Face Detection 1

Vorlesung “Computer Vision für Mensch-Maschine-Interaktion”
WS 2013/14

Rainer Stiefelhagen

18.11.2013

Motivation – Why Face Detection? (1)

A face provides different functions:

- Person identification
- Perception of emotional expressions
- Mouth as source of speech
- Lipreading
- Perception of intention
- Perception of age
- Perception of gender
- Perception of ethnical race
- …

Motivation – Why Face Detection? (2)

Computer-based face perception is important for:

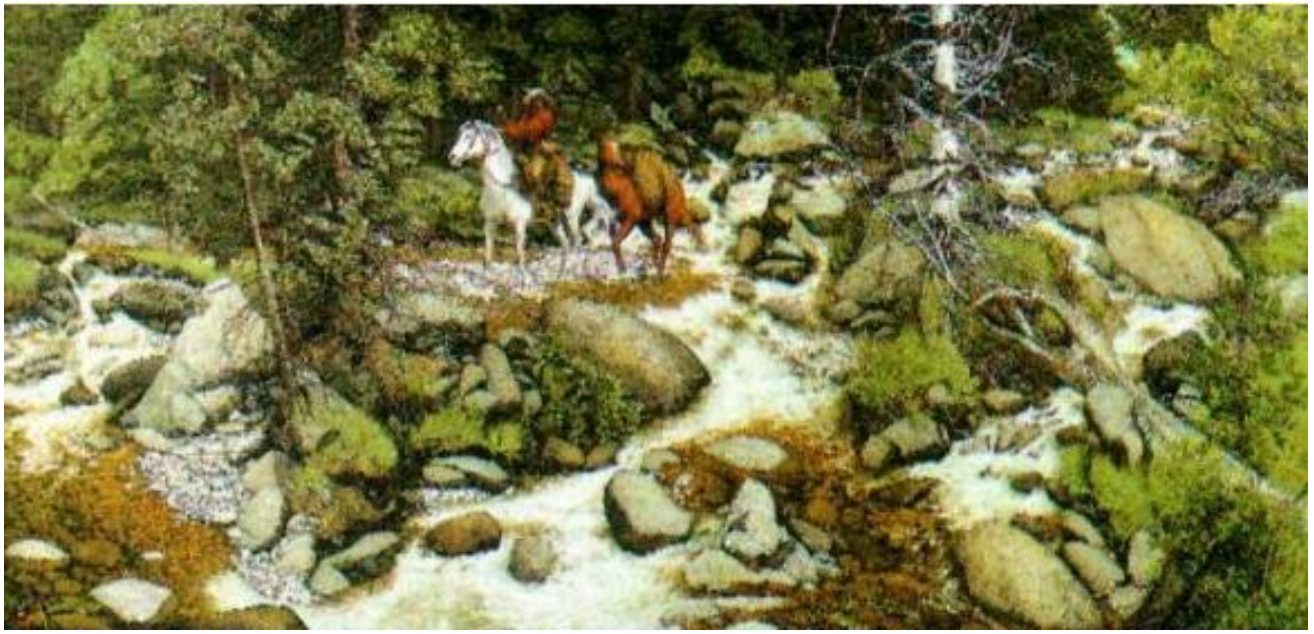
- Human-Machine Interfaces
- Multimedia
- Surveillance
- Security
- Telephone conferences
- Communication
- Animations
- …

Motivation

- Context Aware Environments
  - Smart Meeting / Lecture Rooms
  - Smart Houses
  - Smart Cars
  - Assisted Living
  - Humanoid Robots
- See first lecture!

What makes Face Detection so difficult? (1)

How many faces do you see?



What makes Face Detection so difficult? (2)

Finding faces in unconstrained scenes:

- Humans can do it!!
- We don't know for sure how we do it ☹

What makes Face Detection so difficult? (3)

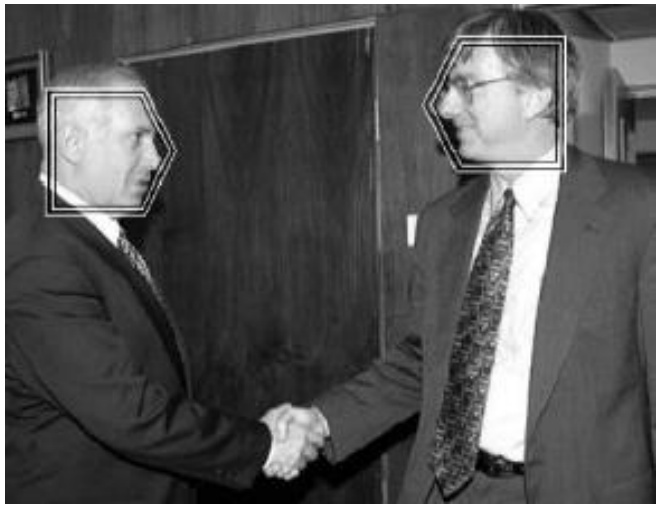
Numerous possible illuminations:

- The same face looks different every time
- There is a lot of work about normalizing shadow effects and reducing illumination changes
- Different people's faces lit from the same direction are more similar than the same person's face lit from different directions

What makes Face Detection so difficult? (4)

Rotation:

- Undoing rotation too expensive
- How to detect rotated faces directly? Different classifiers for different rotations?



