AUTOMATIC HANDWRITING DETECTION
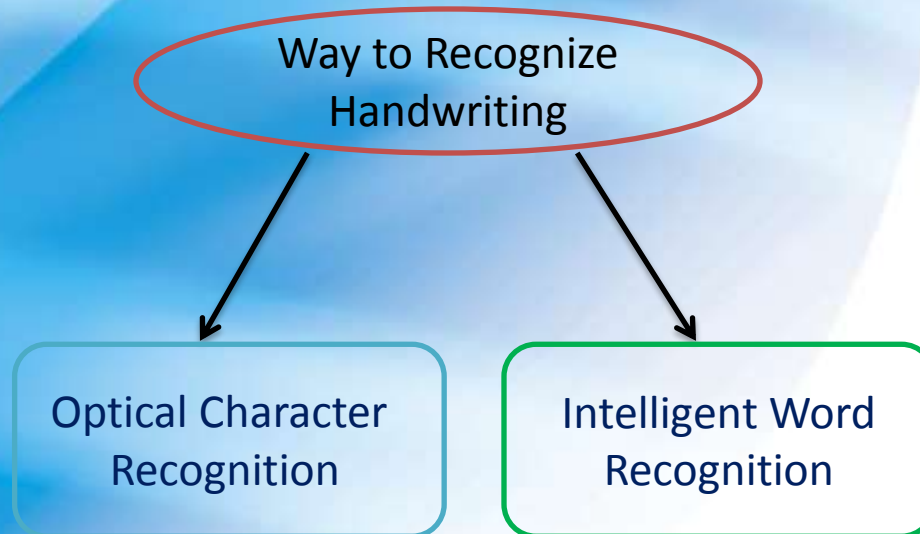
Presentation By:
Bijit Ghosh

Heritage Institute of Technology
Dept. of Computer Science & Engg.

What is Handwriting Detection?

Handwriting Detection is a technique or ability of a Computer to receive and interpret intelligible handwritten input
from source such as paper documents, touch screen, photo graphs etc.

Way to Recognize Handwriting

Optical Character Recognition

Intelligent Word Recognition

Example of Handwriting Images

How it Works?

Query Sample → Identification System → Writer 1

Identification System → Writer n

Identification System ↔ Database with samples of known authorship

Online Recognition

Offline Recognition

Online Recognition

On-line handwriting recognition involves the automatic conversion of text as it is written on a special digitizer or PDA, where a sensor picks up the pen-tip movements as well as pen-up/pen-down switching. That kind of data is known as digital ink and can be regarded as a dynamic representation of handwriting. The obtained signal is converted into letter codes which are usable within computer and text-processing applications. The elements of an on-line handwriting recognition interface typically include:

1) a pen or stylus for the user to write with.

2) a touch sensitive surface, which may be integrated with, or adjacent to, an output display.

3) a software application which interprets the movements of the stylus across the writing surface, translating the resulting strokes into digital text.

Offline Recognition

Off-line handwriting recognition involves the automatic conversion of text in an image into letter codes which are usable within computer and text-processing applications. The data obtained by this form is regarded as a static representation of handwriting. Off-line handwriting recognition is comparatively difficult, as different people have different handwriting styles. And, as of today, OCR engines are primarily focused on machine printed text and ICR for hand "printed" (written in capital letters) text. There is no OCR/ICR engine that supports handwriting recognition as of today.

Methods and Strategies

- Recognition strategies heavily depends on <u>the nature of the data</u> to be recognized.

- In the cursive case, the problem is made complex by the fact that the writing is fundamentally ambiguous as the letters in the word are generally linked together, poorly written and may even be missing.

- On the contrary, hand printed word recognition is more related to printed word recognition, the individual letters composing the word being usually much easier to isolate and to identify.

