

# 多姿态人脸识别\*

陈熙霖<sup>1)</sup> 山世光<sup>1)</sup> 高文<sup>1),2)</sup>

<sup>1)</sup>(哈尔滨工业大学计算机系, 哈尔滨 150001)

<sup>2)</sup>(中国科学院计算技术研究所, 北京 100080)

**摘要** 人脸识别在很多场合都有重要的作用,传统的身份验证是采用某种识别号码等方法,但这难以阻止伪造的发生。由于人的视觉特征如面部、姿态等是相对稳定而且各不相同的,因此采用这些特征进行身份的识别是可行的。本文提出了一种处理多姿态人脸识别的多候选类加权识别方法,为了减少姿态变化的影响,提出了相应的预处理方法。图象质量是影响识别结果的重要因素之一,因此本文提出了一种评价图象质量的方法,实验表明,这种评价的结果是和识别结果紧密相联的,本文提出的识别方法在图象较为清晰的情况下,可以达到很好的识别效果。

**关键词** 人脸识别 多姿态 图象质量评价 特征脸

## 0 引言

在很多场合身份验证是非常重要的,如出入控制、信用卡系统等。现在广泛采用的身份验证形式包括标识号码、磁卡、IC卡等。这些方式的优点是技术已经成熟,并可以采用各种加密手段加以保护,但所有这些都难以阻止伪造。从根本上讲,这些标识验证手段所依赖的都是后天赋予人的信息,不是人所固有的内在信息。当然目前采用指纹、虹膜等的识别技术也已经达到了很好的识别效果<sup>[1]</sup>,但是通常人类的交互并不采用这些方式,更多的是依赖长相(面部特征)、声音、举止动作等加以判断,实际上人类的面部特征是相当稳定的,分别几年甚至几十年的老友即使是在路上匆匆一过也会分辨出来。正是因为如此近年来在国际上有很多研究者对基于人脸和话者等特征的身份识别和验证进行了大量的研究<sup>[2]</sup>。从应用角度讲,自动人脸识别除了可以用于身份识别和验证之外还可用于图象库的检索,根据用户需要,从一个大的脸库中查找某一特定的人,在图象信息查询领域具有广泛的应用前景,可以大大提高工作效率。从识别方法看,目前的人脸识别方法主要分为基于几何特征的方法<sup>[3-7]</sup>、基于模板的人脸识别方法<sup>[7-9]</sup>、主成份分析方法<sup>[10]</sup>和连接主义方法<sup>[11]</sup>等几类。

从识别对象上考虑,大多数研究集中在正面人

脸图象的识别上。在考虑姿态变化方面,Beymer等采用引入虚视点的方法,即将二维的图象和三维的人脸模型相结合,对多姿态的人脸图象进行识别,这种通过合成的方法进行识别要求三维模型必须足够精确,同时计算非常复杂,此外在多视点方法上,Beymer引入了15个不同视角的模型,通过模板匹配进行识别<sup>[12]</sup>。Lades等采用弹性匹配技术识别不同姿态的人脸<sup>[13]</sup>,这种方法的基本假设是要求较为准确地定位特征,由于人脸上缺乏足够的纹理信息,因此,常常难以满足这一假设。

## 1 面部图象的预处理

由于人脸位置、姿态、光照变换的任意性,为了保证具有较好的识别效果,因此必须进行适当的预处理。这些预处理过程包括人脸检测、平面旋转矫正、光照矫正等。人脸检测采用文献<sup>[12]</sup>中的方法,不同之处是为了增强对不同姿态的检测能力,提供了对侧面以及俯、仰和旋转的检测能力,这也解决了在图象平面内平移的问题。

对图象平面内人脸旋转,通过定位出的两个瞳孔点的位置,加以校正。图1(a)是经过瞳孔定位后的图象,利用瞳孔可以估计出面部的中心,经过旋转得到矫正后的图象。这里对面部图象的旋转以人脸的中心为原点,而不是图象的中心为原点,这样可以保证当人脸落在图象的边缘部分时不致于因为旋转

\* 本文的工作得到了国家“863”计划项目的支持(No. 863-306-ZT03-01-2)

