

실시간 화상 인식 서버 기반 감시 카메라 시스템의 최적 구성 연구

이 구 연[°], 김 화 중^{*}

Optimum Configuration of Surveillance Camera System Based on Real Time Image Recognition Server

Goo Yeon Lee[°], Hwa Jong Kim^{*}

요 약

본 논문에서는 화상인식서버 기반의 감시 카메라 네트워크 시스템에서 실시간성을 보장하기 위한 최적의 구성 방법을 연구한다. 화상인식이 이루어져야할 실시간 기준 시간에 대하여 이를 위반하는 화상 프레임의 비율을 구함으로써 서버의 처리용량을 기준으로 한 최적의 화상 프레임 수 및 수용 가능한 카메라 수 등의 파라미터에 대한 분석을 수행한다. 분석된 결과는 화상 감시 카메라 네트워크의 효율적인 운영을 위하여 유용하게 활용될 것으로 판단된다.

Key Words : surveillance, camera, delay, image recognition, network

ABSTRACT

In this paper, we perform an optimum configuration study to guarantee the real time performance in the surveillance camera network system based on the image recognition server. By obtaining the ratio of image frames that violate the real-time reference time for image recognition, we analyze the parameters such as the optimum number of image frames and the corresponding number of cameras based on the processing capacity of the

server. The analyzed results will be useful for efficient operation of the evolving surveillance camera network systems.

I. 서 론

규모가 큰 감시 카메라 시스템은 설치되어 운영하는 카메라의 개수가 수십~수백개 또는 수천개까지 가능하다. 이런 경우 수많은 카메라로부터 전달되어 오는 모든 화상 프레임을 사람이 일일이 실시간으로 모니터링하는 것이 현실적으로 어려운 상황이다. 이에 대한 보완방법으로 전달되어 오는 화상 프레임에 대하여 화상인식 시스템을 설치하여 운영한다면 실시간 모니터링을 보다 효과적으로 수행할 수 있게 된다. 즉 수많은 감시카메라로부터 오는 화상 프레임에 대하여 화상인식시스템이 화상프레임내의 특이사항이 검출되었을 경우 실시간 알람 등으로 통지하게 됨으로써 모든 화상 프레임에 대하여 일일이 사람의 눈으로 확인하는 절차를 대신하게 된다. 그러나 화상인식 시스템에 도착하는 화상 프레임의 수가 너무 많으면 화상인식을 위해 입력 큐에 대기하는 시간이 길어지는 경우가 발생할 수 있다. 이 경우 화상 인식이 너무 늦어지면 감시 카메라 시스템의 실시간성 요구사항을 만족하지 못할 수 있다. 이와 같은 지연은 화상인식 시스템의 입력 큐에서 뿐만 아니라 화상카메라로부터 중앙 감시 시스템까지의 네트워크 전달시간도 영향을 미친다.

화상 프레임 인식의 실시간성을 가능한 보장하기 위해서는 위의 두 가지 지연 요소에 대한 최적화가 필요하다. 그러나 많은 경우 네트워크 전달 시간의 최적화는 이미 네트워크가 구성된 이후 또는 일반적인 best-effort 특성의 공중 인터넷을 사용하는 경우에는 현실적으로 가능하지 않게 된다. 따라서 본 연구에서는 네트워크를 통한 전달 지연이 있다는 전제하에, 화상 프레임 인식의 실시간성을 보장하기 위한 제어 방법으로 화상인식시스템의 입력 큐에서의 지연에 대하여 연구한다.

감시카메라 시스템 구성에 관한 연구는 주로 화상 인식의 정확성이나 또는 화상 프레임의 전송시 품질,

※ 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No. 2018-0-00261, IoT 환경에서 일반개인정보보호규정에 부합(GDPR Compliant)하는 개인정보 관리 기술 개발). 또한 이 논문은 농림축산식품부의 재원으로 농림수산식품 기술기획평가원(첨단생산기술개발사업) 지원사업의 연구결과로 수행되었음(116116-03-1SB010)

•° First and Corresponding Author : (ORCID:0000-0002-1769-6230)Kangwon National University, Dept. of Comp. and Comm. Eng., leegyeon@kangwon.ac.kr, 종신회원

