

# VISIONE

## CORSO DI SISTEMI INTELLIGENTI, CAPITOLO 24

# Outline

- ◇ Percezione in generale
- ◇ Formazione dell'immagine
- ◇ Visione a basso livello
- ◇ 2D → 3D
- ◇ Riconoscimento degli oggetti

# Percezione in generale

Stimolo (percezione)  $S$ , Mondo  $W$

$$S = g(W)$$

P.e.,  $g$  = “computer graphics”. Si può realizzare la visione come l’inverso della computer graphics ?

$$W = g^{-1}(S)$$

# Percezione in generale

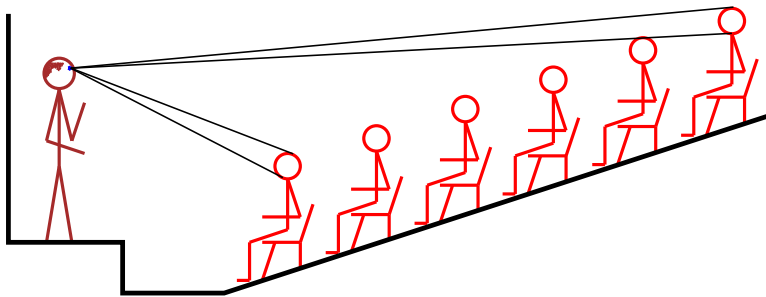
Stimolo (percezione)  $S$ , Mondo  $W$

$$S = g(W)$$

P.e.,  $g$  = “computer graphics”. Si può realizzare la visione come l’inverso della computer graphics ?

$$W = g^{-1}(S)$$

Problema: notevole ambiguità!



# Percezione in generale

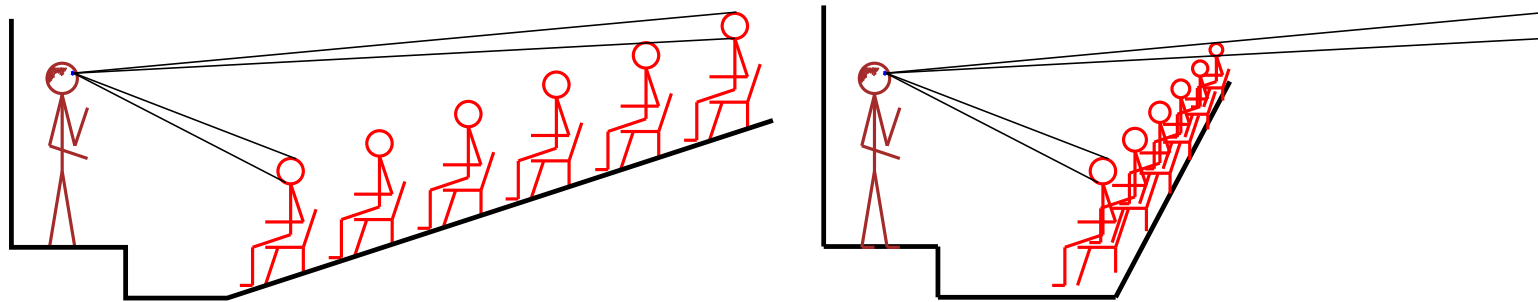
Stimolo (percezione)  $S$ , Mondo  $W$

$$S = g(W)$$

P.e.,  $g$  = “computer graphics”. Si può realizzare la visione come l’inverso della computer graphics ?

$$W = g^{-1}(S)$$

Problema: notevole ambiguità!



# Percezione in generale

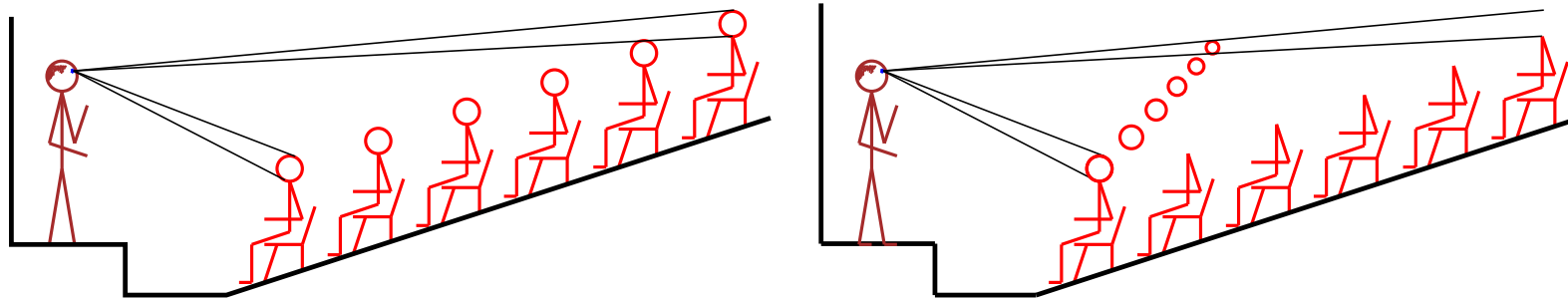
Stimolo (percezione)  $S$ , Mondo  $W$

$$S = g(W)$$

P.e.,  $g$  = “computer graphics”. Si può realizzare la visione come l’inverso della computer graphics ?

$$W = g^{-1}(S)$$

Problema: notevole ambiguità!



# Approcci migliori

Inferenza bayesiana sulle configurazioni del mondo:

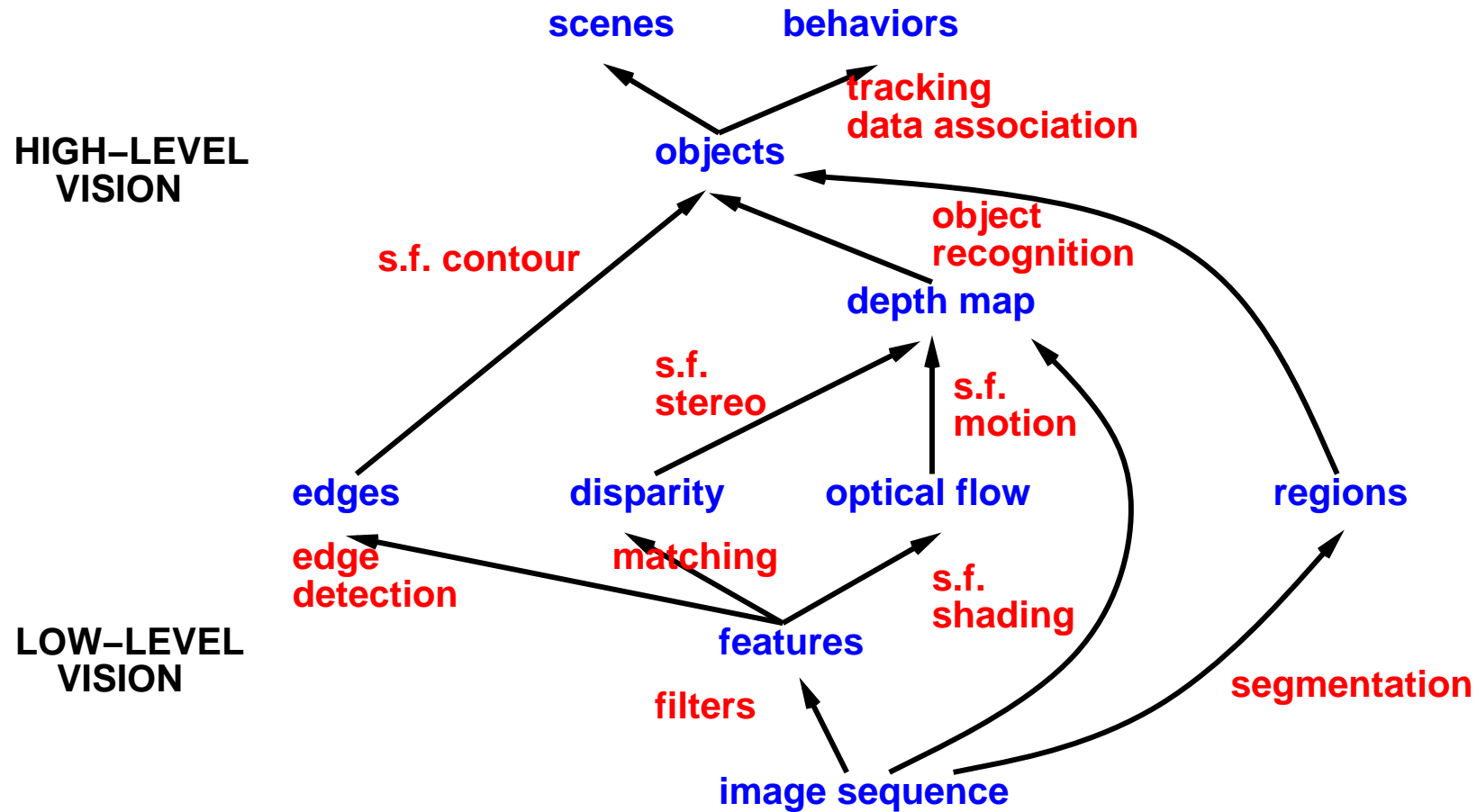
$$P(W|S) = \alpha \underbrace{P(S|W)}_{\text{“graphics”}} \underbrace{P(W)}_{\text{“conoscenza a priori”}}$$

Ancora meglio: nessun bisogno di recuperare la scena esatta!

