Expert Systems With Applications 87 (2017) 304–315

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Sensitive deep convolutional neural network for face recognition at large standoffs with small dataset

Amin Jalali, Rammohan Mallipeddi, Minho Lee*

School of Electronics Engineering, Kyungpook National University, 1370 Sankyuk-Dong, Puk-Gu, Taegu 702-701, Korea

ARTICLE INFO

Article history:
Received 1 December 2016
Revised 14 June 2017
Accepted 15 June 2017
Available online 16 June 2017

Keywords:
Convolutional neural network
Gradient descent
Input-output mapping sensitivity error back propagation
Face recognition at long distances with small dataset
Sensitivity in cost function
Deep neural structures

ABSTRACT

In this paper, we propose a sensitive convolutional neural network which incorporates sensitivity term in the cost function of Convolutional Neural Network (CNN) to emphasize on the slight variations and high frequency components in highly blurred input image samples. The proposed cost function in CNN has a sensitivity part in which the conventional error is divided by the derivative of the activation function, and subsequently the total error is minimized by the gradient descent method during the learning process. Due to the proposed sensitivity term, the data samples at the decision boundaries appear more on the middle band or the high gradient part of the activation function. This highlights the slight changes in the highly blurred input images enabling better feature extraction resulting in better generalization and improved classification performance in the highly blurred images. To study the effect of the proposed sensitivity term, experiments were performed for the face recognition task on small dataset of facial images at different long standoffs in both night-time and day-time modalities.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Recently, deep neural networks such as Convolutional Neural Networks (CNN) have been successfully applied in many recognition problems (Ciresan, Meier, & Schmidhuber, 2012). The conceptual architecture of CNN is inspired by Hubel and Wiesel (1959)'s seminal work on the cat's striate cortex called receptive field. Later, Fukushima (1980) explained the Neocognitron, which defines the layer wise structure of neural networks and explains the spatial invariance characteristic of simple cells and complex cells of visual primary cortex. LeCun introduced the structure of CNN for face and digit recognition (LeCun & Bengio, 1995; LeCun, Bottou, Bengio, & Haffner, 1998; LeCun, Huang, & Bottou, 2004), which demonstrated better recognition results than probability density function methodologies (e.g., Gaussian Bayesian approaches and Gaussian Mixture models) or nonparametric clustering approaches (e.g., K-nearest neighbor classifiers). Rowley, Baluja, and Kanade (1998) utilized CNN for face recognition with three layers and three receptive fields in the first layer. Simard, Steinkraus, and Platt (2003) proposed the implementation of a more efficient subsampling approach in the operation of the convolutional layers instead of a separate subsampling layer leading to a faster algorithm in terms of training.

Simonyan and Zisserman (2015) proved that very deep convolutional networks are effective for large scale image classification. As the depth increases, the performance of the network improves on complex recognition tasks. Szegedy et al. (2015) presented a deep CNN structure called "Inception" which has salient features regarding computing resources of the network. It is accomplished by going deeper with convolutions and increasing the depth and width of the network as the computational cost is kept constant. In addition, it utilizes the Hebbian principle and multi-scale processing in its network architecture. He, Zhang, Ren, and Sun (2016) proposed a deep residual learning framework to simplify the training process of the networks that are deep. It is done by introducing the learning residual functions with reference to the input layers. The accuracy of residual network enhances as the depth increases while having lower complexity. Schroff, Kalenichenko, and Philbin (2015) presented FaceNet which learns a mapping from face samples to a Euclidean space. Those distances represent the similarity and form the feature space. They utilized deep convolutional network trained by triplets of roughly aligned face patches using online mining method. This method could achieve the state-of-the-art recognition performance on the Labeled Faces in the Wild (LFW) and YouTube faces datasets. The studies discussed above prove the effectiveness of deep networks in the recognition tasks.

* Corresponding author.
E-mail addresses: max.jalali@gmail.com (A. Jalali), mallipeddi.ram@gmail.com (R. Mallipeddi), mholee@gmail.com, mholee@knu.ac.kr (M. Lee).

http://dx.doi.org/10.1016/j.eswa.2017.06.025
0957-4174/© 2017 Elsevier Ltd. All rights reserved.

