

PCA 实践

数据集

实验所用到的数据集在下面的链接中, 这些数据是来自剑桥大学提供的 AT&T 人脸数据集, 有 40 个人的人脸图像, 每个人有 10 张不同光照和姿态的照片。



地址: http://www.cl.cam.ac.uk/Research/DTG/attarchive/pub/data/att_faces.tar.Z

实验内容

加载数据集, 利用 PCA 算法对数据集内所有人进行降维和特征提取, 然后将得到的主成分特征向量还原成图像进行观察。这里可以尝试采用不同的降维维度 K 进行操作, 分别观察不同 K 下的特征图像。

实验拓展

尝试对刚降维的特征图像进行 PCA 逆变换, 观察变换前后的图像差异。

实验要求

该实验涉及到的 PCA 操作不允许直接调用接口(如 sklearn 中的 PCA 包), 应当自己实现算法细节。

附录: PMG 文件的加载和预览例程

```
import cv2
import numpy as np

FACE_PATH = "E:/att_faces/orl_faces" # \\ can be ambiguous
PERSON_NUM = 40
PERSON_FACE_NUM = 10
K = 10 # Number of principle components

raw_img = []
```

```
data_set = []
data_set_label = []
```

```
def read_data():
    for i in range(1, PERSON_NUM + 1):
        person_path = FACE_PATH + '/s' + str(i)

        for j in range(1, PERSON_FACE_NUM + 1):
            img = cv2.imread(person_path + '/' + str(j) + '.pgm')
            if j == 1:
                raw_img.append(img)
            img_gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
            height, width = img_gray.shape
            img_col = img_gray.reshape(height * width)
            data_set.append(img_col)
            data_set_label.append(i)
    return height, width
```

```
# Import Data
height, width = read_data()
X = np.array(data_set)
Y = np.array(data_set_label)
n_sample, n_feature = X.shape
```

```
# Print some samples
raw_img = np.hstack(raw_img)
cv2.namedWindow("Image")
cv2.imshow('Image', raw_img)
cv2.waitKey(0)
cv2.destroyAllWindows()
```