Content-based Image and Video Retrieval

Fall 2012/2013

Visual Descriptors

02.10.2012

cv:hci
Computer Vision for Human-Computer Interaction
Research Group
Universität Karlsruhe (TH)
Research University · founded 1825
Fakultät für Informatik

Visual Descriptors

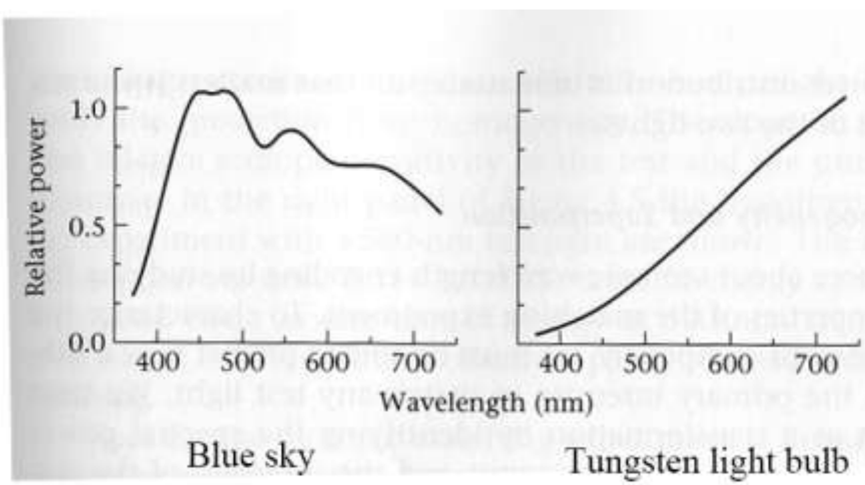
- Descriptors for image retrieval should be
  - Discriminative
  - Robust against image transformations
  - Robust against object transformations, viewpoint and occlusion
  - Efficient to compute

Visual Descriptors

- Color Descriptors
- Texture Descriptors
- Local Descriptors
  - Bag-of-Words

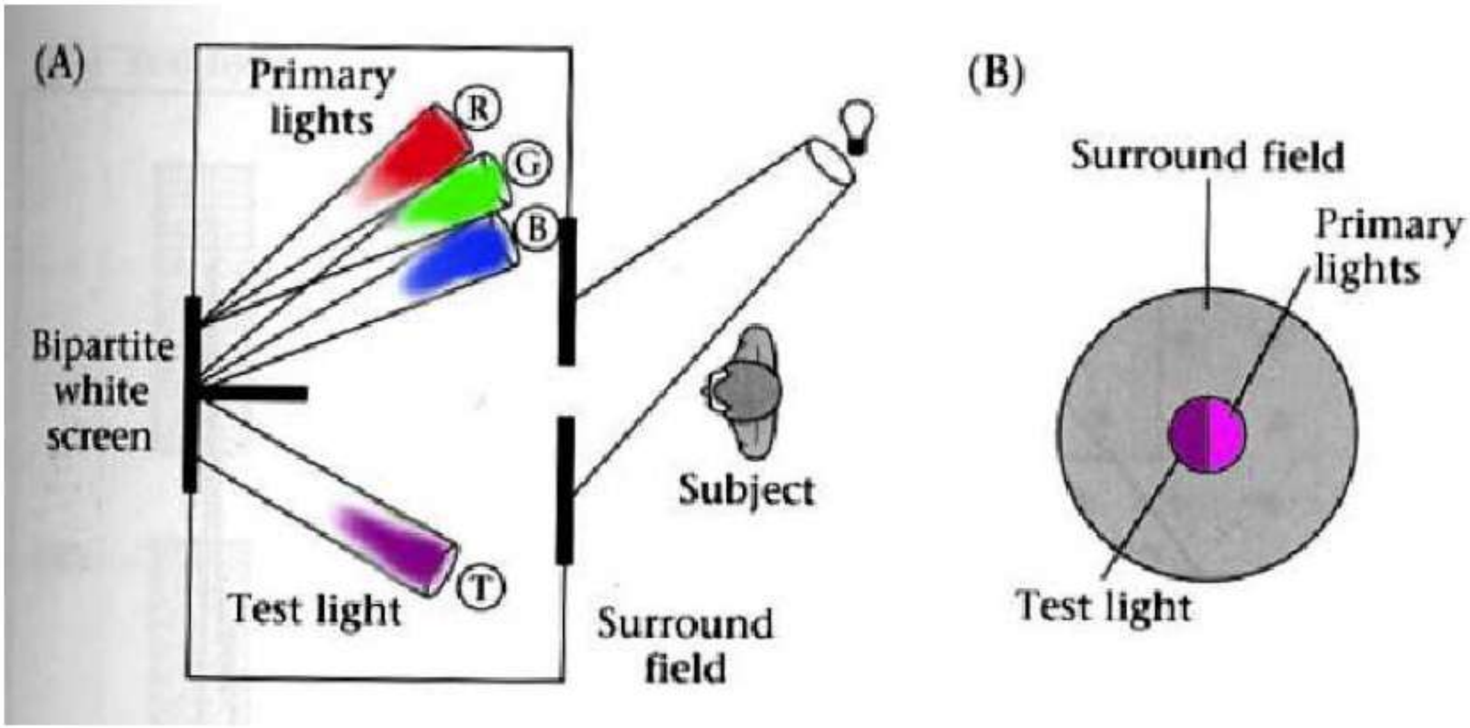
What is color

- A perceptual attribute of objects and scenes constructed by the visual system
- A quantity related to the wavelength of light in the visible spectrum



