

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2017 年 11 月 9 日 (09.11.2017)



(10) 国际公布号
WO 2017/190574 A1

- (51) 国际专利分类号:
G06K 9/00 (2006.01)
- (21) 国际申请号: PCT/CN2017/079519
- (22) 国际申请日: 2017 年 4 月 6 日 (06.04.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201610287163.1 2016 年 5 月 4 日 (04.05.2016) CN
- (71) 申请人: 北京大学深圳研究生院(PEKING UNIVERSITY SHENZHEN GRADUATE SCHOOL) [CN/CN]; 中国广东省深圳市南山区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。
- (72) 发明人: 王文敏(WANG, Wenmin); 中国广东省深圳市南山区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。黄颖(HUANG, Ying); 中国广东省深圳市南山区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。

王荣刚(WANG, Ronggang); 中国广东省深圳市南山区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。李革(LI, Ge); 中国广东省深圳市南山区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。董胜富(DONG, Shengfu); 中国广东省深圳市南山区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。王振宇(WANG, Zhenyu); 中国广东省深圳市南山区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。李英(LI, Ying); 中国广东省深圳市南山区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。赵辉(ZHAO, Hui); 中国广东省深圳市南山区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。高文(GAO, Wen); 中国广东省深圳市南山区西丽镇丽水路深圳大学城北大校区, Guangdong 518055 (CN)。

(74) 代理人: 北京万象新悦知识产权代理事务所(普通合伙)(BEIJING WANXIANGXINYUE INTELLECTUAL PROPERTY OFFICE); 中国北京

(54) Title: FAST PEDESTRIAN DETECTION METHOD BASED ON AGGREGATION CHANNEL FEATURES

(54) 发明名称: 一种基于聚合通道特征的快速行人检测方法

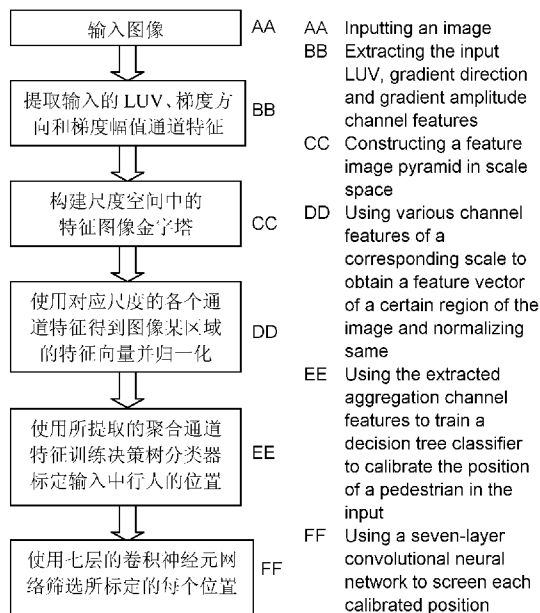


图 1

(57) Abstract: Disclosed is a fast pedestrian detection method based on aggregation channel features, which comprises an early-stage position calibration process and a later-stage position screening process. During the early-stage position calibration process, a plurality of pieces of channel feature information in an input video or image are aggregated; an image pyramid of the input image or video in scale space is built; for each scale, an image of each channel feature is calculated; features serving as a pedestrian existing judgement basis are recognized; and the pedestrian position is initially extracted. During the later-stage position screening process, a convolutional neural network classifier is used for further screening each position calibrated in the early stage; and pedestrians appearing in the image or the video are further detected to obtain a detection result. By using the technical solution of the present invention, where a training data amount is large, the classifier can automatically select features with a good recognition ability to serve as a pedestrian judgement basis. The method has good robustness, and also improves the pedestrian detection precision.

市海淀区颐和园路1号北京资源燕园宾馆
六层1625号, Beijing 100080 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 本发明公布了一种基于聚合通道特征的快速行人检测方法, 包括前期位置标定过程和后期位置筛选过程: 前期位置标定过程聚合输入视频或图像中的多个通道特征信息, 构建输入图像或视频在尺度空间中的图像金字塔, 对每个尺度计算各个通道特征的图像, 识别作为存在行人判别依据的特征, 初步提取行人位置; 后期位置筛选过程使用卷积神经网络分类器对前期标定的每个位置做进一步筛选, 进一步检测图片或视频中出现的行人, 得到检测结果。采用本发明技术方案, 在训练数据量大时, 分类器可自动选择出具有较好识别能力的特征, 作为行人判别依据, 方法的鲁棒性高; 也提高了行人检测的精度。

一种基于聚合通道特征的快速行人检测方法

技术领域

本发明属于计算机视觉技术领域，涉及行人检测方法，尤其涉及一种基于聚合通道特征的快速行人检测方法，针对输入图像或视频，标定其中存在的行人的准确位置。

背景技术

在计算机视觉的研究中，对象检测技术有着比较长的研究历史，其中行人检测问题作为其子问题而存在，是指通过一系列自动化的处理流程完成输入图像或者视频中所存在行人位置标定的过程，其中位置标定由一个矩形框来表示，需要给出该矩形框的左上角和右下角（或矩形框的长、宽）的坐标信息。由于输入图像中行人的姿态、光照的强弱、是否有遮挡、场景的变换以及拍摄角度的不同，都会影响最终的检测效果。因此，使用一种鲁棒的方法检测图像（或视频）输入中存在的行人显得非常重要。

通过获得图像或视频中所包含行人的准确定位，机器人将可以感知周围的状况；无人机可以对特定行人进行跟踪拍摄，或者完成自动送货；无人驾驶则可以根据输入中距离行人的远近提前预判并进行有效避让，提高无人驾驶的安全性；在移动支付领域也可以通过生物特征来完成身份验证；对于多媒体网站则可以实现信息自动过滤，节省大量的人力物力，同时进化网络环境；特别是在公共安全领域，近年来随着恐怖事件的频繁发生而加大了人们对于智能安全系统建设的需求，而行人对象识别与检测技术就可以为其提供基础性的支持，使得对监控视频中的语义信息做进一步筛选成为可能，并在公共安全遭受损害前提前预警，在危机事件发生后对于危险分子快速定位，防止事件蔓延，因此该技术应用范围将非常广阔。

