FaceTracer:

A Search Engine for Large Collections of Images with Faces

Authors:
Neeraj Kumar,
Peter Belhumeur, Shree Nayar
Columbia University

Presented by: Girish Malkarnenkar
19th October 2012
CS395T Visual Recognition

Motivation

- Current search engines use text annotations to find images based on facial appearance.

- **Problems** with this approach:
  1. Manual labeling is time consuming
  2. Textual annotations can be misleading/incorrect
  3. Annotated images are only a small subset of all the images

Google Images then…

Their method then…

Their method now…

Google Images now…

All results
By subject
Personal

Any size
Large
Medium
Icon
Larger than…
Exactly…

Any color
Full color
Black and white

Any type
Face
Photo
Clip art
Line drawing

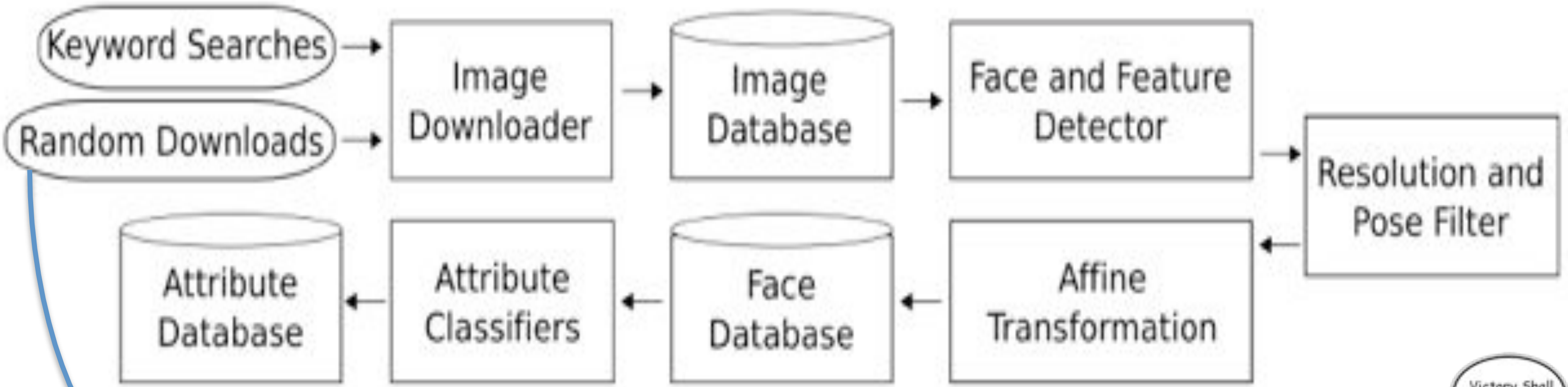
Standard view
Show sizes

PROBLEM STATEMENT:

1. Goal: A search engine based on **both** Facial and Image appearance
2. Since there are billions of images and hundreds of possible attributes, and we can only hope to get a few thousands of manual labels, the labeling of images needs to be done **automatically in a scalable manner**

Database Creation: Downloading images

Celebrity names, Professions, Events etc



Randomly downloaded to permit sampling from a general distribution…

Image from: ECCV 2008 paper, Logos from: link1, link2, link3

Database Creation: Face detection

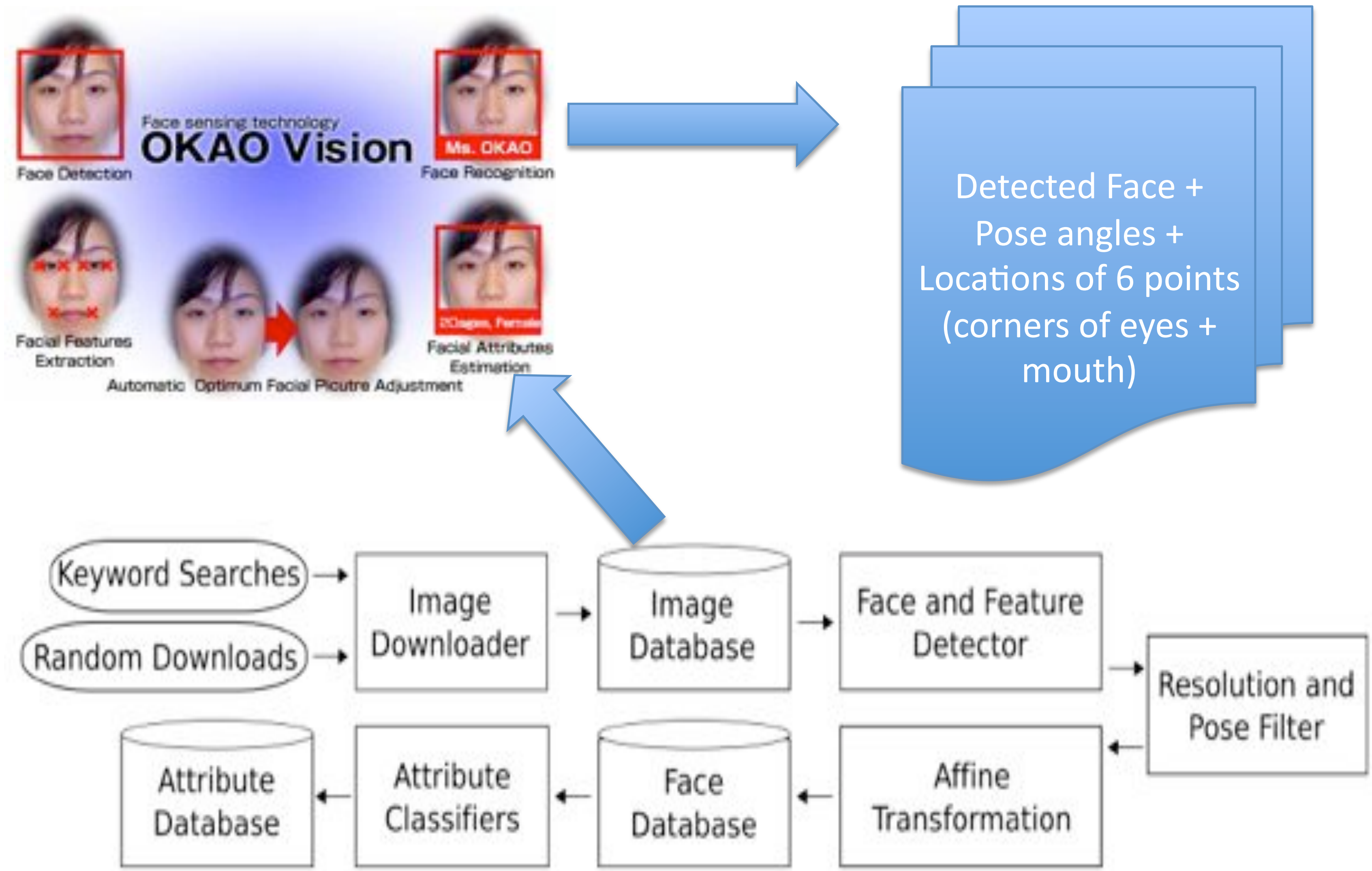
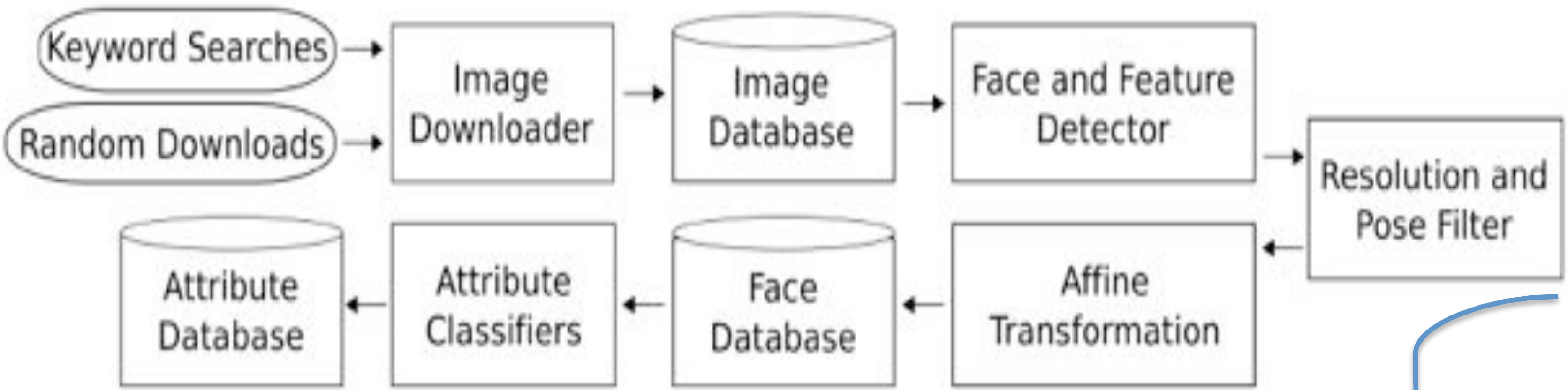


Image from: ECCV 2008 paper, OKAO logo: [link]

Database Creation: Filter/Transformation



Affine transformation to canonical frontal pose using least squares on the 6 points w.r.t a template

Filter detected faces by pose (+/- 10 degrees from front/ center)

Image from: ECCV 2008 paper, [Anakin](), [Luke](), [Affine](),
[Tamara Berg's paper]()

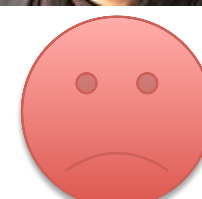


Image Database Statistics

Image Source	# Images	# Faces
Randomly Downloaded	4,248,194	2,124,472
Celebrities	105,568	109,748
Person Names	19,492	12,806
Face-Related Words	13,212	14,424
Event-Related Words	1,429	1,335
Professions	115,808	79,992
Series	7,551	8,585
Camera Defaults	2,153	879
Miscellanous	10,855	16,201
Total	4,539,886	2,373,533

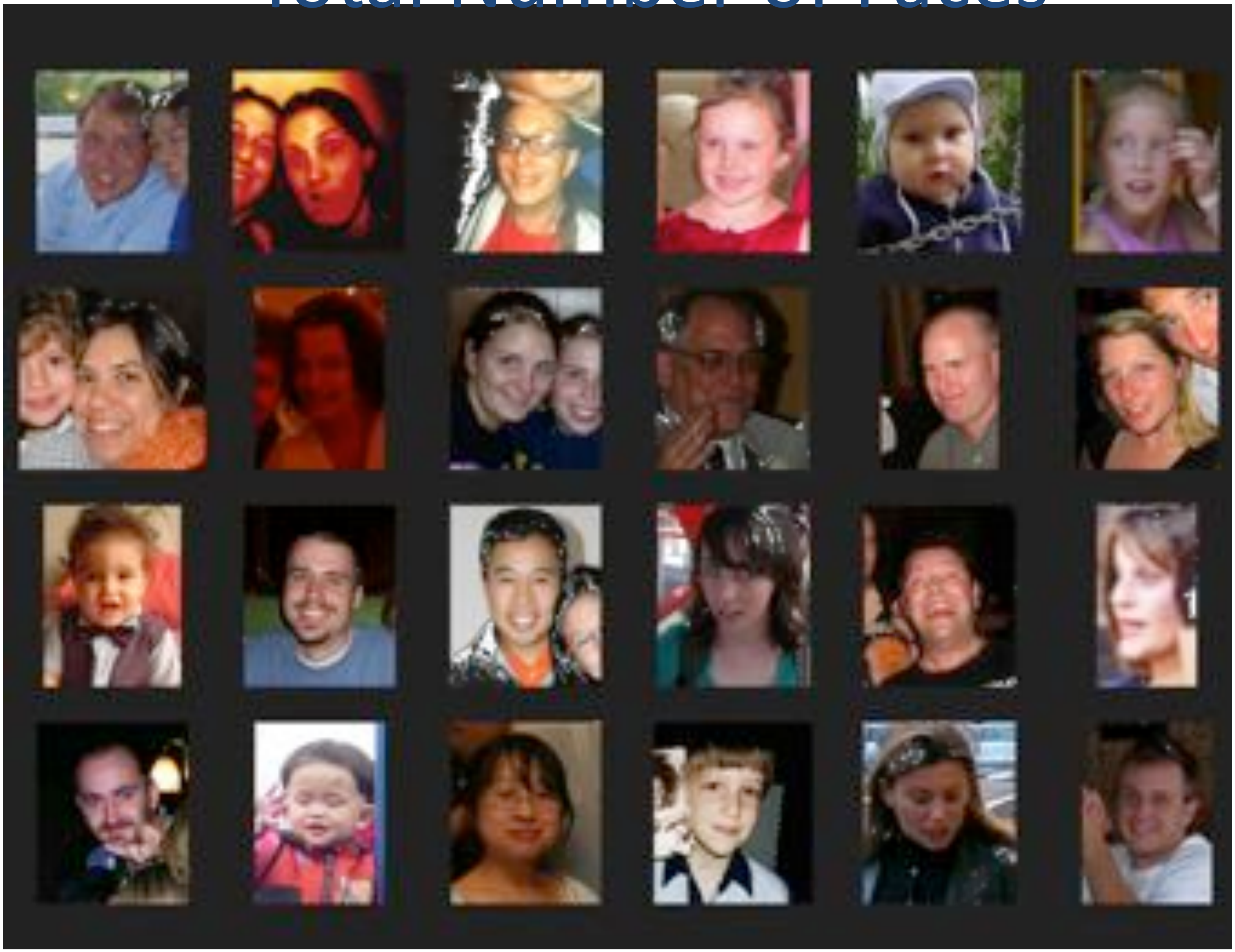
Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Database Size Comparison

Database	# Face Images
MIT+CMU	130
Yale A	165
Yale B	5,760
FERET	14,051
CMU PIE	41,368
FRGC v2.0	50,000
Proposed	2,373,533