Character Recognition

- Character Recognition techniques can be classified according to two criteria:
  - the way preprocessing is performed on the data
  - the type of the decision algorithm
- Preprocessing techniques include :
  - the use of global transforms (correlation, Fourier descriptors, etc.)
  - local comparison (local density, intersections with straight lines, variable masks, etc.)
  - geometrical or topological characteristics (strokes, loops, openings, diacritical marks, skeleton, etc.)
- Decision methods include:
  - various statistical methods,
  - neural networks, structural matching (on trees, chains, etc.)
  - stochastic processing (Markov chains, etc.).

Word Recognition

- Two main types of strategies have been applied to this problem:
  - the holistic approach - recognition is globally performed on the whole representation of words and there is no attempt to identify characters individually.
    - The main advantage of holistic methods is that they avoid word segmentation
  - the analytical approach - deal with several levels of representation corresponding to increasing levels of abstraction (usually the feature level, the grapheme or pseudo-letter level and the word level). Words are not considered as a whole, but as sequences of smaller size units which must be easily related to characters in order to make recognition independent from a specific vocabulary

Feature Extraction

Feature extraction works in a similar fashion to neural network recognizers however, programmers must manually determine the properties they feel are important.
Some example properties might be:

- Aspect Ratio
- Percent of pixels above horizontal half point
- Percent of pixels to right of vertical half point
- Number of strokes
- Average distance from image center
- Is reflected y axis
- Is reflected x axis

An Example: IBM TransNote
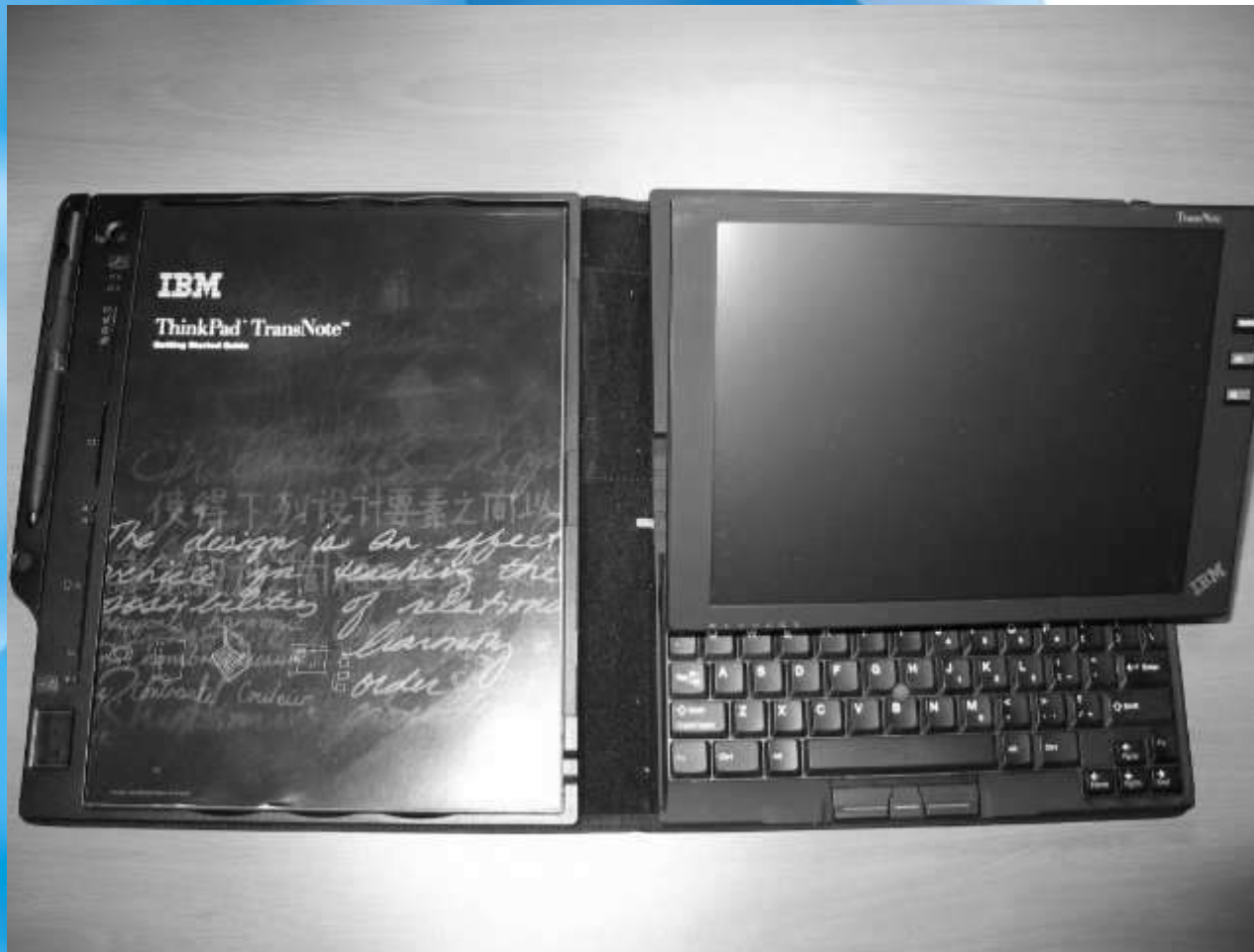
Handwriting Segmentation Algorithm

