Regular paper

Journal of the KIECS. pp. 1207-1214, vol. 14, no. 6, Dec. 31. 2019, t. 98, pISSN 1975-8170 | eISSN 2288-2189
http://dx.doi.org/10.13067/JKIECS.2019.14.6.1207

홈보안 시스템을 위한 CNN 기반 2D와 2.5D 얼굴 인식

MaYing* · 김강철**

CNN Based 2D and 2.5D Face Recognition For Home Security System

MaYing* · Kang-Chul Kim**

요 약

4차 산업혁명의 기술이 우리도 모르는 사이 우리의 삶 속으로 스며들고 있다. CNN이 이미지 인식 분야에서 탁월한 능력을 보여준 이후 많은 IoT 기반 홈보안 시스템은 침입자로부터 가족과 가정을 보호하며 얼굴을 인식하기 위한 좋은 생체인식 방법으로 CNN을 사용하고 있다. 본 논문에서는 2D와 2.5D 이미지에 대하여 여러 종류의 입력 이미지 크기와 필터를 가지고 있는 CNN의 구조를 연구한다. 실험 결과는 50*50 크기를 가진 2.5D 입력 이미지, 2 컨벌류션과 맥스풀링 레이어, 3*3 필터를 가진 CNN 구조가 0.966의 인식률을 보여 주었고, 1개의 입력 이미지에 대하여 가장 긴 CPU 소비시간은 0.057S로 나타났다. 홈보안 시스템은 좋은 얼굴 인식률과 짧은 연산 시간을 요구하므로 본 논문에서 제안한 구조의 CNN은 홈보안 시스템에서 얼굴인식을 기반으로 하는 액추에이터 제어 등에 적합한 방법이 될 것이다.

ABSTRACT

Technologies of the 4th industrial revolution have been unknowingly seeping into our lives. Many IoT based home security systems are using the convolutional neural network(CNN) as good biometrics to recognize a face and protect home and family from intruders since CNN has demonstrated its excellent ability in image recognition. In this paper, three layouts of CNN for 2D and 2.5D image of small dataset with various input image size and filter size are explored. The simulation results show that the layout of CNN with 50*50 input size of 2.5D image, 2 convolution and max pooling layer, and 3*3 filter size for small dataset of 2.5D image is optimal for a home security system with recognition accuracy of 0.966. In addition, the longest CPU time consumption for one input image is 0.057S. The proposed layout of CNN for a face recognition is suitable to control the actuators in the home security system because a home security system requires good face recognition and short recognition time.

Key words

Face Recognition, Convolutional Neural Networks, Smart Home Security System, Face Accuracy
얼굴 인식, 컨벌루션 신경망, 스마트 홈보안 시스템, 얼굴 정확도

* 전남대학교 대학원 컴퓨터공학과 석사과정 (619597401@qq.com)
** 교신저자 : 전남대학교 전기전자컴퓨터공학부
· 접 수 일 : 2019. 09. 24
· 수정완료일 : 2019. 11. 04
· 게재확정일 : 2019. 12. 15

· Received : Sep. 24, 2019, Revised : Nov. 04, 2019, Accepted : Dec. 15, 2019
· Corresponding Author : Kang-Chul Kim
Dept. Electronic Communication, Chonnam Nat'l University,
Email : kkc@chonnam.ac.k

Ⅰ. Introduction

Technologies of the 4th industrial revolution have been unknowingly seeping into our lives and suddenly changing human being's life quality in real life. IoT(: Internet of Things), which is the one of key technologies in the 4th industrial revolution, is the network of devices such as smart phone, vehicles and home electronic appliances. It involves extending internet connectivity beyond standard devices to everyday objects. Embedded in the systems, these devices can communicate and interact over the internet, and they can be remotely monitored and controlled[1-2].

People want to control the home situations at any time in any place. As the technology of communication, sensors and IoT has developed, home security has entered the era of intelligence. An intelligent home security system collects sensor's data, and send the information to the user via the internet, no matter where the user is. The user can monitor the situation of a home with smart phone.

With the rapid development of economy and civilization, most people work at the company and there is no one at home. When children come back home from school or packages are delivered, a security system is required to check them to enter a house. There are some biometrics such as iris, fingerprint, and face recognition to recognize people . At present, face recognition is becoming a hot research field of pattern recognition and artificial intelligence as the recognition ratio has increased with CNN(convolution neural network)[3].

