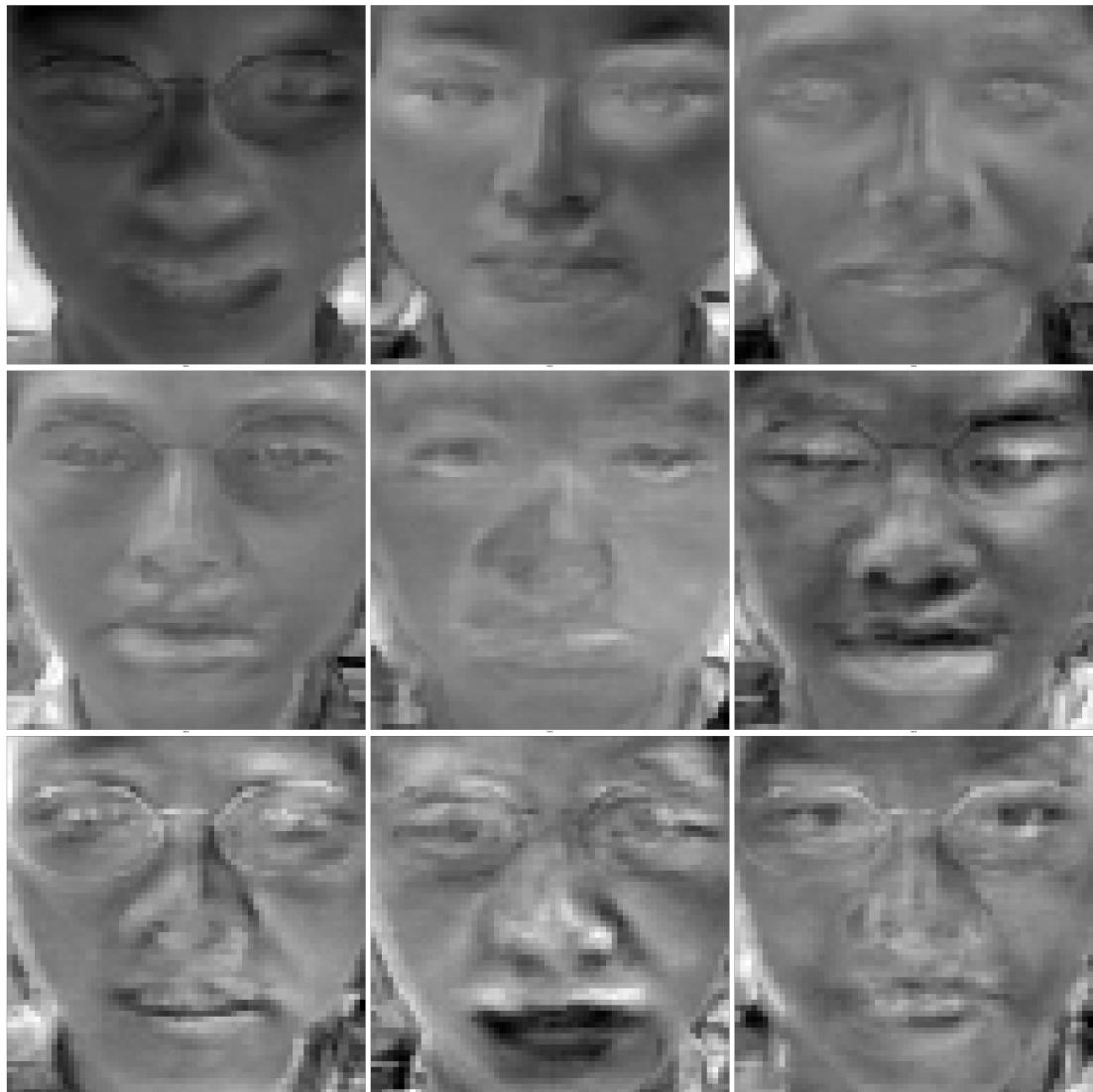


學號：B03902102 系級： 資工三 姓名：廖廷浩

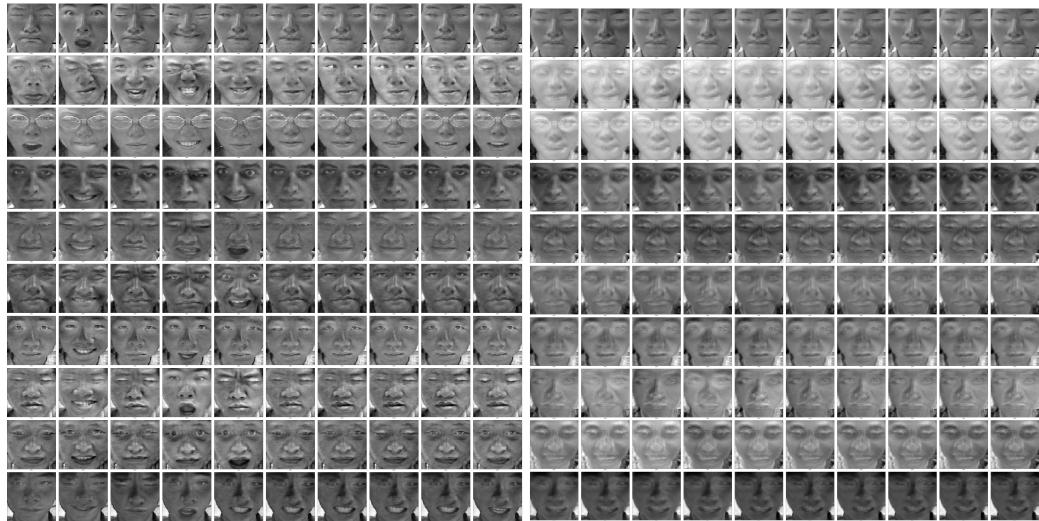
1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

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：(回答 k 是多少)

59

2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

答：

window: 切一段時間內的取一個 window 大小的字數來 train 而非一個一個字

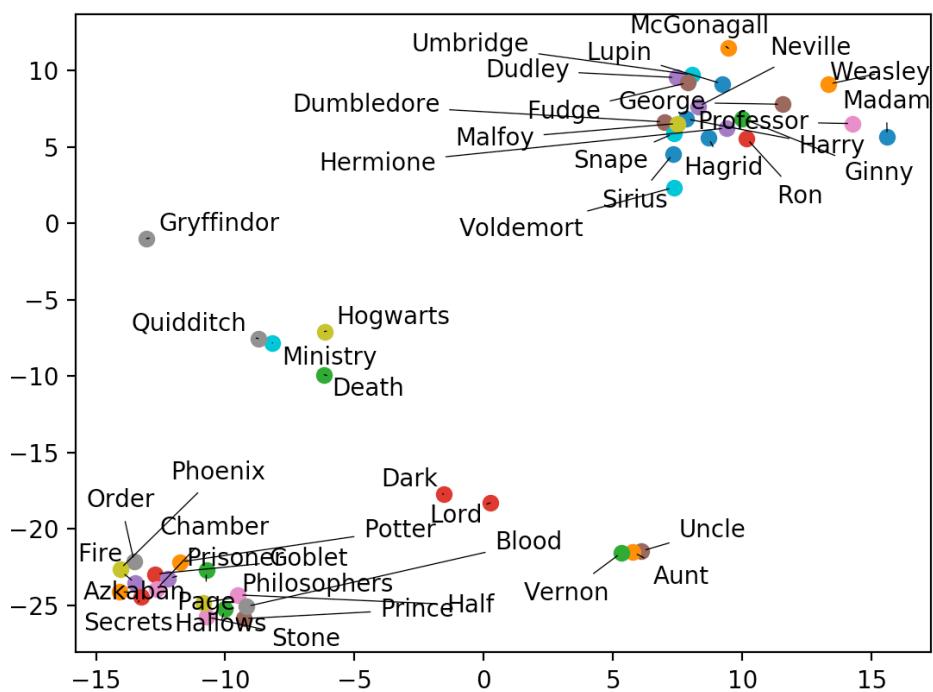
size: 每個 vector 的大小

negative: 提高訓練速度並改善所得詞向量的質量

min_count: 字數太少的字直接忽略掉

2.2. 將 word2vec 的結果投影到 2 維的圖:

答：(圖)



2.3. 從上題視覺化的圖中觀察到了什麼？

答：

較靠近的 word 大概就是書中會比較接近的單字，例如：

Harry, Ron, professor, Lupin 很接近，而我們也可以從故事中得知這些角色常常會同時出現

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：

預估初的原始維度可以用來做出 eigenvalues，而這樣 train 出來的結果的可信度很高，並且利用此預估的 data 作為 X，test_data 當作 y，而其他類似的題目貨問題皆可用此方法來完成

3.2. 將你的方法做在 hand rotation sequence datatset 上得到什麼結果？合理嗎？

請討論之。

答：