

异态汉字识别方法研究*

王 恺, 李成学, 王庆人, 赵 宏, 张 健

(南开大学 计算机与控制工程学院, 天津 300071)

通讯作者: 王恺, E-mail: wangk@nankai.edu.cn

摘 要: 复杂图像文字识别是基于内容图像检索的一个重要研究方向. 针对图像中的文字可能存在倾斜、光照不均、噪音干扰和边缘柔化等多种异态问题, 提出一种有效的异态汉字识别方法, 称作 SC-HOG. 首先, 利用稀疏编码得到基向量和稀疏系数, 通过重构图像滤除噪音、处理边缘柔化; 然后, 利用梯度方向直方图抽取复原图像的汉字边缘梯度特征, 削弱倾斜和光照的影响; 最后, 将获取的特征向量送入分类器, 实现异态汉字的识别. 通过合成数据集和真实数据集两方面的实验来验证 SC-HOG 方法的有效性: 前一方面实验结果表明, SC-HOG 方法对于倾斜、光照不均、噪音干扰和边缘柔化等异态情况有较强的鲁棒性; 后一方面实验结果表明, SC-HOG 方法在原生数字图像和场景图像真实样本集上也能取得较好的结果.

关键词: 稀疏编码; 梯度方向直方图; 增量线性决策分析; 汉字识别; 文档图像处理

中图法分类号: TP391

中文引用格式: 王恺, 李成学, 王庆人, 赵宏, 张健. 异态汉字识别方法研究. 软件学报, 2014, 25(10): 2266-2281. <http://www.jos.org.cn/1000-9825/4502.htm>

英文引用格式: Wang K, Li CX, Wang QR, Zhao H, Zhang J. Research on abnormal Chinese character recognition. Ruan Jian Xue Bao/Journal of Software, 2014, 25(10): 2266-2281 (in Chinese). <http://www.jos.org.cn/1000-9825/4502.htm>

Research on Abnormal Chinese Character Recognition

WANG Kai, LI Cheng-Xue, WANG Qing-Ren, ZHAO Hong, ZHANG Jian

(College of Computer and Control Engineering, Nankai University, Tianjin 300071, China)

Corresponding author: WANG Kai, E-mail: wangk@nankai.edu.cn

Abstract: Recognizing characters from the complex image plays an important role in content-based image retrieval and has been well studied in past decades. The methods for normal characters recognition, however, become inapplicable when characters suffer from skew, uneven illumination, noise and anti-aliasing. A new method, named SC-HOG, is proposed in this paper for recognizing abnormal Chinese characters. Firstly, sparse coding is applied on abnormal character image to smooth noises and reduce anti-aliasing. Secondly, HOG features that help reducing the influence of skew and uneven illumination are extracted. Finally, these features are fed into a well-trained classifier to recognize the character of the given image. Experiments on both synthetic and real data sets show that the proposed method, SC-HOG, achieves high accuracy on abnormal Chinese characters recognition.

Key words: sparse coding (SC); histograms of oriented gradients (HOG); incremental linear discriminant analysis (ILDA); Chinese character recognition; document image processing

图像中的文字是图像的一种重要语义信息, 正确定位和识别这些文字, 对基于内容的图像检索及过滤应用有着重要意义.

待提取文字的图像可分为文档图像、场景图像和原生数字图像这 3 大类. 早期的图像文字提取研究集中在

* 基金项目: 国家自然科学基金(61201424); 天津市自然科学基金(12JCYBJC10100); 中央高校基本科研业务费专项资金(65012131)

收稿时间: 2013-06-22; 定稿时间: 2013-09-09

文档图像方面,在这一阶段,研究内容包括预处理(如二值化、倾斜校正、字符分割等)、字符识别、版面分析、图形符号识别^[1],现有 OCR 软件主要是处理文档图像,对于场景图像和原生数字图像处理效果不佳.最近的研究则集中在复杂图像文字提取(包括场景图像文字提取^[2-30]和原生数字图像文字提取^[31-47])方面:场景图像是指用摄像头所拍摄到的图像,场景中的文字(如路牌、广告牌等)也会作为场景的一部分被拍摄到场景图像中,场景图像中的文字可能夹杂在复杂的背景区域中,也可能存在文本倾斜、光照不均等情况;原生数字图像是指用计算机软件生成并一直以数字形式保存的图像,为了便于在网络传输,原生数字图像具有低分辨率的特点,且通常存在压缩损失和严重的边缘柔化问题(视频字幕是在后期视频编辑时利用计算机软件加上去的,因此视频字幕的提取也属于原生数字图像文字提取的范畴)^[31].

通常,复杂图像文字提取系统的框架如图 1 所示,分为文本定位、文字分割和文字识别这 3 个模块.

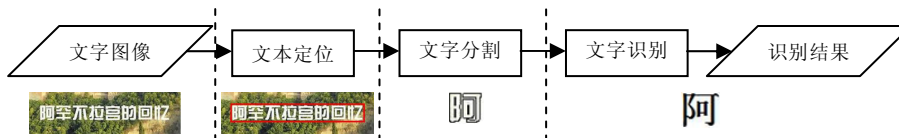


Fig.1 Flow of text extraction from complex image

图 1 复杂图像文字提取系统框架

作为复杂图像文字提取的关键步骤之一,文字识别近年来已取得了一些研究成果.但从以往工作来看,复杂图像中的文字识别研究当前还主要集中在英文方面^[29,30,48-53].在汉字识别领域,我国学者已开展了长期的研究工作,并在印刷体汉字识别、手写体汉字识别(包括联机和脱机)方面取得了丰硕的成果^[54,55].我国自主研发的清华 TH-OCR 与汉王 OCR 在汉字识别方面也已达到很高的性能,能够满足实际应用需求.然而,复杂图像中的汉字识别研究工作还很少,尚处于起步阶段.与传统文档图像中的文字不同,复杂图像中的文字可能会存在倾斜、光照不均、噪音干扰、边缘柔化等多种异态情况.本文的主要工作就是研究异态汉字的识别方法,提升复杂图像汉字提取系统的性能.图 2 是本文研究背景的总结.

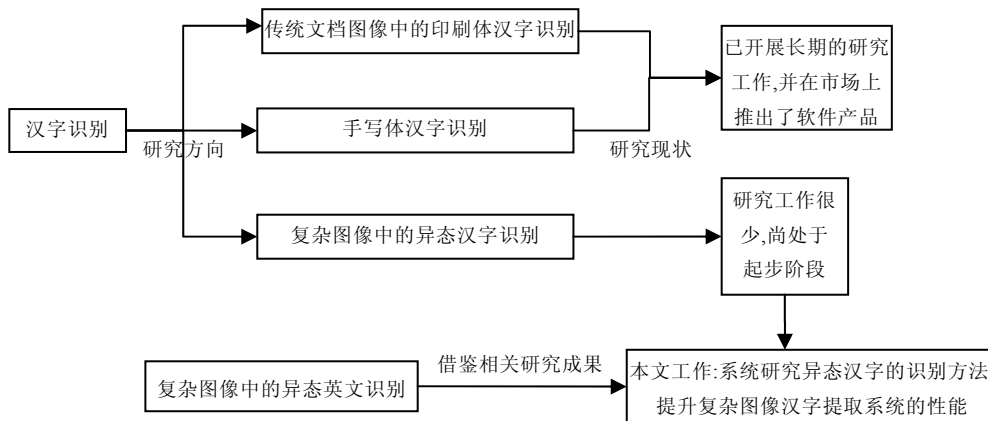


Fig.2 Background summary

图 2 研究背景总结

本文第 1 节介绍复杂图像文字提取方面的相关工作.第 2 节给出用于识别异态汉字的 SC-HOG 方法,分别描述其各个组成部分.第 3 节给出实验结果及分析.第 4 节是本文的总结与下一步工作.

1 复杂图像文字提取方面的工作

关于复杂图像文字提取的工作,国内外学者提出了许多方法,根据研究内容的不同,可以分为 3 大类:

(1) 文本定位方面^[2-24,31-47]

研究如何从图像中得到文本所在的区域,在以往工作中,此类研究最多,根据所采用方法的不同,分为基于连通体的方法、基于纹理的方法和基于角点的方法.

