An Implementation of Video Surveillance Systems with Progressive Quality Improvement Approach

Chaxiong Yukonhiatou[1] Tomoya Kawakami[3] Tomoki Yoshihisa[2] Yuuichi Teranishi[4,2]
Shinji Shimojo[2]

Abstract: Due to the widespread of video surveillance, real-time detection systems for humans or cars are deployed in many places such as airports, banks, and parking spaces. In many of these systems, the cameras continuously send their original quality images to the processing computers via networks such as high resolution images. However, the high quality images require high performance or large communication traffic to be processed and cause serious problems to the processing computers and networks. The quality improvement is one of approaches to reduce the communication traffic in the surveillance systems. For example, the number of installed cameras can be increased as the communication traffic decreases or the processing computer can faster to find a thief recorded in image data got from cameras with a faster transmitting image data from cameras to the processing computers. Therefore, we have proposed the method which reduces the communication time and traffic based on progressive quality improvement (PQI) approach. In this paper, we describe our system implementation and performance experiment of our proposed method.

1. Introduction

Recently, real-time detection systems such as humans and cars detection systems are deployed in several places. For example, in the banks and airports. In most of these systems, the cameras continuously send their original quality image (full quality image) data to the processing computers in real-time via the computer networks even when the target objects for detections are not recorded. This causes a large communication traffic that is a serious problem to the processing computers and networks such as a long communication time, limitation of the number of installed cameras. The progressive quality improvement approach is one of approaches to improve the performance of surveillance systems. For example, the administrators of surveillance systems can faster to receive the notifications from processing computers for a thief recorded in image data got from cameras. For above example, a shorter communication time is a factor to improve the performance of surveillance systems.

A higher quality image gives a higher analysis accuracy. However, a large communication traffic occurs and suppresses the number of installed cameras and transmission frequencies since the transmitted data amount exceeds a upper limit of network bandwidth. For example, in a surveillance system that include 100 cameras with network bandwidth of 100Mbps, each camera send 30 fps of 640x480 image (approx. 5Mbps). The total bandwidth consumption is 500Mbps and this exceeds the network bandwidth. In this case, the maximum number of the cameras in that delay or data loss do not occur is 100Mbps/5Mbps=20. To reduce communication traffic, we have proposed a method in [1]. This method reduces the communication traffic by using progressive quality improvement approach. This paper, we implement a surveillance system by using that method. Some methods to improve communication performance have been proposed in [2], [3] and [4]. These methods have a drawback that the quality of images decrease such as image resolutions and bit-rates reduction.

This paper, we implemented a system to reduce communication traffic in the surveillance systems with the progressive quality improvement approach. In our implementation method, each recording computer can produce images that have some levels, the lowest level has lowest quality and the highest level has the highest quality data. Only the first level (lowest quality image) will be sent to the processing computers in the case that the processing computers no require the higher quality image data. The processing computers can progressively collect the remaining levels only in cases where higher quality images are needed such as humans or cars recorded in the image. Therefore, in our implementation system, the communication traffic and communication time can reduce since the transmitted data amount reduced.

The rest of this paper is organized as follows. In Section 2, we introduce some works that are related to our system. In Section 3, we briefly explain PQI approach. Our implementation is explained in Section 4, and evaluated in Section 5. Finally, we will conclude the paper in Section 6.

2. Related Work

Some methods to reduce communication traffic have been proposed. In [5], the authors suggested a method to improve communication traffic of surveillance systems by controlling the trans-

[1] Graduate School of Information Science and Technology, Osaka University
[2] Cybermedia Center, Osaka University
[3] Nara Institute of Science and Technology
[4] National Institute of Information and Communications Technology

mission packets using priority queuing based on I-Frame and P-Frame packets. However, the communication time in this method can not reduce since the data amount in their method always large. In our method, the lowest quality (it has lower data amount, the communication time is shorter) is sent as the first priority and the remaining quality data can be skipped when the higher quality data are not needed.

