

文章编号: 1004-4353(2017)02-0173-06

基于基本图像特征的中朝文种辨识方法

张鹏, 崔荣一*

(延边大学工学院 计算机科学与技术学科 智能信息处理研究室, 吉林 延吉 133002)

摘要: 提出了一种基于基本图像特征、适用于不同分辨率自然文本图像的中朝文种辨识方法. 在训练阶段, 首先构造标准文字图像库, 然后提取文字的形态学骨架, 最后利用骨架的基本图像特征来训练 BP 神经网络; 在辨识阶段, 首先对自然文本图像进行倾斜校正后做二值化处理以提高字符分割的效果, 然后进行垂直投影、水平投影、字符分割和提取每个字符的骨架信息, 最后基于字符骨架的基本图像特征利用 BP 神经网络来进行文种辨识. 实验表明, 本文提出的方法在中朝文种辨识中综合辨识准确率可达到 87%.

关键词: 文种辨识; 形态学骨架; 基本图像特征; BP 神经网络

中图分类号: TP391.41

文献标识码: A

Chinese and Korean script identification based on basic image features

ZHANG Peng, CUI Rongyi*

(*Intelligent Information Processing Lab., Dept. of Computer Science & Technology,
College of Engineering, Yanbian University, Yanji 133002, China*)

Abstract: A Chinese and Korean script identification method was proposed in this paper based on basic image features, which is suitable for natural images with different resolutions. In the training stage: firstly, the standard text image library was constructed. Secondly, the text morphological skeleton was extracted. finally, the BP neural network was trained using the skeleton's basic image features. In the identifying stage: firstly, the text image binarization was carried out after tilt correction to improve the effect of character segmentation. Secondly, character segmentation were carried out after vertical projection and horizontal projection; and furthermore, the morphological skeleton was extracted. Finally, using the trained BP neural network, the script identification was implemented with the basic image features. Experiment results show that the algorithm achieved an accuracy of 87% in Chinese and Korean script identification.

Keywords: script identification; morphological skeleton; basic image features; BP neural network

0 引言

图像语义理解是进行图像分类、检索、识别的重要步骤, 图像中的文字往往能为图像的语义理解提供关键信息, 而文种辨识是进行图像语义理解的前提步骤. 文种辨识方法大致可分为基于符号匹配的方法、基于统计特征的方法和基于纹理特征的方法^[1]. 基于符号匹配的方法需要准确的字符分割, 对图像质量要求较高, 基于统计特征的方法难以处理字符结构相似的文种, 而基于纹理特征的方法往往计算量较大. 近年来基于纹理特征的文种辨识方法发展迅速, 主要有基于 Gabor 滤波器的方法^[2]、基于灰度共

生矩阵的方法^[3]和基于小波变换的方法^[4]等. 现有的文种辨识方法主要针对文档图像, 并且对文字粗细、大小等有较高的要求. 本文提出一种适用于不同分辨率自然文本图像的中朝文种辨识方法, 该算法辨识准确率较高, 并省去了传统文种辨识方法中需要收集大量不同文字大小和笔划宽度的文本图像作为训练样本的步骤.

1 相关理论

1.1 形态学骨架

骨架是用与原始形状连通性和拓扑结构相一致的细曲线表达物体形状^[5]. 如图 1 所示, 如果 z 是 $S(A)$ 中的一点, 并且 $(D)_z$ 是在 A 内以 z 为中心的一个最大圆, 则不存在位于 A 内的能包含 $(D)_z$ 的更大圆, 圆 $(D)_z$ 称为最大圆盘. 圆 $(D)_z$ 在两个或更多的不同位置上与 A 的边界接触.

