

AI-based Video Analysis Solution for Creating Safe and Secure Society

Surveillance cameras are used around the world as a means of investigating and suppressing terrorism and other forms of criminal activity. However, with surveillance cameras having now been installed in numbers that far surpass the ability of people to monitor visually, a need has arisen for the use of video analysis to improve efficiency and make more effective use of this resource. Through collaborative creation with customers, Hitachi is able to pass on what it has learned from its trialing of video analysis in the form of products that are easier for customers to use. This article describes Hitachi's involvement in collaborative creation with customers who operate airports or other critical infrastructure where people congregate. This is accompanied by an overview of Hitachi's products and value-adding initiatives that make active use of the latest IT and research into video analysis and also its plans for the future.

Hideki Okita

Tomokazu Murakami, Ph. D.

Tatsuya Okubo

Toshiaki Tarui

Yasuhiro Fukuda, P.E. Jp

1. Introduction

A study by the United States Department of State has identified explosives as the most common means of terror attacks, accounting for 54% of all cases⁽¹⁾. This means that when airport security detects a bag or other item left unattended at a terminal, the bag is treated as a potential explosive device intended as a terror attack against airport users. Furthermore, if unable to eliminate the hazard within a given time, airport security needs to put the airport terminal into lockdown, evacuating everyone to ensure their safety and incurring significant operational costs in the process⁽²⁾.

The flight delays or cancellations that accompany these airport terminal lockdowns significantly reduce customer satisfaction. They also impose opportunity costs on the airport's retail and other forms of non-aviation revenue. Accordingly, airport security has a need for ways of shortening the length of time terminals spend in lockdown without compromising safety. As a result, recognizing the advances

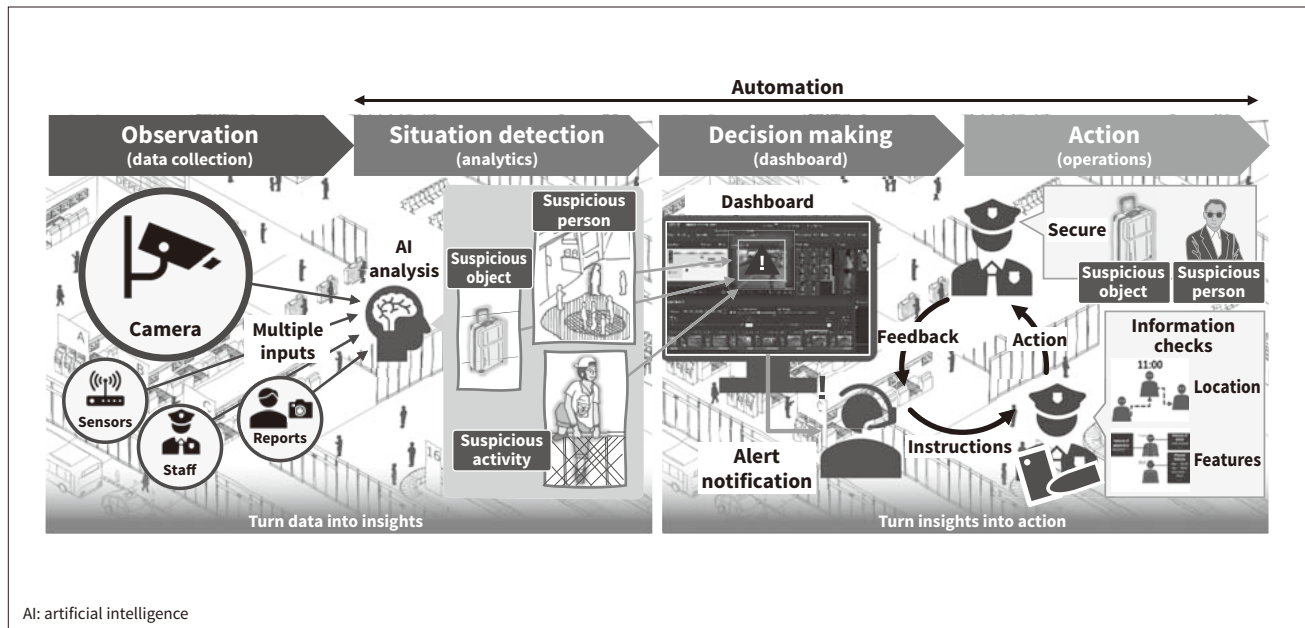
that have been made in video analysis over recent years, many airports have been looking at installing automated systems for detecting unattended baggage through the analysis of surveillance camera video.

