

TR-80-002

Chinese Character Analysis
Using Structural Approach.

by

Ching-Liang Lin

中研院資訊所圖書室



3 0330 03 00004 1

0004

書	考	參
借	外	不

國立台灣大學
電機工程學研究所

碩士學位論文

以結構分析法辨認中文字形

CHINESE CHARACTER ANALYSIS USING
STRUCTURAL APPROACH

林慶良 Ching-Liang Lin

經考試合格特此證明

碩士學位論文考試委員

張 雄
蔣 經

郭 德 剛

指導教授

孫心怡

所主任

中華民國六十九年六月 日

誌 謝

承指導教授陸心怡博士悉心指導，使本論文得以完成。陳諗教授的提供使用影像處理設備，使實驗部份得以順利進行。在此致上衷心的感謝。

中央研究院一年半來的資助，及中央研究院資訊所計算機中心工作人員，王苑華小姐，曾耀正先生，丁之侃先生的協助，在此一併致謝。

謹以本論文獻給我的父母、師長及所有關心我的朋友，沒有他（她）們的鼓勵與支持，本論文將無法完成。

摘 要

本文就結構觀點，提出一種中文字形之辨認方法。分析的過程如下：

首先找出五種基本結構元，再連接俱有均勻性質的結構元形成結構筆形。相對每一結構筆形用一特徵量以描述之。這些結構筆形即被用作建造中文字形之基本圖樣，而中文字形的編碼描述即根據各結構筆形與其在平面之位置關係完成。

本文亦設計了整個軟體程式，並用600個印刷體存與900個手寫體字作實驗。我們共收集了近120個基本特徵碼。

本文對所提出方法所作的估量及其實驗結果都加以討論，並同時附之於文後。

ABSTRACT

In this thesis, we use structural approach for Chinese Character recognition. The analysis procedure is the following:

Five types of structural elements are extracted first, and structural strokes are formed by connecting structural elements. A local feature vector which describes the shape of a structural stroke is associated with each structural stroke. These structural strokes are encoded and used as fundamental building blocks of Chinese characters. A Chinese character is encoded based on structural strokes and their special relations.

The overall program design for this method is included. An experiment using 600 printed and 900 hand-written characters are performed. We have collected about 120 fundamental strokes.

The performance and experimental results of the proposed method are presented and discussed.

目 錄

第一章	緒論	一
第二章	方法描述	四
2-1	文字模型的建立	四
2-2	結構元分析	七
2-3	結構筆形分析	十一
2-4	編碼	十六
2-5	辨認	二十四
第三章	程式設計	二十五
3-1	系統組織	二十五
3-2	資料結構	二十七
3-3	演算程式	三十
第四章	實驗與結果討論	三十六
4-1	格式說明	四十
4-2	討論	五十八
第五章	結論	六十
參考文獻	六十一
附錄		
(一)局部特徵碼轉換表		六十二
(二)字形及其串碼表		六十四

第一章 緒論 (Introduction)

1-1 問題陳述 (Statement of problem)

人們藉助電腦快速的計算能力，競相的把它應用在各個領域內，獲得相當大的成效。

在語言、文字處理方面的應用，已經有許多的研究者致力於此一領域的研究；範圍涵蓋有：阿拉伯數字（0～9）的辨認，各地區本國字母的辨認，語言與語言間的轉譯，……等等，嚐試將人類的智慧溶入電腦中，為人們做最大的服務。

目前，郵件的處理，已使用文字辨認的技巧，用電腦代替人工做檢信、分信的工作。一只輕巧簡便的語言翻譯器，也加入為人類從事語言翻譯的服務；諸如種種，不勝枚舉，均為此一研究成果的展證。

比較中文與外國文字，我們獲得：

(一)文字構造複雜：

(a)字根數目：

阿拉伯數字只有十個字母（0～9），英文共廿六個字母（A～Z），日文共五〇個字母。中文單就字根（字根為字之不可分割部份）就將近數百個。

(b)字根間的關係：

每一英文單字（Vocabulary）由每一個字母，依左而右排列。中文字却是每個字根在二度平面之排列組合，它們彼此間除左右關係外，還有上、下及包含的關係。

由於字根間之組成複雜，使字與字間的區別繁瑣，減慢了中文字處理的速度。

(二)同一字形存在有不同的楷體，楷體間書寫及幾何間距離大，增加統一分類的困難。

如何設計一種兼具考慮速度及文字書寫變化自動的辨認方法，誠是值得研究的問題。

1-2 過去的研究 (Previous Work)

在近廿年內，有許多的研究者從事中文字形辨認的研究，主要可以歸納為

①印刷字體的辨認及②手寫字體的辨認兩類。

(一)印刷字體：第一篇着手研究字形辨認的論文於 1966 年由 R. CASEY 與 G. NAGY [1] 提出，在論文中，作者提出了二階段過程的辨認方法。第一階段依字形之相似性（如字根等），把字加以詳分，然後；在詳內，字與字間以值別的模子附屬於各字，加以細分。字的辨認則依各模子去找找正確的字。S. YAMAMOTO, etc [2] 以層次觀念將此加以推廣。

Y. NAKANO, ect [3] 用投影堆積法（projection profile），把每一個字形在直向或橫向作投影堆積，再利用富氏轉換求出每一個字的富氏轉換特徵頻率。

以字形結構為着眼的有：

PAUL. P. WANG [4] [5] 用拓模觀念，分類及編碼每一中文字。W. W. STALLING [6] 將筆形及筆形間的交點關係，轉換成一圖形（Graph），再定出字根與字根的關係而編碼每一字。（此為第一位把圖形理論（Graph Theory）應用在中文字形辨認的人）

由單一模板之印刷體，近來，進而從事多種模板的研究。

(二)手寫字體：

此一方面的研究，所提出解決的方法，主要有①筆形分析法（Stroke-oriented）②統計法（Statistical）。

(a)筆形分析法：

此一研究方法，先把字形細化（Thinning）找出特徵點（Feature point）或每一筆形（Stroke），再定義相似性量度（Similarity measure），或用編碼方法辨認每一個字。

H. OGAWA, etc [7] 按筆形在平面上之拓模分配特徵，分類，量度每一中文字。MAKOTO, YOSHIDA [8] 先分析字形，後用合成方法以辨認中文字。T. AUGI and H. NAGAHASHI [9] 提出一種以方塊圖（block pattern）為主的中文字編碼方法。

(b)統計法：

Y. L. MA [10] 設計一前向馬可夫動態計劃程式（Forward Markovian Dynamical programming），用統計機率方法判定每一個字。概率之量度是依照每一筆形出現在每一格間的頻率來決定。

(一)建立中文字之書寫模型。

(二)就五種結構元分析中文字，不同於邊界追蹤 (boundary tracing) 及細化 (thinning) 法。

(三)以單筆、直筆及叉筆編碼每一中文字。

各章內容包括：

第二章：將辨認之每一過程詳細的描述。

第三章：介紹程式的設計與演算方法。

第四章：討論及說明實驗所觀察到的結果。

於附錄中可以看到部份字形及其所編成的串碼。

整個分析過程為：對輸入之中文字形，依上至下，左而右處理，並找出各結構元。由各結構元之相對關係位置，串接成各結構筆形；此即為構造中文字形之基本圖樣。根據各結構筆形間之關係，我們用二層次的量對中文字加以描述，並依各串碼長度分類各字形，最後依據所分析的串碼成功的辨認輸入的字形。

2-1 文字模型的建立：

2-1-1：模型描述：

以中國書法的永字八法為基石，本文設計如下的中文字形模型；在此一模型中，我們定出字形的選字法則如下：

〔名詞註釋〕

1 圖(一)所指的點(丶)，橫(一)，斜上(/)，撇(丿)，右短撇(丿)，直(|)，鈎(亅)，捺(㇇)為八種基本筆形，或簡稱為基本筆形。

2 筆形俱有上、下投影重疊關係者，稱為筆形投交。

3 一筆形被分割成兩個或兩個以上的線性部份稱為筆形被分割；一筆形可以任意被分割；其線性且不投交部份稱為結構筆形 (Structural Stroke)

造形法則：

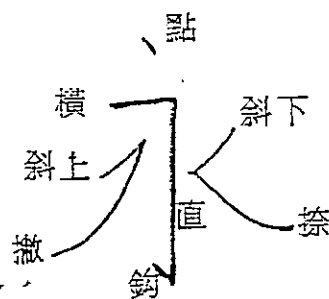
1 任一筆形不投交

2 所有字形為筆形在平面上連續運作，排列的結果。

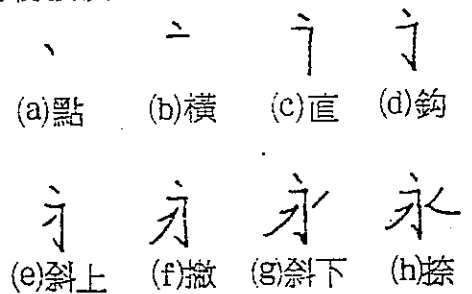
3 每一字形俱有一定的運作方式：如圖(二)表示“永”字的書寫方式。

4 每一筆形自然構成一勻稱部份。

5 字形的區別只能從筆形，字形結構，及筆形間的銜接關係判別。



圖(一) 八種基本筆形



圖(二) (a - h) 說明“永”字之書寫方式

2-1-2：數學特徵分類：

從書寫的方式而論，斜上、撇、及斜下顯然有極大的不同，然而；由幾何的特徵着眼，它們都具有相同的傾斜性，在表(一)中即依照其傾斜性再把八種筆形分類。在編碼時，我們將是以此一分類方法分類。

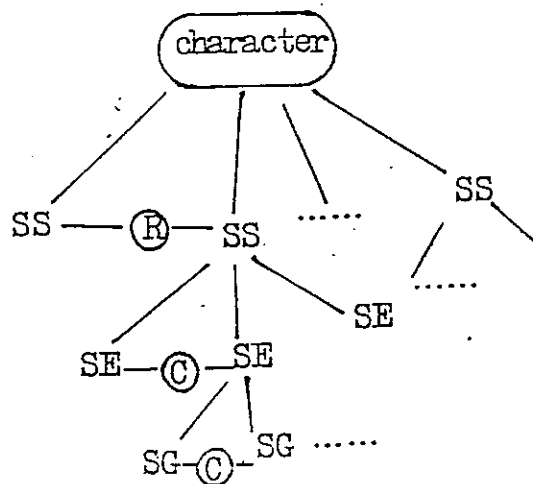
名稱	基本筆形	形狀	傾斜性
點	點	、	
右斜上	1 斜上 2 撇 3 斜下 (右短撇)	、 、 、	斜率 = 正
右斜下	1 捺 2 鈎	、 、	斜率 = 負
橫	橫	—	斜率 = 0
直	直		斜率 = ∞ (直向部份 > 橫向部份)
叉接點		兩筆形交接處	

表(一) 基本筆形分類

※斜率之度量將在 (2 - 4 - 2) 中詳述。

2 - 1 - 3 : 結構元分類

在本文中，我們以結構段 (S G) (Structural Segment) 為中文字處理的最小單位，結構元 (S E) (Structural Element) 由一個或數個結構段之集合，數個結構元構成一結構筆形 (Structural Stroke) ; 圖(三)說明中文字結構段—結構元—結構筆形之層次關係。



圖(四) 字形之描述

(R): SS-SS 間的關係；(上、下、左、右等關係)

(C): SE-SE 及 SG-SG 間之關係，(串接 (Concatenation) 關係)

SS, SE 及 SG 之定義在以後各節中會有進一步的定義。

根據書寫的自然趨向，我們歸納出下列五種結構元。

2 向上延伸

圖(四)：示一向上延伸之結構之，特徵為：

- (1) $P_{U1} \geq P_{U2} + 1, P_{D1} \geq P_{D2}$
- (2) $SLOP = P_1 - P_3$
- (3) $||P_{D1} - P_{D2}|| \leq SLOP, \text{ 且 } ||P_{U1} - P_{U2}|| \leq SLOP$
- (4) $SLOP \leq 4$

3 向下延伸

圖(五)示一向下延伸結構元，特徵為：

- (1) $P_{U1} \leq P_{U2}, P_{D2} \geq P_{D1} + 1$
- (2) $SLOP = P_4 - P_2$
- (3) $||P_{D1} - P_{D2}|| \leq SLOP, \text{ 且 } ||P_{U1} - P_{U2}|| \leq SLOP$
- (4) $SLOP \leq 4$

4 向右發散

圖(六)示一向右發散結構元，特徵為：

- (1) $P_{U1} \geq P_{U2} + 1, \text{ 且 } P_{D1} \leq P_{D2} - 1$
- (2) $||P_{U1} - P_{U2}|| \leq 4 \text{ 且 } ||P_{D1} - P_{D2}|| \leq 4$

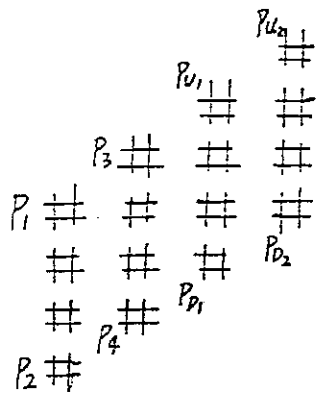
5 向右收縮

圖(七)示一向右收縮結構元，特徵為

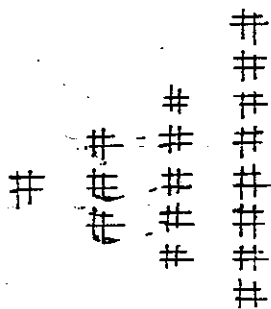
- (1) $P_{U2} \geq P_{U1} \text{ 且 } P_{D1} \geq P_{D2} + 1 \text{ 或 } P_{U2} - 1 \leq P_{U1} \text{ 且 } P_{D2} \leq P_{D1}$
- (2) $||P_{U1} - P_{U2}|| \leq 4, \text{ 且 } ||P_{D1} - P_{D2}|| \leq 4$

6 向右水平延伸

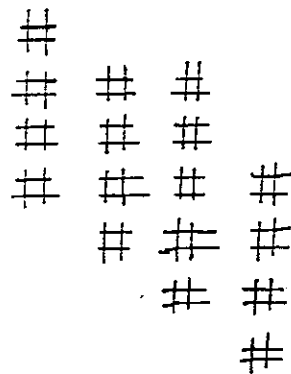
- (1) $P_{U1} = P_{U2} \text{ 且 } P_{D1} = P_{D2}$ (如圖(八))



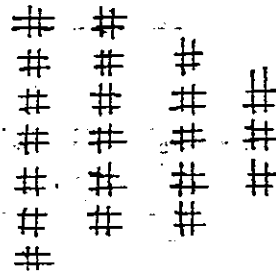
圖(四)向上延伸



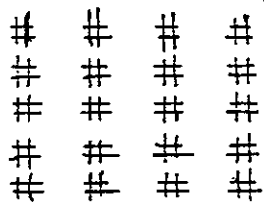
圖(六)向右發散



圖(五)向下延伸



圖(七)向右收縮



圖(八)向右水平延伸

2-2: 結構元分析

定義一:

(一)結構段 (SG): 由一組直向連續的黑點組成的集合 (以黑點代替有字部份, 白點代表空白); 結構段以三組量描述, 記為 $SG(U, V, W)$; U : 直向坐標, V : 橫向最上點坐標, W : 橫向最下點坐標。

(二)結構元 (SE): 由一組橫向相鄰的結構段組成的集合 (在 2 - 1 節中已將其分類), 結構元以六組量描述, 記為 $SE (SG_1 (U, V, W), SG_2 (U_2, V_2, W_2))$

SG_1, SG_2 分別為最左邊及最右邊之結構段, 在以下諸節中

我們把 SG_1 稱為始段， SG_2 稱為尾段。

下例說明此定義的內涵。

2	3	4	
2 #	#		$SG_1 = (2, 2, 6)$
3 #	#	#	$SG_2 = (3, 2, 5)$
4 #	#	#	$SG_3 = (4, 3, 4)$
5 #	#	SG_3	$SE = U \quad SG_1$
6 #	SG_2		$= (SG_1, SG_3)$
SG_1			$= (2, 2, 6, 4, 3, 4)$
	SE		