• 基于连通体的方法^[2-15]

先从图像中抽取连通体,再利用连通体本身和连通体之间的特征进行文本连通体/非文本连通体的分类.连通体的抽取是基于连通体方法要解决的一个关键问题,目前主要利用两种信息:

- ◇ 边缘信息^[2-8,44,45]:先使用边界提取算法(如 canny 算子等)获取边界;再使用平滑算法或形态学方法进行边界连接,由此获取完整的文字边界.基于边缘方法的最大制约因素在于:当背景轮廓和文字轮廓相互交叉粘连时,文字和背景就会被混淆在一起进行处理,最终可能被当作背景区域去除,或者得到一个既包含文字又包含背景的区域.因此,需要结合其他的方法进行验证;
- ◇ 颜色信息^[9-13,31,43]:通过对颜色进行聚类(二值化为颜色聚类的一种特殊情况),将颜色相同或相近的像素点聚合在一起形成候选区域,再对候选区域进行连通体分析得到最终结果.基于颜色方法的难点在于对颜色聚类程度的把握:如果聚类不足,可能将背景与文字混合在一起;而聚类过度,又会影响系统性能;同时,如果图像中存在颜色过度的现象,也将影响颜色聚类的效果.

也有学者综合利用边缘信息和颜色信息,以取得更好的性能^[14,15,46,47].

• 基于纹理的方法^[16-21]

将文本区域视为一种特殊的纹理,通过提取相关纹理特征对候选区域进行筛选,得到文本区域.在少数图像中会存在类似于文本的背景区域,因此,如何选择纹理特征来更好地区分文本和背景,是基于纹理方法的难点所在.另外,基于纹理的方法定位精度低.有研究工作将基于纹理的方法与基于连通体的方法相结合^[22,56],先基于纹理定位文本区域,再在候选文本区域中进行连通体分析滤除非文本区域.

• 基于角点的方法^[23,24]

利用一个字符通常包含很多角点这一特点进行文本区域的定位.此类方法的研究尚处于起步阶段,与基于纹理的方法相比,基于角点的方法具有更快的处理速度,但性能略低.

(2) 文字分割方面^[25-28]

在已经定位到文本区域的基础上,文字分割是从文本区域中将文字像素点分离出来的过程,目前主要利用颜色、边缘等信息.Mishra 等人^[25]基于颜色和边缘特征定义能量函数,应用 graph cuts 求解方法进行像素点的分类(文字点或背景点);Cho 等人^[26]先利用 watershed 算法生成超像素,再根据颜色、边缘、笔画宽度以及每个像素点的局部特征建立条件随机场(conditional random field,简称 CRF)模型,将超像素分类为文字或背景;Wang 等人^[27]提出文本的颜色分布满足高斯模型,先提取可靠的笔画作为基准计算高斯模型参数并进行粗分割,再基于颜色进一步过滤噪音;Wakahara 等人^[28]先在 HSI 颜色空间上进行 K 均值聚类,生成多幅二值化子图像,对每幅子图像进行粗分割后,利用网格特征和 SVM 分类器计算该子图像是字符图像的概率,并取概率最高的子图像作为最终结果.

(3) 文字识别方面^[29,30,48-53]

从文字图像提取特征,并根据这些特征识别出该图像所表示的文字信息.目前,复杂图像文字识别主要利用 SIFT,Gabor,HOG 等特征.Zheng 等人^[29]和 Iwamura 等人^[30]利用 SIFT 特征构建了适用于场景图像文字识别的字符识别器;Weinman 等人^[48]通过抽取字符的 Gabor 特征并进行高斯模糊和降采样识别标识牌字符;Campos 等人^[49]评估了提取不同特征识别场景图像字符的效果;Wang 等人^[50]通过提取 HOG 特征识别场景图像中的字符,再利用建立的词典(lexicon)进行后处理识别出整个单词;Newell 等人^[51]通过改进的 HOG 特征算法进一步提高场景图像字符识别的准确率;Wang 等人^[52]在文献[50]的基础上又构建了一个完整的场景图像文字提取系统;Mishra 等人^[53]则采用自底向上与自顶向下相结合的方式构建的场景图像文字提取系统进一步提高了文字提取的准确率;文献[52,53]均利用 HOG 特征进行字符识别.

目前,我国致力于图像文本信息提取方面研究的主要有清华大学电子工程系^[13]、中国科学院自动化研究

所^[22,27,47,57]、中国科学院计算技术研究所^[58]、上海交通大学信息安全工程学院^[29]、华中科技大学计算机科学与技术学院^[42]、复旦大学媒体计算研究所^[59]、浙江大学人工智能研究所^[60]、南京大学计算机软件新技术国家重点实验室^[61]、武汉大学计算机学院^[62]等。

当前,复杂图像文字识别的研究工作还较少,且主要集中在英文识别方面.本文考虑倾斜、光照不均、噪音干扰、边缘柔化等多种异态情况下的汉字识别问题,通过借鉴异态英文字符识别的相关研究成果,提出稀疏编码-梯度方向直方图相结合的 SC-HOG 方法.本文从“效果实验分析(第 3.1 节~第 3.4 节)”与“综合实验结果(第 3.5 节)”两方面验证了本文提出的基于 SC-HOG 的异态汉字识别方法:前一方面,通过在合成数据集上的实验,其结果表明 SC-HOG 方法对于倾斜、光照不均、噪音干扰和边缘柔化等异态情况有较强的鲁棒性;后一方面,通过在真实数据集上的实验,其结果表明 SC-HOG 方法在原生数字图像和场景图像上也能取得较好的结果.

2 SC-HOG 方法的提出

复杂图像中的汉字可能存在倾斜、光照不均、噪音干扰和边缘柔化等异态情况,如图 3 所示,使得适用于常态汉字的识别方法失效.本文提出 SC-HOG 方法,用来处理异态汉字识别问题.实验结果表明:SC-HOG 方法在抗噪音干扰、光照不均和边缘柔化的同时,能够有效地识别适度倾斜的汉字图像.SC-HOG 方法的基本步骤如下所示:首先,利用稀疏编码(sparse coding,简称 SC)得到基向量和稀疏系数,通过重构图像滤除噪音、处理边缘柔化;然后,利用梯度方向直方图(histograms of oriented gradients,简称 HOG)抽取复原图像的汉字边缘梯度特征,削弱倾斜和光照的影响;最后,将获取的特征向量送入分类器,实现异态汉字的识别.

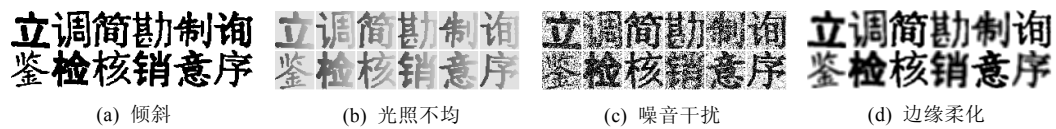


Fig.3 Examples of abnormal Chinese characters

图 3 异态汉字示例

2.1 第 1 步:稀疏编码

Olshausen 等人^[63]发现:稀疏编码提取的基函数与哺乳类动物主视皮层 V1 区简单细胞感受野的响应特性(空间域的局部性、方向性和带通性)类似,自然图像利用 $I(x, y) = \sum_i a_i \varphi_i(x, y)$ (其中, $I(x, y)$ 表示图像, $\varphi_i(x, y)$ 表示决定图像编码的基函数, a_i 表示对应的稀疏系数)得到的编码模型可以建模视觉神经的这种响应特性.为了更好地模拟 V1 区简单细胞感受野,Olshausen 等人^[64]又把超完备基(基函数的个数大于输入数据的维数)引入到稀疏编码中.从数学角度讲,稀疏编码是对图像数据进行线性分解,得到基向量(对应前面提到的基函数)和稀疏系数的过程.

建立稀疏编码模型,对于每个输入图像,只有少量几个基函数有较大的响应输出,能够消除数据之间的高阶相关冗余,因此可用在图像特征提取和滤噪等方面.近年来,稀疏编码在字符识别方面得到了应用:Labusch 等人^[65]首先将手写数字图像分成小块(patch),计算这些块图像稀疏编码的基向量,通过卷积操作得到系数图像,再把系数图像分成规则的无重叠区域,将每个区域的最大和最小系数组成特征向量,最后,利用 SVM 分类器识别手写数字;Barthelem 等人^[66]通过跟踪字符运动轨迹信号,利用稀疏编码算法估计信号的近似值,结合字典学习算法(dictionary learning algorithm,简称 DLA)实现了具有旋转不变性的在线手写字符识别方法;文献[65]和文献[66]都是将稀疏编码用在特征提取上.