(A)   (B)   (C)

Different stages of handwriting segmentation.

**IBM TransNote works based
on two Features...
1) Handwriting Speed
2) Handwriting Wrinkliness**

Images of authentic handwriting sample.

Statistical Experiments

A sample of the calculated measurements.

| Filename	300dpi	600dpi	Wrinkliness	Speed
0101T1	14894	30583	1.03799867	0.11396973
0101T2	8786	18638	1.084968652	0.107457204
0101T3	9258	19764	1.094102493	0.118184103
0202T1	6453	13765	1.092962679	0.093275242
0202T2	6212	13319	1.100356033	0.094080635
0202T3	5824	12722	1.127243231	0.087968122
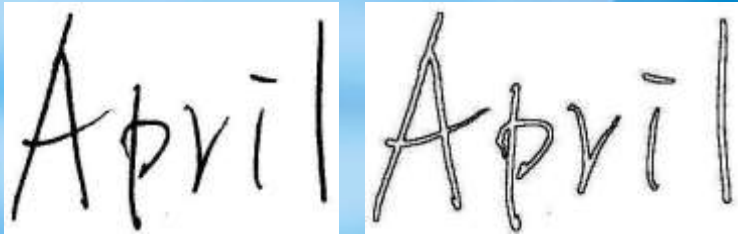
Wrinkliness calculation: a) handwriting
sample, b) edge-detected sample, c) portion of edgedetected
sample at 300 dpi, d) portion at 600 dpi.

Constraints

- Handwriting Recognition aims to design systems which are able to recognize handwriting of natural language

- Methods and recognition rates depend on the level of constraints on handwriting.

- The constraints are mainly characterized by the:
  - types of handwriting
  - number of scriptors
  - size of the vocabulary
  - spatial layout.

Some aspects of Handwriting Recognition

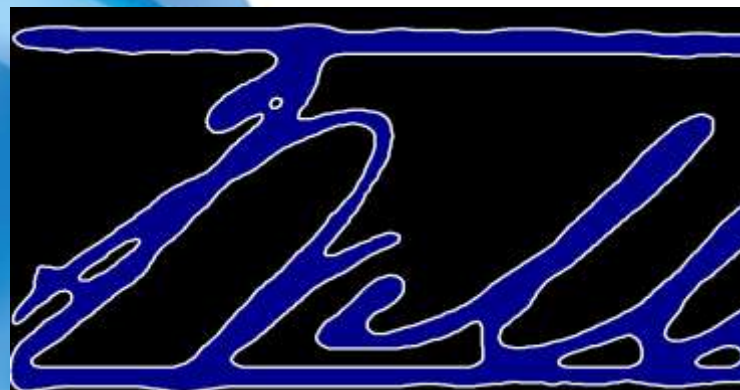*Segmentation problem (can't read word until it is segmented; can't segment word until it is read)*

nr?

