UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

AImage Lab

Vision and Cognitive Systems

Cap. Image processing 1

PIXELS, Images

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Images are bidimensional signals.
IMAGE is a 2D sampled function acquired by specific visual sensors from the external world
I(x,y)=f(x,y): $f(x,y): R^2 \rightarrow R$

[DIGITAL] IMAGE is a MATRIX of (N,M) values, called PIXELs (picture elements) representing the sampled function; an uncompressed image is also called *iconic image.*

M x N= width x height =640 x 480

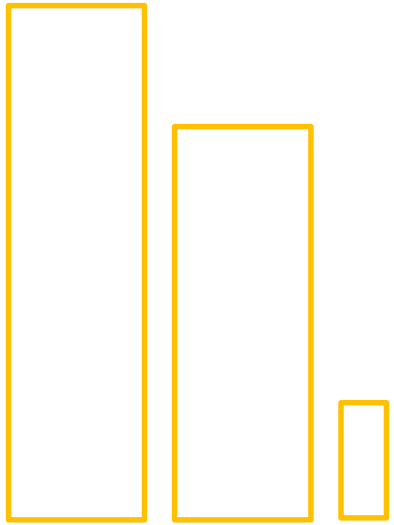
$I(x,y) = f(x,y),\ x \in [x_{min}\ , x_{max}],\ y \in [y_{min}\ , y_{max}]$

The **Pixel is defined as the value of the function represented in the image.** It is usually the sampled and quantized value measured by a sensor but can be also artificially defined in a synthetic way. It can represent either the luminance or the intensity of a set of values of a color space (RGB, HSV, LUV..) or the values of a sensor..

IMAGE and spatial information

The image is the matricial representation of the point values in the **image plane of the camera.** Image pixels in images taken from a camera of the real world give **the spatial information** of the objects in the scene **on the image plane**

Depending on the image formation method, the camera parameters, the lens, represent the 3D spatial distribution projected in a 2D space.



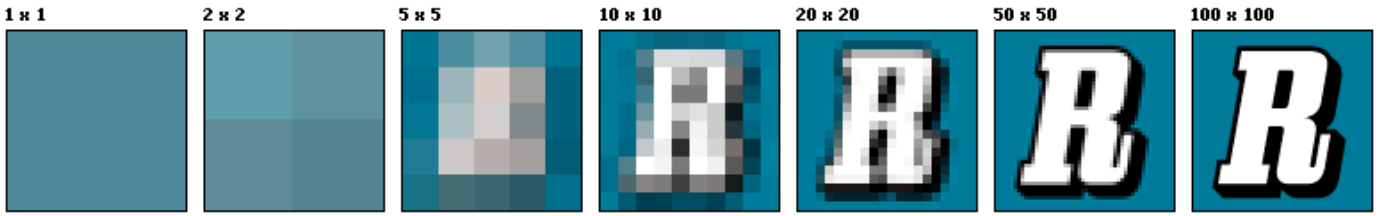
Different or similar height?
The perspective problem

Reconstructing 3D space from the 2D image plane is challenging… we will do it.

RESOLUTION

SPATIAL RESOLUTION (N,M) is the number of samples – pixels in the image plane; depending on the application.

character recognition	64 x 64
automatic inspection	128 x 128, 256 x 256….
Biometric, surveillance	12 x 512, 1024 x 1024 (but target is smaller..)
satellite images	4096 x 4096
Web images	JPEG, other formats..
Now ….	No standard!

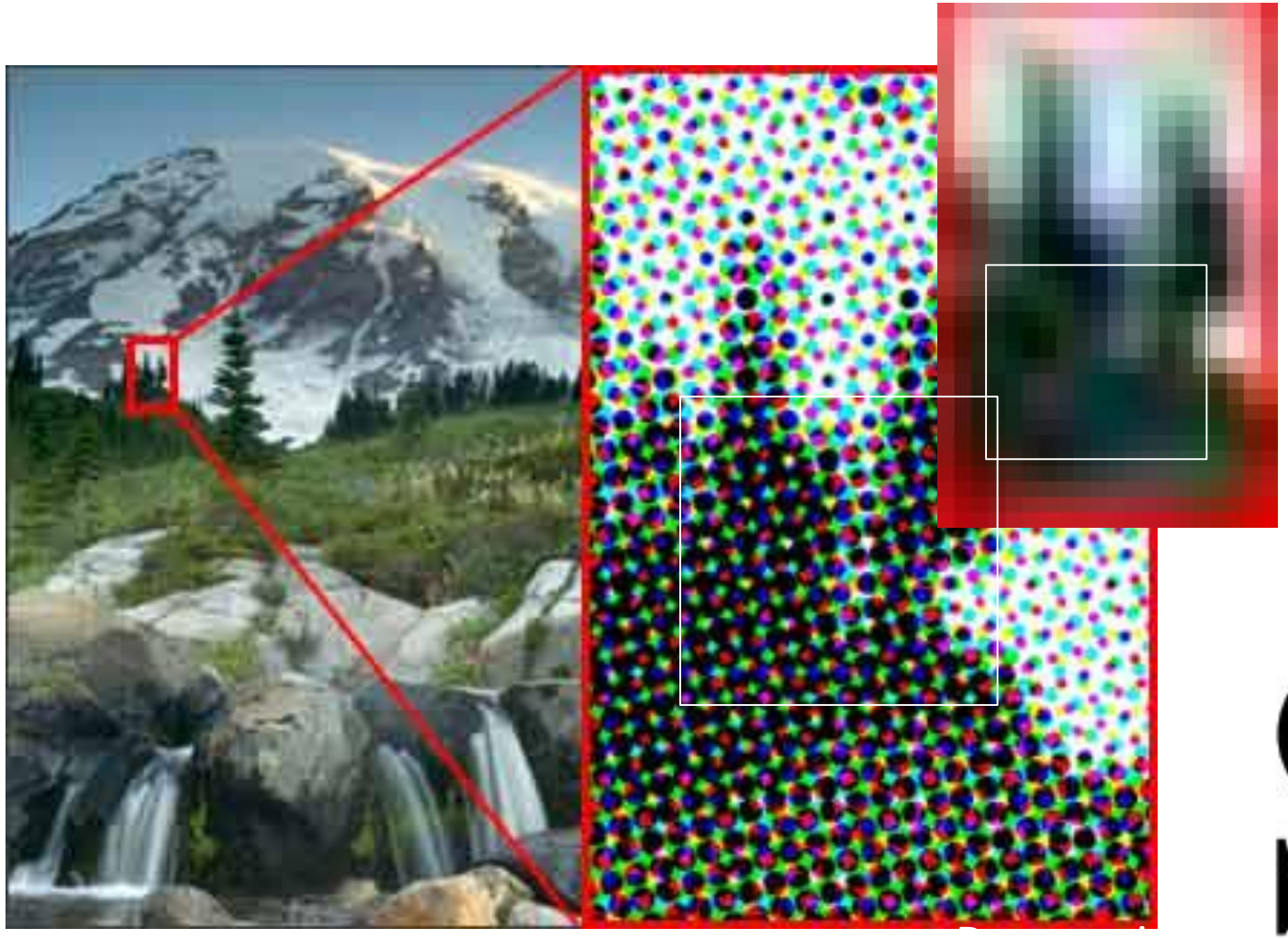


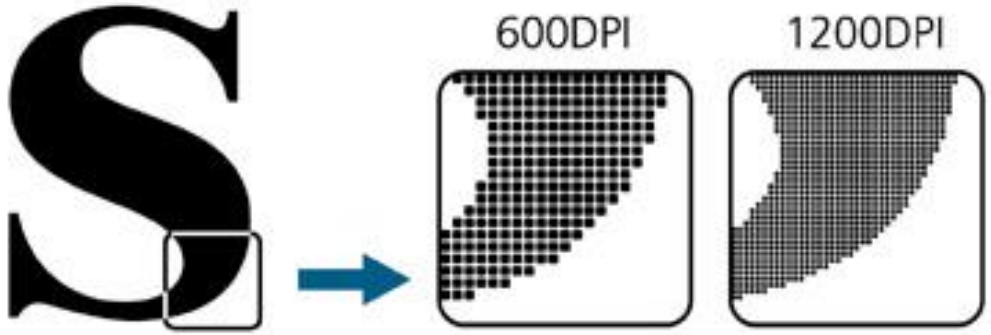
TEMPORAL RESOLUTION is the **frame rate**, measured in fps.

GRAY LEVEL RESOLUTION (radiometric resolution) is the number of gray levels: images use **k** bits for representing the gray level; usually 256 (k=8).

