DOI:10.11992/tis.201604011
网络出版地址:http://kns.cnki.net/kcms/detail/23.1538.TP.20170630.2115.006.html

视觉感知式场景文字检测定位方法

吕国宁[1],高敏[2]
(1.郑州师范学院 网络管理中心,河南 郑州 450044; 2.郑州师范学院 信息科学与技术学院,河南 郑州 450044)

摘 要:针对自然场景中复杂背景干扰检测的问题,本文提出一种基于视觉感知机制的场景文字检测定位方法。人类视觉感知机制通常分为快速并行预注意步骤与慢速串行注意步骤。本文方法基于人类感知机制提出一种场景文字检测定位方法,该方法首先通过两种视觉显著性方法进行预注意步骤,然后利用笔画特征以及文字相互关系实现注意步骤。本文方法在 ICDAR 2013 与场景汉字数据集中均取得较有竞争力的结果,实验表明可以较好地用于复杂背景的自然场景英文和汉字的检测。

关键词:视觉感知;视觉显著性;笔画宽度变换;场景文字;文字检测定位;视觉注意;汉字;英文
中图分类号:TP18;TP39 **文献标志码**:A **文章编号**:1673-4785(2017)04-0563-07

中文引用格式:吕国宁,高敏.视觉感知式场景文字检测定位方法[J]. 智能系统学报, 2017, 12(4): 563-569.
英文引用格式:LYU Guoning, GAO Min. Scene text detection and localization scheme with visual perception mechanism[J]. CAAI transactions on intelligent systems, 2017, 12(4): 563-569.

Scene text detection and localization scheme with visual perception mechanism

LYU Guoning[1], GAO Min[2]
(1.Network Management Center, Zheng Zhou Normal University, Zheng Zhou 450044, China; 2. School of Information Science and Technique, Zheng Zhou Normal University, Zheng Zhou 450044, China)

Abstract: To solve the detection problem with respect to the interference of complex backgrounds in natural scenes, in this paper, we propose a scene text detection and localization scheme based on a visual perception mechanism. The human visual perception mechanism is commonly divided into the fast parallel pre-attention step and the slow serial attention step. In our proposed scheme, we first precedes the pre-attention step with two visual saliency methods and then implement the attention step using a stroke feature and the relationship between characters. Our experimental results show the scheme to be competitive with respect to the ICDAR 2013 and the scene Chinese-character dataset. It is also suitable for English and Chinese character detection of natural scenes under complex background conditions.

Keywords: visual perception; visual saliency; swt; scene text; text detection and localization; visual attention; Chinese text; English text

互联网技术与电子技术的高速发展下,人们逐渐形成以数字图像与视频分享信息交流感情习惯,因此在电子设备与网络中存在着海量的数字图像信息。这些图像信息普遍来自人类生活的自然场景,其中存在着不计其数的关键文字信息。如何有效提取数字图像中的关键文字信息,是有效管理电子设备与网络中的数字图像的重要手段。而有效准确提取数字图像中的关键文字信息是当今一个颇具挑战性的工作,受到研究者的广泛关注。

数字图像中文字的提取根据文字种类分为人工文字和场景文字[1],前者是人们后期添加到图像上的文字,如视频字幕、电影中的说明文字及比赛计分牌等,后者是自然场景中真实存在并通过数字成像设备保存在数字图像中的文字,如交通标示、街道名称、广告海报以及商店招牌等。场景文字的提取因为没有场景先验知识,且受到场景中周围环境、相机参数及光照因素的影响,因而它比人工文字的提取具有更大难度。

场景图像文字定位算法通常分为两类:基于滑动窗口的方法和基于连通域的方法。文献[2-3]隶属基于滑动窗口的方法,首先使用滑动窗口遍历图像各个尺度,分类器判定每一个滑动窗口区域是否包含文字并给出置信度;然后将各个尺度置信度叠加,得到置信图;最后根据置信图分割得到文字区域。文献[4-5]分别利用笔画与最大极值稳定区域获取连通域作为文字候选区域,然后使用分类器对文字候选区域进行验证(保留文字区域,剔除背景区域),最后将单个文字聚合成文本行。基于滑动

收稿日期:2016-04-07. 网络出版日期:2017-06-30.
基金项目:国家自然基金河南人才培养联合基金项目(U1204703, U1304614).
通信作者:吕国宁.E-mail:sjzmdwxqzz@outlook.com.

