

비전기술 기반 방역수칙 모니터링 시스템 개발 연구

이선훈¹, 정준욱¹, 오흥선^{1*}, 김선호²

¹한국기술교육대학교 컴퓨터공학과, ²KIST 안전증강융합연구단

A study on the development of a system for monitoring quarantine rules using vision technology

SeonHoon Lee¹, Jun-Uk Jung¹, Heung-Seon Oh^{1*}, Seonho Kim²

¹School of Computer Science Engineering, KOREATECH

²Research group for Augmented Safety Systems Intelligence Sensing Tracking (ASSIST), KIST

요약 본 연구는 비전기술을 이용한 방역수칙을 모니터링하는 시스템에 관한 논문이다. 본 논문에서 제안하는 시스템은 영상에서 추출한 정보로 사람들이 방역수칙을 제대로 준수하고 있는가를 자동으로 판별하는 것이다. 시스템의 구성은 딥러닝 모델을 이용하여 영상에서 모니터링되는 사람들의 3차원 정보와 방역수칙 위반요소를 추출한 후, 새롭게 정의한 사회적 거리를 통해 방역수칙 준수 여부를 시각화한다. 기존 사회적 거리 정의의 문제점은 물리적 거리 기반으로만 계산하기 때문에 물리적으로 거리만 가깝다면 사회적 거리도 가깝다고 계산된다. 이는 지하철역, 쇼핑몰과 같이 사람들이 상호작용없이 단순히 지나치는 상황에 대해서도 거리가 가깝다는 이유만으로 알람을 발생하기 때문에 지나치게 민감한 경고를 야기하는 문제가 있다. 이러한 문제를 완화하기 위해 개별적으로 판별하던 방역수칙위반 요소들을 크기 비교가 가능한 상대적 거리로 통합한 후, 상대적 거리와 물리적 거리를 통합한 사회적 거리 개념을 제안한다. 또한 상대적 거리 계산을 위한 마스크 검출에 있어 기존 학습 데이터셋과 실제 환경 간의 차이 때문에 마스크 검출 성능이 떨어지는 문제가 존재하는데 학습과 추론에 있어 mosaic과 다중 크기 어그리게이션을 통해 이 문제를 완화하는 방법을 제시한다.

Abstract This study is on a system for monitoring quarantine rules using vision technology. The system proposed in this paper automatically determines whether people are properly following quarantine rules with information extracted from videos. The system extracts three-dimensional information and the elements of a violation of quarantine rules of people monitored from a video using deep learning. It then visualizes compliance with quarantine rules through newly defined social distance. Since the problem of the existing social distance definition is calculated only based on only physical distance, if the distance is small, social distance is also calculated as close. This has a problem of causing overly sensitive alerts because it warns people when they are just physically close, even though they pass without any interaction, such as in a subway station or shopping mall. To alleviate this problem, we propose a concept of social distance that integrates relative distance and physical distance after integrating individually determined factors of quarantine rule violations into relative distance that can be compared. In addition, there is a problem of poor mask detection performance due to differences between the training dataset and the actual environment in detecting masks for calculation of the relative distance. We propose methods to alleviate this problem through mosaic and multi-aggregation in training and inference time.

Keywords : Vision-Based Monitoring, Social-Distance, Mask Detection, 3D Pose Estimation, Epidemic

본 논문은 2022년도 한국기술교육대학교 교수 교육연구진흥과제 지원에 의하여 연구되었음

*Corresponding Author : Heung-Seon Oh(KOREATECH)

email: ohhs@koreatech.ac.kr

Received May 2, 2022

Revised June 16, 2022

Accepted July 7, 2022

Published July 31, 2022

1. 서론

COVID-19의 주요 이슈 중 하나는 의료시스템의 총 용량을 넘어서지 않게 감염 확산을 방지하는 것이다. 이를 달성하기 위해선 방역수칙 준수가 매우 중요하며 최근 이런 추세에 따라 방역수칙 모니터링 자동화 시스템들이 [1-6] 등장하고 있다. 이러한 시스템들의 장점은 사람이 직접 모니터링해야 하는 일을 자동화한다는 것에 있지만 마스크를 착용했는지 혹은 서로 마주보고 대화를 하는 상태인지 등과 같은 방역수칙 위반 판별에 중요한 맥락 (context) 정보를 보지 않고 단순히 두 사람 간의 물리적 거리 (physical distance)를 기반으로 모니터링 하기 때문에 지나치게 많은 거짓알람(false positive)이 발생한다는 문제점이 있다[7]. 본 연구에선 기존 시스템들의 한계점을 해결하고자 비전기술을 이용하여 맥락을 고려하는 상대적 거리라는 개념을 정의했고 이를 탑재한 시스템을 제안하였다.

