The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)
Vol. 22, No. 4, pp.49-57, Aug. 31, 2022. pISSN 2289-0238, eISSN 2289-0246

https://doi.org/10.7236/JIIBC.2022.22.4.49
JIIBC 2022-4-8

공공 다중CCTV 기반에서 재식별 기술을 활용한 특정대상 탐지 및 추적기법 구현

Implementation of Specific Target Detection and Tracking Technique using Re-identification Technology based on public Multi-CCTV

황주성*, 뉴엔탄하이**, 강수경**, 김영규**, 김주용**, 정명석***, 이주연***

Joo-Sung Hwang*, Nguyen Thanh Hai**, Soo-Kyung Kang**, Young-Kyu Kim**, Joo-Yong Kim**, Myoung-Sug Chung***, Jooyeoun Lee***

요 약 정부에서는 전국에 설치된 공공 CCTV를 이용하여 실종아동 등 범죄 예방을 위하여 많은 노력을 하고 있다. 하지만, 운용인력의 부족과 장시간 집중에 따른 집중력 약화 그리고 추적의 어려움 등이 나타나고 있다. 또한, 딥러닝 알고리즘을 통하여 실시간 객체 탐색 및 재인식 그리고 추적을 적용하는 것은 복잡한 신경망 분석의 사유로 파라미터가 증가하고 속도감소 메모리 부족이라는 현상을 나타냈다. 본 논문에서는 실시간 객체 인식이 가능한 Yolo의 적용과 Batch 및 TensorRT 기술 적용을 통하여 신경망을 경량화를 통하여 속도 개선 및 메모리 절약이 가능하도록 설계하였다. 이 논문에서는 이러한 발전된 알고리즘의 연구를 바탕으로 K-reciprocal nearest neighbor 알고리즘, Jaccard distance 비유사도 측정 알고리즘, 산출물 알고리즘 등을 개발하여 공공 CCTV 식별추적시스템 구축을 제시하였다. 그 결과, 비교분석을 통한 알고리즘 조합을 통해 공공 다중CCTV환경에서 실시간으로 객체를 인식하고 재식별하여 객체를 추적할 수 있는 한국형 공공 추적시스템을 제안하였다.

Abstract The government is making great efforts to prevent crimes such as missing children by using public CCTVs. However, there is a shortage of operating manpower, weakening of concentration due to long-term concentration, and difficulty in tracking. In addition, applying real-time object search, re-identification, and tracking through a deep learning algorithm showed a phenomenon of increased parameters and insufficient memory for speed reduction due to complex network analysis. In this paper, we designed the network to improve speed and save memory through the application of Yolo v4, which can recognize real-time objects, and the application of Batch and TensorRT technology. In this thesis, based on the research on these advanced algorithms, OSNet re-ranking and K-reciprocal nearest neighbor for re-identification, Jaccard distance dissimilarity measurement algorithm for correlation, etc. are developed and used in the solution of CCTV national safety identification and tracking system. As a result, we propose a solution that can track objects by recognizing and re-identification objects in real-time within situation of a Korean public multi-CCTV environment through a set of algorithm combinations.

Key Words : AI, Deep Learning, Re-identification, Algorithm, CCTV

*준회원, 아주대학교 과학기술정책전공
**정회원, 유니셈 IOT 사업부
***정회원, 아주대학교 산업공학과 과학기술정책전공(교신저자)
접수일자 2022년 6월 15일, 수정완료 2022년 7월 15일
게재확정일자 2022년 8월 5일

Received: 15 June, 2022 / Revised: 15 July, 2022 / Accepted: 5 August, 2022
*Corresponding Author: mschung333@gmail.com, jooyeoun325@ajou.ac.kr
Department of Industrial Engineering, Ajou University, Korea

Ⅰ. 서 론

폐쇄회로 텔레비전(CCTV: Closed Circuit Television)은 우리가 이동하는 공간 대부분을 녹화하고 중계하고 있다고 해도 과언이 아니다. 정부도 공공 CCTV를 활용하여 범죄 수사 및 예방, 교통관리, 시설물 유지보수 등 많은 분야에 활용하고 있다. 공공기관 CCTV 운영 대수는 2008년 157,197대에서 2020년 1,336,653대로 850% 증가하였고 연평균 20%씩 증가하고 있다[1]. 130만 대의 CCTV를 운용하기 위해서는 많은 인력과 시간 그리고 비용이 소요되며, 운용인력이 모든 CCTV를 분석하여 원하는 성과를 위해 분석하고 추적하는 것은 불가능하다. 하지만, 실종아동(미아), 지적·자폐성·정신장애인, 치매 환자의 실종 접수는 지난 10년(2012~2021)간 440,321건이 발생하였으며 연평균 약 44만 건이 매해 발생하고 있다[2].

