Performance Analysis of Surveillance Camera System with Image Recognition Server

Goo Yeon Lee
leegyeon@kangwon.ac.kr
Dept. of Computer and Communication Eng.
Kangwon National University
Chuncheon-si, Gangwon-do, Korea

Hwa Jong Kim
hjkim@kangwon.ac.kr
Dept. of Computer and Communication Eng.
Kangwon National University
Chuncheon-si, Gangwon-do, Korea

***Abstract*—In this paper, we study a performance analysis of a surveillance camera system with image recognition server based on frame discard rate and server utilization. The states of surveillance cameras are divided into recognition and silence states to analyze the parameters, such as the optimal number of image frames and cameras based on processing capacity of the recognition server. The result of the analysis will be useful for effective operation of the evolving surveillance camera systems.**

Keywords-surveillance camera; image recognition; network

I. INTRODUCTION

The image recognition server located in the central surveillance system can perform well when receiving as many image frames as possible from the surveillance cameras. However, the surveillance system operator may not allow frames to be transmitted continuously because it must pay a lot to the network depending on the number of transmitted image frames. Therefore, in this paper, we consider a system where the image recognition server analyzes the received image frames from surveillance cameras and determine whether there is an object of interest, and accordingly sets the state of each surveillance camera to a *recognition state* for transmitting a large number of frames and a *silence state* for transmitting a small number of frames. In such a surveillance camera network system, the system operator also needs to optimize the number of cameras arranged in accordance with the processing capacity of the image recognition server.

Many researches have been conducted on the number of transmission frames, network bandwidth and security method in surveillance camera systems. In [1], an optimal scheduling has been studied for storing the camera image data based on the network bandwidth and limited processing resources in the CCTV(Closed-Circuit TeleVision) system. In [2], the authors proposed a method to provide streaming service over HTTP (HyperText Transfer Protocol) by interworking with user and server. In this method, the user can predict the next segment resolution by providing current channel resource information.

In this paper, we perform a numerical analysis and discuss results in Section 2 and make a conclusion in Section 3.

II. ANALYTICAL MODEL AND NUMERICAL RESULTS

In the analysis, we assume that *N* surveillance cameras are connected to the image recognition server through the surveillance network. The image recognition server analyzes the arrived image frame and sets the camera to a recognition state with frame rate R_{rec} when the object of interest is recognized, otherwise sets it to a silence state with frame rate R_{sil} ($R_{rec} > R_{sil}$). Figure 1 shows the transition between the two states in surveillance cameras where the recognition and silence states of each camera alternate, and we assume that this transition process is independent of other cameras. This assumption is not appropriate when the surveillance cameras are concentrated at a short distance. However, in this study, we assume that the surveillance cameras are scattered in a large area. Also, for the sake of analysis, it is assumed that the state characteristics of all cameras are the same.

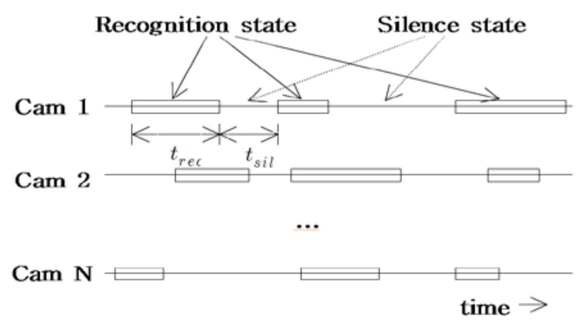

Figure 1. Transitions between recognition and silence states of surveillance cameras.

In Figure 1, we represent the recognition state interval length as a r.v. (random variable) t_{rec} and the silence state interval length as a r.v. t_{sil}. We assume that the number of frames processed per second in the image recognition server (processing capacity) is *C*. Generally, the surveillance system is effective in real-time monitoring, therefore we assume that the image recognition server has a small waiting queue to guarantee real-time performance and we also assume that for the sake of analysis, any frames received beyond the server capacity *C* in seconds are discarded.

