Toward Intelligent Video Surveillance

Hiroyuki Arai[†], Kazuyuki Iso, Akira Kojima, Hitoshi Nakazawa, and Hideki Koike

Abstract

NTT Cyber Space Laboratories is developing image processing technologies that can extract desired information from huge amounts of video data to enable more sophisticated image-based services. It is also developing video surveillance solutions that utilize these technologies. This article overviews the image processing technologies and presents an example of an image processing solution for the financial industry.

1.   Prospects for effective utilization of cameras

Video cameras have been installed in many places in response to the steady rise in concerns about crime and terrorism and corporate security requirements. The benefits that they are expected to provide include acting as a deterrent to crime, enabling analysis after crimes and incidents, and enabling the detection of crimes and dangerous situations and the issuing of warnings. In practice, however, such surveillance systems have not achieved the expected results because the amount of image data generated every day is so huge that it is virtually impossible to check it manually. Thus, there are strong demands for technology that can automatically detect important scenes in the many streams of monitoring images.

2.   Approach to intelligent image monitoring

To meet the requirements of the marketplace, NTT Cyberspace Laboratories has been developing movement and face detection technologies as well as image processing technologies to assist in the monitoring process. These technologies, shown in **Fig. 1**, work as filters that can identify the scenes of greatest interest. Each filter can effectively and automatically

†   NTT Cyber Space Laboratories
    Yokosuka-shi, 239-0847 Japan
    Email: arai.hiroyuki@lab.ntt.co.jp

identify these scenes and quickly generate warnings. These two capabilities raise the efficiency of surveillance systems and thus increase their deterrence effect.

We have also been developing image monitoring solutions to meet the needs of industry in collaboration with operating companies and developing companies at the same time as developing the image processing technologies.

3.   Image processing technologies to assist monitoring

We have developed image processing technologies on the assumption that humans are the target to be monitored. The main technologies are listed below:

(1) Motion detection technology

This can reliably detect motion in an image sequence (**Fig. 2**). While motion detection has recently been provided by many commercial surveillance systems, problems such as false positives and negatives have not been handled well. One of the major factors degrading detection accuracy is changes in lighting. Our technology has no such drawback because it uses feature metrics that are stable in various lighting environments.

