Advanced Topics in Machine Learning (600.692)
Homework 6: Robust Face Recognition with Varying Illumination

Instructor: René Vidal

Due Date: 04/04/2014, 11.59PM Eastern

READING MATERIAL: Chapter 3 and Appendix A of GPCA book.

1. In this exercise you will use a small subset of the Yale B dataset[1], which contains photos of ten individuals under various illumination conditions. Specifically, you will use only images from the first three individuals under ten different illumination conditions. Divide these images into two sets: *Training Set* (images 1-5 from individuals 1 to 3) and *Test Set* (images 6-10 from individuals 1-3). Notice also that there are 5 non-face images (accessible as images 1-5 from individual 4). We will refer to these as *Outlier Set*. Download the file YaleB-Dataset.zip. This file contains the images along with the MATLAB function `loadimage.m`. Decompress the file and type `help loadimage` at the MATLAB prompt to see how to use this function. The function operates as follows.

Function `img=loadimage(individual,condition)`	
Parameters	
`individual`	Number of the individual.
`condition`	Number of the image for that individual.
Returned values	
`img`	The pixel image loaded from the database.
Description	
Read and resize an image from the dataset. The database (directory `images`) must be in the same directory as this file.	

   (a) **Face completion.** Remove uniformly at random $\{0, 10, 20, 30, 40\}\%$ of the entries of all images of individual 1. Apply the Low Rank Matrix Completion (LRMC) algorithm that you implemented in HW4 to these images to compute the mean face and the eigenfaces as well as fill in the missing entries. Note that LRMC does not compute the mean face, so you will need to modify your code from HW4. Plot the mean face and the top three eigenfaces and compare them to what you obtained with PCA in HW3Q2a. Plot also the completed faces and comment on the quality of completion as a function of the percentage of missing entries by visually comparing the original images (before removing the missing entries) to the completed ones. Plot also the error (Frobenius norm) between the original images and the completed ones as a function of the percentage of missing entries and comment your results. Repeat for individuals 2 and 3.

   (b) **Face recognition with missing entries.** Remove uniformly at random $\{0, 10, 20, 30, 40\}\%$ of the entries of all images in the *Training Set* and *Test Set*. Apply the Low Rank Matrix Completion (LRMC) algorithm that you implemented in part (a) to the images in the *Training Set*. Plot the projected training images $\boldsymbol{y} \in \mathbb{R}^d$ for $d = 2$ or $d = 3$ using different colors for the different classes. Do faces of different individuals naturally cluster in different regions of the low-dimensional space? Classify the faces in the *Test Set* by using 1-nearest-neighbor. That is, label an image $\boldsymbol{x}$ as corresponding to individual i if its projected image $\boldsymbol{y}$ is closest to a projected image $\boldsymbol{y}_j$ of individual i. Notice that you will need to develop new code to project an image with missing entries $\boldsymbol{x}$ onto the face subspace you already estimated from the *Training Set*, which you can do as described in Section 3.1 of the GPCA book. Report the percentage of correctly classified face images for $d = 1, \ldots, 10$ and the percentage of missing entries $\{0, 10, 20, 30, 40\}\%$.

[1] `http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html`.

(c) **Face correction.** Remove uniformly at random $\{0, 10, 20, 30, 40\}\%$ of the entries of all images of individual 1 and replace them by arbitrary values chosen uniformly at random from $[0, 255]$. Apply the RPCA algorithm for corrupted entries that you implemented in HW5 to these images to compute the mean face and the eigenfaces as well as correct the corrupted entries. Note that RPCA does not compute the mean face, so you will need to modify your code from HW5. Plot the mean face and the top three eigenfaces and compare them to what you obtained with PCA in HW3Q2a. Plot also the corrected faces and comment on the quality of correction as a function of the percentage of corrupted entries by visually comparing the original images (before removing the missing entries) to the completed ones. Plot also the error (Frobenius norm) between the original images and the corrected ones as a function of the percentage of corrupted entries and comment your results. Repeat for individuals 2 and 3.

(d) **Face recognition with corrupted entries.** Remove uniformly at random $\{0, 10, 20, 30, 40\}\%$ of the entries of all images of individual 1 and replace them by arbitrary values chosen uniformly at random from $[0, 255]$. Apply the RPCA algorithm for corrupted entries that you implemented in part (a) to the images in the *Training Set*. Plot the projected training images $\boldsymbol{y} \in \mathbb{R}^d$ for $d = 2$ or $d = 3$ using different colors for the different classes. Do faces of different individuals naturally cluster in different regions of the low-dimensional space? Classify the faces in the *Test Set* by using 1-nearest-neighbor. That is, label an image $\boldsymbol{x}$ as corresponding to individual i if its projected image $\boldsymbol{y}$ is closest to a projected image $\boldsymbol{y}_j$ of individual i. Notice that you will need to develop new code to project an image with corrupted entries $\boldsymbol{x}$ onto the face subspace you already estimated from the *Training Set*. Report the percentage of correctly classified face images for $d = 1, \ldots, 10$ and the percentage of missing entries $\{0, 10, 20, 30, 40\}\%$.

(e) **Outlier detection.** Augment the images of individual 1 with those from the *Outlier Set*. Apply the RPCA algorithm for data corrupted with outliers that you implemented in HW5 to these images to compute the mean face and the eigenfaces as well as detect the outliers. Note that RPCA does not compute the mean face, so you will need to modify your code from HW5. Plot the mean face and the top three eigenfaces and compare them to what you obtained with PCA in HW3Q2a. Report the percentage of correctly detected outliers.

(f) **Face recognition with corrupted entries.** Apply the RPCA algorithm for for data corrupted with outliers that you implemented in part (e) to to the images in *Training Set* $\cup$ *Outlier Set*. Plot the projected training images $\boldsymbol{y} \in \mathbb{R}^d$ for $d = 2$ or $d = 3$ using different colors for the different classes. Do faces of different individuals naturally cluster in different regions of the low-dimensional space? Classify the faces in the *Test Set* by using 1-nearest-neighbor. That is, label an image $\boldsymbol{x}$ as corresponding to individual i if its projected image $\boldsymbol{y}$ is closest to a projected image $\boldsymbol{y}_j$ of individual i. Report the percentage of correctly detected outliers and the percentage of correctly classified face images for $d = 1, \ldots, 10$ and compare your results to those in HW3Q3.

Submission instructions. Please follow the same instructions as in HW1.