

Mobile Interactive Support System for Time-Critical Document Exploitation

*George Nagy*¹

`nagy@ecse.rpi.edu`

*Daniel Lopresti*²

`lopresti@cse.lehigh.edu`

¹ Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180

² Computer Science and Engineering, Lehigh University, Bethlehem, PA 18015

Pre-Acknowledgment

Thoughts that follow draw heavily on work by Professor George Nagy and his former students at RPI: Dr. Jie Zou, Haimei Jiang, Abhishek Gattani, Borjan Gagoski, Greenie Chang, and Laura Derby.

- "Interactive Visual System," A. Evans, J. Sikorski, P. Thomas, J. Zou, G. Nagy, S.-H. Cha, and C. C. Tappert, Technical Report CSIS NO. 196, Pace University, 2003.
- *Computer Assisted Visual InterActive Recognition*, J. Zou, PhD Thesis, Rensselaer Polytechnic Institute, 2004.
- "Evaluation of Model-Based Interactive Flower Recognition," J. Zou and G. Nagy, *Proceedings of the 17th International Conference on Pattern Recognition*, August 2004.
- "Computer Assisted Visual Interactive Recognition (CAVIAR) Technology," S-H. Cha, A. Evans, A. Gattani, G. Nagy, J. Sikorski, C. Tappert, P. Thomas, and J. Zou, *Proceedings of the IEEE Electro/Information Technology Conference*, May 2005.

Premise

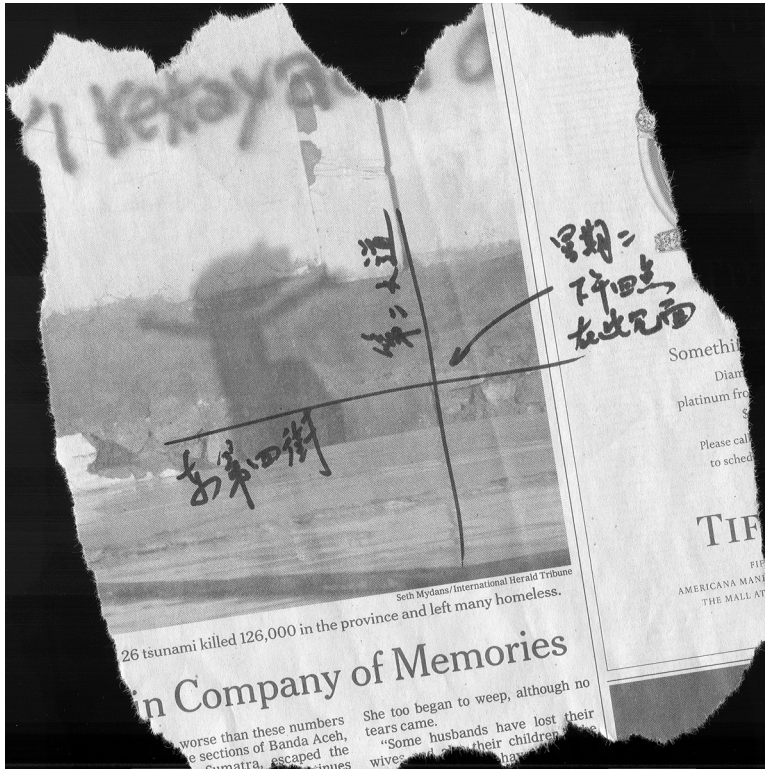
Fully automating the analysis of document images in the most general sense is still an unrealized goal.

Some important applications, though, may not require fully automated analysis:

- Focus is on small number of high-value documents.
- Fully manual processing possible, but too slow.
- Computer and human combine to form “team.”
- Measure of success is to be faster than human alone, more accurate than machine alone.

Scenario

Imagine finding a note scribbled on a discarded scrap of newspaper lying on the street of a large US city ...

