EE 6886: Topics in Signal Processing
-- Multimedia Security System

Lecture 10: Biometric Authentication (III) –Face Recognition

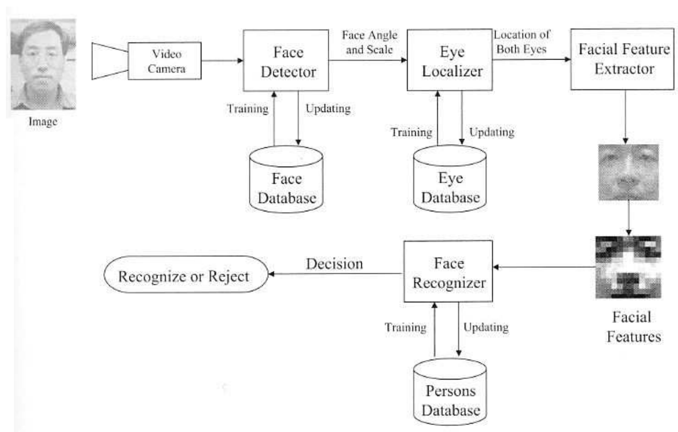
Ching-Yung Lin
Department of Electrical Engineering
Columbia University, New York, NY 10027

Course Outline

- Multimedia Security :
  - Multimedia Standards – Ubiquitous MM
  - Encryption and Key Management – Confidential MM
  - Watermarking – Uninfringible MM
  - Authentication – Trustworthy MM
- Security Applications of Multimedia:
  - Audio-Visual Person Identification – Access Control, Identifying Suspects
  - Surveillance Applications – Abnormality Detection
  - Media Sensor Networks – Event Understanding, Information Aggregation

Face Recognition -- Introduction

- Face Recognition System



Major Challenges on Face Detection and Recognition

- Pose: relative camera-face pose (frontal, 45 degree, profile, upside down)
- Presence or absence of structural components:
  - beards, mustaches, and glasses
  - A great deal of variability among these components
- Facial expression
- Occlusion
- Image orientation
- Imaging conditions:
  - Lighting (spectra, source distribution and intensity)
  - Camera characteristics (sensor response, lenses)

Many Aspects on Face Recognition-Related Topics

- Face localization: determine the image position of a single face, with the assumption that an input image contains only one face.
- Facial Feature Detection: detect the presence and location of features, such as eyes, nose, nostrils, eyebrow, mouth, lips, ears, etc.
- Face Identification: compare an input image against a database and reports a match.
- Face verification (or face authentication): verify the claim of the identity of an individual in an input image.
- Face tracking: continuously estimate the location and possibly the orientation of a face in an image sequence in real time
- Facial expression recognition: identify the affective states (happy, sad, disgusted, etc.) of humans.

Categories of Face Detection Algorithms

- Knowledge-based / Rule-based methods: use known human prior knowledge.
- Feature invariant approaches: aim to find structural features that exist even when the pose, viewpoint or lighting conditions vary, and then use these to locate faces.
- Template matching methods: Several standard face patterns are stored to describe the face as a whole or the facial features separately.
- Appearance-based methods / supervised learning methods: learn models or templates from a set of training images

Categorization of Methods for Face Detection in Image

Categorization of Methods for Face Detection in a Single Image

Approach	Representative Works
Knowledge-based	
	Multiresolution rule-based method
Feature invariant	
– Facial Features	Grouping of edges
– Texture	Space Gray-Level Dependence matrix (SGLD) of face pattern
– Skin Color	Mixture of Gaussian
– Multiple Features	Integration of skin color, size and shape
Template matching	
– Predefined face templates	Shape template
– Deformable Templates	Active Shape Model (ASM)
Appearance-based method	
– Eigenface	Eigenvector decomposition and clustering
– Distribution-based	Gaussian distribution and multilayer perceptron
– Neural Network	Ensemble of neural networks and arbitration schemes
– Support Vector Machine (SVM)	SVM with polynomial kernel
– Naive Bayes Classifier	Joint statistics of local appearance and position
– Hidden Markov Model (HMM)	Higher order statistics with HMM
– Information-Theoretical Approach	Kullback relative information

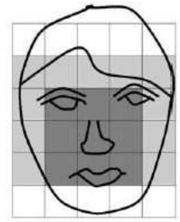
Knowledge-based Methods

- Human-specified rules:
  - A face appears in an image with two eyes that are symmetric to each other, a nose, and a mouth.
  - The relations between features can be represented by their relative distances and positions.
  - Facial features in an input image are extracted first, and face candidates are identified based on the coded rules.
  - A verification process is usually applied to reduce false detections.
- Difficulties of these methods:
  - The trade-off of details and extensibility
  - It is hard to enumerate all possible cases. In a restricted case – heuristics about faces work well in detecting frontal faces in uncluttered scenes.
- Example: (Yang and Huang 1994) Multiresolution Rule-based Methods
  - Three levels of rules:
    - All possible face candidates are found by scanning a window over the input image.
    - The rules at a higher level are general descriptions of what a face looks like
    - The rules at lower levels rely on details of facial features.