Some methods to reduce network bandwidth consumption in the surveillance systems were proposed in [6], [7], [8] and [9]. Method in [6], compresses data before sending to the processing computer, it enables less transmitted data amount. In [7] and [8], video background and foreground subtraction are suggested to reduce the usage of network bandwidth. In [9], a multicast communication method that can send the data in compressed format to multiple clients. These method are similar with our method in the case that the transmitted data are reduced by compressing data in order to reduce network bandwidth consumption. However, these method need to compress and extract data before transmission, this causes further delays.

A real-time stream quality improvement was proposed in [10]. This method improves the transmission data by using cache table to temporarily store the received data and treat the missing data. This method can improve data loss rate, but the communication traffic cannot reduce since this method transmit original data between cameras and processing computers (large data amount to be transmitted). In our method, the data is divided into some levels and the processing computers progressively collect them. The highest quality data in our method is combination of all levels.

The method proposed in [11], reduces the communication traffic since the transmission data is encoded separately and scheduling according to priority. However, this method the quality data cannot improve. In our proposed method can reduce the communication time keeping the application performance.

3. PQI Approach

In this section, we explain our proposed method called PQI (Progressive Quality Improvement). Our method reduces the communication traffic by progressively collecting image data from the lowest quality to the highest quality.

3.1 Summary

In our proposed method, we assume that the cameras or recording computers can generate the recorded images in the progressive JPEG format, which contains multiple images with different qualities (called scans in progressive JPEG). The lowest quality image data (first scan) has the highest priority to be sent to the processing computers. First, the processing computers receive the lowest quality image data and analyze the difference from the previous frame. The processing computers skip collecting and analyzing the next lowest quality image data (next scan) when it is meaningless. Only in the cases where higher-quality image data are needed for analysis (the processing computers judge to request higher quality when objects are recorded in the image data), the processing computers progressively collect the higher quality image data from the cameras or recording computers. Otherwise, the processing computers receive only the lowest quality image data. Therefore, the communication traffic can be kept low if the progress to higher quality image data does not frequently occur.

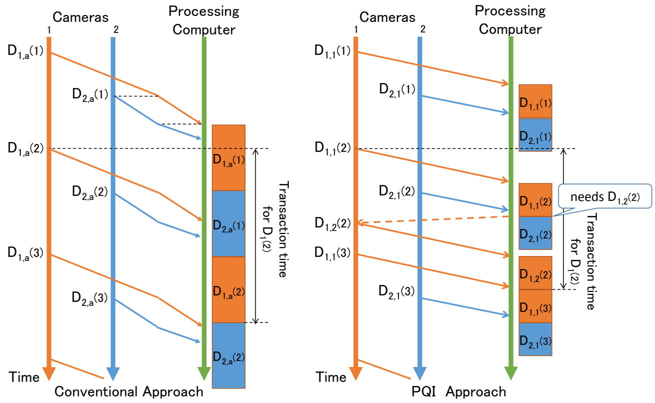

Fig. 1: Structure of our implemented system

3.2 Communication Chart

Figure 1 shows a timing chart for images processing under the conventional approach and the PQI approach. In the PQI approach, the image data $D_{n,a}(t)$ $(n = 1, 2, t = 1, \cdots)$ are divided into some qualities. Here, $D_{n,a}(t)$ is the observed original data of the camera n at the cycle t. $D_{n,q}(t)$ $(q = 1, \cdots, Q)$ is the generated data from $D_{n,a}(t)$ of that quality is Q. Each transaction includes some processes for each divided image data. In the figure, the number of the qualities 2 and the transaction consists of two processes for the divided image data $D_{n,1}(t)$ and $D_{n,2(t)}$. n is the camera number, t is the cycles for data collections, and q is the quality. In the cycle 1, the transaction finishes at the first quality in both cameras. In the cycle 2, the processing computer requires $D_{1,2}(2)$ when it finishes the process for $D_{1,1}(2)$. The camera 1 transmit the required $D_{1,2}(2)$ and the processing computer starts the process for $D_{1,2}(2)$. In this case, the transaction finishes when the processing computer finishes the process for $D_{1,2}(2)$ since the number of the qualities are 2. The transaction time in this case includes generation time, communication time and processing time. $GT_{n,q}(t)$ denotes the generation time of $D_{n,q}(t)$. $CT_{n,q}(t)$ is the communication time (the time start to send the image data from camera to the time to receive image data in the processing computers). $P_{n,q}(t)$ denotes the time required to process it. $TT_{n,q}(t)$ denotes the time to finish transaction. Here, $TT_{n,q}(t) = GT_{n,q}(t) + CT_{n,q}(t) + P_{n,q}(t)$. Therefore, the transaction time is reduced compared with that under the conventional approach as shown in the figure.

