

A Research on the Stroke-Segment-Mesh (SSM) Glyph Depiction

Method of Chinese Character

Min Lin^{1,3}, Rou Song^{1,2}, Shi-Li Ge²

¹College of Computer Science & Technology, Beijing University of Technology, Beijing
100022

²College of Information Sciences, Beijing Language and Culture University, Beijing
100083

³College of Computer & Information Engineering, Inner Mongolia Normal
University, Huhhot 010022

E-MAIL: linmin@blcu.edu.cn, songrou@blcu.edu.cn

Abstract *There are defects in the current depiction method of computerized Chinese character glyph, especially on the aspect of feature selection and computation for comparing glyphs, which cause many problems, such as the input of wrongly written characters, variant forms of characters in ancient literatures, and combined-characters and automatic glyph comparing. This research offers an application-oriented depiction method of Chinese character glyph, which possesses the characteristics of suitable granular degree, ambiguity-free, and standardized basic units and which can be applied in the depiction of any possible different glyph skeletons (including wrongly written characters, variant form of a Chinese character, and combined-characters). This research also offers the classifications of basic strokes, compound strokes, algorithms of automatic stroke extraction and structural relation computation based on stroke-segments-mesh depiction of glyphs. The experiment indicates that this method can be applied in the descriptive input of any glyph and computation for comparing either the whole or part of glyphs.*

1. Introduction

There are only two kinds of Chinese character information in computer, character fonts for output

and internal code for identification. The absence of unified and efficient formal description and supporting computation for comparing Chinese character glyphs causes the difficulty of free input of any possible characters demanded by the method of glyph drawing, especially for those special Chinese characters, such as wrongly written characters, variant forms of characters in ancient literatures, and folk combined-characters. Further research of comparing glyphs with computer is impossible, either. All these problems seriously hinder the application of computer in the fields of Chinese character research, Chinese language teaching and international dissemination^[1] and historical and cultural research.

There are many researches on Chinese character glyph depiction. The depiction methods are usually the classification of Chinese character structure type according to human perception habits, which includes upper-down, left-right, and inclusive structure, etc. and the hierarchical description of Chinese character according to the structure type units in human perception, such as radicals and strokes. The research achievements based on this idea include Chinese Character Information Dictionary, Chinese Character Component Standard, Ideographic Description Sequence (IDS), mathematical expression of Chinese characters, Character Description Language (CDL), and so on^[2-9]. This glyph

depiction method of Chinese character treats characters as recursive combination from large components to small ones, which is powerful in glyph productivity and overcomes the drawbacks of closed large character set plan to some extent. The method is effective on many Chinese characters but since the selection of glyph features depends on human knowledge, the granular degree of the features is too large, the features are not consistent, and there are lots of ambiguities and absences in the description of glyphs, the processing of automatic extraction and comparing computation of glyph features cannot be implemented and the application problems listed above cannot be solved.

The technologies of handwriting Chinese character (HWR) recognition and OCR recognition mainly depend on glyph features of characters. Many character recognition models are constructed in previous researches^[12-16] and these glyph computing models depict Chinese character glyphs with global statistical features extracted from a large quantity of character image sample. The features selected are steady, anti-interference and robust. The models can recognize either complex printed Chinese characters or seriously distorted handwriting Chinese characters with a good recognition performance. However, there are also drawbacks in these glyph models. These models only cover the features of all Chinese characters in the sample whereas the wrongly written characters, variant forms of characters in ancient literatures, and folk combined-characters which are not collected in the sample cannot be precisely described and the related recognition performance cannot be ensured. In addition, it is hard to analyze the features logically or depict part of a character glyph with the global statistical features based on whole Chinese character images. It is hard to set up the corresponding relationship between them and glyph features of human perception, such as stokes, radicals, and structure types. Therefore, these kinds of models cannot be applied in the fields except for whole Chinese

character recognition. It is impossible for these character recognition models to perform the task of extracting different hierarchical character elements from a whole character, such as stokes or radicals, for analysis and computation in Chinese character researching and teaching.

