

Hough 变换 OCR 图象倾斜矫正方法

瞿 洋

(西安空军工程大学工程学院航空管理工程系计算机教研室, 西安 710038)

杨利平

(南京空军司令部气象处, 南京 210018)

摘要 在光学字符识别(OCR)图象扫描输入的过程中,扫描图象或多或少会出现某种程度的倾斜,这种图象的倾斜不仅会给下一步字符的切割造成困难,也影响最终的字符识别精度.通常情况下,为避免用户重新扫描,可以通过软件方法对图象进行矫正,为此提出一种利用 Hough 变换进行图象倾斜矫正的方法.为克服 Hough 变换计算量大的缺点,该方法采用了变分辨率图象金字塔策略.实验结果表明,该方法能快速准确测量出扫描图象的倾斜角度,并且具有很高的抗噪声性和应用适应性.

关键词 光学字符识别 倾斜矫正 Hough 变换 图象金字塔处理

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2001)02-0178-04

Hough Transform OCR Image Slant Correction Method

QU Yang

(Dept of Aeronautical Management Engineering, Institute of Engineering, Air-Force University of Engineering, Xi'an 710038)

YANG Li-ping

(Dept of Meteorology, Air-Force Headquarters of Nanjing, Nanjing 210018)

Abstract In process of OCR image scanning input, images may slant somehow. Severe image slant makes character segmentation next step difficult and lowers overall recognition accuracy of the system. Normally, in order to improve recognition accuracy and save image rescanning time, software correction is used to correct the slant before next step recognition processes of OCR are executing. This paper presents a novel slant correction method called Hough Transform OCR Image Slant Correction Method. The method uses Hough Transform to estimate the slant of scanning images. Because of heavy computing burdens of Hough Transform, the method employs Image Pyramid Processing to overcome the defeat of Hough Transform. Horizontal edge extracting processing is also used to speed the estimating process and to improve estimating accuracy. A great many artificial images and raw scanning images are used to test performances of the methodology. Experiment results show that the method can measure out slanting degree of scanning image efficiently and accurately just in a few seconds, and bears high noise endurance and application adaptability.

Keywords OCR, Slant correction, Hough transform; Image pyramid processing

0 图象倾斜矫正算法的提出

在 OCR 图象扫描录入的过程中,图象或多或少会出现某种程度的倾斜,这种图象的倾斜不仅会给下一步字符的切割造成困难,也影响字符识别的精度,因此有必要设计出一个算法,来测量出图象的旋转角度,以便对图象进行倾斜矫正.

OCR 识别的材料一般包含成段文字(汉字、字母和数字符号)、表格,或两者兼而有之.对于材料包含表格的扫描图象,可以通过跟踪表格线,以找出较长表格横线的方法来确定当前图象的旋转角度.然而,如果表格线由于噪声发生间断,或表格线本身就是虚线时,这种方法就缺乏鲁棒性,还需要设计十分复杂的算法,而且即使设计出这样的算法,它也不能考虑到扫描图象中出现的所有状况.另外,对于文字

材料,这种方法更是无从下手。

大家知道,无论何种材料的图象,虽然人们一眼就能明显地感觉到它真正的水平方向,但是虽不可用其他方法也能够实现人类的这种感觉,并且能进行足够精确的量化? Hough 变换^[1]就能实现人类的这种感觉,并且它还具有良好的抗噪声性能。大家知道, Hough 变换是数字图象技术中一种有效的发现直线的算法。它是先把直角坐标系的目标点映射到极坐标系上进行累积,即它是先使直角坐标系平面上任一直线上的所有点均累积到极坐标系的同一点集中去,然后通过寻找极坐标系中点集的峰值,来发现长的直线特征。由于这种点集是通过累积统计得到的,因而能够容忍直线的间断。