Multiresolution of face image

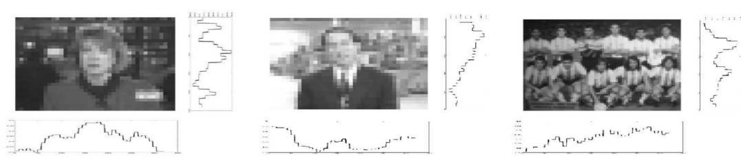
Knowledge-Based Method – Yang and Huang 1994

- Rules at the lowest resolution (Level 1):
  - The center part of the face has four cells with a basically uniform intensity
  - The upper round part of a face has a basically uniform intensity
  - The difference between the average gray values of the center part and the upper round part is significant
- The lowest resolution image is searched for face candidates and these are further processed at finer resolution.
- Rules at the Level 2:
  - Local histogram equalization is performed on the face candidates regions, followed by edge detection.
- Rules at the Level 3,
  - Detail rules of eyes and mouth
- Performance:
  - detect 50 faces out of 60 images.
  - 28 of the detected faces are false alarm
- Multiresolution approach inspired many follow-up researches.



Knowledge-based Method – Kotropoulos and Pitas 1997

- Use horizontal and vertical projections of the pixel intensity.
- The horizontal profile of an input image is obtained first, and then two local minima may correspond to the left and right side of the head.
- The vertical profile is obtained the the local minima are determined for the locations of mouth lips, nose tip, and eyes.
- This method has been tested ona video sequence database of 37 different people.
- Report a detection rate of 86.5%.
- Have difficulty to locate a face in a complex background.



Feature-based Methods

- Detect facial features such as eyebrows, eyes, nose, mouth and hair-line based on edge detectors.
- Based on the extracted features, a statistical model is built to describe their relationships and to verify the existence of face.
- Features other than facial featues:
  - Texture
  - Skin Color
  - Fusion of Multiple Features
- Difficulties:
  - Face features can be severely corrupted due to illumination, noise and occlusion.
  - Feature boundaries can be weakened for faces, while shadows can cause numerous strong edges which render perceptual grouping algorithm useless.

Feature-based Methods

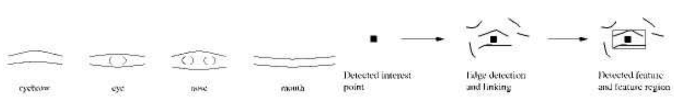
- Sirohey 1993:
  - Use an edge map (Canny detector) and heuristics to remove and group edges so that only the ones on the face contour are preserved.
  - An ellipse is then fit to the boundary between the head region and the background.
  - Detection accuracy: 80% out of a database of 48 images.
- Chetverikov and Lerch 1993:
  - Use blobs and streaks (linear sequences of similarly oriented edges).
  - Use two dark blobs and three light blobs to represent eyes, cheekbones and nose.
  - Use streaks to represent the outlines of the faces, eyebrows and lips.
  - Two triangular configurations are utilized to encode the spatial relationship among the blobs.
  - Procedure:
    - A low resolution Laplacian image is gnerated to facilitate blob detection.
    - The image is scanned to find specific triangular occurences as candidates
    - A face is detected if streaks are identified around a candidate.

Feature-based Methods

- Graf et. al. 1995:
  - Use bandpass filtering and morphological operations
- Leung et. al. 1995:
  - Use a probabilistic method based on local feature detectors and random graph matching
  - Formulate the face localization problem as a search problem in which the goal is to find the arrangement of certain facial features that is most likely to be a face patter.
  - Five features (two eyes, two nostrills, and nose/lip /junction).
  - For any pair of facial features of the same type, their relative distance is computed and modeled by Gaussian.
  - Report a correct rate of 86%.
  - Use statistical theory of shape (Kendall1984, Mardia and Dryden 1989), a joint probability density funciton over N feature points, for the ith feature under the assumption that the original feature points are positioned in the plane according to a general 2N-dim Gaussian.

