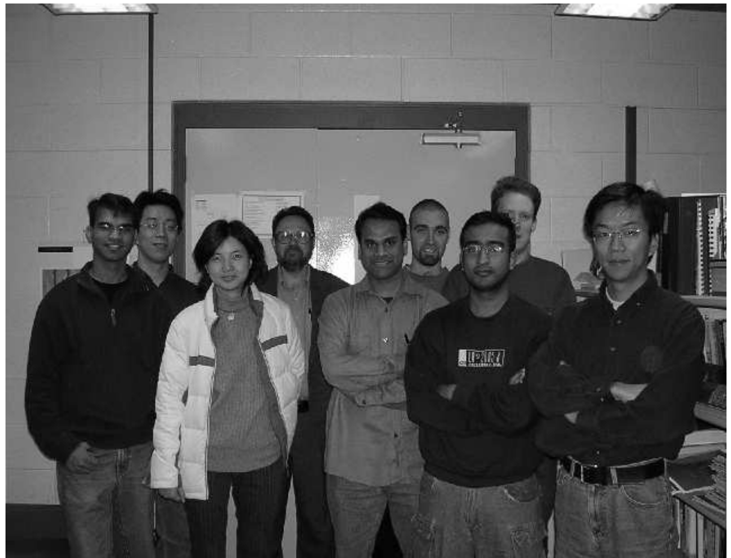
INTERACTIVE, MOBILE, DISTRIBUTED PATTERN RECOGNITION

George Nagy
RPI DocLab

nagy@ecse.rpi.edu

Ack: ex-students **Dr. Jie Zou,**
Haimei Jiang, Abhishek Gattani,
Borjan Gagoski, Greenie Chang,
Laura Derby.



But all the mistakes are my own!

Examples of *visual* pattern recognition

Bar codes (e.g., UPC) √
Normal printed matter √
Motivated hand print √
Fingerprints √
Gross thematic maps from satellite pics √
    Industrial part and assembly inspection ?
    Military targets
    Printed matter in complex formats ?
    Degraded (faxed, copied) printed matter ?
        Sloppy or archaic handwriting
        Detailed thematic maps
        Micrographs, X-rays, skin lesions
        Faces (*lighting, pose, expression, aging*)
        Cryptic cats, birds, fish, flowers, ...

OUTLINE

- Symbolic and Natural patterns
- Interaction
- Mobile recognition
- Pattern recognition networks
- Style and context
- Applications

MESSAGE

- For *natural* patterns,
  consider *interactive recognition*, &
  make your classifiers improve with use.

- For *symbolic* patterns, use as much language and style context as possible

- Keep an eye on cell phones as the pattern recognition platform of the future

SYMBOLIC vs. NATURAL PATTERNS

Symbolic patterns (*glyphs*) evolved for human *communication,* and are therefore distinguishable*.*

However, the distinction is a continuum, not a dichotomy (consider video text, or gene sequences) .

SYMBOLIC PATTERNS

Represent natural or formal languages;

They are images of *2-D* objects
(usually *scanned,* not photographed);

Any reader of the language can perform the
classification manually;

Require high throughput because
every message consists of many patterns;

Many (millions) of samples are available for training;

SYMBOLIC PATTERNS (CONT'D)

A message is an *ordered sequence* of many glyphs: models of *context* and of *style* have been developed;

The error/reject tradeoffs are well understood;

The classes are fixed by an alphabet, syllabary, or lexicon: there are exactly 10 digits and, in Italian, 21 letters of the alphabet;

In feature space, the class centroids are located at the vertices of a regular simplex !

SOME GLYPHS

0 1 2 3 4 5 6 7 8 9

0 1 2 3 4 5 6 7 8 9

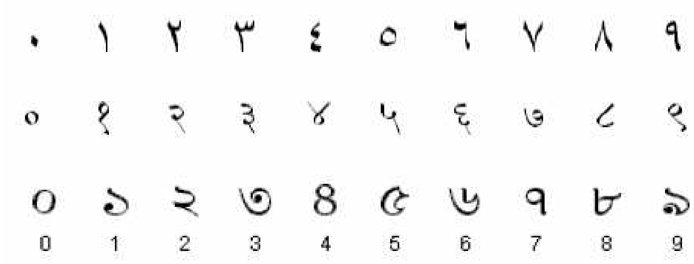
0 1 2 3 4 5 6 7 8 9

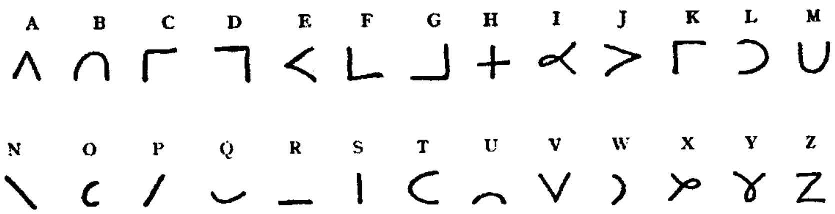
0 1 2 3 4 5 6 7 8 9

Arabic:

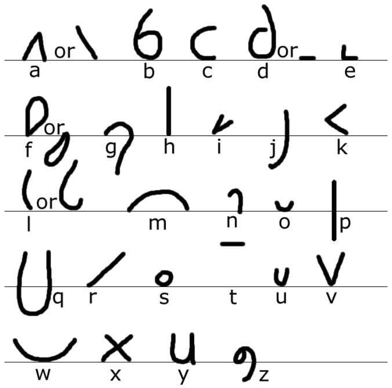
Devnagari:

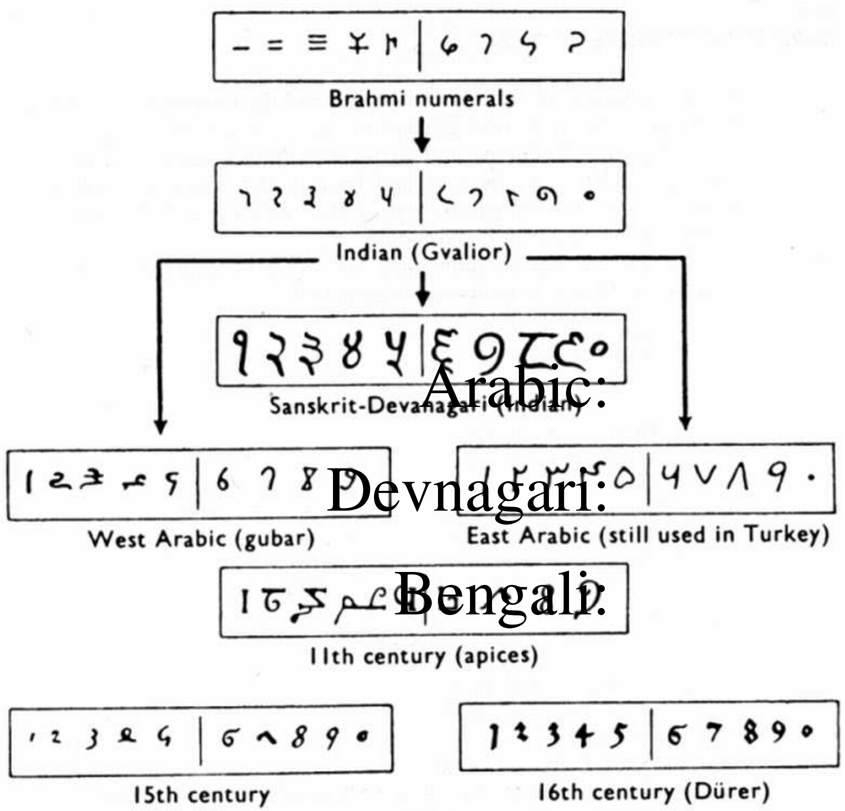
Bengali:





Shorthand symbols





Arabic:
Devnagari:
Bengali:

Figure 2.1(c) The family tree of the Indian numerals. (From K. Menninger, op. cit.)

NATURAL PATTERNS

Lack intrinsic discriminability of symbolic patterns;

Are photographed with varied
pose, expression, lighting;

Must be classified on demand
rather than as part of a work-flow;

Can be recognized only by relatively few experts
(bird-watchers, foresters, physicians);

Often have only small training sets because of the
high cost of labeling

NATURAL PATTERNS (CONT'D)

Occur in arbitrary sequence: seldom have established models of language context;

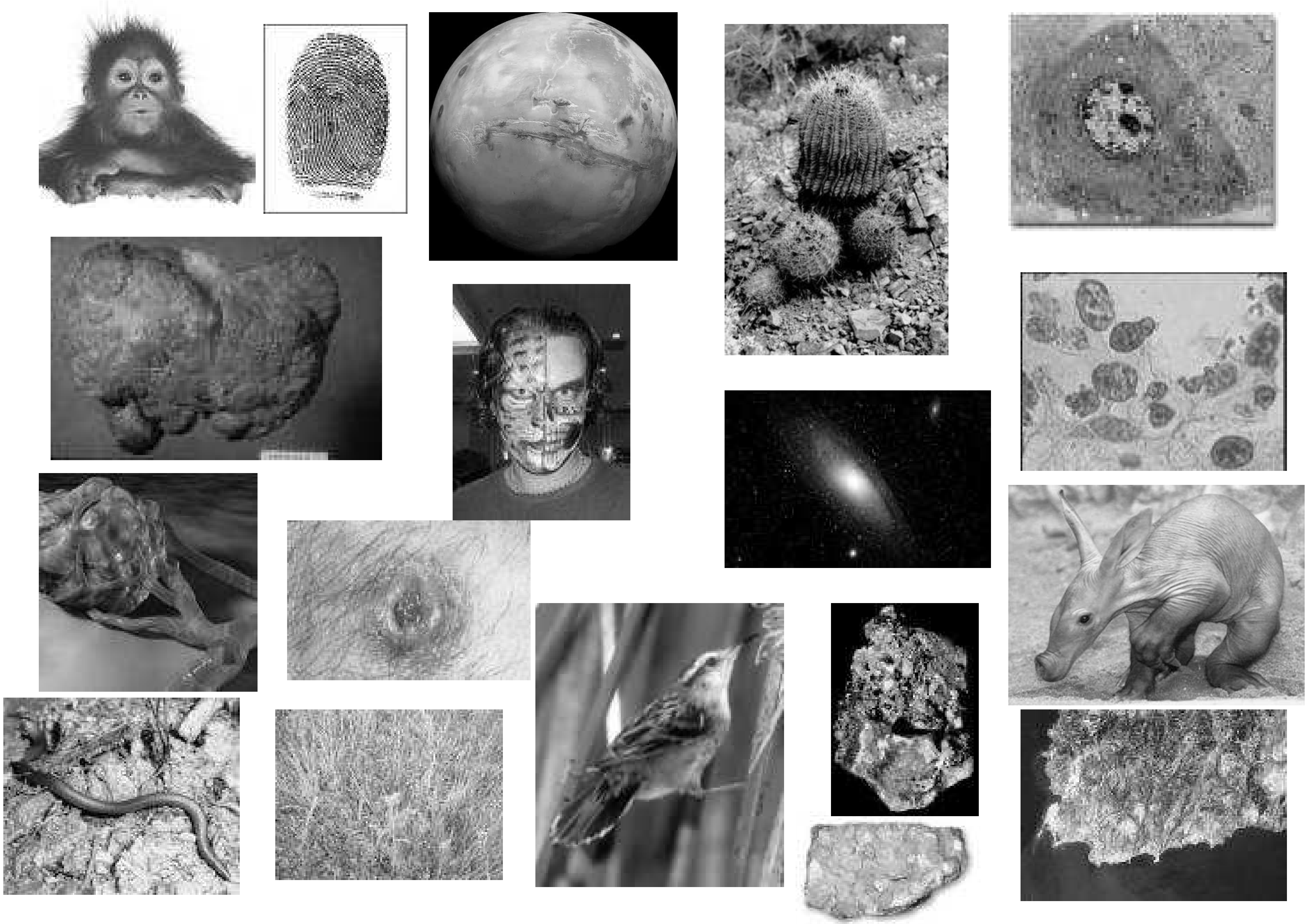
Exhibit a soft, hierarchical class structure, subject to change;

The number of classes is subjective;

Because of the unpredictable cost of errors, every decision must be checked by a human;

Ancillary non-visual information is often required for classification.

SOME NATURAL PATTERNS



INTERACTION WITH NATURAL PATTERNS

DIFFERENCE BETWEEN HUMAN & MACHINE VISUAL CAPABILITIES

With gestalt perception, we can *segment* objects from background

Are aware of *broad context*

Can filter out *correlated* noise

Can judge *pairwise similarity* based on shape, color, and texture

Computers can *store* millions of image-label pairs,

and *compute*

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

THEREFORE:

Segment object (*build model*) with human help if needed

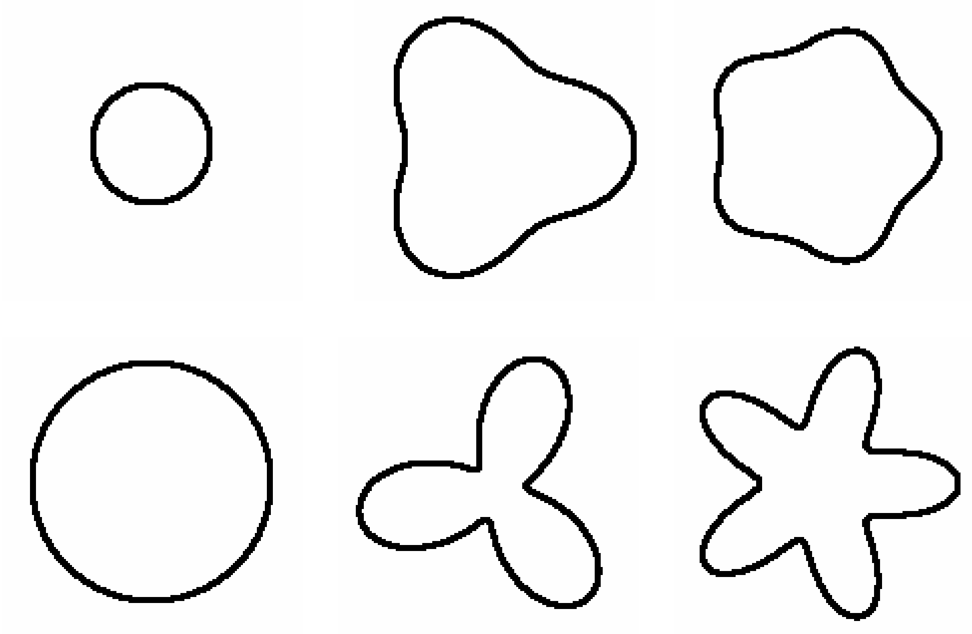
Use a domain-specific *visual model* to mediate between human and computer

Extract features, and rank candidates

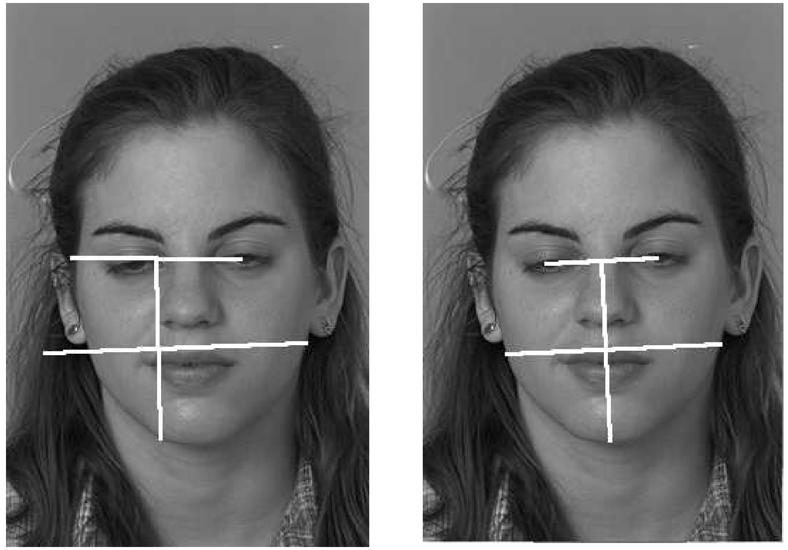
Decide final classification

We have built several experimental CAVIAR
(Computer Assisted Visual Interactive Recognition) systems