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Total Number of Faces



Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

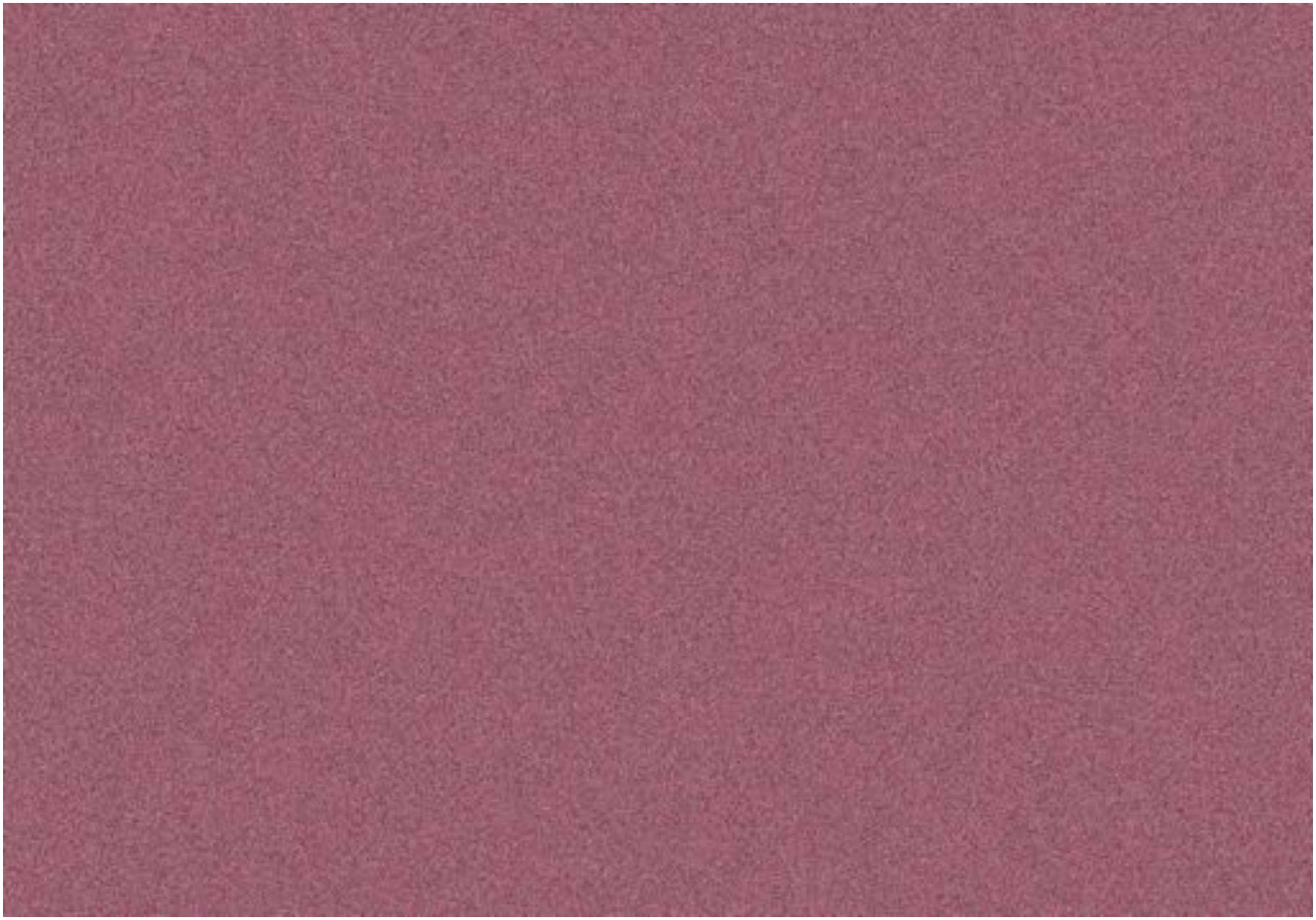
Total Number of Faces

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Total Number of Faces

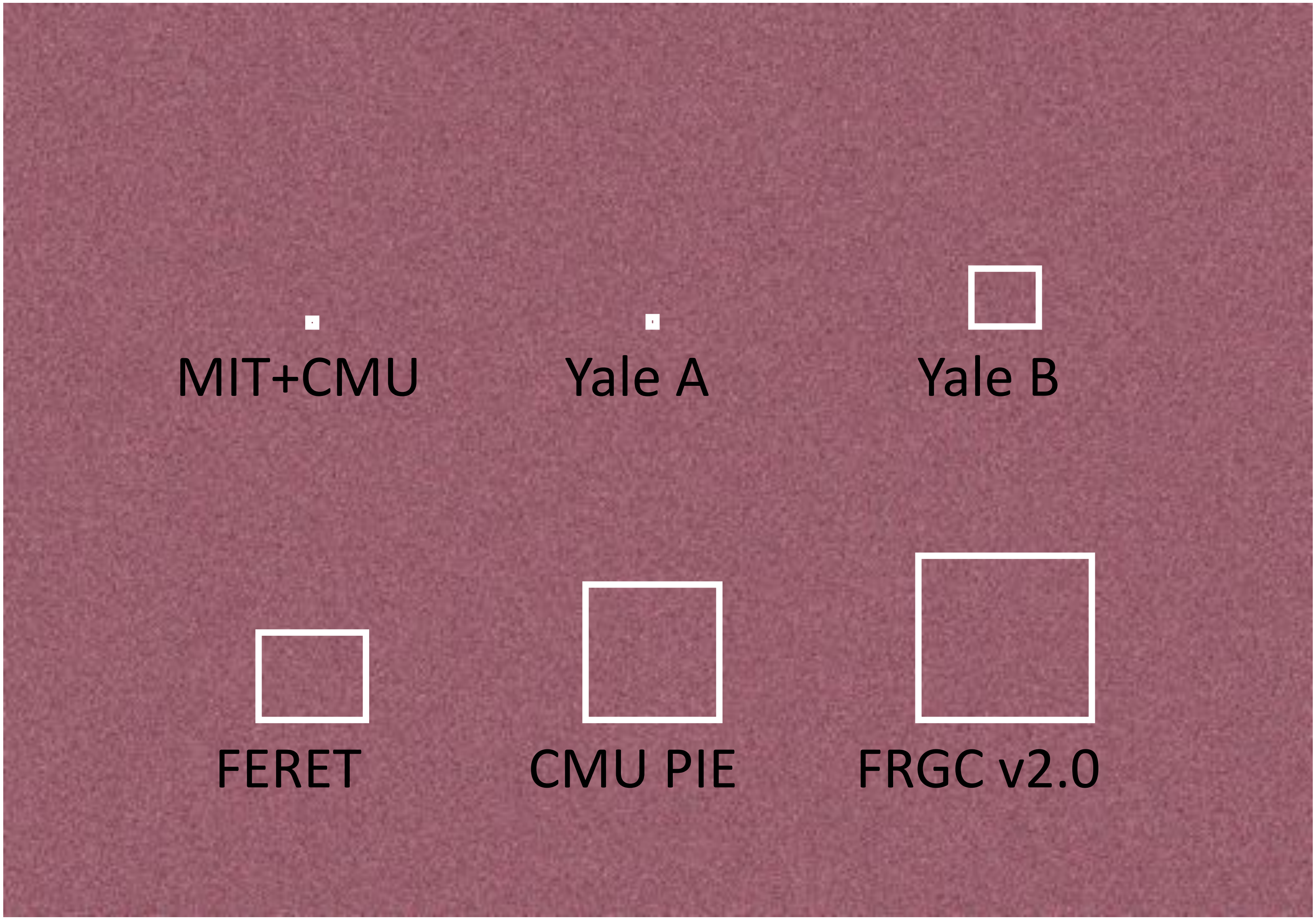
Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Total Number of Faces



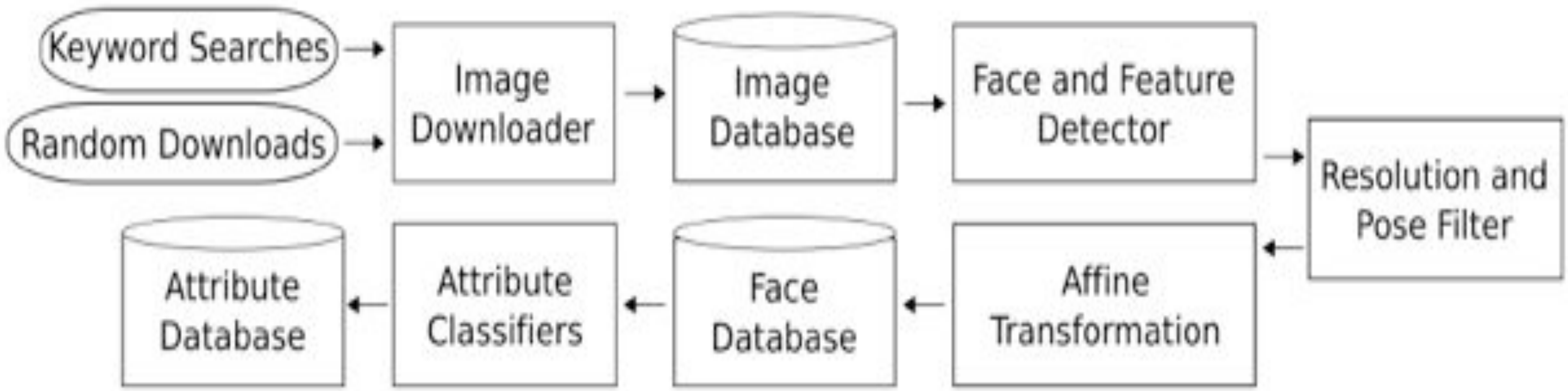
Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Total Number of Faces



Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Manual labeling of attributes…



So at this stage, we **have ~3.1 million images** (at the time of publication in 2008) and we need to train attribute classifier on them for 10 attributes

It is infeasible to manually label all the 3.1M images

BUT

we do need some labeled images for automatically labeling the remaining, so **we manually create ~17,000 attribute labeled images**

Image from: ECCV 2008 paper

Labeled Attribute Statistics

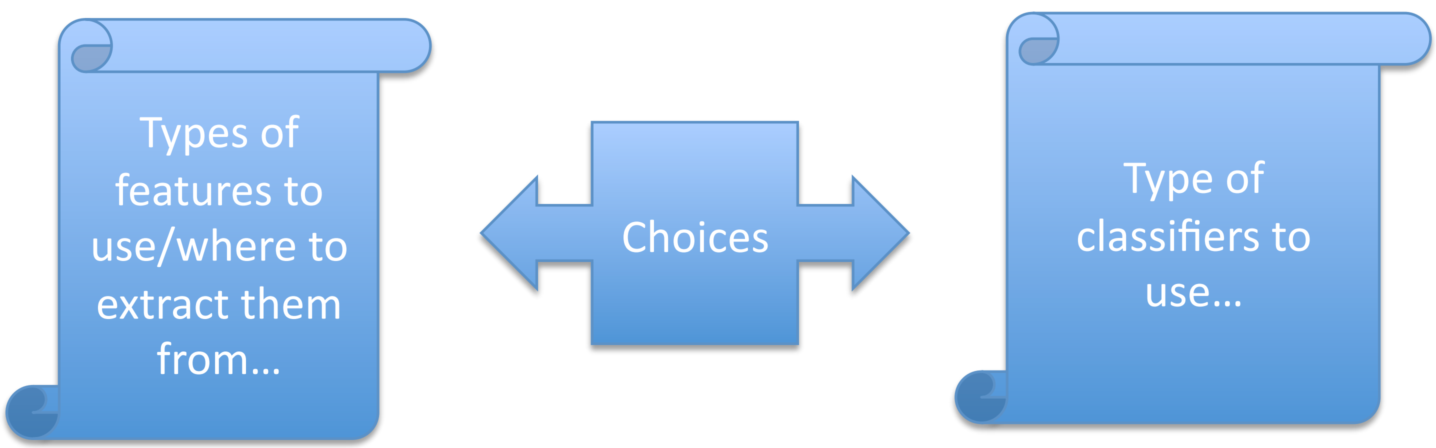
Attribute	# Labeled	Attribute	# Labeled
Gender	1,954	Mustache	1,947
Age	3,301	Smiling	1,571
Race	1,309	Blurry	1,763
Hair Color	1,033	Lighting	633
Eye Wear	2,360	Environment	1,583

Total Number of Labels: 17,454

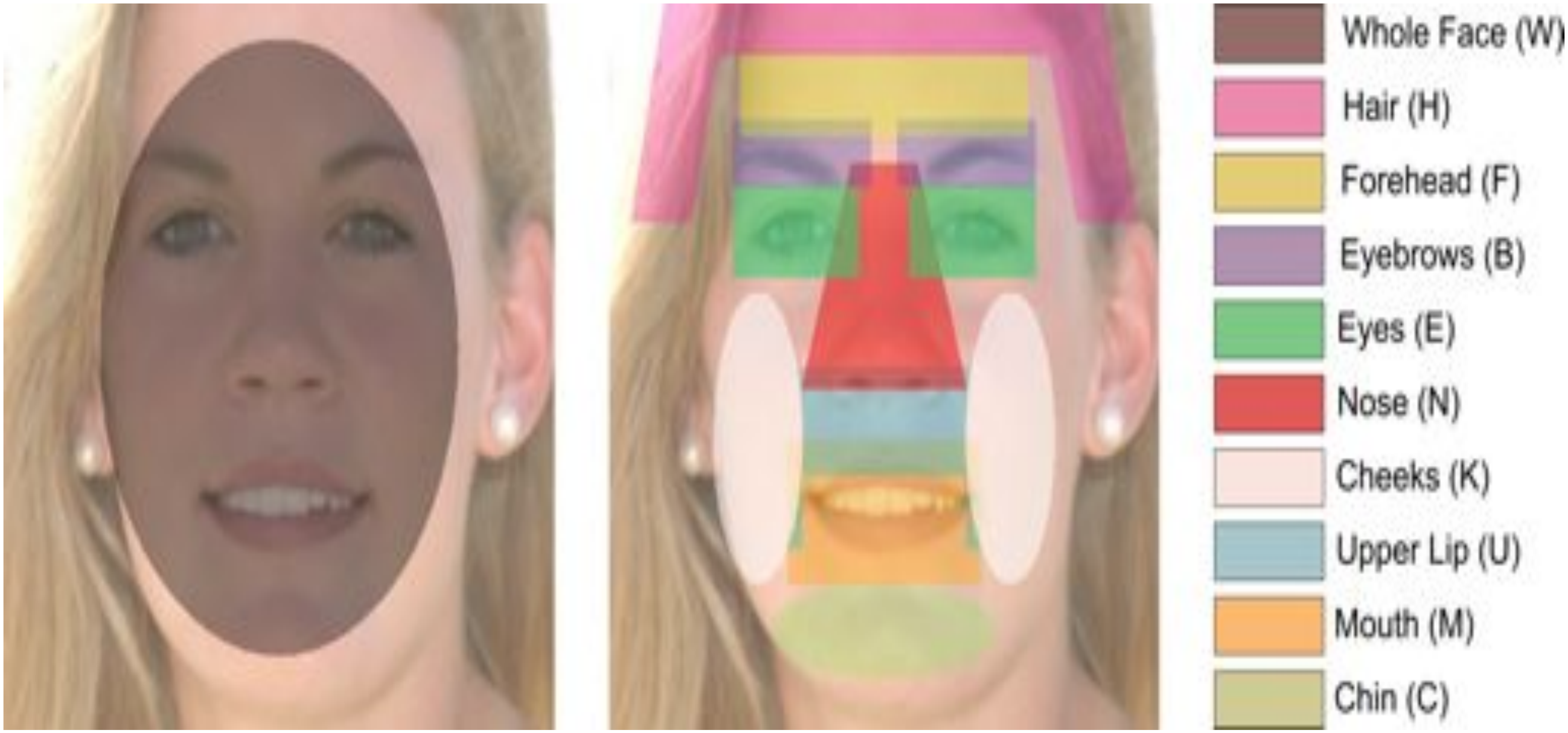
Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

And this is where the fun starts…

- Goal: Given the 17k attribute labels we now need to train attribute classifiers for all 10 attributes to automatically label the remaining images…



Where to extract features from?



