

一种具有旋转鲁棒性的文本图像文种识别方法

顾立娟¹⁾ 平西建¹⁾ 程娟¹⁾ 郝玉保^{2),3)}

¹⁾(解放军信息工程大学, 郑州 450002) ²⁾(信息工程大学 测绘学院, 郑州 450052) ³⁾(75719 部队, 武汉 430074)

摘要 针对目前用于文本图像文种识别的纹理特征描述子对文字行倾斜缺乏不变性,采用可控金字塔变换提取文本图像的纹理特征,通过对特征空间元素重新排列,提出一种对文字行倾斜具有鲁棒性的文本图像文种识别方法。不同倾斜角度文本图像的文种识别结果表明,该算法具有较高的识别准确率并对文字行倾斜具有较强的鲁棒性。

关键词 文种识别 可控金字塔变换 纹理特征 文本图像

中图法分类号: TP391.1 文献标志码: A 文章编号: 1006-8961(2010)06-879-08

A Robust Rotation-invariant Script Identification Method of Document Images

GU Lijuan¹⁾, PING Xijian¹⁾, CHENG Juan¹⁾, HAO Yubao^{2),3)}

¹⁾(PLA Information Engineering University, Zhengzhou 450002)

²⁾(Institute of Surveying and Mapping, Information Engineering University, Zhengzhou 450052)

³⁾(75719 Troops, Wuhan 430074)

Abstract Script identification is significant for attaining information from document images. Most algorithms on texture feature extraction from document images for script identification are inadaptable to the skew of text line presently. For the skew of text line is inevitably, a new algorithm robust to the skew of text line is proposed. Steerable Pyramid transform is used on the document images and the energy statistical features of sub-bands is extracted. Through the realignment of features, the algorithm implements robustness to rotation. Libsvm is used as a classifier. The experiments are conducted on image database containing ten scripts that are scanned from books or magazines. The test samples are rotated with different angles and the results confirm that the algorithm can identify scripts accurately and is robust to the skew of text line simultaneously.

Keywords script identification, Steerable Pyramid transform, texture feature, document images

0 引言

随着网络通信技术和信息处理技术的迅猛发展,人类社会进入了一个数字化信息时代。由文件、档案等文字资料通过图文扫描仪、数码相机等图像数据采集设备生成的,以静止图像编码方式存储的文本图像作为人们传递信息的主要媒介,在信息处

理和传输系统中具有不可替代的作用。随着全球化步伐的加快,网络中广泛使用的文本图像不再由单一的语言文字组成,多种语言文字需要识别和处理。文本图像文字种类的自动识别是对以图像形式存储的各种文字,提取能用于计算机识别的底层特征,实现不同文种的自动划分。在海量的信息处理中,作为文本图像分析的重要组成部分和 OCR(光学字符识别)系统的前端处理技术,文种识别技

基金项目:国家自然科学基金项目(60473022)

收稿日期:2009-01-08;改回日期:2009-03-17

第一作者简介:顾立娟(1982—),女。现为信息工程大学在读信号与信息处理专业硕士研究生。主要研究方向为图像信息处理与模式识别。E-mail: gljmxbd@163.com

术变得尤为重要。

目前,针对语言文字种类识别进行的研究可以划分为基于统计特征、基于符号匹配和基于纹理特征的文种识别三大类。基于统计特征的文种识别算法主要有基于文本行灰度投影法^[1-4]、基于字符上凹面投影法^[5]以及基于字符形状编码法^[6]等。基于符号匹配的文种识别算法主要有基于聚类的模板匹配法^[7]和基于特殊字符的模板匹配法^[1]等。基于统计特征和符号匹配的文种识别算法具有较高的识别准确率,但大多以准确的文本行划分和字符分割为前提,对文本图像的倾斜、噪声等鲁棒性比较差。近年来基于纹理特征的文种识别算法逐渐成为研究热点。主要算法有基于灰度级共生矩阵法(GLCM)^[8]、基于 Gabor 滤波器法^[9]、基于小波变换法^[10]、基于小波的灰度级共生矩阵法^[8]以及基于小波共生直方图的方法^[11]。Gabor 滤波在纹理特征提取上具有良好的效果,但是计算量大;小波滤波计算速度很快,但由于其变换本身不具有旋转不变性,当文字行倾斜时识别率降低;基于小波共生直方图的方法提高了对文字行倾斜文本图像的文种识别率,但是提取了 384 维特征矢量,特征维数较大,给文种分类造成了一定困难。