EXAMPLES OF VISIBLE MODELS



rose curves

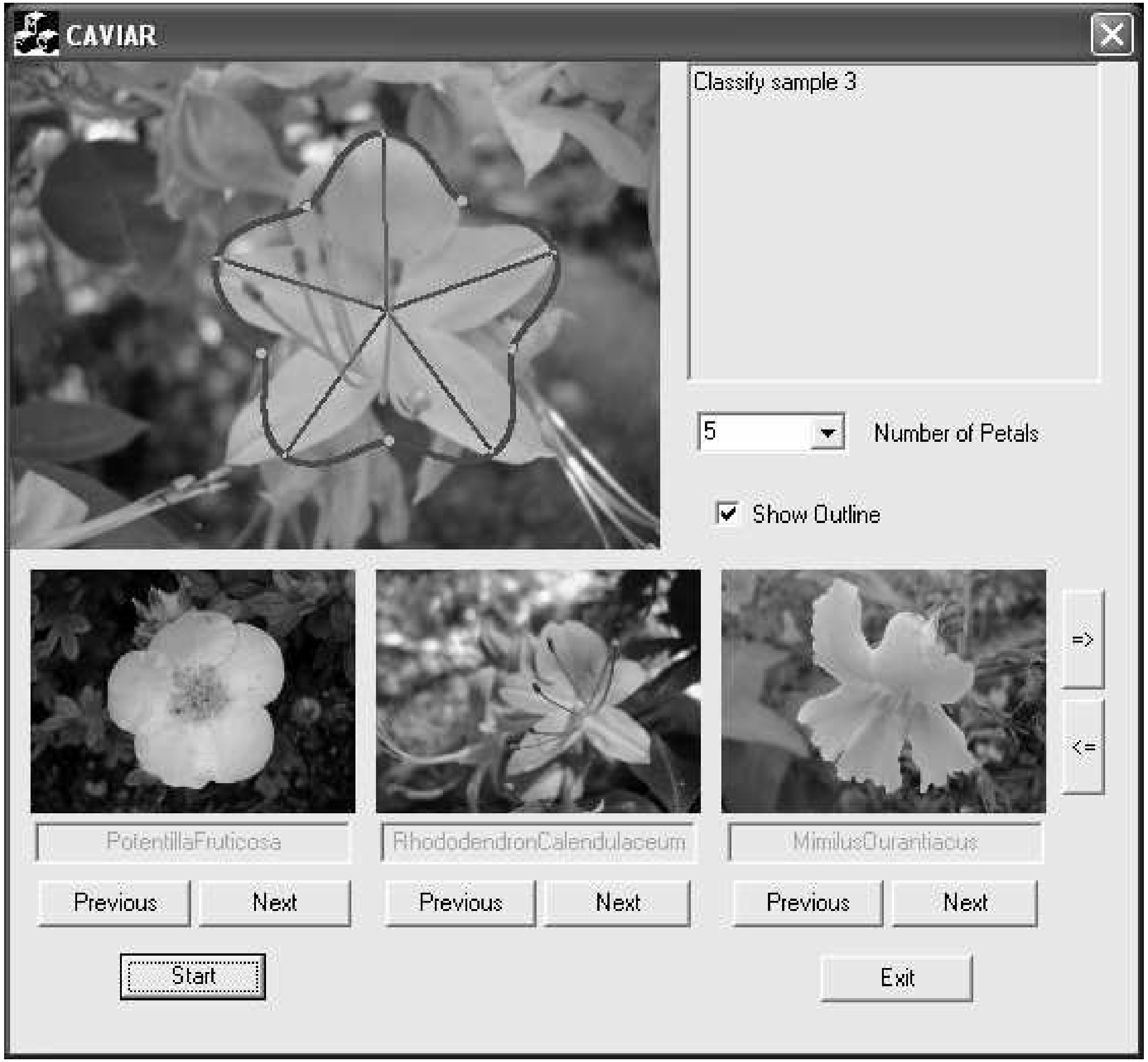


five characteristic points

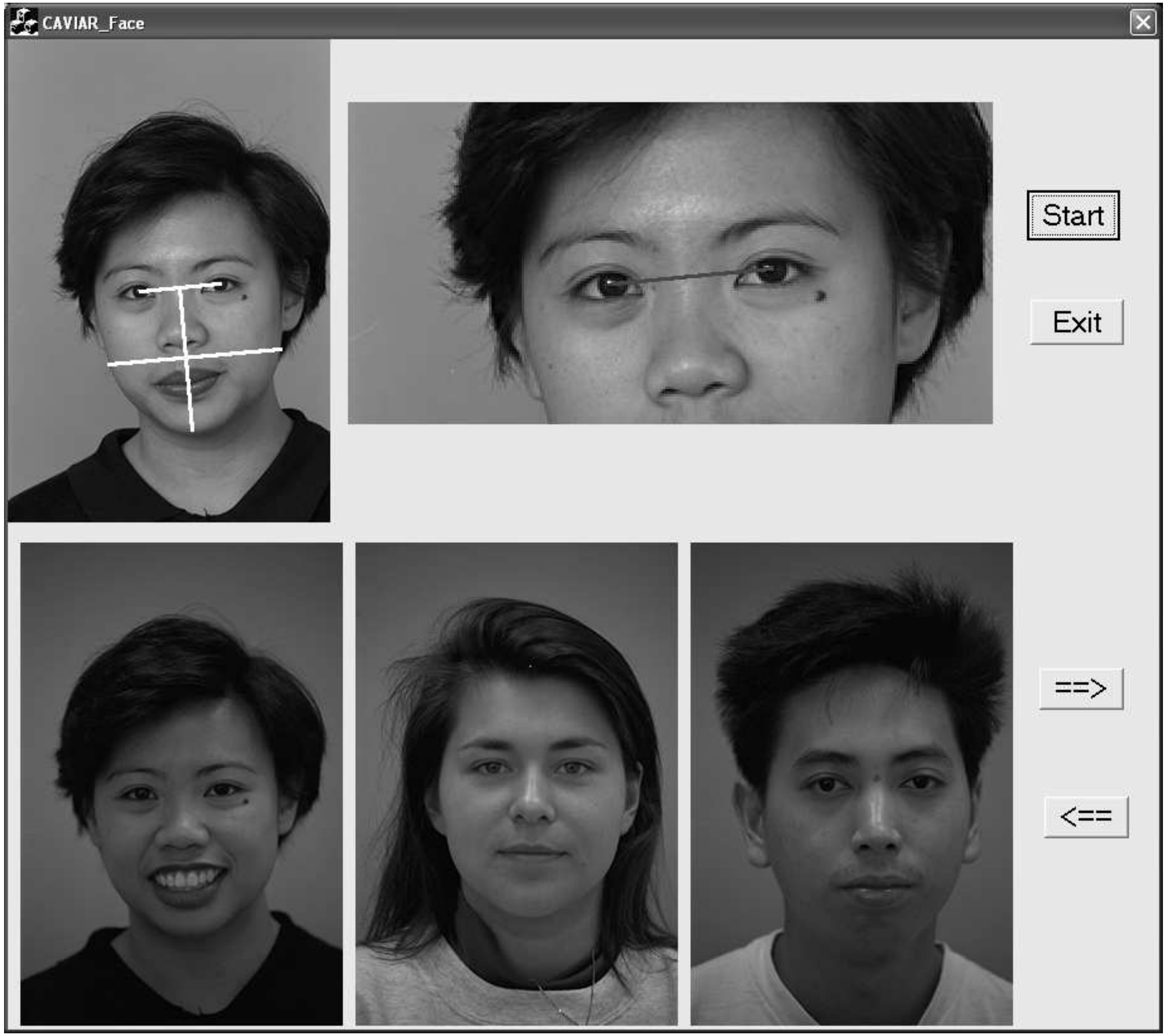
THE **VISIBLE MODEL**

- Mediates between human and computer.
- Domain-specific (different for flowers, faces, fruit, …).
- Constructed by the computer;
  corrected by user if necessary .
- The model guides feature extraction;
  the features are used to **rank order** the classes;
  the reference pictures of the top candidates are displayed.
- The operator selects the reference picture
  most like the unknown picture.
- **The human is always in charge.**

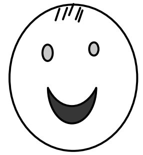
CAVIAR-flower GUI (for outlining petals)



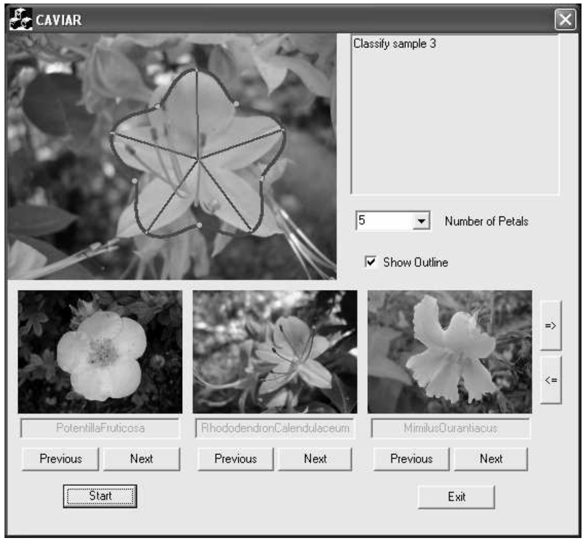
CAVIAR-face GUI (for accurate pupil location)

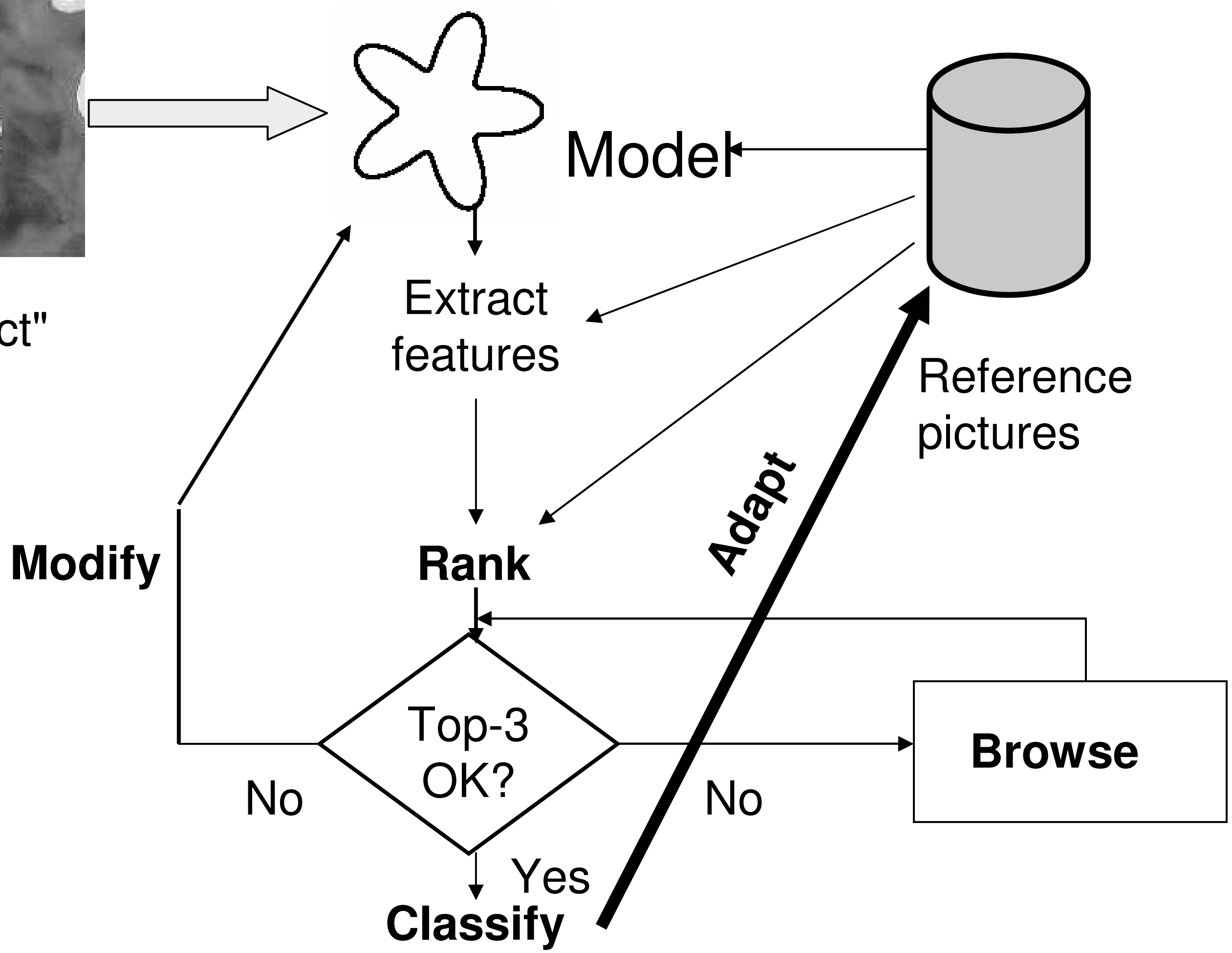


CAVIAR DATA FLOW



Unknown "object"





CAVIAR-FLOWER COMPARED TO MACHINE ALONE AND TO HUMAN ALONE.

102 classes, 102 unknowns, 6 subjects

	Accuracy (%)	Time per flower (seconds)
Interactive	**93** (83 – 99)	**12** (7 – 27)
Machine Alone	**32** (24 – 50)	-
Human Alone	93 (91 - 97)	**26** (18 - 36)

CAVIAR-FACE COMPARED TO MACHINE ALONE AND TO HUMAN ALONE (200 faces)

200 pictures as *gallery*, 50 pictures as *probes*, 6 subjects

	Accuracy (%)	Time per face (seconds)
Interactive	**99.7**	**8**
Machine alone	**47**	--
Human alone	--	**66**

SUMMARY OF OBSERVATIONS

Interactive recognition is twice as **fast** as unaided human, and far more **accurate** than unaided machine (without years of R&D).

Parsimonious interaction *throughout* the process is better than only at the beginning or end.

CAVIAR scales up: it can be initialized with *a single training sample per class,* and *improves with use*.

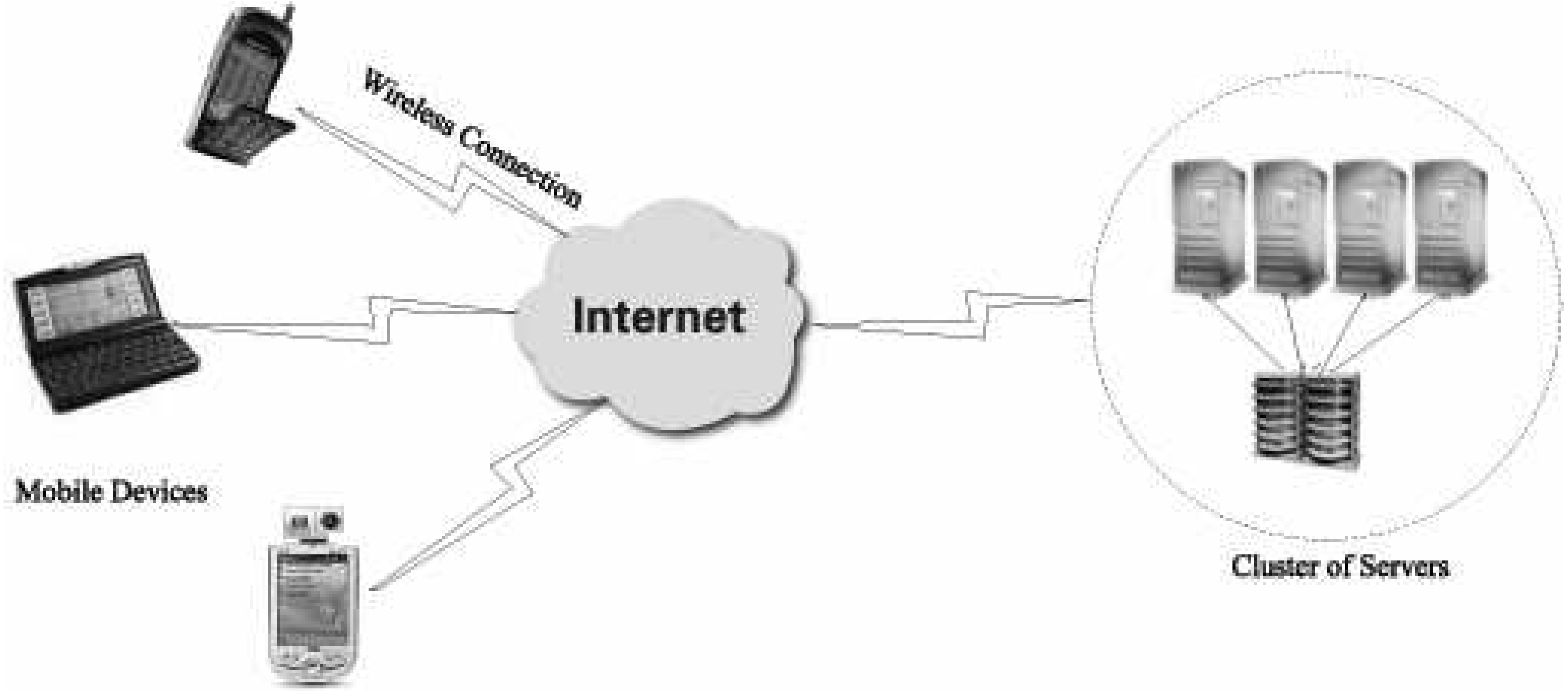
NB:

Our automated classifier for rank-ordering may not be the best.

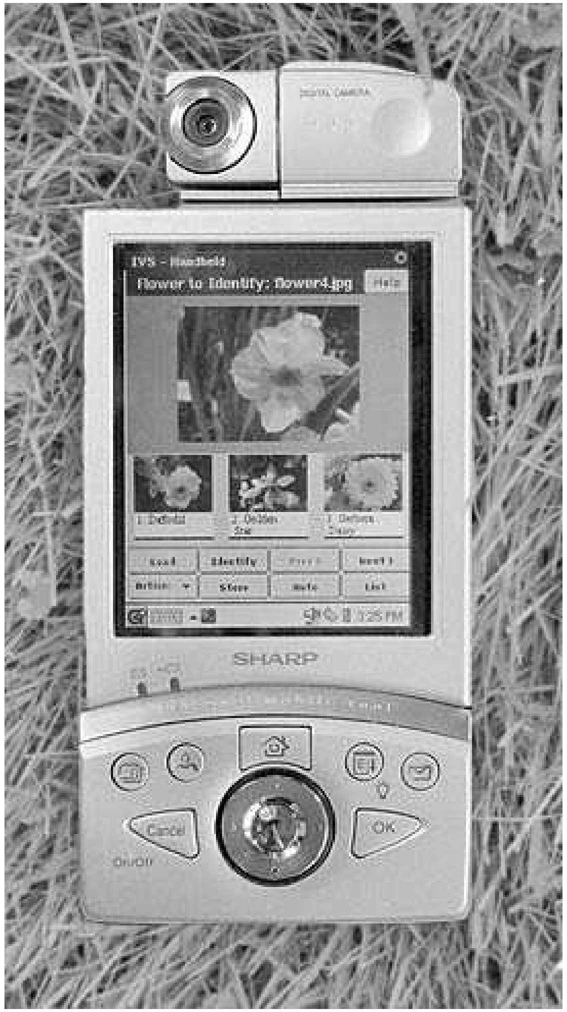
However, better algorithms will reduce interactive time and increase interactive accuracy even further.

