

# 期中大作业报告

## PCA, 2DPCA, & L1-Norm-2DPCA

wins-m

### *I. Introduction*

在基于 PCA (Principal Component Analysis) 的图像特征提取方案中, 无可避免地需要将 2D 的图像数据转化为 1D 向量 (image-as-vector), 这就导致了图形向量映射到了极高维度的空间中, 其协方差矩阵的维度很高, 因此很难通过相对量级较小的训练样本评估协方差矩阵的准确性。尽管通过 SVD 的方法能够计算出特征向量 (从而避免协方差矩阵的生成), 但其计算结果仍然是由协方差矩阵决定的, 即依然是不精确的。

对此, Yang 等人在 2004 年提出了 2DPCA (two-dimensional principal component analysis) 的方法 [1], 直接利用原始的图像矩阵提取 2D 的主成分矩阵。这一算法只需要计算维度较小的协方差矩阵, 这一方面提高了计算的精确度, 另一方面也提高了算法的效率。值得注意的是, 相较于传统一维的 PCA 方法, 2DPCA 可能保留了原图像矩阵更高维度的空间关系, 因此在图像重建和人像识别方面均表现出优势。

但是, 2DPCA 在决定投影方向时采用了 L2-norm distance; 这一方法同传统 PCA 一样, 在离群噪音存在下的表现较差。为解决这类问题, 有人提出了基于 L1-norm 的 PCA 方法, 以及收敛速度较高的 PCA- L1 迭代算法。Li 等人在 2009 年将 PCA-L1 推广到了二维情形下 [2]。该算法 (下文称其为 L2-2DPCA) 在图像特征提取中有如下优势: 1) 使用基于 L1 范数的误差矩阵, 受离群值干扰较小; 2) 使用二维的图像投影, 保留了像素点之间的空间关系; 3) 通过简单的迭代过程, 避免了协方差矩阵和特征向量的大量运算。

本文将 PCA, 2DPCA 和 L1-Based 2DPCA 算法应用到 Yale 人像数据集, 探索三种算法在实际人脸数据中的表现。Yale 数据集包含 15 位个体 164 张在不同表情和光照条件下的肖像。为简化运算量并增加识别难度, 每张图片被压缩为  $72 \times 96$  的像素点。

在训练集中, 每位个体的图像出现 5 次, 图像中的表情是随机排列的, 剩余的图片进入测试集, 用于计算各算法在不同表情 (光照) 条件下人脸识别的准确率。(需要指出的是, 在更加精确的分类任务中有必要将原始图像裁剪为“人脸”的大小, 使得算法通过人脸的特征判断图像类别, 而非基于人脸在浅色背景中的位置; 迫于时间压力, 本报告中并未对齐人脸的坐标位置。) 在分类任务中测试了不同训练集划分比率和有遮挡物噪音存在下的预测精度。

## II. 2DPCA实现

训练集的图像  $\mathbf{X}_i, i = 1, \dots, d$  为  $h \times w$  的矩阵，将其中心化后得到新的图形矩阵  $\mathbf{S}$ ；对于所有的  $\mathbf{S}_i$  计算其协方差矩阵并求平均值（或总和），获得描述训练集总体情况的图形协方差矩阵（Image covariace matrix)  $\mathbf{G}$ ，其大小为  $w \times w$ 。

$$\mathbf{G} = E[(\mathbf{A} - E\mathbf{A})^T(\mathbf{A} - E\mathbf{A})], \quad \mathbf{A} \in \mathbf{S}$$

可以证明，矩阵  $\mathbf{G}$  的特征向量  $\mathbf{V}_1, \dots, \mathbf{V}_d$  就是使得所有图片在其方向上投影的分散程度总和（L2范数）最大的最优映射。对于样本图像  $\mathbf{X}_i$ ，其在方向  $\mathbf{V}_k$  上的主成分（principal component）为

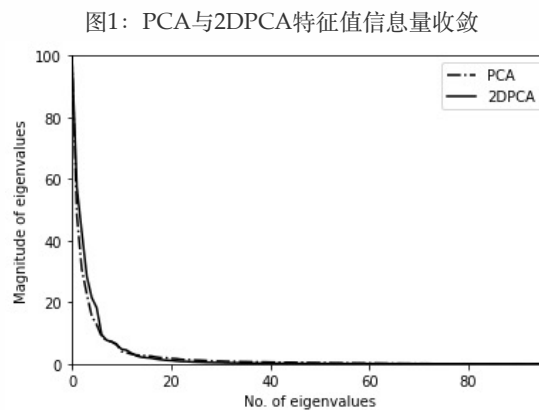
$$\mathbf{Y}_k = \mathbf{X}_i \mathbf{V}_k, \quad k = 1, 2, \dots, w.$$

