



(12)发明专利申请

(10)申请公布号 CN 105765551 A

(43)申请公布日 2016.07.13

(21)申请号 201480053143.7

克里斯多夫·W·斯拉舍

(22)申请日 2014.09.23

简·阿姆特鲁普

(30)优先权数据

61/883,865 2013.09.27 US

14/491,901 2014.09.19 US

(74)专利代理机构 北京市磐华律师事务所

11336

代理人 高伟 王睿

(85)PCT国际申请进入国家阶段日

2016.03.25

(51)Int.Cl.

G06F 15/00(2006.01)

(86)PCT国际申请的申请数据

PCT/US2014/057065 2014.09.23

(87)PCT国际申请的公布数据

W02015/048045 EN 2015.04.02

(71)申请人 柯法克斯公司

地址 美国加利福尼亚州

(72)发明人 安东尼·马乔拉 马继勇

亚历山大·舒斯特洛维奇

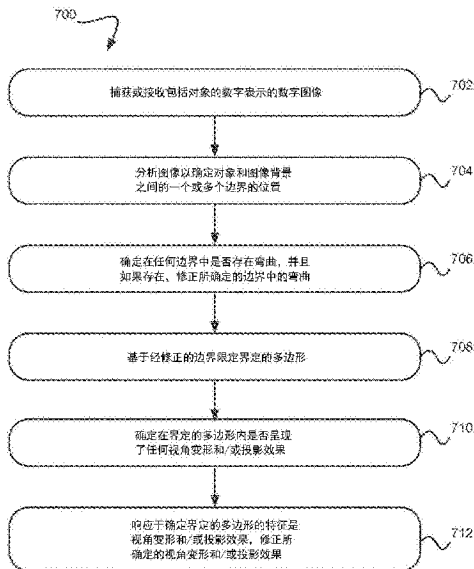
权利要求书2页 说明书30页 附图9页

(54)发明名称

用于三维几何重建所捕获的图像数据的系统和方法

(57)摘要

在多个实施例中,公开了用于处理由移动设备捕获的数字图像的方法、系统和计算机程序产品。大量特征使能和/或帮助利用移动设备处理这种数字图像,否则利用移动设备处理技术上将是不可能或不实际的,此外并且解决了用摄像头而不是传统平台式扫描仪、送纸式扫描仪或多功能外围设备所捕获的图像呈现的独一无二的挑战。尤其是,现在公开的系统和技术使能利用移动设备的摄像头捕获的图像中描绘的对象的三维重建。该重建修正或补偿了由基于摄像头的捕获所引起的视角变形。



1. 一种重建数字图像的方法,所述方法包括:  
接收包括用四角形界定的对象的数字表示的所述数字图像;  
修正所述四角形中的弯曲以形成四边形;以及  
修正所述四边形中的投影效果以形成矩形。
2. 如权利要求1所述的方法,其中所述四边形是平行四边形。
3. 如权利要求1所述的方法,其中修正所述弯曲包括至少部分基于固有四角形坐标转换所述四角形的一个或多个部分。
4. 如权利要求1所述的方法,其中修正所述投影效果包括利用四点算法转换所述四角形。
5. 如权利要求1所述的方法,其中所述数字图像的特征是偏离关于所述对象的法线大约30度或更大的捕获角。
6. 如权利要求1所述的方法,其中修正所述投影效果包括奇异值分解(SVD)。
7. 如权利要求1所述的方法,其中修正所述投影效果包括直接线性变换(DLT)。
8. 如权利要求1所述的方法,其中修正所述投影效果包括估计单应性变换H。
9. 如权利要求8所述的方法,其中估计H包括检测具有 $p_i = (x_i, y_i, 1)^T$ 的一个或多个点对应 $p_i \leftrightarrow P_i'$ 。
10. 如权利要求9所述的方法,其中估计H包括检测四个点对应,每个点对应 $p_i \leftrightarrow P_i'$ 与所述四角形的角位置 $p_i$ 和所述矩形的各自的角位置 $P_i'$ 相对应。
11. 如权利要求8所述的方法,其中H满足表达式 $\lambda p_i = H P_i'$ ,其中 $\lambda$ 等于三维参考坐标位置(X,Y,Z)和捕获设备的相应的参考坐标位置之间的焦距;  
 $p_i = (x_i, y_i, 1)^T$ ;  
 $P_i' = (X_i, Y_i, 1)^T$ ,并且  
 $i$ 是相应的图像中点 $p$ 或 $P'$ 各自的索引特征。
12. 如权利要求1所述的方法,其中修正所述投影效果包括实施非线性细化。
13. 如权利要求1所述的方法,进一步包括输出经重建的所述数字图像,其中经重建的所述数字图像的特征为相对于用垂直于所述对象的捕获角捕获的所述对象的扫描的500-DPI图像,像素位置错误大约5像素或更少。
14. 如权利要求1所述的方法,其中所述四角形的至少一条边的特征为二级多项式或更高级的多项式。
15. 如权利要求14所述的方法,其中所述四角形的每条边的特征为二级多项式或更高级的多项式。
16. 如权利要求1所述的方法,其中修正所述投影效果包括标准化所述数字图像和经重建的图像之间的点对应的坐标。
17. 如权利要求1所述的方法,其中所述对象是特征为已知高宽比的文件。
18. 一种系统,包括:  
处理器,其配置为执行逻辑;以及  
配置为接收包括用四角形界定的对象的数字表示的所述数字图像的逻辑;  
配置为修正所述四角形中的弯曲以形成四边形的逻辑;以及

配置为修正所述四边形中的投影效果以形成矩形的逻辑。

19.一种包括计算机可读存储介质的计算机程序产品,所述计算机可读存储介质具有存储于其上的计算机可读程序代码,所述计算机可读程序代码包括:

配置为接收包括用四角形界定的对象的数字表示的所述数字图像的计算机可读程序代码;

配置为修正所述四角形中的弯曲以形成四边形的计算机可读程序代码;以及

配置为修正所述四边形中的投影效果以形成矩形的计算机可读程序代码。

## 用于三维几何重建所捕获的图像数据的系统和方法

### 技术领域

[0001] 本申请涉及数字图像数据捕获和处理,并且尤其是涉及修正利用摄像头捕获图像所固有的由失真效果引起的图像伪影。

### 背景技术

[0002] 在文件诸如信件、支票、账单、发票等中描绘的数字图像通常利用耦连至计算机工作站诸如膝上型或台式计算机的扫描仪或多功能外围设备而捕获和处理。能够实施这种捕获和处理的方法和系统在本领域是公知的并且很好地适于使用它们的任务。

[0003] 然而,在日常的活动、计算和商业越来越多地利用移动设备来实施的时代,提供类似的文件捕获和处理系统和方法用于在移动平台诸如智能手机、数码相机、平板电脑等上部署和使用是非常有益的。

[0004] 转变传统文件捕获和处理技术中的主要挑战是有限的处理功率和利用移动设备目前可用的硬件能够达到的图像分辨率。这些限制产生了巨大的挑战,因为处理在典型地远比用传统扫描仪能够达到的分辨率低的分辨率下所捕获的图像是不可能的或不切实际的。结果是,基于传统扫描仪的处理算法典型地在利用移动设备捕获的数字图像上实施很差。

[0005] 此外,移动设备上有限的处理和可用的内存使得扫描仪使用的传统图像处理算法在计算成本方面过分昂贵。尝试处理基于传统扫描仪的图像处理算法耗用了太多时间在实际应用在现代移动平台上。

[0006] 进一步的挑战由移动捕获部件(例如,手机、平板电脑等上的摄像头)的性质产生。传统扫描仪能够将物理文件如实地表现在数字图像中,能够将物理文件的纵横比、尺寸和形状严谨地保持在数字图像中,在这些方面移动捕获部件常常不能够生成这样的结果。

[0007] 尤其是,由摄像头所捕获的文件的图像出现了一系列当处理用扫描仪所捕获的图像时所没遇到过的新问题。这一部分是由于获取文件图像的不同方法所固有的差异,以及制造设备的方法所固有的差异。一些扫描仪工作的方法是利用传输装置,其创建了纸张和传感器的线性阵列之间的相对运动。当运动时,这些传感器创建了文件的像素值,并且这些所捕获的像素值的序列形成了图像。因此,通常都具有传感器本身中的噪声决定的水平或垂直的一致性,并且是相同的传感器提供了行间的所有像素。

[0008] 相反,摄像头有很多在非线性阵列中的传感器,例如,典型地布置在矩形中。因此,所有这些个体传感器是独立的,并且渲染的图像数据并不典型地具有水平或垂直一致性。此外,摄像头引入了投影效果,其是拍照片所处的角度的函数。例如,采用如扫描仪中的线性阵列,即使纸张的传输并不是完美垂直于传感器的基准线并且引入了一些偏移,也没有如摄像头中的投影效果。此外,采用摄像头捕获,由于摄像头的光学元件而可能引入非线性失真。

[0009] 由于上面呈现的挑战,提供图像捕获和处理算法及其应用,其补偿和/或修正关联于利用移动设备的图像捕获和处理的问题,而经由有效的处理方法维持低计算成本,将是

有益的。

### 发明内容

[0010] 在一个实施例中,重建数字图像的方法包括:接收包括用四角形界定的对象的数字表示的所述数字图像;修正所述四角形中的弯曲以形成四边形;以及修正所述四边形中的投影效果以形成矩形。

[0011] 一种系统,包括:处理器,其配置为执行逻辑;以及配置为接收包括用四角形界定的对象的数字表示的所述数字图像的逻辑;配置为修正所述四角形中的弯曲以形成四边形的逻辑;以及配置为修正所述四边形中的投影效果以形成矩形的逻辑。

[0012] 一种包括计算机可读存储介质的计算机程序产品,所述计算机可读存储介质具有存储于其上的计算机可读程序代码。所述计算机可读程序代码包括:配置为接收包括用四角形界定的对象的数字表示的所述数字图像的计算机可读程序代码;配置为修正所述四角形中的弯曲以形成四边形的计算机可读程序代码;以及配置为修正所述四边形中的投影效果以形成矩形的计算机可读程序代码。

### 附图说明

[0013] 图1例示了根据一个实施例的网络体系架构。

[0014] 图2示出了根据一个实施例的有代表性的硬件环境,其可以与图1中的服务器和/或客户端相关联。

[0015] 图3A是根据一个实施例的数字图像的示意性表示,所述数字图像包括文件的数字表示。

[0016] 图3B是根据一个实施例的数字图像和多个页面检测分析窗口的示意性表示,所述数字图像包括文件的数字表示。

[0017] 图3C是根据一个实施例的数字图像的示意性表示,所述数字图像包括以多个候选边缘点所表征的文件的数字表示。

[0018] 图3D是根据一个实施例的大分析窗口,其包括数字图像的多个像素,和大分析窗口内的小分析窗口的示意性表示。

[0019] 图4是根据一个实施例的数字图像的示意性表示,所述数字图像包括由目标四边形所界定的文件的数字表示。

[0020] 图5A是根据一个实施例的页面矩形化算法的第一循环的图形表示。

[0021] 图5B是根据一个实施例的到页面矩形化算法的输入的图形表示。

[0022] 图6A根据一个实施例简化示意性示出了用于测量捕获角度的坐标系统。

[0023] 图6B描绘了根据一个实施例的、使用对于对象正常的捕获角度所捕获的矩形对象的示例性示意图。

[0024] 图6C描绘了根据一个实施例的、使用关于对象稍微倾斜的捕获角度所捕获的矩形对象的示例性示意图。

[0025] 图6D描绘了根据一个实施例的、使用关于对象显著倾斜的捕获角度所捕获的矩形对象的示例性示意图。

[0026] 图7是根据一个实施例的方法的流程图。

## 具体实施方式

[0027] 进行下面的说明是为了例示本发明的总原则,并且并不意味着限制本文要求保护的发明性概念。进一步地,本文中描述的特殊特征可以以多种可能的组合和排列的每一种结合描述的其他特征一起使用。

[0028] 除非此文中进行了特别限定,否则所有术语将给出它们最广泛的可能理解,包括说明书隐含的意思以及本领域技术人员理解的意思和/或字典、论文等中限定的意思。

[0029] 还必须注意到,如在说明书和附属的权利要求书中所使用的,除非进行了限定,否则单数形式包括复数指代。

[0030] 本申请涉及图像处理。尤其是,本申请公开了配置为将数字图像中所描绘的对象从非矩形形状转换成大致矩形形状或优选地矩形形状的系统、方法和计算机程序产品。更加优选地,这通过利用两步处理来完成,其中对象中的弯曲部分首先被修正,接着图像中的任何投影效果被修正。

[0031] 根据多个实施例,下面的定义对理解本文中描述的发明性概念是有用的。下面的定义要认为是示例性的,并且为了示例以为本公开提供附加的清楚的目的而提供,但不应该认为是限制本文中公开的发明性概念的范围。

[0032] 此后所称的,“四边形”是四条边的图形,其中(1)每条边是直线形的,以及(2)相邻的边在相交处形成顶点。根据两个示例性的实施例,示例性的四边形在下面的图6C和6D中描绘。

[0033] “平行四边形”是一种特殊类型的四边形,即四条边的图形,其中(1)每条边是直线形的,(2)相对的边是平行的,以及(3)相邻的边不必垂直,这样相邻边相交处的顶点形成其值不必是 $90^\circ$ 的角。

[0034] “矩形”或“矩形的形状”是一种特殊类型的四边形,其限定为四条边的图形,其中(1)每条边是直线形的,(2)相对的边是平行的,以及(3)相邻的边是垂直的,这样在每对邻边之间的顶点处形成的内角是直角,即 $90^\circ$ 的角。根据一个示例性的实施例,图6B中描绘了示例性的矩形。

[0035] 而且,此文中所称的“矩形”和“矩形形状”被认为包括“大致的矩形形状”,其限定为四条边的形状,其中(1)每条边主要是直线形的(例如,在不同的实施例中,每条边的长度的至少90%、95%或99%的特点为一阶多项式(比如 $y=mx+b$ )),(2)每对邻边形成角度值为 $\theta$ 的内角,其中 $\theta$ 在或者(a)两条邻边之间的顶点处,(b)在一条边的主要直线形部分的投影和邻边之间的顶点处,或者(c)在一条边的主要直线形部分的投影和邻边的主要直线形部分的投影之间的顶点处近似为 $90^\circ$ (例如, $\theta$ 满足关系式: $85^\circ \leq \theta \leq 95^\circ$ )。在下面的图7中描绘了示例性的“大致的矩形形状”(注意,所描绘的驾驶证的角是弧形的,这样各自的邻边没有形成分立的顶点,但每对邻边的投影将在它们的交点处产生角度近似为 $90^\circ$ 的顶点)。

[0036] 本文中所涉及的“非矩形形状”包括不是上面限定的“矩形形状”或“大致的矩形形状”的任何形状。在优选的实施例中,“非矩形形状”是“四角形”,如在本文中所涉及的,其是四条边的图形,其中:(1)每条边全部或部分用从所选类别的函数中选择的方程式(例如,从优选地范围从零阶到五阶的多项式,更加优选地从一阶到三阶的多项式,并且尤其更加优选地从一阶到二阶的多项式的类别中选择)表征,并且(2)图形的邻边在它们的交点处形成

顶点。根据一个示例性的实施例,本文中涉及的示例性的四角形在图4中描绘。

[0037] 在当前公开的发明性概念的一个总的实施例中,重建数字图像的方法包括:接收数字图像,其包括由四角形界定的对象的数字表示;修正四角形中的弯曲,以形成四边形;以及修正四边形中的投影效果,以形成矩形。

[0038] 在另一个总的实施例中,系统包括处理器,配置为执行逻辑;以及配置为接收数字图像的的逻辑,所述数字图像包括由四角形界定的对象的数字表示;配置为修正四角形中的弯曲以形成四边形的逻辑;以及配置为修正四边形中的投影效果以形成矩形的逻辑。

[0039] 在仍然另一个总的实施例中,计算机程序产品包括计算机可读存储介质,其具有存储在其上的计算机可读程序代码。计算机可读程序代码包括配置为接收数字图像的代码,所述数字图像包括由四角形界定的对象的数字表示;配置为修正四角形中的弯曲以形成四边形的代码;以及配置为修正四边形中的投影效果以形成矩形的代码。

[0040] 矩形化的以前的方法,诸如在美国专利申请号13/740,127(于2013年1月11日提交)中描述的,很好地修正了在摄像头捕获的图像中观察到的弯曲效果。然而,当摄像头的倾斜和/或转动足够大(例如,大约30度或更多)时,由于图像中存在的严重的投影效果/伪影,该技术修正投影效果的能力常常是不足的。

[0041] 将以前描述的弯曲修正的方法和现在描述的投影效果的修正方法组合到单个双重目的步骤有效地组合了每种方法各自的优点。在一个实施例中,所述方法可以组合如下。首先,弯曲修正部件通过将界定对象的弯曲四角形映射成具有相同角但具有直边(即特点为直线或一阶多项式)的四角形。第二,利用诸如下面描述的4-点方法将这个(现在为直边的)四角形映射成目标矩形。

[0042] 作为明显的优点,尽管重建矩形表示更加准确,但这个双步骤仅影响坐标的两步映射,而图像的实际转换仅发生一次。因为多数处理时间都花在在存储器中操控两个大图像以及组合非整数(x,y)坐标对周围的四个像素,而不是通过任何方法计算正确的坐标,所以相比由传统的台式扫描仪或类似的设备所获得的相应图像,尽管重建的图像中在相对像素位置方面准确度提高了,但关于计算成本和由此的运行时间,该双方法仅比原始方法多花费大约3%。

[0043] 实验测试证实,在具有明显的弯曲和大的投影失真二者的图片上,双重方法相对于单独的基于坐标的方法的错误减小了大约3倍的坐标错误(测量为从矩形化的像素到在相同分辨率下相同文件的经扫描图像中的相同像素的最大距离)。在一个实施例中,剩余的误差在500-DPI分辨率下大约5像素,或者大约一英寸的百分之一。

[0044] 图像(例如,图片、图形、图形示意图、电影的单帧、视频、影片、剪辑等)优选地是用摄像头捕获的数字图像,尤其是移动设备的摄像头。如本文中理解的,移动设备是能够接收数据而不具有经由物理连接(例如,金属线、芯线,电缆等)供应功率并且能够接收数据而没有物理数据连接(例如,金属线、芯线,电缆等)的任何设备。本公开范围内的移动设备包括诸如移动电话、智能手机、平板电脑、个人数字助理、iPod®, iPad®, BLACKBERRY® 设备等的示例性设备。