鉴于以上基于纹理特征的文种识别算法中存在的一些问题,采用可控金字塔(Steerable Pyramid)变换^[12],提出一种新的文本图像文种识别方法。可控金字塔变换是一种多分辨率分析方法,它能独立分解出不同尺度、不同方向的子带信息,没有频谱混叠现象,子带方向具有可调节性。并且,通过对特征空间采取一些操作,可以实现旋转不变性。采用可控金字塔变换对文本图像进行分解,深入研究文本图像分解后细节子带的能量特征,利用可控金字塔子带的可控性,通过对特征空间元素重新排列,提取对旋转具有鲁棒性的 24 维特征矢量。用 Libsvm 分类器进行文本图像的文种识别和分类。以具有不同倾斜角度的多文种文本图像库为测试图像集,实验结果表明,本文方法在保持较高识别性能的前提下对于文本图像的文字行倾斜以及噪声具有较强的鲁棒性。

1 可控金字塔变换原理

可控金字塔变换是一种多分辨率分析方法,类似于 2 维小波多分辨率分析的塔式结构,可控金字

塔的塔式结构可以将图像分解成不同尺度、不同方向的子带信息,其频率分解如图 1 所示。

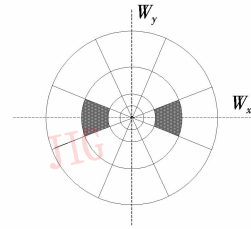


图 1 可控金字塔频率分解示意图

Fig. 1 Decomposition of Steerable Pyramid in frequency domain

可控金字塔分解过程是通过一组方向可控滤波器实现的。方向可控滤波器是一类滤波器,其任意方向的滤波器都是具有较高光滑度的基方向滤波器的线性组合。设 $\psi^\theta(x, y)$ 是函数 $\psi(x, y)$ 旋转 θ 后的函数, $\psi^{\theta_i}(x, y)$ 是 $\psi(x, y)$ 旋转 θ_i 后的一组基函数, $k_i(\theta)$ 是相应的差值函数 ($i = 1, \dots, M$), M 是基函数的个数,则有

$$\psi^\theta(x, y) = \sum_{i=1}^M k_i(\theta) \psi^{\theta_i}(x, y) \quad (1)$$

把 $\psi(x, y)$ 按傅里叶级数展开成极坐标形式:

$$\psi(r, \phi) = \sum_{n=-N}^N a_n(r) e^{in\phi} \quad (2)$$

式中, $r = \sqrt{x^2 + y^2}$, $\phi = \arg(x, y)$, 则存在如下的方向条件:

$$\begin{bmatrix} 1 \\ e^{i\theta} \\ \vdots \\ e^{iN\theta} \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ e^{i\theta_1} & \dots & e^{i\theta_M} \\ \vdots & & \vdots \\ e^{iN\theta_1} & \dots & e^{iN\theta_M} \end{bmatrix} \begin{bmatrix} k_1(\theta) \\ k_2(\theta) \\ \vdots \\ k_M(\theta) \end{bmatrix} \quad (3)$$

当 $k_i(\theta)$ 是式(3)的解时,式(1)成立。同时有

$$\psi^\theta(r, \phi) = \sum_{i=1}^M k_i(\theta) \varphi_i(r, \phi) \quad (4)$$

$\varphi_i(r, \phi)$ 为任意一组基函数, θ_i 在 $(0, \pi)$ 之间等间隔分布。

图 2 为方向可控金字塔的结构, $H_0(w)$ 为高通滤波器, $L_0(w)$ 和 $L_1(w)$ 为低通滤波器, $B(w)$ 为方向带通滤波器。首先通过一个高通和低通滤波器 (H_0, L_0) 将图像分解为高通和低通子带,接下来低通子带被进一步分解为 n 个方向的带通部分 B_1, \dots, B_n 和一个低通部分 L_1 。下一级分解将通过 L_1 行、列下采样后迭代进行。可控金字塔同级滤波带没有下采样操作,具有 $4n/3$ 的计算冗余。

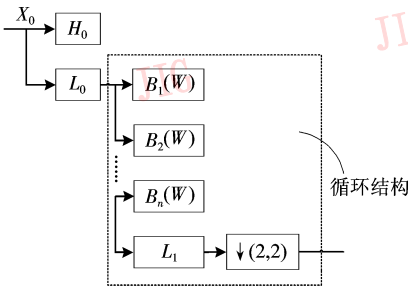


图 2 可控金字塔的多分辨率分解

Fig.2 Multi-resolution decomposition of Steerable Pyramid

在可控金字塔变换频带划分中,各方向子带的频域支撑区均为扇形,更符合人眼视觉特性。没有明显的频谱混淆现象,即同一方向、同一频率的信息一般只在同一个方向子带中出现,从而具有较好的频域局部性和方向选择性,可以通过对变换后方向子带的控制实现对旋转的鲁棒性。