We expect that the interactive system will always outperform both the unaided human and the unaided machine

MOBILE AND NETWORKED CAVIARs



SELF-CONTAINED MOBILE CAVIAR AT PACE UNIVERSITY



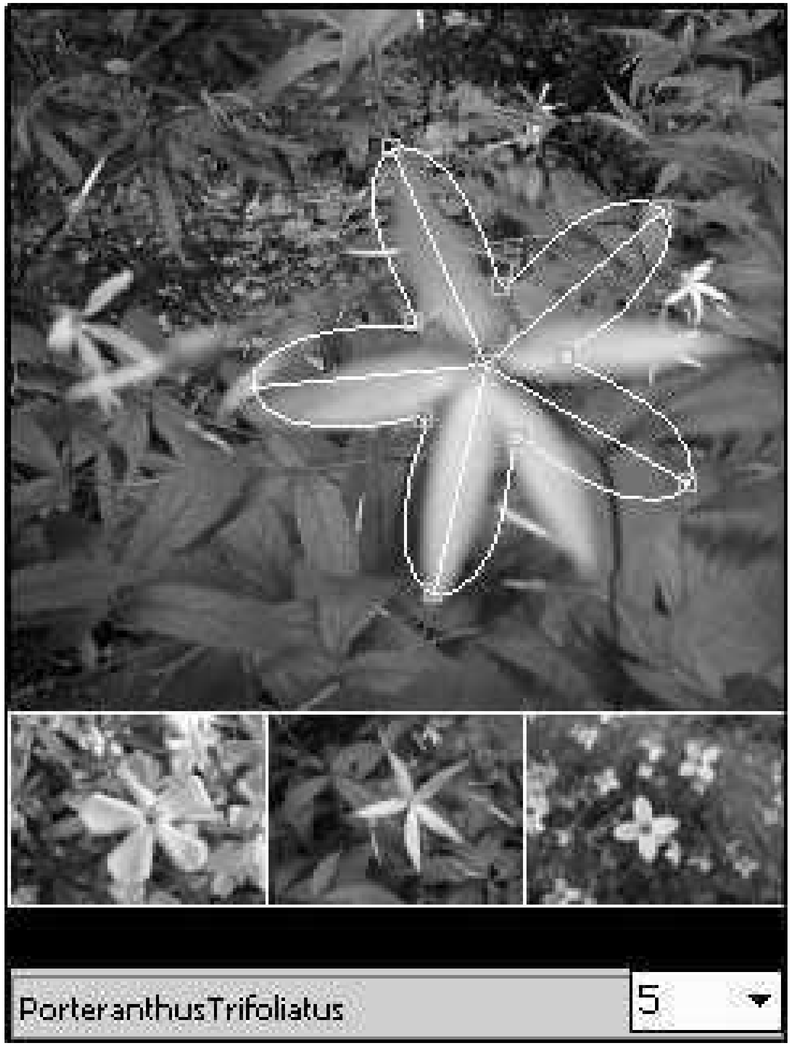
Sharp Zaurus

200 MHz, 64MB

Linux + Personal JAVA

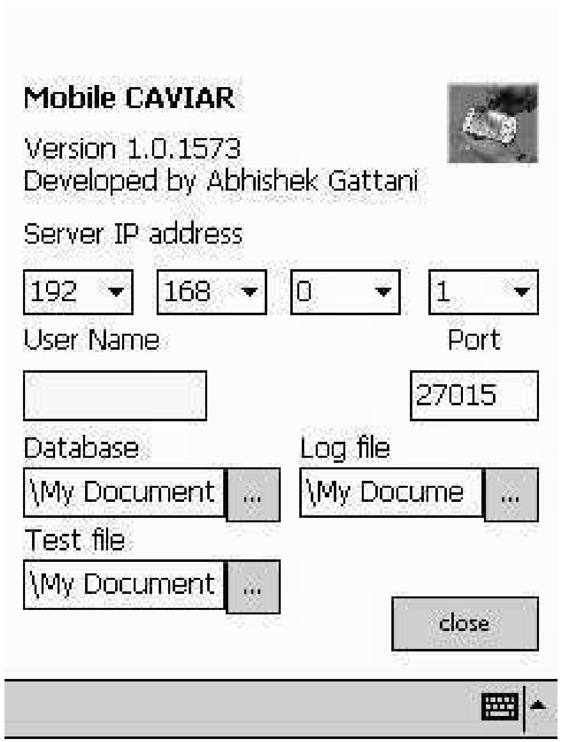
NETWORKED MOBILE CAVIAR AT RENSSELAER

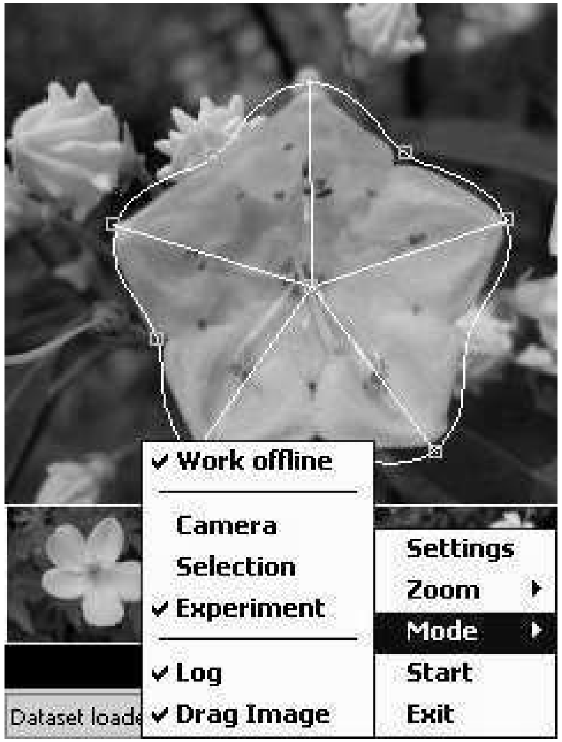
Toshiba, IEEE 802.11b

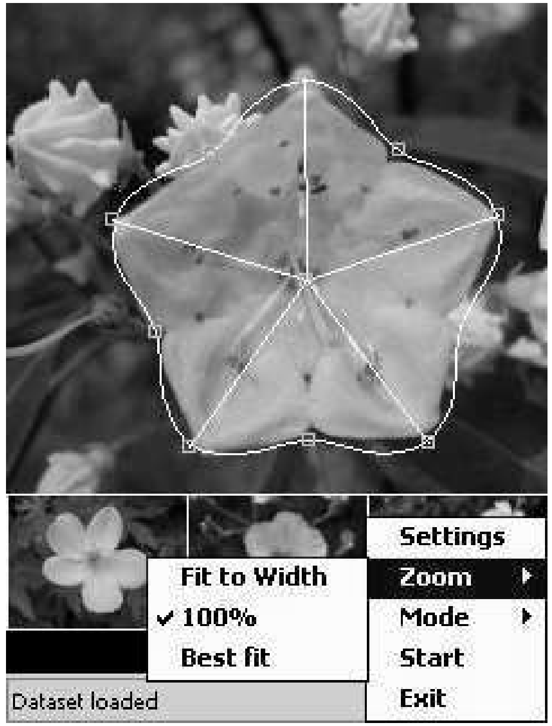


Abhishek Gattani

M-CAVIAR GUI


Mobile CAVIAR
Version 1.0.1573
Developed by Abhishek Gattani
Server IP address
192
168
0
1
User Name
Port
27015
Database
Log file
\My Document
\My Docume
Test file
\My Document
close

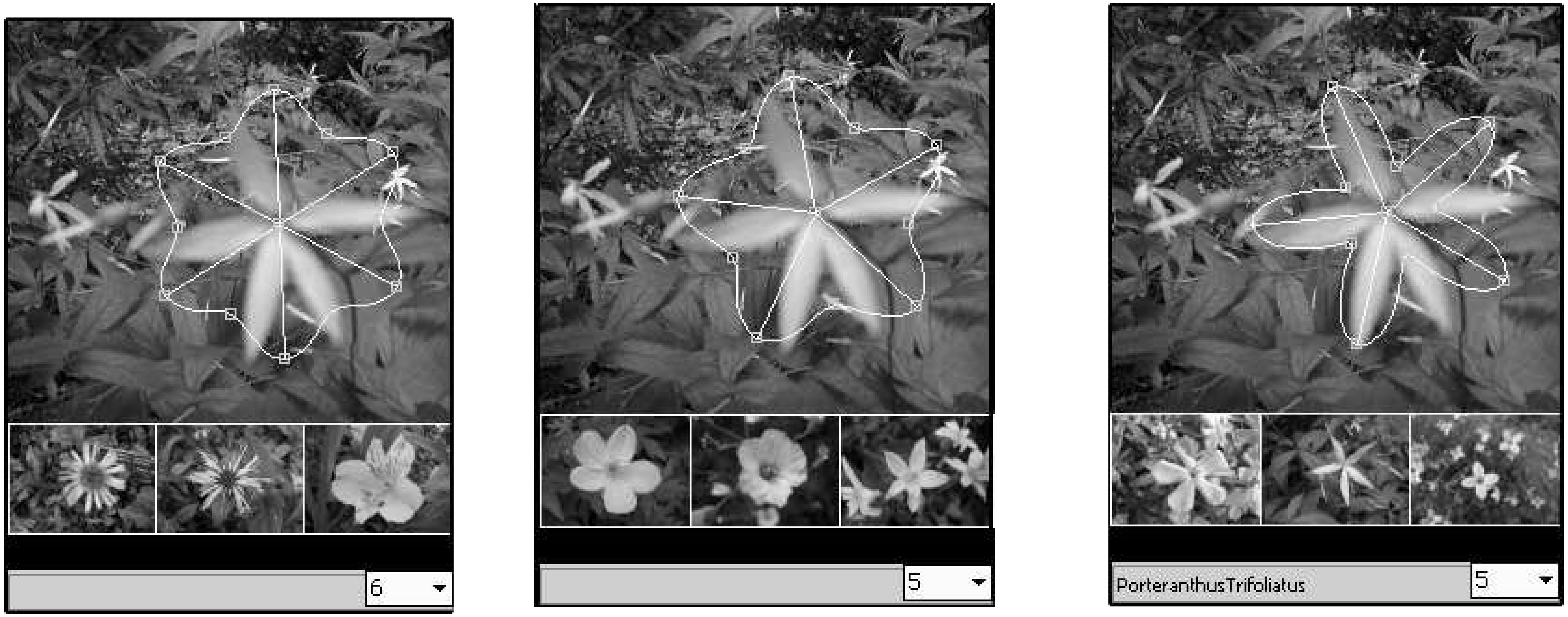

Work offline
Camera
Selection
Experiment
Log
Drag Image
Settings
Zoom
Mode
Start
Exit
Dataset loade


Fit to Width
100%
Best fit
Settings
Zoom
Mode
Start
Exit
Dataset loaded

PDA and Camera Specs

- **Toshiba e800 Specifications**
- CPU Intel PXA263 400 MHz
- Memory 128MB SDRAM Main memory, 32MB CMOS Flash ROM; Application Memory: 32MB NAND Memory (Flash ROM Disk)
- Display 4.0” diagonal, TFT Transective at 65,536 (64K) colors
- Resolution QVGA 240 x 320; VGA 480 x 640
- Graphics Controller ATI Graphics Controller with 2MB internal video memory
- Wireless Integrated Wi-Fi (IEEE 802.11b)
- Expansion 1 Type I/Type II CF Card Slot (3.3V) 1 SD (Secure Digital) card slot Dimensions 135.0 x 77.0 x 16.7 mm
- Weight 198 g
- Operating System Microsoft Mobile Software for Pocket PC 2003 Premium Edition
-
- **Camera Specifications**
- Sensor 1.3 Mega pixels (1280 x 1024 pixels)
- Connection SDIO Slot
- Features 180 Degree Swivel Lens / Adjustable Focus 4x Digital Zoom Preview & Playback) Adjustable Self Timer
- Resolutions 1280x1024, 1024x768, 640 x 480, 320 x 240
- Image Format Standard JPEG
- Color Palette 24-bit Full Color
- Functions Auto Exposure, White Balance and Color Control

M-CAVIAR Classification Example

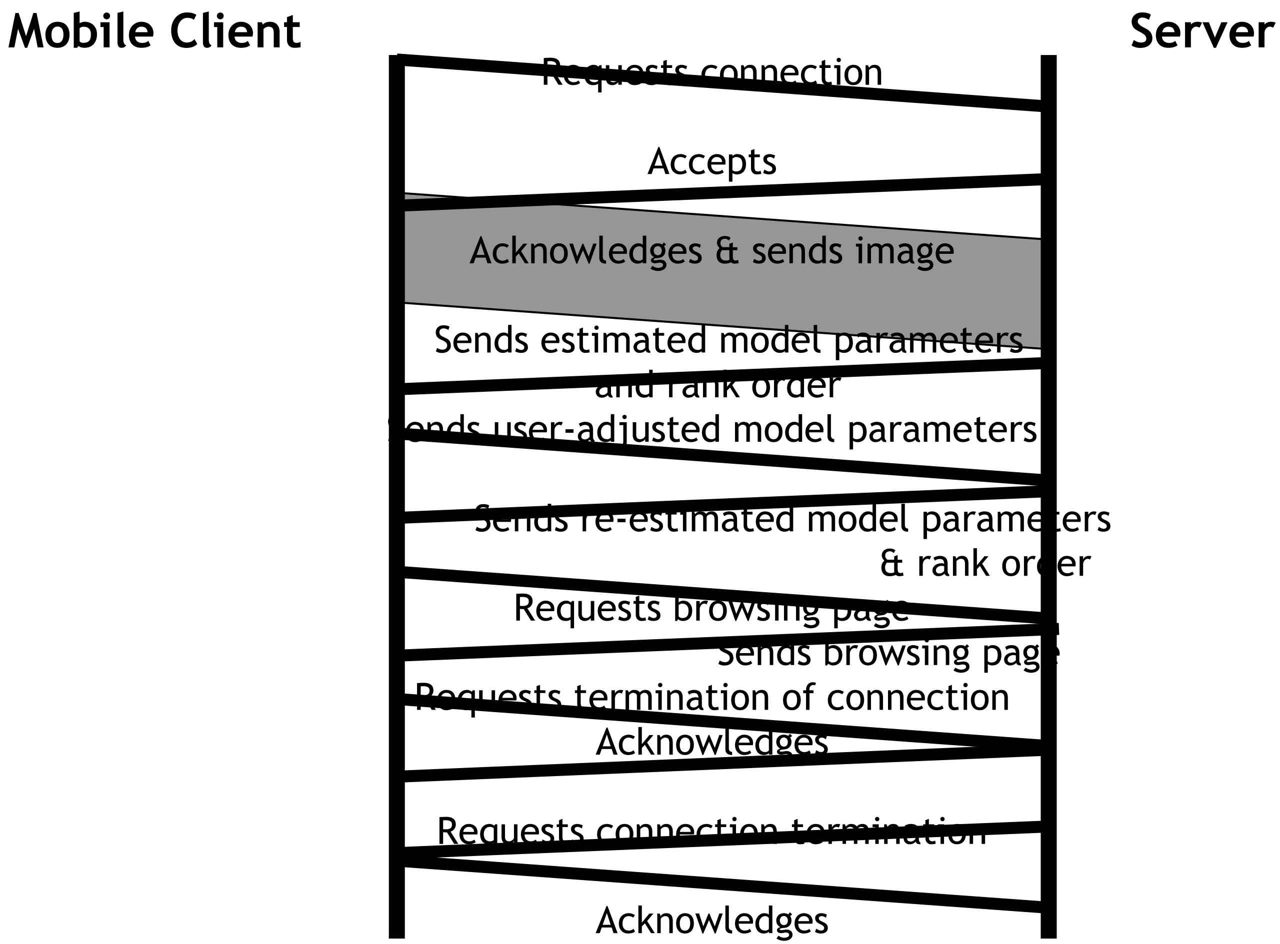


(1) Automatic ordering unsuccessful as the flower is out of focus.

(2) Petal number changed to 5 & the re-estimated rank order
and rose-curve instance are displayed.

(3) The inner radius and phase are changed to fit the flower better and the correct candidate appears.

Communication sequence between the PDA and the server for identifying a test sample



PR NETWORKS for MOBILE PLATFORMS

OPEN MIND initiative – David Stork

Dispersed hierarchy of expert labelers

Multiple labels for ambiguous patterns

Ubiquitous data collection

LARGE training sets

MARIGOLDS

Digital camera

Nikon Coolpix 775

PDA

Veo 130s

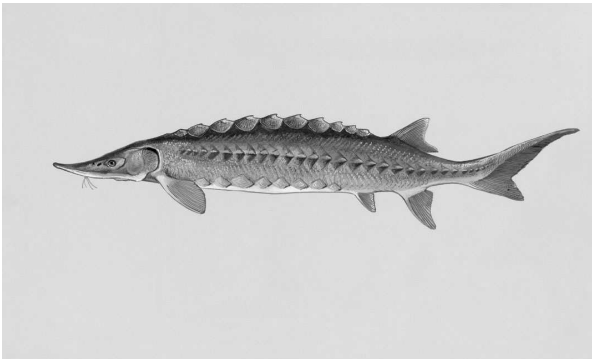
Cell phone

Motorola V400

OTHER APPLICATIONS: FISH ??

