Regular paper | Journal of the KIECS. pp. 1263-1268, vol. 13, no. 6, Dec. 31 2018, t. 92, pISSN 1975-8170 | eISSN 2288-2189
http://dx.doi.org/10.13067/JKIECS.2018.13.6.1263

보안 감시를 위한 심층학습 기반 다채널 영상 분석

박장식* · 마르셀 위라네가라** · 손금영**

Multi-channel Video Analysis Based on Deep Learning for Video Surveillance

Jang-Sik Park* · Marshall Wiranegara** · Geum-Young Son**

요 약

본 논문에서는 영상 보안 감시를 위한 심층학습 객체 검출과 다중 객체 추적을 위한 확률적 데이터연관 필터를 연계한 영상분석 기법을 제안하고, GPU를 이용하여 구현하는 방안을 제시한다. 제안하는 영상분석 기법은 객체 검출과 추적으로 순차적으로 수행한다. 객체 검출을 위한 심층학습은 ResNet을 이용하고, 다중 객체 추적을 위하여 확률적 데이터 연관 필터를 적용한다. 제안하는 영상분석 기법은 임의의 영역으로 불법으로 침입하는 사람을 검출하거나 특정 공간에 출입하는 사람을 계수하는데 응용할 수 있다. 시뮬레이션을 통하여 약 25fps의 속도로 48채널의 영상을 분석할 수 있음을 보이고, RTSP 프로토콜을 통하여 실시간 영상분석이 가능함을 보인다.

ABSTRACT

In this paper, a video analysis is proposed to implement video surveillance system with deep learning object detection and probabilistic data association filter for tracking multiple objects, and suggests its implementation using GPU. The proposed video analysis technique involves object detection and object tracking sequentially. The deep learning network architecture uses ResNet for object detection and applies probabilistic data association filter for multiple objects tracking. The proposed video analysis technique can be used to detect intruders illegally trespassing any restricted area or to count the number of people entering a specified area. As a results of simulations and experiments, 48 channels of videos can be analyzed at a speed of about 27 fps and real-time video analysis is possible through RTSP protocol.

키워드

Video Surveillance, Object detection, Multi-object tracking, Deep learning, Probabilistic data association filter
영상 보안 감시, 객체 검출, 다중 객체 추적, 심층 학습, 확률적 데이터 연관 필터

Ⅰ. 서 론

발전소, 공항, 항만 등의 장소에서 테러 및 무단 침입 등에서 사고가 증가하고 있으며, 시민들의 안전과 보안을 위하여 많은 수의 CCTV 카메라가 설치되고 있다. 많은 수의 CCTV 카메라가 설치되고 있지만, 소수의 관제 인력이 모든 영상을 관제하기 어렵다. 시민의 안전과 주요 시설의 보안에 영향을 주는 상황을

* 교신저자 : 경성대학교 전자공학과(jsipark@ks.ac.kr)
** 경성대학교 전자공학과(marshalltata@naver.com, zhflgk135@ks.ac.kr)
· 접 수 일 : 2018. 11. 16
· 수정완료일 : 2018. 11. 30
· 게재확정일 : 2018. 12. 15
· Received : Nov. 16, 2018, Revised : Nov. 30, 2018, Accepted : Dec. 15, 2018
· Corresponding Author : Jang-Sik Park
Dept. Department of Electronic Engineering, Kyungsung University,
Email : jsipark@ks.ac.kr

자동으로 인지하는 영상분석 기술을 기반으로 하는 지능형 영상 보안 감시 시스템들이 제공되고 있다. 지능형 영상 보안 감시 시스템은 실시간으로 영상을 자동 분석하여 사람이나 물체의 특징을 인식하여 특이 행동을 자동으로 감지하여 사고 예방 또는 사후 처리에 효과적으로 활용할 수 있다.

