

基于 GBDT 和 HOG 特征的人脸关键点定位

张重生, 彭国雯, 于珂珂*

(河南大学计算机与信息工程学院 开封 475001)

摘 要: 人脸关键点检测是计算机视觉领域中的一个重要分支, 其检测精度直接影响人脸识别和表情分析的结果. 为此, 本文提出了一种新的方法来解决人脸关键点检测问题, 即 H-GBDT. H-GBDT 是一种基于 GBDT 决策树和 HOG 特征的人脸关键点检测算法, 该算法是将人脸图像的 HOG 特征作为 GBDT 的输入, 关键点的真实坐标作为 GBDT 的输出来训练预测模型, 在该过程中每个关键点将分纵坐标和横坐标两次在 GBDT 中做回归运算, 并经过不断的调整 GBDT 和 HOG 特征的参数来训练出最佳预测模型. 本文将在 BioID、LFW、LFPW 三种数据集上验证 H-GBDT 算法, 其中 BioID 是正脸数据集, 实验结果表明 H-GBDT 在该数据集上的检测效果最佳, 其检测误差基本上可控制在 2% 以内; LFW 和 LFPW 是自然场景下的数据集, H-GBDT 在这两种数据集上的检测误差一般在 2%~4% 之间.

关键词: 人脸关键点检测; 人脸特征; GBDT; HOG

中图分类号 : TP391

文献标识码: A

文章编号: 1003-4978 (2017) ??-????-??

Facial Points Detection based on GBDT and HOG

Zhang Chong-Sheng, Peng Guo-Wen, Yu Ke-Ke*

(The School of Computer and Information Engineering, Henan University, Kaifeng 475001)

Abstract: The detection of facial points is an important topic in the field of computer vision, its accuracy directly affects the results of face recognition and facial expression analysis. To solve the problem of facial point detection, this paper proposes a new method, which is hereafter referred to as H-GBDT. H-GBDT is an algorithm used for facial points detection based on GBDT and HOG feature. In H-GBDT, the HOG feature is used as the input of GBDT, and the real coordinates of facial points are used as the output of GBDT in order to training the predict model. Each facial point will be divided into vertical and horizontal coordinates to perform regression in GBDT during training, and through the constant adjustment of the parameters of GBDT and HOG to train the best predict model. This paper will verify the performance of H-GBDT on three datasets, which are BioID, LFW and LFPW. BioID is a frontal face dataset, and the experimental results show that the detection effect of H-GBDT is the best on BioID, the error can be controlled under 2%. The detection error of H-GBDT can be controlled between 2% and 4% on the LFW and LFPW datasets, which are datasets in the wild.

Key words: The Detection of Facial Points; the Feature of face; GBDT; HOG

0 引言

人脸关键点检测是指在人脸图像上自动地检测出比较显著区域（如鼻子、眼睛等器官或人脸轮廓）的关键点坐标. 这项检测技术是人脸识别^[1, 2, 3]、人脸追踪^[4, 5, 6]、人脸表情分析^[7, 8, 9]等研究的重要步骤, 特别是在人脸识别和人脸表情分析的研究中非常依赖人脸关键点坐标的精度. 因此这项检测技术在计算机视觉领域受到越来越多的重视并逐渐发展成熟. 目前, 人脸关键点检测技术已经取得了很大的进步, 并在商

收稿日期: 2017-10-10

基金项目: 国家自然科学基金项目“面向大规模基于位置服务的高效索引和空间查询处理技术”(41401466)

作者简介: 张重生, 男, 1982 年生, 博士, 教授, 计算机学会(CCF)会员(51690M), 主要研究领域为数据挖掘、大数据分析、深度学习

*通信作者, E-mail: 1416560660@qq.com

业、社会、国防等各个领域得到了广泛应用。

通常,实现人脸关键点检测的方法主要分为两类:模板拟合法^[10, 11, 12]和回归器学习法^[13, 14, 15]。简单地讲,模板拟合法主要是利用人脸关键点坐标之间的关系建立一个或多个几何模型,再利用模型中的点和该点周围像素的关系预测出该点的真实位置;回归器学习法是指利用一些回归算法(如随机森林, SVM 等)通过不断的迭代拟合方式来训练或学习经过预处理后样本数据生成一个或多个预测模型,然后在测试阶段加载这些预测模型来预测新输入人脸图像的关键点坐标。经过多年的努力,这项研究已经实现各种先进的成果^[14, 15, 16, 17, 18, 19, 20, 21]。其中,ASM^[10, 22, 17](Active Shape Model, 主动形状模型)是一种经典的关键点检测算法,该算法是模板拟合算法之一,并且也是当前最先进的人脸对齐算法之一。AAM^[11, 23](Active Appearance Model, 主动外观模型)是在 ASM 的基础上充分利用全局的纹理信息,并建立了全局灰度模型来代表全局纹理的变化。由于 ASM 和 AAM 都是基于贪心搜索算法关键点定位算法,因此这两种算法都依赖于输入图像的初始模型。后来随机机器学习的发展,人们开始使用级联回归^[23, 24, 14]的方法来解决关键点的检测问题,这种方法主要是使用一些级联函数通过不断的回归与迭代训练样本数据进而得到预测模型,并取得先进的检测效果。