This paper aims to recognize a face by CNN with small dataset for home security that provides a method for users to control the home through a website or a mobile phone. In the following section Ⅱ, the related works of CNN and dataset are introduced. In section Ⅲ, the system configuration and proposed methodology are described. Simulation results and considerations are shown in section Ⅳ, and conclusions are described in section Ⅴ.

Ⅱ. Related works

Deep learning is not a new technology, but an evolutionary techniques that combines and extends existing machine learning techniques[1]. It is part of a broader family of machine learning methods based on artificial neural networks. It is a neural network that is multilayered structure and is being trained by hierarchical abstract learning that distinguish existing general machine learning and existing neural network[4]. It starts with simple character from each layer and the simple characters were combined and become more complicated, and the complicate characters are combined again, and ultimately it recognizes character of the face. The greatest benefit is that most of the part is automated rather than such expectation-driven learning technique which needed professional system and the tasks that require work from human being.

CNN is a kind of deep neural network[5], which is most often used to analyze visual images. CNN uses variants of multi-layer perceptrons, requiring minimal preprocessing. A major feature of CNN is the processing of large data sets of millions of pictures. And it automatically extracts the features of image without any help of an expert.

In recent, as 2D face recognition technigues have disadvantages such as illumination and expression, and 3D camera has been developed, 3D images have been popular. Two representations for modeling 3D faces are 3D and 2.5D images[6]. A 2.5D image is defined as a simplified three dimensional (x,y,z) surface representation that

1) https://en.wikipedia.org/wiki /Deep_learning

contains one depth value for every point in the plain[7].

Some researches have been done to extract the key point with 2.5D images. The 2.5D image is a perfect replacement of pure 3D images for 3D facial recognition to reduce CPU time consumption and increase recognition ratio. Only the face of the head is enough to recognize the face of database in an IoT based home security in this paper.

Recently, geometric shaped facial feature extraction for face recognition(GSF2EFR) is designed for identifying the exact person by finding the center and corners of the eye using eye detection and eye localization modules[8]. It only needs a small space to store geometric feature vectors and has a fast proessing time, but the recognition accuracy is not good enough. A statistical local feature based robust kernel representation(SLF-RKR) model for face recognition was proposed, but its security is low and it can not deal with similar features of face distribution[9]. An improved method for face recognition named statistical Local Line Binary Pattern(SLLBP) was proposed and a statistical analysis on the probability distribution of the gray-level difference values was taken to deal with the gray-level difference, and a mapping function was used to encode the range of these values. But it still can not avoid the shortcomings of face distribution statistical recognition[10]. Primary Component Analysis(PCA) and Support Vector Machine(SVM) makes the neural network get better fitting ability, but its calculation needs a lot of storage space[11-12]. Active Appearance Model(AAM) and Fuzzy Neural Networks(FNN) use a smaller learning rate, and the neural network can fit better, but the convergence rate is relatively slow, which results in a longer calculation time of the neural network[13]. The above face recognition methods have their own advantages and disadvantages, and they are not good enough for smart home security systems. Therefore, CNN is used in our face recognition method. In successful CNN face Recognized system, recognition rate is 98.76%, three convolution layers of CNN method is used[14]. The input is RBG face image, and the face is divided into four parts to calculate the convolution neural network. Although it has a high recognition rate, it is more difficult to calculate the convolution neural network by four times convolution calculation. In face recognition system using CNN method with recognition rate of 98.8%, five convolution layers are used which increases the pressure on the CPU to perform calculations and results in higher costs[15]. So a smart home security system running on a small system like Raspberry Pi needs less computational pressure model.

This paper designs the most optimal layout of CNN for a face recognition with small data sets of 2D and 2.5D image in a home security system. The optimal numbers of the convolution and max pooling, input image size, and filter size are considered which CPU time consumption of CNN is within an allowed time and recognition ratio is high.

Ⅲ. System configuration and the proposed methodology

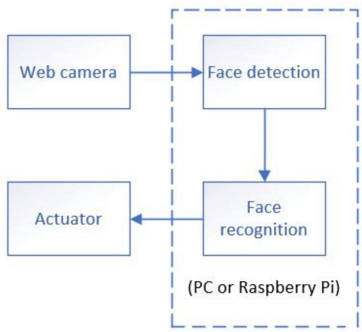


그림 1. 홈보안 시스템
Fig. 1 Home security system

The proposed CNN will be used for a face recognition in a home security system as shown in Fig. 1. The accuracy of facial recognition as well as CPU time consumption are mainly concerned in this paper and the face detection algorithm had been developed in other papers[16].

