of early warning and alarm. If we want to ensure real-time monitoring of abnormal behavior effectively, we need surveillance personnel to monitor the video at all times. Monitoring efficiency depends on the experience and conscientiousness of the monitoring personnel. When surveillance personnel are faced with multiple surveillance videos for a long time, it is difficult to react to anomalies in a timely manner. Therefore, this is an urgent need for intelligent video surveillance, to assist the work of surveillance personnel [24].

A huge monitoring network can instantly generate a huge amount of video data. How to efficiently extract useful information from this huge amount of data becomes a problem to the intelligent video surveillance technology. Specifically, intelligent video surveillance technology means making the computer like a human brain and the cameras like human eyes, analyzing the images to understand useful

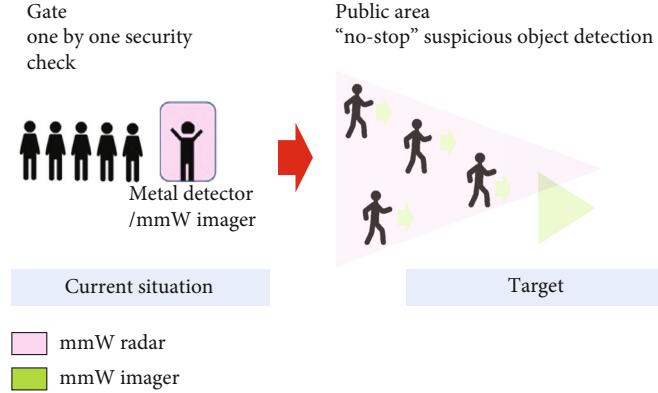


FIGURE 1: The objective of designed system.

content of the scene and behaviors [25]. Our proposed system uses radio waves to capture objects and transform them into usable data for the recognition system. The recognition system can detect suspicious objects and further alarm the security personnel.

2.2. W-Band Unidentified Object Detection. Imaging systems are a core component of millimeter wave security screening equipment. This technology can be divided into passive millimeter wave (PMMW) imaging and active millimeter wave (AMMW) imaging according to the mode of operation [26]. PMMW imaging [27] uses a millimeter wave radiometer to capture the thermal radiation or sky background scattering distribution characteristics of the target and generate images. AMMW imaging [28] uses a transmitter to emit a millimeter wave signal with a certain power to irradiate the object under test and a receiver to capture the signal reflected from the object under test, record its amplitude and phase information, and reconstruct the spatial scattering intensity of the object under test. Compared with passive imaging system, active imaging is less affected by environmental factors, obtains more information, can realize real-time imaging, and has better image quality than passive imaging system. Therefore, AMMW technology has become one of the most promising technologies for imaging human concealed objects to address the current security situation [29].

2.3. Artificial Intelligence for Image Detection. Target recognition, on the other hand, is a popular research direction in computer vision and has been widely used in areas such as face recognition, autonomous driving, target recognition, and tracking.

Regions with CNN features (R-CNN) [30] was proposed at UC Berkeley in 2014, which establishes a rich feature hierarchy for accurate target detection and semantic segmentation, allowing both bottom-up use of large convolutional neural networks (CNN) to localize and segment targets, and also allows supervised pretraining for auxiliary tasks before task-specific tuning when training data is insufficient, resulting in significantly fewer parameters during operation and achieving significant performance improvements that are more efficient than traditional methods. Although this

method has achieved some success, there are many problems with training R-CNNs, i.e., the method must generate a scheme for the training data, then apply the CNN feature extraction to each region, and finally train a support vector machine (SVM) classifier. In 2015, Ross Girshick published Fast R-CNN [31], which is similar to R-CNN in that it uses selective search to generate objects. But unlike R-CNN, which extracts all the features of each region independently and then uses an SVM classifier, fast R-CNN uses a CNN on the entire image and then pools the feature mapping using a “Region of Interest” (RoI). R-CNN finally uses a pre-feedback network for classification and regression. The third iteration of the R-CNN family is as follows: the faster R-CNN. The region proposal network (RPN) was added, moving away from selective search algorithms and allowing end-to-end training.

In 2016, Redmon published a paper on Unified Real-Time Object Detection (YOLO) [32, 33], which presented a simple convolutional neural network approach with good results and speed, and was the first implementation of real-time object detection. The task of the RPN is to output objects based on their attribute scores and then classify them using RoI pooling and fully connected layers.

3. AI-Based W-Band Suspicious Object Detection System for Moving Persons: System Configuration

3.1. Objective. To improve security in public places (especially in densely populated areas), security checks of suspicious objects should be carried out automatically and efficiently. The traditional method is to conduct security checks one by one at each entrance, which leads to over-crowding and inefficiency. Therefore, there is an urgent need for a new security check method capable of automatically performing suspicious object detection on moving people. The objective of this paper is to propose an AI-based W-band suspicious object detection system that can perform nonstop security check for moving persons in open areas. To achieve this goal, a two-stage walkthrough screening system for open areas is proposed. The 1st screening uses millimeter wave radar to detect suspects who may have

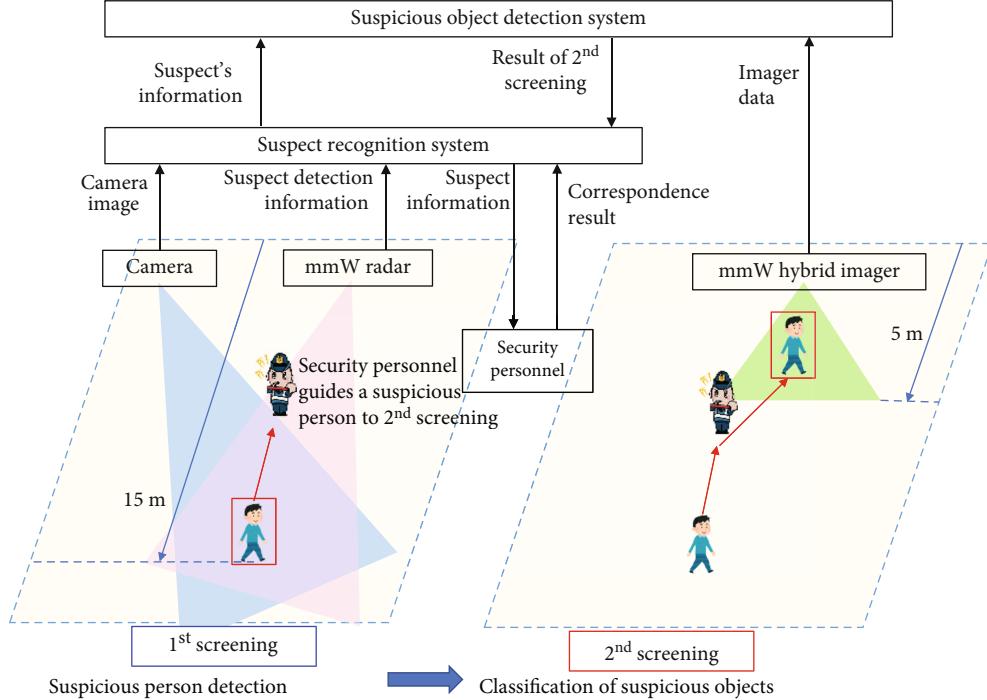


FIGURE 2: System architecture.

