최신 ICT 이슈

III. IBM 100만 명 얼굴 정보 담은 데이터셋 공개, 인식의 공정성이 목표

■ **IBM 기초연구소는 최근 얼굴인식 기술의 공정성과 정확성 향상을 목표로 "얼굴의 다양성(Diversity in Faces: DiF)"이라는 데이터셋을 공개**

- 카메라와 알고리즘을 이용해 얼굴을 인식하여 개인을 식별하는 기술의 실용화는 최근 급속히 확대되고 있으며, 스마트폰의 잠금 해제에 Face ID 등의 얼굴 인식 시스템이 이용되기도 하고, 경찰의 수사 및 경비에 얼굴 인식 시스템이 도입되어 일부 성과를 올리고 있기도 함
- 그러나 얼굴 인식 시스템의 정밀도는 아직 완벽함과는 거리가 있는데, 가령 영국 경찰이 도입한 얼굴 인식 시스템의 경우 오판정률이 90%가 넘는 일도 있었음

<자료> BBC

[그림 1] 축구장에 설치된 자동 얼굴인식 시스템

- 2017년 6월 영국 카디프에서는 UEFA 챔피언스리그 결승전이 열렸는데, 대규모 스포츠 이벤트는 테러의 표적이 될 가능성이 높아 당시 뉴 사우스 웨일즈 경찰은 감시 카메라를 이용한 자동 얼굴인식 시스템으로 범죄 용의자를 탐지해 내겠다는 계획을 세웠음
- 결승전 당일 자동 얼굴인식 시스템에 의해 범죄 용의자로 지목된 사람은 모두 2,470명에 달했는데, 이 중 2,297명이 잘못된 판정으로 드러났음
- 웨일즈 경찰의 사례에서 오판정률이 높았던 원인은 인터폴에서 제공한 용의자 사진의 화질이 떨어졌기 때문인데 그 외에도 얼굴인식 시스템의 정확도는 여러 요인에 의해 영향을 받음
- 오인식 비율이 높아지는 원인 중에는 "얼굴인식 시스템의 편향성"도 있다는 지적이 있는데, 실제로 아프리카계 미국인은 백인보다 5~10% 정도 정확도가 떨어진다는 연구 결과가 있음
- IBM 기초연구소의 연구팀은 실제 얼굴인식 시스템의 성능에 영향을 미치는 것은 "내재적인

* 본 내용과 관련된 사항은 산업분석팀(☎ 042-612-8296)과 최신ICT동향 컬럼리스트 박종훈 집필위원(soma0722@naver.com ☎ 02-576-2600)에게 문의하시기 바랍니다.

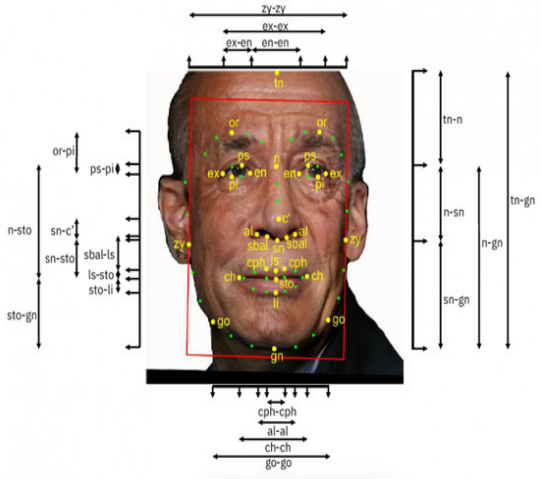
** 본 내용은 필자의 주관적인 의견이며 IITP의 공식적인 입장이 아님을 밝힙니다.

OPEN 출처표시 상업용금지 공공누리 공공저작물 자유이용허락

얼굴의 다양성"이지 인식의 정확도가 개인이나 집단에 따라 달라질 수는 없다는 입장이며, DiF 데이터셋의 공개 목적이 얼굴인식 기술의 '공정성' 향상이라고 한 것은 이런 연유임

■ **이번에 IBM이 공개한 DiF에는 다양한 성별과 인종의 얼굴 이미지와 그에 대한 주석이 포함되어 있으며, 약 100만 명분의 데이터가 포함되어 있다고 함**

- IBM 연구소는 만일 인종이나 성별에 따라 얼굴인식의 정확도에 차이가 발생한다면, 이는 알고리즘을 학습시킬 데이터셋이 특정 성별이나 인종에 편중되어 있기 때문이므로, 데이터셋의 표본을 늘리고 인종과 성별의 샘플링을 균형있게 함으로써 해결 가능하다고 보았음
- DiF 데이터셋의 각각 이미지에는 머리 모양과 얼굴의 대칭, 코의 길이, 이마의 높이 등 얼굴의 객관적 척도와 함께 연령이나 성별 등의 주석 데이터가 라벨링 되어 있다고 함
- 연구팀에 따르면 주석 데이터에는 얼굴 부위 47곳 이상의 크기와 특징이 정리되어 있으며, 이것이 얼굴인식 시스템의 공정성과 정확성을 향상시키고 알고리즘의 성능을 더욱 강력하게 만드는 학습 교사의 역할을 하고 있음



<자료> IBM

[그림 2] DiF 데이터셋의 정보 라벨링

- 연구팀은 IBM 기초연구소에서 보다 공정하고 정확한 얼굴인식 시스템을 개발하기 위한 연구 노력을 계속하고 있지만, 단지 자신들의 연구만으로는 얼굴인식 시스템이 발전할 수 없다고 판단하였음
- DiF를 공개한 것은 이를 통해 다른 곳에서 진행되는 얼굴인식 시스템의 연구에도 기여하는 것이 중요한 과학적 의제를 진전시키는 일이라고 보았기 때문이며, 이번에 공개된 DiF가 얼굴인식 시스템 연구의 새로운 첫 걸음이 되기를 희망하고 있음
- DiF 데이터셋은 전세계에 있는 얼굴인식 시스템 연구 커뮤니티에 제공되는데, DiF에 접속을 원하면 IBM 기초연구소의 설문 조사에 응하고 이메일로 신청하면 됨
- Diversity in Faces Dataset - Trusted AI - IBM Research AI (https://www.research.ibm.com/artificial-intelligence/trusted-ai/diversity-in-faces/)

■ **한편, DiF를 기반으로 한 시스템 개발 연구를 통해 얼굴인식의 공정성과 정확성 개선이라는 목표는 달성될 수 있겠지만, 그에 비례해 기술에 대한 통제 요구도 커질 것으로 보임**

- ▸ 2017년 뉴 사우스 웨일즈 경찰의 얼굴인식 시스템이 2,000명 이상의 시민을 범죄 용의자로 오판한 것이 알려지자, 영국에서는 즉각 시스템 폐지 운동이 전개된 바 있음
- ▸ 영국의 시민단체 '빅브라더 워치(Big Brother Watch)'는 실시간 얼굴 인식은 시민의 자유에 대한 위협일 뿐만 아니라 위험하고 부정확한 정치적 도구라며 강력히 비판하고 있는데, 기술의 정확성 여부와 상관없이 얼굴인식 기술에 대한 근원적 문제제기는 지속될 것으로 보임

[참고문헌]

[1] IBM, 1. 29, https://ibm.co/2Bbr6hc
[2] TechCrunch, 1. 29, https://tcrn.ch/2UrEuVp