Facial Appearance Description Through Facial Landmarks Computation

Na Liu[1], Hao Ge[1], Lei Song[1,2(✉)], and Huixian Duan[1,3(✉)]

[1] Cyber Physical System R&D Center,
The Third Research Institute of Ministry of Public Security, Shanghai, China
songlei93l2@l26.com, hxduan005@l63.com
[2] Shenzhen Key Laboratory of Media Security,
Shenzhen University, Shenzhen, China
[3] Shanghai International Technology & Trade United Co., Ltd., Shanghai, China

Abstract. Face appearance descriptions are semantic meaningful characteristics and beneficial for face recognition and retrieval. In this paper, we propose a facial appearance description method which can describe the whole face, chin, eyebrow, eye, nose and mouth type separately. The description is obtained through facial landmarks computation and geometry shape estimation of each part. Based on this method, semantic search of face images can be achieved on face dataset. What's more, the large scale dataset can be categorized though the facial appearance description before recognition which can help to improve the recognition accuracy and efficiency.

Keywords: Facial appearance description · Landmarks · Facial type

1 Introduction

Face recognition is playing an increasing significant role in public security since it can help to find the object identity quickly based on the captured face images. Recently with the rapid progress of deep learning, face recognition has made great strides and many deep learning based methods performs better than human levels [1–5].

Although the applications of face recognition under controlled conditions have become mature, there are still some problems in real applications, especially under uncontrolled environment. Firstly, the appearance of face image is likely to be affected seriously due to internal or external variations, such as illumination, partial occlusion, expression variations or image quality problem, and the performance will drop dramatically accordingly [6]. Secondly, the precision of searching a person in large scale face dataset can't be guaranteed. Furthermore, the face image is not able to be captured if there is no surveillance camera, and only the description of witness can provide some clues.

Considering the above several problems, facial appearance description is critical when the face images cannot be used for recognition. Firstly, we can search the person through the facial description when the quality of the face image is not so good for recognition or only the witness oral accounts could be used. What's more, facial appearance description can help to filter the large scale dataset and improve the recognition speed and accuracy.

G. Sun and S. Liu (Eds.): ADHIP 2017, LNICST 219, pp. 103–108, 2018.
https://doi.org/10.1007/978-3-319-73317-3_13

Related existing work is face attribute prediction which analyzes a series of face related attributes, including age, gender, hair color, hair style, smile intensity, head pose, eye status, etc. Liu et al. construct two face attribute datasets, i.e. CelebA and LFWA [7], by labeling images selected from the database of Celeb-Faces [8] and LFW [9]. Images in CelebA and LFWA are annotated with 40 facial attributes. There are two categories of face attribute recognition research, holistic and local methods. Holistic methods extract the attribute features from the whole image [10]. The performance of holistic methods will be affected by the deformations of objects. On the other hand, local methods first extract the attribute parts and then do attribute recognition based on each part's features [11–14]. The local methods depend on the precision of face and landmark localization.

As we know, the description of facial organs is more stable than other attributes, such as hair style and color, whether or not wearing accessories, etc. However, there is a lack of methods for detailed description of the facial organs. For example, CelebA and LFWA datasets only include one lips type which is big lip. Therefore, in this paper we propose a facial appearance description method, which includes the whole face, eye, nose, mouth, and eyebrow type description. We first detect the landmarks in each face image and extract the whole face and each organ part based on the landmarks locations. Then, we describe the whole face and facial organs type based on the geometrical shape estimation of each part. Specifically the face type is divided into wide face and long face, and the facial organ types are divided into pointy chin, square chin, small eye, big eye, arched eyebrow, straight eyebrow, big nose, small nose, thick lip and thin lip.

The rest of this paper is organized as follows. Section 2 describes the facial appearance description method. In Sect. 3, extensive experiments are performed to examine the effectiveness of the proposed method. Finally, we conclude the paper in Sect. 4.

2 Proposed Method

In this section, we introduce the description method of facial appearance characteristic. It consists of two major stages: landmark detection and facial appearance description.

For face landmark detection, we apply Zhang's tasks-constrained Deep Model [15], which takes face image as input and output the locations of 68 facial landmarks which contain face, eye, eyebrow, lip and nose contour. Illustrations of landmarks are shown in Fig. 1, in which the landmarks are shown as green dots. The advantage of this method is that it is more robust to occlusion and pose variations compared to existing methods.