지능형 영상 보안 감시 시스템을 구현하기 위해서는 사람, 차량 등의 객체를 검출하는 기술(Object Detection)과 특정 객체를 추적하는 기술(Object Tracking)이 필수적이다. 객체 검출은 고전적인 방법으로는 유사 Haar(Haar-like)[1], HOG(Histogram of Oriented Gradients)[2], LBP(Local Bit Pattern)[3] 등의 특징점을 이용하여 Adaboost[4]과 SVM(Support Vector Machine)[5] 등의 학습 알고리즘을 적용하는 방법이 널리 사용되었으며, GPU를 활용한 병렬 처리를 통하여 고속 검출에 대한 연구도 수행되었다[6]. 최근에는 컨벌루션 신경망(CNN, Convolutional Neural Network)[7]을 기반으로 하는 다양한 심층학습(Deep Learning) 모델이 제안되었다. 객체 검출 심층학습은 R-CNN[8], Fast R-CNN[9] 그리고 Faster R-CNN[10] 등이 제안되었으며, 임베디드 시스템 적용한 YOLO[11], SSD[12]와 같은 객체 검출 모델이 제안되기도 하였다. 일반적으로 심층학습의 계층이 깊어지면 성능이 우수해지지만, 너무 깊은 경우에는 오히려 성능이 저하되게 된다. 이러한 문제를 해결하기 위하여 ResNet[13] 이 제안되었다. 심층학습 구현을 위한 다양한 프레임워크가 개발되고 있다[14].

객체 추적 기술은 칼만 필터(Kalman Filter), 확장 칼만 필터(Extended Kalman Filter) 그리고 파티클 필터(Particle Filter) 등이 개발되었다[15]. 다중 객체 추적을 위하여 확률적 데이터 연관 필터(PDAF, Probabilistic Data Association Filter)가 제안되었다[16].

본 논문에서는 GPU를 이용하여 영상 보안 감시를 위하여 심층학습 기반의 객체 검출과 객체 추적을 연계하는 방안을 제안하고 구현 방법을 제시한다. 객체 검출은 ResNet을 사용하고, 검출된 객체를 추적하기 위하여 PDAF를 순차적으로 적용한다. 검출된 객체를 추적 위하여 확률적 데이터 연관 필터를 사용한다. PDAF는 칼만필 계열로써 검출된 객체의 위치, 이동 속도를 이용하여 다음 프레임의 객체 궤적과 연관될 확률을 계산하고, 객체의 위치를 예측하여 추적한다.

본 논문에서는 다중 채널에 대하여 영상분석을 위하여 엔비디아(NVIDIA)사의 GPU를 활용하여 RTSP로 전송 받은 H.264 영상에 대하여 복원, 객체 검출 추론과 추적 기능을 구현하였다. 48 채널 영상에 대하여 객체 검출 및 추적하는 약 27 fps의 속도로 처리하였으며, 4채널 실시간 영상분석이 가능함을 실험을 통하여 보인다.

II. 심층학습 객체 검출과 추적

영상 분석을 통하여 상황을 인식하거나 특정한 서비스를 하기 위해서는 객체를 검출하고, 추적하는 기능이 필요하다. 성능과 처리 속도 측면에서 우수한 ResNet 이 주로 활용되고 있다. 객체 추적 알고리즘으로 칼만 필터, 파티클 필터 등이 활용되고 있으며 다중 객체 추적을 위하여 PDAF가 효과적이다.

2.1 ResNet

심층학습의 망이 깊어지면, 기울기 소멸(Vanishing Gradient) 또는 포화(Exploding Gradient)와 같은 문제 때문에 학습의 성능이 저하되는 문제(Degradation Problem)가 발생한다[13].