Print resolution

Dot per inch (DPI)





COLOR IMAGES

Image is a set of matrixes one for each dimension of the color space
It is a **data structure** multi.array.
I(x,y) is a 3D Vector in a color space (RGB, HSV)
Video: is a (M+1)D structure
the time dimension!
Rgb: In 1931 CIE Commission Internationale d'Eclairage
For the average standard observer

N.8 MarkRothko

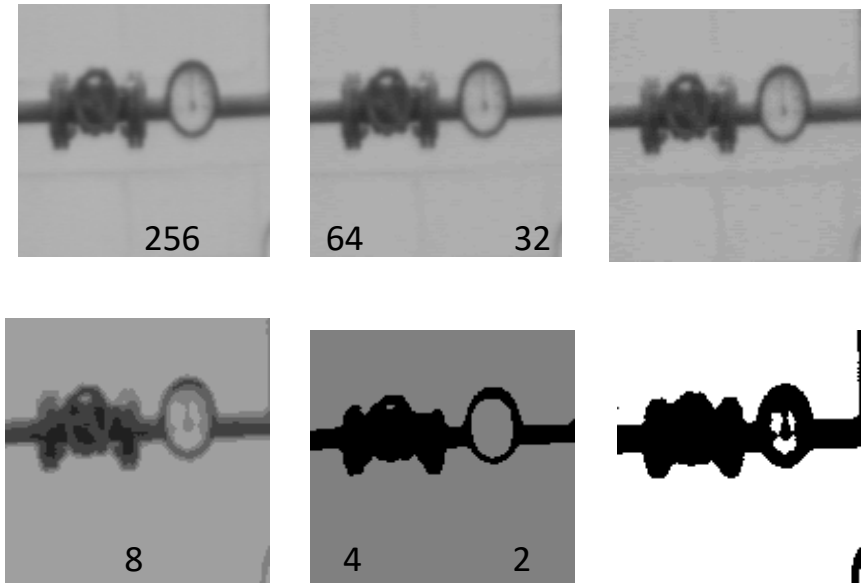
Color

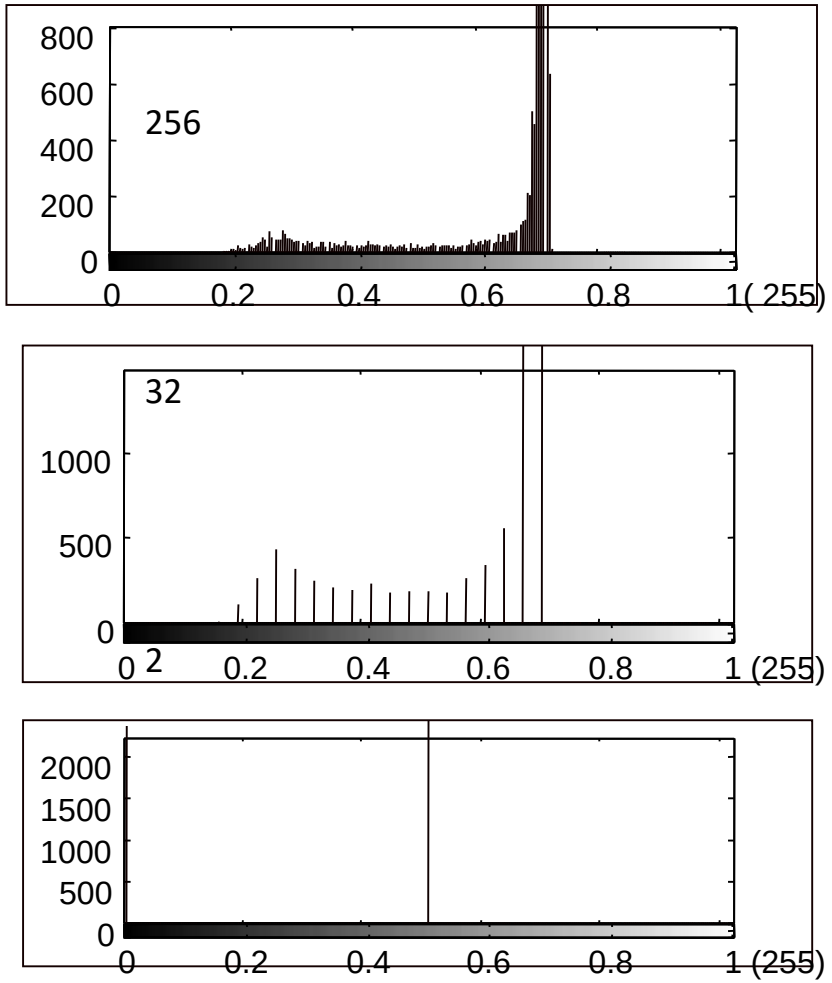
gray scale

binarized

Gray level images

Are 256 levels enough? The human eyes do not need 256 levels and often neither the automatic analysis do.



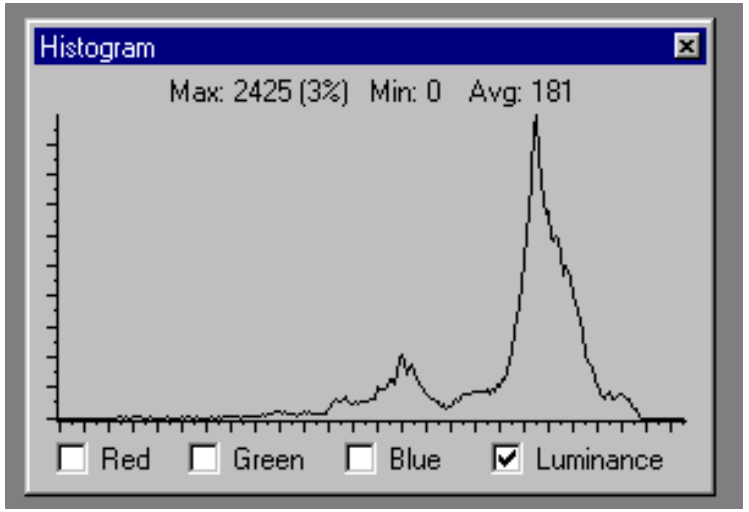


warning: Luminance quantization
is not a Semantic segmentatation

histogram

The **histogram** at gray level is a vector with as many elements (**bins**) as the number of gray levels; the value of each bin is the accumulation of the number of pixels which, in that image, assume the correspondent gray level

```
for (n = 0; n < NLIV; n++)
  hist[n] = 0;
for (i = 0; i < NROW; i++)
  for (j = 0; j < NCOL; j++)
    hist [img [i][j]] ++;
```

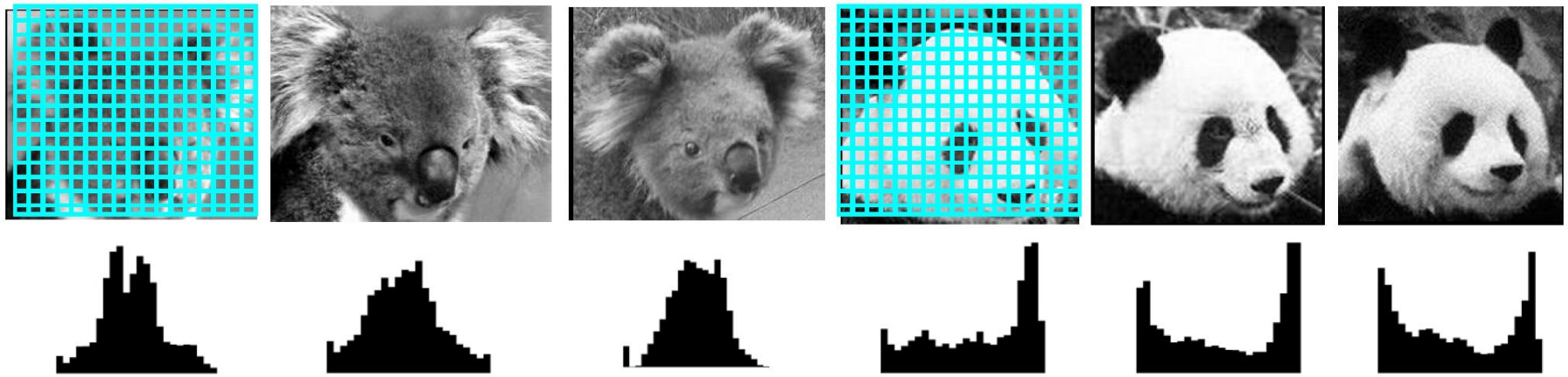


The histogram gives important information for image processing, especially for contrast enhancement and segmentation

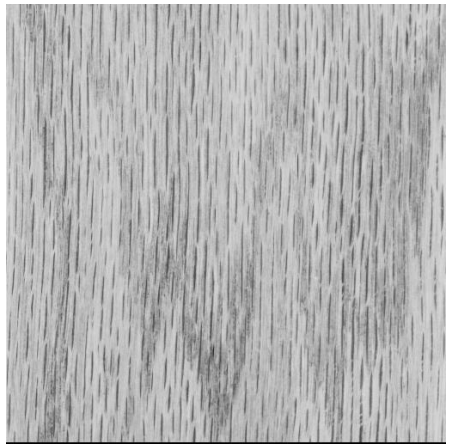
In color images, 3 histograms, or a combined histogram