Feature-based Methods

- Yow and Cipolla 1996:
  - The first stage applies a second derivative Gaussian filter, elongated at an aspect ratio of three to one, to a raw image.
  - Interest points, detected at the local maxima in the filter response, indicate the possible locations of facial features.
  - The second stage examines the edges around these interest points and groups them into regions.
  - Measurements of a region's characteristics, such as edge length, edge strength, and intensity variance are computed and stored in a feature vector.
  - Calculate the distance of candidate feature vectors to the training set.
  - This method can detect faces at different orientations and poses.
  - Report a test accuracy of 85%.



Feature-based Methods -- Texture

- Augusteijn and Skufca 1993:
  - Use second-order statistical features on submiages of 16x16 pixels.
  - Three types of features are considered: skin, hair, and others.
  - Used a cascade correlation neural network for supervised classifications.
- Dai and Nakano1996:
  - Use similar method + color
  - The orange-like parts are enhanced.
  - One advantage is that it can detect faces which are not upright or have features such as beards and glasses.
  - Report a detection rate of 100% for a test set of 30 images with 60 faces.

Feature-based Methods – Skin Color

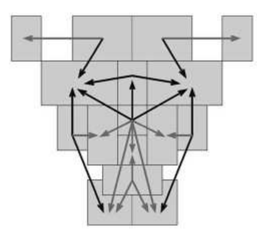
- Many methods have been proposed to build a skin color model.
- The simplest model is to define a region of skin tone pixels using Cr and Cb values by carefully chosen thresholds from the training set.
- Some more complicated models:
  - Histogram intersection
  - Gaussian density functions
  - Gaussian mixture models
- Color appearance is often unstable due to changes in both background and foreground lighting environments.
- If the environment is fixed, then skin colors are effective.
- Several modular systems using a combination of shape analysis, color segmentation and motion information for locating or gtracking heads and faces.

Template-based Methods

- A standard face pattern (usually frontal) is manually predefined or parameterized by a function.
- Given an input image, the correlation values with the standard patterns are computed for the face contour, eyes, nose and mouth independently.
- The existence of a face is determined based on the correlation values.
- Advantage: simple to implement.
- Disadvantage: need to incorporate other methods to improve the performance

Template-based Methods – Predefined templates

- Sinha 1994:
  - Designing the invariant based on the relations of regions.
  - While variations in illumination change the individual brightness of different parts of faces remain large unchanged.
  - Determine the pairwise ratios of the brightness of a few such regions and record them as a template.
  - A face is located if an image satisfies all the pairwise brighter-darker constraints.



Appearance-Based Methods

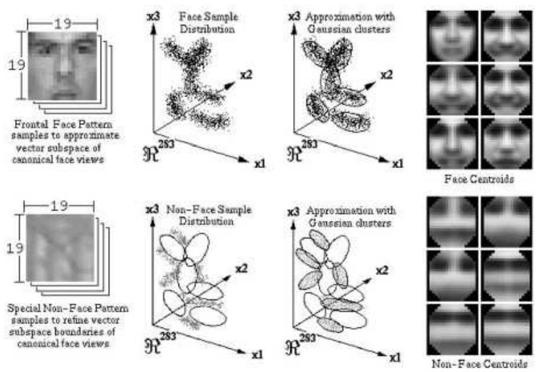
- Supervised learning
- Classification of face / non-face
- Methods:
  - Eigenfaces
  - Distribution-based Methods
  - Neural Networks
  - Support Vector Machines
  - Sparse Network
  - Naïve Bayes Classifier
  - Hidden Markov Model

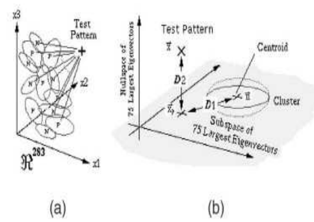
Appearance-Based Method -- Eigenfaces

- Apply eigenvectors in face recognition (Kohonen 1989).
  - Use the eigenvectors of the image's autocorrelation matrix.
  - These eigenvectors were later known as Eigenfaces.
- Images of faces can be linearly encoded using a modest number of basis images.
- These can be found based on the K-L transform or Principal component analysis (PCA).
- Try to find out a set of optimal basis vector eigenpictures.
- Experiments showed that a set of 100 images with face image of 91x50 pixels can be effectively encoded using only 50 eigenpictures.

Appearance-based Methods – Distribution-based Methods

- Sung and Poggio 1996:
  - Each face and nonface example is normalized to a 19x19 pixel image and treated as a 361-dimensional vector or pattern.
  - The patterns are grouped into six face and six nonface clusters using a modified k-means algorithm.

