

使用特征脸方法进行人脸识别

——《计算机组织与科学计算》期末论文

朱恬骅

09300240004 计算机科学与技术

2011年11月

说明 本文中楷体字部分系*Face Recognition Using Eigenfaces*[4]一文的翻译或转述，文中次要的引文省略；宋体字部分系本人就人脸识别和特征脸方法这一论题的评述和论证。

1 问题描述

开发人脸识别的计算模型是很困难的，因为面部是十分复杂的、多维的、有意义的视觉刺激。它们自然而然是一个对象的类别，并且与其它在人类或计算机视觉研究中所采用的人造刺激物有着显而易见的不同。这样，我们不能像多数早期视觉功能那样，对视网膜（或传感器）活动建立一个周详的模型。

我们因而考虑发展一种初步的模式识别方法，它不依赖于包含全部信息的三维模型或细化的外观特征。我们的目的是开发一个快速、准确、合理而简单的人脸识别计算模型，适用于受限的环境（比如办公室或室内）。

虽然人脸识别是一个高层次的视觉问题，仍然有一些结构要求加于这一任务上。基于信息论的方法，我们提出一种利用了这种结构信息的框架，即，对一组人脸图像，寻找一种最为相关的信息编码方法，以便最大程度上将它们区分开来。这种方法将人脸图像转化为一组较小的、包含特征特性的图像，称为“特征脸”（eigenface），世纪也就是对初始的训练图像作主元分析所得到的结果。识别过程通过将一个新图像投影到由特征脸张成的子空间（“脸空间”）上，然后通过比较投影点与各已知人物的人脸投影的位置来获得结果。

现在，经过二十年的发展，人脸识别技术已经有了很大的不同。以往被认为低效且实现复杂的神经网络方法，[6]开始获得了实时学习的能力，并且速度也得到了很大的提高。此外，一批沿着特征脸思想而开发出的新的、

较为简易的识别方法开始涌现，这当中包括了同样较早的Fisherface（“费舍尔脸”）[1]和现在较近的Laplacianface（“拉普拉斯脸”）方法[3]。

1.1 背景和相关工作

许多使用计算机进行人脸识别的相关工作都将注意力集中在使用个体的特征检测，比如眼、鼻、口和头部轮廓，并使用这些特征的相对位置、大小和关系来定义面部模型。在Bledsoe 和Kandade 的早期系统中，一些自动和半自动的人脸识别策略，基于特征点的归一化的位置和比例，对人脸进行建模和识别。最近这种一般的方法由Yuille 等人延续和改进。

这一类方法被证实难以扩展到多视角情况下，并且通常十分脆弱。对人类进行人脸识别的策略的研究，证明了个体特征和它们表面上的关系对成年人进行面部识别来说并没有起到太大作用。然而，这种方法仍然是目前计算机视觉文献中最流行的方法。

面部识别的连接（connectionist）方法则力求捕捉这一任务中构型的、或类似“格式塔”（gestalt）的特征。Fleming 和Cottrel，基于Kohonen 和Lahtio 更早的工作，使用非线性单元来训练一个后向传播的网络来分类人脸图像。这一方法在Stonham 的WISARD 系统中，在人脸的二元识别方面得到了一些成功。大多连接（connectionist）系统将输入图像作为一个二维模式来处理，对面部的构型特征并没有显式地加以应用。目前这一类系统还处于非常简单的阶段，并且不清楚它们能否被应用到更大规模的问题上。

Burt 等人的最近工作使用了基于多分辨率模板匹配的“智能感知（smart sensing）”方法。这一逐层细化的策略使用一台专门的计算机快速计算多个分辨率下的图像，这一方法在对人进行识别的演示中达到了近乎实时的效果。它所采用的面部模型是从面部图像中手工建构的。

这一部分回顾了1991年及以前的人脸识别方法。这些比较简单的方法而今早已不再流行，尤其是当前模式识别越来越自动化，那些基于模板匹配的、或需要人工参与的建模过程早已淘汰。然而这一较为全面的综述为后代的研究者提供了简明的历史追溯。事实上，对人的面部器官提取特征，并计算它们的相对大小和位置等方法，现在仍有其实际意义。例如，在[5]、[2]等工作中，使用对眼、鼻、口等器官进行测量，对面部进行三角剖分的方法进行建模。这一建模结果被应用在了计算机辅助化妆、摄影等诸多领域。

2 使用特征脸进行识别

自动人脸识别领域先前的许多工作忽视了一个问题，那就是人脸刺激的哪些部分对识别特别重要。这提示我们，对人脸图像进行编解码的信息论

方法，在这一任务上可能有不错的前景。通过从人脸图像中提取一些得到加强的特征，我们或许能够较好地解决识别的问题，尽管这些特征不一定和我们对面部特征的直觉感受相同。

用信息论的语言来说，我们想要从面部图像中抓取出相关的一组特征，将他们尽可能高效地进行编码，并且和数据库中一组已经编码的模型进行比较，找出最为相似的那一个。一种简单提取信息的方法，就是在一组面部图像中，寻找某种发生变化的图像，并且这些信息之间相互独立，以此来编码和比较个体的面部图像。

