

用隐马尔可夫模型设计人脸表情识别系统*

尹星云^{**1,2} 王 洵¹ 董兰芳¹ 万寿红¹

(1. 中国科学技术大学计算机科学技术系 合肥 230027; 2. 安徽淮南师范学院信息技术系 安徽 淮南 232001)

【摘要】根据隐马尔可夫模型(HMM)的基本理论和算法设计了一个人脸表情识别系统。该系统由两层HMM组成: 低层由六个HMM组成, 分别对应六种特定表情。人脸表情特征向量进入系统后, 经过低层HMM初步识别, 其结果组成高层HMM的观察向量, 经过高层HMM解码, 确认出表情, 从而提高了系统的识别率, 增强了系统的健壮性。

关键词 隐马尔可夫模型; forward-backward算法; Viterbi算法; Baum-Wellch算法; 人脸表情识别; 人脸表情特征向量

中图分类号 TP391.41 文献标识码 A

Design of Recognition for Facial Expression by Hidden Markov Model

Yin Xingyun^{1,2} Wang Xun¹ Dong Lanfang¹ Wan Shouhong¹

(1. Department of Computer Science & Technology, University of & Technology of China Hefei 230027;
2. Department of Information & Technology, HuaiNan Normal College AnHui HuaiNan 232001)

Abstract Hidden Markov Model(HMM) is a widely used statistical model. This paper deal with the design of a recognition system for facial expression in the light of the principal theory and algorithm of the HMM. The system consists of two levels of HMM, with the low level composed of six HMM, corresponding to six expression-specific. The vectors of the facial expression features, after put into the system, is primarily identified through the low HMM, and result in the observation vectors of the high level of the HMM. Through deciphering of the HMM, the expression are identified, thus enhancing the recognition rate and strengthening the system to a higher level.

Key words hidden Markov model; forward-backward algorithm; Viterbi algorithm; Baum-Wellch algorithm; recognition of facial expression; vectors of features of facial expression

隐马尔可夫模型(Hidden Markov Model, HMM)是一个统计模型, 具有十分丰富健壮的数学结构, 适用于动态过程时间序列建模并具有强大的时序模式分类能力, 理论上可处理任意长度的时序, 它为HMM提供了一个非常广泛的应用范围。目前, HMM已经成功的用于语音识别, 而人脸识别和人脸表情识别是HMM的又一个新的应用领域。人脸表情识别是设计人机智能交互、视频会议和可视电话等系统的关键技术, 有着广泛的应用前景。

1 HMM定义

一个HMM模型由两个随机过程组成, 一个是马尔可夫链, 由初始状态分布概率 $p = \{p_i\}$ 和转移概率矩

2003年1月14日收稿

* 安徽省自然科学基金资助项目, 编号: 2003KJ293

** 女 45 在职硕士生 副教授 主要从事计算机图形图像方面的研究

阵 $A=\{a_{ij}\}$ 描述,输出为状态序列,但是这个输出序列隐藏在模型中;另一个随机过程由 $B=\{b_j(O_t)\}$ 描述,输出为观察序列但不一定是马尔可夫过程,这也是HMM名称的由来。

设有观察序列 $O=O_1O_2\dots O_T$ 和状态集 $S=\{S_1,S_2,\dots,S_N\}$,定义HMM为三元组 $I=(A,B,p)$,其中 $A=\{a_{ij}\}$ 为状态转移概率矩阵, $a_{ij}=P\{S_j \text{ at } t+1|S_i \text{ at } t\}$, $1 \leq i,j \leq N$; $B=\{b_j(O_t)\}$ 为观察符号概率分布,如果 B 是离散的并有 M 个观察值 $\{v_1,v_2,\dots,v_M\}$,则 $b_j(O_t)=P\{O_t=v_k|S_j \text{ at } t\}$, $1 \leq j \leq N, 1 \leq k \leq M$,如果 B 是连续的并用Gaussian混合概率分布,则: $b_j(O_t)=\sum_{k=1}^M c_{jk} N(O_t | \mathbf{m}_{jk}, \mathbf{\Sigma}_{jk})$,其中 c_{jk} 为第 j 次混合系数; $\mathbf{m}_{jk}, \mathbf{\Sigma}_{jk}$ 分别为观察向量的均值和协方差矩阵; p 为初始状态分布概率,记为 $p=\{p_i\}$,其中 $p_i=P\{S_i \text{ at } t=1\}$, $1 \leq i \leq N$ 。

