Face Recognition Benchmark with ID Photos

Dongshun Cui, Guanghao Zhang, Kai Hu, Wei Han and Guang-Bin Huang

Abstract With the development of deep neural networks, researchers have developed lots of algorithms related to face and achieved comparable results to human-level performance on several databases. However, few feature extraction models work well in the real world when the subject which is to be recognized has limited samples, for example, only one ID photo can be obtained before the face recognition task. To our best knowledge, there is no face database which contains ID photos and pictures from the real world for a subject simultaneously. To fill this gap, we collected 100 celebrities' ID photos and their about 1000 stills or life pictures and formed a face database called **FDID**. Besides, we proposed a novel face recognition algorithm and evaluated it with this new database on the real-life videos.

Keywords Face recognition · Face recognition benchmark · ID photos · Total loss function · Real-life face recognition system

D. Cui (✉)
Energy Research Institute @ NTU (ERI@N), Interdisciplinary Graduate School, Singapore, Singapore
e-mail: dcui002@ntu.edu.sg

D. Cui · G. Zhang · W. Han · G.-B. Huang
School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore
e-mail: gzhang009@ntu.edu.sg

W. Han
e-mail: hanwei@ntu.edu.sg

G.-B. Huang
e-mail: egbhuang@ntu.edu.sg

K. Hu
College of Information Engineering, Xiangtan University, Xiangtan, China
e-mail: kaihu@xtu.edu.cn

H. Lu and X. Xu (eds.), *Artificial Intelligence and Robotics*,
Studies in Computational Intelligence 752,
https://doi.org/10.1007/978-3-319-69877-9_4

1 Introduction

Tasks related to face like face recognition have been active research fields for many years. Human-level performance has been achieved by machine-learning based algorithms for face recognition on several existing datasets (*e.g.*, deep neural network [1–3] and extreme learning machines [4]). Some famous and popular face recognition database is summarized and compared in Table 1.

We consider the ID photo of a person is critical for face recognition since the only information we can obtain is ID photo under several specific scenarios before there is a request to recognize some person from the videos or in the real world. For example, it is likely that only ID photos are available before the arrest of criminals. This requires face recognition systems to extract enough useful features from an ID photo and recognize criminals with these representations. Another example is that when offering an intelligent service to some very important persons (*e.g.*, the head of a country) but it is inconvenient to scan them and acquire face samples. The reasons why we choose celebrities as the subjects are:

1. Large quantities. Huge data is the foundation of training a deep neural network [5–7], while the selected celebrities are very famous and large amounts of their stills and life photos (collectively called "non-ID photos") are available on the Internet.
2. Rich scenes. Compared to the ordinary people, celebrities live in much more rich scenes which leads to generating more face images with variable backgrounds and illumination conditions.

We have the largest average image quantity (donated as ♯) for each person among all existing database. ♯ is calculated by $\sharp = Q/N$, while Q is the total images of the database and N is the number of subjects. The comparison between the existing datasets and our FIFD dataset is shown in Table 1.

Table 1 The comparison between existing datasets

Dataset	Year[a]	N	Q	♯	ID photo	Pose variation
LFW	2007	5749	13233	2.30	N	Limited
WDRef	2012	2995	99773	33.31	N	Limited
CACD	2014	2000	163446	81.72	N	Limited
CASIA WebFace	2014	10575	494414	46.75	N	Full
FaceScrub	2014	530	106863	201.63	N	Limited
SFC	2014	4030	4400000	1091.81	N	Limited
CelebA	2015	10177	202599	19.91	N	Limited
IJB-A	2015	500	5712	11.42	N	Full
Megaface	2016	672057	1027060	1.49	N	Full
FIFD (Ours)	2017	100	112839	1128.39	Y	Full

[a]The time of publication of the corresponding papers

The main contributions of our work are:

- We have leveraged a face benchmark which consists of ID photos and non-ID photos of 100 (Chinese) celebrities for face recognition. The way to build the dataset and detailed analysis of the dataset are explained.
- We have proposed a novel architecture for general face recognition and tested it with the real-world videos to illustrate the quality of our FIFD dataset.

