

# 基于 HMM 的单样本可变光照、姿态人脸识别

胡峰松<sup>1)</sup> 张茂军<sup>2)</sup> 邹北骥<sup>3)</sup> 马俊容<sup>1)</sup>

<sup>1)</sup>(湖南大学计算机与通信学院 长沙 410082)

<sup>2)</sup>(国防科学技术大学信息系统与管理学院 长沙 410001)

<sup>3)</sup>(中南大学信息科学与工程学院 长沙 410083)

**摘 要** 提出了一种基于 HMM 的单样本可变光照、姿态人脸识别算法. 该算法首先利用人工配准的训练集对单张正面人脸输入图像与 Candide3 模型进行自动配准, 在配准的基础上重建特定人脸三维模型. 对重建模型进行各种角度的旋转可得到姿态不同的数字人脸, 然后利用球面谐波基图像调整数字人脸的光照系数可产生光照不同的数字人脸. 将产生的光照、姿态不同的数字人脸同原始样本图像一起作为训练数据, 为每个用户建立其独立的人脸隐马尔可夫模型. 将所提算法对现有人脸库进行识别, 并与基于光照补偿和姿态校正的识别方法进行比较. 结果显示, 该算法能有效避免光照补偿、姿态校正方法因对某些光照、姿态校正不理想而造成的识别率低的情况, 能更好地适应光照、姿态不同条件下的人脸识别.

**关键词** 三维人脸重建; 谐波基图像; 隐马尔可夫模型; 人脸识别

中图法分类号 TP391

DOI 号: 10.3724/SP.J.1016.2009.01424

## Pose and Illumination Invariant Face Recognition Based on HMM with One Sample Per Person

HU Feng-Song<sup>1)</sup> ZHANG Mao-Jun<sup>2)</sup> ZOU Bei-Ji<sup>3)</sup> MA Jun-Rong<sup>1)</sup>

<sup>1)</sup>(College of Computer and Communication, Hunan University, Changsha 410082)

<sup>2)</sup>(College of Information System and Management, National University of Defense Technology, Changsha 410001)

<sup>3)</sup>(College of Information Science and Engineering, Central South University, Changsha 410083)

**Abstract** In this paper a novel pose and illumination invariant face recognition algorithm that based on HMM with one sample per person is proposed. Firstly, by learning the train sets that has been fitted with Candide3 model, fitting algorithm automatically fitted the frontal facial input image with Candide3 to get the 3D shape. Then the specifically 3D face model is reconstructed by synthesizing texture to the fitted 3D shape. The new images under different pose can be generated by transform the 3D face model. Decompose the new images to 9 harmonic images' linear combination. By changing the 9 coefficients the new images under different illumination condition can be generated. All of these new images under different pose and different illumination condition make up the train set of individual face hmm. Experimental results show that this method can effectively avoid the recognition rate bring down caused by the pose and illumination normalization is not effective sometimes and can be better fitting the pose and illumination invariant face recognition.

**Keywords** 3D face reconstruction; harmonic image; hidden Markov model; face recognition

收稿日期: 2008-01-19; 最终修改稿收到日期: 2008-09-04. 本课题得到国家自然科学基金(60673093)、国家自然科学基金重大研究计划(90715043)和湖南省自然科学基金(07JJ3125)资助. 胡峰松, 男, 1969 年生, 博士研究生, 副教授, 研究方向为人脸识别、模式识别、图形图像处理等. E-mail: fshu@hnu.cn. 张茂军, 男, 1971 年生, 博士, 教授, 博士生导师, 主要研究领域为虚拟现实、信息系统工程等. 邹北骥, 男, 1961 年生, 博士, 教授, 博士生导师, 主要研究领域为 CAD&CG、数字图像处理等. 马俊容, 女, 1985 年生, 硕士研究生, 主要研究方向为图形图像处理、图像模式识别等.

## 1 引言

作为图像识别与理解研究领域的典型应用, 人脸识别技术经过近 40 年的研究, 取得了长足的发展. 从最初的基于单张图片提取几何特征(如两眼宽度、嘴角宽度等)的识别方法, 发展到基于表观技术、基于三维技术的人脸识别方法. 文献[1-2]分别给出了 1995 年与 2003 年以前的人脸识别领域较为完整的文献综述. 2007 年 3 月, 美国 NIST 报告了 2006 年人脸识别供应商评测(FRVT2006)结果<sup>[3]</sup>, 对控制光照条件下的极高分辨率正面人脸图像, 最小误识率为 0.001 时, 最小误拒率已达到 0.01, 对高分辨率、低分辨率下的正面人脸图像的识别, 这个数据也分别达到了 0.021 与 0.024. 在一定条件下, 有些技术甚至超过了人类的人脸识别能力<sup>[3-4]</sup>. 这表明, 在可控光照、用户配合的条件下, 人脸识别技术已经相当成熟. 但在非理想条件下(如光照可变、姿态可变、部分遮挡、表情变化等)的人脸识别性能还尚待提高. 在 FRVT2006 光照可变的评测中, 极高分辨率的正面人脸图像, 最小误识率为 0.001 时, 最小误拒率只能达到 0.103~0.130<sup>[3]</sup>. 相对于光照不可变的条件, 这个数据下降了一个数量级. 当同时考虑其它因素时, 识别性能会有更为显著的降低. 因此, 非理想条件下尤其是光照、姿态可变的人脸识别, 成为目前主要的研究热点.

