



(12) 发明专利

(10) 授权公告号 CN 102331982 B

(45) 授权公告日 2014. 03. 05

(21) 申请号 201110213589. X

CN 101923723 A, 2010. 12. 22,

(22) 申请日 2011. 07. 28

杨道良, 常明, 任晓霞 . PDF 及其在电子出版领域的应用 . 《计算机应用》. 1999, 第 19 卷 (第 1 期),

(73) 专利权人 深圳万兴信息科技股份有限公司
地址 518057 广东省深圳市南山区高新南一道 006 号 TCL 工业研究院大厦 A 座 9 楼
北侧 A901 室、A 座 8 楼

审查员 朱殿尧

(72) 发明人 解辉

(74) 专利代理机构 广州三环专利代理有限公司
44202

代理人 郝传鑫 熊永强

(51) Int. Cl.

G06F 17/21 (2006. 01)

G06F 17/22 (2006. 01)

(56) 对比文件

CN 101923723 A, 2010. 12. 22,

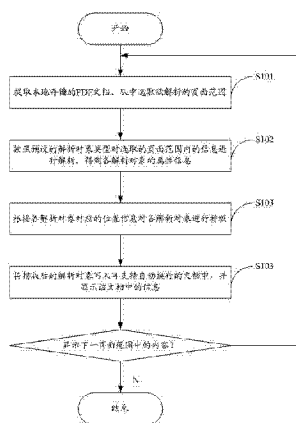
权利要求书2页 说明书4页 附图2页

(54) 发明名称

自适应窗体大小的 PDF 文档显示方法、系统及移动终端

(57) 摘要

本发明适用于电子书应用技术领域, 提供了一种自适应窗体大小的 PDF 文档显示方法、系统及移动终端。方法包括下述步骤: 步骤 A, 提取本地存储的 PDF 文档, 从中选取欲解析的页面范围; 步骤 B, 按照预设的解析对象类型对选取的页面范围内的信息进行解析, 得到各解析对象的属性信息, 各解析对象的属性信息包含有各解析对象的位置信息; 步骤 C, 根据各解析对象对应的位置信息对各解析对象进行排版; 步骤 D, 利用文本流的写入方式将排版后的解析对象写入可支持自动换行的文档中, 并显示该文档中的信息。采用本发明提供的 PDF 文档显示技术, 用户在阅读 PDF 页面时无需左右调整显示窗体, 可以带给用户更好的阅读体验, 使用更加方便。



1. 一种自适应窗体大小的 PDF 文档显示方法,其特征在于,包括下述步骤:

步骤 A,提取本地存储的 PDF 文档,从中选取欲解析的页面范围;

步骤 B,按照预设的解析对象类型对选取的页面范围内的信息进行解析,得到各解析对象的属性信息,所述各解析对象的属性信息包含有各解析对象的位置信息,所述解析对象类型包括文本对象、图形对象、图像对象,所述图形对象的属性信息包含有图形控制信息和图形层次信息;所述图像对象的属性信息包含有图像控制信息和图像层次信息;

所述步骤 B 中对图形对象和图像对象进行解析包括:

步骤 B4,舍弃大图像,即不满足预设标准的图形对象 / 图像对象,对满足预设标准的图形对象 / 图像对象按照所述图形层次信号 / 图像层次信息进行合成;

步骤 B5,将合成后的图形对象 / 图像对象所在的区域位置信息作为合成后的图形 / 图像对应的位置信息;

步骤 C,根据各解析对象对应的位置信息对各解析对象进行排版;

步骤 D,将排版后的解析对象写入可支持自动换行的文档中,并显示该文档中的信息。

2. 如权利要求 1 所述的自适应窗体大小的 PDF 文档显示方法,其特征在于,所述文本对象的属性信息包含有文本格式信息、超链接信息和位置信息。

3. 如权利要求 2 所述的自适应窗体大小的 PDF 文档显示方法,其特征在于:

所述步骤 B 中按照下述步骤 B1 至步骤 B2,或者步骤 B1 至步骤 B3 对文本对象进行解析:

步骤 B1,将文本对象的文本块按照位置信息识别为文本行;