在行人检测的研究历程中出现了很多优秀算法。使用梯度直方图（Histogram of Gradient, HOG）特征加支持向量机（Support Vector Machine, SVM）的滑动窗口算法能够较好的克服光照的影响，因此应用较为广泛，但其在处理遮挡问题时表现不佳。Felzenszwalb 和 Girshick 等人通过研究发现，可以使用局部可形变模型（Deformable Part Model, DPM）来描述待检测行人各个部位以及整体之间的关系，从而可以有效克服遮挡问题，但是该方法计算过程较为复杂，运行速度较慢。

发明内容

为了克服上述现有技术的不足，本发明提供一种基于聚合通道特征的快速行人检测方法，

前期标定通过聚合输入视频和/或图像中的多个通道特征信息，选择出具有较好识别能力的特征作为存在行人的判别依据，增加了方法的鲁棒性，再使用卷积神经网络分类器对前期标定的每个位置做进一步筛选，提高对输入视频和图像中行人进行检测的精度，为智能监控等领域提供基础性技术支持。

本发明的原理是：针对输入图像或视频，行人检测即需要标定其中存在的行人的准确位置，本发明首先聚合输入视频和/或图像中的多个通道特征信息，获取聚合通道特征。通道特征是指对于原始输入图像的某种输出响应，用公式 $C = \Omega(I)$ 来表示。对于灰度图像而言，可以认为其本身即为最朴素的通道，而彩色图像包含了 RGB 三个颜色通道，同时也可以计算其对应的灰度通道或者 CIELUV（色彩空间中的亮度（L）和色度（U,V）通道，也称 LUV 通道）。在此基础上，一阶通道特征可以定义为对某个输入通道在一定区域的像素之和，用 $f(C)$ 表示。而高阶通道特征则定义为使用任意多幅一阶通道特征相互叠加后计算出的结果。利用不同通道的特征进行组合，所得到的结果向量称为聚合通道特征，可作为输入图像某一个区域的特征表示。然后，对于输入图像或者视频，构建其在尺度空间中的图像金字塔，对每个尺度计算其对应的各个通道特征的图像。但是，如果每个尺度的特征图像都要重复计算的话，将会非常耗时，因此，在计算过程中使用快速算法来构建图像特征金字塔，本发明中采用原始尺度的特征图像通过尺度缩放得到该尺度上图像的各个通道特征图像。之后，通过训练决策树模型（使用迭代变换（Iterative Dichotomiser 3, ID3）算法），从所有特征通道中自动选择出具有较好鉴别能力的特征，作为行人的分类依据，并通过该分类器提取各个尺度中行人可能出现的位置，在原始输入的尺度空间中进行标注，完成行人位置的初步提取。最后，本发明通过采用卷积神经网络作为分类器模型，筛选所标注的每个行人位置，检测图片或视频中出现的行人。

本发明提供的技术方案是：

一种基于聚合通道特征的快速行人检测方法，包括前期位置标定过程和后期位置筛选过程；聚合输入视频或图像中的多个通道特征信息，构建输入图像或视频在尺度空间中的图像金字塔，对每个尺度计算各个通道特征的图像，识别作为存在行人判别依据的特征，初步提取行人位置；后期位置筛选过程使用卷积神经网络分类器对前期标定的每个位置做进一步筛选，提高对输入视频和图像中行人进行检测的精度；具体包括如下步骤：

- 1) 对每幅输入图像或视频，计算得到所述图像或视频的多通道的特征图像；
- 2) 构建特征图像金字塔：利用步骤 1) 计算出的特征图像，得到所述图像或视频在尺度

空间内的特征图像金字塔；

3) 根据步骤 2) 得到的特征图像金字塔，对所述图像或视频中待识别的图像目标区域，使用所述区域内所有通道的特征图像，将所述区域表示为统一维度的特征向量，作为聚合通道特征；

4) 使用决策树模型进行训练，从步骤 3 所述特征向量中自动选择出行人鉴别特征作为判别标准，训练得到决策树分类器，通过滑动窗口扫描对不同尺度大小的行人进行位置标定，标定位置表示为在窗口所在区域添加标注；

5) 使用卷积神经网络模型对步骤 4) 得到的每个位置进行筛选，得到最后的检测结果。

针对上述基于聚合通道特征的快速行人检测方法，进一步地，所述多通道包括 LUV 通道、梯度方向通道和梯度幅值通道。

针对上述基于聚合通道特征的快速行人检测方法，进一步地，步骤 2) 具体通过对所述输入图像进行上下采样的方法，构建得到所述输入图像在尺度空间内的特征图像金字塔。

针对上述基于聚合通道特征的快速行人检测方法，进一步地，步骤 3) 中，所述特征图像金字塔中包含的尺度为五个；所述通道为八个；所述聚合通道特征具体通过如下操作得到：

设定输入图像为 I ；待识别图像区域为 R ，在 I 的特征图像金字塔中包含五个尺度，记为 $L_1, L_2, \dots, L_5$ ；各个尺度上八个特征通道的图像记为 $P_{xy}(x = 1, \dots, 5; y = 1, \dots, 8)$ ，每个尺度上待识别图像区域 R 对应的区域记为 $r_1, r_2, \dots, r_5$ ；设 $i=1..5$ ，对尺度 L_i 下的区域 r_i ，将所述各通道图像 $P_{iy}(y = 1, \dots, 8)$ 上的相应区域列在一起，形成区域 r_i 的聚合通道特征。

针对上述基于聚合通道特征的快速行人检测方法，进一步地，步骤 4) 所述训练得到决策树分类器具体为：在使用决策树模型进行训练过程中，计算特征向量中每个维度的特征所带来的信息增益，对信息增益进行排序，依次选取出当前信息增益最大的特征，代表鉴别能力最强的行人鉴别特征，将所述行人鉴别特征作为所述决策树的结点，得到决策树分类器。

在本发明实施例中，所述决策树模型具体为迭代变换算法，输入为训练集 D 、特征集 X 和阈值 ε ，输出为决策树 T ；所述迭代变换算法执行如下操作得到决策树分类器：

41) 通过式 1 计算得到特征向量中每个维度的特征 X_i 为训练集 D 所带来的信息增益

$g(D, X_i)$:

$$g(D, X_i) = H(D) - H(D | X_i) \quad (\text{式 } 1)$$

式 1 中, $H(D) = -\sum_{k=1}^n p_k \log p_k$ 为训练数据集 D 本身的熵, 其中 p_k 为训练数据集 D 中第 k

个类别出现的概率, k 只有两个取值, $k = 1$ 代表行人和 $k = 2$ 代表非行人;

