

# 엣지 디바이스와 카메라 센서 퓨전을 활용한 사람 자세 데이터 자동 수집 시스템

김영근\* · 김승현\* · 김정곤\* · 김원중\*\*

An Automatic Data Collection System for Human Pose  
using Edge Devices and Camera-Based Sensor Fusion

Young-Geun Kim\* · Seung-Hyeon Kim\* · Jung-Kon Kim\* · Won-Jung Kim\*\*

## 요 약

지능형 선별 관제 시스템의 잦은 오탐지로 인해 관제 요원들의 업무 능력 및 시장 신뢰도 저하 문제가 꾸준히 보고되고 있다. 오탐지 문제 개선을 위해 새 AI 모델을 개발하거나 교체하는 것은 기회비용이 크므로, 훈련 데이터 세트 품질을 향상하여 문제를 개선하는 것이 현실적이다. 그러나 소규모 조직은 데이터 세트 수집 및 정제 역량이 부족한 실정이다. 이에 본 논문에서는 사람 자세 추정 모델을 중심으로 엣지 디바이스와 카메라 센서 퓨전을 활용한 사람 자세 데이터 자동 수집 시스템을 제안한다. 이 시스템은 네트워크 말단에서 현장 데이터를 직접 수집하고 레이블링하는 과정을 실시간으로 처리하도록 만들어, 중앙으로 집중되는 연산 부하를 분산시킨다. 또한 현장 데이터를 직접 레이블링하므로 새로운 훈련 데이터 구축에 도움을 준다.

## ABSTRACT

Frequent false positives alarm from the Intelligent Selective Control System have raised significant concerns. These persistent issues have led to declines in operational efficiency and market credibility among agents. Developing a new model or replacing the existing one to mitigate false positives alarm entails substantial opportunity costs; hence, improving the quality of the training dataset is pragmatic. However, smaller organizations face challenges with inadequate capabilities in dataset collection and refinement. This paper proposes an automatic human pose data collection system centered around a human pose estimation model, utilizing camera-based sensor fusion techniques and edge devices. The system facilitates the direct collection and real-time processing of field data at the network periphery, distributing the computational load that typically centralizes. Additionally, by directly labeling field data, it aids in constructing new training datasets.

## 키워드

Intelligent Selective Control, Human Pose Estimation, Edge Device, Camera-based Sensor Fusion, Auto-Labeling  
지능형 선별 관제, 사람 자세 추정, 엣지 디바이스, 카메라 센서 퓨전, 오토 레이블링

\* 대신네트웍스(주)(kimyg96@dsnw.net, skim22@dsnw.net, meatul@dsnw.net)

\*\* 교신저자 : 순천대학교 컴퓨터공학과

• 접수일 : 2023. 11. 02

• 수정완료일 : 2023. 12. 23

• 게재확정일 : 2024. 02. 17

• Received : Nov. 02, 2023, Revised : Dec. 23, 2023, Accepted : Feb. 17, 2024

• Corresponding Author : Won-Jung Kim

Dept. of Computer Engineering, Suncheon National University,

Email : kwj@scnu.ac.kr

## 1. 서론

최근 빈번하게 발생되고 있는 범죄로 인해 치안에 대한 관심이 높아지면서 취약지역을 중심으로 CCTV가 증설되는 추세다[1]. 하지만, 많은 수의 CCTV 영상을 제한된 소수의 인력으로 관제하기는 어렵다[2]. 이에 관련 업체들은 사람을 자동으로 탐지/처리하는 지능형 선별 관제 시스템 개발과 도입에 노력과 자원을 아낌없이 투자하고 있다[3]. 하지만 오탐지 빈도가 잦아 선별 관제 시스템을 온전히 신뢰하기가 어려운 실정이다.

오탐지는 AI 모델의 추론 정밀도(precision)에 직접적인 영향을 받는다. 관제 시스템은 이벤트를 놓치지 않아야 하므로 정밀도를 임의로 낮출 수 없으며, 모델을 다른 것으로 교체하더라도 이 문제에 다시 직면하게 될 가능성이 높다. 따라서 모델의 변경 없이 이 문제를 개선할 방안이 필요하다.

