Singular Value Decomposition for Analysis of Images and Audio

Ashley Batchelor

Abstract

I used Singular Value Decomposition (SVD) to analyze two collections of facial images, and to prepare audio data for machine learning algorithms to identify music genres and artists.

Sec. I. Introduction and Overview

For the first part of this experiment, I analyzed two sets of faces from the Yale Face Database. The first set included 2,414 images of human faces which were cropped to roughly the same size and position for each face. The second set included 167 images of human faces which were uncropped. I used SVD on each set of images. I then compared a sample image with an increasing number of principal modes.

For the second part of this experiment, I analyzed music from three genres: Jazz, Punk, and Baroque, including five artists of each genre. I used samples of songs to train machine learning models and tested the models on other samples.

Sec. II. Theoretical Background

Singular Value Decomposition separates any mxn matrix A (containing data sampled columns) into a product A= $U\Sigma V^*$. The U matrix is an mxn matrix which represents the principal axes of the features in the images. The V matrix is a unitary nxn matrix, where each column represents the principal modes which determine the weight of each feature in an image. The Σ matrix is an nxn matrix with the diagonal elements representing the amplitudes of each of those modes arranged in descending order of magnitude. Data such as images or audio may be approximated by selecting a subset of principal modes which have a large percentage of the energy of the data as defined by the diagonal elements of the Σ matrix.

Sec. III. Algorithm Implementation and Development

Part1

For each set of faces, I converted the image size to 192x168 pixels. I then turned each image into a vector of the rows in series. I stored the vectors as rows of an array. I used this array for SVD. I then selected several different numbers of modes from the Σ matrix (setting the higher modes to zero) and computed a matrix ff as a product of the three matrices U, Σ, and V. From the matrix ff, I reconstructed an arbitrarily selected image by reshaping a single row of data to a matrix.

Part 2

For each of the three sections, I used monaural WAV files sampled at 8000 Hz. I randomly selected 5 second clips of audio, and from each clip I generated a spectrogram using a Gabor filter with a Gaussian kernel, with a Gaussian width of 100 s^{-2}, and a step increment of 0.1 s. I reconstructed the spectrograms as vectors, and stored them in an array. I performed SVD on this array. I then used the V matrix from

the array as training and test data for machine learning using a Naïve Bayes model and a multiclass ECOC model.

a)

For this section, I selected three artists from three genres, Charles Mingus, Siouxsie and the Banshees, and J.S. Bach. From each artist, I used audio data from five songs stored in a single mat file to generate 30 clips. I then used machine learning to identify the artists for 20 trials of training and test data for cross validation.

b)

For this section, I selected three artists from the same genre (punk): The Ramones, Siouxsie and the Banshees, and Joy Division. From each artist, I used audio data from five songs stored in a single mat file to generate 30 clips. I then used machine learning to identify the artists for 20 trials of training and test data for cross validation.

c)

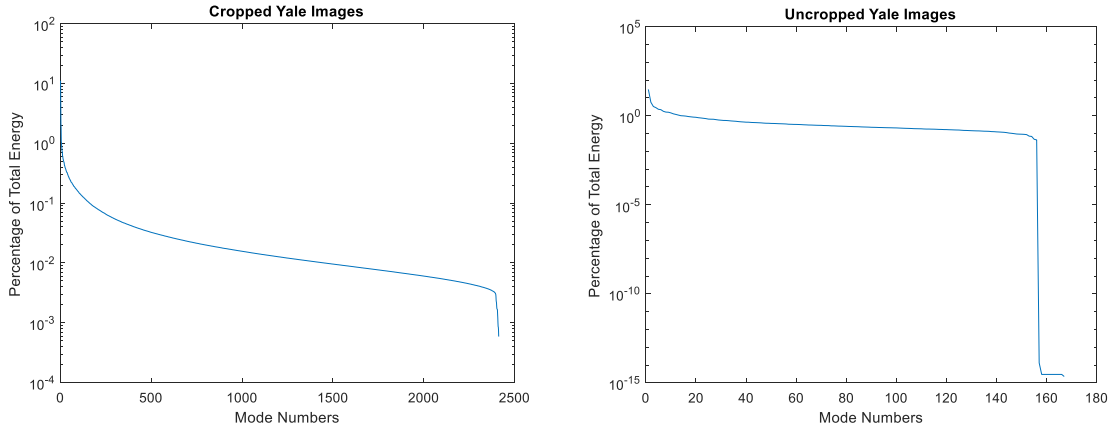
For this section, I selected five artists from three genres Jazz, Punk, and Baroque (artists and songs listed in appendix). From each artist I used one song to generate three mat data files grouped according to genre. From each mat file I generated 30 clips of audio. I then used machine learning to identify the genres for 20 trials of training and test data for cross validation.

