

人脸表情的实时分类

王宇博, 艾海舟, 武勃, 黄畅

(清华大学计算机科学与技术系, 智能技术与系统国家重点实验室, 北京 100084)

E-mail: ahz@mail.tsinghua.edu.cn

摘要: 本文探讨了人脸表情的分类问题。由于人脸表情的多样性和复杂性, 这项工作相当困难。本文提出了一种基于查找表(Look Up Table, LUT)型弱分类器的 Adaboost 方法, 学习人脸表情分类器, 实验表明该方法与支持向量机方法相比, 对于人脸表情分类的正确率相当, 而速度快近 300 倍, 具有实时性, 具有非常明显的应用价值。

关键词: 人脸表情分类; Adaboost; 查找表

中图法分类号: TP391

Real Time Facial Expression Classification

WANG Yubo, AI Haizhou, WU Bo, HUANG Chang

Computer Science and Technology Department, Tsinghua University

State Key Laboratory of Intelligent Technology and Systems

E-mail: ahz@mail.tsinghua.edu.cn

Abstract: In this paper, the problem of facial expression classification is discussed, which is very

本课题得到国家自然科学基金重点项目(批准号: 60332010)的资助。王宇博, 男, 1979 年生, 硕士研究生, 主要研究方向为计算机视觉。艾海舟, 男, 1964 年生, 教授, 主要研究方向为计算机视觉。武勃, 男, 1979 年生, 硕士研究生, 主要研究方向为计算机视觉。黄畅, 男, 1981 年生, 硕士研究生, 主要研究方向为计算机视觉。

difficult because of its diversity and complexity. An Adaboost algorithm based on Look-Up-Table (LUT) weak classifier is presented to train facial expression classifier. The experimental results show that compared to Support Vector Machines (SVMs), our method has almost the same correct rate; and nearly 300 times faster in speed. Our method is almost real time, and has significant value in application.

Key Words: Facial expression classification; Adaboost; Look-Up-Table

1 引言

近年来,随着人机交互技术日益成为研究的热点,情感计算成为一个热门的研究课题,人脸表情的自动分类开始受到广泛的关注。人脸的表情通常分为七类[1]:愤怒(anger),厌恶(disgust),恐惧(fear),喜悦(happiness),中性(neutral),悲哀(sadness)和惊讶(surprise),如图 1 所示。研究人脸表情的自动分类方法对于人机交互技术具有重要意义,在计算机辅助的训练及远程教育等众多领域具有潜在的应用价值。

人脸是人类最重要的生物特征之一,包含着极其丰富的信息。与此同时,在计算机视觉、模式识别和多媒体技术等研究领域,人脸作为图像与视频中的视觉对象之一,占有非常重要的地位。人脸表情是人们之间交流情感和表达意图的最快捷和有力的手段之一,例如可以表达出感兴趣、厌倦和迷惑等认知状态。

人脸表情的自动分类是一项困难的工作。一是由于人脸表情的描述过于语言化,没有明确的形式化的定义;二是到目前为止,人脸表情的心理学研究进展还比较缓慢[2](这些都不是本文所考虑的范畴)。与此同时,对于人脸表情的自动分类系统来说,还有三个需要关注的问题。首先,该系统应该能够处理所有的人脸图像,不论其性别,年龄和种族;第二,

该系统的分类结果应该与人脸的姿态无关，即具有鲁棒性；最后，该系统应该能实时的对人脸表情进行分类。



图 1 人脸的 7 类表情[1]。从左至右分别为愤怒(anger)，厌恶(disgust)，恐惧(fear)，喜悦(happiness)，中性(neutral)，悲哀(sadness)和惊讶(surprise)。

