Weighted SOM-Face: Selecting Local Features for Recognition from Individual Face Image

Xiaoyang Tan[1,2,3], Jun Liu[1], Songcan Chen[1,3], and Fuyan Zhang[2]

[1] Department of Computer Science and Engineering
Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China
{x.tan,s.chen}@nuaa.edu.cn
[2] National Laboratory for Novel Software Technology
Nanjing University, Nanjing 210093, China
fyzhang@nju.edu.cn
[3] Shanghai Key Laboratory of Intelligent Information Processing
Fudan University, Shanghai 200433, China

Abstract. In human face recognition, different facial regions have different degrees of importance, and exploiting such information would hopefully improve the accuracy of the recognition system. A novel method is therefore proposed in this paper to automatically select the facial regions that are important for recognition. Unlike most of previous attempts, the selection is based on the facial appearance of individual subjects, rather than the appearance of all subjects. Hence the recognition process is class-specific. Experiments on the FERET face database show that the proposed methods can automatically and correctly identify those supposed important local features for recognition and thus are much beneficial to improve the recognition accuracy of the recognition system even under the condition of only one single training sample per person.

1 Introduction

Face recognition has been an active research area of computer vision and pattern recognition for decades [1]. Recently, great attention has been paid to the recognition methods using local features of face images due to its robust performance against such variation as noise, occlusion and expression [1-7].

An interesting question that then should be answered in this context is: which portions of a face image are really important for recognition? Empirically, such local regions are eyes, mouth and nose et al. Several authors have tried to manually identify such significant regions with different masks [2, 3]. For example, Brunelli and Poggio [2] used four masks respectively to extract the regions of eyes, mouth and the whole face for recognition, and their experimental results confirmed the discriminating power of such local facial features. Later, Pentland et al. [3] extended Brunelli and Poggio's work by projecting each local feature onto its corresponding eigenspace and using the obtained eigenfeatures for recognition. Both the above works need the involvement of great human endeavor, which is not desired in real applications. Moreover, defining in advance the same regions for all the classes seems to be inconsistent with our intuition that each class should have its own class-specific features, which are really meaningful to the recognition.

Recently, several researchers have also used disrciminant analysis for separating useful from useless facial variation [5,12,13], however, those approaches require a

M. Gallagher, J. Hogan, and F. Maire (Eds.): IDEAL 2005, LNCS 3578, pp. 351–358, 2005.

number of training samples per class in general, while it is not uncommon in practice that there is only one training image per person available to the system (such scenarios including law enforcement, passport or identification card verification, and so on). This suggests additional research needed in this direction. Several works have been developed to attack the *one training sample per person* problem from different respects, including synthesizing virtual samples [6,7], probabilistic matching[7], localizing the single training image [4,5] and neural network method[4].

The self-organizing maps (SOM, [8]) is an artificial neural networks model that implements a characteristic nonlinear projection from the high-dimensional space of signal data into a low-dimensional array of neurons in an orderly and discretely fashion. In the classification of data with a large number of classes such as face recognition, it is very difficult to obtain the hard class boundaries, while the properties of topological preservation and *one to many* mapping supported by SOM are especially useful in such situation.

In this paper, a novel method of automatically selecting some important local features from a single training face image for recognition is proposed. The method is based on a SOM-based face representation model called "SOM-face", which will be briefly reviewed in section 2. We described the proposed method in section 3 and the classification method in section 4. The experiments are reported in section 5. Finally, conclusions are drawn in section 6.

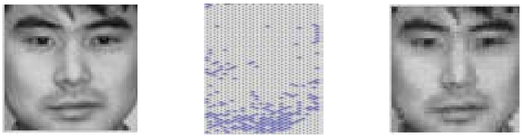
2 The SOM-Face

The essence of SOM-face [4] is to express each face image as a function of local information presented in the image. This is achieved by dividing the face image I into M different local sub-blocks $R_i \Big|_{i=1}^{M}$ at first, each of which potentially preserves some structure information of the image.