Black Crappie

Alabama Shad

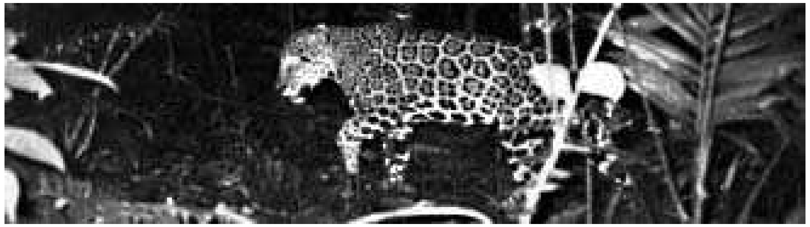


Atlantic Sturgeon

Blue Gill

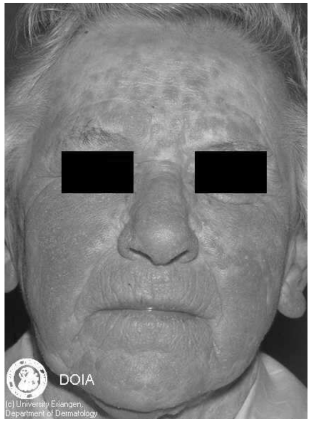
U.S. Fish & wild life service

CRYPTIC CATS ?

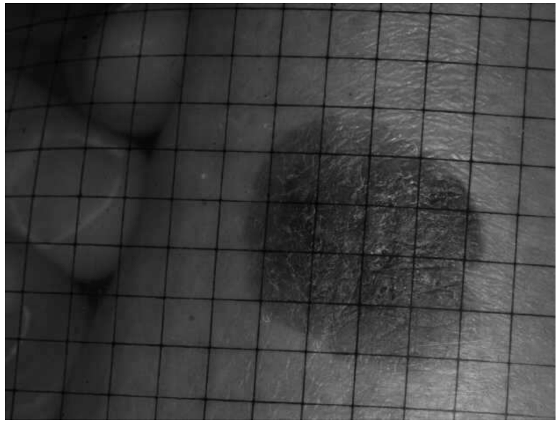


Jan Schipper
NSF-IGERT Fellow
CATIE
Escuela Posgrado
Sede Central 7170
Turrialba, Costa Rica
Central America

Proyecto Conservación del Área Talamanca (ProCAT) is an international project under the umbrella of the Institute of the Rockies.

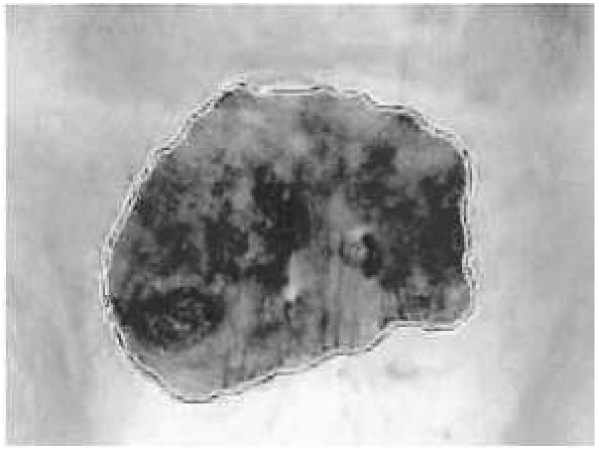


CAVIAR-Derma?



Nearly 1000 diagnoses (classes)

Big image atlases available
- John Hopkins dermatology image atlas
- University of Erlangen, Heidelberg



Color, shape and texture features

Compare with healthy skin patch of same individual

Vary lighting and scale

DERMATOLOGICAL APPLICATONS

Cosmetic dermatology, scar assessment, beauty-aids

Skin cancers: melanoma

Infectious or contagious diseases with spots, e.g. measles

Rashes: hives, eczemas, psoriasis

Accidents: burns, cuts, frostbites

Sexually transmitted diseases

Poisonous plants and bugs: poison ivy, insect bites

Bio-terrorism agents: cutaneous anthrax, plague, tularemia

Potential scenarios for CAVIAR-Derma

When expert unavailable:
military, expeditions, isolated elderly, developing countries

Privacy and convenience

Possibility of collecting additional non-visual info

Photos may be forwarded to health organizations

Training: medical and paramedical personnel

CONTEXT & STYLE

Language context has long been exploited in OCR and ASR through morphological, lexical, and syntactic language models

Style context takes advantage of the common source of patterns (writer, font, printer, copier, scanner).

The way Maria writes “5” can help to recognize whether an ambiguous digit is a “6” or an “8”!

Cf: Sarkar & Nagy, IEEE PAMI, January 2005
Veeramachaneni & Nagy, same issue

LANGUAGE and STYLE CONTEXT

?

Isabella

l40 mm long

LANGUAGE CONTEXT

?

lt47dh1

lt47dhl

STYLE CONTEXT

Inter-pattern Feature Dependence (Style)

Single-class and multi-class style

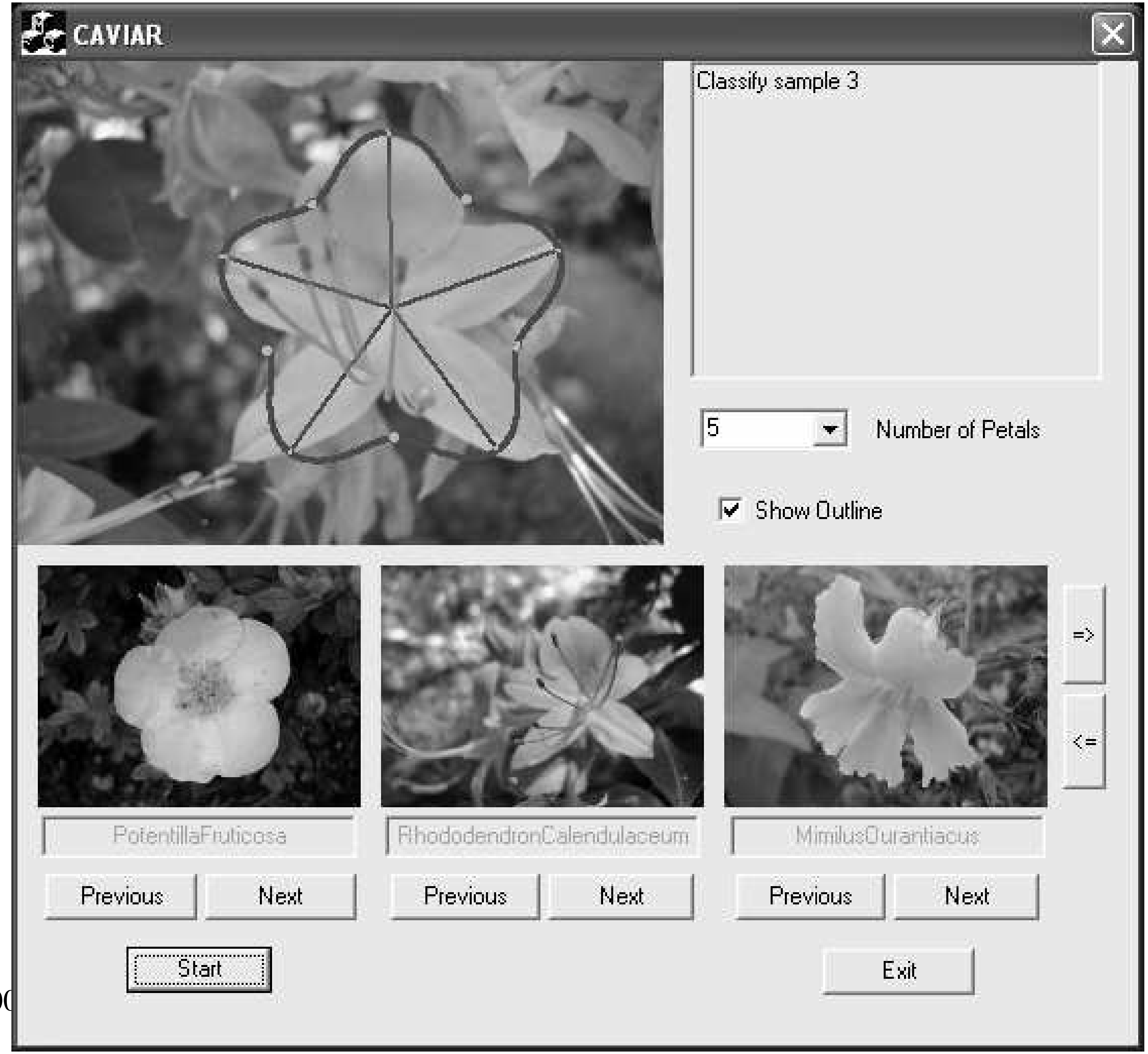
	SINGLE CLASS STYLE	MULTI-CLASS STYLE
Source 1:	29/05/1925	25/07/1922
Source 2:	15/05/1990	**05/05/1925**
Source 3:	21/06/1943	**02/06/1943**
Source 4:	05 /29/1945	02/25/1942

Styles are induced in a collection of documents by multiple sources*.

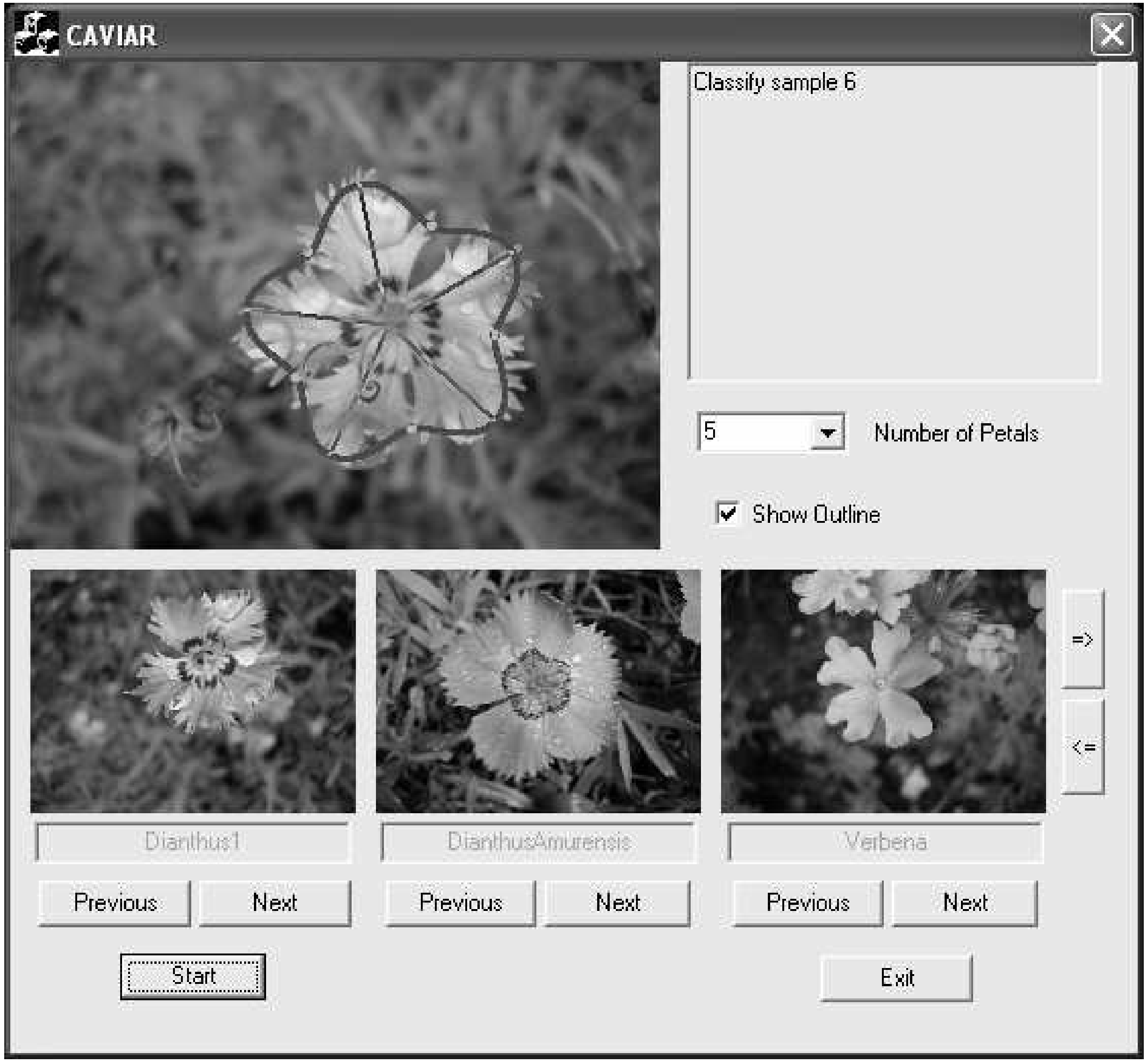
* fonts, printers, scanners, writers, speakers, microphones, ...

CAVIAR-FLOWER

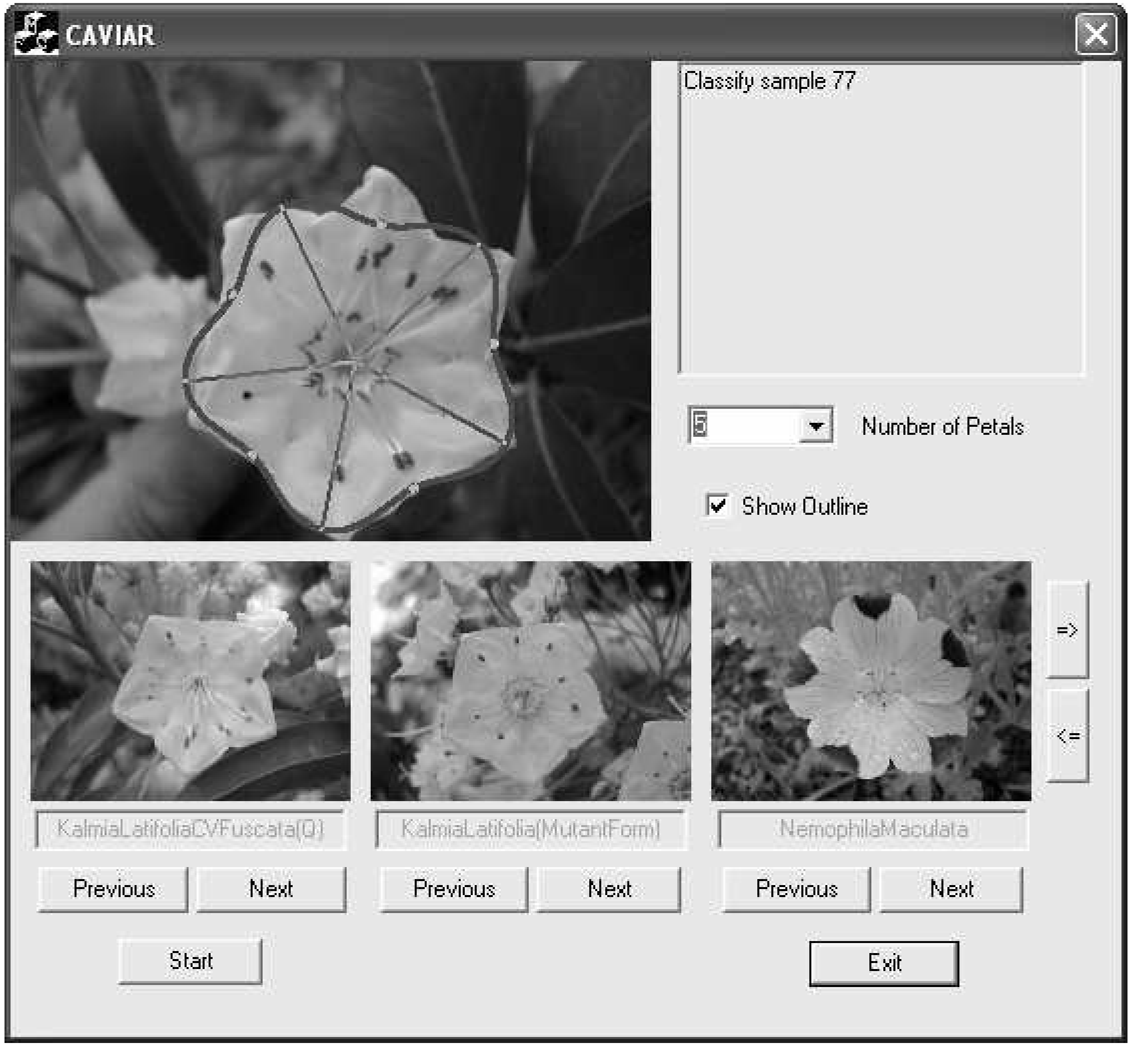
CAVIAR-FLOWER



CAVIAR-FLOWER (continued)



CAVIAR-FLOWER (continued)



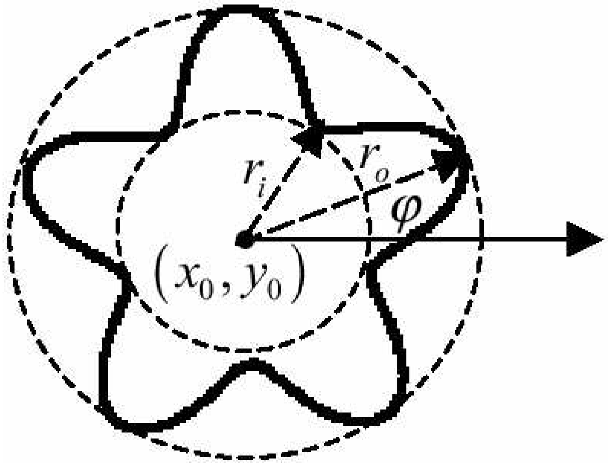
CAVIAR-FLOWER (continued)

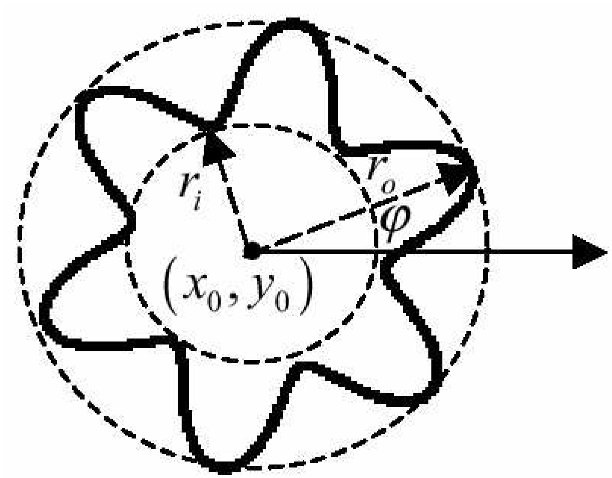
ROSE CURVE MODEL

- Parametric curve with six parameters.

