Master in Computer Vision Barcelona

M1 – Images and Assessment of processing systems

Lecturer: Philippe Salembier, UPC

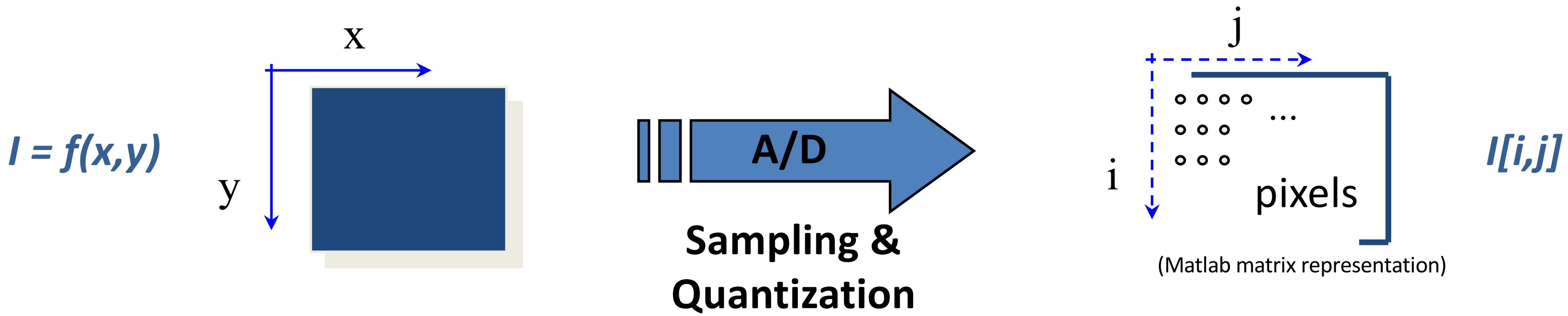
Outline

- **Introduction:**
  - **Image definition**
  - **Image processing systems**
- **User-oriented assessment:**
  - **Objective fidelity criteria**
- **Computer-oriented assessment:**
  - **Objective detection criteria**
- **Summary and Conclusions**

Introduction (I)

Image Definition

- Digital images are multidimensional matrices formed by a process of **sampling** and **quantization** of a continuous and analog function



Pixel = "picture element"

111	81	98	181	179	169	151	147	156	124	141	169	132	124
89	124	141	173	155	173	181	181	166	98	81	94	94	137
143	121	89	90	95	101	121	103	154	175	124	102	99	93
90	123	121	37	156	135	125	175	168	154	122	101	134	137
108	112	122	145	167	156	164	173	153	129	128	130	135	136
109	123	154	164	124	78	78	78	85	85	92	101	97	84

Master in Computer Vision Barcelona

Introduction (II)

Image Definition

- Different types of digital images depending on the nature of what their values represent:
  - **Light intensity**: Scalar value $I[i, j] \in N$
  - **Color**: Vector value (R,G,B) $I[i, j] \in N^3$
  - **Properties of materials**
    - X-ray images: Absorption
    - Ultra-sound images: Density
    - Infrared (IR) images: Temperature
    - Remote Sensing (RS) images: Reflectivity $I[i, j] \in N^B$

Introduction (III)

Image Definition

- Color images are defined in a, usually, 3-dimensional space:
  - **RGB space**: Red, Green and Blue components

R

206	241	195	123	214
222	215	151	237	210
240	238	175	205	255
211	212	158	222	202

G

89	155	152	99	187
96	122	104	211	182
112	138	115	167	223
999	123	98	175	156

B

71	132	120	61	142
81	104	78	178	142
101	122	89	131	181
88	107	74	145	120

- There exist other color spaces with different properties (e.g.):
  - **Perceptual spaces**: HSV (Hue, Saturation, Value), HSL (Luminance), YUV, etc.

Introduction (IV)

Image Processing Systems

Image (input) → Processor →

- **Image (output)**
  - **Enhancement**
  - **Restoration**
  - **Coding / Decoding**

  → **User**

- **Parameters (output)**
  - **Classification**
  - **Analysis**
  - **Modeling**

  → **Computer**

Computer Graphics: Parameters (input) → Image (output)

Introduction (V)

Design criteria for human observers / computers:

Design criteria for a human observer:

- Fidelity
  - Subjective (empirical values):
    - Excellent, Good, Fair, …
  - **Objective**
    - MSE, MAD
    - SNR, PSNR
  - **Objective and somewhat perceptual**
    - SSIM

Design criteria for a computer:

In the classification framework:

- **Precision, Recall**:
  - F measure
- **ROC curve:**
  - Working point
  - Equal Error Rate (EER)

Outline

- Introduction:
  - Image definition
  - Image processing systems
- **User-oriented assessment:**
  - **Objective fidelity criteria**
- **Computer-oriented assessment:**
  - **Objective detection criteria**
- **Summary and Conclusions**

Master in
Computer Vision
Barcelona

User oriented assessment (I)

Objective criteria for human:

- **Mean Square Error (MSE):**
  - Estimation of expectation

$$e[i,j] = I[i,j] - \hat{I}[i,j]$$

$$\sigma_{MSE}^2 \equiv \sigma_e^2 \equiv \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left|e[i,j]\right|^2$$

- **Mean Absolute Difference (MAD):**
  - Faster computation
  - Less sensitive to outliers

$$C_{MAD} \equiv \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left|e[i,j]\right|$$

- **Signal to Noise Ratio (SNR):**
  - Comparison of estimated powers

$$SNR(dB) = 10\log\frac{\sigma_{\hat{I}}^2}{\sigma_e^2}$$

Maximum value of representation

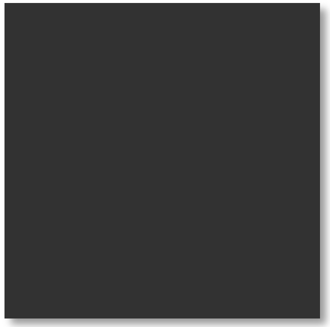
- **Peak Signal to Noise Ratio (PSNR):**
  - Maxv = Maximum possible peak-to-peak value of the representation

$$PSNR(dB) = 10\log\frac{Maxv^2}{\sigma_e^2}$$

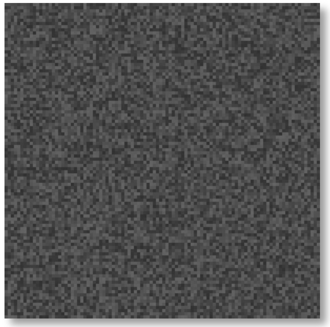
User oriented assessment (I)

SNR versus PSNR:

$$SNR(dB) = 10\log\frac{\sigma_{\hat{I}}^2}{\sigma_e^2} \qquad PSNR(dB) = 10\log\frac{Maxv^2}{\sigma_e^2}$$



