一种基于模糊规则的手写体粘连数字串分割

丁 杰 杨静宇
(南京理工大学计算机科学与技术学院, 南京 210094)

摘 要 手写数字串切分是手写数字 OCR 系统中必不可少的组成部分。实际应用中一般用框格对数字的书写范围进行约束,切分过程比较容易,如果没有框格约束,手写数字串的切分就成为一个难题。针对无约束的手写数字串切分的难点,提出了一种新的粘连数字串切分方法。该方法先使用主曲线实现字符模板的笔画抽取,然后依据字符笔画的模糊特征处理笔画,最后以字符识别器提供的置信度为依据完成切分过程。为验证该新切分方法的效果,对从银行实地采集的 3 000 份真实支票进行了切分实验,其中 363 张支票存在粘连现象,切分正确率为 89.68%。实验结果表明,该算法能够有效地切分多字粘连的手写体数字串。

关键词 主曲线 模糊特征 数字串分割 笔画组合
中图法分类号: TP391 **文献标识码**: A **文章编号**: 1006-8961(2009)11-2292-07

Segmentation of Numeral Strings Based on Fuzzy Features

DING Jie, YANG Jing-yu
(*Department of Computer Science, Nanjing University of Science and Technology, Nanjing* 210094)

Abstract Numeral strings segmentation plays a significant role in the OCR systems. In many applications, numeral strings are filled in preprinted form frames. This makes, the segmentation problem easier. Other wise, the segmentation is difficult. A new segment method for handwritten numeral strings is proposed. Principal curves are selected to extract strokes of characters. The strokes in the initial group are disposed of by the fuzzy features and grouped based on the confidence of the classifiers. On the database composed of 3 000 bank checks with touching digits in 363 checks, the proposed algorithm has been evaluated qualitatively and quantitatively with an the accurate rate of 89.68%.

Keywords principal curve, fuzzy feature, segmentation of numeral strings, stroke grouping

1 引 言

在现有的大多数字符识别系统中,字符的切分是一个不可缺少的步骤,而手写数字串的分割和识别则是字符识别中一个典型的问题,其在一些特定的环境下有十分广泛的应用,如邮政、财务、税务、金融等领域。手写数字因其书写的随意性,字符本身存在断笔以及相邻字符间经常发生相碰、连笔或重叠情况,致使手写数字的切分成为一项非常困难的工作。

目前常用的手写数字分割方法有基于投影和轮廓特征的方法、基于结构特征的方法和基于字符识别的方法。但基于投影和轮廓特征的方法对具体的应用对象缺乏针对性[1],例如投影分析法对于字符发生严重倾斜或交错的情况无法处理,其主要原因是:

(1)垂直投影中连接字符的主体难以确定;

(2)当在数字字符间连笔画为直线的情况下,外轮廓分析法由于找不到轮廓线上的凹点而不太适合使用。

基金项目:国家自然科学基金重点项目(60632050);国家高技术研究发展计划(863)项目(2006AA01Z119)
收稿日期:2008-06-26;**改回日期**:2008-10-16

第一作者简介:丁 杰(1983~),男。南京理工大学模式识别与智能系统实验室计算机专业博士研究生。主要研究方向为模式识别与图像处理。E-mail:dingjie_star@163.com

基于结构特征的方法由于缺乏识别的指导,往往导致分割质量不高而影响后期的单字识别[2]。基于字符识别的方法是采用识别结果的置信度作为分割的度量,可获得较好的识别结果,而选取真实反映识别结果的置信度则成为分割成败的关键[3]。

本文使用推广的多边形主曲线(PLPC)切分算法提取数字骨架,并通过抽取笔画来首先得到初始笔画集合。但初始的笔画集合中存在笔画碎片、过渡笔画以及共用笔画,其不仅增加了笔画组合的复杂度,同时也降低了字符串切分的正确率。模糊方法是近年来的研究热点之一,在人工智能、系统科学、计算机科学研究等领域有着广泛的应用。为了更好地解决手写体数字串的切分问题,本文以笔画的模糊特征为依据来处理初始笔画集合,并依据字符识别器的置信度组合笔画来完成字符串的切分过程(图 1)。大量实验表明,该方法可以很好地应用于粘连数字串的分割工作。