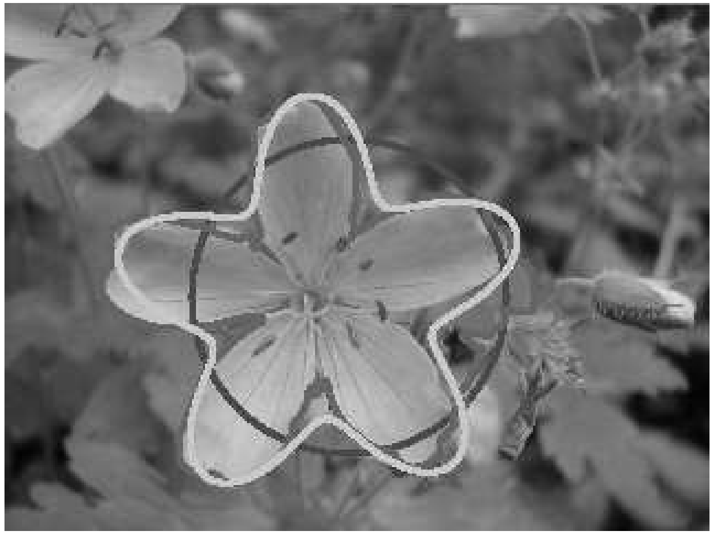
$$\rho = \frac{r_o + r_i}{2} + \frac{r_o - r_i}{2} cos(n\theta + n\varphi)$$
$$= a + b\cos(n\theta + n\varphi)$$

- Flowers are composed of petals, which have circular symmetry.
- When n=0, rose curve reduces to circle.





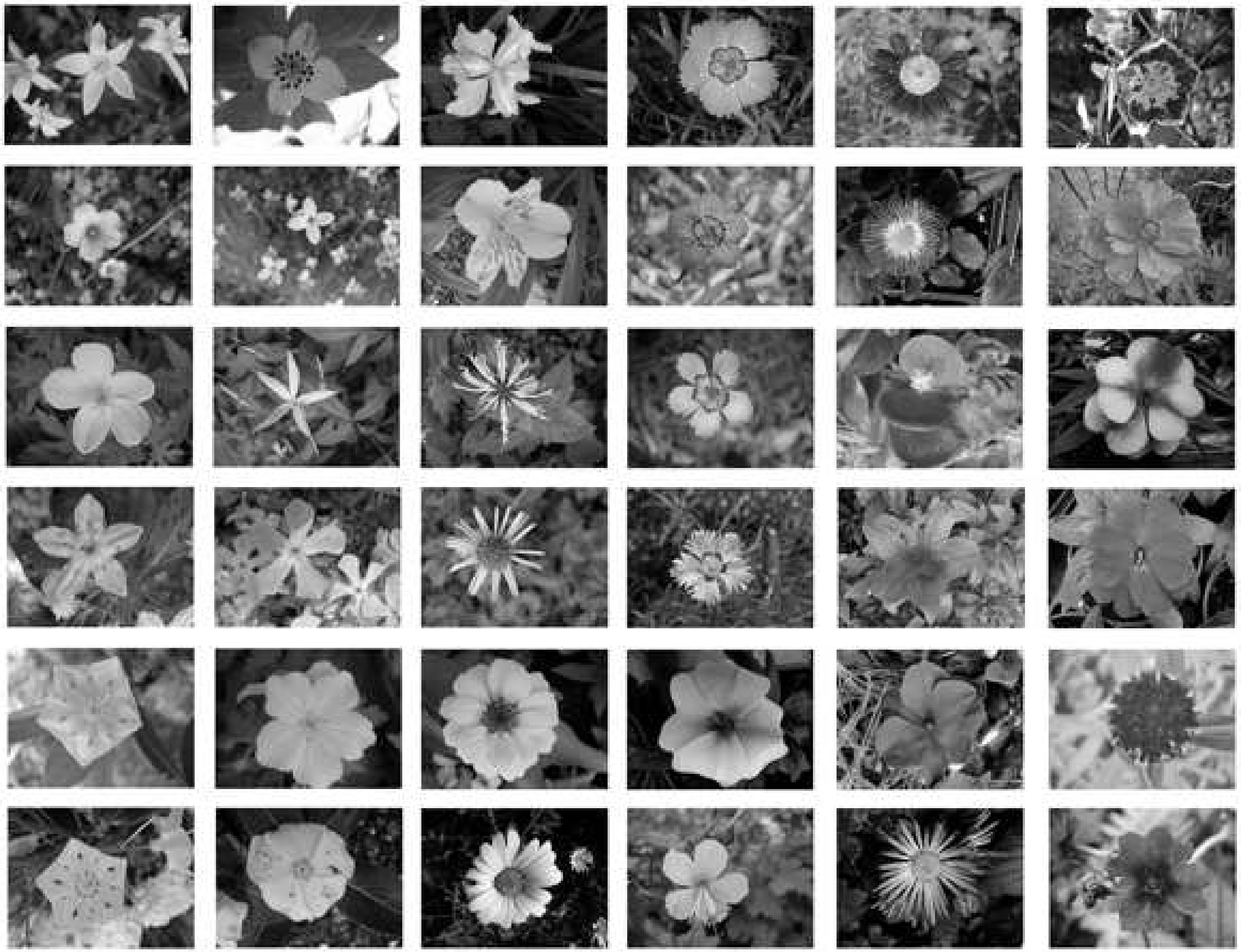
AUTOMATIC MODEL CONSTRUCTION

STRESS FLOWER DATABASE

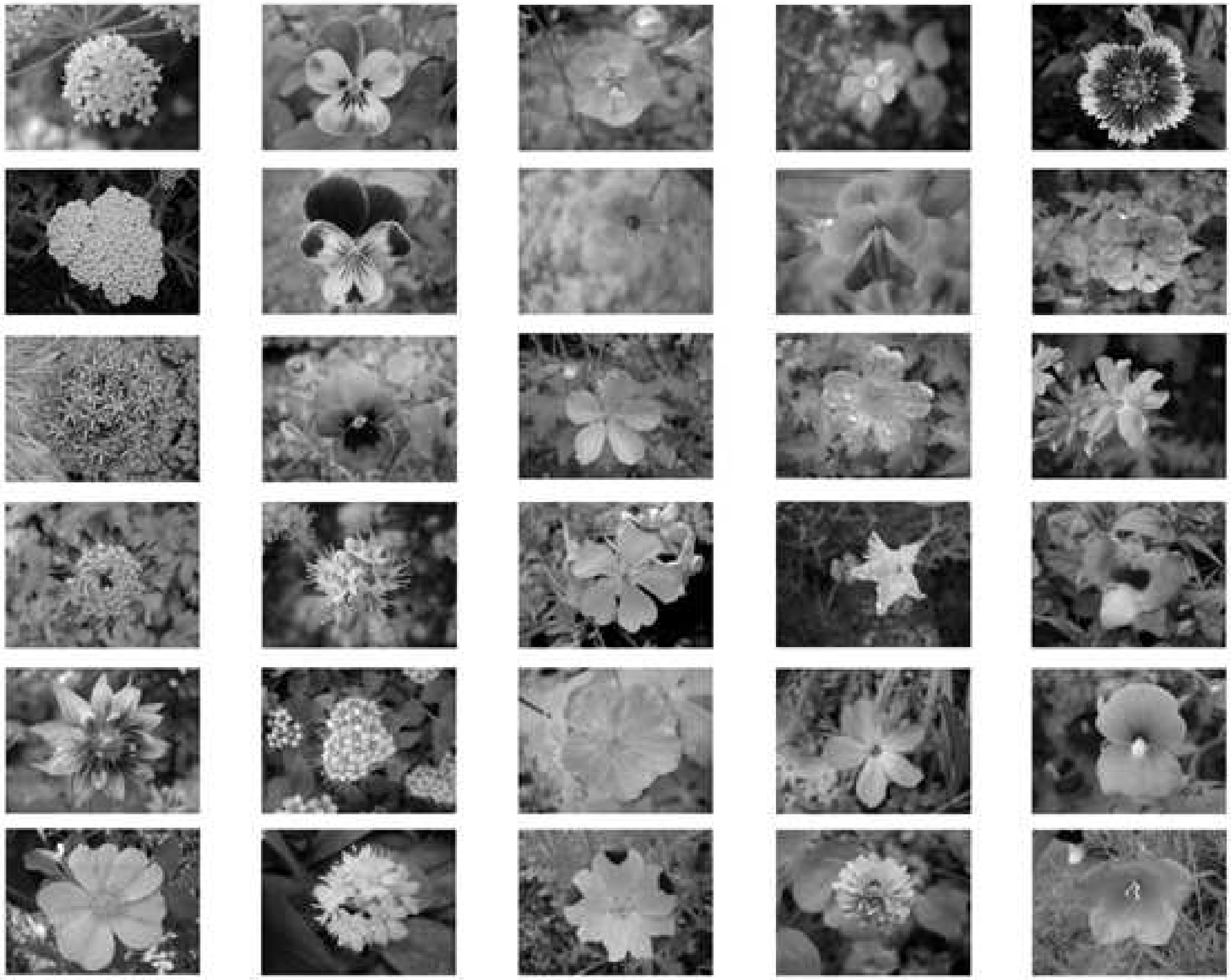
- 320 by 240 pixel pictures
- Highly variable illumination, and complex background
- 216 samples from 29 classes for development
- 612 samples from 102 classes for evaluation
- Most (digital) photos from New England Wildflower Garden

Flower Database (1)



Flower Database (2)

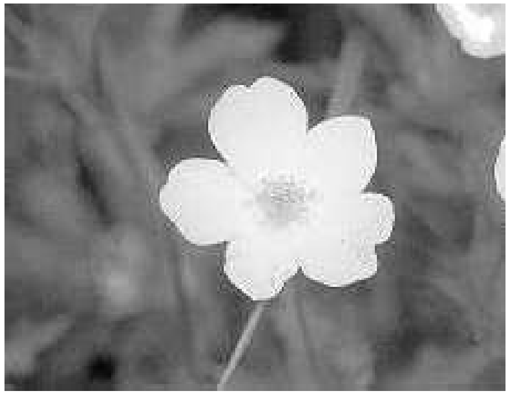
Flower Database (3)



EASILY CONFUSED FLOWERS

Bellis Perennis
Lawn Daisy, English Daisy

LeucanthemumVulgare
Ox-eye Daisy



Anemone Canadensis
Windflower,
Canada Anemone

Viola Canadensis
Canada Violet

CAVIAR Experiments

- 30 subjects
- 612 flower pictures of 102 species
- Every interactive mouse click
  and every automated step
  recorded in LOG files for detailed analysis

CAVIAR Experimental Protocol

Experiment Type	# of Subjects	Training Samples	Test Sample	Notes
I	6	1,2,3,4,5	6	Browsing-only with 5 reference samples
II	6	1,2,3,4,5	6	Interactive with 5 training samples
III	6	1	2,3	Interactive with 1 training sample
IV	6	1,2*,3*	4,5	Interactive with 1 training sample + results of III
V	6	1,2*,3*,4*5 *	6	Interactive with 1 training sample + results of III, IV

* samples initially without labels

Welcome to

Computer Assisted Visual InterActive Recognition (*CAVIAR*)

CAVIAR is an interactive flower classification program. By interacting with the computer, we hope that you can recognize flowers more accurately than a computer can by itself, and faster than you can without computer help.

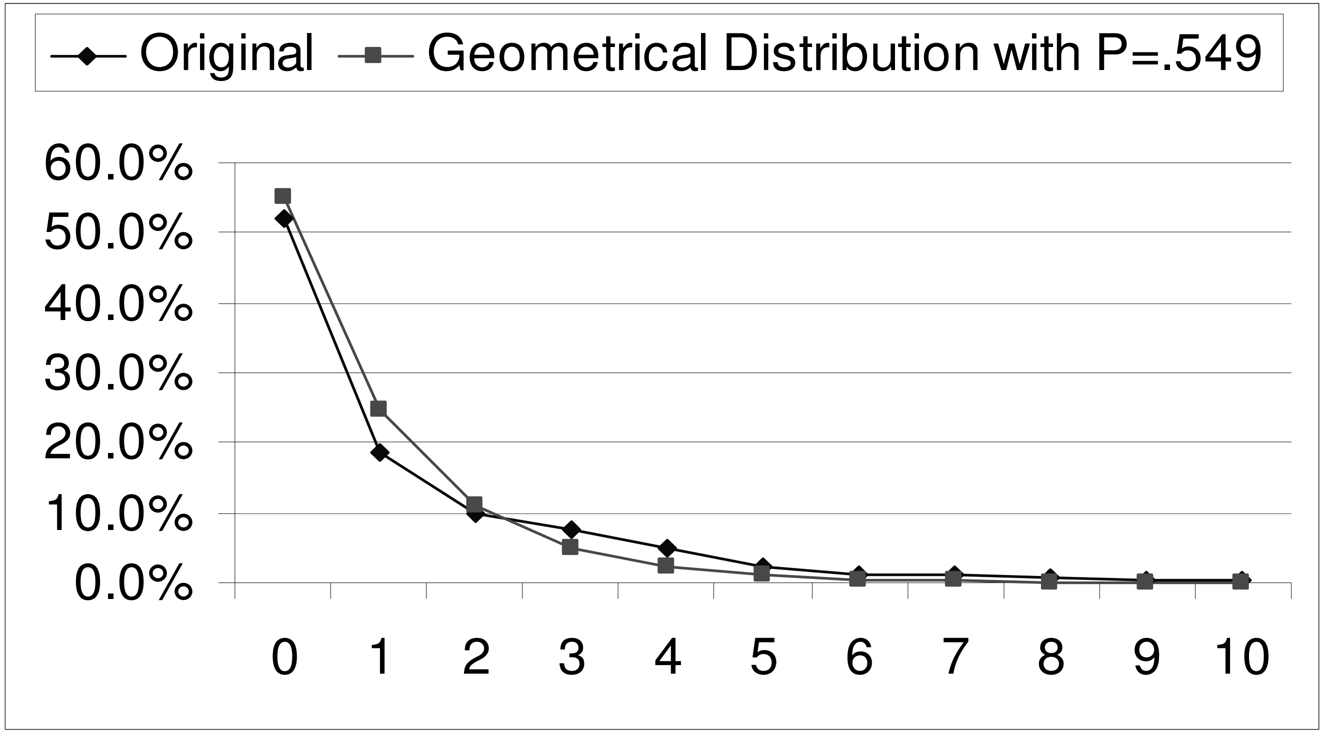
RPI ECSE DocLab

Jie Zou, Borjan Gagoski, George Nagy

INTERACTION COMPARED TO *MACHINE ALONE* AND TO *HUMAN ALONE*.