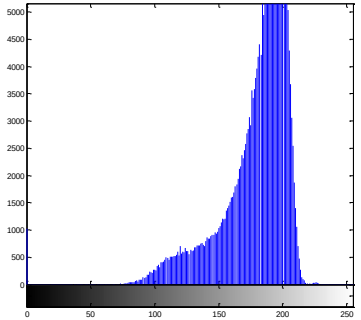
HIstogram

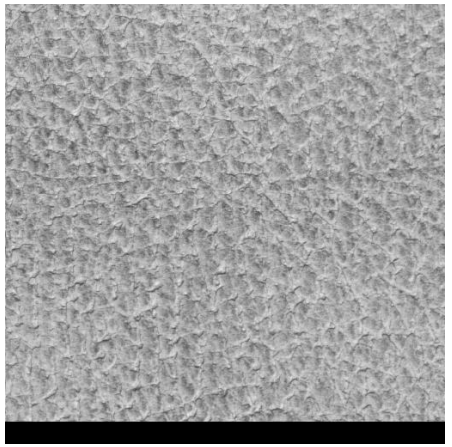
Histogram can be used as a feature vector

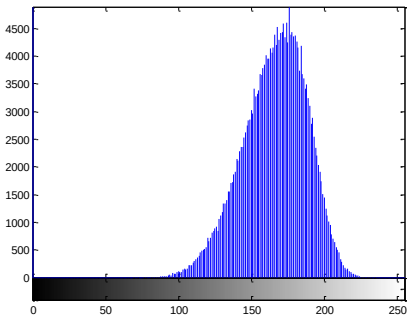


It is a **global feature vector , but** sometime not discriminant enough

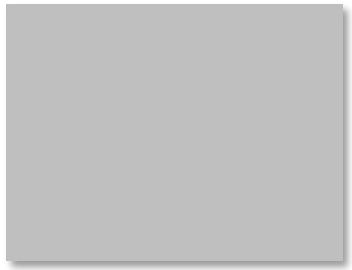


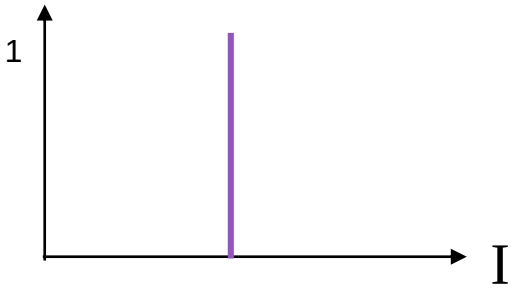


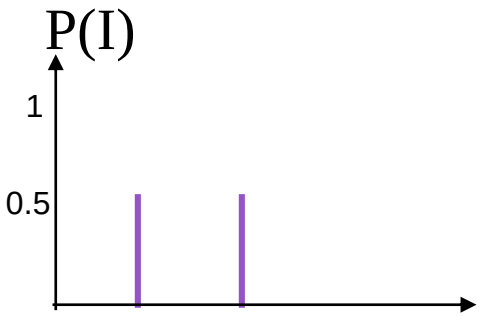


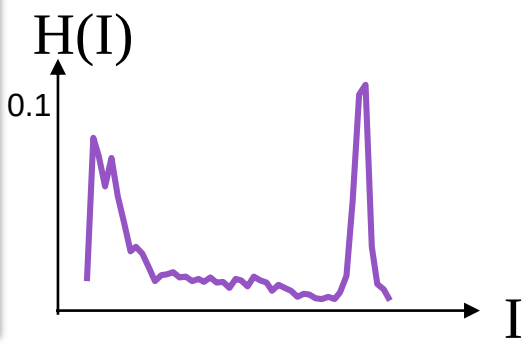


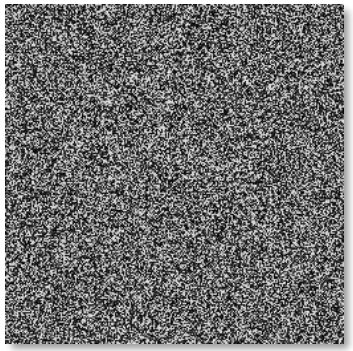
The image histogram does not fully represent the image

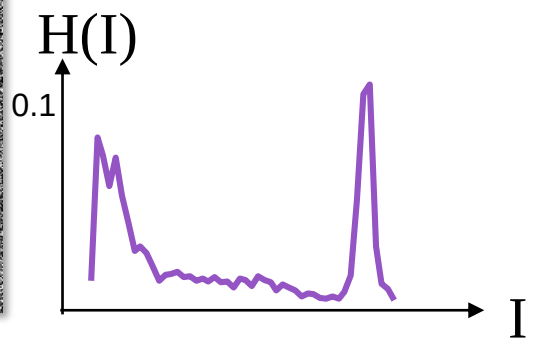








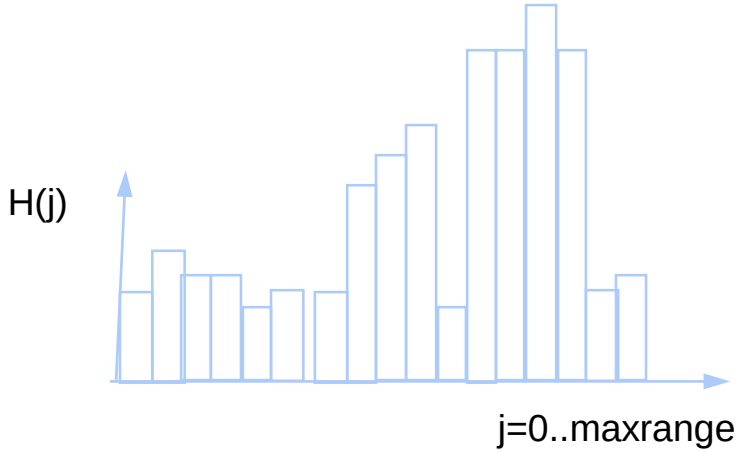




Pixel permutation of the left image

HIstogram

Histogram can be viewed as a discrete approximation of a **pdf distribution of probability** (if the data could be associated with a random distributed variables i.i.d. independent and identically distributed).
Given an image I di n=N x M pixels with L levels
(from 0 to maxrange)



Histogram:
$\boldsymbol{H}(j) = \#\{\boldsymbol{x}: \boldsymbol{I}(\boldsymbol{x}) = j\}$ # operator counts the number of elements

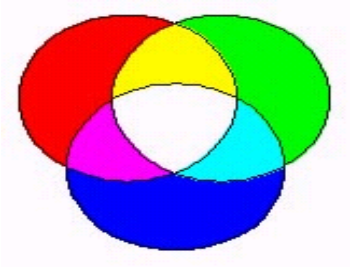
In order to be a pdf , the Normalized
version of histogram must be used
(with the sum of each components equal to 1)

$$\sum_{j=0}^{L-1} p_I(j) = 1$$

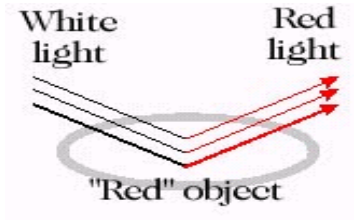
REMIND COLOR THEORY

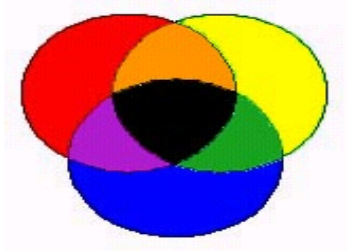
RGB and CMY

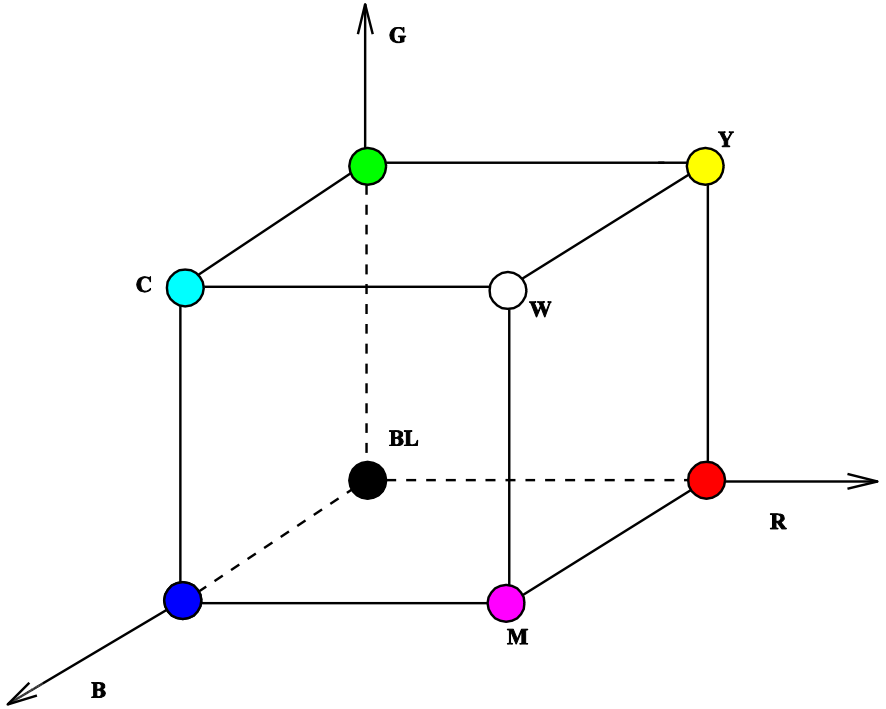
RGB:
Most common – graphics and displays.
Additive and Device Dependent.
Color perception not absolute.