用数学的术语来说，我们希望找到人脸分布的主元，也就是一组人脸图像的协方差矩阵的特征向量。这些特征向量可以被看作是一组特征，它们刻画了人脸图像之间的相互差别。每一个（原始）图像的分布位置多多少少和每一个特征向量有关，所以我们能够将特征向量还原成图像显示出来，这就是我们称之为“特征脸”的原因。

训练集中的每张面部图片能准确地用特征向量的一个线性组合进行表示。可能的特征脸数目与训练集中图像的数目相同。然而面部图像也能够使用少数几个“最佳”的特征脸的线性组合进行近似——也就是对应特征值最大的那些，它们对应了在训练集中变化最大的那些特征。使用少数特征脸的主要原因是计算代价。最好的 M' 个特征向量张成一个 M' 维子空间，亦即所有图像中的一个“脸空间”。它们构成了特征脸分解的一组基向量。

使用特征脸的想法受到Sirovich 和Kirby 开发的一套技术的启发，他们运用主元分析（Principal Component Analysis, PCA）的方法来表示一组面部图像。这提示了我们，大多数面部图像能够通过一组较少的特征图像通过加权的方法得到重建，所以，从已知面部图像中构造一组特征图像，通过比较目标图像和已知图像的权值，可能是一种较为有效和高效的人脸识别方法。

识别过程如下所示：

1. 初始化：获取面部图像训练集，计算特征脸，以此定义脸空间
2. 遇到新的面部图像时，计算基于输入图像和 M 个特征脸的一组权值
3. 检查图像是否足够接近“脸空间”，来判断这一图像是否是人脸
4. 如果是人脸，按权值模式进行分类，确定是哪个人的人脸，或是未知的结果
5. （可选）如果相同的未知脸出现了多次，计算它的特征权值模式，加入到已知的人脸集合中（也就是学习识别它）

2.1 计算特征脸

令图像 $I(x, y)$ 为一个存有灰度值信息的 $N \times N$ 的数组，亦即一个 N^2 维的向量，可以看作是 $\mathbb{Z}_{256}^{N^2}$ 空间中的一个点。人脸图像在这一空间中不是随机分布的，因而可以用一个较低维度的空间代替。使用主元分析寻找最能反映给定图像集特点的特征向量。这些向量都是 N^2 维的，称为“特征脸”。它们构成了一组正交基，组成“脸空间”。

计算特征脸的步骤具体如下所示：

1. 读取训练图像

设输入图像为 $\Gamma_1, \Gamma_2, \dots, \Gamma_M$ ，这一训练集的均值为 $\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i$
记输入图像与平均图像的差值为 $\Phi_i = \Gamma_i - \Psi$

2. 计算协方差矩阵：

$$C = \frac{1}{M} \sum_{i=1}^M \Phi_i \Phi_i^T = AA^T \quad (1)$$

其中 $A = [\Psi_1, \Psi_2, \dots, \Psi_M]$

3. 计算矩阵 C 的特征向量和特征值

根据定义，矩阵 C 是 $N^2 \times N^2$ 大小的，计算如此之大的矩阵的特征向量将花费很长时间。由于我们只需要对应特征值较大的那一部分特征向量，可以计算 $A^T A$ 的特征值和特征向量来代替。

证明：记 $\hat{C} = A^T A$ ， μ_i 为 \hat{C} 、 λ_i 为 C 的第 i 大特征值，它们对应的特征向量分别为 u_i 和 v_i 。则有：

$$A^T A u_i = \mu_i u_i \quad (2)$$

两边同左乘 A ，得到：

$$(A A^T)(A u_i) = \mu_i (A u_i) \quad (3)$$

可以看出， $\mu_i = \lambda_i$ ，且 $v_i = A u_i$ 。

这样，我们就将 $N^2 \times N^2$ 矩阵求特征值和特征向量的问题转化为了求一个 $M \times M$ 矩阵的问题。显然， $M \ll N^2$ 。求得的 v_i 即为特征脸。由于此协方差矩阵是实对称阵，它的特征向量相互正交，所以也就是一组基。

