

레터논문 (Letter Paper)

방송공학회논문지 제27권 제6호, 2022년 11월 (JBE Vol.27, No.6, November 2022)

<https://doi.org/10.5909/JBE.2022.27.6.944>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

데이터 증강을 통한 마스크 착용 얼굴 이미지에 강인한 얼굴 자세 추정

한 경 탁^{a)}, 홍 성 은^{a)†}

Robust Head Pose Estimation for Masked Face Image via Data Augmentation

Kyeongtak Han^{a)} and Sungeun Hong^{a)†}

요 약

최근 코로나바이러스로 인한 마스크 착용이 급증함에 따라 마스크 착용에 대응할 수 있는 기술의 중요성이 증가하고 있다. 얼굴 자세 추정 분야는 운전자 주의, 얼굴 정면화, 시선 감지 등의 다양한 활용성에도 불구하고 마스크 착용에 따른 성능 저하 문제를 해결할 수 있는 연구가 거의 수행되지 않았다. 본 논문은 마스크 착용 유무에 따른 얼굴 자세 추정의 성능 저하에 대한 분석을 토대로, 마스크 크가 없는 얼굴 이미지의 크기 및 자세를 분석하여 마스크 이미지를 합성할 수 있는 데이터 증강 기법을 제안한다. 제안하는 얼굴에 특화된 증강 기법을 활용한 학습은 마스크 착용 여부와 관계없이 얼굴 자세 추정 벤치마크 데이터 세트인 BIWI에서 강인한 성능을 보이며, 특정 모델에 국한되지 않기 때문에 다양한 얼굴 자세 추정 모델에 적용될 수 있다.

Abstract

Due to the coronavirus pandemic, the wearing of a mask has been increasing worldwide; thus, the importance of image analysis on masked face images has become essential. Although head pose estimation can be applied to various face-related applications including driver attention, face frontalization, and gaze detection, few studies have been conducted to address the performance degradation caused by masked faces. This study proposes a new data augmentation that synthesizes the masked face, depending on the face image size and poses, which shows robust performance on BIWI benchmark dataset regardless of mask-wearing. Since the proposed scheme is not limited to the specific model, it can be utilized in various head pose estimation models.

Keyword: Head Pose Estimation, Facial Mask, Facial Data Augmentation

a) 인하대학교 전기컴퓨터공학과(Department of Electrical and Computer Engineering)

† Corresponding Author : 홍성은(Sungeun Hong)

E-mail: csehong@inha.ac.kr

Tel: +82-32-860-7427

ORCID: <https://orcid.org/0000-0003-1774-9168>

※ 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF-2021R1F1A1054569). 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. RS-2022-00155915, 인공지능융합혁신인재양성(인하대학교)).

※ This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2021R1F1A1054569). This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2022-00155915, Artificial Intelligence Convergence Innovation Human Resources Development (Inha University)).

• Manuscript September 5, 2022; Revised October 7, 2022; Accepted October 7, 2022.

Copyright © 2022 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

I. 서론

2019년 이후 코로나바이러스의 팬데믹과 함께 마스크 착용이 강조되었다. 이후 얼굴 인식 분야에서는 마스크 착용에 대한 연구가 활발히 수행되었지만, 얼굴 자세 추정은 얼굴을 활용한 다양한 분야(예: 운전자 주의, 얼굴 정면화, 시선 감지, 딥페이크 등)에 다양하게 활용할 수 있음에도 불구하고 관련 연구가 거의 제안되지 않았다.

본 논문에서는 다양한 얼굴 포즈와 마스크를 착용하지 않은 얼굴 데이터 세트에 적용할 수 있는 얼굴 이미지에 특화된 데이터 증강 기법을 활용한 학습 제안함으로써, 마스크를 착용한 얼굴 이미지에서도 강인하게 동작하는 학습 및 시스템을 제안한다. 제안하는 학습 방법은 얼굴의 자세 추정 및 상대적인 크기 등을 분석하여 마스크를 합성할 수 있는 방법론을 제안한다. 이외에도, Yaw, Pitch, Roll에 대해서 임의로 각도로 이미지를 회전함으로써, 얼굴 이미지의 다양한 각도를 표현할 수 있는 얼굴 특화 증강 기법을 활용한 학습 방법을 제안한다. 제안하는 얼굴에 특화된 증강 기법을 활용하면 강인한 얼굴 자세 추정에 필요한 다양한 각도 이미지와 마스크를 착용한 얼굴 이미지를 생성할 수 있으며, 자체적으로 얼굴 자세 추정 데이터 세트를 구축하고 모델 검증을 수행한다. 본 논문은 마스크 합성을 통한 데이터 증강 기법 도입에 따른 정량적 성능 평가를 제시하고, 그동안 간과되어온 얼굴 데이터 세트의 편향된 얼굴 각도 분포에 따른 문제점을 완화할 수 있는 증강 기법의 효용

