

实时鲁棒的自动人眼状态分类

李召荣, 艾海舟

(清华大学计算机科学与技术系, 北京 100084)

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

摘要: 本文探讨了自动人眼状态分类问题。提出一种基于全局扫描并验证策略的分类框架。该方法采用一种级联结构 (Cascade) 来组织分类器, 采用 Adaboost 算法学习分类器。实验表明这种方法无论在鲁棒性、正确率和速度方面都达到了很好的性能, 具有非常明显的实际应用价值。

关键词: 人眼状态分类; 全局扫描并验证; Adaboost; 级联结构

中图法分类号: TP391

Real-time Robust Automatic Eye State Classification

LI Zhaorong, AI Haizhou

Department of Computer Science and Technology, Tsinghua University, Beijing, 100084

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

Abstract: In this paper, we discuss the problem of automatic human eye state classification. A new classification framework based on a global scan and verification strategy is proposed. This method employs a cascade structure to organize a series of eye state classifiers trained by Adaboost. Experiments over a large dataset show that proposed system reaches a good performance in robustness, correctness and speed, which has significant value in real applications.

Key Words: Eye state classification; Global scan and verification; Adaboost; Cascade

在过去的十到二十年内, 人脸信息处理已经成为计算机视觉和模式识别领域的一个研究热点。作为人脸的主要部分, 人脸器官 (眼睛和嘴巴) 以及它们的状态 (张开和闭合), 包含了丰富的表情和感情信息。它在许多人脸视觉系统中都有重要应用, 比如驾驶员的疲劳驾驶检测, 活性验证, 人脸表情分析等。

但是一直以来, 这个问题并没有引起大家的重视, 也没有专门的、系统的研究文献发表。仅有的少量的研究存在于某些人脸分析系统中, 被当作一个子问题来对待。这些研究也通常是一些经典的模式识别方法的简单应用。以眼睛为例, Chu 等^[1]在研究驾驶员的疲劳检测中, 采用一个 3 层的 BP 网络来进行眼睛睁闭分类。他们采用的特征是一个由 6 个几何参数: 眼宽, 眼高以及 4 个弧度参数构成的 6 维向量。他们在一个自建的包含 124 张图片的数据集上取得了 85% 的正确率。类似地, Tian 等^[2]也采用神经网络来进行分类, 但抽取的特征是包含眼睛区域纹理的 Gabor 变换系数。他们在一个自建的数据集上获得了 93% 的正确率。邓等^[3]在研究活性判别算法中, 使用支持向量机来对直方图均衡化后的人眼灰度块直接分类。他们采用大约 1300 张图片进行实验, 根据选择的核函数不同, 支持向量的个数大概在 60~200 之间, 分类正确率大概在 90%~97% 之间。上述这些方法大多假设特征点定位非常准确, 或者直接采用手工标定的特征点, 而这对通常的自动定位系统来说比较困难。除了分类的方法

本课题得到国家自然科学基金 (60332010, 60673107) 的资助。李召荣, 男, 1982 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, 博士, 教授, 博士生导师, 主要研究方向为计算机视觉。

以外，还有一类基于检测的方法。例如，Veeraraghavan 等^[4]在检测驾驶员疲劳的研究中，直接检测睁眼和闭眼。他们使用基于模板匹配的方法来检测睁眼和闭眼。模板共四种：左眼睁模板，左眼闭模板，右眼睁模板，右眼闭模板。该方法的正确率达到 85% 以上。