由此，如果取前  $m$  个主成分，则可以得到原图像的特征矩阵（feature matrix)  $\mathbf{B} = [\mathbf{Y}_1, \dots, \mathbf{Y}_m]$ 。原始图像集  $\mathbf{X}$  的大小为  $d \times h \times w$ ，利用 2DPCA 降维后的特征矩阵大小为  $d \times h \times m$  ( $m \leq w$ )，可见利用 2DPCA 的方式最终实现了图像的降维。

上述过程的 python 代码如下

```
1 def PCA2D(X, n_dim=None):
2     # Centralize
3     S = X - X.mean(0)
4     # Image covariance (scatter) matrix
5     G = np.array([A.T @ A for A in S]).mean(0)
6     # orthonormal eigenvectors of G
7     D,V = np.linalg.eig(G)
8     # Reorder principal eigenvectors
9     sorted_index = np.argsort(D)[::-1]
10    D,V = D[sorted_index], V[sorted_index]
11    return V[:, :n_dim] if n_dim else V
```

如下图 1 所示，在 (75,72,96) 的训练集图像中分解到的特征值很快逼近到 0，其速度与传统 PCA 接近一致，说明上述降维是有效的。



### III. L1-2DPCA实现

对于训练集  $\mathbf{X}$ ，设置主成分向量为任意初值  $\mathbf{u}$ ，满足  $\|\mathbf{u}\|_2 = 1$ 。用 `iterate_u` 迭代，使得特征空间的 L1 范数  $f(\mathbf{u})$  尽量小，其中

$$f(\mathbf{u}) = \sum_{i=1}^d \sum_{j=1}^h |\mathbf{X}_{ij}\mathbf{u}|. \quad f(\mathbf{u}) = \sum_{i=1}^d \sum_{j=1}^h |\mathbf{X}_{ij}\mathbf{u}|.$$

下述迭代是递减的且快速停止的：

$$\mathbf{u}(t+1) = \frac{\sum_{i=1}^d \sum_{j=1}^h p_{ij} \mathbf{X}_{ij}^T}{\left\| \sum_{i=1}^d \sum_{j=1}^h p_{ij} \mathbf{X}_{ij}^T \right\|},$$
$$\text{where } p_{ij}(t) = \begin{cases} 1, & \text{if } |\mathbf{X}_{ij}\mathbf{u}(t)| > 0 \\ -1, & \text{if } |\mathbf{X}_{ij}\mathbf{u}(t)| \leq 0. \end{cases}$$

如是找到第一个主成分向量  $\mathbf{u}_1$ 。调用 `update_x` 更新  $\mathbf{X}$ ，去除其在  $\mathbf{u}_1$  方向上的信息，继而再次重复上述过程，即可得到后续若干主成分。实现代码如下

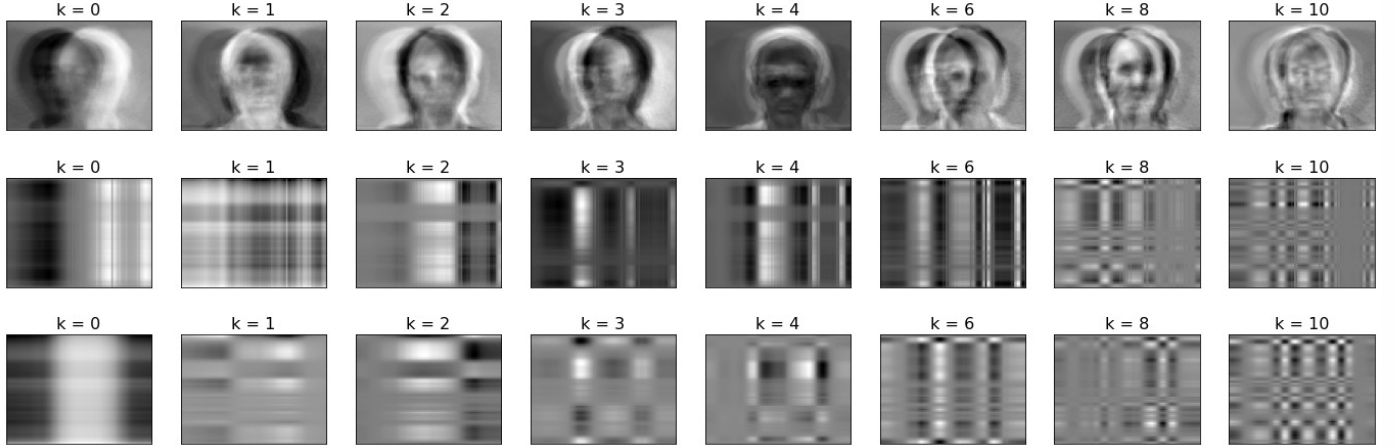
```
1 def PCA2D_L1(X, n_dim=5):
2     # initial value for iteration
3     tmp = np.ones(X.shape[-1])
4     u0 = tmp / np.linalg.norm(tmp)
5     # 1st iteration
6     u = iterate_u(X, u0)
7     # following iterations
8     ret = []
9     ret.append(u)
10    for i in range(n_dim - 1):
11        X = update_x(X, u)
12        u = iterate_u(X, u0)
13        ret.append(u)
14    return np.array(ret).T
15
16 def iterate_u(X, u):
17     delta = 1
18     while delta != 0:
19         Y = X @ u
20         P = (Y>0) * 2 + 1
21         tmp = (P.T * X.T).T.sum(0).sum(0)
22         u = tmp / np.linalg.norm(tmp)
23         delta = np.linalg.norm(X @ u, ord=1) - np.linalg.norm(Y, ord=1)
24     return u
25
26 def update_x(x, u):
27     r = x - (x @ u).reshape(x.shape[0], -1, 1) @ u.reshape(1, -1)
28     return(r)
```