3.3 Data Generation

Figure 2 shows the flow chart of cameras. When the t th cycle starts, each camera n gets $D_{n,a}(t)$ from their local storages. First, they generate $D_{n,1}(t)$ from $D_{n,a}(t)$ and send $D_{n,1}(t)$ to the processing computer. When the camera n receives the request of $D_{n,q}(t)$, it generates $D_{n,q}(t)$ from stored $D_{n,a}(t)$ and sends $D_{n,q}(t)$ to the processing computer.

3.4 Data Processing

Figure 3 shows the flow chart of the processing computer.

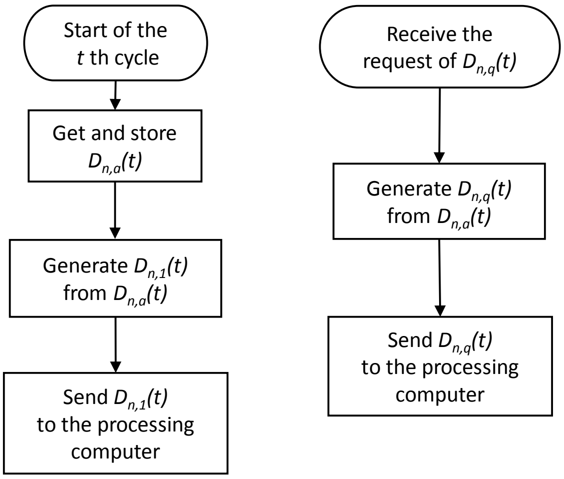

Fig. 2: Data sources generation

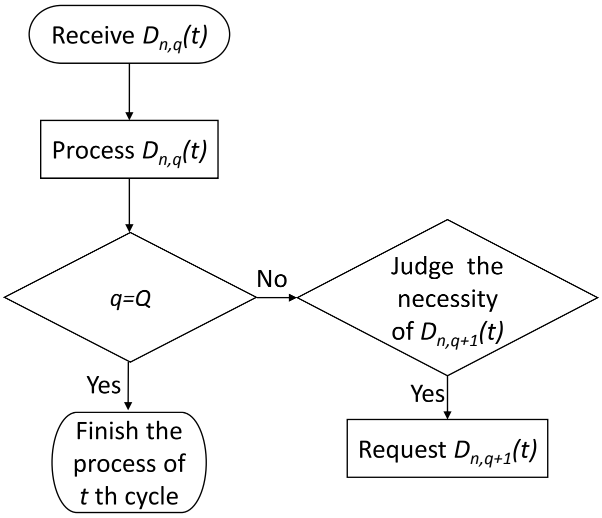

Fig. 3: Data processing in the processing computer

When the processing computer receives $D_{n,q}(t)$, it processes $D_{n,q}(t)$. When $q = Q$ and $D_{n,q}(t)$ is the final quality data, the process of t th cycle finishes. Otherwise, the processing computer judges the necessity of $D_{n,q+1}(t)$. In case that $D_{n,q+1}(t)$ is needed for the process execution, the processing computer requets $D_{n,q}(t)$ to the data source n.

4. Implementation

This section, we describe our implementation for PQI approach.