人脸图像是人脸的内在属性——人脸 3D 形状和人脸光反射率在外界环境因素如光照、光源位置等条件下的成像结果, 它既包括了人脸的内在属性也包含了外界环境因素. 人脸识别系统就是要排除外界因素的干扰, 基于人脸的内在属性进行识别. 基于这一思想, 目前针对光照和姿态变化的人脸识别, 普遍采用“补偿”的方法, 将人脸图像通过某种变换, 将其转换到确定的光照与姿态条件下, 然后将其与样本图像比较以实现识别. 光照的补偿(归一)方法主要有两类, 一是利用数字图像处理技术(如直方图均衡化、直方图规定化、Gamma 灰度校正<sup>[5]</sup>、自商图像、相位图、相对梯度、对数边缘图等)对光照进行归一处理, 文献[6]对这些方法进行了较为全面的分析和比较. 另一类是基于球面谐波基图像的方法<sup>[7]</sup>, 卿来云等给出了一种基于球面谐波基图像的任意光照条件下的人脸识别方法<sup>[8]</sup>, 在人脸形状与光反射率大致相同的假设条件下, 该方法对光照进行估计, 依据光照估计对人脸图像进行光照补偿处理. 对于姿

态的影响, 一种主要的方法是基于三维人脸建模的技术<sup>[9-10]</sup>, 依据输入图像进行特定人脸 3D 形状重建以进行姿态校正. 在我们的前面工作中, 完成了利用 Candide-3 模型对人脸姿态进行校正的方法<sup>[11]</sup>, 在此基础上, 本文尝试另一种思路, 即不对待识别图像进行补偿(或归一、校正), 而是尽可能产生多种姿态与光照的图像, 并以这些图像建立该人脸的隐马尔可夫模型(Hidden Markov Model, HMM), 通过计算由模型产生待识别图像的概率来达到识别目的. 这也是为了适应单样本的要求而提出的方案.

“单样本”可以回溯到最初的基于人脸几何特征进行识别的年代, 该类方法一般都是针对单个人脸样本图像提取几何特征, 将提取的特征保存, 识别时将待识别图像相应特征与保存的特征进行比较完成识别任务. 随着技术的发展, 基于表观技术得到了更多的应用, 而这类技术通常都依赖足够多的样本对系统进行训练. 当只有一个样本时, 基于表观技术的人脸识别方法的识别性能都会有较大的降低, 有些甚至不能完成识别任务<sup>[12]</sup>. 实际上, 包含训练过程的系统都会受到这一问题的困扰. 针对“单样本”问题的人脸识别方法, 可以分为 3 类: 全局方法、局部方法及混合方法. 文献[12]对这一问题进行了较为完整的综述. 本文结合可变光照、姿态, 探讨了单样本限制下的采用混合方法的人脸识别问题.

## 2 系统框架

如前所述, 针对可变光照、姿态的人脸识别, 本文不采用通常的补偿(归一、校正)的方法, 即将输入的图像转换到确定光照与确定姿态下再进行识别, 而是以单一的样本图像为蓝本, 结合光照与姿态的改变, 产生多个光照与姿态不同的数字人脸, 以原始样本与产生的合成人脸作为训练数据, 产生该用户的人脸隐马尔可夫模型, 识别时再将输入图像与模型相比较, 完成识别任务. 该系统的框架如图 1 所示.

针对单样本问题, 主要有两种扩展训练数据的方法<sup>[12]</sup>, 一种是对该样本采用不同的表示方法, 即产生该样本的多种不同的表示. 如面向表示的组件分析(Representational Oriented Component Analysis, ROCA)方法<sup>[13]</sup>、图像扰动方法<sup>[14-15]</sup>等. 这种方法的不足在于, 产生的不同的表示不能作为独立的样本对系统进行训练, 它必需结合原始样本使用. 另外一种方法就是产生具有足够偏移的新的虚拟样本, 这些样本与原始样本在人脸空间中位于不同点. 产生这样的虚拟样本的最简单方法就是对原

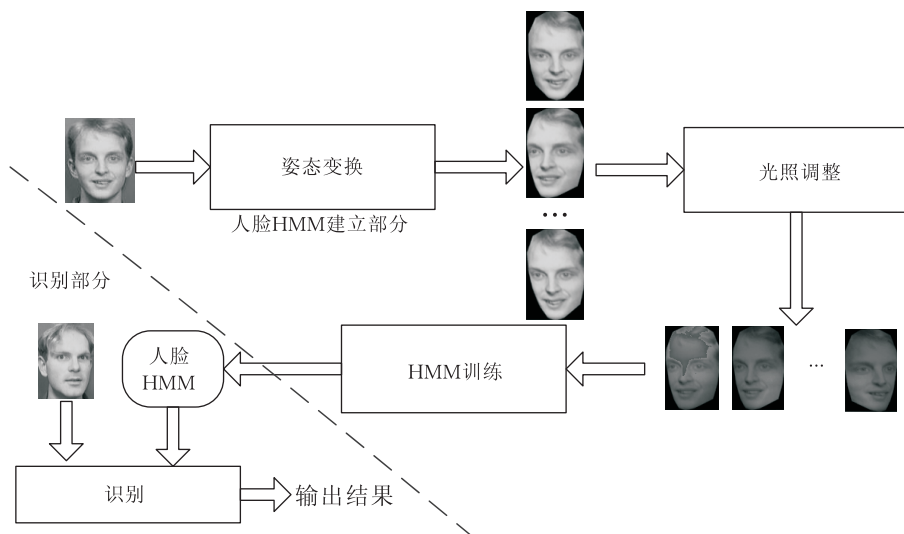


图 1 系统框架

始图像进行几何变换,如旋转、景深、双向对称变换等<sup>[16-17]</sup>.产生新的虚拟样本的更为复杂、也是更有效的方法是:结合人脸图像转换的先验知识产生出具有不同光照、不同姿态(甚至不同表情)的新的人脸图像. Beymer 等早在 1992 年就提出一种所谓的并行变形技术<sup>[18]</sup>,在 2D 图像上合成了新的具有不同姿态的图像.这种方法通过对另外的人脸图像训练集进行人脸图像转换操作的学习,获得相关的先验知识,然后将之应用到人脸转换以产生新的人脸图像.实验结果显示,在用虚拟人脸图像对系统进行训练以后识别正确率从只使用单样本进行训练时的 32.1% 上升到了 85%.但是这种方法也有其不足,首先它需要对转换操作进行学习,需要较多额外的训练集;其次它无法对所产生图像的姿态进行控制,另外,当变化模式数量增大时该方法需要产生更多的人脸图像,这给识别工作带来了更大的困难.对于不同光照的新人脸图像的产生,已有的方法如光照锥方法<sup>[19-20]</sup>等是建立在多个样本进行训练的基础上,无法应用到单样本问题中.