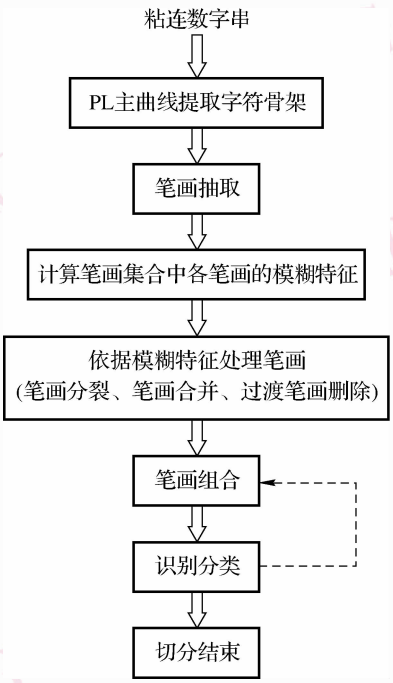


图 1　算法流程

Fig. 1　A general block diagram of our approach

2　笔画抽取

2.1　初始笔画集合

利用骨架表示原始图像,可以在保持图像的重要拓扑结构特征的前提下,减少冗余信息,从而有利于后续模糊特征的抽取。针对字符图像的骨架提取问题,多数提取算法存在以下两方面问题:

(1)对字符目标形状的细节变化比较敏感,从而影响了骨架曲线对结构复杂的拓扑特征的描述能力[4];

(2)骨架信息是以位图形式存储[5-6],体现的是像素点信息,不符合人们对字符按曲线理解的习惯。针对上述问题,主曲线算法通过寻找数据分布“中间”的曲线来反映数据的形态,与字符模板构造一致,由于 PLPC 算法[7]对字符骨架曲线进行拟合和重构,有效地减少了冗余特征,且能够更加准确地反映字符的整体拓扑结构,而且使用主曲线算法所得到的骨架信息是以矢量形式存储的,符合人们对字符的认知习惯,从而有利于特征的抽取过程。

本节采用 PLPC 算法来提取字符骨架,其所生成的字符骨架 $\boldsymbol{G}$ 由一组控制点集 $\boldsymbol{V}$ 和曲线集合 $\boldsymbol{S}$ 构成,即 $\boldsymbol{G}=(\boldsymbol{V},\boldsymbol{S})$,其中,

$$\boldsymbol{V}=\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n\}\quad \boldsymbol{v}_i\subset\mathbf{R}^d\qquad(1\leqslant i\leqslant n)\tag{1}$$

$$\boldsymbol{S}=\{\boldsymbol{s}_1,\cdots,\boldsymbol{s}_m\}\tag{2}$$

$$\boldsymbol{s}_i=\{(\boldsymbol{v}_{i_1},\boldsymbol{v}_{i_2}),(\boldsymbol{v}_{i_2},\boldsymbol{v}_{i_3}),\cdots,(\boldsymbol{v}_{i_{k-1}},\boldsymbol{v}_{i_k})\}$$
$$1<i_1,\cdots,i_k\leqslant n\tag{3}$$