대한민국 정부는 CCTV 모니터링 및 사건·사고에 대비하기 위하여 전국에 통합관제센터를 설치하고 운용하고 있지만 1인당 할당하는 CCTV 수가 너무 많아 대책을 갈구하고 있다. 국민의 안전과 범죄 예방 및 용의자 추적 등을 해결하기 위하여 감시 카메라가 여러 곳에 설치되어 운영되고 있지만, 수많은 영상 데이터를 통한 얼굴 인식 재식별에는 형태, 크기, 포즈, 옷 색깔, 신체 관절 등을 포함한 시각적인 외관의 변화 때문에 연구자들에게도 어려운 과제 중 하나로 여겨져 왔다[3]. 더욱이 보행자의 특징을 파악하여 검출하고 추적한다는 것에는 한계가 있다[4],[5]. 운용인력, 시간, 비용 등 현실적 한계를 해결하기 위하여 AI를 도입하여 추적하는 시스템은 실종아동 등을 찾는데 효율적 방안이 될 수 있다.

최근 인공지능(AI)은 딥러닝(Deep Learning)을 중심으로 머신러닝(Machine Learning)의 한계를 넘어서 인간의 뇌를 모방한 알고리즘(Algorithm)을 통하여 발전해 나가고 있으며, 세계적인 IT 기업인 마이크로소프트, IBM, 구글 등은 딥러닝의 결과물을 이용하여 다양한 분야에 적용시키며 연구를 하고 있다[6]. CCTV 운용에서도 인력의 부족, 장시간 모니터링에 따른 인한 집중력 저하 등 운용의 한계에 대해서 보행자 재식별이 대안으로 나오고 있으며, 지능형 영상 보안 서비스를 CCTV 기반으로 실현하기 위하여 다중카메라에서 촬영되는 여러 객체 중 특정 보행자를 검출, 구별, 추적하는 데 활용된다[7]. 2010년에 시작된 ILSVRC(ImageNet Large Scale Visual Recognition Challenge) 대회를 통하여 재식별 기술에 대한 발전을 거듭하면서 딥러닝을 통한 재식별은 새로운 관심을 불러일으켰다. 최근에 인공지능 딥러닝 기술의 발전을 통한 LBP, Gabor, HOG 등 특징 추출 방법을 이용하여 얼굴인식 기술이 평준화되고, LFW(Labeled Face in the Wild)에서 달성할 수 없었던 높은 얼굴 인식 딥러닝 알고리즘들이 달성되고 있다.

이렇게 발전된 딥러닝 기술을 바탕으로 이 연구 논문은 현재까지 발달한 인식, 재식별, 추적 연구를 탐색하고 비교분석을 통해 분석을 실증하였다. 사람을 검지하고 특성을 추출하여 재식별하고 객체를 추적하는 프로세스에서 최신기법의 선택과 비교를 통해 분석시스템의 경량화를 목표로 하고 있다. 또한, 기존 연구들과 다르게 공공 CCTV를 활용하여 실시간으로 객체의 다중 속성을 통하여 Multi-CCTV 환경에서 구현할 수 있는 알고리즘의 조합을 탐색하고 제시하였다. 이를 통하여 현장에서 사용 가능한 객체를 탐색하고 추적하는 시스템을 구현하였다.

Ⅱ. 관련연구

1. 얼굴 인식 재식별에 대한 기존 연구 사례

객체 인식/분류 기술 성능을 볼 수 있는 ILSVRC (ImageNet Large Scale Visual Recognition Challenge)대회는 2010년 전문가들에 의해서 처음 시작되었고 이후 매년 열리게 되었다. 그러나 처음 시작 당시에 인식 기술의 오차는 28.2%로 사람의 인식 오차 수준인 5% 미만에 상당히 부족하였다[8].

2012년 대회에서 Geoffrey E. Hinton 의 제자인 Alex Krizhevsky가 제안한 GPU기반의 DCNN(Deep Convolutional Neural Network) AlexNet이 에러 15.4%를 기록해 2위 에러율 26.2%를 큰 폭으로 앞서 우승하였는데, 연구의 중심 내용은 인식 기술을 획기적으로 개선을 하였다[9]. 특히 이 연구는 최적화된 GPU 기반의 알고리즘 구현하였고 ReLU(Rectified Linear Unit) 활성화 함수를 통하여 학습속도가 느린 것을 개선하였다. 2013년 ILSVRC 대회에서 우승한 Matthew Zeiler와 Rob Fergus은 ZF Net을 개발하였다. 중심 연구내용은 인식성능을 최적화할 수 있게 제안한 것이다. 이것은 DCNN의 각층에 학습된 내용을 쉽게 이해할 수 있게 하였으며 분류 에러율 13.5% 를 보였다[10]. 2014년 GoogLeNet 팀이 ILSVRC 에서 VGGNet을 제치고 1등을 차지하였다. 이 모델은 깊이를 늘리거나 1x1 컨볼루션층(convolution layer)의 사용으로 모델의 성능을 개

선하는 노력을 했지만 그 보다는 9개 인셉션(Inception) 모듈과 함께한 22개의 계층으로 이루어진 개선된 DCNN 구조를 제시함으로 인하여 오차율은 5.98%까지 감소하였다[11].