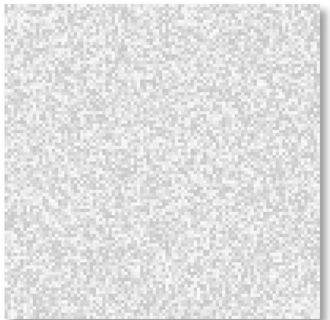
Processing
$\sigma_e = 14$



SNR = -3.6dB

PSNR = -29.1dB

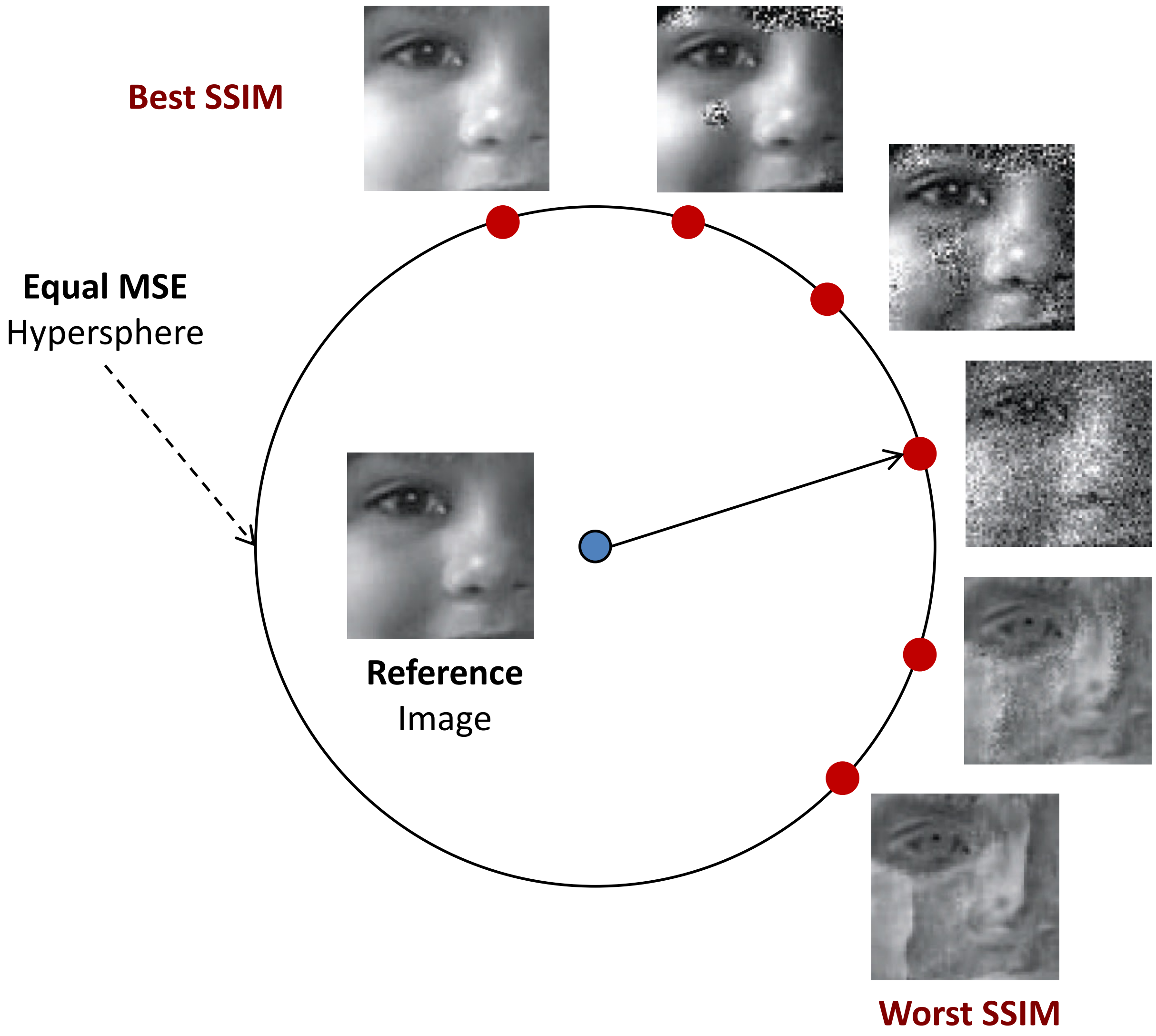
Processing
$\sigma_e = 14$



SNR = -1.0dB

PSNR = -29.1dB

User oriented assessment (III)



Objective and perceptual criteria for human: Researchers are continuously looking for objective measures that model better the subjective behavior of the Human Visual System.

For instance, the **Structural SIMilarity Index** (**SSIM**).

Z. Wang and A.C. Bovik, " Mean Square Error: Love it or Leave it?" *IEEE Signal Processing Magazine*, pp. 98 - 117, January 2009.

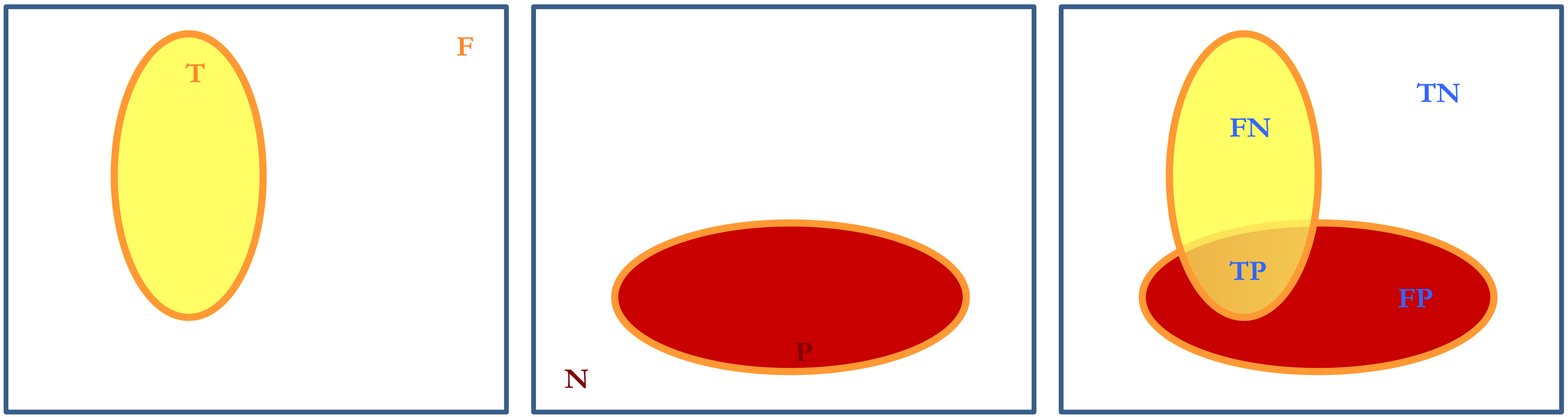
Outline

- Introduction:
  - Image definition
  - Image processing systems
- User-oriented assessment:
  - Objective fidelity criteria
- **Computer-oriented assessment:**
  - **Objective detection criteria**
- **Summary and Conclusions**

Computer-oriented assessment (I)

Problem statement: Given a population of samples (images) that can be classified into two groups, we want to **assess the quality** of an (several) automatic classifier(s). Towards this goal, we have a **ground truth** which is a set of elements that have been manually annotated.

