

<https://doi.org/10.7236/IIBC.2018.18.1.113>

IIBC 2018-1-16

적응적 딥러닝 학습 기반 영상 인식

Image Recognition based on Adaptive Deep Learning

김진우*, 이필규**

Jin-Woo Kim*, Phill-Kyu Rhee**

요약 사람의 감정은 다양한 요소에 의해서 드러난다. 말, 행동, 표정, 옷차림 등등. 하지만 사람은 자신의 감정을 숨길 줄 안다. 따라서 어느 한 가지만으로는 쉽게 그 감성을 짐작할 수 없다. 우리는 이러한 문제를 해결하고 보다 진실한 사람의 감성을 파악하기 위해 행동과 표정에 주의를 기울이기로 하였다. 행동과 표정은 부단한 노력과 훈련이 없으면 쉽게 감출 수 없기 때문이다. 본 논문에서는 딥러닝 방법을 통해 적은 데이터를 가지고 점진적으로 사람의 행동과 표정을 학습하여 두 가지 결과의 조합을 통해 사람의 감성을 추측하는 알고리즘을 제안한다. 이 알고리즘을 통해 우리는 보다 종합적으로 사람의 감성을 파악할 수 있다.

Abstract Human emotions are revealed by various factors. Words, actions, facial expressions, attire and so on. But people know how to hide their feelings. So we can not easily guess its sensitivity using one factor. We decided to pay attention to behaviors and facial expressions in order to solve these problems. Behavior and facial expression can not be easily concealed without constant effort and training. In this paper, we propose an algorithm to estimate human emotion through combination of two results by gradually learning human behavior and facial expression with little data through the deep learning method. Through this algorithm, we can more comprehensively grasp human emotions.

Key Words : Synthetic Emotion Recognition, Deep Learning, Facial Expression Recognition, Action Recognition

1. 서론

자율주행 자동차, 인공지능 스피커, 사물 인터넷(Internet of Things) 등 이제는 익숙해진 기계학습 및 인공지능 기술이 우리의 삶에 스며들고 있다. 이러한 기술은 사람의 얼굴 표정, 음성, 나이, 성별, 행동 등 다양한 것을 인식하는 기술이 발전하면서 가능해졌다.

이러한 인식 기술은 오래 전부터 쭉 연구되어 왔지만 최근 대두된 컨볼루션 신경망^{[12][14]}, 딥러닝 기술^{[13][15]}로 인해 크게 발전하였다. 딥러닝 기술과 빅 데이터의 결합

이 인식 기술을 실용 가능한 영역까지 끌어올렸다.

하지만 모든 영역에 대해 빅 데이터가 존재하는 것은 아니다. 데이터의 부족으로 정확한 인식이 불가능한 경우는 많이 있다. 이를 해결하기 위해 준지도 학습(Semi-Supervised Learning)이나 능동 학습(Active Learning), 능동준지도 학습(Active Semi-Supervised Learning) 등 다양한 시도가 있어왔다^[16].

하지만 이러한 방법을 통해서도 쉽게 파악하기 어려운 영역이 있으니, 그건 바로 사람의 감성이다. 사람은 말, 행동, 옷차림, 표정 등 다양한 요소에 의해서 감정을

*정회원, 인하대학교 컴퓨터공학부

**정회원, 인하대학교 컴퓨터공학부(교신저자)

접수일자: 2017년 11월 28일, 수정완료: 2018년 1월 29일

게재확정일자: 2018년 2월 9일

Received: 28 November, 2018 / Revised: 29 February, 2018

Accepted: 9 February, 2018

**Corresponding Author: pkrhee@inha.ac.kr

Dept. of Computer Engineering, Inha University, Korea

드러낸다. 하지만 사람은 자신의 감정을 숨길 줄 안다. 따라서 어느 한 가지만으로 상대방의 감정을 짐작하기란 쉽지 않은 일이다.

우리는 이러한 문제를 해결하고 보다 진솔한 사람의 감성을 이해하고 파악하기 위해 얼굴 표정과 행동에 주목했다. 행동과 표정은 부단한 노력과 훈련이 없으면 쉽게 감추거나 바꿀 수 없기 때문이다. 이에 본 논문에서는 보다 확실하게 사람의 감정을 추측하기 위해 새로운 기법을 제안하였다.

본 논문의 구성은 다음과 같다. 2장에서는 얼굴 표정 인식과 액션 인식, 점진적 딥러닝 학습 기법에 대한 선행 연구에 대해 알아볼 것이고, 3장에서는 우리가 제안한 시스템에 대한 개념을 설명하고, 4장에서는 점진적 딥러닝 방법을 통한 종합 감성 인식 방법에 대한 실험 환경과 그 결과를, 5장에서는 이 실험을 통해 알게 된 점에 대해서 논의하겠다.

