



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

**MIN-Fakultät**  
**Fachbereich Informatik**  
Arbeitsbereich SAV/BV (KOGS)

# Image Processing 1 (IP1)

## Bildverarbeitung 1

Lecture 3 – Image Understanding  
and Image Formation

Winter Semester 2014/15

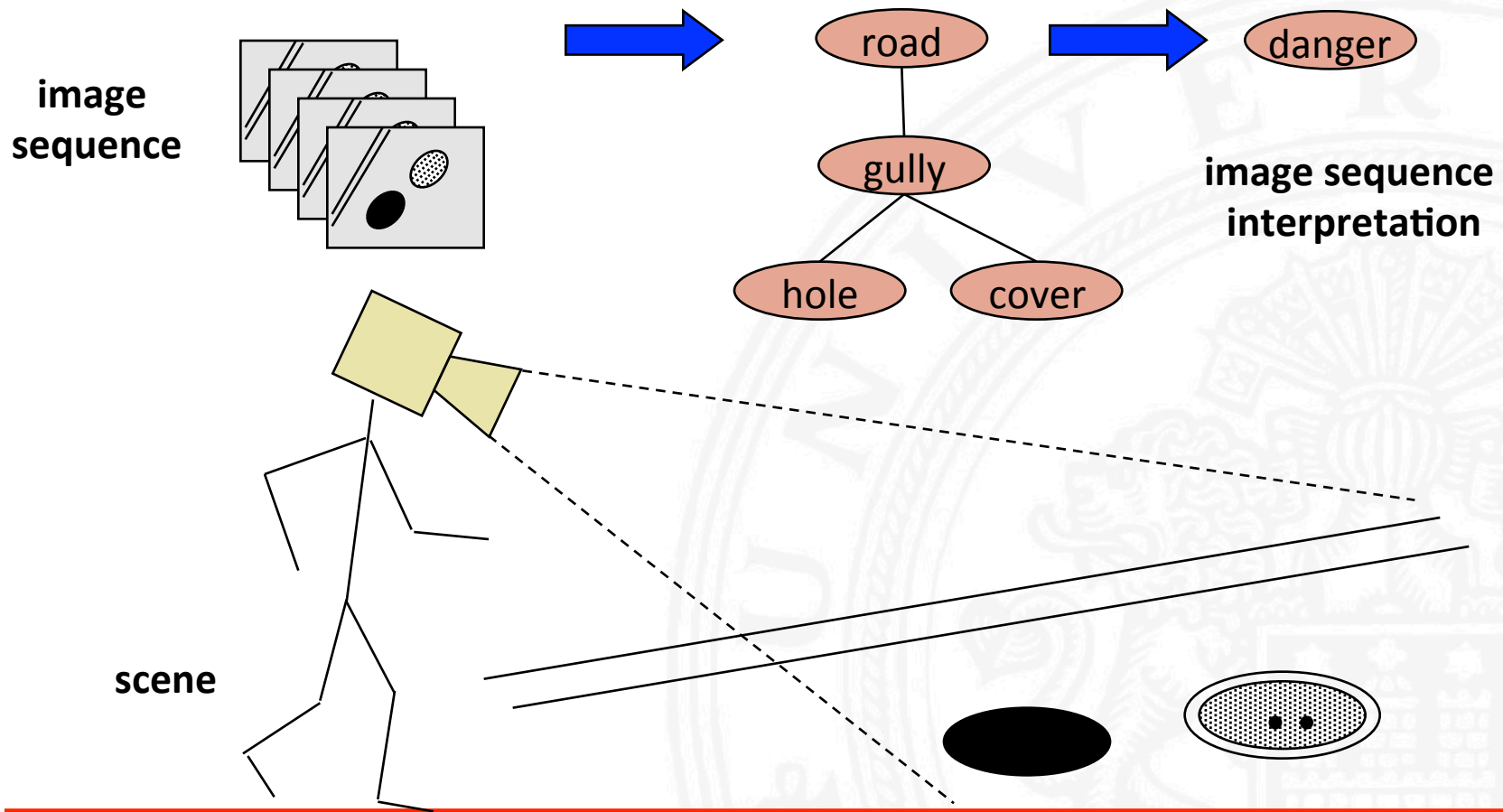
Dr. Benjamin Seppke  
Prof. Siegfried Stiehl