CMY:
Common – graphics and printers.
Subtractive and Device Dependent.
C = 1-R, M = 1-G, Y = 1-B.
Color perception not absolute.







Color spaces

Tridimensional space for each point

- Perceptual Models **CIE (CIELab, XYZ)**
- Hardware oriented models **RGB, CMY**
- Transmission oriented models **YUV, YIQ**
- Compression oriented models **YCbCr**
- Intuitive (perceptive) processing models **HSI, HSV**

Color	Meaning
RGB	Generic Red, Green, Blue unspecified primaries
sRGB	Standard RGB; http://www.srgb.com/
CIE XYZ	Virtual positive primaries, direct relation to spectral power distributions.
CIE L*a*b*	(also CIELAB) standard; http://www.cie.co.at
CIE L*u*v*	(also CIELUV) standard; http://www.cie.co.at
Spherical LAB	L Luminance, AB angular space; [10]
HSV	Hue, Saturation, Value components
HSI	Hue, Saturation, Intensity components
YCbCr	Luminance and chrominance separation for digital video transmission

RGB AND Palette

256 colors

A small block x=100..104, y=100..104

190	190	191	188	186
190	188	188	**188**	188
182	182	187	183	186
167	170	170	174	182
158	158	158	170	173

Link to RGB plaette

Palette:

...			
186	1.0000	0.6118	0.3216
187	1.0000	0.5490	0.3529
188	**1.0000**	**0.5686**	**0.4000**
189	1.0000	0.6353	0.3255
190	1.0000	0.6118	0.4510
191	1.0000	0.6471	0.4196
...			

RGB value in [0.0,1.0]

CIELAB space

CIEXYZ is a reference space which describes the physical nature of color and not its perception by an observator → **CIELAB** is the *de-facto standard in colorimetry*

:

$$L^* = 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 \quad \left(\frac{Y}{Y_n}\right) > 0.008856$$

$$L^* = 903.3\left(\frac{Y}{Y_n}\right) \quad \left(\frac{Y}{Y_n}\right) \leq 0.008856$$

$$a^* = 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right]$$

$$b^* = 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right]$$

$$f(x) = \begin{cases} x^{\frac{1}{3}} & x > 0.008856 \\ 7.787x + \frac{16}{116} & x \leq 0.008856 \end{cases}$$

L is ***Lightness*** CIE 1976 is the prerception of luminosity for a standard observator with a standard illuminator (X_n,Y_n=100,Z_n)

HSV

In each plane with constant brigthness, hue and saturation can be computed to create an Euclidean space
Common among artists.

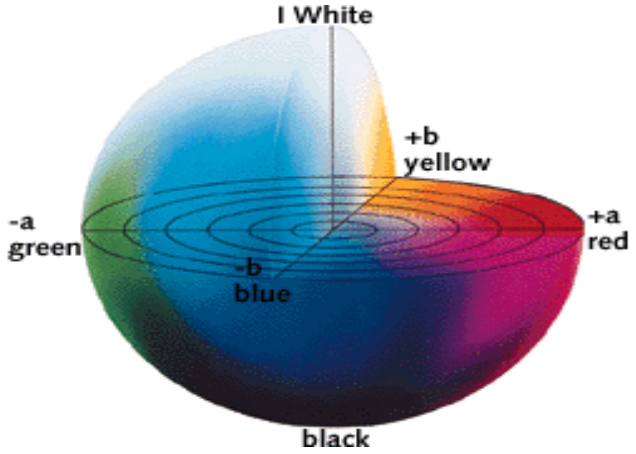
Based on artistic perception.

Hue, Saturation and Value.

Hue = tint of color.
Value = brightness of color.
Saturation = strength of color.

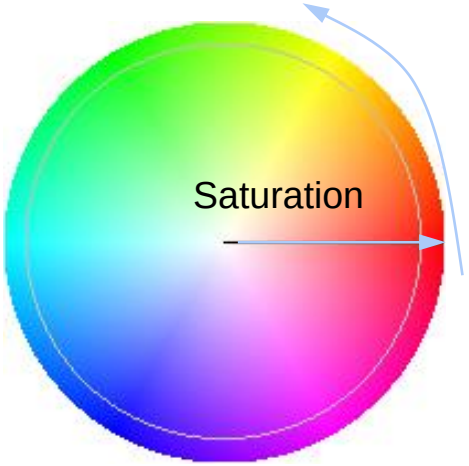
Easy to visualize colors.

$$
\begin{aligned}
H &= \arccos \frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{((R-G)^2+(R-B)(G-B))}} \\
S &= 1-3\frac{\min(R,G,B)}{R+G+B} \\
V &= \frac{1}{3}(R+G+B)
\end{aligned}
$$

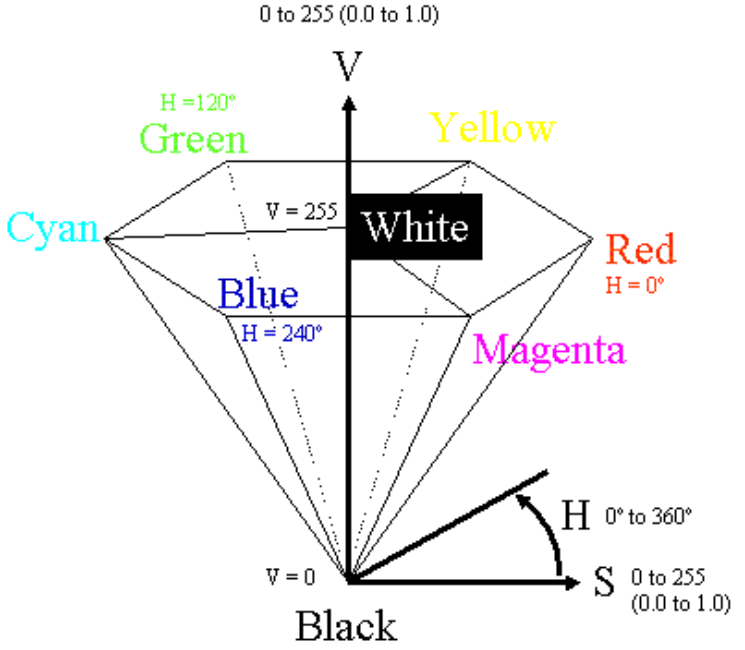


HSV

From RGB to HSV



Hue



$$Max = \max(R, G, B)$$
$$Min = \min(R, G, B)$$

$$V = Max$$

$$S = \frac{(Max - Min)}{Max}$$

$$H = \frac{\pi}{6}\begin{cases} \dfrac{G - B}{(Max - Min)} & Max = R \\ 2 + \dfrac{B - R}{(Max - Min)} & Max = G \\ 4 + \dfrac{R - G}{(Max - Min)} & Max = B \end{cases}$$

YCbCr

YCrCb is an encoded nonlinear RGB signal, commonly used by European television studios and for image compression work.
Linear transformation
Y – Luminance component, C – Chrominance

