Audio-Visual Speech Processing

COM 4110 / COM 6070

Jon Barker

.

j.barker@dcs.shef.ac.uk

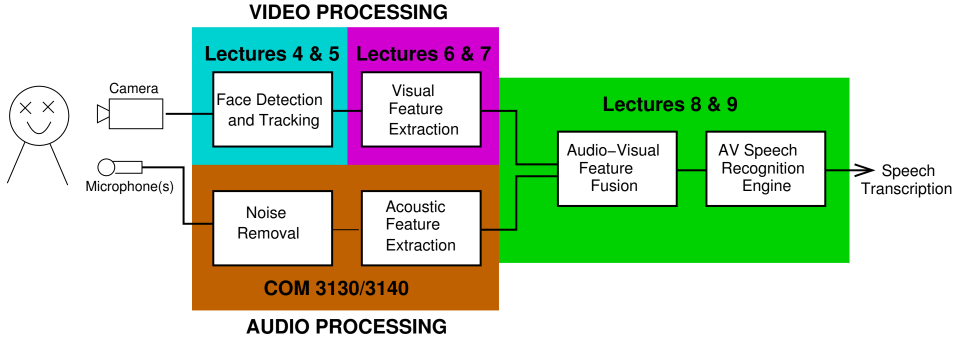
http://www.dcs.shef.ac.uk/~jon

Department of Computer Science

University of Sheffield

Audio-Visual Automatic Speech Recognition (AV-ASR)

The figure shows the basic design of **a *typical* AV-ASR system**.



This lecture and the one that follows will address the first video processing stage - **face detection**.

We will be comparing two systems, that of Carnegie Mellon University (**CMU**) and that of **IBM**. Techniques will be introduced that will be useful again at later points in the processing pipeline.

Lecture 4: Face Detection (Part 1)

Objectives

- To examine methods for finding faces in arbitrary scenes.

Topics

- The Face Detection Problem.
- Neural Network-based systems.
- Chromaticity-based Face Detection.

Reading

- *Neural network-based face detection*, Rowley, Baluja and Kanade, 1998
- *Face and feature finding for a face recognition system*, A.W.Senior, 1999

Lecture 4: Face Detection (Part 1)

Overview

- Face Detection - The Problem.
- Sources of Facial Variability.
- A General Approach to Face Detection.
- The CMU Neural Network-Based Face Detector.
- The IBM Face Detection System:
    - Chromaticity-Based Classification,
    - Fisherfaces - Line Discriminant Analysis, (Lecture 5)
    - Eigenfaces - Principal Component Analysis. (Lecture 5)

Face Detection

What is face detection?

- Face detection systems try to answer the following questions:
  - Given a visual scene, is there a face present?
  - If so, where is the face?

Why is this important?

- Basically, we can't do speechreading unless we can first find the speakers' faces!

Why is face detection a challenging task?

...

Facial Variability

There is large variation in facial appearance:

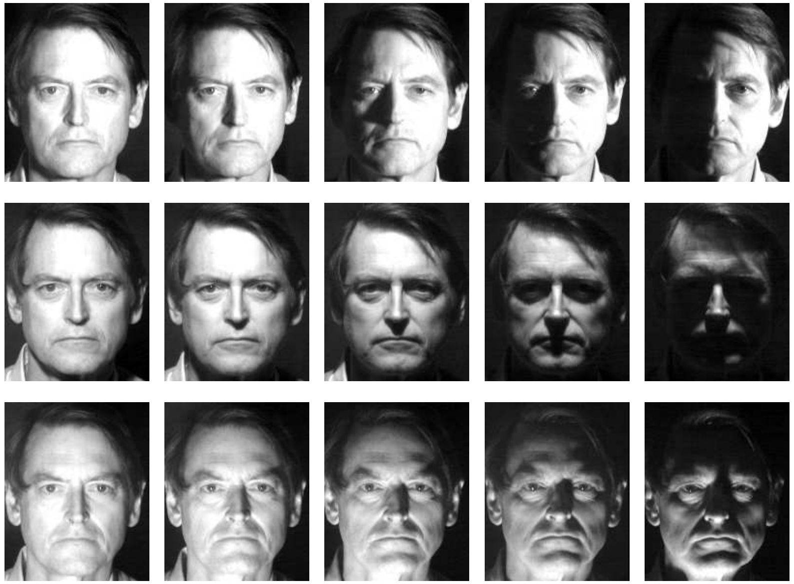
Images from the AT&T Laboratories, Database of Faces, *Samaria and Harter (1995)*.

