

## 显著性纹理结构特征及车载环境下的行人检测\*

肖德贵, 辛晨, 张婷, 朱欢, 李小乐

(湖南大学 信息科学与工程学院 嵌入式与网络计算湖南省重点实验室, 湖南 长沙 410082)

通讯作者: 肖德贵, E-mail: dgxiao@hnu.edu.cn, http://www.hnu.edu.cn

**摘要:** 边缘信息往往是视觉对象检测的关键,已有方法对边缘梯度在各个方向上进行计算,会导致计算冗余.受韦伯局部描述算子的启发,提出一种模拟人眼观察事物的发散性及显著性特点的纹理结构算子.首先,计算邻域像素与中心像素灰度值相对差的总和,除以中心像素的灰度值求出局部显著性因子;然后,通过中心发散灰度共生矩阵提取局部纹理结构;最后,构造二维直方图结合显著性因子和纹理结构,生成一定维数的显著性纹理结构特征描述算子.实验结果表明,该特征算子具有良好的边缘检测能力,应对噪声和明暗变化的鲁棒性以及强大的结构表达能力,其行人检测的准确率优于中心变换直方图和梯度方向直方图.对车辆主动安全有很高的应用价值.

**关键词:** 行人检测;显著性;灰度共生矩阵;车辆主动安全

中图分类号: TP301 文献标识码: A

中文引用格式: 肖德贵,辛晨,张婷,朱欢,李小乐.显著性纹理结构特征及车载环境下的行人检测.软件学报,2014,25(3): 675-689. <http://www.jos.org.cn/1000-9825/4438.htm>

英文引用格式: Xiao DG, Xin C, Zhang T, Zhu H, Li XL. Saliency texture structure descriptor and its application in pedestrian detection. Ruan Jian Xue Bao/Journal of Software, 2014, 25(3): 675-689 (in Chinese). <http://www.jos.org.cn/1000-9825/4438.htm>

## Saliency Texture Structure Descriptor and its Application in Pedestrian Detection

XIAO De-Gui, XIN Chen, ZHANG Ting, ZHU Huan, LI Xiao-Le

(Key Laboratory for Embedded and Network Computing of Hu'nan Province, College of Information Science and Engineering, Hu'nan University, Changsha 410082, China)

Corresponding author: XIAO De-Gui, E-mail: dgxiao@hnu.edu.cn, <http://www.hnu.edu.cn>

**Abstract:** Edge information is often the key to the detection of objects. Traditional edge detection algorithms calculate gradient omnidirectionally, which usually results in calculation of redundancy. Inspired by Weber local descriptor, this study proposes a saliency texture structure descriptor that simulates divergent and significant characteristics of human eyes observing things. First of all, it calculates the sum of relative differences between intensity of a center pixel and those of its neighborhood pixels, and divide the sum by the center pixel's intensity to get its local saliency factor. Then, it extracts its local texture structure through a divergent gray level co-occurrence matrix. At last, it constructs a two dimensional histogram as the feature vectors by combining saliency factor and texture structure. Experimental results show that the saliency texture structure descriptor has the ability of good edge detection and powerful structural expression, and is robust to noise and light and shadow changes. When used in pedestrian detection, the saliency texture structure descriptor gets much higher detection rate than other local descriptors such as CENTRIST and HOG. This descriptor can find its high application value in vehicle active safety system.

**Key words:** pedestrian detection; saliency; gray-level co-occurrence matrix; vehicle active safety

行人检测在计算机视觉领域的许多应用中起着至关重要的作用,例如视频监控、汽车驾驶员辅助系统、人体的运动捕捉系统等.图像的行人检测方法可以分成两大类:轮廓匹配和表观特征.表观特征又被定义成图像特

\* 基金项目: 国家自然科学基金(61272062)

收稿时间: 2012-07-13; 修改时间: 2012-10-24; 定稿时间: 2013-06-09

征空间(也叫做描述算子),它可以分为整体法、局部法、特征点对法。

在整体法中,Papageorgiou 和 Poggio<sup>[1]</sup>提出了 Haar 小波(HWs)特征,并用 SVM 训练行人,其中包括了行人的正面和背面.Viola 和 Jones<sup>[2,3]</sup>采用级联 AdaBoost 学习算法(即选择超过一定阈值的弱分类器组成强分类器的算法)提取基本 Haar-like 特征和扩展的两个 Haar-like 特征,用于视频监控的行人检测.Levi 和 Weiss<sup>[4]</sup>则提出边缘方向直方图(edge orientation histograms,简称 EOHs)进行人脸检测.EOHs 先计算图像梯度强度,然后按梯度方向等分成  $K$  个区间,而特征的表示是通过方向间的统计比得到的一个实数值.Haar-like 和 EOHs 都可以通过图像积分图方法加快运算速度。

Dalal 和 Triggs 等人<sup>[5]</sup>提出了梯度方向直方图特征(histogram of oriented gradient,简称 HOG).HOG 基于梯度信息并允许块间相互重叠,因此对光照变化和偏移不敏感,能有效地刻画人体的边缘特征.然而,HOG 也有其缺点:特征维度高,大量的重叠和直方图统计,使得特征的计算速度慢,进而影响实时性;遮挡处理能力较差;未利用颜色、形状和纹理等特征.针对这些缺点,近年来一些研究者提出了更多<sup>[6-9]</sup>的行人特征,有 COV,Integral Channel Feature,ACF,GGP 等。

Chen 等人<sup>[10]</sup>提出韦伯特征(Weber local descriptor,简称 WLD).WLD 由两部分组成:激励(differential excitation)和方向(orientation),其充分利用人类视觉机制韦伯定理,对明暗变化和噪声干扰有一定的鲁棒性,缺点是方向部分计算复杂。

局部法的主要思想是将人体看成是部位的组合,该方法要解决两个问题:构造有效的部位检测器和建模部位间的几何关系.Mohan 等人<sup>[11]</sup>将人体划分为头、下半身、左右胳膊这 4 个部位,取各个部位分类器的响应值作为支持向量机的输入,构建一个组合的多层次分类器来检测行人。

Edgelet 特征描述的是人体的轮廓特征,但是它描述的是人体局部轮廓的特征,包括的形状有直线、弧线等.它将人体分为几个部分来训练,比如全身、头肩部、腿部和躯干部等,每个部分都使用 adaboost 算法训练一个强分类器;在分类时,利用 4 个部分的联合概率进行决策.该算法采用的是人体的局部特征,所以在出现遮挡的情况下仍然有很好的表现,缺点是特征的计算比较复杂。

Wu<sup>[12]</sup>提取图像的 edgelet 特征用于检测静态图像中的人体,对组成人体的各个部分分别建立模型,每一个 edgelet 描述人体的某个部位的轮廓,然后再用 adaboost 算法筛选出最有效的一组 edgelet 来描述人的整体。

