Convolutional Neural Network와 Stereo Image를 이용한 얼굴 인식

기철민* · 조태훈*

*한국기술교육대학교

Face Recognition Using Convolutional Neural Network and Stereo Images

Cheol-min Ki* · Tai-Hoon Cho**

*Korea University of Technology and Education

E-mail : kigoon@koreatech.ac.kr

요 약

얼굴은 홍채, 지문 등과 같은 사람마다 가진 특수한 정보이다. 얼굴 인식에 대한 연구들은 과거부터 현재까지 지속적으로 진행되고 있으며, 이러한 연구들을 통해 여러 가지의 얼굴 인식 방법들이 나타났다. 이 중에는 스테레오로 구성된 얼굴 데이터를 이용하여 얼굴 인식을 진행하는 알고리즘들이 있다. 본 논문에서는 기계학습의 방법인 Convolutional Neural Network를 이용하여 스테레오로 구성된 얼굴 이미지를 하나의 신경망으로 학습을 진행하였다. 또한 스테레오로 구성된 얼굴 이미지는 카메라 2대를 이용하여 취득하였다. 이 방법은 얼굴 인식에서 보편적으로 많이 사용되는 알고리즘인 PCA를 이용한 스테레오 얼굴 인식의 결과보다 더욱 좋은 성능을 보였다.

ABSTRACT

Face is an information unique to each person such as Iris, fingerprints, etc,. Research on face recognition are in progress continuously from the past to the present. Through these research, various face recognition methods have appeared. Among these methods, there are face recognition algorithms using the face data composed in stereo. In this paper, Convolutional Neural Network with Stereo Images as input was used for face recognition. This method showed better performance than the result of stereo face recognition using PCA that is used frequently in face recognition.

키워드

Machine Learning, Convolutional Neural Network, Face Recognition, Stereo Image

Ⅰ. 서 론

사람이 누구인가를 인식하는 절차는 그 사람의 특징이 일치하는가를 판단하는 것이다. 사람의 고유한 정보는 얼굴, 홍채, 지문 등이 있으며, 얼굴은 사람이 상대방을 인식할 때 시각적인 정보로 누구인지 식별하는 가장 자연스러운 생체 인식 방법이다. 이러한 이유로 과거부터 지금까지 얼굴 인식에 대한 연구는 계속되고 있으며, 이로 인해 여러 가지의 얼굴 인식 방법들이 나타났는데, 여러 얼굴 인식의 방법들 중 2D 영상 1개만을 이용한 얼굴인식 방법들의 단점을 보완하기 위해 2D 영상 2개로 이루어진 Stereo 영상을 이용하는 얼굴 인식의 방법이 나타났다. 기존 Stereo 영상을 사용한 얼굴인식에는 두 이미지의 시차에서 나타나는 disparity 이미지를 이용해서 얼굴을 인식하는 방법이 있다. 이 방법은 좌우 카메라간의 시차를 이용하기 때문에 촬영거리에 민감하게 반응한다. 이를 해결하기 위해 본 논문에서는 Convolutional Neural Network[1,2] 와 Stereo 영상을 사용한 얼굴인식을 제안한다.

본 방법은 2개의 카메라로 촬영한 Stereo 영상을 Convolutional neural Network를 이용하여 학습을 시키고, 판별할 영상이 들어왔을 때 학습시킨 정보를 이용하여 판별할 영상이 어떤 사람인지에 대한 인식을 진행하는 방법이다. 본 방법에

대한 자세한 설명은 본문에서 설명하도록 하겠다.

Ⅱ. 선행 연구

본 논문에서 사용된 Convolutional Neural Network는 이미지에서 고수준의 추상화된 정보를 추출하는 등의 다양한 컴퓨터 비전 분야에서 많은 연구가 진행 중인 인공신경망의 한 종류이다. Convolutional Neural Network와 일반 Neural Network와의 차이는 Layer를 쌓는 과정에서 가중치의 개수를 확연히 줄일 수 있다는 점에서 차이를 보인다. Convolutional Neural Network는 일반적으로 몇 개의 층으로 이루어져 있으며 기본적으로 3가지의 다른 층을 가지고 있다. Convolution layer는 입력한 데이터에 convolution을 적용하여 Convolution feature를 추출하는 층이다. 이를 다시 말하면 유의미한 자질을 추출하는 층이다. 다음으로 Pooling layer가 있는데, Convolutional Neural Network는 보통 이미지에 적용되는데, 이미지의 특성상 픽셀의 개수가 너무 많고, 연산횟수가 많아지기 때문에 영상을 줄여 Sub-sampling하는 과정을 진행하는 층이다. 가장 마지막에 분류를 위해 사용하는 층으로 Feedforward layer가 있는데, 이는 일반적인 neural network와 같이 행동한다. 위에서 나타나는 Convolution layer와 Pooling layer를 번갈아가면서 Convolution feature를 추출하고, 마지막으로 feedforward layer를 이용하여 분류를 한다. 모든 layer에 대해서 추측을 할 때는 Forward propagation, 각 layer의 weight를 갱신할 때는 Backpropagation을 진행한다.[3] <그림 1>은 숫자 데이터의 손 글씨를 인식하기 위한 Convolutional Neural Network의 구조이다. 그림에서 나타나는 과정은 위에서 설명한 것처럼 Convolution layer와 Subsampling을 위한 Pooling layer의 반복으로 이루어져있으며, feedforward layer는 Full connection으로 이어지며 각 레이어의 feature는 120, 84, 10으로 구성된다. output의 개수가 10개인 이유는 0-9까지의 데이터이기 때문이다.