文献[67]提出了一种高性能的稀疏编码计算方法用于特征提取,取得了较好的效果.SC-HOG 方法的第 1 步采用文献[67]的稀疏编码算法处理异态汉字,获取图像的基向量和稀疏系数,通过重构图像能够有效地滤除识别汉字图像中的噪音,同时,能够较好地处理边缘柔化情况.与文献[67]的应用不同,本文将稀疏编码用在图像去噪中.

在训练阶段,根据公式(1)所示的目标函数进行优化问题求解.

$$\begin{aligned} \min f(B,S) &= [\text{重构误差}] + \beta [\text{稀疏惩罚因子}] \\ &= \frac{1}{2\sigma^2} \|X - BS\|_F^2 + \beta \sum_{i,j} \phi(s_{ij}) \\ &= \sum_{i=1}^m \frac{1}{2\sigma^2} \|x_i - \sum_{j=1}^n b_j s_{ij}\|^2 + \beta \sum_{i=1}^m \sum_{j=1}^n \phi(s_{ij}) \end{aligned} \quad (1)$$

其中, $X=(x_1, x_2, \dots, x_m) \in R^{k \times m}$ 表示训练集中的 m 幅汉字图像减去各自的平均灰度值后得到的 m 个 k 维向量集合(k 为每幅图像的像素数); $B=(b_1, b_2, \dots, b_n) \in R^{k \times n}$ 表示经稀疏编码计算得到的 n 个 k 维基向量; $S=[s_{ij}]_{n \times m} \in R^{n \times m}$ 表示稀疏系数矩阵,每一列 $s_i=(s_{i1}, s_{i2}, \dots, s_{in})$ 为一幅图像的 n 维稀疏系数向量; β 是一个常量,用来平衡重构误差与系数的稀疏性; $\phi(s_{ij})$ 表示非线性的稀疏惩罚函数,反映了图像编码的稀疏程度,本文使用公式(2)所示的 ϵ - L_1 惩罚函数:

$$\phi(s_{ij}) = (s_{ij}^2 + \epsilon)^{\frac{1}{2}} \quad (2)$$

以公式(1)为优化目标,利用文献[67]的方法,基于整幅汉字图像(而不是分块)求取所有的基向量 B 和对应的稀疏系数 S . 所有的输入图像均匀地分布在基向量上,且所有基向量上稀疏系数的概率分布相同.

利用公式(1)计算得到的基向量 B 可对任意一幅输入图像 I 求取其稀疏系数 S ,再根据 $I_{new}(x,y)=BS+t$ (t 是图像 I 的平均灰度值)进行重构可得到复原图像 I_{new} ,如图 4 所示.可以看出:通过稀疏编码处理,重构后的异态汉字图像与重构后的训练图像表现出了更强的相似性,噪音和边缘柔化问题已在这一阶段得到了很大程度的缓解,从而保证了系统对这两种异态情况的鲁棒性.重构图像的主要差异体现在倾斜和光照方面,这两个问题将在第 2.2 节利用 HOG 来解决.



Fig.4 Chinese characters reconstructed by sparse coding

图 4 稀疏编码处理后的汉字图像

2.2 第二步:梯度方向直方图

HOG 特征由 Dalal 等人^[68]在 2005 年提出,最初用于行人检测,并达到了很好的性能.随后,HOG 特征在人体检测方面得到进一步发展^[69,70].Chuang 等人^[69]在 HOG 特征中加入人的形状信息(轮廓距离、对称性、梯度密度),形成增量 HOG,结合 AdaBoost 算法提高行人检测的准确率.Wang 等人^[70]结合 HOG 和 LBP(local binary pattern)特征集,通过训练 SVM 分类器,利用 Mean-shift 算法在行人检测中有效地处理部分遮挡问题.由于 HOG 特征对光照变化、适当的倾斜和小位移具有鲁棒性,目前也被广泛应用于其他领域.Albiol 等人^[71]采用 EGBM 方法找到脸部具体的标志点,提取这些点处的 HOG 特征向量进行人脸识别.Creusen 等人^[72]在每个颜色通道上提取 HOG 特征,利用迭代技术训练 SVM 分类器,实现对交通标志的识别.贾慧星等人^[73]将目标分块,并提取每块图像的核函数加权 HOG 特征,利用 Mean-shift 算法实现对车辆、行人等多目标的跟踪.Xiao 等人^[74]通过对提取的 HOG 特征进行 MMC(maximum margin criterion)降维,利用最近邻分类器实现植物叶子的分类.

HOG 特征由 SIFT 特征^[75]发展而来,是一种局部描述符.局部梯度或边缘方向的分布能够很好地表征目标的形状和外形.HOG 通过提取局部区域的梯度方向直方图作为特征,无需考虑图像的整体信息.在实现的过程中,先将图像的检测窗口划分为小的单元格(cell),计算单元格内所有像素点的一维梯度或边缘方向直方图,然后

将若干单元格组成较大区域的块(block),在块内进行对比度归一化,通过滑动窗口密集扫描图像,获取最终的HOG 特征向量.

近年来,HOG 特征在文字提取方面的应用,提高了复杂环境下文本定位^[22,57]和文字识别^[50,51]的准确率.然而,复杂图像上存在较多影响文字识别准确率的干扰因素(如倾斜、光照不均、噪音干扰、边缘柔化等),直接在图像上提取特征识别文字的性能,难以达到满意的效果.为了抗噪音干扰和边缘柔化,我们在第 2.1 节处理后的复原图像上提取HOG 特征.首先,按公式(3)和公式(4)计算点(x,y)的梯度幅值G(x,y)和方向θ(x,y),计算结果如图5所示,图像的梯度图表现了异态汉字的整体轮廓和梯度幅值.

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{3}$$

$$\theta(x,y) = \arctan \left[\frac{G_y(x,y)}{G_x(x,y)} \right] \tag{4}$$

其中,G_x(x,y),G_y(x,y)分别表示该点水平方向、垂直方向的梯度幅值,其计算方法如公式(5)所示.

$$G_x(x,y)=I(x+1,y)-I(x-1,y),G_y(x,y)=I(x,y+1)-I(x,y-1) \tag{5}$$

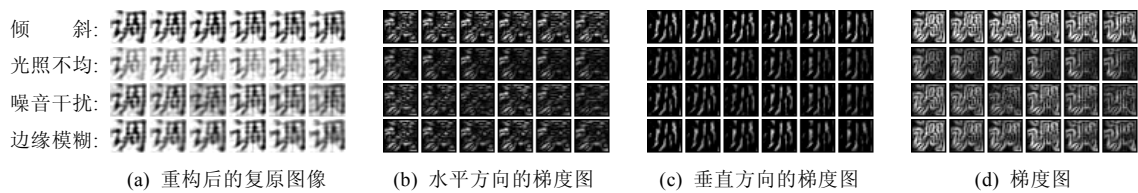


Fig.5 Gradient images of Chinese characters

图 5 汉字图像的梯度图

然后,在块内基于单元格统计直方图.对于单元格内每个像素点的权值,根据其梯度幅值采用高斯加权统计到其梯度方向所属的直方图分组(bin)中.直方图分组的方向范围为 0~π.

为了提高识别性能,削弱图像中光照不均的影响(如图 5(d)所示),对同一块内所有单元格的特征向量进行归一化处理.本文采用 L2-Hys 归一化方式,即以 0.2 为上限值对 L2-norm(计算方法如公式(6)所示)归一化数据修剪后再次归一化.

$$v \rightarrow v / \sqrt{\|v\|_2^2 + \epsilon^2} \tag{6}$$

其中,v 是未归一化的特征向量,ε是一个极小的常量.

归一化处理结果如图 6 所示.可见,虽然不同异态情况下的重构图像在光照(如图 4 所示)和梯度幅值(如图 5 所示)上存在明显的差异,但对图像同一块内的特征向量(24 个)进行归一化后,对应的每个特征值基本相等,从而体现了 HOG 特征对光照不均具有很强的适应能力.

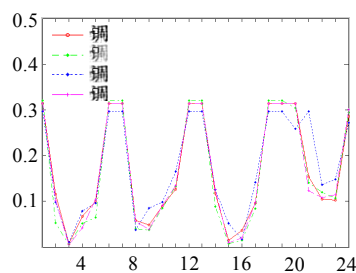


Fig.6 Normalized result

图 6 归一化处理结果

最后,通过滑动窗口(移动间隔为单元格大小)密集扫描图像,将检测窗口中所有块的特征向量依次连接起来,构成 HOG 特征集合.

