

EAN-13 条形码图像的识别

刘艳洋, 曹玉东

(辽宁工业大学 电子与信息工程学院, 辽宁 锦州 121001)

摘要:分析了 EAN-13 条形码的编码和译码方法, 基于 VC++ 平台设计并实现了条形码识别系统。主要包括两大部分: 第一, 对条形码图像的预处理操作。首先对条形码图像进行中值滤波和膨胀操作, 可以确保条形码图像中直线的连通; 然后再对图像进行腐蚀和均值滤波操作, 这样可以进一步消除图像中一些非条形码背景的干扰和影响, 使其易于之后进行的译码识别; 第二, 对预处理后的图像进行译码识别, 其思想是利用相似边距离来判别条形码字符的相似度, 再通过译码并最终得到条形码图像所代表的数字信息。实验结果表明: 该 EAN-13 条形码识别方法可以有效地解决条码印刷模糊、轻微变形等问题, 识别效果较好。

关键词:条形码识别; EAN-13; 图像预处理; 相似边距离法

中图分类号: TP311

文献标识码: A

文章编号: 1673-629X(2015)06-0202-04

doi: 10.3969/j.issn.1673-629X.2015.06.045

EAN-13 Bar-code Image Recognition

LIU Yan-yang, CAO Yu-dong

(School of Electronics and Information Engineering, Liaoning University of Technology,
Jinzhou 121001, China)

Abstract: Encoding and decoding methods for EAN-13 bar code are analyzed in this paper. Design and implement a bar code recognition system based on VC++ including two processing steps. The first step is to preprocess bar code image. Firstly, median filtering and dilation operation are adopted to make sure the line image is continuous. Then, erosion operation and mean filtering are used to clear clutter backgrounds. The second step is to decode the notation of preprocessed bar code image. Firstly, locate the bar code by computing distance of similar edges and then recognize the bar code by decoding and checking methods. The experimental results show that bar code recognition system developed resolves the problems of poor printing quality and a little distortion effectively.

Key words: bar code recognition; EAN-13; image preprocessing; similar edge distance method

0 引言

条形码技术已经被广泛用于工商业、邮政、交通运输、图书出版、仓储管理、医疗卫生、生产控制等诸多行业, 同时条形码是迄今为止最经济、实用的自动识别技术, 一维条形码的广泛应用源于它可靠准确、数据输入快、经济便宜、灵活实用等优点^[1]。

一维条形码是由一组规则排列的条、空以及对应的字符组成的标记^[2]。“条”和“空”分别指条形码图像中黑色的直线和白色的直线。条形码中条和空之间不同的组合表示不同的信息, 因此需要通过一定的方法来进行识别。文中主要研究一维条形码中的 EAN-13 码。

目前一维条形码识别已有许多方法^[2-3], 采用

Hough 变换和神经网络的识别方法^[4-5], 2010 年邓泽峰等提出了图像式一维条码识别的三重编码方法^[6], 徐益峰等提出了基于 EAN-13 条形码识别的改进算法^[7], 2011 年卫权岗等提出使用手机进行 EAN-13 码的识别及应用^[8], 扩大了条形码的应用范围。

在现有方法的基础上, 文中实现了一种条形码识别系统。首先对条形码图像进行预处理, 然后根据相似边距离来判别出条形码字符, 再通过译码、校验来识别条形码。这样将大大降低条形码印刷质量以及条形码污染等各种因素的干扰, 从而有效地提高条形码的识别率, 是一种高精度的条形码图像的识别方法。

收稿日期: 2014-07-31

修回日期: 2014-10-31

网络出版时间: 2015-05-06

基金项目: 国家自然科学基金资助项目(61272214); 辽宁工业大学教师科研启动基金(X201216)

作者简介: 刘艳洋(1989-), 女, 硕士研究生, 研究方向为多媒体信息处理; 曹玉东, 博士, 副教授, 通讯作者, 研究方向为图像处理与识别。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20150506.1645.032.html>