Estrarre solo l'informazione necessaria per

- navigazione
- manipolazione
- riconoscimento/identificazione

# “Sottosistemi” della visione



La visione richiede la combinazione di vari indizi



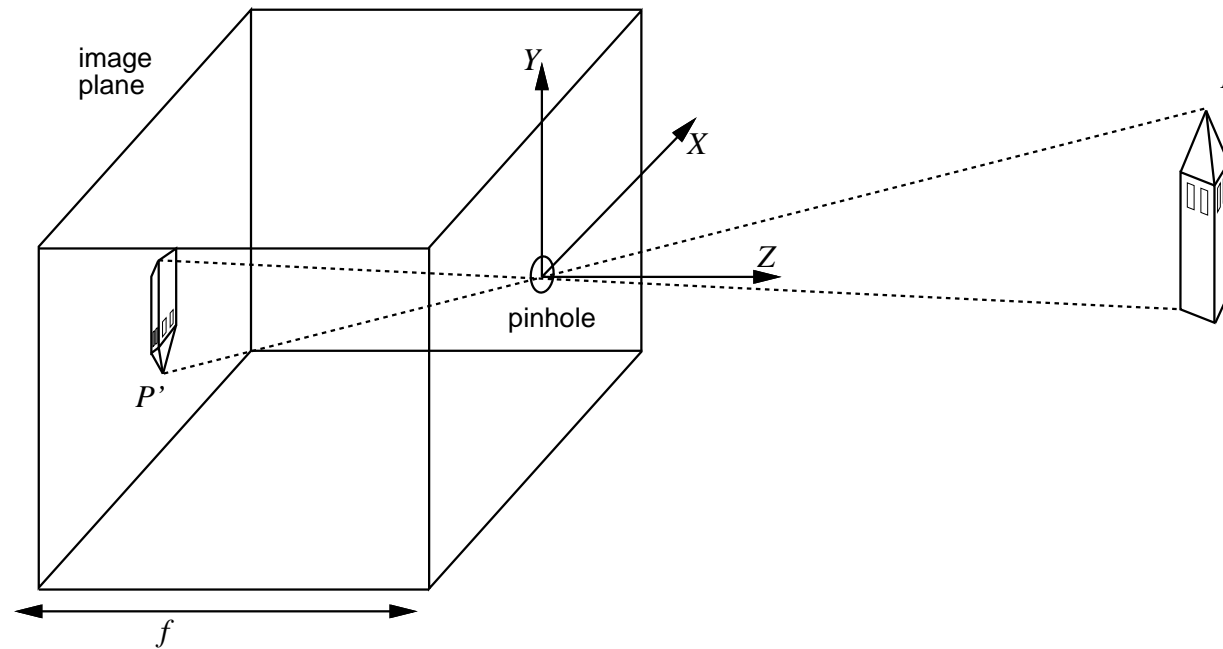
**La visione dipende dalla conoscenza a priori!**



**La visione dipende dalla conoscenza a priori!**



# Formazione dell'immagine



$P$  è un punto nella scena, con coordinate  $(X, Y, Z)$

$P'$  è la sua immagine sul piano dell'immagine, con coordinate  $(x, y, z)$

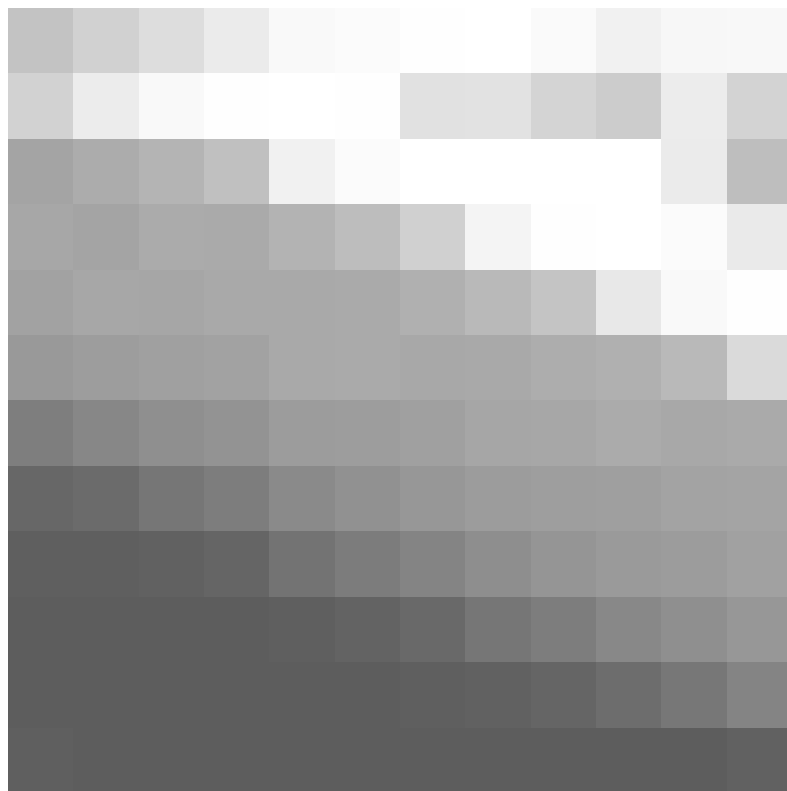
$$x = \frac{-fX}{Z}, \quad y = \frac{-fY}{Z}$$

Il rapporto scala/distanza è indeterminato!

# Immagini



# Immagini



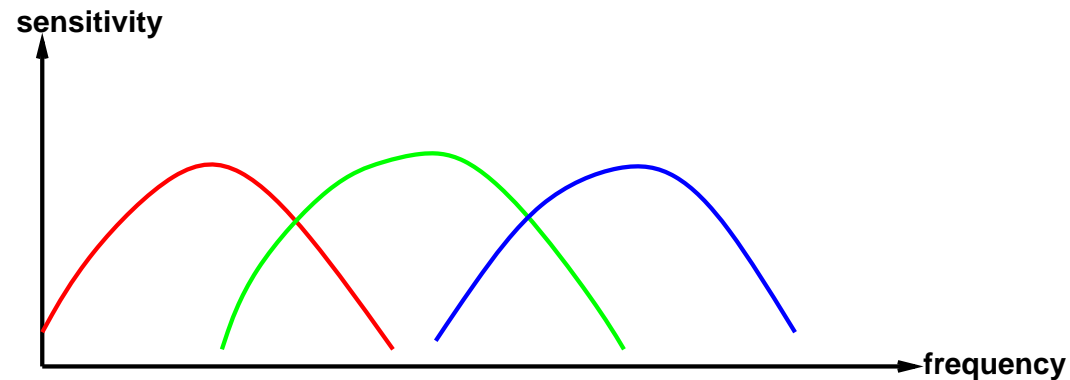
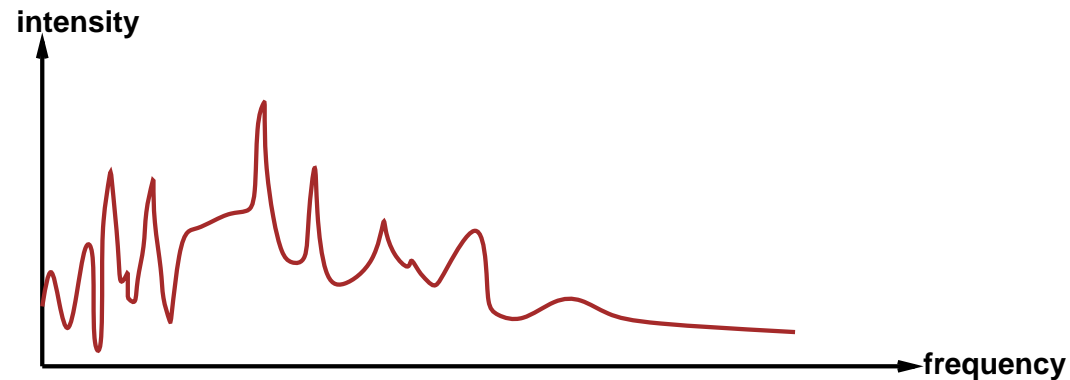
195	209	221	235	249	251	254	255	250	241	247	248
210	236	249	254	255	254	225	226	212	204	236	211
164	172	180	192	241	251	255	255	255	255	235	190
167	164	171	170	179	189	208	244	254	255	251	234
162	167	166	169	169	170	176	185	196	232	249	254
153	157	160	162	169	170	168	169	171	176	185	218
126	135	143	147	156	157	160	166	167	171	168	170
103	107	118	125	133	145	151	156	158	159	163	164
095	095	097	101	115	124	132	142	117	122	124	161
093	093	093	093	095	099	105	118	125	135	143	119
093	093	093	093	093	093	095	097	101	109	119	132
095	093	093	093	093	093	093	093	093	093	093	119

$I(x, y, t)$  è l'intensità in  $(x, y)$  al tempo  $t$

CCD camera  $\approx 1,000,000$  pixels; occhio umano  $\approx 240,000,000$  pixels  
cioè, 0.25 terabits/sec

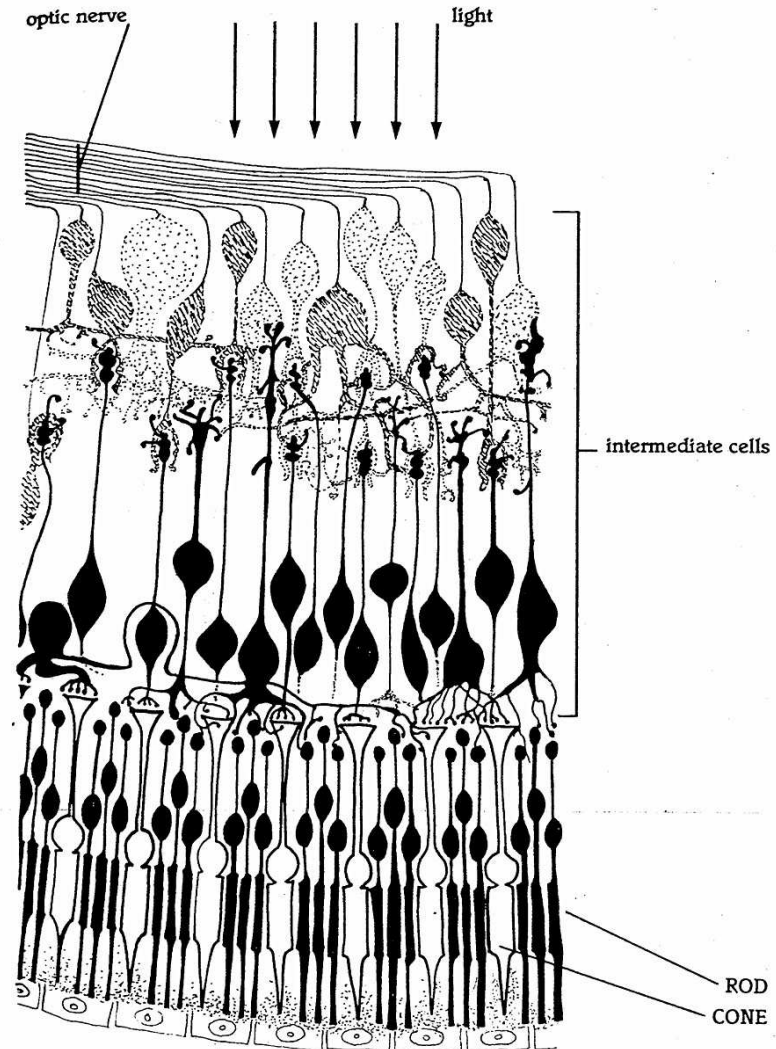
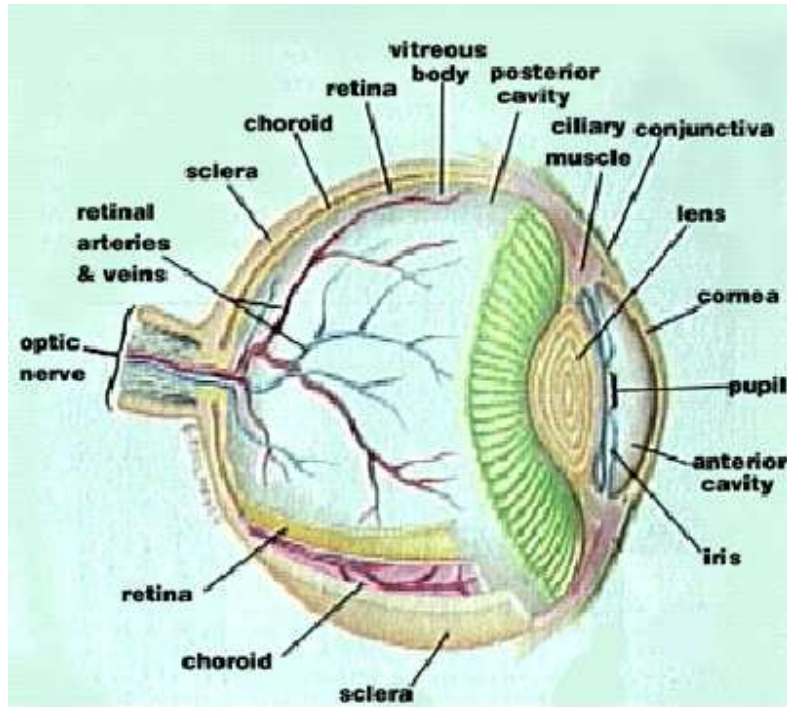
# Visione a colori