II. 관련연구

1. 얼굴 표정 인식



그림 1. 얼굴 표정 인식 시스템
Fig. 1. Facial expression recognition system

얼굴 표정 인식 분야는 컴퓨터 비전 쪽에서 오랫동안 연구되어 온 분야이다^[1]. 보통 인식 과정은 세 가지 단계로 나뉘는데, 전처리, 특징 추출, 결과 분류가 그것이다. 먼저 전처리 과정에서 조명이나 배경 등의 노이즈를 제거하고, 특징 추출에서 얼굴 표정을 대표할 수 있는 특징 벡터를 뽑아내며, 결과 분류에서 추출된 특징들을 비교하여 가장 비슷한 결과를 찾아낸다^{[3][17]}.

얼굴 표정 인식은 보통 미국의 심리학자 폴 에크만이 주장한 기본6정서에 무표정을 더하여 7가지 라벨(분노, 역겨움, 공포, 행복, 슬픔, 놀람, 무표정)로 나타낸다^[2].

2. 행동 인식



그림 2. 행동 인식 시스템
Fig. 2. Action recognition system

행동 인식은 비디오 감시 시스템, 인간 행동 분석, 인간과 기계 간의 커뮤니케이션 등에 특히 중요한 분야로 지난 이십여 년 동안 꾸준히 연구되어 왔다. 주로 HOG (Histogram oriented gradient)^[10], STIP (Space time interest point)^[11], DPM (Deformable part model)^[9] 등의 기술을 사용해오다가 최근에는 딥러닝을 기반으로 하는 행동 인식 알고리즘이 개발되어 큰 성능 향상을 보였다.

3. 적응적 딥러닝 학습 기법

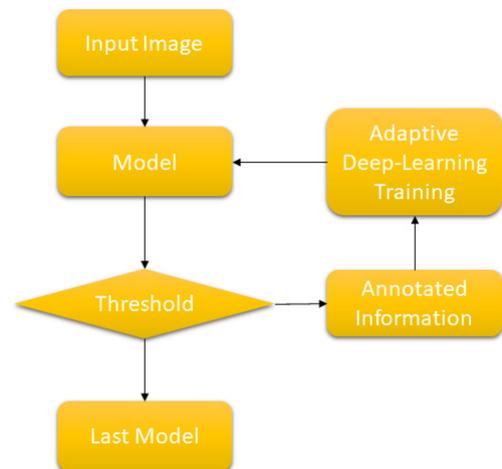


그림 3. 적응적 딥러닝 학습 기법
Fig. 3. Adaptive Deep-Learning Method

딥러닝 학습에 있어서 가장 문제가 되는 것이 바로 오버피팅 문제이다. 오버피팅은 보통 기존의 학습 데이터와 다른 환경에서 실험을 할 때 나타나는데, 이러한 환경 변화에 적응하지 못하고 인식률이 크게 감소하는 것을

뜻한다. 이를 해결하기 위한 가장 단순하고 확실한 방법은 방대한 양의 학습 데이터를 사용하는 것이지만, 현실적으로 그러기가 쉽지 않다. 적응적 딥러닝 학습 기법은 바로 이러한 문제점을 해결해주어 최소한의 노동력으로 최대한의 데이터를 확보하고 학습하도록 도와준다 [6][7][18][19].