鉴于上面提到的原因，本文系统地试验并比较了一些经典的机器学习方法，包括 Fisher 线性鉴别 (Fisher Linear Discriminant, FLD) 分析，支持向量机 (Support Vector Machine, SVM)，以及基于 Haar 型矩形特征的 Adaboost 算法。实验表明，在手工标定特征点的测试集合上，三种经典的分类算法都取得了非常高的分类正确率。然而，不幸的是，在自动定位特征点系统中，三种分类算法的分类正确率都大幅度下降。究其原因，我们发现三种分类算法都对特征点的位置非常敏感，有时甚至 1~2 个像素的偏差可能导致完全相反的结果，而在自动系统中，由于特征点定位算法的局限，不可能产生手工标点那么准确的定位，因而分类性能大幅下降不可避免。一些简单的思想，比如扰动并投票，统计特征点定位的误差分布等，可以在一定程度上缓解这个问题，但并不能令人满意。为了从根本上解决特征点定位不准的影响，我们提出一种全局扫描并验证的策略，其基本思想就是在包含器官的一个大块区域中对所有的子窗口扫描并逐个判断其状态。为了快速高效实施这种策略，受 Viola 等人的人脸检测工作的启发，我们采用 Cascade 结构来组织分类器，采用 Adaboost 方法学习分类器。实验表明这种方法无论鲁棒性、正确率还是速度都达到了很好的性能，具有非常明显的实际应用价值。

为方便叙述，本文的论述都以眼睛为例，嘴的技术完全一样，只是样本不同而已。我们的系统采用同样的算法实现了眼和嘴部状态的分类。

1 人眼状态自动分类系统的基本框架

人眼状态自动分类系统的一般框架包括如下几个步骤。给定一幅包含人脸的图像或者视频序列，首先需要在其中检测出人脸，然后定位人脸特征点再提取人眼样本，最后进行人眼状态自动分类，如图 1 所示。在我们的系统中，人眼状态分为两类：闭眼和睁眼。

我们只考虑正面人脸的眼睛状态分类。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^[5]提出的基于 Haar 型矩形特征，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^[6]，它抽取出六个特征点的位置（双眼眼角和嘴角），其中双眼眼角的位置用于人眼样本的定位和提取。在本文中，我们仅仅详细讨论人眼状态分类模块（图 1 中以粗线标出）。

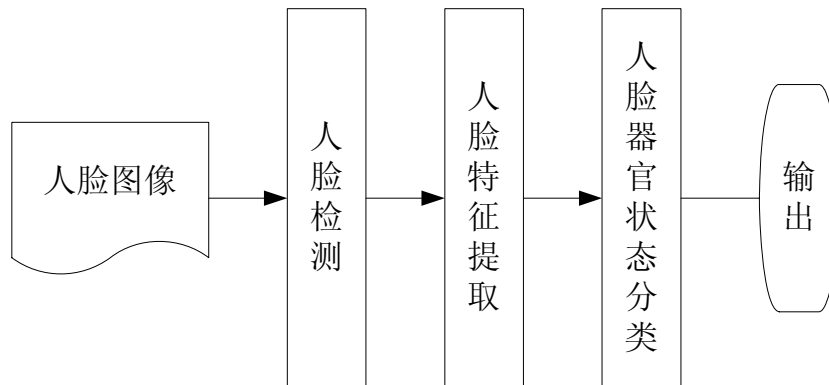


图 1 人眼状态自动分类系统的基本框架

2 经典算法对人眼状态分类的结果

我们实验并比较了 SVM、FLD 和基于 Haar 型矩形特征的 Adaboost 算法。实验的过程如下：首先，在一个手工标定了 6 个特征点（双眼眼角和嘴角）人脸图片的集合上，根据两个眼角的位置提取正方形人眼区域，宽度为人眼实际宽度，并且眼角连线上半部分和下半部分的比例为 2: 3（根据经验），然后进行灰度校正，最后缩放到 24x24 像素大小，如此就得到样本集，然后将灰度向量直接作为特征向量。我们的实验在一个自建的包含 3500 张人脸图片集上进行。数据集包含约 350 个人，男女数量相当，人种为亚洲人，年龄在 15~70 岁之间，室内光照，分辨率较低。图片大小为 240x320 像素，人脸宽度大约 200 像素。其中 1500 张闭眼人脸和 2000 张睁眼人脸，全部经过手工标定好 6 个特征点。通过水平镜像将右眼翻转成左眼，一共切割得到 3000 个闭眼样本和 4000 个睁眼样本。我们挑选出 1000 个闭眼样本和 1000 个睁眼样本用于训练，剩下的 2000 个闭眼样本和 3000 个睁眼样本用于测试。图 2 列出了一些规范化后的左眼样本，大小为 24x24，第一行为闭眼样本，第二行为睁眼样本。



