

# On Averaging Face Images for Recognition under Pose Variations

Xiaozheng Zhang, Sanqiang Zhao, Yongsheng Gao  
Computer Vision and Image Processing Lab,  
Institute for Integrated and Intelligent Systems, Griffith University, Australia  
x.zhang, s.zhao, yongsheng.gao@griffith.edu.au

## Abstract

*Recently, psychological studies showed that averaging human face images greatly improves the performance of face recognition under various pose, illumination, expression, and/or aging conditions. This paper investigates quantitatively the mechanism of the face averaging process in face recognition specifically against pose variations. Facilitated with 3D face dataset, the process of face averaging is tested on face images free from human errors and misalignments. Single images are chosen as gallery and the averaged views as probe in all the experiments. Three different scenarios are experimented, i.e., identification using single gallery images, identification using different gallery images, and averaging using unbalanced range of input image. The experimental results show that the averaging process under pose variations is equivalent to generating a face view in an average pose and the improvement in face recognition is subject to the conditions that the gallery pose is close to the average probe pose.*

## 1. Introduction

Recent surveys of face recognition techniques [5, 11] and vendor tests [8] have revealed that current levels of face recognition is not sufficient for practical use. The major issue lies in the difficulties for face recognition techniques to tolerate image variations brought by pose, illumination, expression, and/or aging variations while distinguishing different identities through face images. These above variations in face recognition have attracted many research efforts and a number of methods have been proposed to tolerate or compensate these variations. For instance, various techniques of virtual view synthesis have been proposed to compensate pose and/or illumination variations, such as [1, 3, 6, 9, 10]. These algorithms are effective to improve the performance of face recognition in the

presence of pose and/or illumination changes. However, they are usually involved in complicated processing for virtual view synthesis and none of them achieved satisfactory accuracy for practical use.

In a recent study of face recognition [7], Jenkins and Burton have proposed to use averaged face images as probe in face recognition based on psychological studies [4] and achieved very promising accuracy. This research reveals that the simple process of averaging face images can dramatically increase the capability of computerized face recognition. In their experiment, a database containing an average of 9 photos for each of 3,628 faces was selected as the gallery and a dataset containing 20 photos for each of 25 faces was selected as the probe. Both of the gallery and probe datasets are images covering a natural range of variations in pose, expression, age and hairstyles and the pose of each image is limited within 30 degrees to make all of the marked facial features visible in the image. The averaged image of the 20 face images per person was used in recognition to match against all the gallery images. In various applications of face recognition, however, the above constraints are often violated, such as single gallery images, single probe images, larger pose variations, and/or unnatural range of image variations. Is the process of averaging still effective in these “extreme” conditions?

This paper intends to investigate the effectiveness of face recognition using averaged images in different conditions and to reveal the mechanism of averaging face images in face recognition. Among various image variations, pose variation is selected as the testing parameter as it is a prominent issue in face recognition [11]. Other variations such as illumination and expression changes can also be tested in face recognition experiments in a similar way. In a dataset of face images under varying poses only, different gallery sets and probe sets are selected and face recognition experiments are performed with or without averaging process to compare the performances of face

recognition in different conditions. This research for the first time tests the universality of face recognition using averaging in different conditions and the results of this research are indicative both to further improve the state-of-the-art face recognition technology and to better understand the mechanism of face recognition through image averaging in face recognition.

## 2. Face averaging and recognition experiment protocol

Because this paper's intention is to investigate the mechanism of face recognition through averaging to pose variations, a 3D face database is selected and the face recognition experiments are tested on synthetic images generated from the 3D face models. In the original process of averaging faces, a number of markers are labeled on each probe image, which may introduce errors unrelated to face averaging. In the proposed experiment protocol, on the other hand, the markers are labeled on the 3D models and are then accurately projected onto 2D face images to exclude these labeling errors. The selection of 3D models in the experiment also excludes possible illumination, expression, and/or aging variations associated with different probe images and ensures that the experimental results specifically reveal the performances of face recognition algorithms to pose variations only. The details of face averaging process on the testing datasets are elaborated as follows.

All of the face recognition experiments are conducted on the USF Human ID 3-D database [2] which contains 136 different 3D face color scans. These 3D face models are cropped to exclude hair and shoulders as shown in Figure 1. 50 face models are randomly selected as testing data in the proposed testing protocols. 34 markers are manually labeled onto each of the 50 probe face models. Both markers and face models are rotated in 3D space to generate views under different viewpoints. In the experiments, the rotation is limited to horizontal in-depth rotation only. Each of the synthetic images with markers' locations is then warped to a fixed template. Both the warped texture image and the original markers' positions of different images of the same person are averaged to form an average texture image and an average markers' positions (template). Then the average texture is warped backward to the average template to synthesize an average image of this person. The warping process used in this experiment is the triangular affine transformation using bi-cubic interpolation. The process of view synthesis and face averaging is illustrated in Figure 2.