收稿日期:1999-01-29;收到修改稿日期:1999-06-11

而使人脸部分受损。图 1(b)是旋转后的图象,图 1 (c)是以图象中心为原点旋转的结果。



(a) 定位瞳孔后的图象

(b) 以人脸中心为原点的旋转

(c) 以图象中心为原点的旋转

图 1 人脸图象的旋转

对由于距离引起的图象中人脸尺寸的变化,通过比例归一化加以校正。这里归一化的目的方面是克服由于距离变化引起的图象大小变化,同时可以在不影响识别效果的前提下降低计算量。归一化后的人脸部分大小为  $64 \times 64$ 。为了突出个性特征,采用了如图 2 的屏蔽掩膜。

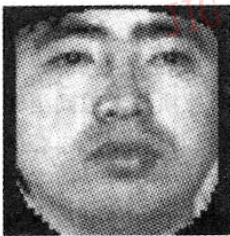
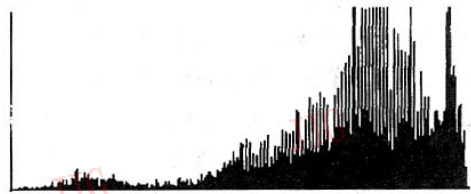


图 2 屏蔽模板

从实验中我们发现虽然采用 Eigenface 方法进行人脸识别可以达到很好的识别效果,但是对光照条件的变化比较敏感。这些变化包括光照、噪声以及偏光等的影响。由于人脸具有相似的三维形状,因此我们认为在相同的光照情况下,人脸图象应具有相似的灰度分布。基于这一假设,我们采用了直方图规格化的方法处理人脸图象,即将不同人脸图象的直方图形状归一。采用图 3(a)的图象作为标准人脸图象(本文中所有图象均为 8bit 的灰度图象),其均值为 190.72,方差为 46.2,直方图分布如图 3(b)。图 4 (a)和(b)是图象直方图归一化前后的对比,均值分别为 141.33 和 191.17,方差分别为 25.35 和 45.01,对应的直方图分别如图 4(c)和(d)。



(a)



(b)

图 3 标准人脸图象及其直方图



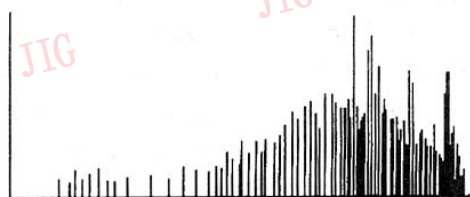
(a) 原图象



(b) 直方图归一后的图象



(c) 图象(a)的直方图



(d) 图象(b)的直方图

图 4 人脸图象的直方图归一

## 2 图象质量的评价

图象的质量是影响识别的重要因素之一,同时对识别结果提供信度支持。但是图象质量的评价在一定程度上是和需要处理的任务相关的。由于人眼的视觉特性,视觉上感觉良好的图象并不一定完全适合计算机自动识别。从计算机处理的角度出发要求图象尽可能清晰,同时具有丰富的内容。对图象质量的衡量可以通过以下几个指标:对比度、噪声、熵等。

设图象为  $F = \{f(i, j) | 0 \leq i < m, 0 \leq j < n, 0 \leq f(i, j) < G\}$ , 其直方图为  $h(k) (0 \leq k < G)$ , 熵定义为:

$$E = - \sum_{i=0}^{G-1} p_i \log p_i \quad (1)$$

其中

$$p_i = \frac{h(i)}{\sum_{j=0}^{G-1} h(j)}$$

在图象中熵的大小反映了图象中灰度的分布,显然一幅清晰的具有视觉意义的图象应当是层次分明的,也就是图象中包含有从低到高各个灰度级的象素,同时这些象素的分布应当尽可能均匀,这样也就说明清晰的图象必然具有较大的熵。这样的一个极端的反例是单一灰度的图象其熵为0,而对于一幅随机生成的图象却可以达到最大,因此熵只是保证了图象在数学意义上的内容丰富。这就如同在运用最大熵原理时必须考虑其他约束条件一样,对图象质量的衡量也必须同时考虑其他条件。

对清晰度的衡量主要反映在图象中噪声的大小。由于噪声是随机的,因此必须对噪声进行估计,噪声的估计基于下面的假设:

- (1) 图象在大多数区域是平滑的;
- (2) 图象的边缘是较为明显的。

对图象中某一点  $f(i, j)$  计算其沿水平和垂直方向的二阶差分  $\partial^2 f(x, y) / \partial x^2$  和  $\partial^2 f(x, y) / \partial y^2$ , 如果  $\partial^2 f(x, y) / \partial x^2 < T_e$  或  $\partial^2 f(x, y) / \partial y^2 > T_e$  ( $T_e$  是边缘阈值), 则表示是边缘, 否则取其平均值作为该点的噪声。最后可以计算得到图象的噪声平均值  $\mu_n$ 。

除了可以将噪声作为清晰度的衡量,熵作为内

容丰富程度的衡量之外,图象中大多数象素的分布范围决定了图象层次是否丰富,尽管这在一定程度上可以通过熵来反映,但是这种关系在熵表达中并不是十分明显,因此定义对比度为:

$$Con = g_1 - g_2 \quad (2)$$

其中:

$$g_2 = \arg \left\{ \sum_{i=0}^{g_2-1} p_i < 0.1, \sum_{i=0}^{g_2} p_i \geq 0.1 \right\},$$

$$g_1 = \arg \left\{ \sum_{i=g_1+1}^{G-1} p_i < 0.1, \sum_{i=g_1}^{G-1} p_i \geq 0.1 \right\}$$

实际上,图象过亮或过暗都会影响识别的结果,因此图象的均值  $\mu$  也是衡量图象质量的指标之一。于是我们定义图象质量为:

$$Quality = \begin{cases} 10 \log \frac{E^2}{\mu_n}, & \text{if } \mu < \mu_{\min} \text{ or } \mu > \mu_{\max} \\ & \text{or } Con < Th_{Con} \\ 20 \log \frac{E^2}{\mu_n}, & \text{otherwise} \end{cases} \quad (3)$$

(3)式中的第一项是考虑到当图象过暗或过亮或者对比不够强烈的话,图象内容的可信度要大大地降低。利用这一衡量方法我们可以得出对识别结果的信度,这在辅助识别或多信息融合识别时尤其重要。在(3)式中涉及三个阈值,这三个阈值的选取是根据以下假设并考虑实验结果得到的,即大多数象素的灰度应当处于集中在中间的1/2灰度范围内,于是取  $\mu_{\max} = 200$ ,  $\mu_{\min} = 64$ , 同时80%以上的象素不能集中在15%以内的范围中,否则或者图象层次不够分明,或者图象内容单一,因此取  $Th_{Con} = 40$ 。利用这一方法对3幅代表性图象评价结果如图5所示。其中图5(a)是在晚间灯光下拍摄的,图5(b)是在白天光线充足的条件下拍摄的,评价的结果分别为14.541952(所有涉及对数的计算均采用自然对数)和25.612938,特别需要说明的是对合成的随机图象,尽管其熵高达5.543,但是其噪声却达到20.00,因此图象质量的最终评价结果仅为8.58。

实际上这里对图象质量的评价的方法可以用于所有需要对识别系统输入图象进行评价的场合,通过引入这种评价机制可以为多识别系统的融合提供单个系统处理结果的信度。



图 5 不同图象的质量评价对比

### 3 多姿态人脸识别

所谓人脸识别是指对给定的包含人脸的输入图象,通过某种方式与已知人脸库中存储的模型进行匹配比较,确定是否是库中某一人物,如果是则给出最佳匹配的库中人物。根据人脸分析的目的不同,至少存在 4 种不同类型的人脸识别任务:

(1) 人脸归类,即将输入人脸按某些特征如种族、性别、年龄段等归类;

(2) 人脸认知,即判断输入人脸是否在已知人脸库中;

(3) 身份验证,即验证输入图象是否是所声明的人;

(4) 完备的人脸识别,即判断是否在已知人脸库中,如果满足前一条件还要判断是已知人脸库中的哪一个人。

由于人脸的本身是一种非刚体,而目前的图象识别技术主要是针对刚体的,同时由于光照、姿态、表情、发型、饰物的多样性以及不同人之间的相似性使得人脸识别成为一个极具挑战性的任务。

