

引用格式:金彦亮,葛飞扬. 用于多光谱行人检测的改进光照权重融合方法[J]. 微电子学与计算机, 2021, 38(1): 27-32. [JIN Y L, GE F Y. Improved fusion method based on ambient illumination condition for multispectral pedestrian detection [J]. Microelectronics & Computer, 2021, 38(1): 27-32.]

用于多光谱行人检测的改进光照权重融合方法

金彦亮, 葛飞扬

(上海大学 通信与信息工程学院, 上海 200444)

摘 要: 着卷积神经网络的发展, 基于多光谱图像的行人检测研究取得了巨大进步, 且应用广泛. 最近研究表明, 融合由多光谱相机(可见光和热成像相机)捕获的图像信息可以使行人在强光和弱光条件下的检测都变得鲁棒. 然而, 如何根据光照条件有效地融合图像信息仍缺乏进一步的研究. 本文提出了一种多层次特征提取方法, 旨在从不同特征层提取有用信息. 同时, 还提出一种置信度融合机制, 测量多光谱图像的光照情况. 采用一个融合函数对双流网络输出的分类结果和 RPN 输出的分类结果进行融合, 提高行人检测的性能. 通过实验将所提出的多光谱光照感知检测 R-CNN(MIAD-RCNN)与现有的多光谱行人检测器进行比较, 该方法在全天候均具有较低的漏检率和较快的速度.

关键词: 多光谱行人检测; 光照感知; 深度神经网络; 置信度融合; 多层次特征

中图分类号: TP391. 41

文献标识码: A

文章编号: 1000-7180(2021)01-0027-06

Improved fusion method based on ambient illumination condition for multispectral pedestrian detection

JIN Yan-liang, GE Fei-yang

(School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China)

Abstract: With the development of convolutional neural network, the research of pedestrian detection based on multi-spectral images has made great progress and has been widely used. Recent studies have shown that the fusion of image information captured by multispectral sensors (visible and thermal imaging cameras) makes pedestrian detection robust in both good and poor lighting conditions. However, there is still a lack of further research on how to effectively fuse image information according to lighting conditions. This paper presents a novel multi-feature extraction method aiming at extracting useful information from different feature levels. Further, this paper also proposes a novel classification score fusion mechanism. By measuring the illumination information of the input multispectral images and using a fusion function, the classification results output from the two-stream classification network and RPN are combined to improve the performance of pedestrian detection. In conclusion, compared with the latest multispectral pedestrian detectors through experiments, the proposed multispectral illumination-aware detection R-CNN (MIAD-RCNN) achieves a lower miss rate and a faster speed during both day time and night time.

Key words: multispectral pedestrian detection; illumination-aware; deep neural networks; score fusion; multi-feature

1 引言

行人检测作为通用目标检测的具体案例之一,已广泛应用于自动驾驶、视频监控和机器人感知等技术中,在过去的几十年中受到越来越多的关注.尽管借助卷积神经网络(convolutional neural networks, CNN)的快速发展在行人检测方面已取得了巨大进展,但是在光照条件差和背景混乱等恶劣情况下检测行人仍然是一个巨大挑战.

目前已经有很多行人检测的方法,其中大多数是基于可见光图像的行人探测并获得显著的性能.文献[1]用实验证明了目标检测 Faster R-CNN^[2]中的 RPN 网络确实可以作为独立的行人检测器使用,而下游分类器由于输入特征图的分辨率不足和缺乏处理难以分类的负样本的方法而导致检测结果变差.因此,作者使用 RPN 网络和提升森林(boosted forests)分类器进行检测,以达到更好的检测结果. SDS-RCNN^[3]同样也使用了与上述类似的结构.基于可见光图像的检测器在恶劣的光照条件下(例如夜间和恶劣天气)很难提取有用的特征.近几年来,也有一些研究人员研究基于多光谱图像的行人检测方法.文献[4]提出了一个用于多光谱行人检测的公开大型数据集(KAIST),使得研究和评估行人检测方法变得更加高效.该文献提出了一种基于多光谱聚合特征(ACF + T + THOG)和基于 Ada-

Boost 的下游分类器进行检测的方法.文献[5-6]在多光谱行人检测中采用了基于深度学习的检测方法,该方法在可见光图像领域已获得成功.研究了四种基于深度学习的融合结构,分别是将双流卷积神经网络在浅层、中层,深层和置信度层依次进行融合并评估行人检测的性能.实验结果表明,将多光谱图像在中间层进行特征融合可获得更鲁棒的结果.文献[7]提出一种新颖的光照感知加权机制来准确地描述环境光照状况,并将 RPN 的分类和回归结果进行融合,得到最终的检测结果.

