Frontal Face Generation from Multiple Low-Resolution Non-frontal Faces for Face Recognition

Yuki Kono[1], Tomokazu Takahashi[1,2], Daisuke Deguchi[1], Ichiro Ide[1], and Hiroshi Murase[1]

[1] Graduate School of Information Science, Nagoya Univesity, Furo-cho, Chikusa-ku, Nagoya, 464-0856, Japan
[2] Faculty of Economics and Information, Gifu Shotoku Gakuen University, Japan

Abstract. We propose a method of frontal face generation from multiple low-resolution non-frontal faces for face recognition. The proposed method achieves an image-based face pose transformation by using the information obtained from multiple input face images without considering three-dimensional face structure. To achieve this, we employ a patch-wise image transformation strategy that calculates small image patches in the output frontal face from patches in the multiple input non-frontal faces by using a face image dataset. The dataset contains faces of a large number of individuals other than the input one. Using frontal face images actually transformed from low-resolution non-frontal face images, two kinds of experiments were conducted. The experimental results demonstrates that increasing the number of input images improves the RMSEs and the recognition rates for low-resolution face images.

1 Introduction

Quality of face images captured by surveillance cameras tends to be poor. Because of restrictions on the installation position and the number of cameras, most of them would be in low-resolution and would not capture a face from desirable angles for person identification. These conditions make it difficult for both humans and computers to identify a person from the obtained face images. Aiming to overcome the difficulty due to low resolution, Baker and Kanade [1] have proposed a "Face Hallucination" method that obtains a high-resolution face image from a low-resolution face image. Since then, many studies related to super-resolution of face images have been reported [2,3,4,5].

Even if we could obtain high-resolution face images from the cameras, another problem still remains that poses of a face in the images are not necessarily desirable. Methods which utilize a 3D face model directly or indirectly to transform face poses have been reported [6,7]. However, it is difficult to apply them to low-resolution face images because they require a large number of accurate point correspondences between a face image and a face model to fit the image to the model. Approaches based on image-based pose transformation have also been reported. A view-transition model (VTM) proposed by Utsumi and Tetsutani [8]

R. Koch et al. (Eds.): ACCV 2010 Workshops, Part I, LNCS 6468, pp. 175–183, 2011.

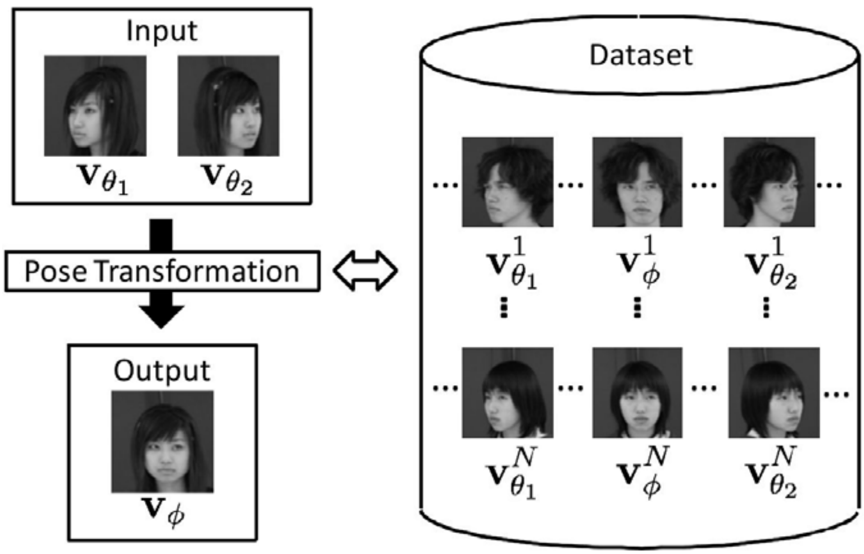

Fig. 1. Frontal face generation from two non-frontal faces

transforms views of an object between different postures by linear transformation of pixel values in images. For each pair of postures, a transformation matrix is calculated from image pairs of the postures of a large number of objects. Another work also took a similar approach [2]. Chai et al. [9] have proposed a Locally Linear Regression (LLR) method for pose-invariant face recognition, which generates a virtual frontal view from a single relatively high-resolution non-frontal face image by applying a patch-wise image transformation method.