本文利用已得到较广泛应用的三维人脸模型——Candide3 模型建立特定人脸模型,以产生具有不同姿态的新的人脸图像.然后将利用球面谐波基图像方法对图像的光照进行调整以产生不同光照的新的人脸图像,从而可以产生不同光照、不同姿态的新的人脸图像.

### 3 不同光照、姿态的数字人脸的产生

#### 3.1 姿态的变化

借助人脸三维模型调整人脸图像中的姿态是很

自然的想法,得到特定人脸的三维模型后,对其进行各种角度的旋转变换即可获得各种不同的姿态.在我们前面的工作中<sup>[11]</sup>,提出了一种利用 Candide3 模型建立特定人脸的三维模型,并在此基础上对输入图像的姿态进行校正,然后进行识别的方法.本文也将利用 Candide3 模型,结合输入的正面人脸图像样本产生特定人脸三维模型,以完成姿态的变化.所不同的是,本文中不是将姿态校正到正面,而是将正面的人脸图像通过姿态调整得到其它不同姿态下的图像.此外,在该方法中,Candide3 模型与特定图像之间的配准借助 ASM 与 AAM(Active Appearance Model,主动表观模型)完成.ASM 与 AMM 分别是局部和全局寻优算法,均需较长收敛时间,两者的结合使得配准时间较长.为了减少配准时间,本文将采用一种具有更小计算复杂度的方法完成配准.下面首先介绍利用 Candide3 完成姿态调整的过程,然后详细介绍在只有一个输入样本的情况下,如何完成输入图像与 Candide3 之间的配准.

Candide 模型是一种参数化的模型,最初的目标是为了服务于基于模型的人脸编码应用,因为只有较少的若干顶点参数,使得应用该模型进行 3D 人脸重建只需要很少的计算时间,从而广泛应用于视频动画和视频传输中.Candide-3 是 Candide 模型的第 3 个版本,它由 113 个顶点和 168 个三角面片组成,图 2 给出了该模型的图示.模型由 113 个顶点的三维坐标给出,每个顶点  $P_i$  由其对应的 3 个坐标值  $x_i, y_i, z_i$  表示,即  $P_i = (x_i, y_i, z_i)^T, 1 \leq i \leq 113$ ,所有 113 个顶点的三维坐标串接而成的矩阵  $V$  给出了整个模型,即  $V = (P_1, P_2, \dots, P_{113})$ ,则  $V$  可以表

示为

$$\mathbf{V} = \mathbf{R}\mathbf{s}(\bar{\mathbf{V}} + \mathbf{S}\sigma + \mathbf{A}\vartheta) + \mathbf{t} \quad (1)$$

其中,  $\bar{\mathbf{V}}$  为标准模型;  $\mathbf{S}$  为形状单元, 它主要用来表示人脸的全局形状;  $\mathbf{A}$  为动作单元, 主要用来控制脸部动作和表情;  $\mathbf{R}$ ,  $\mathbf{s}$  和  $\mathbf{t}$  为全局相似变换控制参数, 即旋转、缩放及位移变换;  $\sigma$  和  $\vartheta$  分别为模型的静态控制参数和动态控制参数. 可以认为任何人脸所对应的 3D 形状模型  $\mathbf{V}$ , 都可以通过对这 5 个参数进行调整由标准模型  $\bar{\mathbf{V}}$  得到. 其中的  $\mathbf{R}$ ,  $\mathbf{s}$  和  $\mathbf{t}$  可以认为是姿态控制参数. 针对输入的图像  $I_i$  对标准模型  $\bar{\mathbf{V}}$  进行参数调整, 以使其各个顶点很好的拟合  $I_i$ , 即得到  $I_i$  所处姿态下的模型  $\mathbf{V}$ , 将  $\mathbf{V}$  与  $I_i$  进行纹理合成, 然后再对纹理合成后的  $\mathbf{V}$  继续进行姿态变换, 即继续调整姿态参数即可得到姿态校正后的合成纹理 3D 形状模型  $\mathbf{V}'$ , 即

$$\mathbf{V}' = \mathbf{R}'\mathbf{s}'\mathbf{V} + \mathbf{t}' \quad (2)$$

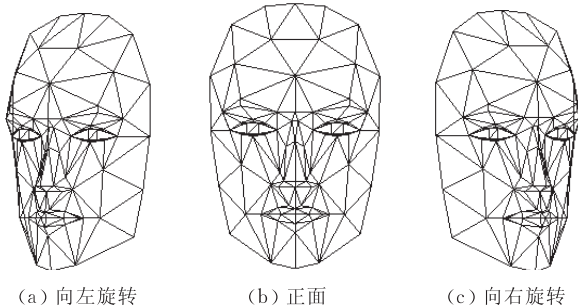


图 2 Candide3 模型

人脸图像与 Candide3 各顶点之间的配准是特定人脸三维重建的关键. 基于人脸图像差与人脸形状差之间存在近似线性关系, 即相似的人脸图像在很大程度上蕴含相似形状这一事实<sup>[21]</sup>, 可以通过对人工配准的样例的学习来完成人脸图像与 Candide3 模型之间的自动配准. 其基本思想是在对正面人脸图像进行大小、灰度归一化后, 输入图像  $I_o$  可以近似表示为训练集中所有图像的线性组合, 即

$$I' = \sum_{j=1}^m \omega_j I_j \quad (3)$$

其中,  $I_1, I_2, \dots, I_m$  为训练集中的图像,  $m$  为训练集中图像总数,  $\omega = (\omega_1, \omega_2, \dots, \omega_m)$  为线性组合系数矢量. 若设  $\omega^*$  为使  $\Delta I = I_o - I'$  取得最小值的线性组合系数, 则与输入图像  $I_o$  自动配准的模型  $\mathbf{V}$  为

$$\mathbf{V} = \omega^*(\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_m)^T \quad (4)$$

