Featured Articles

China Video Solutions: Promoting City Security, Service Quality and Values through Intelligent Visual-perceptions

Peng Yang
Tsutomu Imada
Kai Liu
Daisuke Matsubara

OVERVIEW: Rapid growth of high-quality video devices, video platforms and related services in the smart city environment of China offers huge application and business potential to intelligent video analytics and search. On the other hand, the special aspects of encoded video and visual space (e.g. volume, complexity, diversity) present many obstacles to fully exploring the information and insights in the video content. China Video Solutions, including its video search platform and related platforms, are created to overcome these obstacles with key technologies such as super-high speed video search, intelligent video analytics, strong front-end embedded video platform, light-weight video processing and a flexible video platform. With intensive customer-oriented technical/business practices in China, it shows great potential for making good contributions to the super-secure metropolitan, high-quality city services and value-added applications.

INTRODUCTION

WITH the rapid development of the whole of Chinese society, video-related devices and systems, which can generate multimedia content (especially video content), have increased dramatically. And, the quality of video content generated has been enhanced, which has driven a great increase in the average bit rate and data volume of video content. In the meantime, the informational entropy contained in video content is much higher than other data forms (e.g. text, log). So, more and more services and applications have been built based on video equipment and platforms.

Based on the predictions of CCW in 2015[(1)], this category could support business of over trillion USD in the next five years. As shown in Table 1, the city infrastructure with video-related entities has made great progress. Although huge market and service potential have been witnessed in related fields, this trend imposes great pressure on the capability of cyber systems and technologies. First of all, the resources for encoding, decoding, storing, analyzing, searching and managing of video content must be greatly increased. In order to fully explore the insights within the video content, video/image information analysis and retrieval are required to describe, store, and organize multimedia information. So that, advanced algorithms and platforms are mandated to assist people in detecting realtime events and finding video objects/resources conveniently and quickly.

China Video Solutions are designed and developed to overcome the technical obstacles of applying visual intelligence in a smart city environment. They are composed of a video analytics/search platform, a Chinese product, a strong embedded intelligent video platform, and a large-scale video platform. Many advanced technologies such as intelligent video search, analytics, management and optimization have

Table 1. Migration of Video-related City Infrastructure
City infrastructure with video-related entities is making great progress in China.

Aspect	Previous status	Current status
APP fields	Only safety	Safety, transport, social services, enterprise, etc.
Coverage	Limited area with limited cameras	Whole city with huge number of cameras
Quality	CIF/4CIF	720P/1,080P, 4K soon.
Data volume	2 to 5 Gbyte/camera per day	20 to 50 Gbyte/camera/day, 35 million hours of video content in mid-city
Network	Narrow band	Wire/Wireless broadband
Intelligence	N/A	Wide-spread intelligence
Platforms	Single/Isolated	Integrated/Cloud

APP: application CIF: common intermediate format

been applied to improve the level of safety, quality of services, and value-added applications in the smart city domain.

CURRENT PROGRESS AND PROBLEM STATUSES

Current Status

As indicated in Fig. 1, in the Chinese metropolitan environment, the most rapid development of intelligent video applications and services has been seen mainly in three categories.

Obviously, the fastest growing segment is the industry of public safety, more specifically, intelligent video surveillance. With the threat of terrorism, federal, state and local governments are devoting greater resources to video surveillance. Enterprise and related non-profit entities face the same issues as governments, such as industrial espionage, sabotage or theft. To prevent loss, retail businesses invest in storewide security cameras. Governments and corporations find video surveillance is a vital segment of their security solutions. In this category, entities with large-scale systems, such as the city department of public safety, are more likely to use the full capabilities of video analytics and video search. Meanwhile, other information from the security domain is integrated with the results of video intelligence to enhance the capabilities of security.

Another important category is utilizing video intelligence to enhance the service quality of city life. A huge amount of investment has been witnessed in transport, banks, environment protection, and other sectors. Run-time status in the city environment is more and more important in order to calibrate the resources of public services. Video analytics and search are obviously the most suitable ways to fulfill those requirements. The diversity of features, such as face, car, complicated behavior, and environmental change, based on video analytics is a notable phenomenon.