What makes Face Detection so difficult? (5)

- Intrinsic variations of facial appearance

Source	Possible Tasks
Identity	Classification, known-unknown, verification, identification
Facial expressions	Inference of emotion
Speech	Lip-reading
Sex	Deciding whether male or female
Age	Estimating age

- Extrinsic variations of facial appearance

Source	Effects
Viewing geometry	Pose
Illumination	Shading, color, self-shadowing, specular highlights
Imaging process	Resolution, focus, imaging noise, perspective effects
Other objects	Occlusion, shadowing, indirect illumination

What makes Face Detection so difficult? (6)

- Boundaries of faces not clear: depends on hair styles (although there are approaches trying to do that)
- It is impossible to model intrinsic parameters analytically

What makes Face Detection so difficult? (7)

- Classification face / non face is very complex (we can not model the whole world!)
- Intrinsic and extrinsic parameters do not cover different make-up styles, glasses, jewelery, …
- Biggest problem is head pose & lighting
  - Face appearance changes dramatically with its pose !
- Perception of faces is highly dynamic in space, time and its context
  - A person is more likely to be found behind a desk than on top of a book shelf.

Computerized Face Representation (1)

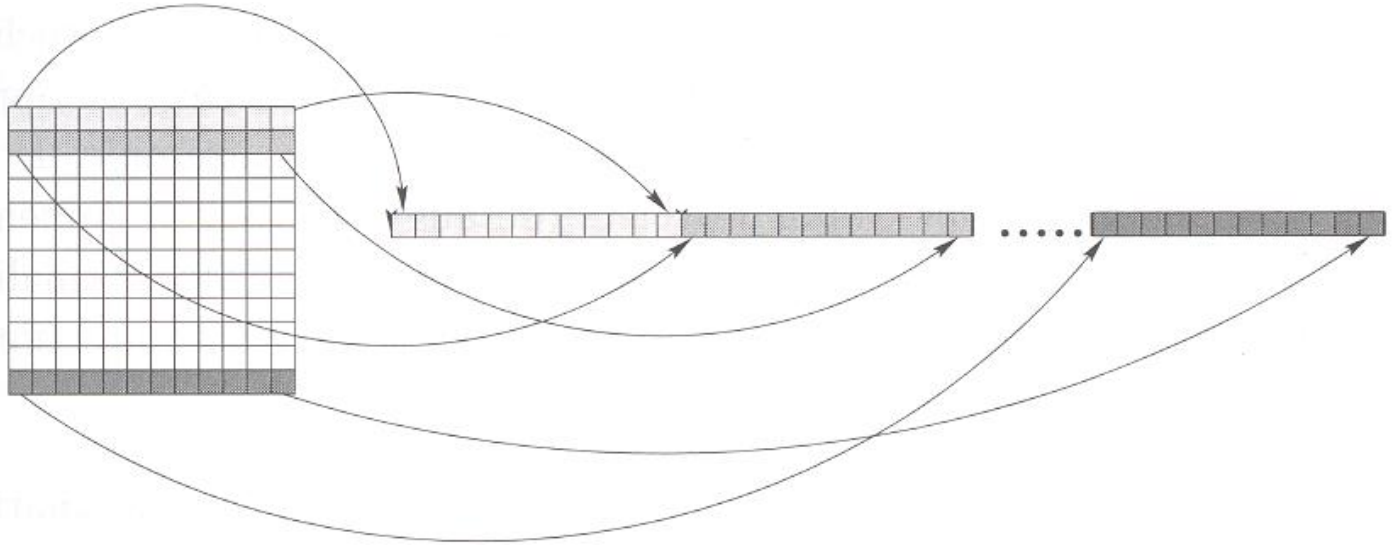
- Local feature-based face representation:
  - „Representation by parts“
  - Find facial features (such as mouth corners, eyes, nostrils, …) and check for well-defined spatial configuration (a priori knowledge)
  - Use of relatively high resolution images (> 60x60 pixels)
    - Humans can detect faces even within 15x15 pixels and lower !
  - Problems with Occlusions: what to do if several parts are not visible?
  - Problem of face detection is divided in multiple detection tasks, each nearly as difficult itself as face detection alone

Computerized Face Representation (2)

- Holistic face representation:
  - No a priori knowledge
  - No subdivision into single parts
  - Relatively low resolution possible
  - Occlusions handled as statistical outliers
  - Variances handled by different preprocessing steps, region of interest limitation and statistical learning
  - Detection can be achieved by
    - Classification between face images / non-face images
    - Application of generic face model on region of interests Detect faces by scoring how well models fit

Computerized Face Representation (3)

- Exemplary holistic feature vector:



Problem: Keep dimensionality low !!

