

# 基于主动轮廓模型的文本检测方法

许肖 顾磊

(南京邮电大学计算机学院 南京 210003)

**摘要** 针对复杂背景下的文本检测问题,提出了一种基于主动轮廓模型的文本检测方法。输入的图像首先经过 sobel-laplacian 锐化后再由 gaussian-laplacian 进行过滤。预处理完成后的图片首先通过改进的主动轮廓模型得到初始轮廓,再通过算法的反复迭代扩大或缩小轮廓线得到最终轮廓,最后通过后处理尽量排除非文本块,从而得到最终文本区。区别于以往检测方法,所提方法最终不但可以框出文本行,还可以框出单个文本,有利于后续分割识别的进行。实验表明所提方法可有效检测出图像中的文本。

**关键词** 文本检测,主动轮廓模型,复杂背景,边缘检测

**中图分类号** TP391.41      **文献标识码** A      **DOI** 10.11896/j.issn.1002-137X.2015.6.060

## Text Detection Method Based on Active Contour Model

XU Xiao GU Lei

(School of Computer Science & Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

**Abstract** To detect text from images with different backgrounds, a text detection method with active contour models was proposed. Before text detection, the sobel-laplacian and gaussian-laplacian were used to sharpen edges and smooth noise, and then iteration algorithm was repeated to enlarge or lessen the contour to get the final contour, ruling out the un-text block at last. The proposed method can box a single text eventually, and it is advantageous to the subsequent segmentation recognition. Experiment shows that the proposed method can effectively detect the text in the image.

**Keywords** Text detection, Active contour models(ACM), Complex background, Edge detection

## 1 引言

随着拍摄功能产品的快速发展以及价格的降低,这类数码产品已经渗透进我们的生活。智能手机、平板电脑等移动设备已经允许我们随时随地使用图片保存我们身边的文本信息,这些图片通常都具有背景复杂的特点。而在方便地保存信息的同时,由于图像中的文本通常具有简短且重要的讯息,因此对拍摄到的图片中文本的检测也具有重要意义及应用价值。

目前,国内外关于复杂背景下的文本检测方法大致可分为基于特征提取的文本检测方法以及基于机器学习的方法两大类。其中基于特征提取的文本检测方法可分为基于连通域、基于边缘、基于纹理 3 种方法。基于机器学习的方法可分为基于神经网络、基于统计学习、基于支持向量机 3 种方法。文献[1]采用了边缘检测、候选文本区检测和细化检测 3 个步骤。首先使用边缘检测获取边缘特征,然后使用聚类分析算法检测到初始文本区域,最后通过以往的经验进行分析以确定最终文本区域。文献[2]提出了一种基于颜色文本检测方法,即通过检测图像的每个像素行来确认该行是否存在文本,与文本像素点颜色相似的像素点应优先检查。文献[3]提出了一种有效的基于边缘的文本检测方法,该方法基于边缘强度、边缘密度、水平分布等稳定特征。文献[4]提出了一种通

过拉普拉斯预处理的骨架化文本检测方法,其分文本检测、连通分量分类、连通分量分割,以及去除非文本 4 步骤进行。

图像分割方法已经较为成熟,如文献[5]提出的 signed pressure force(SPF)分割模型。本文试图将 SPF 模型应用于文本检测,不过将其直接移植到文本检测中的效果并不理想,本文对其进行了改进,提出了新的检测方法。提出的方法可将文本单独框出,而不是框出整个文本区域,并且具有较好的检测效果,这是现有的文本检测方法所不具备的。因此与之前的工作相比,本文有如下两方面的贡献:1)本文提出的方法可以框出单个字符的轮廓,与以往方法均用方框框住一行文本相比,更加精细直观,并且有利于后续识别等工作的展开;2)本文比其他基于边缘或基于颜色的方法更具鲁棒性,检测效果更好。