4.1 System Architecture

Figure 4 shows our implemented system. The recording computers are based on the Raspberry Pi 3 Model B with Camera Module V2 and connect to a processing computer via 100BASE-TX/1000BASE-T network (Allied Telesis CentreCOM GS908GT switch). We used the Python programming language and implemented as a human body detection system. Some progressive JPEG images are prepared beforehand and stored in recording computers' storages. In this case, we use image dataset named "ETH-Bahnhof" in ChangeDetection.Net 2014 Datasets

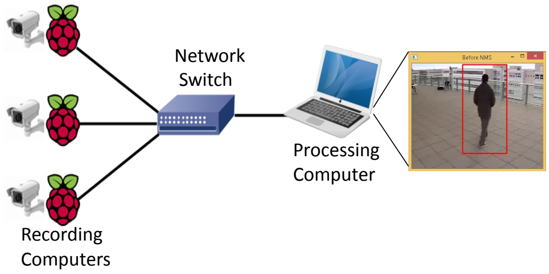

Fig. 4: Structure of our implemented system

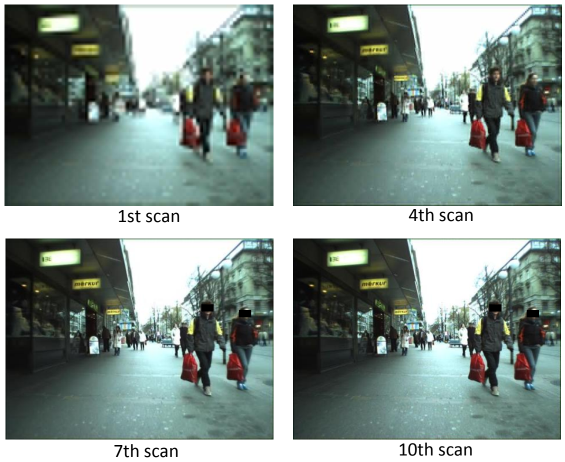

Fig. 5: A example for some scans of progressive-JPG images

[12]. That dataset includes 1000 image data of stand JPEG type with 640x480 resolution. These images are the frames of a video data got from surveillance camera. To extract some qualities data from original image data, we converted them into progressive-JPG type. Each progressive JPG image includes 10 scans (qualities data). In that 10 scans, have different qualities as shown in the example Figure 5. The generated scans are temporarily stored into each storage within the recording computer. Firstly, each recording computer send the lowest scans to the processing computer than other scans. The processing computer tries to detect human bodies from the received scans. If the processing computer detects human bodies in the firstly received scan, the processing computer progressively collects higher-quality scans from the recording computer. Otherwise, the processing computer skips receiving higher-quality scans. Table 1 shows the specifications of our implemented system.

Table 1: Specifications of our implemented system

Items	Details
Recording computer	Raspberry Pi 3 Model B (1.2GHz quad-core ARM Cortex-A53, 1GB memory, 100BASE-TX, Raspbian Ver 9.3. Nov. 2018)
Camera device	Raspberry Pi Camera Module V2
Processing computer (Windows 10 Pro 64bit)	Dell Latitude E7240 (2.10GHz dual-core Intel Core i7-4600U, 8GB memory, Intel HD Graphs 4600, 1000BASE-T
Network	Allied Telesis CentreCOM GS908GT (100BASE-TX/1000BASE-T), CAT5e cables
Cables	CAT5e, 1m length
Num. of recording computers	3
Num. of processing computers	1
Num. of qualities in prog. JPEG	10
Comparison methods	Without PQI approach
Evaluation items	Communication time