一幅图像 $I(x, y)$ 的可控金字塔变换为

$$W_j^k(x, y, \theta_k) = I * \psi_j^n(x, y, \theta_k) \quad (5)$$

式中,可操纵小波 $\psi_j^n(x, y, \theta_k)$ 为一方向性平滑函数 $\beta_j(x, y, \theta_k)$ 的 n 阶导数, j 为尺度参数, $k = 1, 2, \dots, n$, 所以其模为

$$|W_j^k| = \left(\left(\sum_{k=1}^n W_j^k I(x, y, \theta_k) \cos(w_k) \right)^2 + \left(\sum_{k=1}^n W_j^k I(x, y, \theta_k) \sin(w_k) \right)^2 \right)^{1/2} \quad (6)$$

式中, $w_k = k\pi / (n + 1)$, $k = 1, 2, \dots, n$ 。

建立三阶方向可控金字塔,滤波器采用“sp3Filters”滤波器组^[13]。“sp3Filters”是采用加权最小二乘法设计出的具有三阶方向微分性质的 2 维滤波器组,包含高通滤波器 H_0 ,低通滤波器 L_0 和 L_1 ,带通滤波器 B_k ,大小分别为 $15 \times 15, 9 \times 9, 17 \times 17, 9 \times 9$ 。其中 B_k 为 4 个方向的带通滤波器, $k = 1, 2, 3, 4$,方向角度为 $\theta_k = (k - 1)\pi/4$ 。

2 文本图像纹理特征分析

不同文种的文字主要由圆弧形结构基元和线段结构基元按照某种组合关系排列而成。文种的区别主要体现在圆弧的曲率和相对位置,线段的长短和相对方位,以及这二者之间的组合比例上。如,英文字母大多是由竖直笔画和弧形轮廓构成,汉字构字则讲究横竖撇捺,韩国文字既有横竖笔画又存在圆形轮廓,拉丁语系的文字则一般由斜线和弧线构成。

结构基元的纹理特征一定程度上代表了文字的纹理特征,所以这里首先对结构基元在可控金字塔变换域的纹理特征进行分析。

结构基元按照 7 种不同的组合方式排列而成的纹理模式如图 3 所示,将其按式(5)进行三级可控金字塔分解,计算其不同尺度及方向子带的能量和标准差:

$$E_j^k = \sum_x \sum_y |W_j^k(x, y)| \quad (7)$$

$$\sigma_j^k = \sqrt{\frac{1}{N^2} \sum_x \sum_y (|W_j^k(x, y)| - \mu_j^k)^2} \quad (8)$$

式中, W_j^k 为分解后的细节子带, j 为分解级数, k 表示分解方向, N 为子带尺寸。统计每种纹理模式分解子带能量和标准差的条形图。图 3 中,条形图的 12 条柱,前 4 个是第一级分解 4 个子带($0^\circ, 45^\circ, 90^\circ, 135^\circ$)的特征,往右依次为第二、三级分解的 4 个子带的特征。

由图 3 可以看出,不同结构元素的能量和能量标准差在方向上的分布存在差异:结构比较对称的元素方向差异较小,如圆形结构;十字结构在 $45^\circ, 135^\circ$ 方向能量和能量标准差较大; \times 结构在 $0^\circ, 90^\circ$ 方向能量和能量标准差较大。能量大是由于此方向上可以捕捉原图两个方向的能量,而标准差大则是因为滤波方向与原图结构边缘之间的夹角变化较快,能量分布相对分散,可控金字塔变换捕捉到这种差异并将其在相应方向的能量分布中反映出来;对于米字、十字和圆形、 \times 和圆形、米字和圆形组合成的 4 种结构,4 个方向的能量差异很小甚至相同,而它们的不同方向标准差特征之间的差别依然较大:如十字和圆形组合结构在 $45^\circ, 135^\circ$ 方向能量标准差较大, \times 和圆形组合结构在 $0^\circ, 90^\circ$ 方向能量标准差较大,原因在于虽然这些组合结构基元在结构上比较对称,但其能量分布粗糙程度在不同的方向存有差异。所以综合利用能量和能量标准差描述结构基元的特征效果更好。