图 2 规范化后的左眼样本

(1) SVM

SVM 是一种基于结构风险最小化的统计学习理论的学习机，被广泛应用于模式识别各个分支，它最早由 Vapnik 等提出。它特别适用于高维小样本问题，有很好的推广能力。在我们的实验中，使用 RBF 径向基函数核函数，最后训练得到的支持向量个数为 32。实验结果见表 1。

(2) FLD

FLD 试图寻求一个投影方向，使得对两类样本鉴别最好。求得最佳投影方向 \mathbf{w}^* 后，将所有样本投影到最佳投影方向 $y = \mathbf{w}^{*T} \mathbf{x}$ ，并且选择一个阈值 y_0 来划分两类。在实验中，

我们选择的阈值为 $y_0 = \frac{n_1 \tilde{m}_1 + n_2 \tilde{m}_2}{n_1 + n_2}$ ，其中 n_i 是样本个数， \tilde{m}_i ($i=1,2$) 是投影后样本的

均值。实验结果同样在表 1 中。对比 FLD 和 SVM 可以看出，FLD 取得和 SVM 可比的正确率，而且速度更快。

(3) 基于 Haar 型矩形特征的 Adaboost 算法

Adaboost 是一种近年来被广泛应用的学习算法，它最早由 Freund 和 Schapire^[7]等提出。这种算法的主要思想是试图从一个大的弱分类器空间中选出一小部分弱分类器，将它们组合起来构成一个强分类器。大量的实验表明，它拥有与支持向量机可比的分类性能和推广能力，但快得多的计算速度，因而获得广泛应用。

在我们的实验中，采用了 Viola 等提出的基于 Haar 型矩形特征的 Adaboost 算法^[5]，并且采用 Wu 等提出的查找表型弱分类器^[11]。这种 Adaboost 算法的实现快速高效，分类能力强，并且在人脸检测问题上取得成功，很适合我们的实时、高效要求。我们训练得到的弱分类器的个数是 400，实验结果同样列在表 1 中。比较看出，FLD、Adaboost 都取得了和 SVM 可比的正确率（Adaboost 略高于 SVM），但速度快很多。而且对三种算法，闭眼的分类正确

率都要高于睁眼，这说明睁眼样本的分布更复杂，更分散。

表 1. 经典算法对人眼状态分类的结果

算法	闭眼正确率	睁眼正确率	平均计算时间 (ms)
SVM	0.99	0.9463	0.1846
FLD	0.9875	0.9503	0.028
Adaboost	0.9945	0.979	0.05

3 基于全局扫描并验证策略的自动人眼状态分类

从第 2 节的实验结果 (表 1) 可以看出, 3 种分类算法 SVM, FLD, Adaboost 都在手工标定特征点的集合取得了很高的正确率, 特别是 Adaboost 算法, 平均正确率接近 99%, 而且计算速度相当快。因此, 一种很自然的期望就是在自动定位特征点的情况下, 3 种算法或者其中某种算法也能取得相当的分类性能。然而, 不幸的是, 事实并非如此, 通过实验我们发现, 3 种分类算法都对特征点 (眼角) 的位置十分敏感, 甚至 1~2 个像素的偏差都可能使得分类器的输出完全相反。与手动标定的特征点不同, 在自动系统中, 由于特征点定位算法的限制, 不可能产生手工标点那么准确的特征点定位, 误差在所难免, 其结果就是使得分类正确率大幅下降 (图 3 是我们的定位算法在实验数据库上的定位误差的一个分布示意图, 分别统计两个眼角的 x 和 y 坐标, 相对误差定义为: (自动定位坐标 - 手工定位坐标) / 眼睛宽度)。我们在实验数据集上测试了自动分类, 结果证实了这一点。3 种算法的结果都列在表 2 中。可以看出, 不管是那种算法, 都无法克服特征点定位误差的影响。