窗口的方法因为需要遍历图像各个尺度,故速度较慢,但抗干扰能力稍强于基于连通域的方法;基于连通域的方法速度较快,但容易受到复杂背景干扰。

以上算法各有利弊,但都存在复杂背景干扰造成定位效果不佳的问题,并且两类性能远不如人类自身。本文思路来源于文献[6]。针对该问题,本文尝试参照人类视觉感知机制设计算法。人类视觉感知机制按照如下进行:首先进行快速简单的并行预注意过程,此过程能够快速获得显著性目标,消除复杂背景的影响;然后完成一个较慢的复杂的串行注意过程,有意识地剔除无效显著性目标,突出感兴趣的显著性目标。

参考以上两个步骤,本文方法分为 3 个步骤。首先,本文方法采用颜色通道的对比度显著性算法与谱残差显著性算法获得显著性区域;然后,基于显著性区域运用单极性笔画宽度变换获得文字候选区域;最后,根据文字候选区域自身信息与相互之间信息,利用图模型筛选得到文字区域。第一个步骤对应于人类的快速简单的并行预注意过程,后两个步骤相当于较慢的复杂的串行注意过程。

本文创新点在于利用颜色通道的对比度显著性与谱残差显著性获得显著性区域以减少后续算法的虚警率,并根据显著性算法设计单极性笔画宽度变换。

1 视觉显著性算法

本节结合两种显著性模型获得显著性区域,颜色通道的对比度视觉显著性模型侧重基于颜色的对比度较大的区域,而谱残差显著性模型则偏重于边缘丰富的区域。这两种偏好均符合场景文字的对比度突出和边缘丰富的特点,可以较好互补完成文字显著性区域检测。视觉显著性算法流程图如图 1。

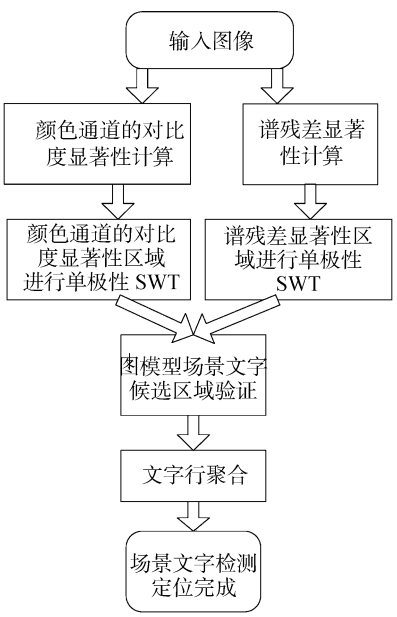


图 1 算法流程图

Fig.1 Algorithm flow chart

1.1 颜色通道的对比度视觉显著性模型

颜色通道的对比度视觉显著性模型是建立在 Opponent Color space 上。式(1)中 L 是 Opponent Color Space 中的亮度分量,RG 是 Opponent Color Space 中红色-绿色分量,BY 是 Opponent Color Space 中蓝色-黄色分量。

$$L = \frac{r + g + b}{3}$$

$$\mathrm{RG} = \frac{r - g}{\max(r,g,b)} \tag{1}$$

$$\mathrm{BY} = \frac{b - \min(r,g)}{\max(r,g,b)}$$

式中:r、g 与 b 代表彩色图像的红色、绿色与蓝色分量。

在以上三通道的基础上,针对每一个通道计算对比度图。对比度图计算方法如式(2)所示是以滑动窗口的方式遍历颜色通道图中每一像素,计算当前像素与周围邻域像素均值的差的绝对值作为相应像素的对比度值。式(2)中 $C(i,j)$ 表示当前颜色通道在位置(i,j)的对比度值,$I(i,j)$是该颜色通道当前位置的强度值,$\bar{I}(i,j)$ 代表该颜色通道当前位置的邻域强度均值。同时,为了增加算法普适性,需要考虑到滑动窗口尺寸问题。

