



(12) 发明专利

(10) 授权公告号 CN 102782706 B

(45) 授权公告日 2014. 07. 23

(21) 申请号 201180013195. 8

G06K 7/10 (2006. 01)

(22) 申请日 2011. 03. 07

(56) 对比文件

(30) 优先权数据

12/720732 2010. 03. 10 US

KR 10-1997-0002420 B1, 1997. 03. 05, 全文.

(85) PCT国际申请进入国家阶段日

2012. 09. 10

KR 10-0480024 B1, 2005. 08. 01, 全文.

JP 2008-187327 A, 2008. 08. 14, 全文.

(86) PCT国际申请的申请数据

PCT/US2011/027439 2011. 03. 07

审查员 刘莹莹

(87) PCT国际申请的公布数据

W02011/112522 EN 2011. 09. 15

(73) 专利权人 微软公司

地址 美国华盛顿州

(72) 发明人 S. 加利奇 D. 尼耶姆塞维奇

B. 德雷塞维奇

(74) 专利代理机构 中国专利代理(香港)有限公司
72001

代理人 李舒 汪扬

(51) Int. Cl.

G06K 9/20 (2006. 01)

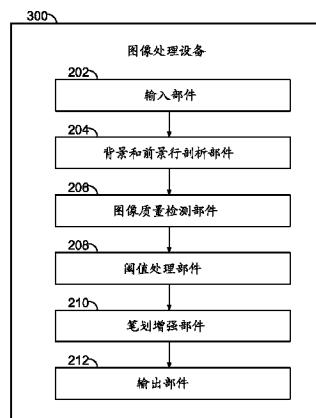
权利要求书2页 说明书12页 附图14页

(54) 发明名称

经历光学字符识别的文本图像的文本增强

(57) 摘要

用于增强文本图像以便经历光学字符识别的方法通过接收包括文本的自然行的图像开始。确定代表沿图像中的自然行的平均背景强度的背景行剖析。同样地,确定代表沿图像中的自然行的平均前景背景强度的前景行剖析。在图像中的像素至少部分根据背景行剖析和前景行剖析被分配给图像的背景或前景部分。被指定给图像的背景部分的像素的强度被调节到最大亮度,以便代表不包括文本的图像的部分。



1. 一种用于增强文本图像以便经历光学字符识别的方法,包括:
接收包括文本的自然行的图像;
确定代表沿图像中的自然行的平均背景强度的背景行剖析;
确定代表沿图像中的自然行的平均前景背景强度的前景行剖析;
至少部分根据背景行剖析和前景行剖析,把图像中的像素分配给图像的背景或前景部分;以及
把被指定给图像的背景部分的像素的强度调节到规定的强度,该规定的强度代表不包括文本的该图像的一部分;
其中分配像素还包括:
把图像中具有小于所选择参数的强度的像素分配给图像的前景部分,其中所选择参数是至少部分从背景行剖析和前景行剖析确定的;
把图像中具有大于所选择参数的强度的像素分配给图像的背景部分;以及
把具有在所选择参数与背景行剖析之间的强度的被选择的像素分配给在边缘空间中表示的自然行的基本已连成分。
2. 权利要求 1 的方法,其中所选择参数是背景行剖析和前景行剖析的平均。
3. 权利要求 1 的方法,其中所选择参数是可调节的。
4. 权利要求 1 的方法,还包括估计代表图像中的文本的已连成分的平均或中值笔划宽度。
5. 权利要求 1 的方法,还包括在边缘空间中表示文本的自然行。
6. 权利要求 5 的方法,还包括检测在边缘空间中表示的自然行中的已连成分。
7. 一种用于增强文本图像以便经历光学字符识别的方法,包括:
接收包括文本的自然行的图像;
确定代表沿图像中的自然行的平均背景强度的背景行剖析;
确定代表沿图像中的自然行的平均前景背景强度的前景行剖析;
估计代表在图像中的文本的已连成分的平均或中值笔划宽度;以及
如果平均或中值笔划宽度小于阈值,则增强文本质量,其中所述增强包括:
把图像中具有小于所选择参数的强度的像素分配给图像的前景部分,其中所选择参数是至少部分从背景行剖析和前景行剖析确定的;
把图像中具有大于背景行剖析的强度的像素分配给图像的背景部分;以及
把具有在所选择参数与背景行剖析之间的强度的被选择的像素分配给在边缘空间中表示的自然行的基本已连成分。
8. 权利要求 7 的方法,还包括通过确定对于被分配给前景强度值的每个像素的局部笔划宽度而估计代表在图像中的文本的已连成分的平均或中值笔划宽度。
9. 权利要求 7 的方法,其中增强文本质量包括执行图像内插,以便提高图像分辨率。
10. 权利要求 9 的方法,其中通过递增地添加在越来越高的强度内的像素的集合,而执行具有在所选择参数与背景行剖析之间的强度的被选择的像素的分配,其中每个集合限定具有在强度区间的有限集中的一个强度区间内的强度的像素。
11. 权利要求 10 的方法,还包括当在边缘像素与基本已连成分的骨架像素之间的距离大于平均或中值笔划宽度的一半时,终止将所选择的像素分配到基本已连成分的至少一

个。

12. 一种增强经历光学字符识别的文本图像的系统,包括:

输入部件,用于接收具有文本的自然行的输入图像;

背景和前景行剖析部件,用于确定分别代表图像中的平均背景和前景强度的背景和前景行剖析;

图像质量检测部件,用于确定图像的质量是否足以把像素分配给图像的背景或前景部分;

阈值处理部件,用于根据背景和前景行剖析把像素分配给图像的背景或前景部分;以及

笔划增强部件,用于如果图像质量检测部件确定图像的质量低于阈值,则增强代表图像中的文本的已连成分的笔划,其中所述增强包括:把图像中具有小于所选择参数的强度的像素分配给图像的前景部分,其中所选择参数是至少部分从背景行剖析和前景行剖析确定的;把图像中具有大于背景行剖析的强度的像素分配给图像的背景部分;以及把具有在所选择参数与背景行剖析之间的强度的被选择的像素分配给在边缘空间中表示的自然行的基本已连成分。

13. 权利要求 12 的系统,其中图像质量检测部件通过估计代表图像中的文本的已连成分的平均或中值笔划宽度并将所述平均或中值笔划宽度与所述阈值进行比较来确定图像的质量。

14. 权利要求 12 的系统,其中背景和前景行剖析部件还被配置成对于沿在边缘空间中表示的自然行中的每个自然行的多个局部化区域确定多个背景和前景强度值,其中每个局部化的区域具有小于或等于预定义值的尺度。

经历光学字符识别的文本图像的文本增强

背景技术

[0001] 光学字符识别(OCR)是文本的图像到如机器可编辑文本的数字形式的基于计算机的转化,通常是按照标准编码方案的。这个过程消除了对将文档人工打字输入到计算机系统的需要。由于图像质量差、由扫描过程引起的缺陷等等,会引发许多不同的问题。例如,常规的OCR引擎可以被耦合到用来扫描文本的页的平板扫描仪。因为页被相对扫描仪的扫描面齐平地放置,所以由扫描仪生成的图像典型地呈现均匀的对比度和照度、减小的扭曲和失真、以及高分辨率。因此,OCR引擎可以容易地把图像中的文本转化成机器可编辑的文本。然而,当图像在对比度、照度、扭曲等等方面具有较差的质量时,OCR引擎的性能可能会恶化,并且处理时间可能会由于对图像中的所有像素进行处理而增加。例如当图像是由基于图像的扫描仪生成时,情形可能是这样,因为文本/图片是从一定的距离、从变化的取向和在变化的照度下被扫描的。即使扫描过程的性能是好的,当质量相对较差的文本页被扫描时,OCR引擎的性能也可能恶化。

发明内容

[0002] 如果原始图像的背景被检测且被滤除,而同时加强剩余的文本像素,则OCR过程的精度可以大大地提高。在一个实现中,背景是通过计算背景和文本彩色剖析(color profile)而被检测和滤波的。文本增强本身是通过使用以高精度识别文本笔划(stroke)像素的笔划恢复技术而被执行的。

[0003] 在一个具体的实现中,用于增强文本图像以便经历光学字符识别的方法通过接收包括文本的自然行(native line)的图像开始。确定代表沿图像中的自然行的平均背景强度的背景行剖析(line profile)。同样地,确定代表沿图像中的自然行的平均前景背景强度的前景行剖析。在图像中的像素至少部分根据背景行剖析或前景行剖析被分配给图像的背景部分或前景部分。被指定给图像的背景部分的像素的强度被调节到最大亮度,以便代表没有包括文本的图像的部分。

[0004] 在再一个具体的实现中,代表图像中的文本的已连成分(connected component)的平均或中值笔划宽度在背景和前景行剖析被确定后才被确定。如果平均或中值笔划宽度小于阈值,则文本质量被增强。例如,文本质量可以通过执行图像内插以增加图像分辨率而被增强。

[0005] 在再一个具体的实现中,对于沿在边缘空间中表示的自然行中的每个自然行的多个局部化区域确定多个背景和前景强度值,其中每个局部化的区域具有小于或等于预定义值的尺度。

[0006] 本概要提供来以简化的形式介绍概念的选择,这些概念将在下面的详细描述中进一步描述。本概要既不打算识别所要求保护的主题的关键特征或必要特征,也不打算被用来帮助确定所要求保护的主题的范围。