本文第 2 节介绍了 SPF 主动轮廓模型的主要思想;第 3 节介绍了本文提出的方法;第 4 节进行了实验结果的对比与讨论;最后总结全文。

## 2 SPF 主动轮廓模型

主动轮廓提取法(Active Contour Models, ACM)在图像分割中是较成熟的方法,有较好的效果<sup>[6]</sup>,将其引入并应用于文本检测是可行的。针对文本检测的特点,对主动轮廓提取法进行改进将得到更好的效果。主动轮廓提取法的基本思想

到稿日期:2014-07-30 返修日期:2014-10-24 本文受国家自然科学基金项目(61302157)资助。

许肖(1989-),女,硕士生,主要研究领域为文本检测,E-mail:903368794@qq.com;顾磊(1978-),男,副教授,硕士生导师,主要研究领域为机器学习、模式识别。

是在某些约束条件下引入一条曲线来圈住目标文本,构建文本轮廓。根据约束条件的性质,现存的 ACM 方法可以分成两类:基于边缘模型<sup>[7]</sup>和基于区域模型<sup>[8]</sup>。

基于边缘模型中最具代表性的是 GAC 模型,它能够利用图像梯度在目标文本边缘来构建边界停止函数(ESF),阻止轮廓继续扩张。该模型引入“能量泛函”(Energy function)这一概念,将定位这一复杂的过程简化为求泛函最小值的数学过程。具体如下:

定义模型中 Energy function 计算公式为

$$F^{GAC}(s) = \int_0^1 g(|\nabla I(S(\rho))|) |S(\rho)| d\rho \quad (1)$$

其中,  $S(\rho)$  为文本区域的外轮廓,  $g$  表示差分运算,  $I$  表示输入的图像。对应着定义一个 balloon force, 这个力位于文本区域内部, 如果初始轮廓比文本区域轮廓小, 则该力就变大, 使得定位轮廓扩大; 如果初始轮廓比文本区域大, 则该力变小, 以缩小定位轮廓的区域。而此时的 level set formulation 计算公式如下:

$$\frac{\partial \theta}{\partial \omega} = g|\nabla \phi| \operatorname{div}(\nabla \phi / |\nabla \phi| + \alpha) + \nabla g \cdot \nabla \phi \quad (2)$$

其中,  $\alpha$  是控制轮廓扩张或收缩的力。基于区域模型中最具代表性的是 C-V 模型, 无论初始轮廓从图像何处开始, C-V 模型都能够自动地检测到所有轮廓。该模型是基于区域的一种方法, 它也含有“能量泛函”这一概念, 不过它与 GAC“能量泛函”定义公式不同。C-V 中的函数定义如下:

$$F^{CV} = \lambda_1 \int_{\text{inside}(c)} |I(x) - c_1|^2 dx + \lambda_2 \int_{\text{outside}(c)} |I(x) - c_2|^2 dx, x \in \Omega \quad (3)$$

其中,  $\lambda_1, \lambda_2$  都是固定参数。可以看出, 式(3)将整个图像  $\Omega$  分为文本和背景两个区域, 而它们各自的系数  $c_1$  和  $c_2$  反映了该区域的特征。在模型中, 它们的值和泛函的最小值相关。所以其计算式如下:

$$c_1(\phi) = \int_{\Omega} I(x) \cdot K(\phi) dx / \int_{\Omega} K(\phi) dx \quad (4)$$

$$c_2(\phi) = \int_{\Omega} I(x) \cdot (1 - K(\phi)) dx / \int_{\Omega} (1 - K(\phi)) dx$$

其中,  $K(\phi)$  为海维塞函数。这样的组合使得泛函得到了最小值。同样, 类似 GAC 模型, 该模型中也可以得到文本区域的 level set formulation 公式:

$$\frac{\partial \theta}{\partial \omega} = \delta(\phi) [\mu \nabla(\nabla \phi / |\nabla \phi|)] - v - \lambda_1 (1 - c_1)^2 + \lambda_2 (1 - c_2)^2 \quad (5)$$