The result of face recognition is used to control the inputs of actuators such as motor and LED.

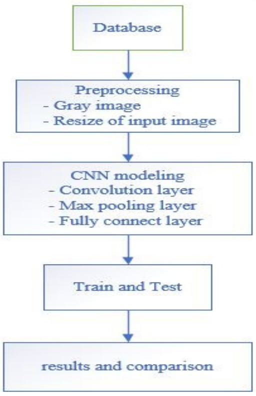


그림 2. 얼굴 인식 처리 과정
Fig. 2 The processing steps of a face recognition

Fig.2 shows the whole process of face recognition for 2D and 2.5D images taken from the dataset. To get the optimal layout of CNN for the face recognition ratio and CPU time consumption, CNN is simulated under various conditions of input image size, convolution and max pooling layer, and filter size. OpenCV is used in the preprocessing steps[17]. RBG color data are transformed into single layer gray data ranging from 0 to 255 to reduce the processing time and the gray images are compressed to the size of 50*50 face images for the input of CNN.

CNN modeling steps are simulated on Keras based on tensorflow[18]. CNN layputs are built in sequential() function in Keras[19]. Add() function in sequential() is used to add convolution layer, max pooling layer and fully connect layer. The CNN with the activation function of ReLu is simulated depending on input size, convolution layer, and filter size after the learning rate and the epoch number are selected. In fully connected layer, Flatten(), Dense() with the activation function of ReLu, Dropout(), and Dense with the activation function of softmax in Keras are executed.

CPU time consumption is one of very important factor for a real-time security system. CPU time highly depends on the input size of CNN, so 6 kinds of input size are considered to get the CNN that has the proper CPU time consumption and recognition ratio. Also the face recognition ratio is affected by the number of convolution layer and pooling layer and the size of filter, so the CNNs with 3 kinds of convolution layers and 3 kinds of filter size are simulated.

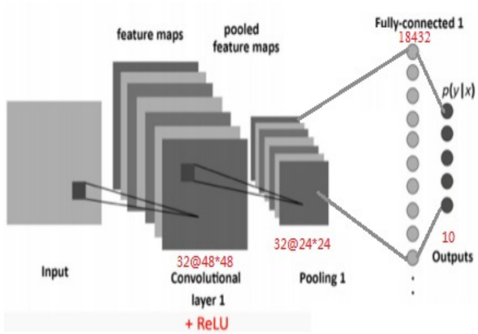


그림 3. 1 컨벌루션 계층과 3*3 필터 CNN
Fig. 3 CNN with one convolution layer and 3*3 filter

Fig. 3 shows an example of CNN with one convolution layer and 3*3 filter and the red letters mean the number of parameters in each layer.

MIT-CBCL-facerec-database which has 7280 images of ten people is used in this paper[20]. These images, which the first and third columns are 2D images and the second and forth are 2.5D images as shown in Fig. 4. 80% and 20% of images are used for training and testing respectively.

그림 4. MIT-CBCL-facerec-database
Fig. 4 MIT-CBCL-facerec-database

Ⅳ. Simulation results and considerations

Table 1 shows the recognition results of two convolution layers and maxpooling layer with 3*3 filter and the learning rate of 0.001. the CPU time decreases and the recognition ratio increase as input image size increases. It is considered that 50*50 input size is optimal for a real-time home security system.

표 1. 입력 이미지 크기에 대한 CPU 시간과 정확도
Table 1. CPU time and accuracy for input images sizes

image size	For test CPU time(s)		Accuracy	
	2D image	2.5D image	2D image	2.5D image
200*200	352.869	356.698	0.925458	0.976652
100*100	142.396	148.664	0.921958	0.969999
90*90	101.324	100.865	0.917526	0.967584
70*70	81.362	81.974	0.917111	0.966951
50*50	52.062	53.534	0.916235	0.966369
30*30	30.125	29.698	0.835621	0.875487

Table 2 shows the recognition ratio for the number of convolutionn layer with the input size of 50*50, 0.0010 learning rate, and 3*3 filter size. From the table 2, the CNN with 2 convolutional layer gets the best recognition accuracy.