本文提出了一种基于外界环境光照感知的多光谱行人检测方法(MIAD-RCNN).该方法以 RPN 网络和独立的下游 CNN 分类器为基础,在特征提取网络部分将多层次特征(浅层、中层和深层特征)融合输入 RPN 网络中用于行人候选框的生成,并建立分类置信度的融合机制.该机制可自适应地生成给定多光谱图像的环境光照感知权重,并通过融合函数有效地融合来自下游双流 CNN 分类器和 RPN 网络输出的置信度.

2 多光谱光照感知检测 R-CNN 方法 (MIAD-RCNN)

基于外界环境光照感知的多光谱行人检测方法的总体框架如图 1 所示.

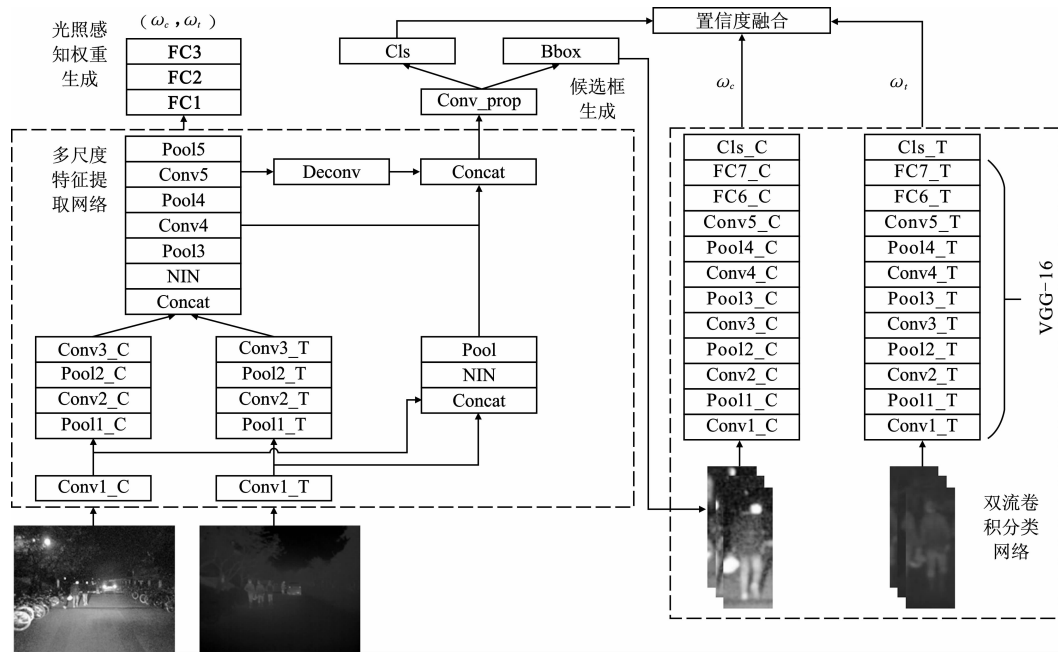


图 1 多光谱光照感知检测 R-CNN 网络的总体结构

本文所提出的改进的光照权重融合方法受到文献[7]的启发,同样引入光照感知信息来辅助网络检

测,但本文进行了以下四点改进.首先,文献[7]所提出的 IATDNN 直接将分类和回归的结果进行加权融合,但没有考虑到可见光相机和红外相机所采集的图像信息在视觉上是互补的,在 3.4 节实验表明了白天时段红外通道和可见光通道提供互补信息,而在夜晚时段可见光通道提供的信息会劣化检测结果,因此最终的行人检测结果在不同光照情况的环境下(白天和夜晚)依赖红外通道和可见光通道的信息各自所占的比重是不同的.其次,IATDNN 所提的特征全都为第五层的深层特征图,并直接用来分类与框回归,深层特征有丰富的语义特征有利于分类,而框回归需要更多的浅层细节特征,因此本文将多层次的特征融合用于分类与框回归任务,同时也有利于小目标检测.再次,IATDNN 中所提的特征为多光谱通道融合后的特征图,后续的白天黑夜的分类与回归分支也是基于这些融合后的特征图,并不能直接体现红外通道与可见光通道所贡献的最终检测结果的比重.最后,本文采用 RPN 和下游分类器互相独立的结构,能更好地挖掘困难样本.