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

Color Matching Process



- Basis for industrial standards

Color Matching

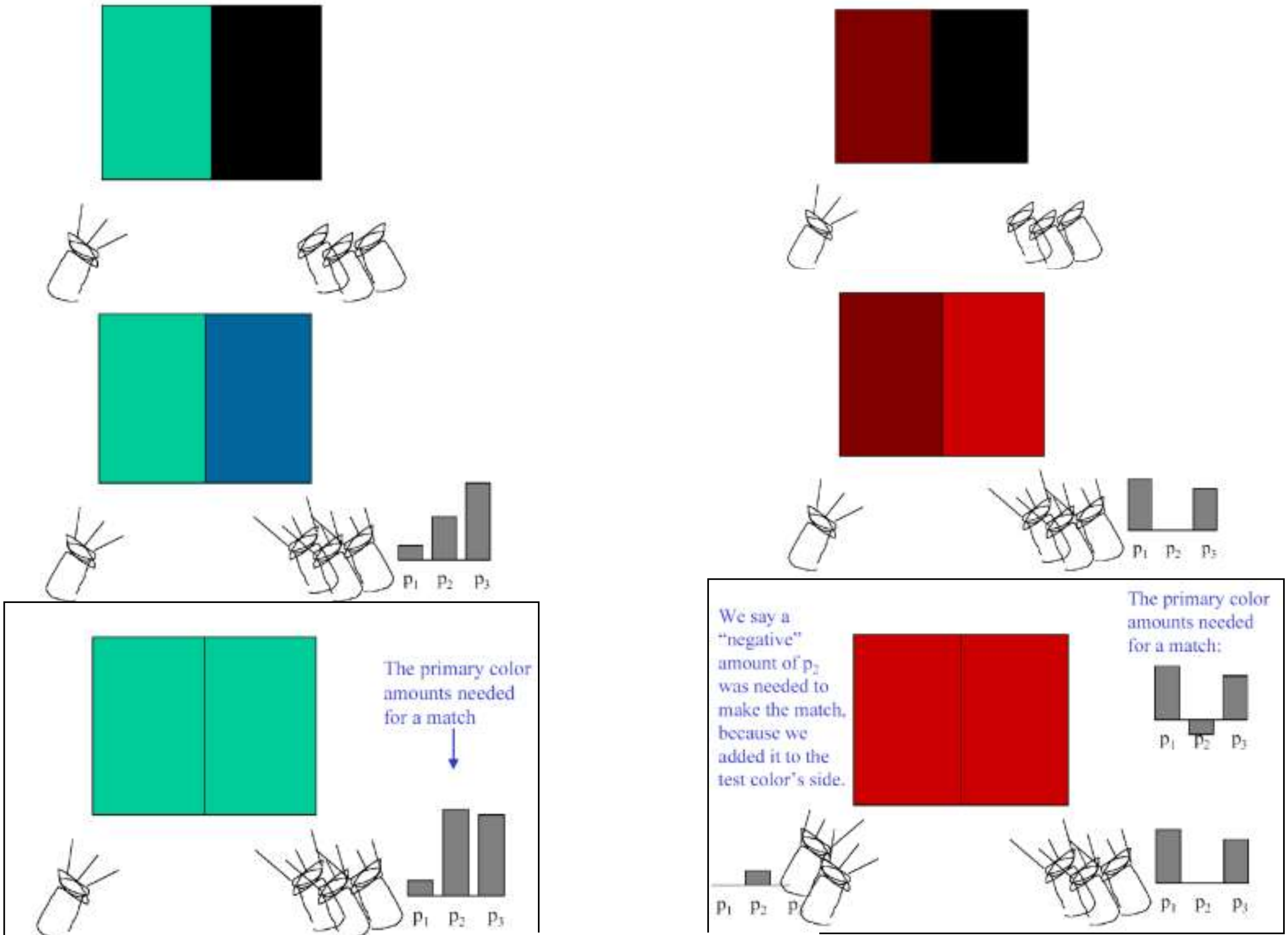


Image courtesy Bill Freeman

Conclusion from Color Matching

- Three primaries are sufficient for most people to reproduce arbitrary colors
  - The human eye normally contains only three types of color receptors, called cone cells
  - Each color receptor responds to different ranges of the color spectrum
  - Humans respond to the light stimulus via a three dimensional sensation, which generally can be modeled as a mixture of three primary colors

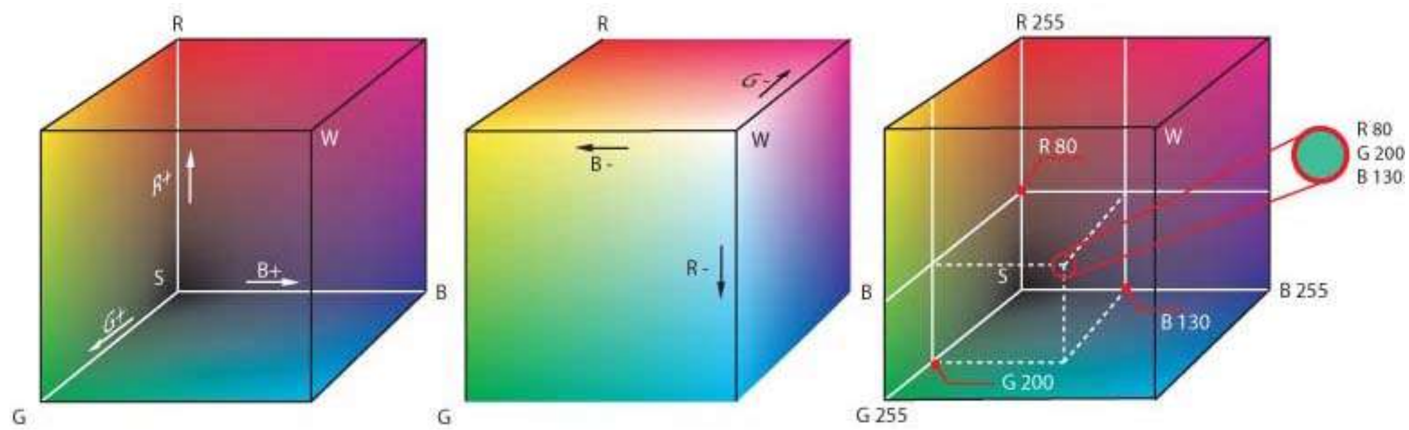
Color Models

- Different ways of parameterizing 3D color space, e.g.
  - RGB
  - XYZ
  - CMY
  - HSV
  - …

