Abstract

It is generally agreed that faces are not recognized only by utilizing some holistic search among all learned faces, but also through a feature analysis that aimed to specify more important features of each specific face. This paper addresses a novel decision strategy that efficiently uses both holistic and facial component (left eye, right eye, nose and mouth) feature analysis to recognize faces. The proposed algorithm uses the whole face features in the first step of recognition task. If the decision machine fails to assign a class (with high confidence) then the individual facial components are processed and the resulting information are combined with those obtained from the whole face to assign the output. Simulation studies justify the superior performance of the proposed method as compared to that of Eigenface method. Experimental results also show that the proposed system is robust against small errors in facial component extractor.

1. Introduction

Identifying people and facial recognition is an important and vital human ability which persists even through the passage of time, changes in appearance and partial occlusion. During the past decade different approaches for face identification are emerged, among them the most widely studied ones are based on statistical learning methods. The objective of statistical learning is to find, based on optimization of a certain separability criteria, a transformation which produce a feature representation with enhanced discriminatory power.

While designing automatic face recognizers has been extensively improved in recent years, there is apparently a large gap between the current systems and the human ability of face recognition. One can even argue that while face recognition in human is a special task (with special cells), most of the current approaches in designing AFR can be considered as engineering solutions for object recognition and AFR is served as a benchmark for developing better methods of object recognition (shouldn't we apply many of the AFR methods for classifying other objects?). On the other hand, it is generally agreed that face recognition involves not only processing information about spatial layout or configuration of facial features (holistic view) but also information about individual facial features (mouth, nose, eyes, etc.). Utilizing the information from facial components is not only inspired by our biological findings but also has the following advantages

- Image variations (due to pose and/or illumination changes) within each component patch is more easily compensated than that of the whole image.
- A facial component can be weighted according to its importance. The component with a large variation can be weighted less in the matching stage.

One of the first studies on the usage of facial components in the human face recognition was carried out by Brunelli and Poggio [1]. Brunelli suggest different mechanism to combine the classification results of each individual component for final decision including added score, weighted added score and the score of the most similar feature. Kim combines the facial features extracted by LDA into a single vector and uses the obtained vector for classification [2]. This paper introduced a hierarchical classifier that mimics the holistic/component behavior of human in recognizing faces. The proposed algorithm uses the whole face features in the first step of recognition task. If the decision machine fails to assign a class with high confidence then the individual facial components are processed and the resulting information are combined with those obtained from the whole face to form the final result. The novelty of our proposed methods comes from efficiently combining the individual facial components (left eye, right eye, nose and mouth) and the whole face features.
features and solving the recognition problem by a hierarchical decision structure.

We want to emphasize that our novel approach can be considered as a practical way of combining the facial components in the decision path of most of the subspace based AFRs. So one can use PCA [3], LDA [4], ICA [5], KPCA [6] or any other statistical approach (and even a combination of them) to extract facial features and utilize our decision structure to design an AFR.

The feasibility of the proposed methods has been successfully tested on ORL and YALE datasets where the images vary in expression and pose. The effectiveness of our methods is shown by comparing the recognition accuracy of the proposed methods with the well-known Eigenface method. This rest of this paper is organized as follows. In section 2 the proposed hierarchical classifier is introduced. Simulation results are described in section 3 and finally, conclusions are given in section 4.

2. Methodology

Different mechanism can be utilized to use the facial component information in the recognition task. For example, one can extract the facial component features using PCA/ICA/LDA/KPCA and then combine these features into a single vector for recognition. As suggested by Brunelli, another approach is to classify each individual component and then added the score of each classifier to form the output. Intuitively, designing separate classifiers for facial components and fuse the results sound more logical than designing a classifier for the combined feature vector. Another advantage of this approach is robustness against pose and illumination because one can hope that some of the facial components are not degraded drastically by the changes in pose and illumination conditions. Biological studies also suggest the possibility of global description serving as a front end for finer, feature-based perception of faces [7].

The proposed block diagram of a recognition system that tries to mimic this behavior is shown in Fig. 1. After decomposing the input image into its components, the whole image and individual component features are extracted and fed into the decision machine. The decision machine first processes the features of the whole image in order to find a match within the stored database. If a match with high confidence is found then there is no need for further processing and the output is assigned. If the decision machine could not decide on the output class by just examining the whole image features, then the individual components are processed.

Different strategies can be devised here,

- Whole image and facial component assign their outputs and the decision machine performs a majority voting for final decision.
- Whole image and facial component assign their outputs and the decision machine performs a weighted majority voting for final decision.
- Facial component modules only searches among the best matching subjects obtained from the whole image classifier and then the decision machine perform a majority voting or a weighted majority voting to assign the output.

The strategy adopted in the reported experiment is the last one (Fig. 2). In our experiments we have utilized the weighted majority voting algorithm for assigning the output. The weights can be determined (from the training set) by using gradient based algorithms or even search algorithms with the leave one out mechanism. For extracting the whole image and facial component features, we utilized the well known Eigenfeature analysis. The whole image classifier selects $k$ candidates for the recognition using a nearest neighbor (k-NN) classifier. If more than $k/2$ of the neighbors belong to the same class and this class contains the nearest neighbor too, then confidence check is passed and the classifier output is the label of that class, otherwise facial component classifiers (which are nearest neighbor classifiers too) compare the input pattern with the stored patterns of the $k$ nearest classes obtained from the whole image classifier. In sequel the experimental results are reported.