2.1 多层次特征提取

多层次特征提取网络旨在从多光谱图像中提取有用的特征信息,提取到的特征可以用于行人的检测和候选框的回归.一对匹配好的可见光图像和红外图像将被送入到以去掉全连接层的 VGG-16 为主干网络的双流网络中去,VGG-16 包含五个卷积块(Conv1~Conv5),每个卷积块中包含 2~3 个 3×3 卷积层,然后接上一个池化层缩小特征图的尺寸,激活函数使用 ReLU 函数.在第三卷积块(Conv3)之后将双流网络通过拼接层(Concat)合并,并通过 Network-in-Network(NIN)层来减小特征维度,使得和合并前的特征的维度相同.此外,这部分还构建了一个多层次特征提取的结构,将第一卷积块(Conv1)、第四卷积块(Conv4)和第五卷积块(Conv5)的输出进行拼接.第一卷积块输出的特征图尺寸较大,通过最大池化层(Pool)进行缩小,而第五层特征图通过反卷积层(Deconv)进行放大.这些特征图通过如上的缩放操作使得和第四层的特征图尺度相同.多层次特征图含有丰富的深层语义信息和浅层细节位置信息,将提取到丰富的多层次特征图传入区域候选框生成模块中.

2.2 区域候选框与光照感知权重生成

区域候选框生成模块通过多层次的特征图来生成准确的行人候选框.该模块的网络结构沿用了经典的 Faster R-CNN 框架中的 RPN 网络,即一个内

核大小为 3×3 的卷积层,接上两个分开的 1×1 的卷积层,分别用于框回归(reg)和分类(cls),最后会输出行人候选框在原图中的位置坐标.对于预设框的选择,本文采用 $N_a = 9$ 种固定长宽比为 0.41 的锚点(anchors).当锚点和任意一个真实样本的交并比大于 0.5 时会被标为正样本,否则就被标为负样本.该部分网络的损失函数定义如下:

$$L = \lambda_1 L_{\text{cls}} + \lambda_2 L_{\text{bbox}}$$

式中, L_{cls} 为交叉熵损失函数, L_{bbox} 使用的是平滑的 L_1 损失函数, λ_i 是正则化系数,本文所有实验都取 $\lambda_i = 1, i = 1, 2$.

对于光照感知权重生成模块,本文以 VGG-16 为基础网络搭建一个二分类网络,并对原始网络结构进行一定改进.为了防止训练时神经元失活,这里激活函数采用 leaky ReLU 替代 ReLU,使得网络更好地拟合.三个全连接层的神经元个数依次设置为 1 024、256 和 2.为了减轻训练时的过拟合,在每个全连接层之后插入比率为 0.5 的 dropout 层.最终生成的权重系数 ω_c 和 ω_i 在经过 softmax 层后输出.该模块网络的损失函数为二分类交叉熵损失函数,当输入的图像为白天拍摄时,真实标签设为 1,否则设为 0. ω_c 和 $\omega_i = 1 - \omega_c$ 为该网络预测的外界光照情况的概率,也作为后面融合函数的两个权重系数, ω_c 越大说明网络预测输入的图片在光照良好的情况下采集的概率越大.

本文从多光谱图像中预测光照感知权重有两个原因.首先,红外图像与可见光图像一起可以提供互补的信息,以实现更好的光照感知权重预测;其次,在确保准确分类光照情况的前提下,光照感知分类网络与区域候选框生成网络共享特征提取网络参数,进而大大减少了网络参数量.

2.3 双流卷积分类与融合

双流卷积分类网络(包含可见光图像流和红外图像流)旨在确定由区域候选框生成网络生成的候选框分类是否为行人.本文构建了一个多光谱分离的下游分类器,并通过采用光照感知权重和一个融合函数融合最终分类的置信度得分,以提高最终行人检测器性能.

多光谱行人候选框会分别输入到两个子网络中,当输入网络时,会进行一次筛选,按照置信度得分从大到小排列,筛选 64 个行人候选框,并将这些筛选后的候选框向外填充 20% 的面积,为分类器提供更多的背景信息,更有利于区分行人目标和背景.由于预测的行人候选框大小不一,在输入到双流卷

积分网络之前,会将所有的候选框调整为 112×112 分辨率. 两个子网络以 VGG-16 作为基础网络, 输出两个置信度. 该部分网络的损失函数定义如下:

$$L = \lambda_1 L_{\text{color}} + \lambda_2 L_{\text{thermal}}$$

式中, L_{color} 和 L_{thermal} 为分类的交叉熵损失函数, λ_i 是正则化系数, 本文所有实验取 $\lambda_i = 1, i = 1, 2$.