Wu 定义了 3 种 edgelet,包括直线型、弧形和对称型.每一个 edgelet 由一组边缘点构成,是一条具有一定形状和位置的线段.对于图像中任意的点,根据该位置是否具有和某 edgelet 形状相似的边缘得到一个响应值.如果边缘的形状与 edgelet 越相似,那么响应值就越高。

这类方法分别检测窗口的局部区域,然后再综合这些区域的检测结果来做最终的判决.优点在于能更好地处理遮挡以及行人姿势的多样性,主要问题在于如何定义局部以及如何整合来自多个部位检测器的信息。

特征点对法是将行人检测问题视为一个广义的霍夫变换:首先,通过局部特征检测器来寻找关键点;然后,在关键点的周围选取一个固定大小的图像块,通过聚类、随机森林或者最大间隔等方法建立图像块的空间分布模式;最后,通过霍夫投票方式寻找图像中的行人位置.典型的方法是 David Lowe 提出的 SIFT 特征<sup>[13]</sup>。

无论是整体、局部还是特征点对法,核心的问题是如何有效表示行人的整体特征、部位特征或者局部块特征.方向梯度直方图特征是日前广泛使用的行人特征表示,但是方向梯度无法刻画人眼视觉敏感度,信息冗余度大.本文针对这一缺点,在中心变换直方图<sup>[14]</sup>特征(census transform histogram,简称 CENTRIST)的基础上,提出一种显著性纹理结构特征,与 CENTRIST 类似于局部二值模式直方图不同,该特征融合了人眼视觉的心理物理学规律,能更好地实现光照波动、背景杂乱等道路环境下的行人检测。

## 1 显著纹理结构(saliency texture structure,简称 STS)

### 1.1 显著性纹理结构特征

根据人眼观察事物的发散性原理,提出的显著性纹理结构特征主要由两部分组成:显著因子和纹理结构。

1.1.1 显著因子

显著因子的意义在于对局部特征的显著性表述,在发散的局部特征中,局部显著性不能由差别阈限的绝对值来进行比较,而应由其相对值进行划定,这种现象可以由著名的韦伯定律解释<sup>[15]</sup>.根据韦伯定律,我们定义局部显著因子为邻域像素值和中心像素值差的总和与中心像素值的比例关系,即:

$$G(x_c) = \frac{\sum_{i=0}^{p-1} (\Delta x_i)}{x_c} = \frac{\sum_{i=0}^{p-1} (x_i - x_c)}{x_c} \tag{1}$$

其中, $G(x_c)$ 是局部显著因子, $x_c$ 是中心像素点灰度值, $x_i$ 是该像素点的邻域像素点灰度值, $p$ 是邻域像素个数.

进一步,定义当前像素点  $x_c$  的显著性因子  $\zeta(x_c)$  为

$$\zeta(x_c) = \arctan[G(x_c)] \tag{2}$$

其中,  $\zeta(x_c) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ .

为了简化,把  $\zeta(x_c)$  划分为  $R$  个区间  $I_r(r=0,1,\dots,R-1)$ .对每个区间  $I_r$ ,显著性因子的权值也不同. $\zeta(x_c)$  值越大,表明显著性越强.

显著算子有方形算子和圆形算子两类,各类算子又可以分成  $3 \times 3$  结构和  $4 \times 4$  结构,如图 1 所示.

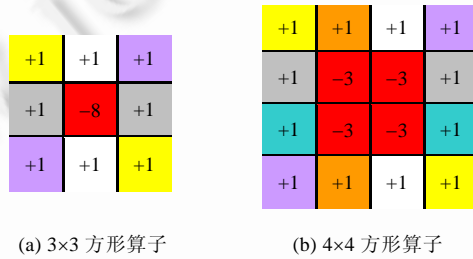


Fig.1 Saliency operators

图 1 显著算子

1.1.2 纹理结构

Haralick<sup>[16]</sup>提出的灰度共生矩阵反映了图像纹理粗糙程度.设  $f(x,y)$  为一幅  $N \times N$  的灰度图像, $d=(dx,dy)$  是一个位移矢量,其中, $dx$  和  $dy$  分别是行和列方向上的位移.灰度共生矩阵定义为从  $f(x,y)$  灰度为  $i$  的像素出发,统计与距离为  $\delta = \sqrt{(d_x^2 + d_y^2)}$ ,灰度为  $j$  的像素同时出现的概率  $P(i,j|d,\theta)$ ,如图 2 所示.

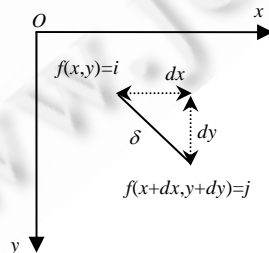


Fig.2 Pixels of the gray level co-occurrence matrix

图 2 灰度共生矩阵的像素对

数学表达式为

$$P(i,j|d,\theta) = \{(x,y) | f(x,y)=i, f(x+dx,y+dy)=j\} \tag{3}$$

根据这个定义,灰度共生矩阵的第  $i$  行第  $j$  列元素表示图像上两个相距为  $\delta$ 、方向为  $\theta$ 、分别具有灰度级  $i$

和  $j$  的像素点对出现的次数.其中, $(x,y)$ 是图像中的像素坐标, $x$  和  $y$  的取值范围为 $[0,N-1]$ ; $i,j$  的取值范围为 $[0,N_g-1],N_g$  为最大灰度级数.

一般而言,根据像素的矩形阵列,像素间方向两两互补角度为  $0^\circ,45^\circ,90^\circ,135^\circ$ ,故  $\theta$ 可以取  $0^\circ,45^\circ,90^\circ,135^\circ$ . 在  $6 \times 6$  的图像块中,图 3 是原图灰度,图 4 是相应的灰度共生矩.