제안하는 시스템은 Fig. 1과 같이 3차원 정보추출, 방역수칙 위반요소 추출, 사회적 거리 추출 및 방역수칙 시각화로 구성된다. 3차원 정보추출은 딥러닝 방식을 이용한 3차원 깊이 정보를 추출하는 방식을 사용했고 방역수칙위반요소 판별은 마스크 착용 유무로 판별했으며 사회적 거리 추출 및 방역수칙 시각화에서는 방역수칙 위반요소들을 바탕으로 상대적 거리 (relative distance)를 추출하고 이를 사회적 거리 (social distance)로 변환 후 방역수칙 준수 여부 하는 방법을 제시한다.

본 논문의 기여점은 다음과 같다.

- 개별적으로 판별하던 방역수칙위반 요소들을 크기 비교가 가능한 상대적 거리로 통합한다.
- 상대적 거리를 물리적 거리와 융합해 거짓 알람에 강한 사회적 거리를 제안하였다.
- 방역수칙시스템을 위한 마스크 검출에 있어 기존의 학습데이터셋과 실제 환경 사이의 차이를 완화는 방법을 제시한다.
- 사람이 직접 CCTV를 보며 사회적 거리를 모니터링하는 업무를 자동화하는 시스템을 제안한다.

논문의 구성은 다음과 같다. 2장에서는 기존 비전기술 기반 방역수칙 시스템의 방법론과 문제점을 살펴보고 3장에서는 전체 시스템의 구성을 기술하였다. 4장에서는 시스템을 적용한 결과를 보여주며, 5장에서는 본 연구의 기대효과 및 향후 필요 연구를 제안하여 비전 기반 방역수칙 모니터링시스템의 방향성을 제시하고자 한다.

2. 관련 연구

2.1 비전기술을 이용한 물리적 거리 연구

비전기술을 기반으로 공간정보를 구하는 방법 중 기하학적 규칙을 이용해 지평면을 조감도(BEV: Bird's Eye View, 이하 BEV)로 변형하는 방법[2-4]이 있다. 변형하는 방법으로는 카메라 파라미터 정보를 이용해 지평면을 왜곡시켜 촬영 관점을 조감도로 변형한다. 이 방법의 장점은 지평면 추출이 정확한 경우 몇 개의 파라미터만으로 조감도를 형성할 수 있지만, 계단같이 평면이 아닌 곳에서는 적용할 수 없으며 지평면 추출에 이상치가 발생하는 경우 치명적인 오류가 발생한다는 단점이 있다.

다른 방법으로 학습을 통해 객체의 2차원 정보를 3차원 정보로 변환시키는 방법[5]이 있다. 이 방법의 장점은 객체별로 3차원 정보를 추출하기 때문에, 특정 객체에서 이상치가 발생해도 전체 시스템에 영향을 주지 않는다는 장점이 있지만, 학습하지 않은 대상에 대해선 성능이 떨어진다는 단점이 존재한다.

2.2 방역수칙위반요소 관련 연구

종래의 방역수칙준수 시스템은 비전기술을 통한 물리적 정보를 획득하여 방역수칙 위반여부를 판정한다 [1,3,4]. 물리적 거리에 기반한 기존 방법론들은 많은 거짓 알람이 발생하는데, 이에 대한 보완책으로 Bertoni 등은 공간학(proxemics)기반 F-formation을 사용하였다[5]. F-formation은 거짓 알람을 줄여준다는 장점이 있지만, 마스크 착용 여부 등 다양한 방역수칙 요소들을 통합하지 못한다는 단점이 있다. 다른 방식으로는 사람들의 밀집도를 판단해 출입량을 조절하는 시스템[2]과 마스크 착용 여부를 방역수칙에 포함시킨 시스템[6]도 있다.

