Development of Intelligent Movie Data Compression Technology for Video Surveillance Systems

Tetsuya TAKAMORI*, Yi HU*, Xuebin HU*,
Yukinori NOGUCHI*, and Makoto YONAHA*

Abstract

We developed "Intelligent Compression", an intelligent movie data compression technology that is useful in video surveillance systems. We expanded our unique image recognition technology, which is used in the digital photography field, in the development of "Intelligent Compression". With its highly accurate recognition capability, this technology realizes a high compression ratio while preserving the information necessary for surveillance. It also makes quick image searching and crowd analysis possible, contributing to wider application of surveillance systems.

1. Introduction

In recent years, surveillance cameras have become a part of our daily lives. In addition to security cameras often introduced by mass media, employee monitoring cameras for compliance control and safety work and traffic surveillance cameras have been increasing. They have been installed in airports, train stations, commercial facilities, financial institutions, hospitals, offices, schools, parks, etc. Everyone is likely to be recorded on video at least once every day.

However, in many cases, the quality of those video images is not very good and it is very rare to be able to immediately identify a wanted person in security camera videos released on TV. Although the quality of images is affected by that of the cameras themselves, the fidelity of video recorders can also be a significant factor that decides the clarity of the images. That is, compression methods for storing image data may be the cause of the degradation of video images. There are intense demands to compress image data. For example, video tape recorder is expected to store one week of data from four cameras in a single video tape. Therefore, a need for compression technology that enables high-resolution, long-duration video recording has always been strong in this domain.

In the meantime, image data sizes have been expanding inevitably because of the trend for high-resolution cameras producing high quality images, large-scale systems supporting network cameras and demands for long-duration recording to monitor compliance in offices. This, together with the aforementioned increase of surveillance cameras and the demand for high quality images, has increased storage costs. We expect that the need for compression technology which can solve those issues will keep increasing in the future.

Thus, we developed a compression technology that enables high-resolution, long-duration recording for surveillance video images and commercialized it as the network video recorder "FUJIFILM Clearvision FVR-100/200" supporting high quality network cameras (Fig. 1)[1].

In this paper, we explain the details and capabilities of the Intelligent Compression technology that is applied in this series of network video recorders.

Fig. 1 FUJIFILM Clearvision FVR-100.

2. Intelligent Compression Technology

2.1 Outline of Intelligent Compression Technology

Image data compression consists of the following three processes: i) Bias the frequency of occurrence of specific data with some signal processing techniques; ii) selection of target data and bias the frequency of their occurrence after removing non-target data; iii) data reduction by code assignment in accordance with data occurrence. In the Intelligent Compression technology, the process described in ii) above is the focus. By enhancing the function of selecting

Original paper (Received November 25, 2011)
* Imaging Technology Center
Research & Development Management Headquarters
FUJIFILM Corporation
Miyanodai, Kaisei-machi, Ashigarakami-gun, Kanagawa
258-8538, Japan

data to be saved in surveillance videos, it enables compression in which target objects can be viewed clearly with high time resolution and spatial resolution. If the targets are human faces, the frame rates and resolution of human face areas are unchanged in compression, while other areas of the images are reduced to a barely recognizable level (Fig. 2).

Fig. 2 Intelligent Compression overview.

Such the Intelligent Compression function requires a detector that selects the targets to be viewed clearly and a compressor that controls the compression process area by area based on the detection results (Fig. 3).

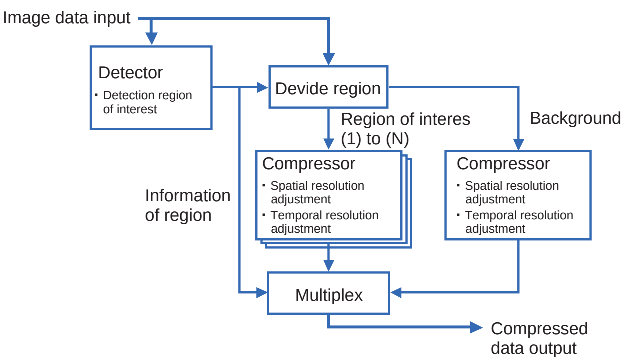

Fig. 3 Block diagram of Intelligent Compression.