#### 3.1 基于 Eigenface 的人脸识别

Eigenface 是从主成份分析(PCA)方法导出的一种人脸识别和描述技术。这种方法将包含人脸的图象区域看作是一种随机向量,因此可以采用 KL 变换获得其正交 KL 基底。对应其中较大特征值的基底具有与人脸相似的形状,因此又称之为特征脸(Eigenface)。利用这些基底的线性组合可以描述、表达和逼近人脸图象,因此可以进行人脸的识别与合成。

利用相对较小的 Eigenface 集描述人脸,这样每幅人脸图象就对应一个维数降低的权向量,理论

分析和实验都表明相似的人脸图象其权向量在由特征脸构成的空间中是簇聚的,因此人脸识别可以在降维后的空间上进行。

将  $N \times N$  图象  $\Gamma$  视为  $N^2$  维的随机列向量,设训练集为  $\{\Gamma_1, \Gamma_2, \dots, \Gamma_M\}$ , 计算其协方差阵对应的特征值和特征向量,由此得到特征脸。由于协方差矩阵的维数为  $N^2 \times N^2$ , 因此直接计算有一定的困难。随机向量集  $\{\Gamma_1, \Gamma_2, \dots, \Gamma_M\}$  的协方差矩阵为:

$$C_{N^2 \times N^2} = \frac{1}{M} \sum_{i=1}^M \Phi_i \Phi_i^T = \frac{1}{M} A A^T \quad (4)$$

其中:

$$A = \{\Phi_1 \Phi_2 \dots \Phi_M\}$$

$$\Phi_i = \Gamma_i - \Psi \quad i = 1, 2, \dots, M$$

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i$$

注意到  $\text{rank}(A A^T) = \text{rank}(A^T A)$ , 因此  $A A^T$  与  $A^T A$  有相同的非 0 特征值, 设  $L = A^T A$ , 因为矩阵  $L$  是一个实对称矩阵, 故最多有  $M$  个非 0 的特征值。显然与原始训练集  $\{\Gamma_1, \Gamma_2, \dots, \Gamma_M\}$  的协方差矩阵相比,  $L$  的维数  $M^2$  通常要小得多(一般情况下  $M$  在 100—1000 之间, 而  $N^2$  在 1024—4096 之间), 因此可以简化计算。由此最多可以得到前维特征脸  $\{E_1, E_2, \dots, E_M\}$ 。

在获得特征脸之后, 可以对每个类的典型样本进行投影, 由此得到每个人脸的投影特征从而构成人脸“特征”库, 作为下一步匹配识别的搜索空间。在识别时, 设预处理后的人脸图象为  $\Gamma$ , 且图象的维数与上述训练过程中的图象相同, 则首先对输入图象相对训练集重心平移, 然后进行规范化, 最后计算在特征脸空间上的投影。即:

(1) 计算差图:  $\Phi = \Gamma - \Psi$ ;

(2) 规范化:  $\Phi = \Phi / \|\Phi\|$ ;

(3) 在人脸子空间上的投影:  $W = U_M^T \Phi$ ; 其中

$U_M^T = [E_1 E_2 \dots E_M]$ ,  $M' \leq M$  为选用的特征脸数目。

(4) 将  $W$  在搜索空间  $\Omega$  中进行匹配, 匹配的准则是最短距离加人脸阈值。设  $\omega_i$  为  $\Omega$  中的第  $i$  个标准样本在人脸子空间上的投影(前  $M'$  维), 则  $W$  与  $\omega_i$  的欧氏距离为  $\Delta_i = \|W - \omega_i\|$ 。

对于简单的识别系统而言, 可以用最小距离准则进行判别。实际上, 由于各种条件如光照、姿态、距离等的变化, 使得最小距离准则未必总能取得好的效果, 特别是在多姿态情况下。因此我们将距离变换为概率, 并利用下面的关系处理多姿态问题。