式中,  $\delta(\phi)$  为单位脉冲响应。文献[5]提到的 spf 模型结合了 GAC 模型和 CV 模型, 使用轮廓边缘内外的统计信息来构建一个基于区域的带有符号的压力公式, 由这个公式来控制检测到的轮廓的变化方向。使用该方法使得目标轮廓内外有着相反的符号, 因此可以检测到的轮廓在目标轮廓外时可收缩, 在目标轮廓内时可扩张。这样当初始轮廓在文本区域的内部或是外部时, 定位轮廓就可以及时缩小或放大。在多次迭代后, 该模型可以得到一个最终解, 使得 level set formulation 得到最小解。

该模型的核心算法思想是引入 spf 函数<sup>[9]</sup>, 定义如下:

$$spf(S(x)) = S(x) - (c_1 + c_2) / 2 / \max(|S(x) - ((c_1 + c_2) / 2)|), x \in \Omega \quad (6)$$

假设文本区域和背景区域的数值符号相反, 可以得到:

$$\operatorname{Min}(S(x)) < (c_1 + c_2) / 2 < \operatorname{Max}(S(x)), x \in \Omega \quad (7)$$

将该 spf 函数定义式代入 level set formulation 计算式, 可得:

$$\frac{\partial \phi}{\partial \omega} = spf(S(x)) \cdot (\operatorname{div}(\nabla \phi / |\nabla \phi|) + \alpha) |\nabla \phi| + \nabla spf(S(x)) \cdot \nabla \phi, x \in \Omega \quad (8)$$

最终, 该模型用数学拉氏变换, 将上面复杂的公式简化为下式, 以便算法的实现, 其中参数可变。

$$\frac{\partial \phi}{\partial \omega} = spf(S(x)) \cdot \alpha |\nabla \phi|, x \in \Omega \quad (9)$$

### 3 改进的主动轮廓检测方法

本文提出的方法由 3 部分组成: 图像预处理, 使用改进的主动轮廓模型提取出文本区域, 提取出的备选文本区的后处理。

#### 3.1 图像预处理

因为图像中的文本与复杂的背景有较低的对比如, 所以突出文本与非文本区别的预处理工作非常重要。本文采用进行 sobel-laplacian<sup>[10]</sup> 变换之后再进一步 gaussian-laplacian<sup>[11]</sup> 变换的两步骤的预处理方法。

在实验结果的检验过程中发现: 一般来说, 当图像本身对比如度不是很高时, 图像的质量不好, 如图 1 所示, 目标和背景之间对比如度差, 目标边缘模糊且噪声大, 直接检测效果不佳。所以使用拉普拉斯变换锐化了这些图像之后, 再使用主动轮廓定位。效果如图 1 所示。

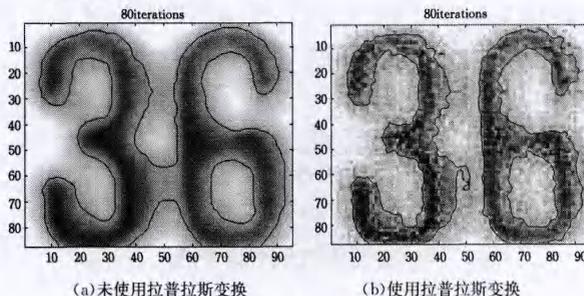


图 1 拉普拉斯锐化后的效果

可以看到之前仿真出来的程序定位模糊, 使用拉氏变换锐化处理图像之后, 定位效果显著提高。

但是锐化也带来了一些麻烦, 因为在锐化过程中, 可能将一些不需要的信息比如噪声等明显化, 会带来不好的影响。故参考文献[4], 提出了 gaussian-laplacian 变换, 首先对图像进行高斯变换(平滑噪声), 再用拉普拉斯变换(锐化对比如度低的图像), 这样的效果就更加完善了。为了加强显示效果, 将文本框颜色设为  $y$ , 宽度设为 2, 效果如图 2 所示。

