Robust face recognition using an ensemble of different methods

Aria Khoshsirat

Introduction

Automatic recognition of humans is a challenging problem which has received much attention during recent years due to its many applications in different fields. Face recognition is one of those challenging problems and to this date, there are no techniques that provide a robust solution for all situations and different applications that face recognition may be used for. The need for an automatic face recognition system especially at the border control, like airports, is becoming very important to strengthen the security of a country.

Generally, feature extraction and classification are the two basic operations of any face recognition system. As a result, to improve the recognition performance of such systems, one has to enhance the accuracy of these operations. In many applications (Especially those related to defense and security), the face recognition task must be robust to missclassifications, meaning it's better to take time and provide a better answer, than to immediately provide (with high risk of error) a classification model from which a fallacious fatal decision would be drawn.

Ensembles are a great way to improve robustness. Using ensembles, we can be sure that the final definite answer is the result of different models (with different rates of error on different data) and can be trusted more than just using a single model. Although many different methods have been published for the task of face recognition, none of them can overcome the others in all types of environments and situations. For instance, one model could use facial accessories' information better to identify a face, while another model could rely on the intensities of different patches.

Using an ensemble of five different methods for face recognition, we've created a very accurate and robust face recognition model which can be used in situations when the misclassification of a face is a dramatic error and robustness is required. The following is a brief summary on the 5 techniques we've used for this project.

Different Methods of Face Recognition

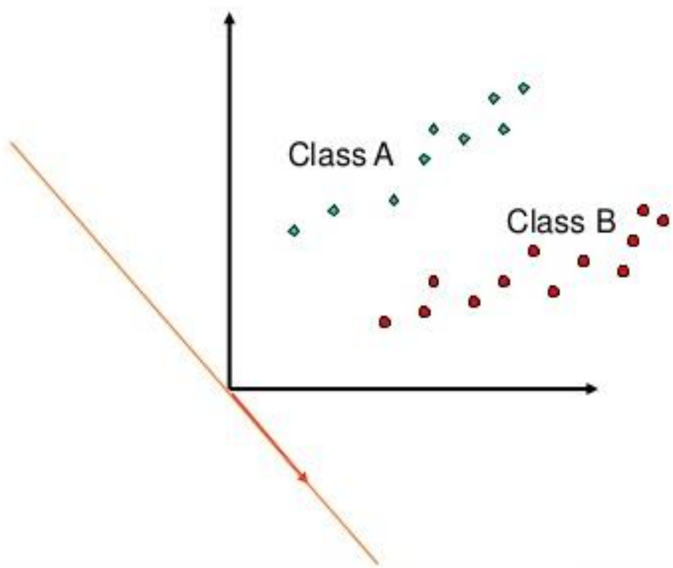
Eigenface (PCA)

The approach of using eigenfaces for recognition was developed by Sirovich and Kirby (1987) and used by Matthew Turk and Alex Pentland in face classification. The eigenvectors are derived from the covariance matrix of the probability distribution over the high-dimensional vector space of face images. The eigenfaces themselves form a basis set of all images used to construct the covariance matrix. To construct the covariance matrix, each face image is transformed into a vector. Each element of the vector corresponds to the pixel intensity. This transformation of the pixel matrix destroys the geometric structure of the image.

Given an s-dimensional vector representation of each face in a training set of images, Principal Component Analysis (PCA) tends to find a t-dimensional subspace whose basis vectors correspond to the maximum variance direction in the original image space. This new subspace is normally lower dimensional ($t \ll s$). If the image elements are considered as random variables, the PCA basis vectors are defined as eigenvectors of the scatter matrix. Classification can be achieved by comparing how faces are represented by the basis set.