本文通过对 GBDT^[26, 27](Gradient Boosting Decision Tree, 梯度提升决策树)算法的研究,了解到 GBDT 算法中的每棵决策树都是回归决策树,因此它几乎可以解决所有的数据回归问题并具有高检测精度的特性。此外,GBDT 在非线性数据处理方面也有着良好的效果。HOG^[28, 29](Histograms of Oriented Gradients, 梯度方向直方图)特征是在图像内的局部单元格上进行提取的特征,所以它对图像几何和光学的形变都能保持良好的不变性。因此,根据对 GBDT 和 HOG 特征性质的了解,本文提出将 GBDT 和 HOG 特征相结合的 H-GBDT 算法来解决人脸关键点的检测问题。H-GBDT 算法是将 GBDT 作为数据的回归器,将人脸图像的 HOG 特征作为 GBDT 算法的输入数据,而相对于人脸图像中心点的关键点坐标作为 GBDT 算法的输出来训练预测模型。本文将在大量数据上使用 GBDT 回归器训练图像的 HOG 特征,并通过不断调整 GBDT 算法中的树个数、学习率两个参数及 HOG 特征提取参数训练出关键点的最佳预测模型。

本文的实验数据是 BioID、LFW 和 LFPW 三个数据集,其中 BioID 和 LFPW 数据集是人脸对齐和人脸关键点检测研究中常用的数据集,LFW 数据集中的每张图像是本人从原 LFW 数据集中抽取的数据集。在 H-GBDT 的实验设计中,本文将随机从数据集中抽取 80% 左右图像作为本文实验的训练数据。实验结果表明 H-GBDT 算法在 BioID 数据集上的检测误差平均值基本控制在 2% 以内,在 LFW 和 LFPW 数据集上的检测误差平均值基本控制在 5% 以内。与 SO-RF^[30]算法和 Face++ 平台提供的人脸关键点检测模型相比,H-GBDT 算法思想和实现简单,且检测精度高,特别是在正脸图像上。

本文章节内容安排如下:引言部分介绍本文的研究背景和意义,并简单介绍了本文提出的 H-GBDT 人脸关键点检测算法的原理;第 1 章介绍 GBDT 算法和 HOG 特征的原理;第 2 章介绍 H-GBDT 算法的具体实现步骤;第 3 章介绍了 H-GBDT 算法的实验设计和 H-GBDT 与 SO-RF 算法和 Face++ 提供的人脸关键点检测模型的检测对比结果;第 4 章是本文的总结。

1 相关工作

1.1 GBDT

GBDT 算法是一种用于数据分类和回归的集成学习算法。该算法是由多棵类似于 CART^[31]的决策树组成,将算法中所有决策树的输出结果累加起来就是 GBDT 的最终输出结果,它在被提出之初就和 SVM^[32, 33](Support Vector Machine, 支持向量机)一起被称为泛化能力较强的算法,并都是数据分析中常到用的学习算法。GBDT 算法是由决策树(Decision Tree, DT)、梯度提升(Gradient Boosting, GB^[26])、Shrinkage 三个基本概念组成。其中,GB 本质上是一个框架,里面可以嵌入不同的其它算法。与传统的 Boosting 算法相比,GB 中每一次新模型的建立都是为了减少上一个模型的残差,所以每个模型的残差都是在梯度方向上减少;每个新模型的输入都是上一个模型输出,输出是下一个模型的输入,预测模型的最终结果是之前所有模型的相加结果,由此可见,GBDT 算法很难实现并行运算。下面公式是 GB 预测模型的数学表达式:

$$F(x; \{\beta_m, \alpha_m\}_{m=1}^M) = F_0(x) + \sum_{m=1}^M \beta_m h(x; \alpha_m)$$

$$F_0(x) = \arg \min \sum_{i=1}^N L(y_i, \gamma)$$

上式中 $h(x; \alpha_m)$ 代表算法的第 m 棵决策树模型, M 代表决策树的棵数; β_m 代表第 m 个模型的权重, α_m 表示第 m 个模型里面的参数; 因此为了优化预测模型 $F(x; \{\beta_m, \alpha_m\}_{m=1}^M)$, 只要优化参数 $\{\alpha, \beta\}$ 即可; $F_0(x)$ 代表算法的初始预测模型, 其值为损失函数的最小值, 其中参数 γ 是常数.