Sources of Inter-Face Variability

Some (not all!) sources of variability between faces:

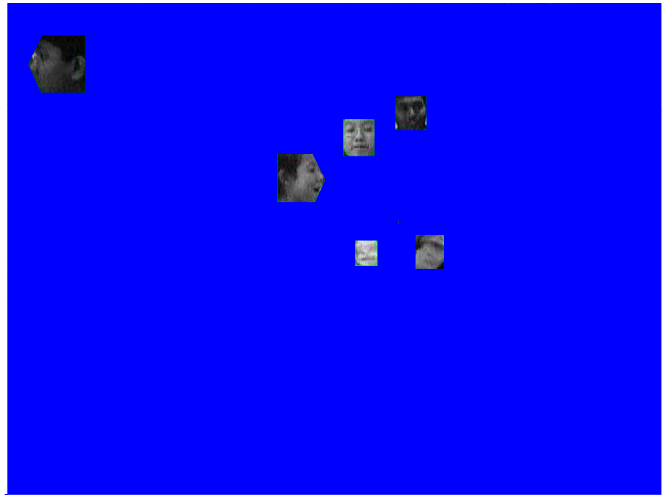
- Differences in the face shape
- Differences between facial features - noses, eyes, lips etc
- Hair colour, length, style,
- Skin tone, (with or without make-up!)
- Beards, moustaches, glasses, hats,
- Distance from camera,
- Angle of head,
- Lighting,
- Visual obstructions

Variability Due to Lighting



Face Detection Demonstration

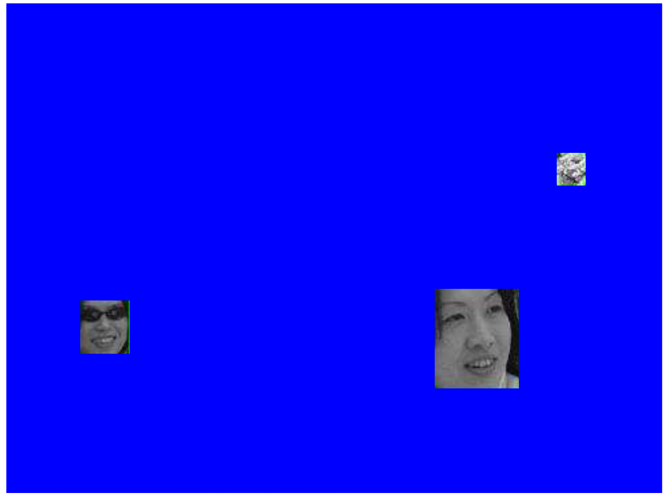
Output of the CMU neural network-based face detector.



Rowley, Baluja and Kanade (1996)

Face Detection Demonstration

Output of the CMU neural network-based face detector.