用 Hough 变换来发现直线的过程,实际上是一个参数猜测过程。这种思想已经得到扩展,即已形成广义 Hough 变换^[2],它可用来发现任意形状的物体。然而 Hough 变换计算量很大,对于一个大的扫描图象,若要获得足够精度图象的旋转角度,以直接进行 Hough 变换,将耗时很长,内存要求也很多。针对该问题,本文采用变分辨率图象金字塔策略^[3,4],该方法类似于人眼“粗看”到“细看”的过程。它首先对原图象作减半采样处理,以得到低一级分辨率的图象;然后对该图象再一次作减半采样,以得到分辨率更低的图象,照此处理,可得到一组金字塔式的变分辨率图象。Hough 变换首先在分辨率较低的图象上进行,以得到粗略的图象倾斜角度,然后再在较高分辨率图象上进行细化,直到获得满意的精度。从上述过程可以看出,倾斜角度粗略估计主要集中在分辨率较低的图象上进行,由于这种图象尺寸较小,因此角度搜索的范围可以选取得很大,步长也可以取得很精细;再在分辨率较高的图象上进行角度细化时,则角度搜索的范围有限,因而这样总体计算量不大,却能够获得较高的精度。另外合适的减半采样过程还有助于提高算法的抗噪声能力。

1 图象倾斜矫正算法

1.1 变分辨率图象金字塔的形成

产生变分辨率图象金字塔实际上是一个减半采样处理过程。由于 Hough 变换具有极强抗噪声能力,因此产生变分辨率二值图象金字塔的原则是保留“黑点”,以便有利于低分辨率图象上倾斜角度的发现,其方法是:

设 L 为图象 $\{f(i, j)\}$ 的变分辨率图象金字塔层数, $B(s, i, j), i, j = 1, 2, \dots, 2^{L-s}$ 表示金字塔的第 s 层(自底至顶,从 0 层到 L 层)的 i 行 j 列的象素值, 0 代表白点, 1 代表黑点, 则第 $s-1$ 层和第 s 层的图象数据关系为

$$B(s, i, j) = B(s-1, 2i, 2j) \cup B(s-1, 2i, 2j+1) \cup B(s-1, 2i+1, 2j) \cup B(s-1, 2i+1, 2j+1) \quad (1)$$

其中,位于最低层,即 $s=0$ 的图象是原始图象, $B(0, i, j) = f(i, j), i, j = 1, 2, \dots, 2^L$ 。

1.2 Hough 变换

Hough 变换的基本策略是:由图象空间目标象素的坐标去计算参数空间中参考点的可能轨迹,并在一个累加器中给计算出的参考点计数。如果参数空间是极坐标 (ρ, θ) , 则坐标变换公式为

$$x \cos \theta + y \sin \theta = \rho \quad (2)$$

这里 x, y 是象素的直角坐标。由此, Hough 变换算法设计为:

- (1) 在 ρ, θ 合适的最大值、最小值之间建立一个离散的参数空间;
- (2) 建立一个累加器 $A(\rho, \theta)$ 矩阵,并置每一个元素为 0;

(3) 选取图象每一个目标象素 (x', y') , 并对每一个 θ 的取值 θ' , 计算 $\rho' = x' \cos \theta' + y' \sin \theta'$, 并在相应累加器 $A(\rho', \theta') = A(\rho', \theta') + 1$ (即把矩阵 A 的元素 $A(\rho', \theta')$ 的值增加 1);

1.3 Hough 变换累积结果的处理及倾斜角度发现

从上述 Hough 变换的过程可以看出,如果图象上存在直线,那么在累加器 $A(\rho, \theta)$ 矩阵中必有一对应的元素为局部极大值。虽然 OCR 识别的文字材料一般不存在真正直线,但文字行间有很强的方向性,因而在累加矩阵 $A(\rho, \theta)$ 中,对应某一个 θ' 列,由于 $A(\rho, \theta')$ 具有局部较大值,因此可以设定一个合适的门限值 T 取

$$T = \gamma \max A(\rho, \theta) \quad (3)$$

这里 $\gamma < 1$, 然后利用该门限值对 $A(\rho, \theta)$ 作如下变换

$$A(\rho, \theta) = 0 \quad \text{if} \quad A(\rho, \theta) \leq T \quad (4)$$

最后,对 $A(\rho, \theta)$ 进行列累加,以得到 $A'(\theta)$ 。此时选取 $A'(\theta)$ 最大的元素所对应的角度为发现的倾斜角度。