The difficulty is that most such incidents are a result of passengers simply forgetting their bags. Accordingly, the systems are unable to do much to reduce the duration of airport terminal lockdowns because airport security is obliged to treat any unattended baggage they detect as a potential explosive until the owner is found.

In response, Hitachi has developed a solution for unattended baggage based on image analysis that can track the movements of a bag owner in real time. This is done using an image search engine that searches through a large number of images at high speed to find those that are similar to a specified image (also called "reverse image search"). By trialing use of the system in the actual security operations of an airport and drawing on the experience of security staff to make improvements, the system is now able to advise security staff on how to quickly find the owner and deal with the unattended baggage.

Figure 1 — Automation of Sequence of Steps from Security Incident Detection to Security Staff Response

Automating the sequence of steps from situation detection to decision making and action reduces the frequency of lockdowns caused by unattended baggage at airport terminals, also shortening the duration of lockdowns when they do occur.



While unattended baggage has long been an issue for airport security, it is anticipated that, by translating decisions made by security staff into practical action on the ground as well as through the use of video analysis to detect security incidents from on-site data, the system will be able both to reduce the number of unattended baggage incidents that result in an airport terminal lockdown and to shorten the duration of lockdowns when they do occur (see Figure 1).

2. Video Analysis Technology that Suits Customer Needs

As noted above, the discovery of a suspicious object can result in part of an airport terminal being shut down if its owner cannot be found. To prevent this, Hitachi has developed a solution that combines technology for detecting unattended items and for tracking people’s movements to help track down the person who left them there.

At events, meanwhile, even when security checks are carried out, considerations of appearance or convenience often make it impractical to cordon off the area on the far side of the gates with high walls. This means it is possible for people from outside to pass items in across the perimeter barriers. To prevent this, there is a need for ways of using surveillance cameras to detect such behavior and track the people involved.

Furthermore, the desire for more reliable security means there are also cases when object as well as people tracking is required in order to determine how large objects are brought in and moved around. For this application, Hitachi has developed a bag tracking technology based on utilizing

the external characteristics of items in image analysis.

Accordingly, while people tracking technology that can reliably track persons of interest will ultimately be the key to security at railway stations, airport, shopping centers, and other public places, also important will be to combine these with other technologies such as for detecting suspicious behavior or tracking the flow of items in order to satisfy specific customer needs.

The following sections respectively describe a behavior detection technology based on posture estimation and a way of tracking large bags.

2.1

Behavior Detection Based on Posture Estimation

This technology works by determining a person’s posture from camera images and can use patterns such as joint positions to identify instances of particular actions. Examples include use in public safety applications to detect when people have fallen or crouched down, and use for security to detect people jumping over or passing items across barriers.

The following explanation uses the example of detecting the act of passing something to another person. The technology detects the locations of arms and legs and other key points on the body of people captured by surveillance camera video and calculates feature values for the distances between points on two people standing on opposite sides of a barrier. Whether or not an item has been passed from one person to another is determined by using a model built beforehand by machine learning to classify these feature values. The features of this technology are that it does not require any additional learning post-installation and that it

Figure 2 — Detection of Item Passed between People

This detects when an item is passed from one person to another by determining their posture from surveillance camera video.



continues to work reliably even when the background scene changes (see **Figure 2**).

