Matching Depth to RGB for Boosting Face Verification

Han Liu, Feixiang He, Qijun Zhao(✉), and Xiangdong Fei

National Key Laboratory of Fundamental Science on Synthetic Vision,
College of Computer Science, Sichuan University, Chengdu, China
qjzhao@scu.edu.cn

Abstract. Low cost RGB-D sensors like Kinect and RealSense enable easy acquisition of both RGB (i.e., texture) and depth images of human faces. Many methods have been proposed to improve the RGB-to-RGB face matcher by fusing it with the Depth-to-Depth face matcher. Yet, few efforts have been devoted to the matching between RGB and Depth face images. In this paper, we propose two deep convolutional neural network (DCNN) based approaches to Depth-to-RGB face recognition, and compare their performance in terms of face verification accuracy. We further combine the Depth-to-RGB matcher with the RGB-to-RGB matcher via score-level fusion. Evaluation experiments on two databases demonstrate that matching depth to RGB does boost face verification accuracy.

Keywords: RGB-D · Deep learning · Score fusion · Face recognition

1 Introduction

In the past decade, face recognition achieved great progress thanks to the recent advances in deep neural networks. The latest works, such as [1–4], not only outperform traditional hand-crafted feature based face recognition, but are also at the brink of human level accuracy. Most of existing approaches rely on texture face images. As a result, they still suffer from serious performance degradation when pose, illumination and expression variations occur to the texture face images in uncontrolled scenarios.

Depth information is believed to be more robust to pose and illumination variations, and provide useful cues for facial expressions. Thanks to the availability of low cost RGB-D sensors (e.g., Kinect and RealSense), increasing research efforts have been denoted to exploring depth information for face recognition [6,7,12]. While most existing RGB-D based face recognition methods conduct RGB-to-RGB and Depth-to-Depth face matching separately, little attention has been paid to Depth-to-RGB face matching, which is demanded in some real-world applications. One example is when a user claims his/her identity in front of a RGB-D sensor using the identity (ID) card that contains his/her 2D face image. See Fig. 1. The deployment of RGB-D sensors at the verification phase

J. Zhou et al. (Eds.): CCBR 2017, LNCS 10568, pp. 127–134, 2017.
https://doi.org/10.1007/978-3-319-69923-3_14

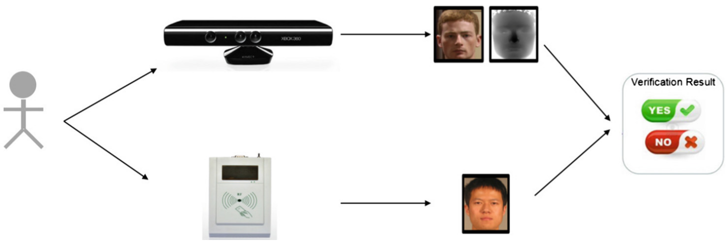


Fig. 1. An example RGB-D vs. RGB based face verification scenario: a user claims his/her identity in front of a RGB-D sensor using the identity (ID) card that contains his/her 2D face image.

can not only enhance the robustness of the face recognition system to spoofing attacks, but also boost the face recognition accuracy if the depth information is effectively utilized for face matching.

In this paper, we propose two different approaches to matching depth to RGB face images, i.e., image-mixing and image-fusion. Both approaches are implemented based on deep convolutional neural networks (DCNNs). The image-mixing approach mixes the RGB and Depth face images into a single training data set and trains a DCNN that takes either RGB or depth face images as input. The image-fusion approach first combines a pair of RGB and Depth face images into a single multi-channel image, and then trains a DCNN to classify such a multi-channel image into either genuine or imposter class with 'genuine' indicating that the source RGB and Depth face images are of the same subject and 'imposter' indicating different subjects. To show the effectiveness of Depth-to-RGB face matching in boosting face recognition accuracy, we further fuse the Depth-to-RGB and RGB-to-RGB face matchers at score level. Evaluation experiments have been conducted on two databases with a state-of-the-art RGB-to-RGB face matcher as baseline. The experimental results demonstrate the benefit of Depth-to-RGB face matching to face verification.

The rest of the paper is organized as follows. In Sect. 2, we discuss the existing RGB-D based face recognition methods and an asymmetric scenario considered in our work. The details of matching Depth-to-RGB face images as well as combining the RGB-to-RGB and Depth-to-RGB matchers are presented in Sect. 3. Evaluation experiments are then reported in Sect. 4. Finally, Sect. 5 concludes the paper.