Combining theoretical and applicational demands, this research offers a stroke-segment-mesh glyph depiction method of Chinese character. This method selects basic unit features with suitable granular degree so as to depict different hierarchical character structure features systematically, to depict the similarity and difference of any possible character skeletons including correctly written characters, wrongly written characters, variant forms of characters in ancient literatures, and combined-characters, and to support the computation for comparing glyphs. This is also a method of input and display Chinese characters in application. This method overcomes the drawbacks of glyph feature selection and computation for comparing glyphs in current glyph description methods, which can be applied in fields such as language and character researching, the glyph corpus creation and the analysis of errors in wrongly written characters in Chinese language teaching and international dissemination, the glyph evolution reseaching in ancient literature management, and historical and cultural studies. This method is also an effective glyph despiction tool for the retrival of uncommon Chinese characters in electronic publication on websites.

2. The Stroke-Segment-Mesh Glyph Depiction Method

2.1. The feature selection of basic units

To ensure the standardization and ambiguity-free of basic units for the description of Chinese character glyphs and to support the computation for comparing glyphs with high efficiency, the following three kinds of segments in a small rectangle are selected as basic unit features

to depict Chinese character glyphs, which are called stroke segments. As shown in Fig 1, every small rectangle contains 14 stroke segments with the definitions as below:

- Upper horizontal border line is divided into 2 segments by the midpoint, namely AE and EB, called horizontal stroke segments;
- Left vertical border line is divided into 2 segments by the midpoint, namely AH and HD, called vertical stroke segments;
- The lines drawn from every vertex to the midpoint of 2 opposite borders, namely AG, AF, BG, BH, CE, CH, DE, DF, and the lines drawn between two pairs of vertexes, namely AC and BD, totally $4 \times 2 + 2 = 10$ segments, called slant stroke segments.

2.2. The stroke-segment-mesh glyph depiction of Chinese characters

$N \times N$ small rectangles of same size containing stroke segments form a matrix of stroke segment mesh, which is called stroke segment mesh plane and used for the depiction of Chinese character glyphs as shown in Fig 2. The lower horizontal border line of each small rectangle in the plane belongs to the small rectangle below and the lower horizontal border line of the small rectangle in the lowest row is excluded in the depiction of glyph skeletons. The right vertical border line of each small rectangle belongs to the small rectangle next to the rectangle itself and the right vertical border line of the small rectangle in the furthest right column is excluded, too. Then, there are totally $14N^2$ stroke segments in the stroke segment mesh plane and these stroke segments form the set

$$G, G = \{S_{i,j}^k \mid 0 \leq i, j < N, 0 \leq k < 14\} \quad (1)$$

where $S_{i,j}^k$ represents a stroke segment; the subscripts, i and j , are the row and column serial number of the small rectangle that the stroke segment belongs to in the stroke segment mesh

plane; N is the row or column number of small rectangles contained in the stroke segment mesh plane; the superscript k is the serial number of stroke segments with different directions in the small rectangles: 0 and 1 representing two kinds of horizontal stroke segments, 2 and 3 representing two kinds of vertical stroke segments, and 4 to 13 representing 10 kinds of stroke segments with different slope directions (illustration see Fig 1).

Fig 1 The stroke segment distribution in the small rectangle

There are two states, effective and ineffective, for every stroke segment. If the segment is stroked, it is effective; if not, it is ineffective.

The definition of stroke-segment-mesh glyph of Chinese character is the set of the effective stroke segments which constitute the character in the stroke segment mesh plane. The set is H ,

$$H \subseteq G, H = \{S_{i,j}^k \mid 0 \leq i, j < N, 0 \leq k < 14, \text{State}(S_{i,j}^k) = 1\} \quad (2)$$

$$\text{State}(S_{i,j}^k) = \begin{cases} 1 & S_{i,j}^k \text{ is the stroke segment stroked} \\ 0 & \text{else} \end{cases}$$

where $\text{State}(S_{i,j}^k)$ is the state function of a stroke segment. If $S_{i,j}^k$ is a stroke segment stroked effectively, the value of the function is 1; else, it is 0. The set of blue effective stroke segments is the depiction of the stroke segment in mesher of the combined character “招财进宝” as shown in Fig 2.