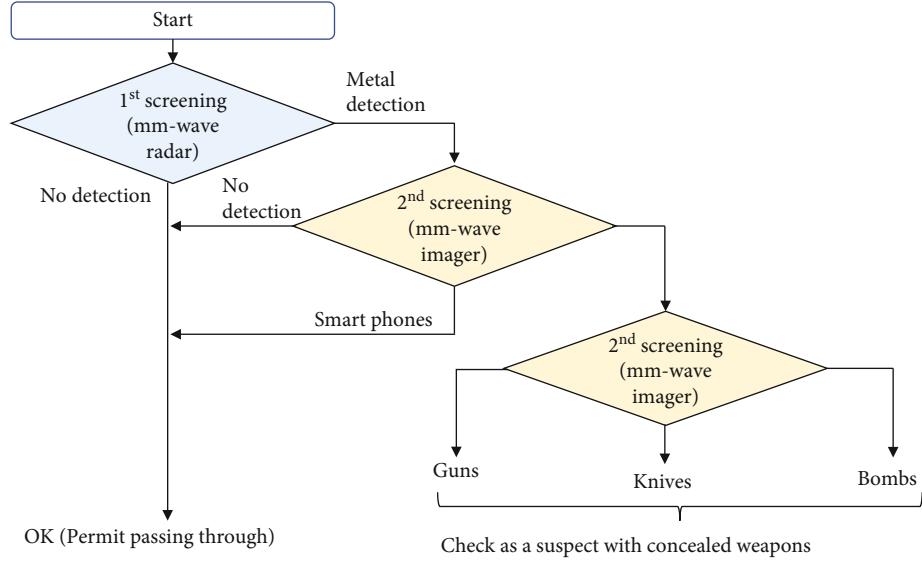


FIGURE 3: Two-stage screening process.

concealed dangerous objects [34]. The radar receives signals returned by the suspects and determines whether this person is potentially in possession of suspicious objects. At the same time, visible light cameras track suspects who are detected as possibly having certain dangerous objects. A 2nd screening uses a hybrid imager with active and passive imaging capabilities to investigate the suspect in detail and identify what suspicious objects he/she has. Also, the system applies artificial intelligence techniques to improve the recognition rate of suspicious objects by this solution. For this purpose, we built a database of suspicious objects for CNN training through simulations and experiments and tried to generate

millimeter wave images for CNN training using GAN and evaluate its feasibility. As shown in Figure 1, the goal of this paper is to develop AI-based sensing/imaging techniques in the W-band (75-110 GHz) to identify suspicious objects on moving people.

3.2. System Architecture. To perform suspicious object detection without stopping the flow of people, the system uses a two-stage screening method (1st screening/2nd screening) to identify suspicious objects hidden in the human body (Figure 2). The use of two-stage screening avoids unnecessary screening of persons without suspicious objects.

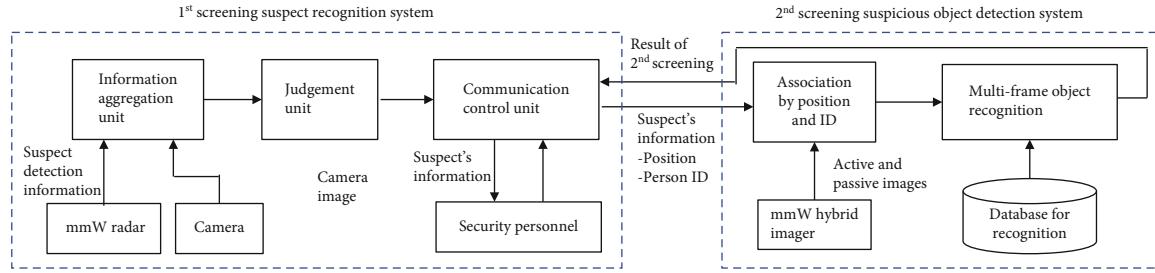


FIGURE 4: Configuration of suspect recognition system and suspicious object detection system.

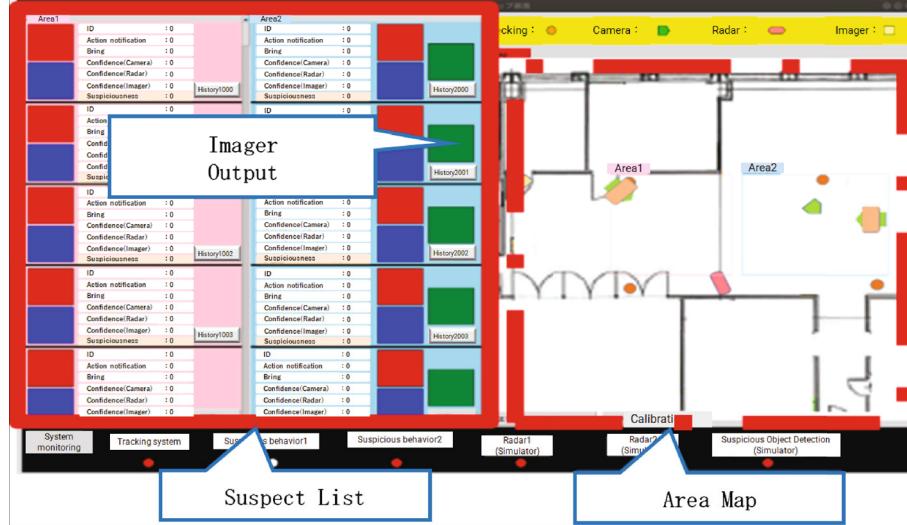


FIGURE 5: Example of a screenshot for the suspect recognition system.

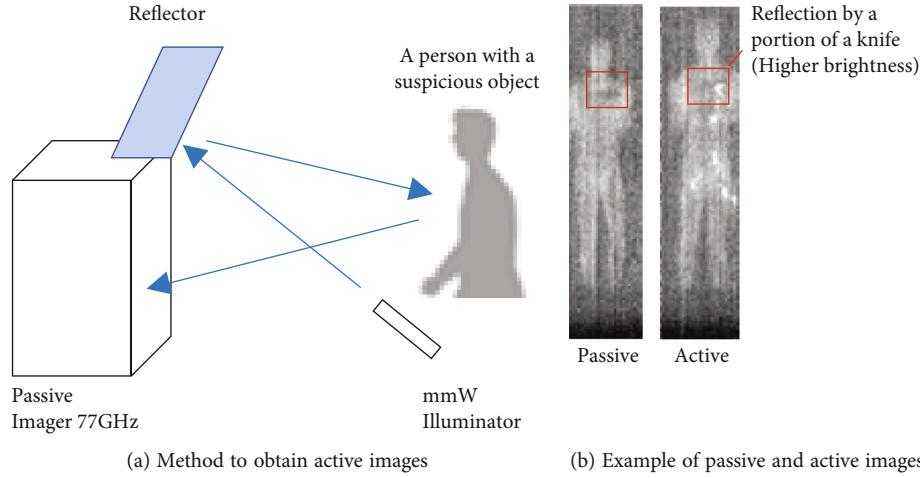


FIGURE 6: Active and passive images.

Detailed screening of suspects can be achieved at the same time. During this two-stage screening process, the mmW radar's focusing length is 15 meters; when a person walks into 1st screening area within 15 m, 1st screening procedure starts. If the detection result is positive on suspicious object, visible camera image pairs with the person, and a 2nd screening procedure will take place. A security personnel guides the person into mmW hybrid imager's range, 5

meters, and the system uses passive and active images to quickly classify the suspicious object. This configuration can increase the throughput of security screening, especially in open areas.

3.2.1. 1st Screening. In the 1st screening, visible light cameras can monitor each person in the surveillance area and record their facial images and related information. Meanwhile, the

TABLE 1: Total system configuration summary.

Screening	Objective	Method
1st screening suspect recognition system	Detect persons with metal objects or suspicious movement for wider area	Detect metal object by using a mmW (78GHz) radar Detect suspicious movement by analyzing camera images
2nd screening suspicious object recognition system	Classify and detect concealed weapon while walking in a pathway	Capture active and passive mmW (77GHz) images by using a hybrid imager Recognize concealed weapon using CNN

W-band millimeter wave radar identifies reflected waves from people within approximately 15 meters and detects people with unusual reflective properties. In the process, the suspect detection system automatically matches the millimeter wave image of the same person detected with the visible image. Once the suspect detection system receives the suspect's information from the radar, the system searches for the suspect's image from the camera image and begins tracking to identify the suspect and determine his/her current location and trajectory and sends the suspect to nearby security personnel for 2nd screening.

3.2.2. 2nd Screening. In the 2nd screening, a millimeter wave hybrid imager consisting of active and passive imagers will be used within 5 meters to identify detailed suspicious objects (knives, guns, scissors, etc.) held by the suspect. In this process, visible light cameras will also be used to help security personnel identify suspicious persons, and the generated visible light images will be correlated with millimeter wave images. From the acquired images, the suspicious object detection system determines if the suspect has some hidden objects and what dangerous objects he/she has. This system can also correspond to the movement of the suspect across one area of the "suspect object detection system" to another. In addition, artificial intelligence technology is used at this stage to increase the probability of identifying suspicious objects.