현장 특성이 잘 반영된 데이터 세트 확보는 모델 변경 없이 오탐지 문제를 개선할 수 있는 가장 합리적인 방안이다. 하지만 데이터 세트 확보 능력이 부족한 대다수 소규모 조직은 실제 현장의 특성과 유사한 오픈 데이터 세트를 AI 모델 훈련에 활용하고 있다[4]. 오픈 데이터 세트는 출처에 따라 품질 수준이 다르며, 실제 현장의 특성을 온전히 반영하지 못하므로, 이를 학습한 AI 모델은 일반화 수준이 높을 수 있으나 현장에 특화된 상태가 아니기에 오탐지 빈도가 잦을 가능성이 높다[5].

데이터 세트의 품질이 AI 모델의 정확성에 직접적인 영향을 미친다는 사실은 이미 잘 알려져 있으며, 다양한 분야에서 이와 관련된 연구를 수행하였다. 비닐하우스 내 작물의 생육 환경 영상 수집과 관련된 연구[6]에서는 원격제어 카트와 멀티 카메라로 구성된 시스템을 제안하였고, 병해충 이미지 수집 시스템을 구현한 연구[7]에서는 전문가 검수 과정을 추가하여 데이터 세트의 신뢰성을 높였다.

한편, 현장 데이터 수집 자체가 어려워 가상의 데이터를 합성하는 연구 유형도 존재한다. 특히 사고 상황을 연출하기 어려운 자율 주행 분야에서 활발히 연구되고 있다. 자율 주행 시뮬레이션 시스템을 구현한 연구[8]에서는 다양한 시나리오로 구성된 가상 환경에서 재현된 데이터를 활용하는 방법을 제안하였다. 생

성 AI 모델을 이용해 차량 내부에서 관찰되는 운전 및 야간 영상을 합성한 연구[9]에서는 합성 데이터가 AI 모델의 추론 정확성에 미치는 영향을 분석하였다.

하지만 CCTV를 이용한 시민 안전 관제와 관련된 데이터 수집 및 품질 향상 연구는 찾아보기 어려웠으며, 오픈 데이터 세트를 활용할 수 있다고 하더라도 앞서 언급한 바와 같이 지속적으로 현장에 특화된 영상 데이터를 수집하고 정제할 방안이 부족한 실정이다.

이에 본 논문에서는 엣지 디바이스와 카메라 센서 퓨전을 활용한 사람 자세 자동 수집 시스템 개발을 통해 현장의 영상 데이터를 직접 수집하고 레이블링하는 방안을 제안한다. 지능형 선별 관제 시스템이 범죄 예방에 집중한다는 사실에 의거하여 연구 범위를 사람 자세 추정 작업에 국한하였다.

논문의 II장에서는 제안한 시스템의 설계에 대한 내용을 기술하였다. III장에서는 구현과 실험, 마지막으로 IV장에서는 결론과 향후 과제에 대하여 기술하였다.

## II. 시스템 설계

시스템은 현장의 데이터 특성 분포를 잘 반영할 수 있도록 현장에서 직접 데이터를 수집할 수 있어야 하고, 데이터 수집뿐만 아니라 레이블링과 같은 처리를 자동으로 수행할 수 있어야 한다.

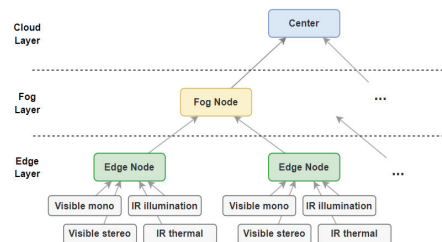


그림 1. 제안 시스템의 아키텍처  
Fig. 1 Architecture of the proposed system

이를 위해 시스템은 엣지 컴퓨팅 방식의 인프라를 갖추어야 한다. 그림 1은 데이터 자동 수집 시스템의 전반적인 컴퓨팅 아키텍처를 나타낸다.