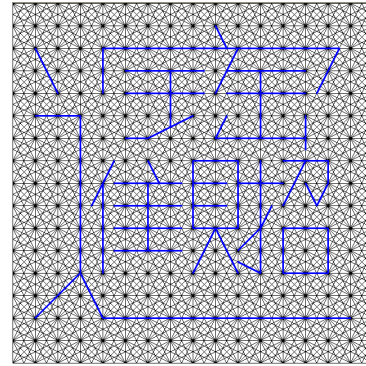


Fig 2 The stroke segment mesh plane and the character glyph in stroke segment mesh

3. The Classification and Extraction of Strokes Based on Stroke Segment Mesh of Glyphs

In order to describe features of Chinese character glyphs on different hierarchies and support the computation for comparing glyphs of whole or part, it is necessary to state the effective combination method of basic strokes and compound strokes constituted by stroke segment basic units, namely, the classified definitions of basic strokes and compound strokes, which are illustrated in detail as below.

3.1. The classification of basic strokes

There are 8 different types in the direction of all stroke segments. For the convenience of description, the direction of horizontal is marked as A; vertical, B; left top to right bottom diagonal, C; right top to left bottom diagonal, D; left top to right bottom level oblique line, E; right top to left bottom level oblique line, F; left top to right bottom vertical oblique line, G; right top to left bottom vertical oblique line, H; as shown in Fig 3.

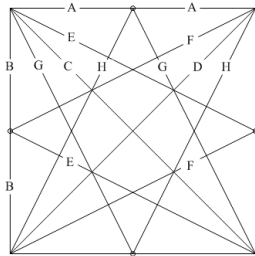


Fig 3 types of stroke segment directions

Definition 1 *simple strokes*: straight line strokes constituted by joint effective stroke segments in the same direction, including:

Horizontal simple stroke a: joint effective stroke segment sequence in direction A;

Vertical simple stroke b: joint effective stroke segment sequence in direction B;

Left oblique simple stroke f: joint effective stroke segment sequence in direction F;

Left oblique simple stroke d: joint effective stroke segment sequence in direction D;

Left oblique simple stroke h: joint effective stroke segment sequence in direction H;

Right oblique simple stroke e: joint effective stroke segment sequence in direction E;

Right oblique simple stroke g: joint effective stroke segment sequence in direction G;

Definition 2 *degree of endpoint*: the degree of an endpoint is the number of effective stroke segments that end in this endpoint. The number of effective stroke segments taking this endpoint as the top or left endpoint is the out degree of the endpoint and as the bottom or right endpoint is the in degree of the endpoint.

Definition 3 *basic strokes*:

Basic stroke Heng: horizontal a type simple strokes;

Basic stroke Shu: vertical b type simple strokes;

Basic stroke Pie: joint simple stroke sequence bhdf from the top down; any of the simple strokes in the sequence can be absent but there must be at least one type of stroke among h, d and f.

Basic stroke Na: joint simple stroke sequence bgce from the top down; any of the simple strokes in the sequence can be absent but there must be at least one type of stroke among g, c and e.

In addition, the basic strokes of Pie and Na types can be further divided into 4 sub-types according to the oblique degree of the effective stroke segments for the detailed depiction of the differences of the same type of strokes. The detailed classification and illustration see Table 1.

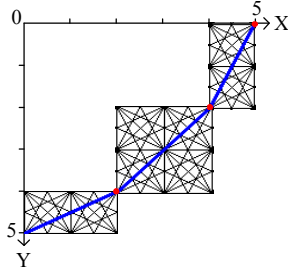
3.2. The extraction algorithm of basic strokes

The joint effective stroke segments in the stroke segment mesh of glyphs are combined into simple strokes; joint simple strokes are combined into basic strokes; then, the extraction of basic strokes is implemented and basic stroke set of stroke segment mesh glyph is achieved. The algorithm is illustrated in detail by the extraction of basic strokes of “Pie” as below.

The description of algorithm:

Step 1 Effective stroke segment series which are connected and have the same direction will be

combined as horizontal (a), vertical (b), left falling (f, d, h), right falling (e, c, g) straight line simple



strokes; Fig 4 shows three left falling simple strokes h, d, f, which are resulted from combining joint effective strokes H, D, F.

Fig 4 the sequence of joint simple strokes of h, d and f and the coordinate system

Step 2 Combining joint left falling (h, d, f) and right falling(g, c, e) simple strokes series to basic strokes “Pie” and “Na” (see Fig 5.)

