CS 231A Computer Vision (Autumn 2012)
Problem Set 1

Due: Oct. 9th, 2012 (2:15 pm)

1 Finding an Approximate Image Basis – EigenFaces (25 points)

In this problem you will implement a solution to a facial recognition task. Specifically, you will determine whether a given image contains a face that belongs to a given set of people. The method you will explore is discussed in Lecture 2, and is known as EigenFaces. EigenFaces relies upon the idea of Principle Component Analysis (PCA).

In this problem, we have high dimensional data and believe that valuable information can be represented in a lower dimensional space. This dimensionality reduction is a typical application of PCA. We will start by representing our data as a line and then decrease the dimensions of our representation.

The images in this problem are in the form of vectors, where $x^{(i)}$ denotes the vector corresponding to the image i. Our representation will consist of a set of linearly independent vectors $a^{(i)}$. The set of vectors, $a^{(i)}$ is known as the facespace. Given a test image as a vector y, we will project it onto the range of the facespace. For our projection, we aim to find a set of weights w to solve

$$\text{min. } \|Aw - y\|_2^2$$

where $A = \left[a^{(1)} \dots a^{(n)}\right]$, or in other words $a^{(i)}$ are the columns of A. Our projection is given by $y_{proj} = Aw_{opt}$. Once we have this projection, we can determine whether y is in our given set of people.

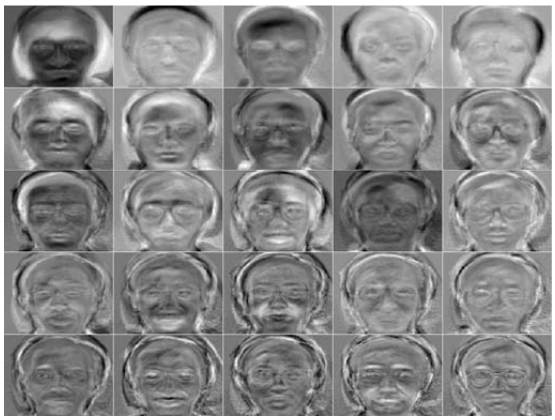

Figure 1: Example facespace

First consider the approximation to our high dimensional data by a line set, L, where $a \in \mathbb{R}^n$, $b \in \mathbb{R}^n$ and a is a unit vector. The set

$$L = \{\ az + b \mid z \in \mathbb{R}\}$$

is a line in $\mathbb{R}^n$. For any vector $x \in \mathbb{R}^n$ the distance between x and L is defined by

$$d(x, L) = \min_{y \in L} \|x - y\|_2$$

where $\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$ (l_2 norm). We have m vectors, and we would like to find the line L that minimizes the total squared distance

$$d_t = \sum_{i=1}^{m} d(x^{(i)}, L)^2$$

In other words we'd like to fit the points to a straight line, as close as possible such that d_t is minimized.

(a) Show that $d(x, L) = \|(I - aa^T)(x - b)\|_2$. *Hint: you might consider using a least squares approach*

(b) Using the result from (a), the total distance squared is given by

$$d_t = \sum_{i=1}^{m} \|(I - aa^T)(x^{(i)} - b)\|_2^2$$

Show that the optimal b, which minimizes d_t while all other variables are held constant, is given by

$$b_{\text{opt}} = \frac{1}{m} \sum_{i=1}^{m} x^{(i)}$$

That is b_{opt} is the mean of the vectors $x^{(i)}$. *Hint:* $P = (I - aa^T)$ *is a projection matrix.*

(c) Using this choice of $b = b_{opt}$, find the optimal value of a which minimizes d_t, in terms of $x^{(i)}$ and b_{opt}? *Hint: Try to formulate the problem as a maximization problem of the form* $\max(a^T \Sigma a)$.

The remaining parts of this problem require programming. We have provided skeleton code in `PS1_data.zip` in the folder `eigenFaces`. The provided code can be executed using Matlab on the computing clusters/personal machines. The appropriate `.m` files have been commented to indicate where your code should be placed and the variables provided by us. All your code for this problem should be located in `eigenface.m`.