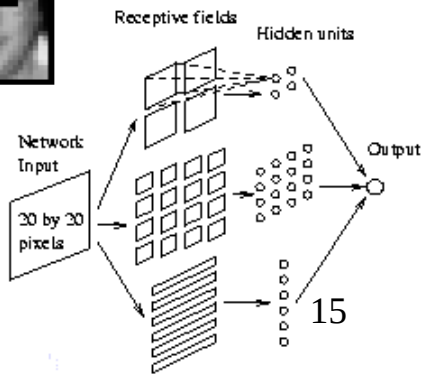
Different Face Detection Approaches

- **Skin-color detection**

- **An elliptical head model**

- **Using Haar-Filter cascades (Viola & Jones)**

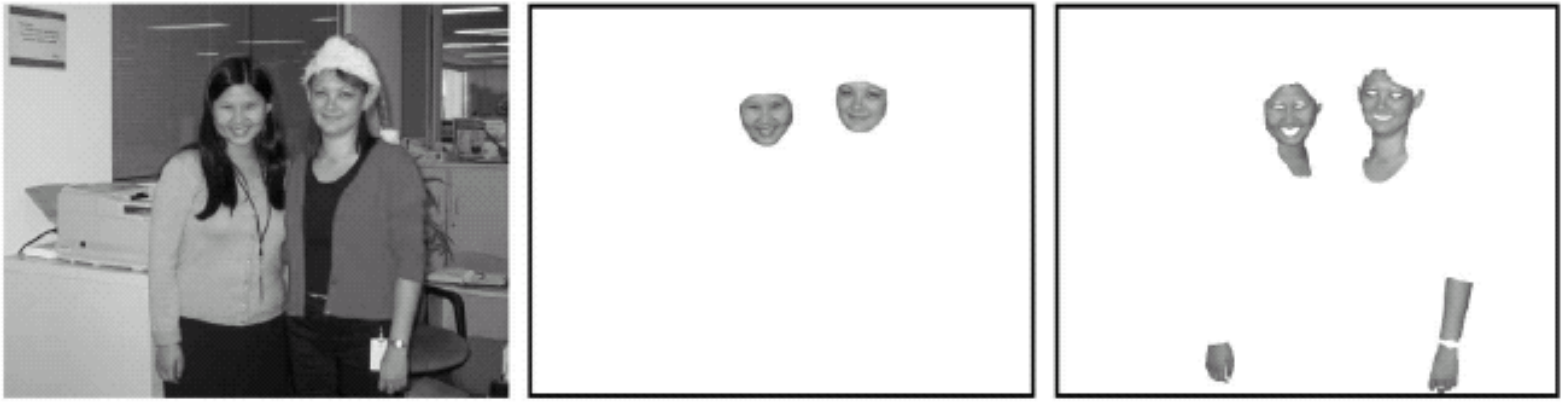
- **Artificial neural networks**



Roadmap

- **Color based face detection**
- An Ellipsoid head model
- Artificial Neural Networks
- Feature-based classifier cascades (Viola & Jones)

Color Based Face Detection



Rationale: human skin has consistent color, which is distinct from many objects ➔ segment skin colored pixels

Advantages:
- fast algorithms
- orientation & scale invariant
- stable against occlusions
- person-independent

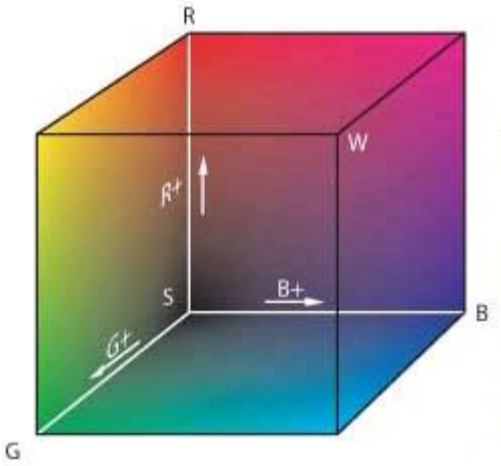
Disadvantages:
- affected by illumination
- does not distinguish heads and hands

Skin Color Segmentation

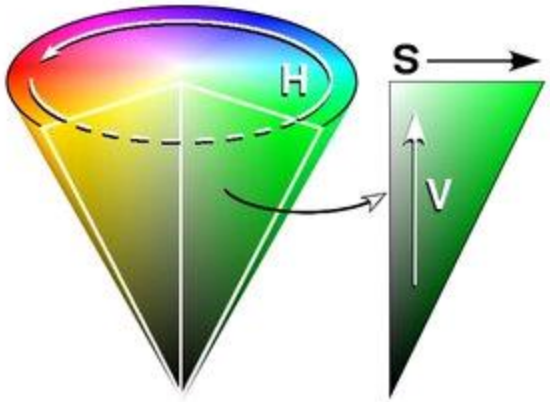
- Choice of color representation
  - RGB, HSV, YCbCr, ...
  - On chrominance only: rg, HS, UV, ...
- Choice of color model
  - Histogram
  - Gaussian Model / Gaussian Mixture Model
- Choice of classifier
  - „Thresholding“, Bayesian, multilayer perceptron, support vector machine, ...

Color Spaces (1)

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

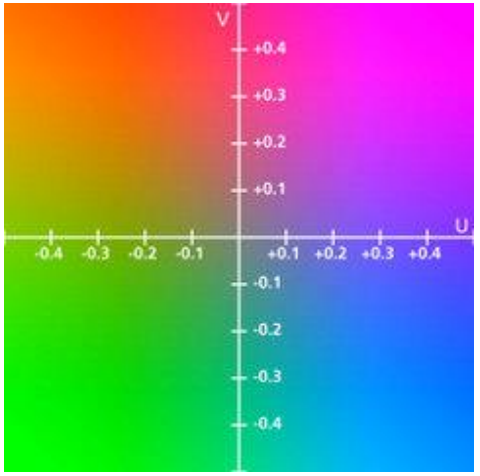


- **HSV/HSI:** hue (H), saturation (S) and value(V)/intensity (I)
  - Closely related to human perception (hue, colorfulness and brightness)

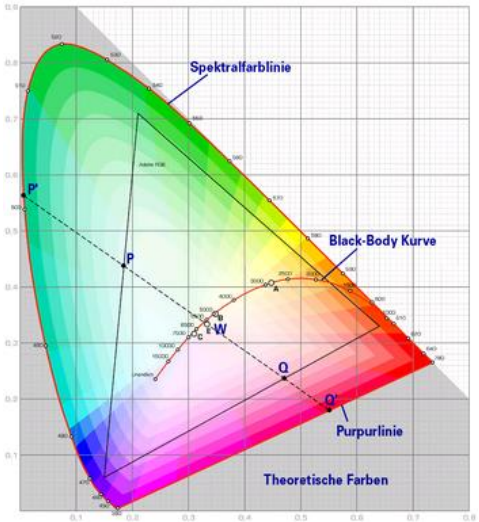


Color Spaces (2)

