Seminar
„Digitale Signalverarbeitung in Multimedia-Geräten“

SS 2003

Man-Machine-Interface (Video)

Computation Engineering Student

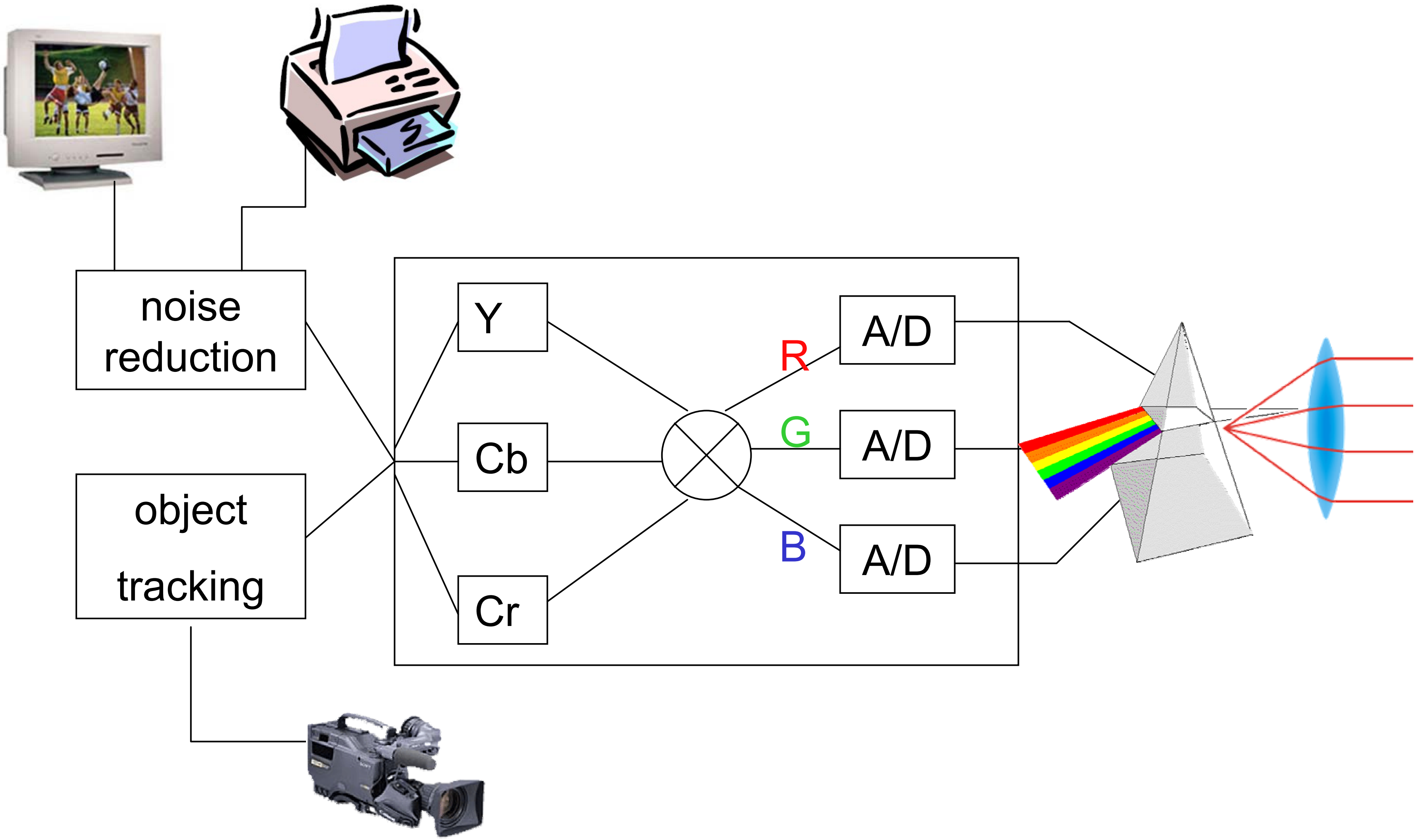
Nataliya Nadtoka

coach: Jens Bialkowski

Outline

1. Processing Scheme
2. Human Visual System
3. Video Representation
    1. Progressive and Interlaced scan
    2. Chrominance subsampling
    3. Color Spaces
4. Object Detection – Face Detection Overview
5. Demonstration

