Karlsruher Institut für Technologie

Computer Vision für Mensch-Maschine Schnittstellen

Vorlesung WS 2017/18

Prof. Dr.-Ing. Rainer Stiefelhagen, Dr. Saquib Sarfraz
Maschinensehen für MMI, Institut für Anthropomatik & Robotik
Studienzentrum für Sehgeschädigte

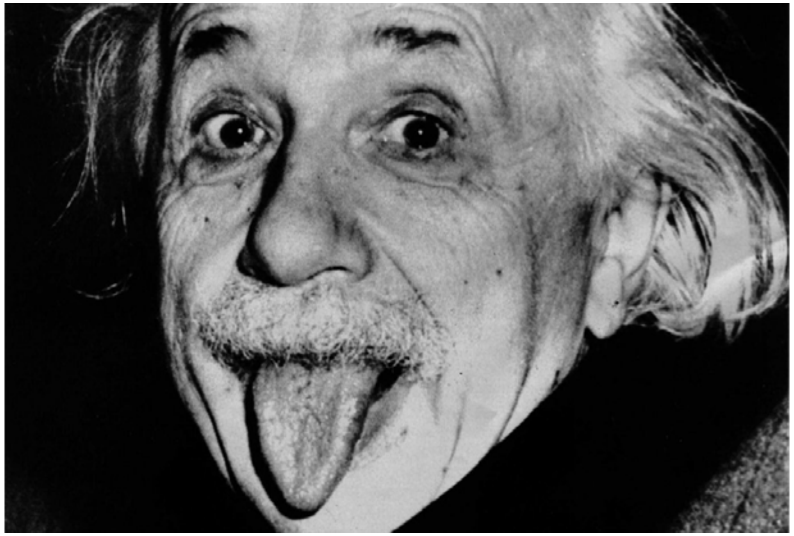
Institut für Anthropomatik und Robotik, Fakultät für Informatik

Lecture 3

FACE DETECTION

Motivation – Why Face Detection? (1)

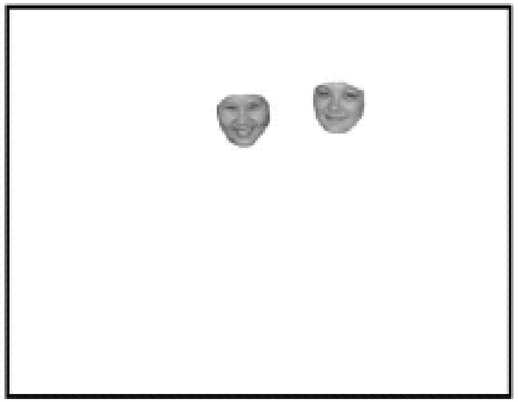
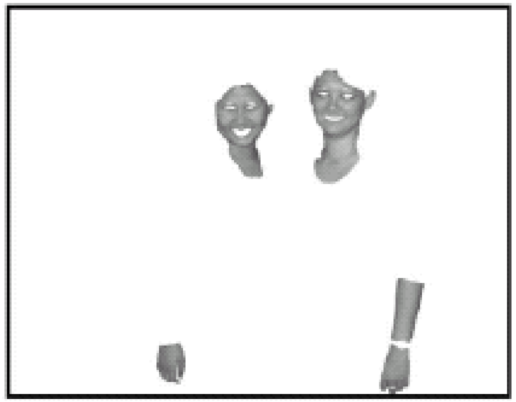
- Person Identification
- Perception of emotional expressions
- Mouth as source of speech
- Lipreading
- Perception of intention & attention
- Perception of age
- Perception of gender
- Perception of ethnical race



How would you approach it ?

- Features ?
- Classifier ?

Color Based Face Detection

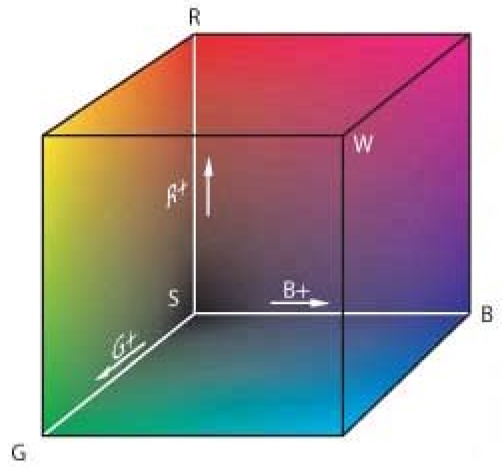
- **Rationale**: human skin has consistent color, which is distinct from many objects
- Possible approach:
  1. find skin colored pixels
  2. Group skin colored pixels
     - (and apply some heuristics) to find the face

Color

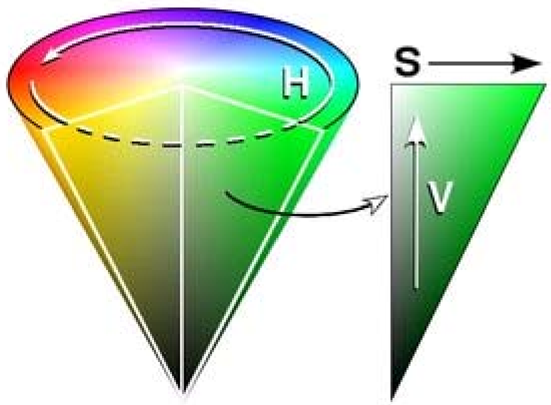
- Grayscale Image: Each pixel represented by one number
  - typically integer between 0 and 255
- Color image: Pixels represented by three numbers
- Different representations exist → „Color Spaces“

Color Spaces

- **RGB**: most widely used, specifies colors in terms of the primary colors red (R), green (G), and blue (B).