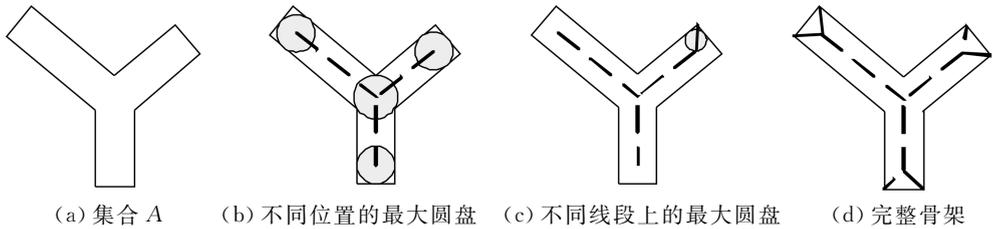


图 1 形态学骨架示意图

A 的骨架可用腐蚀和开操作表达, 即骨架可以表达为

$$S(A) = \bigcup_{k=0}^K S_k(A), \quad (1)$$

而

$$S_k(A) = (A!kB) - (A!kB) \circ B, \quad (2)$$

其中: B 是一个结构元素; $(A!kB)$ 表示对 A 的连续 k 次腐蚀, $(A!kB) = (\dots((A!B)!B)! \dots)!B$; 第 k 次是 A 被腐蚀为空集合前进行的最后一次迭代, 即 $K = \max\{k \mid (A!kB) \neq \emptyset\}$. 式(1)和式(2)说明 $S(A)$ 可以由骨架子集 $S_k(A)$ 的并集得到, 即 A 可以利用式(3)重构:

$$(S_k(A) \oplus kB) = ((\dots(S_k(A) \oplus B) \oplus B) \oplus \dots) \oplus B. \quad (3)$$

1.2 基本图像特征

基本图像特征^[6]是 Griffin 等提出的一种多尺度纹理分析方法, 该方法将图像的纹理基元划分为 7 种不同类型, 这 7 种纹理基元的集合称为基本图像特征. 基本图像特征是根据二阶高斯衍生滤波器组的响应来划分 7 种不同纹理基元. 高斯衍生滤波器组对图像的局部特性十分敏感, 能很好地反应出不同几何结构的纹理区域^[7-9]. 高斯衍生滤波器组由 6 个高斯衍生滤波器组成, 如图 2 所示.

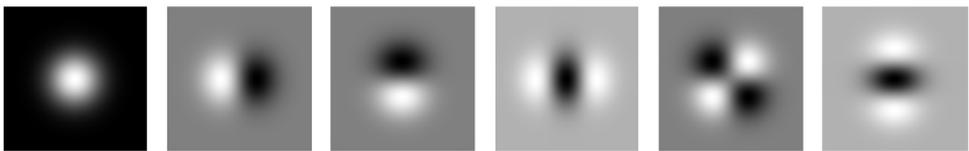


图 2 高斯衍生滤波器组

高斯衍生滤波器的一维表达式为:

$$G^\sigma(x) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{x^2}{2\sigma^2}}, \quad (4)$$

$$G_n^\sigma(x) = \frac{d^n}{dx^n} G^\sigma(x) = \left(\frac{-1}{\sigma\sqrt{2}}\right)^n H_n\left(\frac{x}{\sigma\sqrt{2}}\right) G^\sigma(x), \quad (5)$$

其中上标 σ 为滤波器的尺度参数,下标 n 表示 $G^\sigma(x)$ 的 n 阶导数, H_n 为埃尔米特多项式. 将其推广到二维后,可得

$$G_{mn}^\sigma(x, y) = g_m^\sigma(x)g_n^\sigma(y). \tag{6}$$

图像 $I(x, y)$ 与高斯衍生滤波器 $G_{mn}^\sigma(x, y)$ 的响应记为 $j_{mn}(x, y)$:

$$j_{mn}(x, y) = \langle G_{mn}^\sigma(x, y) | I(x, y) \rangle, 0 \leq m + n \leq 2.$$

计算响应 $j_{mn}(x, y)$ 的 2 个参数: $\lambda = \sigma^2(j_{20} + j_{02})$, $\gamma = \sigma^2 \sqrt{(j_{20} - j_{02})^2 + 4j_{11}^2}$. 然后取公式(7)中 7 个参数中的最大值来确定像素点所对应的纹理单元:

$$\{\epsilon j_{00}, 2\sigma \sqrt{j_{10}^2 + j_{01}^2}, \lambda, -\lambda, 2^{-1/2}(\gamma - \lambda), 2^{-1/2}(\gamma + \lambda), \gamma\}, \tag{7}$$

其中 ϵ 为固定参数,取值范围为(0, 1). 如图 3 所示, ϵj_{00} 对应于图像中的平坦型结构, $2\sigma \sqrt{j_{10}^2 + j_{01}^2}$ 对应于图像中的坡型结构, λ 对应于图像中的暗点型结构, $-\lambda$ 对应于图像中的明点型结构, $2^{-1/2}(\gamma - \lambda)$ 对应于图像中的明线型结构, $2^{-1/2}(\gamma + \lambda)$ 对应于图像中的暗线型结构, γ 对应于图像中的鞍型结构.