- **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



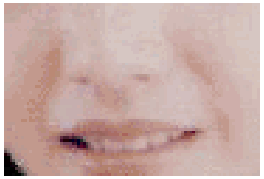
- **Perceptually uniform spaces**: e.g. CIE-Lab, CIE-Luv, ...
  - Perceived color difference is uniform to difference in color values
  - Euclidian distance can be used for color comparison

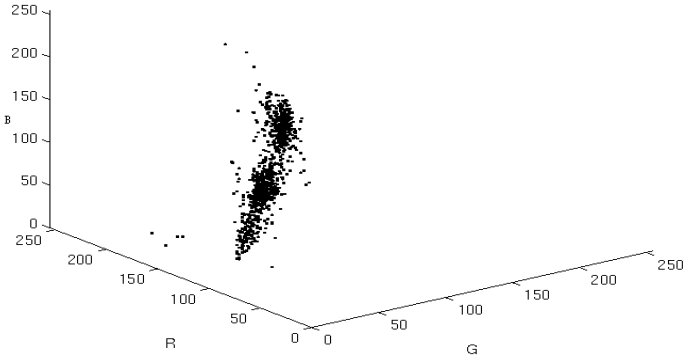


Color Spaces (3)

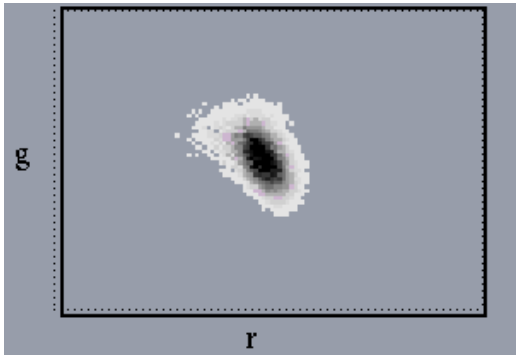
- **Chromatic Color Spaces**:
  - Two color channels containing chrominance (colour) information
  - **HS** (taken from HSV)
  - **UV** (taken from YUV)
  - **Normalized rg** from RGB:
    - r = R / (R+G+B)
    - g = G / (R+G+B)
    - b = B / (R+G+B)
- Motivation: sometimes it is argued that chromatic skin color models are more robust

Skin Color Distribution – Examples (1)



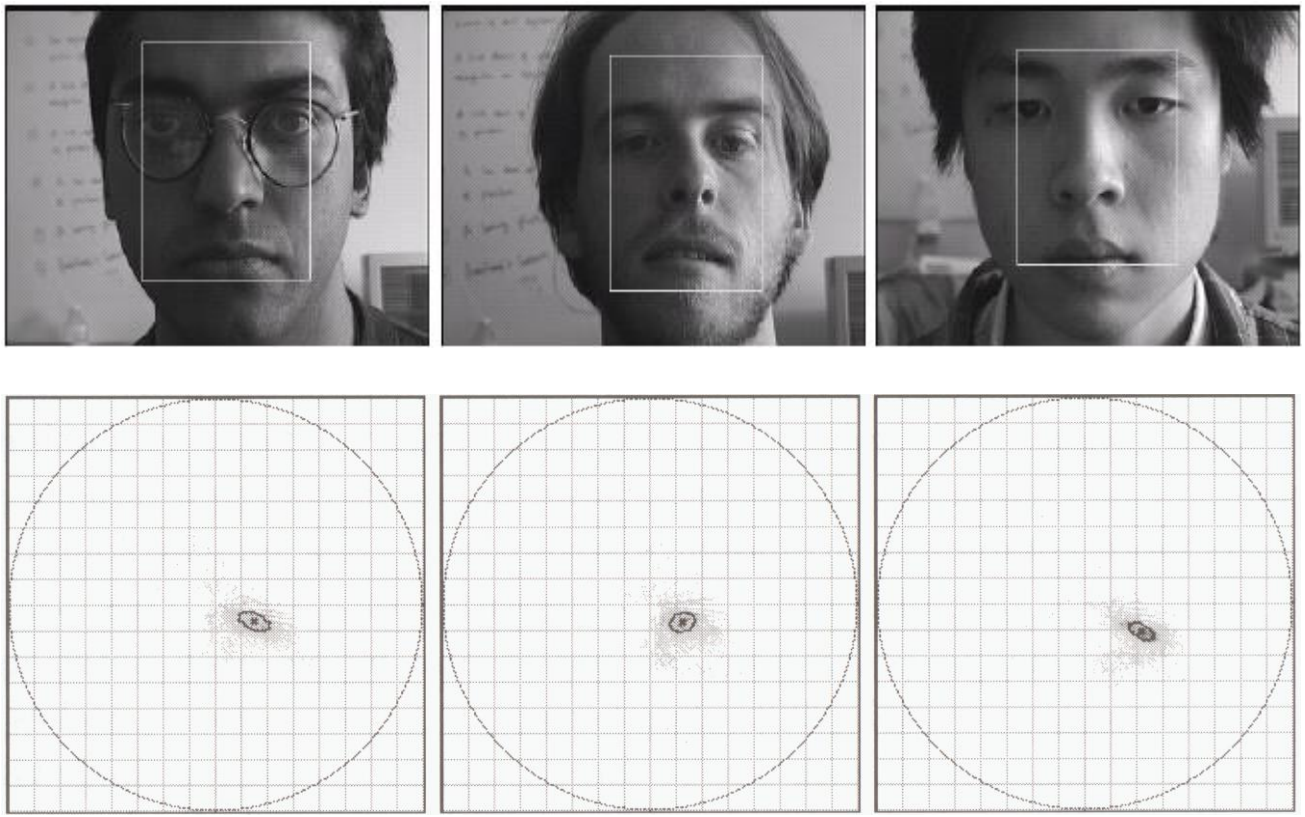


RGB space



rg space

Skin Color Distribution – Examples (2)