표 2. 컨벌루션 계층수에 대한 정확도
Table 2. Accuracy for the number of convolution layer

number of convolution layer	2D recognition accuracy	2.5D recognition accuracy
1	0.912546	0.963325
2	0.916235	0.966369
3	0.831548	0.895482

Table 3 shows the simulation results for the parameters of filter size and learning rate. The recognition accuracy decrease as the filter size decrease, and the best result comes in the learning rate of 0.0010. The CNN with the combination of 3*3 filter and 0.0010 learning rate gets the highest recognition accuracy.

표 3. 파라메타에 대한 정확도
Table 3. Accuracy for parameters

image	learning rate	recognition accuracy		
		3*3 filter	5*5 filter	7*7 filter
2D image	0.0006	0.905481	0.896945	0.886577
	0.0008	0.910526	0.889631	0.895564
	0.0010	0.916235	0.895682	0.882285
	0.0012	0.908655	0.904852	0.879947
	0.0014	0.899999	0.894658	0.892373
2.5D image	0.0006	0.958742	0.956268	0.949658
	0.0008	0.960241	0.960254	0.946842
	0.0010	0.966369	0.956473	0.954644
	0.0012	0.960547	0.952453	0.948699
	0.0014	0.961541	0.958647	0.958467

A part of 2D image's data format is [......(220, 187, 152), (217, 184, 149), (217, 184, 149), (220, 185, 153), (220, 185, 153), (218, 182, 150)......], and a part of 2.5D image's data format is [......(0, 0, 0), (120, 77, 56), (125, 85, 61), (129, 86, 62), (133, 91, 67), (137, 94, 70), (0, 0, 0)......]. 2.5D image has better contrast than 2D image because 2.5D image's data format have a lot of "(0,0,0)", which means the dark area, 2.5D images makes the images more stereoscopic, and enhances the accuracy compared to 2D images. Table 4 shows performance such as f1-score, precisions and recall, and that 2.5D gets higher recognition ratio than 2D.

표 4. 성능 측정
Table 4. Performance metrics

	f1 score	precisions	recall
2D	0.915246	0.913524	0.91786
2.5D	0.962543	0.964172	0.963495

The CPU time consumption for 2 convolution layers, 3*3 filter, and 50*50 input size with Intel(R) Xeon(R) CPU E3-1230 v3@3.30GHz is shown in table 5. The longest CPU time for one image is less than 0.057 sec and considered to be with the allowed time for a real-time home security system.

표 5. 한 개의 이미지에 대한 CPU 시간 소비
Table 5. CPU time consumption for one image

CPU test time(s)	2D image	2.5D image
shortest	0.035	0.036
longest	0.057	0.056
average	0.046	0.047

A home security system in a house has to recognize a visitors' face within one or two seconds with proper recognition ratio, so the proposed face recognition system can be suitable to the home security running on Raspberry Pi.

Ⅴ. Conclusions

In this paper, the performance of CNN is explored depending on the number of a convolution and max pooling layer, input image size, and filter size to select an optimal layout for a home security system. One, two, three convolution layer and max pooling layer, 30*30, 50*50, 70*70, 90*90, 100*100, 200*200 input image size, and 3*3, 5*5, 7*7 filter size are explored to get an best recognition accuracy and CPU time consumption.

This paper proposes that the layout of CNN with 50*50 input image size, 2 convolution and max pooling layer, and 3*3 filter size for small dataset of 2.5D image is most optimal, which shows 0.966 accuracy and the longest CPU time consumption of 0.057S for one input image. The proposed layout of CNN for a face recognition is suitable to control the actuators in the home security system.

References

[1] A. Bahga and V. Madisetti, Internet of Things, *Arshdeep Bahga & Vijay Madsetti*, 2014.

[2] K. Kim, D. Wang, and S. Han, "Home Security System Based on IoT," *J. of the Korea Institute of Electronic Communication Sciences,* vol. 12, no. 1, Feb. 2017, pp. 147-154,

[3] H. Zhang, Z. Qu, L. Yuan, and G. Li, "A face recognition method based on LBP Feature for CNN," *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC),* Chongqing China, Feb. 2017, pp. 544-547.

[4] P. Li, J. Li, and G. Wang, "Application of convolutional neural network in natural language processing," *2018 15th Int. Computer Conference on Wavelet Active Media Technology and Information Processing,* Chengdu, China, Aug. 2018, pp. 120-122.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, no.1, 2012, pp. 1097-1105.

[6] S. Pui and J. Minoi, "Automatic Landmarking on 2.5D Face Range Images," *2014 Fifth Int. Con. on Intelligent System, Modelling and Simulation*, Langkawi, Malaysia, 2014, pp. 261-265.