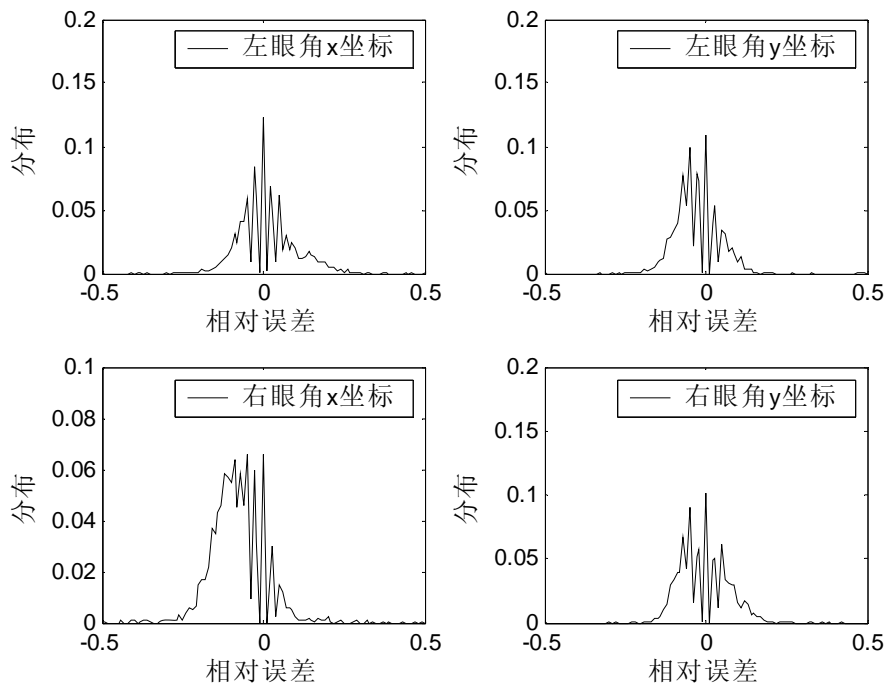


图 3 特征点自动定位算法对左眼定位的相对误差分布图

3.1 简单的改进办法

为了提高自动系统的分类正确率, 最直接的方法就是改进特征点定位算法的精确度, 可是, 从目前这方面的研究现状来看, 这并不是一件容易的事, 而且精确度高的算法往往比较

慢，不符合实时要求。我们的工作重点在于改进分类策略。一种比较容易想到的改进策略是：在训练每种分类器时，根据自动特征点定位算法的误差分布，对手工标定的特征点随机扰动，将扰动后的样本同样加入到训练集中进行训练；而在测试（自动分类）时，对特征点定位算法定位的结果进行扰动，产生一定数量的样本（一般 20 个左右），逐个用分类器判断，然后对分类器的输出进行投票，最后的结果就是得票多的一类。我们对 Adaboost 算法实施了这种策略，然后同样在实验数据库上进行了测试，结果同样列在表 2 中，以方便比较。可以看出，这种策略确实一定程度上提高了分类正确率，但并不能令人满意。究其原因，主要是因为自动系统中，特征点定位的误差是随机的，真实的误差并不知道，因而扰动加投票的目的性不能保证，只能是一种随机近似。而且这种改进方法还依赖于不同的特征点定位算法。

3.2 全局扫描并验证策略和级联分类器结构

通过以上的几组实验，我们应该可以得出以下几条结论：1. 人眼睁开样本的分类并不困难，简单的灰度特征加上简单的线性鉴别分析就可以达到很高的分类正确率；2. 自动分类系统的正确率很低，并不是因为分类器的分类性能不好，而是由于特征点自动定位误差造成的，而且睁眼受到的影响更明显（这主要是因为睁眼样本的分类更复杂，更分散）。为了从根本上克服特征点定位误差的影响，我们提出了一种全局扫描并验证的策略。它的主要思想就是对包含眼睛的纹理块逐个扫描其所有子窗口，并用闭眼分类器和睁眼分类器判断它们是闭眼样本还是睁眼样本，然后分别统计闭眼样本个数和睁眼样本个数，个数多的一类就是最后的结果。它的主要步骤如下：首先，根据自动定位到的眼角的位置（不需要很准），在人脸图片上切割出一块包含双眼的纹理块，旋转并缩放，使得双眼连线水平并且两眼中心的距离大约是 48 个像素；然后，用一个 24x24 像素大小的子窗口在切割出的纹理块中所有可能的位置进行扫描，再利用闭眼分类器和睁眼分类器逐个判断当前扫描到的子窗口是闭眼样本还是睁眼样本。最后，统计扫描到的闭眼样本和睁眼样本的个数，输出个数多的一类作为最后的结果。这个过程的图示显示在图 4 中。