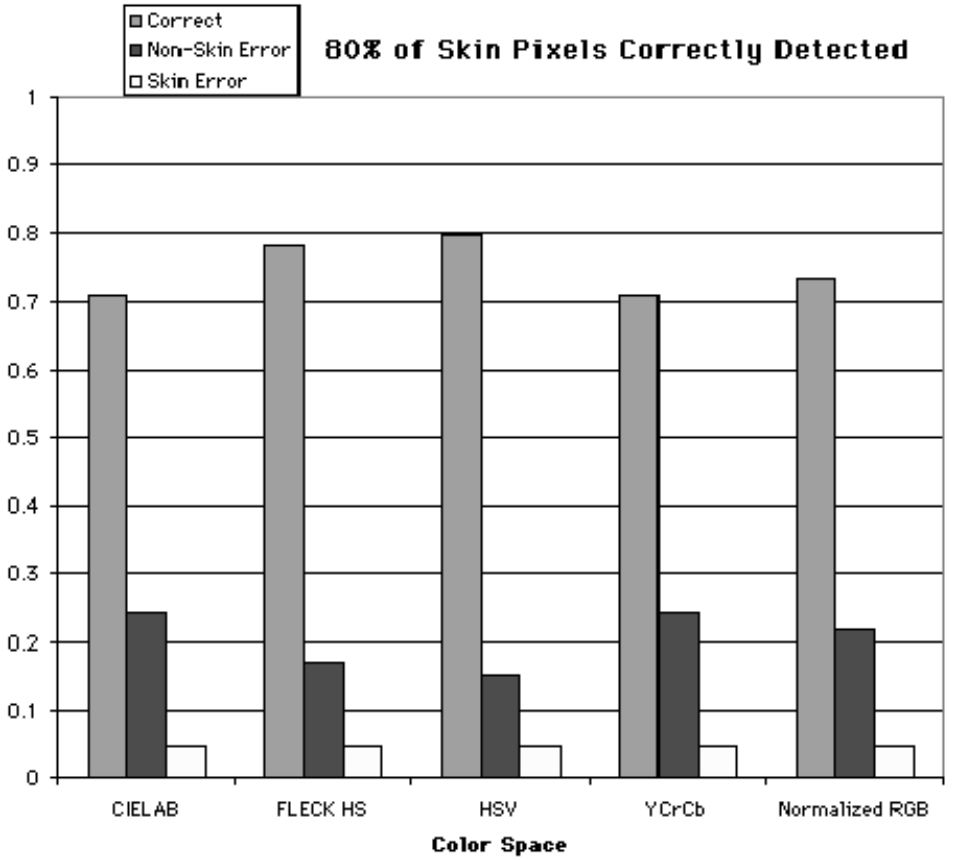
$$Y = 0.299R + 0.587G + 0.114B$$
$$C_r = R - Y$$
$$C_b = B - Y$$

Test on **skin detection**

HSV, HS gave the best results

Normalized rg is not far behind

CIELAB and YCrCb gave poor results



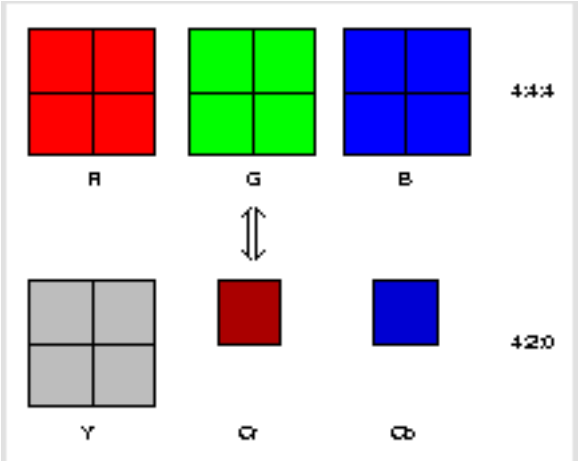
Benjamin D. Zarit, Boaz J. Super, Francis K. H. Quek,
"Comparison of five color models in skin pixel classification"
Works @ ICCV 1999
http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=8885C39DEB3CEAC6544C1D0B473ECA95?doi=10.1.1.129.9666&rep=rep1&type=pdf

JPEG e MPEG

In Jpeg and MPEG colors are converted into luminance Y and chrominance CbCr

Since the eye is more sensitive to Y than chrominance, this is reduced and for 2 or 4 pixels is substituted by the avrerage

Format 4:2:2



See the Multimedia Course.

Image processing operators

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Image processing uses some operators on image pixels which transfotm images into other images.

1)Point operators:

If the value of each pixel of the resulting image depends only on the original pixel in the same image's spatial position (e.g. Thresholding)

$$I'(x,y) = h(I(x,y)) \text{ or in vectorial form } I'(\boldsymbol{x}) = h\big(\boldsymbol{I}(\boldsymbol{x})\big)$$

Also for videos: $I_t{}'(\boldsymbol{x}) = h(I_t(\boldsymbol{x}))$

2)Local (Neighborhood) operators:

If the value of each pixel depends only on the original pixel in the same image's position and in a local Neighborhood (e.g. filters)

$$I'(x,y) = h(I(x,y), \aleph(I(x,y))$$

3) Global operators:

if the value of each pixel depends on all the pixels of the original image (e.g. Fourier transform)

$$I'(x,y) = h(\iint I(x,y)\,)$$

LINEAR POINT OPERATOR

Punctual (point) operators can be applied on one or more images:

$$g(\boldsymbol{x}) = h(f(\boldsymbol{x})) \text{ or } g(\boldsymbol{x}) = h(fo(\boldsymbol{x}), f1(\boldsymbol{x}), .. fn(\boldsymbol{x}),)$$

x is the Domain of the function; for the image, it is the pixel location, represented by the point coordinates in the 2D plane ***x****=(x,y)* or the image coordinates ***x****=(i,j)* .

h() is the operator which transforms an image to another image after an image processing operation
If the h() **transformation** is **linear** it can be written as

$$g(\boldsymbol{x}) = h\big(f(\boldsymbol{x})\big) = s\, f(\boldsymbol{x}) + k$$

s is the **scale factor**, often called also *gain or contrast*
k is the **offset constant** often called also bias or brightness

They are linear operators since they obey the superposition principle

$$h(f0 + f1) = h(f0) + h(f1)$$

Enhanching image perception of a mammography for human experts.

A point-linear operator

$$g(\boldsymbol{x}) = h\big(f(\boldsymbol{x})\big) = s\, f(\boldsymbol{x}) + k$$

A near-neighbor filter

How can we do?

Luminance variation

Luminance variation: Varying the luminance intensity or brightness means only shift the histogram values, by changing proportionally the pixel value in a linearly manner.
Eg. In figure 10% +
s=1.1;
k=0;

$$I'(\boldsymbol{x}) = h(I(\boldsymbol{x})) = s\, I(\boldsymbol{x}) + k$$

Negative operator
s=-1 k=255

LINEAR BLEND

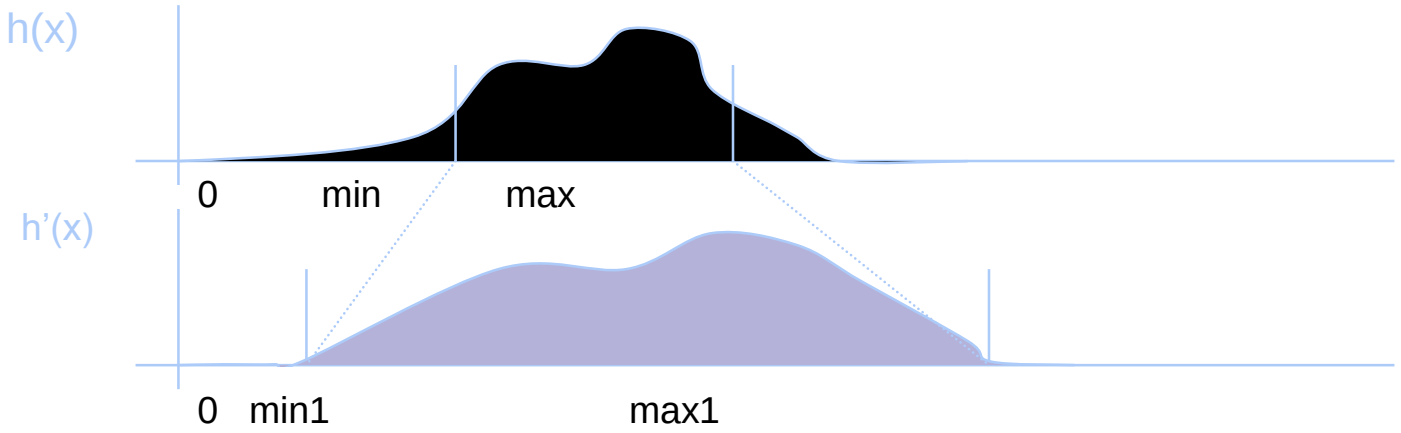
The **linear blend** or **dyadic operator**

Is a linear pixel-level operator combining two images. It is used in video processing for producing visual effects such as a linear transition

$$g(\boldsymbol{x}) = h\big(f0(\boldsymbol{x}), f1(\boldsymbol{x})\big) = (1-\alpha)f0(\boldsymbol{x}) + \alpha f1(\boldsymbol{x})$$

Contrast-stretching

Contrast stretching is the expansion of the gray (or color) level of pixels in a dynamic range given the histogram

