

컨볼루션-순환 딥러닝을 이용한 컬러-깊이-열화상 얼굴 이미지 인식 방법

이범진^{1○}, 이충연¹, 장병탁¹¹서울대학교 컴퓨터공학부

{bjlee, cylee, btzhang}@bi.snu.ac.kr

RGB-D-T Face Recognition using Convolutional-Recursive Deep Learning

Beom-Jin Lee^{1○}, Chung-Yeon Lee¹, Byoung-Tak Zhang¹¹School of Computer Science and Engineering, Seoul National University

요 약

얼굴 인식 문제는 현시점에도 상당한 연구가 필요한 주제이다. 많은 해결 과제 중 가장 큰 문제점은 포즈, 조명에 의한 변화와 더불어 표정변화에 따른 인식을 저하를 들 수 있다. 그리고 기존 연구는 정보량이 적은 2차원 이미지를 사용하여 문제를 해결하려고 시도 하였으나 그 한계를 보이고 있다. 본 논문에서는 시사한 문제의 해결 방법으로 2차원 이미지보다 다양한 모달리티의 정보를 포함하고 있는 RGB-D, Thermal 카메라로부터 획득한 사람의 얼굴 데이터를 사용하여 각 모달리티 별 컨볼루션 신경망(Convolutional Neural Network)을 구축하고 딥 순환 신경망(Deep Recurrent Neural Network)과의 조합을 통해 포즈, 조명, 표정변화에 따른 얼굴 인식 문제의 해결 가능성을 확인한다. 해결하려는 문제를 포함한 26명의 얼굴인식 결과 약 99%의 성능을 보였으며, Deep Learning을 활용한 얼굴 인식 문제의 State of the Art를 뛰어넘을 수 있는 얼굴인식 방법론으로 사용 가능함을 확인하였다.

1. 서 론

신분위조 및 증가하는 범죄의 해결방안으로 강인한(robust) 성능의 얼굴 인식시스템이 각광 받고 있다. 기존 토큰방식, 암호방식의 보안 방법은 도난이나 코드를 잊어버릴 수 있는 잠재적 위험성을 지니고 있기 때문에 생체 데이터를 이용한 접근 제어 방법이나 범인 인식, 보안 시스템들이 선호된다. 이런 생체 데이터를 이용한 시스템은 크게 두 가지 카테고리로 분류 가능하다. 첫 번째는 행동 패턴을 분석하는 시스템과 두 번째는 지문이나, 얼굴 등의 생체학 특징을 분석하는 시스템이다. 이러한 카테고리 가운데 본 논문에서는 생체학 카테고리에 속하고 1960년대부터 활발히 연구가 진행되어온 얼굴인식기술에 대해 새로운 방법을 제시하고 보고한다.

기존 얼굴인식에는 이미지의 값을 사용하는 외형기반 방법론, 특징점 기반 방법론 그리고 하이브리드 방법론, 3가지 접근 방법론이 존재한다. 기존 외형기반 방법은 피셔페이스(Fisherface)[1], 아이겐페이스(Eigenface)[2], 독립성분분석(Independent Component Analysis)[3],

지지벡터머신[4]들이 존재한다. 외형 기반의 접근 방법은 선형의 차원축소 기술들이 적용되어 얼굴 인식을 수행한다. 하지만 해당 방법은 밝기 변화에 따라 성능이 크게 변하여 강인한 성능의 시스템 보장이 어렵다. 두 번째 특징점 기반 방법으로 기존의 기하학 특징점[5]을 이용한 방법, Hidden Markov Model(HMM)[6], Elastic Bunch Graph Matching(EBGM)[7]을 사용하는 방법들이 존재한다. 본 방법들은 입, 눈, 코와 같은 지역적 특징점을 추출하고 특징점들의 위치와의 관계 값 및 확률 값을 구조화된 분류기에 적용하여 얼굴 인식을 수행한다. 이러한 방법의 문제점은 최적 파라미터 탐색, 내재된 필터 조정과 같은 설계인자가 설계자에 의해 구체적으로 최적화된 조정을 필요로 한다. 마지막 하이브리드 방법론은 외형기반 방법론과 특징점 기반 방법론을 결합한 방법으로 하이브리드 지역적 특징 분석[8], 아이겐 페이스와 아이겐 모듈[9]이 존재한다. 해당 방법은 두 가지 방법을 결합하기 위해 설계 복잡도가 증가되었고, 기존 방법과 비교 시 성능 저하가 보고 되어 많은 연구가 이루어지지 않고 있다. 이러한 문제점들을 개선하기 위해 본 논문에서 사용한 방법은

특징점 기반 방법으로 높은 성능을 확보하고 특징점 방법론이 지니고 있는 세부 조정의 문제점을 자동화 함으로써 얼굴인식의 모든 부분에서 성능이 향상된 방법을 제안한다.