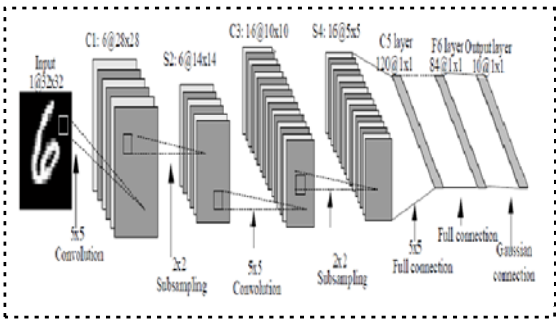

그림 1. MNIST 테스트를 위한 CNN의 구조[4]

비교에 사용한 PCA(Principal Component Analysis)와 Stereo image를 사용한 얼굴인식에는 얼굴을 판별하는데 PCA를 사용한다. PCA는 분포된 데이터들의 주성분(Principal component)을 찾아주는 방법이다. PCA는 n차원 좌표평면에 n개의 점 데이터가 존재할 때, 이 데이터들의 분포 특성을 표현해주는 방법이다. 2차원의 데이터 집합은 <그림 2>와 같이 2개의 주성분 벡터가 나타나며 이러한 특성을 이용해서 얼굴인식에 활용할 수 있는데, 이를 Eigenface라고 한다. Eigenface는 이미지의 픽셀 밝기 값을 일렬로 연결하여 벡터로 만들어 M×N 차원의 벡터가 나타나며, 이 데이터에 PCA를 수행하면 M×N개의 주성분 벡터가 나타난다. 이를 이용하여 새로운 이미지를 입력하여 PCA를 적용하면 M×N차원 내에서 나타난 주성분 벡터와 미리 비교할 대상의 벡터간의 거리를 비교하여 가장 유사한 벡터의 데이터를 같은 데이터라는 판별을 한다.

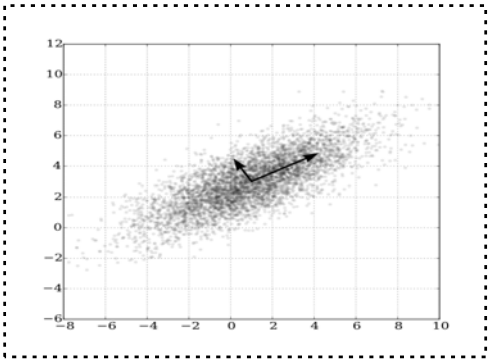

그림 2. 2차원 데이터에서의 주성분 벡터

본 논문에서는 좌, 우 카메라에서 촬영한 얼굴 영상과 두 영상 사이에서 나온 disparity 이미지를 이용한 Eigenface 방법을 비교하였다.[4]

Ⅲ. Convolutional Neural Network와 Stereo Image를 이용한 얼굴인식

본 논문에서는 위에서 설명한 Convolutional Neural Network와 두 장의 이미지로 구성된 Stereo Image를 사용하여 얼굴인식을 진행한다. 먼저 카메라 2대에서 나온 그레이 이미지 2장에서 얼굴 영역만 추출하기 위해 Haar-like Feature를 이용하여 각 카메라에서 촬영되는 얼굴 영역만을 추출한다. 이 후 추출한 얼굴 영역의 사이즈는 제각각이기 때문에 일정한 사이즈로 resize 시켜준다. 영상의 사이즈가 너무 크면 학습 시간이 오래 걸리기 때문에 영상의 사이즈는 29×29의 이미지로 resize를 진행한다. 이 후 1개의 Convolutional Neural Network에서 학습을 진행해야하기 때문에 하나의 영상으로 합쳐주는데, 1개의 3채널 이미지를 생성하고 이미지의 두 채널에

resize한 left, right 이미지를 할당하고 남은 한개의 채널을 비운다. <그림 3>은 resize한 이미지를 하나의 이미지로 할당한 영상들 중 일부를 보여준다.