[0045] 然而,如从多种功能的说明中将变得明显的,现在公开的移动图像处理算法可以应用于,有时有某些修改,来自扫描仪和多功能外围设备(MFP)的图像。类似地,在一些方法中,利用现在公开的处理算法处理的图像可以进一步利用传统扫描仪处理算法进行处理。

[0046] 当然,本文中详细阐述的多个实施例可以利用硬件、软件或它们的任何所期望的组合实现。对此,能够实现本文中详细阐述的不同功能的任何类型的逻辑都可以利用。

[0047] 利用移动设备的一个好处是利用数据方案,基于所捕获的图像的图像处理和信息处理可以用比以前依靠存在扫描仪的方法更加便利、简单化和集成化的方式来完成。然而,使用移动设备作为文件捕获和/或处理设备由于多种原因在此以前被认为是不可行的。

[0048] 在一种方法中,图像可以用移动设备的摄像头捕获。术语“摄像头”应被广泛理解为包括能够捕获设备外面的物理对象的图像的任何类型的设备,所述物理对象诸如一张纸。可以使用任何类型的摄像头。优选的实施例可以使用具有更高分辨率的摄像头,例如8MP或更高,理想地为12MP或更高。图像可以用彩色、灰度、黑白或用任何其他已知的光学效果获取。本文中涉及的术语“图像”意思是包括与摄像头的输出相对应的任何类型的数据,包括未加工的数据、经处理的数据等。

[0049] 本文的说明书被呈现以使得本领域技术人员能够制作和使用本发明,并且在本发明和它们的要求的特殊应用的上下文中被提供。公开的实施例的多种修改对本领域技术人员来说将是轻易显见的,并且本文中限定的总原理可以应用于其他实施例和应用而不脱离本发明的精神和范围。因而,本发明并不意图被限制为示出的实施例,而是被给予和本文中公开的原理和特征相一致的最广泛的范围。

[0050] 尤其是,本文中讨论的本发明的多个实施例利用因特网作为多个计算机系统之间通信的手段来实现。本领域技术人员将认识到,本发明并不被限制为使用因特网作为通信介质,并且本发明的替代性方法可以适应使用私有网、局域网(LAN)、广域网(WAN)或其他通信手段。此外,可以利用有线的、无线的(例如,射频)或光学通信链路的多种组合。

[0051] 可以执行本发明的一个实施例的编程环境示例性地包含一个或多个通用计算机或专用设备,诸如手持式计算机。这种设备(例如,处理器、存储器、数据存储器、输入和输出设备)的细节是已知的并且为了清楚被省略了。

[0052] 还应理解,本发明的方法可以利用多种技术来实现。例如,本文中描述的方法可以用在计算机系统中运行的软件来实现,或使用用于实施操作的一个或多个处理器和逻辑(硬件和/或软件)、专用集成电路、可编程逻辑设备诸如现场可编程门阵列(FPGA)和/或它们的多种组合的硬件来实现。在一个示例性的方法中,本文中描述的方法可以通过居于存储介质诸如物理(例如,非暂态)计算机可读介质上的一系列计算机可执行的指令来实现。此外,虽然本发明的特定实施例可以利用对象导向的软件可编程概念,但本发明并不限于此,并且易于适于利用引导计算机操作的其他形式。

[0053] 本发明还可以用计算机程序产品的形式提供,所述计算机程序产品包括计算机可读存储或信号介质,其上具有计算机代码,所述计算机代码可以由计算设备(例如,处理器)和/或系统执行。计算机可读存储介质可以包括能够在其上存储计算机代码用于由计算设备或系统使用的任何介质,包括光学媒介诸如只读和可写入CD和DVD、磁存储或介质(例如,硬盘驱动、磁带)、半导体存储(例如,FLASH存储和其他便携式内存卡等)、编码在芯片上的固件等。

[0054] 计算机可读信号介质是不适合前述存储介质类别的介质。例如,示例性的计算机可读信号媒介在系统内或系统之间例如经由物理或虚拟网络等通信或否则传送暂态信号。

[0055] 图1例示了根据一个实施例的体系架构100。如图1所示,提供了多个远程网络102,

包括第一远程网络104和第二远程网络106。网关101可以耦合在远程网络102和邻近网络108之间。在本网络体系架构100的上下文中,每个网络104、106可以采用任何形式,包括但不限于LAN、WAN诸如因特网、公共开关电话网络(PSTN)、内部电路网络等。

[0056] 在使用中,网关101用作从远程网络102到邻近网络108的入口点。这样,网关101可以起到路由器和交换机的作用,路由器能够为到达网关101的指定的数据包导向,交换机为指定的数据包提供进出网关101的实际路径。

[0057] 进一步包括至少一个数据服务器114,其耦连至邻近网络108,并且经由网关101从远程网络102易访问。应注意到,数据服务器114可以包括任何类型的计算设备/组件。耦连至每个数据服务器114的是多个用户设备116。这样的用户设备116可以包括台式计算机、膝上型计算机、手持式计算机、打印机或任何其他类型的逻辑。应注意到,在一个实施例中,用户设备111还可以直接耦连至任何网络。

[0058] 外围设备120或外围设备系列120,例如传真机、打印机、网络存储单元等,可以耦连至一个或多个网络104,106,108。应注意到,数据库、服务器,和/或附加的部件可以和耦连至网络104,106,108的任何类型的网络元件一起,或集成到其上被利用。在本说明书的上下文中,网络元件可以指的是任何网络部件。

[0059] 根据一些方法,本文中描述的方法和系统可以用虚拟系统和/或诸如仿真MAC OS环境的UNIX系统、实际上主用MICROSOFT WINDOWS环境的UNIX系统、仿真MAC OS环境的MICROSOFT WINDOWS等仿真一个或多个其他系统的系统实现和/或在其上实现。在一些实施例中,此虚拟化和/或仿真可以通过使用VMWARE软件被加强。

[0060] 在更多的方法中,一个或多个网络104,106,108可以表示通常称为“云”的系统集群。在云计算中,共享的资源,诸如处理功率、外围设备、软件、数据处理和/或存储、服务器等,被提供给云中的任何系统,优选地在按要求的系统中,由此允许跨许多计算系统访问和分配服务。云计算典型地包含因特网或在云中操作的系统之间的其他高速连接(例如,4G LTE、光纤等),但还可以使用连接系统的其他技术。

[0061] 图1例示了根据一个实施例的体系架构100。如图1所示,提供了多个远程网络102,其包括第一远程网络104和第二远程网络106。网关101可以耦合在远程网络102和邻近网络108之间。在本体系架构100的上下文中,每个网络104、106可以采用任何形式,包括但不限于LAN、WAN诸如因特网、公共开关电话网络(PSTN)、内部电路网络等。

[0062] 在使用中,网关101用作从远程网络102到邻近网络108的入口点。这样,网关101可以起到路由器和交换机的作用,路由器能够为到达网关101的指定的数据包导向,交换机为指定的数据包提供进出网关101的实际路径。

[0063] 进一步包括至少一个数据服务器114,其耦连至邻近网络108,并且经由网关101从远程网络102易访问。应注意到,数据服务器114可以包括任何类型的计算设备/组件。耦连至每个数据服务器114的是多个用户设备116。这样的用户设备116可以包括台式计算机、膝上型计算机、手持式计算机、打印机或任何其他类型的逻辑。应注意到,在一个实施例中,用户设备111还可以直接耦连至任何网络。

[0064] 外围设备120或外围设备系列120,例如传真机、打印机、网络存储单元等,可以耦连至一个或多个网络104,106,108。应注意到,数据库和/或附加的部件可以和耦连至网络104,106,108的任何类型的网络元件一起,或集成到其上被利用。在本说明书的上下文中,

网络元件可以指的是任何网络部件。

[0065] 根据一些方法,本文中描述的方法和系统可以用虚拟系统和/或诸如仿真MAC OS环境的UNIX系统、实际上主用MICROSOFT WINDOWS环境的UNIX系统、仿真MAC OS环境的MICROSOFT WINDOWS等仿真一个或多个其他系统的系统实现和/或在其上实现。在一些实施例中,此虚拟化和/或仿真可以通过使用VMWARE软件被加强。

[0066] 在更多的方法中,一个或多个网络104,106,108可以表示通常称为“云”的系统集群。在云计算中,共享的资源,诸如处理功率、外围设备、软件、数据处理和/或存储、服务器等,被提供给云中的任何系统,优选地在按需求的关系中,由此允许跨许多计算系统访问和分配服务。云计算典型地包含因特网或在云中操作的系统之间的其他高速连接(例如,4G LTE、光纤等),但还可以使用连接系统的其他技术。

[0067] 图2示出了根据一个实施例的、与图1的用户设备116和/或服务器114相关联的有代表性的硬件环境。此图例示了具有诸如微处理器的中央处理单元210和经由系统总线212互连的多个其他单元的工作站的典型的硬件配置。

[0068] 图2中示出的工作站包括随机存取存储器(RAM)214,只读存储器(ROM)216,用于连接诸如硬盘存储单元220的外围设备到总线212的I/O适配器218,用于连接键盘224、鼠标226、扬声器228、麦克风232和/或诸如触摸屏和数码相机(未示出)的其他用户接口设备到总线212的用户接口适配器222,用于连接工作站到通信网络235(例如,数据处理网络)的通信适配器234和用于连接总线212到显示设备238的显示适配器236。

[0069] 工作站可以具有常驻其上的操作系统,诸如Microsoft Windows®操作系统(OS)、MAC OS、UNIX OS等。将理解,优选的实施例还可以在除了所提到的那些以外的平台和操作系统上实现。优选的实施例可以利用JAVA、XML、C和/或C++语言,或其他编程语言,和面向对象编程方法一起而写入。可以使用面向对象编程(OOP),其变得越来越多地用于开发复杂应用程序。

#### [0070] 移动图像捕获和处理算法的多个实施例

[0071] 下面描述了处于本公开范围内的移动图像捕获和处理算法以及配置为便于使用这些算法的处理的若干移动应用程序的多个实施例。要领会,下面的每个部分描述了可以在和在其他部分中公开的那些的任何组合中利用的功能,包括本文中描述的任何或多达所有功能。而且,在若干方法中,处理算法实施例以及移动应用程序实施例的功能可以以任何方式组合和/或分布于多个计算资源和/或系统。

[0072] 应用程序可以安装在移动设备上,例如存储在设备的非易失性存储器中。在一个方法中,应用程序包括指令,以在移动设备上实施图像的处理。在另一个方法中,应用程序包括指令,以发送图像到一个或多个非移动设备,例如诸如网络服务器、远程工作站、云计算环境等的远程服务器,如本领域技术人员在阅读本说明书时将理解的。仍然在另一个方法中,应用程序可以包括指令,以确定是否在移动设备上实施一些或所有处理和/或发送图像到远程网站。下面更加详细介绍了图像如何处理的实例。

[0073] 在一个实施例中,在移动设备和远程服务器上实施的处理之间没有差异,除了处理速度、存储器上有效的约束条件等。而且,移动设备上呈现的多个用户接口例如作为移动应用程序的一部分,和与非移动设备通信的显示器上呈现的相应的用户接口之间有一些或没有差异。

[0074] 在其他实施例中,远程服务器可以具有更高的处理功率、更大的能力、更多的处理算法等。仍然在进一步的实施例中,移动设备可以不具有和应用程序相关联的图像处理能力,除了被要求发送图像到远程服务器接收。仍然在另一个实施例中,远程服务器可以不具有和本文中呈现的平台相关的图像处理能力,除了被要求从远程服务器接收经处理的图像。因此,图像可以部分或全部在移动设备上,和/或部分或全部在远程服务器上,和/或部分或全部在云中,和/或部分或全部在总体体系架构的任何部分中处理。而且,一些处理步骤可以在不同的设备上重复。

[0075] 哪个设备实施处理的哪些部分可以由用户限定,可以预确定,可以在运行中确定等。而且,一些处理步骤可以再实施,例如当从用户接收到请求时。因此,未加工的图像数据、经部分处理的图像数据或经全部处理的图像数据可以从移动设备例如利用无线网络传输到远程系统。图像数据随着在远程系统上处理可以返回到移动设备用于输出和/或进一步处理。

[0076] 在进一步的方法中,图像可以被划分,并且多个部分的处理可以分派给多个设备,例如1/2分派给移动设备以及1/2分派给远程服务器,之后经处理的两半被组合。

[0077] 在一个实施例中,选择哪些设备实施处理可以至少部分基于在移动设备本地相对于和服务器通信的相对处理速度。

[0078] 在一个方法中,可以提出处理函数的库,并且移动设备上的应用程序或远程服务器上的应用程序简单地调用这个库,并且基本上调用的意思限定实施哪种处理。然后设备实施那个处理并且输出经处理的图像,或许具有一些相应的元数据。

[0079] 本领域已知的和/或如本文呈现的新的任何类型的图像处理可以在不同的实施例中以任何组合实施。

[0080] 现在参考示例性的图像处理,摄像头可以认为是捕获图像的面积传感器,其中图像可以具有任何数目的投影效果,以及有时是非线性效果。图像被处理以修正这些效果。而且,在处理期间可以找到图像中的文件的位置和边界,例如,围绕页面的、背景中纸的一个或多个实际页面的边界。由于多个实施例的移动本质,那张纸可以放在几乎任何东西上。与处理利用扫描仪生成的文件的图像相比,由于扫描仪背景属性是恒定的并且典型地是已知的,然而移动捕获背景几乎可以根据文件的位置和在图像背景中捕获的相应的周围的纹理而无穷变化,以及由于可变的照明条件,这使图像分析复杂化了。

[0081] 因此,用于由摄像头捕获的纸张所处的的表面的背景的不均匀性提出了一个挑战,并且非线性和投影效果提出了附加的挑战。多个实施例攻克了这些挑战,如很快变得明显的。

[0082] 在一个示例性操作模式中,移动设备上的应用程序可以例如响应于用户请求打开应用程序而开始。例如,可以检测用户选择表示应用程序的图标。

[0083] 在一些方法中,可以请求和/或实施用户认证。例如,可以从用户请求和/或接收用户ID和密码,或任何其他认证信息。

[0084] 在进一步的方法中,可以经由应用程序的图形用户接口使能多个任务。例如,可以给出任务列表。在这种情况下,可以检测由用户选择一个任务,并且可以呈现给用户附加的选择,可以开始预定的任务,摄像头可以开始工作等。

[0085] 可以由移动设备的摄像头捕获图像,优选地当接收到某些类型的用户输入的时

候,诸如检测移动设备屏幕上的轻敲、按压移动设备上的按钮、语言命令、光笔指令等。另一个可能的方案可以包含连续帧的某些级别的分析,例如,从视频流中。在一些方法中,连续帧分析可以由转换跟随,以捕获单个高分辨率的图像帧,其可以自动或由用户触发。而且,该触发可以基于从一个或多个移动设备传感器接收的信息。

[0086] 例如,在一个实施例中,移动设备中或耦合至移动设备的加速度计可以指示摄像头的稳定性,并且应用程序可以分析文件的低分辨率视频帧。如果文件被检测,应用程序可以实施对焦操作并且获取被检测文件的高分辨率图像。

[0087] 低分辨率或高分辨率图像可以进一步处理,但是优选地实施例利用高分辨率图像用于后续处理。在一些方法中,如上面讨论的转换成单帧模式可以是不必要的,尤其是较小的文件诸如名片或收据。为了增加处理效率以及减少处理资源的消耗,文件类型识别可以帮助确定是否转换成单帧模式和/或捕获高分辨率图像用于处理。为了本讨论,假定捕获了一个或多个文件的图像。

[0088] 假定移动设备典型地不具有传统非移动设备的处理功率,一个方法在移动设备上实施一些有限的处理,例如,让用户证实页面已经发现是正确的,图像不是模糊的,和/或照明是充足的,例如,品质的预览。

[0089] 在一个方法中,可以找到由摄像头捕获的图像内的文件。

[0090] 本文中介绍了检测文件的一个或多个边界的附加方法。如果图像中的文件具有非线性或不是矩形,可以应用修正处理。

[0091] 一旦找到图像中的页面,为了使页面成为矩形,一个实施例实施平滑转换,当然假设原始的纸张是矩形的。另一个对图像有用的修正可以是减轻照明的不均匀。

[0092] 在一个示例性方法中,页面检测和矩形化可以大致如下面描述的实施例。

[0093] 移动页面检测的多个实施例

[0094] 例示了用于实施页面检测的示例性方法的一个示例性实施例现在将参考图3A-4描述。参考这些说明,如本文中描述的移动处理算法操控由面积传感器(摄像头)捕获的图像并补偿由此出现的固有困难的优势如何实现将变得清楚。

[0095] 在一个方法中,并且特别参考图3A-3B,边缘检测算法从数字图像300的边界向图像300的中心区域进行,寻找充分不同于已知的背景属性的点。显然,即使由相同的移动设备捕获的图像中的背景304每次也可能是不同的,所以提供了识别图像中的文件的新技术。

[0096] 根据本公开,找到摄像头捕获的图像内的页面边缘有助于适应利用移动设备相对于利用例如扫描仪捕获的图像的属性间的重要差异。例如,由于投影效果,照片中矩形文件的图像看起来可以不是真正的矩形,并且图像中文件的对边可以不具有相同的长度。第二,即使最好的透镜也有一些非线性,其导致对象内的直线,例如大致矩形文件的直边,在所捕获的对象的图像中看起来稍微弯曲。第三,利用摄像头捕获的图像不可避免地趋于在所捕获的图像中引入不均匀的照明效果。这个照明的不均匀使得即使文件放置的表面的完美均匀背景看起来在图像中也具有变化的亮度,并且如果页面不是完全平整的,常常有阴影,尤其是在页面边缘周围。

[0097] 在示例性的方法中,为了避免弄错针对页面边缘的背景内的变化,当前算法利用一个或多个下面的功能。

[0098] 在多个实施例中,图像的框架包含具有周围背景304范围的文件302的数字表示。

在优选的实现方式中,搜索个别页面边缘306可以用跨越(step-over)方法实施,从外到内分析图像的行和列。在一个实施例中,跨越方法可以限定数字图像300内的多个分析窗口308,诸如图3A-3B中示出的。如本文理解的,分析窗口308可以包括一个或多个“背景窗口”,即仅包含描绘数字图像300的背景304的像素的窗口,以及一个或多个“测试窗口”,即包含描绘数字图像300的背景304、文件302的数字表示或两者的像素的窗口。