L'intensità varia con la frequenza  $\rightarrow$  segnale a dimensionalità infinita

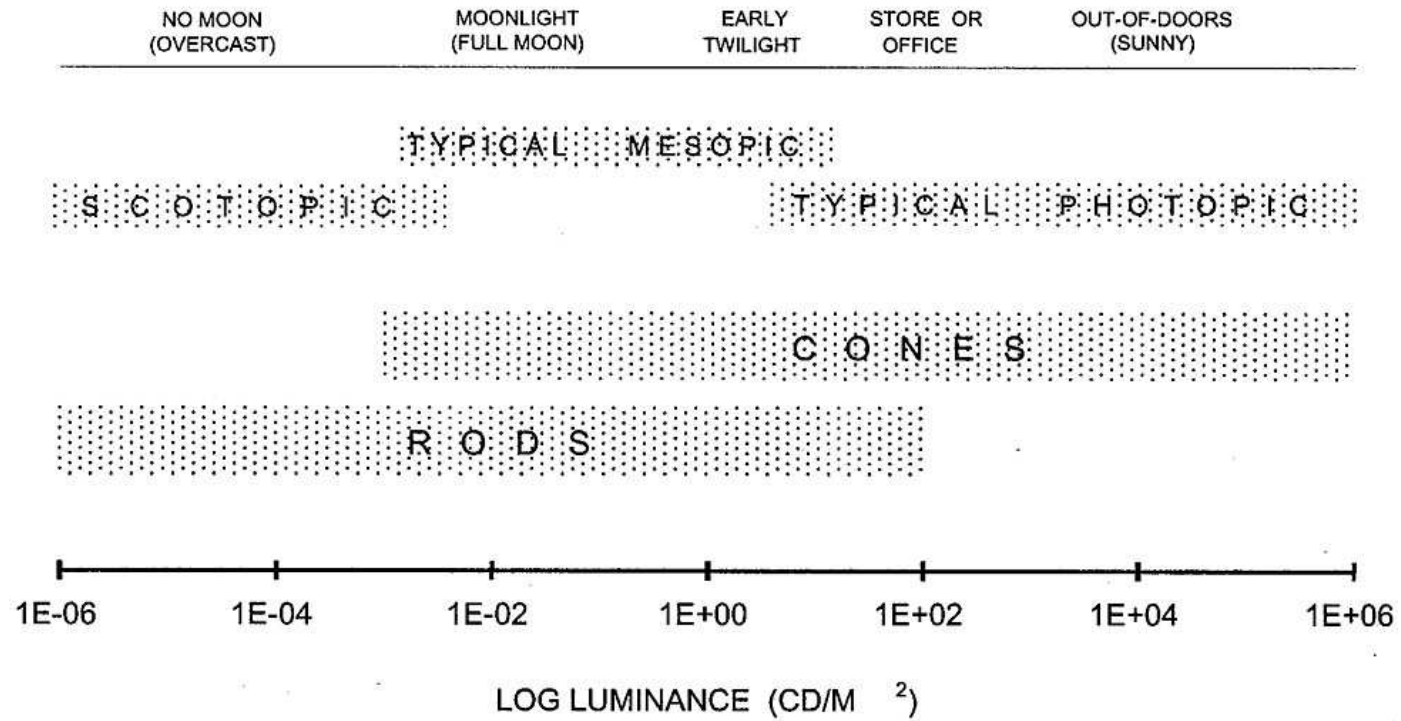


L'occhio umano ha vari tipi di cellule sensibili al colore;  
ognuna integra il segnale  $\Rightarrow$  vettore di intensità a 3 elementi

# Occhio e retina

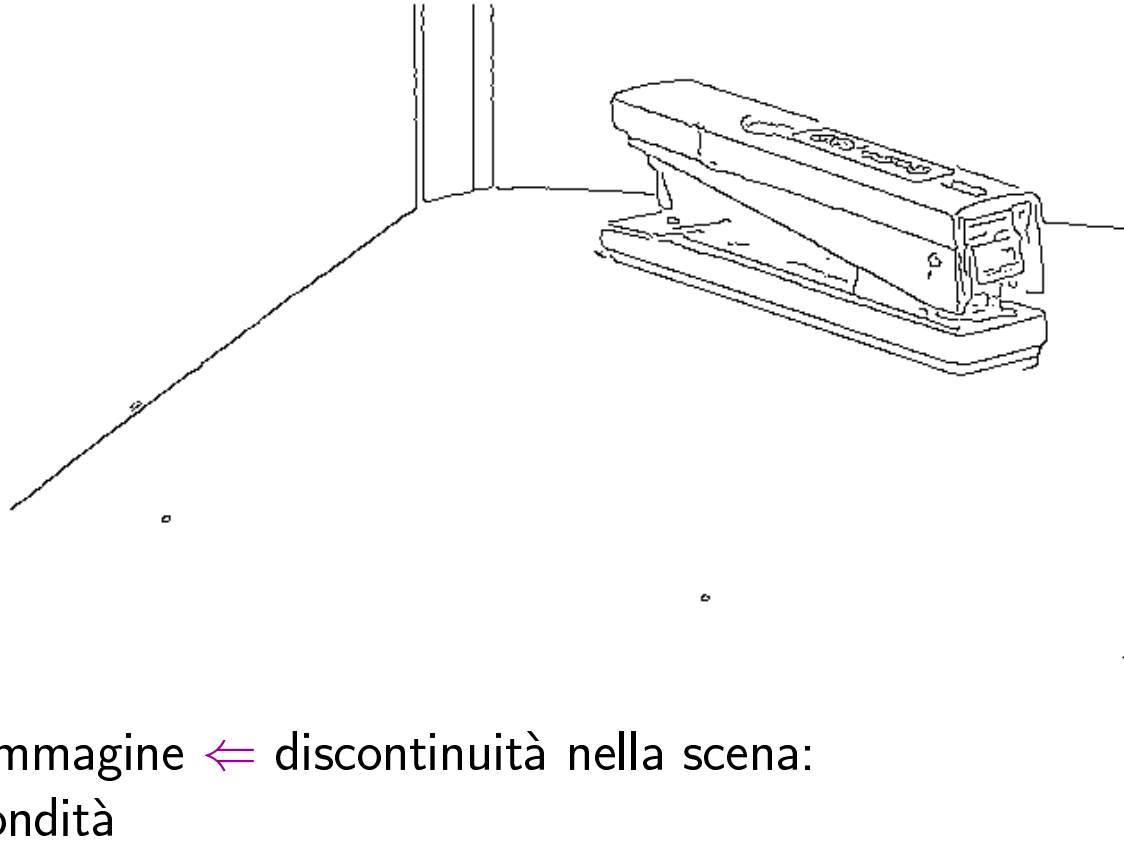


# La retina





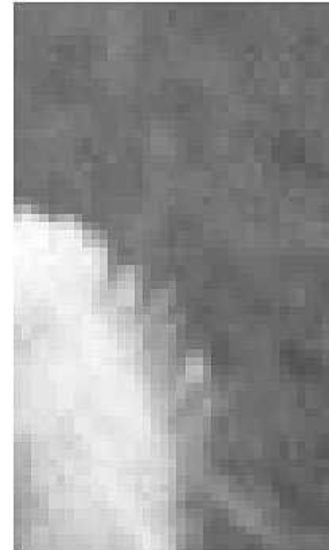
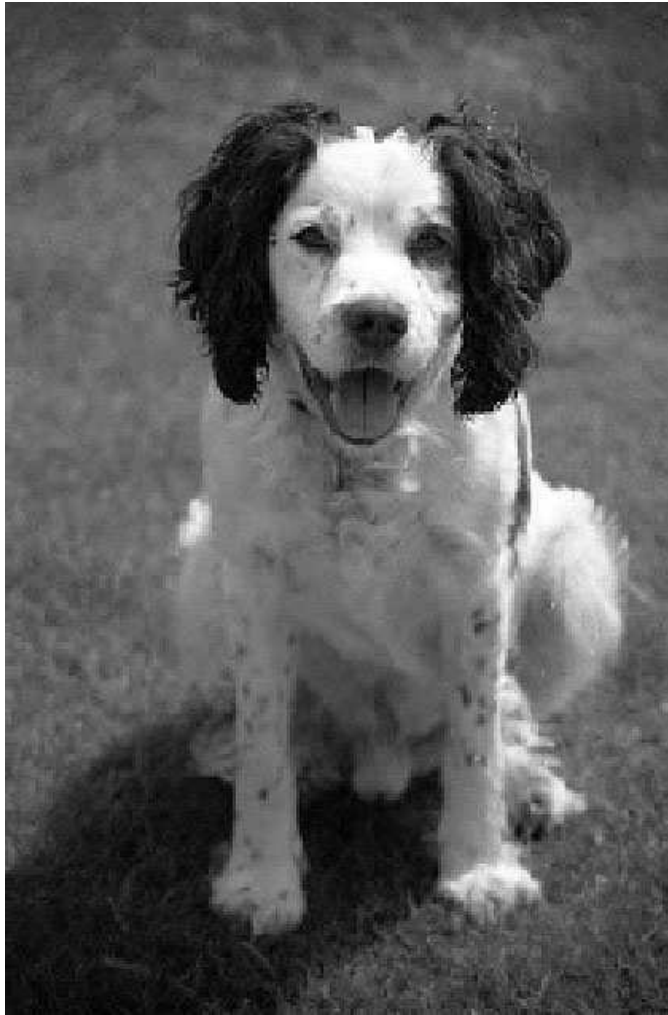
# Rilevamento dei bordi



Bordi nell'immagine  $\Leftarrow$  discontinuità nella scena:

- 1) profondità
- 2) orientamento della superficie
- 3) riflessione (tracce di superficie)
- 4) illuminazione (ombre, etc.)