* (ORCID:0000-0002-3822-390X)Kangwon National University, Dept. of Comp. and Comm. Eng., hjkim@kangwon.ac.kr, 종신회원
논문번호 : 201902-462-B-LU, Received February 12, 2019; Revised April 2, 2019; Accepted April 11, 2019

전송 프레임 수, 네트워크 대역폭, 보안측면에 관한 내용이 주로 연구되어 왔다¹⁻³⁾. [1]에서는 CCTV 시스템에서의 네트워크 대역폭 및 제한된 처리자원에 기반하여 어떤 카메라의 화상 데이터를 저장할 지에 대한 최적의 스케줄링에 대한 연구를 수행하였다. [2]에서는 사용자와 서버가 같이 연동하여 HTTP 상에서 동적으로 스트리밍을 제공하는 방식을 제안하였는데, 이 방식에서는 사용자에게 현재의 채널자원 정보를 제공함으로써 다음 세그먼트의 해상도 등을 예상하고 결정할 수 있게 하였다. [3]에서는 감시 카메라의 인식상태와 유희상태간의 전이특성을 기반으로 화상인식서버의 활용도 및 전달 프레임의 폐기율에 관한 연구를 수행하였다.

II. 분석 모델 및 결과

본 연구에서는 다수(N)의 감시 카메라가 네트워크를 통하여, 화상인식시스템이 구축되어있는 중앙 집중형 감시 시스템에 연결되어 있는 환경을 고려한다. 각 카메라는 화상프레임이 생성될 경우 네트워크를 통하여 중앙 감시센터로 전송한다. 감시 카메라 i ($1 \leq i \leq N$)는 평균 λ_i 개의 화상 프레임을 발생한다고 가정한다. 네트워크의 전달시간은 프레임간 독립적이라고 가정하며, 확률밀도함수가 $f_a(t)$ 인 r.v.(random variable) a 로 표시하며, 평균값을 \bar{a} 라고 표시한다. 중앙 감시 센터로 전달된 화상 프레임은 화상 인식 처리를 위한 입력 큐에 들어간다. 각 화상 프레임에 대한 화상 인식 시스템의 처리 시간은 프레임간에 독립적이라고 가정하며, 확률밀도함수가 $f_c(t)$ 인 r.v. c 로 표시하며 평균값을 \bar{c} 라고 표시한다. N 개의 화상 카메라로부터 전송되어 화상인식 시스템의 입력큐에도 착하는 화상프레임은 $\lambda = \sum_{i=1}^N \lambda_i$ 의 도착률을 갖는 포아슨 분포를 따른다고 가정한다. 이는 독립적으로 발생하는 트래픽인 경우 발생 형태가 어떤 형태라도 발생 소스가 많은 경우에는 그 합은 포아슨 분포에 맞게 되는 특성에 기인함으로, 본 연구에서 가정하고 있는 $N \gg 1$ 인 경우에 대하여서는 현실적으로 적용 가능한 타당한 가정이다. 화상 프레임이 화상 인식 시스템의 입력큐에서의 대기 시간을 확률밀도함수가 $f_b(t)$ 인 r.v. b 로 표시한다. 따라서 화상 프레임이 화상 카메라로부터 출발하여 화상 인식 시스템의 처리가 끝날 때까지의 시간 d 는 $d = a + b + c$ 로 나타낼 수 있다. d 의 확률 밀도 함수는 $f_d(t)$ 로 나타낸다. 화상

프레임의 실시간성 기준을 T 로 표시한다. 즉 $d > T$ 의 지연을 갖는 화상 프레임은 실시간성을 만족시키지 못한 프레임으로 간주하며, 전체 화상 프레임 대비 실시간성을 만족하지 못하는 화상 프레임의 비율을 실시간성 위반 비율 E 라고 표시한다. 실시간성 위반 비율 E 은 $E = \int_T^\infty f_d(t) dt$ 로 구해진다. 화상인식 서버의 입력 큐에서의 지연 b 에 대한 분포는 M/G/1 대기행렬 모델을 적용하여 구한다. $f_b(t)$, $f_c(t)$ 에 대한 LT(Laplace Transform)를 각각 $B^*(s)$, $C^*(s)$ 라고 하면, 서버의 가동률을 $\rho = \lambda \cdot \bar{c}$ 로 정의할 때 다음의 관계식을 얻을 수 있다.