训练拟合后要对整个检测网络进行测试, 本文设计了一个方法通过引入光照感知权重将双流卷积分类网络输出的置信度和区域候选框生成网络输出的分类置信度进行融合. 假设双流卷积分类网络会预测两个二分类置信度: $S^{\text{color}} = \{s_0^c, s_1^c\}$ 和 $S^{\text{thermal}} = \{s_0^t, s_1^t\}$. 那么, 可以通过本文提出的融合函数将这些置信度融合:

$$\begin{cases} s_1 = \frac{\alpha \cdot \omega_c}{\alpha \cdot \omega_c + \omega_t} s_1^c + \left(1 - \frac{\alpha \cdot \omega_c}{\alpha \cdot \omega_c + \omega_t}\right) s_1^t \\ s_0 = 1 - s_1 \end{cases}$$

式中, ω_c 和 ω_t 即前面预测的光照感知权重, 而参数 α 为平衡系数, 用来调节两部分的系数, 本文通过实验将系数设置为 $\alpha = 0.6$ 使得行人检测器获得更好地效果.

得到区域候选框生成网络的分类置信度: $S^{\text{RPN}} = \{s_0^{\text{RPN}}, s_1^{\text{RPN}}\}$, 以及双流卷积分类网络融合后的置信度后, 通过 softmax 函数获得最终分类置信度:

$$s_1^{\text{final}} = \frac{e^{c_{s_1^{\text{RPN}}+s_1}}}{e^{c_{s_0^{\text{RPN}}+s_0}} + e^{c_{s_1^{\text{RPN}}+s_1}}}$$

3 实验结果与分析

3.1 数据集与评价指标

本文使用的 KAIST 数据集是一个公开的多光谱行人检测数据集, 它提供了带有手工标注的大规模可见光和红外图像, 满足实验的需求. 该数据集总共有 95 328 帧多光谱图像, 其中包含 4 类标签、103 128 个标注和 1 182 个不同行人目标. “person”类标签代表独立的行人, “people”类标签代表行人群, “cyclist”类标签代表骑自行车的人, 以及“person?”类标签代表无法明确是否为行人的目标. 本文在训练时仅考虑“person”和“people”类标签, 在测试时会忽略目标像素高度小于 55 像素的目标. 测试时, “Reasonable-all”、“Reasonable-day”和“Reasonable-night”测试集分别代表全时段、白天和夜晚经合理配置的测试集.

对于评估行人检测器的性能指标, 本文采用在每张图片误报率 (false positives per image, FPPI) 范围在 $[0.01, 1]$ 内的对数平均漏检率 (miss rate,

MR), 通过 MATLAB 工具包计算测试结果的平均漏检率.

3.2 网络模型实现细节

本文提出的多光谱光照感知检测 R-CNN 模型在 Tensorflow 深度学习开发平台上实现, 实验训练机器主要配置为单 CPU Intel(R) Xeon(R) E3-1230 v3 @ 3.5GHz, 单 GPU NVIDIA GTX 1080Ti.

训练该模型总共需要三个步骤. 第一步, 使用随机梯度下降 (SGD) 训练特征提取网络和候选框生成模块. 初始学习率为 0.001, 在训练迭代完 40 000 次后学习率下降为 0.0001, 继续训练 20 000 次后终止. 对于每次训练, 会随机抽取 128 个锚点, 正负样本之比为 1:3. 第二步, 固定特征提取网络和区域候选框生成模块的参数, 以学习率为 0.0005 开始训练光照感知权重生成模块, 60 000 次后停止训练. 最后, 将前三个模块训练好的参数固定, 开始训练双流卷积分类网络, 训练参数设置和第一步相同. 对于每次训练, 会随机抽取 64 个行人候选框输入到双流卷积分类网络中, 正负样本之比为 1:3.

3.3 评估光照感知分类的准确性

光照感知权重生成模块输出的光照感知权重对最终的置信度融合机制起到至关重要的作用, 这里使用 KAIST 测试集来评估光照感知分类网络是否能够准确的分类给定多光谱图像的光照情况. KAIST 测试集已经将图像分为白天拍摄的和夜晚拍摄的, 白天拍摄的有 1 455 帧图像, 而夜晚拍摄的有 797 帧图像. 对于光照情况良好的场景 (白天), 光照感知权重 ω_c 大于 0.5, 而光照情况不好的场景 (夜晚), ω_c 小于 0.5. 此外, 本文分别使用不同层级的特征图 (多层次特征图和第五层深层特征图) 比较光照感知分类网络分类的准确性, 以确定哪个特征层级对光照的预测最有帮助, 如表 1 所示. 实验结果表明, 由于越深层的特征图越具有独特的深层语义特征, 有助于预测光照情况, 因此使用第五层特征可以更鲁棒地预测一对给定多光谱图像的光照情况.