Figure 1. 6 examples of 3D face models of USA Human ID 3-D database.

In face recognition experiments, we follow the testing processes of [7], but using only a single gallery image per person for 50 persons. First 50x20 single probe images are matched against gallery images and the performance is recorded as  $r_s$ . Then probe images are averaged in the above process and the 50 average probe images are fed in. The identification rate is denoted as  $r_a$ . The following scenarios are tested in the experiments.

### Scenario 1 – Balanced pose distribution in probe:

This scenario is similar to that tested in [7], where rotation is limited to 30 degrees and probe images are evenly distributed in pose variations. The difference is that single frontal gallery image per person is used in this scenario instead of multiple (7-28) gallery images in [7]. The primary goal of this setting is to reflect the mechanism of averaging faces in face recognition more clearly and to exclude the effect brought by exhaustive gallery searching. 20 input images are randomly selected with uniform distribution with the range of  $\pm 30^\circ$  as shown in Figure 3(a).

**Scenario 2 - Unbalanced pose distributions in probe:** In test 2-4, we put more weight on left rotated probe images than right rotated probe images by using three non-uniform distributions shown in Figure 3(b-d). 20 rotation angles are randomly generated following the specific distributions to synthesise 20 probe images. These probe images contain more left rotated images than right rotated (test 2), only left rotated images ranging 0 to 30 degrees (test 3), and only left rotated images with more images under large rotations than small rotations (test 4).

**Scenario 3 - Rotated gallery images:** The gallery images are then rotated from 0 degree to  $30^\circ$ . Both raw probe images and average images are matched to these gallery images. This scenario is to reveal the dependency of the face recognition using averaging to the conditions of gallery images.

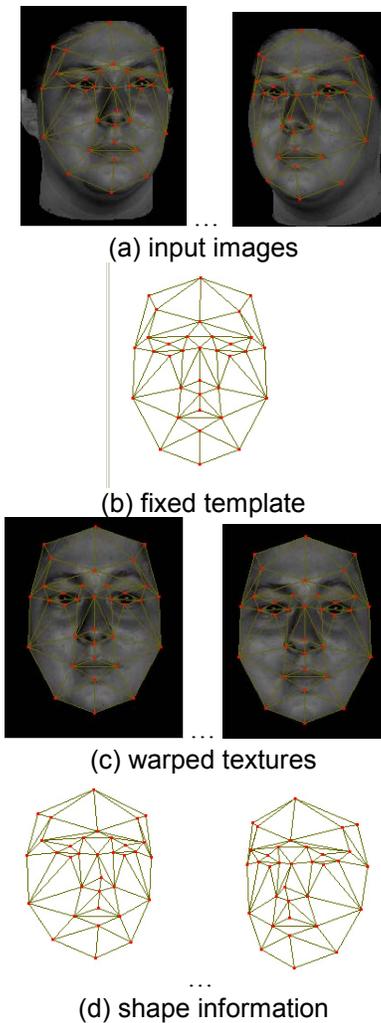


Figure 2-1. The process of generating average image over multiple input images.

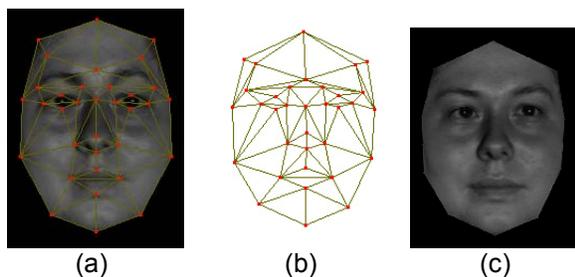


Figure 2-2. The process of generating average image over multiple input images. The averaged warped texture (a) is warped back to the averaged shape grids (b) to form the average face image (c).

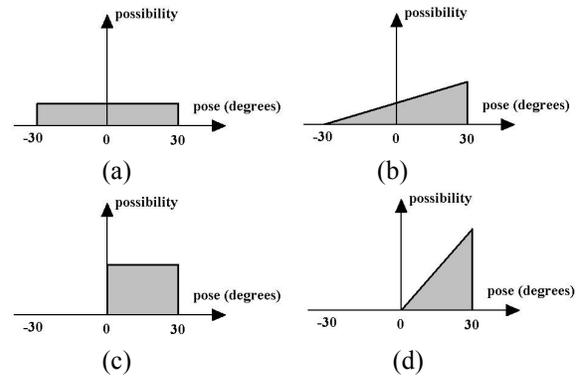


Figure 3. Possibility distributions of input image selection for the first two scenarios.

### 3. Face recognition results

The most widely used Principal Component Analysis (PCA, also known as Eigenfaces) has been selected as the classifier in the face identification experiments. For the original scenario, the identification rates (rank 1) of raw probe images and averaged probe images are listed as in Table 1. For the scenarios of unbalanced input images, the face recognition performances are tabulated in Table 2. For the scenario of rotated gallery images, the performances are shown in Figure 4.

