Abstract

In this paper, we propose a robust face recognition technique based on the principle of eigenfaces. The traditional eigenface recognition (EFR) method works quite well when the input test patterns are cropped faces. However, when confronted with recognizing faces embedded in arbitrary backgrounds, the EFR method fails to discriminate effectively between faces and background patterns, giving rise to many false alarms. In order to improve robustness in the presence of background, we argue in favor of learning the distribution of background patterns. A background space is constructed from the background patterns and this space together with the face space is used for recognizing faces. The proposed method outperforms the traditional EFR technique and gives very good results even on complicated scenes.

Keywords: Face recognition, eigenfaces, face detection, background learning

1. Introduction

In the literature, several works have appeared on the face recognition problem [1, 2, 3, 4, 5]. One of the very successful and well-known face recognition methods is based on the Karhunen-Loeve (KL) expansion [3]. In 1986, Sirovich and Kirby [3] studied the problem of KL representation of faces. They showed that if the eigenvectors corresponding to a set of training face images are obtained, any image in that database can be optimally reconstructed using a weighted combination of these eigenvectors. The paper explored the representation of human faces in a lower dimensional subspace. In 1991, Turk and Pentland [5] used these eigenvectors (or eigenfaces as they are called) for face detection and identification.

Methods such as EFR work quite well provided the input test pattern is a face i.e., the face image has already been cropped and plucked out of a scene. The more general and difficult problem of recognizing faces in a cluttered background has also received some attention in [1, 5]. The authors in [1, 5] propose the use of distance from face space (DFFS) and distance in face space (DIFS) to detect and eliminate non-faces. We show with examples that DFFS and DIFS by themselves (in the absence of any information about the background) are not sufficient to discriminate against arbitrary background patterns. The traditional EFR technique either ends up missing faces or throws up many false alarms, depending on the threshold value. In this paper, we extend the EFR technique to solve the more general problem of robustly recognizing multiple faces in a given scene with background clutter. We explore the possibility of constructing a “background space” which will represent the background images corresponding to a given test image. If the background space is learnt well, it is our claim that patterns belonging to clutter will be closer to the background space than to the face space. This provides a basis for eliminating false alarms which would otherwise have crept in.

2. Effect of Background

The problem involving non-face test images is a difficult one and some attempts have been made to tackle it [1, 5]. In [5], the authors advocate the use of distance from face space to reject non-face images. If \( \hat{T}_i \) is the projection of the mean subtracted image pattern \( T_i \) in the face space, then \( \hat{T}_i \) can be expressed as

\[
\hat{T}_i = \sum_{n=1}^{L'} w_n^i u_n
\]

where \( w_n^i \) is the weight corresponding to eigenface \( u_n \), and \( L' \) is the number of eigenfaces used. The distance from face space (DFFS) is then defined as

\[
||\hat{T}_i - T_i||^2
\]
It has been pointed out in [5] that a threshold \( \theta_{DFSS} \) could be chosen such that it defines the maximum allowable distance from the face space. A test pattern is treated to be a face provided its DFFS value is less than \( \theta_{DFSS} \). In order to perform recognition, the difference error between the weight vector and the weight vector corresponding to every person in the training set is computed. This error is also called the distance in the face space (DIFS). The face class in the training set for which the DIFS is minimum is declared as the recognized face provided the difference error is less than an appropriately chosen threshold \( \theta_{DIFS} \).

However, it is difficult to conceive that by learning just the face class we can segregate any arbitrary background pattern against which the face patterns may appear. As we will show, it may not always be possible to come up with threshold values that will result in no false alarms and yet detect all faces. What would truly be desirable is to have a way of setting the threshold high, so that very few face images are rejected as unknown, while at the same time all incorrect classifications are detected. This is exactly what we attempt to do in this paper. We believe that some properties of the background scene local to a given image must be extracted and utilized for robust face recognition.