$$B^*(s) = \int_0^\infty e^{-st} f_b(t) dt = \frac{s(1-\rho)}{s-\lambda+\lambda C^*(s)} \quad (1)$$

$f_d(t)$ 및 $f_a(t)$ 에 대한 LT를 각각 $D^*(s)$, $A^*(s)$ 라고 할 때 $D^*(s) = A^*(s)B^*(s)C^*(s)$ 의 관계가 성립한다. 본 절에서는 보편적인 경우로서 네트워크의 전달시간 및 화상인식시스템의 프레임 처리시간의 분포가 지수분포를 따르는 경우를 고려한다. 이 경우 다음의 관계식이 주어진다.

$$f_a(t) = \frac{1}{a} e^{-\frac{t}{a}} (t \geq 0), \quad A^*(s) = \frac{1}{as+1} \quad (2)$$

$$f_c(t) = \frac{1}{c} e^{-\frac{t}{c}} (t \geq 0), \quad C^*(s) = \frac{1}{cs+1} \quad (3)$$

위의 식으로부터 $B^*(s)$ 및 $D^*(s)$ 를 구할 수 있다.

$$B^*(s) = \frac{s(1-\rho)}{s-\lambda+\lambda/(\bar{c}s+1)} = \frac{(1-\rho)(\bar{c}s+1)}{\bar{c}s+1-\rho} \quad (4)$$

$$\begin{aligned} D^*(s) &= A^*(s)B^*(s)C^*(s) \\ &= \frac{1-\rho}{\bar{a}(1-\rho)-\bar{c}} \cdot \frac{1}{s+\frac{1}{\bar{a}}} - \frac{1-\rho}{\bar{a}(1-\rho)-\bar{c}} \cdot \frac{1}{s+\frac{1-\rho}{\bar{c}}} \end{aligned} \quad (5)$$

따라서 $f_d(t)$ 및 E 를 다음과 같이 구할 수 있다.

$$f_d(t) = \frac{1-\rho}{\bar{a}(1-\rho)-\bar{c}} \cdot \left[e^{-\frac{t}{\bar{a}}} - e^{-\frac{(1-\rho)t}{\bar{c}}} \right] (t \geq 0) \quad (6)$$

$$E = \frac{1-\rho}{\bar{a}(1-\rho) - \bar{c}} \cdot \left[\bar{a}e^{-\frac{T}{\bar{a}}} - \frac{\bar{c}}{1-\rho}e^{-\frac{(1-\rho)T}{\bar{c}}} \right] \quad (7)$$

그림 1은 $\bar{a}=0.1$, $\bar{c}=0.01$ 및 $T=400\text{ms}$, 600ms , 800ms , 1000ms 인 경우에 프레임 도착률에 따른 E 의 변화(선분으로 표시)를 나타낸다. $\bar{c}=0.01$ 인 경우는 화상 인식 서버의 평균적인 프레임 처리시간이 10ms로서 초당 평균 100개의 화상 프레임을 처리함을 나타내며, $\bar{a}=0.1$ 은 프레임의 네트워크 전달시간이 평균 100ms 임을 나타낸다. 실시간 판단 기준이 T 값이 400ms 인 경우는 민감한 경우로서 극도의 실시간성을 요구하는 경우이며, 반면 T 값이 1000ms인 경우는 1초간의 지연이 있는 경우로서 현실적으로 적용 가능한 값이 된다. 그림 1에서 볼 수 있듯이 $T=400\text{ms}$ 일 때의 도착률에 따른 실시간성 위반 비율 E 가 가장 높으며, 반면 T 가 커질수록 E 가 작게 관찰됨을 알 수 있다. 이는 실시간성 기준을 엄격하게(작게) 설정하면 이를 위반하여 도착하는 프레임의 비율이 커질 수밖에 없는 상식에 기인한다. 수식적인 분석 모델의 개선 방법으로써 실제 운영에서는 입력류로부터 프레임이 화상인식서버로 들어갈 때 이미 실시간성을 만족하지 못하는 것으로 판단되면(즉 이미 프레임의 지연이 T 보다 큰 경우), 이 프레임은 즉각적인 화상인식서버의 처리를 받지 않도록 할 수 있다. 이는 이미 실시간성을 만족하지 못하는 프레임이 서버의 처리를 받을 경우 그 프레임은 여전히 실시간성을 만족하지 못하나, 서버의 처리용량을 바로 사용하게 됨으로써 다른 프레임의 실시간성 만족 가능성을 작아지게 할 수 있기 때문이다. 이러한 프레임은 서버가 유휴상태가 될 때 처리를 받게 할 수 있다. 위와 같은 운영을 할 경우 실시간성 위반 비율 E 는 그렇지 않은 경우보다 향상될