- In the context of classification, several assessment criteria can be defined based on a **few parameters**



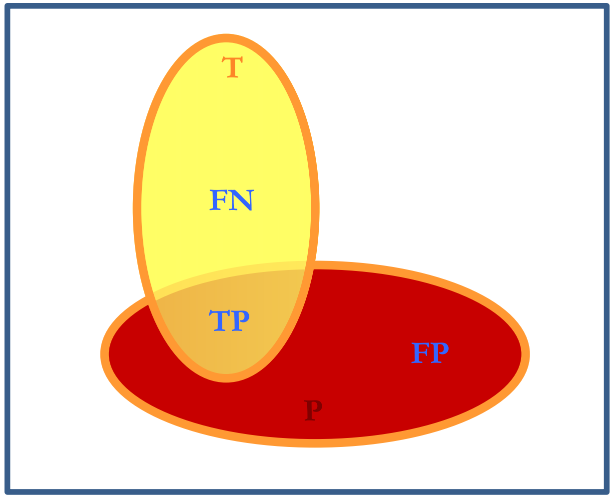
Ground Truth classification:
T: True samples
F: False samples

Automatic classification:
P (Positive): Detected samples
N (Negative): Rejected samples

Assessment result:
TP (True Positive): Correctly detected samples
TN (True Negative): Non-detected false samples
FN (False Negative): Non-detected true samples
FP (False Positive): False detected samples

Computer-oriented assessment (II)

Given the previous parameters, several assessment measures can be defined.



Basic parameters:
T: True samples (ground truth)
P (Positive): Detected samples (automatic)
FN (False Negative): Non-detected samples
TP (True Positive): Correctly detected samples
FP (False Positive): False detected samples

- **Precision:** Fraction of the **detected** samples (P) that have been correctly detected (TP)

$$Precision = \frac{TP}{P} \rightarrow 1$$

- **Recall:** Fraction of the **true** samples (T) that have been correctly detected (TP)

$$Recall = \frac{TP}{T} \rightarrow 1$$

- **Accepted outliers (Fall-out):** Fraction of the false samples (F) that have been incorrectly detected (FP)

$$AO = \frac{FP}{F} \rightarrow 0$$

- **% of false detections:** Fraction of the detected samples (P) that have been incorrectly detected (FP)

$$\%FD = \frac{FP}{P} \rightarrow 0$$

Definition (classification context) [edit]

For classification tasks, the terms **true positives**, **true negatives**, **false positives**, and **false negatives** (see also Type I and type II errors) compare the results of the classifier under test with trusted external judgments. The terms *positive* and *negative* refer to the classifier's prediction (sometimes known as the *expectation*), and the terms *true* and *false* refer to whether that prediction corresponds to the external judgment (sometimes known as the *observation*).

Let us define an experiment from **P** positive instances and **N** negative instances for some condition. The four outcomes can be formulated in a 2×2 *contingency table* or *confusion matrix*, as follows:

		Condition (as determined by "Gold standard")			
	Total population	Condition positive	Condition negative	Prevalence = Σ Condition positive / Σ Total population	
Test outcome	Test outcome positive	**True positive**	**False positive** (Type I error)	Positive predictive value (PPV, Precision) = Σ True positive / Σ Test outcome positive	False discovery rate (FDR) = Σ False positive / Σ Test outcome positive
	Test outcome negative	**False negative** (Type II error)	**True negative**	False omission rate (FOR) = Σ False negative / Σ Test outcome negative	Negative predictive value (NPV) = Σ True negative / Σ Test outcome negative
	Positive likelihood ratio (**LR+**) = TPR/FPR	True positive rate (TPR, Sensitivity, Recall) = Σ True positive / Σ Condition positive	False positive rate (FPR, Fall-out) = Σ False positive / Σ Condition negative	Accuracy (ACC) = Σ True positive + Σ True negative / Σ Total population	
	Negative likelihood ratio (**LR−**) = FNR/TNR	False negative rate (FNR) = Σ False negative / Σ Condition positive	True negative rate (TNR, Specificity, SPC) = Σ True negative / Σ Condition negative		
	Diagnostic odds ratio (**DOR**) = LR+/LR−				

Terminology and derivations from a confusion matrix

true positive (TP)
eqv. with hit
true negative (TN)
eqv. with correct rejection
false positive (FP)
eqv. with false alarm, Type I error
false negative (FN)
eqv. with miss, Type II error

sensitivity or true positive rate (TPR)
eqv. with hit rate, recall
$$\boxed{TPR = TP/P = TP/(TP+FN)}$$
specificity (SPC) or True Negative Rate
$$SPC = TN/N = TN/(FP+TN)$$
precision or positive predictive value (PPV)
$$\boxed{PPV = TP/(TP+FP)}$$
negative predictive value (NPV)
$$NPV = TN/(TN+FN)$$
fall-out or false positive rate (FPR)
$$\boxed{FPR = FP/N = FP/(FP+TN)}$$
false discovery rate (FDR)
$$\boxed{FDR = FP/(FP+TP) = 1 - PPV}$$

accuracy (ACC)
$$ACC = (TP+TN)/(P+N)$$
F1 score
is the harmonic mean of precision and sensitivity
$$F1 = 2TP/(2TP+FP+FN)$$
Matthews correlation coefficient (MCC)
$$\frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Source: Fawcett (2006).[4]

Computer-oriented assessment (IIa). Example

Processing

OCR

PME
p GENE RATIOV
MITTER ANB
j O c

T (True) Letters to be recognized

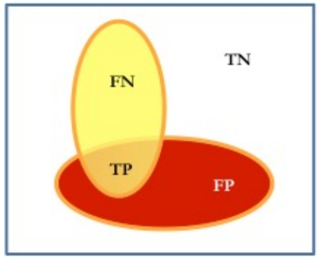
P	M	E		G	E	N	E	R	A	T	I	O	N	M	I	T	T	E	R	R	A	N	D				23

P (Positive) Recognized letters

P	M	E	p	G	E	N	E	R	A	T	I	O	V	M	I	T	T	E	R		A	N	B	j	O	c	26

TP = T∩P (True positive)