Besides the non-profit categories in the city environment, intelligent video applications also show very good potential in profitable fields. Business owners in multiple fields (e.g. commerce, residence, media, transport) utilize results from video intelligence to create new applications and to improve the performance of current ones so that they can directly sell new services or indirectly boost services that are already profitable. As above, a diversity of features is also in demand by customers, and intelligent video

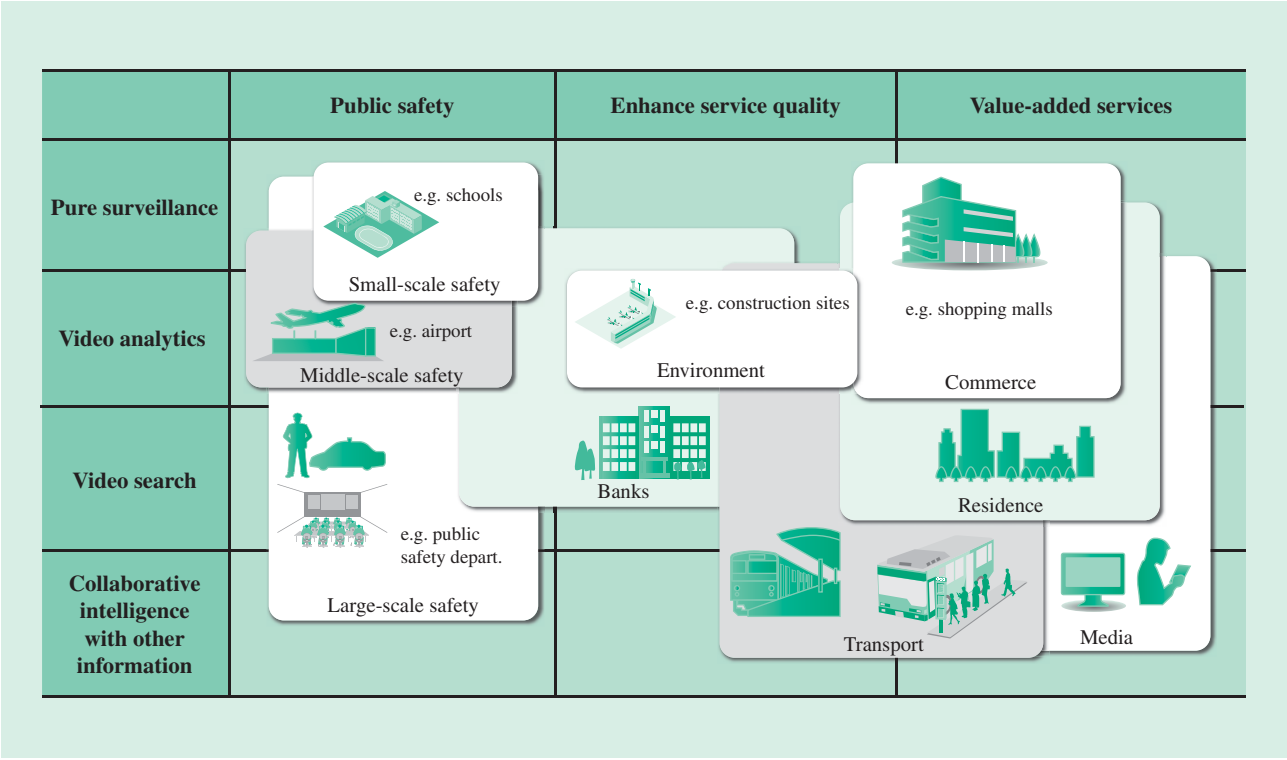


Fig. 1—Correspondence between Application Domains and Technical Categories.
Different fields in a smart city have different preferences for service requirements and technical requirements. Huge diversity has been observed regarding intelligent video technologies.

functions are always integrated with Internet/mobile Internet technologies to increase the usability and penetration for public citizens.

Intelligent video technologies have the virtues of easy installation, low maintenance, and one-stop entry for rich information and features. So, the broad applicability and strong support capability for services are making the number of video applications grow very fast.

On the other hand, in order to fully utilize the potential of intelligent video technologies, there are also many challenges in terms of technical aspects.

