

마스크를 착용한 환경에서 얼굴 인식을 활용한 자동 출석체크 시스템

김영국⁰, 임채현*, 손민지*, 김명호**

⁰숭실대학교 융합소프트웨어학과,

*숭실대학교 융합소프트웨어학과,

**숭실대학교 소프트웨어학과

e-mail: 1101978003@soongsil.ac.kr⁰, {immanuel1995, smgg}@soongsil.ac.kr*, kmh@ssu.ac.kr**

Automatic Attendance Check System Using Face Recognition In A Masked Environment

Young-Kook Kim⁰, Chae-Hyun Lim*, Min-Ji Son*, Myung-Ho Kim**

⁰Dept. of Convergence Software, Soongsil University,

*Dept. of Convergence Software, Soongsil University,

**Dept. of Software, Soongsil University

● 요약 ●

본 논문에서는 CCTV를 통해 얻은 영상에서 얼굴을 인식하여 자동으로 출석 여부를 체크하는 시스템을 소개한다. 이 시스템은 CNN을 바탕으로 RetinaFace 모델을 사용하여 얼굴을 탐지하고, 탐지된 얼굴을 ArcFace 모델로 R^{512} 의 목표 공간으로 임베딩한다. 기존 데이터베이스에 등록된 얼굴과 CCTV를 통해 얻은 얼굴들의 임베딩 벡터 사이의 Angular Cosine Distance를 측정하여 동일 인물인지 판단하는 매칭 알고리즘을 제안한다. 실험을 통해 두 모델을 동시에 사용할 최적의 환경을 파악하고, 마스크 착용으로 얼굴의 하단부가 가려지는 폐색 문제에 더욱 효과적으로 대응하여 매칭 성능을 높이는 방법을 제안한다.

키워드: 딥 러닝(Deep Learning), 얼굴 인식(Face Recognition)

I. Introduction

최근 출결을 효율적으로 확인하고 관리하기 위해 RFID, 블루투스, 클리커 등의 다양한 방법을 사용한 시스템이 구축되고 있다. 하지만 대리 출석 등으로 출석체크를 올바르게 진행하지 못하는 문제점이 발생한다.

딥 러닝의 발전은 이미지를 분석하는 분야에서 두각을 나타내고 있다. 특히 딥 러닝 기법 중 CNN (Convolutional Neural Network)은 얼굴의 위치 정보를 파악하고, 얼굴을 인식하는 분야에서 뛰어난 성능을 보이고 있다.

얼굴의 위치 정보 파악은 얼굴 이미지 분석의 필수 단계이다. 최근 코로나 여파로 마스크를 착용하는 경우 얼굴의 하단부가 가려지는 폐색 문제가 빈번히 발생하고 있다. 이 경우에도 최근 딥 러닝 기술은 얼굴의 위치 정보를 잘 파악한다. 하지만 마스크를 착용한 얼굴을 인식하는 것은 아직 어려운 과제이다.

이에 본 논문에서는 마스크를 착용한 얼굴을 인식하는 방법을 제시하고, 강의실의 CCTV를 통해 얻은 영상에서 얼굴을 탐지하고 기존 데이터베이스에 등록된 얼굴과 비교하여 자동으로 출석을 체크하

는 딥 러닝 시스템을 설계한다.

II. Preliminaries

2. Related works

2.1 얼굴 검출

딥 러닝을 사용한 물체 검출 방법은 CNN 기반의 One-stage, Two-stage 탐지 방법으로 구분한다. 물체를 검출하는 방법은 크게 물체의 위치를 찾는 Localization과 물체를 구별하는 Classification으로 이루어진다. One-stage 탐지 방법은 두 가지를 동시에 수행하고, Two-stage 탐지 방법은 순차적으로 수행한다. One-stage 방법은 Two-stage 방법에 비해 정확도가 부족하지만, 속도가 빠르다는 특징이 있다[1, 2].

얼굴 검출에서는 One-stage 방법으로도 충분한 정확도를 보이기

때문에 우리는 RetinaFace[3]라는 One-stage 기반의 Multi-Task Learning 전략을 사용하는 모델을 사용한다. RetinaFace에서 Multi-Task Learning 전략은 얼굴을 검출하는 바운딩 박스 (bounding box)를 탐색하는 것뿐만 아니라 신뢰도, 얼굴의 눈, 코, 입의 양 끝을 표시하는 5개의 랜드마크(landmark), 3D 위치 및 대응관계를 동시에 예측하는 전략이다. WIDER FACE 데이터셋[4] 등 여러 얼굴 검출 데이터셋에서 최첨단 성능을 달성한다.