## IV. 可视化分析及聚类

对于我们的训练集应用上述两种 2DPCA 算法降维，得到对应的特征向量，与传统的 baseline PCA 进行比较。

取训练集中的随机一张图片  $\mathbf{X}_i$ ，分别投影至三种算法提取的第  $k$  个特征向量，得到特征矩阵  $\mathbf{Y}_k = \mathbf{X}_i \mathbf{V}_k$ ，再由第  $k$  个主成分反射到重建图像  $\hat{\mathbf{X}}_i = \mathbf{Y}_k \mathbf{V}_k^T$ 。分别取  $k \in \{0, 1, 2, 3, 4, 6, 8, 10\}$ ，重建图像如下（图 2）。由上至下分别为 PCA，2DPCA，L1-2DPCA 的结果（取反转色）。

图2: PCA, 2DPCA, L1-2DPCA 第  $k$  个主成分的特征投影（取反色）



在 PCA 的结果中，在  $k = 1, 2$  时可以明显的看到该图像对应的人像色泽最淡（也即取反前颜色最深），可见前两个主成分中涵盖了人脸的大部分区别于其他类型样本的特征。值得注意的是，除了被提取的人脸特征以外，能够明显注意到其他人像的“残影”，其中的原因是 1DPCA 无法抽取二维的空间信息。

在 2DPCA 的结果中，可能由于人像在背景板上的位置存在 variation，报告展示的结果与参考文献 [1] 中的 Fig.2 有一定差距。在前 6 个特征向量上均保留了较多的信息，之后信息逐步衰弱。在第 10 个维度上残余的已经是碎片化的局部信息。

在 L1-2DPCA 的结果中，可以看出衰减速度比 2DPCA 更快，从第三个向量开始已经呈现出局部信息。说明在降维上 L1-2DPCA 更有效率。考虑到，背景中的位置也是一种噪音，这一对比能在一定程度上说明 L1-2DPCA 有对离群值有更好的表现。

重复上述图像重建的过程，将随机一张图像用前  $d$  个向量重建（而非第  $k$  个）。为确保投射向量的大小一致，在二维取前  $d$  个向量时，一维取前  $hd$  个（见下图 3）。

图3: PCA, 2DPCA, L1-2DPCA 前  $d$  个主成分上的特征投影



PCA 的重建效果较差。在  $d$  较小时保留了较多的“残影”，在高维情况下重建出较多的背景“噪点”，说明该算法未能很好地区分人像面部特征和背景板中的干扰特征。对此的解释是——PCA 提取了所有训练集图像平均的主成分，因此对整体样本而言 PCA 的重建结果可能是可靠的；但对于单个样本，由于空间关系被打乱了，重建效果不理想。

2DPCA 可以较好的保留空间关系。的重建结果中可以看出上一部分的结论，即前几个特征向量并未有效提取出图像特征。第一，由前 2 个主成分重建的图像有明显的层次特征，可知其提取的是人像的发际、眉、眼、鼻、唇、下巴等在垂直方向上的空间关系；第二，在  $d$  逐渐增大后，可以发现在图像右侧边缘似乎出现了左半脸，说明在保留较多信息后该算法仍然未能排除“人脸位置”这一噪音信息的干扰。

L1-PCA 的重建表现最好。首先，即使在特征向量较少的情况下仍然能精准定位出人脸在横向和纵向两个维度的位置，说明该算法不受“人脸位置”噪音的影响；其次，人像与背景板的对比度较高，在  $d \geq 50$  的情况下，已经能还原出无噪音的原始图像。这说明该算法很好的消除了离群噪音的影响。

下分别选用前两个特征向量，将训练集图像  $\mathbf{X}$  将至 2 维，并绘制散点图（下图 4），其中颜色对应其真实的种类（即 hum）。在 2DPCA 和 L1-2DPCA 中，我们先计算原图像的特征矩阵  $\mathbf{Y}_k = \mathbf{X}_i \mathbf{V}_k$ ，再分别计算两个维度上的 L2 范数作为该轴上的坐标。进一步，用 k-means 算法进行聚类，绘制相应的分类图（下图 5），并计算调整兰德指数，F 值，准确率，召回率（见下表 1）。