Rowley, Baluja and Kanade (1996)

A General Approach To Face Detection

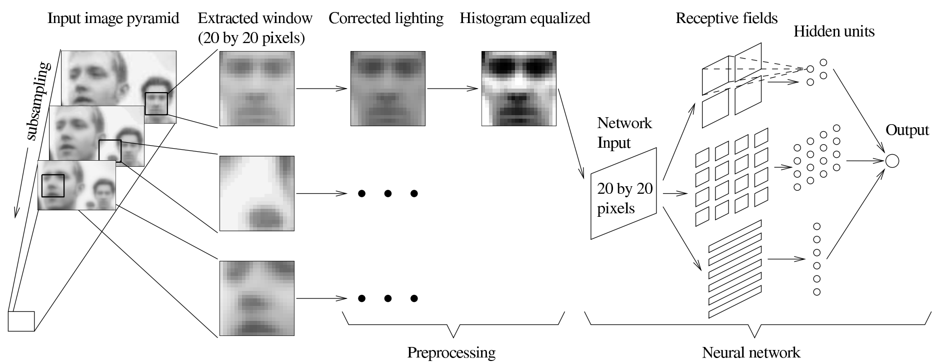
Outline of the general approach:

1. Consider every face-proportioned sub-image in the full image, i.e. consider rectangles at every possible position and scaling.

2. To each sub-image apply two class classification: ‘Face’ or ‘Not Face.’

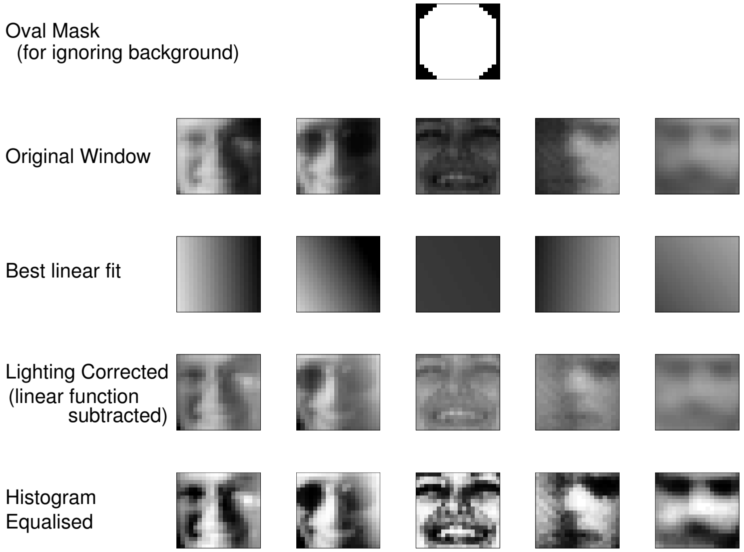
Face Detection Using Neural Networks

Overview of the CMU neural network-based face detection system.



CMU System - Rowley, Baluja and Kanade, PAMI, 1998

Preprocessing Steps for Neural Network Face Detection



Rowley, Baluja and Kanade: Neural Network-Based Face Detection (PAMI, 1998)

Generating Training Data: Examples of Faces

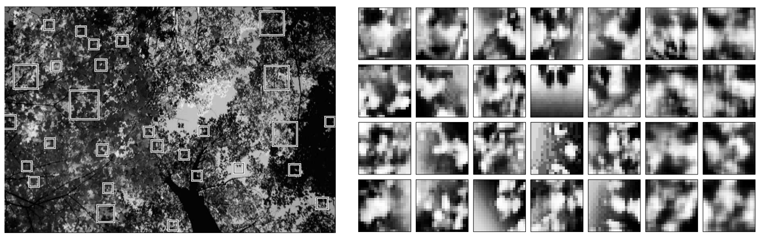
Positive training examples (i.e. faces) are generated as follows:

- 1050 face examples gathered from face databases.
- faces normalised to a fixed 20x20 pixel size and upright orientation.
- 15 copies of each face generated with combinations of the following random transformations:
  - rotations about centre of up to 10 degrees,
  - scalings between 90% and 110%,
  - horizontal or vertical translations of $+/-1/2$ a pixel,
  - horizontal mirroring.