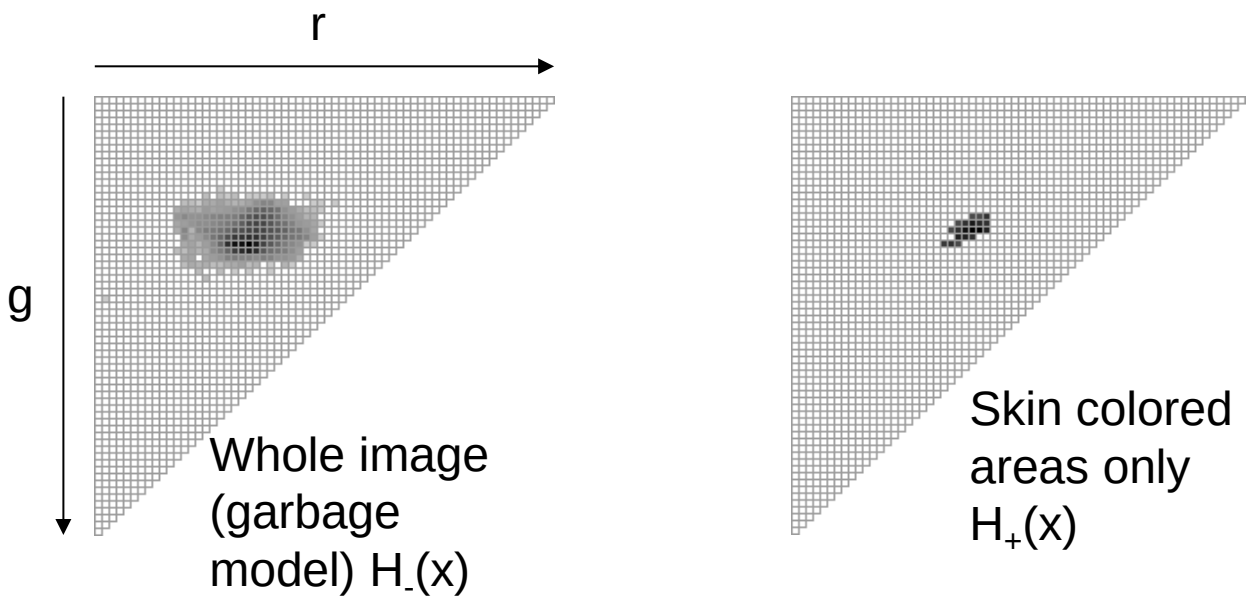


Typical color distributions in HS color space

How to model (specific) colors?

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

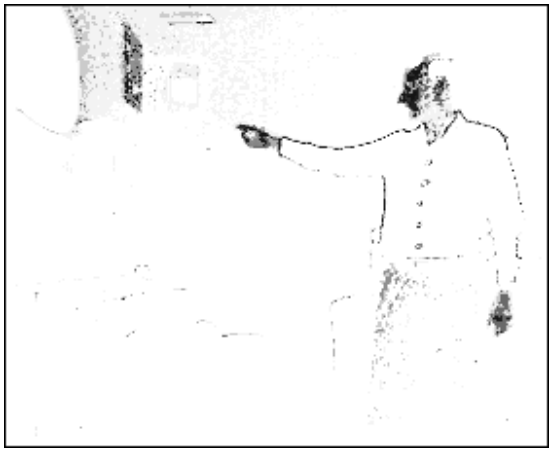
Histogram based skin color model



- 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!

→ Build a histogram of the image within the search window, and compare it to the target histogram.

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**):

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

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

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

Other Classifiers

- **Other classifiers:**
  - Artificial Neural Networks
    - Less memory needed
  - Self-Organized Maps
  - Multilayer Perceptron
  - SVM

Kurzer Einschub: 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:

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

- Example:




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:

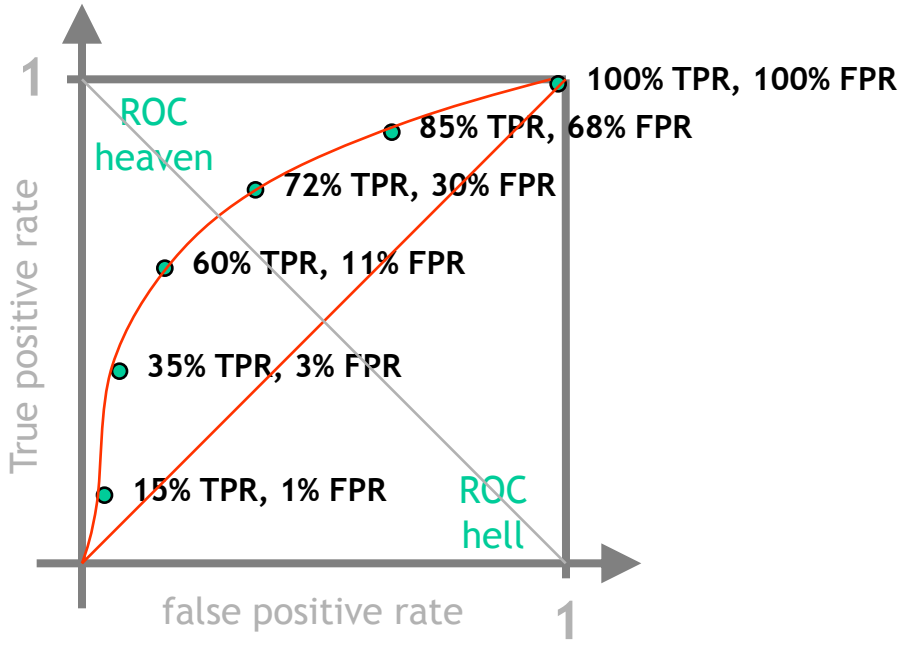
$$\text{true positive rate} = \frac{\text{TP}}{\text{Pos}} = \frac{\text{TP}}{\text{TP+FN}}$$

$$\text{false positive rate} = \frac{\text{FP}}{\text{Neg}} = \frac{\text{FP}}{\text{FP+TN}}$$