2. 2

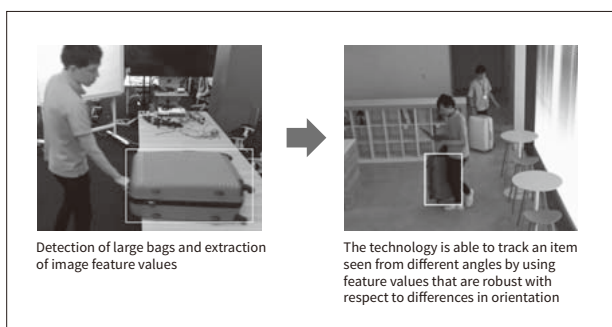
Large Baggage Tracking Technology

This tracking technology provides a way to search for and identify sightings of a particular item across multiple surveillance cameras. It uses a detection model that has been trained using a large number of baggage images to identify suitcases, travel bags, and other kinds of luggage. Next, feature values are calculated using a feature value extraction model (which has also undergone prior training) and stored in a database along with the images and position information. To track the movements of a particular bag of interest, an image showing that bag is specified and a search is performed that compares its feature values with those for all bags that appear in the database images.

By performing the prior training using feature values that are robust with respect to bag orientation, the technology is able to find and track the bag even when it is seen from a different angle than that of the specified image (see **Figure 3**).

Figure 3 — Detection and Tracking of Large Bags

The movements of a bag can be tracked across multiple surveillance cameras by specifying an image of the item of interest.



3. High-speed Person Finder/Tracker Solution and Add-on Function

This section describes the features of Hitachi's high-speed person finder/tracker solution and an add-on function. It also gives examples of their use.

3. 1

Features of High-speed Person Finder/Tracker Solution and Add-on Function

Figure 4 shows an overview of Hitachi's high-speed person finder/tracker solution (left and center) and a multi-query tracking interface available as an add-on function (upper and lower right).

The solution was launched in October 2019, combining the Integrated Physical Security Platform developed by Hitachi Industry & Control Solutions, Ltd. with person detection and analysis plug-ins that can run on the platform. Because the search function does not use facial recognition and is instead based on full body appearance, one of its key features is that it can find people even when the available facial images are of poor quality or not captured at all due to the specifications or locations of surveillance cameras. Search results can be presented as a timeline based on time-stamp and camera information. Even in cases when there is little to go on beyond witness reports, an attribute search function is available that can narrow down "key images" that match specified attributes such as the person's gender or how they are dressed. These key images can then be used as a starting point for searching for or tracking the person of interest.

The multi-query tracking interface available as an add-on provides functions for high-speed people tracking that include displaying search results based on multiple images and automatic updating. Because the airport security application described above requires ways of tracking people in real time, the multi-query interface was developed to augment the above solution by providing an interface that would be easy to use for the urgent tracking of particular individuals in response to a detection alert from an external system. It includes a function (a high-speed people tracking function) that helps determine the time and place of the person of interest's movements by refreshing the tracking screen at regular intervals and automatically switching to cameras that show the most recent potential sightings.

Past video analysis has included technologies that work by use of facial recognition to identify people on a blacklist or by the detection of unattended bags or people entering restricted areas. Hitachi's new solutions can work with the systems that perform this detection, being able to receive an alert from them when a person of interest is detected and then track the location of that person.

3.2

Example Applications of High-speed Person Finder/Tracker Solution

The solution has potential applications for security at airports and other such facilities. Other possibilities include use in criminal investigations, where the solution could show its worth in cases where there are large numbers of images that require visual checking, but only a small team of investigators. That is, Hitachi's solution can significantly reduce workloads and increase productivity by automatically identifying which frames do not show any people and rapidly and automatically narrowing down those images that do show people similar to the person of interest (see Figure 4). The solution also has potential applications for security at shopping centers in situations such as finding lost children or the identification and tracking of suspicious individuals. Provided that the necessary measures for privacy protection are put in place, including disclosing to visitors what is being done, there is also potential for the

solution to help centers with analytics data for marketing or to add more value in other ways by using it to determine the flow of people at the facility or to calculate how long they loiter in particular areas.