엣지 컴퓨팅 아키텍처는 엣지 레이어, 포그(fog) 레이어, 클라우드 레이어 순으로 데이터 흐름이 연결되는 구조를 갖는다[10]. 엣지 레이어와 포그 레이어는 상호 대체될 수 있는 개념이지만, 본 논문에서는 두 개념을 별개로 구분하였다. 엣지 레이어는 데이터 발생 지점에 가까운 네트워크 말단 노드에 분포해 있으며, 포그 레이어는 엣지 레이어와 클라우드 레이어 사이에 위치한다.

엣지 레이어는 제한된 컴퓨팅 능력을 지니고 있으므로 복잡한 AI 연산보다는 카메라로부터 전송된 영상을 간단하게 전처리하고 포그 레이어로 전달하는 역할에 집중한다. 기종과 모달리티가 서로 다른 카메라 모듈들을 사용하므로 해상도 변경, 잡음 제거, 이미지 정합 등이 이 레이어에서 수행된다.

포그 레이어는 엣지 레이어로부터 전달된 프레임 이미지를 실질적으로 분석하는 역할을 담당한다. 자세 추정 AI 모델을 이용하여 이미지에 표현된 사람 객체를 탐색하고 자세 정보를 레이블링하는 작업이 이 레이어에서 수행된다.

마지막으로 클라우드 레이어는 포그 레이어로부터 전송된 프레임 이미지와 관절 좌표 데이터를 취합하고 이를 저장한다.

### III. 구현 및 실험

#### 3.1 엣지 노드 구현

엣지 노드는 NVIDIA® Jetson Nano™ Developer Kit B01을 사용하였다. 이 제품은 128개의 CUDA core로 구성된 Maxwell 아키텍처 기반의 GPU가 포함되어 있고 가격이 저렴하여 임베디드형 AI 프로젝트 진행에 많이 사용되는 제품 중 하나이다. 한편, AI 모델 고도화에 따른 요구 연산 능력의 상한선이 상향되면서, 이 제품만으로는 지능형 프로젝트를 진행하는 것에 한계가 있다.

대신, CUDA를 기반으로 OpenCV와 같은 실시간 영상 처리를 위한 컴퓨터 비전 라이브러리를 빠르게 구동시킬 수 있으므로 카메라 캘리브레이션 및 영상 전처리 작업에 적합하다[11]. 본 논문에서는 다양한 카메라 모듈들을 이 디바이스에 직접 연결하여 엣지 노드를 구성하였다. 운영체제로는 JetPack SDK 4.6.3

이 포함된 Ubuntu 18.04 배포판을, 프로그래밍 언어로는 Python 3.6.9를 사용했다.

카메라는 가시광과 적외선 광원을 동시에 인식할 수 있는 Raspberry Pi Camera Module ver. 2 NoIR과 적외선을 통해 장면에서 온도 분포를 인식할 수 있는 ThermalCam 160B, 그리고 oCamS-1CGN-U 스테레오 카메라 모듈을 장착하였다. 이기종 카메라 모듈들과 Jetson Nano의 연결 및 데이터 처리 과정은 그림 2와 같다.

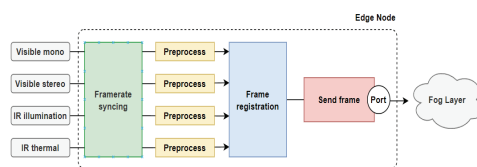


그림 2. 엣지 노드 다이어그램  
Fig. 2 Edge Node configuration diagram

모든 카메라 모듈에 대하여 동일한 시간에 같은 장면을 촬영할 수 있도록 소프트웨어 수준에서 프레임 레이트(frame rate)를 최대한 동기화하였다. 이후 카메라로부터 프레임 이미지를 로드(load)하여 해상도와 컬러 모드를 변경하고, 잡음 제거, 샤프닝 등 영상처리 알고리즘을 적용해 시각적 특성이 잘 드러나도록 처리하였다. 이 과정이 완료되면 각 프레임 이미지를 퓨전한 다음, 포그 노드로 전송한다. 그림 3은 카메라 모듈들이 부착된 엣지 노드 프로토타입의 모습이다.

