Face Detection in Video

Students: Eran Hertzberg & Tal Kenig

Supervisor: Asaf Cohen

Escorting Company: RadVision

Presentation Date: 6/22/2005

Spring 2004, Signal & Image Processing Laboratory

What is Face Detection ?

- Face detection is a domain of image processing, aimed at finding human faces in visual data
- Possible commercial applications for a face detector:
  - Multi Participant Video Conference
  - Cellular Video Conference
  - Variable bit-rate image compression
  - Face recognition system

Project Definition

- Automatic face detection in video stream
- Requirements:
  - Real time implementation
  - Robustness to: Glasses, Mustache, Beard, Hat
  - Race invariant
  - Lighting invariant
  - Color saturation invariant

Project Definition: Simplifying Assumptions

- Only upfront faces

- Single face in an Image

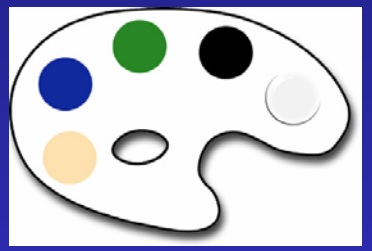
- Whole face in picture

- Decompressed video

Available Solutions

- Linear subspace methods

- Color based analysis



- Motion based analysis & tracking

Rejected Solution: Linear subspace methods

- The method: Determining the likelihood of a certain region in the frame being a human face, based on shape and patterns.

- Why rejected ?

  - Computationally expensive (preprocessing and algorithms)
  - Does not utilize the color and motion data included in color video scenes

- Not suitable for real-time applications
- Ideal for grayscale still images

Chosen Solution:
Color based analysis

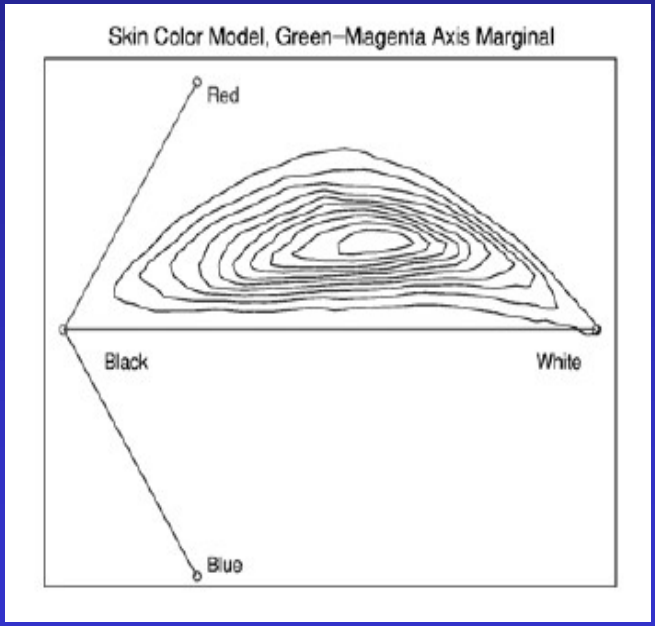
- The method: Determining the likelihood of a certain pixel belonging to human skin.

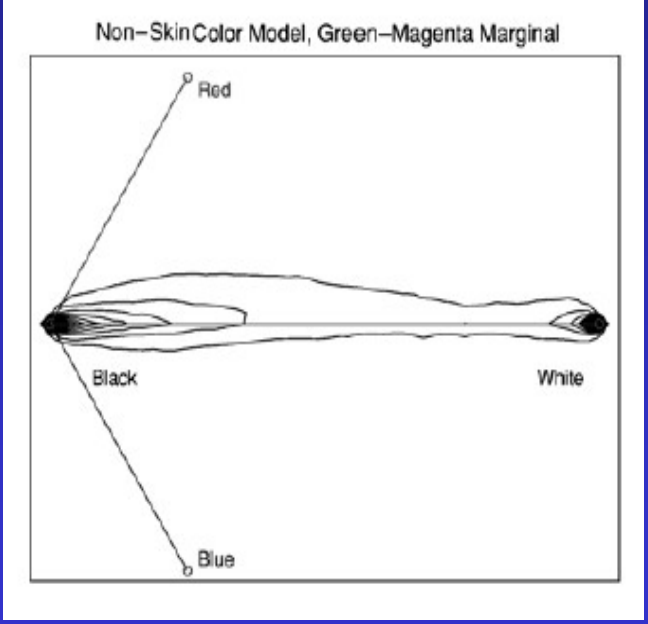
- Chosen technique:
  - Statistical human skin color model by Jones & Rehg