Processing scheme



Human Visual System - Color Perception and representation

Cones

color

low sensitivity

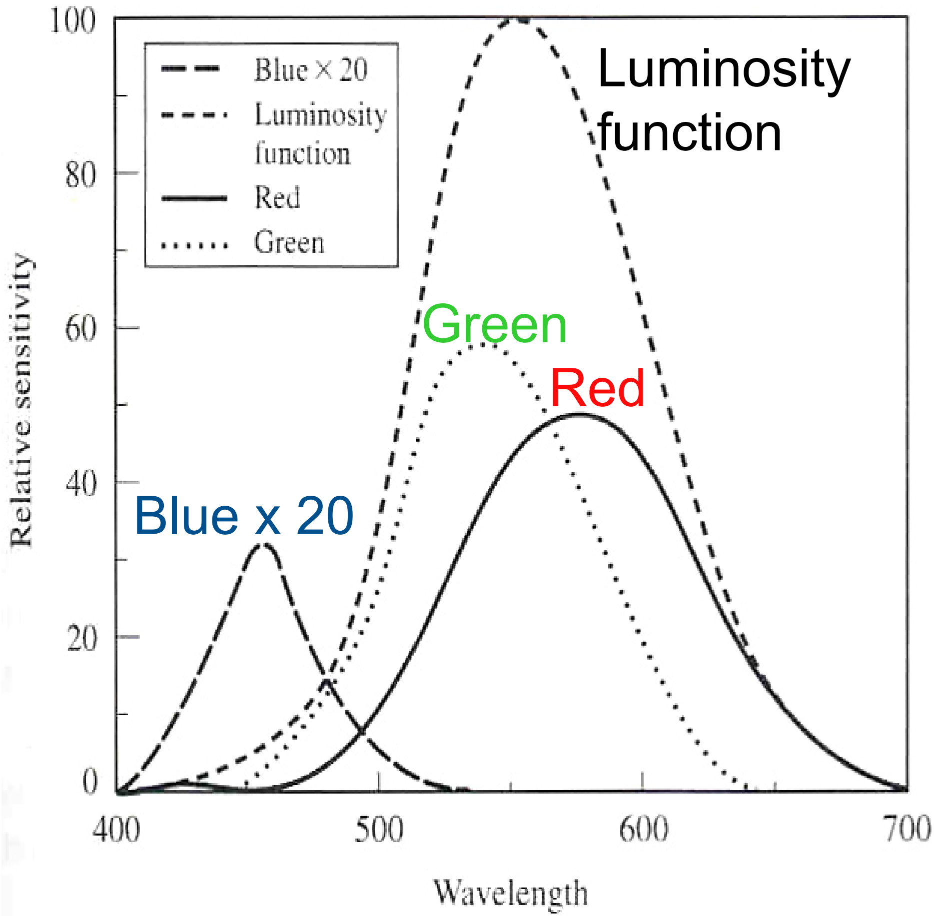
6 Million

Rods

monochrome

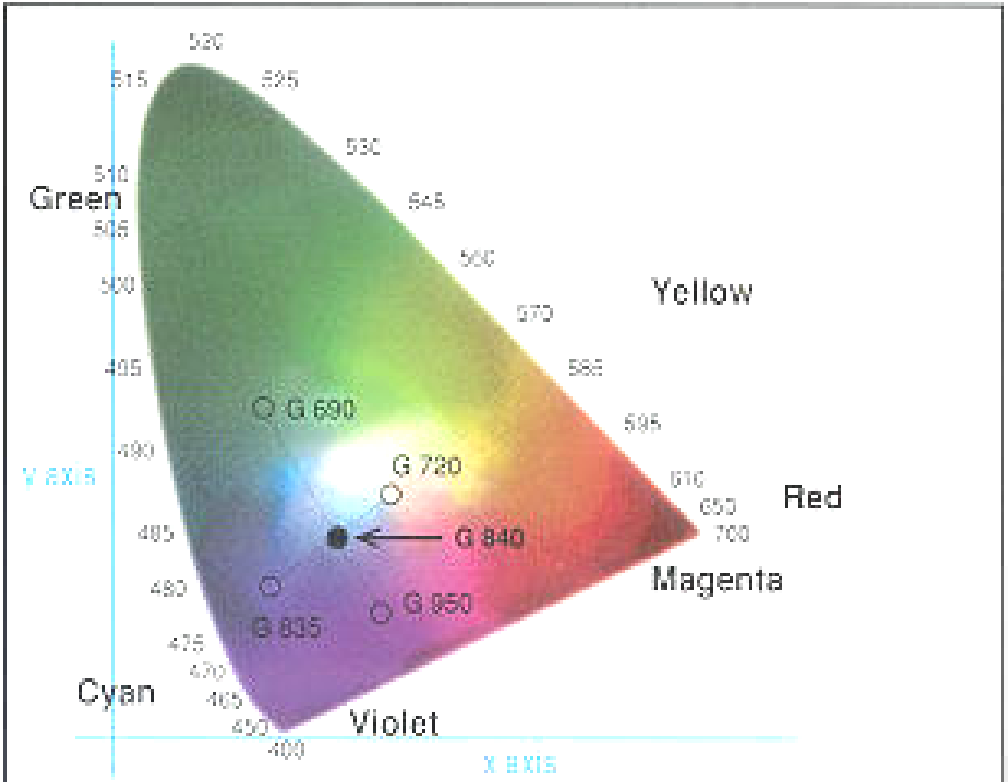
high sensitivity

120 Million



Frequency responses of 3 types of cones in human retina and luminous efficiency function

HVS - Color Gamut