$H(D | X_i) = \sum_{j=1}^m p_j H(D | X_i = x_j)$ 为特征 X_i 给定条件下训练集 D 的条件熵;

42) 对步骤 41) 得到的不同特征向量所得信息增益进行排序, 选取出信息增益最大的特征向量作为行人鉴别特征, 将所述行人鉴别特征作为所述决策树的结点, 得到决策树分类器。

针对上述基于聚合通道特征的快速行人检测方法, 进一步地, 步骤 4) 所述标注形式 (x, y, w, h) , 其中 (x, y) 为所述区域左上角的坐标, (w, h) 为所述区域的宽度和高度。

针对上述基于聚合通道特征的快速行人检测方法, 进一步地, 步骤 5) 所述使用卷积神经网络模型对步骤 4) 得到的每个位置进行筛选, 所述卷积神经网络分类模型采用 AlexNet 网络结构, 包括五个卷基层和两个全连接层; 通过对步骤 4) 得到的图像中所有添加标注的区域, 使用所述 AlexNet 网络进行再次分类, 标注出行人, 完成行人检测。

更进一步地, 所述再次分类包括如下步骤:

51) 通过所述 AlexNet 网络的五个卷基层将要分类的每个区域 r 表示成一个 4096 维的向量, 作为区域 r 的特征向量;

52) 通过一个分类层 Softmax 层, 输入步骤 51) 所述 4096 维的特征向量, 输出一个 2 维向量; 所述输出向量中的第一维表示所述区域 r 识别为行人的概率, 所述输出向量中的第二维表示所述区域 r 识别不为行人的概率;

53) 通过设定概率阈值, 确定所述区域 r 是否为行人, 完成行人检测。

在本发明实施例中, 步骤 53) 所述设定值为 0.9。

与现有技术相比, 本发明的有益效果是:

一种基于聚合通道特征的快速行人检测方法, 包括前期位置标定过程和后期位置筛选过程; 聚合输入视频或图像中的多个通道特征信息, 构建输入图像或视频在尺度空间中的图像

金字塔，对每个尺度计算各个通道特征的图像，识别作为存在行人判别依据的特征，初步提取行人位置；后期位置筛选过程使用卷积神经网络分类器对前期标定的每个位置做进一步筛选，提高对输入视频和图像中行人进行检测的精度。与现有技术相比，本发明技术方案具有如下特点：

（一）本发明利用了输入中的多个通道特征的信息，因此在拥有很多训练数据时，分类器可以自动选择出具有较好识别能力的特征，将其作为判别某个位置是否可能存在行人的判别依据，增加了方法的鲁棒性。

（二）本发明在得到最终结果前，使用卷积神经网络分类器对前期标定的每个位置做进一步筛选，提高了检测精度。

附图说明

图 1 是本发明提供的基于聚合通道特征的快速行人检测方法的流程框图。

图 2 是本发明实施例通过滑动窗口扫描对不同尺度大小的行人进行位置标定示意图。

图 3 是本发明实施例使用到的聚合通道特征；

其中，(a) 为输入的图像；(b) 为 L 颜色通道的图像；(c) 为 U 颜色通道的图像；(d) 为 V 颜色通道的图像；(e) 为梯度幅值通道的图像；(f) 为第一个梯度方向通道的图像；(g) 为第二个梯度方向通道的图像；(h) 为第三个梯度方向通道的图像；(i) 为第四个梯度方向通道的图像。

图 4 是本发明实施例根据多通道特征图像在尺度空间构建得到特征图像金字塔的示意图；

其中，(a) 是原始尺度梯度幅值通道的图像；(b) 是 (a) 缩放 75% 后对应尺度上梯度幅值通道的图像；(c) 是 (a) 缩放 50% 或对应尺度上梯度幅值通道的图像；(d) 是 (a) 缩放 25% 后对应尺度上梯度幅值通道的图像。

图 5 是本发明实施例使用多通道特征图像生成图像区域块特征向量的示意图；

其中，(a) 是输入图像；(b) - (i) 对应图 3 说明的各个通道，包括 LUV 通道、梯度方向通道和梯度幅值通道，同时用矩形框标注出其中待识别区域；(j) 是图像中待识别区域的聚合通道特征。

图 6 是本发明实施例使用卷积神经网络对前期提取的行人位置进行分类过程的示意图。

图 7 是本发明实施例在 INRIA 行人数据集上采用本发明方法与其它现有方法的性能指标对比图；

其中，横坐标为平均误检数，单位为个/帧；纵坐标为漏检率，单位为%；OURS 为本发

明提供方法；VJ 为 OpenCV 中使用的方法；HOG 为梯度直方图（Histogram of Gradient）；ACF 为聚合通道特征（Aggregate Channel Features）。

具体实施方式

下面结合附图，通过实施例进一步描述本发明，但不以任何方式限制本发明的范围。

本发明提供一种基于聚合通道特征的快速行人检测方法，包括前期位置标定过程和后期位置筛选过程；图 1 是本发明实施例中的基于聚合通道特征的快速行人检测方法的流程框图。前期标定通过聚合输入视频和/或图像中的多个通道特征信息；通过构建输入图像或者视频在尺度空间中的图像金字塔，对每个尺度计算其对应的各个通道特征的图像；选择出具有较好识别能力的特征作为存在行人的判别依据，增加了方法的鲁棒性，初步提取行人位置；后期位置筛选过程使用卷积神经网络分类器对前期标定的每个位置做进一步筛选，提高对输入视频和图像中行人进行检测的精度，为智能监控等领域提供基础性技术支持。其中，图像尺度空间金字塔是对于一幅图像进行缩放后得到的一组图像，他们是对不同距离和不同清晰度条件下输入图像的映射。卷积神经网络是计算机视觉中所使用的一种分类器模型，它的作用是将不同类别的图片区分开来。

基于多通道聚合特征完成行人检测过程，需要对不同尺度大小的行人完成位置标定，整个过程称为滑动窗口扫描，图 2 是本发明实施例通过滑动窗口扫描对不同尺度大小的行人进行位置标定示意图。滑动窗口扫描过程中，需要对图像中不同大小的区域块进行判定，通过不断调整区域块大小并穷尽搜索整幅图像，得到其中行人。本发明提供一种基于聚合通道特征的快速行人检测方法，在具体实施中，主要包括如下步骤（见图 1）：