(d) Run the script `readYaleFaces`. This will load a matrix A, whose columns are the reshaped images of the data set, into the Matlab workspace. Find the mean of these images (as in the Lecture 2 notes) and subtract it from all the images to form a matrix B (of same dimension as A). Write code that finds the r largest singular values and the corresponding left singular vectors of B. These singular vectors are the eigenfaces corresponding to our data set. Turn in the 5 largest singular values. Please put your code is the m-file `eigenface.m`. *Note: that this is a big matrix, and just typing* `svd` *may not do what you want. Check the help for the* `svd` *command in Matlab.*

(e) For facial recognition, we project the given image onto the facespace. If the resulting error is less than a given threshold, the image is recognized as belonging to the facespace. For this part use 25 eigenfaces, *i.e.* choose $r = 25$ above. You are given twenty images, i.e. `image1.mat` ... `image20.mat`. Images `image1.mat`, `image3.mat`, `image6.mat`, `image12.mat`, and `image14.mat` are not in the facespace. To determine if an image y is in the facespace we use a threshold τ such that if $\|y - y_{proj}\|_2^2 \geq \tau$ the given image is not in the facespace and if $\|y - y_{proj}\|_2^2 < \tau$ the given image is in the facespace. This threshold τ should maximize the number of correct classifications. To find a value for τ use the following threshold values and report the value which yields the maximum correct classifications $\tau = \{0.009\text{E}8,\ 0.09\text{E}8,\ 0.9\text{E}8,\ 9\text{E}8\}$. Also report the number of correct detections this optimal value of τ achieves. In the skeleton code there is code that will calculate the number of correct detections for you.

2 Steerable filters (25 points)

Images can often be characterized by an aggregation of local information contained in the derivative or gradient. State of the art image descriptors often rely on such local gradients. In this problem, you will derive one local operator, steerable filters, which provides the directional derivative of an image. Due to the efficiency of local operators they are often used in real-time computer vision tasks such as motion estimation and face recognition. Steerable filters also provide the flexibility to look for derivatives you might be expecting in a specific direction.

Let $G^0(x, y)$ be some 2-dimension Linear Shift Invariant (LSI) filter, a function of the cartesian coordinates x and y. Let $G^\theta(x, y)$ be a rotation of $G^0(x, y)$ by θ radians about the origin in the counter-clockwise direction.

(a) Show that

$$G^\theta(x, y) = G^0(r\cos(\phi - \theta), r\sin(\phi - \theta))$$

where $r = \sqrt{x^2 + y^2}$ and $\tan\phi = y/x$.

(b) Using the fact that $G^\theta(x, y)$ can be written as

$$G^\theta(x, y) = G^0\left(r\cos(\phi - \theta), r\sin(\phi - \theta)\right)$$

write an expression for $G^\theta(x, y)$ in terms of r, θ, ϕ, given that $G^0(x, y) = -2xe^{-(x^2+y^2)}$.

Let $F^\theta(x, y) = I(x, y) \star G^\theta(x, y)$ where $\star$ denotes convolution. Using your expression for $G^\theta(x, y)$ show that $F^\theta(x, y) = a(I(x, y) \star G^0(x, y)) + b(I(x, y) \star G^{\pi/2}(x, y))$ where $a, b \in \mathbb{R}$ i.e. $F^\theta(x, y)$ can be written as a linear combination of $I(x, y) \star G^0(x, y)$ and $I(x, y) \star G^{\pi/2}(x, y)$. State the value of a, b explicitly.

(c) Find the direction of maximum response at a point (x,y) of the image $I(x, y)$ to the steerable filter $G^\theta(x, y)$. The direction of maximum response is the θ, such that $F^\theta(x, y)$ has the largest magnitude. Give your answer in terms of the image $I(x, y)$ and the responses $F^0(x, y)$, $F^{\pi/2}(x, y)$ to the two steerable basis filters $G^0(x, y)$, $G^{\pi/2}(x, y)$.

Remark: Once this maximum response is found you can use it to steer your filter $G^\theta(x, y)$ such that it will produce a large response when an image similar to the original $I(x, y)$ passes through the filter.

3 Digital Matting (25 points)

In digital matting, a foreground element is extracted from a background image by estimating a color and opacity for the foreground element at each pixel. Thus we can think of digital matting as a form of background subtraction. The opacity value at each pixel is typically called its α. Matting is used in order to composite the foreground element into a new scene. Matting and compositing were originally developed for film and video production. The most common compositing operation is the over operation, which is summarized by the compositing equation

$$C = \alpha F + (1 - \alpha)B$$

where $C \in \mathbb{R}^3$, $F \in \mathbb{R}^3$, and $B \in \mathbb{R}^3$ are the pixel's composite, foreground, and background color vectors , respectively, and $\alpha \in \mathbb{R}$ is the pixel's opacity component used to linearly blend between foreground and background.

