Facial Recognition using Singular Value Decomposition

Leif Wesche

1-23-2018

Abstract:

Two data sets consisting of images of different subjects faces were analyzed and recognized using Singular Value Decomposition with the goal of creating an algorithm that would be able to distinguish between the subjects. The singular value spectrum, obtained by organizing all of the images in a data set into a single large matrix and decomposing the matrix into U, Σ, and V matrices, was analyzed to interpret the range of modes constructed to fit the data. The appropriate range of modes was selected based on which ranges were found to describe the specific features of each face. Next, the average face for each subject was calculated and each of these faces was projected onto the U matrix to form a "key" for each person. These keys were compared to the projection of any image across the appropriate range to determine which key an image most closely fit, and their corresponding identity.

I. Introduction and Overview

The purpose of this homework was to explore the ways SVD decomposition can be used to analyze a set of data. In this assignment, the data consisted of two sets of images taken at Yale University and provided to us. The first set consisted of photos of 38 different peoples faces taken under various lighting conditions, all pre-cropped and aligned. The second set consisted of un-cropped and unaligned photos of 15 subjects faces with various facial expressions. The SVD decomposition was used to analyze the data in terms of the singular value spectrum and the associated modes, reconstruct the photos, and eventually develop an algorithm which was used to recognize subject's faces.

II. Theoretical Background

Singular Value Decomposition can be thought of as representing a set of data in terms of it's most ideal coordinate systems which aim to minimize redundancy. SVD is a decomposition technique which is guaranteed to exist for any matrix. It separates a matrix A into three subsequent matrices, U, Σ, and V. The U and V matrices are two orthonormal bases which can be thought of as rotational" matrices acting on Σ. The U matrix corresponds to the columns of the decomposed matrix A, while the V matrix corresponds to the rows. The U and V matrices are also unitary, which ensures that their transpose is equal to their inverse. Their inverse and transpose are often represented by the same variable, U^* and V^* respectively. The matrix Σ can be thought of the "stretch factor" of the data set; it is a square diagonal matrix containing the singular values of the matrix A. These singular values are closely related to the eigen values of matrix A. The matrix A can be written in terms of the three matrices U, Σ, and V as shown below in Equation 1.

$$A = U \times \Sigma \times V^* \tag{1}$$

The Pythagoearn norm was also used as part of an algorithm to distinguish between faces in the group. The Pythagorean theorem was to measure the Euclidian distance between two points in n-th dimensional space the same way the distance would be measured in normal three dimensional space. The difference was taken between each component, those differences were squared, each component was summed, and the square root of the sum was calculated.

III. Algorithm Implementation and Development

The assignment was broken up into 10 sections. The first section loaded the greyscale image data into Matlab in the form of one matrix of double precision numbers for each image. The images were reduced in resolution to ensure that the following sections would run smoothly on my laptop. In both the cropped and un-cropped data sets, the images were organized into cell arrays. Each cell contained every photo taken of a single person. There were 38 people in the

un-cropped data set and 64 photos were taken of each person, reduced to a resolution of 96x84. There were 15 people in the un-cropped data set and 11 photos of each person, reduced to a resolution of 75x50.

In Section 2, each image was reshaped into a single column vector, and each of these column vectors were compiled row-wise into a large database matrix D. For the cropped image data set the size of the matrix D was 8064x2432, with 8064 denoting the total number of pixels in each image and 2432 denoting the total number of images. In Section 3, this large matrix D was decomposed into U, Σ, and V matrices using Matlab's built in "svd" command.

In Sections 4 and 5, the singular value spectrum was plotted along with several mode faces corresponding to specific singular values. The Σ matrix consists of singular values along the diagonal and zeroes everywhere else, with the singular values organized from highest to lowest along the diagonal. For each data set, the number of singular values generated was equal to the number of photos in the set. So for the cropped data set, there were 2432 images and 2432 singular values generated. Each singular value has a corresponding mode. These modes are the columns of the matrix U. The matrix U was of interest, not the matrix V, because of the fact that in Section 2 the image data was organized into columns, not rows. Because of the column orientation of the images, the matrix U contains mode shapes that are related to the images, and the matrix V contains mode shapes that are not significant in this case. The relative size of each singular value is of interest because a larger singular value means the corresponding mode face describes a larger amount of the data.

