ECE624

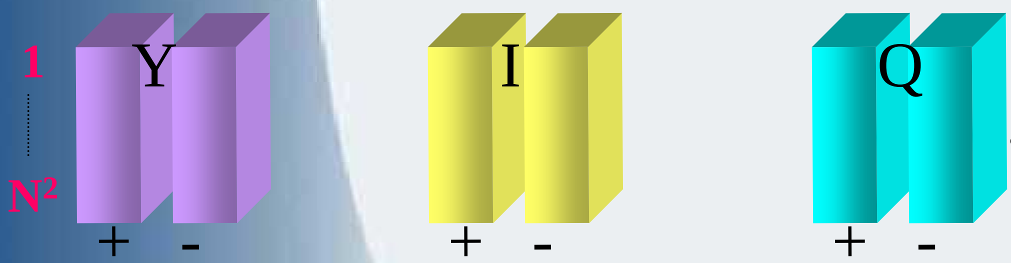
Week 3-a

Existing Technical Approaches

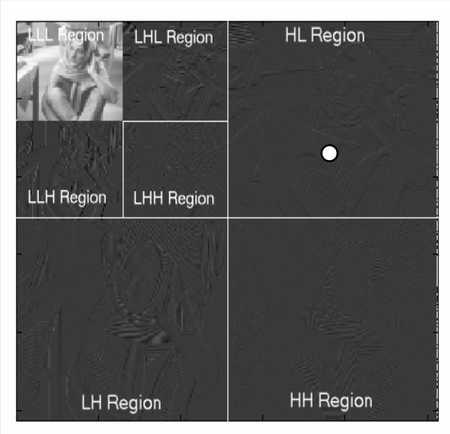
- Pixel-by-pixel comparison
- Compare complex/structured features
  - Color histogram
  - Frequency domain characteristics
    - Wavelet Transform

Example:
Fast Multiresolution Querying (U. Wash)

- Haar wavelets, color space (YIQ)
- Feature vector: 40-60 most significant coefficients from each color space
- Search space organized as 6 linear arrays