2 应用HMM识别表情的基本原理

人脸表情识别系统中的状态 $S_i(i=1,2,\dots,6)$ 是表情,观察序列 $O=O_1O_2\dots O_T$ 由表情特征向量组成,其概率分布 B 用Gaussian混合概率分布。

用Baum—Welch算法训练HMM:取一个初始HMM $I=(A,B,p)$,用

$$\bar{p}_i = \mathbf{g}_i(i) \quad 1 \leq i \leq N$$

$$\bar{a}_{ij} = [\sum_{t=1}^{T-1} \mathbf{x}_t(i,j)] / \sum_{t=1}^{T-1} \mathbf{g}_t(i) \text{ 和 } \bar{b}_j(k) = [\sum_{t=1, O_t=v_k}^T \mathbf{g}_t(j)] / \sum_{t=1}^T \mathbf{g}_t(j), \quad 1 \leq i,j \leq N \text{ 迭代, 而}$$

$$\mathbf{x}_t(i,j) = [\mathbf{a}_t(i)a_{ij}b_j(O_{t+1})\mathbf{b}_{t+1}(j)] / \sum_{i=1}^N \sum_{j=1}^N \mathbf{a}_t(i)a_{ij}b_j(O_{t+1})\mathbf{b}_{t+1}(j), \quad \mathbf{g}_t(i) = \sum_{j=1}^N \mathbf{x}_t(i,j)$$

式中 $1 \leq t \leq T$, \mathbf{a} 和 \mathbf{b} 为forward-backward变量。直到 $R(O|I)$ 不再变化时 I 即为所求。

用 $R(O|I)$ 评估HMM模型的好坏, $R(O|I)$ 由Forward-backward算法计算:

$$\mathbf{a}_1(i) = p_i b_i(O_1) \quad 1 \leq i \leq N$$

递推

$$\mathbf{a}_{t+1}(j) = [\sum_{i=1}^N \mathbf{a}_t(i)a_{ij}] b_j(O_{t+1}) \quad 1 \leq j \leq N, 1 \leq t \leq T-1$$

终止

$$P(O|I) = \sum_{i=1}^N \mathbf{a}_T(i)$$

观察序列进入系统后由Viterbi算法解码:

$$\mathbf{d}_1(i) = p_i b_i(O_1), \quad \mathbf{y}_1(i) = 0 \quad 1 \leq i \leq N$$

递推

$$\mathbf{d}_t(j) = \max_{1 \leq i \leq N} [\mathbf{d}_{t-1}(i)a_{ij}] b_j(O_t) \quad 2 \leq t \leq T, 1 \leq j \leq N$$

$$\mathbf{y}_t(j) = \operatorname{argmax}_{1 \leq i \leq N} [\mathbf{d}_{t-1}(i)a_{ij}] \quad 2 \leq t \leq T, 1 \leq j \leq N$$

终止

$$P^* = \max_{1 \leq i \leq N} [\mathbf{d}_T(i)]$$

$$(q_T^* = \operatorname{argmax}_{1 \leq i \leq N} [\mathbf{d}_T(i)])$$

路径回溯 $q_t^* = \mathbf{y}_{t+1}(q_{t+1}^*), t=T-1, T-2, \dots, 1$, $q_1^*, t=T, T-1, \dots, 1$, 就是识别出的表情状态序列。

随着 t 的增大,实现算法时计算精度将超出计算机的处理能力而产生溢出。解决它的办法是在计算过程中引入一个缩放因子,计算结束时消去,不影响计算结果。另外不能直接求 $P(O|\bar{I})$,而用对数式求 $\log[P(O|I)]$,用Viterbi算法解码时如果用对数式就不用乘缩放因子^[1-4]。

3 用HMM设计的人脸表情识别系统

3.1 系统结构

人脸表情可以归纳为六类: {高兴Happy (1), 生气Angry (2), 吃惊Surprise (3), 厌恶Disgust (4), 恐惧Fear (5), 悲哀Sad (6)}。为每种表情设计一个HMM,每个HMM是一个带返回的从左到右模型,六个这样的HMM构成一个人脸表情分类器,如图1所示。对每个HMM求 $P(O|I_c)$, $1 \leq c \leq 6$,如果取 $c^* = \operatorname{argmax}_{1 \leq c \leq 6} [P(O|I_c)]$,则 c^* 就是系统识别出的人脸表情编号。但是由于人脸表情有上千种,简单地归纳为六类,使得有些表情很难准确地归为某一类,这就增加了系统识别的难度。如,人脸上呈现的表情接近Happy和Surprise时,通过低层HMM获得的 $P(O|I_1)$ 和 $P(O|I_3)$ 实际上非常接近。由于 $P(O|I)$ 的概率特性

