

文章编号: 1671-8860(2006)07-0573-03

文献标志码: A

# 人脸识别中嵌入式隐 Markov 模型 结构的优化算法研究

王 晖<sup>1</sup> 卢 健<sup>1</sup> 孙小芳<sup>1</sup>

(1 武汉大学遥感信息工程学院, 武汉市珞喻路 129 号, 430079)

**摘要:** 提出了一种优化嵌入式隐 Markov 模型状态数的搜索算法, 并且由此算法得到改进的嵌入式隐 Markov 模型结构。实验证明, 这种结构既能提高人脸的识别率, 最高可达到 100%, 又比传统的结构减少 30% 左右的训练时间和识别时间。

**关键词:** 嵌入式隐 Markov 模型; 结构; 人脸识别

中图法分类号: T P751; P237.4

隐 Markov 模型是人脸识别的有效模型, 它先是应用在语音识别上<sup>[1]</sup>, 后来由 Samaria 首先应用到人脸图像的模式识别上<sup>[2]</sup>, 提出用从上到下的 5 个状态的隐 Markov 模型来表示人脸; Neffian 发展为嵌入式的隐 Markov 模型<sup>[3]</sup>, 他提出的模型由 5 个超状态构成, 每个超状态内部分别嵌入(3, 6, 6, 6, 3)个子状态, 但是他并没有给出得到这个结构的方法, 只是凭经验确定。这种结构后来被越来越多的研究人员所使用<sup>[4, 5]</sup>。本文给出了一个优化模型结构的算法, 并且提供了一个更优的结构, 使识别率更高, 训练和识别的时间大为缩短, 而且占用更少的内存。

## 1 嵌入式隐 Markov 模型

嵌入式隐 Markov 模型是隐 Markov 模型的扩展, 它是在一维隐 Markov 模型的每个隐状态里面又嵌入一个独立的一维隐 Markov 模型, 这样, 嵌入式的隐 Markov 模型就由竖向的超状态和一系列横向的嵌入状态构成。因此, 一个嵌入式隐 Markov 模型可记为:

$$\lambda = (\pi, A, \Lambda)$$

其中,  $\pi$  为超状态初始概率分布,  $\pi = \{\pi_i, 1 \leq i \leq N\}$ ,  $N$  为垂直方向的超状态数;  $A$  为超状态转移概率矩阵,  $A = \{a_{ij}, 1 \leq i, j \leq N\}$ ;  $\Lambda$  为超状态

内嵌入的水平方向的子 Markov 模型序列, 即嵌入的 Markov 模型,  $\Lambda = \{\Lambda^i, 1 \leq i \leq N\}$ ; 第  $i$  个超状态可表示为  $\Lambda^i = \{\pi^i, A^i, B^i\}$ ,  $\pi^i$  为第  $i$  个超状态中嵌入状态的初始概率分布,  $\pi^i = \{\pi_k^i, 1 \leq k \leq N^i\}$ ,  $N^i$  为第  $i$  个超状态中嵌入的状态数, 子状态  $S^i = \{S_k^i, 1 \leq k \leq N^i\}$ ;  $A^i$  为第  $i$  个超状态中嵌入状态的转移概率矩阵,  $A^i = \{a_{kl}^i, 1 \leq k, l \leq N^i\}$ ;  $B^i$  为第  $i$  个超状态中嵌入的子状态的观察值概率密度函数,  $B^i = \{b_k^i(O_{t_0, t_1})\}$ ,  $O_{t_0, t_1}$  表示  $t_0$  行、 $t_1$  列的观测向量 ( $t_0 = 1, \dots, T_0$ ;  $t_1 = 1, \dots, T_1$ )。 $B^i$  通常用多维混合高斯概率密度函数来近似表示:

$$b_k^i(O_{t_0, t_1}) = \sum_{f=1}^F C_{kf}^i N(O_{t_0, t_1}, \mu_{kf}^i, U_{kf}^i),$$
$$1 \leq k \leq N^i$$

式中,  $F$  为组成  $b_k^i(O_{t_0, t_1})$  的混合概率密度函数的个数;  $C_{kf}^i$  是第  $i$  个超状态中嵌入的第  $k$  个状态的第  $f$  个混合量的权系数;  $N(O_{t_0, t_1}, \mu_{kf}^i, U_{kf}^i)$  为多维高斯概率密度函数, 其中,  $\mu_{kf}^i$  表示第  $i$  个超状态中嵌入的第  $k$  个状态的第  $f$  个组合量的均值向量;  $U_{kf}^i$  表示第  $i$  个超状态中嵌入的第  $k$  个状态的第  $f$  个混合量的协方差矩阵。