For the development that follows, we will assume that our input image has already been segmented into three regions: background, foreground, and unknown, with the background and foreground regions having been delineated conservatively. The goal then, is using local image statistics to solve for the foreground color F, background color B, and opacity α given the observed color C for each pixel within the unknown region of the image.

The method you will derive uses a continuously sliding window for local neighborhood definitions, marches inward from the foreground and background regions, and utilizes nearby computed F, B, and α values (in addition to these values from known regions) in constructing oriented Gaussian distributions. Using a statistical framework we will leverage these distributions to improve the accuracy of the foreground and background within the initially unknown region. Specifically we will use maximum a posteriori (MAP) estimation to improve our foreground and background segmentation. The MAP estimate is given by

$$\arg\max_{F,B,\alpha} \; P(F, B, \alpha \mid C)$$

Remark : If you need a refresher on MAP estimation, log likelihood functions, and/or maximum likelihood estimation you can refer to Chapter 3 of Pattern Classification Duda, Hart, and Stork which we have posted on the course webpage. We also gave a brief introduction to maximum likelihood estimation in Problem Set 0 Problem 5.

(a) Generally we don't know the distribution $P(F, B, \alpha \mid C)$. In this problem we will use the simplifying assumption that F, B and α are independent. Using this assumption rewrite

(a) Input (b) Opacity α (c) Composite

Figure 2: Example of digital matting

the MAP estimate as the sum of log-likelihood functions, $\arg\max_{F,B,\alpha} \ell(F, B, \alpha \mid C) =$?. Make sure there are no terms in your solution that are probabilities given C.

(b) We can model the log-likelihood function of our observation C as

$$\ell(C \mid F, B, \alpha) = -\frac{\|C - \alpha F - (1-\alpha)B\|^2}{\sigma_C^2}$$

where we know σ_C. Also we model the log-likelihood function of foreground color F as

$$\ell(F) = -(F - \bar{F})^T \Sigma_F^{-1} (F - \bar{F})$$

where for the purposes of this problem you can assume $\bar{F}$ and Σ_F have already been computed. The log-likelihood for the background B is

$$\ell(B) = -(B - \bar{B})^T \Sigma_B^{-1} (B - \bar{B})$$

again, for the purposes of this problem you can assume $\bar{B}$ and Σ_B have already been computed. Also, Σ_F^{-1} and Σ_B^{-1} are symmetric matrices. To solve for the MAP estimate, we need to split this problem into two sub-problems. To solve the first sub-problem we hold α constant while optimizing $\ell(F, B, \alpha)$ over F and B. For this part of the problem hold α constant and show that the optimal F and B are given by the linear equation

$$\begin{bmatrix} \Sigma_F^{-1} + I\alpha^2/\sigma_C^2 & I\alpha(1-\alpha)/\sigma_C^2 \\ I\alpha(1-\alpha)/\sigma_C^2 & \Sigma_B^{-1} + I(1-\alpha)^2/\sigma_C^2 \end{bmatrix} \begin{bmatrix} F \\ B \end{bmatrix} = \begin{bmatrix} \Sigma_F^{-1}\bar{F} + C\alpha/\sigma_C^2 \\ \Sigma_B^{-1}\bar{B} + C(1-\alpha)/\sigma_C^2 \end{bmatrix}$$

(c) Now, we solve the second sub-problem by optimizing $\ell(F, B, \alpha)$ over α while holding F and B constant. In this problem we assume that $\ell(\alpha)$ is a constant. Show that, under this assumption, the optimal α is given by

$$\alpha = \frac{(B - F)^T (B - C)}{\|B - F\|^2}$$

Remark: To obtain the final estimates for F,B and α in this problem we must iterate between these two solutions.

4 Content-Aware Image Resizing (25 points)

For this exercise, you will implement a version of the content-aware image resizing technique described in Shai Avidan and Ariel Shamirs SIGGRAPH 2007 paper, Seam Carving for Content-Aware Image Resizing. In this problem all the code and data can be found in `PS1_data.zip` in the `seamCarving` folder. First read through the paper, with emphasis on sections 3, 4.1, and 4.3.