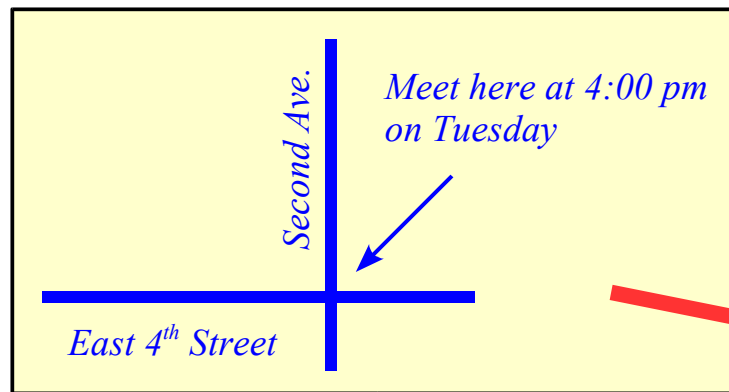


What can be gleaned from a simple source such as this?

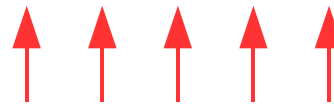
Scenario

Given context:

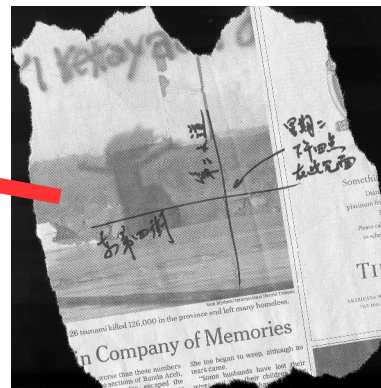
- Map likely NY City
- Tuesday likely April 12th



Chinese to English translation

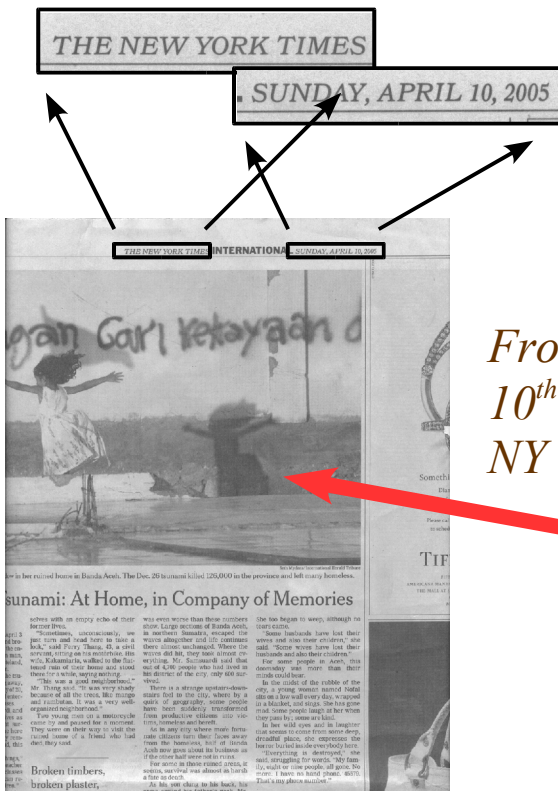


From April 10th issue of NY Times



Also:

- ID of writer
- Other notes authored by same writer



NY Times 4/10/05

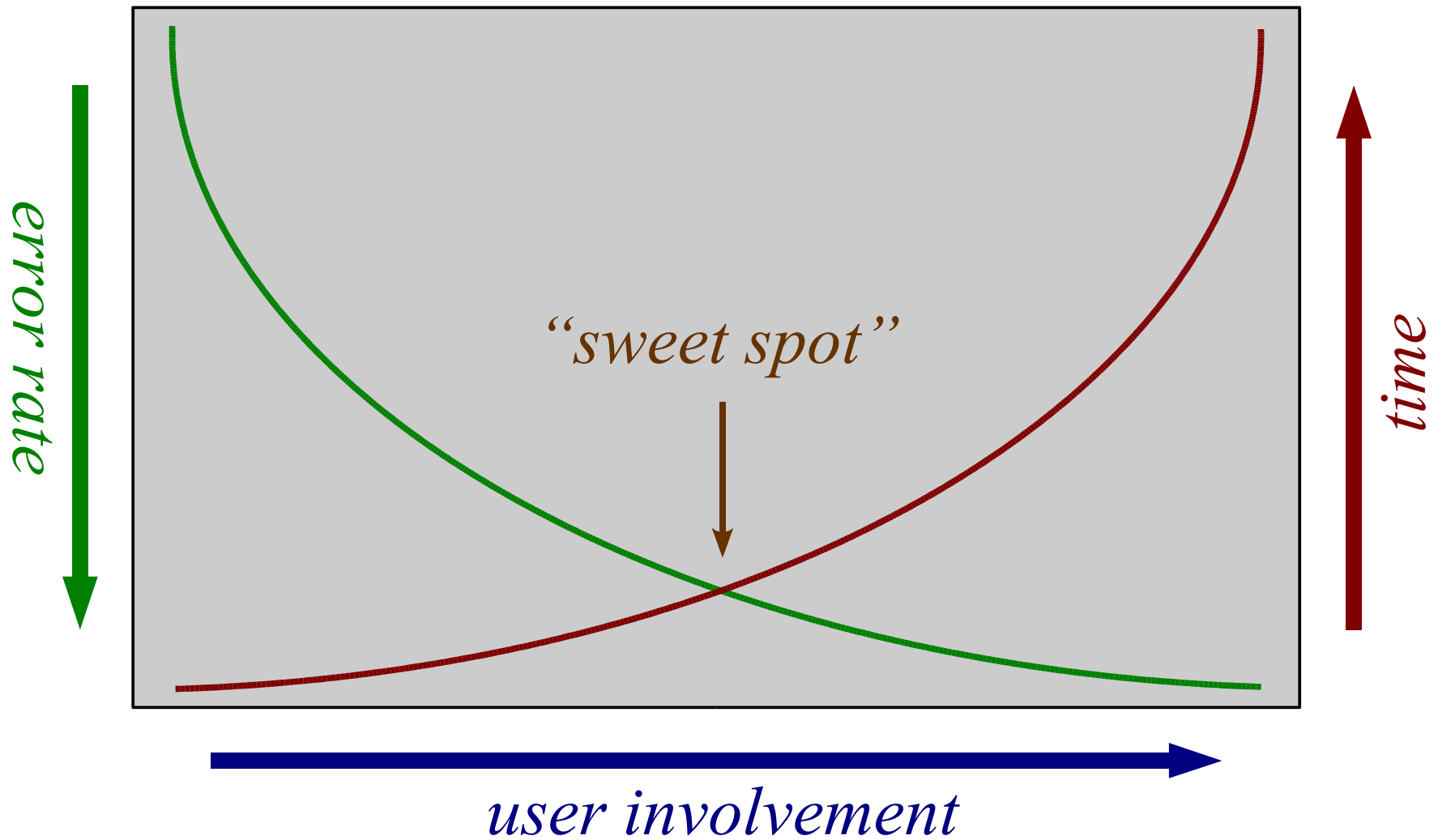
Map of lower Manhattan

Mobile Interactive Document Analysis

- A concept: untested to date in realm of documents.
- Motivated by work on *Computer Assisted Visual Interactive Recognition* by Nagy, et al.
- Mobile CAVIAR runs on camera-equipped PDA at RPI.
- Task is flower classification.
- Direct manipulation of abstract model mediates human-computer interaction.
- Semi-automated recognition.



What CAVIAR Strives For



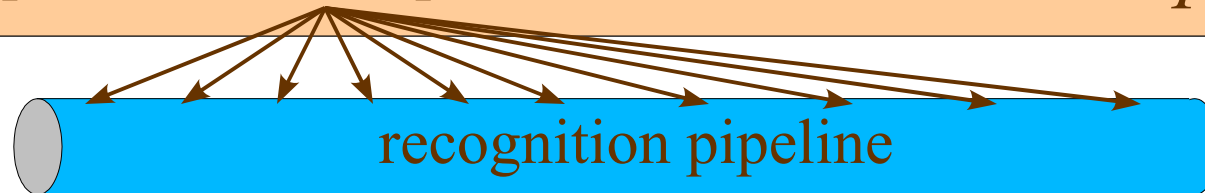
Getting the User Involved

Traditional approach (current OCR systems):

