

一种改进的小波变换域的字符定位方法

张 阳, 王嘉梅

(云南民族大学 电气信息工程学院, 云南 昆明 350011)

摘 要: 在分析图像中字符特点的基础上,研究其在小波变换域中的特征,依据这些特征,分析频率加权阈值法和利用块间信息的平均阈值法的优缺点,将二者结合进行图像字符定位。实验表明,该算法可进行图像中的字符定位,同时又可以弥补两种方法定位时的不足。

关键词: 字符定位; 小波变换; 加权系数

中图分类号: TP391.4

文献标识码: A

文章编号: 1674-7720(2011)18-0035-03

An improved algorithm of character localization based on wavelet transform domain

Zhang Yang, Wang Jiamei

(School of Electrical and Information Technology, Yunnan University of Nationalities, Kunming 350011, China)

Abstract: On the basis of analyzing characteristics in images, the paper studies its wavelet transform domain. According to these characteristics, the paper analysis the advantages and disadvantages of frequency weighted threshold value method and utilization of information between blocks average threshold value method, and combines them to do the image character positioning. Experiments show that this algorithm can undertake image of character positioning; At the same time, also can make up for the two methods lack when positioning.

Key words: character localization; wavelet transform; weighted frequency

随着计算机性能的提升和数字化存储技术的不断发展,从海量的视频资料中找到想要的信息已经成为十分迫切的需求,所以对视频内容的分析、建立基于内容的索引以及视频内容的结构化描述成为近年来研究的热点。字符的分析与处理是视频分析的重要内容,进行图像字符分析处理的基础就是探测图像中字符的出现及定位其准确坐标。

目前对图像中的字符定位方法大致可以分为两大类:一类是在原像素域内的处理方法^[1-2],其主要是根据字符的纹理特征或结构特征来定位字符;另一类是基于压缩域的处理方法^[3-5],主要是 DCT 变换域和小波域,主要依据字符的笔划边缘在频域中表现较大的系数和方向性,利用这些系数来进行字符定位。在压缩域中,这些方法能够充分利用图像中的字符在频率变换域中的特点,且只需要进行简单的解码,其处理的速度较快,具有较高的正确检测率,是目前研究的热点。参考文献[6]提出了小波变换域的图像字符定位方法,并通过对大量图像字符定位的研究表明,该方法适应性强,但处理背景复杂的图像效果稍微差一些。而参考文献[7]在定位时

采用加权频率的方法,在对背景复杂度高的图片进行定位时有较好的效果,但是该方法处理一些较小的英文字符时漏检较多。本文在分析参考文献[6-7]的基础上,先对小波变换系数进行非线性加权,然后再采用自适应动态阈值及结合块间信息来提高检测的正确率及算法的适应性。

1 小波变换域能量表征及分类

字符可能出现在图像和视频中的任何地方,但并不是所有出现的字符对视频和图像都有内容上的相关。将视频中出现的字符分为人工字符和背景字符两类^[1]。人工字符本身有特定的笔划结构,在空间上表示为水平、竖直以及对角线方向的笔划。字符灰度变化不大,但边缘变化较大,有突出的水平、竖直以及斜向的纹理特征。本文主要研究人工字符定位,利用人工字符的这些特征可以将字符区间和背景分开,对字符进行定位。

对图像信号进行小波变换,可以将其分解为平均部分、垂直细节部分、水平细节部分以及对角线细节部分,这些刚好对应于图像中字符的笔划特点。在小波变换域上进行字符定位的主要步骤就是给出标志字符中笔划

图形、图像与多媒体

走向对应的频率变化的能量。设图像 $f(i,j)$ 经小波变换后对应的小波系数为 $C(i,j)$, 如图 1 所示。

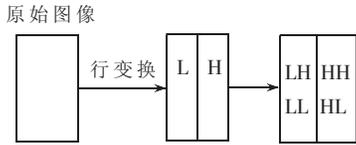


图 1 图像小波分解及区域定义

图像中一块区域 R 经变换后得到 3 个高频区域, HH、LL、HL 的能量定义如式(1)所示, 即对块中小波系数求和。

$$E = \sum_{i,j \in \Omega} |C(i,j)| \quad (\Omega \text{ 为 HH、LL 或 HL}) \quad (1)$$

2 小波变换域中确定图像字符坐标方法

将带有字符的图片进行小波变换, 得到对应于平均、水平、竖直和对角线方向的 4 个分解图, 其中每个分解图的行和列为原图的一半。

本文在设定阈值前对变换的小波系数进行非线性加权:

$$E = \sum_{i,j \in \Omega} |4 \times i \times j \times C(i,j)| \quad (2)$$

其中, i, j 为单元块内的行和列。

利用字块的平均信息定义各个块的能量, 对于水平方向能量只在水平方向上进行平均, 水平方向能量定义为:

$$E_h = \sum_{m=-2}^2 E(i+m,j)/5 \quad (3)$$

水平方向能量分布如图 2 所示。



图 2 水平方向能量分布图

类似地, 竖直方向能量定义为:

$$E_v = \sum_{m=-2}^2 E(i,j+m)/5 \quad (4)$$

竖直方向能量分布如图 3 所示。

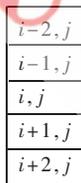


图 3 竖直方向能量分布图

对角线方向能量定义为:

$$E_x = \sum_{m=-1}^1 \sum_{n=-1}^1 E(i+m,j+n)/5 \quad (m \neq 0 \text{ 或 } n \neq 0) \quad (5)$$