RGB color model

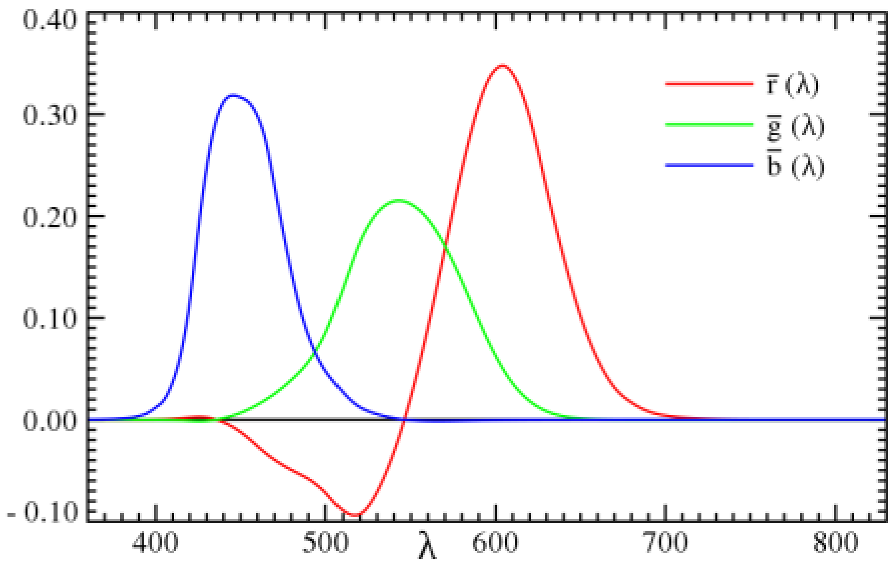
- Official standard:
  - R=645nm, G=526nm, B=444nm

- RGB Color Cube



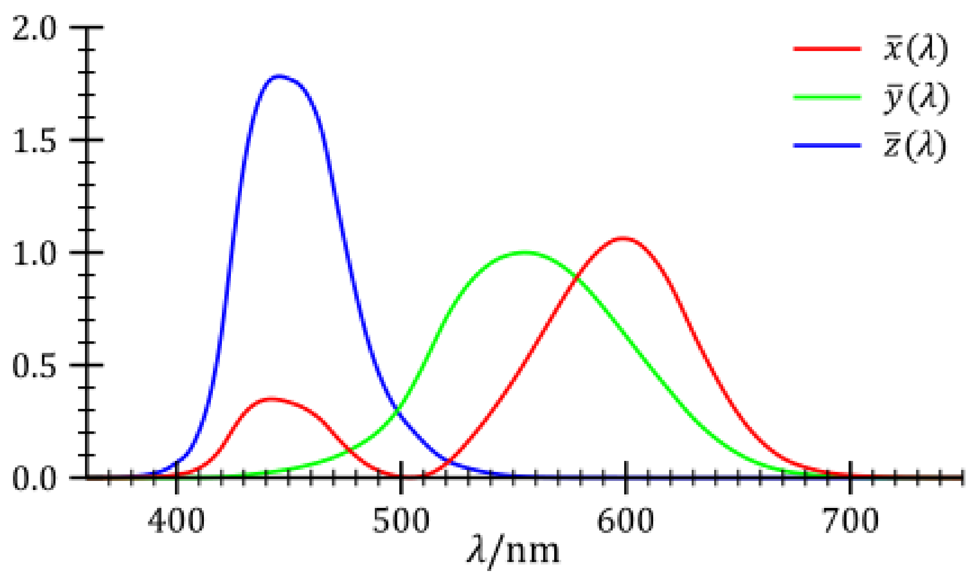
RGB Spectral Colors

- Amounts of RGB primaries needed to display spectral colors



XYZ Color Model (CIE)

- XYZ colorspace is a linear transform of RGB so that all pure wavelengths have positive values
- XYZ spectral colors



CIE: Commission Internationale de l'Eclairage

XYZ and RGB

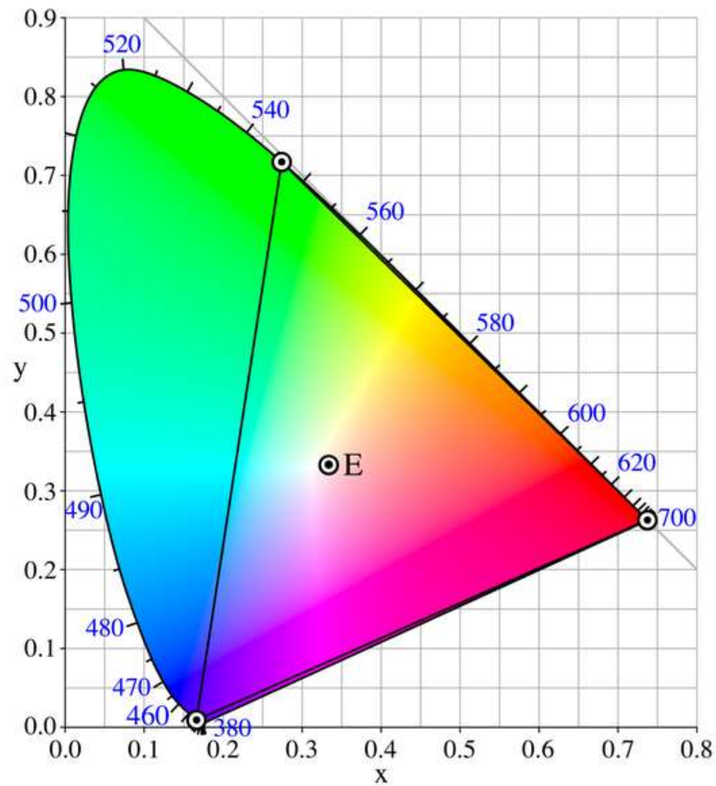
- Linear transformation reparameterizes color space

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.17697} \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$x = \frac{X}{X+Y+Z}$$

$$y = \frac{Y}{X+Y+Z}$$

$$z = 1 - x - y$$



CIE *xy* chromaticity diagram

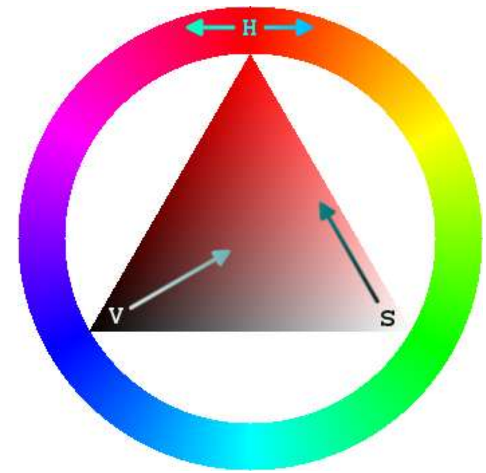
HSV Color Space

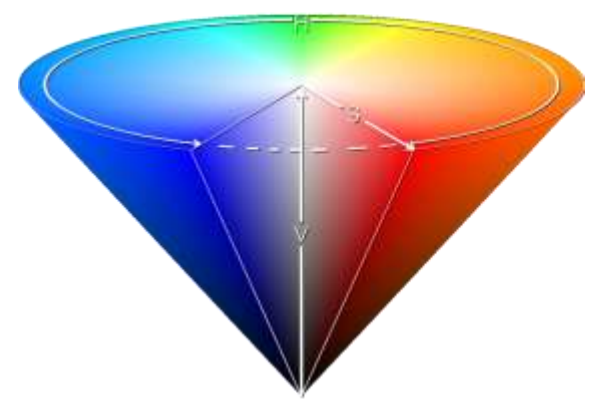
- H(Hue), S(Saturation), V(Value)
  - Closely related to human perception (hue, colorfulness and brightness)