1983년부터 매년 6월 열리는 국제 학술대회인 2015년 CVPR(Computer Vision and Pattern Recognition)에서 구글은 FaceNet 을 발표하였다. 중심 사항은 Anchor와 Positive 사이는 가깝게, LFW에서 Anchor-Negative 사이는 멀게 하는 Triplet Loss를 이용하여 99.63%의 결과를 이뤘다[12]. 또한, 하와이에서 열렸던 2017년 CVPR 대회인 IEEE Conference에서 Zhun Zhong 팀은 그의 연구논문에서 Person re-identification을 위한 이미지 검색의 성능을 높이기 위해 환경을 고려하여 feature distance를 향상시켰다. reciprocal neighborhood를 제안하는데, 서로에게 가장 가까이 있는 이웃일수록 같은 identity일 가능성이 크다는 가정을 가지고 이미지 검색을 수행하였다[13].

Ⅲ. 연구방법

본 연구는 CCTV 영상 분석을 통하여 딥러닝 객체 인식 및 특정 대상 탐지 및 추적을 위한 솔루션의 딥러닝 기반 검색 방법을 제시하는 것을 목표로 한다. 딥러닝은 객체 및 속성 인식, 재식별, 추적의 3가지 프로세스를 두고 개발 및 대구 수성구 데이터를 통해 실증하였다.

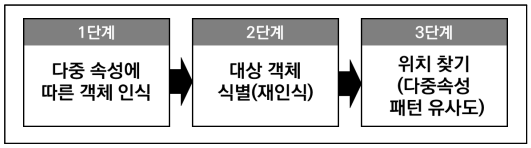

그림 1. 연구체계도
Fig. 1. Research Process

실종 대상(사람)의 선택된 객체(사람) 또는 입력 사진 그리고 속성 입력값과 유사도가 높은 객체를 딥러닝 알고리즘을 통해 다중 CCTV에서 전송된 NVR(Network Video Recorder)에서 영상의 프레임을 추출하여 객체를 시간순으로 탐색하는 솔루션을 개발하였다. 다중 CCTV 영상에서 다중 속성을 통한 객체 1)인식, 2)재식별, 3)위치 찾기(유사도)를 수행하기 위한 딥러닝 알고리즘 조합을 제시하였다.

1단계 인식에서는 분석 경량화를 통한 다중 CCTV 환경에서 다중 속성을 실시간으로 탐색할 수 있는 객체 식별(검지)을 목표로 하였다. 2단계에서는 제시된 사람(미아 등)과 동일한 인물을 찾아내는 재식별(재인식)단계로써 다양한 딥러닝 모델 중 최적화된 모델을 제시하는 것으로 목표로 한다. 3단계에서는 제시된 다중 속성과 이미지/영상에서 객체의 동선을 파악해낼 수 있도록 적용 가능한 유사도 측정 알고리즘을 탐색하였다.

실험은 수성구청의 시험데이터를 통해 수성구청 내의 AI실증랩에서 진행되었으며, 서버는 MS windows 10 pro(64bit), CPU: AMD 7302 processor(2ea), NVIDIA A100 40GB(3ea), Memory256gb, Disk: 8TB(SSD)를 사양으로 하여 실증을 하였다. 실증에서 사용된 CCTV 이미지 데이터는 표 1과 같다.

표 1. 데이터 개요
Table 1. Data Set

	gallery/ train images	vaildata	test/ query
검지(yolo)	13,000	950	-
속성추출	352,374	31,713	35,237
재식별	16,441	-	3,441

1. 객체 식별/재인식 알고리즘

가. 객체 식별(검지) 알고리즘 적용

CCTV를 활용하여 실종자 탐색을 위하여 제1차적으로 제시된 객체(사람) 속성 정보를 바탕으로 연관 객체(사람) 이미지를 찾는 과정이 필요하다. 본 연구에서는 방향, 성별, 옷차림 등 표 2와 같이 객체의 속성을 분류하여 탐색을 시도하였다. 주어진 실종자 속성정보를 기반으로 크롭된 객체의 이미지를 입력받아 사람 이미지마다 속성을 추출하는 딥러닝 모델을 시도하였으며, CCTV 또는 NVR을 통한 영상 스트림에서 AI 모델을 사용하여 객체를 검지하는 기능으로서 검지 된 객체 이미지를 DB에 추가하고 객체 검지모델을 구현했다.

객체 식별을 위한 알고리즘을 적용하기 위하여 현재 딥러닝의 모델을 비교분석하여 딥러닝 최신 모델인 Yolo v4를 채택하여 인식하였다. Yolo는 객체 인식프레임워크로서 영상에서 실시간으로 객체를 인식하는데 많이 활용되고 있다[14]. 기존 객체 인식 방법인 Adaboos, Haar, HOG, SVM 등의 방법은 느린 속도와 최적화의 문제로 객체인식을 사람 시각 시스템을 모방하기에는 한계가 있다. 반면, Yolo를 통하여 객체 인식을 훨씬 높은 수준에서 영상 전체를 한 번에 분석할 수 있다[15]. 딥러닝

객체 식별에 Yolo v4인 MS COCO(오픈 데이터셋)를 선정함으로 객체 식별 성능을 최대로 끌어올리도록 하였다[16].

