



로봇환경에서 3차원 CNN을 이용한 비디오 기반 얼굴 인식

Video-based Face Recognition Using an 3D Convolutional Neural Network in Robot Environment

저자 (Authors)	변영현, 곽근창 Yeong-Hyeon Byeon, Keun-Chang Kwak
출처 (Source)	Proceedings of KIIT Summer Conference , 2014.5, 26-29 (4 pages)
발행처 (Publisher)	한국정보기술학회 Korean Institute of Information Technology
URL	http://www.dbpia.co.kr/Article/NODE06086084
APA Style	변영현, 곽근창 (2014). 로봇환경에서 3차원 CNN을 이용한 비디오 기반 얼굴 인식. Proceedings of KIIT Summer Conference, 26-29.
이용정보 (Accessed)	한성과학고등학교 1.234.117.*** 2018/08/26 00:52 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

로봇환경에서 3차원 CNN을 이용한 비디오 기반 얼굴 인식

변영현*, 객근창**

Video-based Face Recognition Using an 3D Convolutional Neural Network in Robot Environment

Yeong-Hyeon Byeon*, Keun-Chang Kwak**

본 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임.
(NRF-2013R1A1A2012127)

요 약

최근 빅데이터와 오버피팅 방지, 하드웨어 성능향상으로 딥러닝(deep learning)이 떠오르고 있다. 그 중에 CNN(Convolutional Neural Networks)은 이동, 크기, 회전에 내성을 갖기 때문에 로봇환경에서의 얼굴인식에 적합하다. 본 논문은 3차원 CNN을 이용하여 비디오 기반 얼굴인식을 수행하였으며, ETRI Face Video 데이터 베이스를 가지고 실험을 하였다. 1m 데이터 50장을 학습하여, 1m 50장, 2m 200장, 3m 100장을 검증한 결과로 특징 맵 수에 따라 최고 100%, 88%, 73% 인식률을 얻었다.

Abstract

Recently deep learning outstands because of big data, prevention of over-fitting and improvement of hardware. One of them an CNN(Convolutional Neural Networks) is available to robot environment because it achieves some degree of shift, deformation and scale invariance. In this paper, we experiment video-based face recognition using 3D CNN on the ETRI face video. Training data is 50pictures captured from 1m away, and checking data are 50, 200, 100 pictures captured from 1m, 2m, 3m away and The rates of true recognition are 100%, 88% and 73% as maximum by varying the number of feature maps.

Key words

3D convolutional neural network, video-based face recognition, deep learning, robot environment

* 조선대학교 제어계측공학과 석사과정

** 조선대학교 제어계측공학과 교수 (교신저자)

I. 서 론

카메라로 연속된 영상을 획득하는 로봇환경에서 얼굴을 인식하는 기술로 주로 눈과 눈, 눈과 코 사이의 거리를 파악하는 방법으로 연구되어 왔다. 하지만 시간의 차를 갖고 연속으로 획득된 영상들은 개별마다 얼굴의 위치와 크기, 방향에 차이가 있어서, 앞서의 방법으로 인식하는데 어려움이 있다. 연구자들은 동물의 중추신경계를 흉내 낸 신경망 분석을 통해 얼굴을 더 정확히 분석하는 딥러닝에 주목하고 있다. 최근에 페이스북은 딥러닝을 이용하여 얼굴인식을 하는 딥페이스를 개발중이라고 하였으며, 그 성능은 인간보다 0.25% 정확도가 낮은 97.25%로 인식한다[1].

딥러닝의 CNN(Convolutional Neural Network)은 컨벌루션 단계와 서브샘플링 단계를 통해 특징을 추출하고, 이동과 크기, 방향에 내성을 갖도록 하기 때문에 필기체 숫자인식[2], 행동인식[3], 문자인식[4], 문서인식[5] 등에 성공적으로 응용되었다. 또한, 2차원 컨벌루션에 국한되지 않고 3차원을 이용한 연구도 발표되어 그 응용이 다양해지고 있다.

본 논문에서는 3차원 CNN을 이용하여 일련의 영상들에 대한 특징을 추출하고, 인식 대상에 대한 특징벡터와의 거리를 계산하여 가까운 범주로 분류한다. 실험에 사용된 데이터베이스는 ETRI Face Video이며, CNN을 이용한 얼굴비디오 인식에 적용된 사례가 없어 향후 로봇환경에서 CNN을 이용한 얼굴 비디오 인식에 도움이 될 것이다.

