

像素邻域差向量协方差特征的多视角人脸检测

吕 晶,薛亚非,谷静平

(南京师范大学中北学院,江苏 丹阳 212300)

[摘要] 像素差向量特征是一阶统计量,而研究其高阶统计特性仍是一个未解问题. 针对该问题,研究了像素差向量在局部邻域中的协方差统计特性,提出了一种新颖的像素局部差向量协方差特征,显著提高了多视角人脸检测的性能. 该方法首先计算输入图像的多通道特征图,并计算特征图的像素差向量;然后在指定尺寸的邻域内计算像素差向量的协方差矩阵,获得差向量中任意两个元素间的相关信息,用于表示人脸的局部特征;最后设计了一种人脸候选区域提取方法,对该人脸检测器作进一步优化,实现了实时运行速度. 实验结果表明,本文方法优于当前其他主流方法,且可部署在低功耗的边缘计算设备上.

[关键词] 人脸检测,像素差向量特征,像素差向量协方差特征,人脸候选区域提取

[中图分类号] TP391.41 [文献标志码] A [文章编号] 1672-1292(2022)02-0073-07

Multi-view Face Detection Based on the Pixel Differential Covariance Feature

Lü Jing, Xue Yafei, Gu Jingping

(Zhongbei College, Nanjing Normal University, Danyang 212300, China)

Abstract: PDV is a discriminative first order operator, how to encode high order statistics into PDV is still unaddressed. In this paper, we investigate the covariance of the PDV in a local region, and propose a novel pixel differential covariance feature (PDCF), significantly improving the performance for multi-view face detection. Following the PDV based object detection pipeline, multiple channel maps are initially prepared before calculating pixel differential feature. Afterwards, covariance matrix of the PDV is computed to obtain the pair-wise correlation between arbitrary two elements. Finally, a face proposal mechanism is designed to generate face candidates and a PDCF based detector is further refined, achieving real-time running speed. Experimental results show that, the proposed method outperforms other state-of-the-art methods and can be deployed on the edge-computing device.

Key words: face detection, pixel differential feature, pixel differential covariance feature, face proposal

人脸检测是计算机视觉领域非常重要的一个研究分支,被广泛应用于生物特征验证、视频监控、目标跟踪和面部表情识别等领域. 姿态变化、人群遮挡、面部装饰、视角差异、光照不均衡等问题都使得人脸检测面临重大挑战. 虽已有许多研究工作来解决这些问题,但复杂条件下的人脸检测性能仍不能满足实际应用的要求.

早期的人脸检测研究将人脸检测定义为一个刚体检测问题,主要围绕设计不同的手工特征或学习方案,以获得更满意的结果. 具体来说,Haarlike^[1]、MB-LBP^[2-3]、ICF^[4]和 SURF^[5]等手工特性结合集成学习算法,可在实时运行速度下获得令人满意的结果. 在高端 GPU 设施的高性能服务器的帮助下,一些商业产品可以在图像或视频中非常准确地捕捉人脸.

本文的目标是设计一个轻量化的人脸检测系统,可以部署在移动或嵌入式设备中,并具有较强的检测性能. 由于通道特征(ICF)^[6-7]运行速度快、性能高,可用于高效的人脸特征表示. 在此基础上,像素差向量(PDV)特征^[8]的提出极大地提升了性能,但其忽略了不同像素之间的关系建模. 而 PDV 作为一种一阶

收稿日期:2021-06-24.

基金项目:江苏省高等学校自然科学基金项目(19KJB52004).

通讯作者:吕晶,讲师,研究方向:图像处理. E-mail:51818715@qq.com

运算符,如何将其高阶统计信息应用于特征编码以提升特征表达能力,仍未得到很好的解决. 针对此问题,本文研究了 PDV 在局部区域内差向量分量之间的协方差信息,提出了一种新的像素差协方差特征,可显著提高多视角人脸检测的性能.