附图说明

- [0007] 图 1 显示用于光学字符识别的系统的一个说明性例子。
- [0008] 图 2 显示不适合于处理的文本行的例子。
- [0009] 图 3 显示更适合于 OCR 处理的、在图 2 上的文本行的变换。
- [0010] 图 4 显示用来保证文本比背景更暗的图像变换。
- [0011] 图 5 显示在边缘空间中文本行表示的例子。
- [0012] 图 6 是在图 2 上的图像的灰度版本。
- [0013] 图 7 显示图 6 上的自然行的最左部分,其中在具有 $2 \times h \times h$ 的尺度的矩形上执行局部处理。
- [0014] 图 8 显示在边缘空间中表示的、在图 7 上的自然行。
- [0015] 图 9 显示围绕图 8 所示的已连成分的框,它们规定包含边缘的一个区域(深色)和不包含边缘的另一个区域(浅色)。
- [0016] 图 10a 是对于图 7 的矩形的局部化区域的强度的直方图;图 10b 是图示对于图 9 的深灰色区域的、像素相对于它们的彩色强度的分布的直方图;图 10c 是通过从图 10a 上直方图的相应值中减去图 10b 上第二直方图的值而得到的直方图。
- [0017] 图 11 显示对于图 6 的文本行的、计算的背景(上面的线)和前景(下面的线)行剖析。
- [0018] 图 12 显示把求逆运算符应用到深色背景上的浅色文本的结果。
- [0019] 图 13 显示低质量文本图像的例子。
- [0020] 图 14 显示把阈值处理规程应用到图 13 所示的图像的结果。
- [0021] 图 15 显示为箭头发源于其中的单个前景像素确定局部笔划宽度的方式。
- [0022] 图 16 显示像素被映射到要被内插的图像中的一组四个像素。
- [0023] 图 17 图示在内插过程中的步骤。
- [0024] 图 18 图示把内插过程应用到图的上面部分所描绘的文本图像的结果。
- [0025] 图 19 图示将对其应用文本增强过程的单词。
- [0026] 图 20 描绘了通过把像素添加到图 20 所示的单词中的第一个字母而进行的文本增强过程。
- [0027] 图 21a 显示已连成分的例子;图 21b 显示对于图 21a 上的已连成分的边缘像素;图 21c 显示图 21a 上的已连成分的骨架(skeleton)。
- [0028] 图 22 显示把图 20 的像素添加过程应用到已连成分的骨架的结果。
- [0029] 图 23 显示把文本增强过程应用到图 13 所描绘的图像的结果。
- [0030] 图 24 显示可以对经历光学字符识别的文本图像执行文本增强的图像处理设备的一个例子。
- [0031] 图 25 是显示用于增强要经历光学字符识别的文本图像的总体方法中的一个例子的流程图。

具体实施方式

[0032] 图 1 显示用于在图像上进行光学字符识别(OCR)的系统 5 的一个说明性例子,所述系统包括用来生成文档 15 的图像的数据捕获装置(例如,扫描仪 10)。扫描仪 10 可以是基于图像的扫描仪,它利用电荷耦合器件作为图像传感器来生成图像。扫描仪 10 处理图像来

生成输入数据,并把输入数据传送到处理装置(例如,OCR引擎20),以便进行图像内的字符识别。在这个具体的例子中,OCR引擎20被合并到扫描仪10中。然而,在其它的例子中,OCR引擎20可以是单独的单元,诸如独立的单元,或者是被合并到诸如PC、服务器等等的另外的设备的单元。

[0033] 如果原始图像的背景被检测且被滤除,而同时加强剩余的文本像素,则OCR过程的精度可以大大地提高。正如下面详细说明的,背景是通过计算背景和文本彩色剖析而被检测和滤波的。文本增强本身是通过使用以高精度识别文本笔划像素的笔划恢复技术而被执行的。

[0034] 这个文本增强过程的输入是文本的自然行(见下面的对于“自然行”的定义)的图像。然而,在着手描述实际的过程之前,将介绍某些数学概念。这些概念将贯穿本讨论被普遍地使用以便便于本发明的理解,并且在一些情形下定形(formalize)某些解释说明。

[0035] 定义1:宽度 w 和高度 h 的数字彩色图像是两个自变量的向量函数 $I:W \times H \rightarrow GS^3$,其中 $GS=[0, 1, \dots, 255]$, $W=[0, 1, \dots, w-1]$, $H=[0, 1, \dots, h-1]$,以及 $\times$ 表示笛卡尔乘积。

[0036] 这个定义是从RGB彩色系统推导的,以及在 $I(r, g, b)$ 中的分量 r, g, b 分别对应于红、绿和蓝色分量。

[0037] 定义2:宽度 W 和高度 H 的数字灰度图像是两个自变量的标量函数 $I:W \times H \rightarrow GS$,其中 GS 可以是:

[0038] - $GS=[g1, g2]$,其中灰度图像被称为二元、二级或二色调图像

[0039] - $GS=[g1, g2, g3, \dots, g16]$,其中灰度图像被称为16级灰度图像

[0040] - $GS=[g1, g2, g3, \dots, g256]$,其中灰度图像被称为256级灰度图像。

[0041] 在这一点,应当介绍贯穿文档使用的一个约定。由于图像被看作为函数,所以应当定义它的图形呈现的坐标系。通常,图像的左上角被视为参考点。这里将遵循这个约定,所以,图像的左上角将按被放置在原点对待。

[0042] 定义3:三元组 $(I(x, y), x, y)$ 被称为像素。对 (x, y) 被称为像素坐标,而 $I(x, y)$ 被称为像素值。

[0043] 通常,术语“像素”可互换地被使用于坐标、值以及坐标和值两者。在这里,如果不可能引发混淆的话,术语“像素”也将这样地使用,否则将使用准确的术语。同样,当不可能混淆时,符号 $I(x, y)$ 将可互换地使用。

[0044] 先前的三个定义提供了对于数字图像是什么的理解。图像处理的通常任务是一系列变换,这些变换导致原始图像的某种呈现,该呈现对于进一步的分析更便利,使得可以得出各种结论。以下的定义为这些变换的定形提供数学手段。

[0045] 定义4:令 Ω 为具有尺度 w 和 h 的所有图像的集合。函数 $T:\Omega^n \rightarrow \Omega$ 被称为 n 元图像运算符。如果 $n=1$,运算符是一元的,而对于 $n=2$,运算符是二元的。

[0046] 先前的定义暗示,运算符是通过使用某组变换法则把一个图像(或几个图像)变换成另一个图像的函数。最重要的图像运算符是基于滤波器的运算符。滤波器(有时称为核或掩膜)是 $n \times m$ 尺寸的矩阵 A_{nm}

[0047]

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

[0048] 通常, n 和 m 是相等的并且是奇数, 从而产生例如 3×3 、 5×5 、 7×7 滤波器。基于滤波器的运算符通过使用下述法则来变换输入图像, 即: 在输出图像中的像素 $I_0(x, y)$ 通过使用以下公式进行计算:

[0049]

$$I_0(x, y) = \sum_{i=1}^n \sum_{j=1}^m a_{ij} I(x - \frac{n}{2} + i - 1, y - \frac{m}{2} + j - 1)$$

[0050] 其中所有的除法是整数除法。

[0051] 换句话说, 在输出图像中的像素是通过将输入图像中的对应像素的邻域与滤波器进行卷积而被构造的。

[0052] 定义 5: 令 I 为宽度 w 和高度 h 的图像, 以及令 $I(x, y)$ 为任意像素。像素 $\{I(x+1, y), I(x-1, y), I(x, y+1), I(x, y-1)\}$ 的集合被称为 $I(x, y)$ 的 4-邻居。相似地, 像素 $\{I(x+1, y), I(x-1, y), I(x, y+1), I(x, y-1), I(x-1, y-1), I(x-1, y+1), I(x+1, y-1), I(x+1, y+1)\}$ 的集合被称为 $I(x, y)$ 的 8-邻居。

[0053] 在文献中对于毗邻有不同的定义, 但将使用以下的定义, 因为它对于随后的讨论是最方便的。

[0054] 定义 6: 如果像素 $I(x_2, y_2)$ 是像素 $I(x_1, y_1)$ 的 8-邻居集合的成员并且它们的像素值是“相似的”, 则这两个像素 $I(x_1, y_1)$ 和 $I(x_2, y_2)$ 是毗邻的。

[0055] 单词相似的被加上引号, 是因为不存在相似性的严格定义。相反, 这个定义是按照特定应用的要求而被采用的。例如, 可以说, 如果两个像素的像素值是相同的, 则它们是相似的。贯穿本文档, 都将采用这个定义, 如果不是则另外阐述。

[0056] 定义 7: 如果存在集合 $\{I(x_2, y_2), I(x_3, y_3), \cdots, I(x_{n-1}, y_{n-1})\}$, 使得对于 $i=1, 2, \cdots, n-1$, $I(x_i, y_i)$ 和 $I(x_{i+1}, y_{i+1})$ 是毗邻的, 则两个像素 $I(x_1, y_1)$ 和 $I(x_n, y_n)$ 是已连接的。

[0057] 定义 8: 已连成分是图像像素的子集, 其中来自该集合的每个像素是与来自该集合的所有其余像素已连接的。

[0058] 为了滤波文本图像中的背景, 需要了解文本与背景之间的关系, 以便在干净的 (例如, 白色) 背景上提供灰度文本。这个要求可以参照图 1 弄明白。

[0059] 图 2 上描绘的文本行的图像不适合于连续的处理。如果它可以被变换成图 3 所示的图像, 则在 OCR 过程中随后的步骤多半会更简单和更精确。而且, 方便的是假设: 文本比背景彩色更深。如果事实上情形正是这样, 则所有的文本-背景组合可以一致地处理。因此, 当必要时, 应当执行图 4 上描绘的图像变换。

[0060] 先前的讨论应当澄清了: 为了滤波文本图像的背景, 应当确定文本-背景关系 (在彩色方面); 以及文本增强过程产生了其中深色文本呈现在白色背景上的图像。在下一节中, 描述了文本-背景关系检测。

[0061] 行剖析

[0062] 要注意的第一件事情是文本行图像具有两种类型的像素,一种源自于文本,而另一种源自于背景。另外,在文本与背景之间的边缘处有突然的彩色改变。这导致在边缘空间中可以观察到文本行的结论。通过使用各种各样熟知的边缘检测技术中的任意技术,可以得到边缘空间中的行表示。边缘空间中的行表示的例子显示于图 5。

