

Video Analytics AI for Public Safety and Security

Use of security cameras and video analytics AI to maintain safety and security in public facilities is growing. Hitachi is already contributing to the safety and security of public spaces through its commercialization of solutions that are able to search for missing children or suspicious persons in real time analyzing large quantities of security camera footage. Against a background that includes the spread of COVID-19 and working style reforms, video analytics AI is being called upon to expand its capabilities for supporting the people whose job it is to maintain safety and security. This article describes the application of AI to anti-infection measures and people flow analysis in public places as well as current work and future plans for video analytics that aims to deliver even better support through a deep understanding of people and the environments they inhabit.

Tomoaki Yoshinaga

Yasuhiro Fukuda, P.E. Jp

Yuki Watanabe, Ph.D.

Atsushi Hiroike, Ph.D.

1. Introduction

The maintenance of high levels of safety and security is essential for achieving ongoing urban growth with a high quality of life (QoL). Security cameras are widely deployed as a means of maintaining public order at railway stations, airports, and shopping malls as well as in city streets and other public spaces. Artificial intelligence (AI) also helps maintain security through its use to rapidly identify unauthorized intrusions or other such incidents from the analysis of large quantities of security camera footage.

Hitachi has been a leader in the development of techniques that use AI to support urban safety and security, including the 2006 release of video surveillance recorders with facial recognition and a high-speed search technique

for finding faces in footage from 100 or more cameras that became available in 2009. With the 2019 release of the High-speed Person Finder/Tracker Solution⁽¹⁾ that features use of deep learning, Hitachi also provides support for large-scale security and safety operations at social infrastructure and public places through a solution that can perform rapid and precise searches for specific individuals in large quantities of security camera footage.

The primary objective of video analytics AI has in recent years been shifting away from threat identification and toward providing support for facility users and for the personnel who operate and manage large facilities with low staffing levels. This article describes the latest solutions and technologies that assist with safe and secure facility operation as well as what Hitachi is doing to further enhance its AI technologies.

2. Solutions that Use Video Analytics AI to Enhance Surveillance and Security

Hitachi has developed new solutions that use security cameras and video analytics AI to support facility operations as well as for the prevention or detection of crime. This section describes a solution for operating facilities while maintaining social distancing amid the COVID-19 pandemic that has spread globally since it first emerged in early 2020, and another for visualizing the flow of people that can be used for evacuation planning and other management decision-making at such sites.

2.1

Infection Prevention Supporting Solution

With government agencies around the world indicating that they will persevere with basic infection control measures such as mask-wearing and social distancing, at least until their COVID-19 vaccination programs are complete, life is not expected to get back to normal for several years. Accordingly, to assist with anti-infection measures, a new version (v1.3^{(2),*1}) of the add-on functions for the High-speed Person Finder/Tracker Solution that Hitachi markets in Japan has been released that features enhanced functionality for detecting instances of people congregating in large numbers.

This works by expressing the locations of people identified in video images as two-dimensional coordinates and analyzing this to provide an indication of social distancing. By only counting people who are closer than a specified distance to one another, rather than everyone who appears in the security camera footage, this function is able to identify instances of crowding and raise an alarm (see **Figure 1**).

Even after the COVID-19 pandemic has passed, and because this same functionality can also be used to detect

*1 Marketed outside Japan by Hitachi Vantara Corporation as Hitachi Multifeature Video Search.

the crowds that congregate when an incident occurs, it is likely that it will continue to prove useful in security applications as a means of quickly identifying when something abnormal has happened.

The new version also incorporates functions for identifying and tracking persons of interest (POI) and vehicles of interest (VOI), featuring an expanded interface that enables use in tandem with thermal imaging cameras for the detection and subsequent tracking of people with elevated body temperatures and a new function that can search for cars, motorbikes, or other vehicles.

While the solution is used for security surveillance at public facilities or in urban neighborhoods where people congregate in large numbers, mainly for finding or tracking missing children or suspicious persons at the associated site or location, it also has wider potential applications. One such example would be its use to generate “origin and destination” (OD) data on where passengers board and disembark public transportation services that operate by cash or ticket. It would also likely prove useful in applications where camera footage is analyzed by AI and utilized as Internet of Things (IoT) data, such as identifying people who change their clothes in a toilet, for example. The solution can also help promote working style reform and accelerate the digital transformation that is the key to resolving societal challenges. This involves reducing workloads at sites where past practice has been to review large amounts of camera footage visually, treating this as IoT data and processing it automatically, and providing the capability to rapidly and automatically narrow down specific people who resemble a person of interest, boosting efficiency by establishing human-in-the-loop practices that incorporate human decision making.