Chosen Solution: Statistical human skin color model by Jones & Rehg

- Histogram based color model:

256^3 bins RGB color space

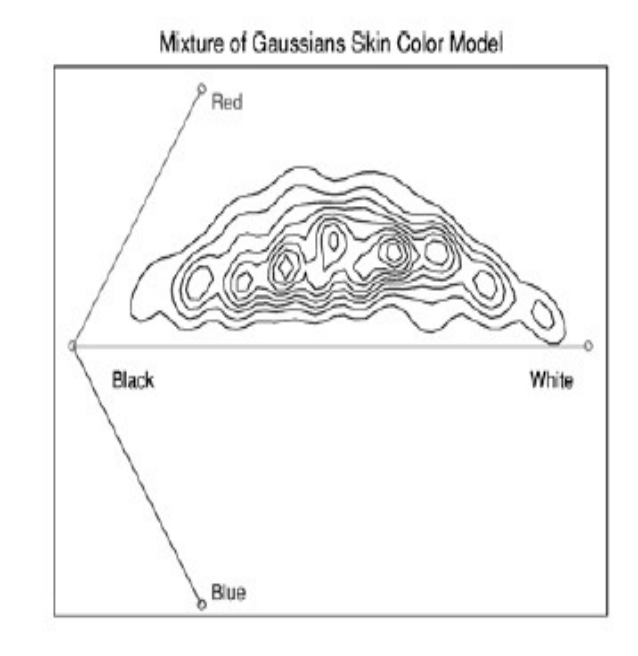


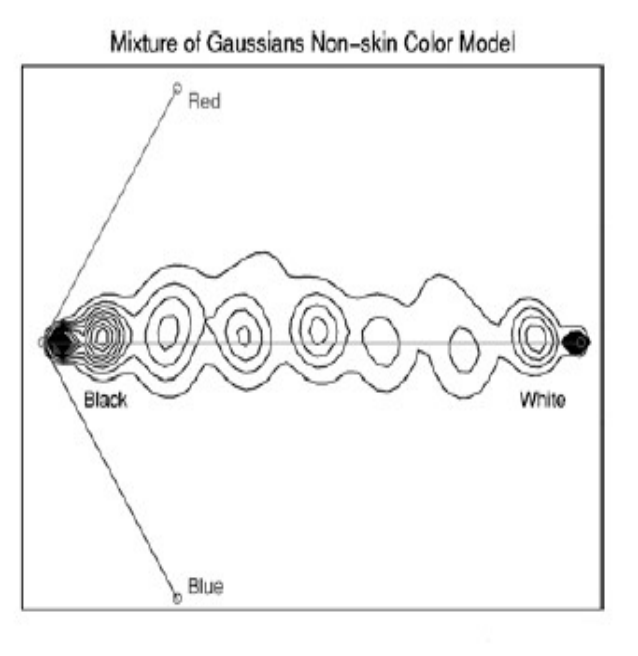


Chosen Solution: Statistical human skin color model by Jones & Rehg

- Gaussian mixture model:

Originally 256^3 bins RGB color space, we utilize a 64^3 bins RGB reduced color space





Chosen Solution: Motion based analysis & tracking

- The method: Human faces are more likely to exhibit high motion than the background