同时我们注意到, \times 结构相对于十字结构、十字和圆形的组合相对于 \times 和圆形的组合,均旋转了 45° ,而它们的能量和能量标准差特征值也对应发生相应的循环移位。如果将能量值最大的方向子带的特征值移到第 1 位,其他的顺次循环,则这种差异将被大大减小甚至消除。由此可见,对可控金字塔变换的旋转不变性,将特征空间依据能量最大的方向重新进行排序,将会得到对旋转具有鲁棒性的特征

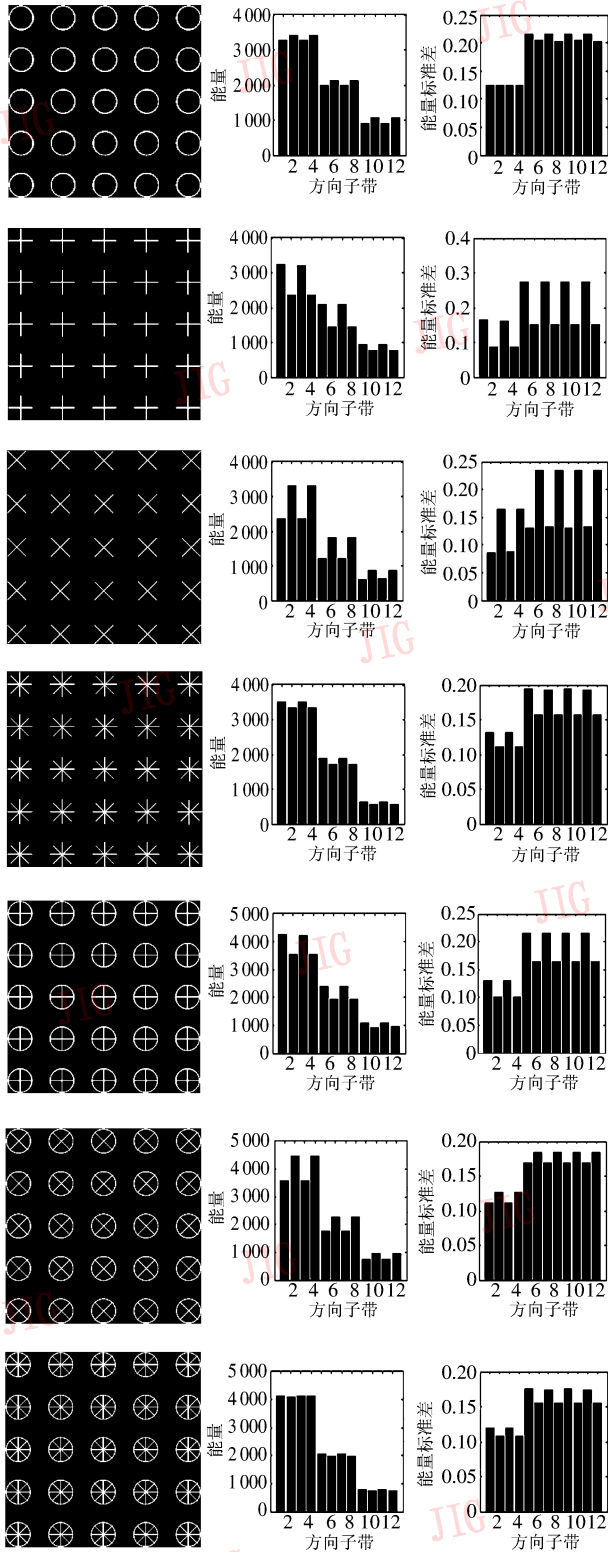


图 3 文字结构基元(左)及其可控金字塔分解细节子带的能量和能量标准差条形图(右)

Fig. 3 Character structure elements(left) and the bars of energy mean and standard variance of Steerable Pyramid sub-bands(right)

值,而这种特性对于文本图像文字行出现倾斜时的文种识别具有重要意义。

3 文本图像纹理特征提取和文种识别

3.1 文本图像的可控金字塔能量统计纹理特征

设有文本图像 $I(x, y)$, $x, y = 1, 2, \dots, N$, 对其进行能量归一化处理:

$$H(x, y) = \frac{I(x, y)}{\left(\frac{1}{N^2} \sum_{x=1}^N \sum_{y=1}^N I^2(x, y)\right)^{1/2}} \quad (9)$$

按照式(5)对 $H(x, y)$ 进行三级可控金字塔分解,每一级分解带通部分取 4 个方向,共得到 12 个细节子带。图 4 为一幅英文图像进行三级 4 个方向可控金字塔分解的结果,从右上角到左下角依次为 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 方向子带。