C. I. E. Chromaticity Map Cross reference for G 840

Tristimulus values:

X, Y, Z – intensities of Red, Green and Blue

Chromaticity coordinates:

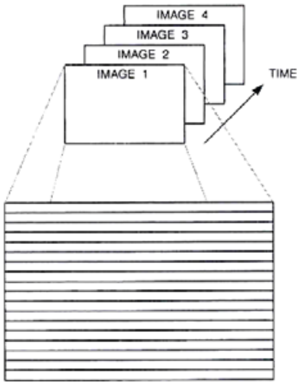
$$x = \frac{X}{X+Y+Z} = \frac{red}{red+green+blue}$$

$$y = \frac{Y}{X+Y+Z} = \frac{green}{red+green+blue}$$

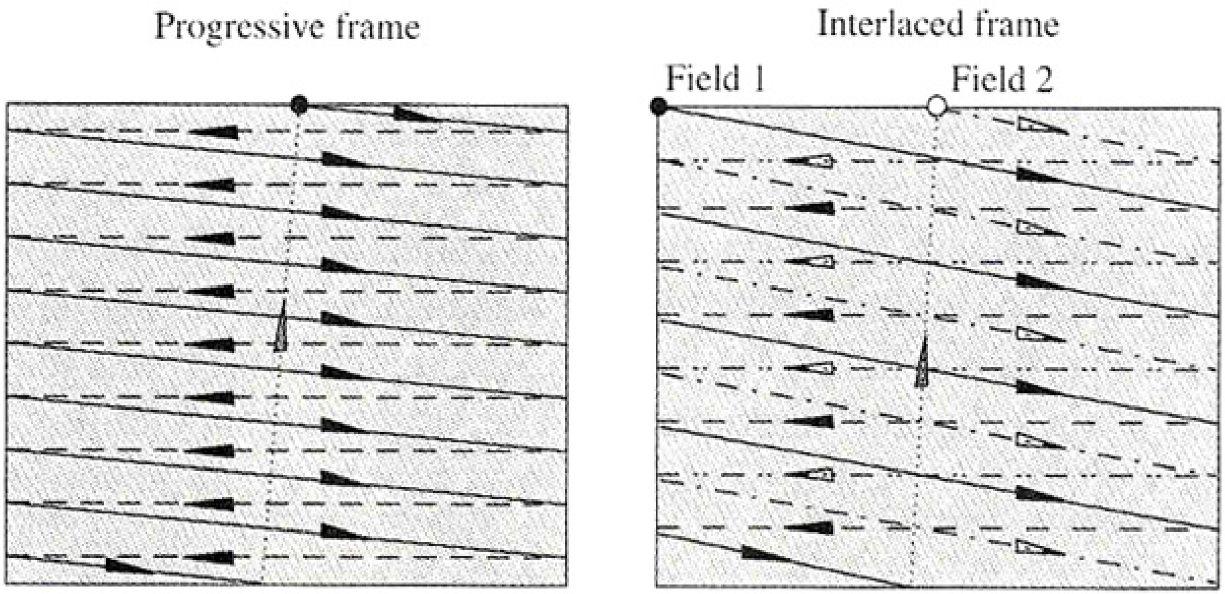
$$z = \frac{Z}{X+Y+Z} = \frac{blue}{red+green+blue}$$