2.3 整体流程及参数设置

SC-HOG 方法的流程如图 7 所示.训练阶段的参数设置如下:

- Step 1. 将训练样本归一化为 48×48 pixels 的图像,利用稀疏编码算法对所有的训练样本图像进行重构.为了提高效率,同时使重构的图像尽可能清晰,实验参数设置为 $\sigma=1, \beta=0.4, \epsilon=0.01$,基函数为 192 个,迭代次数为 1 000 次(参见公式(1)和公式(2)).
- Step 2. 提取复原图像的 HOG 特征向量.根据 Dalal 的理论^[69],采用一定的重叠块(窗口滑动扫描图像时),能大幅增强特征的鲁棒性,因此,实验参数分别为 $\text{block}=16 \times 16$ pixels, $\text{cell}=8 \times 8$ pixels, $\text{bin}=6$,检测窗口为图片大小,滑动窗口为 block 大小,以 8 pixels(单元格大小)的距离滑动,抽取到的特征向量维数为 600 维.
- Step 3. 为了支持在线学习,本文采用 ILDA^[76]进行汉字识别:将提取到的特征向量送入 ILDA 训练,得到最佳的变换矩阵,投影变换也是降维的过程,将利于分类识别的特征向量保留下来,消除冗余的非必要信息.降维后的特征向量为 96 维.

测试阶段与训练阶段的参数设置相同.

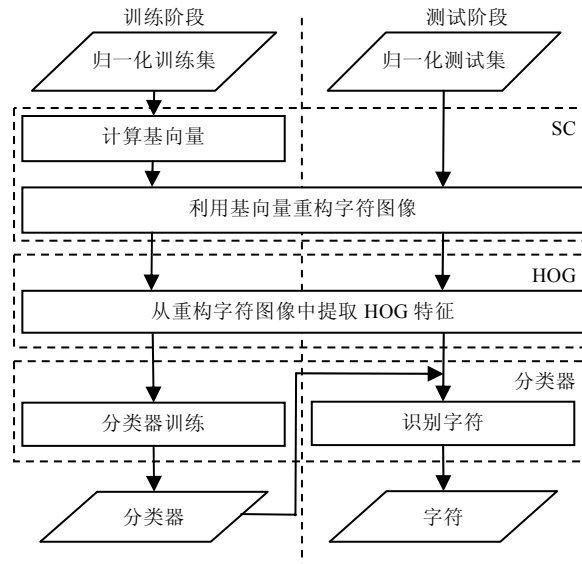


Fig.7 An overview of SC-HOG

图 7 SC-HOG 流程

3 实验结果及分析

本文的实验分为“效果实验分析(第 3.1 节~第 3.4 节)”与“综合实验结果(第 3.5 节)”两部分:前一部分在合成样本集上做实验,用于验证 SC-HOG 方法对于倾斜、光照不均、噪音干扰和边缘柔化等异态情况的鲁棒性;后一部分在现实世界的样本集上做综合实验,用于验证 SC-HOG 方法在实际应用中的有效性.

本文实验字符集包括一级汉字和部分常用二级汉字(共 3 879 类),所使用的训练样本集包括无任何干扰因素的 15 516(=3879 类 \times 4 种字体)幅汉字图像,字体包括宋体、仿宋、黑体和楷体,测试样本集包括对训练样本做处理得到的 4 个合成样本集和实际采集的 2 个真实样本集.

在早期关于印刷体汉字识别的研究工作中,统计识别与结构识别相结合的网格化特征由于具有较强的抗

干扰性,在实际中得到了广泛应用^[77].近期研究工作^[50,51]已经通过实验表明,HOG 在复杂图像文字识别中具有最好的性能.因此,在本文的比对实验中,选择网格化(grid)特征和 HOG 特征作为比对,对本文所述方法 SC-HOG 进行校验,见表 1.比对实验中均采用 ILDA^[76]进行汉字识别.

Table 1 Methods used in comparative experiments

表 1 比对实验中所用的方法

特征方法	方法描述
Grid	网格化特征.每个网格包含 2×2pixels,通过计算每个网格像素点的灰度平均值生成 576 维特征向量
HOG96	提取图像的 HOG 作为特征向量,参数为 block=32×32pixels,cell=16×16pixels,bin=6,检测窗口为图片大小,滑动窗口为 block 大小,以 16pixels 的距离滑动,特征向量的维数为 96 维
HOG216	提取图像的 HOG 作为特征向量,参数为 block=24×24pixels,cell=12×12pixels,bin=6,检测窗口为图片大小,滑动窗口为 block 大小,以 12pixels 的距离滑动,特征向量的维数为 216 维
HOG600	提取图像的 HOG 特征,参数为 block=16×16pixels,cell=8×8pixels,bin=6,检测窗口为图片大小,滑动窗口为 block 大小,以 8pixels 的距离滑动,特征向量的维数为 600 维

3.1 关于倾斜问题

对训练集中的每一幅图像,按顺时针、逆时针两个方向每隔 1°进行旋转,生成测试集.对训练集中的每一幅汉字图像,测试集中对应倾斜角度分别为 1°~3°的汉字图像共 6 幅,如图 8 所示.表 2 分别列出了各方法对测试集的认识率.

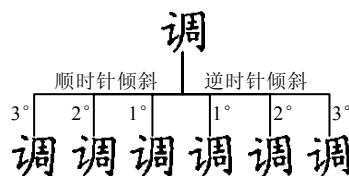


Fig.8 Examples of Chinese characters with skew

图 8 倾斜汉字图像示例

Table 2 Comparison of different methods on the recognition accuracy of Chinese characters with skew

表 2 各方法对倾斜汉字的识别率

倾斜角度	Grid (%)	HOG96 (%)	HOG216 (%)	HOG600 (%)	SC-HOG (%)
1°	99.27	99.08	99.77	99.93	99.98
2°	98.06	98.90	99.74	99.88	99.95
3°	95.60	98.53	99.58	99.82	99.85

从表 2 可以看出,Grid 能够识别适度倾斜的汉字图像,但当倾斜角度增大时,其识别率有较大程度的下降;HOG 则具有较强的抗倾斜能力;本文方法利用了 HOG 的优势,对适度的倾斜具有较强的鲁棒性,同时,SC 处理所造成的信息损失并没有对识别造成负面影响.

3.2 关于光照不均问题

将原始样本图像和第 3.1 节生成的倾斜图像通过加深背景色、笔画像素灰度值,以每列 2pixels 递增的方法得到光照不均的汉字图像(其中,每一套图像中包括汉字无倾斜图像 1 幅,倾斜角度为 1°~3°的图像共 6 幅),如图 9 所示.表 3 分别列出了各方法对测试集的认识率.

从表 3 可以看出,Grid 对光照不均的汉字图像识别性能明显下降.这是由于光照引起图像的前背景像素点灰度发生变化,从而使灰度特征失效;对于 HOG 特征,设置较大的滑动窗口(HOG96)识别率很低,通过调小滑动窗口,增加窗口之间的重叠度,可以削弱光照和倾斜的影响,性能得到大幅提升(HOG216 和 HOG600);本文方法利用了 HOG 的优势,对光照和适度的倾斜具有较强的鲁棒性,同时,SC 处理所造成的信息损失对识别所带来的负面影响较小.

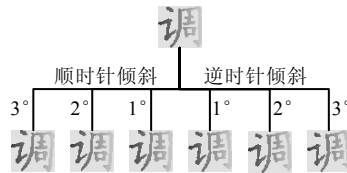


Fig.9 Examples of Chinese characters with uneven illumination

图 9 光照不均汉字图像

Table 3 Comparison of different methods on the recognition accuracy of Chinese characters with uneven illumination

表 3 各方法对光照不均汉字的识别率

倾斜角度	Grid (%)	HOG96 (%)	HOG216 (%)	HOG600 (%)	SC-HOG (%)
0°	20.39	98.32	99.77	99.97	99.95
1°	17.03	97.70	99.66	99.91	99.92
2°	11.27	97.11	99.57	99.89	99.85
3°	6.95	96.19	99.35	99.81	99.77

3.3 关于噪音干扰问题

将原始样本图像和第 3.1 节生成的倾斜图像分别加入方差 σ^2 为 0.10, 0.20, 0.50 的高斯噪声, 得到噪音干扰的汉字图像(其中, 每一套结果图像中包括汉字无倾斜图像 1 幅, 倾斜角度为 1°~3° 的图像共 6 幅), 如图 10 所示. 表 4 分别列出了不同程度噪音下各方法对测试集的识别率.

