Elements of Modern Face Recognition

Erik Learned-Miller

Joint work with Vidit Jain, Gary Huang, Andras Ferencz, and others

Face Recognition Technology is Here to Stay

Questions about Face Recognition

- How hard can it be?
- What's it good for today?
  - What about in the near future?
- What are the underlying technologies?
  - Hyperfeatures for recognition.
  - Congealing for alignment.
- How well does it work?
  - How do we characterize performance?

Questions about Face Recognition

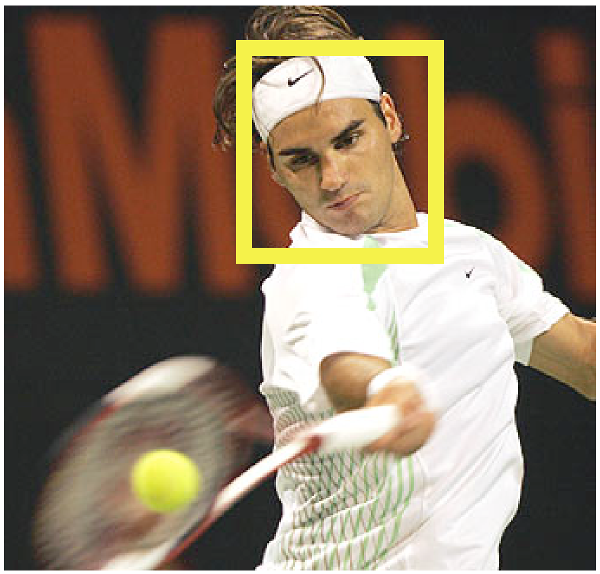
- How hard can it be?
- What's it good for today?
  - What about in the near future?
- What are the underlying technologies?
  - Hyperfeatures for recognition
  - Congealing for alignment
- How well does it work?
  - How do we characterize performance?

How hard is face recognition?

- Humans think of face recognition as trivial.
- For machines, it is much harder than
  - Playing chess,
  - Doing large integrals.
- Failures of human face recognition illustrate some of the difficulties.

Detection-Alignment-Recognition Pipeline

Detection



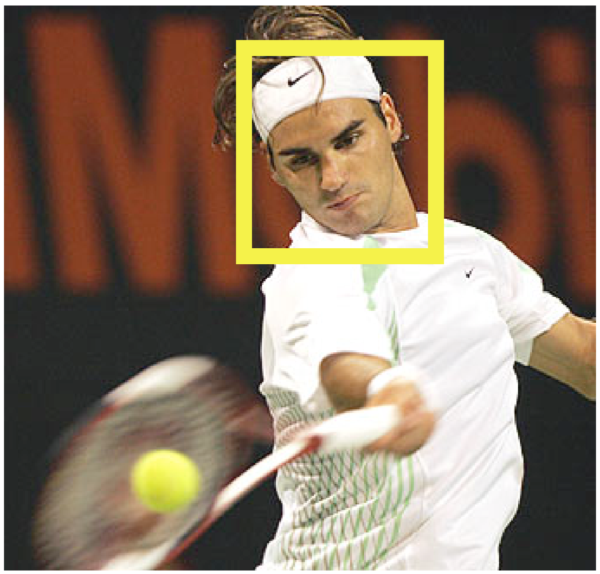
Alignment

Recognition

“Same”

Detection-Alignment-Recognition Pipeline

Detection

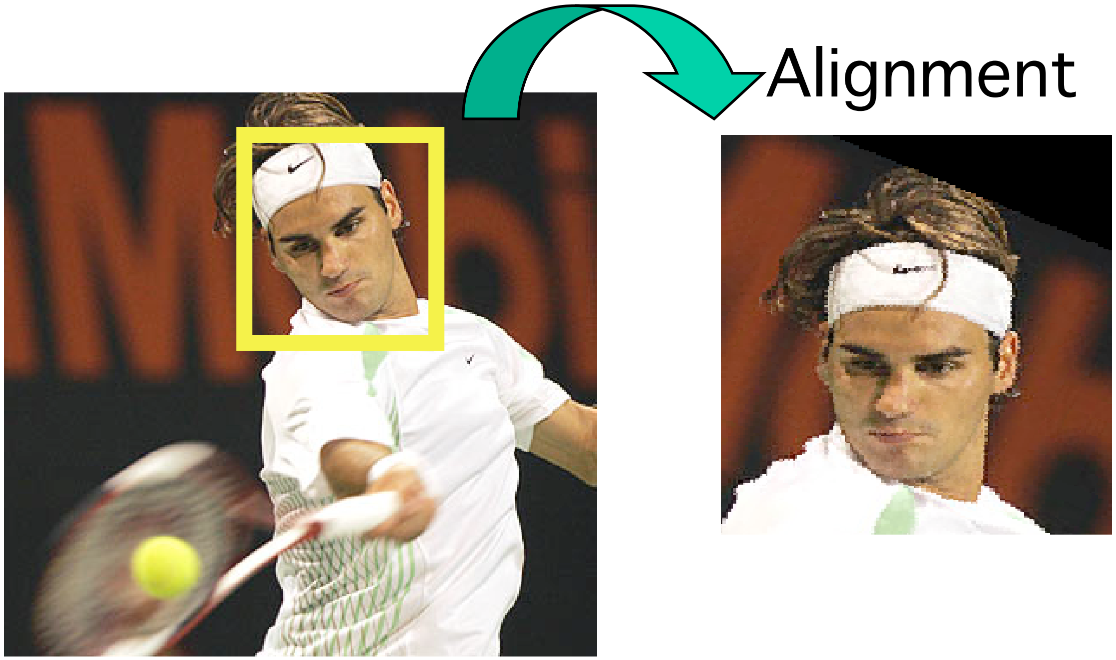


Harder than you might expect...

Face detection

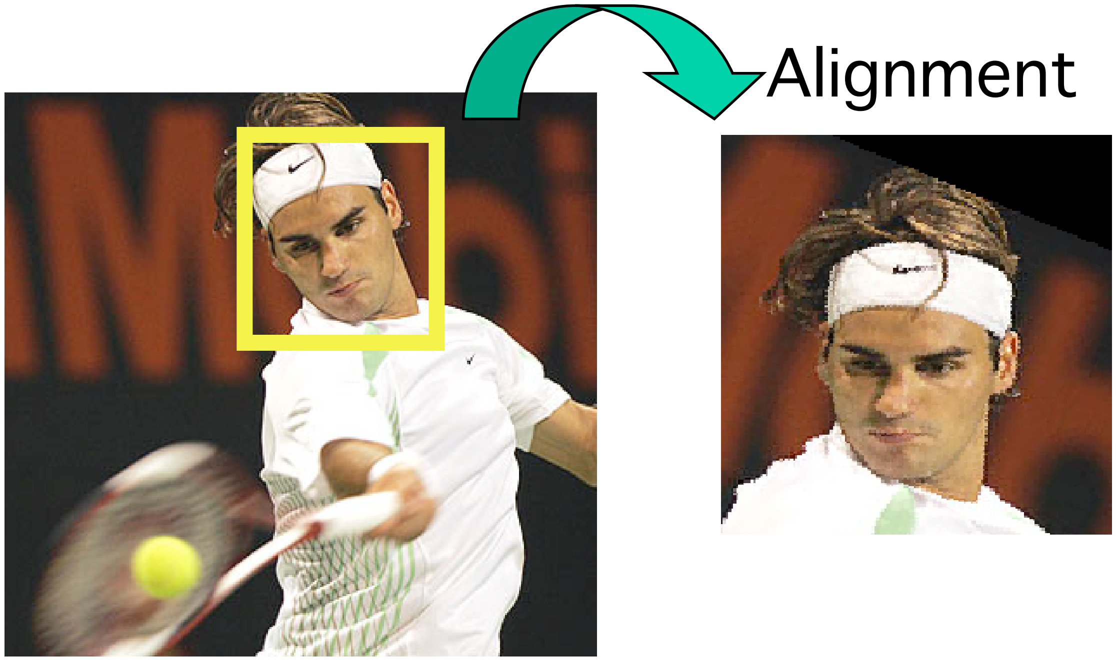
Image by "Furitsu"
From Michael's "Visual Phenomena & Optical Illusions"

Alignment



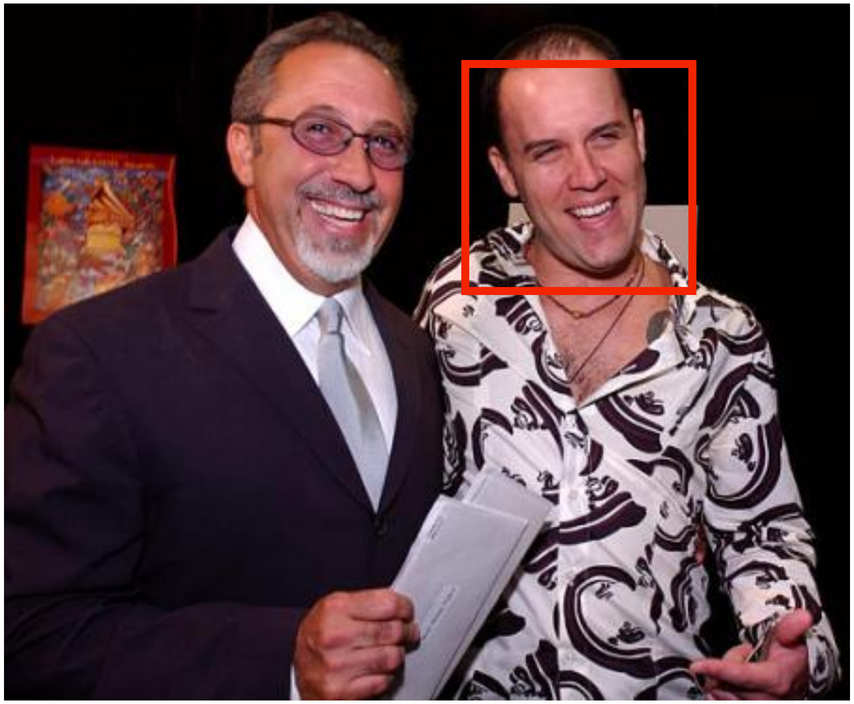
Alignment

- *Surprisingly important for recognition algorithms...*



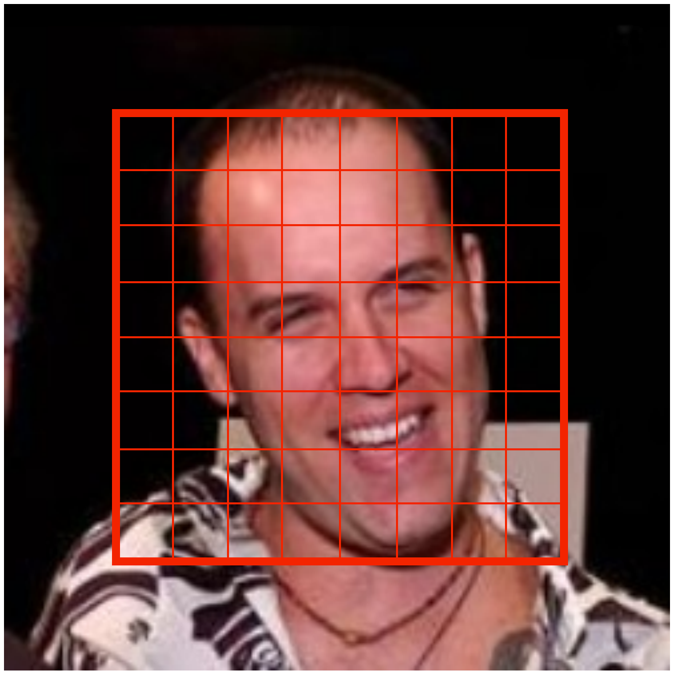
Original pictures...

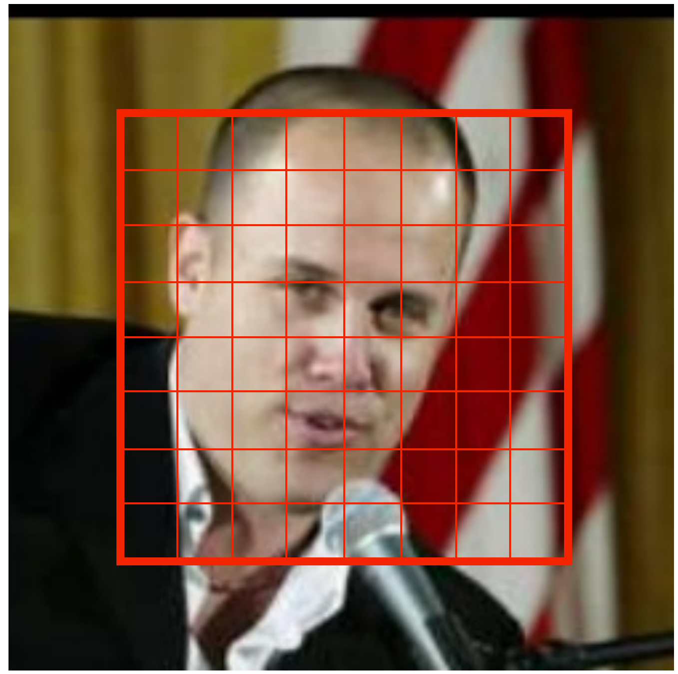
After detection...



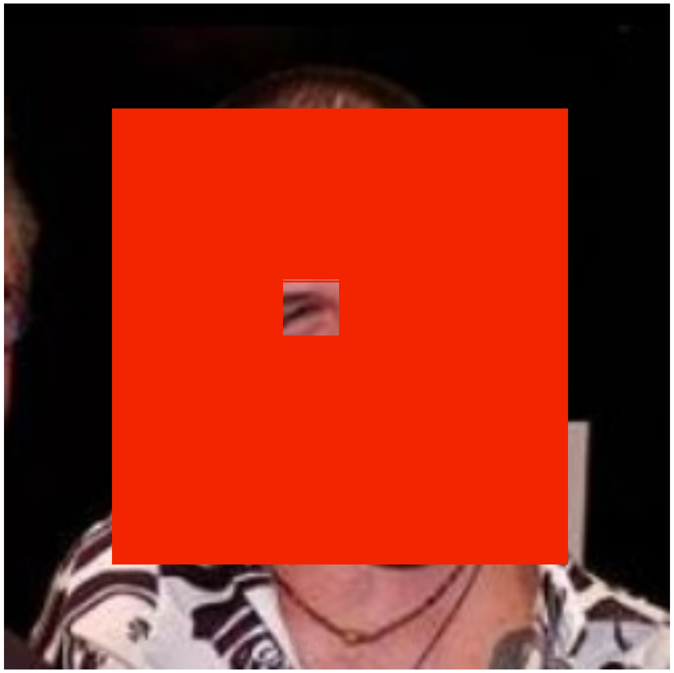
Cropping...

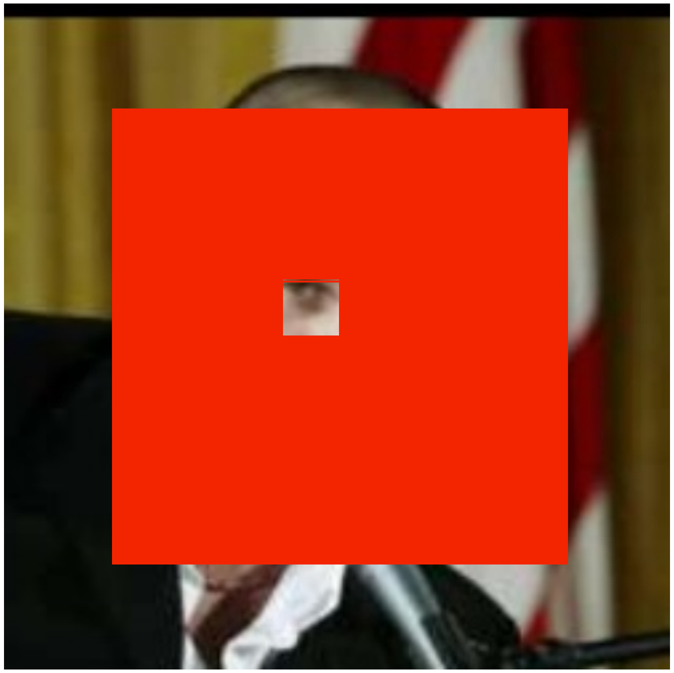
Patchwise comparison...