初始笔画集合如图 2 所示。

图 2　初始笔画集合示例

Fig. 2　Examples of initial stroke sets

2.2 模糊特征抽取

本节给出了 9 种特征片段(表 1),以及相应的 11 种模糊特征[8],包括直线特征:笔画段与直线的相似程度 *MSTR*;直线类型特征:笔画段与垂直线的相似程度 *MVL*、笔画段与水平线的相似程度 *MHL*、笔画段与正斜线的相似程度 *MPS*、笔画段与反斜线的相似程度 *MNS*;弧线特征:笔画段的弯曲程度 *MARC*;弧线类型特征:笔画段与 A 型弧的相似程度 *MAL*、笔画段与 U 型弧的相似程度 *MUL*、笔画段与 C 型弧的相似程度 *MCL*、笔画段与 D 型弧的相似程度 *MDL*、笔画段与 O 型弧的相似程度 *MOL*(表 2)。此外,相邻笔画段的重叠程度 *MOP* 表示相邻笔画的水平方向重叠特征,笔画段的长短程度 *MLEN* 表示笔画长度的模糊特征。

表 1 特征片段

Tab. 1 The meaningful segments

直线(straight lines)片段	弧线(arcs)片段
竖线(vertical line)	A 型弧(A-like)
横线(horizontal line)	U 型弧(U-like)
斜线(positive slant)	C 型弧(C-like)
反斜线(negative slant)	D 型弧(D-like)
	O 型弧(O-like)

表 2 模糊特征

Tab. 2 The fuzzy features

直线特征(straightness)	直线类型特征(line-types)	弧线特征(arc-ness)	弧线类型特征(curve-types)
MSTR(μ straightness)	*MVL*(μ vertical line) *MHL*(μ horizontal line) *MPS*(μ positive slant) *MNS*(μ negative slant)	*MARC*(μ arc-ness)	*MAL*(μ A-like) *MUL*(μ U-like) *MCL*(μ C-like) *MDL*(μ D-like) *MOL*(μ O-like)

设笔画 $\boldsymbol{s}_i,\boldsymbol{s}_j(1\leqslant i,j\leqslant m)$ 的左、右界分别为 $\boldsymbol{x}_i^{\mathrm{L}}$,$\boldsymbol{x}_i^{\mathrm{R}}$ 和 $\boldsymbol{x}_j^{\mathrm{L}}$,$\boldsymbol{x}_j^{\mathrm{R}}$,将笔画 $\boldsymbol{s}_i$ 与 $\boldsymbol{s}_j$ 在水平方向重叠的模糊特征记为 $MOP_{i,j}$,则相应的模糊测度为

$$\mu_{i,j}=\frac{\Lambda(\boldsymbol{x}_i^{\mathrm{R}}-\boldsymbol{x}_j^{\mathrm{L}},\boldsymbol{x}_j^{\mathrm{R}}-\boldsymbol{x}_i^{\mathrm{L}})}{\max(\boldsymbol{x}_i^{\mathrm{R}}-\boldsymbol{x}_j^{\mathrm{L}},\boldsymbol{x}_j^{\mathrm{R}}-\boldsymbol{x}_i^{\mathrm{L}})} \tag{4}$$

$$\Lambda(x,y)=\max(\min(x,y),0) \tag{5}$$

若将 $\boldsymbol{s}_i$ 的笔画长度的模糊特征记为 $MLEN_i$,则相应的模糊测度为

$$\mu_i^{\mathrm{Len}}=\frac{\sum_{l=1}^{N-1}d_{i_l,i_{l+1}}}{\boldsymbol{x}_k^{\mathrm{R}}-\boldsymbol{x}_k^{\mathrm{L}}} \tag{6}$$

其中 $d_{i_l,i_{l+1}}$ 表示笔画 $\boldsymbol{s}_i$ 上控制点 $\boldsymbol{v}_{i_l}$ 与 $\boldsymbol{v}_{i_{l+1}}$ 之间的欧氏距离。

与以上 13 种模糊特征对应的模糊函数的隶属度的 7 个子区间分别为 VS(very small),S(small),MS(middle small),M(middle),ML(middle large),L(large),VL(very large)(如图 3 所示),其值域按隶属度由低到高表示为 $\{VS,S,SM,M,ML,L,VL\}$。

