

基于二维 Gabor 小波的人脸识别算法

曹林^{①②} 王东峰^① 刘小军^① 邹谋炎^{①②}

^①(中国科学院电子学研究所 北京 100080)

^②(中国科学院研究生院 北京 100039)

^③(北京信息工程学院信息与通信工程系 北京 100101)

摘要 该文提出了一种基于二维 Gabor 小波的人脸识别算法。该算法先对人脸图像进行多分辨率的 Gabor 小波变换,然后在图像上放置一组网格结点,每个结点用该结点处的多尺度 Gabor 幅度特征描述,采用主元分析法对每个结点进行去相关、降维,最后形成特征结。把每个特征结作为观测向量,对隐马尔可夫模型进行训练,并把优化的模型参数用于人脸识别。实验结果表明,该方法识别率高,复杂度较低。

关键词 人脸识别, Gabor 小波变换, 主元分析, 隐马尔可夫模型

中图分类号: TP391.41 文献标识码: A 文章编号: 1009-5896(2006)03-0490-05

Face Recognition Based on Two-Dimensional Gabor Wavelets

Cao Lin^{①②} Wang Dong-feng^① Liu Xiao-jun^① Zou Mou-yan^{①②}

^①(Institute of Electronics, Chinese Academy of Sciences, Beijing 100080, China)

^②(Graduate School of the Chinese Academy of Sciences, Beijing 100039, China)

^③(Dept. of Information and Telecommunication, Beijing Information Technology Inst., Beijing 100101, China)

Abstract A new approach based on two-dimensional Gabor wavelets transform for face recognition is presented. The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels. A set of vectors called nodes, over a dense grid of image points are formed, and each node is labeled with a set of complex Gabor wavelets coefficients. The magnitudes of the coefficients are used for recognition. Principal component analysis is a decorrelation technique and its primary goal is to project the high dimensional vectors into a lower dimensional space. Feature nodes, as observation vectors of HMM, is derived by using principal component analysis. A set of images representing different instances of the same person is used to train each HMM, and each individual in the database is represented by an optimal HMM face model. Experimental results show that the proposed algorithm has a high recognition rate with relatively low complexity.

Key words Face recognition, Gabor wavelets transform, Principal component analysis, Hidden Markov Model(HMM)

1 引言

近十年来,人脸识别的研究有了很大进展。与指纹、语音等其他人体生物特征相比,人脸识别更加直接、友好,在身份识别,视频检索,安全监控等方面有着广泛的应用,是当前模式识别和人工智能领域的一个研究热点^[1]。特征提取在人脸识别中的作用至关重要,如何根据人的视觉机制提取有效的特征一直是模式识别领域的研究热点。早期的研究中有人用Gabor小波对大脑皮层的视觉感知细胞的性态进行建模^[2],即可以把每个视觉细胞看作一个具有一定方向和尺度的Gabor滤波器。当外界刺激(例如图像信号)输入到视觉细胞

时,视觉细胞的输出响应就是图像与Gabor滤波器的卷积,而这个输出信号经大脑的进一步处理后形成最后的认知映像。由于这种模型能较好地解释人的视觉对图像尺度的伸缩和方向变化的容忍度,因此它被广泛的应用到人脸识别中^[3]。

众所周知,一种人脸识别方法的效果如何,取决于它在多大程度上利用了图像的原始信息,于是可以利用一组数值特征来描述人脸的各个器官,并且利用这组数值特征对人脸进行识别。但是简单地利用一组数值特征不能满意地解决人脸识别问题。另一方面,视觉识别人脸的机制是十分微妙的,人们对此的认识还非常肤浅。事实上,人脸应当作为一个整

体来描述, 不仅仅包括各个器官的数值特征, 还应当包括各个器官的不同表象和相互关联。用于语音识别的隐马尔可夫模型^[4](Hidden Markov Model, HMM)则为我们提供了解决这一问题的良好机制, 按照这种模型, 观测到的特征被看成是另一组不可观测“状态”产生的一系列实现。因而可以将不同的人用不同的HMM参数来表征, 而同一个人由于姿态和表情变化产生的多个观测序列可以通过同一个HMM模型来表征, 这在理论上有了很大的进步。