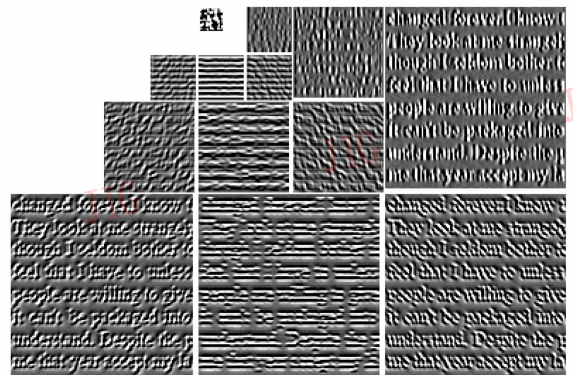


图 4 英文图像三级可控金字塔分解

Fig. 4 Three levels decomposition of Steerable Pyramid for English script image

为了缩小特征值之间的数量差异,取能量均值和标准差作为文本图像的纹理特征:

$$\mu_j^k = \frac{1}{N^2} E_j^k \quad (10)$$

据式(7)、(8)、(10)计算各子带的能量均值和标准差:

$$R = \begin{bmatrix} \mu_1^1 & \sigma_1^1 & \mu_1^2 & \sigma_1^2 & \mu_1^3 & \sigma_1^3 & \mu_1^4 & \sigma_1^4 \\ \mu_2^1 & \sigma_2^1 & \mu_2^2 & \sigma_2^2 & \mu_2^3 & \sigma_2^3 & \mu_2^4 & \sigma_2^4 \\ \mu_3^1 & \sigma_3^1 & \mu_3^2 & \sigma_3^2 & \mu_3^3 & \sigma_3^3 & \mu_3^4 & \sigma_3^4 \end{bmatrix} \quad (11)$$

由于图像的旋转对应其可控金字塔变换不同方向子带的能量发生平移,所以计算细节子带能量最大的方向作为主控方向 DO_i :

$$DO_i = \max \{E^1, E^2, E^3, E^4\} \quad (12)$$

$$E^k = \sum_{j=1}^3 E_j^k, k = 1, 2, 3, 4 \quad (13)$$

i 表示能量最大方向的索引(假设 $i=2$)。

以主控方向为基准对 R 进行循环移位:

$$R' = \begin{bmatrix} \mu_1^2 & \sigma_1^2 & \mu_1^3 & \sigma_1^3 & \mu_1^4 & \sigma_1^4 & \mu_1^1 & \sigma_1^1 \\ \mu_2^2 & \sigma_2^2 & \mu_2^3 & \sigma_2^3 & \mu_2^4 & \sigma_2^4 & \mu_2^1 & \sigma_2^1 \\ \mu_3^2 & \sigma_3^2 & \mu_3^3 & \sigma_3^3 & \mu_3^4 & \sigma_3^4 & \mu_3^1 & \sigma_3^1 \end{bmatrix} \quad (14)$$

将 R' 按行展开,即可得到 24 维可控金字塔能量统计纹理特征矢量。

3.2 可控金字塔能量统计纹理特征对文字行倾斜的适应能力

据以上分析,经过循环移位的可控金字塔能量统计纹理特征对文字行的旋转具有鲁棒性。建立包含英、中、俄、日、韩、阿拉伯、印地、缅甸、柬埔寨、藏 10 种文种的图像库。从中抽取英文图像 50 幅作为测试图像,检验可控金字塔能量统计纹理特征对旋转的鲁棒性。对测试图像做某一角度的旋转,据式(14)计算每幅测试图像和旋转后图像的可控金字塔能量统计纹理特征矢量。据下式计算旋转图像与测试图像的平均欧氏距离 J :

$$J = S_b / S_w \quad (15)$$

$$S_b = ((m_1 - m_2)(m_1 - m_2)^T)^{1/2} \quad (16)$$

$$S_w = \frac{1}{2} \sum_{i=1}^2 \left(\frac{1}{50} \sum_{k=1}^{50} ((x_k^{(i)} - m_i)(x_k^{(i)} - m_i)^T) \right)^{1/2} \quad (17)$$

$$m_i = \frac{1}{50} \sum_{k=1}^{50} x_k^{(i)}, i = 1, 2 \quad (18)$$

$x_k^{(i)}$ 表示 i 类,第 k 个样本的特征矢量, $m_i, i = 1, 2$ 分别表示测试图像和旋转后图像的样本均值, S_w 为类内离散度, S_b 为类间离散度。把旋转图像与测试图像看做两个类,则 J 值越小,特征对旋转的鲁棒性越强。