$$x+y+z=1$$

Raster Scan - Progressive and Interlaced Scan



In a raster scan a camera captures a video sequence by sampling it in both temporal and vertical directions.



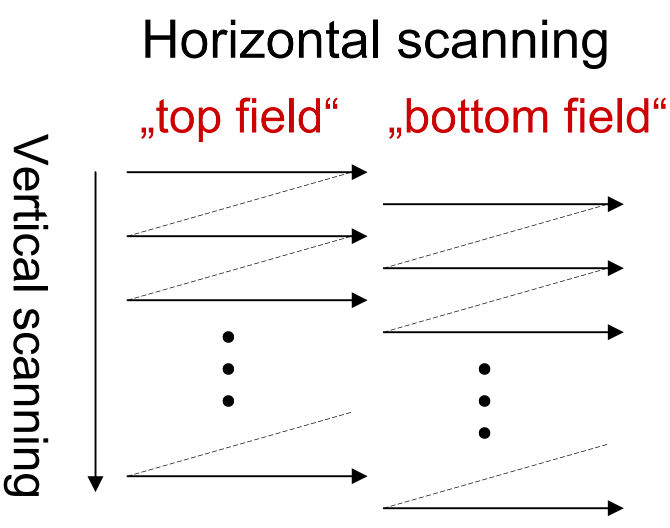
(a) (b)

Electronic or optic beam of an analog video camera continuously scans the imaged region from the top to bottom and then back to the top

Progressive and Interlaced scan (continued)

Progressive scan: horizontal lines are scanned successively

Interlaced scan: each frame is scanned in 2 fields



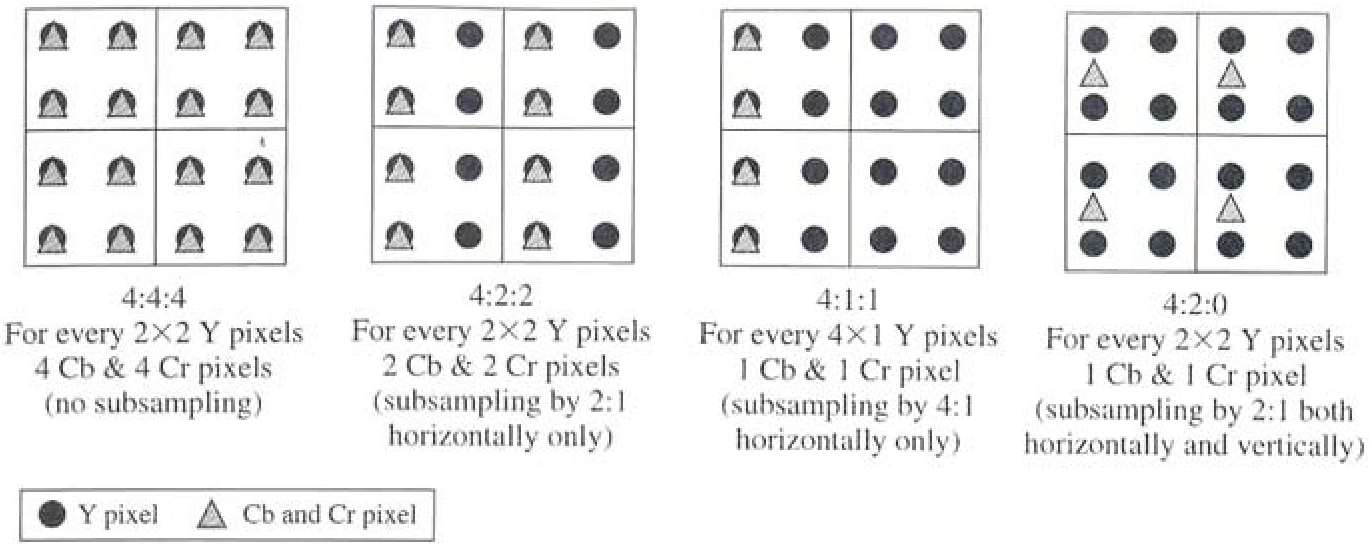
Motivation: trade-off the vertical resolution for an enhanced temporal resolution given the total number of lines that can be recorded within time