Aiming to improve the accuracy of pose transformation and face recognition for low-resolution face images by using the information obtained from multiple input images, we propose a method for frontal face generation from multiple low-resolution non-frontal faces. The proposed method transforms multiple non-frontal input face images to a frontal face as shown in Figure 1. To achieve this, the proposed method uses a general image dataset consisting of faces of a large number of individuals viewed from various angles other than the input individual. The face pose transformation is achieved not by considering its three-dimensional structure, but by synthesizing a face image with a different pose from partial face image patches calculated from a large number of general individual's faces. This patch-wise image transformation, which we name Local VTM (LVTM), is achieved based on the VTM method.

2 Frontal Face Generation from Multiple Non-frontal Faces

The proposed method can be applied to frontal face recognition by using not only two input images but also one or any number of input images. However, in

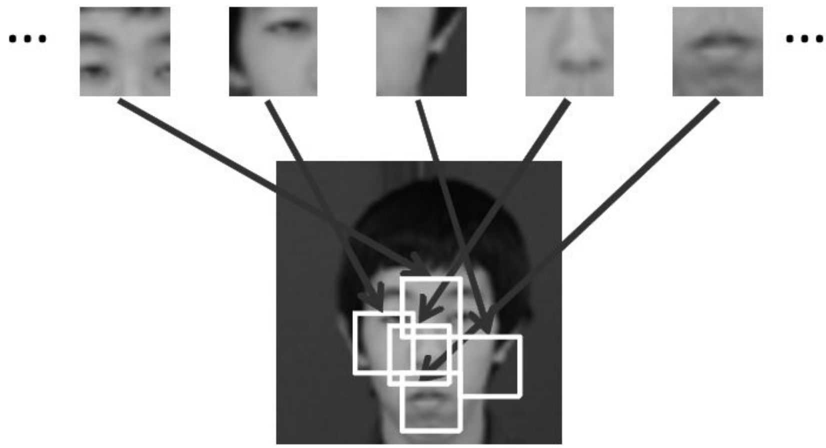

Fig. 2. Synthesis of face patches

the interest of simplicity, we describe a frontal face generation algorithm for two non-frontal face input images.

As shown in Figure 1, the proposed method transforms two input non-frontal face images $\mathbf{v}_{\theta_1}$ and $\mathbf{v}_{\theta_2}$ to an output frontal face image $\mathbf{v}_{\phi}$. Here, θ_1 and θ_2 represent poses of the input faces, and ϕ represents the frontal pose of the output face. To achieve this, the proposed method uses a dataset that contains $\mathbf{v}^n_{\theta_1}$, $\mathbf{v}^n_{\theta_2}$ and $\mathbf{v}^n_{\phi}$ $(n = 1, ..., N)$. Here, n represents an individual in the dataset.

Faces of two persons have similar parts although these faces are not totally similar. Transforming the input face image using the information of the entire face image of other individuals might degrade the characteristics of the input individual's face. Therefore, instead of directly transforming the entire face image, we transform face patches that are partial images of a face image for each location in the face image. Then, as shown in Figure 2, an output frontal face image is synthesized from the transformed patches. The proposed method is summarized as follows.

1. Establish the correspondence of each patch position between frontal and non-frontal poses.
2. Transform each pair of images in two non-frontal patches to the corresponding frontal patch using LVTM.
3. Synthesize the transformed patches to obtain the entire frontal face image as shown in Figure 2.

We describe each step in detail below.

