1. 绪论

1.1 课题研究背景及意义

随着计算机科学技术的不断发展，人工智能的概念被提了出来，使得计算机可以在很多场景上代替人类。机器视觉就是人工智能的一个主要应用领域，其主要目标便是通过对图像的处理和分析，模拟人类的视觉识别能力并做出相应的判断。

今年来，随着城市建设的不断发展，以及人们对高品质生活的不断追求，机器视觉领域得到了快速的发展，市场也有巨大的需求。行人检测作为机器视觉领域的一个重要研究方向，该技术在道路交通，智能安防以及娱乐产业等很多领域都有着及其重大的研究意义。其主要应用在以下几个领域:

1. 智能驾驶

我国的汽车保有量增长迅速，相应的交通事故也不断增长，交通事故的主要原因也是人为疏忽导致，行人检测装置检测到行人时就会提醒司机或者自动采取制动措施，从而避免交通事故的发生。行人检测在智能驾驶的作用主要体现在辅助驾驶和无人驾驶上。行人检测技术需要相关的图像采集设备或者雷达等，设备采集到数据后处理器应用相应的图像识别算法分析数据最终得出是否有行人的结论，无人驾驶汽车则要做出相应的措施去避让行人，辅助驾驶系统则通过语音提示的方式告诉司机，司机再做出合适的处理。

1. 安防监控

在很多时候我们需要对特定场合的行人进行监控，比如监狱外围区域，私人住宅的行人进行监控，发现可疑人物就会及时报警；军事基地还可以监控外来人员非法入境；或者监控中发现行人出现异常举动，就可以通知公安、消防等快速介入，保障生命财产不受损失。尤其在金融机构、人员流动大的公共场所更加需要智能化的安防监控系统。早期，监控主要通过人来对显示屏上的图像进行监视，不仅需要大量的人力和物力，而且监控效率并不能得到保证。因此，行人检测技术就有了用武之地，系统可以自动完成行人检测和人的行为分析，实现了对可疑人物、危险行为的准确预警。

1. 机器人导航

通过行人检测技术，机器人可以找到感兴趣的区域，然后导航到目的区域，典型的应用就是现在的无人机可以对行人进行跟随，可以在完全无人操作的情况，主动跟随目标，或者在行人的周围定点环绕。

本文主要研究的是行人检测技术在智能驾驶上的应用，ADAS(高级驾驶辅助系统)现在已经开始广泛的应用在我们的汽车上，为我们驾驶过程中提供了很多辅助功能。行人检测效果图如图1-1，行人识别就是辅助驾驶系统的一个主要功能，行人识别技术应用在辅助驾驶的难点是检测的精度和速度。所以人们对行人检测技术的算法性能提出了更高的要求，在保证检测的精度的情况下，还必须提高检测的速度，保证检测的实时性。在早期提出的算法中，鉴于汽车设备环境的限制，在精度和实时性上面都不适合应用在辅助驾驶上，在最近几年一些优秀的检测算法被提出了，其中基于CENTRIST特征的行人检测算法不管是检测精度和速度上都让行人检测技术应用在汽车辅助驾驶系统有了较好的效果。而且GPU的并行运算可以进一步优化检测的速度，从而保证了检测系统的实时性。

1.2 课题研究现状分析

行人检测算法在最近的十几年发展迅速，国内外的研究人员对行人检测算法的准确性和实时性进行了大量研究，接下来对这些研究工作进行分析。其中主要的研究来自于国内外的高校和科研机构，国内主要有清华大学、南京大学、中科院等，其中中科院计算机科学重点实验室及自动化研究所都取得了巨大成绩。除此之外，国内许多科技公司也在该领域投入了大量的人力物力，其中包括目前比较火热的百度，成立了百度大脑，并将行人检测技术应用在了无人驾驶上，还有类似商汤科技在人脸识别和目标识别领域都取得了巨大成就，并将技术转换为商业产品应用在了我们生活中，为我们的生活提供了便捷。国外的加州理工，麻省理工，谷歌，脸书等高校和企业都在进行行人检测课题的研究；每年都有很多优秀的论文发表在计算机视觉权威的机构上。



图1-1

1.2.1 行人检测算法的研究现状

行人检测目前最常用的方法是基于统计学习的方法，根据大量的样本构建行人检测分类器。提取特征主要有目标的灰度，边缘，纹理，颜色，梯度直方图等信息。分类器主要包括神经网络、SVM、adaboost以及当前十分火热的深度学习。统计学习目前的难点有：行人的姿态、服饰各不相同、复杂的背景、不同的行人尺度以及不同的关照环境；提取的特征在特征空间中的分布不够紧凑；分类器的性能受训练样本的影响较大；离线训练时的负样本无法涵盖所有真实应用场景的情况。