其中,  $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_m$  为训练集中各图像手工配准的模型. 结合式(1), 自动配准模型  $\mathbf{V}$  中所包含的参数  $\mathbf{R}$ ,  $\mathbf{s}$ ,  $\sigma$  和  $\mathbf{t}$  可分别按下式计算.

$$\begin{cases} \mathbf{R} = \omega^*(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_m)^T \\ \mathbf{s} = \omega^*(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m)^T \\ \sigma = \omega^*(\sigma_1, \sigma_2, \dots, \sigma_m)^T \\ \mathbf{t} = \omega^*(\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_m)^T \end{cases} \quad (5)$$

$\omega^*$  可以按如下方法计算求得. 首先将训练集图像矩阵数据转换为图像矢量数据, 训练集中的全部图像的矢量构成训练集图像矩阵, 记为  $\mathbf{A} = (a_{ij})_{m \times n}$ ; 然后将配准图像  $I_o$  转换为图像矢量, 记为  $\mathbf{B} = (b_1, b_2, \dots, b_n)$ ,  $n$  为  $I_o$  的像素总数. 则  $\omega\mathbf{A}$  为  $\mathbf{B}$  的线性近似, 两者之间存在误差:  $E = \|\mathbf{B} - \omega\mathbf{A}\|^2$ . 对  $\omega$  的求解转换为求解的最小化问题:

$$\omega^* = \min_{\omega} E \quad (6)$$

因此, 很容易得到

$$\omega^* = \mathbf{B}\mathbf{A}^\perp \quad (7)$$

其中,  $\mathbf{A}^\perp = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$  为  $\mathbf{A}$  的逆转矩阵. 图 3 给出了 orl 人脸数据库<sup>①</sup>中 5 个输入图像与该图像在训练集上线性组合图像的对比, 第 1 行为输入图像, 第 2 行为组合图像, 第 3 行为两者的差, 训练集由 orl 人脸库中 20 个人的正面图像组成. 给出的 5 个图像, 其中前 3 个图像在训练集中没有类内样本, 后两个图像在训练集中有其类内样本.



图 3 输入图像、线性组合图像及其差

求得  $\omega^*$  后, 即可用式(5)结合训练集中手工标定的配准模型的参数, 计算输入图像的配准模型的参数, 以完成输入图像与模型之间的自动配准. 图 4 给出了图 3 中 5 个输入图像的自动配准结果. 完成模型的自动配准后, 进行纹理合成即可得到该输入图像的特定三维模型, 然后对该模型进行旋转变换即可得到各种姿态, 图 5 给出了图 3 中第一个输入图像的 8 种不同的姿态.

① ORL face database. <http://www.cam-orl.co.uk/facedatabase.html>





图 4 5 个输入图像的自动配准示例



图 5 姿态变化示例

### 3.2 光照的变化

目前已有的产生光照不同的新的人脸图像的方法,如 ASM 方法、光照锥方法等,都是基于对同一个人的多个样本的训练来完成的. 本文将采用文献[8]介绍的基于球面谐波基图像的光照调整方法. 在文献[8]中,首先对图像的光照进行估计,然后再根据估计对光照进行补偿,以将图像的光照调整到确定的光照环境中. 因为我们不采用补偿的方法,而是根据输入的图像产生多个光照不同的数字人脸,因此可以省略其中的估计步骤,直接对 9 个光照系数进行调整来达到目的. 假设人脸是朗伯凸表面,则人脸图像可以表示为

$$I(x, y) \approx \sum_{l=0}^2 \sum_{m=-l}^l L_l^m \lambda(x, y) A_l Y_l^m(\alpha(x, y), \beta(x, y))$$

$$= \sum_{l=0}^2 \sum_{m=-l}^l L_l^m b_l^m(x, y) \quad (8)$$

其中,  $b_l^m(x, y) = \lambda(x, y) A_l Y_l^m(\alpha(x, y), \beta(x, y))$  为 人脸的球面谐波基图像.  $A_l$  为朗伯反射函数的球面谐波系数,  $Y_l^m$  为球面谐波函数,  $\lambda(x, y)$  为 人脸图像中某点的光反射率,  $\alpha(x, y)$ ,  $\beta(x, y)$  为 该点的法线方向,  $L_l^m$  为光照系数. 显然,求得各点的光反射率及法线方向即可求得人脸图像的 9 个球面谐波基图像,从而可以通过改变  $L_l^m$  来实现光照调整的目的. 在 3.1 节的描述中已经介绍了特定人脸 3D 模型的

建立方法,得到人脸图像的 3D 形状后就很容易计算各点的法线方向. 图 6 给出了图 3 中 5 个输入图像的几个光照、姿态不同的数字人脸示例.