4. Technologies Needed for Video Analysis

This section describes the features of the deep learning technology used to perform image analysis by artificial intelligence (AI).

4.1

Overview of High-speed Person Finder/Tracker Solution

Figure 5 shows a diagram of the solution. The solution uses the person detection and analysis plug-in running on the Integrated Physical Security Platform to perform the visible feature extraction and attribute estimation functions used to detect people in camera images and track their movements,

Figure 4 — Overview of High-speed Person Finder/Tracker Solution and Add-on Function

By incorporating the multi-query tracking interface as a solution add-on, real-time tracking of a person of interest can be initiated by alerts from other systems.

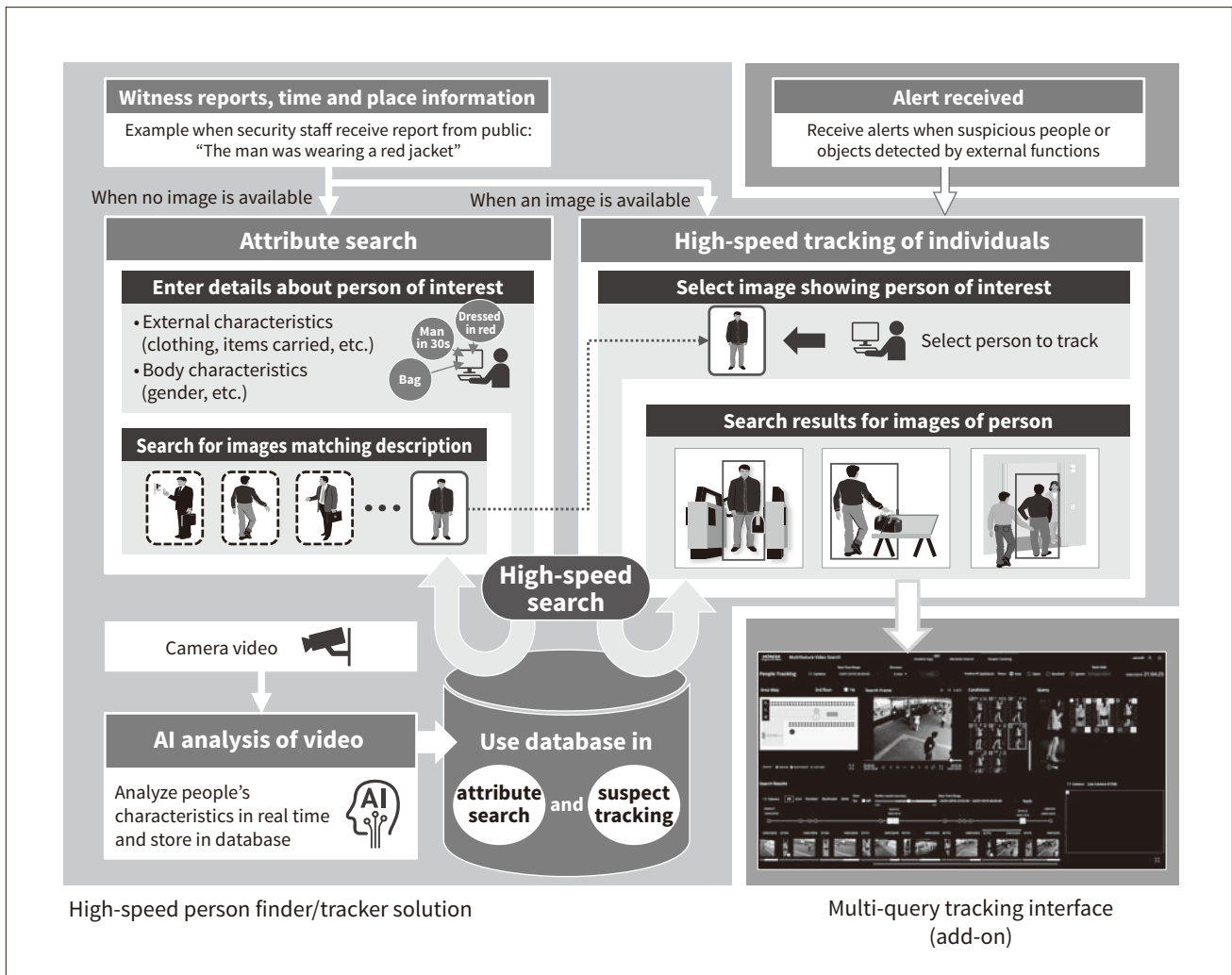
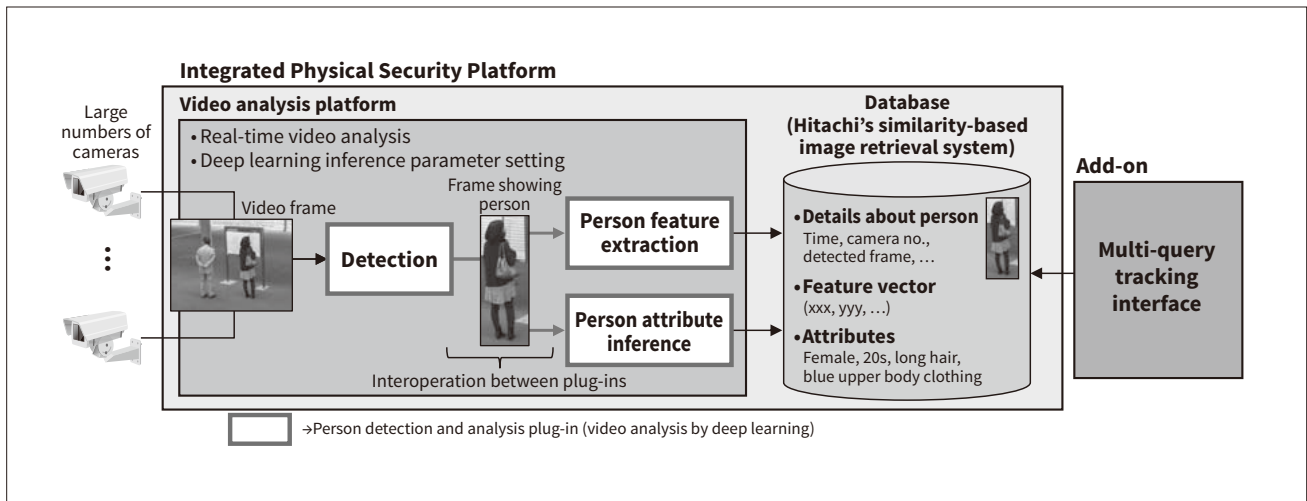


Figure 5 — Architecture of High-speed Person Finder/Tracker Solution

Surveillance video can be analyzed in real time and stored in a database using the person detection and analysis plug-in that runs on the Integrated Physical Security Platform.



storing the results in a database. The analysis and database storage of video from multiple cameras installed over a wide area is performed in real time, with Hitachi's high-speed similarity-based image retrieval system^{*1} being used to provide instant updates on the movements of the person being tracked.

4.2

Technologies Used to Perform Deep Learning in Real Time

The Integrated Physical Security Platform uses Hitachi's original video analysis framework for the real-time analysis of video from large numbers of cameras. The analysis framework has the following features.

(1) Interoperation between video analysis modules

The video analysis modules have a standard interface and run as plug-ins, providing the flexibility to select and combine modules as required. For example, interoperation with baggage tracking can be realized by this feature.

