# Dynamic Stroke Information Analysis for Video-Based Handwritten Chinese Character Recognition

Feng Lin and Xiaoou Tang
*Department of Information Engineering*
*The Chinese University of Hong Kong*
*Shatin, Hong Kong*
*{flin0,xtang}@ie.cuhk.edu.hk*

## Abstract

*Video-based handwritten Character Recognition (VCR) system is a new type of character recognition system with many unique advantages over on-line character recognition system. Its main problem is to effectively extract stroke dynamic information from video data for character recognition. In this paper, we propose a new stroke extraction algorithm through dynamic stroke information analysis for a VCR system. The experimental results on over 3000 video character sequences show that our system can extract the Chinese character stroke dynamic information similar to an on-line system.*

## 1. Introduction

For a pen-based computer, the input is recognized by the computer through automatic handwritten character recognition. The handwritten character recognition systems are traditionally categorized into off-line and on-line systems based on the data input [1]. The off-line recognition is performed on character images converted by optical scanners [2][3][4]. The on-line system captures temporal or dynamic information of the writing, including the number of strokes, the order and run direction of each stroke, through an electronic tablet [5][6][7]. The on-line system has much higher recognition rate than its off-line counterpart. However, for traditional on-line system, writing on the slippery surface of the writing pad is cumbersome and time consuming. The extra cost of the pen and pad, as well as the inconvenience of using them, greatly undermines the popularity of pen-based on-line system.

We recently develop a video based handwritten Chinese character recognition system that combines the advantages of both the online and offline approaches [8]. Since writing on paper is the most natural way for handwriting, the system allows users to write on any regular paper just like using the off-line system. However, instead of waiting for a whole page to be written and then scanned into a computer, we use a video camera attached on the computer to capture a sequence of the character images while it is being written on the paper. Then using video process techniques, the temporal information of each stroke is extracted. Thus we can capture the stroke temporal information similar to the on-line system. This allows the system to achieve high recognition accuracy like an on-line system, but without requesting users to write awkwardly on the slippery surface of an on-line writing board.

The key step of the VCR system is to effectively extract the stroke dynamic information from the video sequence. Once the stroke dynamic information is extracted, the recognition process is exactly the same as the traditional on-line system. Two dynamic stroke information extraction schemes have been proposed recently. One scheme simulates the on-line system to directly track the pen-tip motions and has been applied in signature verification [8].The second scheme is based on the comparison and tracking of ink pixels (the results of pen-down motions) along the video sequence temporal direction [9][10].

Both approaches struggle to accurately extract the dynamic information for *every single ink pixel*, which is often mixed with the pen, hand, and their shadows. The computational cost is very high and the result is often unstable. In addition, when a character is written too fast to track every pixel, this pixel-by-pixel tracking method may produce incorrect pixel connections and fail to recover all the strokes. In this paper we propose a novel dynamic stroke information extraction method based on *stroke* analysis instead of pixel analysis. The experimental results on a large data set show that our approach can correctly recover Chinese character strokes and achieve recognition accuracy comparable to the on-line recognition system.

## 2. System Overview

Fig. 1 illustrates the functional diagram of our VCR system. A sequence of images is captured by a video camera (Fig. 1(b)) and passed to the *pre-processing* module. The output of the pre-processing module is a binary sequence containing both the written characters

and the background noise caused by the pen, hand, and their shadows, as shown in Fig. 1(c). The complete written character skeleton shown in Fig. 1(d) is obtained from the last frame with a clear and finished character after the hand and pen left the character location. An off-line system would recognize this thinned character directly. Because of the lack of the character dynamic information, the off-line system usually performs poorly. However, the stroke tracing and extraction techniques developed for the off-line systems can be helpful for us to organize the spatial information of the ink pixels, which is the goal of the *Static Information Analysis* module in Fig. 1(a). We have developed an efficient stroke extraction method based on a bi-directional graph to extract off-line handwritten Chinese character strokes [11]. Applying it on the character shown in Fig. 1(d), we obtain labelled stroke results in Fig. 1(e). This paper focuses on the *Dynamic Information Analysis* module, which extracts the stroke dynamic information, such as the run direction and the drawing order from the video sequence. The output of this block is shown in Fig. 1(f) with all the information required by the *On-line Character Recognition*.