2.2 Detector

In surveillance videos, human figures and moving objects are the targets that need to be clear. As shown in Fig. 4, the detection of those objects can be covered almost completely by using motion detection technology and human head detection technology. In the field of motion detection technology, there already exist many useful technologies. Therefore, the development of human head detection technology was the key to realize a detector that can achieve required performance.

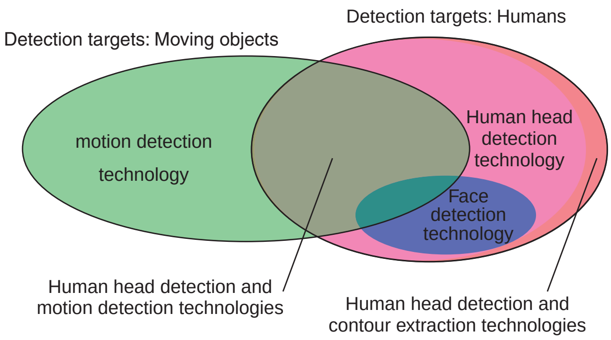

Fig. 4 Relation between detection target and technical components.

2.2.1 Overview of Human Head Detection

Fig. 5 illustrates the configuration of our human head detection process. As shown, the core of the process is the human head detector. We developed it based on computer vision technology.

We have been involved in the development of computer vision technologies for many years and achieved high accuracy recognition, specially at the point of false positive. Among such technologies is the face detection function applied in our digital cameras for focus and exposure adjustment. However, the same function can achieve only 10% accuracy in human head detection for videos from surveillance cameras. The reason is that face angles shot with surveillance cameras are different from those taken with digital cameras. In most cases, the surveillance images are taken downward from near ceilings. Front faces are seldom seen in surveillance videos while, with digital cameras, faces are taken from the near front. To solve this issue, we developed a novel technology that enables the detection of the human head from any angle.

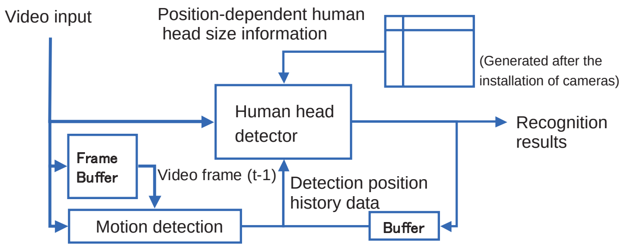

Fig. 5 Block diagram of head detector.

The human head detector searches and recognizes human heads within each image frame with video inputs. The recognition is made by comparing the image data with templates of human heads derived from multiple image characteristics. Human heads are detected by scanning onto images those templates on every size and rotation angle, and assessing the matching results (Fig. 6).

The face detection technique applied in our digital cameras is used in this head detection. However, human heads have more image characteristics variation than faces and many objects whose image characteristics are similar to those of human heads can exist in videos. To achieve high detection accuracy, a greater number of image characteristics are incorporated into templates used in head detection than in face detection.

Fig. 6 Scanning procedure for head detection.

2.2.2 Restrictions on the Search Range for Head Detection

Restrictingthe search range for head detection can increase its processing speed and can reduce its false positive rate. That is an important element in the development of the head detection process. Restrictions on the search range are realized by the following two techniques.

Based on the fact that most surveillance cameras are installed pointing downwards and their viewing angles are fixed for a considerable length of time, the first technique utilizes the relationship between human head positions and their sizes within images.

If the camera is installed pointing downwards, the longer the distance from the camera, the smaller and the higher up the objects of the same size appear in the image. At the same time, in surveillance camera shooting, human heads remain within fixed ranges in height from the floor and in size. Therefore, depending on the position in the image, possible human head sizes also remain within a specific range (Fig. 7).