Sec IV. Computational Results

Part 1

The cropped faces required approximately 200 modes in order to construct a recognizable image, and approximately 600 modes in order to construct a high-quality image (79.9 % of the total energy).

The uncropped faces required approximately 100 modes in order to construct a recognizable image, and approximately 140 modes in order to construct a high-quality image (98.5 % of the total energy).



Figures 2 a) and b), Percentage of energy of each mode number of the SVD of the cropped Yale Images and the uncropped Yale images.

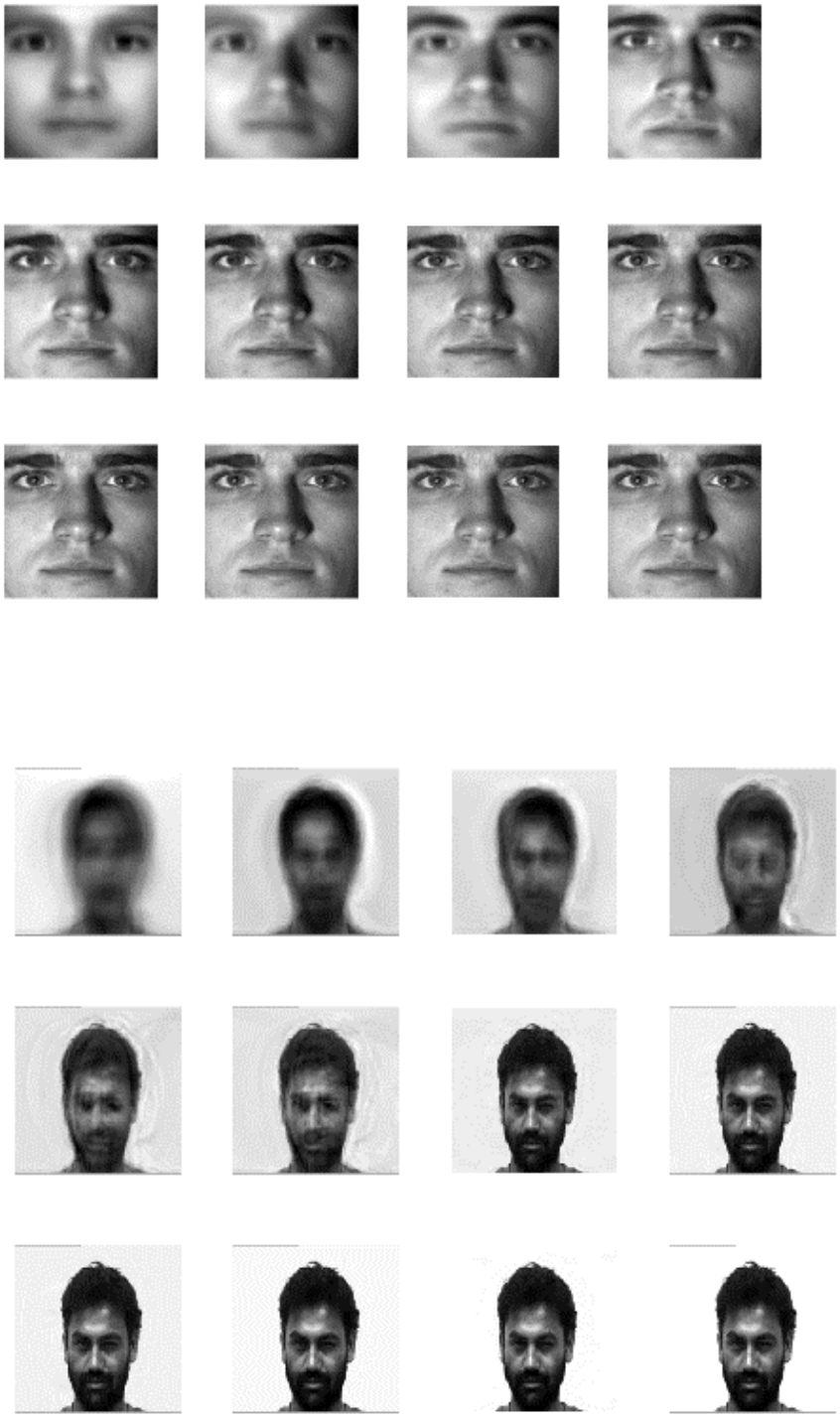


Figure 2 - Cropped Yale image with 1, 5, 10, 100, 200, 300, 500, 600, 700, 800, 900, and 1000 modes. Left to right, top to bottom.

Figure 3 - Uncropped Yale image with 1, 5, 10, 25, 50, 75, 100, 110, 120, 130, 140, and 150 modes. Left to right, top to bottom.

Part 2

For the Naïve Bayes model, part a) resulted in a total accuracy of 57.7 %. Part b) resulted in a total accuracy of 65.3 %. Part c) resulted in a total accuracy of 59.3 %. The confusion matrices for each section are shown in Figures 4 a), b), and c).

		Predicted		
Actual		Mingus	Siouxsie	Bach
	Mingus	36	26	38
	Siouxsie	15	70	15
	Bach	15	18	67

		Predicted		
Actual		Ramones	Siouxsie	Joy Division
	Ramones	91	5	4
	Siouxsie	35	30	35
	Joy Division	15	10	75

		Predicted		
Actual		Jazz	Punk	Baroque
	Jazz	18	35	47
	Punk	5	93	2
	Baroque	15	18	67