通常背景应当从训练图像中去除，从而特征脸中，面部以外的区域中应该都是零值。

2.2 使用特征脸分类人脸图像

一个新的人脸图像 Γ 通过下面这一简单的操作变换到特征脸组成（也就是投影到脸空间）： $\omega_k = \vec{v}_k^T(\Gamma - \Psi)$, $k = 1, \dots, M'$ 。向量 $\Omega^T = [\omega_1, \omega_2, \dots, \omega_{M'}]$ 中的权重描述了每一个特征脸在表示输入图像中的贡献，将这组特征脸作为一组基。这一向量被用来在一组已知的人脸类别中寻找（如果有的话）中最为接近的那一类 k^* ，亦即 $k^* = \arg \min_k \epsilon_k$ ，其中 $\epsilon_k = \|\Omega - \Omega_k\|$ 。设置阈值 θ_ϵ ，若最小的 ϵ_{k^*} 大于这一数值，则认为这一人脸图像结果为“未知”。

在实践中，设有 K 个人，每个人有 n_k 张训练图像 $\Pi_{k,i}$ ($k = 1, 2, \dots, K, i = 1, 2, \dots, n_k$)，在使用 Π 计算特征脸的同时，计算每个人自己的平均脸 $\Omega_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \Pi_{k,i}$ 。

2.3 对原始方法的改进

上述提出的原始方法有几个缺点：首先，如果新增加了一个识别对象（人脸），重新进行训练时不仅需要新对象的人脸图像，还需要保存原来所有对象的人脸图像，造成了重复的计算；其次，出于同样的原因，当识别对象增加时训练的时间将大大增加。这将不利于在实践中随时增减识别对象。

为了解决上述这两个缺点，考虑由 M' 个特征脸张成的脸空间。令重建的图像 $\hat{\Omega} = \sum_{i=1}^{M'} \omega_i \vec{v}_i = \sum_{i=1}^{M'} \vec{v}_i \vec{v}_i^T (\Gamma - \Psi) + \Psi$ ，残差 $r = \|\Omega - \hat{\Omega}\|$ 实际表示的是图像空间中源图像到脸空间的距离。

现在，我们对每个识别对象单独创建一个脸空间，那么对新的输入图像 Ω 的分类问题就转化为了寻找 $k^* = \arg \min_k r_k$ ，其中 r_k 为 Ω 到第 k 个脸空间的距离。图1展示了一个特征脸分解的例子。

组装矩阵 $W_k = [\vec{v}_1, \vec{v}_2, \dots, \vec{v}_{M'_k}]$ ，则 $r_k = \|W_k W_k^T (\Gamma - \Omega_k) - (\Gamma - \Omega_k)\| = \|(W_k W_k^T - I)(\Gamma - \Omega_k)\|$ 。

令 $n_k \equiv M/K$ ， $M'_k \equiv M/2K$ ，则对 K 个对象进行识别需要进行 $KN^2M/2K = N^2M/2$ 次乘法。即识别时的复杂度不变。在训练时，原方法需要计算一个 $M \times M$ 对称阵的特征向量，而现在需要计算 K 个 $M/K \times M/K$ 矩阵的特征向量，前者的复杂度为 $O(M^3)$ ，后者复杂度为 $O(M^3/K^2)$ ，得到了优化。并且，后者可以允许运行时动态地添加或删除识别对象或重新训练，而不必保存其它对象的训练集。进一步的实验也显示，这一改进方法的准确度要优于原始方法。



Figure 1: 改进的特征脸识别方法



Figure 2: ORL脸库中的例子

3 数值实验

3.1 测试集

本文使用ORL人脸数据库¹。该数据库包含了40个人的脸图片。每个人有十张不同时间拍摄的图像。图2显示了其中四个人的图片。在面部表情（比如睁眼闭眼、微笑与否）和面部细节（比如是否戴眼镜）上有所不同。所有这些照片都是在深色背景前拍摄的，人物基本上正面对照相机。

对于每个人，选取前五张照片为训练集，后五张照片为测试集进行测试。

¹<http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>

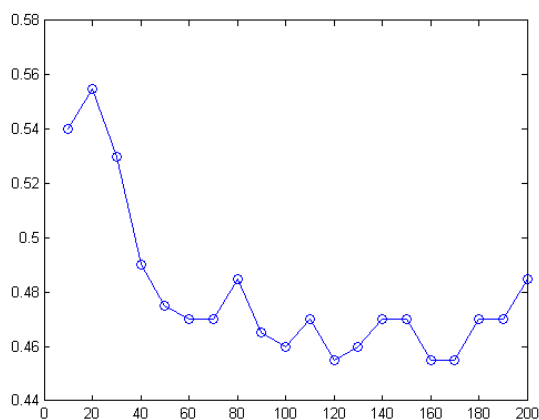


Figure 3: 原始方法特征脸个数与错误率关系图

3.2 特征脸个数的选取

图3显示了原始方法中特征脸个数和错误率之间的关系。可以看出，错误率并不随特征脸个数单调下降，在局部反而上升。这同时提示了过多的特征脸并不能显著改善原始方法的正确率。在此令接受阈值 $\theta_e = +\infty$ 。