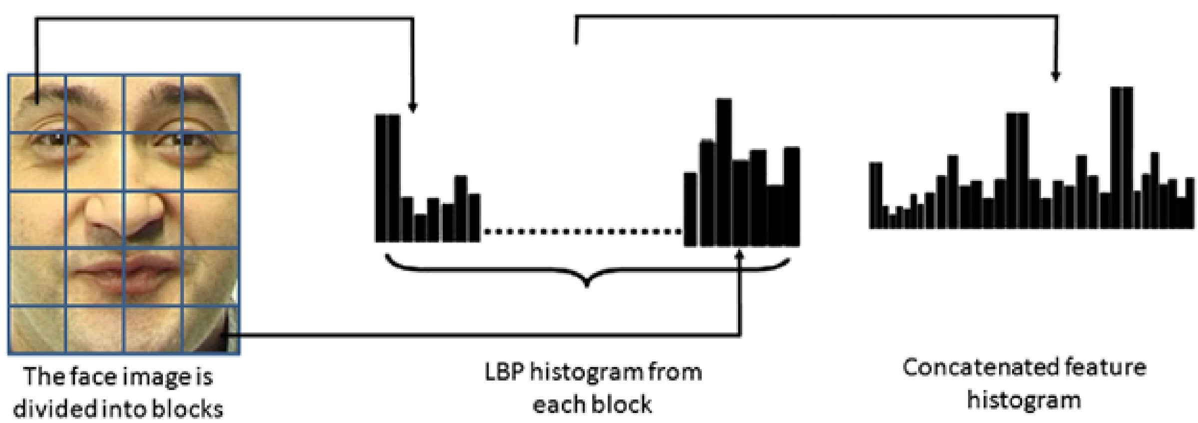
Fisherface (LDA)

Linear Discriminant Analysis (LDA) an idea suggested by R.A. Fisher in 1936, is used to find the subspace representation of a set of face images, the resulting basis vectors defining that space are known as Fisherfaces. This is another technique that is also constructed from linear decomposition. While PCA builds a subspace to represent, in an optimal way, “only” the object “face”, LDA constructs a discriminant subspace to distinguish, in an optimal way, the faces of different people. LDA, also called “Fisher linear discriminant” analysis, is one of the most widely-used approaches for face recognition. It uses the reduction criterion based on the concept of the separability of data per class. LDA includes two stages: the original space reduction by the PCA and the vectors of the final projection space, called “Fisher faces”. The latter are calculated on the basis of the classes’ separability criterion, but in the reduced space. This need for the input space reduction is caused by the total scattering matrix singularity criterion of the LDA approach. Comparative studies show that methods based on the LDA usually give better results than those based on PCA.



Local Binary Patterns (LBP)

There exist several methods for extracting the most useful features from face images to perform face recognition. One of these feature extraction methods is the Local Binary Pattern (LBP) method. This relative new approach was introduced in 1996 by Ojala et al. With LBP it is possible to describe the texture and shape of a digital image. The procedure consists of using the texture descriptor to build several local descriptions of the face and combining them into a global description. These features consist of binary patterns that describe the surroundings of pixels in the regions. Instead of striving for a holistic description, this approach was motivated by two reasons: the local feature-based or hybrid approaches to face recognition have been gaining interest lately. These local feature-based and hybrid methods seem to be more robust against variations in pose or illumination than holistic methods. The facial image is divided into local regions and texture descriptors (histograms) are extracted from each region independently. The descriptors are then concatenated to form a global description of the face. This feature vector forms an efficient representation of the face and is used to measure similarities between images. According to several studies face recognition using the LBP method provides very good results, both in terms of speed and discrimination performance. Because of the way the texture and shape of images is described, the method seems to be quite robust against face images with different facial expressions, different lightening conditions, image rotation and aging of persons.