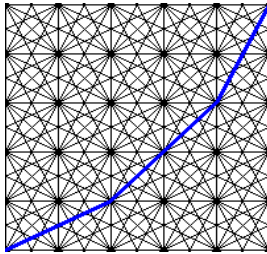


Fig 5 the basic stroke “Pie” combined by simple strokes of h, d and f types

The simple strokes combining algorithms based on graph-adjacent list.

2.1) Taking simple strokes as sides, and endpoints as vertexes, the graph-adjacent list is constructed, among which vertexes are sequenced upper-down to form the linked list, whereas for strokes which share the same upper endpoint will be sequenced left to right to form a linked list. Fig 6 is a graph-adjacent list based on simple strokes of Fig 4

2.2) Joint left falling (h, d, f) and right falling (g, c, e) simple strokes sequences are searched, which is based on graph-adjacent list, and combined into basic strokes as “Pie” and “Na”.

2.2.1) Find an unprocessed left falling or right falling simple strokes p;

2.2.2) Find an unprocessed simple strokes q, which taking lower endpoint of p as its upper endpoint, and have the same type with

p, and without any angles' increasing in direction compared to p.

2.2.3) Update adjacent list; combine q with p, update the lower endpoint of p as the lower point of q, p is marked as processed, the lower endpoint degree of q minus 1, then q is deleted from adjacent list.

2.2.4) goto to step 2.1), and keep going until no unprocessed left falling and right falling simple strokes are left in the adjacent list.

Step 3 For joint connected basic strokes “Shu” and “Pie”, or “Shu” and “Na”, combine them further as vertical left-falling “Pie” and vertical right-falling “Na” type basic stroke.

Based on adjacent list, searching basic strokes “Shu” and “Pie”, or “Shu” and “Na”, and then combine them.

3.1) Find an unprocessed basic stroke “Shu” p;

3.2) Find basic stroke q with type “Pie” or “Na”, which takes the lower endpoint of p as its upper endpoint.

3.3) Update adjacent list: combine q with p, update the lower endpoint of p as q's lower endpoint, p is marked as processed and q's upper endpoint degree minus 1, then q is deleted from adjacent list.

3.4) goto step 3.1) until no unprocessed basic strokes “Shu” is left in the adjacent list.

Step 4 Delete zero degree endpoints from vertex linked list of the adjacent list.

Step 5 Output basic strokes from vertex linked list of the adjacent list, resulting basic strokes list sequenced top-down based on endpoints.

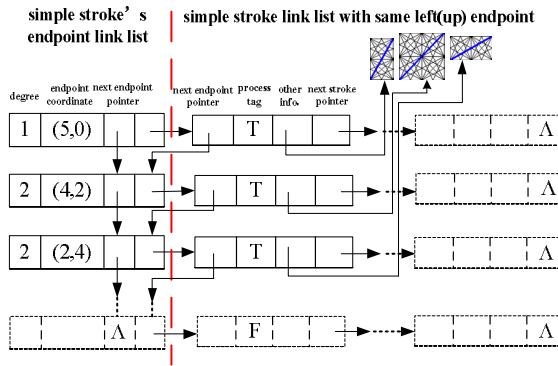


Fig 6 Graph-adjacent list structure based on simple strokes

3.3. Compound strokes classification

17 simple strokes' effective joint types were delineated, based on simple strokes types, for further classification, see Table 2.

3.4. Extraction Algorithm of Compound Strokes Based on Topological Sort

For compound strokes extraction, a series of jointed basic strokes could have various effective ways of combination. In accordance with habitual "upper-down, left-right" writing directions, the basic strokes can be designated as directed arc; taking the endpoint of strokes as vertex to construct mesh glyph's directed graph; taking the starting arc from the zero in-degree vertex as the starting stroke; using directed graph topological sort to realize the chinese writing principle "upper-down , left-right", so as to disambiguate the effective combining method of basic strokes, and achieving definitive compound strokes finally.