步骤 B2,将文本行识别为文本段;

步骤 B3,按照文本位置识别该文本是否有超链接。

4. 如权利要求 1 所述的自适应窗体大小的 PDF 文档显示方法,其特征在于,所述步骤 C 中在排版时,如果有图形对象或图像对象的位置信息对应于文本对象的右方,则将该图形对象或图像对象排在该文本对象的下方。

5. 一种自适应窗体大小的 PDF 文档显示系统,其特征在于,包括:

页面范围选取单元,用于提取本地存储的 PDF 文档,并从中选取欲解析的页面范围;

PDF 页面信息解析单元,用于按照预设的解析对象类型对选取的页面范围内的信息进行解析,得到各解析对象的属性信息,所述各解析对象的属性信息包含有各解析对象的位置信息,所述解析对象类型包括文本对象、图形对象、图像对象,所述图形对象的属性信息包含有图形控制信息和图形层次信息;所述图像对象的属性信息包含有图像控制信息和图像层次信息;

所述 PDF 页面信息解析单元按照如下方式对图形对象和图像对象进行解析:舍弃大图像,即不满足预设标准的图形对象 / 图像对象,对满足预设标准的图形对象 / 图像对象按照所述图形层次信号 / 图像层次信息进行合成,将合成后的图形对象 / 图像对象所在的区域位置信息作为合成后的图形 / 图像对应的位置信息;

解析对象排版单元,用于根据各解析对象对应的位置信息对各解析对象进行排版;

显示控制单元,用于将排版后的解析对象写入可支持自动换行的文档中,并控制显示该文档中的信息。

6. 如权利要求 5 所述的自适应窗体大小的 PDF 文档显示系统,其特征在于,所述文本对

象的属性信息包含有文本格式信息、超链接信息和位置信息。

7. 如权利要求 6 所述的自适应窗体大小的 PDF 文档显示系统,其特征在于:

所述 PDF 页面信息解析单元按照如下方式对文本对象进行解析:将文本对象的文本块按照位置信息识别为文本行;将文本行识别为文本段;按照超链接区域进行超链接识别。

8. 如权利要求 5 所述的自适应窗体大小的 PDF 文档显示系统,其特征在于,所述解析对象排版单元在排版时,如果有图形对象或图像对象的位置信息对应于文本对象的右方,则将该图形对象或图像对象排在该文本对象的下方。

9. 一种移动终端,其特征在于,包括如权利要求 5 至 8 任一项所述的自适应窗体大小的 PDF 文档显示系统。

自适应窗体大小的 PDF 文档显示方法、系统及移动终端

技术领域

[0001] 本发明属于电子书应用技术领域,尤其涉及一种自适应窗体大小的 PDF 文档显示方法、系统及移动终端。

背景技术

[0002] 随着移动终端的大量普及,在其上进行 PDF 文档阅读的需求与日俱增。由于 PDF 文档的自身格式限制,不能根据移动终端的窗体大小自动排版,所以造成了人们在阅读的时候需要频繁拉动滚动条来移动页面才可以看全页面信息。因此如何能析取出 PDF 文档中的信息,重新排版适应设备窗体大小、方便人们阅读成为了一个迫切的需求。

[0003] 针对上述技术问题,现有技术中存在以下两种解决方案:

[0004] 1、将 PDF 页面信息截取出一页进行显示。这种显示方式不能按照窗体大小进行适应显示,并且不能在放大页面的时候重新渲染,显示效果不够清晰。

[0005] 2、按照 PDF 内部结构,依次将每个对象渲染在显示窗体上。虽然克服了在显示时页面不够清晰的缺点,但是依然不能使得显示信息适应窗体,阅读体验不好。

发明内容

[0006] 本发明所要解决的技术问题在于提供一种自适应窗体大小的 PDF 文档显示方法、系统及移动终端,以在显示 PDF 文档信息时可以使页面信息能自适应窗体大小,给用户以更好的阅读体验。

[0007] 本发明是这样实现的,一种自适应窗体大小的 PDF 文档显示方法,包括下述步骤:

[0008] 步骤 A,提取本地存储的 PDF 文档,从中选取欲解析的页面范围;

[0009] 步骤 B,按照预设的解析对象类型对选取的页面范围内的信息进行解析,得到各解析对象的属性信息,所述各解析对象的属性信息包含有各解析对象的位置信息;

[0010] 步骤 C,根据各解析对象对应的位置信息对各解析对象进行排版;

[0011] 步骤 D,将排版后的解析对象写入可支持自动换行的文档中,并显示该文档中的信息。

[0012] 进一步地,所述解析对象类型包括文本对象、图形对象、图像对象。

[0013] 进一步地,所述文本对象的属性信息包含有文本格式信息、超链接信息和位置信息;所述图形对象的属性信息包含有图形控制信息和图形层次信息;所述图像对象的属性信息包含有图像控制信息和图像层次信息。

[0014] 进一步地,所述步骤 B 中按照下述步骤 B1 至步骤 B2,或者步骤 B1 至步骤 B3 对文本对象进行解析:

[0015] 步骤 B1,将文本块按照位置信息识别为文本行;

[0016] 步骤 B2,将文本行识别为文本段;

[0017] 步骤 B3,按照文本位置识别该文本是否有超链接;

[0018] 所述步骤 B 中按照如下方式对图形对象和图像对象进行解析:

[0019] 步骤 B4,对满足预设标准的小尺寸图形对象 / 图像对象按照所述图形层次信号 / 图像层次信息进行合成 ;

[0020] 步骤 B5,将合成后的图形对象 / 图像对象所在的区域位置信息作为合成后的图形 / 图像对应的位置信息。

[0021] 进一步地,所述步骤 C 中在排版时,如果有图形对象或图像对象的位置信息对应于文本对象的右方,则将该图形对象或图像对象排在该文本对象的下方。

[0022] 本发明还提供了一种自适应窗体大小的 PDF 文档显示系统,包括 :

[0023] 页面范围选取单元,用于提取本地存储的 PDF 文档,并从中选取欲解析的页面范围 ;

[0024] PDF 页面信息解析单元,用于按照预设的解析对象类型对选取的页面范围内的信息进行解析,得到各解析对象的属性信息,所述各解析对象的属性信息包含有各解析对象的位置信息 ;

[0025] 解析对象排版单元,用于根据各解析对象对应的位置信息对各解析对象进行排版 ;

[0026] 显示控制单元,用于将排版后的解析对象写入可支持自动换行的文档中,并控制显示该文档中的信息。

[0027] 进一步地,所述 PDF 页面信息解析单元按照如下方式对文本对象进行解析 :将文本块按照位置信息识别为文本行 ;将文本行识别为文本段 ;按照超链接区域进行超链接识别 ;所述 PDF 页面信息解析单元按照如下方式对图形对象和图像对象进行解析 :对满足预设标准的小尺寸图形对象 / 图像对象按照所述图形层次信号 / 图像层次信息进行合成,将合成后的图形对象 / 图像对象所在的区域位置信息作为合成后的图形 / 图像对应的位置信息。

[0028] 进一步地,所述解析对象排版单元在排版时,如果有图形对象或图像对象的位置信息对应于文本对象的右方,则将该图形对象或图像对象排在该文本对象的下方。

[0029] 本发明还提供了一种移动终端,其特征在于,包括如上所述的自适应窗体大小的 PDF 文档显示系统。

[0030] 采用本发明提供的 PDF 文档显示技术,用户在阅读 PDF 页面时无需左右调整显示窗体,可以带给用户更好的阅读体验,使用更加方便,由于适用于小尺寸的手机终端。

附图说明

[0031] 图 1 是本发明实施例提供的自适应窗体大小的 PDF 文档显示方法的显示流程图 ;

[0032] 图 2 是本发明实施例提供的自适应窗体大小的 PDF 文档显示系统的结构原理图。

具体实施方式

[0033] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0034] 本发明实施例中,首先按照 PDF 的文档结构,解析出 PDF 页面中的对象,其次按照解析对象的类型和位置信息对其进行重新排版,完后将重新排版后的得到的信息以文本流

