Virtual samples construction using image-block-stretching for face recognition

Yingnan Zhao[1,2,3], Xiangjian He[3], Beijing Chen[1, 2]

[1] Nanjing University of Information Science & Technology, Jiangsu Engineering Center of Network Monitoring, 210024 Nanjing, China
[2] Nanjing University of Information Science & Technology, School of Computer & Software, 210044 Nanjing, China
[3] University of Technology, Sydney, School of Computing and Communications, Sydney, Australia

ann_zhao_99@163.com, xiangjian.he@uts.edu.au, nbutimage@126.com

Abstract. Face recognition encounters the problem that multiple samples of the same object may be very different owing to the deformation of appearances. To synthesizing reasonable virtual samples is a good way to solve it. In this paper, we introduce the idea of image-block-stretching to generate virtual images for deformable faces. It allows the neighbored image blocks to be stretching randomly to reflect possible variations of the appearance of faces. We demonstrate that virtual images obtained using image-block-stretching and original images are complementary in representing faces. Extensive classification experiments on face databases show that the proposed virtual image scheme is very competent and can be combined with a number of classifiers, such as the sparse representation classification, to achieve surprising accuracy improvement.

Keywords: Face recognition, virtual image, sparse representation

1 Introduction

As one of the most active branches of biometrics, face recognition is attracting more and more attention [1, 2]. However, it is still faced with a number of challenges, such as varying illumination, facial expression and poses [3-6]. These appearance deformations lead to the fact that the same position of different images of the same object may have varying pixel intensities. It thereby increases the difficulty to recognize the object accurately [7]. Since more training samples are able to reveal more possible variation of the deformations, they are consequently beneficial for correct face classification. However, in real-world applications, there are usually only a limited number of available training samples, due to limited storage space or capturing training samples in a short time. In fact, non-sufficient training samples have become one bottleneck of face recognition [8, 9].

On feasible way is to construct virtual samples to obtain better face recognition result. The literatures have proposed several schemes to synthesize new samples. From the viewpoint of applications, they can be categorized into two kinds, i.e. 2D virtual face images and 3D ones. As for 2D virtual face images, Tang et. al. [10] used prototype faces and an optic flow and expression ratio image based method to generate virtual facial expression. Thian et. al. [11] used simple geometric transformations to generate virtual samples. Ryu et. al. [12] exploited the distribution of the given training set to generate virtual samples. Beymer et. al. [13] and Vetter et. al. [14] synthesized new face samples with virtual views. Jung et. al. [15] exploited the noise to synthesize new face samples. Sharma et. al. [16] synthesize multiple virtual views of a person under different poses and illumination from a single face image and exploited extended training samples to classify the face. In order to overcome the small sample size problem of face recognition, Liu et. al. [17] represented each single image as a subspace spanned by its synthesized (shifted) samples. Xu et. al. [18-20] proposed a series of virtual sample construction methods based on symmetry, mirror samples and average samples. These virtual samples are easy to obtain, and the relative algorithms do yield attractive accuracy improvement in face recognition. On the other hand, 3D face synthesis also plays an important role in computer vision and virtual reality [21]. These studies offer good solutions to produce more available samples by exploiting the characteristics of face images.

In this paper, we propose to exploit the image-block-stretching method to generate new training samples and be combined with a sparse representation based method to perform face recognition. The new training samples indeed reflect some possible appearance of the face. The sparse representation based method simultaneously uses the original and new training samples to perform a face classification. This method also takes advantages of the score level fusion, which has proven to be very competent and is usually better than the decision level and feature level fusion.

Classification is a basic task in the fields of pattern recognition and computer vision. Various classification algorithms have been proposed with different viewpoints. For face recognition, sparse representation classification (SRC) algorithm is one of the best algorithms and has obtained very good performance [22]. However, Zhang et al. pointed out that it was collaborative representation (CR) rather than l_1 norm sparsity that contributes to the final classification accuracy [23]. Based on the non-sparse l_2 norm, Collaborate representation classification (CRC) [23] could lead to similar recognition but a significantly highly computational speed. It is noted that all the feature elements both in SRC and CRC share the same coding vector over their associated sub-dictionaries. This requirement ignores the fact that the feature elements in a pattern not only share similarities but also have differences. Therefore, Yang et al. presented a relaxed collaborative representation (RCR) [24] model to effectively exploit the similarity and distinctiveness of features. Liner Regression Classification (LRC) [25] can be referred as l_2 norm based on the linear regression model. Representation algorithms with the l_2 minimization have analytic solutions whereas conventional SRC algorithms should use iterative procedures to determine the solutions. Moreover, the formers always have lower computational costs than the conventional SRC algorithm [26].