그림 1의 (a)는 일반적인 학습망의 구조이며, (b)는 잔차 학습(Residual Learning)의 기본 구조이다. 일반적인 학습망에서 $H(x)$가 최적화되도록 학습을 수행한다. 잔차 학습은 개념을 달리하여 $H(x)-x$를 최적화하도록 한다. 여기서, $F(x)$를 식 (1)로 정의한다.

$$F(x)=H(x)-x \tag{1}$$

따라서, 그림 1의 (a)는 (b)로 변형이 된다. (b)와 같은 구조에서, $H(x)-x$의 최적화는 $F(x)$가 0으로 수렴하는 것이 되기 때문에 학습이 방향이 미리 결정된다. 이것은 사전 조정(Pre-conditioning)의 역할을 하게 된다[13]. $F(x)$가 거의 0이 되는 방향으로 학습하게 되고, 식 (1) 과 같이 출력과 입력의 차로 학습을 하기 때문에 잔차 학습 이라고 하고, 그림 1의 (b)의 구조를 기본적으로 사용하는 심층학습 모델을 ResNet 이라고 한다[13].

ResNet은 그림 1의 (b)와 같이 단순히 출력에 입력을 더하는 단축 연결(Shortcut Connection)이 추가 때문에 계수가 증가하지 않으며, 덧셈 계산만 추가된다. 단축 연결을 통하여 깊은 망에서도 성능 저하없이 최적화가 가능하고, 깊어진 망으로 정확도를 개선할 수 있다.

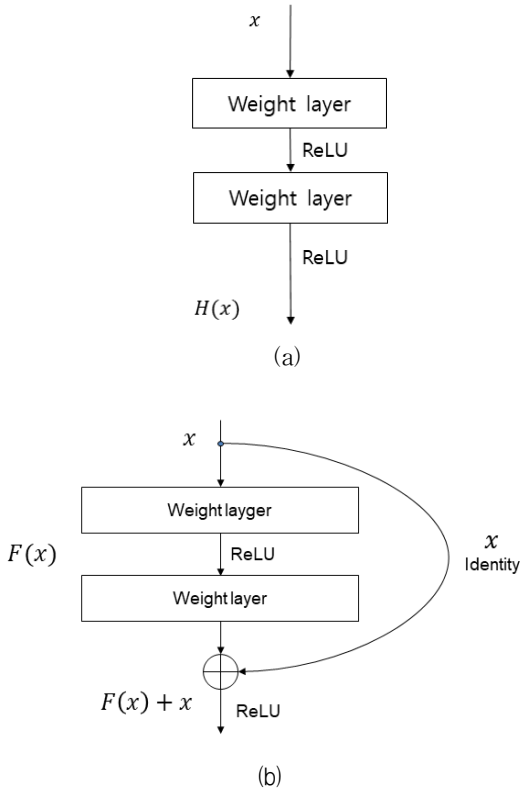

그림 1. 일반적인 학습과 잔차 학습 기본 구조
Fig. 1 Basic structure of general learning and residual learning

2.2 다중 객체 추적

PDAF는 칼만필터를 기반으로 검출된 객체의 위치, 이동 속도를 이용하여 다음 객체의 궤적과 연관될 확률을 계산하고, 객체의 위치를 예측한다[14]. 그림 2는 PDAF의 흐름도를 나타낸다. PDAF는 예측 단계와 유효한 측정값을 검증하는 단계, 데이터의 연관성을 확인하는 단계, 마지막으로 상태를 추정하는 단계로 구성된다.

추적 객체의 상태벡터(State Vector)는 식 (2)과 같이 정의되며, 이는 추적 객체의 중심 좌표 (p_x, p_y)와 속도(v_x, v_y)로 이루어져 있다. 식 (2)은 일정한 속도의 모델을 고려해야 한다. 속도가 일정하지 않을 경우 이산 시간 평균 0 (Discrete-time Zero-mean)인 백색 잡음(White Noise)으로 가정한다.

$$\mathrm{x} = [p_x \; v_x \; p_y \; v_y]^{\mathrm{T}} \tag{2}$$

PDAF는 데이터 잡음(Data Clutter)이 있더라도 추적이 가능하다. 주로 PDAF가 검증 데이터에 대한 의존도가 낮고, 상태 오차 공분산(State Error Covariance)의 증가로 인해 가능하다[16]

