

特征脸：Eigenface

周强

2017 年 7 月 20 日

特征脸 (Eigenface) 是指用于机器视觉领域中的人脸识别问题的一组**特征向量**。使用特征脸进行人脸识别的方法首先由 Sirovich and Kirby (1987) [1] 提出, 并由 Matthew Turk 和 Alex Pentland [2] 用于人脸分类。该方法被认为是第一种有效的人脸识别方法。这些特征向量是从高维矢量空间的人脸图像的协方差矩阵计算而来。这些特征脸本身形成了用于构造协方差矩阵的所有图像的**基础集合 (basis set)**。这样, 通过使用较小的一组基础图像 or 特征脸来表示原始训练图像, 从而减小尺寸。通过比较 faces 如何用基础集合表达, 从而实现分类。

1 历史

特征脸方法开始于探索人脸图像的低维表示。Sirovich 和 Kirby (1987) 表明, **主成分分析**可以用于收集人脸图像以形成一组基本特征。这些基本图像 (称为 Eigenpictures) 可以线性组合以重建原始训练集中的图像。如果训练集由 M 个图像组成, 主成分分析可以形成 N 个图像的基本集合, 其中 $N < M$ 。通过增加本征图像 (Eigenpictures) 的数量可以减少重建误差, 然而所需的数量总是选择小于 M 。例如, 如果需要为 M 个人脸图像的训练集生成 N 个特征脸 (eigenfaces), 则可以说每个脸部图像可以由所有这些 K 个”特征”或”特征脸”的比例组成: $Face\ image_1 = (23\% \text{ of } E_1) + (2\% \text{ of } E_2) + (51\% \text{ of } E_3) + \dots + (1\% \text{ of } E_n)$ 。

1991 年, M. Turk 和 A. Pentland 扩大了这些结果, 并提出了人脸识别的特征脸法 [3]。除了设计基于特征脸的自动人脸识别系统之外, 他们还显示了一种计算协方差矩阵的特征向量的方式, 使得当时的计算机可以在大量人脸图片上执行特征分解。人脸图像通常占据高维空间, 传统的主成分分析在这些数据集上是棘手的。Turk 和 Pentland 的论文展示了基于通过图片数量而不是图片像素数量大小的矩阵来提取特征向量的方法。

一旦建立, 特征脸法扩展到包括预处理方法, 以提高精度。**多重流形 (multiple manifold)**方法也用于构建不同对象和不同特征 (比如眼睛等) 的特征集合 [5] [6]。

2 生成特征脸

一组特征脸可以通过在一大组描述不同人脸的图像上进行**主成分分析 (PCA)** 获得。非正式地, 特征面可以被认为是一组 “标准化的面部成分”, 源自许多面部照片的**统计分析**。任意一张人脸图像都可以被认为是这些标准脸的组合。例如, 一张人脸图像可能**平均脸 (average face)** 加上特征脸 1 的 10%, 加上特征脸 2 的 55%, 再减去特征脸 3 的 3%。显然, 它并没有将许多特征脸组合在一起, 以实现公平的近似大多数面孔。另外, 由于人脸是通过一系列向量 (每个特征脸一个比例值) 而不是数字图像进行保存, 可以节省很多存储空间。

间。

创建的特征脸将显示为以特定图案排列的亮和暗区域。这种模式是一个面孔的不同特征被单独评估和评分。将有一种模式来评估对称性，如果有任何风格的面部毛发，发丝是哪里，或评估鼻子或嘴巴的大小。其他特征脸具有不太容易识别的图案，而特征脸的图像可能看起来很像脸。

用于创建特征脸和使用它们进行识别的技术也在人脸识别之外使用。该技术还可用于[手写分析](#)，[唇读](#)，[语音识别](#)，[手语/手势解释](#)和[医学影像分析](#)。因此，有些不使用术语特征脸，但更喜欢使用“特征图像”。