표 2. 객체 속성 분류표
Table 2. Object Attribute Data Specification

속성		
방향	전면, 측면, 후면 인식불가	
성별	남성, 여성, 인식불가	
옷차림	상의	타입(예:긴팔, 반팔, 원피스...)
		색상(예: 갈색, 빨강, 주황...)
	하의	타입(긴바지, 짧은바지, 치마..)
		색상(예: 갈색, 빨강, 주황...)
장구류	머리카락	남성(예:스포츠, 장발, 머리없음...)
		여성(예:단발, 긴머리...)
	모자류(예:모자, 헬맷, 미착용)	
	안경류(예:일반안경, 선글라스, 미착용)	
	가방(예: 백팩, 손가방, 크로스백, 미착용)	
	신발류 컬러(예: 갈색, 빨강, 주황...)	
장소	도심	
	주택	
	도심 외곽	
시간	일출	
	낮	
	일몰	
	밤	
날씨	맑음	
	흐림	
	비	
	눈	
	안개	
옵션	나이(어린이, 청소년, 청장년, 중년, 노년)	
	키	

CNN(Convolutional Neural Network) 딥러닝 모델은 이미지 기반의 객체 탐색에 있어 성능과 속도가 좋다고 판정되어 객체(사람) 식별 알고리즘 성능고도화에 필요하다고 판단된다. 또한, 반복 실험과 파라미터(parameters) 업데이트에서 최소의 loss를 위해 학습의 결과를 반영 및 평가를 진행하였다.

또한, 다중 영상에 대해서 객체를 탐색하고 추론하기 위해서는 딥러닝 모델의 최적화가 필요하다. 이를 위해서 본 연구에서 모델 최적화를 위해 딥러닝 모델에 TensorRT을 적용하였다. 즉, 다중 영상에서 객체를 탐색하고 추적할 때 병렬 배치(batch)를 통해 동시에 탐색 및 추론하도록 하였다. 따라서 하나의 영상에서 솔루션을 진행하는 것보다 높은 탐색 및 추적이 가능하여 본 연구에서도 병렬 배치(batch)추론이 가능하도록 알고리즘을 설계하였다.

나. 객체 재식별 모델의 선정

Market1501, CUHK03, DUKE, MSMT17의 데이터에서 Backbone(ResNet 기반) 모델보다도 OSNet 모델이 대부분 좋은 성능을 보인다(Table 3)[17]. 본 연구의 재식별(RE-ID)에서도 최고성능이라고 평가되고 있는 ResNet보다 OSNet이 더 적합한지 판단하고 적용하기 위해 Resnet-50과 OSNet 딥러닝 모델을 통하여 비교분석을 실시하였다. 또한, re-ranking의 전후를 비교하여 본 연구에 최적화된 재식별 모델을 발굴하는 것을 목표로 하였다.

2. 객체 추적 시스템(유사도)

찾고자 하는 객체(실종자, 미아 특정대상 등)의 사진 또는 특성을 토대로 대략적 시간과 지역 등 정보를 기반으로 영상에서 일차적으로 객체를 식별한다. 다음으로 유사도가 높은 객체를 찾아내고 해당 객체가 포함된 영상이 탐지될 수 있도록 분석하고 실증하였다. 이를 통해서 찾아낸 객체의 메타 데이터를 통해 위치 시간 데이터를 사용하여 해당 대상(객체)의 위치 및 시간대별 경로를 보여줄 수 있다. 객체를 추적한 데이터를 통해 운영인력의 시간과 노력을 최소화하고 빠른 대응이 가능할 수 있음이 기대된다.

가. 객체 추적 알고리즘 설계

다음으로 객체 추적 알고리즘은 재식별(re-identification) 과정과 객체(사람) 추적(Person Detect)과정으로 시스템이 구성된다. 재식별(Re-ID)과정은 다중 속성(attributes)을 추출하여 RDB(Relational Database)에 추출 값을 저장한다. 솔루션 프로세스에서 특정 객체 탐색의 Query로부터 RDB에 저장된 객체(사람)의 속성정보를 가지고 재순위(re-ranking) 결과를 나타내고 그 결과를 솔루션에 전달한다. 재식별(re-identification)과정과 객체(사람) 추적(Person Detect)과정은 필요에 따라 작업을 요청하고 응답한다.

이를 위해서 속성 및 특징(Feature)을 추출하는 방법을 Mang ye외(2021)는 Feature 추출에서 CNN기법을 활용한 주요 방법을 ①Global Feature Representation Learning, ②Local Feature Representation Learning,

③Auxiliary Feature Representation Learning, ④ Feature Representation Learning 4가지로 분류하여 제시하였다[18] .본 연구의 가장 큰 특징은 다중 영상에서 특징(속성)을 추출하는 것이다. 이를 위해 비디오에서 다중프레임을 사용하는 Feature Representation Learning은 특징(속성) 정보를 더 강력하게 추출하는 특장점이 있는 방법을 채택하였다.