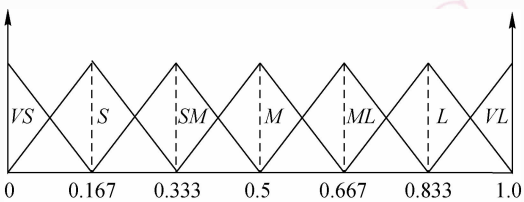


图 3 模糊函数

Fig. 3 Fuzzy membership function

2.3 基于模糊特征的笔画集合处理

2.3.1 笔画分裂

通过实际观察发现,字符间通过过渡笔画形成粘连是字符串粘连的主要方式,称为过渡粘连。过渡笔画的端点分为连接点与非连接点两类。连接点是字符骨架图中度数大于 2 的点。不难发现,具有非连接点的过渡笔画往往包含了相邻两个字符的笔段(如图 2(b)原图所示),需要予以分割。文献[9]针对单个字符提出了依据笔画方向分裂笔画,即认为分割点处的笔画书写方向应有较大的变化。本节针对过渡笔画的特点,提出了适合过渡笔画的分割规则。具体做法为以笔画左端点为起始点遍历笔画,依据规则 1

确定当前遍历点是否为分割点(见图 4)。

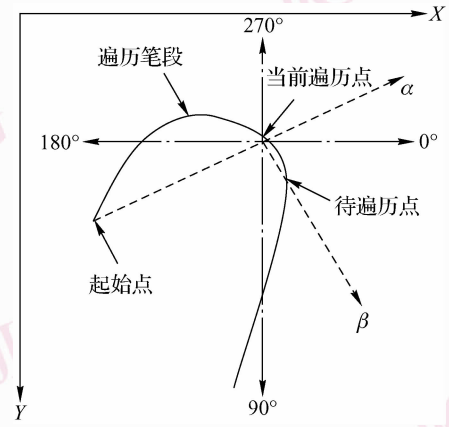


图 4　笔画遍历示例

Fig. 4　Example of stroke traversal

规则 1　若 $|\alpha-\beta|>\theta$,则以当前遍历点为分割点分割笔画。其中 α 表示起始点与当前遍历点之间连线的方向;β 表示当前遍历点与相邻的待遍历点之间连线的方向;θ 为阈值,可依据表 3 取值。由图 4 可见,需予以分裂的过渡笔画通常具有较大的弧度,即其模糊特征 *MAL* 的取值较大,分裂点选取为笔画段中笔画方向的突变点,若实验中的笔画方向变化大于 30°,即认为是分裂点。

表 3　笔画分割规则

Tab. 3　The rules of stroke segmentation

模糊特征		角度区间(°)		阈值 θ
过渡笔画	当前笔段	α	β	
MAL = *M* 或 *L* 或 *VL*	*MSTR* = *L* 或 *VL*	(270°,360°)	[0°,45)	30°
			[315°,360°)	30°
			[45°,315°)	120°

2.3.2　笔画合并

由于初始笔画集合中的笔画数目巨大,这一方面增加了笔画组合的复杂度,另一方面降低了字符串分割的正确率,因此需要对相关笔画进行合并。本文采用基于连接点的笔画合并方式,即笔画先按连接点分组,然后将笔画在组内进行合并。

设公式(2)的 $\boldsymbol{S}$ 为笔画集合,则与连接点 v 对应的笔画子集为

$$\boldsymbol{S}(v)=\{s_i \mid (\boldsymbol{v}_{i_1}=\boldsymbol{v}\vee\boldsymbol{v}_{i_N}=\boldsymbol{v}),\ \boldsymbol{s}_i\in\boldsymbol{S}\quad 1\leqslant i\leqslant n\}\tag{7}$$

规则 2　(O 型合并) $\boldsymbol{s}_i,\boldsymbol{s}_j\in\boldsymbol{S}(v)$,如果满足表 4 给出的规则,则予以合并。