[0099] 在优选的实施例中,文件的数字表示可以通过将第一分析窗口308即背景分析窗口限定在与文件放置在其上的表面的背景304相对应的图像的范围而在数字图像中被检测。在第一分析窗口308内,可以在第一分析窗口308内限定多个小分析窗口(例如,如图3D中示出的测试窗口312)。利用多个测试窗口312,可以估计描述背景304的一个或多个统计属性的一个或多个分布。

[0100] 继续参考紧接上面讨论的优选的实施例,检测文件的数字表示的边界的下一步可以包括在数字图像内限定多个测试窗口312,并且分析数字图像的相应区域。对每个测试窗口312,可以计算一个或多个描述相应的图像区域的统计值。进一步地,这些统计值可以和相应的描述背景304的统计分布相比较。

[0101] 在优选的方法中,可以沿路径尤其是直线路径限定多个测试窗口312。在尤其优选的方法中,多个测试窗口312可以在水平方向和/或垂直方向,即沿着数字图像的行和列进行限定。而且,可以利用逐步进行法(stepwise progression)沿行和/或列之间的路径限定窗口312。在一些实施例中,如本领域普通技术人员根据读本说明书将理解的,利用逐步进行法可以有利地增加文件检测过程的计算效率。

[0102] 而且,在一些实施例中,开始步骤的规模可以基于图像的分辨率或像素大小估计,但如果文件边的可靠检测是有利的,此步骤可以减少,如下面进一步讨论的。

[0103] 在更多的实施例中,该算法估计描述图像属性的若干统计数据分布,所述图像属性在位于文件周围的背景内的大分析窗口308中找到。在一个方法中,可以在大分析窗口308内限定多个小窗口312,并且可以估计描述小测试窗口312的统计数据的分布。在一个实施例中,在数字图像的背景区域限定大分析窗口308,诸如图像的左上角。

[0104] 描述背景像素的统计数据可以包括任何可以由数字图像数据生成的统计数值,诸如最小值、最大值、中间值、平均值、值的扩展或范围、方差、标准偏差等,如本领域普通技术人员在阅读本说明书时将理解的。数值可以从任何描述数字图像300的数据中采样,诸如一个或多个颜色通道的亮度值,例如红-绿-蓝或RGB、青-品红、黄、黑或CMYK、色调饱和值或HSV等,如本领域普通技术人员在阅读本说明书时将理解的。

[0105] 如图3D所示,小分析窗口312的每个可以包括大分析窗口308内的多个像素的子集。而且,小分析窗口312可以是能够适于大分析窗口308的边界内的任何大小和/或形状。在优选的实施例中,小分析窗口312可以以矩形形状为特征,并且更加优选地是特征为在第一方向(例如,高度)为3像素长并且在第二方向(例如,宽度)是7像素长的矩形。当然,其他小分析窗口的大小、形状和维度在现在公开的处理算法中也是适于实现的。

[0106] 在一个实施例中,可以利用测试窗口分析图像并且检测图像中描绘的文件的数字表示的边界。背景窗口用于估计背景的原始统计属性和/或重新估计背景的局部统计属性。为了解决诸如不均匀照明和/或背景纹理变化的伪影问题,重新估计是必要的和/或有利的。

[0107] 优选地,在一些方法中,可以对文件页面外面范围内的大分析窗口308中的一些或所有的多个小分析窗口312实施统计估计。这种估计可以利用逐步移动大分析窗口308内的小分析窗口312实施,并且逐步移动可以以任何合适的增量进行,以便改变对指定像素采样的数目。例如,为了提升计算效率,分析过程可以限定大分析窗口308内的足够保证每个像素318被采样一次的多个小分析窗口312。因而,在此计算效率方法中限定的多个小分析窗口312将共享共用边界但不重叠。

[0108] 在另一种方法中,设计为提升统计估计的稳定性,分析过程可以限定大分析窗口308内的足够保证每个像素318被采样最大次数的多个小分析窗口312,例如通过减少步幅以在顺序限定的小分析窗口312之间在指定方向上仅产生单个像素偏移。当然,在现在公开的处理算法的多个实施例中可以利用任何步幅增量,如本领域普通技术人员在阅读本说明书时将理解的。

[0109] 技术人员将领会,利用来重新估计数字图像中的局部背景的统计数据的大分析窗口308以及测试窗口可以以任何期望的方式放置在数字图像中。

[0110] 例如,根据图3A中示出的一个实施例,对指定的行*i*的左边边缘的搜索从计算邻近以指定行*i*为中心的图像左边的框架边界的大分析窗口308中的上述统计数据开始。

[0111] 仍然在更多的实施例中,当算法从图像外面的区域朝向其内部区域前进时,当遇到可能的非背景测试窗口(例如,所估计的统计数据的测试窗口不同于最后已知的局部背景的统计特点的分布)时,为了重新评估小分析窗口312内的已选择的统计数据和大分析窗口308内的相应统计数据的局部分布之间的差异的有效性,在一些实施例中,该算法可以退回到之前确定的背景区域,形成新的大分析窗口308并且重新估计背景统计数据的分布。

[0112] 如本领域普通技术人员在阅读本说明书时将领会的,该算法可以以多种方式从图像300的外部区域行进到图像300的内部区域。例如,在一个方法中,该算法以大致螺旋样式行进限定测试窗口312。在其他方法中,该样式可以是沿垂直或水平方向的大致蛇形。仍然在更多的方法中,该样式可以是大致地砂砾状样式。该样式还可以由部分或全部覆盖数字图像300的“顺序掩膜”限定,诸如棋盘样式、垂直地、水平地,或斜线条样式、同心形状等,如本领域普通技术人员在阅读本说明书时将理解的。在其他实施例中,根据一些已限定的程序,诸如大分析窗口308和/或小分析窗口312的分析窗口可以以随机方式、伪随机方式、推测地等遍布数字图像300被限定,如本领域普通技术人员在阅读本说明书时将理解的。该算法可以以任何期望的方式采用测试窗口的顺序行进,只要该路径允许退回到已知的背景,并且该路径用期望的粒度覆盖整个图像。

[0113] 有利地,用这种方式重新计算统计数据有助于适应数字图像300和/或背景304固有的任何照明偏移,否则其会导致图像中的非背景点(例如,图3C中所示的离群候选边缘点316)的错误识别。

[0114] 仍然在更多的实施例中,当该差异统计上有效时,为了再次核查并且因而忽视背景304的纹理中的小变化,诸如木材纹理、表面上的刮痕、表面的图案、小阴影等,该算法可以进一步沿着它的路径跳过一定距离。

[0115] 在附加的和/或可替代的实施例中,在发现可能的非背景点以后,该算法确定该点是否位于阴影的边缘(其可能性,尤其是如果页面的边缘被提升到背景表面之上)并且尝试到达真实的页面边缘。这个过程依靠观察阴影朝向真实边缘一侧通常变暗,跟着图像突然

变亮。

[0116] 利用上面描述的页面边缘检测的方法是因为由于某些原因,使用标准页面检测器可能是不必须的并且甚至不期望的。首先,大多数标准页面检测器包含耗费时间的操作,并且第二,该目前的算法并不涉及附加要求,像监测边缘多么窄、它们跟随哪个方向等。更加重要的是,寻找页面边缘306不必包含页面检测本身,即根据本公开的页面边缘检测可以以并不搜索文件边界(例如,页面边缘306)而是搜索与从背景到文件的转换相关联的图像特点的方式来实施。例如,该转换的特征可以是平衡光滑纸内的白亮度,即通过纹理的改变而不是平均灰度或彩色等级的改变。

[0117] 结果是,得到候选边缘点(例如,图3C所示的候选边缘点314)是可能的,其基本上是网格上每行和每列中的第一和最后的非背景像素。为了消除随机离群点(例如,图3C所示的离群候选边缘点316)并确定哪些候选边缘点314和页面的每条边相对应,在一个方法中分析相邻的候选边缘点是有用的。

[0118] 在一个实施例中,“点”可以认为是数字图像内的任何区域,诸如像素、像素间的位置(例如,具有分数坐标的点,诸如2像素乘2像素正方形的中心)、像素的小窗口等,如本领域普通技术人员在阅读本说明书时将理解的。在优选的实施例中,候选边缘点与测试窗口(例如,3像素乘7像素窗口)的中心相关联,所述测试窗口已经发现由被确定不同于描述局部背景的统计数据的分布的统计数据所表征。

[0119] 如本文中理解的,“相邻的”候选边缘点或“相邻的”像素分别认为是兴趣点或兴趣像素(例如,像素318)附近的或邻近其的点或像素,例如,位于至少部分沿兴趣点或兴趣像素的边界的点或像素、位于兴趣点或兴趣像素的阈值距离内的点或像素(诸如在指定方向、位于兴趣点或兴趣像素的一行内、位于兴趣点或兴趣像素的一列内的2,10,64像素等以内)等,如本领域普通技术人员在阅读本说明书时将理解的。在优选的方法中,“相邻的”点或像素可以是最接近沿特定方向的兴趣点的候选边缘点,所述特定方向例如水平方向和/或垂直方向。

[0120] 每个“好的”边缘点理想地具有至少两个直接相邻的点(每边一个)并且不偏离连接这些相邻点和“好的”边缘点的直线段很远,例如候选边缘点和至少两个直接相邻的点可以适于线性回归,并且其结果的特征可以是确定系数( $R^2$ )不小于0.95。这条直线段关于一个或多个数字图像的边界的角度,和它的相对位置一起确定边缘点是否分配到页面的顶、左、右或底边。在优选的实施例中,候选的边缘点和两个相邻的边缘点可以分配到各自的三角形的角上。如果候选边缘点处三角形的角度接近180度,则候选边缘点可以认为是“好的”候选边缘点。如果候选边缘点处三角形的角度偏离180度大于阈值(诸如20度或更多),则候选边缘点可以从“好的”候选边缘点集合中排除。此启发式算法背后的基本原理是基于期望在确定行和列内的第一和最后的非背景像素时去掉随机错误。这些像素不可能存在于连续的直线中,所以在一些方法中,根据距离和方向核查相邻的点是尤其有利的。

[0121] 为了速度,此网格的步幅可以从大的数字诸如32开始,但其可以减小2倍并且重复搜索边缘点,直到有足够的点来确定页面边缘的基于最小均方差(LMS)的方程(见如下)。如果这个过程即使利用所有行和列之后也不能可靠地确定边,则放弃并将整个图像当作页面。

[0122] 在一个实施例中,页面边的方程如下确定。首先,该算法利用将最差的离群点扔掉

直到所有剩余的支持边缘点都位于离LMS直线的小距离以内的策略将最好的LMS直线适于每条边。例如,具有离沿着文件的特定边界连接多个候选边缘点的大致直线最大距离的点可以被指定为“最差的”离群点。此步骤可以迭代重复以从多个候选边缘点中指定和/或移除一个或多个“最差的”离群点。在一些方法中,候选边缘点偏离连接多个候选边缘点的线的距离至少部分基于数字图像的大小和/或分辨率。

[0123] 如果这条线沿其所有延长线都不被很好地支持,该算法可以尝试将最佳二级多项式(抛物线)适于相同的原始候选点。找到最佳抛物线相比找到最佳直线之间的算法差异是较小的:代替确定方向和线的偏移量的两个未知系数,有确定弯曲度、方向和抛物线的偏移量的三个系数;然而,在一个实施例中,在其他方面过程基本上相同。

[0124] 如果抛物线的支持比直线的支持强,尤其是更接近候选边缘范围的末端,则结论是该算法应该偏向抛物线为图像中页面边的更好模型。否则,在多种方法中,线性模型被利用。

[0125] 为了找到页面四角形(例如,图4中示出的四角形400和下面进一步详细讨论的)的角(可能稍微弯曲),可以计算文件的已找到的四条边的交点。在优选的实现方式中,为了做这个,有必要考虑三种情况:计算两条直线的交点、计算直线和抛物线的交点以及计算两条抛物线的交点。

[0126] 第一种情况有单个解(因为顶和底页面边缘306多半水平延伸,而左和右页面边缘306多半垂直延伸,相应的LMS线不能是平行的),并且该方案确定相应的页面角的坐标。

[0127] 第二种情况计算直线和抛物线的交点稍微更复杂些:作为二次方程的结果可能有0,1或2个解。如果没有交点,可以表明页面检测的致命问题,并且它的结果可能被拒绝。单个解有些不大可能,但不会有进一步的问题。两个交点呈现了选择,在这种情况下更接近框架的相应角的交点是更好的候选点——实际上,方程式的其他解可能离图像框架的候选范围非常远。

[0128] 第三种情况,计算两条抛物线的交点,导致四级多项式方程,其(原则上)可以分析求解。然而,实际上获得解所必需的计算数目可以比用近似迭代算法更大,近似迭代算法也能保证期望的子像素精确度。

[0129] 根据一个方法,下面参考文件302的数字表示的矩形化详细描述了用于此目的的一个示例性的步骤。

[0130] 作为结果的目标四角形(例如,如下面关于图4进一步详细讨论的四角形400)的有效性有若干限制条件。即,四角形优选不太小(例如,在任何期望值的预定阈值以下,诸如图像的总面积的25%),四角形的角优选并不位于图像框架外面太远处(例如,少于100像素远),并且角本身应该优选地可解释为左上、右上、左下和右下处,其对角线交叉于四角形的内部等。在一些实施例中,如果不满足这些限制条件,给出的页面检测结果可以被拒绝。

[0131] 在一个示例性的实施例中,其中所检测到的文件302的数字表示的四角形是有效的,该算法可以确定目标矩形。目标矩形的宽度和高度可以分别设置为四角形的顶边和底边的平均值以及左边和右边的平均值。

[0132] 在一个实施例中,如果实施了倾斜修正,目标矩形的倾斜角度可以设置为0,这样页面边将变得水平和垂直。否则,倾斜角度可以设置为顶边和底边相对于水平轴的角度平均值以及左边和右边相对于垂直轴的角度平均值。

[0133] 在附加的实施例中,用类似的方式,如果实施了裁剪修正,可以指定目标矩形的中心,以便匹配四角形的四个角的坐标的平均值;否则可以计算该中心,以便目标矩形在图像框架的左上方结束。

[0134] 在一些方法中,如果页面检测结果由于任何原因被拒绝,为了得到更多的候选边缘点并且有利地得到更加貌似真实的结果,本文中描述的过程的一些或所有步骤可以用较少的步幅增量重复。在最坏情况的情境下,其中即使采用了允许的最小步幅问题继续存在,则所检测的页面可以设置为整个图像框架并且原始图像可以原样保留。

[0135] 现在详细参考本文描述的发明性页面检测实施例的示例性实现方式,在一个方法中,页面检测包括实施诸如下面描述的方法。如本领域普通技术人员在阅读本说明书时将领会的,该方法可以在任何环境中实施,包括文本描述的那些以及和本公开一起提供的任何附图中表现的那些。

[0136] 在一个实施例中,该方法包括操作,其中限定了和从数字图像背景到文件的数字表示的转换相对应的多个候选边缘点。

[0137] 在多个实施例中,在操作中限定多个候选边缘点包括一个或多个附加的操作,诸如下面描述的操作。

[0138] 在一个操作中并且根据一个实施例,大分析窗口(例如图3A-3B和3D所示的大分析窗口308)在数字图像300内限定。优选地,为了获得数字图像背景304的信息特征用于和非背景的信息特征(例如,文件302的数字表示,诸如下面参考操作进一步详细讨论的背景统计数据)作比较和对比,第一大分析窗口在描绘数字图像背景304的多个像素但不描绘非背景(例如,文件302的数字表示)的区域中限定。例如,第一大分析窗口308可以限定在数字图像300的角(诸如左上角)处。当然,第一大分析窗口可以限定在数字图像300的任何部分而不脱离本公开的范围。

[0139] 而且,如本领域普通技术人员在阅读本说明书时将理解的,大分析窗口308可以是任何大小和/或特征化为任何适合的维度,但在优选的实施例中,大分析窗口308近似40像素高以及近似40像素宽。

[0140] 在特定优选的方法中,大分析窗口308可以限定在数字图像的角区域。例如,参考图3A,示出了数字图像300,数字图像300包括具有多条边306的文件302和背景304的数字表示。如上面参考操作所描述的,大分析窗口308可以限定在包括多个背景像素并且不包括与文件302的数字表示相对应的像素的区域。而且,在一些方法中,大分析窗口308可以限定在数字图像300的角处。

[0141] 在另一个操作中,根据一个实施例,在数字图像300内诸如在大分析窗口308内可以限定多个小分析窗口312。小分析窗口312可以至少部分和一个或多个其他小分析窗口312重叠,这样其特征为包括一个或多个重叠区域320,如图3D所示。在优选的方法中,所有可能的小分析窗口312都限定在大分析窗口308中。当然,小分析窗口可以限定在数字图像的任何部分,诸如图3B中示出的,并且优选地,可以限定小分析窗口以便每个小分析窗口的由单个中心像素表征。

[0142] 仍然在另一个操作中,根据一个实施例,计算一个或多个小分析窗口312(例如,大分析窗口308内的一个或多个小分析窗口312)的一个或多个统计数据并且估计相应的统计数据的一个或多个分布(例如,估计遍及多个小分析窗口312的统计数据的分布)。在另一个

实施例中,可以遍及一个或多个大分析窗口308估计统计数据分布并且可以可选地混合。

[0143] 而且,数值可以描述与数字图像的背景相关联的任何特征,诸如背景亮度值、背景颜色通道值、背景纹理值、背景色调值、背景对比度值、背景锐度值等,如本领域普通技术人员在阅读本说明书时将理解的。仍然而且,统计数据可以包括大分析窗口308内的多个小分析窗口312上描述数字图像背景304的多个像素的一个或多个颜色通道中的亮度值的最小值、最大值和/或范围。

[0144] 仍然在另一个操作中,并且再根据一个示例性的实施例,估计背景统计数据的一个或多个分布。通过估计统计数据的分布,可以得到特征化例如大分析窗口308内数字图像300的背景304的属性的描述性分布。

[0145] 该分布优选地与所计算的每个小分析窗口的背景统计数据相对应,并且可以包括例如亮度最小值的分布、亮度最大值的分布等,由其可以得到分布统计描述符,诸如最小亮度值的最小值和/或最大值、最大亮度值的最小值和/或最大值、亮度值的最小和/或最大范围、最小颜色通道值的最小值和/或最大值、最大颜色通道值的最小值和/或最大值、颜色通道值的最小和/或最大范围等,如本领域普通技术人员在阅读本说明书时将领会的。当然,任何计算出的背景统计数据(例如,用于亮度值、颜色通道值、对比度值、纹理值、色调值、锐度值等的)都可以集成分布并且可以利用任何描述分布的数值而不脱离本公开的范围。

