

几何信息与 SIFT 特征相结合的 特定人手写关键词检测

张文超¹, 吕岳¹, 文颖¹, 黄志敏²

(1.华东师范大学 计算机科学与技术系,上海 200241; 2.公安部第三研究所,上海 200032)

摘 要:中文汉字类别繁多,书写随意性大,使得中文的手写体关键词检测具有很大的挑战性。提出一种基于文字几何信息和 SIFT 特征相结合的手写体关键词检测方法,通过计算文本图像特征的匹配度来检测特定书写人的手写关键词。尺度不变特征转换(scale invariance feature transform,SIFT)局部特征具有良好的稳定性和独特性,既能适应同一书写人手写汉字的差异,又能区分不同书写人的书写笔迹。结合文字的几何信息,通过滑动窗口和最大团查找方法可以有效地删除误匹配点,极大地提高关键词检测的成功率。对大量手写体文本图像的实验结果表明,该方法能够有效检测同一书写人的相同关键词,具有较高的召回率和准确率。

关键词:关键词检测;SIFT;滑动窗口;最大团查找

中图分类号: TP18 **文献标志码:** A **文章编号:** 1673-4785(2014)05-0544-07

中文引用格式:张文超,吕岳,文颖,等. 几何信息与 SIFT 特征相结合的特定人手写关键词检测[J]. 智能系统学报, 2014, 9(5): 544-550.

英文引用格式:ZHANG Wenchao,LYU Yue,WEN Ying,et al. Specific handwritten keyword spotting using geometric information and SIFT feature[J]. CAAI Transactions on Intelligent Systems, 2014, 9(5): 544-550.

Specific handwritten keyword spotting using geometric information and SIFT feature

ZHANG Wenchao¹, LYU Yue¹, WEN Ying¹, HUANG Zhimin²

(1. Department of Computer Science and Technology, East China Normal University, Shanghai 200241, China; 2. The Third Research Institute of Ministry of Public Security, Shanghai 200032, China)

Abstract: Large variety of Chinese characters and handwriting styles leads to a big challenge for keyword spotting in Chinese handwritten documents. A new method combining the character geometric information and SIFT feature is proposed for detecting handwritten keywords of specific handwritten. It is proven that SIFT is a stable and distinctive local feature, which can perform well in distinguishing different handwriting styles. Combined with character geometric information and maximum clique matching, the proposed method can effectively remove miss-matching feature points and improve the precision rate of detection. Experimental results in handwriting document images show that the method can efficiently detect keywords of particular writers and remain high recall rate and high precision rate.

Keywords: keyword spotting; SIFT; sliding window; maximum clique matching; geometric information

手写关键词检测是在大量手写文本图像中检测

特定书写人的特定关键词。随着机器视觉和人工智能的发展,手写关键词检测受到了越来越多的关注。当今社会中存在大量的手写文本,如历史文献、笔记等,这些文本往往通过扫描以图像的形式存储,从中

检测人们关注的关键词,有其应用需求。由于手写体文字随意性大,书写形式自由,容易引起文字倾斜、模糊或粘连,所以特定人手写关键词检测技术的发展面临很大的困难。早期 Rath 和 Manmatha^[1]提出的 DTW (dynamic time warping) 方法在英文手写体文本的关键词检测中取得了较好的效果。此后产生了大量的基于统计和基于结构的方法^[2],使得关键词检测技术得到了快速发展。近几年刘成林等^[3-5]在手写体汉字识别上做了大量工作,在汉字识别和语义分析的基础上进行关键词检测,取得了不错的结果。但中文汉字类别多,书写风格多变,基于识别的关键词检测方法很大程度上受制于手写体汉字识别的效果,更重要的是这种依靠识别结果的方法丢失了文本图像中的笔迹信息,无法区分关键词是来自相同书写人还是不同书写人。