2.3 마스크 검출 연구

COVID-19 이후 마스크 착용이 필수가 되면서 마스크 검출 연구가 활발히 이루어지고 있다. 최근 연구들은 딥러닝 기반 객체 검출 모델을 기반으로 마스크 데이터셋을 학습시켜 연구를 진행한다. 예를 들어, Jiang 등의 연구에서는 자체 마스크 데이터셋 Properly Wearing Masked Face Detection Dataset (PWFMD) 를 구축하고 이를 이용하여 객체검출 모델 yolo-v3 모델을 학습하였다[8,9]. RetinaFaceMask [10]는 마스크 데이터셋 MAsked FAcEs for Face Mask Detection (MAFA-FMD) [10]를 이용하여 최신 얼굴 검출 모델 RetinaFace [11] 모델을 학습하였다.

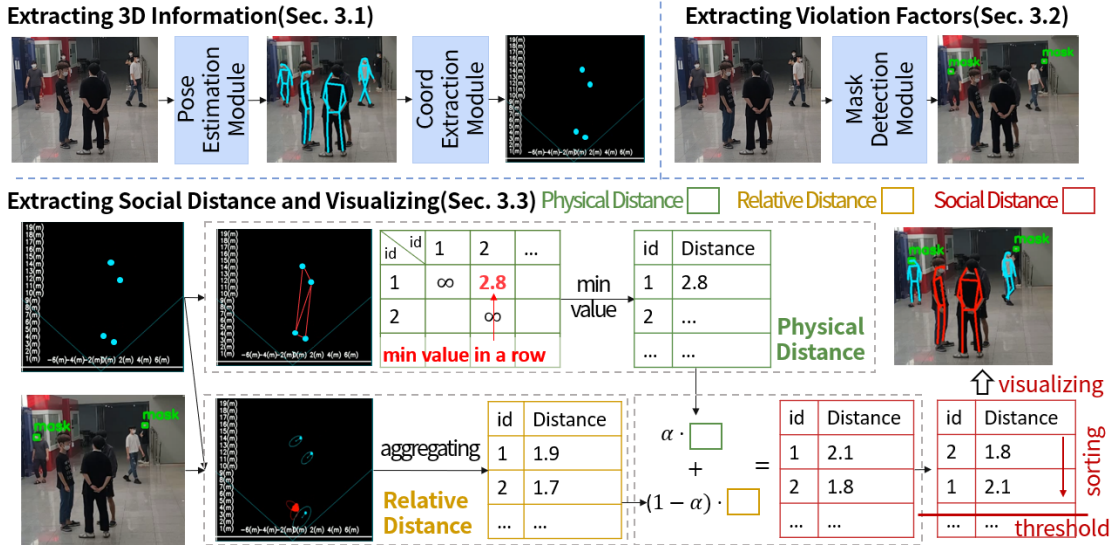


Fig. 1. quarantine rules monitoring system

3. 방역수칙 모니터링 시스템

제안하는 시스템은 Fig. 1과 같이 단일 이미지에서 사람들의 물리적 거리와 상대적 거리를 측정한다. 여기에서 물리적 거리는 실제 사람들 간의 유클리드 거리이며, 상대적 거리는 마스크 착용 여부 등과 같은 비물리적 방역수칙 위반요소들을 수치화 가능한 거리로 변환한 것을 의미한다. 최종적으로 물리적 거리와 상대적 거리를 조합하여 사회적 거리를 계산하고 이를 이용해 방역수칙 준수 여부를 판단한다.

3.1 3차원 정보 추출

본 연구의 3차원 정보추출 방법은 monoLoco[12]를 사용했고, 이미지에서 사람의 자세정보를 추출한 후 좌표추출 모듈을 통해 각 개인별 3차원 정보를 추출하는 과정을 거친다. 이후 3차원 정보 중 높이 정보를 제거하여 거리 공간(metric space)이 유지되는 조감도(BEV) 형태로 정보를 변환해 각 id 간의 물리적 거리를 측정한다. 이때 Fig. 1 하단부의 Physical Distance를 볼 수 있듯이 각 id 간의 물리적 거리는 대칭행렬(symetric matrix)이기 때문에 상삼각행렬(upper triangular matrix)로 만들었고 대각 성분은 자기 자신과의 거리이므로 크기를 무한대로 설정하여 시스템상에서 무시되게 설정했다.