[0146] 仍然在另一个操作中,根据一个实施例,在数字图像300内限定大分析窗口,诸如图3A-3B所示的分析窗口308。

[0147] 而且,窗口形状可以通过正向地设置窗口边界作为数字图像300的一部分来限定,可以通过反向地例如通过给数字图像300应用掩膜并且限定未被掩膜的数字图像300的区域作为分析窗口来限定。仍然而且,窗口可以根据样式来限定,尤其在通过给数字图像300应用掩膜来反向限定窗口的实施例中。当然,可以利用用于限定窗口的其他方式而不脱离本公开的范围。

[0148] 在更多的实施例中,该方法可以包括实施操作,其中计算分析窗口312的一个或多个统计数据。而且,在优选的实施例中,每个分析窗口统计数据与在操作中估计的大分析窗口308的背景统计数据的分布相对应。例如,在一个实施例中,最大亮度和背景亮度最大值的分布相对应,最小亮度和背景亮度最小值的分布相对应,亮度范围和背景亮度范围的分布相对应等,如本领域普通技术人员在阅读本说明书时将理解的。

[0149] 在更多的实施例中,该方法包括确定至少一个分析窗口统计数据和相应的背景统计数据的分布之间是否存在统计上的重大差异。如本领域普通技术人员在阅读本说明书时将领会的,确定是否存在统计上的重大差异可以利用任何已知的统计学的重要评估方法或度量标准来实施,诸如p值(p-value)、z检验(z-test)、卡方相关性(chi-squared correlation)等,如技术人员在阅读本说明书时将领会的。

[0150] 在附加的和/或可替代的方法中,该方法包括在分析窗口中指定一个或多个点(例如,最中心像素318或点),其中描述像素318的值和相应的背景统计数据的分布之间存在统计上的重大差异的点被指定为候选边缘点。该指定可以通过本领域已知的任何合适的方法,诸如设置与像素相对应的标识、分类像素的坐标、制作像素坐标的阵列、修改描述像素318的一个或多个值(诸如亮度、色调、对比度等)或通过任何其他合适的手段来完成。

[0151] 在操作中,根据一个实施例,一个或多个操作可以重复一次或多次。在优选的实施

例中,可以实施多个这样的重复,其中每个重复在数字图像的不同部分上实施。优选地,该重复可以被实施,直到文件的数字表示的每条边都已经被评估。在多个方法中,限定分析窗口308、312可以导致共享一个或多个边界的多个分析窗口308、312,整个或部分重叠的多个分析窗口308、312,和/或不共享任何共同边界且不重叠的多个分析窗口308、312等,如本领域普通技术人员在阅读本说明书时将理解的。

[0152] 在尤其优选的实施例中,多次重复可以在检测可能的非背景窗口(例如,包含候选边缘点的窗口或包含诸如不均匀照明、背景纹理变化等的伪影的窗口)时被用于重新估计局部背景统计数据的方式来实施。

[0153] 在操作中,根据一个实施例,四角形400的四条边可以基于多个候选边缘点来限定。优选地,四角形400的四条边包含数字图像300中的文件302的数字表示的边缘306。在一些方法中,限定四角形400的各边可以包括实施一次或多次最小均方差(LMS)近似。

[0154] 在更多的方法中,限定四角形400的各边可以包括识别一个或多个离群候选边缘点,并且从多个候选边缘点中移除一个或多个离群候选边缘点。进一步地,限定四角形400的各边可以包括实施至少一次排除一个或多个离群候选边缘点的附加LMS近似。

[0155] 仍然进一步地,在一个实施例中,四角形400的每条边用从一类函数中选择的方程式表征,并且实施至少一次LMS近似包括确定每个方程式的一个或多个系数,在优选的实现方式中诸如二级多项式的最佳系数。根据这些方法,限定四角形400的各边可以包括确定文件的数字表示的每条边是否落入指定的函数类别内,诸如二级多项式或诸如线性函数而不是二级多项式的简单函数。

[0156] 在优选的方法中,实施方法可以准确限定文件的四条优势边周围的四角形,而忽略偏离文件的优势边的一个或多个点,诸如图3A-3C和图4中描绘的裂口310和/或突耳320。

[0157] 现在公开的四角形400的附加的和/或可替代的实施例的特征可以是具有四条边并且每条边的特征是一个或多个诸如上面讨论的多项式函数的方程式。例如,四角形400的各边以不止一个方程式为特征的实施例可以包含将一条或多条边分成多个部分,每部分的特征是诸如上面讨论的多项式函数的方程式。

[0158] 在多个实施例中,限定四角形400可以可替代地和/或附加地包括限定四角形400的一个或多个角。例如,在计算出多个交点的情况下,四角形400的角可以通过计算四角形400的邻边之间的一个或多个交点并且从一个或多个计算出的交点中指定合适的交点来限定。仍然在更多的实施例中,限定角可以包括解出一个或多个方程式,其中每个方程式的特征是属于所选择的一类函数,诸如N级多项式等,如本领域普通技术人员在阅读本说明书时将理解的。

[0159] 在多个实施例中,四角形400的角可以通过一个或多个的下列项来确定:四角形400的两条弯曲邻边的交点;两条大致直线的交点;以及一条大致直线和一条大致曲线的交点。

[0160] 仍然在更多的实施例中,该方法可以包括附加的和/或可替代的操作,其中文件302的数字表示和四角形400被输出到移动设备的显示器。输出可以以任何方式实施,并且可以依赖于移动设备硬件和/或软件的配置。

[0161] 而且,输出可以用多种方法实施,以便帮助进一步的处理和/或用户和输出交互。例如,在一个实施例中,四角形400可以以设计为将四角形400和数字图像300的其他特征区

分开的方式来显示,例如通过用特定颜色、图案、照明主题、作为动画片等显示四角形400,如本领域普通技术人员在阅读本说明书时将理解的。

[0162] 仍然进一步的,在一些实施例中,输出四角形400和文件302的数字表示可以帮助用户以任何合适的方式手动调节和/或限定四角形400。例如,用户可以和移动设备的显示器进行交互以转换四角形400,即在一个或多个方向上移动四角形400的位置而维持四角形400的纵横比、形状、边长、面积等。附加地和/或可替代地,用户可以和移动设备的显示器进行交互以手动限定或调解四角形400的角的位置,例如敲击四角形400的角并且将该角拖拽到数字图像300内的期望位置,诸如文件302的数字表示的角。

[0163] 再次参考图4,描绘了页面检测的理想结果的一个特定示例,其示出了数字图像300内文件302的数字表示,并且具有包围文件302的数字表示的边缘的四角形400。

[0164] 在一些方法中,诸如上面描述的页面检测方法可以包括一个或多个附加的和/或可替代的操作,诸如将在下面描述的。

[0165] 在一个方法中,页面检测可以进一步包括捕获一个或多个包含文件的数字表示的图像数据和关于文件的数字表示的音频数据。捕获可以利用一个或多个耦连至移动设备的诸如麦克风、摄像头、加速度计、传感器等捕获部件来实施,如本领域普通技术人员在阅读本说明书时将理解的。

[0166] 在另一个方法中,一旦确定存在统计上的重大差异,页面检测可以包括限定新的大分析窗口309和重新估计新的大分析窗口309的背景统计数据的分布,即本质上是在已识别出可能的非背景点的点附近,诸如文件的边缘306的一个附近,的数字图像300的不同区域中重复操作1908和/或1910。

[0167] 在若干示例性的实施例中,大分析窗口308可以定位在行中最左侧的非背景像素附近或在该处或定位在行中最右侧的非背景像素附近或在该处,可以定位在列中最顶侧的非背景像素附近或在该处或定位在列中最底侧的非背景像素附近。

[0168] 包含这种重新估计的方法可以进一步包括确定至少一个小分析窗口统计数据 and 相应的经重新估计的大分析窗口统计数据的分布之间是否存在统计上的重大差异。以这种方式,对是否存在统计上的重大差异得到更可信的确定是可能的,并且因此更好地区分从数字图像背景到文件的数字表示的真正转换,而不是例如纹理的变化、照明不均匀和/或数字图像内的其他伪影到数字表示的转换。

[0169] 而且,实施或不实施如上描述的重新估计可以帮助该方法避免数字图像中诸如照明和/或背景纹理等的变化的伪影,所述伪影不与从数字如下背景到文件的数字表示的真正转换相对应。在一些方法中,避免伪影可以采取绕过数字图像的一个或多个区域的形式(该区域的特征为将该区域与真正的背景区分开的纹理、变化等)。

[0170] 在一些方法中,一旦确定在所估计的大分析窗口308的统计分布和所计算的小分析窗口312的相应的统计数据之间存在统计上的重大差异,在小分析窗口附近限定新的大分析窗口,重新估计新的大分析窗口的统计数据的分布,并且确定在重新估计的统计分布和所计算的小分析窗口312的相应的统计数据之间不存在统计上的重大差异,则可以绕过一个或多个区域。

[0171] 在其他的方法中,一旦确定此经核查的窗口的统计数据没有明显不同于已知的背景统计属性,例如,如通过测试统计意义来表明,则绕过可以通过进一步沿路径核查另一

个分析窗口312并且重新开始搜索向非背景的转换来完成。

[0172] 如技术人员在阅读本公开时将领会的,绕过可以通过进一步沿路径核查另一个分析窗口来完成。

[0173] 仍然在进一步的方法中,页面检测可以附加地和/或可替代地包括确定四角形400是否满足一个或多个质量控制度量标准;并且一旦确定四角形400不满足一个或多个质量控制度量标准,则拒绝四角形400。而且,质量控制度量标准可以包括测量诸如LMS支持的度量标准、最小四角形400面积度量标准、四角形400角的位置度量标准以及四角形对角线交点位置度量标准。

[0174] 实际上,确定四角形400是否满足一个或多个的这些度量标准起到核查该方法的表现的作用。例如,核查可以包括确定四角形400是否覆盖至少全部数字图像面积的阈值,即四角形400是否包括总图像面积的至少25%。

[0175] 而且,核查可以包括确定四角形400的对角线是否交叉于四角形400的边界内部,确定是否由足够数据计算出一个或多个的LMS近似值,以对从其得出的统计数据有强的可信度,即LMS近似是否有足够的“支持”(在多个方法中,诸如从至少5个数据点或总数目的至少四分之一的数据点计算出近似值),和/或确定四角形400角的位置(由表征四角形400的每个相应边的方程所限定)是否存在于数字图像的边缘的阈值距离以内,例如,四角形400的角是否位于离指定方向的数字图像的边缘多于100像素远。当然,可以利用其它质量度量标准和/或核查方法而不脱离这些公开的范围,如本领域普通技术人员在阅读本说明书时将领会的。

[0176] 在一个方法中,质量度量标准和/或核查方法可以帮助拒绝非最理想的四角形400的限定,并且进一步帮助改进四角形400的边的限定。例如,一个方法包含接收基于多个候选边缘点限定四角形400的四条边未能限定有效的四角形400,即未能满足一个或多个的质量控制度量标准的指示;以及重新限定多个候选边缘点。尤其是,在这个实施例中,一旦尝试失败,重新限定多个候选边缘点包括采样位于数字图像内的比之前所采样的点的数目更大数目的点。在一个方法中,为了分析更大数目的候选边缘点,这可以通过减少数字图像的一个或多个的行和列之上的步幅并且重复算法的所有步骤来完成。步幅可以在垂直方向、水平方向或两方向减少。当然,可以利用重新限定候选边缘点和/或重新采用数字图像内的点的方法而不脱离本公开的范围。

[0177] 仍然进一步的,页面检测可以包括指定整个数字图像为文件的数字表示,尤其是多次重复方法也未能限定有效的四角形400,甚至在数字图像分析中采用明显连续减少的步幅。在一个方法中,指定整个数字图像为文件的数字表示可以包括限定图像的角为文件的角,限定图像的边为文件的边等,如本领域普通技术人员在阅读本说明书时将理解的。

[0178] 如本文中描述的,四角形400的对角线的特征可以是第一条线连接所计算的四角形400的左顶角和所计算的四角形400的右底角,并且第二条线连接所计算的四角形400的右顶角和所计算的四角形400的左底角。而且,第一条线和第二条线优选地交叉于四角形400的内部。

[0179] 在多个方法中,一个或多个的前述操作可以利用处理器实施,并且处理器可以是移动设备的一部分,尤其是具有集成摄像头的移动设备。

[0180] 弯曲修正

[0181] 现在描述的发明性概念包括修正数字图像中的弯曲,并且尤其是修正文件的数字表示的边的弯曲。修正弯曲的多种方法将在下面示例性参考图5A-5B详细描述。

[0182] 在一个实施例中,弯曲修正算法的目标是将四角形400平滑转换成四边形。尤其是,四角形400以多个方程式为特征,每个方程式与四角形400的一条边相对应并且从所选择的一类函数中选择。例如,四角形400的每条边可以以一级多项式、二级多项式、三级多项式等为特征,如技术人员在阅读本说明书时将领会的。

[0183] 在一个方法中,四角形400的边可以用方程式描述,并且在优选的实施例中四角形400的左边以二级多项式 $x = a_2*y^2 + a_1*y + a_0$ 为特征;四角形400的右边以二级多项式 $x = b_2*y^2 + b_1*y + b_0$ 为特征;四角形400的顶边以二级多项式 $y = c_2*x^2 + c_1*x + c_0$ 为特征;并且四角形400的底边以二级多项式 $x = d_2*x^2 + d_1*x + d_0$ 为特征。

[0184] 本文中呈现的弯曲修正的说明利用四角形内的多个基于四角形的固有坐标对(p, q)的定义,每个固有坐标对(p, q)与顶到底的曲线和左到右的曲线的交点相对应,所述顶到底的曲线通过按照顶到底曲线p比1-p的系数比组合所有相应系数来从其左边和右边的方程所得到的方程式来表征,所述左到右的曲线通过按照左到右曲线q比1-q的系数比组合所有相应系数来从其顶边和底边的方程所得到的方程式来表征,其中 $0 \leq p \leq 1$ ,并且 $0 \leq q \leq 1$ 。

[0185] 在优选的实施例中,其中四角形400的边以二级多项式方程式为特征,与固有坐标p相对应的顶到底曲线将以方程式 $x = ((1-p)*a_2 + p*b_2)*y^2 + ((1-p)*a_1 + p*b_1)*y + ((1-p)*a_0 + p*b_0)$ 为特征,并且与固有坐标q相对应的左到右曲线将以方程式 $y = ((1-q)*c_2 + q*d_2)*x^2 + ((1-q)*c_1 + q*d_1)*x + ((1-q)*c_0 + q*d_0)$ 为特征。当然,其他方程式可以是上面描述的任何边和/或曲线的特征,如本领域普通技术人员在阅读本说明书时将领会的。

[0186] 对于平行四边形,固有坐标变得特别简单:在平行四边形内,每个固有坐标对(p, q)与平行于平行四边形的左边和平行四边形的右边的每个的线和平行于平行四边形的顶边和平行四边形的底边的每个的线的交点相对应,平行于平行四边形的左边和平行四边形的右边的每个的线例如以p比1-p的比例分割顶边和底边两者的线;平行于平行四边形的顶边和平行四边形的底边的每个的线例如以q比1-q的比例分割顶边和底边两者的线,其中 $0 \leq p \leq 1$ ,并且 $0 \leq q \leq 1$ 。在另一种特定情况下,当四角形是单位正方形,即边长为1的正方形,固有坐标正好是普通直角坐标:坐标为(p, q)的点是垂直的线 $x = p$ 和水平的线 $y = q$ 的交点。

[0187] 下面描述的弯曲修正算法的目标是将经弯曲修正的图像中的每个点与原始图像中的相应的点相匹配,并且按照将四角形400的四条边的每条边转换成连接它现在的角的大致直线这样的方式进行;然而,这样的技术可以将由它的四条边的方程式描述的任何四角形转换成任何其他这样的四角形。

[0188] 下面描述的坐标映射算法的主要观点是通过如下方式完成这个目标:首先,计算目的图像中的每个点P(未示出)的固有坐标(p, q),第二,将这些坐标与原始图像中的固有坐标的相同坐标对(p, q)相匹配,第三,计算分别与这些固有坐标相对应的左到右曲线和顶到底曲线的交点坐标,以及最后,将在原始图像中找到的点处的颜色值或灰度值分配到点P。

[0189] 现在参考图5A,其描绘了根据一个实施例的、页面弯曲修正算法的第一次循环的图形表示。如图5A所示,数字图像500中的每个点可以与对应于与点相关联的固有坐标的顶

到底曲线504和左到右曲线506(曲线可以包括直线、弯曲的线,例如抛物线等,如本领域普通技术人员在阅读本说明书时将理解的)的交点对应。

[0190] 如从本说明书将变得明显的,弯曲修正可以包含多条这样的左到右曲线506和顶到底曲线504。

[0191] 而且,弯曲修正可以包括将目标固有坐标与文件502的数字表示的原始固有坐标相匹配。

[0192] 如图5A所示,这个匹配可以包括循环搜索指定的左到右曲线506和指定的顶到底曲线504的交点。图5A示出了本公开范围内的示例性循环搜索的第一次循环。

[0193] 根据下面进一步详细讨论的一个方法,循环搜索包括指定具有坐标 $(x_0, y_0)$ 的开始点508,开始点508可以位于文件502的数字表示内的任何地方,但优选地位于目标四边形的中心处或其附近。

[0194] 循环搜索可以包括将开始点508投影到两条相交的曲线504,506中的一条上。虽然开始点可以被投影到曲线504,506中的一条上,在一个方法中,循环搜索中第一次循环的第一半包括将开始点508投影到顶到底曲线上,以得到下个点的x坐标 $(x_1)$ ,在图5A中投影结果用点510表示,其具有坐标 $(x_1, y_0)$ 。类似地,在一些实施例中,循环搜索中第一次循环的第二半包括将点510投影到左到右曲线506上,以得到下个点的y坐标 $(y_1)$ ,在图5A中投影结果用点512表示,其具有坐标 $(x_1, y_1)$ 。

[0195] 图5B是根据一个实施例的、在将文件502的数字表示分成由多条顶到底的曲线504和多条左到右的曲线506限定的多个大小相等的部分之后,页面弯曲修正算法的开始点的图形表示。

[0196] 在一些实施例中,进一步的循环可以利用类似的方法,诸如下面进一步详细描述。

[0197] 继续参考图5A-5B,根据一个实施例,用于修改数字图像中的文件的数字表示的一个或多个空间特征的方法包括一个或多个的下列操作。如本领域普通技术人员在阅读本说明书时将领会的,该方法可以在任何合适的环节中实施,包括那些在本公开的附图和相应的说明中示出和/或描述的。