1 EAN-13 码的编码方法

EAN-13 是标准商品条码^[9],它是一种(7,2)码,即每个字符的总宽度为7个模块,由两个条和两个空交替组成。标准型 EAN 代码由13位数字组成,称为 EAN-13 代码,根据 EAN 规范^[10],这13位数字分别赋予了不同的含义,分别是3位前缀码,4位厂商代码,5位商品项目代码和1位校验字符。

EAN-13 条码字符集包括10个数字字符,即0~9。每个数字字符包含模块总数为7,是由二进制表示的,其中“1”表示条的模块,“0”表示空的模块。表1给出了 A、B、C 三个条码字符集的代表形式。

表1 条形码字符集表示

数字字符	A 子集	B 子集	C 子集
0	0001101	0100111	1110010
1	0011001	0110011	1100110
2	0010011	0011011	1101100
3	0111101	0100001	1000010
4	0100011	0011101	1011100
5	0110001	0111001	1001110
6	0101000	0000101	1010111
7	0111011	0010001	1000100
8	0110111	0001001	1001000
9	0001011	0010111	1110100

EAN-13 条码符号结构参见文献[2],包含模块总数为113个,EAN-13 条码的前置码不用条码表示,也不包括在左侧数据符中。条码字符构成方式包括奇排列和偶排列,正是这种构成方式决定了左侧数据符,即前置码之后的6位数字。

2 EAN-13 码图像的识别

通常,在图像识别之前,都需要图像预处理操作。不同的图像识别任务往往需要不同的图像预处理技术。一般比较常见的图像预处理技术包括图像归一化、二值化、滤波去噪、图像增强锐化、直方图均衡化等方法。

2.1 条码图像预处理

(1) 条形码图像的二值化。

图像的二值化就是将图像通过一定的阈值进行一定的对比之后变为黑白图像。因此在条形码图像中,分别用白色和黑色来表示条码线,用白色来表示背景部分。阈值的确定是二值化操作中最关键的步骤。抽象出对阈值起作用的因素,阈值可以由下面的函数表示:

$$R = F[(i, j), g(i, j), S(i, j)] \tag{1}$$

式中, R 表示阈值; (i, j) 表示像素; $g(i, j)$ 表示像素 (i, j) 的自身性质(例如灰度); $S(i, j)$ 表示像素 $(i,$

$j)$ 邻近区域内的某种性质。

根据条码图像灰度直方图的灰度范围,确定一个最佳的阈值。如果图像像素点的灰度值已经归一化,根据对条形码图像的反复测试得出:在[0.40, 0.60]范围内确定阈值 R 比较合适。因此,需要进一步比较阈值 R 和图像中像素点的灰度值大小,如果图像中像素点的灰度值大于阈值 R ,则该像素点的灰度值变为255,否则变为0。根据式(1)可得到条码图像的二值化图像 $g(i, j)$ 。

$$g(x, y) = \begin{cases} 0 & f(x, y) < R \\ 255 & f(x, y) \geq R \end{cases} \tag{2}$$

(2) 条形码图像的滤波。

利用中值滤波法^[11]可以去除图像中孤立的噪声点,将当前像素点与邻域内周围像素进行比较,假如差值超过模板中灰度变化范围的一半,就改变原来的像素值。例如,某一灰度值为 X 的像素点 (x, y) ,其去噪过程可表示为:

$$g(x, y) = \begin{cases} \text{NOT}(X), & |X - \bar{X}| > 0.5w \\ X, & \text{其他} \end{cases} \tag{3}$$