The description of algorithm:

```

While(zero in-degree vertex have a unprocessed
outgoing arc p){
  Switch(the stroke type of arc p){
    Case "Heng"
      if identify"Heng+ZuoShuGou/YouShuGou"
        then break
      if identify "Heng+ShuHengGou" then break
      if identify "HengShu" then break
      if identify "Heng+PieTi/PieGou" then break

```

```

      if identify "Heng+PieHengGou" then break
      if identify "Heng+PieNaGou/PieNaHengGou"
        then break
      if identify "HengPie" then break
      identify "Heng"
      break

```

Case "Shu"

```

      if identify "ShuGou/ShuTi" then break
      if identify "ShuHeng/ShuHengGou"
        then break
      identify "Shu"
      break

```

Case "Pie"

```

      if identify "PieTi/PieGou" then break
      if identify "PieHeng/PieHengGou"
        then break
      if identify "PieNa/PieNaGou/PieNaHengGou"
        then break
      identify "Pie"
      break

```

Case "Na"

```

      if identify "NaGou" then break
      if identify "NaHengGou" then break
      if identify "Na+ShuTi/NaShuGou" then break
      if identify "Na+PieTi/PieNaGou" then break
      identify "Na"
      break
    }
  }
}

```

Fig 7 is the DFA to identify compound strokes with "Heng" as the starting basic stroke.

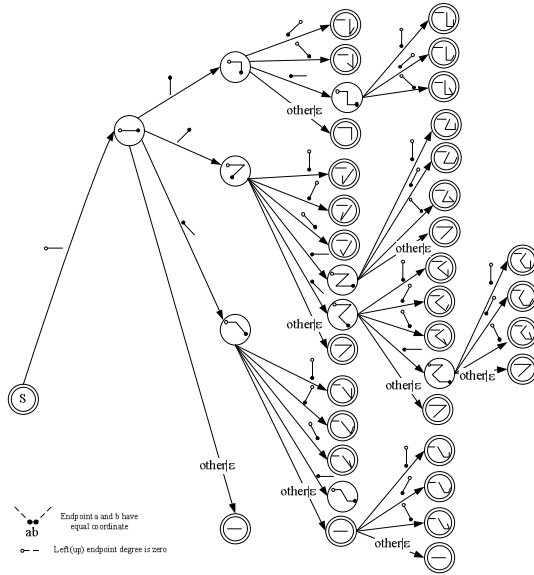


Fig 7 Identification of compound strokes' DFA with “Heng” as the starting basic stroke

4. the Computation of Stokes Structure Relation Based on Stokes-segment-mesh Glyph

Chinese character is of two-dimensional image, the relative positional relationship among strokes plays a key role in glyph depiction. The positional relationship among strokes is depicted by relation matrix^[19]. Taking A as a free stroke in glyph mesh, its relative position to the other strokes is depicted as: taking A's each minimum bounding rectangle side and its extending line as borderline, dissecting the mesh surface into a $3 \times 3 = 9$ -plane laid out as Chinese character “井”, which are centered around A, which represent 8 directions surrounding A, as shown in Fig 8. Each plane is numbered upper-down and left-right, such as: the first row, and first column represents the upper-left area of stroke A, the third row, and third column represents the down-right area of stroke A, forming a 3×3 boolean value matrix R_{AB} , the element $R_{AB}[i,j]$'s value represents whether the direction of the i row and j column around stroke A is jointed with stroke B's bounding rectangle area: if jointed, $R_{AB}[i,j]=1$, or $R_{AB}[i,j]=0$, the value of matrix R_{AB} represents the positional relationship between

strokes A and B. Matrix R_{AB} is named as the direction relation matrix of strokes A and B. Fig 8 represents the second point of character “心” (marked as area A), its bounding rectangle sides and extending lines dissect the mesh surface into 9 areas. Its value of direction relation matrix with other strokes (marked as area B, C, D) are as follows:

$$R_{AB} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, R_{AC} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, R_{AD} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

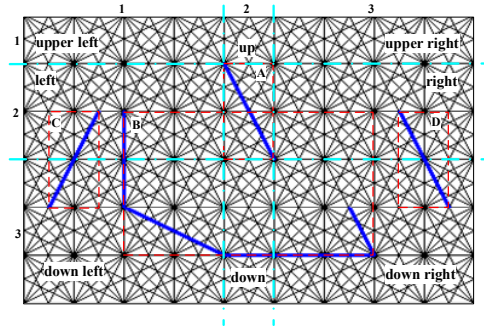


Fig 8 The strokes direction relation illustration of character “心”

Any two strokes' direction relation matrix in glyph can be computed in similar ways. The difference value between direction relation matrixes of relative strokes for two glyphs represents the difference degree of character structure. Moreover, the joint relation between strokes is depicted directly by intersection coordinates. The difference value of relative strokes intersection points represents the difference degree of intersection relation. Intersection relation is regarded as a special case of positional relation.