(2) Face detection technology

This can detect a human face in an image and identify the location of the face (**Fig. 3**). Face detection technology has already been developed to support
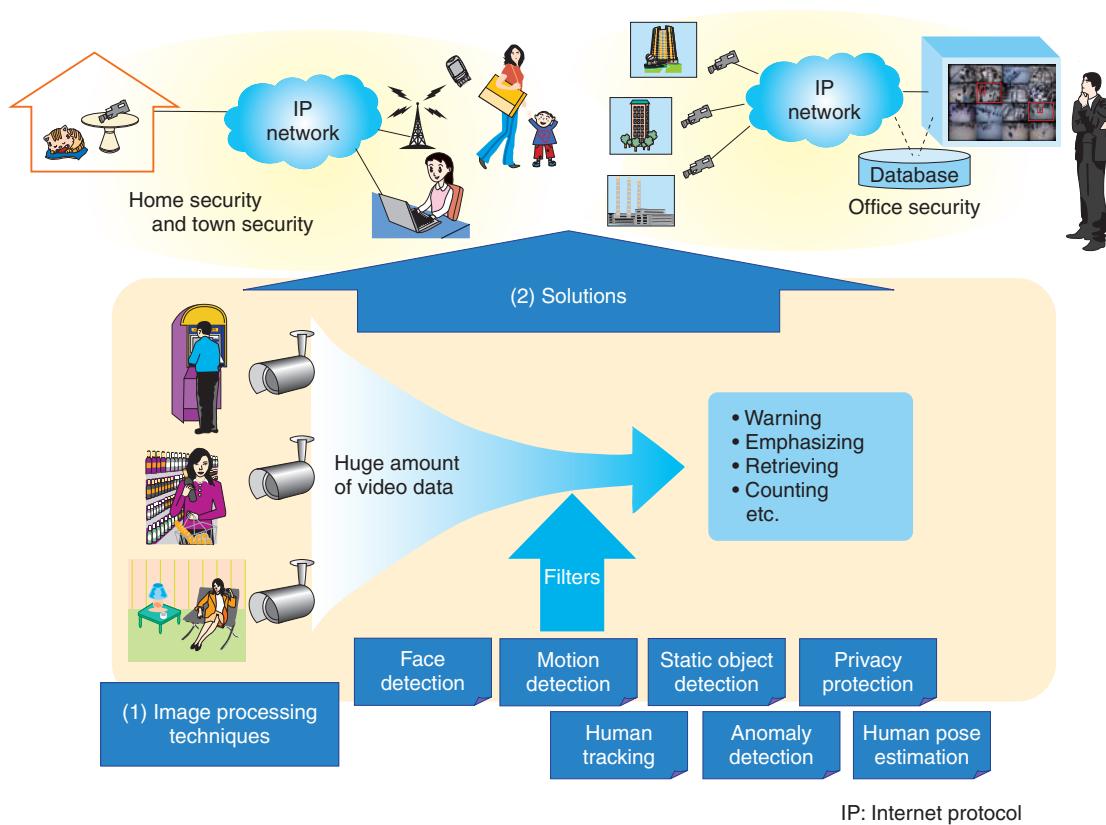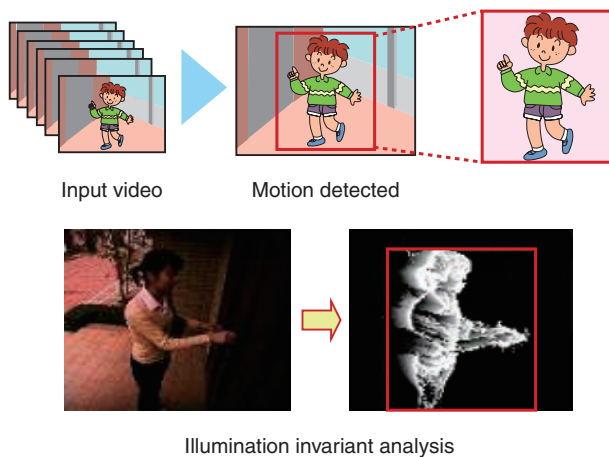
Fig. 1. Overview of approaches.
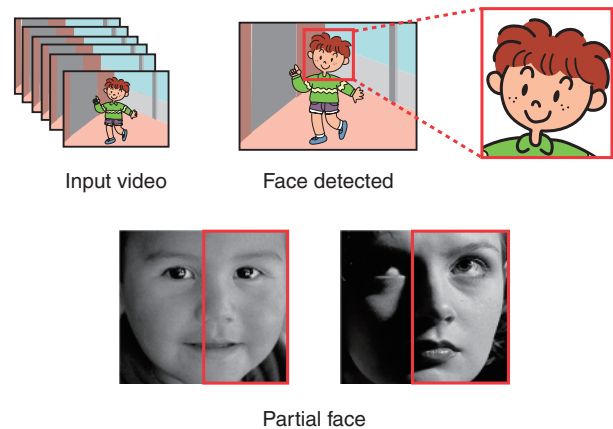

Fig. 2. Motion detection technology.

Fig. 3. Face detection technology.

image quality control in digital cameras. However, authentication systems and digital cameras generally target full-front faces, which makes face detection relatively easy. Until now, there has been no truly effective face detection technology for surveillance systems that capture faces from every possible angle.

Our approach is to focus on the partial face, which is half the face including the nose. As a result, we have developed a face detection method that is stable against variations in face direction. We are refining this technology to further raise its accuracy.

(3) Static object detection technology

Fig. 4. Static object detection technology.

Fig. 5. Privacy protection technology.

This can detect things that have been left behind or removed (**Fig. 4**). It is based on detecting a partial change in an image and then detecting the subsequent succession of changes. Since the technology uses an algorithm that can discriminate short- and long-term changes in images of a scene, it can be applied to very busy areas such as railway stations.

(4) Privacy protection technology

This can detect a moving object in an image and blur only this specific part (**Fig. 5**). It is needed when the video monitoring is used for purposes other than security. For example, in a fast food restaurant with seats upstairs, this technology can show people downstairs how many vacant seats are available upstairs while protecting the privacy of the customers by blurring their faces. The number of such applications is increasing as more cameras are being installed in public areas, and third parties will observe the resulting images. In such circumstances, our technology can prevent a loss of privacy.

(5) Human tracking technology

This can trace human movement in a three-dimensional space. Human movement is recognized by detecting the position of a human at each instant of time. This ability is needed not only for security applications such as detecting suspicious behavior in facilities, but also for marketing applications such as analyzing customer behavior in shops. This technology is introduced in the third article in this Special Feature: "3D Human Tracking for Visual Monitoring" [1].