和计算机的存储误差, $P(O | I_1)$ 很可能等于 $P(O | I_3)$,从而不能正确地识别这两种表情。为了提高系统识别复杂表情的能力,用两层HMM构造人脸表情识别系统,如图2所示。高层HMM图中的编号含义如下:(1)Happy, (2) Angry, (3)Surprise, (4)Disgust, (5)Fear, (6)Sad, (7)中性Neutral。

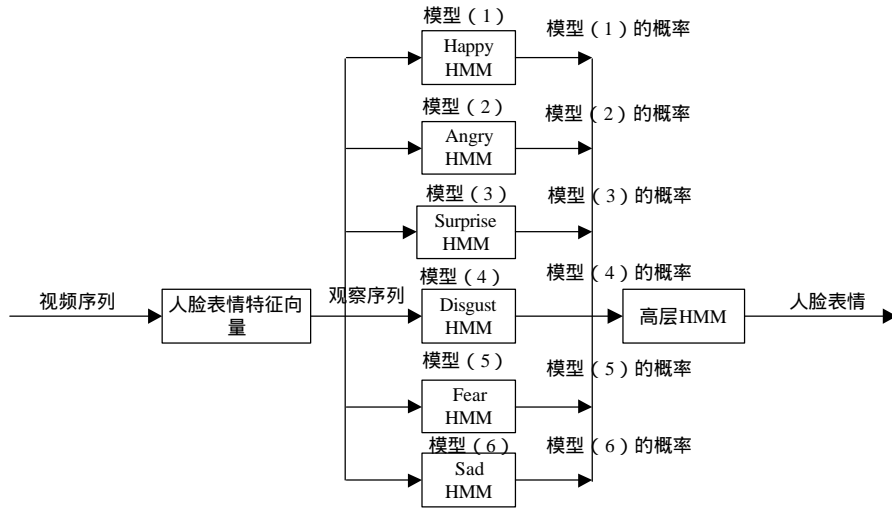


图1 低层HMMs结构示意图

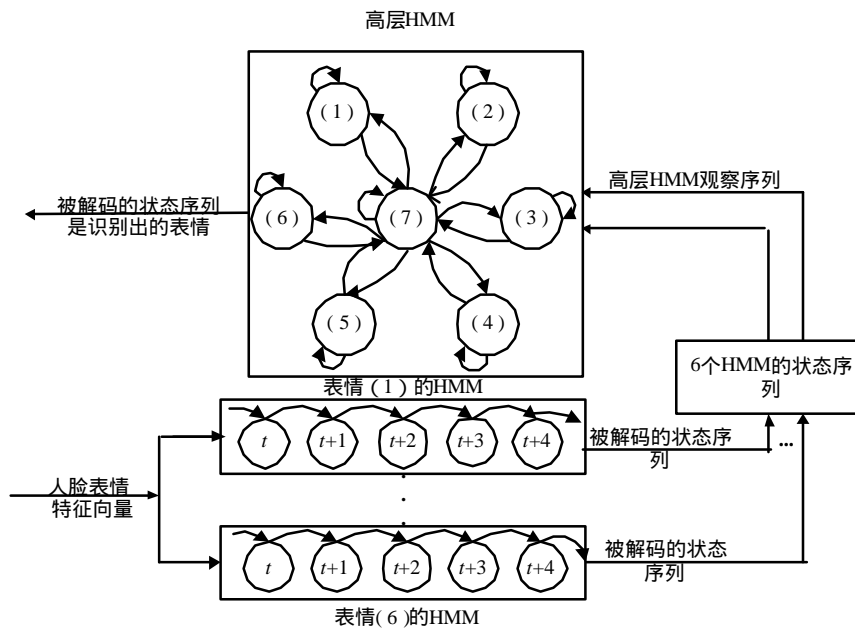


图2 系统结构示意图

低层HMM的参数非常直观,其参数如下:初始表情状态分布概率 $\mathbf{p} = \{p_i\}$, $p_7=1$, $p_i=0, 1 \leq i \leq 6$ 。表情状态转移概率矩阵 $A = \{a_{ij}\}$, a_{ij} 表示从第 i 种表情状态转移到第 j 种表情状态的概率,并规定当 i, j 均不为7且 i 不等于 j 时, a_{ij} 表示从第 i 种表情状态先转移到中性表情状态再转移到第 j 种表情状态的概率,即假设两种非中性状态的表情之间不能直接转移,这种假设是合理的,如:从“Sad”不可能不经过中性状态直接转移到“Happy”。