Technical Challenges

(1) Efficiency issues when searching a large volume of video content

Since the related architecture has been enhanced so much, the amount of video content generated has increased dramatically. In this sense, how to find a target object or specific video clips in a huge volume of content is a big issue. For instance, in a large-scale video surveillance system, such as on a city scale, searching for a human face could take more than 1,200 man/months. It is a huge burden for system operators and end users, and pretty much diminishes the usability and service quality. Some users created a system to add multiple tags in order to describe the video content with text. This way is okay for small-scale systems. However, in large-scale systems, creating the tags themselves will consume a huge amount of resources. And, a large portion of the video content is really difficult to describe. Lastly, tags for the same content could be different from person to person. So, differences in human understanding also devastate the usability of text-based video search.

In this sense, high speed and accurate content-based video retrieval are mandatory.

(2) Challenges for IT resources

One of the major technical challenges to deploying large-scale intelligent video applications is the high consumption of information technology (IT) resources, including computational resources [such as central processing unit (CPU)], storage resources (e.g. memory, disks), and networking resources. For example, one way of high-definition (HD) video stream can consume 6 Mbit/s for transmission over the network, one enterprise-class CPU core, and 512 Mbyte of memory for video decoding, one to two CPU cores, and 1 Gbyte of memory for video analytics with one feature, 50-Gbyte hard-disk for one day of storage, and related backup resources to provide reliable service. In large-scale systems, where hundreds of thousands of video cameras could be available in one city, the related costs as mentioned above will be a huge burden.

In the meantime, there are a number of circumstances where high-performance IT equipment cannot be deployed. Sites with high dynamics, such as running vehicles from public transportation or construction areas, have no conditions for providing good space, enough power, and the necessary cooling to support servers and storage. But, customers with such kinds of facilities still want to use video intelligence.

In this sense, the video systems and applications should not only be optimized to save resources, but also be customized to survive all kinds of local environments.

(3) Diverse requirements on multiple aspects

Video content potentially stores extremely rich content compared with text and voice. Mankind can explore much richer information from coded video. In different industries, the events or objects sought in video content could be quite different.

The visual perception of space in human beings shows great diversity, which is much more than what we can describe and depict. The same principle applies in the field of intelligent video as well. Based on a full market investigation of the necessary features in customers' minds, the result spans quite a board feature space, such as the bio-features of human beings, the behaviors of a single person or group of persons, vehicle features, city environment features, and the combination of or transition between the features mentioned above. Development of all these features by a single company could be a huge liability in terms of cost and time.

Meanwhile, all kinds of video systems have been deployed in the city environment for a long time. Most of them have no intelligent genes. While the new intelligent video functions are deployed, the integration with legacy systems could be quite challenging. Interfaces with all kinds of front-end devices, existing video management systems, a large number of applications/service systems from multiple customers, existing IT facilities, and management platforms must be considered.

As a short summary, although intelligent video technologies have brought great potential to high security, high quality, and high value in a smart city. Many technical challenges have been noticed as well. China Video Solutions are designed and developed

accordingly to promote city security, service quality, and value through intelligent visual-perceptions.

CHINA VIDEO SOLUTIONS

The overall architecture of a China Video Solutions is mainly composed of video search platform functions and strong embedded front-end platform with light-weight analytics (see Fig. 2).

Inside the video search platform domain:

(1) Data Ingest Engine is mainly in charge of extracting, transforming, and loading (ETL) of video content from all kinds of devices and systems at static sites in the city environment.

(2) The Video Analytics Engine is the major component for doing all kinds of video analytics. It fetches the buffered key-images from the Data Ingest Engine. Intelligent analytics can be done based on JPEG frames, and feature vectors are extracted as results. Intelligent analytics functions should be defined and supported based on the application scenario. For example, in the category of public security, human faces, cars, or car license plates could be required to be detected.