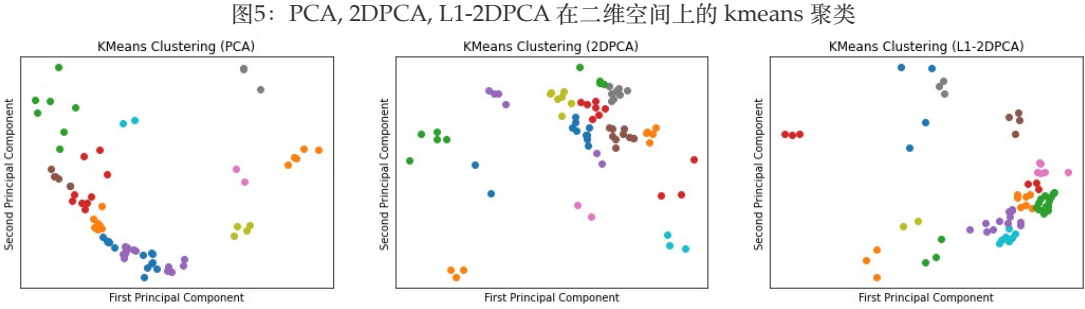
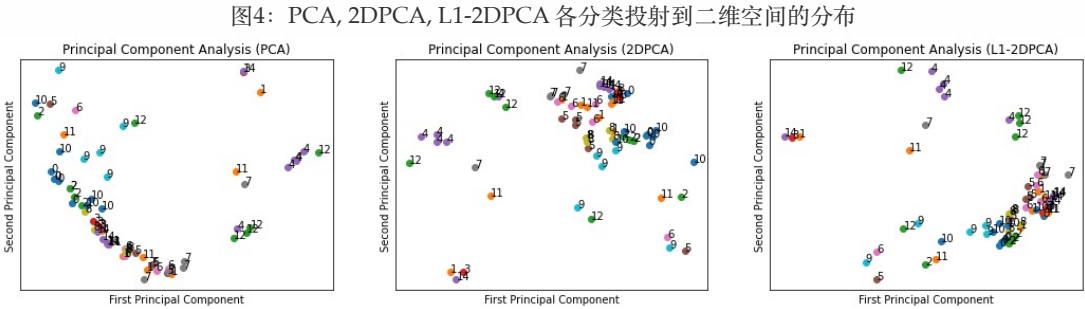


表1: kmeans 聚类结果对比

	ARI	F-Score	Precision	Recall
PCA	0.2784	0.2448	0.2216	0.2733
2DPCA	0.3118	0.2581	0.2304	0.2933
L1-2DPCA	0.3509	0.3017	0.2596	0.3600

上述聚类结果表明，L1-2DPCA 在 F 值，准确率，召回率各方面均明显优于 PCA 和 2DPCA，说明在样本内分类上 L1-2DPCA 具有优势；2DPCA 略优于 baseline PCA，但二者表现较为接近，说明根据前两个主成分 2DPCA 仅仅是在一维的 PCA 方法上略有提升，影响其误差的主导因素是图像中的离群噪音的部分。

## V. 分类对比

上述算法应用中，将 Yale 数据集以 5:6 划分为了训练集和测试集；后续分类任务中，将调整这一比率。

现利用训练集中提取的主成分向量对测试集图像进行分类。采用如下近邻分类器，一维 PCA 用欧式距离 (Euclidean distance) 分类，对二维特征矩阵采用定义的如下矩阵距离：

- Define the distance btw. two arbitrary feature matrices,  $B_i = [Y_1^{(i)}, Y_2^{(i)}, \dots, Y_d^{(i)}]$  and  $B_j = [Y_1^{(j)}, Y_2^{(j)}, \dots, Y_d^{(j)}]$ , as

$$d(B_i, B_j) = \sum_{k=1}^d \left\| Y_k^{(i)} - Y_k^{(j)} \right\|_2,$$

where  $\left\| Y_k^{(i)} - Y_k^{(j)} \right\|_2$  denotes the Euclidean distance btw. the two principal component vectors  $Y_k^{(i)}$  and  $Y_k^{(j)}$

- The training samples are  $B_1, B_2, \dots, B_M$  from  $x\_train$ . Each of these samples is assigned a given class  $\omega_k$  from  $y\_train$ .
- Given a test sample  $B$  from  $x\_test$ , if  $d(B, B_l) = \min_j d(B, B_j)$  and  $B_l \in \omega_k$ , then the resulting decision is  $B \in \omega_k$