2005年法国研究人员Dalal在2005的CVPR发表的HOG+SVM的行人检测算法，基于HOG特征的行人检测提出，使得行人检测领域有了较大突破，HOG+SVM作为经典算法也别集成到opencv里面去了，可以直接调用实现行人检测。

基于统计学习方法的行人检测技术很重要的一点就是样本集，而当前关于行人检测主要的数据库有[MIT数据库](http://cbcl.mit.edu/software-datasets/PedestrianData.html" \t "_blank)，INRIA数据库，[Caltech行人数据库](http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/" \t "_blank)等。其中最常用的是INRIA和Caltech数据库，INRIA数据库是目前使用的最多的静态行人检测数据库，提供原始图片及相应的标注文件。训练集有正样本614张（包含2416个行人），负样本1218张；测试集有正样本288张（包含1126个行人），负样本453张。Caltech数据库是目前规模较大的数据库，采用车载摄像头拍摄，约10个小时左右，视频的分辨率为640x480，30帧/秒。标注了约250,000帧（约137分钟），350000个矩形框，2300个行人，另外还对矩形框之间的时间对应关系及其遮挡的情况进行标注。数据集分为set00~set10，其中set00~set05为训练集，set06~set10为测试集（标注信息尚未公开）。

HOG特征描述了一幅图像的方向梯度信息，提取该特征采用的滑动窗口扫描图像的方式，将窗口划分为大小相同的图像块（block），以块为单位提取图片的特征，这种方法可以保证提取到稳定的HOG特征，HOG+SVM的结合在INRIA数据库上的检测精度接近90%，但是HOG特征的提取计算量高，所以此方法的检测速度很慢，无法实现实时检测。

2009年，Dollar等提出了积分通道特征（Integral channel feature），该方法使用了梯度幅度特征，图像的LUV特征，及梯度直方图特征，为了简化计算使用了积分图像的方法实现特征的快速计算，该方法由于提取了多种特征来描述图像，所以检测精度可以达到86%，而且检测速度也可以得到提高。但是其复杂的特征提取导致检测还是无法做到实时性。

2011年吴建鑫等提出了基于轮廓信息的CENTRIST特征，作者通过大量的实验证明了获取行人边缘轮廓信息就能够准确描述一个行人的特征，而且CENTRIST特征的计算也比较简单，检测效果与积分通道特征相比也相差无几，而且作者在分类器上使用的是级联分类器，首先是一个线性SVM分类器快速找出可能的行人区域，再将结果交给第二层的HIK SVM分类器检测，因此保证了算法的检测精度，所以此方法是一个检测精度和速度都比较可行的一个算法。

1.3 本文组织结构及主要工作

由于主要研究的是行人检测技术应用在汽车辅助驾驶系统上，然而在这种应用场景下，汽车上的设备有限，所以很难做到实时性检测，这就要求我们设计出一套实时性高的行人检测系统。针对这个问题，我们提出了基于BING和CENTRIST的行人检测系统，并通过GPU加速计算，从而达到实时检测的目的。本文的主要研究内容如下：

（1）绪论介绍了行人检测技术的国内外现状，以及其常见的应用领域，以及行人检测技术遇到的一些难点。

（2）介绍了BING算法代替滑动窗口检测目标的原理，以及BING算法是如何帮助我们减少不必要的重复计算的。

（3）介绍了基于CENTRIST特征的检测算法，并与基于HOG特征的行人检测算法做了比较，证明了CENTRIST特征不但可以减少计算量，还可以进一步提高检测的准确率。

（4）介绍了GPU并行运算架构，并深入分析了算法模块中并行化计算的可行性，用CUDA编程实现了算法的并行化实现，节省了大量时间，提高了算法的运行效率。

第三章 BING似物性算法原理

本文提出的是一种结合似物性检测和统计学习的行人检测算法。传统的行人检测算法大都采用滑动窗口产生候选框，然后将候选框的特征输入到训练好的分类器去进行分类，从而判定一个候选框内是否有行人，BING是通过对图像进行预处理，产生带评分的建议窗口通过分类模型进行评估得分，实现检测行人的功能。

