

基于线结构分析的档案图像倾斜校正

杨 有^{1,2} 李晓虹²

(北京航空航天大学计算机学院数字媒体实验室 北京 100083)¹

(重庆师范大学数学与计算机科学学院 重庆 400047)²

摘要 档案数码化为政府档案资源深度利用提供了途径,数码化后的档案图像倾斜校正是数字档案质量的必然要求,也是数字图像的预处理过程。通过引入一个多尺度的高斯滤波器,可以对档案图像的全局特征进行增强,继而在增强的图像中,定义基于梯度函数的结构分析张量,提取像素点的线结构似然信息,对于大于一定阈值的线结构信息,计算其倾角,最后利用倾角直方图的统计特性确定档案图像的倾斜度。实验表明,该方法对文本富裕档案具有普遍适用性。

关键词 档案图像,倾斜校正,线结构分析张量

Slant Correction of Document Image Based on Line Structural Analysis

YANG You^{1,2} LI Xiao-Hong²

(Digital Media Laboratory, Computer College of Beihang University, Peking 100083)¹

(College of Mathematics and Computer Science, Chongqing Normal University, Chongqing 400047)²

Abstract Document digitalizing provides a method to utilize the archival resources of government comprehensively. Slant correction of digitalized document is necessary for the quality of digital image, and it's a preprocessing of any other procedure. By a Gaussian filter of multi-scale, the global structure of the original document image was enhanced. Based on gradients of enhanced image, a structural analysis tensor is also defined for the detection of line structure at the each pixel. For the line-likeness value at every pixel, which is greater than a given number, the corresponding angle is calculated. From the histogram of these angles, document skew can be determined. Our experiments show that the proposed method is adaptive to various documents with rich text.

Keywords Document image, Slant correction, Line structural analysis tensor

1 概述

自1960年,光电文字识别 OCR(Optical Character Recognition)开始用于档案图像的处理,就开启了计算机在档案图像数码化方面应用的先河。目前,伴随 Internet 技术的发展,诸如工商企业档案、国土资源档案等大量政府资源档案均以数字图像方式实现共享和检索,极大地改变了传统的政府办公模式,树立了良好的政府形象。

档案数码化就是将纸质档案(含照片)资料通过扫描、OCR 识别、排序标引转化为基于文字和图像信息的数字档案,并备份到光盘或磁盘设备中,用于长期保存。由于数码化过程中人为操作和数码化设备的差异均可能使数码化后的档案图像出现倾斜,图像纠偏是图像处理的第一步。

档案图像倾斜校正主要有四种方法^[1]:投影、特征点分布、Hough 变换和方向敏感特征统计。方向敏感特征统计方法首先从图像中提取方向敏感局部特征,并计算局部特征的倾角,再对倾角进行统计表决,最后得到档案图像的倾斜度。该方法在英文文本档案^[2]和手写体文本档案^[3]方面已有应用。

本文使用方向敏感特征方法,用于政府资源档案图像。文章的第2部分定义倾斜检测的计算公式,说明其原理;第3部分描述相应的算法;第4部分分析实验结果;最后指出算法

的不足和改进的方向。

2 线结构分析

2.1 多尺度图像滤波

高斯滤波器是忽视图像局部细微结构、增强全局结构的常用方法^[4]。对于文本型的档案图像,单个文字是局部结构,而文字串就构成全局结构。适当方差的高斯滤波后,可以使图像中由文字串构成的线结构得以显现。设原始倾斜图像为 $f(x, y)$,通过方差为 $t = \sigma^2$ 的高斯滤波器进行滤波(也称扩散或模糊),其计算公式为:

$$u(x, y, t_k) = f(x, y) * G_{t_k} \quad (1)$$

其中

$$G_{t_k} = \frac{1}{2\pi t_k} \exp\left\{-\frac{(x^2 + y^2)}{2t_k}\right\}$$

$u(x, y, t_k)$ 代表扩散后的图像, $k=1, 2, \dots, n$ 对应着 n 个尺度 t_k ,符号“*”表示卷积运算。

通过对尺度 t_k 的控制,可以得到具有理想线结构信息的图像。如图1所示,随着尺度 t_k 的逐渐增大, $u(x, y, t_k)$ 图像的模糊程度越来越严重,线结构在 t_k 为某个优化值出现,如图1(b),转而随 t_k 的进一步增大而消失,如图1(c)所示。