定義二：我們定義兩相鄰結構段的上差 (D_u) 及下差 (D_d) 為：

設 $S_1 = (U_1, V_1, W_1)$, $S_2 = (U_2, V_2, W_2)$

上差 (D_u) = $||V_1 - V_2||$

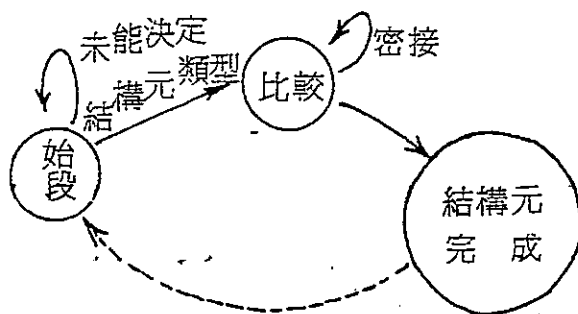
下差 (D_d) = $||W_1 - W_2||$

長度 ($||S_1||$) = $||V_1 - W_1||$

2-2-1：概念描述：

圖(九)之遷度圖示 (transition diagram)，說明結構元萃取的過程：

由 SG_1, SG_2 決是結構元的始段，再作繼續提取，直至尾段完成。



圖(九) 概念遷渡圖

2-2-2：分析

引理：設 $SG_1(U_1, V_1, W_1)$ 為第一結構段， $SG_N(U_N, V_N, W_N)$ 為同 SG_1 相鄰之右結構段 (即 $U_N = U_1 + 1$)

則 SG_1 與 SG_N 之接鄰情形只有：

- (1) $SG_N = \phi$; 即 SG_1 沒有相鄰段
- (2) $|(SG_N)| = 1$; 即 SG_1 只有一相鄰段
- (3) SG_N 為已組成之結構元的尾段

(4) $SGN \geq 2$; 即 SG_1 有多於一個的相鄰段。

說明：

(1) $SGN = \phi$; SG_1 為存在之結構元的邊界，不是一結構元的始段。如圖(+))

(2) $\#(SGN) = 1$;

① $\|SGN\| > \|SG_1\|$ 且 $Du > 4$ 或 $Dd > 4$

SG_1 為結構元之尾段或許多結構元的尾段的共同鄰段。或是一段始。

分別如下：

(1) SG_1 是結構元的始段；當有一結構元之尾段同 SG_1 相鄰而且存在另一結構元之始段同 SE 之尾段相鄰。如圖(+) - i

(2) SG_1 同一個或多個結構元之尾段相鄰， SG_1 為此結構元之共同尾段。

如圖(+) - ii

② $\|SG_1\| > \|SGN\|$ 且 $Du >$ 或 $Dd > 4$

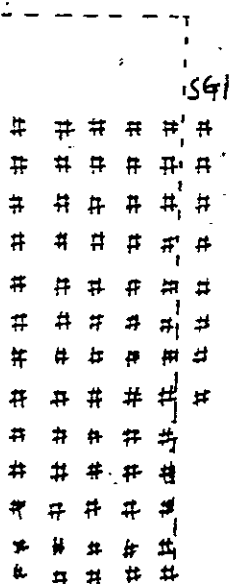
SG_1 同另一結構元之尾段相鄰；將 SG_1 預置成結構元之始段，其坐標值 $= (U_1, U_N, U_N)$ 如圖(+))

③ $\|SG_1\|, \|SGN\|$ 滿足條件； SG_1 是一結構段的始段。如圖(+))

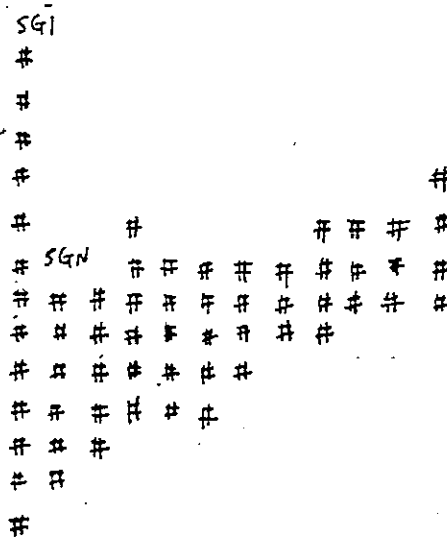
(3) $\#(SGN) \geq 2$; SG_1 為結構元之尾段。切劃

SG 與 SGN 構成結構元之始段。如圖(+))

(4) SGN 為已完成的結構元 SE 的尾段； SGN 為另一結構元之始段， SG_1 為結構元之尾段。預置 SE 之尾段。如圖(+))

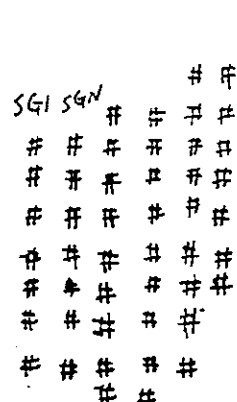


圖(+) $SGN = \phi$



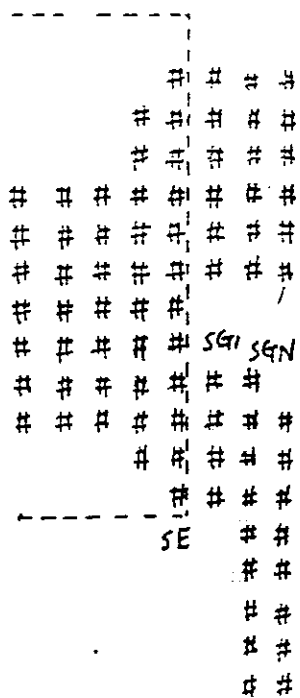
圖(ii) $\|SG_1\| > \|SGN\|$

且 $Du > 4$ 或 $Dd > 4$

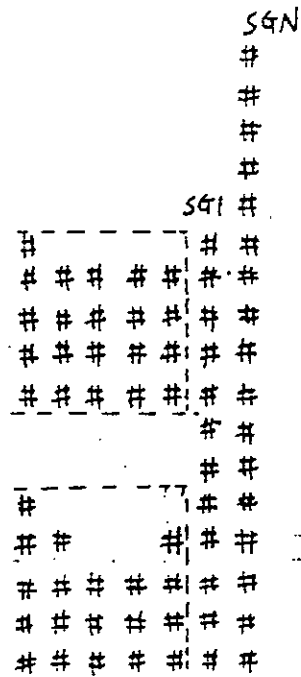


圖(iii) SG_1, SGN 滿足

結構元條件

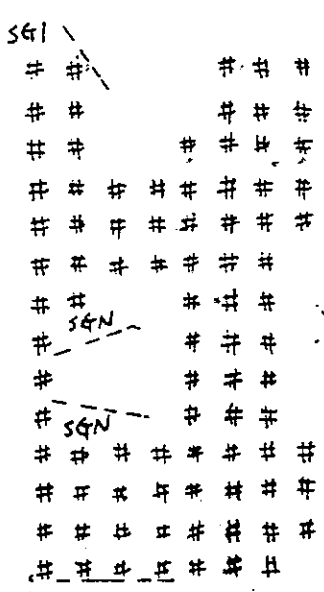


(i)

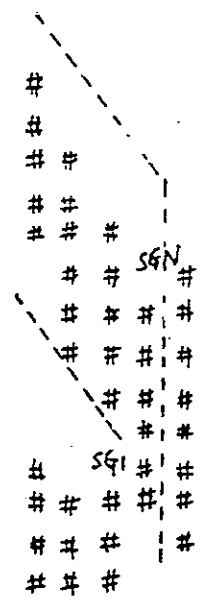


(ii)

圖(十) $||SGN|| > ||SG1||$ 且 $Du > 4$ 或 $Dd > 4$



圖(十一) $(SGN) \geq 2$



圖(十二) SGN 為已找出結構元之尾段

2-3 : 結構筆形分析

2-3-1 性質

按(2-1-1)之定義一結構筆形爲一具有線性均勻部份，而且不投交者。在此節中，我們將再其詳細定義成：

定義(-)：

(一)結構筆形(ST)：由1組或兩組以上的結構元，依其線性程度串接而成，我們用六種量描述；設

$$ST = \bigcup_{i=1}^l SE_i, \quad \text{且} \quad SE_i = \bigcup_{j=1}^N SG_j$$

$$SE_i = (U_i, V_i, W_i, U_N, V_N, W_N)$$

其中 $(U_i, V_i, W_i) \in SG_j$

則 $ST = (U_1, V_1, W_1, U_l, V_l, W_l)$

其中 $(U_1, V_1, W_1) \in SE_1$ 且 (U_1, V_1, W_1) 爲 SE_1 之始段。

$(U_l, V_l, W_l) \in SE_l$ 且 (U_l, V_l, W_l) 爲 SE_l 之尾段。

(二)兩結構元 $SE_1 = (U_1, V_1, W_1, U_N, V_N, W_N)$, $SE_2 = (U_1, V_1, W_1, U_N, V_N, W_N)$

如果 $(V_N, W_N) \cap (V_1, W_1) \neq \phi$

則稱此兩結構元水平投合 (horizontal projection overlap)

* (a, b) 表示由 a, b 構成之區間。

引理：兩相鄰(水平投合)之結構元，構成一結構筆形的充分且必要條件爲：

- (1) 任意相鄰之結構元存在一個且僅有一個的相鄰結構元。
- (2) 兩相鄰之結構元不造成一很大的轉點(我們謂長度有2倍以上關係者謂之)

說明：

- (1) 因爲結構筆形爲一不投交筆形，如果存在有兩個或兩個以上的相鄰結構元，即成爲一投交筆形，故得證。
- (2) 結構筆形俱有最大的均勻性質，故長度差異太大應分成兩個結構筆形。

2-3-2 : 建立連接圖形 (Adjacency Graph)

定義(-)：設 $G = (V, E, A)$ 之集合爲一連接圖形其中；

V (vertex) 意爲所有節所構成的集合 = $\{ST_i\}$

E (edge) 節與節中存在的連接關係： $R(ST_i, ST_j)$

A (Attribute) 節所附有的屬性； $A(ST) = \{(X, Y, Z)\}$

並且 $X = ST_i$ 之種類

$Y = \sum (E)_L$; 同節有左方連接關係之總數。

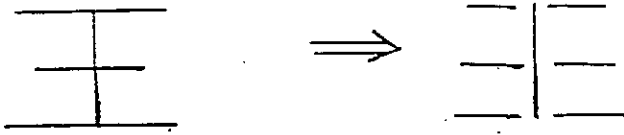
$Z = \sum (E)_R$; 同節有右方連接關係之總數。

按照此一定義，我們很容易的利用前節所述的方法，把中文字轉換成一接圖形。

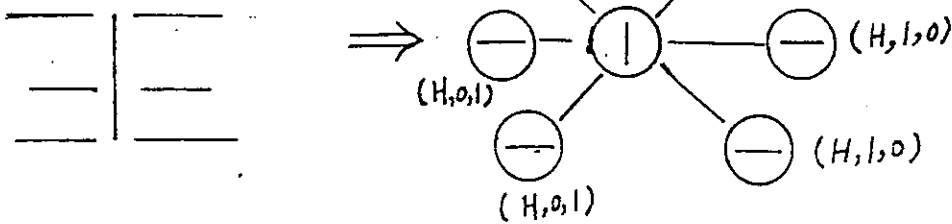
下例：說明其部份過程

例(1)：建立“王”字之連接圖：

一：分析結構筆形（圖中用隔間分開者即為結構筆形）

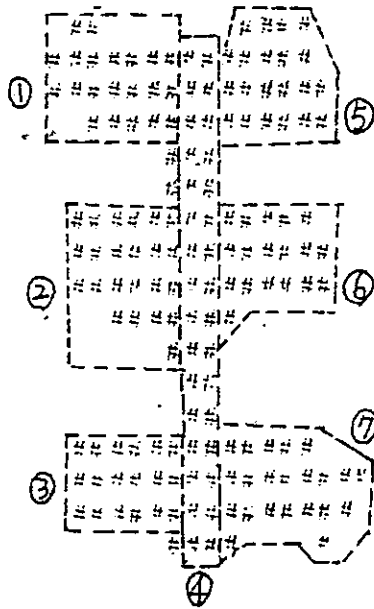


二：轉換成連接圖形：



V：直筆 H：橫筆

下二實例為“王”及“木”二字經過程式執行後之結果；其中用虛線框出的四方區域及標示的代號，顯示各結構筆形（格式之說明請參閱（4-1）節）。



STRUCTURAL ELEMENTS :

FORMAT IS : <ID. , COORDINATE , COORDINATE , TYPE>

ID.	COORDINATE 1	COORDINATE 2	COORDINATE 3	TYPE
1	23	16	17	2
2	24	21	23	3
3	24	28	30	2
4	25	15	18	6
5	26	21	24	2
6	27	16	18	3
7	30	16	31	3
8	32	16	18	2
9	32	21	24	4
10	32	28	31	6
11	34	15	18	2
12	34	21	23	2
13	34	28	30	3
14	37	29	31	6

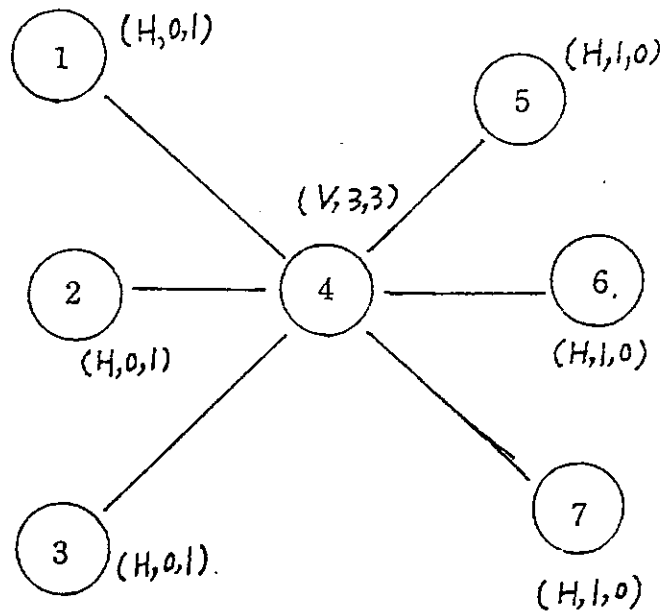
STRUCTURAL STROKES :

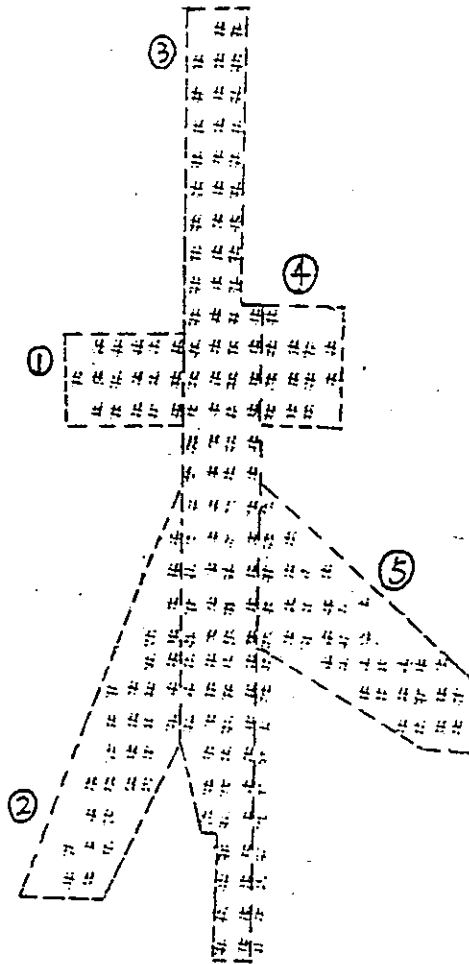
FORMAT IS <ID. , ELEMENT>

ID.	ELEMENT
1	4 5
2	5 3
3	2 3
4	7 8
5	8 11
6	9 12
7	13 14

Adjacency matrix

1	0	0	0	4	0	0	0	0
2	0	0	0	4	0	0	0	0
3	0	0	0	4	0	0	0	0
4	1	2	3	0	5	6	7	0
5	0	0	0	4	0	0	0	0
6	0	0	0	4	0	0	0	0
7	0	0	0	4	0	0	0	0