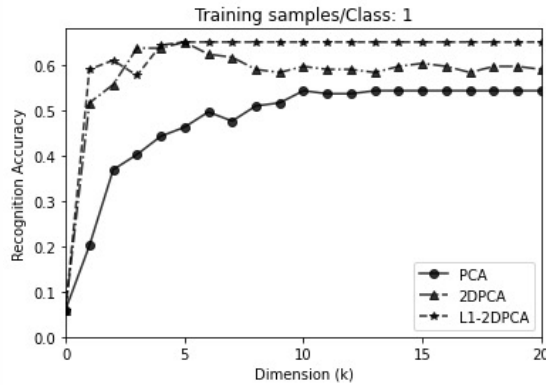
现对每一类图像取一个随机的图形作为训练集，获得如下  $1 \times 15$  的训练集：

图6: Yales 人脸数据训练集



应用三种算法到训练集，得到前 10 个特征向量进行样本外预测，结果如上图 7 所示。容易看出 2DPCA 和 L-2DPCA 算法的识别结果相似，但 L1-2DPCA 表现更加稳定——2DPCA 在主成分向量较多（多于 5 个）时表现下降。PCA 的识别准确率则相对较低，最终稳定在  $0.5 \sim 0.6$  之间。因此，可以认为，在分类任务中，L1-2DPCA 优于 2DPCA 优于 PCA。

图7: 外样本预测精度



后续实验的第  $k$  个测试中，使用每个类别的前  $k$  个图像作为训练数据，剩余的作为测试数据，计算三种算法降维后的识别准确度，其中  $k$  依次取 2, 3, 4, 5，结果见下图 8 与表 2. 可以看出，随着训练集的扩大，三种算法的外样本分类准确性均提高了。囿于算力所限，本次报告中仅计算了前 20 个特征向量的情况，后序的趋势中也可能出现更高的准确率。但是，我们注意到，随着训练集的丰富以及提取特征维度的增加，带来的分类准确性提升已经相当有限：当每个类型取 5 个样本进入训练集时，三种算法的最高分类准确率都在 80% 以上；就最优识别率而言，2DPCA 和 L2-2DPCA 难分伯仲。当然，在实际预测任务中，无法确定外样本预测时应该选择的特征维数，因此 L2-2DPCA 仍具有 2DPCA 无可比拟的稳定性优势。

图8：外样本预测精度（不同大小的测试集）

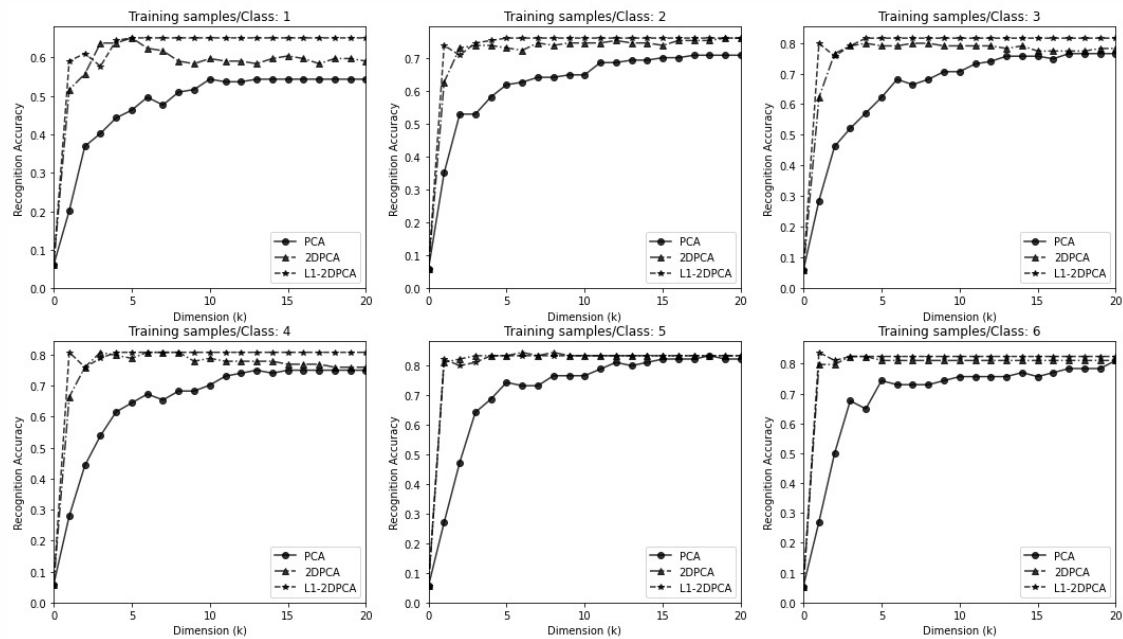


表2：最优分类准确度（%）对比

sample per class	1	2	3	4	5	6
PCA	54.36 (10)	70.90 (17)	76.47 (17)	75.00 (13)	83.15 (18)	81.08 (20)
2DPCA	65.10 (96,5)	76.12 (96,19)	79.83 (96,4)	80.77 (96,3)	84.27 (96,6)	82.43 (96,3)
L1-2DPCA	65.10 (96,5)	76.12 (96,5)	81.51 (96,4)	80.77 (96,1)	83.15 (96,4)	83.78 (96,1)

