

# 영유아 이상징후 감지를 위한 표정 인식 알고리즘 개선

## The improved facial expression recognition algorithm for detecting abnormal symptoms in infants and young children

김윤수\*, 이수인\*, 석종원\*★

Yun-Su Kim\*, Su-In Lee\*, Jong-Won Seok\*★

### Abstract

The non-contact body temperature measurement system is one of the key factors, which is manage febrile diseases in mass facilities using optical and thermal imaging cameras. Conventional systems can only be used for simple body temperature measurement in the face area, because it is used only a deep learning-based face detection algorithm. So, there is a limit to detecting abnormal symptoms of the infants and young children, who have difficulty expressing their opinions. This paper proposes an improved facial expression recognition algorithm for detecting abnormal symptoms in infants and young children. The proposed method uses an object detection model to detect infants and young children in an image, then It acquires the coordinates of the eyes, nose, and mouth, which are key elements of facial expression recognition. Finally, facial expression recognition is performed by applying a selective sharpening filter based on the obtained coordinates. According to the experimental results, the proposed algorithm improved by 2.52%, 1.12%, and 2.29%, respectively, for the three expressions of neutral, happy, and sad in the UTK dataset.

### 요 약

비접촉형 체온 측정 시스템은 광학 및 열화상 카메라를 활용하여 집단시설의 발열성 질병을 관리하는 핵심 요소 중 하나이다. 기존 체온 측정 시스템은 딥러닝 기반 얼굴검출 알고리즘이 사용되어 얼굴영역의 단순 체온 측정에는 활용할 수 있지만, 의사표현이 어려운 영유아의 이상 징후를 인지하는데 한계가 있다. 본 논문에서는 기존의 체온 측정 시스템에서 영유아의 이상징후 감지를 위해 표정인식 알고리즘을 개선한다. 제안된 방법은 객체탐지 모델을 사용하여 영상에서 영유아를 검출한 후 얼굴영역을 추출하고 표정인식의 핵심 요소인 눈, 코, 입의 좌표를 획득한다. 이후 획득된 좌표를 기반으로 선택적 샤프닝 필터를 적용하여 표정인식을 진행한다. 실험결과에 따르면 제안된 알고리즘은 UTK 데이터셋에서 무표정, 웃음, 슬픔 3가지 표정에 대해 각각 2.52%, 1.12%, 2.29%가 향상되었다.

*Key words : Expression Recognition, Selective Sharpening, Edge Map, Object Detection, Face Detection*

\* Dept. of Information and Communication Engineering, Changwon National University

★ Corresponding author

E-mail : jwseok@changwon.ac.kr Tel : +82-55-213-3836

※ Acknowledgment

This work was supported by the Technology development Program(S3017456) funded by the Ministry of SMEs and Startups(MSS, Korea)

Manuscript received Aug. 22, 2021; revised Sep. 13, 2021; accepted Sep. 17, 2021.

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

### 1. 서론

2019년 12월에 발병한 COVID-19 바이러스와 같은 발열성 질병이 급격히 유행한 이후 음식점, 건물 및 행사장 입구 등 인원이 밀집되는 공간에 발열성 질병에 대한 다양한 측정 방법들이 사용되고 있다. 그 중 발열성 질병 조기 발견에 대중적으로 운용되는 방법으로 비접촉식 체온 측정 방식이 있으며, 이는 적외선 센서를 활용하여 체온 측정을 진행한다. 일반적으로 사용되는 방식으로 일정한 거리를 둔 상태에서 적외선 센서가 체온을 측정하며, 카메라를 활용한 방법과 센서만 설치하여 측정하는 방식이 있다. 카메라를 활용한 방법의 경우 신체 전체를 고려하지 않고, 주로 피부가 노출되는 얼굴 영역을 검출한 뒤 측정을 진행한다.

얼굴검출(Face Detection) 알고리즘의 경우, 초기 얼굴 모양에 대한 환경적인 영향으로 얼굴 내 변동성이 크기 때문에 단계별로 비얼굴 영역을 배제할 수 있는 기법이 선택 되어졌다. 그 중, HOG의 히스토그램을 기반으로 Cascade 분류기를 통해 사람의 얼굴에 해당하는 특징들을 학습시켜 인지할 수 있도록 구현하였다[1]. 하지만 이 방법은 모든 픽셀의 RGB 값을 보고 영상을 처리하는 방식으로 진행되므로 성능 대비 많은 연산시간이 소요되는 단점이 있어 Haar와 같이 특징을 기반으로 영상을 고속처리 하는 방법도 연구되어왔다[2].