(2) Real-time video analysis

Real-time operation is achieved by switching video from multiple cameras between the video analysis modules.

(3) Deep learning parameter setting

This involves setting execution parameters to tune the efficiency of general-purpose computing on graphics processing units (GPGPU)^{*2} based on the current requirements (number of people detected, required frame rate) and server resources (performance and installation conditions). These settings include the number of deep learning batches and how many video analysis threads to run concurrently on the central processing unit (CPU).

*1 A technology developed by Hitachi that can, in less than a second, identify images similar to a specified image from large data sets containing in the order of a million images.

*2 A technology that uses GPUs for general-purpose applications (other than image processing).

5. Considerations for Designing Deep Learning Runtime Performance

The hardware performance requirements for executing deep learning are determined by running a number of tests. Ruggedized servers able to withstand environmental changes have increasingly been adopted in recent years as edge computing devices for Internet of Things (IoT) sensors. Performance tests were conducted using the NVIDIA Tesla T4^{*3} to identify the important factors to consider when installing such systems in the field [what Hitachi refers to as "operational technology" (OT) applications] or in on site at public facilities, for example. Consumer-model servers were installed in temperature-controlled rooms to test them under the real-world conditions found outside data centers. **Figure 6** shows time-series data for graphics processing unit (GPU) usage, power consumption, temperature, and frequency measured during the execution of a number of real-time video analyses.

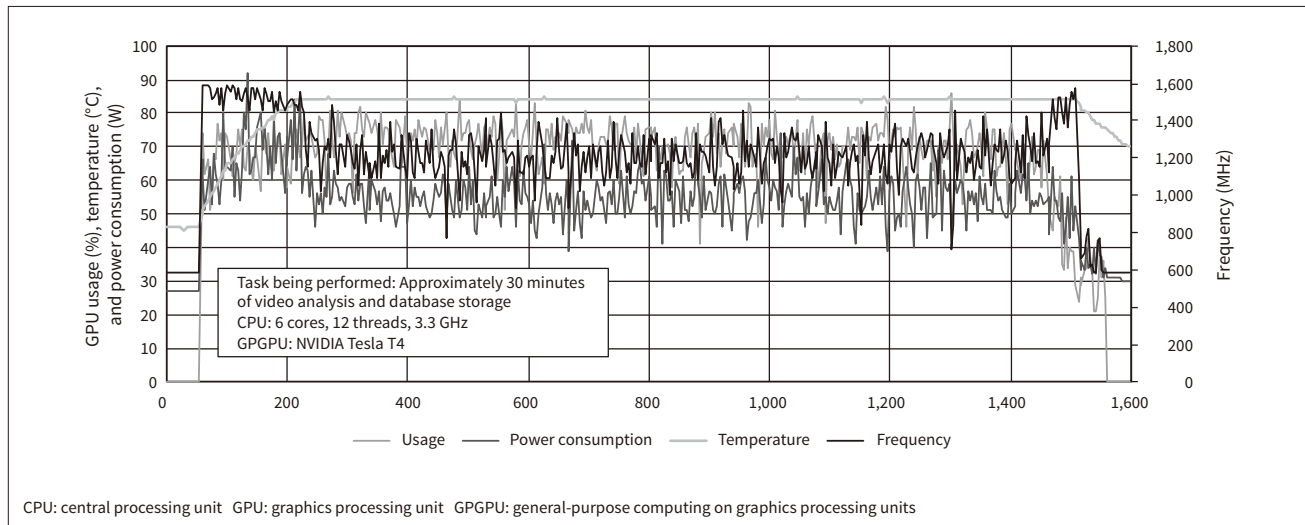
The results show GPU temperature initially increasing, reaching 84°C after 200 s. At this point, the frequency dropped from around 1,500 MHz to the 1,200-MHz range, presumably due to activation of GPU protection mechanisms. The slowing of the GPU clock frequency reduced deep learning performance by about 20%.