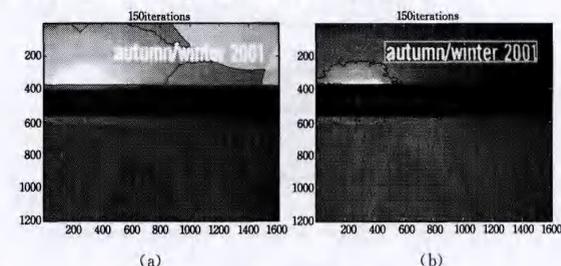


图 2 gaussian-laplacian 变换后的效果

由于复杂背景的特点, 有些图片较为模糊或文本与背景

融合,本文叠加使用 sobel-laplacian 变换来锐化图片,提高检测效果。对复杂背景下的文本检测效果如图 3 所示。

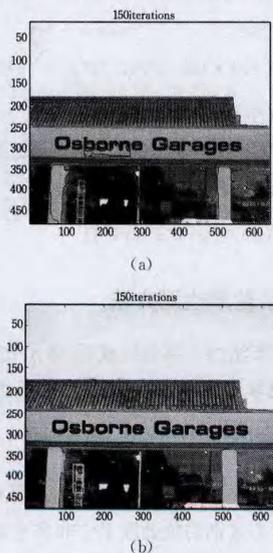


图 3 sobel-laplacian 变换后的效果

从图 3 可以看出,使用 sobel-laplacian 变换后再进行 gaussian-laplacian 变换的两步骤的预处理的效果,比未使用时更为理想,可以检测出对比度低或较模糊的图片。

图 4 从左到右依次是未加本文所提出预处理、加入 gaussian-laplacian 处理、加入 sobel-laplacian 处理后的结果图,从图中可明显看到两次预处理变换对文本检测所起的效果。

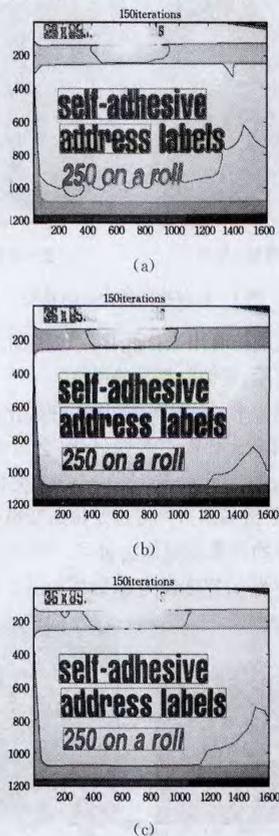


图 4 预处理变换的效果

### 3.2 改进的轮廓提取算法

虽然 SPF 算法在图像分割中具有良好表现,但是直接将

其用于文本检测得不到很好的效果,效果图可见 4.2 节。本文对其进行改进,对初始轮廓的选取进行优化,使检测效果更加理想。具体做法如下:图像中的一个像素与周围的像素之间的差值大于一定阈值之后,就令其为 1;否则令其为 -1。这样,初始轮廓就是由一些点阵组成,而该点阵本身的二值化,就已经能够一定程度地反映一幅图像中的文本信息。初始轮廓的选取方法如式(10)所示。

$$\begin{aligned}
 & [P(x-w,y)-P(x,y)>t] \text{AND} [P(x+w,y)-P(x,y) >t] \\
 & \text{OR} \\
 & [P(x,y-w)-P(x,y)>t] \text{AND} [P(x,y+w)-P(x,y) >t] \\
 & \text{OR} \\
 & [P(x+d,y+d)-P(x,y)>t] \text{AND} [P(x-d,y-d)-P(x,y)>t] \\
 & \text{OR} \\
 & [P(x-d,y+d)-P(x,y)>t] \text{AND} [P(x+d,y-d)-P(x,y)>t]
 \end{aligned} \quad (10)$$

其中, $P(x,y)$ 代表 $(x,y)$ 点像素值的大小, $d,w$ 代表像素间相互比较时的范围大小, $t$ 代表阈值。如果满足上面的条件,则令此处为 1,否则为 -1。显然, $t,d$ 与  $w$  的取值和初始轮廓的效果有关。取几组  $d,w,t$  的值进行比较观察可得,两个参数的变化对最后算子的影响,可以进行如下定性的分析: $t$  越大,表明算子中比较的阈值越大,太小或太大都区别不了背景与文本; $w$  与  $d$  越大,表明算子中相互比较的两个像素的间隔就越大,那么就可能导致最后图像的边界出现虚空,边界不完整。如图 5 所示,可直观看出,使用  $t=30,w=3,d=3$  是比较好的参数设置。