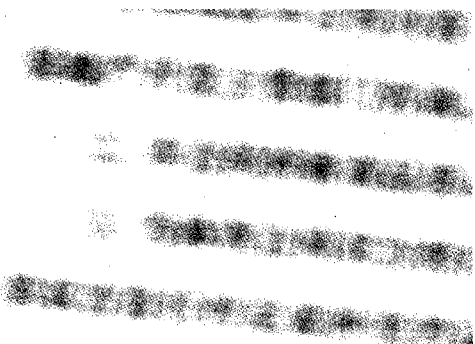
2.2 线结构的定义

因为图像灰度 $f(x, y)$ 的梯度向量 $(f_x, f_y)^T =$

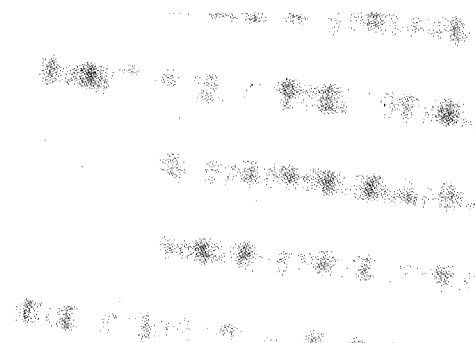
$(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})^T$ 能够反映线结构和方向信息, 因此, 在某像素点的一个邻近区域内, 如果梯度向量有统一的方向, 就定义图像在该像素点具有线结构^[5]。

房地产开发主管部门咨询。
2. 本合同所称商品房
3. 为体现合同双方的
供双方自行约定或补充约

(a) 原始图像



(b) $t=26$ 时的图像



(c) $t=60$ 时的图像

图 1

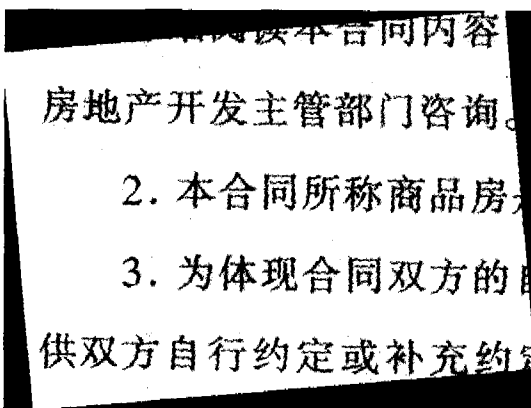


图 2 纠偏后的图像

为了在像素点的邻近区域对线结构信息进行分析, 利用梯度函数 f_x 和 f_y 定义线结构分析张量如下:

$$J_\rho(f(x, y)) = \begin{pmatrix} G(x, y, \rho^2) * (f_x)^2 & G(x, y, \rho^2) * f_x f_y \\ G(x, y, \rho^2) * f_x f_y & G(x, y, \rho^2) * (f_y)^2 \end{pmatrix} \quad (2)$$

其中 $G(x, y, \rho^2)$ 是方差为 ρ^2 的高斯函数。 ρ 的取值决定邻近区域尺寸的大小。 该分析张量的特征向量反映了梯度向量在其邻近区域的两个基本方向, 其特征值反映梯度分布的叛离程度。

假设用 λ_1 和 λ_2 ($\lambda_1 > \lambda_2$) 表示线结构分析张量的特征值。 则当 $\lambda_1 \gg \lambda_2$ 时, 说明梯度向量的方向分布迥然不同, 即意味着图像在 (x, y) 点很可能存在一个线结构。 这种可能的程度, 我们用线似然值 $S(x, y)$ 来表示, 定义如下:

$$S(x, y) = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \quad (3)$$

显然, $S(x, y)$ 取值在 $[0, 1]$ 之间。 当 $S(x, y) \approx 1$ 时, 像素点 (x, y) 及其邻近区域在很大程度上存在线结构; 而当 $S(x, y) = 0$ 时, 情况正好相反, 即 (x, y) 及其邻近区域存在线结构的可能性很小。 一般来讲, $S(x, y)$ 值的大小与 (x, y) 及其邻近区域存在线结构的可能性成正比。