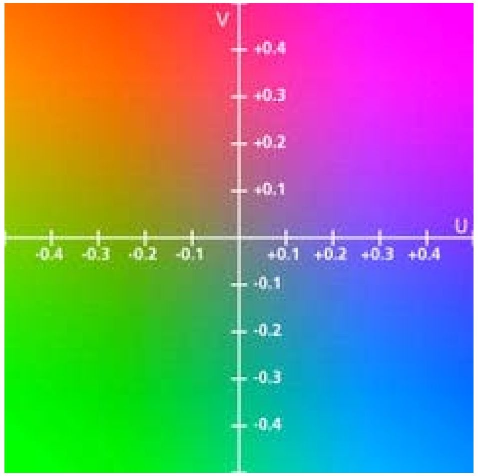


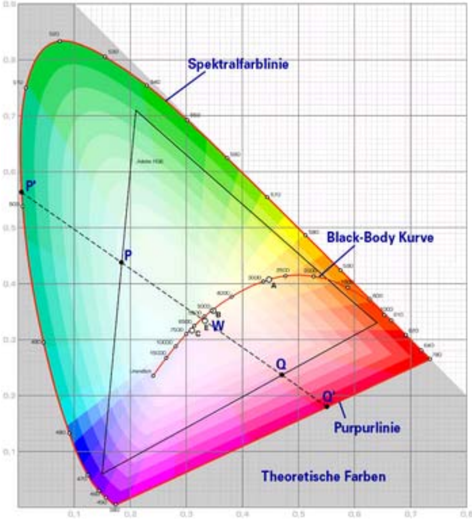
- **HSV/HSI:** hue (H), saturation (S) and value(V)/intensity (I)
  - Closely related to human perception (hue, colorfulness and brightness)



Color Spaces (2)

- **Class Y spaces**: YCbCr (Digital Video), YIQ (NTSC), YUV (PAL)
  - Y channel contains brightness, other two channels store chrominance (U=B-Y, V=R-Y)
  - Conversion from RGB to Yxx is a linear transformation
- **Perceptually uniform spaces**: e.g. CIE-Lab, CIE-Luv, ...
  - Perceived color difference is uniform to difference in color values
  - Euclidian distance can be used for color comparison





Color Spaces (3)

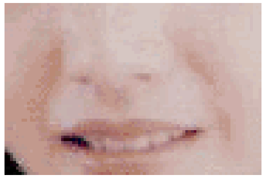
- Chromatic Color Spaces
  - Two color channels containing chrominance (colour) information

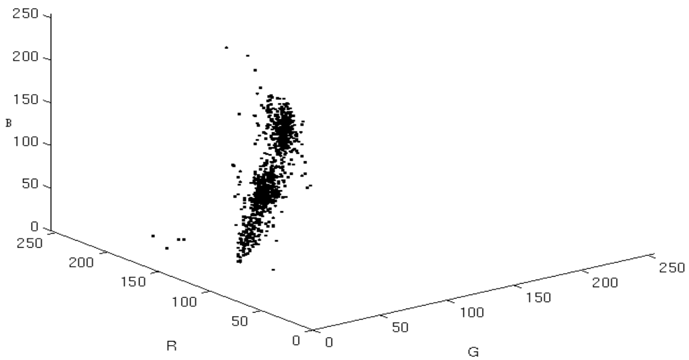
  - **HS** (taken from HSV)
  - **UV** (taken from YUV)

  - **Normalized rg** from RGB:
    - r = R / (R+G+B)
    - g = G / (R+G+B)
    - b = B / (R+G+B)

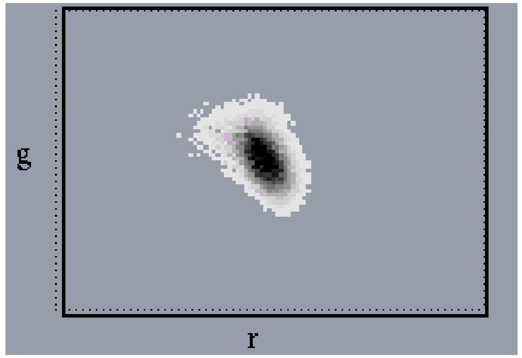
- Motivation: sometimes it is argued that chromatic skin color models are more robust

Skin Color Distribution – Examples (1)