[0198] 在一个实施例中,该方法包括将四角形400转换成四边形的操作。尤其是,该四角形400以多个方程式为特征,每个方程式与四角形400的边相对应并且从所选择的一类函数中选择。例如,四角形400的每条边可以以一级多项式、二级多项式、三级多项式等为特征,如技术人员在阅读本说明书时将领会的。

[0199] 在一个实施例中,四角形400的边可以用方程式描述,并且在优选的实施例中,四角形400的左边以二级多项式 $x = a_2*y^2 + a_1*y + a_0$ 为特征;四角形400的右边以二级多项式 $x = b_2*y^2 + b_1*y + b_0$ 为特征;四角形400的顶边以二级多项式 $y = c_2*x^2 + c_1*x + c_0$ 为特征;并且四角形400的底边以二级多项式 $x = d_2*x^2 + d_1*x + d_0$ 为特征。而且,顶到底曲线的方程式是: $x = ((1-p)*a_2 + p*b_2)*y^2 + ((1-p)*a_1 + p*b_1)*y + ((1-p)*a_0 + p*b_0)$ ,并且左到右曲线的方程式是: $y = ((1-q)*c_2 + q*d_2)*x^2 + ((1-q)*c_1 + q*d_1)*x + ((1-q)*c_0 + q*d_0)$ 。当然,其他方程式可以是上面描述的任何边和/或曲线的特征,如本领域普通技术人员在阅读本说明书时将领会的。

[0200] 在一个实施例中,曲线504,506可以用适于一个或多个的下列一般形式的示例性的多项式函数描述。

[0201]  $x_1 = u_2 * y_0^2 + u_1 * y_0 + u_0;$

[0202]  $y_1 = v_2 * x_1^2 + v_1 * x_1 + v_0,$

[0203] 其中,  $u_i = (1-p) * a_i + p * b_i$ , 并且  $v_i = (1-q) * c_i + q * d_i$ , 并且其中,  $a_i$  是四角形的左边的方程式的系数,  $b_i$  是四角形的右边的方程式的系数,  $c_i$  是四角形的顶边的方程式的系数,  $d_i$  是四角形的底边的方程式的系数, 并且  $p$  和  $q$  是与曲线 504, 506 相对应的基于四角形的固有坐标。在一些方法中, 诸如  $a_i, b_i, c_i, d_i$  等的系数可以由在实施页面检测过程中完成的计算、估计和/或确定获得, 诸如上面讨论的关于页面检测的页面检测方法。

[0204] 当然, 如本领域普通技术人员将理解的, 将四角形 400 转换成四边形可以包括一个或多个附加的操作, 诸如将在下面更加详细描述。

[0205] 在一个实施例中, 该方法可以附加地和/或可替代地包括将四角形 400 的一个或多个区域用足够平滑以避免引入附加的伪影的方式延伸成作为结果的四边形。

[0206] 在一些方法中, 将四角形 400 转换成矩形可以包括确定矩形的高度、矩形的宽度、矩形的倾斜角和/或矩形的中心位置。例如, 这种转换可以包括将目标矩形的宽度限定为四角形 400 的顶边的宽度和底边的宽度的平均值; 将目标矩形的高度限定为四角形 400 的左边的高度和右边的高度的平均值; 依赖于图像中的矩形的期望位置限定目标矩形的中心; 并且限定目标矩形的倾斜角度, 例如响应于用户请求抗倾斜文件的数字表示。

[0207] 一旦得到(直边的)四边形, 现在公开的算法可以用投影修正组件进行将四边形转换成目标矩形, 有利地具有通过比较经修正的图像中多个对象元件的像素位置和扫描的图像中各自的对象元件的相应像素位置所测量的非常低的错误。在优选的实施例中, 残余的错误在具有大约 500 点每英寸(DPI)的分辨率的图像中可以是大约 5 像素或更少。这不与对象的“经修正的图像”上的像素相对应, 该对象在任何方向离该对象的扫描图像中的相同像素的相应位置多于 5 像素远。

[0208] 投影修正

[0209] 在一个方法中, 本文中描述的投影修正基本上包括通过上面描述的弯曲修正算法将产生的四边形转换成真正的矩形。优选地, 当从特定视角(例如, 在该对象的正常角度, 诸如如果用传统平台式扫描仪、多功能设备等扫描对象将是捕获角度, 如本领域普通技术人员在阅读本说明书时将理解的)观察时, 该矩形大致表示在数字图像中捕获的对象的实际尺寸、纵横比等。

[0210] 在图 6A-6D 中示意性地表明了多个捕获角和相关联的投影效果。

[0211] 在一些方法中, 投影修正可以包括对图像数据应用诸如四点算法的算法。在多个实施例中, 这样的算法可以依赖或可以不依赖于一个或多个的下列假设: 1) 从捕获角度的视角, 沿每个宽度大小和高度大小, 3D 对象的厚度是 0, 并且所捕获的 3D 对象的大小是非零; 2) 宽度大小和高度大小的纵横比是已知的。纵横比的值不必精确已知, 可以容忍小的测量错误, 其会影响 3D 重建错误。在优选的实施例中, 特征为错误小于大约 10% 的测量是可接受的(例如, 与诸如信件、法定的、A4、A5、驾驶证、信用卡、销售收据、名片等的预定文件类型相对应的纵横比); 3) 捕获的图像中四个角的像素的 2D 像素位置利用图像分段技术是可估计的, 所述图像分段技术诸如本文中描述的页面检测中所利用的; 4) 所捕获的图像和参考图像中的对象角的位置(其与对象的“真实世界的坐标”相对应)通过参考图像和捕获的图像中的四个像素/角的预建立的对应来描述; 以及 5) 3D 重建实现了和从真正对象的特定视角

观察的那些大致相同的重建图像中的像素位置,例如,如在利用垂直于对象捕获角度捕获的500DPI图像中观察到的,诸如由从相同视角扫描对象的2D表示将创建的。

[0212] 在一个实施例中,现在描述的页面检测算法可以用于估计文件的角的2D像素位置。为了找到(可能稍微弯曲的)页面四角形(例如,如图4中示出的四角形400)的角,计算文件的所找到的四条边的交点。

[0213] 在另一个实施例中,并且特别参考参考图像和所捕获的图像的像素坐标/位置(尤其是角的坐标/位置)之间的对应,参考图像中对象(例如,文件)左顶角、左底角、右底角和右顶角的坐标优选地与所捕获的图像中各自的对象角的坐标/位置相对应。在确定中,确定每个对应中的精确的相互关系可以用纹理和/或图像特征作为参考点。例如,在一个实施例中,纹理方向和文件纵横比可以用于确定所捕获的文件的方向。

[0214] 仅为了示例性目的,下面的说明将例示在驾驶证的数字图像上实施的视角修正的一个实施例。驾驶证的参考图像可以优选地利用扫描仪、多功能打印机或本领域已知的其他设备来捕获,以不将视角倾斜或变形引入用其捕获的图像中。参考图像可以优选地已利用已知分辨率的扫描仪捕获,更加优选地分辨率为近似500DPI,以及以已知的捕获角度,最优选地以文件的正常角度(例如,90度的捕获角度,诸如上面图6B中示出的)。

[0215] 利用参考图像,以及优选地当用关于对象的“真正的”形状、大小、维度、纹理等先验知识补充时,在一定程度上重建对象是可能的,这样在重建的图像中表示的该对象具有同参考图像相同的或几乎相同的特征。在一些实施例中,重建的图像和参考图像可以是不同的但可以按比例缩放。在这种情况下,将缩放操作应用于重建的图像可以消除比例尺的任何差异,这样重建的对象具有如参考图像中描述的对象相同的或几乎相同的特征。

[0216] 用户可以利用移动设备捕获他们的驾驶证的图像,并且可能是在陡峭的捕获角度(例如偏离正常的大约30度或更多的捕获角度)。结果是,在捕获的图像中驾驶证的表示的特征是3D视角变形,使得大致的矩形文件形状上看起来是梯形的。在极端情况下,诸如当利用远离法线大于30度的捕获角度观察时,离捕获设备最远的边缘的长度看起来会比离捕获设备最近的边缘的长度更短等,如本领域普通技术人员在阅读本说明书时将理解的。在一个实施例中,远边的长度看起来会明显更短,例如,和近边的长度的50%一样小,尤其是当利用陡峭的捕获角度时(例如,偏离法线30度或更大)。

[0217] 在多个实施例中,3D重建优选地最小化了在重建原始图像中引入任何变形,以生成所捕获的对象/文件的矩形表示。

[0218] 在一个实施例中,视角修正可以包括利用移动设备捕获图像,识别所捕获的图像内的四个点、像素等,每个点/像素与所检测的驾驶证(或其他四角形文件)的可能的角相对应,以及至少部分基于四个角的像素建立3D转换。当形成图像的边界的四条边的多边形被分段时,四个像素的位置可以估计。

[0219] 平面单应性/投影变换是两个平面之间的非奇异线性关系。在这种情况下,单应性变换限定了所捕获的图像和图像平面之间的四个角的像素/位置的线性映射。在一些方法中,摄像头参数的计算可以利用单应性变换H的估计,诸如方程式(1)中示出的。

[0220] 
$$\lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}}_{\text{homography } H} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}, \quad (1)$$

[0221] 如在上面的方程式(1)中描述的:

[0222] ● $\lambda$ 是“参考”或“真实世界”坐标系中位置(X,Y,Z)的焦深,(例如,来自参考图像的坐标系,诸如在上面的图6B和7中示出的)。给与另一种方式, $\lambda$ 可以认为是参考坐标系中的点(X,Y,Z)和捕获设备之间的直线距离;

[0223] ●(x,y,z)是所捕获的图像中指定像素位置的坐标;并且

[0224] ●H是具有元素 $h_{ij}$ 的(3×3)矩阵,其中i和j分别限定相应的行和列

[0225] 索引。

[0226] 在一个方法中,方程式1中描绘的坐标(x,y)和(X,Y)与所捕获的图像平面和参考图像(例如,图6B中所示的)中各自的点的坐标相对应。坐标Z设置为0,对应于假设在每个位置描述的对象沿着单个(例如X-Y)平面具有零厚度。在一个实施例中,上述计算中我们可以省略方程式1中的z值,因为它在确定单应性矩阵中不必发挥任何作用。

[0227] 因而,单应性矩阵H可以通过检测四个点对应 $p_i \leftrightarrow P_i'$ 来估计,其中 $p_i = (x_i, y_i, 1)^T$ 是所捕获的图像平面中四个角的位置;以及 $P_i' = (X_i, Y_i, 1)^T$ 是相对应的四个角的点的坐标,其中在下面的讨论中i是范围从1到n的点的索引值。利用之前引入的标记法,方程式(1)可以写成如下面的方程式(2)所示。

[0228] 
$$\lambda p_i = H P_i'. \quad (2)$$

[0229] 为了消除缩放因素,在一个实施例中,计算方程式(2)的每项的叉积是可能的,如方程式(3)所示的:

[0230] 
$$p_i \times (\lambda p_i) = p_i \times (H P_i'), \quad (3)$$

[0231] 因为 $p_i \times p_i = 0_3$ ,方程式(3)可以写成如下面的方程式(4)所示。

[0232] 
$$p_i \times H P_i' = 0_3. \quad (4)$$

[0233] 因而,矩阵结果 $H P_i'$ 可以表示为方程式(5)。

[0234] 
$$H P_i' = \begin{bmatrix} h^{1T} P_i' \\ h^{2T} P_i' \\ h^{3T} P_i' \end{bmatrix}, \quad (5)$$

[0235] 其中, $h^{mT}$ 是H的第m行的转置矩阵(例如, $h^{1T}$ 是H的第一行的转置矩阵, $h^{2T}$ 是H的第二行的转置矩阵等)。因此,重写方程式(4)如下是可能的:

[0236] 
$$p_i \times H P_i' = \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} \times \begin{bmatrix} h^{1T} P_i' \\ h^{2T} P_i' \\ h^{3T} P_i' \end{bmatrix} = \begin{bmatrix} y_i h^{2T} P_i' - h^{2T} P_i' \\ h^{1T} P_i' - x_i h^{3T} P_i' \\ x_i h^{3T} P_i' - y_i h^{1T} P_i' \end{bmatrix} = 0_3. \quad (6)$$

[0237] 尤其是,方程式(6)在 $h^{mT}$ 和 $h^{mT} P_i' = P_i'^T h^m$ 中是线性的。因而,方程式(6)可以重新表示为如下方程式(7)所示的形式:

[0238] 
$$\begin{bmatrix} 0_3^T & -P_i'^T & y_i P_i'^T \\ P_i'^T & 0_3^T & -x_i P_i'^T \\ -y_i P_i'^T & x_i P_i'^T & 0_3^T \end{bmatrix} \begin{bmatrix} h^1 \\ h^2 \\ h^3 \end{bmatrix} = 0_3. \quad (7)$$

[0239] 注意到,方程式(7)中示出的矩阵的行不是线性独立的。例如,在一个实施例中,第

三行是 $-x_i$ 乘以第一行和 $-y_i$ 乘以第二行的和。因而,对每个点对应,方程式(7)提供了两个线性独立的方程式。两个第一行优选地用于解出H。由于单应性变换利用单应坐标写成齐次坐标写成,在一个实施例中单应性矩阵H可以用8个参数加上齐次缩放因子(其可以视为第9个自由参数)进行限定。在这样的实施例中,提供了8个方程式的至少4个点对应可以用于计算单应性矩阵。实际上,并且根据一个示例性的实施例,优选地利用更多数目的对应,以便获得过定线性系统,其导致更加可靠的结果(例如,相对像素位置的较低错误)。通过用向量形式重写H为 $h=[h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}]^T$ , n对点对应使能创建 $2n \times 9$ 线性系统,其用方程式(8)表示。

[0240]

$$\underbrace{\begin{pmatrix} 0 & 0 & 0 & -X_1 & -Y_1 & -1 & y_1 X_1 & y_1 X_1 & y_1 \\ X_1 & Y_1 & 1 & 0 & 0 & 0 & -x_1 X_1 & -x_1 Y_1 & -x_1 \\ 0 & 0 & 0 & -X_2 & -Y_2 & -1 & y_2 X_2 & y_2 X_2 & y_2 \\ X_2 & Y_2 & 1 & 0 & 0 & 0 & -x_2 X_2 & -x_2 Y_2 & -x_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & -X_n & -Y_n & -1 & y_n X_n & y_n X_n & y_n \\ X_n & Y_n & 1 & 0 & 0 & 0 & -x_n X_n & -x_n Y_n & -x_n \end{pmatrix}}_C \begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{pmatrix} = 0_{2n} \tag{8}$$

[0241] 如方程式8所示,前两行与第一特征点对应,如用坐标X,Y,x,y的下标值所指示的。在这种情况下下标值是1。第二个两行与第二特征点对应,如用下标值2所指示的,最后两行与第n个特征点对应。对于四点算法,n是4,并且特征点是文件页面的四个角。

[0242] 解出这个线性系统包含计算奇异值分解(SVD)。这个SVD与将矩阵重做成矩阵乘积 $C=UDV^T$ 的形式相对应,其中,解h与矩阵C的最小特征值的特征向量相对应,在一个实施例中,当特征值按降序分类时,其可以位于矩阵V的最后一列。

[0243] 当使用两个或更多摄像头时,诸如传统上实施的用于立体机视觉系统的,值得注意矩阵C不同于用于估计基本矩阵的八点算法中所用的典型矩阵。尤其是,虽然传统上用在八点算法中的元素组成了投影到两个摄像头平面上的特征点,现在描述的矩阵C中的元素组成了投影到仅单个摄像头平面上的特征点和3D对象上的相应特征点。

[0244] 在一个实施例中,为了避免数值不稳定,点对应的坐标优选地可以标准化。这可以例如利用称为标准化的直接线性变换(DLT)算法技术来完成。例如,在一个实施例中,在估计出单应性矩阵之后,方程式(1)可以用于计算指定值(X,Y)的每个像素位置(x,y)。在实际应用中,挑战包含当(x,y)的值是给定的或已知的先验时计算(X,Y)。如方程式(1)所示,并且在优选的实施例中,(x,y)和(X,Y)是对称的(即,当(x,y)和(X,Y)的值被对换时,方程式1的有效性依然有效)。在这种情况下,可以估计“逆”单应性矩阵,并且给出在所捕获的图像中例如相机视图中描绘的对象的相应2D坐标后,这个“逆”单应性矩阵可以用于重建对象的3D(即“参考”或“真实世界”)坐标。

[0245] 基于前述,实现现在描述的四点算法(以及其任何等价变化和/或修改,如技术人员在阅读这些说明时将领会的)是可能的,其可以用于多个实施例中,以高效并有效地将特征为至少一些视角变形的数字图像重建成免除任何这样的视角变形的经修正的数字图像,其中经修正的图像的特征为像素位置错误大约为5像素或更小。

[0246] 多个实施例可以附加地和/或可替代地包括利用前述的数据、计算、结果和/或概念来获取进一步的关于所捕获图像、对象等的有用信息。例如，在多个实施例中，确定所捕获的对象和捕获设备、捕获设备的俯仰角和/或翻滚角等之间的距离是可能的，如本领域普通技术人员在阅读本说明书时将理解的。

[0247] 在估计出(X,Y)值之后，方程式1中的表达可以描述如下：

$$[0248] \quad \lambda = h_{31}X + h_{32}Y + h_{33} \quad (9)$$

[0249] 因此，在一个实施例中，焦深，也称为3D(即“参考”或“真实世界”)坐标系统和捕获设备中每个点(X,Y,Z)之间的距离，可以利用上面的方程式9计算。

[0250] 对象的旋转矩阵的确定

[0251] 在估计出3D对象的位置以后，在所捕获的图像中我们有每个像素的(X,Y)和 $\lambda$ 。注意到(X,Y)是世界坐标系的坐标，而 $\lambda$ 是在摄像头坐标系中到点(X,Y)的距离。如果假设3D对象为刚性物体，我们将提出一种算法以估计从世界坐标系到摄像头坐标系的旋转矩阵。下面的方程式为旋转和转化点(X,Y,0)而提出：

$$[0252] \quad \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = R \begin{pmatrix} X \\ Y \\ 0 \end{pmatrix} + t \quad (10)$$

[0253] 其中，(X<sub>c</sub>, Y<sub>c</sub>, Z<sub>c</sub>)相对摄像头坐标系的坐标，其通过用旋转矩阵R和转化向量t旋转世界坐标系中的点(X,Y,Z)而得到，其中t是独立于(X,Y)的常量。注意到，Z<sub>c</sub>的值与 $\lambda$ 的值相同，如之前利用方程式9所估计的。