### 3.2 多姿态与多候选类识别

由于人脸的姿态的多样性, 因此必须考虑多姿态问题。这里的多姿态是相对于正面标准像而言的,

同时要求是有意义的视角, 如不能是顶视或后视图象, 而且必须保证两眼是可见的。人脸姿态的变化可以有 6 个自由度, 即沿  $X$ 、 $Y$ 、 $Z$  轴的平移和绕  $X$ 、 $Y$ 、 $Z$  轴的旋转, 如图 6 所示。对沿  $X$ 、 $Y$  的平移, 在图象上表现为人脸位置的变化, 可以通过适当的检测方法处理, 对沿  $Z$  轴的平移, 在图象上表现为比例的变化, 这可以通过前述的比例归一化处理。前面讨论的利用眼球旋转就是为了处理绕  $Z$  轴的旋转。因此上述 6 个自由度的变化只有两个难以直接从图象上确定, 即绕  $X$  和  $Y$  轴的旋转。对于这两个自由度的变化可以通过增加子类解决, 考虑到识别的对象——人脸必须是有意义的视角, 取 9 个不同的位置, 如图 7 所示。

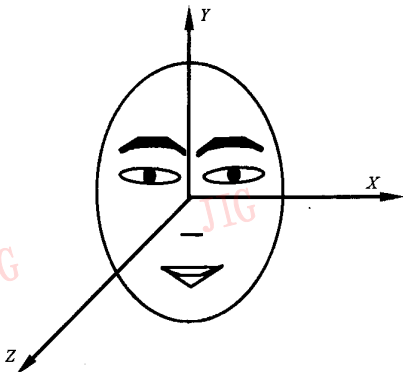


图 6 人脸变化的自由度



图 7 多姿态人脸库构成

虽然构成了多姿态的人脸库, 由于同一人的不同姿态具有较大的相关性, 因此在识别时这些子类间也是相互支持的, 利用这种相关性可以大大地提高识别率。设识别算法  $A$  可以识别的不同类集合为  $\Omega^{(A)} = \{\omega_{1,1}^A, \omega_{1,2}^A, \dots, \omega_{1,n_1}^A, \omega_{2,1}^A, \omega_{2,2}^A, \dots, \omega_{2,n_2}^A, \dots, \omega_{m,1}^A, \omega_{m,2}^A, \dots, \omega_{m,n_m}^A\}$

对其中第  $j$  类分为  $n_j$  个子类。

设输入图象导致的识别结果中前  $k$  个候选子类为  $\{\phi_1^A, \phi_2^A, \dots, \phi_k^A\}$ , 对应的概率为  $q_1^A \geq q_2^A \geq \dots \geq q_k^A$ 。对识别问题而言, 可以忽略子类间的差异, 因此实际包含的类别可能小于  $k$  个。将其中同类的各子类合

并, 若  $\phi_1^A, \phi_2^A, \dots, \phi_j^A$  属于同一类, 对应概率为  $q_1^A \geq q_2^A \geq \dots \geq q_j^A$ , 则合并后的概率对应为  $p_i^A = q_1^A + \eta_1 q_2^A + \dots + \eta_{j-1} q_j^A$ 。最终得到的实际候选类为  $\{\omega_1^A, \omega_2^A, \dots, \omega_k^A\}$ , 对应的概率为  $p_1^A \geq p_2^A \geq \dots \geq p_k^A$ 。于是得到最终的识别结果, 这里的关键是合并概率时的参数  $\eta_i$  的确定, 在我们的系统中是通过对训练集图象的识别结果训练得到的。在识别过程中关键的做法是多个子类的合并, 合并概率是基于这样的假设, 即人脸在不同姿态间是连续变化的, 同时同一人的不同视角的图象是具有相似性的。

## 4 实验结果分析

在测试的人脸库中共有 60 人, 每人 20 幅图象, 我们取其中的 400 幅图象进行了识别效果的测试, 识别集与训练集没有重叠。在不采用多子类候选加权识别的情况下, 识别正确的图象共有 356 幅。采用第 3 节中的方法, 取  $\eta_1=0.2, \eta_2=0.2\eta_1, \dots$ , 正确识

别的图象为 384 幅, 错误识别的图象为 16 幅。从实验中可以验证上述方法是有效的。同时实验表明, 对于错误识别的 16 幅图象, 其质量评价结果多在 12 以下, 质量评价结果在 20 以上的图象只有一幅未能正确识别。这一结果表明在采用多候选类加权识别的情况下, 只要图象层次分明、清晰, 这一方法可以达到很高的识别率。(如图 8 所示)。