A					
0	1	2	3	0	1
1	2	3	0	1	2
2	3	0	1	2	3
3	0	1	2	3	0
0	1	2	3	0	1
1	2	3	0	1	2

B					
0	0	0	0	1	1
0	0	0	0	1	1
0	0	0	0	1	1
0	0	0	0	1	1
2	2	2	2	3	3
2	2	2	2	3	3

Fig.3 Source images A,B

图 3 原图 A、原图 B

	0	1	2	3
0	$P(0,0)$	$P(0,1)$	$P(0,2)$	$P(0,3)$
1	$P(1,0)$	$P(1,1)$	$P(1,2)$	$P(1,3)$
2	$P(2,0)$	$P(2,1)$	$P(2,2)$	$P(2,3)$
3	$P(3,0)$	$P(3,1)$	$P(3,2)$	$P(3,3)$

Fig.4 Gray level co-occurrence matrix

图 4 灰度共生矩结果

由公式(3)可以计算出:当  $d=1$  时, $0^\circ,45^\circ$ 的灰度共生矩阵如图 5 所示.

$$P_A(0^\circ) = \begin{bmatrix} 0 & 8 & 0 & 7 \\ 8 & 0 & 8 & 0 \\ 0 & 8 & 0 & 7 \\ 7 & 0 & 7 & 0 \end{bmatrix} \quad P_B(0^\circ) = \begin{bmatrix} 24 & 4 & 0 & 0 \\ 4 & 8 & 0 & 0 \\ 0 & 0 & 12 & 2 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

$$P_A(45^\circ) = \begin{bmatrix} 12 & 0 & 0 & 0 \\ 0 & 14 & 0 & 0 \\ 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 12 \end{bmatrix} \quad P_B(45^\circ) = \begin{bmatrix} 18 & 3 & 3 & 0 \\ 3 & 6 & 1 & 1 \\ 3 & 1 & 6 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix}$$

Fig.5 Transformed matrix

图 5 变换后矩阵

从结果可以看出:对角线元素全为 0,表明同行灰度变化快;对角线元素较大,表明纹理较粗.灰度共生矩阵反映了图像灰度分布关于方向、邻域和变化幅度的综合信息,但它并不能直接提供区别纹理的特性.

因此,由发散特性,定义提取出描述纹理结构特征的中心发散灰度共生矩阵如下:

$$PJ(i,j|d)=\{(x,y)|f(x,y)=i,f(x+dx,y+dy)=j\} \tag{4}$$

以  $4 \times 4$  方形算子为例,固定  $d$  为 1,在  $4 \times 4$  的窗口内计算中心发散结构局部边缘特征如图 6 所示.

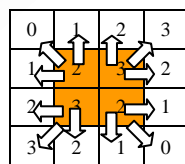


Fig.6 The model of divergent co-occurrence matrix

图 6 发散型共生矩阵模型

对于图 6,由公式(4)可以计算出:当  $d=1$ ,方向取四周灰度像素的中心发散灰度共生矩见表 1.

在中心发散灰度共生矩中,行列元素对应的是当前灰度与下一级灰度的统计概率.可以采用二维转换成一维向量将共生矩用作特征表达,形成一个  $N_g \times N_g$  大小的特征向量.实验过程中发现,这样出来的特征稀疏性较大,维数太高,难以区分数据的主要特征.最终确定以灰度共生矩的边缘概率方式作为纹理结构特征,并在实验中得到很好的效果.对于同样大小的灰度图像,生成的特征向量大小为  $2 \times N_g$ ,维数得到降低,同时,其体现了不同灰度级共生矩的统计特性.

表 1 边缘概率结果见表 2.

**Table 1** Center divergent matrix

**表 1** 中心发散共生矩

	0	1	2	3
0	0	0	2	0
1	0	0	4	0
2	0	0	0	4
3	0	0	0	2

**Table 2** With edge probability

**表 2** 边缘概率结果

	0	1	2	3	e
0	0	0	2	0	2
1	0	0	4	0	4
2	0	0	0	4	4
3	0	0	0	2	2
f	0	0	6	6	

为方便计算,其纹理结构可以由边缘概率的特征向量(2,4,4,2,0,0,6,6)来表示.

### 1.1.3 显著性纹理结构算子

结合显著性因子和中心发散灰度共生矩,构造 STS 的二维直方图,这个二维直方图将 STS 的两个组成部分显著因子和纹理结构特征联系到了一起.

直方图统计公式见公式(5):

$$STh[i]=\{(x,y)|E(x,y)=i\}, 0 \leq i \leq 15 \tag{5}$$

其中,  $(x,y)$  为像素坐标;  $E(x,y)$  为对角二值模式值,即对角四邻域像素与灰度阈值进行比较得到的二值编码.

首先,使用公式(2)计算每个局部区域的显著因子  $\zeta(x_c)$ ;同时,用公式(4)计算纹理结构  $PJ$ ,从而可以得到两幅新的图像——显著因子图像和纹理结构图像,根据这两幅图像计算二维直方图(如图 7 所示).换句话说,在二维直方图里,每一列对应的是一个纹理分支,每一行对应的是一个显著区间.因此,每一小块对应的值是一个确定的显著区间和纹理分支的频率.这样,通过构造二维直方图,将显著性因子和纹理结构特征联系到了一起,得到显著性纹理结构特征描述算子.该算子更精确了图像关键纹理信息的表达,很好地处理了图像分类问题中的特征区分力,并在行人检测、人脸识别和场景分类中起到良好的图像表达作用.

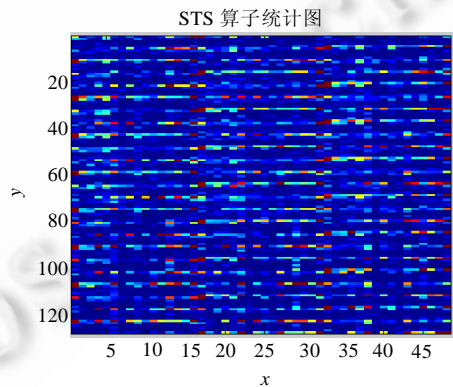


Fig.7 STS 2D histogram

图 7 STS 二维直方图

### 1.2 算子性能分析

提出的 STS 描述算子是一种基于韦伯定律的描述算子,其有以下几个优点:精准的检测边缘信息、噪声和明暗变化的鲁棒性及强大的表达能力.

- 如图 8 所示,如果采用一般边缘提取算子(图 8(a)),通过一个选定的阈值分割后,可能左边的边缘将会被忽略而只保留下面的那条边缘;而用韦伯定律计算比值的 STS(图 8(b)),任何差值都是与自身做相除运算,所以完好地保留了左边缘和下边缘.这正好可以解决行人边缘阈值选择的难题,体现其在边缘提取过程中的保全性;
- STS 算子对图像中的噪声具有鲁棒性.具体来说,STS 像图像滤波器一样减少噪声的影响.正如图 1 所

示,显著因子的计算是通过周围邻域的差值总和与当前像素值得到的,因此减少了个别噪声像素的影响.此外,采用除法操作也一定程度上减少噪声在图像上的影响;

- STS 在明暗变化上有一定的处理效果:一方面,光线明暗引起的变化相当于整体加减一个常量值,而 STS 计算周围邻域和当前像素的差分,因此平均下来将消除常量变化带来的影响;另一方面,常量倍数变化带来的影响又通过相除减少其差异性.因此,这个描述算子对于光照变化具有鲁棒性.