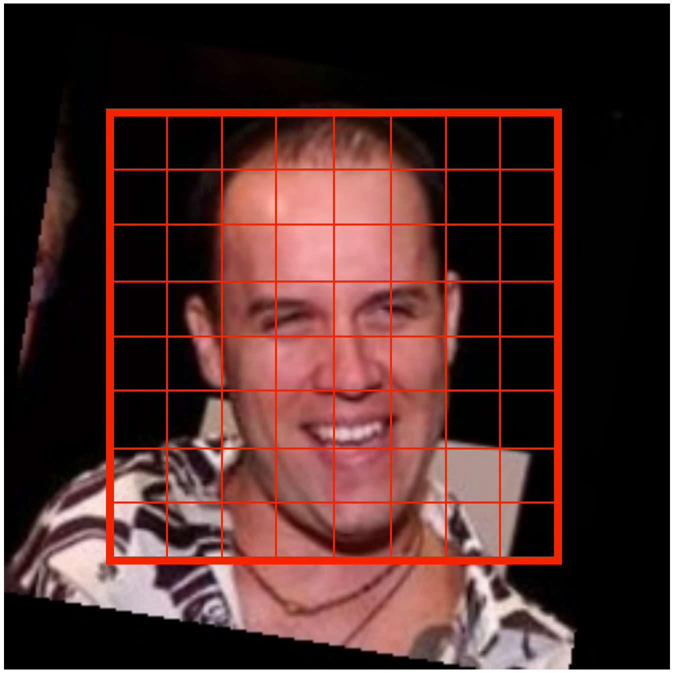
Differences are too large for successful recognition

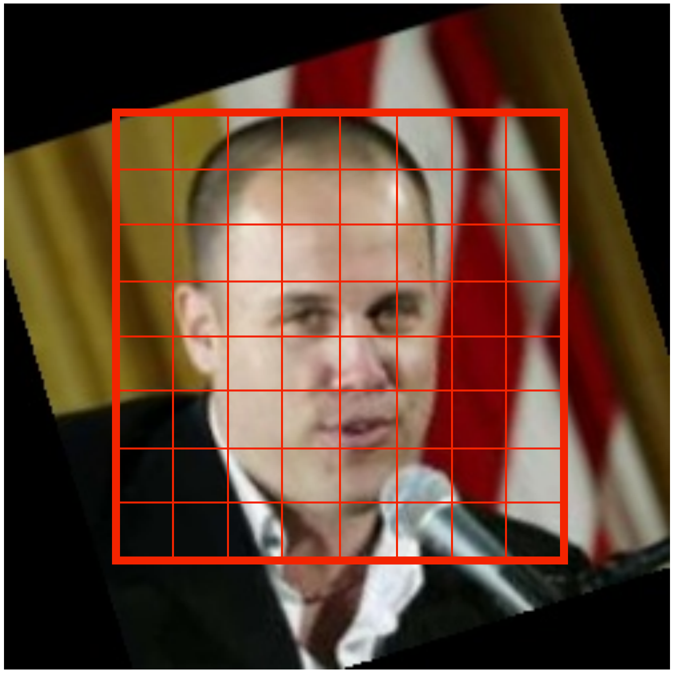




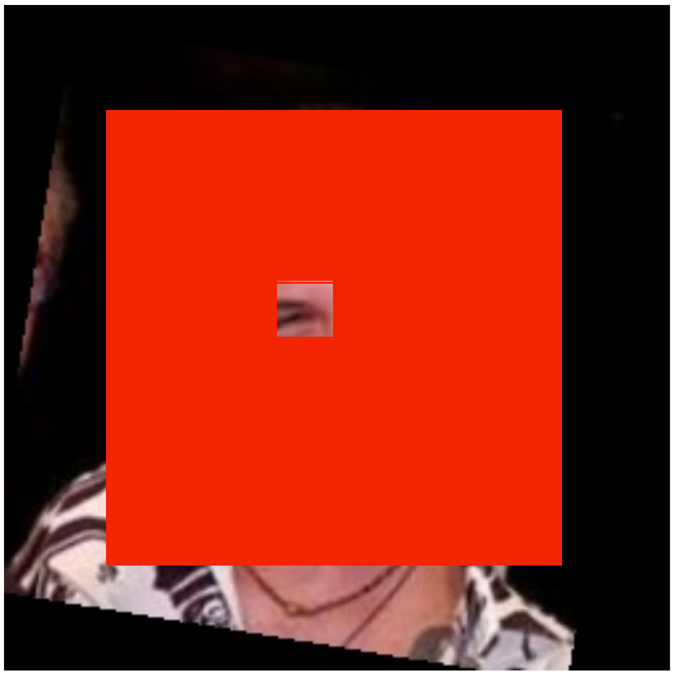
Cropping...

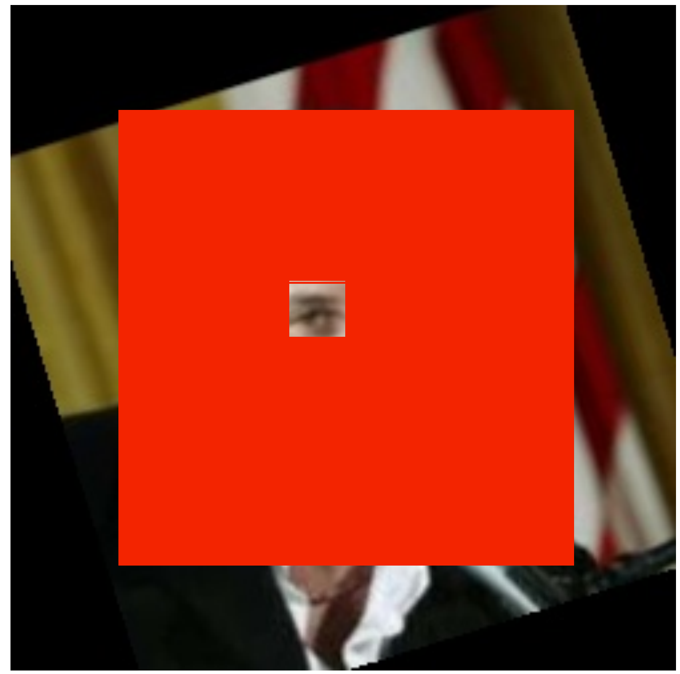
Improved alignment





Recognition greatly improved...





Alignment and human perception

- We’re not usually aware of it, but alignment can dramatically affect our ability to interpret images.

Does alignment affect human recognition?

Schwaninger et al., 2003

Does alignment affect human recognition?

Schwaninger
et al., 2003

Recognition

Recognition: More Than Meets the Eye

from www.coolopticalillusions.com

So far

- Three main subtasks in face processing:
  - Detection
  - Alignment
  - Recognition
- For humans:
  - Usually easy, but each has its failure modes
- For machines:
  - All areas still far inferior to humans.

Questions about Face Recognition

- How hard can it be?
- What's it good for today?
  - What about in the near future?
- What are the underlying technologies?
  - Hyperfeatures for recognition.
  - Congealing for alignment.
- How well does it work?
  - How do we characterize performance?

Some Applications

- Access control →

- Watch lists 
- Organizing personal photo collections

- Image retrieval: web search for pictures

- many others...

Errors

- Applications must be tolerant to errors.
- Different applications have different failure modes.
- Error types have different costs.
- Can usually make trade-offs between error types.

Example 1

- Personal photo collections:
  - Type 1 error: Failed to label a picture of Erik.
  - Type 2 error: Identified Steve as Erik.
- Both of these are cheap to fix.
- When confidence is low, don’t label.

Example 2

- Watch lists: looking for Osama Bin Laden at JFK.
- Error types
  - False positive: Identify John Doe as Bin Laden.
  - False negative: Fail to identify Osama Bin Laden.
- Error costs:
  - False positive: Easily corrected by human.
  - False negative: Extremely costly.
- Problem: no good trade-off.
  - Either we miss Bin Laden 90% of the time, or we make millions and millions of false ID's.

Current trends

- Algorithms are slowly becoming more accurate.
- Manufacturers looking hard for apps that can tolerate errors.
- Apps that require very high accuracy rates will rely on humans for the foreseeable future.

Questions about Face Recognition

- How hard can it be?
- What's it good for today?
  - What about in the near future?
- What are the underlying technologies?
  - Hyperfeatures for recognition.
  - Congealing for alignment.
- How do we characterize performance?

Face Recognition Paradigms

- Two dominant paradigms:
  - Recognition with a fixed “gallery”
  - Face verification or pair matching

Gallery Recognition

Fixed Gallery of Registered People

Yanlei

Erik

Hanna

Sridhar

Rick

Question: Is this person one of the registered people?

Gallery Recognition

Fixed Gallery of Registered People

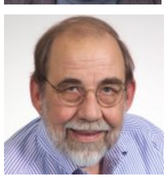
Yanlei

Erik

Hanna

Sridhar

Rick



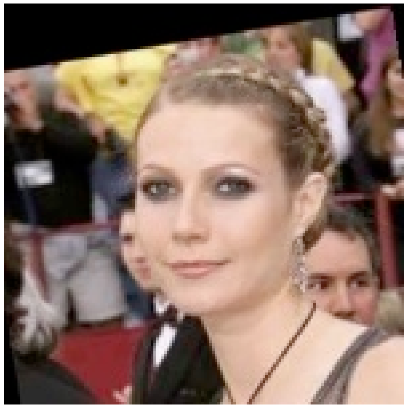
Question: Is this person one of the registered people?

becomes much more difficult as gallery size increases.

Face Verification (pair matching)

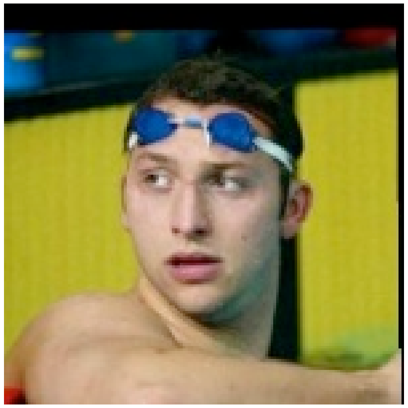
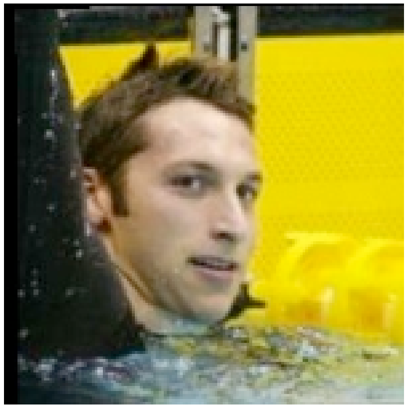
Are these the same person?

Sample Face Verification Problem



Sample Face Verification Problem

Sample Face Verification Problem

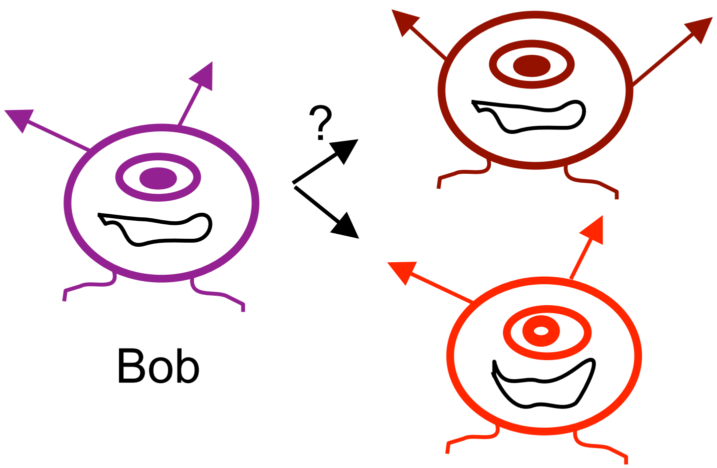
Some problems are just too hard to solve

Face Verification

- How can we teach a computer to do this by giving it examples?
- How do humans learn such a task?

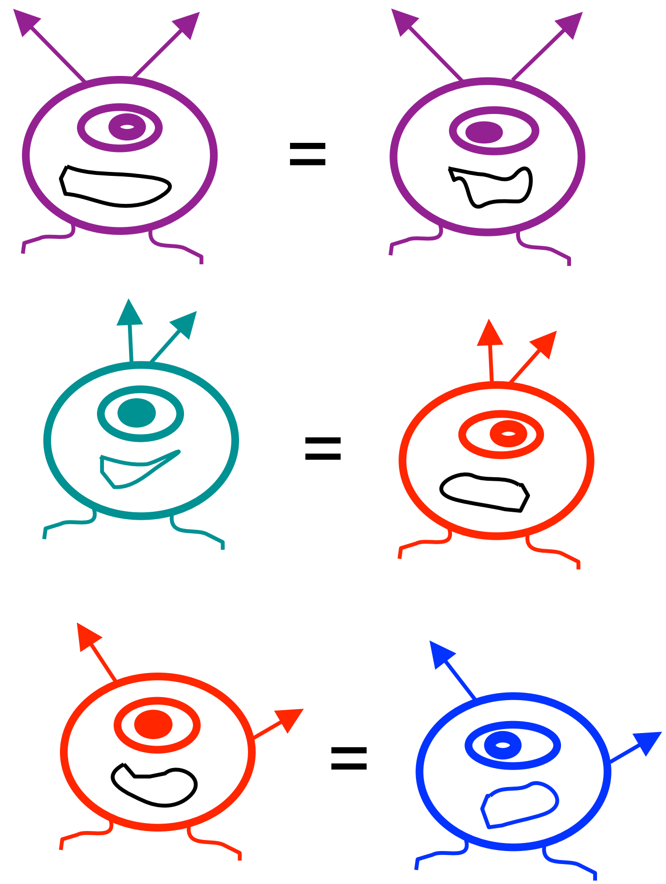
Crash Course on Martian Verification

Test: Find Bob after one meeting



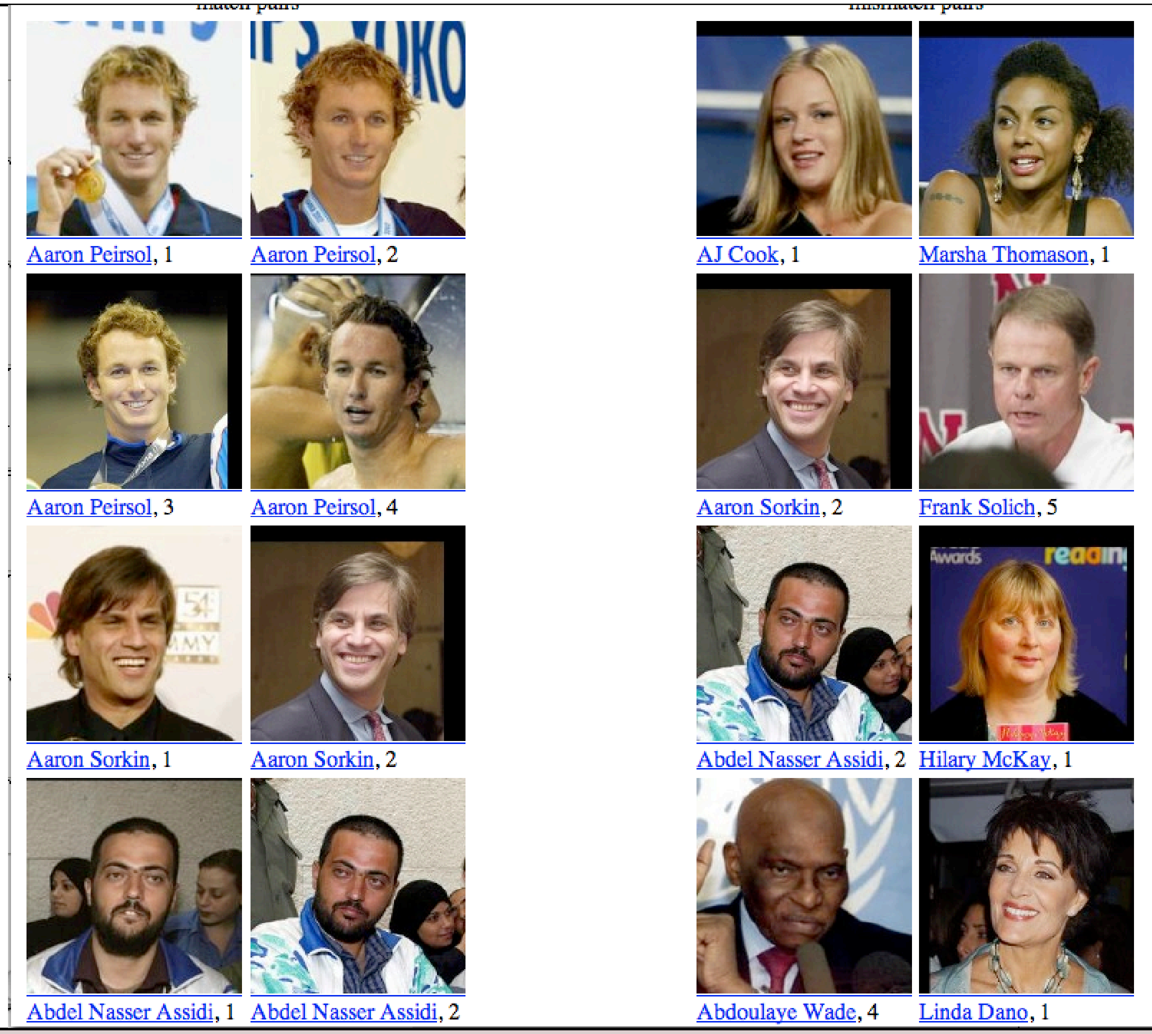
Bob

Martian training set



Training Data for Human Verification

"same"