[0063] 很明显,在图 5 上保留了所有的重要的文本特征。字母主要被表示为边缘空间中的已连成分。而且,在边缘空间中的文本特征的呈现是与文本-背景关系无关的(例如,哪个彩色更深是不相关的)。所以,可以制定以下的定义。

[0064] 定义 9:自然行是在文本行图像的边缘空间中的所有已连成分的集合。

[0065] 如果 CC_i 是第 i 个已连成分以及 nl 是自然行,则下式成立:

[0066] $nl = \{CC_1, CC_2, \dots, CC_n\}$ 。

[0067] 行剖析规程是在自然行上进行的,产生以下定义中规定的行剖析。

[0068] 定义 10. 令 nl 是自然行,以及令 $I(x, y); x \in \{0, 1, \dots, w-1\} \wedge y \in \{0, 1, \dots, h-1\}$ 是具有宽度 w 和高度 h 的对应的自然行图像。然后,如果满足以下的条件,则函数 $FLF_{nl}: W \rightarrow GS$ 和 $BLF_{nl}: W \rightarrow GS$ (其中 $W = \{0, 1, \dots, w-1\}$ 和 $GS = \{0, 1, \dots, 255\}$) 被称为前景行剖析和背景行剖析:

[0069]

$$- FLP_{nl}(x_0) = \frac{\sum_{j=0}^{h-1} I(x_0, j) |_{I(x_0, j) \in FP(x_0)}}{card(FP(x_0))}$$

$$- BLP_{nl}(x_0) = \frac{\sum_{j=0}^{h-1} I(x_0, j) |_{I(x_0, j) \in BP(x_0)}}{card(BP(x_0))}$$

[0070] 其中 $FP(x_0)$ 是来自自然行图像的、具有坐标 $x=x_0$ 的所有前景像素的集合,以及 $BP(x_0)$ 是来自自然行图像的、具有坐标 $x=x_0$ 的所有背景像素的集合。

[0071] 换句话说,剖析函数反映对于给定的图像列的平均背景或前景强度。

[0072] 在参照作为图 2 上的图像的灰度版本的图 6 以后,应当明白引入行剖析的重要性。虽然常常合理地假设:文本彩色强度和背景彩色强度沿水平轴是相当均匀的,但这个假设在图 6 的情形下是不成立的。在图像的左侧的平均背景强度约为 190,而在图像的右侧,平均背景强度约为 100 (差值大于整个灰度范围 0-255 的三分之一)。

[0073] 因此,在图 6 的情形下,跨整个图像来用单个值近似背景强度将在精度上造成很大的损失,从而阻碍随后的处理步骤得到最可信的信息。这是引入行剖析的原因。行剖析规程是用于估计行剖析的方法,且将在下面描述。

[0074] 在定义 10 中介绍的行剖析很难使用给定的公式来计算,因为在过程中的这一点

处缺乏关于前景和背景像素的知识。所以,应当执行某种估计。

[0075] 要观察的第一件事是,虽然在整个自然行上可能存在彩色的很大变化,但在相对较小的局部化区域中,彩色应当是相当恒定的。如果该局部化区域被选择为例如具有等于几倍 h 的宽度和等于 h 的高度的矩形(其中 h 是行的高度),则可以对于该小的部分执行局部处理。

[0076] 矩形的尺度被选择为遵从以下假设:在与几个字符的宽度成比例的距离上,彩色变化不能相对较大。局部处理暗示,在矩形上执行的处理产生两个值,一个用于主导的背景彩色,一个用于主导的前景彩色。这两个结果可被使用来近似在对应的矩形上的行剖析值。一个这样的矩形的例子被描绘于图 7。

[0077] 图 7 显示图 6 的自然行的最左面部分,其中所选择的矩形具有尺度 $2*h \times h$ 。如果局部处理产生分别为主导的前景和背景彩色的 df 和 db ,则可以使用以下的近似:

$$FLP_m(x) \cong df; x = \{0, 1, \dots, 2 * h\}$$

[0078]

$$BLP_m(x) \cong bf; x = \{0, 1, \dots, 2 * h\}$$

[0079] 这个近似导致对于行剖析的阶梯状(step-like)近似。为了计算 df 和 db 的值,图 7 的矩形应当在边缘空间中表示,如图 8 所描绘的。也可以使用其它近似。例如,分段线性近似可以给出更好的结果,虽然它可能需要增多的处理时间。

[0080] 图 8 上的三个内部矩形对应于已连成分的边界框。这些框限定两个不同的区域,一个区域包含边缘,另一个区域不包含边缘。这些区域被图示于图 9。在图 9 上的深灰色区域对应于边缘区域,而浅灰色区域对应于非边缘区域。

[0081] 接着,计算直方图,其举例说明对于图 7 的矩形局部化区域,像素相对于它们的彩色强度的分布(distribution)。计算另一个直方图,其举例说明仅仅对于图 9 的深灰色区域,像素相对于它们的彩色强度的分布。通过从第一个直方图的对应值中减去第二个直方图的值,得到再一个直方图。所计算的直方图被显示于图 10a、10b 和 10c。

[0082] 图 10c 上的直方图是图 9 上浅灰色区域的直方图(这是因为它是通过上述的减法过程而被创建的)。这是不包括边缘的区域,这暗示不带有字母的区域,因为只在字母-背景边界处才发现边缘。换句话说,它是对应于背景的区域。这将通过更详细地讨论图 10a 和 10b 的直方图而被展示。

[0083] 图 10a 上的直方图代表对于图 7 的整个矩形的像素强度的分布。如图所示,有一个由背景像素引起的占主导的峰。严格地说,这个占主导的峰的存在并不是惯例(考虑完全粗体化的文本),所以还不会假设这实际上代表了背景。在峰的左面还有直方图的相对平坦的部分,其是由文本和背景以及它们的混合引起的。

[0084] 直方图 10b 代表在图 9 所示的内部矩形内的像素的像素强度的分布,其对应于边缘区域。除了背景峰不是那样占主导之外,这个直方图看起来相当类似于 10a 上的直方图。第三直方图(图 10c)是最重要的。如上所指出的,它对应于背景像素。由于直方图减法过程,起源于文本和文本-背景混合的成分被完全抵消(注意,缺少来自图 10a 的平坦区域)。在直方图上对应于峰的最大值的特定 bin(仓室)给出占主导的背景彩色的强度。在本例中,它是 190。

[0085] 从图 10c 的直方图也可以得到背景动态范围。它可以通过检查在 0 开始的强度并找出对于具有非零数量的像素的第一 bin 的强度值而被找到。在图 10c 上, 这个值是 150。然后, 对于其中像素的数量返回到零的随后的 bin, 找出强度值。在图 10c 上, 这个值是 230。因此, 动态范围是 [150, 230]。在直方图 10b 上的、在这个范围以外的强度值是由前景文本像素导致的。在图 10b 的直方图上对应于从前景像素引起的第二峰的最大值的 bin 对应于占主导的前景彩色的强度, 在本例中, 它是 47。

[0086] 对于前景和背景行剖析的估计可以通过把这个过程应用到自然行的每个局部化区域(例如, 图 7 上的矩形)而得到。应当指出, 得到这些估计所需要的所有处理依赖于简单的直方图处理; 即, 直方图计算、直方图减法和最大值提取。不执行复杂的操作(例如, 峰检测、峰宽度估计、平坦区域检测), 因此几乎不留机会来让潜在的错误由于复杂的分析引发。对于图 6 上文本行的计算的背景和前景行剖析被描绘于图 11, 其中上面的线代表背景行剖析和下面的线代表前景行剖析。

[0087] 在这一点, 应当执行一个附加的步骤。即, 通过观察在两个行剖析之间的关系, 可以容易地提取文本-背景关系, 以便确定哪一个彩色更深。如果在深色背景上找到浅色文本, 则可以应用图像求逆运算符 $Inv(I(x, y))$ 。这个运算符被定义为:

[0088] $Inv(I(x, y)) = 255 - I(x, y)$ 。

[0089] 应用这个运算符的结果被描绘于图 12。在执行这个求逆后, 当必要时, 可以安全地假设深色文本出现在浅色背景上。一旦背景和前景行剖析被计算(以及当必要时执行图像求逆), 就可以通过使用阈值处理(thresholding)而滤除背景, 这将在下面描述。

[0090] 阈值处理

[0091] 一旦行剖析被计算, 滤波背景的过程就相当直截了当了。函数 $Thresh(I(x, y)) : GS \rightarrow GS$ 被应用到每个像素, 且被定义为:

[0092]
$$Thresh(I(x, y)) = \begin{cases} 255, & I(x, y) > (FLP(x) + BLP(x)) / 2 \\ I(x, y), & \text{否则} \end{cases}$$
。

[0093] 换句话说, 具有小于前景和背景剖析的平均的像素强度值的每个像素仍维持原样, 而其它像素被改变以使得它们看起来具有最大强度。也就是, 它们被改变为白色。这种阈值处理可被看作为根据它们的像素值把像素分类为前景或背景的过程。如果像素被分类为前景像素, 则它的值不被修改, 而如果像素被分类为背景像素, 则它的值被改变为白色。

[0094] 更一般地, 在一些实现中, 不是把阈值确立为前景和背景剖析的平均, 而是可以选择某个其它的参数, 其是从前景和背景剖析确定的。

[0095] 图 3 描绘的图像是通过使用上述的阈值处理技术而得到的。这个技术产生非常好的结果。然而, 它在低质量图像上的应用并不总是令人满意的, 以及某种文本增强技术可被应用到这样的图像, 正如下面讨论的。

[0096] 低质量文本增强

[0097] 图 13 显示低质量文本图像的例子。在这个图像上应用前述的阈值处理规程的结果显示于图 14。显然, 作为结果的图像具有差的质量。任何随后的处理步骤注定要作出错误的判决, 因为图像即使对于人来说都几乎是难辨认的。劣质结果的原因是明显的。组成图 13 的文本笔划的像素值具有大的变化。相比于前景像素, 它们中的某些甚至更类似于背