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

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 comparison, IEEE PAMI, Vol.27 (1), Jan. 2005

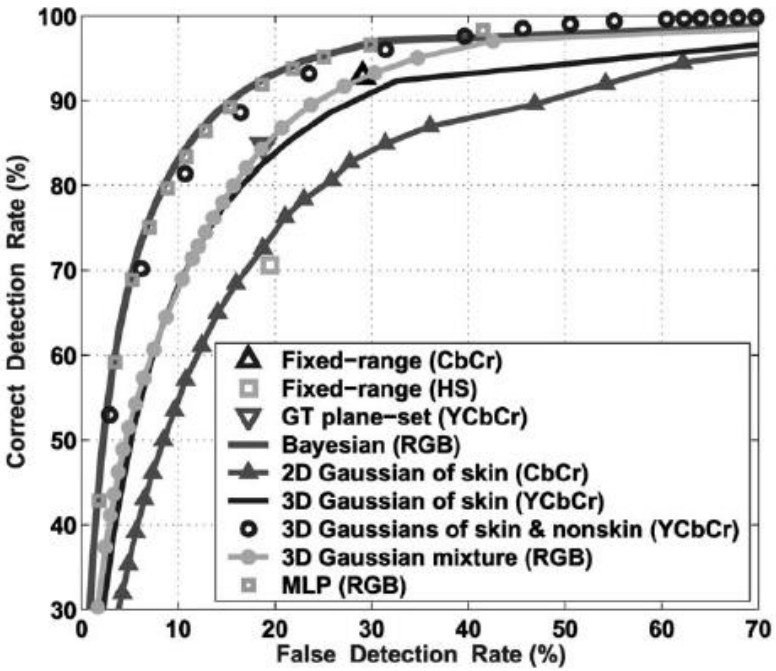
- Database: ECU face and skin detection database
  - 4000 images, mainly from the Web, diversity of background, illumination, face & skin types
  - (12.000 face images, 2000 landscape images)

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

Investigated classifiers (from Phung et al. 2005)

Phung et al., 2005: Results & Conclusions

- 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



ROC curves (from Phung et al.)

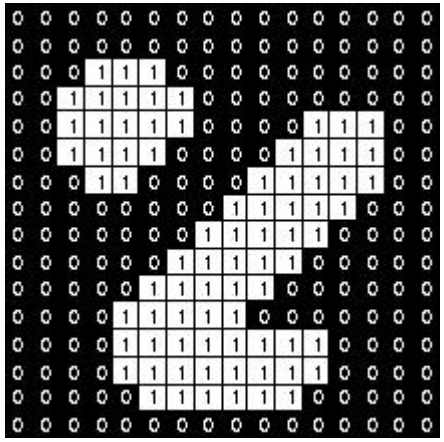
From Skin-Colored Pixels to Faces

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

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

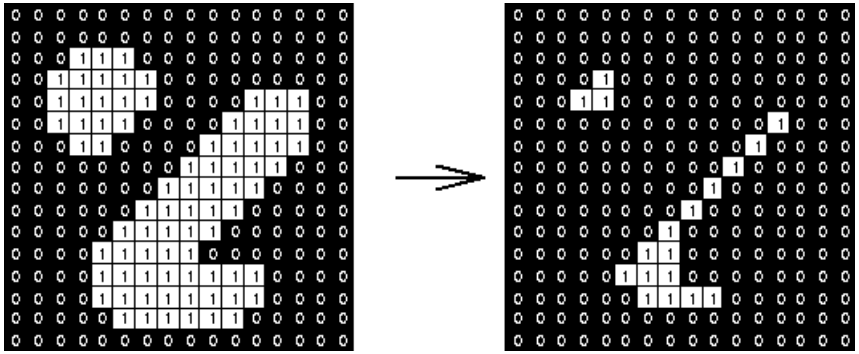
Perceptual Grouping (1)

- 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



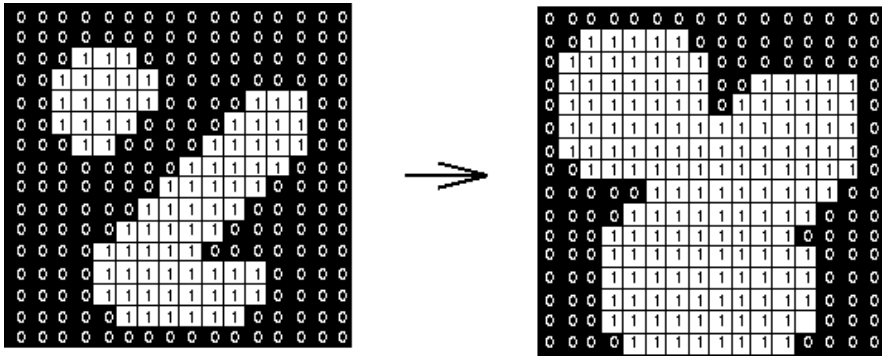
Perceptual Grouping (2)