The flow of the 1st screening and the 2nd screening is shown in Figure 3. The 1st screening will primarily detect people possibly carrying suspicious objects through millimeter wave radar, and the 2nd screening will capture suspicious objects carried in clothing through data analysis from a hybrid imager. The final classification will be divided into two groups: those who do not carry suspicious items and those who will undergo final inspection by security personnel.

3.3. Suspect Recognition System (1st Screening). The suspect recognition system is primarily used for the 1st screening, detecting moving persons with suspicious objects, because of the limitation of the radar focusing with a length of 15 meters. When a detected person enters the illuminated area of the millimeter wave radar (typically 15 m), the radar identifies the person and begins to detect if the person has a suspicious object. If the millimeter wave radar determines that the person may be concealing a suspicious object, it notifies the suspect recognition system. The system associ-

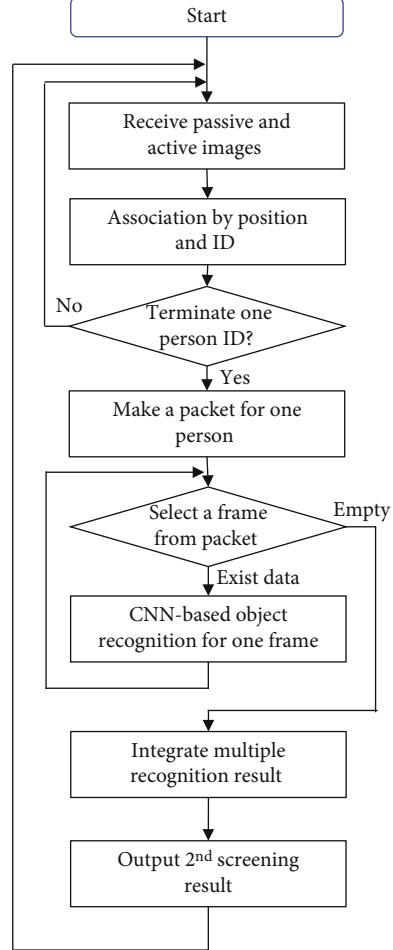


FIGURE 7: Flow of multiframe object recognition.

ates the suspect's camera image with relevant information. It then sends an alert combining this information to nearby security personnel and the suspicious object detection system. The suspect recognition system consists of an information aggregation unit, a judgment unit, and a communication control unit, as shown in Figure 4. The information aggregation unit receives information from the millimeter-wave radar about the millimeter-wave radar ID of the potential suspect detected, the time of detection, his/her location, direction of movement, speed of movement, and the relevant information is sent once every second. Note that the suspect's certainty is calculated based on the received reflected wave characteristics of the

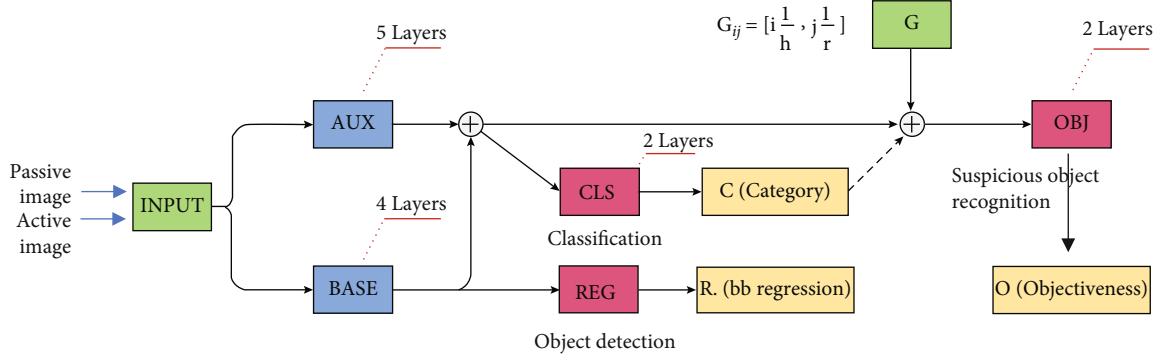


FIGURE 8: Neural networks to classify objects.

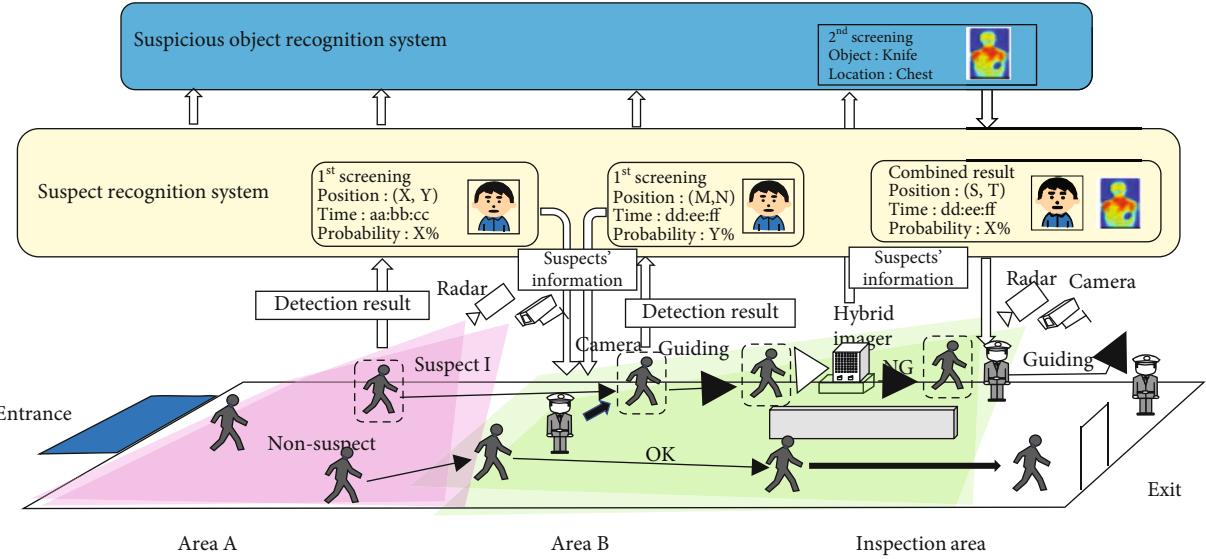


FIGURE 9: The whole system outline.

millimeter wave radar. This unit then matches the received information with the suspect's camera image and assigns a uniform ID to each person. All information will be transmitted to the judgment unit for suspicion determination and contacted with the suspect object detection system and security personnel through the communication control unit. The security personnel will pass the suspect through the suspicious object detection system for a 2nd screening and receive the results of the suspect's investigation.

Figure 5 shows an example of a screenshot for the suspect recognition system. The table in the left half shows the suspect's personal information, the camera image, and the certainty of carrying dangerous objects. The right half plots a map of the surveillance area, including the suspect's current location and trajectory. Each suspect is distinguished by ID and color on the map, making it easy for security personnel to monitor in real time.

3.4. Suspicious Object Detection System (2nd Screening). Suspicious object detection systems identify suspicious objects held by people by capturing them walking in that environ-

ment and inputting the acquired passive and active images into an AI-based neural network. Among different AI techniques, convolutional neural networks (CNNs) are representative deep learning techniques for image recognition and image classification. Two key features of CNNs distinguish them from other neural networks, namely, reducing computational complexity and ensuring translational invariance. The first part is used to extract features and includes a convolutional layer, a pooling layer, and a batch normalization layer. The second part works in the same way as the neural network and is used to perform classification, including the flatten layer and the fully connected layer. In this paper, we use CNN technology directly, so the performance of the AI part is determined by the CNN. The technical details of the CNN will not be described here.