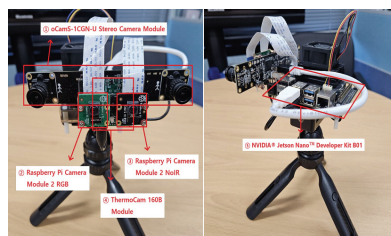


그림 3. 엣지 노드 프로토타입  
Fig. 3 Edge Node prototype

#### 3.2 포그 노드 구현

NVIDIA® Jetson Orin™ Nano Developer Kit을 포그 노드 엣지 디바이스로 사용하였다. 이 제품은 1,024개의 CUDA core와 32개의 Tensor core로 구성

된 Ampere 아키텍처 기반의 GPU가 탑재되어 있다. 이 모델은 엣지 노드의 Jetson Nano 보다 연산 성능이 월등히 뛰어나, AI 모델에 따라서는 경량화를 거치지 않아도 구동시킬 수 있다. 엣지 계층으로부터 전달된 카메라 센서 퓨전 기반의 정합 영상에서 사람을 찾고 자세를 추정하는, 실질적인 지능적 처리를 수행한다. 운영체제로는 JetPack SDK 5.1.0이 포함된 Ubuntu 20.04 배포판을, 프로그래밍 언어로는 Python 3.8.10을 사용했다.

TRT Pose는 NVIDIA의 딥러닝 추론 최적화 라이브러리인 TensorRT에서 실행될 수 있도록 최적화된, 실시간 자세 추정을 위한 AI 모델이다. 본 논문에서는 DenseNet을 TRT Pose의 백본(backbone) 네트워크로 두고 훈련한, 사전 학습된 가중치를 사용하였다. 전송된 프레임 이미지로부터 자세를 추정하는 과정은 그림 4와 같다.

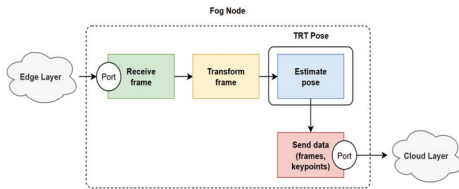


그림 4. 포그 노드 다이어그램  
Fig. 4 Fog Node configuration diagram

이미지를 가로세로 중 짧은 것을 변의 길이로 갖는 정사각형이 되도록 중앙 크롭(crop)한 후, 256x256으로 크기를 조정하고, 각 RGB 픽셀값을 평균 0.485, 0.456, 0.406, 표준 편차 0.229, 0.224, 0.225로 설정하여 정규화하였다.

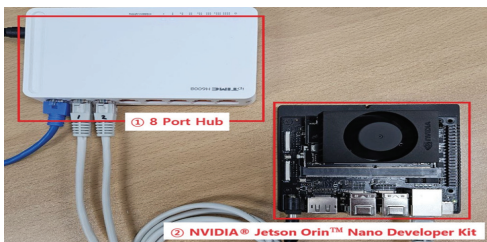


그림 5. 포그 노드 프로토타입  
Fig. 5 Fog Node prototype

자세 추정 모델은 전처리된 입력 프레임 이미지로

부터 사람의 관절 위치를 추론하여 좌표를 출력한다. 그림 5는 상기한 과정을 수행하는 포그 노드 프로토타입의 모습이다.

### 3.3 클라우드 노드 구현

클라우드 노드는 중앙 집중식 인프라의 센터 서버와 동일한 개념이다. 포그 레이어로부터 전송된 프레임 이미지와 자세 정보 레이블링 데이터를 저장하고, 모니터링 장치를 통해 결과를 디스플레이하는 역할을 수행한다.