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



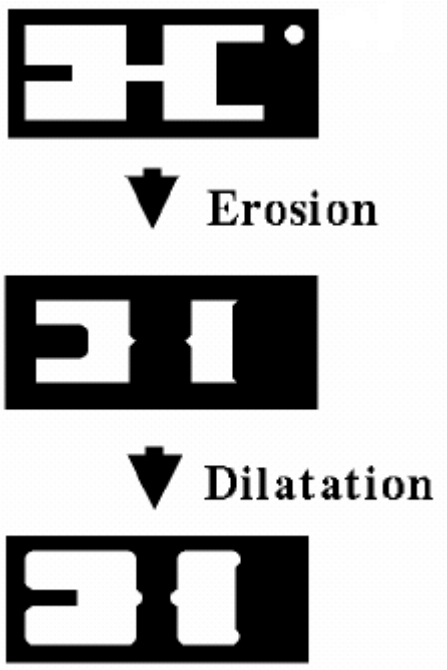
Perceptual Grouping (3)

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



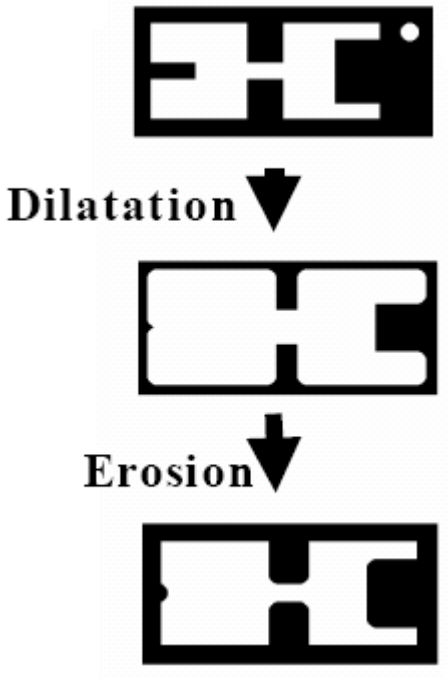
Perceptual Grouping (4)

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

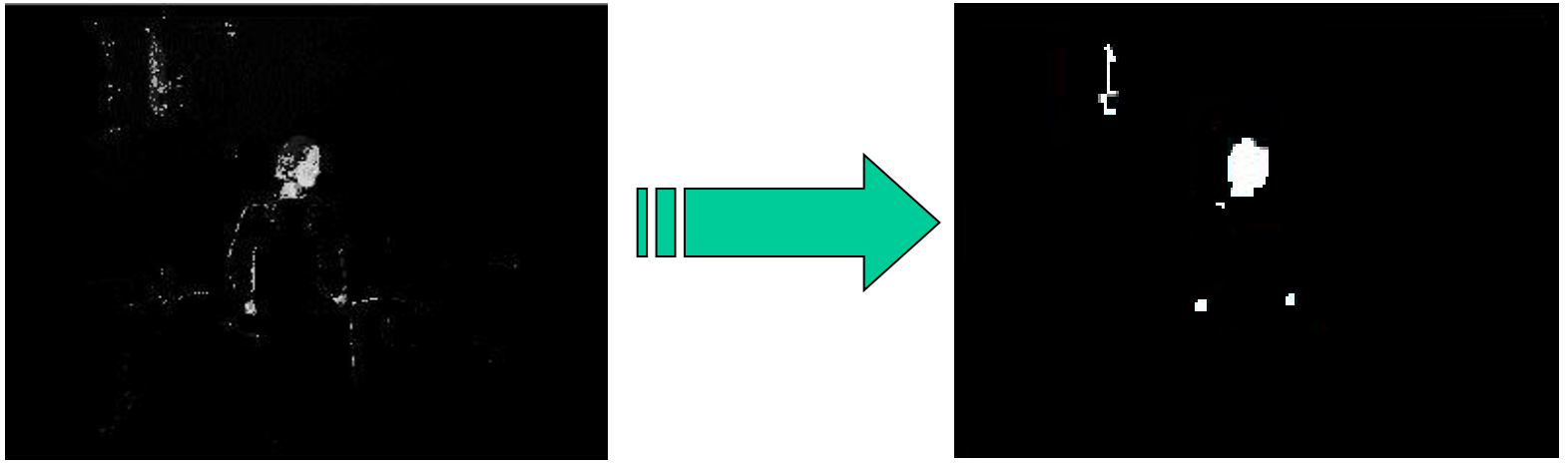


Perceptual Grouping (5)

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



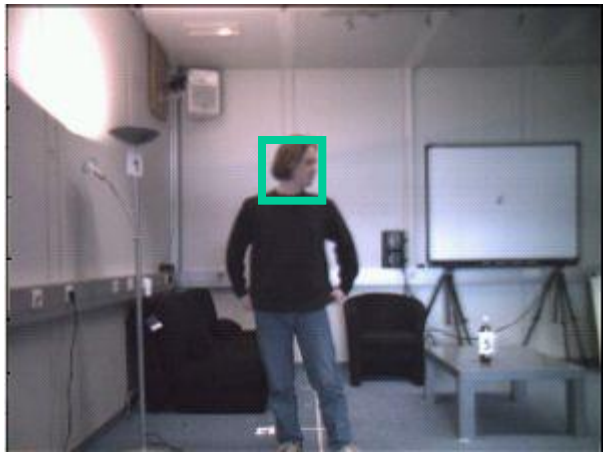
Perceptual Grouping (6)