2 Related Work

Existing RGB-D based face recognition methods mostly assume that RGB-D face data are captured at both enrollment and recognition stages. They usually utilize the depth information in two different ways. The first way is to match the probe Depth face image with the enrolled Depth face image, and combine the result with the matching result of probe and enrolled RGB face images [12].

A variety of feature representations have been applied to both RGB and Depth face images, including handcrafted features (e.g., LBP features [8], eigenface features [9]) and learning-based features (e.g., extracted features based on CNN [22]). Based on these features, various classifiers have been employed, such as Cosine distance metric [23], Joint Bayesian [2], SVM [12]. The second way is to render new face images at different pose angles from the RGB-D face data to enlarge the gallery database such that faces can be effectively recognized in a larger range of pose angles [11]. Unlike these existing RGB-D based face recognition methods, we in this paper consider an asymmetric scenario in which RGB-D face data need to be matched to RGB face images. Matching RGB-D face data to RGB face images, particularly matching Depth to RGB (i.e., Texture) face images, is essentially a heterogeneous face recognition problem. However, to the best knowledge of the authors, no research work on the problem of matching Depth to Texture has been published in the face recognition literature.

A number of RGB-D face databases (e.g., Lock3DFace [19]) are available in the public domain. However, all these RGB-D face databases are constructed for research on RGB-D to RGB-D face recognition. In this paper, we will choose from these public databases the RGB components captured in one session as gallery, and the RGB-D data captured in other sessions as probe to evaluate the RGB-D to RGB face recognition performance.

3 Proposed Methods

In this section, we first introduce in detail the proposed image-mixing and image-fusion approaches to matching Depth-to-RGB face images, and then present a score-fusion based RGB-D vs. RGB face verification method.

3.1 Image-Mixing Approach

Our proposed methods are based on a state-of-the-art deep neural network for face recognition, specifically, LightCNN [14]. LightCNN was initially designed for extracting facial features from RGB face images. It can achieve comparable face recognition accuracy with other complex networks by using a significantly smaller set of parameters. Therefore, we employ it as the baseline in our experiments. In order to enable LightCNN to extract discriminative features from both RGB and Depth face images, the image-mixing approach simply mixes the RGB and Depth face images to form a single training set. This training set is then used to finetune the original LightCNN. See Fig. 2(a).

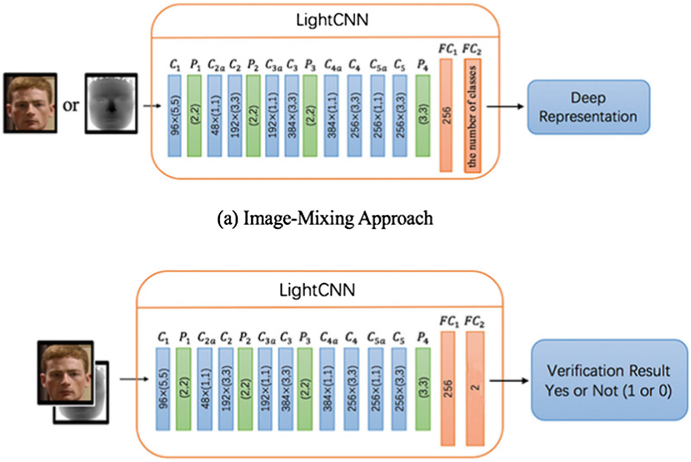
Once the LightCNN is finetuned by the mixed training set of RGB and Depth face images, it can be used to extract features from RGB or Depth face images by exporting the output of the FC_1 layer. Based on the extracted features, the Cosine distance metric [21] is used to measure the similarity between a pair of RGB and Depth face images.

3.2 Image-Fusion Approach

In the above image-mixing approach, we neither revise the structure of LightCNN, nor change its loss function. In this section, we propose another approach, namely image-fusion approach. This approach differs from the image-mixing approach mainly in two aspects. First, pairs of RGB and Depth face images are fused across channels, resulting in multi-channel face images, which are used as the input to the network in the image-fusion approach. See Fig. 2(b). Before the fusion, the pixel values in depth images are normalized via min-max normalization to the same range as the pixel values in RGB images. Second, the LightCNN is adapted for verification, rather than identification, in the image-fusion approach. To this end, the adapted LightCNN performs a two-class classification on the input multi-channel face image, i.e., whether the source pair of RGB and Depth face images are from the same subject (genuine class) or two different subjects (imposter class). It outputs a probability for each of the two classes, and assigns the input data into the class corresponding to the larger probability. Obviously, the image-mixing approach aims to seek for unified feature representations that are suitable for both RGB and Depth face images. In contrast, the image-fusion approach fulfills an end-to-end face verification through explicitly exploring the correlation between the RGB and Depth face images.

