

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G06F 17/24 (2006.01)

G06K 9/34 (2006.01)



[12] 发明专利说明书

专利号 ZL 200510056178.9

[45] 授权公告日 2009 年 3 月 4 日

[11] 授权公告号 CN 100465945C

[22] 申请日 2005.3.24

[21] 申请号 200510056178.9

[30] 优先权

[32] 2004.3.24 [33] US [31] 10/808,194

[73] 专利权人 微软公司

地址 美国华盛顿州

[72] 发明人 C·E·何里 K·H·切拉皮拉

P·A·维奥拉

T·T·克里斯蒂安松

[56] 参考文献

CN1339775A 2003.3.13

JP10-198688A 1998.7.31

US2002/0101626A1 2002.8.1

US2003/0095709A1 2003.5.22

CN1375782A 2002.10.23

JP10-78997A 1998.3.24

审查员 郑嘉青

[74] 专利代理机构 上海专利商标事务所有限公司

代理人 张政权

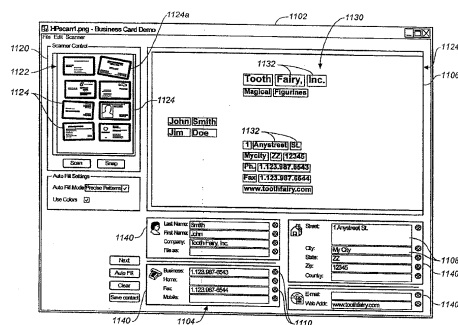
权利要求书 5 页 说明书 35 页 附图 19 页

[54] 发明名称

从扫描文档填充电子表格的方法和装置

[57] 摘要

提供了一种用于填充来自电子图像的电子表格的计算机可执行的方法和装置。该方法和装置识别电子图像内一个对象的大小、方向和位置，并且识别来自与对象对应的图像内的像素的信息元素。电子表格的字段连同识别的信息元素一起通过图形用户界面显示给用户。这些信息元素被解析为不同信息类型的标记组。用标记组填充电子表格的至少一些字段以产生一个填充的表格。允许用户通过图形用户界面编辑填充的表格。



1. 一种用于从电子图像填充电子表格的计算机可实现方法，该方法包括：
 - (a) 标识电子图像内具有任意方向的第一对象的大小、方向和位置；
 - (b) 从对应于第一对象的电子图像内的像素标识信息元素；
 - (c) 通过图形用户界面向用户显示所述电子表格的字段和所标识的信息元素；
 - (d) 把信息元素解析到不同信息类型的标记组；以及
 - (e) 用所述标记组填充所述电子表格的字段以产生一填充的表格，且允许用户通过图形用户界面编辑填充的字段，

其中，(a) 包括：

分类步骤，分类图像内的每个像素以产生像素分类数据，所述分类步骤包括分类每个像素为数据像素或者背景像素中的一种；

定义步骤，按照特定方向数据像素的总和定义一图像函数以处理像素分类数据；

第一计算步骤，计算第一方向的图像函数以产生第一数据集；

第二计算步骤，计算第二方向的图像函数以产生第二数据集；

搜索步骤，搜索第一方向的图像函数中的差异区域和第二方向的图像函数中的差异区域，其中所述差异区域是图像中有极少数据像素或没有数据像素的区域；

划分步骤，基于所述图像函数中的差异区域把图像划分为子图像；和

处理步骤，处理子图像以确定包括所述第一对象的所述电子图像内的每个对象的大小、方向和位置。

2. 如权利要求 1 所述的方法，其特征在于，所述处理子图像包括：重复所述分类步骤、所述定义步骤、所述第一计算步骤、所述第二计算步骤、所述搜索步骤以及所述划分步骤，直到每个子图像包含单个对象或不能再被划分。

3. 如权利要求 1 所述的方法，其特征在于：

所述步骤 (b) 进一步包括使用光学字符识别来标识第一对象内的文本块和该文本块的两维位置；以及

所述步骤 (c) 进一步包括通过图形用户界面向用户显示电子表格的字段的同

时显示所标识的文本块。

4. 如权利要求 3 所述的方法，其特征在于：

所述步骤（d）进一步包括用指示不同信息类型的视觉指示符突出显示第一对象的图像内的不同信息类型的标记组。

5. 如权利要求 1 所述的方法，其特征在于：

所述步骤（d）进一步包括从未标记媒体存储器接收信息元素作为未标记媒体，并解析所述未标记媒体以标识信息元素的信息类型；

所述步骤（e）进一步包括用所述信息元素中的至少一个，基于该元素的信息类型来自动地填充字段中的至少一个，

其中，所述方法还包括：

（f）通过对象数据图形用户界面显示所述信息元素和通过表格图形用户界面显示填充的字段和未填充的字段。

6. 如权利要求 5 所述的方法，其特征在于，所述步骤（f）进一步包括在对象数据图形用户界面内采用视觉指示符来指示信息元素与表格内特定字段相容。

7. 如权利要求 5 所述的方法，其特征在于，所述步骤（f）进一步包括指示与表格内至少一个填充的字段内的至少一个信息元素相关联的状态级别。

8. 如权利要求 7 所述的方法，其特征在于，所述步骤（f）进一步包括提示用户基于指示的状态级别校验和/或纠正所述至少一个填充的字段内的至少一个信息元素。

9. 如权利要求 8 所述的方法，其特征在于，所述步骤（d）进一步包括基于用户对所述至少一个信息元素的校验和/或纠正更新所述填充的表格的其他字段。

10. 如权利要求 1 所述的方法，进一步包括：

（f）将从对任何填充的字段所作的编辑中收集的边缘信息写入到边缘信息存储器；和

（g）基于所述边缘信息，重新解析所述信息元素到不同信息类型的标记组，以及重新填充至少一个其他的字段。

11. 如权利要求 1 所述的方法，其特征在于，电子图像包括多个对象并且该方法进一步包括：

（f）为每个对象执行（a）到（e），使得对于每个对象，用从该对象内的像素获得的信息元素来至少部分地填充相应电子表格的字段。

12. 如权利要求 11 所述的方法, 其特征在于, 所述多个对象包括不同类型的对象。

13. 一种用于至少部分地填充电子表格的系统, 该系统包括:

一对象探测和提取模块, 它处理电子图像内的像素以标识电子图像内具有任意方向的第一对象的大小、方向和位置;

一光学字符识别模块, 它从与第一对象相对应的电子图像内的像素标识信息元素;

一图形用户界面, 它同时向用户显示电子表格的字段和标识的信息元素; 和

一解析模块, 它将信息元素解析到不同信息类型的标记组, 并用所述标记组至少部分地填充字段以产生填充的电子表格,

其中, 所述对象探测和提取模块包括:

一数据像素探测模块, 该数据像素探测模块将所述图像的每个像素分类为数据像素或者背景像素中的一种, 按照一特定方向上数据像素的总和定义一图像函数, 在第一方向上计算所述图像函数以产生第一数据集, 在第二方向上计算所述图像函数以产生第二数据集, 以及搜索第一方向的图像函数中的差异区域和第二方向的图像函数中的差异区域, 其中, 所述差异区域是图像中有极少数据像素或没有数据像素的区域;

一能够基于所述图像函数中的差异区域划分图像成较小的子图像的分割模块; 和

一处理所述子图像以探测图像内的对象的数量、大小、方向和位置的单个对象提取模块。

14. 如权利要求 13 所述的系统, 其特征在于, 所述图形用户界面允许用户编辑填充的字段和任何未填充的字段。

15. 如权利要求 13 所述的系统, 其特征在于:

所述光学字符识别模块包括用于使用光学字符识别标识第一对象内的文本块和该文本块的两维位置的装置; 和

所述图形用户界面包括用于通过图形用户界面向用户显示第一对象的图像内的标识的文本块以及用指示所述不同信息类型的视觉指示符突出显示第一对象的图像内的不同信息类型的标记组的装置。

16. 如权利要求 13 所述的系统, 其特征在于:

所述解析模块包括用于从未标记媒体存储器接收信息元素作为未标记媒体、解析所述未标记媒体以标识信息元素的信息类型、和自动地用所述信息元素中的至少一个、基于该元素的信息类型填充所述字段中的至少一个的装置；和

所述图形用户界面包括用于通过对象数据图形用户界面显示信息元素并通过表格图形用户界面显示填充的字段和任何未填充的字段的装置。

17. 如权利要求 16 所述的系统，其特征在于，所述对象数据图形用户界面包含视觉指示符以指示与表格内特定字段相容的信息元素。

18. 如权利要求 16 所述的系统，其特征在于，所述表格图形用户界面包含与表格内至少一个填充的字段内的至少一个信息元素相关联的状态级别指示符。

19. 如权利要求 17 所述的系统，其特征在于，所述图形用户界面包含用于提示用户基于与字段相关联的状态级别指示符校验和/或纠正至少一个填充的字段内的至少一个信息元素的装置。

20. 如权利要求 14 所述的系统，其特征在于，所述解析模块包含用于根据对填充的字段和未填充的字段所作的编辑更新所述填充的电子表格的其他字段的装置。

21. 一种用于从具有不同信息类型的第一和第二对象的电子图像填充电子表格的方法，该方法包含：

标识电子图像内的所述第一和第二对象的大小、方向和位置，并把电子图像划分为与每个对象的大小、方向和位置相关联的电子图像内的像素相对应的子图像，这是通过：

分类图像内的每个像素以产生像素分类数据，所述分类包括分类每个像素为数据像素或者背景像素中的一种；

按照特定方向中数据像素的总和定义一图像函数以处理像素分类数据；

计算第一方向的图像函数以产生第一数据集；

计算第二方向的图像函数以产生第二数据集；

搜索第一方向的图像函数中的差异区域和第二方向的图像函数中的差异区域，其中所述差异区域是图像中有极少数据像素或没有数据像素的区域；

基于所述图像函数中的差异区域把图像划分为子图像；和

处理子图像以确定每个对象的大小、方向和位置，

对每个子图像执行光学字符识别以标识在相应对象内的未标记信息元素；

对于每个子图像，解析所述未标记信息元素成为标记信息元素；

用从第一对象的子图像标识的标记信息元素填充第一电子表格类型内的字段以产生第一填充表格；

用从第二对象的子图像标识的标记信息元素填充第二电子表格类型内的字段以产生第二填充表格；和

通过图形用户界面向用户显示第一和第二填充表格和未标记信息元素并允许用户通过图形用户界面编辑所述第一和第二填充表格。

22. 如权利要求 21 所述的方法，其特征在于：

所述第一对象包括名片，而第二对象包括购货收据；和

所述第一电子表格包含软件地址簿的联络信息记录而所述第二电子表格包含软件金融应用的电子金融记录。

从扫描文档填充电子表格的方法和装置

有关申请的交叉参考

本申请要求于 2003 年 1 月 29 日提交的美国专利申请第 10/354, 500 号、题为“在数字图像数据中自动探测和提取对象的系统和方法 (SYSTEM AND METHOD FOR AUTOMATICALLY DETECTING AND EXTRACTING OBJECTS IN DIGITAL IMAGE DATA)”的优先权并且是它的部分继续申请, 还要求于 2004 年 3 月 3 日提交的美国专利申请第 10/792, 519 号、题为“辅助填表 (ASSISTEND FORM FILLING)”的优先权。

技术领域

本发明涉及一种用于自动填充从扫描文档或其他电子图像填充电子表格的计算机可执行的方法和装置。

背景技术

从电子图像输入数据, 比如从扫描文档输入数据是一个艰苦的任务。往往需要的不只是一个图像的电子副本, 比如扫描件, 而是也需要表格形式的数据或其他文本信息。大多数用于帮助完成计算机产生的表格的现有技术系统使用光学字符识别、自然语言处理和其他用于标识扫描文档内信息元素的特定类型的人工智能技术。一旦该信息元素被标识, 它们被放置在所选表格上一个适当的字段或位置内。然而, 这些方法被广泛地认为是非常不可靠的。

此外, 现有技术的系统一次仅能处理一个文档, 这更增加了与填充电子文档相关联的体力和时间耗费。同样的, 用于扫描文档并帮助完成计算机产生的表格的硬件需要文档有一个预先确定的大小和方向以便它们可以被适当地扫描。这会限制系统的多功能性而且可能需要购买用于扫描特定类型文档的特定的硬件。例如, 现在市场上有名片扫描仪, 它允许用户将名片放入扫描仪, 一次一张名片, 提取联络信息用于填充地址簿。扫描仪的大小可以接受有预定大小和方向的名片。这些扫描仪不能用于扫描其他类型和大小的文档, 比如购货收据和帐单。同样地, 一次只能扫描一张名片, 这就降低了效率。其他名片专

用扫描仪，比如 Hotcard Technology Pte Ltd 所销售的，可以一次扫描多张名片，但名片在扫描仪上必须有特定的方向。

发明内容

所以，填充表格是单调乏味的、费时的而且非常容易出现人为错误。因此，本领域存在对帮助更快且更准确的填充表格的系统和方法的需要。渴望改进的方法和装置从扫描文档或其他电子图像来填充电子表格。

本发明提供了一种用于从电子图像填充电子表格的计算机可实现方法，该方法包括：

- (a) 标识电子图像内具有任意方向的第一对象的大小、方向和位置；
- (b) 从对应于第一对象的电子图像内的像素标识信息元素；
- (c) 通过图形用户界面向用户显示所述电子表格的字段和所标识的信息元素；
- (d) 把信息元素解析到不同信息类型的标记组；以及
- (e) 用所述标记组填充所述电子表格的字段以产生一填充的表格，且允许用户通过图形用户界面编辑填充的字段，

其中，(a) 包括：

分类图像内的每个像素以产生像素分类数据，所述分类包括分类每个像素为数据像素或者背景像素中的一种；

按照特定方向数据像素的总和定义一图像函数以处理像素分类数据；

计算第一方向的图像函数以产生第一数据集；

计算第二方向的图像函数以产生第二数据集；

搜索第一方向的图像函数中的差异区域和第二方向的图像函数中的差异区域，其中所述差异区域是图像中有极少数据像素或没有数据像素的区域；

基于所述图像函数中的差异区域把图像划分为子图像；和

处理子图像以确定包括所述第一对象的所述电子图像内的每个对象的大小、方向和位置。

本发明还提供了一种用于至少部分地填充电子表格的系统，该系统包括：

一对象探测和提取模块，它处理电子图像内的像素以标识电子图像内具有任意方向的对象的大小、方向和位置；

一光学字符识别模块，它从与第一对象相对应的电子图像内的像素标识信息元素；

一图形用户界面，它同时向用户显示电子表格的字段和标识的信息元素；和

一解析模块，它将信息元素解析到不同信息类型的标记组，并用所述标记组至少部分地填充字段以产生填充的电子表格，

其中，所述对象探测和提取模块包括：

一数据像素探测模块，该数据像素探测模块将所述图像的每个像素分类为数据像素或者背景像素中的一种，按照一特定方向上数据像素的总和定义一图像函数，在第一方向上计算所述图像函数以产生第一数据集，在第二方向上计算所述图像函数以产生第二数据集，以及搜索第一方向的图像函数中的差异区域和第二方向的图像函数中的差异区域，其中，所述差异区域是图像中有极少数据像素或没有数据像素的区域；

一能够基于所述图像函数中的差异区域划分图像成较小的子图像的分割模块；和

一处理所述子图像以探测图像内的对象的数量、大小、方向和位置的单个对象提取模块。

本发明还提供了一种用于从具有不同信息类型的第一和第二对象的电子图像填充电子表格的方法，该方法包含：

标识电子图像内的所述第一和第二对象的大小、方向和位置，并把电子图像划分为与每个对象的大小、方向和位置相关联的电子图像内的像素相对应的子图像，这是通过：

分类图像内的每个像素以产生像素分类数据，所述分类包括分类每个像素为数据像素或者背景像素中的一种；

按照特定方向中数据像素的总和定义一图像函数以处理像素分类数据；

计算第一方向的图像函数以产生第一数据集；

计算第二方向的图像函数以产生第二数据集；

搜索第一方向的图像函数中的差异区域和第二方向的图像函数中的差异区域，其中所述差异区域是图像中有极少数据像素或没有数据像素的区域；

基于所述图像函数中的差异区域把图像划分为子图像；和

处理子图像以确定每个对象的大小、方向和位置，
对每个子图像执行光学字符识别以标识在相应对象内的未标记信息元素；
对于每个子图像，解析所述未标记信息元素成为标记信息元素；
用从第一对象的子图像标识的标记信息元素填充第一电子表格类型内的字段以产生第一填充表格；
用从第二对象的子图像标识的标记信息元素填充第二电子表格类型内的字段以产生第二填充表格；和
通过图形用户界面向用户显示第一和第二填充表格和未标记信息元素并允许用户通过图形用户界面编辑所述第一和第二填充表格。

附图说明

图 1 是依照本发明的一个实施例、以常规的个人电脑的形式实现本发明的示例性系统的框图。

图 2 是纳入了本发明所描述的对象提取系统和方法的图像处理系统的示例性实现的概括框图。

图 3 是说明图 2 所示的对象提取系统的组件或模块的框图/流程图。

图 4 是说明图 3 所示的单个对象提取模块的详细框图。

图 5 是说明图 4 所示的对象探测和提取过程的进一步细节的综合流程图。

图 6 所示是使用对象探测和提取方法查找图像内的单个对象的第一工作示例。

图 7 所示是与图 6 的对象具有相同大小但不同方向的对象。

图 8 所示是使用对象探测和提取方法查找图像中的多个对象的第二工作示例。

图 9 所示是图 8 显示的图像的子图像的处理。

图 10 是对每个对象图像中的文本进行光学识别并聚集所识别的文本的方法的流程图。

图 11 是以图 10 所示的方法聚集识别的文本时所执行步骤的流程图。

图 12 所示是一填表界面，该界面根据本发明的一个实施例便于辅助用户填充电子表格内的字段。

图 13 是根据本发明另一实施例的填表界面。

图 14 所示是一系统，该系统便于通过图 12 和 13 所示的界面根据本发明的实施例来辅助填表。

图 15 是对示例性隐藏 Markov 模型，它便于辅助用户填充表格内的字段。

图 16 是一个柱状图，它说明了本发明在辅助用户填充表格方面的功效。

图 17 是根据本发明的一个实施例填充表格的方法的流程图。

图 18 是依照本发明的另一个实施例填充表格的方法的流程图。

图 19 是依照本发明的另一个实施例填充表格的方法的流程图。

具体实施方式

本发明的实施例提供了一种用于帮助用户用从对象如名片、帐单和购货收据的电子图像获得的数据填充电子表格的方法和装置。该电子图像可以从任何来源获得，比如从电子文件或数字成像设备中获得。在一个实施例中，从一个常规用途的扫描仪或数字照相机中获得图像。每张图像可以包括一个或多个具有未知大小、方向和位置的对象。图像中的每个对象包括特定信息类型的未标记信息元素，比如名片上的姓名和联络信息或购货收据的卖主、日期和金额。

图像中的各个对象被分割成部分，并且经分割的对象中的信息元素被标识。该系统能够识别并分割在同一个图像中被一起扫描的许多小文档。对图像中的每个对象而言，系统识别对象内的文本数据、基于特定信息类型解析文本数据并自动填充目标应用或电子表格内的字段。例如，如果目标应用是地址簿中的联络信息，用户可以一次扫描一张或多张名片并且系统会从各个分割的名片中提取姓名、电话号码、email 地址和其他信息。包含十个数字的一串文本可能是美国的电话号码，xxxx@yyyy.zzz形式的字符串可能是一个 email 地址。来自每张名片的信息元素被用于自动填充用户的联络表。图像可以保留以作参考。