With this paper, we present a novel scheme to generate virtual images for deformable faces. Our scheme first divides an original image into non-overlapping blocks. Then for two neighbored blocks in a row, we stretch them randomly. After all blocks are dealt with, we treat the new image as a virtual sample. The rationale of image-block-stretching in our scheme is based on the following fact. The deformation of the appearance of faces makes pixel intensities within the same block changeable. For example, different facial expressions of a person will lead to varying image block sat the same position of different face images. A slight smile will cause the pixel shift of some blocks in comparison with the original image. The experiments show that the proposed virtual images provide important information of deformable faces and can win very satisfactory accuracy for face recognition. This paper has the following contributes. 1) It proposes a kind of competent virtual images for deformable faces, which is very beneficial to reduce the data uncertainty. 2) The proposed scheme is simple, easy to understand and implement.

The rest of this paper is organized as follows. Section 2 first presents the proposed novel scheme to generate virtual images and interprets the reasonability, then describes the algorithm to integrate original images and virtual images. Section 3 performs experiments and Section 4 concludes the paper.

2 The proposed scheme and algorithm

2.1 Scheme to produce virtual samples using image-block-stretching

Suppose the size of each original training sample is $a \times b$. The proposed scheme to produce virtual samples includes the following steps.

1. Each original training sample is divided into non-overlapping image blocks, each block having the size of a' by b' pixels.
2. For all blocks within each training sample, an array $t[i, j]$ is generated, where $i = 1,2,3,\cdots m$, $j = 1,2,3,\cdots n$. Here, $m = [a / a']$ and $n = [b / b']$.
3. For each row of the array t, from the left to the right, the two neighbored blocks, denoted as $t[i, j]$ and $t[i, j+1]$, are stretched randomly. Suppose the stretching scale is s. Then the two stretching results is $a' \times (b' + s)$, $a' \times (b' - s)$ or $a' \times (b' - s)$, $a' \times (b' + s)$. Note that the interpolation method is the nearest neighbor algorithm.
4. When all the blocks are dealt with, we get a virtual sample. Every virtual image is converted into a vector.

Fig.1 shows some examples of original images and the corresponding virtual images. We see that the virtual image is still a natural image. Because it is a modification of the corresponding original image, it has obvious deviation from the original image and represents some possible variations of the object. As a result, the simultaneous use of the original images and virtual images allows more information of the object to be provided, and better recognition accuracy can be achieved. An intuitive explanation of the benefit of more available training samples is as follows. Because the object is deformable and the number of samples in the form of images is always smaller than the dimension of samples, the difference between images of the same object is usually great. Consequently, the test sample may be very different from the limited training samples of the same object. However, when virtual images are also used as training samples, the possibility that the training sample with the minimum distance to the test sample is from the same object as the test sample will be increased. As we know, if the training sample with the minimum distance to the test sample is from the same object as the test sample, we can obtain the correct recognition result of the test sample via the nearest neighbor classifier. Thus, more accurate recognition of objects can be obtained under the condition that the original images and virtual images are simultaneously used as training samples and the nearest neighbor classifier is exploited. We also say that the simultaneous use of the original images and virtual images enables the data uncertainty to be reduced.

Fig. 1. Examples of original images (the top row) and the corresponding virtual images (the bottom row).

2.2 Algorithm to integrate original images and virtual images

First of all, it should be pointed out that our algorithm exploits three kinds of virtual training samples. The first kind of virtual training samples is the mirror face image presented in [19].The second kind of virtual training samples is obtained by applying our image-block-stretching scheme to the original face image. The third kind of virtual training samples is obtained by applying our image-block-stretching scheme to the mirror face image.