성을 정량적, 정성적으로 입증한다.

II. 본론

본 연구에서는 마스크 착용 증강 기법과 rotation matrix를 활용하여 한정된 Yaw, Pitch, Roll 분포를 다양한 분포로 표현하는 데이터 증강을 제안함으로써 마스크 착용 여부와 관계없이 강인한 성능을 보이는 학습 방법을 제안한다.

1. 마스크 착용 증강 기법

제안하는 마스크 착용 증강 기법은 그림 1에서와 같은 파이프라인에 따라서 수행된다. 입력 이미지가 주어지면 얼굴 영역 및 랜드마크 포인트를 검출하며, 이때 사용되는 모델은 사전 학습된 Retina-Face^[1]를 활용한다. 이후 탐지된 얼굴과 랜드마크를 활용해서 마스크의 키포인트를 계산하고 얼굴의 회전 각도 계산한다. 회전 각도의 일정 임계치에 따라서 템플릿에 존재하는 좌측 회전, 우측 회전, 정상의 마스크 중 하나를 선택하며 이후 추정된 얼굴에서의 마스크 착용 포인트와 마스크의 키 포인트를 Perspective Transformation 기법^[2]을 활용하여 합성한다. 제안하는 Perspective Transformation을 활용하면 얼굴 영역에 대한 변환 없이 마스크 영역만을 얼굴 마스크 포인트에 맞춰서 변환을 수행할 수 있기 때문에 원본 얼굴 이미지의 지속성을 유지할 수 있는 장점을 지닌다.