写入方式写入 HTML 等可支持自动换行的文档,最后用 WebView 显示文档信息。

[0035] 图 1 示出了本发明提供的自适应窗体大小的 PDF 文档显示方法的实现流程,详述如下。

[0036] 在步骤 S101 中,提取本地存储的 PDF 文档,从中选取欲解析的页面范围。

[0037] 本发明实施例所针对的 PDF 文档为存储在本地的文档,目的在于加快显示设备对本地 PDF 文档的访问。其中,页面范围可选取一整页 PDF 页面或者一页 PDF 页面中的部分内容,也可以选取多页 PDF 页面。

[0038] 在步骤 S102 中,按照预设的解析对象类型对选取的页面范围内的信息进行解析,得到各解析对象的属性信息。

[0039] 其中,各解析对象的属性信息包含有各解析对象的位置信息,针对不同的解析对象,此属性信息中具体的信息类型也不同。在本发明实施例中,解析对象类型包括文本对象、图形对象、图像对象三种。对于文本对象,其属性信息包含有文本格式信息、超链接信息和位置信息,其中文本格式信息主要指文本的字体、字号、颜色、删除线和下划线等信息,超链接信息指外部网页链接等,而位置信息则为本文本对象相对于页面左上角的位置。对于图形对象,其属性信息包含有图形控制信息和图形层次信息,相应地,图像对象的属性信息包含有图像控制信息和图像层次信息,图形控制信息为图形颜色、图形透明度、图形线条宽度、图形边框颜色等,图像控制信息为图像颜色、图像透明度等,图形层次信息或图像层次信息为该图形或图像在页面对象的层次,即层次越高越靠上,高层次对象盖在低层次对象上。

[0040] 本发明实施例中,对于不同的解析对象,具体的解析方式也不同。对于文本对象,解析方式如下:首先将文本块按照位置信息识别为文本行,具体识别过程为:首先设定一个阈值,然后将本页中文本块纵坐标相差值在这个阈值内的文本块定义为一个文本行集合,然后再根据这个集合中的文本块的横坐标进行排列连接为一个文本行;再将文本行识别为文本段,具体识别文本段方式为:将一页中的文本行按照纵坐标的大小排列连接为一个文本段。考虑到文本可能还包含有超级链接属性,因此还包括一个识别超级链接的步骤,具体可按照文本的位置识别该文本是否有超链接属性,具体识别方式为:在 PDF 文档中每个超链接解析出来后为一个矩形区域,每个区域中包含哪些文本就需要按位置匹配前面的文本块,匹配到的文本块就设置成此超链接属性。而对于图形对象和图像对象,考虑到大图像会在生成的文档中占有比较大的面积,严重影响文本的阅读,因此本发明中将大图像舍弃而只考虑合成在原 PDF 文档中叠加在一起的小图像。解析方式为:对满足预设标准的小尺寸图形对象/图像对象按照图像层次信息进行合成,将合成后的图形对象/图像对象所在的区域位置信息作为合成后的图形/图像对应的位置信息。

[0041] 在步骤 S103 中,根据各解析对象对应的位置信息对各解析对象进行排版。

[0042] 排版规则是从左到右,从上到下,如果图形对象或图像对象在文本段的右边,就将其新位置设为文本对象的下方。

[0043] 在步骤 S104 中,将排版后的解析对象写入可支持自动换行的文档中,并显示该文档中的信息。

[0044] 例如可采用以文本流的写入方式生成 HTML 文件或 text 文件等,然后通过 WebView (网页视图)载入该 HTML 文件或 text 文件进行显示。此处生成的页面会按照窗体大小进行

自动排版。

[0045] 上述步骤示出了一个完整的显示流程,如果需要进一步显示其他 PDF 页面范围内的内容,则重复上述步骤即可。

[0046] 本领域普通技术人员可以理解实现上述各实施例提供的方法中的全部或部分步骤可以通过程序来指令相关的硬件来完成,所述的程序可以存储于一计算机可读取存储介质中,该存储介质可以为 ROM/RAM、磁盘、光盘等。