表 4　O 型合并规则

Tab. 4　The rules of O-like combination

合并类型	模糊特征		
	$\boldsymbol{s}_i$	$\boldsymbol{s}_j$	$\boldsymbol{s}_i\cup\boldsymbol{s}_j$
O 型	*MAL* = *L*,*VL*	*MUL* = *L*,*VL*	*MOL* = *VL*
	MCL = *L*,*VL*	*MDL* = *L*,*VL*	
	MSTR = *VL*	*MAL* = *VL*	
		MUL = *VL*	
		MCL = *VL*	
		MDL = *VL*	

规则 3　(线型合并) $\boldsymbol{s}_i,\boldsymbol{s}_j\in\boldsymbol{S}(v)$,如果满足表 5 给出的规则,则予以合并。

表 5　线型合并规则

Tab. 5　The rules of line-like combination

合并类型	模糊特征		
	$\boldsymbol{s}_i$	$\boldsymbol{s}_j$	$\boldsymbol{s}_i\cup\boldsymbol{s}_j$
线型	*MSTR* = *VL*, *MVL* = *ML*,*L*,*VL*	*MSTR* = *VL*, *MVL* = *ML*,*L*,*VL*	*MSTR* = *L*,*VL*
	MSTR = *VL*, *MHL* = *ML*,*L*,*VL*	*MSTR* = *VL*, *MHL* = *ML*,*L*,*VL*	
	MLEN = *VS*,*S*, *SM*,*M*	*MLEN* = *VS*,*S*, *SM*,*M*	

规则 4　(>型合并) $\boldsymbol{s}_i,\boldsymbol{s}_j\in\boldsymbol{S}(v)$,如果满足表 6 给出的规则,则予以合并。

表 6　>型合并规则

Tab. 6　The rules of >-like combination

合并类型	模糊特征		
	$\boldsymbol{s}_i$	$\boldsymbol{s}_j$	$\boldsymbol{s}_i\cup\boldsymbol{s}_j$
>型	*MSTR* = *VL*, *MNS* = *L*,*VL*	*MSTR* = *VL*, *MPS* = *ML*,*L*,*VL*	*MDL* = *L*,*VL*
	MSTR = *VL*, *MPS* = *L*,*VL*	*MSTR* = *VL*, *MNS* = *ML*,*L*,*VL*	

2.3.3　过渡笔画删除

过渡粘连是定位格内字符的主要粘连方式,这类粘连的主要特征是粘连字符间存在过渡笔画。这一笔画有可能是字符的笔画及其延长部分(见图 5(a)),也有可能不属于其中任何一个字符(见图 5(b))。对于第 2 种情况,需要将该笔画删除。

通过实际观察发现,需要删除的过渡笔画与其左边粘连的数字有关。这些数字一般为 0(普通 0,开口的 0)、3(普通 3,下部有孔的 3)和 6(普通 6,开口的 6)。

(a) (b)

图 5 两类过渡笔画示例

Fig. 5 Two types of transitional connection

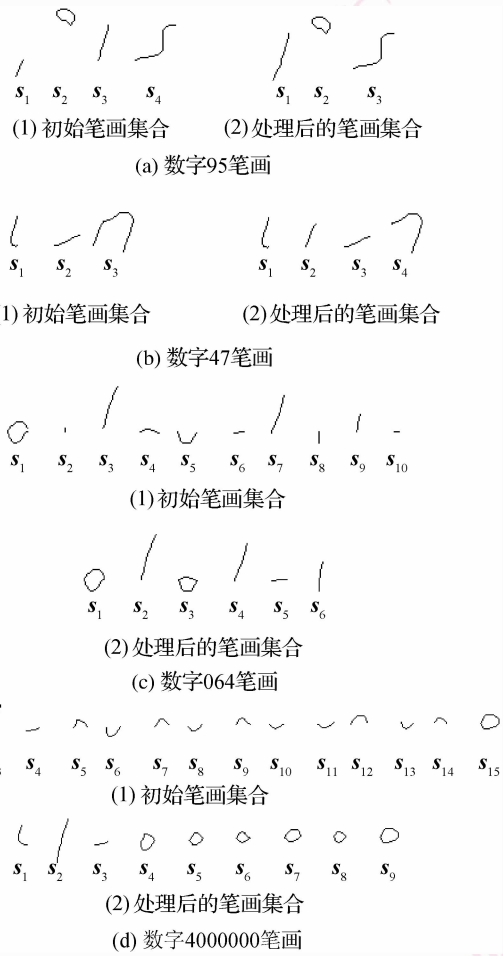
规则 5 笔画 s_i 及其左邻笔画 $s_j(1\leqslant i,j\leqslant m)$，如果满足表 7 给出的规则，则删除 s_i。