[0254] 考虑到单应性矩阵H与固有的摄像头参数矩阵A和r<sub>1</sub>、r<sub>2</sub>的相互关系，其中r<sub>1</sub>、r<sub>2</sub>分别是第一和第二列向量，揭示了下面的相互关系：

$$[0255] \quad H = \sigma A(r_{1,2},) \quad (11)$$

[0256] 其中 $\sigma$ 是常量且A是固有摄像头参数矩阵，限定为：

$$[0257] \quad A = \begin{pmatrix} a & c & d \\ & b & e \\ & & 1 \end{pmatrix} \quad (12)$$

[0258] 其中a和b是缩放因子，其包括摄像头的焦距信息， $a = f/d_x$ ，并且 $b = f/d_y$ ，其中d<sub>x</sub>、d<sub>y</sub>是图像的缩放因子，c是关于两个图像轴的倾斜参数，以及(d,e)是相应的主要点的坐标。

[0259] 在估计单应性矩阵H之后，矩阵A可以如下估计：

$$[0260] \quad a = \sqrt{w/B_{11}}; \quad (12.1)$$

$$[0261] \quad b = \sqrt{wB_{11}(B_{11}B_{22} - B_{12}^2)}; \quad (12.2)$$

$$[0262] \quad c = -B_{12}a^2b/w; d = \frac{vv_0}{b} - B_{13}a^2/w; \quad (12.3)$$

$$[0263] \quad v = -B_{12}a^2b/w; \quad (12.4)$$

$$[0264] \quad e = (B_{12}B_{13} - B_{11}B_{23})/(B_{11}B_{22} - B_{12}^2); \quad (12.5)$$

$$[0265] \quad w = B_{33} - (B_{13}^2 + e(B_{12}B_{13} - B_{11}B_{23}))/B_{11}. \quad (12.6)$$

[0266] 在上面的相互关系中，未知参数是B<sub>ij</sub>。这些值通过下面的方程式估计：

$$[0267] \quad \begin{pmatrix} v_{12}^t \\ (v_{11} - v_{22})^t \end{pmatrix} G = 0, \quad (12.7)$$

[0268] 其中,G是上述方程式的答案,可替代地表达为:

$$[0269] \quad G = (B_{11,12,22,13,23,33})^t, \quad (12.8)$$

$$[0270] \quad \text{其中, } v_{ij} = (h_{i1}h_{j1}, i_1h_{j2}+h_{i2}h_{j1}, i_2h_{j2}, i_3h_{j1}+h_{i1}h_{j3}, i_3h_{j2}+h_{i2}h_{j3}, i_3h_{j3})^t \quad (12.9)$$

[0271] 注意到,在传统的四点算法中,因为准确估计缩放因子a,b是可能的,倾斜因子c假设为0,其意味着可以忽略摄像头的倾斜变形。在一个实施例中,进一步有用的是假设d和e的值为0( $d=0, e=0$ )。

[0272] 从方程式(11), $B = (r_1 r_2 t)$ ,其中 $\sigma^{-1}A^{-1}H = B$ 。利用这个相互关系使能新的方法以由方程式 $C = (r_1 r_2 \ 0)$ 估计 $r_1 r_2$ ,其中C的第一和第二列向量是B的第一和第二列向量,并且C的第三列向量为0。

[0273] 首先,用SVD(奇异值分解)方法分解矩阵C, $C = U \Sigma V^t$ ,其中U是3乘3正交矩阵,其中,V是3乘3正交矩阵。则 $r_1$ 和 $r_2$ 由下面的方程式估计:

$$[0274] \quad (r_1 \ r_2 \ 0) = U \begin{pmatrix} W \\ 0 \end{pmatrix} \quad (13)$$

[0275] 其中W是2乘3矩阵,其第一和第二行向量分别是 $V^t$ 的第一和第二行向量。在上面的计算中,假设 $\sigma$ 是1。这个缩放因子不影响U和W的值,并且因此不影响 $r_1$ 和 $r_2$ 的估计。在估计出 $r_1$ 和 $r_2$ (例如,利用方程式13)以后,利用R是估计 $r_3$ 的旋转矩阵的事实是有用的,其是符号(1或-1)确定的 $r_1$ 和 $r_2$ 的叉积。有两个R的可能答案。在一个利用右手坐标系的例子中, $r_3$ 的值是 $r_1$ 和 $r_2$ 的叉积值。

[0276] 由旋转矩阵确定偏移角(yaw)、俯仰角(pitch)和翻滚角(roll)

[0277] 在一个实施例中,偏移角、俯仰角和翻滚角(分别用 $\alpha$ 、 $\beta$ 和 $\gamma$ 表示)也称为欧拉角,其分别限定为绕z轴、y轴和x轴的转角。根据这个方法,方程式10中的旋转矩阵R可以表示为:

$$[0278] \quad R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \quad (14)$$

[0279] 其中,每个r是矩阵R的元素。

[0280] 直接由指定的旋转矩阵R确定 $\alpha$ 、 $\beta$ 和 $\gamma$ 参数通常是便利的。在一个实施例中,翻滚角可以由下面的方程式估计(当 $r_{33}$ 不等于0时):

$$[0281] \quad \gamma = \text{atan2}(r_{32}, r_{33}) \quad (15)$$

[0282] 类似地,在另一种方法中,俯仰角可以由下面的方程式估计:

$$[0283] \quad \beta = \text{atan2}(-r_{31}, \sqrt{r_{11}^2 + r_{21}^2}) \quad (16)$$

[0284] 仍然在更多的方法中,偏移角可以由下面的方程式估计(例如,当 $r_{11}$ 是非零时):

$$[0285] \quad \alpha = \text{atan2}(r_{21}, r_{11}) \quad (17)$$

[0286] 尤其是,在一些方法中,当 $r_{11}, r_{33}$ 或 $\sqrt{r_{11}^2 + r_{21}^2}$ 的值在0附近时(例如, $0 < r_{11} < \epsilon, 0 < r_{33} < \epsilon$ ,或 $0 < \sqrt{r_{11}^2 + r_{21}^2} < \epsilon$ ,考虑到数值的稳定性,其中 $\epsilon$ 值设置为合理值,在一个实施例中,诸如 $0 < \epsilon \leq 0.01$ ,以及在尤其优选的实施例中, $\epsilon = 0.0001$ 。总之, $\epsilon$ 的值可以完全或部分基于有限的计算机字长等确定,如本领域普通技术人员在阅读本说明书时将理解的),这与旋转矩阵R的退化相对应,特殊公式用于估计偏移角、俯仰角和翻滚角的值。

[0287] 估计对象和捕获设备之间的距离

[0288] 仍然在更多的实施例中,利用诸如摄像头的固有参数(例如,焦距、图像平面上(u, v)的缩放因子)的信息估计对象和捕获设备之间的距离是可能的,无需了解对象大小。

[0289] 在一种方法中,这种算法的要是可以总结如下:1)用于所捕获的图像的摄像头的焦距可以由设备的API调用(例如,安卓设备提供了API调用以取得所捕获的图像的焦距信息)来提供和访问;2)缩放因子dx和dy用方程式12.1和12.2中的算法来估计。

[0290] 这意味着我们可以估计一种类型设备的缩放因子dx和dy,并且我们不需要分别针对每个设备估计缩放因子dx和dy。例如,在一个示例性的实施例中,使用苹果iPhone®4智能手机,利用上面呈现的算法用已知大小的对象来估计缩放因子是可能的。此后,两个缩放因子可以假设为对相同的设备类型是相同的。

[0291] 根据一个示例性的方法,估计对象到摄像头的距离的算法描述如下:下面我们标准化方程式中的(u, v), (X, Y):

$$[0292] \quad \lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = H \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \quad (18)$$

[0293] 注意到,方程式18等于方程式1,除了我们用方程式18中的(u, v)代替方程式1中的(x, y)。

[0294] 假设 $\tilde{u} = u/L_u$ ,  $\tilde{v} = v/L_v$ ;  $\tilde{x} = X/L_x$ ;  $\tilde{y} = Y/L_y$ ; 其中,  $L_u, L_v$ 是坐标u和v处的图像大小,并且 $L_x, L_y$ 是要确定的对象大小。

[0295] 则方程式18可以表达为:

$$[0296] \quad \lambda \begin{pmatrix} \tilde{u} \\ \tilde{v} \\ 1 \end{pmatrix} = \tilde{H} \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ 1 \end{pmatrix} \quad (19)$$

[0297] 其中:

$$[0298] \quad \tilde{H} = \begin{pmatrix} 1/L_u & & \\ & 1/L_v & \\ & & 1 \end{pmatrix} H \begin{pmatrix} L_x & & \\ & L_y & \\ & & 1 \end{pmatrix} \quad (20)$$

[0299] 标准化的单应性矩阵 $\tilde{H}$ 可以由方程式(20)估计。注意到,由方程式11我们有

[0300]  $H = oA(r_1 \ r_2 \ t)$  (21)并且摄像头的固有参数矩阵用下面的简单形式假设:

$$[0301] \quad A = \begin{pmatrix} f/dx & c & d \\ & f/dy & e \\ & & 1 \end{pmatrix} \quad (22)$$

[0302] 其中, f是摄像头的焦距, dx, dy是摄像头的缩放因子,其可以由从[00182]到[00209]部分呈现的算法来估计。

[0303] 由方程式(19), (20)和(21),因而:

$$[0304] \quad \sigma A \begin{pmatrix} r_1 & r_2 & t \\ L_x & L_y & 1 \end{pmatrix} = \tilde{H} \quad (23)$$

$$[0305] \quad \text{其中 } \tilde{H} = \begin{pmatrix} L_u & & \\ & L_v & \\ & & 1 \end{pmatrix} \tilde{H}$$

[0306] 因为A是已知的,由方程式(23)我们有:

$$[0307] \quad \sigma \begin{pmatrix} r_1 & r_2 & t \\ L_x & L_y & 1 \end{pmatrix} = A^{-1} \tilde{H} \quad (24)$$

[0308] 表示  $K = A^{-1} \tilde{H}$ ,  $K = (k_1, k_2, k_3)$ , 由方程式(24)我们有:

$$[0309] \quad \sigma r_1 L_x = k_1 \quad (25)$$

$$[0310] \quad \sigma r_2 L_y = k_2 \quad (26)$$

$$[0311] \quad \sigma t = k_3 \quad (27)$$

[0312] 因为方程式(27)中的t是对象相对于摄像头的变换向量。t的L2范数(欧几里得范数)如下:

$$[0313] \quad ||t|| = ||k_3|| / \sigma \quad (28)$$

[0314] 是对象的左顶角到摄像头的距离。

[0315] 因为  $||r_1|| = ||r_2|| = 1$ , 由方程式(8)和(9), 我们有

$$[0316] \quad L_x = ||k_1|| / \sigma \quad (29)$$

$$[0317] \quad L_y = ||k_2|| / \sigma \quad (30)$$

[0318] 方程式(29)和(30)可以用于估计沿X和Y坐标方向文件的大小。利用这种方法, 缩放因子可以依然未知。

[0319] 注意到, 上面描述的估计旋转矩阵的算法不需要缩放因子 $\sigma$ 。而是, 在一些方法中, 适合于假设 $\sigma=1$ 。我们可以用上面呈现的算法估计翻滚角、俯仰角和偏移角。由方程式(29)和(30), 我们也可以估计对象的纵横比如下:

$$[0320] \quad \text{aspectratio} = L_x / L_y = ||k_1|| / ||k_2|| \quad (31)$$

[0321] 由假设的矩形估计俯仰角和翻滚角

[0322] 实际上, 最普通的情况是矩形文件的摄像头捕获, 诸如标准大小的纸张、名片、驾驶证和其他证件等。因为摄像头的焦距不变化, 并且因为偏移角的知识与已讨论类型的文件图像处理不相干, 为了矩形化文件的相应图像, 仅需要确定摄像头相对于文件平面的翻滚角和俯仰角。

[0323] 算法的理念简单地是对于任何相关的俯仰角-翻滚角组合, 可以计算对应于图片中找到的四角形的文件的对象坐标(直到缩放、旋转和偏移)。当使用俯仰和翻滚的修正值时, 这个对象坐标中的所计算的四角形的特征是90度角, 并且偏移的特征是四个角差异的平方的总和。这个标准是有用的, 因为其缓和并有效处罚了个别的大偏离。

[0324] 基于这个标准的梯度下降程序可以在大约毫秒以内找到好的俯仰角-翻滚角对。这已经在图片中的四角形被正确确定的实例中被实验证实。这个方法所用的偏移角等于0

并且到对象的距离是任意固定值,因为这些值的变化仅增加对象坐标的附加的正交变换。该方法在坐标变换的计算中也使用已知的摄像头焦距,但如果所有四个角都已经找到并且有三个独立的角,则可以使用相同的标准和稍微更加复杂的梯度下降程序来估计除俯仰和翻滚以外的焦距—当进来的图片具有或不具有任何关于它们使用的什么摄像头的信息时,这对于基于服务器的处理可以是有用的。

[0325] 有趣的是,当页面检测错误时,即使最佳的俯仰角-翻滚角对也会留下相当大的残余角错误(1度或更大),或至少如果页面仅仅平行于自身被裁剪,从找到的对象坐标得到的纵横比与真正的纵横比不匹配。

[0326] 此外,即使当文件的所检测的边中的一个的位置是可疑的或完全丢失(例如,文件的那条边被部分或完全阻隔,没有描绘,或模糊得无法识别等),应用该算法也是可能的。在梯度下降程序中,为了达到期望的结果,修改上面限定的标准以仅使用两个角例如邻近边的那些,是有用的。以这种方式,该算法仍可以被用于由具有假的和/或不可检测的左顶角和右顶角的图片四角形来估计俯仰角和翻滚角。

[0327] 在一个示例中,左边和右边上接近图像框架顶部的任意点都可以指定为左顶角和右顶角。最佳估计的俯仰角-翻滚角将在对象坐标中创建同等假的左顶角和右顶角,但该文件将仍然被正确矩形化。文件的丢失的(例如,顶部的)边的方向可以被重建,因为它应该大致平行于其对侧的(例如,底部的)边,并且与邻近的(例如,左边的和/或右边的)边正交。

[0328] 其余问题是在作为整体的图像的背景中的何处放置丢失的边,并且如果纵横比已知,则丢失的边的补偿可以很好地估计,以及如果未知,则可以推到框架的边缘,只是不要丢掉任何数据。当图片沿其一条边仅包含部分文件,例如包含存款单的发票的底部,该算法的这种变化可以解决重要的用户情况。在这种形势下,文件的底边、左边和右边可以正确确定并且用于估计俯仰角和翻滚角;这些角和焦距一起可以用于矩形化文件的可见部分。

[0329] 因而,在图7所描绘的方法700所例示的一个一般方法中,在操作702,重建包括捕获或接收包括对象的数字表示的数字图像,优选地具有已知特征(大小、纹理、颜色配置等)的文件或其他对象。

[0330] 在操作704,分析所捕获或接收的图像,以确定将对象的数字表示和图像背景或表示在图像中的其他对象分离的一个或多个边界的位置。

[0331] 在操作706,分析边界,以确定在一个或多个的边界中是否存在任何弯曲(即非线性的区域)。可以确定在一个边界中存在弯曲,例如通过确定多项式表达式,其描述边界的特征适于特定类别的函数,诸如一级多项式、二级多项式、三级多项式、四级多项式等,其每个都可以是不同类别的函数。如果确定存在弯曲,优选地被修正以生成沿整个边界长度具有大致线性特征的边界。

[0332] 一旦任何经确定的弯曲被修正,在操作708,分析和/或推算线性边缘的边界,以限定界定的多边形,优选地为界定的四角形,并且更加优选地是界定的平行四边形、梯形或矩形。

[0333] 在操作710,分析数字图像和/或界定的多边形,以确定在界定的多边形内是否呈现了任何视角变形和/或投影效果。

[0334] 在操作712,并且响应于确定界定的多边形、数字图像等内存在视角变形和/或投影效果,修正视角变形和/或投影效果以生成经重建的多边形。优选地,界定的多边形是四

边形并且经重建的多边形是矩形。

[0335] 当然,示范性方法700的前述公开可以具体化为配置为执行逻辑的系统 and/或配置为实施大致类似于任何本文中描述的那些的包括计算机可读程序代码的计算机程序产品。

[0336] 类似地,本文中讨论的所有发明性概念、特征、技术、部件、系统、产品等应视为模块化的,并且可以以任何合适的方式组合,其将由本领域普通技术人员在阅读这些说明时领会。尤其是,这些组合包括上面公开的任何可能的特征集合,唯一除了技术人员在阅读这些说明时将认识到是相互排斥或不相容的那些特征。因而,本发明性概念的多个实施例可以通过引用上面公开的任何数目的特征进行限定,包括或排除任何个别特征。

[0337] 例如,在一个方法中,重建数字图像的方法包括接收包括用四角形界定的对象的数字表示的数字图像;修正四角形中的弯曲以形成四边形;以及修正四边形中的投影效果以形成矩形。四边形可以是平行四边形。修正弯曲可以包括至少部分基于固有四角形坐标转换四角形的一个或多个部分;并且修正投影效果可以包括利用四点算法转换四角形。尤其是,修正投影效果可以包括一个或多个的奇异值分解(SVD)、直接线性变换(DLT)、非线性细化(NLR),标准化数字图像和经重建的图像之间的点对应的坐标,和/或估计单应性变换(H)。估计H可以包括检测具有 $p_i = (x_i, y_i, 1)^T$ 的一个或多个点对应 $p_i \leftrightarrow P_i'$ ,和/或检测四个点对应,每个点对应 $p_i \leftrightarrow P_i'$ 与四角形的角位置 $p_i$ 和矩形的各自的角位置 $P_i'$ 相对应。数字图像可选地特征为偏离关于对象法线大约30度或更多的捕获角。H首选地满足表达式 $\lambda p_i = H P_i'$ ,其中 $\lambda$ 是三维参考坐标位置(X,Y,Z)和捕获设备的相应的参考坐标位置之间的焦距; $p_i = (x_i, y_i, 1)^T$ ;  $P_i' = (X_i, Y_i, 1)^T$ ,并且 $i$ 是相应的图像中点 $p$ 或 $P'$ 各自的索引特征。该方法还可以包括输出经重建的数字图像,其中经重建的数字图像的特征为相对于用垂直于对象的捕获角捕获的对象的扫描的500-DPI图像,像素位置错误大约5像素或更少。四角形的边的特征可以是多项式,其可以是二级多项式或比二级多项式更高级的多项式。最优选地,对象是特征为已知高宽比的文件。

