

# 얼굴 유사성을 이용한 딥러닝 기반 특정인 식별 시스템 구현

서정윤, 박성준, 이하림

금오공과대학교

darkjyuk@naver.com, tjwjddb4018@naver.com, hrlee@kumoh.ac.kr

## Implementation of Deep-learning based Specific Person Identification System using Facial Similarity

Seo, Jeongyun, Park, Sungjun, Lee, Harim

Kumoh National Institute of Technology

### 요약

과학 기술의 발전에도 불구하고, 실종자나 범죄자와 같은 특정 개인을 수색하는 작업은 여전히 경찰과 같은 전문가에 의해 직접 수행되며 상당한 인력과 시간을 필요로 한다. 특히 2022년에 18세 미만 아동 실종 신고 건수는 26,416건 발생하였으며 대부분은 해결되었지만, 1년 동안 찾지 못한 사례 또한 871건 (2022년 4월 기준)에 달한다. 이러한 미해결 실종 사건에 대해서는 지속적인 인력 및 시간 투입이 어렵기 때문에 해당 문제를 해결할 자동화 시스템의 개발이 절실하다. 본 논문에서는 딥러닝 기반의 얼굴 탐지 모듈과 얼굴 간 유사도 측정 모듈을 구현함으로써 특정 인물을 식별할 수 있는 시스템을 개발한다. 얼굴 탐지 모듈은 YOLOv6에 집중 처리 기법인 CBAM을 적용하여 구현하였으며 얼굴 유사도 측정 모듈은 GroupFace와 벡터 특징 처리에 특화된 캡슐 네트워크를 결합하여 개발하였다. 해당 시스템은 Astra 카메라, NVIDIA Jetson Xavier NX, 및 모니터로 구성되며 모든 구성 요소는 Robot Operating System을 통해 연결된다. 얼굴 탐지 모듈의 경우 집중 처리 기법을 적용하였을 때 향상된 얼굴 탐지 성능을 보였고 얼굴 유사도 측정 모듈의 경우 상위 5개 정확도가 93.5%임을 확인하였다. 마지막으로 구현한 시스템의 시연 영상을 제공함으로써 제안 시스템의 효용성을 보인다.

### I. 서론

과학 기술의 발전에도 특정 인물을 탐색하는 작업은 여전히 많은 인적, 물적 자원 및 시간이 요구된다. 게다가, 수색 기간이 길어지는 경우에는 해당 자원들을 지속적으로 투입하는 것에 한계가 있기 때문에 상기 작업을 자동화할 수 있는 새로운 방법이 제시되어야만 한다. 이에 본 논문에서는 딥러닝 기반의 인물 탐색 모듈 및 얼굴 유사도 측정 모듈을 활용하여 자동화된 인물 탐색 수단을 제시하고자 하며 인물 탐색 작업에 대한 새로운 패러다임을 제시하고자 한다.

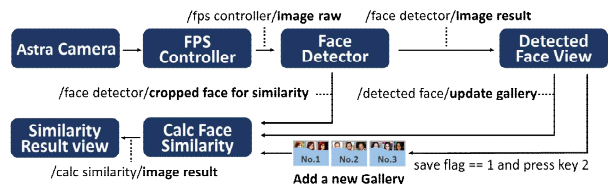


그림 2. 인물 탐색 자동화 시스템 ROS 노드 구조도

이터베이스 내에 존재하는 목표 인물의 얼굴과 추출된 인물의 얼굴 간의 유사도를 계산하여 두 인물의 일치 여부를 판단한다.