The first 20 largest singular values were normalized by the sum of all the values and plotted as a scatter plot. This was done to show what percentage of the entire database of faces was made up of the largest modes corresponding the largest singular values. Next, the entire singular value spectrum normalized by the sum of the spectrum was plotted on a logarithmic axes to get a picture of the range of singular values. These plots are shown below in Figures 1a and 1b for the cropped photo data.

Several of the mode faces were plotted and compared. Images were constructed by re-sizing the columns of the U matrix. Mode faces corresponding to singular values 1, 20, 50, and 500 are shown below in Figure 1c, 1d, 1e, and 1f respectively. The mode faces across the spectrum for both the cropped and un-cropped images were analyzed to determine how the modes and their significance to the data set changed across the spectrum.

In Section 6, the picture database was reconstructed by a reduced number of modes and singular values. This was done to explore how the number of modes used effected the quality of image reconstruction. The database of images was reconstructed using N nodes by reducing the U and V matrices down to the N desired rows, and reducing the diagonal Σ matrix down to the desired NxN size. These new matrices, U_N, V_N, and Σ_N were then manipulated as shown below to produce a new database matrix D_N which was the same size as the original database matrix D.

$$D_N = U_N \times \Sigma_N \times V_N^* \tag{2}$$

The database was reconstructed using N=10, 20, 100, 500, and 1000 modes. Figure 2 shows the original Image 1 of Person 1, along with the same reconstructed using the various number of nodes. These images were plotted to show how the detail and quality of the image reconstruction improved as the number of modes increased, and at what point the quality improvement began to drop off.

The average faces of each person was calculated in Section 7. This was done by adding each of the images of the same person together and then dividing by the number of photos added. For future convenience these average face images were reshaped into column vector form and stored in a cell array.

In Section 8, each of the average faces were projected onto the orthogonal basis U by multiplying the average faces in vector form with U. This was done to develop a "key" corresponding to each average face. Next in Section 9, each of the images were projected onto the U matrix individually. Figure 3 shows the first picture of the first face in the cropped data set along with its projection onto U, followed by the average picture of the first person and its projection.

Finally, Section 10 consists of an algorithm that identifies which person is in any photo in either the cropped or un-cropped data sets. It works by selecting a photo to test, then projecting that photo onto the U matrix. This projection is compared to each of the "keys" generated for the average faces by subtracting the key from the projection, then by calculating the Pythagorean length of the resulting vectors.

Only specific sections of the "key" and the test photo's projection were used to calculate calculate the specific length. Only the center parts of each vector was used to calculate the Pythagorean length. The early section, about the first 20 values, were left out since those images represented the most prevalent features among all of the faces. When trying to distinguish between faces, it proved most effective to classify faces according to the differences in the faces that set them apart from each other. The early parts of the key and test projection, corresponding to the most common mode shapes among the set, would not be productive to compare since they represent the features that are most common among all of the faces, not features that are specific to some faces. The ends of the key and projection were also left out since around halfway through the data, the singular values get extremely small. This means that the ends of the keys and projections represent modes that are insignificantly, representing details that are closer to random noise than specific features.

The key that was found to be the "closest" to the photos projection onto the U matrix, or in other words returned the minimum Pythagorean length, was said to be the best match. The second and third best matches were also recorded. Figure 4 shows an example input and output of the algorithm.

IV. Computational Results

The first 20 singular values of the cropped data were normalized by the sum of the singular values and plotted, as shown in Figure 1a below. The highest of these values reached just over 12% of the total sum of singular values, while soon after the spectrum quickly dropped off to around the 1% mark. Figure 1 shows the entire singular value spectrum plotted on a log scale. The plot shows the singular values drop of quickly at the start, followed by a middle range of about 500 values values around the 1-0.1%, trailed by about 2000 values that make up less than 0.01% of the data set. Similar trends were seen in the un-cropped images, with the highest value mode composing over 30% of the data, a mid range of about 100 values composing 1-0.1% of the data set, and a about 50 modes following that that drop quite suddenly to extremely low values, approaching machine precision. When trying to distinguish between different faces, the mid range of values in both sets will prove most useful in comparing the different faces and features.