For a test sample, an arbitrary representation based classification algorithm is respectively applied to the original training samples and virtual training samples. Suppose that there are C classes. Let $d_1,\cdots,d_C$ denote the class residuals of the test sample with respect to the original training samples of the first to the C-th classes. Let $e_1,\cdots,e_C$ denote the class residuals of the test sample with respect to the first kind of virtual training samples of the first to the C-th classes. Let $f_1,\cdots,f_C$ and $g_1,\cdots,g_C$ denote the class residuals of the test sample with respect to the second and third kinds of virtual training samples.

The class residuals of the test sample are integrated using

$$q_j = t_1 d_j + t_2 e_j + t_3 f_j + t_4 g_j,\ j = 1, \cdots, C. \quad (1)$$

Let $r = \arg\min_j q_j$. We assign the test sample to the r-th class.

As for the weight, we use an automatic procedure to determine it. The automatic procedure is as follows. Let $d_1, \cdots, d_C$ stand for the sorted results of $d_1, \cdots, d_C$ and suppose that $d_1 \le \cdots \le d_C$. Accordingly, $e_1, \cdots, e_C$, $f_1, \cdots, f_C$, and $g_1, \cdots, g_C$ have the similar meanings. We define $t_{10} = d_2 - d_1$ and $t_{20} = e_2 - e_1$. t_{30} and t_{40} are defined as $t_{30} = f_2 - f_1$ and $t_{40} = g_2 - g_1$. We respectively define

$$t_1 = \frac{t_{10}}{t_{10} + t_{20} + t_{30} + t_{40}}, \quad (1)$$

$$t_2 = \frac{t_{20}}{t_{10} + t_{20} + t_{30} + t_{40}}, \quad (2)$$

$$t_3 = \frac{t_{30}}{t_{10} + t_{20} + t_{30} + t_{40}}, \quad (3)$$

$$t_4 = \frac{t_{40}}{t_{10} + t_{20} + t_{30} + t_{40}}. \quad (4)$$

We would like to point out that a similar weight setting algorithm was proposed in [27], which is the first completely adaptive weighted fusion algorithm and obtained an almost perfect fusion result. However, the algorithm in [27] can fuse only two kinds of scores.

3 Experimental results

In experiments, SRC and CRC are respectively used as the classifier. Because our scheme stretches the neighbored two blocks of an original image randomly, so we run it five times and take the average of the rates of classification errors as the final result. In the following tables CRC and SRC denote the algorithms of naïve CRC presented in [23] and SRC presented in [22], respectively. In the context “Our method integrated with CRC” is used to present the result of applying our method to improve

CRC. “Our method integrated with SRC” is used to present the result of applying our method to improve SRC. Note that the stretching scale s is 1for all the experiments.

3.1 Experiment on the FERET database

We first use a subset of the FERET face dataset to perform an experiment. The subset contains 1400 face images from 200 subjects and each subject has seven different face images. This subset was composed of images in the original FERET face dataset whose names are marked with two-character strings: ‘ba’, ‘bj’, ‘bk’, ‘be’, ‘bf’, ‘bd’, and ‘bg’. Every face image was resized to a 40 by 40 image. We respectively select the first 2,3and 4 face images of each subject as original training samples and use the remaining face images as testing samples. Fig. 2 presents some cropped face images from the FERET face database. The size of the image block in our method is set to $a = 2$ and $b = 2$. Table 1 shows the rates of classification errors (%) of different methods on the FERET database. We see that our method outperforms naïve CRC and SRC.

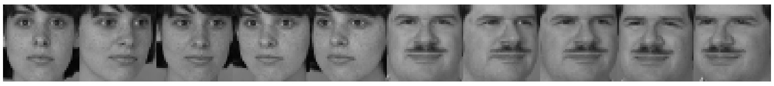

Fig. 2. Some face images from the FERET face database.

Table 1. Experimental results on the FERET database.