为了解决决策树训练过拟的问题, GBDT 中使用 Shrinkage^[36] 算法, 该算法是防止过度拟合中常用的一种方法, 其认为迈一小步可以更加接近真实值, 而迈一大步可以虽然更快的接近真实值, 但是也可能因为步子太大造成结果误差更大. Shrinkage 算法避免了这种情况大大提高了 GBDT 的泛化能力, 提高 GBDT 检测数据的精度. 该算法的数学表示式如下:

$$F_{m+1}(x) = F_m(x) + v \cdot \gamma_{m+1} \cdot h_{m+1}(x)$$

其中, 参数 v 是学习率, 其取值范围为 $[0, 1]$. 一般学习率越低模型的性能越好, 但是训练模型的速度也将越慢, v 通常被设置为 0.1.

算法 1. 梯度提升算法.

输入: 训练集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

输出: $F_M(x)$

1. 使用常量值初始化模型

$$F_0(x) = \arg \min \sum_{i=1}^N L(y_i, \gamma)$$

其中, γ 是常数.

2. 迭代训练模型

for $m = 1:M$

a. 计算损失函数的负梯度在当前模型的值, 将它作为残差的估计:

$$r_{im} = -\left[\frac{(L(y_i, F(x_i)))'}{F(x_i)'}\right]_{F(x)=F_{m-1}(x)}$$

其中 $i = 1, 2, \dots, n$

b. 将残差 r_{im} 拟合到回归树 $h_m(x)$ 中, 即 $h_m(x)$ 的训练集为 $\{(x_i, r_{im})\}_{i=1}^n$

c. 计算乘数 γ

$$\gamma_m = \arg \min \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

d. 更新模型

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

end

3. 输出模型 $F_M(x)$

1.2 HOG 特征

图像特征是描述图像信息的一种形式, 通常这种描述方式形式对于计算机来说更能简单有效地描述图像信息. 在计算机视觉和图像处理领域中常用的图像特征有 HOG^[28, 29]、LBP^[34]、Gabor^[35]等. 经过对这些特征的研究, 本文最终使用 HOG 特征作为本文的图像特征. HOG 特征是用来进行目标检测的一种特征描述子, 它是通过计算、统计图像局部区域内的梯度方向直方图而构成的一种局部特征. 该特征经常与 SVM 分类器

结合用来进行人脸识别。法国研究人员 Dalal 在 2005 年提出了将 HOG 特征和 SVM 分类器结合的方法用于行人检测，并获得了极好的检测效果。HOG 特征的主要思想：图像局部区域内的目标外表和形状可以通过局部梯度或者目标边缘方向的密度分布来描述。提取 HOG 特征之前首先需要对灰度图像进行 gamma 变换，该变换不但可以降低光照和阴影的影响，而且也有去噪的效果。然后将变换后图像分成小的连通区域，这些连通区域被称为细胞单元 cell；然后以 cell 为单位，再将 cell 划分为多个 block，计算和统计每个 block 中每个像素的梯度或者边缘方向的直方图。图像梯度方向范围为 0~360，HOG 中将图像梯度方向分成 9 个方向块。每个像素的梯度和方向的计算过程如下：

1) 分别计算像素 (x, y) 在纵向和横向两个方向梯度：

$$G_x(x, y) = I(x + 1, y) - I(x - 1, y)$$

$$G_y(x, y) = I(x, y + 1) - I(x, y - 1)$$

2) 计算该像素的梯度幅值 $G(x, y)$ 和梯度方向 θ ：

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}$$

$$\theta = \arctan\left(\frac{G_y(x, y)}{G_x(x, y)}\right)$$

将每个 cell 中所有 block 的方向直方图串联成一维 cell 数组，最后再将整张图像中所有 cell 数组组成该图像的 HOG 特征。

2 算法设计

通过第 2 章对 GBDT 和 HOG 特征性质的研究，本文提出将 GBDT 和 HOG 特征相结合的 H-GBDT 算法来实现人脸关键点的检测。相对于其它的关键点检测算法，本文的 H-GBDT 算法的思想和实现都较简单，且检测精度高，其具体实现步骤如下：

1. 图像预处理。将人脸从原图像中剪切下来，对人脸图像去噪处理、归一化处理等，再将所有人脸图像的关键点坐标转化为相对于该图像中心点的坐标。

2. 提取人脸图像特征。在灰度图上提取 HOG 特征，在彩色数据集上先将 RGB 颜色空间转化到 HSV 颜色空间，然后在 V 通道上提取 HOG 特征，将这两种方式提取的特征统称为 HOG 特征。其中，彩色 V 通道像素值的取值范围为 0~1。

3. 训练预测模型。每个关键点将两次调用 GBDT 做回归运算，分别是将该点坐标的纵坐标和横坐标作为 GBDT 的输出，每次调用 GBDT 的输入都是步骤 2 中提取的 HOG 特征。将每次 GBDT 回归运算训练的模型组合在一起构成 H-GBDT 的最终预测模型。本步骤中，当检测多个关键点坐标时，需要 2 倍于关键点的输出。每个关键点（图像中的一个二维坐标点）需要两次调用 GBDT，输出两个结果（每个结果对应一个坐标点取值的预测），形成两个模型，将这两个模型的预测结果组合起来，用于预测一个对应关键点的坐标。