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

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

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





Channels in RGB and HSV

red

green

blue

hue

saturation

Intensity (value)

Chromatic Color Spaces

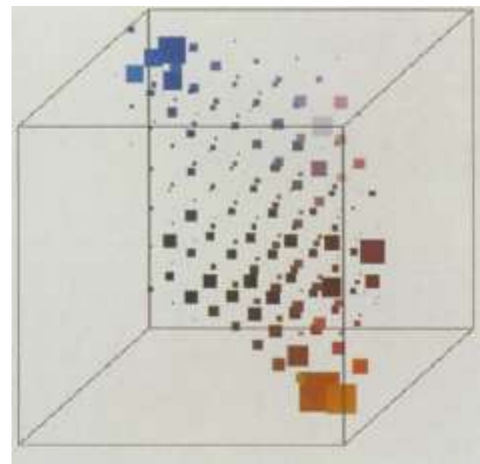
- Two color channels containing chrominance (color) information
  - HS (take from HSV)
  - Normalized rg from RGB
    - r = R / (R+G+B)
    - g = G / (R+G+B)
    - b = B / (R+G+B)
- Motivation: sometimes it is argued that chromatic color models are more robust against illumination variations such as highlighting, shade, shadow, etc.

Color Histogram

- A color histogram represents the distribution of colors where each histogram bin corresponds to a color in the quantized color space
- Color histogram as feature descriptor
  - Color space selection
  - Color space quantization
  - Histogram computation
  - Histogram distance metrics

Computer Vision for Human-Computer Interaction
Research Group - Universität Karlsruhe (TH)

Color Histogram



Pros and Cons

- Pros
  - Easy and fast to compute
  - Compact representation of color information
  - Can easily be normalized so that different image histograms can be compared
- Cons
  - Local information (not able to extract spatial localized feature)
  - fixed-size structures, cannot achieve a balance between expressiveness and efficiency

Color Moments

- Central moments are statistics
  - First order = mean
  - Second order = variance
  - Third order = skew
  - Fourth order = kurtosis
  - High order moments are less intuitive

$$m_d = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^d$$

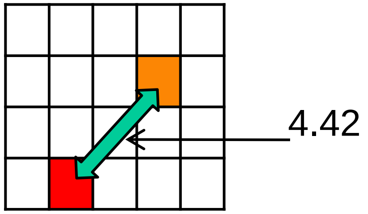
$$m_1 = 132.4$$
$$m_2 = 2008.2$$
$$m_3 = 4226$$
$$m_4 = 12.6 \times 10^6$$

- For color images, take moments of each band

Color Correlogram

- Describe global distribution of local spatial correlation of colors
- A table indexed by color pairs, $\mathrm{P}(c_i, c_j, d)$ specifies the probability of finding a pixel of color c_j at a distance *d* from a pixel of color c_i in the image

  e.g. P (Red, Orange, 4.42) = Probability of



- An *Autocorrelogram* captures spatial correlation between identical colors => subset of correlograms

AutoCorrelogram

image 1

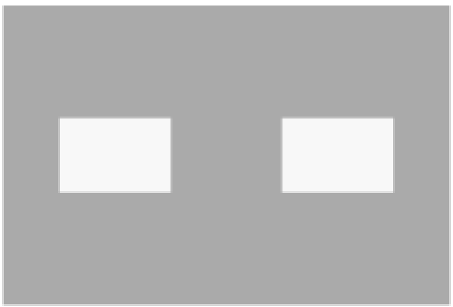
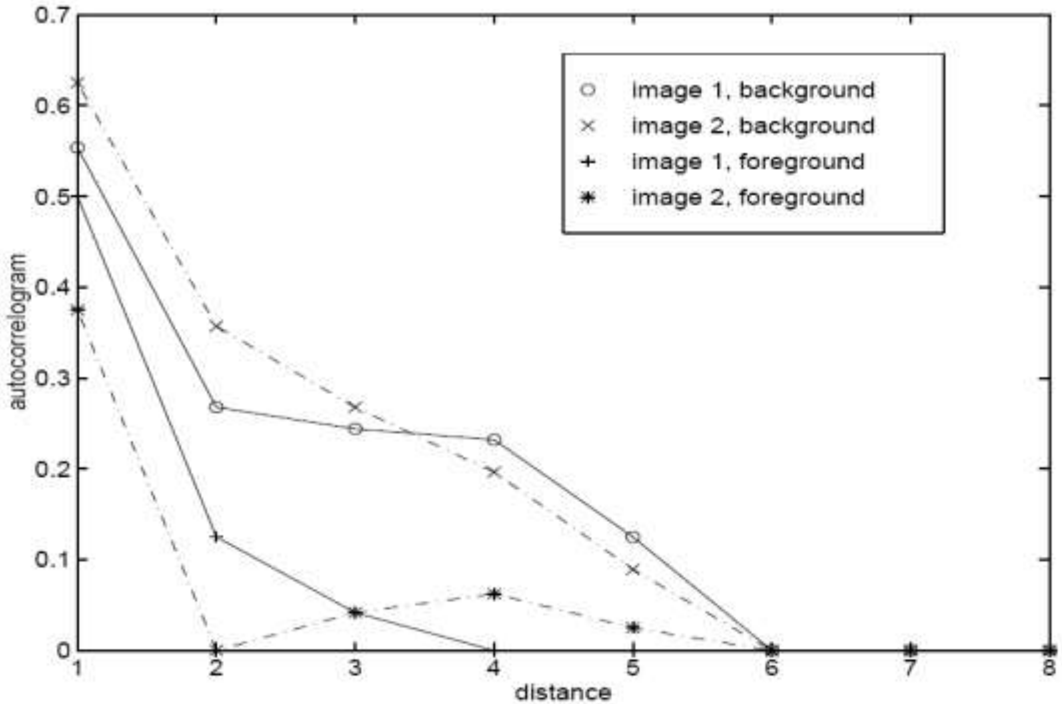
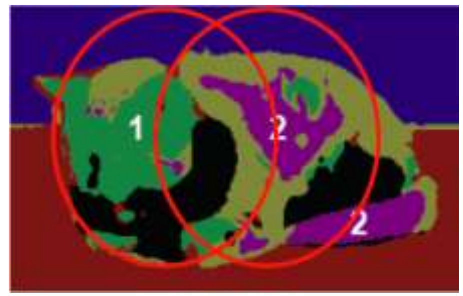
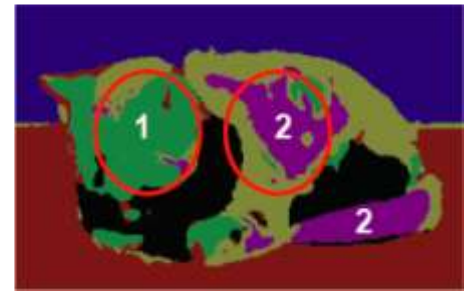