# Definition of Image Understanding

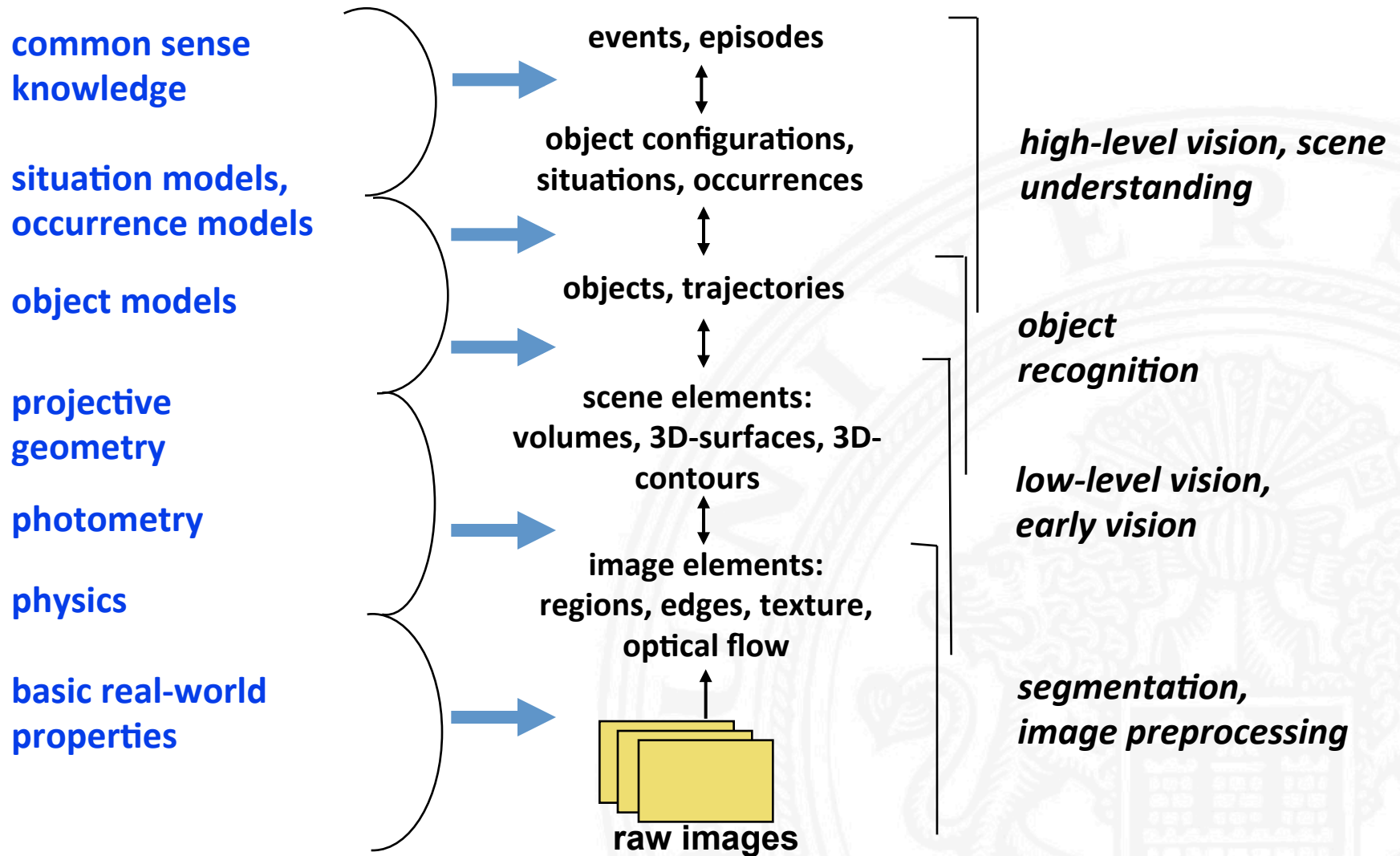
**Image understanding is the task-oriented reconstruction and interpretation of a scene by means of images**

- "scene": section of the real world
  - stationary (3D) or
  - moving (4D)
- "image": view of a scene
  - projection, density image (2D)
  - depth image (2 1/2D)
  - image sequence (3D)
- "reconstruction and interpretation": computer-internal scene description
  - Quantitative
  - Qualitative
  - Symbolic
- "task-oriented": for a purpose, to fulfil a particular task
  - context-dependent,
  - supporting actions of an agent