3.2 마스크 검출 모듈

마스크 착용 여부 판별을 위해서 PWMFD [8]를 이용하여 최신 CNN 기반 검출 모델인 yolov4 [13]를 baseline 모델로 학습하였다. PWMFD는 마스크 착용, 미착용, 그리고 오착용의 3가지 클래스를 가지고 있다. 자체 분석결과 PWMFD 데이터셋은 개별 이미지 내 얼굴 개수 분포의 불균형과 Fig. 2와 같이 학습데이터와 대상 데이터와의 불일치 문제가 존재한다. 본 연구에서는 이 문제를 완화하기 위해 학습 시 이미지 4장을 합쳐 사용하는 mosaic [13] 방법을 사용하였고 학습과 추론 모두 다중 크기 어그멘테이션 [14] 방법을 적용하여 실제 환경처럼 작은 크기의 마스크도 검출 가능케 하였다.

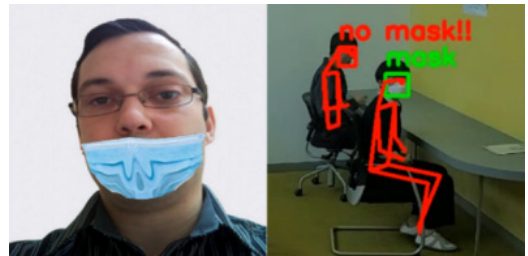


Fig. 2. left: mask data in PWMFD[8], right: masks detected in CCTV

3.3 사회적 거리측정

본 연구에서는 물리적 거리와 상대적 거리를 명시적으

로 구분하여 계산하고 그 결과를 조합하여 방역수칙 준수 여부를 판단할 사회적 거리를 계산한다. 관측되고 있는 사람의 사회적 거리를 S , 물리적 거리를 S_p , 상대적 거리를 S_R 이라고 할 때 사회적 거리 S 를 계산하는 식은 Eq. 1과 같다.

$$S = \alpha S_p + (1 - \alpha) S_R \quad (1)$$

여기서 α 는 물리적 거리 반영률을 의미한다. 물리적 거리 S_p 는 사람과 사람 간의 유클리드 거리이며 주변 사람 중 가장 가까운 사람과의 거리가 물리적 거리가 된다. 상대적 거리 S_R 는 Eq. 2와 같이 계산한다.

$$S_R = R \cdot \exp(-A_\mu) \quad (2)$$

Fig. 3처럼 두 사람의 상대적 거리는 타원의 겹친 면적이 클수록 가깝다. 계산의 중심이 되는 μ 는 실제 사람의 위치보다 전면부에 배치하는데 이를 기준으로 x 지점의 위험도를 계산한다. 이러한 배치 의도는 사람의 전면부 점수를 후면부 점수보다 더 높게 부여하기 위함이다. x 지점은 겹친 영역에 해당하는 좌표이고 이러한 x 지점이 모여 Eq. 2와 같이 겹친 영역을 계산한다. 여기서 R 은 상대적 거리의 스케일, A_μ 는 겹친 영역의 점수이며 Eq. 3과 같이 구하고 지점 x 에 대한 거리점수인 $S(x, \mu)$ 는 Mahalanobis 거리로써 Eq. 4와 같이 구한다.

$$A_\mu = \sum_x S(x, \mu) \quad (3)$$

$$S(x, \mu) = E \cdot \exp\left(- (x - \mu)^T \begin{bmatrix} \lambda & 0 \\ 0 & 4\lambda \end{bmatrix} (x - \mu)\right) \quad (4)$$

이때 E 는 환경 위험도를 의미하는데 사람의 밀집도가 높으면 감염 위험도가 크고 밀집도가 낮으면 감염 위험도가 낮아짐을 반영한다. λ 는 타원의 장축과 단축의