自从Nefian^[5]采用在矩形抽样窗口进行 2D-DCT变换来提取HMM所需要的观测向量以来, 大部分学者在利用HMM进行人脸识别时, 都采用在矩形窗口进行某种变换来提取特征, 常见的变换有 2D-DCT, KLT, 小波变换, 奇异值分解等。当 1D-HMM识别率不高时, 就采用更复杂的嵌入式隐马尔可夫模型(Embedded Hidden Markov Model, EHMM)^[5] 或 2D-HMM^[6], 研究的热点主要是如何简化这些复杂的模型, 而忽略了对有效特征抽取的研究。本文提出了基于Gabor小波和HMM的人脸识别算法, 在人脸图像上放置一组网格结点, 每个结点的特征用该结点处的多尺度Gabor幅度特征描述, 从而在每个结点处形成一个特征矢量, 并把把这个特征矢量作为HMM的观测向量, 然后用 1D-HMM来描述这些结点之间的关联。实验结果表明, 与EHMM相比, 本文所述的方法与其识别率相当, 但由于仅采用了 1D-HMM, 所以计算复杂度低。

2 特征提取

二维Gabor小波的核函数定义为^[3]

$$\psi_{v,\mu}(z) = \frac{\|k_{v,\mu}\|^2}{\sigma^2} e^{-\frac{\|k_{v,\mu}\|^2 \|z\|^2}{2\sigma^2}} \left(e^{ik_{v,\mu}z} - e^{-\sigma^2/2} \right) \quad (1)$$

$$k_{v,\mu} = (k_v \cos \varphi_\mu, k_v \sin \varphi_\mu) \quad (2)$$

其中 φ_μ, k_v 分别定义了波向量 $k_{v,\mu}$ 的方向和尺度, $z = (x, y)$, $\|\square\|$ 定义了向量范数。在式(2)中 $k_v = k_{\max} / f^v$, $\varphi_\mu = \pi\mu/8$ 。 f 为频率域中的采样步长, 通常取 $f = \sqrt{2}$ 。 k_{\max} 对应最大的采样频率, 取 $k_{\max} = \pi/2$ 。参数 σ 决定了高斯窗的宽度与波向量长度的比率, 在本文中取 $\sigma = 7\pi/4$ 。

用 $I(z) = I(x, y)$ 表示图像的灰度分布, 则图像 I 与 Gabor 小波 $\psi_{v,\mu}$ 的卷积为

$$O_{v,\mu}(z) = I(z) \otimes \psi_{v,\mu}(z) \quad (3)$$

符号 \otimes 表示二维卷积。一个图像经过 40 个 Gabor 小波滤波

后形成的集合为

$$S = \{O_{v,\mu}(z) : v \in \{0, \dots, 4\}, \mu \in \{0, \dots, 7\}\} \quad (4)$$

应用卷积定理, 能够通过快速傅里叶变换(FFT)计算 $O_{v,\mu}(z)$:

$$\mathfrak{F}\{O_{v,\mu}(z)\} = \mathfrak{F}\{I(z)\} \mathfrak{F}\{\psi_{v,\mu}(z)\} \quad (5)$$

$$O_{v,\mu}(z) = \mathfrak{F}^{-1}\{\mathfrak{F}\{I(z)\} \mathfrak{F}\{\psi_{v,\mu}(z)\}\} \quad (6)$$

其中符号 \mathfrak{F} 和 \mathfrak{F}^{-1} 分别定义了 FFT 和 IFFT。

由式(3)可知, 对于图像上任意一点 $z_i = (x_i, y_i)$, 经过 40 个 Gabor 滤波器后形成了一个特征矢量, 我们把这个特征矢量定义为“结”, 用 $J(z_i)$ 表示, 即

$$J(z_i) = \{J_{v,\mu}(z_i) = |O_{v,\mu}(z)|_{z=z_i}, v \in \{0, \dots, 4\}, \mu \in \{0, \dots, 7\}\} \quad (7)$$