Face divided into 10 functional regions…

Image from: ECCV 2008 paper

Feature Types

Pixel Value Type	Normalizations	Aggregation
RGB (r)	None (n)	None (n)
HSV (h)	Mean-Norm (m)	Histogram (h)
Image Intensity (i)	Energy-Norm (e)	Statistics (s)
Edge Magnitude (m)		
Edge Orientation (o)		

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Feature Types

Pixel Value Type	Normalizations	Aggregation
RGB (r)	None (n)	None (n)
HSV (h)	Mean-Norm (m)	Histogram (h)
Image Intensity (i)	Energy-Norm (e)	Statistics (s)
Edge Magnitude (m)		
Edge Orientation (o)		

RGB, Mean Norm., No Aggreg. (r.m.n)

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Feature Types

Pixel Value Type	Normalizations	Aggregation
RGB (r)	None (n)	None (n)
HSV (h)	Mean-Norm (m)	Histogram (h)
Image Intensity (i)	Energy-Norm (e)	Statistics (s)
Edge Magnitude (m)		
Edge Orientation (o)		

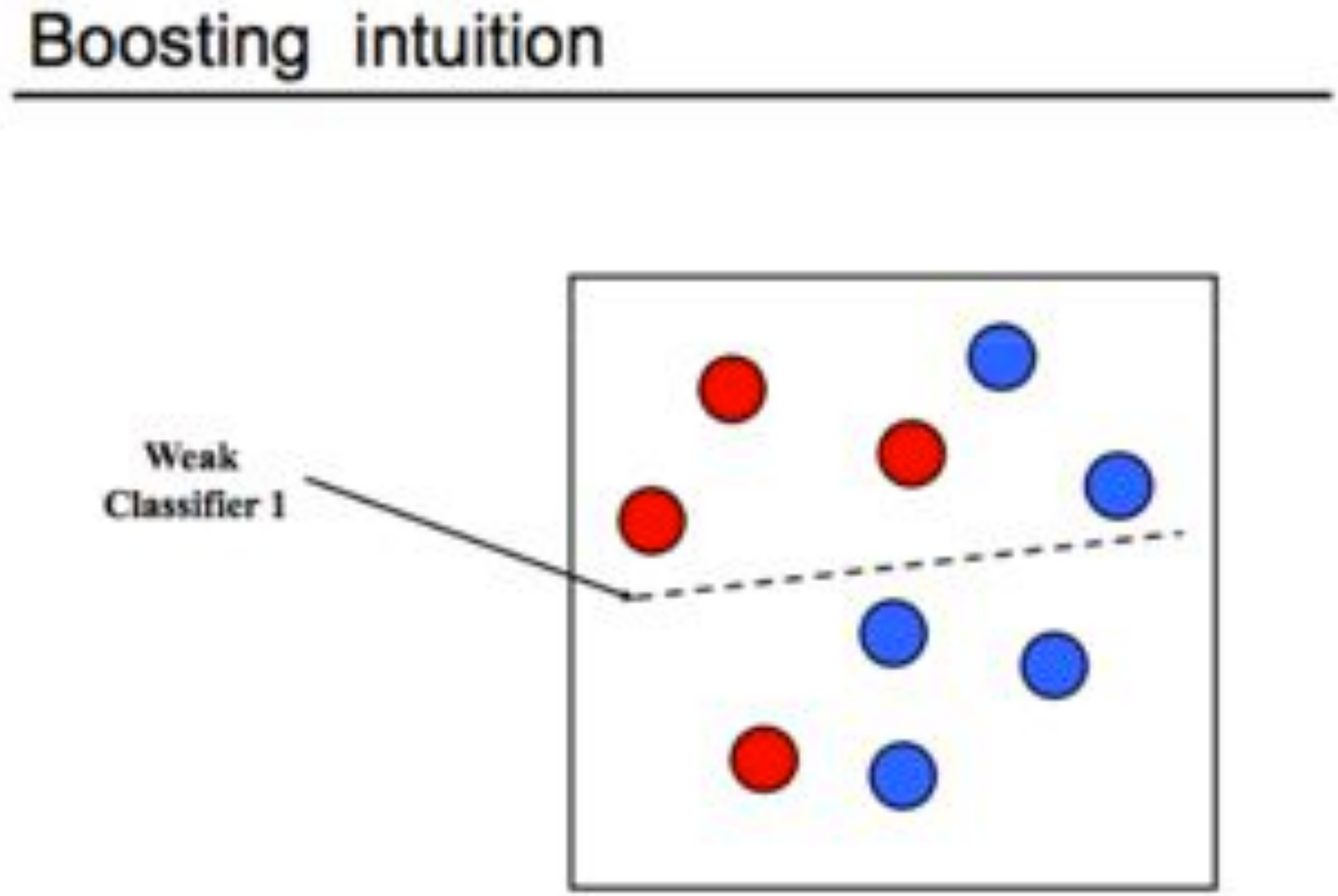
Edge Orientations, No Norm, Histogram (o.n.h)

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Classifier architecture…

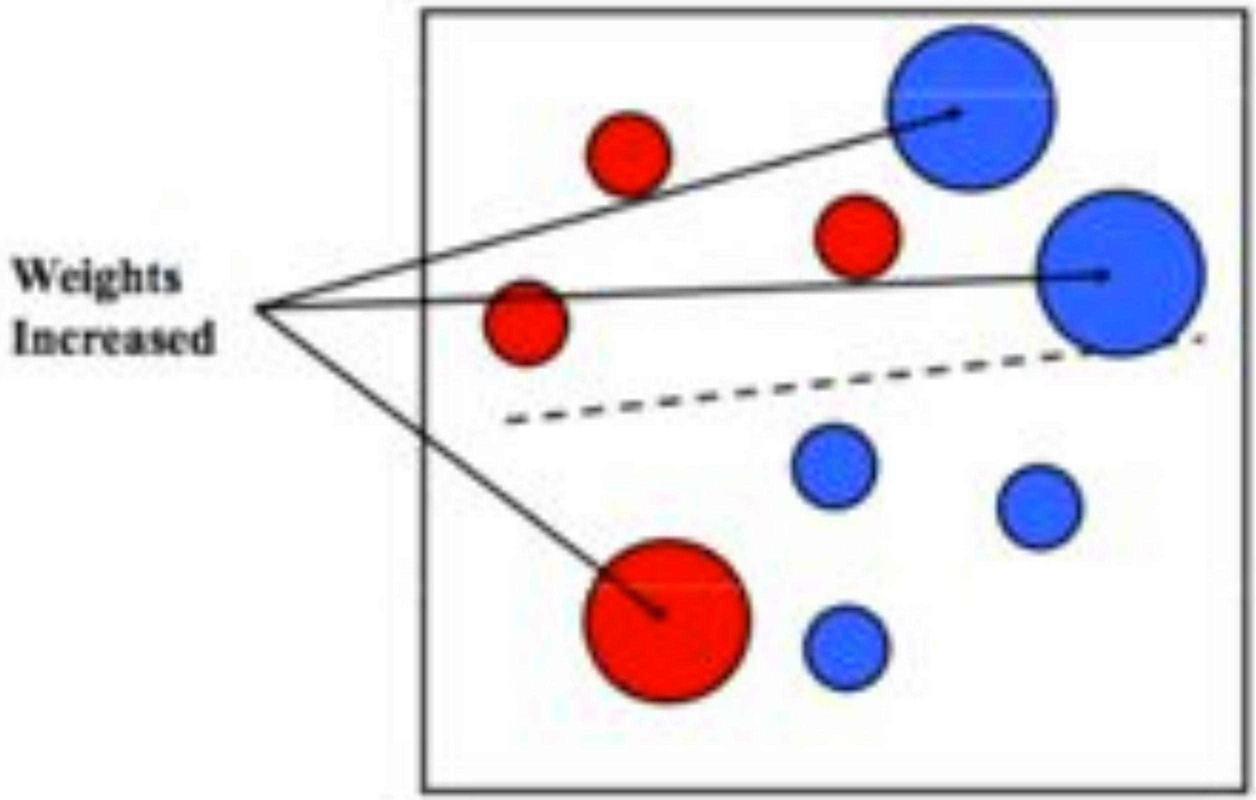
- Recent state of the art results in classification have mainly been achieved with SVMs
- The problem with SVMs is that **irrelevant features** can confuse/over-train the classifier…
- E.g. It might not make sense to use all facial pixels for training a classifier for just "is smiling"
- Given the large set of types of features/regions, we need a good way of selecting an optimal combination of features for each attribute…
- Enter Adaboost…

Quick Review of Boosting…



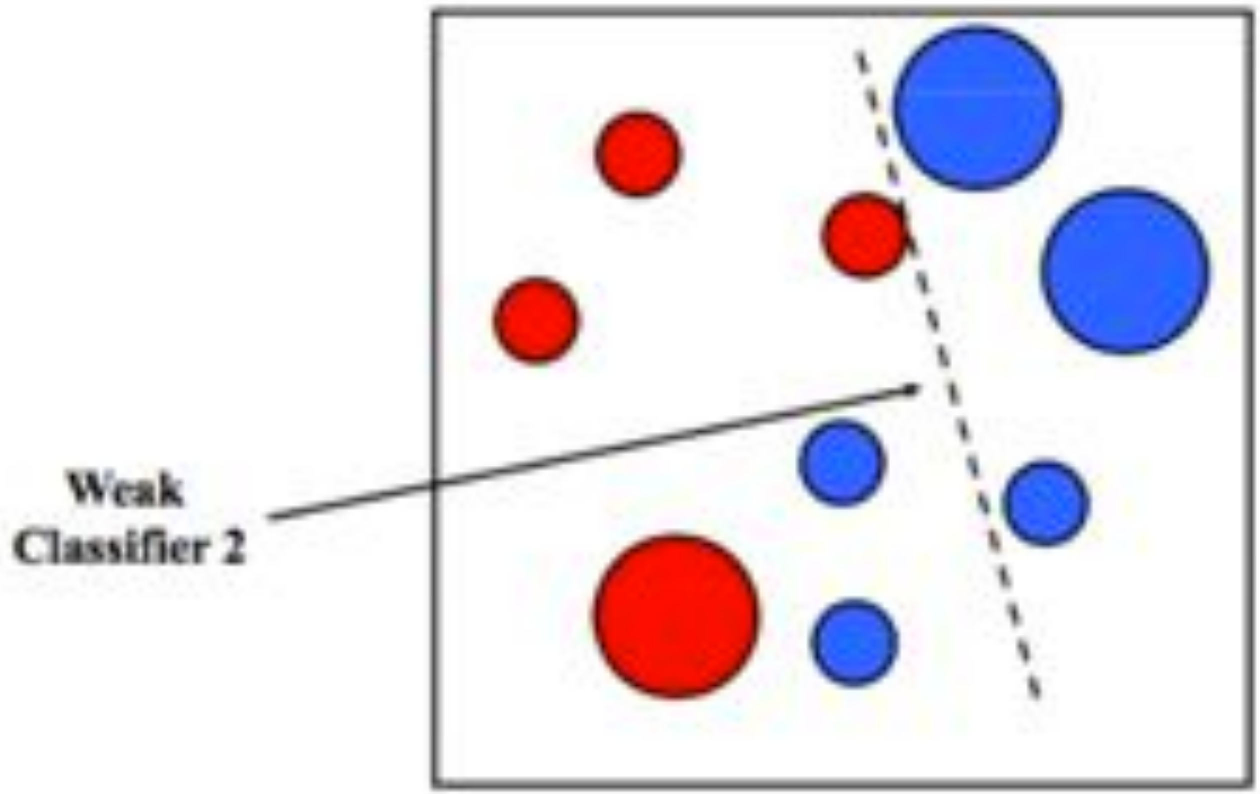
Source: http://www.cs.utexas.edu/~cv-fall2012/slides/fall2012_04_categories_part1.pdf

Boosting illustration



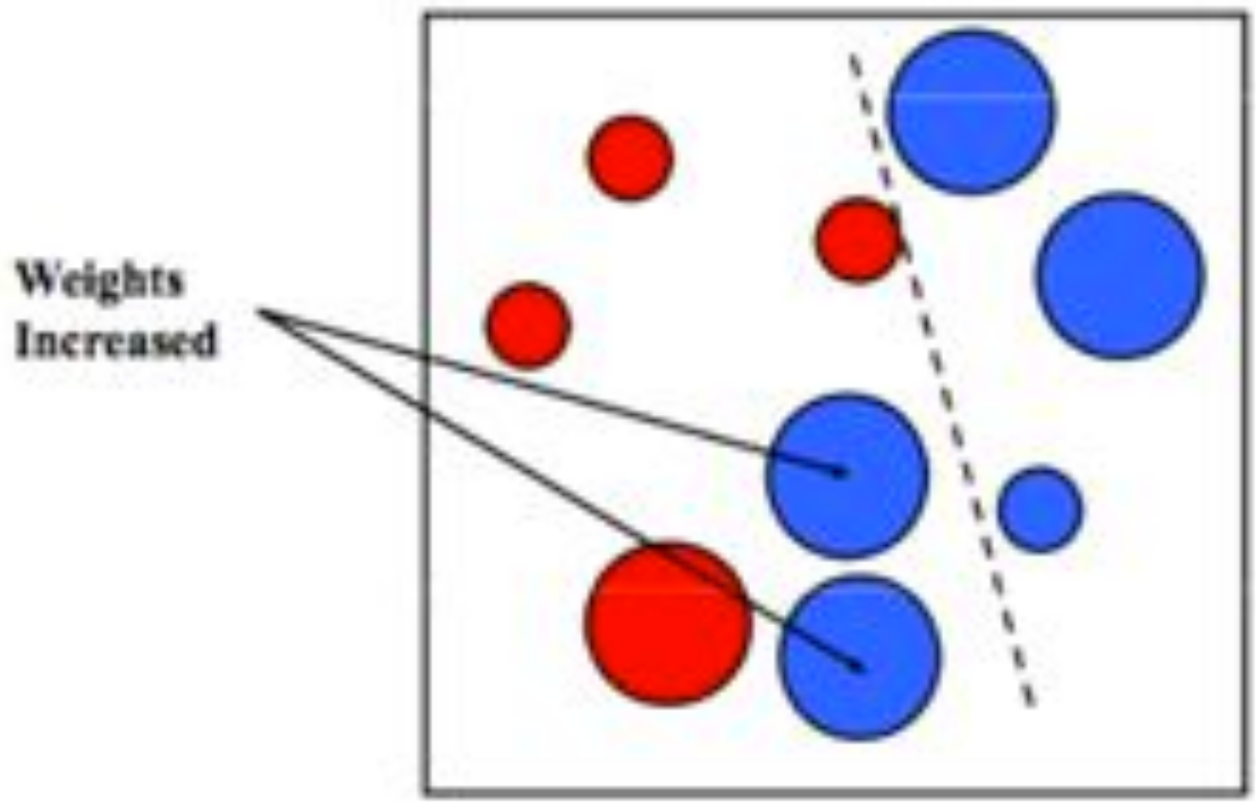
Source: http://www.cs.utexas.edu/~cv-fall2012/slides/fall2012_04_categories_part1.pdf