original movie

motion vectors norm

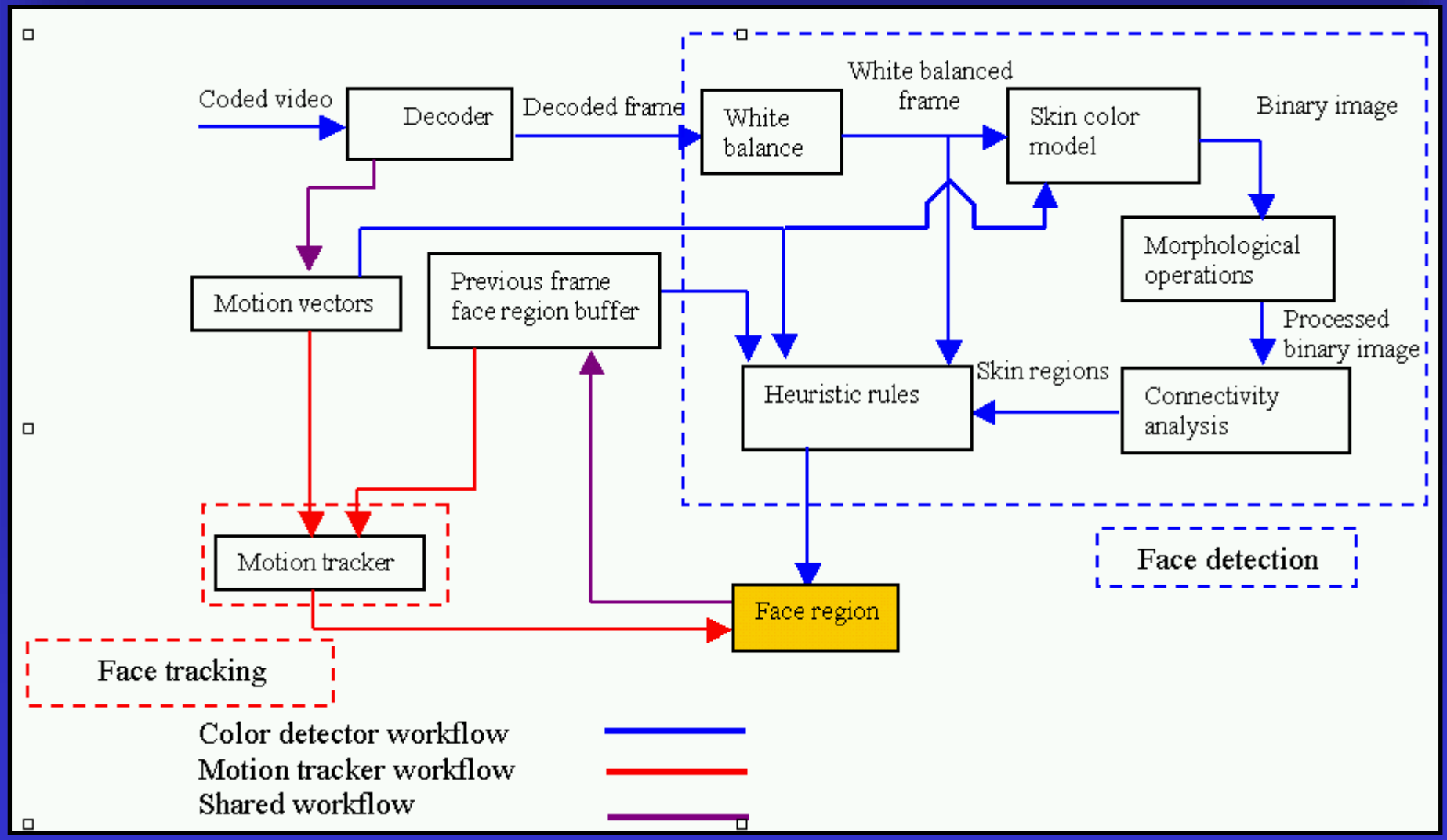
- The use of decompressed video stream allows us access to motion vectors without computational cost
- MVs are used by the proposed system in two different ways:
  - Checking if a suspected skin pixel is in motion
  - Tracking an identified face region

Chosen Solutions: A demonstration of the proposed system

We proudly present:

Archadi

The proposed system: general description



The proposed system: a detailed description

Preprocessing: white balance unit

- We used an algorithm called "Normalize to White“.

- White balance coefficients are calculated only once, based on the first frame.