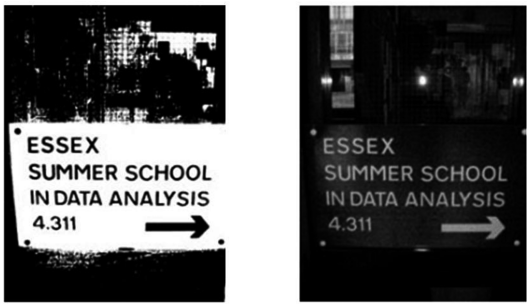
$$C(i,j) = \mathrm{abs}(I(i,j) - \bar{I}(i,j)) \tag{2}$$

$$ws = (1/2^{\sigma}) \times \min(w,h) \tag{3}$$

式中:σ = [4 5 6 7 8] 是滑动窗口的尺度因子,w 与 h 为图像的宽度与高度。

最后,将不同尺寸的滑动窗口下得到的对比度图进行线性叠加并进行归一化得到颜色通道的显著性图。本节选取了红色-绿色通道与蓝色-黄色通道进行对比度显著性计算,并逐像素对二者取几何平均与高斯滤波,如图 2。

(a)原图



(b)红色-绿色通道与蓝色-黄色通道原图

(c)$\sigma=5$

(d)$\sigma=8$

(e)两个尺度结合

(f)两个通道显著性图结合

图 2　颜色通道的显著性效果图

Fig.2　Saliency map of color channel

1.2　谱残差视觉显著性模型

谱残差视觉显著性算法[7]是快速可靠且无需先验知识的显著性算法,它分为 3 步:1)将彩色图像灰度化并进行适当缩放和预处理;2)对前一步产生的灰度图像傅里叶幅度对数谱进行卷积均值滤波;3)从图像傅里叶幅度对数谱中减去上一步的均值滤波结果,最终得到显著性图 S。式(4)描述谱残差视觉显著性模型的求解

$$S = \log(A(I)) - h(I) * \log(A(I)) \qquad (4)$$

式中:$A(I)$ 表示图像的傅里叶幅度谱,$\log(A(I))$ 表示图像的傅里叶幅度对数谱,$h(I)$ 表示均值滤波。

图 3 显示的是利用谱残差视觉显著性模型得到的场景文字显著性图。上面一行图像是场景文字的原图,下面一行图像是对应的谱残差显著性图,图像亮度代表显著性程度。谱残差视觉显著性算法有效检测自然场景中包含文字的边缘丰富区域,但同时也会因为环境中其他边缘丰富的元素产生虚警率。

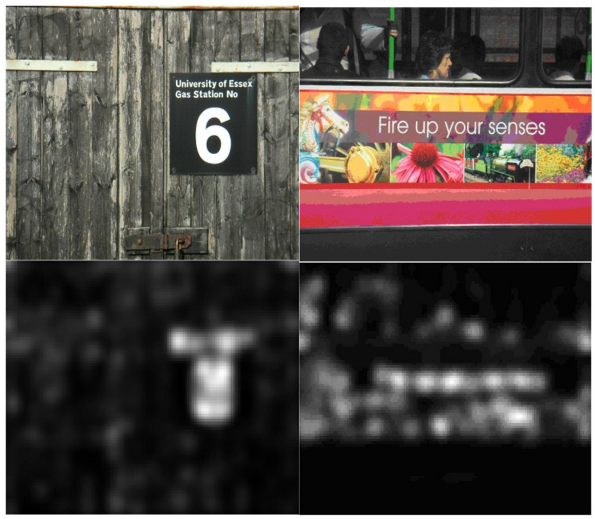


图 3　谱残差显著性效果图

Fig.3　Spectral residue saliency map

1.3　显著性区域

以上两种显著性图的取值范围是介于 0~1 之间,对二者计算显著性图,本质是进行二值化。因此可以使用改进的大津法求取显著图的二值化阈值 T'_s,二值化阈值 T'_s 将显著图分为显著性区域与非显著性区域。