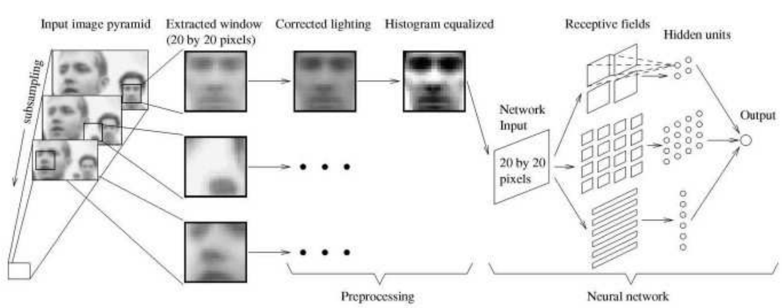




Distances are measured in the 75-dimensional subspace

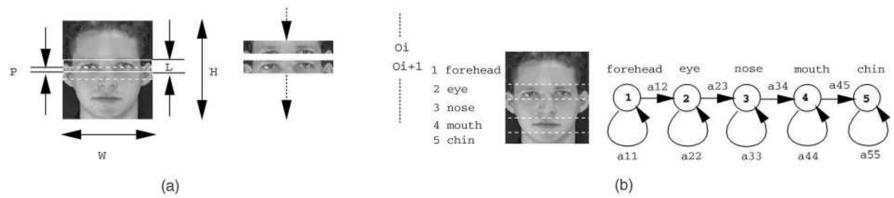
Appearance-based Methods – Neural Network

- Rowley 1996:
  - The first component is a neural network that receives a 20 x 20 pixel region and outputs a score ranging from -1 to 1.
  - Nearly 1050 face samples are used for training.



Appearance-based Method -- HMM

- The goal of training an HMM is to maximize the probability of observing the training data by adjusting the parameters in an HMM model.



Each face sample is converted to a sequence of observation vectors.

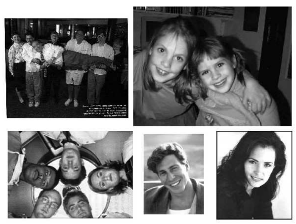
Face Image Database

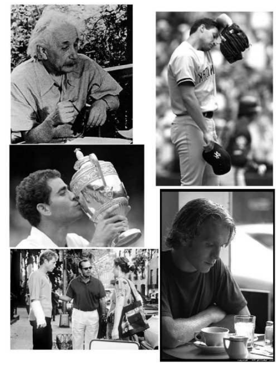
- These are commonly used face image databases

Data Set	Location	Description
MIT Database [163]	ftp://whitechapel.media.mit.edu/pub/images/	Faces of 16 people, 27 of each person under various illumination conditions, scale and head orientation.
FERET Database [115]	http://www.nist.gov/humanid/feret	A large collection of male and female faces. Each image contains a single person with certain expression.
UMIST Database [56]	http://images.ee.umist.ac.uk/danny/database.html	564 images of 20 subjects. Each subject covers a range of poses from profile to frontal views.
University of Bern Database	ftp://iamftp.unibe.ch/pub/Images/FaceImages/	300 frontal face images of 30 people (10 images per person) and 150 profile face images (5 images per person).
Yale Database [7]	http://cvc.yale.edu	Face images with expressions, glasses under different illumination conditions.
AT&T (Olivetti) Database [136]	http://www.uk.research.att.com	40 subjects, 10 images per subject.
Harvard Database [57]	ftp://ftp.hrl.harvard.edu/pub/faces/	Cropped, masked face images under a wide range of lighting conditions.
M2VTS Database [116]	http://poseidon.csd.auth.gr/M2VTS/index.html	A multimodal database containing various image sequences.
Purdue AR Database [96]	http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html	3,276 face images with different facial expressions and occlusions under different illuminations.

Benchmark for Face Detection

- Commonly used test sets





Data Set	Location	Description
MIT Test Set [154]	http://www.cs.cmu.edu/~har	Two sets of high and low resolution gray scale images with multiple faces in complex background.
CMU Test Set [128]	http://www.cs.cmu.edu/~har	130 gray scale images with a total of 507 frontal faces.
CMU Profile Face Test Set [141]	ftp://eyes.ius.cs.cmu.edu/usr20/ ftp/testing_face_images.tar.gz	208 gray scale images with faces in profile views.
Kodak Data Set [94]	Eastman Kodak Corporation	Faces of multiple size, pose and under varying illumination in color images. Designed for face detection and recognition.