4. 预测新输入图像的关键点坐标。在测试阶段，加载上步骤 3 中得到的 GBDT 训练模型，在该模型中将输入新图像的 HOG 特征，输出新图像的关键点坐标，然后将关键点坐标还原到原图中便是该图像关键点的预测坐标。

在步骤 3 中，本文通过不断反复地调整 GBDT 和 HOG 特征的参数，直到训练的预测模型在测试数据上的检测误差最小为止，并将该模型作为 H-GBDT 算法的最佳模型，其对应的参数作为 H-GBDT 的最佳参数。

3 实验设计

3.1 数据集

BioID^[37] 数据集中包含 23 个不同测试人的各种人脸姿态正面视图，该数据集中人脸图像都是灰度图。本文从中选取 1465 张图像作为本文的实验数据，其中随机选取 1100 张作为训练数据，剩余的 365 张作为测试数据；每张图像上包含 20 个人脸关键点坐标，如图 1(a) 所示。BioID 数据集常被用来做人脸关键点或人脸对齐的研究。

LFW 中的图像全来自互联网, 并且每张图像都是在自然场景下拍摄. LFW 数据集共包含 13000 张图像, 本文选取了 1987 张人脸图像, 并为每张图像标注了 20 个关键点, 如图 1(b). 每个关键点的位置意义与 BioID 中的相同. 本文将不考虑与两只外侧眼角相邻的点, 将剩余的 18 个点作为本文实验的关键点. 然后这 1987 张图像中随机选取了 1500 张作为训练数据, 剩下的 487 张作为实验的测试数据. 该数据集中的图像分辨率不高, 所以它经常被用来做图像检索.

LFPW^[38] 是自然场景下的数据集, 其中包含 811 张训练图像, 224 张测试数据, 每张图像包含 68 关键点坐标, 如图 1(c), 大部分研究人员选择其中的部分关键点作为研究数据, 本文选取其中 18 个关键点作为本文的测试点. 该数据集经常作为挑战性数据集用于人脸关键点检测或人脸对齐.



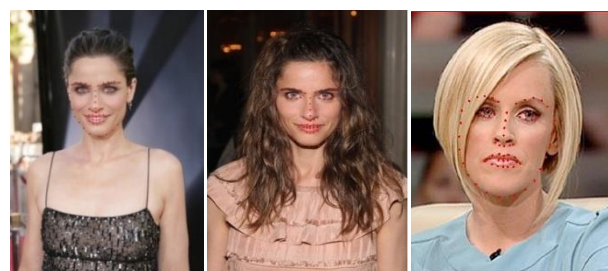
(a) BioID 数据集的关键点坐标

(a) the key point coordinates of the BioID dataset



(b) LFW 数据集的关键点坐标

(b) the key point coordinates of the LFW dataset



(c) LFPW 数据集的关键点坐标

(c) the key point coordinates of the LFPW dataset

图 1 BioID、LFW 和 LFPW 数据集的关键点坐标

Fig. 1 Key points for the BioID, LFW, and LFPW datasets

3.2 S0-RF 和 Face++

S0-RF^[30] (Structured-Output Regression Forests, 结构输出回归森林) 是 2012 年杨恒博士提出的人脸关键点检测算法, 该算法是将人脸关键点的空间结构划分为与单个关键点相关联的星型图, 决策树中的每个叶子学习了单个点的回归模型以及星形图中各点之间的相互依赖关系模型. 在测试阶段, 将对应不同模型的两个投票结果确定为一个点的位置. S0-RF 算法实现了 20 个关键点的检测, 其实验结果证明: S0-RF 算法在 BioID 数据集上有着较高的检测精度. S0-RF 算法做关键点检测时需要加载多个预测模型才能提高 S0-RF 的检测精度. 本文在灰度数据集上检测 20 个关键点, 在彩色数据集上检测 18 个关键点作为与本文 H-GBDT 的对比点.

Face++ 是在北京旷视科技有限公司旗下的新型视觉服务平台, 该平台提供了用于人脸关键点检测的模型, 该模型可以检测 83 个甚至更多关键点坐标, 其中包括眼睛、嘴巴、眉毛、鼻子等器官和人脸轮廓. 通常 Face++ 检测的高精度人脸关键点几乎能完美贴合人脸. Face++ 检测的关键点中有 11 个关键点与 H-GBDT

算法中 11 个关键点的位置意义相同, 这 11 个关键点分别是: 2 个嘴角、4 个眼角、4 个模型眉角, 一个鼻尖. 因此, 本文将在 H-GBDT 和 Face++ 的模型上对比在这 11 个关键点的检测效果.

3.3 实验结果比较

本节展示 H-GBDT 与 S0-RF、Face++ 的关键点检测误差的比较结果. H-GBDT 算法参数有 GBDT 最佳参数: 决策树棵数为 1000, 学习率为 0.03, 树的最大深度为 3; HOG 特征最佳参数是: cell 大小为 12×12 , block 为 2×2 , cell 窗体的移动步长为 2.