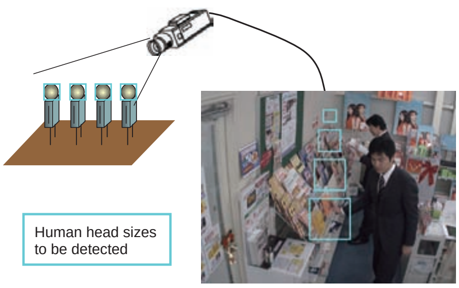

Fig. 7 Relation between the head size and position on screen.

With this principle, by narrowing down human head sizes to be detected depending on the position, the high speed and low false positive technique was realized.

However, for commercialization, this technique required a function to recognize the relationship between human head positions and their sizes within images that varies depending on the installation position of each surveillance camera. Therefore, we also developed a system that derives and sets the said relationship automatically as values by analyzing the videos shot with the installed surveillance cameras.

The second technique utilizes the motion of objects in images, which is also based on the fact that the viewing angles of surveillance cameras are fixed for a considerable length of time.

Moving objects are detected in image areas using data that differs between frames, and areas without any motion are excluded from the search range for head detection. To eliminate false positive caused by the subtle movement of cameras and image noise, data are extracted from open areas in images to recognize the motion of objects.

Because simply excluding motionless areas may cause failure in detecting still humans, a function to correct the search range based on the analysis of past data has been added.

2.3 Compressor

A compressor needs to reduce the size of data while retaining the required image quality of important areas. However, in general motion picture compression techniques, the whole screen image or image file is compressed evenly and this often blurs the details of the targets such as facial features in shots that have many fine textures and sudden motion while still shots are unnecessarily clear.

The detector that we developed for this series of network video recorders has been configured so that spatial resolution and difference in both time and space quantization can be controlled from area to area. This enables stable compression of each target, retaining appropriate image quality in any shot (Fig. 3).

3. Capability of Intelligent Compression

3.1 Image Quality-Compression Ratio Capability

Image quality and compression ratio forms a trade-off relationship. If the compression rate is increased, image quality is degraded and vice versa (Fig. 8). Because image

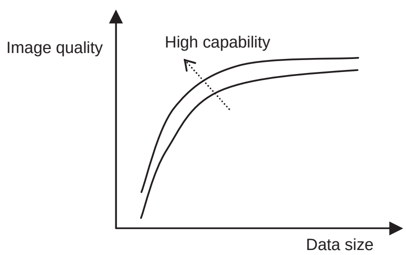

Fig. 8 Compression rate-image quality curve.

quality and data sizes can vary greatly, image quality comparison is made with the data sizes fixed and data size comparison with the image quality fixed.

Fig. 9 shows the comparison of two images compressed into the same size: one from a surveillance camera digital video recorder (DVR) on the market and another from our network video recorder FVR-100 that incorporates the Intelligent Compression technology. As seen, in the light of human identification, the quality of the image from FVR-100 far exceeded that from the general DVR. It was thus confirmed that the Intelligent Compression is more capable of high-quality image compression than are general compression methods.

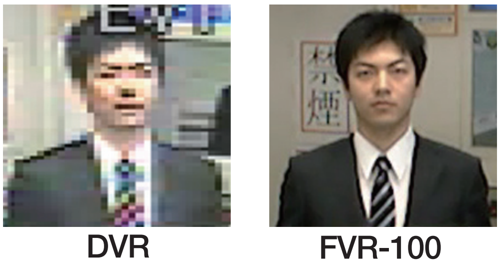

Fig. 9 Comparison of image quality between different compression rates.

On the other hand, in Fig. 10, the data sizes of compressed images are compared. They have been compressed with three different techniques, namely JPEG compression, H. 264 compression (the latest motion picture compression technique applied in 1seg and Blu-ray) and the Intelligent Compression, into the same quality levels in the light of human identification. The data sizes indicated are the averages of values gained via demonstration tests at multiple financial institutions and commercial facilities. The quality of cameras connected and shooting environments may affect their compression ratios, but it is still clear that the Intelligent Compression can achieve a higher ratio than the other two techniques.