- User interacts with recognition process at very beginning (to set thresholds, deskew page, etc.) or at very end (to proofread output and correct errors).
- Interfaces are complex – rarely interactive.
- Input given early is done blindly. Input given later can't be fed back to improve algorithm.

potential user input here

CAVIAR philosophy



Human vs. Machine Capabilities

Humans employ gestalt perception and can:

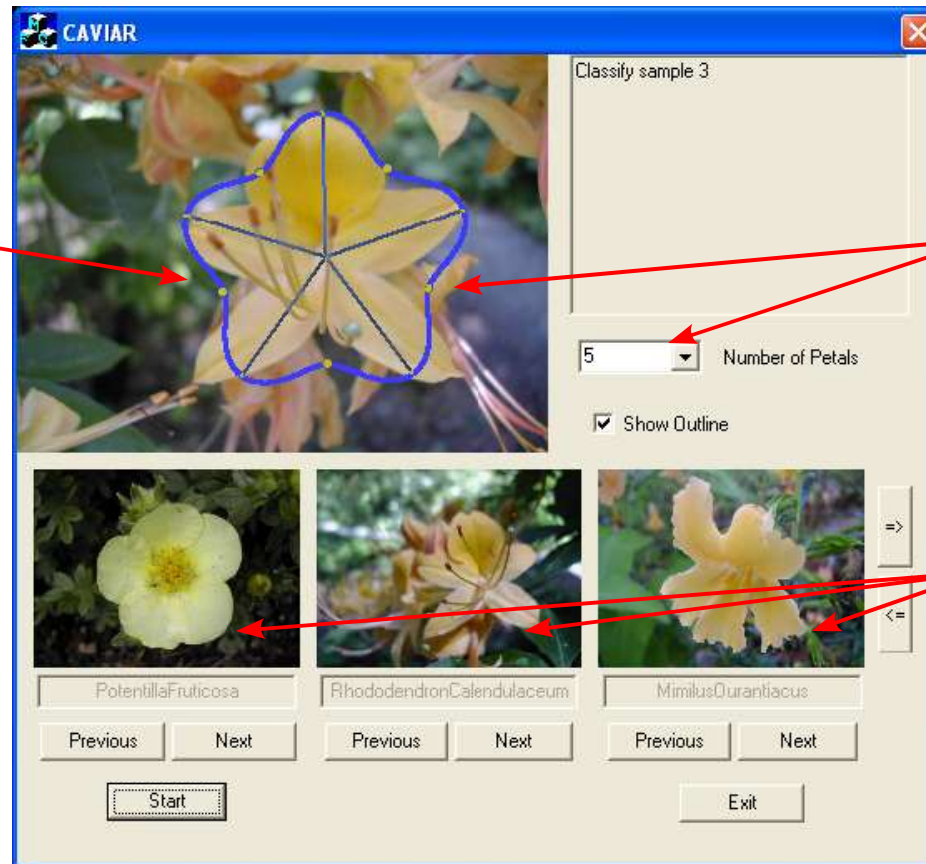
- segment objects from background,
- recognize broad context and bring it to bear,
- filter out correlated noise,
- judge similarity based on shape, color, and texture.

Computers train on enormous datasets and can:

- compute geometrical moments, spatial frequencies, topological properties, multivariate parameter estimates, posterior probabilities, etc.

CAVIAR for Flowers

*Rose curve
model for flower
segmentation*



*Model can be
adjusted via
stylus or mouse*

*CAVIAR
presents top
matches, user
browses to select*

Direct manipulation of model is key – avoids forcing user to deal in “magic” numbers which are unintuitive.