## 2 嵌入式隐 Markov 模型的训练

嵌入式隐 Markov 模型的三个基本算法如下。

收稿日期: 2006-04-22。

项目来源: 国家自然科学基金资助项目(40171032)。

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

1) 给定模型参数  $\lambda = (\pi, A, \Lambda)$  和观测向量序列  $O$ , 如何计算概率  $p(O|\lambda)$ , 即观测向量序列与模型吻合的概率。这是一个评估问题, 决定出在多个模型中选择与观测向量序列最佳匹配的模型。本文采用向前向后算法<sup>[4]</sup>。

2) 给定模型参数  $\lambda = (\pi, A, \Lambda)$  和观测向量序列  $O$ , 如何选择最佳状态序列来解释观测向量序列, 即揭开模型的隐含部分, 在一优化准则下找到最优状态序列。本文采用双重嵌套的 Viterbi 算法<sup>[4]</sup>。

3) 给定观测向量序列  $O$ , 如何调整模型参数, 使  $p(O|\lambda)$  最大, 即模型参数估计问题。本文采用 Baum-Welch 算法<sup>[4]</sup>。

训练过程如图 1 所示, 首先把观测向量均匀分割到  $N$  个从上到下的超状态中, 然后在每个超状态中的观测向量又从左到右分到  $N^i$  个嵌入状态中。在下一次迭代过程中, 这些均匀分割的观测向量通过双重嵌套的 Viterbi 算法进行重新分割, 然后通过双重  $K$  均值分割算法进行模型参数的重新估计。当前后两次 Viterbi 分割的似然度之差小于给定的阈值后, 算法结束。

上文提到的观测向量的提取是根据文献[5]中的研究结果, 在原始图像上开一个采样小窗口, 使窗口从左到右、从上到下进行采样, 相邻的图像块有一定的覆盖, 本文取图像块的二维离散余弦变换的部分系数作为观测向量。

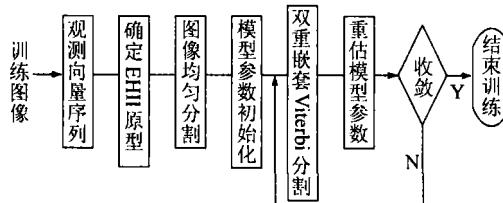


图 1 嵌入式隐 Markov 模型的训练过程

Fig. 1 Training Process of Embedded Hidden Markov Model

### 3 结构优化算法

如图 2(a) 所示, 它从上到下分为 5 个超状态, 每个超状态内的嵌入状态数分别是  $(3, 6, 6, 6, 3)$ 。文献[3]中并没有给出得到状态结构  $(3, 6, 6, 6, 3)$  的算法, 而是凭经验选取。本文提出了一种状态数搜索算法。为了优化结构, 必须提出所谓优化的准则。本文认为, 优化的标准有两个:

嵌入状态数尽量少; 识别率尽量高。算法如下:

初始化搜索起点, 即状态数  $(1, 1, 1, 1, 1)$ 。

© 1994-2015 China Academic Journal Electronic Publishing House. All Rights Reserved. <http://www.cnki.net>

为当前状态位置; 计算当前状态的识别率; 计算当前状态的相邻状态的识别率, 如状态  $(1, 1, 1, 1, 2)$ 、 $(1, 1, 1, 2, 1)$ 、 $(1, 1, 2, 1, 1)$  等; 将相邻状态中最大的识别率与当前状态的识别率比较, 如果当前状态的识别率大, 则搜索结束; 否则, 置识别率最大的状态为当前状态, 转步骤 。

由本搜索算法计算的结果可知, 状态  $(3, 2, 1, 4, 6)$  有最高的识别率, 为 100%, 并且嵌入的状态数由 24 个下降到 16 个, 如图 2(b) 所示, 在纵向还是分为 5 个超状态, 但是嵌入的状态数调整为  $(3, 2, 1, 4, 6)$ 。

