高技术通讯 2013 年 第23 卷 第4 期:379 ~ 386

doi:10.3772/j.issn.1002-0470.2013.04.008

基于几何模糊的复杂场景图像关键字识别①

李 鹏② 崔 刚

(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

摘 要 针对复杂场景中受各种形变和噪声干扰的中文文字图像关键字,提出了一种基于几何模糊的鲁棒识别方法。该方法首先根据图像边缘采样特征点的几何模糊描述符得到样本与目标图像间的初始位置对应关系,然后基于相同文字对应的形变图像间的匹配特征点位置容易出现漂移,且通常并不存在严格意义的匹配位置的原因,通过对匹配点的邻域布局相似性进行判断滤除潜在的误匹配,进而通过分析样本图像中的未匹配点的布局范围以区分存在相同偏旁的中文文字,从而进一步提高匹配准确性。而且,基于上述方法,给出了一种直接发现图像中是否存在某特定关键字的方法。实验结果表明,上述方法对受形变等噪声干扰的图像关键字的识别具有较高的准确性。

关键词 几何模糊,局部特征,形变文字,噪声干扰,图像关键字识别

0 引 言

随着图像型垃圾邮件的泛滥和数字化图书馆中文档图像应用规模的不断扩大,图像关键字识别已成为文档图像信息检索和垃圾邮件图像过滤领域的关键技术之一[1]。本文研究了基于几何模糊描述符识别复杂干扰场景下的形变文字图像的技术,在此基础上,提出了一种不需要进行文本定位过程就能直接发现图像中是否存在特定关键字的方法。实验表明,该方法对受各种噪声干扰的图像和关键字,具有较高的识别准确性。

1 相关工作

目前的图像关键字识别方法主要分为两类。第一类方法利用光学字符识别(optical character recognition, OCR)工具将图像中的文本内容转换为ASCII 码形式,然后在此基础上进行检索和过滤。虽然 OCR 已在图像文本内容提取方面取得了显著成效,但它的缺点是计算复杂度高,不适用于对实时性要求较高的应用。此外,OCR 通常只对于排版整齐和受噪声干扰少的图像识别效果好,并且需要提前预知文本的语言类型。第二类方法则直接利用图像关键字的边缘形状和连通域等视觉特征进行匹配。利用这类方法,用户可以首先指定关键字,然后借助对应的图像关键字样本与目标图像的匹配结果进行识别。这类方法避免了 OCR 中复杂度较高的版面分析、文本区域定位、字符分割和识别等过程,具有更高的灵活性和更广泛的应用领域。

目前,第二类图像关键字识别方法也分为两类。第一类需要对图像进行分割,借助图像中的连通域特征进行识别。如 Bai 等[2]提出了利用英文字符的上下基线等特征生成单词的形状编码,然后用于英文文档图像内容识别。Roy 等[3]提出一种利用连通域构成的字符基元进行关键字识别的方法。Lu 等[4]提出了一种根据汉字笔画邻域特征发现文档图像中是否包含特定关键字的方法。第二类则不需要进行图像分割,借助稳定区域特征和滑动窗口等方法进行识别。如 Gatos 等[5]提出对图像进行分块,然后提取每块的5 个特征向量描述符进行匹配。Rusinol 等[6]提出利用滑动窗口的方法将图像分为不同的模板,然后提取尺度不变特征变换(scale invariant feature transform, SIFT)描述符用于匹配。Aouadi 等[7]提出了一种利用通用霍夫变换进行阿拉伯手写历史文档图像关键字识别的方法。Burl 等[8]根据手写文字笔画的方向提取关键点,然后利

① 国家自然科学基金(61171193)资助项目。

② 男,1984 年生,博士生;研究方向:模式识别,图像处理;联系人,E-mail: lipenggongda@gmail.com

(收稿日期:2012-05-28)