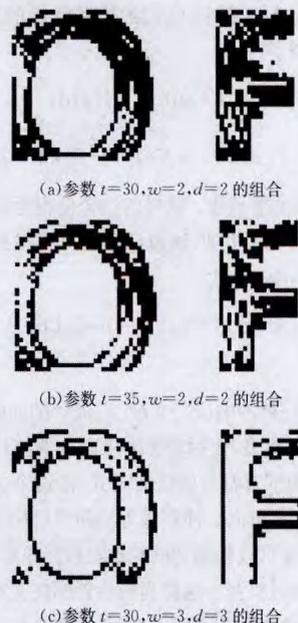


图 5

### 3.3 图像后处理

通过如下两点后处理排除非文本区域,使得检测更加准确:1)长度、宽度小于一定值的被检测图像块很难被人眼识别,基本可以排除是文本区域的可能性,如图 6 区域 a 所示。长宽大于一定值的被检测图像块,不符合图像中文本的体征,

可以排除是文本区域的可能性,如图6区域b所示。2)轮廓达到最大或最小长宽值,即轮廓连向边缘的可以排除是文本区域的可能性。本文后处理中所选取的参数为  $maxscale_x = 2/3$ ,  $minscale = 25$ ,  $xdis = 20$ ,  $ydis = 10$ 。其中  $maxscale_x$ ,  $minscale$  为最大最小宽度值;  $xdis$ ,  $ydis$  为进行轮廓合并时判定相邻文本区是否为文本行的距离依据,如图6区域c所示。其中区域a为长度宽度同时小于  $M/25$ ,  $N/25$  的备选区域,是非文本区;区域b为长度大于  $2/3 * M$  或宽度大于  $2/3 * N$  的备选区域,是非文本区;区域c为两个文本之间的水平距离小于  $xdis$  同时两个文本之间的垂直距离小于  $ydis$  的备选文本区,最终作为一个文本行连接在一起。

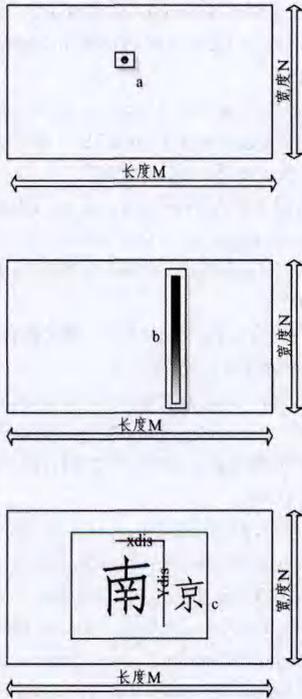


图6 后处理示意图

## 4 实验结果

### 4.1 数据集

本文选择了ICDAR2005(International Conference On Document Analysis And Recognition)<sup>[12]</sup>复杂彩色图像的文本定位竞赛中公开的图像库进行实验。ICDAR图像库包含249幅通过数码相机拍摄得到的场景图像,其内容以门牌、广告、指示标识为主。图像中的文本包括各种不同的字体、颜色、大小、对比度等,背景具有不同的复杂度,语言是英语。字符的大小变化极大,小的小于10个像素高,大的大于800个像素高。

### 4.2 本文提出检测方式的特点

文本检测大部分是文本分割、文本识别的准备工作,以往的检测均为以框出文本行为目的,这样做能检测出文本,但并不利于后续工作的展开。本文提出的检测方式为了配合现行的评价标准,将原本单独分割开的文本通过一系列判断方法连在一起形成文本行。为了加强显示效果,将文本框颜色设为y,宽度设为2,单独文本框效果图如图7所示。

直接将SPF主动轮廓模型用于文本检测,效果并不理想,如图8(b)所示;使用本文提出的改进的轮廓模型后,定位

效果明显改善,实验结果如图8(c)所示;为便于进行算法评估,将轮廓合并,结果如图8(d)所示。



图7 单独文本框效果(左为原图,右为实验结果图)