1 相关工作

人脸检测研究的历史可以追溯到至少 50 年前,文献[1]提供了对早期和当前研究的详细介绍. 人脸检测方法可以根据输入图像类型分为灰度图像和彩色图像两大类. 在早期的研究中,大多数方法都属于第一类. 文献[5]对多尺度的图像使用一组基于神经网络的滤波器,然后将检测结果合并为最终的输出. 作为人脸检测领域的里程碑工作,V&J 检测器利用 Haar 小波^[1],对位于指定位置的相邻区域不同大小的平均像素差值进行编码. 通过 AdaBoost 算法学习 Haar 特征,并级联实现实时运行. 许多后续研究对这项工作进行了改进,如扩展的 Haarlike 或不相交的 Haarlike feature^[1]. 有学者基于分块 LBP 特征^[2]用于正面人脸检测,以较少的特征数目显著提升了检测精度. 文献[3]提出了一种新的基于分布的特征,将识别信息嵌入到特征中以提高人脸检测器的性能. 这些方法均基于图像亮度信息,对光照变化敏感,缺乏颜色信息. 另一类方法利用颜色信息,在目标检测中也被证明是非常有效的. HOG 特征首先被提出用于人体特征建模,也可用于人脸检测. 近年来,基于形变部件模型的人脸检测方法已成为人脸检测的主要方法. 该方法在输入图像分辨率较高的情况下取得了良好的效果. 最近,采用 V&J 的检测方法流程,积分通道特征也被用于人脸特征表示,在一些公共数据集上取得了较好的结果. 在行人检测领域广泛应用的多特征图滤波策略的成功推动下,类似的思想也被应用于人脸检测中,达到与当前水平相当的性能. 文献[8]对多通道映射中的像素差分关系进行建模,实现了快速准确的人脸检测.

与此同时,基于深度特征的方法^[9-12]取得了巨大的性能提升,这得益于最近通用目标识别的进展. 深度特征表示在处理其他视觉任务时也非常有效,但依赖于极为昂贵的 GPU 和较高的计算复杂度.

2 方法的提出

2.1 像素差向量(PDV)

与局部二进制模式(LBP)编码方式相似,像素差向量的目的是建立局部区域中心像素与其相邻像素之间的关系. 如图 1(a)所示,像素差向量是计算位于中心的锚像素 a_c 与周边 3×3 邻域内像素 $a_i (i=0,1,2,\dots,7)$ 的差值. 多尺度像素差向量是其一种扩展版本,为锚定像素引入了邻域半径 r ,用于表示不同尺度的像素局部统计信息. 在图 1(b)中,半径 $r=1$ 的邻域用蓝色高亮显示,半径 $r=2$ 的邻域用黄色方块标注. 对于单个半径 r ,仅对原始 PDV 可视化 8 个不同的方向. 而对于 $W \times H$ 图像,可以提取不同锚像素 $(W-r) \times (H-r) / (s \times s)$,其中 s 表示步幅. 每个半径为 r 的块共包含 $(2r+1) \times (2r+1)$ 像素.

2.2 像素差向量协方差特征(PDCF)

像素差分运算是一种一阶算子,可以有效地对局部块内的像素关系进行建模. 众所周知,高阶统计量具有更强的鉴别能力,可用于目标检测领域. 基于此,本文提出了一种像素差分协方差特征,利用 PDV 中不同分量之间的相关性,提升像素特征的表达能力. 协方差矩阵能够捕获共存的判别模式,有利于模式分类. 从直观上看,两种模式的相关性等共存模式比单一模式更具鉴别力. 具体而言,人脸具有特殊的对称结构和几何结构关系,如眼睛沿着鼻子对称,嘴总是在鼻子下面,面部区域几乎相同. 基于这些先验知识,本文设计了单半径和双半径 PDV 两种不同情形,分别如图 2(a)和(b)所示.

这两个版本的 PDCF 旨在捕获具有相同或不同半径 PDV 的不同分量之间的相关性. 如图 2(a)所示,对于 $(2r+1) \times (2r+1)$ 像素的图像块,4 个分量分别表示为 d_1, d_2, d_3 和 d_4 . PDV 的各分量在不同方向上得到了差分关系,反映了图像局部区域纹理的变化. 计算完 PDV 后,建立协方差矩阵来对 PDV 分量之间的相关性进行建模. 可以发现,不同半径的 PDV 捕获的纹理变化不同,会对人脸检测的性能产生影响.