Given the likelihood that edge computing will be required in a wide range of IoT sensor applications, it cannot be assumed that the servers used for deep learning will be installed in data centers. Accordingly, safety margins need to be included when performing server performance design for the runtime requirements of deep learning to allow for factors such as the performance degradation

*3 NVIDIA and Tesla are trademarks or registered trademarks of NVIDIA Corporation in the United States and other countries.

Figure 6 — GPGPU Performance During Real-time Video Analysis

The graph plots time-series data for the measured GPU usage, power consumption, and temperature (left axis) and frequency (right axis).



observed in the test that was due to installation conditions. Taking heed of this insight, reliable operation will be ensured by adopting system design techniques and an optimal combination of Hitachi PC servers based on the environmental requirements that customers need to run deep learning.

6. Future Plans

Along with the spread of technologies for analyzing video that are being developed as solutions for dealing with terrorism and other forms of criminal activity at critical social infrastructure in airport security and other such applications, surveillance cameras have come to be recognized for their potential use as IoT sensors. However, extracting useful information from video will in some cases require changes to the location or configuration of these cameras,

with the accompanying problem that it may be difficult to find enough time to make the necessary adjustments given the rising number of cameras being installed. Hitachi is addressing these issues by developing solutions that utilize three-dimensional (3D) digital design simulator platform. This platform can be used to visualize the customer's facility using real-life 3D model and render live simulated video surveillance camera views prior to installation. Along with the benefits of shortening the time taken to select camera types and optimized locations, with adjustment work completed more quickly and with less rework, it also provides the ability to specify the camera field of view and configuration settings to best suit the objective of video analysis (see **Figure 7**).

Meanwhile, the requirements for facility security include the information handled by command centers that monitor the video from surveillance cameras and issue instructions to the field, and this information is becoming more diverse. The analysis of information obtained from surveillance

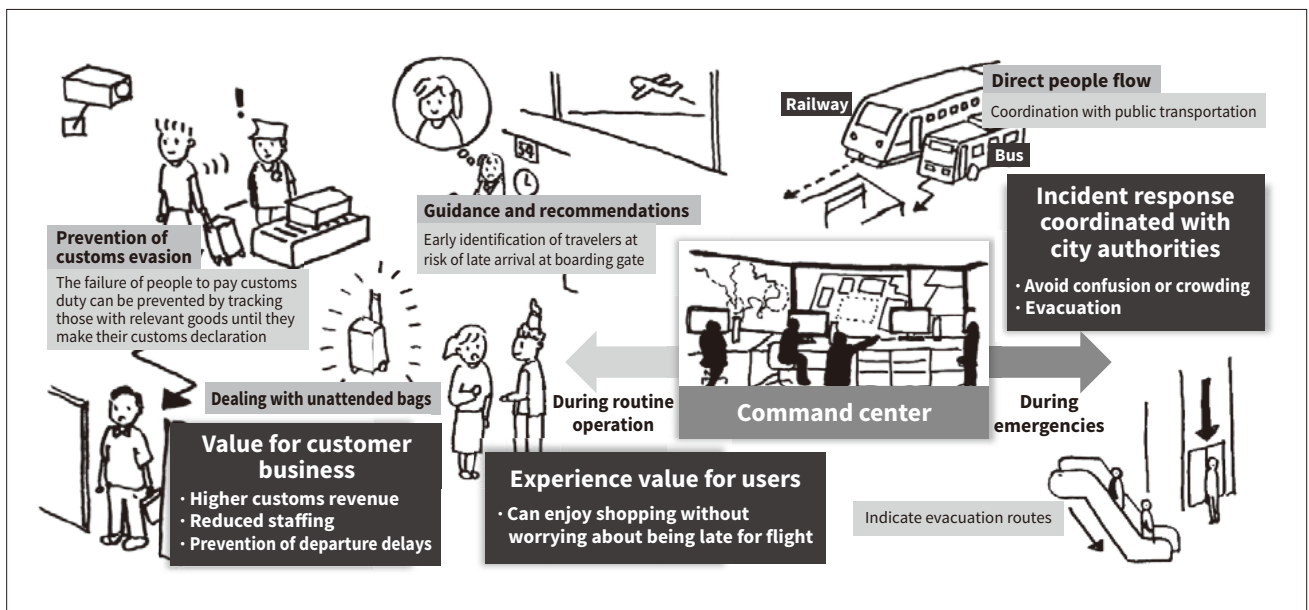
Figure 7 — Service Providing 3D Simulation of Camera Video