## 3. Stroke Dynamic Information Analysis

### 3.1. Clues to retrieve stroke dynamic information

The strokes in the form of skeletons, derived from the *Static Information Analysis*, indicate the spatial relationship of character skeleton pixels. With the help of an image sequence, we can further recover the dynamic information for each stroke. In fact, the frames that describe a stroke-writing process can provide the full dynamic information of the stroke if the shooting camera is mounted properly. An example of such frames is shown in Fig. 2(a). The corresponding binary image is shown in Fig. 2(b). The final stroke trajectory, i.e., the result of the static analysis, is marked by the dotted line. We can see that this stroke is written from point "x" to point "+". The stroke drawing timeslot, i.e. its pen-down to pen-up slot, can be located by analyzing the video sequence containing frames similar to Fig. 2(a). By comparing the drawing timeslots of all character strokes, the drawing order of each stroke can be recovered.
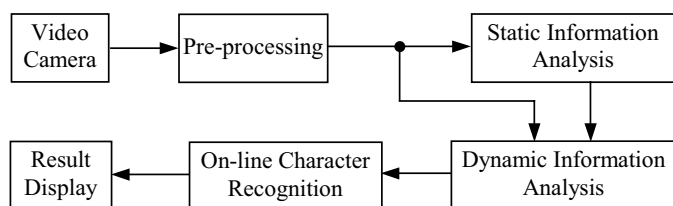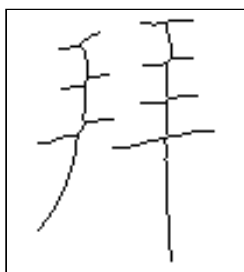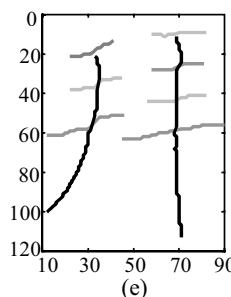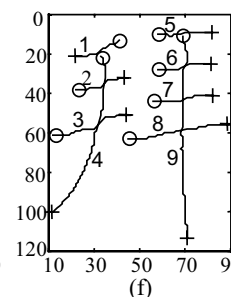


(a)



(b)



(c)



(d)



(e)



(f)

Fig. 1. (a) The VCR system block diagram. (b) An original video image sequence. (c) The binary sequence of (b). (d) The thinned result of the last image in sequence (c), which is the reference image for the following analysis. (e) The output of *Static Information Analysis*, where the character static stroke information is extracted from the reference image (d). (f) The output of *Dynamic Information Analysis*, where the character stroke dynamic information is completely recovered.

### 3.2. Mathematical descriptions

**(a) Ink pixel descriptions**

Using the illustration of a binary frame at time $t$ shown in Fig. 2, we first define some terms that we will use often.

An ink pixel is a pixel on a stroke that has been drawn so far. It must be a black pixel, however a black pixel may not be an ink pixel. A black pixel may also be produced by the pen, hand, or their shadows.

An ink skeleton pixel (ISP) is an ink pixel on the skeleton of a stroke that has been drawn so far.

A black skeleton pixel (BSP) is a black pixel on the skeleton of a completed stroke extracted from the final frame. An ISP must be a BSP. However, a BSP is not necessarily an ISP on frame $t$ as it may be a skeleton pixel of a stroke that has not been drawn but has turned black due to the pen, hand or their shadows.

The stroke trajectory can be derived from the ISPs. In what follows, the set of ISPs is denoted by *IS*, and the set of BSPs by *BS*.

From the above analysis, we have

$$IS \subset BS . \qquad (1)$$

The profile width of an ink pixel is defined as the width of all black pixels around the ink pixel, as shown in Fig. 2(b). The dynamic ranges of the profile widths of ink pixels are bounded by a threshold if the pixels are not covered by other objects such as the pen or hand. This boundary threshold can be estimated using the clean character in the last frame as shown in Fig. 1(c).

We need to estimate the boundary threshold of profile width of ink pixels. For a general ball-pen, each stroke usually has the same width throughout, so we assume that the profile widths of all ink pixels on the same stroke are approximately the same. The boundary threshold can be estimated from the final stroke

If a pixel is a BSP and its profile width is within boundary threshold, then it is called a width-limited pixel (WLP). The set of WLPs is denoted by *WS*. Obviously,

$$WS \subset BS . \qquad (2)$$

Note that an ISP may not be a WLP since it may be covered by a shadow during the writing process, rendering its profile width outside the boundary threshold at a particular time. Similarly, a black pixel caused by the shadow of the pen or hand may accidentally satisfy the requirement to become a WLP although it is definitely not an ISP.