나. 다중 속성에서 재식별 알고리즘 개발[13]

객체의 다중 속성(색상, 형태, 얼굴, 옷 등)의 유사도가 높은 순으로 탐지하기 위하여 re-ranking 및 유사도 측정 알고리즘 적용이 필요로 한다. 기존 연구들에서는 K-nearest neighbors(k-NN)를 통해 re-ranking 하는 것이 주로 채택된 방법이다. 이는 이미지들이 true-match라는 것을 기본으로 한다. 하지만, Zhun Zhong 외(2017)은 K-reciprocal encoding method를 활용한 재인식(re-identification) 결과물에 대한 순위 재조정(re-ranking)을 제언하였다. 그 이유로서 Zhun은 true-match임에도 불구하고 다른 이미지가 true-match 속에 있을 거짓 가능성 때문이다. 이는 k-NN의 결과를 순위 재조정에 바로 사용하는 것은 적절하지 않을 수 있다는 가능성을 시사하였다. 이러한 시사점은 k-NN에 대한 문제제기로서 유사도 측정을 위한 거리에 대해 새로운 해석방법인 K-reciprocal nearest neighbor(k-RNN)를 제언한 것이다. 두 객체가 k-RNN의 관계에 해당한다는 것은 하나의 객체가 probe였을 때 k-NN결과 안에 다른 해당 객체가 있어야 하고 반대 상황에서도 동일하게 k-NN 서로 다른 두 객체가 속해 있어야 한다. 그리하여 Weighted K-reciprocal neighbor 집합을 vector로 인코딩하여 K-reciprocal vectors를 이용하여 자카드거리(Jaccard distance)를 계산한다. 재인식 결과값에 대한 정확도를 향상을 위해 Local query로 확장한다. 최종거리는 원래 거리와 자카드 거리 간의 집합으로 구한다. 본 연구에서도 비유사도 측정알고리즘인 자카드 거리를 통하여 k-NN가 가지고 있을 수 있는 오류를 수정할 K-reciprocal nearest neighbor 알고리즘을 적용하였다.

다. Loss Function

학습 과정에 있어서 무한한 학습은 시간적 한계와 일정 수준 이상의 학습은 과적합의 문제를 가지고 있어 학습을 정지시킬 손실함수를 모델 학습 시 적용해 최적화하였다. 크로스 엔트로피(CE: Cross Entropy) Loss는 보통 Multi -Class Classification에서 매우 효과적인 Loss funtion으로서 주로 적용한다. 그리고 Softmax를 통하여 값들의 범위는 0과1, 총 합은 1로 한다. 해당(정답) 라벨과의 CE를 통해 Loss를 구한다. 또한 정답 클래스에 해당 스코어에 대해서만 로그 합을 하면 최종 Loss를 구하게 된다.

$$CE = -\sum_{i}^{C} t_i log(f(s)_i)$$

$$CE = -\sum_{i=1}^{C'=2} t_i log(f(s_i)) = -t_1 log(f(s_1)) - (1-t_1) log(1-f(s_1))$$

t_i: ground truth (1)
Si: 각 i의 CNN 마지막 층 output score 백터의 i번째요소

Triplet 손실을 통하여 상대적 거리 차이를 최적 학습한다. 수행목표는 기준과 Positive의 거리는 최소화하고, 기준과 Negative와의 거리는 최대화한다(Anchor는 기준이 되는 이미지이며, Positive는 기준인 Anchor와 동일한 ID의 이미지이며, Negative는 Anchor와 다른 ID의 이미지를 뜻함). 최종적으로 이를 통해 유의미한 결과를 도출할 수 있도록 적용하였다.

Ⅳ. 실험결과 및 고찰

1. 객체(사람) 인식

라벨링 데이터를 확보하기 위해 데이터수집 및 NVR에 녹화된 영상을 통해 이미지를 수집한 뒤 라벨링을 수

표 3. OSNet을 통한 속성 추출 모델 결과
Table 3. Attributes Extraction Model as OSNet

학습 환경	OS: Windows10, GPUs: A100 3개
파라미터	Batch size: 512, learning rate: 0.065, epoch: 150
Label	[No-Color Model] attributes: 28 (upper colors, lower colors, shoes colors 제외)
Quantity	Train : 352,374 / Test set [10%], Validate set [10%]
학습 결과	---------- subset|#images ---------- train|352374 val|31713 test|35237 total:19324 ---------- #testpersons:35237 (instance-based)accuracy:92.5% (instance-based)precition:95.8% (instance-based)recall:95.9% (instance-based)f1-score:95.8% **(label-based)meanaccuracy:89.6%**

행하여 데이터셋을 확보하여 학습하였다. 다중카메라 상황에서 대상 객체에 대해 식별/재인식 하는 딥러닝 알고리즘 솔루션 개발에서 사람 검지 모델 Yolo v4는 90.12%정확도를 나타냈다. 다중 속성하에서 속성추출 모델 OSNet 89.6%의 정확도를 나타냈다.

2. OSNet 기반 객체 재인식(재식별)

Re-ranking을 적용한 Resnet-50과 OSNet 알고리즘에서 mAP 및 Rank 1부터 Rank 20까지의 성능을 비교하였다. 그 결과, OSNet에서의 mAP는 Resnet-50보다 13.9%p 높은 94.7%의 성능을 보였으며, 모든 Rank에서 우수한 성능 향상을 보였다.