景像素,且某些有噪声的(noisy)像素具有类似于前景像素值的值。另外,笔划宽度极小,导致很少的文本完整性。

[0098] 以上描述的阈值处理规程对于阈值函数使用固定值,其常常是在局部背景与前景剖析的中间。克服关于低质量文本图像的问题的一个方法是使得阈值处理规程更自适应,以使得阈值对于不同的图像是可调节的。不幸地,为了有效地完成这一点通常需要更多的上下文信息。阈值处理过程的一个替换例牵涉到使用更智能的(和更密集的)处理来检测低质量图像,以便提供更好的背景滤波。这样的替换过程将在讨论可以如何检测低质量文本图像以后进行描述。

[0099] 低质量文本检测

[0100] 检测低质量文本图像的第一步骤牵涉到识别使这样的图像质量低的那些特征。如前所述,这样的文本图像的特征在于小的笔划宽度和大的文本彩色变化。如果笔划宽度是较大的,则以上描述的阈值处理过程将产生字符的“变薄的”再现,但它们仍旧是可辨认的。这对于小的彩色变化也是成立的。可以得出结论,低质量文本图像的关键特征是小的笔划宽度与大的彩色强度变化的组合。

[0101] 要指出的另一件事情是图 14 描绘的阈值处理规程的结果。显然,这个规程在字符的一些部分中产生具有小的宽度(典型地小于两个像素)的笔划。通过计算在这个图像上的“笔划”宽度的中值,可以评估文本质量。

[0102] 为了描述用于估计笔划宽度的规程,需要以下的定义。

[0103] 定义 11. 令 $cc = \{(x_1, y_1), \dots, (x_n, y_n)\}$ 是已连成分。每个这样的 $(x_i, y_i) \in cc$, 即其使得存在有属于 (x_i, y_i) 的 8-邻居集合的 $(x_j, y_j) \in cc$, 被称为 cc 的边缘像素。

[0104] 笔划宽度估计规程通过计算图 14 上的每个边缘前景像素的局部笔划宽度开始。这可以通过计算在所有可能的方向上的笔划宽度而完成,如图 15 中对于箭头发源于其中的单个前景像素所例示的。

[0105] 从选择的前景像素开始,检查在 90 度方向的相邻像素。在垂直方向上,在到达背景像素之前有一串 3 个连续的前景像素。类似地,在 0 度方向有 2 个前景像素,在 45 度方向有 1 个前景像素,以及在 135 度方向延伸了 3 个前景像素。从这四个值中间,选择最小的(在这种情形下是 1)作为对于所选择的像素的局部笔划宽度。这个相同的规程被应用到所有的边缘像素,由此产生一组局部笔划宽度。对于文本的总的笔划宽度的良好估计是这组值的中值,其被表示为 SW_{or} 。

[0106] 一旦计算 SW_{or} ,所有的必须的数据就可得到来检查文本图像的质量。如果 SW_{or} 小于某个预定义的阈值 SW_{tr} ,则图像可被看作为低质量图像。 SW_{tr} 的选择影响过程的其余部分。因为选择的 SW_{tr} 的值越小,被看作为低质量的自然行图像越少。虽然这暗示较小数量的随后处理,但它也使得某些自然行图像未被增强。另一方面,如果为 SW_{tr} 选择较大的值,则将以更多的处理功率为代价来增强更多的自然行。在某些实现中,适当的折衷值是在 $SW_{tr} \in [2, 4]$ 的范围内的。

[0107] 一旦低质量图像被检测到,就可以应用文本增强方法。下面描述了用于文本增强的两个说明性方法。第一个方法是非常精确的,但在需要的处理功率方面是昂贵的,而第二个方法是更经济的,但多多少少不太精确。

[0108] 精确的文本增强

[0109] 被传递到文本增强部件的图像具有小于某个预定义的阈值的笔划宽度 SW_{or} 。这暗示在处理中只有很小的自由度。克服这个问题的一种方式执行图像内插。图像内插是一种用于增加图像的分辨率的技术。虽然有大量的内插技术是可用的, 但一个合适的例子是双线性内插, 它是非常简单的然而又是足够强有力的来服务于文本增强过程的需要。

[0110] 作为例子, 假设图像要被内插以使得它的分辨率在两个 (x 和 y) 方向上增加到两倍。这意味着, 在原始图像上的每个像素被映射到内插的图像上的一组四个像素。这个过程例示于图 16。

[0111] 在双线性内插过程中的第一步骤是把原始图像上的坐标 (x, y) 处的像素值映射到输出图像上的坐标 ($2x, 2y$)。这个过程描绘于图 17。包括小的画阴影的矩形的像素是没有被分配以值的像素。这样, 输出像素的四分之一被分配以恰当的像素值。应当指出, 映射或复制在输出图像上产生位于两个已分配的像素 (这些像素是具有至少一个偶坐标的像素) 之间的某些像素。这些像素的值可以被计算为被分配给它们的相邻像素的平均值 (见图 17b)。在这一点, 只有在输出图像上具有奇坐标的像素仍旧是未被分配的。这些像素的值可以被计算为在它的 4- 邻居集合中的像素的所有像素值的平均。在双线性内插过程中的这个最后步骤产生内插的图像 (见图 17c)。图 17c 上的最后图像有某些像素不带有分配的值, 因为它们依赖于图上未示出的像素值。

[0112] 图 18 的底部部分显示把内插过程应用到该图的上部部分所描绘的文本图像的结果。

[0113] 把图像内插应用到低质量文本图像所引发的一个有利的副作用是内插过程的固有的平滑影响。这产生具有文本彩色强度的更小方差的图像, 其是对于低质量文本的问题之一。

[0114] 在这一点, 在文本增强过程中, 知道以下的关于内插的图像的信息:

[0115] - 背景行剖析 $BLP(x)$

[0116] - 前景行剖析 $FLP(x)$

[0117] - 笔划宽度 $SW_{hr} = 2 * SW_{or} + 1$ 。

[0118] 由于阈值处理算法产生文本的受侵蚀 (erode) 的版本, 因此小于 $((BLP(x) + FLP(x)) / 2)$ 的所有像素值可以被安全地看作为文本的一部分, 并被分配以 0 的值。同样地, 具有大于 $BLP(x)$ 的值的像素可被分类为背景像素, 并被分配以 255 的值。然而, 具有在从 $((BLP(x) + FLP(x)) / 2$ 到 $BLP(x) = (L_1(x), L_2(x))$ 的范围内的值的像素在分类方面是成问题的, 并且需要某种附加处理来确定它们是属于文本还是属于背景。

[0119] 为了以易管理的和方便的方式分类这些像素, 像素彩色分辨率可以通过使用以下等式而被划分成在前述的不确定范围上的有限数量的强度区间 n :

[0120]

$$I(x, y) = \begin{cases} I_1 = \Delta I, & I(x, y) \in (L_1(x), L_1(x) + \Delta L) \\ I_2 = 2\Delta I, & I(x, y) \in (L_1(x) + \Delta L, L_1(x) + 2\Delta L) \\ \dots & \dots \\ I_{n-1} = (n-1)\Delta I, & I(x, y) \in (L_1(x) + (n-1)\Delta L, L_1(x) + n\Delta L) \end{cases}$$

[0121] 其中 $\Delta I = 255/n$ 和 $\Delta L = (L_2(x) - L_1(x))/n$ 。

[0122] 确定强度区间的尺寸的参数是 n ，它应当被选择成产生少量的强度区间。在一些实现中，好的选择可以是例如 $n \in [3, 5]$ 。通过把强度划分成 n 个区间来减小彩色分辨率，可被使用来以某种方式增强文本的质量，该方式将结合图 19 和 20 描述。

[0123] 现在将把文本增强过程应用到图 19 上的字母 F。图 20 通过把像素加到这个字母而描绘了该过程。具有在第一区间 ΔI 内的强度的像素被添加到图 20 的最左面的字母。具有在第二区间 ΔI 内的强度的像素被添加到图 20 的从左面数的第二个字母。这个过程通过把在每个随后更高的强度区间内的像素添加到图上每个随后的字母而重复进行。

[0124] 图 20 清楚地例示了以这种方式添加像素的结果。具体地，随着像素被添加，字母 F 变得越来越明确。虽然明显地这个过程可被使用来增强或重建各个字母，但当字母开始“渗漏”到背景或开始互相接触时引发了问题。为了防止这一点，必须采用某种停止准则来确定何时不应当添加附加像素。以下的定义在得到适当的停止准则时是有用的。

[0125] 定义 12. 令 cc 是已连成分。属于已连成分的、与 cc 的两个不同的边缘像素等距离的一组像素被称为 cc 的骨架。

[0126] 图 21a 显示已连成分的例子。图 21b 上较深色的像素是对于这个已连成分的边缘像素，图 21c 的交叉影线 (cross-hatched) 像素代表已连成分的骨架。

[0127] 要观察的一件事情是像素添加过程对于骨架几乎没有影响。这可以在图 22 中看到，其描绘了对应于图 20 中的像素添加过程的已连成分的骨架。在每个像素添加步骤，已连成分的骨架几乎保持为相同的。确实改变的东西是骨架像素与边缘像素之间的距离。添加的像素越多，骨架像素与边缘像素之间的距离变得越大。

[0128] 在这一点，重要的是记起这个已连成分的笔划宽度 SW_{hr} 已经被计算。合理的是要求像素添加过程在某小数量的骨架像素离边缘像素超过 $SW_{hr}/2$ 的距离时终止。如果这个值要被超过，则已连成分将开始“渗漏”到背景。因此，在这一点不再有像素应当添加到字母。