Interlaced	Progressive
• each scan line is refreshed half as often	• no limit on the line-to-line changes
• limited line-to-line changes	• high resolution image (vertically)

„Z- effect“ illustration

„Z- effect“ illustration (continued)

Color Coordinates and Chrominance Subsampling



BT.601 chrominance subsampling formats.

Reason: human vision has a higher resolution for luminance than for chrominance components

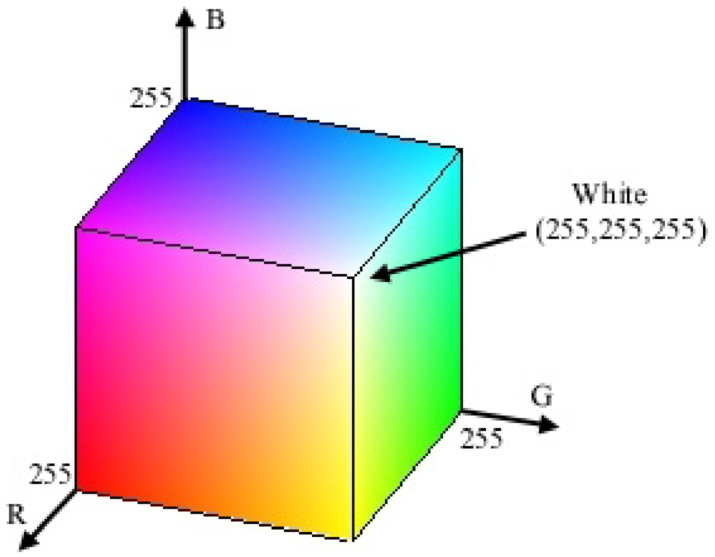
Color Spaces

A Color Space is a mathematical representation of a set of colors.

The most popular color models:

RGB	computer graphics, cameras, scanners
YUV	PAL, NTSC, SECAM (Europe) television
YCbCr	compression in video systems (JPEG, MPEG 1-4)
HSV	„artists work“ postprocessing
CMYK	printers

RGB Color Space



RGB cube

Red, Green and Blue are *primary additive colors*

- used as phosphors by CRTs
- basic colors for computer graphics, digital cameras

Drawbacks:

- equal bandwidth requirements
- high computational effort
- luminance and chrominance

YUV Color Space

- Used in NTSC, PAL TV standards (Europe):
  - black & white systems are supported by Y luminance component
  - color (U and V) was added to display color picture
- conversion works on gamma corrected RGB signal (R`G`B`):