3.3.1 H-GBDT 与 S0-RF 比较分析

为了公平比较 H-GBDT 和 S0-RF 两种算法, 本文将使用相同的训练数据在 H-GBDT 和 S0-RF 上训练预测模型, 并用相同的测试数据在训练好的预测模型上预测图像的关键点坐标.

图 2、图 3、图 4 展示了 H-GBDT 和 S0-RF 在 BioID、LFW、LFPW 数据集上每个关键点检测误差平均值的对比结果. 这三张图表的横坐标代表人脸关键点, 纵坐标代表每个关键点对应的检测误差平均值.

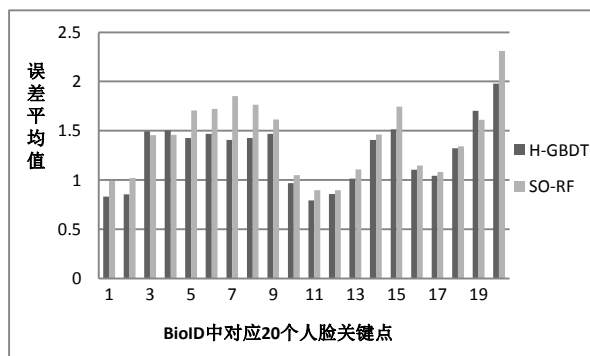


图 2 H-GBDT 与 S0-RF 在 BioID 数据集上 20 个关键点的检测误差平均值的比较结果

Fig. 2 Comparison of the mean of the detection errors of 20 key points for H-GBDT and S0-RF on the BioID dataset

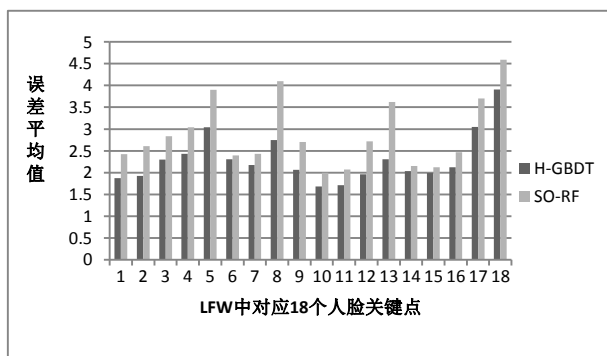


图 3 H-GBDT 与 S0-RF 在 LFW 数据集上 18 个关键点的检测误差平均值的比较结果

Fig. 3 H-GBDT and S0-RF in the LFW data set on the 18 key points of the average error detection results

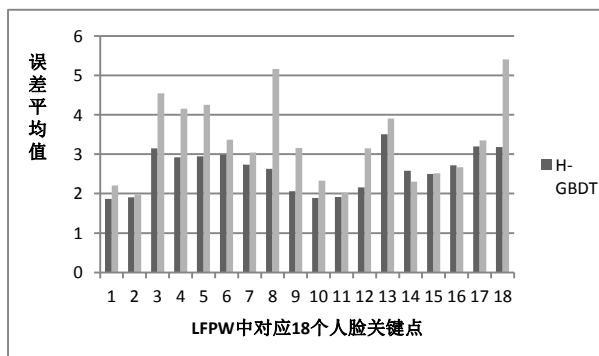


图 4 H-GBDT 与 S0-RF 在 LFPW 数据集上 18 个关键点的检测误差平均值的比较结果

Fig. 4 H-GBDT and S0-RF in the LFPW data set on the 18 key points of the detection error of the

average comparison results

图 2 表明,在 BioID 数据集上,H-GBDT 算法对每个关键点检测误差平均值在 0.8%~2%之间. 图 3 表明,在 LFW 数据集上,除了最后一个关键点在两种算法上的检测误差都比较大之外,其它点在 H-GBDT 中的检测误差平均值在 1.6%~3%之间. 图 4 表明,在 LFPW 数据集上,在 H-GBDT 中的检测误差平均值在 1.8%~3.5%之间. 图 2、图 3、图 4 三张图表的误差平均值的对比结果表明 H-GBDT 算法在单个关键点检测误差平均值低于 S0-RF 算法,特别是在 LFW 和 LFPW 数据集上,H-GBDT 算法在单个点检测效果上明显优于 S0-RF.

表 1 H-GBDT 和 S0-RF 在三中数据集上的检测误差平均值

Tab.1 H-GBDT and S0-RF in the three data sets on the detection error average

	BioID	LFW	LFPW
H-GBDT	1.27	2.31	2.6
S0-RF	1.4	2.88	3.31

表 1 显示 H-GBDT 和 S0-RF 在 BioID、LFW、LFPW 三种数据集上的检测误差平均值. 由表可知,与 S0-RF 相比,在 BioID 数据集上,H-GBDT 的检测精度提高了将近 9.2%,而在 LFW 和 LFPW 数据集上检测精度提高了将近 20%. 因此,H-GBDT 相对于 S0-RF 算法具有绝对的优势.

