

采用 EPS 技术的高精度彩色出版

朱崇湘 赵文耘 孙永学

(复旦大学计算机科学系, 上海, 200433)

摘 要 从大型数据库管理数据信息到自动的高精度排版输出, 是当前信息服务行业的迫切需要, 在作者的研究成果——《天翼结构化彩色号簿出版系统》中, 这一技术得到了实现。此文对该系统实现中的关键技术之一——EPS 彩色页面输出技术进行了研究, 介绍了 PostScript 语言的特性和基本概貌, 并重点研究了页面描述、坐标变换、图形嵌入和文字处理等几个方面的内容。最后对这一技术的应用前景做了展望。

关键词 自动出版 彩色出版 大型数据库 Post Script 语言 EPS 标准

分类号 TP399

1 前 言

随着计算机技术的飞速发展, 出版行业出现了跨时代的进步, 基于计算机的文字处理及出版系统成为了人们研究的热点。同时, 由于信息在社会经济发展中所起的作用越来越重要, 各行各业中以数据库形式存储在计算机中的数据, 正在以难以想象的速度膨胀, 怎样快速正确地把信息从大型数据库转化到印刷品上, 成为了摆在人们面前的一个难题。当前, 国内外在文字处理、出版方面的工具软件都是以手工方式进行排版。例如, Microsoft Word、PageMaker 以及国内的方正等都是这方面的优秀软件。由于手工操作的限制, 这类软件无法解决以上提出的问题。以电话号簿出版业为例, 目前上海市有数百万电话用户, 由这些用户的各种信息组成的数据库有几十千兆字节, 每年需以这个数据库为依据出版各种各样的号簿数十种, 手工排版是完全不可能的。基于上述原因, 我们于 1995 至 1996 年间开发了《天翼结构化彩色号簿出版系统》(见 1997 年 1 月 22 日文汇报), 并采用此系统成功地出版了亚洲第一本彩色电话号簿——'96 上海黄页电话号簿。在本文中, 我们对该系统中的关键技术之一——EPS 页面输出技术进行了讨论。

2 EPS 技术的提出

出版印刷业中, 彩色出版可归结为 CMYK4 原色的分色及合成。而对于一个出版系统而言, 印刷质量是系统实现的关键, 衡量其优劣的首要标准为 CMYK4 色片的精度。在国内, 方正系统等自行研究了汉字曲线字库等技术, 采用了一套自行定义的语言描述及字库

上海市科委“九五”攻关项目部分资助。 1997-09-30 收稿

朱崇湘: 硕士研究生, 研究方向: 软件工程。

赵文耘: 副教授, 主要研究领域: 软件工程。

孙永学: 硕士研究生, 研究方向: 软件工程。

描述,解决了汉字的印刷问题。而在国外,中高档印刷业普遍采用的标准为 PostScript 页面描述语言。不管是台式机出版系统还是专业出版系统,均采用 PostScript 语言来表示所希望的页面输出结果,它实际上已成为一种工业标准,许多激光打印机和所有的照排机均配有 PostScript 解释器。由于图形是通过程序设计语言描述产生的,同样的页面描述,传送到不同的输出设备就可得到不同的输出质量,从而保证了描述的设备无关性。而为了实现不同软件的数据交换,国际上采用的跨平台图形文件交换技术标准 of EPS (Encapsulated Postscript) 技术。因此,为了保证印刷品输出的高精度,我们选择采用了 PostScript 语言作为我们的输出描述语言,这样,彩色号簿就变成了号簿文本信息与彩色图形广告的有机合成品,如图 1 所示。

为实现该技术,我们开发了一套完整的适合出版要求的 PostScript 原语来完成:文字、填充项、线、图象等的输出、缩放、填色、合成等操作以及其它一些出版中的特殊要求,如:补字(用预选设计的汉字图形文件自动填补照排机国际字库以外的汉字)等。

