

doi:10.16652/j.issn.1004-373x.2016.04.028

一种基于角点与BP神经网络的文本检测及定位算法

唐思源, 高琦, 邢俊凤

(内蒙古科技大学包头医学院 计算机科学与技术系, 内蒙古 包头 014040)

摘要: 检测并提取视频图像中的文本信息对视频图像和内容的理解意义重大。以现有的文本检测算法为基础, 提出一种基于角点与BP神经网络相结合的文本检测算法。该算法首先应用多尺度角点算法提取文本角点信息并初步定位文本行, 接着提取文本特征, 最后应用BP神经网络精确定位文本。实验结果表明, 此算法与经典方法相比具有更高的正确率和鲁棒性, 视频中文本的正确检测率达到90.3%。

关键词: 文本检测算法; 多尺度角点算法; BP神经网络; 文本信息提取

中图分类号: TN911-34; TP317.4

文献标识码: A

文章编号: 1004-373X(2016)04-0112-04

Text detection and location algorithm based on angular-point and BP neural network

TANG Siyuan, GAO Qi, XING Junfeng

(Department of Computer Science and Technology, Baotou Medical College, Inner Mongolia University of Science and Technology, Baotou 014040, China)

Abstract: The detection and extraction of text information in video image has important significance on understanding the video image and content. On the basis of the current text detection algorithm, an text detection algorithm based on the integration of angular-point and BP neural network is proposed. The multi-scale angular-point algorithm is used to extract the text angular-point information and primarily locate the text line, and then extract the text features. Finally the text is precisely located by BP neural network. The experimental results prove this method has higher accuracy and robustness in comparison with the classical method. The correct detection rate of text in the video can reach up to 90.3%.

Keywords: text detection algorithm; multi-scale angular-point algorithm; BP neural network; text information extraction

0 引言

在数字视频中, 常会有一些文本的信息内容, 这些文本信息大部分都是对视频内容起到补充及重要内容提示的作用。比如, 对人物姓名, 产品介绍, 交通指示标牌等文字的说明, 这些文字说明有助于人们更好地理解和分析相关的视频内容。因此, 正确地检测并提取出视频中的文本信息, 对图像的理解、分析有重要的研究意义。在对视频帧图像进行文本检测时, 选取一个合适的文本区域检测方法非常重要。因为在视频帧中, 文字与其他背景图像存在特征差异, 所以, 研究者们用得最多的方法是基于特征的方法。可以根据文字特有的特征, 比如: 文字的排列多为水平、垂直方向, 颜色多为单色, 文字的大小是在肉眼观察范围内, 以及文字的笔画比较丰富等来检测并定位文本区域。研究者们可以根据这些特征, 利用基于边缘检测或基于角点检测^[1]等方法来

确定文本区域。但应用上述等方法只能检测出一些规范条件的文本, 随着制作视频的复杂及处理图像的进步, 视频中的文本也越来越多样化, 基于特定规则的传统方法已经无法检测出比较复杂的文本, 所以, 本文提出了应用基于特征的方法与基于学习的方法相结合的思想来更好地检测文本区域。基于学习的方法^[2]是通过训练学习来获取可靠的参数模型, 通过此模型, 可以检测出在变化因素, 如不同尺寸、字体、语种下的文本区域, 从而提高检测结果的精度, 同时可以增加训练样本, 扩大学习区域来扩充检测范围, 提高检测率。

1 视频中文本的检测与定位

本文首先应用基于角点的方法^[3]对潜在文本区域进行粗检测, 由于采集的视频中, 主要是中文文字, 中文的文字笔画数比较多, 具有大量的角点信息, 而且相同的文本区域中字体和大小都是统一的, 应用该方法能生成准确的候选文本区域。由于背景复杂等原因, 有些和背景相似的文字区域可能遗漏, 本文应用BP神经网络^[4]实现文本区域的精确检测, 精确定位出文本区域。视频中

收稿日期: 2015-08-19

基金项目: 内蒙古自治区高等学校科学研究项目(NJZY215)

