Viewpoint-Invariant Face Recognition Based on View-Based Representation

Jinyun Chung, Juho Lee, Hyun Jin Park and Hyun Seung Yang

Korea Advanced Institute of Science and Technology, 373-1, Guseong-dong, Yuseong-gu, Daejeon, 305-701, Republic of Korea
{mage,jhlee,hjpark,hsyang}@paradise.kaist.ac.kr

Abstract. In this paper, we suggest a viewpoint-invariant face recognition model based on view-based representation. The suggested model has four stages: view-based representation, viewpoint classification, frontal face estimation and face recognition. For view-based representation, we obtained the feature space by using independent subspace analysis, the bases of which are grouped like the neurons in the brain's visual area. The viewpoint of a facial image can be easily classified by a single-layer perceptron due to view-dependent activation characteristic of the feature space. To estimate the independent subspace analysis representation of frontal face, a radial basis neural network learns to generalize the relation of the bases between two viewpoints. Face recognition relies on a normalized correlation for selecting the most similar frontal faces in a gallery. Through our face recognition experiment on XM2VTS [9], we obtained a face recognition rate of 89.33%.

1 Introduction

Face recognition has been an active research area due to its importance in surveillance systems. There have been great technological advances and nowadays some researchers try to develop commercial products with face recognition technology. In most face recognition methods, however, environmental variations such as variations in the viewpoint and illumination cause significant degradation of performance. As a result, many researchers now concentrate on developing face recognition algorithms that overcome those problems. We focus here on ways of overcoming the viewpoint variation problem.

A trivial solution to the viewpoint variation problem is to remember the facial images of every viewpoint, though facial images of every viewpoint are generally not available. We therefore need to develop a face recognition algorithm that relies on a limited number of facial images.

To develop a face recognition algorithm that achieves this purpose, other researchers have used a 3-D facial model to get view-invariant facial features or they have tried to develop a computational transformation model between the facial images of two different viewpoints. However, these approaches assume that the surface of the face is a Lambertian surface [1] or that the face is a plane [2]. To compute 3-D facial model parameters or transformation parameters, we should find the corresponding

points between two images of different viewpoints but this task is vulnerable to illumination variation.

To solve the viewpoint variation problem from another viewpoint, we examine the characteristics of the human visual system, which is robust to viewpoint variation.

Neurophysiological study of the brain reveals that neurons in the inferior temporal area respond to the facial image of a certain viewpoint [3]. Such specialization of neurons progresses as a result of unsupervised learning, and we deduce that this characteristic of neurons will be useful for classifying the viewpoint of an observed facial image and for learning the relation between the responses of neurons. Hence, we focus on this characteristic of the human brain and suggested a viewpoint-invariant face recognition model. To simulate unsupervised learning of view-specific neurons, we use independent subspace analysis (ISA). A facial image is represented in a subspace obtained through ISA and this representation shows a view-specific activation characteristic.

After using the ISA subspace to produce representations of the facial images, then their viewpoints are classified with the aid of a single-layer perceptron (SLP). According to the results of the viewpoint classification, a neural network is selected to estimate the ISA representation of a frontal face image from that viewpoint. The neural network that we use to estimate the representation of the frontal face image was a radial basis function neural network (RBF-NN), which is trained with the aid of support vector regression (SVR). The recognition is performed by comparing the normalized correlation between the estimated facial image and the gallery of facial images with a frontal viewpoint.

The proposed model needs no manually constructed model because unsupervised learning is used to obtain the view-based representation. Other face recognition methods need facial models whose feature points are selected manually [1, 2]. Furthermore, although the eigensignature method estimates the facial representation of an unfamiliar view in an eigenspace [4], we do so in view-specific subspaces. In view-based representation, the relation between bases is simpler. Moreover, it is advantageous to classify viewpoints and to regress the relation between two facial images of different viewpoints.

2 A Viewpoint-Invariant Face Recognition Model Based on View-Dependent Representation

The proposed face recognition model has four stages: view-dependent representation on an ISA subspace, viewpoint classification, estimation of frontal face representation, and face recognition with a normalized correlation. In this section, we describe each stage.