As shown in Figure 4, the suspicious object detection system uses imager data from the millimeter wave hybrid imager (Figure 6) to identify the type of suspicious object within a 5-meter range. Figure 6(a) and Table 1 show the combination of a 77GHz mmW imager and millimeter wave illuminator; this active imaging method can provide active and passive images. Figure 6(b) shows example images from passive and active

methods. Objects are illuminated brighter in active images. Figure 7 shows the flow of multiframe object recognition. Imager data for the same ID is associated and selected by position in a packet. The multiple frames in the packet are processed to detect and recognized by using CNN classifier as shown in Figure 8. The CNN outputs the type of suspicious object (knife, gun, bomb, etc.) as the “C (category)” in Figure 8 and the probability of identification (the “O (objectiveness)” in Figure 8). The recognition results from multiple frames are selected based on majority voting among data with higher O (objectiveness) in the last part of Figure 7. Finally, the identification results are sent to the suspect recognition system.

3.5. System Implementation. The entire system is implemented in two consecutive areas (areas A and B) for two-stage screening (Figure 9). The suspect recognition system is implemented in two PCs: a PC for handling output data of the radar and camera images (OS: Ubuntu 18.04, CPU: Intel Core i7-10750H 2.60GHz, memory: 32GB memory, GPU: NVIDIA GeForce RTX2080) and a PC for display of results (OS: Ubuntu 18.04, CPU: Intel Core i7-9550H 2.60GHz, memory: 32GB memory). The suspicious object detection system is developed by Python 3.8 on Ubuntu 18.04 with a high-performance PC (CPU: Intel Xeon Gold 6246 3.3GHz [2 PCs], memory: 192GB, GPU: NVIDIA Quadro RTX8000 48GB). The entire Area A and one part of Area B use millimeter wave radar and visible light cameras for the 1st screening (suspect recognition system) to search for suspects who may be hiding suspicious objects, as shown in Figures 3 and 4. The other part of Area B uses millimeter wave hybrid imagers for the 2nd screening (suspect object detection system) to determine if a suspect is carrying a dangerous object and what dangerous objects he/she has, as shown in Figure 7. In this system integration test, security personnel in Area B direct suspects identified by the 1st screening to the scanning area of the hybrid imager for a 2nd screening. If it is determined that the suspect is not carrying a dangerous object, he/she will be released.

4. Performance Evaluation

4.1. Two-Stage Screening

4.1.1. Suspect Recognition System (1st Screening). The experimental setup for the functional test of the suspect recognition system (1st screening) is shown hereafter. If the millimeter-wave radar determines that the suspect may have some suspicious objects, the millimeter-wave radar informs the detection result to suspect person detection system. The system associates the suspect’s camera image with the suspect’s information. It then sends an alert combining the information to a nearby security personnel for a 2nd detection for the suspect. As shown in Figure 10, for the 1st screening, we set up a millimeter-wave radar and camera 15 meters away from the person and let the suspect carrying a handgun approach slowly to test the operation of suspect recognition system. The specific experimental scenario of preliminary test for 1st screening is shown in Figure 11, which was conducted in a radio anechoic chamber to check

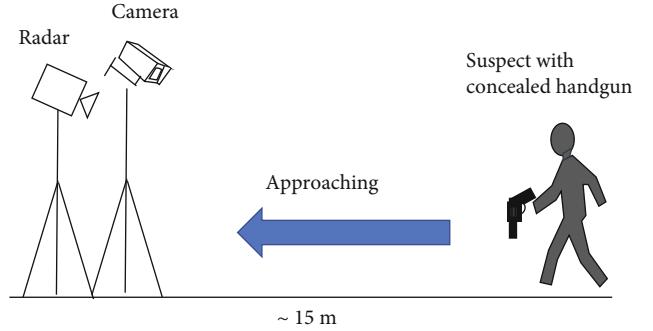


FIGURE 10: Experimental principle of preliminary test for the 1st screening.

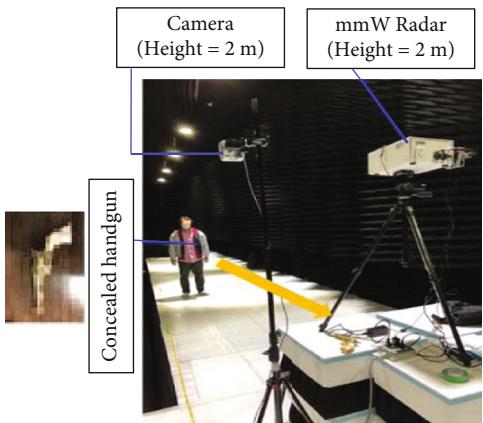


FIGURE 11: Experimental scenario of the preliminary test for the 1st screening.

TABLE 2: Experimental parameters of the preliminary test for the first screening.

Item	Values
Environment	Anechoic chamber
mmW radar	
Frequency	78 GHz band
Modulation	FM-CW
Bandwidth	~4 GHz
Ant. Height	2 m
Illumination area	~15 m

the basic operation of the system in a controlled environment (i.e., no multipath environment). Due to the limited area inside the radio anechoic chamber, only a 1st screening area was set up. The specific experimental parameters are illustrated in Table 2, where we use a 1D radar device in the 78 GHz band and Frequency-Modulated Continuous Wave (FMCW) modulation. Figure 12 shows a screenshot of the suspect recognition system. We have examined the trajectories of the suspects detected by the millimeter wave radar, the camera images of the suspects, and the alerts sent to the security personnel through the system.

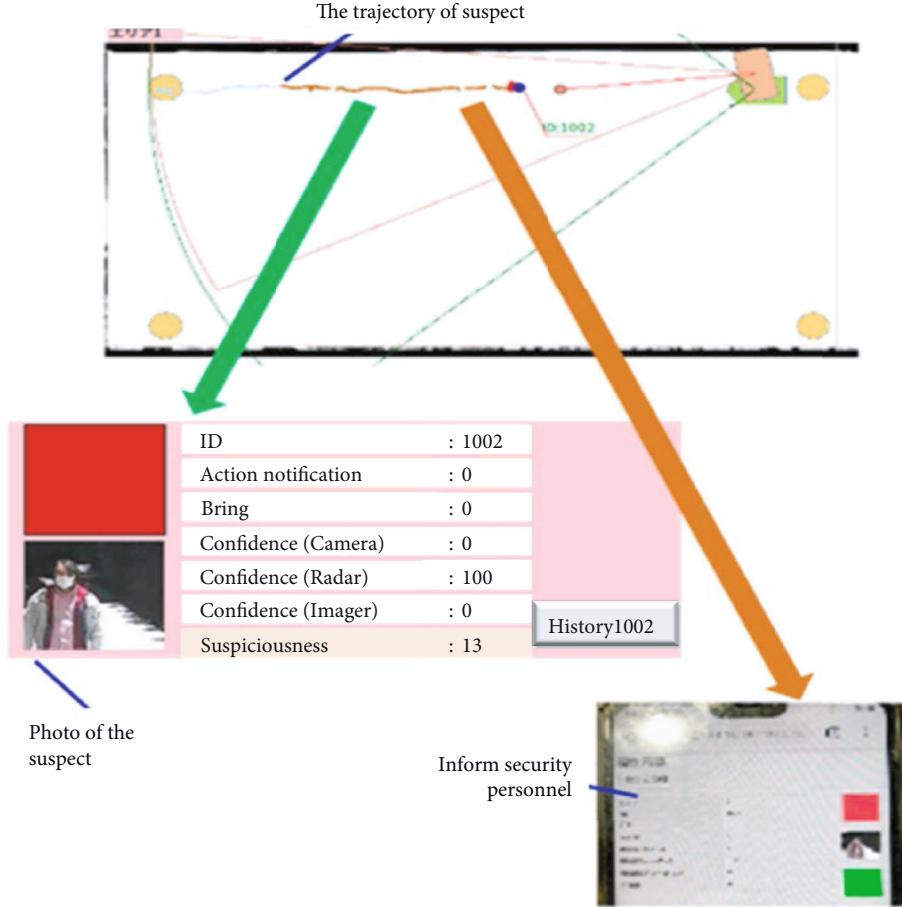


FIGURE 12: System screenshot of the 1st screening.

