Pattern Recognition Letters 107 (2018) 25–32

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Face biometric quality assessment via light CNN

Jun Yu, Kejia Sun, Fei Gao*, Suguo Zhu

Key Laboratory of Complex Systems Modeling and Simulation, School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China

ARTICLE INFO

Article history:
Available online 1 August 2017

JEL classification:
41A05
41A10
65D05
65D17

Keywords:
Biometric quality assessment
Convolutional neural networks (CNNs)
Face recognition
Image quality assessment
Video surveillance

ABSTRACT

In this paper, we proposed a novel biometric quality assessment (BQA) method for face images and explored its applications in face recognition. Here, we considered five categories of common homogeneous distortion in video suvillance applications, i.e. low-resolution, blurring, additive Gaussian white noise, salt and pepper noise, and Poisson noise. In the BQA model, we first learnt a classifier to simultaneously predict the categories and degrees of the degradation in a face image. Because the quality labels are often ambiguous and inaccurate, we used a light convolutional neural network with the Max-Feature-Map units to make the BQA model robust to noisy labels. Afterwards, we calculated the biometric quality score by pooling such predictions based on the recognition confidence of each degradation class. Finally, we proposed one promising strategy for developing reliable face recognition systems based on this BQA method. Thorough experiments have been conducted on the CASIA, FLW, and YouTube databases. The results demonstrate the effectiveness of the proposed BQA method.

1. Introduction

Face recognition system plays a significant role in the public security area. Unfortunately, in practical video surveillance scenarios, face images may suffer from various degradations during the process of capture, compression, and transformations [7]. For example, face images might be of low-resolution, blur, or noisy. Such degradations lead to great challenges in the face recognition system and tremendously decrease the face recognition performance. It is critical to deploy reliable face recognition systems for video surveillance applications [6].

For this purpose, one possible way is using biometric quality assessment (BQA) [4] techniques. The goal of BQA is estimating the ability of an image to function as a biometric. For face recognition systems, BQA algorithms can automatically predict whether a face image is proper for person identification or verification. Besides, BQA algorithms can be used to monitor the video surveillance cameras, to optimize the face image processing algorithms [15], to select "good" images from videos [30,35], and to design proper face recognition algorithms [26], etc. All these utilizations can improve both the effectiveness and robustness of a face recognition system.

Initially, international researchers and organizations devote great efforts to propose standards of acquiring proper face images for identification and verification [25]. These standards contain a number of instructions for judging the quality of face image based on various factors, e.g. brightness, facial pose, emotion, occlusion, etc. Accordingly, researchers define the biometric quality as the degree one face image departures from the standard frontal face image. They typically use hand-crafted features, e.g. the Local Binary Pattern (LBP) and Scale Invariant Feature Transform (SIFT), to represent a face image and evaluate the biometric quality based on these instructions [3,12,21,38].

Another pipeline of BQA is to evaluate the visual quality of a face image, following the idea of image quality assessment (IQA) [9,11,16]. Such methods are adept at estimating visual quality degradations, such as noise, compression artifacts, and blurring [17,26]. For example, Gunasekaret al. [15] release a face quality assessment database and use the statistics of Discrete Cosine Transform (DCT) coefficients to characterize quality degradations. However, face images are rather specific in term of content. It is critical to explore specific features for developing reliable face BQA models.

Lately, Beveridge et al. [2], Phillips et al. [24] find that, given a face image, its biometric quality is not only dependent on its own visual quality but also highly related to both the test image and the query image. As a result, researchers propose to denote the biometric quality as the matching between the test image and a reference image [1,21]. In the face recognition system, the neutral frontal face image always achieves better recognition performance than those with variations, it is thus regarded as the reference.

* Corresponding author.
E-mail address: gaofei@hdu.edu.cn (F. Gao).

http://dx.doi.org/10.1016/j.patrec.2017.07.015
0167-8655/

Recently, deep learning techniques, especially Convolutional Neural Networks (CNNs), have lead to great success in solving various image processing problems [10]. Researchers are therefore motivated to learn BQA models using CNNs. For example, Vignesh et al. [30] use the matching score between a given face image and the reference image as the quality index. Specially, they use both the local binary pattern (LBP) and histogram of gradient (HOG) for feature extraction, and use the mutual subspace method (MSM) to calculate the matching score. Finally, they adopt such matching scores as the quality labels and learn a quality assessment model via an eight-layer CNN. More recently, Pan et al. [22] adopt the Deep Face network [23] to extract the feature representation, and use the Probabilistic Linear Discriminant Analysis (PLDA) [5] to estimate the distance between a given image and the reference image. Finally, they learn a quality assessment model via the VGG-16 network [27]. However, the reliability of the generated quality ratings is curious. Besides, the size of quality-labeled face images is relatively small.