2장에서는 2차원 CNN을 설명하고 3차원에 대해 간단히 언급하며, 3장에서는 실험방법과 실험결과를 보고하며, 끝으로 4장에서 결론을 짓는다.

II. Convolutional Neural Network

GitHub에서 다운로드가 가능한 Matlab으로 작성된 딥러닝 라이브러리 내의 CNN은 필기체 숫자인식에 관한 것이다. 그곳에서 CNN은 5층으로 구성되어 있다. 첫 번째 층은 입력층이고, 2층은 컨벌루션 층이고, 3층은 서브샘플링 층이고, 4층은 컨벌루션 층이고, 5층은 서브샘플링 층이다. 입력으로 들

어온 영상을 특정 범위로 임의 생성된 커널과 컨벌루션을 하여 특징을 추출한다. 그 다음 서브샘플링을 거쳐 회전, 크기, 이동에 내성을 갖도록 한다. 이러한 과정을 거치게 되면 처음 입력으로 들어온 영상의 크기는 상당히 줄어들게 되며, 그 영상을 1차원 벡터로 구성하여 특징벡터가 된다. 여기에서 사용된 컨벌루션은 2차원 컨벌루션으로 영상과 커널이 2차원으로 구성되어 있다. 3차원 CNN은 영상과 커널이 3차원으로 구성되어 있으며, 첫 번째와 두 번째 차원은 영상을 나타내고, 세 번째 차원은 시간을 나타낸다. 이렇게 하면 시간에 따른 변화의 특징도 추출이 가능하다. 이렇게 추출된 특징은 퍼셉트론으로 커널과 바이어스, 가중치들을 학습시켜 인식한다[6]-[8].

III. 실험방법 및 결과

본 논문에서 실험에 사용된 CNN은 5층으로 구성되어 있다. 1층은 입력층이고, 2층은 컨벌루션 층이고, 3층은 서브샘플링 층이고, 4층은 컨벌루션 층이고, 5층은 서브샘플링 층이다. 2층과 4층의 컨벌루션은 3차원 컨벌루션을 수행하며, 커널은 특정 영역의 임의 값으로 초기화 된다. 5층을 거치게 되면 특징벡터를 얻게 되고, 분류는 특징벡터간의 거리가 가까운 범주로 하였다.

그림 1은 실험에 사용된 CNN의 구조를 보여주고 있다. 실험에 사용된 데이터는 ETRI Face Video이고 거리에 따른 10명의 얼굴데이터를 가지고 있다. 각 영상은 웨버라는 로봇에 장착된 카메라로부터 촬영되었고 얼굴만 추출되어 그레이스케일 및 히스토그램 평활화를 거친 것이다.

그림 2는 ETRI Face Video의 예시를 보여준다[9]. 그리고 CNN에 맞추기 위해 원래 45×40인 영상을 44×40 크기로 변경했으며, 이러한 영상을 연속된 10 프레임씩 묶으면 입력 층에 44×40×10의 3차원 영상이 들어간다. 2층에서 5×5 커널과 컨벌루션 되어 40×36×6이 되고, 2×2 서브샘플링을 거쳐 20×18×6이 되고, 5×5 커널과 컨벌루션 되어 16×14×1이 되고, 2×2 서브샘플링을 거쳐 8×7×1이 된다. 한 맵이 56차원의 특징벡터를 생성한다[3].