$$Y = 0.299\,R' + 0.587\,G' + 0.114\,B'$$

$$U = -0.147\,R' - 0.289\,G' + 0.436\,B' = 0.492\,(B' - Y)$$

$$V = 0.615\,R' - 0.515\,G' - 0.100\,B' = 0.877\,(R' - Y)$$

YCbCr Color Space

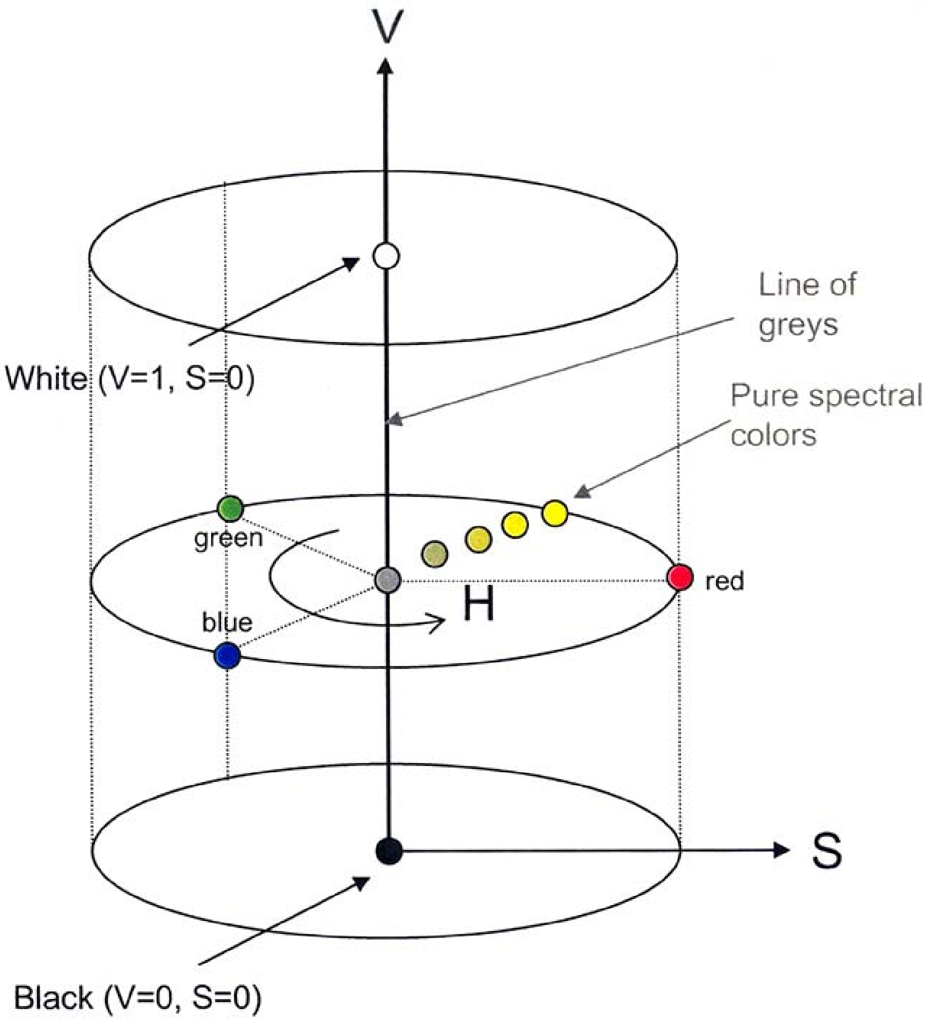
- part of ITU-R BT.601 world wide digital component video standard
- scaled and offset version of YUV(luminance and chrominance are scaled additionally)
- Y [16..235], Cb and Cr [16..240]
- comes in different subsampling formats (4:4:4, 4:2:2, 4:2:0)
- used in compression MPEG 1-4, JPEG

$$Y_{601} = 0.299\,R' + 0.587\,G' + 0.114\,B'$$

$$C_b = -0.172\,R' - 0.339\,G' + 0.511\,B' + 128$$

$$C_r = 0.511\,R' - 0.428\,G' - 0.083\,B' + 128$$

HSV Color Space



H(Hue)

S(Saturation)

V(Value of intensity) „brightness“

non-linear transform from RGB tristimulus to color cylinder

HSV color system defined by C.I.E.(International Commission for Illumination)

Object detection – Face Detection

Goal: make man – machine interface more humane

Research in face processing includes:

- Face Recognition
- Face Tracking
- Pose Estimation
- Expression Recognition
- Gesture Recognition