1)首先采用大津法得到阈值 t,然后在训练数据集中设定显著区域中文字召回率的阈值 T_R,初始化系数 α 为 1,以 0.01 为步长递减系数 α,直到首次显著区域中文字召回率 R 首次达到阈值 T_R 即停止,

最终通过式(5)计算得到阈值 T'_s。颜色通道的对比度显著性算法系数为 $\alpha_c = 1$,谱残差显著性算法系数为 $\alpha_s = 0.73$。在得到两种显著性区域后,分别进行数学形态学操作,并填补去除显著性区域中的孔洞。

$$T'_s = \alpha \times t \tag{5}$$

2 单极性笔画宽度变换算法

笔画(Stroke)是图像中相邻的能够形成近似恒定宽度的条带部分[8]。而"笔画宽度"则被定义为近似恒定宽度的条带边缘之间的距离,即图 4 中 p 与 q 像素之间的距离 w。

笔画宽度变换[8](SWT)为数字图像中所有像素计算对应的笔画宽度。此种变换最终结果是笔画宽度图,图中每一像素值是其笔画宽度。

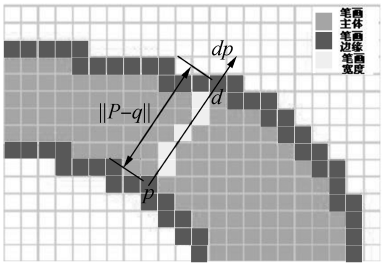
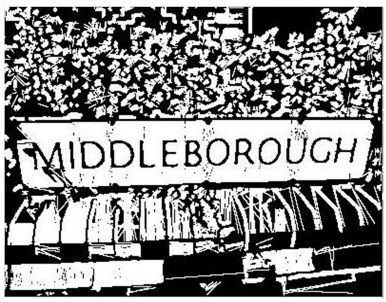


图 4 笔画宽度计算方法图

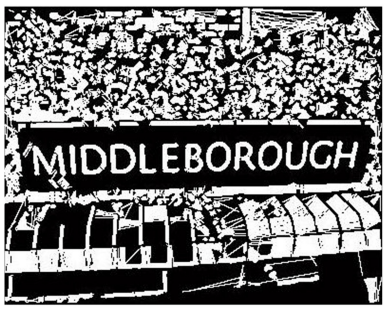
Fig.4 Stroke width map

通常自然场景中的文字存在黑暗背景明亮文字与黑暗文字明亮背景两种极性,因此在无任何先验知识情况下需要沿边缘像素的梯度方向与反梯度方向进行两次 SWT。图 5(b)中 SWT 的方向与场景文字极性不符,图 5(b)中 SWT 的方向与场景文字极性相符。可看出,两次 SWT 固然可以保证自然场景中两种极性的文字不遗漏,但也增加大量非文字区域的虚警。对此,本节基于视觉显著性提出两种极性判断条件,并据此设计单极性 SWT 算法。图 5(d)、(e)是分别对应(b)、(c)的笔画宽度直方图,从中可看出,当极性正确情况下笔画宽度直方图更加集中。