在人脸表情的分析方面，文献[2]综述了人脸表情的分类方法，可见许多经典的模式识别方法都可用于人脸表情的分类。例如，M.J. Lyons 等人[3]利用主分量分析和线性判别分析在 JAFFE 表情库上得到了很好的结果（正确率达到 92%）；C. Padgett 等人[4]使用反向传播算法训练人工神经网络在 Ekman 人脸表情库[5]上能够达到 86%的正确率；M. Pantic 等人[6]使用专家系统方法在 265 张双视角的人脸表情图片上的正确率达到了 91%；T. Otsuka 等人[7]采用隐马尔可夫模型的方法也达到了比较好的结果（该论文中并没有提到此方法的正确率）。最近，支持向量机被成功地用于人脸图像处理的各个方面，例如人脸检测[8]，人脸的姿态估计[9]和人脸识别[10]等，取得了优于传统模式识别方法的性能。与此同时，Viola 和 Jones[11]提出了一种基于 Haar 特征的 AdaBoost 方法，可以进行实时的人脸检测；M.S. Bartlett 等人[12]使用基于 Gabor 特征的 AdaSVM 方法在 Cohn-Kanade 的人脸表情数据库[13]上取得了相当好的分类结果。我们受其启发，使用基于查找表(Look-Up-Table, LUT)型弱分类器的 Adaboost 方法[14]，进行人脸表情分类器的学习。实验表明该方法与 SVM 方法相比，对于

人脸表情分类的正确率相当，而速度快近 300 倍，接近于实时。

2 人脸表情自动分类系统的基本框架

人脸表情自动分类的一般框架包括如下几个步骤。给定一幅图像或者视频序列，首先需要其中检测出人脸，然后提取人脸特征和进行人脸样本的规范化，最后进行人脸表情的自动分类，如图 2 所示。

我们只考虑正面人脸的表情分类。由于正面人脸检测问题已经基本得到解决，最具代表性的工作是 Viola 等人[11]提出的基于 Haar 特征，使用 Adaboost 方法得到的瀑布型分类器，我们也采用类似的方法进行人脸检测。人脸特征抽取过程我们使用了基于纹理到形状映射的方法[15]，它抽取出三个特征点（双眼中心和嘴中心）用于样本的几何规范化。规范化后的样本用于人脸表情分类器的学习。