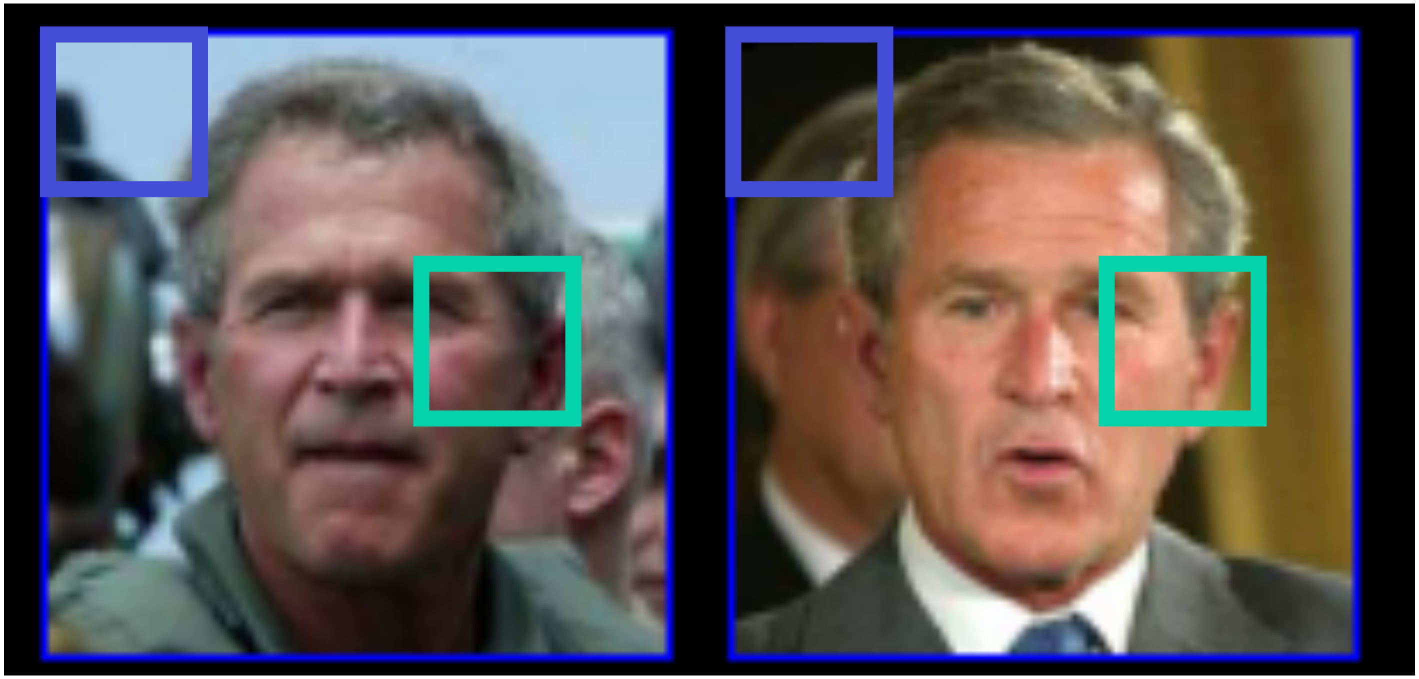


"different"

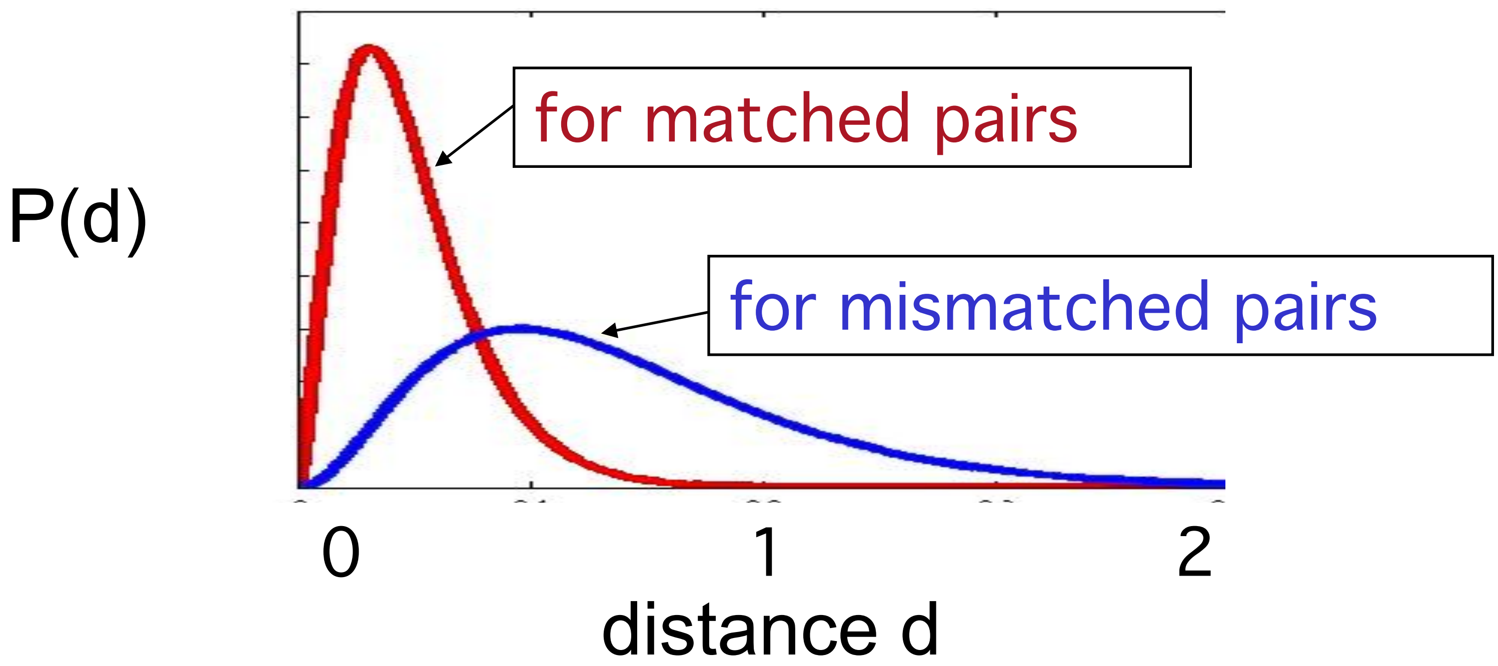
Consider pairs of patches

Let d be "distance" between a pair of patches.

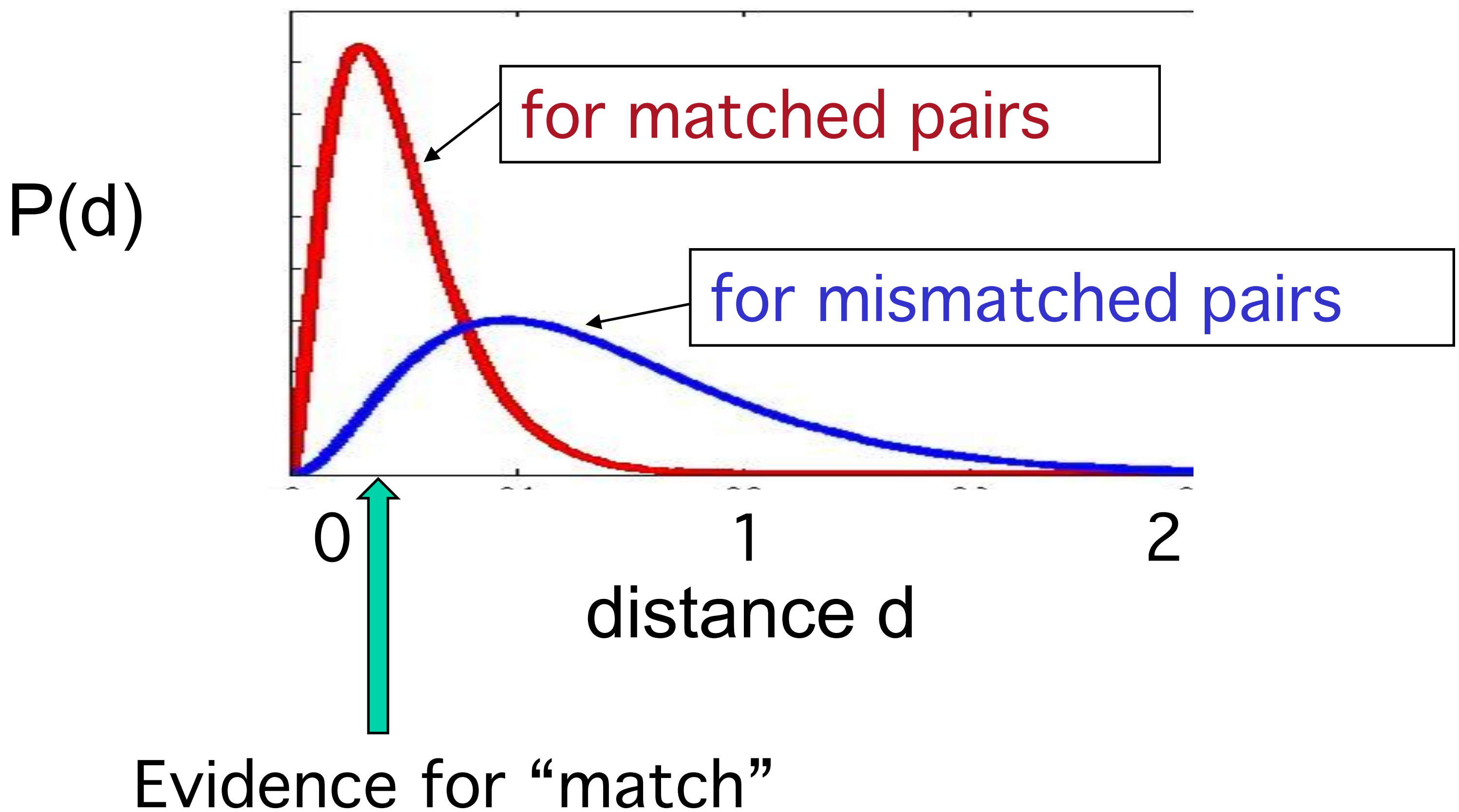
d = 1-correlation(patch1, patch2).



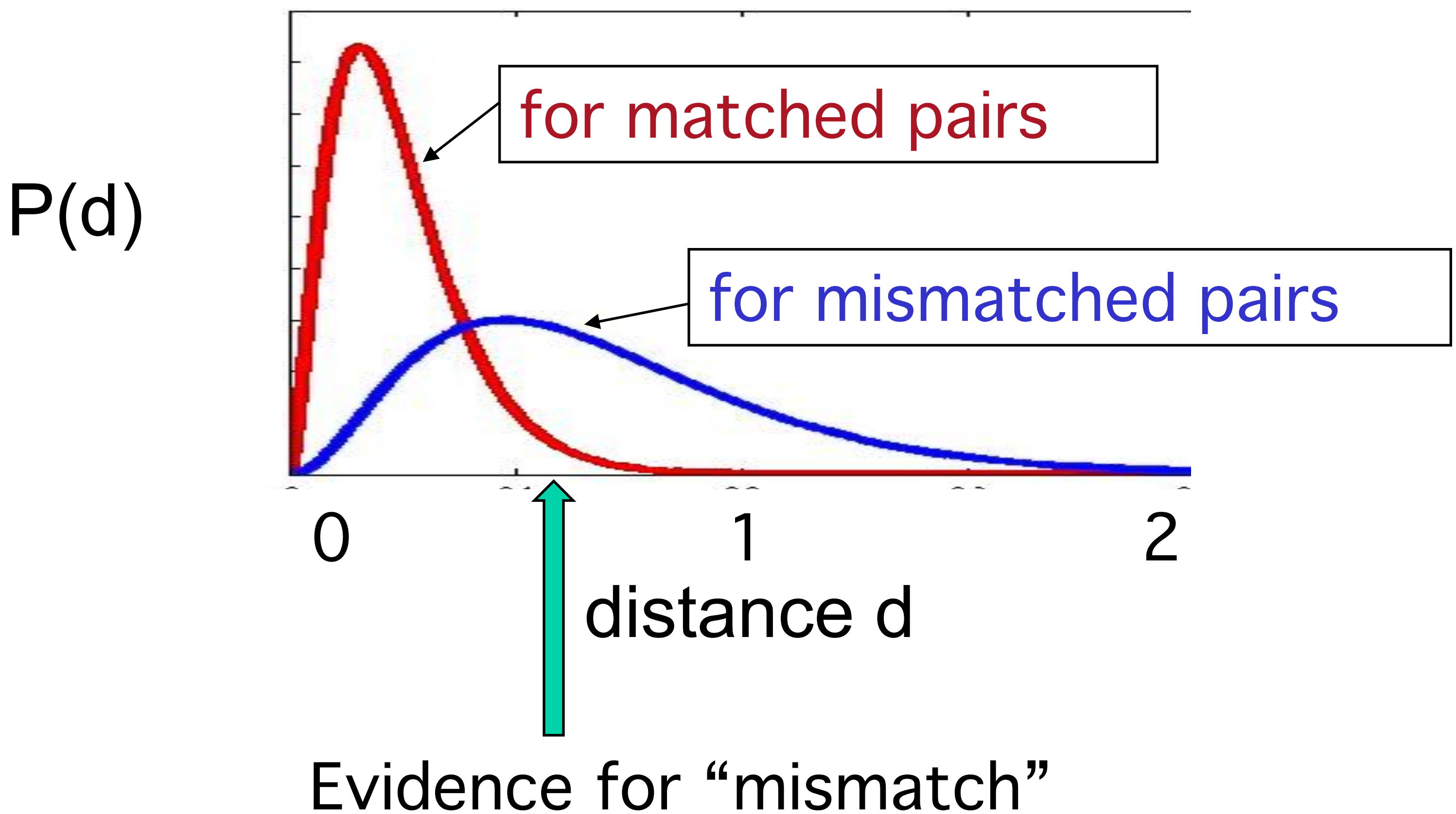
Distributions over distances after "learning"



For a new image pair…



or…


P(d)
for matched pairs
for mismatched pairs
0
1
2
distance d
Evidence for “mismatch”

Computing the likelihood of same and different

N: number of patches in an image

d_i: distance between the ith pair of patches

$D = \{d_1, d_2, ..., d_N\}$: set of d_i for all patches

$$Prob(D|\texttt{same}) = \prod_{i=1}^{N} (d_i|\texttt{same})$$

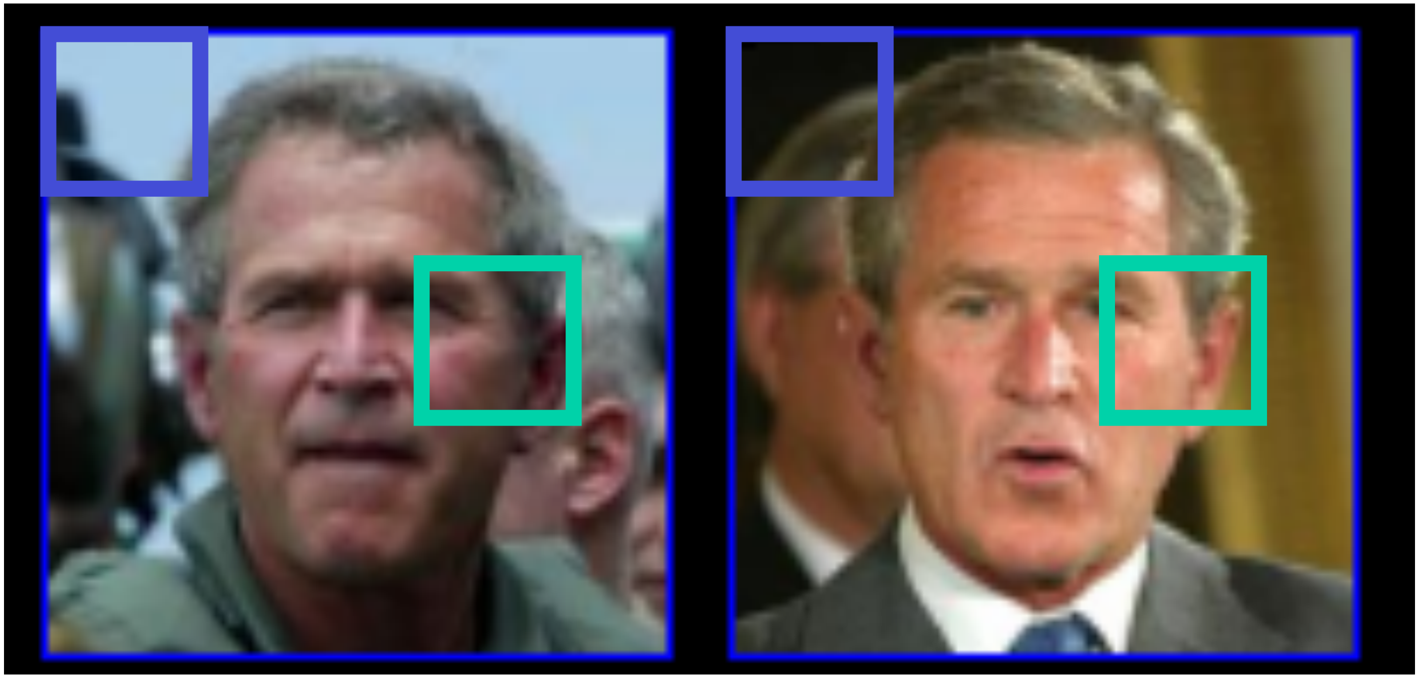
$$Prob(D|\texttt{diff}) = \prod_{i=1}^{N} (d_i|\texttt{diff})$$

Three models for distributions over distances

1. Universal patch model:
   P(d|same)
   P(d|different)
2. Spatially dependent patch model:
   P(d|same,x,y)
   P(d|different,x,y)
3. Hyper-feature dependent model:
   1. P(d|same,x,y,appearance of left image)
   2. P(d|different,x,y,appearance of left image)

Universal Patch Model

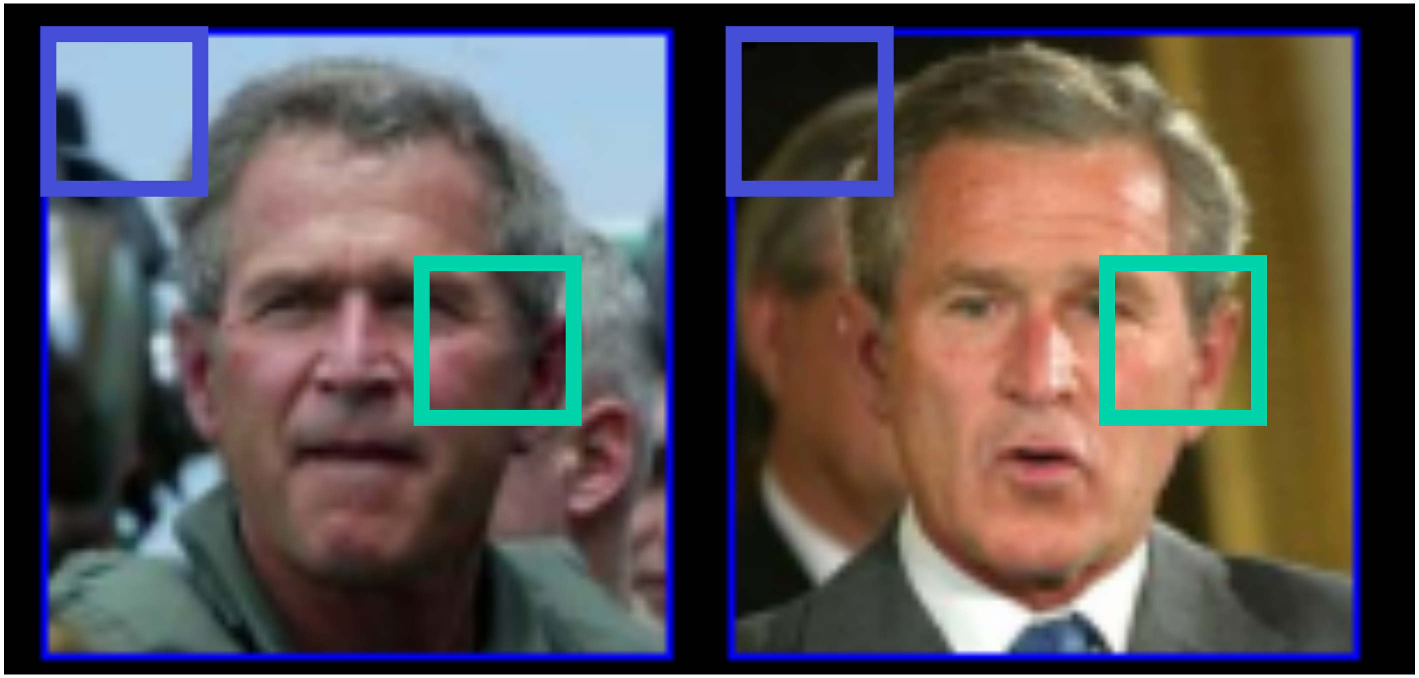
A single P(d | same) for all patches



Different blue patches are evidence against a match!