3.1 似物性采样的概念

受到生物学的启发，发现人在面对复杂场景时，人眼可以快速，准确的将注意力放在比较明显的视觉对象上，并且优先理解显著性物体。这种基于视觉显著性的物体检测框架受到越来越多关注。算法首先是能够很好地定位图中的显著区域，识别出主要物体。似物性算法通常是采用训练学习或对输入图像定性分析，对不同大小尺寸的窗口进行评分排序，然后提取钱N个独立窗口的坐标信息，最后将得到的候选框交给目标检测算法进行进一步分类。因此，用似物性检测产生的建议窗口，是行人检测预处理阶段的输出结果，使得后面的分类算法的计算量更小。

3.1.1 似物性采样算法分析

目前使用较多，性能较好的似物性算法主要是基于窗口的检测，此方法通过一个公式来打分，根据分数来确定窗口包含物体的可信度，然后把窗口的分数进行排序。下面我们主要会介绍程明明老师提出的BING算法。

BING算法使用梯度作为输入特征，首先训练一个一级分类器，再根据尺寸大小和一级分类器得分训练出36种不同尺寸的二级分类器。BING算法采用了特殊的计算方式，将梯度二值化，使得算法的检测速度加快，实验证明，BING算法可以很高效的产生一系列类别独立，高分辨率的对象窗口，通过使用1000个建议窗口，我们的对象检测准确率高达96.2%，通过增加建议窗口数量或者考虑颜色空间来计算BING特征，实验结果可以提高到99.5%。

3.2 BING算法分析

BING（Binarized normed gradient）是由程明明教授在2014年在CVPR上提出的一种高效的目标建议生成算法，算法的效率可以达到300fps。在VOC2007数据集可以达到96.2%的探测率。

BING算法是基于一个独立事物都拥有很好的闭合边缘特征和中心，相反背景都是杂乱的没有统一特征和纹理。作者提出了无论什么形状和类别的图像，只要是一个独立的物体，那么把它的图像响应窗口调整到一个小的固定大小的尺寸（常用的是8\*8的尺寸），那么它的NG特征（二值化梯度幅值）会有很明显的闭合边缘共性如图3-1所示。而且BING算法采用了一定的技巧使得只需要一些原子操作计算（例如加法，按位移动等）就能完成部分计算。

BING特征启发于能够在识别对象之间感知它的人类视觉系统，因此引入了一个64维的梯度特征幅值特征，二值化的梯度幅值特征（BING）很有效的获取到图像窗口的对象状态。为了能够找到图像中的一般对象，我们扫描一个定义好的量化窗口（依据尺度或者是纵横比）。

每一个窗口通过一个线性模型获得得分：

（1）

l=(i,x,y) （2）

代表过滤器得分，代表NG特征，l代表坐标，i表示尺度，（x,y）表示窗口位置。运用非极大值抑制（NMS），我们为每个尺度提供一些建议窗口。相对于其他窗口（例如：100\*100），一些尺度（10\*500）的窗口包含对象的可能性是很小的。因此我们定义对象状态得分（校准过滤器得分）：

（3）

其中，，针对不同窗口尺度i的窗口，得到不同的独立学习系数。使用校准的非常快的，通常只需要在最终的建议窗口重排之后进行。

3.3 BING特征计算原理、

为了使得计算速度算法提出了一个NG特征的加速版，二值化梯度幅值，加速特征的提取和检测过程。我们学习的线性模型参数可以近似表示为一系列基向量的组合:

其中表示基向量的个数，

第二章 基于CENTRIST特征的行人检测算法原理

2.1引言

本章首先阐述了CENTRIST特征的特点，以及基于CENTRIST特征的行人检测算法的主要步骤，首先是如何提取图像的CENTRIST特征，以及采用级联分类器对图片进行滑动窗口检测，检测图片中是否有行人，最后比较了CENTRIST特征和HOG特征在检测准确率和速度方面的差距。

2.2 CENTRIST特征简介

2.2.1 CENTRIST特征的原理

CENTRIST（C4）特征能够很好的描述行人的特征，而且特征的提取相对传统方法计算更加简单，待特征提取完毕后就可以直接交给分类器进行判断的出结果。C4在目前最高精确度下可以达到20帧每秒的检测速度，并且是在只使用一个处理线程和不使用GPU等硬件的情况下达到的。能达到实时而精确的检测源于以下两点：第一，相邻像素差值的符号是描述轮廓的关键信息；第二，CENTRIST描述子非常适合做人体检测，因为它编码了符号信息并且可以隐式地表达全局轮廓。使用CENTRIST描述子和线性分类器，我们提出了一种不需要显式生成特征向量的计算方法，它不需要图像的预处理或特征向量的归一化，只需要O(1)时间去测试一个图片区域。