III. 제안 시스템 개요

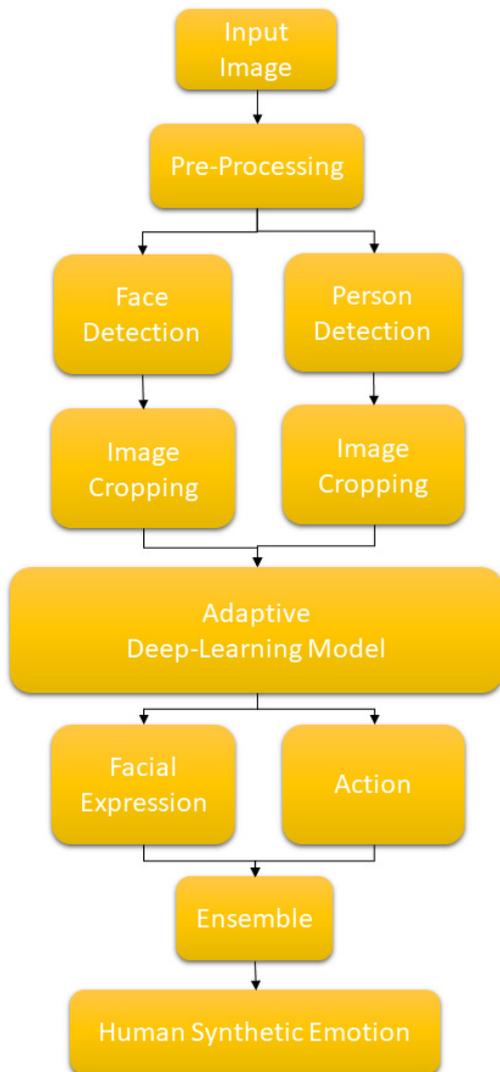


그림 4. 영상 인식 시스템 개요도
 Fig. 4. Image recognition system

그림 4는 본 논문에서 제안하는 시스템의 전체적인 개요를 나타낸다. 먼저 입력받은 이미지에서 사람과 얼굴을 잘 검출할 수 있도록 전처리 과정을 거친다. 그 후, 얼굴 검출기와 사람 검출기를 통해 각각 사각형 영역을 찾아낸다. 그렇게 찾아낸 영역을 잘라내어 적응적 딥러닝 학습 모델로 얼굴 표정과 행동을 예측한다. 예측이 끝나면 결과 값의 조합에 따라 종합적인 감성을 판단한다.

IV. 실험 및 결과



그림 5. 영상 인식 화면
 Fig. 5. Image recognition scene

본 논문에서 VGG-16 네트워크 [4][5][8]를 사용하여 각각 행동 인식과 얼굴 표정 인식 모델을 학습하였다. 그 후, 종합적인 감성을 파악하기 위해 분노, 역겨움, 공포, 행복, 슬픔, 놀람, 무표정의 일곱 가지 감정 라벨과 전화 중, 책 읽는 중, 컴퓨터 사용 중, 사진 찍는 중, 점프 중의 다섯 가지 행동 라벨을 갖는 실험 데이터를 준비하여, 적응적 딥러닝 학습 기법을 통해 학습하였다.

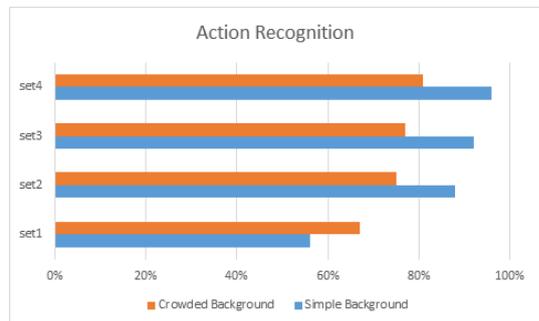


그림 6. 액션 인식 그래프
 Fig. 6. Action recognition graph

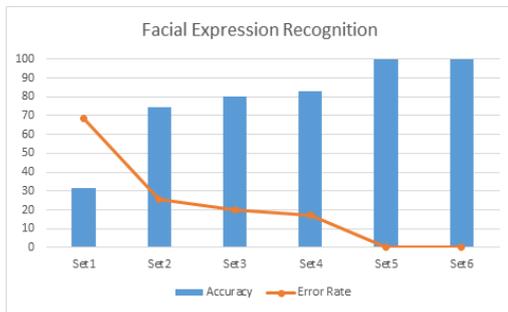


그림 7. 얼굴 표정 인식 그래프
Fig. 7. Facial expression recognition graph

그림 6과 같이 행동 인식의 경우, 뒤의 배경이 단순하느냐 복잡하느냐에 따라 그 인식률이 차이가 컸다. 영상에서 사람의 모습만 정확히 추출해낼 수 없었기 때문에 배경이 강력한 노이즈가 되어 인식을 방해했기 때문이다. 하지만 적응적 딥러닝 학습 기법을 통해 80% 근처까지 성능을 향상시킬 수 있었다.

그림 7은 얼굴 표정 인식에 대한 그래프인데, 이 경우 얼굴선 안에서 이미지가 추출되므로 배경이 전혀 들어가지 않는다. 이를 통해 우리는 얼굴 표정 인식률을 100%에 가깝게 끌어올릴 수 있었다.

V. 결론

이미 오랫동안 사람의 표정을 인식하는 알고리즘이나 액션을 인식하는 알고리즘은 연구되어 왔다. 본 논문에서는 인식된 얼굴 표정과 액션을 조합하여 사람의 감성을 추론하는 알고리즘을 제안한다. 또한 본 논문에서 제안한 알고리즘은 적응적 딥러닝 학습 기법을 활용하므로 보다 다양한 환경에서 적용이 가능할 것으로 기대된다.