另一个实施例中，用户可以扫描多个收据，拖放日期、金额和/或其他文本块到金融软件应用中适当的字段，比如费用报告应用、电子数据表或货币管理软件，比如 Microsoft Money™ 中的字段。收据的图像可以存储以作参考和/或与费用报告一并发送。对费用报告文件系统而言，使用付费方的公共密匙可以加密图像文件的密码散列以防止篡改数字图像。

该系统通过图形用户界面将解析的文本和填充的字段呈现给用户并免于出错，因为识别的文本聚集可被拖放到适当的字段。同样地，用户可以直接输入数据到任一字段。即使光学字符识别（OCR）未能正确地识别名片上的文本块，比如公司名称，但它可能已经聚集了那块文本。于是，用户可以拖拉那块文本到适当的字

段。这对于诸如收据的文档被扫描的应用而言是非常有用的。在收据上可以有許多块数字和文本,用户通常仅对输入卖主姓名、日期、最后金额和可能的税金感兴趣。只要是聚集了对象上的文本,用户就可以拖拉适当的块到表格或目标应用中适当的字段。

图 1 说明的是一个适合的计算机系统环境 100 的例子,在其上本发明的一些实施例可以执行。计算机系统环境 100 只是一个适合的计算环境的例子并没有限制本发明的用途或功能性范围的意图。该计算环境 100 也不应该理解为对示例性操作环境中图示的任一组件或其结合具有依赖性,有要求。

本发明可在许多其他通用或专用计算机系统环境或配置中操作。适合用于本发明使用的熟知计算系统、环境和/或配置的例子包括,但不限于个人电脑、服务器计算机、手持或膝上设备、多处理器系统、基于多处理器的系统、机顶盒、可编程的消费电子设备、网络 PC、小型机、大型机、包括任何以上系统或设备的分布式计算环境等等。

本发明可以计算机可执行指令,比如由计算机执行的程序模块的一般上下文来描述。一般地,程序模块包括执行特定任务或实现特定抽象数据类型的例行程序、程序、对象、组件、数据结构等等。本发明也可以在通过通信网络连接的远端处理设备执行任务的分布式计算环境中实践。分布式计算环境中,程序模块可以位于包括存储器存储设备的本地和远端计算机存储介质中。

参照图 1,用于实现本发明的一个示例性系统包括计算机 100 形式的通用计算设备。计算机 110 的组件可以包括但不限于:处理单元 120、系统存储器 130 和把包括系统存储器在内各种系统组件耦合到处理单元 120 的系统总线 121。该系统总线 121 可以是若干种类型的总线结构中的任何类型,包括存储器总线或存储器控制器、外围总线和使用任何一种总线结构的本地总线。作为例子,而非限制,这样的结构包括工业标准结构 (ISA) 总线、微通道结构 (MCA) 总线、扩展工业标准结构总线 (EISA)、视频电子标准协会 (VESA) 本地总线和周边元件扩展端口 (PCI) 总线,也称为 Mezzanine 总线。

计算机 110 通常包括多种计算机可读介质。计算机可读介质可以是任何可由计算机 110 访问的可用介质,包括易失性的和非易失性介质、可移动的和不可移动的介质。作为例子而非限制,计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括易失性的和非易失性的、可移动的和不可移动的介质,这些介

质可以在用于存储比如计算机可读指令、数据结构、程序模块或其他数据的信息的任何方法或技术中来实现。计算机存储介质包括但不限于：RAM、ROM、EEPROM、闪存或其他存储技术、CD-ROM、数字化视频光盘（DVD）或其他光磁盘存储器、磁性盒式磁带、磁性磁带、磁盘存储器或其他磁性存储设备，或任何其他可用于存储想要的信息并可由计算机 100 访问的介质。通信介质通常以诸如载波 WAV 或其他传送机制的调制数据信号的形式包含计算机可读指令、数据结构、程序模块或其他数据，且包括任何信息传输介质。术语“调制数据信号”是指在该信号中有一个或多个特征被以在信号中对信息进行编码的方式设定或改变的信号。作为例子但非限制，通信介质包括有线介质，如有线网络或直接的有线连接和比诸声音的、射频、红外线和无线介质。以上的任意组合也包括在计算机可读介质的范围内。

系统存储器 130 包括易失的和/或非易失的存储器形式的计算机存储介质，比如只读存储器（ROM）131 和随机存取存储器（RAM）132。典型地，基本输入/输出系统 133（BIOS）存储在 ROM 131 中，BIOS 包括如在启动期间帮助在计算机 110 的元件之间传递信息的基本程序。典型地，RAM 132 包括可立即访问的和/或当前处理单元 120 在其上操作的数据和/或程序模块。作为例子但非限制，图 1 例示操作系统 134、应用程序 135、其他的程序模块 136 和程序数据 137。

计算机 110 也可以包括其他可移动的/不可移动的、易失的/非易失的计算机存储介质。只是作为例子，图 1 说明了一个可以从不可移动的、非易失的磁性介质中读取或写入的硬盘驱动器 141、一个可以从不可移动的、非易失的磁盘 152 中读取或写入的磁盘驱动器 151 和一个可以从不可移动的、非易失的比如 CD ROM 或其他光学介质的光盘 156 中读取或写入的光盘驱动器 155。其他可用于示例性操作环境中的可移动的/不可移动的、易失的/非易失的计算机存储介质包括但不限于：磁性盒式磁带、闪存卡、数字化视频光盘、数字视频磁带、固态 RAM、固态 ROM 等。典型地，硬盘驱动器 141 通过一个不可移动的存储器接口，如接口 140 连接到系统总线 121，磁盘驱动器 151 和光盘驱动器 155 通常由不可移动的存储器接口，如接口 150 连接到系统总线 121。

以上讨论并在图 1 中例示的驱动器及其关联计算机存储介质为计算机 110 的计算机可读指令、数据结构、程序模块和其他数据提供了存储。比如，图 1 中，硬盘驱动器 141 存储有操作系统 144、应用程序 145、其他程序模块 146 和程序数据 147。注意，这些组件可与操作系统 134、应用程序 135、其他程序模块 136 以及程

序数据 137 相同,也可以不同。操作系统 144、应用程序 145、其他程序模块 146 和程序数据 147 在此给出了不同的标记以说明,最低限度他们是不同的拷贝。

用户可以通过输入设备如定位设备 161、键盘 162、麦克风 163 和数字成像设备 164 向计算机 110 输入命令和信息。例如,定位设备 161 可以包括鼠标、轨迹球和触垫。其他输入设备(未示出)可以包括操纵杆、游戏垫、圆盘式卫星电视天线等等。这些或其他的输入设备经常通过与系统总线相耦合的用户输入接口 160 连接到处理单元 120,但也可以经由其他接口和总线结构相连接,比如并行端口、游戏端口或通用串行总线(USB)。监视器 191 或其他类型的显示设备也通过如视频接口 190 的接口连接到系统总线。除了监视器,计算机也可以包括其他外围输出设备,比如扬声器 197 和打印机 196,它们通过输出外围接口 190 进行连接。

计算机 110 可以在使用到一个或多个远端计算机,比如远端计算机 180 的逻辑连接的网络环境中操作。远端计算机 180 可以是个人计算机、手持设备、服务器、路由器、网络 PC、对等设备或其他公共网络节点,并且通常包括与计算机 110 相关的以上描述中的许多或所有元件。图 1 中描述的逻辑连接包括局域网(LAN) 171 和广域网(WAN) 173,但也可包括其他网络。这样的网络环境在办公室、企业范围的计算机网络、企业内部互联网、因特网中是很普通的。

当用在 LAN 网络环境中时,计算机 110 通过网络接口或适配器 170 连接到 LAN 171。当用在 WAN 网络环境中时,计算机 110 通常包括调制解调器 172 或用于通过诸如因特网的 WAN 173 建立通信的其他手段。可内置或外置的调制解调器 172 可以通过用户输入接口 160 或其他适合的手段连接到系统总线 121。在网络环境中,与计算机 110 相关描述的程序模块或其中一部分可以存储在远端存储器存储设备中。作为例子而非限制,图 1 中有驻留于远端计算机 180 上的远端应用程序 185。应该注意到,所显示的网络连接是示例性,而且在计算机之间建立通信连接其他方式也是可以使用的。

在本发明的一些实施例的上下文中,诸如名片、帐单和收据等对象的电子图像可以从诸如数字成像设备 164 的来源或从通过诸如调制解调器 172 或 LAN 171 的远端来源中在本地获得。图像也可以从存储在诸如以上讨论和/或在图 1 例示的任何易失的或非易失的存储器存储介质上的文件中获得。数字成像设备 164 可以包括诸如通用或专用扫描仪、复印机、数字照相机或数字摄像机。其他类型的成像设备也可以用于获取一个或多个感兴趣的对象的电子图像。

I. 从单个图像分割多个对象

光电扫描仪和其他成像设备用于获得包含印刷信息（比如文本、插图或照片）的对象，并且将信息转换为计算机可以使用的数字形式。一般而言，用户将要扫描的对象放在扫描仪的压盘上。扫描头扫过压盘的范围且所得图像被分成多个像素。为每个像素位置分配一个取决于像素颜色或强度的值。然后，所得位矩阵（叫做位图）可以存储在文件中、显示在显示器上、及用软件应用处理。

如上提及，用户经常会需要扫描多个对象。作为例子，用户可能想扫描多张名片、帐单或者收据。为了节省时间，一次想对一个以上的对象进行扫描。因此，用户会将多个对象放在扫描仪的压盘上并且在一个过程中扫描它们。这节约了时间和精力，因为用户不需要为每个对象重复这样的过程：将对象放在扫描仪压盘上、盖上盖子并连接到扫描软件界面。

同时扫描多个对象带来的一个问题是，对象是作为一个单一的位图表现在扫描图像内的。这意味着当扫描图像保存为文件、在显示器上显示和被软件应用处理时，认为该图像是一个单一的图像或对象。用户经常会想将每个对象存为一个单独的文档。一些扫描应用确实允许用户手动选择每个对象的边界并将对象存为一个单独的文档。然而，这个手动隔离扫描图像内每个对象的过程是重复的、乏味的和费时的。

然而，本发明的一个实施例为从扫描图像中探测和提取多个对象提供了一个简单并健壮的系统和方法。该系统和方法允许用户在扫描仪上放置多个对象，识别扫描仪上的对象数目并询问用户是否原意将每个对象保存为一个单独的文档或用来填充独立的多个电子表格。这样的系统和方法使得扫描过程更为快速和有效并且减轻了用户手动分割扫描图像中每个对象的负担。

A. 系统综述

这里描述的对象探测和提取系统和方法能够在数字图像数据中自动查找想要的对象并将那些想要的对象从其他对象和任何背景中隔离开来。在仍然保留单次过程可以扫描多个对象的优势的同时，允许将每个对象考虑成自己单独的对象。因此，每个单独的对象可以存为自己的文档或由软件应用单独处理而独立于包括在扫描图像中的其他对象。例如，当多张名片被平台扫描仪一次扫描时，该系统和方法可以区别彼此相邻排列的多张名片。

一般而言，对象探测和提取系统和方法能够探测并提取具有已知形状但未知大小、方向和数量的对象。这部分地是通过沿每个对象的每个方向或维度定义“图像

函数”来实现的。这个图像函数是原始图像内的数据的一个函数并代表该数据。作为例子，假定图像包括矩形的两维（2-D）对象。进一步假定希望确定显示在图像中的矩形对象的数量和对象的尺寸、方向和位置。为了确定这种信息，该对象探测和提取系统和方法定义了两个耦合的一维（1-D）图像特性函数。大多数情况下，从这些函数中可以确定对象的数量和它们的大小、方向和位置。

每个图像函数都有某些要求。一个要求是，当只有一个所需类型的单独对象呈现在图像中时，该函数应该有一个特殊的可识别的特征。例如，如果该对象类型为矩形并且该对象特征函数是沿某个特定方向的位于该对象内的像素的总和（称作数据像素），可识别特征是该函数是梯形。当然，其他想要的对象类型和其他的对象特征函数会产生其他可识别特征。典型地，可识别特征是形状，但在其他的实施例中，该特征可能是，例如，像素颜色或像素强度。

对象特征函数是沿两个或更多不同方向计算出的，并且每当数据像素中出现间隙或差异（disparity）之处，该图像被分为子图像。这些间隙指示在那个位置沿其中一个方向缺乏需要的对象。子图像的细分以一种重复方式继续，直到该对象可识别特征的特征函数指示两种可能性中的一种。第一种可能性是子图像包括单个需要的对象（例如单张矩形名片）。另一种可能性是没有发现单个需要的对象并且没有进一步细分的可能。如果后一种情况出现，系统通知用户不能确定对象的全部数目、大小、方向和位置。

B. 图像处理系统

图2是一个结合进了以上描述的对象探测和提取系统及方法的图像处理系统示例性实现200的全面框图。一般而言，数字图像数据由一个对象探测和提取系统202进行处理以确定对象的数量和包括在数字图像数据中的每个对象的大小、方向和位置。该系统202通过确定每个图像的边界并自动将该对象隔离为单独的图像对象而实现此目的。这就节约了用户手动隔离每个对象的时间和精力。

用户将多个对象（比如名片或收据）， $0(1)$ ， $0(2)$ 和 $0(3)$ 放置在扫描设备206（比如平台扫描仪或图1中的其他数字成像设备164）的压盘204上。图2中显示的虚线代表压盘204包括在扫描设备206中。然后用户对位于压盘204上的对象进行扫描并获得数字图像数据210。该数字图像数据210是包含每个对象 $0(1)$ ， $0(2)$ 和 $0(3)$ 以及背景数据212的单个数字图像。由阴影

线显示在图 2 中的背景数据通常代表了在扫描期间覆盖压盘的扫描设备 206 盖子（未示出）的颜色。在示例性实现中，假定背景颜色是已知的或者是可以估计或确定的。

该对象探测和提取系统 202 位于计算设备 214 之上，比如图 1 显示的计算系统环境内。如下面详细描述，该数字图像数据 210 发送到对象探测和提取系统 202 中并进行处理。该对象探测和提取系统 202 查找到数字图像数据 210 中的每个对象 $0(1)$ ， $0(2)$ 和 $0(3)$ 并从数据 210 中提取每个对象。一旦提取完毕，这些对象可以作为独立于其他对象和背景数据 212 之外的单独图像对象进行处理。

C. 对象探测和提取系统

对象探测和提取系统 202 包括显示在图 3 中的许多程序模块，它们允许系统自动区别数字图像数据 210 中的一个或多个对象。对象探测和提取系统 202 包括数据像素探测模块 300、分割模块 310 和单个对象提取模块 320。

图像 330（比如显示在图 2 中的图像数据 210）由数据像素探测模块 300 接收，并且模块 300 对图像内的每个像素进行分析和分类以获得像素数据 340。该像素数据 340 包括诸如该像素是数据像素还是背景像素的信息。数据像素是位于图像 330 中的任何对象中的像素。另一方面，背景像素是对象之外并在背景之中的像素。此外，像素数据 340 包括诸如沿图像 330 的两个或更多方向的数据像素数量的信息。该数据像素探测模块还定义一图像函数用于处理像素数据。例如，如果图像函数被定义为总计图像的一个方向的数据像素，该像素数据 340 将包括沿描述图像 330 的坐标系统的一个轴的数据像素的数量和沿坐标体系另一个轴的数据像素的数量。

下一步，像素数据 340 发送到分割模块 310。分割模块 310 确定在图像函数和像素数据 340 中是否存在差异或间隙。如下面详细说明的，这些差异通常是图像 330 中某些只有很少的数据像素（相对于周围区域）或根本没有数据像素的区域。然后确定该图像 330 是否能够基于是否查找到差异而被分割开（框 350）。如果是，图像 330 是能够被分割开的并会根据相应差异而被分割开。其效果是可以把图像 330 拆分成多块或多个子图像（框 360）。然后每个子图像提交给数据像素探测模块 300 用于处理（框 370）并且用图像 330 的一部分（即每个子图像）取代图像 330，递归进程重新开始。对每个子图像的重复处

理一直继续直到子图像只包括单个对象或对子图像再无分割开的可能为止。第一种情况下,子图像被发送到单个对象提取模块 320 中用于处理。第二种情况下,该系统和方法通知用户不能确定子图像内对象的数量、大小、方向和位置。然而,后一种情况很少发生,因为该系统和方法是非常健壮的。因此,该方法通过递归地将图像分解为子图像来定位并分割每个对象。这种分解一直继续直到每个子图像包括单个对象或不能再进一步被分解为止。

如以上陈述,如果不存在差异之处,那么图像 330 不能被拆分的部分被发送给单个对象提取模块 320。单个对象提取模块 320 处理图像 330,从而可以检测并提取图像 330 内的对象并且找到图像 330 内的对象的数量、大小、方向和位置。提取的对象 380 从对象探测和提取系统 110 中输出。例如,提取的对象可以包括总图像 330 内的单张名片或收据的子图像。

图 4 是说明显示在图 3 中的单个对象提取模块 320 的细节的框图。一般而言,单个对象提取模块 320 检查可能包含单个对象或者根本没有对象的子图像,且如果有对象的话,定位该对象。模块 320 在确定子图像不能被进一步拆分后处理来自主图像 330 的每个子图像。或者,如果确定主图像不能被拆分,模块 320 处理主图像。

单个对象提取模块 320 包括像素分析模块 400、校验模块 410 和对象位置输出模块 420。有可能包括单个对象的子图像 430 由像素分析模块 400 接收并且产生像素数据。基于像素数据,计算子图像内对象位置的估计坐标。估计坐标发送到校验模块 410。该校验模块 410 将每个估计坐标与主图像 330 进行比较,子图像 430 是该主图像的一部分。注意,图像 330 可能与子图像 430 是相同的。该比较用来确定任何的估计坐标是图像 330 的似乎可信匹配(plausible fit)并用于校验子图像 430 内对象的存在。如果找到似乎可信的匹配,那么正确的坐标被发送到对象位置输出模块 420 然后作为输出发送(框 440)。根据坐标,可以从子图像 430 中分割并提取对象。如果没有找到似乎可信的匹配,那么校验模块将此情况通知对象位置输出模块 420。这种情况下,对象位置输出模块 420 不会输出单个对象的坐标,而是相反会输出一个说明在子图像 430 内没有找到对象的消息。

D. 对象探测和提取系统的综合流程图

图 5 是说明图 2-4 中显示的依照本发明的实施例的对象探测和提取方法详

细例子的流程图。在步骤 500 收到一个图像。在步骤 501，计算第一方向的数据像素的数量用于产生第一数据集。类似地，在步骤 502，计算第二方向的数据像素的数量用于产生第二数据集。作为例子，典型的图像是一个包含像素行和像素列的扫描的矩形图像。图像函数可被定义为一个方向数据像素的总和。这种情况下，为图像中的每一行计算行中数据像素的数量。同样地，为图像中的每一列计算列中数据像素的数量。第一数据集包含数据像素在图像行上的分布而第二数据集包括数据像素在图像列上的分布。

接下来，在步骤 503 搜索该第一和第二数据集，以便在步骤 504 确定是否存在任何差异区域。这些差异区域，或间隙是图像中那些只有极少数据像素或根本没有数据像素的区域。如果出现差异，那么在步骤 505，沿差异区域定义一条数据差异的线。例如，如果图像中的一行不包含数据像素，则沿着那一行定义一数据差异线。基于数据差异线，在步骤 506 图像被分为或被分割为子图像。一旦创建了这些子图像，就认为它们是独立的图像，是与产生它们的图像分离的。然后，在步骤 507，再次分别处理每个子图像。如此，为每个子图像在迭代过程中重复框 501—506。