Boosting illustration



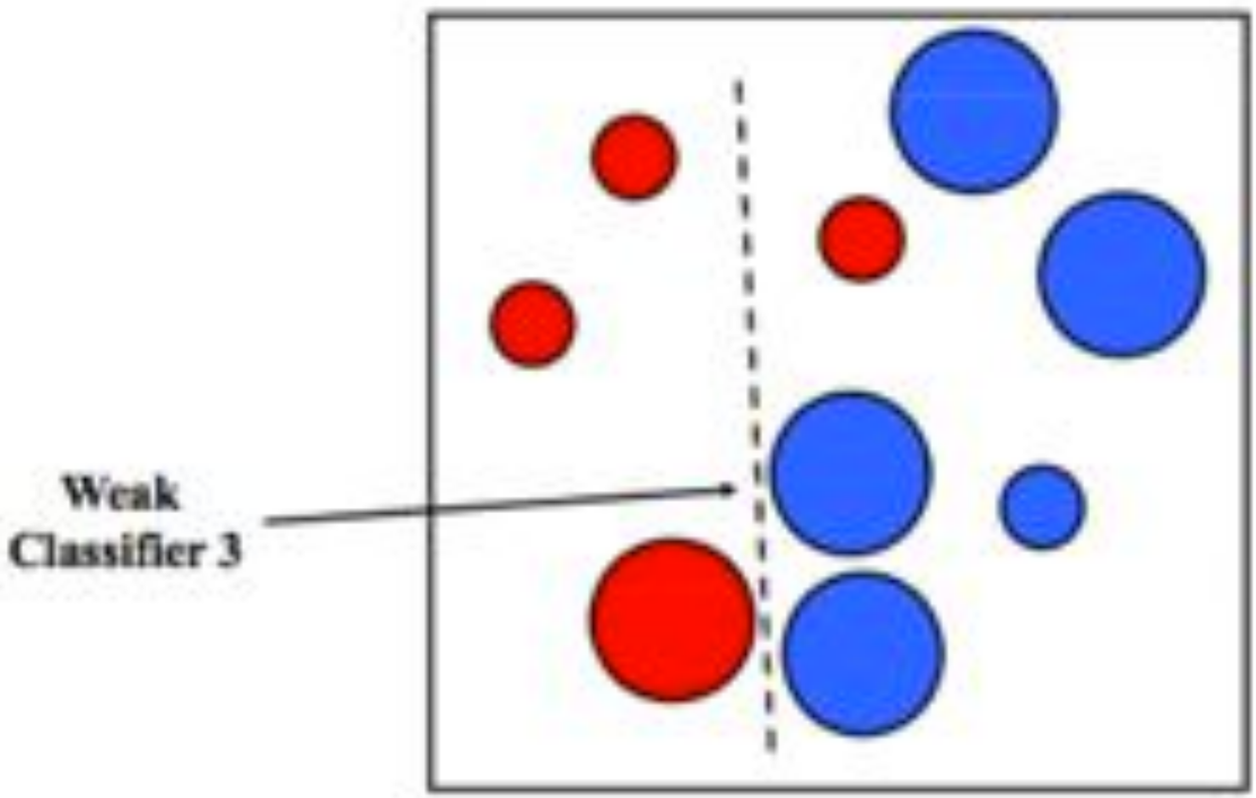
Source: http://www.cs.utexas.edu/~cv-fall2012/slides/fall2012_04_categories_part1.pdf

Boosting illustration



Source: http://www.cs.utexas.edu/~cv-fall2012/slides/fall2012_04_categories_part1.pdf

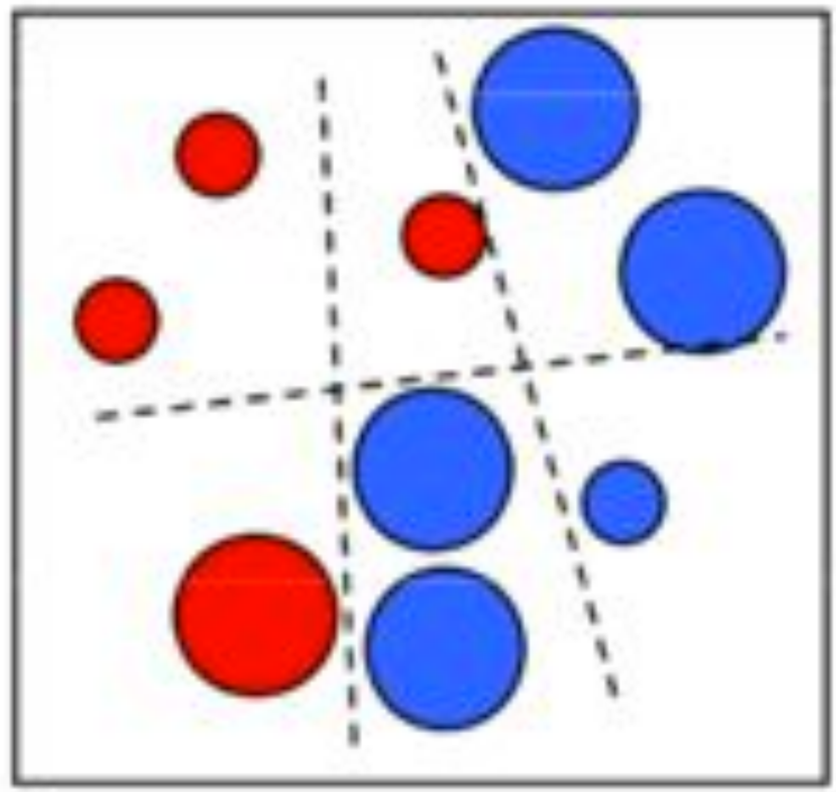
Boosting illustration



Source: http://www.cs.utexas.edu/~cv-fall2012/slides/fall2012_04_categories_part1.pdf

Boosting illustration

Final classifier is a combination of weak classifiers



Source: http://www.cs.utexas.edu/~cv-fall2012/slides/fall2012_04_categories_part1.pdf

Combining Boosting with SVMs…

- The idea is to **construct a "local" SVM for every possible combination** of region, feature types and SVM parameters (LibSVM)
- And then to **use Adaboost** to create an optimal classifier using a **linear combination of these local SVMs**
- The usual Adaboost algorithm is modified so that no retraining is needed at the beginning of each round (since these SVMs are either powerful/useless classifiers depending on the relevance of the features used)

Discussion…

- Boosting is meant to turn weak learners into strong learners. Does using boosting in this scenario where you have pre-trained SVMs make sense? Wouldn't using some feature selection approach be better?
- Performance degradation in boosting, (Wickramaratna, J. and Holden, S. and Buxton, B., Multiple Classifier Systems, **2001**) shows that boosting strong learners can cause performance degradation

Discussion…

- While in this paper, they assumed that since SVMs were either powerful/useless learners, the normal retraining step in Adaboost wasn't needed, they have a related follow-up work (Attribute and Simile Classifiers for Face Verification, N. Kumar, A. Berg, P. Belhumeur, S. Nayar. ICCV 2009) where they use forward feature selection instead of Adaboost.
- While they don't get much better results, their system isn't restricted to only frontal poses.