3.3.2 H-GBDT 与 Face++比较分析

本文从 BioID、LFW、LFPW 数据集中随机抽取别抽取 200、200、147 张图像作为验证 H-GBDT 和 Face++ 的测试数据. 表 1 显示的是这三种数据集在 H-GBDT 和 Face++ 上 11 个关键点的检测误差平均值.

表 2 H-GBDT 和 Face++在 BioID、LFW、LFPW 数据集上 11 个关键点的检测误差平均值

Tab. 2 H-GBDT and Face ++ The average of the detection errors for 11 key points on the BioID, LFW, LFPW datasets

	BioID		LFW		LFPW	
	H-GBD	Face++	H-GBDT	Face++	H-GBDT	Face++
T						
1	1.32	2.392	2.402	2.799	3.337	4.007
2	1.038	3.758	2.01	1.976	3.017	2.834
3	1.1	2.823	1.87	2.228	3.142	3.864
4	1.334	2.893	2.427	2.87	3.6	2.544
5	1.993	1.598	2.510	1.874	5.208	2.798
6	1.949	1.627	2.754	1.611	4.981	3.453
7	2.076	2.886	2.625	2.113	5.776	3.136
8	1.875	2.764	3.501	3.949	4.697	4.339
9	1.952	3.135	2.667	2.409	4.889	4.481
10	1.816	2.574	2.553	2.309	4.469	3.855
11	1.877	3.012	3.218	3.628	4.051	5.417

下面,我们分别在 BioID、LFW、LFPW 数据集的抽取数据中随机选取 25 张图像,使用 H-GBDT 和 Face++ 检测这些图像. 图 5、图 6、图 7 横坐标代表这 25 张图像,纵坐标代表每张图像关键点的检测误差平均值.

图 5 表明,在 BioID 数据集中的这 25 张图像上,H-GBDT 算法在每张图像的误差平均值基本在 2%以内,而 Face++模型在每张图像的检测误差平均值在 1.5%~3.5%之间. 因此,H-GBDT 比 Face++的检测模型在正脸的关键点检测上具有显著地优势.因此,就这 11 个关键点而言,H-GBDT 比 Face++ 提供的检测模型更适合用于正脸图像的关键点检测.

图 6、图 7 上的关键与图 5 相同. 需要说明的是,在 LFPW 的部分图像上,Face++的检测误差平均值较低,或与 H-GBDT 相当.

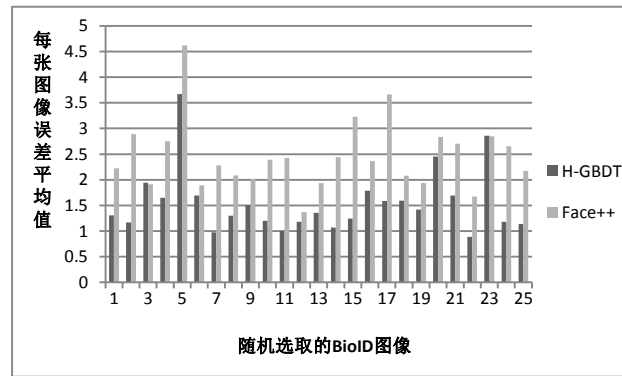


图 5 H-GBDT 和 Face++ 在 BioID 数据集中 25 张图像检测误差平均值的比较结果

Fig. 5 H-GBDT and Face++ Comparison of 25 image detection error averages in the BioID dataset

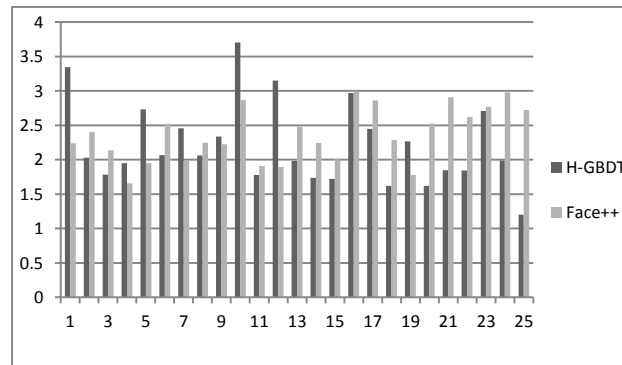


图 6 H-GBDT 和 Face++ 在 LFW 数据集中 25 张图像检测误差平均值的比较结果

Fig. 6 H-GBDT and Face++ in the LFW data set 25 image detection error comparison of the average results