### 3.4 자세 추정 모델의 레이블링 성능 검증

성능 검증을 위해 COCO 데이터 세트의 val 2017 키포인트 데이터를 벤치마킹 데이터 세트로 사용하였다. 이 데이터 세트는 이미지 5,000장과 각 이미지에 대한 레이블링 데이터로 구성되며, 용량은 약 800MB이다. 벤치마킹 결과는 표 1과 같다.

표 1. 자세 추정 모델 성능  
Table 1. Performance of pose estimation model

AP/AR	IoU	Area	maxDets	Result
AP	0.50 : 0.95	all	20	0.213
AP	0.50	all	20	0.507
AP	0.75	all	20	0.209
AP	0.50 : 0.95	medium	20	0.208
AP	0.50 : 0.95	large	20	0.329
AR	0.50 : 0.95	all	20	0.311
AR	0.50	all	20	0.576
AR	0.75	all	20	0.324
AR	0.50 : 0.95	medium	20	0.240
AR	0.50 : 0.95	large	20	0.421

AP(Average Precision)는 재현율을 0.1부터 1.0까지 0.1 간격으로 변화시킬 때 관찰되는 정밀도 값들을 평균한 것이며, AR(Average Recall)은 탐지 개수에 대한 재현율 값을 평균한 것이다. IoU(Intersection over Union) 0.50 : 0.95는 IoU Threshold를 0.50부터 0.95까지 0.05 간격으로 변화시킨다는 의미이며, IoU 0.50 또는 0.75는 해당 값에 대해서만 검증했음을 의미한다. Area는 객체 바운딩 박스 크기에 대한 것으로, 32<sup>2</sup> 미만의 객체를 'small', 32<sup>2</sup> 이상 92<sup>2</sup> 미만의 객체를 'medium', 그리고 92<sup>2</sup> 이상인 객체를 'large'로

구분한다. maxDets은 탐지 가능한 객체 최대 수량을 나타내며, 본 논문에서는 20개로 설정하였다.

표 1에서 IoU 0.50 : 0.95 - Area all에 대한 AP 검증 결과를 살펴보면, 모든 신체 부위에 대하여 평균 정밀도 0.213(21%)을 달성하였음을 알 수 있다. 이 수치는 동일 벤치마킹 데이터 세트에 대한 기성 자세 추정 모델(OpenPose, AlphaPose 등)들의 AP 값 0.6 ~ 0.7의 절반에도 미치지 못한다. 그러나 상기한 검증 결과는 해당 벤치마킹 데이터 세트에 한정된 성능 수치이며, 단순히 AP 수치만으로 모델의 실용성을 따지기는 어렵다. 이는 모델이 실제 구동될 여러 환경적인 요소를 함께 고려하여야 하기 때문이다. 예를 들어, 충분한 연산 자원을 갖춘 머신을 가정하여 설계된 기성 모델들은 엣지 디바이스 같은, 자원이 한정적인 머신 환경에서 구동조차 되지 않을 가능성이 높으므로 실용적인 모델이라고 판단하기 어렵다.