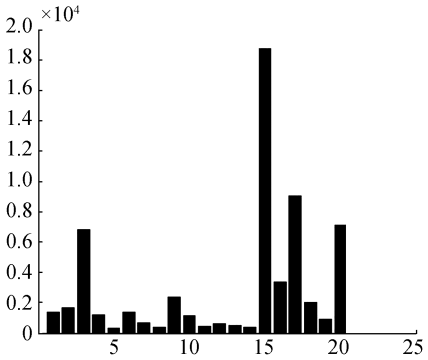
(a)原图



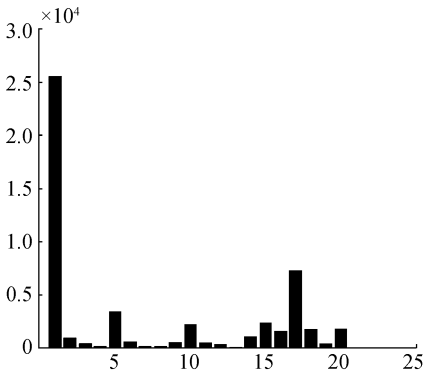
(b)SWT 方向与文字极性不符



(c)SWT 方向与文字极性相符



(d)图(b)对应笔画宽度直方图



(e)图(c)对应笔画宽度直方图

图 5 场景文字极性与笔画宽度直方图关系

Fig.5 The relation between the pole of scene text and stroke width histogram

极性判断条件:

①起始阶段不做极性判断,任意选择一种极性在显著性区域进行 SWT。若其间,任一边缘像素的射线越过显著性区域边界,则此显著性区域为相反极性。

②如果两种极性 SWT 计算中均未发生边缘像素的射线越过显著性区域边界情况,则对该显著性区域两种极性的笔画宽度图求直方图。按照式(6)计算两种极性的笔画宽度直方图的集中度,集中度较大的极性为此显著性区域的极性。

$$f(h) = \frac{1}{N}\sum_{i=2}^{N}(h(i) - h(i-1)) \quad (6)$$

式中:h 代表笔画宽度直方图,N 是划分的 bin 数目,i 代表 bin 的编号。

3　基于图模型的文字候选区域验证与文字行聚合

无向图模型通常被用于图像分割,本节尝试将其表示文字候选区域相互之间的关系,并将文字候选区域使用最大流/最小割方法标注为文字与背景。

在文字候选区域的无向图 $G = \{V,E\}$ 中,顶点 V 是文字候选区域,边缘 E 连接着顶点 V,表示着文字候选区域的相互关系。当文字候选区域满足如式(7)关系则二者相邻。其中 x_i、x_j 分别代表两个文字候选区域的位置,w_i、w_j 分别代表两个文字候选区域的宽度,h_i、h_j 分别代表两个文字候选区域的高度,$\mathrm{dist}(x_i,x_j)$ 分别代表两个文字候选区域的实际距离。

$$\begin{aligned}\mathrm{dist}(x_i,x_j) &< 2\times\min(\max(w_i,h_i),\max(w_j,h_j))\\ &\wedge \min(w_i,w_j)/\max(w_i,w_j) > 0.4\\ &\wedge \min(h_i,h_j)/\max(h_i,h_j) > 0.4\end{aligned} \quad (7)$$

无向图 G 的代价函数如式(8)所示。

$$E(A) = \sum_{p=1}^{P} U_p(A) + \sum_{\{p,q\}\in N} B_{\{p,q\}}(A) \quad (8)$$

式中:U 是一元代价函数,B 是二元代价函数。一元代价函数是使用如表 1 中 5 个特征根据随机森林分类器输出得到。

$$B_{\{p,q\}} = \exp\left(-\frac{0.5\times \mathrm{Dis}_{\mathrm{color}} + 0.5\times \mathrm{Dis}_{\mathrm{stroke}}}{2\times\sigma^2}\right) \quad (9)$$

式(9)是二元代价函数,$\mathrm{Dis}_{\mathrm{col}}$ 与 $\mathrm{Dis}_{\mathrm{stroke}}$ 分别代表两个相邻文字候选区域的颜色差值与笔画宽度差值。

最终,图模型求解即文字候选区域的标注则采用文献[9]的最大流/最小割算法。

在进行文字候选区域验证后,根据文字高度的相似性、笔画宽度的相似性、颜色的相似性与相对位置关系采用启发规则进行文字行的聚合。

表 1　图模型用到的特征

Table1　The feature used in graph model

一元代价函数特征	二元代价函数特征
宽高比 w/h	颜色
占有率 $N_{cc}/(w*h)$	笔画宽度
笔画特征 1strokeWidth/max(w,h)	
笔画特征 2var(strokewidth)/mean(strokewidth)	
边缘强度 $N_{\mathrm{edge}}/(w*h)$	