1) 对每幅输入图像，计算其各个通道的特征图像；

本实施例中，对每幅输入图像计算了图像的 LUV 通道、梯度方向通道和梯度幅值通道的特征；图 3 表示了本实施使用到的聚合通道特征，其中，(a) 为输入的图像；(b) 为 L 颜色通道的图像；(c) 为 U 颜色通道的图像；(d) 为 V 颜色通道的图像；(e) 为梯度幅值通道的图像；(f) 为第一个梯度方向通道的图像；(g) 为第二个梯度方向通道的图像；(h) 为第三个梯度方向通道的图像；(i) 为第四个梯度方向通道的图像。

2) 构建特征图像金字塔：利用上一步计算出的特征图像，得到其在尺度空间内的特征图像金字塔；

此过程直接通过将原始特征图像上下采样的形式，构建尺度空间内特征图像金字塔，该特征图像金字塔作为下一步骤的输入；图 4 是本发明实施例根据多通道特征图像在尺度空间

构建得到特征图像金字塔的示意图；其中，(a) 是原始尺度梯度幅值通道的图像；(b) 是 (a) 缩放 75% 后对应尺度上梯度幅值通道的图像；(c) 是 (a) 缩放 50% 或对应尺度上梯度幅值通道的图像；(d) 是 (a) 缩放 25% 后对应尺度上梯度幅值通道的图像。

3) 对待识别的图像目标区域，使用该区域内所有通道的特征图像将该区域表示为统一维度的特征向量，作为聚合通道特征；

设定输入图像为 I ，待识别图像区域为 R ，在 I 的特征图像金字塔中包含五个尺度 $(L_1, L_2, \dots, L_5)$ 上十个特征通道的图像 $P_{xy}(x = 1, \dots, 5; y = 1, \dots, 10)$ ，每个尺度上待识别图像区域 R 对应的区域为 $r_1, r_2, \dots, r_5$ 。对尺度 L_i 下的区域 r_i ，取各通道图像 $P_{iy}(y = 1, \dots, 10)$ 上的对应区域列在一起，形成 r_i 的聚合通道特征。图 5 示意了本实施例使用各通道特征图像生成图像区域块特征向量；其中，(a) 是输入图像；(b) - (i) 对应图 3 说明的各个通道，包括 LUV 通道、梯度方向通道（图中 (f) - (i) 显示了 6 个梯度方向通道中的 4 个作为示例）和梯度幅值通道，同时用矩形框标注出其中待识别区域；(j) 是图像中待识别区域的聚合通道特征。本实施例中，对于不同大小的图像区域，将其归一化到 128x64 维度的向量上进行表示。

4) 使用决策树模型（本发明实施例采用迭代变换 (Iterative Dichotomiser 3, ID3) 算法）自动选择出其中具有鉴别能力的特征（行人鉴别特征）作为判别标准，训练得到决策树分类器，通过滑动窗口扫描对不同尺度大小的行人进行位置标定，获得待测试图像中行人可能出现的位置；

在训练过程中，ID3 算法首先计算特征向量中每个维度的特征 X_i 为训练集 D 所带来的信息增益 $g(D, X_i)$ （亦称为互信息），通过如下公式（式 1）计算得到：

$$g(D, X_i) = H(D) - H(D | X_i) \quad (\text{式 1})$$

其中， $H(D) = -\sum_{k=1}^n p_k \log p_k$ 为训练数据集 D 本身的熵， p_k 为训练数据集 D 中第 k 个类别出现的概率，此处 k 只有两个取值， $k = 1$ （行人）和 $k = 2$ （非行人）；

$H(D | X_i) = \sum_{j=1}^m p_j H(D | X_i = x_j)$ 为特征 X_i 给定条件下训练集 D 的条件熵；

之后对不同特征所得信息增益进行排序，选取出当前鉴别能力最强的特征（信息增益最

大)作为决策树的结点,得到决策树分类器。ID3 算法的输入为训练集 D 、特征集 X 和阈值 ε ,输出为决策树 T ,具体计算过程如下:

(a) 若 D 中所有实例属于同一个类 C_k ,则 T 为单结点树,并将 C_k 作为该结点的类标记,返回 T ;

(b) 若 $X = \Phi$,则 T 为单结点树,将 D 中实例数最多的类 C_k 作为该结点的类标记,返回 T ;

(c) 否则,当 $X \neq \Phi$ 时,计算 X 中各个特征对 D 的信息增益,选择信息增益最大的特征 X_i ,若 X_i 的信息增益小于 ε ,则 T 为单结点树,并将 D 中实例数最多的类 C_k 作为该结点的类标记,返回 T ;

(d) 若 X_i 的信息增益大于等于 ε ,对 X_i 的每一个可能值 x_j ,按照 $X_i = x_j$,将 D 划分为若干非空子集 D_j ,将 D_j 中实例数最大的类作为标记,构建子结点,由结点及其子结点构成树 T ,返回 T ;

(e) 对第 j 个子结点,以 D_j 为训练集,以 $X - \{X_i\}$ 为特征集,递归地调用(a)-(d),得到子树 T_j ,返回 T_j 。

对测试图像,通过滑动窗口的方法获得当前待判别区域,使用训练好的决策树模型判别该区域是否是行人,若判别结果为正,则为该窗口所在区域添加标注,标注形式为 (x, y, w, h) ,其中 (x, y) 为该区域左上角的坐标, (w, h) 为该区域的宽度和高度;

5) 使用卷积神经网络模型对上述得到的每个位置进行筛选,得到最后的检测结果。

对上一步得到的图像中所有标注过的区域,使用卷积神经网络分类模型对此区域进行再次分类。本发明在实验中采用了 AlexNet 网络结构(实际实施过程可不限于该网络结构),此网络中有五个卷积层:卷积层 1(conv1, 包含 96 个 $11 \times 11 \times 3$ 的卷积核),卷积层 2(conv2, 包含 256 个 5×5 的卷积核),卷积层 3(conv3, 包含 384 个 3×3 的卷积核),卷积层 4(conv4, 包含 384 个 3×3 的卷积核),卷积层 5(conv5, 包含 256 个 3×3 的卷积核);两个全连接层:全连接层 1(fc1, 输入为 $6 \times 6 \times 256$ 维的卷积特征向量,输出为 4096 维)和全连接层 2(fc2, 输入和输出均为 4096 维)。这个过程中对于要分类的每个区域 r ,经过此网络的计算后,将会得到一个 4096 维(fc2 的输出维度)的特征向量作为此区域的特征表达,使用该表达,通过一个分类层(Softmax 层,该层依据本发明任务添加,它的输入为 fc2 层所输出的 4096 维