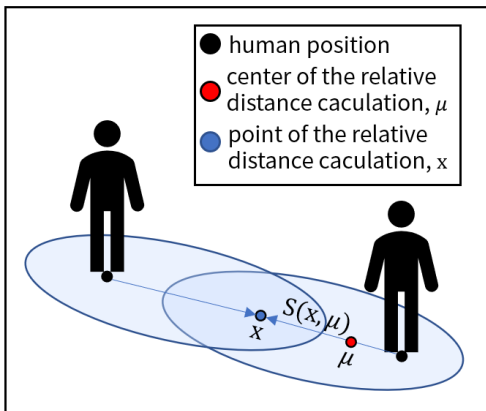


Fig. 3. the method of relative distance measurement

길이를 결정하는 인수으로써 각 값이 작아질수록 타원의 크기가 커진다. 예를 들어 마스크를 미착용하는 경우 값이 작아지면서 타원이 커지고, 타원이 커지는 것에 따라 겹치는 면적도 넓어지면서 상대적 거리가 가까워진다.

4. 방역수칙 모니터링 시스템 적용

4.1 마스크 검출 시스템 실험결과

Table 1는 PWMFD [8] 에서 보고한 마스크 검출 시스템의 성능을 나타낸다. YOLOv4-P5(our*)는 추론 시 다중 크기 어그멘테이션을 적용한 것을 의미한다. 다중 크기 어그멘테이션의 스케일 비율은 [0.67, 0.83, 1] 3가지로 적용하였다. Table 1에서 mAP₅₀의 성능은 YOLOv3+SE보다 1.4% 떨어지지만 엄격하게 검출영역을 잡는 지표인 mAP_{50:95}가 YOLOv3+SE보다 3% 향상된 것을 알 수 있다. 또한 다중 어그멘테이션을 사용했을 때가 사용하지 않았을 때보다 mAP_{50:95}이 2.8% 높아진 것을 보아 본 연구에서 적용한 다중어그멘테이션 방법이 마스크 검출에 효과적임을 알 수 있다.

Table 1. Face mask detection result

method	Backbone	Input-Size	mAP _{50:95}	mAP ₅₀
RetinaNet[15]	ResNet-50[18]	500	46.3%	94.2%
RetinaNet[15]	ResNet-50[18]	800	60.7%	94.9%
EfficientDet-D0[16]	Efficient-B0[19]	512	49.9%	84.5%
EfficientDet-D1[16]	Efficient-B1[19]	608	54.6%	85.1%
SSD[17]	VGG-16[20]	512	41.7%	92.7%
YOLOv3[9]	Darknet53	608	64.9%	96.1%
YOLOv3+SE[8]	SE-Darknet53	608	73.7%	99.5%
YOLOv4-P5(our)	CSPDarknet53	640	74.3%	97.7%
YOLOv4-P5(our*)	CSPDarknet53	640	76.1%	98.1%

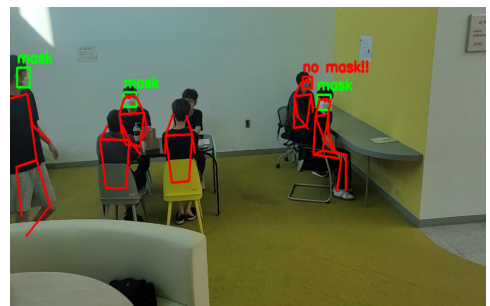


Fig. 4. result of mask detection

Fig. 4는 마스크 검출 결과를 보여준다. 실제 환경에서 촬영된 CCTV 영상 속 마스크 크기는 매우 작으며, Fig. 4의 마스크 평균 크기는 $33.75(px) \times 33.75(px)$ 이다. 이는 객체검출 분야에 있어 정확한 위치검출 능력을 요구하고 본 연구에서 제시하는 마스크 검출 모델이 이 능력을 충족함을 알 수 있다.