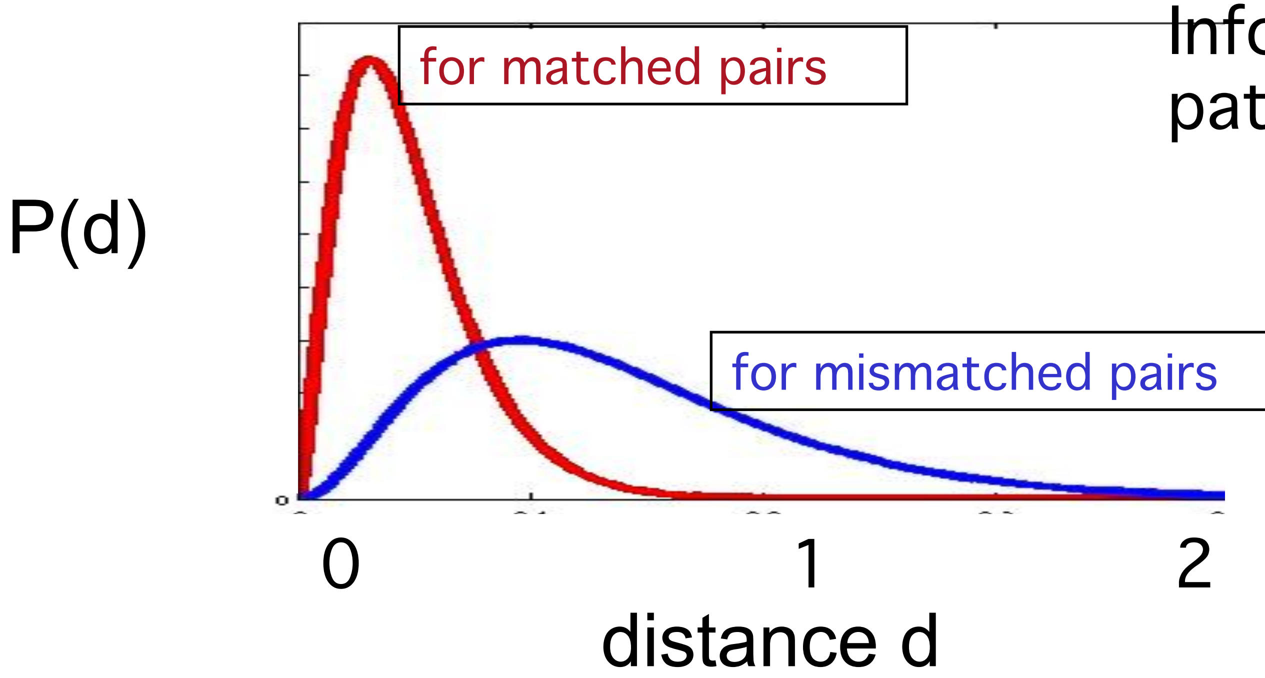
Spatial Patch Model

P(d|same,x_1,y_1) estimated separately from P(d|same,x_2,y_2)

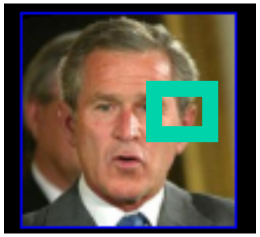


Greatly increases discriminative power of model.

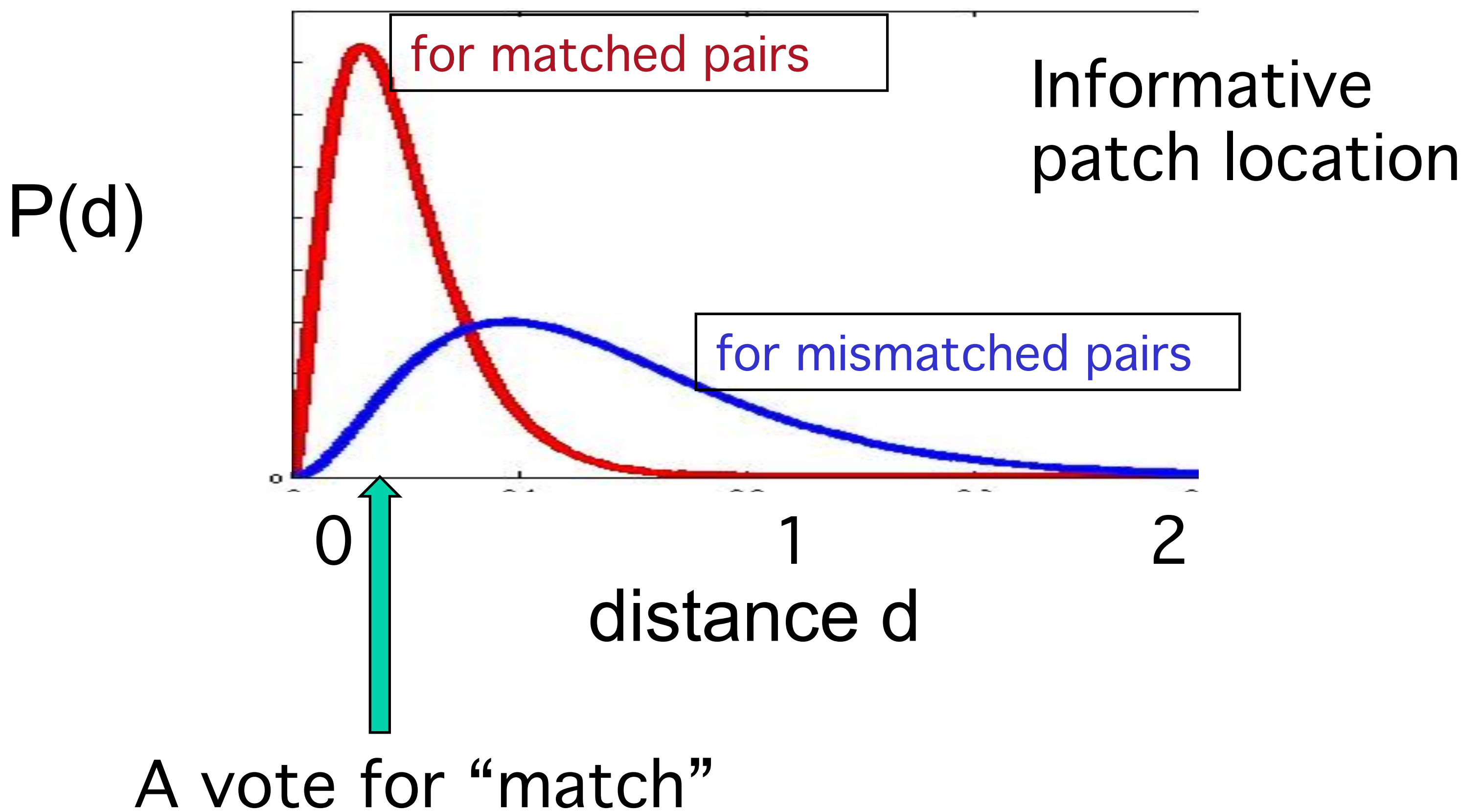
Distributions over distances for specific locations



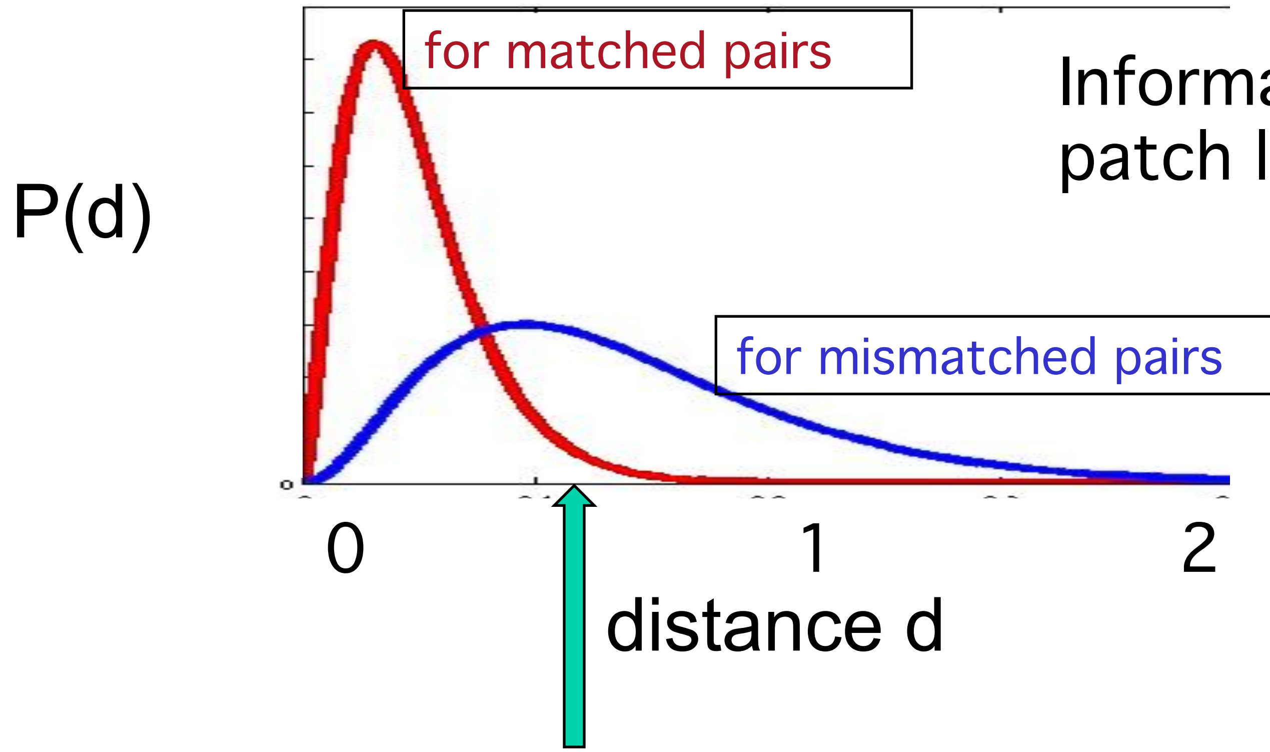
Informative
patch location



Distributions over distances for specific locations



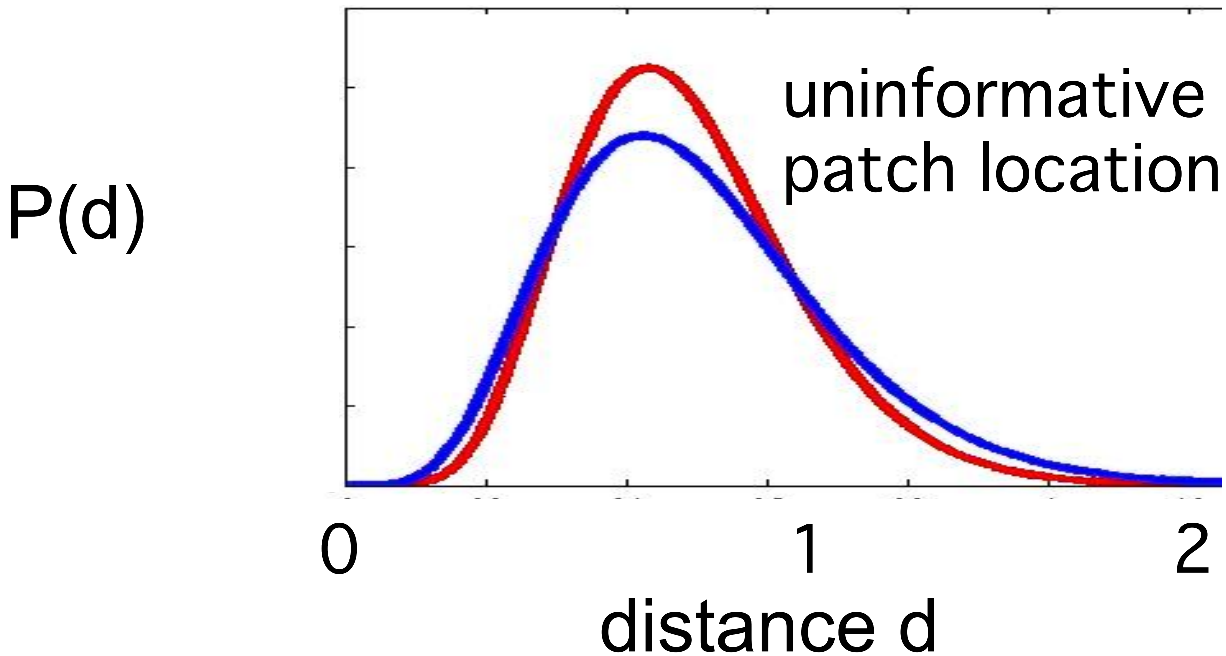
Distributions over distances for specific locations



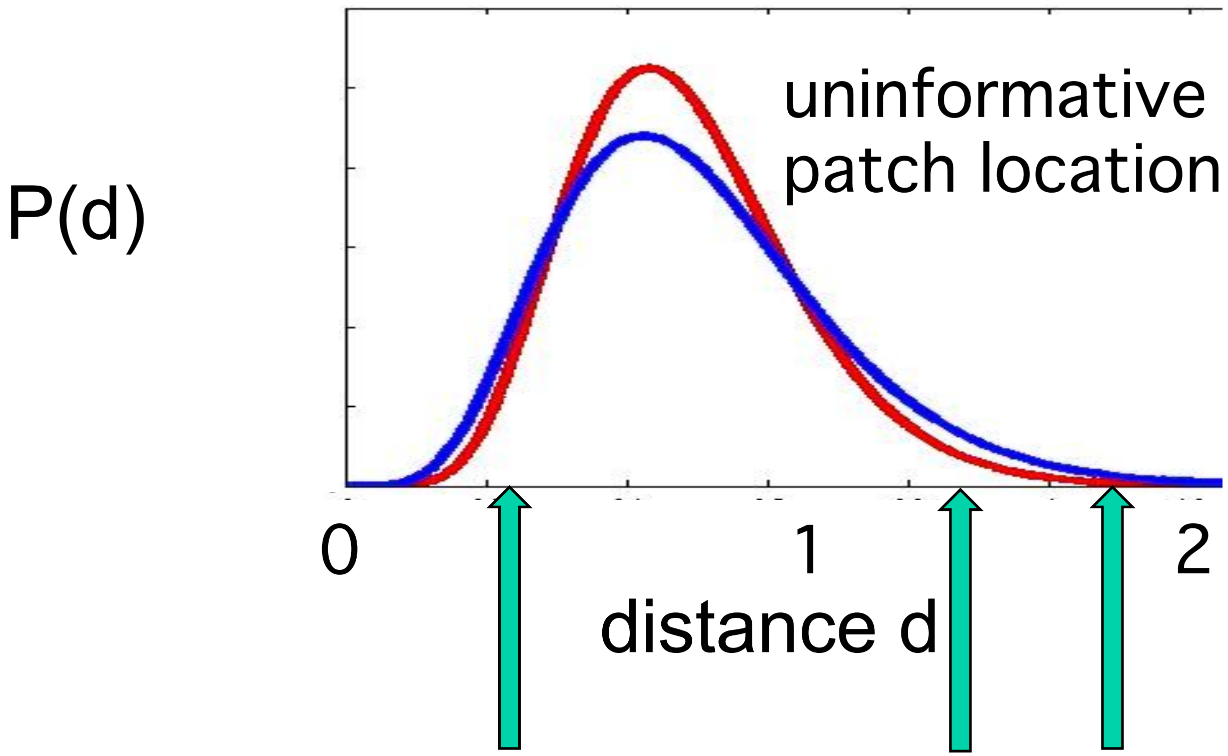
Informative
patch location

A vote for "mismatch"