[0129] 记住前面的讨论，文本增强算法可以被制定为如下：

[0130] 1) 以两倍高的分辨率内插低质量文本图像。

[0131] 2) 通过规定有限数量的强度区间来降低文本图像的彩色分辨率。

[0132] 3) 识别在图像中仅仅由最深色的像素 (例如，具有零强度的那些像素) 形成的所有的已连成分。这些已连成分将被称为基本已连成分。

[0133] 4) 对于每个基本已连成分执行以下步骤：

[0134] a. 计算已连成分的骨架。

[0135] b. 如果在骨架中的像素离边缘像素小于距离 $SW_{hr}/2$ ，则通过把属于下一个强度区间 (当前的 $I + \Delta I$) 的相邻像素添加到当前的已连成分而创建新的已连成分。

[0136] c. 重复步骤 a 和 b，直至满足停止准则为止。也就是说，当有离边缘像素大于 $SW_{hr}/2$ 的距离的骨架像素时，这个过程停止。

[0137] 一旦所有的基本已连成分都以这种方式被处理，文本增强过程就结束。把这个过程应用到如图 13 描绘的图像的结果例示于图 23。

[0138] 经济的文本增强

[0139] 先前描述的用于通过使用笔划恢复来增强文本图像的算法常常可以提供非常好的结果。然而，一个缺点是它的计算复杂性。在这一节，将描述该算法的一个变例，它不太

昂贵(就处理功率而言),但具有类似的质量。

[0140] 如在先前的方法中那样,这个方法执行图像内插,且把彩色分辨率按以上描述的方式分成有限数量的区间。然而,在前一节中描述的算法是费时的,因为对基本已连成分应用了递归过程,其牵涉到骨架的计算。由于骨架是停止准则的组成部分,所以可以利用停止准则的某个其它变例来降低算法的计算复杂性。下面描述一个方法。

[0141] 在减小彩色分辨率后,留下一小组的彩色。具有 255 的强度的所有像素(白色像素)被认为是背景的一部分。具有零的强度的所有像素(黑色像素)被认为是文本像素。是灰色像素才成问题。再次参照图 19,一个避免需要执行骨架的精确计算而同时得到它的估计的方式是(对于每个灰色像素)找出以该像素为中心的、完全由非白色像素组成的最大矩形。令 $a(x, y)$ 是对于像素 $I(x, y)$ 的矩形的尺度。具有满足 $a(x, y) \geq SW_{thr}$ 的矩形的所有像素都几乎肯定是文本的一部分或是背景中的平坦的有噪声区域。这个集合被表示为 $A = \{(x_1, y_1), \dots, (x_j, y_j)\}$ 。现在,使得 $\forall (x_i, y_i), d((x_k, y_k), (x_i, y_i)) \leq SW_{thr}$ (其中, d 是欧几里得距离)的所有像素 $(x_k, y_k) \in A$ 被添加到该集合中。换句话说,来自集合 A 内的

的像素的邻近区域的所有像素被添加,这是对于上述的算法中像素添加规程的替换例。为了保证不发生到背景的“渗漏”,集合 A 可以经历滤波。这样,如果来自该集合的像素离最接近的边缘的距离大于 SW_{thr} ,则可以消除这些像素。这样,平坦的有噪声区域从文本中被消除。

[0142] 这个算法类似于上述的更精确的算法。然而,不是在添加区间内的每组像素之前执行递归计算,而是应用一步近似,从而导致较小的处理时间量。通过对低质量文本图像应用这个算法而达到的结果被发现类似于通过应用精确的算法而达到的结果,所述结果被描绘于图 23。

[0143] 图 24 显示图像处理设备 300 的一个例子,其可以对经历光学字符识别的文本图像执行上述的文本增强过程。可以被合并并在 OCR 引擎中的该设备包括输入部件 202,用于接收具有文本的自然行的输入图像。该设备还包括背景和前景行剖析部件 204、图像质量检测部件 206、阈值处理部件 208 和笔划增强部件 210。输出部件 212 生成增强的图像作为输出,其可以被 OCR 引擎的随后的部件利用。

[0144] 图 25 是显示用于增强要经历光学字符识别的文本图像的总体的方法的一个例子的流程图。该方法可以通过图 24 所示的图像处理设备实施。该方法在步骤 305 当由输入部件 202 接收到文本的自然行的图像时开始。在步骤 310,行剖析部件 204 然后对自然行执行边缘检测,以产生在边缘空间中的自然行的表示。在步骤 315,行剖析部件 204 接着识别在边缘空间中的自然行的表示中出现的每个已连成分。通过使用在步骤 310 和 315 中得到的信息,行剖析部件 204 在步骤 320 中以上述的方式确定背景和前景行剖析。一旦行剖析被确定,图像质量检测部件 206 就在步骤 325 中根据行剖析确定图像的质量是否足以分配像素给图像的背景或前景部分。如上所述,可被使用来确定图像质量的一个准则是已连成分的笔划宽度。如果在判决步骤 330 中发现图像质量是足够高的,则在步骤 335 中由阈值处理部件 208 把像素分配给图像的前景或背景部分。另一方面,如果图像质量是不够高的,则在步骤 340 中由笔划增强部件 210 执行图像增强,以便增强已连成分的笔划。最后,在步骤 345 中,输出部件 212 生成增强的图像,其可以被 OCR 引擎的随后部件利用。

[0145] 当在本申请中使用时,术语“部件”、“模块”、“系统”、“设备”、“接口”等等通常打

算指与计算机有关的实体,或者是硬件、硬件与软件的组合、软件,或者是执行中的软件。例如,部件可以是,但不限于,在处理器上运行的进程、处理器、对象、可执行文件、执行的线程、程序和 / 或计算机。作为说明,在控制器上运行的应用和该控制器都可以是部件。一个或多个部件可以驻留在执行的线程和 / 或进程内,以及部件可以被定位在一个计算机上和 / 或被分布在两个或更多个计算机之间。

[0146] 而且,所要求保护的主体可以通过使用标准编程和 / 或工程技术产生软件、固件、硬件或它们的任何组合,而被实施为方法、设备或制品,以便控制计算机实施所公开的主题。当在这里使用时,术语“制品”打算包括从任何计算机可读装置、载体、或媒体可访问的计算机程序。例如,计算机可读媒体可包括,但不限于,磁存储装置(例如,硬盘、软盘、磁条…)、光盘(例如,紧凑盘(CD)、数字多功能盘(DVD)…)、智能卡和闪速存储器装置(例如,卡、棒、键驱动(key drive)…)。当然,本领域技术人员将会认识到,可以对于本配置做出许多修改而不背离所要求保护的主体范围或精神。

[0147] 虽然本主题是以对于结构特征和 / 或方法动作特定的语言描述的,但应当明白,在所附权利要求中限定的主题不是必须限于以上描述的特定特征或动作。相反,以上描述的特定特征或动作是作为实施权利要求的示例性形式被公开的。

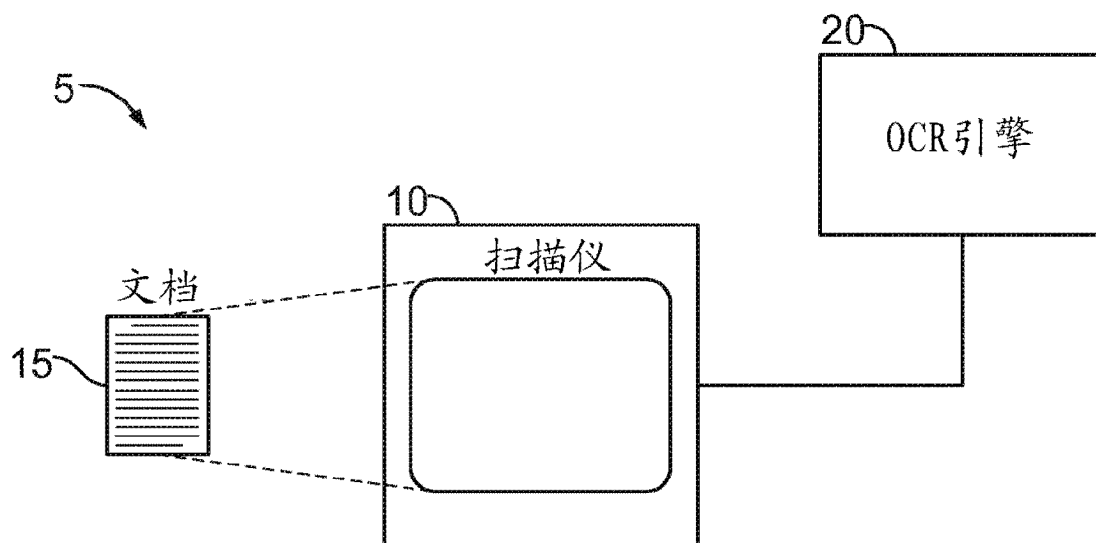


图 1

opportunity also to forward

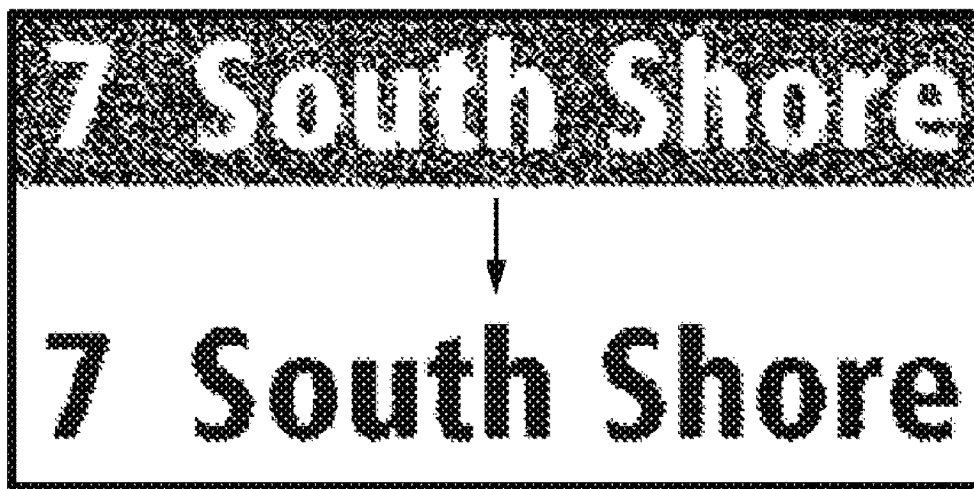
(样本文本)