假设用 v_1 和 v_2 表示特征值 λ_1 和 λ_2 对应的特征向量, 则当 $S(x, y) \approx 1$ 时, 特征向量 v_1 的方向与 (x, y) 处线结构的方向垂直, 而特征向量 v_2 的方向与 (x, y) 处线结构的方向平行。 因此, 图像在 (x, y) 处, 线结构的方向就等于 v_2 的方向。 其计算公式为:

$$\text{angle}(x, y) = \arctg \frac{v_2(x, y, 2)}{v_2(x, y, 1)} \quad (4)$$

2.3 多尺度线结构分析

由于不同滤波参数 t 会导致不同效果的滤波图像, 而线结构的检测是基于滤波后的图像的, 因此, 像素点的线似然值与 t 相关, 可用 $S(x, y; t)$ 表示, 其中 x 和 y 代表像素点的坐标值, t 代表不同尺度。 因此 $S(x, y; t)$ 是多尺度下的一个线结构信息。

多数资源档案中, 字体的大小和方向是一致的, 这样 $S(x, y; t)$ 只有一个峰值, 对应着同一个线结构方向。 取图 1(a) 中坐标为 (230, 360) 的像素点, 对应的线结构似然值函数 $S(x, y; t)$ 在 $t=26$ 时取得最大值, 对应的滤波图像为图 1(b)。

如果档案图像具有多级层次结构, 文本行同时存在几种大小的字体, 则线结构似然函数 $S(x, y; t)$ 可能存在多个峰值, 每个峰值代表一级结构。 对于这种情况, 把峰值按从大到小次序排序, 将所有峰值对应的滤波器对图像依次进行处理, 一个峰值就代表某个方向的全局结构化处理, 最后得到的图像能代表各个方向的全局特征, 依上面的方法同样可以进行后续步骤的计算。

3 倾斜检测算法

基于上面的定义和分析, 档案图像倾斜检测算法描述如下:

- (1) 对原始图像像素 $f(x, y)$, 通过公式(1)计算其滤波图像像素 $g(x, y; t_k), k=1, 2, \dots, n$ 。
- (2) 对滤波图像像素 $g(x, y; t_k)$, 通过公式(2)计算其结构分析张量 $J_{\rho_k}(x, y)$, 其中, $k=1, 2, \dots, n$, 且 $\rho_k = 2\sqrt{t_k}$ 。
- (3) 计算每个像素对应的结构分析张量 $J_{\rho_k}(x, y)$ 的特征值和特征向量, 且用 λ_1^k 和 λ_2^k 来表示对应的特征值, 且假设 λ_1^k

$\geq \lambda_0$ 。

(4)对每个像素点,通过公式(3)计算线结构的似然值 $S(x, y, \rho_k)$,且设 $S_{opt}(x, y) = \max_{k \in \{1, 2, \dots, n\}} S(x, y, \rho_k)$ 。

(5)设 d 为线结构似然阈值,对于 $S_{opt}(x, y) \geq d$ 的像素点,通过公式(4)计算该像素点的倾角 $angle(x, y)$ 。

(6)如果图像所有像素点还未处理完,回到步骤(1)。

(7)计算 $angle(x, y)$ 直方图,直方图峰值对应的角度就是档案图像的倾斜角。

4 实验结果

选用北京市朝阳区的国土及房产资源档案进行实验,原始图像如图1(a),它是“京朝其04字00899”号档案的第15页的区域截取子图,具有 500×700 像素,经上面描述的算法处理后,得到档案图像的倾斜度为 5° ,纠偏后的图像见图4。由图可见,该图像没有视觉倾斜感,说明算法的纠偏效果令人满意。

结论 本文基于线结构分析对政府资源档案图像的倾斜进行检测。使用像素点及其邻域内的梯度分布信息作为方向敏感特征,通过定义高斯函数的方差,实现邻域大小的调控,从而在多尺度下找到像素点的最佳线结构似然信息,进而再用线结构似然信息计算各像素点的倾角,最后用大于一定阈

值的倾角进行统计表决,以倾角直方图的峰值作为档案图像的倾角。

在算法的第二步中,线结构分析张量需要对不同的 ρ 值逐点计算,而每次线结构分析张量的计算也与 ρ^2 成正比,因此算法的时间复杂性为 $O(\sigma^2 N)$,其中 $\rho = 2\sqrt{t}$, N 是图像尺寸。为减少计算量,一种办法是不对整幅图像进行计算,而只取图像中几处文本富裕区进行计算;另一种办法是利用对档案图像已知的先验信息(比如字体大小),把 ρ_k 的取值范围缩小。当然,这两种办法会降低算法的适应性。