关键词检测的另一方法是基于图像匹配,即将文本图像表述为特征,再对特征区域或者特征点进行相似度匹配,避免了手写体汉字的识别问题而将关键词检测转化为图像匹配问题。有许多特征可以描述文本图像,如统计特征、结构特征、局部特征等。尺度不变特征转换 (scale invariance feature transform, SIFT) 是一种旋转、尺度、光照不变的局部特征^[6],具有良好的稳定性和独特性。从文本图像上提取的 SIFT 特征点,保留了文字局部的细节信息。而结合文字的几何信息,又能够有效地避免汉字重复结构对特征点匹配带来的影响,极大提高关键词检测的成功率。以往对关键词检测的研究并没有对书写人加以区分,并且目前对基于匹配方法的研究还比较初步。因此,本文提出一种基于 SIFT 特征匹配的手写关键词检测方法,能够对特定人的中文手写关键词进行检测,同时结合文字几何信息来提高检测的准确率。该方法对预留的特定作者书写的关键词图像和待检测文本图像分别进行 SIFT 特征提取,使用滑动窗口在待检测文本图像进行滑动并将窗口内文本图像特征与关键词图像特征进行匹配,根据几何信息对匹配点建立几何约束图,通过最大团查找方法删除误匹配特征点,最终筛选和合并窗口来定位特定的手写关键词。实验表明,本方法能够有效地检测手写体关键词,并具有较高的召回率和准确率。

1 SIFT 特征提取

SIFT^[6]算法由 Lowe 在 1999 年提出并在 2004 年将其完善,该算法在机器视觉领域得到了迅速的

发展和应用。SIFT 特征在图像旋转、尺度变换、仿射变换和视角变化条件下都有良好的不变性。SIFT 特征包括了 SIFT 检测算子和描述算子,前者能够检测图像局部稳定的关键点,后者对检测到的关键点邻域生成特征描述矢量。该算法主要分 3 步:1) 尺度空间的建立和关键点的检测;2) 关键点主方向计算;3) 关键点特征描述。

Lindeberg 等^[7]已经证明了高斯核是实现尺度变换的唯一变换核,所以将图像的尺度空间表示成函数 $L(x,y,\sigma)$,它由图像 $I(x,y)$ 和一个变尺度的高斯函数 $G(x,y,\sigma)$ 卷积产生,即

$$L(x,y,\sigma) = G(x,y,\sigma) \otimes I(x,y)$$

式中, \otimes 表示在 x 和 y 两个方向进行卷积操作, $G(x,y,\sigma)$ 为

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

SIFT 算法通过对两个相邻高斯尺度空间图像相减,得到 DoG (difference of Gaussians) 图像来近似 LoG (Laplace of Guassian) 方法。通过对高斯差分图像 $D(x,y,\sigma)$ 进行局部极值搜索,在位置空间和尺度空间定位关键点,其中

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma)$$

式中 k 为相邻尺度空间尺度值的倍数。

DoG 方法是通过构建图像金字塔实现的,将图像金字塔分为 O 组,每组 S 层。通过对每组上下相邻两层的高斯尺度空间图像相减得到 DoG 结果。为了寻找尺度空间的关键点,每个内部的 DoG 图像层像素点需要和它同尺度内的 8 邻域点和上下相邻尺度空间的 9×2 个点一共 26 个点比较,确保在尺度空间和二维图像位置空间都检测到极值点。

为了实现图像特征的旋转不变性,需要计算图像梯度来求取特征点的主方向。每个已经检测到特征点对应一个尺度值 σ ,在该尺度下的高斯图像中,以特征点为中心,计算 $3 \times 1.5\sigma$ 为半径区域内图像梯度的幅值和幅角 (梯度方向):

$$m(x,y) = \sqrt{L(x+y,y) - L(x-1,y) + L(x,y+1) - (x,y-1)}$$
$$\theta(x,y) = \arctan\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right)$$

计算某一特征点邻域内高斯图像的梯度之后,使用直方图统计梯度方向和幅值。将 $0^\circ \sim 360^\circ$ 的方向角分为 36 个区间,当梯度方向落入到某一个区间时,将该方向角对应的幅值乘以圆形高斯加权函数再累加,最终形成梯度直方统计图。将其中最大峰值对应的梯度方向角作为特征点的主方向。