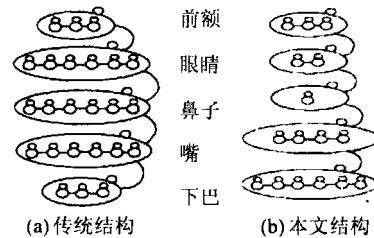


图 2 两种模型结构

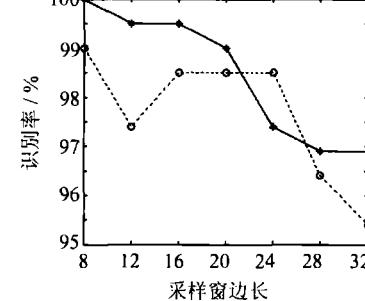
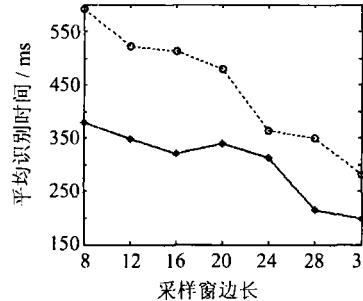
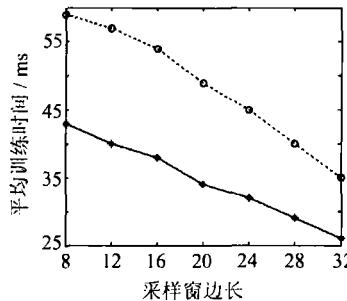
Fig. 2 Structures of Two Models

### 4 实验

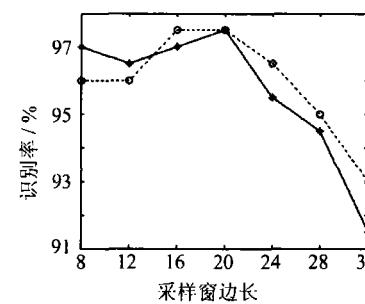
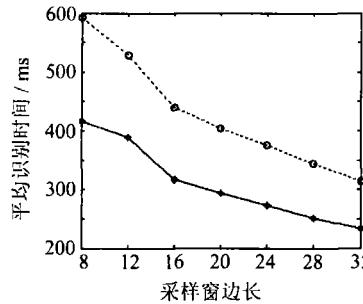
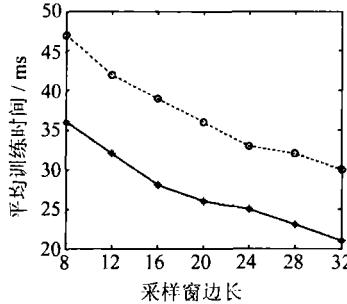
为了能和文献[3]的结果进行比较, 用相同的训练和测试数据集, 即 ORL(Olivetti research laboratory) 人脸数据库<sup>[6]</sup>。ORL 人脸数据库包含 40 个人, 每人 10 幅图像, 共 400 幅人脸图像。每幅人脸图像的大小为  $92 \times 112$ 。和文献[3]一样, 本文采用每个人的 5 幅图像进行训练, 另外 5 幅图像进行识别。识别方法是: 计算待识别图像的观测向量序列与已经训练好的模型之间的似然度, 待识别图像被识别为具有最高似然度的那个模型所代表的人脸。

实验平台为 IBM R50 53C 笔记本电脑, 迅驰 1.4 GHz CPU, 256 M 内存, 30 G 硬盘, 操作系统为 Windows XP home 版。实验结果如表 1 所示。由表 1 可知, 本文提出的模型结构具有更高的识别率。由于文献[3]并没有给出算法的时间复杂度, 本文无法比较, 而且用的平台不同, 也没有可比性。但是在同样的平台上, 本文对这两个模型结构进行了全面的比较。图 3 反映了两种结构分别在 5 幅图像训练、5 幅图像识别和 4 幅图像训练、5 幅图像识别两种情况下的实验结果, 图中实线代表  $(3, 2, 1, 4, 6)$  结构, 虚线代表  $(3, 6, 6, 6, 3)$  结构, 每组实验的采样窗大小分别为  $8 \times 8$  到  $32 \times 32$  的 7 种情况。实验结果在训练时间、识别时间和识别率方面进行了比较。从图 3(a) 可以

看出, 结构(3, 2, 1, 4, 6)在识别率上普遍比(3, 6, 6, 6, 3)高, 其中最高的识别率达到100%; 在图3(b)中, 两者的识别率相当。但不管哪组实验, 在识别时间和训练时间上, (3, 2, 1, 4, 6)都比(3, 6, 6, 6, 3)大为缩短, 在训练时间上, 平均节省28.4%; 在识别时间上, 平均节省30.9%。