文本的检测与定位算法结构图如图1所示。

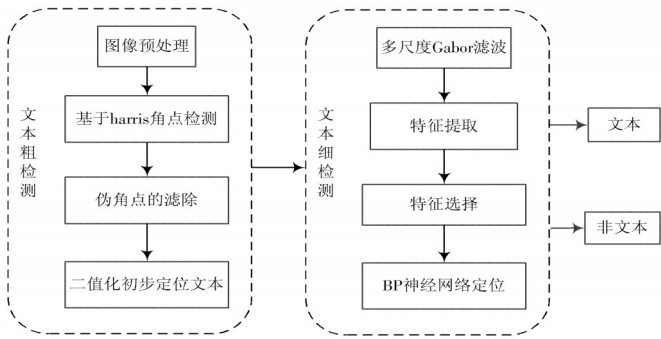


图1 视频中文本的检测与定位算法结构图

1.1 基于角点检测的粗定位算法

角点和边缘是图像比较明显的特征,视频图像中文字笔画的边缘及角点信息比较突出,本文通过检测视频图像中角点信息来粗略地定位文本区域。

1.1.1 角点检测算法的选取

比较经典的角点检测算法有: moravec, susan, mic 和 harris 等^[5], 本文选取 harris 角点算法来检测图像中的角点。传统方法中检测图像的角点信息完全通过阈值来判断, 阈值大了部分角点会检测不到, 阈值小了会把一些非角点信息检测成角点。为了改进传统的算法, 本文提出了多尺度角点检测方法, 该方法把图像分割成不同尺度的图像, 在不同尺度的图像上检测角点信息, 打破了单靠阈值来检测角点的局限性。通常, 尺寸小的参数可以检测出细小的灰度变化, 能反映一些特殊点的信息, 但对噪声比较敏感; 而尺寸大的参数可以检测出差异大的灰度变化, 能反映一些剧烈变化点的信息, 对噪声不敏感。

1.1.2 多尺度角点算法实现

本文获取的图像都是彩色图像, 彩色图像色彩变化大, 获取角点比较困难。首先将彩色图像转化成灰度图像。本文应用平均值法将彩色图像转化成灰度图像, 如图2所示。

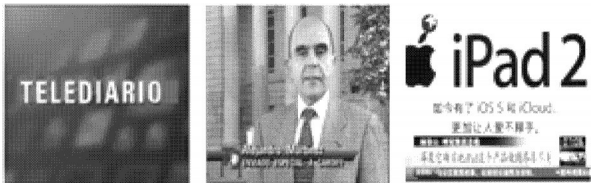


图2 灰度化处理后的图像

多尺度角点检测算法流程图如图3所示。角点检测后图像的角点分布图如图4所示。

1.1.3 伪角点的滤除

由于文本区域角点分布比较集中、有一定规律, 而非文本区域的角点分布杂乱无章。根据文本的特征, 本

文通过角点滤波算法^[6], 去除伪角点。

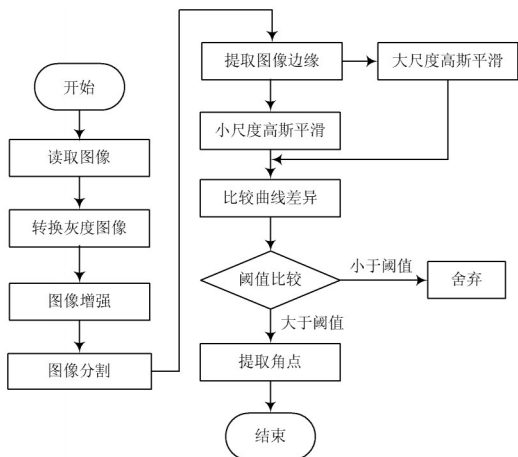


图3 多尺度角点检测算法流程图

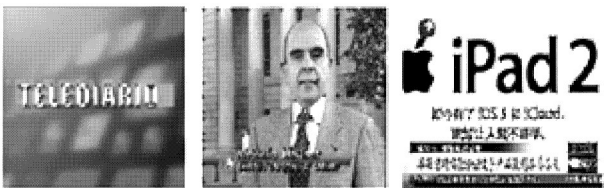


图4 角点分布图

其步骤如下：

Step1: 文本信息基本都是按行排列, 本文设定阈值 $n=5$ 。用 $corners(i)$ 表示统计后每一行角点的数目, 如果 $corners(i) \leq N$, 则第 i 行为非文本角点, 否则为文本行角点。