Distributions over distances for specific locations



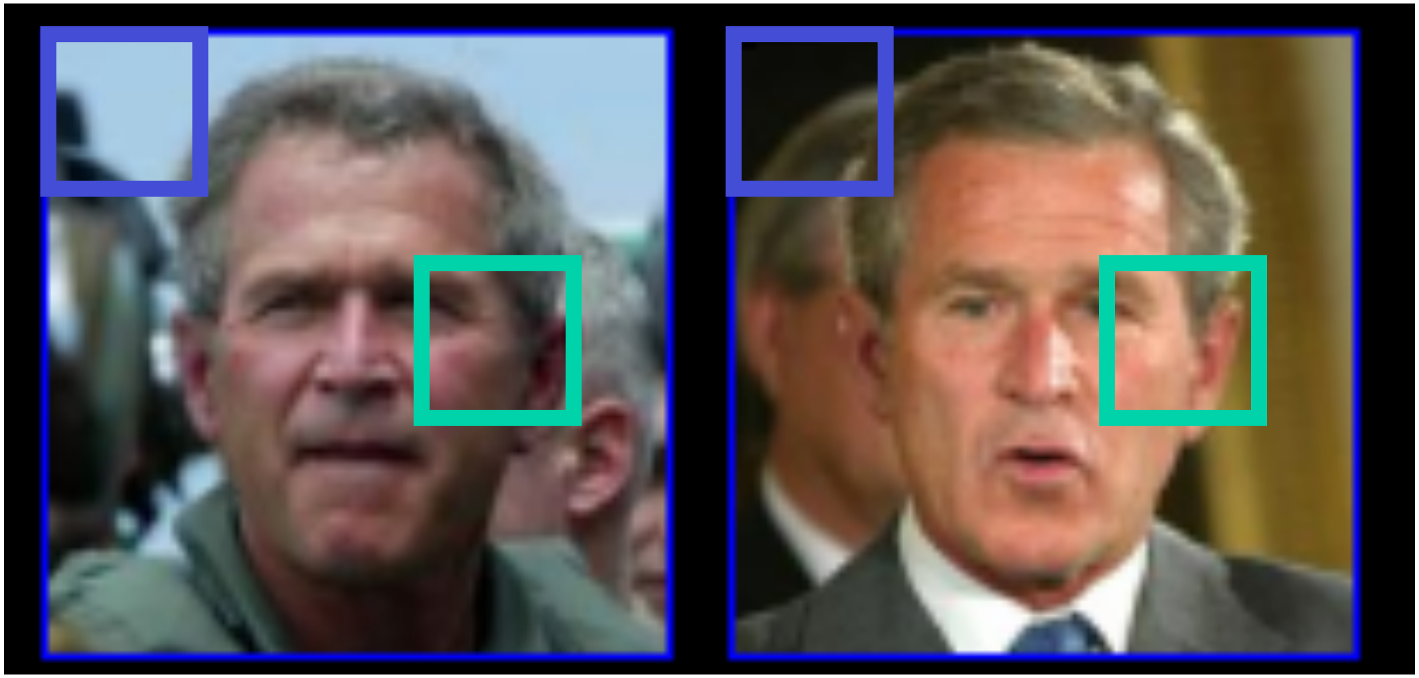
Distributions over distances for specific locations



Most values of d are not informative

Spatial Patch Model

$P(d|\text{same},x_1,y_1)$ estimated separately from $P(d|\text{same},x_2,y_2)$



Avoid drawing wrong conclusion from blue patches.

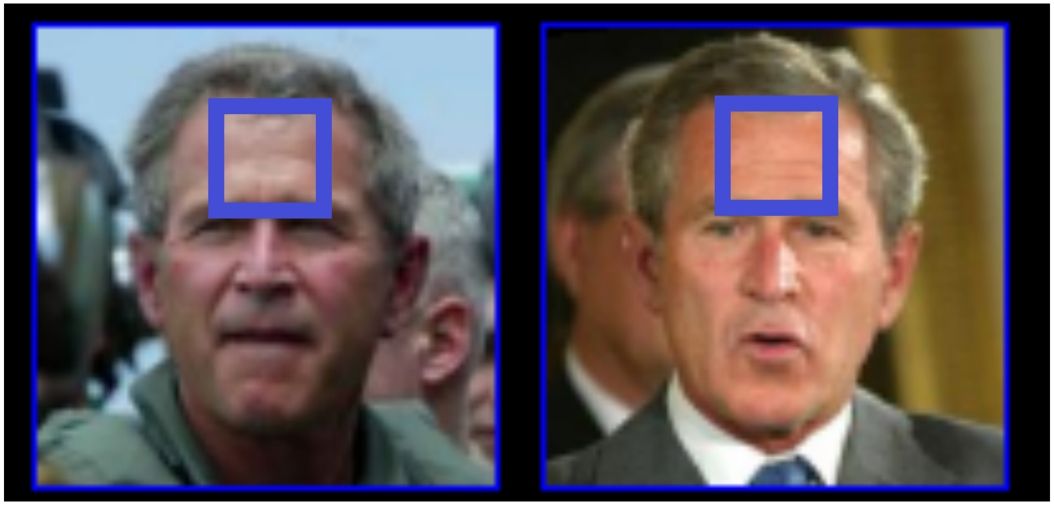
Spatial patch model summary

- By building separate models for each location in the image, we put appropriate emphasis on each region.
- Must estimate separate distribution over d for each position in image.
  - Need more training data.

Hyper-Feature Patch Model

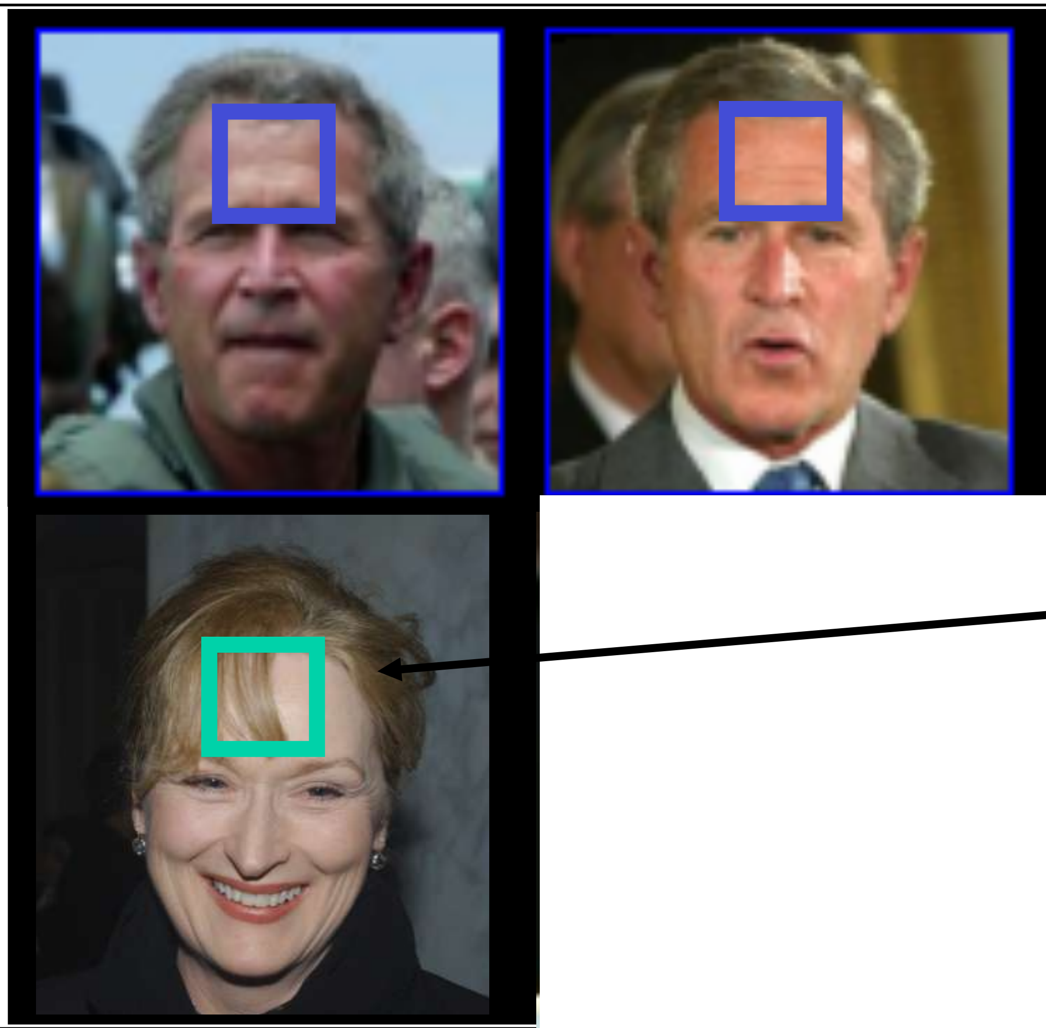
Is the patch from a matching face going to match this patch?

Hyper-Feature Patch Model



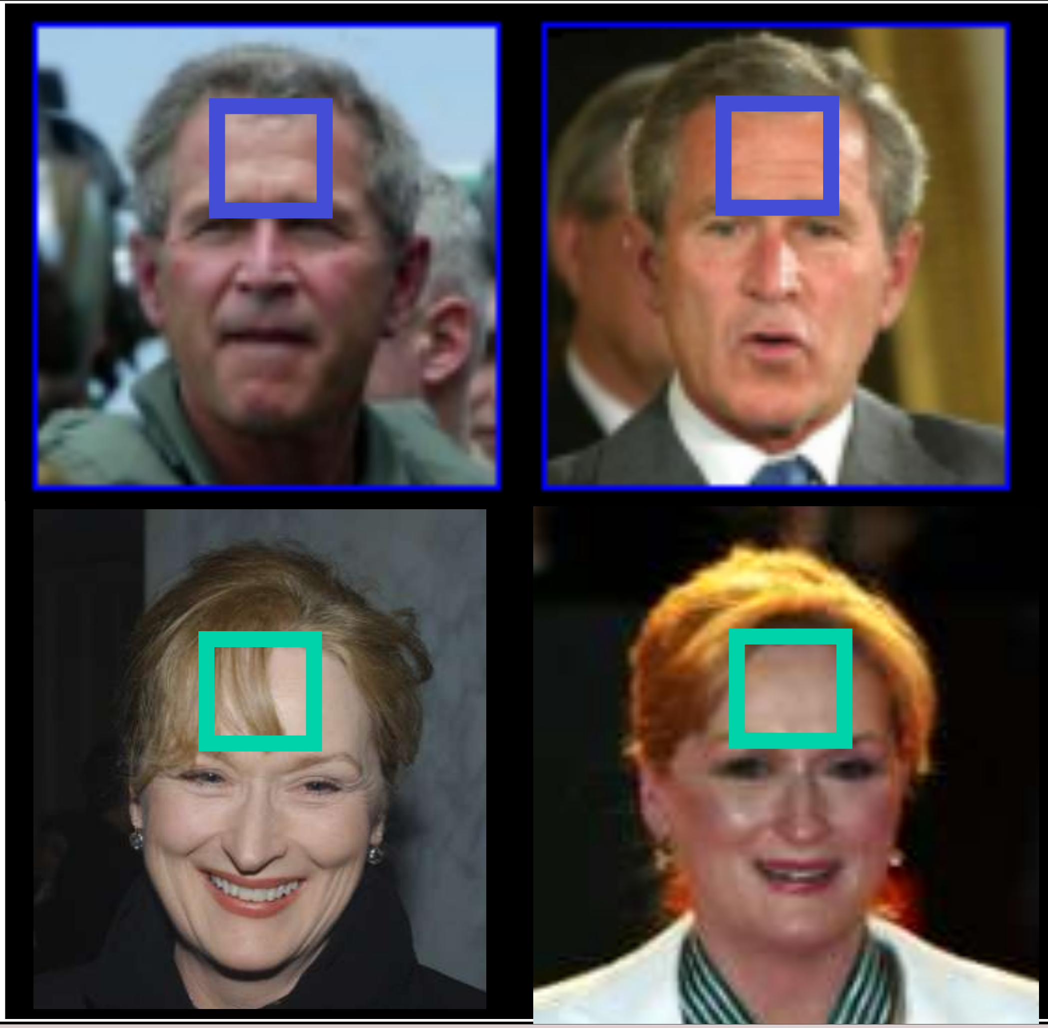
Is the patch from a matching face going to match this patch? Probably yes

Hyper-Feature Patch Model



What about this patch?

Hyper-Feature Patch Model



What about this patch? Probably not.

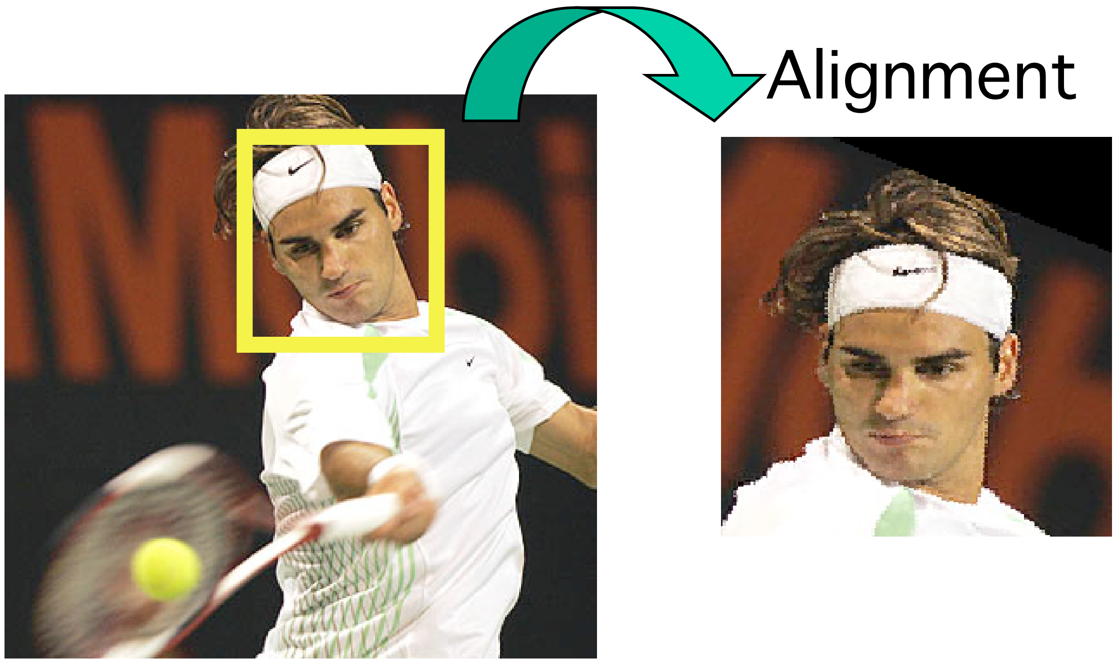
Hyper-features

- Conditioning features for distance distributions
  - P(d|same, x, y, edge energy, contrast, color, etc.)
  - P(d|differ, x, y, edge energy, contrast, color, etc.)
- Greatly improves precision of model.
- Only problem:
  - Must estimate many many distributions. Need a lot of data.
  - Mitigated by using a generalized linear model to share parameters among estimates.

Summary of Recognition work

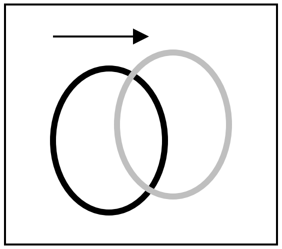
- “Independent" patch model built on probabilities of patch differences.
- Build special conditional distributions of differences based on the appearance of one of the images.
- Around 2005 was state-of-the-art.
  - Has since been surpassed by many others.

Back to Alignment



Strategies of Alignment

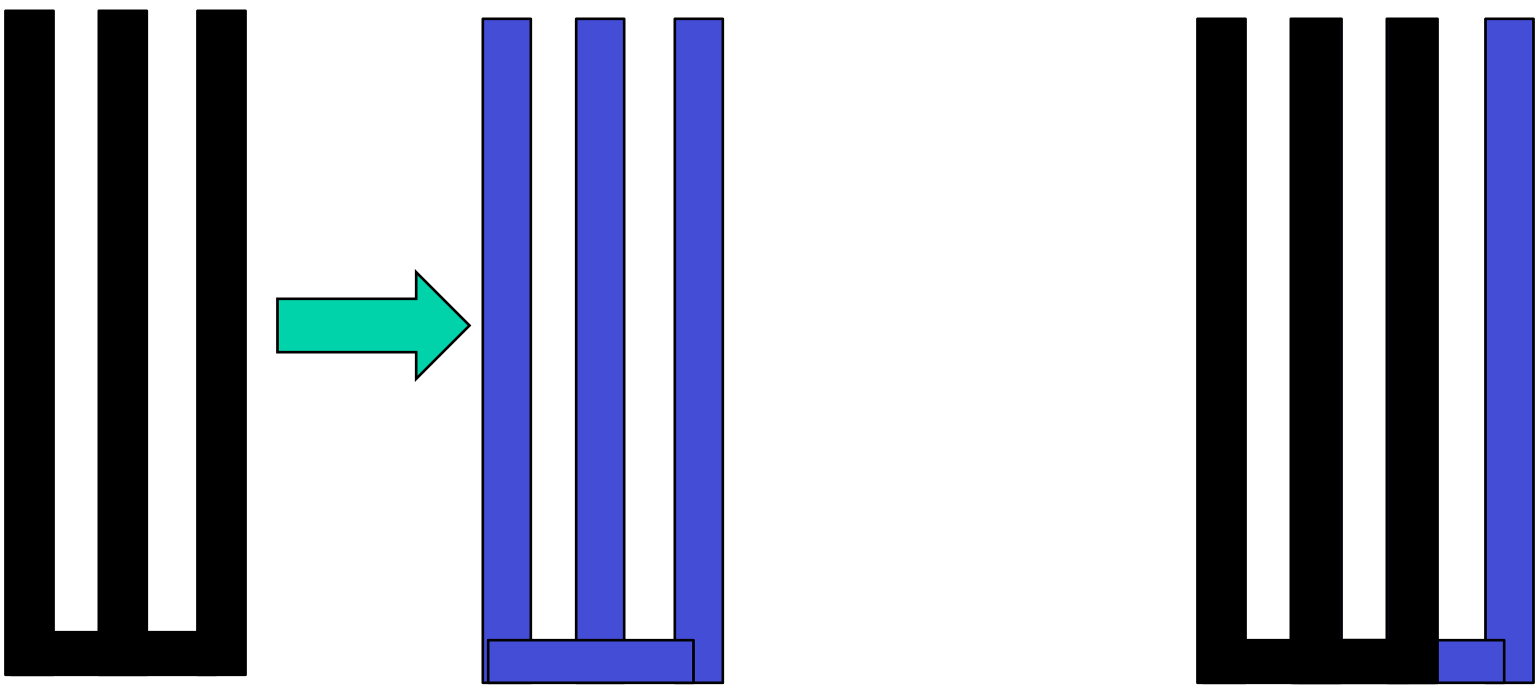
- Find parts: eyes, nose, mouth
  - Put parts in standard positions.
  - Difficulty: a whole new recognition problem!
  - Often certain parts are invisible. Would like to be robust to the missing parts.
- Try to align each new face to some "standard face" via gradient descent.
  - Prone to getting stuck in local optima.