그러나 위와 같은 객체 검출 혹은 얼굴 인식 등과 같은 고수준의 작업의 경우 머신러닝 기법은 분명한 한계가 존재하였다. 이를 영상의 특징을 사람이 직접 설계 및 추출해주어 머신러닝 알고리즘으로 학습을 진행하는 구조의 문제라 판단하였다. 따라서 목표에 적합한 특징을 알아서 추출하고 학습까지 진행하는 심층 인공지능망, 즉 딥러닝 모델을 적용시켰고, 전보다 더욱 높은 성능을 보여주었다[3]. 그중 특히 합성곱 신경망(CNN, Convolutional Neural Network)과 같은 이미지 처리 특화 인공지능망 기반 모델들이 물체 검출(Object Detection)을 위해 고안되었다. R-CNN은 선택적 검색>Selective search) 알고리즘을 사용해 관심 영역을 추출해 후보 영역들을 CNN 모델에 입력하는 방식으로 물체 검출을 구현하였으며[4], 후보 영역들을 모두 모델에 입력시켜 출력률 얻어 내야하기 때문에 속도 측면에서 느려지는 단점을 극복하기위해 후보 관심

영역의 추출 없이 입력에서 다중 객체를 한 번에 검출하도록 구성된 YOLO(You Only Look Once) 모델로 성능은 유지하며 높은 속도의 객체검출이 가능하게 되었다[5]. 나아가 하나의 특징맵만을 출력하는 YOLO와 달리 다중 특징맵(Multi-Feature map)을 출력하도록 한 SSD(Single Shot Detector)가 개발되어 속도와 정확도면에서 더욱 향상된 검출 성능을 보여주었다[6]. 이와 같은 객체검출 모델들을 얼굴이 포함된 데이터셋에 훈련시켜 실생활에 적용이 가능한 얼굴인식기반 비접촉 체온 측정 시스템이 개발될 수 있도록 하였다.

일반적인 경우에선 얼굴영역을 검출하여 온도 측정으로 건물로 출입할 때 발열성 질병의 감염 여부를 조기 진단할 수 있고 의사표현이 원활한 성인의 경우 몸이 아프거나 발열 발생시, 주변 사람에게 상태를 말하고 조치를 진행할 수 있다.

하지만 집단생활을 하는 사회적 약자(소아 및 청소년)의 경우 자기표현에 서툴기 때문에 의사소통이 어려워 주변 담당자에게 이상징후를 전달함에 어려움이 많다.

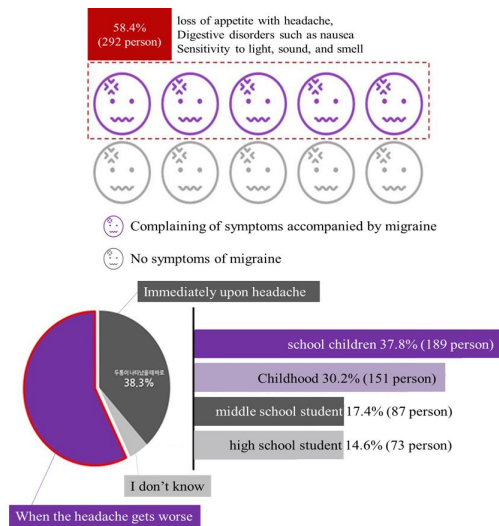


Fig. 1. Survey on the childhood and adolescent headaches [The Korean Headache Society].

그림 1. 소아 청소년 두통 실태 조사 [대한두통학회]

그림 1과 같이 대한두통학회의 소아 청소년 두통 실태 조사 결과 학동기 및 학동전기의 경우 참다가 두통이 심해지면 주변 담당자에게 이상징후를 전달하므로 체온 측정만으로 조기 검출에 애로가 있고, 발열 전염병의 경우 증상 악화 후 대처를 진행했을 때 집단감염으로 발전될 가능성이 농후하다.