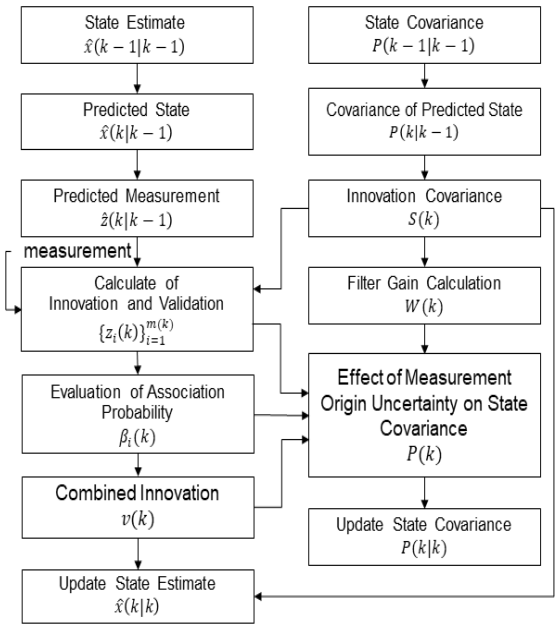

그림 2. PDAF 구조
Fig. 2 Structure of PDAF

III. 보안 감시 시스템 구현

본 논문에서 영상기반의 보안 감시 시스템 구현을 그림 4와 같이 CCTV 카메라로부터 RSTP 프로토콜로 영상 정보를 수신하여 압축된 영상을 복원하고, 객체 검출을 위하여 ResNet 심층학습을 수행하고,

PDAF 추적을 순차적으로 수행한다. 미디어 데이터의 연결(Pipe)은 gstreamer 를 활용한다.영상분석을 위하여 RTSP 프로토콜 수신 데이터에 대하여 nvvideocodec API를 이용하여 복원한다.

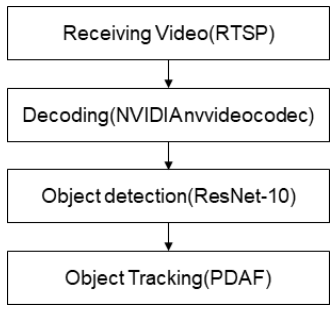

그림 3. 제안하는 영상보안 시스템 구성
Fig. 3 Configuration of the proposed video surveillance system

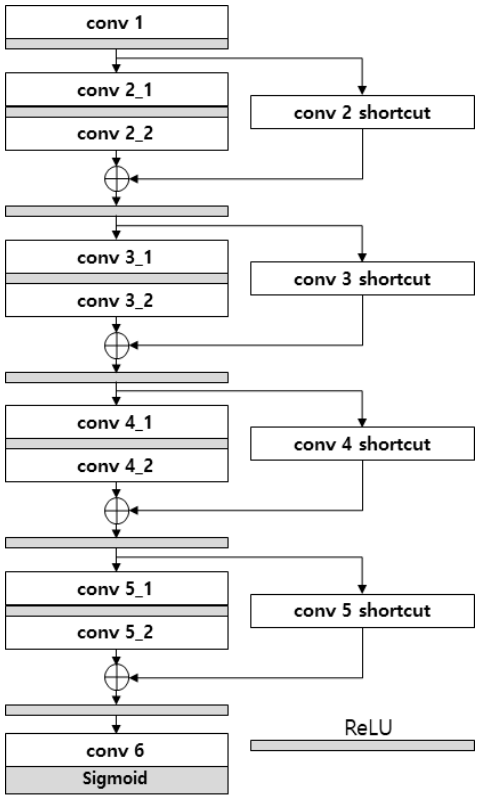

그림 4. ResNet-10 모델의 구조
Fig. 4 Architecture of ResNet-10 Model.