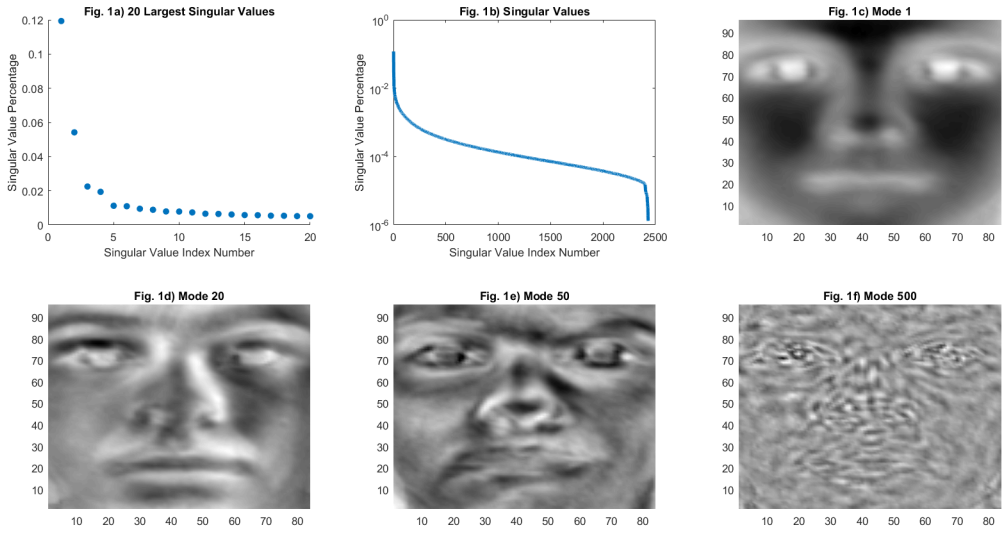

Figure 1

Figure 1c shows the first mode, corresponding to the largest singular value and reconstructed by organizing the first column of the U matrix into a 96x84 matrix. In the image shows a basic shape of a mouth, two eyes, and a nose, but lacks recognizable features. Figures 1d, 1e, and 1f shows subsequent modes, or faces corresponding to lower singular values. Figures 1d and 1e show faces in the "middle range" of the spectrum. In these images, the building blocks of distinguishing features begin to emerge in the faces. Figure 1f shows a mode near the low end of the singular value spectrum. This image only vaguely looks like a face. By this point in the range, the modes represent more noise than data pertaining to distinguishing features in specific faces.

The image database matrix D was reconstructed using various numbers of modes. Figure 2a below shows the first image of the first person, and the following sub-figures show the same photo reconstructed from N=10, 20, 100, 500, and 1000 modes respectively.