• coefficient(i,j) --> Ima

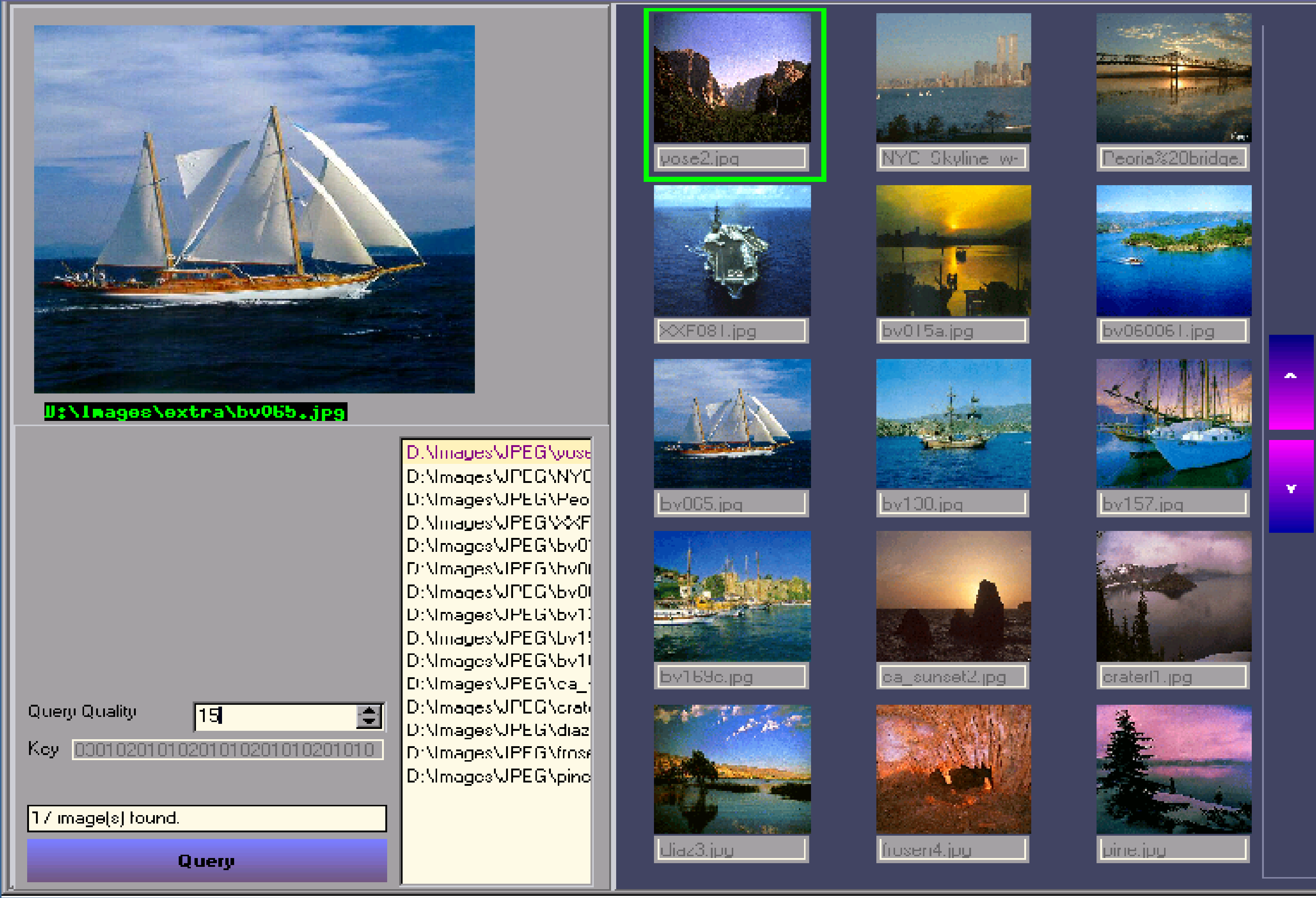


- Query time depends on the image contents (48 seconds on a SUN SPARC for a database of 1093 images)

Observations

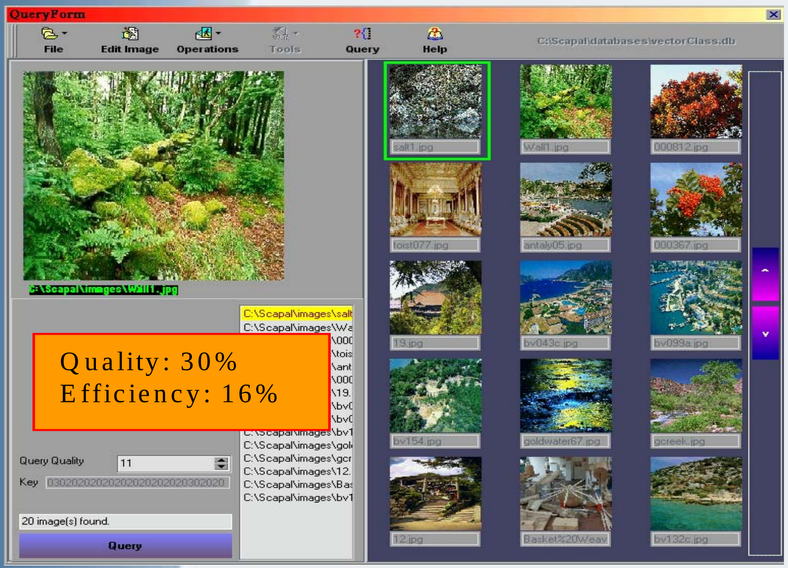
- Separate indexing of each color plane, increasing dimensionality of the search space
- Long image feature keys (typically > 256bytes)
- Query time linearly scales with the database size.
- Selection of the feature space and indexing structure as two separate problems.