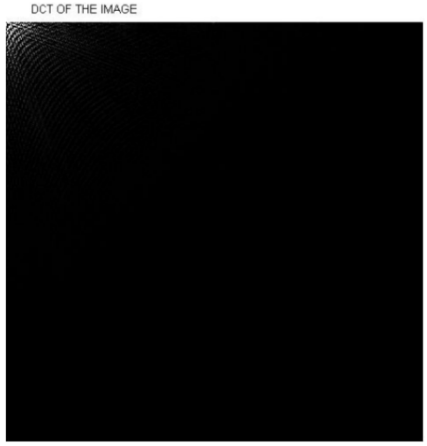


Discrete Cosine Transform (DCT)

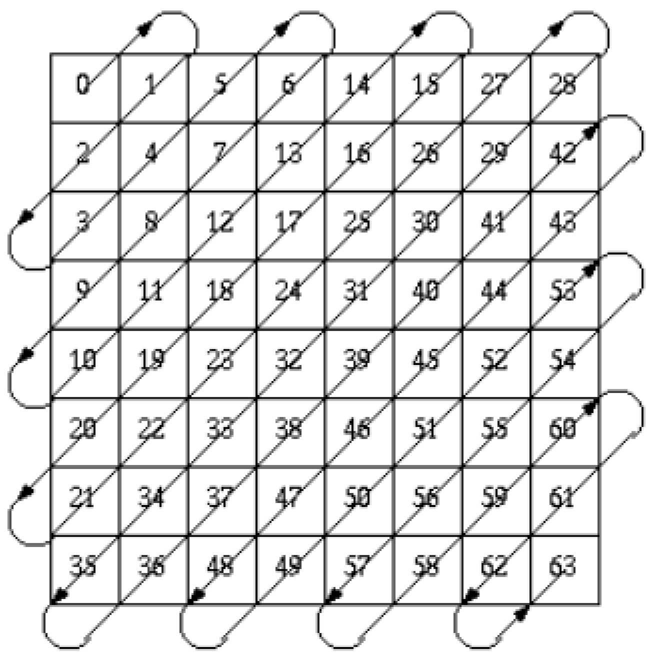
A discrete cosine transform (DCT) expresses a finite sequence of data points in terms of a sum of cosine functions oscillating at different frequencies. DCTs are important to numerous applications in science and engineering, from lossy compression of audio (e.g. MP3) and images (e.g. JPEG) (where small high frequency components can be discarded), to spectral methods for the numerical solution of partial differential equations. In particular, a DCT is a Fourier-related transform similar to the discrete Fourier transform (DFT), but using only real numbers. The DCTs are generally related to Fourier Series coefficients of a periodically and symmetrically extended sequence whereas DFTs are related to Fourier Series coefficients of a periodically extended sequence.

To obtain the feature vector representing a face, its DCT is computed, and only a subset of the obtained coefficients is retained. The size of this subset is chosen such that it can sufficiently represent a face. The DCT of the image has the same size as the original image. But the coefficients with large magnitude are mainly located in the upper left corner of the DCT matrix. Low frequency coefficients are related to illumination variation and smooth regions (like forehead cheek etc.) of the face. High frequency coefficients represent noise and detailed information about the edges in the image. The mid frequency region coefficients represent the general structure of the face in the image. Hence we can’t ignore all the low frequency components for achieving illumination invariance and also we can’t truncate all the high frequency components for removing noise as they are responsible for edges and finer details.

DCT OF THE IMAGE

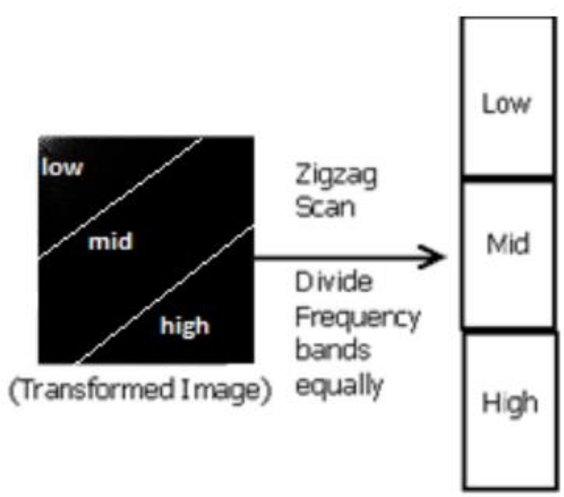


In our approach, We take DCT of the image. Here our image size is 200 x 200. Next we convert the DCT of the image into a one dimensional vector by zigzag scanning. We do a zigzag scanning so that in the vector the components are arranged according to increasing value of frequency.