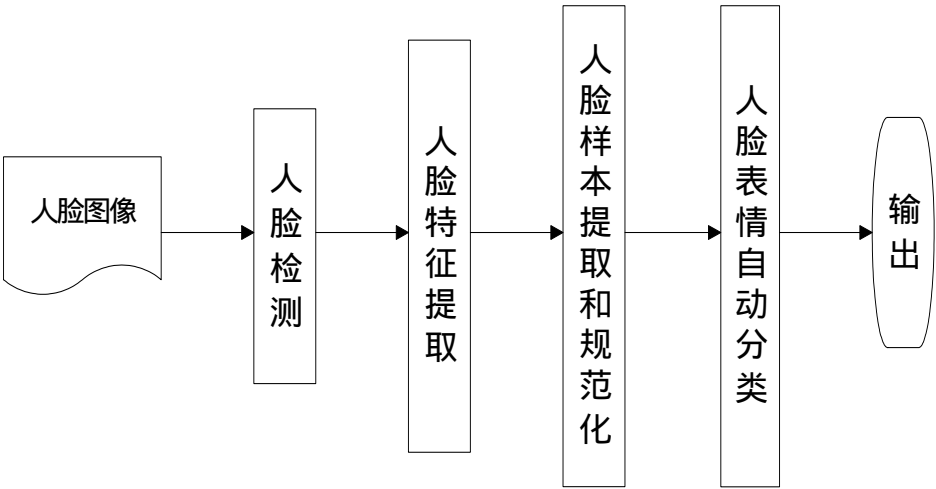


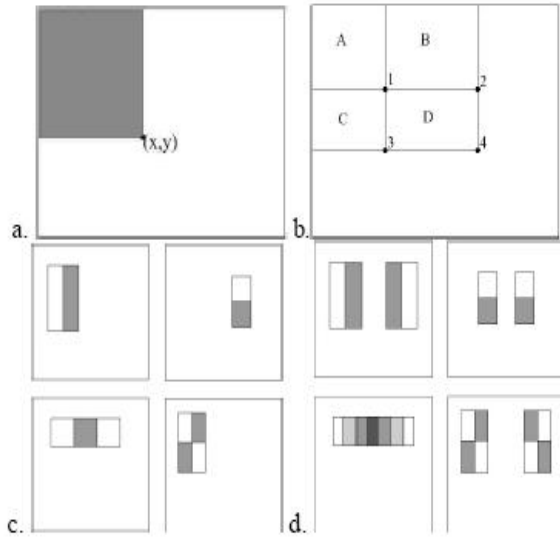
图 2 人脸表情自动分类系统的基本框架

3 基于查找表型弱分类器的连续 Adaboost 方法

3.1 Haar 特征和积分图像

Haar 型特征是 Viola 等人[11]提出的一种简单矩形特征，因类似于 Haar 小波的计算而得

名。如图 3 所示，Haar 型特征的定义是黑色矩形和白色矩形在图像子窗口中对应的区域的灰度级总和之差，Haar 型特征反映了图像局部的灰度变化。



a. b. : 积分图像在 1 点的值表示区域 A 的灰度级总和，简记为 A，在 2 点的值为 A+B，在 3 点的值为 A+C，在 4 点的值为 A+B+C+D。于是，由 1, 2, 3, 4 围成的矩形区域 D 的灰度级总和可以表示为 1+4-2-3。(方便起见，i 即表示积分图像在 i 点的值)；
c. d. : 外部方框表示图像的子区域，内部小矩形表示 Haar 特征。Haar 特征值为白色区域对应块的灰度级总和减去黑色区域对应块的灰度级总和。

图 3 Haar 特征和积分图像。

Haar 型特征的计算是通过积分图像实现的。灰度图像 $I(x, y)$ 的积分图像 $\tilde{I}(u, v)$ 定义为

$$\tilde{I}(u, v) = \int_{x=0}^u \int_{y=0}^v I(x, y) dx dy$$

可见，积分图像中一点 P 的值就是灰度图像中以原点和 P 为对角点的矩形区域的灰度总和。借助积分图像，可以通过类似查表的方式快速计算 Haar 型特征，见图 3。

3.2 连续的多分类 Adaboost 算法

Adaboost 是一种机器学习算法[16]，能够自动的从整个弱分类空间中挑选出若干个组成一个强分类器，所谓弱分类器是指正确率稍大于 50% 的分类器，即稍好于随机猜测，通常我们很容易产生一系列的这样的弱分类器，构成弱分类器空间。最终的强分类器具有如下的

形式： $h(x) = \text{sgn}(\sum_{i=1}^T a_i h_i(x) - b)$ 。其中， h_i 是弱分类器， T 为弱分类器的个数， b 为阈值。

可以看出，最终强分类器在形式上类似于线性感知机。

Adaboost 是一个贪婪算法，每一轮根据当前的样本概率分布 D_t 选取使

$r_t = \sum_{i=1}^m D_t(i) y_i h_t(\mathbf{x}_i)$ (其中 m 为样本总数) 最大化的弱分类器。在离散的情况下, 即弱分

类器只输出 ± 1 时, $1 - r_t$ 可以看作是 h_t 在 D_t 下的错误率。此时 r_t 最大化就是使错误率最小化。在找到当前最优弱分类器后, Adaboost 动态调整样本的概率分布, 增加错分样本的权重, 减小正确样本的权重, 这样在下一轮中当前错分的样本会得到更多的重视。连续的 Adaboost 算法要求弱分类器能够输出一个表示置信度的连续值, 这种连续的置信度能够更精确的反应样本的概率特性。关于连续 Adaboost 算法的收敛性有如下不等式[16]:

$$\frac{1}{m} |\{i: H(\mathbf{x}_i) \neq y_i\}| \leq \prod_{t=1}^T Z_t$$

当弱分类器 h_t 的正确率大于 50% 时, Z_t 总小于 1[16]。

对于多类情况, 假定共有 k 类, 记 $\mathcal{Y} = \{1, \dots, k\}$ 。将弱分类器看作是 $\mathcal{X} \times \mathcal{Y} \rightarrow [-1, 1]$ 的映射。定义指标函数

$$Y(i, l) = \begin{cases} 1 & y_i = l \\ -1 & y_i \neq l \end{cases}$$

借助上述弱分类器和指标函数的概念就可以将 Adaboost 推广到多类情况。连续多分类 Adaboost 算法[16]如下所示:

-
- 给定训练样本集合 $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)$, 其中 $\mathbf{x}_i \in \mathcal{X}$ 为样本向量, $y_i \in \mathcal{Y}$ 为类别标签, m 为样本总数。初始化样本概率分布 $D_1(i, l) = 1/(mk)$, $i=1, \dots, m$, $l=1, \dots, k$ 。
 - 对 $t = 1, \dots, T$ (T 为要选择的弱分类器个数):
 1. 在分布 D_t 下, 选择一个弱分类器 $h_t(\mathbf{x}_i, l)$, 使 $r_t = \sum_{i,l} D_t(i, l) Y(i, l) h_t(\mathbf{x}_i, l)$ 最大化;
 2. 令 $\alpha_t = \frac{1}{2} \ln \left(\frac{1+r_t}{1-r_t} \right)$;
 3. 更新样本概率分布

$$D_{t+1}(i,l) = \frac{D_t(i,l) \exp(-\alpha_t Y(i,l) h_t(\mathbf{x}_i, l))}{Z_t}, \text{ 其中 } Z_t \text{ 是归一化因子。}$$

- 最终的强分类器为

$$H(\mathbf{x}) = \arg \max_l \left(\sum_{t=1}^T \alpha_t h_t(\mathbf{x}, l) \right).$$

多类 Adaboost 算法中的 r_t 称为 Hamming 损失，在 Hamming 损失意义下可以保证弱分类器的正确率总大于 50%。文献[16]证明了多类连续 Adaboost 的一个误差上界为

$$\sqrt{k-1} \prod_{t=1}^T Z_t.$$

3.3 LUT 型弱分类器

Adaboost 只是一个学习过程，最终强分类器的性能还是要依赖于原始的弱分类器空间的特征。Viola 和 Jones[11]采用了阈值型的弱分类器（图 4），即 $h(\mathbf{x}) = \text{sgn}[f_{\text{Haar}}(\mathbf{x}) - b]$ ，其中 f_{Haar} 为 Haar 型特征值， b 为阈值。其中每个弱分类器对应于一个 Haar 特征，统计这个 Haar 特征在正例样本和反例样本上的特征值，找到一个最佳分类阈值。阈值型弱分类器虽然计算简单，但是太过于局限，基本上只能适用于两类都满足单高斯分布的情况。但是在连续 Adaboost 的过程中，样本的权重更倾向于不规则的类似多高斯模型的分布，同时弱分类器必须能给出一个表示置信度的实数值。为此，我们采用了 LUT 型的弱分类器（图 4），它几乎可以模拟任何概率分布，这样就不必为样本的分布做很强的假设。