5. Experiment Result

The stroke-segment-mesh glyph depiction experiment is processed by self-developed hand-draw tools. Now, it has finished the stroke-segment-mesh glyph depiction for 6763 Chinese characters in GB2312 character set, nearly 1000 wrongly written characters in Chinese composition of overseas students, and part of variant forms of Chinese characters. Automatic

extraction of both basic and compound strokes is reached for glyph, and the accuracy degree of extraction reached 100%. The computation for strokes positional relationship can be processed automatically as well. The experiment proves, the $n=12$ strokes segment mesh can be used to depict the glyph of most Chinese characters, the $n=16$ mesh can be used to depict all the glyph for Chinese characters. These experiment results show that strokes segment mesh of glyph is a reliable, effective computational glyph depiction model with strong extension potential.

6. Conclusion

Strokes-segment-mesh glyph depiction model, with definitive limited direction segments—strokes as basic depiction unit, presents uniform features and appropriate granular degree, and overcomes existing problems of present Chinese characters depiction system. The problems are as follows:

1) Taking structure type, radicals, and strokes as depiction units, the granular degree of features are too big, the collection of feature values is fixed and closed, which makes it hard to represent open Chinese characters collection, and impossible to compare similarities and difference between some similar form characters.

2) For Bitmap Font or TrueType Curve-outlined Chinese Character^[20], their depiction basic unit granular degree is too small, is lack of regulatory rules, and includes many difference resulted from written tools used and aesthetic perspectives. These differences just mentioned are not the essential difference of Chinese character glyphs.

3) The Chinese character identification model adopted overall statistical features, which makes it hard to depict parts of the glyph while using logic analysis, and difficult to form corresponding relationship with human identified glyph features (such as strokes, radicals, structure type), and

difficult to be applied in other areas except whole character identification.

Considering the aspect of supporting glyph features automatic computation, owing to its standardized depiction basic units, ambiguity-free, and appropriate granular degree, and Chinese characters skeleton depicted only without nonessential calligraph features, all these measures which effectively limited the variant types of glyph features. As a result, the structure feature stability of glyph depiction is ensured, and makes it possible to accurately extract part glyph features from a whole glyph (such as basic strokes, compound strokes etc.), and correctly identify feature types. Thus, for one aspect, the complication degree of glyph feature computation is decreased, and for another, the computation result is more reliable, and it is easier to set up a corresponding relationship with present glyph depiction system.

Considering aspect of usability, for foreigners who has no knowledge of Chinese characters and typist who has a below-average literacy level, once they can correctly identify the glyph, they will be able to use the stroke-segment-mesh glyph depiction method to depict the Chinese characters that he has seen or thought of.

Therefore, this method is an input tool for both uncommon characters and wrongly written characters. It is a glyph depiction method which can extract glyph feature and structure relationship from various levels as to demand, and support glyph automatic comparative computation. On these basis, the further working purpose is to realize the automatic extraction of radicals, and glyph searching etc, so as to provide an effective tool for studies such as, creating Chinese characters glyph corpus, and providing services for studies and teaching of Chinese characters.