本文在进行特征提取后, 将使用 HMM 表示这些结之间的关联, 并使用图 1 所示方法对图像进行均匀采样。假设人脸图像的宽度为 W , 高度为 H , 本文用宽度为 L_x , 高度为 L_y 的滑动采样窗($L_x \times L_y$)对图像从上到下, 从左到右进行采样, 采样窗水平方向的重叠为 P_x , 垂直方向的重叠为 P_y 。采样时只把采样窗口的中心点作为人脸图像的“结”, 忽略窗口内的其他点, 也就是说, 一个采样窗对应一个“结”。采样个数由下式给出:

$$T_0 = \left\lfloor \frac{H - L_y}{L_y - P_y} \right\rfloor + 1 \quad (8)$$

$$T_1 = \left\lfloor \frac{W - L_x}{L_x - P_x} \right\rfloor + 1 \quad (9)$$

其中 T_0 表示垂直方向采样点的个数, T_1 表示水平方向采样点的个数, 符号 $\lfloor \cdot \rfloor$ 表示下取整。本文中采样窗口的大小为 7×7 , 9×9 , 11×11 。重叠为 $P_x = 0$, $P_y = 0$ 。

由式(7)可知, $J(z_i) \in \square^D$, 其中 $D = 40$ 。文献[7]指出人脑总是在低维空间中完成诸如相似性判别的感知任务, 因此低维对于机器学习也同样非常重要。此外维数的降低也减少了下一步使用 HMM 对人脸图像进行训练的计算量。另一

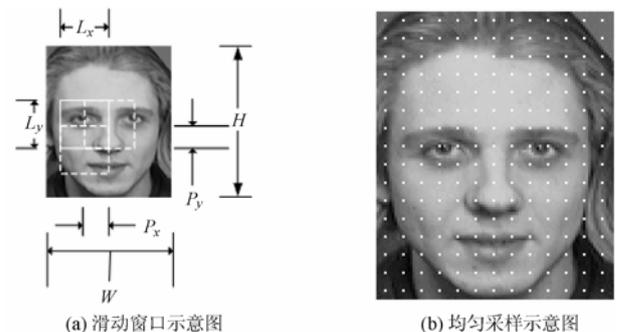


图1 均匀采样方案

方面, 每个结之间存在相关性, 这将影响系统的识别性能, 图2说明了降维和去相关的必要性。主元分析法^[8](Principal Component Analysis, PCA)是一种良好的去相关方法, 是在最小均方差准则下把高维数据投影到低维空间的最优方法。利用它形成降维矩阵 \mathbf{Y} , 其中 $\mathbf{Y} \in \mathbb{R}^{D \times d}$, 则降维表达式为

$$\mathbf{J}_{\mathbf{Y}}(\mathbf{z}_i) = \mathbf{Y}^T \mathbf{J}(\mathbf{z}_i) \quad (10)$$

低维向量 $\mathbf{J}_{\mathbf{Y}}(\mathbf{z}_i) \in \mathbb{R}^d$ 表达了原始向量 $\mathbf{J}(\mathbf{z}_i)$ 的本质属性, 我们把经过变换后的结 $\mathbf{J}_{\mathbf{Y}}(\mathbf{z}_i)$ 定义为“特征结”, 并把它作为下一步HMM的观测向量。

3 隐马尔可夫模型

3.1 基本概念

隐马尔可夫模型(HMM)是一种用参数表示的, 用于描述随机过程统计特性的概率模型^[4]。它由两部分组成: 一个是隐含的马尔可夫链称为隐含层; 另一个是实际的观测量称为观测层。HMM的基本参数组成为: N_i 表示隐含状态数, $\mathbf{S} = \{S_j\}$ 表示隐含状态, $\mathbf{O} = \{o_i\}$ 表示观测序列, E 表示观测序列的长度, M 表示高斯概率混合成分的个数, $\mathbf{A} = \{a_{ij}\}$ 表示状态转移矩阵, $\mathbf{B} = \{b_j(o_i)\}$ 表示观测层的概率分布矩阵, $\boldsymbol{\Pi} = \{\pi_i\}$ 为初始状态概率分布。下文中, 把HMM简记为: $\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\Pi})$ 。

3.2 HMM 训练

Samaria^[9]在最初提出基于HMM的人脸识别方法时考虑如下: 人脸图像包含头发、额头、眼睛、鼻子和嘴巴5个显著的特征区域, 即使头部有一些偏转或倾斜, 它们的次序从上到下保持不变, 可以认为这5个显著区域隐含5个状态。事实上, 在求HMM的初始参数时, 我们可以进行类似的均匀分割, 但是最终求得的状态是不受这些区域限制的, 而且也不一定非要把人脸分成5个状态。因而, 可以更进一步理解, 我们观测到的序列是由若干个状态产生的, 这些状态是抽象的, 不具有具体的意义, 只能通过观测序列对它进行估计。