SIFT 描述子 d 是对特征点邻域内高斯图像梯度统计结果的一种表示。计算特征描述矢量的步骤如下:

1) 对某一特征点,在其所在的尺度 σ 对应的高斯图像中,以该点为中心将图像旋转 θ 角度。

2) 将其 $3\sigma \times 3\sigma$ 的邻域划分成 4×4 个子区域,并对每一个子区域计算梯度统计直方图。这里将 $0^\circ \sim 360^\circ$ 的方向角划分成 8 个区间,每 45° 一个区间。最终形成 $4 \times 4 \times 8 = 128$ 维的 SIFT 特征矢量。

3) 为了去除光照变化的影响,对 128 维特征矢量进行归一化处理。

通过上述方法,文本图像或者图像区域可以描述为 SIFT 特征点的集合 $S = \{F_i\}$,每个特征点 $F_i = (x, y, \sigma, \theta, d)$ 由 5 个数据组成,分别表示特征点在图像中的横坐标、纵坐标、尺度值、方向角和特征矢量。通过对关键词图像提取 SIFT 特征点集,存储在特征库文件中,便于后期对特征点的快速读取。

2 关键词检测

对关键词图像和文本图像提取 SIFT 特征点后,需要将两者特征点进行匹配,得到相似的特征点对集合。但是由于汉字结构的相似性,使得汉字之间的特征点相似度匹配存在不小误差,关键词图像上的特征点直接在文本图像特征点集合匹配得到的结果错误比较多。Rodriguez 等^[8]通过提取滑动窗口内的特征进行英文的手写体关键词检测,郑琪等^[9]使用滑动窗口的方法对自然场景中的文字进行定位。利用滑动窗口算法能够缩小特征点匹配的范围,进而有效增加正确匹配的特征点数量。

2.1 滑动窗口提取

对文本图像进行滑动窗口提取,首先构造一个高度等同于文本行,宽度为 w 的窗口,并将该窗口在文本图像上从左至右滑动,滑动间隔为 s 。每一次滑动,窗口都会覆盖文本图像一部分特征点,关键词图像特征点依次与各个窗口内的特征点匹配。

用集合 $Wind(p, q)$ 表示窗口,其中 p 和 q 分别表示窗口的起始坐标和终止坐标。设文本图像宽度为 W_d ,窗口数量为 n ,如图 1 所示。

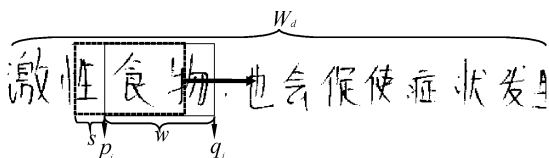


图 1 滑动窗口

Fig.1 Sliding window

按照如下方法来提取滑动窗口。

- 1) 若 $W_d < w$, 则 $n = 1, p_1 = 0, q_1 = W_d$;
- 2) 否则,按照式(1)计算窗口的坐标:

$$n = 1 + \lceil (W_d - w) / s \rceil$$

$$p_i = p_{i-1} + s \quad (1)$$

$$q_i = q_{i-1} + s$$

- 3) 若 $q_i > W_d$, 则 $q_i = W_d$, 即最后一段不足宽度也作为一个窗口。

从文本图像的左侧开始到右侧末端,就像一个从左至右滑动的窗口。每次窗口内的匹配只针对横坐标 x 落在 p_i 和 q_i 之间的文本特征点集合。参数窗口宽度 w 和移动步长 s 影响窗口内的特征点匹配数量。窗口过宽会失去很多正确匹配点,过窄则会产生大量误匹配点。适当加宽窗口可以抑制更多的误匹配点对。而移动步长过大会错过正确的匹配窗口或割开关键词,过小又会使窗口数量增多而降低匹配效率。