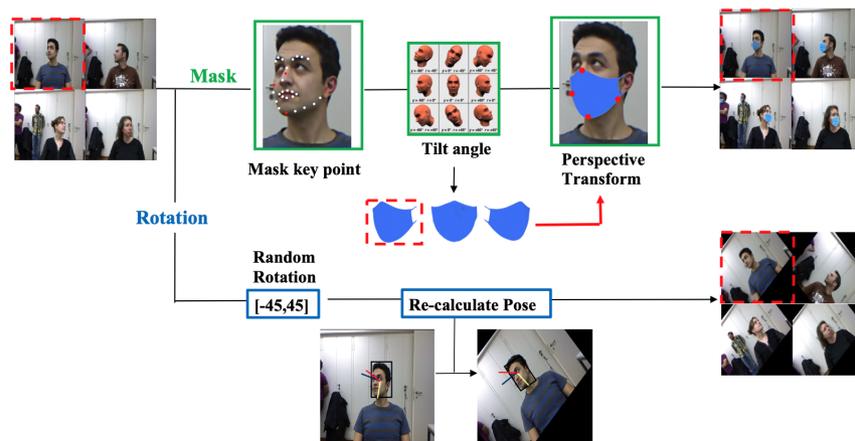


그림 1. 데이터 증강 기법 프레임워크

Fig. 1. Proposed data augmentation framework

2. 회전에 의한 증강 기법

기존 얼굴 자세 추정에서 널리 쓰이는 벤치마크 데이터 세트는 그림 2-(a)에서 볼 수 있듯이 매우 한정된 Yaw, Pitch, Roll로 표현되며, 이러한 한정된 각도의 라벨 분포는 얼굴 자세 추정 모델이 정면에 가까운 얼굴 위주로 과적합하게 만드는 문제점이 있다. 따라서 제한된 각도 분포의 데이터만을 학습에 활용하게 되면, 성능 개선에 제한적인 문제가 있다. 본 연구는 회전에 의한 데이터 증강 기법을 활용함으로써 그림 2-(b)와 같이 다양한 Yaw, Pitch, Roll 분포가 되게끔 한다^[3]. 제안 방법은 Yaw, Pitch, Roll마다 회전 각도 [-45, 45]를 임의로 선택한 후 해당 각도에 맞춰서 이미지를 회전시킨다. re-calculate pose 방법에서는 새로운 라벨 Yaw, Pitch, Roll은 변환 전 각도들을 활용해서 수식(1)번 처럼 Euler angle을 rotation 행렬 (R)로 구성하고, 수식(2)번 처럼 z축 기준으로 이미지 회전 각도만큼 rotation시키는 3x3 행렬을 구성하여 R 행렬과 행렬 곱셈을 수행하여 새로운 rotation 행렬 (R')을 구성한다. rotation 행렬 (R')을 활용하여 수식(3)처럼 최종 변환된 각도를 획득하며, 최종 변환된 각도는 얼굴 자세 추정의 라벨로서 활용된다. 그림

2-(a)에서의 기존 라벨 분포가 그림 2-(b)와 같이 변경되며, 그림 2-(c)는 회전에 의해 증강된 이미지 예시이다.

III. 실험 결과

제안하는 데이터 증강 기법의 효용성을 검증하기 위해서 얼굴 자세 추정 벤치마크 데이터 세트인 BIWI^[4]를 활용하였다. 실험 프로토콜은 1) 마스크 미착용 시나리오 (BIWI: 4,710장) 2) 마스크 착용 시나리오 (BIWI_masked: 10,474장)이다. 모델 학습 시 두가지 시나리오에서 동일하게 마스크 미착용 데이터 (BIWI-train: 10,990장)를 학습으로 활용하였으며, 6DRepNet^[5]에서 제안된 증강기법을 단계적으로 적용하여 성능 비교를 수행하였다. 그림 3은 제안하는 증강 기법을 활용한 합성 마스크 착용 예시이다. 성능 평가지표는 MAE를 활용한다. 표 1은 마스크 미착용 시나리오에서의 제안하는 기법의 정량적 성능 비교를 보여주며 표 2는 마스크 착용 시나리오에서의 성능 비교 결과이다. 표 1의 2행-4행에서 확인할 수 있듯이 제안하는 증강 기법을 단계적으로 적용하면 마스크 미착용 상황에서 표 1-4행처럼 마스크 미착용 상황에서 표1-2행의 베이스보다 1.45 MAE 성

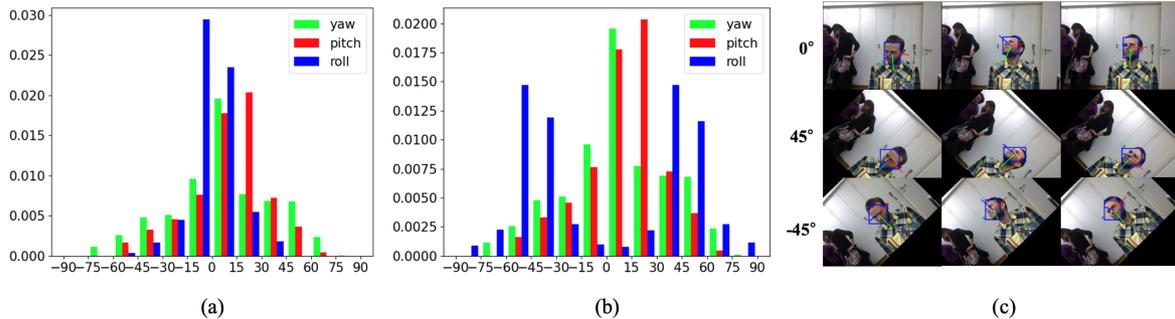


그림 2. (a) 회전 증강 기법 적용 전 각도 라벨 분포, (b) 회전 증강 기법 적용 후의 각도 라벨 분포, (c) 회전 증강 기법 적용 이미지 예시
 Fig. 2. (a) Distribution of angle labels before rotation enhancement, (b) Distribution of angle labels after rotation enhancement, (c) Example images of rotation enhancement

$$R = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\phi) & 0 & -\sin(\phi) \\ 0 & 1 & 0 \\ -\sin(\phi) & 0 & \cos(\phi) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\psi) & \sin(\psi) \\ 0 & -\sin(\psi) & \cos(\psi) \end{bmatrix}, \quad (1)$$

$$R' = R \cdot \begin{bmatrix} \cos(\gamma) & \sin(\gamma) & 0 \\ -\sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

$$\phi' = \arcsin R'_{31}, \quad \psi' = -\arctan \frac{R'_{32}}{R'_{33}}, \quad \theta' = -\arctan \frac{R'_{21}}{R'_{11}}, \quad (3)$$