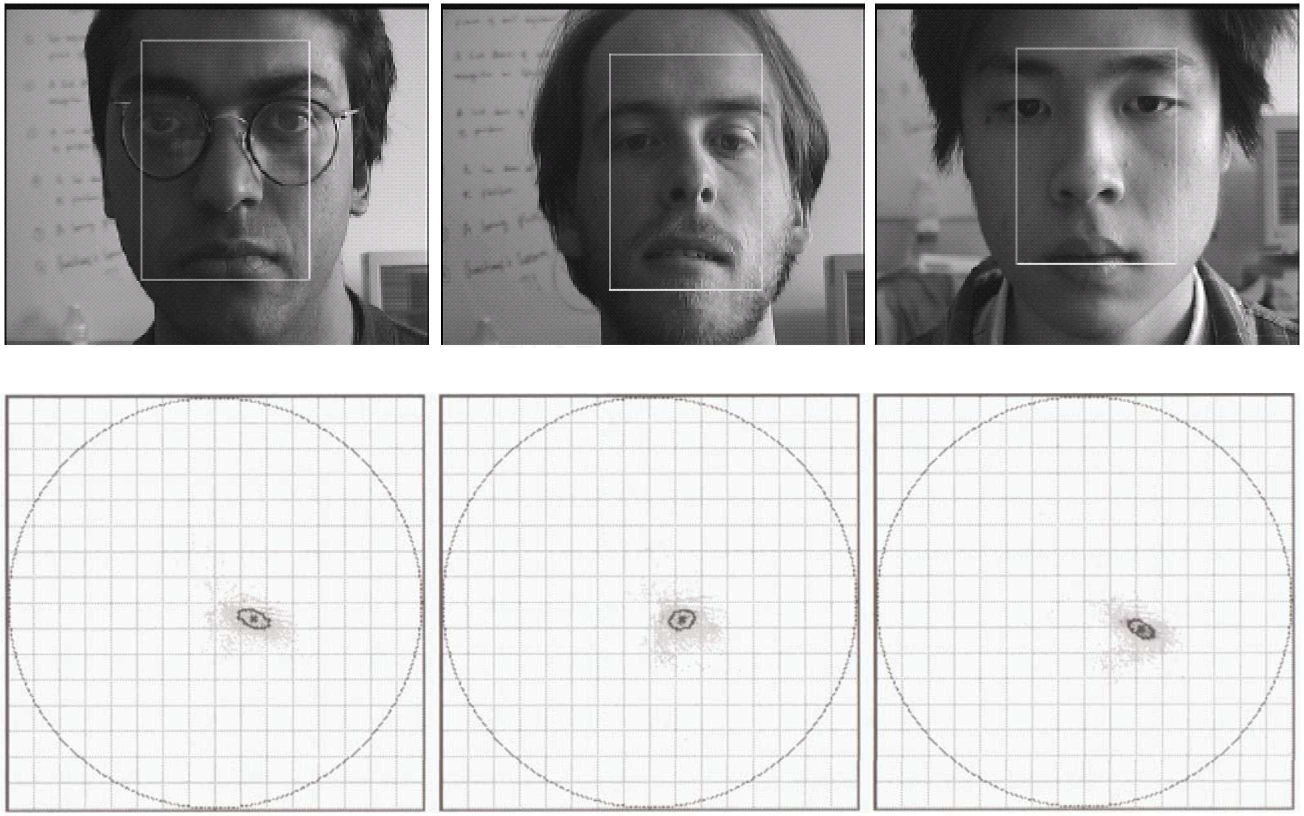


RGB space



rg space

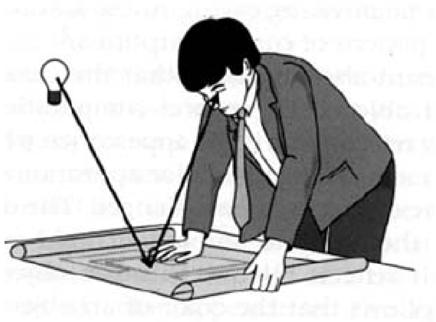
Skin Color Distribution – Examples (2)

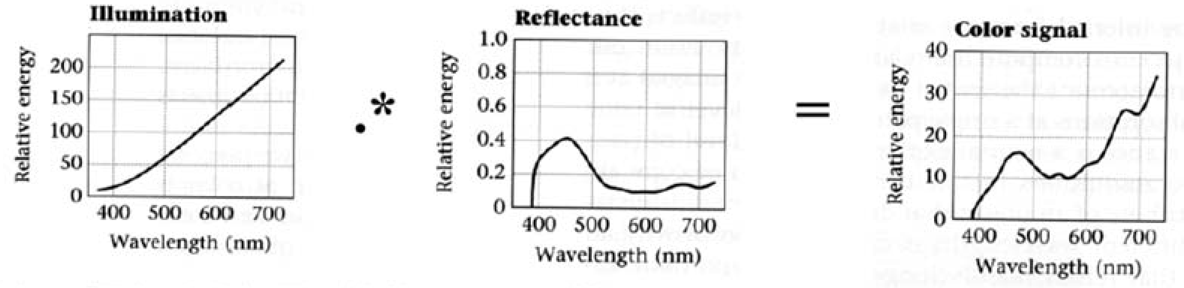


Typical color distributions in HS color space

Problems!

- Reflected color depends on spectrum of the light source (and properties of the object / surface)

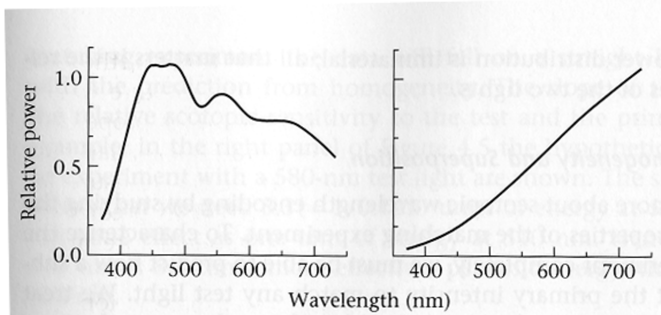




Foundations of Vision, by Brian Wandell, Sinauer Assoc., 1995

- If the light source / illumination changes, the reflected color signal changes!