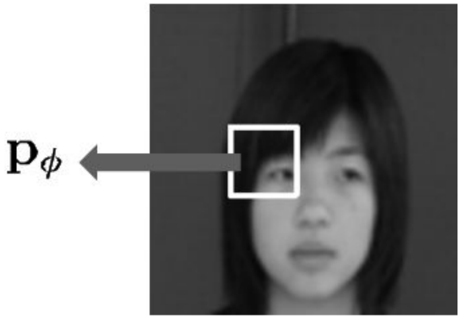

Fig. 3. Patch $\mathbf{p}_\phi$ in $\mathbf{v}_\phi$

2.1 Face Patch Correspondence

The proposed method transforms an entire face by patch-wise image transformation. Therefore we have to find the correspondence of similar facial texture patches between different poses. There are some algorithm to achieve this. For example, the LLR method [9] employs a cylinder face model for finding the correspondence between different poses. In our case, we employ a simple affine model calculated from three points that are the centers of the eyeballs and the lower tip of the nose. In other words, patch correspondences between different poses are found by transforming facial images so that the three points are located at the same positions.

2.2 Patch-Wise Face Image Transformation

A face patch is represented as $\mathbf{p}_\phi$, which is located in the output frontal image $\mathbf{v}_\phi$(Figure 3). Similarly, we represent the face patches corresponding to $\mathbf{p}_\phi$ in input non-frontal images $\mathbf{v}_{\theta_1}$ and $\mathbf{v}_{\theta_2}$ as $\mathbf{p}_{\theta_1}$ and $\mathbf{p}_{\theta_2}$, respectively. On the other hand, we represent face patches in $\mathbf{v}^n_{\theta_1}$, $\mathbf{v}^n_{\theta_2}$ and $\mathbf{v}^n_\phi$ in the dataset as $\mathbf{q}^n_{\theta_1}$, $\mathbf{q}^n_{\theta_2}$, and $\mathbf{q}^n_\phi$ $(n = 1, ..., N)$, respectively. The size of each patch is $W \times H$ pixels. We summarize below the symbols for the patches used in the algorithm.

Patch	Symbols
Input	$\mathbf{p}_{\theta_1}$, $\mathbf{p}_{\theta_2}$
Output	$\mathbf{p}_\phi$
Dataset	$\mathbf{q}^n_{\theta_1}$, $\mathbf{q}^n_{\theta_2}$, $\mathbf{q}^n_\phi$ $(n = 1, ..., N)$

This method transforms $\mathbf{p}_{\theta_1}$ and $\mathbf{p}_{\theta_2}$ to $\mathbf{p}_\phi$ using a transformation matrix $\mathbf{T}$ based on the VTM method [8]. The LVTM method proposed here, transforms each local area of an image while the VTM method transforms the entire area of an image.

The LVTM method calculates $\mathbf{p}_\phi$ as follows:

$$\mathbf{p}_\phi = \mathbf{T} \begin{bmatrix} \mathbf{p}_{\theta_1} \\ \mathbf{p}_{\theta_2} \end{bmatrix}, \tag{1}$$

where $\mathbf{p}$ also represents the vector form of the patch image, which is a column vector that has pixel values of the image as its elements. $\mathbf{T}$ is a $WH \times 2WH$ matrix which transforms $\mathbf{p}_{\theta_1}$ and $\mathbf{p}_{\theta_2}$ to $\mathbf{p}_{\phi}$. We calculate beforehand $\mathbf{T}$ using the dataset patches for each face patch position by solving the following equation:

$$\left[\, \mathbf{q}_{\phi}^{1} \cdots \mathbf{q}_{\phi}^{N} \,\right] = \mathbf{T} \begin{bmatrix} \mathbf{q}_{\theta_1}^{1} & \cdots & \mathbf{q}_{\theta_1}^{N} \\ \mathbf{q}_{\theta_2}^{1} & \cdots & \mathbf{q}_{\theta_2}^{N} \end{bmatrix}, \tag{2}$$

in the same manner as in [8].