用 Dryden-Mardia 概率模型对关键点的空间分布特征进行相似性判断。

上述方法在图像背景“干净”的前提下效果较好,但对于受噪声干扰的复杂场景中的文字识别性能有限。然而,一些特殊应用中的文字图像经常受到各种噪声或形变等干扰。如为了对抗过滤系统,垃圾邮件图像中的文字通常经过了形变,并且添加各种噪声干扰[9]。历史文档图像中也经常包含许多噪声[10]。因此,出现了一些其它用于复杂场景中的图像关键字识别方法。如 Leydier 等[10]提出一种利用关键字图像的兴趣区域及其位置关系进行中世纪手稿图像检索的方法。随后,作者又对兴趣区域的提取方法进行了改进[11]。这种方法不需要对图像进行布局分析和字符分割等预处理,对于噪声干扰下的文字识别具有较好的鲁棒性,但其方法不适用于中文文字图像识别。Llados 等[12]提出利用形状上下文进行文档图像关键字识别。此外,Thayananthan 等[13]提出了一种广义形状上下文,并结合传统形状上下文进行验证码图像文字识别,但这种方法需要借助如线性规划等方法进行最优匹配估计,计算复杂度较高。此外,形状上下文需要进行字符分割,且易受噪声影响,不适用于噪声干扰下的文字图像识别。Mori 等[14]提出利用 Chamfer 匹配识别混杂场景下的验证码图像文字。这种方法不需要进行字符分割,对噪声干扰具有很高的鲁棒性,但是需要大量的匹配模板以应对形状的尺度、旋转等变换。Campos 等[15]提出利用分类器识别自然场景中的英文字符。随后,在此基础上,Wang 等[16]提出一种识别被全自动区分计算机和人的公开图灵测试(Completely Automated Public Turing Test to Tell Computers and Humans Apart, CAPTCHA)干扰的英文单词的方法。大小写英文字符共 52 个,因此可以通过提取每个字符的特征,然后利用多类别分类器进行识别,最后在此基础上进行英文单词的识别。但汉字由独立个体构成,并且常用汉字有 4000 多个,噪声干扰下的字符之间容易存在特征交叉,因此不适合利用分类器进行识别。

不同语言的图像关键字通常具有不同的轮廓特征,中文字符的轮廓特征则更加复杂,并且目标图像中的相同文字可能对应不同的字体、笔画宽度和噪声干扰等。不同噪声干扰下的图像关键字通常由多个不同的连通域构成,并且部分干扰下的文字难以分割。这些都使得复杂干扰场景下的图像关键字识别仍然面临诸多挑战。

2 本文工作概述

本文工作分为文字图像识别和图像关键字发现两部分。

文字图像识别:给定文字图像集,找出与样本图像匹配的文字图像,即对应于相同的文字。理想情况下,期望通过样本与目标图像对应匹配的局部特征进行识别。但部分文字图像仅可以提取较少的 SIFT 等局部特征点,并且这些点通常并不位于文字的边缘上。同时受形变和其它噪声干扰,导致目标与样本图像的匹配特征点数量有限。因此,需要一种更有效的局部特征描述方法。

形变文字的识别与目标识别具有相似性。本文借鉴了其中的几何模糊描述符进行图像关键字检测,它对于低形变下的目标识别具有较好的效果。通过采样图像边缘能量特征点,并用几何模糊进行描述,从而得到对应的匹配点,这是本文进行文字图像识别的基础。

图像关键字发现:给定目标图像,判断其中是否存在特定的关键字。图像关键字发现更具有实际意义。虽然可以利用本文方法对定位后的文本区域进行识别,但是复杂场景中的文本区域定位仍然具有挑战性。因此,我们在上述工作基础上提出一种图像关键字发现方法。根据匹配的局部特征确定潜在的文字区域,然后截取后进行详细验证,从而避免文本定位过程。

3 中文文字图像识别

3.1 几何模糊描述符

图 1 给出了汉字“票”对应的多种不同字体,以及受不同类型噪声干扰的样本图像。图像关键字识别系统通常难以获取同一关键字的所有不同图像样本,但需要在有限样本的情况下尽可能地识别最多的相关结果,所以其中的关键步骤在于稳健特征的提取。

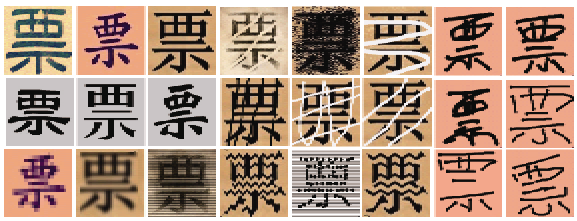

图 1 文字“票”对应的不同文字图像样例

通常,目标在发生形变的情况下,我们很难进行精确匹配。Berg 等[17,18] 提出了一种低形变状态下的形状匹配方法,即先对原信号进行几何模糊,通过对比不同物体形状几何模糊后的差别来判定两个目标是否相似。图像几何模糊特征提取的具体步骤如下:首先,通过对图像进行滤波得到 4 个方向(0°,45°,90°,135°)的边缘能量稀疏信号;然后,根据空间位置变化利用可变核对图像进行高斯模糊,离基准点越远,图像信号模糊的越厉害;最后,以基准点为中心,对几何模糊后得到的形状进行离散点采样,生成几何模糊描述符。