Figures 4 a), b), and c) - Confusion matrices for each of the three sections of Part 2 using a Naïve Bayes model.

For the multiclass ECOC model, part a) resulted in a total accuracy of 70.0 %. Part b) resulted in a total accuracy of 70.0 %. Part c) resulted in a total accuracy of 49.7 %. The confusion matrices for each section are shown in Figures 5 a), b), and c).

		Predicted		
Actual		Mingus	Siouxsie	Bach
	Mingus	64	22	14
	Siouxsie	21	72	7
	Bach	4	22	74

		Predicted		
Actual		Ramones	Siouxsie	Joy Division
	Ramones	59	18	23
	Siouxsie	15	69	16
	Joy Division	15	3	82

		Predicted		
Actual		Jazz	Punk	Baroque
	Jazz	54	2	44
	Punk	34	30	36
	Baroque	32	3	65

Figures 5 a), b), and c) - Confusion matrices for each of the three sections of Part 2 using a multiclass ECOC model.

Sec. V. Summary and Conclusions

The uncropped image set required a significantly higher percentage of the total energy in order to reproduce a good quality image. The lower quality images showed a significant amount of ghosting. The first mode of the uncropped image set only vaguely resembled a face, whereas the first mode of the cropped image set showed distinct facial features.

For music recognition, the multiclass ECOC model gave better results for the first and second sections (recognizing artist or genre based on single artist examples). The Naïve Bayes model gave better results for the third section (recognizing genre based on multiple artists).