Based on the above assumptions, the probability of a surveillance camera being in the recognition state and in the silence state are given as $P_{rec} = \frac{E(t_{rec})}{E(t_{rec})+E(t_{sil})}$ and $P_{sil} = 1 - P_{rec}$, where $E(\cdot)$ is the expected value. Also, the probability that *k* cameras are in the recognition state can be calculated as $P_k = \binom{N}{k}(P_{rec})^k(1-P_{rec})^{N-k}$, ($0 \le k \le N$). In terms of reasonable operation, the minimum total number of transmitted frames per second should be equal to or less than the processing capacity *C*. Also, processing capacity *C* should be equal to or less than the maximum total number of transmitted frames per second. That is $N \cdot R_{sil} \le C \le N \cdot R_{rec}$.

 ISBN: 978-1-61208-721-4

The average total number of frames transmitted per second is $N{\cdot}P_{rec}{\cdot}R_{rec} + N{\cdot}(1 - P_{rec}){\cdot}R_{sil}$. If the server processing capacity C is designed to be equal to $N{\cdot}R_{rec}$, there will be no frames that are not processed by the server and discarded. However, in this case, the waste of processing capacity of the server is on average $C - N{\cdot}P_{rec}{\cdot}R_{rec} - N{\cdot}(1 - P_{rec}){\cdot}R_{sil}$ frames per second. Therefore, it is necessary to increase the utilization of the server by allowing some frames to be discarded. Thus, we analyze the relationship between the processing capacity of the server, the number of discarded frames, and the number of cameras. Given the server capacity C, the corresponding number of cameras m can be calculated using $C{=}m{\cdot}R_{rec}{+}(N - m){\cdot}R_{sil}$ where m is an integer that is equal to $m = \left\lfloor \frac{C - N\cdot R_{sil}}{R_{rec} - R_{sil}} \right\rfloor$. Now the expected average number of frames that are discarded without processing at the server is:

$$F_{discarded} = \sum_{k=m+1}^{N} P_k \cdot [k \cdot R_{rec} + (N - k) \cdot R_{sil} - C] \quad (1)$$

In this study, two performance measures are considered. The first is the ratio of the number of frames that are discarded at the server to the server's processing capacity that we call Discarded Ratio (*DR*). The discarded frames consume network resources without being processed at the server. So the larger the number, the greater the inefficiency. The second is the utilization of the server, ρ, which is expressed as a ratio of the number of frames actually processed by the server to the capacity of the server. High utilization means the high operation efficiency of the server.

DR is calculated as $\frac{F_{discarded}}{C}$, and we have

$$DR = \frac{1}{C} \cdot \sum_{k=m+1}^{N} P_k \cdot [k \cdot R_{rec} + (N - k) \cdot R_{sil} - C] \quad (2)$$
$$= \sum_{k=m+1}^{N} \binom{N}{k} (P_{rec})^k (1 - P_{rec})^{N-k} \left[\frac{k \cdot R_{rec} + (N-k) \cdot R_{sil}}{C} - 1\right].$$

Then ρ can be calculated as

$$\rho = \frac{1}{C} \cdot [\sum_{k=0}^{m} P_k \cdot [k \cdot R_{rec} + (N - k) \cdot R_{sil}] + \sum_{k=m+1}^{N} P_k \cdot C]$$
$$= 1 - \sum_{k=0}^{m} \binom{N}{k} (P_{rec})^k (1 - P_{rec})^{N-k} \left[1 - \frac{k \cdot R_{rec} + (N-k) \cdot R_{sil}}{C}\right]. \quad (3)$$

Figures 2 and 3 are numerical results of the analysis when C=100, R_{rec} =5, R_{sil} = 1 and P_{rec} = 0.1, 0.2, 0.3, 0.4, 0.5.