Step2: 有一些成片的干扰噪声, 通过模板运算去除。定义2个模板: M_1 为 8×10 , M_2 为 8×20 , 用 N_i 表示每个模板对应的阈值, 本文设 $N_1=3, N_2=4$ 。用 $corners(P, M_i)$ 表示在角点 P 处落在模板 M_i 里的总角点数, 如果 $corners(P, M_i) < N_i$, 证明角点 P 处不是文字角点, 反之则为文字角点。

如图5所示为经过角点滤波算法去除伪角点的角点分布图, 根据去除前后对比, 大部分伪角点信息都被去除, 证明该算法是合理可行的。

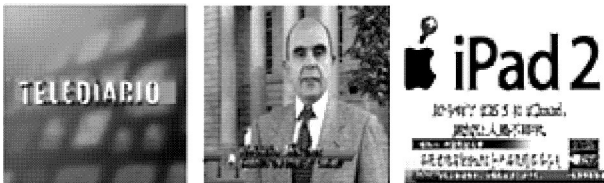


图5 滤除伪角点分布图

1.1.4 二值化处理

图像二值化^[7]是把图像分割成只有背景(黑色)和对象(白色)的一种操作, 通过设定阈值来对图像进行二值化。本文通过反复试验的方法, 来选取阈值, 直到图像

产生比较好的区分结果为止。通过创建一个图形控件来改变阈值的大小,确定阈值后,像素灰度值与阈值之差小于30,将像素置为0,即为黑色,否则置为255,即为白色。处理后的结果如图6所示。

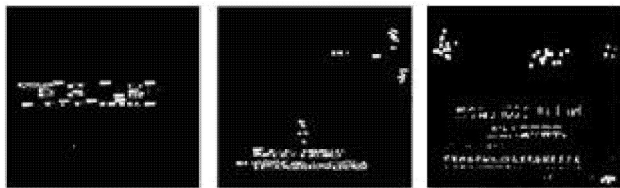


图6 二值化后结果图

1.1.5 文本区域的粗定位

经过二值化处理后,可以观察到文本区域的角点比较分散,为了将角点聚集成块,对其进行形态学运算^[8]。通过形态学的膨胀运算将分散的角点连成片,再进行腐蚀运算去除周边及内部的孤立点,如图7所示,经过粗定位后,已经去除了大部分背景区域,也尽可能多地留下文本区域,而且在一定程度上保证了文本区域不会被漏掉。

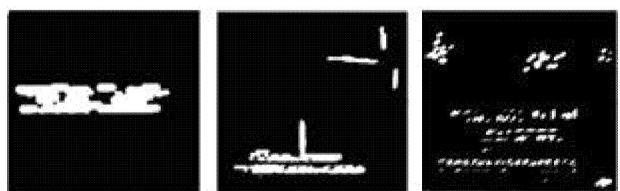


图7 膨胀腐蚀后处理结果图

1.2 基于BP神经网络文本区域的精确定位

在完成文本区域的粗检测后,有一些和文字笔画及特征相似的区域被误判为文本区域。本文继续用纹理分析的方法去掉干扰的背景区域。最后经过BP神经网络的鉴别,真实的文本将被定位出来。

1.2.1 Gabor滤波及文本特征的提取

Gabor^[9]是一种窄带带通滤波器,主要是在方向上提取图像的纹理特征,特别对文本的笔画方向比较敏感。一般文字的笔画特征比较丰富,笔画方向基本都集中在垂直、水平、两个对角线方向,且表现出一定的周期性,文字的笔画可以看成是一种特殊的纹理,利用Gabor滤波器提取文字方向特性并结合一些其他的纹理特征可以很好地将文本与背景分开。