Table 1. Face identification performance in Scenario 1

	Raw probe images $r_s$	Averaged probe images $r_a$
Test 1	70.64%	79.40%

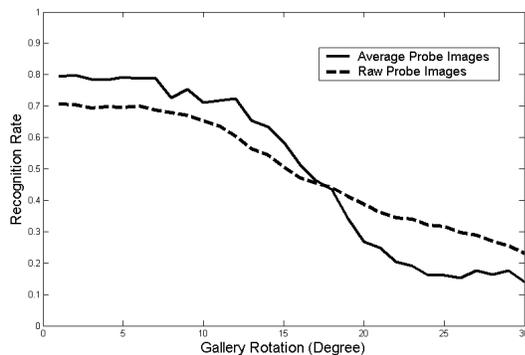
When matching against single frontal gallery images and averaging from balanced range of probe images, the average process improve the identification performance by nearly 9% in terms of pose variations only. It is different from the claim of [7] and that could be explained that the experiment settings in [7] are far more difficult than those in this paper. In [7], all of the image variations including pose, illumination, expression, aging, and etc. are considered, while in this paper, only pose variation is handled. The use of multiple gallery images could also be helpful for averaging process to reach 100% accuracy, while in this paper, only single gallery images are used to test the algorithm more rigorously.

When averaging from unbalanced range of probe images, the performance of averaging decreases as the unbalance condition getting severer. If the probe images are all rotated to one side only (e.g., to the left), the averaging process achieved lower performance than the recognition using raw probe images. This

finding reveals that under pose variations, the averaging process is in principle to find the mean of the variations and the view of the averaged pose is expected to be the pose in the average image. As results, the performance is optimal when the gallery pose and the averaged pose are identical.

**Table 2. Face identification performance in Scenario 2**

	Raw probe images $r_s$	Averaged probe images $r_a$
Test 2	74.05%	77.20%
Test 3	70.20%	65.00%
Test 4	60.36%	45.80%



**Figure 4. Face identification performance in Scenario 3**

Figure 4 demonstrates the above findings in a different way while altering gallery poses instead of altering the distributions of the probe poses. It shows that the improvements achieved by averaging process decreases as the gallery image is rotated. When the gallery image pose is larger than  $17^\circ$  rotation, the ability of averaging process to improve performance stops and it begins to deteriorate the separability of face images.

#### 4. Conclusion

In this paper, we tested face recognition using averaging in three different scenarios under pose variations, i.e., single gallery scenario, unbalanced probe pose distributions, and non-frontal gallery pose scenarios. Experimental results show that the averaging process is to generate face views in an averaged pose. The improvement achieved by averaging process is largely due to the conditions that the averaged probe pose is identical to the gallery pose. When this prerequisite is violated, the performance of face recognition using averaging decreases and

becomes even worse than the recognition using raw probe images. The future work is to investigate the mechanism of face recognition using averaging in terms of other image variations such as illumination, aging, and/or expression.

#### Acknowledgement

The research described in this paper was supported by Australian Research Council (ARC) Discovery Grant DP0451091.

#### References

- [1] D. Beymer and T. Poggio, "Face recognition from one example view," *ICCV*, pp. 500-507, Cambridge, MA, USA, 1995.
- [2] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," *SIGGRAPH*, pp. 187-194, New York, NY, USA, 1999.
- [3] V. Blanz, P. Grother, P. J. Phillips, and T. Vetter, "Face recognition based on frontal views generated from non-frontal images," *CVPR*, 2005.
- [4] A. M. Burton, R. Jenkins, P. J. B. Hancock, and D. White, "Robust representations for face recognition: The power of averages," *Cognitive Psychology*, vol. 51, pp. 256-284, 2005.
- [5] S. Chen, X. Tan, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *PR*, vol. 39, no. 9, pp. 1725-1745, 2006.
- [6] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *PAMI*, vol. 23, no. 6, pp. 643-660, 2001.
- [7] R. Jenkins and A. M. Burton, "100% Accuracy in Automatic Face Recognition," *Science*, vol. 319, no. 5862, pp. 435, 2008.
- [8] P. J. Phillips, P. Grother, R. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, "Face recognition vendor test 2002," *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 44, Nice, France, 2003.
- [9] X. Zhang, Y. Gao, and M. K. H. Leung, "Automatic texture synthesis for face recognition from single views," *ICPR*, vol. 3, pp. 1151-1154, 2006.
- [10] X. Zhang, Y. Gao, and B.-L. Zhang, "Recognising rotated faces from two orthogonal views in mugshot databases," *ICPR*, vol. 1, pp. 195-198, 2006.
- [11] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-459, 2003.