手写体文本书写的随意性造成了汉字之间疏密程度的不同,这使得窗口宽度和移动步长很难选取。经过大量实验发现,当窗口宽度与关键词图像宽度的比值等于关键词图像与文本图像特征点尺度均值的比值时,匹配的结果比较好。通过特征点尺度来调节窗口宽度,可以适应图像的尺度变化。本方法中 w 和 s 按照公式(2)计算:

$$w = W_s \times \frac{\text{Avg}(\sigma_d)}{\text{Avg}(\sigma_s)} \quad (2)$$

$$s = \frac{w}{6}$$

式中: W_s 为关键词图像宽度, $\text{Avg}(\sigma_s)$ 为关键词图像特征点尺度的均值, $\text{Avg}(\sigma_d)$ 为文本图像特征点的均值。

在每个窗口中,所有关键词图像特征点可利用 Kd-树算法^[10]查找到文本图像中 5 个最近邻的特征点,形成初始的匹配点对。汉字之间存在大量重复的结构和相似的笔画,而且 SIFT 局部特征缺乏特征点在空间分布上的几何性质,这使得初始的特征点匹配结果存在大量的错误匹配。利用 SIFT 局部特征本身带有的几何信息,如位置、方向、尺度信息,构造几何约束关系,可以进一步筛选错误匹配点对,提高滑动窗口内特征点的匹配成功率。

2.2 几何信息约束和最大团查找

利用特征点的方向、尺度等信息对匹配点对进行初步的筛选,如果不考虑文本图像的旋转、扭曲和缩放,两个相似的特征点之间的方向角、尺度值差异就不会特别大。另外,特征矢量的欧氏距离特别大的匹配点对也会被删除,因为特征矢量表示特征点

周围的梯度统计信息。利用式(3)可以对匹配特征点对进行初步筛选:

$$\begin{aligned} &|\theta_s - \theta_w| < p_1 \\ &p_2 < \frac{\sigma_s}{\sigma_w} < \frac{1}{p_2} \\ &\|d_s - d_w\|^2 < p_3 \end{aligned} \tag{3}$$

式中: θ_s 、 σ_s 、 d_s 和 θ_w 、 σ_w 、 d_w 分别表示关键词图像和窗口图像的特征点方向角、尺度、特征矢量。 p_1 、 p_2 、 p_3 分别是对应的阈值。匹配点经过初步筛选后的结果如图 2 所示。



图 2 匹配特征点初步筛选后结果

Fig.2 Feature matching after initial selection

图中截取了滑动窗口在文本图像中的一个片段,矩形框为窗口的位置,下面显示的是关键词的图像,细直线连接了两幅图像中匹配的特征点。可以看到图中依然存在很多误匹配的点对,这是由汉字重复的结构信息引起的。利用特征点在文字的空间分布建立几何约束,可以进一步删除误匹配点对。几何约束原理就是:关键词图像上存在一定位置关系的两个特征点,它们在窗口图像中对应的两个匹配点也具有相似的位置关系。通过这个原理可以在特征点对之间建立几何约束图,再通过图论中的最大团查找算法获得正确的匹配点对集合。几何约束图(geometric constraint graph,GCG)建立步骤如下

1) 设 $S = \{s_1, s_2, \dots, s_m\}$ 为关键词图像特征点集, $W = \{w_1, w_2, \dots, w_m\}$ 为文本图像中与之匹配的特征点集。无向图 $G = (V, E)$ 表示几何约束图。 $V = \{v_1, v_2, \dots, v_m\}$ 为 G 的顶点集合,其中 $v_i = (s_i, w_i)$, 表示 S 和 W 中每一对匹配点就对应 G 中的一个顶点。 $E \subseteq V \times V$ 为 G 边的集合。

2) 按照如下方法添加 G 的边,对于任意两个特征点对 $v_i = (s_i, w_i)$ 和 $v_j = (s_j, w_j)$, 满足约束条件

$$\begin{aligned} &|(x_{s_i} - x_{s_j}) - (x_{w_i} - x_{w_j})| < p_4 \times \text{Avg}(\sigma_s) \\ &|(y_{s_i} - y_{s_j}) - (y_{w_i} - y_{w_j})| < p_5 \times \text{Avg}(\sigma_s) \end{aligned} \tag{4}$$