따라서 본 논문에서는 사회적 약자인 영유아의 의사표현의 어려움을 고려하여 집단감염을 조기 예방할 수 있는 개선된 표정 인식 알고리즘을 제안한다.

## II. 본론

### 1. Relation Method

#### 가. YOLO

YOLO 알고리즘은 입력 영상에 그리드를 적용하여 경계 상자를 다차원적으로 배치하고, 확률 맵을 구성하여 최종 객체를 검출한다. 아래와 같은 단계로 구성되며 입력된 영상을 448×448 크기로 영상 크기 조절을 우선 수행한다.

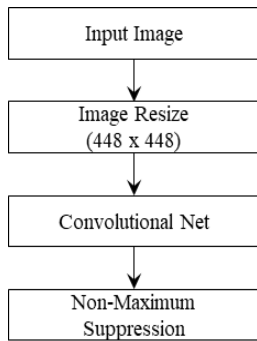


Fig. 2. YOLO flow chart.  
그림 2. YOLO 순서도

#### (1) Network Architecture

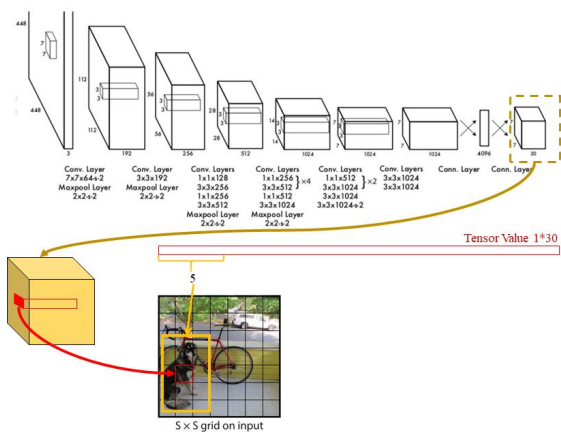


Fig. 3. YOLO network structure.  
그림 3. YOLO 네트워크 구조

YOLO 네트워크 구조는 GoogLeNet 모델을 기반으로 영상 분류기가 운영되며 24개 컨볼루션 계층과 2개의 완전 결합 계층으로 구성된다.

그림 3번과 같이 네트워크의 최종 출력인 7×7×30 특징맵에서 입력 영상에 적용된 그리드별 경계 상자와 신뢰도 지수, 클래스별 예측 값을 확인 할 수 있다.

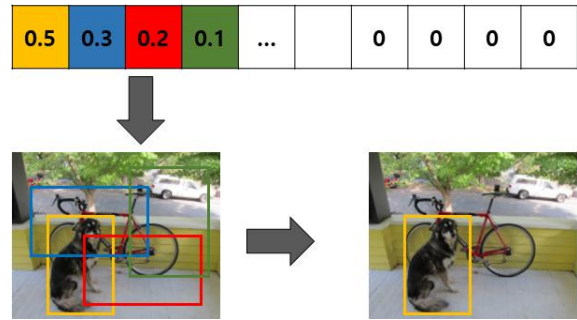


Fig. 4. Application Example of Non-Maximum Suppression.  
그림 4. Non-Maximum Suppression 적용예시

그림 4와 같이 최종적으로 Non-Maximum Suppression 진행하여 동일한 물체에 중복 생성된 경계 박스 중에서 가장 확률이 높은 것만 그려준다.

#### 나. MTCNN

MTCNN 알고리즘은 아래와 같은 단계로 구성되며 각 단계별 상이한 영상크기에 맞추어 영상 크기 조절을 우선 수행한다.

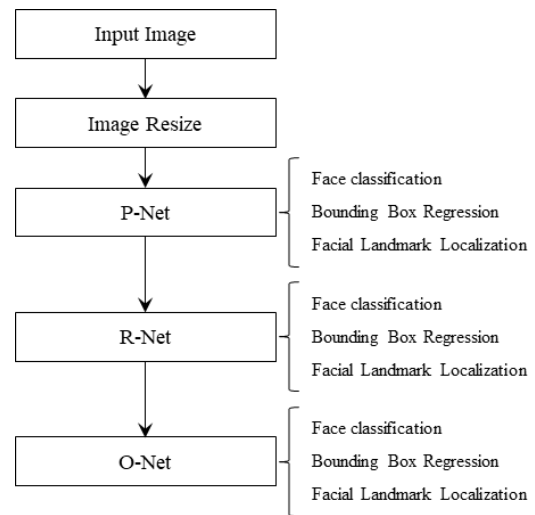


Fig. 5. MTCNN flow chart.  
그림 5. MTCNN 순서도

#### (1) Face classification

후보 창과 경계 상자 회귀 벡터를 얻기 위해 Deep Dense Face Detector(DDFD)와 유사한 컨볼루션

네트워크를 활용한다[7].