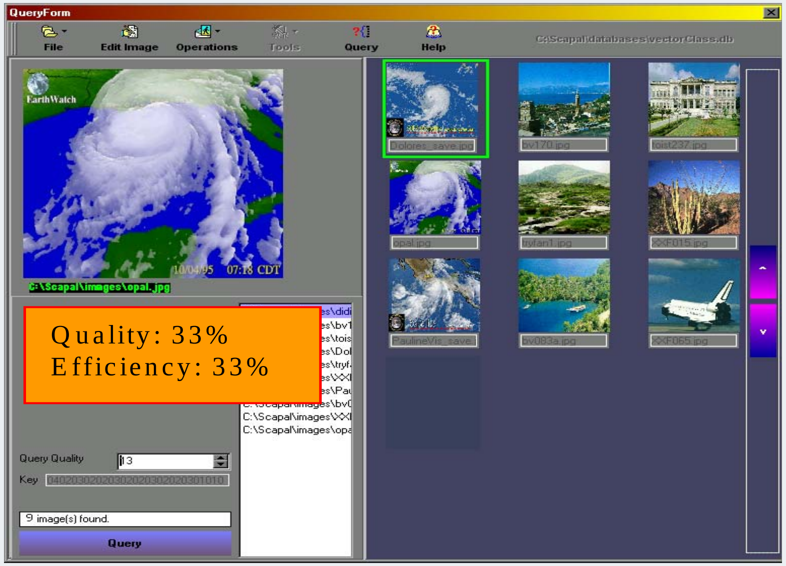
Wavelet Based Texture Query on 1610 Images

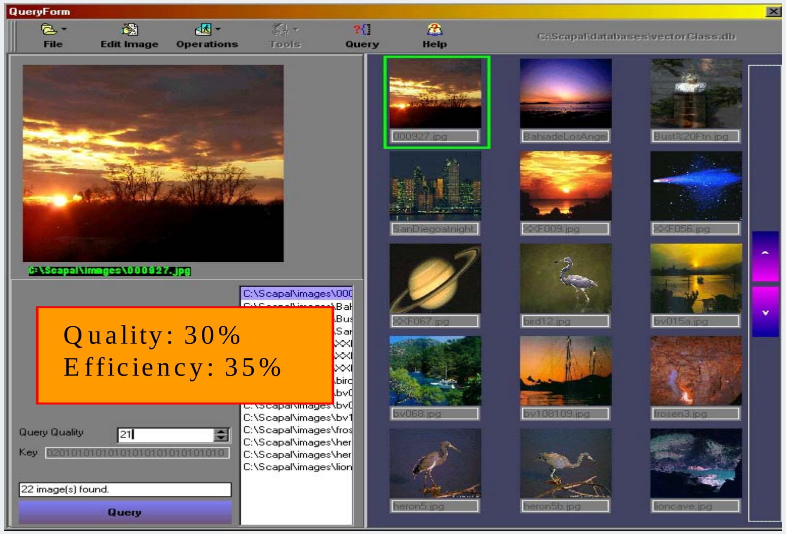


Color and Texture Results

Color and Texture Results







Query Response Time

System	Image Size	Database Size	Feature Vector Size	Search Time
LiangKuo	192x128	2119	212	NA
WBIIS Stanford University	128x128	10,000	≥ 768 bytes	3.3secs
IBM QBIC	100x100	1000	NA	2-40secs
University of Washington	128x128	1093	O(m) bytes	47.46secs

Object Recognition Layer

- Features are analyzed to recognize objects and faces in an image database.

- Features are matched with object models stored in a knowledge base.
  - Each template is inspected to find the closest match.
  - Exact matches are usually impossible and generally computationally extensive.
  - Occlusion of objects and the existence of spurious features in the image can further diminish the success of matching strategies.

Template Matching Techniques

- Fixed Template matching
  - Useful if object shapes do not change with respect to the viewing angle of the camera.
  - Image subtraction and correlation have been used in fixed template matching.
    - Image Subtraction: Difference in the intensity levels between the image and the template is used in object recognition. Performs well in restricted environments where imaging conditions (such as image intensity) between the image and the template are the same (X-Ray Images).
    - Matching by correlation utilizes the position of the normalized cross- correlation peak between a template and image. These methods are generally immune to noise and illumination effects in the image, but suffer from high computational complexity caused by summations over the entire template.