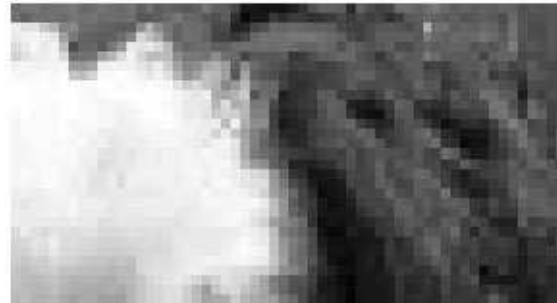
# Bordi



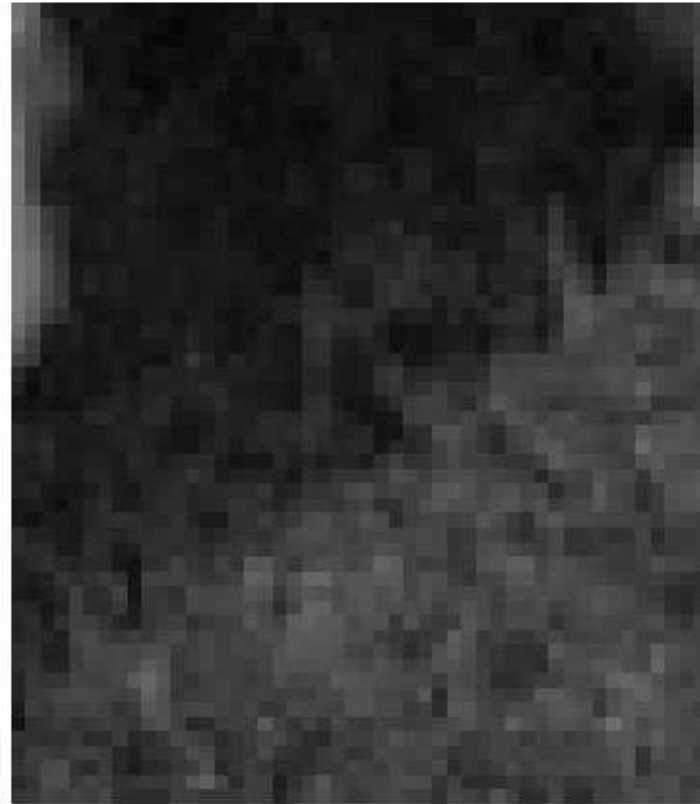
# Bordi



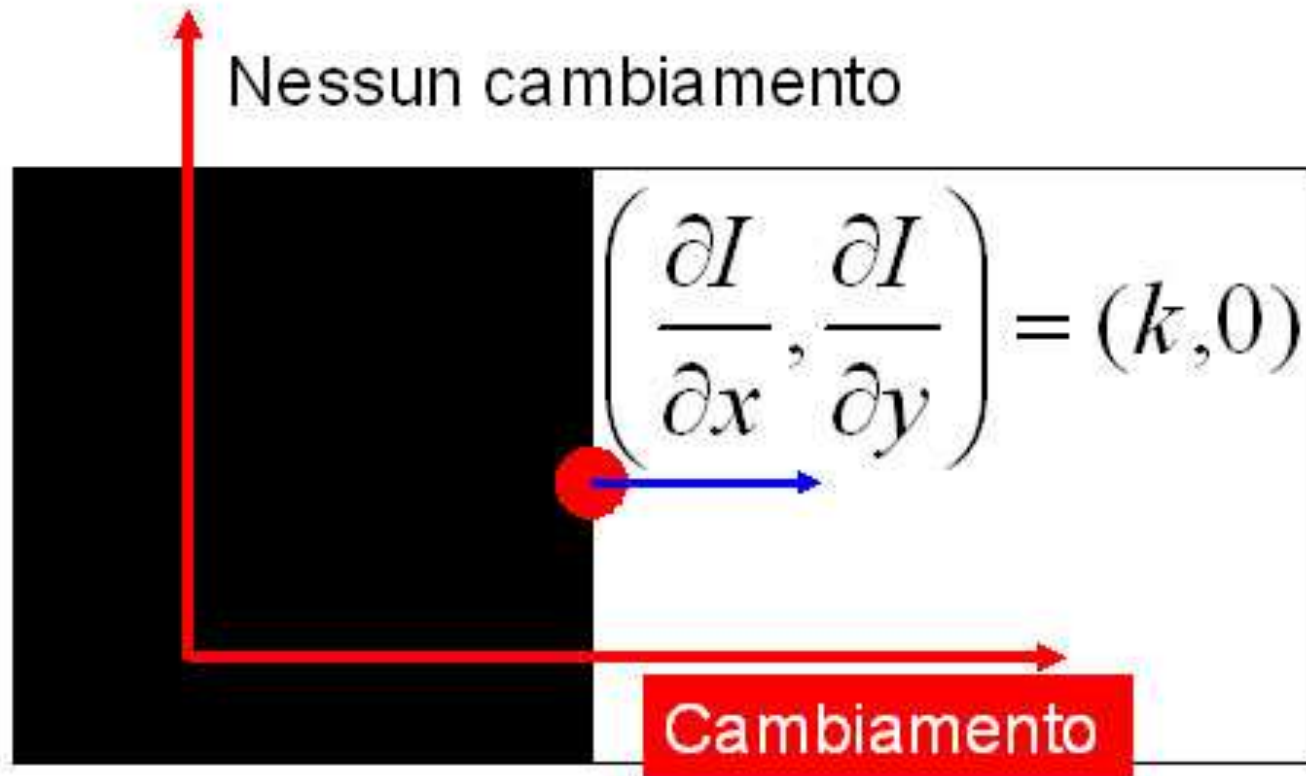
# Bordi



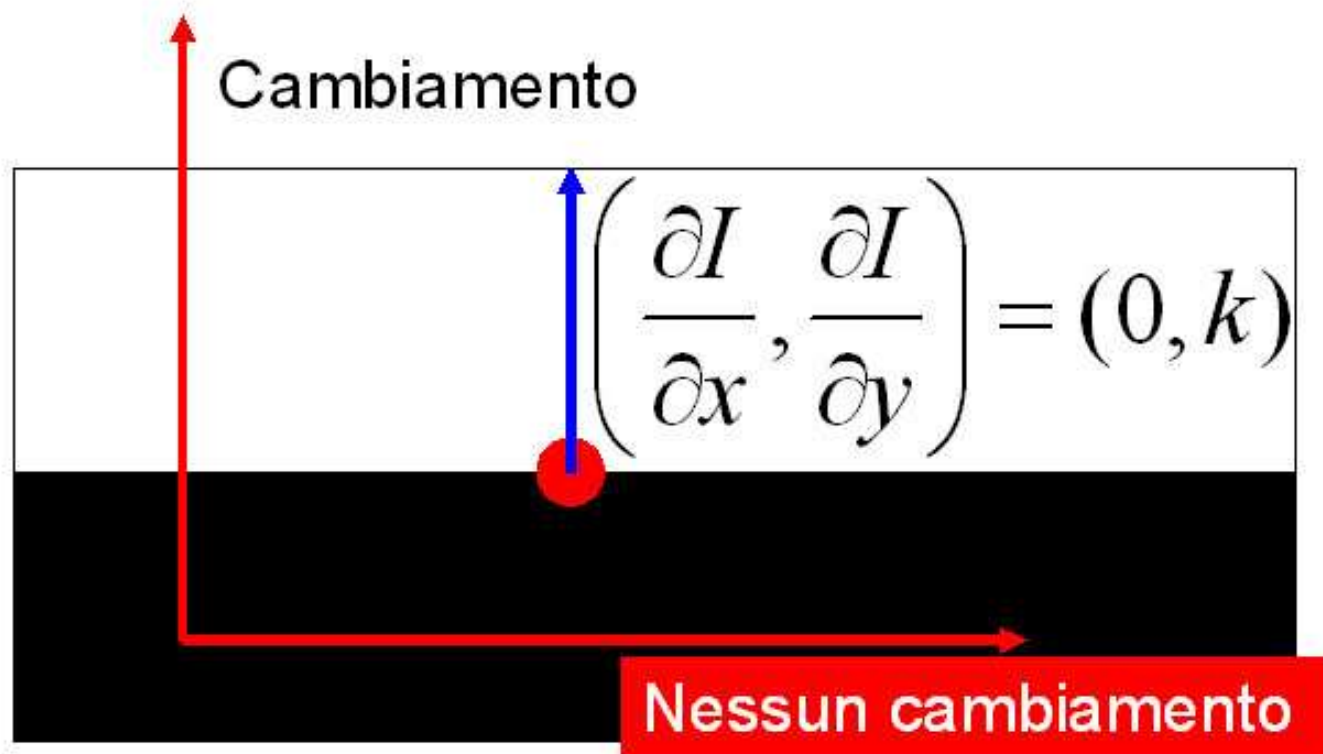
# Bordi



# Gradiente



# Gradiente



# Gradiente

## Caso più generale

$$\left( \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right) = (k1, k2)$$

La direzione del gradiente è  
perpendicolare al bordo

Minor cambiamento

La magnitudine del gradiente  
misura la "forza" del bordo

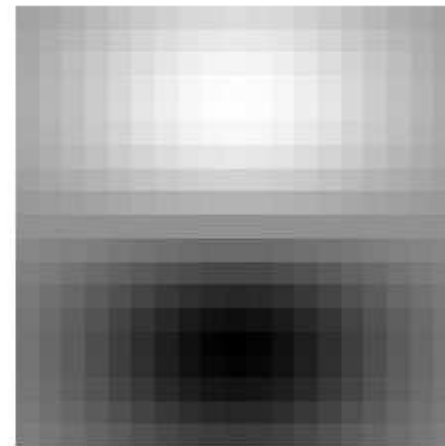
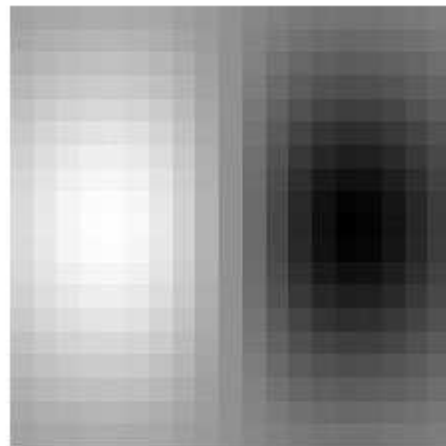
Maggior cambiamento