4.1.2. Suspicious Object Detection System (2nd Screening). The suspicious object detection system identifies suspicious objects by capturing people walking in that environment and feeding the acquired hybrid imager data into an artificial intelligence-based neural network. We selected results for four objects (including knives, guns, smartphones, and nonpossessed objects) and used recall and precision to estimate the overall performance of the two-stage screening method. The usability and safety of the system are demonstrated by recognition rate or both recall and precision rate. In the results, there can lead to 4 different poles: true positive, true negative, false positive, and false negative. The “recall rate,” “precision rate,” and “accuracy rate” are defined as follows.

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (2)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (3)$$

where TP and TN stand for “true positive” and “true negative” and FP and FN stand for “false positive” and “false negative.” For the 2nd screening, we used actual walking

samples to evaluate the performance. As shown in Figure 13, we used a multiple frames of hybrid images for people walking at distances ranging from 2.8 to 1.6 meters. Table 3 [35] lists the recall values of 72.2%, 77.3%, 50.0%, and 85.7% for knives, guns, smartphones, and nonconcealed items, respectively. In addition, the precision values for knives, firearms, smartphones, and nonpossession were 76.5%, 81.0%, 66.7%, and 100.0%, respectively. The results show that these values vary over a wide range of 50% to 100%, but all values are above 50%. They perform reasonably well as a first step.

4.1.3. Performance Evaluation. This suspicious object detection system uses two-stage walkthrough screening, with the 1st screening using millimeter-wave radar to determine whether the suspect is carrying suspicious objects and the 2nd screening using millimeter-wave hybrid imager to distinguish between several types of suspicious objects (knife, gun, smartphone, and nonpossession). In this evaluation, the system performance was evaluated primarily by recall and precision. The recall represents the percentage of correct responses and precision is the percentage of correct predictions. Here, we define the recall and precision for the 1st screening as R_1 and P_1 , respectively; the recall and precision for the 2nd screening as R_2 and P_2 , respectively; and the

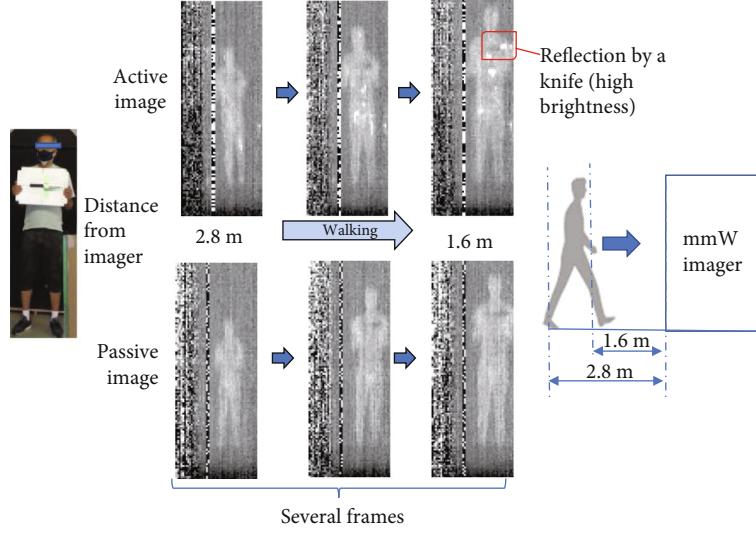


FIGURE 13: Experimental principle of preliminary test for the 2nd screening.

TABLE 3: Recall and precision for the 2nd screening.

Objects	Recall	Precision
Knives	72.2%	76.5%
Guns	77.3%	81.0%
Smartphones	50.0%	66.7%
Nonprocessed objects	85.7%	100.0%

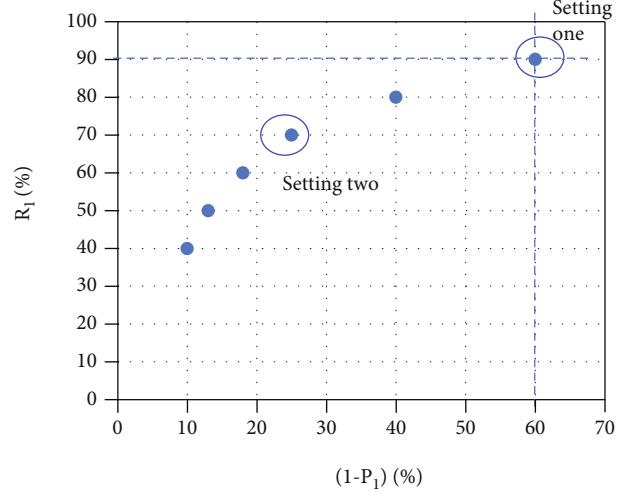
overall recall and precision for both screenings as RT and PT , and are calculated as follows:

$$RT = R1 * R2, \quad (4)$$

$$PT = 1 - (1 - P1) * (1 - P2). \quad (5)$$

According to [36], a receiver operational characteristic (ROC) curve about $R1$ and $(1-P1)$ can be obtained (Figure 14). $R1$ and $P1$ can be adjusted by changing the power threshold or the reflected area of the radar device. First, we set $R1$ to a higher value of 90%, and with Figure 14, we can obtain $P1$ to a lower value of 40%. A higher value of recall can reduce suspicious object misses, while a lower value of precision can lead to suspicious object false positives. Through the experiment, we get $R2$ and $P2$ for suspicious objects (knife, gun, smartphone, and nonpossession) in the 2nd screening, and according to Equations (4) and (5), we can calculate RT and PT . The details are shown in Tables 4 and 5, and the overall recall and precision are 64.2% and 88.6%, respectively [37]. As can be seen, the recall rate by the two-stage screening in this way changed from 90% in the 1st screening to the final 64.2%, and the precision rate increased from 40% in the 1st screening to 88.6%.

Next, we set $R1$ to 70%, and according to Figure 14, we can get $P1$ as 75%. Through the experiment, we get the $R2$ and $P2$ of suspicious objects (knife, gun, smartphone, and nonpossession) in the 2nd screening, and according to Equations (4) and (5), we can calculate the RT and PT .

FIGURE 14: ROC curve (relationship between $R1$ and $(1-P1)$).

The details are shown in Tables 6 and 7, and the overall recall and precision are 49.9% and 95.3%, respectively [37]. It is seen that the recall rate by this method of two-stage screening changed from 70% in the 1st screening to the final 49.9%, and the precision rate increased from 75% in the 1st screening to 95.3%. We can see that the two-stage screening method has the characteristic of maintaining a higher precision rate and the precision rate is a very important indicator for this system to increase the prediction efficiency.

The suspect recognition system for 1st screening outputs data to the 2nd screening in real time at 1 data/second. The suspicious object recognition for 2nd screening can receive the data from the hybrid imager at a maximum of 10 fps, and a maximum of 20 frames for the same person are stored in a packet. The packet is recognized together after passing through the hybrid imager. The average processing time of suspicious object recognition for one frame is measured as 0.054565 seconds/frame; the recognition result is produced in 1.1 seconds after the person passes.

TABLE 4: An example of recall for two-stage screening (R1 = 90%).

Objects	1st screening		2nd screening		Total recall RT
	Detection result	Recall R1	Classification result	Recall R2	
Knives			Knives	72.2%	65.0%
Guns	Metal detection	90%	Guns	77.3%	69.6%
Smartphones			Smartphones	50%	45.0%
Nonpossession	No metal	90%	Nonpossession	85.7%	77.1%
Average		90%		71.3%	64.2%

TABLE 5: An example of precision for two-stage screening (P1 = 40%).

Objects	1st screening		2nd screening		Total PT
	Detection result	P1	Classification result	P2	
Knives			Knives	76.5%	85.9%
Guns	Metal detection	40%	Guns	81.0%	88.6%
Smartphones			Smartphones	66.7%	80.0%
Nonpossession	No metal	40%	Nonpossession	100.0%	100.0%
Average		40%		81.1%	88.6%

TABLE 6: An example of recall for two-stage screening (R1 = 70%).

Objects	1st screening		2nd screening		Total recall RT
	Detection result	Recall R1	Classification result	Recall R2	
Knives			Knives	72.2%	50.5%
Guns	Metal detection	70%	Guns	77.3%	54.1%
Smartphones			Smartphones	50%	35.0%
Nonpossession	No metal	70%	Nonpossession	85.7%	60.0%
Average		70%		71.3%	49.9%

TABLE 7: An example of precision for two-stage screening (P1 = 75%).