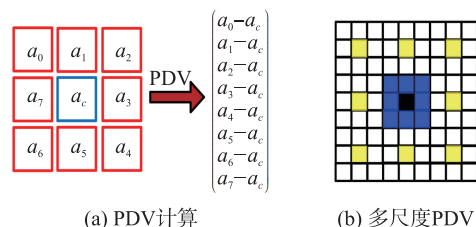


图 1 像素差向量

Fig. 1 PDV

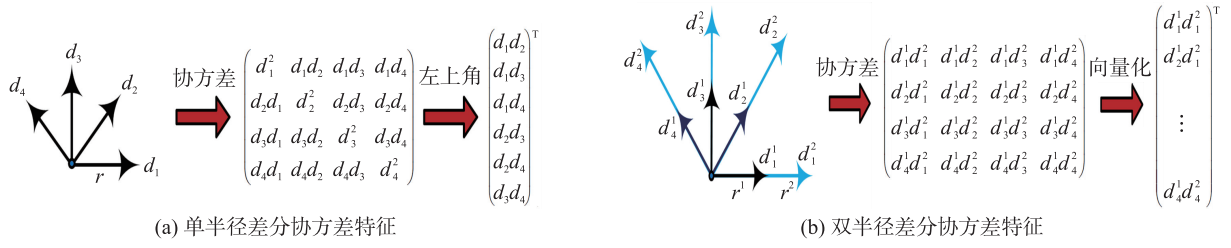


图 2 两种不同类型的 PDCF 方法

Fig. 2 Two different types of PDCF methods

2.2.1 单半径的 PDCF(PDCF-S)

与计算协方差矩阵的步骤相同,PDCF-S 建立在半径为 r 的 PDV 上. 如图 2(a) 所示,为清晰起见,本文只对 PDV 的 4 个分量进行了可视化. PDCF-S 的目标是在相同大小的局部区域内建立不同元素之间的相关性,并在相同锚点像素下探索不同元素之间的关系. 通过计算协方差矩阵得到每个锚点像素的高阶统计量,可以有效捕获纹理信息. 此外,由于对称特性,PDCF-S 特征的维度为 $(d^2-d)/2$,其中 d 通常设置为 4 或 8.

2.2.2 双半径的 PDCF(PDCF-D)

PDCF-D 建立在多尺度 PDV 基础上,其中 PDV 有两个不同的半径 r_1 和 r_2 . PDCF-D 的目的是探索不同尺寸的局部区域像素差之间的相关性. PDCF-D 不仅可以捕获每个锚点像素处不同方位的差值,还可以接收到相同方位差值之间的相关信息. PDCF-D 对 PDCF-S 具有互补效应. 如图 2(b) 所示,PDCF-D 特征的维度为 d^2 ,高于 PDCF-S 特征.

2.3 复杂性分析

假设检测窗口尺寸 $W \times H \times k$ 是像素,步长为 s ,半径 r ,多个通道地图的最终大小为 $k \times \left(\frac{N-2r}{s}\right) \times \left(\frac{M-2r}{s}\right)$,其中 W, H, k 分别表示检测窗口的宽度、高度和通道数. 当 PDV 的维数为 d 时,PDCF 的维数为 $k \times \left(\frac{N-2r}{s}\right) \times \left(\frac{M-2r}{s}\right) \times d^2$. 具体而言,当 $W=80, H=80, s=4, r=3, k=10, d=8$ 时,PDCF 的最终维数为 109 520.

2.4 检测框架

对于 PDCF 检测器的训练,可以按照流程,将 PDCF 集成到 Boosting 框架中. 与 PDV 方法不同,本文方法的 PDCF 在 PDF 层之后又增加了一个层. 为了提高人脸检测系统的速度,本文引入了级联结构,该方法在保证检测性能的前提下,能够降低计算复杂度. 整个检测流程如图 3 所示,其中人脸检测流程包括两个模块:人脸区域提取模块和 PDCF 检测模块. 前者基于 PDV 的检测器,运行速度快,召回率高. 后者为本文提出的 PDCF 检测器,利用 PDV 的高阶统计量对候选人脸进行细化,能够高效地去除误检.