## Derivata Gaussiana

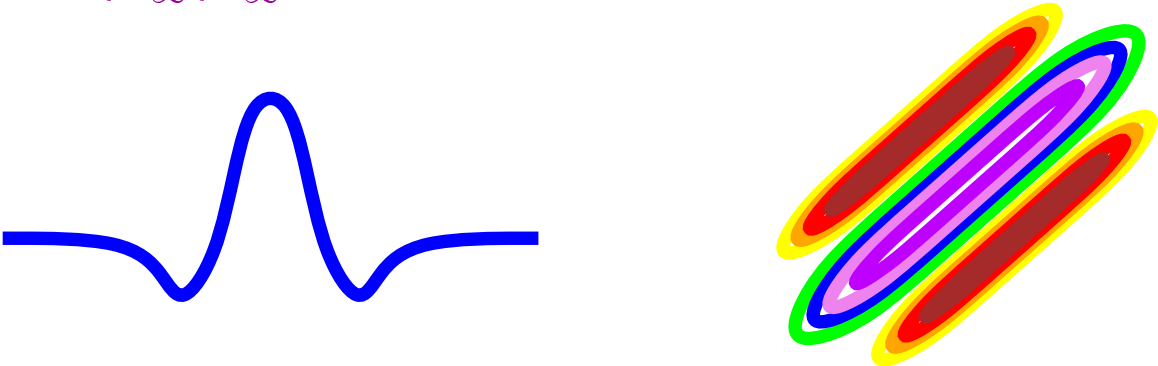
- Usare due derivate, in direzione x e y
- Si può usare la derivata di una Gaussiana



# Rilevamento dei bordi

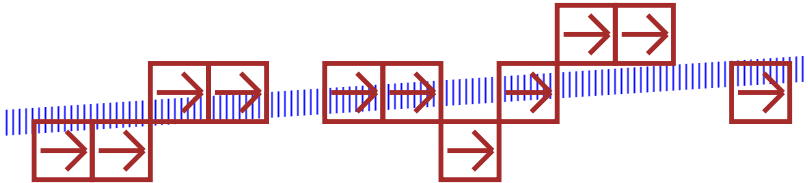
1) Convolvere l'immagine con filtri orientati spazialmente (possibilmente multi-scala)

$$E_{\theta}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\theta}(u, v) I(x + u, y + v) du dv$$

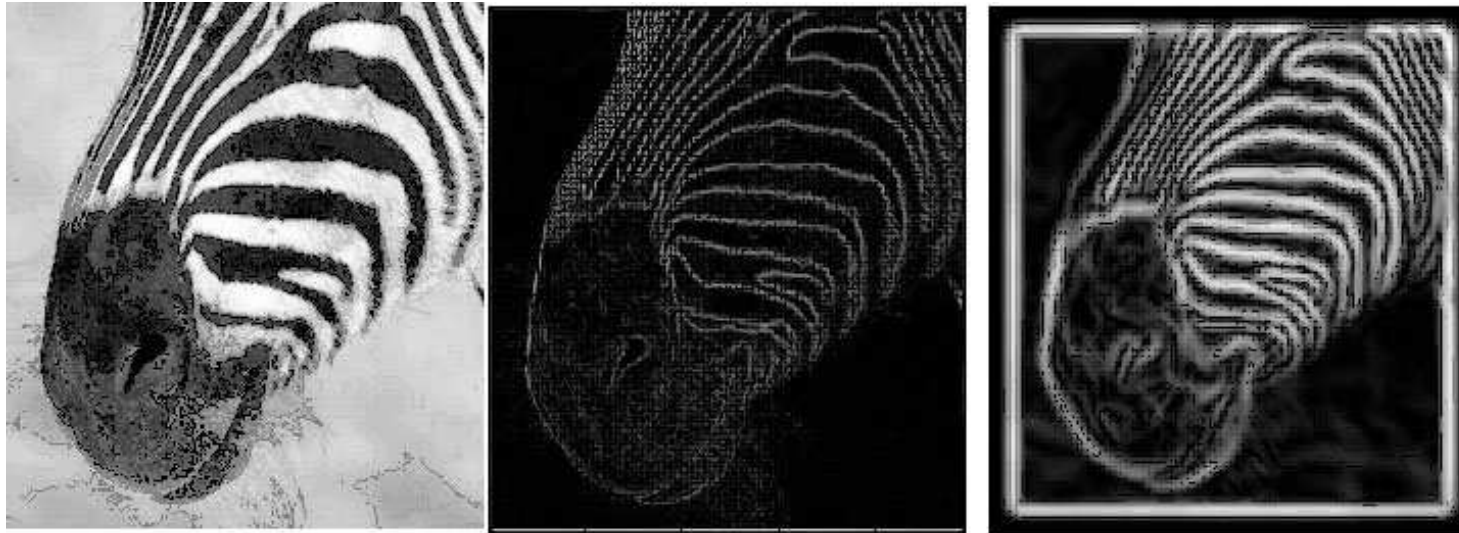


2) Etichettare i pixel al di sopra della soglia con l'orientamento del bordo

3) Inferire segmenti lineari combinando pixel di bordo che si riferiscono allo stesso orientamento



# Scala



## Scala

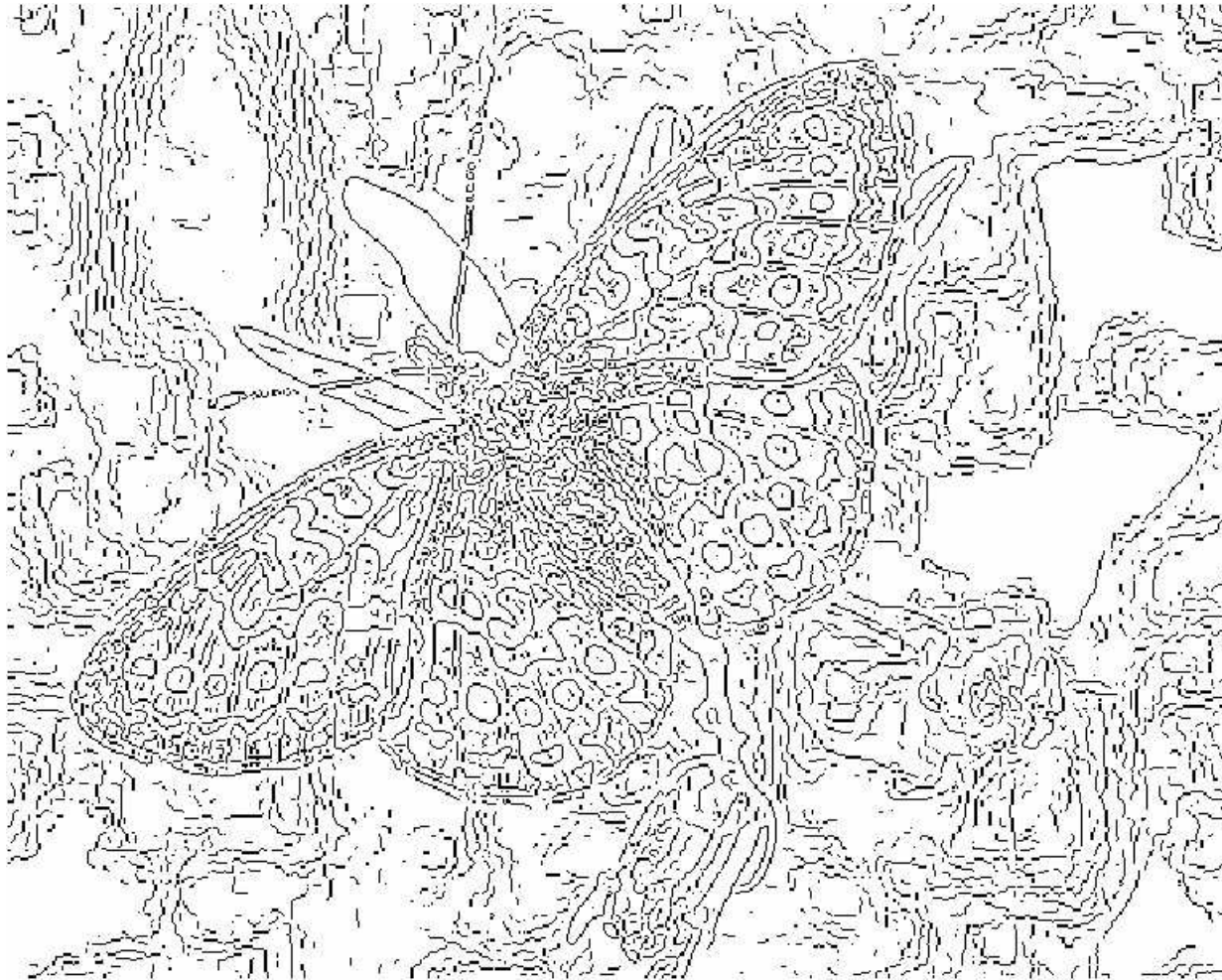
### Smoothing

- Elimina bordi rumorosi
- Rende i bordi più "smooth"
- Rimuove i dettagli "fini"