HMM 训练就是要为每个类别确定一组经过优化的HMM 参数, 每个模型可以用单幅或多幅图像进行训练, 训练步骤如下:

- (1) 对人脸进行 Gabor 小波变换, 求出特征结, 并将其作为观测向量, 即 $o_i = \mathbf{J}_{\mathbf{Y}}(\mathbf{z}_i)$ 。
- (2) 建立一个通用的模型 $\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\Pi})$, 确定模型的状态数, 允许的状态转移和高斯混合概率成分的个数。
- (3) 将训练数据均匀分割, 与 N_i 个状态对应, 计算模型

的初始参数。对于状态转移矩阵 $\mathbf{A} = \{a_{ij}\}$, 可以使 $a_{ij} = 0$, 当 $j < i$ 或 $j > i + 1$ 。对于初始概率分布 $\boldsymbol{\Pi} = \{\pi_i\}$, 可以使 $\pi_1 = 0$, $\pi_i = 0 (i \neq 1)$, 即HMM从第一个状态开始^[9]。

概率分布矩阵 $\mathbf{B} = \{b_j(o_i)\}$ 可依据下面的公式计算:

$$b_j(o_i) = \sum_{k=1}^M c_{jk} (2\pi)^{-d/2} |\boldsymbol{\Sigma}_{jk}|^{-1/2} \cdot \exp(-(\mathbf{o}_i - \hat{\boldsymbol{\mu}}_{jk})^T \boldsymbol{\Sigma}_{jk}^{-1} (\mathbf{o}_i - \hat{\boldsymbol{\mu}}_{jk}) / 2) \quad (11)$$

其中 c_{jk} 为混合成分的比例因子, 即用高斯混合模型来对概率分布矩阵 \mathbf{B} 建模。 $\hat{\boldsymbol{\mu}}_{jk}$ 和 $\boldsymbol{\Sigma}_{jk}$ 分别为高斯混合模型的均值和协方差矩阵,

$$\hat{\boldsymbol{\mu}}_{jk} = \left(\sum_{i=1}^{E_{j,k}} \mathbf{o}_i^{(j,k)} \right) / E_{j,k} \quad (12)$$

$$\boldsymbol{\Sigma}_{jk} = \left(\sum_{i=1}^{E_{j,k}} (\mathbf{o}_i^{(j,k)} - \hat{\boldsymbol{\mu}}_{jk})(\mathbf{o}_i^{(j,k)} - \hat{\boldsymbol{\mu}}_{jk})^T \right) / E_{j,k} \quad (13)$$

其中 $E_{j,k}$ 是均匀分割后, 状态 j 的第 k 个混合成分对应的序列长度, $\mathbf{o}_i^{(j,k)}$ 为相应的观测向量。

(4) 用Viterbi分割取代均匀分割^[4], 并利用分段K均值聚类方法^[4]求出高斯混合模型的参数, 迭代调整初始模型参数。

(5) 采用Baum-Welch算法^[4]对参数进行重新估计, 并用最后得到的模型优化参数表示人脸数据库中的某个类别。

3.3 HMM 人脸识别

在人脸识别阶段, 首先要对待识别的人脸图像 k 进行 Gabor 变换, 计算它的特征结, 形成观测序列 $\mathbf{O}^{(k)}$, 然后使用前项-后项算法^[4]计算每个训练模型 λ_i 产生该序列的概率 $P(\mathbf{O}^{(k)} | \lambda_i)$, 最大值所对应的模型就是待识别人脸所属的类别, 可以用公式表达为

$$\lambda_n = \arg \max_i P(\mathbf{O}^{(k)} | \lambda_i) \quad (14)$$

即: 如果第 n 个模型 λ_n 产生序列 $\mathbf{O}^{(k)}$ 的概率最大, 则将图像 k 归入第 n 类。

4 实验结果及分析

本文使用 ORL 人脸数据库, 该数据库包含 40 个人, 每人 10 幅图像, 共 400 幅图像, 每个图像大小为 $W = 92$, $L = 112$ 。实验中每人取 5 幅图像, 共 200 幅图像进行训练, 用另外 200 幅进行识别。

图2显示了不进行 PCA 变换, 直接用 $\mathbf{J}(\mathbf{z}_i)$ 作为 HMM 的观测向量时, 观测向量维数与识别率的关系。从图中可见, 最高识别率低于 80%, 这说明利用 PCA 进行去相关和降维