STRUCTURAL ELEMENTS ::

FORMAT IS: <ID., COORDINATE, COORDINATE, TYPE>

ID.	COORDINATE 1	COORDINATE 2	COORDINATE 3	COORDINATE 4	COORDINATE 5	COORDINATE 6	TYPE
1	13	13	13	14	12	14	4
2	13	28	29	14	26	20	2
3	15	12	14	18	12	14	3
4	15	23	28	16	23	26	6
5	17	21	26	18	18	24	2
6	19	3	24	20	2	27	4
7	21	11	31	22	11	31	3
8	23	11	14	24	12	14	6
9	23	18	21	24	18	21	2
10	25	12	14	26	12	13	6
11	25	19	21	25	19	22	1
12	27	20	22	28	20	23	1
13	29	21	23	30	22	24	1
14	21	22	24	32	22	24	3

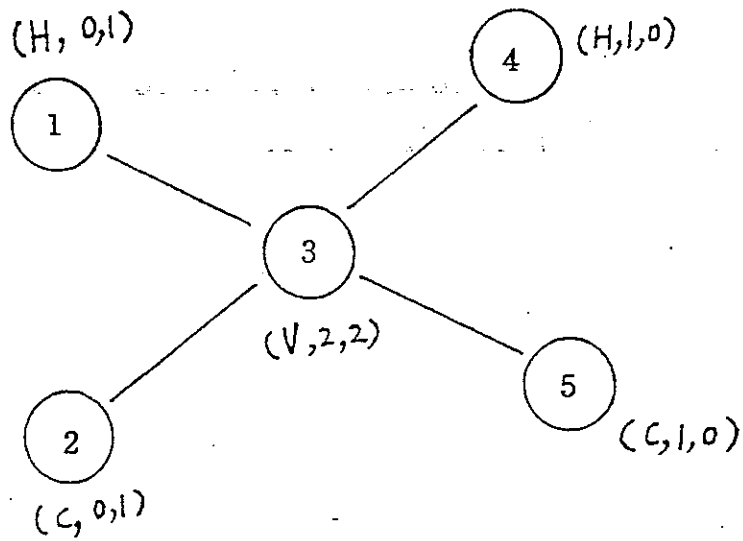
STRUCTURAL STROKES :

FORMAT IS <ID., ELEMENT>

ID.	ELEMENT 1	ELEMENT 2	ELEMENT 3	ELEMENT 4	ELEMENT 5	ELEMENT 6
1<	1	3	0	0	0	0
2<	2	4	5	0	0	0
3<	6	7	0	0	0	0
4<	8	10	0	0	0	0
5<	9	11	12	13	14	14

Adjacency matrix

1	0	0	3	0	0	0
2	0	0	3	0	0	0
3	1	2	0	4	5	0
4	0	0	3	0	0	0
5	0	0	3	0	0	0



c : 斜 筆

2-4 編碼

在此一論文中，我們以兩種方式對每一個中文字加以層次的描述。以局部特徵碼描述局部之特徵，再由整體碼對整個字作整體的描述。

2-4-1；“直”筆形與“叉”筆形的特徵：

(a)直筆形

(1)幾何性質：斜筆 $=\infty$ ，兩邊長度 \gg 水平展開寬度。

(2)相連性：有較多的連接筆形。

(b)叉筆

(1)幾何性質：由兩筆形交叉而成，幾何性質不明顯。

(2)相連性：存在有大於或等於2的左或右之連接筆。

例： 中間虛線部份即為一叉筆，其左，右各有兩斜筆。

2-4-2：局部特徵向量 (local feature vector)

每一局部特徵碼以一特徵向量 (F_1, F_2, \dots, F_7) 表示。

[n, m]：每一節之屬性，所有相連性均對中心結構筆形而言。

F_1 ：左相連之橫筆形 $ST(t, n, m)$ 其中 $n \neq 0$

F_3 ：右相連之橫筆形 $ST(t, n, m)$ 其中 $m \neq 0$

F_2 ：橫過中心結構筆形之橫筆。

F_4 ：左方相連之斜上筆 $ST(t, n, m)$

F_5 ：左方相連斜下筆 $ST(t, n, m)$

F_6 ：右方相連之斜上筆 $ST(t, n, m)$

F_7 ：右方相連之斜下筆 $ST(t, n, m)$

各特徵向量度為：

(1)橫筆： (F_2)

設 ST 為中心結構筆形， ST_1, ST_2 分別為其左相連及右相連的結構筆形；則 ST_1, ST_2 構成一橫筆的條件為

①水平投合

②傾斜性 $=0$ (或 $|\theta| < 15^\circ$)

在判定 ST_1, ST_2 為橫筆，可利用的方法有平方誤差，直線密合 (least square error line fitting) 法，在此論文用，我們用一種質心法：

設 $ST_l = \{SE_1, SE_2, \dots, SE_N\}$

$ST_r = \{SE_1, SE_2, \dots, SE_N\}$

$SE_j = (U_j, V_j, W_j, R_j, S_j, T_j)$

左方結構筆形之中心 = $R^1(X^1, Y^1)$

$$X^1 \equiv \left(\sum_i \frac{(u_i^1 + r_i^1)(u_i^1 - r_i^1)}{2} \right) / \sum_j (u_j^1 - r_j^1)$$

$$Y^1 \equiv \left(\sum_j (v_j^1 + w_j^1 + S_j^1 + t_j^1) / 4(u_j^1 - r_j^1) \right) / \sum_j (u_j^1 - v_j^1)$$

$R^1(X^1, Y^1), R^r(X^r, Y^r)$ 之直線

$$Y = m(X - X^1) + Y^1; m = \frac{Y^1 - Y^r}{X^1 - X^r}$$

ST_i, ST_i 橫成一橫筆的條件為：

$$(t_n^1 - m(Y_n^1 - X^1) - Y^1)(S_n^1 - m(Y_n^1 - X^1) - Y^1) < 0$$

且 $|m| < 0.267 (\sim 15^\circ)$

(2) 特徵量 $f_1, f_3, f_4, f_5, f_6, f_7$ 之量度。

設 $ST = [SE_1, SE_2, \dots, SE_r]$

$$m \equiv \left[\frac{1}{2}(S_N + t_N) - \frac{1}{2}(V_1 + W_1) \right] / (r_N - U_1)$$

① 如果 $|m| < 0.267 (\sim 15^\circ)$

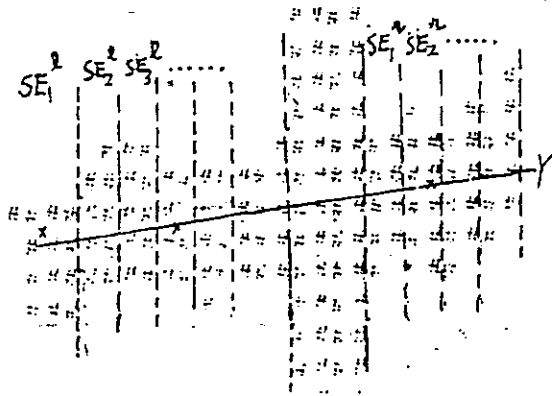
則 f_1 或 f_3 加 1；即為一橫筆

② 如果 $m > 0.267$

則 f_4 或 f_6 加 1；即為一上斜筆

③ 如果 $m < -0.267$

則 f_5 或 f_7 加 1；即為一下斜筆



$$SE_j = (U_j, V_j, W_j, R_j, T_j, S_j)$$

圖(六) F_2 量之度量

2-4-3 局部碼之次序：

(2-4-2) 節中局部碼並沒有加入位置的意義；然而在中文字形中各局部特徵在平面上的分配情形，關係文字的整體描述，於此節中，我們對每一局部碼附加以位置描述，並且依位置關係之優先次序將各局部碼作更完整的排列。

A：局部特徵碼之坐標表示：

每一局部特徵碼附以坐標值以制定其在平面之位置；以 (X, Y_1, Y_2) 表之， X, Y_1, Y_2 之定義如下：

(1) 局部碼只有單一結構筆形構成：

$$\text{設 } ST = (SE_1, SE_2, \dots, SE_N)$$

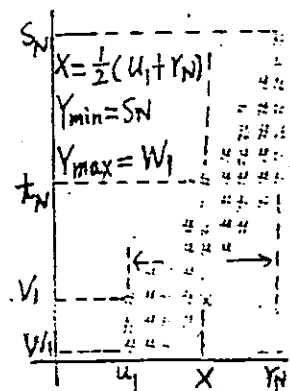
$$SE_j = (u_j, v_j, w_j, r_j, s_j, t_j)$$

$$X = \frac{1}{2} (u_1 + r_N)$$

$$Y_2 = \text{MAX}(w_2, t_N)$$

$$Y_1 = \text{MIN}(v_1, s_N)$$

圖(2)說明此一意義。



圖(2)單筆結構筆形之位置定向。

(2) 局部碼由二個或二個以上結構筆形構成：以中心之結構筆形之位置定之。

設 中心結構筆形為 ST

$$ST = (SE_1, SE_2, \dots, SE_N)$$

$$SE_j = (U_j, V_j, W_j, r_j, s_j, t_j)$$

$$X = U_1 \quad X = r_N$$

$$Y_1 = V_1 \quad \text{或} \quad Y_1 = s_N$$

$$Y_2 = W_1 \quad Y_2 = t_N$$

B：關係：

設兩局部碼 $L(X_1^1, Y_1^1, Y_2^1)$ 及 $L(X_1^2, Y_1^2, Y_2^2)$ ，在平面間的關係有上、下、左、右關係。 $R_L(l_1, l_2)$ 表示 l_2 在 l_1 之上方，當其滿足下列諸條件時。

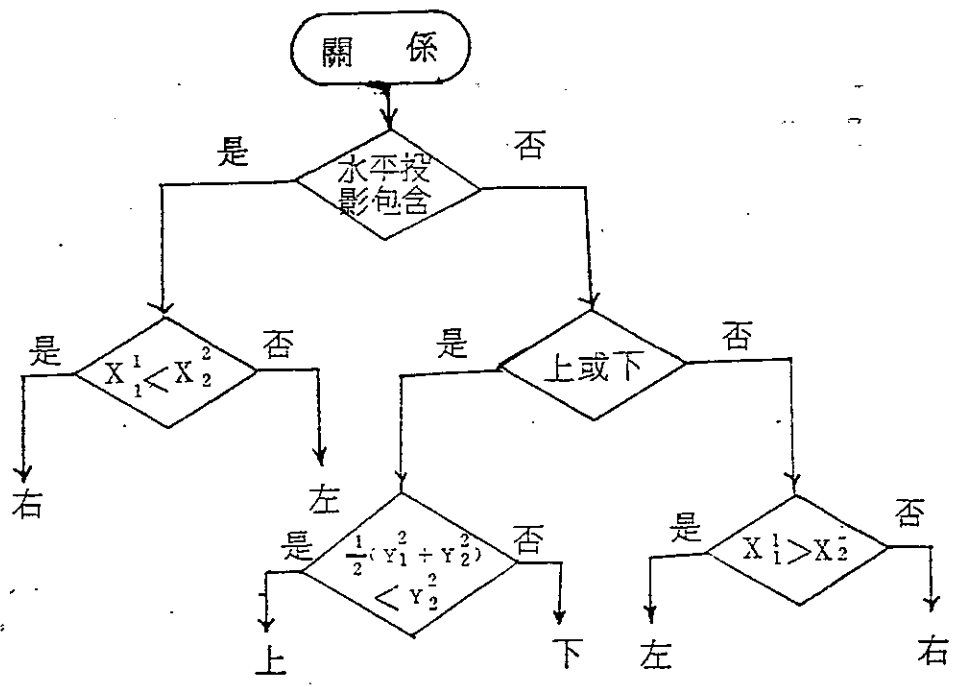
(1) 右方 $R_{\text{右}}(l_1, l_2) = R_{\text{左}}(l_2, l_1)$

(a) if $(Y_1^1, Y_2^1) \subset (Y_1^2, Y_2^2)$ 或 $(Y_1^2, Y_2^2) \subset (Y_1^1, Y_2^1)$
 且 $X_1^1 < X_1^2$
 或 (b) $(Y_1^1, Y_2^1) \not\subset (Y_1^2, Y_2^2)$ $[\subseteq: \subset \text{ 或 } \supset]$
 且 $X_1^1 < X_1^2$

(2) 左方: $R_{\text{右}}(l_1, l_2) = R_{\text{左}}(l_2, l_1)$
 (3) 上方: $R_{\text{上}}(l_1, l_2) = R_{\text{下}}(l_2, l_1)$ (下之定義同上)

(a) $(Y_1^1, Y_2^1) \not\subset (Y_1^2, Y_2^2)$
 (b) $\frac{1}{2}(Y_1^2 + Y_2^2) < Y_1^1$

整個過程之決定步驟如圖(六)所示



圖(六)局部上碼之關係決定圖

依局部碼之關係，整體局之局部碼構成的優先次序為上，左、右及下
 如：

- 局部圖樣：局部特徵向量
- 副 $(0, 0, 0, 0, 0, 0, 0) \dots\dots(1)$
 - ⌈ $(0, 0, 2, 0, 0, 0, 0) \dots\dots(2)$
 - E $(0, 0, 3, 0, 0, 0, 0) \dots\dots(3)$
 - 王 $(0, 3, 0, 0, 0, 0, 0) \dots\dots(4)$

$$\exists (3, 0, 0, 0, 0, 0, 0) \dots\dots(5)$$

$$\sqsupset (2, 0, 0, 0, 0, 0, 0) \dots\dots(6)$$

$$\downarrow (0, 0, 0, 0, 0, 0, 0) \dots\dots(7)$$

$$\downarrow (0, 0, 0, 0, 1, 0, 0) \dots\dots(8)$$

按各局部特徵間的關係，可以列成下表：

設表中之值：1 = R 上，-1 = R 下，2 = R 左，-2 = R 右之關係

係		R (l_1, l_2)								
		l_2	1	2	3	4	5	6	7	8
l_1	1		0	-1	-1	-1	-1	-1	-1	-2
	2		1	0	-1	-1	-1	-2	-2	-2
	3		1	1	0	-2	-2	1	1	1
	4		1	1	2	0	-2	1	1	1
	5		1	1	2	2	0	1	1	1
	6		1	2	-1	-1	-1	0	-2	-2
	7		1	2	-1	-1	-1	2	0	-2
	8		2	2	-1	-1	-1	2	2	0

1, 2, 3, 4, 5, 6, 7, 8, 9 為局部碼之代號如上所標記

由關係矩陣轉換成串碼，其方法如：

步驟一：在各行中找尋其中一行各矩陣元量都是 1 或是 2（亦即僅為上與左之關係）

步驟二：去除步驟一之所在行列值。

步驟三：如果還有行或列未取盡，重覆，步驟一；否則轉換完畢。

依此一方法上表之整個轉換過程即如

① 第一行一之值恰好剛為 1，2 故第一次找到對應碼 = 1

② 去除第一行及第一列得矩陣為

$R(l_1, l_2)$	$l_1 \backslash l_2$	2	3	4	5	6	7	8
2		0	1	-1	-1	-2	-2	-2
3		1	0	-2	-2	1	1	1
4		1	2	0	-2	1	1	1
5		1	2	2	0	1	1	1
6		2	-1	-1	-1	0	-2	-2
7		2	-1	-1	-1	2	0	-2
8		2	-1	-1	-1	2	2	0