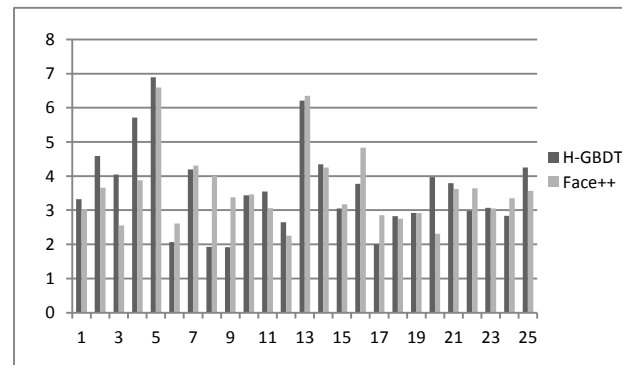


图 7 H-GBDT 和 Face++ 在 LFPW 数据集中 25 张图像检测误差平均值的比较结果

Fig. 7 H-GBDT and Face++ Comparison of 25 image detection error averages in the LFPW dataset

本研究主要对比分析不同方法在检测精度方面的性能。而关于不同方法的运行时间效率, H-GBDT 基于 Python 编程语言实现, 并调用开源软件 Scikit-learn 中的 GBDT 算法。S0-RF 基于 C/C++ 语言实现; 而 Face++ 是商业软件, 因此仅对外提供了调用接口, 但无法获知其运行平台和开发语言及环境的情况, 因此无法客观对比不同方法的时间效率。

本文开源了 H-GBDT 相关的源程序, 读者可以通过如下链接下载试用: <https://github.com/chongshengzhang/H-GBDT/>.

4 总结

本文通过对 GBDT 和 HOG 特征的研究，提出了 H-GBDT 算法来解决人脸关键点检测问题，并对该算法在 BioID、LFW、LFPW 数据集上通过不断对其调整参数训练出最佳预测模型。与 SO-RF 算法相比，在正脸数据集上，H-GBDT 的检测精度提高 9.2%，LFW 和 LFPW 数据集上，H-GBDT 的检测精度平均提高 20%左右。Face++ 提供的关键点检测模型共可以检测出 83 个甚至更多个关键点，但是有 11 个关键点的位置意义与本文算法中的关键点相对应，因此本文抽取部分测试图像作为 Face++ 的测试数据如图 5、图 6 和图 7 所示。实验结果表明，与 Face++ 模型相比，H-GBDT 在正脸图像上有着更高的检测精度，因此 H-GBDT 比 Face++ 的检测模型更适合用于正脸数据的检测。

参 考 文 献:

- [1] M. Pantic and L. Rothkrantz. Expert system for automatic analysis of facial expressions. *Image and Vision Computing Journal*, , 2000, 18(11):881–905.
- [2] Zhenyao Zhu, Ping Luo, Xiaogang Wang, Xiaoou Tang. Deep Learning Identity-Preserving Face Space, *IEEE International Conference on Computer Vision*, 2013:113-120
- [3] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. *IEEE Conference on Computer Vision and Pattern Recognition*, 2014:1701–1708
- [4] 刘明宝, 姚鸿勋, 高文. 彩色图像的实时人脸跟踪方法. *计算机学报*, 1998, 21(6):527-532
- [5] DS Chevitere, MND Santos. Real-Time Face Tracking and Recognition on IBM Neuromorphic Chip. *IEEE International Symposium on Multimedia*, 2017:667-672
- [6] J Sung, T Kanade, D Kim. Pose Robust Face Tracking by Combining Active Appearance Models and Cylinder Head Models. *International Journal of Computer Vision*, 2008, 80(2):260-274
- [7] S. Koelstra, M. Pantic, and I. Y. Patras. A dynamic texturebased approach to recognition of facial actions and their temporal models. *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, 2010, 32(11):1940–1954
- [8] Y. Yacoob and L. S. Davis. Recognizing human facial expressions from long image sequences using optical flow. *TPAMI*, 18(6):636–642, 1996.
- [9] M Pantic. *Automatic Analysis of Facial Expressions*. Springer US, 2015, 22(12):128-134.
- [10] D Ramanan, X Zhu. Face detection, pose estimation, and landmark localization in the wild. *Computer Vision and Pattern Recognition (CVPR)*. 2012, 157(10):2879-2886.
- [11] Cootes, Edwards, Taylor, etc. Active appearance models. *Pattern Analysis and Machine Intelligence*, 2001, 23(6): 681–685.
- [12] X Yu, J Huang, S Zhang, W Yan. Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model. *IEEE International Conference on Computer Vision (ICCV)*, 2013:1944-1951.
- [13] M Valstar, B Martinez, X Binefa, etc.. Facial point detection using boosted regression and graph models. *IEEE Conference on Computer Vision & Pattern Recognition*, 2010, 199(5):2729 – 2736.
- [14] XP Burgos-Artizzu, P Perona, P Dollar. Robust face landmark estimation under occlusion[C]. *IEEE International Conference on Computer Vision*, 2013:1513 – 1520.
- [15] X Cao, Y Wei, F Wen, J Sun. Face alignment by explicit shape regression[J]. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, 107(2):2887 – 2894.
- [16] Yang H., Patras I. Sieving regression forest votes for facial feature detection in the wild. *2013 IEEE International Conference on Computer Vision*. 2013, 1936–1943.
- [17] M. Seise, S. McKenna, I. Ricketts, and C. Wigderowitz. Learning active shape models for bifurcating contours. *IEEE Trans. Medical Imaging*, 2007, 26(5):666–677.
- [18] R Girshick, J Shotton, P Kohli, A Criminisi, etc.. Efficient regression of general-activity human poses from depth images. *International Conference on Computer Vision*, 2011, 6669(5):415-422.
- [19] C Sagonas, G Tzimiropoulos, S Zafeiriou, M Pantic. 300 Faces in-the-Wild Challenge: The First Facial Landmark Localization Challenge. *IEEE International Conference on Computer Vision*, 2013, 91(4):397–403