式中, \bar{X} 表示 X 的8邻域像素点灰度值的平均值;

w 表示图像灰度值的变化范围。

因此利用中值滤波法能够有效地滤除条形码图像中的小颗粒噪声及毛刺,同时又可以保留条形码的边缘信息,更有利于条形码图像的识别。

均值滤波也称为线性滤波^[12],主要方法为邻域平均法。均值滤波的基本思想就是利用邻域的均值来代替邻域中心的像素值。对于当前像素点 (x, y) ,通常选择 3×3 的模板,求模板中像素的均值。此时,当前像素点 (x, y) 的像素值为所求得的均值,公式表示如下:

$$g(x, y) = \frac{1}{m} \sum_{i=1}^m f(x, y) \tag{4}$$

式中, m 为该模板中包含当前像素在内的像素总数。

均值滤波可以起到抑制噪声的作用,可以消除非条形码边缘的直线对边缘定位的干扰。

(3) 条形码图像的形态学处理。

膨胀和腐蚀是数学形态学图像处理的基础,并且是两个互为对偶的运算。即膨胀是将目标图像扩大的过程,而腐蚀是将目标图像收缩的过程。

膨胀运算是合并图像周围的背景点^[13],使目标图像的边界向外扩张的操作。利用膨胀操作,可以消除目标图像中的小颗粒噪声,同时可以有效地填补图像分割后不连续的边缘和目标图像之间的空隙。

设当前图像为 F ,结构元素为 B ,当结构元素 B

的原点移动到图像中的某点 (x, y) 时, 把它记作 B_{xy} 。此时图像 F 被结构元素 B 膨胀的公式表示如下:

$$D = F \oplus B = \{x, y \mid B_{xy} \cap F \neq \emptyset\} \quad (5)$$

式(5)可以理解为, 当结构元素 B 的原点移动到 (x, y) 时, 如果在 B 中像素值为1的点的个数至少有1个时, 那么膨胀后图像上该点的像素值为1, 否则为0。实验中可以看到膨胀后条形码图像中白色直线中间的空隙能够很好地被填补。

腐蚀运算可以消除边界点, 并使目标图像的边界向内部收缩。利用腐蚀操作并选取足够大的结构元素可以腐蚀掉图像中的两个细小的连通。

设当前图像为 F , 结构元素为 B , 此时图像被结构元素 B 腐蚀的运算公式表示如下:

$$E = F \ominus B = \{x, y \mid B_{xy} \subseteq F\} \quad (6)$$

式(6)可以理解为, 当结构元素 B 的原点移动到 (x, y) 时, 如果 B 完全包含在 F 中, 则在腐蚀后的图像上该点为1, 否则为0。实验中可以看到腐蚀后条形码图像中黑白直线之间的细小连通很好地被腐蚀掉。腐蚀运算示例可参见文献[14]。

条形码图像是由许多某一高度的黑、白条直线组成的, 在其通过中值滤波和膨胀操作后, 可以确保条形码图像中直线的连通, 同时也消除了包含在直线中的小颗粒噪声, 然后再对条形码图像进行腐蚀和均值滤波操作, 这样便可消除一些非条形码背景的干扰和影响, 并且很好地保证了条形码区域的清晰可靠。

图1给出了条形码图像中值滤波前后的效果图。



(a) 滤波前的条形码 (b) 滤波后的条形码

图1 滤波前后的条码图像对比

2.2 条形码的译码

利用相似边距离法译码、纠错、识别字符的优点明显, 即使条形码符号的印刷质量欠佳, 甚至出现字体倾斜、字符扫描不均匀等问题, 并且理论值和测量值之间会有相对比较的偏差, 但利用相似边距离依然可以达到正确地译码并识别的目的。

(1) 译码算法。

设条码字符一个扫描行的连续条空宽度为 $c_1, c_2, \dots, c_{59}, (y_1, x_1), (y_1, x_2), \dots, (y_1, x_{60})$ 为一行扫描记录的坐标点, 则每行扫描的条空宽度分别为: $c_1 = x_2 - x_1, c_2 = x_3 - x_2, \dots, c_{59} = x_{60} - x_{59}$ 。逐行扫描整幅图像, 记录下满足条件的扫描行数及各扫描行变化点的横坐标值。