Then, a self-organizing map (SOM) neural network is trained using all the obtained sub-blocks from all the available training images. After the SOM map has been trained, each sub-block R_i from the same face image I can be mapped to its corresponding Best Matching Units (BMUs) by a nearest neighbor strategy, whose location in the 2D SOM topological space is denoted as a location vector $l_i = \{x_i, y_i\}$. We can group all the location vectors from the same face as a set, i.e., $I = \{l_i\}_{i=1}^{M} = \{x_i, y_i\}_{i=1}^{M}$, which is called the face's "SOM-face" representation. Note that such a representation is both compact and robust: on one hand, the possible faults like noise in the original face image can be eliminated in the process of SOM training, on the other hand, the "SOM-face" representation is different from other SOM-based VQ methods in that only the location vector of each sub-block will be as prototype for later recognition purpose, while the weight vector is not used here.

Moreover, 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 or vector, so the common features of different classes can be easily identified. This merit is extremely useful for the work in this paper.

Fig.1 shows an example of an original image, its projection ("SOM-face") and the reconstructed image with the corresponding weight vectors.



a) Original face b) SOM-face c) Reconstructed face

Fig. 1. Example of an original face image, its projection and the reconstructed image

3 Identifying Important Local Features Adaptively

In human face recognition, different facial regions have different degrees of importance. Hopefully, exploiting such information would improve the accuracy of the recognition system [7]. However, many local face recognition approaches do not consider that, thus a test image can be easily misclassified. Here we present a novel method able to extract such important information from the SOM-face representation of a face image automatically.

In the proposed method, each neuron in SOM topological space is regarded as a *code word* denoting some local area of the original face image. So, each original face image can be described by a specializing subset of those code words, and all the face images share the same dictionary which is actually all the neurons existing in the SOM topological space. This novel interpretation of the SOM output neurons make it feasible for us to analyze the degrees of importance of different local areas by computing the weights of corresponding neurons.

Consider a face image I that has been partitioned into M non-overlapping sub-blocks as described in Section 2. To measure the degree of significance of each sub-block in I, we present all the sub-blocks of I to the trained SOM network once again, counting for each activated neuron the number of sub-blocks that are attracted. Formally, let the number of sub-blocks of I attracted by the j-th neuron be tf_j, then tf_j can be regarded as some measure of the expressive power of the neuron for I: the more sub-blocks it attracts, the stronger the expressive power it has. However, most expressive features may not be the most discriminating features, and the distribution of the neurons in the whole face space should also be considered.

Let the number of different classes attracted by the j-th neuron be n_j, which is considered here as an indicator of the distribution of the sub-blocks in SOM topological space. Big n_j values indicate much overlap between the distributions for different classes and hence low discriminability, whereas small n_j values indicate little overlap and hence high discriminability.

Combining both the above factors (i.e. tf_j and n_j), we can evaluate the importance of a neuron to a given face. And the importance of each sub-block to a given face can therefore be calculated indirectly by evaluating the significance of its corresponding neuron (in other words, BMU). Formally, let the corresponding neuron of the i-th sub-block of face I be j, then the sub-block's degree of importance w_i can be evaluated as follows [9]:

$$w_i = tf_j * \log(C/n_j + 1) \tag{1}$$

where C is the total number of classes to be recognized.

4 Recognition Based on Weighted SOM-Face

Now we describe a soft-*k*NN-based classification method (called weighted SOM-face) used to give a label to a testing face. The soft-*k*NN decision strategy is used here to reduce *k*NN's sensitivity to noise and to exploit the SOM's topological preservation property. Note that the selection of important local facial features becomes a natural derivation of the algorithm, just by setting the weights of unimportant features to zero.

Formally, suppose that the probe face is divided into *M* sub-blocks, denoted as $I_{probe} = \{R_1, R_2, ..., R_M\}$, where R_r is the *r*-th sub-block, whose BMU's locations in the 2D SOM topological space is $l_r = \{x_r, y_r\}$. Then we denote the set of *k* nearest neighbors of l_r as $N_k(l_r) = \{l_{r,1}, l_{r,2}, ..., l_{r,k}\}$, where each element $l_{r,i}\Big|_{i=1}^{k}$ is a prototype vector of the *r*-th sub-block from one of the *C* classes.