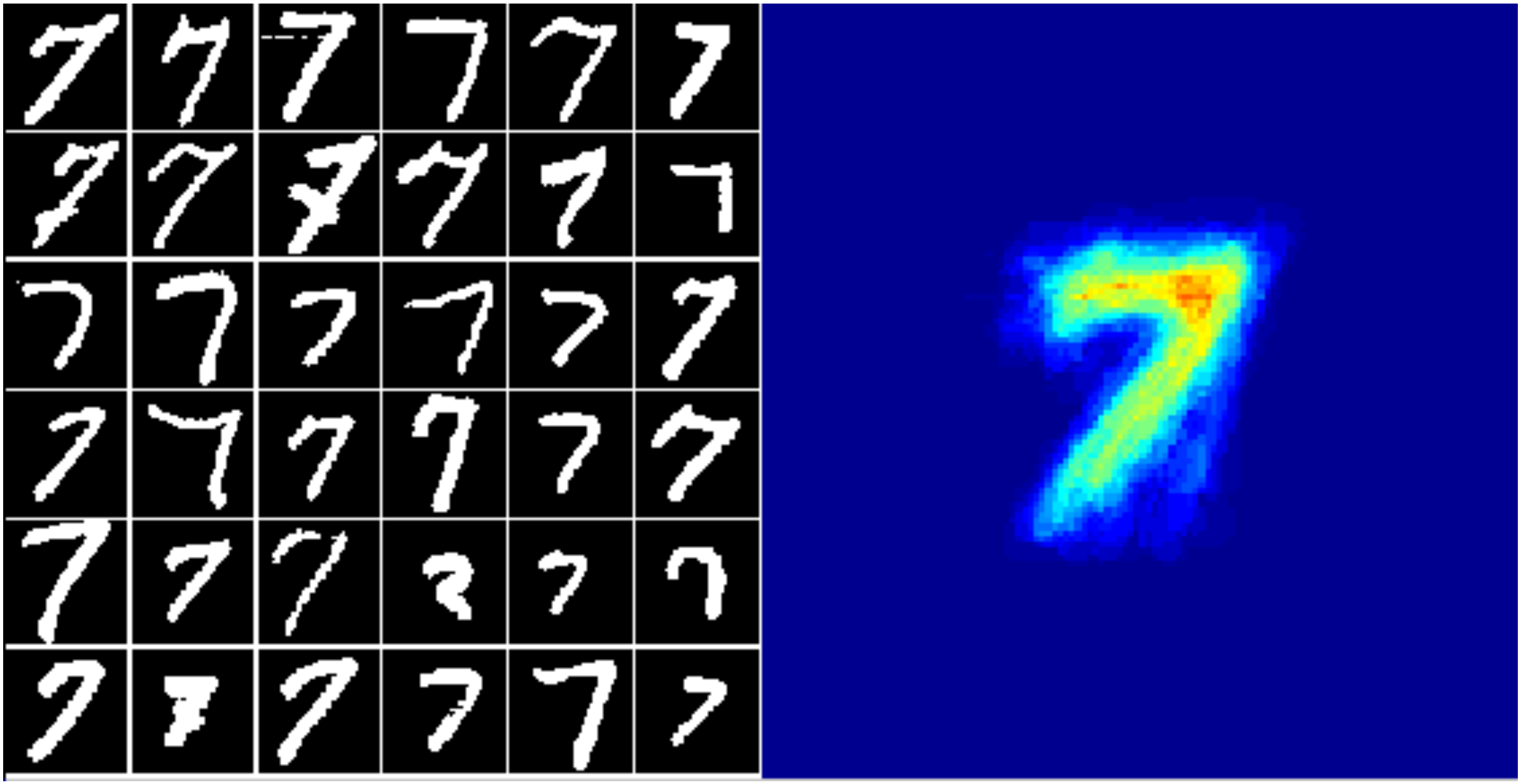
Local optimum problem in alignment

Unaligned

Stuck in a local optimum

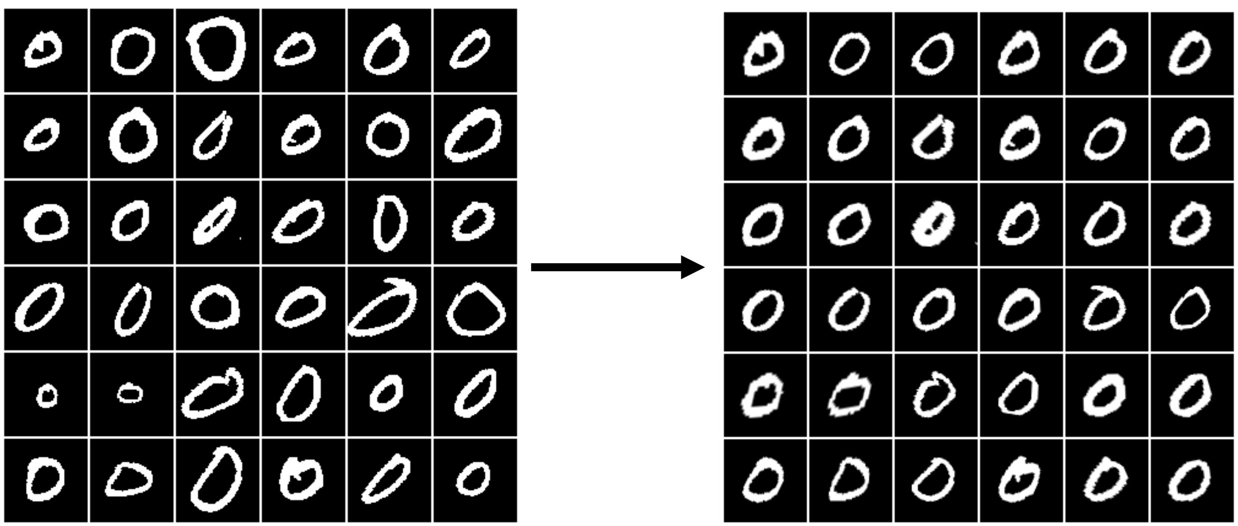


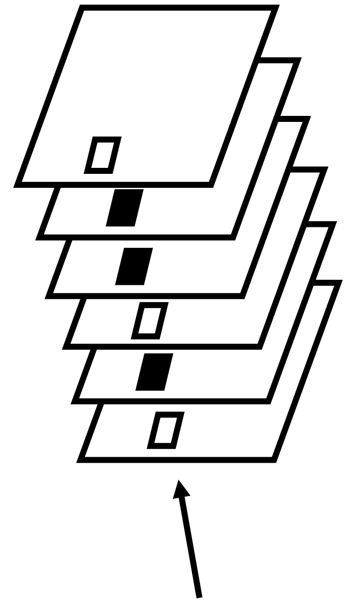
Congealing *(CVPR 2000)*



Criterion of Joint Alignment

- Minimize sum of pixel stack entropies by transforming each image.





A pixel stack

Congealing Interpretations

- Make each image as likely as possible with respect to all other images.
  - Joint maximum likelihood
- Find hidden variables (transformations) that make data as “compact” (low entropy) as possible.

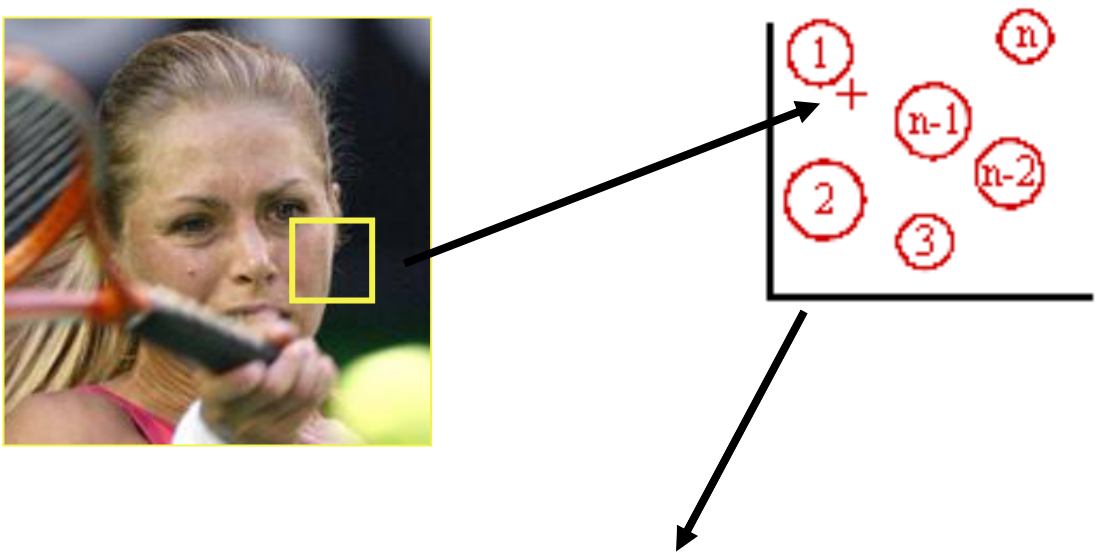
Congealing Complex Images

- Extend congealing to deal with noise in real world images
  - Complex and variable lighting effects
  - Occlusions
  - Highly varied foreground objects (hair, hats, glasses…)
  - Highly varied backgrounds

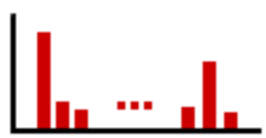
Difficulties with complex images

- Congealing requires estimating a probability distribution at each pixel from N images.
  - Let N=100.
  - If images are binary
    - Each pixel can take on only two values.
    - Estimate of Bernoulli random variable is good.
  - If images are color
    - Each pixel can take on 16 million values.
    - Estimate of 16 million-valued multinomial from 100 samples is horrible.

Congealing Complex Images: Feature Clustering



Feature clusters



vector representing
probability of each cluster,
or “mixture” of clusters

Convert face images to arrays of multinomials

- Start with data set of faces
- Compute SIFT at each pixel
- Cluster SIFT vectors (16 clusters)
- At each pixel, form posterior (multinomial) over clusters
- Distribution of pixel stack is mean of multinomial vectors
- Now, do congealing over these multinomial vectors

Questions about Face Recognition

- How hard can it be?
- What's it good for today?
  - What about in the near future?
- What are the underlying technologies?
  - Hyperfeatures for recognition.
  - Congealing for alignment.
- How do we characterize performance?

A lot of misinformation

- How well does your product work?
  - "It achieves 99% accuracy!"
    - On what problem?
    - How big is the gallery?
    - How difficult are the images?
    - Does this include detection?
- Industry is the major culprit in exaggerated claims.
  - Superbowl example.
  - London surveillance example.
- However...

The Bernie Madoff of Face Recognition

"100% Accuracy in Automatic Face Recognition" [!!!]

Science 25 January 2008

The Bernie Madoff of Face Recognition

"100% Accuracy in Automatic Face Recognition" [!!!]

Science 25 January 2008

If someone's results are too good, you should be highly skeptical.

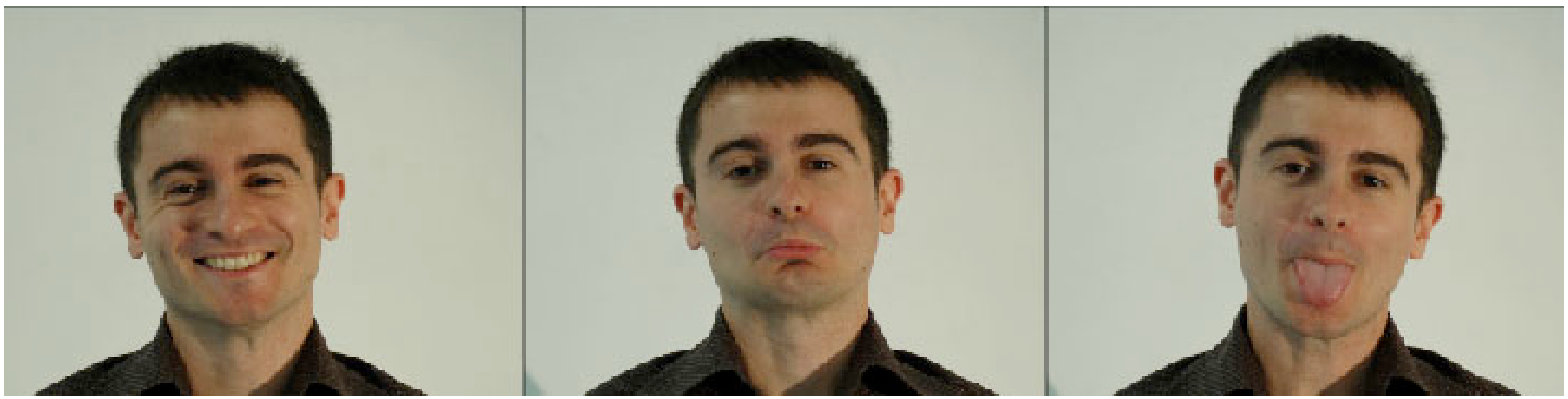
What's needed

- Careful problem definitions.
  - Validation given detection and alignment.
  - Validation given detection, but not alignment.
  - Validation given neither.
  - Intruder detection with alignment and neutral pose.
  - etc.
- Carefully defined test suites.
  - GTAV Database
  - UMass Amherst database