2.3 Synthesis of Frontal Face

For each face patch position, the proposed method calculates $\mathbf{p}_{\phi}$ using the input patches and the dataset patches. After this, the proposed method synthesizes $\mathbf{v}_{\phi}$ from all $\mathbf{p}_{\phi}$. The pixel values of regions where face patches are overlapped are calculated by averaging the pixel values of the overlapped patches.

3 Experiment

To demonstrate the effectiveness of the proposed method, two experiments were conducted. In the first experiment, we transformed non-frontal face images to a frontal face image. Then we calculated the RMSE between each transformed face and the ground-truth frontal face while changing the number of input images and the input poses. In the second experiment, we input the transformed images to a system that recognizes individuals from the frontal face images. From the results, we evaluated the effectiveness of using multiple input images for frontal face generation from low-resolution non-frontal faces.

3.1 Frontal Face Generation

We used a face image dataset provided by SOFTPIA JAPAN [10], which contains face images of 300 individuals taken from horizontal angles of 0 (front), ± 10, ± 20, ± 30 and ± 40 degrees. We transformed all images by affine transformation with three points (the centers of the eyeballs and the lower tip of the nose) to find the patch correspondences between different poses. Figure 4 shows samples of images used for frontal face generation. The image size was 32×32 pixels and the face patch size was set to 16×16 pixels. Face images of 150 individuals were used for LVTM's training set and others were transformed. For the quantitative evaluation, we calculated the average RMSE to the ground-truth frontal faces in the cases of 1, 2 and 3 inputs.

3.2 Face Recognition

We input the transformed images to a recognition system that recognizes a person from a frontal face. The eigenspace method [11] was employed as the recognition strategy. We constructed an eigenspace from 150 actual frontal face images of the same individuals used for the input images. For the purpose of comparison, we calculated the average recognition rate in the cases of 1, 2 and 3 inputs.

Fig. 4. Samples of images used for frontal face generation

Table 1. Overall average of RMSEs and recognition rates

Number of inputs	1	2	3
RMSE	19.1	17.7	16.7
Recognition rate [%]	61	73	77

4 Result and Discussion

Table 1 shows the average RMSEs and recognition rates. These averages were calculated from 150 individuals and all combinations of the input degrees for the cases of 1, 2 and 3 inputs. These results show that the more input images were used, the better RMSE and recognition rate were obtained. From this, we confirmed the effectiveness of the proposed method that uses multiple input images for frontal face generation. Figure 5 shows samples of the transformed

Table 2. RMSEs and recognition rates for each set of input degrees

(a)RMSE

Input angle[°]	−40	−30	−20	−10	+10	+ 20	+30	+40
−40	21.0	19.1	17.9	17.3	17.1	17.3	17.7	18.6
−30	-	19.5	18.0	17.2	17.0	17.2	17.8	18.0
−20	-	-	18.6	17.0	16.2	16.5	17.0	17.4
−10	-	-	-	17.6	15.9	16.2	16.7	16.8
+10	-	-	-	-	18.0	17.0	17.0	17.0
+20	-	-	-	-	-	19.0	18.0	17.9
+30	-	-	-	-	-	-	19.0	18.7
+40	-	-	-	-	-	-	-	20.2

(b)Recognition rate [%]

Input angle[°]	−40	−30	−20	−10	+10	+20	+30	+40
−40	45	55	69	73	73	71	67	65
−30	-	55	67	76	77	74	70	67
−20	-	-	68	79	83	80	75	69
−10	-	-	-	72	82	80	81	77
+10	-	-	-	-	70	80	80	72
+20	-	-	-	-	-	70	70	67
+30	-	-	-	-	-	-	60	59
+40	-	-	-	-	-	-	-	46

The number of input	1	2	3	Ground truth
				
RMSE	15.1	13.6	13.5	0.0

(a)Results with input angles of +30 for 1 input, ±30 for 2 inputs, and ±30 and −20 for 3 inputs