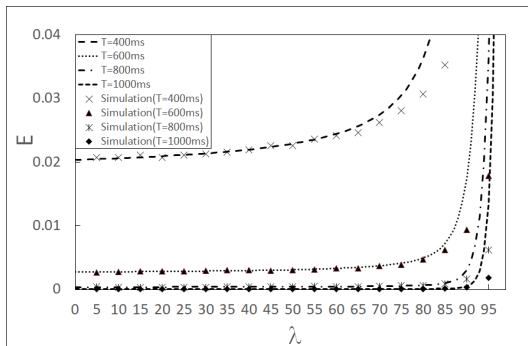


그림 1. 실시간성 기준 시간 및 프레임 도착률에 따른 E
Fig. 1. E according to real-time reference time and frame arrival rate

것으로 예상할 수 있으며, 모의 실험을 통한 그림 1의 결과(표식으로 표시)로부터 항상 정도를 알 수 있다.

화상 감시 시스템을 설계할 때 적용 도메인에 따라 실시간성에 기반하여 위반 비율을 어느 정도로 유지할 지에 대한 결정을 하게 된다. 이에 따라 총 도착률을 결정하고, 화상카메라의 수를 조정하게 된다. 분석 결과로부터 E 를 1%로 정하는 경우 수용가능한 총 도착률은 $T=400\text{ms}$ 때는 불가이며, $T=600\text{ms}$, 800ms , 1000ms 일 때 각각 87, 92, 94가 된다. $T=400\text{ms}$ 가 불가로 계산되는 이유는 네트워크를 통한 전달 지연이 400ms이상일 확률이 2%이기 때문이다. 모든 화상 카메라의 평균 프레임 발생 비율이 초당 0.5라는 가정하는 경우 즉 $\lambda_i = \lambda/N = 0.5$ ($1 \leq i \leq N$)의 경우, 수용 가능한 화상 카메라의 수는 $T=600\text{ms}$, 800ms , 1000ms 일 때 각각 174대, 184대, 188대로 계산된다. 만약 λ_i 가 i 에 따라 달라지는 경우에는 λ 에 대하여 $\lambda = \sum_{i=1}^N \lambda_i$ 인 N 로서 화상카메라의 수를 구하면 된다.

모의실험에서는 $T=600\text{ms}$, 800ms , 1000ms 일 때 수용 가능한 총 도착률이 각각 90, 97, 98이 되어 수식적인 분석모델보다 향상됨을 알 수 있다.

III. 결 론

본 연구에서는 다수의 감시 카메라가 연결된 화상 인식서버 기반의 실시간 감시 카메라 네트워크 시스템에서 실시간 기준 시간 대비 위반 화상프레임의 비율을 구함으로써 수용 가능한 최적의 화상프레임 수 및 카메라 수를 구하였다. 이렇게 구해진 분석 결과는 최근 광범위하게 보급되고, 또한 진화되고 있는 화상 감시 카메라 네트워크의 효율적인 운영을 위하여 유용하게 활용될 것으로 판단된다.

References

- [1] K. J. Lin, T. K. Hou, and R. J. Chiu, "Jitter-constrained camera scheduling in CCTV surveillance networks," *2016 IEEE Int. Conf. Sign. and Image Process.*, 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. IEEE*

WiMob, Rome, Italy, pp. 1-8, Nov. 2017.

- [3] Y. Shqairat and G. Y. Lee, "Performance analysis of surveillance camera system based on image recognition server," *J. KICS*, vol. 42 no. 4, pp. 816-818, Apr. 2017.