表 7 删除规则

Tab. 7 The rules of deletion

s_i 的模糊特征	$MOP_{i,j}$	s_j 的模糊特征
$MSTR=VL$, $MHL=ML,L,VL$	VS, MS	$MOL=VL$ $MUL=L,VL$ $MOL=S,MS,M$ $MDL=L,VL$ $MOL=S,MS,M$

初始笔画集合经过笔画分割、合并以及过渡笔画删除后的笔画集合如图 6 所示。



(1) 初始笔画集合 (2) 处理后的笔画集合

(a) 数字95笔画

(1) 初始笔画集合 (2) 处理后的笔画集合

(b) 数字47笔画

(1) 初始笔画集合

(2) 处理后的笔画集合

(c) 数字064笔画

(1) 初始笔画集合

(2) 处理后的笔画集合

(d) 数字4000000笔画

图 6 图 1(a) ~ 图 1(d)的初始笔画集合处理示例

Fig. 6 Stroke disposal result of stroke sets in Fig. 1(a) ~ Fig. 1(d)

3 字符串分割

3.1 笔画组合

将式(2)笔画集合 $\boldsymbol{S}$ 中的各笔画，按其水平位置排列形成的笔画序列 $\hat{s}_1,\hat{s}_2\cdots,\hat{s}_m$ 记为 $\boldsymbol{Seq}^{S}_{1,m}$，其子序列 $\hat{s}_i,\hat{s}_{i+1},\cdots,\hat{s}_j(1\leqslant i,j\leqslant m)$，记为 $\boldsymbol{Seq}^{S}_{i,j}$。本节使用动态规划算法完成笔画组合(见图 7)。其中，$N_k(0\leqslant k\leqslant n)$表示第 k 层节点，n 为粘连数字串中的数字个数，依据波形分析估计[9-10]，$C(\cdot)$表示字符识别器提供的置信度，$\delta_{i,j}$表示将笔画序列 $\boldsymbol{Seq}^{S}_{1,j}$ 切分为 i 个独立字符的权重，ζ_i 表示第 i 个字符的首笔画在笔画序列中的下标，笔画组合结果由 $\zeta_i(1\leqslant i\leqslant n)$表征，则最优笔画组合满足式(8) ~ 式(10)。

$$\delta_{0,0}=1 \tag{8}$$

$$\delta_{i,j}=\max_{i,k=1}^{j-1}(\min(\delta_{i-1,k},C(\boldsymbol{Seq}^{S}_{k+1,j}))) \quad 1\leqslant i\leqslant n \tag{9}$$

$$\zeta_i=\arg(\max_{j=1}^{m}\delta_{i,j}) \tag{10}$$

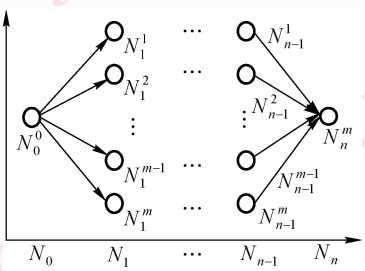


图 7 笔画组合图

Fig. 7 Stroke grouping

3.2 置信度修正

由于笔画组合具有一定的随意性和复杂性，因而识别器提供的置信度往往不能真实地反映识别结果。由于一些非字符模式的笔画组合也可能具有较高的置信度，因此识别器的可信度成为切分过程的关键和瓶颈。本节依据字符笔画的模糊特征对识别器的置信度进行修正，以获取更为可靠的分割依据。