- This is problematic!

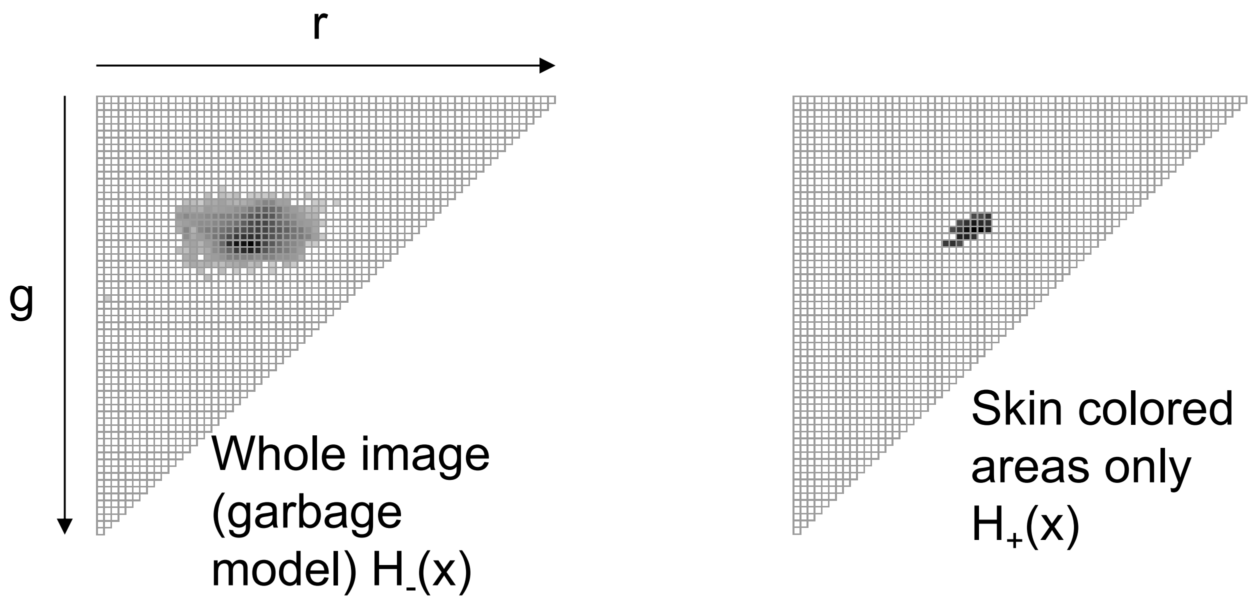


4.4 THE SPECTRAL POWER DISTRIBUTION of two important light sources are shown: (left) blue skylight and (right) a tungsten bulb.

How to model skin color ?

- Non-parametric models
  - → typically histograms
- Parametric models
  - Gaussian Model
  - Gaussian Mixture Model
- Or just learn decision boundaries between classes
  - → discriminative model
  - ANN, SVM, …

Histogram as skin color model



- Works very well in practice
- Memory size quickly gets high
  - 128MB if n=256 (RGB space),
  - 256KB if n=32 (RGB space)
- A large number of labelled skin and non-skin samples is needed !

Histogram Backprojection

- The simplest (and fastest) way to utilize histogram information is the histogram backprojection
- Each pixel in the backprojection is set to the value of the (skin-color) histogram bin indexed by the color of the respective pixel
  - A color x is considered as skin color if $H_+(x) > \theta$

Histogram Matching

- Backprojection is good, when the color distribution of the target is monomodal.
- Backprojection is not optimal, when the target is multi colored!
- Solution: Build a histogram of the image within the search window, and compare it to the target histogram.
- Many distance metrics for histograms, e.g.:
  - Battacharya distance
  - Histogram intersection
  - Earth-movers distance,…

Histogram Backprojection vs. Matching

- Histogram Backprojection
  - Compares color of a single pixel with color model
  - Fast and simple
  - Can only cope well with mono-modal distributions
  - sufficient for skin-color classification
- Histogram Matching / Intersection
  - Compares color histogram of image patch with color model
  - Better performance
  - Can cope with multi-modal distributions
  - Computationally expensive

Other Models:Gaussian Density Models

- Gaussian Densities
  - Assume that the distribution of skin colors *p(x)* has a parametric functional form
  - Most common funtion: Gaussion function G(**x**; μ, **C**):

$$p(x|skin) = G(\mathbf{x}; \mu, \mathbf{C}) = (2\pi)^{-d/2} |\mathbf{C}|^{-1/2} \exp\{-1/2\, (\mathbf{x} - \mu)^T \mathbf{C}^{-1} (\mathbf{x} - \mu)\}$$

  - Mean μ and covariance matrix **C** are estimated from a training set of skin colors S = $\{x_1, x_2, ..., x_N\}$:
    - $\mu = E\{\mathbf{x}\}$, $\mathbf{C} = E\{(\mathbf{x} - \mu)\,(\mathbf{x} - \mu)\}$
  - A color is considered as skin color if
    - $p(x|skin) > \theta$ or
    - $p(x|skin) > p(x|non\text{-}skin)$