4　实验与分析

本文实验图像来自 ICDAR 2013 场景文字定位竞赛数据集。ICDAR2013 场景文字定位竞赛数据集是目前英语文字定位算法的主流测试数据集,它取代了 2011 年之前的主流数据集即 ICDAR 2005 场景文字定位竞赛数据集。ICDAR 2013 场景文字定位竞赛数据集包含训练与测试两部分,本文随机森林分类器的训练数据集来自 ICDAR 2013 场景文字定位竞赛数据集的训练集,算法评估则在测试集上完成,结果如表 2。表 2 中的 R 代表召回率,P 代表准确率,F 代表综合性能,评价方法按照竞赛标准[10]。从表 2 可以看出本文算法与竞赛大多数算法相比是具有竞争力的,3 个性能指标(召回率、准确率与综合性能)分别比表 2 中算法第一名的 3 项指标分别高 1.48%、0.45%与 0.82%。

本文同时对自然场景汉字进行了测试,使用的数据集如文献[12]描述,评价标准参照文献[11],实验结果如表 3 所示。如文献[12]是 2012~2013 年间国内研究者算法性能,可以看出本文算法远好于以上两种算法。值得说明,因为国际研究者鲜有公开的受到研究者一致认可的场景汉字数据集,所以可参照的算法与数据集不多。

表 2　ICDAR 2013 文字定位竞赛数据集实验结果

Table2　The result in ICDAR 2013 Task2 dataset　%

方法	算法性能		
	R	*P*	*F*
USTB_TexStar	66.45	88.47	75.90
TextSpotter	64.84	87.51	74.49
CASIA_NLPR	68.24	78.89	73.18
本文算法	67.93	88.92	76.72

表 3 场景汉字数据集实验结果

Table3 The result in Chinese scene text dataset %

方法	算法性能		
	R	*P*	*F*
文献[12]	72	88	76
文献[13]	73	68	71
本文算法	74	89	79

实验在 Intel E7400/2G RAM,MATLAB 混合编程情况下完成,实验中单幅图像均保持长宽比归一化高度为 480,每幅图像平均耗时 1.2 s。场景文字验证阶段的随机森林分类器由 150 棵树组成,采用交叉验证的方法进行训练,轮流用 2/3 训练样本训练和 1/3 样本验证。

图 6 是本文方法效果图,可以看出本文方法取得不错效果,较好排除背景干扰,有效检测定位图像中的场景英文和场景汉字。本文方法是对英文与汉字同时有效。

图 6 算法效果图

Fig.6 Algorithm result

5 结论与展望

本文提出一种视觉感知式场景文字检测定位方法。该方法首先利用颜色通道的对比度显著性与谱残差显著性获得显著性区域,然后在显著兴趣区域中采用单极性笔画宽度变换得到文字候选区域,最后再根据文字候选区域自身信息与相互之间信息基于图模型筛选得到文字区域。第 1 个步骤对应于视觉感知机制的预注意过程,后两个步骤对应于视觉感知机制的注意过程。实验表明,本文方法在 ICDAR 2013 与 ICDAR 2005 竞赛数据集中取得较有竞争力的结果。本文创新点在于利用颜色通道的对比度显著性与谱残差显著性获得显著性区域以减少后续算法的虚警率,并根据显著性算法设计单极性笔画宽度变换。

参考文献:

[1]JUNG K, KIM K I, JAIN A K. Text information extraction in images and video: a survey[J]. Pattern recognition, 2004, 37(5): 977-997.

[2]BAI Bo, YIN Fei, LIU Chenglin. Scene text localization using gradient local correlation[C]//International Conference on Document Analysis and Recognition, Washington DC, 2013: 1412-1416.

[3]姜维, 卢朝阳, 李静, 等. 针对场景文字的基于视觉显著性和提升框架的背景抑制方法[J]. 电子与信息学报, 2014, 36(3): 617-623.
JIANG Wei, LU Zhaoyang, LI Jing, et al. Visual saliency and boosting based background suppression for scene text [J]. Journal of electronics & information technology, 2014, 36(3): 617-623.