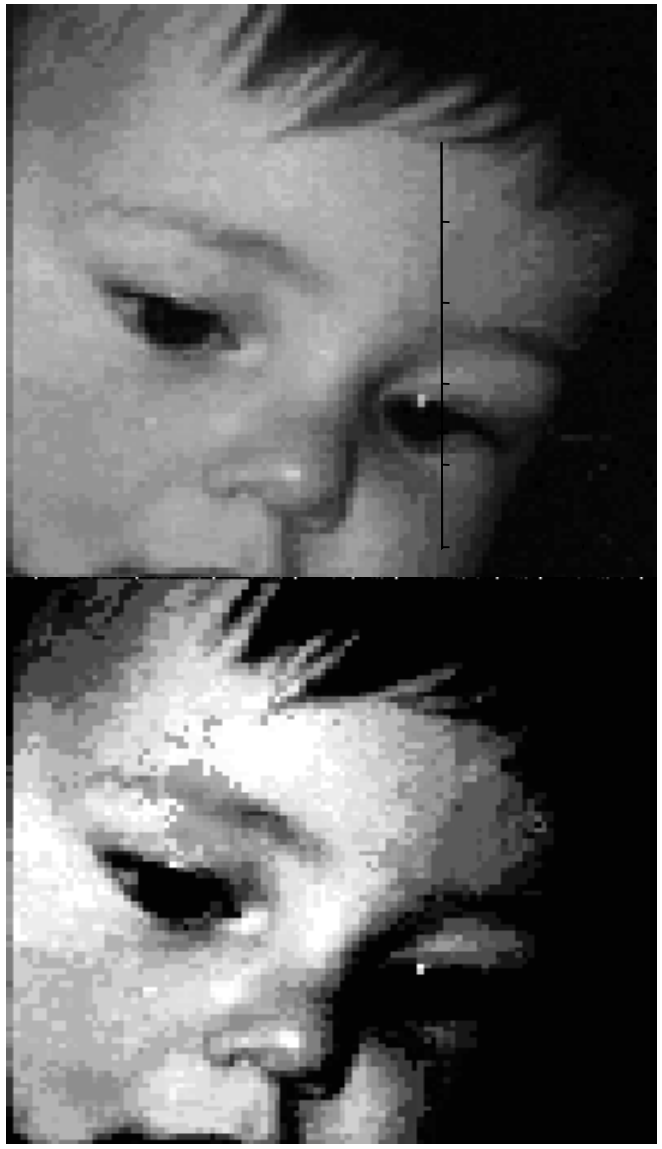


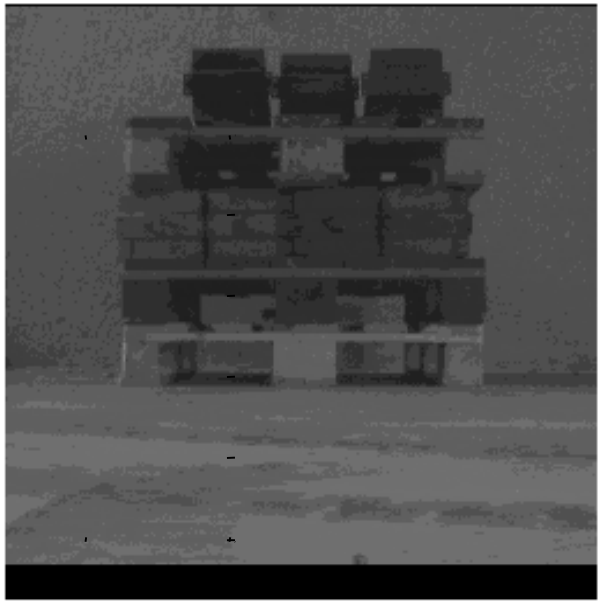
for each pixel p : (x,y) it computes I'(p):

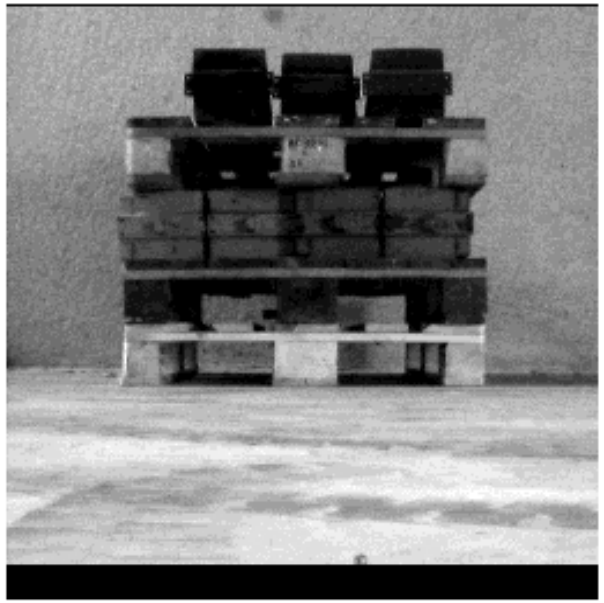
```
ScaleFactor= (max1-min1)/(max-min)

if (I(p)<=min)    I'(p)=min1;
if (I(p)>=max)    I'(p)= max1;
if ((I(p)>min)&(I(p)<max))
                    I'(p)= (I(p)-min)* ScaleFactor +min1;
```

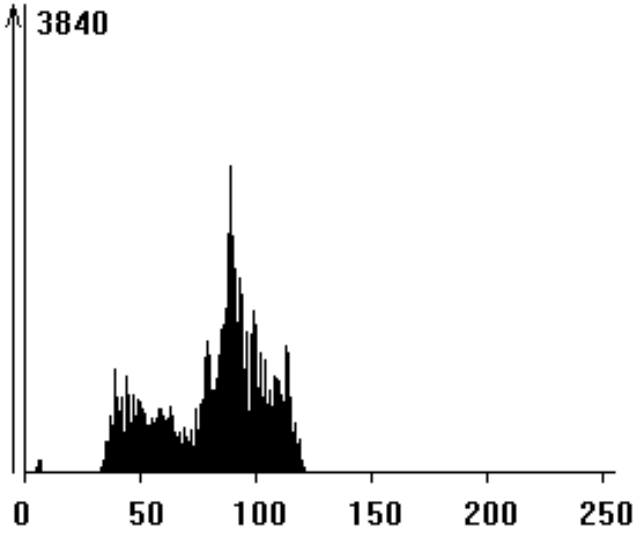
always $I'(\boldsymbol{p}) = h(I(\boldsymbol{p})) = s\, I(\boldsymbol{p}) + k$

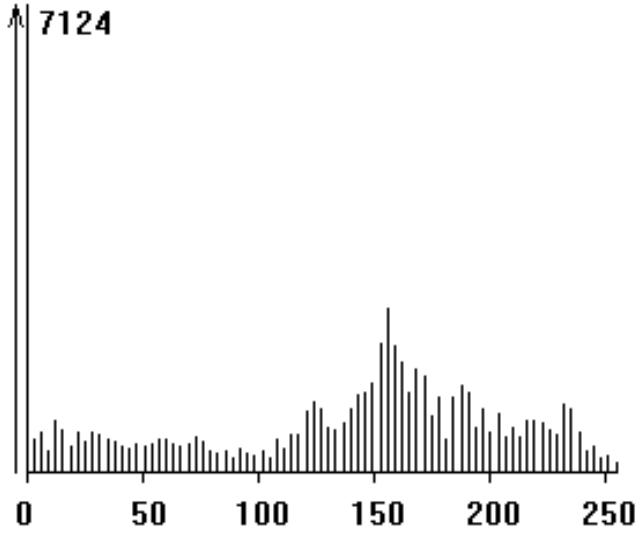
Examples







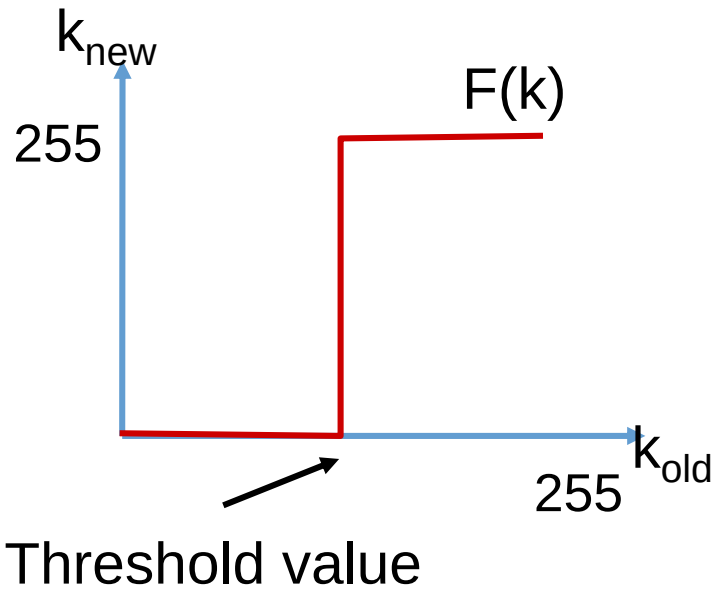


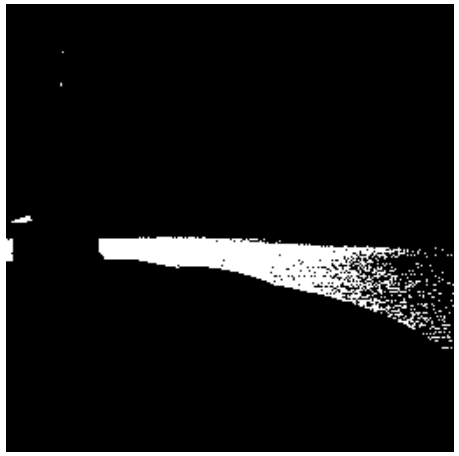


Thresholding

Thresholding: the first «segmentation operator»
Thresholding is the task to find (or give) a solution of a problem of transforming a gray level image into a binary image.
Thresholding becomes a segmentation operator is the semantic associated with the perception is in such a way preserved or enhanced.
The fist image has many gray levels buy we would be able to see only two levels: black (the target object) and whithe (the background)