(2) Bounding Box Regression

Resize된 각 영상에 대하여 컨볼루션 네트워크 이후 얼굴영역으로 추정된 경계 상자에 회귀 벡터를 사용한다. 이후 각 후보를 보정하고, 각 후보 창에 대해 가장 가까운 경계 상자의 왼쪽 상단, 높이 및 너비와의 오프셋을 예측하며 수식은 다음과 같다.

$$L_i^{Bbox} = \|\hat{y}_i^{Bbox} - y_i^{Bbox}\|_2^2 \tag{1}$$

여기서  $\hat{y}_i^{Bbox}$  는 컨볼루션 네트워크에서 획득한 회귀 타겟,  $y_i^{Bbox}$  실제 좌표이다.

(3) Facial Landmark Localization

앞서 획득된 얼굴영역에서 눈(2), 코(1), 입꼬리(2)의 x, y 좌표 landmark localization 값을 획득하며 예시는 아래와 같다.

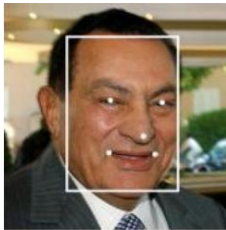


Fig. 6. Landmark localization of face region [8].  
그림 6. 얼굴영역에서 landmark localization 검출 [ 8]

landmark localization 수식은 다음과 같다.

$$L_i^{mark} = \|\hat{y}_i^{mark} - y_i^{mark}\|_2^2 \tag{2}$$

여기서  $\hat{y}_i^{mark}$  는 컨볼루션 네트워크에서 획득한 Facial landmark 좌표,  $y_i^{mark}$  실제 좌표이다.

Resize 된 각 영상에서 획득된 좌표 값을 원영상의 좌표로 변환해주며, Non-Maximum Suppression을 적용하여 가장 높은 확률만 남기고 제거를 수행한다.

다. MicroExpNet

고성능 얼굴 표정인식 모델들이 CK+, Oulu-Casia와 같은 공개된 데이터셋에서 사람보다 높은 인식률이 검증되었다. 그러므로 인식 정확도를 높이기보다는 연산량을 절감하기 위해 모델의 경량

화 연구가 지속되었고, MicroExpNet은 지식 증류(Knowledge Distillation)를 차용한다[9].

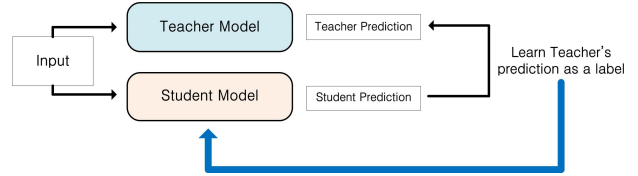


Fig. 7. Concept of Knowledge Distillation.

그림 7. 지식 증류 개념도

그림 7은 지식 증류 학습 방법에 대한 개념도이며 전체 모델은 Teacher 모델, Student 모델 2가지 서브모델로 나뉘어진다. Teacher 모델은 Go-ogLeNet과 같은 크고 무거운 모델, Student 모델은 Teacher 모델보다 상대적으로 훨씬 작고 가벼운 모델로 구성되며, 수식은 다음과 같다.

$$p_t = \frac{e^{z_i/T}}{\sum_j e^{z_j/T}}, p'_s = \frac{e^{v_i/T}}{\sum_j e^{v_j/T}}, p_s = \frac{e^{v_i}}{\sum_j e^{v_j}} \tag{3}$$

여기서  $p_t$  는 Teacher softmax의 soft 출력,  $z_i$  는 Teacher의 logit,  $p_s$  와  $p'_s$  는 Student softmax의 hard와 soft 출력,  $T$  는 온도이다. 각 모델은 입력된 데이터에 해당하는 클래스를 예측하며 학습을 통해 사전의 학습된 복잡한 모델의 가중치를 증류하여 보다 가벼운 모델에 옮기는 것이 가능하게 된다.