C4特征是针对行人的局部轮廓提出的，相邻像素的大小关系是编码的关键，具体像素的值对该特征的计算影响并不大。图片中的行人的衣服上的图像，绘画等纹理性特征对行人检测有较大影响。因此首先用Sobel边缘检测算子对这些纹理性的特征进行消除，同时又能较好的保留行人的整体轮廓。所以我们使用Sobel算子先对图像进行平滑处理，用Sobel梯度值代替像素的大小，这样就可以消除这些局部的高频纹理特征带来的不理影响。因此通过Sobel算子平滑后的原始图像*Ｉ*可以得到一个新的图像*Ｉ*’ ，*Ｉ*和*Ｉ*’的对应关系如下图2-1：



图 2-1

如图2-2所示，原始图像（a）经过Sobel进行平滑处理后，图像很好地过滤掉了一些不重要的纹理特征，（b）很好的描述了行人的边缘轮廓信息。



图 2-2

C4特征中一个重要的概念是CT（Census Transform）值，图像上某个像素的CT值由该像素的灰度值和其九邻域的像素的灰度值比较而得来，具体计算的方法如图2-1所示。若相邻的像素的灰度值比中心像素的灰度值大，则把这个位置的值置为0，否则置为1。然后将邻域内的八个数字按照从上到下从左往右的一个顺序排列起来得到一个八位的二进制数，最后将这个二进制数转为十进制数，这个十进制数就是中间那个像素的CT值。而且由于是一个八位的二进制数得到的，所以这个CT值的取值范围就是0到255。



图 2-3

我们通过计算CT值来代替图像的灰度值，从而得到新的图像，我们称为CT图像。为了更好的描述，在遍历一张图像时，使用一个256维度的CT值直方图来统计每个CT值的在图像中出现的频率。最终得到的这个直方图就称之为CENTRIST特征。

在实际应用中，把一个图像分为若干个小的图像块，我们取的是一个108\*36大小的检测窗口，然后把这个块分割为9\*4的图像小块（block），每2\*2的邻接小块组成一个超级块（super-block），所以通过计算得到一共有24个超级块，计算特征时就以一个超级块为单位，则整张图片的CENTRIST描述特征就是256\*24=6144维大小的直方图。（由于CT值的计算是与3\*3的模板进行匹配，因此超级块的边缘像素不会参与CT值的计算）。

2.2.2 CENTRIST特征的计算过程

上一节我们介绍了特征的计算原理，本节我们会介绍提取一副图像的CENTRIST特征的过程，包含以下步骤：

1. 读取原始图片到内存中；
2. 采用滑动国定大小的检测窗口对图片提取CENTRIST特征。但是由于真实图像中行人的大小各异，为了能够不漏检，因此我们必须要对图像进行缩放，使得各个尺度的行人能够被滑动窗口检测到，如图2-5所示，红色的矩形框就是检测窗口。
3. 使用Sobel算子对图像进行处理，构建Sobel图像；
4. 根据前面介绍的 计算CT值的方法计算CT图像，从而提取行人的轮廓信息；
5. 检测窗口在CT图像上滑动，并计算得到窗口的CENTRIST特征；

整个流程如图2-6所示，接下来我们会介绍如何利用我们提取到的CENTRIST特征对行人进行检测。



图 2-5



图 2-6

2.3 级联分类器

基于CENTRIST特征的行人检测算法使用了线性SVM分类器和HIK SVM分类器串联得到的级联分类器来实现快速、准确的检测。首先要对分类器进行训练。

2.3.1 SVM原理

支持向量机（SVM）是90年代中期发展起来的基于统计学习理论的一种机器学习方法，通过寻求结构化风险最小来提高学习机泛化能力，实现经验风险和置信范围的最小化，从而达到在统计样本量较少的情况下，亦能获得良好统计规律的目的。通俗来讲，它是一种二类分类模型，其基本模型定义为特征空间上的间隔最大的线性分类器，即支持向量机的学习策略便是间隔最大化，最终可转化为一个凸二次规划问题的求解。2005年，Dalal等研究员，将HOG特征和支持向量机结合起来应用在行人检测取得了巨大成功。