的向量，输出为 2 维的向量，输出向量中每个维度的取值范围在[0,1]区间内，其中第一维表示此区域为行人的概率，第二维表示此区域不为行人的概率)，从而得到该区域是否为行人。图 6 是本发明实施例使用卷积神经网络对前期提取的行人位置进行分类过程的示意图。如图 6 所示，对所得概率值大于 0.9 的区域，标注其为行人，标注形式为(x, y, w, h, s)，其中(x, y, w, h)与步骤 4 中标注意义相同，s 为所得概率值，从而得到最后的检测结果。

在本发明实施例中，我们采用了输入图像的颜色特征通道，具体使用图像的色彩空间而非原始的 RGB 空间，因为 CIELUV 色彩空间中光照的变化对所提取的特征干扰较小。对于每个点所计算得到的梯度方向，将其量化到六个角度范围内，并用该方向梯度幅值对其进行加权，得到六个梯度方向的特征通道。六个角度范围为：

$[0, \pi / 3], (\pi / 3, 2\pi / 3], (2\pi / 3, \pi], (\pi, 4\pi / 3], (4\pi / 3, 5\pi / 3], (5\pi / 3, 2\pi]$ 。

最后，还将梯度幅值本身单独作为一个特征通道。因此，对于输入有总共十个特征通道的数据作为其特征表示。

对于检测性能的评测，采用平均误检数(False Positive Per Image, FPPI)与漏检率(Miss rate)指标作为最终评测指标，该值表示固定误检数时，每幅图像的漏检率是多少，该指标数值越低说明算法的检测性能越好。图 7 显示了本发明与其他一些成熟算法在通用的 INRIA 行人数据集下检测性能指标的结果对比，其中横坐标为平均误检数（单位：个/帧），纵坐标为漏检率（%）。OURs 为本发明方法的检测性能，从图中可看出，本发明方法 OURs 的检测性能优于 VJ（OpenCV 中使用的方法）、梯度直方图(Histogram of Gradient, HOG)算法和聚合通道特征(Aggregate Channel Features, ACF)算法的检测性能。

需要注意的是，公布实施例的目的在于帮助进一步理解本发明，但是本领域的技术人员可以理解：在不脱离本发明及所附权利要求的精神和范围内，各种替换和修改都是可能的。因此，本发明不应局限于实施例所公开的内容，本发明要求保护的范围以权利要求书界定的范围为准。

权 利 要 求 书

1、一种基于聚合通道特征的快速行人检测方法，包括前期位置标定过程和后期位置筛选过程；前期位置标定过程聚合输入视频或图像中的多个通道特征信息，构建输入图像或视频在尺度空间中的图像金字塔，对每个尺度计算各个通道特征的图像，识别作为存在行人判别依据的特征，初步提取行人位置；后期位置筛选过程使用卷积神经网络分类器对前期标定的每个位置做进一步筛选，提高对输入视频和图像中行人进行检测的精度；具体包括如下步骤：

1) 对每幅输入图像或视频，计算得到所述图像或视频的多通道的特征图像；

2) 构建特征图像金字塔：利用步骤 1) 计算出的特征图像，得到所述图像或视频在尺度空间内的特征图像金字塔；

3) 根据步骤 2) 得到的特征图像金字塔，对所述图像或视频中待识别的图像目标区域，使用所述区域内所有通道的特征图像，将所述区域表示为统一维度的特征向量，作为聚合通道特征；

4) 使用决策树模型进行训练，从步骤 3 所述特征向量中自动选择出行人鉴别特征作为判别标准，训练得到决策树分类器，通过滑动窗口扫描对不同尺度大小的行人进行位置标定，标定位置表示为在窗口所在区域添加标注；

5) 使用卷积神经网络模型对步骤 4) 得到的每个位置进行筛选，得到最后的检测结果。

2、如权利要求 1 所述基于聚合通道特征的快速行人检测方法，其特征是，所述多通道包括 LUV 通道、梯度方向通道和梯度幅值通道。

3、如权利要求 1 所述基于聚合通道特征的快速行人检测方法，其特征是，步骤 2) 具体通过对所述输入图像进行上下采样的方法，构建得到所述输入图像在尺度空间内的特征图像金字塔。

4、如权利要求 1 所述基于聚合通道特征的快速行人检测方法，其特征是，步骤 3) 中，所述特征图像金字塔中包含的尺度为五个；所述通道为十个；所述聚合通道特征具体通过如下操作得到：

设定输入图像为 I ；待识别图像区域为 R ，在 I 的特征图像金字塔中包含五个尺度，记为 $L_1, L_2, \dots, L_5$ ；各个尺度上十个特征通道的图像记为 $P_{xy}(x = 1, \dots, 5; y = 1, \dots, 10)$ ，每个尺

度上待识别图像区域R对应的区域记为 $r_1, r_2, \dots, r_5$ ；设 $i=1..5$ ，对尺度 L_i 下的区域 r_i ，将所述各通道图像 $P_{iy}(y=1, \dots, 10)$ 上的相应区域列在一起，形成区域 r_i 的聚合通道特征。

5、如权利要求1所述基于聚合通道特征的快速行人检测方法，其特征是，步骤4)所述训练得到决策树分类器具体为：在使用决策树模型进行训练过程中，计算特征向量中每个维度的特征所带来的信息增益，对信息增益进行排序，依次选取出当前信息增益最大的特征，代表鉴别能力最强的行人鉴别特征，将所述行人鉴别特征作为所述决策树的结点，得到决策树分类器。

6、如权利要求5所述基于聚合通道特征的快速行人检测方法，其特征是，所述决策树模型具体为迭代变换算法，输入为训练集D、特征集X和阈值 ϵ ，输出为决策树T；所述迭代变换算法执行如下操作得到决策树分类器：

41) 通过式1计算得到特征向量中每个维度的特征 X_i 为训练集D所带来的信息增益 $g(D, X_i)$ ：

$$g(D, X_i) = H(D) - H(D | X_i) \quad (\text{式 1})$$

式1中， $H(D) = -\sum_{k=1}^n p_k \log p_k$ 为训练数据集D本身的熵，其中 p_k 为训练数据集D中第k个类别出现的概率，k只有两个取值， $k=1$ 代表行人和 $k=2$ 代表非行人；