本文把图像分成4×4个子区域,应用多尺度Gabor滤波器对候选区的4个方向进行滤波。对每个子区域的候选区域提取均值、方差、能量、熵、惯量和局部统一性6个纹理特征^[10]。计算Gabor特征向量为4×4×6=64维;同理,提取6个纹理特征的向量为4×4×6=96维,对这两类特征进行归一化后,将这两类特征级联就可以得到最终

文本特征向量,该特征向量的维数为160维。

1.2.2 特征选择

一共提取了160个特征,但并不是把所有提取的特征都放入特征训练集中,有一些特征相近的特征值选取其中一到两个放入训练集中就可以,这样可以减少特征提取的时间。本文通过生成图像的傅里叶变换,并对变换数据进行分类,以便得到一组量度。通过量度值来选择文本的纹理特征,经过测试一共选取24个特征来训练神经网络。

1.2.3 BP神经网络定位及分类

BP神经网络^[11]过程图如图8所示。

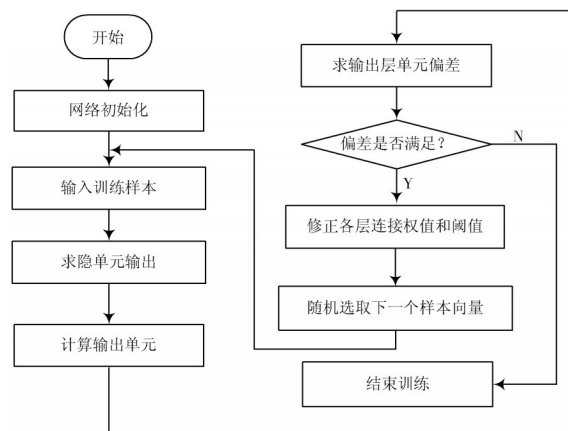


图8 神经网络训练过程图

在本课题的实验中,应用三层BP神经网络结构实现对文本的定位。BP网络输入层和输出层的神经元个数由试验得出均为5;隐含层神经元数目的确定一直没有通用的方法。因此,设计一个可以在一定范围内调整隐层神经元个数的BP网络结构,采用函数trainidx对网络进行训练,该函数的学习算法是梯度下降动量法,而且学习速率是自适应的。通过误差对比,确定最佳的隐含层神经元个数,得出网络的训练误差,如表1所示。

表1 网络训练误差

神经元个数	12	15	19	21	24	26	27
网络误差	1.041 2	0.729 4	0.176 7	0.144 9	0.180 7	0.144 2	0.162 1

表1表明,在经过2 000次训练后,隐含层神经元为26的BP网络对函数的逼近效果最好,它的误差最小,而且网络经过232次训练就达到了目标误差。在BP反向传播学习算法中,初始值采用随机分配值的方法,这样做是为了取得分布均匀的小数,但是这样就为测试带来了不便,因为没有有一个基准点进行比较。因此,本文对误差反向传播训练算法进行改进,根据文本的特征,采取如下测试方法:

取50次学习平均值,运行程序50次,设定学习速率0.5,动向量0.5,输入24节点,输出2节点,误差 $e<0.000\ 01$,学习结束。本文共收集样本161份,其中彩色图像55份、视频图像52份、场景图像54份。在彩色图像中取35份、视频图像中取27份、场景图像中抽取25份作为训练的样本集,其余部分用作测试样本集,使用改进后的BP算法进行试验对比,括号内的是改进的算法结果,如表2所示。

表2 BP网络学习结果对比表

学习速率	动向量	学习次数	训练样本正确识别率/%	测试样本正确识别率/%	总体样本识别率/%
0.3	0.3	657	81(93.1)	61(82.4)	142(88.1)
0.3	0.5	512	80(91.9)	63(85.1)	143(88.8)
0.3	0.8	400	83(95.4)	64(86.4)	147(91.3)
0.5	0.3	465	82(94.2)	62(83.8)	144(89.4)
0.5	0.5	413	84(96.6)	65(87.8)	149(92.5)
0.5	0.8	336	81(93.2)	64(86.4)	145(90.1)