### II. 특정 인물 식별 자동화 시스템

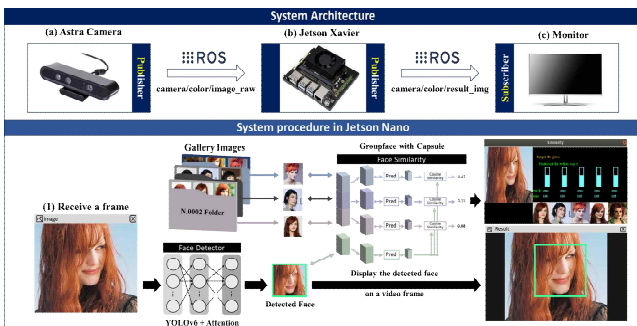


그림 1. 인물 탐색 효율성 증대를 위한 자동화 시스템

제안 시스템은 실시 감시 카메라에 촬영된 임의의 인물을 탐지하여 해당 인물의 얼굴을 추출한다. 이후, 얼굴 유사도 측정 알고리즘에 따라 데

#### 2.1 전체적인 동작

제안 시스템은 카메라, 임베디드 컴퓨터, 디스플레이 장치로 구성되며 각 요소는 ROS(Robot Operating System)로 연결된다. ROS 노드는 FPS controller, face detector, detected face viewer, calc face similarity, similarity result viewer로 이루어져 있으며 노드 간 송수신하는 데이터의 흐름을 그림 2에서 확인할 수 있다.

먼저, 카메라로 촬영한 영상을 딥러닝 네트워크가 탑재된 임베디드 컴퓨터로 전송한다. 임베디드 컴퓨터는 얼굴 탐지 모듈을 활용해 수신한 영상에서 사람 얼굴을 추출하며 추출된 얼굴은 데이터베이스 내의 목표 얼굴과 짝지어진다. 이후, 얼굴 유사도 계산 모듈을 통해 벡터화된 얼굴 이미지 간의 유사도를 측정함으로써 최종적으로 디스플레이 장치에 유사도 점수 및 결과 이미지를 출력한다.

이미지 크기	mAP@0.65				
	FPS	미적용	채널	공간	채널-공간
1280	48~51	0.6040	-	0.6620	-
960	54~61	0.6540	0.3786	0.6700	0.6230
640	55~63	0.5670	0.5400	0.5380	0.5290
320	55~65	0.3200	0.3320	0.3320	0.3360

표 1. 집중 기법 적용 여부 및 입력 이미지 크기에 따른 YOLOv6 네트워크의 성능

데이터 수	코사인 유사도 점수						
	20	40	60	80	100	120	140
상위 1개	0.764	0.811	0.843	0.834	0.848	0.859	0.865
상위 3개	0.845	0.878	0.896	0.893	0.905	0.910	0.919
상위 5개	0.878	0.903	0.919	0.915	0.925	0.931	0.935

표 2. 데이터베이스 내 존재하는 타겟 얼굴 데이터 수에 따른 코사인 유사도 점수

## 2.2 얼굴 탐지 모듈 및 얼굴 유사도 계산 모듈

얼굴 탐지 모듈은 사람의 얼굴을 탐지하고 추출하기 위해 YOLOv6를 사용하였다. 본 연구는 기존 YOLOv6에 집중 처리 기법을 추가적으로 적용하여 성능 향상을 야기하고자 하였고 다양한 집중 처리 기법 중 CBAM(Convolutional Block Attention Module)을 적용하였다. 본래 CBAM은 채널-공간에 대한 집중 처리 기법이나, 채널과 공간을 개별적으로 적용해봄으로써 채널, 공간, 채널-공간 세 가지 집중 모듈 중 가장 적합한 집중 처리 기법이 무엇인지 탐구하고 해당 기법을 적용하였다.

얼굴 유사도 계산 모듈은 얼굴 탐지 모듈에서 검출된 얼굴 이미지와 데이터베이스 내 목표 얼굴 이미지로부터 특징 벡터를 추출한다. 이후, 코사인 유사도 기법을 통해 추출된 두 벡터의 유사도를 계산하고 이를 근거로 탐지된 인물과 목표 인물이 일치하는지 판별한다. 얼굴 이미지로부터 특징 벡터를 추출하는 얼굴 이미지 벡터화 네트워크는 GroupFace와 캡슐 네트워크로 구성하였다. 기존 GroupFace는 ResNet과 선형네트워크로 이루어져 있으나, 특징 벡터화에 특화된 캡슐 네트워크를 ResNet 후미에 적용하여 벡터화된 특징을 다루는 데에서 이점을 취할 수 있도록 하였다.