3.3 改进方法的表现

图4展示了改进方法在ORL脸库上的表现。横轴表示训练的对象（类别）数，纵轴表示错误率。可以看出，改进方法较原始方法正确率得到了很大的提高。

3.4 一个实时识别的系统

使用OpenCV和Python进行编程。运行时界面如5所示。

4 结论

在思想上，特征脸方法实际上保存的是其作为“脸部”的特征而非差异。这是由PCA 方法本身的特性造成的。相较而言，基于线性判定分析 (Linear Discriminant Analysis, LDA)的Fisherface 方法能在ORL测试集上达到更高（约96%）的准确率。

特征脸方法的优点是显而易见的：它的实现十分方便。在实践中，特征脸方法的缺点有：

1. 对光线要求高

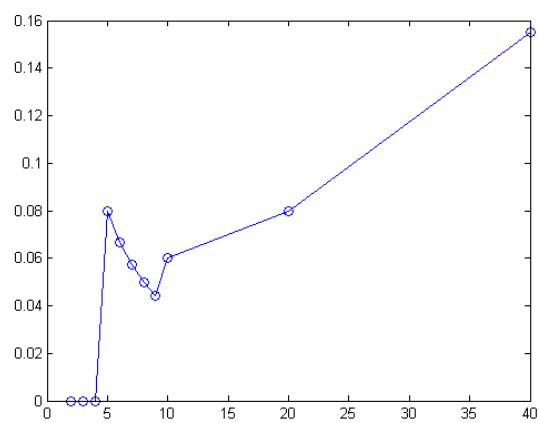


Figure 4: 改进方法的识别对象个数与错误率关系图

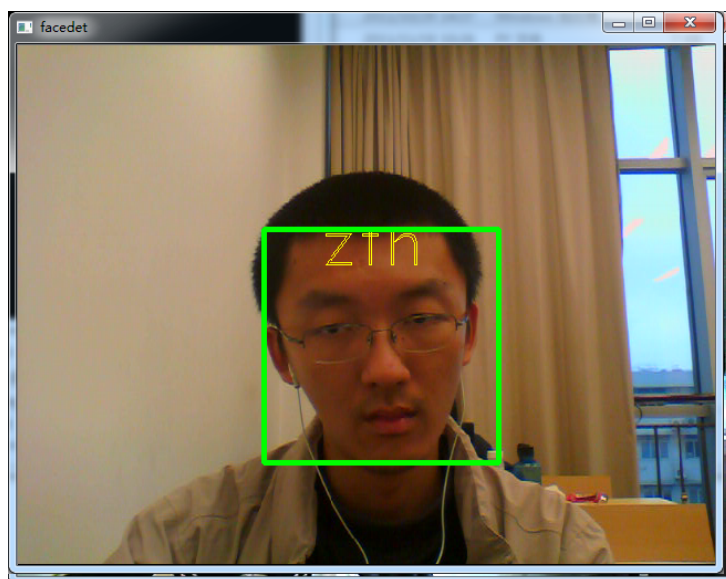


Figure 5: 实时识别系统的截图



Figure 6: 第一行：原始图像；第二行：未归一化情形下的重建图像；第三行归一化后的重建图像

光线变化将极大地降低识别准确率。可以对输入图像进行预处理，但效果有限。为了减少光线的区别，需要对图像进行归一化。图6中对比了是否进行归一化的结果。

2. 脸的转动（摆头和摇头）或表情变化会降低准确率

在原始方法中，令 $\theta_e = +\infty$ ，在光线变化的情况下得到了平均约96%的准确率，在方向变化上达到85%，在大小变化上为64%。

当令 θ_e 为某一有限值使识别准确率（不包括“未知”）达到100%时，识别为未知的比例在光线变化的情况下达到19%、方向变化为39%、大小变化为60%。

试验显示 θ_e 的增加将同时增加准确率。

参考文献

- [1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Patt. Anal. Mach. Intell.*, 19(7):711–720.
- [2] Yael Eishental, Gideon Dror, and Eytan Ruppin. Facial attractiveness: Beauty and the machine. *Neural Computation*, 18(1):119–142, 2006.
- [3] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, and Hong-Jiang Zhang. Face recognition using laplacianfaces. *IEEE Trans. Patt. Anal. Mach. Intell.*, 27(3):328–340, 2005.

- [4] M. Turk and A. Pentland. Face recognition using eigenface. In *IEEE CVPR*, pages 586–591, 1991.
- [5] David Zhang, Qijun Zhao, and Fangmei Chen. Quantitative analysis of human facial beauty using geometric features. *Pattern Recognition*, 44(4):940–950, 2011.
- [6] Jun Zhang, Yong Yan, and Martin Lades. Face recognition: Eigenface, elastic matching, and neural nets. *Proceedings of the IEEE*, 85(9), 1997.