An experimental results on the performance

- A test of several algorithms

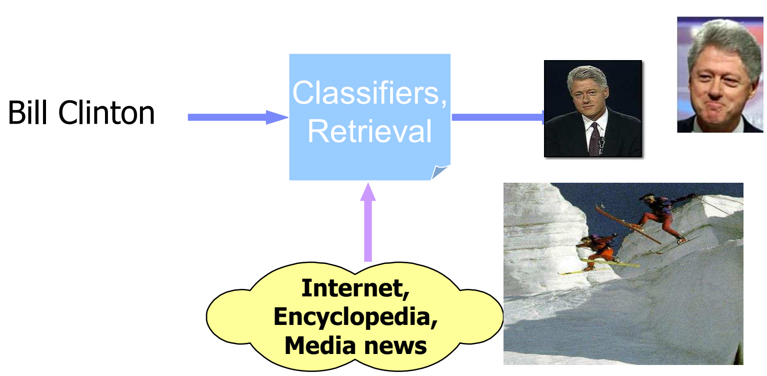
Experimental Results on Images from Test Set 1 (125 Images with 483 Faces) and Test Set 2 (23 Images with 136 Faces) (See Text for Details)

Method	Test Set 1 Detection Rate	Test Set 1 False Detections	Test Set 2 Detection Rate	Test Set 2 False Detections
Distribution based [154]	N/A	N/A	81.9%	13
Neural network [128]	92.5%	862	90.3%	42
Naive Bayes classifier [140]	93.0%	88	91.2%	12
Kullback relative information [24]	98.0%	12758	N/A	N/A
Support vector machine [107]	N/A	N/A	74.2%	20
Mixture of factor analyzers [175]	92.3%	82	89.4%	3
Fisher linear discriminant [175]	93.6%	74	91.5%	1
SNoW with primitive features [176]	94.2%	84	93.6%	3
SNoW with multi-scale features [176]	94.8%	78	94.1%	3
Inductive learning [38]	90%	N/A	N/A	N/A

Face Recognition System

- Use similar techniques as mentioned in the appearance-based face detection methods.
- Building personal models based on the training examples.
- Existing deployed systems:
  - SmartGate at Sydney's airport:
    - Successfully processed 62,000 transactions of 4,200 enrolled Qantas Air crew in a controlled environment.
    - Accuracy > 95%
  - Indetix and Viisage systems at Boston's Logan Airport 2002:
    - Fail to detect suspected criminals 38% of a time.
    - An earlier test at Palm Beach airport fails 53% of the time.

Learning Face Models using Autonomous Learning



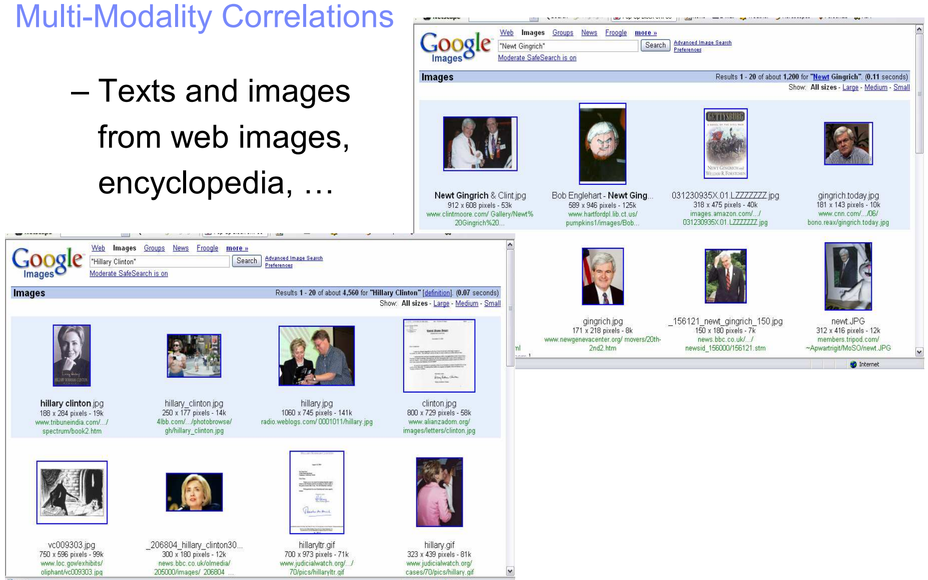
- Automatically build visual models associate with concepts (Bill Clinton, weather news, …)