	Accuracy (%)	Time per flower (seconds)
Interactive	**93** (83 – 99)	**12** (7.23 – 27.13)
Machine Alone	**32** (24 – 50)	-
Human Alone	93 (91 - 97)	**26** (18 - 36)

Finite State Machine model of interaction



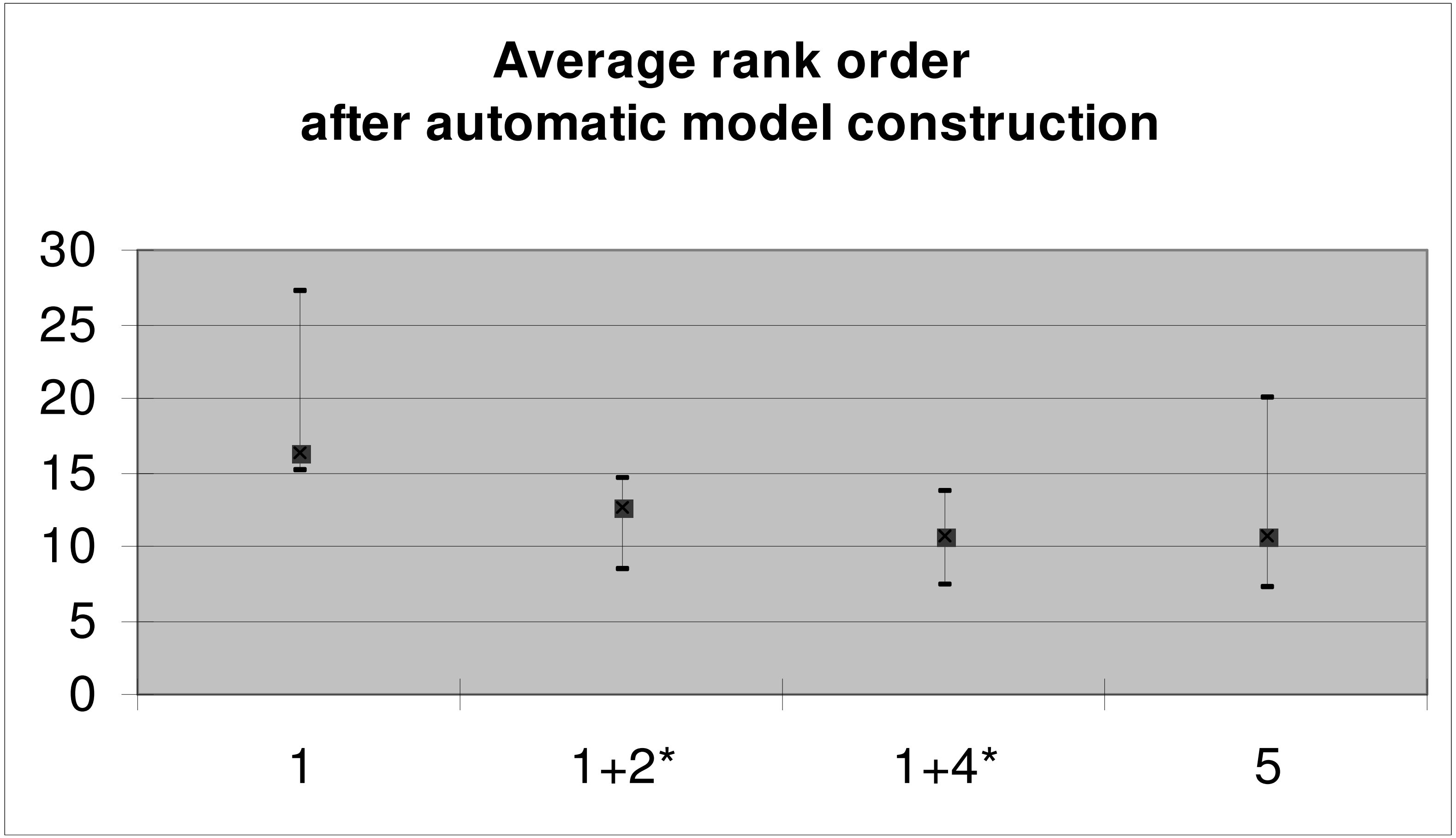
- 52% samples are immediately confirmed.
- 90% samples are identified after 3 adjustments.
- The probability of success on each adjustment is ~0.5.

DECISION-DIRECTED ADAPTATION

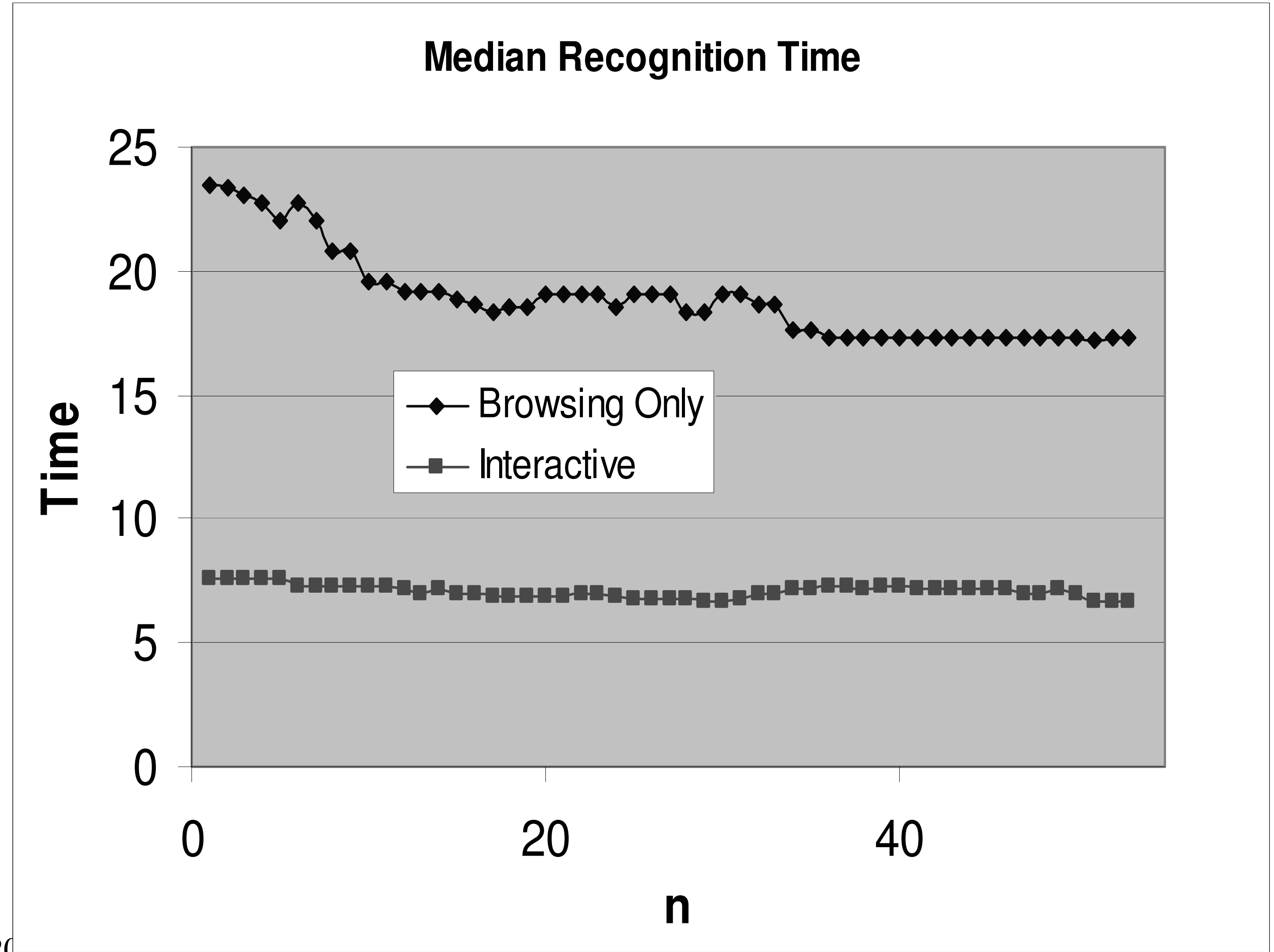
RESULTS:

Year	Collaborator	Data	# classes	d	Gain
1966	Shelton	12-font typescript	26	96	5.0X
1994	Baird	100-font print	96	512	2.5X
2002	Harsha V.	NIST hand-print	10	50	1.8X
2003	El-Nasan	cursive handwriting	100	42	4.0X
2004	Zou	flowers	102	8	1.2X

SYSTEM ADAPTATION



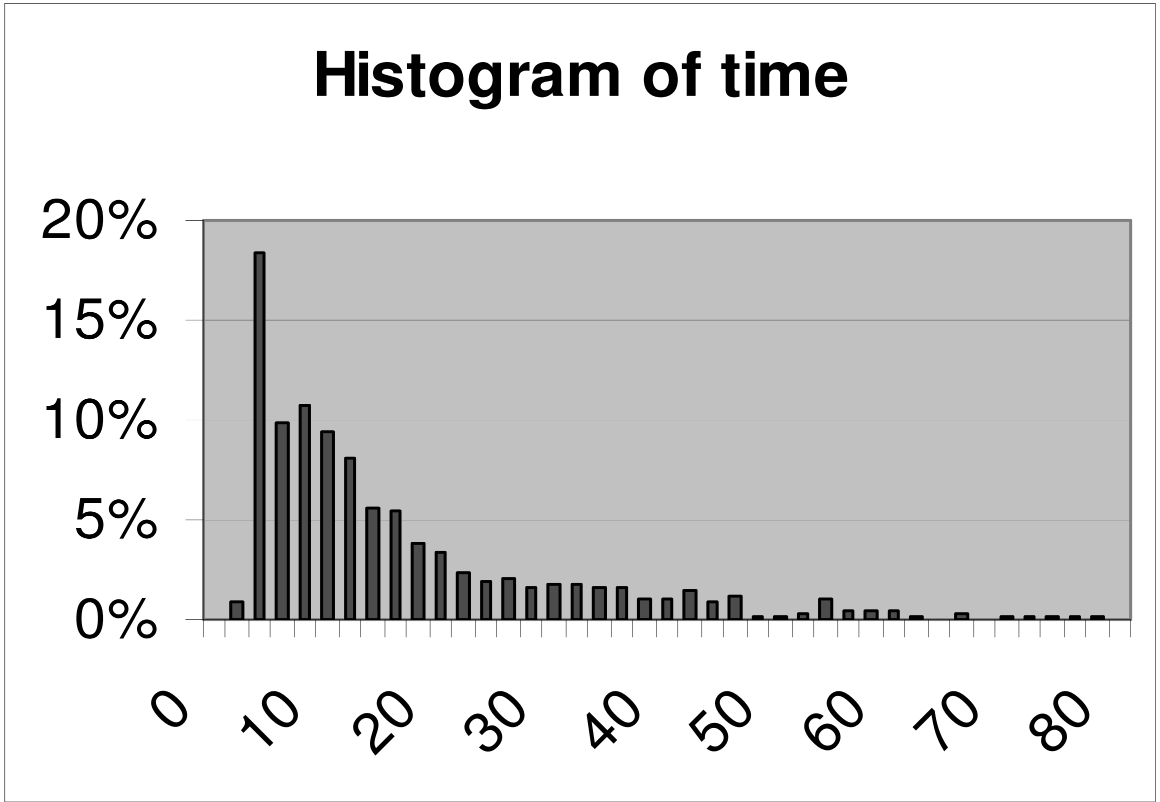
HUMAN LEARNING



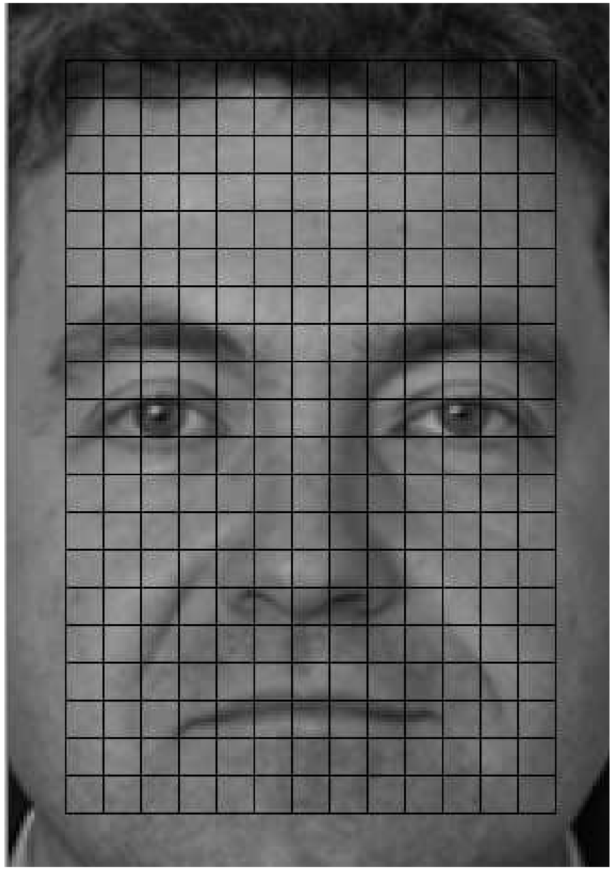
ENROLLMENT:

REFERENCE DATA SEGMENTED WITH INTERACTIVE CORRECTION

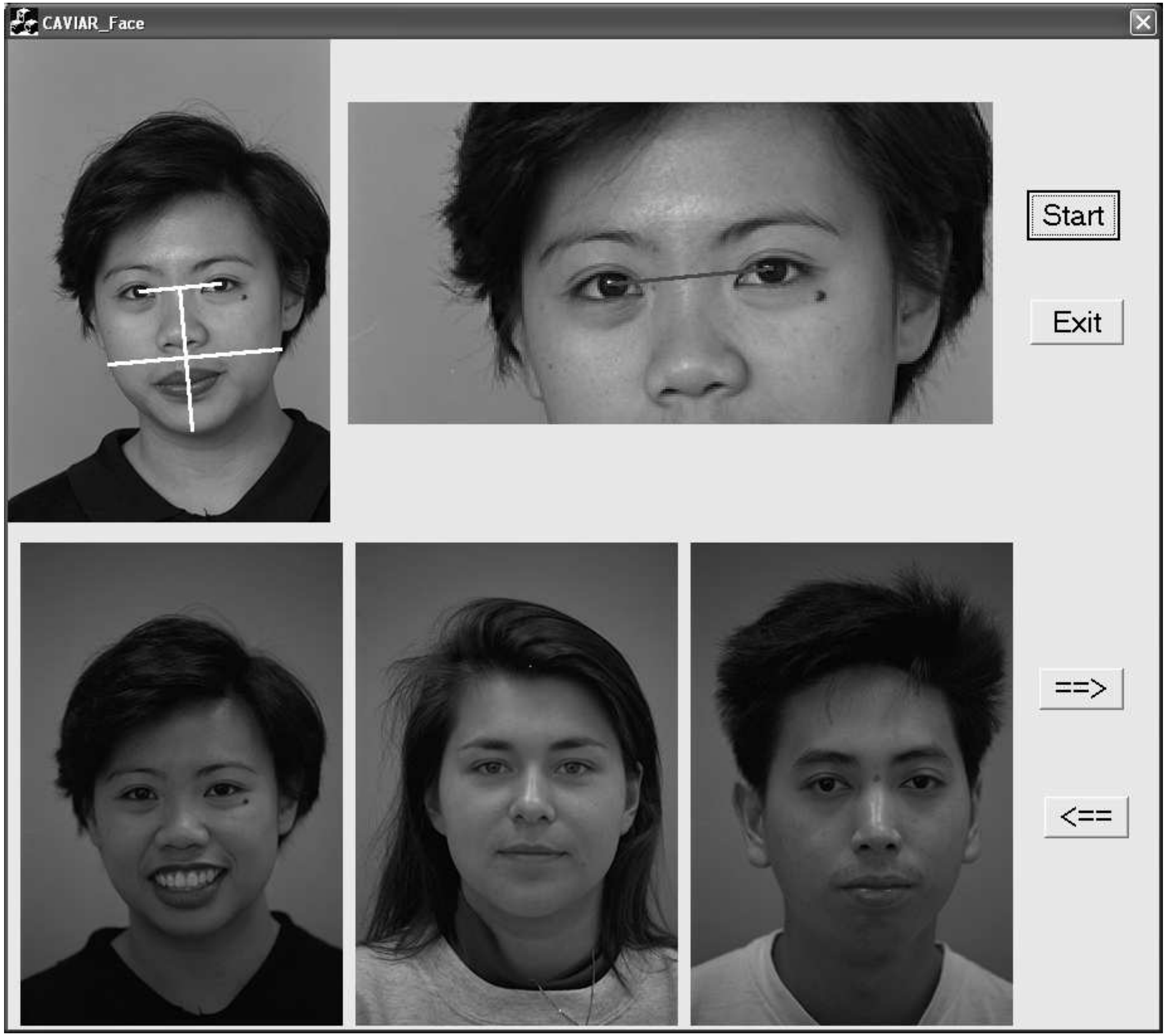
- 15.2 seconds per picture (5.7 seed pixels),
- 1078 flowers from 113 species