# Scala

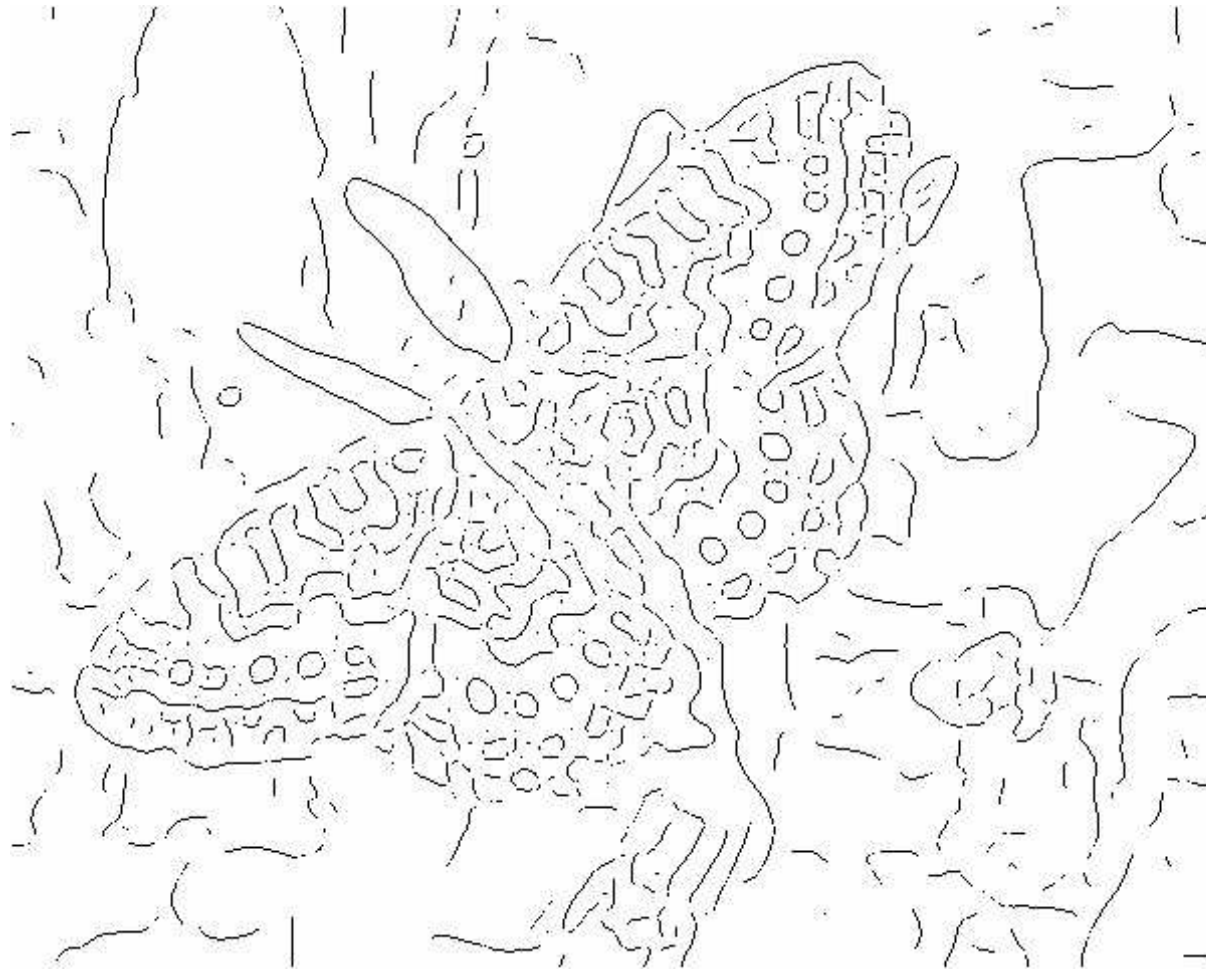


# Scala



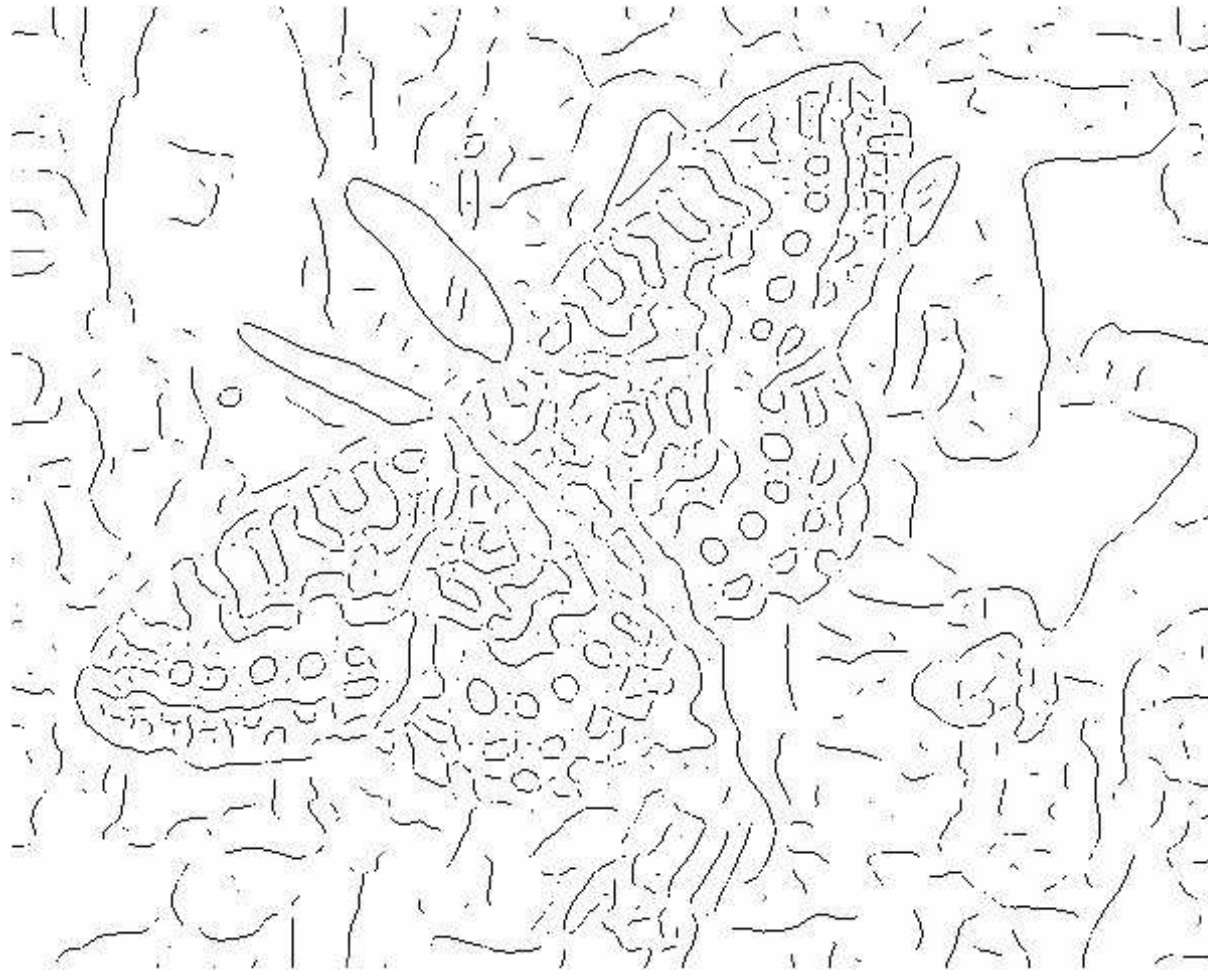
scala fine  
soglia  
alta

# Scala



scala rozza,  
soglia alta

# Scala



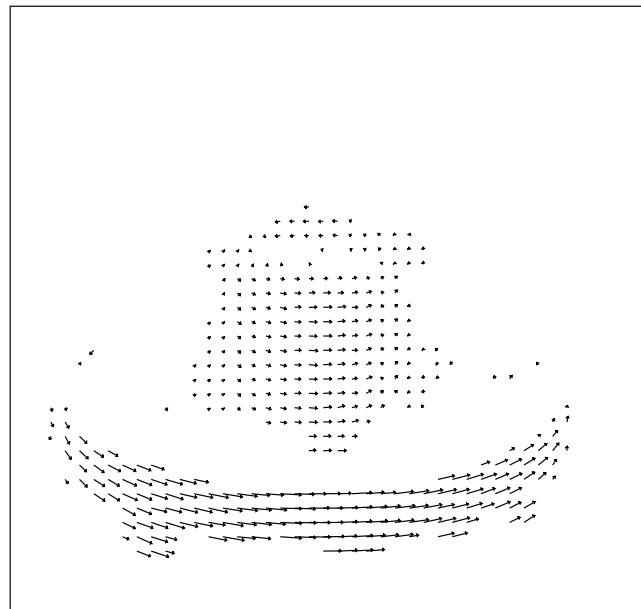
scala rozza,  
soglia bassa

## Indizi da conoscenza a priori

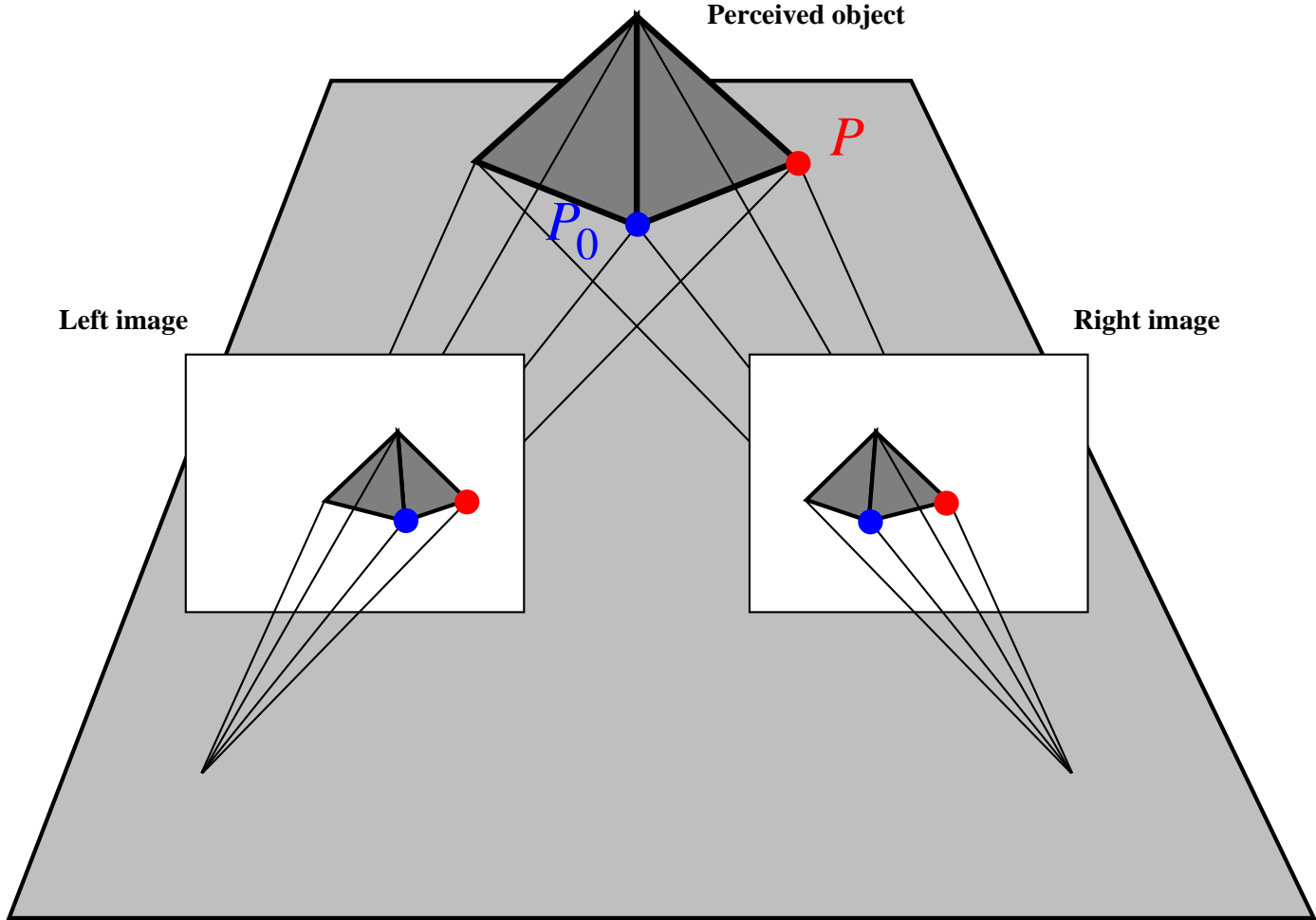
Forma da...	Assume
movimento	corpi rigidi, moto continuo
stereo	corpi solidi, contigui, non-ripetitivi
trama (texture)	trama uniforme
ombreggiatura	riflesso
contorno	curvatura minima