In this paper, we proposed a novel BQA method for face images and explored its applications in face recognition. Specially, we considered five common homogeneous distortion categories in video surveillance applications, i.e. low-resolution, blurring, additive Gaussian white noise, salt and pepper noise, and Poisson noise. In the BQA model, we first learnt a classifier to simultaneously predict the categories and degrees of the degradation in a face image. Specially, we used the light convolutional neural network (CNN) with the Max-Feature-Map units, because the quality labels are often ambiguous and inaccurate. Afterwards, we calculated the biometric quality score by pooling such predictions based on the recognition confidence of each degradation class. Here, we evaluated the recognition confidence of one degradation class by statistically measuring its impact on the recognition performance. Finally, we proposed one promising strategy for developing reliable face recognition systems based on this BQA method. Thorough experiments were conducted on the CASIA, FLW, and YouTube databases. The results demonstrate the effectiveness of the proposed BQA method.

Our contributions are briefly four-folds:

(1) We learnt a robust BQA model from noisy degradation labels by employing the light CNN model with Max-Feature-Map;
(2) We statistically and numerically considered the relationships between the quality degradation and the face recognition performance in the estimation of face image quality;
(3) We proposed one potential use of BQA for improving the reliability of face recognition systems; and
(4) The proposed BQA algorithms is highly consistent with human perception.

The rest of the paper is organized as follows. Section 2 details the framework of the proposed BQA method. Afterwards, the experiment settings and results are presented in Sections 3. Finally, Section 4 concludes this paper with directions for future work.

2. BQA for face recognition

The pipeline of the proposed BQA method is shown in Fig. 1. In this work, we first formulated several common distortions and correspondingly generated a large number of degraded face images. Afterwards, we adopted a light CNN with MFM units as our network prototype, and learnt a classifier to simultaneously predict the categories and degrees of the degradation in a face image. Then, we calculated the biometric quality score by pooling these predictions based on the impact of each degradation on the recognition performance. Finally, we proposed to embed this BQA model into the face recognition system to improve its effectiveness. We will sequentially introduce each stage in the following subsections.

2.1. Degradations

In practical video surveillance scenarios, face images may suffer from various degradations during the process of capture, compression, and transformation. In this paper, we considered five common homogeneous distortion categories, i.e. low-resolution, blurring, additive Gaussian white noise, Salt and pepper noise, and Poisson noise. Because there is a small number of face images with quality notations, we formulated these distortions and correspondingly generated a large number of degraded face images for learning and evaluating BQA models.

Preprocessing: We first detected and extracted the face area in a given image. All face images are converted to gray-scale and normalized to 128×128 via landmarks. The normalized image is illustrated in Fig. 2. According to the 5 facial points extracted by Sun et al. [28] with manual adjustment, we rotated two eye points to be horizontal, which can overcome the pose variations in roll angle. The distance between the midpoint of eyes and the midpoint of mouth, as well as the y axis of the midpoint of eyes, are set to be 40 pixels, because the distance between the midpoint of eyes and the midpoint of mouth is relatively invariant to pose variations in yaw angle.

Afterwards, we generated the degraded versions of each face image as the following.

- *Low Resolution with Nearest Neighbor Interpolation (LR)*: Large distance and low resolution face are two of the most challenging issues in video surveillance. To formulate the quality degradation caused by low resolution, we down-sampled the original face image with a factor r_{LR}. Here, we chose $r_{LR} \in \{0.7, 0.55, 0.4\}$ and generated three low resolution face images from each original image. In addition, we resized the low resolution image to 128×128 by using the nearest neighbor interpolation method, because the input of the CNN should be fixed to 128×128.
- *Gaussian Blurring (Gblur)*: We filtered the gray image by using a circular-symmetric 2-D Gaussian kernel of size $w \times w$ pixels and standard deviation σ_B. In the implementation, we empirically set $(w, \sigma_B) \in \{(3, 0), (5, 0), (7, 0)\}$, and generated three blurring versions of each face image. It can be observed that the blurring operation weakens the details in the image. As shown in Fig. 2, it is notable that images with Gblur are different from the low-resolution images, in terms of both the image content and the visual quality.
- *Additive Gaussian White Noise (AWGN)*: In the process of capture, transformation, and processing, additive Gaussian white noise might be introduced into the face image. Besides, most additive noise can be approximated by AWGN. We thus formulated AWGN and added AWGN to each face image. The strength of AWGN is controlled by the standard deviation σ_A. In the implementation, we empirically set $\sigma_A \in \{20, 40, 60\}$ and generated three noisy versions of each face image.
- *Salt and Pepper Noise (SPN)*: In the process of capture, transformation, and compression, salt and pepper noise might occur in a face image. SPN is shown as black or white points in an image. The strength of SPN is measured by the percent of noisy points, corresponding to the total number of pixels in the image. Specially, we set $p_s \in \{10, 20, 30\}$ in the implementation.
- *Poisson Noise (PN)*: The camera in the endoscope is a photon counter, and thus introduces noise that follows the Poisson distribution. We formulated the Poisson noise here. Specially, for each pixel, the corresponding output is generated by Poisson distribution with mean of a luminance value l_{PN}. Specially, we set $l_{PN} \in \{30, 50, 70\}$ in the implementation.

As a result, we generated 15 degraded versions (5 distortions $\times$ 3 degrees/distortion) of each face image. Fig. 2 illustrates one