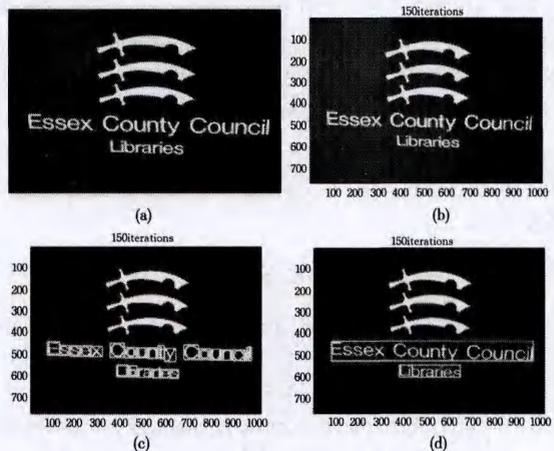


图8 算法改进效果图

### 4.3 实验结果与评估

本文选择了文献[4]中使用的区域级评估方法进行评估,以下为评估标准。

正确检测区(TDB):一个被检测到的区域至少包含一个正确字符。因此,一个TDB有可能完全覆盖一个文本行。

错误检测区(FDB):一个被检测到的区域没有包含文本。

不完整检测区(MDB):一个被检测到的区域丢失一行文本20%以上的字符(MDB是TDB的子集)。根据文献[13],文本区域覆盖检测80%以上的正确区域被认为是准确的。

对于数据集中的每幅图片,手动计数实际文本区(ATB)的数量。

$$Recall(R) = TDA/ATB$$

$$Precision(P) = TDB/ATB$$

$$F\text{-measure}(F) = 2 * P * R / (P + R)$$

$$Mis\text{detection Rate}(MDR) = MDB/TDB$$

为了比较结果,本文采用了4种现存的方法进行对比。文献[1]中方法用Edge1表示,从4个Sobel边缘图中提取了6种统计特征。文献[3]中方法用Edge2来表示,在YUV色空间采用Sobel算子检测。文献[2]中方法用Color表示,在

Lab 色彩空间采用分层聚类来定位统一颜色的文本行。对同一个图像库使用统一的参数检测,对比的文章使用的参数均根据其推荐值确定,本文所提出算法的参数根据经验设定。实验结果如表 1 所列。

表 1 ICDAR 图像库实验结果

Method	R	P	F
Edge1	0.53	0.61	0.57
Edge2	0.67	0.33	0.44
Color	0.60	0.44	0.45
本文提出算法	0.77	0.58	0.62

注:R-召回率,P-精确度,F-综合指数。

通过表 1 中结果可得知,本文提出的基于主动轮廓模型的方法有较好的检测效果,因为在 ICDAR 库中其召回率有较突出的表现。本文在精确度方面还有上升空间,但是由于文本与非文本来自同一个场景,消除非文本时很容易将文本一同删掉,因此这并非易事。

**结束语** 本文提出了一种主动轮廓模型的文本检测方法,该方法在自然场景和人工场景中均可应用。所提出的方法的初始轮廓不再是一个简单的矩形框,而是让初始轮廓就能够反映图像中的文本信息。这样做不但能得到较好的检测效果,而且检测后将文本单独框出,有利于后续分割识别的进行,使整个文本检测识别过程更加紧密。sobel-laplacian 与 gaussian-laplacian 两步骤的文本预处理,使文本与非文本界限突出,提升了检测结果。实验结果表明,本文所提出的方法效果良好。

## 参考文献

[1] Liu C, Wang C, Dai R. Text detection in images based on unsupervised classification of edge-based features [C]// Document Analysis and Recognition. 2005;610-614

[2] Mariano V Y, Kasturi R. Locating uniform-colored text in video frames [C]// 15th International Conference on Pattern Recognition. 2000;539-542

[3] Cai M, Song J, Lyu M R. A new approach for video text detection [C]// 2002 International Conference on Image Processing.