m?
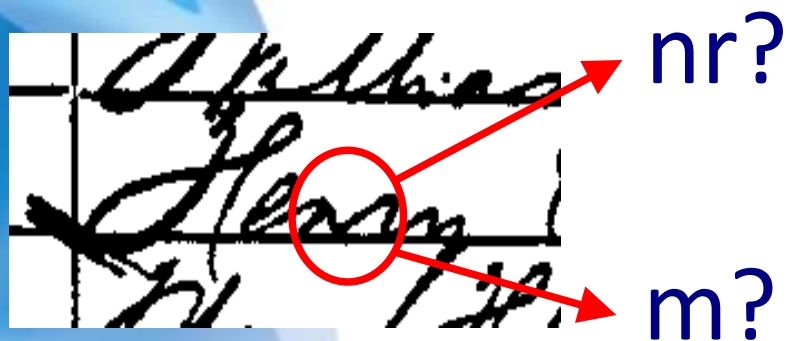
— *Different handwriting styles*

— *Use of dictionary to correct for errors in reading*

Srnitb --> Smith
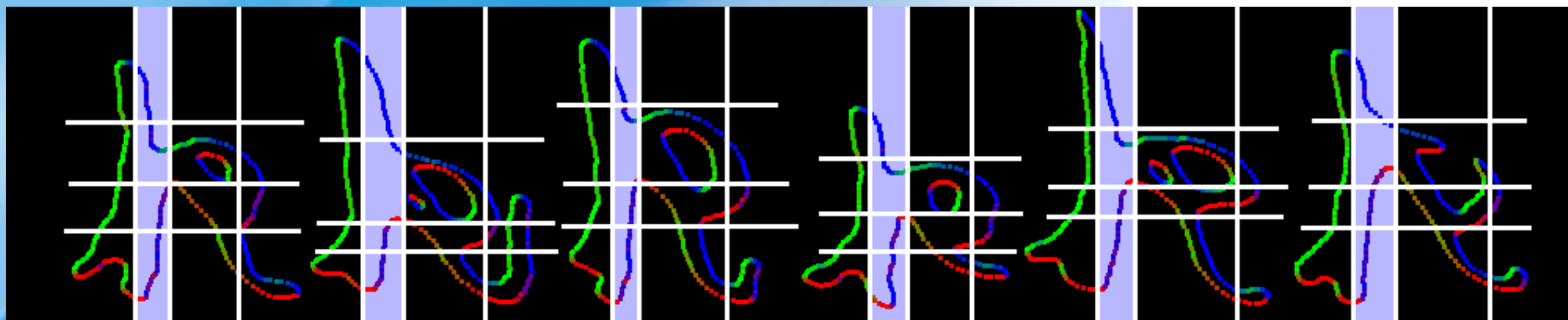
Outlines of word are traced and smoothed:

Handwriting slope is corrected for automatically:

Segmentation

- Goal: robustly cut letters into segments
- Match multiple segments to detect letters
- Easier than matching whole letter

Advantages & Disadvantages

ADVANTAGES:
The most important advantage of speech
over handwriting is the speed of data entry. This is because it is much easier to dictate the
machine than to write

DISADVANTAGES:
Has also drawbacks, such as it is noisy to hear
someone sitting next to us and talking to his machine. Moreover, anyone who wants to input
confidential data to his/her computer is not willing to do it in public places. Most importantly, it
is not possible to speak to a machine in a natural way due to constraints such as out of vocabulary
wards, background noise, cross-talk, accented speech and so on.

Conclusion

So after all these stuff...
We have a brief Idea about Automatic Handwriting Detection..
1) online and offline detection is available
2)Online procedure is easier then Offline procedure(difficult one)
3)Today's business world need some computerized authentication for security purpose
the (AHD) fulfill their need.
4)Handwriting recognition is important for genealogy...
...but it is hard
5)Current methods don't work very well...
...and they don't operate much like the human brain

Thank You..