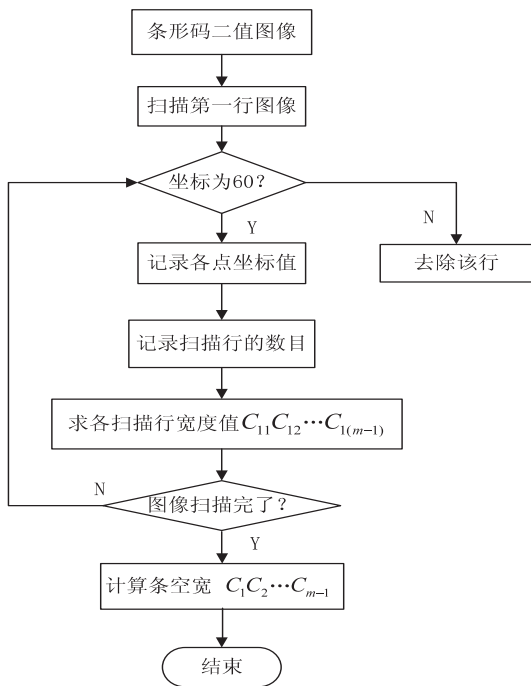


图2 条形码的条和空的计算流程图

设扫描行数为 n , 每个扫描行变化点的横坐标值为1到 m ($m = 60$), 则得到如下的 $n \times m$ 矩阵:

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}$$

(2) 条码条空宽度的计算。

系统利用平均值法来计算条码条空宽度, 与其他算法相比具有可靠性高、适用性广的优点。经过 n 行扫描后, 设条码连续各条空宽度为: C_1, C_2, \dots, C_{m-1} , 分别由以下各式计算:

$$C_1 = \frac{1}{n} \sum_{i=1}^n C_{i1}$$

$$C_2 = \frac{1}{n} \sum_{i=1}^n C_{i2}$$

$$\dots\dots$$

$$C_{m-1} = \frac{1}{n} \sum_{i=1}^n C_{i(m-1)}$$

综上所述, 由图像条码译码原理归纳条码条空宽度计算过程如图2所示。

2.3 条码字符的判别方法

在2.2节中已经描述了条形码的每一个条、空的计算方法, 得到了条形码的每一个条和空的宽度。如图3所示, 可知条码总的宽度为 $T = C_1 + C_2 + \dots + C_{m-1}$, 则每一个模块的宽度为 $mx = T/59$ 。每个字符是由两个条和两个空组成, 所以每个字符有三个相似边, 每个相似边的距离为:

$$T_1 = C_1 + C_2, T_2 = C_2 + C_3, T_3 = C_3 + C_4$$

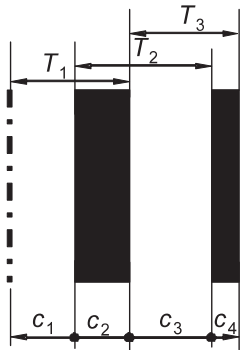


图3 一个字符的相似边距离

可以计算出每个相似边所包括的模块数目为 $A_1 = \lceil T_1/mx \rceil$, $A_2 = \lceil T_2/mx \rceil$, $A_3 = \lceil T_3/mx \rceil$ 。求出 A_1 、 A_2 、 A_3 后, 就可以根据它们不同的宽度模式得到这四个码条所对应的数字字符值。宽度模式对应的数字字符和编码规则如表 2 所示。条形码译码方法通过选择相似边距离归一化后的字符值, 识别效果比较明显并且可以得到比较高的识别率。

表2 EAN-13 条码编码与归一化值

字符值	左奇字符 (A) 编码	左偶字符 (B) 编码	右偶字符 (C) 编码	$A_1A_2A_3$ (左 A、C)	$A_1A_2A_3$ (左 B)
0	0001101	0100111	1110010	532	235
1	0011001	0110011	1100110	443	344
2	0010011	0011011	1101100	334	433
3	0111101	0100001	1000010	552	255
4	0100011	0011101	1011100	245	542
5	0110001	0111001	1001110	354	453
6	0101111	0000101	1010000	225	522
7	0111011	0010001	1000100	443	344
8	0110111	0001001	1001000	334	433
9	0001011	0010111	1110100	423	324