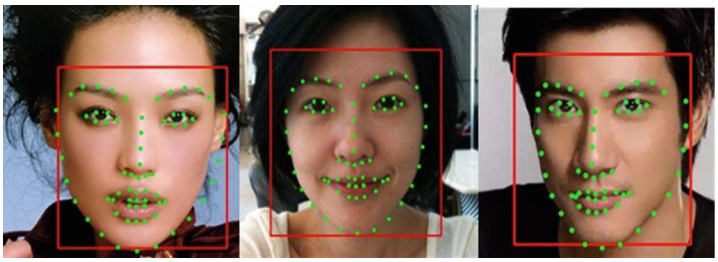

Fig. 1. Illustration of facial landmarks detection (Color figure online)

After the detection of 68 facial landmarks, we extract each face organs and do quantitative analysis based on the landmarks locations. For the computation convenience, we first define some important keypoints which include glabella point, gnathion, gonion, zygion. In this paper, we use the detected landmarks to approximate these keypoints, especially the glabella point is approximated by the two inner corners of the eyebrows. The illustration of these keypoints are shown on Fig. 2.

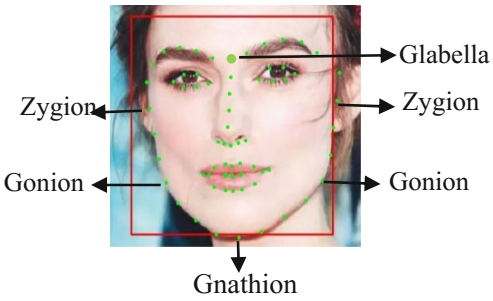

Fig. 2. Illustration of the defined keypoints

After the extraction of each part, we compute the index according to the geometry ratio of each part. The computation is done as following.

Facial index $= \frac{\text{facial height}}{\text{facial breadth}}$, where facial height is the distance from the glabella point to the gnathion, facial breadth is the distance between the left and the right zygion.

Chin index $= \frac{\text{gonion breadth}}{\text{facial breadth}}$, where facial breadth is defined as above and gonion breadth denotes the distance between the left and the right gonion.

Eyebrow index $= \frac{\text{eyebrow height}}{\text{eyebrow breadth}}$, where eyebrow height is the distance between the top and the bottom landmark on the eyebrow, and eyebrow breadth is the distance between the two brow corner points.

Eye index $= \frac{\text{eye height}}{\text{eye breadth}}$, where eye height denotes the distance between the top and the bottom landmark on eyes, eye breadth is the distance from the endocanthion to the ectocanthion.

Nose index $= \frac{1}{2}(\frac{\text{nose breadth}}{\text{facial breadth}} + \frac{\text{nose height}}{\text{facial height}})$, where facial breadth and heigth are defined as in facial index computation formula, nose breadth is the distance of two ala nasi points and nose height is defined as the distance between nasion point and rhinion point.

Mouth index $= \frac{\text{mouth height}}{\text{mouth breadth}}$, where mouth height is the mean of the upper lip and under lip height, mouth breadth is the distance between the two corner points of the mouth.

Based on the index value, we can classify the type of each part. The classes are as follows: the shape of face is divided into wide face and long face, the chin type is divided into pointy chin and square chin, the eye type is divided into big eye and small

eye, the eyebrow type is divided into straight brow and arched brow, the nose type is divided into big nose and small nose, the mouth is divided into thick mouth and thin mouth. In other words, the facial characteristics are transformed to standard semantic description through the above method. And on the basis of this, the person can be searched through the facial description when there is only facial characteristic feature can be used. What's more, the large scale dataset can be classified via the description and the recognition task can be done in a smaller set, so the recognition speed and accuracy can be improved accordingly.

3 Experiments

To verify the performance of the proposed method. We first construct a facial description database, by labeling images downloaded from the internet. There are 800 images in total. For each facial type, there are at least 50 images. Figure 3 shows the sample images of each type in the database.