[0338] 示例性的系统可以包括一个或多个处理器、图像捕获设备,以及相应的功能逻辑,其配置为实施紧接在前的段落中详细阐述的任何一个或多个的操作,并且也配置为利用任何数目的特定特征,诸如对象是具有矩形形状和已知高宽比、 $\lambda$ 、 $p_i$ 、 $P_i'$ 等的文件,如本领域普通技术人员在阅读本说明时将理解的。

[0339] 示范性的计算机程序产品可以包括计算机可读存储介质,其具有嵌入其上的计算机可读程序代码,以实施紧接在前的段落中详细阐述的任何一个或多个的操作,以及也利用任何数目的特定特征,诸如对象是具有矩形形状和已知高宽比、 $\lambda$ 、 $p_i$ 、 $P_i'$ 等的文件,如本领域普通技术人员在阅读本说明时将理解的。

[0340] 虽然上面描述了多个实施例,应理解,它们仅以示例的方式呈现,并且不是限制。因而,本发明的实施例的宽度和范围不应被任何上述的示范性实施例限制,而是仅依照下面的权利要求和它们的等价物进行限制。

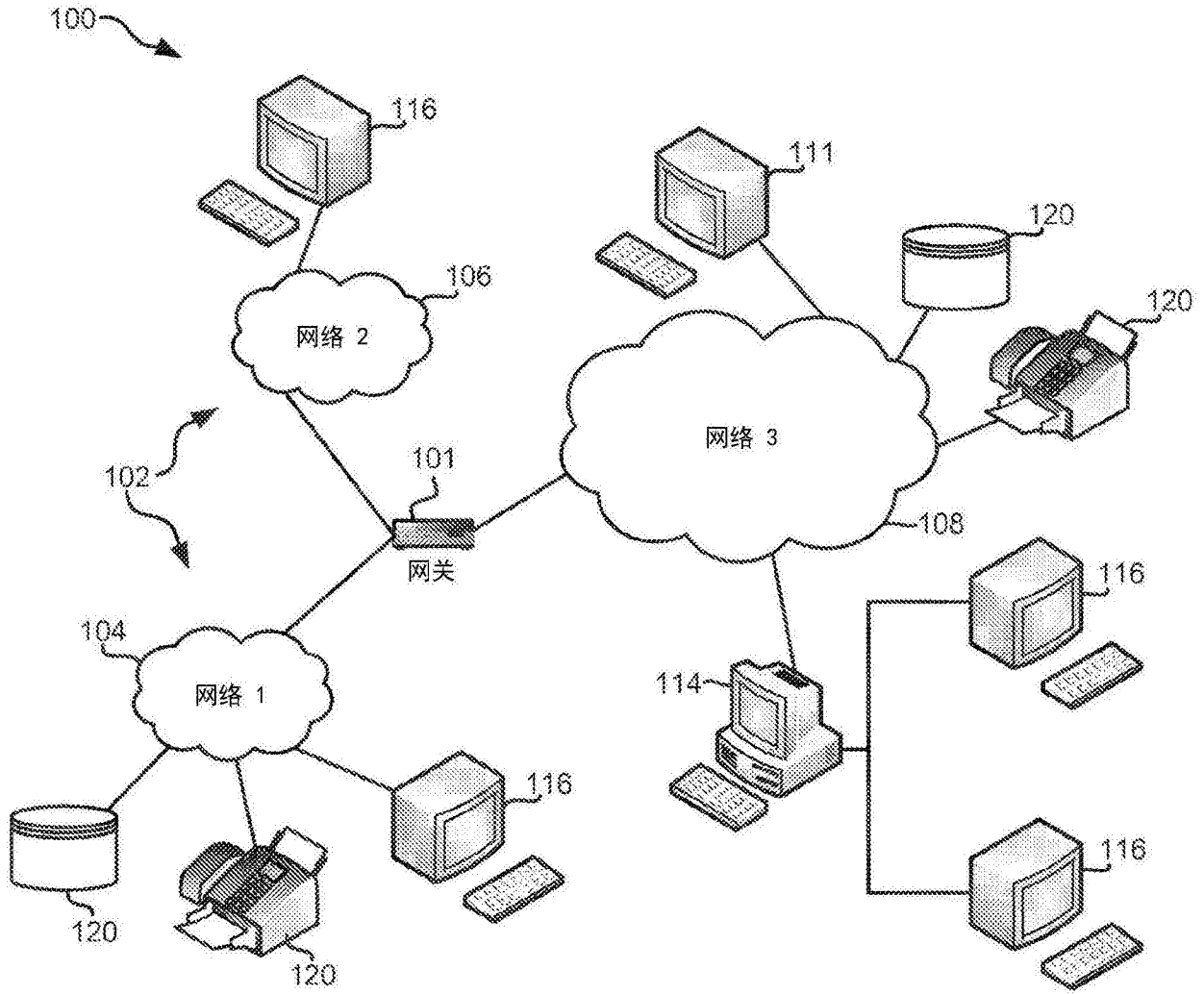


图1

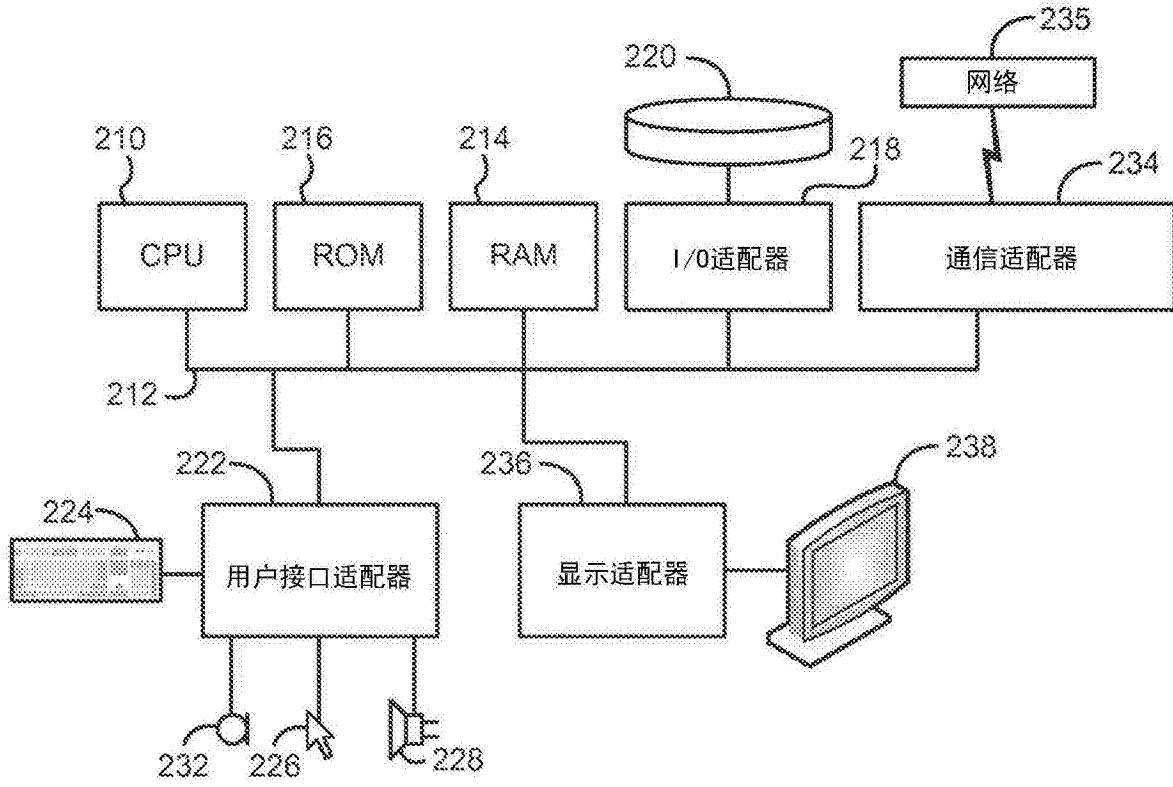


图2