3. Proposed Method

제안된 알고리즘의 경우 그림 8과 같은 순서도로 구성되고, (1) 객체검출, (2) 얼굴영역 검출, (3) 선택적 샤프닝, (4) 표정인식으로 진행되며 객체검출과 얼굴영역 검출의 경우 각각 YOLO, MTCNN 알고리즘을 활용한다[5, 8].

객체검출 수행 시, 원본 영상에서 배경이 제거된 객체를 도출할 수 있으며, MTCNN 알고리즘을 활용하여 얼굴영역을 검출할 수 있다. 얼굴영역 검출 진행 시, Facial Landmark Localization을 적용하여 표정인식의 핵심요소인 “눈, 코, 입”의 위치정보를 획득한다.

획득된 위치정보를 기반으로 그림 9와 같이 관심영역(ROI, Region Of Interest)을 표시한다. 표정인식에 있어 눈, 코, 입 영역이 핵심요소이며 그 외의 특징이 두드러질 경우 인식을 저하의 요인이 되므

로, 원영상에 대한 선택적 샤프닝을 다음과 같이 수행한다.

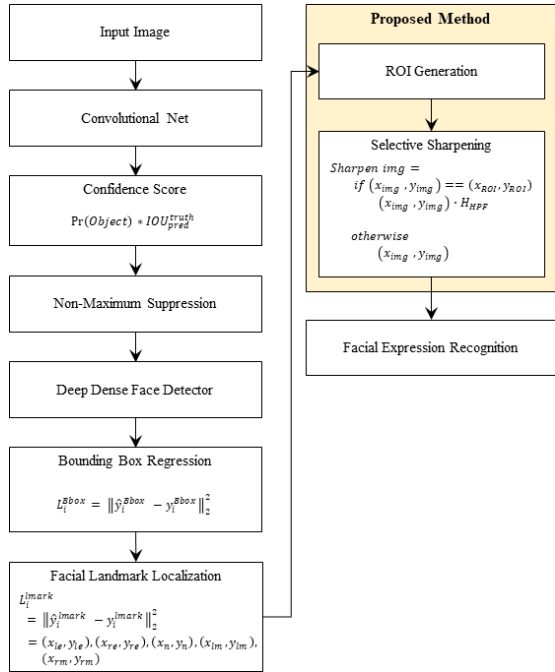


Fig. 8. Block Diagram of Proposing Algorithm.  
그림 8. 제안된 알고리즘 순서도

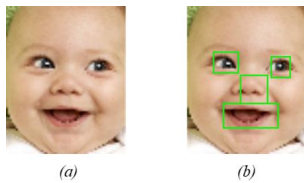


Fig. 9. Eye, nose, mouse location information based Region of Interest (a) Face detected image, (b) Region of interest displayed image..

그림 9. 눈, 코, 입 위치정보 기반 관심영역 표시  
(a) 얼굴검출 영상, (b) 관심영역 표시영상

$$\text{Sharpening } img = \begin{cases} (x_{img}, y_{img}) = (x_{ROI}, y_{ROI}) \\ (x_{img}, y_{img}) \cdot H_{HPF} \\ \text{otherwise} \\ (x_{img}, y_{img}) \end{cases} \quad (3)$$

여기서  $x_{img}, y_{img}$  는 각각 원영상의 좌표,  $x_{ROI}, y_{ROI}$  는 관심영역의 좌표,  $H_{HPF}$  는 고대역 필터이다. 고대역 필터는 가장 기본적으로 사용되는 필터를 적용하였으며, 다음과 같이 구성된다.

$$H_{HPF} = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{pmatrix} \quad (4)$$

선택적 샤프닝을 수행한 결과 그림 10과 같이 관심영역의 Edge가 향상되는 것을 확인할 수 있다.