[0047] 图 2 示出了本发明实施例提供的自适应窗体大小的 PDF 文档显示系统的结构原理,为了便于描述,仅示出了与本发明相关的部分。此 PDF 文档显示系统可以为内置于具有显示功能的移动终端中的软件单元、硬件单元或软硬件结合的单元。

[0048] 参照图 2,本发明实施例提供的自适应窗体大小的 PDF 文档显示系统包括页面范围选取单元 21、PDF 页面信息解析单元 22、解析对象排版单元 23、显示控制单元 24。其中,页面范围选取单元 21 用于提取本地存储的 PDF 文档,并从中选取欲解析的页面范围,PDF 页面信息解析单元 22 按照预设的解析对象类型对选取的页面范围内的信息进行解析,得到各解析对象的属性信息,所述各解析对象的属性信息包含有各解析对象的位置信息,解析对象排版单元 23 再根据各解析对象对应的位置信息对各解析对象进行排版,最后由显示控制单元 24 将排版后的解析对象写入可支持自动换行的文档中,并控制显示该文档中的信息,例如可采用文本流的写入方式。

[0049] 上述解析对象类型包括文本对象、图形对象、图像对象;所述文本对象的属性信息包含有文本格式信息、超链接信息和位置信息;所述图形对象的属性信息包含有图形控制信息和图形层次信息;所述图像对象的属性信息包含有图像控制信息和图像层次信息。

[0050] 上述 PDF 页面信息解析单元 22 按照如下方式对文本对象进行解析:将文本块按照位置信息识别为文本行;将文本行识别为文本段;按照超链接区域进行超链接识别;PDF 页面信息解析单元 22 按照如下方式对图形对象和图像对象进行解析:对满足预设标准的小尺寸图形对象/图像对象进行合成,将合成后的图形对象/图像对象所在的区域位置信息作为合成后的图形/图像对应的位置信息。

[0051] 进一步地,解析对象排版单元 23 在排版时,如果有图形对象或图像对象的位置信息对应于文本对象的右方,则将该图形对象或图像对象排在该文本对象的下方。

[0052] 本发明实施例还提供了一种内置有上述自适应窗体大小的 PDF 文档显示系统的移动终端。

[0053] 上述自适应窗体大小的 PDF 文档显示方式不仅可适用于具有小尺寸屏幕的移动终端,还可适用于具有较大尺寸屏幕的终端,如计算机显示器等,例如当具有较大尺寸屏幕的终端在显示 PDF 文档,可能需要使原先处于最大化状态的窗体变小来配合其他的操作,此时即可采用如上所述的自适应窗体大小的 PDF 文档显示方式,而在此较小的窗体内无需让用户通过操作滚动条浏览 PDF 文档。