- [20] ShaoqingRen, Xudong Cao, Yichen Wei, etc.. Face Alignment at 3000 FPS via Regressing Local Binary Features. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, 25(3):1685-1692.
- [21] Z Zhang, P Luo, CL Chen, X Tang. Facial Landmark Detection by Deep Multi-task Learning. European Conference on Computer Vision, 2014, 8694:94-108
- [22] D. Cristinacce, T. Cootes. Boosted regression active shape models. Proc. British Machine Vision Conference, 2007: 880–889.
- [23] GJ Edwards, TF Cootes, CJ Taylor. Face recognition using active appearance models. European Conference on Computer Vision, 1998, 1407(6):940-943
- [24] Michel Valstar, Brais Martinez, MajaPantic, Xavier Binefa. Facial point detection using boosted regression and graph models. IEEE Conference on Computer Vision and Pattern Recognition, 2010, 119(5):2729-2736
- [25] E Zhou, H Fan, Z Cao, Y Jiang. Extensive Facial Landmark Localization with Coarse-to-Fine Convolutional Network Cascade. IEEE International Conference on Computer Vision, 2013:386-391
- [26] Gradient boosting. Wikipedia [EB/OL]. https://en.wikipedia.org/wiki/Gradient_boosting. 2017.3.15
- [27] 柯国霖. 梯度提升决策树(GBDT)并行学习算法研究. 厦门:厦门大学, 2016
- [28] Histogram of oriented gradients [EB/OL]. https://en.wikipedia.org/wiki/Histogram_of_oriented_gradients. 2017.3.3
- [29] Zhu Q, Yeh M C, Cheng K T, et al. Fast human detection using a cascade of histograms of oriented gradients. Conference on Computer Vision and Pattern Recognition, 2006, 2: 1491-1498
- [30] Heng Yang, IoannisPatras. Face Parts Localization Using Structured-output Regression Forests. Asian Conference on Computer Vision, 2012, 7725:667-679.
- [31] PL Gutierrez, S Siva. Classification and Regression Tree (CART). Springer Netherlands, 2008, 8(3):370-370
- [32] Support vector machine. [EB/OL]. https://en.wikipedia.org/wiki/Support_vector_machine.
- [33] V Rapp, T Senechal, K Bailly, L Prevost. Multiple kernel learning SVM and statistical validation for facial landmark detection. IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, 2011:265-271
- [34] Local binary patterns. [EB/OL]. https://en.wikipedia.org/wiki/Local_binary_patterns.
- [35] Gabor filter. [EB/OL]. https://en.wikipedia.org/wiki/Gabor_filter.
- [36] Shrinkage.[EB/OL].[https://en.wikipedia.org/wiki/Shrinkage_\(statistics\)](https://en.wikipedia.org/wiki/Shrinkage_(statistics))
- [37] <http://www.datatang.com/data/3045>
- [38] <http://ibug.doc.ic.ac.uk/resources/300-W/>

Background

Facial points detection is used to automatically detect the significant feature points on the face, such as the nose, mouth and other organs or face contours. With the development of computer vision, facial points detection has become an independent research topic in the computer vision field, and becomes more and more mature. The realization of facial points detection is very complex, especially in complex natural scenes. The factors such as light, facial color, pose, shelter make the detection more difficult.

There are many classical algorithms for facial points detection, such as ASM. ASM is one of the most commonly used and most advanced algorithms for face landmark detection. But ASM is very dependent on its initial shape model. This paper proposes to combine Gradient Boosting Decision Tree (GBDT) and HOG in order to further improve the accuracy of facial point detection. GBDT is a machine learning technique for regression and classification, which produces a prediction model in the form of an ensemble of weak prediction models. The iterations of GBDT are similar to those in Boosting, but GBDT can be used to optimize the loss function that can be differentiated.

This paper uses BioID, LFW, LFPW datasets as experimental data. The experimental results show that compared with SO-RF, the detection accuracy of H-GBDT is improved by 9.2% on the BioID, and 20% on LFW and LFPW. When compared with Face++, the detection accuracy of H-GBDT is improved by 37.6% on the BioID dataset, which is significant.