2.2

Wide-Area People Flow Analysis Solution

This section describes a solution for analyzing the flow of people over a wide area using a technique for tracking their movements.

Figure 1 — Function for Detecting Crowding

Crowding is identified by detecting the people visible in video images and determining the distance between them. This will have applications even after the COVID-19 pandemic has passed, such as grasping an occurrence of abnormal situations.

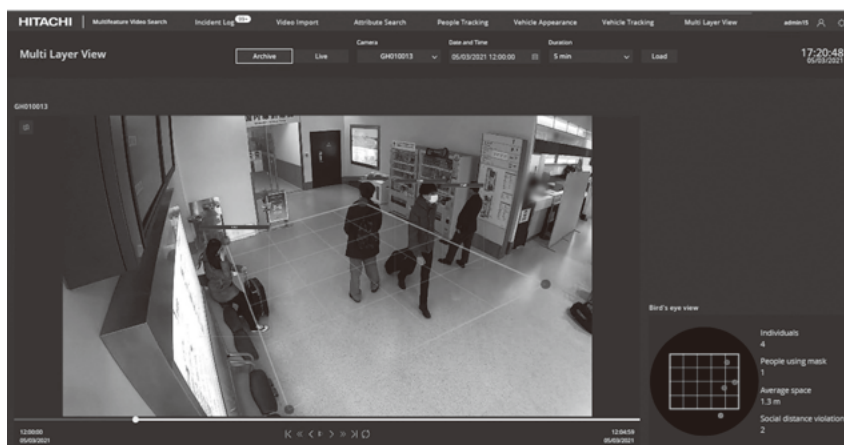




Figure 2 — Example Dashboard from Wide-area People Flow Analysis Solution

The dashboard analyzes the flow of people at a shopping mall and facilitates and supports situation assessment in the event of an accident or other incident.

Figure 2 shows an example dashboard screen of the wide-area people flow analysis.

The dashboard is able to display: (1) People flows superimposed on a wide-area map, (2) A trend graph of crowding levels, (3) A breakdown of where people enter the site, (4) An array of images that show people, (5) The movements of people in a camera field-of-view, and (6) Timelines for specific people. Of these, (2), (4), and (5) visualize the results of detecting and tracking people within a single camera field-of-view, the sort of people flow analysis that is available on most video analytics products. However, Hitachi has developed a technique that tracks people across multiple cameras. Combining this technique into the people flow analysis allows it to analyze the flow of people across wide areas that span numerous cameras. The dashboard can also indicate from which areas the people visible in a particular camera image entered the scene, presenting this superimposed on a wide-area map (1) and in a breakdown chart (3). This provides a way to assess what has caused a crowd to gather. The timeline display (6), meanwhile, provides information about the movements of individuals from one camera to another and can be used to analyze gatherings of people as a group or to assess how long it takes people to move between different areas.

The results of analyzing the flow of people across wide areas can be utilized in big data analytics to provide a comfortable and uncrowded environment at locations such as shopping malls or amusement parks. The intuitive visualization of people's movements is also useful during sudden accidents or other incidents for determining what has happened and responding quickly. Through its wide-area people flow analysis, Hitachi contributes to making public places pleasant, safe, and secure.

3. Next-generation Video Analytics AI that Supports Facility Operations

As described above, video analytics is becoming an essential tool, not only for crime prevention, but also for managing

large public facilities efficiently with a small staff. If the people who work at facilities like these are to deliver more attentive and higher quality services despite low staffing levels, then it will be important for AI to have a deep understanding of what is happening with everyone present at such sites. It is only when AI has such a deep understanding of people that it can provide accurate information to facility staff and users in an optimal manner and take over certain tasks.

This section presents such a next-generation video analytics technique that can search for specific individuals based on an in-depth understanding of how people act and what they carry around with them. It also describes a three-dimensional (3D) simulation technique that was useful in the development of this video analytics AI.