本文使用的模糊函数为

$$B_{x_0}(x) = S_{\alpha|x|+\beta}(x_0 - x) \tag{1}$$

其中,常量 α 和 β 用于控制几何模糊的程度,x_0 是基准点的位置。对于每个基准点,在 4 个通道中根据不同的距离均采样 1,8,8,10,12 和 12 个点,然后拼接 4 个通道的采样点值从而得到最终的几何模糊描述符。我们采样一定数量的边缘点作为基准点(本文中也将其称为特征点),然后提取几何模糊描述符用于匹配。特征点采样选择统一均匀采样。图 2 给出了一对匹配特征点的几何模糊描述符样例。

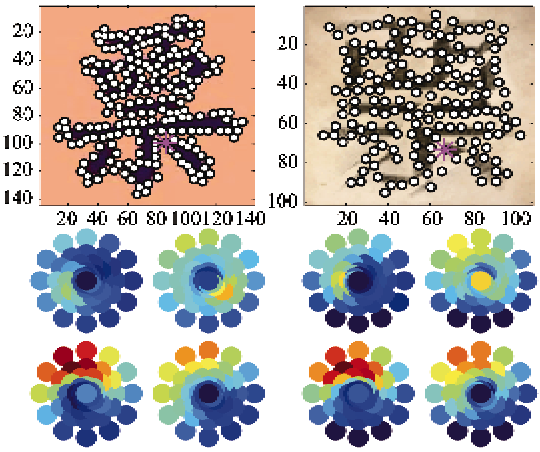

图 2　匹配特征点样例及其对应的几何模糊描述符

假设 $P = \{p_i \mid 1 \leqslant i \leqslant n\}$ 和 $Q = \{q_j \mid 1 \leqslant j \leqslant m\}$ 分别为样本与目标图像的特征点描述符集合。对于 $\forall p_i \in P$,如果其与目标图像中的最近邻点 $q_{i'}$ 的相似性小于阈值 τ,则 $q_{i'}$ 为 p_i 的匹配点,即

$$q_{i'} = \arg_{q_j} \min_{q_j \in Q} d(p_i, q_j), \quad d(p_i, q_j) \leqslant \tau \tag{2}$$

否则,p_i 在目标图像中不存在匹配点。其中 $d(\cdot,\cdot)$ 为相似性度量函数。据此,我们可以将 P 分为匹配与未匹配两个点集,即

$$P = \{p_i\} = \{p_j\}_M \cup \{p_k\}_{UM} \tag{3}$$

这两个集合将用于后续的匹配关系验证。图 3 给出了部分利用几何模糊描述符得到的目标与样本图像特征点匹配样例。可见即使当图像在受到较大噪声干扰时,仍能够得到大量准确的匹配特征点。

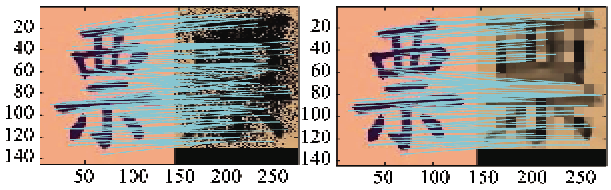

图 3　文字图像特征点匹配样例

3.2　非严格性误匹配特征点检测

局部特征点匹配中的一个重要问题是容易存在误匹配。准确的特征点对应关系是正确识别的基础。利用几何模糊描述符虽然能够得到较多的匹配点,但其中仍可能存在误匹配。对于近似复制图像检测而言,由于图像之间存在完全相同的部分,因此匹配特征点存在严格意义上的正确与误匹配[19] 之分。但由图 1 可知,文字图像之间可能存在不同的形变和干扰,所以需要区别判断。这里我们首先讨论匹配特征点漂移问题,然后给出误匹配滤除方法。

受文字图像形变、噪声干扰以及特征点采样等因素的影响,样本文字图像与目标图像的特征点之间通常并不存在严格意义的对应关系。例如手写文字就可能引起匹配特征点的漂移。如图 4 所示,对于样本图像中的点 p_i,从人的视觉感知角度而言,其与 $q_{i'}$ 匹配更合适,但真正的匹配结果确为 $q_{i''}$,即匹配点位置存在漂移现象。所以,如果 $q_{i'}$ 在目标图像中位于某个限定区域内,即认为 p_i 与 $q_{i'}$ 为正确匹配,否则视为误匹配。为了显示的简洁性,这里仅给出部分匹配点。