图 6 不同姿态的光照调整示例

## 4 特征提取与人脸 HMM

单样本限制下,通常有 3 类识别方法,即全局方法、局部方法及混和方法. 全局方法一般是将人脸图像中各点的像素灰度值按照位置串接为一个特征矢量,这使得应用小样本训练高维度的特征矢量的问题——“维度灾难”问题更为加剧,此外,因为只有一个特征矢量,从而无法直接获取类内偏差信息,这两个不足是全局方法本身所无法克服的. 局部方法是使用局部脸部特征完成人脸识别的方法,相对于全局方法,局部方法更适合于单样本下的人脸识别,它有以下 3 个优点:(1)局部方法用低维度的特征矢量集表示原始人脸图像,有助于降低“维度灾难”问题的影响;(2)局部方法提供了利用某个局部区域完成人脸识别的可能,使得一些类内共有的特征更易被利用;(3)不同的局部特征有助于分类器的多样性,即为识别过程使用多种不同的分类器提供了可能,这显然有利于提高识别率. 尽管局部方法具有以上 3 点,但将局部与全局相结合的混合方法仍然较之局部方法具有更多的优势,因此在局部方法中或多或少都通过不同的形式将全局方法结合到其中. 本文通过建立人脸 HMM 完成这一任务. 下面先介绍局部定义及特征提取方法,然后介绍人脸 HMM 的建立以及 HMM 如何混合全局与局部方法.

### 4.1 局部定义与特征提取

对原始人脸图像及通过调整光照、姿态产生的

数字人脸图像进行划块,对每一块分别提取特征.如图 7 所示,将人脸图像划分为若干方块( $m$  行  $n$  列),针对每一个方块,将方块中的每个像素的灰度值按位置串接为局部特征矢量  $\mathbf{e}=(v_1, v_2, \dots, v_l)^T$ ,  $l$  为方块中像素总数,记图像中像素总数为  $p$ ,则  $l=p/(m \times n)$ .  $m, n$  的取值反映局部定义的粒度大小,实验中对  $m, n$  的取值进行了比较.所有方块的特征矢量构成特征矩阵  $\mathbf{E}=(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{m \times n})$ .

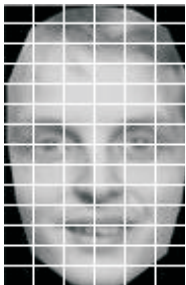


图 7 人脸图像的局部定义

## 4.2 人脸 HMM

HMM 作为典型的统计学模式识别方法,因为完善的数学理论的支撑和在自动语音识别系统中的成功应用,近十几年来越来越受到广泛关注,也已经有人尝试应用 HMM 完成人脸识别任务<sup>[22-24]</sup>.从理论上讲, HMM 所能代表的概率分布类型没有理论上的限制<sup>[10]</sup>. 一个恰当构建的 HMM 对类内偏移具有较强的鲁棒性,同时对类间模式的区别有较强的敏感性. HMM 是一个双重随机过程,其中之一是马尔可夫链,这是基本的随机过程,描述状态的转移,另一个是观测值在各状态中的分布概率,它描述状态和观测值之间的统计对应关系.用两个随机过程更能有效地描述客观世界中事件发生的复杂关系.按照观测值概率取值与状态转移描述方式的不同,可以分为离散 HMM (DHMM)、半连续 HMM (SCHMM) 与连续 HMM (CHMM). 相对来说半连续、连续 HMM 都能更细致地反映事件发生的关系,但也需要更多的训练用数据,本文采用 DHMM,记为  $\lambda=(\pi, \mathbf{A}, \mathbf{B})$ , 其中,  $\pi=(\pi_1, \pi_2, \dots, \pi_N)$  是初始概率矢量,  $N$  为状态数;  $\mathbf{A}=(a_{ij})_{N \times N}$  是状态概率转移矩阵;  $\mathbf{B}=(b_j(v_k))_{N \times M}$  是观测值概率矩阵.

HMM 具有多种不同的逻辑结构,在人脸识别应用中得到较多应用的主要是网状和链状结构. 本文采用“左右型”链状结构,并且对状态间的转移做了较为严格的限制——由左至右有跳转. 确定了

模型的状态转移的逻辑结构以后,还有一个重要的参数就是状态数的确定. 作为探讨,对所有人脸使用了相同的状态数,在实验中对不同状态数对性能的影响做了比较.

对一幅人脸图像(包括原始图像与产生的数字图像)完成特征提取后,图像被表示为  $\mathbf{E}$ , 对  $\mathbf{E}$  中的列向量进行向量量化得到该图像的观测值序列  $O=\{b_1, b_2, \dots, b_{mn}\}$ , 训练集(原始图像及产生的姿态、光照不同的人脸图像)中的全部观测值序列  $\{O_1, O_2, \dots, O_s\}$  ( $s$  为训练集中图像总数)作为训练数据,使用 Baum-Welch 算法对人脸 DHMM 进行训练. 用户在注册时输入的每人一幅图像,经过上述过程得到人脸 DHMM 库  $\{\lambda_1, \lambda_2, \dots, \lambda_l\}$ ,  $l$  为注册用户数.

识别时的任务就是对给定的观测值序列  $O$  和所有人脸 DHMM, 计算由各模型产生  $O$  的概率,并取其中最大概率对应的模型为识别结果. 文中采用前向-后向算法. 为了避免数值太过密集采用通常的做法——对计算的概率取自然对数,即

$$P' = \lg(P(O | \lambda)) \quad (9)$$

识别过程如下:

1. 对输入人脸图像  $I$  进行特征提取得到其对应  $\mathbf{E}$ ;
2. 对  $\mathbf{E}$  中的列向量进行向量量化得到该图像的观测值序列  $O=\{b_1, b_2, \dots, b_{mn}\}$ ;
3. 利用式 (9) 计算由人脸 DHMM 库中各人脸 DHMM:  $\lambda_i, i \in [1, L]$ , 产生  $O$  的概率:  $P'_i, i \in [1, L]$ ;
4. 取  $\{P'_1, P'_2, \dots, P'_l\}$  中最大值  $P'_j = \max\{P'_1, P'_2, \dots, P'_l\}$ , 如果  $P'_j < T$  ( $T$  为一确定阈值), 拒绝识别该图像, 否则  $j$  即为识别出的候选人编号.

