



# Tecnologie digitali per il suono e l'immagine 2020/21

Vincenzo Lombardo

Corso di Laurea in DAMS

Università di Torino

Mutuato in parte da Elaborazione audio e musica

(Laurea Magistrale di Informatica)

# La percezione visiva

Psicofisiologia della visione

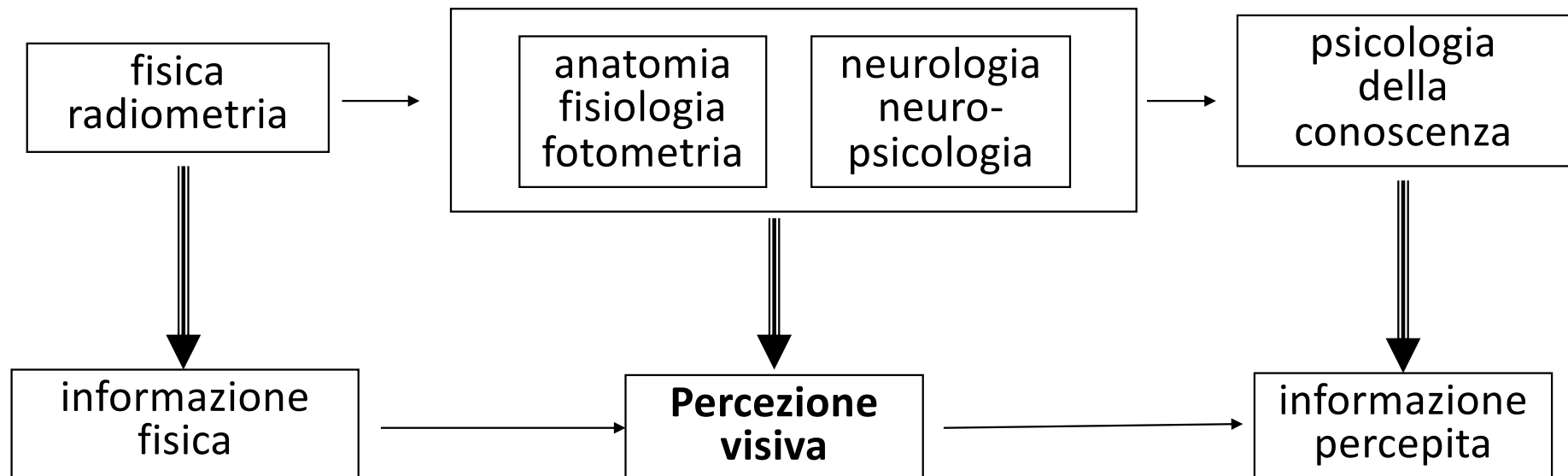
La ricostruzione della scena in 3-D

# La visione

“ ... Il cuore vero della percezione visiva è l’inferenza derivata dalla struttura di un’immagine sulla struttura del mondo reale esterno. La teoria della visione è esattamente la teoria di come è possibile fare ciò ... ”