2.2 얼굴 인식

얼굴 인식을 위해 CNN으로 얼굴을 표현하는 임베딩을 학습하여 같은 사람의 얼굴을 임베딩한 벡터 간의 거리는 짧고, 다른 사람의 얼굴을 임베딩한 벡터 간 거리는 멀도록 학습한다. ArcFace[5]는 기존의 Cosine Loss[6], Triplet Loss[7] 등 Loss함수를 개선하는 Additive Angular Margin Loss를 도입하여 Hypersphere의 Geodesic Distance Margin을 최적화한다. 이런 특징으로 ArcFace는 대규모 영상 데이터셋을 포함한 여러 얼굴 인식 벤치마크들에서 최첨단 성능을 달성한다.

III. The Proposed Scheme

얼굴 탐지 및 인식 모델을 통한 자동 출석체크 시스템[8]을 제안한다. 시스템의 구조는 다음 Fig. 1과 같다.

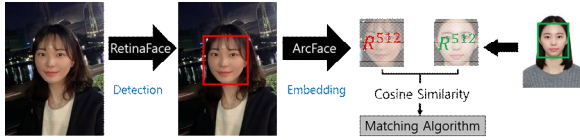


Fig. 1. 출석 시스템 구조

먼저 영상에서 RetinaFace를 통해 얼굴을 탐지하고, 기존 데이터베이스 즉 출석부의 얼굴 사진 및 탐지한 얼굴을 ArcFace를 통해 R^{512} 의 목표 공간으로 임베딩한다. 임베딩하여 생성된 두 벡터 사이의 거리를 Angular Cosine Distance로 측정하여 가장 가까운 두 사진을 서로 매칭한다. 두 벡터 사이의 Angular Cosine Distance는 다음과 같이 측정한다.

$$Angular\ Cosine\ Distance = \frac{\arccos(Cosine\ similarity)}{\pi}$$

매칭 알고리즘은 다음 Table 1과 같다.

Table 1. 매칭 알고리즘

Matching Algorithm
입력 : $\{f_1, \dots, f_m\}$ 출력 : 중복값 제거된 인덱스 리스트 S 1. $F_{m \times m}$ 생성 2. repeat (rows) 3. for ($i=1$ to m) in ($row \leftarrow F_{i \times m}$) { 4. $s_i \leftarrow \arg\min(row)$ 5. if ($s_i = s_{i+1}$) { 6. $s \leftarrow \min(s_i, s_{i+1})$ 7. 큰 값을 제외처리 8. $S_{row} \leftarrow s$ 9. } 10. } 11. until(중복값이 없을 때까지)

알고리즘의 입력은 임베딩된 이미지들의 집합으로, 각각의 f_j ($1 \leq j \leq m$)는 크기 (1, 512)의 벡터이다. 출력은 중복값이 제거된 출석 정보를 담은 리스트 S 로, 크기는 $1 \times m$ 이다. 먼저 임베딩 벡터들 사이의 유사도를 구해 $m \times m$ 크기의 행렬 F 를 생성한다. 이 행렬은 모든 경우의 출석을 비교하기 위해 사용된다. 알고리즘은 $s_i = s_{i+1}$ 과 같은 중복값이 발생하지 않을 때까지 모든 행(rows)을 반복하며, 중복값이 발생할 경우 해당 행(row)에 제외처리를 한다.

검출된 얼굴이 마스크를 착용한 경우 얼굴의 하단부가 가려지는 폐색 문제가 발생한다. 이 경우 출석부의 사진과 탐지된 얼굴의 동일성을 판단하기 어렵다. 이를 해결하려는 방법으로 Arcface의 입력으로 사용할 2가지 이미지 생성 방식을 제안한다. 첫 번째 방식은 RetinaFace를 통해 검출된 바운딩 박스에서 코를 지나는 수평선 하단부를 모두 흰색으로 바꾸어 새로운 이미지를 생성한다. 하단부에서는 동일한 특성을 추출하고, 상단부만 다른 특성을 추출하여 상단부 위주의 거리를 측정한다. 두 번째 방식은 바운딩 박스에서 코를 지나는 수평선 하단부를 잘라내고 상단부만 남긴 이미지를 생성한다. 폐색이 일어난 것으로 예측되는 부분을 제외한 상단부만의 특징을 추출하여 거리를 측정한다. 폐색이 일어나지 않은 부분의 특징에 가중치를 주기 위해 각각 거리를 측정하여 기존 바운딩 박스의 거리와 결합한다.