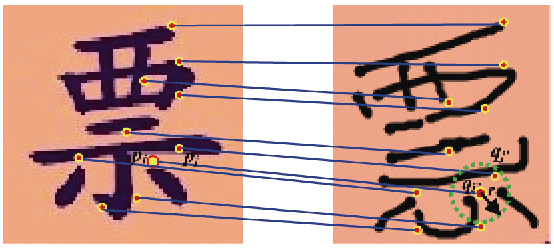

图 4　匹配特征点的位置漂移问题

为了提高匹配的准确性,我们结合特征点的位置漂移现象,从以下两个方面考虑滤除潜在的误匹配:

(1) 每对匹配特征点在两副图像中以 r 为半径的圆形邻域内也应存在其它匹配特征点,否则可能

为误匹配。由于本文的特征点通过对边缘等间隔采样得到，所以匹配特征点往往不是独立存在的。由于可能存在匹配点的位置漂移，因此这里不对邻域内匹配点之间的相对位置关系做严格判断。

(2) 每对匹配特征点在两幅图像中以 r 为半径的圆形邻域内的所有对应匹配点之间的连线应该近似平行。这些连线应该存在一个主方向 α，这对匹配点的连线方向与主方向之间的角度差应小于限定阈值，否则可能为误匹配。

综上，对于匹配点对 $(p_i,q_{i'})$，我们利用其 r 邻域范围内其它匹配特征点的位置关系进行误匹配验证。假设 p_i 与 $q_{i'}$ 的连线 $\overrightarrow{p_iq_{i'}}$ 的角度为 $\alpha_{\overrightarrow{p_iq_{i'}}}$，同理，可以得到分别位于 p_i 与 $q_{i'}$ 的 r 邻域内的所有匹配点的连线边的方向及其主方向角 α，以及匹配点的总数 $n_{p_iq_{i'}}$，则 p_i 与 $q_{i'}$ 的误匹配检测判断结果如下：

$$\text{判断结果}(p_i,q_{i'}) = \begin{cases} \text{正确匹配}, & |\alpha_{\overrightarrow{p_iq_{i'}}} - \alpha| \leqslant \varepsilon \,\&\, n_{p_iq_{i'}} \geqslant \eta \\ \text{误匹配}, & \text{其他} \end{cases} \tag{4}$$

其中：ε 和 η 为阈值，本文中分别取 10° 和 8。特征点邻域半径 r 取样本图像宽度的 1/6。实际中，r 和 η 还可根据样本图像大小与特征点采样间隔进行调整。

3.3 共享偏旁的结构相近文字图像区分

偏旁是由笔画组成的具有组配汉字功能的构字单位，是汉字合体字的结构单位。从前称合体字的左方为“偏”，右方为“旁”；现在把合体字的组成部分统称为偏旁。如汉字“票 要 覆 粟 贾”中都包含偏旁“西”。包含相同偏旁的文字图像之间可能存在大量的匹配点，但确实不是相同的文字，即不能完全根据匹配的点对数量来判断是否为正确匹配的文字图像。

样本与目标图像的匹配特征点集中在相同的偏旁区域，而不同的区域则通常不存在匹配点。因此，可以利用样本图像中未匹配特征点的布局来区分具有相同偏旁的文字。对于样本图像中的未匹配特征点集合 $\{p_k\}_{UM}$，利用邻接性将其分为多个不同的区域。对于 $\forall p_k,p_m \in \{p_k\}_{UM}$，如果 $e(p_k,p_m) \leqslant \sigma$，则 p_k 与 p_m 邻接，记作 $p_k \propto p_m$，其中 $e(p_k,p_m)$ 为 p_k 与 p_m 的距离。邻接关系具有传递性，如果 $p_k \propto p_m$ 和 $p_m \propto p_h$，则 $p_k \propto p_h$。据此，我们可以得到所有未匹配特征点之间的邻接关系。最后，将所有彼此邻接的点归为一个区域。如图 5 给出了一个样本图像中未匹配点区域的样例，其中每个区域都利用红色的方框给以标注。