Zig-zag Scanning

Now we divide the whole range of frequency into three equal sections and derive the coefficient of feature vector from each section. In case of low frequency section we reject the 1st three terms and consider the next 2000 terms. We reject the 1st three terms to achieve illumination invariance.



In case of mid and high frequency section we obtain the coefficient from those position and include them in the feature vector. Here in our case we are considering 2000 coefficient from each section. So for each image we have obtained a feature vector of size 6000. This feature vector contains all the good information that DCT provides for face recognition and is quite less in size than the original 200*200 feature matrix. For recognizing a new face, we compute the euclidean distance between each training face's DCT feature vector and the new face's, and find the min distance. Then the new face is classified in the group (person) in which the training face belongs to.

Nearest centroid classifier

Briefly, the method computes a standardized centroid for each person. Nearest centroid classification takes the face profile of a new sample, and compares it to each of these class centroids. The class whose centroid that it is closest to, in euclidean distance, is the predicted class for that new face image. This method is in fact quite good for our application, since we usually take up to 50 images per person for the training phase and each image is taken after a 0.3 second interval. This allows individuals to have different facial expression and orientation (size & rotation). Therefore computing the mean of the faces of one individual can very well provide the necessary information to present that person.

Capturing and detecting faces

Capturing images is done via webcam (possibly a Raspberry Pi). The process is done in less than a minute so the new person who wants to donate his/her face to the app doesn't have to wait for long. The algorithm takes up to 50 images of a person and then detect their faces from those images. Face detection is done using the frontal face Haar cascade to identify where the face is in an image. The detected faces are then saved to a folder with the name of the new individual. This name is then used for face recognition. To improve accuracy and lessening the effect of illumination each face is then recieves the histogram equalization treatment for later use in the face recognition phase.

Ensemble and Face Recognition

As mentioned above, our approach is an ensemble of five different face recognition methods. These methods are : Eigenface, Fisherface, LBP, DCT and NCC which are mostly different from each other, so they can be good elements of an ensemble to improve the final model's robustness. This method includes classifying a face with each of these five different methods and then saving the models' answers. Afterwards, by using a majority vote algorithm we decide which answer is the final decision of the whole ensemble. To improve robustness we ignore the major votes which are less or equal to 2, meaning even if two different methods classifies a test sample to a class but the rest of the models (3 others) classify it to different classes, the answer cannot be trusted and should be ignored. The minimum major vote could be set to 4 for even more robustness if required, although it requires more testing time for the models to recognize the new face. The final classification of the model is done whenever the major vote of the ensemble classifies the new face to a class for 30 times, unsequentially. This is also done to improve the final answer's accuracy and robustness. It's good to note that although 30 times of testing sounds like a long time, the testing phase is actually quite fast and capturing and recognizing a new face is done in less than a second.

References

Acharya, Tilak and Sahoo, Sunit Kumar (2013) *Face Recognition Using PCA and DCT Based Approach.* BTech thesis.

Face Recognition by Independent Component Analysis
Marian Stewart Bartlett, Member, IEEE, Javier R. Movellan, Member, IEEE, and Terrence J. Sejnowski, Fellow, IEEE

A Survey of 2D Face Recognition Techniques
Mejda Chihaoui, Akram Elkefi, Wajdi Bellil, Chokri Ben Amar – Computers – 2016

Face Description with Local Binary Patterns: Application to Face Recognition
T. Ahonen ; A. Hadid ; M. Pietikainen
IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 28, Issue: 12, Dec. 2006)

Face Recognition Using the Discrete Cosine Transform
Ziad M. Hafed - Martin D. Levine
International Journal of Computer Vision - July 2001, Volume 43, Issue 3, pp 167–188

Face Image Analysis by Unsupervised Learning
Authors: Bartlett, Marian Stewart - 2001