본 논문은 아래와 같이 구성되었다. 섹션 2에서는 각 모듈별 Convolutional Neural Network 구성 방법 및 각 모듈별 CNN을 DRNN과 결합하는 방법에 대하여 서술하고 섹션 3에서는 RGB-D-T 데이터를 사용한 실험 결과에 대해 서술하고 토의한다.

2. 연구 내용 및 방법

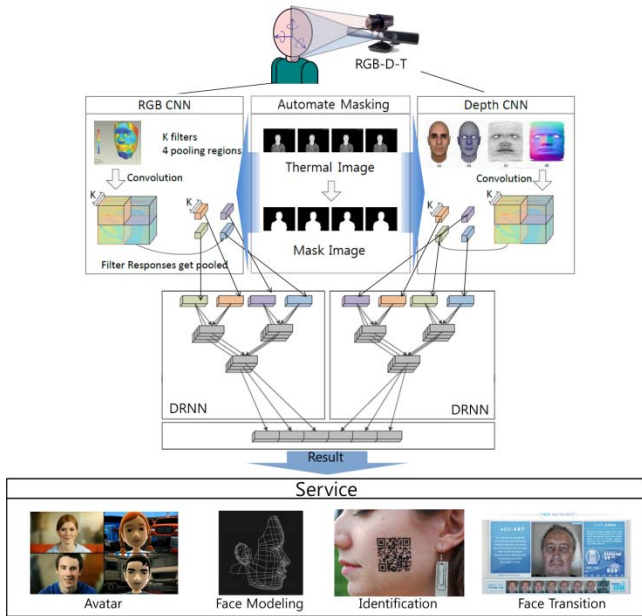


그림 1. 전체 시스템 구조도

그림1은 RGB-D-T 카메라를 이용하여 취득한 얼굴 이미지 데이터로부터 특징점을 자동적으로 추출 가능한 Convolutional Neural Network(CNN)과 얼굴인식에서 얼굴을 추론하는 모델인 Deep Recurrent Neural Network의 구조[10]를 설명하고 있다. 또한 제안하는 기술의 결과를 다양한 플랫폼에 탑재하여 사용자의 얼굴을 인식을 기반으로 하는 서비스를 제공가능 함을 확인할 수 있다.

2.1 Automatic Masking

CNN을 통한 컨볼루션 벡터를 추출하기 위해서는 각 모듈 별 이미지와 그에 상응 하는 마스크 이미지가 필요하다. 기존 방법들은 모두 수작업을 통하여 이미지의 마스크를 만들었지만 본 논문에서는 열화상 이미지를 통해 자동적인 마스크 이미지를 생성한다. 마스크 이미지는 열화상 이미지에서 한계점 이상의 값을 갖을 때 마스크로 변환하는 방법을 사용한다.

2.2 Convolutional Neural Network

컨볼루션의 필터들을 학습시키기 위해 먼저 RGB

이미지와 깊이 이미지에서 임의의 이미지 패치(patch)를 추출한다. 이런 패치는 정규화 및 화이트닝을 거친 후 k-means 알고리즘을 적용하여 클러스터링 한다. 이를 통해 패치는 엣지정보와 색상, 깊이 특징을 추출한다. 이러한 전처리를 마친 후 단일 계층의 CNN을 통해 다음 Recurrent Neural Network(RNN)의 입력으로 사용될 특징점 벡터들을 추출한다. 본 논문에 사용된 CNN은 [11]에 사용된 방법과 동일한 방법을 취하고 있으며, 이는 이미지 정규화 과정을 거친 후 컨볼루션을 수행한다.