3 实际实现

要创建一组特征脸，必须：

1. 准备一个训练集的人脸图像。构成训练集的图片需要在相同的照明条件下拍摄的，并将所有图像的眼睛和嘴对齐。他们还必须在预处理阶段就重采样到一个共同的像素分辨率 ($R \times C$)。现在，简单地将原始图像的每一行的像素串联在一起，产生一个具有 $R \times C$ 个元素的行向量，每个图像被视为一个向量。现在，假定所有的训练集的图像被存储在一个单一的矩阵 T 中，矩阵的每一行是一个图像。
2. 减去均值向量。均值向量 a 要首先计算，并且 T 中的每一个图像都要减掉均值向量。
3. 计算协方差矩阵 S 的特征值和特征向量。每一个特征向量的维数与原始图像的一致，因此可以被看作是一个图像。因此这些向量被称作特征脸。他们代表了图像与均值图像 (mean image) 差别的不同方向。通常来说，这个过程的计算代价很高 (如果可以计算的话)。但是特征脸的实际适用性源于有效地计算 S 的特征向量的可能性，而无需明确计算 S ，如下所述。
4. 选择主成份。一个 $D \times D$ 的协方差矩阵会产生 D 个特征向量，每一个对应 $R \times C$ 图像空间中的一个方向。实际上，特征值对应着特征向量表示的方向上的方差。具有较大特征值的特征向量会被保留下来，一般选择最大的 N 个，或者按照特征值的比例进行保存，如保留前 95%。
5. k 是满足 $\frac{n(\lambda_1 + \lambda_2 + \dots + \lambda_k)}{n(\lambda_1 + \lambda_2 + \dots + \lambda_n)} > \epsilon$ 的最小整数，其中 k 是选择的主成分数目， n 是图片数目， ϵ 是设置在总方差上的阈值。

这些特征脸现在可以用于标识已有的和新的人脸：我们可以将一个新的人脸图像 (先要减去均值图像) 投影到特征脸上，以此来记录这个图像与平均图像的偏差。每一个特征向量的特征值代表了训练集合的图像与均值图像在该方向上的偏差有多大。将图像投影到特征向量的子集上可能丢失信息，但是通过保留那些具有较大特征值的特征向量的方法可以减少这个损失。例如，如果当前处理一个 100×100 的图像，就会得到 10000 个特征向量。在实际使用中，大多数的图像可以投影到 100 到 150 个特征向量上进行识别，因此，10000 个特征向量的绝大多数可以丢弃。

4 计算特征向量

直接在图像的协方差矩阵上进行 PCA 计算在计算量上是不可行的。如果图像比较小，如 100×100 的灰度图像，则每个图像是一个 10000 维空间的一个点，协方差矩阵 S 则具有 $10,000 \times 10,000 = 10^8$ 个元素。然而，协方差矩阵的秩受到训练图像的限制：如果有 N 个训练样本，则最多有 $N - 1$ 个对应非零特征值的特征向量。如果训练样本的数目比图像的维数低，则可以通过如下方法简化主成分的计算。

设 T 是预处理图像的矩阵，每一列对应一个减去均值图像之后的图像。则，协方差矩阵为 $S = TT^T$ ，并且对 S 的特征值分解为

$$Sv_i = TT^Tv_i = \lambda_i v_i \quad (1)$$

然而， TT^T 是一个非常大的矩阵。因此，如果转而使用如下的特征值分解

$$T^T T u_i = \lambda_i u_i \quad (2)$$

此时，我们发现如果在等式两边乘以 T ，可得到

$$TT^T T u_i = \lambda_i T u_i \quad (3)$$

这就意味着，如果 u_i 是 $T^T T$ 的一个特征向量，则 $v_i = T u_i$ 是 S 的一个特征向量。假设我们的训练集有 300 张 100×100 像素的图像，则 $T^T T$ 是一个 300×300 的矩阵，这就比原先的 $10,000 \times 10,000$ 协方差矩阵要容易处理许多。需要注意的是，上面的特征向量 v_i 没有进行归一化，如果需要，应该在后面进行处理。