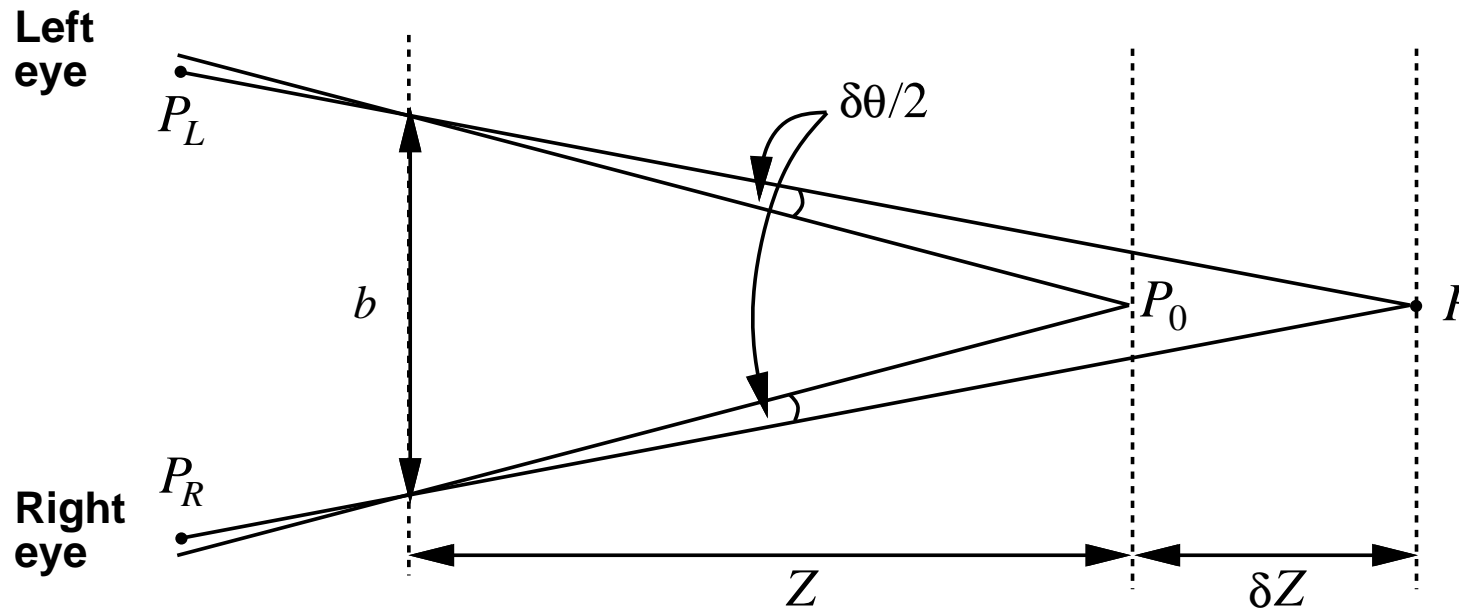
# Movimento



# Visione stereo



## Stima della profondità



Geometria semplice:  $\delta Z = Z^2 \delta\theta / (-b)$

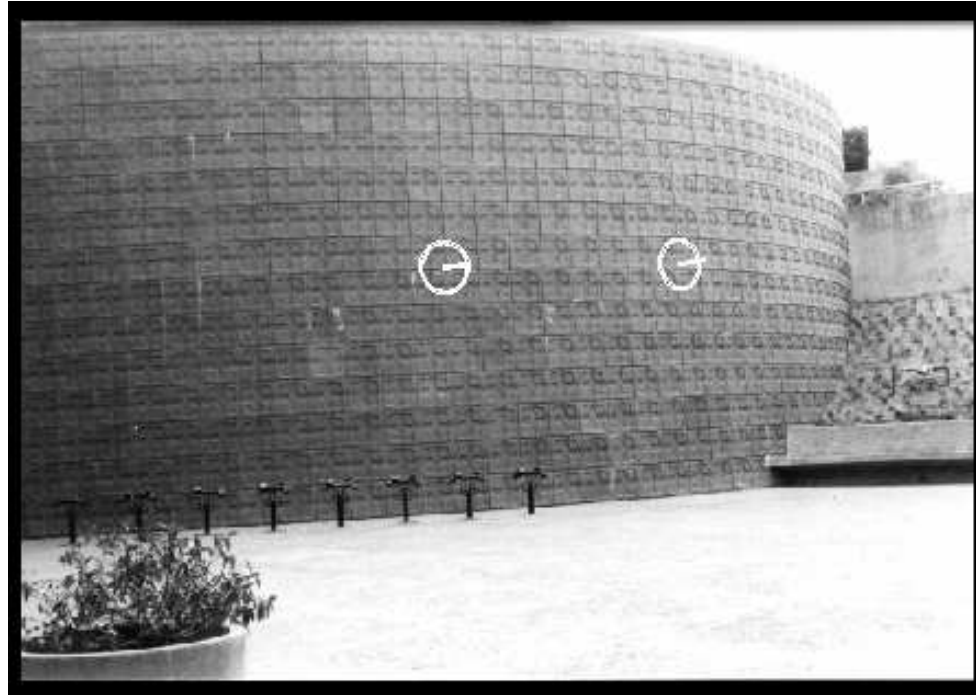
Fisiologia:  $\delta\theta \geq 2.42 \times 10^{-5}$  radianti,  $b = 6\text{cm}$

$Z = 30\text{cm} \Rightarrow \delta Z \approx 0.04\text{mm}$

$Z = 30\text{m} \Rightarrow \delta Z \approx 40\text{cm}$

$b$  grande  $\Rightarrow$  migliore risoluzione!

## Trama (Texture)

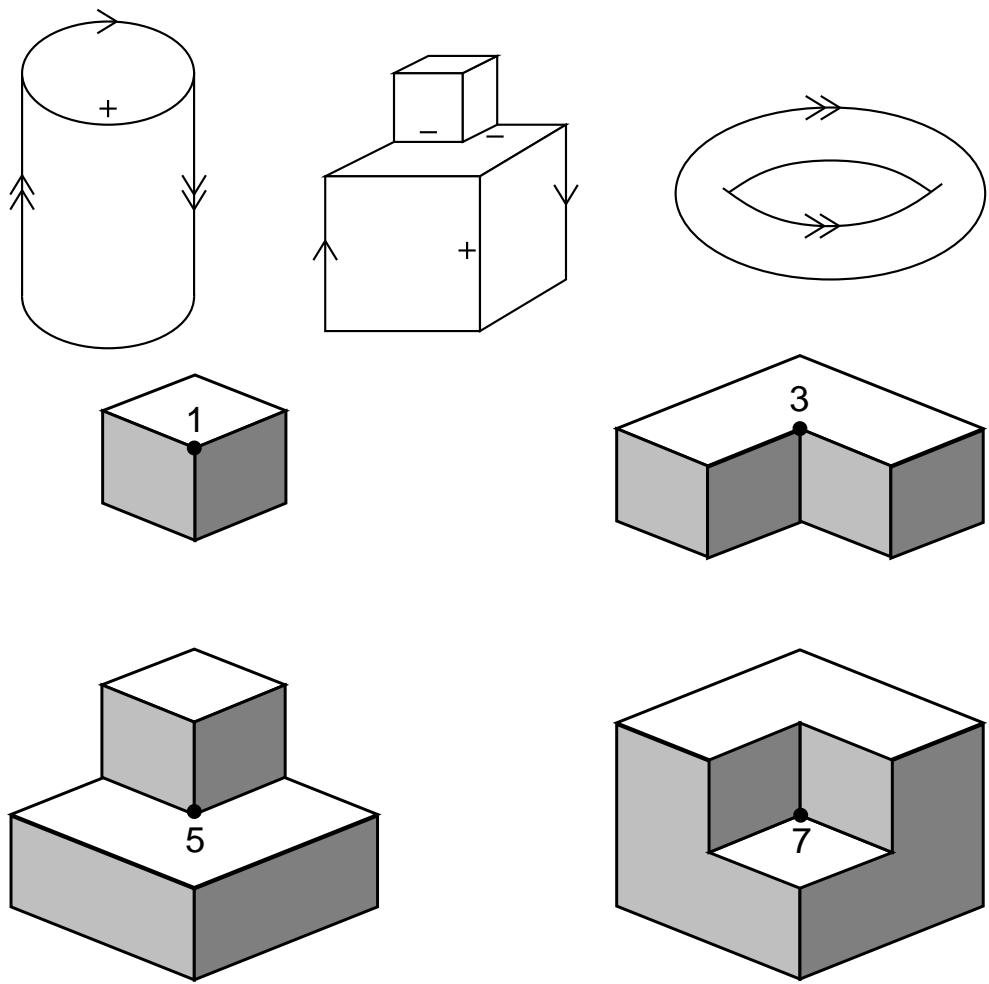


Idea: assumere che la trama sia uniforme, calcolare la forma della superficie che avrebbe prodotto la deformazione

