CSC411: Project #1

Due on Monday, January 29, 2018

Xiangyu Kong

Problem 1

Dataset description

The dataset consists of 1691 gray-scale 32 × 32-pixel of images of 12 different actor/actress's faces. The angle, lighting and expressions for the faces are all different (as shown in Figure 1a and Figure 1b). Some faces aren't fully shown (as shown in Figure 1c). After cropping their faces out, examining the dataset, most of the faces are correctly cropped given the coordinates (Figure 1d), but there are still incorrect coordinates that leads to incorrect faces (Figure 1e). Also, some cropped faces are not aligned together. e.g. Some are center aligned and some are not (Figure 1d and Figure 1f).

(a) front

(b) side

(c) blocked face

(d) correctly cropped

(e) incorrect coordinates for face

(f) face not aligned to the center

Figure 1

Problem 2

Separating the sets.
The algorithm for separating the sets of data is to first shuffle the data as a list, then through list slicing, first pick 10 images as the test set, then pick 10 images as the validation set, and finally the rest are used as the training set.