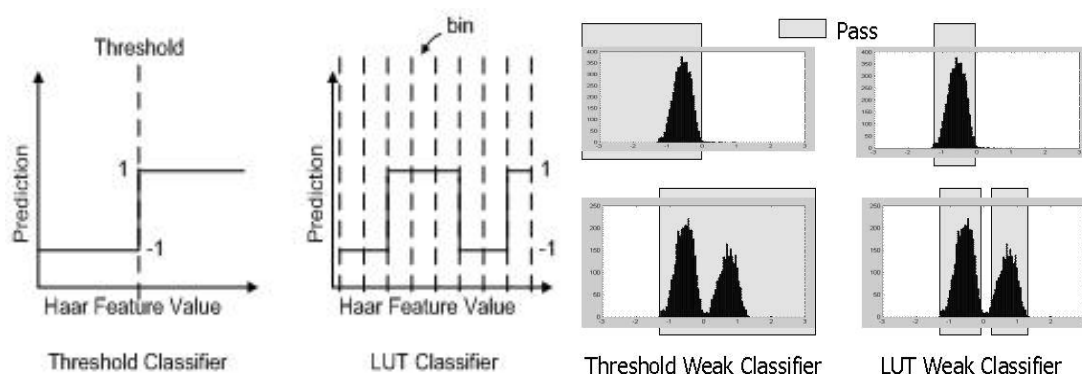


图 4 阈值型弱分类器和 LUT 型弱分类器。直方图表示样本的概率分布。

假定 Haar 特征值 f_{Haar} 已经归一化到 $[0, 1]$,将其值域 n 等分 ,记 $bin_j = [(j-1)/n, j/n)$, $j=1, \dots, n$ 。对于两类(w_1, w_2)问题 ,若 $f_{Haar}(\mathbf{x}) \in bin_j$,则 $h(\mathbf{x}) = P_1^{(j)} - P_2^{(j)}$,其中 $P_1^{(j)}, P_2^{(j)}$ 分别为两类样本的 Haar 特征值落在区间 bin_j 内的条件概率 , 即 $P_i^{(j)} = P(\mathbf{x} \in w_i | f_{Haar}(\mathbf{x}) \in bin_j)$, $i=1, 2, j=1, \dots, n$ 。定义示性函数

$$B_n^j(u) = \begin{cases} 1 & u \in [j-1/n, j/n) \\ 0 & u \notin [j-1/n, j/n) \end{cases}, j=1, \dots, n$$

则 LUT 型弱分类器可以表示为阶梯函数的形式

$$h_{LUT}(\mathbf{x}) = \sum_{j=1}^n (P_1^{(j)} - P_2^{(j)}) B_n^j(f_{LUT}(\mathbf{x}))$$

对于 k 分类(w_1, \dots, w_k)且 $k>2$ 的情况 ,可以采用二维 LUT 的形式 ,若 $f_{Haar}(\mathbf{x}) \in bin_j$, 则 $h(\mathbf{x}, l) = 2P_l^{(j)} - 1, l=1, \dots, k$ 。二维 LUT 也可以类似的写为阶梯函数的形式。定义二元示性函数

$$B_n^{j,l}(u, y) = \begin{cases} 1 & u \in [j-1/n, j/n) \wedge y = l \\ 0 & otherwise \end{cases}, j=1, \dots, n, l=1, \dots, k$$

则二维 LUT 弱分类器可以形式地写为

$$h_{LUT}(\mathbf{x}, y) = \sum_{j=1}^n \sum_{l=1}^k (2P_l^{(j)} - 1) B_n^{j,l}(f_{LUT}(\mathbf{x}), y)。$$

4 人脸表情分类器的实现

我们使用 JAFFE(Japanese Female Facial Expression)数据集[1]来进行人脸表情分类器的训练。这个数据集包括 10 个日本女性的 7 种表情 , 共 213 张图片 , 从中抽取人脸表情样本并作规范化 , 如图 5 所示。为了提高 Adaboost 的泛化能力 , 定义如下变换 : (1)左右镜像 , (2)纵向拉伸 1.1 倍 , (3)尺度放大 1.1 倍 , (4)左右旋转 5 度 , 这些变换及其组合用于生成伪样

本，由每个人脸生成 24 个不同的原始人脸样本，这样共计有 5112 个人脸表情样本。采用多类连续 Adaboost 算法训练人脸表情分类器，此时 $k=7$ 。在 Adaboost 算法中，选择 400 个基于 Haar 特征的弱分类器。