Training the Neural-Network

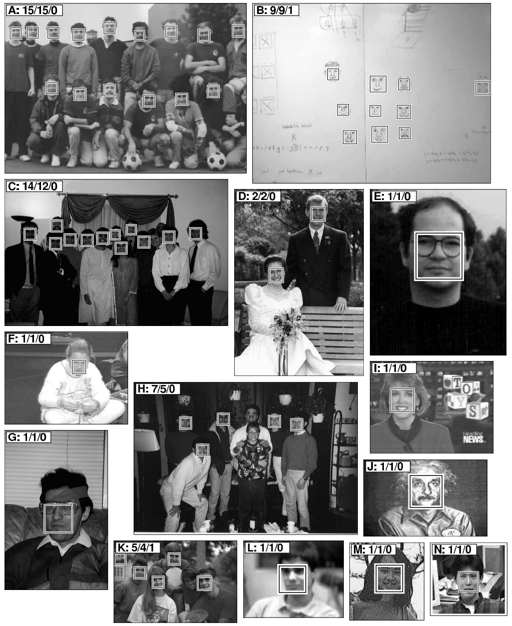
The network is trained using the following scheme:

1. Create an initial set of 1000 random non-face images.
2. Train network to output 1 for face examples, and -1 for nonfaces.
3. Run system on scenery *containing no faces*. Collect sub-images incorrectly identified as faces.
4. Select 250 of these sub-images and add them to the non-face training set. **Go to step 2**.

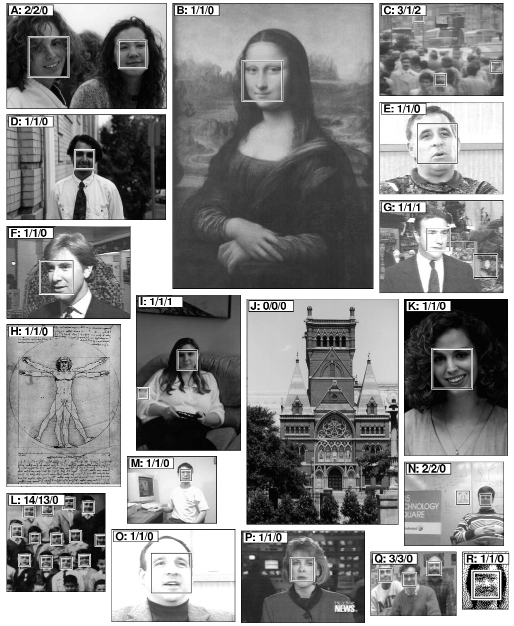


Examples Taken From *Rowley, Baluja and Kanade (1998)*

Examples Taken From *Rowley, Baluja and Kanade (1998)*

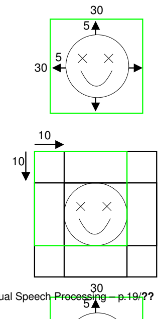


Examples Taken From *Rowley, Baluja and Kanade (1998)*



Computational Cost of the CMU System

- The **basic CMU system** is fairly slow:
  - A 320x240 pixel image contain 246766 sub-images!
  - Processing a single image on a 200MHz R4400 SGI Indigo 2 takes 383 seconds (>6 minutes). (Maybe about 10 time faster on a modern desktop).
- A **'fast' version** of the system can be made by training on faces that are off-centre by up to 5 pixels in any direction.



Face Detection for Video Sequences

When applying face detection of **video data** there are extra constraints that can be exploited to help:

- **Continuity**: As a face moves about its location in one frame is a strong predictor of its location in the next frame (**Face Tracking**). i.e. Tracking a face that has been previously detected is an easier problem than detecting a face in the first place.
- **Sound location**: If we have stereo audio (or better, the output of a **microphone array**), we can estimate the face position using the apparent direction of the speaker's voice.

These sources of information can be used to **focus the search on a specific region**, i.e. only a portion of the whole image has to be processed.

Computationally Efficient Face Detection

- Even with a focused search the problem remains that there may be a large number of possible sub-images that need to be classified.
- For a **realtime** system face detection must be performed on at least 25 frames a second. i.e **40 ms processing time per frame**!
- The classification step must be **computationally cheap**... but to avoid errors it must also be **reliable**.

How can we achieve this?