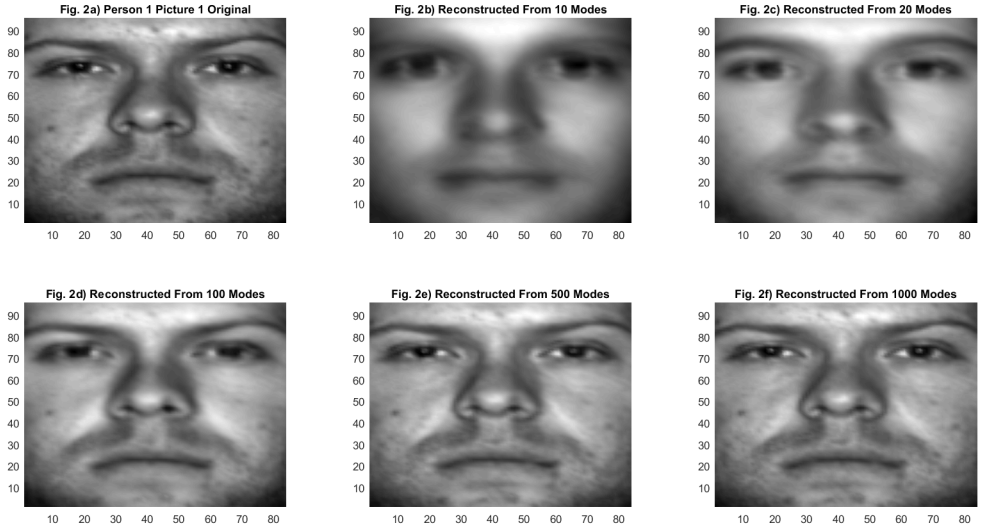

Figure 2

In Figures 2b and 2c, while the reconstructed images is clearly a face in both cases, the two faces don't including many distinct features yet. Using around 100 modes to reconstruct the image, seen in Figure 2d, seems to be about the point where there is enough visible detail in the reconstructed face to match it with the original by eye. The image reconstructed with 500 modes, as seen in Figure 2e, looks about identical to the original image, and the image reconstructed with 1000 modes, shown in Figure 2e, looks about the same. This results agrees with the results observed from the singular value spectrum shown in Figure 1b. The first, most prevalent modes, seems to construct the basic outline of the face, while the mid-range modes add details, and the modes at the end of the spectrum do little to influence the reconstructed photos.

After the average faces for each person was calculated, they were projected onto the U matrix, and this average face projection was used as a "key" for each persons face. Figure 3b below shows the average face of person 1, and Figure 3d below shows the first 20 values of the projection associated with that image.

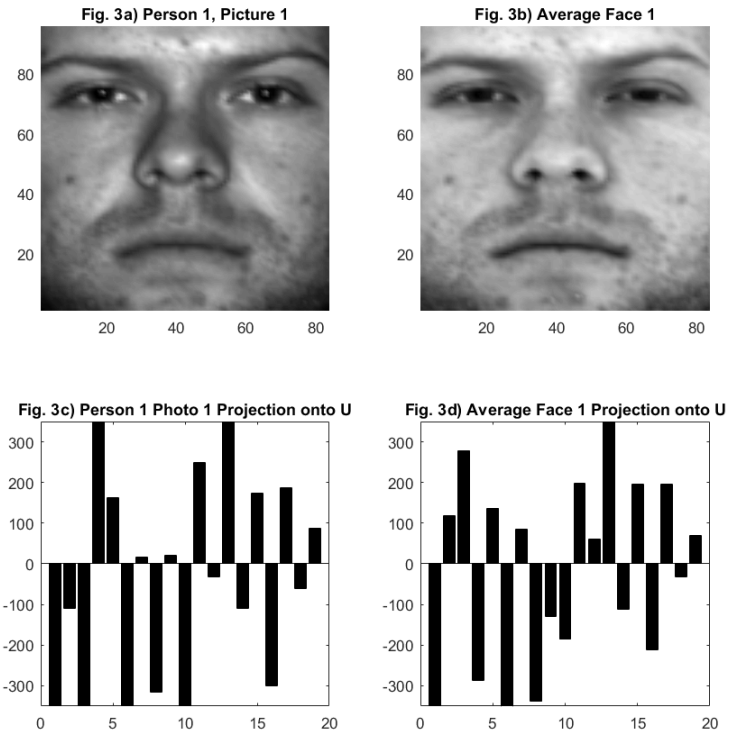

Figure 3

Each individual image in both data sets was projected onto their respective U matrices as well. Figures 3a and 3b below show the first photo of person one, along with the first 20 values of that photos projection onto the U matrix.

In Section 10, an algorithm was used to compare any one of the images with the set of average faces. The algorithm worked by calculating the Euclidean distance between points in the average faces "keys", an example of which is shown in Figure 3d, and corresponding points in the image under test's projection onto the U matrix, an example of which is shown in Figure 3c. The "key" that was found to be closest to the test photos projection was said to be the best match. The algorithm output the best three matches for each test.

But not every value in the key and the projection were compared. In fact, only the mid-range points were compared, and the first points and end points were completely ignored by the recognition algorithm. Only the mid points, ranging from the 50th value to the 250th value in the cropped images and from about the 10th to 100th value in the un-cropped images, were used because these were the values corresponding to singular values which contributed to the unique features of each face. Values before these in the spectrum corresponded to mode shapes which had too much in common with each face, and values after these corresponded to mostly noise. These ranges were determined by comparing facial reconstructions similar to the ones shown in Figure 2 to see at which ranges the faces began gaining more unique features and at what point adding modes to the reconstruction had little effect.