image 2



Circular kernel Correlogram

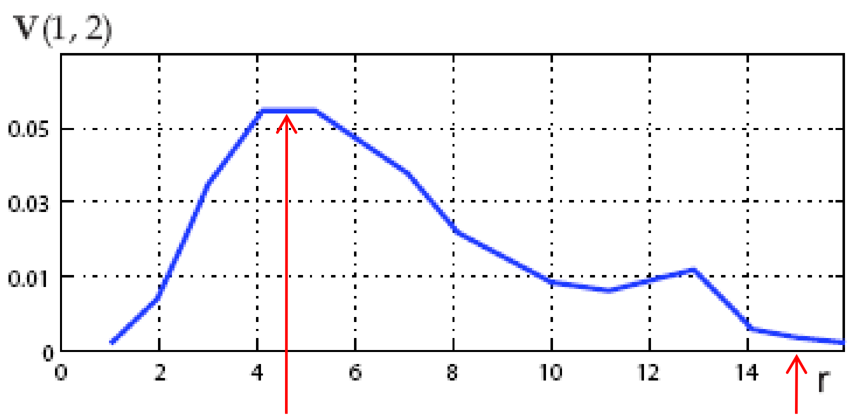




radius r = distance between regions

other radii

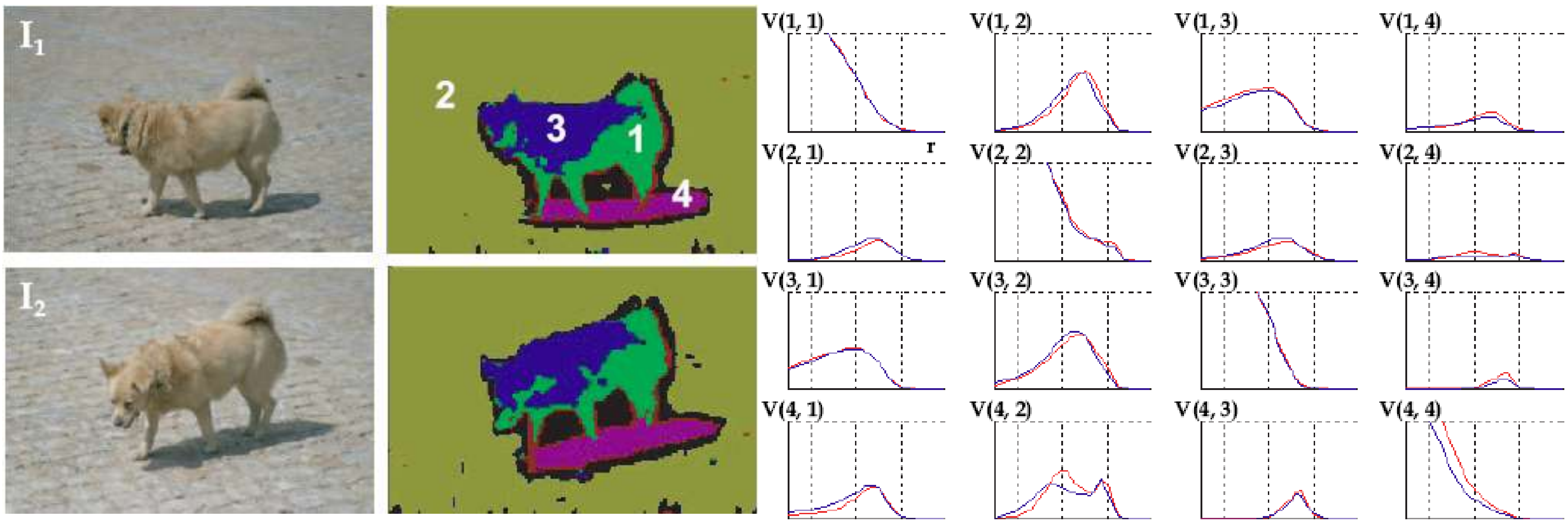
Correlation = pixel match between two regions covered by kernel



Maximum Correlation: (radius of kernel r = distance between regions)