对角线方向能量分布如图 4 所示。

采用自适应阈值法进行分类, 阈值公式为:

$$W.Threshold = Aver + (Max - Min) \times k \quad (6)$$

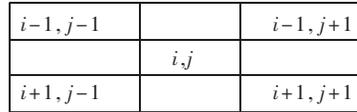


图 4 对角线方向能量分布图

其中, $Aver$ 为该方向上的平均能量, Max 和 Min 分别为该图像内最大的块能量和最小的块能量, k 为经验值。

字符主要是由水平和竖直的笔划组成, 对角线方向笔划较少, 即对角线方向能量较小, 字符只在统计意义上具有一定的对角线方向能量^[4]。具体的算法步骤如下:

- (1) 对输入图像进行小波分解;
- (2) 将小波分解后的子图划分成 4×4 的单元块并计算单元块的能量;
- (3) 计算各个方向的阈值;
- (4) 选出单元块水平能量或者竖直能量大于阈值的单元块作为候选目标字符区;
- (5) 采用形态学滤波的方法去掉单个的噪声点, 并填充候选区域;
- (6) 用最小矩形框表示出候选区域, 求其平均对角线方向能量, 大于阈值则确定为字符区, 否则删除;
- (7) 对选出的字符区进行水平、竖直方向投影, 去除与字符粘连的非字符背景。

使用上述步骤进行定位的过程如图 5 所示。



图 5 字符定位过程示意图

3 实验结果及讨论

实验中选择的图像分别来自视频截图以及 Web 图像等, 水平和竖直方向的实验参数 k 取 0.2~0.23 之间的值效果较好, 对角线方向 k 取 0.13 效果较好。三种方法的检测率比较结果如表 1 所示。

《微型机与应用》2011 年 第 30 卷 第 18 期

表 1 三种方法检测比较结果

	漏检	误检	检测率/%	误检率/%
参考文献[6]方法	43	58	91.60	16
参考文献[7]方法	49	54	92.30	15.10
本文方法	28	49	92.50	13.70

部分实验定位效果如图 6 所示, 实验中选择的图像共包含 360 个字符区, 其中实际检测到的字符区 332 个, 正确检测率为 92.5%; 未检测到的字符区 28 个, 漏检率为 5.8%; 检测到的错误字符区 49 个, 错误检测率为 13.7%。参考文献[6]的方法主要针对背景复杂的图像, 采用平均能量的方法使字符区间的能量削弱了, 从而产生误检。参考文献[6]的误检主要来自于背景复杂度较高的图像, 而其在定位时, 小于 8 像素的英文字符在定位过程中会被误认为是噪点删除; 在图像中出现多个英文字符区域时, 该算法优先选择了斜向能量较大的几个字符区域, 一些字符区域被视为背景, 造成漏检。



图 6 字符区定位部分结果

改进后的方法对小波系数进行了非线性加权, 然后进行块间能量平均, 在背景复杂度较高的图像处理上, 该方法定位效果比较好; 在处理一些包含多字符区域的图像或者包含低于 8 像素的英文字符区域时, 该方法能很好地完成定位。从实验结果来看, 对于尺寸很大或者

字符间隙很大的字符还是没能很好地定位。当一些非文本区域具有与文本相似的纹理时, 本算法会产生虚警, 文本定位的效果不好。这些压缩域字符定位的通病, 也是今后着重研究的问题。

本文提出了一种改进的小波变换域的图像中字符定位的方法, 该方法在充分利用字符笔划在小波变换域中的特点的基础上, 对小波系数进行加权, 突出了字符区间的能量。算法实现简单, 具有较强的适应性。实验结果表明, 本文算法具有较高的正确检测率, 总体而言, 相对于视频、图像上人工字符定位来说, 效果还是比较理想的。今后要进一步研究字符分割及识别技术。

参考文献

- [1] 王辰, 松杨, 胡晓峰. 视频中的文字探测[J]. 小型微型计算机系统, 2002, 23(4): 478-481.
- [2] WU V, MANMATHA R, RISEMAN E M. An automatic system to detect and recognize text in images[C]. IEEE Transaction on Pattern Analysis and Machine Intelligence 1999:6.
- [3] ZHONG Y, HANG H J, JAIN A K. Automatic caption localization in compressed video[C]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000:8.
- [4] 黄祥林, 沈兰荪. 基于 DCT 压缩域的图像字符定位[J]. 中国图像图形学报, 2002, 7(001): 22-26.
- [5] 黄晓东, 周源华. 用小波变换及颜色聚类提取的视频图像内中文字幕[J]. 计算机工程, 2003(1): 43-44.
- [6] 张二虎, 张绪进, 张志刚. 小波变换域中图像字符的定位提取方法[J]. 应用科学学报, 2006(2): 135-139.
- [7] 孙惠平, 刘党辉, 沈兰荪. 基于 DCT 压缩域的快速字符定位算法研究[J]. 电子学报, 2006(4): 751-754.

(收稿日期: 2011-07-18)

作者简介:

张阳, 男, 1986 年生, 硕士研究生, 主要研究方向: 图像通信。

王嘉梅, 女, 1964 年生, 本科, 教授, 主要研究方向: 电子信息学科。