Feature Selection for High Dimensional Face Image Using Self-Organizing Maps

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] Shanghai Key Laboratory of Intelligent Information Processing
Fudan University, Shanghai 200433, China
{x.tan,s.chen}@nuaa.edu.cn {zhouzh,fyzhang}@nju.edu.cn

Abstract: While feature selection is very difficult for high dimensional, unstructured data such as face image, it may be much easier to do if the data can be faithfully transformed into lower dimensional space. In this paper, a new method is proposed to transform the high dimensional face images into low-dimensional SOM topological space, and then identify important local features of face images for face recognition automatically using simple statistics computed from the class distribution of the face image data. The effectiveness of the proposed method are demonstrated by the experiments on AR face databases, which reveal that up to 80% local features can be pruned with only slightly loss of the classification accuracy.

1. Introduction

Face recognition has been an active research area of computer vision and pattern recognition for decades. Many classical face recognition methods have been proposed [1] to date and have obtained success. In most of the subspace-type face recognition method [1,2], each face is represented by a single vector formed by concatenating pixels of the face image in row scan way. Such a representation makes the dimensionality of the feature space very high. On the other hand, the number of available training samples is generally very limited. In some extreme case, only one image is available per person [3-7]. This makes it important to investigate feature selection to improve the generalization of the recognition system.

This, however, can be a very difficult task for some complex data such as face image due to the sparseness nature of the high dimensional feature space. To circumvent that problem, we therefore propose to transform the high dimensional, unstructured face image data to lower dimensional space first, and then select features in the latter space. The task of feature selection may be much easier due to the simplification of the feature space. In previous work, we have found that the SOM (self-organizing maps, [8]) topological space is suitable for face representation [6], and an SOM-based face representation model called "SOM-face" has been proposed.

In this paper, we extended the work [6] in two aspects: (1) a novel method of automatically selecting important local features from the face image for recognition is proposed, and (2) investigate the problem of how much of the local features can be reduced without losing useful information in class prediction. The paper proceeds as follows. After briefly reviewing the SOM-face model in section 2, we described the proposed method in section 3. The experiments are reported in section 4. Finally, conclusions are drawn in section 5.

2. The SOM-face Model

The essence of SOM-face [6] is to express each face image as a function of local information presented in the image. This is achieved by dividing the image into k different local sub-blocks at first, each of which potentially preserves some structure information of the image. Then, a self-organizing map (SOM) neural network is trained using the obtained sub-blocks. The reconstruction of a face in the SOM topological space is called SOM-face (see Fig.1). Note that any face localized in the same way above can be projected onto the quantized lower dimensional space to obtain its compact but robust representation. The main advantage of such a representation is that, in the SOM-face, the information contained in the face is distributed in an orderly way and represented by several neurons instead of only one neuron corresponding to a weight vector, so the common features of different classes can be easily identified.

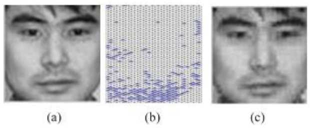


Fig. 1. Example of an original image, its projection and the reconstructed image a) Original face image. b) The distribution of image in the topological space. c) “SOM face” reconstructed

3. Feature selection in the SOM topological space

SOM mapping makes it feasible for us to analyze the degrees of importance of different local areas based on simple statistics computed from the empirical distribution. Here, three criterions are proposed to measure the goodness of sub-blocks of each face, i.e., face frequency (FF), a χ^2 statistic (CHI) and neuron purity (NP).

- Face frequency criterion (FF)

Face frequency is derived from the concept of document frequency in automatic text analysis field [10]. Here it means the number of different faces a neuron attracts. Based on this simple statistics, two strategies can be applied to perform feature selection. The first strategy (FF-1) is based on the assumption that the rare neurons (i.e., the neurons with low FF value) may be either non-informative for recognition or not influential in global performance. The degree of importance of a neuron can be calculated as a *non-decreasing* function of its FF value.

On the other hand, the FF-value can be regarded as an indicator of the distribution of the sub-blocks of all the faces in SOM topological space. That is, big FF-values indicate much overlap among the distributions of different classes and hence inducing low discriminability, whereas small values indicate little overlap and hence high discriminability. In this sense, the degree of importance of a neuron can also be calculated as a *non-increasing* function of its FF value. This strategy is named FF-2.

- χ^2 statistic (CHI)

The χ^2 statistic tries to measure the dependence between class and term. Consider a two-way contingency table of a neuron *t* and a face *c*, where *A* is the number of times the *t* and *c* co-occur, *B* is the number of time the *t* occurs without *c*, *C* is the number of times the *c* occurs without *t*, and *N* is the total number of faces, then the χ^2 statistic between the *t* and *c* is defined to be:

$$\chi^2(t,c) = \frac{N \times (AD - CB)^2}{(A+C)\times(B+D)\times(A+B)\times(C+D)} \quad (1)$$

The χ^2 statistic is zero if t and c are independent. In this study, we computed the pairwise χ^2 statistic between each neuron and each training face, and then measured the final neuron-goodness score according to the maximal rule:

$$\chi_{\max}{}^2(t,c) = \max_{i=1}^{C}\{\chi^2(t,c_i)\} \quad (2)$$

- Neuron purity (NP,[9])

Neuron purity is a measure to quantify the degree of separability of a neuron. Let the number of sub-blocks of class c attracted by neuron t be λ_c, then the probability of class c in neuron t is given as $p_{ct} = \frac{\lambda_{ct}}{\sum_{c\in C}\lambda_{ct}}$ and the degree of separability (or the purity) of a neuron t is defined to be:

$$NP_t = \sqrt{\frac{K_t}{K_t - 1}\sum_{i=1}^{K_t}(P_{it} - 1/K)^2} \quad (3)$$

where K_t is number of classes attracted by neuron t.