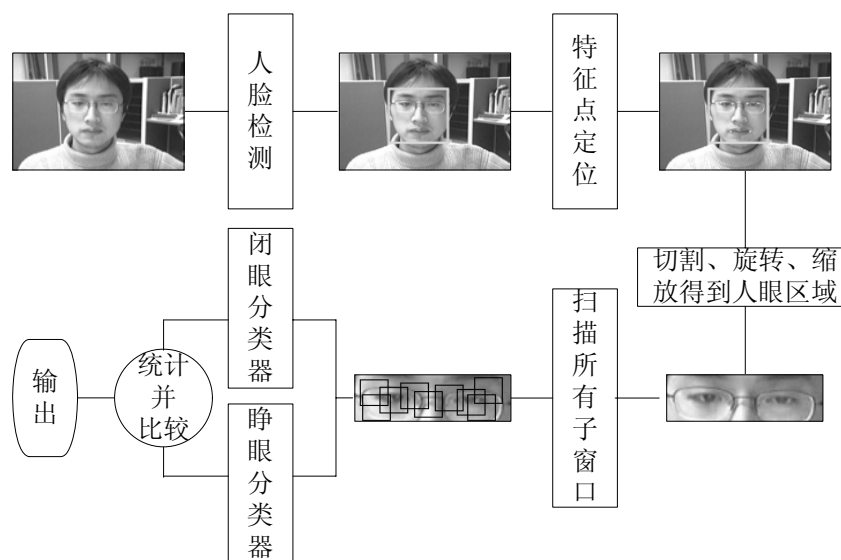


图 4 扫描并验证过程示意图

为了实施这么一个过程，我们需要考虑很多问题。例如怎么快速高效实施扫描，怎么训练分类器都是很棘手的问题。所幸的是，Viola 等人的人脸检测框架给我提供了一个很好的范例。受他们的启发，我们同样采用级联结构来组织分类器。这种结构如图 5 所示，每个数字对应级联分类器的一层，越靠前的层，分类器越简单，处理速度越快，而且每一层都保证几乎让全部正例通过而拒绝大部分反例，由于在整个搜索区域中，正例样本很少，这样大部

分子窗口就很快在前面被排除掉了，其结果就使总的计算时间大大减少，同时获得了很高的分类正确率。具体对我们的问题而言，我们需要训练两个级联分类器，一个用于检测闭眼，另一个用于检测睁眼。与人脸级联分类器（人脸检测器）的训练作一个类比，我们在训练闭眼级联分类器时，将闭眼作为正例（人脸检测中的“人脸”），将睁眼和眼睛周围的其他区域作为反例（人脸检测中的“背景”）；类似地，在训练睁眼级联分类器时，将睁眼作为正例（人脸检测中的“人脸”），将闭眼和眼睛周围的其他区域作为反例（人脸检测中的“背景”）。