[0054] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

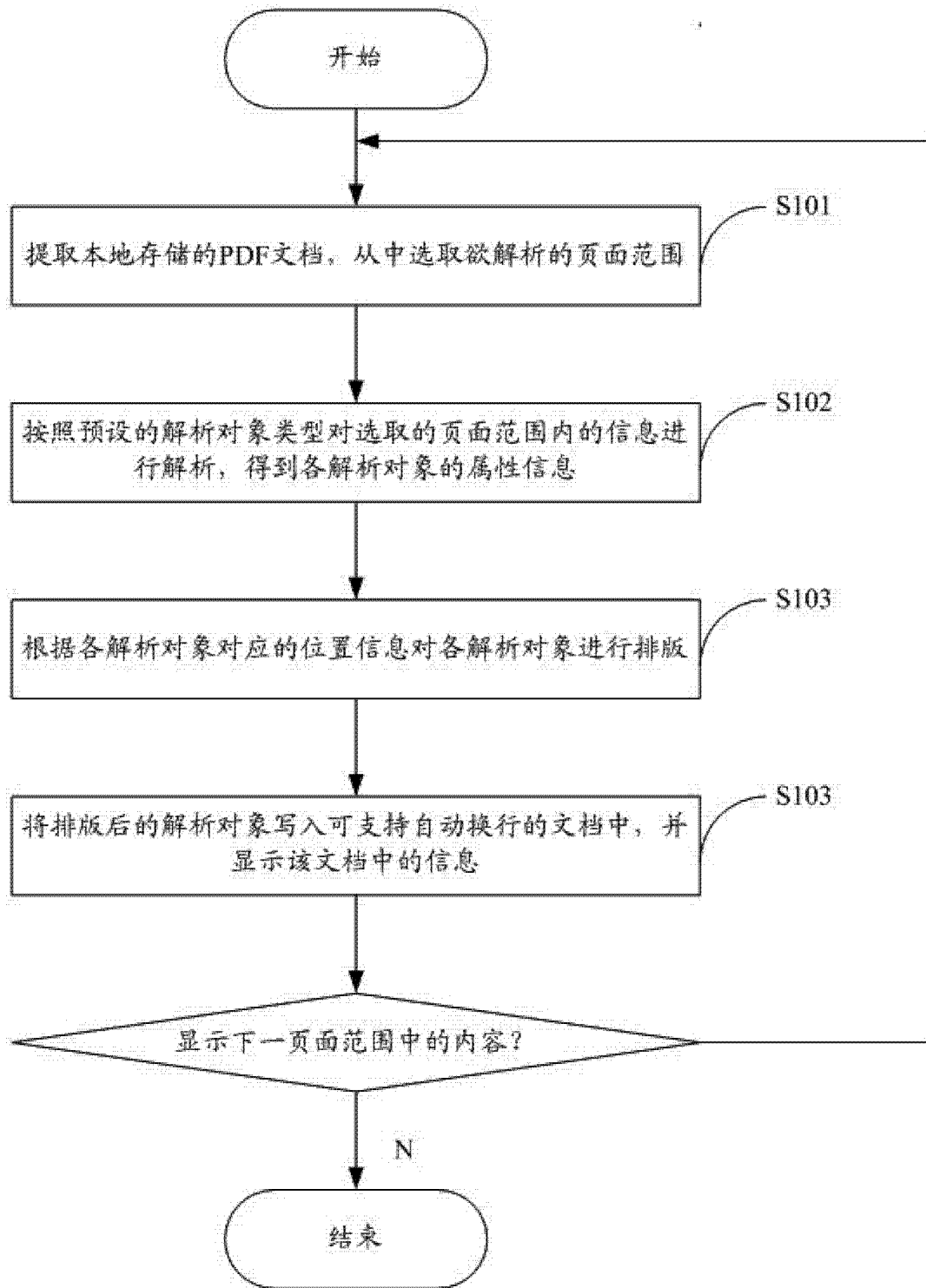


图 1

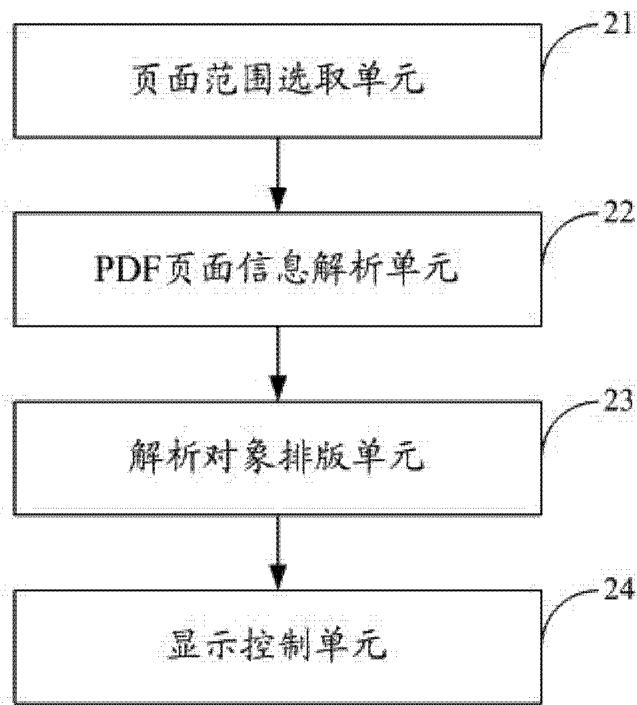


图 2