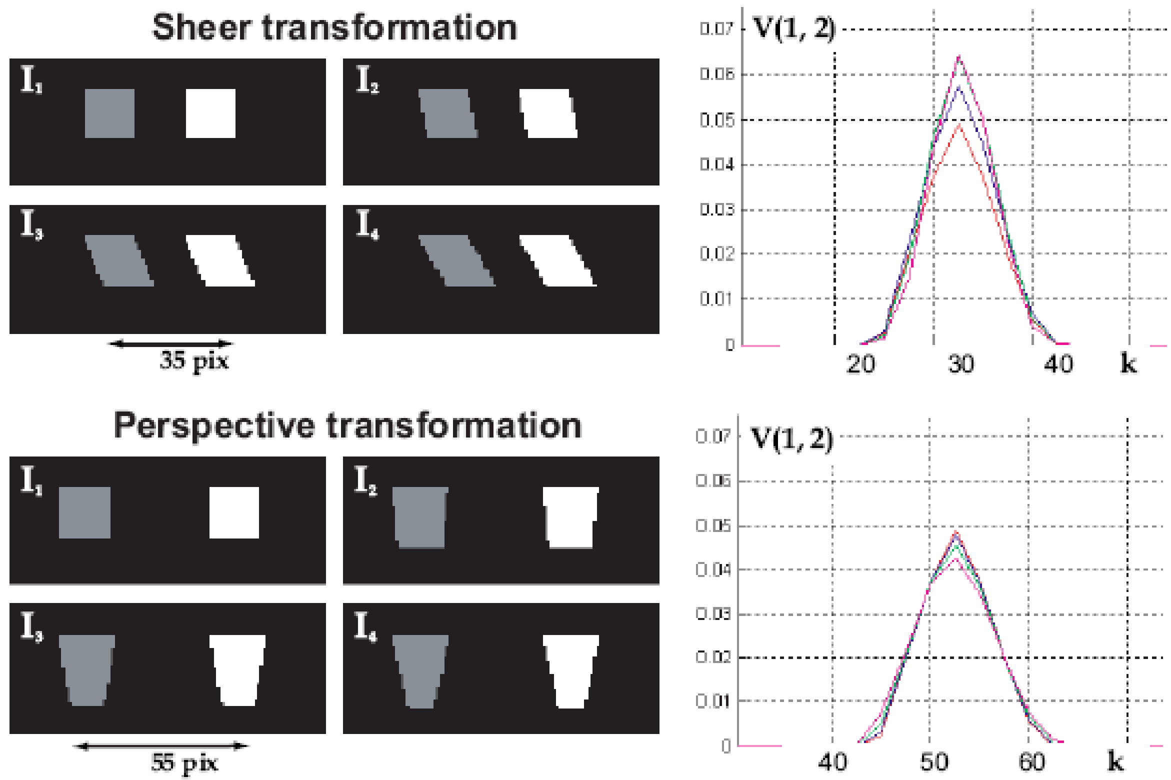
Correlation decreases as difference between radius and distance increase

Robust to Pose Changes



(See Savarese et al.)

Invariant to Geometric Transformations



Properties of Correlogram

- Invariant to translation
- Circular kernels induce rotational invariance
  - Rectangular kernels computationally more efficient
- Robust to affine, perspective transformations
- Robust to general object pose changes
- Not invariant with respect to scale! => learn with multiple training images at multiple scales
- Can be computed in $O(K*N^2)$, K<<N

Texture

What is Texture?

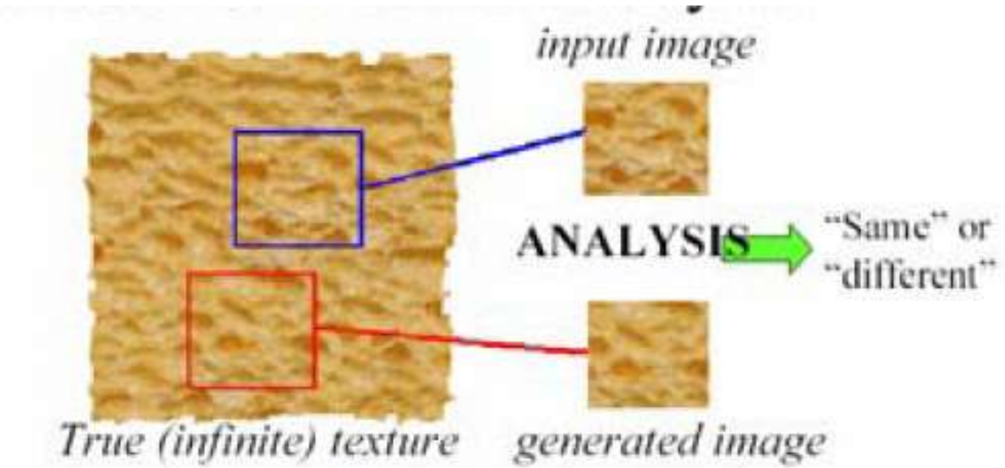
- Often used to represent all the “details” in the image
- One or more basic local patterns that are repeated in a periodic manner

Texture with repeated local patterns

Local pattern

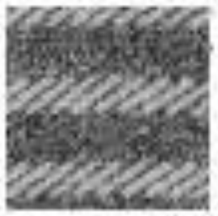
Discrimination

- Goal of texture analysis

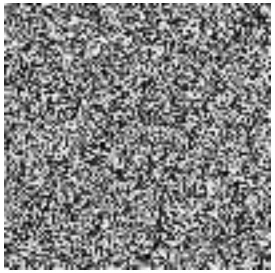


- Compare textures and decide if they're made of the same thing

What is Texture?

repetition

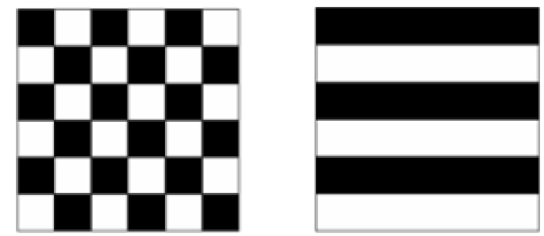
stochastic

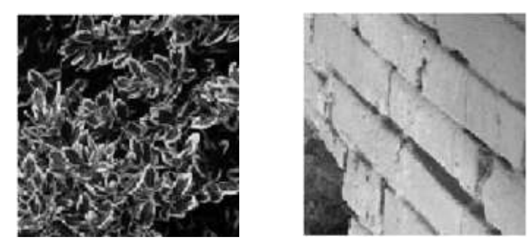
both

Texture Analysis

- Two approaches for texture analysis:
  - Structural (top down)
    - Decompose image into basic elements or texels (textons)
    - Convenient for artificial textures
  - Statistical (bottom up)
    - Texture is property that can be derived from the statistics of small group of pixels, such as mean and variance
    - Convenient for natural textures



Artificial textures



Natural textures

Statistical Approach

- Not always easy to define and segment out texels, especially for natural scenes

Look like similar, but difficult to see a texel structure; might have similar statistics