2.1 Unsupervised Learning for View-Based Representation

Neurons in the visual area of the brain are responsive to specific facial viewpoints, such as frontal or profile views. The neurons that respond to similar viewpoints are

spatially proximate to each other. The neural networks of the cerebral visual cortex tend to organize themselves in the direction of increasing independency of the neural coding [5]. We can apply this unsupervised learning principle to learn view-based representation in a manner that is similar to the brain's view-tuned neural coding scheme [6].

The ISA computational algorithm is adequate for unsupervised learning of view-based representation [7]. Through ISA, independent subspaces, groups of bases, can be obtained, although independent bases can be obtained through independent component analysis. By maximizing the following equation, we can find the linear transformation, $\mathbf{W}$, which maximizes the independency between groups of output variables:

$$\log P(\mathbf{X} \mid \mathbf{W}) = \sum_{t=1}^{T} \sum_{j=1}^{J} G(\sum_{i \in S_j} \mathbf{w}_i^T \mathbf{x}(t)) + T \log \mid \det \mathbf{W} \mid, \tag{1}$$

where $\mathbf{x}$ is a feature vector of a sensory signal, $\mathbf{w}$ is the basis of an ISA subspace, G is a non-Gaussian function, T is the number of samples, J is the number of subspaces, and S_j is a jth subspace.

A facial image can be projected to an ISA subspace and we can get its view-dependent representation. As shown in Fig. 1, when the bases are grouped according to their viewpoints, the projection norm to a certain group of bases that represent the same viewpoint of a given facial image is higher than the projection norm to other groups. This characteristic is very useful for classifying the viewpoint and for obtaining the relation of bases.
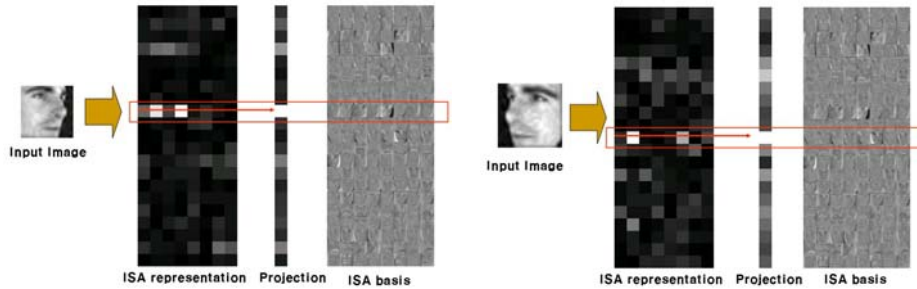

Fig. 1. View-specific activation of an ISA representation. Two facial images with a different viewpoint show high activation values in different groups of ISA bases

2.2 Viewpoint Classification from View-Based Representation

When we use ISA, the proposed view-based representation shows a view-dependent activation characteristic. Hence, we thought the viewpoint could be easily classified with the aid of an SLP. The SLP can learn which bases are related to a certain viewpoint and which bases are not. We therefore quantized the viewpoint of a horizontal rotation from -45° to +45° and a vertical rotation from -15° to +15° to 21 units at intervals of 15°. The input layer of the SLP is an ISA representation of an observed facial image and the output layer consists of 21 neurons. To determine the viewpoint

of a given image, we selected the maximum output neuron with the highest activation value.

2.3 Estimation of the Representation of a Frontal Face Image

The human brain can learn an association of viewpoints by remembering representative examples. If a person's facial image and a remembered representative face from frontal viewpoint resemble each other, his or her facial image and the remembered representative face from profile viewpoint may also resemble each other. Hence, as shown in Fig. 2, we simulated a neural association between the facial images of two other viewpoints as an RBF-NN. The hidden layer of the RBF-NN has representative examples and the output layer shows the estimated representation of the frontal face.