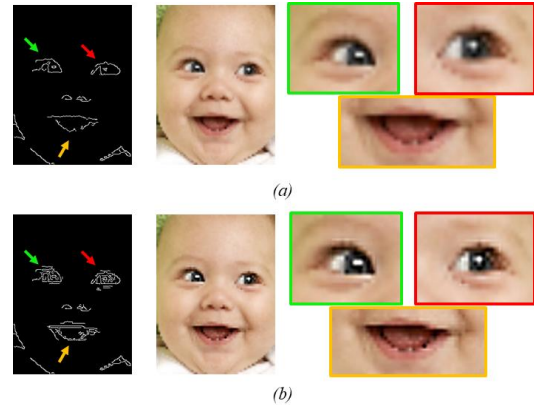


Fig. 10. Application result of ROI region sharpening  
(a) Original image, (b) Selective sharpened image.  
그림 10. ROI 영역 샤프닝 적용 결과  
(a) 원영상, (b) 선택적 샤프닝 적용영상

#### 4. Experimental results

본 논문에서 제안하는 알고리즘의 성능을 시험함에 있어 UTK face dataset에서 임의로 테스트 케이스 10개를 선별한다. 하나의 테스트 케이스에 레이블(감정) 당 10개, 총 30개의 데이터가 준비된다. 따라서 전체 데이터셋을 300개를 선별하여 성능을 검증하였다.



Fig. 11. Example of test dataset samples.  
그림 11. 테스트 데이터 샘플 예시

그림 11은 10개의 테스트 케이스 중 하나의 예시를 보여준다. 감정 표현이 명확하지 않은 영유아의 특성상 무표정(Neutral), 슬픔(Sad), 행복(Happy) 3가지의 뚜렷한 감정만으로 데이터셋 레이블을 구성하였다.

시험에 사용된 모델은 Oulu-CASIA 데이터셋에 사전 훈련된 MicroExpNet[10]을 사용하였다. Teacher 네트워크는 Inception v3, Student 네트워크는 CNN, FC(Fully-Connected) 층 순으로 각 2개씩 이어져 구성되어 있다.

Table 1. Average recognition rate of Conventional Method.

표 1. 기존 방법의 평균 인식률

Case	Neutral	Sad	Happy
CM(Test Case 1)	98.76	98.84	96.71
CM(Test Case 2)	94.14	93.65	89.93
CM(Test Case 3)	97.38	93.86	93.48
CM(Test Case 4)	93.11	95.37	91.79
CM(Test Case 5)	93.15	94.82	92.59
CM(Test Case 6)	94.89	92.47	95.92
CM(Test Case 7)	88.13	93.35	91.97
CM(Test Case 8)	85.11	96.59	95.23
CM(Test Case 9)	96.66	93.43	89.19
CM(Test Case 10)	90.94	98.02	94.34
Average	93.23	95.04	93.12

Table 2. Average Recognition rate of Proposing Method.

표 2. 제안하는 방법의 평균 인식률

Case	Neutral	Sad	Happy
PM(Test Case 1)	97.92	<b>99.27</b>	94.55
PM(Test Case 2)	<b>98.78</b>	<b>95.81</b>	<b>92.20</b>
PM(Test Case 3)	<b>97.69</b>	<b>97.54</b>	93.36
PM(Test Case 4)	<b>96.35</b>	<b>97.26</b>	<b>93.99</b>
PM(Test Case 5)	<b>97.58</b>	<b>96.87</b>	<b>98.05</b>
PM(Test Case 6)	<b>91.57</b>	88.56	<b>97.60</b>
PM(Test Case 7)	<b>96.51</b>	91.98	<b>93.99</b>
PM(Test Case 8)	<b>88.47</b>	<b>99.55</b>	<b>99.61</b>
PM(Test Case 9)	<b>98.87</b>	<b>98.16</b>	<b>96.37</b>
PM(Test Case 10)	<b>93.79</b>	96.60	<b>94.38</b>
Average	<b>95.75</b>	<b>96.16</b>	<b>95.41</b>

표 1, 2는 영유아를 대상으로 검출된 얼굴영역을 그대로 표정 분류 모델에 입력시키는 기존방식과 선택적 샤프닝을 적용시키는 제안된 방식의 표정 인식률 실험결과를 보여준다. 위 결과와 같이 기존 방식 보다 제안방식이 전체 테스트 케이스에서 무표정, 슬픔, 행복 순으로 2.52%, 1.12%, 2.29%의 인

식률이 향상되었으며, 모든 감정에 대해 평균 약 2%의 성능이 개선되었음을 확인할 수 있다.