(6) Anomaly detection technology

This can automatically detect unusual events contained in the stored image data created by video surveillance systems. This is necessary because we cannot predict and locate all suspicious events in advance. This technology is introduced in the second article: "Detecting the Degree of Anomaly in Security Videos" [2].

(7) Human pose estimation technology

This technology can detect head and body posture (direction and approximate arrangement). We have combined a pattern recognition technology and a three-dimensional information extraction technology to develop an algorithm that can extract information related to head and body posture. This technology is introduced in the fourth article: "Human Pose Estimation for Image Monitoring" [3].

The first four technologies mentioned above have been developed, while the other three have been verified in basic experiments.

4. Image monitoring solution for the financial market

4.1 Market requirements

The image monitoring market is being stimulated by requirements that monitoring images from cameras should be kept for at least a few years and rapidly retrieved and analyzed when required at a date considerably later than the recording date instead of soon afterwards.

In financial institutions, the main services based on image monitoring are status verification at ATMs (automated teller machines (also known as cash points)) and branch offices, image recording, and post-event searching for background information and image submission. Existing systems using video tape recorders (VTRs) or digital video recorders (DVRs) are not perfect because it is troublesome to change the recording media and check the recording status; moreover, it is not easy to identify the desired images. The need for higher levels of security at ATMs and branch offices is growing because criminals have installed small cameras in ATMs to steal card information and crime has been increasing overall in recent years.

Under these circumstances, the financial market has set three requirements for image monitoring:
(1) No loss of recorded data
(2) Significant overall cost reductions (e.g., management, operating, and system costs)
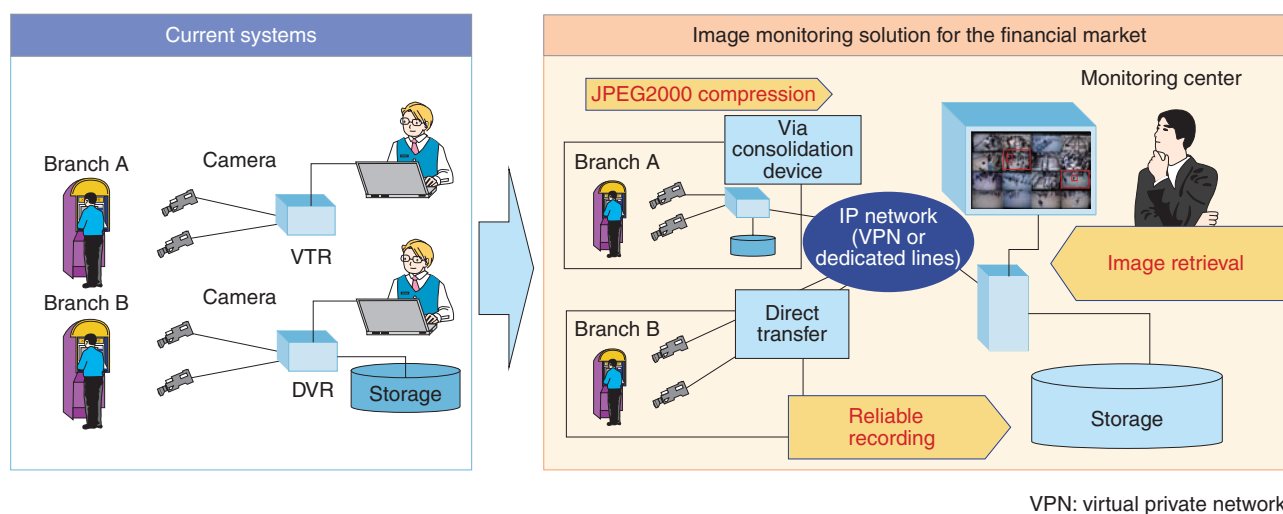(3) Higher levels of security

Fig. 6. Overview of image monitoring solutions for financial market.

4.2 Overview of solutions

To meet the above requirements, we have developed solutions to achieve overall cost reductions and enhanced security levels through the use of image processing technologies and highly reliable consolidation of camera images at a center via networks (**Fig. 6**).

(1) Image consolidation without image loss

To reliably consolidate and record images even via an inexpensive best-effort network, it is necessary to handle temporary network failures and bandwidth fluctuations. In our system, consolidation devices, placed close to the cameras, temporarily store the images output by the cameras and then send them to the server when triggered by the server. Each consolidation device has sufficient memory to store data for several hours, which guards against loss due to brief network outages. These devices can also work in best-effort networks because they can tolerate bandwidth fluctuations. This consolidation technique helps to make the cost of the overall surveillance system much lower. For more reliable image storage, the system can transfer the images to the center directly if one or more consolidation devices fail.