③在②之矩陣內重覆①之選取，得到對應碼 = 2

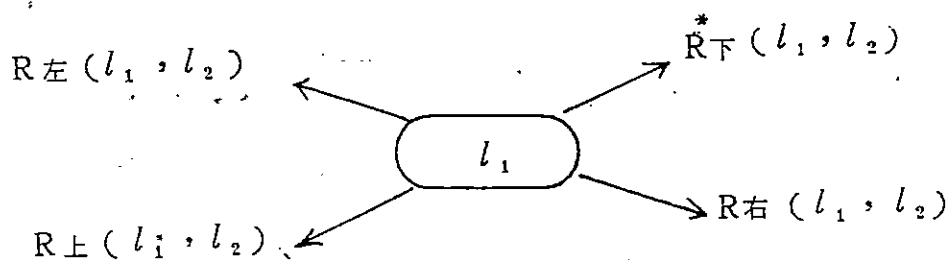
④依次繼續，所找到的對應數碼為：6, 7, 8, 3, 4, 5

⑤找出之串碼為 1, 2, 6, 7, 8, 3, 4, 5 對應 " - c o l j e m e" "

(c) 簡化之二元關係樹 (Binary relation tree) 編碼：

由(b)節所述之關係矩陣編號，在程式執行速度上將花費較多的時間，於本文中我們另外設計一種二元樹的編碼方法。利用此一方法去近似(b)中之方法，由下列子我們知道亦可達成的。

二元關係樹之概念，就是把原來之二元關係加以特徵的推廣，而成上、下、右、左四種關係以別於一般之大於與小於關係，圖(4)即說明各節 (Node) 與其關係間的流向：



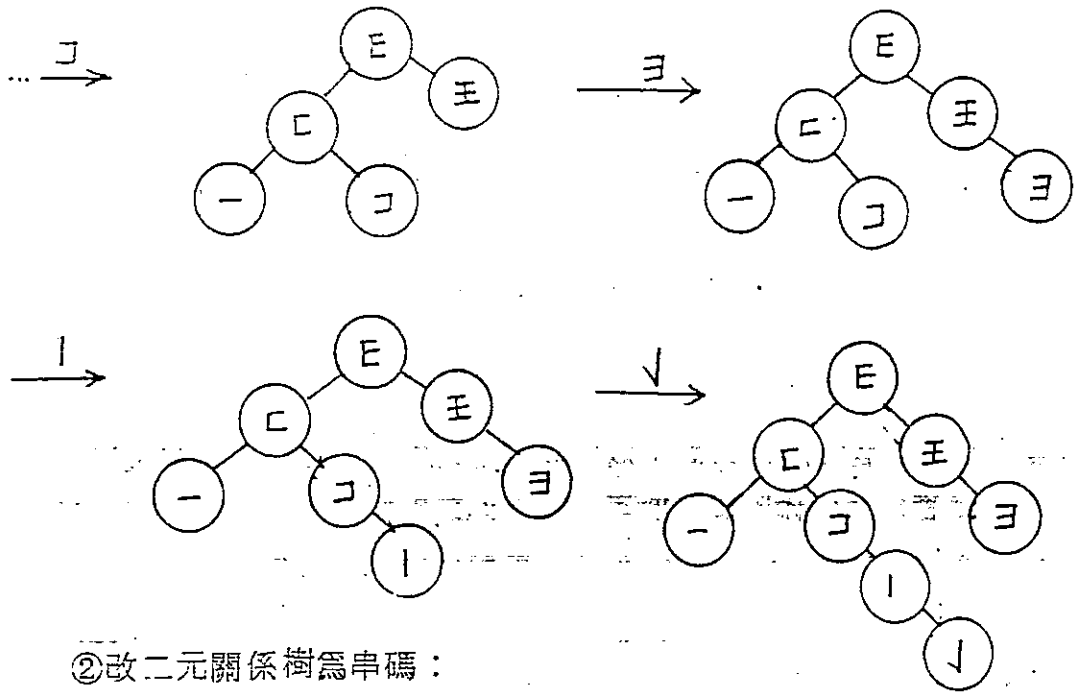
* $R_{\text{下}}(l_1, l_2)$ 意為 l_2 在 l_1 的下方。

圖(4)二元關係樹之關係流向

二元關係樹之長成乃依照如下方式：

輸入：欲加入之結構筆形 (局部碼)；在此內稱為局部碼。

步一：訪問根，若根 (root) 為空 (empty)，則加入根，否則



②改二元關係樹為串碼：

依(c)節所述之方法，很快的即可改為“-CコIJEHO”

此與用關係矩陣所做者完全相同。

例：只字之編碼：

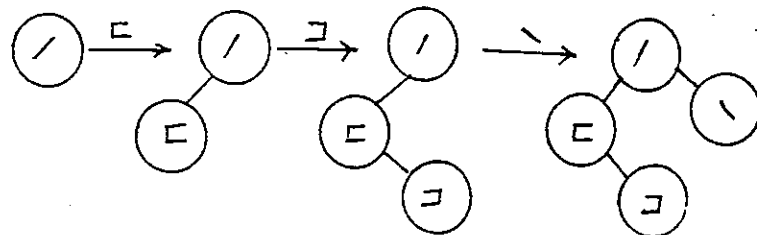
1 局部碼（按取出之順序）“/ C コ \”

2 二元樹關係：

(a)關係矩陣

	/	C	コ	\
/	0	1	1	-2
C	-1	0	-2	-1
コ	-1	2	0	-1
\	2	1	1	0

(b)二元關係樹



(c) 串碼

□ □ / \

3. 比較同(b)節之關係矩陣編碼

利用(b)節之方法把(a)之關係矩陣編碼，即為

□ □ / \

此與用二元樹編碼者相同。

2 - 5 : 辨認 :

按整體碼的長度，將中文字加以分類，屬於同一類者置於同一檔內；辨認的過程是首先計算整體碼的長度，再搜尋其相對的檔案資料。

步驟一：〔推〕 Push：將整體碼推入堆 (Stack) 中，並記錄指標。

步驟二：〔開檔〕：依照指標開啟所對應的檔。

步驟三：〔密合〕：讀取檔中的資料，依次同堆內者比較，如果搜尋不着，為外來字，則去除 (reject)

第三章 程式設計 (Program Design)

整個系統之程式流程及資料結構將在本章內作深入的描述。

3-1: 系統組態 (System Configuration)

圖(廿)為本系統之整體流程，在此一系統中主要分為三部份(1)預先處理(Pre-Processing) (2)編碼(coding) (3)辨認(Recognition)，每一程式均以模塊(Modularity)設計。

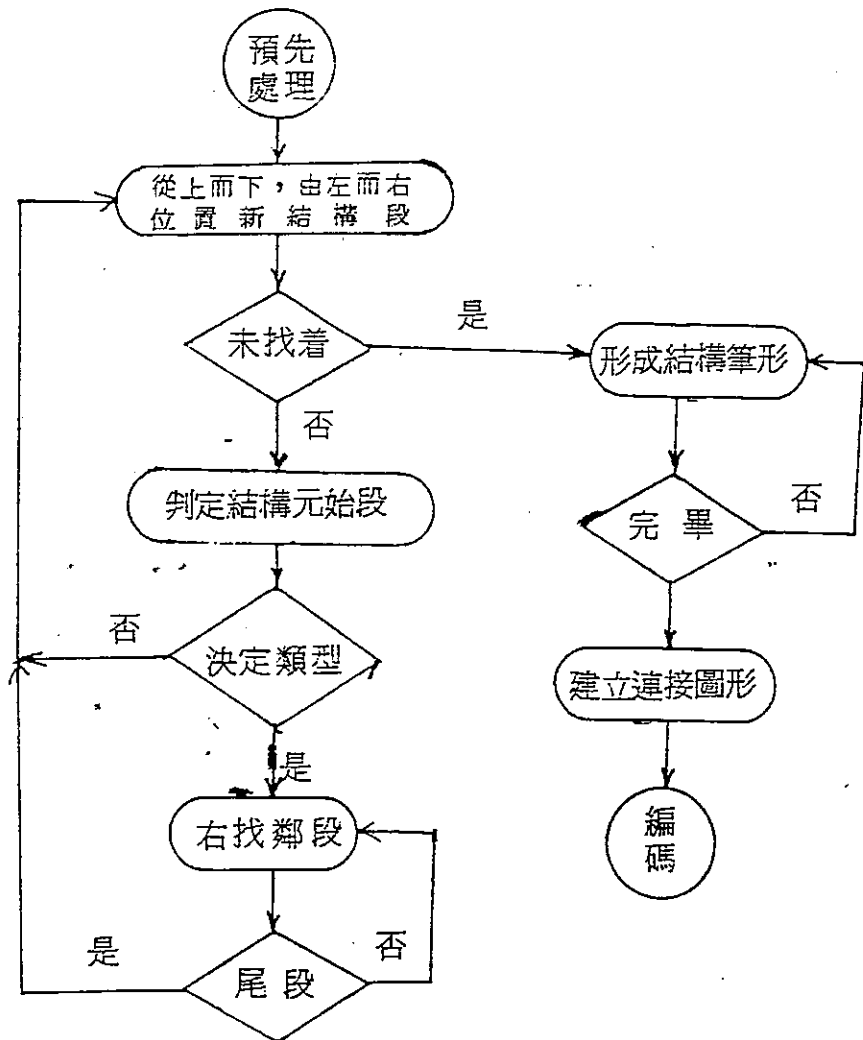
預先處理：

由輸入之中文字形，取出結構元，結構筆形，及建立結構筆形間的連接圖形(Adjacency graph)

編碼：

把連接圖形，依其左右連接性質，轉換成局部特徵碼。按照各局部特徵碼的平面排列關係再轉換成整體碼。

辨認：在字庫中找尋所儲存的字。



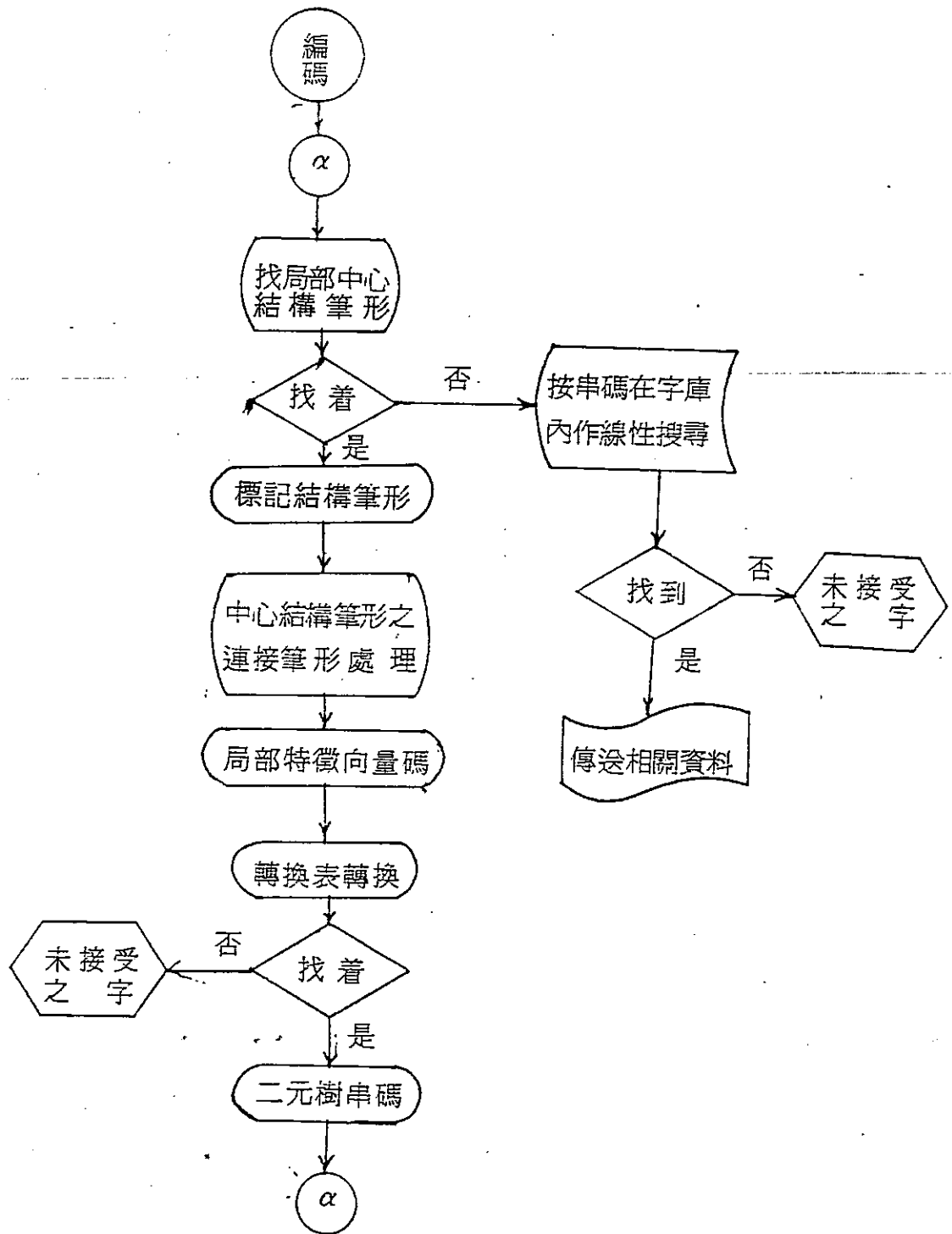


圖 (廿) 系統流程圖

3 - 2 : 資料結構 :

結構元，結構筆形，連接圖形於程式中全以表 (Table) 之形式存起，於此節中我們將其一一的陳述。

(1) IBUF(80,10) : 結構元表

存錄結構元之種類，位置，連接組合同一結構筆形之結構元，並且連接尾段同在一向值的結構元族。格式 (format) 如表(三)

type	INIX	INUY	INDY	INFX	INFUY	INFDY	PTRY	STPT	x
⋮	⋮				⋮	⋮			

表(三) 結構元表之格式

說明：type : 結構元種類

INIX : 起始結構段直向值

INUY : 起始結構段橫向上值

INDY : 起始結構段橫向下值

INFX : 尾結構段直向值

INFUY : 尾結構段橫向上值

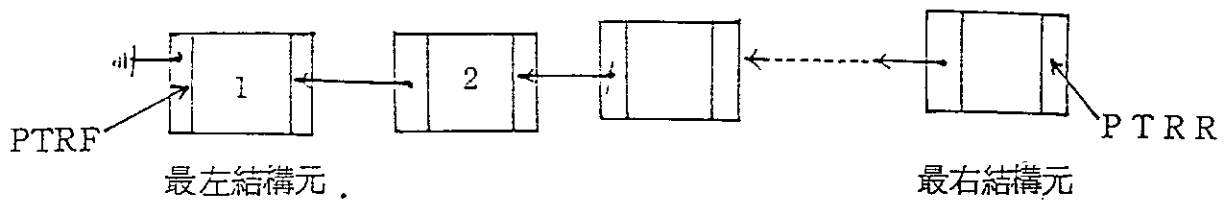
INFDY : 尾結構段橫向下值

PTRY : 尾段同一直向值之相連接指標

STPT : 構成一結構筆形之結構之接連指標

(2) ITABL(40,2): 結構筆形表 :