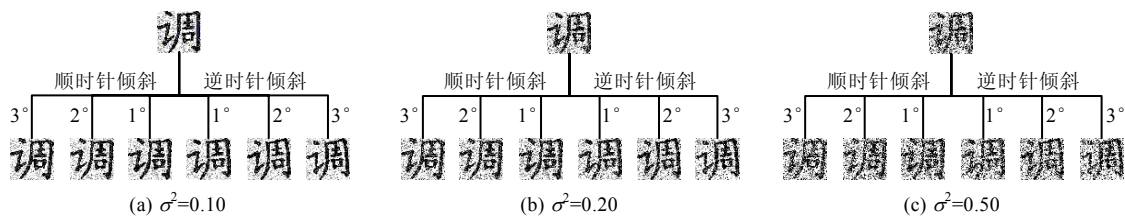


Fig.10 Examples of Chinese characters with noises

图 10 噪音干扰汉字图像

Table 4 Comparison of different methods on the recognition accuracy of Chinese characters with noises

表 4 各方法对噪音干扰汉字的识别率

参数	倾斜角度	Grid (%)	HOG96 (%)	HOG216 (%)	HOG600 (%)	SC-HOG (%)
$\sigma^2=0.10$	0°	99.36	65.33	71.63	65.85	99.93
	1°	98.38	59.66	68.80	61.79	99.89
	2°	95.73	56.28	65.94	59.62	99.81
	3°	88.65	52.20	62.13	55.90	99.80
$\sigma^2=0.20$	0°	97.89	14.79	21.61	23.27	99.80
	1°	95.64	13.12	19.61	21.13	99.69
	2°	90.46	11.85	17.96	19.43	99.52
	3°	79.52	10.44	15.94	17.11	99.15
$\sigma^2=0.50$	0°	77.87	0.54	1.13	1.71	94.99
	1°	70.09	0.55	0.96	1.60	94.86
	2°	58.61	0.51	0.97	1.57	93.89
	3°	43.87	0.45	0.94	1.42	92.02

从表 4 可以看出, Grid 能够识别适度噪音干扰的汉字图像, 但当噪音较多时, 其性能下降明显; 由于 HOG 表征的是汉字边缘特征, 而噪音严重干扰了图像中边缘梯度变化, 从而使该特征失效; 本文方法利用 SC 的滤噪性能, 同时利用 HOG 的抗倾斜性能, 保证了倾斜时噪音干扰图像的识别性能.

3.4 关于边缘柔化问题

将原始样本图像和第 3.1 节生成的倾斜图像分别利用标准差 σ 为 2,2,5,卷积模板 τ 为 $3 \times 3, 5 \times 5, 5 \times 5$ 的高斯模糊得到边缘柔化的汉字图像(其中,每一套结果图像均包括汉字无倾斜图像 1 幅,倾斜角度为 $1^\circ \sim 3^\circ$ 的图像共 6 幅),如图 11 所示.表 5 分别列出了不同参数下各方法对测试集的认识率.

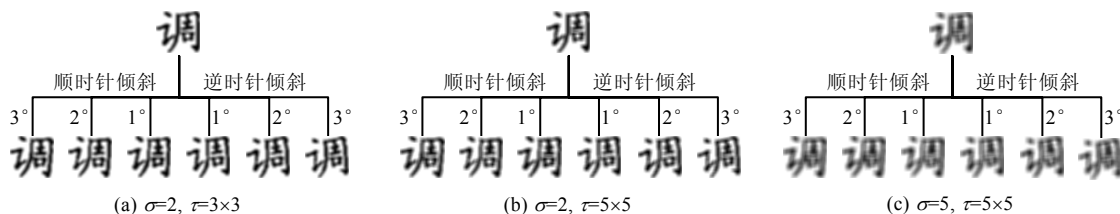


Fig.11 Examples of Chinese characters with anti-aliasing

图 11 边缘柔化汉字图像

Table 5 Comparison of different methods on the recognition accuracy of Chinese characters with anti-aliasing

表 5 各方法对边缘柔化汉字的认识率

参数	倾斜角度	Grid (%)	HOG96 (%)	HOG216 (%)	HOG600 (%)	SC-HOG (%)
$\sigma=2, \tau=3 \times 3$	0°	99.50	98.59	99.65	99.93	99.98
	1°	99.27	98.52	99.63	99.90	99.98
	2°	98.42	98.21	99.57	99.88	99.94
	3°	95.04	97.66	99.39	99.79	99.87
$\sigma=2, \tau=5 \times 5$	0°	99.30	97.71	99.50	99.89	99.99
	1°	99.06	97.59	99.42	99.84	99.97
	2°	98.27	97.00	99.34	99.81	99.92
	3°	94.98	96.27	99.18	99.71	99.83
$\sigma=5, \tau=5 \times 5$	0°	99.22	97.18	99.42	99.88	99.99
	1°	98.97	97.02	99.32	99.86	99.96
	2°	98.19	96.41	99.22	99.79	99.91
	3°	94.94	95.44	99.09	99.68	99.81

从表 5 可以看出,Grid 能够识别边缘柔化的汉字图像;对于 HOG 特征,设置较大的滑动窗口(HOG96),其认识率很低,通过调小滑动窗口,增加窗口之间的重叠度,可以削弱边缘柔化和倾斜的影响,性能得到大幅提升(HOG216 和 HOG600);本文方法利用了 HOG 的优势,对光照和适度的倾斜具有较强的鲁棒性,同时,SC 处理所造成的信息损失并没有给识别带来负面影响.

从第 3.1 节~第 3.4 节的实验可以看出:

- Grid:适合处理边缘柔化、轻度倾斜和轻度噪音的情况;当倾斜角度或噪音干扰增加时,处理效果下降明显;不适合识别光照不均的汉字图像;
- HOG:具有较强的抗倾斜能力,能够较好地处理光照不均和边缘柔化情况,对噪音较敏感;
- SC-HOG:具有较强的抗倾斜能力,能够较好地处理光照不均和边缘柔化情况,噪音严重时也能得到较好的处理结果.

在合成数据集上的这些实验结果表明,SC-HOG 方法对于倾斜、光照不均、噪音干扰和边缘柔化等异态情况有较强的鲁棒性.

3.5 真实数据集上的实验

本文以原生数字图像和场景图像中的文字识别为例,对本文所提出的汉字识别方法进行校验.实验中共采集了 400 幅原生数字图像和 376 幅场景图像,分别包括 5 320 和 8 335 个汉字.表 6 分别给出了 HOG600 和 SC-HOG 在真实数据集上的认识率,图 12 是其图形化表示.可见,SC-HOG 方法在真实数据集上也展现出了较好的结果.

Table 6 Comparison of HOG600 and SC-HOG on the real dataset

表 6 HOG600 和 SC-HOG 在真实数据集上的识别率

数据集	特征方法	候选字数								
		1	10	20	30	40	60	80	100	
原生数字图像	HOG600 (%)	44.15	63.03	67.56	69.85	71.65	74.02	75.70	76.97	
	SC-HOG (%)	47.35	65.94	70.15	72.69	74.38	76.41	77.76	78.91	
场景图像	HOG600 (%)	59.36	74.74	77.63	79.31	80.59	82.13	83.12	83.86	
	SC-HOG (%)	65.80	81.49	84.29	85.70	86.73	88.27	89.18	89.84	

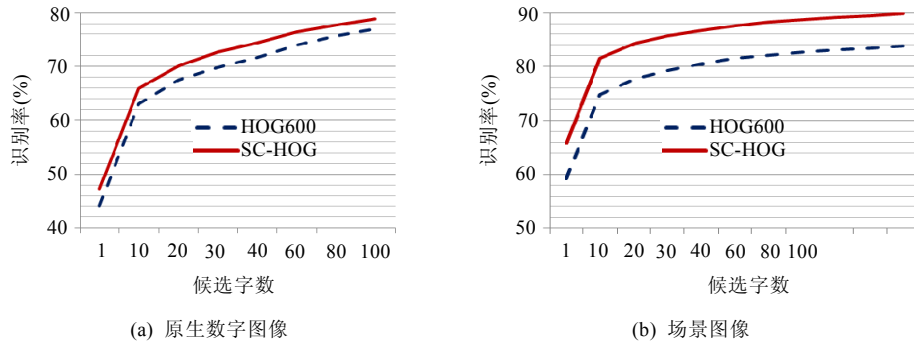


Fig.12 Comparison of HOG600 and SC-HOG on the real dataset

图 12 HOG600 和 SC-HOG 在真实数据集上的识别率

图 13 和图 14 中分别给出了利用本文方法识别正确和错误汉字图像的示例.从图 14 可以看出(相应地,左图为测试图像,右图为匹配的训练图像),识别错误的原因主要在于:(1) 字形相似;(2) 文字模糊不清;(3) 字体特殊.