P	M	E		G	E	N	E	R	A	T	I	O		M	I	T	T	E	R		A	N					20



Computer-oriented assessment (IIb). Example

Processing

OCR

PME
p GENE RATIOV
MITTER ANB
j O c

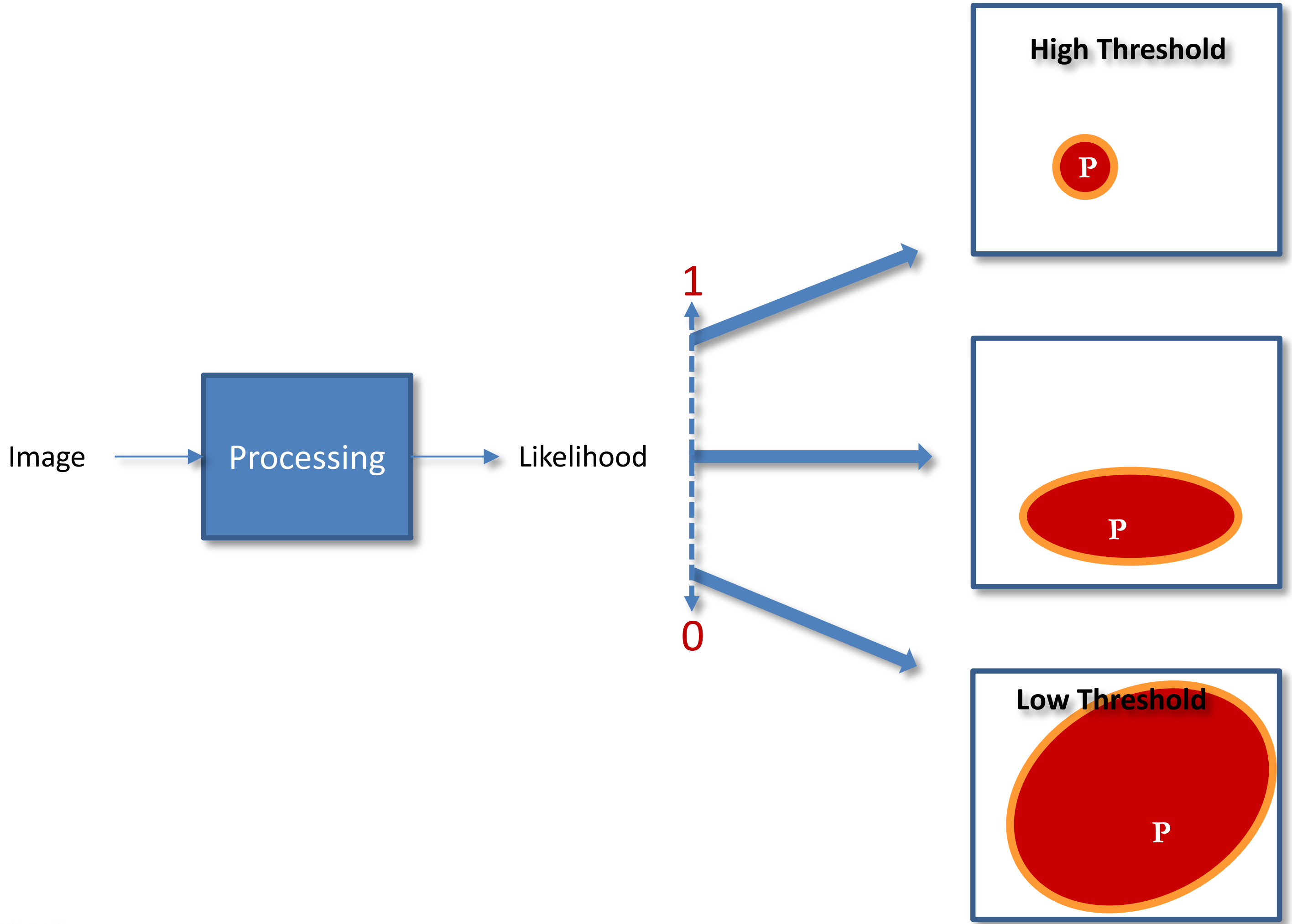
T = 23
P = 26
TP = 20

$$\text{Precision} = \frac{TP}{P} = \frac{TP}{TP+FP} = \frac{20}{26} \approx 0.77 \Rightarrow 77\%$$

$$\text{Recall} = \text{true positive rate} = \frac{TP}{T} = \frac{TP}{TP+FN} = \frac{20}{23} \approx 0.87 \Rightarrow 87\%$$

Computer-oriented assessment (III)

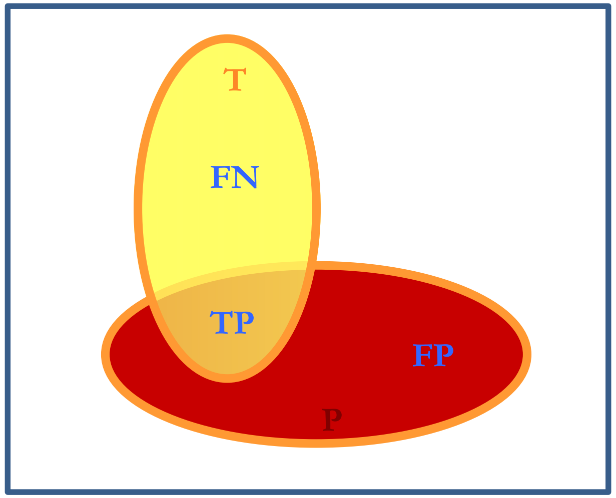
A given classifier depends on a **detection threshold** over an estimated parameter that, typically, may assess either a distance or a similarity. If we assume that a **similarity is used**, above the detection threshold value the sample is accepted as being an element of the given class.

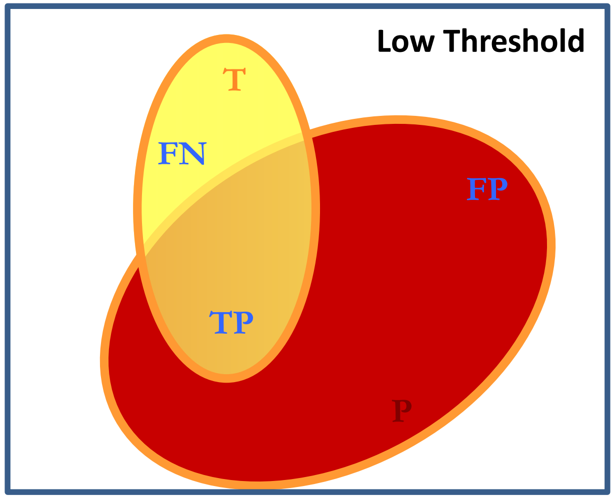


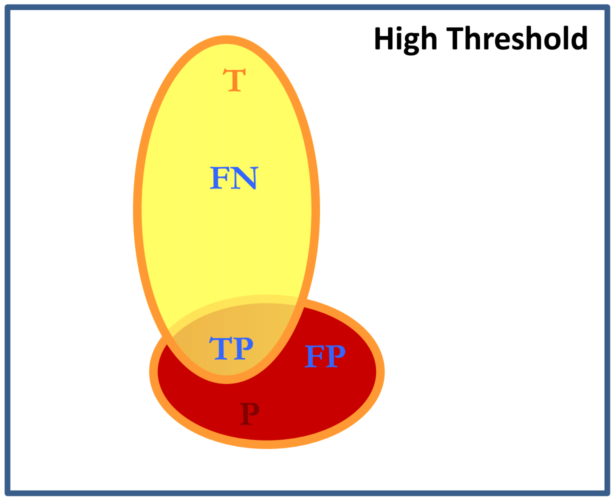
Computer-oriented assessment (III)

A given classifier depends on a **detection threshold** over an estimated parameter that, typically, may assess either a distance or a similarity. If we assume that a **similarity is used**, below (above) the detection threshold value the sample is rejected (accepted) as being an element of the given class.