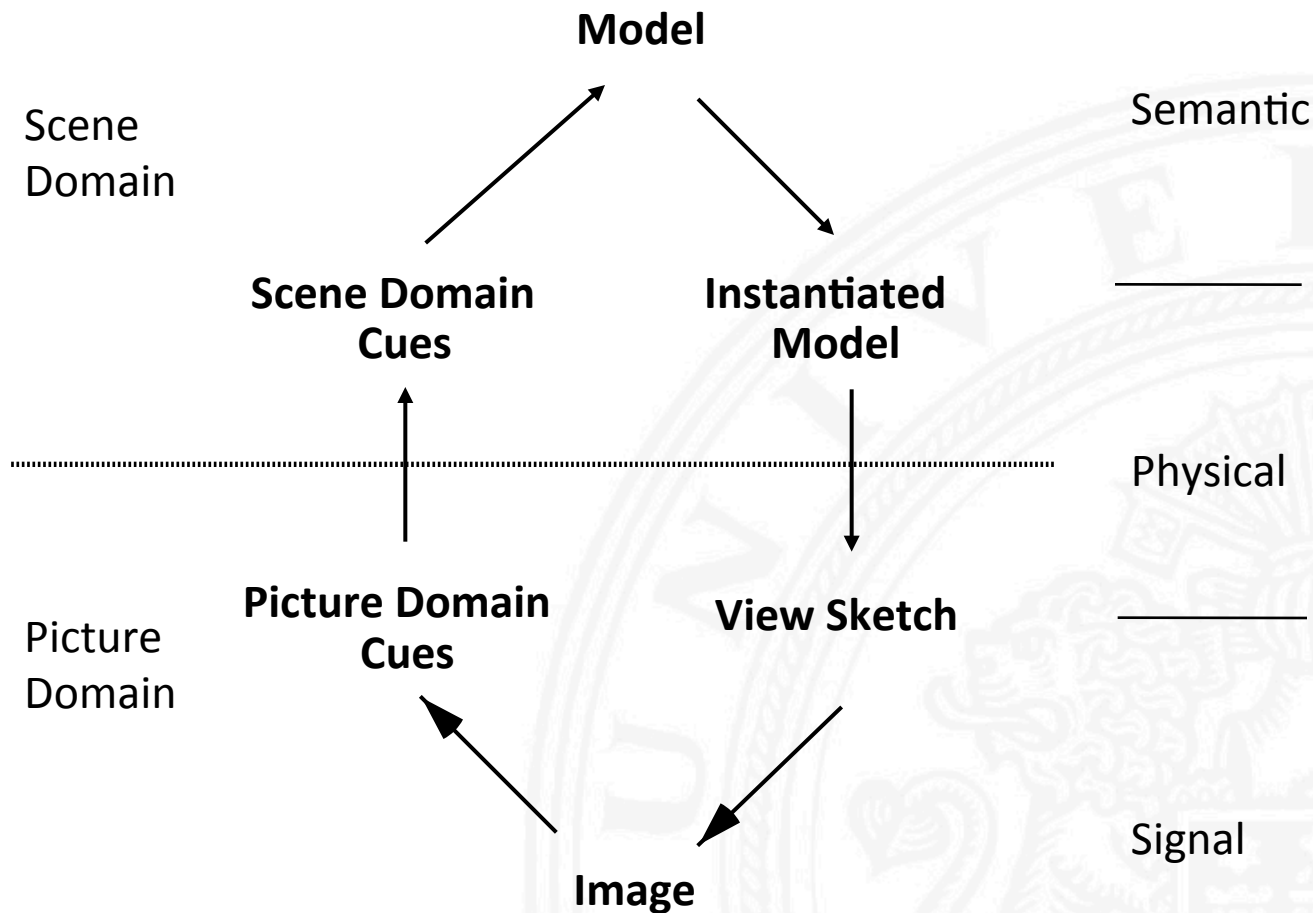
# Illustration of Image Understanding



# Image Understanding as a Knowledge-based Process



# A Model of Scene Analysis (Kanade 78)



# Abstraction Levels for the Description of Computer Vision Systems

## Knowledge level

*What knowledge or information enters a process? What knowledge or information is obtained by a process?*

*What are the laws and constraints which determine the behavior of a process?*

---

## Algorithmic level

*How is the relevant information represented?*

*What algorithms are used to process the information?*

---

## Implementation level

*What programming language is used?*

*What computer hardware is used?*

# Example for Knowledge-level Analysis

What knowledge or information enters a process? What knowledge or information is obtained by a process?

What are the laws and constraints which determine the behavior of a process?

- Consider shape-from-shading:



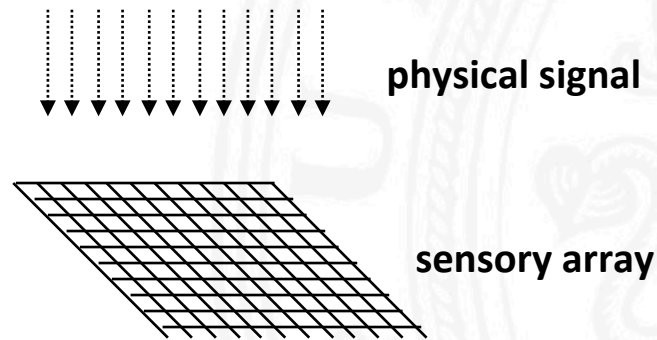
In order to obtain the 3D shape of an object, it is necessary to:

- state what knowledge is available (greyvalues, surface properties, illumination direction, etc.)
- state what information is desired (e.g. qualitative vs. quantitative)
- exploit knowledge about the physics of image formation

# Image Formation

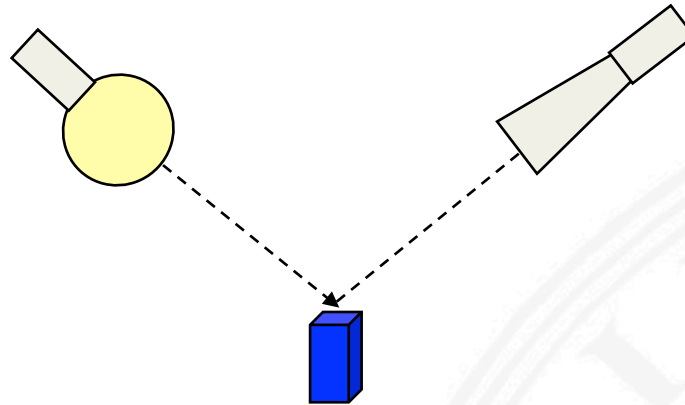
Images can be generated by various processes:

- Illumination of surfaces, measurement of reflections ← **Natural Images**
- Illumination of translucent material, measurement of irradiation
- measurement of heat (infrared) radiation
- X-ray of material, computation of density map
- measurement of any features by means of a sensory array





# Formation of Natural Images



## Intensity (brightness) of a pixel depends on

1. illumination (spectral energy, secondary illumination)
2. object surface properties (reflectivity)
3. sensor properties
4. geometry of light-source, object and sensor constellation (angles, distances)
5. transparency of irradiated medium (mistiness, dustiness)

# Qualitative Surface Properties

When light hits a surface, it may be

- absorbed
  - Reflected
  - scattered
- } in general, all effects may be mixed

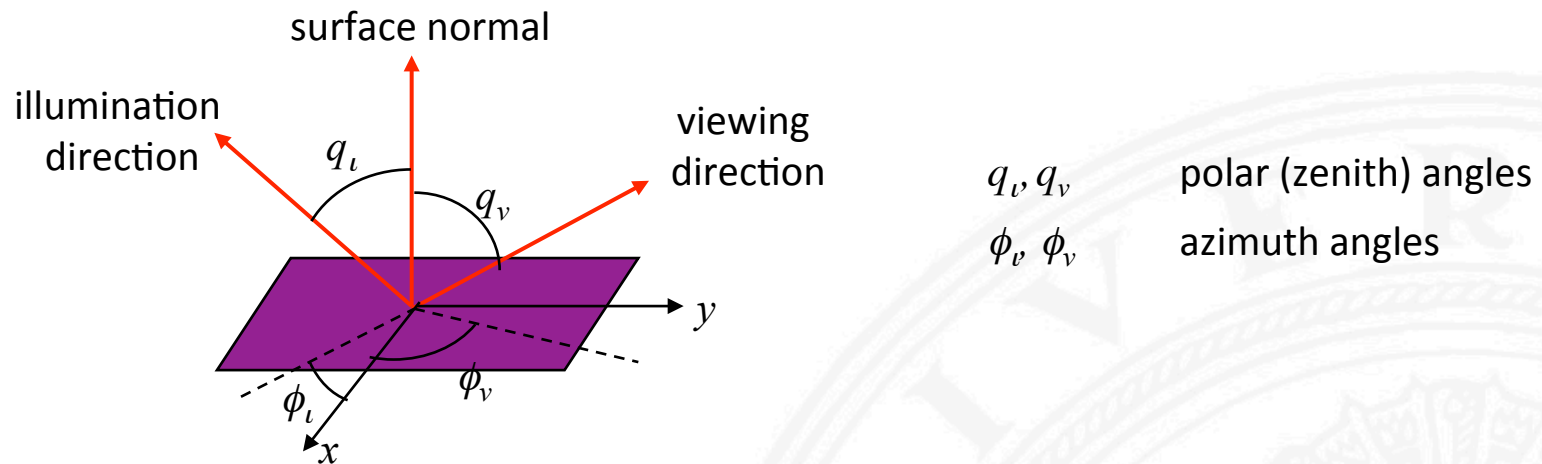
Simplifying assumptions:

- Radiance leaving at a point is due to radiance arriving at this point
- All light leaving the surface at a wavelength is due to light arriving at the same wavelength
- Surface does not generate light internally

The "amount" of reflected light may depend on:

- the "amount" of incoming light
- the angles of the incoming light w.r.t. to the surface orientation
- the angles of the outgoing light w.r.t. to the surface orientation

# Photometric Surface Properties



In general, the ability of a surface to reflect light is given by the Bi-directional Reflectance Distribution Function (BRDF)  $r$ :

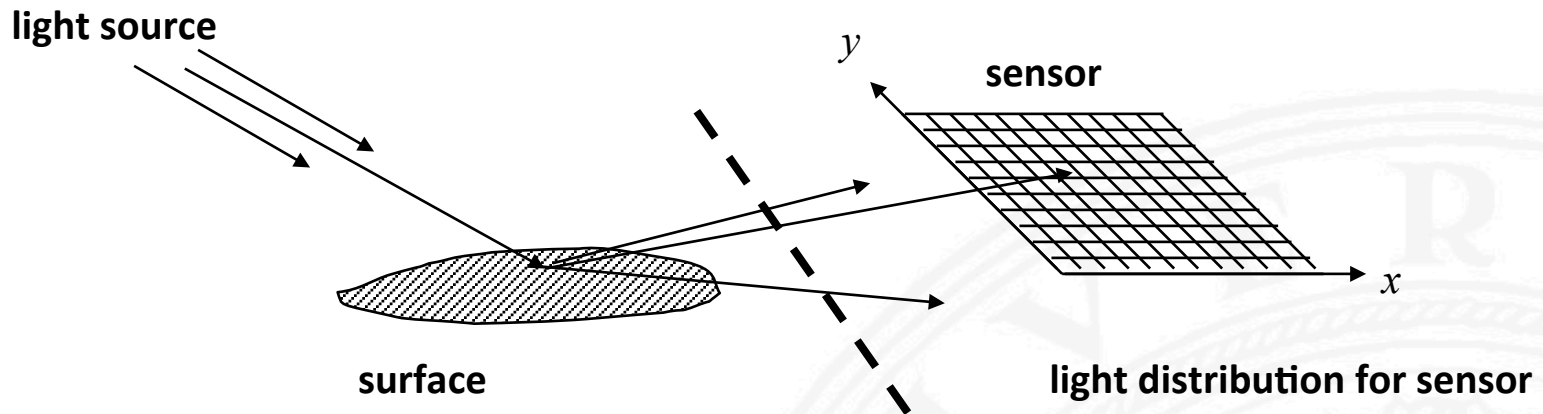
$$r(\theta_l, \phi_l; \theta_v, \phi_v) = \frac{\delta_L(\theta_v, \phi_v)}{\delta_E(\theta_l, \phi_l)}$$

radiance of surface patch towards viewer

irradiance of light source received by the surface patch

For many materials the reflectance properties are rotation invariant, in this case the BRDF depends on  $q_l, q_v, \phi$ , where  $\phi = \phi_l - \phi_v$ .

# Intensity of Sensor Signals



Intensities of sensor signals depend on

- location  $x, y$  on sensor plane
- instance of time  $t$
- frequency of incoming light wave  $\lambda$
- spectral sensitivity of sensor

$$f(x, y, t) = \int_0^{\infty} C(x, y, t, \lambda) S(\lambda) d\lambda$$

sensitivity function of sensor  
spectral energy distribution

# Multispectral Images

- Sensors with  $n$  separate channels of different spectral sensitivities generate multispectral images:

$$f_1(x, y, t) = \int_0^{\infty} C(x, y, t, \lambda) S_1(\lambda) d\lambda$$

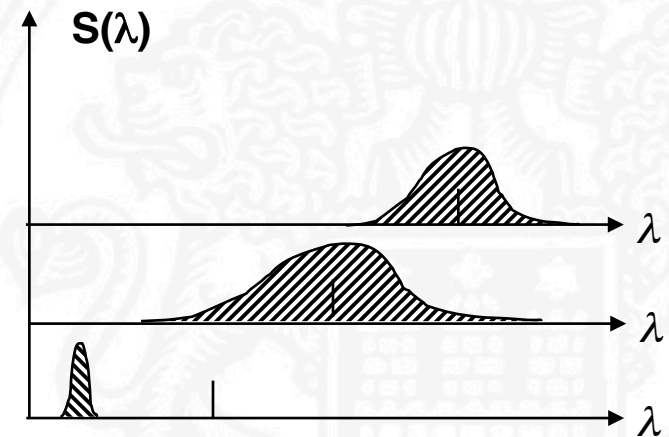
$$f_2(x, y, t) = \int_0^{\infty} C(x, y, t, \lambda) S_2(\lambda) d\lambda$$

