얼굴 특징 추출 및 클러스터링을 활용한 얼굴 검색 기법

*신준호 **김종환 **조숙희 **김정학 *고영준
*충남대학교 컴퓨터공학과, **한국전자통신연구원
*{junhoShin@o.cnu.ac.kr, yjkoh@cnu.ac.kr} **{jonghwan, shee, junghak}@etri.re.kr

Face Search Method Based on Face Feature Extraction and Clustering

*Junho Shin **Jong-hwan Kim **Sukhee Cho **Junghak Kim *Yeong Jun Koh
*Department of Computer Science & Engineering, Chungnam National University
**Electronics and Telecommunications Research Institute

요약

최근 미디어의 발전으로 빠른 속도로 많은 양의 사람들의 얼굴이 포함된 사진, 동영상들이 인터넷에 업로드 되고 있다. 이러한 현상에 맞춰 인공지능을 활용한 얼굴 인식 기술의 놀라운 발전이 있었으나, 대규모 데이터셋에서 임의의 인물을 검색하는 경우에서는 연산량과 저장공간의 부담이 존재한다. 특히, 인터넷에 존재하는 수많은 불법 촬영물에서 피해자를 정확하고 신속하게 검색하기 위해서는 효율적인 얼굴 검색 시스템이 필요하다. 따라서, 본 논문은 얼굴 특징 추출과 클러스터링을 활용하여 방대한 양의 불법 촬영물 셋에서 피해자 동영상을 효율적으로 검색할 수 있는 기법을 제안한다. 불법 촬영물 동영상 검색 실험 환경을 만들기 위해 YouTube Faces [1] 데이터셋으로 유사 동영상 셋을 만들고 이 환경에서 실험을 진행한다. 얼굴 특징 추출 모델은 ResNet100 네트워크를 CosFace 손실함수와 Glint360K 데이터셋으로 학습시킨 모델 [2]을 사용한다. 추출된 얼굴 특징들을 HAC(Hierarchical Agglomerative Clustering) 알고리즘으로 클러스터링 한 후, 클러스터 대푯값을 통해 얼굴 검색 실험을 했을 때의 실험 결과를 분석한다.

1. 서론

미디어의 발전으로 사람의 얼굴 이미지가 인터넷에 업로드 되고 있고, 이는 시간이 지나면서 증가폭이 점점 커지고 있다. 이 얼굴 이미지들을 정확하게 인식하고자 하는 선행 연구들이 있었고, 뛰어난 정확성으로 사람 얼굴을 구분할 수 있게 되었다. 하지만, 대규모 얼굴 데이터셋에 임의의 얼굴 이미지를 검색하는 경우, 방대한 데이터와 이에 따른 연산량 문제로 현실에 적용하기 어려운 문제가 생긴다. 예를 들면, 수천 개의 동영상 중에서 임의의 인물이 등장하는지 알고자 할 때, 각 동영상의 등장인물들의 얼굴 이미지가 다량 존재하는데, 이를 전부 검색하기에는 비효율적이므로 각 인물의 대푯값을 추출해내는 것이 필요하게 된다. 우리는 방대한 양의 불법 촬영물 셋에 피해자가 어떤 동영상에 존재하는지 검색하는 것을 가정한다. 본 논문에서는 얼굴 이미지들을 특징 공간에서 클러스터링 한 후, 클러스터의 대푯값을 통해 얼굴을 검색하는 방법을 소개한다.

2. 본론

2.1. 얼굴 특징 추출 네트워크

얼굴 인식에서는 Closed-Set과 Open-Set이라는 두 가지의 환경이 있다. 전자는 시험 환경에서 학습에 등장하지 않았던 인물이 입력되지 않는 것이 보장되고, 후자는 시험 환경에서 학습 데이터에는 없던 제3의 인물이 입력으로 등장할 수 있는 환경이다. 본 논문은 Open-Set 환경을 다룬다. 학습에서 없는 인물이 시험 환경에서 나올 수 있다는 점 때문에, 이를 해결하기 위해 Metric learning [3, 4]이 얼굴 인식 분야의 주류가 되었다. 우리는 이전에 제안된 모델 중, ResNet100 네트워크를 CosFace 손실함수와 Glint360k 데이터셋으로 학습한 네트워크를 사용한다. CosFace의 특성상 클러스터링과 얼굴 검색 실험의 거리는 코사인 거리를 사용한다.