3 EAN-13 条形码系统的实现

系统基于 Visual C++ 开发平台实现, 条形码图像的识别系统的界面如图 4 所示。按下“输出识别结果”按钮, 对应的输出结果显示在图 4 的编辑框内。



图4 条形码识别系统的界面

4 结束语

文中介绍了 EAN-13 条形码的格式、编码规则以及条码图像生成与识别的基本原理, 并用 Visual C++ 实现了条码图像识别系统。尽管关于一维条码识别的设备很多, 但多数都是针对于光电识别应用的。光电识别设备善于识别印刷质量好的条码, 而通过基于图像处理技术辨识一维条码, 将条形码的使用范围扩大到带有摄像头的手机等设备中, 能够识别质量稍差的条形码, 识别效果好, 具有一定的借鉴意义。

参考文献:

[1] 陈丹晖, 刘 红. 条码技术与应用[M]. 北京: 化学工业出版社, 2006.

[2] 董华冰. 一维图像条形码识别方法的研究[D]. 广州: 华南理工大学, 2012.

[3] 王再超. 一维条码技术原理及应用探究[J]. 计算机光盘软件与应用, 2011(23): 140-140.

[4] Muniz R, Junco L, Otero A. A robust software barcode reader using the hough transform[C]//Proc of 1999 international conference on information intelligence and systems. Bethesda: IEEE, 1999: 313-319.

[5] Youssef S M, Salem R M. Automated barcode recognition for smart identification and inspection automation[J]. Expert Systems with Applications, 2007, 33(4): 986-977.

[6] 邓泽峰, 熊有伦, 黄小鹏. 图像式一维条码识别的三重编码方法[J]. 光电工程, 2010, 37(3): 133-137.

[7] 徐益峰, 金恒宏, 戴曙光. 基于 EAN-13 条形码识别的改进算法[J]. 计算机与数字工程, 2011, 39(7): 127-131.

[8] 卫权岗, 马建红, 吴非洋, 等. 使用手机进行 ENA-13 码的识别及应用[J]. 电脑编程技巧与维护, 2011(16): 107-109.

[9] 税午阳, 周明全, 耿国华, 等. 基于数字图像处理的 EAN13 码识别[J]. 计算机工程与应用, 2008, 44(26): 177-179.

[10] Joseph E, Pavlidis T. Waveform recognition with application to bar codes[C]//Proceedings of 1991 IEEE international conference on system, man, and cybernetics. Charlottesville: IEEE, 1991: 129-134.

[11] Liu Ningzhong, Yang Jingyu. Recognition of two-dimensional bar code based on Fourier transform[J]. Journal of Image and Graphics, 2003, 8(8): 877-882.

[12] 黄小英. 基于图象的条形码识别的算法研究及设计[J]. 电子技术, 2011, 38(5): 21-22.

[13] 张 伟, 王军锋, 王 涛, 等. 一种基于改进算子的形态学边缘检测算法[J]. 计算机技术与发展, 2013, 23(6): 23-26.

[14] 严 军, 沈建华. 基于膨胀-腐蚀算法的 PDF417 码纠错机制的研究[J]. 计算机技术与发展, 2013, 23(6): 136-138.

EAN-13条形码图像的识别

作者: [刘艳洋](#), [曹玉东](#), [LIU Yan-yang](#), [CAO Yu-dong](#)
作者单位: [辽宁工业大学 电子与信息工程学院, 辽宁 锦州, 121001](#)
刊名: [计算机技术与发展](#) 
英文刊名: [Computer Technology and Development](#)
年, 卷(期): 2015 (6)

引用本文格式: [刘艳洋](#), [曹玉东](#), [LIU Yan-yang](#), [CAO Yu-dong](#) [EAN-13条形码图像的识别](#) [期刊论文] - [计算机技术与发](#)
[展](#) 2015 (6)