3.3 RGB-D vs. RGB Based Face Verification

In the scenario of RGB-D vs. RGB based face verification, we propose to conduct RGB-to-RGB face matching by using a conventional 2D face matcher, and mean-



(a) Image-Mixing Approach

(b) Image-Fusion Approach

Fig. 2. Matching Depth to RGB face images, (a) is Image-Mixing Approach, (b) is Image-Fusion Approach. Note that RGB images are converted to gray images.

while perform Depth-to-RGB face matching by using the methods presented in the past two sections. The matching results of the RGB-to-RGB and Depth-to-RGB face matchers are combined at the score level with a weighted sum fusion rule:

$$s = \lambda_1 s_r + \lambda_2 s_d \tag{1}$$

where s_r is the score of RGB-to-RGB face matching, s_d is the score of Depth-to-RGB face matching, λ_1 is the weight given to RGB-to-RGB matching, λ_2 is the weight of Depth-to-RGB matching, and s is the combined score employing the two models for the final verification.

Note that the similarity score output by the image-mixing approach is defined by the classifier (i.e., Cosine distance based nearest neighbor classifier in this paper) over the extracted features, and the similarity score output by the image-fusion approach is defined as the probability of the genuine class.

4 Experiments

4.1 Databases

Lock3DFace [19], a public RGB-D face database, is used in our experiments. In order to simulate the proposed scenario, we choose the neutral frontal RGB images as gallery set, and the remained RGB-D images as probe set which contain pose, illumination and expression variations.

The gallery RGB face images and the probe RGB-D face data in the above database is acquired by using the same devices in the lab. Therefore, we construct a private database by ourselves in real-world applications. We use the RealSense scanner mounted on a gate at the entrance of a railway station. When a subject passes through the gate, he/she is asked to stand in front of the gate and have his/her RGB-D face data collected. We finally acquire 100 RGB-D face images for each of 2,293 subjects. We call this database as Realistic RGB-D face database (in short, Realistic-RGB-D).

Following LightCNN training protocol, all face images are converted to gray-scale and resized to 144 × 144 based on the face rectangle detected by MTCNN. Afterwards, RGB face images are divided by 255 while Depth maps are divided by its maximum distance, both of which are normalized to $[0, 1]$. Due to deep learning is robust to translation and scale but not to rotation, we rotate two eye points to be horizontal according to the 5 facial points detected by [20], which can overcome the pose variations in roll angle. The training data sets are further enlarged by applying various tricks, including random cropping, mirroring and scale jittering.

4.2 Evaluation Protocol

To train the proposed Depth-to-RGB face matchers, we choose a subset from both Lock3DFace and Realistic-RGB-D as training data. Specifically, the training set contains 450 subjects and the left 50 subjects are used for testing on

Lock3DFace while 2,164 subjects are used for training and 129 subjects for testing on Realistic-RGB-D.

In order to avoid imbalanced data and overfitting, the training of the image-fusion based Depth-to-RGB face matcher requires that one half of the image pairs are genuine and the other half of the image pairs are impostor, produced by randomly pairing images of different subjects. For testing, we randomly form 1,500 genuine pairs and 1,500 imposter pairs for Lock3DFace which mainly contains expression and pose variations, and 1,040 genuine pairs and 1,040 imposter pairs for Realistic-RGB-D, which mainly contains pose and illumination variations.

4.3 Results

Table 1 shows the accuracy of the proposed two approaches in terms of the percentage of correctly recognized testing image pairs. Note that for the image-mixing approach, a similarity threshold is required to determine whether an image pair is genuine or imposter. Here, we choose the threshold that maximizes the accuracy. As can be seen, the image-fusion approach performs better than the image-mixing approach on both of the two databases. This proves the importance of exploring the correlation between the depth and texture (i.e., RGB) information of faces, which is useful for boosting face recognition accuracy.

Table 1. Face verification accuracy of the proposed Depth-to-RGB matchers on Realistic-RGB-D and Lock3DFace databases.

Test set	Method	Accuracy
Realistic-RGB-D	Image-Mixing	79.3%
	Image-Fusion	86.7%
Lock3DFace	Image-Mixing	79%
	Image-Fusion	87.4%

To evaluate the contribution of depth information to existing RGB-based face matchers, we combine our proposed Depth-to-RGB face matchers with the baseline RGB-to-RGB LightCNN matcher via score level weight sum fusion (refer to Sect. 3.3). Table 2 lists the obtained recognition accuracy (i.e., True Positive Rate or TPR when False Positive Rate or FPR is 1% or 0.1%) on the Realistic-RGB-D and Lock3DFace databases. From these results, we can clearly see that the proposed image-fusion approach, compared with the image-mixing approach, makes a larger improvement thanks to its more effective utilization of the depth information.

