Robust Face Recognition from a Single Training Image per Person with Kernel-based SOM-Face

Xiaoyang Tan [1,2] Songcan Chen[2,3] Zhi-Hua Zhou [1] Fuyan Zhang [1]

[1]National Laboratory for Novel Software Technology,
Nanjing University, Nanjing 210093, China
[2]Department of Computer Science and Engineering,
Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China
[3] National Laboratory of Pattern Recognition,
Institution of Automation, Chinese Academy of Sciences, Beijing 100080, China
{txy123,s.chen}@nuaa.edu.cn {zhouzh,fyzhang}@nju.edu.cn

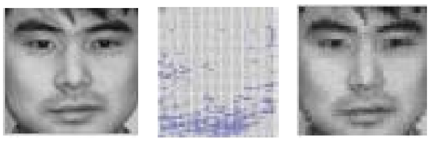
Abstract. In this paper, a kernel-based SOM-face method is proposed to recognize expression variant faces under the situation of only one training image per person. Based on the localization of the face, an unsupervised kernel-SOM learning procedure is carried out to capture the common local features and the non-Euclidean structure of the image data, so that a compact and robust representation of the face can be obtained. Experiments on the FERET face database show that the Kernel-based SOM-face method can obtain higher recognition performance than the regular SOM-face method.

1 Introduction

One of the factors that have strongly affected the performance of face recognition is the face representation model. For example, the eigenface, as a classical representation model for face recognition, tries to find a linear mapping that mostly keeps the variation between the face images [1]. However, due to its linearity in nature, this PCA-based representation can not always capture the non-linear structure of face images. Fisherface is also widely used, which aims to extract the most discriminant features from the face image [2]. However, like other LDA-based methods, Fisherface suffers from the *small sample problem* and will fail in the situation where there is only one training image per person available [3-5]. The limitations discussed above suggest additional research needed on the representation of face image.

Recently, several researchers have tried to model the inherent nonlinear structure of the complex image data using nonlinear methods, such as neural networks [6], support vector machines [7], and kernel methods [8]. In a previous work [5], an SOM-based face representation model called “SOM-face” (see Fig.1) has been proposed to deal with both the nonlinear problem and the small sample problem in face recognition. In this paper, motivated by the success of kernel methods in pattern regression, a kernel-based SOM-face method is proposed. This method generalizes the strength of the kernel method and SOM network while at the same time overcomes some of the shortcomings of regular SOM-face method.

The paper proceeds as follows. The kernel SOM algorithm is introduced in section 2. The proposed method is presented in section 3. The experiments are reported in section 4. Finally, conclusions are drawn in section 5.



a) b) c)

Fig. 1. Example of an original image, its projection and the reconstructed image. a)Original face image. b) The distribution of image in the topological space. c) "SOM face" reconstructed

2 Kernel SOM

The basic idea of Self-Organizing Map [9] is to find and adapt the winner neuron and its topological neighbors according to the current input vector so as to reveal the hidden statistical structures of the input space. However, the commonly used distance measure in the regular SOM algorithm is the Euclidean norm, thus the non-Euclidean neighborhood structure in the input space can hardly be revealed.

Recently, Pan et al. [10] and Andras [11] independently proposed the kernel-SOM algorithm,whose key feature is that the updates of the high dimensional weight vectors are made indirectly by updating the low dimensional vectors in the original space, thus the method can be intepreted as a way to induce different non-Euclidean distance measures for the original space using different kernel fuctions. Common cases for the kernel fuctions are the polynomial, RBF and logarithmic kernels, etc.

In particular, for the RBF kernel (**K**(x, y)=exp(-$||x$-$y||^2$/2 σ^2)) and logarithmic kernel (**K**(x, y)=log((1+$||x$-$y||^2$)/ σ^2)), we have the following kernel-SOM updating rules respectively:

$$\Delta w(t) = \eta(t)h(t)\, {}^{2e^{-\|x-w(t)\|^2/2\sigma^2}}\!\Big/_{\sigma^2} \left(x - w(t)\right) \triangleq \eta(t)h(t)\rho_{d,\sigma}(t)\left(x - w(t)\right) \quad \textbf{(1)}$$

$$\Delta w(t) = \eta(t)h(t)\, {}^{1}\!\Big/_{\sigma^2 \cdot ((1+\|x-w(t)\|^2)/\sigma^2)} (x - w(t)) \triangleq \eta(t)h(t)\rho'_{d,\sigma}(t)\left(x - w(t)\right) \quad \textbf{(2)}$$

By comparing the new updating rules (1) or (2) with the regular SOM updating rules Ä$w(t)$=ç(t)$h(t)(x$-$w(t))$, it can be found that a scale factor ñ$_{d∥\sigma}(t)$ is added in the kernel version, which is affected by the distance d between the input vector x and the winner neuron in the input space under certain kernel paremeter σ at time t. The value of the scale factor decreases as d increases. This means that the attraction of the winner neuron and its neighborhood to the outlier or noisy data will be lessened.