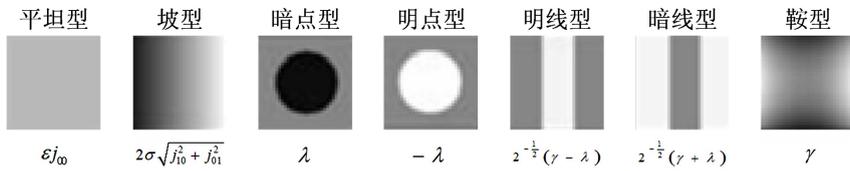


图 3 7 种基本图像特征

2 算法设计

基本图像特征具有尺度不变性、旋转不变性和仿射不变性,在目标识别、纹理分类等领域得到了广泛的应用,并取得了很好的效果. 本文利用中朝文字纹理结构的差异性,将基本图像特征应用到文种辨识领域,以此来描述这两种语言文字的结构特点,并提出基于基本图像特征的文种辨识方法. 本文提出的文种辨识方法流程图如图 4 所示.

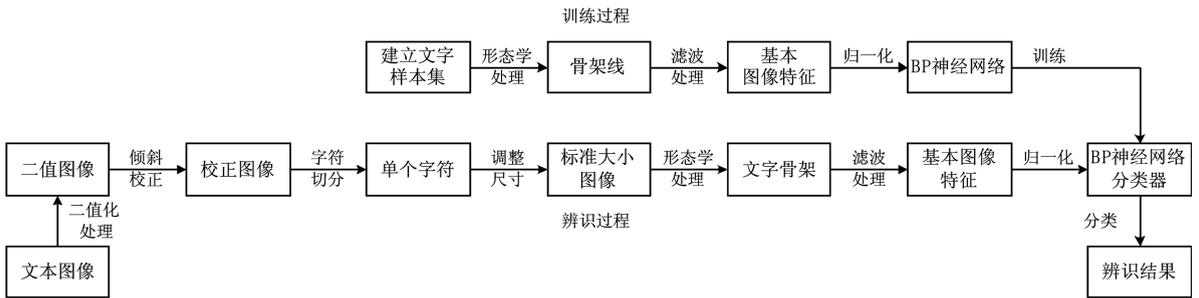


图 4 文种辨识实验流程图

2.1 训练过程

训练过程主要分为 3 部分,即生成标准文字样本图像库、标准文字样本特征提取与训练 BP 神经网络,如图 5 所示. 训练过程的具体步骤如下:

- 1) 根据字频统计分别选取中文和朝鲜文中前 1000 个常用字;
- 2) 使用程序自动生成中文和朝鲜文每个常用字的 100×100 分辨率图像;
- 3) 通过形态学处理提取每个标准文本图像的骨架形成文字骨架;
- 4) 使用高斯衍生滤波器提取每个标准文字骨架的基本图像特征,为方便特征表示与处理,以 1 表示平坦型结构,2 表示坡型结构,以此类推;
- 5) 将基本图像特征归一化后作为训练数据输入 BP 神经网络进行训练.

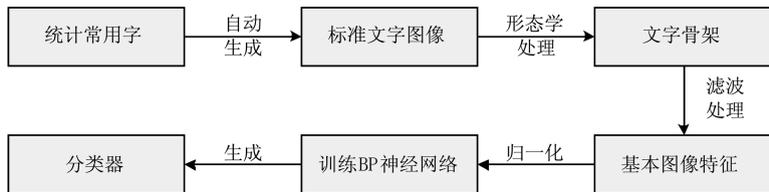


图 5 训练过程

2.2 辨识过程

辨识过程主要分为 5 部分,即倾斜校正、字符分割、骨架提取、基本图像特征提取及 BP 神经网络分类,如图 6 所示. 辨识过程的具体步骤如下:

- 1) 将自然文字进行倾斜校正处理,得到文字按水平排列的图像;
- 2) 把步骤 1)中得到的图像进行二值化处理得到二值化图像;
- 3) 把步骤 2)中得到的二值化图像分别进行垂直投影和水平投影,根据投影的平均值进行字符分割,得到单个字符图像;
- 4) 将每个字符图像的大小调整为与标准文字图像大小相同的 100×100 分辨率;
- 5) 对单个字符图像进行形态学处理,形成文字骨架;
- 6) 使用高斯衍生滤波器提取每个字符文字骨架的基本图像特征;
- 7) 将每个字符的基本图像特征归一化后输入 BP 神经网络进行分类;
- 8) 以多数类别决定该文字的文种.