5 Connection with SVD

令 X 表示 $d \times n$ 数据矩阵，其中列 x_i 表示减去均值后的图像矢量。那么，

$$\text{covariance}(X) = \frac{XX^T}{n} \quad (4)$$

Let the [singular value decomposition \(SVD\)](#) of X be:

$$X = U\Sigma V^T \quad (5)$$

那么 XX^T 的特征值分解为：

$$XX^T = U\Sigma\Sigma^T U^T = U\Lambda U^T \quad (6)$$

其中， $\Lambda = \text{diag}(\text{eigenvalues of } XX^T)$

因此，我们可以看到：

- The eigenfaces = U (the left singular vectors of X);
- The i th eigenvalue of $XX^T = \frac{1}{n}(\text{ith singular value of } X)^2$;

在数据矩阵 X 上使用 SVD，我们不需要计算实际协方差矩阵就可以得到特征脸 (eigenfaces)。

6 在人脸识别中的应用

特征脸的最直接的应用就是人脸识别。在这个需求下，特征脸相比其他手段在效率方面比较有优势，因为特征脸的计算速度非常快，短时间就可以处理大量人脸。但是，特征脸在实际使用时有个问题，就是在不同的光照条件和成像角度时，会导致识别率大幅下降。因此，使用特征脸需限制使用者在统一的光照条件下使用正面图像进行识别。

特征脸主要是尺寸减小方法 (dimension reduction method), 可以使用相对较小的数据集来表示目标。为了识别人脸, 脸部图像被保存为描述每个特征脸对该图像的贡献的权重集合。当将新的脸部图像呈现给系统进行分类时, 通过将图像投影到特征脸集合来找到自己的权重。这提供了一组描述该脸部图像的权重。然后将这些权重与图库集中的所有权重进行分类, 以找到最接近的匹配项。最近邻法是找到两个向量之间的欧几里德距离的简单方法, 其中最小值可以被归类为最接近的目标。

伪代码:

- Given input image vector $U \in \mathbb{R}^n$, the mean image vector from the database M , calculate the weight of the k th eigenface as:

$$w_k = V_k^T (U - M)$$

Then form a weight vector $W = [w_1, w_2, \dots, w_k, \dots, w_n]$

- Compare W with weight vectors W_m of images in the database. Find the Euclidean distance.

$$d = \|W - W_m\|^2$$

- If $d < \epsilon_1$, then the m th entry in the database is a candidate of recognition.
- If $\epsilon_1 < d < \epsilon_2$, then U may be an unknown face and can be added to the database.
- If $d > \epsilon_2$, U is not a face image.

7 回顾

Eigenface 提供了一种简单而便宜的方式来实现人脸识别:

- 其训练过程是完全自动的, 易于编码。
- 特征脸充分降低面部图像表现的统计学复杂性。
- 一旦计算了训练数据集的特征脸, 可以实时地实现人脸识别。
- Eigenface 可以处理大型数据集。

然而, 特征脸法的缺陷也是显而易见的:

- 对照明, 尺寸和平移非常敏感; 需要高度控制的环境。
- 特征脸难以捕获表情变化。
- 最重要的特征主要是关于照明编码, 并且不提供关于实际面部的有用信息。

为了在实践中应对照明干扰, 特征脸方法通常丢弃数据集中的前三个特征脸。由于照明通常是面部图像最大变化背后的原因, 前三个特征脸将主要捕获三维照明变化的信息, 对脸部识别几乎没有贡献。通过舍弃这三个特征面, 人脸识别的准确性将会有相当大的提升, 但是 Fisherface 和 Linear 空间等其他方法仍然比特征脸法更具有优势。

参考文献

- [1] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A*, 4(3):519–524, Mar 1987.
- [2] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, pages 586–591. IEEE, June 1991.