

基于 EPS 格式条形码的图形识别

陈青^{1,2}, 吴敏², 谢俊喜², 潘鹏¹

(1.湖南工业大学 计算机与通信学院, 湖南 株洲 412008; 2.中南大学 信息科学与工程学院, 湖南 长沙 410083)

摘要: 由于 EPS 格式为矢量类文件, 无分辨率可言, 很难利用数字图像处理技术来识别。针对 EPS 格式条形码识别过程中存在诸如受黑墨扩展因素造成了条码的条和空的比例不符, 无法准确识别出正确的条码等问题, 提出了一种基于二进制法解析 EPS 格式条形码, 并对其进行识别, 判断条码的准确度, 这样可以提前检测出 EPS 格式条码的错误。对不同放大系数的条形码进行分析, 从而避免生成胶片和印刷时的误差和浪费。实验结果证明效果显著。

关键词: EPS 格式; 图形识别; 条形码

中图分类号: TB482, TP391.41

文献标识码: A

文章编号: 1673-9833(2008)02-0055-04

Graphic Recognition of Bar Code Based on EPS Format

Chen Qing^{1,2}, Wu Min², Xie Junxi², Pan Peng¹

(1. School of Computer and Communication Hunan University of Technology, Zhuzhou, Hunan, 412008, China;

2. School of Information Science and Engineering, Central South University, Changsha, 410083, China)

Abstract: In view of vector class and without resolution rate of EPS format, it is difficult to use digital image processing technology for recognition. There are many influence factors to bar code recognition, such as ink spread during the printing process, bar codes are read as sequences of widths of bars and spaces, it makes wrong to compare such widths with each other. The EPS format bar code based on binary method which can judge the bar code accuracy; reduce the printing error; analyzed different amplifying coefficients of bar code; avoid generate film and printing error and waste is proposed. The result proves good.

Key words: EPS format; graphic recognition; bar code

0 引言

自动识别技术是信息数据自动识别、自动输入计算机的重要方法和手段。条码技术是以计算机、光电技术和通信技术为基础的一项综合性科学技术, 它的使用极大地提高了数据采集和信息处理的工作效率。在众多自动识别技术中, 条码技术越来越被人们所认识和接受, 成为当今主要的计算机自动识别技术之一。

目前在条形码的制作过程中, 主要利用一个硬件电路扫描包装上的条形码, 根据采集到的条码图像, 利用条码的条和空的对比度不同, 测得一个电路波形, 然后利用一个整形电路, 尽可能地将边缘恢复出

来并且测得条和空的比例^[1, 2], 再根据条形码的编码规则来分析出条形码的符号^[3, 4], 如图 1 所示。

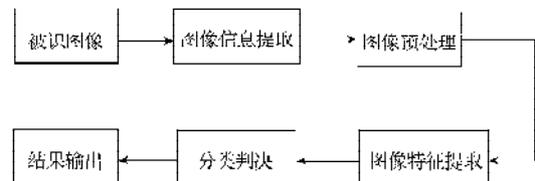


图 1 图像识别的基本流程

Fig. 1 Basic process of image recognition

文中以一维条码为研究对象, 它是人们日常生活

收稿日期: 2007-12-17

作者简介: 陈青(1967-), 女, 江苏镇江人, 湖南工业大学副教授, 中南大学博士生, 主要研究方向为图象识别。

中最常见的商品代码。一维条码主要有 EAN 和 UPC 两种,其中 EAN-13 码是我国主要采取的编码标准。EAN 是欧洲物品条码(European Article Number Bar Code)的英文缩写,是以消费资料为使用对象的国际统一商品代码。研究发现,在生成 EPS 格式的 EAN-13 条码中,会出现黑墨扩展、条码字符有误等错误,在被识别图像中就会出现较大的误差,这样后续的处理、识别过程就变得毫无意义,造成了在制作胶片和印刷成本上的大量浪费。对于条形码的制作主要是由以下几步来完成,如图 2 所示。

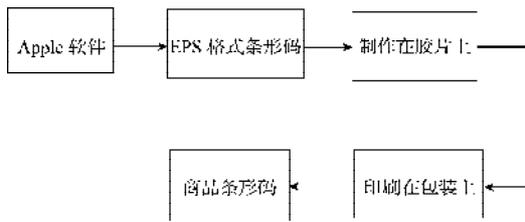


图 2 EPS 格式条形码制作流程

Fig. 2 Making process of the EPS format bar code

本文设计的方法主要是对未印在胶片上的 EPS 格式的条形码图形进行识别,判断其准确度。这样可以减少由于软件和机器自身的缺陷造成的不必要的损失,从而减少在制作胶片和印刷在包装上的浪费。由于在电脑上生成的条形码是 EPS 格式的, EPS 格式属于矢量类图像文件,没有像素和分辨率可言,因此不能利用数字图像处理技术来定位其像素和条形码来比较宽度。但是 EPS 格式图像可以用 ASCII 编码保存,也可以用二进制(BinaV)编码记录图像数据。因此当采用二进制形式对图像数据进行编码时,图像的存储容量和转换到打印机所需的工作量将是采用 ASCII 编码方法的一半。本文提出了一种基于二进制法操作来分析 EPS 格式的条形码。