由表2可知,改进后的BP算法的总体样本识别率及正确识别率都有所提高,表明改进的算法的网络泛化能力比较好。

2 实验结果

本文在Matlab 7.0开发环境中实现,实验数据集中随机选取新闻类、不规则文本类和广告类三类,视频数据图像共500余张,在视频文本定位中通过查全率 R 和精确率 P 来度量定位算法的性能^[12]:查全率 R =正确检测到的文本数/视频中的文本总数;精确率 P =检测正确文本区域数/检测文本区域总数。经过实验后,对前后改进的定位算法数据进行了对比,其结果如表3所示。

表3 改进前后算法检测结果比较

数据类别	定位方法	正确行数	漏检行数	误检行数	查全率/%	精确率/%
新闻类	改进前	42	3	4	94.9	91.3
	改进后	44	1	2	97.1	95.6
不规则文本类	改进前	51	9	6	85	89.4
	改进后	56	4	2	93.3	96.5
广告类	改进前	45	9	11	83.3	80.3
	改进后	49	5	6	90.9	89.1

由表3分析测试结果可见,该算法可以提取出大部分的文本区域,特别是新闻类视频文本具有较高的准确性,从数据看出,本文改进的算法对广告类视频定位的准确率同样优于改进前的定位结果,改进前与改进后部

分对比图如图9、图10所示。

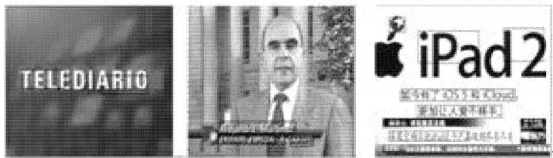


图9 改进前视频中的文本检测结果



图10 改进后视频中的文本检测结果

由图9、图10可知,改进后的定位算法很好地提取了大小不一的字体,避免了较大字符的漏检,提高了定位精度,同时提高了对背景复杂的文字区域的检测精度,但也存在对模糊文本区域的漏检情况。实验表明,本文提出的算法对于检测定位文本区域具有良好的准确性和鲁棒性,平均正确的检测率达到了90.3%。

参 考 文 献

[1] WU H C, CHANG C C. A novel digital image watermarking scheme based on the vector quantization technique [J]. Computers & security, 2005, 24(6): 460-471.

[2] LIANG C W, CHEN P Y. Dwt based text localization [J]. International journal of applied science and engineering, 2014, 2 (1): 105-116.

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

[4] 王学武,谭得健.神经网络的应用与发展趋势[J].计算机工程与应用,2003,39(3):98-100.

[5] 陈又新,刘长松,丁晓青.复杂彩色文本图像中字符的提取[J].中文信息学报,2003,17(5):55-59.

[6] 章毓晋.图像处理与分析技术[M].北京:清华大学出版社,2000.

[7] 王晓丹.基于Matlab的系统分析与设计[M].西安:西安电子科技大学出版社,2000.

[8] 杨淑莹.图像模式识别[M].北京:清华大学出版社,2005.

[9] 冈萨雷斯.数字图像处理[M].北京:电子工业出版社,2004.

[10] 郝文化,董秀芳.Matlab图形图像处理应用教程[M].北京:中国水利水电出版社,2004.

[11] 欧文武,朱军民,刘昌平.视频文本定位[J].计算机工程与应用,2004,40(30):65-67.

[12] 陈又新,刘长松,丁晓青.复杂彩色文本图像中字符的提取[J].中文信息学报,2003,17(5):55-59.

[13] 张引,潘云鹤.面向彩色图像和视频的文本提取新方法[J].计算机辅助设计与图形学学报,2002,14(1):36-40.

作者简介:唐思源(1981—),女,硕士研究生,讲师。研究方向为数字图像处理。
高琦,讲师。研究方向为计算机应用。
邢俊凤,副教授,硕士。研究方向为计算机应用。