3. The Background Space

We argue in favor of learning the distribution of background images specific to a given scene. It is to be expected that background distribution will favor background images while the distribution of faces would favor the face patterns. In any given image, the number of background patterns usually far outnumbers the faces. To learn the distribution of the background, we need to generate sufficient number of observation samples from the given test image. We use simple thresholding to separate background patterns using the a priori statistical knowledge base of faces or the face space. Let \( \mu_1, \mu_2, \ldots, \mu_b \) be the mean values of the weights corresponding to each face class in the training set. Here \( q \) is the number of face classes or people in the training set. In the face space, let the weight vector of the test subimage \( \mathbf{z} \) be given by \( \mathbf{W} \). Then, the pattern \( \mathbf{z} \) is treated as a background image if the Euclidean distance of its weight vector from each of the class mean weights is greater than a predefined threshold \( \theta_b \) i.e.,

\[
\text{If } \| \mathbf{W} - \mu_i \| > \theta_b \quad \forall i, \ i = 1, \ldots, q
\]

then the image pattern is considered to be a non-face image. For high confidence, this threshold is chosen to be large enough. Sufficient number of background patterns can be obtained from the given test image in this manner. These patterns would represent a reasonable sampling of the background scene. The mean and covariance estimated from the samples obtained via (2) allow us to effectively extrapolate to other background patterns as well. A background image reconstructed with the eigenbackground images can be expected to have smaller error as compared to the case when it is reconstructed using eigenfaces.

We group the background patterns into \( K \) different clusters by the classical K-means algorithm where each cluster contains one pattern center. Each pattern center is treated to be representative of all the samples within its cluster. Thus, we can significantly reduce the number of background images that we have to deal with.

The pattern centers returned by the K-means algorithm are used as training images for learning the background space. Although the pattern centers belong to different clusters, they are not totally uncorrelated and further dimensionality reduction is possible. The procedure that we follow is similar to that used to create the face space. We first find the principal components (KL expansion) of the background pattern centers or the eigenvectors of the covariance matrix \( C_b \) of the set of background pattern centers. The space spanned by the eigenvectors corresponding to the largest \( K' \) eigenvalues of the covariance matrix \( C_b \) is called the background space. The significant eigenvectors of the matrix \( C_b \), which we call ‘eigenbackground images’, form a basis for the background image patterns.

4. Robust Face Recognition

In this section, we propose a robust face recognition scheme that finds faces by searching a given test image for patches of image patterns of faces embedded in a cluttered background and finally classification. Training data samples of image patterns of faces are first used to create the face space. Given a test image, the background is then learnt ‘on the fly’ and the background space corresponding to that test image is derived. Finally, the system classifies a subimage as being either a known face or as a background pattern by using the knowledge of both the face space and the background space.

Once face space and the background space are learnt, the test image is examined again, but now for the presence of faces at all points in the image. Let the subimage pattern under consideration in the test image be denoted as \( \mathbf{z} \). The vector \( \mathbf{z} \) is projected onto the face space as well as the background space to yield estimates of \( \mathbf{z} \) as \( \hat{\mathbf{z}}_f \) and \( \hat{\mathbf{z}}_b \), respectively. The test pattern \( \mathbf{z} \) is classified as belonging to the ‘face class’ if

\[
\| \mathbf{z} - \hat{\mathbf{z}}_f \|^2 < \| \mathbf{z} - \hat{\mathbf{z}}_b \|^2
\]

and

\[
\| \mathbf{z} - \hat{\mathbf{z}}_f \|^2 < \theta_{DFSS}
\]

where \( \theta_{DFSS} \) is an appropriately chosen threshold. Recognition of \( \mathbf{z} \) is then carried out as follows. The weight vector \( \mathbf{W} \) corresponding to pattern \( \mathbf{z} \) in the face space is compared...
with the pre-stored mean weights of each of the face classes. The pattern \( z \) is recognized as belonging to the \( i^{th} \) person if

\[
i = \min_j \| W - m_j \|^2, \quad j = 1, \ldots, q
\]

and

\[
\| W - m_i \|^2 < \theta_{DIFS}
\]

where \( q \) is the number of face classes or people in the database and \( \theta_{DIFS} \) is a suitably chosen threshold.