Template Matching Techniques

- Deformable Template Matching
  - More suitable for cases where the objects in the database may vary due to rigid and non-rigid deformations.
  - Template is represented as a bitmap describing the characteristic contour/edges of an object shape.
  - An objective function with transformation parameters which alter the shape of the template is formulated reflecting the cost of such transformation.
  - An objective function is minimized by iteratively updating the transformations parameters to best match the object.
  - Applications include: handwritten character recognition and motion detection of objects in video frames.

Spatial Modeling and Knowledge Representation layer

- Maintain the domain knowledge for representing spatial semantics associated with image databases.
- At this level, queries are generally descriptive in nature, focus mostly on semantics and concepts present in image databases.
- Semantics at this level are based on “spatial events” describing the relative locations of multiple objects.
- An example involving such semantics is a range query which involves spatial concepts such as *close by, in the vicinity, larger than* etc.
  - For example: “Retrieve all images that contain a large tumor in the brain”

Spatial Modeling and Knowledge Representation layer

- The general approach for modeling spatial semantics for such applications is based on identifying spatial relationships among objects once they are recognized and marked by the lower layer using bounding boxes or volumes.
- Several techniques have been proposed to formally represent spatial knowledge at this layer.
  - Semantic networks
  - Mathematical logics
  - Constraints
  - Inclusion hierarchies
  - Frames

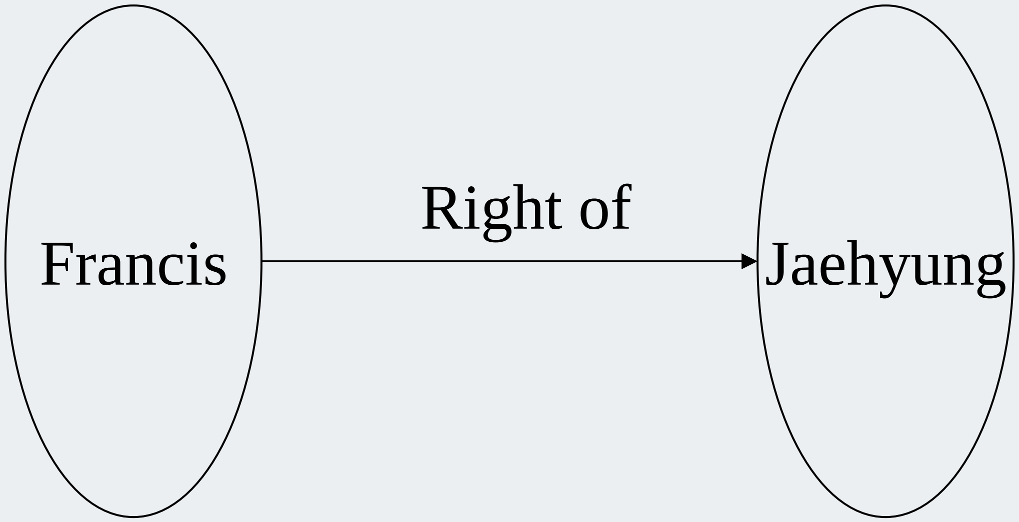
Constraint - based Methodology

- Domain knowledge is represented using a set of constraints in conjunction with formal expressions such as predicate calculus or graphs.
- A constraint is a relationship between two or more objects that needs to be satisfied.
  - Example: PICTION system
    - Its architecture consists of a natural language processing module (NLP), an image understanding module (IU), and a control module.
    - A set of constraints is derived by the NLP module from the picture captions. These constraints (Called Visual Semantics by the author) are used with the faces recognized in the picture by the IU module too identify the spatial relationships among people.
    - The control module maintains the constraints generated by the NLP module and acts as a knowledge-base for the IU module to perform face recognition functions.

In front of EE building, Francis day is standing to the right of Jaehyung


Francis
Jaehyung

NLP- Module Generated Constrain Graph



Inclusion Hierarchies

- The approach is object-oriented and uses concept classes and attributes to represent domain knowledge.