Mixture of Gaussians Models

- Mixture of Gaussians
  - One Gaussian might not be sufficient to describe the distribution of skin colors (e.g. in HS-space)

$$p(x) = \sum_{i=1}^{K} \pi_i G(x, \mu_i, C_i)$$

  - Parameter set Φ can be estimated using the EM algorithm
    - Iteratively changes parameters so as to maximize the log-likelihood of the training set:

$$L = \log \prod_{i=1}^{N} p(x_i \mid \Phi)$$

  - A color is considered as skin color if
    - *p(x|skin)* $\geq \theta$
    - or *p(x|skin)* > *p(x|non-skin)*

Bayes Classifier

- Skin Classification using Bayes Decision Rule
  - Minimum cost decision rule
  - Classify pixel to skin class if *P(Skin|x) > P(Non-Skin|x)*
  - Decision Rule: $$\frac{p(\mathbf{x} \mid Skin)}{p(\mathbf{x} \mid Non-Skin)} \geq \frac{P(Non-Skin)}{P(Skin)}$$
  - The classconditionals p(x|ω) can be estimated from the corresponding histograms:

$$p(x \mid \omega_i) = h_i(x) / \sum_x h_i(x),$$

where $h_i(x)$ is the count of pixels from class ω_i that have value x

Discriminative Models / Classifiers

- Artificial Neural Networks
- Support Vector Machine
- …

Many possible approaches …

- Histogram backprojection / matching
- Gaussian / Gaussian Mixture model
- Bayes Classification
- Discriminative approaches: SVM, ANN

- Different color models
  - 3D color models: RGB, HSV, …
  - Chromatic color models (2D)

- <u>Which approach is best ?</u>

Performance Measures

Performance Measures

- Measuring the performance of object recognition algorithms is not trivial
- There are different measures depending on the application

1. For classification (i.e. yes/no decision, if object is present or not)
   - ROC (Receiver-Operating-Characteristic)

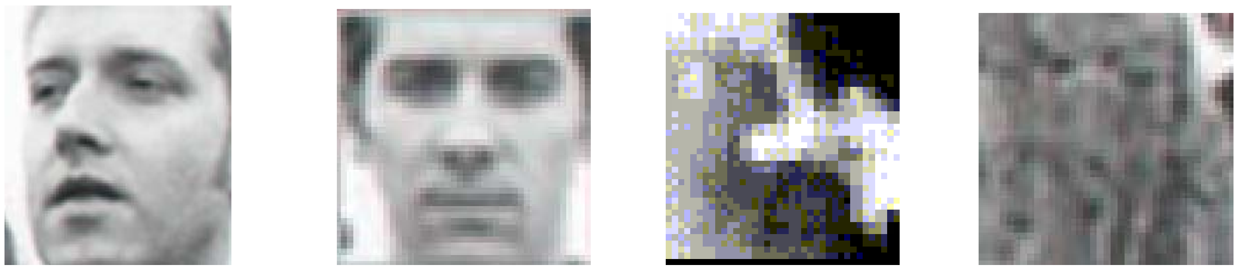
2. For localization (i.e. detecting the object's position)
   - RPC (Recall-Precision-Curve)
   - DET (Detection Error Trade-Off)

Classifying a hypothesis

- When comparing recognition hypotheses with ground-truth annotations have to consider four cases:

- Example:

	Predicted positive	Predicted negative
Positive examples (Pos)	*True positive* (TP)	*False negative* (FN)
Negative example (Neg)	*False positive* (FP)	*True negatives* (TN)



Prediction:	Yes	No	Yes	No
Case:	TP	FN	FP	TN

ROC

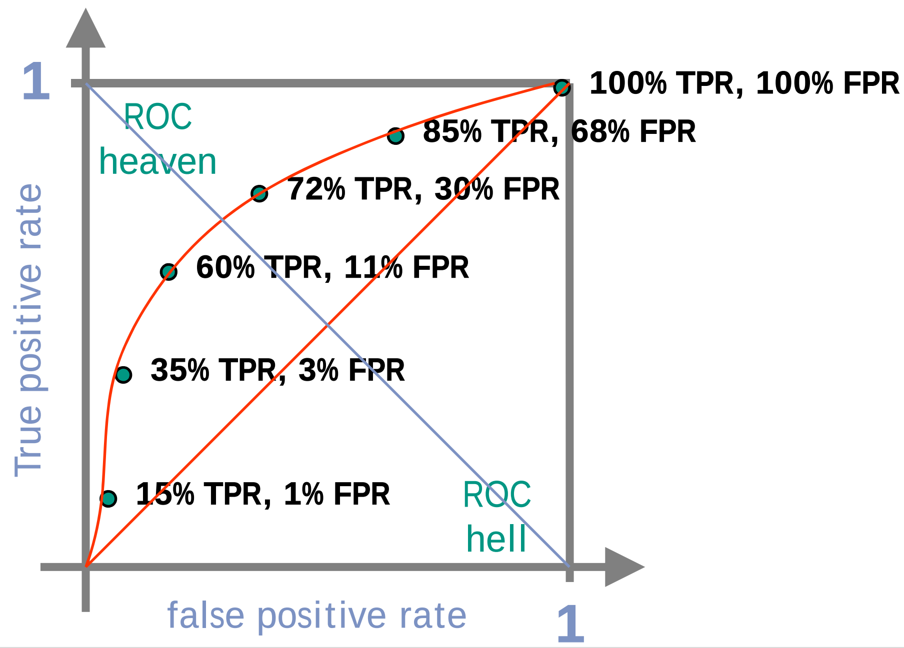
- Used for the task of classification
- Measures the trade-off between true positive rate and false positive rate:

$$\text{true positive rate} = \frac{\mathrm{TP}}{\mathrm{Pos}} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$$
$$\text{false positive rate} = \frac{\mathrm{FP}}{\mathrm{Neg}} = \frac{\mathrm{FP}}{\mathrm{FP+TN}}$$