Idee simili funzionano per l'ombra—assumere riflesso uniforme, etc.—**ma** interferenze di riflesso implicano calcolo non locale delle intensità percepite

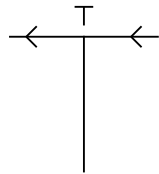
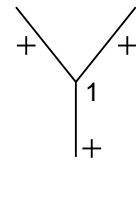
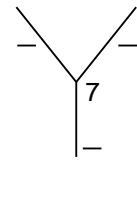
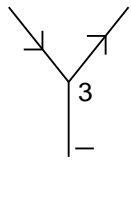
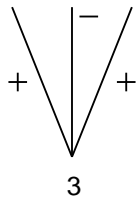
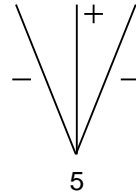
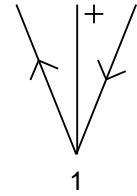
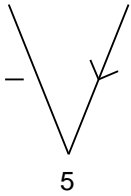
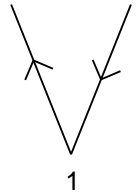
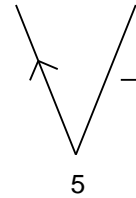
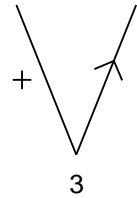
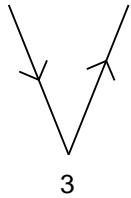
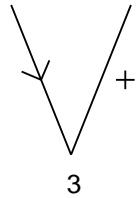
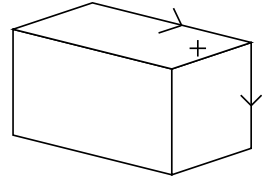
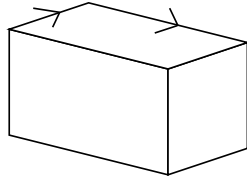
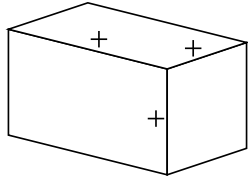
⇒ le cavità sembrano meno profonde di quanto siano

# Tipi di bordi e vertici

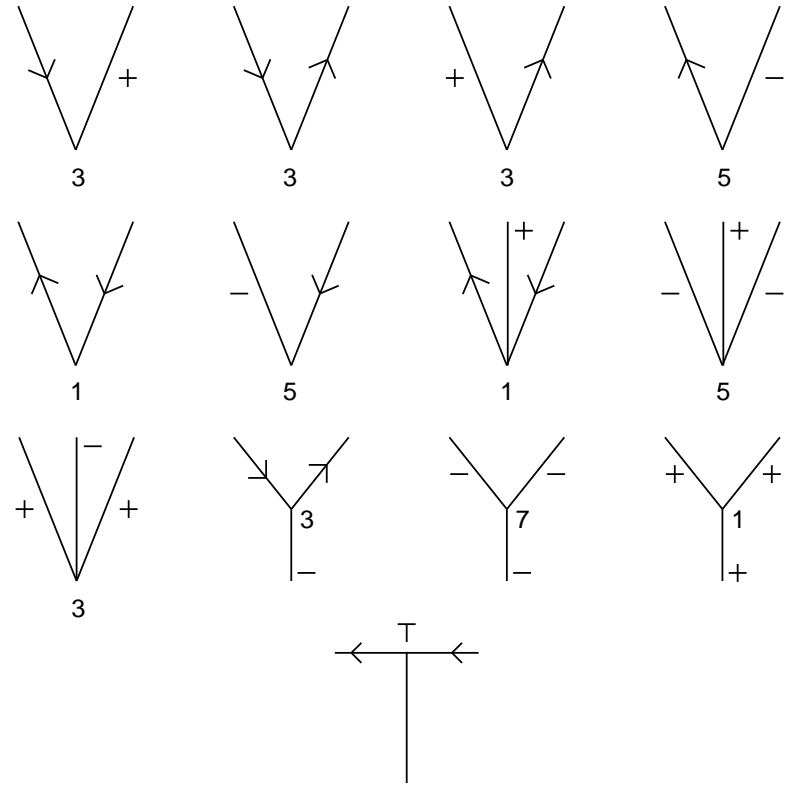
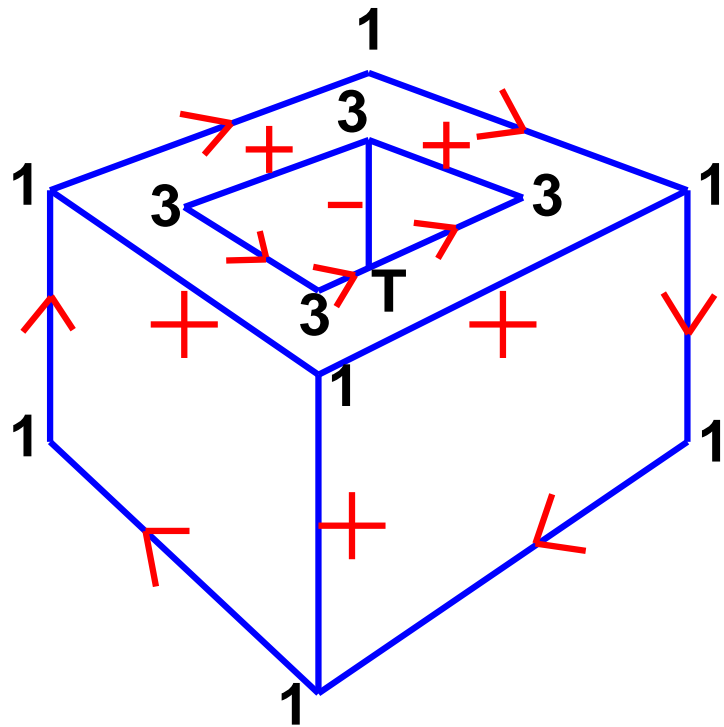


Assumiamo mondo di oggetti solidi a forma di poliedro con vertici triedali

# Etichette vertice/bordo



# Esempio di etichettatura vertice/bordo



Sistema di vincoli: variabili = bordi, vincoli = possibili configurazioni dei vertici

# Riconoscimento degli oggetti

Semplice idea:

- estrarre forme 3-D dall'immagine
- confrontare con “libreria di forme”

Problemi:

- estrarre superfici curve dall'immagine
- rappresentare la forma degli oggetti estratti
- rappresentare la forma e variabilità delle classi di oggetti nella libreria
- segmentazione impropria, occlusioni
- illuminazione sconosciuta, ombre, tracce, rumore, complessità, etc.

Approcci:

- indicizzare la libreria tramite misure di proprietà invarianti degli oggetti
- allineamento delle caratteristiche dell'immagine con quelle dell'oggetto in libreria
- confronto dell'immagine con viste multiple (**aspetti**) dell'oggetto in libreria
- metodi di apprendimento automatico basati sulle statistiche dell'immagine



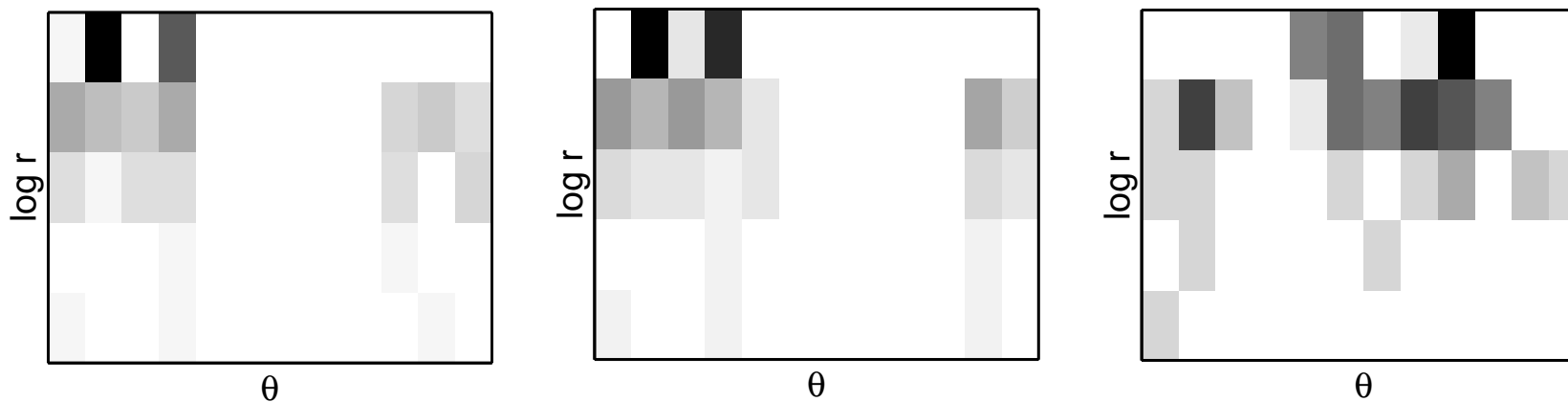
## Confronto forma-contesto

Idea base: convertire la **forma** (un concetto relazionale) in un insieme fisso di **attributi** usando il **contesto spaziale** di ognuno di un insieme fissato di punti sulla superficie della forma.



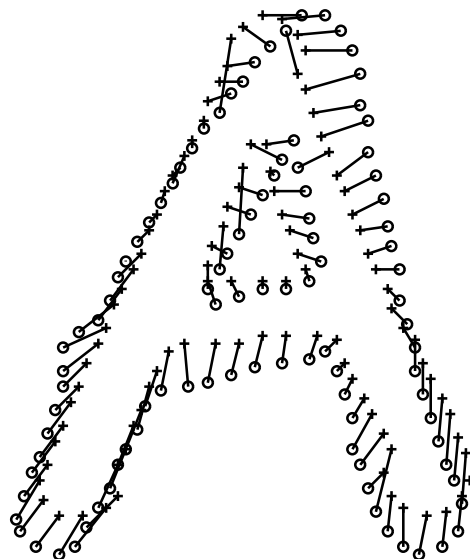
## Confronto forma-contesto

Ogni punto è descritto per mezzo del suo istogramma locale di contesto (numero di punti che cadono in ogni cella di una griglia log-polare)



## Confronto forma-contesto

Determinare la distanza totale tra forme sommando le distanze di punti corrispondenti al miglior "matching" possibile



## Riassunto

La visione è difficile — rumore, ambiguità, complessità

Conoscenza a priori è essenziale per vincolare il problema

Necessità di combinare insieme più indizi: movimento, contorni, ombreggiatura, trama, stereo, ...

“Libreria” di rappresentazioni di oggetti: forma vs. aspetti

Confronto immagine/oggetto: caratteristiche, linee, regioni, etc.