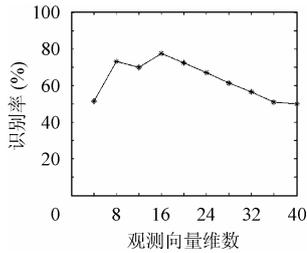


图2 不进行PCA变换,识别率与观测向量维数的关系

是非常必要的。另外当维数 $D \in \{8 \dots 16\}$ 时, Gabor 小波的尺度小, 高斯窗采样频率大, 识别率较高, 在此范围外, 识别率递减。这说明相对较小尺度的 Gabor 变换包含较多的人脸特征, 观测向量之间的相关性较弱, 对识别率的改善作用较大。但是这并不等于大尺度的 Gabor 变换没有用处, 实际上高频信息描述了图像的局部细节, 低频信息描述了图像的全局特征, 而 PCA 的作用是把两种信息进行有效的融合, 去除相关性, 提高识别率。

图3显示了经过PCA变换后, 识别率与不同窗口观测向量维数的关系。图中的图例表示HMM状态数和高斯概率混合成分的个数。例如 5×3 表示本条曲线是在状态数 $N_i = 5$, 混合成分个数 $M = 3$ 的条件下绘制的。从图中可见, 在3种不同的采样窗口下, 当观测向量维数 $d \in \{6 \dots 10\}$ 时, 识别率较高, 而且在 $d = 6$ 附近识别率最高。当 $d > 10$ 时, 随着维数的增高, 识别率降低; 当 $d < 6$ 时, 随着维数的增高, 识别率递增。这是因为维数太少, 包含的人脸特征也少, 所以识别率降低; 而维数过多, 一方面由于PCA只利用了图像的二阶统计信息, 不可能完全去相关, 所以随着维数的增加, 观测向量之间的相关性增大, 识别率降低; 另一方面, 图像的本质特征总是包含在低维空间中^[7], 所以观测向量的维数越高, 识别率越低。此外采样滑动窗口越小, 观测序列

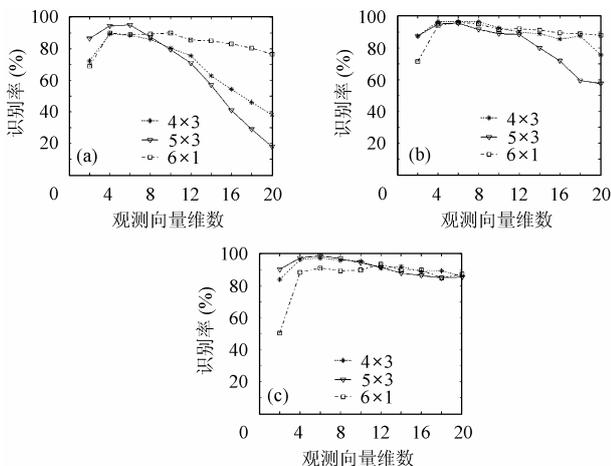


图3 识别率与不同采样窗口观测向量维数的关系

(a) 11×11 窗口 (b) 9×9 窗口 (c) 7×7 窗口

的长度越大, HMM 将拥有更多的训练数据, 所以整体识别率增高。例如 7×7 窗口下的整体识别率最高, 但是窗口越小, 观测序列的长度越大, 训练 HMM 的时间越长, 因此窗口的选择也要适当。

表1显示了观测向量维数 $d = 6$ 时, 状态个数 N_i 和高斯概率混合成分个数 M 对识别率的影响。从表中可见, 随着 M 增大, 识别性能得到改善, 但当 M 增大到一定量时, 此时若再增大 M , 则识别率降低。由于用高斯混合模型描述概率分布矩阵 B , 而矩阵 B 描述了每个状态的概率统计属性, 适当地增大 M 可以增加高斯混合模型的自由度, 因此可以更好地表现每个状态的统计属性, 进而识别率增高。但是如果 M 增大过量, 就使高斯混合模型过分依赖于训练数据, 对当前训练数据的描述过于精确, 以至于测试数据与训练数据稍有不同时, 就无法识别出测试数据, 这实际上是过学习的结果。