Autonomous Learning

- Correlations between different modalities offers the possibility of autonomous learning
  - Audio, close captions, and visual data in video sequences

Multi-Modality Correlations

– Texts and images from web images, encyclopedia, …



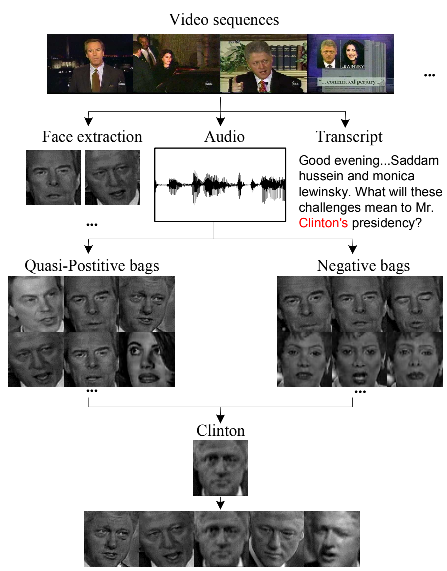
Difficulties

- Uncertainty in the automatic association from the multi-modality data
- Why we think it may still be feasible?

  Consistent cross-modality correlations of the correct association in different occasions

Learning Face Model from Cross-Modality Information

- Objective:
  - Generate Visual Concept Models without Annotation
    - Cross-Modality Training using Automatic Speech Recognition Transcription
    - Continuously Learning Concepts from Broadcasting TVs
    - Large Scale of Visual Concept Learning from Sub-Optimal Annotations such as Google Images
- Methodologies:
  - Multiple Instance Learning with Quasi-Positive Bags
  - Support Vector Machine Regression



Example Training Face Models

Multiple-Instance Learning

- The trainer only labels collections of examples (bags).
- A bag is labeled
  - "negative" if all the examples in it are negative.
  - "positive" if there is at least one positive example in it.

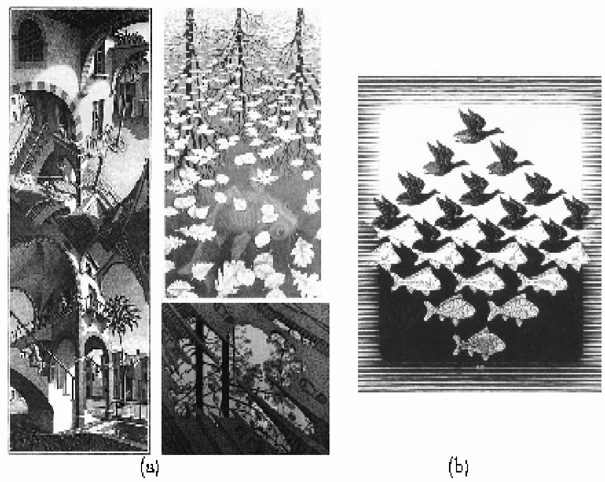


Figure: Some (a) positive and (b) negative training examples

Actual positive instance: Trees [Maron 98]

Diverse Density (DD)

- A method to solve MIL problems
  - Looking for an instance which is close to the instances in different positive bags and far from all the instances in the negative bags

$$\arg\max_t \prod_i \Pr\left(t \mid B_i^+\right) \prod_i \Pr\left(t \mid B_i^-\right)$$

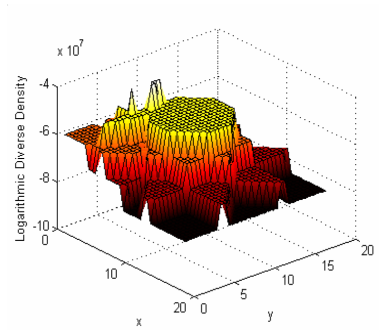
$$\Pr\left(t \mid B_i^+\right) = 1 - \prod_j \left(1 - \Pr\left(t \mid B_{ij}^+\right)\right)$$

$$\Pr\left(t \mid B_i^-\right) = \prod_j \left(1 - \Pr\left(t \mid B_{ij}^-\right)\right)$$

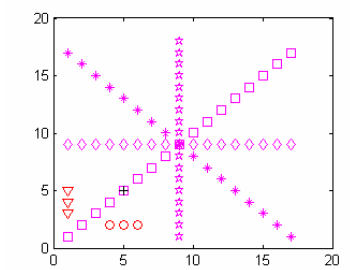
where
$$\Pr\left(t \mid B_{ij}^+\right) = \exp\left(-\left\|B_{ij}^+ - t\right\|^2\right)$$

$$\Pr\left(t \mid B_{ij}^-\right) = \exp\left(-\left\|B_{ij}^- - t\right\|^2\right)$$