3.1

Search Based on People's Behavior and Baggage

If security and facility management is to become more efficient, it requires a rapid detection capability for criminal acts such as purse-snatching or people entering off-limits areas, and for identifying people in distress, such as someone who has collapsed. Meanwhile, what happens to bags is also important as it is common for people to have a backpack

Figure 3 — Behavior Search

Using a technique able to identify nine different bodily actions likely to occur in public places, the function can search for images that show particular actions. Deep learning can also be used to expand the range of actions able to be identified.

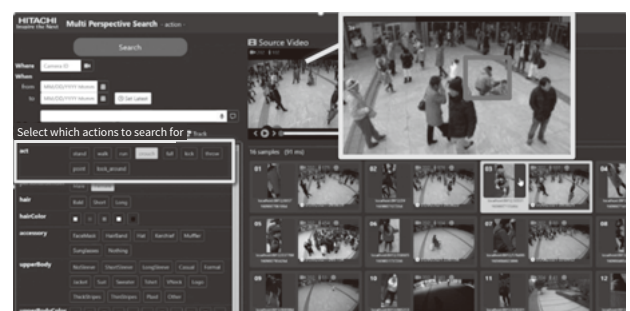


Figure 4 — Baggage Belonging Recognition

The lines linking people and bags indicate the results of inferring which bag belongs to which person. The attribution of ownership is achieved accurately despite the crowded conditions where other people also come in close proximity to the bags.



or suitcase with them when they enter a public facility. Accordingly, Hitachi has developed techniques for searching based on the behavior of people who appear in security camera footage or the bags they carry.

Behavior search works by classifying the actions of people who appear in security camera footage and storing the results in a database. This data can then be searched for instances of people performing a specified action (see **Figure 3**). Nine such actions can currently be identified using a network model that was created using deep learning (walking, running, falling over, standing upright, squatting, kicking, throwing, pointing, and looking around). While performing a person-of-interest search on the basis of clothing or physique alone is problematic, this technique has the potential to make searches more efficient and can also be used for collecting statistics on facility usage (such as the number of people who fall over each day).

Poor accuracy has been a problem with previous techniques used in security camera applications due to how much the shapes of people's skeletal frames vary due to individual differences and depending on where they are and which direction they are facing. Accordingly, Hitachi developed a deep learning model that considers not only skeletal frame shape but also changes in the movements of the joints that make up a person's frame⁽³⁾. As shown in **Figure 3**, the search function can identify specific actions even in scenes where there is considerable variety in people's posture and orientation.

A feature of the baggage search is its ability to search in a way that links a bag to its owner. This technique can detect people and their belongings (suitcases, backpacks, and so on) while also identifying who owns or is carrying which bag (see **Figure 4**). By tracking ownership relationships in this way, the technique is better able to accurately identify events such as bags being stolen or passed from one person to another. In cases where a bag is handled by more

than one person, it can also search for the different people involved. As previous techniques established ownership solely on the relative positions of people and bags, they tended to misidentify ownership in cases where a number of people entered the vicinity of the bag concerned. Along with detecting people and the items they carry, this technique is also able to infer the relation between them, whether in the hand or on the shoulder⁽⁴⁾. Thereby, the new technique has been able to figure ownership relationships highly accurately, reducing more than 30% in misidentifications.

While these techniques are still at the research stage, their accuracy has already been enhanced to the level where they can be used in practice. Because these techniques for identifying people's actions and who owns which bag provide an in-depth understanding of what is happening at a site, they have the potential to provide users with a safe and pleasant experience by enabling services such as the timely identification of abandoned bags and being able to quickly identify and offer support if a person collapses. The technology also has potential uses beyond public places, such as collecting activity logs at factories and other workplaces or facilitating workplace safety by determining whether workers are following safe practices and using tools and machinery correctly.

3.2

Use of 3D Simulation to Assist AI Development

Data is essential to improving the accuracy of these AI techniques with sophisticated capabilities for the recognition of human activity. As the latest deep learning techniques are available as open-source software, many different problems can be solved given the availability of a suitable dataset. Unfortunately, the cost of acquiring such data is increasing. For the human recognition task described above, it is often not possible to obtain the necessary datasets because of problems with privacy and the rights to the video. While the problems can be resolved by recording and annotating video separately, this is very costly and time-consuming.