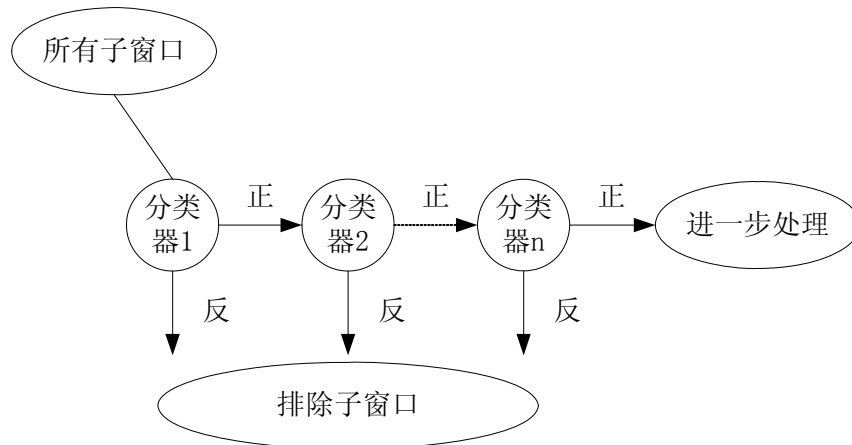


图 5 级联结构

3.3 Adaboost 级联分类器的实现

从第 2 节的实验结果可以看出，基于 Haar 型矩形特征的 Adaboost 算法具有最高的分类性能和速度优势，再加上它是一个线性组合的强分类器，因此非常适合于组织级联结构。参考 Viola 等人的人脸级联分类器的训练，我们训练两个 Adaboost 级联分类器，分别用来检测闭眼和睁眼。以训练睁眼级联分类器（检测器）为例，我们首先需要收集三个集合的样本：正例样本集 P（睁眼样本集），初始反例样本集 N（闭眼样本+非眼区域），“背景”图片集 B（不包含睁眼的人脸纹理块）。实验用的数据库同第 2 节中用到的，所有图片旋转至双眼连线水平。我们在手工标定好特征点的 2000 张睁眼图片上，通过扰动眼角 1~4 个像素（之所以扰动，主要是想弥补样本数目的不足），我们一共得到 8000 个睁眼样本，规范化到 24x24 大小，构成集合 P；从 1500 张闭眼图片集合上选出 500 张，通过扰动 1~4 个像素，得到 12000 个闭眼样本，再加上收集 8000 个非眼样本（不包含眼睛的人脸块），一共 20000 个，都规范化到 24x24 大小，构成集合 N；在剩下的 1000 张闭眼图片上，切割出人脸区域，构成集合 B。训练每一层，使用正例样本集 P 和反例样本集 N 训练出当前层的强分类器，调整其阈值，使得正例样本的通过率达到预定的目标，然后使用已经训练出来的部分层叠分类器在“背景”图片集 B 上收集误报，作为反例样本，加到反例集合 N 中，用于后面一层的强分类器的训练，当发现从“背景”图片集合上搜索不到足够的反例时，训练终止。我们训练的睁眼级联分类器共 17 层，每层的弱分类器的个数分别为：3, 6, 8, 13, 20, 30, 45, 60, 75, 90, 100, 110, 110, 110, 110, 110, 110。闭眼级联分类器的训练类似。我们同样在实验数据库上测试了级联结构的 Adaboost 算法，结果也列在表 2 中方便比较。比较可以看出，基于扫描并验证策略的级联 Adaboost 算法与其他算法相比，大大提高了分类的正确率，接近 100%。而且速度与简单的 Adaboost 算法相比，并没有明显下降。在我们的 P41.7GHZ 的机器上，完整的系统达到了 14 帧/秒的速度，基本达到实时要求。图 6 列出了一些实验结果，分别是 CMU-PIE^[12]子集上，Internet 上收集的，以及摄像头的结果（为清楚起见，状态标注黑底白字表示闭，白底黑字表示开）。