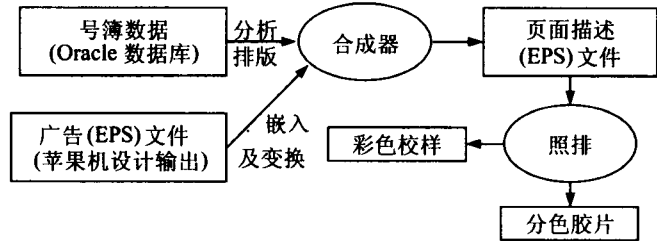


图 1 页面合成示意图

3 EPS 技术的实现

3.1 PostScript 语言简介

PostScript 语言是由 Adobe System 公司设计和实现的一种描述页面输出的程序设计语言。它有多种可以任意组合的图形操作符,包含有变量,并允许把许多操作符组合成复杂的过程或函数。一个 PostScript 格式的文件就是一个可以由解释器来执行的程序。PostScript 语言具有描述能力强、提供可编程元素、支持多种彩色模型 (RGB, CIE, CMYK)、支持范围广、可移植性强等特点。因此,在高精度图文输出方面具有非常广泛的应用。

3.2 页面描述

PostScript 既是一种程序设计语言,又是一种页面描述语言,其映象模型是一个简单统一的 2 维图形观点:图象可以通过把颜料放置在页面上的某个特定区域来建立,着色的区域可以是字母形状、一般的填充形状、线或半色调表示的照片。

生成打印页面的 PostScript 程序包含下面几个方面的内容:使用路径构造操作符建立一个路径;设置隐式参数;完成着色操作。程序中还可以通过裁剪路径对当前页面上的各种图形进行裁剪。由于 PostScript 的这种机制,使得输出图形和文字(事实上字库也是用 PostScript 语言描述的图形)都是由程序勾勒出来的。因此,在任何缩放、旋转、变换等操作下,图形都不会发生任何失真和精度的损失。

PostScript 解释器支持一个被称为图形状态的数据结构,它用于保存当前图形控制参数,这些参数定义了图形操作符执行的上下文关系,如:当前彩色参数、当前路径、当前字体等。

图形状态本身并不是一个对象,不能够直接存取。但它包含有很多对象,所有这些对象都能由图形状态操作符读和改变。图形状态保存在堆栈中,PostScript 程序可以用 gsave

操作符把当前图形状态压入到图形状态栈，这样就完成了图形状态的保存。恢复堆栈顶部保存的图形状态可以通过用 `grestore` 弹出栈顶元素来完成。这种设施使得页面描述的元素所产生的影响局部化，对实现图形嵌入等操作非常有利。

下面我们给出一段 `PostScript` 页面描述，该描述在页面上绘出图形如图 2 所示。

3.3 坐标系变换与图形嵌入

在彩色出版物中，彩色图形广告是很常见的。这些广告一般总是由美工人员使用图形广告设计软件（如 `Freehand`、`Photoshop` 等）进行设计，最后以一定的图形文件格式嵌入到版面中。由于当前广告设计软件平台中 `Macintosh` 占据较大的优势，而 `Macintosh` 文件格式中二进制表示与 `Intel` 系列微机的二进制表示有着高低字节相反的差别，这就使得二者之间的数据交换需要通用性强的文件格式，`EPS` 文件格式出色地完成了这一任务。因此，我们采用了 `EPS` 作为交换文件格式，实现了变换和嵌入功能，使得异种平台的文件互操作易如反掌。