After

Before

The proposed system: a detailed description

Skin color model: our implementation

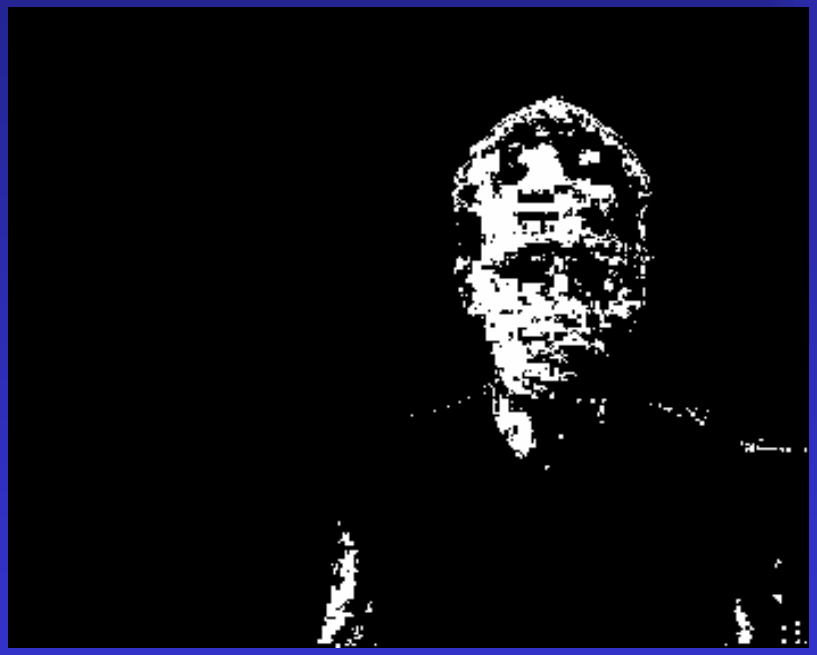
- The classifier function is set to:
  $C(R,G,B)=P_{skin}(R,G,B)/P_{non\text{-}skin}(R,G,B)$
- We use an adaptive threshold value based on the mean value of C(R,G,B) calculated over the previous frame.

The proposed system: a detailed description

Skin color model: our implementation

- The output of the skin color unit is a binary image of the detected skin pixels.

original frame



skin pixels binary image

The proposed system: a detailed description

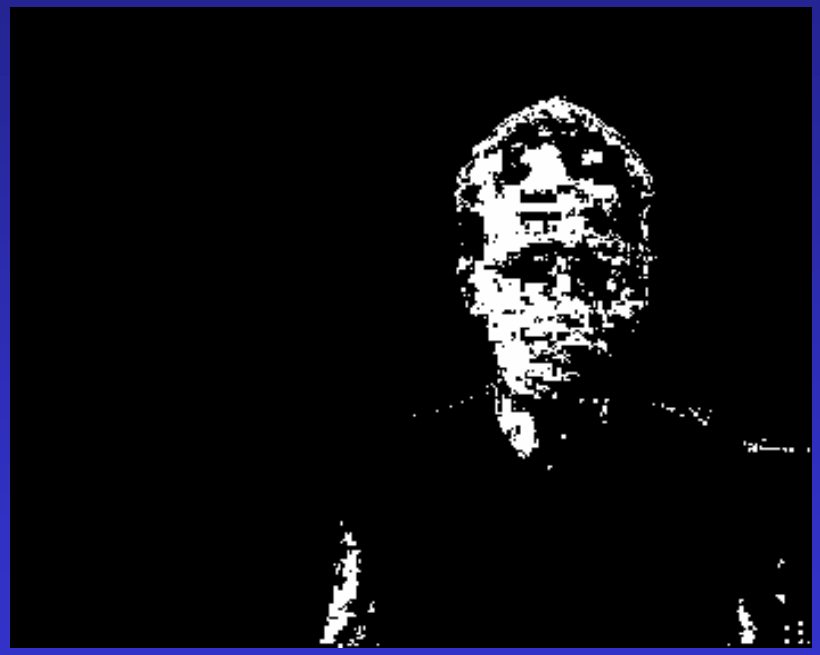
Connectivity analysis

- The purpose: extracting geometrically connected skin regions from the binary skin image
- The binary image is first morphologically processed
- A novel computationally inexpensive connectivity analysis algorithm is applied.

The proposed system: a detailed description

Connectivity analysis

- The output of the connectivity analysis unit is a list of rectangular connected skin regions



skin pixels binary image

rectangular skin regions

The proposed system: a detailed description

Heuristic rules

- The purpose: selecting the face region, given a list of rectangular skin regions
- The method: Applying heuristic rules
  - Height to Width ratio
  - Region size
  - Percentage of skin pixels in region
  - Distance from previous face region
  - Location of region

The output : detected face region

Face Detection Conclusions

- Real-time implementation is the most crucial consideration in the system design
- Varying lighting conditions and image capture devices force us into using adaptive, image derived, classifiers and parameters
- The usage of adaptive classifiers creates a problem in establishing a reliable confidence criteria
- There is a constant trade off between coherent tracking & recovery ability

And for the Grand Finale:

We proudly present:

Alex