객체 검출을 위한 ResNet-10의 구조는 그림 4와 같다. 컨벌루션 계층 1과 6은 일반적인 컨벌루션을 수행하고, 컨벌루션 2, 3, 4 그리고 5는 ResNet 기본 블록으로 구성된다. 기본 블록은 2개의 컨벌루션 계층과 1개의 단축 컨벌루션 계층(Shortcut Layer)으로 구성되고, 일반 컨벌루션과 단축 컨벌루션 계층의 합에 대하여 ReLU 활성함수(Activation Function)를 적용한다.

검출된 객체에 대하여 PDAF를 통하여 추적을 적용한 결과는 그림 5와 같다. 검출된 객체의 중심을 추적하도록 설정하고, 12 객체를 추적할 수 있도록 설정한다.

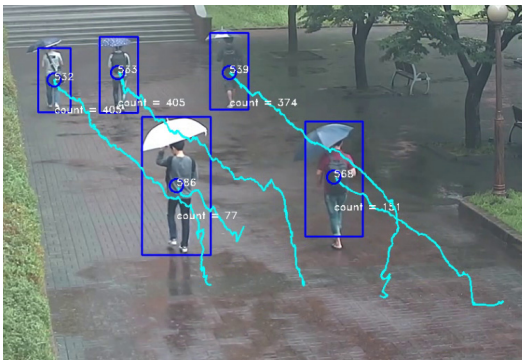

그림 5. PDAF에 의한 객체 추적 결과
Fig. 5 A result of object tracking with PDAF

Ⅳ. 시뮬레이션 결과 및 검토

저장 영상 및 실시간 영상분석 처리를 위하여 Intel Xeon E5-2650 CPU와 시스템 메모리 64GB를 갖춘 워크스테이션에 16G 메모리를 가진 Quadro P5000 그래픽카드를 실장한 워크스테이션을 사용한다. 개발 환경은 엔비디아 nvvideocodec 라이브러리, CUDA 9.2, TensorRT-4.0, OpenCV 3.4.1을 사용한다.

다채널 영상분석 성능을 확인하기 위하여 현장에서 녹화한 비디오를 44채널로 나누어 읽고, 4채널은 현장에 설치된 카메라로부터 RTSP 프로토콜로 수신한 영상에 대하여 객체 검출 및 추적을 수행하였다. 수행결과는 그림 6과 같이 다채널에 대하여 동시에 객체 검출 및 추적을 수행하는 것을 확인할 수 있다.

그림 6. 영상 분석 결과(4채널 RTSP전송 포함)
Fig. 6 A result of video analysis(including 4 channels RTSP video transmission)

그림 7은 실시간 영상분석 처리는 4채널만 나타낸 것으로 사람 또는 차량을 적절이 검출하고 추적하는 것을 확인할 수 있다.

그림 7. 실시간 영상분석 결과(4채널 RTSP전송)
Fig. 7 A result of real-time video analysis(4 channel RTSP video transmission)

표 1은 처리하는 채널 수에 대하여 처리 속도를 측정한 결과이다.

표 1. 채널 수에 따른 영상분석 평균 처리 속도
Table 1. Average speed of video analysis as the number of channels

No. of channels	Average frame per second
10	30.00
20	29.86
30	29.86
48	27.50

10 채널에 대해서 평균 30fps로 처리하고, 48채널에 대하여 평균 약 27fps 처리한다. 일반적인 영상통합관제센터에서 약 15fps 정도로 영상 관제를 하고 있어 GPU를 가진 그래픽 카드를 실장한 워크스테이션으로 다채널 영상분석이 가능함을 확인할 수 있다.

Ⅴ. 결 론

본 논문에서는 영상 보안 감시를 위하여 다중 객체 검출, 추적을 방법을 제안한다. 객체 검출을 위하여 ResNet를 적용하고 PDAF를 이용하여 다중 객체를 추적하는 방안을 제안한다. 48채널의 영상에 대하여 영상분석 결과 약 27fps를 처리할 수 있음을 확인하였다. 4채널의 카메라 영상에 대하여 실시간 처리할 수 있음을 보인다.