1.4 提高倾斜角度发现精度

由于文字笔画都有一定的粗细,而这种笔画粗细对倾斜角度发现精度有一定影响,通过研究及大

量实验证明,在 Hough 变换之前,先提取图象的水平边缘,然后把水平边缘作为 Hough 变换的目标象素,即可消除笔画粗细对倾斜角度发现精度的影响.另外,对于较高分辨率图象,应适当压缩 Hough 变换中的参数 ρ ,以保证累加矩阵 $A(\rho, \theta)$ 出现较高峰值.

这里水平边缘提取方法为:首先对图象进行列扫描,然后当象素值发生变化时,使该象素值为 1;若象素值未发生变化时,则使该象素值为 0.

1.5 倾斜矫正算法的流程

该算法过程步骤如下:

- (1) 利用式(1)生成图象金字塔;
- (2) 取最高层图象;
- (3) 提取图象的水平边缘,利用式(2)进行 Hough 变换;
- (4) 利用式(3)(4)对 Hough 变换累积结果进行处理,寻找并计算倾斜角度;
- (5) 若获得的倾斜角度满足精度要求,则转到第(7)步;
- (6) 角度细化,取下一层图象,转到第(3)步;
- (7) 旋转图象;
- (8) 结束.

2 实验结果

在用人造图象和实际扫描图象进行实验中,应用本文算法来生成 3 级图象金字塔的过程如下:首次 Hough 变换,角度的搜索范围为 $[-45^\circ, 45^\circ]$,步长为 1° ,共 91 个值;第 2 次 Hough 变换,角度的搜索范围为 4° ,步长为 0.2° ,共 21 个值;第 3 次 Hough 变换,角度的搜索的范围为 0.4° ,步长为 0.04° ,共 11 个值;Hough 变换累积结果处理参数 $\gamma = 0.66$;参数 ρ 在第 1 次 Hough 变换时不压缩,在第 2 次 Hough 变换时压缩 2 倍,在第 3 次 Hough 变换时压缩 4 倍.应当说明的是,参数取值并不严格,图象金字塔级数和角度搜索步长的选取则相对重要,因为不同取值会影响精度和速度,对一般图象,生成三级图象金字塔就足够了, γ 取值在 0.3 到 0.8 之间.

图 1 为一幅人造文字图象,并人为旋转了 10° .若直接使用所有的文字象素进行 Hough 变换,则算法测定的倾斜角度为 9.8° ;若使用图象的水平边缘象素进行 Hough 变换,则算法测定的倾斜角度为 10° .这说明提取水平边缘可提高算法精度.

图 2 为一幅扫描的英文和汉字混合的文本图



图 1 人造文字图象

象;图 3 是对图 2 形成的金字塔图象的第三级(顶级)提取水平边缘所得到的图象(即图 2 的左上角处理后的图象);图 2 经人工测算倾斜角度为 2.55° ,而用本算法测定的倾斜角度为 2.60° .图 4 为一幅扫描英文文本图象(英文文本图象弧线较多,对算法考验最为严峻).经人工测算的倾斜角度为 $-5.36^\circ \sim -5.83^\circ$,而用本算法测定的倾斜角度为 -5.72° .应当说明的是,图 1 至图 4 所显示的图象并不是它们的原始尺寸,为文稿显示方便,他们的放缩比例也不相同.