## 5 识别实验

为了对所提识别方法的性能进行评价,我们在 orl 人脸库上进行了识别实验,并比较  $m, n$  以及状态数取不同值对识别方法的影响. 实验首先将人脸库中前 10 人的 100 个图像作为自动配准的训练集,对其进行手工 Candide3 模型配准. 然后以其他 30 人作为测试数据集,其中每个人的正面图像作为其产生不同姿态、光照数字人脸图像的样本,每个人的模型的训练图像总数为 100. 这样,需要识别的人脸图像总数为 270, 实验结果如表 1 所示,从表中可以看出,最好情况下,仅有 3 个图像识别错误,识别率为 98.9%.

表 1 取不同状态数及  $m \times n$  值时对 orl 库中部分图像的认识结果

状态数	识别率/%							
	$m \times n$ =10	$m \times n$ =20	$m \times n$ =30	$m \times n$ =40	$m \times n$ =50	$m \times n$ =60	$m \times n$ =70	$m \times n$ =80
3	78.9	83.9	79.9	81.8	84.6	79.3	76.0	75.8
4	79.5	84.2	81.5	88.3	87.9	81.2	76.5	77.3
5	78.2	84.3	84.2	88.1	85.4	88.6	78.3	77.4
6	76.1	84.1	85.8	90.2	90.5	94.5	78.2	76.6
7	77.6	82.9	85.8	90.3	94.3	96.5	83.9	78.1
8	74.5	81.1	80.5	86.5	93.1	<b>98.9</b>	91.7	78.6
9	74.4	82.3	83.8	87.1	93.0	94.9	83.1	78.6
10	74.8	79.0	77.4	84.5	90.3	<b>93.3</b>	<b>92.7</b>	78.2
11	72.3	79.1	78.9	83.6	85.4	94.4	88.3	80.0
12	70.1	76.7	75.2	82.3	88.5	<b>92.5</b>	84.3	79.1

$m, n$  的取值涉及处理问题的粒度选择策略. 一方面它与提取的特征矢量的维度大小密切相关, 另一方面它又与特征矢量序列长度, 即观测值序列长度密切相关, 而这又与模型状态数相关. 当  $m \times n$  的取值太小时, 必然使得特征矢量的维度增大, 序列长度减小. 极端情况下,  $m \times n$  取 1 时, 序列长度为 1, 特征矢量维度为图像像素总数, 局部方法退化为全局方法. 当  $m \times n$  的取值太大, 观测值序列过长, 必然使得模型必需有更大的状态数, 增加了模型的复杂程度, 从而使得模型的训练数据过大. 因此选择合适的  $m \times n$  值, 同时配合选择恰当的状态数能较大程度提高所提方法的识别性能. 表 1 中, 最优识别率出现在  $m \times n$  值为 60, 状态数为 6~8 的区域中充分说明了这一点.

此实验中, 我们同时记录了各个计算过程的运算时间. 具体说主要有以下两个部分中的共计 6 个过程. (1) 人脸 DHMM 生成部分. 主要包括以下 4 个过程: 配准、光照变化、数字人脸生成和人脸 DHMM 生成. (2) 识别部分. 主要包括以下 2 个过程: 配准和识别. 其中配准过程主要涉及到式(7)和式(5)的计算, 光照变化主要是式(8)的计算, 其中每个输入图像的 9 个球面谐波基图像的计算只需进行一次, 在已知 9 个基图像的基础上, 光照变化实际是调整光照系统的过程. 每个数字人脸生成过程的两个主要计算: 一是姿态变化也即是姿态参数的调整; 二是光照系数的调整, 每个输入人脸图像的  $n$  个数字人脸的生成是一个  $n$  次的循环. 人脸 DHMM 的生成过程涉及到特征矢量获取、矢量量化及利用训练数据进行模型的训练是所有过程中最为耗时的过程. 识别过程包括特征矢量获取、矢量量化及  $l$  次式(9)的计算. 表 2 记录了所有这些过程的平均耗时, 其中每个输入图像生成数字人脸图像数为 100,  $m \times n$  值为 60, 状态数为 6,  $l$  为 30.

表 2 各过程运算时间

	耗时/s		耗时/s
配准	0.012	人脸 DHMM 生成	49.18
光照变化	0.035	识别	8.16
数字人脸生成	1.780		

因为 orl 库中人脸变化模式少, 大多数系统的识别率均能达到 90% 以上. 为了进一步检验所提方法的性能, 我们对 CMU-PIE 人脸库<sup>[25]</sup>进行了识别实验, 并与文献[10]中采用的光照补偿、姿态校正方法的识别结果进行了比较. 为了便于比较, 首先对只有姿态变化的图像子集进行识别, 然后对姿态、光照均有变化的图像子集进行识别, 每个人的脸 DHMM 的训练图像总数仍为 100, 样本为闪光灯序号“11”的姿态 27 的均衡光照正面图像.  $m \times n$  值为 60, 状态数为 8. 自动配准的训练集使用 AR 人脸库中的均衡光照、自然表情的图像子集.

对 CMU-PIE 的 4 个姿态子集: 姿态集合 05、姿态集合 29、姿态集合 37、姿态集合 11 进行识别, 模型训练的数据不包含产生的光照不同的数字人脸图像. 识别结果及与文献[10]应用 Gabor PCA 和 LDA 相结合的识别结果列于表 3. 从表 3 数据可以看出, 尽管对每个用户只使用了一个样本, 识别性能相对采用姿态校正后的 Gabor PCA+LDA 识别方法的性能没有明显降低. 因为采用 HMM 模型避免了姿态校正对某些姿态校正效果欠佳的不足, 使得对 4 个姿态的识别正确率相对更稳定.