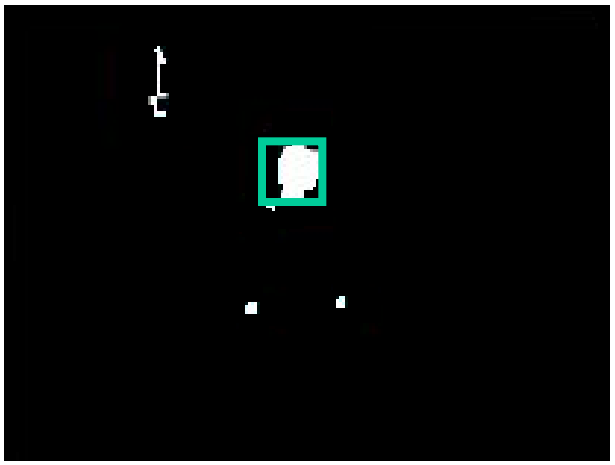


- Apply morphological closing then morphological opening
- Resulting image is reduced to connected regions of skin color (blobs)

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 Face Tracker

Problems with unimodal color based face detection

- Distraction resulting from skin-colored background

- Managing multiple persons

Color Based Face Detection - Summary

- Many color spaces and skin color classifiers can be used
- Bayesian classifier and MLP work very well and are widely used
  - Sufficient training data is needed for modeling the pdf, in particular for Bayesian approach (positive & negative pdfs learned)
- Choice of color space depends on task:
  - All approaches are sensitive to illumination changes (different light sources)
  - Chromatic color space introduces some robustness against changes in illumination intensity
  - Using all colour channels seems to work better in some cases (see Phung et al., 2005)
- Skin-colored background reduces performance
- Detecting multiple persons difficult

Roadmap

- Color based face detection
- **An ellipsoid head model**
- Artificial Neural Networks
- Feature-based classifier cascades (Viola & Jones)

Ellipsoid Head Silhouette Detection (1)

- Motivation:
  - Managing out of plane motion
  - Managing head rotation requires bi-model color modelling (face & hair)
  - Align bounding box around face, no matter if face is partially occluded
  - Head shape can be approximated by an ellipsoid

Ellipsoid Head Silhouette Detection (2)

- Define ellipsoid state $s=(x,y,\sigma)$
  - at position (x,y)
  - size σ (length of minor axis)
  - fixed aspect ratio (e.g. 1.2)

- Head shape is approximated on computed edge image of current frame (e.g. using Sobel operator)

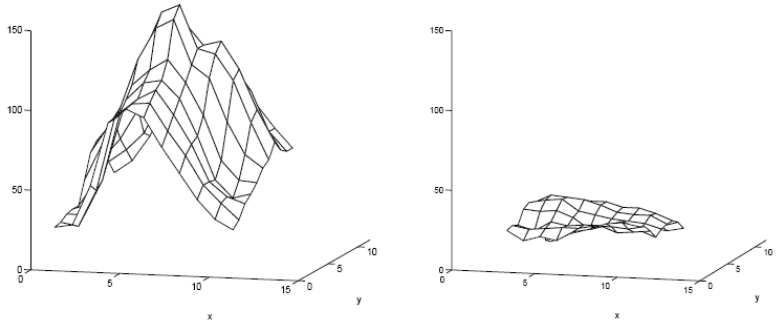
Ellipsoid Head Silhouette Detection (3)

- Search for head candidate s*:
  - Maximize the normalized sum of gradient magnitude around the perimeter of an ellipsoid:

$$s^* = \arg\max_{s \in S} \left\{ \frac{\sum_{i=1}^{N_\sigma} |g_i|}{N_\sigma} \right\}$$

Ellipsoid Head Silhouette Detection (4)

- Problem with background:
  - To stabilize against strong edges of a cluttered background scene, gradients may be thresholded ➔ stronger and weaker gradients are handeled equally. Disadvantage?
  - If background clutter is too high, ellipse tends to drift and gets stuck on background objects
    - Fig. a) shows the normalized sum in an unambiguous environment, Fig. b) depicts how the tracker is distracted from a cluttered background scene)



(a) (b)

Ellipsoid Head Silhouette Detection (5)

- Enhancing maximization criterion to include affiliation of perimeter‘s pixels to a specified color model
  - Color model contains both hair and skin color

$$\phi_c(s) = \frac{\sum_{i=1}^{N} \min(I_S(i), M(i))}{\sum_{i=1}^{N} I_S(i)}$$

- Enhancing criterion to further take care of gradient direction instead of magnitudes only
  - Gradients underneath ellipsoid are to be checked if their directions match ellipsoid‘s shape (check with normal for each pixel on the ellipsoid‘s outline)

$$\phi_g(s) = \frac{1}{N_\sigma} \sum_{i=1}^{N_\sigma} \left| n_\sigma(i) \cdot g_s(i) \right|$$

- New maximization criterion: $s^* = \arg\max_{s \in S} \left\{ \phi_c(s) + \phi_g(s) \right\}$

Ellipsoid Head Silhouette Detection (6)

- Due to shape modelling, the system is more robust against partial occlusions

- As long as the background scene does not provide ellipsoidal gradients, color & gradient score maximization keeps invariant to cluttered environments.

Summary

- Face detection is important
  - Basic building block for further analysis, such as face recognition, facial expressions, head pose, person tracking, etc.
- Face detection is challenging
  - Illumination, rotation, occlusion, …
- Colour based face detection
  - Color-spaces
  - Models: histograms, Gaussian Models, Mixture of Gaussians Model
  - Histogram-backprojection / Histogram matching
  - Bayes classifier
  - Extension: Color + ellipse model
- Advantages: Fast, rotation & scale invariant, robust against occlusions
- Disadvantages: Cannot distinguish head and hands, affected by illumination

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.

Stan Birchfield, *Elliptical Head Tracking Using Intensity Gradients and Color Histograms*, IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, California, June 1998.

Michael J. Swain and Dana H. Ballard, *Color Indexing*, International Journal on Computer Vision, 7:1, 11-32 (1991).