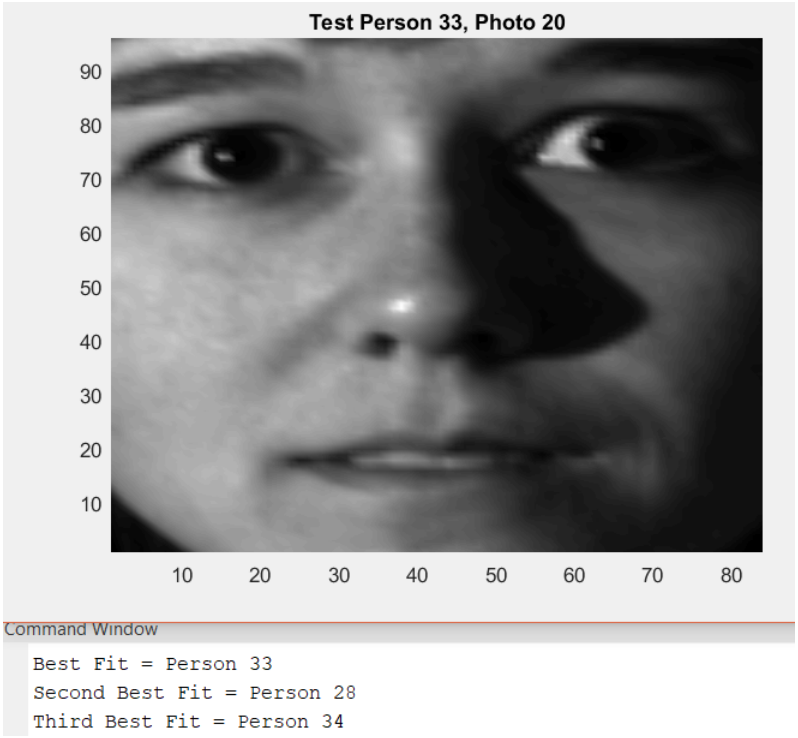

Figure 4

Figure 4 above shows an example output of the test algorithm. It was found that by comparing only the middle range of the "keys" with test face projections, the algorithm returned by far the best results in both the cropped and un-cropped image data sets. The algorithm was much more likely to return a false Best Fit if the first or last values, corresponding the the most common and least common mode shapes, were taken into account.

The algorithm was found to work extremely well for the cropped images. Using a well-lit photo like the first photo of every person, an example of which is Figure 3a, the algorithm worked for every person. It worked just as well using even a half-lit photo, such as the one shown in Figure 4. The only time the algorithm made any mistakes was in recognizing some of the darkest photos in the set, but even those were correct most of the time. For all of the 38 subjects, the algorithm was able to recognize the well lit photos for 38/38 subjects, and rarely made mistakes with the darker photos.

Surprisingly enough, the algorithm was able to recognize the un-cropped images nearly as well as the cropped images. When using a well-lit straight photo with a black facial expression, such as the first image, the algorithm was able to identify 14/15 people successfully. It very well with other photos in the set too, as the varying facial expressions and lighting of the subjects seemed to have little effect on the output. For more than half of the 15 subjects in the un-cropped data set, the algorithm was able to identify each photo completely accurately.

V. Summary and Conclusions

The SVD decomposition was used to separate image data sets into three unique matrices, the U, Σ, and V matrices. The singular value spectrum was analyzed for each database along with the values corresponding mode faces to determine the significance of each portion of the spectrum. The faces were reconstructed from varying numbers of modes to see how many modes were needed for accurate facial reconstruction and how modes from different points in the spectrum contributed to the reconstructions. The average face of each person in the data was computed and projected onto the U bases to form a "key", or unique average projection, for each person. Finally, a facial recognition algorithm using a optimum range of the spectrum was developed by comparing each "key" in the database with any photo. The algorithm was tested, and found to work successfully with well-lit photos for 38/38 subjects in the cropped data set and 14/15 subjects in the un-cropped data set.