存錄結構筆形之最左結構之最右結構元之指標每結構元間以連鎖 (link list) 相連。



格式：

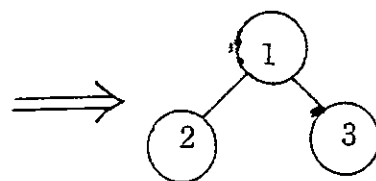
PTRF	PTRR
⋮	⋮

PTRF : 指向 IBUF 中, 為最左結構元指標

PTRR : 指向 IBUF 中, 為最右結構之指標

(3) IADJ(40, 40) : 連接圖形以矩陣表格表之, 每一值存入相連接的結構筆形代碼。IADJ(m, n)=n

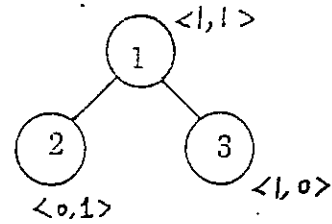
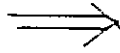
	1	2	3
1	0	2	3
2	1	0	0
3	1	0	0



(4) LR(40, 2) : 結構筆形屬性表 (Attribute Table)

存錄結構筆形之左右連接結構筆形個數起始值 = -1

	L	R
1	1	1
2	1	1
3	-1	-1
	⋮	⋮



(5) JTREE(30,8) : 局部特徵碼關係表 (local feature table)
有錄局部特徵碼及其關係格式為

Code	LS	RS	FA	X	YU	YD	CN
⋮	⋮						

Code : 局部特徵轉換碼 (請參閱附錄一)

LS : 左兒子 (Leftson)

RS : 右兒子 (rightson)

FA : 父親 (father)

X : 直向坐標

YU : 橫向最上坐標值

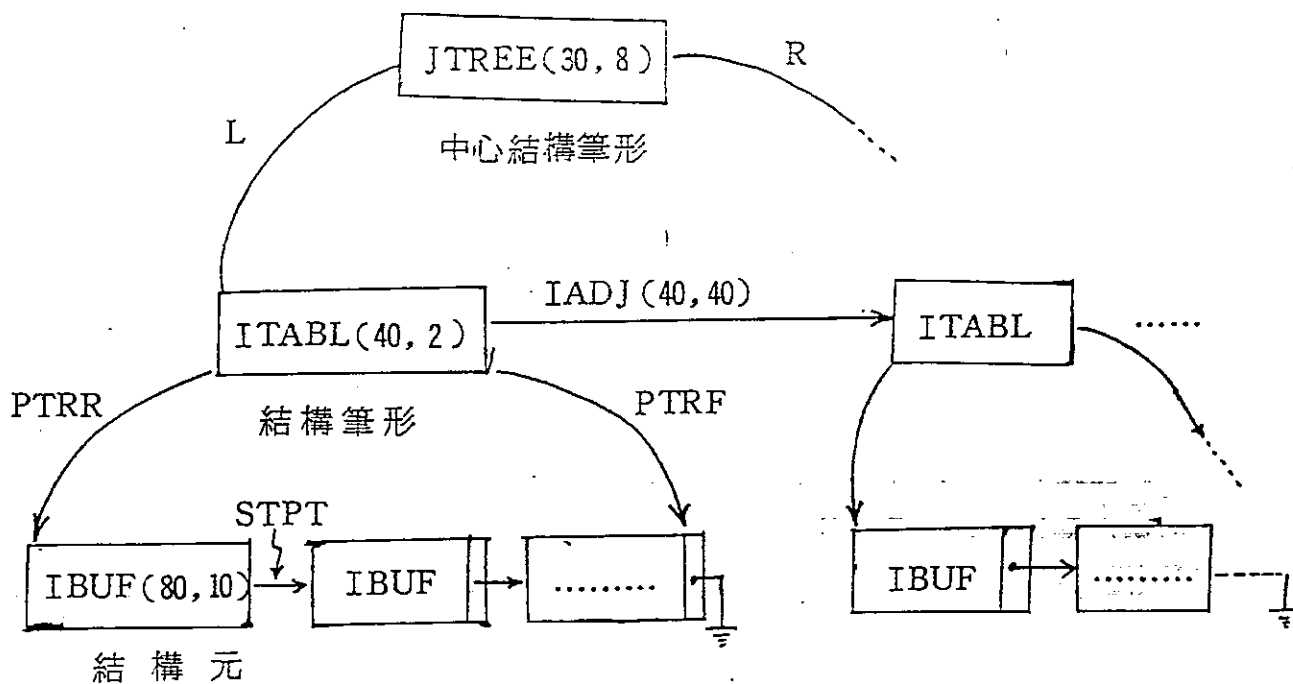
YD : 橫向最下坐標值

CN : 中心結構筆形代碼

(6) 資料檔 (File)

存錄局部特徵對應碼及整個碼

我們可以把此系統之局部特徵碼的資料結構，用層次關係圖(表)之成



圖(三) 資料結構之連鎖關係圖

3-3 : 演算程式 :

整體之演算程式，將在此節中詳細說明，我們依設計之各模樸 (modularity) ... 加以描述。

部份編碼之程式，已在第二章中說及，於此節內我們未重覆書寫。

找結構段 (LCAT)

輸入：起始點之坐標 (IX, IY)

輸出：結構段的位置坐標 (X, YU, YD)

步驟一：[掃瞄] 由 (IX, IY) 向下搜尋，找到結構段時做步驟(三)，否則做步驟(二)

$IY = IY + 1$

步驟二：[測量試邊界] 從右鄰行繼續掃瞄 (由上而下)，如果超過圖樣邊界，停止操作。

否則執行步驟(-)

$IX=IX+1, IY=1$

步驟三：〔找到〕：送回所找着的結構段坐標(X, YU, YD)

決定始段(ESTM)

輸入：前兩結構段的坐標($SG1, SG2$)

輸出：所計算完之結果。

狀況一：右結構段已是結構元之尾段。

狀況二： $SG1$ 及 $SG2$ 滿足結構元構成的條件。

狀況三： $|1+SG111| < |1+SG211|$ ，不滿足結構元構成的條件。

狀況四： $|1+SG111| > |1+SG211|$ ，不滿足結構元構成的條件。

步驟一：〔量度〕：比較 $SG1$ 及 $SG2$ 的長度，並執行相對的狀況。

狀況一：〔重寫〕：把尾段是 $SG2$ 的結構元重置，並將所有與此一 $SG1$ 有尾段相鄰的結構元，改寫其尾段坐標值。

狀況二：〔滿足〕

狀況三：〔重寫〕：如果存在一結構元的始段同 $SG1$ 有相同的相鄰結構元之尾段，把 $SG1$ 當作下之結構元的始段，否則將所有同 $SG1$ 相鄰之結構元尾段改至 $SG1$ 之直向坐標。

狀況四：〔修飾〕以 $SG1$ 之直向坐標為結構元之直向坐標值， $SG2$ 橫向坐標為結構元之橫向坐標值。

向上延伸結構元 (N2)

輸入：結構段之坐標及斜率($SG1, X, INY, IFY$)

輸出：下一鄰段的坐標值。

步驟一：〔預先測試〕如果($IX+1, IFY+1$) 點是黑點，停止並送回沒有密合的狀況重返程式中。

步驟二：〔下點坐標〕：由($IX+1, IFY$) 點由下往上掃瞄至($IX+1, INY$) 點如果沒找到黑點，送回沒有密合的狀況，否則記錄下點之坐標($IX+1, IDY$)

步驟三：〔上點坐標〕：由步驟二所得之下點坐標向上掃瞄，直找到白

點爲止，記錄其上點坐標 ($1X+1, 1UY$)

步驟四：〔長度控制〕：設算 IUY 及 INY 之差值，($K=INUY-INY$) 如果 K 大於或 K 小於斜率，則爲不密接狀況，否則繼續步驟一。

* 向下，等狀況同向上延伸大同小異，於此文中不詳加敘述。

(形成結構筆形) (MERGE) :

輸入：結構元表 (IBUF)

輸出：結構筆形表 (IQ 爲結構筆形之編號指示)

步驟一：〔起始〕設 $I = 1, IQ = 1$

步驟二：〔線性找尋〕：在結構元表內找出未形成結構筆形的結構元，並以其爲結構筆形之起始結構元，如果沒有找着，工作完成。
 $LS=I, LS=LS+1, ITABL=I$

步驟三：〔找鄰結構元〕計算結構元相鄰之右方結構筆形數目，如果大於 2 或等於 0，均不合條件。

步驟四：〔返鎖找尋〕：由步驟三找到的結構元，再找尋，同此結構元相鄰之左結構元，並計其數目，如果數目 ≥ 2 ，不合條件，並執行步驟二，重新另一結構筆形之開始。

步驟五：〔長度比較〕：比較兩結構元的長度，不得差異有二倍以上之關係，否則是結構筆形之終止，執行步驟二，重新另一結構筆形之開始。

連結結構筆形間連接圖形 (GENG):

輸入：結構筆形表

輸出：結構筆形連接圖形及屬性表 (LR)

步驟一：〔起始〕設

$I = 1, LR = -1$ ，取出第一結構筆形

步驟二：〔線性搜尋〕在結構筆形表內依次找各結構筆形是此一結構筆形的相鄰筆形。

步驟三：〔預增〕取出下一結構筆形，如果結構筆形已用盡，則停止操作。

$I=I+1$ ，執行步驟二。

找局部碼中心結構筆形 (NODE) :

輸入：連接圖表及其屬性表

輸出：局部特徵碼之中心結構筆形

步驟一：〔起始〕設 $I=1$

步驟二：〔找尋〕找出尚未標示過 (Mark old) 之結構筆形及

其屬性 $K=I+1$ ， $[n,m]=LR$

步驟三：〔決定〕

如果 $\langle n,m \rangle = \langle 0,0 \rangle$ 且尚未標示，則找到只有一結構筆形的局部特徵碼。否則

如果 $\langle n,m \rangle = \langle 0,1 \rangle$ 則找出 $ST1$ 右鄰近結構筆形 ($ST2$) 及其屬性 $\langle 1,t \rangle = \langle 1,1 \rangle$ 比較 $ST1$ 及 $ST2$ 之長度 (直向)，比較長的是局部特徵碼之中心結構筆形
否則

如果 $m=1$ ，再取 $ST2$ 之右鄰近筆形 ($ST3$)，並作長度比較，比較長者為中心結構筆形。

否則 $ST2$ 是中心結構筆形

否則 $ST1$ 是中心結構筆形

找局部中心結構筆形之各連接筆形 (BPR)

輸入：所有同中心結構筆形 (ID)

輸出：所有同中心結構筆形相連之結構筆形。

步驟一：(取出) 取出同中心結構筆形相鄰的右相鄰結構筆形。

($ST1$)，如果沒找着，執行步驟八

$SCAN=ID\#+1$ ， $\langle n,m \rangle = LR(SCAN)$

步驟二：〔決定〕

如果 $n \neq 1$ 則執行步驟三

否則 如果 $m=1$ 則 $IP=ST1$ 並執行步驟四

否則 如果 $m \neq 0$ ，至步驟七，否則至步驟六

步驟三〔扭〕加入適當的資料，並返回執行步驟(一)，(八)

步驟四：〔取〕取 ST1 之右連接系

步驟五：〔決定〕

如果 $11ST211 > 11ST1$

步驟六，否則執行步驟七

步驟六：〔推〕將結構筆形推入堆
值，執行步驟(一)，(八)

步驟七：〔連合〕把所有結構筆形(TP)相鄰的結構筆形
中心結構筆形相鄰，並把這些結構筆形作適當的改動
執行步驟(一)，(八)

步驟八：〔取〕取出同結構筆形 ST1 左相鄰之結構筆形

如果沒找到 則停止執行

$SCAN = ID\# - 1, \langle n, m \rangle = LR(ST1)$

步驟九：〔決定〕

如果 $m \leq 1$ 則執行步驟(三)

否則 如果 $n \leq 1$ 則執行步驟六

否則 執行步驟七

$SCAN = SCAN - 1$

編碼局部特徵是 (CODE)

輸入：局部特徵之各結構筆形(置於堆中)

輸出：特徵向量

步驟一：〔移〕取出中心結構筆形左相鄰之結構筆形，如果沒找着，執行步驟五

步驟二：〔最大重合〕：在與中心結構筆形右相鄰之結構筆形中，找出同步驟(一)所找之結構筆形有水平重合者，如果沒找着，執行步驟(四)

步驟三：〔線性度量〕：計算(用質心法)兩結構元的線性關係，如果沒有線性關係則執行步驟二，否則記錄在特徵向量內

步驟四：〔重寫〕計算每一結構筆形的斜率，及其特徵向量並作適當的記錄。返回執行步驟(一)，(五)

步驟五：〔移〕取出尚未處理之右相鄰結構筆形，如果沒找着則

。 停止操作，否則執行步驟四

<註> 在程式中，在步驟後有兩個數目，如步驟(三)，(四)表示至由進入之執行地方。

第四章 實驗與結果討論 (Experimental Result and Discussion)

本論文中所提出的中文辨認方法，已用 Fortran 語言寫成程式，並且在 PDP-11/70 上作過實驗。輸入的中文字形包括印刷體及手寫正楷體（為作者自己參閱字典逐一寫成），總計印刷體有六百字，手寫體有近九百字，實驗之結果將在本章內逐一描述及討論。

字形之設計，我們取用印刷體及手寫體，印刷體直接由中文打字機（萬國語言機器）打在紙上；手寫體之製作則是按 $1\text{cm} \times 1\text{cm}$ 之方格設計大小，再將字寫入方格內。

紙上的字，經由 C-4100 掃瞄機將其按色度值以拜特 (Byte) 讀入，掃瞄機共分 256 個色度值。實驗時是將每一個字取出，並取適當的界限值 (threshold) 將色度圖轉換成黑與白之二元圖 (Binary pattern)，然後利用所設計的程式，去分析辨認每一個字樣。

圖 (廿二) 為本論文實驗所製作的字形。

仄什亮亭亡些井互乳乙乍丹丁仇仗仙仲仿伊伐休伯伴伸伺佇位低住佐恤愜 惹恩恰悅悶倍倉
券副劣授掛掉掠拔拓拒拍拉刑刊刀冷冬冗冒册典兵兒僉兌戶打扣技擅拙抽拆慶做倨佔侯侵促俱
卡危厄叔句叨操攻絞救敦敗散昂旺旱旬冒旦斥味呼呢吹吸咨吐吏吃叱史召叫只威晒哉哥昔昇昧
已巧巨活沙汰汽汕汗毛母寺宇宮欠武女士匆夕央在板松束村朱札未木困址坐培曲朋朝晶景普响
己巾床序古安甘州江北西山東京林吉川日月年万拾百千中華民國石斗升合里步弓丈尺寸氏比每
用田由生玉令以任企但佈何火求永水代介今仁人享亞云于予上下不且世主乃久之乏乞也了止正
全內入克光先兄允共公凡出切白目知示力勿化半去又及友立否向名同司台可另右口反 老至自
平千布市已左小少尤居局工計言定好如子字存守官行衣良失夫太天大多外固因回土在堂而耳育
走買弟引式店往心必志足身車較迎近成我或手才托投門長金重支方早昌明易星是李本末有更日
隻雨陪阿阻阪羊系邑邪空竹辛禾赤矢眉皂皇皮谷豆 申甲瓦犯王訂牽牛牙片父災焚衫次果材
耶考肉肥肩肯肴胡雪青革飽飯臣舟航芝芳茂茅茶骨鳥鳴鼎鼎鼎鼎鼓鼓黃麻麥鹿壹貳叁肆伍陸
蓬蔚蔭蕨蔽堯蕪蕭薪薰藍藝藻蘊蘭鴻鶴醫麟麪麻黎齋齡兩分厘吊元角第銀洋號斤 \$ 零拾玖捌柒

(22 - 1)

圖 (1 - 4) 為實驗製作之字形 ((22 - 1) 印刷體，(22 - 2) ~ (22 - 4) 手寫體)