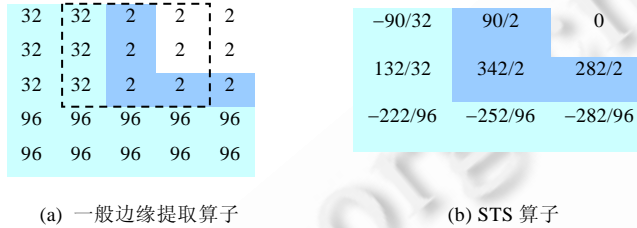


Fig.8 Comparison of different edge detectors

图 8 边缘提取对比分析

1.2.1 STS 对重要信息的保留

STS 算子的优异之处在于模仿人眼视觉注意机制,提取目标对象中显著部分进行编码,从而避免了不必要信息的编码.例如,图 9 是应用在 15 种场景分类<sup>[18]</sup>中的显著因子提取结果.图 9(c)中,柜子和茶几轮廓的显示明显优于图 9(b)中的显示,一般人从两幅图中查找某个目标物体,除了艺术鉴赏外,都会觉得图 9(c)更适合人眼观察特点.在图 9(e)中,山峰已经被多余的纹理覆盖,完全看不出具体东西;而图 9(f)则真实地还原了山峰的轮廓信息.图 9(h)相对图 9(i)包含了更多的天空和云层等内容,道路的主体边缘没有得到体现.

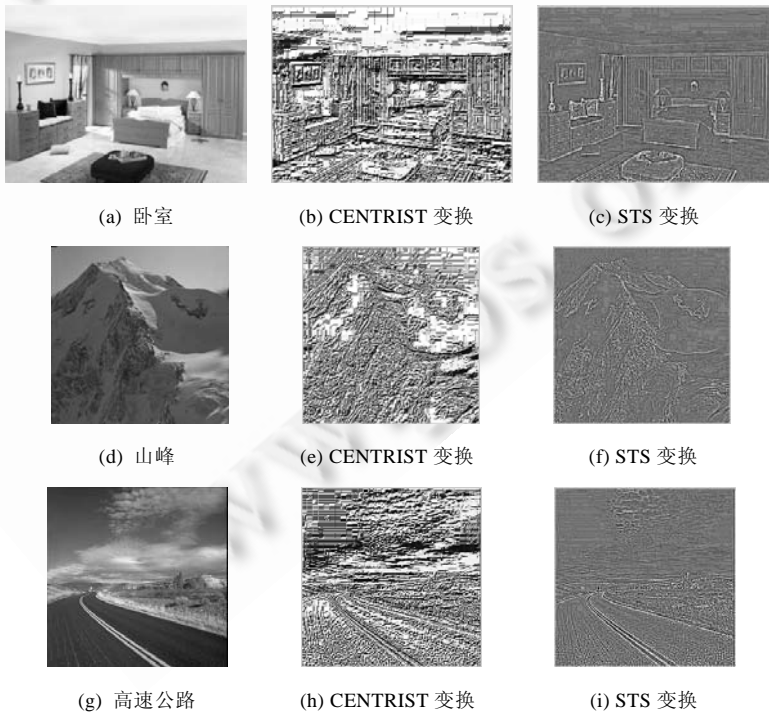


Fig.9 Sample scene images and their CENTRIST and STS transformed results

图 9 15 种场景分类部分变换



1.2.2 STS 对结构的编码

为了理解 STS 为什么能高效捕捉目标结构的原因,进一步分析 STS 值的分布情况和特征向量的组成部分.通过对瑞典叶子结构的数据集<sup>[17]</sup>测试得知,在采用 3×3 的模板下,有 3 种共生矩出现次数最多,如图 10 所示.它们大多类似尖角或弧线边缘结构.通过统计这些潜在局部结构,就可以表征图像语义特征.其中,左侧为 3 种结构及典型特征形式,右侧为实验中出现的用例说明.

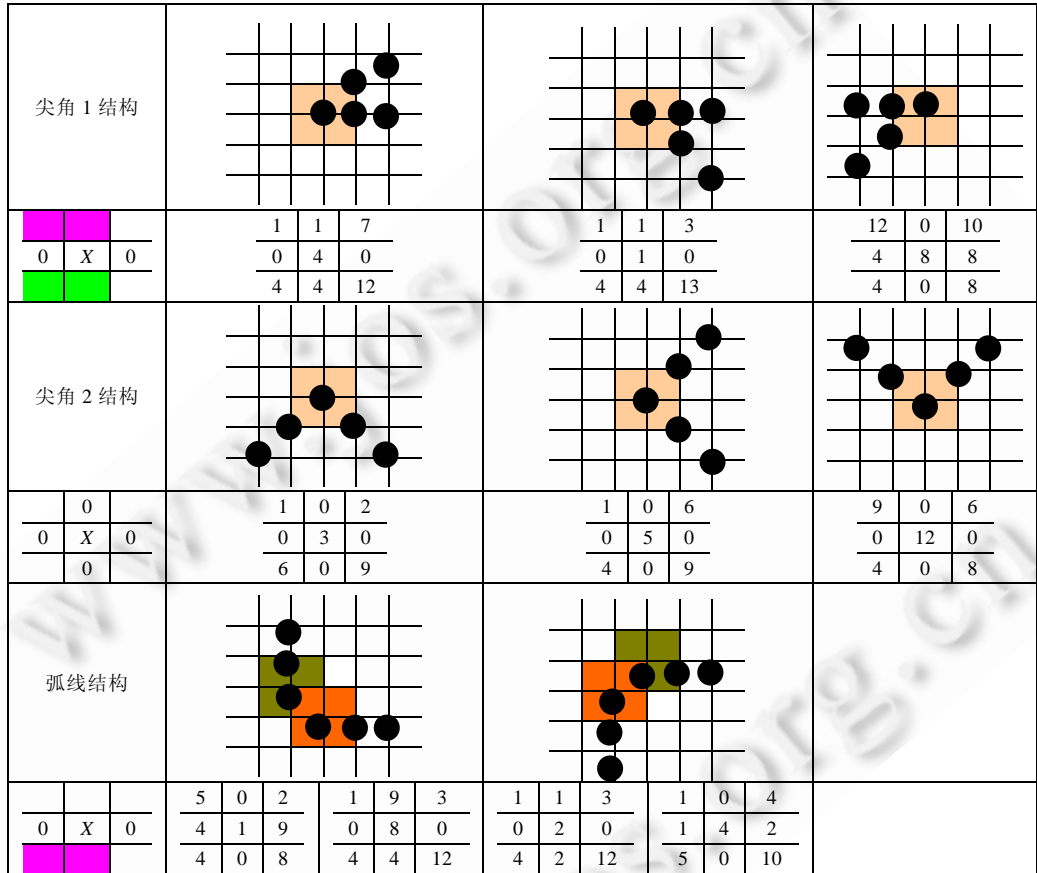


Fig.10 STS transformation analysis

图 10 STS 变换分析

尖角形结构和弧线结构的多次出现,体现了叶子图像多齿和多弧的结构特点,这与我们对树叶的结构认识是一致的.通过比对左侧结构的特点形式,即可以在很大程度上描述目标图像为叶形物体.