在步骤 504，如果正被处理的子图像中的一个没有出现差异，那么子图像再次进行个别处理。这包括在步骤 508 计算第一方向的子图像的数据像素的数量以产生第三数据集和在步骤 509 计算第二方向的子图像的数据像素的数量以产生第四数据集。

应该注意到，如果在该方法最初（或第一次）迭代中没有查找到差异，那么不需要执行框 508 和 509。这是因为第一方向的数据像素的数量和第二方向的数据像素的数量可能已经在框 501 和 502 计算过了。图 5 中用虚线框画出了步骤 508 和 509 的轮廓来表示这一点。

一旦已经计算了像素数据，在步骤 510 使用数据的拐点（inflection point）来确定对象可能的坐标。可能存在多个与像素数据相对应的对象。为此，在步骤 511 对照输入的图像去核对可能的坐标以确定哪一个（如果有的话）可能的坐标是输入图像似乎可信的匹配。如果在步骤 512 的确定是肯定的并且可能的坐标中的一组是似乎可信的匹配，那么那些坐标就在步骤 513 作为输出发送。一旦知道了图像内对象的坐标和位置，该对象就可以从图像中分割并提取出来。在步骤 514，如果不存在图像的可能坐标的似乎可信的匹配，那么确定

图像中不能查找到对象。

E. 工作示例

为了说明对象探测和提取方法的细节，以下给出两个工作例。

1. 单个对象案

图 6 说明了扫描图像 602 中的一个对象 600。在这个工作例子中，对象 600 是一个矩形对象，比如收据。假定对象 600 具有预定形状，比如是一个矩形。然而，对象 600 的大小、方向和位置是未知的。

提取对象的第一步是将扫描图像 602 内的每个像素分类为背景像素或数据像素。在这个工作例子中，通过检查每个像素的颜色来执行这种分类。背景像素是位于对象 600 之外的像素。另一方面，数据像素是位于对象 600 之内的像素。假定背景的颜色 b （即对象 600 之外的像素值）是已知的或可以估算的。此外，假定至少对象 600 内的大部分像素是不同于 b 的，其差距超过一阈值。以数学语言表述，扫描图像 602 中的任何像素，如果其

$$|Im(i,j) - b| > \text{阈值}$$

该像素被定义为一个数据像素，且所有其他像素被定义为背景像素。应该注意到，可以使用一种颜色而非灰度方法来区别数据和背景像素，并且这个判定可以是基于一种比单独使用阈值更为复杂的方法。

接下来，使用建立在扫描图像 602 之上的轴对数据像素进行合计。这个工作例子中，一个二维直角坐标系 604 建立在扫描图像 602 上，这样 i 轴对应于水平方向（或行）而 j 轴对应于垂直方向（或列）。首先，计算每行内的数据像素的数量。方法是：对于固定的 j 值，对沿 i 轴的数据像素进行合计，指定为 $P(j)$ （这里 $P(j)$ 是一个在行或 i 方向的图像函数）。对 j 的所有值都这样执行。所得 $P(j)$ 的图表（在第 j 行的数据像素的总和）是第一梯形 620。其次，计算每列内的数据像素的数量。为此目的，对一个固定的 i 值，对沿 j 轴的数据像素进行合计，指定为 $Q(i)$ （这里 $Q(i)$ 是一个在列或 j 方向的图像函数）。对 j 的所有值都这样执行。所得 $Q(i)$ 的图表（在第 i 行的数据像素的总和）是第二梯形 630。

然后，在第一和第二梯形 620 和 630 进行初等几何计算。根据该几何运算可以断定 $P(j)$ 图表的顶部等于 $x \cos(\theta)$ 并且 $Q(i)$ 图表的顶部等于 $y \sin(\theta)$ ，其中 x

和 y 是对象 600 的维度而 θ 是方向的角度。对象 600 的角是四个坐标点 (g,a) , (h,c) , (f,d) 和 (e,b) , 它们对应于第一梯形 $P(j)$ 和第二梯形 $Q(i)$ 的拐点。

应该注意, 存在另一种情况, 其中扫描图像 810 中的对象会产生相同的 $P(j)$ 图表 (第一梯形 620) 和相同的 $Q(i)$ 图表 (第二梯形 630)。这种可能性在图 7 中显示。这种情况下, 第二对象 700 位于第二扫描图像 702 内。第二对象 700 和图 6 中显示的第一对象 600 具有相同大小, 但却有相反的方向 (即, 第二对象 700 朝向的角度是 $(-\theta)$ 而不是角度 (θ))。第二对象 700 有坐标 (h,b) , (g,d) , (e,c) 和 (f,a) 并且是唯一产生同样的梯形 620 和 630 的其他可能对象。

在这个单个对象的案例中, 可以确定要么是第一对象 600 要么是第二对象 700 出现在扫描图像 602 和 702 中。然而, 必须检查出现的是哪个对象。为了确定出现的是哪个对象, 对照扫描图像数据检查每个对象的顶点。然后使用最适合数据的对象而抛弃另外一个。换句话说, 分析每个矩形以确定该大小、位置和方向的矩形实际上包含几乎所有的其 $|\text{Im}(i,j)-b|$ 比指定的阈值大的像素。使用单个阈值区别数据和背景像素的情况是作为例子来使用的。更为复杂的策略是, 例如在彩色图像中使用所有三种颜色, 而不是灰度图像中的单种颜色可以产生更佳的结果。

2. 多个对象案

以上揭示的用于单个对象案的对象提取方法可以扩展到多个对象的案。总体上, 这涉及将多个对象案拆分为多个单个对象案, 后者可以通过如上描述来解决。如图 8 所示, 在第二工作例子中, 扫描的图像 800 包括多个对象, 即第一对象 801, 第二对象 802 和第三对象 803。在这个多对象案中, 使用与以上揭示的对象提取方法相同非方法, 不过是以递归的方式。

特别地, 与单个对象案相似, 扫描图像 800 中的每个像素被分类为数据像素或者背景像素。这种分类基于像素颜色而执行。接下来, 图像被定义为某个特定方向的数据像素的总和。在这个工作例子中, 计算沿轴线的数据像素的总和, 所得 $P(j)$ 的图表 (第 j 行的数据像素总和) 是第一梯形 810 而所得 $Q(i)$ 的图表 (第 i 行的数据像素的总和) 是第二梯形 812。应该注意, 在该案例中, 当扫描图像 800 是由多个对象组成, $P(j)$ 和 $Q(i)$ 的数量是由每个单个对象产生

的梯形的总和组成。

不做一些简化就来估算梯形 810 和 812 的参数是很困难的。然而，观察到第一梯形 810 中， $P(j)$ 图表在 j_0 处数据具有差异（或间隙）， j_0 是 $P(j)$ 等于 0 的位置。这说明此位置没有任何图像数据而且因此， j_0 行之上和之下的扫描图像 800 的各部分是被独立处理的。利用这一事实，对象探测和提取方法将扫描图像 800 划分为两个子图像：（1）顶部子图像 820（ j_0 之上的行）；和（2）底部子图像 822（ j_0 之下的行）。

一旦扫描图像 800 被划分，再次使用以上描述的对象探测和提取方法来处理每个子图像 820 和 822。特别地，在顶部子图像 820 和底部子图像 822 上计算两个方向($P(j)$ 和 $Q(i)$)的图像函数。参照图 8，可以看到顶部子图像 820 包含单个矩形（第一对象 801），如此，这个问题分解为以上所述的单个对象案。因此，，通过使用以上描述的用于单个对象案例的方法查找第一对象 801 的坐标。

底部子图像 822 包括第二对象 802 和第三对象 804。执行该对象探测和提取方法的另一次重复，底部子图像 822 内的每个像素基于像素颜色被分类为数据像素或背景像素。图 9 给出了这次重复的处理。特别地，计算 $P(j)$ 和 $Q(i)$ 的数量。所得 $P(j)$ 的图表是第一梯形 830，所得 $Q(i)$ 的图表是第二梯形 832。如图 9 所示，在 i_1 位置， $Q(i)$ 存在数据差异（或间隙）。这表明通过取 i_1 左边那些列（左边的子图像 834）和 i_1 右边那些列（右边的子图像 836），底部子图像 822 可被进一步划分。

F. 伪代码示例

作为例子而非限制，下面的伪代码描述了对象探测和提取方法的一个可能的实现：

```
function procMult(Im);
I0 = 0; j0 = 0; i1 = leni; j1 = lenj;
[P, Q] = getProjections(Im);
[gapsi, gapsj] = getGaps(P, Q);
if ((length(gapsi)-2)+(length(gapsj)-2)<1)
    drawObject(Im, P, Q);
```

```

else
    for m = 0:length(gapsi)-2
        for n = 0:length(gapsj)-2
            procMult(Im(gapsi(m):gapsi(m+1),
                gapsj(n):gapsj(n+1)))
        end
    end
end
end
end

```

调用的函数如下：

`[P, Q] = getProjections(Im)`

在图像区域计算 $P(j)$ 、 $Q(i)$ 的例程

`[gapsi, gapsj] = getGaps(P,Q)`

确定 $P(j)$ 、 $Q(i)$ 中的任何间隙的方位。对图 6 中图像的响应是

$\text{gapsi} = [0, i_{\max}]$ 和 $\text{gapsj} = [0, j_{\max}]$, 对于图 8, 是 $\text{gapsi} = [0, i_{\max}]$ 和 $\text{gapsj} = [0, j_{\max}]$.

`drawObject(Im, P, Q)`

检查梯形的 $P(j)$ 和 $Q(i)$, 估算它们的参数并确定是否任何矩形适合该数据。如果是, 增加顶点到全局列表。

G. 实现问题

以上的讨论假定在背景像素和数据像素之间没有混淆。在这种理想的情况下, 由图表 $P(j)$ 和 $Q(i)$ 构成的矩形会是理想的矩形并且拐点可以很容易地有把握地确定。

然而, 实践中, 有可能无法对所有的像素进行正确的分类。这种不精确性所产生的效应是, 梯形可能由于例如噪音而不同于理想状态。然而, 幸运的是, 因为图像函数 $P(j)$ 和 $Q(i)$ 被定义为一个方向上所有像素的总和, $P(j)$ 和 $Q(i)$ 函数本身就很健壮。此外, 因为这些矩形的上底通常是最多数的公值 (the most common value), 很容易从柱状图进行健壮估算。于是, 可以估算拐点为公值阈值之内的点。而且, 当确定是否在 $P(j)$ 和 $Q(i)$ 函数中操纵数据的差异或间

隙时,经常发生的情况是,噪声或背景颜色的误估计保证 $P(j)$ 和 $Q(i)$ 很少准确等于 0。

尽管用于这个工作例子的图像函数 $P(j)$ 和 $Q(i)$ 被定义为两个或多个不同方向的数据像素的总和,应该注意,其他的定义也是可以使用的。作为例子,图像函数 $R(j)$ 可以被定义为等于最右边的数据像素的列位置减去最左边的数据像素的列位置,另一个图像函数 $S(i)$ 可被定义为等于最顶端的数据像素的行位置减去最底端的数据像素的行位置。这种情况下, $R(j)$ 和 $S(i)$ 也能使对象探测和提取系统和方法有效地运作。实际上,在没有噪声时,应该注意,当图像是由单个矩形对象组成时, $P(j) = R(j)$ 并且 $Q(i) = S(i)$ 。

应该注意,查找到子图像内的单个对象的具体方法只是许多可能方法中的一种。一旦子图像被确定为只包括单个对象,许多其他方法都产生令人满意的结果。

H. 附加的实施例

一个实施例中,对象探测和提取过程被应用于图像的二次抽样版本。使用图像的二次抽样版本的优势是这避免了处理高分辨率的图像数据。

另一个实施例中,一旦确定子图像可能只包含单个对象,使用一个匹配算法来估算 $P(j)$ 和 $Q(i)$ 函数梯形最佳匹配。那么,最适合数据的梯形的拐点(或膝盖点)被用于形成对象顶点的估计。

在另一个实施例中,一旦已经查找到子图像内单个对象顶点的估计,对子图像内容的单个对象最佳匹配就确定了。是通过使用一种确定一个使子图像内的实际数据和所提议的矩形匹配之间的均方(或其他的计量)最小化的矩形的技术来达到这一目标的。

在再一个实施例中,如果证明不可能自动地确定扫描仪压盘的背景颜色,则用户可以用如鼠标的定位设备来指向一背景像素以帮助该进程。

另一个实施例中,如果算法没能正确地分割对象,用户可以指示对象的边界或角以帮助该进程。

II. 每个分割对象的光学字符识别和聚类

一旦对象探测和提取系统 202 (图 2 所示的) 已经输出每个标识的对象的

坐标并且这些对象已经从整个图像中提取出来，可对每个对象的图像进行处理以标识包含在对象中的有用的信息元素。这些信息元素然后可以聚集并提供给一模块用于帮助用户填充相关联的电子表格，比如在地址簿内的联络信息条目或在费用报告中的条目。

图 10 是说明依照本发明的一个实施例用于标识信息元素并聚集这些元素的方法 900 的流程图。在步骤 901，由对象探测和提取系统 202 标识的各个对象从总图像中提取出来。一个实施例中，每个对象是一个任意方向的矩形。这些矩形可以对应于名片、收据或其他类型的对象。

在步骤 902，每个对象被旋转到水平朝向，正面向上。如上所述，该对象可以以任意的取向随机地放在扫描仪上，比如正面向上、向一旁、颠倒的或介于其间的任何角度。

在步骤 903，在所有四个正交方向上使用光学字符识别（OCR）模块对每个旋转的对象的图像进行处理，以防万一在步骤 902 中对象被旋转到颠倒或朝向一侧。这些方向包括从假定的正面向上的水平位置旋转到 0 度、90 度、180 度和 270 度的方向。步骤 903 用于确定对象的方向、以及文本的上下文和位置信息。步骤 903 的输出是对象上识别的文本块及其它们的两维（2-D）位置的列表。

文本块可以包括任何信息元素，比如字母数字字符或其他符号的字符串。这些元素可以采用任何可用的形式，比如字词、数字或其他图形信息。

在步骤 904，在步骤 902 识别的文本块被聚集起来以标识文本区域。文本区域的例子包括：1）姓名和抬头（比如在名片的上部）；2）住宅、工作和手机号码以及传真信息；3）email 和 web URL 信息；和 4）图标和公司名称等等。这些文本区域的特点是这些区域内部的词间距在平均水平上小于不同的文本区域之间的词间距。这些文本区是不相重叠的。以下参照图 11 对如何将词聚集成文本区域的过程作更为详细的描述。

在步骤 905，每个特定对象的文本区域用适当的文本区域分隔符串接成单个文本流。例如，文本区域可以从对象的左上角到右下角按序串接。例如，文本区域分隔符可以包括换行符和分段符。然后，聚集的、串连的文本区域传给

帮助填充表格的模块用于填充相关联的电子表格，比如地址簿中的联络信息记录。

图 11 是说明聚集图 10 的步骤 904 中的文本区域的字词的方法的流程图。如上提及，图 10 的步骤 903 中识别的字词被聚集以标识文本区域。在步骤 910，该方法标识对象内两个最靠近的词作为一个文本区域。一个实施例中，基于对象内词的边界框之间的 x-和 y-距离来定义接近与否。例如，这些距离可以表示 Euclidean 和 Manhattan 距离。在步骤 911，聚集过程计算包含在文本区域中的词的平均 x 和平均 y 距离。

在步骤 912，该聚集过程查找到一个与文本区域最为接近的并且未包括在文本区域中的词。文本区域和未包含在文本区域中的词之间的距离的一个例子被定义为文本区域的词和任意词之间的最小距离。另一个例子是文本区域的词和边界框之间的距离。

在步骤 913，聚集模块确定最靠近文本区域的词的 x-和 y-距离是否比文本区域的平均 x 和平均 y 的特定倍数要小。独立因数可以用于 x- 和 y-距离。如果这样的话，在步骤 914，该词即加到在文本区域中并且聚集模块返回到步骤 912。如果不是这样，在步骤 915，从对象上的一组词中提取文本区域，且聚集模块返回到步骤 910 以查找对象上剩下的词中的下两个最靠近的词。这个过程一直重复直到对象上所有的词已经聚集到文本区域中为止。

III. 辅助的填充表格

识别的文本块的聚集文本区域然后存储在未标记媒体数据存储设备中，诸如图 1 中的本地或远端存储器设备中的一个上以供填充表格模块使用。

图 12 是系统 1000，用于依照本发明的一个实施例的产生和操作一个表格填充器界面 (FFI) 1002 以便于辅助电子表格的填充。此例中，表格是地址簿中的联络信息记录。填充表格模块将表格填充器界面 1002 呈现给用户以帮助用户通过转变来自聚集的文本区域的未标记数据成为已标记数据（比如 XML 格式的数据）或到一个数据库中来帮助用户填充表格。FFI 1002（在这里也称为计算机“屏幕”）包括表格数据图形用户界面 (GUI) 1004（这里也称为“表格”）和对象数据图形用户界面 1006（这里也称为“文本框”），它们可以彼

此相邻以方便用户。

表格数据 GUI 1004 包含很多字段 1008, 使得每个字段保留给一个特定信息块 (比如姓、名、街道地址、邮政编码等等) 使用。状态指示符 1010 可以与每个字段相关联以便通知用户特定的字段中信息的当前状态。置信指示符 1012 也可以与每个字段 1008 相关联以便通知用户与字段 1008 中的信息正确性相关联的概率。此外, 表格填充器界面 1002 可以显示要处理对象的图像 (未在图 12 中示出)。

图 12 中说明的 FFI 1002 是作为用于输入地址簿中的联络信息的界面的例子。最初, 表格数据 GUI 1004 和空白文本框 1006 显示给用户。用户可以从数据存储器中复制聚集的文本区域到文本框 1006 (比如, 通过来自屏幕上的显示窗口中的剪切和粘帖)。或者, 填充表格模块可以自动插入从关联的对象 (比如名片、帐单或收据) 中获得的聚集的文本区域到文本框 1006。例如, 如果原始图像包含多个对象, 该填充表格模块为每个对象填充一个联络信息记录。

该填充表格模块可以尝试分类或解析未标记的对象数据以标识文本框 1006 内的对象数据内的信息元素。一旦该对象数据已经被解析, 该模块用标识的元素来填充表格 1004 的字段 1008。文本框 1006 和表格 104 内原始未标记的对象数据可以同时显示在屏幕 1002 上, 并且现在已标记对象数据可以是扩张的 (augmented) 以便以可视方式指出关联性 (比如, 使用颜色编码或其他视觉指示符)。例如, 系统 1000 可以使用紫色来指示文本中的已经被用于填充表格 1004 中的地址字段的某些元素。根据例子, 可以使用一种单独的颜色 (比如橙色) 来指示模块已经确定哪个特定文本是有潜在兴趣的, 但信用级别还未高到足以将它分配给字段, 因此用户可以决定特定的文本是否应该分配给特定的字段。

根据本发明的一个实施例, 用户可以填充表格 1004 的一部分, 并且该填充表格模块可以搜索文本框 1006 中可用的对象数据、定位潜在的输入字段候选者、显示定位的元素并填充表格中的剩余字段。以这种方式, 可以执行部分自动填充。