表 1 不同层级的特征图的光照感知分类的准确性比较

	白天	夜晚
多层次特征图	96.96%	94.48%
第五层特征图	98.90%	97.99%

3.4 评估 MIAD-RCNN 性能

为了进一步评估光照信息是否可以有效地用于置信度的融合机制, 并提高多光谱行人探测器的性能, 本文和原始的以 VGG-16 为基础特征提取网络

的 RPN+下游 CNN 分类器的方法进行对比,如图 2(a,b)所示. 表 2 所示的实验结果表明白天时段的基于可见光图像的行人检测方法略好于基于红外图像的行人检测方法. 但是,由于仅仅通过使用可见光图像难以区分夜间的行人目标,因此在夜晚时段红外图像对检测行人更有帮助. 可以看出,单独的可见光图像流分类网络可能会在恶劣的照明条件下降低整个行人检测器的性能,而红外图像流在任何光照条件下都可以较为正常地工作. 利用本文所提出的置信度融合机制来自适应地生成光照感知权重来融合双流卷积分类网络和区域候选框分类网络的置信度得分,并且也探究了多层次的特征提取网络是否能提高检测效果. 所有的对比结果都在表 2 中展示,实验结果表明,本文提出的置信度融合方法可以有效地提高行人检测器的性能,并且多层次特征提取网络可以有效地融合浅层、中层和深层的神经网络特征,来生成更准确的行人候选框.

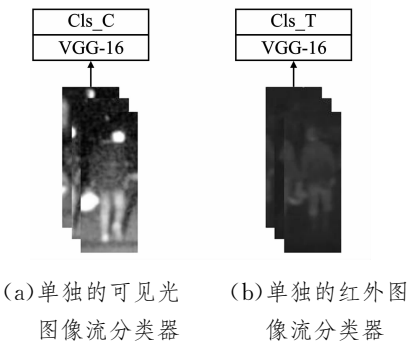


图 2 不同方法对比图

表 2 和原始 RPN+下游 CNN 分类器的漏检率对比

	Reasonable- all	Reasonable- day	Reasonable- night
可见光图像流分类	39.89%	32.75%	54.53%
红外图像流分类	32.50%	34.10%	30.24%
置信度融合	25.67%	26.75%	24.84%
置信度融合+多层次特征提取	24.92%	25.21%	24.17%

本文所提出的 MIAD-RCNN 行人检测方法和目前多光谱行人检测精度较高的 5 种方法进行对比: Halfway Fusion^[6]、ACF+T+THOG^[4]、Fusion RPN+BDT^[8]、FRPN-Sum+TSS^[9] 和 IATDNN+IAMSS^[7], 如表 3 所示, 相比其他顶尖的多光谱行人检测方法, 本文所提出的方法无论实在白天场景还是在夜间场景下都有着更低的漏检率. 此外, 本文提出的检测方法可以使用 GPU 并行加速计算, 其

检测速度也超过了表 4 中所示的现有的检测方法, 在使用单个 NVIDIA Geforce 1080Ti 显卡的平均检测每张多光谱图像速度为 0.218 秒.

表 3 和原始 RPN+下游 CNN 分类器的漏检率对比

	Reasonable- all	Reasonable- day	Reasonable- night
ACF+T+THOG	54.80%	51.97%	61.19%
Halfway Fusion	37.19%	37.12%	35.33%
Fusion RPN+BDT	29.68%	30.51%	27.62%
FRPN-Sum+TSS	26.67%	26.75%	25.24%
IATDNN+IAMSS	26.37%	27.29%	24.41%
MIAD-RCNN	24.92%	25.21%	24.17%

表 4 不同检测方法的的速度对比

方法	Halfway Fusion	Fusion RPN+BDT	FRPN- Sum+TSS	IATDNN+ IAMSS	MIAD- RCNN
时间	0.398 s	0.780 s	0.230 s	0.226s	0.218 s