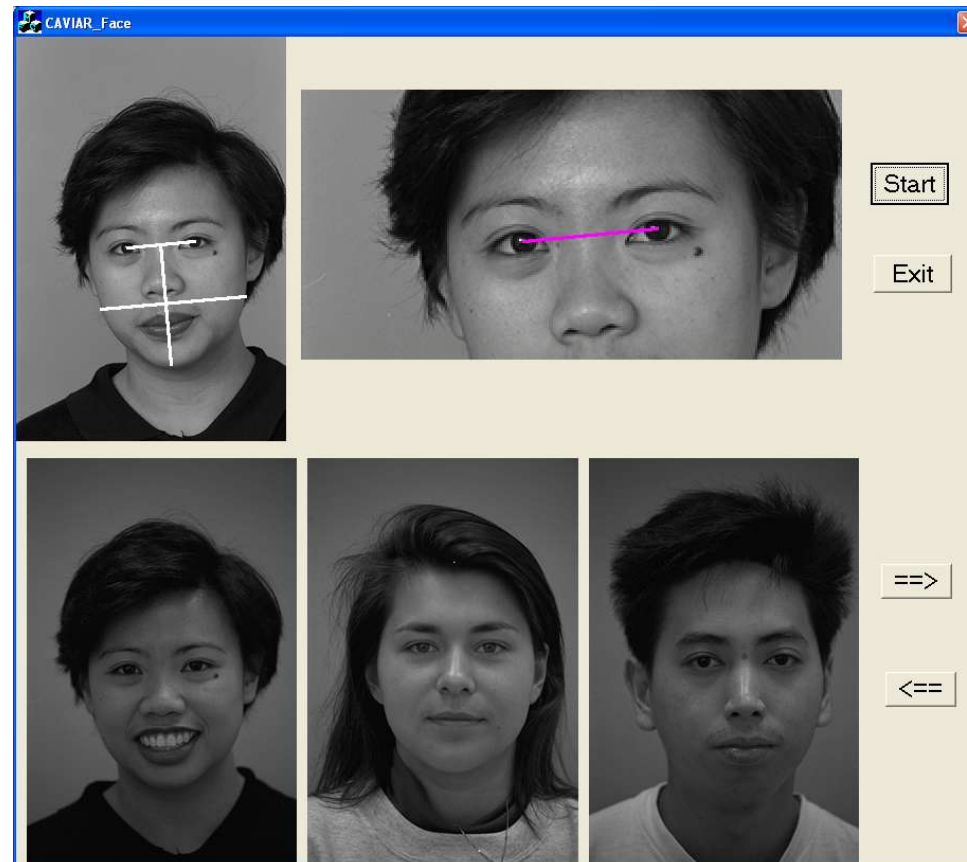
CAVIAR for Flowers

Evaluation of CAVIAR for flowers performed at RPI:

- 30 human subjects
- 612 flower photos / 102 species

	<i>Accuracy</i>	<i>Time / flower</i>
<i>Interactive</i>	93% (83% – 99%)	12 secs. (7 secs. – 27 secs.)
<i>Machine Alone</i>	32% (24% – 50%)	~0 secs.
<i>Human Alone</i>	93% (91% – 97%)	26 secs. (18 secs. – 36 secs.)

CAVIAR for Faces



Similar work done at RPI for face recognition. In this case, GUI designed for accurate pupil location.

CAVIAR for Faces

Evaluation of CAVIAR for faces performed at RPI:

- 6 human subjects
- 200 face photos as gallery / 50 face photos as probes

	<i>Accuracy</i>	<i>Time / face</i>
<i>Interactive</i>	99.7%	7.6 secs.
<i>Machine Alone</i>	47.0%	~0 secs.
<i>Human Alone</i>	100%	66.3 secs.

CAVIAR for Documents

While there are many existing interactive packages that perform document analysis functions, and even mobile systems for capturing images of, say, street signs and translating the scene text located therein, the CAVIAR approach appears relatively unique.

- Direct manipulation requires a realizable model.
- System must respond to updates interactively.
- To be a “win,” application must go beyond readily discernible text (manual data entry is always an option). This makes developing testing scenarios a challenge (CAVIAR tests must include human users).

Technological Issues

Reasons to be optimistic about prospects for mobile interactive document analysis:

- Rapid development of high-quality, low cost digital cameras (now nearly ubiquitous in cell phones).
- Increasing screen resolutions, storage capacities, and CPU power of PDA's and other mobile devices.
- Widespread consumer usage of messaging services has led to improved text entry techniques.
- Need for network connectivity depends on application, but growing availability and rising bandwidths portend well.