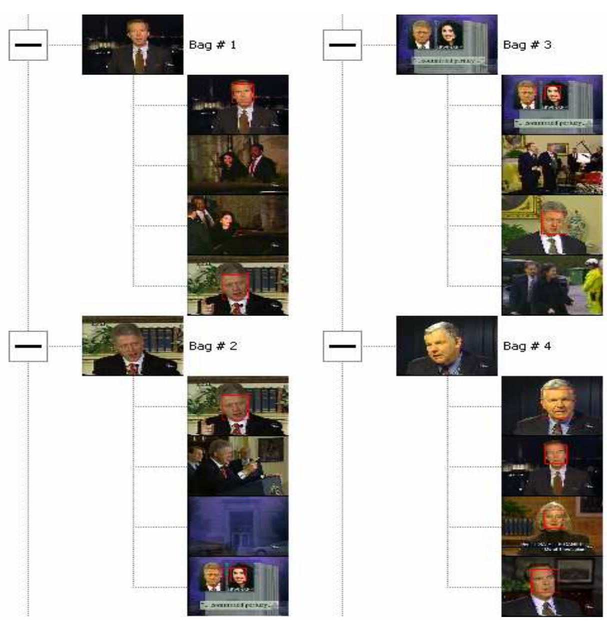
Problem of Diverse Density



- With a false-positive bag, $\Pr\left(t \mid B_i^+\right)$ will be very small or even zero.
- These outliers will influence the result significantly due to the multiplication of the probabilities.



Generalized Multiple Instance Learning



Autonomous Learning Scheme

Input: Named Concept

Database

Dictionaries, Internet

Imperfect Labeling

Bags of candidate images / regions

Cross-Modality Analysis

Uncertainty Pruning

GMIL +ULD

Instances with Relevance Scores

Relevance Modeling

SVR

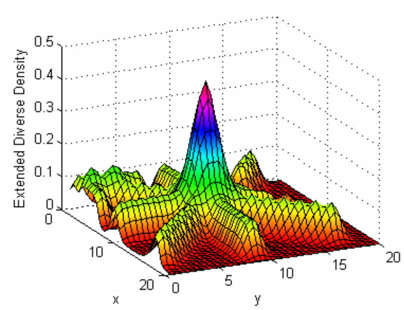
Output: Semantic Visual Models

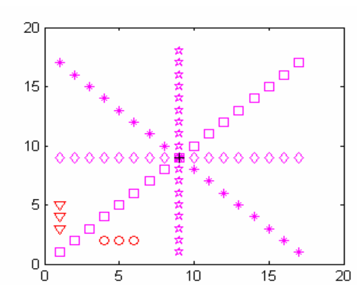
- Uncertainty Pruning:
  - Generalized Multiple Instance Learning (GMIL)
  - Uncertainty Labeling Density (ULD)
- Relevance Modeling:
  - Support Vector Regression (SVR)

Uncertain Labeling Density (ULD)

$$\arg\max \frac{\sum_i P\left(t \mid B_i^+\right) \cdot \prod_i \Pr\left(t \mid B_i^-\right)}{Z}$$

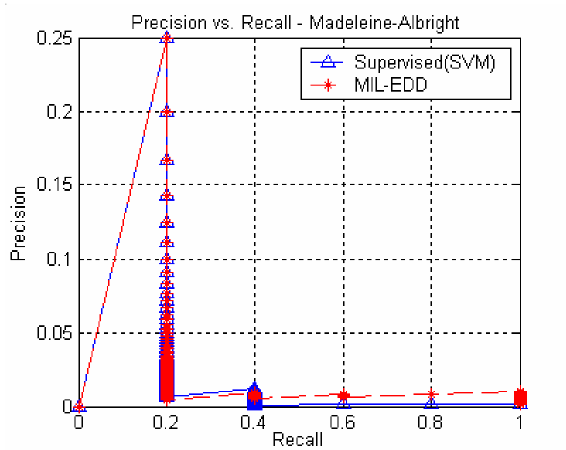
Z is a normalization constant to keep ULD in [0,1].