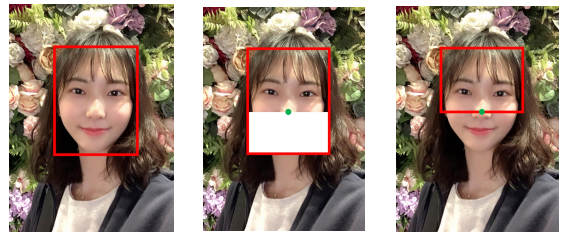


Fig. 2. 거리를 측정할 이미지 예시

IV. Experiments

4.1 실험 환경

RetinaFace는 ResNet-50[9]을 backbone으로 WIDER FACE 데이터셋에서 사전 훈련된 모델을, ArcFace는 ResNet-50을 backbone으로 MS Celeb 1M 데이터셋[10]에서 사전 훈련된 모델을 사용한다.

실험 데이터는 The 'Celebrity Together' Dataset의 훈련 데이터셋에서 임의의 100개 클래스 중 정면 사진을 골라 마스크를 씌운 새로운 데이터셋을 만들었다. 기존 데이터베이스 즉 출처부의 사진처럼 사용할 데이터는 D1, CCTV 환경에서 검출된 이미지처럼 사용할 데이터는 D2, D3는 D2의 이미지에서 마스크를 씌워 만든 이미지 데이터이다. 실험1은 D1과 D2, 실험2는 D1과 D3를 매칭하는 것으로 명칭 한다.

4.2 RetinaFace와 ArcFace 최적화

RetinaFace를 통해 검출된 얼굴의 바운딩 박스의 크기와 데이터베이스의 사진 크기의 관계를 파악한다. 동일 인물을 약 45cm 간격으로 5장씩 촬영하여 서로 다른 해상도의 이미지에서 동일 인물 간 거리를 계산하면 Fig. 3과 같이 비슷한 크기의 사진에서는 0.2 근처의 값으로 측정되는 반면 크기가 많이 차이 나는 경우 0.4 근처의 값으로 차이가 크다. 따라서 우리의 시스템에서는 출처부의 사진을 CCTV를 통해 측정되는 바운딩 박스의 크기와 비슷하게 한다.

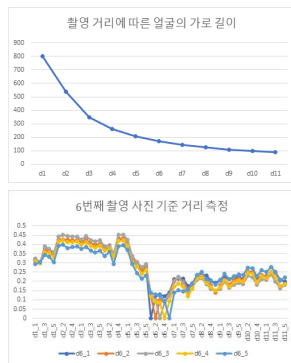


Fig. 3. 촬영 거리에 따른 얼굴의 가로 길이와 거리

4.3 유사도 측정 방법

본 논문에서 거리를 측정할 3가지 이미지를 다음과 같이 명칭 한다. F는 얼굴 이미지 전체, W는 얼굴 이미지의 하단부를 흰색으로 변경한 이미지, N은 얼굴 상단부 이미지를 의미한다. 3가지 이미지를 임베딩하여 실험하여 Rank-1, 매칭 정확도를 측정한 실험 결과는 Table 2와 같다.

Table 2. F, W, N 기반의 Rank-1, 매칭 정확도(%)

	F	W	N
실험1	98,100	82,93	67,69
실험2	69,75	79,89	58,59

실험1에서 W 이미지는 F 이미지에 비해 좋지 않은 성능을 기록하지만, 마스크를 착용한 실험 2에서는 더 좋은 성능을 기록한다. 이를 조합하여 마스크를 착용하지 않은 경우의 정확도를 떨어뜨리지 않고, 마스크를 착용한 경우의 정확도만 높이기 위해 다음 식 (1)과 같은 거리 결합 방법을 제안한다.

$$Distance = \alpha \times F + \beta \times W + \gamma \times N \quad \cdots (1)$$

본 논문에서는 $\alpha = 1.0, \beta = 0.5, \gamma = 0.2$ 의 값으로 측정하였다.

4.4 실험 결과

식 (1)의 거리를 3개의 이미지를 조합하여 Rank-1, 매칭 정확도를 측정한 실험 결과는 Table 3과 같다. 3개 중 2개만 조합하여 측정한 경우에도 α, β, γ 의 값은 위와 같다.