1.2.3 计算复杂度分析

在基于局部特征的目标检测分类算法中,HOG 算法是一种比较著名的方法,CENTRIST 算法是最近几年出现的精度高的算法之一,通过分析对比 STS 与 HOG 和 CENTRIST 在单个描述子的构造时间,可以直观地反映三者的计算时间复杂度.

1. 特征值计算

在 HOG 算法中,特征值通过对梯度强度计算得出.由于 HOG 特征值计算要求开方和乘方运算,同时,HOG 特征需要计算梯度方向,且要根据方向的值在直方图中进行线性插值,所以必须计算反正切函数.这样,导致 HOG 特征的计算量非常大、复杂度高,需要耗费大量的计算时间;在 CENTRIST 算法中,特征值只需要加法即可实现;在 STS 算法中,除显著因子的计算需要用到反正切函数进行归一化外,其他的只涉及加法运算.

## 2. 描述子构造

本文采用实验对比 3 种算法的描述子构造时间.对于实验样本,先将所有图像归一化到  $32 \times 32$  像素,每个样本将被划分为 49 个  $8 \times 8$  像素的块.对于 HOG 而言,每个块分为 4 个方格,每个方格统计一个梯度方向直方图,直方图将  $0^\circ \sim 360^\circ$  划分为 8 个区间.在 CENTRIST 算法中,需将特征值先进行统计,再分配到  $4 \times 4 \times 16$  的区域. STS 算法与 CENTRIST 类似.为了体现比较实验的客观性,对同一组图像进行实验,消耗时间为描述算子完成一次运算的平均时间.表 3 为 STS, HOG 和 CENTRIST 单个描述子计算时间的对比,从结果可以看出, STS 在描述子计算时间上比 HOG 有明显的提高,比 CENTRIST 略低,但基本能满足应用的时间复杂度要求.

**Table 3** Computation time of different descriptors

表 3 描述子计算时间比较

算法	描述子计算时间(ns)
HOG	18.058
CENTRIST	0.724
STS	5.468

### 1.3 STS与CENTRIST和HOG的相似度分析

对于分类的评估方法而言,目标在于选择一种描述目标特征的向量,该向量的特性应该与相同类型目标具有相似性质,同时与相异型目标之间差异性最大.对于样本  $X$ ,我们将计算它与各类样本的相似程度. $x_{in}$  表示在  $X$  同类目标中与  $X$  最相近的样本向量;同理, $x_{out}$  表示其他类中与  $X$  最相近的样本向量.

显然,我们希望  $Snn = S(x, x_{out}) - S(x, x_{in})$  结果为正值,且越大越好.其中,  $S(x, y)$  表示  $x$  与  $y$  之间的相似度.正的结果依据最近邻法则(1-NN)表示正确分类.因此,  $Snn$  参数可以很好的评价描述算子的性能.

将 STS 与典型的检测算子 CENTRIST 和 HOG 进行比较,如图 11 所示为在 INRIA 数据集(都是 Sobel 图像)<sup>[7]</sup> 上的性能比较.图中使用了所有 2 416 个正样本和 2 436 个负样本.左图显示的是 STS 和 CENTRIST 的  $Snn$  分布图(直方图).相似度被归一化为  $[0, 1]$  区间,负  $Snn$  (如 0 坐标的左边)表示 1-NN 的错分类. STS 的错分类率为 1.7%,而 CENTRIST 的错分类率为 2.9%,高了 STS 两倍之多.从右图中看出, STS 和 CENTRIST 的  $Snn$  值大部分都为正值,但从幅值上看, STS 的幅值比 CENTRIST 的高很多,所以, STS 相对于 CENTRIST 能够很好地区分物体. HOG 的 1-NN 识别相对于前两种算子错误率更高,几乎达到一半(46%的 1-NN 错误率),但在 1-NN 幅值上与 STS 相近.

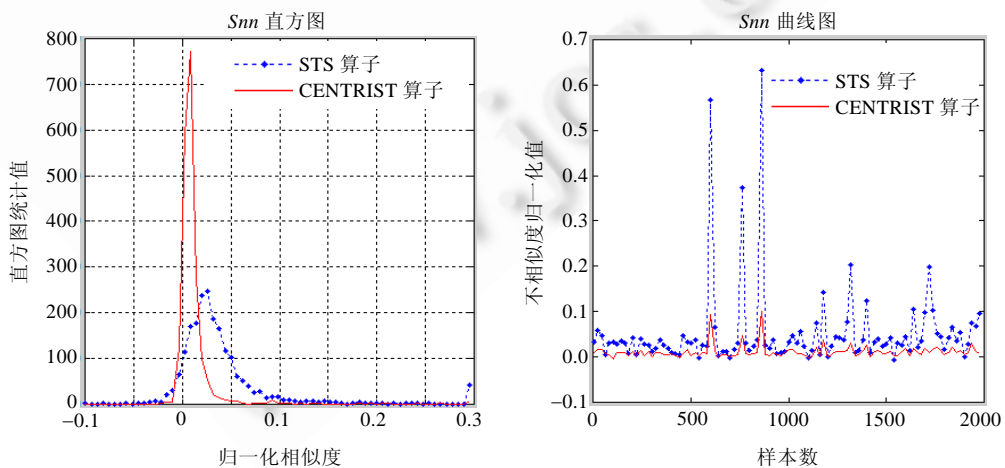


Fig.11 Histogram and curve of similarity score difference

图 11 相似度差值直方图和曲线图



## 2 实 验

显著性纹理结构描述算子在以下 3 个数据集中进行测试:瑞典叶子<sup>[17]</sup>、15 种场景<sup>[18]</sup>和 INRIA 的行人数据库<sup>[7]</sup>.每个数据集随机分离出训练集和测试集,重复 5 次取平均值作为结果.这里仅用到数据中的像素强度值,忽略颜色信息.分别采用文献[14]中的金字塔分裂统计和文献[5]中的间隔统计方式,二级降维金字塔式特征向量将变成  $40 \times (25+5+1)$  维.间隔式中根据目标的先验知识,分别采用  $40 \times 120$  检测窗口大小,将窗口分成  $5 \times 10$  个基本块(每个基本块含有 96 个像素).将任何相邻的  $2 \times 2$  个基本块组成一个超级块并从超级块中提取一个 STS 描述符.这样一共有  $9 \times 4 = 36$  个超级块,所以,特征向量维数为  $40 \times 36 = 1440$  维.因为计算时采用  $3 \times 3$  区域算子,所以每个超级块边缘第 1 个像素不包含在内.