[4]CONG Yao, et al. Detecting texts of arbitrary orientations in natural images[C]//IEEE Conference on Computer Vision and Pattern Recognition, Providence. 2012: 1083-1090.

[5]LI Yao, JIA Wenjing, SHEN Chunhua, et al. Characterness: an indicator of text in the wild[J]. IEEE transactions on image processing, 2014, 23(4): 1666-1677.

[6]赵春晖, 王佳, 王玉磊. 采用背景抑制和自适应阈值分

割的高光谱异常目标检测[J]. 哈尔滨工程大学学报, 2016, 37(2): 278-283.
ZHAO Chunhui, WANG Jia, WANG Yulei. Hyperspectral anomaly detection based on background suppression and adaptive threshold segmentation[J]. Journal of Harbin engineering university, 2016, 37(2): 278-283.[7] HOU X D, ZHANG L Q. Saliency detection: a spectral residual approach[C]//IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, 2007: 1-8.

[8] EPSHTEIN B, OFEK E, WEXLER Y. Detecting text in natural scenes with stroke width transform[C]//IEEE International Conference on Computer Vision and Pattern Recognition. San Francisco, 2010: 2963-2970.

[9] BOYKOV Y, KOLMOGOROV V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision[J]. IEEE transaction pattern analysis and machine intelligence, 2004, 26(9): 1124-1137.

[10] KARATZAS D, SHAFAIT F, UCHIDA S, et al. ICDAR 2013 Robust Reading Competition[C]//IEEE International Conference on Document Analysis and Recognition. Washington DC, 2013: 1484-1493.

[11] LUCAS S M. ICDAR 2005 text locating competition results [C]//8th International Conference on Document Analysis and Recognition. 2005: 80-84.

[12] 姜维,卢朝阳,李静,等. 基于角点类别特征和边缘幅值方向梯度直方图统计特征的复杂场景文字定位算法[J]. 吉林大学学报: 工学版, 2013, 43(1): 250-255.
JIANG Wei, LU Zhaoyang, LI Jing, et al. Text localization algorithm in complex scene based on corner-type feature and histogram of oriented gradients of edge magnitude statistical feature[J]. Journal of Jilin University: engineering and technology edition, 2013, 43(1): 250-255.

作者简介:

吕国宁,男,1981 年生,讲师,主要研究方向为人工智能和大数据。

2017 机器人及机电一体化国际会议 (ICRoM 2017)
2017 the International Conference on Robotics and Mechantronics (ICRoM 2017)

2017 the International Conference on Robotics and Mechantronics (ICRoM 2017) will be held during December 12-14, 2017, Hongkong.

Topics of interest include all aspects , but not limited to:

Mechatronics and Robotics
Actuator design, robotic mechanisms and design, robot kinematics and dynamics
Agile Manufacturing
Agriculture, construction, industrial automation, manufacturing process
Automation and control systems, middleware
Biomedical and rehabilitation engineering, welfare robotics and mechatronics
Cellular Manufacturing
Concurrent Engineering
Design for Manufacture and Assembly
Distributed Control Systems
Flexible Manufacturing Systems
FMS Artificial Intelligence
Humanoid robots, service robots
Human-robot interaction, semi - autonomous systems, telerobotics
Information Technology Applied to
Knowledge Based Systems
Lean Manufacturing Logistics
Machine Vision
Management of Technology
Manufacturing Mining robotics Mobile robotics
Modeling and Simulation Scheduling
Nano/micro systems and applications, biological and medical applications
Navigation, localization, manipulation
Operations Management
Rapid Prototype
Rescue, hazardous environments
Robot intelligence and learning
Robot vision and audition
Robots and Automation
Sensor design, sensor fusion, sensor networks
Sensor development Sensors and Applications
Sustainability, energy conservation, ecology
Universal design and services, ubiquitous robots and devices