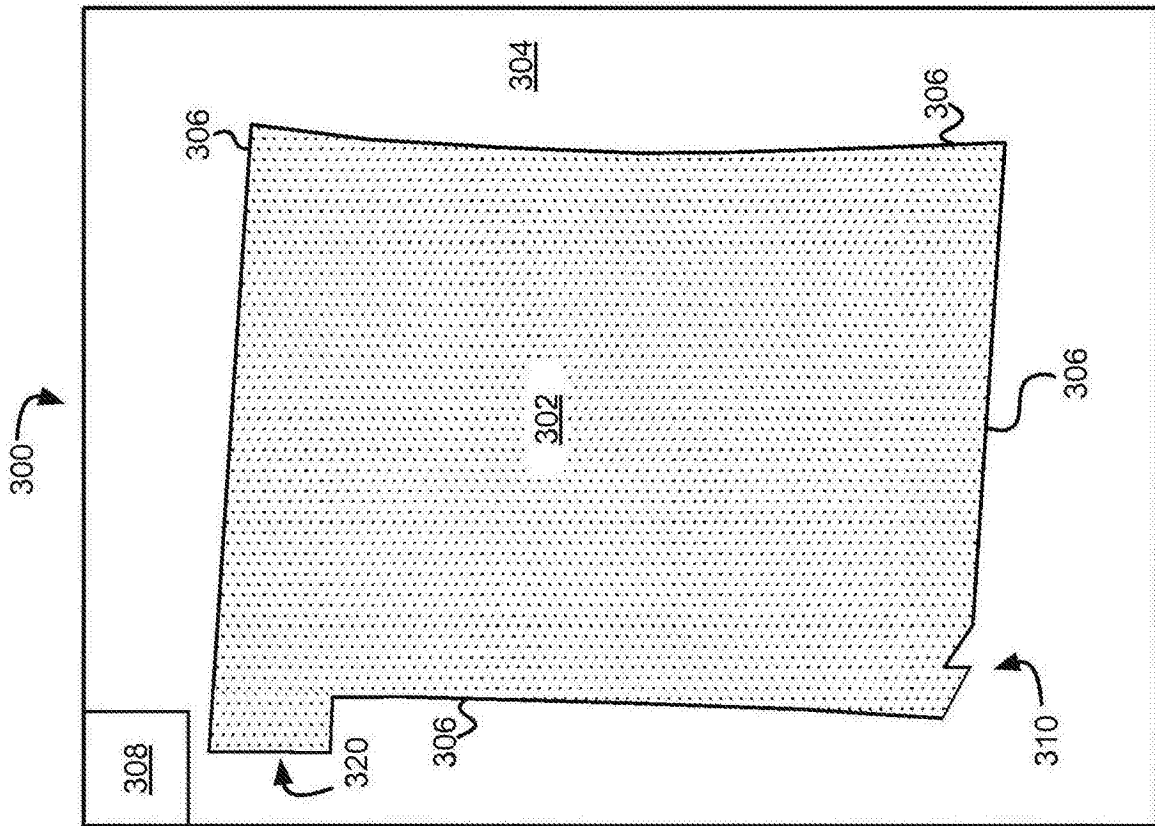


图3A

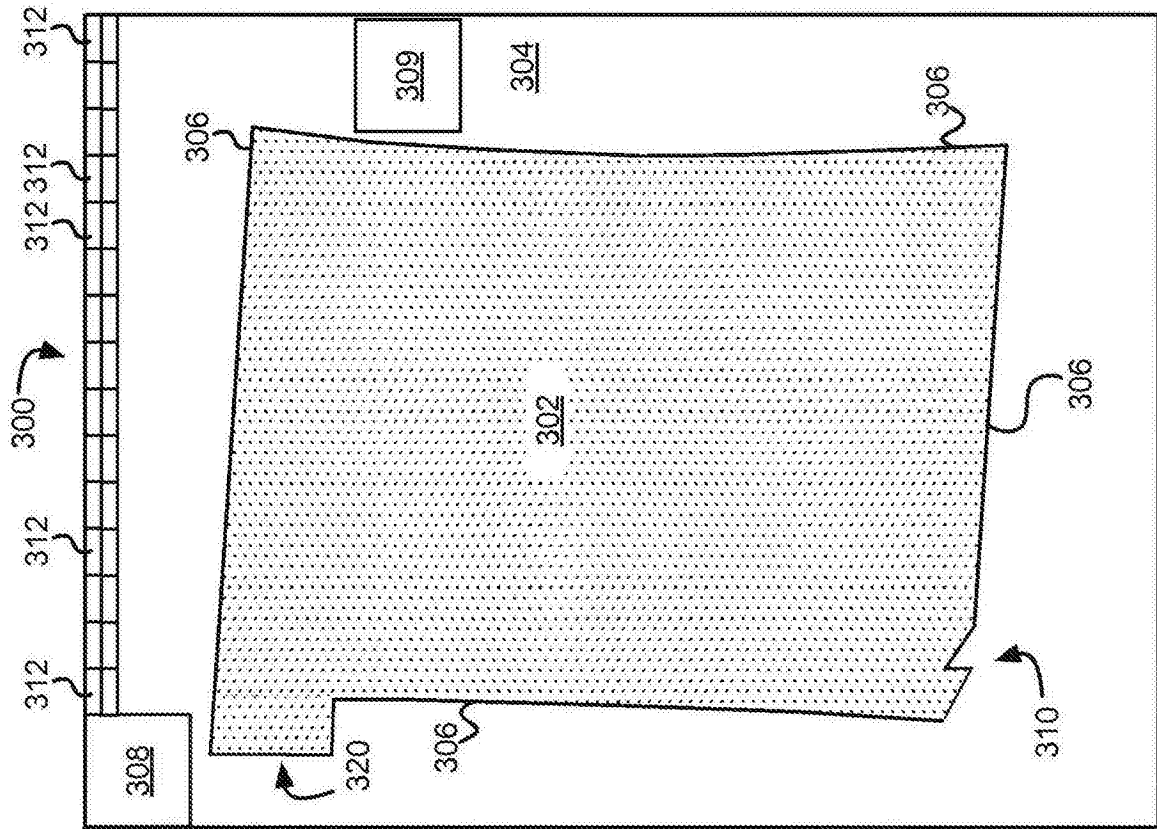


图3B

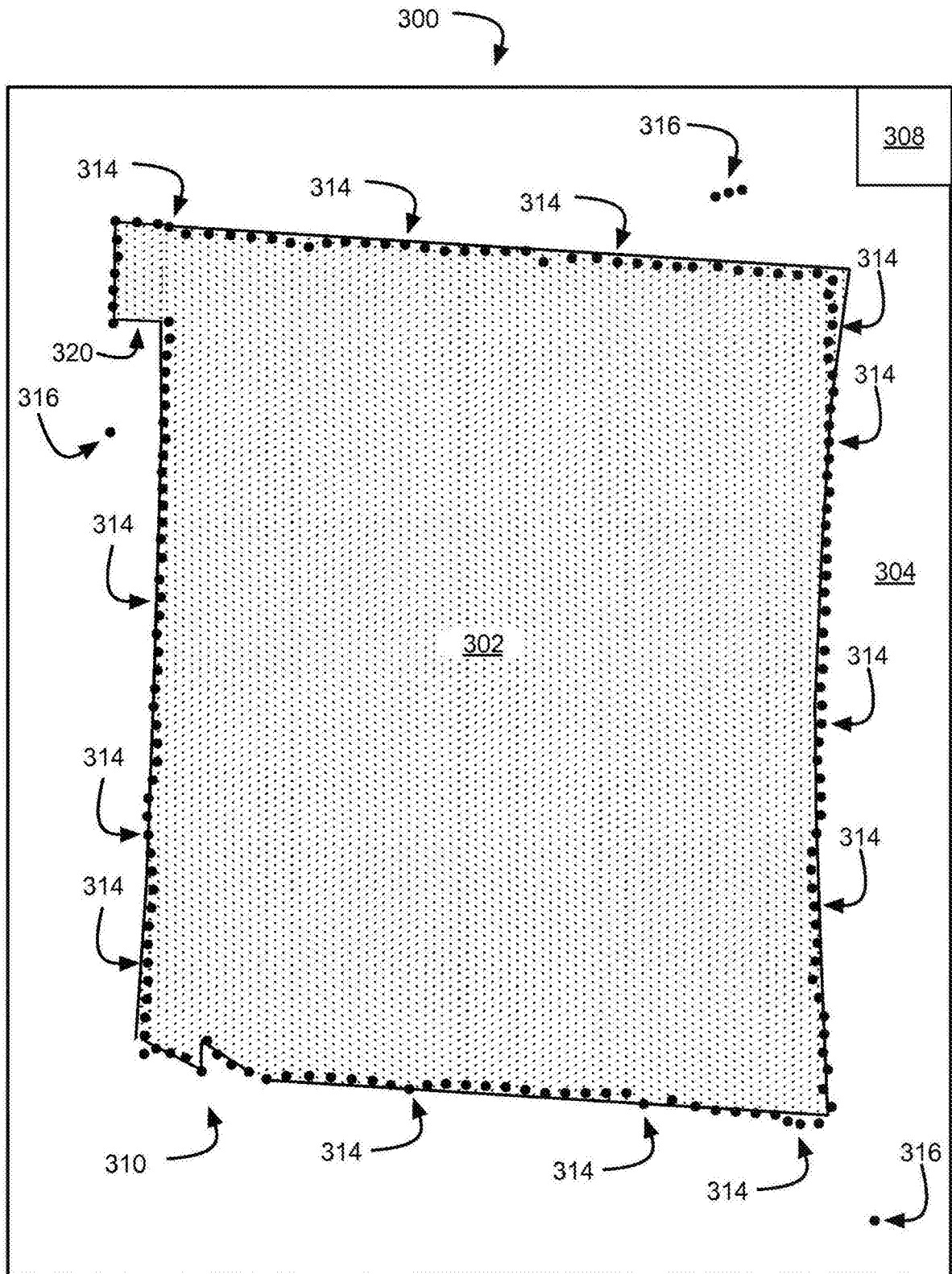


图3C

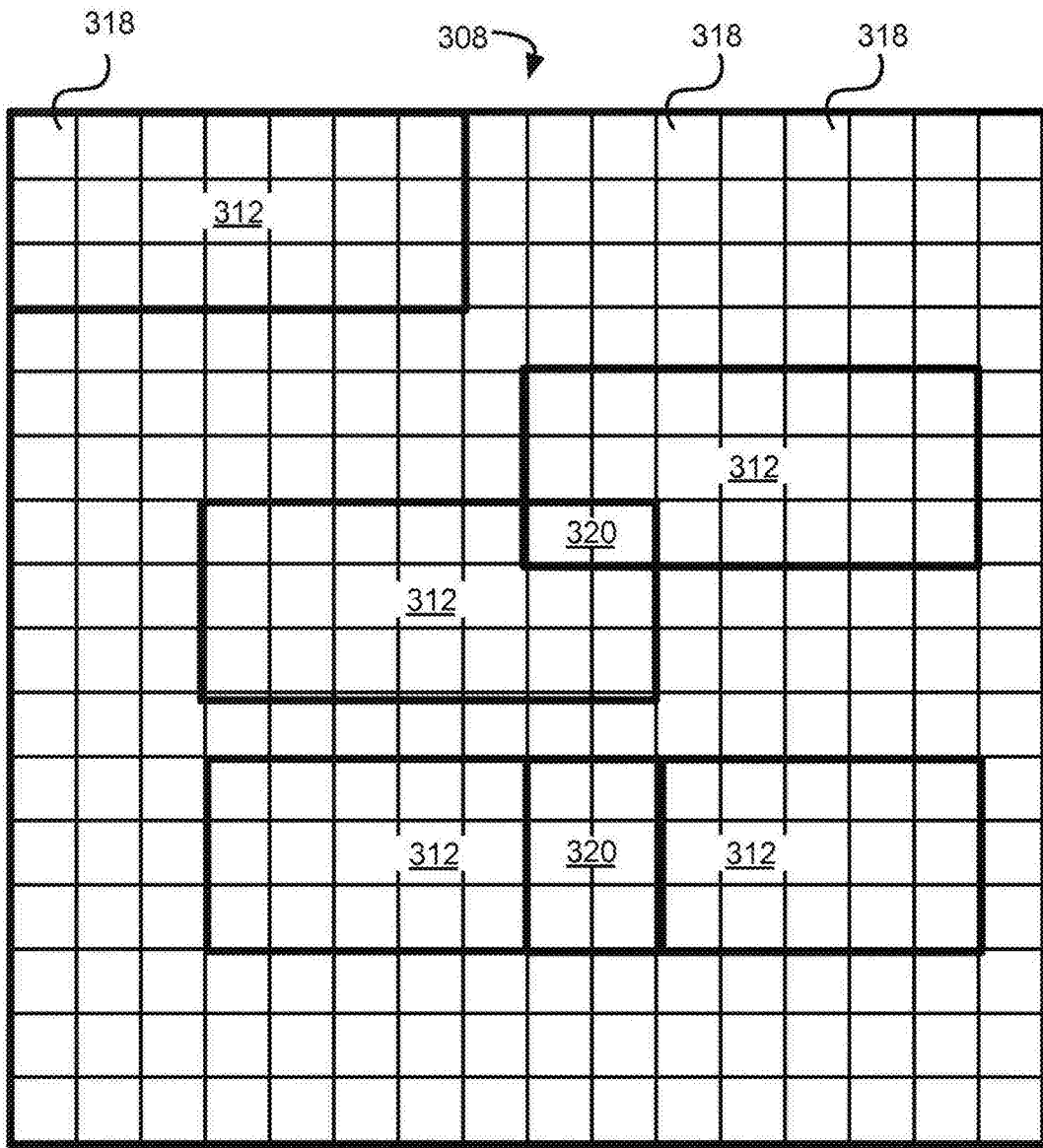


图3D

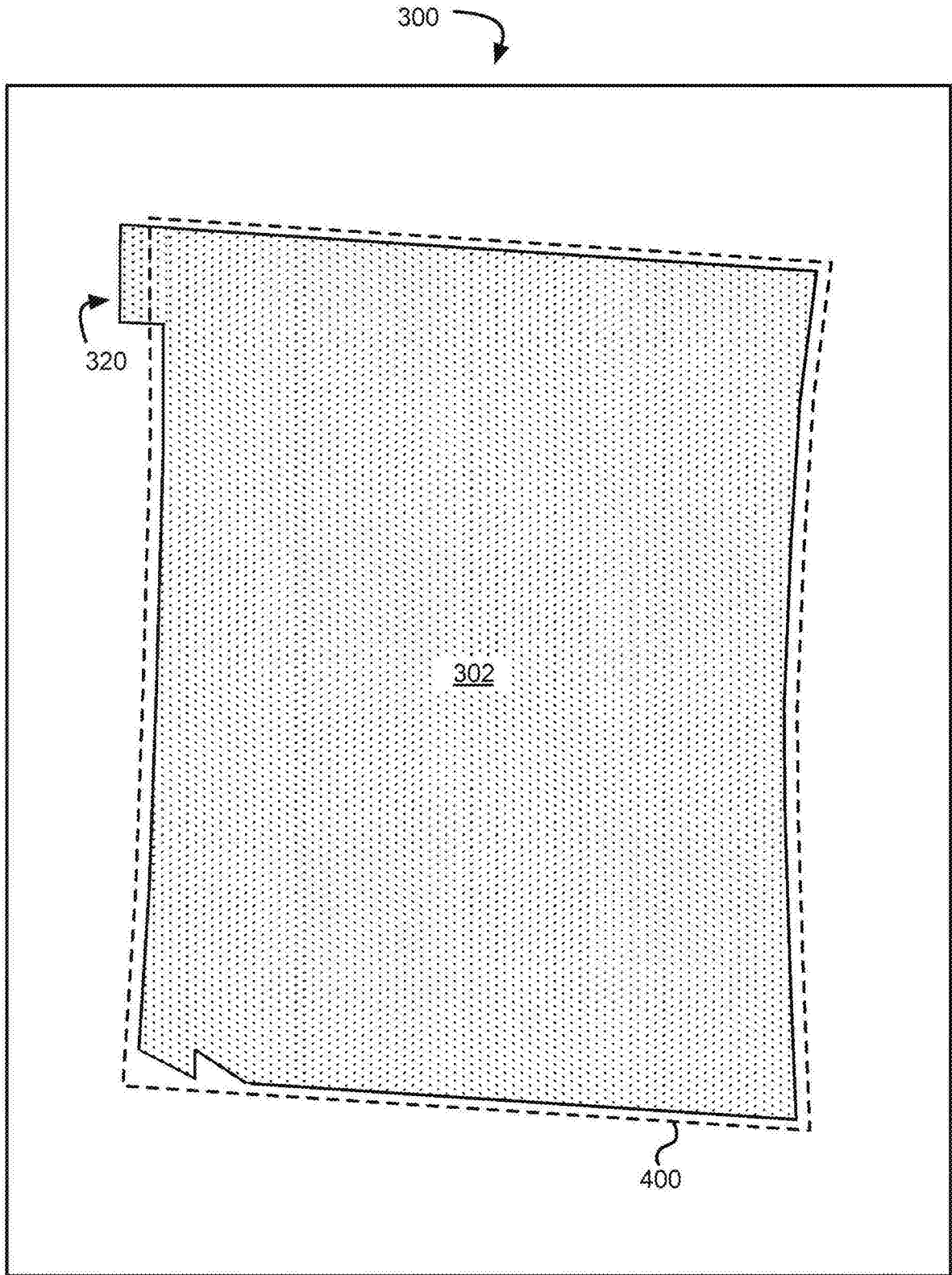


图4

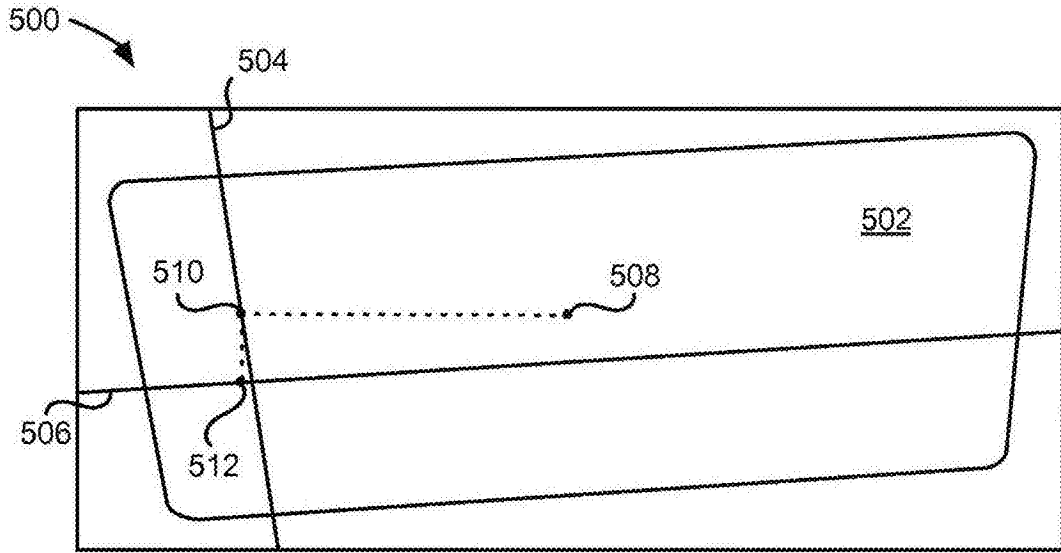


图5A

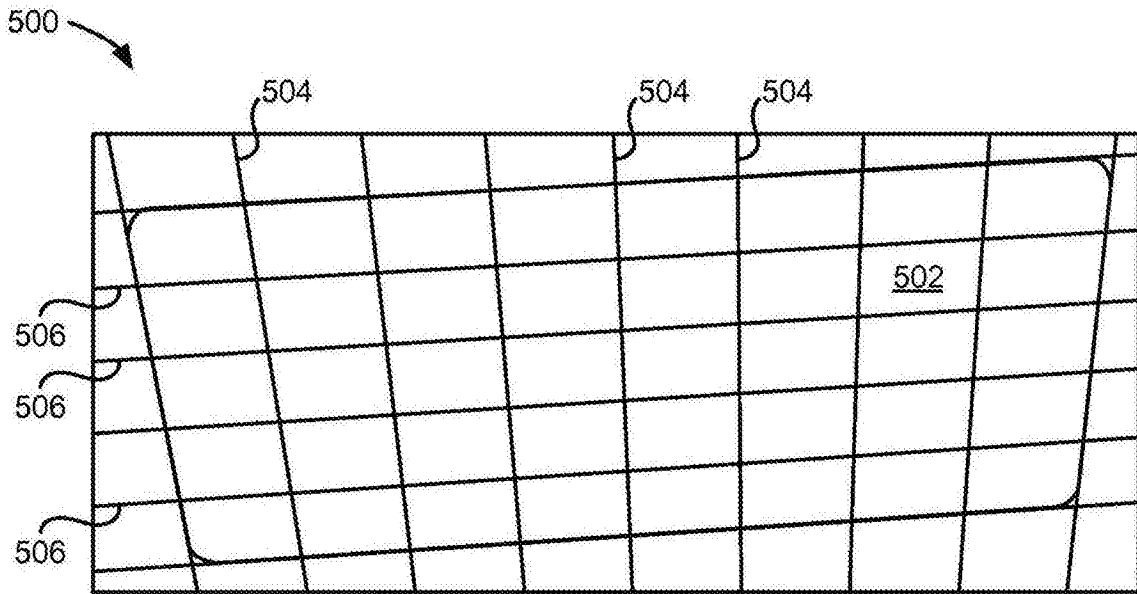


图5B

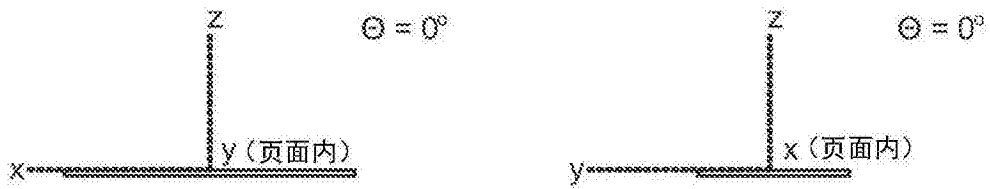


图6A

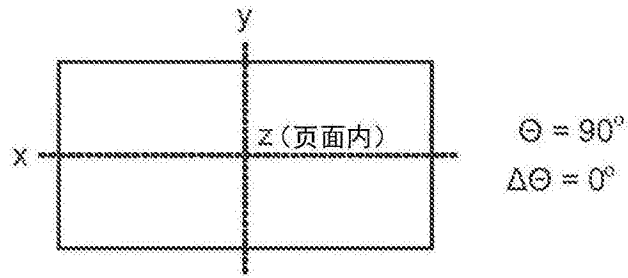


图6B

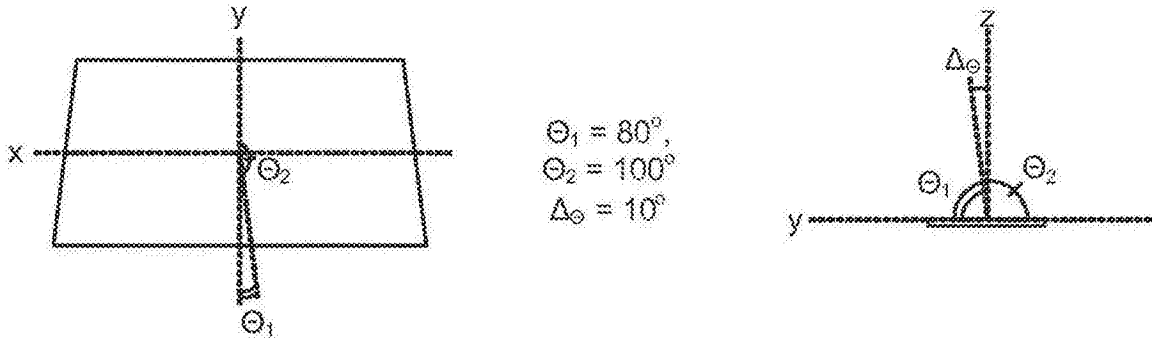


图6C

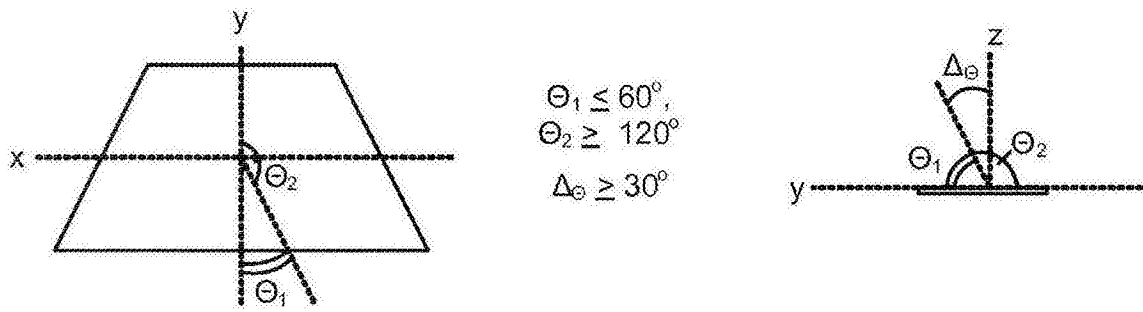


图6D

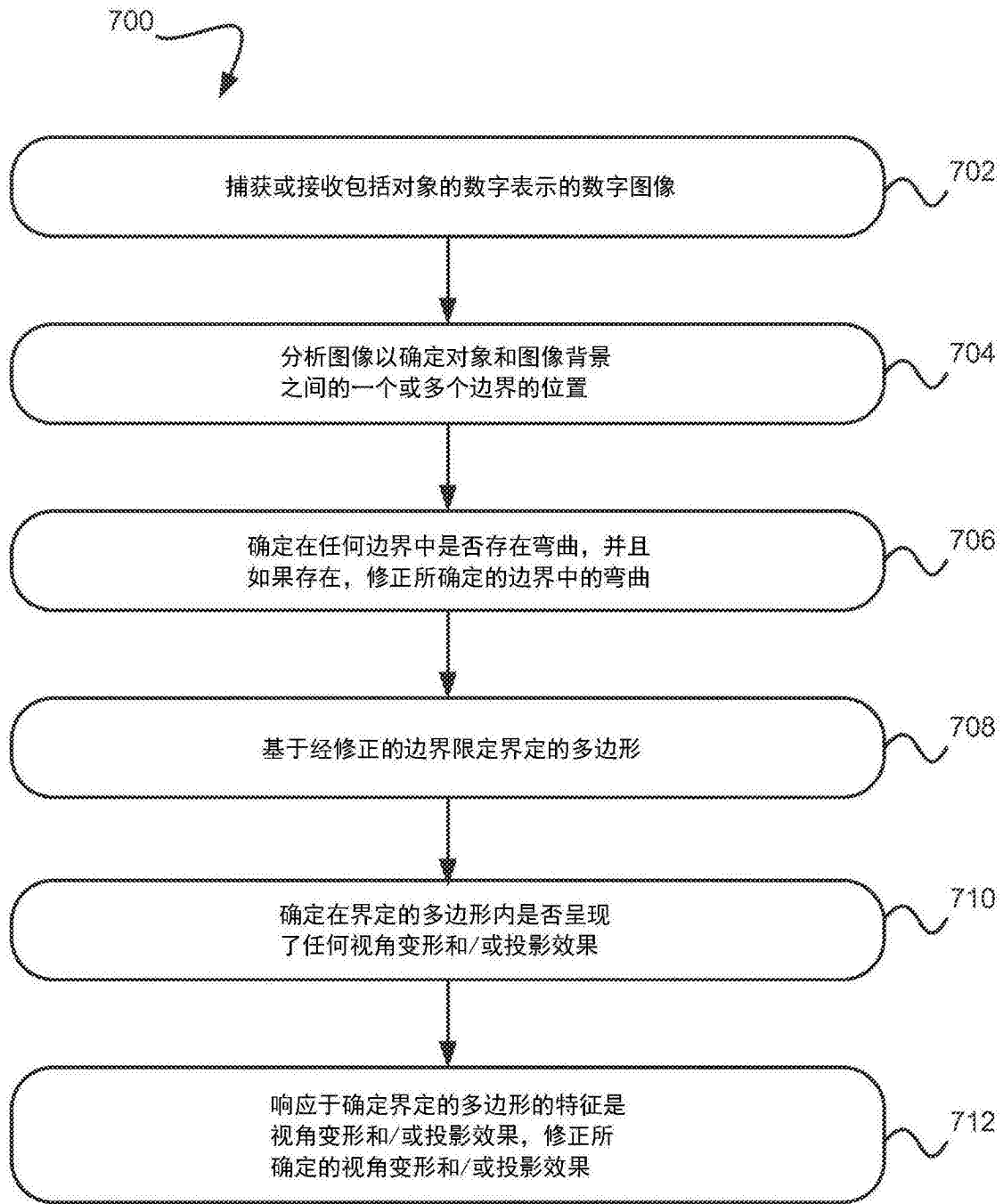


图7