Facial Recognition Using PCA

David Huie
Harvey Mudd College

Dustin Rodrigues and **Alex Ruch**
Pomona College

Abstract

Facial recognition is an important area of research of computer science with obvious implications. As one would imagine, it is not a simple task. A large number of variables come into play including facial expressions, hairstyles, camera angle, and lighting. Our paper in particular examines Principal Component Analysis (PCA). This method of facial recogniiton is sensitive to variations of the images, so we used face databases that attempted to minimizae these differences. Our results were FILL IN RESULTS.

Introduction

Facial recognition is a varied area of study with applications to security, biometrics, and personal use. In terms of security, one could imagine a situation where an authority has a database of images of faces and security footage of a criminal. It is rarely feasible to manually comb through a database, but a sophisitcated enough facial recogniiton system could possibly determine the identity of the criminal. In biometrics, using a camera to confirm a subject's identity is cheaper than a fingerprint reader or an iris scanner, and is more convenient than having to manually enter a password. The personal uses of facial recognition include training some images of a photo album based on user-based tags and then automatically tagging any new photos with known people.

However, all of these applications have clear deficiencies, most of which come from variations of facial expressions, hairstyles, camera angles, and lighting. These factors contribute to a significant amount of inaccuracy in facial recognition. For personal uses such photo tagging, a wrongly tagged photo is not a very large issue. However, if an authorized user cannot log into their own system because they recently got a haircut, or even worse, an unauthorized user was able to log in, this is a significant security deficiency.

One of the challenges of facial recognition is the extraordinarly large space of possible images. For instance, a 100×100 image will have $10\,000$ pixels. A naïve method would be to take an image we want classified and compare it against known images using some notion of distance. However, given the aforementioned size of the data, this will be a very computationally intensive task. Instead, Principal Component Analysis (PCA) is used to extract the features with the most amount of variance. Using PCA, one can reduce the space from a dimension of $10\,000$ to something much more managable, such as 10 or 20 dimensions. Once these principal components are found, we can project an image to be classified onto the principal components and determine which face it most closely resembles.

Algorithm

The PCA Algorithm begins by taking a two-dimensional matrix representing the image and flattening it into a long one-dimensional vector. Each vector is then mean-centered. The algorithm then finds the set of orthonormal vectors which has the largest projection onto the mean-centered vectors. These vectors are given by the eigenvectors of the covariance matrix of WW^T, where W is the matrix of the column vectors of the mean-centered images placed next to each other. The eigenvectors corresponding to the largest eigenvalues will produce the greatest variance. To classify an image, we project the image onto a selected number of eigenvectors with the largest eigenvalues and then find the distance (Euclidean, Manhattan, or otherwise) to each face class to see which one most closely resembles the image.

Results

Conclusion