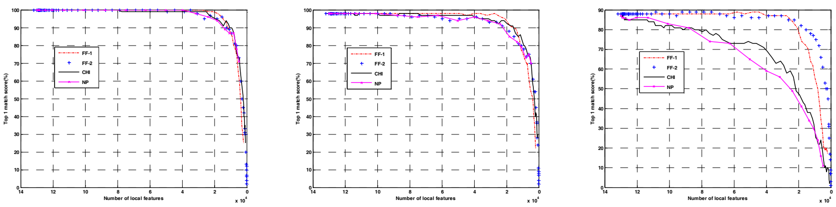
4. Experiments

The AR face database [11] is used in the experiments, which contains over 4,000 color face images of 126 people's faces, among them, a subset of 400 images from 100 different subjects was used. Some sample images are shown in Fig.2.

Fig. 2. Sample images for one subject of the AR database [10].

Before the recognition process, each image was cropped and resized to 120x165 pixels and then converted to gray-level images, which were then processed by a histogram equalization algorithm. A sub-block size of 5x3 was used and only the neutral expressions images (Fig.2a) of the 100 individuals were used for training, while the other three were used for testing.

We ranked the neurons according to the goodness value obtained with different criterions, and removed those neurons whose goodness values are below some predefined threshold. Since different faces have different set of neurons, the resulting neurons after pruning are also different face by face and only the left neurons would be used for classification. The relationship between the top 1 recognition rate and the total remaining local features of the probe set are displayed in Fig.3.



(a) (b) (c)

Fig. 3. Top 1 recognition rate vs. total remaining unique local feature count (a)-(c) are corresponding to the smile, anger and scream images (Figs.3b, 3c, and 3d)

It can be observed from Fig. 3 that on the first two probe sets, the four compared criterions (FF-1, FF-2, CHI and NP) have similar effect on the performance of the recognition system, and 80% or more of the sub-blocks can be safely eliminated with almost no loss in recognition accuracy. On more challenging probe sets (Fig.2d), two FF-type criterions perform much better than the other two (CHI and NP). This observation indicates that the criterions which make use of the class information do not necessarily lead to excellent performance.

5. Conclusions

In this paper, a novel feature selection method for high dimensional face images based on their SOM transformation is proposed. Experiments on AR database reveal that up to 80% sub-blocks of a face can be removed from the probe set without loss of the classification accuracy. This could be particularly useful when a compact representation of face is needed, such as in the application of smart card, where the storage capability is very limited.

Acknowledgement

This work was supported by the National Science Foundation of China under the Grant No. 60271017, the National Outstanding Youth Foundation of China under the Grant No. 60325207, and the Jiangsu Science Foundation Key Project under the Grant No. BK2004001.

References

1. Zhao W., Chellappa R., Phillips P. J., and Rosenfeld A.. Face Recognition: A Literature Survey. *ACM Computing Survey*, 2003,35(4): 399-458.
2. Turk M. and Pentland A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 1991, 3(1): 71-86.
3. Wu J. and Zhou Z.-H. Face recognition with one training image per person. *Pattern Recognition Letters*, 2002, 23(14): 1711-1719.
4. Chen S. C., Zhang, D. Q., and Zhou Z.-H. Enhanced $(PC)^2A$ for face recognition with one training image per person. *Patt. Recog. Lett.*, 2004, 25(10):1173-1181.
5. Chen S. C., Liu J., and Zhou Z.-H. Making FLDA applicable to face recognition with one sample per person. *Pattern Recognition*, 2004,37(7): 1553-1555
6. Tan X., Chen S.C., Zhou Z.-H., and Zhang F. Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft kNN ensemble. IEEE Trans. Neural Networks, in press.
7. Martinez A.M. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE TPAMI, 2002, 25(6): 748-763
8. Kohonen T. *Self-Organizing Map*, 2nd edition, Berlin: Springer-Verlag, 1997.
9. Singh S., Singh M., and Markou M. Feature selection for face recognition based on data partitioning. In: *Proc. 15th ICPR*, Quebec, Canada, 2002, 680-683.
10. Yang Y. and Pedersen J. O. A comparative study on feature selection in text categorization. In: *Proc. 14th ICML*, Nashville, TN, 1997, 412-420.
11. Martínez A. M. and Benavente R. The AR face database. *CVC Technical Report*, no.24, June 1998.