图 2

opportunity also to forward

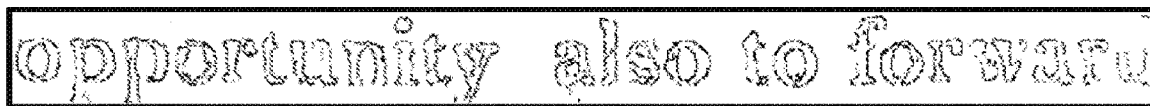
(在白色背景上的灰度文本)

图 3



(深色背景变换)

图 4



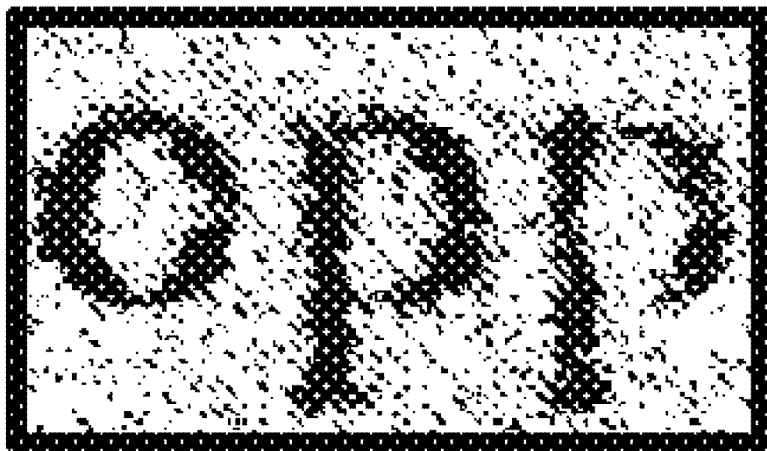
(在边缘空间中的文本行表示)

图 5



(文本图像的灰度版本)

图 6



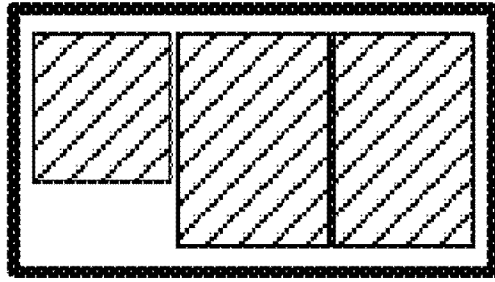
(自然行的片段)

图 7



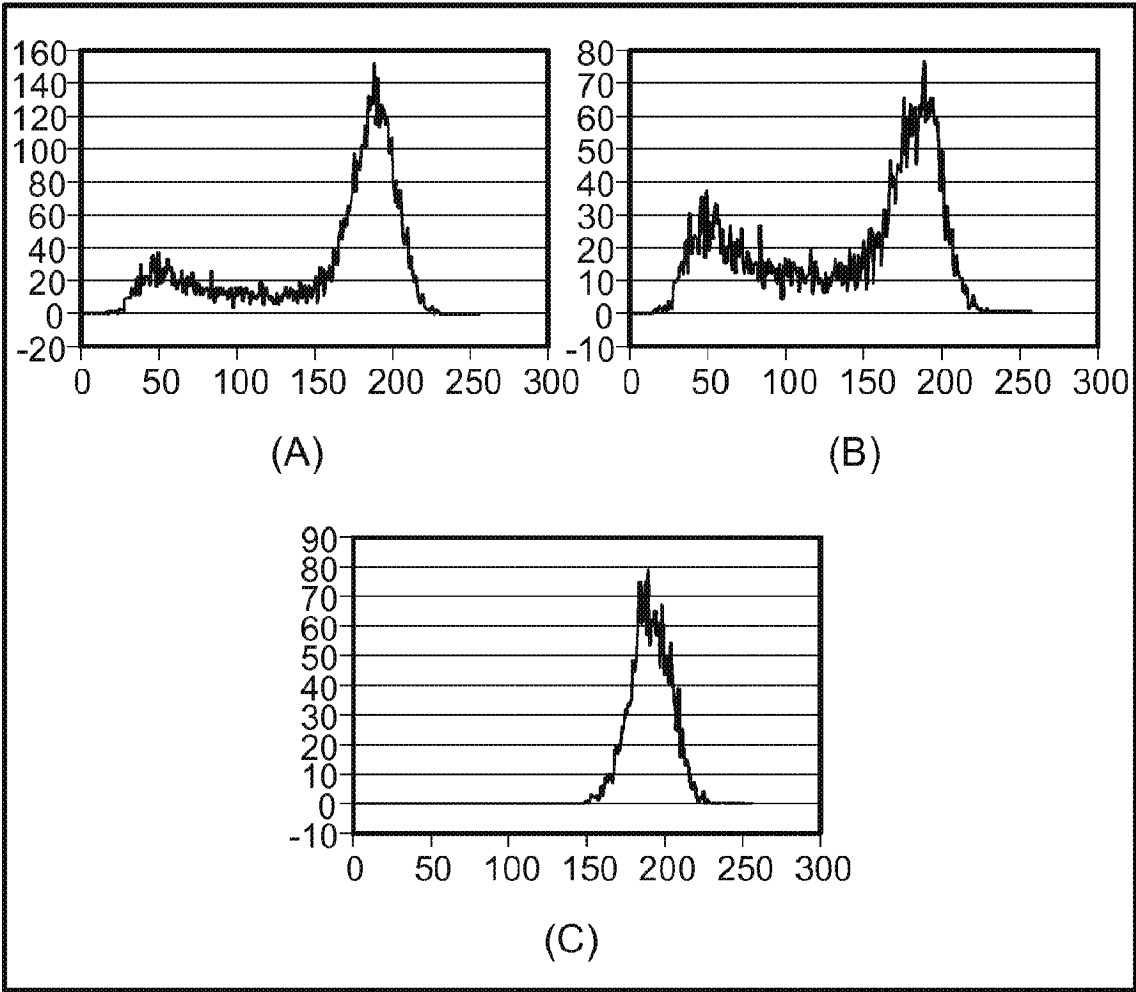
(在边缘空间中的自然行片段)

图 8



(在边缘空间中的自然行片段)

图 9



(直方图)

图 10

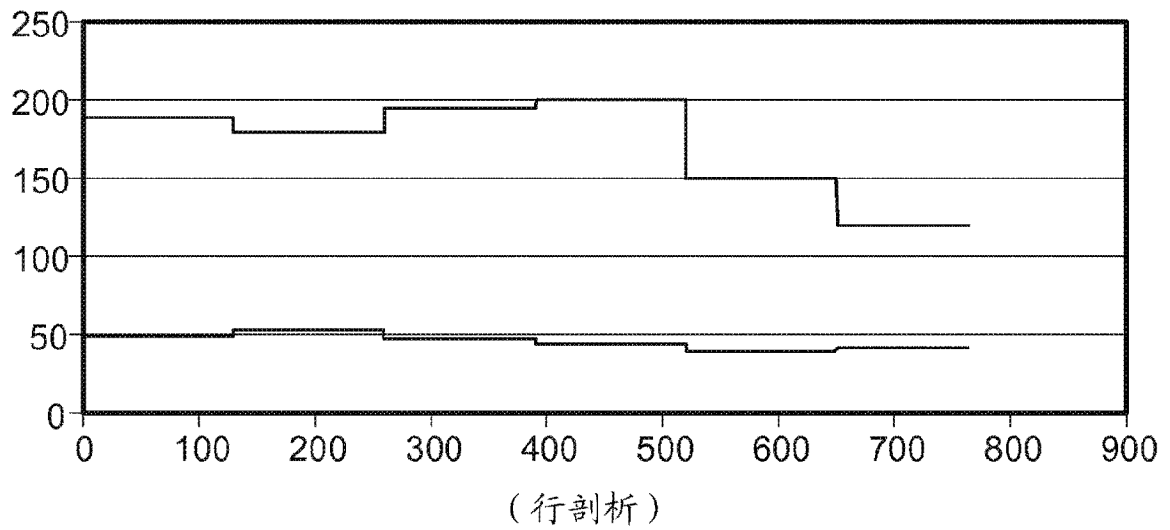
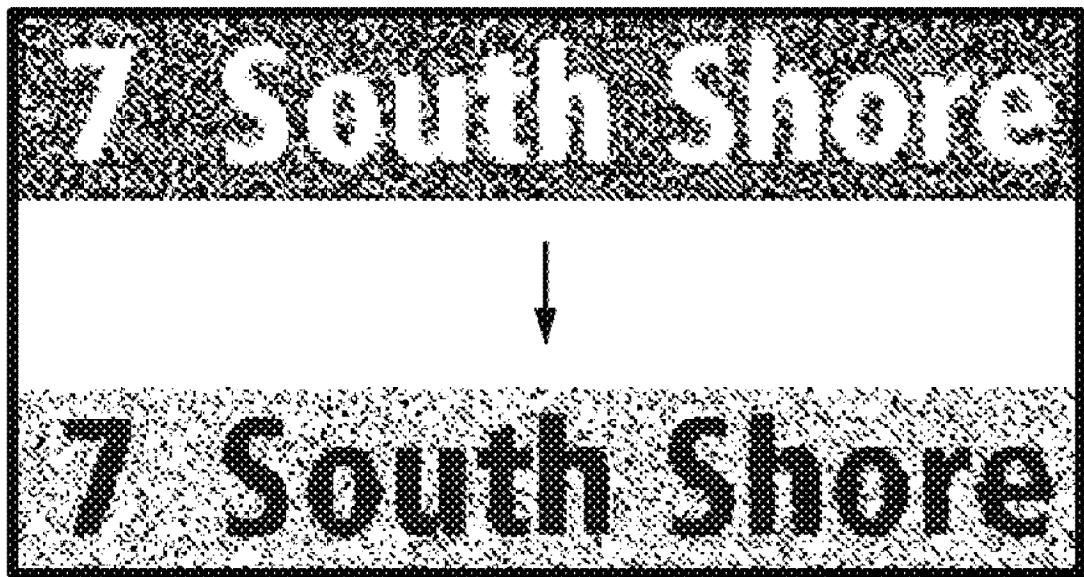


