

高质量 Bezier 曲线描述轮廓库 自动生成算法

马小虎 潘志庚

(浙江大学 CAD & CG 国家重点实验室 310027)

关键词: 轮廓字库, Bezier, 参数型最小二乘法.

1 基本概念

随着排版和印刷系统的蓬勃发展 and 广泛应用,人们迫切希望有高质量的汉字输出,而汉字输出质量取决于汉字字型的表示方法.在点阵法、向量法和轮廓法^[1]三种方法中,轮廓法描述是实现高质量低成本输出的重要途径,一般都采用低次(不大于三次)Bezier曲线来描述字型轮廓.本文提出一种Bezier曲线描述轮廓库自动生成算法,为后面的叙述方便,下面先给出一组定义.

定义 1. 构成轮廓描述汉字的直线段称为直线笔划,可用始点和终点坐标表示为(S, E);而把构成轮廓描述的曲线段称为 Bezier 笔划,用导引点(P_0, P_1, P_2, P_3)表示.

定义 2. 一个汉字 a ,把对应轮廓描述汉字的所有直线笔划所组成的集合 $S(a)$ 和所有 Bezier 笔划所组成的集合 $B(a)$ 称为汉字 a 的轮廓信息,记为 $I_c(a)$,有

$$I_c(a) = (S(a), B(a))$$

定义 3. 一个汉字集 A 的所有汉字轮廓信息构成的集合称为该汉字集 A 的轮廓库,记为 $L_c(A)$,有

$$L_c(A) = \{I_c(a), | a \in A\}$$

定义 4. 求与轮廓线拟合的连续曲线的过程称为轮廓线的连续化过程,简称连续化.显然,连续化时要把轮廓线适当划分,并逐段进行.

定义 5. 轮廓段是轮廓线的一部分,表示为 \widehat{AB} , 其中 A 和 B 是两个端点.

定义 6. 轮廓段 \widehat{AB} 的拱极点定义为该轮廓线上的点到 \overline{AB} (直线段)的距离最大的点, \widehat{AB} 的拱高定义为该点到 \overline{AB} 的距离.

定义 7. 凹点是指交叉笔处、笔端粘连处和折笔拐弯处(内侧)的点,凸点是指笔端所

在处和折笔拐弯处(外侧)的点. 凹凸点的确定采用求轮廓线上各点的 1-曲率、2-曲率和 3-曲率的方法¹⁾.

2 生成汉字轮廓库算法

2.1 算法 AL: 生成汉字轮廓库

AL1. 对汉字集 A 中的每个汉字做 AL2.

AL2. 根据汉字 a 的点阵信息提取轮廓线, 并对每条封闭轮廓线做 AL3.

AL3. 将整个封闭轮廓线分为若干轮廓段, 对每一轮廓段做 AL4.

AL4. 若该轮廓段的拱高 d_m 足够小, 则抽取直线笔划的始点和终点 (S, E), 并转 AL8.

AL5. 求拟合于给定轮廓段的不超过三次的 Bezier 笔划.

AL6. 检查求出的 Bezier 笔划与原轮廓段的拟合程度, 若误差较大则做 AL7, 否则转 AL8.

AL7. 将轮廓段细分, 对划分后的每段轮廓做 AL5.

AL8. 把求得的直线笔划增加到 $S(a)$ 中, Bezier 笔划增加到 $B(a)$ 中, 在单个汉字轮廓连续化结束时, 把 $S(a)$ 、 $B(a)$ 加到 $I_c(a)$ 集合中.

本算法的执行结果为自动生成轮廓线汉字库 $L_c(A)$.

2.2 算法 B: 轮廓线提取(算法 AL 中 AL2 的细化)

设 MTR 是存放原点阵的 $N \times N$ 维数组 (N 为汉字点阵水平向点数), MTR[i, j] 取值 0 或 1, R 是轮廓点列集合, b 是轮廓跟踪始点, c 是被检查的当前点, d 是搜索方向(如图 1(a) 所示). first 是一标志, 当找到轮廓上下一点时它为 TRUE.

B1. 选轮廓上一点 b , 使其 4-近邻(有关 N -近邻的定义如图 1(b))^[2] 不在集合 R 中.

B2. $c = a$; $d = 6$; first = TRUE;

B3. while (($c \neq b$) or (first = TRUE)) do

begin

found = FALSE; count = 3;

while ((found = FALSE) and (count > 0)) do

begin

count = count - 1;

if (c 的 $(d - 1)$ -近邻 e 在 R 内)

then begin $c = e$; $d = (d - 2)$; found = TRUE; end

else if (c 的 d -近邻 e 在 R 内)

then begin $c = e$; found = TRUE; end

else if (c 的 $(d + 1)$ -近邻 e 在 R 内)