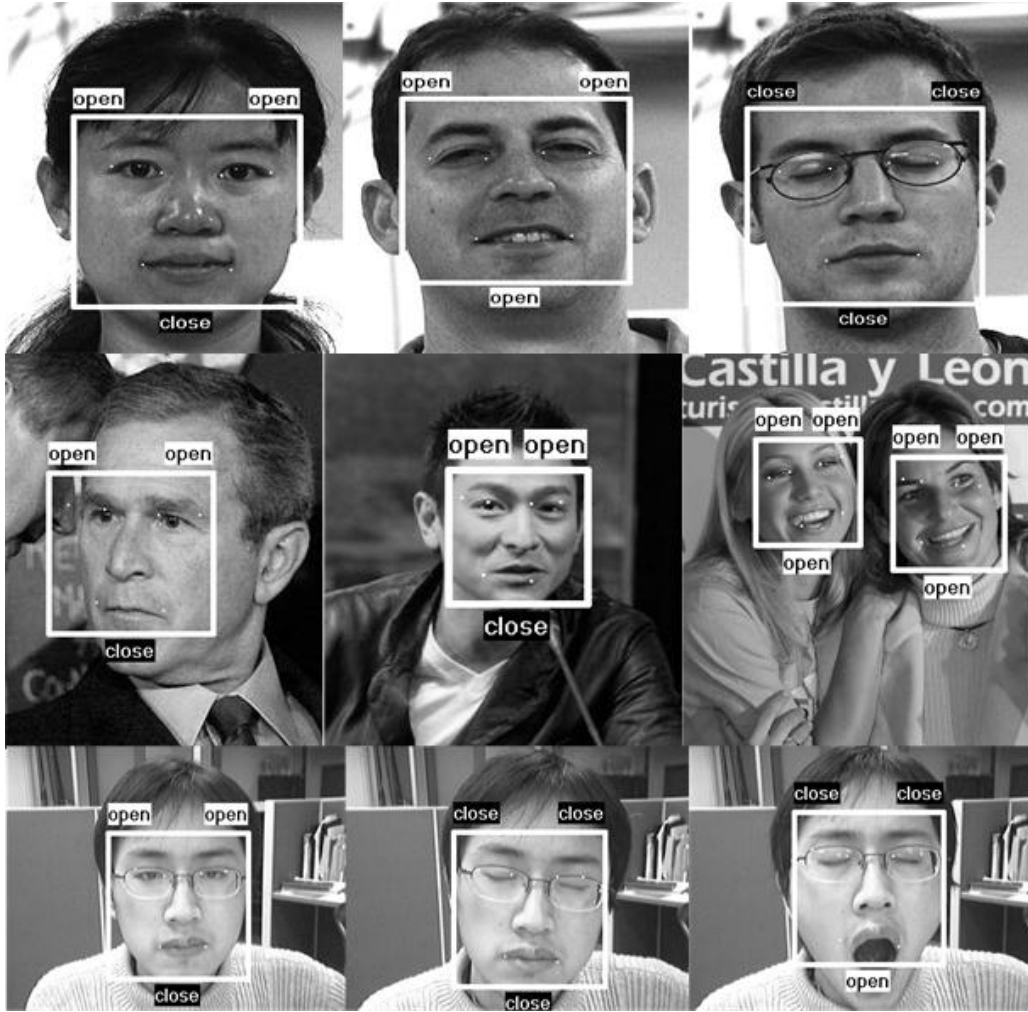


图 6 一些实验结果

为了验证我们的算法对特征点定位的鲁棒性，我们对自动定位的特征点的位置进行随机扰动，扰动的范围为双眼距离的 10% 以内（对我们的数据库，双眼距离大概 80 像素，所以扰动 1~8 个像素），然后同样在实验数据库上进行了测试，结果同样列在表 2 中，对比可以发现，特征点的扰动并未对分类性能造成明显下降，从而也从侧面证明了我们算法克服了受特征点定位误差影响这一关键性困难，达到了预期效果。

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；二是因为特征点定位算法失败，与真实位置差的太远，这导致算法错误地旋转和缩放测试图片，从而导致最终的分类错误。解决的办法是彻底去掉特征点定位这一过程，直接利用人脸检测的结果，同时训练多分辨率的带平面内旋转的人眼状态分类器，这是下一步待研究的问题。

表 2. 自动人眼状态分类结果

算法	闭眼正确率	睁眼正确率	平均计算时间 (ms)
SVM	0.856287	0.649306	9.8
FLD	0.931138	0.516898	1.5
Adaboost	0.926647	0.680324	3.2
Adaboost+简单改进	0.932635	0.845602	16.5
级联 Adaboost	0.994012	0.999074	25.1
级联 Adaboost+扰动测试	0.992515	0.997917	25.1