The Many Faces of Face Recognition

GTAV Face Database

The Many Faces of Face Recognition



GTAV Face Database

The Many Faces of Face Recognition

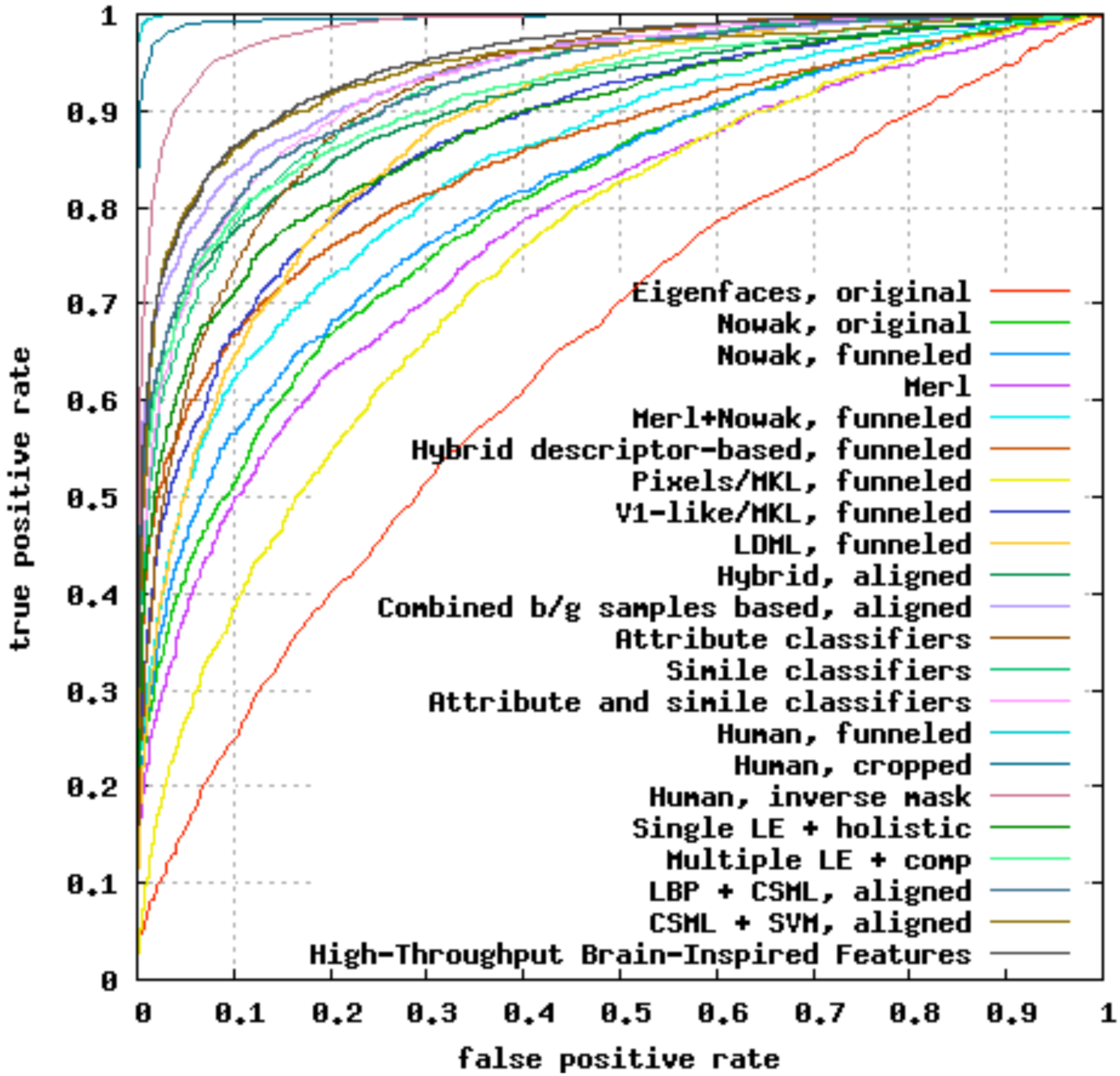
GTAV Face Database

UMass Database

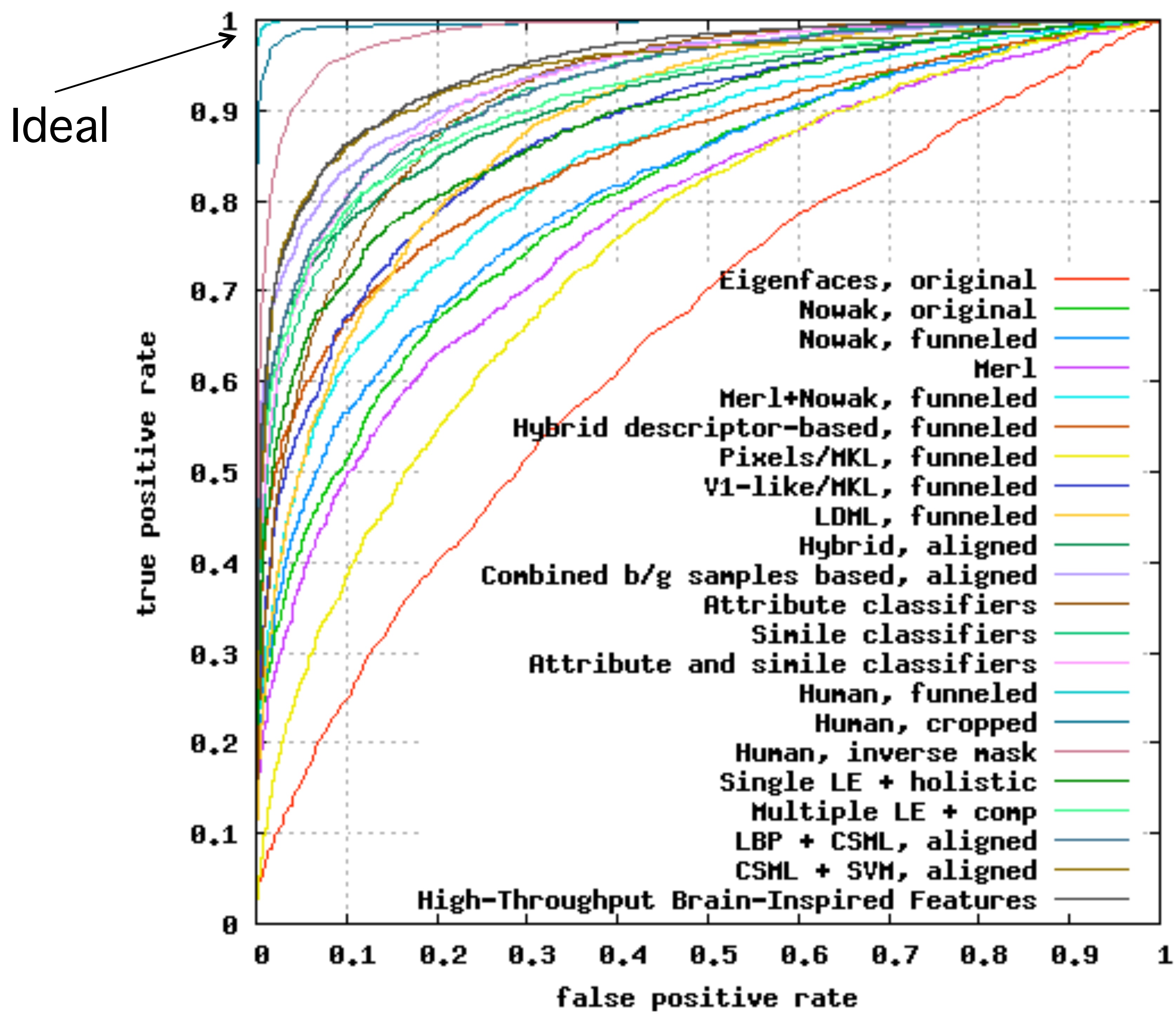
- Real-world images: “wild”
- 13,233 images, with name of each person
- 5749 people
- 1680 people with 2 or more images

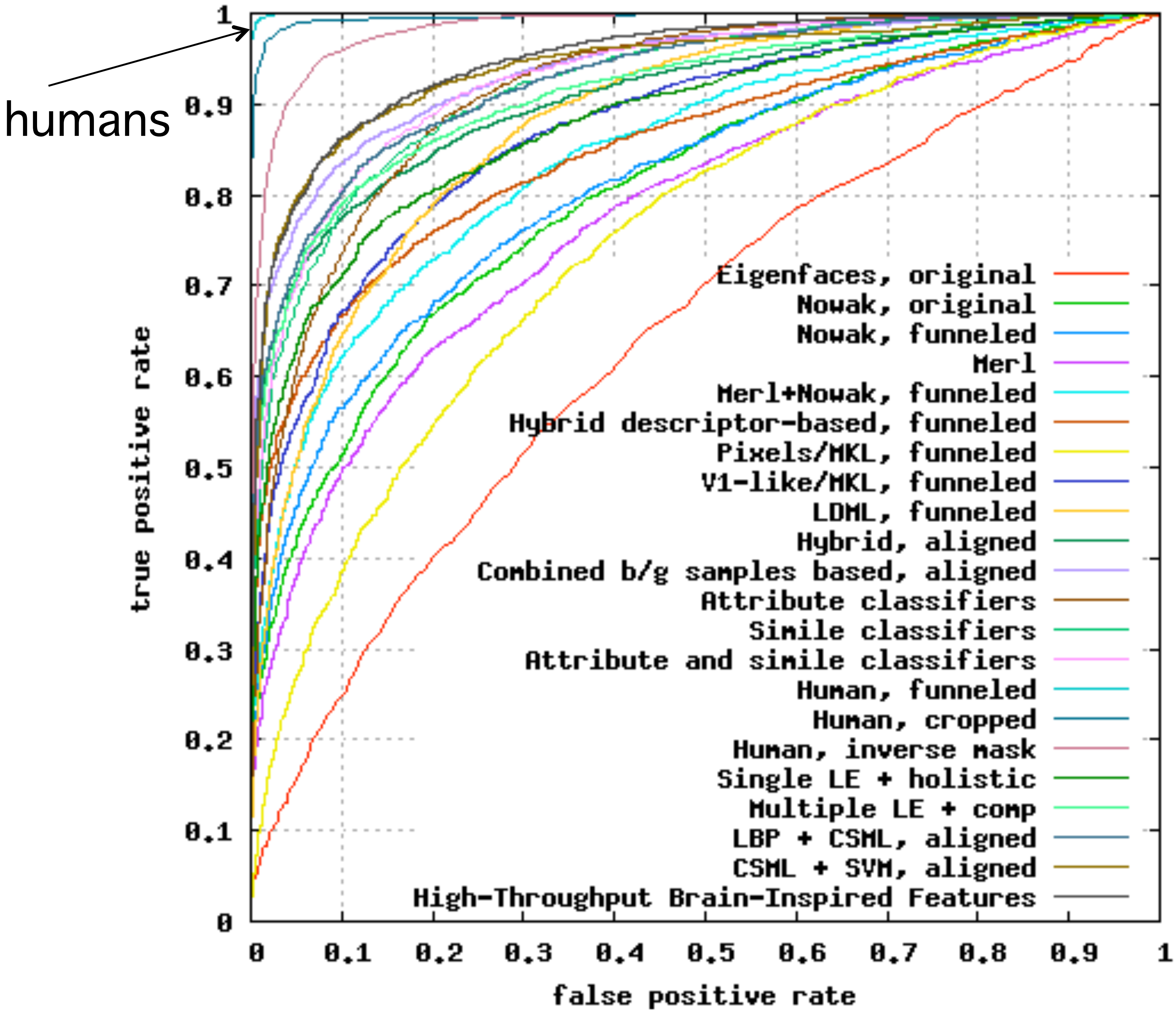
- Designed for the face verification problem.
- Best machine accuracy: currently about 86%!
- Human accuracy: about 99.8%

Sample UMass database images

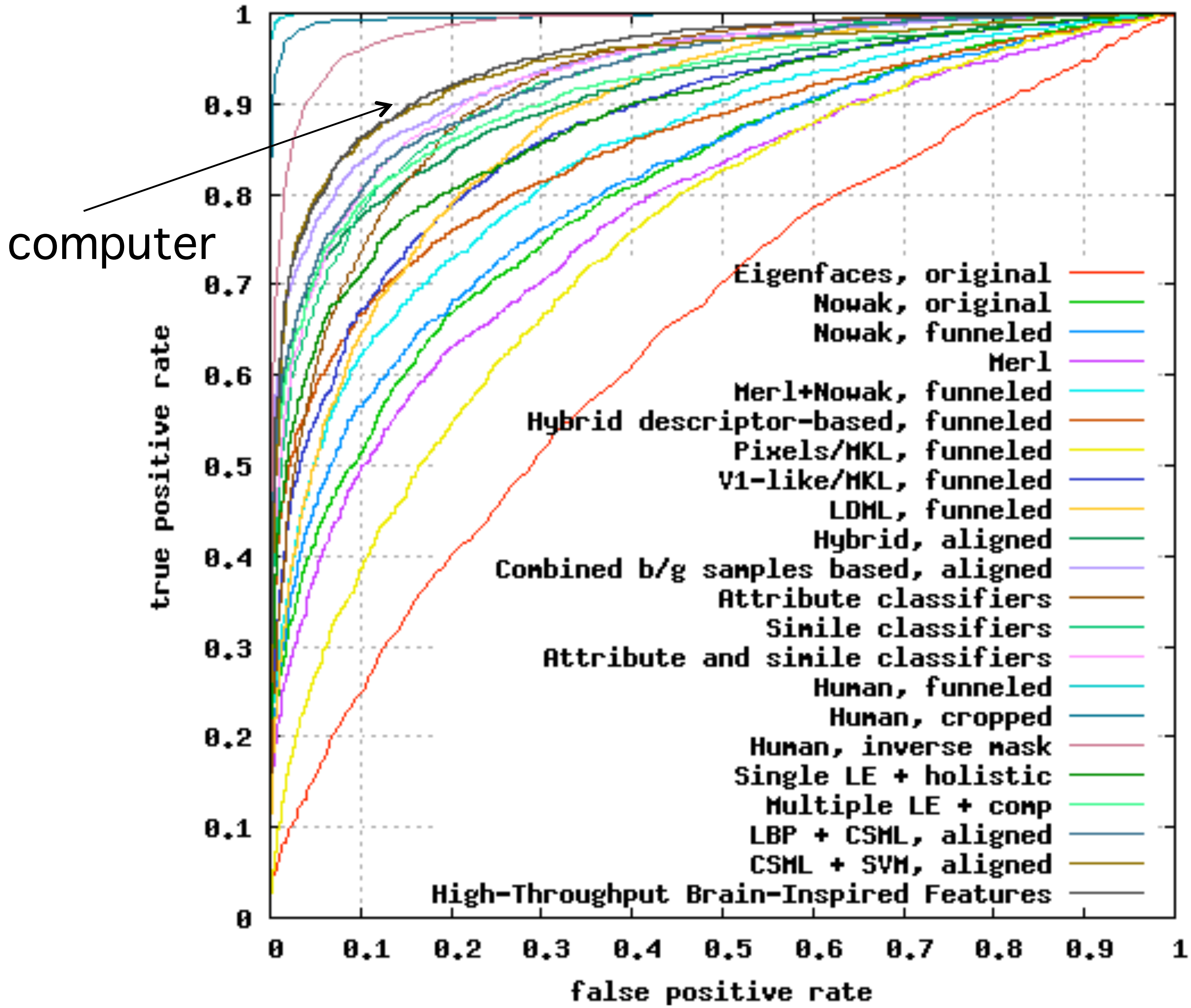


true positive rate
false positive rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Eigenfaces, original
Nowak, original
Nowak, funneled
Merl
Merl+Nowak, funneled
Hybrid descriptor-based, funneled
Pixels/MKL, funneled
V1-like/MKL, funneled
LDML, funneled
Hybrid, aligned
Combined b/g samples based, aligned
Attribute classifiers
Simile classifiers
Attribute and simile classifiers
Human, funneled
Human, cropped
Human, inverse mask
Single LE + holistic
Multiple LE + comp
LBP + CSML, aligned
CSML + SVM, aligned
High-Throughput Brain-Inspired Features


Ideal
true positive rate
false positive rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Eigenfaces, original
Nowak, original
Nowak, funneled
Merl
Merl+Nowak, funneled
Hybrid descriptor-based, funneled
Pixels/MKL, funneled
V1-like/MKL, funneled
LDML, funneled
Hybrid, aligned
Combined b/g samples based, aligned
Attribute classifiers
Simile classifiers
Attribute and simile classifiers
Human, funneled
Human, cropped
Human, inverse mask
Single LE + holistic
Multiple LE + comp
LBP + CSML, aligned
CSML + SVM, aligned
High-Throughput Brain-Inspired Features



humans
true positive rate
false positive rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Eigenfaces, original
Nowak, original
Nowak, funneled
Merl
Merl+Nowak, funneled
Hybrid descriptor-based, funneled
Pixels/MKL, funneled
V1-like/MKL, funneled
LDML, funneled
Hybrid, aligned
Combined b/g samples based, aligned
Attribute classifiers
Simile classifiers
Attribute and simile classifiers
Human, funneled
Human, cropped
Human, inverse mask
Single LE + holistic
Multiple LE + comp
LBP + CSML, aligned
CSML + SVM, aligned
High-Throughput Brain-Inspired Features


computer
true positive rate
false positive rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Eigenfaces, original
Nowak, original
Nowak, funneled
Merl
Merl+Nowak, funneled
Hybrid descriptor-based, funneled
Pixels/MKL, funneled
V1-like/MKL, funneled
LDML, funneled
Hybrid, aligned
Combined b/g samples based, aligned
Attribute classifiers
Simile classifiers
Attribute and simile classifiers
Human, funneled
Human, cropped
Human, inverse mask
Single LE + holistic
Multiple LE + comp
LBP + CSML, aligned
CSML + SVM, aligned
High-Throughput Brain-Inspired Features

Best Match Scores

Worst Match Scores

Best Mismatch Scores

Best Mismatch Scores

Worst Mismatch Scores

Summary of Current Accuracy

- Need careful evaluations on well-defined problems.
- Difficult to assess true state of the art
  - lack of industry participation in benchmarks.
  - But clearly far inferior to humans in most settings.
- Cheating is common on benchmarks.

Questions about Face Recognition

- How hard can it be?
- What's it good for today?
  - What about in the near future?
- What are the underlying technologies?
  - Hyperfeatures for recognition.
  - Congealing for alignment.
- How do we characterize performance?

Thanks!