支持向量机的基本思想：定义一个线性最优超平面，使得超平面可以很好的将数据集分开。如图2-7是一个线性二分类的例子，一个二维平面，平面中有两种不同的点，分别用两种不同的颜色表示，一种为红色的点，另一种为蓝色的点，红色的直线是一个超平面，由图可见直线将平面内的点完全分开，这条直线就是我们要求解的超平面，下侧的全是-1，上侧的全是1。



图 2-7

接着我们可以假设超平面为：

其中 是法向量，为偏置项，()决定了超平面方程，样本到超平面的平均距离为：

如果()表示的超平面可以对样本集进行分类，即，如果=1则，

如果=-1，则。令

所有满足上公式取等号的样本点我们称之为支持向量，正负两类的支持向量到超平面的距离之和为:

其中为点到平面的间隔，为了是最大，就是求满足公式的前提下求解最优参数()，也就是如下所示

s.t.

因此当最大，就可以使得分类间隔最大。等价于让取最小值，因此公式(4-5)可以改写为：

s.t.

因此问题就转化为了求在约束条件下求最优解的问题，而且这是一个凸二次优化问题，我们使用拉格朗日乘子法和KKT条件来求w和b，首先构造拉格朗日函数，具体的计算方法此处就不做介绍了，请查阅相关资料了解。

之前我们介绍了支持向量机的基本原理，但是对于线性不可分的样本集，我们需要将数据集映射到高维空间去，从而将数据线性分开，但是由于映射到高维空间，数据的维度会爆发式的增加，使得计算量很大，所以需要引入核函数来降低计算量。那么接下来就是要在高维空间学习找到线性可分的超平面，经常使用的核函数主要有以下几种：

（1）线性核函数：

（2）多项式核函数：

（3）高斯径向核函数：

（4）Sigmoid核函数：

2.3.2 HIK SVM原理

在目标检测领域，最重要的应该是分类器模型，优秀的分类器模型对系统能够出色完成分类任务是最关键的一步，之前SVM取得了较好的分类效果，也成功应用在了目标检测的众多领域。但是在实际应用中，许多情况核函数的选择才是关键，因此如何选择核函数是分类模型的重要一步。

当前在计算机视觉领域，人们大量运用直方图作为图像特征描述符。其中以直方图交叉核函数（HIK）作为核函数的SVM分类器被证明在处理直方图特征的分类能够获得更好的效果。直方图交叉核成功的原因主要有以下两点：

（1）第一，在计算机视觉领域，有很多特征描述都是使用直方图结构进行表述，如常见的：SIFT，HOG，CENTRIST特征等，由于直方图简单明了，可以有效的统计特征，因此获得了广泛的应用。直方图交叉核，以下简称HIK，可以用于比较直方图之间相似度，已经被证明比其他常用的方法有着更加好的效果。

（2）第二，在常用的分类器模型中，SVM的训练速度快，需要的样本数据量小，分类效果好，但是缺点是分类速度慢，而决策树模型则在分类速度上有着明显优势。然而，使用HIK核函数的SVM则兼具两者的优点。

在近年来基于HIK的SVM分类器被广泛引用在了许多快速检测方法中，首先介绍以下直方图相交核函数相关的公式：

*=*

上式就是直方图交叉核函数的数学表达式，其中直方图相交就是在求两个直方图的对应bin中的较小的值并把所有比较得到的较小值累加起来就得到了最后的结果。由于最后的求和相当于对其归一化的得到的结果，所以的结果越大就表明两者之间就越相似。正式因为直方图交叉核函数在直方图匹配运算时的优越性，所以近年来很多快速检测算法都使用基于HIK的SVM分类器。

2.3.3 直方图相交核函数的正定性

要使用HIK作为SVM的核函数就必须证明核函数必须要满足正定性。Boughorbel证明在正实数的情况下HIK是正定的，下面我们将介绍HIK的半正定性。

首先假设非负的实数集{x>=0|x},接下来需要证明HIK核函数具备半正定性n表示样本点的个数，d表示特征的维度，作为特征的第j维向量。首先当d=1的时候，有n个样本点，，…,,并且假设其中,所以核函数有以下性质：

由此可以得到

2.3.3 训练级联分类器

(1)线性SVM的训练

训练分类器时正样本集P我们用的是108\*36的包含行人的图像，负样本集我们使用的是同样大小的不包含人的图像，把正负样本集输入训练线性分类器训练分类器H1，然后用H1分类样本集N，从而获得一个新的负样本集，再次使用P和训练得到分类器H2，如此反复训练多个分类器，使负样本N中的所有区域都被至少一个分类器认为是负样本。然后使用P和所有的负样本集训练最终的支持向量机分类器。