参考文献

- 1 Hull J J. Document Image Skew Detection; Survey and Annotated Bibliography [J]. Document Analysis System II, World Scientific, 1998. 40~64
- 2 Dasari L, Bloomberg D S. Rapid detection of page orientation [P]. U. S. Patent 5276742. January 1994
- 3 Rundle A. Optimum scan angle determining means [P]. U. S. Patent 3831146. August 1974
- 4 Lindeberg T. Scale-Space Theory in Computer Vision [M]. Kluwer, Boston, 1994
- 5 Bigun J, Granlund G H, Wiklund J. Multidimensional orientation estimation with applications to texture analysis and optical flow [J]. IEEE Trans. Pattern Anal. Machine Intell, 1991, 13(8): 775~790

(上接第171页)

5)代理层。主要由教师代理、学习者代理和学习伙伴代理组成。其中教师代理接收教师的请求与信息通过决策系统实现智能学习环境的建立与更新,同时根据学习者的学习请求在决策系统和工具系统的支持下代理智能学习环境的老师完成概念陈述、示范举例、答疑、布置作业、评价学生;学习者代理既可以是真实学习者的代理,也可以是教师以学生身份注册参与学生学习的代理,学生代理接收学习者的注册要求通过决策系统注册,接收学生信息通过决策系统提取学生特征信息,通过工具系统代理智能学习环境的学习者展示自己的学习过程;学习伙伴代理根据学习者的特征和特征信息库的信息代理智能学习环境中不同的学习伙伴,为学习者的合作学习提供支持。如图5所示。

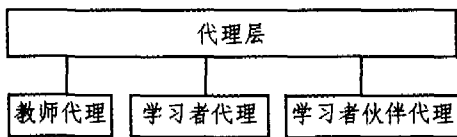


图5 代理层组成

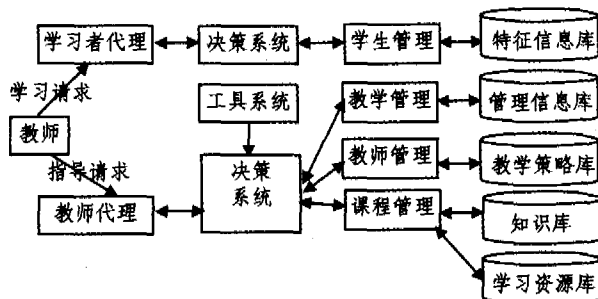


图6 教师工作流程图

4 通用智能学习环境模型的工作流程

通用智能学习环境模型的工作流程以教师工作流程和学

习者工作流程图示如下。工具系统中的导航工具被自动调用为学习者学习过程进行适应性导航,数据挖掘工具被自动调用生成学习者相应的学习材料,通信服务工具被调用帮助学习者收集网格中相应的学习资料。

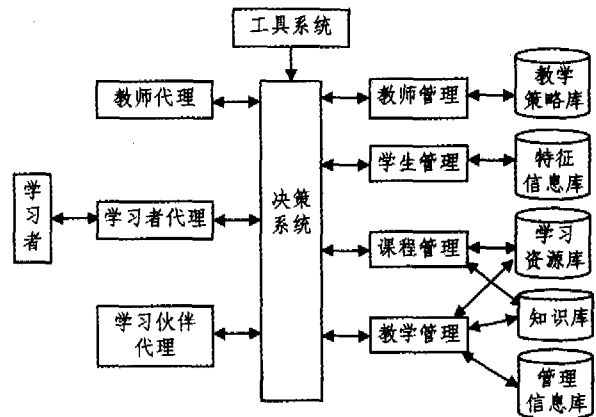


图7 学习者工作流程图

结束语 本文主要从建构主义学习理论、混合学习理论、现代教学理论出发,以学生学习为中心,构建了一个智能性、开放式、集成化学习环境模型,它既支持学生学习的自主建构,又提供适时的学习指导。其实现细节和关键技术还在研究过程中。

参考文献

- 1 李宗树,曾艳. 网格——教育信息化新视野. 见:中国人工智能学会计算机辅助教育专业委员会2005年第十二届学会年会论文集, 2005. 88~90
- 2 王陆. 虚拟学习社区原理与应用. 高等教育出版社, 2004. 63~89
- 3 Vicari R M, Flores C D, Silvestre A M, Louise J. et al. A multi-agent Intelligent Environment For Medical Knowledge, 2003
- 4 Stathacopoulou R, Magoulas G D, Grigoriadou M, et al. Neuro-fuzzy knowledge processing in intelligent learning environments for improved student diagnosis. 2005
- 5 Bouras C, Triantafilou V, Tsiatsos T. A framework for intelligent Virtual Training Environment, The steps from Specification to Design. 2002