$$\vdots$$

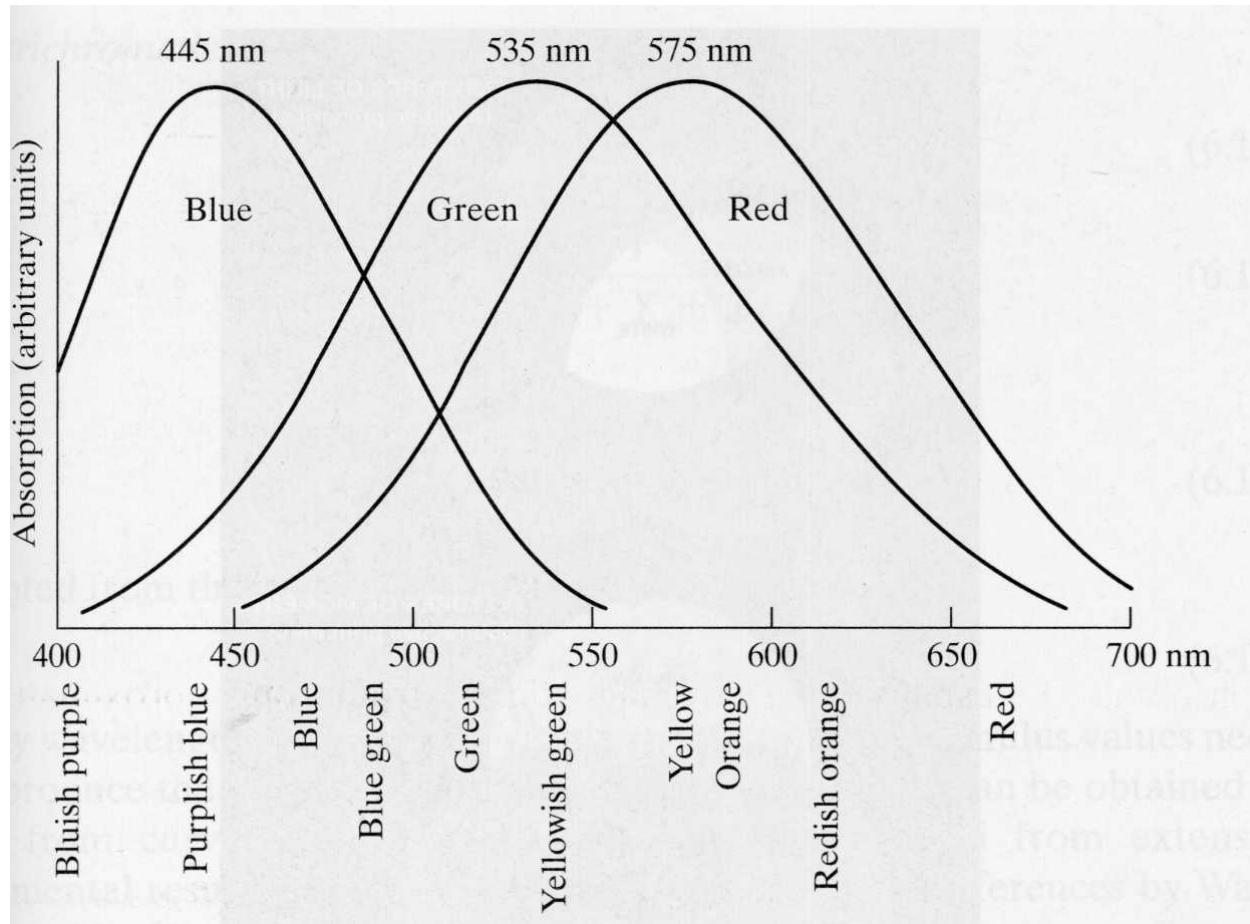
$$f_n(x, y, t) = \int_0^{\infty} C(x, y, t, \lambda) S_n(\lambda) d\lambda$$

- Example:

- R (red)                      650 nm center frequency
- G (green)                    530 nm center frequency
- B (blue)                     410 nm center frequency



# Spectral Sensitivity of Human Eyes



Standardized wavelengths: red = 700 nm, green = 546.1 nm, blue = 435.8 nm

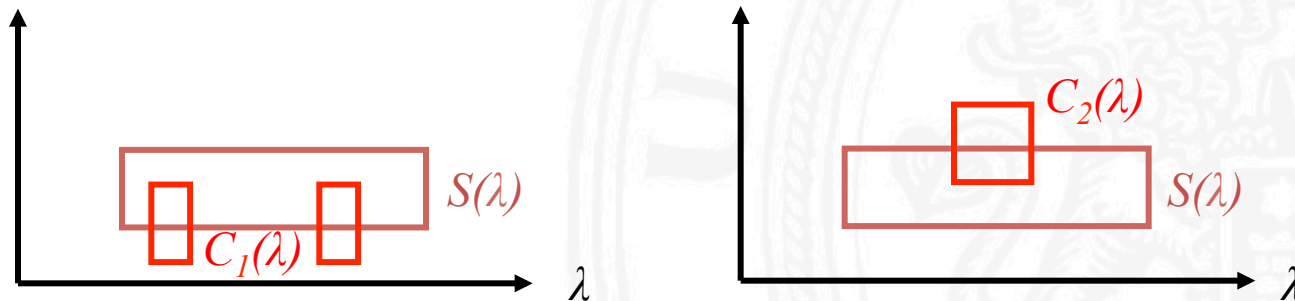
# Non-unique Sensor Response

Different spectral distributions may lead to identical sensor responses and hence cannot be distinguished