표 4. 객체 재인식에서 ResNet-50과 OSNet 비교 결과표
Table 4. Compare ResNet-50 and OSNet

Applying person re-ranking		
	ResNet-50	OSNet
mAP	85.7%	94.7%
Rank-1	96.7%	99.1%
Rank-5	98.8%	99.8%
Rank-10	99.3%	99.9%
Rank-20	99.7%	99.9%
query	3,441	
gallery	16,441	

또한, OSNet 모델에서 Re-Ranking 알고리즘 전과 후의 성능을 비교하였다. Re-ranking 알고리즘 적용 여부에 따른 mAP와 Rank 1부터 Rank 20까지의 결과값을 통하여 성능을 비교하였다. mAP는 re-ranking 적용 전 80.8%로 나타났으나 적용 후 94.7%로 mAP는 13.9%p 높은 결과값을 보였다. 또한 re-ranking 적용 전 Rank1에서는 93.3% 정확도에서 99.1%로 5.8%p의 결과값이 향상되는 결과를 나타냈으며, Rank 20에서는 re-ranking 전 99.0%에서 적용 후 99.9% 0.9%p 향상으로 미비하지만 re-ranking 후 모든 mAP와 Rank에서 향상된 결과값을 보였다.

표 5. OSNet에서 Re-ranking 전/후 비교 결과
Table 5. Compare Before and After Re-ranking in OSNet

OSNet		
	before re-ranking	after re-ranking
mAP	80.8%	94.7%
Rank-1	93.3%	99.1%
Rank-5	97.7%	99.8%
Rank-10	98.5%	99.9%
Rank-20	99.0%	99.9%
query	3,441	
gallery	16,441	

3. 다중 속성인식과 객체 재식별 구현

앞선 재식별에 대한 알고리즘 비교와 re-ranking 전후 비교를 통하여 OSNet re-ranking을 Re-ID(재식별)에 적용하였다. 그림2와 같이 객체의 속성 인식에서 투입 속성요소로서 얼굴(정면), 성별(남성), 윗옷(긴팔), 바지(긴바지), 가방(핸드백), 안경(착용), 머리카락(긴 머리)을 투입하면 여러 이미지 갤러리에서 해당하는 객체를 식별하는 시스템을 구축하였다. 높은 수준의 인식률을 보였다. 작업 산출물로 투입된 객체에 대한 적합한 로그와 Ranked Images의 일치도를 확인할 수 있다(Table 4).

그림 2. 다중속성 값을 활용한 인섹 모델 및 결과 예시
Fig. 2. Attribute Extraction Model and Output

표 6. 속성별 추출 결과값
Table 6. Attribute Result

	face -측면	gender-남	up -긴팔	low -긴바지	bag -미착용	glasses-미착용	hair -노말
rank1	94.7%	98.4%	97.5%	98.9%	97.3%	5.7%	95.3%
rank2	96.8%	97.6%	95%	98.4%	88.4%	22.4%	92.2%
rank3	71.7%	50.6%	94.2%	91.1%	36.7%	3.4%	11.3%
rank4	99%	54.4%	38.8%	52.2%	73.7%	42.4%	48.6%
rank5	57.2%	93.9%	45.8%	43.5%	89.8%	2.6%	66.4%

사람의 특징 즉 성별, 옷차림(상의, 하의), 장구류(머리카락, 모자, 안경 등), 장소, 시간, 날씨 등으로 분류된 속성 라벨로 데이터셋을 구성한다. 각 부분의 특징에 대한 속성값을 학습하여 식별 및 재인식을 수행하는 것이 효과적이다. 그리고, 사람별로 라벨을 추출하고 CCTV 영상에 검지 된 객체(사람)의 라벨을 입력된 라벨과 비교하

여 유사도가 가장 높은 데이터 순(re-ranking)으로 결과를 나타낼 수 있도록 구현하였다.

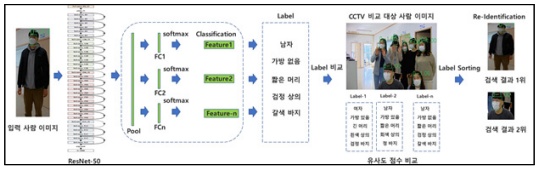

그림 3. 재인식 및 re-ranking 과정
Fig. 3. Design of Re-Identification Using Properties

4. 객체(사람) 추적

Muti-CCTV 환경에서 다양한 속성을 가지고 있는 객체(사람, 미아, 대상)에 대해서 추적하는 것이 본 연구의 가장 주요 목표이다. 이를 실현하기 위하여 제시된 속성 정보와 이미지를 통하여 동선을 추정을 실험하였다. k-RNN을 통해 재식별 결과를 re-ranking 함으로써 K-neareast neighbors의 한계를 보완하여 정확도를 높였다. 또한, Cross-Entropy Loss Function과 Triplet Loss Function을 적용하여 분석 시간을 단축 할 수 있는 최적화된 알고리즘을 설계하였다.