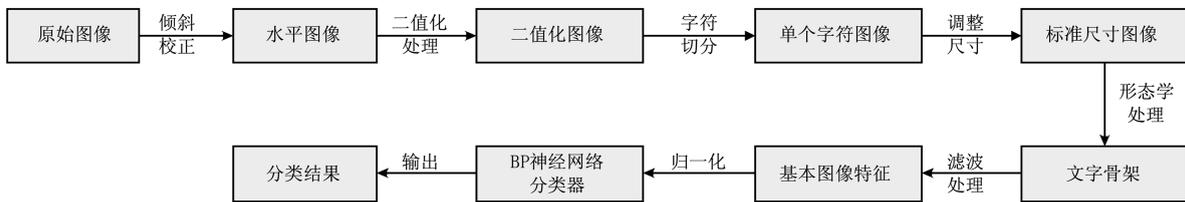


图 6 辨识过程

3 实验结果及分析

实验平台为 Win10,程序开发环境为 Matlab 2016b,实验对象为不同分辨率的自然文本图像,如图 7 所示.



图 7 部分实验图像

3.1 训练阶段

首先使用程序自动生成 100×100 分辨率的常用字图像,然后进行形态学处理提取文字骨架,再使用高斯衍生滤波器提取每个字符文字骨架的基本图像特征,最后将基本图像特征归一化后输入 BP 神经网络进行训练,如图 8 所示.

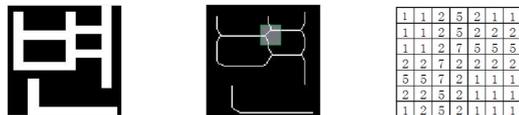


图 8 基本图像特征提取

经过高斯衍生滤波器处理得到的基本图像特征

为 100×100 的特征矩阵, 归一化为 $1 \times 10\,000$ 的一维特征向量. BP 神经网络的传递函数选用 Sigmoid 函数, 学习算法为弹性梯度下降法, 最大训练次数为 5 000, 学习率为 0.05, 目标精度为 0.005, 隐含层节点数使用经验公式(13)^[10] 确定为 14.

$$n = \log_2 x, \quad (13)$$

其中 n 为隐含层节点数目, x 为输入层节点数目.

3.2 辨识阶段

对测试的自然文本图像, 首先进行倾斜校正处理, 如图 9 所示. 对倾斜校正后的图像进行二值化处理, 然后进行垂直投影和水平投影, 最后根据字符平均宽度、高度及质心对字符进行分割, 所得结果如图 10 所示. 得到分割好的字符后, 将字符图像尺寸调整为 100×100 分辨率, 然后进行形态学处理提取骨架信息, 得到字符骨架如图 11 所示. 使用高斯衍生滤波器分别对每幅字符骨架提取基本图像特征, 如图 12 所示. 提取每幅字符骨架的基本图像特征并归一化后, 输入到训练好的 BP 神经网络中进行分类, 根据多数类别来决定最终的文种.



(a) 原始图像 (b) 倾斜校正后图像

图 9 图像倾斜校正



图 10 字符分割结果

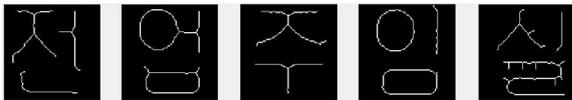


图 11 字符骨架



(a) 文字骨架 (b) 基本图像特征

图 12 基本图像特征提取

3.3 结果分析

将本文算法与文献[11]中的算法进行实验对比, 实验对象为包含中文和朝鲜文的自然文本图像各 100 幅, 实验对比结果如表 1 所示. 从表 1 可看出, 本文提出的文种辨识算法对中文的辨识准确率为 88%, 对朝鲜文的辨识准确率为 86%, 综合辨识准确率为 87%, 均高于文献[11]中的算法, 这表明该算法具有较高的辨识度. 中文和朝鲜文的辨识准确率不同的原因是: 中文和朝鲜文结构的相似性、调整字符图像尺寸时文字可能产生的变形以及文字骨架的提取效果不同.