Appendix A MATLAB functions used and brief implementation explanations

confusionmat() – This returns a confusion matrix for predicted and actual data.

fitcnb() – This fits a Naïve Bayes model to the training data.

fitcecoc() – This fits a multiclass ECOC model to the training data.

imread() – This reads an image into a MATLAB array.

imresize() – This resizes an image to a specified size.

randperm() – This returns a random permutation of a set of numbers.

reshape() – This reshapes an array, e.g. a vector to a matrix or vice versa.

svd() – This calculates the three matrices of SVD based on the input matrix.

Appendix B MATLAB codes

Part 1

```
cd 'C:\Users\Ashley\Documents\MATLAB\Yale Faces Cropped\CroppedYale\';
faceData = dir('**/*.pgm');
%faceData = dir('C:\Users\Ashley\Documents\MATLAB\Yale Faces
Uncropped\yalefaces\*');
facesArray = zeros(192*168,length(faceData));
for i = 3:numel(faceData)
filename = [faceData(i).folder '\' faceData(i).name];
currentImage = imread(filename); % read in image file to matrix
% resize images
currentImage = imresize(currentImage, [192,168]);

facesColumn = reshape(currentImage, [], 1); %reshape 192x168 matrix to column
vector
facesArray(:, i) = facesColumn; %put column vector into matrix
end

[u,s,v] = svd(facesArray, 'econ'); %do SVD on array of faces data

%Part 3 What does the singular value spectrum look like and how many modes
are necessary for good image reconstructions? (i.e. what is the rank r of the
face space?)
semilogy(diag(s)/sum(diag(s)));

%images with various modes
modes = [1,5,10,100,200,300,500,600,700,800,900,1000];
%modes = [1,5,10,25,50,75,100,110,120,130,140,150];
for j=1:12
```

```
index=1979;
index=79;
ff=u(:,1:modes(j))*s(1:modes(j),1:modes(j))*v(:,1:modes(j)).';
testImageVector = ff(:,index);
testImage = reshape(testImageVector,[192,168]);
subplot(3,4,j);
imagesc(testImage), colormap(gray), axis square, axis off
end
```