The set *IS* is difficult to obtain from the binary image sequence (Fig. 1(c)) due to the sensitivity of ISP's profile width to noise, whereas the set *WS* is computable. To capture the character-writing information, our system requires the user not to shadow the pen tip during his writing, thus an ISP currently being written is not covered and must be a WLP. Therefore, when a stroke is being written, a WLP is usually an ISP on the stroke. So we can use *WS* instead of *IS* to recover the stroke dynamic information.

**(b) Stroke dynamic information descriptions**

Unlike previous methods that try to recover dynamic information for each pixel, we directly recover the dynamic information for each character stroke extracted by the *Static Information Analysis* module. We focus on the evolution of a character stroke $S$ obtained by checking each frame of the corresponding binary video sequence as shown in Fig. 1(c).

We first build a switch curve, denoted by SwitchCurve or $sc(t)$, to describe the evolution of stroke $S$ in Definition 1.

**Definition 1:** The SwitchCurve is defined as the percentage of WLP on stroke $S$,

$$sc(t) = \frac{n_s(t)}{N_{str}} ,$$

where $n_s(t)$ is the number of WLPs on stroke $S$ in frame $t$ and $N_{str}$ is the total skeleton pixel number of $S$.

When a stroke is written, the value of the SwitchCurve gradually increases from 0 to 1. After the



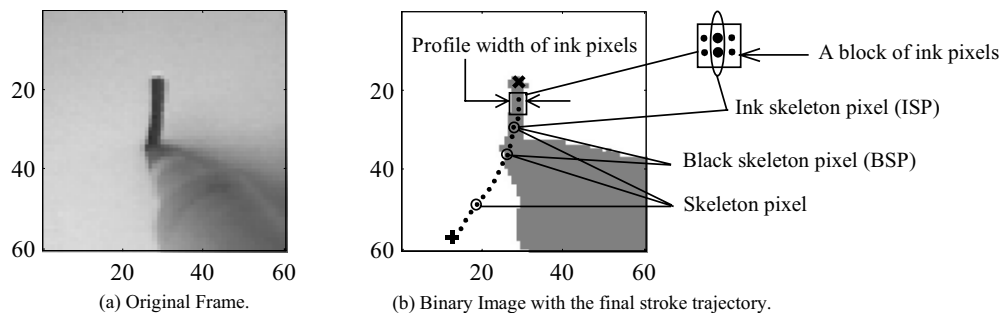(a) Original Frame.  (b) Binary Image with the final stroke trajectory.

Fig. 2. Dynamic clue in image sequence.

stroke is finished, the SwitchCurve may still fluctuate due to the variation of the WLP number caused by the shadow noise, as illustrated by the dotted curve in Fig. 3(a). Therefore, only the first monotonically ascending curve corresponds to the stroke's drawing slot, i.e., the time slot between pen-down and pen-up.

The run direction of stroke $S$ is another kind of dynamic information we need to recover. As illuminated in Fig. 2(b), the stroke writing direction, i.e., run direction, is from 'x' to '+'. However, the static analysis result may either be in the same direction or the opposite direction of the writing direction. Therefore, to recover the actual writing direction of the statically tracked stroke, we need to identify these two possibilities. Toward this, we design a centroid shift curve, denoted by CentrShiftCurve or $csc(t)$.

**Definition 2:** The CentroidShiftCurve is defined as

$$csc(t) = \frac{n_c(t)}{N_{str}+1},$$

where $n_c(t)$ is the centroid position of all WLPs on stroke $S$ in frame $t$.

If the WLPs are the first $n$ skeleton pixels of stroke $S$ according to the order of static stroke analysis in frame $t$, their centroid location is $n_c(t) = \frac{n+1}{2}$, and thus we have $csc(t) = \frac{n+1}{2(N_{str}+1)}$. After a stroke is completely drawn, with $n = N_{str}$, $csc(t)$ becomes 0.5. Therefore the locus of the centroid will shift from 0 to 0.5 in the stroke's drawing slot. Similarly, if the WLPs are assembled in the reverse order of static analysis, the track of the centroid will be from 1 to 0.5.