*David Marr*

# Aspetti coinvolti

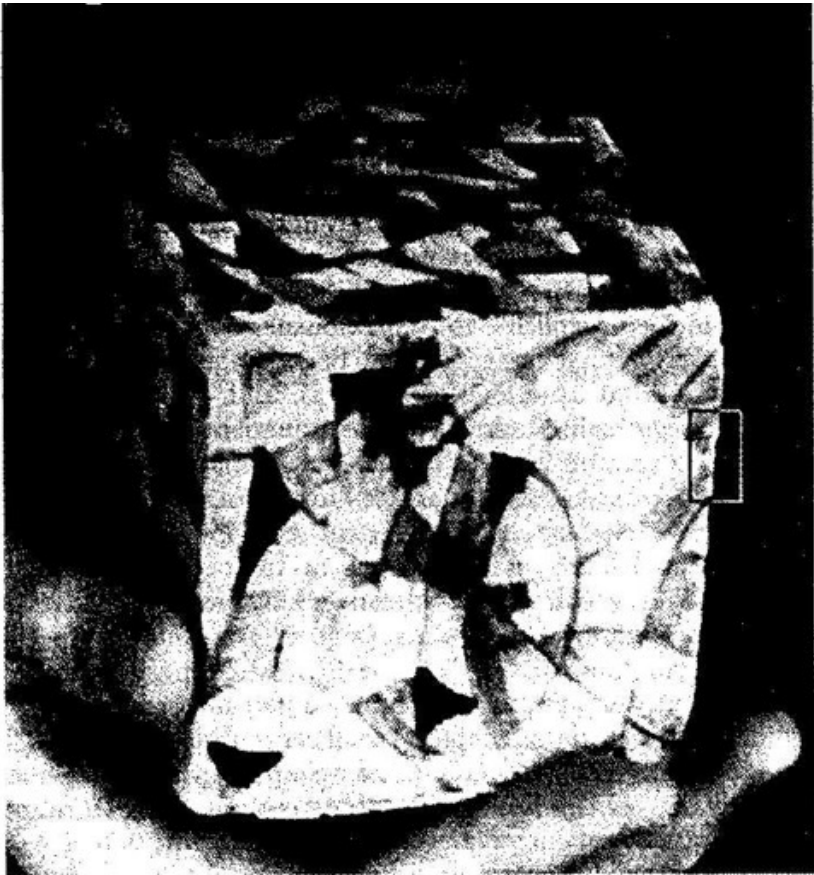




# Obiettivi della visione

- Ricca comprensione di ciò che è nel mondo
- Localizzazione degli oggetti
- Come gli oggetti cambiano nel tempo
- Conseguenze per un sistema:
  - Capacità di riconoscere e manipolare gli oggetti
  - Capacità di interagire fisicamente con l'ambiente

# Input



225	221	216	219	219	214	207	218	219	220	207	155	136	135	130	131	125
213	206	213	223	208	217	223	221	223	216	195	156	141	130	128	138	123
206	217	210	216	224	223	228	230	234	216	207	157	136	132	137	130	128
211	213	221	223	220	222	237	216	219	220	176	149	137	132	125	136	121
216	210	231	227	224	228	231	210	195	227	181	141	131	133	131	124	122
223	229	218	230	228	214	213	209	198	224	161	140	133	127	133	122	133
220	219	224	220	219	215	215	206	206	221	159	143	133	131	129	127	127
221	215	211	214	220	218	221	212	218	204	148	141	131	130	128	129	118
214	211	211	218	214	220	226	216	223	209	143	141	141	124	121	132	125
211	208	223	213	216	226	231	230	241	199	153	141	136	125	131	125	136
200	224	219	215	217	224	232	241	240	211	150	139	128	132	129	124	132
204	206	208	205	233	241	241	252	242	192	151	141	133	130	127	129	129
200	205	201	216	232	248	255	246	231	210	149	141	132	126	134	128	139
191	194	209	238	245	255	249	235	238	197	146	139	130	132	129	132	123
189	199	200	227	239	237	235	236	247	192	145	142	124	133	125	138	128
198	196	209	211	210	215	236	240	232	177	142	137	135	124	129	132	128
198	203	205	208	211	224	226	240	210	160	139	132	129	130	122	124	131
216	209	214	220	210	231	245	219	169	143	148	129	128	136	124	128	123
211	210	217	218	214	227	244	221	162	140	139	129	133	131	122	126	128
215	210	216	216	209	220	248	200	156	139	131	129	139	128	123	130	128
219	220	211	208	205	209	240	217	154	141	127	130	124	142	134	128	129
229	224	212	214	220	229	234	208	151	145	128	128	142	122	126	132	124
252	224	222	224	233	244	228	213	143	141	135	128	131	129	128	124	131
255	235	230	249	253	240	228	193	147	139	132	128	136	125	125	128	119
250	245	238	245	246	235	235	190	139	136	134	135	126	130	126	137	132
240	238	233	232	235	255	246	168	156	141	129	127	136	134	135	130	126
241	242	225	219	225	255	255	183	139	141	126	139	128	137	128	128	130
234	218	221	217	211	252	242	166	144	139	132	130	128	129	127	121	132
231	221	219	214	218	225	238	171	145	141	124	134	131	134	131	126	131
228	212	214	214	213	208	209	159	134	136	139	134	126	127	127	124	122
219	213	215	215	205	215	222	161	135	141	128	129	131	128	125	128	127