卅	半	弁	卒	協	卓	南	卜	占	卦	卯	印	危	卽	卷
匆	化	北	匪	區	匡	匠	匣	匪	匹	區	十	千	午	升
力	加	功	劣	劫	助	努	劬	勒	勳	勻	勾	勿	包	匆
刷	刺	到	刮	制	剝	前	刺	則	剖	剔	剛	副	割	剩
刃	刃	分	切	刈	刊	列	刑	划	勿	別	判	利	創	刻
冥	冬	冰	治	冷	冽	凍	凋	几	凱	凶	凹	出	凸	刀
八	六	兮	公	共	具	典	兼	舟	冊	再	冒	胃	最	冗
價	儷	諸	兄	光	兇	兆	先	克	克	兢	入	內	全	兩
做	偉	偶	偎	偕	偵	側	偷	博	備	傑	傘	債	僅	僥
借	倚	倒	們	諾	俚	俚	倨	俱	倡	個	候	倘	倪	匿
俟	俗	俚	俄	係	俚	俚	倨	倍	倣	倣	倣	倣	倣	倣
佰	併	侈	佩	佻	侖	侖	信	侵	倣	倣	倣	倣	倣	倣
佃	佔	但	佣	作	伯	低	余	佯	使	佬	供	例	來	侃
仗	代	令	仙	仞	仿	伉	伙	伊	休	伏	仲	件	任	仰
亭	亮	人	仁	仆	仇	仍	今	介	仄	以	付	仔	仕	他
云	井	互	五	互	些	亞	亟	亡	交	亦	亥	亨	亨	京
乎	兵	兵	乖	乘	乙	九	也	乞	乜	乳	乾	子	事	于
丞	丟	丫	中	串	丸	凡	丹	主	乃	久	之	尹	乍	之
一	守	七	三	下	丈	上	丑	丐	不	丙	世	丕	且	丘

桌	杵	枹	梧	棋	森	棺	止	正	武	肯	歪	歹	死	歿
杵	杷	柶	梧	枯	柄	柏	某	柑	樞	柱	柵	相	柶	桂
材	村	杜	東	杵	杯	東	松	板	枉	析	果	杭	枝	香
朔	朕	其	木	木	未	末	本	札	朱	朽	朴	杉	季	杏
晝	晤	普	景	晴	晶	日	曲	書	曹	最	月	有	朋	胡
旱	盱	昂	昌	明	易	昔	旺	星	映	昨	昭	是	味	時
斗	斜	斤	斤	欣	欣	斯	方	於	放	日	旦	旨	早	旭
拳	扌	持	指	扌	收	改	攻	政	故	效	教	散	敦	文
拋	抽	拓	扌	扌	扌	扌	扌	扌	扌	扌	扌	扌	扌	扌
扶	批	技	扌	扌	扌	扌	扌	扌	扌	扌	扌	扌	扌	扌
扌	房	所	肩	手	扎	扌	打	扌	扌	扌	扌	扌	扌	扌
惶	惶	慚	戈	戊	戊	戊	戎	戎	戎	戎	戎	戎	戎	戎
怒	恤	恨	恩	息	恰	恍	恣	恪	恫	恬	悅	惜	懼	
思	怠	急	性	怨	怪	怎	恰	怯	怔	快	恬	愧	恃	恆
心	必	忌	忍	忘	忙	志	忑	忠	快	忽	忡	忡	忡	怕
那	邦	邗	邗	邗	邗	邗	邗	邗	邗	邗	邗	邗	邗	邗
阱	阻	附	阿	阡	陬	降	陣	陪	陳	陸	陶	隄	隅	隕
佯	徐	徒	得	復	街	徨	衙	德	阡	防	阮	阪	阮	阮
彗	形	彤	彥	行	役	方	彼	往	征	徠	待	很	徊	律
弓	吊	引	弗	弗	弘	弛	弟	弦	弧	弭	張	彈	弓	當

庄	庇	床	底	店	府	庚	唐	廁	廂	廟	廿	廿	弄	式
帆	希	帖	帝	帟	帟	帥	帥	帳	帽	幅	干	平	罕	幻
左	巧	巨	巫	禾	差	貢	己	巳	巴	目	巷	巾	匝	布
爰	岐	岳	岩	岸	岫	岢	岬	峒	峙	嵩	豈	川	州	工
少	尔	尖	肖	尤	尸	尼	尺	局	居	屈	屈	咫	山	屹
守	宋	完	宗	官	定	宜	宙	宝	宣	家	宿	封	射	專
奴	奶	奸	好	如	妓	妹	妹	姐	姝	姝	孟	它	亢	宇
夏	夕	外	多	大	天	太	夫	夫	央	失	夾	奇	奈	女
坏	坂	坤	坦	幸	甘	坪	基	吉	士	壯	志	賈	冬	各
去	在	地	寺	圭	圪	圪	圪	圪	坊	坐	址	均	圾	坎
單	渣	嗎	嘶	嘻	囡	回	因	困	勿	固	采	國	圓	土
咪	咩	哨	唐	喧	哥	哲	咒	嘯	哩	哨	唱	啥	善	喧
哀	咨	哉	咸	唳	咳	哇	晒	咽	咪	品	哄	哈	尺	咱
吟	兒	咆	呼	咐	呱	嗽	和	咚	呢	周	咋	命	咎	咬
吼	呀	吱	含	吭	味	呵	咖	呸	咕	哎	咀	呻	呷	咄
呆	妮	吳	星	呂	君	吩	告	吹	吻	吸	吮	吵	呐	吠
吁	吋	各	向	名	合	吃	后	吒	各	吞	吾	否	呶	吧
叨	司	叫	另	只	史	叱	台	句	吟	吉	吏	同	吊	吐
卸	反	取	叔	受	叛	叟	曼	口	可	古	右	召	叮	叩
卸	卸	卸	卸	卸	卸	卸	卸	卸	卸	卸	卸	卸	卸	卸

(22-4)

4-1: 格式說明:

字以 $25\mu\text{m} \times 25\mu\text{m}$ 讀入，縮小五倍，使原來 $1\text{cm} \times 1\text{cm}$ 的字點化成 80×80 的字樣 ($1\text{pixel} = 125\mu\text{m} \times 125\mu\text{m}$)。

以「偉」字為例依次說明所印之格式的意義:

整個格式，除了字形外，我們另以三部份印出其結果:

第一部份：印出所取出之結構元；其格式為:

[$\pm D_{\#}$ ，始結構段坐標，末結構段坐標，結構元之種類]

$\pm D_{\#}$ ：依結構元取出之先後順序而編成的代號。

始(末)結構段坐標：構成結構元之始(末)結構段之直向坐標值及橫向最上與最下坐標值。

結構元種類：為本之所取出之五種結構元。依次編號為:

1：向下延伸

2：向上延伸

6：向右收縮

4：向右發散

3：向右水平延伸

第二部份：印出所取出之結構筆形；其格式為:

[$I D_{\#}$ ，構成結構筆形之結構元]

$I D_{\#}$ ：依結構筆形取出之順序所編成的代號。

構成結構筆形之結構元：包括所有結構元其組成此一結構筆形，順序為依左而右。而其相對應的結構元就是第一部份所印出者。

第三部份：印出經由編碼處理後的結果。其格式依次說明如下:

1 局部特徵碼之特徵向量：格式為 ($U_1, U_2, U_3, U_4, U_5, U_6, U_7$)

U_1 ：中心結構筆形左方之單橫筆

U_2 ：橫過中心結構筆形之橫筆

U_3 ：中心結構筆形右方之單橫筆

U_4 ：中心結構筆形左方之斜上筆

U_5 ：中心結構筆形左方之斜下筆

U_6 ：中心結構筆形右方之斜上筆

U_7 ：中心結構筆形右方之斜下筆

如第一組數據 (0, 0, 0, 1, 0, 1, 0)，意為此一中心結構筆形左方有一斜上筆，右方有一斜上筆 (就是「イ」之代碼)。

2. "MATCHED ASCID VALUE" 經由局部特徵碼表轉換後的代碼。「
」我們給它定為 9 (此並無任何意義；對於為何要給「イ」之碼定為 9)。

3. NODE = 2, ... at (18, 30, 76)

局部中心結構筆形的代號及其構成之最左方結構段的相對坐標值。

4. left branch ; 局部中心結構筆形左方的結構筆形代號。

right branch ; 局部中心結構筆形右方的結構筆形代號。

在此下之兩排資料，表示左方之代碼及其斜率之度量值，其中代碼等於 0，表示已構成一橫過中心結構筆形之橫筆。

至此；每一局部特徵碼描述完畢，而其餘之特徵碼則依次依上格式印出

5. 依次重覆，至最後一排中之兩行資料，就是經由簡化的二元樹編碼法後的次序，依次由上而下安排。

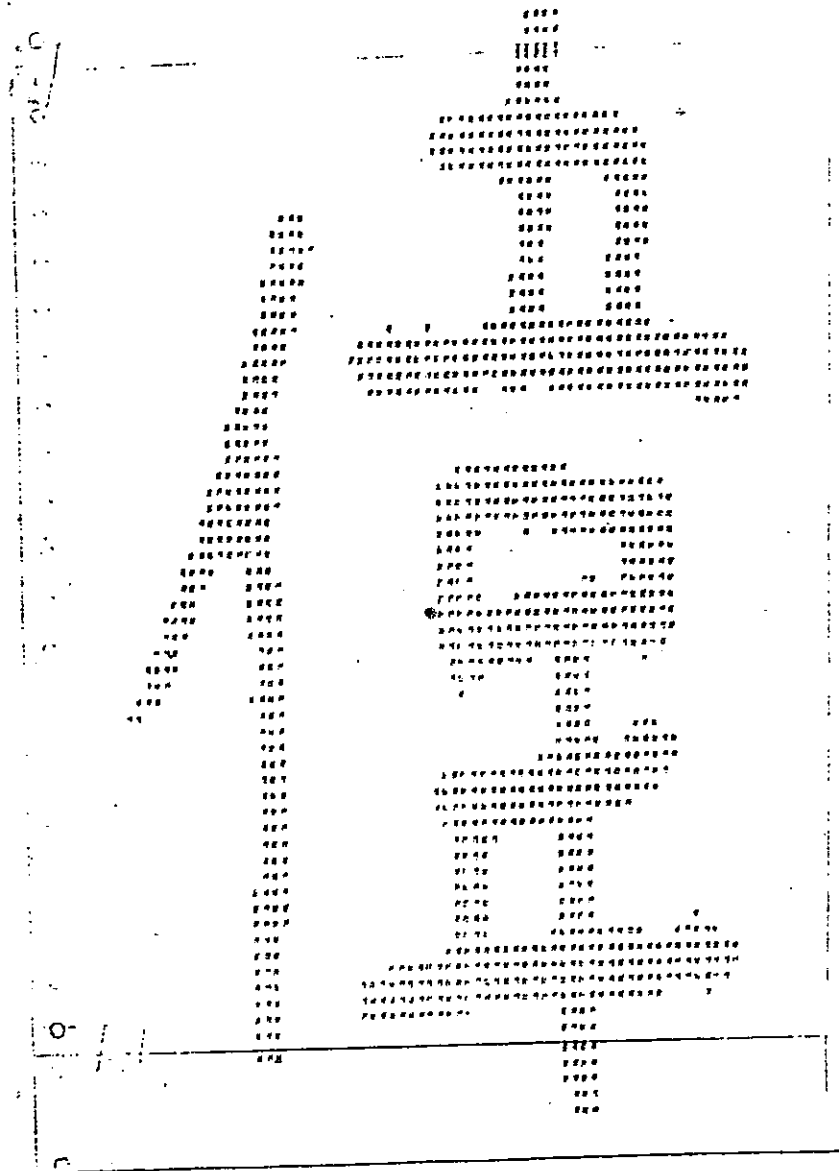
第一行代表中心結構筆形之代號。第二行為相對應的碼。以此字為例，我們即可參照附錄一之轉換表轉換，得到如下的結果：

14	71	±
22	7	ㄣ
2	9	イ
8	6	ㄇ
23	7	コ
9	71	±
17	91	ㄚ

改成橫向串列就是“±ㄣイㄇコ±ㄚ”；此一串列同理論上之編碼之次序完全相符合。

6. 顯示之各字分別為印刷體及手寫體的結果。在附錄中，我們另附部份字

形之處理結果與其所編成的串碼。



STRUCTURAL ELEMENTS:

FORMAT IS: <ID, COORDINATE, COORDINATE, TYPE>

ID	COORDINATE 1	COORDINATE 2	COORDINATE 3	COORDINATE 4	COORDINATE 5	COORDINATE 6	COORDINATE 7	COORDINATE 8	TYPE
1	4	44	54	5	53	54			2
2	6	41	53	7	50	53			2
3	8	49	52	9	47	51			2
4	10	45	49	11	44	48			2
5	12	42	46	14	39	44			2
6	15	36	44	17	35	44			2
7	18	30	76	19	27	76			2
8	20	24	34	21	23	32			2
9	22	23	30	24	15	75			4
10	25	22	32	25	21	75			4
11	25	71	73	30	71	73			2
12	30	31	34	31	31	74			2
13	31	70	73	37	70	73			2
14	32	70	74	33	31	74			2

15	(34,	31,	34)	(35,	31,	34)	3
16	(36,	30,	34)	(37,	31,	34)	6
17	(36,	59,	60)	(37,	58,	61)	4
18	(37,	18,	19)	(38,	17,	20)	4
19	(37,	40,	50)	(38,	40,	52)	1
20	(38,	31,	34)	(41,	31,	34)	3
21	(38,	58,	73)	(39,	58,	73)	3
22	(39,	17,	20)	(43,	17,	20)	3
23	(39,	39,	53)	(40,	39,	52)	6
24	(40,	58,	72)	(41,	58,	72)	2
25	(41,	39,	43)	(42,	39,	42)	6
26	(41,	47,	52)	(42,	48,	51)	6
27	(42,	30,	33)	(43,	30,	33)	3
28	(42,	58,	62)	(43,	58,	61)	6
29	(42,	69,	72)	(48,	69,	72)	3
30	(42,	39,	42)	(45,	39,	42)	3
31	(43,	42,	51)	(44,	49,	51)	3
32	(44,	17,	21)	(45,	16,	21)	2
33	(44,	30,	34)	(45,	27,	34)	2
34	(44,	59,	61)	(46,	58,	61)	3
35	(45,	47,	51)	(46,	47,	51)	3
36	(46,	17,	34)	(47,	10,	33)	2
37	(46,	39,	43)	(47,	39,	42)	6
38	(47,	47,	50)	(48,	47,	50)	3
39	(47,	57,	61)	(48,	57,	61)	3
40	(48,	10,	24)	(49,	10,	24)	3
41	(48,	30,	23)	(49,	20,	34)	1
42	(48,	39,	42)	(49,	39,	43)	1
43	(49,	47,	78)	(50,	47,	80)	1
44	(50,	16,	20)	(51,	17,	20)	6
45	(50,	30,	24)	(55,	30,	24)	3
46	(50,	39,	43)	(51,	40,	43)	6
47	(51,	47,	80)	(52,	46,	80)	2
48	(52,	17,	20)	(55,	17,	20)	2
49	(52,	40,	43)	(55,	40,	43)	3
50	(53,	46,	50)	(55,	47,	50)	6
51	(53,	56,	60)	(54,	57,	60)	6
52	(53,	69,	72)	(57,	68,	72)	3
53	(55,	57,	60)	(57,	55,	59)	2
54	(56,	17,	34)	(57,	18,	34)	6
55	(56,	40,	50)	(57,	40,	50)	3
56	(58,	19,	25)	(59,	19,	25)	3
57	(58,	30,	34)	(59,	30,	34)	3
58	(58,	40,	51)	(59,	40,	50)	6
59	(58,	55,	59)	(59,	55,	59)	3
60	(58,	69,	72)	(59,	69,	72)	3
61	(60,	31,	34)	(63,	31,	34)	3
62	(60,	40,	50)	(61,	41,	49)	5
63	(60,	56,	58)	(61,	56,	57)	6
64	(60,	69,	71)	(61,	68,	71)	2
65	(62,	68,	71)	(63,	67,	71)	2
66	(64,	31,	35)	(67,	31,	35)	3
67	(64,	68,	72)	(67,	69,	70)	6
68	(65,	32,	25)	(69,	32,	24)	5


```

( 0 0 0 0 0 0 0 0 )
MATCHED ASSIC VALUE =..... 71
MODE= REFOUND AT 29 52 72
LEFT BRANCH:
0.830000E+01 0.000000E+00
0.801000E+01 0.000000E+00
RIGHT BRANCH
0.000000E+00 0.000000E+00
0.000000E+00 0.000000E+00