### 2.1 瑞典叶子结构实验

瑞典叶子数据<sup>[17]</sup>收集了来自瑞典的 15 种不同的叶子,每个叶子含有 75 个样本数据.选取 25 张图像作为训练集,剩下的 50 张作为测试集.这个数据集实验的主要目的是评估描述算子的结构表达能力.

每个叶子图像数据只给出了 128 个点的坐标信息,连接各点获得叶子轮廓图,图 12 给出了 STS 算子和 SENTRIST 算子的对比.在图 12(a)和图 12(b)中,左边为 CENTRIST 变换图,右边为 STS 变换图.可以看出,STS 变换的轮廓图具有线条更粗实和辨识度更清晰的优点,能够很好地表达边缘轮廓结构.

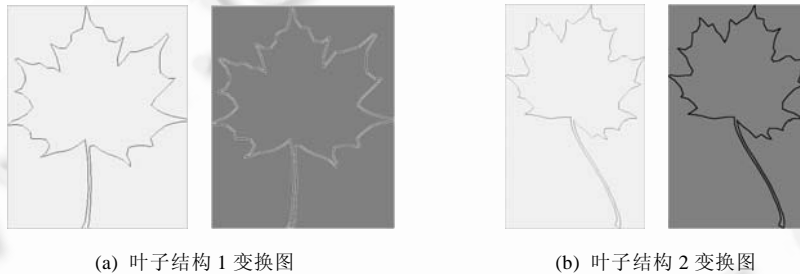


Fig.12 Transformed Swedish leaf images

图 12 叶子结构变换对比图

### 2.2 场景识别实验

15 类场景分类数据是由 Oliva 和 Torralba(8 种)<sup>[19]</sup>、Fei-Fei 和 Perona(13 种)<sup>[20]</sup>、还有 Lazebnik 等人(15 种)<sup>[18]</sup> 提供,场景包括了办公室、商店、海边等.图像分辨率为  $300 \times 250$  左右,每类大约包含 210~410 张图片.数据涵盖了室内室外很宽泛的内容.该实验主要检验 STS 算子在噪声和明暗变化情况下的边缘信息检测能力.

图 9 给出了两种算子对场景的变换效果图.由图 9 可知,CENTRIST 对场景的结构和纹理都有明显的表达,STS 算子更注重显著差异的部分,如房屋基本结构框架、家具的典型轮廓的描述,不注重细节纹理的刻画.这些不同光照场景的实验,验证了 STS 算子具有鲁棒地显著边缘检测能力,同时有利于降维和剔除冗余.

### 2.3 行人检测实验

采用 INRIA 的行人数据库,部分行人检测样本如图 13(a)所示.该数据库是目前静态图像的行人检测研究最具挑战性的测试集,包括各种不同光照、行走姿态、穿着和视角等变化的行人.实验中的训练集和测试集所采用的样本数见表 4.实验结果采用召回率(公式(6))和准确率(公式(7))来评价.

$$\text{召回率} = \frac{\text{正确检测的行人样本数}}{\text{总的行人样本数}} \times 100\% \quad (6)$$

$$\text{准确率} = \frac{\text{正确检测的行人样本数}}{\text{正确检测的行人样本数} + \text{误判的行人样本数}} \times 100\% \quad (7)$$

**Table 4** Number of training and testing sets

**表 4** 实验训练样本数和测试样本数

	正样本数	负样本数
训练集	2 416	22 111
测试集	1 126	15 264

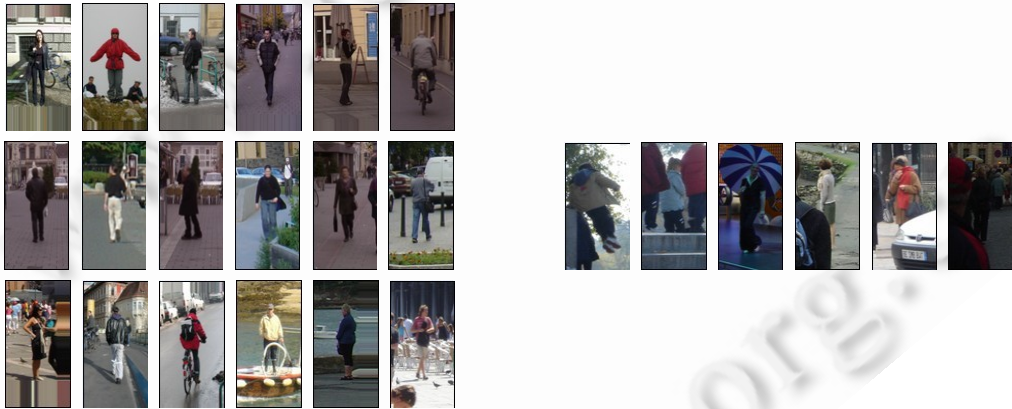
表 5 列出了利用 HOG,CENTRIST 和 STS 这 3 种特征在 INRIA 上的实验结果.从召回率来看,CENTRIST 和 STS 均好于 HOG,CENTRIST 与 STS 二者的性能相当;从准确率来看,STS 均好于 CENTRIST 和 HOG.

**Table 5** Performance comparison of STS, CENTRIST and HOG

**表 5** STS,CENTRIST 和 HOG 的性能比较

	HOG (%)	CENTRIST (%)	STS (%)
召回率	89.3	93.8	94.2
准确率	90.7	91.5	93.6

部分漏检行人样本如图 13(b)所示,当光线较暗、背景复杂、衣着纹理复杂或部分遮挡时,这些因素使得图像容易产生大量的虚假信息,所以容易漏检.采用 STS 算子的行人检测算法各部分耗时分布见表 6.从表 6 也可以看出,采用 STS 算子的行人检测算法还有较大的改进空间,如优化加速.



(a) 部分训练样本集

(b) 部分漏检测试样本

**Fig.13** Sample data set for pedestrian detection

图 13 行人检测数据集

**Table 6** Computation time in different parts

**表 6** 计算时间分布

处理模块	用时百分比(%)
Sobel 梯度算子	16.55
STS 算子计算	9.36
图像积分	44.65
图像尺度变化	5.68
扫描	23.75
后处理	0.02

通过上述数据集的实验比较得知,STS 具有良好的结构表达、显著的边缘检测和纹理区分功能,更注重人眼观察的显著性信息,在非刚性物体的检测与识别上有其特有的优势,适合于以轮廓特征为标志的行人检测.

### 3 车载环境行人检测应用

我们将 STS 描述算子应用到车载环境的行人检测,行人检测的系统结构如图 14 所示.检测采用滑动窗口重选搜索.由于滑动窗口相对独立性,故采用级联方法加快检测时间.对于窗口搜索过程中,首先快速排除图像纹理一致及不符合车载环境下平面假设条件的窗口,减少分类器的运行比例,将更少的候选窗口留给计算复杂度高的分类运算,从整体上改观系统性能.