文献[11]通过抽取各训练样本的特征片段(表 1)，以特征片段的模糊特征创建知识库，并依据模糊评判规则 max-min 方法计算待识样本与训练样本间的相似度[12]。相似度计算如下：

设有 c 类模式 $\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,\cdots,\boldsymbol{\omega}_c$，第 i 个模式类 $\boldsymbol{\omega}_i$ 的训练样本有 n_i 个，记为 $\boldsymbol{X}_{i,s}(1\leqslant s\leqslant n_i)$。给定待

识样本 $\boldsymbol{X}$,其含有 m 个特征片段 $\boldsymbol{F}_1,\boldsymbol{F}_2,\cdots,\boldsymbol{F}_m$,$R(\boldsymbol{X})$表示 3.1 节字符识别器的识别结果,其对应的模式类别记为 $\boldsymbol{\omega}_r$,则 $\boldsymbol{X}$ 的相似度为

$$S(\boldsymbol{X})=\max_{s=1}^{n}(\min(\text{min-max}(\boldsymbol{X}_{F_1},\boldsymbol{X}_{r,s,F_1}),\cdots,\text{min-max}(\boldsymbol{X}_{F_m},\boldsymbol{X}_{r,s,F_m})))\tag{11}$$

式中,n 表示第 r 个模式类 $\boldsymbol{\omega}_r$ 中特征片段数为 m 的样本个数。修正后的字符识别器的置信度如下:

$$\delta(\boldsymbol{X})=\begin{cases}1 & S(\boldsymbol{X})>0.5\\ 0 & S(\boldsymbol{X})\leqslant 0.5\end{cases}\tag{12}$$

修正后的识别结果为

$$\hat{C}(\boldsymbol{X})=C(\boldsymbol{X})\delta(\boldsymbol{X})\tag{13}$$

4　实验结果及分析

4.1　本文算法的实验结果

实验采用的数据是从我国银行实地采集的 3 000张现行支票,其中 363 张存在粘连现象,共计有 352 个两字粘连字串和 113 个多字粘连字串。应用本文提出的方法对其中存在字符粘连现象的支票进行处理,部分分割结果见图 8,实验的分割结果统计见表 8。分割失败的原因主要如下:

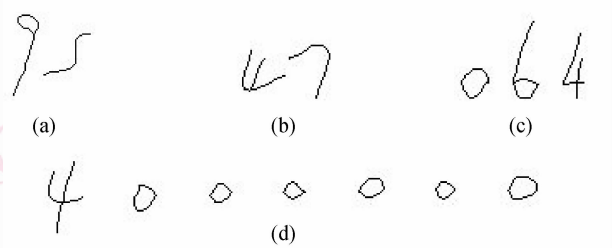


图 8　与图 1(a) ~ 图 1(d)对应的分割结果

Fig. 8　Segmentation result of strings in Fig. 1(a) ~ Fig. 1(d)

表 8　本文分割方法的分割结果统计

Tab. 8　The statistical radio of method used in the paper

	粘连数字串数目	正确分割率(%)	错误分割率(%)
两字粘连	352	91.19	8.81
多字粘连	113	84.96	15.04
总计	465	89.68	10.32

(1) 图像质量较差,大量噪声干扰导致字符笔画存在严重断裂和缺失。

(2) 字符笔画序列中存在错位笔画。笔画集合经过笔画合并,笔画碎片大幅减少,这虽然在一定程度上减少了序列中所包含的错位笔画,却仍然无法避免错位笔画的出现。由于无法预知笔画序列错位的位置,任一笔画均有可能成为错位笔画,因此对其进行处理将大幅增加算法的复杂度。