```

```

( 0 0 0 0 0 0 0 0 )
MATCHED ASSIC VALUE =..... 6
MODE= REFOUND AT 27 40 50
RIGHT BRANCH
0.100000E+02 0.000000E+00
0.100000E+02 0.000000E+00

```

```

( 0 0 0 0 0 0 0 0 )
MATCHED ASSIC VALUE =..... 7
MODE= REFOUND AT 50 40 50
LEFT BRANCH:
0.100000E+02 0.000000E+00
0.100000E+02 0.000000E+00

```

```

( 0 0 0 0 0 0 0 0 )
MATCHED ASSIC VALUE =..... 91
MODE= REFOUND AT 49 47 78
LEFT BRANCH:
0.100000E+02 0.000000E+00
0.100000E+02 0.000000E+00
0.100000E+02 0.000000E+00
RIGHT BRANCH
0.000000E+00 0.000000E+00
0.000000E+00 0.000000E+00
0.000000E+00 0.000000E+00

```

```

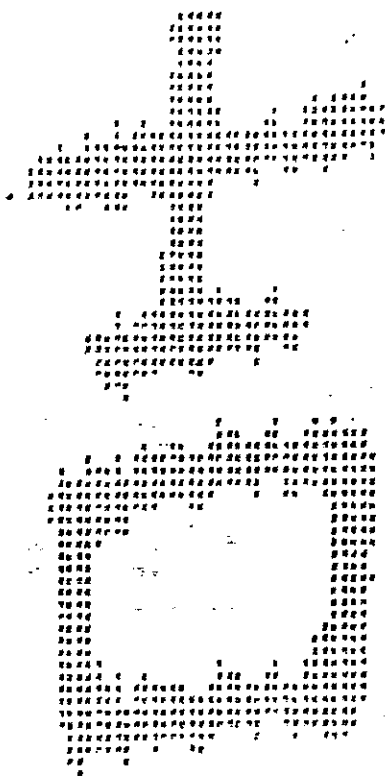
( 0 0 0 0 0 0 0 0 )
MATCHED ASSIC VALUE =..... 7
MODE= REFOUND AT 56 17 34
LEFT BRANCH:
0.100000E+02 0.000000E+00
0.100000E+02 0.000000E+00
RIGHT BRANCH
0.000000E+00 0.000000E+00

```

```

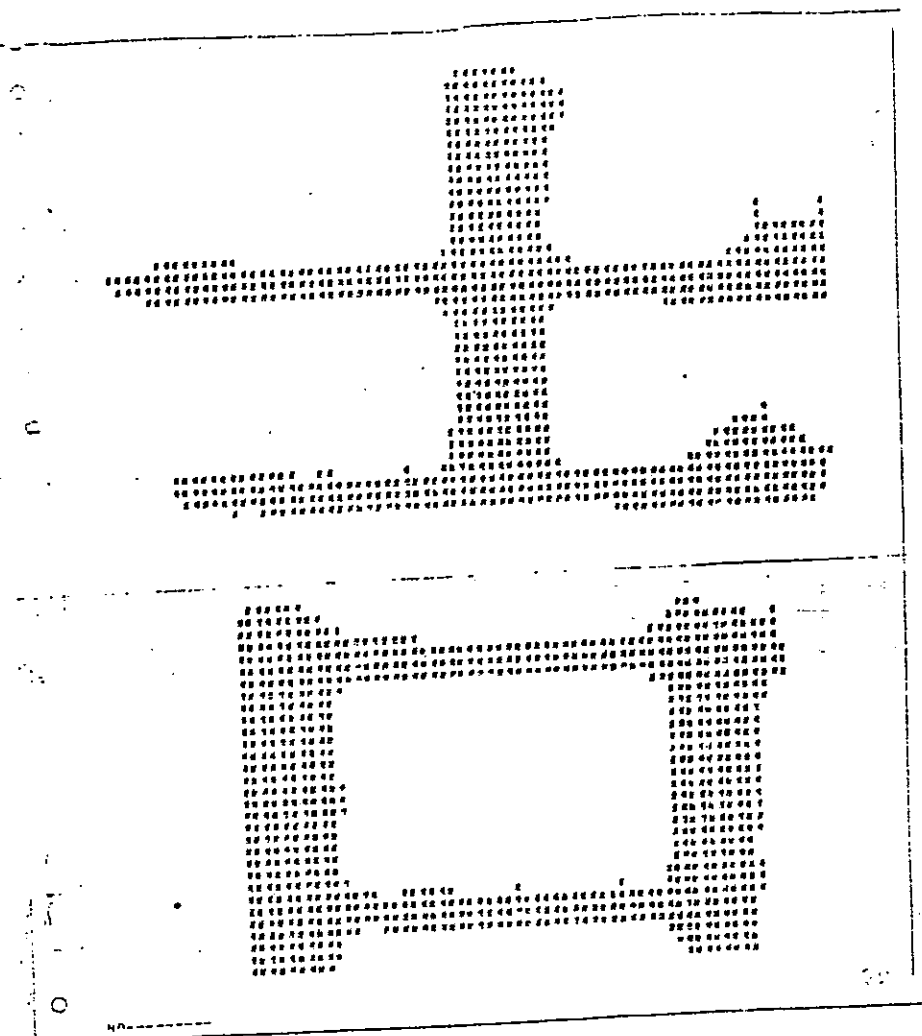
14 71
17 7
0 0
0 6
23 7
0 71
17 91

```



STRUCTURAL ELEMENTS ::

ID	COORDINATE	COORDINATE	TYPE
1	13	15	17
2	13	14	17
3	15	14	17
4	16	14	61
5	17	14	18
6	19	12	17
7	19	29	30
8	19	39	40
9	19	40	46
10	19	46	63
11	21	18	18
12	21	40	45
13	21	53	53
14	22	17	13
15	23	18	18
16	23	41	45
17	22	0	64
18	24	0	32
19	26	0	17
20	25	0	43
21	26	0	43
22	28	0	42
23	28	0	62
24	29	0	31
25	20	0	43
26	31	0	43
27	31	0	17
28	31	0	32
29	14	17	2
30	13	17	1
31	18	18	3
32	13	17	6
33	14	18	3
34	19	17	6
35	29	33	1
36	40	46	3
37	56	53	3
38	11	18	3
39	30	45	3
40	37	63	3
41	0	34	1
42	11	16	2
43	30	43	3
44	7	62	3
45	7	32	3
46	12	17	3
47	30	42	6
48	30	42	6
49	30	62	3
50	30	62	3
51	2	32	3
52	30	43	3
53	30	63	3
54	2	17	3
55	2	31	3



STRUCTURAL ELEMENTS	FORMAT IS:	<ID	COORDINATE	COORDINATE	TYPE
1	(5	21	22	1
2	(7	21	22	1
3	(9	20	23	1
4	(11	20	23	1
5	(17	38	40	1
6	(11	38	40	1
7	(15	38	40	1
8	(17	38	40	1
9	(17	49	79	1
10	(37	22	23	1
11	(19	22	41	1
12	(22	40	79	1
13	(22	40	41	1
14	(22	40	41	1
15	(24	38	79	1
16	(24	38	41	1
17	(26	50	41	1
18	(26	51	56	1
19	(26	51	66	1
20	(28	77	75	1
21	(33	33	41	1
22	(34	33	55	1
23	(30	77	76	1
24	(32	73	76	1
25	(37	73	41	1
26	(35	73	55	1
27	(58	33	41	1
28	(37	33	41	1

STRUCTURAL ELEMENTS ::

	FORMAT	IS	<ID	, COORDINATE	, COORDINATE	, TYPE				
1	(2,	45,	75)	(3,	46,	77)	(1
2	(4,	7,	7)	(5,	4,	8)	(4
3	(4,	4,	77)	(5,	45,	75)	(2
4	(6,	4,	9)	(7,	3,	9)	(2
5	(6,	45,	50)	(8,	45,	50)	(3
6	(6,	57,	62)	(8,	57,	62)	(3
7	(6,	68,	75)	(9,	68,	75)	(3
8	(8,	4,	9)	(9,	4,	10)	(1
9	(8,	16,	38)	(9,	16,	41)	(1
10	(8,	72,	76)	(9,	72,	77)	(1
11	(9,	45,	51)	(10,	45,	50)	(6
12	(9,	57,	53)	(10,	58,	63)	(6
13	(10,	3,	10)	(11,	4,	10)	(6
14	(10,	16,	40)	(11,	17,	40)	(6
15	(10,	71,	76)	(13,	71,	76)	(6
16	(11,	45,	50)	(13,	45,	48)	(6
17	(11,	5,	63)	(12,	5,	63)	(6
18	(12,	4,	10)	(13,	4,	9)	(6
19	(12,	17,	22)	(13,	17,	22)	(6
20	(12,	33,	40)	(16,	33,	37)	(6
21	(13,	5,	52)	(14,	5,	62)	(6
22	(14,	3,	9)	(15,	3,	11)	(1
23	(14,	33,	72)	(15,	19,	23)	(1
24	(14,	42,	48)	(16,	44,	49)	(6
25	(14,	0,	76)	(16,	71,	76)	(6
26	(14,	5,	52)	(16,	5,	62)	(6
27	(14,	5,	10)	(17,	5,	10)	(6
28	(16,	19,	23)	(17,	18,	22)	(6
29	(17,	33,	37)	(18,	22,	36)	(6
30	(17,	44,	76)	(18,	43,	76)	(6
31	(18,	4,	9)	(26,	4,	9)	(6
32	(19,	17,	22)	(22,	17,	22)	(6
33	(19,	33,	36)	(20,	33,	36)	(6
34	(20,	44,	76)	(21,	43,	76)	(6
35	(21,	3,	10)	(22,	3,	10)	(6
36	(21,	33,	37)	(22,	33,	37)	(6
37	(22,	43,	48)	(21,	43,	48)	(6
38	(22,	77,	63)	(23,	57,	62)	(6
39	(22,	71,	76)	(24,	71,	76)	(6
40	(23,	4,	9)	(24,	4,	9)	(6
41	(23,	17,	21)	(24,	17,	21)	(6
42	(23,	33,	36)	(24,	31,	36)	(6
43	(24,	42,	48)	(25,	42,	48)	(6
44	(24,	57,	62)	(25,	58,	62)	(6
45	(25,	4,	10)	(26,	4,	9)	(6
46	(25,	17,	22)	(30,	17,	22)	(6
47	(25,	32,	37)	(26,	32,	36)	(6
48	(25,	71,	77)	(26,	71,	76)	(6
49	(26,	42,	47)	(30,	42,	47)	(6
50	(26,	58,	62)	(27,	57,	62)	(6
51	(27,	3,	9)	(28,	4,	10)	(1
52	(27,	31,	36)	(28,	31,	36)	(6
53	(27,	70,	76)	(28,	71,	76)	(6
54	(28,	57,	62)	(29,	56,	62)	(6
55	(29,	4,	10)	(31,	4,	10)	(6
56	(29,	31,	27)	(30,	30,	40)	(4
57	(29,	71,	76)	(32,	71,	76)	(6
58	(30,	57,	63)	(32,	58,	63)	(6
59	(31,	17,	39)	(32,	17,	39)	(6
60	(31,	42,	48)	(32,	42,	48)	(6
61	(32,	4,	9)	(33,	4,	9)	(6
62	(33,	18,	39)	(34,	18,	39)	(6
63	(33,	43,	75)	(34,	43,	76)	(6
64	(34,	4,	10)	(35,	4,	10)	(6
65	(35,	17,	27)	(36,	17,	27)	(6

(2 0 0 0 0 0 0) ...
 MATCHED ASSIC VALUE = 7
 NODE= 13 FOUND AT 31 17 39
 LEFT BRANCH:

0.800000E+01 0.000000E+00
 0.700000E+01 0.000000E+00

(3 0 0 0 0 0 0) ...
 MATCHED ASSIC VALUE = 5
 NODE= 14 FOUND AT 33 43 75
 LEFT BRANCH:

0.120000E+02 0.000000E+00
 0.110000E+02 0.000000E+00
 0.100000E+02 0.000000E+00

A SINGLE STROKE AT: (52 23 48)

(0 0 0 0 0 0 0) ...
 MATCHED ASSIC VALUE = 1
 NODE= 17 FOUND AT 75 1 80
 LEFT BRANCH:

0.160000E+02 0.000000E+00

2 0
 6 6
 13 7
 15 0
 1 4
 0 31
 14 5
 17 1

.....

STRUCTURAL ELEMENTS::

	FORMAT	IS:	<ID.	,COORDINATE	,COORDINATE	,TYPE>			
1	(3,	7,	8)	(4,	6,	8)	2
2	(5,	7,	8)	(6,	7,	8)	3
3	(5,	41,	78)	(8,	41,	78)	3
4	(7,	6,	8)	(11,	6,	8)	3
5	(9,	42,	63)	(10,	42,	63)	3
6	(9,	59,	77)	(13,	71,	73)	6
7	(10,	15,	37)	(13,	15,	37)	3
8	(11,	42,	47)	(12,	43,	46)	6
9	(11,	56,	60)	(13,	57,	59)	6
10	(12,	7,	8)	(26,	7,	8)	3
11	(13,	43,	46)	(17,	43,	46)	3
12	(14,	16,	36)	(15,	16,	35)	6
13	(14,	57,	59)	(23,	57,	59)	2
14	(14,	71,	73)	(21,	71,	73)	2
15	(16,	17,	21)	(17,	17,	21)	3
16	(16,	29,	34)	(18,	29,	34)	6
17	(18,	13,	20)	(34,	13,	20)	3
18	(18,	44,	46)	(19,	44,	46)	3
19	(19,	31,	34)	(20,	31,	34)	3
20	(20,	43,	46)	(21,	43,	46)	3
21	(21,	32,	34)	(22,	32,	34)	3
22	(22,	44,	46)	(23,	44,	47)	1
23	(22,	70,	73)	(23,	70,	73)	3
24	(23,	31,	34)	(24,	31,	34)	3
25	(24,	44,	73)	(28,	44,	73)	3
26	(25,	32,	34)	(32,	32,	34)	2
27	(27,	7,	9)	(41,	7,	9)	3
28	(29,	44,	50)	(30,	44,	50)	2
29	(29,	55,	61)	(30,	55,	61)	3
30	(29,	70,	73)	(30,	71,	73)	6
31	(31,	44,	47)	(33,	45,	46)	6
32	(31,	57,	60)	(34,	57,	60)	3
33	(31,	71,	73)	(43,	71,	73)	3
34	(33,	31,	34)	(34,	32,	34)	6
35	(34,	45,	46)	(35,	45,	46)	3
36	(35,	18,	21)	(39,	18,	21)	3
37	(35,	32,	34)	(37,	32,	34)	3
38	(35,	53,	60)	(41,	53,	60)	3
39	(36,	44,	47)	(37,	44,	46)	6
40	(38,	31,	35)	(39,	30,	35)	7
41	(38,	44,	46)	(39,	44,	46)	3
42	(40,	17,	37)	(41,	16,	38)	4
43	(40,	44,	47)	(43,	44,	47)	3
44	(42,	6,	9)	(43,	6,	9)	3
45	(42,	16,	38)	(43,	16,	38)	3
46	(42,	57,	60)	(43,	56,	61)	4
47	(44,	5,	9)	(45,	4,	9)	2
48	(44,	17,	37)	(45,	17,	36)	6
49	(44,	43,	78)	(45,	42,	79)	4
50	(46,	4,	9)	(48,	4,	9)	3