- Example:
  - Algorithm X detects 80% of all faces (true positive rate), while making 25% error on images not containing faces

ROC

- Each prediction hypothesis has generally an associated probability value or score
- The performance values can therefore plotted into a graph for each possible score as a threshold



Skin-color: Analysis and Comparison

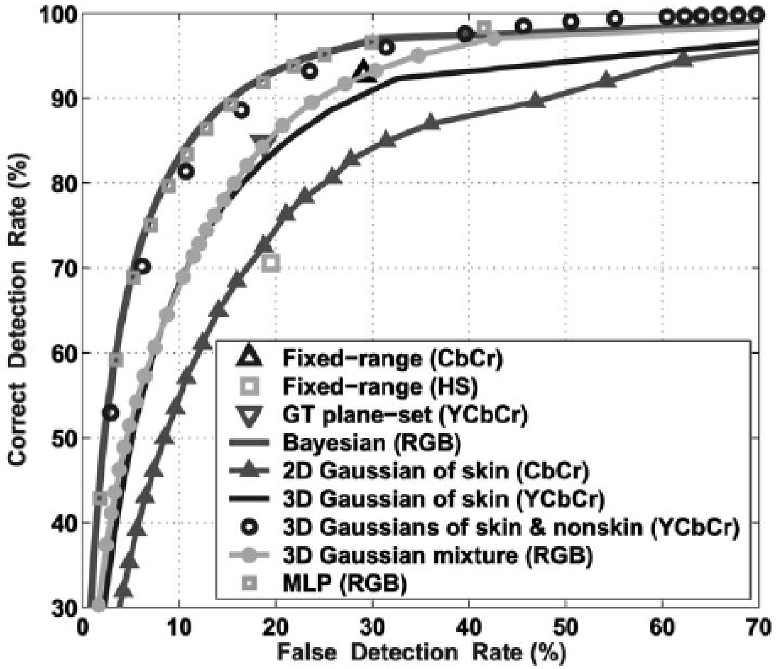
- Phung et al., Skin segmentation using color pixel classification: Analysis and comparision, IEEE PAMI, Vol.27 (1), Jan. 2005

- Database: ECU face and skin detection database
  - 4000 images, mainly from the Web, diversity of background, illumination, face & skin types
  - 12.000 face images, 2000 landscape images

ID	Classifier	Classifier Parameters	Color Representation
CbCr-fixed	CbCr fixed-range: $77 \leq Cb \leq 127$ and $133 \leq Cr \leq 173$	[9]	CbCr
HS-fixed	HS fixed-range: $0.23 \leq S \leq 0.68$ and $0 \leq H \leq 50^o$	[11]	HS
GT plane-set	Garcia & Tziritas' plane set: skin cluster by 8 planes in YCbCr	[10]	YCbCr
Bayesian	Bayesian classifier with the histogram technique: 256^3 bins	trained	RGB
2DG-pos	2-D unimodal Gaussian of skin	trained	CbCr
3DG-pos	3-D unimodal Gaussian of skin	trained	YCbCr
3DG-pos/neg	3-D unimodal Gaussians of skin and nonskin	trained	YCbCr
3DGM	3-D Gaussian mixture of skin and nonskin	[4]	RGB
MLP	Multilayer perceptron	trained	RGB

Skin-color: Analysis and Comparison (2)

- Conclusions
  - Bayesian approach and MLP worked best
    - Bayesian approach needs much more memory
  - Approach is largely unaffected by choice of color space, but
  - Results degraded when only chrominance channels were used



Phung et al., Skin segmentation using color pixel classification: Analysis and comparision, IEEE PAMI, Vol.27 (1), Jan. 2005

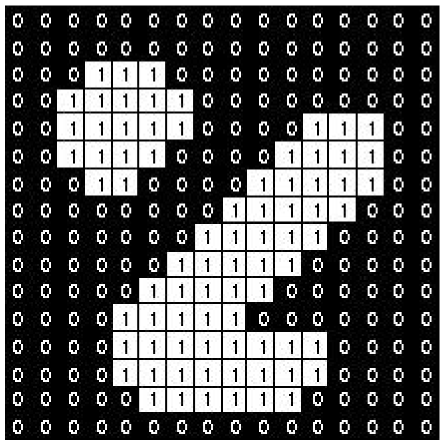
From Skin-Colored Pixels to Faces

- Skin-colored pixels need to be grouped into object representations

- Problems: skin-colored background, further skin-colored body parts (hands, arms, …), Noise, …