The rest of the paper is organized as follow: Sect. 2 introduces the related work on existing scientific methods on creating face databases and the state-of-the-art face recognition algorithms. Then we explain the details on how we collect our database and data diversity is shown by using the images from our database. In Section, we propose a novel face recognition architecture and show the results of our methods on the real-life videos by training a model with our database. A conclusion is given, and future work is claimed in Sect. 5.

2 Related Work

2.1 Protocol of Building Face Recognition Database

The database is essential for training face recognition algorithms to achieve models and evaluating their performance. And there is no doubt that algorithms benefit a lot from a comprehensive and exquisite database which requests a scientific process. We have introduced the size of the existing popular datasets in Sect. 1, and here the systematic procedure of creating a face recognition database is to be introduced.

The general steps of building a face database are collection, cleaning, and arrangement of face images. Experimental and real-life environment are the two sources for gathering images, and the Internet provides a convenient way to collect real world pictures. Face image cleaning consists of face detection, face alignment, and duplication elimination. All of these fields have been explored for over 20 years, but still provide no perfect solutions [8–10]. Finally, face image annotation (manually or automatically) and stored by the order are done.

2.2 Existing Face Recognition Algorithms

Here we list some state-of-the-art face recognition algorithms on LFW dataset. DeepFace is the first wide accepted face recognition algorithm that approaches human-level performance on LFW (97.35%) [1]. It follows the general process pipeline of face detection, face alignment, face representation and face verification. DeepFace trains an effective nine-layer deep neural network (DNN) and tune 120 million parameters with no weight sharing between local connections. Later,

DeepID, DeepID2, DeepID2+, and DeepID3 are proposed by modifying the structure of the network (for example, DeepID adopts Convolutional Network.) and increase the accuracy further [2, 11–13].

FaceNet is proposed by using triplet loss (each pair of triplet consists of an anchor sample x^a, a positive sample x^p, and a negative sample x^n) and increase the accuracy to 99.63%. The loss function is

$$\sum_{i}^{N} \left[\|x_i^a - x_i^p\|_2^2 + \alpha - \|x_i^a - x_i^n\|_2^2 \right]. \tag{1}$$

Here, α is a constant which guarantees the distance between positive and negative pairs.

3 Our Database

There are massive image data on the Internet, and we can find a huge amount of stills and life photos for most celebrities. In our task, only the celebrities of whom we can obtain a clear ID photo will be put into our dataset.

3.1 Collection Rules and Flowchart

The target of or work is to build a face recognition database which contains one ID photo and as more non-ID photos as possible. These non-ID photos should be diverse enough to simulate the real-world scenario. With the help of search engines (like Google, Baidu), we have obtained the ID photos of 100 Chinese celebrities so far. Folders are created and named after the celebrities, and ID photos are stored in the corresponding folders.

3.2 Resources of the ID and Non-ID Photos

There are lots of pictures on the Internet for the celebrities we can collect, and the first step is to find celebrities' ID photos. So far, we have gathered 100 celebrities' ID photos and stored these ID photos separately. With these celebrities' name and the full consideration of data diversities, we continue to search their stills and life photos mainly from the popular movie databases (*e.g.*, IMDb, Mtime), famous image search engines (*e.g.*, Google Images and Baidu Image), and some social networks (*e.g.*, Baidu Tieba, Sina Weibo).

Fig. 1 Some of the ID photos in our collected database. Note that these are Chinese ID-card photos, and people can have slight smiles when they take photos

3.3 Insight of the Proposed Database

We have created the face database (FDID) with the ID photos of the celebrities from the internet and as many of their corresponding non-ID photos as possible. We have randomly selected 36 celebrities (18 males and 18 females) and shown their ID photos in Fig. 1.