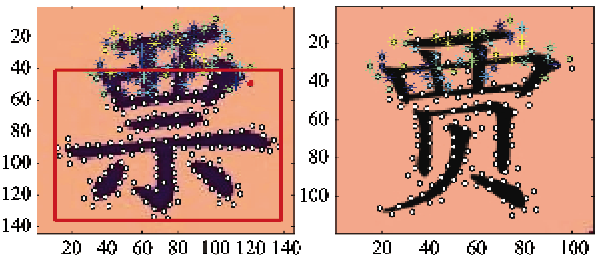

图 5 样本图像中未匹配特征点分区的样例

最后，根据未匹配特征点集合 $\{p_k\}_{UM}$ 的规模和对应的区域面积总和与样本图像的面积比例确定样本图像与目标图像中是否存在不相同的偏旁成分。P 和 Q 的最终相似性得分为

$$C(P,Q) = \lambda \cdot \frac{|\{p_k\}_{UM}|}{|P|} + (1-\lambda) \cdot \frac{\sum A_i}{A_P} \tag{5}$$

其中：A_i 为样本图像 P 中未匹配点的第 i 个区域的面积，A_P 为 P 的总面积，λ 为权重。当 $C(P,Q)$ 大于限定阈值时，P 与 Q 不匹配；否则匹配。

需要注意的是：如果偏旁是一个文字，但它又在某个文字中出现，如“票”和“漂”。此时，如果存在匹配，则该关键字确实出现了。OCR 工具也可能将部分合体字识别为不同偏旁对应的汉字。由于较少以一个偏旁作为关键字进行检索，因此我们不考虑该问题。

4 图像关键字发现

实际应用中更关心的是某幅图像中是否包含特定的敏感关键字。虽然可以利用第 3 节的方法对定位后的文本区域进行关键字匹配，但杂乱场景下的文本定位仍然非常具有挑战性。如果直接提取目标图像的几何模糊描述符与样本图像进行匹配，则由于几何模糊描述符采样点范围较大，目标图像中文本边缘区域特征点的描述符易受干扰，从而影响匹配的准确性。鉴于上述问题，本文使用如下方法进行图像关键字发现。

具体算法流程如图 6 所示：首先提取关键字图像样本与目标图像的几何模糊描述符，获取匹配特征点；然后根据这些特征点的位置和样本大小确定潜在的匹配文字区域，截取后，借助第 3 节中的方法进行详细判断。具体的潜在文字区域定位方法如下：假设样本图像大小为 (W,H)，点 $p_i(x_i,y_i)$ 与 $q_{i'}(x_{i'},y_{i'})$ 匹配，则目标图像中以点 $(x_{i'} - x_i,\ y_{i'}$

$-y_i$)与($x_{i'}-x_i+W$, $y_{i'}-y_i+H$)为对角线点,W、H为高和宽所构成的矩形区域就是潜在的关键字区域。这样就可以绕过复杂背景图像中的文本区域定位过程。在这个过程中,我们在目标图像中提取较多的采样点用于匹配并定位潜在的关键字区域。如果目标图像尺度较大,则可以采用滑动窗口的方法,对每个窗口区域借助上述方法进行判断,从而有助于提高检测的准确性。

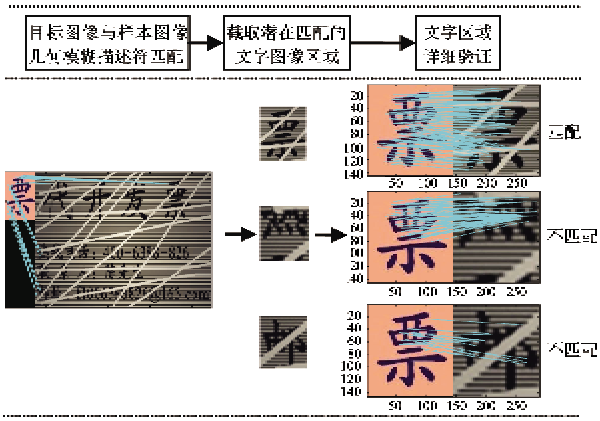

图6　图像关键字发现算法流程

由于几何模糊描述符需要利用可变核对图像进行平滑,所以受文字结构影响,当分辨率较小时,平滑后的图像视觉特征可区分性降低。因此,我们要求文字图像的分辨率最小为90×90。并且,由于几何模糊描述符对尺度和旋转的鲁棒性有限。这里,可识别的文字与样本的尺度变化范围为90%到110%,旋转角度差小于5°。

5　实验验证与分析

实验分为两部分:(1)测试本文方法识别文字图像的性能;(2)测试本文方法发现图像中特定关键字的性能。下面首先介绍本文的测试图像库。

第一类测试图像为受形变和噪声干扰的文字图像。我们选择50个汉字作为测试的基准,每个汉字对应20种不同的干扰情况,干扰类型与图1中"票"对应的文字图像相似。此外,随机选择其它500幅文字图像和部分自然图像截图用于共同构成最终的测试图像库,总规模为1600幅。第二类是包含65个文字并且受不同噪声干扰的10幅图像,每幅图像主要对应一种干扰类型。这些图像用于关键字发现性能测试。