Objects	1st screening		2nd screening		Total PT
	Detection result	P1	Classification result	P2	
Knives			Knives	76.5%	94.1%
Guns	Metal detection	75%	Guns	81.0%	95.3%
Smartphones			Smartphones	66.7%	91.7%
Nonpossession	No metal	75%	Nonpossession	100.0%	100.0%
Average		75%		81.1%	95.3%

Based on the above method, we performed the same experiment by setting the recall of the 1st screening to 40%, 50%, 60%, and 80% to obtain the overall recall and precision (Figure 15). This figure shows the relationship between the recall of the 1st screening and the total system performance (overall recall and precision). We can control the overall performance of the entire system by adjusting the parameter of recall for the first screening. The higher recall represents higher security and fewer missed tests. Higher precision indicates higher availability and fewer false positives. By understanding these characteristics, the system can be designed appropriately based on operational strategies such as emphasizing security or availability.

4.2. Recognition Optimization by GAN. In this paper, we use artificial intelligence techniques to help suspicious object detection system to improve the recognition rate, the key of which is the construction of a suspicious object database. It can build the database by capturing images of dangerous objects generated by active/passive imagers, but it is impractical to rely entirely on this approach because it requires a huge number of images to help train the neural network. Generating images by generative adversarial networks (GAN) is a new idea to build a suspicious objects database. GAN is a deep learning model that is one of the most promising methods for unsupervised learning of complex distributions in recent years. Therefore, we try to generate some

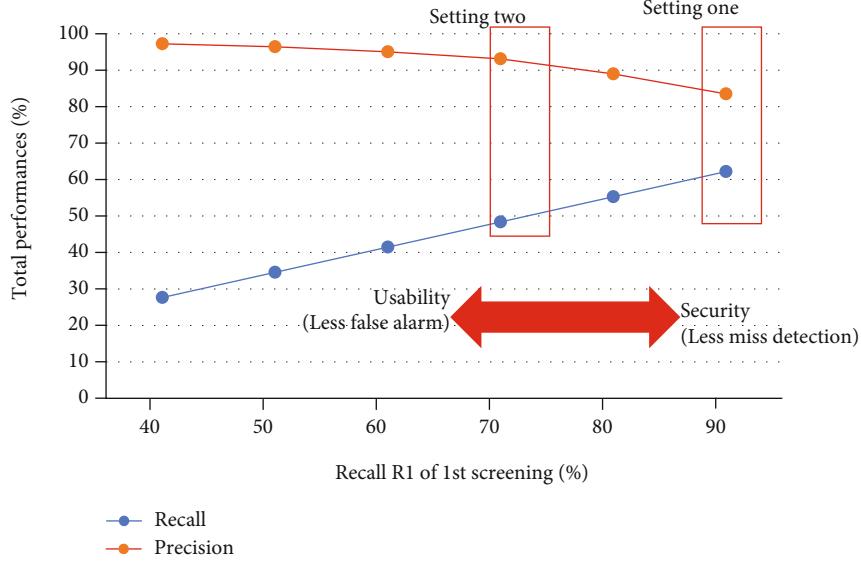


FIGURE 15: Relationship between the recall of the 1st screening and the total system performance.

images by GAN to complement the suspicious object database and evaluate its usability. The whole evaluation process was carried out in a CNN environment with two major categories: (1) comparing the differences in the final recognition rates by training the CNN with GAN-generated images and the original images and (2) identifying the factors that affect the recognition rates if the CNN is trained with GAN-generated images. Throughout the experiment, we used four suspicious items such as gun, knife, scissors, and other items to train and evaluate the CNN, and the network parameters of each layer of the CNN are shown in Table 8.

4.2.1. Comparing the Differences in the Final Recognition Rates by Training the CNN with GAN-Generated Images and the Original Images. To evaluate the effect of GAN on the recognition rate of the system, we conducted two sets of experiments. In the first set of experiments, we trained the CNN with the GAN generated images and the original images, respectively, and compared the final recognition accuracy. For the GAN-generated images, we used the real experimental images obtained by the active/passive imager which included 572 knives, 118 guns, 43 scissors, and 139 other items (66%, 13%, 5%, and 16%, respectively) for GAN training and generated 1000 knives, 1000 guns, 1000 scissors, and 1000 other items for CNN training. Here, we choose epoch 1000 and batch_size 32. For the original images, we use the same proportion of images obtained by the active/passive imager which includes 572 knives, 118 guns, 43 scissors and 139 other items (66%, 13%, 5%, and 16%, respectively) for the CNN training directly. In the CNN evaluation, we used the same proportions of 285 knives, 58 guns, 21 scissors, and 69 other items (66%, 13%, 5%, and 16%, respectively) as in the GAN training. It is worth emphasizing that the real experimental images used for CNN training and CNN evaluation are completely differ-

TABLE 8: Network parameters.

Layer (type)	Output shape	Param #
conv2d_18 (Conv2D)	(None, 32, 32, 64)	640
max_pooling2d_17 (MaxPooling)	(None, 16, 16, 64)	0
batch_normalization_14 (BatchNorm)	(None, 16, 16, 64)	256
flatten_17 (Flatten)	(None, 16384)	0
dense_33 (Dense)	(None, 256)	4194560
dense_34 (Dense)	(None, 4)	1028

Total params: 4196484. Trainable params: 4196356. Nontrainable params: 128.

ent. The experimental results are shown in Figure 16. When the GAN images are used for CNN training, the recognition accuracy remains in a stable state in the range of 0.8-0.9; when the original images are used for CNN training, the recognition accuracy gradually increases with the increase of epoch number and finally remains in a stable state in the range of 0.9-1.0. When the number of CNN epochs is small, the accuracy of CNN training with GAN images is higher; as the number of epochs increases, the accuracy of using original images will exceed that of using GAN images.

In the second set of experiments, we also trained the CNN with the GAN-generated images and the original images and compared the final recognition accuracy. For the first GAN-generated images, we used the real experimental images obtained by the active/passive imager which included 43 knives, 43 guns, 43 scissors, and 43 other items (with proportions of 25%, 25%, 25%, and 25%, respectively) for GAN training and generated 1000 knives, 1000 guns, 1000 scissors, and 1000 other items for CNN training. Here, we choose epoch of 1000 and batch_size of 32. For the original images, we use the same scale of real experimental images obtained by the active/passive imager which includes

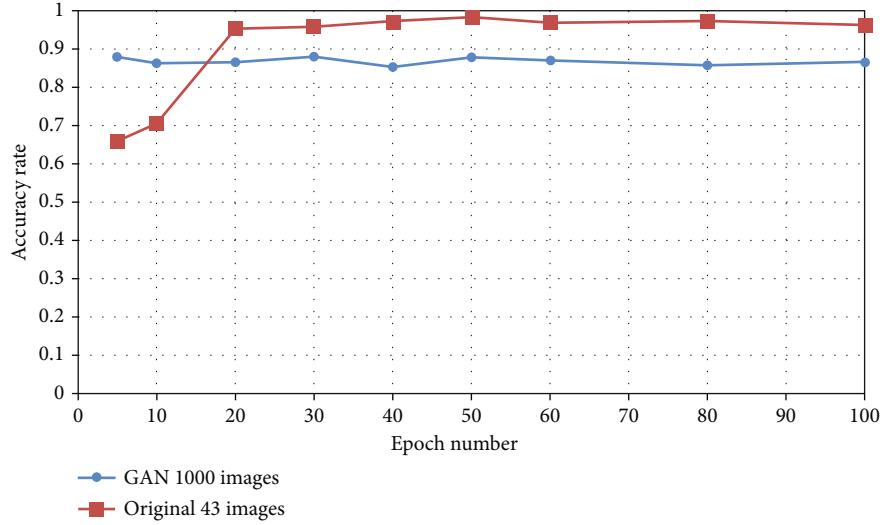


FIGURE 16: Differences in the final recognition rates by training the CNN with GAN-generated images and the original images (43 knives, 43 guns, 43 scissors, and 43 other items).