作为对比,用文献[11]提出的方法对测试图像做同样的实验。

对测试图像分别做 $0^\circ, 5^\circ, 15^\circ, \dots, 165^\circ, 175^\circ$ 及 $90^\circ, 180^\circ$ 的旋转,实验结果如图 5 所示。

由图 5 可见,由于可控金字塔变换带通部分取 4 个方向,每个子带捕获 45° 的方向信息,基于可控金字塔能量统计纹理特征的 J 在 $45^\circ, 135^\circ$ 附近值较小,在 $0^\circ, 90^\circ, 180^\circ$ 处出现极小值,且各个角度 J 值均小于文献[11]采用的方法,对旋转具有很强的鲁棒性。

3.3 基于 Libsvm 的文种识别

支持向量机^[14](SVM)通过非线性变换将输入

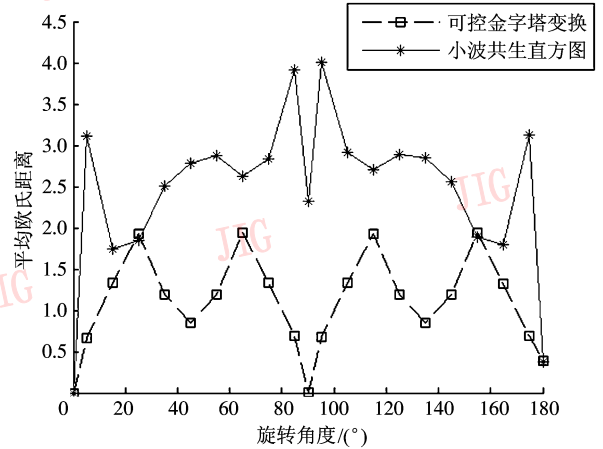


图 5 平均欧氏距离比较图

Fig.5 Comparison of average distance between Steerable Pyramid and Wavelet co-occurrence histogram

空间变换到一个高维空间,根据结构风险最小化准则,在使训练样本分类误差极小化的前提下,尽量提高分类器的泛化推广能力。Vapnik 等人在研究中发现,核函数当中的具体参数和误差惩罚因子 c 是影响 SVM 性能的关键因素。Libsvm 是一个 SVM 工具^[15],通过交叉确认法确定最佳的参数,提供平行的栅格搜索法进行参数迭代,输出每次迭代的 c , γ 值以及相对应的 CV(交叉验证)准确度的对数值。这一过程不仅可以取得最优参数,而且能够全面考虑各种输入数据的不同属性特征,避免欠学习、过学习问题的产生。同时采用分隔面模式,有效地克服了样本分布、冗余特征以及过拟合等因素的影响,具有很好的泛化能力,占有效果和稳定性上的优势。对于可控金字塔变换提取的 24 维特征矢量,对多文种分类,无论从泛化能力还是识别精度考虑,Libsvm 都是理想的分类工具。所以本实验采用 Libsvm 进行分类,将径向基函数(RBF)作为核函数。通过交叉验证法得到精确的最优参数,提高识别准确率。径向基函数为

$$K(x, x_i) = \exp\left(-\frac{1}{2\sigma^2} \|x - x_i\|^2\right) \quad (19)$$

首先对图像库中的文本图像做归一化处理,然后进行三级可控金字塔分解,得到 12 个细节子带,计算每个子带的能量均值和标准差特征,并将特征值按照主控方向重新进行排序,得到 24 维可控金字塔能量统计纹理特征矢量。用 Libsvm 对训练样本进行参数优化,并用得到的参数对测试样本进行测试,得到文种的分类结果。图 6 为整个实验流程图。