遮挡物干扰下的分类表现

理论上，L1-2DPCA 在离群噪音存在下表现较好；在文献 [2] [3] 中，使用类似二维码的黑白各点矩形对图像进行了遮挡，分类任务的结果显示出 L1 算法相对于 L2 具有更高的准确率。为了更好地对比算法在类似的离群噪音影响下的表现，在原有 Yale 人脸图像上随机生成矩形遮挡物，效果如下图 9 所示。

在人脸识别的分类测试中发现，遮挡物的颜色会对算法的表现产生影响。因此，分别选取黑、白、灰三种颜色的矩阵在相同的随机位置对数据集进行遮挡，每个类别分别取  $k = 1, \dots, 6$  个图像进入训练集，使用上文描述的近邻分类器判断测试集图像所属的类别。使用前 20 个特征向量投影后的最佳准确率见下表 3，括号内是达到最佳值时选用的特征向量维数。

图9: 遮挡后的 Yales 人脸数据集示例

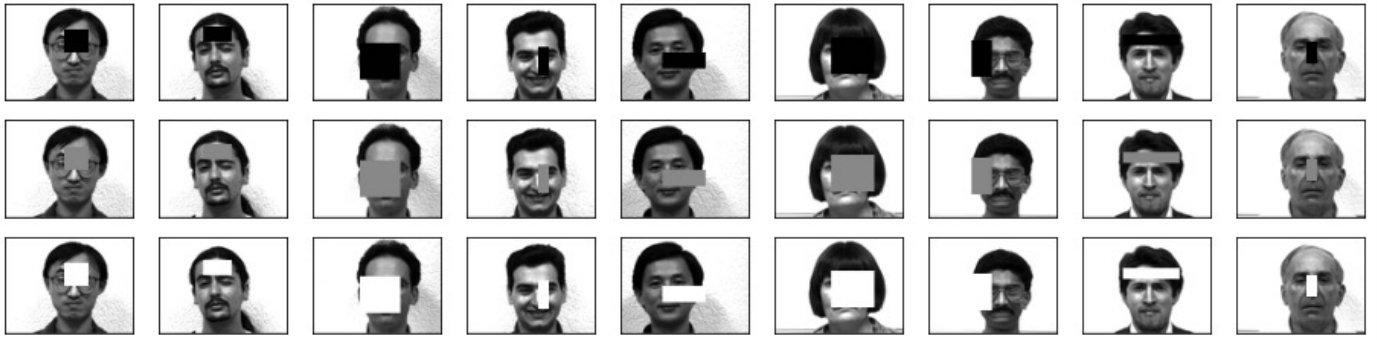


表3: 不同遮挡颜色下各算法的分类表现

sample per class	1	2	3	4	5	6
<i>Black Occluder</i>						
PCA	51.01 (13)	58.21 (13)	64.71 (18)	66.35 (6)	69.66 (13)	68.92 (20)
2DPCA	53.69 (96,17)	67.91 (96,20)	71.43 (96,14)	73.08 (96,13)	79.78 (96,13)	78.38 (96,4)
L1-2DPCA	48.99 (96,5)	60.45 (96,4)	68.07 (96,4)	72.12 (96,4)	77.53 (96,4)	78.38 (96,4)
<i>Gray Occluder</i>						
PCA	55.03 (14)	70.90 (19)	76.47 (17)	75.96 (18)	82.02 (18)	79.73 (19)
2DPCA	62.42 (96,4)	75.37 (96,7)	80.67 (96,3)	79.81 (96,7)	84.27 (96,7)	82.43 (96,4)
L1-2DPCA	62.42 (96,4)	75.37 (96,5)	79.83 (96,5)	78.85 (96,3)	82.02 (96,5)	81.08 (96,1)
<i>White Occluder</i>						
PCA	46.98 (14)	60.45 (20)	65.55 (19)	67.31 (19)	70.79 (19)	71.62 (14)
2DPCA	51.68 (96,13)	70.90 (96,19)	75.63 (96,14)	72.12 (96,13)	79.78 (96,20)	78.38 (96,13)
L1-2DPCA	44.97 (96,3)	61.19 (96,4)	66.39 (96,4)	66.35 (96,4)	71.91 (96,4)	75.68 (96,4)

对比上表中的算法表现, 在有遮挡的情况下, L1-PCA 在前 5 个主成分向量上已经达到了最优的准确率, 而PCA 和 2DPCA 则需要在较多的主成分上展开才到达最为准确的预测, 说明 L1- 算法受到离群噪音的干扰小, 表现稳定。

但是, L1- 算法的效果在特定类型的遮挡下并不一定是最佳的。下图 10 - 图 12 描述了取前  $k \leq 3$  个特征向量时的准确率。当遮挡物颜色为纯黑色 (`#000000`) 和纯白色 (`#ffffff`) 时, 2DPCA 所能实现的分类准确率高于 L1-2DPCA, 且随着主成分向量数的增长继续保持递增趋势。这一现象与参考文献 [1] [2] 是矛盾的吗?