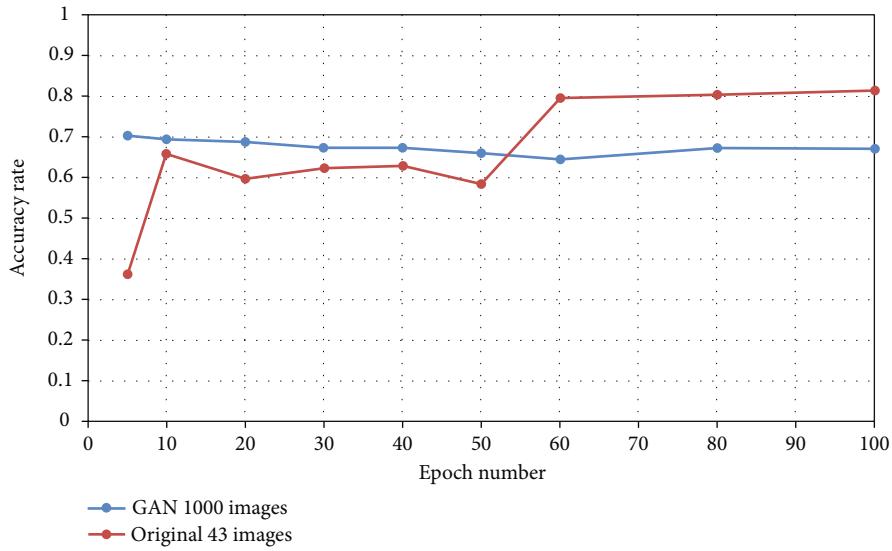


FIGURE 17: Differences in the final recognition rates by training the CNN with GAN-generated images and the original images (572 knives, 118 guns, 43 scissors, and 139 other items).

43 knives, 43 guns, 43 scissors, and 43 other items (25%, 25%, 25%, 25%, and 25%, respectively) for CNN training directly. In the CNN evaluation, we used the same proportions of 285 knives, 58 guns, 21 scissors, and 69 other items (66%, 13%, 5%, and 16%, respectively) as in the first set of experiments. It is worth emphasizing that the real experimental images used for CNN training and CNN evaluation are completely different. The experimental results are shown in Figure 17. When the GAN images are used for CNN training, the recognition accuracy remains in a stable state in the range of 0.6-0.7; when the original images are used for CNN training, the recognition accuracy gradually increases with the increase of epoch number and finally remains in a stable state in the range of 0.8-0.9. When the

number of CNN epochs is small, the accuracy of CNN training with GAN images is higher; as the number of epochs increases, the accuracy of using original images will exceed that of using GAN images. Comparing Figures 16 and 17, it can be concluded that the more original images used for CNN training, the higher the accuracy of the original images will be and surpass the accuracy of the GAN images faster.

4.2.2. Identifying the Factors That Affect the Recognition Rates If the CNN Is Trained with GAN-Generated Images. In our proposed system, it is important to improve the object recognition accuracy by using proper materials for training. Compared with other pattern recognition tasks like character recognition, it is difficult to acquire large scale of

TABLE 9: Experiment results when epoch = 5 (572 knives, 118 guns, 43 scissors, and 139 other items (66%, 13%, 5%, and 16%, respectively) for GAN training).

	Gun	Evaluation results		Scissor	Total	Recall	Average recall
Input		Knife	Other				
Knife	3	269	12	1	285	0.94386	
Gun	53	1	4	0	58	0.913793	0.6828397
Scissor	0	19	1	1	21	0.047619	
Other	9	3	57	0	69	0.826087	
Total	65	292	74	2	433		
Precision	0.815385	0.921233	0.77027	0.5			Accuracy rate
Average precision		0.751722					0.8775982

TABLE 10: Experiment results when epoch = 5 (43 knives, 43 guns, 43 scissors, and 43 other items (25%, 25%, 25%, and 5%, respectively) for GAN training).

	Gun	Evaluation results		Scissor	Total	Recall	Average recall
Input		Knife	Other				
Knife	5	193	33	54	285	0.677193	
Gun	42	1	14	1	58	0.724138	0.6867717
Scissor	1	5	5	10	21	0.47619	
Other	5	4	60	0	69	0.869565	
Total	53	203	112	65	433		
Precision	0.792453	0.950739	0.535714	0.153846			Accuracy rate
Average precision		0.608188					0.704388

training and evaluation data in the field, operated and selected by human. We choose GAN to generate training data. To evaluate the factors affecting the recognition rate in GAN-generated images for CNN training, we focused on the effect of the number and proportion of original images used for GAN training on the recognition accuracy. First, we used the original images obtained from the active/passive imager, which included 572 knives, 118 guns, 43 scissors, and 139 other items (66%, 13%, 5%, and 16%, respectively) for GAN training and generated 1000 knives, 1000 guns, 1000 scissors, and 1000 other items for CNN training, respectively. Here we choose epoch of 1000 and batch_size of 32. In the CNN evaluation, we use the same proportion of 285 knives, 58 guns, 21 scissors, and 69 other items (66%, 13%, 5%, and 16%, respectively) as in the GAN training. It is worth emphasizing that the real experimental images used for CNN training and CNN evaluation are completely different. Table 9 shows recall, precision, and precision for the evaluation results. The average recall for CNN evaluation at epoch 5 is 0.6828397, with 0.94386, 0.913793, 0.047619, and 0.826087 for knife, gun, scissors, and other items, respectively, followed by an average recognition accuracy of 0.8775982. Although the proportion of guns and other items in GAN training is not high (13% and 16%, respectively), they obtain better recall (0.913793 and 0.826087, respectively). It is worth noting that the obtained scissor recall is too low (only 0.047619), and in

the next step, we need to verify if it is due to the low percentage of scissors in the GAN training (only 5%). Also, we want to test whether the number of original images used for GAN training affects the recognition rate.

For the above hypothesis, we generate 1000 knives, 1000 guns, 1000 scissors, and 1000 other items for CNN training using original images obtained with active/passive imagers including 43 knives, 43 guns, 43 scissors, and 43 other items (25%, 25%, 25%, 25%, and 25%, respectively) for GAN training. Here, we choose epoch as the number of items. Here, we choose epoch of 1000 and batch_size of 32. In the CNN evaluation, we use the same proportion of 285 knives, 58 guns, 21 scissors, and 69 other items (66%, 13%, 5%, and 16%, respectively). It is worth emphasizing that the real experimental images used for CNN training and CNN evaluation are completely different. As shown in Table 10, the average recall for CNN evaluation at epoch 5 is 0.6867717, with 0.677193, 0.724138, 0.47619, and 0.869565 for knife, gun, scissors, and other items, respectively, followed by an average recognition accuracy of 0.704388. It can be seen that as the number of original images used for GAN training decreases (the total number decreases from 433 to 172), the average recognition accuracy of CNN evaluation decreases from 0.8775982 to 0.704388. The recall of scissors in CNN evaluation increases from 0.047619 to 0.47619 as the percentage of scissors in GAN training increases from 5% to 25%.

5. Conclusion

Densely populated public places are high-risk areas for terrorist attacks and are always exposed to various forms of terrorist threats. To cope with this problem, this paper proposes an AI-based W-band suspicious object detection system, which is capable of uninterrupted security check for moving persons in open areas. To achieve this goal, a two-stage screening system is proposed. The 1st screening uses millimeter wave radar to detect suspects who may have concealed threatening objects such as firearms. The radar receives signals returned by the suspects and detects whether they have threatening objects. At the same time, visible light cameras track suspects who are detected as possibly having certain dangerous objects. A 2nd screening uses a hybrid imager with active and passive imaging capabilities to investigate the suspect in detail and identify what dangerous objects he/she has. Also, the system applies artificial intelligence techniques to improve the recognition rate of suspicious objects by this solution. Finally, this paper experimentally verifies the usability and security of the system, which has affected the efficiency of identifying suspicious objects.

Data Availability

Data available on request.