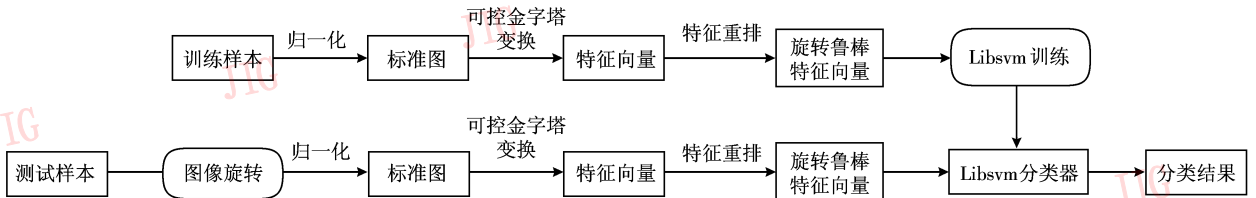


图 6 实验流程

Fig. 6 Flew chart of the whole procedures

4 实验结果与分析

4.1 实验方案

为了检验算法对无变形文本图像库的识别性能以及对旋转的鲁棒性,自建了一个包含英、中、俄、日、韩、阿拉伯、印地、缅甸、柬埔寨、藏 10 种文种各

300 幅图像的文本图像库,图像为 128 × 128 像素大小的灰度图像。库中图像从报纸、杂志和书籍上扫描得到,在采集过程中有轻微的噪声、笔画断裂等质量退化现象。图 7 展示了库中 10 幅不同文种的文本图像。每种文种取 100 幅用于训练,余下的 200 幅用于测试。



图 7 图像库中的文本图像

Fig. 7 Document images from image database

采用以下两组实验方案:

- 1) 测试图像分别做 45°, 90°, 135°, 180° 的旋转。
- 2) 测试图像无旋转以及进行 5°, 10°, 15°, ..., 40°, 45° 范围以内不等角度的旋转。

4.2 结果分析

用训练样本对 Libsvm 进行训练,得到最佳参数,用此参数对两组实验的测试样本进行测试。并与文献[11]中的方法进行比较。

实验方案 1 的识别结果如表 1 所示,实验方案 2 的识别结果如表 2 所示。

对实验方案 2 基于可控金字塔变换文种识别结果的错分样本分布进行统计。统计数据取不同的测试样本共 2 000 幅。统计结果如表 3 所示。其中, number 值为 2000 幅 A 文种图像中识别为 B 文种的

样本数量(A 为列对应的文种,B 为行对应的文种),空单元格表示数量值为 0。

由表 1 可以看出,本文算法对图像 45°, 135° 的旋转识别率相对较低,对图像 90°, 180° 的旋转具有很高的识别率。识别精度远远优于文献[11]中所用方法。

表 1 实验方案 1 的识别结果 Rate

Tab. 1 Result of the first experiment

方法	旋转角度			
	45°	90°	135°	180°
本文	41.8	93.4	45.4	93.9
文献[11]	11.4	44.8	11.1	75

表 2 实验方案 2 的识别结果
Tab.2 Result of the second experiment

文种类型	0°		5°		10°		15°		20°		25°		30°		35°		40°		45°	
	本文方法	文献[11]方法	本文方法	文献[11]方法	本文方法	文献[11]方法	本文方法	文献[11]方法	本文方法	文献[11]方法	本文方法	文献[11]方法	本文方法	文献[11]方法	本文方法	文献[11]方法	本文方法	文献[11]方法	本文方法	文献[11]方法
英文	97.5	94	75	58	71	51	66	44	54.5	39.5	48	24	49	23	49.5	22	48	17	33.5	12.5
中文	94	94	90	81.5	90	79	82.5	75	77	68	75.5	62.5	67	61	67	52	55	41.5	51	37
俄文	94.5	93.5	59	49	57.5	48.5	50	38	44	35	35.5	30.5	40	31	39	20.5	40.5	18	43.5	16.5
日文	96	92.5	94	85	91	82	87.5	77.5	79.5	71	75.5	70.5	76	71	78	66.5	77.5	58	73	46
韩文	98.5	95.5	85.5	77	81	72.5	77.5	62	69	55.5	66.5	51	67.5	56	65	48.5	60.5	47	68.5	38
阿拉伯文	95	93	91	80	91.5	72	83	70	77.5	63	75	62.5	66.5	63	57.5	38	59	32.5	57	27
印地文	92	95	86.5	73	83.5	64	76	44	72	27	69.5	17	68	22.5	70.5	15	65.5	12	63	13
缅甸文	93	98	90	85	87.5	80	85	76.5	78.5	71.5	73.5	62	74	63.5	67.5	55	69	48.5	67.5	44
柬埔寨文	95.5	91.5	70	89	57.5	82	52.5	78.5	47	70.5	45	69	44.5	73.5	35	66	37.5	60.5	39	60
藏文	97	95	82	65	73.5	62	63	49.5	61.5	43	57.5	40.5	56.5	45	58	33	55	27	54	24
Rate	95.3	94.2	82.3	74.25	78.4	69.3	72.3	61.5	66.05	54.4	62.15	48.95	60.9	50.95	58.7	41.65	56.75	36.2	55	31.8