The RBF-NN could be effectively trained with the aid of SVR [8]. When SVR is performed with an RBF kernel, the SVR result is the same as that of an RBF-NN in which the nodes of the hidden layer are representative examples. However, because SVR has only a single output unit, we had to repeat SVR for each basis. In our experiment, there were 168 ISA bases; hence, we obtained 168 SVR results for a single RBF-NN.
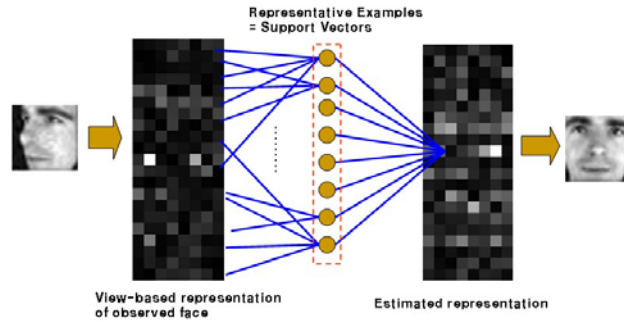

Fig. 2. Estimation of a front face representation by using an RBF-NN trained by SVR

2.4 Face Recognition with a Normalized Correlation

Face recognition results are obtained by comparing the estimated representation of an observed facial image with the representations of gallery images, all of which have a frontal viewpoint. For the comparison, we computed the degree of similarity by using the normalized correlation of the following equation:

$$similarity = \left(\frac{\mathbf{x}_i^{\mathrm{T}} \mathbf{x}}{|\mathbf{x}_i \| \mathbf{x}|} \right), \tag{2}$$

where $\mathbf{x}$ is the estimated representation of an observed faical image and x_i is the representation of a gallery image.

We obtained the recognition result by selecting the gallery representation with the highest degree of similarity.

3 Experimental Results

To validate our model, we performed experiments on viewpoint classification and face recognition from unknown viewpoints. By way of preparation, we trained the view-based representation, view-classification network, and a view-prediction network.

First, we rendered the face images by means of a rendering textured 3-D mesh model in the virtual reality markup language included in XM2VTS DB [9]. We generated 21 viewpoint images for each of the 291 face models, resulting in 5061 gray-scale images. The dimensions of the gray-scale images were 32×32 pixels.

View-based representation was learned as explained in section 2.1. After using principal component analysis to reduce the dimensions of the facial images to 168 dimensions, we maximized Eq. (1) by using a gradient descent and we obtained 168 bases. There were 21 groups and eight bases in each group. The detailed ISA procedures are explained in [7].

3.1 Viewpoint Classification

For a viewpoint-classification network, we divided the face images into roughly 3000 as a training set and 2000 as a test set. We then used the standard perceptron rule to train the network. The network subsequently took the basis activation as the input and gave one out of 21 quantized viewpoints. The pure test recognition rate was 90.2%, though the major classification errors were small, suggesting that the classified viewpoints were not far from the correct answer. Ninety-five percent of the false classifications were within 15 degrees of the correct viewpoint, either horizontally or vertically. This graceful error property is helpful at the prediction step.

3.2 Viewpoint-Invariant Face Recognition

After the viewpoint classification, an RBF-NN is selected to estimate the frontal face representation. Because each face image has 21 viewpoints, we prepared 20 RBF-NNs to estimate the frontal face representation from an arbitrary non-frontal face representation. For the SVR, we used Gunn's SVM open source [8]. Moreover, we used the face data of 241 people for the training, and we used the face data of 50 people for the face recognition testing.

Use of the selected RBF-NN enabled the frontal face representation to be estimated from the representation of the observed non-frontal facial image. The estimation result is shown in the gray image of Fig. 3. The estimated frontal face images look similar to the original frontal images. However, when the viewpoint difference is large, the estimated frontal face images have some artifacts. This phenomenon means that the relation between two widely different viewpoints is more complex than that of two narrowly different viewpoints. For a more accurate estimation, either more examples are needed for the training of the RBF-NN or the facial features should be simplified.
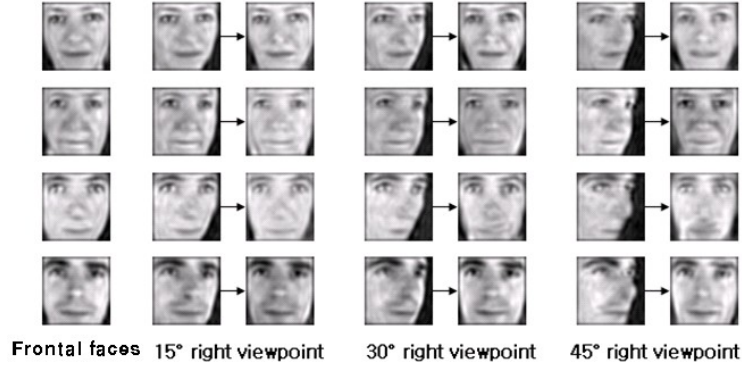
Fig. 3. Estimated frontal face images from non-frontal face images. The leftmost images are the original frontal face images. The images on the right are various viewpoint images and their estimation results