then begin $c = e$; found = TRUE; end

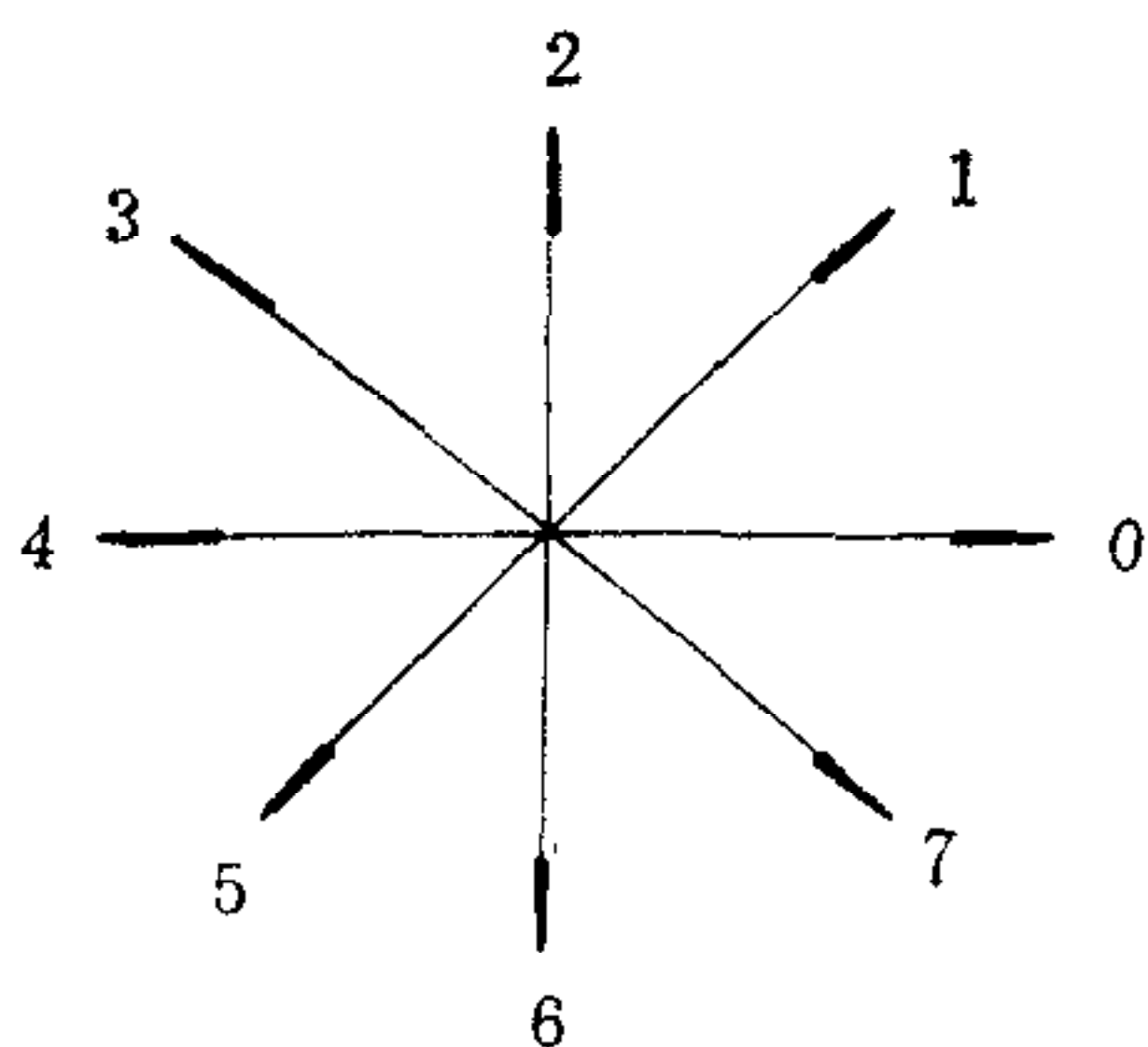
1) 参见马小虎“使用 Bezier 曲线拟合汉字字型研究”(南京大学硕士论文) 1991 年 7 月.

```

else d = d + 2;
end
end
end

```

B4. 算法结束.



B_3	B_2	B_1
B_4	C	B_0
B_5	B_6	B_7

(a) 遍历方向

(b) 象素 C 的 N-近邻 (0 ≤ N ≤ 7) 象素 B_N 的位置

图 1 N-近邻定义和遍历方向

算法说明:

- (i) 算法中所有对 d 的加减操作为模 8 运算.
- (ii) 该算法除对外轮廓线使用一次外, 还须对区域的每个孔洞使用一次.

2.3 轮廓线初始划分点集的确定

根据 Bezier 曲线的性质和所要拟合的汉字字型的特点, 在确定轮廓线上的初始划分点集时, 制定以下规则:

- 对楷体进行拟合, 初始划分点集定于: 轮廓线上凸点的两侧和轮廓线上凹点的两侧.
- 对黑体进行拟合, 初始划分点集定于: 轮廓线上凸点处和轮廓线上凹点的两侧.