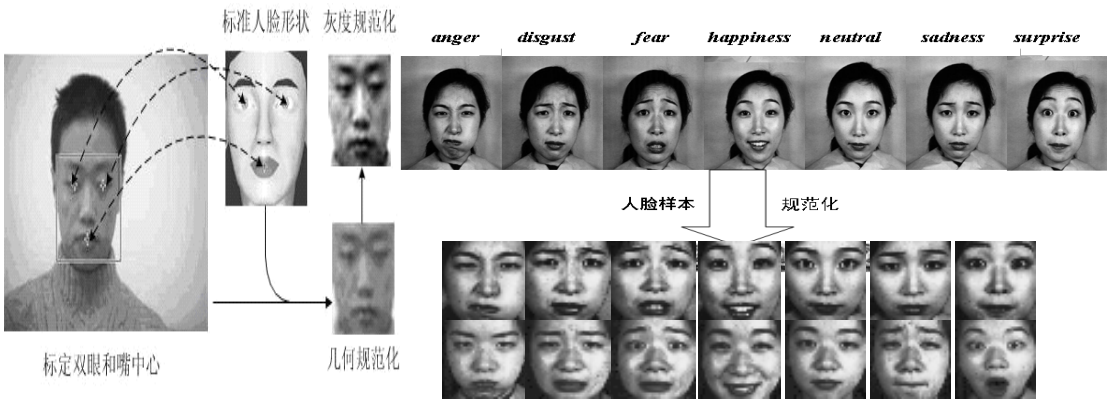


图 5 人脸表情样本的规范化

为了进行比较，我们使用支持向量机（Support Vector Machines，SVMs）方法[17]进行训练，选择径向基函数为核函数，用上述 7 类共 5112 个人脸表情的样本训练 SVM 分类器，得到了共 1870 个支持向量。

在得到人脸表情分类器之后，我们首先在 JAFFE 数据集[1]上进行了测试，达到了 99.1% 的正确率，由于这部分属于训练集合内，所以获得的正确率非常高。为了对算法的推广性能进行评价，我们在独立于训练集的图像库（包括扫描的普通照片、Internet 上下载的图片、摄像机抓取的图片等，这些图片共有 206 幅，包括了 385 个正面人脸表情）上进行了测试。实验中使用一台 CPU 为 Pentium IV 2.53GHz，内存为 512MB 的计算机，实验结果如表 1 所示。可以看出，基于 Adaboost 的方法运行速度快，比基于 SVM 的方法快近 300 倍，基本达到实时，且性能几乎与基于 SVM 的方法相当。图 6 是使用人脸表情分类器在真实图片上进行人脸表情识别的一些结果。

表 1 人脸表情的分类结果

方法	RBF-kernel SVM	LUT Adaboost
----	----------------	--------------

正确率	92.6%	91.4%
每幅图平均处理时间	28.7 毫秒	0.11 毫秒



图 6 在真实图片上人脸表情识别的一些结果

与此同时，我们研究了一些错分样本的实例。究其原因，一是样本的分辨率太低；二是样本距离表情类别的边界很近（如图 7 所示，这三幅图片的表情从左至右分别为中性、惊讶、惊讶；但是我们的表情分类器把它们分类为厌恶、喜悦和恐惧）。尤其是第二种样本，主要因为人脸的表情是受人心理活动的控制，有时表现出的表情很复杂，连人都很难判断出人脸确切的表情。这些错分的样本大都来自互联网图片。在真实照片中这种情况几乎不可避免，如何增强算法对这类情况的鲁棒性也是进一步要考虑的问题。



图 7 一些错分样本的实例

上述实验结果表明，在多分类的情况下，基于 LUT 型弱分类器的连续 Adaboost 算法在

分类正确率上完全可以与核函数为 RBF 的 SVM 相当,而后者被认为是目前在正确率上最优的分类算法。由于通过积分图像可以快速计算 Haar 特征,Adaboost 分类器的形式也比较简单,从而,在速度上 Adaboost 分类器比 SVM 分类器快近 300 倍,基本能够达到实时处理。所以,基于 LUT 型弱分类器的连续 Adaboost 方法是一种高效的图像模式识别方法,具有非常优越的性能。

5 总结及展望

本文提出了一种基于 LUT 型弱分类器的连续 Adaboost 方法,并将这种方法应用于人脸表情的分类问题。LUT 型弱分类器几乎能够近似模拟任何形式的概率分布,这就使得我们不必对样本概率分布做很强的假设。Adaboost 是一种通用的学习算法,将若干个弱分类器整合成类似于线性感知机的强分类器,而后者的性能远远优于前者。Adaboost 方法在正确率上与 SVM 相当,而速度上比 SVM 快近 300 倍,基本可以达到实时,在实时监控系统上有很好的应用前景。