The reason why we pay more attention to the ID photo is that it usually provides one person's most information among all his face images. Here we have listed some properties of an ID photo in the passport [14]:

1. No head pose.
2. Neutral expression.
3. No occlusions (*e.g.*, glasses and marks) on the face.

Obviously, these rules echo the main challenges of real-world face recognition applications since non-ID photos always don't fulfill one or several of these rules. To make an intuitive comparison, we randomly select four celebrities (two males and two females) to show the diversity contents of our database. Examples of face images with large-angle head pose, various expressions, and occlusions from our database are shown in Fig. 2a, b, c respectively.

(a) Different head poses.

(b) Various facial expressions.

(c) Occlusions.

Fig. 2 Examples of face images from our database

4 Proposed Method and Results

Based on our database including ID photos and non-ID photos, we re-design the architecture of the face recognition as shown in Fig. 3.

In this section, we will introduce a novel and general face recognition loss function for our database. We point two kinds of intra-person variances based on the assumptions of the same person's non-ID photo should be similar to his/her ID photo and the same person should look similar to his/her other non-ID photo.

We use x_{ij} indicates the j-th photo of the i-th person in the dataset. To distinguish the ID photo and non-ID photos, we assume $j = 0$ denotes the former while $j \in [1, m]$ indexes the latter. Assume $r(x)$ as the representation of sample x, it can be manually descriptors (*e.g.*, LBP, Gabor, and Eigenvector) or learned by a neural network (*e.g.*, AlexNet, PCANet, and CNN). Dissimilarity function is expressed as $d(\cdot)$. These two kinds of intra-person variances yield the following two loss functions.

The first intra-person loss function is:

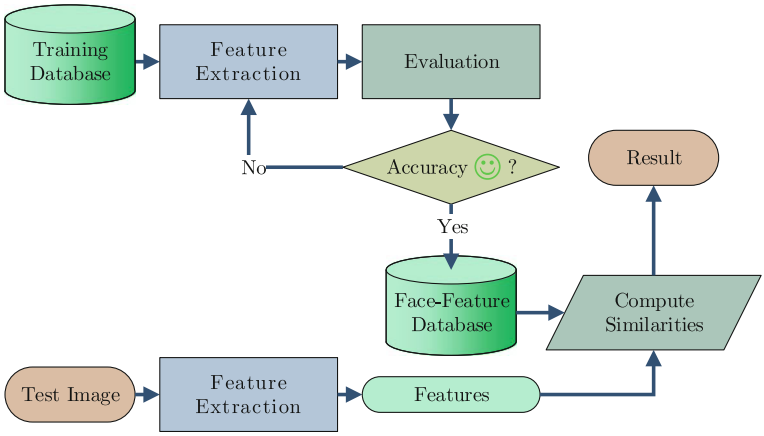

Fig. 3 The proposed architecture of the face recognition

$$\mathcal{L}_1 = \sum_{i\in[1,n]} \sum_{j\in[1,m]} d_1(r(x_{i0}), r(x_{ij})) \tag{2}$$

The second intra-person loss function is:

$$\mathcal{L}_2 = \sum_{i\in[1,n]} \sum_{j_1,j_2\in[1,m]} d_2(r(x_{ij_1}), r(x_{ij_2})). \tag{3}$$

Intra-person loss functions constraint the similarity between the photos of the same person, and for face recognition, we also need to constraint the dissimilarity between the photos of the different persons.

The inter-person loss function is:

$$\mathcal{L}_3 = \sum_{i_1,i_2\in[1,n]} \sum_{j_1,j_2\in[1,m]} d_3(r(x_{i_1j_1}), r(x_{i_2j_2})). \tag{4}$$

So, the total loss function $\mathcal{L}_{total}$ for the set of face images is:

$$\mathcal{L}_{total} = \lambda_1\mathcal{L}_1 + \lambda_2\mathcal{L}_2 - \lambda_3\mathcal{L}_3 \tag{5}$$

where $\lambda_1, \lambda_2, \lambda_3$ are the weights of $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$ respectively. Our target is to minimize this loss function by learning an effective model to represent the face images, which is expressed as:

$$\mathcal{T}(r, \lambda_i) = min(\mathcal{L}_{total}). \tag{6}$$

Currently, most face recognition models are the instances of Eq. 6 by setting $\lambda_1 = \lambda_2$.