Opportunities for Interaction

Binarization / thresholding:

- Many automated techniques presume some known distribution of foreground/background pixels.
- User could define window using “point-and-drag.”

Text-line finding / skew detection:

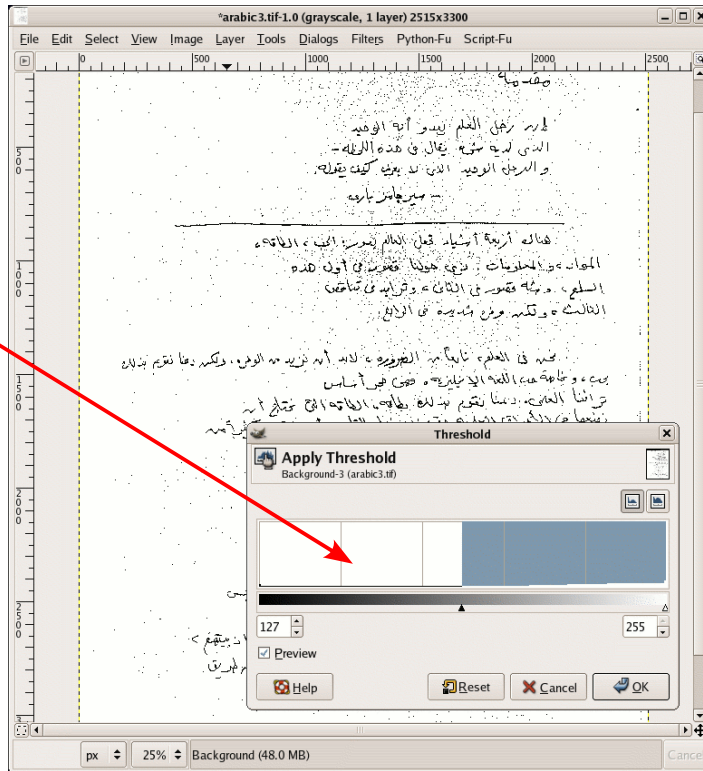
- Can be challenging, especially for handwriting.
- User could correct by rotating baselines as needed.

Word / character segmentation:

- Again, handwriting especially difficult here.
- User could apply broad range of lexical knowledge.