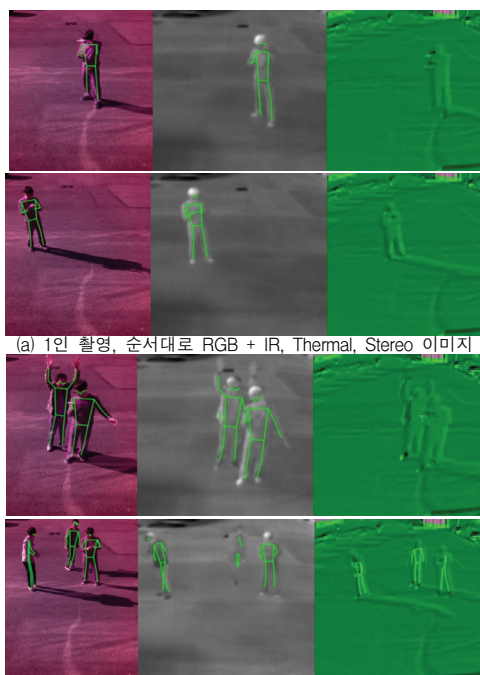


그림 6. 실제 현장에서 자세 추정 수행 결과 화면  
Fig. 6 Pose estimation result on site

한편, 표 1에서 IoU가 0.50일 때 AP와 AR 성능은 각각 0.507(50%), 0.576(57%)을 달성하였으며, IoU

0.50 : 0.95 - Area large의 AP 값이 Area all 보다 상대적으로 높은 성능을 달성하였음을 고려한다면, 모델의 IoU Threshold를 0.5 근처로 세팅하고 피사체가 92<sup>2</sup> 픽셀 이상의 크기가 될 수 있는 가까운 거리에서 촬영할 경우, 충분히 실용적인 퍼포먼스를 달성할 수 있을 것으로 판단된다. 또한, AP와 AR이 균형을 이루므로 오탐지 발생 문제를 개선할 수 있다. 그림 6은 실제 현장에서 엣지 노드와 포그 노드를 통해 레이블링된 데이터를 클라우드 노드에서 시각화하여 실시간으로 출력한 화면이다.

네트워크 환경에 따른 지연(실험 환경에서 2~3초)은 발생하였으나 프레임 레이트를 안정적으로 유지하였으며 준수한 실시간 처리 성능을 보여주었다. RGB + IR 이미지에서 인식률이 가장 높았고, 스테레오 이미지는 잘 인식하지 못하였다. 촬영 각도가 적절할 경우 10~15m 밖의 사람 또한 잘 인식하였다. 단, 사람이 나란히 서 있거나 겹친 경우, 관절 위치를 잘못 인식하는 경우가 많았다.

#### IV. 결론 및 향후 과제

본 논문은 CCTV 기반 지능형 선별 관제 시스템의 오탐지 문제 개선을 위해 고품질 데이터를 자동으로 획득하고 레이블링하는, 데이터 자동 수집 시스템을 제안하였다. 카메라 센서 모듈을 퓨전하여 데이터 다양성을 확보하였고, 엣지 컴퓨팅 인프라를 통해 현장의 특성 분포를 잘 반영한, 고품질의 데이터를 트래픽 문제없이 대량으로 획득할 수 있을 것으로 판단된다. 구현된 시스템의 레이블링 성능은 벤치마킹 데이터 세트에서 기성 모델보다 낮게 측정되었지만, 실용성 측면에서 충분히 유용하게 활용할 수 있을 것이라 판단된다.

단, 이기종 카메라 센서 퓨전에 따른 몇 가지 한계 점을 확인할 수 있었다. 예를 들어, 해상도, 프레임 레이트, 화각(FoV, Field of View) 등 하드웨어적 스펙이 다를 경우, 이를 소프트웨어적인 방법으로 동기화하기가 어려웠으며 연산 과정 추가로 인한 지연이 발생하였다. 향후 카메라의 하드웨어적 특성을 잘 고려하여 성능을 개선하는 연구가 필요하다.

감사의 글

이 논문은 2023년도 교육부 재원으로 한국산업  
기술진흥원의 지원을 받아 수행된 연구임.

References

[1] S.-O. Yoon, M.-S. Kim, and G.-H. Seok, "Inter-Module Interworking Evaluation of TDMA-Based Wireless IP Video Transmission System," *J. of The Korea Institute of Electronic Communication Sciences*, vol. 18, no. 1, 2023, pp. 1-10.

[2] J.-S. Park, M. Wiranegara, and G.-Y. Son, "Multi-channel Video Analysis Based on Deep Learning for Video Surveillance," *J. of The Korea Institute of Electronic Communication Sciences*, vol. 13, no. 6, 2018, pp. 1263-1268.