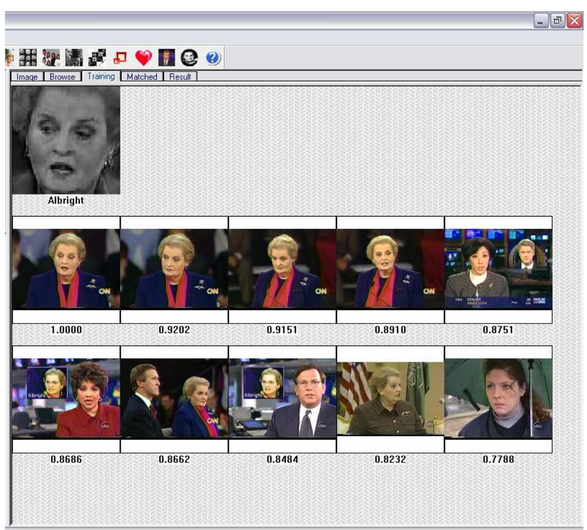


Face-Name Association Performance

Baseline: Supervised learning with Support Vector Machine (SVM)

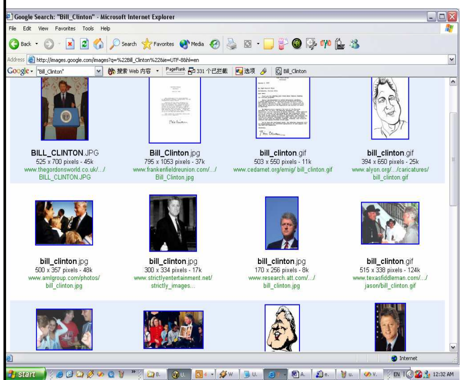


"Albright" Retrieval Rank List

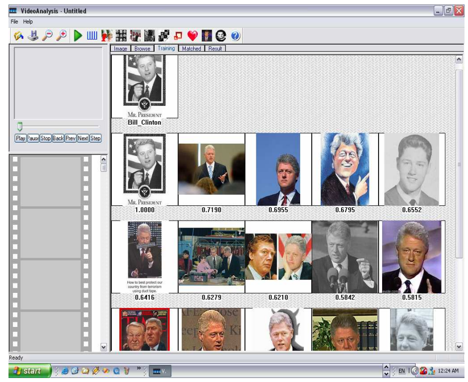


Another Application Example

- Improving Google image search

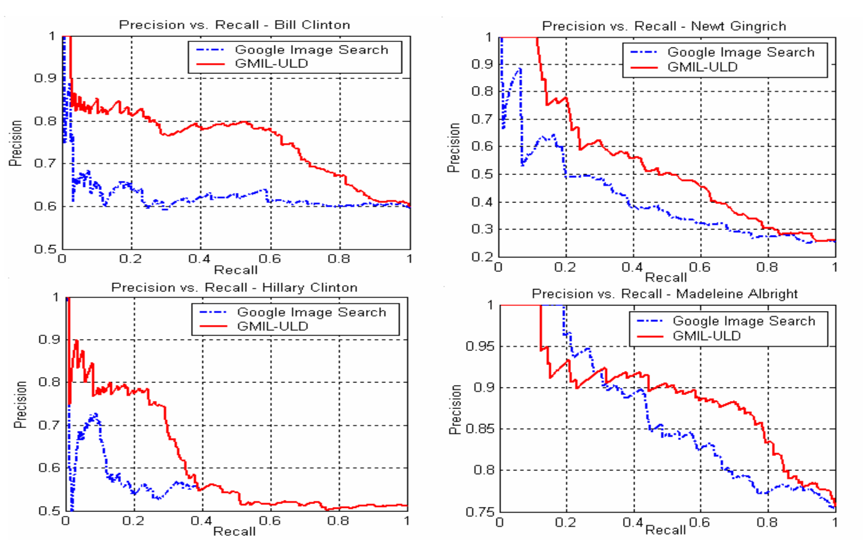


Google image Search
-- Bill Clinton



Improved Google search by cross-modality automatic learning Text & image

Preliminary Detection Results based on Web Annotations



References

- M.-H. Yang, D. Kriegman and N. Ahuja, “Detecting Faces in Images: A Survey”, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, Jan 2002.
- S.Y. Kung, Mak, and S. H. Lin, “Biometric Authentication”, Prentice-Hall, 2005. (Chapter 8)
- Xiaodan Song, Ching-Yung Lin and Ming-Ting Sun, "**Automatic Visual Concept Training Using Imperfect Cross-Modality Information**," Y.P. Tan, K.H. Yap and L. Wang, editors, Intelligent Multimedia Processing with Soft Computing, Springer, 2004.
- Xiaodan Song, Ching-Yung Lin and Ming-Ting Sun, "**Cross-modality automatic face model training from large video databases**," The First IEEE CVPR Workshop on Face Processing in Video (FPIV'04), June 2004.