- Numeric quantities or statistics that describe a texture can be computed from the gray level values (or colors) alone
- Less intuitive, but computationally efficient
- It can be used for both classification and segmentation

- Alternatives:
  - Co-occurrence matrices
  - Edge histogram
  - Wavelets
  - etc…

Texture Models

- Transform an image window into a set of numbers (feature vector)
- Patches of the same texture should cluster in feature space

- Textures are made up of repeated subelements, with similar statistical properties
- Problem: find the subelements and represent their statistics

Identifying subelements

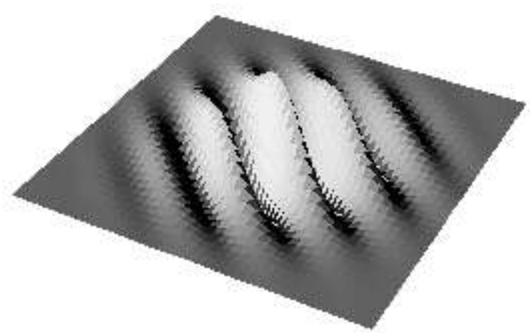
- Look for specific shapes
  - But: no known canonical set of textons
- Use a set of filters that capture simple pattern elements
- Human vision suggests spots and oriented filters at different scales
  - Spots: typically symmetric Gaussians
  - Bars: typically oriented Gaussians

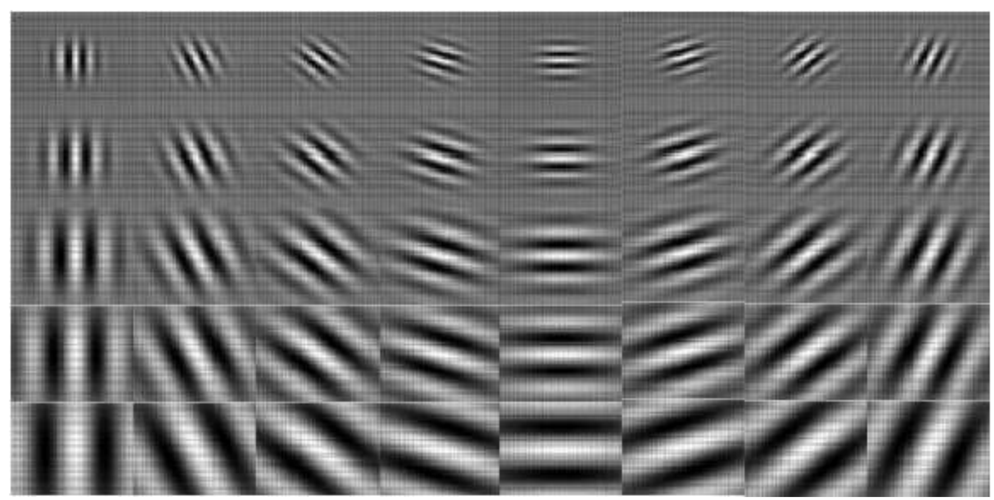
Choice of Filters

- No obvious advantage to any type of oriented filters
  - Weighted sum of Gaussians
  - Gabor filters
  - Wavelets
- How many filters?
  - Literature suggests 4-11 scales and 2-18 orientations (typically 6 orientations suffice)
- Tradeoff: detail of representation versus cost of computation

Gabor Filters

- 2-D sine waves modulated by a Gaussian envelope.
- Good models of the receptive fields found in simple cells of the primary visual cortex.
- 2D Gabor filter at different scales and orientations (spatial domain):





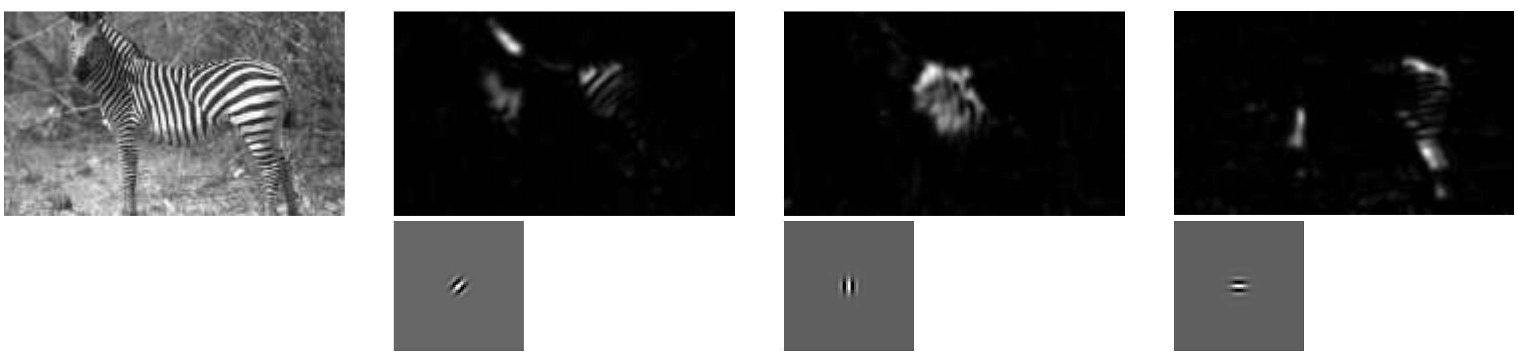
Typical: 5 scales, 8 orientations

Gabor Wavelet Transform

- For a given image *I(x,y)*, perform convolution with gabor filter kernels

$$G_{mn}(x,y) = \sum_{s=0}^{S}\sum_{t=0}^{T} I(x-s, y-t)\, g_{m,n}(s,t)$$

- Compute magnitude with real and imaginary part responses



Example filter response in magnitude with corresponding gabor filter (real part)

Gabor Filter Feature

- Each channel in the filter bank filters a specific type of texture
- Gabor feature descriptor
  - Computes the energy and energy deviation for each channel
  - Computes mean and standard variation of frequency coefficients

$$F = \{f_{DC}, f_{SD}, \mu_1, \ldots, \mu_N, \sigma_1, \ldots, \sigma_N\}$$

- Distance metric
  - L1 distance (in MPEG-7)