则在顶点 v_i 和 v_j 之间添加一条边。式中 x_s 和 x_w 表示特征点在关键词图像和窗口图像内的横坐标,而 y_s 、 y_w 则表示其纵坐标,参考图 3 示例。 p_4 和 p_5 是两个阈值,为了适应不同关键词图像的尺寸,式(4)中的右值应当与其平均尺度值成正比。该式约束条

件的几何意义是,分布于汉字上的特征点随笔画的随意性移动的范围不会超过一定的阈值。因为相同的汉字具有相同的笔画结构和笔顺,相应的特征点也应当具有类似的分布。

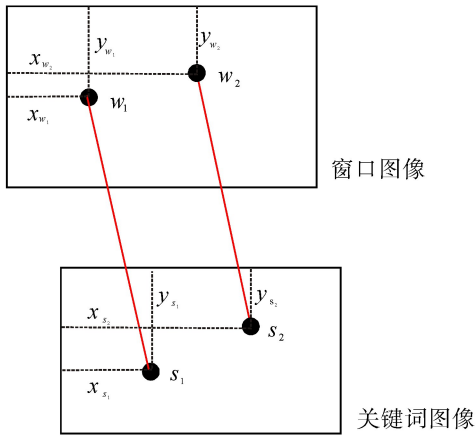


图 3 窗口内匹配点的分布信息

Fig.3 Location information of matching points

当存在越多的正确匹配点时,GCG 图中互相连接的顶点将会越多。为了得到一个存在最多正确匹配点对的子集,就要在 GCG 中找到一个最大子图,在这个子图中所有的顶点都是互相连接的,可使用最大团查找(maximal clique matching,MCM)算法来实现,其具体过程如下

- 1) 初始化候选顶点集 $C = V$, 最大团顶点集为 $M = \varphi$;
- 2) 计算 V 中每个顶点的度,记为集合 $\text{Deg}(V)$;
- 3) 选择 C 中 $\text{Deg}(V)$ 最大的顶点 v_0 , 将其加入 M , 即 $M = M \cup \{v_0\}$;
- 4) 将 v_0 从 C 中删除,并且只在 C 中保留和 v_0 相连接的顶点,即 $C = C \setminus v_0$, $C = C \cup N(v_0)$ 。其中 $N(v_0)$ 表示 v_0 的邻接顶点集。
- 5) 如果 $C = \varphi$, 算法结束,否则转到重复上述过程。

通过建立几何约束图和最大团查找算法,可以删除大量误匹配点对,图 2 中的初始匹配经过几何约束处理后结果如图 4 所示。线连接了两幅图中匹配的特征点对,从图 4 中可以看到,笔画上的特征点基本上是正确的。

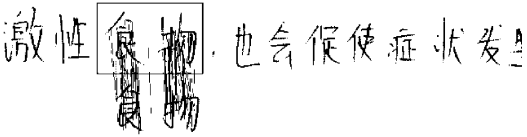


图 4 匹配特征点进一步筛选结果

Fig.4 Location information of matching points

2.3 窗口筛选和合并

结合几何约束对匹配点进行筛选之后,每个窗

口都会对应一个匹配点的数量,数量过少的窗口则被删除。由于汉字繁简程度不同导致其自身的特征点数量也不同,使得不同汉字的匹配点数量差异很大,本文采用自适应阈值来筛选窗口,即

$$tn = k \times S_{\text{num}}$$

tn 为阈值, S_{num} 为关键词图像特征点数, k 根据经验值取 0.27。

当关键词中的某一个字出现在文本图像中时,往往匹配的点也较多,比如“食物”和“食品”、“粮食”,它们之中都有“食”字,但实际上是不应该被检测出来的。对该问题采用两种策略解决,一是基于重心偏移的判决,二是基于点分布的判决。当匹配的特征点大部分集中在一个字上时,关键词图像特征点的重心横坐标和水平方向的分布会发生比较大的偏移。当前后两次偏移满足:

$$|XCent(S_0) - XCent(S_x)| > tc$$

$$|LPoi(S_0) - LPoi(S_x)| > tp$$

则当前窗口被删除。其中 $XCent(S)$ 表示关键词图像特征点的重心横坐标, $LPoi(S)$ 表示关键词图像特征点在左半侧的分布率,即左半边的点占有总点数的比例。 tc 和 tp 是两个阈值,取经验值 0.15 和 0.18。当某一窗口特征点集合与关键词图像特征点进行匹配时,会得到初始匹配点集和筛选后匹配点集两个集合,若经过误匹配点筛选后剩余点的重心横坐标和原始坐标偏移很远的时候,删除该窗口;若经过误匹配点筛选后剩余点的在关键词图像左侧分布的比例和原始比例偏离很大的时候,删除该窗口。该方法能够有效判别单个字产生误匹配的情况。如图 5 所示,两个圆圈表示关键词图像的重心,从中间的位置移动到了左侧的位置,因为筛选误匹配点以后,匹配点基本都集中在了汉字“食”上,虽然匹配点数目还比较多,但经过重心偏移的判决,该窗口仍然会被删除,“性食”也不会当做“食物”被检测出来。

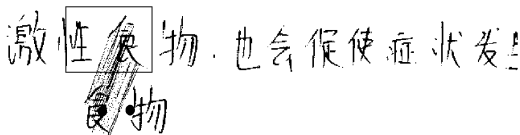


图 5 关键词图像特征点的重心偏移

Fig.5 The center deviation of feature points of keyword images

经过筛选之后,对剩下的窗口进行合并,因为有些重叠的窗口均包含了同样的关键词。当两个窗口重叠部分占整个窗口的 60% 以上,则将它们合并为一个窗口。合并之后的窗口就是最终检测到的候选

关键词窗口。

3 实验结果和分析

本文实验采用 CASIA 提供的离线手写体文本图像库 HWDB2.1^[11],该图像库由 300 位不同作者书写,每人 5 封文本图像,一共 1 500 封图像。所有的图像以 300 DP 的分辨率扫描,以 8 位灰度级图像存储。为了避免行分割对实验的影响,同时为了便于滑动窗口的提取,我们将文本图像预先按行分割,每一个文本图像分为 10 个左右的单行图像,最终得到 15 000 封图像。从中选取 50 位作者的 2 764 封图像作为实验中的待测文本图像。

评价实验结果的标准主要有召回率 R 、准确率 P 和 F 值。召回率能体现本文方法对关键词检测能力,越高越好;准确率表明检测结果中正确检测结果的比例,体现该方法的检测效率,越高越好; F 值则是对上述两个标准的综合考量。假设检测图像 Z 封,共检测到 S 个关键词,其中正确检测 Y 个,误检 N 个, $S=Y+N$,应当检测到的关键词有 T 个。上述 3 个评价标准按照式(5)计算:

$$\begin{aligned} R &= \frac{Y}{T} \\ P &= \frac{Y}{S} \\ F &= \frac{2}{1/R + 1/P} \end{aligned} \quad (5)$$

首先对 tn 和 p_4, p_5 参数进行实验。 p_4 和 p_5 表示对匹配点横纵坐标的约束程度,直接影响正确匹配点对的数目;而 tn 作为匹配点总数的阈值决定是否删除窗口,当窗口内匹配点数超过阈值 $tn = k \times S_{\text{num}}$ 则保留窗口。这对参数对最终关键词检测的结果影响较大,在算法中十分关键。在固定窗口大小和滑动步长的情况下对不同的约束阈值和筛选阈值进行实验,测试结果如表 1 所示。从表 1 数据横向分析,随着约束阈值增大,召回率逐渐升高,但准确率逐渐降低。这是因为对匹配点对的几何约束更加严格时,会有效删除更多的误匹配点对。纵向分析,随着筛选阈值 k 不断增大,召回率逐渐降低而准确率逐渐升高。这是因为较小的 k 保留了更多的正确窗口,但随之也引入很多误检测的窗口。整体来看, F 值在横向和纵向都呈现先升高后降低的趋势,在 k 为 0.27, p_4 和 p_5 为 15 和 12 时取得最大值 88.02%。