The service scans the site with a three-dimensional (3D) camera prior to system installation and performs a simulation replicating the actual surveillance camera video to show things like the field of view provided by the planned camera locations. Undertaking this preliminary check shortens the time taken for camera installation, avoids rework, and enables tuning of the video analysis software to be completed more quickly.



Figure 8 — Command Centers that Add Value and Can Potentially Serve as Social Infrastructure During Emergencies

This example shows how routine work at an airport can be made more efficient and how the flow of people during emergencies can be managed.



video not only has the potential to improve the efficiency of routine activities, but can also contribute to facility operations in ways that add value. With regard to what command centers were originally intended for, it is also anticipated that systems for providing appropriate information to facility users during an emergency or that serve as a platform for integration with external public transportation services and associated parties will also become a public safety requirement in the future (see **Figure 8**).

7. Conclusions

To enable the public to feel safe, reliable video analysis technology will need to expand to cover all corners of society. Moreover, expanding the scope of video analysis utilizing methods like deep learning will also require technologies for building the IT infrastructure for cloud platforms and computers like those at high-performance computing facilities. Hitachi intends to contribute to creating a safe and secure society by developing products that are easier to use and by using proven technologies for the construction of large-scale IT infrastructure.

References

- 1) E. Miller, “Annex of Statistical Information: Country Reports on Terrorism 2016,” National Consortium for the Study of Terrorism and Responses to Terrorism (Jul. 2017).
- 2) O. Nwofia et al., “A Methodology for Designing Airports for Enhanced Security Using Simulation,” *Journal of Aviation Technology and Engineering*, Vol. 3, Iss. 1, pp. 2–8 (Oct. 2013).
- 3) Y. Ikeura et al., “Multimodal Deep Learning Platform for IoT Data,” *Hitachi Review*, 69, pp. 694–698 (Sep. 2020).

Authors



Hideki Okita

Value Creation Project, Global Center for Social Innovation – Tokyo, Research & Development Group, Hitachi, Ltd. *Current work and research:* Research and development of digital smart city solutions. *Society memberships:* A senior member of the Institute of Electronics, Information and Communication Engineers (IEICE) and a member of IEEE.



Tomokazu Murakami, Ph. D.

Intelligent Vision Research Department, Center for Technology Innovation – Artificial Intelligence, Research & Development Group, Hitachi, Ltd. *Current work and research:* Research and development of image recognition and image processing technology. *Society memberships:* IEICE, the Institute of Image Information and Television Engineers (ITE), and the Virtual Reality Society of Japan (VRSJ).



Tatsuya Okubo

Public Safety Department 2, Government & Public Corporation Information System Division, Social Infrastructure Systems Business Unit, Hitachi, Ltd. *Current work and research:* Promoting solution business for public safety businesses.



Toshiaki Tarui

Intelligent Vision Research Department, Center for Technology Innovation – Artificial Intelligence, Research & Development Group, Hitachi, Ltd. Temporarily assigned to Hitachi Industry & Control Solutions, Ltd. *Current work and research:* Application and productization of video analysis technologies.



Yasuhiro Fukuda, P.E. Jp

Public Safety Department 2, Government & Public Corporation Information System Division, Social Infrastructure Systems Business Unit, Hitachi, Ltd. *Current work and research:* Productization of video analysis technologies.