표 7. 추적 시험 및 결과
Table 7. Test & Result of Tracking

시험 목표	제시된 사람 속성 정보와 이미지(문제 데이터)를 바탕으로 동선 추정
시험 기준	o 측정지표 및 산식 -측정지표 : Rank-k Accuracy (k:1부터 5 사이의 정수) -산식(k개 중에 정답이 있으면 정답) ·Rank-k Accuracy(%)=n÷h×100 (n:정답수, h:사람등장횟수) ·총10문제 별로 Rank-k Accuracy(%)를 구한 후 평균산출
시험 데이터	· 갤러리 데이터: 10개의 CCTV 원본 영상(약 5분 분량) ※각 영상은 서로다른 위치의 CCTV에서 획득 ·문제 데이터: 특정사람(찾을대상)의 속성정보 및 크롭 이미지
실험 결과	제시된 사람 속성 정보와 이미지(문제 데이터)를 바탕으로 동선 추정 성능은 Rank-1 Acc:81.9%,Rank-2Acc:84.4%,Rank-3Acc:84.4%,Rank-4Acc:84.4%, Rank-5Acc:84.4%로확인됨

5. 인식 및 재식별 딥러닝 수행속도

연구를 통해 제시한 알고리즘의 조합과 프로세스는 기존 연구의 한계인 실시간 다중카메라 인식 및 추적이 불가능한 형태에서 보다 발전된 결과물을 보여준다. 다수의 동영상을 프레임 당 객체 식별은 70ms(15fps)로 고속 처리를 통해 실시간 구현이 가능했다. 특히, 10개 CCTV의 동영상은 15fps ~ 30fps로 제공되어 최소 70ms 이내로 객체(사람)을 인식하였다. 제시된 사람이미지인 100문제에 대해서 이미지 갤러리(3000장) 속 동일 인물 이미지를 찾는 처리시간 20분(4ms)이 소요되었다.

V. 결 론

기존 CCTV 활용과 관련한 연구에서는 단일속성(색, 얼굴 등)을 통하여 단일 카메라에서 녹화된 영상을 분석하고 추적하는 것이 주로 수행되었다. 본 연구는 객체를 검색하는 모델을 Yolo v4를 적용하였다. 이를 통해, 기존의 고부하 객체 인식a 방식을 개선하여 실시간 객체 인식이 국내 공공 CCTV 환경에서 가능한 것을 확인하였다. 재인식 모델에 대한 재평가와 추적시스템을 다양화하여 최적의 모델을 선택하였다. 이로 인해 다중 CCTV 영상에서 실시간으로 객체(사람)을 추적하는 확인하였다.

본 연구에서 제시한 알고리즘 및 모델을 통해 객체 인식-객체 식별/재인식-객체 추적에 있어서 1) 다중카메라 간 대상 객체에 대해 식별 및 재인식할 수 있는 딥러닝 알고리즘과 2) 다중카메라 간 대상 객체에 대해 추적할 수 있는 알고리즘 그리고 3) 다수의 동영상을 실시간으로 처리하기 위한 프레임 당 객체(사람) 식별 속도 향상 기술을 대구 수성구 공공데이터와 실증랩을 통해 실증하였으며, 적용 가능한 한국형 CCTV관제 AI 시스템을 제시하였다.

하지만, 근본적인 CCTV관제 운용에 있어서 인력의 한계를 보완하기 위해서는 시각화 또는 프로그램화하여 AI에 대한 비전문가도 사용할 수 있는 장치의 개발연구가 필요하다. 또한, 본 연구는 대구 수성구의 CCTV 환경에서만 실증되었다는 지역적 한계를 가지고 있다. 특히, 경찰청, 지자체 등 CCTV의 실시간 연동이 어려운 한계점이 있었다. 또한, 향후 보다 넓은 지역적 범위 즉, 다수 지자체 혹은 전국단위의 연동을 통한 연구가 수행되어야 할 필요성이 있다.

References

[1] S.Korea Personal Information Protection Commission, Data on the current situation of the survey and personal information protection comprehensive support system, 2020. https://www.index.go.kr/potal/stts/idxMain/selectPoS

ttsIdxMainPrint.do?idx_cd=2855&board_cd=INDX_001

[2] S.Korea National Police Agency, Status of reporting and handling of missing children, etc, 2022. https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1610

[3] S. Gong, M. Cristani, S. Yan, C. C. Loy, "Person Re-Identification", Springer London, 2014. DOI: https://doi.org/10.1007/978-1-4471-6296-4

[4] Jae-kyu Park, ki-Bong Nam, Dong-kyun Kim, Kyeong-Seok Han, "Color Mapping track design based on background Intensity difference model for pedestrian detection in CCTV", Journal of The Korea Society of Information Techonology Policy & Management, Vol.10, No.8, pp.273-279, Oct 2016.

[5] Jae-kyu Park, "A Study on Design and Experimental Verification of Deep Learning Based CCTV Pedestrian Detection and Tracking System : Focused on CIC7P Model", Soongsil University, Seoul, 2017

[6] Hyuk-Ro Park, "Systran International", Communications of the Korean Institute of Information Scientists and Engineers, Vol.33, No.10, pp.53-54, Oct 2015.