1 EPS 格式简介

EPS 全称为 Encapsulated PostScript,译为封装的 PS 文件^[5]。PS 语言原来只为打印机而设计的,但是 PS 语言中的缩放和转移的功能,使得这种文件可以内置在其他的 PS 格式文件的页面上^[6]。EPS 是 Adobe 公司的 Post Script 页面描述语言的文件格式,这种语言用于描述矢量图形。由于桌面出版大多使用 Post Script 页面描述语言^[7]打印输出,因此,几乎所有的图形图像处理软件和桌面出版软件都支持 EPS 格式。

2 二进制法解析 EPS 格式

利用二进制法分析 EPS 格式条形码,主要是由 Post Script 语言描述的条形码的条和空的宽度及其在页面

中的位置。利用下面式(1)可以得出条和空的宽度,利用式(2)可以得出条和空之间的比例对应关系。

$$W = (x_2 - x_1 - k_{i-1} / 2 - k_{i+1} / 2) \quad (1)$$

式(1)中: x_1 表示一个条或空起始位置;

x_2 表示一个条或空的终止位置;

w 表示条或空中单个模块的宽度;

k 表示条或空对应单个模块的比例系数;

W 表示条形码中条或空的实际宽度;

k_i 表示由数据对应得当前条或空的宽度;

$$k = W / w \quad (2)$$

式(1)中 k_{i-1} 由数据表示的前一个条或空的宽度, k_{i+1} 表示由数据表示的后一个条和空的宽度。

主要过程是,利用二进制法对 EPS 格式的条码进行分析,准确定位出描述条形码条和空的数据。一个 EPS 格式的文件中可能有很多个条形码,通过 VC++6.0 结合数据库,将每一个条码的数据分别存储在各自的数据库中,再利用上面公式(2)对其分析,就可以得出条码的比例宽度。由于受黑墨扩展因素^[8]的影响,会出现条码字符与模块宽度比例不符的条形码。文中对出现错误的条码,采用字母标识号可以很明显地辨别出来,即如果字符对应的条码模块比例不符则用字母“A”来代替。

3 应用示例

下面主要以 EPS 格式的 EAN-13 条形码为对象来进行分析。如图 3 所示,对于条码 EAN-13 图:起始符由 3 个模块组成,左侧 6 位数据符由 42 个模块组成,中间分隔符由 5 个模块组成,右侧 5 位数据符由 35 个模块组成,终止符由 3 个模块组成。每个数据字符由 7 个模块组成,由 7 位的 0 和 1 来表示。其中在条形码的模块中,模块 0 和 1 组成了条码的条和空;其中一个单一的条和空可以由 1~4 个模块组成。对 EPS 格式的 EAN-13 一维条形码进行分析发现,一个一维条形码的符号为: 6923 4693 8038 1,如图 3 所示。

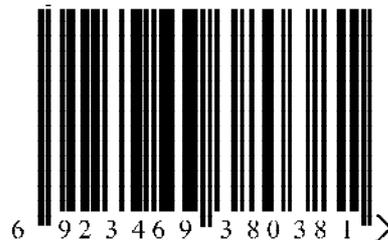


图 3 EAN-13 条形码

Fig. 3 EAN-13 bar code

在解析 EPS 文件时发现,条形码绘制是以二进制数据表示的。数据表示形式及分析出条码字符的数据结果如表 1 所示:

表1 EPS格式条码的数据形式

Tab. 1 The EPS format bar code data form

x_1	x_2	k_i
1 787. 514 1 782. 129 m	1 791.256 1 782.129 m	1.785 7 w
1 791. 256 1 782. 129 m	1 798.739 1 782.129 m	1.785 7 w
1 798. 739 1 782. 129 m	1 803.416 1 782.129 m	1.785 7 w
1 803. 416 1 782. 129 m	1 803.416 1 782.129 m	3. 656 5 w
1 810. 827 1 782. 129 m	1 816.439 1 782.129 m	3. 800 4 w
1 816. 439 1 782. 129 m	1 821.188 1 782.129 m	3. 800 4 w
1 821. 188 1 782. 129 m	1 830.542 1 782.129 m	1.785 7 w
1 830. 542 1 782. 129 m	1 838.026 1 782.129 m	1.785 7 w
1 838. 026 1 782. 129 m	1 843.638 1 782.129 m	5. 527 3 w
1 843. 638 1 782. 129 m	1 847.380 1 782.129 m	1.785 7 w
1 847. 380 1 782. 129 m	1 853.927 1 782.129 m	1.785 7 w
1 853. 927 1 782. 129 m	1 864.217 1 782.129 m	7. 398 1 w
1 864. 217 1 782. 129 m	1 868.894 1 782.129 m	1.785 7 w
1 868. 894 1 782. 129 m	1 873.571 1 782.129 m	3. 656 5 w
1 873. 571 1 782. 129 m	1 877.312 1 782.129 m	1.785 7 w
1 877. 312 1 782. 129 m	1 881.054 1 782.129 m	1.785 7 w
1 881.054 1 782. 129 m	1 890.408 1 782.129 m	1.785 7 w
1 890. 408 1 782. 129 m	1 894.078 1 782.129 m	1.785 7 w
1 894. 078 1 782. 129 m	1 899.690 1 782.129 m	1.641 7 w
1 899. 690 1 782. 129 m	1 909.116 1 782.129 m	1.641 7 w
1 909. 116 1 782. 129 m	1 916.599 1 782.129 m	5.527 3 w
1 916. 599 1 782. 129 m	1 920.341 1 782.129 m	1.785 7 w
1 920. 341 1 782. 129 m	1 929.695 1 782.129 m	1.785 7 w
1 929. 695 1 782. 129 m	1 933.364 1 782.129 m	1.785 7 w
1 933. 364 1 782. 129 m	1 938.976 1 782.129 m	1.641 7 w
1 938. 976 1 782. 129 m	1 947.539 1 782.129 m	1. 641 7 w
1 947. 539 1 782. 129 m	1 955.022 1 782.129 m	3. 800 4 w
1 955. 022 1 782. 129 m	1 959.627 1 782.129 m	3. 800 4 w
1 959. 627 1 782. 129 m	1 963.369 1 782.129 m	1.785 7 w

注：表中 x_1 表示起止位置； x_2 表示终止位置。

通过对上面的数据分析发现，PostScript语言主要是通过 m 、 w 来描述条形码的条和空。其中 m 表示 move to 即起始点或终止点， w 表示 setlinewidth 即绘制条和空的宽度。在译码的过程中，利用计算机对于二进制数据的处理方法，主要是提取出其中的数据，解析出其中的条码符号所对应的条和空对应的比例。表1中， k_i 表示由数据对应应当前条或空的宽度。利用公式(1)和公式(2)来对表1的数据进行计算。首先根据公式(1)，得到 x_2 表示终止位置、 x_1 表示起始位置和 k_i 的值，这就可以得到每个条或空的实际宽度。由于本条形码中第一个条的宽度为单位模块宽度，其宽度为1.785 7，所以设定 $w=1.785 7$ 为本条形码中单个模块的宽度，于是可以利用公式(2)计算出每个条码字符的条空比例。

对表1的分析结果如表2所示：其中条为1，空为-1。

表2 利用公式(1)、(2)对表1进行分析的结果

Tab. 2 Using formula(1、2) analyzed table 1

条码字符	条	空	比例
前置符	6	1 -1	1
第一位	9	-3 1 -1	2
第二位	2	-2 2 -1	2
第三位	3	-1 1 -4	1
第四位	4	-2 3 -1	1
第五位	6	-1 1 -1	4
第六位	9	-3 1 -1	2
中间分割符	-1	1 -1	1 -1
第一位	3	1 -4	1 -1
第二位	8	1 -2	1 -3
第三位	0	3 -2	1 -1
第四位	3	1 -4	1 -1
第五位	8	1 -2	1 -3
校验符	1	2	
终止符		1 -1	1

通过分析发现，由于条和空的颜色不同，设定第一个条为黑色，设定的条件是当与其他的条或空相比较的颜色相同时为正数(在输出的结果中，输出 n 个1)，当颜色不同时为负数(在输出的结果中，输出 n 个-1)。但是由于计算机自身的原因，在生成条形码的过程中，会受黑墨点扩展(printing with ink spread)的影响。对于一个由7个模块组成的字符如表3所示。

表3 模块与数字的对应关系

Tab. 3 Corresponding relation of the module and figure

条和空的类型	模块的序号						
	1	2	3	4	5	6	7
正确的条和空	1	1	1	-1	-1	1	-1
错误的条和空	1	1	-1	-1	-1	1	-1