Other Models: Co-occurrence Matrix

- The intensity histogram is very limited in describing a texture (e.g - checkerboard versus white-black regions.
- Use higher-level statistics: Co-occurrence Matrix (Pairs distribution).
- Let $f(m,n)$ be a gray-level image. Then the co-occurrence matrix $C_{\boldsymbol{d}}$ is defined as follows:

$$C_{\boldsymbol{d}}(i,j) = |\{(m,n)|f(m,n) = i \text{ and } f(m+dm, n+dn) = j\}|$$

where $\boldsymbol{d} = (dm, dn)$ is displacement, the value of $C_{\boldsymbol{d}}(i,j)$ indicates how many times value *i* co-occurs with value *j* in some designated spatial relationship

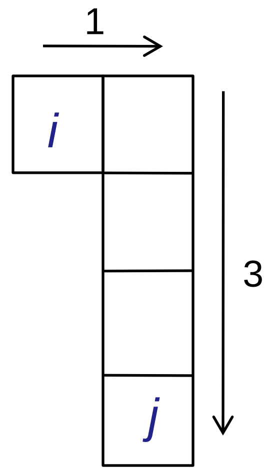
Computer Vision for Human-Computer Interaction
Research Group - Universität Karlsruhe (TH)

cv:hci

Co-occurrence Matrix

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \end{bmatrix}$$

gray level image



$\boxed{d=(3,1)}$

$$\begin{array}{c c} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \begin{bmatrix} 1 & 0 & 3 \\ 2 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \quad C_d$$

co-occurrence matrix

Variations

- Two variations:
  - Normalized co-occurrence:

$$N_d(i,j) = \frac{C_d(i,j)}{\sum_i \sum_j C_d(i,j)} \quad \Rightarrow \quad 0 \le N_d \le 1$$

N_d can be thought of as probabilities

  - Symmetric co-occurrence:

$$S_d(i,j) = C_d(i,j) + C_{-d}(i,j)$$

groups pairs of symmetric adjacencies

Co-occurrence Matrices

- Co-occurrence matrices capture properties of a texture, but they are not compact; instead, numeric features can be computed from them to be used in further analysis
  - Energy = $\sum_i \sum_j N_d^2(i,j)$
  - Entropy = $-\sum_i \sum_j N_d(i,j) \log_2 N_d(i,j)$
  - Contrast = $\sum_i \sum_j (i-j)^2 N_d(i,j)$
  - Homogeneity = $\sum_i \sum_j \frac{N_d(i,j)}{1+|i-j|}$
  - Correlation = $\sum_i \sum_j \frac{(i-\mu_i)(j-\mu_j) N_d(i,j)}{\sigma_i \sigma_j}$

μ_i , μ_j , σ_i, σ_j are means and standard deviations of the rows and column sums:

$$N_d(i) = \sum_j N_d(i,j)$$

- How to choose *d*?
  - One solution is to select value(s) of *d* that have the most structure; i.e., to maximize

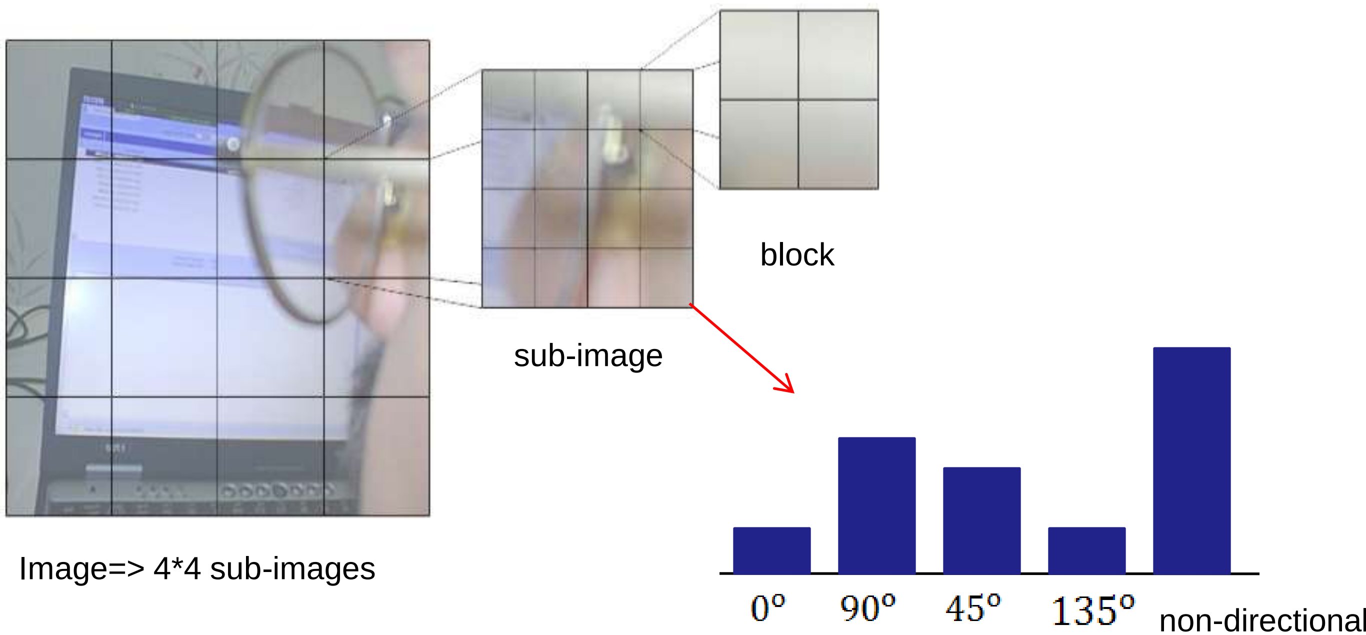
$$\sum \sum \frac{N_d^2(i,j)}{N_d(i) N_d(j)} - 1$$

Other Models: Edge Histogram

- Captures the spatial distribution of different types of edges
  - Partition image into large local regions
  - Partition each local region into small image patches
  - Quantize each patch as horizontal, vertical, diagonal
    - Use gradient filters
  - Collect results into local region histograms
  - Perform matching based on histograms

Edge Histogram Descriptor

- Edge Histogram Descriptor (EHD) in MPEG-7 for calculating frame similarity



A 5-bin histogram in for each sub-image

Spatial Information

- In which image areas should the descriptors be computed?
  - Whole image?
  - Sub-windows?

- Some spatial information can be kept by extracting descriptors on sub-windows

- Other approaches
  - Local descriptors / key point detection
  - Image segmentation
    - → next week

…