表1 $d = 6$ 状态个数和高斯概率混合成分个数对识别率的影响

(a) 9×9 窗口						
状态个数	高斯概率混合成分个数					
	1	3	5	7	9	11
3	93.0	96.0	96.0	90.0	91.0	84.5
4	93.0	96.5	93.5	89.5	89.5	75.0
5	92.5	96.0	92.5	89.0	80.0	65.0
6	96.0	95.0	89.0	83.5	70.5	58.0

(b) 7×7 窗口						
状态个数	高斯概率混合成分个数					
	1	3	5	7	9	11
3	83.5	95.0	96.5	95.0	94.5	94.0
4	85.5	97.5	95.5	95.0	93.0	93.0
5	90.0	99.0	97.5	96.5	93.5	91.5
6	91.0	97.5	96.5	93.0	90.5	89.5

表2列出了相关方法的实验结果对比。从表中可以看出 Samaria^[9]直接取图像亮度作为观测向量, 观测向量的维数高, 采用一维HMM识别只有 84%。即使采用伪二维模型 PHMM^[9], 识别也只有 94.5%, 而且状态个数多, 观测序列的长度大。Nefian^[5]采用 2D-DCT系数作为观测向量, 降低了观测向量的维数, 并取得了与Samaria相当的识别效果, 识别率为 86%。尽管Nefian采用的EHMM, 具有很高的识别率, 但是它实际上是伪 2D-HMM模型, 观测向量的个数多, 状态总数多, 所以对模型进行训练所需的时间较长。本文采用的特征抽取方案观测向量所需的维数小, 观测序列的长度适中, 识别率较高, 因而易于工程应用。

表 2 相关方法实验结果对比

方法	观测向量维数	状态个数	观测序列长度	识别率 (%)
1D-HMM-Lum ^[9]	920	5	103	84.0
2D-PHMM-Lum ^[9]	80	24	2236	94.5
1D-HMM-DCT ^[5]	39	5	52	86.0
2D-EHMM-DCT ^[5]	6	24	2236	99.0
本文方法	6	5	208	99.0

5 结束语

本文提出了一种基于 Gabor 小波变换和 1D-HMM 的人脸识别方法, 并同 Samaria 和 Nefian 的方法进行了比较。实验结果表明, 本文方法识别率高, 复杂度较低。

本文算法在特征提取上所需的计算量较大, 在今后的研究中, 可采用快速算法降低复杂度, 使本算法更有利于工程应用。总之, 基于 Gabor 小波变换和 HMM 的人脸识别方法具有较高的识别率, 工程上易于实现, 是一种具有发展潜力的人脸识别方法。

参 考 文 献

- [1] Gong S, Psarrou A. Dynamic Vision: from Images to Face Recognition[M]. London: Imperial College Press, 2000: 5 – 20.
- [2] Daugman J. Two-dimensional spectral analysis of cortical receptive field profiles[J]. *Vision Research*, 1980, 20(10): 847 – 856.
- [3] Lades M, Vorbruggen J C, Buhmann J. Distortion invariant object recognition in the dynamic link architecture[J]. *IEEE*

Trans. on computers, 1993, 42(3): 300 – 311.

- [4] Rabiner L. A tutorial on hidden Markov models and selected application in speech recognition[J]. *Proce. IEEE*, 1989,77(2): 257 – 286.
- [5] Nefian A. A hidden Markov model-based approach for face detection and recognition[D/D]. Georgia: Georgia Institute of Technology, 1999: 38 – 108.
- [6] Othman H, Aboulnasr T. A separable low complexity 2D HMM with application to face recognition[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2003, 25(10): 1229 – 1238.
- [7] Helmuth L. Objection recognition: where the brain tells a face from a place[J]. *Science*, 2001, 292(5515): 196 – 198.
- [8] Duda R, Hart P, Stork D. Pattern Classification, second edition[M]. New York: Wiley-Interscience, 2000: 114 – 139.
- [9] Samaria F. Face recognition using hidden Markov model[D/D]. Cambridge: University of Cambridge, 1994: 27 – 82.

曹 林: 男, 1977 年生, 博士生, 研究方向为图像处理和模式识别。

王东峰: 男, 1974 年生, 博士, 助理研究员, 研究方向为信号和图像处理。

刘小军: 男, 1972 年生, 博士, 副研究员, 研究方向为信号和图像处理。

邹谋炎: 男, 1941 年生, 博士, 研究员, 博士生导师, 研究方向为信号和图像处理以及宽带无线通信。