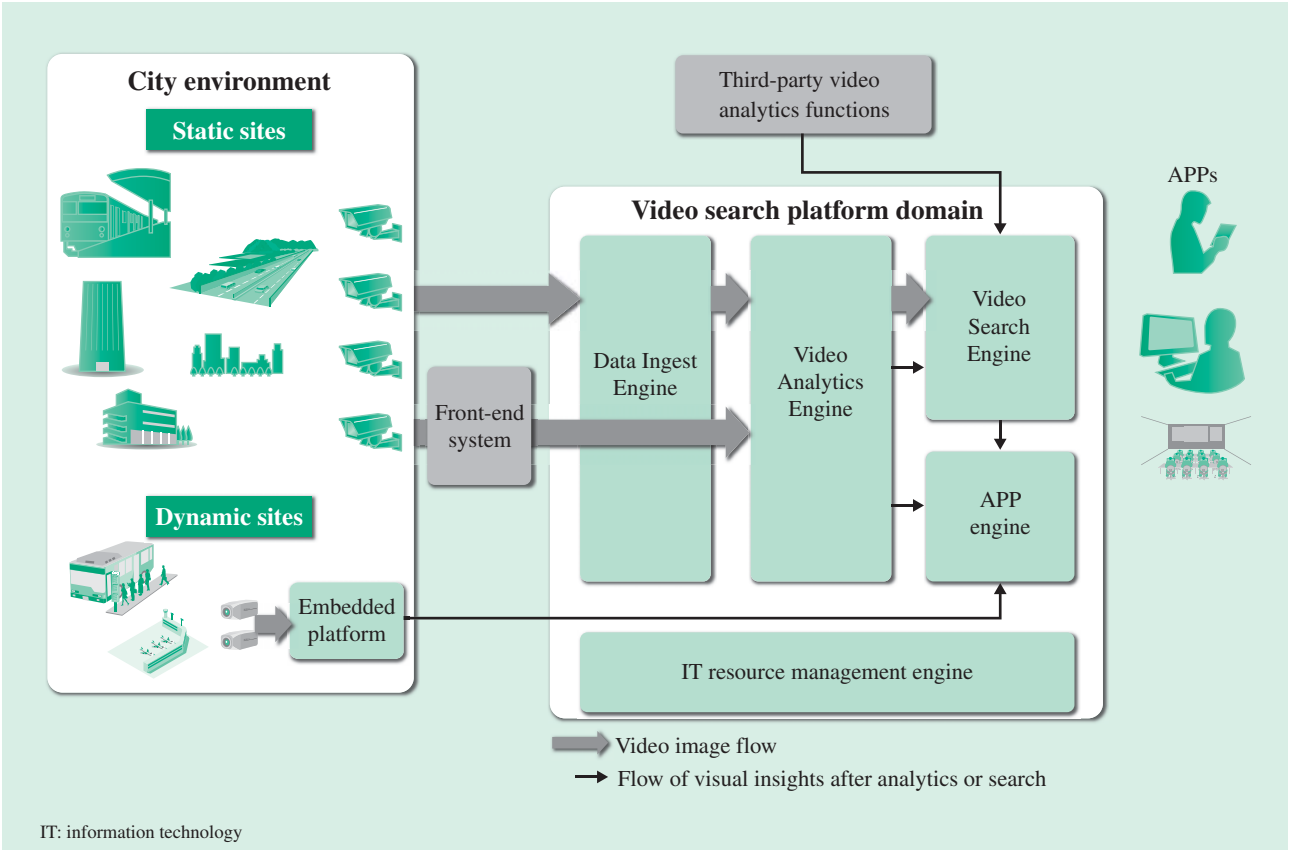
All these features, created by analytics modules and related mathematical models, will be described by feature vectors, and registered in the feature database of the Video Search Engine, the main component for searching.

(3) The IT resource management engine is the central controller for all IT resources for static sites. The application (APP) engine is the portal for external applications and services.

In the meantime, the embedded front-end platform is created to do all kinds of video analytics at dynamic sites. In order to combat the technical challenges mentioned in "Technical Challenges," many key technologies have been developed.

Enhanced Content-based Video Search

An initial large-scale video search system has been developed by Hitachi's Center for Technology Innovation.[2] Its strength is in performing high speed video searches in large-scale systems. With regular personal computer (PC) servers, a single content-based search of 100 million key images could be



IT: information technology

Fig. 2—Overview of the Solution Architecture.
Static sites are locations with video cameras in a static spot, reliable wired networks, and a server-friendly environment.
Dynamic sites could be vehicles or construction sites, which could be moving, wireless networks, and a harsh environment.