表 1 文种辨识结果

算法	正确辨识率/%		综合辨识率/%
	中文	朝鲜文	
本文算法	88	86	87
文献[11]算法	81	81	81.5

4 结论

本文提出的基于基本图像特征、适用于不同分辨率自然文本图像的中朝文种辨识方法, 其中朝文种辨识中综合辨识准确率达到 87%, 优于文献[11]的结果. 该算法不需要人工收集大量的训练样本, 对自然文本图像限制小, 提高了文种辨识的可用性与易用性. 如何针对手写字体、不规整字体实现有效的文种辨识, 是今后进一步研究的内容.

参考文献:

[1] 顾立娟, 平西建, 程娟, 等. 一种具有旋转鲁棒性的文字文种识别方法[J]. 中国图象图形学报, 2010, 15(6): 879-886.

- [2] 陆小川,伊兵哲,平西建,等. 含噪文本图像的中英文文种识别研究[J]. 计算机工程与设计, 2007, 28(21): 5150-5152.
- [3] Busch A, Boles W W, Sridharan S. Texture for script identification[J]. IEEE Trans Pattern Anal Mach Intell, 2005, 27(11): 1720-1751.
- [4] 金璟璇,崔荣一,崔旭. 基于小波统计特征的行块级朝汉文种辨识[J]. 延边大学学报(自然科学版), 2013, 39(4): 277-280.
- [5] Rafael C Gonzalez, Richard E Woods. 阮秋琦译. 数字图像处理[M]. 2 版. 北京: 电子工业出版社, 2007.
- [6] 郭龙,平西建,周林,等. 基本图像特征用于文本图像文种识别[J]. 应用科学学报, 2011, 29(1): 56-60.
- [7] Griffin L D. The second order local-image-structure solid[J]. IEEE Trans Pattern Anal Mach Intell, 2007, 29(8): 1355-1366.
- [8] Griffin L D, Lillholm M. Symmetry sensitivities of derivative-of-gaussian filters[J]. IEEE Trans Pattern Anal Mach Intell, 2010, 32(6): 1072-1082.
- [9] Crosier M, Griffin L D. Texture classification with a dictionary of basic image features[C]//2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008). Anchorage, Alaska, USA: 2008.
- [10] Bello M G. Enhanced training algorithms, and integrated training/architecture selection for multilayer perceptron networks[J]. IEEE Trans Neural Networks, 1992, 3(6): 864-875.
- [11] 朴明姬. 自然图像中文字语种辨识方法的研究[D]. 延吉: 延边大学, 2014: 25-29.

~~~~~  
(上接第 172 页)

- [5] 陈文兵,成海燕,陈允杰. 一种基于轮廓特征点的图像检索方法[J]. 计算机工程, 2012, 38(12): 197-200.
- [6] 吕彬. 基于全局直方图及其空间关系的图像检索技术[D]. 太原: 太原理工大学, 2014: 23-25.
- [7] 周同驰,程旭,吴镇扬. 分层树结构字典编码的行为识别[J]. 中国图象图形学报, 2014, 19(7): 1054-1061.
- [8] Rao L Koteswara, Rao D Venkata. Local quantized extrema patterns for content-based natural and texture image retrieval[J]. Human-Centric Computing and Information Sciences, 2015, 26(5): 1-24.
- [9] Cohen Andrew R, Vitanyi Paul M B. Normalized compression distance of multiples[J]. Physics, 2012, 41(13): 121-130.
- [10] Costa Paulo, Barroso João, Fernandes Hugo, et al. Using Peano-Hilbert space filling curves for fast bidimensional ensemble EMD realization[J]. EURASIP Journal on Advances in Signal Processing, 2012, 2012(1): 113-125.
- [11] Yang Peng, Xie Li, Liu Jilin. Simultaneous celestial positioning and orientation for the lunar rover[J]. Aerospace Science and Technology, 2014, 34(3): 45-54.
- [12] Besiris D, Zigouris E. Dictionary-based color image retrieval using multiset theory[J]. Journal of Visual Communication and Image Representation, 2013, 24(7): 1155-1167.
- [13] 谭文明,李斌,张文聪. 基于中心对称局部二值模式的背景建模方法研究[J]. 中国科学技术大学学报, 2010, 40(11): 1112-1117.