## III. 실험 및 시스템 구현

### 3.1 데이터셋

해당 시스템의 얼굴 탐지 모듈 네트워크는 WiderFace를, 얼굴 유사도 계산 모듈 네트워크는 정제된 VGGFace2를 사용하였다. 정제된 VGGFace2는 학습된 얼굴 탐지 네트워크를 적용함으로써 기존의 데이터에서 배경을 제외한 얼굴 정보만을 추출하여 재구성하였다.

### 3.2 각 모듈별 네트워크 성능

표 1은 기존 YOLOv6 네트워크에 적용한 집중 처리 기법별 이미지 크기에 대한 얼굴 탐지 성능 및 초당 프레임 처리 수를 보여준다. 성능 측정 지표로는 IoU 문턱 조건을 0.65로 하는 mAP@0.65를 사용하였으며, 객체 검출에 있어 입력 이미지의 크기에 따라 네트워크의 성능 및 연산 속도가 달라지기에 다양한 크기의 이미지에서 mAP 성능 및 초당 프레임 처리 수를 복합적으로 고려하였다.

얼굴 유사도 계산 네트워크의 성능을 평가하기 위해 상위 n 정확도(top n accuracy) 평가지표를 사용했다. n 값이 클수록 정확도가 증가하며, 데이터베이스 내 목표 사람 별 얼굴 이미지 수가 많을수록 성능이 향상되는 것을 표2에서 확인할 수 있다. 다만, 데이터 수가 80개 이상인 경우 성능 향상이 크지 않으므로 대략 80개의 이미지만으로 강인한 시스템을 구축할 수 있을 것이다.

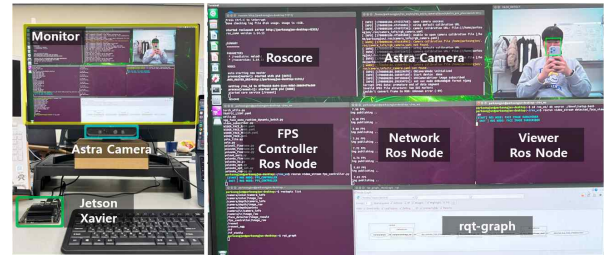


그림 4. 제안 시스템 구성 요소 및 실제 동작 모습

## 3.4 구현 시스템의 시연

그림 3의 좌측은 상기 시스템의 구성 요소를 보여준다. 시스템 구현을 위한 임베디드 컴퓨터로 NVIDIA Jetson Xavier NX를 사용했고 카메라는 Astra Camera Pro Plus를 사용하였다. 우측은 해당 시스템을 구성하는 ROS 노드가 각 터미널에서 실행되는 것을 보여준다. 시연 영상과 관련된 링크는 다음과 같다.

(<https://www.youtube.com/watch?v=3rGlCjDPHTc&t=21s>)

## IV. 결론

본 시스템의 목적은 카메라로부터 받아오는 얼굴 이미지와 데이터베이스 내에 존재하는 목표 인물을 비교함으로써 사람의 신원 일치 여부를 판단하여 효율적으로 인물 탐색 작업을 진행할 수 있도록 하는 것이다. 해당 시스템은 실종 아동, 재난 상황에서의 생존자, 지명 수배범, 탈영병 수색 등 인물 탐색 작업에서 범용적으로 적용될 수 있을 것이며, 인물 탐색 작업에 새로운 패러다임을 제시할 수 있을 것이라 기대된다.

## ACKNOWLEDGMENT

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1G1A1007058).

## 참고 문헌

- [1] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei, and X. Wei, "YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications," CoRR, arXiv:2209.02976, 2022.
- [2] Y. Kim, W. Park, M.-C. Roh, and J. Shin, "GroupFace: Learning Latent Groups and Constructing Group-Based Representations for Face Recognition," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020.
- [3] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," Proceedings of the European conference on computer vision (ECCV), 2018.
- [4] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules," 2017 Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 2017.