固定筛选阈值和约束阈值为上述两个最佳值,再针对不同的窗口宽度和滑动步长进行实验。原始的窗口宽度为 $W = W_s \times \text{Avg}(\text{ScI}_d) / \text{Avg}(\text{ScI}_s)$,在

此基础上分别进行增大和减小。其中 S 为关键词图 像宽度的 $1/6$ 。测试结果如表 2 所示。

表 1 不同筛选阈值和约束阈值的检测结果

Table 1 Detection results of different selection thresholds and restriction thresholds												
滑动 步长	窗口宽度											
	$W-S$			W			$W+S$			$W+2S$		
	R	P	F	R	P	F	R	P	F	R	P	F
$W/3$	77.02	85.82	81.18	81.69	89.62	85.47	78.17	93.81	85.28	77.85	94.23	85.26
$W/4$	86.07	82.03	84.01	86.58	90.42	88.46	83.80	92.69	88.02	74.52	93.73	83.03
$W/5$	83.80	86.63	85.19	83.39	86.30	84.82	81.19	91.42	86.00	77.48	93.49	84.74
$W/6$	89.26	82.65	85.83	86.59	85.49	86.04	86.25	92.12	89.08	77.85	92.56	84.57
$W/7$	86.96	82.20	84.51	86.07	88.51	87.27	84.37	89.22	86.73	77.25	93.08	84.43

表 2 不同窗口宽度和滑动步长的检测结果

Table 2 Detection results of different window widths and sliding steps										
筛选阈值 k	约束阈值 p_4, p_5									
	14, 11			15, 12			16, 13			
	R	P	F	R	P	F	R	P	F	
0.24	86.58	85.84	86.21	90.47	74.09	81.46	92.69	64.46	76.04	
0.25	84.73	90.44	87.49	87.13	83.92	85.49	87.88	73.63	80.13	
0.26	83.38	92.56	87.73	86.39	86.17	86.28	87.27	82.47	84.80	
0.27	78.12	95.80	86.06	83.80	92.69	88.02	87.27	87.43	87.35	
0.28	74.79	97.18	84.52	79.65	94.60	86.49	80.53	90.15	85.07	

从表 2 数据横向分析,随着窗口宽度的增大,召回率呈现先升高后降低的趋势,而准确率逐渐升高,这是因为太小的窗口宽度可能损失一部分正确匹配点,而太大的窗口又会引入很多误匹配点,但是由于结构约束能够筛选窗口,使得准确率大幅提升。只有适当窗口大小才能得到比较好的结果。纵向分析,滑动步长的大小对召回率和准确率的影响出现了波动,这是由于手写体汉字间距的随意性导致。整体来看, F 值在横向也表现出先升高后降低的趋势,在纵向有一定波动性。在窗口宽度取 $W+S$,滑动步长取 $W/6$ 时, F 值取得最大值 89.08%。

固定窗口大小和滑动步长为上述两个最佳值,再次对筛选阈值和约束阈值进行实验。测试结果如

表 3 所示。从表 3 和表 1 的对比可以看出,所有的召回率都有提高,大部分准确率下降,由于召回率提高的幅度更大,所以整体的 F 值有所提高,并且在相同的参数下取得最大值。本文方法为了提高关键词检测的成功率,在基于 SIFT 特征进行文字图像匹配之后,利用了文字几何信息对匹配点对进行筛选。为了验证文字结构信息对检测的效果,对结合文字几何信息之前和之后进行了对比实验,结果如表 4 所示。很显然,在没有几何信息约束的情况下仅仅依靠 SIFT 特征进行匹配,大部分图像存在误检的情况。这是由于汉字大量的重复结构造成,汉字笔画类别单一,只有依靠笔画之间的结构约束才能精准地匹配。