由于受黑墨扩展因素的影响，在生成的条和空中，空可能变成了条。例如[1 -2 1 -3]数字分别代表了1个模块宽度的条，2个模块宽度的空，1个模块的条，3个模块的空。由条码对应关系得其表示的符号字符为8，但可能读出的比例结果是[2 -1 1 -3]。这样虽然也是7个模块，但是表示了另外的符号数。[1 -2 1 -3]的模块表示为：1 -1 -1 1 -1 -1 -1，[2 -1 1 -3]的模块表示为：1 1 -1 1 -1 -1 -1，从而造成了错误的条形码。此时就可以用字母“A”来标识此条码字符。基于这些错误情况的考虑，下面利用基于二进制法操作来分析EPS格式的条形码。对于每个码字的7个模块单独进行分析，利用vc程序指令来读取其长度，放入一个数组中，求出其比例来，然后利用条形码所对应的编码规则与数组中的值进行分析。由于用Ghostscript可以描述条码模块的尺寸，可在此设定第一个尺寸为条码模块的单位宽度，以后的尺寸与此类比，这样就可以得到条码的整体的宽度比例。然后利

用“A、B、C”子集来确定前置符的数字符号，可以得到一个数组，将其存入计算机中，得出其条码的符号。在一张EPS格式的文件中有多个条形码，可以利用计算机技术把它们单独分割开来，把这些条码的符号分别填入到Access的数据库中。然后利用ODBC建立数据库连接，利用Query语句查询出与一维条码相对应的符号字符。则可以方便地读取EPS文件中的条形码字符，并对其进行识别。

4 结论

本文提出的利用二进制法识别EPS格式条形码方法与传统的条形码检测算法相比，可以更早地检测出条码的准确度，从而避免制作在胶片上、印刷在包装上的一些不必要的浪费，节省了印刷费用和胶片制作成本。由于受墨点扩展及其它一些因素的影响，可能出现条码字符与条码条和空不对应的条形码，利用二进制法分析EPS格式的条形码，可以对条码的尺寸比例进行分析，其精确度达到百分之一。这样可以更精确地定位条形码的精度，尽量减少由内在因素引起的错误。实验中利用VC++6.0完成对软件的设计，对500多张不同放大比例的EPS格式条形码文件进行分析，发现有10多个错误的条码符号。这样可以在条形码制成胶片、印刷在包装之前检测出条码模块排布上的错误，降低检测人员利用硬件电路检查的工作量^[9,10]。基于本文采用的算法，利用VC++6.0开发的软件已在中国物品编码中心湖南分中心运用，经过3个多月的测试，取得了很好的效果，可以准确地判断出是否存在错误的条形码，节省了用户在印刷成本上的费用。

参考文献:

- [1] Joseph E, Pavlidis T. Waveform recognition with application to bar code [C]// In: Proc of 1991 IEEE Int'l Conf on Systems, Man, and Cybernetics. Charlottesville: IEEE Press, 1991: 129-134.
- [2] Shellhammer S J, Goren D P, Pavlidis T. Novel signal-processing techniques in barcode scanning[J]. IEEE Robot&Automation Magazine, 1999, 6(issue1): 57-65.
- [3] 中国物品编码中心. 条码技术与应用[M]. 北京: 清华大学出版社, 2003.
- [4] 陈丹晖, 刘红. 条码技术与应用[M]. 北京: 化学工业出版社, 2005.
- [5] 蔡慧萍. 理解EPS——一种重要的图形格式漫谈[J]. 南昌职业技术师范学院学报, 2000(3): 56-61.
- [6] Russell Lang, Ghostgum Software Rty Ltd. GhostScript Ghostview and GSview [EB/OL].[2002-05-11]. <http://www.cs.wisc.edu/~ghost/>.
- [7] AllenBraunsodorf. Thecomp Long postscriptFAQv3.0[EB/OL]. [1998-02-02]. <http://www.Postscript.org/FAQs/language/FAQ.html>.
- [8] Pavlidis T, Swartz J, Wang Y P. Fundamentals of bar code information theory[J]. IEEE Computer, 1990, 23(4): 74-86.
- [9] Mallat S G. A Theory for Multiresolution Signal Decomposition: the Wavelet Representation[C]// IEEE Transactions Pattern Analysis Machine Intelligence 1989, [S.L.]:[S.N]: 674-693.
- [10] Joseph E, Pavlidis T. Bar Code Waveform Recognition Using Peak Locations[J]. IEEE Trans PAMI, 1994, 16(6): 630-640.

(责任编辑: 罗立宇)