- These concepts may represent image feature, high-level semantics, semantic operators and conditions.

Frames

- A frame usually consists of a name and a list of attribute-value pairs. A frame can be associated with a class of objects or with a class of concepts.

- Frame abstractions allow encapsulation of file names, features, and relevant attributes of image objects.

Facial Data

Face Recognition in Image/Video Databases

Three Steps:

1. Face Detection: Locating face inside an image
2. Feature Extraction: Detection of various parts of a face
3. Face Recognition: Identifying a person

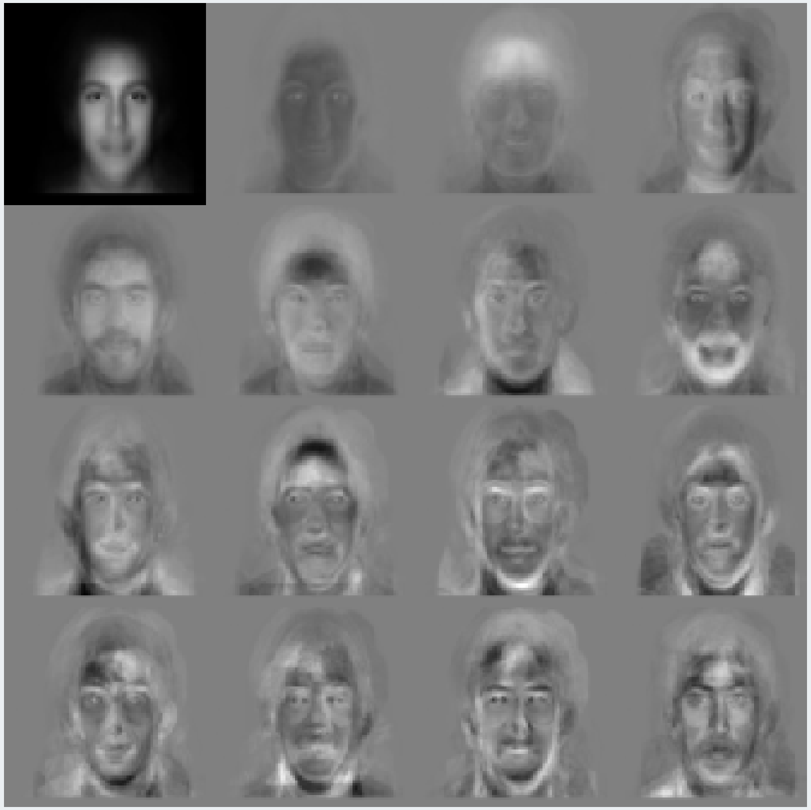
Face Detection Schemas

- Information Theoretic Approach
- Deformable Templates Approach
- Mosaic Images Approach

Face Detection
Information Theoretic Approach

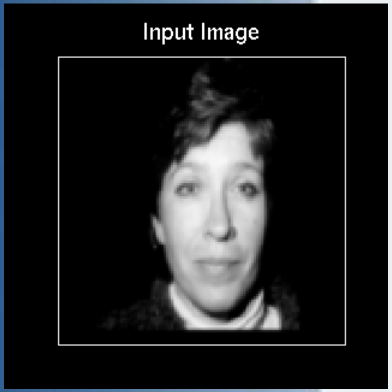
- Uses Principle Component Analysis
  1. Calculate the covariance matrix of some known images and find dominant eigen vectors (eigen faces) and store them
  2. To locate a face, scan the image with a local sub-image
  3. Calculate the projection of each sub-image on the eigen faces.
  4. Calculate the euclidean distance between the original image and its projection
  5. Small distance indicates the presence of a face.