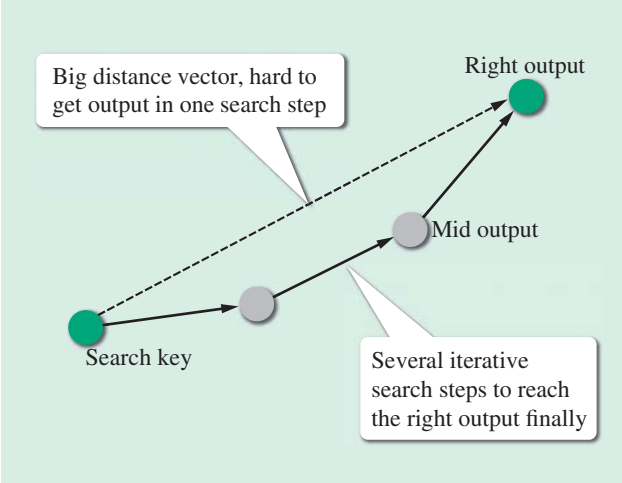


Fig. 3— Principles of Iterative Search.
In large-scale video search systems, a huge number of feature vectors, which are created as the result of video analytics, bring in huge noise. Iterative search is used to bypass the impact of the noise, and get the correct results in a limited number of steps.

finished within one second. The results of onsite trials show the efficiency of quick video search could be enhanced by more than 99%, when users try to find objects with key images in a large video database.

In China, the following new technical features have been developed to deal with super-large-scale video surveillance environments:

(1) Iterative search

In super-large-scale video systems, quite a few factors can contribute a lot of noise to video analytics. So that, the search accuracy over a huge amount of feature vectors will degenerate. Iterative search is developed to maintain the same search accuracy by taking several exploratory steps, as indicated in Fig 3.

(2) Scene-based filtering function

In search results, the same objects in continuous frames of the same scene are more likely to be shown together. It will decrease the valuable information when users want to show objects in multiple scenes. Therefore, a filter function has been developed in the APP engine to extract a single instance in one scene, while filtering out duplicate instances.

(3) Dynamic resource scheduling in the cloud

Because the scale of the video system is going to be big, private clouds are chosen by customers as the IT infrastructure. As shown in "Technical Challenges," intelligent video technologies could consume a huge amount of IT resources. So that, the scheduling of resources based on virtual machine management is necessary. In trials, with advanced algorithms for IT scheduling, 20% of resources could be saved.

Strong Embedded Platform and Light-weight Video Analytics

As indicated in "Current Progress and Problem Statuses" chapter, there are a number of dynamic sites where traditional analytics equipment (e.g. servers) cannot be deployed.

China's first all-in-one equipment has been developed to deal with this issue. It is an open platform based on embedded Linux[*1] with enough computational resources [CPU+digital signal processor (DSP)], video recording, and rich networks (WiFi,[*2] 2G, 3G). And, it is strong enough for harsh environments.

Meanwhile, to save cost, one device supports multiple analytics functions for multiple video channels. So, light-weight analytics algorithms, which traditionally cost a huge amount of IT resources, have been developed for this purpose.

The platform is designed to be open to any new light-weight algorithms in order to quickly and significantly accelerate the procedure for going to market.

Rich Interfaces

Rich interfaces have been developed to handle the obstacles of diversities as mentioned in "Current Progress and Problem Statuses" chapter.

(1) Interfaces to front-end devices/systems

In the Data Digest Engine, it supports interfacing with cameras from hundreds of camera vendors in China. And, it also can upload offline content from various storages.

(2) Interfaces to third-party video analytics

In fact, due to the quick industrial growth, many companies are working on algorithms for video analytics. This interface allows the Video Search Engine to register the result of external intelligent functions. Then, it can fully utilize functions from the entire industry.

(3) Web API for diverse applications

All kinds of application platforms can utilize the capabilities of this solution in a short time using this interface.

APPLICATION CASES IN CHINA

Large-scale Video Search for Public Safety

With the large scale involved in public safety, how to find a target person or car in a huge video database

*1 Linux is a registered trademark of Linus Torvalds.
*2 Wi-Fi is a registered trademarks of Wi-Fi Alliance.