图10：黑色遮挡物下的分类精度

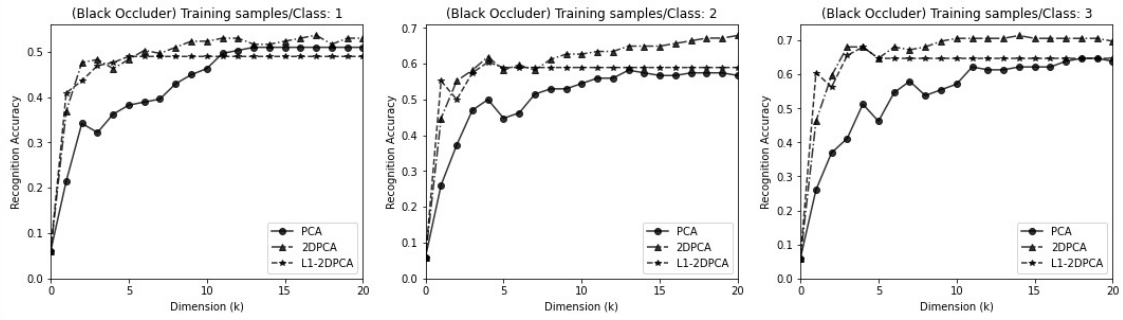


图11：灰色遮挡物下的分类精度

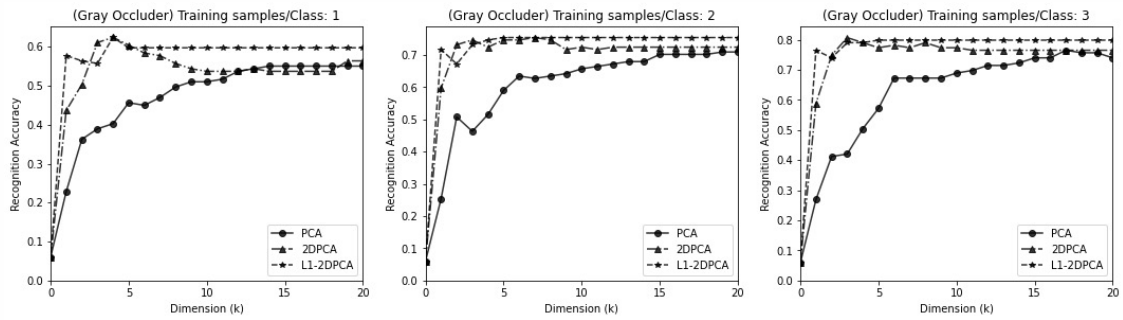
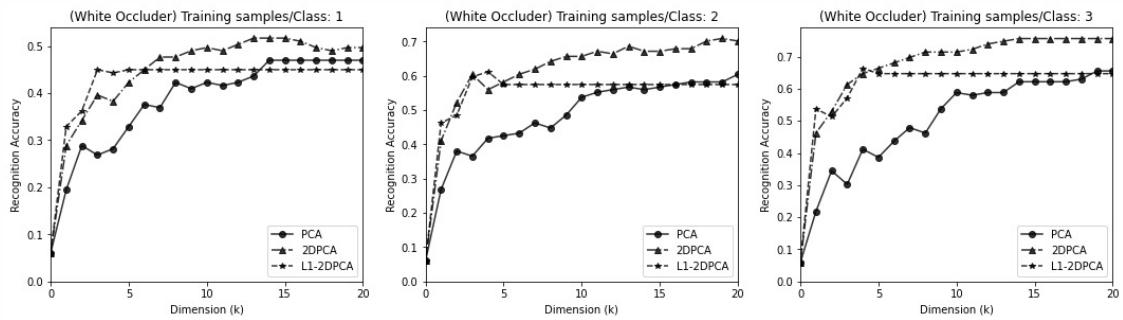
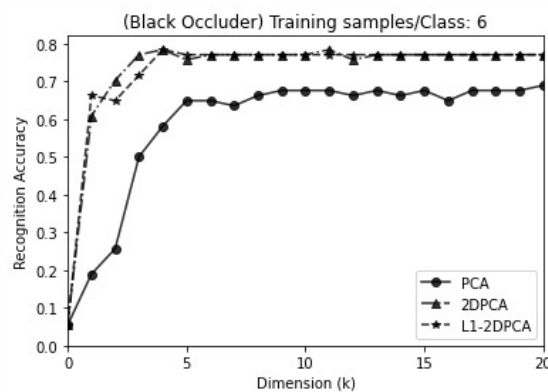