Table 1 shows the face recognition results of 50 people. The face recognition method, which is explained in section 2.4, obtained an overall face recognition rate of 89.33%. The face recognition rate dropped when the viewpoint was rotated horizontally by $\pm 45^\circ$ and also when the viewpoint was rotated vertically.

Table 1. Recognition rate of viewpoint-invariant face recognition

	-45 (L)	-30	-15	0	15	30	45 (R)
Up 15	84%	94%	92%	98%	94%	88%	74%
Front	90%	96%	96%	100%	100%	94%	74%
Down 15	70%	86%	96%	100%	94%	90%	66%

4 Conclusion

Inspired by the biology of human vision, we proposed viewpoint-invariant face recognition model based on view-based representation. We used ISA to simulate view-based representation and we showed that this type of representation has view-dependent characteristics. Thanks to these view-dependent characteristics, our model uses an SLP to easily classify each viewpoint. Furthermore, the association between different viewpoints can be learned by using an RBF-NN that has been trained with SVR. When given a non-frontal face image, the model uses the viewpoint classification to select an RBF-NN for the estimation of the frontal face representation. The RBF-NN successfully estimates the representation of frontal face images from the representation of non-frontal face images. The overall recognition rate was 89.33%.

The results highlight the potential of our biologically inspired model, though the recognition rate dropped for widely different viewpoints. For better results, we plan to find a simpler feature space where associations between viewpoints can be learned more easily. Furthermore, because our experiments were performed with synthesized

facial images under ideal illumination conditions, we plan try to gather and experiment on real images.

References

1. Georghiades, A.S., Belhumeur, P. N., Kriegman, D. J.: Illumination-Based Image Synthesis: Creating Novel Images of Human Faces Under Differing Pose and Lighting. IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes Fort Collins, Colorado (1999) 47-54
2. Okada, K., Steffans, J., Maurer, T., Hong, H., Elagin, E., Neven, H., et al.: The Bochum/USC Face Recognition System and How it Fared in the FERET Phase III Test. Face Recognition: From Theory to Applications, Springer-Verlag (1998) 186-205
3. Desimone, R., Albright, T.D., Gross, C.G., Bruce, C.: Stimulus-Selective Properties of Inferior Temporal Neurons in the Macaque. Journal of Neuroscience, 4(8), (1984) 2051-2062.
4. Graham, D., Allinson, N.: Face Recognition from Unfamiliar Views: Subspace Methods and Pose Dependency. In 3rd Int. conf. on automatic face and gesture recognition (1998) 348-353
5. Hinton, G. E., Sejnowski, T. J.: Unsupervised Learning. Bradford Book (1998)
6. Li, S., Lu, X., Zhang, H.: Unsupervised Learning of View-Subspaces and View-based Clustering of Appearance. MS Research Technical Report, MSR-TR-2001-08 (2001)
7. Hyvärinen, A., Hoyer, P. O.: Emergence of Phase and Shift Invariant Features by Decomposition of Natural Images into Independent Feature Subspaces. Neural Computation,12(7): (2000) 1705-1720,
8. Gunn, S. R.: Support Vector Machines for Classification and Regression. Technical Report ISIS-1-98, Department of Electronics and Computer Science, University of Southampton (1998)
9. Messer, K., Matas, J., Kittler, J., Luettin, J., Maitre, G.: XM2VTS DB: The Extended XMVTS Database. Int. conf. on Audio and Video-based Biometric Person Authentication, Washington D.C. (1999)