The **performance** of the classifier should be evaluated taking into account jointly two opposite assessment criteria. These criteria values change when the detection threshold changes .





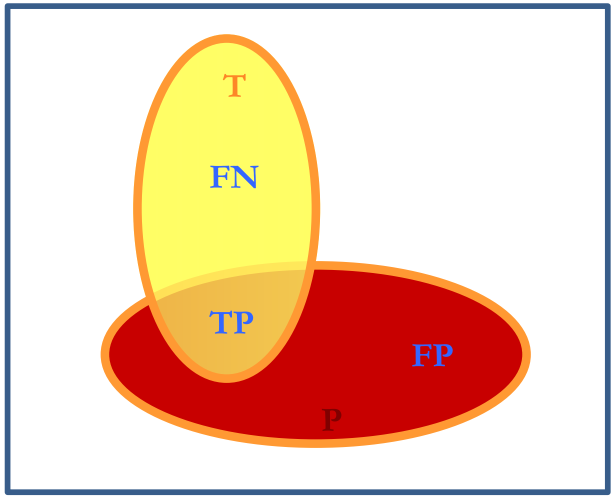


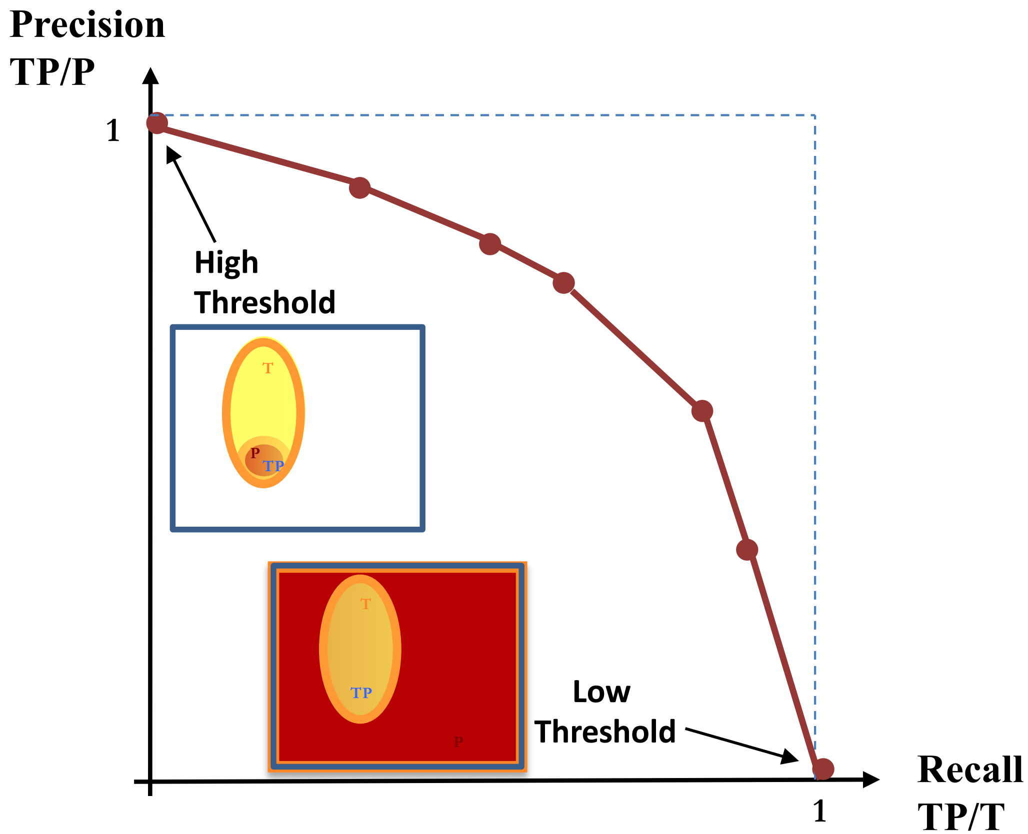
As the detection threshold is reduced, detected samples (P), correct detections (TP) and false detections (FP) increase but in a different way. There is a trade-off.

Computer-oriented assessment (IV)

Precision & Recall curve:

To **assess a classifier**, its behavior for all possible detection threshold values has to be studied. For each threshold value, the Precision and recall are computed and plotted, forming a curve.





$$\frac{TP}{T} = \frac{True\,Positives}{True\,Samples}$$ Recall

$$\frac{TP}{P} = \frac{True\,Positives}{Positive\,Samples}$$ Precision

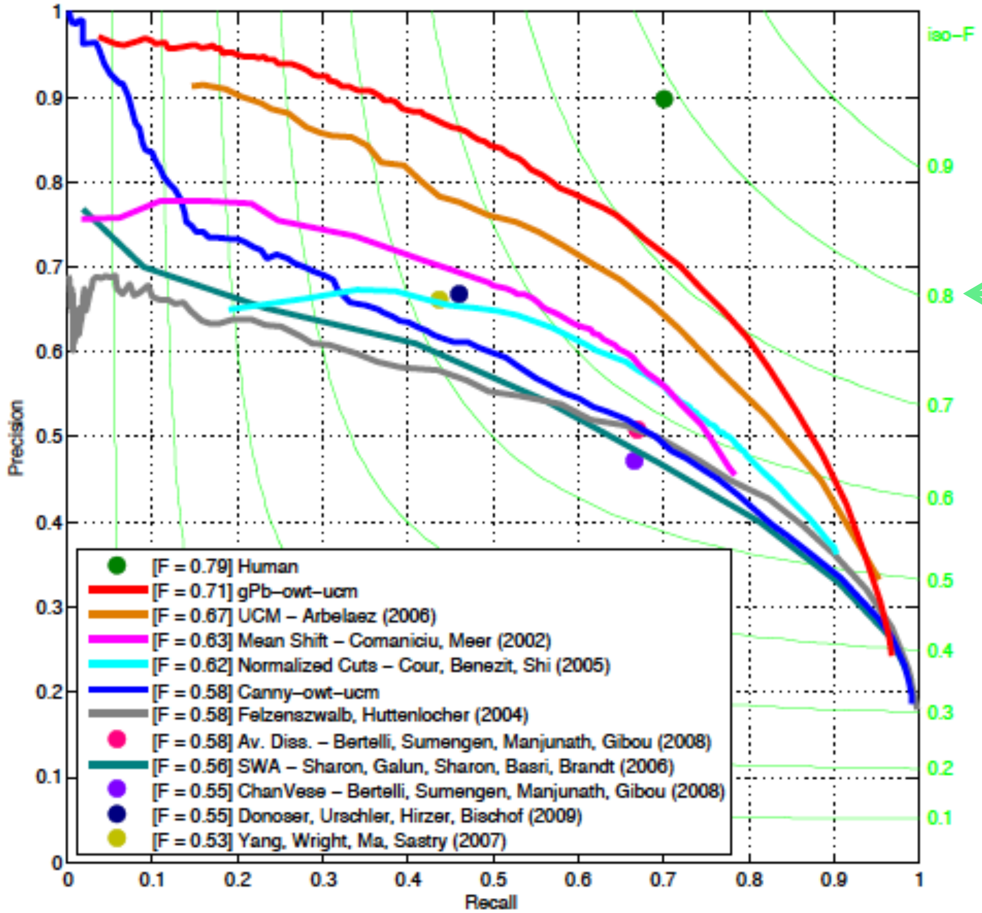
***F* measure:** the maximum of the harmonic mean can be used to set the threshold value:

$$F_{measure} = 2\frac{Prec \cdot Rec}{Prec + Rec}$$

Computer-oriented assessment (V)

Precision – Recall

Example



Iso-F curves

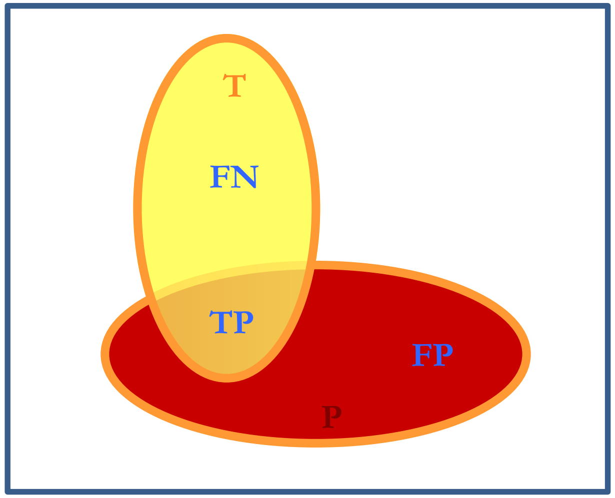
Fig. 2. **Evaluation of segmentation algorithms on the BSDS300 Benchmark.** Paired with our gPb contour

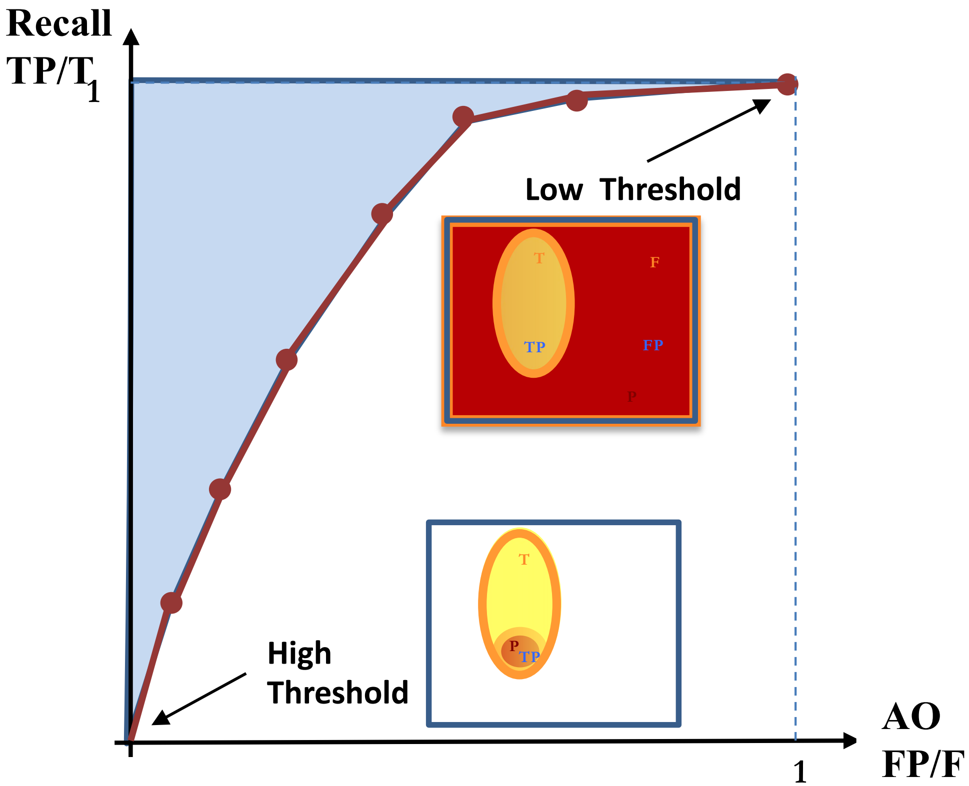
(from: *Contour Detection and Hierarchical Image Segmentation* P. Arbelaez, M. Maire, C. Fowlkes and J. Malik. IEEE TPAMI, Vol. 33, No. 5, pp. 898-916, May 2011)

Computer-oriented assessment (VI)

Receiver Operating Characteristics (ROC):

This curve generally represents the recall value as a function of the accepted outliers (but variations exist). The ideal system will have recall≈1, AO≈0.





$$\frac{TP}{T} = \frac{True\,Positives}{True\,Samples}$$ Recall

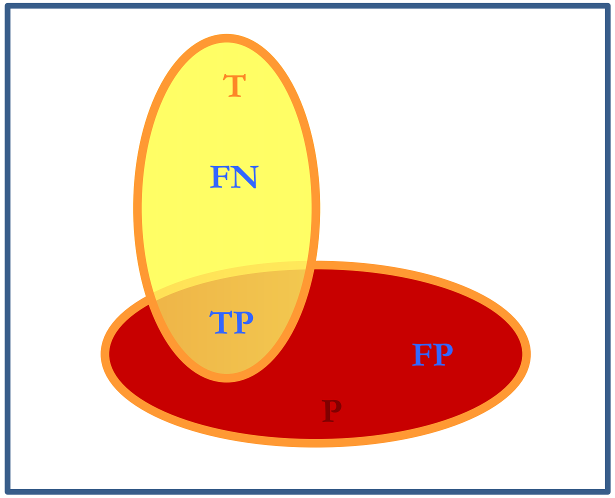
$$\frac{FP}{F} = \frac{False\,Positives}{False\,Samples}$$ AO