[3] C.-S. Chung, "A Case Study on the Operation Enhancement of Integrated CCTV Control Center at Busan Metropolitan City," *J. of Korean Association for Regional Information Society*, vol. 18, no. 3, 2015, pp. 123-154.

[4] K.-S. Lim, and G.-W. Kim, "Cloud-based intelligent video surveillance platform for providing empirical-based deep learning video analysis technology," *Review of Korea Institute of Information Security & Cryptology*, vol. 29, no. 3, 2019, pp. 37-43.

[5] Y.-H. Kim, "Demonstration case study of deep learning-based smart city video surveillance solution in Osan," *Information and Communications Mag.*, vol. 37, no. 5, 2020, pp. 42-48.

[6] H.-K. Lyu, and S. Yun, "Agricultural environment image data collection system development for artificial intelligence learning," *Summer Annual Conf. of The Institute of Electronics and Information Engineers, Gwangju, Korea*, 2020, pp. 887-888.

[7] D.-W. Jeong, S.-J. Yoo, Y.-H. Gu, K.-H. Lee,

and J.-H. Park, "Pests Image Collect and Inspect System Design and Implementation for the Image-Based Pests Automatic Diagnosis," *Summer Annual Conf. of The Institute of Electronics and Information Engineers, Jeju, Korea*, 2016, pp. 1694-1697.

[8] E.-O. Joo, D.-Y. Kim, B.-S. Cho, and M.-S. Kim, "Implementation of Autonomous Driving Simulation System using 3D Spatial Data for Various Imagery Data Collection," *J. of Korean Society for Geospatial Information Science*, vol. 29, no. 3, 2021, pp. 3-12.

[9] T.-H. An, J. Kang, and K.-W. Min, "Analysis of the Effects of Data Augmentation Techniques with CycleGAN on Semantic Segmentation in Night and Rain Environments," *Summer Annual Conf. of The Institute of Electronics and Information Engineers, Jeju, Korea*, 2021, pp. 2237-2239.

[10] N.M. Fasial, and T. Helmy, "A Real-Time Deep Learning-based Smart Surveillance Using Fog Computing: A Complete Architecture," *Procedia Computer Science*, vol. 218, no. 107, 2023, pp. 1102-1111.

[11] D.-H. Kim, and S.-Y. Kim, "A Study on Risk Situation Recognition Using OpenCV," *J. of The Korea Institute of Electronic Communication Sciences*, vol. 16, no. 2, 2021, pp. 211-217.

저자 소개



김영근(Young-Geun Kim)

2001년 한려대학교 전자계산학과 졸업(공학사)  
2012년 순천대학교 대학원 컴퓨터과학과 졸업(이학석사)

2014년 순천대학교 대학원 컴퓨터과학과 박사수료  
2022년 ~ 현재 대신네트웍스(주) 기술연구소장

※ 관심분야 : 비전AI, 빅데이터, AIoT



**김승현(Seung-Hyeon Kim)**

2014년 순천대학교 컴퓨터공학과  
졸업(공학사)

2016년 순천대학교 대학원 컴퓨  
터과학과 졸업(이학석사)

2019년 순천대학교 대학원 컴퓨터공학과 박사수료

2022년 ~현재 대신네트웍스(주) 책임연구원

※ 관심분야 : 비전AI



**김정곤(Jung-Gon Kim)**

2023년 순천대학교 대학원 컴퓨  
터공학과 졸업(공학석사)

2023년 순천대학교 대학원 컴퓨  
터공학과 재학중(공학박사)

1997년 ~현재 대신네트웍스(주) 대표이사

※ 관심분야 : 비전AI, 컴퓨터 보안



**김원중(Won-Jung Kim)**

1987년 전남대학교 계산통계학과  
졸업(이학사)

1989년 전남대학교 대학원 전산  
통계학과 졸업(이학석사)

1991년 전남대학교 대학원 전산통계학과 졸업(이  
학박사)

1992년 ~현재 순천대학교 컴퓨터공학과 교수

※ 관심분야 : 비전AI, 빅데이터, 분산시스템