$$f_1(x, y, t) = \int_0^{\infty} C_1(x, y, t, \lambda) S(\lambda) d\lambda = \int_0^{\infty} C_2(x, y, t, \lambda) S(\lambda) d\lambda$$

different spectral energy distributions

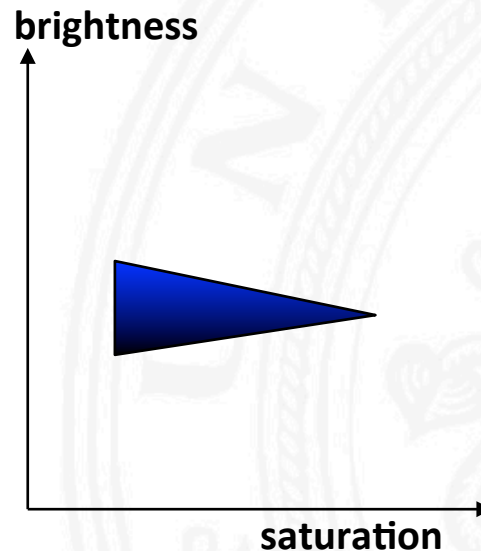
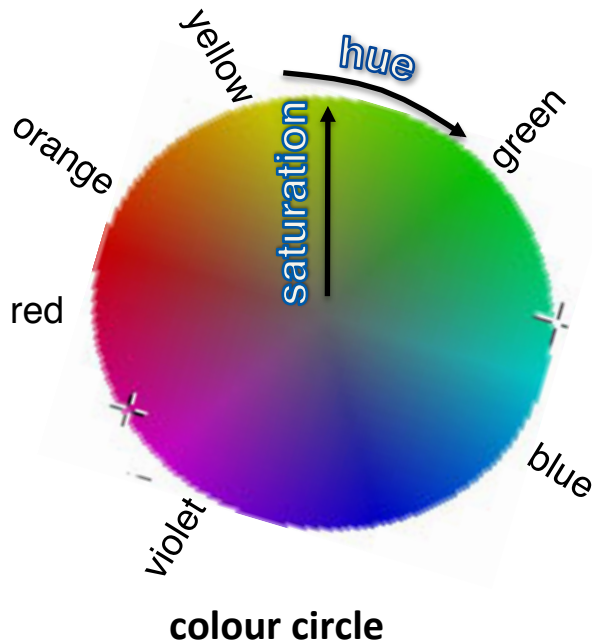
**Example:**



# Dimensions of Colour

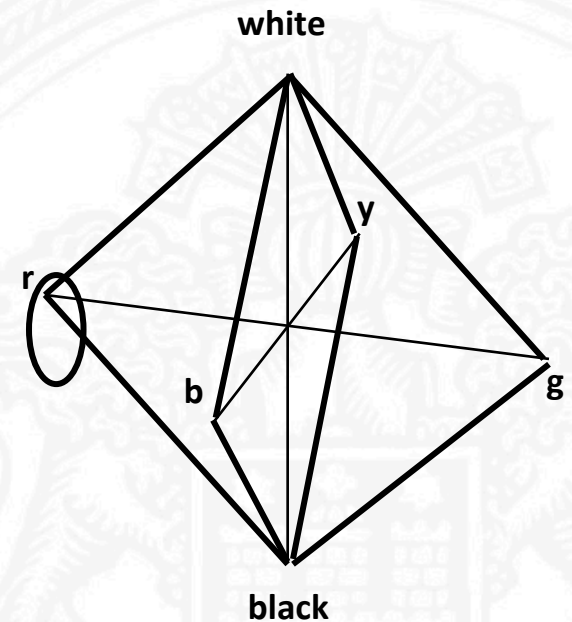
Human perception of colour distinguishes between 3 dimensions:

- Hue
  - Saturation
  - Brightness
- } chromaticity



## NCS\* colour spindle

\* Swedish Natural Colour System

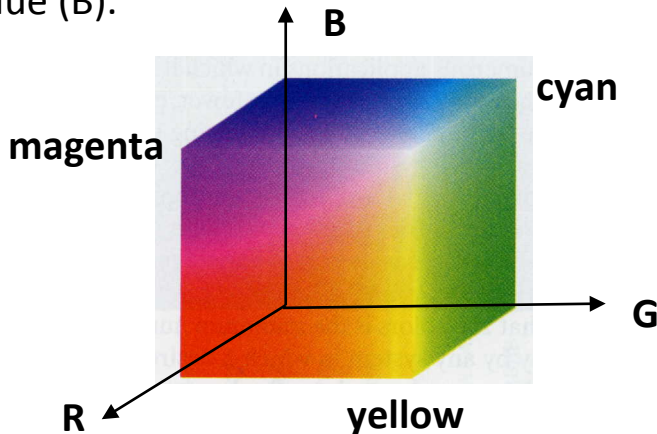




# Computer Vision Colour Models

## RGB colour model

Different colors are generated by adding different portions of red (R), green (G), and blue (B).



RGB is the most commonly used color space in Computer Vision.