Part 2

```
%Example of routine to put multiple wav files into a single array of data z.

clear all; close all; clc

z=[]

list= dir('C:\Users\Ashley\Desktop\HW4\Part 3\Punk\*');
for j=3:length(list)
    currentFileName = strcat('C:\Users\Ashley\Desktop\HW4\Part 3\Punk\',
list(j).name);
    y = audioread(currentFileName)
    z = [z; y]
end


%Note: For efficiency for testing the processing algorithms, I saved each
file above as a .mat file to use in the code below and executed the code
above and the code below as separate scripts

dataSet=zeros(90,40000);
%populate dataSet values 1-30
%load('CharlesMingus.mat')
%load('Ramones.mat')
load('Jazz.mat');
currentAudio = z;
for i=1:30
    currentPos = round((length(currentAudio)-40000)*rand);
    currentClip = currentAudio(currentPos:currentPos+39999);
    dataSet(i,:)=currentClip(:);
end

%populate dataSet values 31-60
%load('Siouxsie.mat')
%load('Siouxsie.mat')
load('Punk.mat');
currentAudio = z;
for i=31:60
    currentPos = round((length(currentAudio)-40000)*rand);
    currentClip = currentAudio(currentPos:currentPos+39999);
    dataSet(i,:)=currentClip(:);
end


%populate dataSet values 61-90
```

```
%load('Bach.mat')
%load('JoyDivision.mat')
load('Baroque.mat');
currentAudio = z;
for i=61:90
    currentPos = round((length(currentAudio)-40000)*rand);
    currentClip = currentAudio(currentPos:currentPos+39999);
    dataSet(i,:)=currentClip(:);
end

dataSpec=[];
for i=1:90
    v=dataSet(i,:);
    %Take a spectrogram of each clip and store it
    L=5; n=40000;
    t2=linspace(0,L,n+1); t=t2(1:n);
    k=(2*pi/L)*[0:n/2-1 -n/2:-1]; ks=fftshift(k);

    windowWidth=5000;
    translationStep = 0.1;
    kmax = n/(40*L);

    Vgt_spec=[];
    tslide=0:translationStep:L;
    for j=1:length(tslide)
        g=exp(-1*windowWidth*(t-tslide(j)).^2); % Gaussian Kernel
        Vg=g.*v; Vgt=fft(Vg);
        Vgt_spec=[Vgt_spec; abs(fftshift(Vgt))];
    end

    concatVgt_spec=[];
    for p=1:51
        concatVgt_spec = [concatVgt_spec, squeeze(Vgt_spec(p,:))];
    end
    dataSetSpec(:,i)=concatVgt_spec(:);

end


%Apply SVD to each spectrogram and store the data
[u,s,v]=svd(dataSetSpec,'econ')



totalConfusionMatrix=zeros(3);
for numberOfTrials = 1:20
    q1=randperm(30);
    q2=randperm(30);
    q3=randperm(30);
    xGroup1=v(1:30,1:50);
    xGroup2=v(31:60,1:50);
    xGroup3=v(61:90,1:50);
    xtrain=[xGroup1(q1(1:25),:); xGroup2(q2(1:25),:); xGroup3(q3(1:25),:)];
```

```
    xtest=[xGroup1(q1(26:end),:); xGroup2(q2(26:end),:);
xGroup3(q3(26:end),:)];
    ctrain=[ones(25,1); 2*ones(25,1);3*ones(25,1)];
    actual=[ones(5,1);2*ones(5,1);3*ones(5,1)];
    %Apply Naive Bayes Classifier and test the model
    nb=fitcnb(xtrain,ctrain)
    pre=nb.predict(xtest);
    %Apply multiclass ECOC Classifier and test the model
    %model=fitcecoc(xtrain,ctrain);
    %pre=model.predict(xtest);
    %Create the current confusion matrix
    currentConfusionMatrix = confusionmat(actual,pre);
    %Update the total confusion matrix
    totalConfusionMatrix = totalConfusionMatrix + currentConfusionMatrix;
end
```

Appendix C Music List

a) **Charles Mingus (from Mingus Ah Um)**
"Better Git It in Your Soul"
"Goodbye Pork Pie Hat"
"Boogie Stop Shuffle"
"Self-Portrait in Three Colors"
"Open Letter to Duke"
Siouxsie and the Banshees (from Juju)
"Spellbound"
"Into the Light"
"Arabian Knights"
"Halloween"
"Monitor"
J.S. Bach (Brandenburg Concerto)
Movements 1-5

b) **The Ramones (from Hey Ho Let's Go The Anthology Disc 1)**
"Beat on the Brat"
"I Wanna Be Your Boyfriend"
"53rd and 3rd"
"Now I Wanna Sniff Some Glue"
"I Wanna Be Sedated"

Siouxsie and the Banshees (from Juju)
"Spellbound"
"Into the Light"
"Arabian Knights"
"Halloween"
"Monitor"

Joy Division (from Unknown Pleasures)
"Disorder"
"New Dawn Fades"
"She's Lost Control"
"Shadowplay"
"Interzone"

c) **Charles Mingus –** "Self-Portrait in Three Colors"
John Coltrane – "Blue Train"
Miles Davis – "Sanctuary"
Thelonious Monk – "Misterioso"
Cecil Taylor – "Unit Structure / As of Now / Section"

Siouxsie and the Banshees – "Spellbound"
The Ramones – "I Wanna be Sedated"
Joy Division – "Shadowplay"
Social Distortion – "So Far Away"
Dead Kennedys – "Macho Insecurity"

J.S. Bach – "Brandenburg Concerto No. 1, Allegro"
Domenico Scarlatti – "Sonata in D Minor K.5"
Handel - "Sarabande"
Pachelbel – "Canon in D Minor"
Buxtehude – "Passaglia"