(1) 文字图像检索

为了测试本文方法对于各种干扰的鲁棒性,对于50个关键字,我们都利用一个样本图像进行检索。本文方法对于所有关键字检索结果的平均准确率约为89%,平均召回率约为72%。此外,表1给出了利用国内著名的汉王OCR(HW-OCR)[20]和Google的Tesseract-OCR[21]对所有50组文字图像进行识别的准确率。可见,Tesseract-OCR的识别准确率非常低,而HW-OCR的准确性则较高。但通过分析识别结果可知,HW-OCR对于受噪声干扰的文字图像正确识别率很低,而背景干净的文字则正确识别率较高。

表1　OCR工具的识别性能

OCR工具	准确率
Tesseract-OCR	9%
HW-OCR	56%

图7给出了14个随机选择的文字样本图像检索结果的准确率-召回率(P-R)曲线。P-R关系反映了调整算法中不同的参数(如几何描述符匹配阈值和样本图像未匹配的点所占的面积比例等)后的准确率和召回率的变化关系。

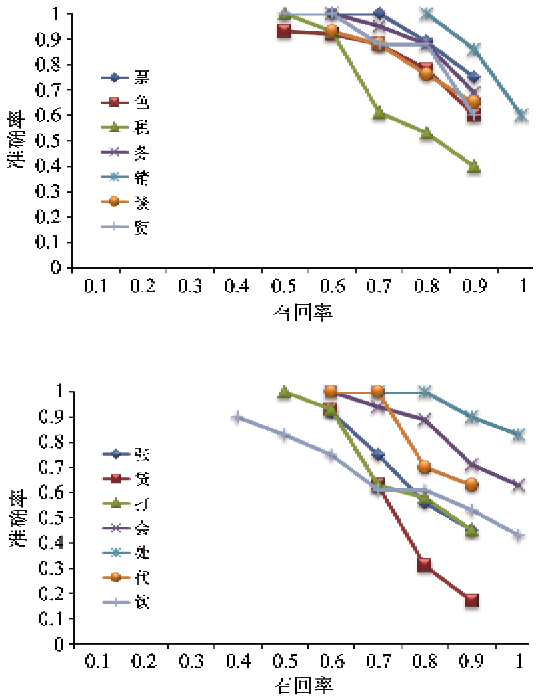

图7　14个文字的检索结果P-R曲线

图8给出了利用汉字"贸"的样本图像得到的检索结果,其中正确、错误和漏检的图像分别利用绿色、红色和紫色点框标出。

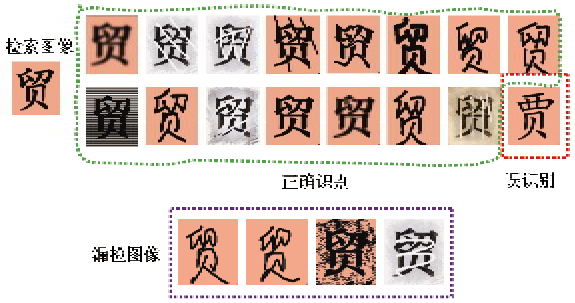

图 8 “贸”检索结果样例

由检索结果可知,本文方法对于形变和各种噪声干扰下的文字检索具有较高的准确性。而 OCR 工具对于背景噪声干扰较大的文字识别的准确率较低。这是由于 OCR 主要用于对背景干净、受噪声少的图像进行内容识别,所以对于复杂干扰下的文本内容识别效果不佳。但是本文方法也存在部分误识别和漏检的情况。由图 7 可知,当召回率高于 70% 后,部分关键字的检索准确率会有较明显下降,如“税”和“贷”。这是由于部分测试图像受噪声干扰较大,或者存在部分边缘缺失,导致提取的边缘采样点与样本图像存在较大差异。对于一些结构相近的图像也容易存在误匹配,如图 8 中的“贸”和“贾”。当求几何模糊描述符时,边缘信号模糊后的图像特征可区分性下降,导致这些文字图像间存在匹配的特征点,却并不是正确匹配的文字。但是可见这两个文字对应的图像视觉特征确实具有相似性。OCR 通常也难以区分这些文字图像。由漏检的 4 幅图像可知,前两幅手写文字与样本图像笔画宽度存在差异,并且笔画粘连严重,而后两副图像存在的噪声干扰较大,边缘采样点和对应的描述符存在差异,从而导致了漏检。由图 7 中关键字“贸”的检索结果 P-R 曲线可知:通过放松匹配限制可以提高召回率,但是匹配准确性会有所下降。