Number of training samples per person	2	3	4
CRC	41.60	55.63	44.67
Our method +CRC	39.48	46.58	43.28
SRC	41.90	49.25	36.67
Our method +SRC	36.15	37.39	31.83

3.2 Experiment on the GT database

In the GT face database there are 750 face images from 50 persons and every person provided 15 face images. The original images are color images with clutter background taken at the resolution of 640×480 pixels and show frontal and tilted faces with different facial expressions, lighting conditions and scales. We used the manually cropped and labeled face images to conduct the experiment. Fig. 3 shows some of these images. We further resized them to obtain images with the resolution of 40×30 pixels. They were then converted into gray images. In our experiment the first 3-10 face images of every person are used as training samples and the other images are employed as test samples. The size of the image block in our method is set to $a = 2$ and $b = 3$. Table 2 shows the rates of classification errors (%) of different methods on the GT database. We see that our method also performs better than original CRC and SRC.

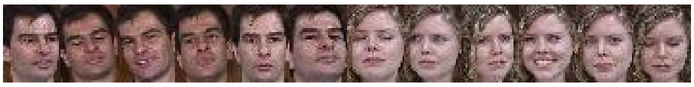

Fig. 3. Some face images from the GT face database.

Table 2. Experimental results on the GT database.

Number of training samples per person	3	4	5	6	7	8	9	10
CRC	54.67	52.91	51.20	44.44	41.50	40.57	39.00	36.00
Our method +CRC	51.63	48.52	46.38	42.04	36.29	35.83	32.40	29.78
SRC	55.33	53.82	51.20	44.44	41.75	41.43	37.33	34.40
Our method +SRC	54.18	52.26	49.37	41.60	38.44	37.62	34.50	33.48

3.3 Experiments on the ORL database

The ORL face database includes 400 face images taken from 40 subjects each having 10 face images. In this database some images were taken at different times and with various lightings, facial expressions (open/closed eyes, smiling/not smiling), and facial details (glasses/no glasses). All images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). We resized every image to form a 56 by 46 image matrix. Fig. 4 shows some images from this ORL dataset. The first 1, 2, 3, 4, 5 and 6 face images of each subject were used as training samples and the others were exploited as test samples, respectively. The size of the image block in our method is set to $a = 4$ and $b = 2$. Table 3 shows rates of classification errors (%). It again tells us that our method is better than naïve CRC and SRC.

Fig.4. Some images from this ORL dataset.

Table 3. Experimental results on the ORL database.

Number of training samples per person	1	2	3	4	5	6
CRC	31.94	16.56	13.93	10.83	11.50	8.13
Our method + CRC	31.49	16.46	13.40	11.29	11.41	8.81
SRC	32.78	17.81	16.07	12.50	12.00	10.00
Our method + SRC	30.18	16.62	15.38	11.27	10.82	9.63

4 Conclusions

In this paper, the proposed novel scheme not only can efficiently generate natural virtual images for deformable faces but also can lead to notable accuracy improvement. Moreover, the rationale of the proposed scheme is easy to understand. As shown earlier, image-block-stretching of an original image can reflect possible change of the appearance of a deformable face, so the obtained virtual image is a proper representation of the object. Besides the proposed scheme is applicable for faces, it can also provide useful representation for a general object because of the following factors. Within a small enough region of a general image, in most cases the difference of the values of pixels is usually little. As a result, image-block-stretching of a small region of an image will obtain reasonable and new representation of this region. It should be pointed out that a non-deformable object also usually has varying images owing to changeable illuminations, imaging distances and views. Thus, after we use the proposed scheme to produce virtual images for general objects, we can also integrate the virtual images and original images to obtain better classification performance.

Acknowledgments. This work is supported in part by the PAPD of Jiangsu Higher Education Institutions, Natural Science Foundation of China (No. 61572258, No.

61103141 and No. 51505234), and the Natural Science Foundation of Jiangsu Province (No. BK20151530).