The two typical styles of $csc(t)$ are shown in Fig. 3(b). Fig. 3(b.1) indicates the skeleton pixels are written down in the same direction of static stroke analysis,



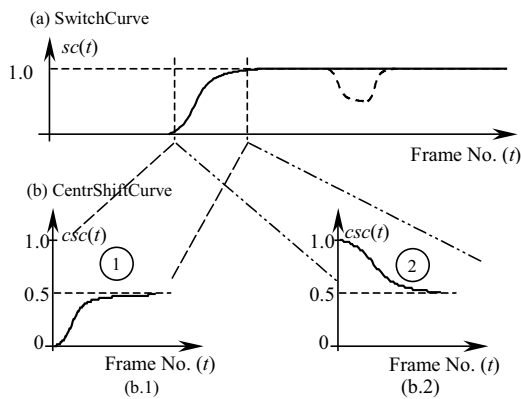(a) SwitchCurve

(b) CentrShiftCurve

Fig. 3. Dynamic information analysis in the mathematical model.

while Fig. 3(b.2) means the opposite. The run direction of a stroke can thus be decided.

### 3.3. Stroke's Drawing Slot Analysis

A stroke's SwitchCurve can be employed to detect its drawing slot through the following steps.

Step I. Locate the midst point $T_{mid}$, where the SwitchCurve passes through 0.5 for the first time, i.e. roughly half of the stroke has been drawn.

Step II. Search backward from $T_{mid}$ to find the point $T_{start}$ where the SwitchCurve begins rising, i.e. the pen-down moment.

Step III. Search forward from $T_{mid}$ to find the point $T_{end}$ where the SwitchCurve stops ascending, i.e. the pen-up moment.

Therefore, $T_{mid}$, $T_{start}$, and $T_{end}$ are the feature points of the SwitchCurve. First of all, $T_{mid}$ must be accurately determined. Since WLP is sensitive to the shadow noise, the SwitchCurve may fluctuate before pen-down moment or after pen-up moment. To alleviate the problem, an auxiliary curve, stroke mask curve, denoted by MaskCurve or $mc(t)$, is introduced in Definition 3.

**Definition 3:** The MaskCurve is defined as, the percentage of BSP on stroke $S$,

$$mc(t) = \frac{n_m(t)}{N_{str}},$$

where $n_m(t)$ is the number of BSPs on stroke $S$ in frame $t$.

Notice that $mc(t)$ counts all the black skeleton pixels in stroke $S$, no matter whether they are due to acutual writing or shadow pixels. Obviously, after a stroke is completely written down, its number of BPSs will remain to be $N_{str}$, i.e., MaskCurve will stay at 1, as the solid line shown in Fig. 4(a). From the MaskCurve, we can roughly estimate the half-past-stroke moment, $T_{mkmid}$, and its starting moment, $T_{mkstart}$. Then the pen-down moment $T_{start}$ and the pen-up moment $T_{end}$ can be detected by the following five steps.

Step 1. $T_{mkmid} = T_1 + 1$, where $T_1$ is the last point below 0.5 in the MaskCurve.

Step 2. $T_{mkstart} = T_2 + 1$, where $T_2$ is the first point below $1/N_{str}$ backward from $T_{mkmid}$ in the MaskCurve.

Since a WLP is always a BSP, but a BSP may not be a WLP, the MaskCurve is always higher than a SwitchCurve. Therefore, we have the following two inequalities.

$$T_{mid} \geq T_{mkmid} , \qquad (3)$$

$$T_{mkstar} \leq T_{start} . \qquad (4)$$

Using $T_{mkmid}$ and $T_{mkstart}$ as references, with inequality (3) and (4), we can locate the feature points $T_{mid}$ and $T_{start}$ on the SwitchCurve.

Step 3. Searching forward from $T_{mkmid}$, we find the first point above 0.5 in the SwitchCurve as $T_{mid}$.
Step 4. $T_{start} = T_3 + 1$, $T_3$ is the last point below $1/N_{str}$ in the SwitchCurve in the region [$T_{mkstart}$, $T_{mid}$].

To locate the point $T_{end}$, we define a new smooth curve f-SwitchCurve as follows.