Disclosure

This article was presented in part at the 2020 ITU Kaleidoscope: Industry-Driven Digital Transformation (ITU K), Online, December 2020 [22].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by a research grant for expanding radio wave resources (JPJ000254) of the Ministry of Internal Affairs and Communications under the contract for "Research and development of radar fundamental technology for advanced recognition of moving objects for security enhancement." The authors would like to thank Dr. Yonemoto and his team from Electronic Navigation Research Institute for their support during the experiment conducted at their radio anechoic chamber.

References

- [1] Y. Liu, Z. Wang, and G. Zhang, "Application of multi-domain stereo prevention and control technology in counter-terrorism," in *2019 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pp. 230–230, Shenzhen, China, 2019.
- [2] S. Gupta, R. S. Kartha, E. Ramirez-Asis, J. Flores-Albornoz, R. Vilchez-Vásquez, and S. K. Singh, "A novel traffic surveillance system using an uncalibrated camera," *Journal of Advanced Transportation*, vol. 2022, Article ID 5867524, 10 pages, 2022.
- [3] F. DeLucia, A. Samuels, R. Harmon et al., "Laser-induced breakdown spectroscopy (libs): a promising versatile chemical sensor technology for hazardous material detection," *IEEE Sensors Journal*, vol. 5, no. 4, pp. 681–689, 2005.
- [4] C. Nelson, V. Chaudhary, J. Edman, and P. Kantor, "Walk-through metal detectors for stadium security," in *2016 IEEE Symposium on Technologies for Homeland Security (HST)*, pp. 1–6, Waltham, MA, USA, 2016.
- [5] M. Schieber, A. Zuck, H. Gilboa, and G. Zentai, "Reviewing polycrystalline mercuric iodide x-ray detectors," *IEEE Transactions on Nuclear Science*, vol. 53, no. 4, pp. 2385–2391, 2006.
- [6] D. Mery, D. Saavedra, and M. Prasad, "X-ray baggage inspection with computer vision: a survey," *IEEE Access*, vol. 8, pp. 145620–145633, 2020.
- [7] H. Yin, "Public security video image detection system construction platform in cloud computing environment," *Computational Intelligence and Neuroscience*, vol. 2022, 13 pages, 2022.
- [8] L. Kong and R. Dai, "Object-detection-based video compression for wireless surveillance systems," *IEEE Multimedia*, vol. 24, no. 2, pp. 76–85, 2017.
- [9] D. Sheen, D. McMakin, and T. Hall, "Three-dimensional millimeter-wave imaging for concealed weapon detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 9, pp. 1581–1592, 2001.
- [10] G. Zhao, S. Li, B. Ren, Q. Qiu, and H. Sun, "Cylindrical three-dimensional millimeter-wave imaging via compressive sensing," *International Journal of Antennas and Propagation*, vol. 2015, Article ID 218751, 6 pages, 2015.
- [11] T. Hong, J. Yao, C. Liu, and F. Qi, "Mmwave measurement of rf reflectors for 5g green communications," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 8217839, 10 pages, 2018.
- [12] C. Chen, J. Jiang, R. Fu, L. Chen, C. Li, and S. Wan, "An intelligent caching strategy considering time-space characteristics in vehicular named data networks," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2021, <https://ieeexplore.ieee.org/document/9658313>.
- [13] S. Wan, X. Xu, T. Wang, and Z. Gu, "An intelligent video analysis method for abnormal event detection in intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4487–4495, 2021.
- [14] D. Wilson, "Use of modeling and simulation to support airport security," *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 8, pp. 3–6, 2005.
- [15] T. S. Rappaport, S. Sun, R. Mayzus et al., "Millimeter wave mobile communications for 5g cellular: it will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.
- [16] Y. Meng, C. Lin, J. Zang, and A. Qing, "Mmw holographic imaging for security check," in *2019 IEEE Asia-Pacific Microwave Conference (APMC)*, pp. 569–571, Singapore, Singapore, 2019.
- [17] S. Narejo, B. Pandey, C. Rodriguez, M. R. Anjum et al., "Weapon detection using yolo v3 for smart surveillance system," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9975700, 9 pages, 2021.
- [18] H. Jain, A. Vikram, A. K. Mohana, and A. Jain, "Weapon detection using artificial intelligence and deep learning for security applications," in *2020 International Conference on*

- Electronics and Sustainable Communication Systems (ICESC)*, pp. 193–198, Coimbatore, India, 2020.
- [19] S. H. Myint, Y. Katsuyama, T. Sato et al., “Deep learning based concealed object recognition in active millimeter wave imaging,” in *2021 IEEE Asia-Pacific Microwave Conference (APMC)*, pp. 434–436, Brisbane, Australia, 2021.
- [20] H. Zhang, T. Xu, H. Li et al., “Stackgan: text to photo-realistic image synthesis with stacked generative adversarial networks,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 5908–5916, Venice, Italy, 2017.
- [21] C. Yuan, C. Sun, X. Tang, and R. Liu, “FLGC-fusion GAN: an enhanced fusion GAN model by importing fully learnable group convolution,” *Mathematical Problems in Engineering*, vol. 2020, Article ID 6384831, 13 pages, 2020.
- [22] Y. Katsuyama, K. Yu, S. H. Myint, T. Sato, Z. Wen, and X. Qi, “Ai-based w-band suspicious object detection system for moving persons using gan: solutions, performance evaluation and standardization activities,” in *2020 ITU Kaleidoscope: Industry-Driven Digital Transformation (ITU K)*, pp. 1–7, Ha Noi, Vietnam, 2020.
- [23] J. Yu, H. Chen, K. Wu, Z. Cai, and J. Cui, “A distributed storage system for robust, privacy-preserving surveillance cameras,” in *2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS)*, pp. 1195–1196, Singapore, Singapore, 2020.
- [24] M.-C. Chen and Y.-M. Liu, “An indoor video surveillance system with intelligent fall detection capability,” *Mathematical Problems in Engineering*, vol. 2013, Article ID 839124, 8 pages, 2013.
- [25] F. Ding, G. Zhu, Y. Li, X. Zhang, P. K. Atrey, and S. Lyu, “Anti-forensics for face swapping videos via adversarial training,” *IEEE Transactions on Multimedia*, 2021, <https://ieeexplore.ieee.org/document/9495178>.
- [26] M. Xiao, S. Mumtaz, Y. Huang et al., “Millimeter wave communications for future mobile networks,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1909–1935, 2017.
- [27] L. Yujiri, M. Shoucri, and P. Moffa, “Passive millimeter wave imaging,” *IEEE Microwave Magazine*, vol. 4, no. 3, pp. 39–50, 2003.
- [28] D. Goshi, Y. Liu, K. Mai, L. Bui, and Y. Shih, “Cable imaging with an active w-band millimeter-wave sensor,” in *2010 IEEE MTT-S International Microwave Symposium*, pp. 1620–1623, Anaheim, CA, USA, 2010.
- [29] D. M. Sheen, D. L. McMakin, and T. E. Hall, “Active millimeter-wave and sub-millimeter-wave imaging for security applications,” in *2011 International Conference on Infrared, Millimeter, and Terahertz Waves*, pp. 1–3, Houston, TX, USA, 2011.
- [30] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, 2014.
- [31] R. Girshick, “Fast r-cnn,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, Santiago, Chile, 2015.
- [32] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: unified, real-time object detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, Las Vegas, NV, USA, 2016.
- [33] Z. Cheng and F. Zhang, “Flower end-to-end detection based on yolov 4 using a mobile device,” *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 8870649, 9 pages, 2020.
- [34] J. Feng, Y. Sun, and X. Ji, “High-resolution ISAR imaging based on improved sparse signal recovery algorithm,” *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5541116, 7 pages, 2021.
- [35] Y. Katsuyama, K. Yu, T. Sato et al., “Ai based suspicious object recognition system using w-band hybrid imager in public space walking environment,” in *2021 24th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Okayama, Japan, 2021.
- [36] J. Hausner and N. West, “Radar based concealed threat detector,” in *2008 IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems*, pp. 1–8, Tel-Aviv, Israel, 2008.
- [37] H. Dobashi, “A two-stage walkthrough screening of persons with concealed threat objects using a millimeter-wave radar and an imager,” in *2021 24th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Okayama, Japan, 2021.