Table 1 Basic stroke classification and illustration






















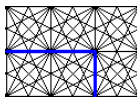

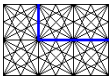

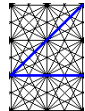
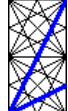
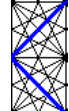

































stroke type	Heng	Shu	Pie				Na			
semantic	横	竖	平撇	斜撇	立撇	竖撇	平捺	斜捺	立捺	竖捺
type code	0	1	2				3			
Typical type of basic stroke sample										
										
										

Table 2 Compound strokes classification and illustration

stroke type	ZuoShuGou		HengShu	HengPie	ShuHeng	YouShuGou	PieHeng	PieTi	PieNa
semantic	竖钩		横折	横撇	竖折	竖提	撇折	撇折	撇点
type code	1g		01	02	10	1t	20	2t	23
typical type of compound d stroke sample									
stroke type	PieGou	PieNaGou	PieHengGou	PieNaHengGou	NaShuGou	NaPieGou	NaGou	NaHengGou	ShuHengGou
semantic	撇钩	右弯钩	右弯钩	右弯钩	左弯钩	左弯钩	斜钩	卧钩	竖弯钩
type code	2g	23g	20g	230g	31g	32g	3g	30g	10g
typical type of compound stroke sample									
									
									
									
									
									

References

[1] Wang,Ning.the Motivation of Construction of Chinese Characters and Modern Chinese Character Components Separation[J]. Language Planning(in Chinese),1997(3):4-9

[2] Chinese Character Coding Group, Shanghai Jiao Tong University. Chinese Character Information Dictionary[M]. Beijing: Science Press, 1988.

[3] National Language Committee. GF3001-1997 Chinese Character Component Standard of GB 13000.1

- Character Set for Information Processing. Beijing: Language & Culture Press, 1997.12
- [4] Ideographic Description, <http://www.unicode.org/versions/Unicode4.0.0/ch11.pdf>:307-309
- [5] <http://www.sinica.edu.tw/~cdp/> Taiwan's Academia Sinica Institute of Information Science, Literature processing laboratory
- [6] Sun Xing-Ming, Yin Jian-Ping, Chen Huo-Wang. On mathematical expression of a Chinese character. Journal of computer research and development(in Chinese), 2002, 39(6): 707-711
- [7] Zhang Wen-Yin, Sun Xing-Ming, Zeng Zhen-Bing. Automatic Generation of Mathematical Expression of Chinese Characters. Journal of computer research and development(in Chinese), 2004, 41(5):848-852
- [8] Richard Cook. A Specification for CDL(Character Description Language):an extract of [PhD Dissertation]. UC Berkeley, Dept.of Linguistics, 2003
- [9] <http://www.wenlin.com/cdl/> Wenlin Institute, University of California, Berkeley
- [10] Su Pei-Cheng. Essential of the modern Chinese characters (update). Peking University Press, 2001
- [11] Liang Yan-Min. Features of Chinese Character Components and Chinese Character Instruction to Foreign Students[J]. Language Teaching and Linguistic Studies(in Chinese), 2004(4):76-80
- [12] Ding, Xiao-Qing. Chinese Character Recognition: A Review[J]. ACTA ELECTRONICA SINICA, 2002, 30(9): 1364-1368.
- [13] Lin, Zhi-Qing & Guo, Jun. An Algorithm for the Recognition of Similar Chinese Characters[J]. Journal of Chinese Information Processing, 2002, 16(5): 44-48.
- [14] Liu Xia-Bi, Jia Yun-De. The Formation and Extraction of Chinese Character Sub-Stroke[J]. Chinese Journal of Computers, 2004, 27(3):389-395
- [15] Chen Liang-Yu, Zeng Zhen-Bing, Zhang Wen-Yin. Analysis and Recognition of the Structure of Chinese Characters[J]. Journal of Shanghai Institute & Electric Power, 2005, 21(1):63-65
- [16] <http://www.xiaoyaobi.com/> Beijing xiaoyaobi Pattern Recognition Workstations
- [17] Feng Zhi-Wei. Description of Chinese Character Structure by Context Free Grammar[J]. Linguistic Sciences, 2006, 5(3):14-23
- [18] Wang Kai-Shou, Wang Ying-Wei. The Relations Stable Principle of Chinese Characters Script[J]. Journal of Chinese Information Processing, 1996, 10(4):24-31
- [19] Du Shi-Hong, Wang Qiao etc. A Qualitative Description Model of Detailed Direction Relations[J]. Journal of Image and Graphics, 2004, 9(12):1496-1503
- [20] Wen Li-Xin, He Er-Gong etc. The Design for Chinese Character Pattern CAD System Based on the Chinese Character Generator[J]. Journal of Chinese Information Processing, 1996, 10(1):14-20