### Typical discretization:

8 bits per colour dimension

→ 16.777.216 colours

## HSI colour model

Different colors are described by Hue (H), Saturation (S), and Intensity (I). Can be derived from RGB model:

$$H = \begin{cases} Q & \text{if } B \leq G \\ 360 - Q & \text{if } B > G \end{cases}$$

$$Q = \arccos \left( \frac{\frac{1}{2}[(R-G) + (R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right)$$

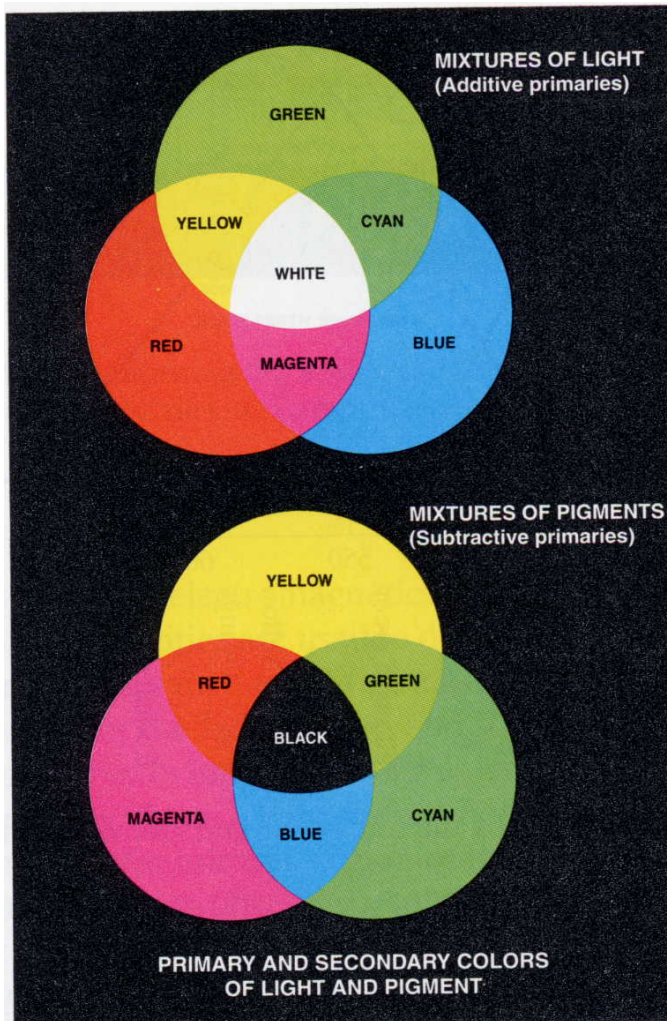
$$S = 1 - \frac{3}{R+G+B} \min(R, G, B)$$

$$I = \frac{R+G+B}{3}$$

Closer to human perception

Better choice e.g. for selecting colors!

# Primary and Secondary Colours



## Primary colours:

red, green, blue

## Secondary colours:

magenta = red + blue

cyan = green + blue

yellow = red + green

from: **Gonzales & Woods**  
**Digital Image Processing**  
**Prentice-Hall 2002**

# RGB Images of a Natural Scene

Here, single colour images are rendered as greyvalue intensity images:  
stronger spectral intensity = more brightness

**R+G+B**



**R**



**G**



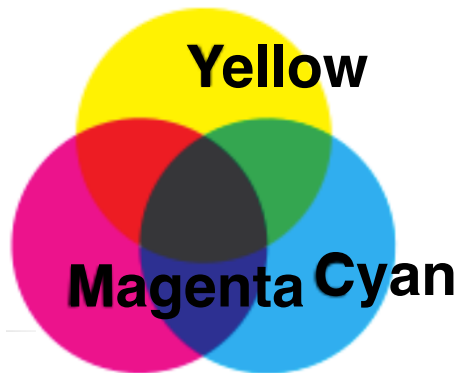
**B**



# Printing Color Models

Subtractive color model because of white background!

CMY colour model



$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

CMYK colour model

A combination of C, M, and Y typically cannot produce a clear, dark black. Therefore a fourth black ink (K for 'key' or blac'k') is used in addition.



# Discretization of Images

Image functions must be discretized for computer processing:

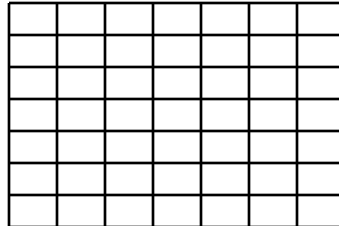
- spatial quantization  
the image plane is represented by a 2D array of picture cells
- greyvalue quantization  
each greyvalue is taken from a discrete value range
- temporal quantization  
greyvalues are taken at discrete time intervals

$$f(x, y, t) \Rightarrow \left\{ \begin{array}{l} f_s(x_1, y_1, t_1), f_s(x_2, y_2, t_1), f_s(x_3, y_3, t_1), \dots \\ f_s(x_1, y_1, t_2), f_s(x_2, y_2, t_2), f_s(x_3, y_3, t_2), \dots \\ f_s(x_1, y_1, t_3), f_s(x_2, y_2, t_3), f_s(x_3, y_3, t_3), \dots \end{array} \right\}$$

**A single value of the discretized image function is called a pixel (picture element).**

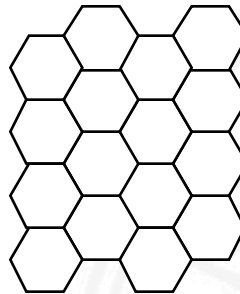
# Spatial Quantization

## Rectangular grid



Greyvalues represent the quantized value of the signal power falling into a grid cell.

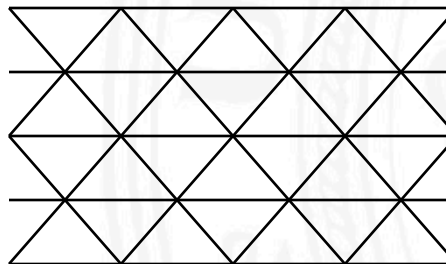
## Hexagonal grid



Note that samples of a hexagonal grid are equally spaced along rows, with successive rows shifted by half a sampling interval.