Table 2. Face verification accuracy of the baseline LightCNN matcher and its fusion with the proposed Depth-to-RGB matcher on Realistic-RGB-D and Lock3DFace databases.

Test set	Method	TPR@FPR = 1%	TPR@FPR = 0.1%
Realistic-RGB-D	LightCNN	97.8%	94.5%
	Image-Mixing + LightCNN	98%	95.7%
	Image-Fusion + LightCNN	98.4%	96.2%
Lock3DFace	LightCNN	95.5%	91%
	Image-Mixing + LightCNN	95.6%	91.2%
	Image-Fusion + LightCNN	96%	92.3%

5 Conclusions

In this paper, we have proposed two methods, referred to as image-mixing and image-fusion approaches, to solve the problem of matching Depth to RGB face images. Image-fusion approach proves to be more effective than image-mixing approach. This indicates the advantage of exploring the correlation between Depth and RGB face images in boosting face recognition accuracy. Furthermore, depth maps, as a complementary type of information, improve the verification accuracy of state-of-the-art face recognition methods, as shown in the experiments on two databases. The two approaches proposed in this paper essentially perform on the image data level. In the future, we are going to collect ID photos as our gallery to complete our experiments. Furthermore, we will tackle the Depth-to-RGB face matching problem from a feature transform perspective.

Acknowledgments. This work is supported by National Key Research and Development Program of China (2017YFB0802303, 2016YFC0801100) and the National Key Scientific Instrument and Equipment Development Projects of China (2013YQ49087904).

References

1. Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1891–1898 (2014)
2. Sun, Y., Wang, X., Tang, X.: Deeply learned face representations are sparse, selective, and robust. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2892–2900 (2015)
3. Taigman, Y., Yang, M., Ranzato, M.A., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)
4. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: European Conference on Computer Vision, pp. 499–515 (2016)

5. Lin, C.P., Wang, C.Y., Chen, H.R., Chu, W.C., Chen, M.Y.: RealSense: directional interaction for proximate mobile sharing using built-in orientation sensors. In: 21st ACM International Conference on Multimedia, pp. 777–780 (2013)
6. Li, W., Li, X., Goldberg, M., Zhu, Z.: Face recognition by 3D registration for the visually impaired using a RGB-D sensor. In: European Conference on Computer Vision, pp. 763–777 (2014)
7. Hayat, M., Bennamoun, M., El-Sallam, A.A.: An RGB based image set classification for robust face recognition from Kinect data. Neurocomputing **171**, 889–900 (2016)
8. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
9. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**(1), 71–86 (1991)
10. Goswami, G., Vatsa, M., Singh, R.: RGB-D face recognition with texture and attribute features. IEEE Trans. Inf. Forensics Secur. **9**(10), 1629–1640 (2014)
11. Ciaccio, C., Wen, L., Guo, G.: Face recognition robust to head pose changes based on the RGB-D sensor. In: 6th International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–6 (2013)
12. Lee, Y., Chen, J., Tseng, C.W., Lai, S.H.: Accurate and robust face recognition from RGB-D images with a deep learning approach. In: British Machine Vision Conference (2016)
13. Sarfraz, M.S., Stiefelhagen, R.: Deep perceptual mapping for thermal to visible face recognition (2015). arXiv preprint arXiv:1507.02879
14. Wu, X., He, R., Sun, Z., Tan, T.: A light CNN for deep face representation with noisy labels (2016). arXiv preprint arXiv:1511.02683v2
15. Goodfellow, I.J., Warde-Farley, D., Mirza, M., Courville, A., Bengio, Y.: Maxout networks (2013). arXiv preprint arXiv:1302.4389
16. Lin, M., Chen, Q., Yan, S.: Network in network (2013). arXiv preprint arXiv:1312.4400
17. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 539–546 (2005)
18. Bloch, I.: Information combination operators for data fusion: a comparative review with classification. IEEE Trans. Syst. Man Cybern.-Part A: Syst. Hum. **26**(1), 52–67 (1996)
19. Zhang, J., Huang, D., Wang, Y., Sun, J.: Lock3DFace: a large-scale database of low-cost Kinect 3D faces. In: International Conference on Biometrics (ICB), pp. 1–8 (2016)
20. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Sig. Process. Lett. **23**(10), 1499–1503 (2016)
21. Nguyen, H., Bai, L.: Cosine similarity metric learning for face verification. In: Asian Conference on Computer Vision, pp. 709–720 (2010)
22. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
23. 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)