2.2. 클러스터링과 대푯값

클러스터링은 HAC(Hierarchical Agglomerative Clustering) 알고리즘을 사용한다. HAC는 모든 데이터를 클러스터로 간주하는 것으로 시작해서, 거리가 가까운 클러스터들을 합쳐가며 클러스터 사이의 거리가 임계값에 도달했을 때 멈추는 방법이다. HAC에 사용되는 임계값은 실험을 통해 결정된 값을 사용한다. 각 클러스터에 포함된 특징 벡터들의 평균을 클러스터 대푯값으로 설정하고, 이 대푯값을 얼굴 검색에 이용한다. 이에 맞게, HAC 알고리즘의 연결 방식도 평균을 사용한다.

2.3. 데이터셋

본 논문에서는 YouTube Faces 데이터셋을 사용한다. YouTube Faces는 1,595명의 인물에 대해 총 3,425개의 동영상으로 구축되어 있다. 각 동영상은 평균 181.3 프레임으로 구성된다. 인물의 얼굴이 프레임의 중앙에 바로 서있도록 정렬된 버전의 데이터셋을 사용하였으며, RetinaFace를 사용해 얼굴을 검출한 뒤, 얼굴만 떼어내서 사용한다.

우리는 YouTube Faces 데이터셋을 불법 촬영물 검색 실험 환경과 유사하게 재구축했다. 전체 인물 중 동영상이 3개 이상 존재하는 인물 533명을 사용한다. 이 533명 중 4~5명을 무작위로 고르고, 각 인물의 동영상을 한 개씩 골라 모은다. 동영상을 한 개씩 가져올 때, 전체 프레임을 가져오는 대신, 그 절반을 무작위로 골라서 가져온다. 이것을 유사 동영상 이라고 하며, 이 유사 동영상을 5,000개 만들어서 검색 대상이 되는 동영상 셋으로 사용한다.

3. 실험

3.1. 클러스터링 임계값

클러스터링이 어떤 순간에 멈춰야 할지 정하기 위해서는 임계값을 설정해야 한다. 임계값에 따른 성능 평가를 위한 지표가 필요한데, 우리는 BCubed Precision과 BCubed Recall을 사용한다.

$$\text{Precision} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j \in C(i)} \frac{Correctness(i,j)}{|C(i)|}$$

$$\text{Recall} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j \in L(i)} \frac{Correctness(i,j)}{|L(i)|}$$

어떤 특징 i에 대해서 $C(i)$는 같은 클러스터로 분류된 특징들의 집합이고, $L(i)$는 실제로 같은 클래스를 가지는 특징들의 집합이다. $Correctness(i,j)$는 i와 j가 같은 클래스이고, 분류 결과에서도 같은 클러스터일 때 1, 아니면 0이 된다. 우리는 약간의 오버클러스터링을 허용하기 때문에, precision을 중요한 지표로 임계값을 결정했다. 표 1은 유사 동영상 전체를 각각 클러스터링 한 후, 평가지표를 계산해서 평균한 결과이다. 결과에 따라 HAC의 거리 임계값을 0.59로 사용했다.

Threshold	Precision	Recall
0.58	1.0000	0.9779
0.59	1.0000	0.9801
0.60	0.9999	0.9823

표 1 유사 동영상 데이터 셋에 대한 클러스터링 성능

3.2. 얼굴 검색 실험

유사 동영상을 클러스터링 한 뒤, 입력 이미지를 유사 동영상의 클러스터 대푯값과 거리를 계산한다. 가장 가까운 클러스터와의 거리가 검색 실험을 위한 임계값보다 작다면 입력 이미지는 해당 클러스터에 해당하는 인물로 유사 동영상에 존재한다는 뜻이 된다. 이 때, 실제로 입력 이미지의 인물과 해당 클러스터의 대푯값에 가장 가까운 얼굴 이미지가 같은 인물이라면 이것은 잘 예측한 것이고, 다른 인물이라면 틀리게 예측한 경우이다. 반대로, 가장 가까운 클러스터와의 거리가 임계값보다 크다면 이 인물은 해당 유사 동영상에 존재하지 않는다는 뜻이 된다.