![Fig. 1. Block diagram of the proposed recognition system.](image-url)
3. Experimental Results

We assess the performance of the proposed methods using two datasets: A) ORL dataset that contains 400 images corresponding to 40 subjects and B) YALE dataset containing 165 images from 15 individuals. For both dataset we have used the k-NN classifier with \( k = 3 \) as whole image classifier. We also used the well-known Eigenface method for feature extraction and compare the recognition curves of the proposed methods with Eigenface algorithm. We want to reemphasize that our novel approach can be considered as a practical way of combining the facial components in the decision path of most of the subspace based AFRs. Eigenfeature analysis [8], integral projection [1], multiresolutional analysis [9], and etc. are examples among different approaches devised for facial component extraction during the past years. As this paper is focused on the applications of facial components in boosting the recognition performance, the facial components are extracted manually. Later we will show how the performance of the proposed algorithm is affected by the accuracy of facial component extractor. Some examples of the extracted components for the first subject of YALE database is shown in Fig. 3.

In the sequel we assess the efficiency of our proposed methods against Eigenface method by comparing the recognition rate curves.

3.1. Tests on ORL dataset

The ORL face database (developed at the Olivetti Research Laboratory, Cambridge, U.K.) is composed of 400 images with ten different images for each of the 40 distinct subjects. The variations of the images are across pose, size, time, and facial expression. All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position, with tolerance for some tilting and rotation of up to about 20°. There is some variation in scale of up to about 10%. The spatial and grey-level resolutions of the images are 92×112 and 256, respectively. Some of the images of this database are shown in Fig. 4.

The first three images (from the ten available images for each subject) are chosen for training, while the remaining images are used for testing. In general, the performance of the Eigenface method varies with the number of principal components. Fig. 5 shows the recognition performance of the facial components and the whole image for the ORL database. Among the facial features, left eye component had the maximum discrimination power (near 65% correct classification). The facial components can be sorted by decreasing performance as follows

1. Left Eye
2. Mouth
3. Nose
4. Right Eye

It is interesting to note that the differences between recognition performances of the selected facial components are less than 7%, and there is a large gap between these components and whole image as the performance of whole image is always higher than the facial components by more than 20%.
Fig. 5. Recognition performance for each facial component and the whole face on ORL database.

Fig. 6 shows the results of applying our proposed method to the ORL database. The relative performance of the algorithms is self evident in this figure. Comparing the recognition curves reveals that our proposed method outperformed the Eigenface algorithm for all feature dimensions. The maximum recognition is obtained by using 90 features (the feature dimension is the identical for whole image and facial components) which is 89.29%. The maximum difference between performances is 3.6% and again obtained when 90 features are used.

The Hierarchical decision machine uses the weights tabulated in Table 1 for fusing the results of whole image and facial component classifiers. These weights are found by forcing the weights to be an integer in the interval $[0,3]$ and brute force searching the weight space.

### Table 1. Hierarchical decision machine weights for ORL dataset

<table>
<thead>
<tr>
<th>Name</th>
<th>Whole image</th>
<th>Left eye</th>
<th>Right eye</th>
<th>Nose</th>
<th>Mouth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

### 3.2. Tests on YALE dataset

The Yale face database contains 165 images with 11 different images for each of the 15 distinct subjects. The 11 images per subject are taken under different facial expression or configuration: center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and wink. Some of the images of this database are shown in Fig. 7. The training set contains center-light, normal and without glasses of each subject and with glasses, happy, sad, sleepy, surprised and wink are used for testing.

Fig. 7. Some examples of the YALE database.

Fig. 8 shows the recognition performance of the facial components and the whole image for the YALE database. Among the facial features, nose component had the maximum discrimination power (near 36% correct classification). The facial components can be sorted by decreasing performance as follows

1. Nose
2. Mouth
3. Right Eye
4. Left Eye

It is again interesting to note that the performance of whole image is always higher than the facial components by more than 45%. This is mainly because the differences between facial component images of each subject (due to lighting condition and the facial expression) and are higher than that of ORL database.
Table 2 shows the hierarchical decision machine weights which are again found by brute force searching the weight space. As it is evident from this table, the decision machine weights nose component more than the other components because it has higher recognition performance.

<table>
<thead>
<tr>
<th>Name</th>
<th>Whole image</th>
<th>Left eye</th>
<th>Right eye</th>
<th>Nose</th>
<th>Mouth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

In Fig. 9, the recognition rate of the proposed algorithm is compared to that of Eigenface method. Again for all feature dimensions the proposed method outperforms the Eigenface algorithm. The maximum recognition is obtained by using 19 features which is 93.33%. The maximum difference between performances is 4.4% and obtained when 15 features are used.

3.3. Robustness against facial component extractor

In order to evaluate the dependence of the proposed algorithm on the accuracy of facial component, the following simple experiment is carried out on the YALE data-base. A uniform noise is added to the position of each facial component patch in the training set and test set. The power of the noise is dependent on the window size of facial component respectively. Fig. 10 shows the recognition rate of the proposed algorithm under noise disturbance. The performance of the hierarchical system outperforms the Eigenface method up to noises which can move the facial component window by ±20% window size randomly. If the noise power increases to ±30% of window size, the performance is converge to that of Eigenface method for higher number of features which shows that the facial components are no more useful how ever for small number of features still the proposed method outperforms the Eigenface algorithm.

4. Conclusion

In this paper we proposed a hierarchical system which tries to mimic the human recognition behavior by efficiently combining the facial components in the recognition task. The decision system uses the weighted majority strategy to fuse the results of whole image and facial component classifier. Simulation studies justify the superior performance of the proposed method as compared to that of Eigenface method. Experimental results also show that the proposed system is robust against small errors in facial component extractor. Our next goal is to design an adaptive system which learns the weights for each subject in the training set.
5. References