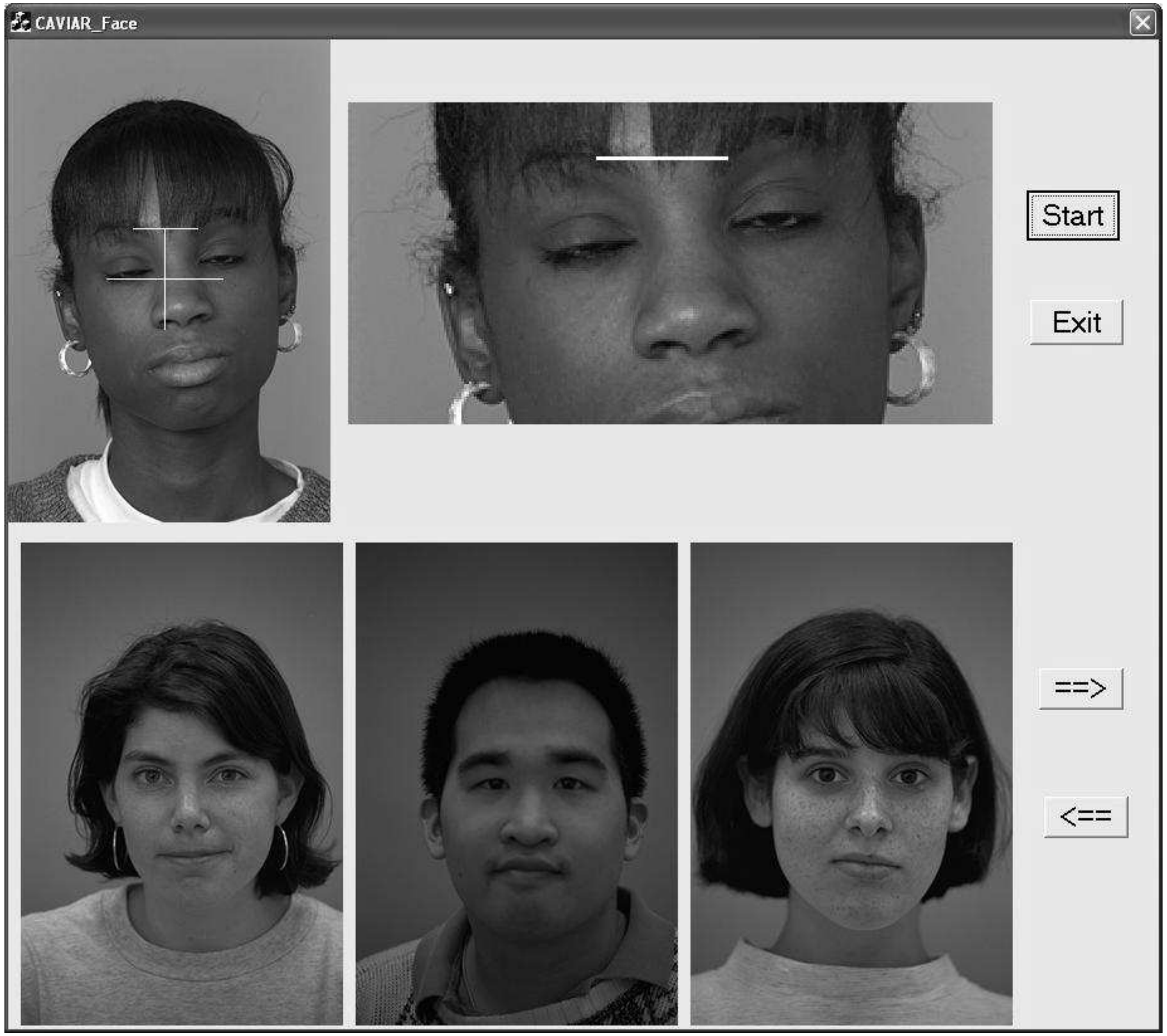
CAVIAR-FACE



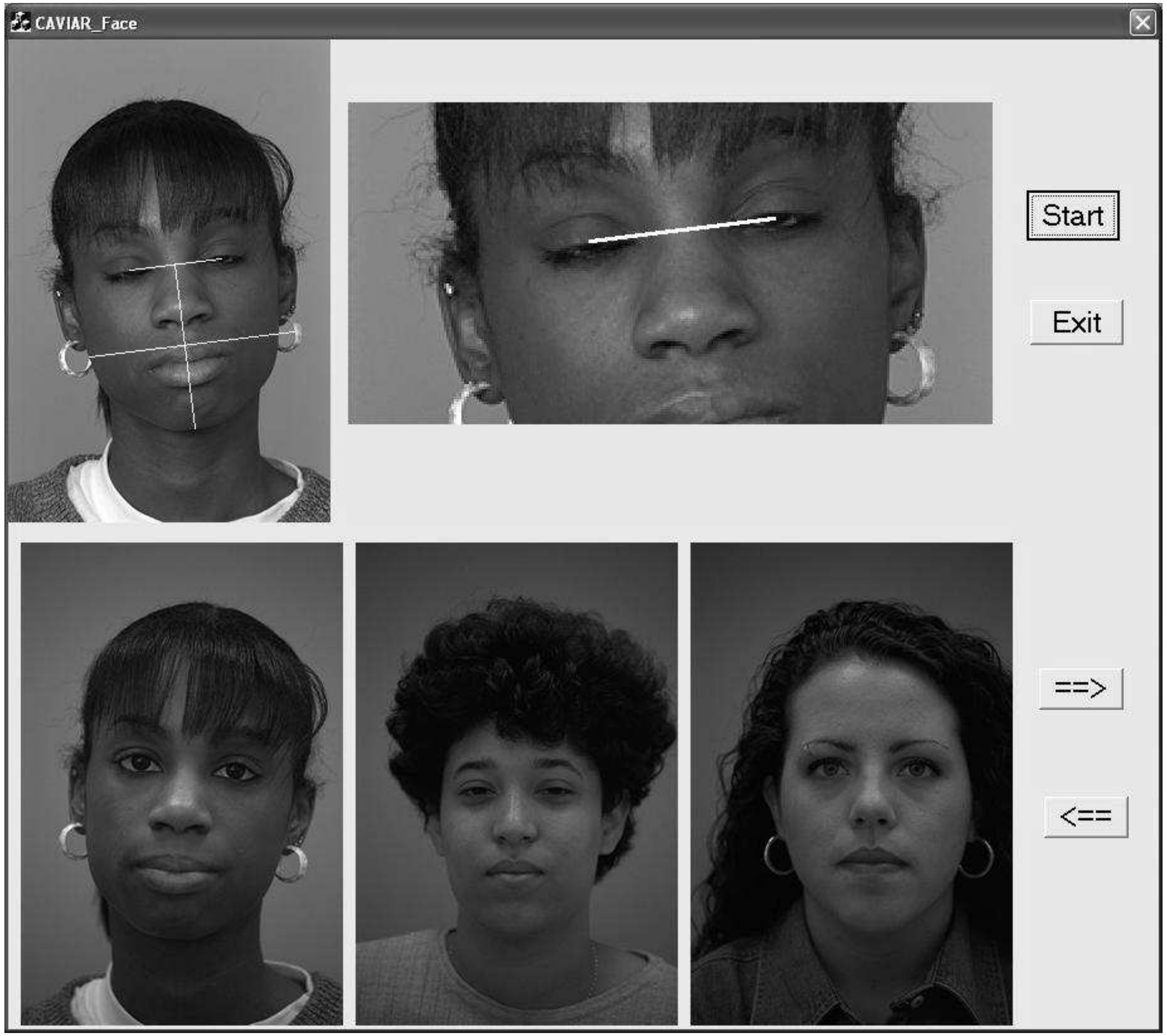
GUI designed for accurate pupil location



GUI before model adjustment



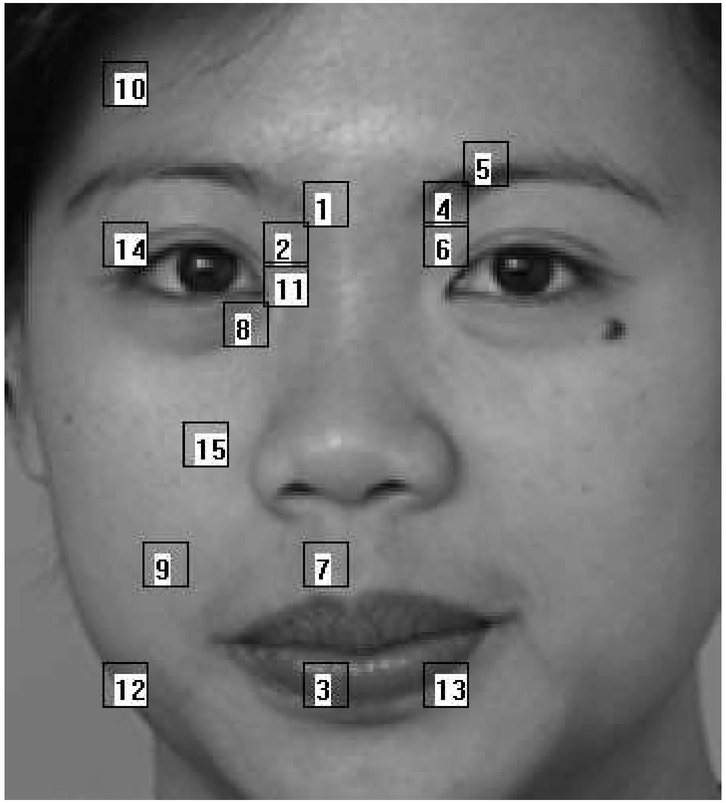
GUI after model adjustment

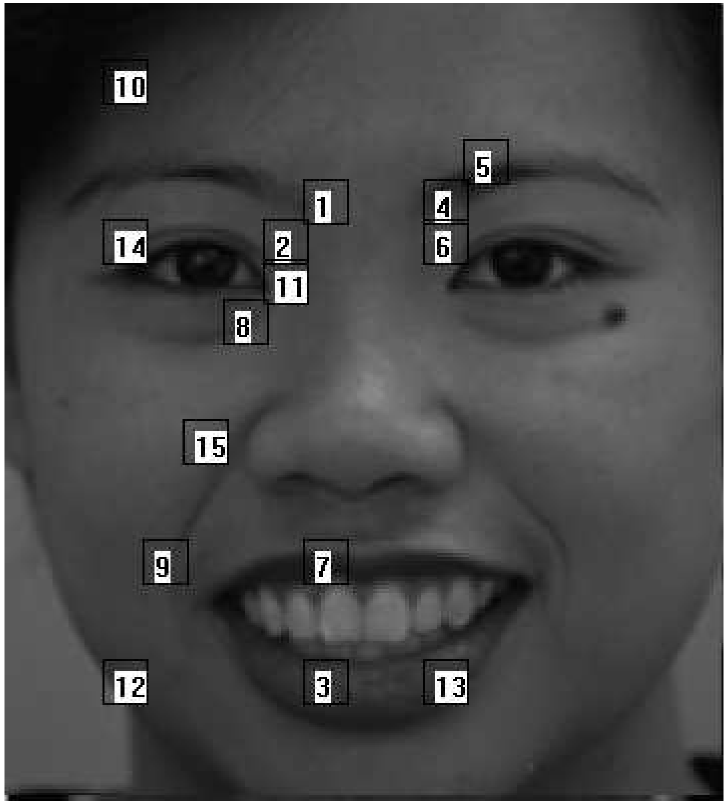


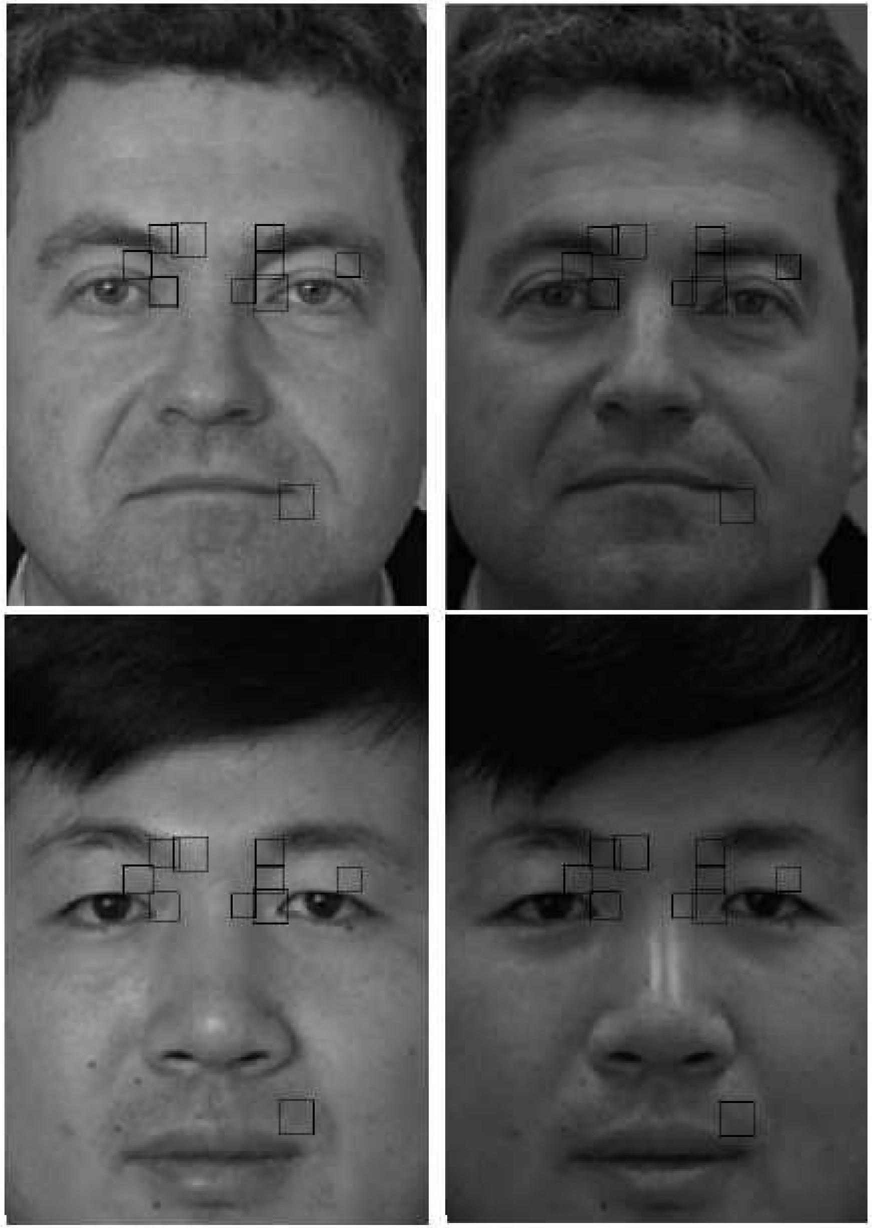
FEATURE TEMPLATES
(best 15 of 240 candidates)

Most discriminating features near, but not on, eyes.

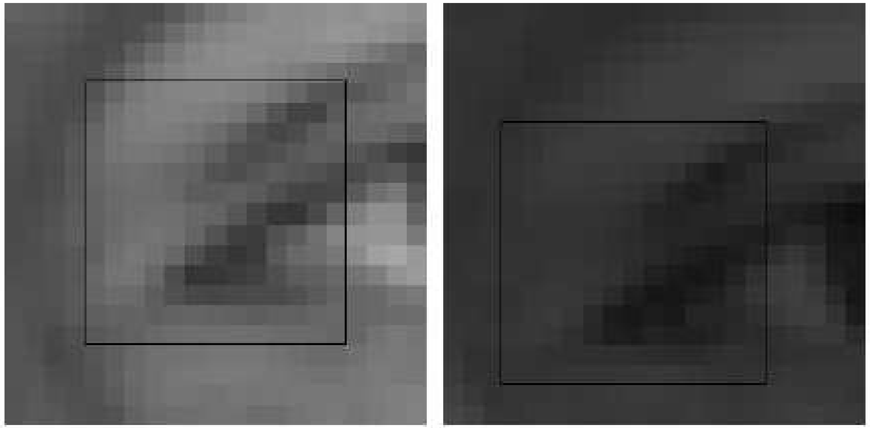
Single best feature yields 40% accuracy on 200 classes!

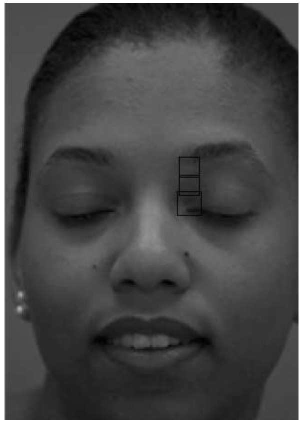




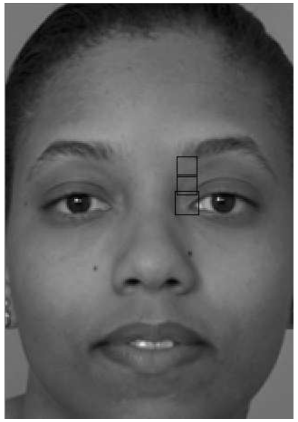


Search over a 5x5 window





GalleryEASY AND DIFFICULT FERET PAIRS




Probe

G1 G4

Gallery

TEMPLATES

	Gallery (reference) faces G1		G2		G3		G4		G5	
	Similarity	Rank	Similarity	Rank	Similarity	Rank	Similarity	Rank	Similarity	Rank
P1	0.999501	1	0.997885	5	0.997886	4	0.998195	2	0.998056	3
P2	0.997412	2	0.997273	3	0.997989	1	0.996801	5	0.997120	4
P3	0.970771	2	0.960403	5	0.964492	4	0.975555	1	0.970332	3
Borda Count		5		13		9		8		10
Final Rank		1		5		3		2		4

FEATURE EXTRACTION AND CLASSIFICATION

Affine size normalization based on model

Local histogram equalization on template surround

Cosine similarity measure on 11x11 feature templates

5x5 search window for each template

Features selected by agglomerative search

Borda Count classifier based on rank order

(usually only five features required for Top-3)

Difficult face-pairs require more features,

but only extracted from leading candidates

Other experiments on pose, expression, aging, …

CAVIAR-FACE INTERACTIONS (6 subjects, 200 faces)

CAVIAR-FACE COMPARED TO MACHINE ALONE AND TO HUMAN ALONE (200 faces)

200 BK pictures as *gallery*, 50 BA pictures as *probes*, 6 subjects

	Accuracy (%)	Time per face (seconds)
Interactive	99.7	7.6
Machine Alone	47.0	~0
Human Alone	--	66.3

COMPUTER BASED INTERACTIVE RETRIEVAL vs. CAVIAR

CBIR	CAVIAR
Subjective retrieval	Objective classification
User judges retrieval results	Statistical decision boundary
User weights features	Machine weights features
Broad domain	Narrow domain
Relevance feedback	Relevance feedback
	Model adjustment

(EXPANDED) MESSAGE

Interactive recognition is faster than unaided human, and more accurate than unaided machine (without years of R&D).

Parsimonious interaction *throughout* the process is better than only at the beginning or end.

Interactive systems can be initialized with *a single training sample per class,* and *improve with use*.

Interaction with images requires a *visible model* that is accessible to both man and machine.

Let both do what they do best: let human help in segmentation.

Leave the *human in charge*.

Read *IEEE-PAMI* diligently.