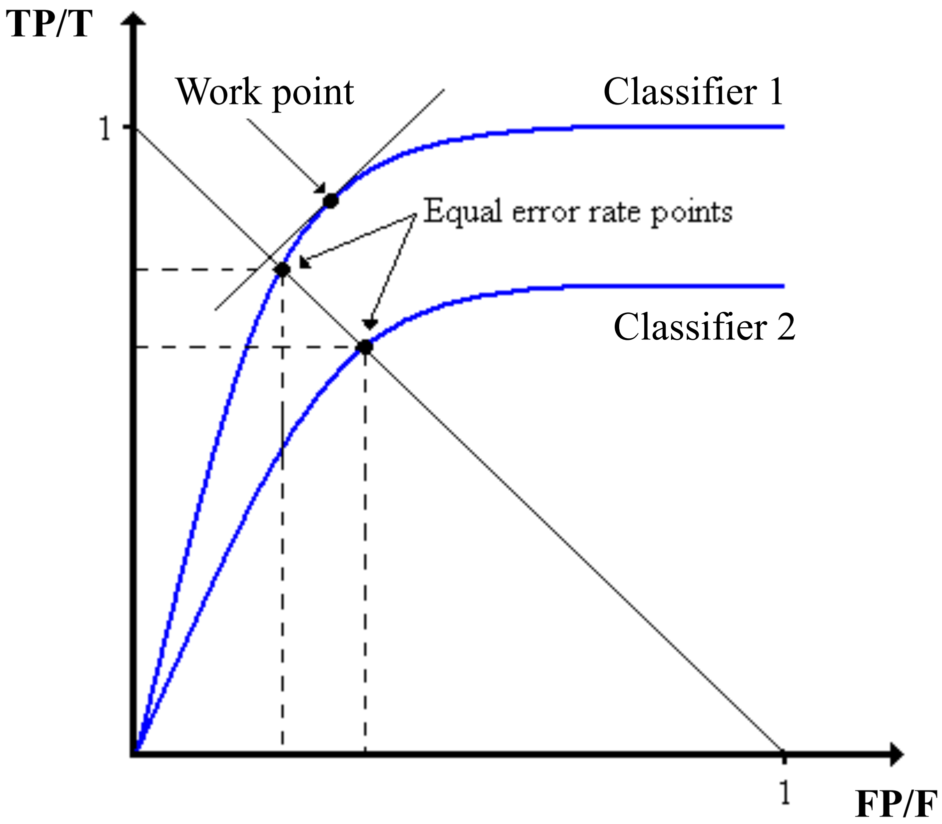
A classifier performance can be computed using the **area above the curve**: the smaller, the better.

Computer-oriented assessment (VII)

Receiver Operating Characteristics (ROC):

Given a classifier, **to select the threshold value** to be used there exist different approaches. Depending on the requirements of the application, one approach can be more suitable than the others.





- **Work point:** curve slope = 1. For smaller threshold values increases larger in FP than TP.
- **Equal error rate (EER):** FN/T equals FP/F (small value for good detectors).

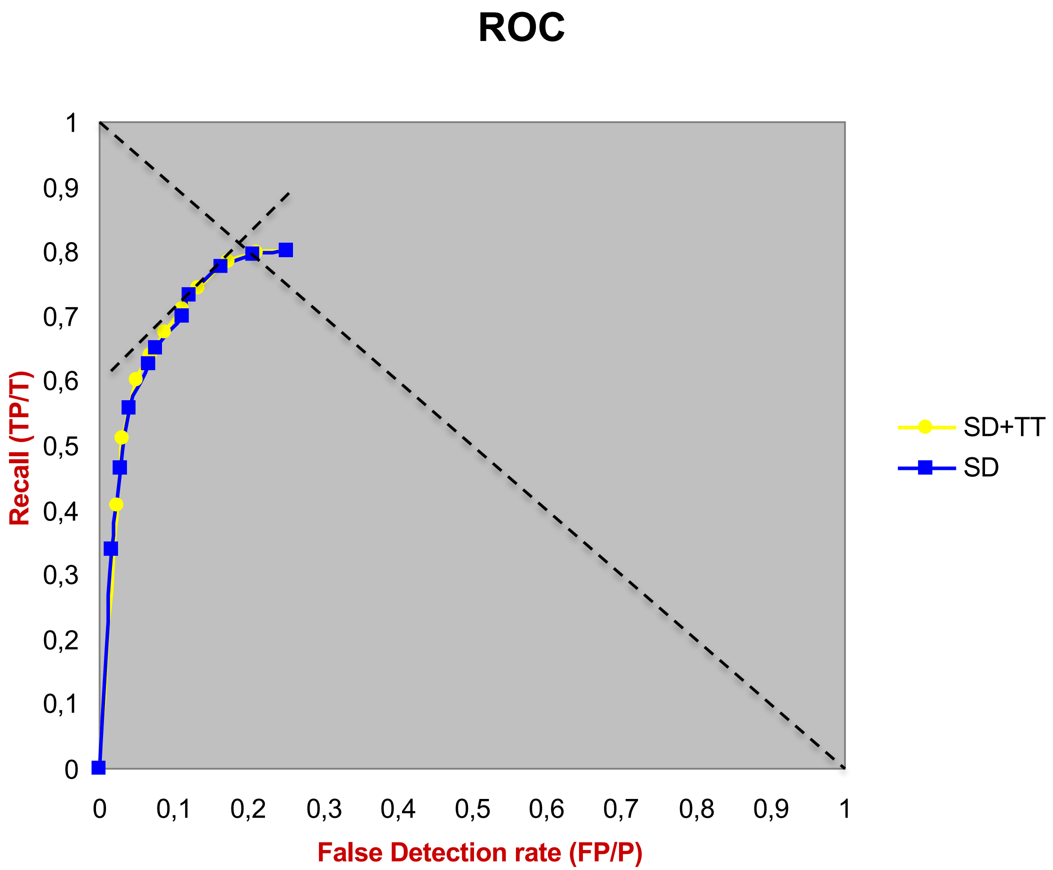
$$\frac{TP}{T} + \frac{FP}{F} = 1$$

$$\frac{T - FN}{T} + \frac{FP}{F} = 1$$

$$\frac{FP}{F} = \frac{FN}{T}$$

Computer-oriented assessment (VIII)

Example 1: Estimating density of people in a crowd (camera 1)