4 总结及展望

本文提出了一种鲁棒的基于扫描并验证策略的自动人眼状态分类算法，实验表明速度快、正确率高。对于人眼状态分类问题，我们系统地比较了几种经典的模式分类算法支持向量机、线性鉴别分析、基于 Haar 型矩形特征的 Adaboost 的分类性能，发现这几种算法都能达到很高的分类正确率。然而，我们发现，在自动分类系统中，由于受到特征点定位误差的影响，所有算法的正确率都大幅度下降。并且，一些简单的改进策略并不能有效地解决这个问题。为此，我们提出一种全局扫描并验证的策略，参考 Viola 等人的人脸检测工作，采用级联结构来组织分类器，给出了 Adaboost 算法在该结构上的实现。实验表明，我们提出的改进算法框架很好地排除了特征点定位误差的影响，分类性能有大幅提高，并且速度并没有明显下降，基本达到实时。而且该算法的存储量很小，大约 2M 左右，很适合移动设备的小容量要求。因而，潜在的应用前景看好。

5 参考文献

- [1] Jiangwei Chu, Lisheng Jin, Lie Guo, et al. Driver's Eye State Detecting Method Design Based on Eye Geometry Feature [C]. In IEEE Intelligent Vehicles Symposium, Parma, Italy, 2004, 357~362
- [2] Ying-li Tian, Takeo Kanade, and Jeffrey F. Cohn. Eye-State Action Unit Detection by Gabor Wavelets [C]. In Proceeding of the Third International Conference on Advances in Multimodal Interfaces, Beijing, China, 2000, 143~150
- [3] Deng Gang, Cao Bo, Miao Jun, et al. Liveness Check Algorithm Based on Eye Movement Model Using SVM [J]. Journal of Computer-Aided Design & Computer Graphics, 2003, 15 (7): 853~857 (in Chinese)
邓刚, 曹波, 苗军, 等. 基于支持向量机眼动模型的活性判别算法[J]. 计算机辅助设计与图形学学报, 2003, 15(7): 853~857
- [4] Harini Veeraraghavan, Nikolaos P. Papanikolopoulos. Detecting Driver Fatigue Through the use of Advanced Face Monitoring Techniques [R]. Minnesota, USA: Institute Center for Transportation Studies, University of Minnesota-Twin Cities, 2001
- [5] P. Viola and M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features [C]. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, 2001: 511-518
- [6] Tong Wang, Haizhou Ai and Gaofeng Huang. A Two-Stage Approach to Automatic Face Alignment [C]. In Proceedings of SPIE, Vol.5286, Third International Symposium on Multispectral Image Processing and Pattern Recognition, Beijing, China, 2003,558-563
- [7] R.E.Schapire, The Boosting Approach to Machine Learning An Overview [C], MSRI Workshop on Nonlinear Estimation and Classification, Berkeley, 2002: 149-172
- [8] Y.Freund and R.E.Schapire. A decision-theoretic generalization of on-line learning and an application to boosting [J]. Journal of Computer and System Sciences, 1997, 55 (1):119~139
- [9] Y.Freund and R.E.Schapire. Experiments with a new Boosting Algorithm [C]. In Proceedings of the 13th International Conference on Machine Learning, San Francisco: Morgan Kaufmann, 1996: 148-156.
- [10] R.E. Schapire and Y. Singer. Improved Boosting Algorithm Using Confidence-rated Predictions [J]. Machine Learning, 1999, 37(3): 297-336

- [11] Bo Wu, Haizhou Ai, and Chang Huang. LUT-based Adaboost for Gender Classification [C]. In Proceedings of International Conference on Audio and Video Based Biometric Person Authentication, Guildford, 2003: 104-110
- [12] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination and expression dataset[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(12): 1615-1618

	<p>李召荣，男，1982年7月生。 2004年获得中国科学技术大学学士学位，现在清华大学计算机科学与技术系攻读硕士研究生，研究方向为计算机视觉。 E-mail: lizr04@mails.tsinghua.edu.cn</p>
	<p>艾海舟，男，1964年1月生。 1985、1988、1991年分别获得清华大学学士、硕士、博士学位，1994.9-1996.8在比利时布鲁塞尔自由大学作博士后研究，现为清华大学计算机科学与技术系教授，博士生导师。研究方向为计算机视觉。 E-mail: ahz@mail.tsinghua.edu.cn</p>

联系人：艾海舟

通信地址：清华大学计算机系

邮编：100084

E-mail 地址：ahz@mail.tsinghua.edu.cn

联系电话：010-62795495