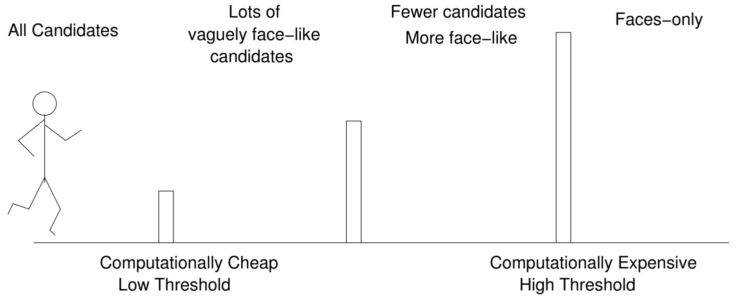
Lecture 4: Face Detection (Part 1)

Overview

- Face Detection - The Problem.
- Sources of Facial Variability.
- A General Approach to Face Detection.
- The CMU Neural Network-Based Face Detector.
- **The IBM Face Detection System:**
  - Chromaticity-Based Classification,
  - Fisherfaces - Line Discriminant Analysis, (Lecture 5)
  - Eigenfaces - Principal Component Analysis. (Lecture 5)

Computationally Efficient Classification

How can we make the face search both **computationally cheap** and **reliable**?



For each classification arrange a series of 'hurdles':

- Most sub-images will look very different from faces and can be rejected cheaply.
- Expensive classification techniques are only applied to a small number of good candidates.

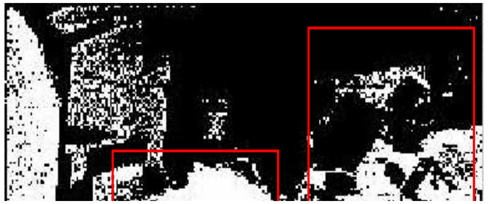
Classification Techniques Employed in the IBM System *Senior*

The IBM system uses a combination of **three** separate classification techniques:

- **Skin-tone-Based Classification.**
  - Cheap but unreliable.
- **The Fisher Linear Discriminant.** (*next lecture*).
  - More expensive, but more reliable.
- **Distance From Face Space** (*next lecture*).
  - Computationally expensive.

Skin-Tone-Based Classification

1. **Mark** pixels whose colour is similar to **skin-tone**.



Summary

- Face detection is challenging due to the large degree of variability in the appearance of faces.
- The general approach is to break the scene into a large number of sub-images and classify each sub-image as face or non-face.
- Neural-networks can perform very well but are computationally expensive.
- Face tracking can be employed to focus the search space.
- Efficient systems can be built by cascading a variety of increasingly sophisticated classification techniques.
- Skin-tone-based classification can be used to cheaply reject sub-images that do not have a high proportion of skin coloured pixels.

Next Lecture: Face Detection (Part 2)

Objectives

- To examine more sophisticated face detection techniques.

Topics

- Linear Discriminant Analysis - Fisherfaces
- Principal Component Analysis - Eigenfaces

Reading

- *Eigenfaces for recognition*, Turk and Pentland, 1991
- *Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection*, Belhumeur, Hespanha and Kriegman, 1997

References

- Duda and Hart (1973) *Pattern classification and scene analysis*, John Wiley and Sons, New York.
- Rowley, Baluja and Kanade, (1996) Neural network-based face detection, In *IEEE Transactions on Pattern Analysis and Machine Intelligence,* San Francisco, CA, pp. 203-207.
- Samaria and Harter (1994) Parameterisation of a Stochastic Model for Human Face Identification, In *Proceedings of 2nd IEEE Workshop on Applications of Computer Vision*, Sarasota FL.
- A.W.Senior (1999) Face and feature finding for a face recognition system, In *Proc. Second International Conference on Audio- and Video-based Biometric Person Authentication*, 154–159, Washington, 1999.
- CMU Face detection demo: http://www.vasc.ri.cmu.edu/cgi-bin/demos/findface.cgi