MESSAGE (cont'd)

Make use of language models at all possible levels

Exploit single-pattern style (i.e. consistency) using multimodal classifiers and *adaptation*

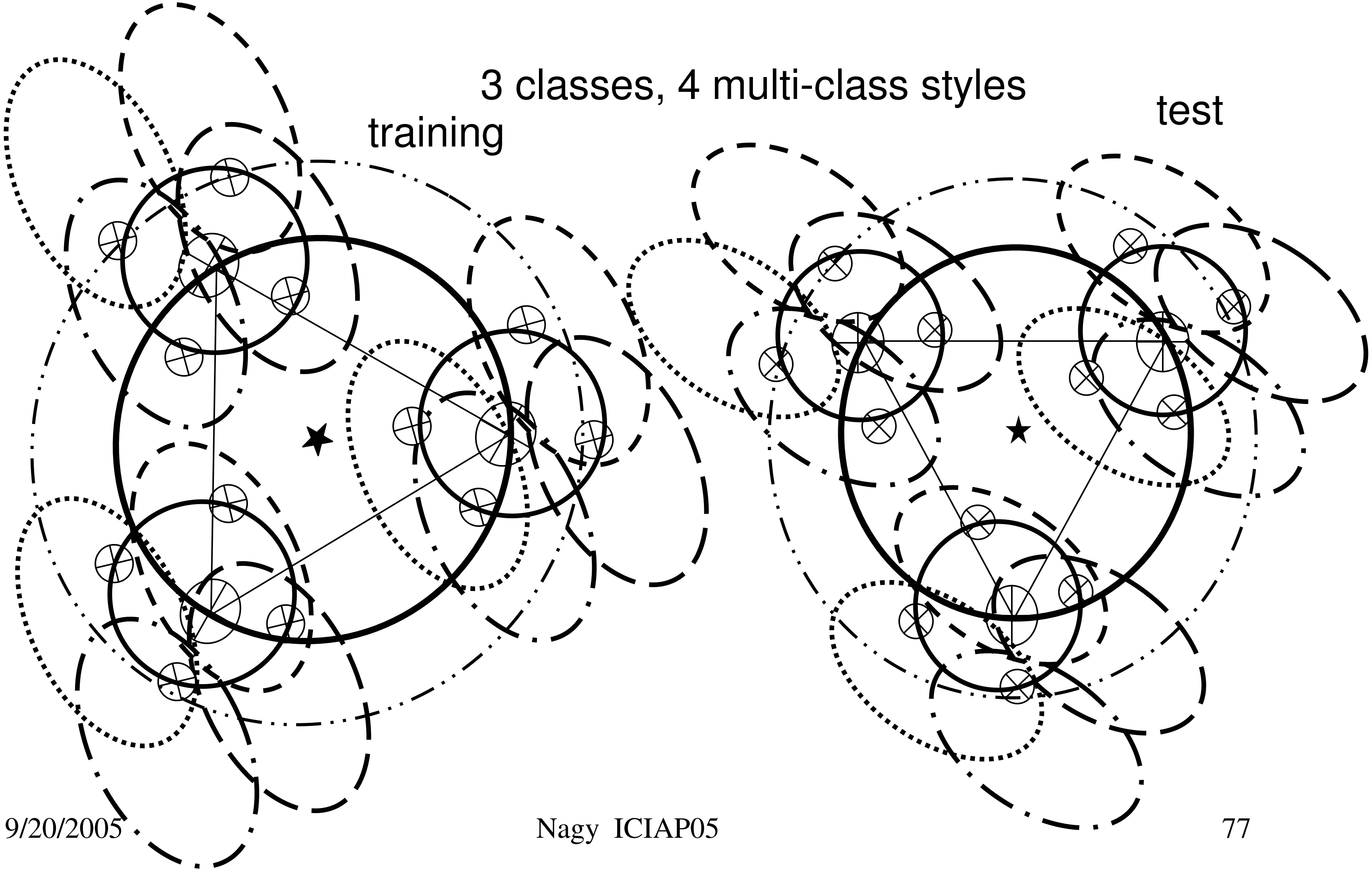
Classify entire fields to exploit multi-pattern style

Thank you

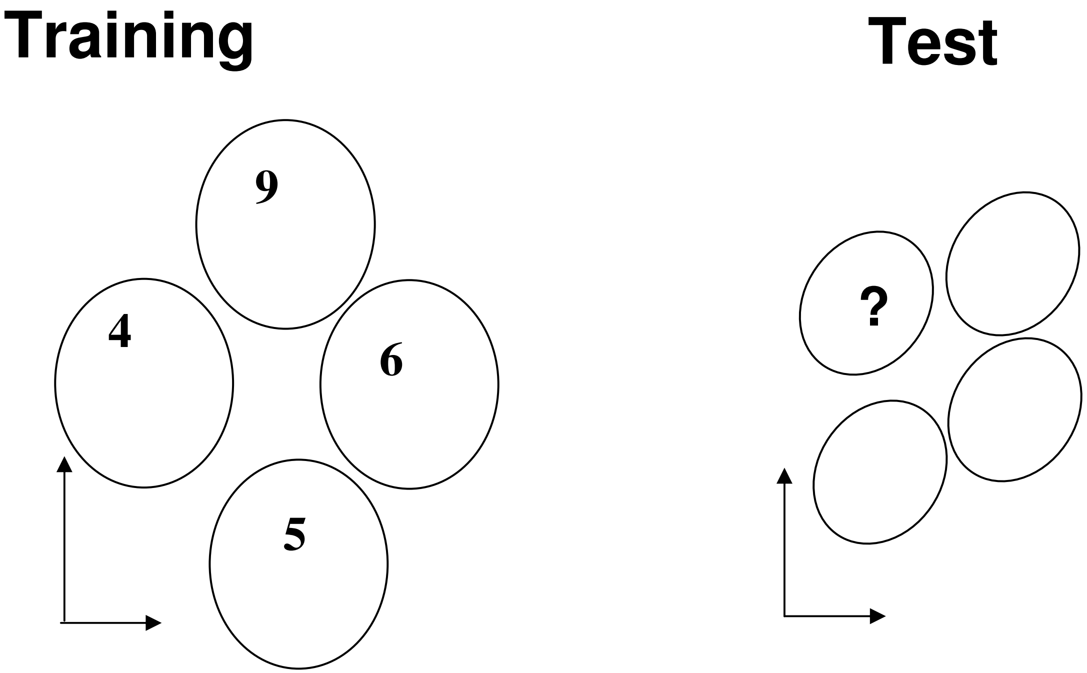
www.ecse.rpi.edu/doclab/vpr.pdf

WEAKLY CONSTRAINED DATA

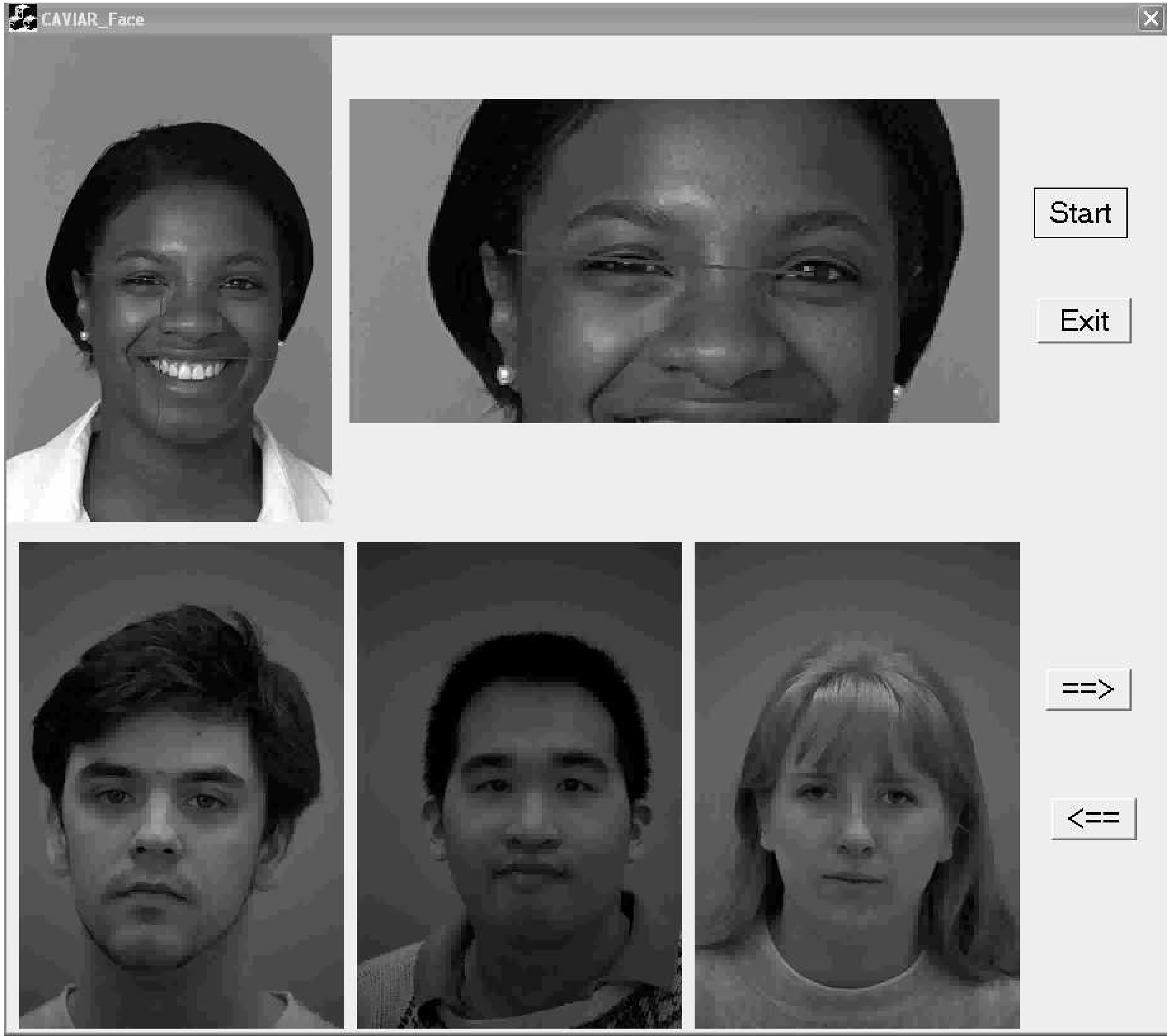
given $p(\mathbf{x})$, find $p(\mathbf{y})$, where $\mathbf{y}=g(\mathbf{x})$



Are weak constraints enough?

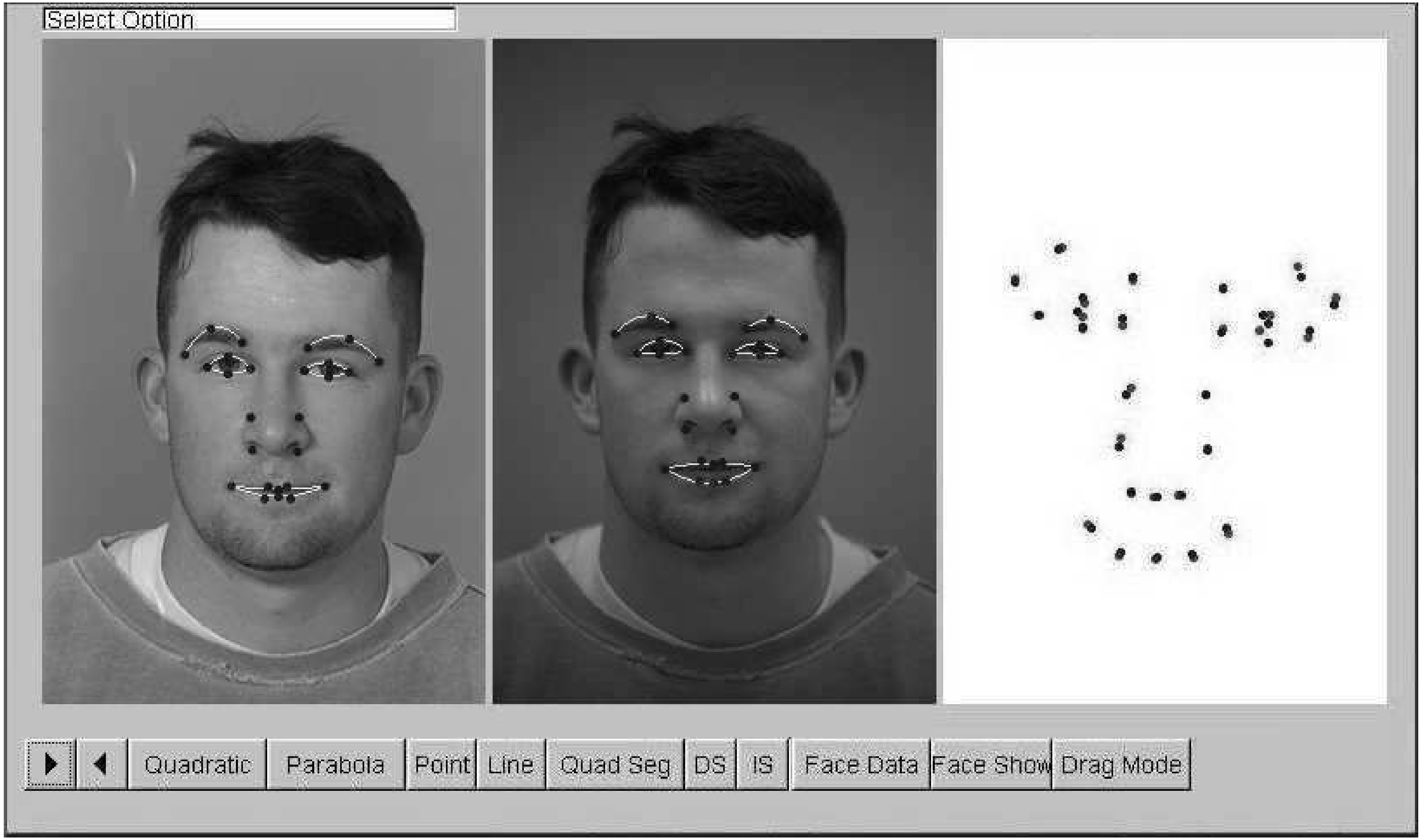


GUI (continued)

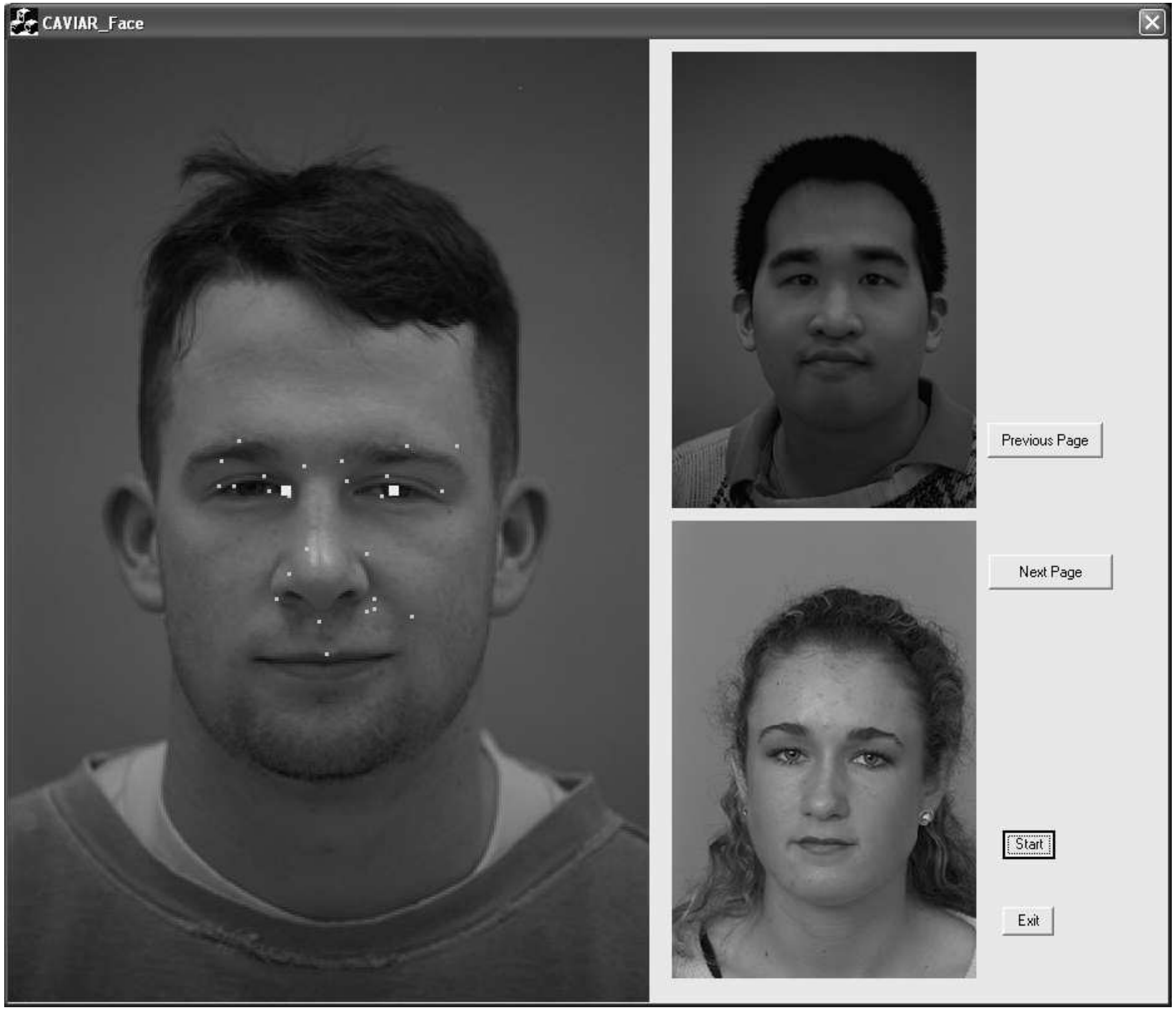


CAVIAR-FACE: FIDUCIAL POINTS AFTER SIMILARITY TRANSFORM

Matt Green



CAVIAR-FACE (BAD PUPIL LOCATION)



CAVIAR-FACE (GOOD PUPIL LOCATION)

MISRECOGNIZED FACES