4 结束语

为提高全时段的行人检测的准确性和鲁棒性, 本文提出了一种用于多光谱行人检测基于光照条件的融合方法, 并构造了一个强大的多光谱光照感知检测 R-CNN 网络模型 (MIAD-RCNN). 该网络可以自发地根据外界环境光照条件有效地融合最终的分类置信度, 以提高检测性能. 在 KAIST 数据集上的实验结果表明, 该方法在全时段均获得较好的漏检率, 白天和夜晚场景均获得了明显的提升, 并且以平均每张多光谱图像 0.218 秒获得了更快的检测速度. 虽然本方法可以通过 GPU 进行并行加速运算, 但是网络模型仍旧非常复杂, 难以部署到计算资源受限的嵌入式移动设备上运行. 因此, 为满足自动驾驶实时性要求, 下一步计划研究模型的压缩与蒸馏, 在不影响检测效果的情况下, 尽量提高检测速度及降低模型的参数量, 增强其实用性.

参考文献:

[1] ZHANG L L, LIN L, LIANG X D, et al. Is faster R-CNN doing well for pedestrian detection? [C]//14th European Conference on Computer Vision. Amsterdam: Springer, 2016: 443-457. DOI: 10.1007/978-3-319-46475-6_28.

[2] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Proceedings of the 28th International Conference on Neural Information Processing

- Systems. Cambridge: ACM, 2015: 91-99.
- [3] BRAZIL G, YIN X, LIU X M. Illuminating pedestrians via simultaneous detection and segmentation[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 4950-4959. DOI: 10.1109/ICCV.2017.530.
- [4] HWANG S, PARK J, KIM N, et al. Multispectral pedestrian detection: benchmark dataset and baseline [C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 1037-1045. DOI: 10.1109/CVPR.2015.7298706.
- [5] WAGNER J, FISCHER V, HERMAN M, et al. Multispectral pedestrian detection using deep fusion convolutional neural networks [C]//24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges: ES-ANN, 2016.
- [6] LIU J J, ZHANG S T, WANG S, et al. Multispectral deep neural networks for pedestrian detection[M]//WILSON R C, HANCOCK E R, SMITH W A P. Proceedings of the British Machine Vision Conference. British: BMVA Press, 2016. DOI: 10.5244/C.30.73.
- [7] GUAN D Y, CAO Y P, YANG J X, et al. Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection[J]. Information Fusion, 2019, 50: 148-157. DOI: 10.1016/j.inffus.2018.11.017.
- [8] KONIG D, ADAM M, JARVERS C, et al. Fully convolutional region proposal networks for multispectral person detection[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu: IEEE, 2017: 49-56. DOI: 10.1109/CVPRW.2017.36.
- [9] GUAN D Y, CAO Y P, YANG J X, et al. Exploiting fusion architectures for multispectral pedestrian detection and segmentation[J]. Applied Optics, 2018, 57(18): D108-D116.

作者简介:

金彦亮 男,(1974-),博士,副教授.研究方向为无线传感器网络、无线多媒体传感器网络、无线宽带接入和信号处理.

葛飞扬(通讯作者) 男,(1994-),硕士研究生.研究方向为目标检测、行人检测. E-mail: gfy0921@qq.com.

(上接第 26 页)

- [7] HUANG K H, ABRAHAM J A. Algorithm-based fault tolerance for matrix operations[J]. IEEE Transactions on Computers, 1984, C-33(6): 518-528. DOI: 10.1109/TC.1984.1676475.
- [8] CHEN J Y, LIANG X, CHEN Z Z. Online algorithm-based fault tolerance for choleskydecomposition on heterogeneous systems with GPUs[C]//Proceedings of 2016 IEEE International Parallel and Distributed Processing Symposium. Chicago, IL, USA: IEEE, 2016: 993-1002. DOI: 10.1109/IPDPS.2016.81.
- [9] BRAUN C, HALDER S, WUNDERLICH H. A-ABFT: autonomous algorithm-based fault tolerance for matrix multiplications on graphics processing units [C]//Proceedings of the 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. Atlanta, GA, USA: IEEE, 2014: 443-454. DOI: 10.1109/DSN.2014.48.
- [10] WHITEHEAD N, FIT-FLOREA A. Precision and performance: floating point and IEEE 754 compliance for NVIDIA GPUs[R]. Santa Clara, USA: NVIDIA Corporation, 2011.

作者简介:

包冲 男,(1979-),硕士,高级工程师.研究方向为高性能计算及计算机可靠性. E-mail:17148897@qq.com.

张善从 男,(1977-),博士,二级研究员.研究方向为高可靠计算机应用、航天电子学工程.