Face detection

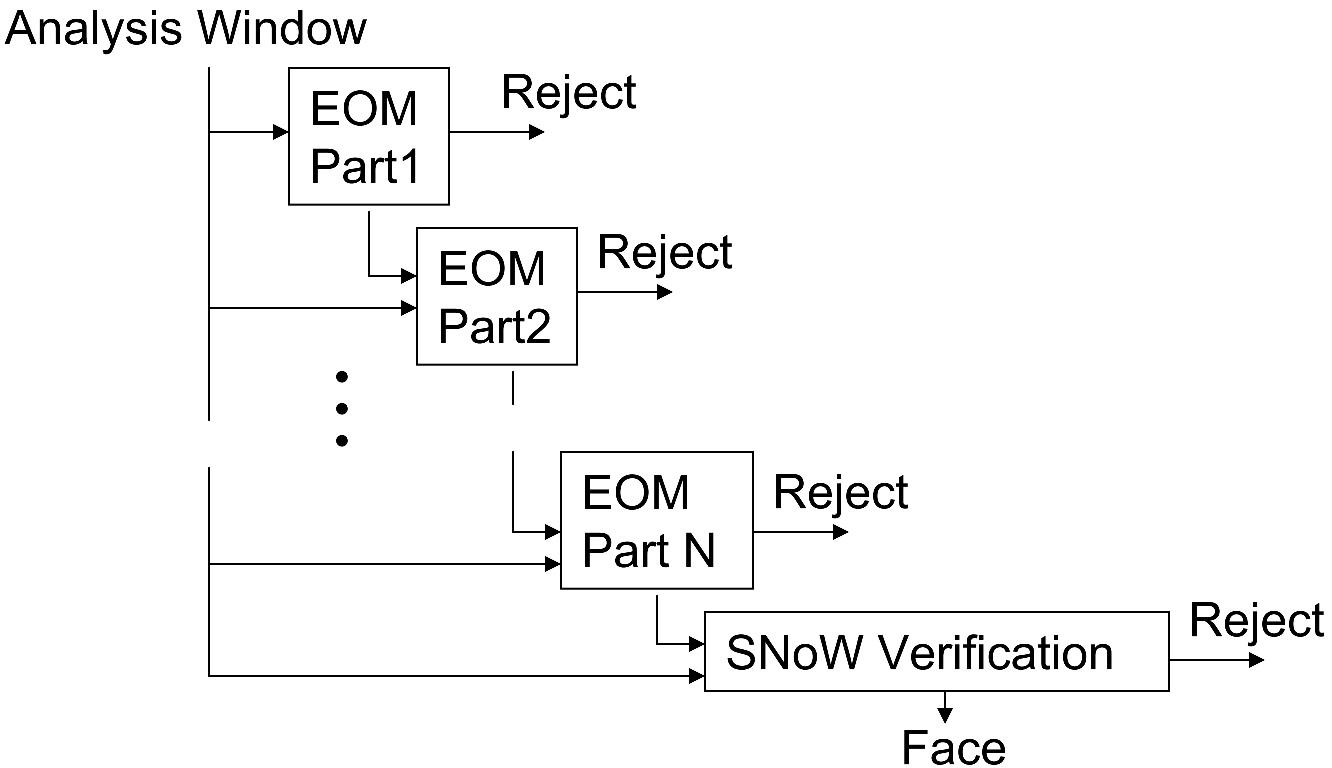
- ⇨ **Given:** a single image or sequence of images
- ⇨ **Goal:** identify all image regions containing face regardless its three dimensional position and orientation and the lighting conditions
- ⇨ **Challenges:**
  - **Pose** (frontal, 45 degree, profile, upside down)
  - **Presence or absence of structural components** (beards, mustaches, glasses)
  - **Facial expression**
  - **Occlusion**
  - **Image orientation**
  - **Imaging Conditions** (lighting, camera characteristics)