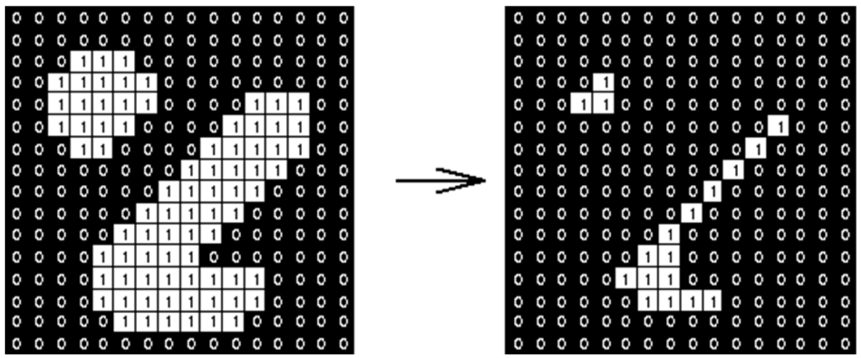
Perceptual Grouping (1)

- Morphological Operators: Operators performing an action on shapes where the input and output is a binary image.
- Threshold each pixel‘s skin affiliation ➔ Binary Image



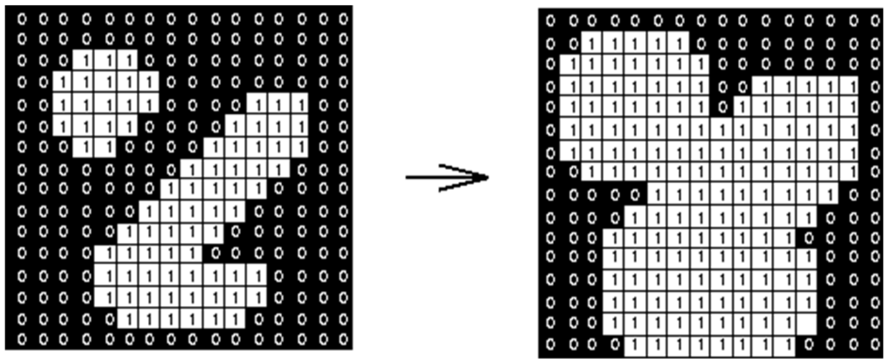
Perceptual Grouping (2)

- Morphological Erosion:
  - Remove pixels from edges of objects
  - Set pixel value to min value of surrounding pixels



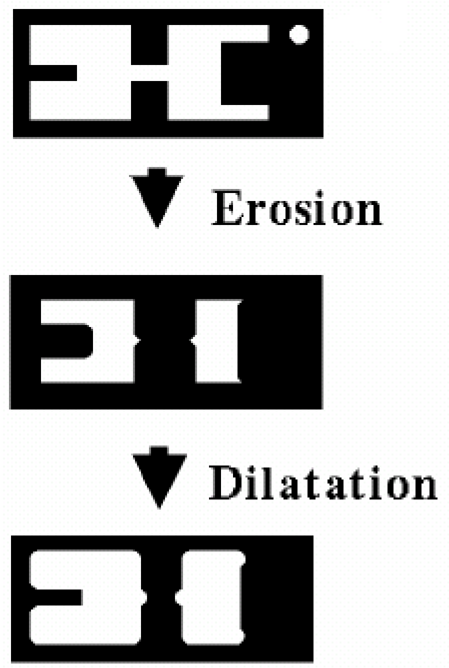
Perceptual Grouping (3)

- Morphological Dilatation:
  - Add pixels to edges of objects
  - Set pixel value to max value of surrounding pixels



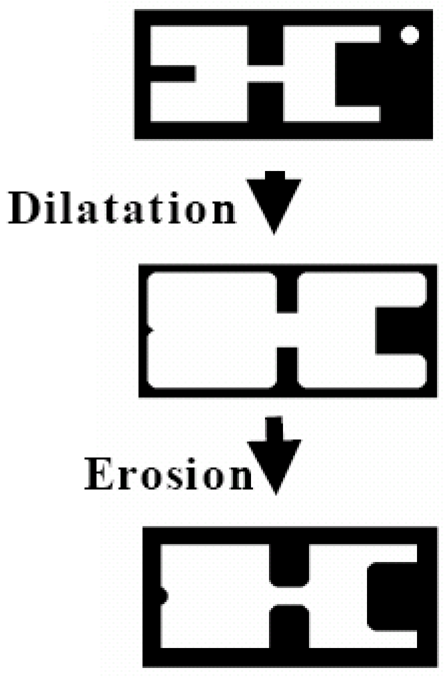
Perceptual Grouping (4)

- **Morphological Opening:**
  - Apply erosion, then dilatation
  - Goal:
    - Smooth outline
    - Open small bridges
    - Eliminate outliers