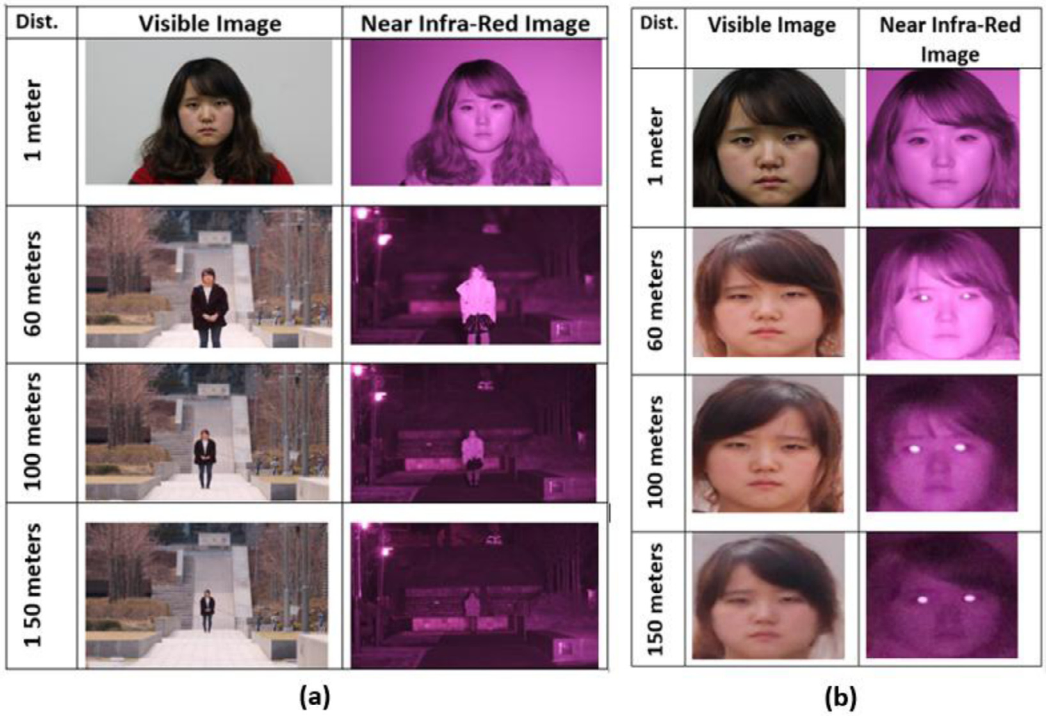

Fig. 1. Examples of long distance and near infrared facial images. (a) Eight images in LDHF database (four near infrared images and four visible images at 1 m, 60 m, 100 m, and 150 m, respectively) for each subject. (b) Cropped sample faces of a subject in LDHF dataset are shown.

In some practical applications such as satellite imaging, long distance object recognition, and public safety control, images are generally captured in unconstrained illumination and environments with complexities such as fog, cloudy, long distance and blurry conditions. Therefore, the quality of the images obtained in such systems is of degraded quality and hence the recognition becomes a challenging task. Generally, major manual processing combined with automatic image recognition methods are applied on the captured images to get acceptable accuracy and efficiency. However, the recognition task becomes further complicated when the number of samples are small. Therefore, for a given dataset with small number of samples and problem complexity, the design of the learning algorithm to have appropriate generalization is the issue of this study.

The objective of this paper is to propose an improved feature extraction and generalization for highly blurred images by considering the sensitive cost function for training process of deep-CNN. The sensitivity regularization term in training algorithm considers the small changes of input images to make the CNN more sensitive to images with high intra-class variance and low inter-class variance. To evaluate the efficiency of the proposed approach the Long Distance Heterogeneous Face dataset (LDHF) (Kang, Han, Jain, & Lee, 2014) is used. This dataset contain few sample images with different face sizes, different modalities (night-time and day-time), and different quality degradations (e.g., blur and noise) at different distances. A sample subject of the LDHF dataset with its corresponding cropped faces is shown in Fig. 1. The incorporated sensitivity term is expected to highlight the small variations in the pixels of the blurred low illumination images during the learning process by better feature extraction. In other words, by incorporating the sensitivity term in the CNN cost function, the neural activations of the hidden layers are located at high gradient region of the activation function for areas with small variations. This results in the modifications of the corresponding weights of the structure during training. To implement the sensitivity in CNN structure, a high pass filter is inherently incorporated in cost function of the error back propagation learning algorithm to highlight the edges and high frequency variations in the blurred images. Throughout the training process, as the error flows back to the first layer, the weights updating takes into consideration the small changes in the input images due to the presence of high pass filter resulting in better classification.

To demonstrate the effectiveness of the proposed sensitivity term, experiments were performed using a face dataset consisting different scenarios. First, the recognition performance in images that are degraded, blurred and with low illumination obtained from a specific distance during day-time (Table 4) and night-time (Table 5) is considered. The proposed method introduces a regularization method to make the deep structure more sensitive to images with high intra-class variance and low inter-class variance in datasets consisting very few data samples with high complexities. It is noticed that the augmented dataset is too small including 6 samples per distance. The intention of proposed sensitivity approach is to highlight the small variations and emphasize the high frequency components for better internal representation of features and further generalization while the dataset has just a few samples. The second scenario investigates the recognition task of images of one modality (either day-time (Table 6) or night-time (Table 7)) for a particular distance compared with a pool of other distances in which the complexity of the dataset increases. Images from one distance are selected for test process and images from other standoffs are integrated during the training process. Due to very small number of samples in the dataset, we integrate the images from other distances during the training process to make the dataset larger. Furthermore, integration of samples makes it more complicated due to high difference between the samples as the distance increases. This experiment illustrates the impact of the proposed method on the performance of deep-CNN by increasing both complexity and data samples. Third experimental setup is the recognition task in the pool of integrated images of night-time and day-time at different standoffs (Table 8) in which images of different modalities and distances are all mixed together to evaluate the

Download English Version:

https://daneshyari.com/en/article/4943060

Download Persian Version:

https://daneshyari.com/article/4943060

Daneshyari.com