(2) 图像中特定关键字的发现

当图像受噪声干扰时,是否能够发现图像中包含特定关键字至关重要。这里,我们仍然选择关键字“票”进行测试,并且利用一个样本图像用于检索。每张目标图像中都包含 6 个该关键字。“票”发现结果的平均准确率和召回率分别为 96.9% 和 91.5%。图 9 给出了部分目标图像中的发现定位结果,其中红色和绿色框标注的都为潜在的关键字区域。经过最后的验证,正确的区域由绿色框标出,而红色框标注的则为被排除的区域。

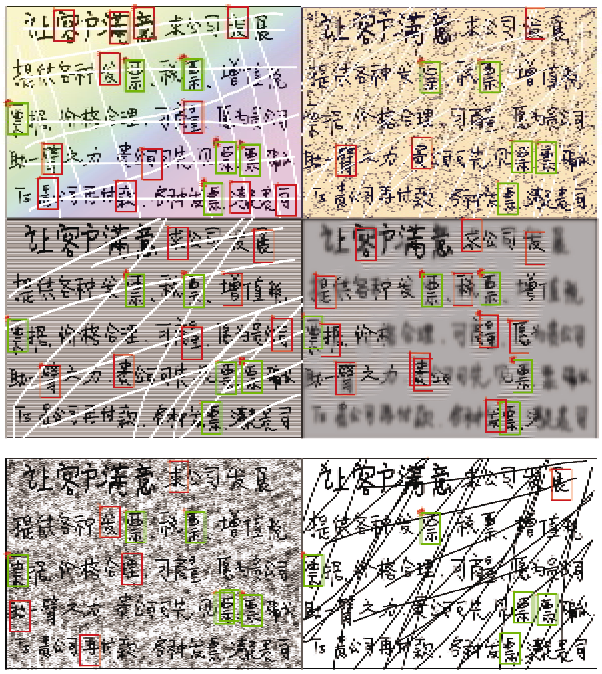

图 9 文字“票”在部分目标图像中的发现结果

从结果中可以发现,虽然存在部分误识别的情况,但大多数相关的关键字都被正确发现。我们可以通过调整几何模糊描述符的匹配参数来提高检测的准确性。此外,两个 OCR 工具在所有测试图像中都仅识别出 1 个正确的关键字“票”,可见噪声干扰会严重影响 OCR 的识别性能。虽然 HW-OCR 在关键字图像识别中取得了较高的准确性,但是其仍然难以应用于实际图像中受噪声和形变干扰的特定关键字发现。

综上可知,本文方法对于复杂干扰下的关键字识别具有较高的准确性。由于中文文字的结构较复杂,并且存在具有结构相似性的文字,因此结果中存在部分误识别。此时,对于一个检索关键字,可以借助多个不同的样本图像进行识别,以提高图像关键字发现的准确性。

我们在 Matlab R2010b 中实现了本文方法。机器配置为 Intel ® Core(TM)2 Duo CPU 2.0GHz, 2GB 内存,Windows Vista 操作系统。对于图 9 所示的分辨率为 1487 × 881 的测试图像,关键字“票”的检测时间约为 480ms。而 HW-OCR 和 Tesseract-OCR 对于每幅图像的识别时间则约分别为 0.3s 和 8s。实际环境下,本文方法的计算复杂度随测试图像大小、特征点采样数量的不同而变化。由于 Matlab 主要用于算法原型验证,对于循环控制类程序执行效率较低。因此,可以通过利用 C/C++ 来进一步提高执行效率。

6　结　论

本文中,我们提出了一种基于几何模糊描述符的复杂干扰场景下的形变文字图像识别方法,并且分析了形变文字间匹配特征点的漂移现象和区分共享偏旁文字的方法。在此基础上,我们给出了一种直接发现图像中是否存在特定关键字的方法,并且不需要进行文本定位过程。实验结果也表明,本文方法对于受各种噪声干扰下的文字识别和发现具有较高的准确性。对于一些需要过滤图像中是否包含特定敏感关键字的应用领域,本文提供了一种有效的参考方法。但是,几何模糊描述符不具有尺度和旋转的完全不变性,并且对于分辨率较小的文字可区分性能仍然有限,这也是我们下一步需要解决的问题。