2.3 Deep Recurrent Neural Network

CNN으로부터 추출된 컨볼루션 정보는 RNN 구성의 가장 기본인 3차원 블록 단위의 Tree로 구성된다. RNN은 같은 Neural Network(NN)을 재귀적으로 적용함으로써 계층적 특징을 표현 가능하게 하기 위함으로써, 구성된 블록은 수식 (1)을 통해 부모 벡터로 묶이게 되고, 이는 모든 블록에 대해 딥 방식의 재귀적 호출을 반복하여 하나의 부모만 남게 될 때까지 반복된다. 해당 학습 방법은 전향 전파방식의 무감독 학습을 통해 이루어 지고, 가장 상위 부모를 Softmax 분류기의 특징점 벡터로 사용되게 된다.

$$p = f\left(W \begin{bmatrix} x_1 \\ \vdots \\ x_{b,2} \end{bmatrix}\right) \quad (1)$$

f(): 비선형 tanh 함수, W: 랜덤 가중치, b: 블록 사이즈

3. 실험 및 결과 토의

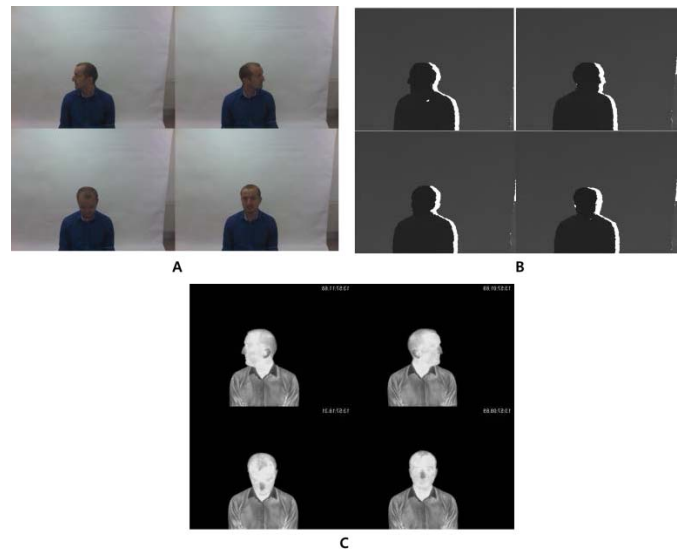


그림 2. RGB-D-T 데이터 셋 예. 정면, 오른쪽, 왼쪽, 아래, 위, 밝기 변화, 표정변화 이미지 취득 A) RGB, B) 깊이, C) 열화상 이미지

3.1 실험 구성

제안하는 방법의 성능을 확인하기 위해 RGB-D-T based Face Recognition 데이터 셋[12]으로 실험을 구성하였다. 얼굴 인식을 수행할 실험자는 총 26명이며

각 사람은 300개의 RGB, 깊이, 열화상 이미지로 구성되어 있다. 총 데이터 개수는 7800개로 이루어져 있고 각 인물 데이터의 2:1비율로 학습과 얼굴인식을 수행하였다.

그림 2에서 볼 수 있듯이 이미지는 본 논문에서 시사한대로 표정변화, 얼굴 각도 변화, 이미지 밝기 변화를 모두 고려한 이미지 데이터 셋이다.

3.2 얼굴인식 비교실험 및 결과

방법	RGB 성능	깊이 성능	RGB+깊이 성능
제안방법	97%	99%	99.5%
SVM	85%	80%	불가

표 1. 비교 실험 결과

비교실험은 이미지를 Histogram of Oriented Gradients (HOG)로 특징점을 추출하고 이를 1대 다(One-vs-All) 선형 분류 기술을 적용한 SVM으로 얼굴 인식을 수행하였다. 기존 [12]에서 모든 변화를 구분하여 얼굴 인식을 수행한 것과 달리 본 실험에서는 26명에 대해서 모두 시사한 문제점을 모두 포함시켜 학습, 인식을 수행하였다. 그 결과 본 논문에서 사용한 방법이 기존 얼굴 인식을 수행한 결과보다 성능을 월등히 우수한 것을 확인할 수 있다.