$H(D | X_i) = \sum_{j=1}^m p_j H(D | X_i = x_j)$ 为特征 X_i 给定条件下训练集D的条件熵；

42) 对步骤41)得到的不同特征向量所得信息增益进行排序，选取出信息增益最大的特征向量作为行人鉴别特征，将所述行人鉴别特征作为所述决策树的结点，得到决策树分类器。

7、如权利要求1所述基于聚合通道特征的快速行人检测方法，其特征是，步骤4)所述标注的形式为 (x, y, w, h) ；其中， (x, y) 为所述区域左上角的坐标， (w, h) 为所述区域的宽度和高度。

8、如权利要求1所述基于聚合通道特征的快速行人检测方法，其特征是，步骤5)所述使用卷积神经网络模型对步骤4)得到的每个位置进行筛选，所述卷积神经网络分类模型采用AlexNet网络结构，包括五个卷积层和两个全连接层；通过对步骤4)得到的图像中所有添加标注的区域，使用所述AlexNet网络进行再次分类，标注出行人，完成行人检测。

9、如权利要求 8 所述基于聚合通道特征的快速行人检测方法，其特征是，所述再次分类包括如下步骤：

51) 通过所述 AlexNet 网络的五个卷基层将要分类的每个区域 r 表示成一个 4096 维的向量，作为区域 r 的特征向量；

52) 通过一个分类层 Softmax 层，输入步骤 51) 所述 4096 维的特征向量，输出一个 2 维向量；所述输出向量中的第一维表示所述区域 r 识别为行人的概率，所述输出向量中的第二维表示所述区域 r 识别不为行人的概率；