그림 3. 제안 기법을 활용한 마스크 착용 합성 이미지 예시
 Fig. 3. Examples of mask-wearing synthetic images using the proposed augmentation

능 개선을 보였다. 이러한 결과는 회전에 의한 증강기법을 통한 다양한 각도에 대한 학습이 마스크 미착용 상황에서의 얼굴 자세 추정에 대한 일반화 능력을 개선시킨다는 점을 보여준다. 표 2에서 마스크 착용 상황에서도 증강 기법을 단계적으로 적용 시 많은 성능 개선을 보임을 확인할 수 있다. 결과적으로, 제안하는 얼굴에 특화된 증강 기법을 얼굴 자세 추정에 적용하면 마스크 착용 여부와 상관없이 일관되게 성능 개선을 보이며, 특히 실환경을 고려한 마스크 착용 시나리오에서는 성능 개선폭이 매우 큼을 확인할 수 있다.

표 1. 마스크 미착용 데이터 (BIWI)에서의 성능 비교
 Table 1. Performance comparison of the proposed augmentation on BIWI dataset

Model (증강 기법)	Yaw ↓	Pitch ↓	Roll ↓	MAE ↓
6DRepNet ^[6]	7.50	5.34	4.78	5.87
6DRepNet (Masked)	5.58	3.93	4.46	4.66
6DRepNet (Masked + Rotation)	5.43	4.17	3.74	4.45(-1.45)

표 2. 마스크 착용 데이터 (BIWI_masked)에서의 성능 비교
 Table 2. Performance comparison of the proposed augmentation on BIWI_masked dataset

Model (증강 기법)	Yaw ↓	Pitch ↓	Roll ↓	MAE ↓
6DRepNet ^[7]	13.95	16.74	8.64	13.11
6DRepNet (Masked)	5.41	4.74	3.39	4.51
6DRepNet (Masked + Rotation)	4.45	2.97	2.90	3.44(-9.67)

IV. 결 론

본 논문은 마스크 착용 상황에서 발생하는 성능 저하를 해결하기 위한 얼굴에 특화된 마스크 합성 기법과 회전에 의한 데이터 증강 기법을 소개한다. 실험에서 제안하는 얼굴에 특화된 증강 기법의 효용성을 BIWI 벤치마크 데이터 세트에서 정량적, 정성적으로 보인다. 제안하는 얼굴에 특화된 데이터 증강 기법은 특정 모델에 국한되지 않는 장점을 지니며 마스크 착용 여부와 관련 없이 일관된 성능 향상을 보이고, 특히 최근 주목 받고 있는 마스크 착용 시나리오에서는 기존 모델 대비 9.67 MAE의 성능 향상을 보였다.

참 고 문 헌 (References)

- [1] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild," Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, June, 2020.
doi: <https://doi.org/10.1109/CVPR42600.2020.00525>
- [2] A. Anwar and A. Raychowdhury, "Masked Face Recognition for Secure Authentication," arXiv preprint arXiv:2008.11104, 2020.
doi: <https://doi.org/10.48550/arXiv.2008.11104>
- [3] A. Sheka and V. Samun, "Rotation Augmentation for Head Pose Estimation Problem," IEEE Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology, pp. 0308-0311, 2021.
doi: <https://doi.org/10.1109/USBREIT51232.2021.9455063>
- [4] G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L. V. Gool, "Random forests for real time 3d face analysis," International journal of computer vision, vol. 101, no. 3, pp. 437-458, 2013
doi: <https://doi.org/10.1007/s11263-012-0549-0>
- [5] T. Hempel, A. A. Abdelrahman, and A AL-hamadi, "6D Rotation Representation For Unconstrained Head Pose Estimation," arXiv preprint arXiv:2202.12555, 2022.
doi: <https://doi.org/10.48550/arXiv.2202.12555>