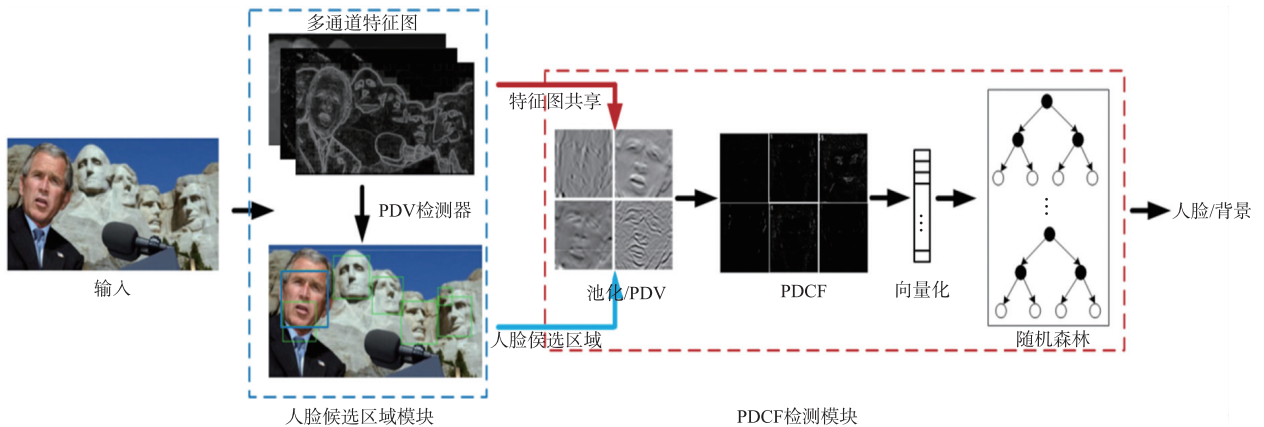


图 3 人脸检测流程

Fig. 3 The process of face detection

首先,根据输入图像 I 计算多通道特征图 F ,用于计算差分像素特征和协方差特征. 这两种类型的特征都共享 F ,从而降低了模型的计算复杂度. 其次,利用 PDV 检测器生成人脸候选区域模块,从而获得疑似人脸区域集合. 对候选人脸区域进行池化并计算 PDV 特征,在此基础上进一步计算协方差特征并将矩阵展平成特征向量. 最后利用训练好的随机森林对输入的特征向量进行处理得到最终的决策结果. 由于 PDV 与 PDCF 相比计算复杂度低,因此适用于获取高质量的候选人脸特征;而利用高阶统计信息的 PDCF 具有更高的特征判别能力,可以用于过滤候选人脸集合中的困难样本,实现高效的人脸检测.

3 实验及讨论

3.1 实验设置

本文所采用的人脸数据集的详细信息如表 1 所示,实现的多视图人脸检测器在 AFLW 数据库上进行训练. 人脸窗口大小设置为 80×80 像素. 根据不同的侧视角度共训练了 5 个人脸检测器,分别代表 $(-\infty, -60^\circ]$ 、 $(-60^\circ, -20^\circ]$ 、 $(-20^\circ, 20^\circ]$ 、 $[20^\circ, 60^\circ]$ 和 $[60^\circ, \infty)$ 的侧视角度区间. 每个视图的平均人脸如图 4 所示. 俯仰角和平面内旋转角度均限制在 $[-35^\circ, 35^\circ]$. 根据侧视角度,每个检测器的训练正样本数目分别为 3 949、8 818、19 724、8 818、3 949 张,从 PASCAL VOC2007 数据集中收集了 5 770 张不包含人脸的图像作为负样本. 本文实现的检测器利用 Boosting 算法进行训练,最终每个检测器由 2 048 个弱分类器组成.

本文采用查准率和查全率曲线及平均查全率两种度量方法对公共数据集上的不同方法进行性能评价,利用文献[4]提供的工具箱进行实验评估.

表 1 人脸检测公开数据集

Table 1 Face detection public dataset

数据集名称	数据集描述
AFLW	http://lrs.icg.tugraz.at/research/aflw/ ; 训练集: 约 2.5 万张标注的人脸图片
AFW	http://www.ics.uci.edu/~xzhua/face/AFW.zip ; 测试集: 205 张图片共包含 468 个标注人脸
PASCAL Face	PASCAL VOC2008_test, VOC2009_test, VOC2010_test, VOC2011_test; 测试集: 851 张图片包含 1 341 个人脸