图 8 正确识别的部分图象

## 5 结 论

人脸识别是自动身份验证的重要手段之一, 从本文的结果不难看出对于相对较小的识别集(100 人左右), 采用人脸识别可以达到较好的效果, 实际上利用与其他信息的融合可以处理更大的识别集。相对姿态变化而言, 图象质量对识别结果的影响更大, 因此在需要进行自动识别的场合, 适当的光源控制是必须的。本文提出的方法对于多姿态人脸识别是较为有效的一种方法。

### 参 考 文 献

- 1 Wildes R P. Iris recognition: An emerging biometric technology. In: Proceedings of the IEEE, 1997, 85(9): 1348~1363.
- 2 Brunelli R, Falavigana D. Person identification using multiple cues. IEEE Transactions on PAMI, 1995, 17(10): 955~966.

- 3 Kanade T. Picture processing by computer complex and recognition of human faces. Tech Report, Kyoto Univ Dept of Information Science, 1973.
- 4 Craw I, Cameron P. Parameterizing images for recognition and reconstruction. In: Proc of British Machine Vision Conference, 1991, 367~370.
- 5 Chen C W, Huang L H. Human face recognition from a single front view. Int J of Pattern Recognition, 1992, 6(4): 571~593.
- 6 Wong K H, Law H H M, Tsang P W M. A system for recognizing human faces. In: Proc of Int Conf ASSP, 1989, 1638~1642.
- 7 Brunelli R, Poggio T. Face recognition: Features vs templates. IEEE Trans PAMI, 1993, 15(10): 1042~1052.
- 8 Beymer D J. Pose-invariant face recognition using real and virtual views. MIT AI-Lab Tech Report No. 1574, 1996.
- 9 Nastar C, Pentland A. Matching and recognition using deformable intensity surfaces. MIT MediaLab Tech Report, No. 334, 1995.

- 10 Turk M, Pentland A. Eigen-faces for recognition. *J of Cognitive Neuroscience*, 1991, 3(1): 71~86.
- 11 Valentin D, Abdi H, O'Toole A J, Cottrell G W. Connectionist models of face processing: A survey. *Pattern Recognition*, 1994, 27: 1209~1230.
- 12 Beymer D. Pose-invariant face recognition using real and virtual views[dissertation]. AI Lab, MIT, 1996.
- 13 Lades M, Vorbruggen *et al.* Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans Computer*, 1993, 42(3): 300~311.
- 14 Liu Mingbao, Gao Wen. A hierarchical approach to human face

detection in a complex background. In: *The First International Conference on Multimodal Interface*, Beijing, 1996.



山世光 博士研究生。主要研究领域为图象理解,多模式人机接口。



陈熙霖 博士,教授。主要研究领域为图象理解,多模人机式接口,数据压缩系统以及虚拟现实等。已出版专著1本,发表论文30余篇。



高文 教授,博士生导师。现任中国科学院计算技术研究所所长,国家“863”计划智能计算机主题专家组组长(首席专家),联想中央研究院院长。主要研究领域为人工智能应用和多媒体技术,侧重于计算机视觉,模式识别与图象处理,多媒体数据压缩,多模式接口及虚拟现实。

## Pose-Variied Face Recognition

Chen Xilin<sup>1)</sup>, Shan Shiguang<sup>1)</sup> and Gao Wen<sup>1),2)</sup>

<sup>1)</sup>(Harbin Institute of Technology, Harbin 150001)

<sup>2)</sup>(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080)

**Abstract** Automatic face recognition/identification is very useful in some areas. Usually, an ID number is applied for that purpose. This can't prevent the forge ID to be used. The visual feature (such as face, gesture etc.) is relatively stable, and different between each other. So these information can be used to identify the human being. In this paper, a multi-candidate based approach for pose varied face recognition is given. To reduce the affection of pose variety, some pre-process methods are also applied. Actually, the quality of the image is a very important factor for automatic recognition, so a method for image quality estimation is proposed. The experiments show that the image quality is associated with the recognition result. When the input image is relatively clarity, it can reach a good recognition result.

**Keywords** Face recognition, Pose varied, Image quality estimation, Eigenface