3.3 토의

본 논문에서는 기존 얼굴 인식 방법론 중 특징점 기반 방법론에서 CNN과 DRNN을 결합한 모델을 사용하여 얼굴인식을 수행한 결과에 대해 보고하였다. 기존 얼굴 인식 분야에서 가장 큰 문제로 삼고 있던 얼굴 각도 변화, 표정변화, 밝기변화 문제에 대해서 본 논문이 사용한 방법을 사용한다면 해결 가능 할 수 있음을 확인할 수 있었다. 하지만 본 방법론의 문제는 새로운 인물이 추가 될 시 비싼 계산력과 높은 계산 시간을 다시 수행해야 한다는 문제점을 지니고 있다. 특히 CNN에서 컨볼루션 벡터를 추출할 때 계산 시간이 굉장히 높다. 이를 해결하기 위해서 GPU 프로그래밍과 같은 분산 컴퓨팅 방법을 적용한다면 시간과 성능 모두 만족 시키는 방법론이 될 수 있을 것이다.

향후 연구로써 재학습이 필요하지 않는 점진학습 적용 방법, 다양성을 추가한 데이터의 성능 확인 및 실생활에서 얼굴인식을 수행하여 그 결과에 맞춰 사용자에게 적절한 서비스를 제공하는 실생활 어플리케이션 [13] 개발을 목표로 하고 있다.

Acknowledgement

본 논문은 삼성전자의 지원을 받아 수행된 연구이며 미래창조과학부(NRF-2010-0017734-Videome) 및 정보통신기술진흥센터의 정보통신·방송 연구개발사업(10035348-mLife, 14-824-09-014-Machine Learning Center, 10044009-HRI.MESSI)의 지원을 일부 받았음.

참고문헌

- [1] X. Wang and X. Tang, "Dual-Space Linear Discriminant Analysis for Face Recognition," Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 564-569, 2004.
- [2] N. Sun, H.-x. Wang, Z.-h. Ji, C.-r. Zou, and L. Zhao, "An Efficient Algorithm for Kernel Two-Dimensional Principal Component Analysis," In Neural Computing and Applications, Vol. 17, pp. 59-64, 2008.
- [3] D. N. Chandrappa and M. Ravishankar, "Automatic face recognition in a crowded scene using multi layered clutter filtering and independent component analysis," In 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 552-556, 2012.
- [4] G. Guo, S. Z. Li, and K. Chan, "Face Recognition by Support Vector Machines," Proc. of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 196-201, 2000.
- [5] T. Moe Ma Ma and M. M. Sein, "Multi Triangle based Automatic Face Recognition System by using 3D Geometric Face Feature," In Instrumentation and Measurement Technology, pp. 895-899, 2009.
- [6] H. Othman and T. Aboulnasr, "Hybrid Hidden Markov Model for Face Recognition," Proc. of Image Analysis and Interpretation, 4th IEEE Southwest Symposium, pp. 36-40, 2000.
- [7] A. Z. Pervaiz, "Real Time Face Recognition System Based on EBGMM Framework," In 12th International Conference Computer Modelling and Simulation (UKSim), pp. 262-266, 2010.
- [8] C. Cunjian, "Decision Level Fusion of Hybrid Local Features for Face Recognition," In 2008 International Conference on Neural Networks and Signal Processing, pp. 199-204, 2008.
- [9] J. F. Pereira, G. D. C. Cavalcanti, and R. Tsang Ing, "Modular Image Principal Component Analysis for face recognition," In International Joint Conference on Neural Networks(IJCNN), pp. 2481-2486, 2009.
- [10] R. Socher, B. Huval, B. Bath, C. D. Manning and A. Y. Ng, "Convolutional-recursive Deep Learning for 3D Object Classification," In Advances in Neural Information Processing Systems, pp. 665-673, 2012.
- [11] K. Jarrett, K. Kavukcuoglu, M. Ranzato and Y. LeCun, "What is the Best Multi-Stage Architecture for Object Recognition?," In IEEE 12th International Conference on Computer Vision, 2009, pp. 2146-2153, 2009.
- [12] O. Nikisins, K. Nasrollahi, M. Greitans, T. B. Moeslund, "RGB-D-T based Face Recognition," Proc. of International Conference on Pattern Recognition, 2014.
- [13] B.-T. Zhang, "Ontogenesis of agency in machines: A multidisciplinary review," In AAAI 2014 Fall Symposium on The Nature of Humans and Machines: A Multidisciplinary Discourse, Arlington, 2014.