4.2 시스템 적용 결과

Fig. 5은 CCTV 이미지에 시스템을 적용한 결과를 시각화한 그림이다. 정상적으로 방역수칙을 지키고 있는 사람들은 청록색으로 자세 추정이 되었으며 방역수칙을 위반한 사람들은 빨간색으로 자세 추정이 되었다. Fig. 5 오른쪽 그림의 타원들은 사람들이 서로 겹치는 정도를 조감도로써 보여준다. 우측 상단부를 보면 타원이 서로 겹쳐 있다. 좌측 사진을 보면 서로 마주 보는 상황이 아니며 거리도 충분하기에 방역수칙위반 상황이 아니지만, 만일 기존 시스템처럼 단순히 영역이 겹쳤거나 일정 거리 안으로 들어왔을 때 경고를 하는 방식이었다면 거짓 알람이 발생할 것이다. 본 연구는 상대적 거리 개념을 적용하여 기존 시스템이 지나치게 민감한 경고를 완화하여 모니터링의 피로도를 감소시켰다.

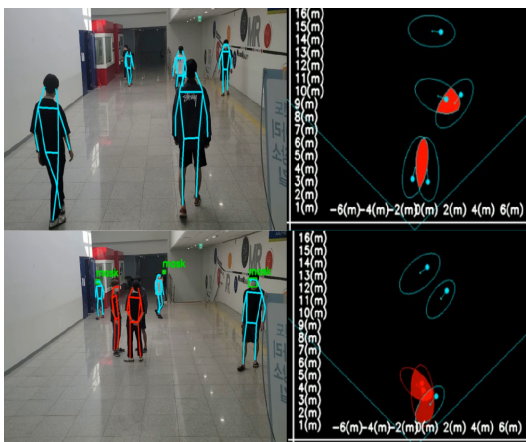


Fig. 5. result of system outputs

5. 결론

본 연구에서는 비전기술을 통해 실시간 방역수칙 모니터링 시스템을 연구하였다. 이는 사람이 직접 CCTV를 보며 사회적 거리를 모니터링하는 부담을 줄여주고 또한 눈대중으로 측정하던 거리를 시스템적으로 해결할 수 있

게 하였다.

기존 시스템들은 지나치게 많은 거짓 알람을 발생시키는 문제점을 본 연구의 시스템은 상대적 거리 개념을 통해 해결했다. 또한 상대적 거리 추출 과정에서 마스크 검출 모듈의 학습 환경과 실제 환경 간의 차이를 다중어그멘테이션과 mosaic을 통해 해결했다.

단안 카메라 기반 거리 측정 방법은 깊이 측정을 직접적으로 할 수 없기에 상황에 따라 거리 측정이 다소 부정확할 수 있다는 단점이 존재한다. 향후 연구에서는 단안 카메라와 ToF(Time of Flight)센서 등을 융합하여 보다 정밀한 거리 측정을 통해 방역수칙모니터링 연구를 지속해 나가기를 기대한다.

References

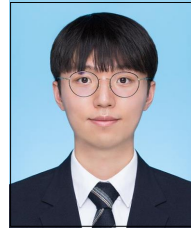
- [1] Ahmed, Imran, et al. "A deep learning-based social distance monitoring framework for COVID-19." *Sustainable Cities and Society* 65 (2021): 102571. DOI: <https://doi.org/10.1016/j.scs.2020.102571>
- [2] Yang, Dongfang, et al. "A vision-based social distancing and critical density detection system for COVID-19." *Sensors* 21.13 (2021): 4608. DOI: <https://doi.org/10.3390/s21134608>
- [3] Rezaei, Mahdi, and M. Azarmi. "Deepsocial: Social distancing monitoring and infection risk assessment in covid-19 pandemic." *Applied Sciences* 10.21 (2020): 7514. DOI: <https://doi.org/10.3390/app10217514>
- [4] Ahamad, et al. "Person detection for social distancing and safety violation alert based on segmented ROI." 2020 10th IEEE international conference on control system, computing and engineering (ICCSCE). IEEE, 2020. DOI: <http://dx.doi.org/10.1109/ICCSCE50387.2020.9204934>
- [5] Bertoni, Lorenzo, Sven Kreiss, and Alexandre Alahi. "Perceiving humans: from monocular 3d localization to social distancing." *IEEE Transactions on Intelligent Transportation Systems* (2021). DOI: <http://dx.doi.org/10.1109/TITS.2021.3069376>
- [6] Ramadass, et al. "Applying deep learning algorithm to maintain social distance in public place through drone technology." *International Journal of Pervasive Computing and Communications* (2020). DOI: <http://dx.doi.org/10.1108/IJPC-05-2020-0046>
- [7] M. Cristani, et al. A. (2020). The visual social distancing problem. *Ieee Access*, 8, 126876-126886. DOI: <http://dx.doi.org/10.1109/ACCESS.2020.3008370>
- [8] Xinbei, J, et al. "Real-time face mask detection method based on YOLOv3." *Electronics* 10.7 (2021): 837.