(上接第 281 页)

## 参考文献

[1] Dinh H T, Lee C, Niyato D, et al. A survey of mobile cloud computing: architecture, applications, and approaches [J]. Wireless communications and mobile computing, 2013, 13(18):1587-1611

[2] Garg S K, Versteeg S, Buyya R. A framework for ranking of cloud computing services [J]. Future Generation Computer Systems, 2013, 29(4):1012-1023

[3] Iosup A, Epema D. On the Gamification of a Graduate Course on Cloud Computing [C]// The International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE. 2013

[4] Venkata Krishna P. Honey bee behavior inspired load balancing of tasks in cloud computing environments [J]. Applied Soft Computing, 2013, 13(5):2292-2303

[5] Ryan M D. Cloud computing security: The scientific challenge, and a survey of solutions [J]. Journal of Systems and Software, 2013, 86(9):2263-2268

2002;117-120

[4] Shivakumara P, Phan T Q. A laplacian approach to multi-oriented text detection in video [J]. IEEE Transactions on Software Engineering, 2011, 33(2)

[5] Zhang Kai-hua, Zhang Lei, Song Hui-hui. Active contours with selective local or global segmentation: a new formulation and level set method [J]. Image and Vision Computing, 2010, 28(4):668-676

[6] Zhu G P, Zhang S H Q, Zeng Q S H, et al. Boundary-based image segmentation using binary level set method [J]. Optical Engineering, 2007, 46(5):050501

[7] Li C M, Xu C Y, Gui C F, et al. Level set evolution without re-initialization: a new variational formulation [C]// IEEE Conference on Computer Vision and Pattern Recognition. San Diego, 2005;430-436

[8] Li C M, Kao C, Gore J, et al. Implicit active contours driven by local binary fitting energy [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2007

[9] Xu C Y, Yezzi J R A, Prince J L. On the relationship between parametric and geometric active contours [C]// Processing of 34th Asilomar Conference on Signals Systems and Computers. 2000;483-489

[10] 刘海波, 沈晶, 岳振勋. Visual C++ 数字图像处理技术详解 [M]. 北京:机械工业出版社, 2014

Liu Hai-bo, Shen Jing, Yue Zhen-xun. Digital image processing using visual C++ [M]. Beijing: China Machine Press, 2014

[11] 蒋先刚. 数字图像模式识别工程软件设计 [M]. 北京:中国水利水电出版社, 2008

Jiang Xian-gang. Digital image pattern recognition engineering software design [M]. Beijing: China WaterPower Press, 2008

[12] Lucas S M. ICDAR 2005 text locating competition results [C]// Eighth International Conference on Document Analysis and Recognition, 2005(ICDAR 2005). 2005

[13] Chen D, Odobez J M, Thiran J P. A localization / verification scheme for finding text in images and video frames based on contrast independent features and machine learning [J]. Image Communication, 2004, 19:205-217

[6] Szymanski T H. Low latency energy efficient communications in global-scale cloud computing systems [C]// Proceedings of the 2013 workshop on Energy efficient high performance parallel and distributed computing. ACM, 2013;13-22

[7] Pérez O, Amaya I, Correa R. Numerical solution of certain exponential and non-linear Diophantine systems of equations by using a discrete particle swarm optimization algorithm [J]. Applied Mathematics and Computation, 2013, 225:737-746

[8] Mandal D, Kar R, Ghoshal S P. Digital FIR filter design using fitness based hybrid adaptive differential evolution with particle swarm optimization [J]. Natural Computing, 2014, 13(1):55-64

[9] Belmecheri F, Prins C, Yalaoui F, et al. Particle swarm optimization algorithm for a vehicle routing problem with heterogeneous fleet, mixed backhauls, and time windows [J]. Journal of intelligent manufacturing, 2013, 24(4):775-789

[10] Katherasan D, Elias J V, Sathiyaraj P, et al. Simulation and parameter optimization of flux cored arc welding using artificial neural network and particle swarm optimization algorithm [J]. Journal of Intelligent Manufacturing, 2014, 25(1):67-76