(a) 5幅图像训练、5幅图像识别的结果



(b) 4幅图像训练、4幅图像识别的结果

图3 两种模型结构的比较

Fig. 3 Comparison of Two Kinds of Model Structures

从模型的结构上看, (3, 2, 1, 4, 6)总共有16个嵌入状态, 比(3, 6, 6, 6, 3)的24个状态数少了1/3, 因而它占用的内存数更少, 嵌入状态的运算量也少。然而状态数的减少并没有使识别率降低, 反而提高了, 这说明嵌入式隐Markov模型的识别能力与被识别对象本身的性质有关, 而与状态的多少没有必然联系, 只有最符合对象特性的模型结构才能产生最好的识别效果。从识别率曲线上看到, 随着采样窗口的增大, 识别率有降低的趋势。其原因是, 随着采样窗口的变大, 所产生的观测序列变短, 决定了嵌入的HMM的训练次数变少。从图3(b)的比较中可以看出, 当用4幅图像训练时, 本文的结构和传统的结构有相当的识别率; 但当用5幅图像训练时, 本文结构的识别率大多超过了传统结构的识别率, 这说明本文的模型结构对训练图像的数量反应敏感、收敛快。

## 参考文献

- [1] Bahl L, Jelinek F, Mercer R. A Maximum Likelihood Approach to Continuous Speech Recognition. *IEEE Trans Pattern Anal Machine Intell*, 1983 (5): 179-190

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

表1 两种模型结构识别率的比较

Tab. 1 Comparison of Recognition Rate of

### Two Structures

模型结构	识别率/%
文献[3] (3, 6, 6, 6, 3)	98
本文(3, 2, 1, 4, 6)	100

第一作者简介: 王晖, 博士生。现从事图像处理与理解、人脸识别研究。

E-mail: zhengrong01@263.net

(下转第581页)

## Spatial-Temporal Evolvement Trends of Mesoscale Convective Systems over Qinghai-Tibetan Plateau

LIN Hui<sup>1</sup> JIANG Jixi<sup>2</sup> YANG Yubin<sup>1,3</sup> FANG Zhaobao<sup>4</sup>

(1 Institute of Space and Earth Information Science, Chinese University of Hong Kong, Shatin, N. T., Hong Kong)

(2 National Satellite Meteorological Center, China Meteorological Administration,  
46 South Street, Zhongguancun, Beijing 100081, China)

(3 State Key Lab. for Novel Software Technology, Nanjing University, 22 Hankou Road, Nanjing 210093, China)

(4 The P. L. A. 91561 Command, Guangzhou 510320, China)

**Abstract:** An automatic meteorological data mining approach is presented based on two kinds of heterogeneous satellite datasets, temperature black body (TBB) remote sensing imagery and high resolution limited area analysis and forecasting system (HLAFS) data, aiming to analyze the meteorological evolvement trends and patterns of MCS and discover the hidden correlations and causalities between the environmental physical fields and the eastward evolvement trends of MCS. Research results are provided in the form of decision rules and environmental physical model graphs to express the hidden knowledge effectively and understandably.

**Key words:** spatial data mining; mesoscale convective systems; Qinghai-Tibetan plateau

**About the first author:** LIN Hui, Ph. D, professor, Academician of International Eurasian Academy of Sciences. His major research areas include cloudy and rainy area remote sensing, virtual geographic environments, spatially integrated humanities and social science.

E-mail: hulin@cuhk.edu.hk

(上接第 575 页)

## An Improved Algorithm Based on Embedded Hidden Markov Model Structure for Face Recognition

WANG Hui<sup>1</sup> LU Jian<sup>1</sup> SUN Xiaofang<sup>1</sup>

(1 School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

**Abstract:** An improved embedded hidden Markov model structure for face recognition is presented. The model not only improves the recognition rate, but also reduces the train and recognition time costs dramatically. Experimental results show that the highest recognition rate of 100% is achieved. The most exciting thing is that the structure can shrink 30% time costs than traditional structure both on train and recognition stage.

**Key words:** embedded hidden Markov model; structure; face recognition

**About the first author:** WANG Hui, Ph. D candidate, majors in image processing and interpretation, face recognition.

E-mail: zhargrong01@263.net