Fig. 4 Results of the proposed method on real-life videos (frame sequences) by training on our database. Each row shows the face recognition result of the same celebrity

We train a model with our dataset by adopting a similar network structure proposed in [1] by replacing the softmax loss function with ours and test it with our database and some real-life videos. ID and non-ID photos of 80 of our collected celebrities are randomly selected for training and ten celebrities' photos for verifying, and the best feature model (mainly tuning the parameters of the networks) and dissimilarity computation algorithm are achieved. The remaining 10 celebrities' non-ID photos are used to test, and we achieve an accuracy of near 80%. Besides, we collect these 10 celebrities' videos from YouTube and transfer them into frames. Then we implement the same procedures of preprocessing (*e.g.*, face detection), and extract features with the optimized model. Finally, we compute the dissimilarity between features of the input face images and all the candidate ID photos and output the celebrity's name which corresponded to the minimum dissimilarity value. We have an ID photo and extract features from it. Then for the input videos, we do the face detection, feature extraction, and dissimilarity computation. Finally, we output the face identification results which are demonstrated in Fig. 4.

5 Conclusion

We have created a new database which contains ID photos of 100 celebrities with their over 1000 stills and life photos. Details on how to build collect this database are introduced, and the data diversity on faces with different head poses, facial expressions, occlusions, illuminations, degrees of toilette, and ages are shown. Besides, we proposed a novel architecture on face recognition when the ID photos are available, and new total loss function which contains two intra-person loss functions and one inter-person loss function are presented. Models are trained with the proposed database using our target function, and video-level experiments are performed to demonstrate the meaning of our database and the effectiveness of the proposed method.

References

1. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)
2. Sun, Y., Wang, X., Tang, X.: Deeply learned face representations are sparse, selective, and robust. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2892–2900 (2015)
3. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823 (2015)
4. Tang, J., Deng, C., Huang, G.-B.: Extreme learning machine for multilayer perceptron. IEEE trans. Neural Netw. Learn. Syst. **27**(4), 809–821 (2016)
5. Lu, H., Zhang, L., Serikawa, S.: Maximum local energy: an effective approach for multisensor image fusion in beyond wavelet transform domain. Comput. Math. Appl. **64**(5), 996–1003 (2012)
6. Xu, X., He, L., Shimada, A., Taniguchi, R.-I., Lu, H.: Learning unified binary codes for cross-modal retrieval via latent semantic hashing. Neurocomputing **213**, 191–203 (2016)
7. Lu, H., Li, B., Zhu, J., Li, Y., Li, Y., Xu, X., He, L., Li, X., Li, J., Serikawa, S.: Wound intensity correction and segmentation with convolutional neural networks. Concurr. Comput. Pract. Exp. **29**(6) (2017)
8. Kawulok, M., Celebi, M.E., Smolka, B.: Advances in Face Detection and Facial Image Analysis. Springer (2016)
9. Liu, Q., Deng, J., Tao, D.: Dual sparse constrained cascade regression for robust face alignment. IEEE Trans. Image Process. **25**(2), 700–712 (2016)
10. Tang, J., Li, Z., Wang, M., Zhao, R.: Neighborhood discriminant hashing for large-scale image retrieval. IEEE Trans Image Process. **24**(9), 2827–2840 (2015)
11. Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1891–1898 (2014)
12. Sun, Y., Chen, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification-verification. In: Advances in Neural Information Processing Systems, pp. 1988–1996 (2014)
13. Sun, Y., Liang, D., Wang, X., Tang, X.: Deepid3: face recognition with very deep neural networks (2015). arXiv:1502.00873
14. Rules for passport photos. https://www.gov.uk/photos-for-passports/photo-requirements. Accessed 27 Feb 2017