4.2 Communication Protocol

In this implementation, RTP/RTCP (Real-time Transport

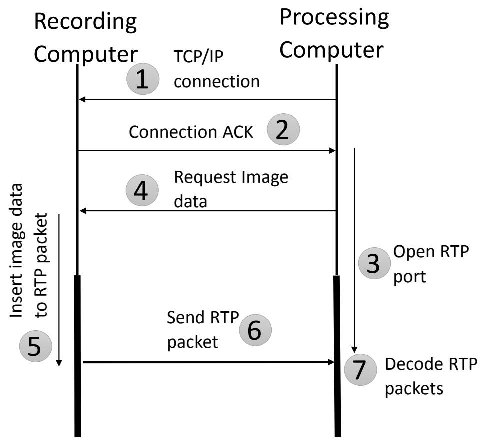

Fig. 6: System communication

Protocol/Real-time Control Protocol) are used for sending image data and controlling communication between recording computers and processing computer. Figure 6 shows communication of our implemented system. In our system, two channels are used, communication channel and data channel. First step, the processing computer connects to the recording computer via a TCP/IP communication (TCP socket). The recording computer confirms connection to the processing computer in second step. If the connection completed, the processing computer opens RTP channel and waiting for receiving RTP packets in the third step. If the connection confirmation is established, the processing computer requests for image data is started in step fourth. In step fifth, when the recording computer receives requests for image data, it encodes each image data into RTP packets and sends to the processing computer via data channel in step sixth. Finally, the processing computer receives RTP packet and decodes for getting image data. When the higher qualities data are needed the processing computer sends requests to the recording computers for higher quality image data via communication channel (How to request a higher image quality data is explained in Section 4.3).

4.3 How to Request Data

The progressive JPG images contain some scans (quality image data). In our implementation, they contain 10 scans every.

Figure 7 shows how to request a higher quality image from the processing computer. In this figure, the first scan (the lowest quality image) of an progressive JPG image is sent to the processing computer, and then the processing computer checks whether the hum bodies are recorded in that the lowest image or not. The processing computer skips collecting the remaining quality image 2nd to 10th when the 1st scan has no human bodies recorded. The higher remaining quality image will be progressively collected only when the processing computer found the human bodies are recorded in the lowest quality image. The highest quality image that the processing computer received is the combination of first sent scan and the remaining scans.

5. Evaluation Results

5.1 Evaluation Setup

Some parameters used in this implementation are show in Table 2.

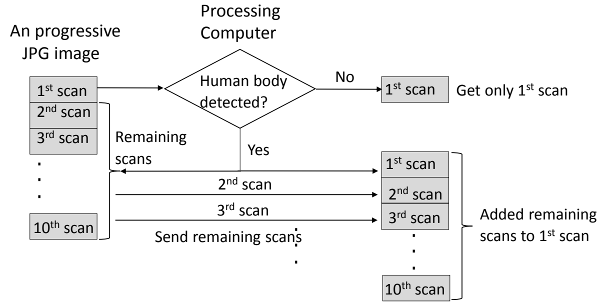

Fig. 7: How to request a higher quality image from the processing computer

Table 2: Parameter values

Communication Bandwidth	approx.100 [Mbps]
Communication Protocol	RTP over TCP
1000 frames Data	changedetection.net
Image Resolution	640x480
Progressive JPG	10 scans
Frame Rate	30 fps
Image Analysis	full body detection

5.2 Communication Time

We investigate the communication traffic under PQI approach and that under the conventional approach. We measure the communication time by sending some images from the recording computers to the processing computers. Sending a larger data amount to the processing computer causes longer communication since the network bandwidth has a upper limit. In this section, we show the communication time $CT_{n,q}$ by calculating the time to start sending from recording computer n to the time to receive image quality q (some low quality image data) in the processing computer. In conventional, the recording computers send original image quality Q.

Figure 8 shows communication time for one-to-one (only one recording computer on the network) communication of a recording computer and a processing computer. To reduce the influences of the initial stages, we take out 30 frames and measure the communication time. The horizontal axis is the number of images' frame data [num.] and the vertical axis is the communication time [sec.]. We can see that the communication time under our PQI approach is shorter than that under that the conventional approach in the case that no higher quality image data are needed since the transmitted data amount under PQI is smaller than in that under the conventional approach. The communication time in the frame number of 24 and 25 of PQI approach is longer than the communication time in conventional approach because the processing computer needs higher quality image data for that image has human bodies recorded.