### III. 결론

본 논문에서는 영유아 이상징후 감지를 위한 표정 인식 알고리즘 개선 방법을 제시하였다. 제안된 방법은 표정인식을 수행하기 전, 관심영역을 생성하고 핵심 요소인 눈, 코, 입의 영역을 고려하여 중요 영역만을 개선하기 위해 좌표 값 기반 선택적 샤프닝을 수행하였다. 우리는 감정 표현이 명확하지 않은 영유아의 특성을 고려하여 무표정, 슬픔, 행복 3가지의 뚜렷한 감정 영상을 선별 후 테스트를 수행했으며, 제안된 방법은 10개의 테스트 케이스에서 무표정, 슬픔, 행복에 대한 표정인식률이 기존 방법 대비 평균적으로 각각 2.52%, 1.12%, 2.29% 향상됨을 확인하였다. 실험을 살펴보면 관심영역의 좌표 값을 고려한 선택적 샤프닝 기법이 표정인식률 향상에 효과적이지만, 성능 최적화를 위해서 약한 특징점도 향상할 수 있는 추가 연구가 필요하다고 판단된다. 향후 현재 개발중인 임베디드 기반 광학 및 열화상 카메라에 포팅하여 테스트한다면 영유아뿐만 아니라 취약계층(노인, 장애인 등)에도 활용할 수 있을 것으로 기대된다.

### References

[1] H Yang, XA Wang, "Cascade classifier for face detection," *Journal of Algorithms & Computational Technology*, Vol.10, No.3, pp.187-197, 2016. DOI: 10.1177/1748301816649073

[2] YUSTIAWATI, Ratna, et al. "Analyzing of Different Features Using Haar Cascade Classifier." *2018 International Conference on Electrical Engineering and Computer Science (ICECOS)*, pp.129-134, 2018. DOI: 10.1109/ICECOS.2018.8605266

[3] I. Masi, Y. Wu, T. Hassner and P. Natarajan, "Deep Face Recognition: A Survey," 2018 31st SIBGRAP Conference on Graphics, Patterns and Images (SIBGRAP), pp.471-478, 2018. DOI: 10.1109/SIBGRAP.2018.00067.

- [4] R. Girshick, J. Donahue, T. Darrell, and J. Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation," *In CVPR(IEEE Conference on Computer Vision and Pattern Recognition)*, 2014.  
DOI: 10.1109/CVPR.2014.81
- [5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. "You only look once: Unified, real-time object detection" arXiv preprint arXiv:1506.02640, 2015.
- [6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. "SSD: Single shot multibox detector," *In Eur. Conf. Comp. Vis. (ECCV)*, pp.21-37, 2016.  
DOI: 10.1007/978-3-319-46448-0\_2
- [7] FARFADE, Sachin Sudhakar; SABERIAN, Mohammad J.; LI, Li-Jia. "Multi-view face detection using deep convolutional neural networks." *In: Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*. pp.643-650, 2015. DOI: 10.1145/2671188.2749408
- [8] ZHANG, Kaipeng, et al. "Joint face detection and alignment using multitask cascaded convolutional networks." *IEEE Signal Processing Letters*, Vol.23, No.10, pp.1499-1503, 2016.  
DOI: 10.1109/LSP.2016.2603342
- [9] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," *in WACV*, 2016.  
DOI: 10.1109/WACV.2016.7477450
- [10] C. Ilke, S. Eren, A. Emre, "MicroExpNet: An Extremely Small and Fast Model For Expression Recognition From Face Images," *Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp.1-6, 2019.  
DOI: 10.1109/IPTA.2019.8936114.

## BIOGRAPHY

### Yun-Su Kim (Member)



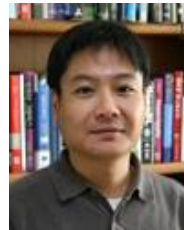
2020 : BS degree in Dept. of Information & Communication Engineering, Changwon National University.  
2020~ : MS degree in Dept. of Information & Communication Engineering, Changwon National University.

### Su-In Lee (Member)



2018~present: Senior Research Engineer, Software R&D Team, YOUANDI  
2020~present: Integrated PhD program student in Dept. of Information & Communication Engineering, Changwon National University.

### Jong-Won Seok (Member)



1995 : MS degree in Electronics Engineering, Kyungpook National University  
1999 : Ph.D degree in Electronics Engineering, Kyungpook National University  
2004~ : Professor, Changwon National University.