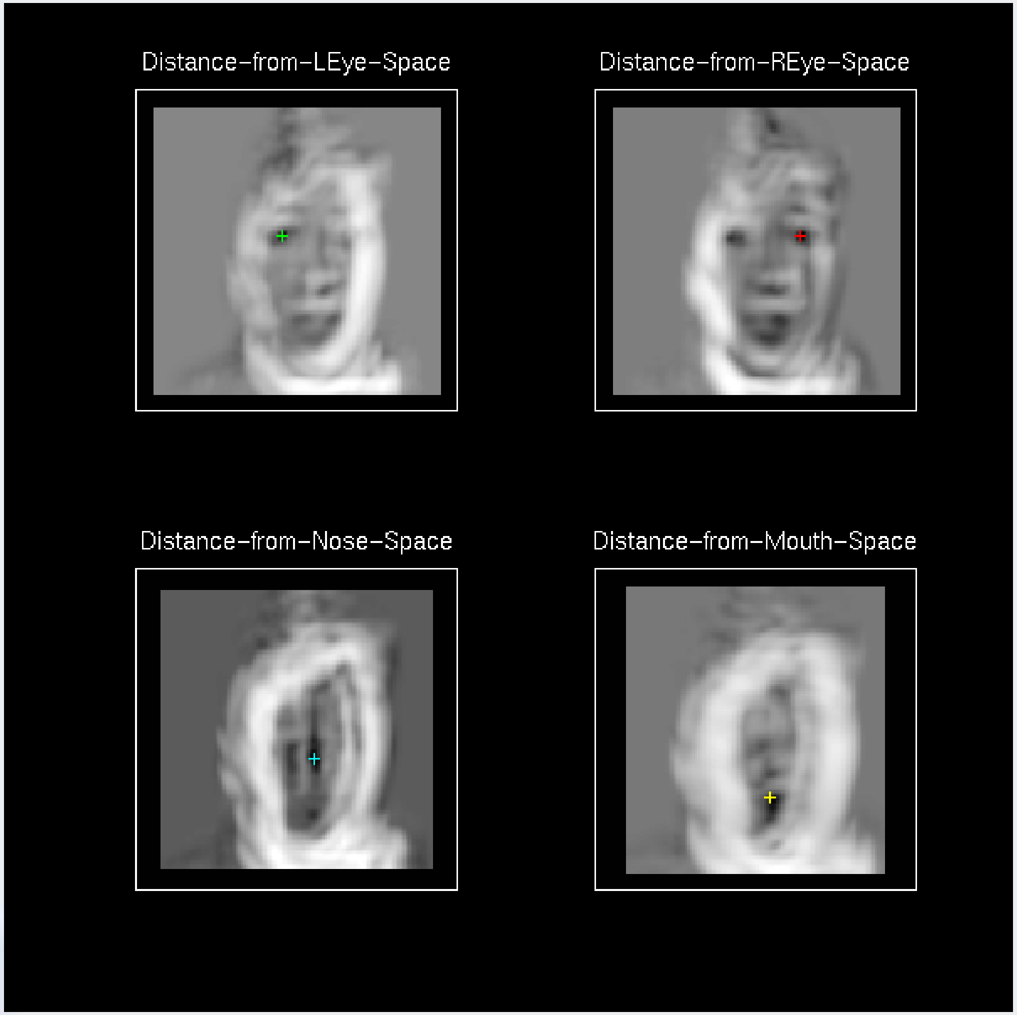
Standard Eigenfaces



Ref: http://vismod.media.mit.edu/vismod/demos/facerec/basic.html

Distance Features





Ref: http://vismod.media.mit.edu/vismod/demos/facerec/basic.html

Face Detection
Deformable Templates Approach

- Parameterized templates similar to various parts of face is used, namely: eyes, nose, lips, and mouth
- Rules can be applied to enhance the detection algorithm, for example

1. Distance between individual features should be proportional
2. Face must contain two eyes, below them exist a single nose and below it a single mouth

Face Detection
Mosaic Approach

Rule Based technique

- Construct from the given image, a set of images with different resolutions
- Rule
  - if, at a certain resolution, in a region composed of 4x4 cells with relatively light grey level and above it, another region with relatively dark then this is a possible face region
- Other rules can be applied to filter out irrelevant regions