图12：白色遮挡物下的分类精度



需要指出，白色是该数据集的背景颜色，而黑色则与人脸中头发、眼珠的成分接近。因此，在每个类型进入训练样本的数量  $k$  较少时，L1- 算法会将遮挡物的位置分布也作为人脸特征的一部分；而在每个人脸类型进入训练样本的数量  $k$  增大后（如上表），2DPCA 与 L1-2DPCA 的分类表现接近。下图描述了黑色遮挡物（图13），每个分类抽取 6 个图像进入训练集的测试结果。

图13：黑色遮挡物下的预测精度（前 6 个主成分）



在所选取的遮挡物与原始图像中的元素有较大差异时，例如参考文献中黑白点阵纹理和本实验 Gray Occluder 的遮挡物，可以发现 L1- 算法的分类表现优于 L2- 算法。之后的探索中，可以尝试旋转、位移、白噪音、拼接等其他的离群干扰方案，对比 L1- 与 L2- 两种类型算法的表现。

## VI. PCA, 2DPCA 及 L1-2DPCA 算法各自优劣势

### PCA

优势：

- 应用广泛，在降维和误差校正中有较为广泛的应用；
- 当样本量大而单个样本的信息量较小时，PCA 的降维效果往往较好：因为 PCA 能够兼顾所有像素点的特征，即使不同类别的样本的特征差异较大，例如 MNIST 数据集；而在 2DPCA 中，由于对全体样本的列（或行）计算特征投影，其中求均值的过程会丢失各类型在列（或行）上的细节特征；

劣势：

- 在图像数据矩阵较大而样本数量相对较小的任务中，将很难计算出高纬度的协方差矩阵，难以保证计算的精确度；
- 此外，对较大的二维数据，拉伸至一维将使得协方差矩阵计算量庞大，算法的效率较低；
- 当存在离群值（噪音）时，例如污染、遮挡、阴影等，很难提取数据的真实特征，效果较差；

在单一图像的特征提取中，PCA 方法通常是便捷有效的。但是，在本报告涉及的聚类 and 分类等任务中，由于算法需要将二维空间矩阵拉伸成一维向量处理，使得最后的特征空间损失了像素点之间的位置关系，最终的表现较差。

### 2DPCA

优势：

- 更精确——在图像样本量小而单个样本数据大的情况下（例如人脸识别），直接提取较小的二维矩阵特征，可以精确地计算出协方差矩阵，保护总体方差；
- 由于计算量小，效率更高——协方差矩阵维数较小，缩短了计算特征值和特征向量的时间。例如，若图像的大小为  $d = h \times w$ ，对 PCA 而言，协方差矩阵大小为  $d \times d$ ，但是对于 2DPCA 而言，协方差矩阵大小为  $w \times w$ ；
- 图像重建的效率高、效果好——从表 2 中可以看出，2DPCA 特征向量在分类任务达到最优时所用的特征向量数远少于 1D 的情况，使得重建计算量更小；L1-2D 的情况下这一优势更加明显；

劣势：

- 并不能够提取对人脸识别非常重要的局部特征；
- 在图像中存在离群值（噪音）的情况下效果较差。

### L1-2DPCA

优势：

- 由于使用了 L1-norm-based matrix，对离群值 (outliers) 噪音的鲁棒性 (robustness) 好；
- 由于使用了二维的表示形式，减少了原始图像信息的损失；避免了大量的奇异值计算，迭代过程比较简单，保证算法效率和重构的效率；

劣势：

- 计算效率不如 2DPCA：在训练样本较多时 L1-2DPCA 需要迭代运算，速度较慢（见下表 3）；
- 性能不稳定：只在噪音信息与原图像其他部分特征对比明显的情况下优于 L2PCA，当噪音信息与背景或人脸有相似性，且单个类型进入训练集的数量较少时（在上一部分有详细讨论），L1-2DPCA 很难区分噪音特征与人脸特征；
- 需要迭代，存在初值设定等问题：可能会收敛到局部最优，在分类任务中最终稳定的准确率水平可能不如 2DPCA；
- 文献 [2] 中计算效率较高的迭代方法并没有完备的数学性质：将对偶问题直觉地转化，实际上并不是等价。

表4：Yale人脸数据15\*72\*96中提取特征向量的耗时（CPU: Apple M1）

sample per class	1	2	3	4	5
PCA	654 ns	653 ns	669 ns	658 ns	663 ns
2DPCA	9.47 ms	11.2 ms	11.8 ms	13.6 ms	14.1 ms
L1-2DPCA	8.96 ms	17.7 ms	28.6 ms	37 ms	54.1 ms

## Reference

[1] Yang, Jian, et al. "Two-dimensional PCA: a new approach to appearance-based face representation and recognition." IEEE transactions on pattern analysis and machine intelligence 26.1 (2004): 131-137.

[2] Li, Xuelong, YanweiPang, and Yuan Yuan. "L1-norm-based 2DPCA." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 40.4 (2010): 1170-1175.

[3] Kwak, Nojun. "Principal component analysis based on L1-norm maximization." IEEE transactions on pattern analysis and machine intelligence 30.9 (2008): 1672-1680.