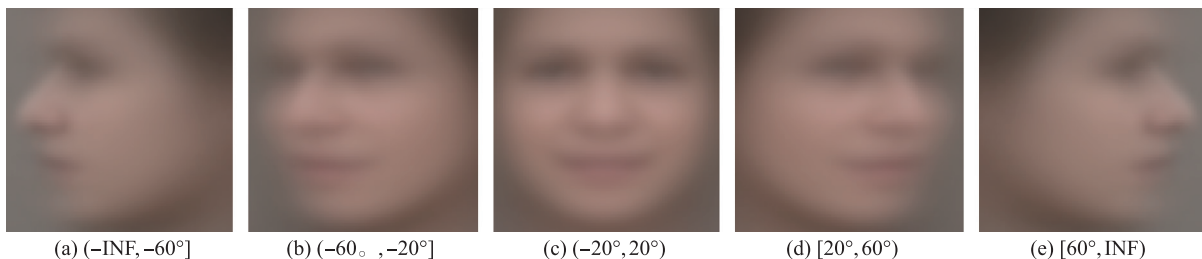


图 4 5 种不同视图的平均人脸

Fig. 4 Average face of five different views

3.2 关于 PDCF 的讨论

3.2.1 不同半径的参数选择

本节对不同半径的 PDCF 展开了详细的参数寻优实验,同时也对 ACF^[7] 和 PDV^[8] 两种经典方法进行了比较,具体结果如表 2 所示. 可以发现,合适尺寸的 PDV 半径对检测器的平均精度存在一定的影响. 当设置 $R=3$ 像素时,深度 3 的决策树具有最高的精度. 采用这种设置的 PDCF 也明显优于 ACF 和 PDV 方法,说明了本文方法具有有效的特征表示能力. 此外,PDCF-S 性能略优于 PDCF-D,因此在接下来的实验中选择 PDCF-S 作为默认特征.

3.2.2 PDCF 的分布分析

图 5 所示为检测模型参数训练中每个通道前 6 个学习所得特征的可视化表示. 并不是 PDCF 的每个特性维度都同等重要,Adaboost 算法对 d_2d_4 、 d_2d_8

表 2 PASCAL 数据集上 PDCF 的参数寻优实验

Table 2 Parameter optimization experiment of PDCF

方法	半径(R)	树深度(D)	平均精度(AP)/%	
			PDCF-S	PDCF-D
PDCF	2	2	88.19	87.23
	2	3	89.46	88.04
	2	4	87.82	87.56
	3	2	88.56	87.98
	3	3	89.75	88.34
	3	4	87.34	86.18
	4	2	86.26	86.43
	4	3	86.14	85.98
ACF	—	2	81.85	
PDV	3	2	87.62	

和 d_3d_7 这 3 个特征分量选择的频率较高,其他分量的选择频率相对较低.这说明在模型训练中,PDV 在水平和垂直方向上的不相邻元素和对称分量更受青睐.同时还可以观察到,通道 4(梯度大小)、通道 6(梯度方向为 30°)和通道 10(梯度方向为 150°)的特征选择优先级高于其他通道特征.

3.3 与主流方法的对比实验

将本文方法在两个公共数据集上与其他主流方法进行比较实验,本文方法记为 PDCF-Ours.实验采用与 PDV-Ours 和 ACF-Ours 相同的实验设置,以保证公正性.

3.3.1 AFW 数据集的对比实验

从图 6 和表 3 可以看出,本文方法在所有特征中 mAP 性能表现最优,分别比 DPM 和 HeadHunter 方法高出了 0.56% 和 0.63%.此外,PDV-Ours 和 ACF-Ours 这两个基础算法也具有较好的性能表现,比 SquaresChnFtrs-5 分别高出了 1.13% 和 0.19%.在训练数据和参数设置相同的情况下,PDCF-Ours 的性能比 PDV-Ours 和 ACF-Ours 分别高出约 1.4% 和 2.34%,从而验证了本文方法的有效性.此外,本文方法与当前先进的深度学习方法^[10,12]相比仍有一些差距,但略优于 Faceness-Net 方法^[9].由于深度学习模型依赖于额外海量的数据用于模型预训练,还需要高端的 GPU 进行并行处理,而本文方法属于经典机器学习方法,具有模型复杂度低的特点,对设备算力和功耗的要求会大大降低.