参考文献

[1] Murugappan A, Ramachandran B, Dhavachelvan P. A survey of keyword spotting techniques for printed document images. *Artificial Intelligence Review*, 2011, 35(2):119-136

[2] Bai S Y, Li L L, Tan C L. Keyword spotting in document images through word shape coding. In: Proceedings of the International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009. 331-335

[3] Roy P P, Ramel J, Ragot N. Word retrieval in historical document using character-primitives. In: Proceedings of the International Conference on Document Analysis and Recognition, Beijing, China, 2011. 678-682

[4] Lu Y, Tan C L. Word spotting in Chinese document images without layout analysis. In: Proceedings of the International Conference on Pattern Recognition, Quebec, Canada, 2002. 57-60

[5] Gatos B, Pratikakis I. Segmentation-free word spotting in historical printed documents. In: Proceedings of the International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009. 271-275

[6] Rusinol M, Aldavert D, Toledo R, et al. Browsing heterogeneous document collections by a segmentation- free word spotting method. In: Proceedings of the International Conference on Document Analysis and Recognition, Beijing, China, 2011. 63-67

[7] Aouadi N, Kacem A. Word spotting for Arabic handwritten historical document retrieval using generalized Hough transform. In: Proceedings of the Third International Conference on Pervasive Patterns and Applications, Rome, Italy, 2011. 67-71

[8] Burl M C, Perona P. Using hierarchical shape models to spot keywords in cursive handwriting data. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Santa Barbara, USA, 1998. 535-540

[9] Biggio B, Fumera G, Pillai I, et al. Image spam filtering by content obscuring detection. In: Proceedings of the 4th Conference on Email and Anti-Spam, Mountain View, USA, 2007. 2-3

[10] Leydier Y, Lebourgeois F, Emptoz H. Text search for medieval manuscript images. *Pattern Recognition*, 2007, 40(12):3552-3567

[11] Leydier Y, Ouji A, Lebourgeois F, et al. Towards an omnilingual word retrieval system for ancient manuscripts. *Pattern Recognition*, 2009, 42(9):2089-2105

[12] Llados J, Roy P P, Rodriguez J A, et al. Word spotting in archive documents using shape contexts. In: Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis, Girona, Spain, 2007. 290-297

[13] Thayananthan A, Stenger B, Torr P H S, et al. Shape context and chamfer matching in cluttered scenes. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, USA, 2003. 127-133

[14] Mori G, Malik J. Recognizing objects in adversarial clutter: breaking a visual CAPTCHA. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, USA, 2003. 134-144

[15] Campos T E D, Babu B R. Character recognition in natural images. In: Proceedings of the 4th International Conference on Computer Vision Theory and Applications, Lisboa, Portugal, 2009. 273-280

[16] Wang K, Belongie S. Word spotting in the wild. In: Proceedings of the 11th European Conference on Computer Vision, Crete, Greece, 2010. 591-604

[17] Berg A C, Malik J. Geometric blur for template matching. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, USA, 2001. 607-617

[18] Berg A C, Berg T L, Malik J. Shape matching and object recognition using low distortion correspondences. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005. 26-33

[19] Li P, Yan H B, Cui G, et al. Image local invariant features matching using global information. In: Proceedings of the International Conference on Information Science

and Technology, Wuhan, China, 2012. 627-633

[20] 汉王 OCR 6.0. http://www.hanwang.com.cn: Hanwang, 2012

[21] Tesseract OCR. http://code.google.com/p/tesseract-ocr/: Google, 2012

Image keyword spotting in cluttered scenes based on geometric blur

Li Peng, Cui Gang

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)

Abstract

Image keyword spotting was thoroughly studied because it is a key technique for document image retrieval, spam image filtering, etc, and a method for spotting of deformable Chinese image keywords in cluttered scenes based on geometric blur was proposed. The method first obtains the sampled feature points and the corresponding geometric blur descriptors from the multi-channels of oriented edge energy images, after that the initial correspondences between local features can be obtained. Then, it filters out the potential mismatches using the neighborhood information for improving the matching accuracy based on the consideration that there are no rigid correspondences between the local features of deformable keyword images because the matched feature points often float to their neighbors. Next, it uses the ratio of the area of the no-match feature points in the sample image to that of the whole image to distinguish the keyword images sharing the same character part. Furthermore, a method to spot the specific keywords in an image based on the method above was proposed. The experimental results show that the proposed spotting method can recognize and spot the keyword images with the higher accuracy.

Key words: geometric blur, local feature, deformable character, noise interruption, image keywords spotting