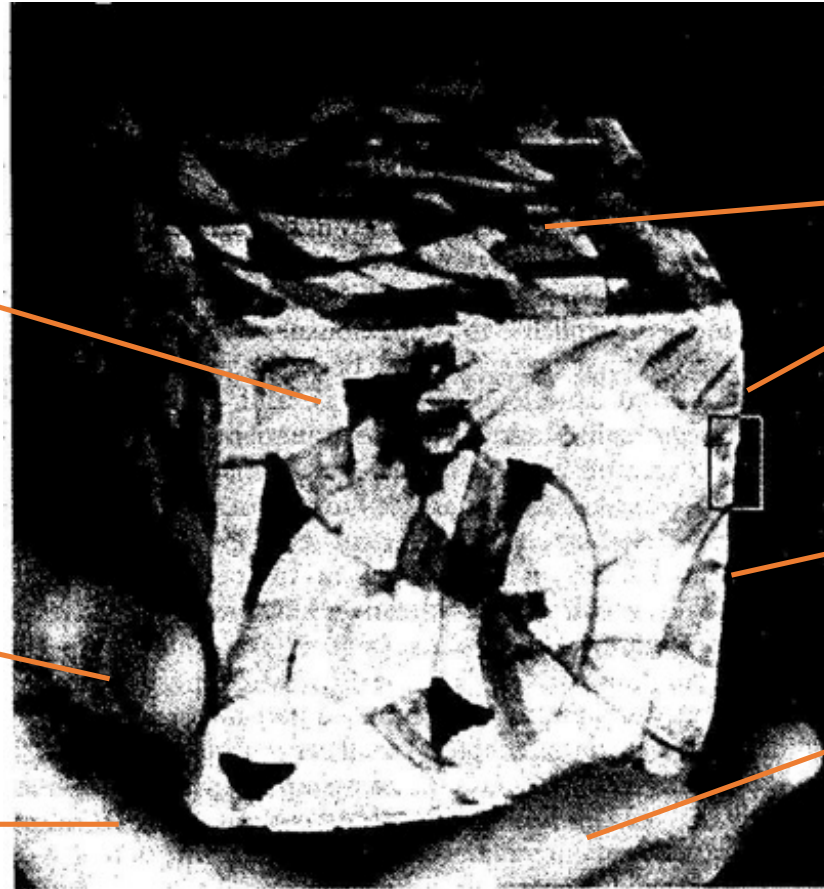
[Johnson-Laird - Computer and the Mind]

# Output

*Cubo di un qualche  
materiale  
(cristallo?)*

*Pollice*

*Mano (dx)*



*Superfici  
perpendicolari  
tra loro*

*Superficie nascosta*

*La mano sos(tiene) il cubo*

Rappresentazione della scena 3D: Posizione sistema VS posizione oggetti della scena  
Posizione relativa degli oggetti  
Colori, forme, composizione, cinematica della scena  
Identificazione di oggetti noti e meno noti