Train Classifiers

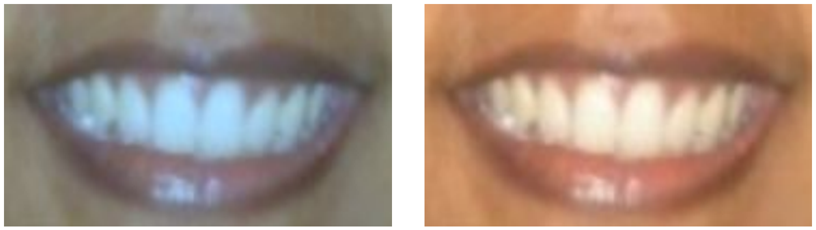
Mouth

Raw RGB

Pool of Classifiers

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Train Classifiers



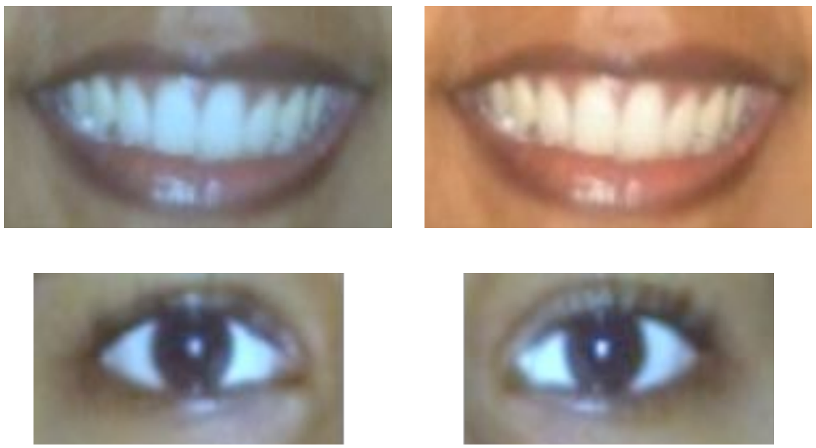
Eyes

Mean-Normalized RGB

Pool of Classifiers

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Train Classifiers



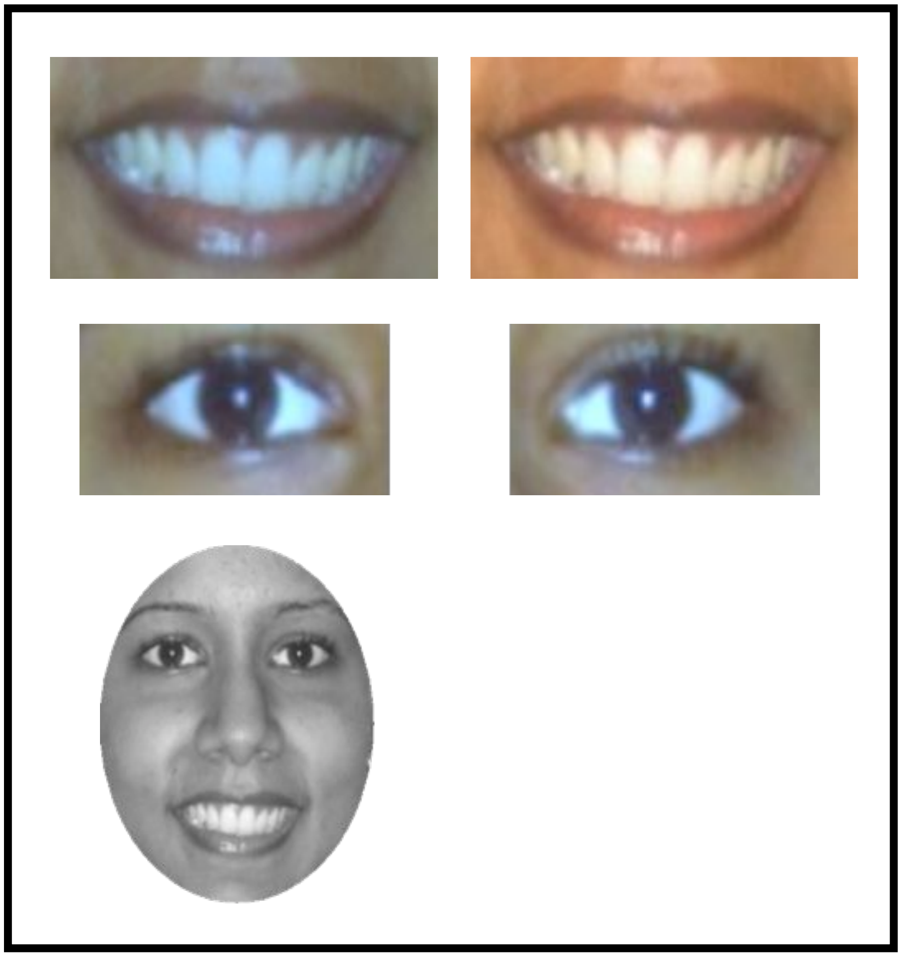
Pool of Classifiers

Whole Face

Raw Intensity

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Train Classifiers



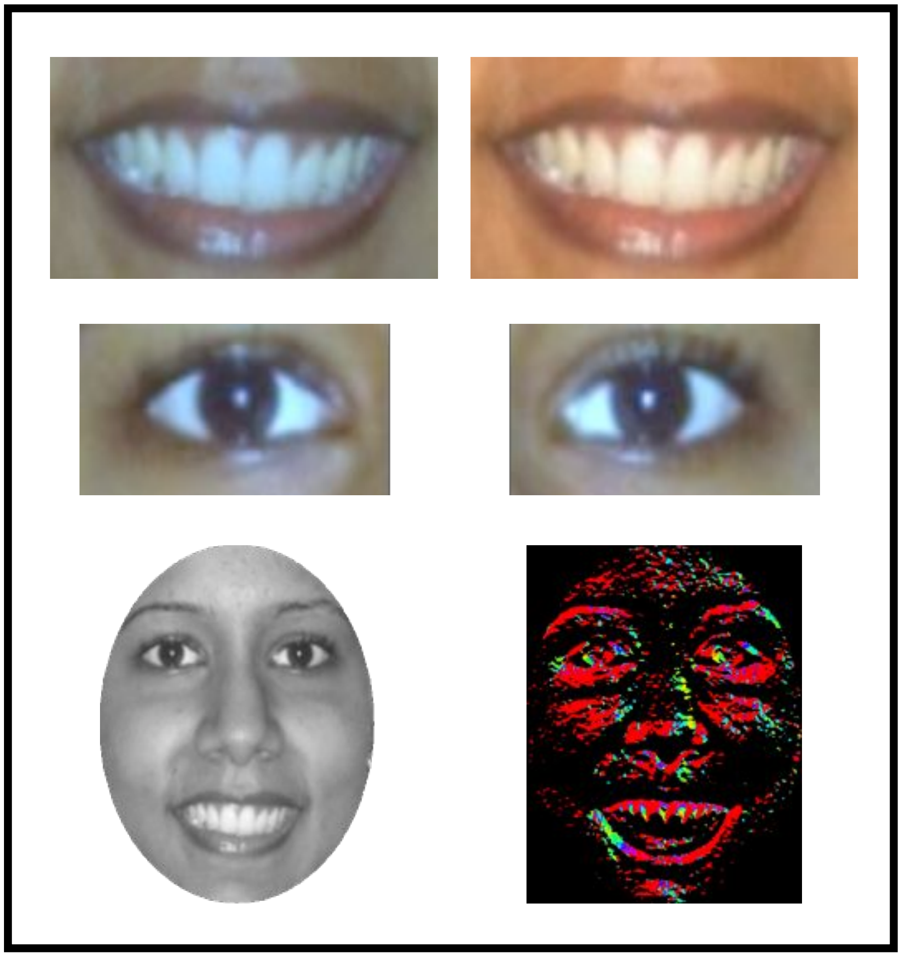
Pool of Classifiers

Whole Face

Gradient Directions

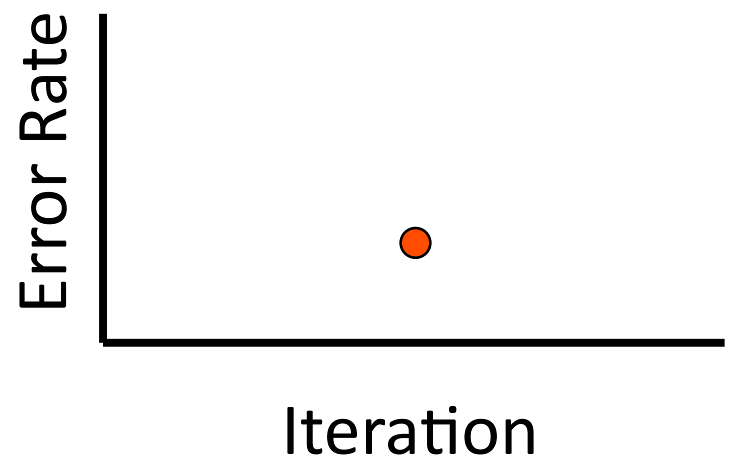
Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Select Classifiers



Pool of Classifiers

Selected Classifiers



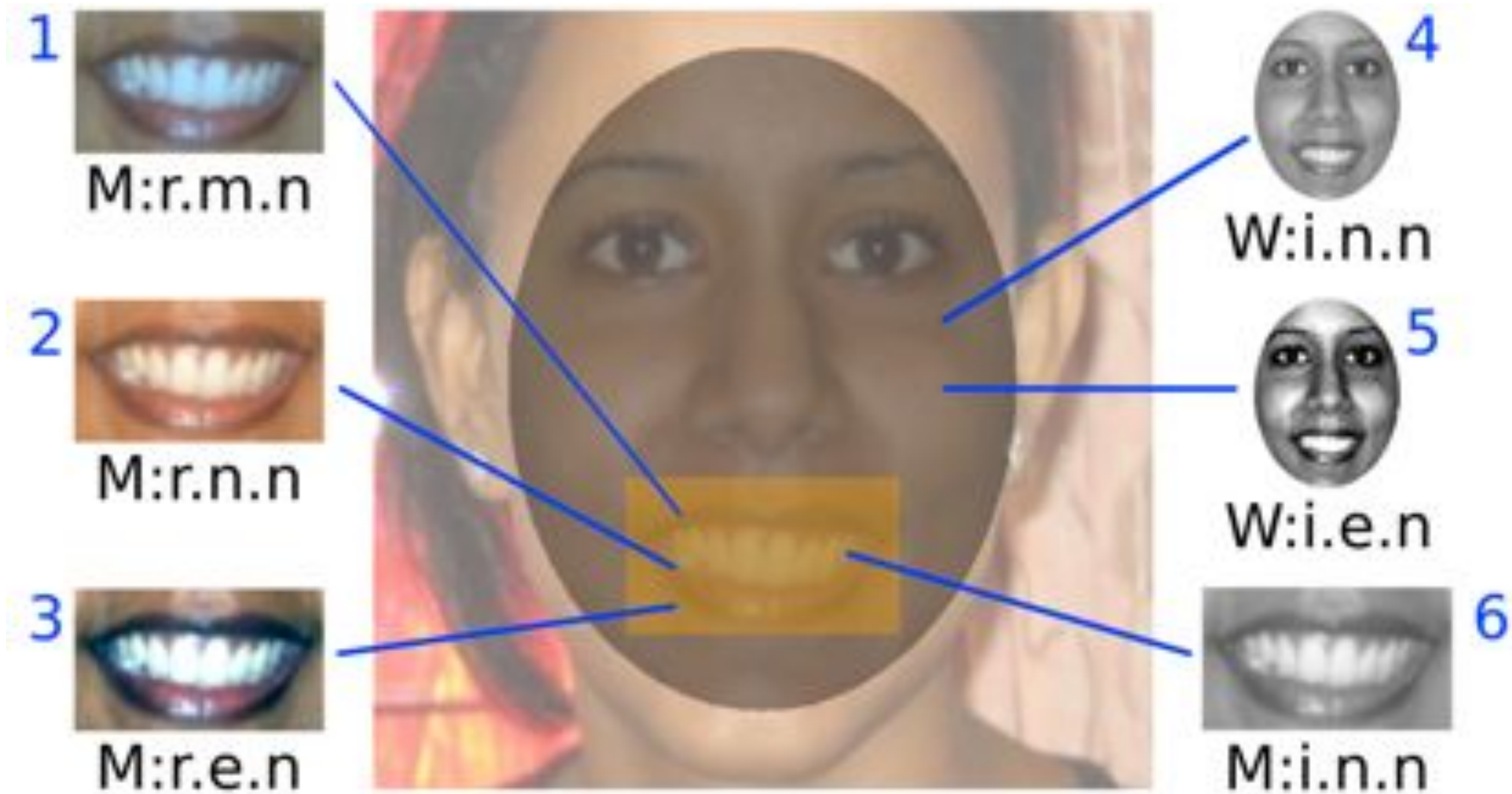
Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Feature Selection: Smiling

1. Mouth: RGB, Mean Norm., No Aggreg. (M:r.m.n)
2. Mouth: RGB, No Norm., No Aggreg. (M:r.n.n)
3. Mouth: RGB, Energy Norm., No Aggreg. (M:r.e.n)
4. Whole Face: Intensity, No Norm., No Aggreg. (W:i.n.n)
5. …