3 The Proposed Method

A high-level block diagram of the proposed method is shown in Fig. 2. The details of the method are described in the following subsections.

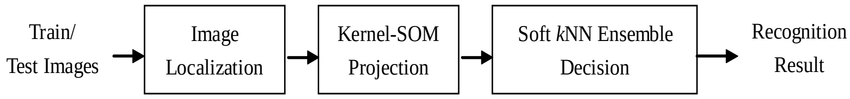


Fig. 2. Block diagram of the proposed method for face recognition

3.1 Localizing Face Image

Since local features are relatively less sensitive to the occlusion and variation in face (such as expression, pose, and illumination) than global features [12], we pay much attention to the local representation of face. One of the simplest ways to localize the face image is to divide the image into M different non-overlapping local sub-blocks with equal sizes, each of which potentially represents specific local information of the image. As a result of the process, a set of sub-block vectors (SBV) is obtained.

In this way, the information of face image is distributed and represented by several low dimensional local features instead of only one high dimensional vector. This helps relax the small sample problem. Furthermore, the sub-pattern dividing process can also help increase the diversity of the SOM-faces [13], which is useful for the soft *k*NN ensemble in the later step.

3.2 The Kernel-SOM Projection

There are two main problems to be solved then, i.e. 1) Since localization process more likely results in many identical sub-blocks from different face classes, that is, some sub-blocks may belong to or be shared by several different classes at the same time, it usually causes the so-called one-to-many mapping problem. 2) In the situation of classification of data with a large number of classes, it is indeed difficult to find a hard class boundary for each class. The first problem makes most supervised learning methods such as radial basis function network (RBF) and multi-layer perceptron network (MLP) fail in the sense of their one-to-one mapping characteristic.

In our previous work [5], we found Self-Organizing Maps (SOM) is a suitable option for those problems, because the neurons of SOM may have multiple class labels at the same time, thus providing the capability of *one-to-many* mapping, on the other hand, the topological preservation property of SOM make it possible to represent the content of each class in a nonlinear way [9]. In this work, we replace the regular SOM with kernel-SOM. The use of kernel method helps find a simpler class boundary structure, which, accordingly, leads to a more robust classification.

The training of kernel-SOM is similar to that of the regular SOM. Its training process is divided into two phases as recommended by [9], that is, an ordering phase and a fine-adjustment phase. We notice that the scale factor $\tilde{n}_{d\| \sigma}(t)$ in (1) or (2) may hurt the ordering performance because of its very small values in the initial phase. Thus we do not make any modification to the regular learning rule in the ordering phase, and the reference vectors are adjusted according to (1) or (2) only in the second phase.

3.3 Soft Nearest Neighbor Decision

Since the SOM makes similar input patterns clustered to adjacent neurons, the relationship between spatially adjacent neurons could be used to improve the classification performance. This leads to a soft *k*NN ensemble decision scheme. A separate soft *k*NN classifier is constructed for each sub-block of the face image to calculate the confidence value for its membership in every class according to:

$$c_{jk} = \log(\tau+1) / \log(d_{jk}+1) \quad (3)$$

where $d_{j\delta k}$ is distance between the *j*-th neuron of the face and its *k*-th nearest neighbors, and ô is the minimum among all the distances. This defines a confidence value C_{jk} for the *j*-th sub-block's membership in *k*-th class, that is, the higher the confidence value for a class, the more likely a sub-block will belong to that class.

Then, the label of the test image can be obtained through majority voting, as follows,

$$Label = \arg\max_k (\sum_{j=1}^{M} c_{jk}), k = 1, \dots C \quad (4)$$

where *M* is the total number of sub-blocks of a face and *C* is the number of face class.

4 Experiments

The experimental face database used in this work comprises 400 gray-level frontal view face images from 200 persons, with the size of 256×384. Each person has two images (**fa** and **fb,** used as training gallery and probes respectively) with different facial expressions. All the images are randomly selected from the FERET face database [14] without any special criterion set forth beforehand. Before the recognition process, the raw images are normalized 60×60 pixels and the inter-ocular distance is 28 pixels.

The details of the experiments are given below. In the localization phase, a block size of 3×3 is used. Then two kind of kernel SOM, i.e. RBF-kernel SOM and log-kernel SOM are trained respectively with 100 updates in the first phase and 400 updates in the second one, with σ =2. The initial weights of all neurons are set to the greatest eigenvectors of the training data, and the learning parameter and the neighborhood widths of the neurons converge exponentially to 0 with the time of training.