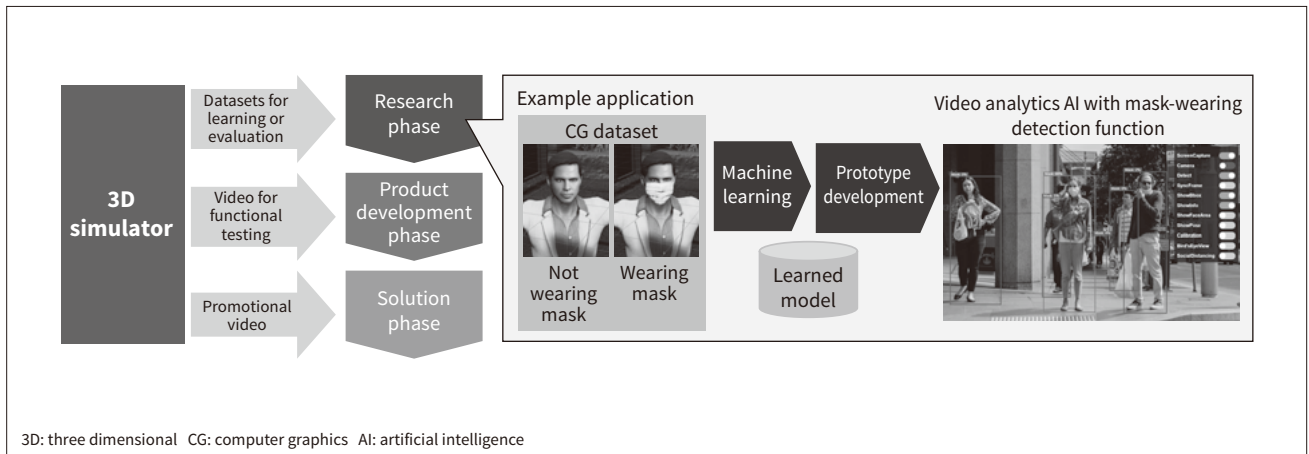
Accordingly, Hitachi has adopted the practice of using 3D simulators in AI development (see **Figure 5**).

A wide variety of different scenes can be replicated by making videos showing people and background objects in a simulator. Moreover, accurate annotation can be achieved automatically by extracting the state of the different objects manipulated by the simulator. Cost can be reduced by a factor of 20 and time by a factor of 16 compared to recording and annotating actual video for this purpose, for example.

The simulator was used in the development of an AI for detecting whether people are wearing masks. The introduction of requirements to wear masks in public places has brought a rapid rise in demand for automatic detection by AI. On the other hand, early in the pandemic, regions where mask-wearing was not a common practice suffered from

Figure 5 — Use of 3D Simulator in Video AI Development

Computer graphics provides useful data for various different phases of development, not only the research and development phase.



a lack of suitable real-world video, while bringing people together to shoot and then annotate such video was likewise problematic. A prototype was able to be developed quickly, however, by instead using datasets generated by a simulator.

While this section has mainly focused on the use of simulators in the learning phase, Hitachi believes that 3D simulators are also effective across many other phases of AI development, including functional testing and marketing promotion. In the future, the company intends to respond to rapidly changing societal needs and deliver solutions in a timely manner by putting simulators to use across the various different phases of video AI development.

4. International Activities Aimed at Enhancing Video Analytics AI

This section looks at some of the work being done by Hitachi to improve its techniques for video analytics AI that support safety and security.

Video analytics AI is evolving rapidly with competition to develop the technology picking up pace around the world. China and the USA in particular are forging ahead with development. Continuing to pursue the solutions and developments described in this article while seeking ways to overcome the challenges facing customers and wider society will require participation in open innovation with customers, universities, and other business partners as well as Hitachi’s ongoing strengthening of its own technical capabilities. This section looks in particular at Hitachi’s achievements in international competitions among these efforts.

Hitachi has since 2015 been a regular entrant in the annual Text Retrieval Conference (TREC) Video Retrieval Evaluation (TRECVID) competition run by the National Institute of Standards and Technology (NIST). Numerous such competitions have been organized in recent years and Kaggle² has made a name for itself as a platform for this

²Kaggle is a registered trademark or trademark of Google LLC.

community, with participation by individuals interested in the technology as well as by universities and companies engaged in their own research and development. Rather than treating them as a form of public relations or a way to enhance its own expertise, Hitachi uses these competitions as part of its own activities motivated by the goal of resolving societal challenges. Accordingly, Hitachi selects tasks that are closely aligned with its own businesses and prioritizes those competitions that, rather than just achieving a ranking, are also about comparing different methodologies through workshops and other such activities. While doing its best to perform well in these competitions and seeking to enhance its skills, Hitachi has worked on developments with a view to their practical application without overly concerning itself with immediate rankings. The technique described in section 3.1 for searching based on people’s behaviors is one example of an outcome that involved participation in a competition to hone Hitachi’s own skills and keep up with technology trends. Hitachi also achieved top-level accuracy in the 2020 Disaster Scene Description and Indexing (DSDI) task involving the application of recognition techniques to aerial imaging⁽⁵⁾ (see **Figure 6**).