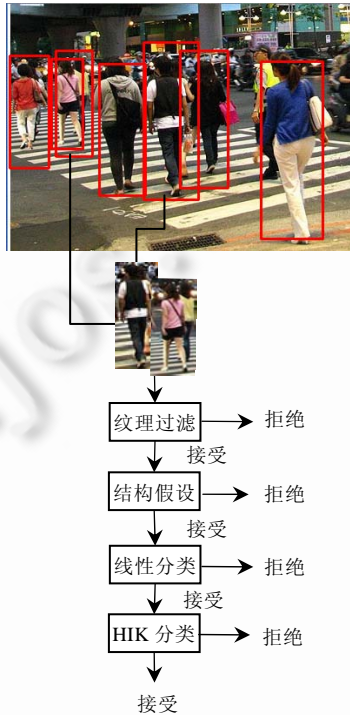


Fig.14 Pedestrian detection framework

图 14 行人检测框架图

#### (1) 纹理协方差过滤

图像纹理的一致性可以用图像协方差来表示.如图 15 所示,区域 1 的纹理一致性低于包含目标图像块的纹理一致性,因此,该区域可以被认定为背景信息.矩形区域 3 包含目标图像块.我们采用图像积分的方式计算子窗口内图像的协方差值来对图像一致性进行计算,排除掉协方差值高于某个阈值  $\sigma_{\min}^2$  的区域,以加快非目标对象的过滤.但这种方法无法过滤结构有序的包含非对象的图像块,如图 15 中矩形区域 2.

计算图像方差方法见公式(8):

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad (8)$$

推导出公式(9):

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \mu^2 \quad (9)$$

其中,  $x_i$  表示图像中的第  $i$  个像素,  $n$  表示像素的个数,  $\mu$  定义如下:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (10)$$

对于大小为  $n$  的图像块,上面公式计算  $\sigma^2$  需要  $n$  大小的查找表内存.为充分利用图像的重叠区域,可以共享内存存储空间,图像  $I$  通过积分变换后  $I'$  只需要 8 个内存表. $I'$  积分图中的值  $(x,y)$  表示点  $(1,1)$  到  $(x,y)$  的累加值,公式如下:

$$I'(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y') \quad (11)$$

$B$  矩阵内的计算即可以用公式(12)表示:

$$\sum_{i=1}^n x_i = I'(B) \quad (12)$$

将等式右边改成平方数的累加,得到:

$$I''(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y')^2 \quad (13)$$

同理得到:

$$\sum_{i=1}^n x_i^2 = I''(B) \quad (14)$$

联合上述公式,可以得到:

$$\sigma^2 = \frac{1}{n} I''(B) - \left[ \frac{1}{n} I'(B) \right]^2 \quad (15)$$

对于  $\sigma_{\min}^2$  的计算,我们用初始块的方差的一半.



Fig.15 Analysis of texture covariance

图 15 纹理协方差分析

## (2) 道路结构假设

根据道路本身的特性,如消失点、道路方向,可以估计行人可能的位置和精确尺度以进行有效遍历.

如图 16 所示,寻找某类目标,在没有考虑场景三维结构的情况下,所有行人可能位置和尺度如图 16(b)所示.如果能够估算出道路平面的方向,利用这个信息可以修正寻找位置<sup>[21]</sup>,如图 16(d)所示.同理,如果能估算出摄像头的视角线,便可以得到图 16(f)所示区域.结合上述几种线索,最终得到图 16(g).

## (3) 分类器的选择

分类器采用线性 SVM 分类器,该线性 SVM 分类器虽然速度很快,但准确率大大降低,故结合文献[22]中提出的更适合于直方图统计类型数据的 HIK SVM.两类分类器进行级联得到最终分类器.

## (4) 检测结果

实验视频利用我们的无人驾驶汽车实验平台的车载摄像头采集于长沙市的城市道路.行人检测算法在笔记本上实现,部分序列检测结果如图 17 所示.