51	(46,	18,	22)	(47,	18,	22)	3
52	(46,	42,	78)	(48,	43,	76)	6
53	(49,	5,	10)	(50,	6,	10)	6
54	(49,	44,	50)	(51,	46,	47)	6
55	(51,	6,	10)	(53,	8,	9)	6
56	(58,	71,	72)	(59,	71,	73)	1
57	(59,	8,	64)	(60,	7,	64)	2
58	(60,	71,	73)	(61,	72,	73)	6
59	(61,	7,	64)	(63,	7,	62)	6
60	(62,	72,	73)	(63,	72,	73)	3
61	(64,	3,	13)	(65,	8,	13)	3
62	(64,	72,	74)	(66,	73,	76)	1
63	(67,	73,	79)	(68,	72,	79)	2
64	(69,	72,	79)	(71,	71,	78)	2
65	(72,	3,	77)	(73,	3,	75)	6
66	(74,	3,	71)	(75,	3,	67)	6
67	(76,	4,	7)	(77,	4,	7)	2

STRUCTURAL STROKES :

	FORMAT	IS	<ID.	FL	EMPH	T>
1<	1 2 4 10 27 44 47 50 53 55	0	0	0	0	0
2<	3 0 0 0 0 0 0 0 0 0	0	0	0	0	0
3<	5 0 0 0 0 0 0 0 0 0	0	0	0	0	0
4<	6 14 23 0 0 0 0 0 0 0	0	0	0	0	0
5<	7 12 0 0 0 0 0 0 0 0	0	0	0	0	0
6<	8 11 18 20 23 0 0 0 0 0	0	0	0	0	0
7<	0 13 0 0 0 0 0 0 0 0	0	0	0	0	0
8<	15 17 36 0 0 0 0 0 0 0	0	0	0	0	0
9<	16 10 21 24 26 34 37 40	0	0	0	0	0
10<	25 0 0 0 0 0 0 0 0 0	0	0	0	0	0
11<	20 31 35 39 41 43	0	0	0	0	0
12<	29 32 38 46 0 0 0 0 0 0	0	0	0	0	0
13<	30 33 0 0 0 0 0 0 0 0	0	0	0	0	0
14<	42 45 49 0 0 0 0 0 0 0	0	0	0	0	0
15<	49 52 0 0 0 0 0 0 0 0	0	0	0	0	0
16<	51 0 0 0 0 0 0 0 0 0	0	0	0	0	0
17<	54 0 0 0 0 0 0 0 0 0	0	0	0	0	0
18<	56 58 60 62 63 64	0	0	0	0	0
19<	57 59 0 0 0 0 0 0 0 0	0	0	0	0	0
20<	61 0 0 0 0 0 0 0 0 0	0	0	0	0	0
21<	65 66 0 0 0 0 0 0 0 0	0	0	0	0	0
22<	67 0 0 0 0 0 0 0 0 0	0	0	0	0	0

A SINGLE STROKE AT: (3 7 8)

(0 0 3 0 0 0 0)
 MATCHED ASSIC VALUE = 4
 MODE= 2FOUND AT 5 41 78COVER
 RIGHT BRANCH

0.400000E+01 0.000000E+00
 0.600000E+01 0.000000E+00
 0.700000E+01 0.000000E+00

(0 3 0 0 0 0 0)
 MATCHED ASSIC VALUE = 31
 MODE= 16 FOUND AT 24 44 73
 LEFT BRANCH:

0.700000E+01 0.000000E+00
 0.600000E+01 0.000000E+00
 0.400000E+01 0.000000E+00

RIGHT BRANCH

0.000000E+00 0.000000E+00
 0.000000E+00 0.000000E+00
 0.000000E+00 0.000000E+00

(0 0 2 0 0 0 0)
 MATCHED ASSIC VALUE = 0
 MODE= 5 FOUND AT 10 15 37
 RIGHT BRANCH

0.000000E+01 0.000000E+00
 0.000000E+01 0.000000E+00

(0 0 0 0 0 0 0)
 MATCHED ASSIC VALUE = 7
 MODE= 14 FOUND AT 40 17 37
 LEFT BRANCH:

0.000000E+01 0.000000E+00
 0.000000E+01 0.000000E+00

(3 0 0 0 0 0 0)
 MATCHED ASSIC VALUE = 5
 MODE= 15 FOUND AT 44 43 78
 LEFT BRANCH:

0.120000E+02 0.000000E+00
 0.120000E+02 0.000000E+00
 0.110000E+02 0.000000E+00

(0 0 0 0 0 0 0)
 MATCHED ASSIC VALUE = 1
 MODE= 21 FOUND AT 72 3 77
 LEFT BRANCH:

0.100000E+02 0.000000E+00

(6 0 0 0 0 0 0)
 MATCHED ASSIC VALUE = 1
 MODE= 19 FOUND AT 59 2 64

1 0
 5 6
 14 7
 19 1
 21 1
 2 4
 10 01
 15 5

4-2 討 論

總和所設計的字形之實驗分析結果，我們發現：

(一)字之分析能力：

以附表中的局部特徵碼作為基本碼，對手寫體（約300字）作一實驗統計；整個過程是先由此方法分析字形，針對各局部碼在表內作局部碼轉換，如果局部碼未能轉換，我們將之列於失敗。如此一來，我們得到：

1 局部中心筆形數目同預測者不合者：

實驗結果，此法可以正確的找出各局部中心筆形。

2 輸出字碼結果錯誤：

在300字中，佔有8.5%；究其原因

(1)斜率特徵之度量所引致的誤差。

(2)字之相銜性未如定義者。

3 字內之局部碼未在表內：

在此一統計中，共有16.5%的字局部碼未在列表中。究其原因：

(1)多筆形連接，形成多重筆形。如“貝”字與“貝”字僅差在一斜撇之相連部份。

(2)字之筆劃太多，各筆形間由於機器（scanner）之鑑別率；致使字形模糊。

由此一實驗結果，證實此一方法真正的辨認能力僅75%，但不可忽視的是在僅有一百多個局部碼中，就有此一辨析能力，使我們發覺對於正確的書寫方式；我們可以利用為數不多的局部碼去編串文字。

(二)字之處理速度：

整個軟體特式，在PDP-11/70 RSX-11M之作業系統內執行，輸入之字形以 80×80 之方陣處理。從字開始分析至輸出號碼，其間處理之時間少者約為五秒，字形筆劃多者要到十秒以至廿秒；亦即其間之浮動很大。試以 40×40 作實驗，字之處理時間約在五秒內。

字形之縮小，處理速度隨之亦快，但對於組合字形，往往由於縮小而致成二種不同之字；對於處理之速度，字形之大小影響因素頗大。在適當的設計之下，當會有助於速度的提高。

(三)字之編碼辨別：

本文所設計的編碼方法，我們發現：

- (1)以有限數目的特徵碼，即可編串文字描述。
- (2)字碼之長度較短，一般如“兼”之複新字，其長度亦只是“7”（即
八、 手子、 幸、 ）。
- (3)字碼短，易於儲存，節省記憶單位。

然而，其存在之缺點是：

- (1)不同的字會產生相同的字碼：如“土^上干工”按本文所提之方法，其描述結果均為一直筆，被二橫筆跨過。
- (2)同一字形可能俱有不同的字碼：對於印刷體，有正確的書寫方式，此一結果反少。對於手寫體，由於書寫的不一致，如“不”及“丩”即構成二種碼。

(四)字之幾何轉換：

- (1)放大及縮小：在合理的範圍內，沒有影響。
- (2)字之旋轉：大角度的旋轉，如木 \Rightarrow 𣎵，顯然本文所設計的方法不能輸出相同的碼。

(五)印刷體與手寫體：

按照一定的書寫方式，使各局部特徵碼間的平面位置關係保持不變，由(4~4)之附例中我們可以看到手寫體“吉”字與印刷體“吉”字經過二元樹編串的碼是相同的。然而；一般而言，手寫體間各特徵的關係常不易維持，因此；同一字將出現有兩種不同的串碼，為補足此一缺陷，我們認為可以在字庫中存不同的串碼以代表同一字。

第五章 結 論 (Conclusion)

由結構方法分析中文字形在本文中已作深入的探討，從實驗的結果而言，本文所提出的方法，具有下述的優點：

- 1 符合中文字形結構形式，適當加入語意，使辨認之潛力增加。
- 2 方法簡單，省却繁瑣的數值計算。

然而從實用的觀點而言，要達到真正的用機器辨認中文字形，我們認為尚須作如下之努力及突破：

- 1 正規化中文字形之標準書寫方式。
- 2 去除筆形與筆形間的連接條件限制，尋求一種更有效的編碼方法，並對處理所花費的時間不致負荷過巨。

參 考 文 獻

1. R. CASEY and G. NAGY. "Recognition of Printed Chinese Characters", IEEE TRANS ON EC-15 NO. 1. 1966
2. S.YAMAMOTO. etc. "Chinese Character Recognition By Hierarchical Pattern Matching" IJGPR. 1973
3. Y. NAKANO. etc. "Improvement of Chinese Character Recognition Using Projection Profiles" IJGPR, 1973
4. PAUL. P. WANG. "The Topological Analysis, Classification, and Encoding of Chinese Characters for Digital Computer Interface-part I. CCIOS 1973, VOL. 1
5. PAUL. P. WANG. "The Topological Analysis, Classification, and Encodig of Chinese Character for Digital Computer Interface-part II. IJCPR. 1973
6. W. W. STALLING "Chinese Character Recognition". Syntactic Patten Recognition, Applications. ED. K. S. FU. Springer Verlag, 1977
7. H. OGAWA and K. TANIGUGHI, "A method of Handwritten Chinese Character Recognition "ICS, '77
8. MAKOTO. YOSHIOA "Hand written Chinese Character Recognition by an Analysis-bySynthesis Method" PPR
9. T. AGUI and H. NAGAHASHI "A Coding Methos of Chinese Characters" IEEE TRANS on PAMI Vol. PAMI-1, NO. 4, 1979
10. Y.L. MA "Computer-Aided Chinese Character Recognition by Forward Morkovian Dynamic Programming" CCIOS, 1973
11. AHO.H.V., AND J.D. ULLMAN "The Theory of Parsing, Translation, and Compiling, Volume 1: parsing" Prentice-Hall, Englewood Cliffs, N. J., 1972
12. K. S. FU. "Syntactic Methods in Pattern Recognition" Academic Press, 1974
13. S.Y.LU and K.S.FU "Applicability of Pattern Handling Methods to Chinese Language Processing" in Published.
14. 辭源 臺灣商務印書館。

附 錄

(一) 局部特徵碼轉換表：

使用說明：每一表按局部特徵向量（參閱（2-4-2））之 F_2 值分類。Page (N) 意即 $F_2 = N$ 之局部特徵向量。在諸表中各量之安排次序爲：< 左方接連之橫筆數，左方接連之斜上筆數，左方接連之斜下筆數，右方接連之橫筆數，右方接連之斜上筆數，右方接連之斜下筆數 >。在 ASSIGNED CODE 欄內每一特徵向量相對應一給定值。代表的字形在 PATTERN 欄內可以找着。

PAGE	FEATURE	ASSIGNED CODE	PATTERN
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5
((6	6
((7	7
((8	8
((9	9
((0	0
((1	1
((2	2
((3	3
((4	4
((5	5

ㄟ字形及串碼表：

本節節錄字形（計 147 個）經由程式處理後之串碼（其中字形取自圖(三)之(2)與(4)的樣本內。

字 形	串 碼	字 形	串 碼
丁	41	偃	9,46,91,5,49,58
丟	95,6,7	冶	0,0,6,7,67
兵	78,71,0	切	49,59,7
井	71,71	刮	75,0,6,7,1
代	9,42,0,7	劣	0,0,0,0,59,7
佔	9,41,6,7	匝	1,6,7,71,7
併	9,91,71(9,0,0,71,71)	卒	0,0,6,6,41
俗	0,0,9,7,6,7	串	6,7,6,101,7
參	6,7,59,0,0	作	6,9,41
加	59,7,6,7,0	借	41,41,9,59,6,7
化	9,6,7	借	1,1,9,4,5
丫	2	共	71,71,0,0
兵	53,71,0	刈	3,0,7
人	6	侖	0,0,6,71,71,7
企	0,41,41	具	51,52,0,0
但	9,4,5,0	刊	71,0,7
偶	9,4,1,11,5,6,71,7	卓	41,45,71
功	45,45,7	九	59,6,7
卉	41,59,41	亞	6,6,7,7,7,6
中	6,71,7	俎	6,6,64,63
佣	9,4,91,5	典	46,91,91,47,0,0
佩	9,31,1,7,71,7	* 南	71,6,91,7(71,6,0,0,71,7)
們	9,4,5,5	丹	41,0,45

也	6,41,7,7	叵	1,6,7
伙	9,0,0,6	區	1,6,7,6,7,7,7
何	9,6,7,41	反	7,6,7,43
倨	6,7,9,41,6,7	司	0,7,6,7
偷	0,0,9,4,5,1,7	时	6,7,0,41
兼	0,0,101,47,0,101,0	呀	6,7,49,82
刑	71,71,1,7	咒	6,7,6,7,31,13,7
* 刺	6,92,59,1,1(6,92,7,1,1)	喳	75,6,7,4,5,0
劬	6,7,59,7,6,7	又	32,6
匣	6,4,5	宋	0,44,7
* 丐	9,71,7(6,71,7)	爾(尔)	6,7,0,7,0
主	0,91	希	43,59,6,41,7
估	6,71,6,7	取	55,63,7,6,43
* 佯	9,101(9,0,0,91)	各	6,3,6,7
* 侵	41,9,6,0,7,43(41,9,6,0,7,3)	岐	6,7,71,7,6,43
* 俱	7,51,52,0,0(9,51,52,0,0)	坤	45,4,91,5
則	25,26,0,0,0,7	外	6,32,16
訖	41,6,7,6,7	尖	0,1,0,71
交	0,0,0,0,3	床	31,7,43
休	9,44	叔	6,7,0,41,0,43
佐	9,59,71	向	49,6,7,7
侯	41,9,73,7	呈	6,7,81
* 倡	7,4,5,4,5(9,4,5,4,5)	寺	71,0,41
匹	6,59,41	肖	0,1,0,4,5
卦	71,6,45	巫	6,71,6
久	6,7,6	只	6,7,0,0
個	9,6,45,7,6,7	* 呂	6,7,71,7,6(6,7,71,6,7)
再	46,101,47	附	6,7,9,0,41
別	6,7,53,7,1,7	信	0,0,6,7,0,0,6,7

圭	71,71	*咽	6,7,6,71,7(6,7,6,43,7)
大	43	已	4,7,7
尤	59,0,6,7	叭	6,7,0,0
店	7,6,41,6,7	啞	6,7,71,6,7
合	0,0,6,7	哩	4,5,6,7,111
囚	6,6,7	壬	81
圯	71,7,6,7	央	6,73,7,41
天	73	宣	6,7,0,45,0
叟	6,41,7,6,7,43	居	6,7,71,6,7
叱	6,7,24,7	去	41,6,7
呵	6,7,6,7,42	古	71,6,7
哇	71,6,7,71	咬	41,41,6,7,3
*回	41,6,7,7(6,6,7,7)	品	6,7,6,7,6,7
太	43,0	固	6,91,7,6,7
台	6,7,6,7	註*：爲電腦之結果同想像者不同 而正確者在括號內。	
后	7,0,6,7		
告	6,71,6,7		
咖	59,6,7,7,7		
和	48,6,7		
哂	71,6,7,71,6,7		
因	6,43,7		
夫	73		
宙	0,7,4,91,5		
己	7,6,7		
口	6,7		
句	6,6,7,7		
吒	6,7,57,7		
吓	6,7,75,0		
咚	7,6,6,7,43,0		

作者簡歷

林慶良於民國四十二年十一月五日出生於台灣省屏東縣琉球鄉。
民國六十五年六月於清華大學物理系取得理學士學位。