[7] Kwang Ho Song, Yoo Sung Kim, "Person Re-Identification Using an Attention Pyramid for Local Multiscale Feature Embedding Extracted from a Person's Image", Journal of KIISE, Vol.48, No.12, pp.1305-1317, Dec 2021. DOI: https://doi.org/10.5626/JOK.2021.48.12.13

[8] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge", International Journal of Computer Vision, Vol.115, Issue3, pp211-252, Dec 2015. DOI: https://doi.org/10.1007/s11263-015-0816-y

[9] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks" Communications of the ACM, Vol.60, No.6, pp.84-90, Jun 2017. DOI: https://doi.org/10.1145/3065386

[10] Matthew D. Zeiler and Rob Fergus, "Visualizing and Understanding Convolutional Networks", European conference on computer vision. Springer, Cham, pp.818-833, Sep 2014. DOI: https://doi.org/10.1007/978-3-319-10590-1_53

[11] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, ""Going Deeper with Convolutions"", Proceedings of the IEEE conference on computer vision and pattern recognition, pp1-9, Sep 2014. DOI: https://doi.org/10.48550/arXiv.1409.4842

[12] Florian Schroff, Dmitry Kalenichenko, James Philbin, "FaceNet, A Unified Embedding for Face Recognition and Clustering", Computer Vision and Pattern Recognition, pp.1-10, Jun 2015. DOI: https://doi.org/10.1109/CVPR.2015.7298682

[13] Zhun Zhong, Liang Zheng, Donglin Cao, Shaozi Li, "Re-ranking Person Re-identification with κ-reciprocal Encoding", Computer Vision and Pattern Recognition, pp.3652-3660, May 2017. DOI: https://doi.org/10.48550/arXiv.1701.08398

[14] Jae-Young Chang, Gyu Lee, "Automatic Walking Guide for Visually Impaired People Utilizing an Object Recognition Technology", The Journal of The Institute of Internet, Broadcasting and Communication, Vol. 22, No. 2, pp.115-121, Apr 2022. DOI: https://doi.org/10.7236/JIIBC.2022.22.2.115

[15] Sang Bum Nam, "Deep Learning-Based Real-time Object Tracking on CCTV", KWANGWOON UNIVERSITY, Seoul, 2018.

[16] Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao, "Scaled-YOLOv4: Scaling Cross Stage Partial Network", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp13029-13038, 2021. DOI: https://doi.org/10.1109/CVPR46437.2021.01283

[17] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, Tao Xiang, "Omni-scale feature learning for person re-identification", Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.3702-3712, Nov 2019. DOI: https://doi.org/10.1109/ICCV.2019.00380

[18] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, Steven C.H. Hoi, "Deep Learning for Person Re-identification: A Survey and Outlook", IIEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.44, Issue.6, Jan 2021. DOI: https://doi.org/10.1109/TPAMI.2021.3054775

저 자 소 개

황 주 성(준회원)

- 2021 ~ 현재 : 아주대학교 과학기술정책전공 박사과정
- 2020 ~ 2020 : 울산연구원 혁신성장연구실 전문연구원

Nguyen Thanh Hai(정회원)

- 2011 ~ 2015 : 동국대학교 전자전기공학과 졸업 박사
- 2016 ~ 2018 : ㈜에스에이티 연구소 선임
- 2018 ~ 현재 : ㈜유니셈 IOT 기술연구소 AI팀 과장

강 수 경(정회원)

- 2010 ~ 2012 : 가톨릭대학교 공과대학 졸업 석사
- 2014 ~ 2018 : 이에스이㈜ WCS사업부 대리
- 2018 ~ 현재 : ㈜유니셈 IOT사업부 기술연구소 AI 선임연구원

김 영 규(정회원)

- 2004 ~ 2006 : 명지대 공과대학 졸업 석사
- 2006 ~ 2018 : 건아정보기술 연구소 과장
- 2018 ~ 현재 : ㈜유니셈 IOT 기술연구소 AI 파트장

김 주 용(정회원)

- 2006 ~ 2008 : 성균관대학교 경영전문대학원 수료
- 2007 ~ 2008 : 국립한밭대학교 전자공학과(야간) 겸임교수 활동(겸직)
- 2009 ~ 2020 : The World X-Zone Corporation, Smart City Group, C.T.O.
- 2021 ~ 현재 : 유니셈㈜ IoT사업부 기술연구소 소장/실장

정 명 석(정회원)

- 2018 ~ 현재 : 아주대학교 공과대학 산업공학과 특임교수
- 2015 ~ 2018 : 아주대학교 산업공학과 졸업, 박사
- 2000 ~ 2002 : 서강대학교 대학원 졸업, 경영학석사

이 주 연(정회원)

- 2012 ~ 현재 아주대학교 산업공학과 교수
- 2020.3 ~ 현재 : 중소기업 4차산업혁명위원회 위원장(중소기업중앙회)
- 2020 ~ 2021 : 산업통상부 산업융합촉진 국가옴부즈만(기업호민관)

※ 본 연구는 정보통신산업진흥원의 AI융합 국민안전 확보 및 신속대응 지원사업[2000- 2033-329]으로 수행한 연구결과입니다.
※ 본 연구는 국토교통부의 2022년 국토교통기술사업화지원사업[22TBIP-C162280-02]의 지원을 받아 수행된 연구입니다.
※ 본 연구는 과학기술정보통신부 '과학기술정책 전문인력 양성 및 지원사업'[N2020-0295]의 지원을 받아 수행된 연구입니다.