Table 3. 결합한 거리 기반의 Rank-1, 매칭 정확도(%)

	F+W	F+N	F+W+N
실험1	97,100	98,100	97,100
실험2	80,88	75,82	85,94

마스크를 착용하지 않은 실험 1의 경우 우리가 제안한 방법을 적용했을 때 얼굴 검출한 바운딩 박스만 사용하여 거리를 측정한 경우와 크게 다르지 않은 결과를 보였다. 하지만 모두 마스크를 착용한 실험 2의 경우 새로 제안한 방법과 조합할수록 성능이 향상되었다. F 이미지에서 80%의 Rank-1, 88%의 매칭 정확도를 F, W, N 이미지에서 거리를 결합하여 각각 85%의 Rank-1, 94%의 매칭 정확도로 향상시켰다. 실제 강의실 환경에서 10명을 대상으로 결과는 다음과 같다. 실험1은 마스크를 쓰지 않은 환경에서, 실험2는 마스크를 착용한 환경에서 출처부의 사진과 매칭하는 것으로 명칭 한다.



Fig. 4. 실제 강의실 환경에서 실험 사진

Table 4. Fig. 4의 Rank-1, 매칭 정확도(%)

	F	F+W+N
실험1	90,100	90,100
실험2	50,40	60,100

V. Conclusions

자동 출석체크 시스템과 관련해 다양한 접근방법이 제시되고 있다. 본 논문에서는 CCTV 영상을 통한 자동 출석체크 시스템을 제안한다. 영상에서 CNN을 활용하여 얼굴을 검출 및 임베딩하여 기존 데이터베이스의 사진과 비교하여 출석 여부를 판단한다. 실제 환경에 적용하기 위해 경량의 backbone 네트워크를 사용하였으며, 정면 얼굴이 검출될 수 있는 환경으로 제한하였다.

마스크를 쓴 경우 동일성을 잘 판단하기 위해 얼굴 하단부를 흰색으로 만든 이미지, 얼굴의 상단부 이미지로 임베딩된 벡터로 거리를 측정하고 얼굴 전체 이미지에서 얻은 거리와 결합 방법을 제시한다. 그 결과 100명을 대상으로 마스크를 착용하지 않은 경우의 정확도를 유지하면서, 마스크를 쓴 상황에서 기존 방법보다 Rank-1 정확도를 16%, 매칭 정확도를 19% 증가시켰다. 우리가 제안하는 방법 즉 비교할 두 사진에서 폐색이 있는 부분에 동일한 처리 방법을 통해 폐색을 지우거나, 폐색이 일어나지 않은 부분만으로 거리를 측정하여 전체 이미지에서 측정한 거리와 결합하여 매칭하는 것이 선글라스, 모자 등으로 얼굴의 다른 부분을 가려 생기는 폐색 문제에 잘 대응할 것으로 기대한다.

ACKNOWLEDGEMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT연구센터지원사업의 연구결과로 수행되었음" (IITP-2020-2018-0-01419)

REFERENCES

[1] Joseph Redmon, "You Only Look Once: Unified, Real-Time Object Detection," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016.

[2] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," IEEE Conference on Computer Vision and Pattern Recognition, pp. 580-587, 2014.

[3] Jiankang Deng, "RetinaFace: Single-stage Dense Face Localisation in the Wild," Computer Vision and Pattern Recognition (CVPR), arXiv preprint arXiv:1905.00641, 2019.

[4] S. Yang, P. Luo, C. Loy, and X. Tang, "Wider face: A face detection benchmark," IEEE conference on computer vision and pattern recognition, pp. 5525-5533, 2016.

[5] Jiankang Deng, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4690-4699, 2019.

[6] Wang, Hao, "Cosface: Large margin cosine loss for deep face recognition," IEEE Conference on Computer Vision and Pattern Recognition, pp. 5265-5274, 2018.

[7] Florian Schroff, Dmitry Kalenichenko and James Philbin, "Facenet: A unified embedding for face recognition and clustering," IEEE conference on computer vision and pattern recognition, pp. 815-823, 2015.

[8] Mhlee, grseo, hnmam, kwkim and yhkim, "An Automated Seat Checking Using Image Object Detection," Korea Software Congress, 1645-1647, 2019.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.

[10] Y. Guo, L. Zhang, Y. Hu, X. He and J., "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," In European conference on computer vision, pp. 87-102. Springer, Cham, 2016.