将低层HMMs解码的六个表情状态序列组成高层HMM的一个六维观察向量 $O_t^h = [q_t^{(1)}, q_t^{(2)}, \dots, q_t^{(6)}]$,其中 $q_t^{(i)}$ 为第 i 个HMM的状态,观察符号概率分布 $B^{(h)} = \{b_k(O_t^h)\} = \{\prod_{i=1}^6 [P(q_j^{(i)} | S_k)]\}$,其中, $P(q_j^{(i)} | S_k) = \{\text{低层第 } i \text{ 个HMM的第 } j \text{ 个状态在高层HMM的第 } k \text{ 个状态的期望概率}\}, 1 \leq j \leq 6$ 。

3.2 系统实现

首先对人脸图象进行预处理,把体现人脸表情的部分从人脸图象中分离出来,然后进行离散余弦变换^[5]。设 $X_{ex}(k_1, k_2)$ 、 $X_{nu}(k_1, k_2)$, 分别为同一个人的某种表情状态和中性表情状态的离散余弦系数, $k_1, k_2 = 1, 2, \dots, N$, N 为图象的大小。 $X_{ex} - X_{nu}$ 的值(称为表情系数)组成一个 $N \times N$ 的矩阵,将该方阵按行排列组成一维的人脸表情特征向量 $Y_i = [y_{i0}, y_{i1}, \dots, y_{iN}]$, i 表示为第 i 张人脸图象,将 Y_i 作为低层HMM的观察向量^[5]。

系统训练过程:

1) 从手工分割的视频序列中提取出六种表情的特征向量组成观察序列,用 Baum—Welch 算法分别训练六个HMM,得到六个特定表情的HMM。

2) 低层HMM训练好后,将人脸表情序列同时输入进这六个特定表情的HMM,用 Viterbi 算法得到六个被解码的人脸表情状态序列。

3) 将低层的六个人脸表情状态序列组成高层HMM的一个六维观察向量 $O_i^h = [q_i^{(1)}, q_i^{(2)}, \dots, q_i^{(6)}]$, $q_i^{(i)}$ 为第 i 个HMM的状态。

4) 对观察序列 O_i^h 再次用 Baum—Welch 算法训练高层HMM。

系统的使用和训练过程类似,任意一个人的人脸视频序列进入系统后,由系统自动提取出人脸表情特征向量。这些特征向量进入低层HMM后,由 Viterbi 算法解码得到表情状态序列,表情状态序列被组织成高层HMM 的观察序列再次由 Viterbi 算法解码,获得视频序列中人脸对应的表情。

4 结 束 语

目前,对手工分割的视频序列用该系统实现了人脸表情识别,其中Happy的识别率最好,几乎达到100%,Sad的识别率最差,只有89.16%,平均识别率达到97.2%。结果表明HMM是一种非常有效的基于统计的识别表情的方法。其缺点是离散余弦变换虽然具有压缩比高且能较好地保持原始图象信息的优点,但是在人脸表情识别系统中作为人脸表情特征向量,数据量仍然太大,影响系统速度。

下一步的研究工作重点放在系统的前端,为系统选择一个更合适的人脸表情特征的描述方式和人脸特征跟踪的方法,实现从连续的视频序列自动地识别人脸的功能。

参 考 文 献

- [1] Lawrence R R. A tutorial on hidden markov models and selected application in speech recognition[J]. Proceeding of the IEEE, 1989, 77(2): 257-286
- [2] Ira C. Automatic facial expression recognition from video sequences using temporal information: [Master Thesis][D]. USA: University of Illinois at Urbana-Champaign, Dept. of Electrical Engineering, 2000
- [3] Alexander V L, Mark B. GeneMark.hmm: new solutions for gene finding[J]. Nucleic Acids Research, 1998, 26(4): 1107-1115
- [4] Otsuka T, Ohya J. Recognizing multiple persons' facial expression using HMM based on automatic extraction of significant frames from image sequences[C]. In: Proc. Int. Conf. on Image Processing (ICIP-97). USA, 1997. 546-549
- [5] Xiao Y, Chandrasiri N P, Tadokoro Y M Oda. Recognition of facial expressions using 2D DCT and neural network[J]. Electronics and Communications in Japan, 1999, 82(7): 1-11

编辑 漆 蓉