万一填充表格模块未能正确地标识一块文本, 比如名片上的公司名称, 就

有可能聚集那块文本区域。用户可以从文本框 1006 中拖拉这块文本到适当的字段 1008，例如，使用一种定位设备。对于扫描诸如收据的文档的应用来说，这是非常有用的。收据上可能有许多块数字和文本，其中用户只对如卖主姓名、日期、最后金额、可能还有税金的输入字段感兴趣。只要文本区域被聚集并显示在文本框 1006 中，用户就可以拖拉适当的文本块到适当的字段。

用户可以迅速校验解析的正确性。如果解析有错误，用户可以纠正它们，比如通过从文本框 1006 中拖拉元素并放置到表格 1004 相应的字段 1008 中、通过直接键入到字段 1008、和通过纠正文本框 1006 中的文本。此外，解析协议可以利用边缘信息（side information），比如用户提供的先前的修正或添加。例如，如果用户已经输入信息到一个字段或纠正了最初的解析，用户可以指令系统重新解析对象数据并依靠由用户提供的边缘信息（通过单击图 12 中标做“Auto Fill（自动填充）”的按钮）。

例如，如果姓名为“John Smith”从名片中提取出来，这暗示着“John”是一个特定联络信息的名而且“Smith”是姓。然而，用户可以看到名和姓在最初的对象中，无论是偶然还是其他，已经被调换过并且可以使用以上描述的拖放技术来移动“John”到名字段。此外，可以提供带有下拉菜单的字段，这样其中显示在文本框 1006 中的对象数据包含一个以上的名，例如，其中一个名可以显示在名的字段中，其他的可以在下拉菜单中被提供。用户可以简单地打开菜单（比如单击或停留在该字段上），且如果该字段需要纠正时，选择一个替代的姓名。

在这样做后，系统可以自动移动“Smith”到姓字段，从而减少了填充表格所需要的用户动作的数量，同时为姓字段提高了置信用水平，因为所基于的事实是，用户纠正了“John”为联络信息中的名并且因此不是联络信息中的姓。这样的自动化用户动作后字段填充是纠正增殖（correction propagation）的一个例子。

一些情况下，允许用户指定哪些字段可以用作边缘信息是有利的。例如，这些字段可以包括那些由用户填充或纠正过的。用户可以指定其他字段可由系统重写。因为有状态指示符 1010，这种准许是方便的，状态指示符 1010 可以

指示用户没有作用于该字段、或已经校验、纠正和/或输入信息到该字段的状态指示符 1010。每个字段的状态可以是，例如，“未填充和未校验”、“自动填充但未校验”或者“用户填充或自动填充并已校验”。

例如，“未填充和未校验”的字段可以有第一种颜色（比如红色）的状态指示符 1010。如果系统 1000 填充了该字段（比如自动填充该字段），那么状态指示符可被升级到第二种状态指示符的颜色（比如黄色）以提醒用户该字段已经被自动填充了但未校验。这样一个指示符可以提醒用户需要用户校验但却未必要纠正的一种状态，如“John Smith”的例子。如果用户校验了该字段内的信息是正确的，状态指示符可被升级到第三种颜色（比如绿色）以指示“填充并校验”的状态。进一步看这个例子，如果用户输入信息到具有红色的状态指示符的字段，那么状态指示符可以直接升级到绿色，因为用户通过这样做已经填充了字段并校验了该信息是正确的。因此字段现在为“填充并校验”。此外，另一个字段的或另一些字段的置信读可以通过用户对第一个字段的校验和/或纠正而被更新和/或改善。例如，在“John Smith”的例子中，如果哪一个名称是名哪一个名称是姓未经校验的话，名和姓的字段都有一个黄色的状态指示符。如果用户校验了“John”是正确的名，那么模块可以升级名的字段的状态到“用户填充并校验”（比如用一个绿色的状态指示符）。因为用户已经校验了“John”是名（并且因此不是姓），系统可以保留“Smith”在姓的字段，并且因此姓的字段的置信指示符也可以从黄色升级到绿色（比如自动填充并校验）。

依照本发明的有关方面，颜色编码的置信指示符 1012（比如，图 12 中显示的围绕字段画出的框或者字段的边界颜色、字段和/或文本的背景颜色等等）可以与特定的字段 1008 关联。例如，对系统 1000 而言，很难用高置信因数填充的字段能够依照可以指示用户字段内的信息是低于期望置信度阈值的颜色模式来标记。置信指示符可以表示一个从 0 到 1 的值，是颜色不同色度。而且，这个例子中的置信指示符 1012 可以是，例如，实心指示符、闪烁指示符、全亮度和对比度等淡入淡出指示符、或任何其他适合的可以指示有关字段的各种置信级别的指示符方案。

例如,包含“@.”或“.com”的一块信息可以自动地插入到表格中的“email”字段。类似地,一块具有(nnn) nnn-nnnn、 nnn-nnn-nnnn、 或 nnn-nnnn 等等格式的信息(其中 n 是一个整数)可以自动地以高置信度插入到电话号码字段。应该意识到,高置信读标记可以与可插入信息的有关字段的其他类型的信息相关联,且这种信息的自动插入不限于 email 和/或电话号码字段。

图 13 是对与依照本发明的可替代实施例的填充表格界面的说明。与图 12 的实施例类似,填表界面 1102 包括一带有多个对应于不同信息类型的字段 1108 表格数据 GUI 1104 和一用于每一字段的状态指示符 1110。然而,填表 GUI 1102 进一步包括一个用于显示一个诸如从光电扫描仪获得的电子图像 1122 的预览窗格(GUI) 1120。图像 1122 包括从图像 1120 中由图 3 中的对象探测和提取模块 202 分割的多个对象,比如名片 1124。一个实施例中,每个对象 1124 由一个围绕对象的彩色边界突出显示。例如,每个对象可由红色边界加以突出显示。

因为各个对象 1124 已经从整个图像 1122 中分割出来,用户可以通过例如在特定对象 1124 上移动光标和单击该对象来个别选择每个对象。然后,对象可以显示在对象窗格(GUI) 1106 内。对象窗格 1106 与图 12 中的对象数据 GUI 类似,但已作了修改以用由文本块 1130 表示的对象数据(解析的或未解析的)显示选中的对象 1124。

每个文本块在对象窗格 1106 内通过例如围绕相关联文本的彩色框 1132 来标识。属于相同信息类型的文本块可以用带有相同颜色的框 1132 来突出显示。因此,来自不同集群的文本块会有不同颜色的框 1132。这个颜色可与任何使用的颜色协调以标识不同的信息区域 1140。例如,文字“Tooth Fairy, Inc.”标识了一个公司名称并且可用蓝色框 1132 来突出显示,它可以与相应信息区域 1140 协调相同的颜色。文本块中的每个词或记号可以有自己颜色框 1132,如图 13 所示,或者所有同一文本块的词可用单种彩色框 132 来突出显示。类似地,街道地址文本块可以有不同颜色的彩色框,比如紫色。未使用的文本块,比如“Magical Figurines”可以用另一种其它颜色框 1132 来突出显示。

与图 12 的实施例类似,颜色关联和突出显示的文本帮助用户校验解析的

数据并更新或纠正已经由填表模块填充的任何字段 1108。用户可以直接在字段 1108 内打字、拖放来自对象窗格 1106 的信息元素到字段、或者通过字段中的下拉菜单从多个信息元素中进行选择。例如，填表模块可以标识“John Smith”和“Jim Doe”为两套不同的名和姓。如果填表模块在“First Name（名）”和“Last Name（姓）”字段 1108 中输入了不正确的姓名，用户可以简单地通过以上方法的一种来选择正确的姓名。在作出纠正时，填表模块可以重新解析文本块以利用新的“边缘信息”，使得相关字段可以自动更新。

同样，状态指示符 1110 指示特定字段内的任何信息的状态。这些指示符可以指例如，“未填充和未校验”、“自动填充但未校验”，或者“填充并校验”。

图 13 也说明了一个例子，其中图像 1122 中的其中一个对象 1124a 朝向任意角度。此例中，对象 1124a 的朝向相对于水平方向成锐角。当被选中时，对象 1124a 显示在对象窗格 1106 内。然而，因为对象探测和提取系统 202（图 2 所示）已经标识了每个对象 1124 的坐标并且这些对象已经从整个图像中被“提取”出来，每个对象的子图像可以旋转到朝向垂直方向、正面向上，如对象窗格 1106 内所示。这同样允许 OCR 模块在每个对象上执行可靠的字符识别并聚集识别的文本块用于由填表模块进行解析。因此，对象在整个图像 1122 内可以有任意的方向，比如介于并包括 0 度到 360 度的任何角度。

图 14 是对通过例如填表界面 1002（图 12 中所示）或填表界面 1102（图 13 中所示）而便于辅助填充表格的填表模块或系统 1200 的说明。来自图 12 和 13 的标号被包括在图 14 中以表明与两个实施例中相类似的元素。系统 1200 包括一个控制组件 1202、解析组件 1208、未标记媒体数据存储 1210、表格数据存储 1212 和边缘信息存储器 1214。控制组件 1202 可操作地耦合到数据 GUI 1004、1106、对象数据 GUI 1006、1106 和解析组件 1208。在本发明中，术语“组件”指的是计算机相关实体，是硬件、软硬件的结合、软件或者执行中的软件。例如，组件可以包括但不限于：运行于处理器上的进程、处理器、对象、可执行的程序、执行的线程、程序、和/或计算机。作为说明，运行于服务器上的应用和服务器都可以是计算机组件。一个或多个组件可以驻留于一进

程和/或执行线程中，并且组件可以局限于一台计算机上和/或分布在两台或更多台计算机之中。“线程”是操作系统的核心在进程中调度来执行的一个实体。如本领域所熟知的，每个线程有一个相关联的“上下文”，它是与线程执行有关的易失性数据。线程的上下文包括系统寄存器的内容和属于该线程进程的虚拟地址。因此，构成线程上下文的实际数据随它的执行而改变。

控制组件 1202 可以接收并分析未标记对象数据以便方便填充表格内的字段。这样的未标记数据可以通过对象数据 GUI 1006、1106 呈现给用户。未标记数据可以是，例如，从名片、发票或购货收据中识别的文本。聚集在文本区域内的未标记数据可被存储在未标记媒体存储器 1210 中。解析组件 1208 解析存储在未标记媒体存储器 1210 中的未标记数据以标识信息类型并确定潜在的表格填充器数据。如上提及，表格填充器数据可以包括专有名词，比如姓名、数字数据集、地址、电话号码、邮政编码等等，它们随后可以被存储在表格数据存储器 1212 内。存储在表格数据存储器 1212 内的数据可以用于填充表格内的字段，并通过表格数据 GUI 1004、1104 呈现给用户。同样地，对象数据 GUI 1006、1106 内的标记的、解析的对象数据可以用视觉指示符来突出显示以标识特定的信息类型或该数据所属的字段。

如参照图 12 和 13 所描述的，然后用户可以校验和/或纠正表格内的各个字段，并且这种校验和/或纠正可以作为边缘信息存储在边缘信息存储器 1214 内。解析组件 1208 可以根据用户所作的校验和/或改变，使用存储的边缘信息以更新表格数据存储器 1212。以这种方式，文本分类和/或标注可以更新，这允许与自动填充的字段相关联的状态级别响应于用户对字段的校验和/或纠正而被升级，从而方便了纠正增殖。

应该意识到，这里描述的数据存储器（比如存储器）组件可以包括，例如，参照图 1 所描述的任何本地或远端存储器，并且可以包括易失性存储器或非易失性存储器，或者可以同时包括易失性存储器和非易失性存储器。

本发明的一个实施例中，解析组件 1208 包括一个人工智能（AI）组件，该组件可以就最适当的可以输入一块特定数据的字段做出推论。本发明中，术语“推论”一般是指就一组由事件和/或数据捕获的观察资料进行的推理过程、

或者系统、环境和/或用户从一组由事件和/或数据捕获的观察资料进行推断的状态。推论可以用于，例如标识具体的上下文或动作，或者产生一个对状态的概率分布。该推论是盖然性的。也就是说，推论可以包括基于对数据和事件的考虑对诸状态计算一个概率分布。推论也可以指用于从一组事件和/或数据构建更高级别事件的技术。这种推论导致了从一组观察的事件和/或存储的事件数据构造出新事件或新动作，无论这些事件在时间方面是否彼此紧密相关，并且无论这些事件和数据是来自于一个还是许多个事件和数据来源。各种不同的分类方案或系统，比如支持矢量机器、神经网络、专家系统、Bayesian 信仰网络、模糊逻辑和数据融合引擎，可以用于结合本发明在执行和/或推断动作中应用。此外，在本发明的一个实施例中，推论可以基于例如隐藏 Markov 模型（HMM）来作出。

图 15 是说明依照本发明的一个实施例使用 HMM 来方便辅助填表的图表 1300。HMM 和其他盖然性的模型可被应用于从用户界面到解析器“后信道（back-channel）”信息以便于纠正增殖，这允许当单个字段被用户纠正时其相邻字段得以修正。HMM 是一个具有一组状态 Q 、一输出字母表 O 、跃进概率 A 、输出概率 B 和初始状态概率 Π 的有限状态机的变型。典型地，当前状态是不可测见的。相反，每个状态可以以某种或然性 B 产生一个输出。通常状态 Q 和输出 O 是易懂的，因此 HMM 可以说是一个三变量 (A, B, Π) ，带有下列属性：

$$A = [a_{ij} = P(q_j \text{ at } t+1 \mid q_i \text{ at } t)]$$

$$B = [b_{ik} = P(o_k \mid q_i)],$$

$$\Pi = [p_i = P(q_i \text{ at } t=1)].$$

符号 $P(a \mid b)$ 表示“a”在给定“b”时的条件概率。以上的方程式中， A 是转变到下一个状态“ q_j ”（在时间 $t+1$ ）的概率，假定当前状态为“ q_i ”（在时间 t ），这里 $q_i \in Q$ 。 B 是输出是 O_k 的概率，假定当前状态是 q_i ， $O_k \in O$ 。对每个状态指数“ i ”， Π 是在时间 $t=1$ 时，状态为 q_i 的概率。

依照图 15，图示了从 X_1 到 X_n 的各种随机变量，它们可以代表表格内的字段。这样的字段可以是包括{名，后缀，姓，街道地址号码，街道名称，城市，州，邮政编码，电话号码，email 地址等等}的一组字段的部分。应该懂得，该 X 字

段的组和可输入其中的信息元素 Y 并不限于以上描述示例性信息字段，而是可以包括任何其他适合的信息和/或字段。 Y 可以代表与给定的 X 相应的实际信息元素，使得如果 Y_1 等于“John”并且 $X_1 = \text{“名 (first name)”}$ 为真（比如， $P(X_1 = \text{名 (first name)}) = 0.23$, $P(X_1 = \text{姓 (last name)}) = 0.03$, $P(X_1 = \text{城市 (city name)}) = 0.093$ 等等），则使得可以选择展示最高分的标签（比如，本例中的“first name”）。这种推论方便了查找隐藏变量最佳设定。在 Hidden Markov 模型的情况下，可以查找到最可能的状态序列。例如：

$$\arg \max_{x_1, \dots, x_n} p(Y_1 \dots Y_n | X_1 = x_1, \dots, X_n = x_n)$$

根据一类似的例子，特定的 X 可能具有与其相关联的条件“5 个阿拉伯数字”，这样如果 Y 有七个数字（比如 555-1234），那么对于讨论中特定的 X ，会登记一个低概率（比如， $P(Y=555-1234|X)=0.00001$ ）。相反地，包括比如 12345 的信息的 Y 会为特定的 X 记录一个高的概率（比如， $P(Y=555-1234|X)=0.9989$ ）并且可以插入到表格内相关联的字段中。类似地，七个数字 Y 会为具有条件“7 个数字”的 X 记录一个高的概率。本发明可以使用任何数目的适合变量或测试以确定哪一个特定的 Y 满足与特定的 X 有关联的条件，以方便辅助填表。

本发明的一些实施例可以通过将隐藏变量 (X_s) 设定为与特定字段的标签相应的状态来利用盖然性模型，比如以上讨论的 HMM 的优势，该模型包括隐藏的和观察的随机变量。例如，以上描述的 HMM 中的随机变量 Y 是“观察的”随机变量，其中每个变量对应于一个记号。记号是在记号分隔符（比如，空格、破折号、逗号等等）之间的一段文本。例如，文本串“this-is a, test”可以被记为：

“this” = token 1

“is” = token 2

“a” = token 3

“test” = token 4

隐藏变量 X_s 代表了记号具有每个被允许的标签（比如，记号在标签上分布）的概率。在信息提取领域内，最经常地，众 x 仍为未观察到的，因为没有使用“边缘信息”。为了强制一个盖然性模型使用边缘信息（比如以带有用户提供的文本的文本字段的形式），对应于用户提供的文本的记号可以被搜索到

并且相应的隐藏变量 x 可以被设定为与字段标签相一致的状态。这可以被认为是设定 $p(X1 = \text{First Name}) = 1$ 和 $p(X1 = \text{Last Name}) = 0$ 等等，并且在推论期间不更新。例如，如果用户打字“Smith”到表格的姓字段，可以在所有的记号内进行搜索以寻找“Smith”。然后，设定 $p(X2 = \text{LastName}) = 1$ ，并且在推论期间不用为 $p(X2)$ 更新概率分布。

纠正增殖可以进一步获得自用户界面到解析器后信道的信息。以这种方式，当用户纠正单个字段时，其相邻字段可以被填充。例如，本发明可以使用一种基于规则的解析方法，其中一个规则的简化版本规定“如果用户设定 LastName（姓）的字段，那么在未标记文本中搜索姓并且将紧接在姓前面的字标记为名”。同样存在名的相应的规则。以这种方式，姓的纠正“增殖”到名。应该懂得，如这里所述的纠正增殖并不限于名和姓，而是可以应用于任何和所有有关类型的信息、文本等等。

此外，本发明的一些实施例可以采用条件随机字段（CRF），它们是 HMM 和最大熵模型的概括。CRF 允许引入任意的非本地特征并且捕获标签间的依赖性，允许评估解析信息块的置信度。以这种方式，当信息具有高置信度级别时，本发明可以自动分配一块解析的信息到一个字段，并且可以标记信息元素为具有低置信度级别供用户审视和/或纠正。

图 16 是一个柱状图 1400，它显示了在纠正随机错误字段之前和之后的 CRF 之间的关系。正如下面更为详细的描述，在图 16 中，根据每个表格中包含错误的字段的数量进行了表格分组。实心棒指示修正之前的 CRF，空心棒指示随机错误字段已经被修正后的分布。在填表过程中，与字段具有和纠正有关的用户行为可以通过许多用户交互模型（UIM）来预期和/或建模。例如，在一个简单情形，UIM1 中，用户可以被呈现一个自动填充的表格并且被要求纠正所有的错误（比如，没有执行纠正增殖）。因此，所要求的用户动作的数量等于发生在自动填充期间错误的总数量。

根据第二种情况，UIM2，假定最初是自动的字段分配，并且用户执行了单个随意选择的纠正，基于该纠正，系统可以发动纠正增殖。如此反复直到所有字段都是正确的。

根据第三种情况，UIM3，假定最初是自动的字段分配，并且用户在最低置信度的错误字段上执行了纠正。例如，用户可以被以视觉方式以置信度次序提醒了到各字段，比如通过图 12 中的置信指示符 1012，直到错误被查找到为止。纠正增殖可以依照最低置信度字段的纠正执行，并且可以提示用户纠正任何剩余的错误。