图 11



(图像求逆)

图 12



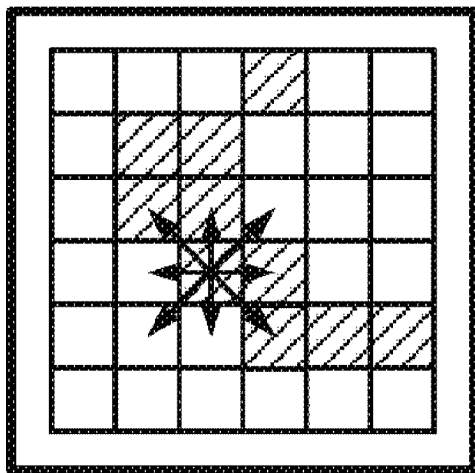
(低质量文本图像)

图 13



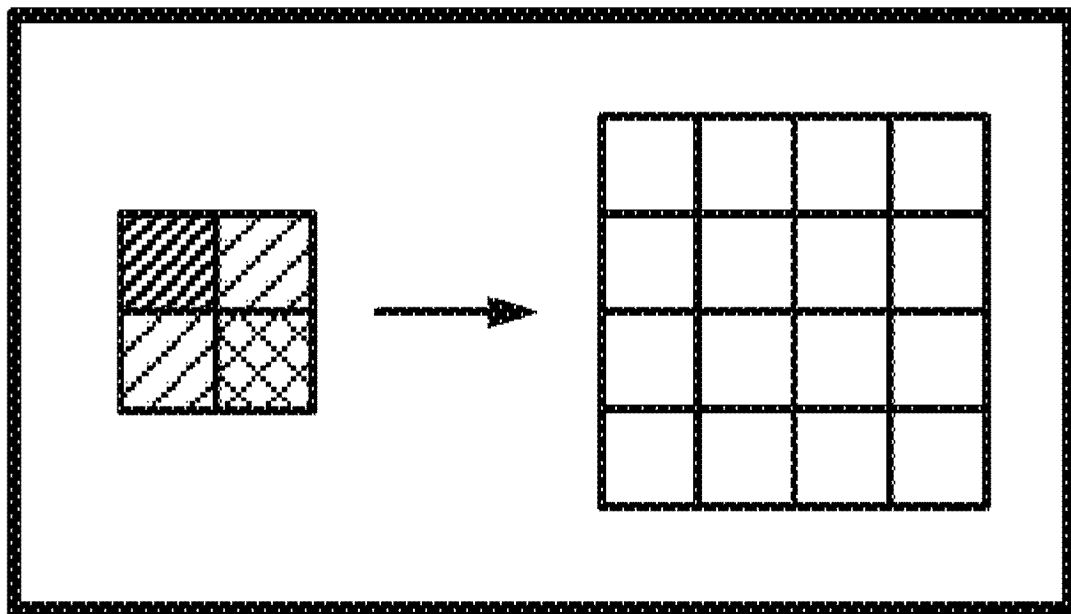
(在低质量文本图像上的阈值处理的结果)

图 14



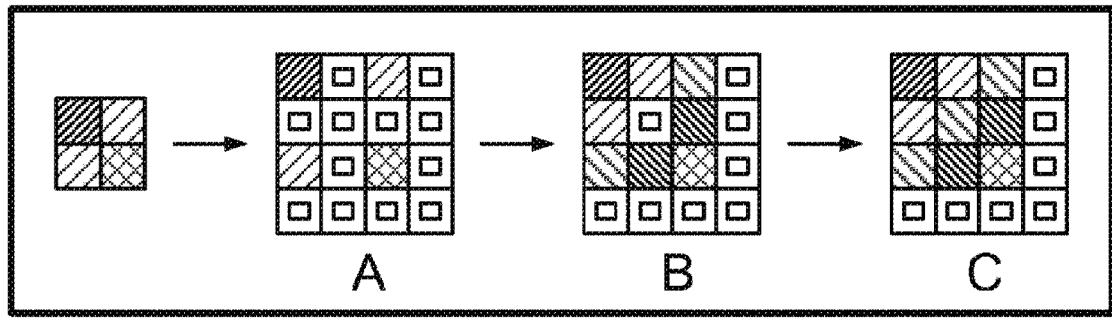
(对于单个前景像素的笔划宽度计算)

图 15



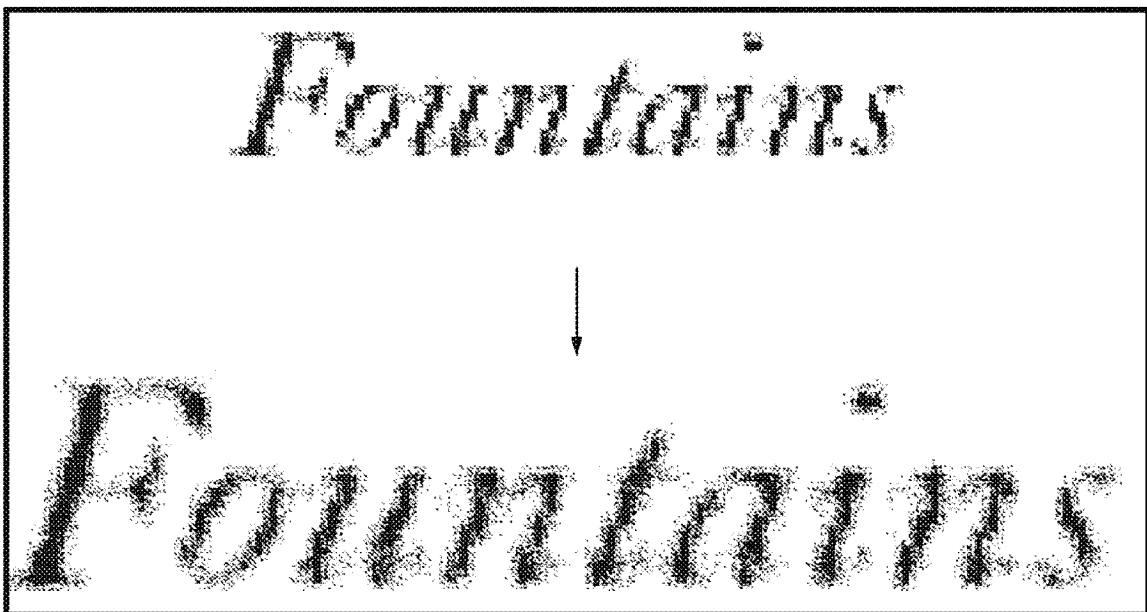
(图像内插)

图 16



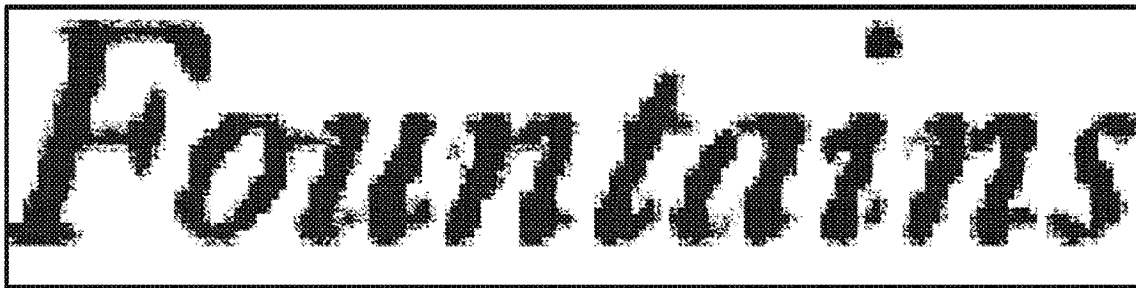
(内插步骤)

图 17



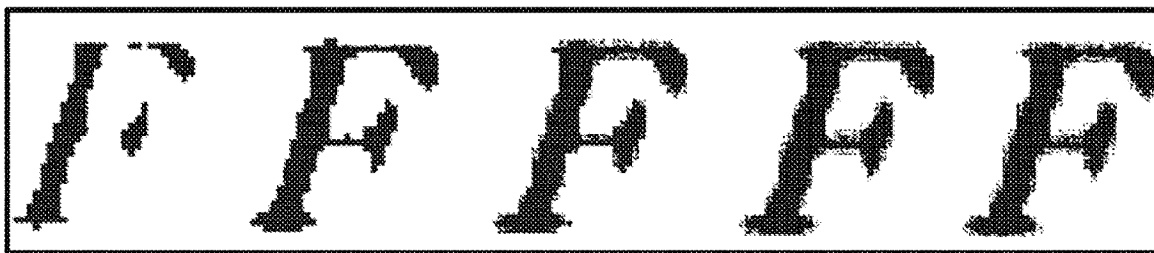
(文本图像内插)