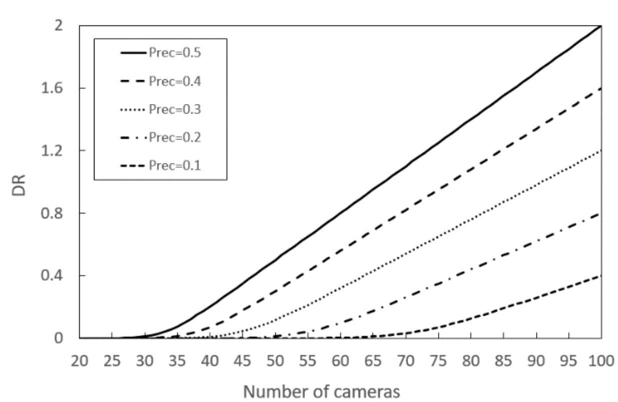

Figure 2. Relation between *N* and *DR* when *C*=100, R_{rec} =5, R_{sil} = 1, P_{rec} = 0.1, 0.2, 0.3, 0.4, 0.5.

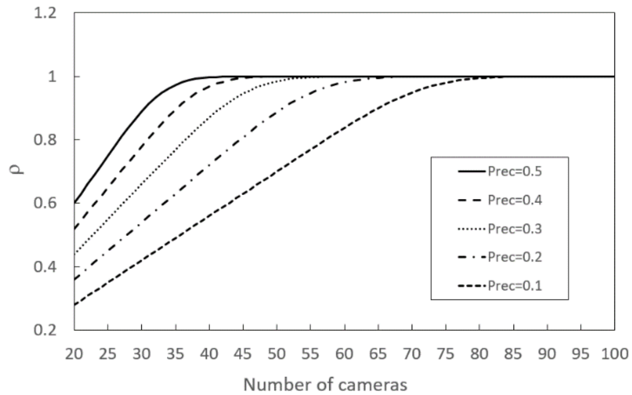

Figure 3. Relation between *N* and ρ when *C*=100, R_{rec} =5, R_{sil} = 1, P_{rec} = 0.1, 0.2, 0.3, 0.4, 0.5.

Figure 2 shows the relationship between *DR* and *N*, and Figure 3 shows the relationship between ρ and *N*. As shown in Figures 2 and 3, when designing *DR* = 0 which means no frames to be discarded at the server, the number of surveillance cameras is limited to *N* = 20. However, in this case, the utilization, ρ, is 28%, 36%, 44% 52%, 60% for P_{rec} = 0.1, 0.2, 0.3, 0.4, 0.5 respectively, which means the server efficiency is not high as one expects. To increase the server efficiency, we need to allow some value of *DR*. For example, when *DR*=0.01 (1%), and P_{rec}= 0.1, 0.2, 0.3, 0.4, 0.5, the number of cameras can be increased to *N*= 64, 49, 40, 34, 29, respectively. In this case, the server utilization is 88.8%, 87.3%, 87.1%, 87.4%, 86.4% respectively. Hence, as we allow the number of unprocessed and thus discarded frames at the server to be 1% of the server processing capacity, the utilization of the server rapidly increases to 86% through 89%.

III. Conclusion

In this study, we analyzed two performance factors that are required for an efficient surveillance camera system with image recognition server, i.e., server utilization and frame discard rate. We believe the results obtained in this study will be useful for the efficient design and operation of surveillance camera system, which has been widely used recently.

Acknowledgment

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No: 2018-0-00261, GDPR Compliant Personal Identifiable Information Management Technology for IoT Environment)

References

[1] K. J. Lin, T. K. Hou, and R. J. Chiu, "Jitter-Constrained Camera Scheduling in CCTV Surveillance Networks", 2016 IEEE International Conference on Signal and Image Processing(ICSIP), pp. 650-654, Aug. 2016.

[2] E. Khorov, A. Krasilov, M. Liubogoshchev, and S. Tang, "SEBRA: SAND-Enabled Bitrate and Resource Allocation algorithm for network assisted video streaming", in Proc. of IEEE WiMob, Rome, Italy, pp. 1-8, Nov. 2017.