图 2 `PostScript` 页面描述
输出结果示例

<code>%! PS</code>	
<code>/inch {72mul} def</code>	%定义了一个 inch 子程序
<code>/width 4 inch def</code>	%定义矩形的宽度
<code>/height 7 inch def</code>	%定义矩形的高度
<code>gsave</code>	%保存当前图形状态
<code>1 inch 7 inch moveto</code>	%设置当前点
<code>width 0 rlineto</code>	%把矩形 1 右边线加入到当前路径，
	%在此这前当前路径为空
<code>0 height rlineto</code>	%把矩形 1 右边线加入到当前路径
<code>width neg 0 rlineto</code>	%把矩形 1 顶边线加入到当前路径
<code>closepath</code>	%关闭路径
<code>7.5 inch 10 inch moveto</code>	%设置当前点
<code>width neg 0 rlineto</code>	%把矩形 2 顶边线加入到当前路径
<code>0 height neg rlineto</code>	%把矩形 2 左边线加入到当前路径
<code>width 0 rlineto</code>	%把矩形 2 底边线加入到当前路径
<code>closepath</code>	%关闭路径
<code>0.50 setgray</code>	%设置填充形状（隐式参数）
<code>fill</code>	%完成着色操作
<code>grestore</code>	%恢复上次保存的图形状态

图 3 `PostScript` 页面描述程序代码示例

在 `PostScript` 程序中，语言定义了一个标准的、默认的坐标系统，称为用户空间，程序根据它来对页面上的点进行定位。该用户空间是一个虚拟坐标系，不管用来产生输出的设备如何，它与当前页面有相同的关系。初始情况下，用户空间的原点在输出页面的左下角；`x` 轴为水平方向，正 `x` 轴向右延伸；`y` 轴为垂直方向，正 `y` 轴向上延伸；`x` 轴与 `y` 轴的单位长度为 `1/72` 英寸。称这样的用户空间为默认用户空间。

出现在 PostScript 程序中的坐标系与当前页的关系可以通过坐标变换来改变,坐标变换可以通过操作符如 translate (平移)、rotate (旋转) 和 scale (缩放) 来完成。坐标变换不仅增强了设备独立性,而且是一个非常有用的工具。例如,原来占据一整页的页面描述可以在不作任何变化的情况下,把它作为另一个页面的一个元素放到另一个页面描述中。我们正是利用这一特性来实现图形嵌入的。

要把一个图形文件嵌入到当前页面的指定位置和指定范围内,必须得到被嵌入图形的大小。这可以通过读取 EPS 文件头部的结构化注释来得到。在 EPS 规范中规定文件头部必须包含该文件的必要信息,这些信息以 “%%” 开始的行来表示,PostScript 解释器将这些行作为注释行忽略掉。下面是一个 EPS 文件头的例子:

其中 “%%BoundingBox: 0 0 595 807”

一行包含了该文件的边框信息。

嵌入操作包含以下几个步骤:

1) 使用 gsave 保存当前的图形状态;
2) 把原点平移 (translate) 到嵌入位置的左下角;

3) 通过缩放 (scale) 改变坐标轴单位,比例因子设定为嵌入位置与嵌入图形的大小比例,这时嵌入位置变为了被嵌入图形的整个页面的坐标空间;

4) 把被嵌入图形的 PostScript 描述代码作为一般描述加入到页面描述中;

5) 使用 grestore 恢复嵌入前的图形状态。

在嵌入时如果适当改变变换矩阵的参数,还可以得到旋转、拉伸和扭曲等特殊效果,这里不做详述。

3.4 文字处理

文字处理确切地讲就是处理字库中的字符。字符被定义为一个普通的几何图形,并且可以对它进行所有的图形处理。考虑到大多数页面描述中正文的重要性,PostScript 提供了一些高级设施来允许对字符进行任意地和有效地描述、选择和输出。这些设施包括:字库字典、用于匹配字符和字符描述的编码机制、对字库可用的矩阵信息以及对字库的高速缓存。

在 PostScript 上下文关系中,字库是一个字典,每种字体都有与之对应的字库。一般的 PostScript 解释器不配备汉字库,要具备汉字输出功能,就必须购买汉字库来扩充 PostScript 解释器。MonoType 在这一市场占据了绝对的主导地位,它为用户提供了几十种高质量的 PostScript 汉字库。字体可以有两种编码方法,即 ASCII (内码) 和 8 进制编码,两种编码之间可以相互转换,这主要是为了解决一些特殊字符以及语言保留字的输出问题。