Figure 9 shows the communication time increasing the number of recording computers. We show one of communication time under running three recording computers. The image resolutions and their frame rates are the same. In this figure, we use three recording computers run together in the network and send their

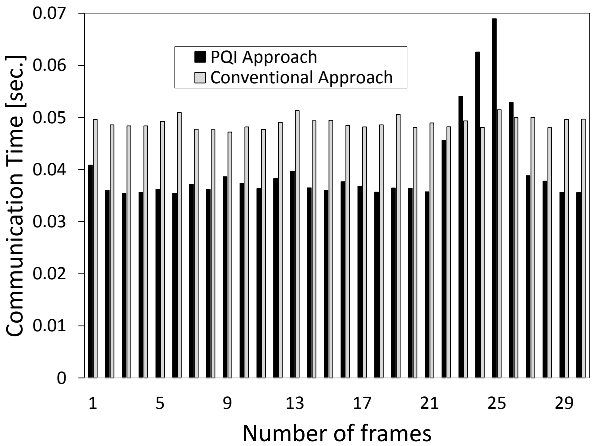

Fig. 8: The communication time between one recording computer and one processing computer

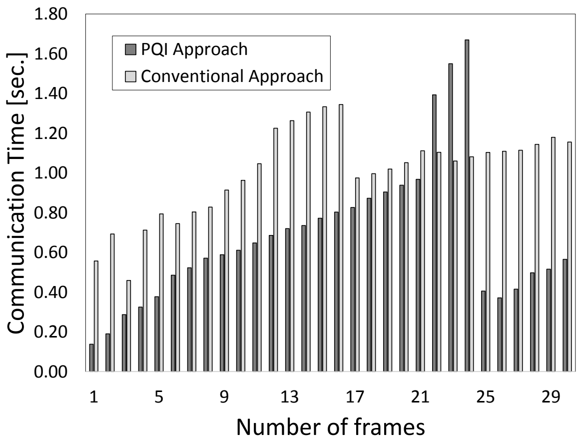

Fig. 9: The communication time between three recording computers and one processing computer

own images to the processing computer. The number of sending images, resolutions and frame rate are the same for all of them. The horizontal axis is the number of images' frame data [num.] and the vertical axis is the communication time [sec.]. We can see that the communication time in our propose method is shorter compared with that under the communication time of conventional approach. The shorter communication time at the beginning because the network bandwidth has more capacity for communication and the network bandwidth after shared to all of three recording computers the communication time increases since the network bandwidth reduces. The communication time shapely reduces in the frame number of 25 of PQI approach since sometimes all of frames are sent in lowest qualities.

5.3 Intervals Evaluation

A large transmitted data causes the number of data received reduces since the communication traffic suppress its sending frequencies. We check the data interval in the processing computer.

Figure 10 shows the data interval in PQI and conventional approach. The horizontal axis is the number of received images' frame data [num.] and the vertical axis is the time to receive the next image frame data [sec.]. The image resolutions and their frame rates are the same. In this cases, we can see that almost

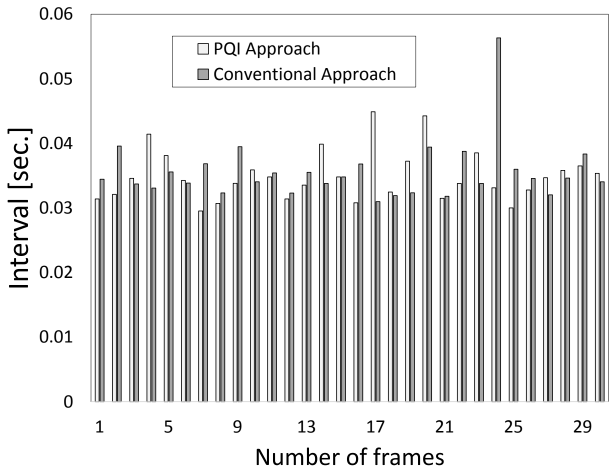

Fig. 10: The intervals for image data collection under one recording computer and one processing computer

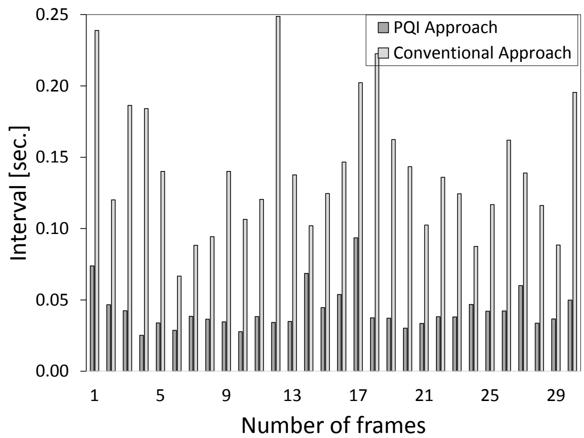

Fig. 11: The intervals for image data collection under three recording computers and one processing computer

of received image's frame data are similar in both of PQI and conventional approach since the network bandwidth has enough capacity for their transmission data.