表 3 AFW 数据集上经典机器学习模型与深度学习模型的平均精度比较

Table 3 Comparison of average accuracy between classical machine learning model and deep learning model on AFW dataset

Methods	HeadHunter	DPM	PDCF	Faceness-Net ^[9]	STN ^[10]	RetinaFace ^[12]
AP/%	97.14	97.21	97.77	97.20	98.35	99.90

3.3.2 PASCAL 数据集的对比实验

PASCAL 人脸数据集的实验结果如图 7 所示.在 mAP 方面,本文方法的性能明显优于 SquareChnFtr、Structured Model 和 ACF 方法,性能提升达到了 4.87%、6.57% 和 8.59%,也分别比 DPM、HeadHunter 和 PDV 方法分别高出约 0.15%、0.81% 和 2.82%.因而,本文提出的 PDCF 方法在两个公共数据集上均优于其他基于手工设计的特征.这一结论也与 AFW 数据集上的评测结果一致.

如表 4 所示,本文方法与当前先进的深度学习方法相比尚有较大的差距,最新的 RetinaFace 方法在该数据集上精度达到了 99% 以上,而基于 FasterRCNN 模型的 STN 方法和 Faceness-net 方法比本文方法的平均精度分别高出 3.66% 和 1.67%.

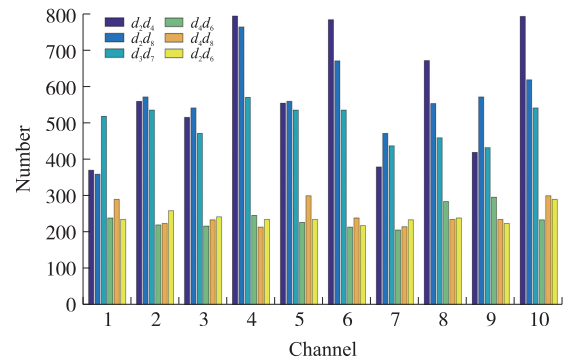


图 5 PDCF 特征分布分析

Fig. 5 Characteristic distribution analysis of PDCF

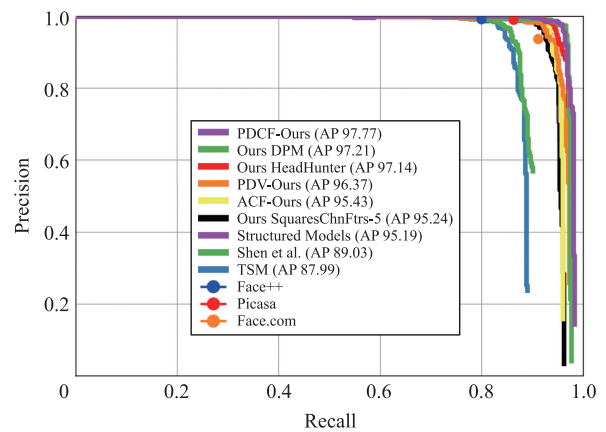


图 6 AFW 数据集上的对比实验

Fig. 6 Comparative experiment of AFW dataset

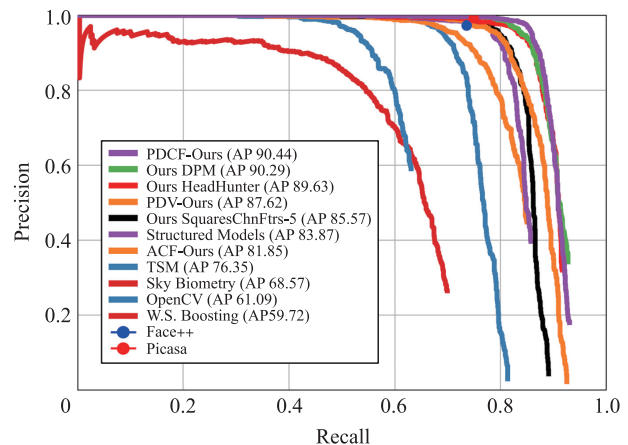


图 7 PASCAL 数据集的结果

Fig. 7 Results of PASCAL dataset

表 4 PASCAL 数据集上经典机器学习模型与深度学习模型的平均精度比较

Table 4 Comparison of average accuracy between classical machine learning model and deep learning model on PASCAL dataset						
Methods	HeadHunter	DPM	PDCF	Faceness-Net ^[9]	STN ^[10]	RetinaFace ^[12]
AP/%	89.63	90.29	90.44	92.11	94.10	99.45

3.4 检测结果比较

AFW 和 PASCAL 数据集的图片检测结果如图 8 所示. 可以发现, 本文的检测器在两个公共数据集的精度和召回率方面表现得非常好. 此外, 在手和下巴处存在一些误检, 对于漏检场景主要归因于图像中存在的遮挡和模糊问题.