**Definition 4:** A f-SwitchCurve is a SwitchCurve smoothed by a Gaussian filter.

We can locate $T_{end}$ by Step 5.

Step 5. Forward from $T_{mid}$, the first local maximum point on the f-SwitchCurve is $T_{end}$.

The relationship of SwitchCurve, MaskCurve, f-SwitchCurve, and their feature points are depicted in Fig. 4(a) and the corresponding CentrShiftCurve is shown in Fig. 4(b).

We use the mean value of the frame numbers within the stroke-drawing slot to represent the writing time of the stroke. Ranking the writing times of all strokes, we can recover the final stroke writing order in the character.

## 4. Experiments

To test our recognition system, ten students were invited to write Chinese characters. Their handwritings were captured by a video camera in real time. Totally 3039 character sequences were acquired. We extracted 18581 strokes from the data set, larger than most traditional on-line recognition studies. Fig. 5 shows some final results of the stroke extraction, where both the run direction and drawing order of each stroke are labelled by our algorithm automatically and correctly. About 99% of the 18581 strokes were obtained correctly. The result clearly demonstrates the robustness of our algorithm.

In the experiment, most of the dynamic analysis errors are due to incorrect estimation of the stroke run direction.

To evaluate our character stroke extraction results, a character recognition experiment is conducted. Since our stroke extraction results have all the information required by an on-line recognition system, a traditional on-line method could be directly applied. Here we adopt the attributed relational graph (ARG) matching method [7] for our recognition experiment. In the experiment, 2947 out of 3039 characters (97% of the whole dataset) can be recognized with top 5 selections. Among the errors, only 9 characters (0.3% of the whole dataset) were incorrectly recognized because of our stroke extraction algorithm (including static and dynamic analysis). Others were caused by the limitation of the ARG matching algorithm. The results show that the performance of our VCR system is comparable to that of the traditional on-line system.
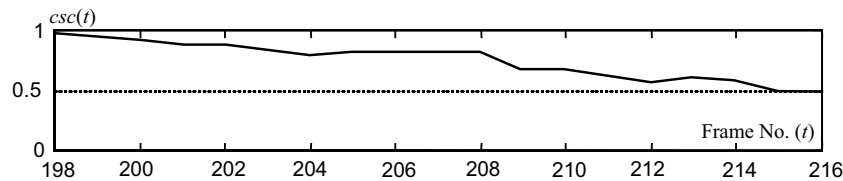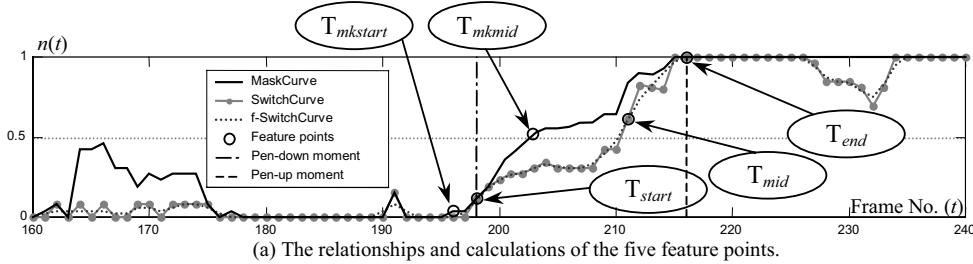


(a) The relationships and calculations of the five feature points.

(b) The CentrShiftCurve in the corresponding pen-down to pen-up timeslot.

Fig. 4. The drawing slot analysis on a character stroke.

## 5. Conclusion

In this paper, we present a novel scheme for the video-based handwritten Chinese character recognition system, the crux of which is how to robustly recover the character stroke dynamic information from the video data. An effective approach based on stroke analysis is proposed to accurately extract the stroke dynamic information. Our experiment on a large data set shows that the system can retrieve the dynamic information of Chinese characters efficiently. About 99% of the strokes were extracted correctly, and 97% of recognition accuracy was reached by using the ARG matching method, which is comparable to the performance of the on-line system.

## Acknowledgments

## References

[1] R. Plamondon and S. N. Srihari, "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey," *IEEE Trans. PAMI*, vol. 22, no. 1, pp. 63-84, Jan. 2000.