DOI: <http://dx.doi.org/10.3390/electronics10070837>

- [9] Redmon, et al. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).
- [10] X. F. et al. "RetinaFaceMask: A Single Stage Face Mask Detector for Assisting Control of the COVID-19 Pandemic." 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2021. DOI: <http://dx.doi.org/10.1109/SMC52423.2021.9659271>
- [11] J Deng, et al. "Retinaface: Single-shot multi-level face localisation in the wild." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. DOI: <http://dx.doi.org/10.1109/CVPR42600.2020.00525>
- [12] Bertoni, et al. "Monoloco: Monocular 3d pedestrian localization and uncertainty estimation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019. DOI: <http://dx.doi.org/10.1109/ICCV.2019.00696>
- [13] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934 (2020).
- [14] J. Park, S. Lee, et al. "Uncertainty-based Deep Object Detection from Aerial Images" Journal of Institute of Control, Robotics and Systems 26, 11 (2020) : 891-899. DOI: <http://dx.doi.org/10.5302/J.ICROS.2020.20.0131>
- [15] T.Y. Lin, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017. DOI: <http://dx.doi.org/10.1109/ICCV.2017.324>
- [16] M. Tan, et al. "Efficientdet: Scalable and efficient object detection." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. DOI: <http://dx.doi.org/10.1109/CVPR42600.2020.01079>
- [17] W. Liu, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016. DOI: http://dx.doi.org/10.1007/978-3-319-46448-0_2
- [18] K. He, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. DOI: <http://dx.doi.org/10.1109/CVPR.2016.90>
- [19] M. Tan, and L. Quoc. "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.
- [20] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition". In ICLR, 2015. 2, 3, 5, 6

이 선 훈(SeonHoon Lee)

[정회원]



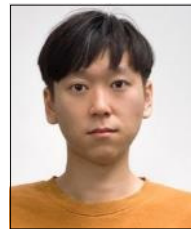
- 2020년 2월 : 한국기술교육대학교 컴퓨터공학과 (공학사)
- 2020년 3월 ~ 현재 : 한국기술교육대학교 컴퓨터공학과 (석박사 통합 재학)

<관심분야>

컴퓨터 비전, 3D object detection, visual SLAM

정 준 욱(Jun-Uk Jung)

[준회원]



- 2020년 2월 : 한국기술교육대학교 컴퓨터공학과 (공학사)
- 2020년 9월 ~ 현재 : 한국기술교육대학교 컴퓨터공학과 (석사재학)

<관심분야>

인공지능, 기계학습, 딥러닝, 영상처리, 얼굴인식

오 흥 선(Heung-Seon Oh)

[정회원]



- 2006년 2월 : 항공대학교 컴퓨터 공학 (공학사)
- 2009년 2월 : 한국과학기술원 전산학 (공학석사)
- 2014년 2월 : 한국과학기술원 전산학 (공학박사)
- 2018년 3월 ~ 현재 : 한국기술교육대학교 컴퓨터공학부 조교수

<관심분야>

딥러닝, 강화학습, 컴퓨터비전, 자연어처리

김 선 호(Seonho Kim)

[정회원]



- 1995년 2월 : 연세대학교 전산학과 (공학학사)
- 1997년 2월 : 포스텍 정보통신대학원 정보통신학과 (공학석사)
- 2008년 5월 : Virginia Tech, 공과대학, 컴퓨터과학 (공학박사)
- 1997년 3월 ~ 현재 : 한국과학기술정보연구원 (현재 책임연구원)

〈관심분야〉

자연어처리, 기계학습, 인공지능, 디지털콘텐츠