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

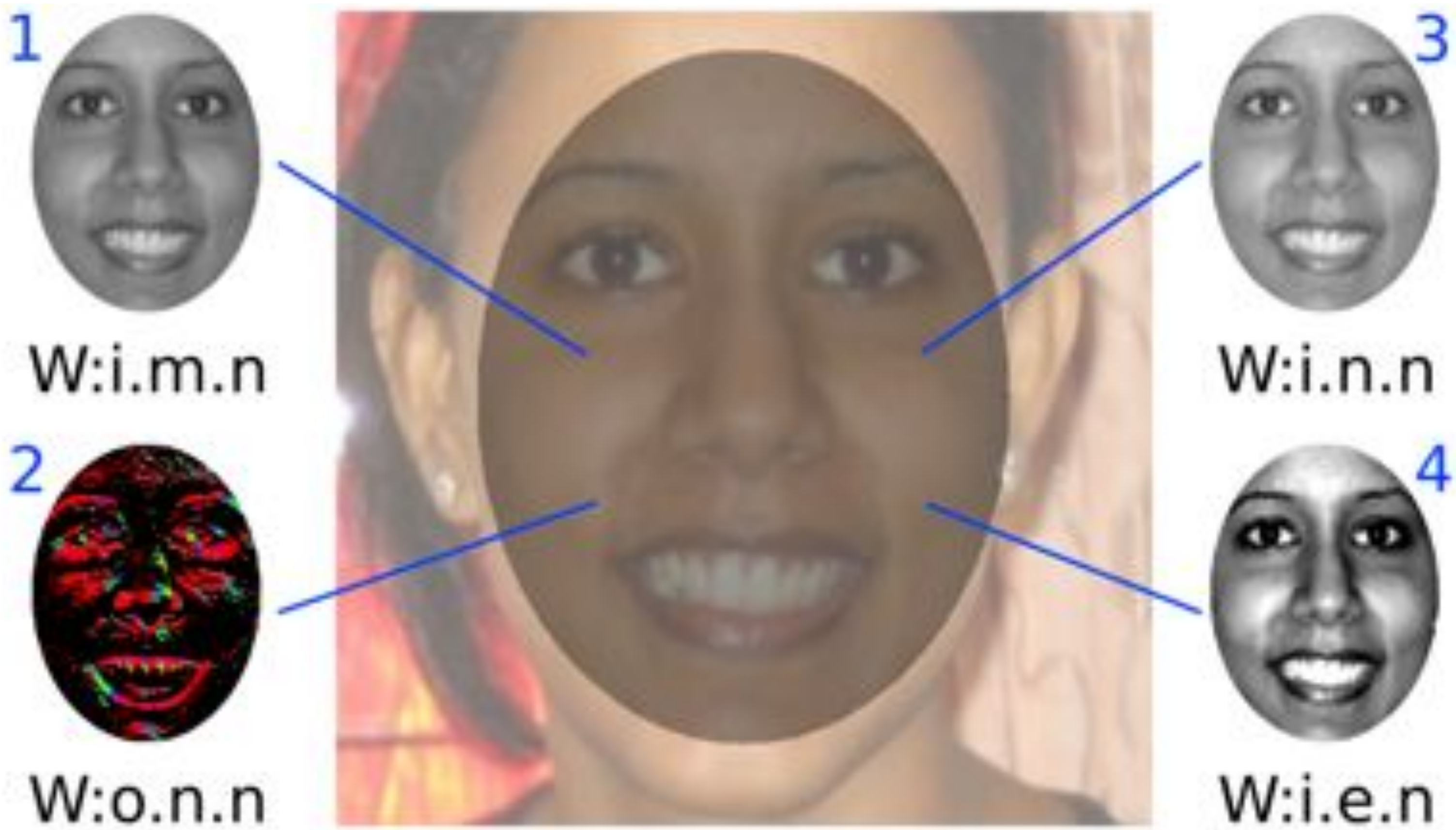
Selected Features



Smiling

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

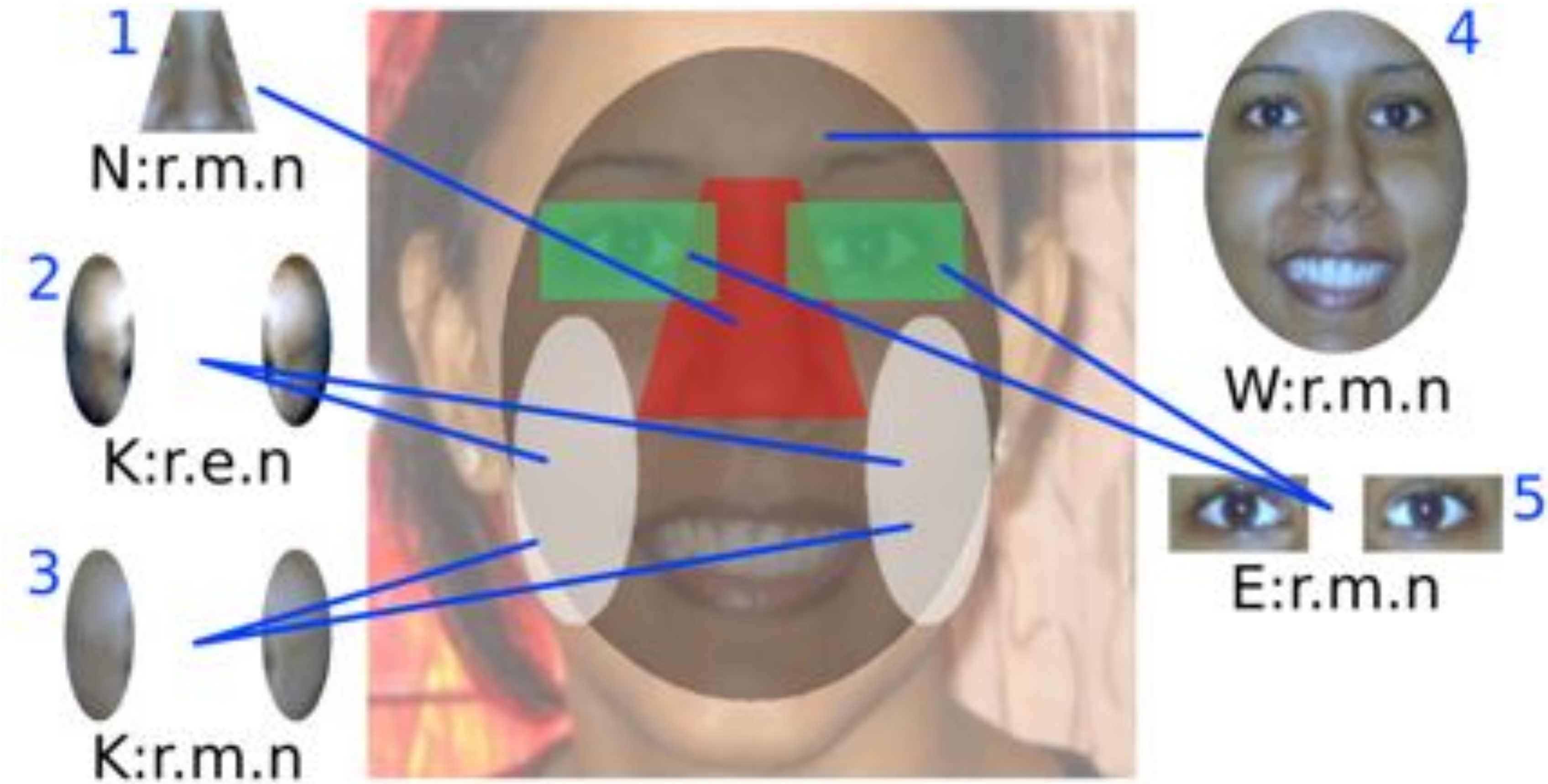
Selected Features



Gender

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

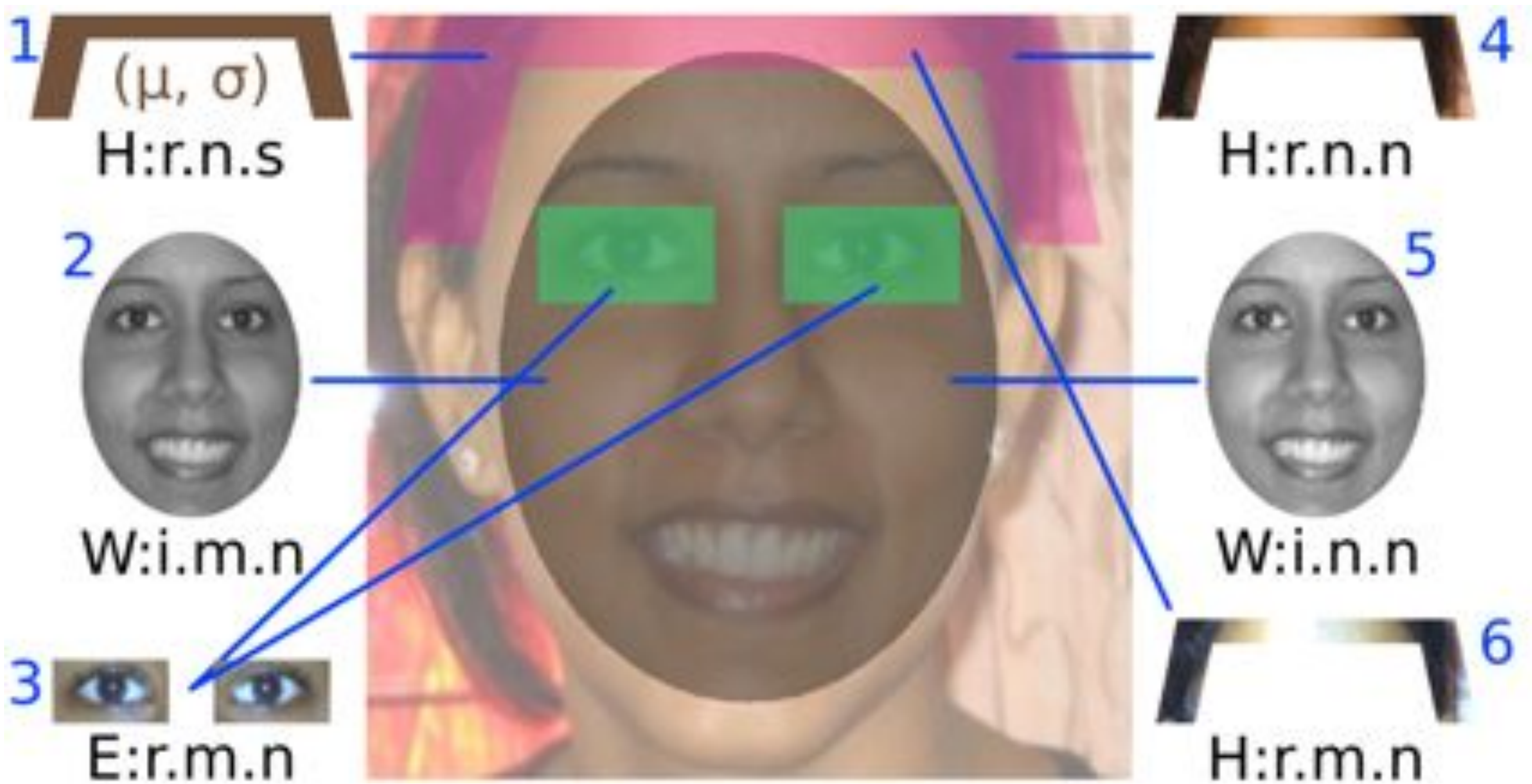
Selected Features



Indoor/Outdoor

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Selected Features



Hair Color

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

One global SVM to rule them all…

- The **drawback** of such a local SVM based architecture is that it requires **keeping a large number of SVMs in memory and evaluating all** of them for every new input image…
- This is solved by training **one global SVM** on the **union** of the features from the top N highest weighted (by Adaboost) SVMs
- The next slide shows the comparative error rates for the local v/s global SVM approach…

Discussion…

- While concatenating the local feature sets (for creating the global SVM from the local top ranked SVMs) they do not seem to use the Adaboost weights/scores for those feature-region sets. Could this help?

Classification Accuracy

Attribute	Error Rates for Attribute-Tuned Local SVMs	Error Rates for Attribute-Tuned Global SVM	Top Feature Combinations in Ranked Order Each combination is represented as Region:pixtype.norm.aggreg
Gender	9.42%	8.62%	W:i.m.n | W:o.n.n | W:i.n.n | W:i.e.n
Age	17.34%	16.65%	W:i.m.n | W:i.n.n | H:r.e.n | E:r.m.n | H:r.e.s | W:o.n.n
Race	7.75%	6.49%	W:i.m.n | E:r.e.n | C:o.n.n | M:r.m.n | W:o.n.n
Hair Color	7.85%	5.54%	H:r.n.s | W:i.m.n | E:r.m.n | H:r.n.n | W:i.n.n | H:r.m.n
Eye Wear	6.22%	5.14%	W:m.n.n | W:i.n.n | K:o.n.h | W:m.m.n | N:r.n.n
Mustache	6.42%	4.61%	U:r.e.n | M:r.m.n
Smiling	4.60%	4.60%	M:r.m.n | M:r.n.n | M:r.e.n | W:i.n.n | W:i.e.n | M:i.n.n
Blurry	3.94%	3.41%	W:m.m.n | H:m.n.n | W:m.n.n | H:m.m.n | M:m.m.n
Lighting	2.82%	1.61%	W:i.n.n | W:i.e.n | K:r.n.n | C:o.n.n | E:o.n.n
Environment	12.25%	12.15%	N:r.m.n | K:r.e.n | K:r.m.n | W:r.m.n | E:r.m.n

Source of slide: The corresponding ECCV 2008 paper

Comparison to State-of-the-Art

Method	Gender Error Rate	Smiling Error Rate
Proposed	8.62%	4.60%
Adaboost (pixel comparison feats) Baluja & Rowley, IJCV 2007	13.13%	7.41%
Adaboost (Haar-like features) Shakhnarovich et al., ICAFGR 2002	12.88%	6.40%
Full face SVM Moghaddam & Yang, TPAMI 2002	9.52%	13.54%

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Discussion…

- This approach seems to boil down to feature selection. Could their improved results be because of their rich set of features rather than the Boosting+SVMs approach?

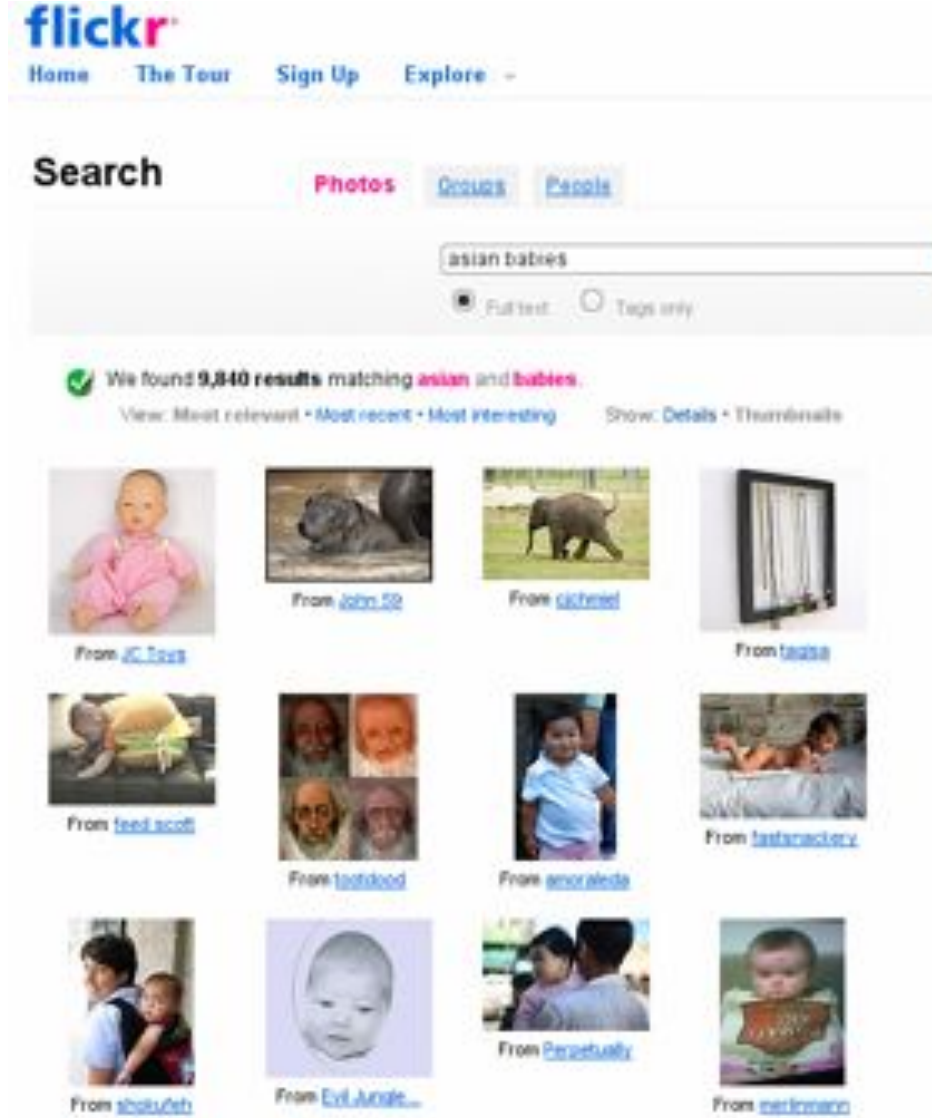
FaceTracer Engine…

- Step1: [Offline] Training attribute-tuned global SVM classifiers for each of the 10 attributes
- Step2: [Offline] All the images in the database are sent through these classifiers to be labeled
- These labels are stored for fast **online** search.
- The search interface accepts simple text based queries and maps these onto the labels by sing a dictionary of terms.
- It returns the results in the order of decreasing confidence.
- Applications: law enforcements, social networking sites, personal snap collection management

Discussion…

- They mention that for multiple-attribute query terms, they convert the classifier confidences into probabilities and then use the product of these probabilities for scoring/ranking. Is this the right approach? Don't these different attribute classification scores need to be calibrated properly?
- Multi-Attribute Spaces: Calibration for Attribute Fusion and Similarity Search. W. Scheirer, N. Kumar, P. Belhumeur, T. Boult. CVPR 2012

“Asian Babies”

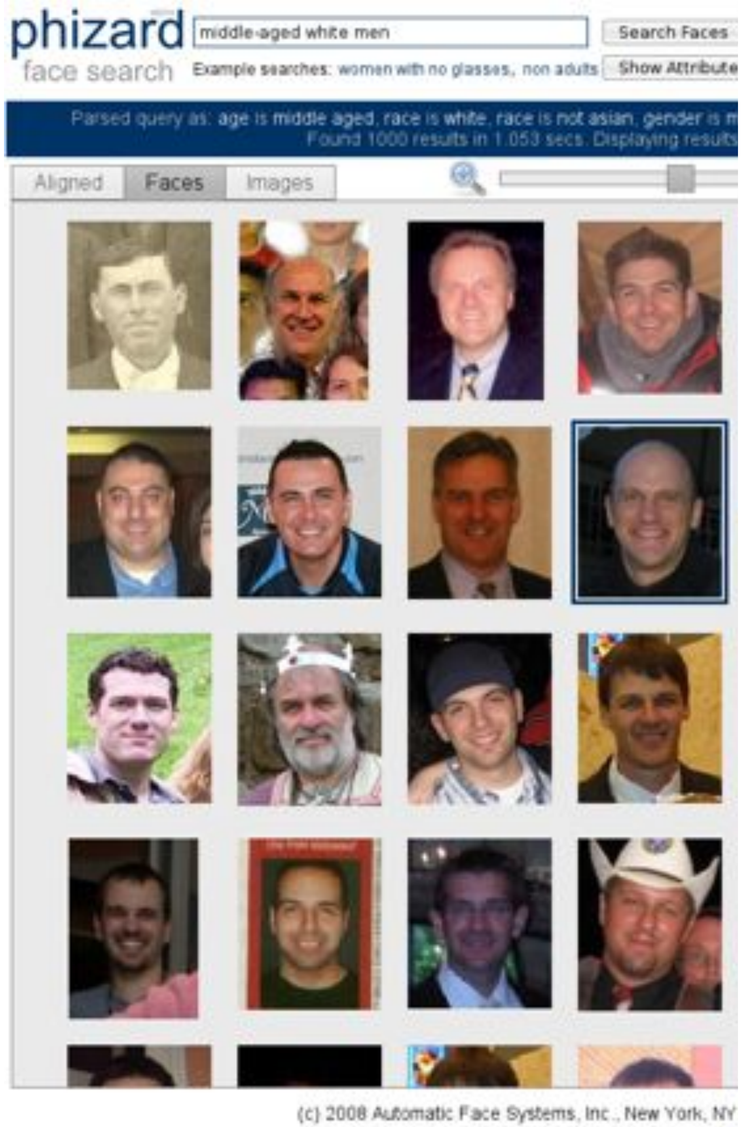


Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

“Adults Outside”