Figure 6 — Disaster Situation Assessment from Aerial Photography

The technique can be used for rapid situation assessment during disasters or to assist with visual assessment by humans.

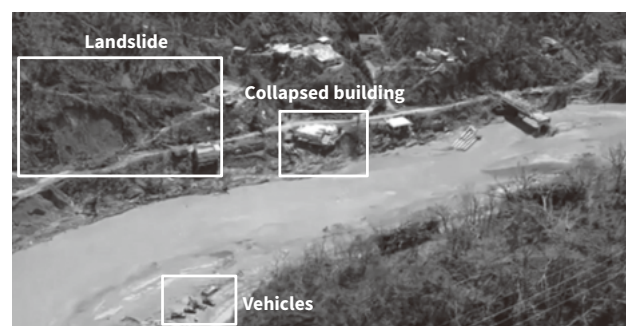
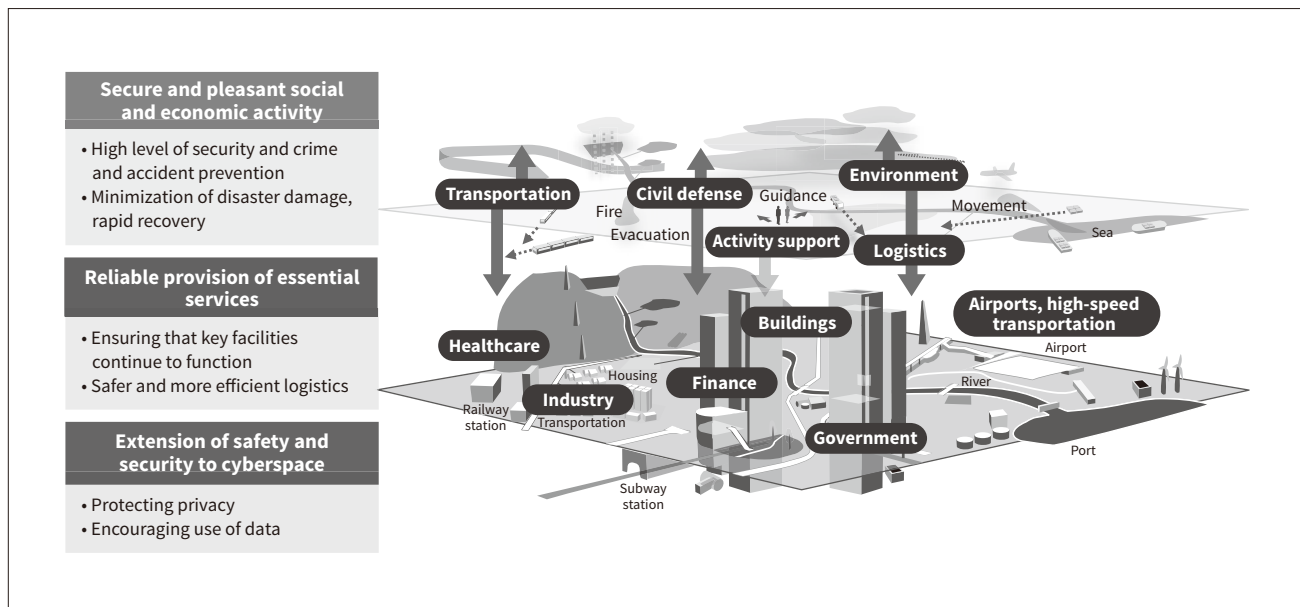


Figure 7 — Hitachi Video Analytics AI for Safety and Security at Organizations that Support Human Life and Social and Economic Activity

The AI helps people in their daily lives and movements and supports the urban management activities of companies through an in-depth understanding of the behavior of people out on the street or in other public places. It also offers extensive support for things like disaster prevention and environmental protection by providing a broad analysis of the conditions under which people go about their lives.