The confidence value $C_{r,i}$ describing the probability for the *r*-th sub-block's membership in the C_i can be evaluated based on their pairwise distance in the SOM topological space, as follows:

$$c_{r,i} = f(d(l_r, l_{r,i})) \tag{2}$$

where *d(,)* is the function to compute Euclidean distance, and *f(.)* is a real-valued monotonous decreasing function, satisfying the condition *f(0)=1*. In this paper, the *f(.)* function adopted is as follows:

$$f(d) = \frac{\log(\tau(d)+1)}{\log(d+1)} \quad d > 0 \tag{3}$$

where $\tau(d) = \min\{d(l_r, l_{r,i}), j = 1, ..., k\}$, that is, the minimum pairwise distance between *r*-th block and its *k* nearest neighbors in the SOM topological space.

Finally, a weighted linearly-summed voting scheme (Eq.4) is employed to assign the final label of the test image as the class with the maximum total confidence value. The weight (importance) w_i of each neuron obtained in Section 3 is of course incorporated into the calculation, as follows:

$$label = \arg\max_j \left(\sum_{i=1}^{M} w_i c_{r,j}\right) \quad j = 1, 2, ..., C \tag{4}$$

where w_i is subject to the condition: $\sum_{i=1}^{M} w_i = 1$.

5 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. There are 71 females and 129 males. Each person has two images (**fa** and **fb**) with different facial expressions. The **fa** images are used as gallery for training while the **fb** images as probes for testing. All the images are randomly selected from the FERET face database [11]. Some samples of the database are shown in Fig.2. Before the recognition process, the raw images were normalized and cropped to a size of 60×60 pixels.

In the localizing phase, the training images are partitioned into non-overlapping sub-block with size of 3×3. Then a single SOM map with the size of 88×16 using the sub-blocks obtained from partitioning all images is trained. The training process is divided into two phases as recommended by Kohonen [4], that is, an ordering phase and a fine-adjustment phase. 1000 updates are performed in the first phase, while 2000 times in the second one. 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 1 with the time of training.

Fig. 2. Some raw images in the FERET database

The analysis described in Section 3 indicates that the performance of the recognition system may be improved by exploiting the discriminating information contained in different local features. In order to verify this hypothesis, we first perform experiments to compare the performance of the proposed methods with that of some other approaches dealing with face recognition with one training image per person, such as eigenface [2], Enhanced (PC)^{2}A algorithm(E(PC)^{2}A,[8]) and Matrix Fisher Linear Discriminant Analysis (MatFLDA, [5]).

The comparison result is tabulated in Table.1, which reveals that when the top 1 match rate is concerned, the weighted SOM-face method achieves the best performance among the compared approaches. This result indicates that the weights of different sub-blocks are indeed informative for face recognition.

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

Method	Accuracy
Standard Eigenface	83.0
E(PC)^{2}A[6]	85.5
MatFLDA[5]	86.5
Regular SOM-face	87.5
Weighted SOM-face	**89.5**

To further study the behavior of the weighted methods, another set of experiments are conducted to compare the performance of the weighted and non-weighted SOM-face methods, concerning different *k* value used in the soft *k*NN decision. The results are shown in Fig.3, with eigenface as the benchmark.

It can be observed from Fig.3 that the weighted strategy outperforms the non-weighted SOM-face on the whole. In particular, when *k=1*, the top1 match rate of the weighted strategy is only 77.0%. However, with the increase of *k*-value, the matching rate rises as well. In particular, when *k* gradually increases to 50 (i.e. about half of the image database size), the weighted method began to perform better than its non-weighted counterpart, and since then, a large performance margin between the two methods can be observed. This again reveals the usefulness of the weight information to the recognition system.

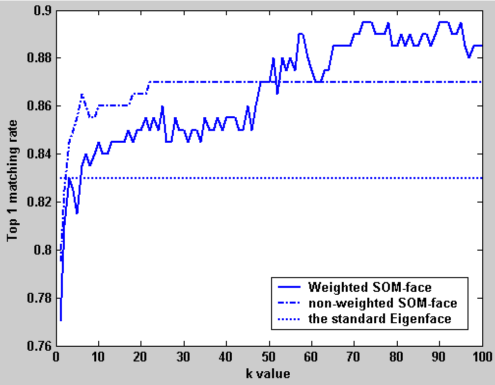
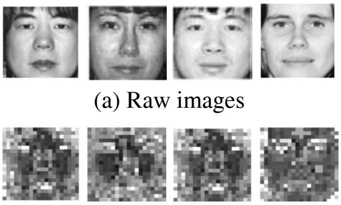


Fig. 3. Top 1 match rate as a function of *k*-value with different methods

Now we shall visualize the important local features selected by the weighted method to help us interpret the obtained results. Fig.4 shows some images and its corresponding important local features. The brighter the color of a sub-block, the higher degree of importance it has.