表格填充通常要求完美的精确度。因此，每当减少了填表时间、减少了用户可感知的负担或二者皆有时，都是可以认识到其益处的。除其他标准性能量度外，本发明的一个实施例使用功效量度，叫做期望的用户动作数(ENUA)。ENUA 被定义为用户纠正填充表格内所有字段所需要动作（比如，鼠标点击等等）的数量。ENUA 可以因 UIM 而变，如上所述。为了表示 ENUA，使用符号 $P(i;j)$ ，它是在 i 个手动纠正之后在错误 j 的数量上的概率分布。这样的分布由图 16 的柱状图表示。

UIM1 中，例如，ENUA 为：

$$ENUA = \sum_{n=0}^{\infty} nP(0;n)$$

其中 $P(0;n)$ 是在错误字段数量上的分布。

根据模型 UIM2 和 UIM3，例如，ENUA 是

$$ENUA^1 = (1-P(0;0)) + \sum_n nP(1;n)$$

其中 $P(0;0)$ 是所有字段最初被正确地分配的概率， $P(1;n)$ 是一个字段被纠正后在表格内错误字段数量之上的分布。根据使用哪一个 UIM，可以导致不同的分布。ENUA¹ 的上标 1 指示纠正增殖已经执行了一次。

仍参照图 16，根据每个表格内包含错误字段的数量，表格被分组。实心棒指示在任何修正之前使用基于 CRF 的剖析器的结果，空心棒指示在一个任意错误字段已被纠正之后的分布。可以使用这种信息来分别估计 $P(0;n)$ 和 $P(1;n)$ 。

图 17 是说明依照本发明的一个实施例的自动填充表格辅助的方法 1500 的流程图。尽管一个或多个方法是作为一系列动作或步骤被显示或描述的，但应该懂得，本发明不被该步骤的次序所限制，因为根据本发明，一些步骤能以一种与此处显示或描述的不同的次序和/或同时与其他步骤发生。例如，本领域

的技术人员会懂得，一个方法可另表示为一系列相关的状态或者事件，比如在一个状态图中。而且，并非所有说明的步骤对于实现一个依照本发明的方法而言都是需要的。

在 1502，选择的未标记数据被插入到对象数据 GUI 的文本框中。在图 12 中显示的例子中，未标记数据显示在对象数据 GUI 1006 中。在图 13 显示的例子中，对象数据显示在对象数据 GUI 1106 内的对象图像中。在 1504，对象数据被解析以确定能够潜在地被使用以填充表格内的特定字段的元素。可以在 1506 分配状态给输入到字段内的元素并指示给用户。例如，已选择的未标记数据比如“John Smith”和“Jane Doe”包含两个名和两个姓。如果“John”用于填充例如联络名单的“名”字段，那么它可以具有一个相关联的状态指示符（比如“填充但未核对”），其可以提醒用户注意“John”可能不是名的字段内的正确输入这一事实。此外，通过下拉菜单，用户可以使用“Jane”从而方便了潜在的用户对名字段的修正。指示符可以是，例如，“名”字段旁边的彩色编码的状态指示符“灯”。进一步看这个例子，可以使用红—黄—绿协议来指示不同的状态级别，其中红色指示填充但未核实，且绿色指示填充（自动地或者由用户）并核实的字段。当前例子中，First Name（名）字段可以有一个黄色状态指示符，它指示了 First Name 字段被填充但名“John”未经核实。

一个实施例中，该方法可以直接前进到步骤 1510，其中提示用户校验或纠正展示了任何低于例如，绿色状态（比如，其中绿色指示填充并校验的状态）的字段。另一个实施例中，该方法首先前进到 1508，其中做出一个关于是否所有的字段展示了一个最高可能的状态（比如，是否所有字段是“填充并校验”）的决定。如果在 1508，所有的字段显示“填充并校验”状态，那么不需要提示用户采取行动并且进程终止。

然而，如果任何字段展示了低于“填充并校验”状态，那么该方法可以前进到 1510，那里提示用户纠正和/或校验任何可疑的字段。在 1512，做出一个关于是否所有的用户已经纠正（比如改变）了任何信息的决定。依照本发明，如果“John”不是“First Name”字段内需要的输入，那么用户可以在文本框（或对象窗格）中单击“Jane”并拖拉“Jane”到 First Name 字段以纠正输入。或

者，可以从一个已经显示在 First Name 字段内的下拉菜单中选择“Jane”。如果用户已经纠正了任何信息，那么该方法可以前进到 1514，那里可以根据用户输入来更新一个或多个字段，并且文本框 1006（图 12）内或对象窗格 1106（图 13）内的未标记数据可被重新解析。该方法然后可以回到 1506 用于状态升级和数据输入到表格字段，这可以就用户输入而发生。

如果用户没有在 1512 纠正信息，那么在 1516 作出一个关于用户是否已经校验了字段输入的决定。在 1516，如果用户没有校验带有低于需要的状态的字段输入，那么该方法回到 1510 用于进一步提示用户采取行动。如果用户在 1516 校验了正确的信息，那么字段和它们相应的状态可以在 1518 更新。例如，如果“John”对于 First Name 字段是需要的输入，那么状态指示符可以从黄色升级到绿色。

图 18 是说明依照本发明的实施例的方法 1600 的流程图。在 1602，未标记数据被解析。在 1604，隐藏 Markov 模型（HMM）被应用以确定一个特定元素可以输入的适当的字段。在 1606，元素可以显示在带有状态指示符的已确定的适当字段内。在 1608，可以提示用户校验和/或纠正在字段内输入的信息。在 1610，作出关于是否探测到用户纠正的决定。如果是，那么在 1612，用户纠正的字段可以通过纠正增殖与其他字段一起更新，并且它们相应的状态指示符可以因此而升级。然后该方法可以回到 1606，在那里，可以依照用户输入显示元素并指示状态。如果在 1610 未探测到纠正，那么在 1614 作出有关用户纠正是否已经发生的决定。如果用户没有校验输入的需要纠正的信息，那么该方法可以回到 1608 用于进一步提示用户采取行动。如果在 1614，确定用户已经校验了可疑字段需要纠正的信息，那么该方法可以前进到 1616，其中校验的元素被显示在适当的字段内并且升级的状态被显示。

图 19 是说明依照本发明的另一个实施例的方法 1700 的流程图。在 1702，未标记对象数据被读入未标记媒体存储器。在 1704，边缘信息（比如，从用户动作，比如数据输入、确认、纠正等收集的信息）被读入边缘信息存储器。在 1706，未标记数据被解析以标识能够潜在地填充表格字段的元素。标识的元素可在 1708 被写入到表格数据存储器中。然后，在 1710，标识的元素可在表格

GUI 内的表格字段内显示给用户。在 1712, 对象数据 GUI 内的对象数据可以用方便了辅助用户填充表格字段的视觉指示符来显示。例如, 对象数据 GUI 内的名可以是以特定颜色 (比如橙色) 彩色编码的以便指示它们可以被输入到表格 GUI 内的 First Name 字段, 该字段也可以是橙色彩色代码。依照另一个例子, 解析的包括 “@” 符号的对象数据可以用例如蓝色编码以指示这种文本可以输入到表格 GUI 的 “email” 字段, 该字段也可以是蓝色的。

在 1714, 在表格 GUI 内提示用户对元素到字段的分配进行校验和/或纠正。然后, 在 1716, 作出再次解析对象数据的决定。如果作出这样一个决定, 那么在 1718, 用户输入被添加到边缘信息存储器, 并且该方法回到 1706 用于未标记数据解析和元素标识的再次重复。如果确定在 1716 不需要附加的解析, 那么在 1720, 该表格数据存储器的内容可被写入到数据库或文件。

显示在图 17-19 的方法可以对从整个扫描图像中提取的每个单个对象而言执行。这个例子中, 其中每个对象是名片, 每张名片上的文本信息被解析并且用于填充相应的表格, 比如地址簿中的联络信息记录的表格字段。因此, 为每张包括在图像中的名片创建一个联络信息记录。每张名片提取的图像也可以与联络信息记录一起存储。

这些例子中, 用户不需要分别扫描每张名片。而是可以一次扫描许多张名片。从整个图像中, 系统提取每张名片的图像, 然后标识每张名片上的信息元素并且帮助用户分配这些元素到各个联络信息记录中的相应字段。这大大提高了从无数名片输入数据的效率。

在每个对象是购货收据的例子中, 收据上的文本块被聚集并被显示在未标记的文本框 1006 (图 12 显示) 中。图 13 显示的例子中, 每个收据可在预览窗格 1120 内被分别选择并被显示在对象窗格 1106 内。收据内的文本块被解析并被用于填充相应的表格, 比如费用报告或其他金融软件应用内适当的字段。收据上可能存在多块数字和文本, 其中用户只对比如卖主姓名、日期、最后金额和可能还有税金的输入字段感兴趣。只要这些文本区域被标识并显示在对象数据 GUI 1006 或 1106 中, 用户就可以拖拉适当的文本块到适当的字段 1008 (图 12) 或 1108 (图 13)。

一个实施例中，用户可以一次扫描几个收据，拖放日期、金额和/或每个收据内的其他文本块到金融软件应用，比如费用报告应用、电子数据表或货币管理软件比如 Microsoft Money™内的适当字段。收据的图像可以存储以作参考和/或与费用报告一起发送。对于费用报告填充系统，图像文件的一个用密码写的散列可以使用付费方的公共密钥加密以防止篡改数字图像。

另一个实施例中，系统能够从单个图像中提取多个不同类型的对象。例如，一次可以扫描多张名片和收据并且从整个图像中提取出每个对象。一旦每个对象的文本元素已被标识和/或被聚集，文本元素可被处理以确定被扫描的对象类型。具有联络信息的对象，比如公司名称、个人姓名、住址、电话号码、email 地址等等，有可能是名片。具有卖主姓名、日期和在代表金额栏内有数字的则可能是收据。其他类型的对象也可以被扫描。基于特定类型的对象，系统帮助用户输入文本到适当的电子表格。例如，在图 13 显示的实施例中，系统显示了对象数据 GUI 1106 中选择的对象和表格 GUI 1104 内的适当表格的字段 1108。或者，系统可以在对象数据 GUI 1106 内显示对象的图像，并提示用户在显示适当的电子表格的字段来完成之前，标识对象的类型。

应该意识到，本发明的系统和/或方法可以用在 web 爬行搜索系统，促进计算机组件以及非计算机相关组件。更进一步，本领域的技术人员会承认，本发明的系统和/或方法在大量的电子相关技术中是可以应用的，这些技术包括但不限于计算机、服务器和/或手持电子设备等等，它们可以是有线的和/或无线的等等。

以上所描述的包括本发明的例子。当然，为了描述本发明，不可能描述每一种组件和方法的可以想见的组，但本领域普通技术人员可以认识到，本发明的其它组合和改变是可能的。因此，本发明意在包括在所附权利要求书的精神和范围之内所有这样的改变、修改和变更。

尽管本发明已参照有关首选实施例作了描述，本领域的技术人员会认可，形式和细节的变化是可以作出的，而不会脱离本发明的精神和范围。例如，表格可以从一个或多个对象的任何电子图像中被填充。该图像可以通过任何类型的数字图像设备，比如光电扫描仪或数字照相机而获得。这个对象或这些对象