[2] N. Arica and F. T. Yarman-Vural, "An Overview of Character Recognition Focused on Off-Line Handwriting," *IEEE Trans. Syst. Man. Cybern. Part C*, vol. 31, no. 2, pp. 216-233, May 2001.

[3] S. Madhvanath and V. Govindaraju, "The Role of Holistic Paradigms in Handwritten Word Recognition," *IEEE Trans. PAMI*, vol. 23, no. 2, pp. 149-164, Feb. 2001.

[4] N. Arica and F. T. Yarman-Vural, "Optical Character Recognition for Cursive Handwriting," *IEEE Trans. PAMI*, vol. 24, no. 6, pp. 801-813, Jun. 2002.

[5] C. C. Tappert, C. Y. Suen and T. Wakahara, "The State of the Art in On-Line Handwriting Recognition," *IEEE Trans. PAMI*, vol. 12, no. 8, pp. 787-808, Aug. 1990.

[6] S. D. Connell and A. K. Jain, "Writer Adaptation for Online Handwriting Recognition," *IEEE Trans. PAMI*, vol. 24, no. 3, pp. 329-346, Mar. 2002.

[7] J. Liu, W. K. Cham and M. M. Y. Chang, "Online Chinese character recognition using attributed relational graph matching," *IEE Proc. Vision Image Signal Process.*, vol. 143, no. 2. pp. 125-131, 1996.

[8] M. E. Munich and P. Perona, "Visual Identification by Signature Tracking," *IEEE Trans. PAMI*, vol. 25, no. 2, pp. 200-217, Feb. 2003.

[9] X. Tang, C. Yung and J. Liu, "Handwritten Chinese character recognition through a video camera," *Proc. Int'l Conf. Image Processing*, pp. 692-695, 2000.

[10] X. Tang and F. Lin, "Video-based handwritten character recognition," *Proc. Int'l Conf. Acoustics, Speech, and Signal Processing*, pp. 3748 -3751, 2002.

[11] F. Lin and X. Tang, "Off-line handwritten Chinese character stroke extraction," *Proc. Int'l Conf. Pattern Recognition*, pp.249 –252, 2002.

(a) The original character frames.
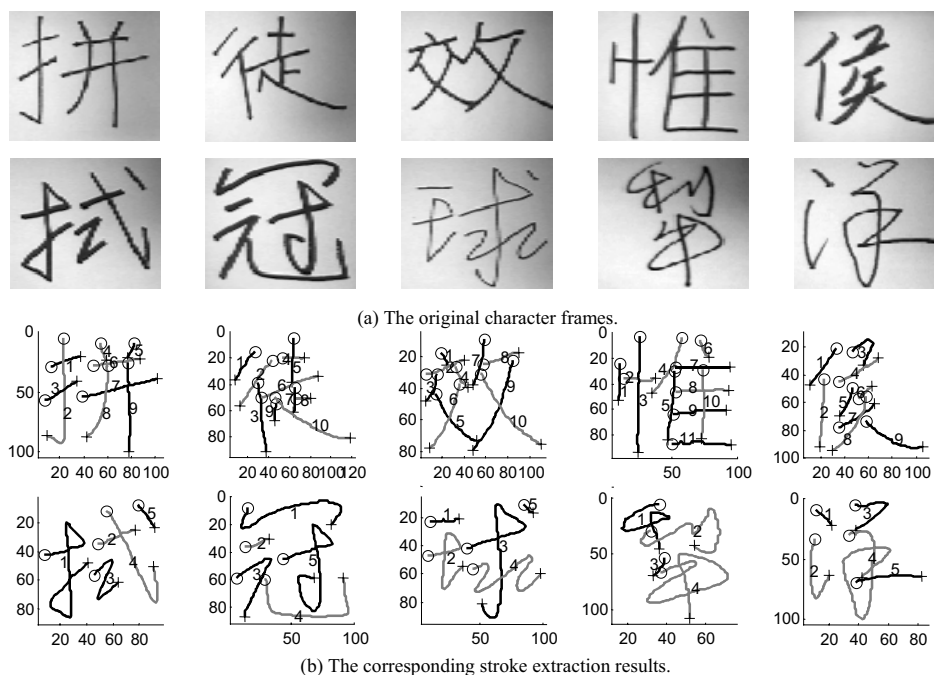


(b) The corresponding stroke extraction results.

Fig.5. Some stroke extraction results and the original character frames.