(2) Cost reduction using JPEG2000

The volume of image data greatly affects the network and storage costs. It is important to compress data as much as possible while keeping the image quality required for monitoring*. JPEG2000 has the advantage of keeping the required quality while achieving stronger compression than the widely used JPEG scheme. Our system offers users a choice of either JPEG2000 or JPEG.

(3) Image retrieval based on image processing

Our system can detect movement, faces, and static objects within the image data held by the center for effective image retrieval; this enables the results to be output as metadata. This function supports post-event searches, such as checking scenes that include faces near a safe on a specified date, so it contributes to greater security.

5. Future work

We have been developing sophisticated and practical image processing technologies with the goal of making possible a safer and more relaxed society. We have started to develop a crowd analysis technology that can support security in crowded areas such as stations and airports and a technology for extracting and analyzing human actions. Future reports will cover non-steady-state estimation technology, human tracking technology, and human posture estimation technology. We are also examining a technology that can prevent tampering with monitoring images and an encryption tool to ensure that monitoring images are used only for the intended purposes.

References

[1] T. Osawa, X. Wu, K. Wakabayashi, and H. Koike, "3D Human Tracking for Visual Monitoring," NTT Technical Review, Vol. 5, No. 11, 2007.

* Within the recognition conditions of the evaluation chart for security image systems as defined by the Japanese Security System Association.

https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr2007
11sf3.html

[2] K. Sudo, T. Osawa, K. Wakabayashi, and H. Koike, "Detecting the
Degree of Anomaly in Security Videos," NTT Technical Review, Vol.
5, No. 11, 2007.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr2007
11sf2.html

[3] S. Ando, X. Wu, A. Suzuki, K. Wakabayashi, and H. Koike, "Human
Pose Estimation for Image Monitoring," NTT Technical Review, Vol.
5, No. 11, 2007.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr2007
11sf4.html

Hiroyuki Arai
Senior Research Engineer, Visual Media Communications Project, NTT Cyber Space Laboratories.
He received the M.S. degree in physics from Hokkaido University, Hokkaido, in 1991. He joined NTT in 1991 and engaged in research on a map recognition system. He was transferred to NTT DATA in 2001 and developed image processing techniques. He was transferred to NTT Cyber Space Labs. in 2005. From 2000 to 2005, he was a fellowship-researcher in the "Natural Vision Project" of the National Institute of Information and Communications Technology. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan and the Institute of Image Information and Television Engineers.

Kazuyuki Iso
Research Engineer, Visual Media Communications Project, NTT Cyber Space Laboratories.
He received the M.S. degree in knowledge science from Japan Advanced Institute of Science and Technology in 2000. He joined NTT Cyber Space Laboratories in 2000 and engaged in research on visual communication systems. He was transferred to NTT Resonant Inc. in 2004 and worked to develop videoconference systems. He was transferred to NTT Cyber Space Labs. in 2006. He is a member of the Information Processing Society of Japan and the Virtual Reality Society of Japan.

Akira Kojima
Senior Research Engineer, Visual Media Communications Project, NTT Cyber Space Laboratories.
He received the B.E. and M.E. degrees in mathematical engineering and information physics from the University of Tokyo, Tokyo, in 1988 and 1990, respectively. He joined NTT Labs. in 1990, where he was engaged in research on a video database system and multimedia information retrieval. From 1996 to 1999, he worked on a digital library project at the Business Communications Headquarters of NTT East. He took up his present post in 2006. He is a member of IEICE and the Association for Computing Machinery.

Hitoshi Nakazawa
Senior Research Engineer, Supervisor, NTT Cyber Space Laboratories.
He received the B.S. degree in electronic engineering from Ibaraki University, Ibaraki, in 1984. Since joining Nippon Telegraph and Telephone Public Corporation (now NTT) in 1984, he has mainly been engaged in R&D of facsimile intelligent communication systems (F-net), digital rights management systems, desktop conference systems, and intelligent monitoring systems (NiMSA). He is a member of IEICE.

Hideki Koike
Senior Research Engineer, Supervisor, Group Leader, Visual Media Communications Project, NTT Cyber Space Laboratories.
He received the M.S. degree in mathematics from Tohoku University, Miyagi, in 1985. He joined NTT Labs. in 1985 and engaged in research on image processing. He was transferred to NTT COMWARE in 2001 and engaged in research on RFID. He moved to NTT Cyber Space Labs. in 2007 and is engaged in research on computer vision. He is a member of IEICE.