Ground truth				
Transformed face				
RMSE	19.6	21.5	14.8	17.2

(b)Results for various persons in the 2 input case with input angles of ±30

Input degree[]	+10	+20	+30	+40	Ground truth
					
RMSE	14.8	16.3	21.5	17.5	0.0

(c)Results for various input angles in the 1 input case

Fig. 5. Examples of transformed facial images

frontal face images and their RMSEs. The samples of the results for each of the 1, 2 and 3 input cases are shown in Figure 5 (a). On the other hand, Figure 5 (b) and (c) show the samples of the results for various individuals and various input angles, respectively. Additionally, Table 2 shows the average RMSEs and face recognition rates for each set of input degrees in the cases of 1 and 2 inputs. In the tables, the values in the fields where two input angles are the same are the 1 input case. These results show that the inputs with angles close to the front tend to achieve small RMSEs and high recognition rates. Increasing the number of input images increases the information of the input individual for frontal face generation. However, some cases where adding another input image degraded the accuracy were observed. For example, the recognition rate of the input degrees 30 and 40 was 59%, while that of the input degree 30 was 60%. Such situations were also observed in the case of 3 inputs. This is because appropriate patch correspondences would be difficult to obtain by the simple method when two poses are distant. The inappropriate patch correspondences might degrade the accuracy of the frontal face generation.

5 Conclusion

We proposed a method for frontal face generation from multiple low-resolution non-frontal faces for face recognition. The proposed method achieves the image-based face pose transformation by using the information obtained from multiple input face images without considering three-dimensional face structure. Using frontal face images actually transformed from low-resolution non-frontal face images, two kinds of experiments were conducted. The experimental results demonstrated that increasing the number of input images generally improves the RMSEs and the recognition rates for low-resolution face images.

Future work includes making use of the knowledge on the movement of face parts according to face pose change in order to obtain appropriate correspondences between face patches.

Acknowledgement

This work was supported by "R&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society," Special Coordination Fund for Promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government.

References

1. Baker, S., Kanade, T.: Hallucinating faces. In: Proc. International Conference on Automatic Face and Gesture Recognition, pp. 83–88 (2000)
2. Jia, K., Gong, S.: Hallucinating multiple occluded face images of different resolutions. Pattern Recognition Letters 27, 1768–1775 (2006)

3. Dedeoğlu, G., Baker, S., Kanade, T.: Resolution-aware fitting of active appearance models to low resolution images. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 83–97. Springer, Heidelberg (2006)
4. Lan, C., Hu, R., Lu, T., Luo, D., Han, Z.: Global face super resolution and contour region constraints. In: Zhang, L., Lu, B.-L., Kwok, J. (eds.) ISNN 2010. LNCS, vol. 6064, pp. 120–127. Springer, Heidelberg (2010)
5. Huang, H., He, H., Fan, X., Zhang, J.: Super-resolution of human face image using canonical correlation analysis. IEEE Trans. Pattern Recognition 43, 2532–2543 (2010)
6. Beymer, D., Poggio, T.: Face recognition from one example view. In: Proc. 5th International Conference on Computer Vision, pp. 500–507 (1995)
7. Blanz, V.G., Phillips, P.J., Vetter, T.: Face recognition based on frontal views generated from non-frontal images. In: Proc. International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 454–461 (2005)
8. Utsumi, A., Tetsutani, N.: Adaptation of appearance model for human tracking using geometrical pixel value distribution. In: Proc. 6th Asian Conference on Computer Vision, vol. 2, pp. 794–799 (2004)
9. Chai, X., Shan, S., Chen, X., Gao, W.: Locally linear regression for pose-invariant face recognition. IEEE Trans. Image Processing 16, 1716–1725 (2007)
10. SOFTPIA JAPAN.: HOIP-F (Face Image Dataset), `http://www.softopia.or.jp/rd/facedb.html`
11. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Proc. International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 586–591 (1991)