[17] 图式识别, 4种识别方法对比, 1979年第1期, 62-70.
 [18] 王仲威, 刘利军, 刘永, 刘永, 刘永, 1984.
 [19] 王仲威, 刘利军, 刘永, 刘永, 刘永, 1984.
 [20] 王仲威, 刘利军, 刘永, 刘永, 刘永, 1984.
 [21] 王仲威, 刘利军, 刘永, 刘永, 刘永, 1984.
 [22] 王仲威, 刘利军, 刘永, 刘永, 刘永, 1984.
 [23] 王仲威, 刘利军, 刘永, 刘永, 刘永, 1984.
 [24] 王仲威, 刘利军, 刘永, 刘永, 刘永, 1984.
 [25] 王仲威, 刘利军, 刘永, 刘永, 刘永, 1984.
 [26] 王仲威, 刘利军, 刘永, 刘永, 刘永, 1984.
 [27] 王仲威, 刘利军, 刘永, 刘永, 刘永, 1984.
 [28] 王仲威, 刘利军, 刘永, 刘永, 刘永, 1984.
 [29] 王仲威, 刘利军, 刘永, 刘永, 刘永, 1984.
 [30] 王仲威, 刘利军, 刘永, 刘永, 刘永, 1984.

图 2 扫描的英汉混合的文本图象

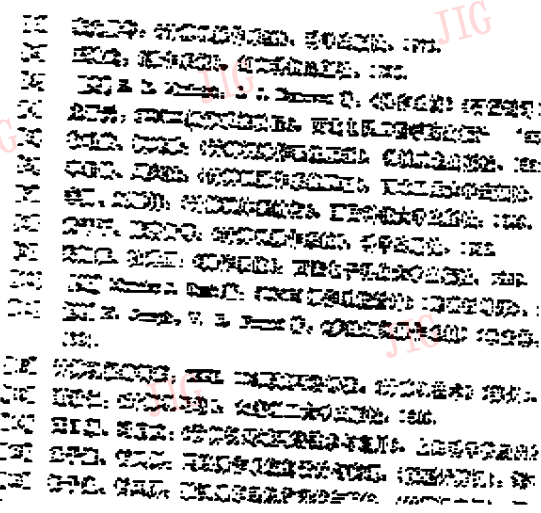


图 3 图 2 的左上角金字塔第三级水平边缘

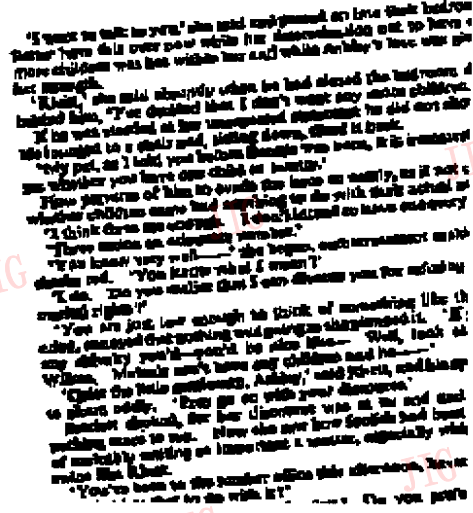


图 4 扫描英文文本图象

本文还测试了带表格的扫描文本图象,对于上述各类图象,本文算法均能准确测出各类图象的倾斜角度.从实验结果来看,本文算法优于目前若干个商业 OCR 软件.本文算法处理一个普通 $2\,000 \times 3\,000$ pixel 的扫描图象,在 Pentium133 微机,角度测量时间小于 2s.

上述实验结果表明,本文提出的利用 Hough 变换进行 OCR 图象倾斜矫正的方法,能快速准确测量扫描图象的倾斜角度,并且具有很高的抗噪声性和应用适应性.

参 考 文 献

- 1 徐建华. 图象处理与分析. 北京: 科学出版社, 1992 :194195.
- 2 Ballard D H. Generalizing the hough transform to detect arbitrary shapes. Pattern Recognition, 1981, 13(4):111122.

- 3 Zheng Qinfen, Rama Chellappa. A computational vision approach to image registration. IEEE Trans. Image Processing, 1993, 2(3):311325.
- 4 Li Ze-Nian. On edge preservation in multiresolution images. CVGIP, 1992, 54(6):461472.

瞿 洋 1966 年生,工学博士,副教授.研究领域为人工智能、图形图象处理、机器视觉.

杨利平 1968 年生,工学硕士,室副主任.研究领域为人工智能、图象处理.