Since a background pattern will be better approximated by the eigenbackground images than by the eigenface images, it is to be expected that \( \| z - \hat{z}_j \|^2 \) would be less than \( \| z - \hat{z}_j \|^2 \) for a background pattern \( z \). On the other hand, if \( z \) is a face pattern, then it will be better represented by the face space than the background space. Thus learning the background space helps to reduce the false alarms considerably and imparts robustness to the EFR technique.

5. Experimental Results

In this section, we demonstrate the performance of the proposed scheme on two different datasets i) the standard Yale face database and ii) face database generated in our laboratory. The Yale database consists of 165 gray scale frontal images of 15 subjects. These are taken under different lighting conditions and facial expressions, and our intention is to test the proposed method under different conditions. For our experiments, we selected 15 individuals and 10 training images for each individual. The images were cropped to 33 \times 33\text{ pixel arrays}. The face space was constructed from this training set offline. After some experimentation, the number of significant eigenvectors was found to be 40 for satisfactory performance. The database created in our laboratory consists of images of 8 subjects with 10 images per subject. The face images were cropped to 21 \times 21\text{ pixel arrays} for training. The number of significant eigenfaces used to create the eigenface space for this database was chosen to be 20.

The system was first tested by artificially embedding images of some of the subjects from the Yale database at random locations in different test images of size 128 \times 128\text{ pixels} against a background scene that included trees, roads and building structures. The test image was scanned for the presence of faces at all points in the image. If a face pattern is found at any location in the test image, a white box is drawn at that location. For the second set of experiments, test images were captured in our laboratory and the subjects appear naturally in these real images. The background consisted of computers, furniture, etc. These images serve to represent real face recognition situations. A black box is drawn at the location where the system finds a face.

For the proposed method, the eigenbackground space was learnt ‘on the fly’ for each test image using the methodology discussed in Section 3. Thresholds \( \theta_{DIFS} \) and \( \theta_{DIFS} \) were chosen to be the maximum of all the DFFS and DIFS values, respectively, among the correctly recognized faces in the training set. The number of background pattern centers was chosen to be 600 while the number of eigenbackground images was chosen to be 150. The number of eigenbackground images was arrived at based on the accuracy of reconstruction of the background patterns.

Results corresponding to Yale database for the two methods are shown in Fig. 1. The figures are quite self-explanatory. The traditional EFR incurs many alarms when it attempts to recognize all the faces in the image. On the other hand, the proposed method detects all the faces without false alarms. Results obtained on real images captured in the laboratory are given in Figs. 2 - 3. Our method utilizes the background information quite effectively in order to discard non-face patterns, whereas the traditional EFR throws up false alarms.

6. Conclusion

In the literature, the eigenface technique has been demonstrated to be very useful for face recognition. However, when the scheme is directly extended to recognize faces embedded in background clutter, its performance degrades as it cannot satisfactorily discriminate against non-face patterns. In this paper, we have presented a robust scheme for recognizing multiple faces in still images of natural scenes against a cluttered background. We argue in favor of constructing a background space from the background images of a given scene. With moderate computational complexity, the scheme outperforms the traditional EFR technique and gives accurate recognition results on real images with almost no false alarms even on fairly complicated scenes.

References

Figure 1. (a) A test image with faces embedded in it. (b) Recognition results corresponding to traditional EFR using both DFFS and DIFS. Even though the faces are correctly recognized, there are a lot of false alarms in the upper right corner. (c) Output results for the proposed EFR method. There are no false alarms and both the faces are correctly recognized.

Figure 2. (a) A real test image where a person appears naturally against a cluttered scene. (b) Face recognition results for the traditional EFR technique using both DFFS and DIFS. (c) Recognition results with proposed method.

Figure 3. (a) Test image consisting of desks and computers as background clutter. Recognition results for (b) traditional EFR, and (c) proposed method. Note that traditional EFR throws up many false alarms.