图 8 AFW 和 PASCAL 数据集上的检测结果

Fig. 8 Test results on AFW and PASCAL datasets

3.5 运行时间比较

表 5 为 640×480 像素的输入图像在不同检测窗口尺寸下的运行时间比较. 3 种方法均使用相同的训练数据, 在相同的实验平台上 (DELL T7610 服务器, 双 16 核 CPU 2.6 GHz, 内存 64G) 运行. 在窗口大小为 80×80 和 40×40 像素的情况下, PDV 方法比 ACF 和 PDCF 方法有更快的运行速度. 此外, PDCF 方法具有高辨别能力, 但同时也提高了计算复杂度, 因此略慢于 ACF 方法, 但该方法可滤除图像中大部分的困难负样本. 引入 PDV 方法作为人脸候选区域提取, 可以较好地平衡检测器的精度和速度, 最终检测速度可达 20 帧/s.

表 5 640×480 图像的运行时间比较

方法	运行速度/(帧/s)		方法	运行速度/(帧/s)	
	80×80 像素	40×40 像素		80×80 像素	40×40 像素
ACF	19.61	4.31	PDCF	18.24	4.18
PDV	24.67	6.24	PDV+PDCF	20.25	5.12

4 结论

依赖于 PDV 高阶统计信息的高判别能力,本文提出了一种用于多视角人脸检测的像素差向量协方差特征,该方法有效地扩展了像素差向量特征的一阶统计特性,提升了特征表达能力,提高了多视角人脸检测的性能,实现了一种基于人脸后续区域提取的实时多视角人脸检测系统,对人脸姿态变化具有较强的鲁棒性。

此外,本文引入人脸区域候选模块,可进一步提升检测速度,在不使用 GPU 加速的情况下,处理分辨率为 640×480 的图像时,处理速度可达 20 帧/s,非常适合于部署在低功耗边缘计算设备上。

[参考文献] (References)

- [1] 曾建凡. 多角度人脸检测与识别方法研究[J]. 电子设计工程, 2017, 25(11): 41–44.
- [2] YANG B, YAN J J, LEI Z, et al. Aggregate channel features for multi-view face detection[C]//Proceedings of the 2014 International Joint Conference on Biometrics. Clearwater, USA: IEEE, 2014.
- [3] LI J G, ZHANG Y M. Learning SURF cascade for fast and accurate object detection[C]//Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE, 2013.
- [4] SHEN J F, ZUO X, LI J, et al. A novel pixel neighborhood differential statistic feature for pedestrian and face detection[J]. Pattern Recognition, 2017, 63: 127–138.
- [5] 张明浩, 杨耀权, 靳渤文. 基于图像增强技术的 SURF 特征匹配算法研究[J]. 自动化与仪表, 2019, 34(9): 98–102.
- [6] UEHARA K, SAKANASHI H, NOSATO H, et al. Object detection of satellite images using multi-channel higher-order local autocorrelation[C]//2017 IEEE International Conference on Systems, Man, and Cybernetics. Banff, Canada: IEEE, 2017.
- [7] MATHIAS M, PEDERSOLI R B M, VAN R B M. Cool, face detection without bells and whistles[C]//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014.
- [8] 耿渊哲. 基于 LBP 采样学习的人脸识别研究[J]. 计算机与数字工程, 2018, 46(2): 371–374.
- [9] YANG S, LUO P, LOY C C, et al. Faceness-Net: face detection through deep facial part responses[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(8): 1845–1859.
- [10] CHEN D, HUA G, WEN F, et al. Supervised transformer network for efficient face detection[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016.
- [11] LI J, WANG Y B, WANG C A, et al. DSFD: dual shot face detector[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019.
- [12] ZHANG S, CHI C, LEI Z, et al. RefineFace: refinement neural network for high performance face detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(11): 4008–4020.

[责任编辑: 严海琳]