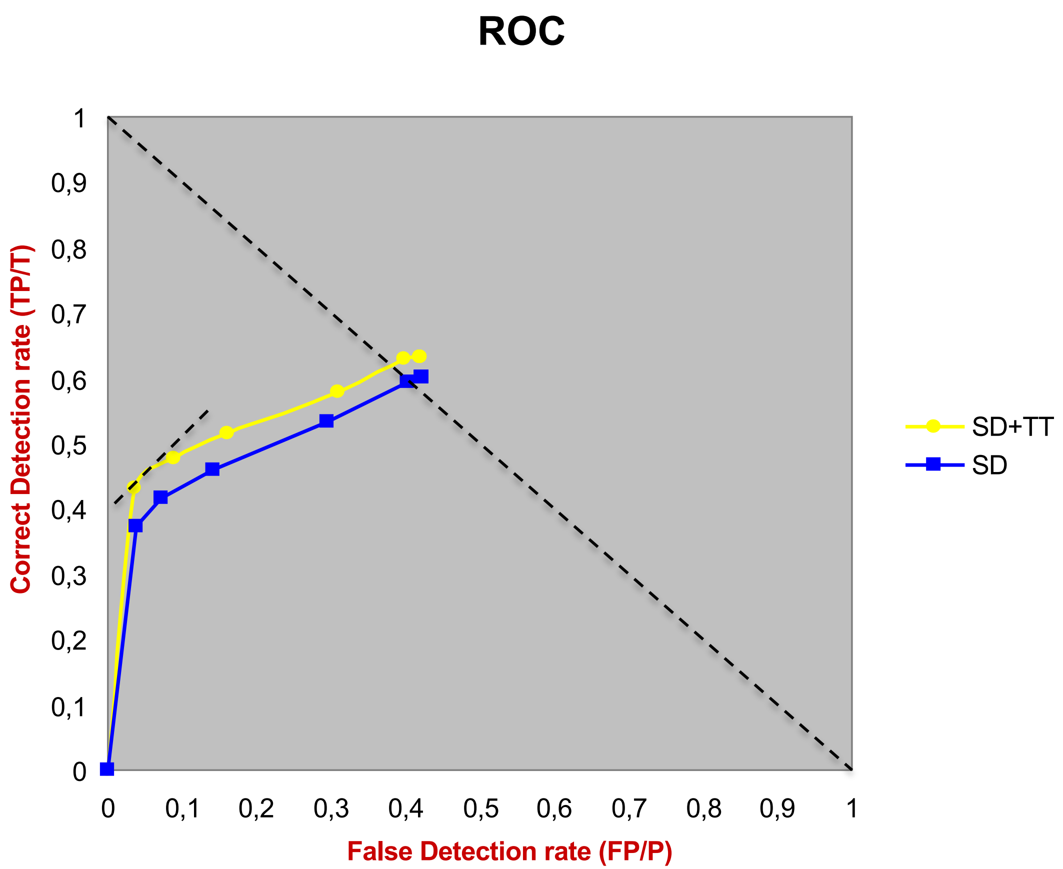
Aleix Puig, Pere Puig, *Estimació de la densitat de persones en multituds,* PFC ETSETB/UPC, 2003/04

- **Work point**
  - TP/T = 74%
  - FP/P = 13%
- **EER**
  - TP/T = 79%
  - FP/P = 20%

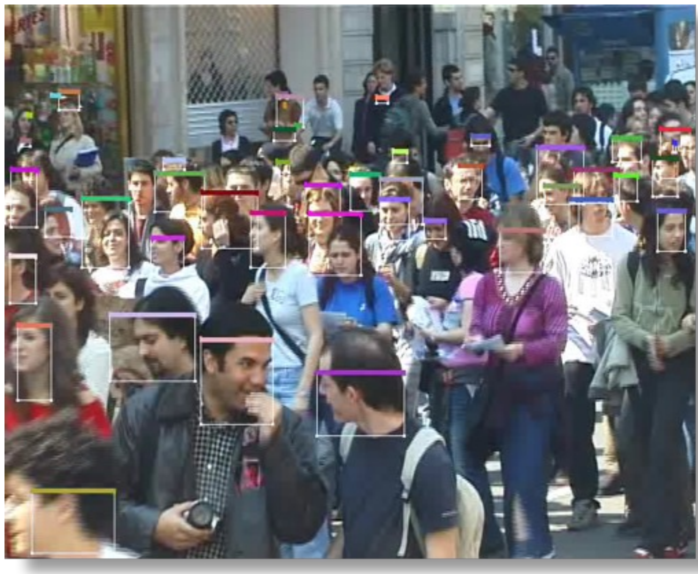
Master in Computer Vision Barcelona

Computer-oriented assessment (IX)

Example 1: Estimating density of people in a crowd (camera 2)

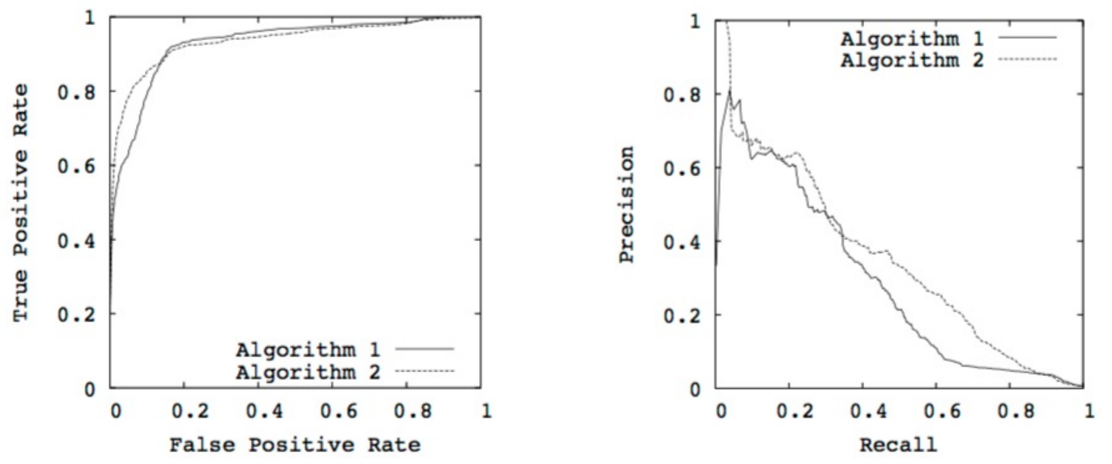


- **Work point**
  - TP/T = 48%
  - FP/P = 9%
- **EER**
  - TP/T = 61%
  - FP/P = 37%



Aleix Puig, Pere Puig, *Estimació de la densitat de persones en multituds,* PFC ETSETB/UPC, 2003/04

Computer-oriented assessment (X)



(a) Comparison in ROC space

(b) Comparison in PR space

Essentially equivalent representations but give a different perception in particular in case of highly skewed classes

Davis, J., Burnside, E., Dutra, I., Page, D., Ramakrishnan, R., Costa, V. S., & Shavlik, J. (2005). View learning for statistical relational learning: With an application to mammography. Proceeding of the 19th International Joint Conference on Artificial Intelligence. Edinburgh, Scotlan

Outline

- Introduction:
  - Image definition
  - Image processing systems
- User-oriented assessment:
  - Objective fidelity criteria
- Computer-oriented assessment:
  - Objective detection criteria
- **Summary and Conclusions**

Summary and Conclusions

- Image processing systems are **assessed with different parameters** depending of the application:
  - **Human-oriented** versus **computer-oriented** applications
- In human-oriented applications, measures should take into account **human perception**.
- In **classification** (computer-oriented) **applications**, classifier's performance depend on a threshold:
  - The classifier has to be assessed over the **whole set of threshold values**
  - The selected threshold value **depends on the application**