可以包括任何类型的具有有用文本信息，比如名片、帐单或购买收据的文档。

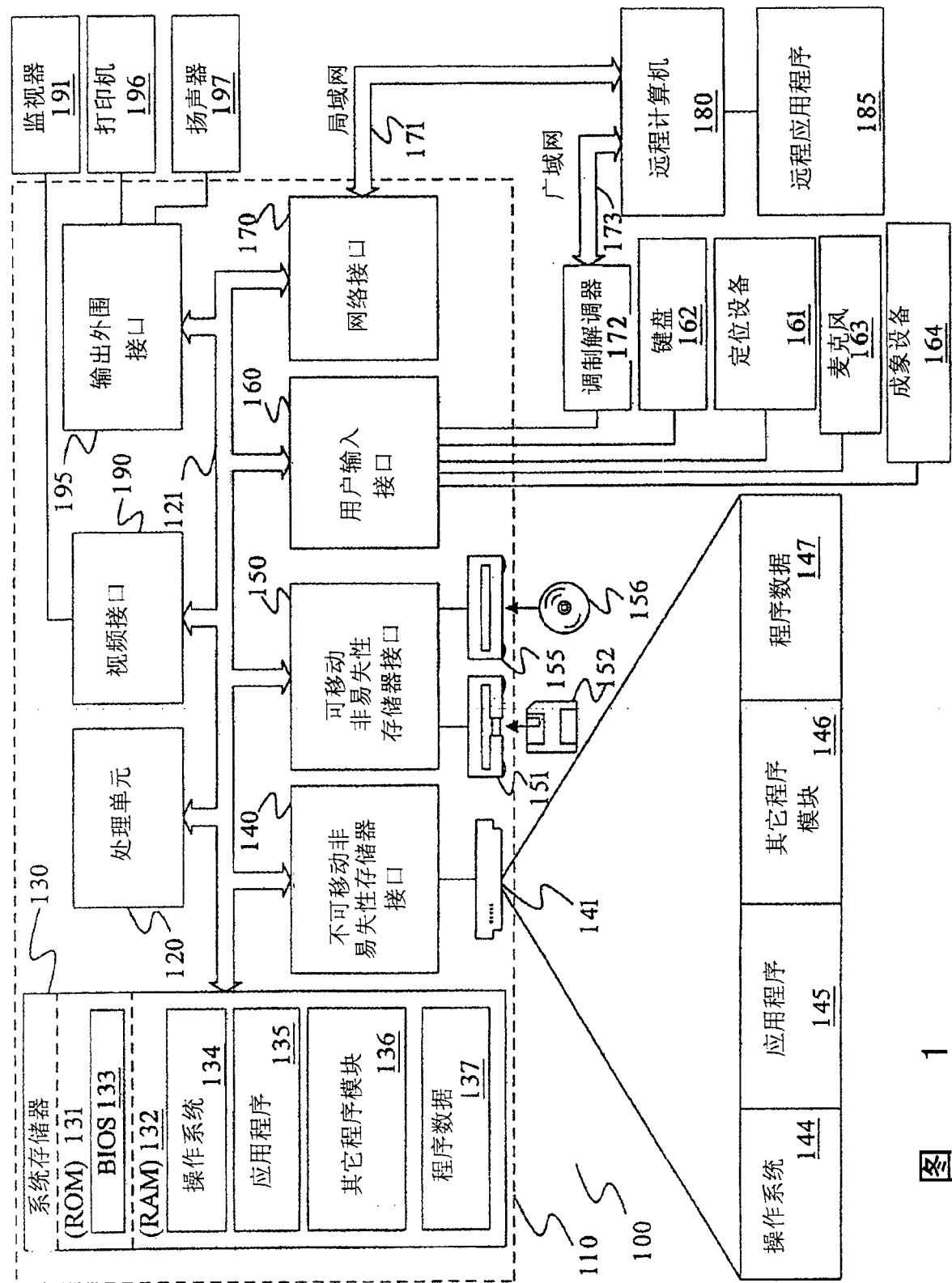


图 1

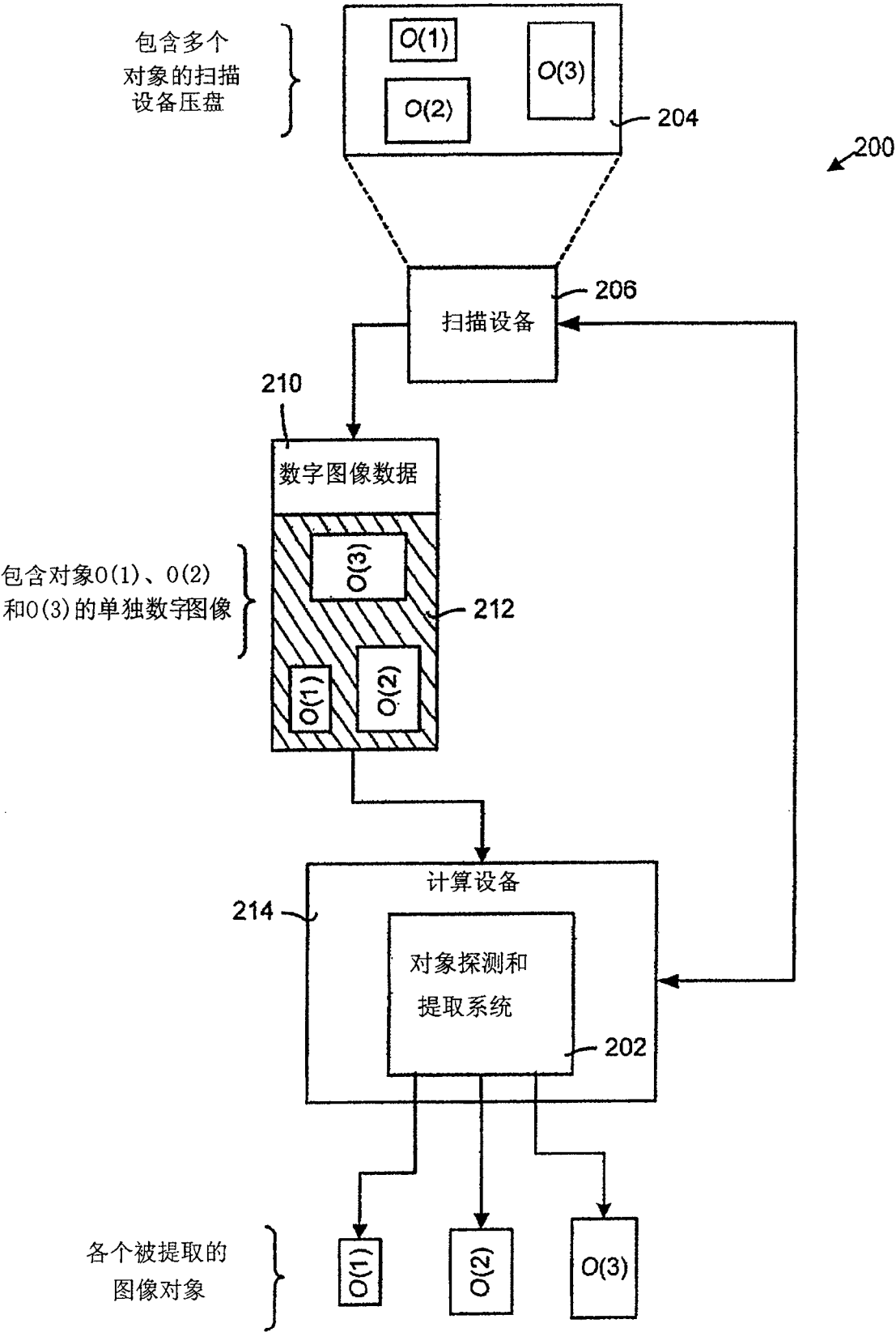


图 2

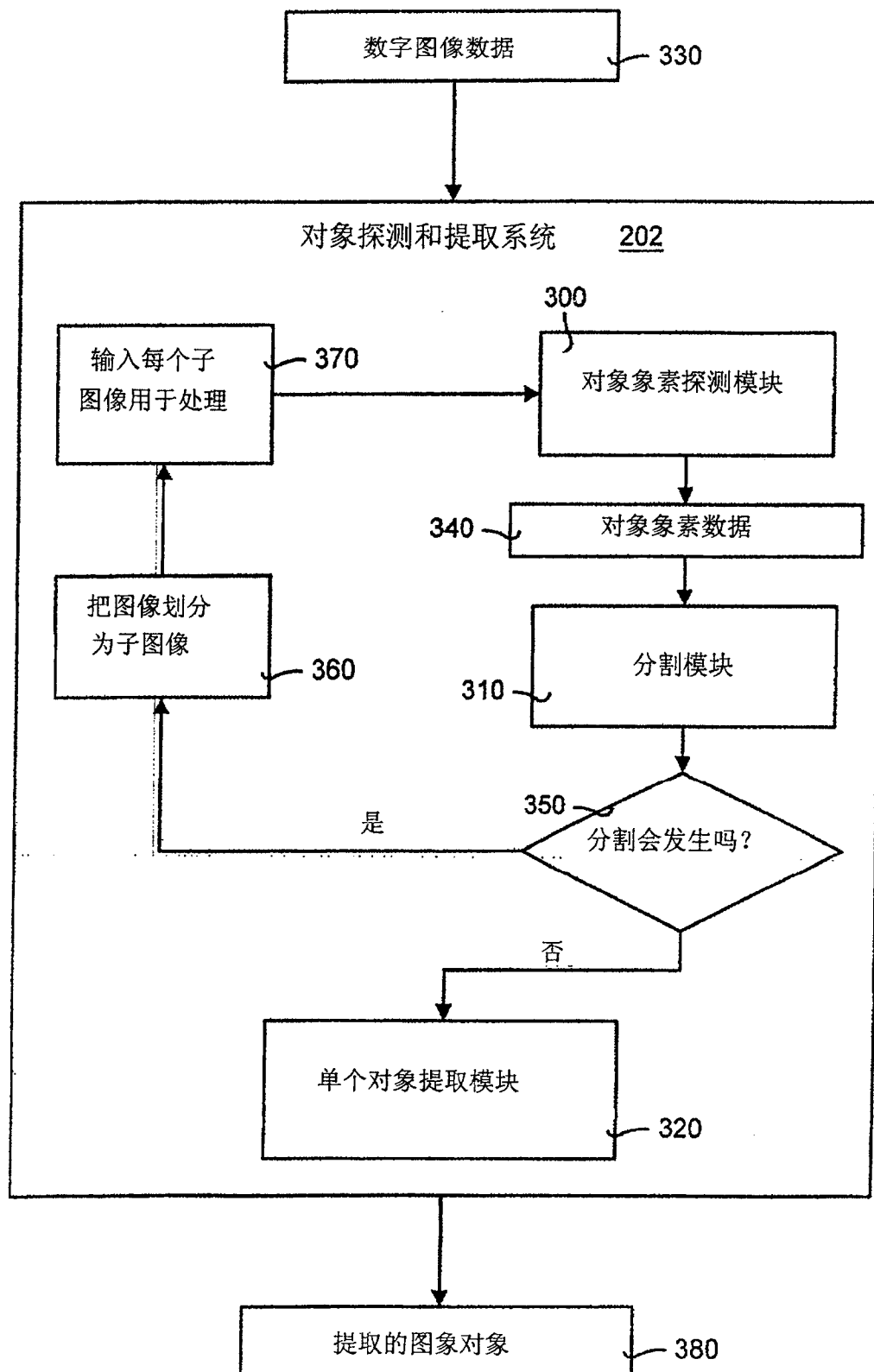


图 3

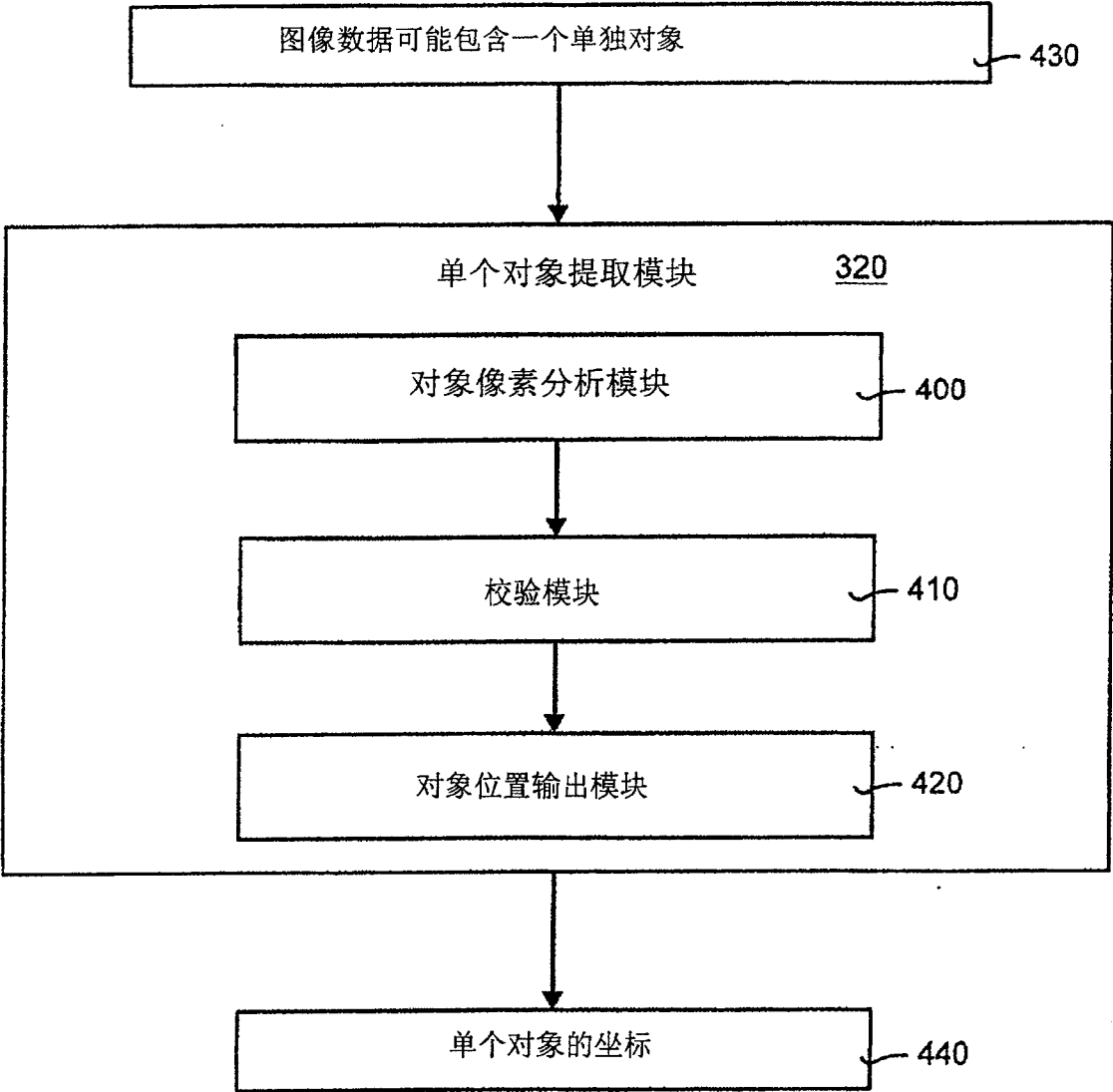


图 4

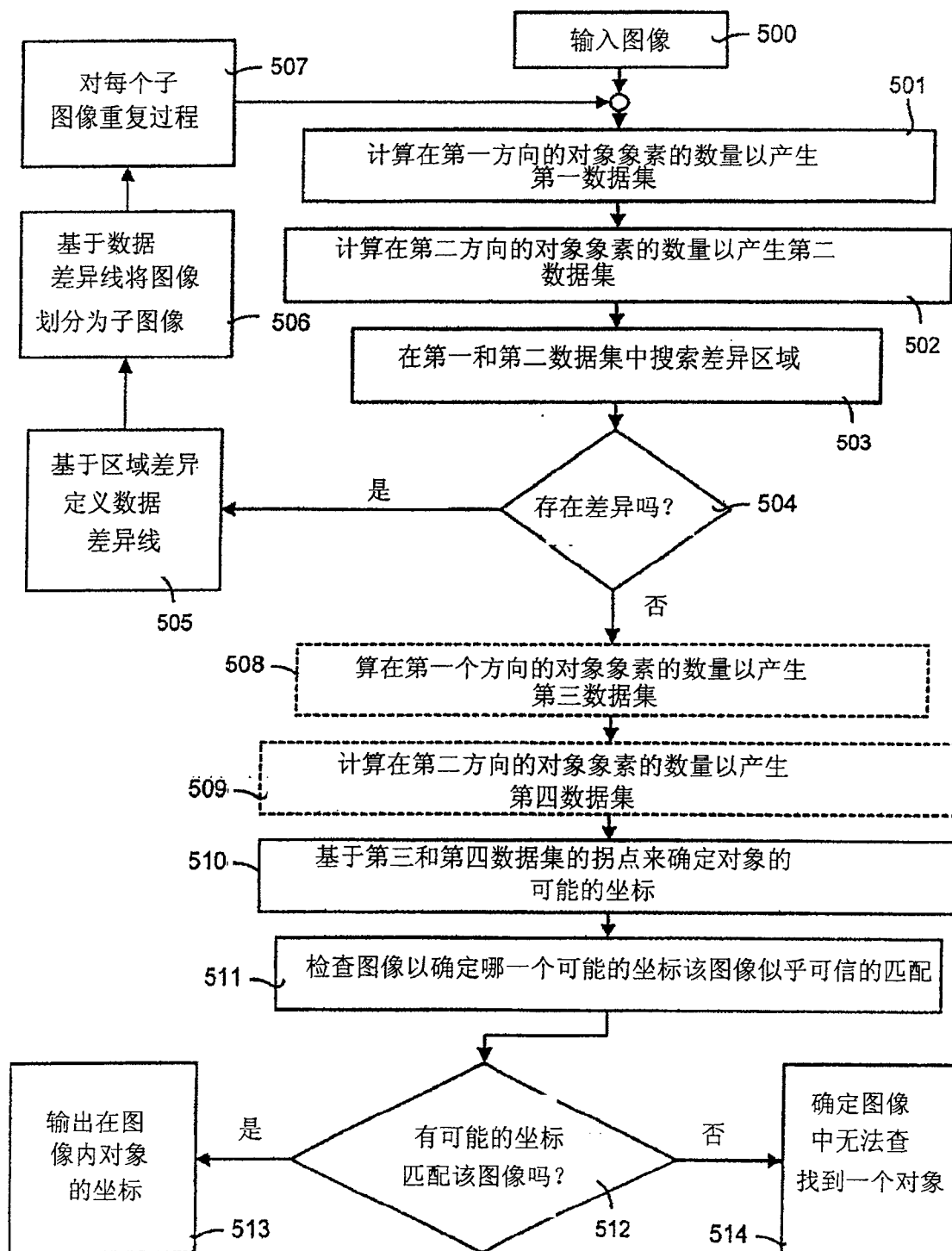


图 5

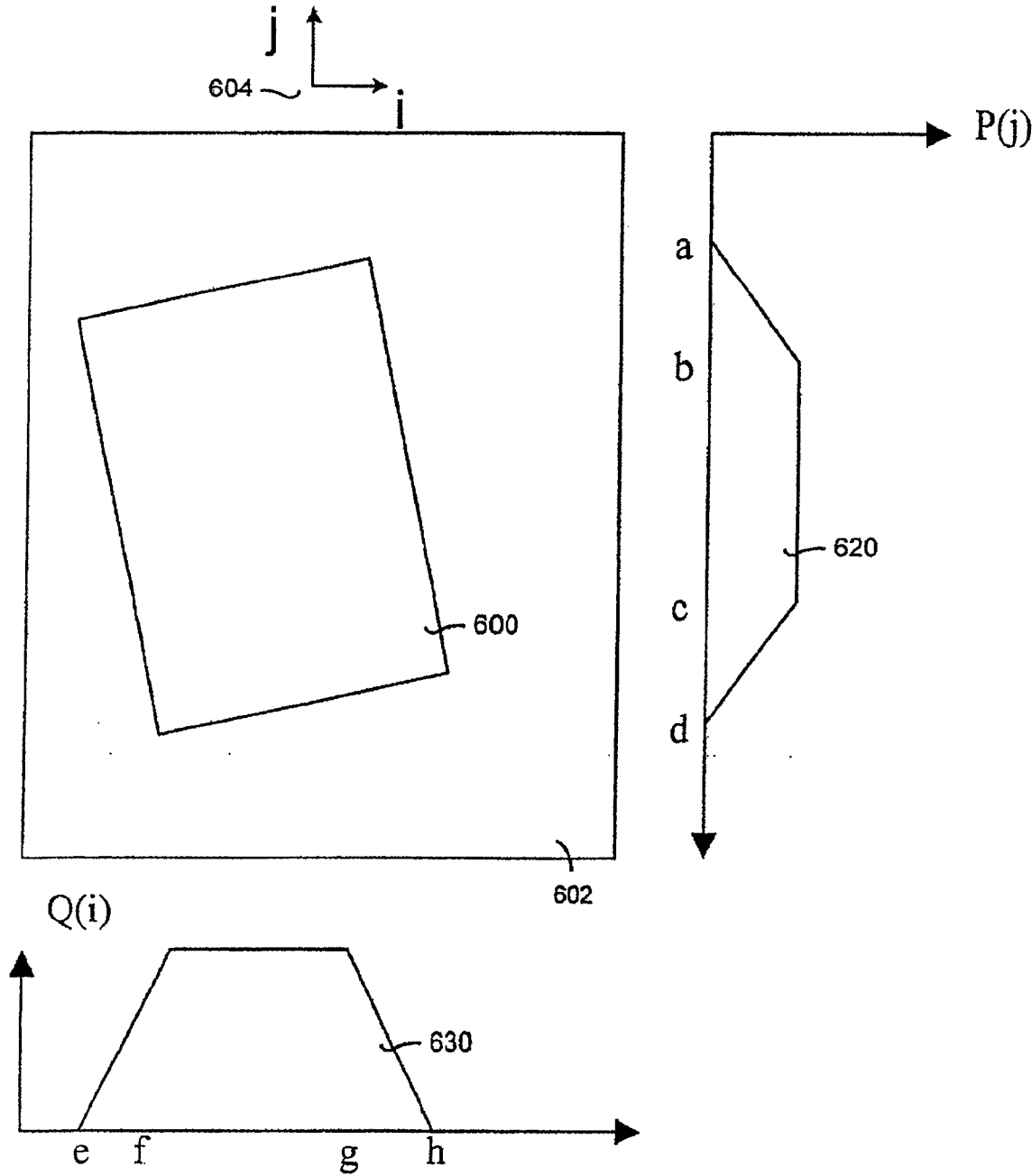


图 6

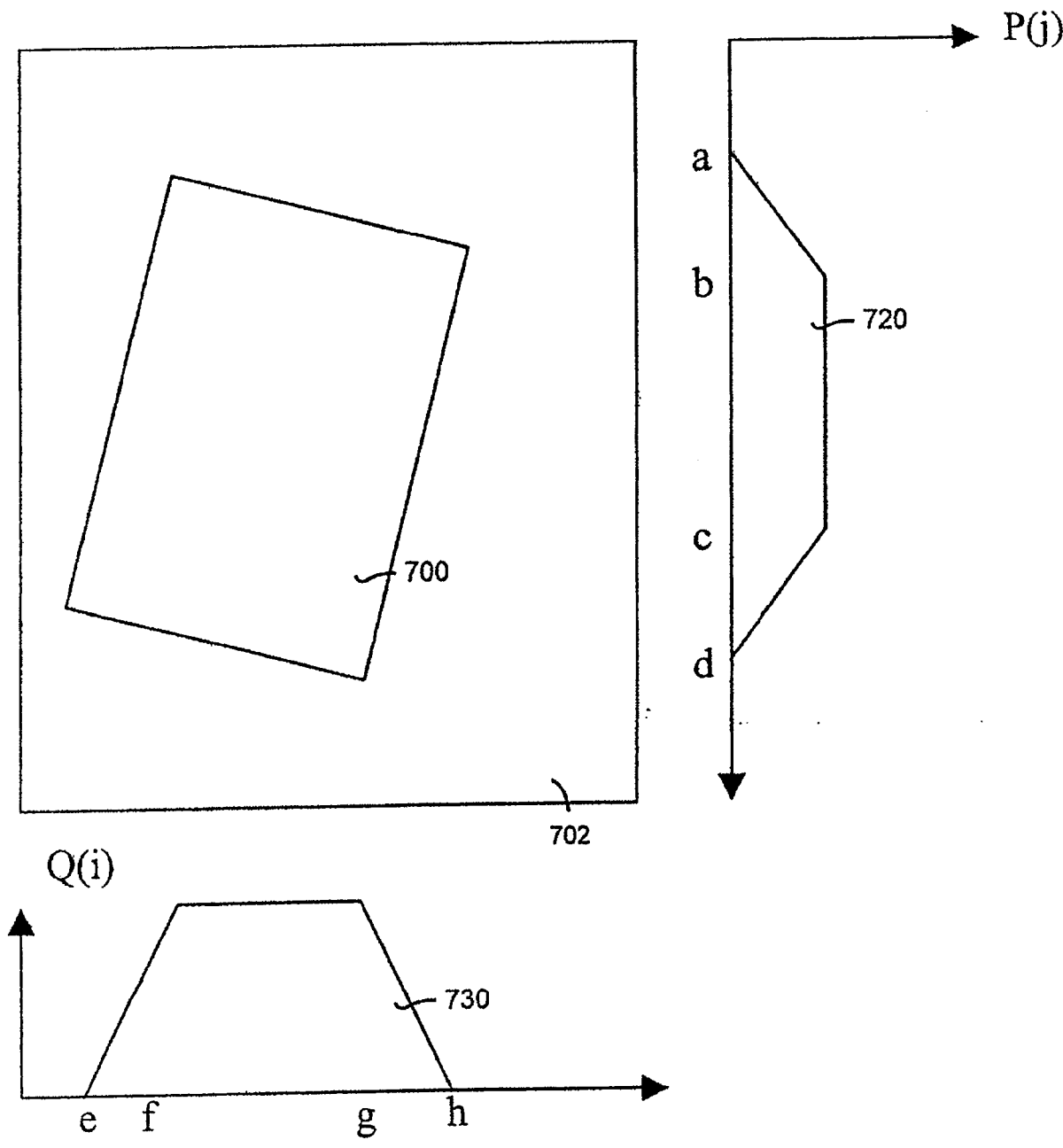


图 7

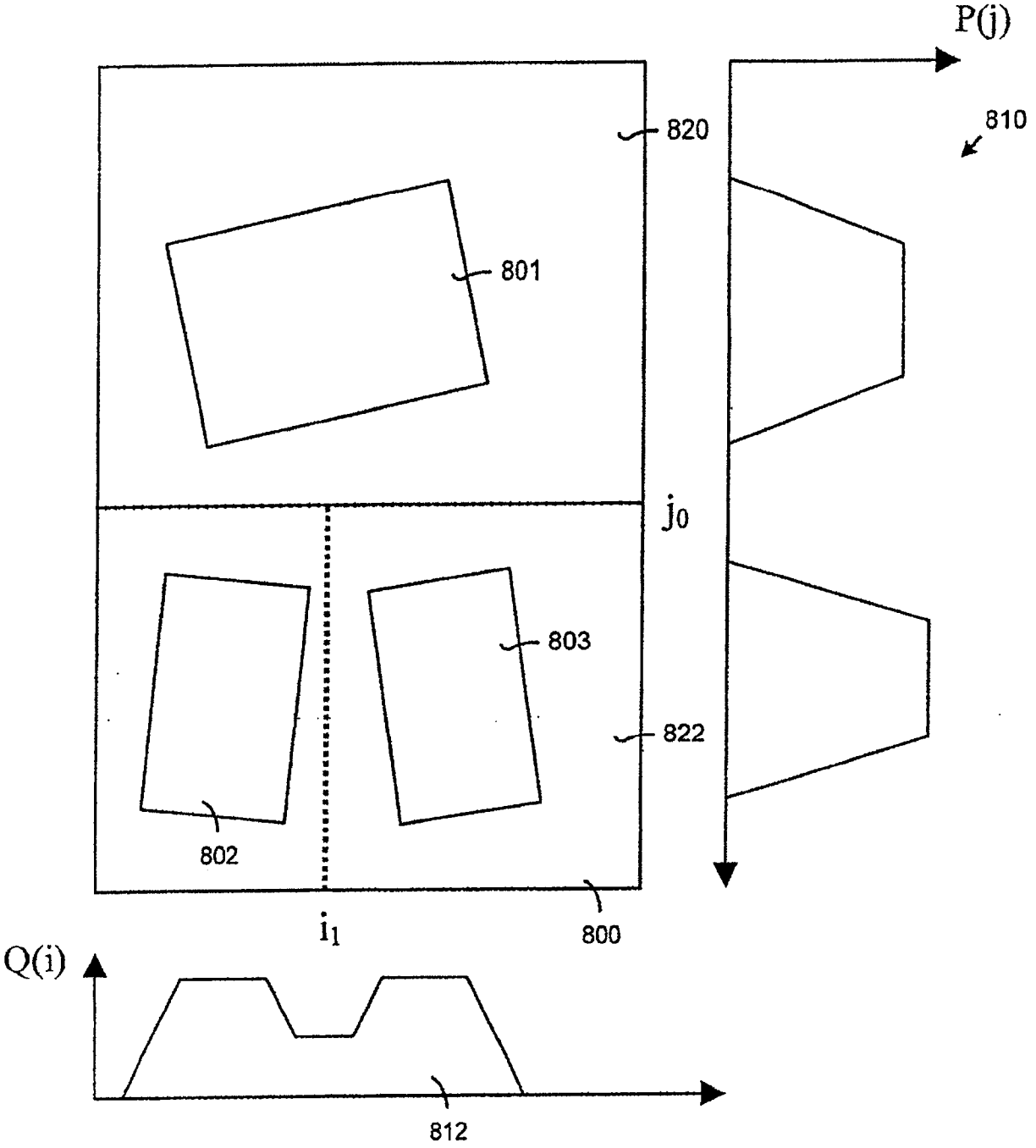


图 8

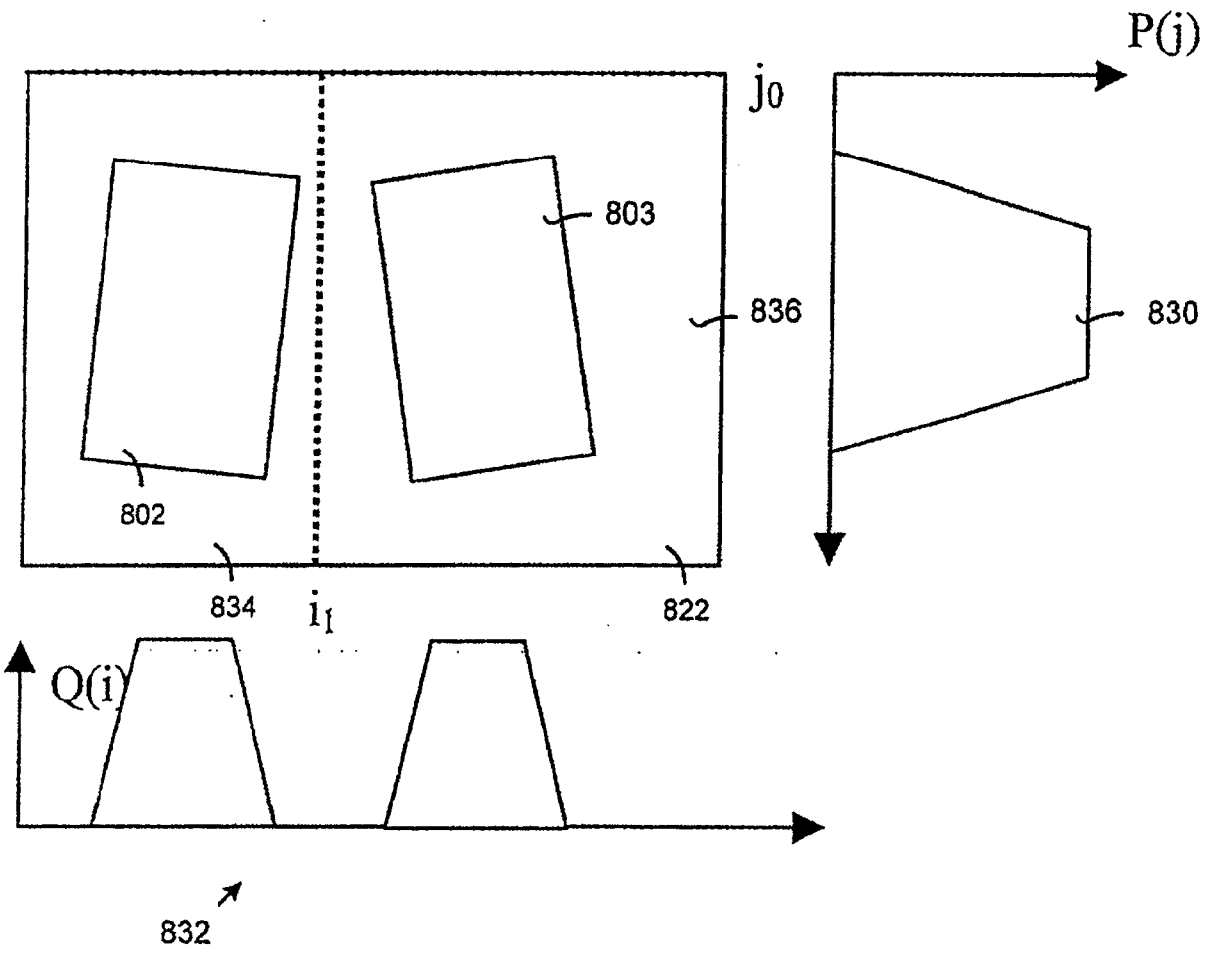


图 9