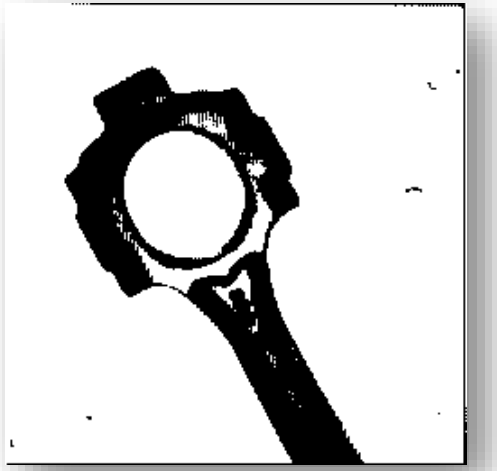
Thresholding: the task of finding a “ good” threshold value

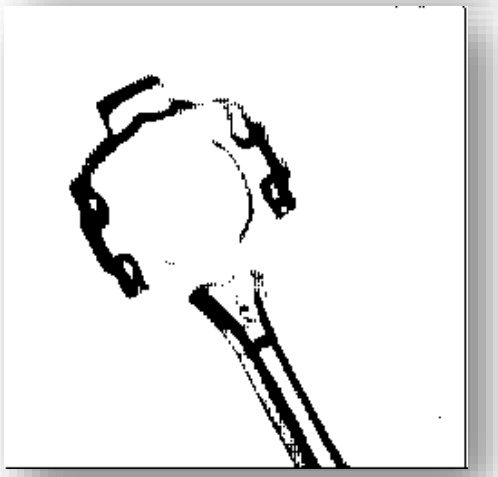




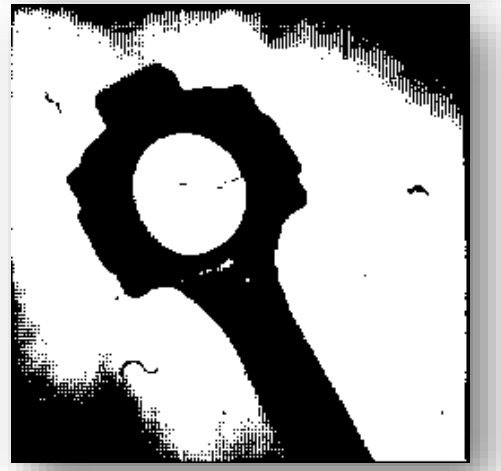
Original Image

Binary Image





Threshold too low



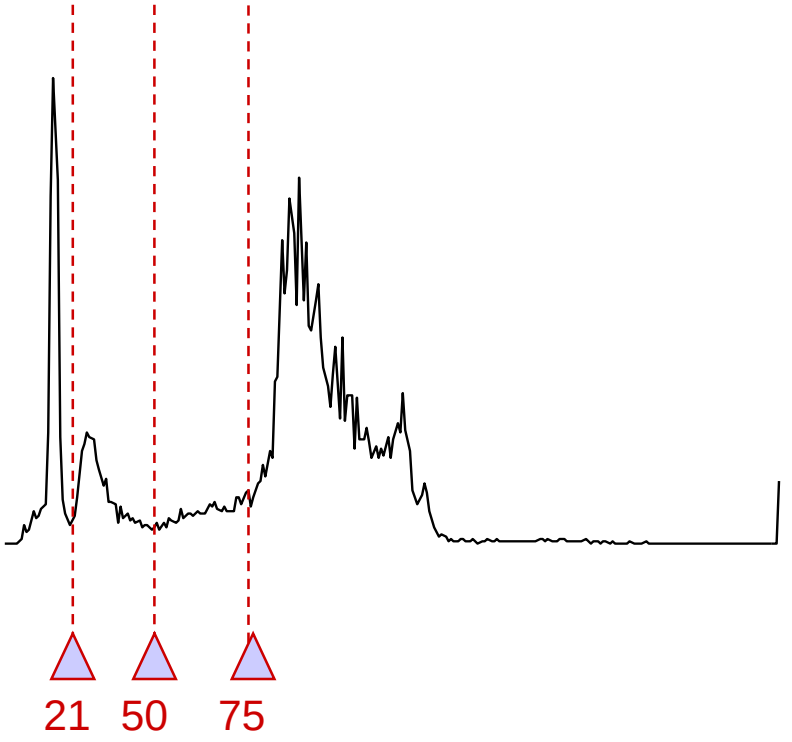
Threshold too high

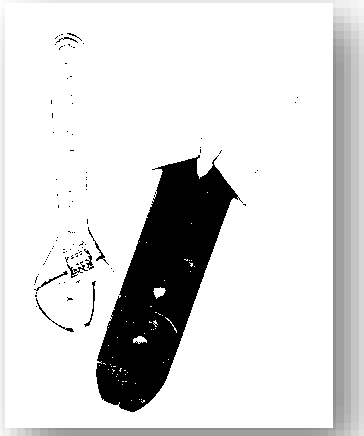
Segmentation using Thresholding

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

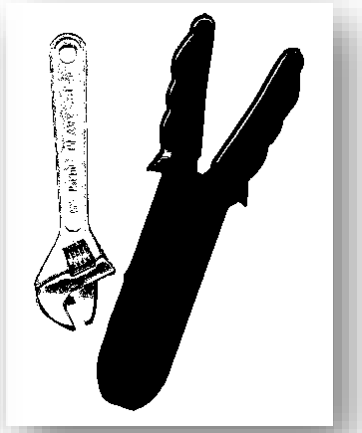
Original

Histogram

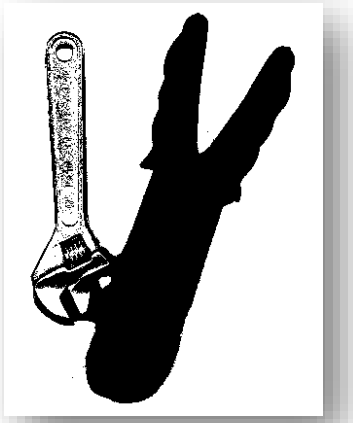




Threshold = 21



Threshold = 50



Threshold = 75

Thresholding

Thresholding: the task consists in the selection of a value T of brightness (intensity) capable of dividing the image into 2 regions of pixels with intensity greater or less than T .
It is a punctual operator which transforms a grey level image into a binary image (thresholding-based binarization)
Given an image I(**x**)=f(**x**) (**x**)=(i,j) , it is transformed to g(**x**)

$$g(\boldsymbol{x}) = Thresh(f(\boldsymbol{x}), T)$$

```
For each (i)
   For each (j)
     if (f(i,j)>=T)    g(i,j)=1;
                     else g(i,j)=0;
```

Global threshold if T= T(f)
adaptive threshold if T=T(f,W(i,j))
if it depends on the window W where it is computed

Automatic thresholding

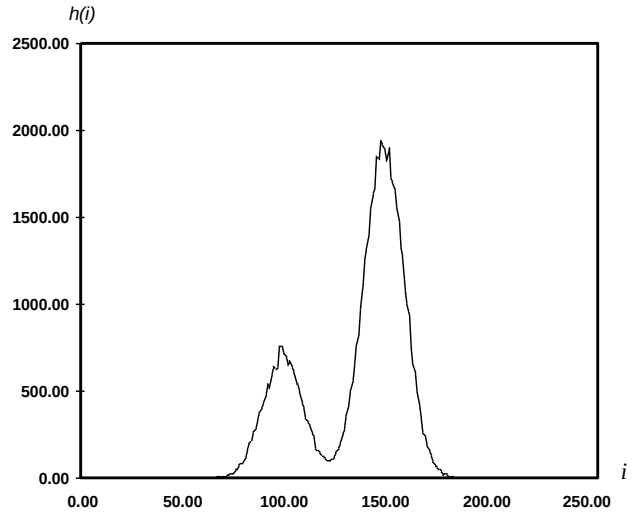
How to define T automatically?:

Without knowing the objects of interest, that is without any prior let's give the computer the chance of seeing, measuring the statistic proprieties of the histogram.