表 3 对 CMU-PIE 人脸库中 4 种姿态的识别结果比较

Pose set	识别率/%	
	本文算法	文献[10]中的算法
Pose29	97.3	97.1
Pose05	97.1	98.5
Pose11	96.3	95.6
Pose37	92.6	88.2

在模型训练的数据中加入产生的光照不同的数字人脸图像, 重新对所有模型进行训练, 然后对 CMU-PIE 数据库中的“illum”集合中的部分图像, 即姿态集合为 05、29, 闪光灯序号为 02~22 的图像集合, 共 2856 幅图像, 进行识别实验. 实验结果如图 8 所示. 全部 21 个闪光灯下图像的平均识别率为

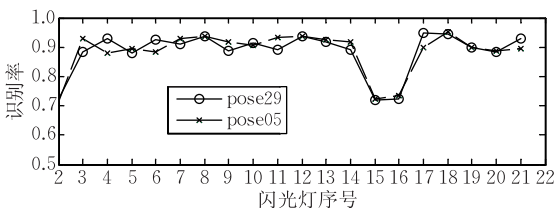


图 8 对“illum”子集的识别结果



80.3%。从图中可以看出,闪光灯 02、15、16 下的识别率较低只略高于 70%,但相对文献[10]中这 3 组图像的不到 50%的识别率还是提高了近 20%。

## 6 总 结

光照与姿态问题是目前人脸识别中亟待解决的关键问题,本文所提出的算法在单样本的限制下,利用特定人脸的 3D 模型,得到光照和姿态不同的新的人脸图像,再构建用户的独特人脸 HMM 以实现识别。将不同光照与姿态的人脸图像作为模型的训练数据,使模型中包含有不同光照、姿态下的观测值分布与状态转移的信息,从而能够有效地识别各种不同姿态与光照条件下的人脸图像。模型构建完成后,识别时无需对待识别图像进行姿态校正和光照补偿,显著降低了识别处理时间,更能适应于大规模、实时人脸识别应用。

在本文中直接将图像像素的灰度值作为获得观测序列的基础元素。实际上,产生的虚拟样本与输入的真实样本一样,都是由像素构成的人脸图像,因此可以提取到所有真实样本所能提取到的特征。如使用二维小波分析获取人脸特征表示向量的方法<sup>[26]</sup>、Zernike 矩特征方法、维数约减及特征提取方法(如 Principal Component Analysis(PCA), Independent Component Analysis (ICA), Fisher Discriminant Analysis(FDA))、广义散度差矩阵特征向量<sup>[27]</sup>等均可适用于虚拟样本。

因为有单样本的限制,在模型配准时只能通过对人工配准的样例学习来实现输入图像的自动配准,而且自动配准的精度,在较大程度上受到可供学习的人工配准样例的数量的限制,因此,今后的工作将主要集中于单样本限制下模型配准的研究工作中,以进一步提高识别的性能。此外,HMM 是一个较复杂的模型,针对不同的应用,有较多的参数需要进行设置,如何更有效地将 HMM 应用到人脸识别中将是今后工作的另一个重点。

## 参 考 文 献

- [1] Chellappa R, Wilson C L, Sirohey S. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 1995, 83(5): 705-741
- [2] Zhao W, Chellappa R, Phillips P J et al. Face recognition: A literature survey. *ACM Computing Survey*, 2003, (10): 399-458
- [3] Phillips P J, Scruggs W T, O'Toole A J et al. FRVT2006 and ICE2006 large-scale results. *FRVT2006: Evaluation Report*. March, 2007
- [4] O'Toole A J, Phillips P J, Jiang Fang et al. Face recognition algorithms surpass humans matching faces over changes in illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(8): 532-541
- [5] Shan Shiguang, Gao Wen, Cao Bo et al. Illumination normalization for robust face recognition against varying lighting conditions//*Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*. France, 2003: 157-164
- [6] Du Bo, Shan Shiguang, Qing Laiyun et al. Empirical comparisons of several preprocessing methods for illumination insensitive face recognition//*Proceedings of the ICASSP2005*, 2005: 981-984
- [7] Zhao W Y, Chellappa R. Illumination-insensitive face recognition using symmetric shape-from-shading//*Proceedings of the CVPR'2000*, 2000: 1286-1293
- [8] Qin Lai-Yun, Shan Shi-Guang, Chen Xi-Lin et al. Face recognition under varying lighting based on the harmonic images. *Chinese Journal of Computers*, 2006, 29(5): 760-768 (in Chinese)  
(卿来云, 山世光, 陈熙霖等. 基于球面谐波基图像的任意光照下的人脸识别. *计算机学报*, 2006, 29(5): 760-768)
- [9] Blanz V, Vetter T. Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(9): 1-12
- [10] Chai Xiu-Juan, Shan Shi-Guang, Qing Lai-Yun et al. Pose and illumination invariant face recognition based on 3D face reconstruction. *Journal of Software*, 2006, 17(3): 525-534 (in Chinese)  
(柴秀娟, 山世光, 卿来云等. 基于 3D 人脸重建的光照、姿态不变人脸识别. *软件学报*, 2006, 17(3): 525-534)
- [11] Hu Feng-Song, Lin Ya-Pin et al. Individual 3D face generation based on Candide-3 for face recognition. *Journal of Hunan University (Natural Sciences)*, 2008, 35(11): 69-73 (in Chinese)  
(胡峰松, 林亚平等. 应用于人脸识别的基于 Candide-3 特定人脸三维重建. *湖南大学学报*, 2008, 35(11): 69-73)
- [12] Tan X, Chen S, Zhou Z-H et al. Face recognition from a single image per person: A survey. *Pattern Recognition*, 2006, 39(9): 1725-1745
- [13] Frade F, De la Torre, Gross R et al. Representational oriented component analysis (ROCA) for face recognition with one sample image per training class//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005: 266-273
- [14] Martinez A M. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 25(6): 748-763