[7] H. Hu, S. A. Shah, M. Bennamoun, and M. Molton, "2D and 3D Face recognition Using Convolutional Neural Network," *Proc. of the 2017 IEEE regional 10 Conference,* Penang, Malaysia, Nov. 2017. pp. 133-138.

[8] S. R. Benedict and J. S. Kumar, "Geometric Shaped Facial Feature Extraction for Face

Recognition," *2016 IEEE Int. Conf. on Advances in Computer Applications (ICACA)*, Coimbatore, India, Apr. 2016, pp. 275-278.

[9] M. Yang, L. Zhang, S. Shiu, and D. Zhang, "Robust Kernel Representation With Statistical Local Features for Face Recognition," *IEEE, transactions on neural networks and learning systems*, vol. 24, no. 6, June 2013, pp. 900-912.

[10] Z. Sun and J. Cheng, "Face Recognition with statistical Local Line Binary Pattern," *IEEE, Seventh Int. Conf. on Computational Intelligence and Security*, Hainan, China, Apr. 2011, pp. 1114-1117.

[11] L. X. Wei, Y. Sheng, W. Qi, and L. Ming, "Face Recognition Based on Wavelet Transform and PCA," *IEEE, Pacific-Asia Conf. on Knowledge Engineering and Software Engineering*, Shenzhen, China, Apr. 2009, pp. 136-138.

[12] Z. Lihong, S. Ying, Z. Yushi, Z. Cheng, and Z. Yi, "Face Recognition based on multi-class SVM," ***2009 Chinese Control and Decision Conference***, Guilin, China, Sept. 2009, pp. 5871-5873.

[13] Q. xiaoyan, "Face Recognition Algorithm Based on AAM and FNN," *2014 Fifth Int. Conf. on Intelligent Systems Design and Engineering Applications*, Hunan, China, Sept. 2014, pp. 464-468.

[14] M. Coskun, A. Ucar, O. Yildirim, and Y. Demir, "Face Recognition Based on Convolutional Neural Network," *2017 Int. Conf. on Modern Electrical and Energy Systems (MEES)*, Kremenchuk, Ukraine, Feb. 2017, pp. 376-379.

[15] A. Adil, J. Hind, and J. Cheng, " Localized Deep CNN Structure for Face Recognition," *2018 11th Int. Conf. on Developments in eSystems Engineering (DeSE)*, Cambridge, United Kingdom, May 2018, pp. 52-57.

[16] K. Kim and H. Wei, "Development of a Face Detection and Recognition System Using a RaspberryPi," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 12, no. 05, 2017, pp. 859-864.

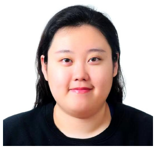
[17] D. Baggio, Daniel , D. Escriva, N. Mahmood, R. Shilkrot, S. Emami, K. Ievgen, and J. Saragih, Mastering OpenCV with practical computer vision projects. *Packt Publishing*, 2012.

[18] F. Ertam and G. Aydin, "Data Classification with Deep Learning using Tensorflow," *2017 Int. Conf. on Computer Science and Engineering (UBMK)*, Antalya, Turkey, Sept. 2017, pp. 755-758.

[19] A. Nagisetty and G. P. Gupta, "Framework for Detection of Malicious Activities in IoT Networks using Keras Deep Learning Library," 2019 3rd Int. Conf. on Computing Methodologies and Communication, Erode, India, Apr. 2019, pp. 633-637.

[20] B. Weyrauch, J. Huang, B. Heisele, and V. Blanz, "Component-based Face Recognition with 3D Morphable Models," *2004 Conf. on Computer Vision and Pattern Recognition Workshop*, Washington, DC, USA, 2004, pp. 27-34.

저자 소개



MaYing

2016년. Beijing Institute of Petrochemical Technology, Communication Engineering 졸업(공학사)
2019. 현재 전남대학교 대학원 컴퓨터공학과 재학
※ 관심분야 : Face Recognition, IoT

김강철(Kang-Chul Kim)

1981년 서강대학교 전자공학과 학사
1983년 서강대학교 전자공학과 석사
1996년 경상대학교 전자공학과 박사
현재 전남대학교 전기전자통신컴퓨터공학부 교수
※ 관심분야 : 임베디드시스템, NoC, IoT Pattern Recognition