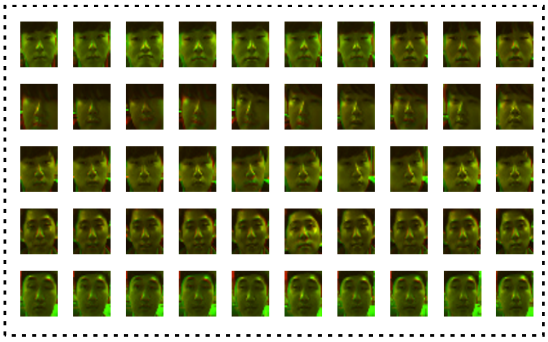

그림 3 . red, green 채널에 할당한 이미지

이 후 <그림 3>에서 나타난 이미지에서 red, green 채널의 데이터를 바이너리파일로 변환한 데이터를 입력 이미지로 Convolutional Neural Network를 이용하여 학습을 시킨다. 초기 입력 이미지는 29x29의 이미지이며 입력 이미지가 input layer로 입력이 되면 첫 번째 Convolution layer에서 6개의 Feature map을 만들고 13×13으로 Subsampling을 진행한다. 두 번째 Convolution layer에서 50개의 Feature map을 만들고 5×5로 Subsampling을 진행한다. 세 번째 layer에서 100개의 Neuron이 나타나며 마지막 output layer에서는 분류하고자 하는 얼굴의 라벨 개수만큼의 Neuron이 나타난다. 본 논문에서는 10명의 사람을 분류하기 위해 10개의 Neuron이 나타난다. 위에서 설명한 것처럼 2개의 카메라에서 촬영한 영상을 Haar-like Feature를 이용하여 얼굴을 찾고 얼굴의 관심영역을 29×29로 resize한다. 이후 영상을 데이터 포맷에 맞춰 변환하여 Convolutional Neural Network를 이용해 인식하는 과정을 시스템 흐름도로 표현하면 <그림 6>과 같다. 아래 그림에서 Image Convert 과정은 resize부터 데이터 변환까지를 묶어 표현했다.

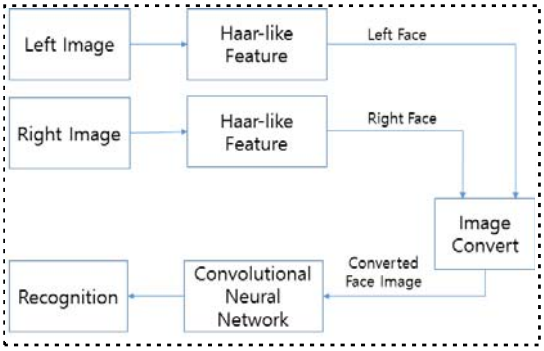

그림 4. Convolutional Neural Network와 Stereo Image를 이용한 얼굴인식 과정

Ⅳ. 실험 결과

본 논문의 실험 환경은 Microsoft Windows 7(64bit), IDE는 Visual Studio 2013을 사용하였고 openCV 2.4.9와 Firefly SDK를 사용하였다. 또한 카메라는 Pointgrey사의 Firefly MV FFMV-03M2M 카메라 2대를 이용하였고 렌즈는 초점거리가 16mm인 렌즈를 사용하여 얼굴 영상을 촬영했다.

Convolutional Neural Network와 Stereo Image를 이용한 얼굴인식의 결과는 다음에 나타난 <표 1>과 같다.

표 1. Convolutional Neural Network와 Stereo Image를 이용한 얼굴인식의 데이터 구성 및 정확도

	사람 수	총 영상 개수	학습 횟수	정확도
학습세트	10	200	89	100%
테스트세트		100		

Stereo Image는 학습세트에 10명의 얼굴을 각각 20장씩 총 200장의 이미지를 학습세트로 사용하였고, 테스트세트에는 10명의 얼굴을 각 10장씩 총 100장의 이미지를 테스트세트로 사용하였다. 학습 횟수는 MSE의 변화가 거의 없을 때까지 진행하여 89번의 학습을 진행하였고, 100%의 결과가 나타났다. 또한 PCA와 Stereo Image를 이용한 얼굴인식의 결과는 <표 2>와 같다.