Face detection approach

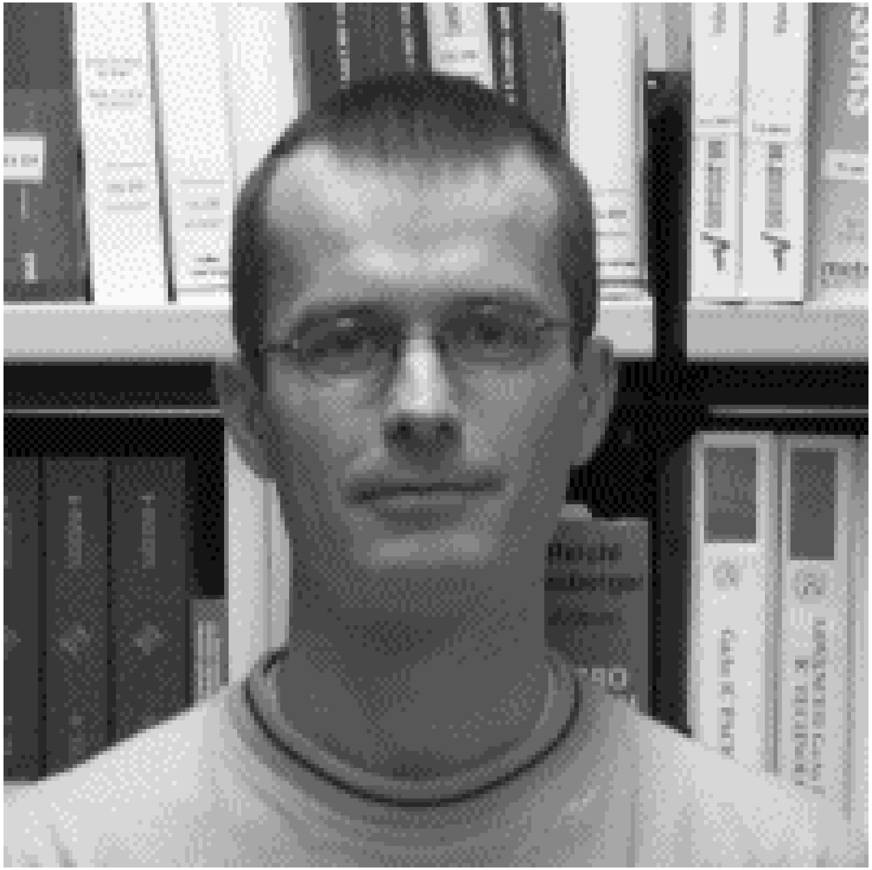
- Knowledge based - multilevel rule-based method with mosaicing
- Feature invariant
  - Facial features — grouping of edges
  - Texture — Space Gray-Level Dependence Matrix of face
  - Skin Color — Mixture of Gaussian
  - Multiple features — Integration of skin color, size and shape
- Template matching — Human defined face templates
- Appearance based method
  - Eigenface — Eigenvector decomposition and clustering
  - Distribution based — Gaussian distribution and multilayer perception
  - Neural Network — Ensemble of neural networks and arbitration
  - Support Vector Machine — Training SVM with RBF kernel
  - Bayesian approach — Naive Bayes Classifier on local appearance
  - Hidden Markow Model — Higher order statistics with HMM
  - Information-Theoretical Approach — Kullback relative information

Face Detection at Video Frame Rate Based on Edge Orientation Features (B. Fröba and C. Küblbeck, Fraunhofer Gesellschaft)

- **Works** with still images and video streams
- Uses a combination of two approaches
  - Edge Orientation Matching
  - Appearance based method (called SNoW)



Demo (Christian Küblbeck, Fraunhofer-Gesellschaft)



Conclusions

- ⇨ Video recording is physical representation (voltage)
- ⇨ Further digital processing with respect to Human Visual System, e.g.:
  - For trading the amount of data, e.g.
    - Progressive vs. Interlaced Scan
    - Color Subsampling
  - For representing Color and Brightness:
    - RGB
    - YUV
    - YCbCr
    - HSV
  - Object detection
  - And more: e.g. quality improvement (e.g. Denoising, Edge Enhancement, …)
- ⇨ Man-Machine-Interface brings together video recording and human visual system

Reference

- ⇨ Neith Jack: " Video Demystified" 3rd Ed.
- ⇨ Wang, Ostermann, Zhang: "Video Processing and Communications", Prentice Hall, 2001
- ⇨ Kaup, Script Multimedia Communications, Uni Erlangen-Nürnberg
- ⇨ Ming-Hsuan Yang, David Kriegman, Narendra Ahuja "Detecting Faces in Images: A Survey", University of Illinois
- ⇨ Bernhard Fröba, Christian Küblbeck, "Robust Face Detection at Video Frame Based on Edge Orientation Features", Fraunhofer Institute for Integrated Circuits
- ⇨ www.iis.fraunhofer.de/bv/