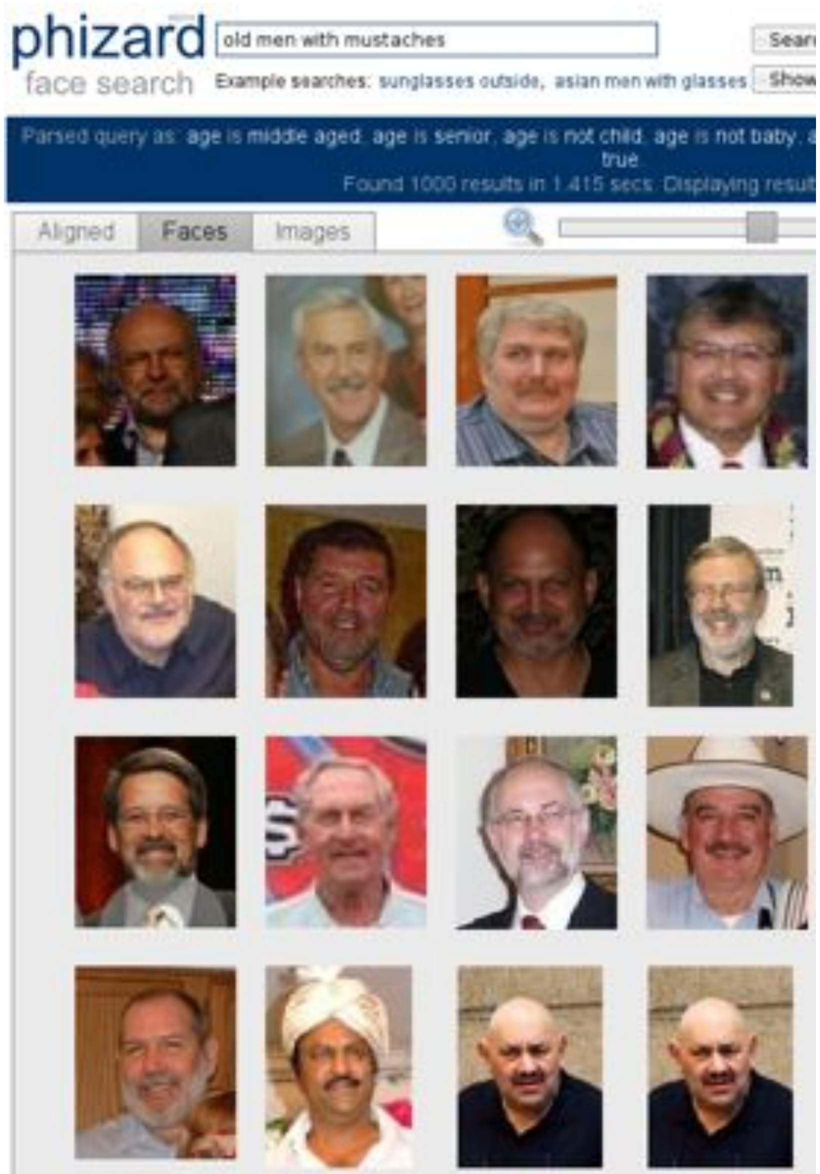
Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

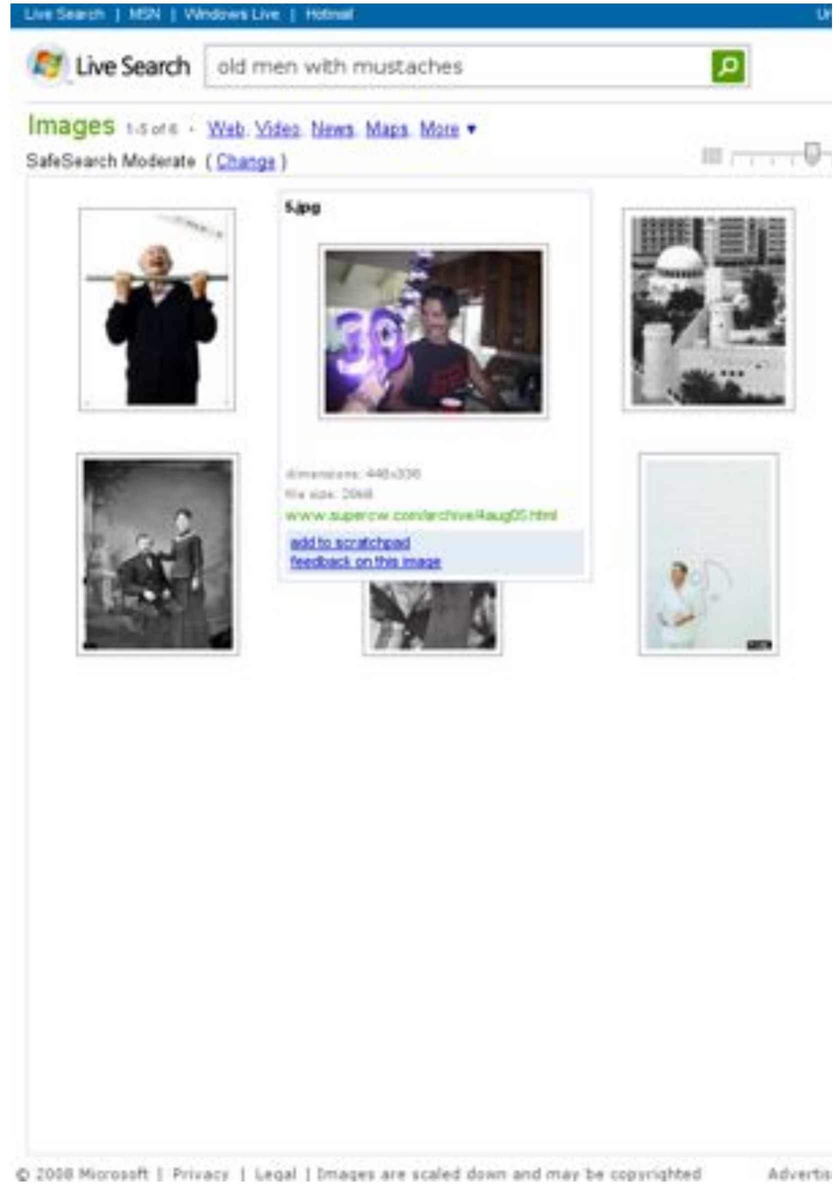
“Middle-Aged White Men”



Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

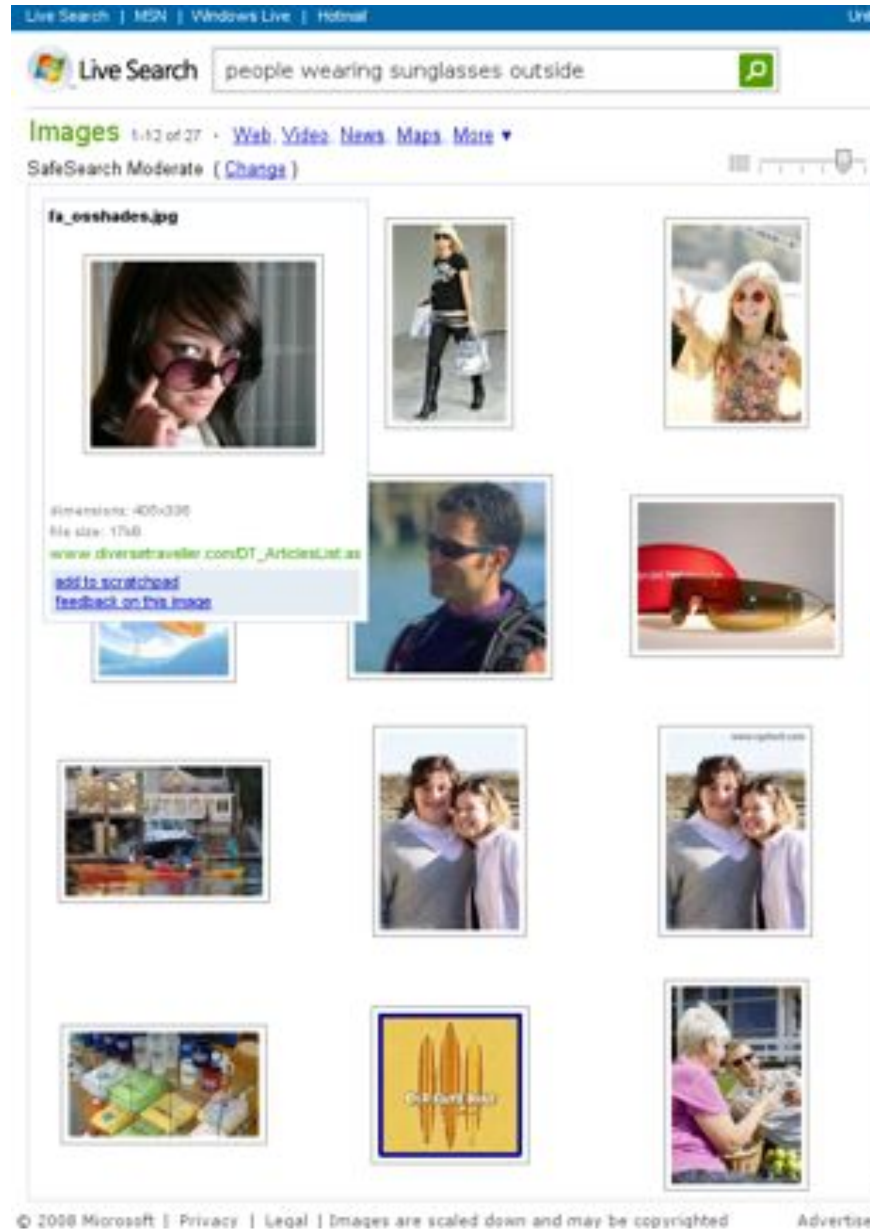
"Old Men With Mustaches"





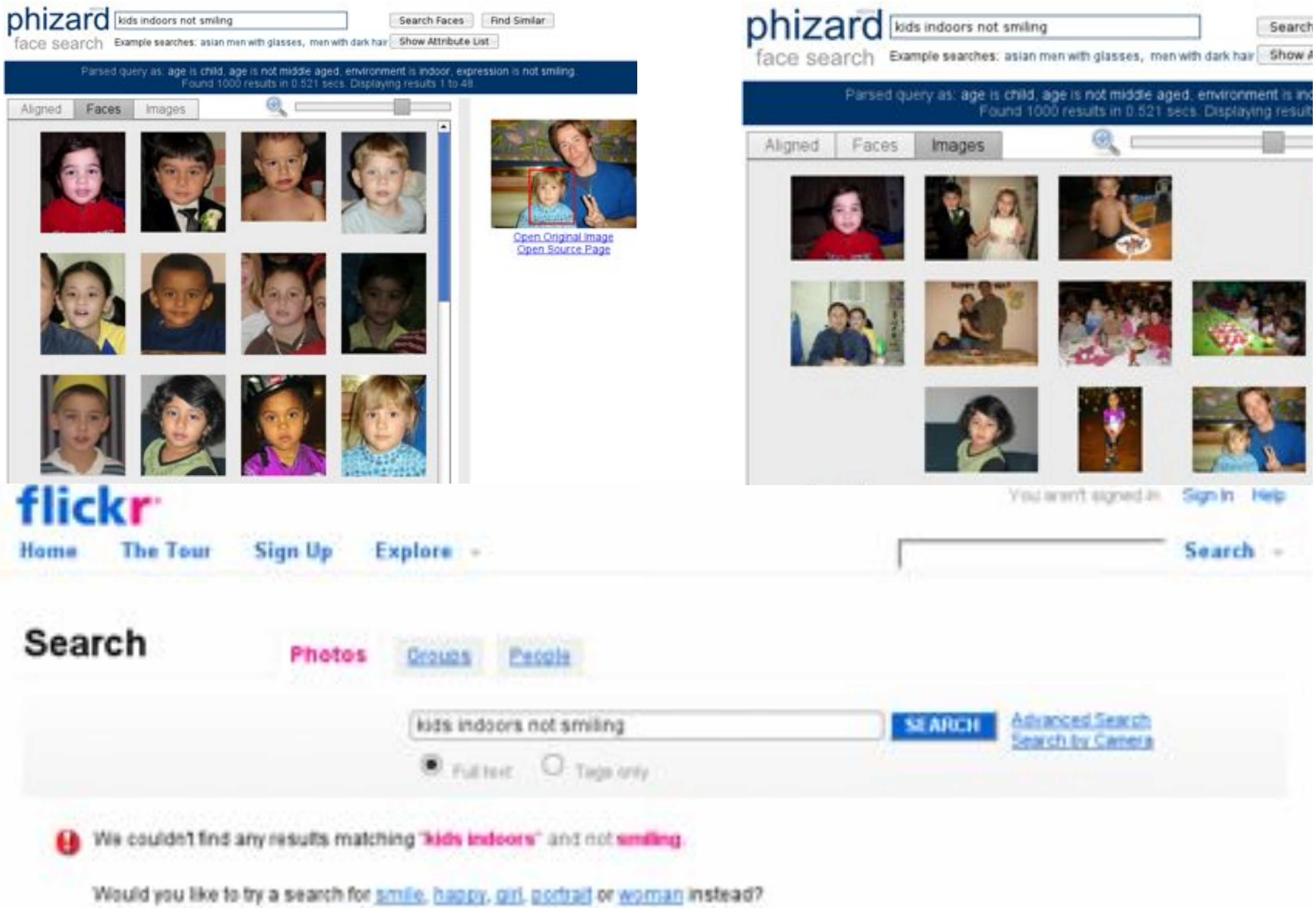
Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

“People Wearing Sunglasses Outside”



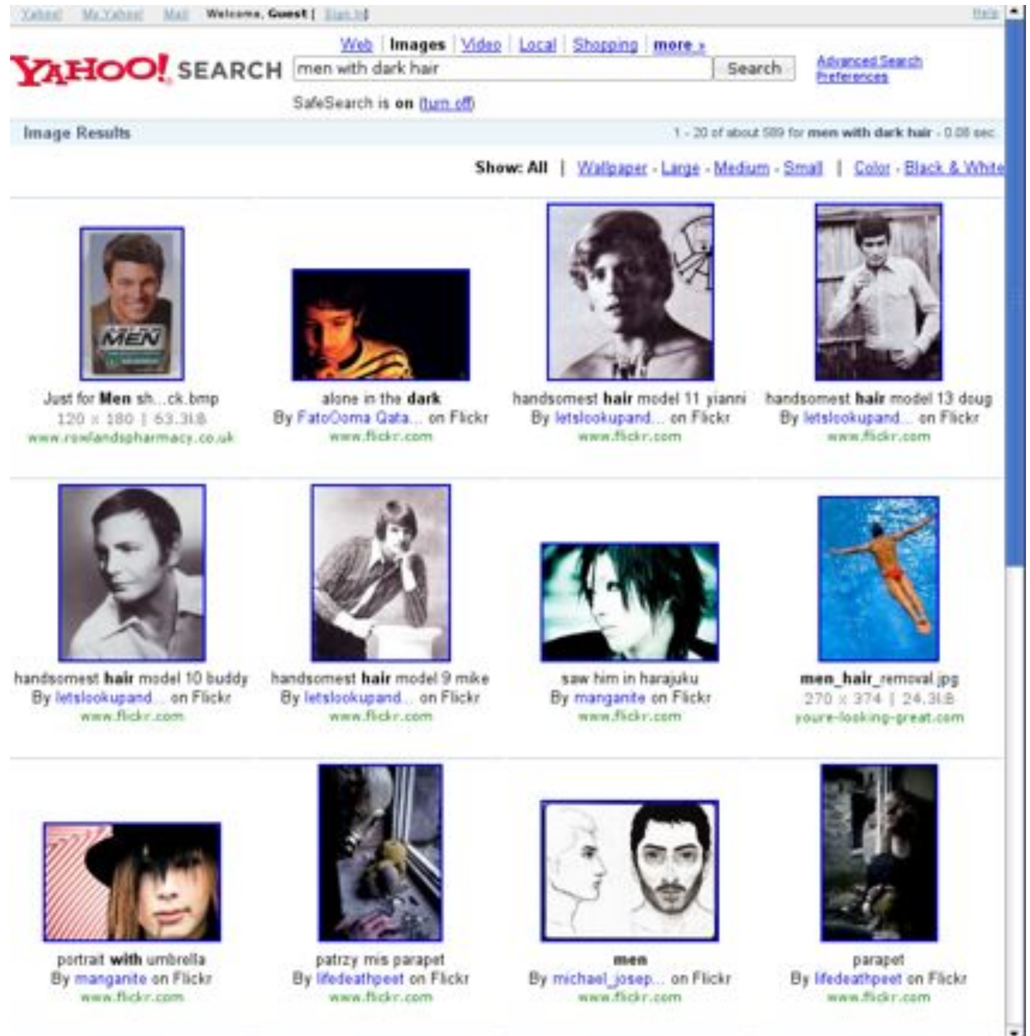
Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