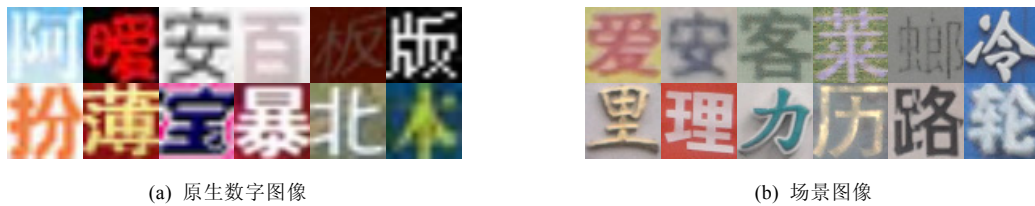


Fig.13 Examples of correctly recognized Chinese characters in the real dataset

图 13 真实数据集中识别正确的汉字图像示例



Fig.14 Examples of mis-recognized Chinese characters in the real dataset

图 14 真实数据集中识别错误的汉字图像示例

4 结束语

针对复杂图像中文字可能存在的倾斜、光照不均、噪音干扰、边缘柔化等问题,本文提出一种异态汉字识别方法 SC-HOG.该方法先利用稀疏编码滤除噪音和处理边缘柔化,再通过 HOG 削弱光照影响并增强抗倾斜能力,从而有效地识别异态汉字.本文从“效果实验分析(第 3.1 节~第 3.4 节)”与“综合实验结果(第 3.5 节)”两方面验证了本文提出的基于 SC-HOG 的异态汉字识别方法:前一方面实验结果表明,SC-HOG 方法对于倾斜、光照不均、噪音干扰和边缘柔化等异态情况有较强的鲁棒性;后一方面实验结果表明,SC-HOG 方法在原生数字图像和场景图像真实样本集上也能取得较好的结果.

文档图像处理领域的知名国际会议 ICDAR 在 2003 年、2005 年和 2011 年分别举办了 3 次场景图像文字提取竞赛、在 2011 年首次举办了原生数字图像文字提取竞赛,极大地促进了复杂图像文字提取方面的研究工作.国内外学者基于竞赛提供的带标注的样本集开展了大量研究工作,取得了丰硕的成果.然而,竞赛中所使用的样本集都是以英文字符作为提取目标,目前还缺少可应用于复杂图像汉字提取研究的公开数据集,这严重阻碍了相关研究工作的开展.本文所采集的真实数据集规模还较小,我们下一步工作是进一步扩大真实数据集的规模、形成公开数据集并建设自动评测环境,为复杂图像汉字提取方面的研究提供一个良好的平台.另外,我们也将参考英文文本区域定位方面的工作,研究适用于汉字的文本区域定位方法.

References:

- [1] Nagy G. Twenty years of document image analysis in PAMI. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000, 22(1):38–62. [doi: 10.1109/34.824820]
- [2] Liu XQ, Samarabandu J. Multiscale edge-based text extraction from complex images. In: *Proc. of the Int'l Conf. on Multimedia and Expo. Toronto: IEEE*, 2006. 1721–1724. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4036951> [doi: 10.1109/ICME.2006.262882]
- [3] Hasan YMY, Karam LJ. Morphological text extraction from images. *IEEE Trans. on Image Processing*, 2000,9(11):1978–1983. [doi: 10.1109/83.877220]
- [4] Phan TQ, Shivakumara P, Tan CL. A Laplacian method for video text detection. In: *Proc. of the 10th Int'l Conf. on Document Analysis and Recognition. Barcelona: IEEE*, 2009. 66–70. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5277790> [doi: 10.1109/ICDAR.2009.153]
- [5] Kim W, Kim C. A new approach for overlay text detection and extraction from complex video scene. *IEEE Trans. on Image Processing*, 2009,18(2):401–411. [doi: 10.1109/TIP.2008.2008225]
- [6] Zhang J, Kasturi R. Character energy and link energy-based text extraction in scene images. In: Kimmel R, Klette R, Sugimoto A, eds. *Proc. of the 10th Asian Conf. on Computer Vision, Part II. Queenstown: Springer-Verlag*, 2011. 308–320. <http://www.springerlink.com/content/21132457782g5419/> [doi: 10.1007/978-3-642-19309-5_24]
- [7] Epshtein B, Ofek E, Wexler Y. Detecting text in natural scenes with stroke width transform. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. San Francisco: IEEE*, 2010. 2963–2970. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5540041> [doi: 10.1109/CVPR.2010.5540041]
- [8] Peng X, Cao H, Prasad R, Natarajan P. Text extraction from video using conditional random fields. In: *Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE*, 2011. 1029–1033. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065466> [doi: 10.1109/ICDAR.2011.208]
- [9] Jain AK, Yu B. Automatic text location in images and video frames. *Pattern Recognition*, 1998,31(12):2055–2076. [doi: 10.1016/S0031-3203(98)00067-3]
- [10] Ohya J, Shio A, Akamatsu S. Recognizing characters in scene images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1994,16(2):214–224. [doi: 10.1109/34.273729]
- [11] Hase H, Shinokawa T, Yoneda M, Suen CY. Character string extraction from color documents. *Pattern Recognition*, 2001,34(7):1349–1365. [doi: 10.1016/S0031-3203(00)00081-9]
- [12] Garg R, Hassan E, Chaudhury S, Gopal M. A CRF based scheme for overlapping multi-colored text graphics separation. In: *Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE*, 2011. 1215–1219. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065503> [doi: 10.1109/ICDAR.2011.245]

- [13] Zhang HW, Liu CS, Yang C, Ding XQ, Wang KQ. An improved scene text extraction method using conditional random field and optical character recognition. In: Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE, 2011. 708–712. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065403> [doi: 10.1109/ICDAR.2011.148]
- [14] Zhong Y, Karu K, Jain AK. Locating text in complex color images. Pattern Recognition, 1995,28(10):1523–1535. [doi: 10.1016/0031-3203(95)00030-4]
- [15] Yi CC, Tian YL. Text string detection from natural scenes by structure-based partition and grouping. IEEE Trans. on Image Processing, 2011,20(9):2594–2605. [doi: 10.1109/TIP.2011.2126586]
- [16] Mao WG, Chung FL, Lanm KKM, Siu WC. Hybrid Chinese/English text detection in images and video frames. In: Proc. of the 16th Int'l Conf. on Pattern Recognition. Quebec: IEEE, 2002. 1015–1018. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1048210> [doi: 10.1109/ICPR.2002.1048210]
- [17] Jain AK, Karu K. Learning texture discrimination masks. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1996,18(2): 195–205. [doi: 10.1109/34.481543]
- [18] Jung K. Neural network-based text location in color images. Pattern Recognition Letter, 2001,22(14):1503–1515. [doi: 10.1016/S0167-8655(01)00096-4]
- [19] Li H, Doerman D, Kia O. Automatic text detection and tracking in digital video. IEEE Trans. on Image Process, 2000,9(1): 147–156. [doi: 10.1109/83.817607]
- [20] Kim KI, Jung K, Kim JH. Texture-Based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003,25(12):1631–1639. [doi: 10.1109/TPAMI.2003.1251157]
- [21] Lee JJ, Lee PH, Lee SW, Yuille A, Koch C. AdaBoost for text detection in natural scene. In: Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE, 2011. 429–434. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065348> [doi: 10.1109/ICDAR.2011.93]
- [22] Pan YF, Hou XW, Liu CL. A hybrid approach to detect and localize texts in natural scene images. IEEE Trans. on Image Processing, 2011,20(3):800–813. [doi: 10.1109/TIP.2010.2070803]
- [23] Zhao X, Lin KH, Fu Y, Hu Y, Liu Y, Huang TS. Text from corners: A novel approach to detect text and caption in videos. IEEE Trans. on Image Processing, 2011,20(3):790–799. [doi: 10.1109/TIP.2010.2068553]
- [24] Uchida S, Shigeyoshi Y, Kunishige Y, Yaokai F. A keypoint-based approach toward scenery character detection. In: Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE, 2011. 819–823. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065425> [doi: 10.1109/ICDAR.2011.168]
- [25] Mishra A, Alahari K, Jawahar CV. An MRF model for binarization of natural scene text. In: Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE, 2011. 11–16. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065267> [doi: 10.1109/ICDAR.2011.12]
- [26] Cho MS, Seok JH, Lee S, Kim JH. Scene text extraction by superpixel CRFs combining multiple character features. In: Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE, 2011. 1034–1038. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065467> [doi: 10.1109/ICDAR.2011.209]
- [27] Wang XF, Huang L, Liu CP. A novel method for embedded text segmentation based on stroke and color. In: Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE, 2011. 151–155. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065294> [doi: 10.1109/ICDAR.2011.39]
- [28] Wakahara T, Kita K. Binarization of color character strings in scene images using K -means clustering and support vector machines. In: Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE, 2011. 274–278. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065318> [doi: 10.1109/ICDAR.2011.63]
- [29] Zheng Q, Chen K, Zhou Y, Gu CC, Guan HB. Text localization and recognition in complex scenes using local features. In: Kimmel R, Klette R, Sugimoto A, eds. Proc. of the 10th Asian Conf. on Computer Vision, Part III. 2010. 121–132. <http://www.springerlink.com/content/2462601p175422g3/> [doi: 10.1007/978-3-642-19318-7_10]
- [30] Iwamura M, Kobayashi T, Kise K. Recognition of multiple characters in a scene image using arrangement of local features. In: Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE, 2011. 1409–1413. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065542> [doi: 10.1109/ICDAR.2011.283]
- [31] Karatzas D, Mestre SR, Mas J, Nourbakhsh F, Roy PP. ICDAR 2011 robust reading competition challenge 1: Reading text in born-digital images (Web and Email). In: Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE, 2011. 1485–1490. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065555> [doi: 10.1109/ICDAR.2011.295]