Ties with financial industry

- Alignment
  - Correlation up to time shift
  - Instruments may appear independent due to a time shift, but are really dependent.
- Conditional models.
  - Searching for right conditioning variables
- Other work
  - Finding independent causes in data

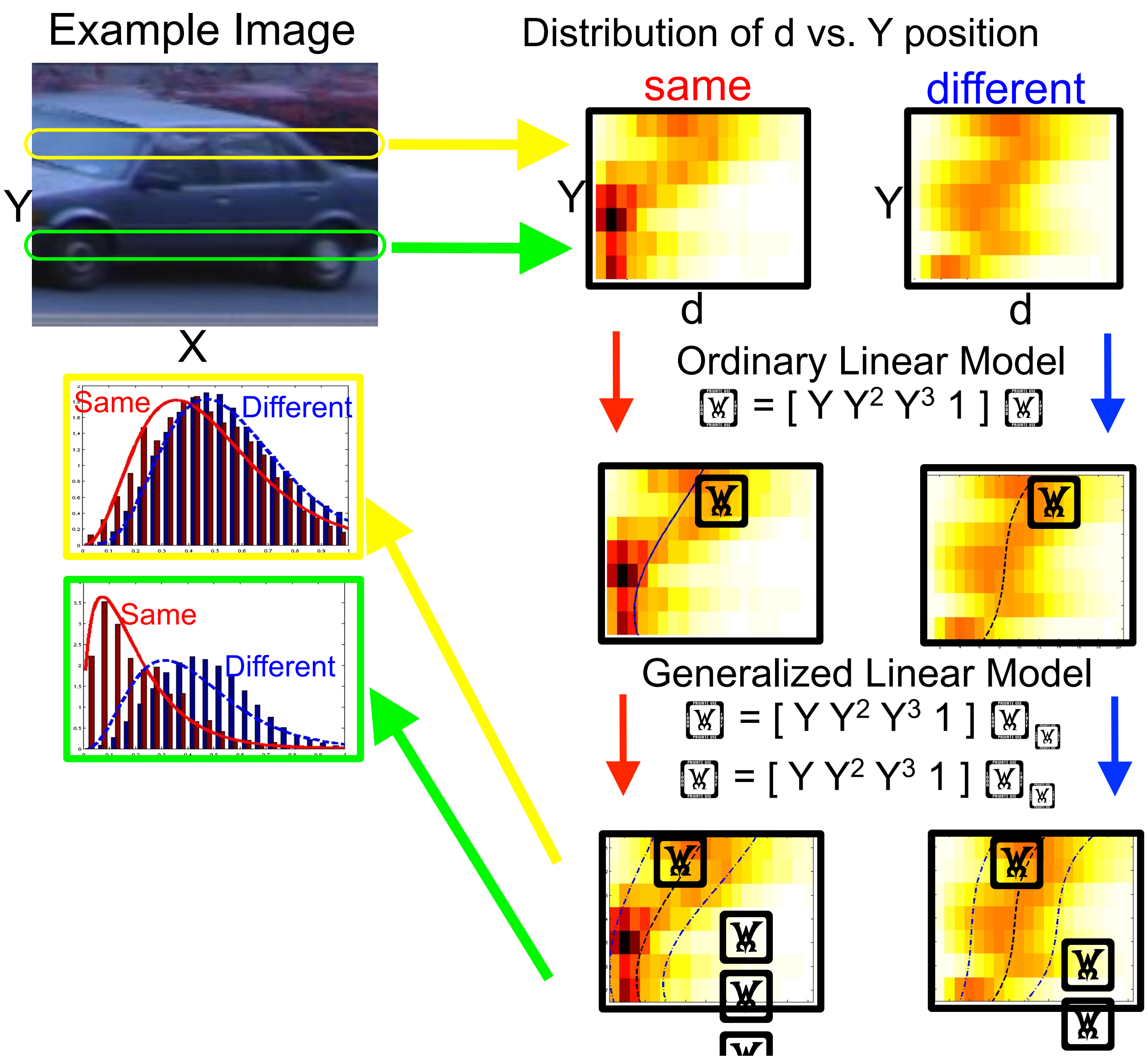
Overall Summary

- Face recognition is complex and difficult.
- Humans are masters at taking advantage of
  - context
  - image structure
  - dynamic alignment

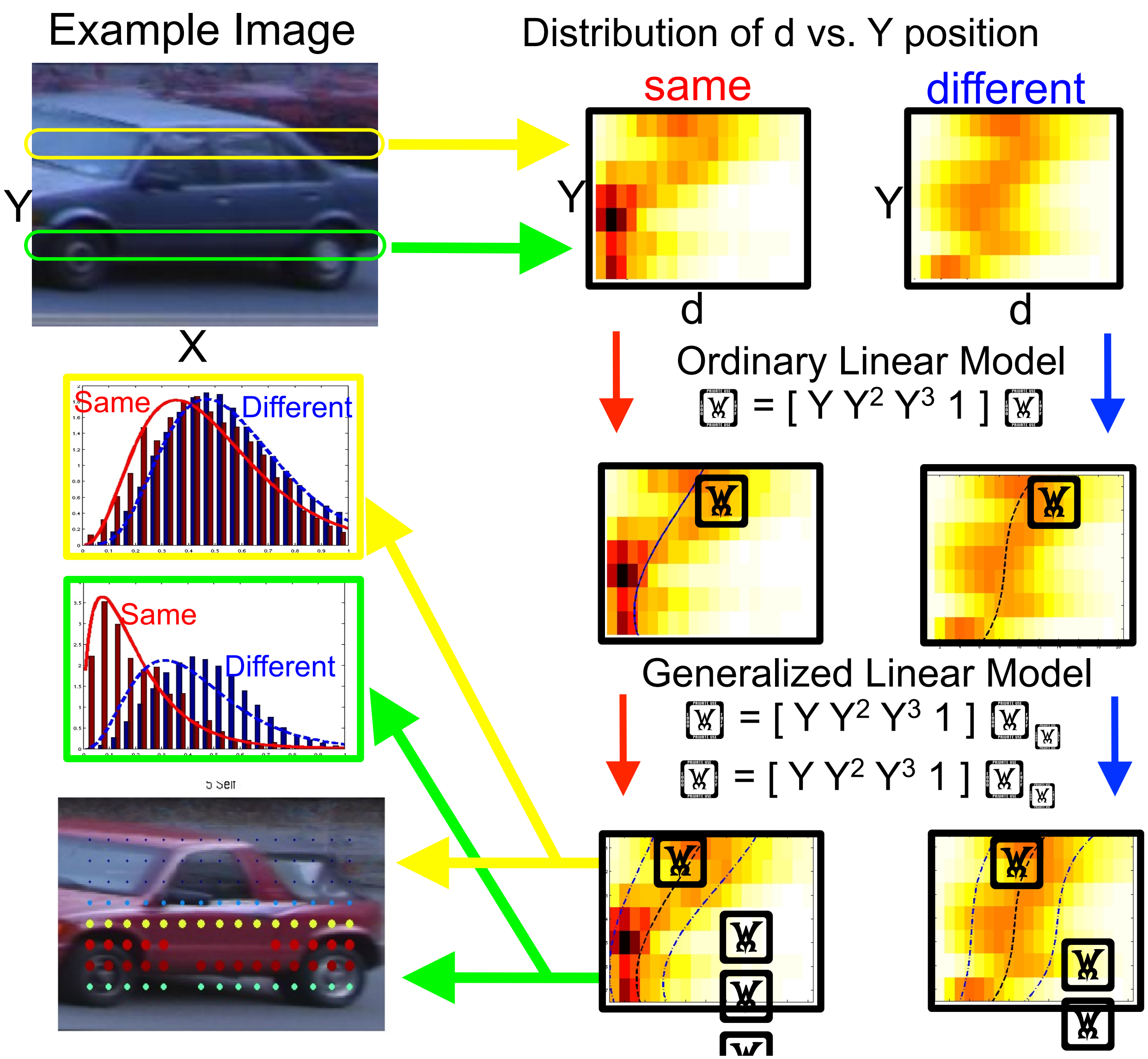
Do we need separate methods for each step?

- Theoretically:
  - No. Evaluate recognition algorithm for every pose and position. Pick best match.
- Practically:
  - Maybe. Performance may be dramatically improved by specializing for each task.

Generalized Linear Model



Generalized Linear Model

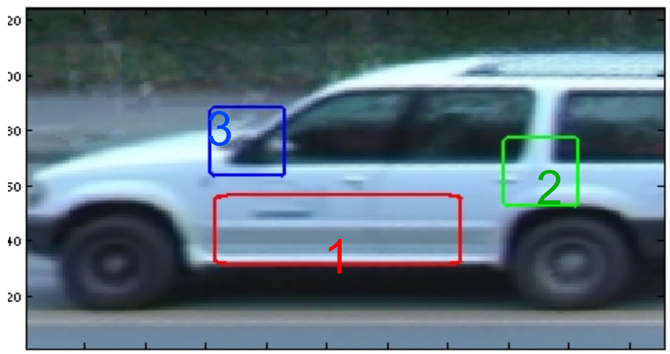


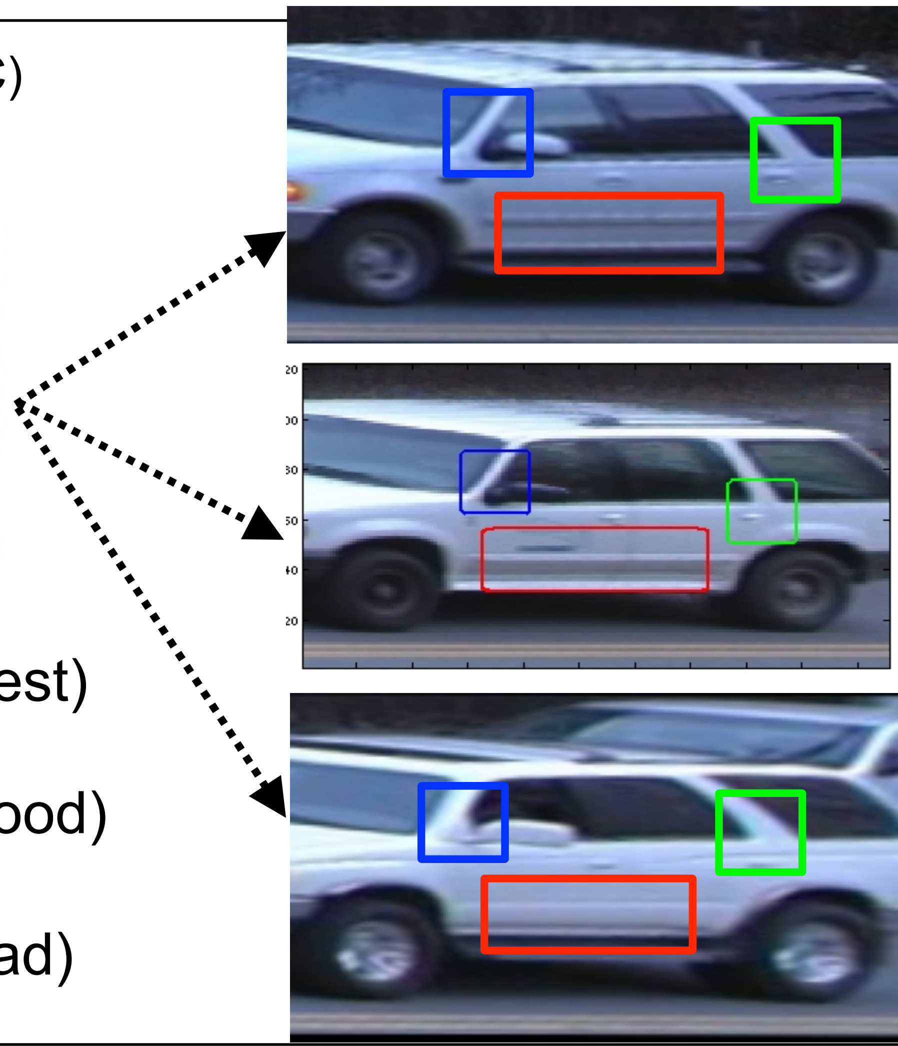
Cascade for Efficient Matching

- Using hyper-feature model, can predict likely utility of a patch
  - Note that even after we observe the “left” image, D is still a random variable, as it also depends on the “right” image.
  - Mutual information: I(D; C)
- For test image, sort patches according to utility.
- Compare against other images only until decision is reached.

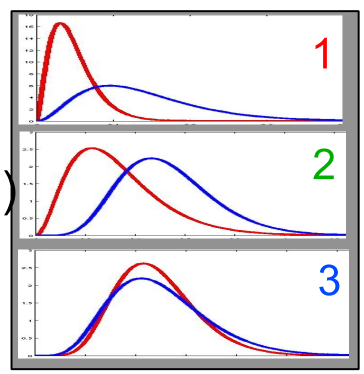
Estimating Saliency

Saliency = Mutual Information $I(D_j;C)$
where C={same,diff}





P(D)



D

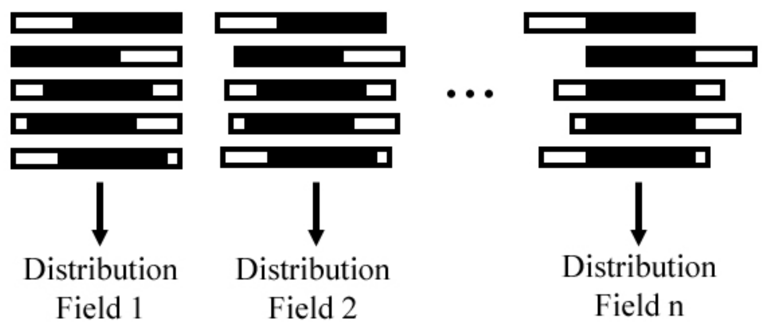
$I(D_1;C)$ = .39 (best)

$I(D_2;C)$ = .23 (good)

$I(D_3;C)$ = .01 (bad)

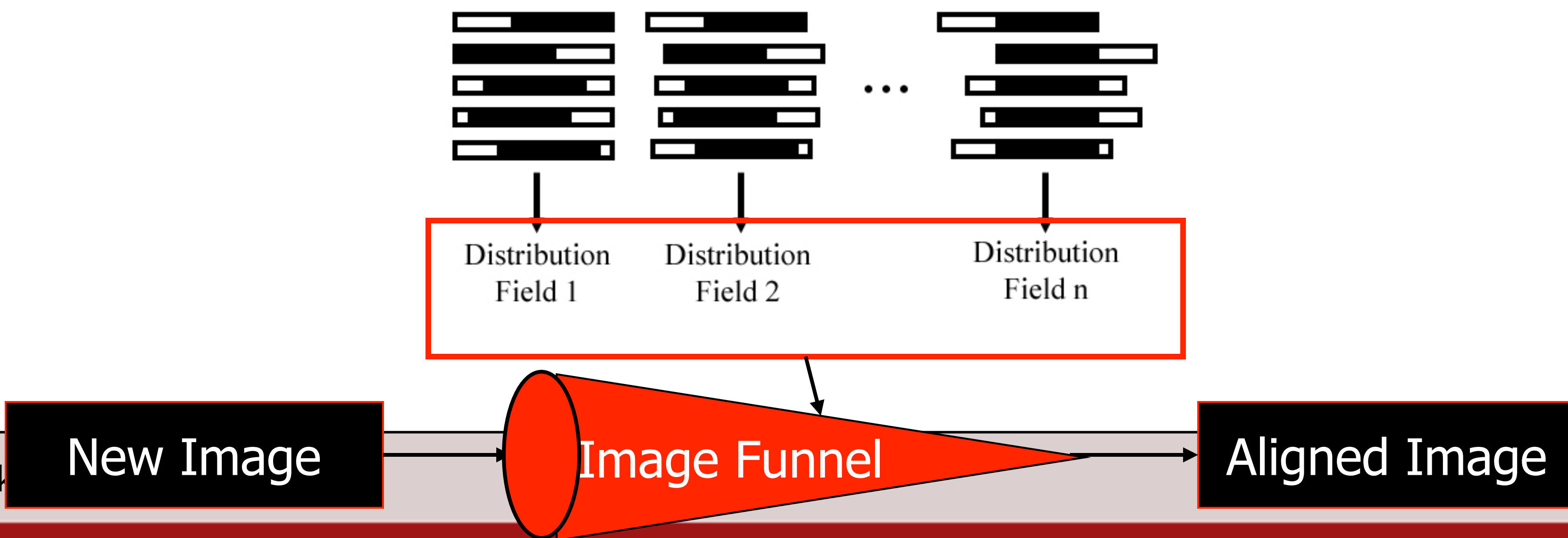
Congealing


update distribution field from transformed images
increase likelihood of image with respect to distribution field


Distribution Field 1
Distribution Field 2
...
Distribution Field n

How to align a new image after congealing?

- Insert into training set, re-run algorithm
- More efficient to save sequence of distribution fields from congealing
  - High entropy to low entropy sequence → “Image Funnel”
- Funneling: increase likelihood of new image at each iteration according to corresponding distribution field



Comparing Patches

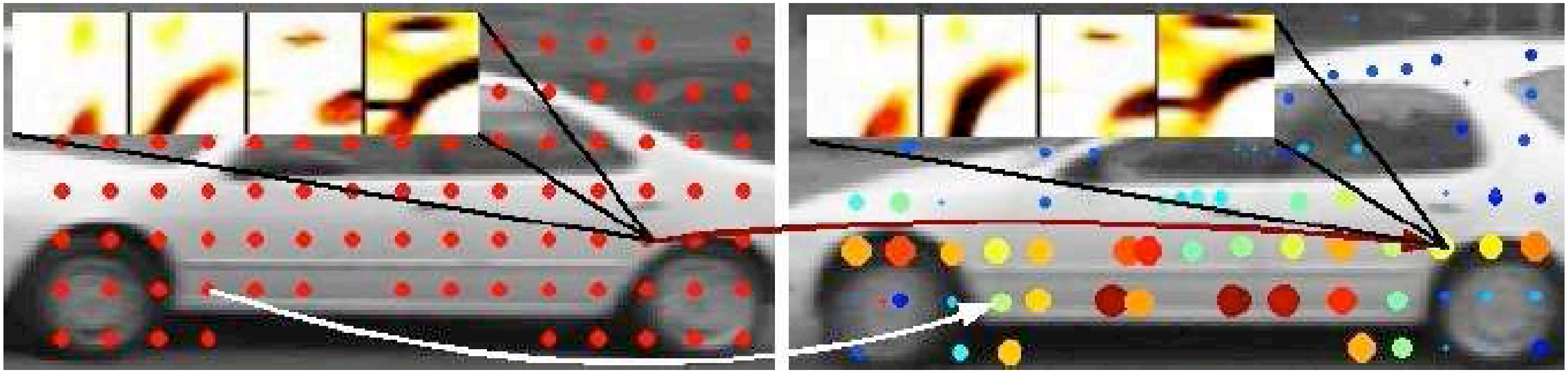


Figure 3: *Patch Matching:* The left image is sampled (red dots) by patches encoded as oriented filter channels (for one patch, four channels are shown). Each patch is matched to the best point in the other image by maximizing the appearance similarity between the patches. For each pair of patches (bi-patch), the size and color of the matching circles indicates similarity in appearance. (Larger and redder is more similar.)

$$d_j = 1 - CorrCoef(F_j^L, F_j^R)$$

Congealing of curves