Figure 11 shows one of the data interval in PQI and conventional approach under running three recording computers. We set the image resolutions and frame rate are the same. The horizontal axis is the number of received images' frame data [num.] and the vertical axis is the time to receive the next image frame data [sec.]. The data interval in PQI approach is faster than that under conventional approach since sometime the communication time of PQI approach is shorter than the under conventional approach. Therefore, in PQI approach can collect more data.

5.4 Detection Accuracy Evaluation

A low quality image decreases detection accuracy. We check whether human body detected in original images are detected in lowest quality in PQI approach or not. The position of human bodies often move from the position in original image quality. Therefore, we give an error margin. When the accepted error margin is e, the position (x, y) and size of detected of human bodies in original image quality is ex, ey and wxh, respectively. The error margin is $\frac{ew}{100}$% and $\frac{eh}{100}$%. We judge the correct detected

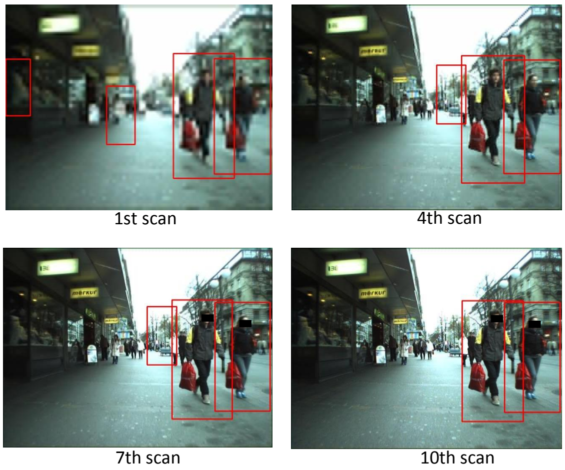

Fig. 12: The accuracy by changing qualities

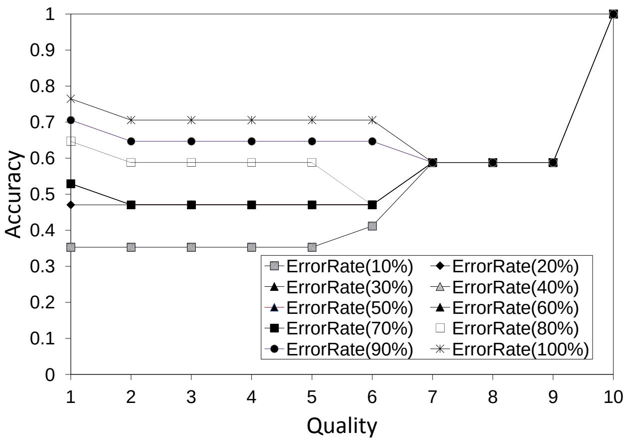

Fig. 13: DetectionAccuracy

human bodies when the position of detected hum body (x', y') in in lowest quality image satisfies $x - \frac{ew}{100} < x' < x + \frac{ew}{100}$ and $y - \frac{eh}{100} < y' < y + \frac{eh}{100}$. We use HOGDescriptor Algorithm in OpenCV for human bodies detection. It's a library for the computer vision field. In Figure 12, shows the example of human bodies detection. The detection position in the lowest quality image often move from the position of original image quality (the highest quality).