2.4 求拟合轮廓段的 Bezier 曲线(算法 AL 中 AL5 细化)

已知轮廓段 \widehat{AB} 的 $n + 1$ 个离散点 Q_0, Q_1, \dots, Q_n , 求拟合于 \widehat{AB} 的低次(≤ 3)Bezier 曲线使其满足(该 Bezier 曲线 P 导引点设为 P_0, P_1, P_2, P_3):

- (i) $P(t)$ 与 $Q_i = (x_i, y_i)$ ($i = 0, 1, \dots, n$) 之间的最大偏差小于 σ (σ 为给定常数, 如取 $\sigma = 0.5$).
- (ii) 相邻的两个笔划轮廓段的拟合曲线交接处达到 C^1 连续.

求解这一问题的方法有反求顶点法、几何作图法^[3]等, 本文给出一种新方法即参数型最小二乘法, 下面描述这一方法.

取逼近函数为 $F(t)$, $F(t)$ 是三次 Bezier 曲线 $P(t)$:

$$P(t) = (1 - t)^3 P_0 + 3(1 - t)^2 t P_1 + 3(1 - t) t^2 P_2 + t^3 P_3$$

则基函数为:

$$\begin{cases} \varphi_0(t) = (1 - t)^3 \\ \varphi_1(t) = 3(1 - t)^2 t \\ \varphi_2(t) = 3(1 - t) t^2 \\ \varphi_3(t) = t^3 \end{cases} \quad (0 \leq t \leq 1)$$

待定系数为 P_1, P_2 , 其中 $P_0 = A = Q_0, P_3 = B = Q_n$ 为端点约束条件. 按累加弦长法计算参数 t 值:

设 $l_i = |Q_i Q_{i-1}| (i = 1, 2, \dots, n); l_0 = 0;$

$$L = \sum_{i=0}^n l_i$$

$$\text{则 } t_i = \sum_{k=0}^i \frac{l_k}{L}$$

显然有 $t_0 = 0, t_1 = 1$

用方差来衡量逼近程度, 有:

$$\phi(P_1, P_2) = \sum_{k=0}^n d_k \left| \sum_{j=0}^3 P_j \varphi_j(t_k) - Q_k \right|^2 \quad (2.1)$$

其中 d_k 为点 Q_k 处的加权系数. 要使加权方差(2.1)达到最小, 必须:

$$\frac{\partial \phi}{\partial P_i} = 2 \sum_{k=0}^n d_k \left[\sum_{j=0}^3 P_j \varphi_j(t_k) - Q_k \right] \varphi_i(t_k) = 0 \quad (i = 1, 2)$$

即:

$$\sum_{j=0}^3 P_j \sum_{k=0}^n d_k \varphi_j(t_k) \varphi_i(t_k) = \sum_{k=0}^n d_k Q_k \varphi_i(t_k)$$

令

$$\begin{cases} \sum_{k=0}^n d_k \varphi_j(t_k) \varphi_i(t_k) = S_{ij} \\ \sum_{k=0}^n d_k Q_k \varphi_i(t_k) = T_i \end{cases} \quad (i = 1, 2)$$

则得方程组:

$$\sum_{j=0}^3 P_j S_{ij} = T_i \quad (i = 1, 2) \quad (2.2)$$

求解线性方程组(2.2)即可求得 P_1, P_2 的值.

3 算法应用和讨论

本文提出的算法已在 Sun 3/60 工作站上用 C 语言实现, 用该算法抽取的汉字轮廓库可增加到我们实现的汉字 PostScript 解释器¹⁾中, 该字库支持汉字的任意旋转和无级变倍, 在放大倍数很大时仍保持不变形. 图 2 是利用本文提出的方法所生成的汉字轮廓库进行汉字输出的例子, 输出的软件为汉字 PostScript 解释器.

本文提出的对给定轮廓段 \widehat{AB} 求拟合于它的三次 Bezier 曲线的导引点方法比几何作图法^[3]要优越. 使用几何作图法的结果是会使一部分的轮廓段的拟合效果较差, 使得为了保证拟合效果需要细分更多较短的轮廓段, 从而占用较多的存贮量; 而本文的参数型

1) 参见潘志庚“网络环境下窗口系统的研究与应用”(南京大学硕士论文), 1990.7.



图 2 汉字输出图例

最小二乘法不存在这样的问题,可用于较好地拟合“S”型曲线而不必再分段。

致谢: 本文的研究工作在张福炎教授指导下完成,在此表示感谢。

参 考 文 献

- [1] Jurgen Schonhut, Page Description Languages, in: Advance in Computer Graphics V, pp 120—190, Springer-Verlag, 1990.
- [2] Theo Pavlidis “Algorithm For Graphics and Image Processing” Computer Science Press INC, 1982.
- [3] 董蕴美,卫平,一个将黑白图象数据由离散化为连续的方法,计算机学报, 12(1989),(8).

THE AUTOMATIC GENERATION ALGORITHM OF HIGH-QUALITY OUTLINE FONT USING BEZIER CURVE

MA XIAOHU PAN ZHIGENG

(State Key Lab. of CAD&CG, Zhejiang Univ., Hangzhou, 310027)

Key words: Outline font, Bezier, paramatic least square method