53) 通过设定概率阈值，确定所述区域 r 是否为行人，完成行人检测。

10、如权利要求 9 所述基于聚合通道特征的快速行人检测方法，其特征是，步骤 53) 所述概率阈值设定为 0.9。

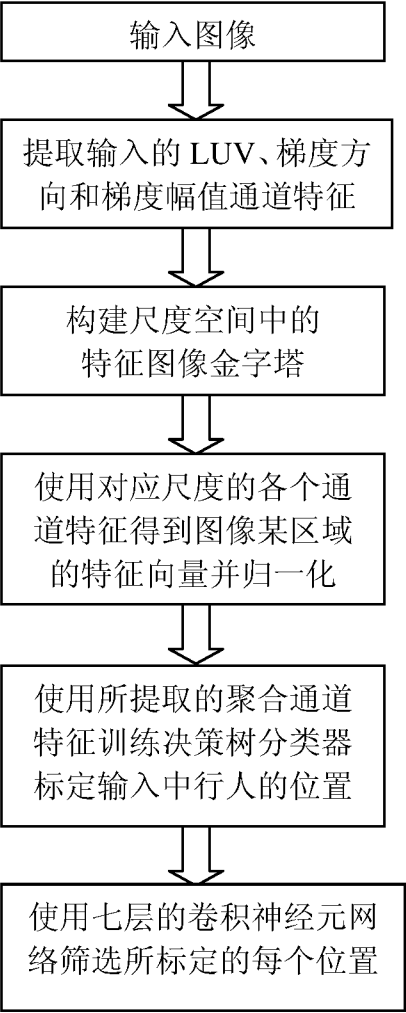


图 1

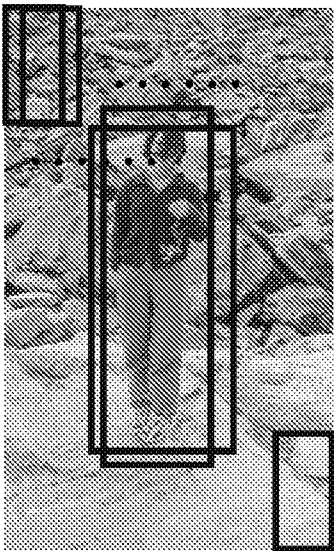


图 2

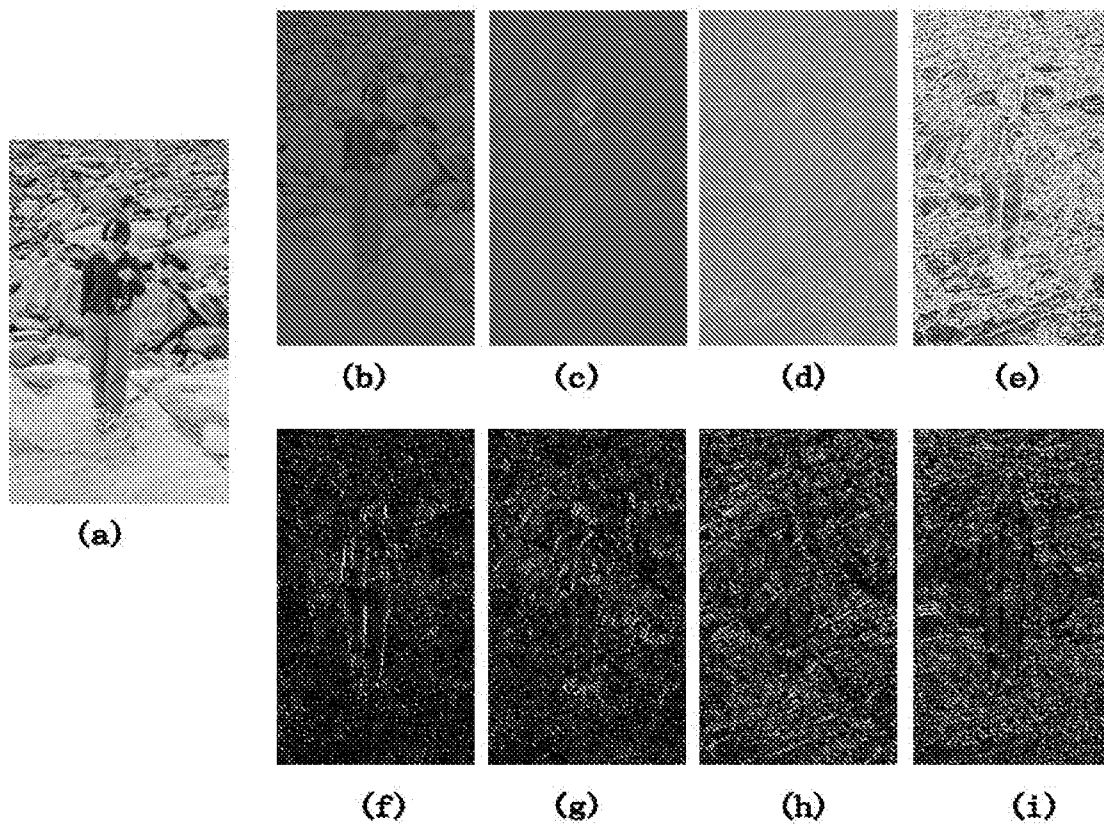


图 3

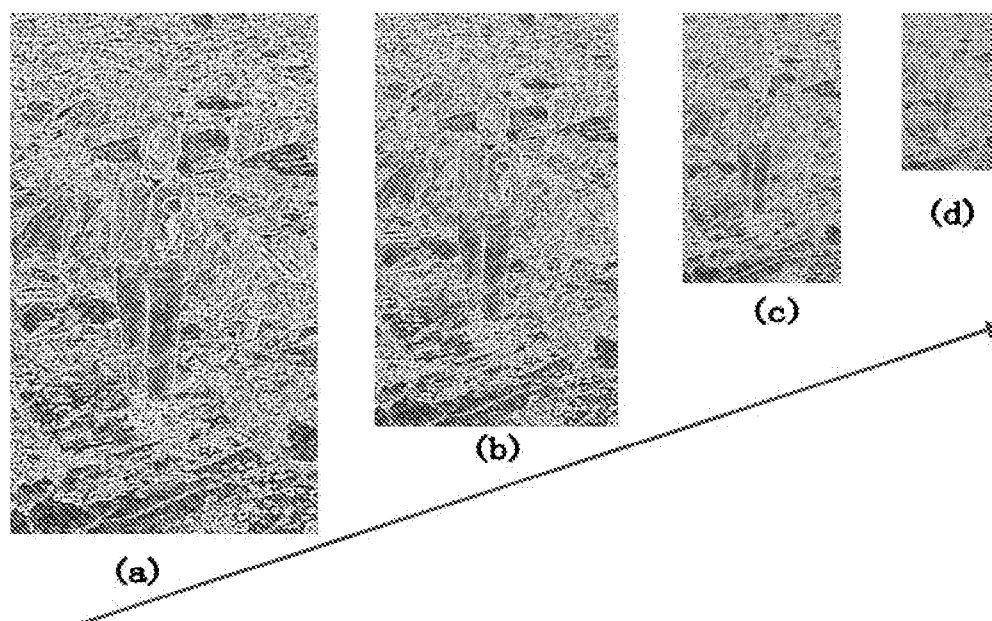


图 4

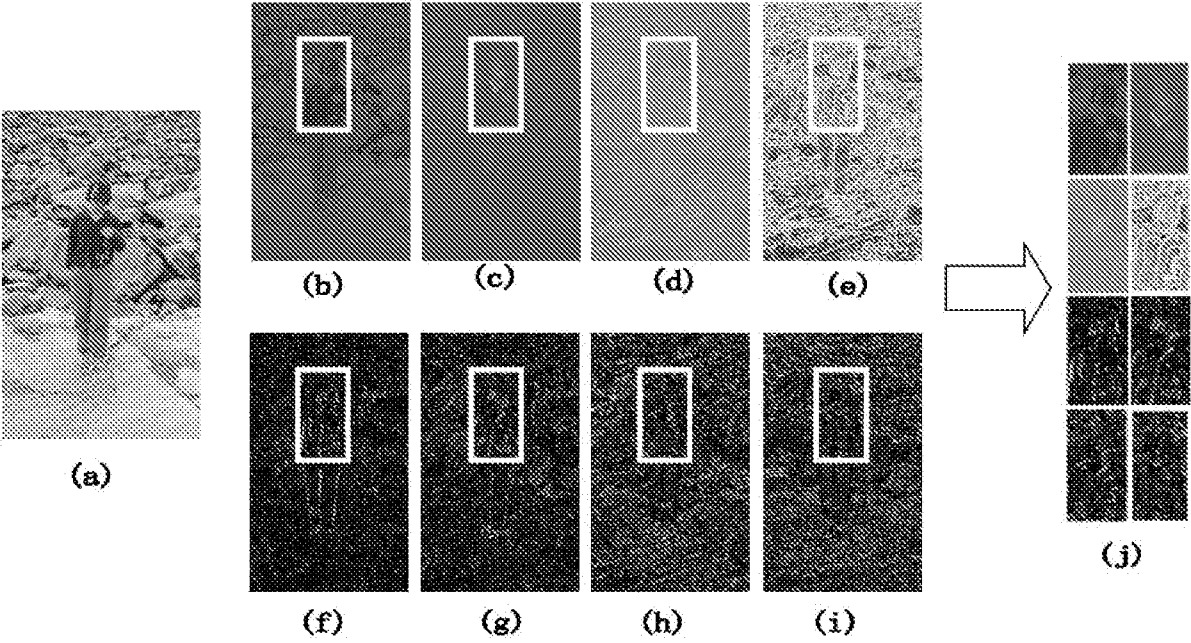


图 5

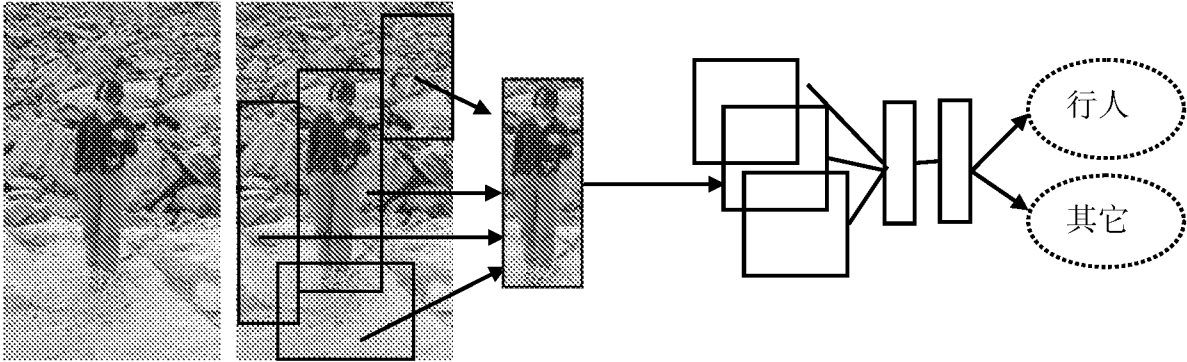


图 6

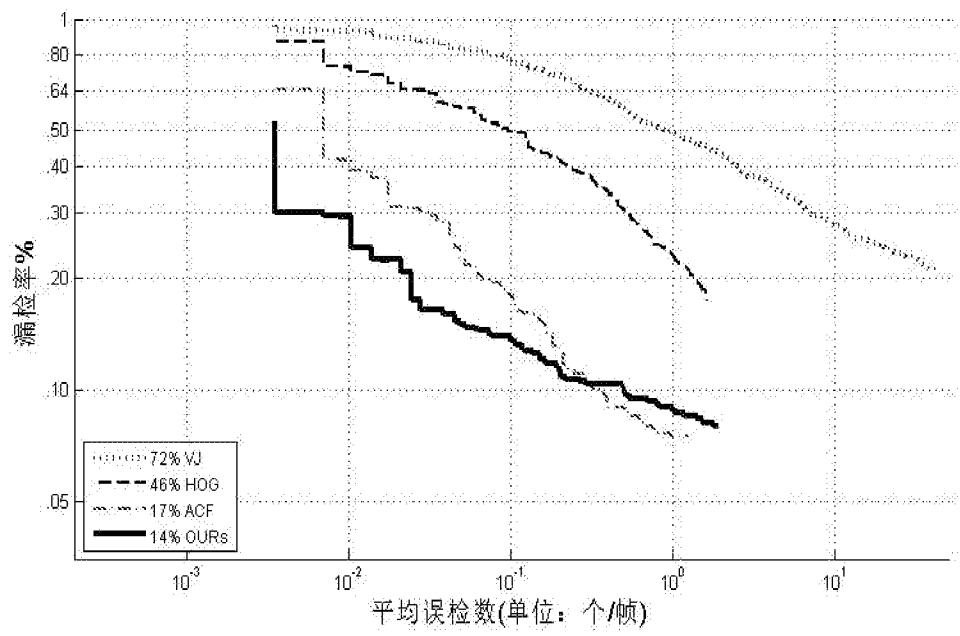


图 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/079519

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC: polymerise, channel, characteristic, pedestrian, detect, determine, position, utilize, convolute, neural, network, model, pyramid, robustness, vector, extract, discretion, structure, channel, image, video, input, screen

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 105975929 A (PEKING UNIVERSITY SHENZHEN) 28 September 2016 (28.09.2016) claims 1-10	1-10
PX	CN 106339686 A (JIANGSU UNIVERSITY) 18 January 2017 (18.01.2017) description, pages 1-4	1-10
PX	WAN, Wei, "The Object Detection Algorithm Research And Application Based on Deep Learning" The Master Dissertation of University of Electronic Science and Technology of China, 01 September 2016 (01.09.2016), the text, pages 37-60	1-10
Y	LI, Qingwu et al., "Real Time Pedestrian Detection based on Multi-scale Aggregation Channel Features" Journal of Electronic Measurement and Instrumentation, volume 29, number 11, 30 November 2015 (30.11.2015), the text, pages 1691-1695	1-10
Y	CN 105046197 A (XIDIAN UNIVERSITY) 11 November 2015 (11.11.2015) description, pages 3-5	1-10
A	CN 103679187 A (SHENZHEN INSTITUTE OF ADVANCED TECHNOLOGY) 26 March 2014 (26.03.2014) the whole document	1-10

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	
"E" earlier application or patent but published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	"&" document member of the same patent family

Date of the actual completion of the international search 08 June 2017	Date of mailing of the international search report 28 June 2017
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer LI, Kai Telephone No. (86-10) 62413120