Figure 13 shows the detection accuracy of PQI approach changing the qualities. The accuracies are the number of correctly detected human bodies in the lower qualities divided by the number of the human bodies detected in original image quality. The horizontal axis is the number of qualities, from 1 to 10. and the vertical axis is the accuracy. We shows the accuracies under different accepted error margin rates. We see that the accuracies in the lowest quality increase as given error margin increase since a larger error margin rate can satisfy the moved position of human bodies from original increases. The accuracies reduce in the quality number 7, 8 and 9 because OpenCV sometimes detected a similar to human bodies in the lowest quality increases compare with the highest quality as shown in 12. In the lowest quality image detected 4 regions while the 7th quality image detects only 3. Therefore, in PQI approach, we can adjust the acceptance error margins for detecting human bodies in the lowest quality images.

6. Conclusion

In this paper, we implemented a video surveillance system with progressive quality improvement. In our implementation, there are three Raspberry Pi as recording computer and a laptop acts a processing computer. The results show that our proposed method can reduce the communication traffic when the progress to higher quality image data does not frequently occur and keeping the detection quality as an actual surveillance system. We have evaluated the detection accuracies under our PQI approach. The processing computer can detect human bodies in the lowest quality image with changing an error margin. In the future, we plan to implement a system to improve the transaction rates by using progressive quality improvement approach.

Acknowledgement

This research was supported by a Grants-in-Aid for Scientific Research(C) numbered JP17K00146 and JP18K11316, and by I-O DATA Foundation.

References

[1] C. Yukonhiatou, T. Yoshihisa, Y. Teranishi, Y. Ishi, T. Kawakami, and S. Shimojo, A Scheme to Improve Stream Data Analysis Frequency for Real-time IoT Applications, Multimedia, Distributed, Cooperative and Mobile (DICOMO 2018) Symposium, pp. 1205-1211, July, 2018.

[2] J. Xu, Y. Andrepoulos, Y. Xiao and M. van der Schaar, "Non-Stationary Resource Allocation Policies for Delay-Constrained Video Streaming: Application to Video over Internet-of-Things-Enabled Networks," in IEEE Journal on Selected Areas in Communications, vol. 32, no. 4, pp. 782-794, April 2014.

[3] P. Zhao, W. Yu, X. Yang, D. Meng and L. Wang,"Buffer Data-Driven Adaptation of Mobile Video Streaming Over Heterogeneous Wireless Networks," in IEEE Internet of Things Journal, vol. 5, no. 5, pp. 3430-3441, Oct. 2018.

[4] S. Molina-Giraldo et al.,"People detection in video streams using background subtraction and spatial-based scene modeling," 2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA), Bogota, pp. 1-6, Nov. 2015.

[5] N. Abbas and F. Yu, "Design and Implementation of a Video Surveillance System for Linear Wireless Multimedia Sensor Networks," 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, pp. 524-527, Oct. 2018.

[6] U. A. Agrawal and P. V. Jani,"Performance analysis of real time object tracking system based on compressive sensing," 2017 4th International Conference on Signal Processing, Computing and Control (ISPCC), Solan, pp. 187-193,Jan. 2017.

[7] X. Lu, C. Ye, J. Yu and Y. Zhang, "A Real-Time Distributed Intelligent Traffic Video-Surveillance System on Embedded Smart Cameras," 2013 Fourth International Conference on Networking and Distributed Computing, Los Angeles, CA, pp. 51-55, Oct. 2013.

[8] L. M. Brown, Max Lu, Chiao-Fe Shu, Ying-li Tian and A. Hampapur, "Improving Performance via Post Track Analysis," 2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, Beijing, 2005, pp. 341-347.

[9] Chia-Hui Wang, Ray-I Chang and Jan-Ming Ho, "An effective communication model for collaborative commerce of Web-based surveillance services," EEE International Conference on E-Commerce, 2003. CEC 2003., Newport Beach, CA, USA, pp. 40-44, July 2003.

[10] Q. Sun, W. Qi and J. Zhang, "Quality Improvement of Real-Time Video in Mobile Network," 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Exeter, United Kingdom, pp. 1039-1042, June 2018.

[11] K. Lee and S. Lee, "Perceptual priority based packet scheduling for quality constrained low delay live surveillance system," 2011 IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS), Seoul, pp. 1-4, Sep. 2011.

[12] Image dataset for human detections "pedestrians" in ChangeDetection.Net 2014. http://www.changedetection.net. [Accessed: 29-May-2018].