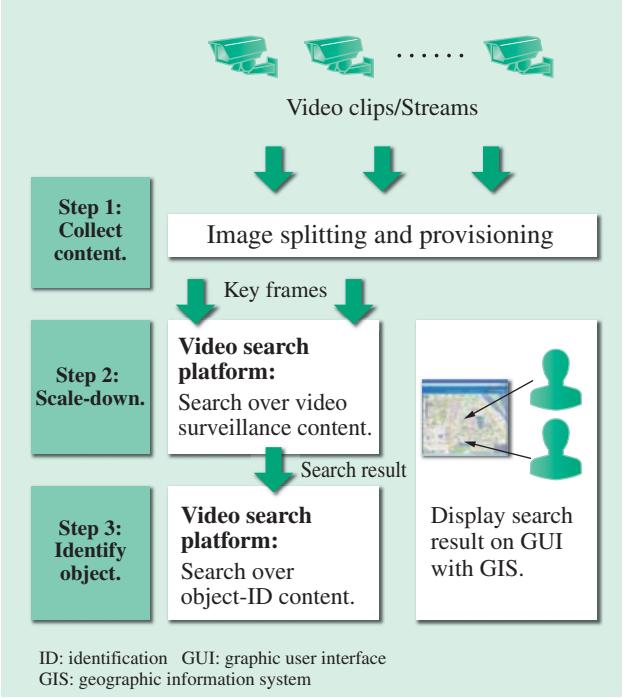


Fig. 4—Typical Work Flow in Video Search for Public Safety. The video/image search is used to help the users filter the target objects (e.g. people, cars) quickly and accurately from a huge amount of video content.

must be a big issue. A typical service flow can be seen in Fig. 4. Basically, the search functions are utilized in order to narrow down the scope of the search and to find the target objects quickly and accurately.

In face search and car search trials in two cities in China (Table 2), this solution can achieve a quick search with 90%+ accuracy. Search speed has been increased by more than 99% in real practice.

Intelligent Vehicle Platform to Improve Service Quality and Add Value for Public Transportation

Although scheduling of public transportation has been practiced for many years, the usability of transport resources in a smart city is still not optimized. The major issue is that the current scheduling mechanisms are mainly based on information outside of vehicles, without considering the in-vehicle status of passengers.

Passenger density and number detection through video analytics based on the embedded platform have been achieved and operated in several cities in China. In accord with the diverse statuses of passengers in different buses/subways, public transportation resources (such as buses, vehicles, subways, trunks) could be further optimized.

TABLE 2. Onsite Results of Video Search for Public Safety
Video search can achieve quick search results with 90%+ accuracy, and its search speed has been increased by more than 99% in real practice.

Camera	City 1	City 2
1920×1080	100% (1 to 2 iterations)	N/A
1280×720	100% (1 to 2 iterations)	N/A
720×576	92% (2 to 3 iterations)	90% (1 to 3 iterations)

FUTURE DIRECTIONS

Although China Video Solutions have been recognized as having good potential in the Chinese market, there are still a number of fields that need to be further explored. Basically, this solution could be further expanded to more vertical fields in the smart city environment. In order to achieve this target, more analytics features should be developed or integrated in the near future. In the meantime, the strong intelligent embedded platform can be used in many other fields with dynamic environments. So that, more light-weight video analytics functions should be migrated to this platform.

Lastly, over-consumption of IT resources is still one of the major bottlenecks of intelligent video applications in China. Further optimization in algorithms, platforms, and management could be the focus of future effort.

CONCLUSIONS

Targeting the trillion-dollar market of China related to intelligent video technologies, China Video Solutions were created to further enhance the security, quality of service, and value of application in a smart city environment. With the world's leading capabilities in video search and video analytics, hybrid video platforms, rich interfaces and high-profile optimization capabilities, these solutions show good potential to further explore the usability of visual perceptions of mankind.

Friendly partnerships in a technical portfolio will provide good contributions to the entire industry in the next step. Based on these solutions, it is believed that more promising applications and services can be created in Chinese metropolitan areas.

REFERENCES

(1) "China Intelligent Video Market and Necessary Technologies," CCW research report 2015.
(2) A. Hiroike and Y. Obuchi, "Information Retrieval for Large-scale Image and Audio Archive," Hitachi Hyoron **95**, pp. 200–205 (Feb. 2013) in Japanese.

ABOUT THE AUTHORS

Peng Yang
CSI China Customer Co-Creation Project 2, Hitachi (China) Research & Development Corporation. He was engaged in developing intelligent multimedia solutions for urban security projects at the time of writing.

Tsutomu Imada
Business Strategic Office, Hitachi Beijing Tech Information Systems Co., Ltd. He is currently engaged in business development of image search solutions in China and APAC area markets.

Kai Liu
Products Solution Division, Hitachi Beijing Tech Information Systems Co., Ltd. He is currently engaged in business development of image search solutions in the China market.

Daisuke Matsubara
Global Center for Social Innovation - Tokyo, Research & Development Group, Hitachi, Ltd. He is currently engaged in urban security projects. Mr. Matsubara is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).