Achieving rapid situation assessment during disasters is a serious societal challenge, and the desire to address this challenge was a major motivation behind this technology.

As well as participating in competitions, Hitachi also organized the MMAAct Challenge as part of the Activity Net workshop held at the Conference on Computer Vision and Pattern Recognition (CVPR) in 2021. More than 20 teams from universities, companies, and other organizations participated in the MMAAct Challenge, where the task they were set involved the detection and classification of actions in simulated security camera video. Taking on the role of competition organizer allowed Hitachi to establish open networks with others working in the same field. In the future, Hitachi hopes to continue sharing diverse ideas and increasing the speed of technology development.

Along with the ongoing development of its own technologies and businesses, Hitachi in its efforts to become a leading global company will also need to push ahead with next-generation technologies that can resolve societal challenges that currently lack solutions. By playing a lead role in the industry as a participant or organizer of global competitions, sometimes competing, other times engaging in open collaboration, Hitachi intends to create the more advanced generation of AI technology.

5. Conclusions

This article has described solutions and technique that use video analytics AI to assist with security and other tasks involved in running a large facility.

Hitachi believes that the use of video analytics AI in place of people to identify small and easily overlooked changes, such as people’s movements or who bags belong to, can lead to the provision of services that are more in tune with people. Considering safety and security in a broader context, there is also a need for situation assessment during disasters and other such incidents. In this sense, expanding the scope of application of video analytics should enable its use to support people’s way of life and the citywide maintenance of safety and security (see **Figure 7**).

When deployed in practice, the AI techniques and solutions described in this article are administered in accordance with Hitachi’s Principles for the Ethical Use of AI in its Social Innovation Business to ensure that they are not detrimental to people’s interests and are used in ways that take care not to invade people’s privacy. Through the widespread and citywide deployment of highly reliable AI solutions that perform accurately and are suitable for use under a wide range of conditions, Hitachi is contributing to the creation of a safe and secure society.

References

- 1) Hitachi News Release, “Commencing Sales of ‘High-Speed Person Finder/Tracker Solution’ that Uses AI Image Analysis Technology to Support More Efficient and Sophisticated Monitoring and Security Operations at Stations, Airports, Commercial Facilities, and Public Institutions” (Oct. 2019) in Japanese, <https://www.hitachi.co.jp/New/cnews/month/2019/10/1004a.html>

- 2) Hitachi News Release, “Enhancing the ‘High-Speed Person Finder/Tracker Solution’ to Support the Sophistication of Surveillance and Security Operations Utilizing AI Image Analysis Technology” (Mar. 2021) in Japanese, <https://www.hitachi.co.jp/New/cnews/month/2021/03/0325.html>
- 3) Y. Ito et al., “Multi-Stream Adaptive Graph Convolutional Network Using Inter- and Intra-Body Graphs for Two-Person Interaction Recognition,” *IEEE Access*, Vol. 9, pp. 110670–110682 (Aug. 2021).
- 4) M. Tamura et al., “QPIC: Query-Based Pairwise Human-Object Interaction Detection with Image-Wide Contextual Information,” *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR) 2021*, pp. 10410–10419 (Mar. 2021).
- 5) Hitachi News Release, “Hitachi Develops Technology for AI-Based Disaster Video Recognition” (Feb. 2021), <https://www.hitachi.com/New/cnews/month/2021/02/210219.html>

Authors



Tomoaki Yoshinaga

Video Analytics Solution Project, Lumada Data Science Laboratory, Research & Development Group, Hitachi, Ltd. *Current work and research:* Research and development of video analytics AI technologies for public safety. *Society memberships:* The Institute of Electronics, Information and Communication Engineers (IEICE).



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 technology.



Yuki Watanabe, Ph.D.

Video Analytics Solution Project, Lumada Data Science Laboratory, Research & Development Group, Hitachi, Ltd. *Current work and research:* Research and development of video analytics AI technologies for public safety.



Atsushi Hiroike, Ph.D.

Video Analytics Solution Project, Lumada Data Science Laboratory, Research & Development Group, Hitachi, Ltd. *Current work and research:* Research and development of video Analytics AI technologies for public safety. *Society memberships:* The Japanese Psychological Association (JPA).