4.2　与滴水算法的比较

滴水(drop-falling)算法[13]是模拟水滴从高处向低处滴落的过程来进行字符串的切分,由于水滴的滴落过程限定水滴只能向下或水平移动,因此水滴经过的路径就是切分路径。滴水算法是一种简单有效的手写字符串的切分算法,已得到比较广泛的应用。但是用滴水算法搜索切分路径时,盲目性较大,在噪声干扰或字符间存在连笔时,往往切分效果不佳。本文算法与传统的滴水算法的分割结果比较见表 9,由表 9 可见,在噪声或字符间存在连笔的情况下,本文算法有效地提高了字符串切分的正确率。

表 9　本文分割方法与传统滴水算法分割结果比较

Tab. 9　The comparative result with the segmentation method using drop-falling algorithm

分割算法	粘连数字串数目	正确分割率(%)	错误分割率(%)
滴水算法	465	82.58	17.42
本文方法		89.68	10.32

5　结　论

如何有效进行非约束手写数字串的切分是比较困难的问题,由于字符的粘连情况复杂,因此实际可以应用的切分算法并不多。本文方法先通过主曲线分析的方法抽取字符笔画,然后在此基础上依据模糊规则进行笔画处理,这一方面降低了笔画组合的复杂度,另一方面也提高了分割的正确率,如何进一步降低算法的复杂度将是下一步工作的研究重点。

参考文献(References)

1　Lu Yi, Shridhar M. Character segmentation in handwritten words-an overview[J]. Pattern Recognition, 1996, **29**(1):77-96.

2　Chen Y K, Wang J F. Segmentation of single-or multiple-touching handwritten numeral string using background and forground analysis [J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2000, **22**(11):1304-1317.

3 Casey R, Lecolinet E. A survey of methods and strategies in character segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, **18**(7):690-706.

4 Choi W P, Lan K M, Siu W C. An efficient algorithm for extraction of a Euclidean skeleton[A]. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal[C], New York, USA, 2002: 3241-3244.

5 Rockett P I. An improved rotation-invariant thinning algorithm[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, **27**(10): 1671-1674.

6 Ahmed M, Ward R. A rotation invariant rule-based thing algorithm for character recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, **24**(12): 1672-1678.

7 Kegl B, Krzyzak A. Piecewise linear skeletonization using principal curves[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, **24**(1): 59-74.

8 Batuwita K B M R, Bandara G E M D C. Meaningful segmentation of offline individual handwritten numeric characters [A]. In: Proceedings of IEEE Internatinal Conference on Fuzzy Systems[C], Vancouver BC, Canada, 2006: 1500-1505.

9 Lu Zhong-kang, Chi Zhe-ru, Siu Wan-chi. Length estimation of digit string using neural networks with structure based[J]. SPIE/IS&T Journal of Electronic Imaging, 1998, **7**(1): 79-85.

10 Zhang Chuang, Wu Ming, Guo Jun. Algorithm of the length estimation of unconstrained handwritten connected numeral strings[J]. Journal of Beijing University of Posts and Telecommunications, 2004, **27**(3): 63-67. [张闯,吴铭,郭军. 非限制自由手写体粘连数字个数的判断[J]. 北京邮电大学学报, 2004, **27**(3): 63-67.]

11 Bandara G E M D C, Pathirana S D, Ranawana R M. Use of fuzzy feature descriptions to recognize handwritten alphanumeric characters [A]. In: Proceedings of 1st Conference on Fuzzy Systems and Knowledge Discovery[C], Singapore, 2002: 1586-1591.

12 Bandara G E M D C, Batuwita K B M R. Fuzzy recognition of offline handwritten numeric characters [A]. In: Proceedings of IEEE Internatinal Conference on Cybernetics and Intelligent Systems[C], Bangkok, Thailand, 2006:1-5.

13 Congedo G, Dimauro G, Impedovo S, *et al*. Segmentation of numeral strings[A]. In: Proceedings of 3th International Conference on Document Analysis and Recognition[C], Montreal Que, Canada, 1995: 1038-1041.