(a) Write a Matlab function which computes the energy of an image, where Energy is defined to be

$$e(x, y) = \sum_{c \in \{R,G,B\}} \left| \frac{\partial I_c(x, y)}{\partial x} \right| + \left| \frac{\partial I_c(x, y)}{\partial y} \right|$$

where $I_c(x, y)$ is the pixel in color channel c located at (x, y). This function should be written in `computeEnergy.m`, take an RGB image in and return a 2D array of equal size which contains the derivative. In this case, we will use Sobel filters to take the horizontal and vertical derivatives, then create our energy from the sum of the absolute value. The sobel filters are defined as:

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

Also we have included a sample image `parta_sampleinput.jpg`, and the output of our version of `computeEnergy.m`, `parta_sampleoutput.mat` to help you debug your code. You can use the `testEnergy.m` to show the difference between your output and our output. *Hint: you may find the Matlab function* `filter2` *useful.*

(b) The energy of the optimal seam is defined in Equation 4 of Avidan et al as:

$$s^* = \min_z \sum_{i=1}^{n} e(\mathbf{I}(s_i))$$

where $\mathbf{I}(si)$ are the pixels of a path of a seam. In this problem we will compare two different methods of solving this problem. Turn in your code for both sub problems.

(i.) Implement a greedy algorithm for finding the optimal seam by sequentially adding each pixel in the path depending on the energies of immediate options. Use the skeleton code found in `findSeam_Greedy.m` to write your code. We have given you our seam carving results of the greedy algorithm on `stanford.jpg`, as `stanford_resize_gdy.jpg` to help you debug your final seam carving code in part(c).

(ii.) Implement the dynamic programming solution given in the paper with the following update step:

$$M(i, j) = e(i, j) + \min(M(i-1, j-1), M(i-1, j), M(i-1, j+1))$$

Where M is the cumulative energy function, as mentioned in the paper. Use the skeleton code found in `findSeam_Dyn.m` to help you write your code. The results of the dynamic programming is provided as `stanford_resize_dyn.jpg`. You can use it to debug your final code in part(c).

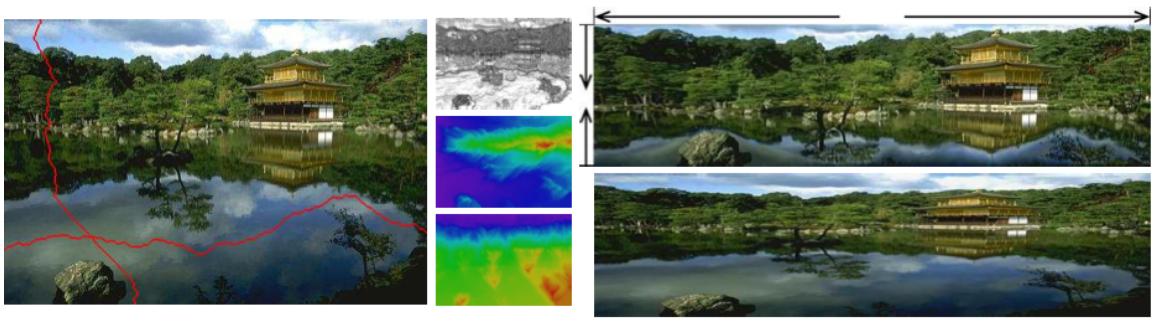

Figure 3: Example of content-aware image resize compared to standard rescale

(c) Use the skeleton code provided in `reduceWidth.m` to combine your parts (a) and (b) into a functional seam carving application.

(d) Now test your code by running the following tests using both the dynamic programming and greedy techniques. Submit all images generated from this part:

```
reduceWidth('stanford.jpg', 200)
reduceWidth('dogs.jpg', 200)
reduceWidth('pool.jpg', 200)
```

We have given you our results on the `stanford.jpg` image as `stanford_resize_dyn.jpg` and `stanford_resize_gdy.jpg` for the dynamic programing and greedy algorithms respectively to help you debug your code. Comment on where the inferiority of the greedy methodology is most apparent and how this is solved by dynamic programming. Also comment on what types of images are sensitive to this manipulation and which images can have many lines removed without a perceptible difference. Find 2 images of your own and run experiments to show both a good and bad performance using the dynamic programming approach.