References

- [1] Xiaoming Zhao, Shiqing Zhang, "A Review on Facial Expression Recognition - Feature Extraction and Classification", IETE Technical Review, 2016.
DOI : <http://dx.doi.org/10.1080/02564602.2015.1117403>
- [2] H. Rowley, S. Baluja, and T. Kanade, "Human face detection in visual scenes." Technical Report CMU-CS-95-158R, School of Computer Science, Carnegie Mellon University, Nov. 1995.
- [3] Andre Teixeira Lopes, Edilson de Aguiar, Thiago Oliveira-Santos, "A Facial Expression Recognition System Using Convolutional Networks", 28th SIBGRAPI Conference on Aug. 2015.
DOI : <http://dx.doi.org/10.1109/SIBGRAPI.2015.14>
- [4] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, Andrew Zisserman, "Return of the Devil in the Details: Delving Deep into Convolutional Nets", arXiv:1405.3531v4 [cs.CV] 5 Nov 2014.
- [5] Karen Simonyan, Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv:1409.1556v6 [cs.CV] 10 Apr 2015.
- [6] Cohn, D., Ghahramani, Z., Jordan, M., "Active learning with statistical models." Journal of Artificial Intelligence Research, Vol.4, pp. 129-145, 1999.
DOI : <https://doi.org/10.1613/jair.295>
- [7] Riccardi, G. and Hankkani-Tur, D., "Active learning: theory and applications to automatic speech recognition", IEEE Transactions on Speech and Audio Processing, Vol.13, No.4, pp. 504-511, 2005.
DOI : <https://doi.org/10.1109/TSA.2005.848882>
- [8] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps," Iclr, p. 1-, 2014.
- [9] P. F. Felzenszwalb, R. B. Girshick, D. Mcallester, and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models." Proceedings of the IEEE CVPR, 2008.
- [10] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection.", Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.2005.
DOI : <http://dx.doi.org/0.1109/CVPR.2005.177>
- [11] I. Laptev, "On space-time interest points," in

- International Journal of Computer Vision, 2005.
DOI : <http://dx.doi.org/10.1007/s11263-005-1838-7>
- [12] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional Two-Stream Network Fusion for Video Action Recognition," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit, pp. 1933 - 1941, 2016. arXiv:1604.06573 [cs.CV]
- [13] G. Gkioxari, B. Hariharan, R. Girshick, and J. Malik, "R-CNNs for Pose Estimation and Action Detection," arXiv Prepr. arXiv1406.5212, pp. 1 - 8, 2014.
DOI : [arXiv:1406.5212](https://arxiv.org/abs/1406.5212) [cs.CV]
- [14] A. Richard, "A BoW-equivalent Recurrent Neural Network for Action Recognition Bag-of-Words Model as Neural Network," British Machine Vision Conference, 2015.
DOI : <https://dx.doi.org/10.1016/j.cviu.2016.10.014>
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," IEEE Conf. Comput. Vis. Pattern Recognit., pp. 770 - 778, 2016. arXiv:1512.03385 [cs.CV]
- [16] B. Settles, "Active Learning," Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 6, no. 1, 2012.
DOI : <https://doi.org/10.2200/S00429ED1V01Y201207AIM018>
- [17] Alvin Poernomo, Dae-Ki Kang, "Content-Aware Convolutional Neural Network for Object Recognition Task," IJASC, Vol. 5, No. 3 pp. 1-7 May 2016.
DOI : <https://doi.org/10.7236/IJASC.2016.5.3.1>
- [18] J.W.Kim, P.K.Rhee, "Active Semi-Supervised Learning, Facial Expression, Recognition, Deep Learning", The Journal of The Institute of Internet, Broadcasting and Communication(JIIBC), VOL. 17 NO. 2, pp.165-171, 2017.
- [19] P.K.Rhee, Enkhbayar Erdenee, Shin Dong Kyun, Minhaz Uddin Ahmed, SongGuo Jin, "Active and semi-supervised learning for object detection with imperfect data" Cognitive Systems Research 45: 109-123 (2017)

저자 소개

김진우(정회원)



- 2013년 8월 : 인하대학교 컴퓨터공학부 학사
- 2017년 8월 : 인하대학교 컴퓨터공학부 석사

이필규(정회원)



- 서울대학교 전기공학부 학사
- ETSU(East Texas State University) 컴퓨터공학부 석사
- 루이지애나대학교 라피엣캠퍼스 컴퓨터공학부 박사
- 1992년 ~ 현재 : 인하대학교 교수

※ 이 논문은 과학기술정보통신부 및 정보통신기술진흥센터의 ICT융합산업원천기술개발의 일환으로 수행하였음(2017-0-00543, 보행자 위치공간 인지 증강 및 스포츠 경기력 분석을 위한 정밀 측위 원천기술개발)