향후 학습 데이터셋을 보안하여 학습을 수행하고, 실시간 객체 검출 및 추적을 위한 GPU를 탑재한 임베디드 시스템으로 구현할 계획이다.

감사의 글

본 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구(No. B0717-16-0107, 국민참여형 사회안전서비스를 위한 영상 크라우드 소싱 핵심기술개발)와 BB21+ 사업의 성과임

References

[1] P. Viola and M. J. Jones, “Rapid object detection using a boosted cascade of simple features,” In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition,* Kauai, USA, Feb. 2001, pp. 511-518.

[2] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Diego, USA, June 2005, pp. 886-893.

[3] T. Ahonen, A. Hadid, and M. Pietikainen, “Face recognition with local binary patterns: application to face recognition,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, Dec. 2006, pp. 2037-2041.

[4] Y. Freund and R. E. Schapire. "Experiments with a new boosting algorithm in machine learning", In *Proc. of 13th Int. Conf. In Machine Learning*, San Francisco, USA, 1996, pp. 148-156.
[5] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, issue 3, Sept. 1995, pp. 273-297.
[6] J. Park, "Comparison speed of pedestrian detection with parallel processing graphic processor and general purpose processor," *J. of Korean Institute of Electronic Communication Society,* vol. 10, no. 2, 2015, pp.239-246.
[7] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation,* vol. 1, issue 4, 1989, pp. 541-551.
[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierachies for accurate object detection and semantic segmentation," In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Ohio, USA, June 2014, pp. 580-587.
[9] R. Girshick, "Fast R-CNN," In *Proc. IEEE Int. conf. on Computer Vision*, Santiago, Chile, 2015, pp. 1440-1448.
[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, 2017, pp. 1137-1149.
[11] J. Redmon, S. Divvala, R. Girshik, and A. Farhadi, "You only look once: unified, real-time object detection," In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, USA, June 2016, pp. 779-788.
[12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: single shot multibox detector," In *Proc. European Conf. on Computer Vision*, Amsterdam, Netherlands, Oct., 2016, pp. 21-37.
[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition,* Las Vegas, USA, June 2016, pp. 770-778.
[14] Y. S. Lee and P. J. Moon, "A comparison and analysis of deep learning framework," *J. of Korean Institute of Electronic Communication Society,* vol. 12, no. 1, 2017, pp.115-122.
[15] B. Choi, J. Park, J. Song, and B. Yoon, "Object detection and tracking with infrared videos at night-time," *J. of Korean Institute of Electronic Communication Society,* vol. 10, no. 2, 2015, pp.183-188..
[16] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Systems Magazine*, vol. 29, issue 6, 2009, pp. 82-100.

저자 소개

박장식(Jang-Sik Park)

1992년 부산대학교 전자공학과 졸업(학사)
1994년 부산대학교 대학원 전자공학과 졸업(석사)
1999년 부산대학교 대학원 전자공학과 졸업(박사)
1997년 ~ 2011년 동의과학대학교 전자학과 교수
2011년 ~ 현재 경성대학교 전자공학과 교수
2016년 ~ 현재 TTA PG-427 CCTV표준 의장
※ 관심분야 : 적응신호처리, 음성 및 음향신호처리, 컴퓨터비전, 딥러닝

Marshall Wiranegara

2017년 경성대학교 전자공학과 졸업(학사)
2017년 ~ 현재 재영소프트 연구원
※ 관심분야 : 컴퓨터비전, 딥러닝

손금영(Geum-Young Son)

2016년 동의대학교 멀티미디어공학과 졸업(학사)
2017년 ~ 현재 경성대학교 대학원 전기전자공학과 석사과정
※ 관심분야 : 영상처리, 컴퓨터비전, 딥러닝