表 3 实验方案 2 本文方法识别结果的 number 统计值

Tab.3 Statistics of number in the second experiment result with the proposed method

文种	英文	中文	俄文	日文	韩文	阿拉伯文	印地文	缅甸文	柬埔寨文	藏文
英文	1 184		582	7		64	36	17	48	
中文		1 498	31	44	237		133		47	96
俄文	341	182	1 007				306			60
日文	1	131	56	1 656	145			82	72	
韩文			42		1 479	37		63		
阿拉伯文	1		90			1 506			56	74
印地文	397	187	184	123	56	283	1 493		449	382
缅甸文	46			10			11	1 571		72
柬埔寨文						36			1 047	
藏文	30	2	8	160	83	74	21	267	281	1 316

由表 2 可以看出,当图像库中的图像没有旋转时,两种算法的识别率均比较高,而当文本图像文字行出现不同角度范围内的倾斜时,文献[11]中所用方法识别率迅速下降,本文算法则表现出了较强的鲁棒性,无论是识别率还是稳定性均为最优。由于英文和俄文纹理特征相近,文字行倾斜时两种算法识别率均降低,但本文算法相对较好。对柬埔寨文,由于文字特征主要体现在结构上,本文算法识别效果较差。表 3 反映了错分样本的分布情况,错分样本主要集中在英文、俄文和印地文。从算法的复杂性角度分析,由于可控金字塔变换存在冗余性,本文算法的特征提取速度比文献[11]中所用方法慢,但所提特征维数仅占其特征维数的 24/384,用 Libsvm

进行分类时,大大降低了参数优化的复杂度,分类效率高。

5 结 论

提出了一种基于可控金字塔变换的文本图像文种识别方法,并通过特征空间的循环移位操作,实现了对文字行倾斜的鲁棒性。将 Libsvm 用于文种分类,大大提高了文种识别率。

参考文献 (References)

[1] Nakayama T, Spitz A L. European language determination from

- image [C]//Proceedings of the International Conference on Document Analysis and Recognition. Tsukuba, Japan: University of Tsukuba, 1993: 159-162.
- [2] Spitz A L. Script and language determination from document images [C]//Proceedings of Third Annual Symplic Document Analysis Information Retrieval. Las Vegas, America: University of Las Vegas, 1994: 229-235.
- [3] Elgammal A M, Ismail M A. Techniques for language identification for hybrid Arabic-English document images [C]//Proceedings of Sixth International Conference on Document Analysis and Recognition. Seattle, Washington DC, America: University of Seattle, 2001: 1100-1104.
- [4] Ding J, Lam L, Suen C Y. Classification of oriental and European scripts by using characteristic features [C]//Proceedings of ICDAR [C]. Ulm, Germany: IEEE Computer Society, 1997: 1023-1027.
- [5] Pal U, Chaudhuri B B. Identification of different script lines from multi-script documents [J]. Image and Vision computing, 2002, 20(13-14): 945-954.
- [6] Spitz A L. Determination of the script and language content of document images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(3): 235-245.
- [7] Hochberg J, Kelly P, Thomas T. Automatic script identification from images using cluster-based templates [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(2): 176-181.
- [8] Busch A, Bolse W W, Sridharan S. Texture for script identification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(11): 1720-1732.
- [9] Tan T. Rotation invariant texture features and their use in automatic script identification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(7): 751-756.
- [10] Zeng Li, Tang Yuanyan, Chen Tinghui. Multi-scale wavelet texture-based script identification method [J]. Chinese J. Computers, 2000, 23(7): 699-704. [曾理,唐远炎,陈廷槐. 基于多尺度小波纹理分析的文字种类自动识别 [J]. 计算机学报, 2000, 23(7): 699-704.]
- [11] Hiremath P S, Shivashankar S. Wavelet based co-occurrence histogram features for texture classification with an application to script identification in a document image [J]. Pattern Recognition Letters, 2008, 29: 1182-1189.
- [12] Simoncelli E P, Freeman W T, Adelson E H, et al. Shiftable multiscale transforms [J]. IEEE Transactions on Information Theory, 1992, 38(2): 587-607.
- [13] Freeman W T, Adelson E H. The design and use of steerable filter [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(9): 891-906.
- [14] Vapnik V. The Nature of Statistical Learning Theory [M]. New York: Springer-Verlag, 1995.
- [15] Chang Chihchung, Lin Chihjen. Libsvm: a library for support vector machines [EB/OL]. [2008-09-03]. <http://www.csie.ntu.edu.Tw/~cjlin/libsvm>. 2001/2006