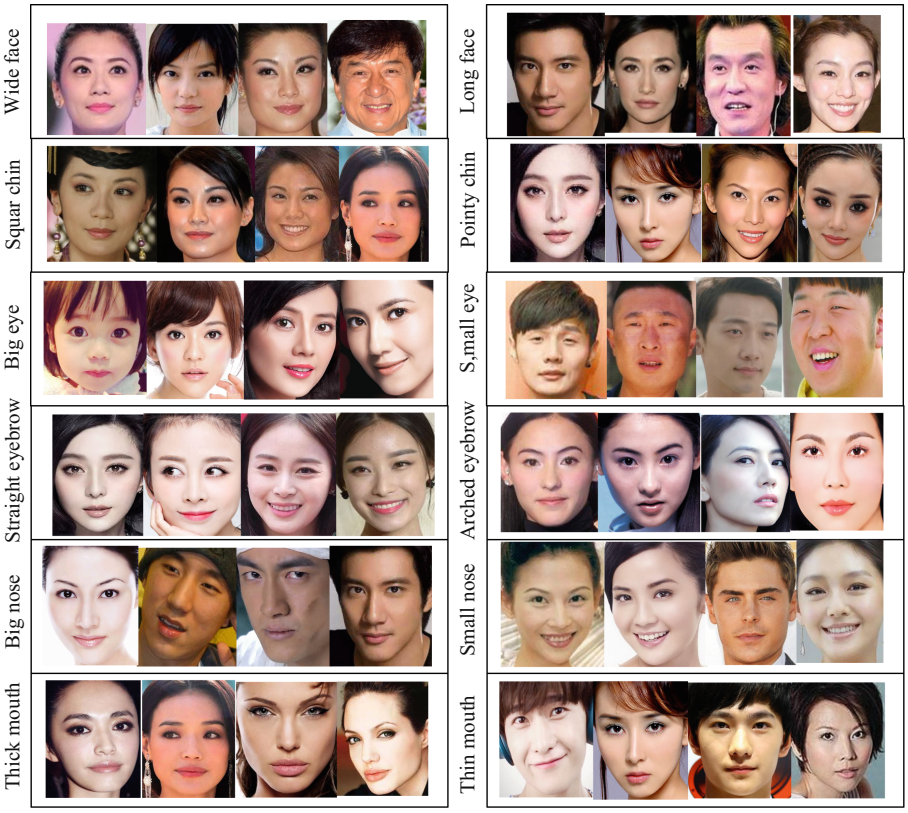

Fig. 3. Sample images in the database

For each type, we selected 30 images for training the description parameters. Then given an image, we can compute the index of each part and determine which type it belongs to. Some experimental results are illustrated in Fig. 4. From the results we can see that through the facial description, the face image can be expressed through standard semantic description. And we can use the descriptions to search the object or classify the large scale database. It should be noted that the facial appearance descriptions may not be accurate in every type classification due to internal or external variations. This is because the appearance of the face image changes due to illumination, partial occlusion, expression variations or image quality problem. So in some cases, the obvious facial features might be correctly described, but the others descriptions may be less accurate. So we should use the more effective descriptions in real applications. And we will do more research on this work.

Long face	Long face	Wide face	Long face
Pointy chin	Pointy chin	Square chin	Pointy chin
Straight eyebrow	Straight eyebrow	Arched eyebrow	Straight eyebrow
Big eye	Big eye	Big eye	Big eye
Big nose	Big nose	Big nose	Small nose
Thick mouth	Thick mouth	Thick mouth	Thin mouth

Fig. 4. Illustration of the experiment results

4 Conclusion

In this paper, we propose a facial appearance description method which is computed through the landmarks computation and geometry shape estimation of each part. It contains the whole face, chine, eyebrow, eye, nose and mouth type description. This work can help to recognize the identity of a person through semantic retrieval when there is only facial feature description can be utilized. In the future, we will study the more elaborate description method and its application in real security applications.

Acknowledgement. The authors of this paper are members of Shanghai Engineering Research Center of Intelligent Video Surveillance. This work is sponsored by the National Natural Science Foundation of China (61403084, 61402116); by the Project of the Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University (ESSCKF 2015-03); and by the Shanghai Rising-Star Program (17QB1401000).

References

1. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: Proceedings of CVPR (2014)
2. Sun, Y., Liang, D., Wang, X., Tang, X.: Deepid3: face recognition with very deep neural networks. arXiv:1502.00873 (2015)
3. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: a unified embedding for face recognition and clustering. In: Proceedings of CVPR (2015)
4. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: Proceedings of BMVC (2015)
5. Wen, Y., Zhang, K., Li, Z., Qiao, Yu.: A discriminative feature learning approach for deep face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 499–515. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_31
6. Liu, N., Lai, J.H., Zheng, W.S.: A facial sparse descriptor for single image based face recognition, **93**, 77–87 (2012)
7. Ouyang, W., Li, H., Zeng, X., Wang, X.: Deep learning face attributes in the wild. In: Proceedings of ICCV (2015)
8. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07–49, University of Massachusetts, Amherst, October 2007
9. Sun, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification-verification. In: NIPS (2014)
10. Razavian, A.S., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. arXiv:1403.6382 (2014)
11. Berg, T., Belhumeur, P.N.: Poof: part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation. In: Proceedings of CVPR (2013)
12. Luo, P., Wang, X., Tang, X.: A deep sum-product architecture for robust facial attributes analysis. In: Proceedings of ICCV (2013)
13. Zhang, N., Paluri, M., Ranzato, M., Darrell, T., Bourdev, L.: Panda: pose aligned networks for deep attribute modeling. In: Proceedings of CVPR (2014)
14. Bourdev, L., Maji, S., Malik, J.: Describing people: a poselet-based approach to attribute classification. In: Proceedings of ICCV (2011)
15. Zhang, Z., Luo, P., Loy, C.C., Tang, X.: Facial landmark detection by deep multi-task learning. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8694, pp. 94–108. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10599-4_7