References

1. L. Zhang, S. Chen, L. Qiao: Graph optimization for dimensionality reduction with sparsity constraints. Pattern Recognition, vol. 45, no. 3, pp. 1205-1210 (2012)
2. Z. Fan, Y. Xu, D. Zhang: Local linear discriminant analysis framework using sample neighbors. IEEE Transactions on Neural Networks, vol. 22, no. 7, pp. 1119-1132 (2011)
3. S. J. Wang, J. Yang, M. -F. Sun, et al.: Sparse tensor discriminant color space for face verification. IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 6, pp. 876-888 (2012)
4. D. Zhang, F. Song, Y. Xu, et al.: Advanced pattern recognition technologies with applications to biometrics. New York, Medical Information Science Reference (2009)
5. X. Zhang, Y. Gao：Face recognition across pose: A review. Pattern Recognition, vol. 42, no. 11, pp. 2876-2896 (2009)
6. S. N. Kautkar, G. A. Atkinson, M. L. Smith: Face recognition in 2D and 2.5D using ridgelets and photometric stereo. Pattern recognition, vol. 45, no. 9, pp. 3317-3327 (2012)
7. Ali Shahrokni, François Fleuret, Pascal Fua: Classifier-based Contour Tracking for Rigid and Deformable Objects, British Machine Vision Conference, No. CVLAB-CONF-2005-014 (2005)
8. P. Zhang, X. You, W. Ou, et al.: Sparse discriminative multi-manifold embedding for one-sample face identification. Pattern Recognition, vol. 52, no., pp. 249-259 (2016)
9. Z. L. Sun, L. Shang: A local spectral feature based face recognition approach for the one-sample-per-person problem. Neurocomputing, (2016)
10. B. Tang, S. Luo, H. Huang: High performance face recognition system by creating virtual sample. Proceedings of the International Conference on Neural Networks and Signal Processing, pp. 972-975 (2003)
11. N. P. H. Thian, S. Marcel, S. Bengio: Improving face authentication using virtual samples. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 6-10 (2003)
12. Y.-S. Ryu, S.-Y. Oh: Simple hybrid classifier for face recognition with adaptively generated virtual data. Pattern Recognition Letters, vol. 23, no. 7, pp. 833-841 (2002)
13. D. Beymer, T. Poggio: Face recognition from one example view. Proceedings of the Fifth International Conference on Computer Vision, pp. 500-507 (1995)
14. T. Vetter: Synthesis of novel views from a single face image. International Journal of Computer Vision, vol. 28, no. 2, pp. 102-116 (1998)
15. H. –C. Jung: Authenticating corrupted face image based on noise model. Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 272-277 (2004)
16. A. Sharma, A. Dubey, P. Tripathi, et al.: Pose invariant virtual classifiers from single training image using novel hybrid-eigenfaces. Neurocomputing, vol. 73, no. 10-12, pp. 1868-1880 (2010)
17. J. Liu, S. Chen, Z.-H, Zhou, et al.: Single image subspace for face recognition. Proceedings of the 3rd International Conference on Analysis and Modeling of Faces and Gestures, pp. 205-219 (2007)

18. Y. Xu, X. Zhu, Z. Li, et al.: Using the original and 'symmetrical face' training samples to perform representation based two-step face recognition. Pattern recognition, vol.46, no. 4, pp. 1151-1158 (2013)
19. Y. Xu, X. Li, J. Yang: Integrate the original face image and its mirror image for face recognition. Neurocomputing, vol. 131, pp. 191-199 (2014)
20. Y. Xu, X. Fagn, X. Li, et al.: Data uncertainty in face recognition. IEEE Transactions on Cybernetics. vol. 44, no. 10, pp. 1950-1961 (2014)
21. H. T. Nguyen, E. P. Ong, A. Niswar, et al.: Automatic and real-time 3D face synthesis, VRCAI, pp. 103-106 (2009)
22. Wright, John, Allen Y. Yang, Arvind Ganesh, Shankar S. Sastry, and Yi Ma: Robust face recognition via sparse representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 31, no. 2, pp. 210-227 (2009)
23. D. Zhang, M. Yang, X.-Ch Feng: Sparse representation or collaborative representation: Which helps face recognition? In Computer Vision (ICCV), 2011 IEEE International Conference on, pp. 471-478 (2011)
24. Y. Meng, D. Zhang, S. Wang: Relaxed collaborative representation for pattern classification. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 2224-2231 (2012)
25. N. Imran, R. Togneri, M. Bennamoun: Linear regression for face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on vol. 32, no. 11, pp. 2106-2112 (2010)
26. Z. Zhang, Y. Xu, J. Yang, et al.: A Survey of Sparse Representation: Algorithms and Applications. IEEE Access, 3, pp. 490-530 (2015)
27. Y. Xu, Y. Lu: Adaptive weighted fusion: A novel fusion approach for image classification. Neurocomputing, vol. 168, pp. 566-574 (2015)