“Kids Indoors Not Smiling”



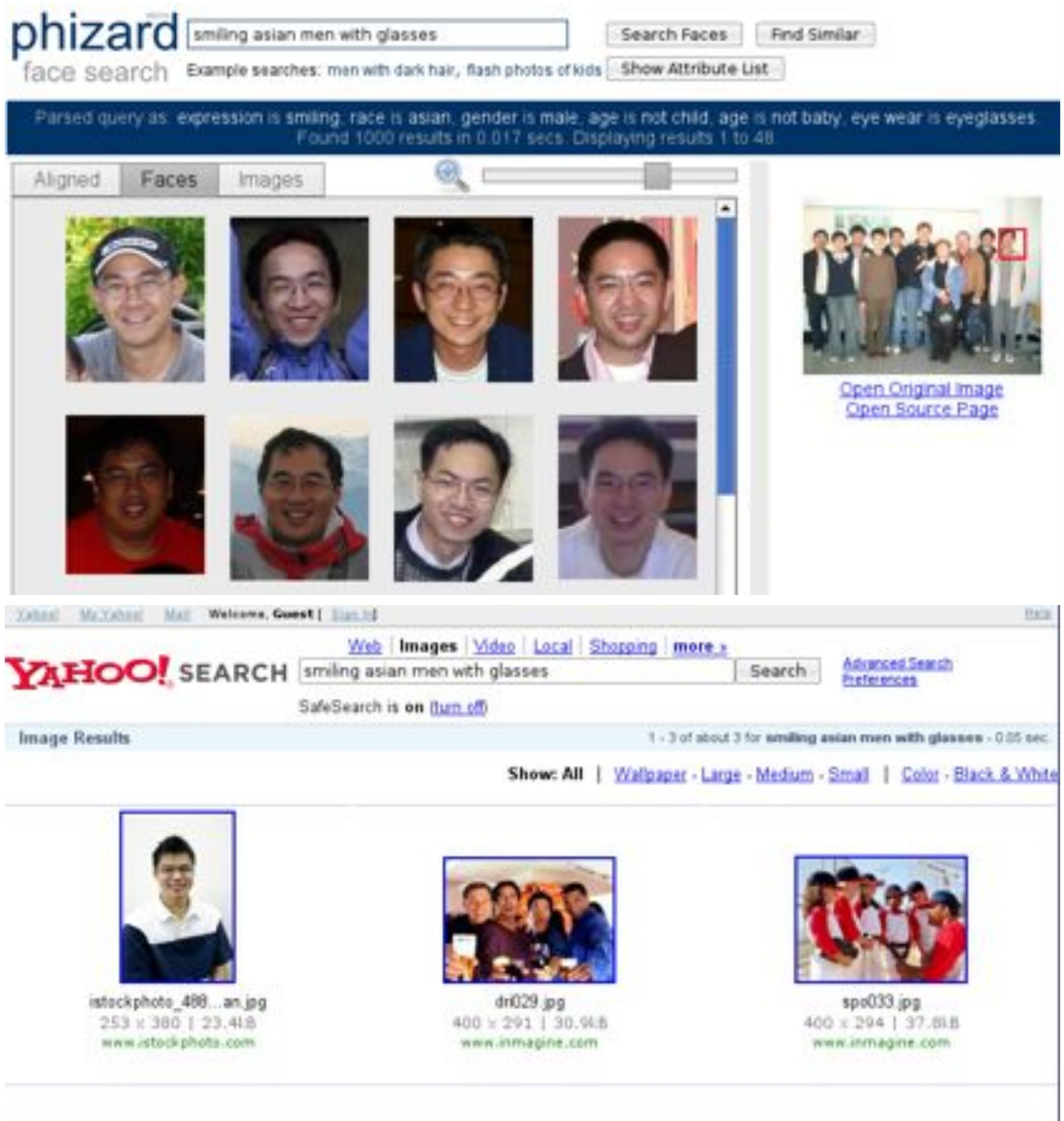
Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

“Men With Dark Hair”



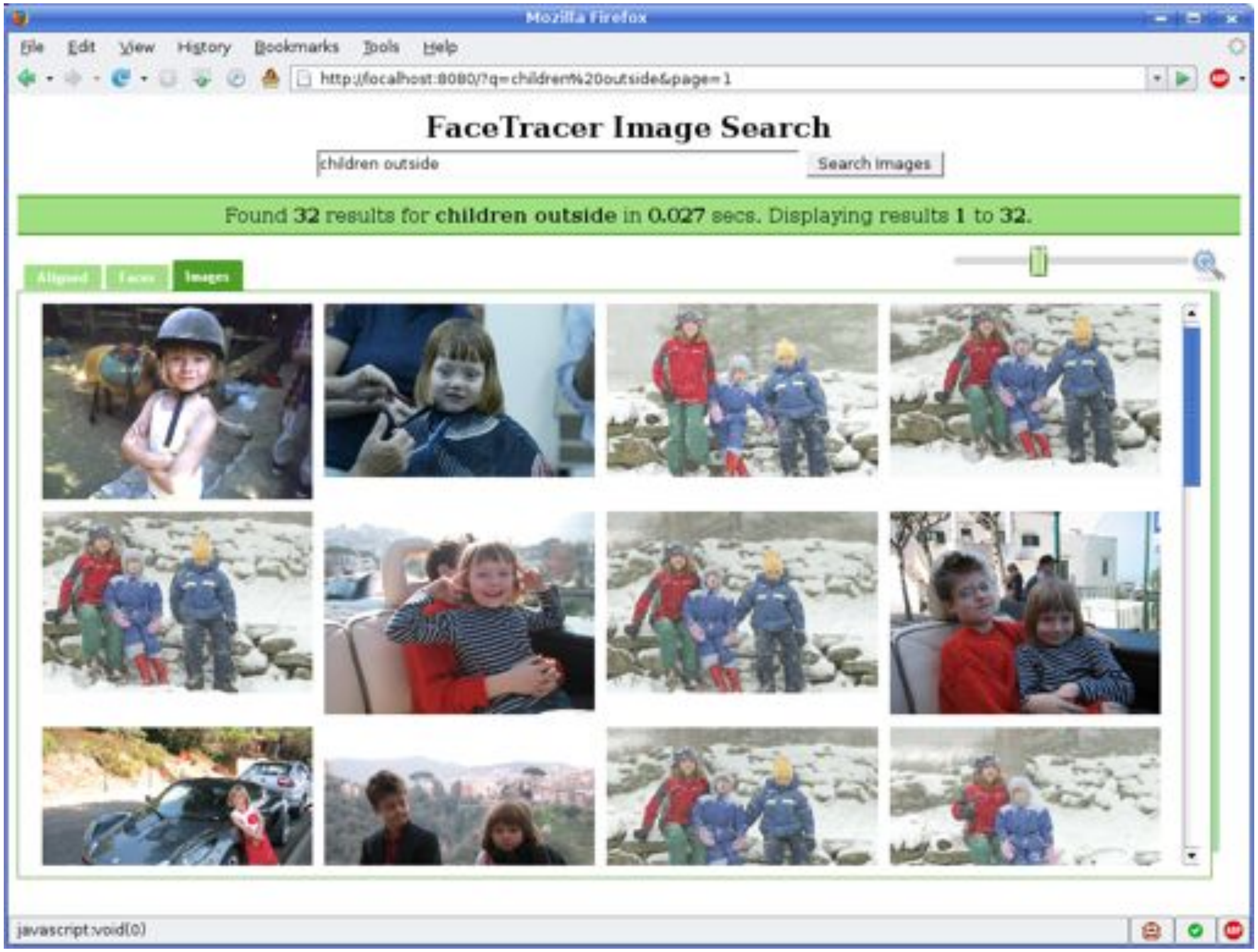
Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

“Smiling Asian Men With Glasses”



Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Personal FaceTracer Search



“Children outside”

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Discussion…

- While this method works very well for descriptive attribute based search on facial images, is **it scalable to the general type of queries** that most image search engines use?

Major contribution of this paper

The idea of allowing people to search for faces with descriptive terms by learning nameable semantic attributes for facial images.

Strengths

- The idea of combining Boosting with SVMs
- This helps on two fronts:

1. SVMs are powerful classifiers unlike the usual "weak" classifiers that Boosting is used on. This can help where Boosting usually fails.
2. Boosting helps in selecting the optimal set of features from the variety of feature/region choices available.

- Combining Boosting with SVMs seems to obtain better results than the state-of-the-art approaches that use solely Boosting or SVMs.

Strengths

- Finds an optimal set of relevant features/ regions for training for each attribute
- Approach implemented on the largest collection of images of "real-world" faces.
- Easily extensible to new attributes
- Handles both facial attributes & image attributes

Weaknesses

- Limited to frontal poses only
- The methods they compare with do not use the same/similar set of features and instead use a relatively impoverished set of features. As such, they do not seem to be fair baselines

Related work

- "Attribute and Simile Classifiers for Face Verification," (oral presentation)
- Neeraj Kumar, Alexander C. Berg, Peter N. Belhumeur, Shree K. Nayar, Proceedings of the 12th IEEE International Conference on Computer Vision (ICCV),
- October 2009.

Prior approaches

Images → Low-level features → Verification

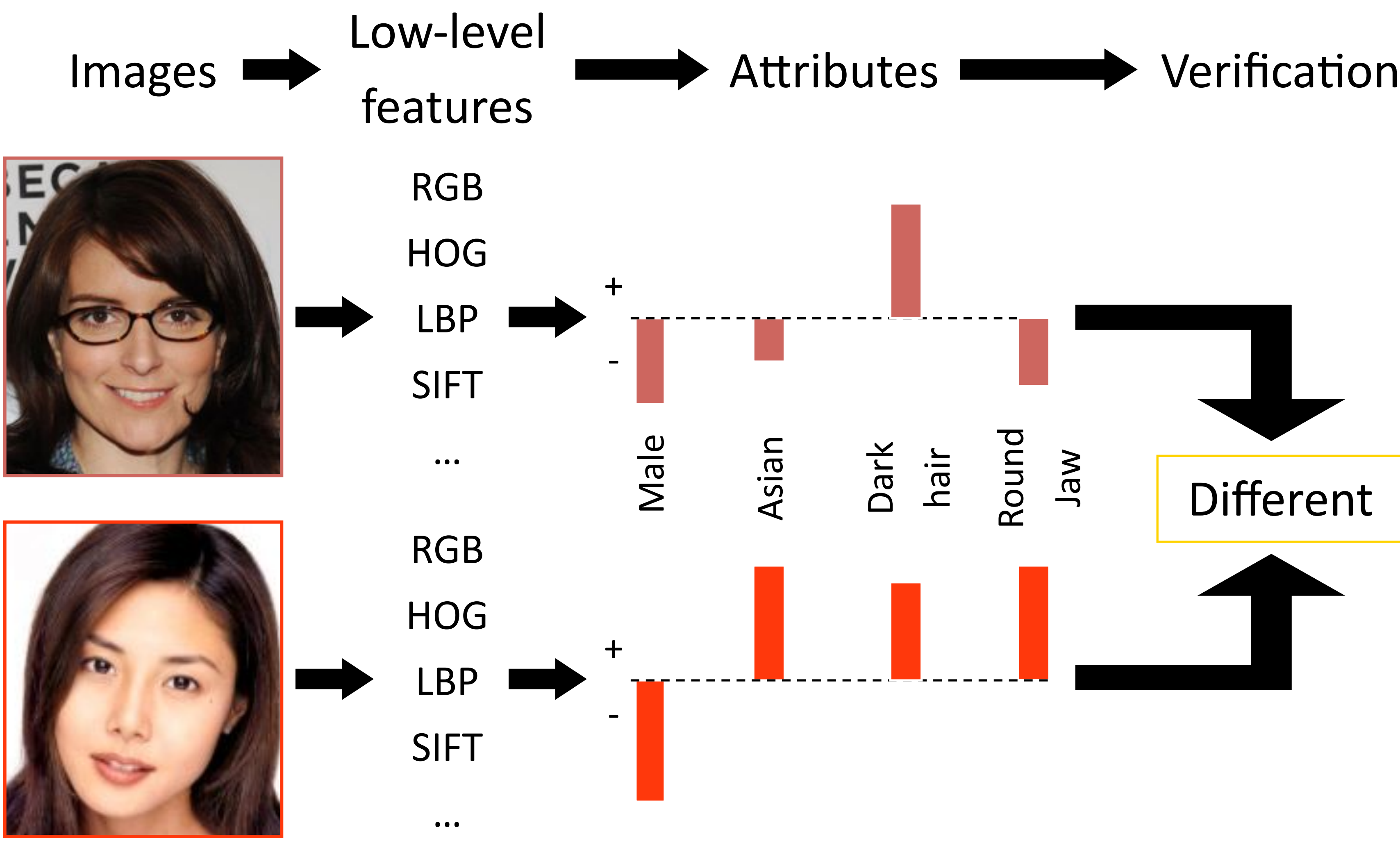
RGB
HOG
LBP
SIFT
…

RGB
HOG
LBP
SIFT
…

Different

Source of slide: http://homes.cs.washington.edu/~neeraj/projects/faceverification/

Their approach: attributes



Source of slide: http://homes.cs.washington.edu/~neeraj/projects/faceverification/

References/Resources

- http://homes.cs.washington.edu/~neeraj/projects/facesearch/
- Attribute and Simile Classifiers for Face Verification, N. Kumar, A. Berg, P. Belhumeur, S. Nayar. ICCV 2009
- Multi-Attribute Spaces: Calibration for Attribute Fusion and Similarity Search. W. Scheirer, N. Kumar, P. Belhumeur, T. Boult. CVPR 2012
- FaceTracer: A Search Engine for Large Collections of Images with Faces. N. Kumar, P. Belhumeur, and S. Nayar. ECCV 2008