PostScript 提供一个称为字库高速缓存 (fontcache) 的数据结构,用于提高字符打印的效率。由于汉字库所需的存储空间太大,同时缓存多种汉字库一般是不可能的。汉字库一般都存储在硬盘中(而英文字库一般是固化在解释器 ROM 之中的),解释器在缓存与硬盘之间进行字体的调度是自动完成的,但是这种调度效率相对较低。因此,在描述页面

```
%! PS-Adobe-3.0 EPSF-3.0
%%Title: PS11. EPS
%%Creator: Software Engineering Laboratory, FDU.
%%CreationDate: Thu Sep 11 13: 20: 25 1997
%%For: Shanghai Telephone Directory Corp.
%%BoundingBox: 0 0 595 807
%%DocumentNeededResources: (atend)
%%DocumentSuppliedResources: (atend)
%%CMYKCustomColor: (atend)
%%DocumentCustomColors: (atend)
%%DocumentProcessColors: (atend)
```

图 4 EPS 文件头示例

时,频繁更换字体对输出效率将有很大的影响。然而,在一些应用中,频繁更换字体有是必要的,例如电话号簿的页面输出格式为:

复旦大学

邯郸路……65642222

在这里,一家用户的信息由户名、地址、电话构成,它们以不同的字体输出。整个页面由成百上千的用户组成。在这种情况下,如果简单地顺序输出,将在一个页面输出过程中进行几千次字体选择,效率极低。为解决这一问题,我们在输出页面时设计了文字输出缓冲队列。队列按照不同的字体来组织,文字输出时不是直接进行描述,而是先存储在它与使用的字体相对应的队列中,在页面结束前一次性把各个队列中的文字输出到页面描述中。通过这种技术,我们把页面在激光照排机上的输出速度提高了10倍以上。

4 小 结

《天翼结构化彩色号簿出版系统》已在1996年1月通过专家鉴定并经国际联网检索。专家认定该系统在“号簿结构化彩色出版方面,属国内首创并填补了空白”(摘自鉴定委员会鉴定意见)。该系统现已在号簿生产中得到充分应用,1996年一年出版亚洲第一本彩色号簿数十万本,新增产值近二千万元,其中的EPS技术的应用可以说是站在巨人的肩膀上,充分利用国外技术,达到完善的技术效果。我们的软件系统可描述1/1000磅的间隔(当然这已超过了最高级的排版精度),这样输出设备越高级,其出版质量也就越高。而采用大型数据库ORACLE为核心的自动出版可将国内众多的企业从手工排版、非数据库管理的境况中解放出来。也许辞书、电话号簿等大容量出版物的又一次革命即将到来。

参 考 文 献

- 1 PostScript Language Reference Manual (Second Edition) Addison-Wesley Publishing Company, Inc.
- 2 PostScript By Example Henry McGilton. Mary Campione
- 3 潘志庚, 胡长源, 张明敏. PostScript 及程序实例集. 北京科海培训中心
- 4 方正排版系统用户手册
- 5 潘志庚. Windows 环境下图形图象程序设计. 北京: 清华大学出版社

High Precision Colorful Publishing Using EPS Technology

Zhu Chongxiang Zhao Wenyun Sun Yongxue
(Fudan University, hangHai, 200433)

ABSTRACT It's a pressing demand of ISP that convert data managed by RDBMS to high precision publication automatically. In our achievement of research, "Tian Yi Structured Colorful Telephone Directory Publishing System", this problem is resolved. In this paper, we discuss the EPS technology that is one of the key technology of this publishing system. At first, we outline the characteristic and basic features of Post Script language. Then, discuss page description, coordinate transformation, graphics insertion and words processing in detail. Finally, the future application of this technology is given.

KEY WORDS Automatic publishing Colorful publishing RDBMS Post Script language EPS standard