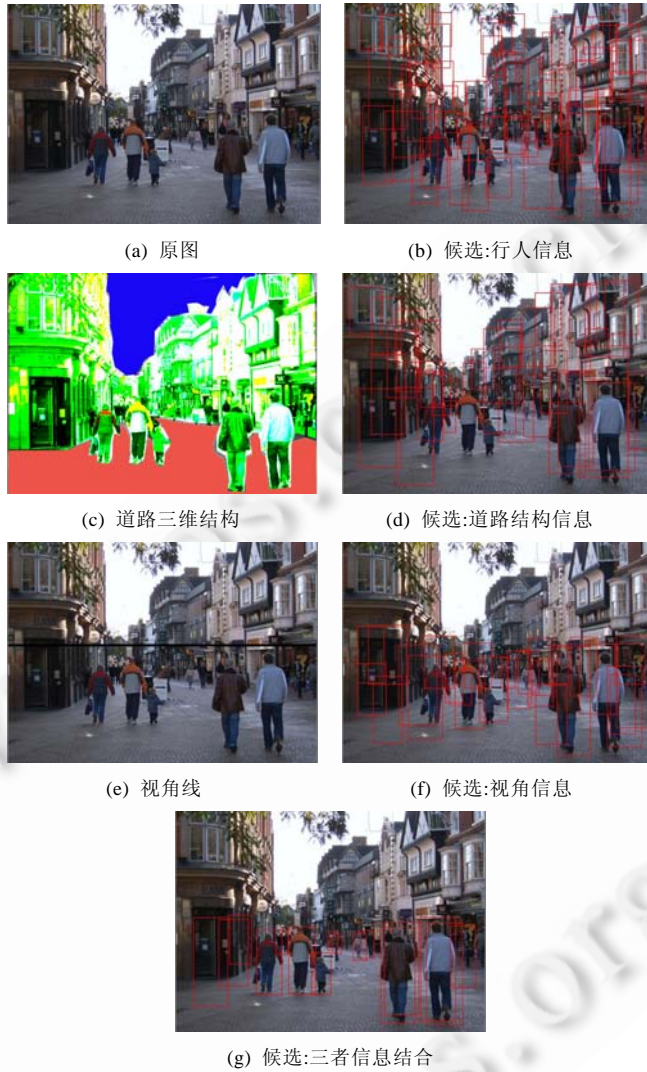


Fig.16 Road structure and pedestrian detection strategy

图 16 道路结构与检测策略

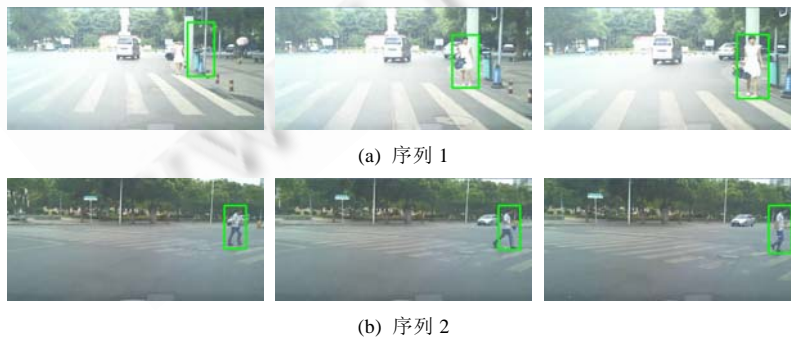


Fig.17 Pedestrian detection results in a real scene

图 17 城市道路的行人检测结果



## 4 结束语

行人检测在计算机视觉领域的许多应用中起着至关重要的作用.图像的行人检测方法的一个核心问题是如何有效表示行人的整体特征、部位特征或者局部块特征.方向梯度直方图特征是目前广泛使用的行人特征表示,但是方向梯度无法刻画人眼视觉敏感度,信息冗余度大.本文提出了一种融合视觉显著性和灰度共生矩的纹理特征提取方法,采用显著性因子与灰度共生矩二维统计量描述纹理,在保留图像梯度信息的同时,减少非显著成分的冗余编码,提高了纹理描述能力,在行人检测中得到很好的效果,优于 CENTRIST 和 HOG.尽管本文提出的特征提取方法主要是针对行人检测,但是瑞典叶子结构、典型场景识别实验验证 STS 具有较强的结构表达能力、边缘信息准确提取能力,在非刚性物体的检测与识别上有其特有的优势.运用 STS 进行真实城市环境的车载行人检测取得了不错的效果,只是算法的实现还是基于传统计算机.下一步的研究是改进 STS,将其运用到多目标组合分类以及嵌入式平台.

## References:

- [1] Papageorgiou C, Poggio T. A trainable system for object detection. *Int'l Journal of Computer Vision*, 2000,38(1):15–33. [doi: 10.1023/A:1008162616689]
- [2] Viola P, Jones MJ, Snow D. Detecting pedestrians using patterns of motion and appearance. In: *Proc. of the Int'l Conf. on Computer Vision*. 2003. 734–741. [doi: 10.1109/ICCV.2003.1238422]
- [3] Jones MJ, Snow D. Pedestrian detection using boosted features over many frames. In: *Proc. of the IEEE Conf. Computer Vision and Pattern Recognition*. 2008. 1–4. [doi: 10.1109/ICPR.2008.4761703]
- [4] Levi K, Weiss Y. Learning object detection from a small number of examples: The importance of good features. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2004. 53–60. [doi: 10.1109/CVPR.2004.1315144]
- [5] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2005. 886–893. [doi: 10.1109/CVPR.2005.177]
- [6] Tuzel O, Porikli F, Meer P. Pedestrian detection via classification on riemannian manifolds. *IEEE Trans. on PAMI*, 2008,30(10): 1713–1727. [doi: 10.1109/TPAMI.2008.75]
- [7] Dollar P, Tu Z, Perona P, Belongie S. Integral channel features. In: *Proc. of the British Machine Vision Conf*. 2009. 1–11.
- [8] Gao W, Ai H, Lao S. Adaptive contour features in oriented granular space for human detection and segmentation. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2009. 1786–1793. [doi: 10.1109/CVPR.2009.5206762]
- [9] Liu YZ, Shan SG, Zhang WC, Chen XL, Gao W. Granularity-Tunable gradients partition (GGP) descriptors for human detection. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2009. 1255–1262. [doi: 10.1109/CVPR.2009.5206724]
- [10] Chen J, Shan SG, He C, Zhao GY. WLD: A robust local image descriptor. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2010, 32(9):1705–1720. [doi: 10.1109/TPAMI.2009.155]
- [11] Mohan A, Papageorgiou C, Poggio T. Example-Based object detection in images by components. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2001,23(4):349–361 [doi: 10.1109/34.917571]
- [12] Wu B, Nevatia R, Li Y. Segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses. *Int'l Journal of Computer Vision*, 2009,82:185–204. [doi: 10.1007/s11263-008-0194-9]
- [13] Lowe DG. Distinctive image features from scale-invariant keypoints. *Int'l Journal of Computer Vision*, 2004,60(2):91–110. [doi: 10.1023/B:VISI.0000029664.99615.94]
- [14] Wu JX, Reh JM. CENTRIST: A visual descriptor for scene categorization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2011,33(8):1489–1501. [doi: 10.1109/TPAMI.2010.224]
- [15] Shen J. Weber's law and weberized TV restoration. *Elsvier Physica*, 2003,175:241–251.
- [16] Haralick RM, Shanmugam K, Dinstein IH. Texture features for image classification. *IEEE Trans. on Systems, Man, and Cybernetics*, 1973,3(6):610–621. [doi: 10.1109/TSMC.1973.4309314]
- [17] Söderkvist OJO. Computer vision classification of leaves from Swedish trees [MS. Thesis]. Linköping University, 2001.
- [18] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2006. 2169–2178. [doi: 10.1109/CVPR.2006.68]



- [19] Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int'l Journal of Computer Vision*, 2001,42(3):145-175. [doi: 10.1023/A:1011139631724]
- [20] Li FF, Perona P. A Bayesian hierarchical model for learning natural scene categories. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2005. 524-531. [doi: 10.1109/CVPR.2005.16]
- [21] Hoiem D, Efros AA, Hebert M. Putting objects in perspective. *Int'l Journal of Computer Vision*, 2008,80(1):3-15. [doi: 10.1007/s11263-008-0137-5]
- [22] Wu J. A fast dual method for HIK SVM learning. In: *Proc. of the European Conf. on Computer Vision*. LNCS 6312, 2010. 552-565. [doi: 10.1007/978-3-642-15552-9\_40]



肖德贵(1972-),男,湖南南县人,博士,副教授,CCF 会员,主要研究领域为智能视觉计算,嵌入式与网络计算,可视化.  
E-mail: dgxiao@hnu.edu.cn



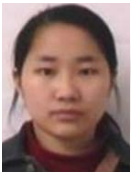
朱欢(1987-),女,硕士,主要研究领域为图像处理,模式识别.  
E-mail: 657393021@qq.com



辛晨(1987-),男,硕士,主要研究领域为机器视觉,模式识别.  
E-mail: 1532740822@qq.com



李小乐(1988-),男,硕士,主要研究领域机器视觉,模式识别.  
E-mail: 173497449@qq.com



张婷(1988-),女,硕士,主要研究领域为机器视觉,模式识别.  
E-mail: 793684129@qq.com