- [32] Antonacopoulos A, Karatzas D. An anthropocentric approach to text extraction from WWW images. In: Proc. of the 4th Int'l Workshop on Document Analysis Systems. Rio de Janeiro: Springer-Verlag, 2000. 515–525. http://eprints.soton.ac.uk/263495/1/DAS2000_Antonacopoulos.pdf
- [33] Antonacopoulos A, Karatzas D. Text extraction from Web images based on human perception and fuzzy inference. In: Proc. of the 1st Int'l Workshop on Web Document Analysis. Seattle: PRImA Press, 2001. 35–38. http://eprints.soton.ac.uk/263510/1/WDA2001_Antonacopoulos.pdf
- [34] Antonacopoulos A, Karatzas D. Fuzzy segmentation of characters in Web images based on human colour perception. In: Lopresti DP, Hu JY, Kashi RS, eds. Proc. of the 5th Int'l Workshop on Document Analysis Systems. Princeton: Springer-Verlag, 2002. 295–306. <http://www.springerlink.com/content/wl3wr9rkgd9f1e1e4/> [doi: 10.1007/3-540-45869-7_35]
- [35] Karatzas D, Antonacopoulos A. Two approaches for text segmentation in Web images. In: Proc. of the 7th Int'l Conf. on Document Analysis and Recognition. Edinburgh: IEEE, 2003. 131–136. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1227646> [doi: 10.1109/ICDAR.2003.1227646]
- [36] Karatzas D, Antonacopoulos A. Text extraction from Web images based on a split-and-merge segmentation method using colour perception. In: Proc. of the 17th Int'l Conf. on Pattern Recognition. Cambridge: IEEE, 2004. 634–637. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1334328> [doi: 10.1109/ICPR.2004.1334328]
- [37] Karatzas D, Antonacopoulos A. Colour text segmentation in Web images based on human perception. *Image and Vision Computing*, 2007,25(5):564–577. [doi: 10.1016/j.imavis.2006.05.003]
- [38] Sun J, Wang ZL, Yu H, Nishino F, Katsuyama Y, Naoi S. Effective text extraction and recognition for WWW images. In: Proc. of the 2003 ACM Symp. on Document Engineering. Grenoble: ACM Press, 2003. 115–117. http://dl.acm.org/ft_gateway.cfm?id=958241&ftid=236575&dwn=1&CFID=143626486&CFTOKEN=12231636 [doi: 10.1145/958220.958241]
- [39] Aradhye HB, Myers GK, Herson JA. Image analysis for efficient categorization of image-based spam E-mail. In: Proc. of the 8th Int'l Conf. on Document Analysis and Recognition. Seoul: IEEE, 2005. 914–918. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1575677> [doi: 10.1109/ICDAR.2005.135]
- [40] Zhou JY, Lopresti D. Extracting text from WWW images. In: Proc. of the 4th Int'l Conf. on Document Analysis and Recognition. Ulm: IEEE, 1997. 248–252. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=619850> [doi: 10.1109/ICDAR.1997.619850]
- [41] Lopresti D, Zhou JY. Locating and recognizing text in WWW images. *Information Retrieval*, 2000,2:177–206. [doi: 10.1023/A:1009954710479]
- [42] Liu F, Peng X, Wang TJ, Lu SF. A density-based approach for text extraction in images. In: Proc. of the 19th Int'l Conf. on Pattern Recognition. Tampa: IEEE, 2008. 1–4. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4761637> [doi: 10.1109/ICPR.2008.4761637]
- [43] Situ LJ, Liu RZ, Tan CL. Text localization in Web images using probabilistic candidate selection model. In: Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE, 2011. 1359–1363. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065532> [doi: 10.1109/ICDAR.2011.273]
- [44] Perantonis SJ, Gatos B, Maragos V. A novel Web image processing algorithm for text area identification that helps commercial OCR engines to improve their Web image recognition efficiency. In: Proc. of the 2nd Int'l Workshop on Web Document Analysis. Edinburgh, 2003. 61–64. <http://users.iit.demokritos.gr/~bgat/wda2003>
- [45] Perantonis SJ, Gatos B, Maragos V, Karkaletsis V, Petasis G. Text area identification in Web images. In: Vouros GA, Panayiotopoulos T, eds. Proc. of the Methods and Applications of Artificial Intelligence. Samos: Springer-Verlag, 2004. 82–92. <http://www.springerlink.com/content/1k3etxdannyybkeu/> [doi:10.1007/978-3-540-24674-9_10]
- [46] He JY, Li SF. Hybrid Chinese/English text identification in Web images. In: Proc. of the 3rd Int'l Conf. on Image and Graphics. Hong Kong: IEEE, 2004. 361–364. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1410459> [doi: 10.1109/ICIG.2004.78]
- [47] Liu J, Zhang SW, Li HP, Liang W. A Chinese character localization method based on intergrating structure and CC-clustering for advertising images. In: Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE, 2011. 1044–1048. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065469> [doi: 10.1109/ICDAR.2011.211]
- [48] Weinman JJ, Learned-Miller E, Hanson A. Scene text recognition using similarity and a lexicon with sparse belief propagation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2009,31(10):1733–1746. [doi: 10.1109/TPAMI.2009.38]
- [49] de Campos TE, Babu BR, Varma M. Character recognition in natural images. In: Ranchordas A, Araújo H, eds. Proc. of the Int'l Conf. on Computer Vision Theory and Application. Lisbon: INSTICC Press, 2009. <http://eprints.pascal-network.org/archive/00009132/01/deCampos-VISAPP-2009.pdf>