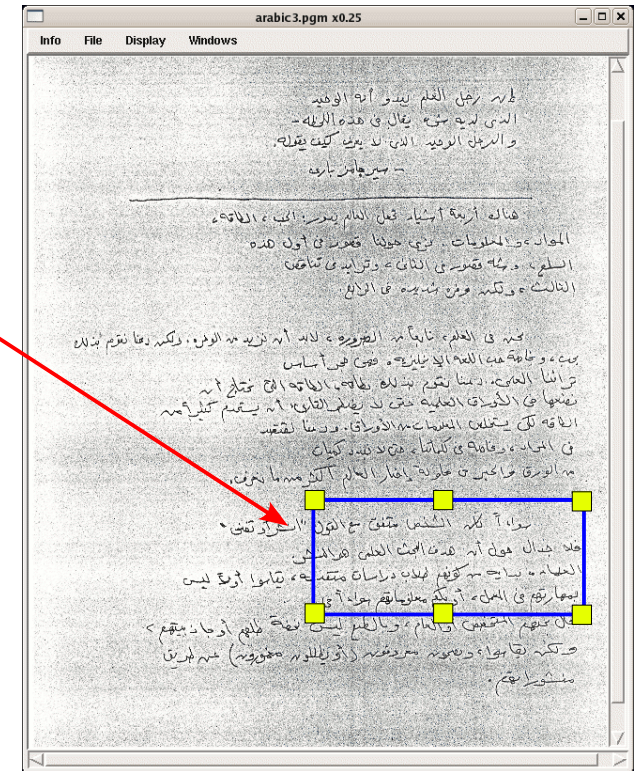
Interaction Opportunities: Binarization

Histogram with slider



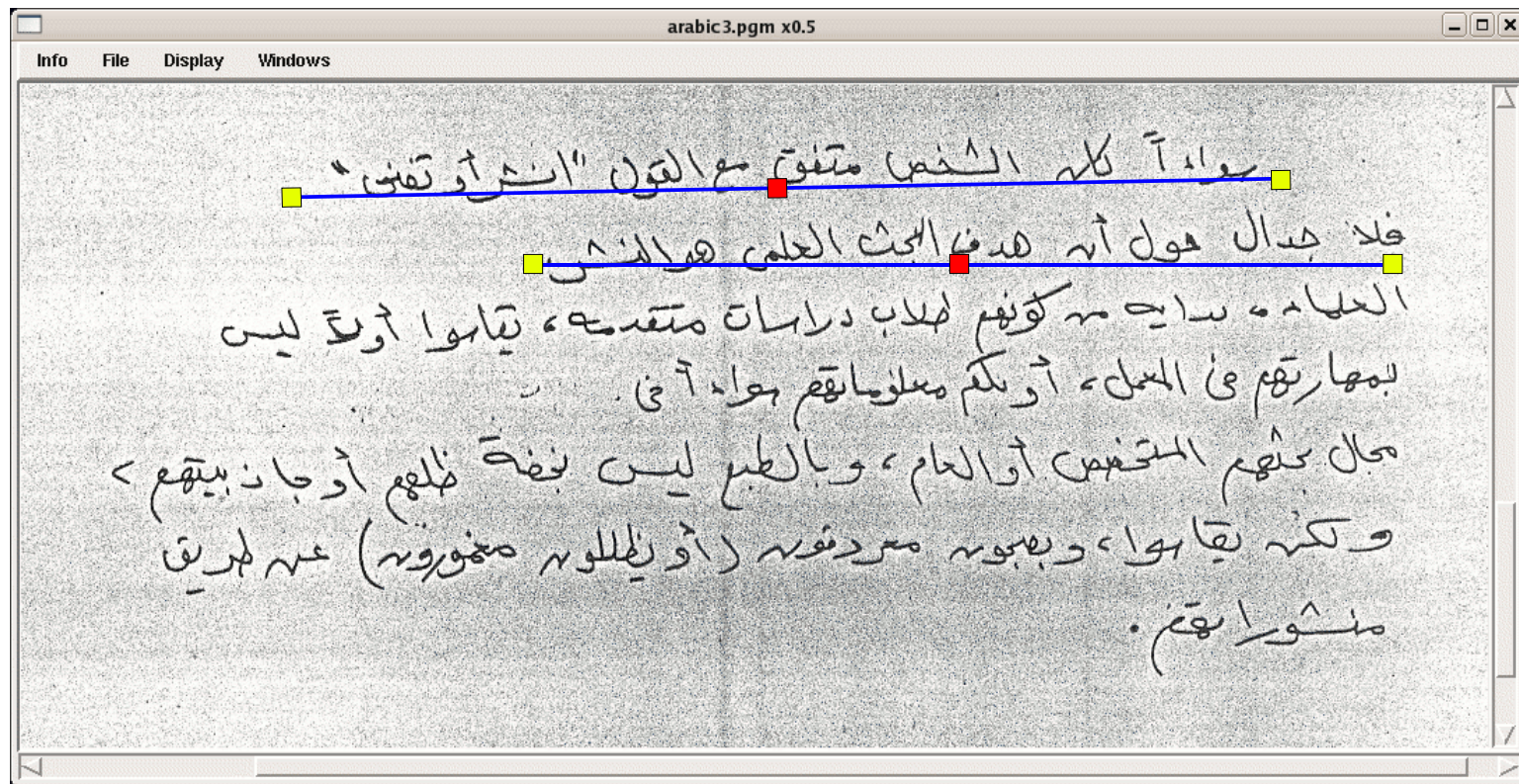
Traditional approach ("gimp")

User sketches window with both text and background



CAVIAR approach

Interaction Opportunities: Deskewing



Conclusions

- Interactive recognition is faster than unaided human, and more accurate than unaided machine.
- Parsimonious interaction throughout process is better than only at beginning or end.
- Interactive systems can be initialized with single training sample per class, and improve with use.
- CAVIAR interaction paradigm requires visible model that is accessible to both human and machine. Finding the right “widget” is very important.
- Applications to document analysis worth examining.