Table 1 presents the performance of the proposed method with reference to other template-based approaches, such as nearest neighbor (1-NN), eigenface[1], and $E(PC)^2A$[4] concerning the *top 1 match rate*. Table 1 reveals that the proposed method .achieves higher recognition accuracy than other approaches in dealing with the *one image per person problem*, such as eigenface and $E(PC)^2A$.

Table 1. Comparison of recognition accuracies (%) for six approaches

Method	Accuracy
1-NN	84.0
Eigenface	83.0
$E(PC)^2A$	85.5
Regular SOM-face	87.5
RBF-kernel-SOM-face	88.5
LOG-kernel-SOM-face	89.5

Next, we study the influence of the *k* value. Experimental results are presented in Fig. 3. It can be observed that when the *k* value is small (e.g. *k*<30), the regular SOM-face method performs better than the two kernel-SOM-face methods. However, the overall performance is not so good (<86.5%). When *k* gradually increases to the range between 60 and 120, the performances of all the three compared methods increase as well, ranging from 86% to 89.5%. Among them, the log-kernel SOM-face performs the best, next the RBF-kernel SOM-face, and both the kernel SOM-face methods perform better than the regular SOM-face method.

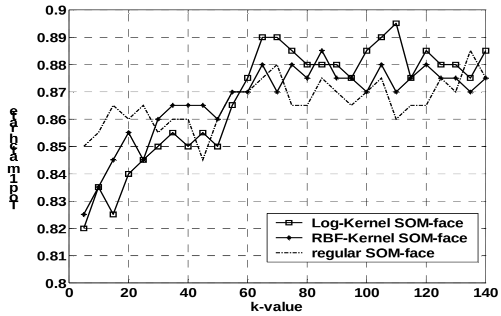


Fig. 3. Top 1 matching rate as a function of *k*-value.

5 Conclusion

In this paper, a novel face representation and recognition approach is presented, where faces are first localized, then represented by their kernel-SOM-based proximities with non-Euclidean distance measure embedded in. Experimental results show the superiority of using kernel-SOM-face to using regular SOM-face in scenarios where only one training image per person is available.

Acknowledgement

This work was supported by the National Natural Science Foundation of China under the Grant No. 60271017, the National Outstanding Youth Foundation of China under the Grant No. 60325237, the Jiangsu Science Foundation under the Grant No. BK2002092, and the Jiangsu Science Foundation Key Project. Portions of the research in this paper use the FERET database of facial images collected under the FERET program.

References

1. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience **3** (1991) 71-86
2. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Analysis and Machine Intelligence **19** (1997) 711-720
3. Wu, J., Zhou, Z.-H.: Face recognition with one training image per person. Pattern Recognition Letters **23** (2002) 1711-1719
4. Chen, S.C., Zhang, D.Q, Zhou, Z.-H.: Enhanced $(PC)^2A$ for face recognition with one training image per person. Pattern Recognition Letters, in press
5. Tan, X.Y, Chen, S.C., Zhou Z.-H., Zhang, F.Y.: Recognizing seriously occluded, expression variant faces from single training image per person with SOM-based kNN ensemble. Technical Report, Computer Science & Engineering Department, Nanjing University of Aeronautics and Astronautics, Nanjing, China, Dec. 2003
6. Raytchev, B., Murase, H.: Unsupervised face recognition by associative chaining. Pattern Recognition **36** (2003) 245-257
7. Pang, S., Kim, D., Bang, S.Y.: Membership authentication in the dynamic group by face classification using SVM ensemble. Pattern Recognition Letters **24** (2003) 215-225
8. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Face recognition using kernel direct discriminant analysis algorithms. IEEE Trans. Neural Networks **14** (2003) 117-126
9. Kohonen, T., Self-Organizing Map. 2nd edition. Springer-Verlag, Berlin (1997)
10. Pan, Z.S., Chen, S.C., Zhang, D.Q.: A kernel-based SOM classification in input space. Acta Electronica Sinica **32** (2004) 227-231 (in Chinese)
11. Andras, P.: Kernel-kohonen networks. International Journal of Neural Systems, **12** (2002) 117-135
12. Pentland, A., Moghaddam, B., Starner, T.: View-based and modular eigenspaces for face recognition. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Seattle, WA, (1994) 84-91
13. Kuncheva, L.I, Whitaker, C.J.: Feature subsets for classifier combination: an enumerative experiment. In: Kittler, J., Roli, F. (eds.): Lecture Notes in Computer Science, Vol. 2096. Springer, Berlin (2001) 228-237
14. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.J.: The FERET database and evaluation procedure for face recognition algorithms. Image and Vision Computing **16** (1998) 295-306