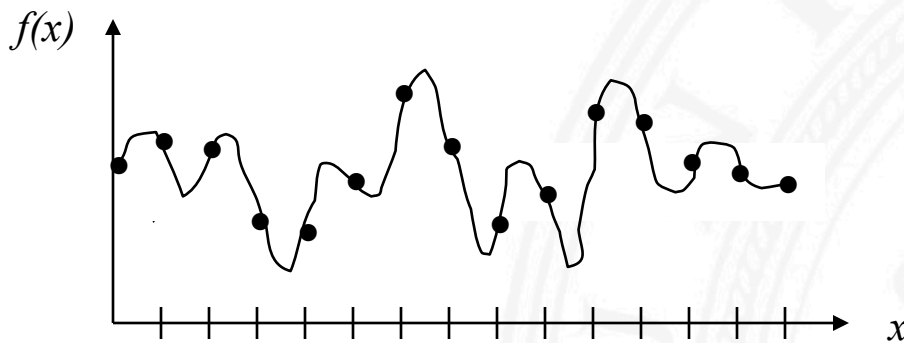


## Triangular grid



# Reconstruction from Samples

- Under what conditions can the original (continuous) signal be reconstructed from its sampled version?
- Consider a 1-dimensional function  $f(x)$ :



- Reconstruction is only possible, if "variability" of function is restricted.

# Sampling Theorem

Shannon's Sampling Theorem:

A bandlimited function with bandwidth  $W$  can be exactly reconstructed from equally spaced samples, if the sampling distance is not larger than  $\frac{1}{2W}$ .

Bandwidth = largest frequency contained in signal  
(=> Fourier decomposition of a signal)

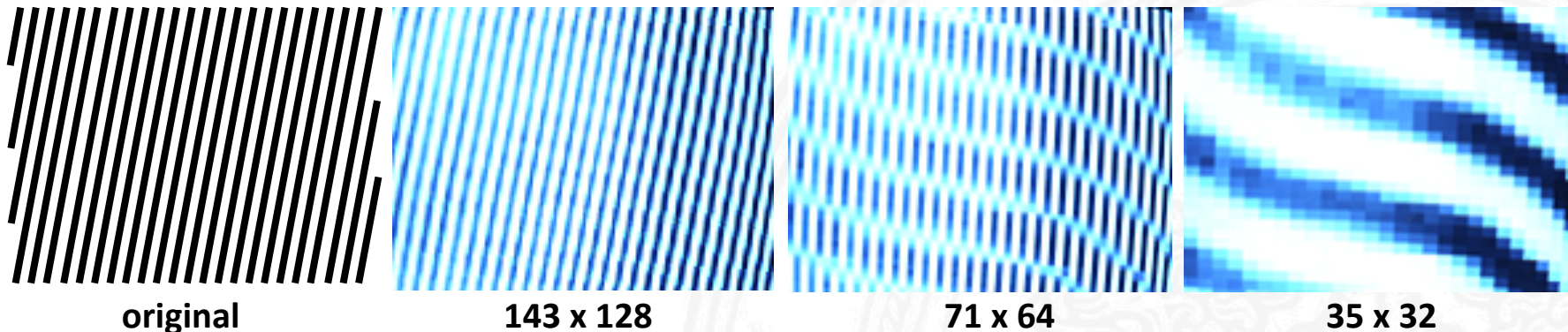
Analogous theorem holds for 2D signals with limited spatial frequencies  $W_x$  and  $W_y$



# Aliasing

Sampling an image with fewer samples than required by the sampling theorem may cause "aliasing" (artificial structures).

Example:



To avoid aliasing, bandwidth of image must be reduced prior to sampling ( $\Rightarrow$  low-pass filtering)

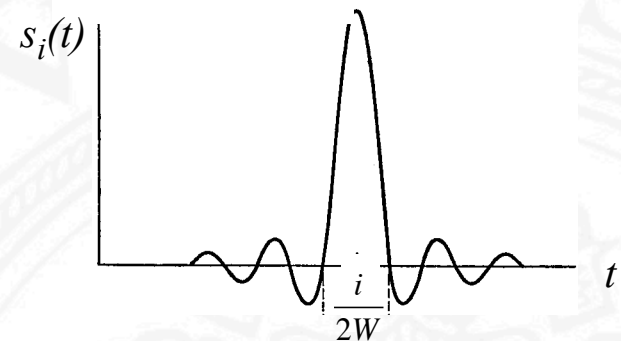
# Reconstructing the Image Function from Samples

Formally, a continuous function  $f(t)$  with bandwidth  $W$  can be exactly reconstructed using sampling functions  $s_i(t)$ :

$$s_i(t) = \sqrt{2W} \frac{\sin 2\pi W [t - i / (2W)]}{2\pi W [t - i / (2W)]}$$

$$x(t) = \sum_{i=-\infty}^{\infty} \sqrt{\frac{1}{2W}} \underbrace{\left( \frac{i}{2W} \right)}_{\text{sample values}} S_i(t)$$

sample values



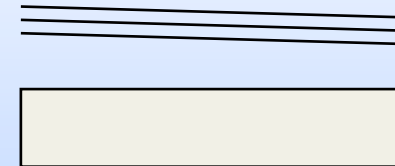
An analogous equation holds for 2D.

In practice, image functions are generated from samples by interpolation.

# Sampling TV Signals

## PAL standard:

- picture format 3 : 4
- 25 full frames (50 half frames) per second
- interlaced rows: 1, 3, 5, ... , 2, 4, 6, ...
- 625 rows per full frame, 576 visible
- 64 ms per row, 52 ms visible
- 5 MHz bandwidth



Only 1D sampling is required because of fixed row structure.

Sampling intervals of  $Dt = 1/(2W) = 10^{-7}s = 100 \text{ ns}$  give maximal possible resolution.

With  $Dt = 100 \text{ ns}$ , a row of duration 52 ms gives rise to 520 samples.

In practice, one often chooses 512 pixels per TV row.

→  $576 \times 512 = 294912$  pixels per full frame

→ rectangular pixel size with  $\text{width/height} = \frac{\left(\frac{4}{512}\right)}{\left(\frac{3}{576}\right)} = 1.5$