표 2. PCA와 Stereo Image를 이용한 얼굴인식의 데이터 구성 및 정확도

	사람 수	총 영상 개수	정확도
학습세트	10	200	74%
테스트세트		100	

PCA에서는 Left Image와 Right Image, 둘 사이의 시차에서 나온 Disparity Image 각각에 PCA를 적용하여 나온 고유벡터들을 이용하며, 이 고유벡터와 인식할 테스트데이터와의 거리를 계산하여 나온 값에 가중치를 주고, 가장 작은 값이 나온 데이터를 찾는 방식을 사용하였다. 이 방식이 사용된 논문[5]에서는 Left Image와 Disparity Image의 거리를 계산해서 각각 가중치를 0.7, 0.3을 주어 얼굴인식을 진행했지만, 본 논문에서는 Left Image, Right Image, Disparity Image 3개의 이미지를 모두 사용하며, 각각의 가중치는 0.4, 0.4, 0.2의 가중치를 주었다. PCA에서의 입력이미지의 크기는 100x100의 얼굴 영상을 사용하는 점만 Convolutional Neural Network와 다르며, 이미지의 개수와 학습세트, 테스트세트의 구성은 동일하다.

Convolutional Neural Network와 Left Image를

이용한 얼굴인식의 결과는 <표 3>과 같다.

표 3. Convolutional Neural Network와 Left Image를 이용한 얼굴인식의 데이터 구성 및 정확도

	사람 수	총 영상 개수	학습 횟수	정확도
학습세트	10	200	125	94%
테스트세트		100		

<표 3>은 Stereo Image를 이용한 결과와 Mono Image를 이용한 결과를 비교하기 위한 테스트 결과이다. Convolutional Neural Network의 layer 구성과 학습세트, 테스트세트 등은 모두 동일하며 Input Image는 Left Image만을 이용한다.

종합적으로 Convolutional Neural Network와 PCA에서의 얼굴인식 정확도는 <표 4>과 같다.

표 4. 테스트 결과의 비교

	CNN + Stereo	PCA + Stereo	CNN + Mono
정확도	100%	74%	94%

Convolutional Neural Network와 Stereo Image를 이용한 얼굴인식의 결과가 상당히 좋게 나온 것은 많은 수의 데이터를 사용하지 않았기 때문이지만, <표 4>의 결과로 보아 PCA를 사용한 얼굴인식의 결과보다 Convolutional Neural Network를 사용한 얼굴인식이 정확도가 높은 것을 확인할 수 있다. 또한 본 논문에서 사용한 방법은 거리에 민감한 disparity 이미지를 사용한 PCA보다 거리와 영상의 rotation 등에 강인함을 가졌다. 하지만 PCA와 Convolutional Neural Network에서 공통적으로 나타난 문제는 빛의 변화에 민감하여 학습세트와 테스트세트가 밝기 값이 큰 차이가 나면 정확도가 떨어지는 문제가 나타났다. Convolutional Neural Network에서는 밝기 값이 다른 여러 이미지들로 학습 세트를 구성하여 지도학습 하는 방법으로 진행하는 방법이 있다. 하지만 모든 밝기에 대해 학습을 진행하는 것은 제한이 있으므로 위에서 언급한 데이터의 추가와 밝기에 강인한 검출은 추후에 보완해야할 문제이다.

Ⅴ. 결론

보편적으로 얼굴인식에서 사용되는 PCA를 이용한 얼굴인식과 Convolutional Neural Network와 Mono Image를 사용한 얼굴인식보다 본 논문에서 제시한 Convolutional Neural Network와 Stereo Image를 사용한 얼굴인식의 결과가 높은 정확도를 얻을 수 있었다. 하지만 조명변화가 심할 때 정확도가 떨어지는 모습을 보였다. 이 부분을 해결할 수 있다면 특별한 동작이나 행위 없이 자연스러운 생체인식이 가능할 것이며, 여러 응용분야에서 널리 사용될 수 있을 것으로 보인다.

참고문헌

[1] Y. LeCun, L. Bottou, Y. Bengio, and P.Haffner, "Gradient-Based Learning Applied to Document Recognition", Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[2] Patrice Y. Simard, Dave Steinkraus, John Platt, "Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis" , External Link International Conference on Document Analysis and Recognition (ICDAR), IEEE Computer Society, Los Alamitos, pp. 958-962, 2003.

[3] http://www.deepcumen.com/2015/04/Convolutional-neural-network

[4] http://www.ommegaonline.org/article-details/A-MATLAB-based-Convolutional-Neural-Network-Approach-for-Face-Recognition-System/464

[5] Sergey Kosov, Thorsten Thormahlen, Hans-Peter Seidel, “Rapid Stereo Vision Enhanced Face Recognition”, ICIP, 2010.