위의 시나리오에 따라서 실험 데이터를 구성한다. 유사 동영상의 각 인물마다 해당 인물의 이미지 30장을 positive 샘플로 사용한다. 30장은 각 인물이 YouTube Faces에 있는 다수의 동영상 중 유사 동영상 구축에 사용된 같은 동영상, 다른 동영상 구분 없이 무작위로 구성된다. 이는 인터넷 상의 불법 촬영물들이 중복되는 것이 많다는 점을 고려해서 중복을 허용한 것이다. negative 샘플은 유사 동영상에 포함되지 않은 인물의 이미지를 1장씩 가져와서 positive 샘플 개수와 동일하게 구성한다. 실험 결과를 측정하기 위한 혼동 행렬을 다음과 같이 정의한다.

- TP: 입력 사람이 대상에 존재하고, 잘 분류된 경우
- FP: 입력 사람이 대상에 없고, 특정 클러스터로 분류한 경우
- FN: 입력 사람이 대상에 존재하고, 다른 클래스로 분류하거나 없다고 예측한 경우
- TN: 입력 사람이 대상에 없고, 실제로 없다고 예측한 경우

임계값을 점차 늘려가며 precision과 recall을 계산하고, PR 곡선을 그리면 그림 1과 같은 결과가 나온다.

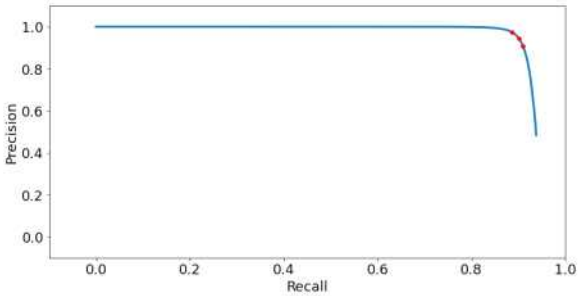

그림 1 얼굴 검색 실험의 PR 곡선

Threshold	Precision	Recall	F-score
0.79	0.9730	0.8731	0.9204
0.82	0.9389	0.8913	0.9145
0.84	0.8989	0.9001	0.8995

표 2 precision과 recall의 가중치에 따른 변화

결과에 의해 임계값이 0.79일 때, 최대의 F-score를 가지게 되는 것을 알 수 있다. 여기서 recall을 좀 더 중요한 지표로써 생각한다면 임계값을 조금 더 높여서 사용 할 수 있다. 임계값에 따른 지표 변화가 표 2에 나타나 있다.

4. 결론 및 향후 연구

본 논문은 얼굴 특징을 클러스터링 한 후, 그 클러스터들의 대푯값을 통해 인물이 존재하는지 검색하는 실험을 진행했다. HAC 알고리즘으로 충분한 클러스터링 성능을 낸다는 것을 보였고, 이 대푯값으로 실험을 했을 때, 높은 정확도가 나온다는 것을 보였다. 향후 연구에서는 검색 실험의 정확도를 더 높이는 데 집중할 것이다.

Acknowledgement

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2020-0-01891, 불법촬영물에서 특정 얼굴 검색 기술 개발)

참고문헌

[1] Wolf Lior, Tal Hassner, and Itay Maoz, "Face recognition in unconstrained videos with matched background similarity," in CVPR, 2011.

[2] Xiang An, Xuhan Zhu, Yuan Gao, Yang Xiao, Yongle Zhao, Ziyong Feng, Lan Wu, Bin Qin, Ming Zhang, Debing Zhang and Ying Fu, "Partial FC: Training 10 million identities on a single machine," in ICCV, 2021.

[3] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li and Wei Liu, "Cosface: Large margin cosine loss for deep face recognition," in CVPR, 2018.

[4] Jinkang Deng, Jia Guo, Niannan Xue and Stefanos Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in CVPR, 2019.