- [50] Wang K, Belongie S. Word spotting in the wild. In: Daniilidis K, Maragos P, Paragios N, eds. Proc. of the European Conf. on Computer Vision. Heraklion: Springer-Verlag, 2010. 591–604. http://vision.ucsd.edu/~kai/pubs/wang_eccv2010.pdf [doi: 10.1007/978-3-642-15549-9_43]
- [51] Newell AJ, Griffin LD. Multiscale histogram of oriented gradient descriptors for robust character recognition. In: Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE, 2011. 1085–1089. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065477> [doi: 10.1109/ICDAR.2011.219]
- [52] Wang K, Babenko B, Belongie S. End-to-End scene text recognition. In: Metaxas DN, Quan L, Sanfeliu A, Van Gool LJ, eds. Proc. of the 13th Int'l Conf. on Computer Vision. Barcelona: IEEE, 2011. 1457–1464. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6126402> [doi: 10.1109/ICCV.2011.6126402]
- [53] Mishra A, Alahari K, Jawahar CV. Top-Down and bottom-up cues for scene text recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Providence: IEEE, 2012. 2687–2694. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6247990> [doi: 10.1109/CVPR.2012.6247990]
- [54] Ding XQ. Chinese character recognition: A review. Acta Electronica Sinica, 2002,30(9):1364–1368 (in Chinese with English abstract).
- [55] Zhao JY, Zheng RR, Wu BC, Li M. A review of off-line handwritten Chinese character recognition. Acta Electronica Sinica, 2010, 38(2):405–415 (in Chinese with English abstract).
- [56] Neumann L, Matas J. Real-Time scene text localization and recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Providence: IEEE, 2012. 3538–3545. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6248097> [doi: 10.1109/CVPR.2012.6248097]
- [57] Pan YF, Hou X, Liu CL. A robust system to detect and localize texts in natural scene images. In: Kise K, Sako H, eds. Proc. of the 8th Int'l Workshop on Document Analysis Systems. Nara: IEEE, 2008. 35–42. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4669943> [doi: 10.1109/DAS.2008.42]
- [58] Wang WQ, Fu LB, Gao W, Huang QM, Jiang SQ. Text detection based on stroke features. Journal on Communications, 2007, 28(12):116–120 (in Chinese with English abstract).
- [59] Liu Y, Xue XY, Lu H, Guo YF. A video text detecting method based on edge detection and line features. Chinese Journal of Computers, 2005,28(3):427–433 (in Chinese with English abstract).
- [60] Zhang Y, Pan YH. Design of a new color edge detector for text extraction under complex background. Ruan Jian Xue Bao/Journal of Software, 2001,12(8):1229–1235 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/12/1229.htm>
- [61] Zhao Y, Lu T, Liao WJ. A robust color-independent text detection method from complex videos. In: Proc. of the 11th Int'l Conf. on Document Analysis and Recognition. Beijing: IEEE, 2011. 374–378. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065338> [doi: 10.1109/ICDAR.2011.83]
- [62] Cai B, Zhou TR, Zhu YY. Digital video global text extraction techniques based on line extraction. Journal of Wuhan University of Hydraulic and Electric Engineering, 2005,38(4):104–108 (in Chinese with English abstract).
- [63] Olshausen BA, Field DJ. Emergence of simple-cell-receptive-field properties by learning a sparse code for natural images. Nature, 1996,381:607–609. [doi: 10.1038/381607a0]
- [64] Olshausen BA, Field DJ. Sparse coding with an overcomplete basis set: A strategy employed by V1. Vision Research, 1997,37(23): 3311–3325.
- [65] Labusch K, Barth E, Martinetz T. Simple method for high-performance digit recognition based on sparse coding. IEEE Trans. on Neural Networks, 2008,19(11):1985–1989. [doi: 10.1109/TNN.2008.2005830]
- [66] Barthelemy Q, Larue A, Mayoue A, Mercier D, Mars JI. Shift & 2D rotation invariant sparse coding for multivariate signals. IEEE Trans. on Signal Processing, 2012,60(4):1597–1611. [doi: 10.1109/TSP.2012.2183129]
- [67] Lee H, Battle A, Raina R, Ng AY. Efficient sparse coding algorithms. In: Schölkopf B, Platt JC, Hoffman T, eds. Proc. of the 2006 Conf. Advances in Neural Information Processing Systems. Vancouver: MIT Press, 2007. 801–808. http://books.nips.cc/papers/files/nips19/NIPS2006_0878.pdf
- [68] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. San Diego: IEEE, 2005. 886–893. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1467360> [doi: 10.1109/CVPR.2005.177]
- [69] Chuang CH, Huang SS, Fu LC, Hsiao PY. Monocular multi-human detection using augmented histograms of oriented gradients. In: Proc. of the 19th Int'l Conf. on Pattern Recognition. Tampa: IEEE, 2008. 1–4. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4761500> [doi: 10.1109/ICPR.2008.4761500]

- [70] Wang XY, Han TX, Yan SC. An HOG-LBP human detector with partial occlusion handling. In: Proc. of the 12th Int'l Conf. on Computer Vision. Kyoto: IEEE, 2009. 32–39. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5459207> [doi: 10.1109/ICCV.2009.5459207]
- [71] Albiol A, Monzo D, Martin A, Sastre J, Albiol A. Face recognition using HOG-EBGM. Pattern Recognition Letters, 2008,29(10): 1537–1543. [doi: 10.1016/j.patrec.2008.03.017]
- [72] Creusen IM, Wijnhoven RGJ, Herbschleb E, de With PHN. Color exploitation in hog-based traffic sign detection. In: Proc. of the 17th Int'l Conf. on Image Processing. Hong Kong: IEEE, 2010. 2669–2672. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5651637> [doi: 10.1109/ICIP.2010.5651637]
- [73] Jia HX, Zhang YJ. Multiple kernels based object tracking using histograms of oriented gradients. Acta Automatica Sinica, 2009, 35(10):1283–1289 (in Chinese with English abstract). [doi: 10.3724/SP.J.1004.2009.01283]
- [74] Xiao XY, Hu RX, Zhang SW, Wang XF. HOG-Based approach for leaf classification. In: Huang DS, Zhang X, Alberto C, García R, Zhang L, eds. Proc. of the 6th Int'l Conf. on Intelligent Computing. Changsha: Springer-Verlag, 2010. 149–155. <http://www.springerlink.com/content/vr74w9201702x687/> [doi: 10.1007/978-3-642-14932-0_19]
- [75] 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]
- [76] Pang S, Ozawa S, Kasabov N. Incremental linear discriminant analysis for classification of data streams. IEEE Trans. on Systems, Man and Cybernetics—Part B: Cybernetics, 2005,35(5):905–914. [doi: 10.1109/TSMCB.2005.847744]
- [77] Liu CS, Guo FX, Ding XQ, Guo H. A review of machine-printed Chinese character recognition. China InfoWorld, 1997,663: 141–145 (in Chinese with English abstract).

附中文参考文献:

- [54] 丁晓青. 汉字识别研究的回顾. 电子学报, 2002,30(9):1364–1368.
- [55] 赵继印, 郑蕊蕊, 吴宝春, 李敏. 脱机手写体汉字识别综述. 电子学报, 2010,38(2):405–415.
- [58] 王伟强, 付立波, 高文, 黄庆明, 蒋树强. 基于笔画特征的叠加文字检测方法. 通信学报, 2007,28(12):116–120.
- [59] 刘洋, 薛向阳, 路红, 郭跃飞. 一种基于边缘检测和线条特征的视频字符检测算法. 计算机学报, 2005,28(3):427–433.
- [60] 张引, 潘云鹤. 复杂背景下文本提取的彩色边缘检测算子设计. 软件学报, 2001,12(8):1229–1235. <http://www.jos.org.cn/1000-9825/12/1229.htm>
- [62] 蔡波, 周洞汝, 朱映映. 基于直线抽取的数字视频全局文字提取的研究. 武汉大学学报(工学版), 2005,38(4):104–108.
- [73] 贾慧星, 章毓晋. 基于梯度方向直方图特征的多核跟踪. 自动化学报, 2009,35(10):1283–1289. [doi: 10.3724/SP.J.1004.2009.01283]
- [77] 刘长松, 郭繁夏, 丁晓青, 郭宏. 印刷汉字识别方法综述. 中国计算机报, 1997,663:141–145.



王恺(1979—),男,吉林长春人,博士,副教授,CCF 会员,主要研究领域为图像处理,模式识别.

E-mail: wangk@nankai.edu.cn



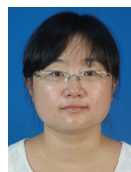
赵宏(1967—),女,博士,副教授,CCF 会员,主要研究领域为机器学习.

E-mail: zhaoh@nankai.edu.cn



李成学(1988—),男,硕士生,主要研究领域为文档图像处理.

E-mail: lichengxuelcx@126.com



张健(1978—),女,博士生,主要研究领域为文档图像处理.

E-mail: nkzhangjian@163.com



王庆人(1944—),男,博士,教授,博士生导师,主要研究领域为图像处理,模式识别,软件工程.

E-mail: qrwang@experexchange.com