The images of Fig.4 reveal that the selected important local features contain in most case the regions of the face supposed important for classification: eyes, mouth, nose and hair, while the cheek is *not* considered important. These results are in accord with the observations reported by other researchers [2,3]. In addition, the figure shows that the important local areas of different classes are also different. This confirms our intuition that the local areas important to face recognition should be class-specific. This motivates us to further study on this issue to fully exploit the potential discriminative power of the local facial features.



(a) Raw images

(b)Relative importance of facial features

Fig. 4. Some images and the distribution of important local features weighted by the proposed method

Finally, it is worthy to take some discusses about the size of sub-blocks. The choice of sub-block size reflects the balance between generalization and specialization. Specialization means the unique properties of the given face image (not necessarily of the individual himself), while generalization means generic properties that are less sensitive to the minor changes of a given image, such as noise, illumination, etc. Generally, as the sub-block gets smaller, the degree of generalization grows higher, at the same time the degree of specialization becomes lower. In this sense, to improve the robustness of the system, a smaller sub-block size may be desired, nevertheless, it is nonsense if too small size is used because each face consists of the same set of gray-value pixels in some view. More experimental details please refer to [4].It is also worthy to note that in this implementation, only the appearance-type feature (i.e., the grey values of each pix in a sub-block) is used, while other more complex invariant feature such as Gabor wavelet can be readily used. However, in that case, the size of sub-block should be relatively large to make the feature extraction possible. This will be the focus of our future research.

6 Conclusions

In this paper, a feature weighting strategy is proposed to calculate the importance of different local facial features from individual face image, based on simple statistics computed from its classification distributions on SOM surfaces. In light of the ideas from the automatic text analysis field, the proposed method assigns higher weights to local features which are both expressive and discriminative to the face image. Experiments on FERET database show that the proposed method can automatically identify those supposed important local features as eyes, mouth, nose and hair for recognition and furthermore, exploiting such information is much beneficial to improve the recognition accuracy of the recognition system even under the condition of only one single training sample per person.

Acknowledgement

This work was partially supported by the Jiangsu Science Foundation Key Project (BK2004001). Portions of the research in this paper use the FERET database of facial images collected under the FERET program.

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. Brunelli R. and Poggio T. Face recognition: features versus templates. IEEE TPAMI, 1993,15(10): 1042-1062
3. Pentland A., Moghaddam B., and Starner T. View-based and modular eigenspaces for face recognition. In: Proc. of the IEEE Inter. Conf. CVPR, Seattle, WA, 1994, 84-91
4. Tan, X.Y., 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 Transactions on Neural Networks, in press
5. Chen, S. C., Liu, J., and Zhou, Z.-H. Making FLDA applicable to face recognition with one sample per person. Patt. Recog., 2004, 37(7): 1553-1555
6. 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, 25:1173-1181
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. Grossman D. A. and Frieder O. Information Retrieval: Algorithms and Heuristics, Boston, MA: Kluwer, 1998.

10. Singh, S., Singh, M. and Markou, M., “ Feature Selection for face Recognition based on Data Partitioning”, Proc. 15th Int. Conf. Patt. Recog., ICPR'02, 11-15 August, 2002
11. Phillips P. J., Wechsler H., et al. The FERET database and evaluation procedure for face recognition algorithms. Image and Vision Computing, 1998, 16(5): 295-306
12. Moghaddam, B., Pentland, A., 1997. Probabilistic visual learning for object representation. IEEE Trans. on Pattern Analysis and Machine Intelligence 19(7), 696-710.
13. Belhumeur P., Hespanha J., and Kriegman, D. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE TPAMI, 1997, 19(7): 711-720.