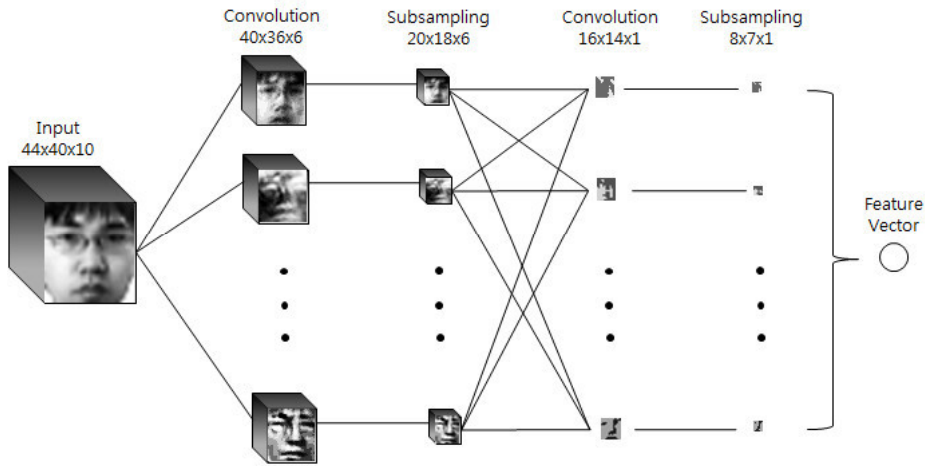


그림 1. CNN의 구조
Fig. 1. Structure of CNN

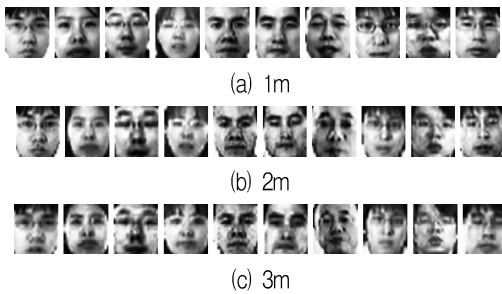


그림 2. 거리 변화에 따른 얼굴영상
Fig. 2. Face database according to distance variation

IV. 결 론

3차원 CNN을 이용하여 얼굴비디오 인식을 하였다. 맵 수에 따라 다양한 인식률을 얻었으며 최고 1m에서 100%, 2m에서 88%, 3m에서 73%를 얻었다. 분류 알고리즘을 통상 퍼셉트론을 사용하는데, GPU를 사용하지 못해 실행시간이 오래 걸려 충분한 실험을 하지 못하였다. 향후, GPU프로그래밍을 통해 여러 가지 분류 알고리즘을 비롯한 다양한 실험을 통해 인식률을 높일 것이다.

참 고 문 헌

- [1] HuffingtonPostKorea, http://www.huffingtonpost.kr/2014/03/21/story_n_4996658.html?ir=Korea&utm_hp_ref=korea
- [2] X. X. Niu and C. Y. Suen, "A novel hybrid CNN-SVM classifier for recognizing handwritten digits", Pattern Recognition, Vol. 45, Issue 4, pp. 1318-1325, 2012.
- [3] Shuiwang Ji, Wei Xu, and Ming Yang, "3D Convolutional Neural Networks for Human Action Recognition", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 35, Issue 1, pp. 221-231, 2013.

표 1. CNN을 이용한 얼굴인식 결과

Table 1. Results of face recognition using CNN

	맵의 수		인식률
	2층	4층	
1m 검증	1	13	100%
2m 검증	28	3	88%
3m 검증	8	1	73%

1m에서 촬영된 10명의 얼굴 데이터를 5세트씩 총 50세트를 학습으로 정의하고, 검증은 1m-50세트, 2m-200세트, 3m-100세트로 각각 거리가 가까운 범주로 분류를 하였다. 실험에서 CNN의 맵 수는 2층에서 1부터 30까지, 4층에서 1부터 60까지 변화를 주어 실험하였다. 표 1은 CNN을 이용한 얼굴인식의 실험결과이며, 최고의 인식률과 그 때의 맵 수를 나타내었다.

- [4] G. Lv, "Recognition of multi-fontstyle characters based on Convolutional neural network", International Symposium on Computation Intelligence and Design, Vol. 2, pp. 223-225, 2011.
- [5] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition", Proceeding of IEEE, Vol. 86, Issue 11, pp. 2278-2324, 1998.
- [6] The MathWorks, Inc., <http://www.mathworks.com/matlabcentral/fileexchange/38310-deep-learning-toolbox>(Deep Learning Toolbox)
- [7] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face Recognition: A Convolutional Neural-Network Approach", IEEE Trans. on Neural Networks, Vol. 8, Issue 1, pp. 98-113. 1997.
- [8] Y. H. Han and K. C. Kwak, "Face representation and recognition using third-order Tensor-based MPCA method", Journal of Korean Institute of Information Technology, Vol. 9, Issue 6, pp. 147-154, 2011.