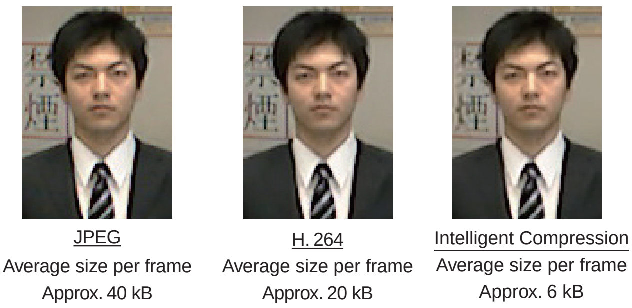

Fig. 10 Comparison between compression rates.

3.2 Compression Speed

The human head detection process applied in the Intelligent Compression is generally time-consuming because it requires many calculations in its image recognition process. However, the development of technology that limits the search range for human head detection has increased its processing speed dramatically.

The speed of the Intelligent Compression including the detection process is about 160 ms per frame on average on PCs (Intel Core 2, 2 GHz CPU with 2 GB memory). The Intelligent Compression technology is feasible on PCs/servers.

4. Intelligent Search

Because of its high compression ratio, the Intelligent Compression is suitable for long-duration recording. In that application, a function to retrieve the required data quickly is indispensable. We therefore developed an intelligent search function, too.

The intelligent search function is realized by making it possible to store, frame by frame, the position data of detected objects generated in the process of the Intelligent Compression within the compressed data itself. An original position data format was designed to keep its influence on the overall compression ratio to a minimum.

For example, a shot in which the detection target (e.g., human) enters a specified position can be searched quickly by comparing the data of the specified shooting area and the stored position data. By saving the search results as cueing bookmarks, the video can be skipped from one shot to another that require confirmation. This search function considerably reduces the time required for video checks, compared with fast-forwarding (Fig. 11).

Fig. 11 Intelligent search.

5. Analysis of the Degree of Congestion

With the human head detection data, the number of people in a shooting area can be derived. As shown in Fig. 12, the time-scale summary of the detection data can indicate the flow of humans including the degree of congestion in the area.

We are interested in such an analytical function because it can produce a valuable new feature for customers by extracting useful marketing information from surveillance videos. Customer expectations of that secondary benefit are also very high. We are going to extend that function by enabling the integration of multiple camera information in the future.

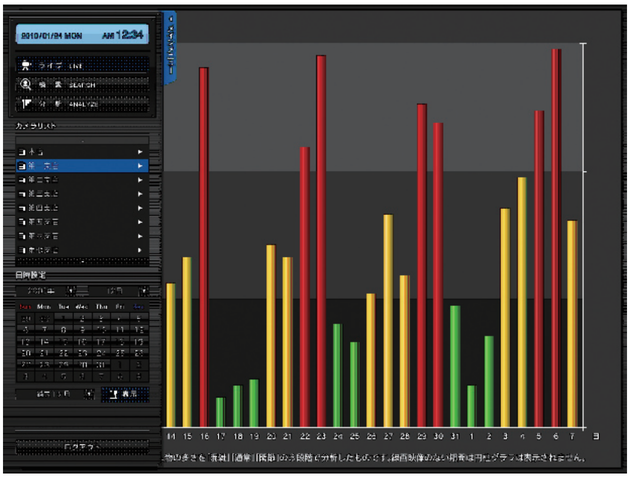

Fig. 12 Number of people count.

6. Conclusion

In this paper, we explained the features of the Intelligent Compression technology for surveillance videos. With this technology, human image data area can be compressed without quality degradation and, in addition, storage costs for surveillance video recording can be reduced. Having incorporated this compression technology, the network video recorder FVR-100 achieves a size reduction of image data to one-thirtieth or less of Motion JPEG that is generally used in this domain. This capability is useful in various other situations that require human image data communication such as TV conferences and remote monitoring. By extending the detection targets to those other than human figures, we can respond to the demands of reducing the size of various other data without quality degradation. We expect that such demands will become stronger in parallel with advances in information and communication technology and spread to fields other than video surveillance. To respond to those diversified needs, we are going to further improve the functions and capabilities of this technology.

Reference

1) Hirata, Kenji. High-resolution and High-compression Engine Next-generation Network Surveillance Video Recorder, "FUJIFILM Clearvision FVR-100". Image Laboratory. **20** (11), 72-76 (2009).