基于 LUT 型弱分类器的连续 Adaboost 方法是一种标准的通用的方法,不仅可以应用于人脸表情的分类,对大多数两类和多类的图像模式识别问题也同样适用。

6 参考文献

- [1] Lyons M J. The Japanese Female Facial Expression (JAFPE) Database [DB], <http://www.mis.atr.co.jp/~mlyons/jaffe.html>, 1998.
- [2] Pantic M and Rothkrantz Leon J M. Automatic Analysis of Facial Expressions: The State of Art [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(12): 1424 ~ 1445.

- [3] Lyons M J, Budynek J, and Akamatsu S. Automatic Classification of Single Facial Images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(12): 1357 ~ 1362.
- [4] Padgett C and Cottrell G W. Representing Face Images for Emotion Classification. In: Proceedings of Conference Advances in Neural Information Processing Systems [A], Denver, USA, 1996. 894 ~ 900.
- [5] Ekman P and Friesen W V. Unmasking the Face [M]. New Jersey: Prentice Hall, 1975.
- [6] Pantic M and Rothkrantz L J M. Expert System for Automatic Analysis of Facial Expression [J]. Image and Vision Computing J., 2000, 18(11): 881 ~ 905.
- [7] Otsuka T and Ohya J. Spotting Segments Displaying Facial Expressions from Image Sequences Using HMM [A]. In: IEEE Proceedings International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998. 442 ~ 447.
- [8] Liang Luhong, Ai Haizhou, Xiao Xipan, et al. Face Detection Based on Template Matching and Support Vector Machines [J]. Chinese Journal of Computers, 2002, 25(1): 22 ~ 29. (in Chinese)
- (梁路宏 , 艾海舟 , 肖习攀等。基于模板匹配与支持矢量机的人脸检测[J]。计算机学报 , 2002, 25(1): 22 ~ 29。)
- [9] Huang J, Shao X, and Wechsler H. Face pose discrimination using support vector machines [A]. In: IEEE Proceedings International Conference on Pattern Recognition, Brisbane, Australia, 1998. 154 ~ 156.

- [10] Guo Guodong, Li S Z, Chan Kapluk. Face recognition by support vector machines [A]. In: IEEE Proceedings International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000. 196 ~ 201.
- [11] Viola P and Jones M. Rapid Object Detection using a Boosted Cascade of Simple Features [A]. In: IEEE Proceedings International Conference on Computer Vision and Pattern Recognition, Hawaii, USA, 2001. 511 ~ 518.
- [12] Bartlett M S, Littlewort G, Fasel I, et al. Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction [A]. In: IEEE Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction, Wisconsin, USA, 2003.
- [13] Kanade T, Cohn J F, and Tian Y L. A comprehensive database for facial expression analysis [A]. In IEEE Proceedings International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000. 46 ~ 53.
- [14] Wu Bo, Ai Haizhou, Huang Chang. LUT-Based Adaboost for Gender Classification [A]. In: International Conference on Audio- and Video-Based Biometric Person Authentication, Guildford, UK, 2003. 104 ~ 110.
- [15] Ai Haizhou, Xiao Xipan, and Xu Guangyou. Face Detection and Retrieval [J]. Chinese Journal of Computers, 2003, 26(7): 874 ~ 881. (in Chinese)
(艾海舟 ,肖习攀 ,徐光祐。人脸检测与检索 [J]。计算机学报 ,2003, 26(7): 874 ~ 881。)
- [16] Schapire R E and Singer Y. Improved boosting algorithms using confidence-rated predictions [J]. Machine Learning, 1999, 37(3): 297 ~ 336.

- [17] Moghaddam B and Yang M H. Gender Classification with Support Vector Machines [J],
IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 707 ~ 711.