- [15] Martinez A M. Recognizing expression variant faces from a single sample image per class//Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), 2003: 353-358
- [16] Yang Q, Ding X. Symmetrical PAC in face recognition//Proceedings of the ICIP02, 2002, 2: 97-1000
- [17] Gutta S, Wechsler H. Face recognition using asymmetric faces//Proceedings of the ICBA2004, 2004: 162-168
- [18] Beymer D, Poggio T. Face recognition from one example view. *Science*, 1996, 272(5250)
- [19] Geghiades A S, Belhumeur P N et al. Illumination-based image synthesis: Creating novel images of human faces under differing pose and lighting//Proceedings of the Workshop on Multi-View Modeling and Analysis of Visual Scenes, 1999: 47-54
- [20] Geghiades A S, Belhumeur P N et al. From few to many: Illuminationcone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23: 643-660
- [21] Chai Xiu-Juan, Shan Shi-Guang, Gao Wen et al. Example-based learning for automatic face alignment. *Journal of Software*, 2005, 16(5): 718-726(in Chinese)  
(柴秀娟, 山世光, 高文等. 基于样例学习的面部特征自动标定算法. *软件学报*, 2005, 16(5): 718-726)
- [22] Le H-S, Li H. Recognizing frontal face images using hidden Markov models with one training image per person//Proceedings of the 17th International Conference on Pattern Recognition (ICPR2004), 2004, 1: 318-321
- [23] Martinez A. Face image retrieval using HMMs//Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries. *IEEE Computer Society*, 1999: 35-39
- [24] Liu Xiao-Jun, Wang Dong-Feng, Zhang Li-Fei et al. An approach for face recognition based on singular value decomposition and hidden Markov model. *Chinese Journal of Computers*, 2003, 26(3): 340-344(in Chinese)  
(刘小军, 王东峰, 张丽飞等. 一种基于奇异值分解和隐马尔可夫模型的人脸识别方法. *计算机学报*, 2003, 26(3): 340-344)
- [25] Sim T, Baker S, Bsat M. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(12): 1615-1618
- [26] Aleix M. Face image retrieval using HMMs//Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries. Washington D. C., USA: *IEEE Computer Society*, 1999: 35-39
- [27] Liu Xiao-Jun, Wang Dong-Feng, Zhang Li-Fei et al. An approach for face recognition based on singular value decomposition and hidden Markov model. *Chinese Journal of Computers*, 2003, 26(3): 340-344(in Chinese)  
(刘小军, 王东峰, 张丽飞等. 一种基于奇异值分解和隐马尔可夫模型的人脸识别方法. *计算机学报*, 2003, 26(3): 340-344)
- [28] Sim T, Baker S, Bsat M. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(12): 1615-1618
- [29] Li Wen-Shu, Zhou Chang-Le, Xu Jia-Tuo. A novel face recognition method with feature combination. *Journal of Zhejiang University Science*, 2005, 6A(5): 454-459
- [30] Song Feng-Xi, Yang Jing-Yu, Liu Shu-Hai et al. Face representation based on the multiple-class maximum scatter difference. *Acta Automatica Sinica*, 2006, 32(3): 378-385(in Chinese)  
(宋枫溪, 杨静宇, 刘树海等. 基于多类最大散底差的人脸表示方法. *自动化学报*, 2006, 32(3): 378-385)



**HU Feng-Song**, born in 1969, Ph.D. candidate, associate professor. His research interests include face recognition, pattern analysis, graphics and image processing.

**ZHANG Mao-Jun**, born in 1971, Ph. D., professor,

Ph.D. supervisor. His research interests include virtual reality, information system engineering.

**ZOU Bei-Ji**, born in 1961, Ph. D., professor, Ph. D. supervisor. His research interests include CAD & CG, graph and image processing.

**MA Jun-Rong**, born in 1985, master candidate. Her research interests include graphics and image processing, image pattern analysis.

## Background

This work is accomplished by the union of face recognition researchers from Hunan University, Central South University and National University of Defense Technology. This work is partly supported by the National Natural Science Foundation of China under project “Trace and Analyses

Method of Body Empty Out Roll Movement based on Video” with grant No. 60673093, National Natural Science Foundation of China under project “Key Technology Research on Major Engineering Disaster oriented Numerical Value Integrated Simulate Plat” with grant No. 90715043 and Natural

Science Foundation of Hunan Province in China under project “Research on Video Analyses Method and Its Application of Body’s Bounce Movement” with grant No. 07JJ3125.

The group has been concentrated in Patter recognition, image processing and image recognition for decades. In the past several years, plenty of research work of the group has been done on the basic theories in face recognition,as well as the practical engineering techniques. There were several papers published or accepted by some well known journals or proceedings,and many of the research fruits have been successfully applied to practical applications and are going to be industrialized.

This paper solved the problem of face recognition under

variant illumination and poses effectively; the face recognition rate is 20% higher than the recognition rate of the four images in the newest paper in the research region. The theory base of this paper is individual 3D face generation method proposed by authors, which is published in Journal of Hunan University. This paper make use of Candide3 model, combining an importation of frontal face image samples to produce individual person’s face 3D model, thus obtaining the variety of poses. Experimental results show that this method can effectively avoid the recognition rate bring down caused by the possible inefficient pose and illumination normalization, and can be better fitting the pose and illumination invariant face recognition.