图 18



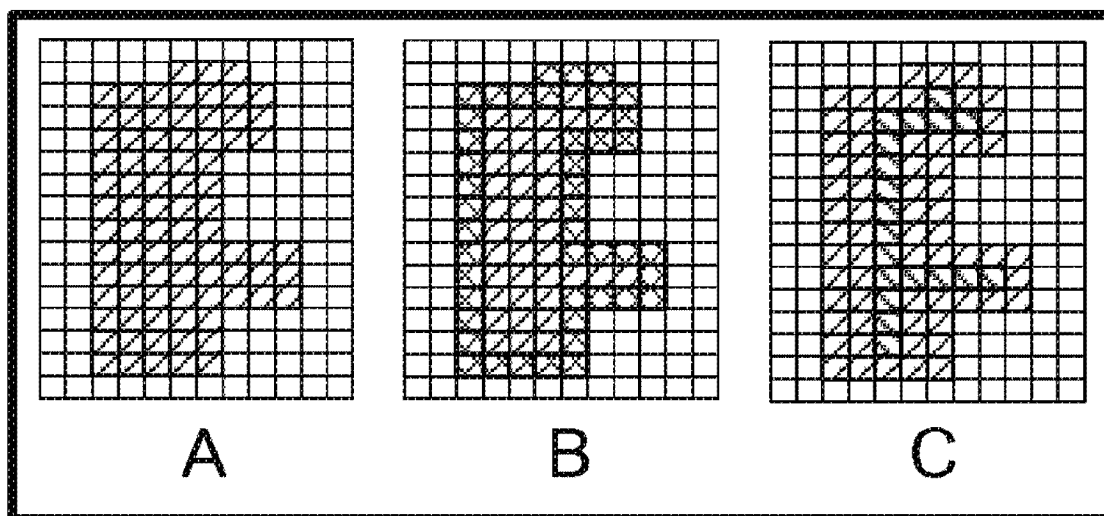
(具有降低的彩色分辨率的文本图像)

图 19



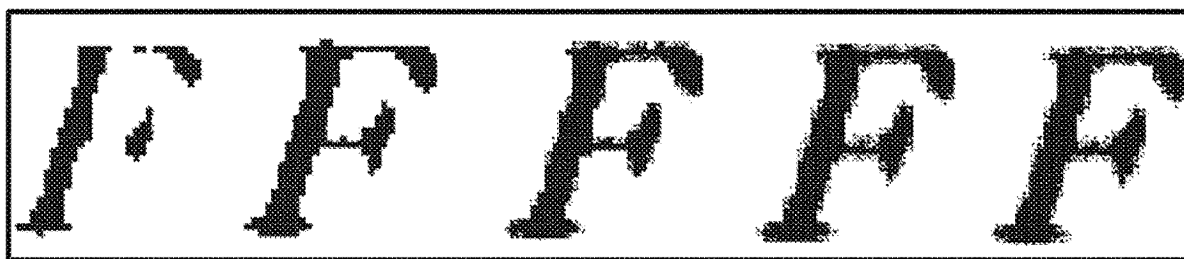
(像素添加过程)

图 20



(骨架例示)

图 21



(字母骨架)

图 22



(增强的文本)

图 23

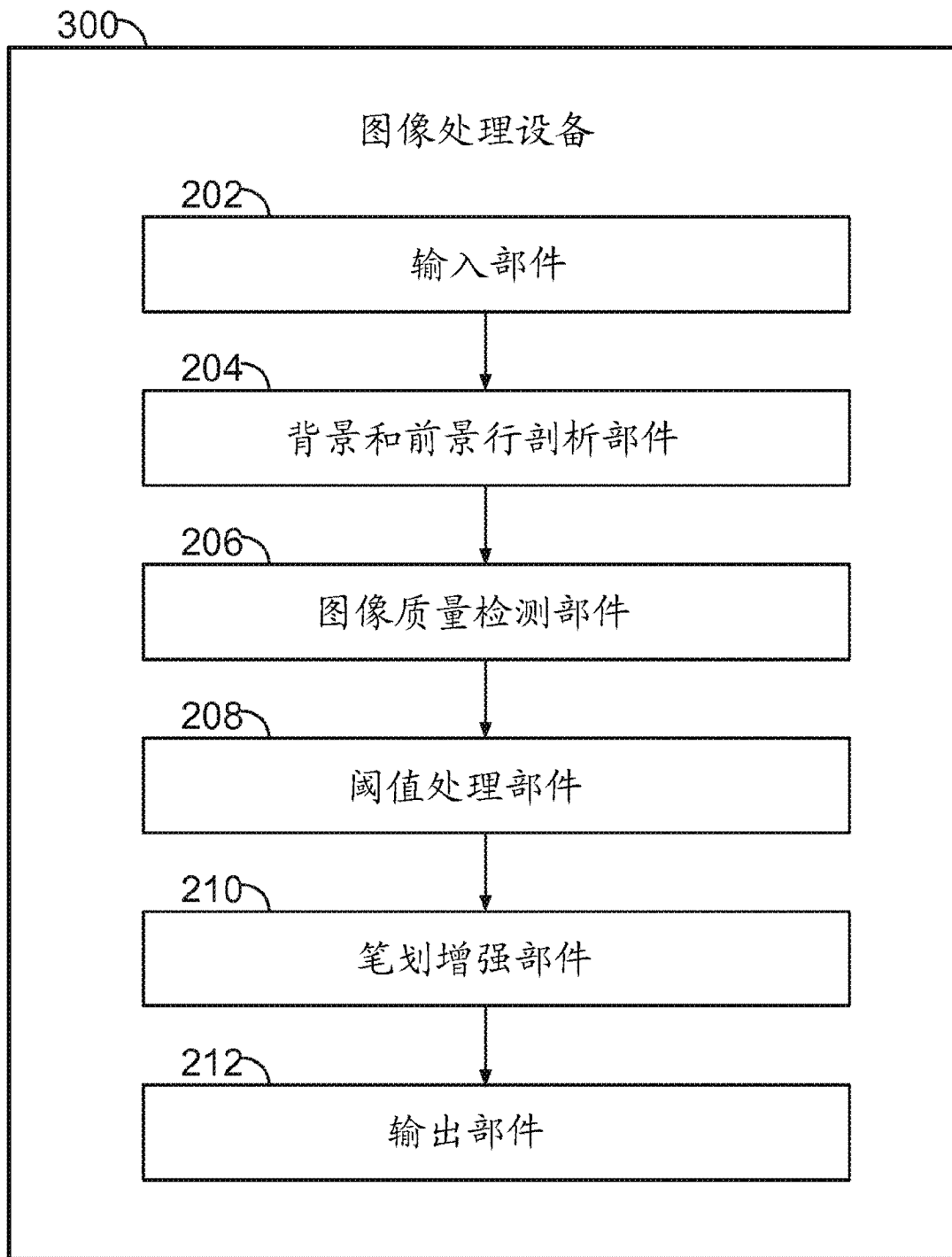


图 24

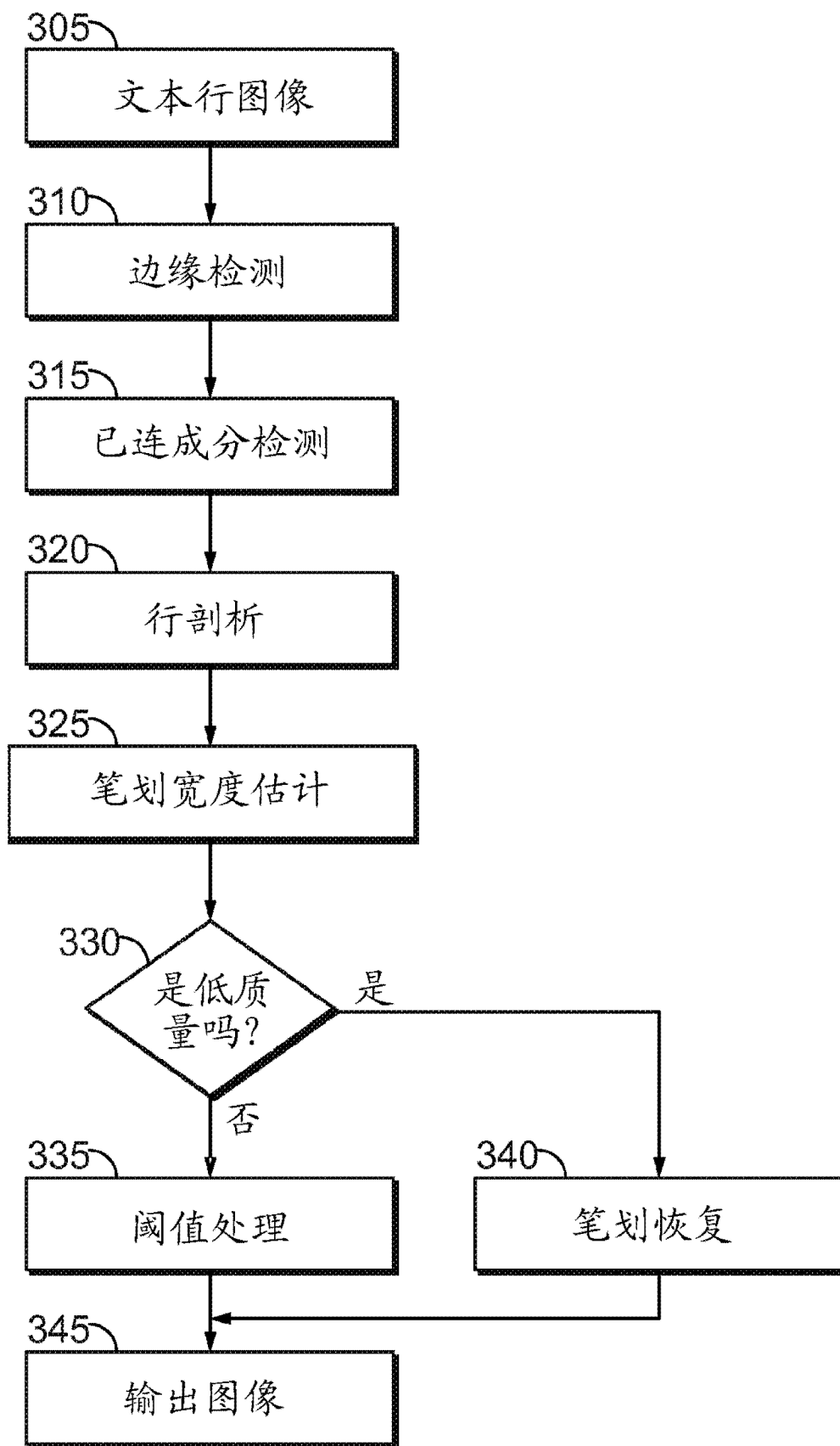


图 25