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/CN2017/079519

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 105975929 A	28 September 2016	None	
CN 106339686 A	18 January 2017	None	
CN 105046197 A	11 November 2015	None	
CN 103679187 A	26 March 2014	None	

A. 主题的分类 G06K 9/00 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类		
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06K 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNPAT, CNKI, WPI, EPDOC: 金字塔, 神经元, 分类器, 鲁棒性, 通道, 检测, 筛选, 输入, 视频, 图像, 构建, 尺度, 识别, 判 别, 提取, 卷积, 向量; polymerise, channel, characteristic, pedestrian, detect, determine, position, utilise, convolute, neural, network, model		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 105975929 A (北京大学深圳研究生院) 2016年 9月 28日 (2016 - 09 - 28) 权利要求1-10	1-10
PX	CN 106339686 A (江苏大学) 2017年 1月 18日 (2017 - 01 - 18) 说明书第1-4页	1-10
PX	万维. "《基于深度学习的目标检测算法研究及应用》" 《电子科技大学 硕士学位论文》, 2016年 9月 1日 (2016 - 09 - 01), 正文第37-60页	1-10
Y	李庆武 等. "《基于多尺度聚合通道特征的实时行人检测》" 《电子测量与仪器学报》, 第29卷, 第11期, 2015年 11月 30日 (2015 - 11 - 30), 正文第1691-1695页	1-10
Y	CN 105046197 A (西安电子科技大学) 2015年 11月 11日 (2015 - 11 - 11) 说明书第3-5页	1-10
A	CN 103679187 A (深圳先进技术研究院) 2014年 3月 26日 (2014 - 03 - 26) 全文	1-10
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: "A" 认为不特别相关的表示了现有技术一般状态的文件 "E" 在国际申请日的当天或之后公布的在先申请或专利 "L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的 公布日而引用的或者因其他特殊理由而引用的文件 (如具体 说明的) "O" 涉及口头公开、使用、展览或其他方式公开的文件 "P" 公布日先于国际申请日但迟于所要求的优先权日的文件 "T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解 发明之理论或原理的在后文件 "X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是 新颖的或不具有创造性 "Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并 且这种结合对于本领域技术人员为显而易见时, 要求保护的发 明不具有创造性 "&" 同族专利的文件		
国际检索实际完成的日期 2017年 6月 8日		国际检索报告邮寄日期 2017年 6月 28日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10) 62019451		受权官员 李凯 电话号码 (86-10) 62413120

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/079519

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	105975929	A	2016年 9月 28日	无	
CN	106339686	A	2017年 1月 18日	无	
CN	105046197	A	2015年 11月 11日	无	
CN	103679187	A	2014年 3月 26日	无	