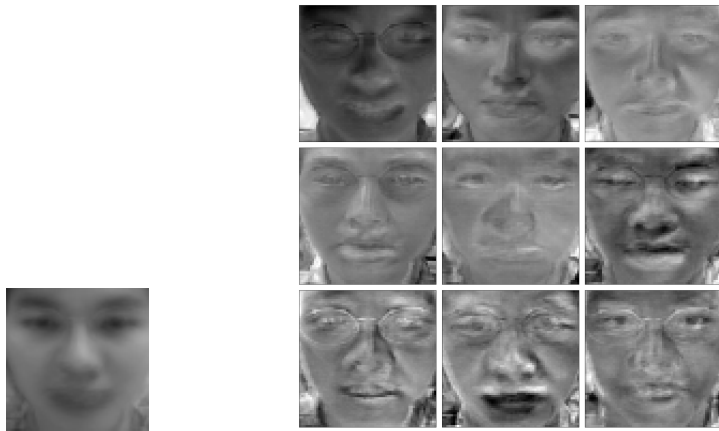


學號：R05922096 系級：資工碩一 姓名：李哲安

1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

答：(左圖平均臉，右圖為 3x3 格狀 eigenfaces, 順序為 左到右再上到下)



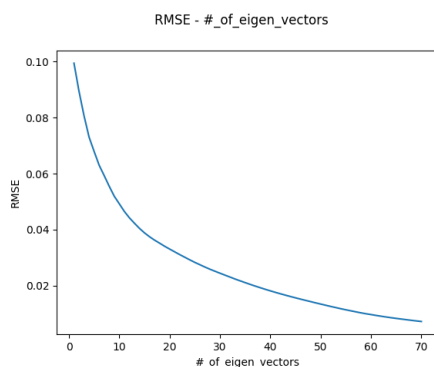
1.2. Dataset 中前 10 個人的前 10 張照片的原始圖片和 reconstruct 圖 (用前 5 個 eigenfaces):

答：(左右各為 10x10 格狀的圖, 順序一樣是左到右再上到下)



1.3. Dataset 中前 10 個人的前 10 張照片投影到 top k eigenfaces 時就可以達到 < 1% 的 reconstruction error.

答：k = 59 時 RMSE = 0.00996823197098 < 0.01

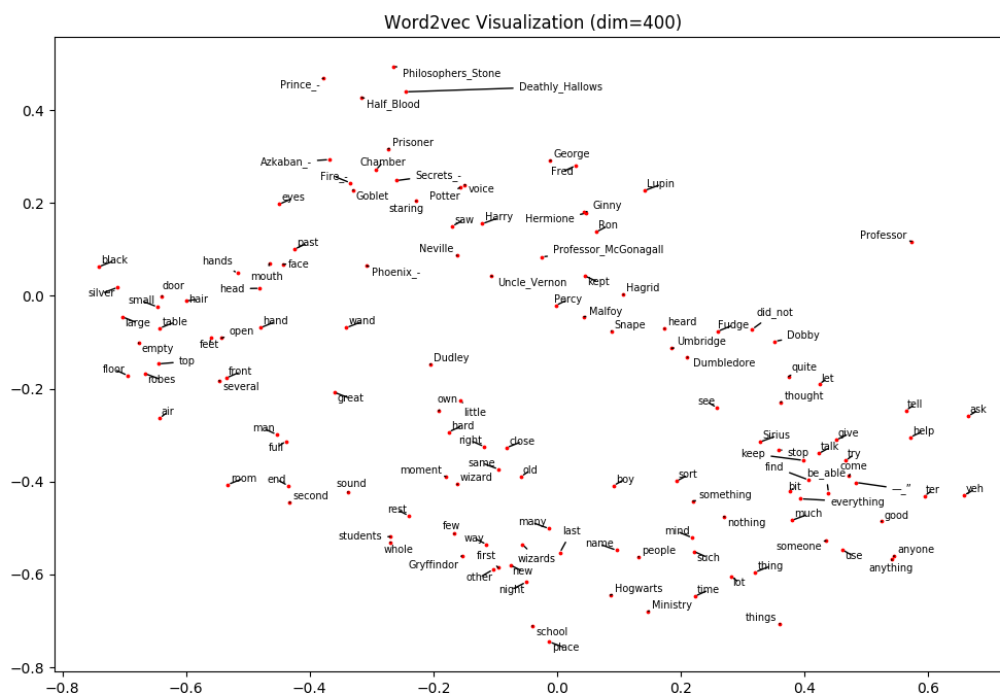


2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

word2vec.word2vec()中

-output 是 training data 路徑

2.2. 將 word2vec 的結果投影到 2 維的圖:



答：

相似的字在二維圖中會傾向聚集在一起。

3.1. 請詳加解釋你估計原始維度的原理、合理性，這方法的通用性如何？

答：

參考 Paper 和 source code

Paper: “Maximum Likelihood Estimation of Intrinsic Dimension, Eliza veta Levina, Peter J.Bicke”

Source Code: https://github.com/OFAI/hub-toolbox-python3/blob/master/hub_toolbox/IntrinsicDim.py

原理為計算 dataset 每筆 data 的 k nearest neighborhood distanse，並利用 Maximum Likelihood Extimation 來估計 dataset 的 Intrinsic Dimension。但在利用 gen.py 的 data 進行測試後發現對於 Intrinsic Dimension 高於 20 的 dataset 預測出來的結果都會偏低，所以利用 gen.py 的來對 source coed 做了係數的修正，最後在 kaggle 上得到 0.1006 的分數。

3.2. 將你的方法做在 hand rotation sequence dataset 上得到什麼結果？合理嗎？請討論之。

答：

對 hand rotation sequence dataset 的預測結果為 50 dimension。

根據 paper 的實驗結果，和在 gen.py 上的 data 測試結果，大於 20 dim 的預測結果很可能都是預測不準確的(預測偏低)。