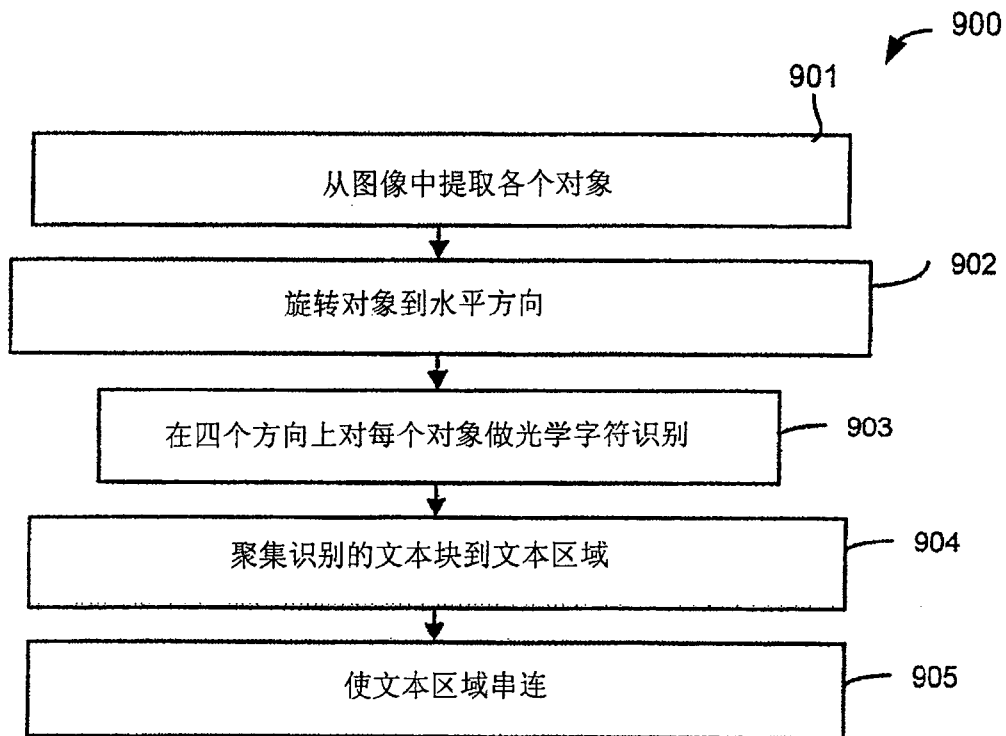


图 10

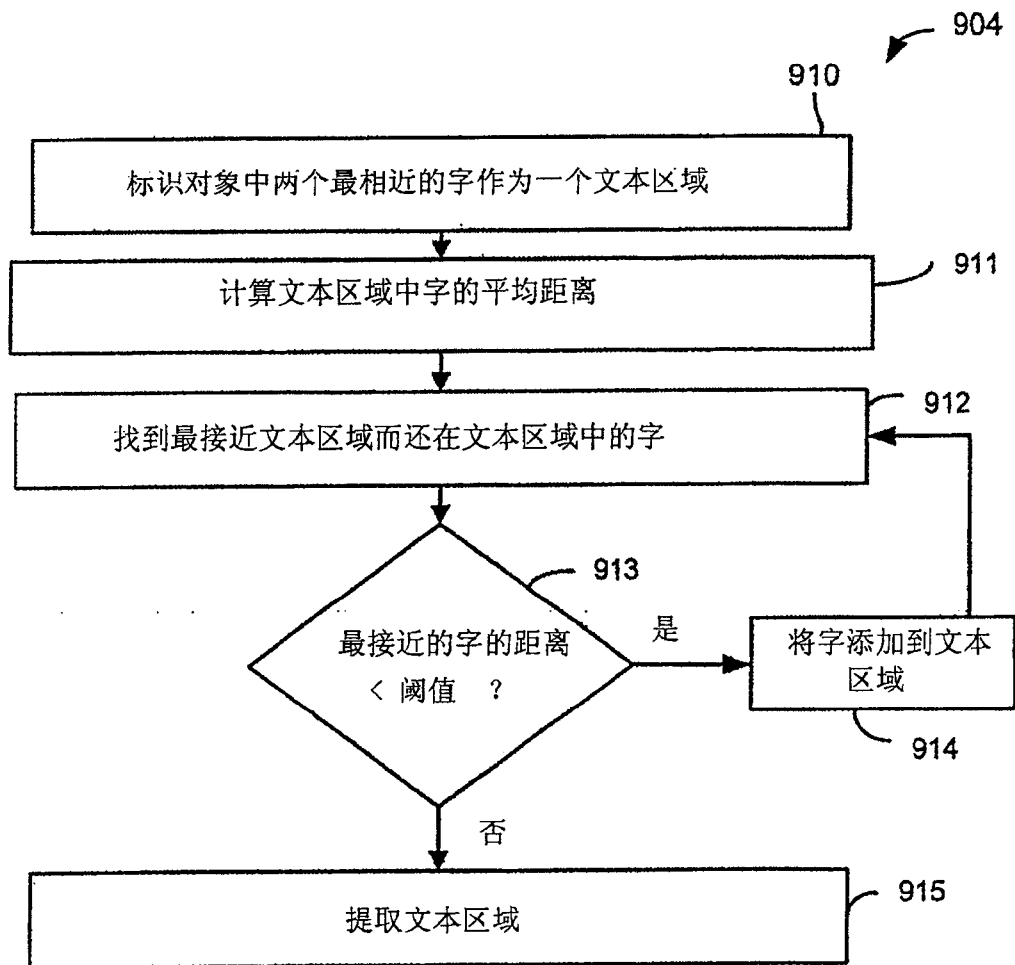


图 11

Form Filler

Personal Information:

Last Name: First Name:

Company: File as:

Business Information:

Business: Home:

Fax: Mobile:

Address:

Street:

City: State: Zip:

Country:

Contact Information:

Email: Web Addr:

Preview Area (1006):

John Smith

jsmith@abodefg.name

jsmith@xyz.name

Work: 1+(123) 987-6543

One Anystreet

Mycty ZZ 12345

USA

Form Controls:

Auto Fill Mode ☒ Precise Patterns ☒ Use Colors ☒

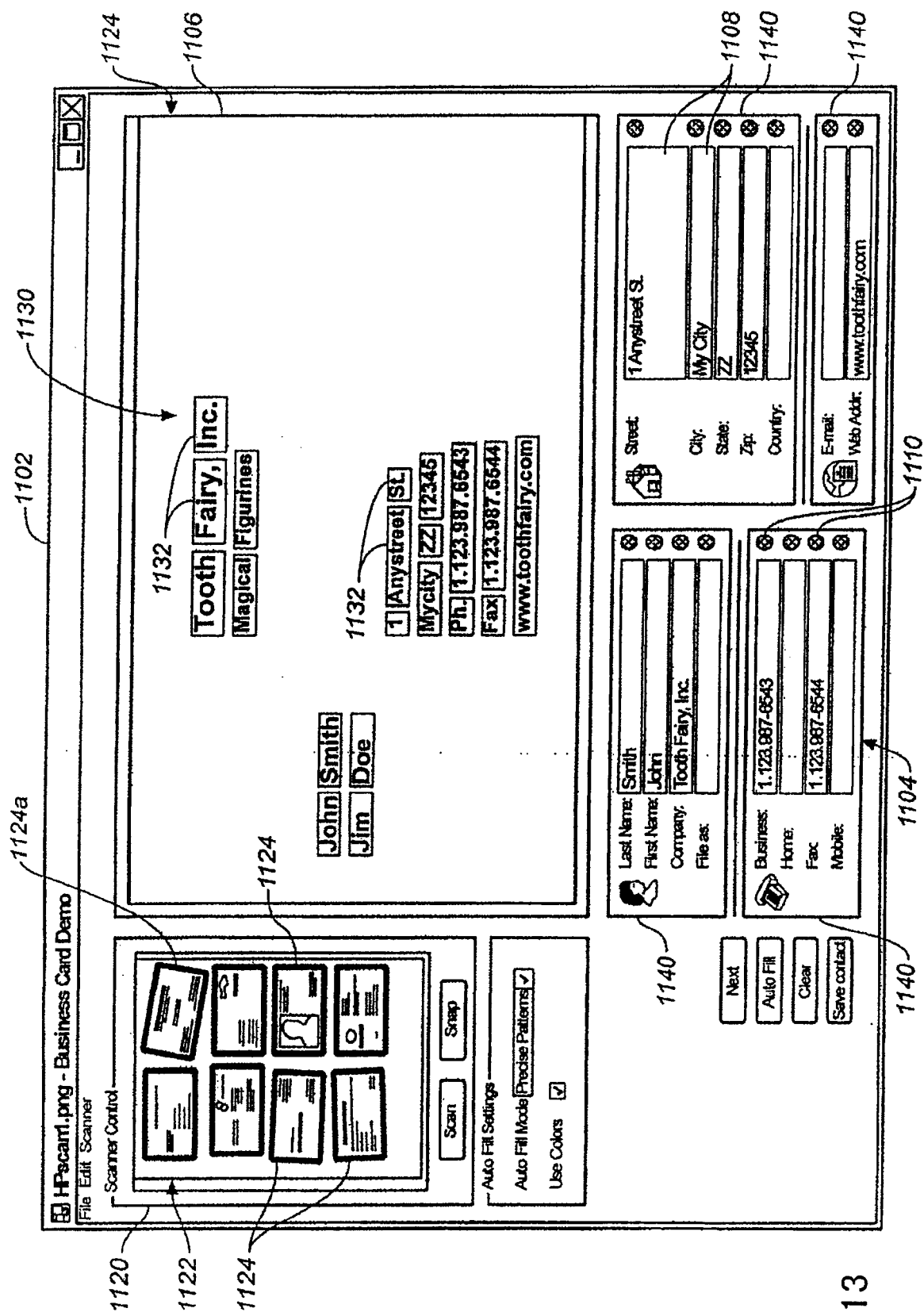


图 13

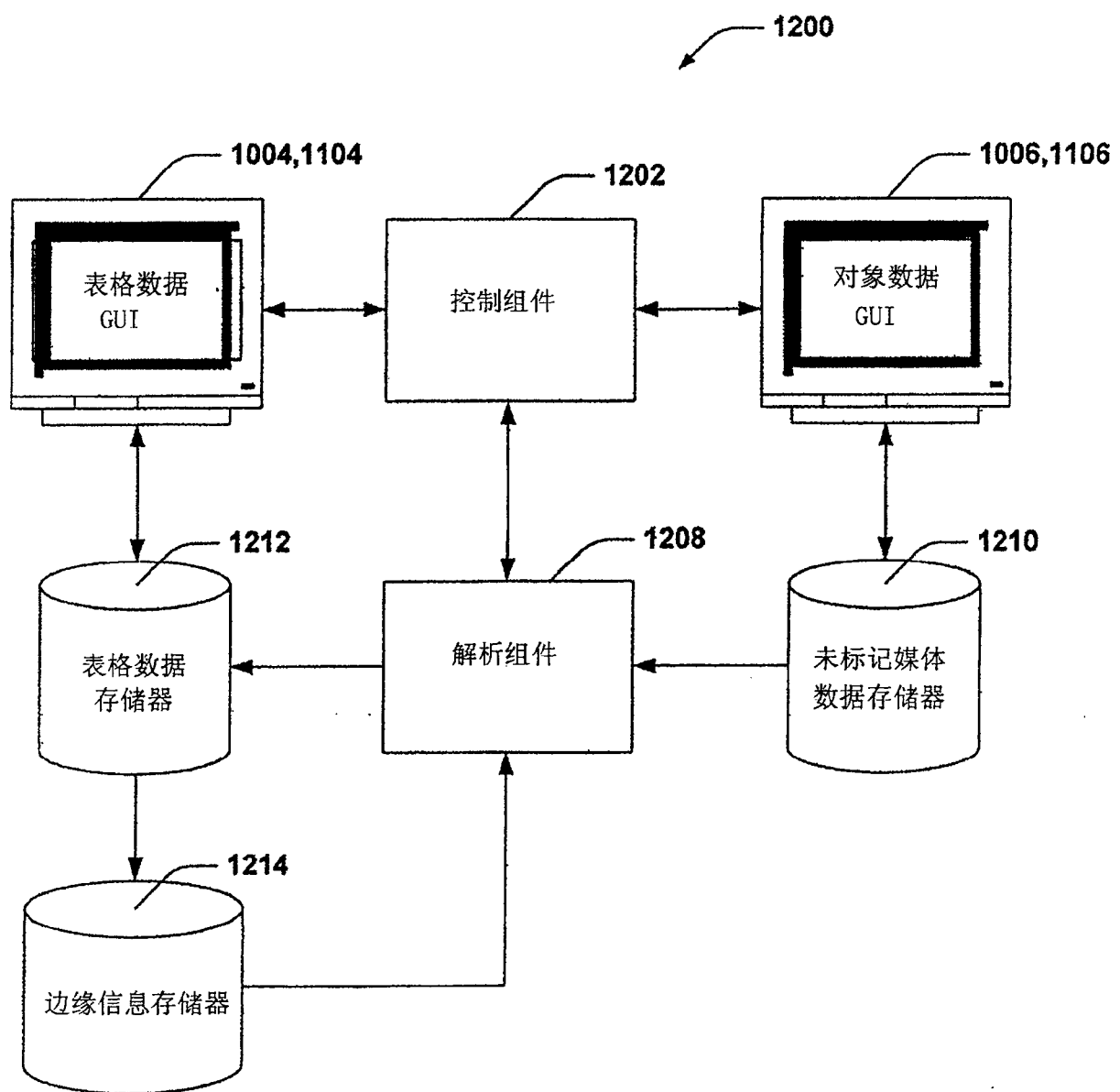


图 14

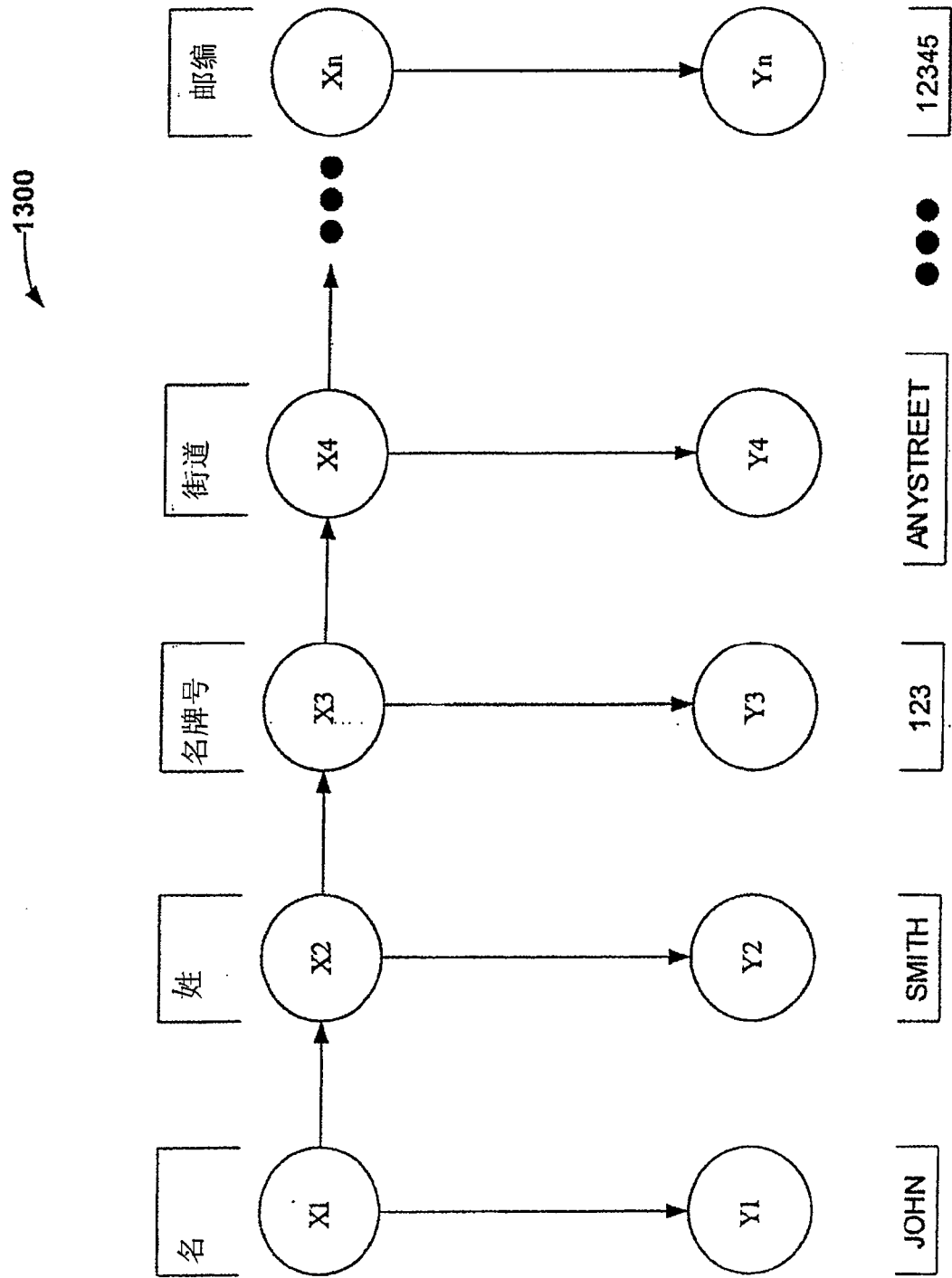


图 15

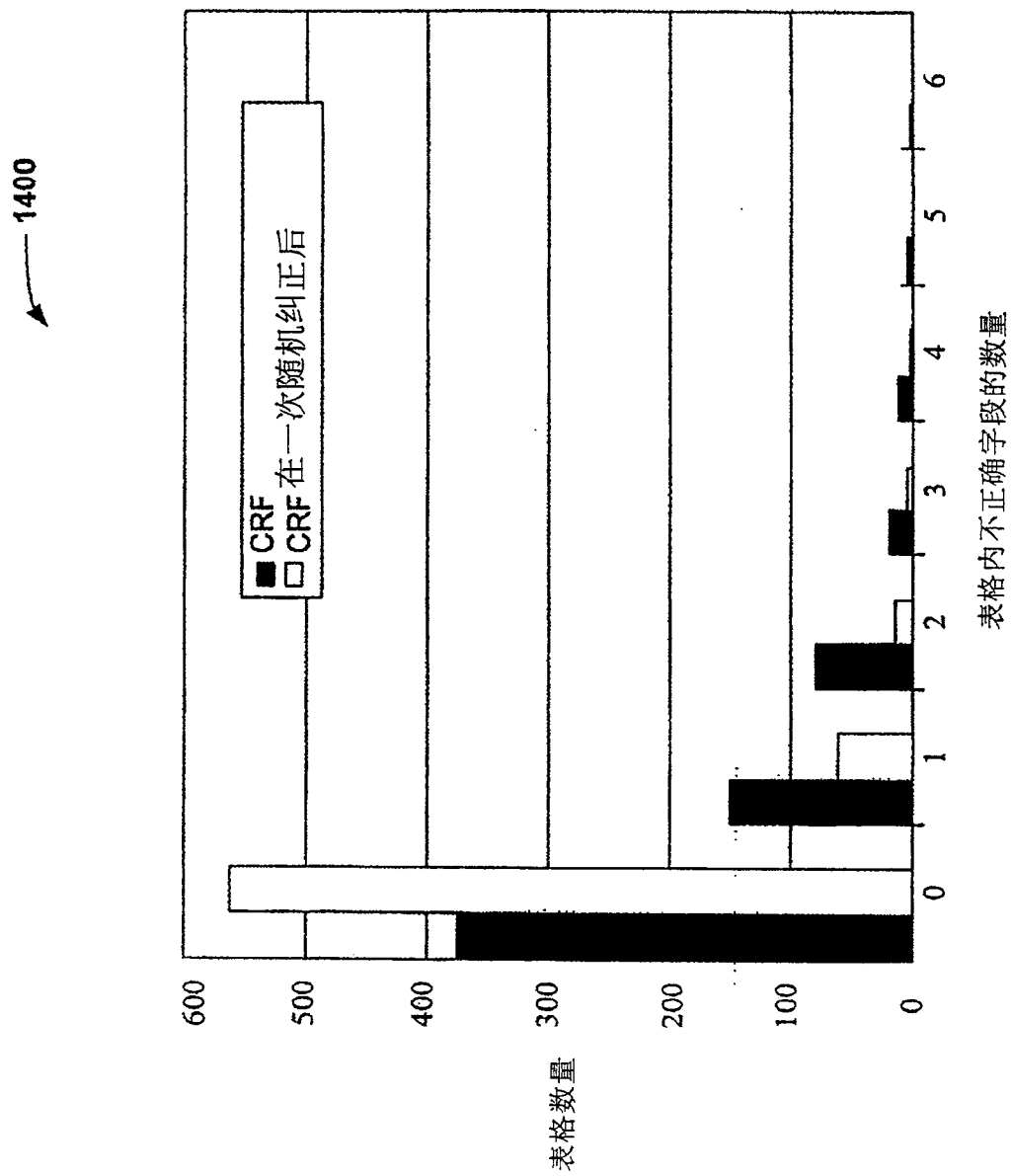


图 16

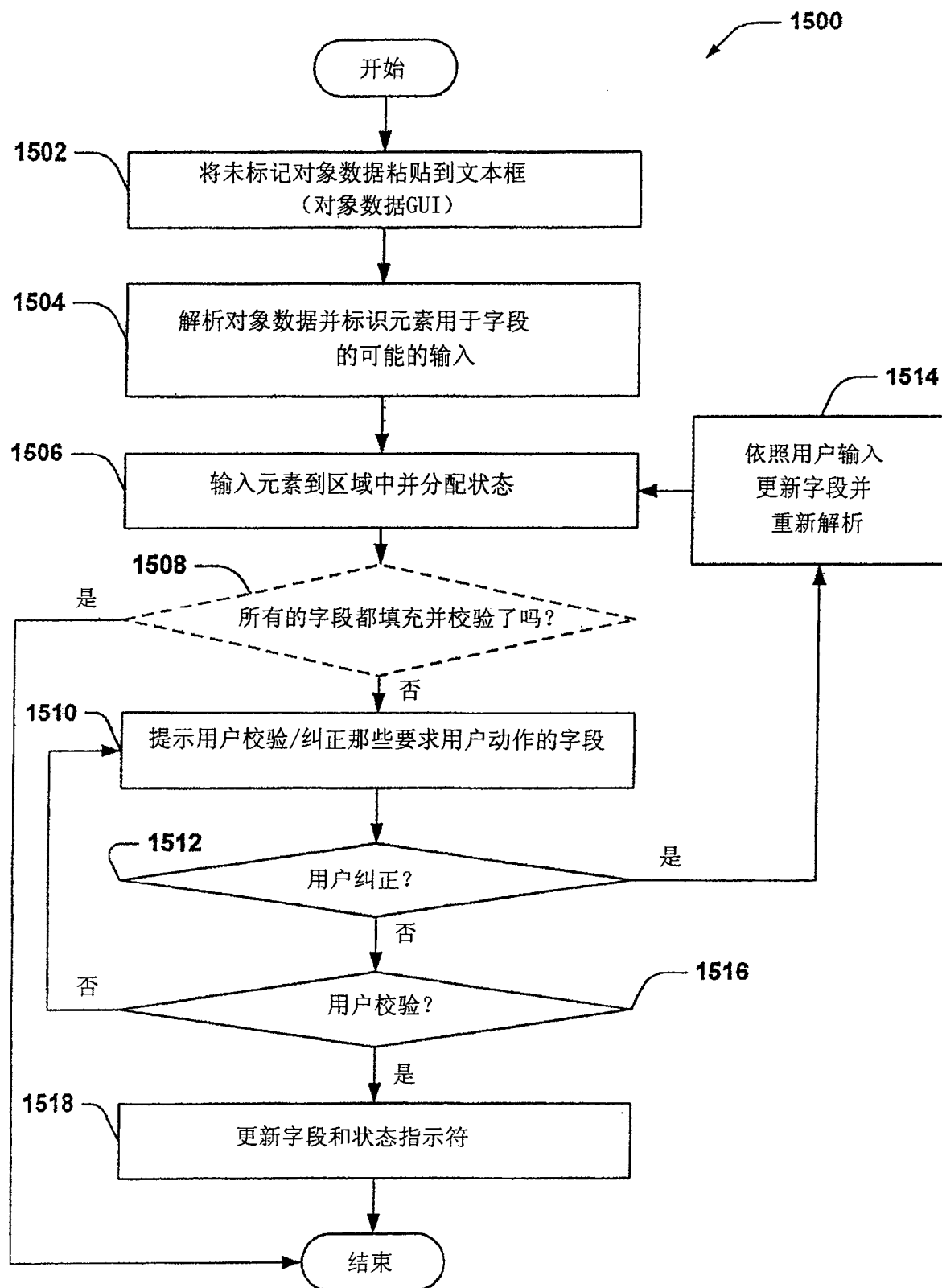


图 17

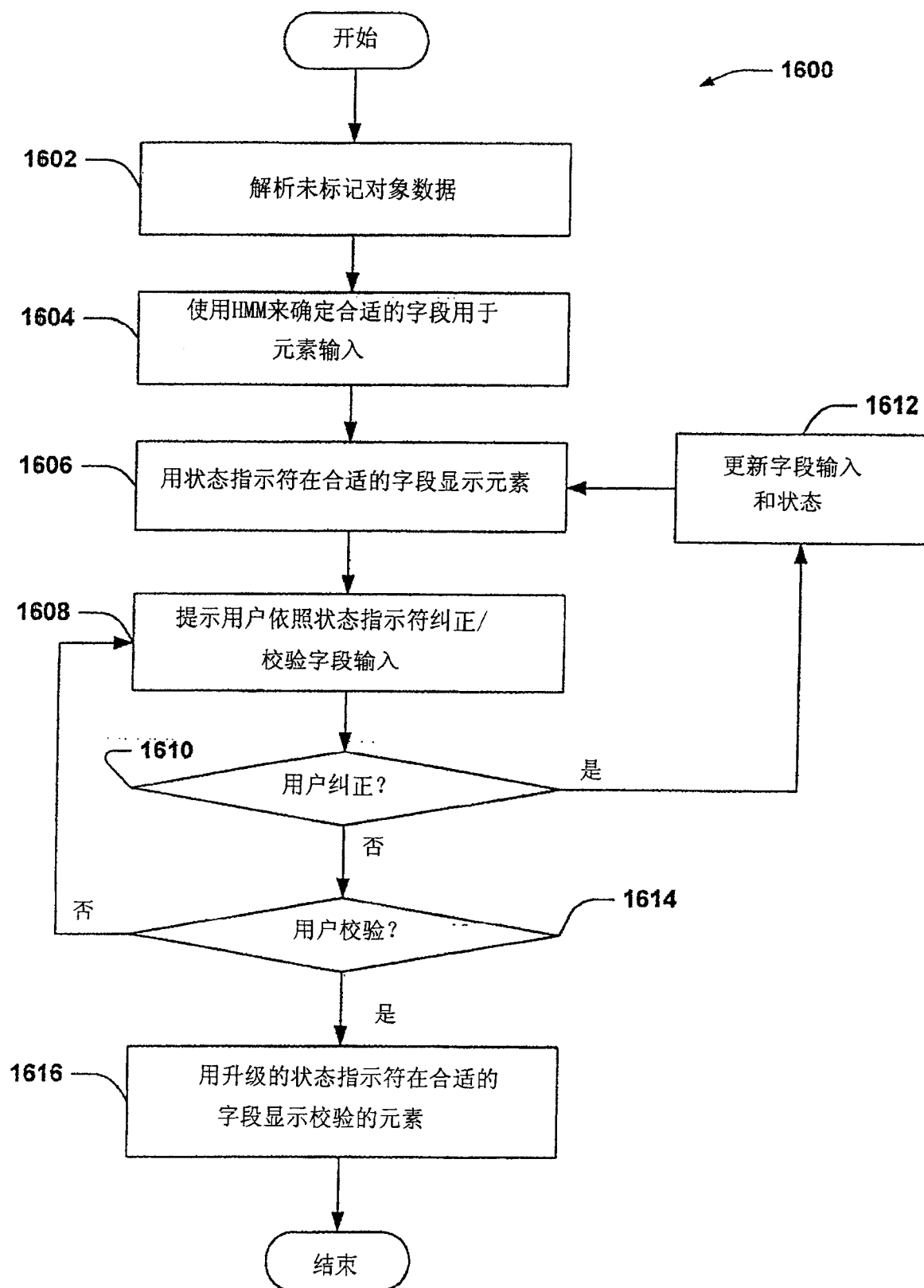


图 18

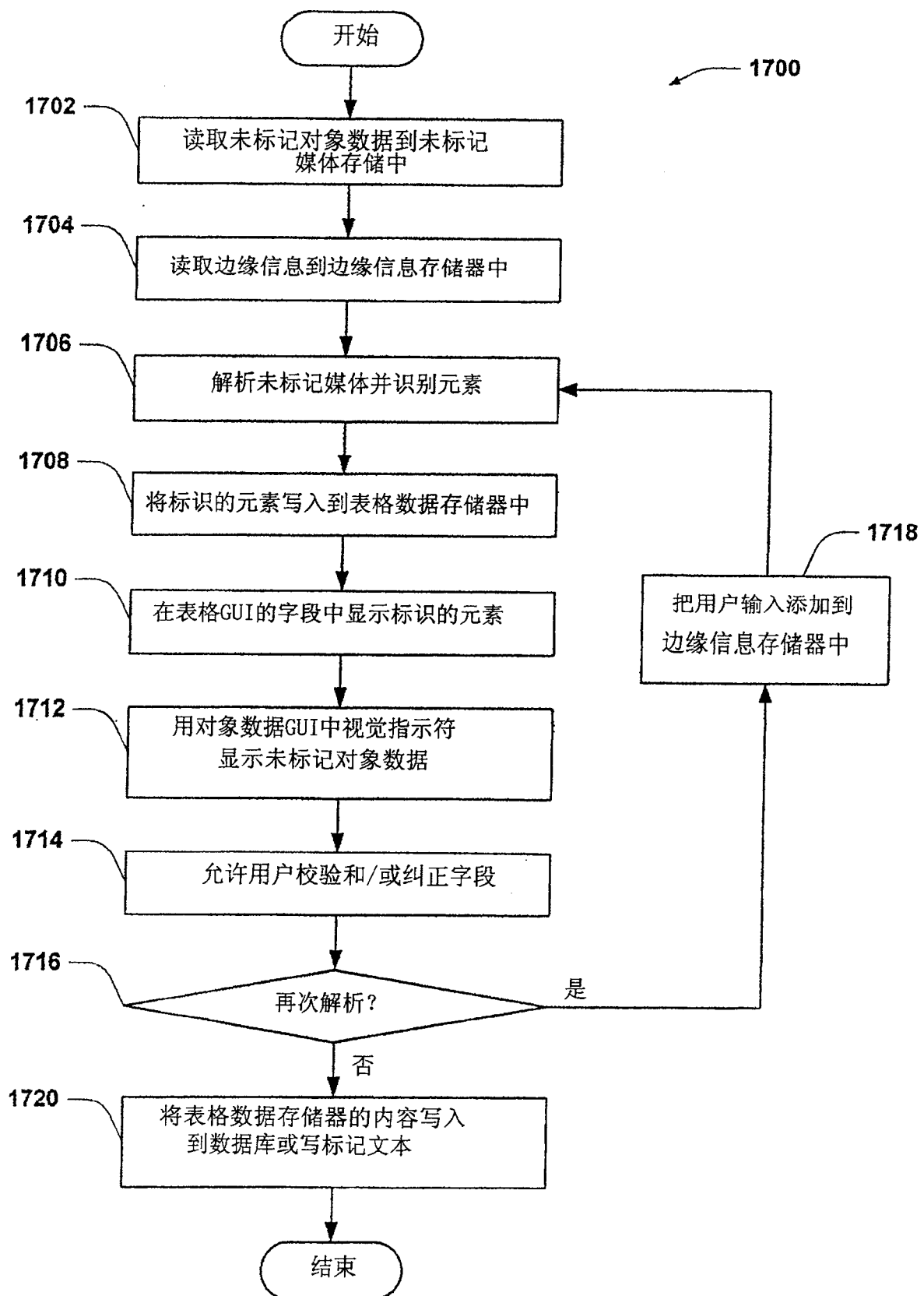


图 19