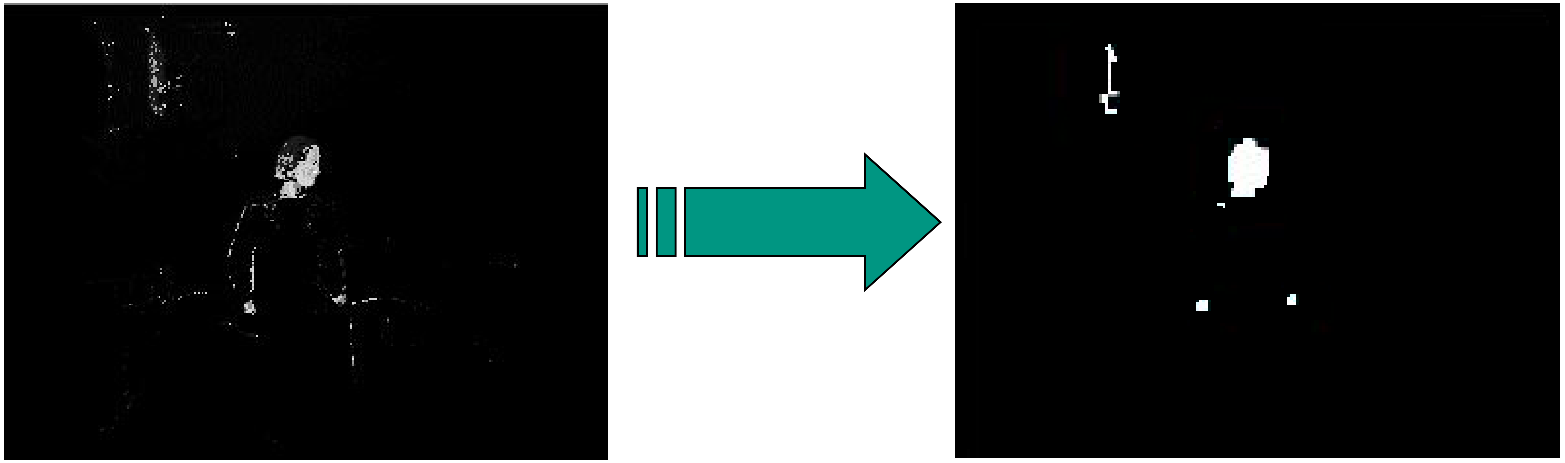


Perceptual Grouping (5)

- **Morphological Closing:**
  - Apply dilatation, then erosion
  - Goal:
    - Smooth inner edges
    - Connect small distances
    - Fill unwanted holes



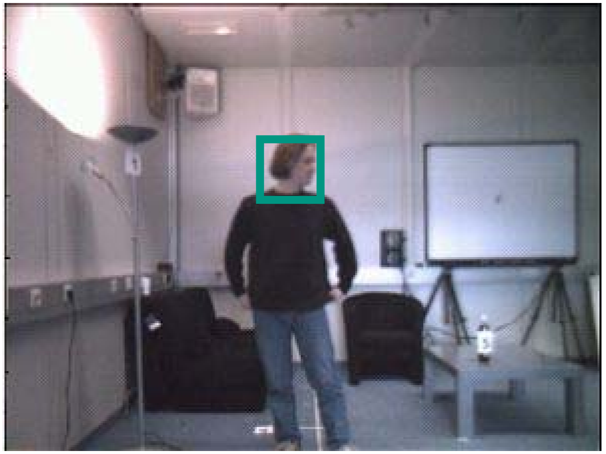
Perceptual Grouping (6)

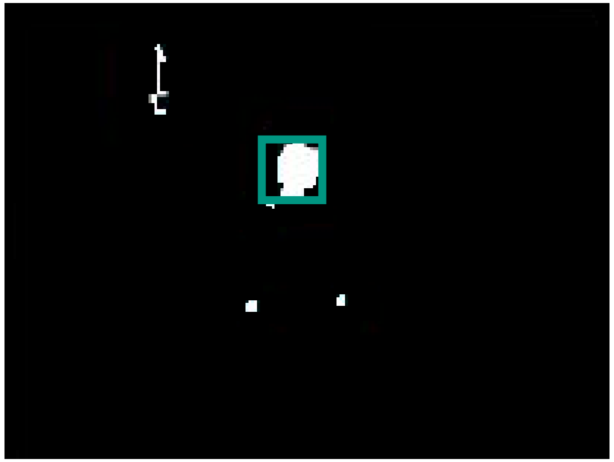


- Apply morphological closing then morphological opening
- Resulting image is reduced to connected regions of skin color (blobs)

From Skin Blobs To Faces

- Goal: align bounding box around face candidate
- Important for:
  - Face Recognition
  - Head Pose Estimation
- Different approaches:
  - Choose cluster with biggest size
  - Ellipse fitting (approximate face region by ellipse)
  - Heuristics to distinguish between different skin clusters
  - Use temporal information (tracking)
  - Facial Feature Detection
  - …





Real-time color-based face tracking (1996)

J. Yang & A. Waibel, A real-time face tracker, WACV 1996

- Chromatic skin color model (r,g)
- Color modelled using one Gaussian
- Color-model gradually adapted to cope with illumination changes
- Active camera control
- Runs @ 15-30 fps (in 1996!), depending on face size

Summary – Skin color classification

- Different color spaces and classifiers can be used
  - Models: histograms, Gaussian Models, Mixture of Gaussians Model
  - Histogram-backprojection / Histogram matching
  - Bayes classifier
  - Discriminative Classifiers (ANN, SVM)

- Bayesian classifier and ANN seem to work well
  - Sufficient training data is needed for modeling the pdf, in particular for Bayesian approach (positive & negative pdfs learned)

- Advantages: Fast, rotation & scale invariant, robust against occlusions

- Disadvantages:
  - Affected by illumination
  - Cannot distinguish head and hands
  - Skin-colored objects in the background problematic

- Metric: ROC curve used to compare classification results / methods
  - True positive rate vs. false positive rate

References

- Phung et al, *Skin Segmentation Using Color Pixel Classification: Analysis and Comparison,* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, 1, Jan. 2005
- Michael J. Swain and Dana H. Ballard, *Color Indexing*, International Journal on Computer Vision, 7:1, 11-32 (1991)