This is very unusual (in general we know what we are looking at). If it is not, the segmentation of the object w.r.t. the background can be done only in simple cases where there is a high contract and a bimodal histogram

Hp: the points of the target object have a grey level different from the background gray level (**bimodal histogram**)

➔ T must divide the two modes



Otsu thresholding

Goal: automatically identify the optimal threshold dividing 2 groups of luminosity intensity.

Otsu Threshold (1979):

Any threshold separates the histogram into 2 groups with each group having its own statistics (mean, variance).

The homogeneity of each group is measured by the within group variance. The optimum threshold is that threshold which minimizes the within group variance thus maximizing the homogeneity of each group.

Is it useful?

It is useful only if the assumption of bimodal histogram is true

◆Original

◆Otsu (89)

◆manual(128)

◆manual(64)

Othsu Algorithm

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Algorithm step-by-step:

1) Compute for a given threshold t:

q1(t), q2(t) the prior probabilities that a pixel belongs to group 1 and 2

$$q_1(t) = \sum_{i=1}^{t} P(i) \qquad q_2(t) = \sum_{i=t+1}^{L} P(i)$$

Where P(i) is the posteriori probability computed by histogram (the normalized histogram)

Given n pixels p of the image $P(i) = \#\{\boldsymbol{p}: I(\boldsymbol{p}) = i\}/n$

$$P(i) = H(i)/n$$

OTSU

2) Compute mean and variance

$$\mu_1(t) = \sum_{i=1}^{t} iP(i)/q_1(t) \qquad \sigma_1^2(t) = \sum_{i=1}^{t} [i-\mu_1(t)]^2 P(i)/q_1(t)$$

$$\mu_2(t) = \sum_{i=t+1}^{L} iP(i)/q_2(t) \qquad \sigma_2^2(t) = \sum_{i=t+1}^{L} [i-\mu_2(t)]^2 P(i)/q_2(t)$$

3) Compute the within group variance

$$\sigma_W^2(t) = q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t)$$

It is the **within group variance:** the smaller is the wgv, the more compact to the mean value are the two groups

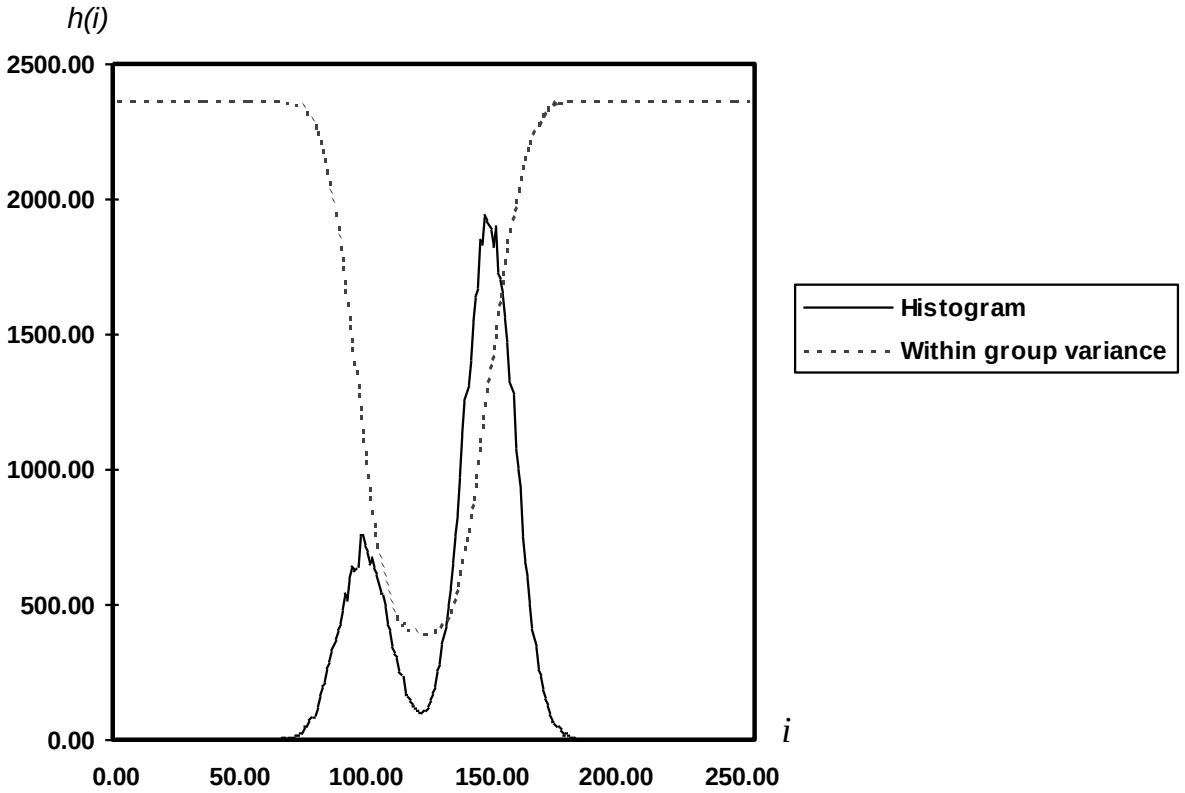
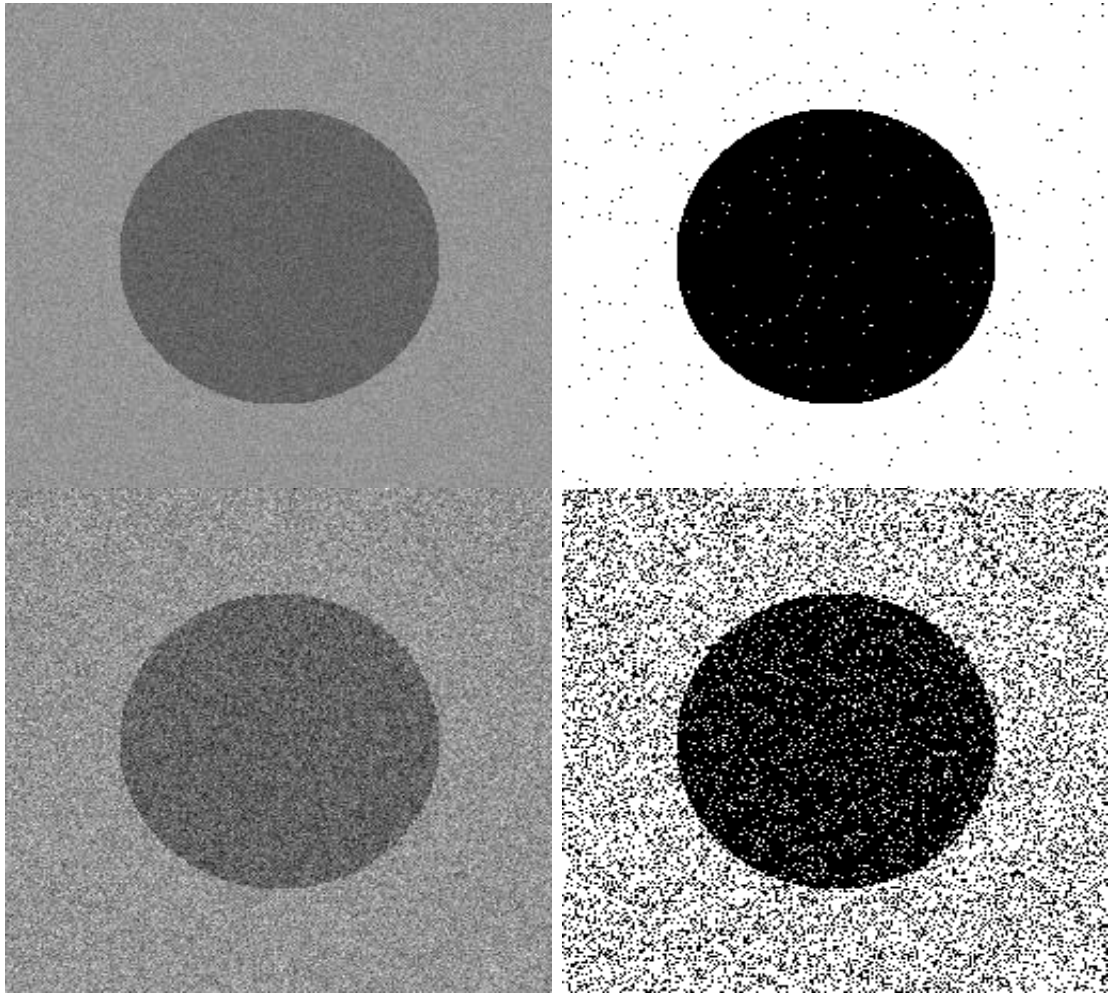
4) Minimization:

for each t=1..L , search the minimum of 3)

The wgv must be minimized: exaustive procedure

Only requires L=256 comparisons for and 8-bit grey level image

Otsu's Threshold example



T=124