表 3 固定参数后不同筛选阈值和约束阈值的检测结果

Table 3 detection results with different selection threshold and restriction threshold after parameters fixed										
筛选阈值 k	约束阈值 p_4, p_5									
	14, 11			15, 12			16, 13			
	R	P	F	R	P	F	R	P	F	
0.25	87.17	89.11	88.13	92.54	82.40	87.18	93.28	70.62	80.83	
0.26	85.82	90.98	88.33	89.58	86.50	88.01	91.57	78.54	84.55	
0.27	79.23	94.76	86.30	86.25	92.12	89.08	91.57	85.57	88.47	
0.28	76.64	96.37	85.38	79.65	92.48	85.59	81.12	89.18	84.96	

表 4 结合文字几何信息之前和之后的检测结果
Table 4 Detection results before/after geometric constraint

	<i>R</i>	<i>P</i>	<i>F</i>
几何约束之前	87.58	0.17	0.34
几何约束之后	86.25	92.12	89.08

4 结束语

结合文字几何信息和 SIFT 局部特征的关键词检测方法能够有效检测特定书写人的手写关键词,并且具有较好的召回率和准确率。该方法先通过对文本图像提取 SIFT 特征,再利用滑动窗口进行分块的特征点匹配,并通过几何信息约束来筛选匹配点,最终得到检测结果。SIFT 特征良好的稳定性保证了相同书写人的相同关键词能够被检测,而利用几何关系对特征点分布的约束解决了汉字重复结构带来的误匹配问题。但是手写体汉字的随意性和书写疏密程度的多变性给检测带来了不可避免的困难,这是造成某些关键词无法检测的主要原因。

参考文献:

[1] RATH T M, MANMATHA R. Word image matching using dynamic time warping [C]//Proceedings of Conference on Vision and Pattern Recognition. Madison, USA, 2003: 521-527.

[2] LLADOS J, RUSINOL M, FORNES A, et al. On the influence of word representations for handwritten word spotting in historical documents [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2012, 26(5): 1-25.

[3] ZHANG H, WANG D H, LIU C L. Keyword spotting from online Chinese handwritten documents using one-vs-all trained character classifier [C]//2010 International Conference on Frontiers in Handwriting Recognition. Kolkata, India, 2010: 271-276.

[4] ZHANG H, LIU C L. A lattice-based method for keyword spotting in online Chinese handwriting [C]//2011 International Conference on Document Analysis and Recognition. Beijing, Chian, 2011: 1064-1068.

[5] ZHANG H, WANG D H, LIU C L. A confidence-based method for keyword spotting in online Chinese handwritten documents [C]//2012 21st International Conference on Pattern Recognition. Tsukuba, Japan, 2012: 525-528.

[6] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[7] LINDEBERG T. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention [J]. International Journal of Computer Vision, 1993, 11(3): 283-318.

[8] RODRIGUEZ J A, PERRONNIN F. Local gradient histogram features for word spotting in unconstrained handwritten documents [C]//2008 International Conference on Frontiers in Handwriting Recognition. Montréal, Canada, 2008: 18-25.

[9] 郑琪, 管海兵, 陈凯. 基于局部特征的自然场景图片中文字定位和识别方法的研究 [D]. 上海: 上海交通大学, 2011.

ZHENG Qi, GUAN Haibing, CHEN Kai. Research on text localization and recognition in image with complex scenes using local features [D]. Shanghai: Shanghai Jiaotong University, 2011: 33-45.

[10] BEIS J S, LOWE D G. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces [C]//Proceedings of Conference on Computer Vision and Pattern Recognition. San Juan, Puerto Rico, 1997: 1000-1006.

[11] LIU C L, YIN F, WANG D H, et al. CASIA online and offline Chinese handwriting databases [C]//2011 International Conference on Document Analysis and Recognition. Beijing, China, 2011: 37-41.

作者简介:



张文超,男,1988 年生,硕士研究生,主要研究方向为图像处理与模式识别。



吕岳,男,1968 年生,博士,教授、博士生导师,主要研究方向为模式识别、图像处理、智能系统。



文颖,女,1975 年生,博士,副教授,主要研究方向为图像处理、模式识别、机器学习。