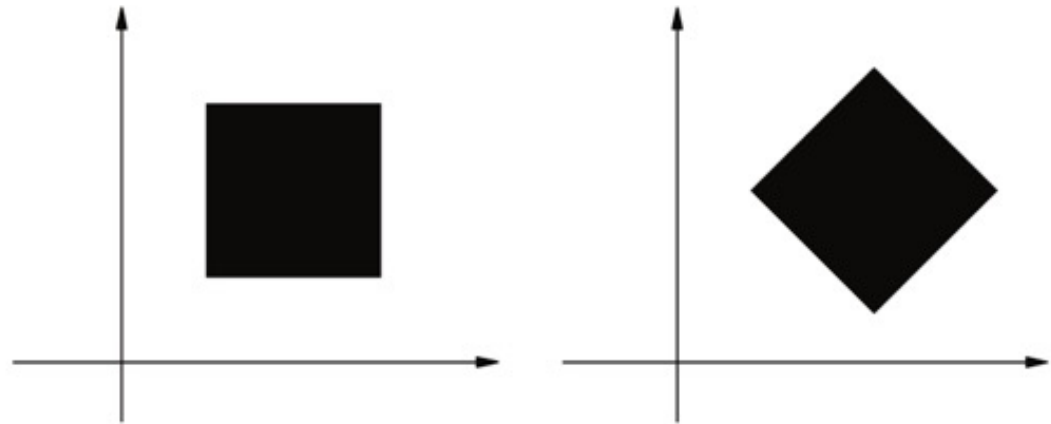
# Principio di indeterminazione

Limiti della percezione visiva: realtà fenomenica e percezioni non veridiche



Triangolo di Kanizsa

Oggetto fenomenico e  
oggetto fisico

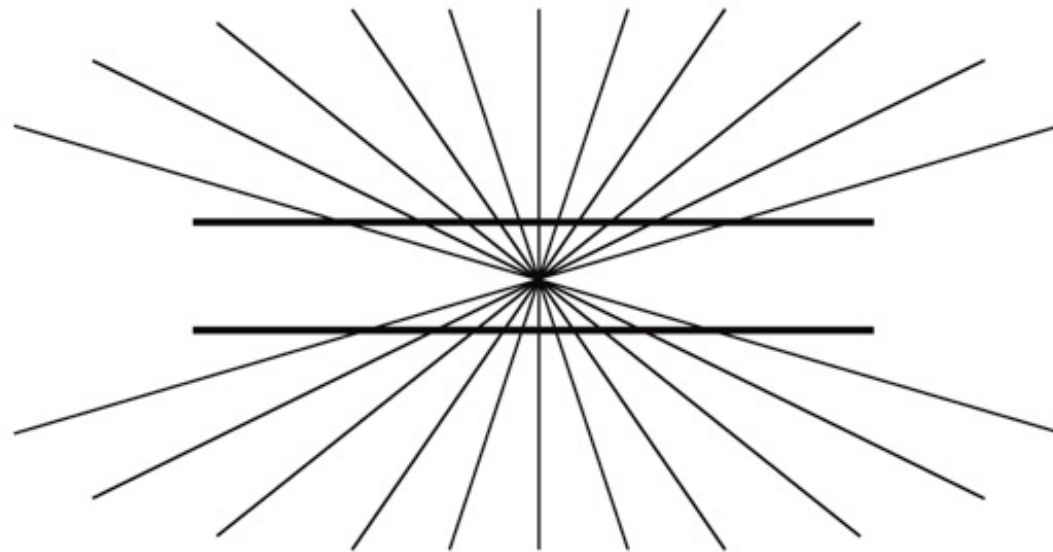


Quadrato di Mach

Allineamento con gli assi cardinali  
influenza la percezione di forma e dimensione

# Illusioni visive

Oggetti di interesse: dipendono dalla natura dei meccanismi sensoriali e elaborazioni del sistema visivo



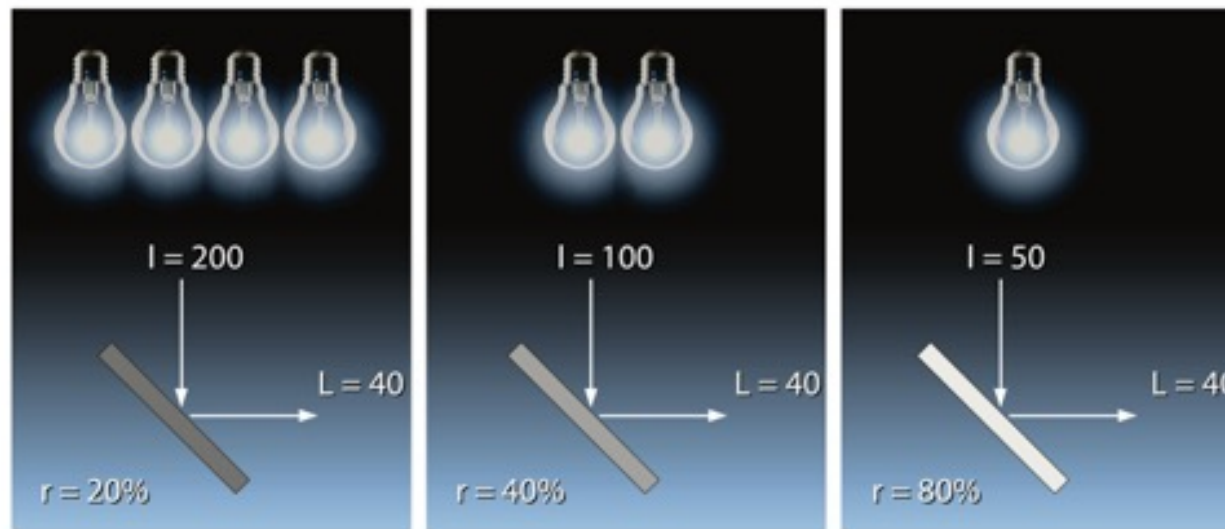
Illusione di Hering

# Indeterminazione

- Causata dal dispositivo di cattura
- Da immagine retinica e rappresentazione in codice neurale a cosa l'abbia determinata
- Trasformazione ricostruttiva: il sistema visivo cerca di recuperare proprietà non veicolate dalla luce

# Indeterminazione di tipo fotometrico

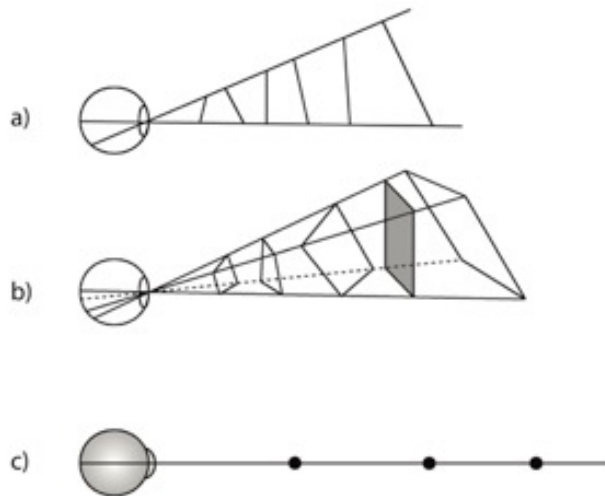
Confusione tra componenti separate nel mondo fisico (illuminazione + proprietà delle superfici)



La stessa luminanza risultante  $L$  può essere frutto di diversi prodotti  $I \times r$ .

# Indeterminazione di tipo geometrico

Eliminazione dell'informazione di profondità (da 3D a 2D)



Visione bidimensionale

Visione prospettica

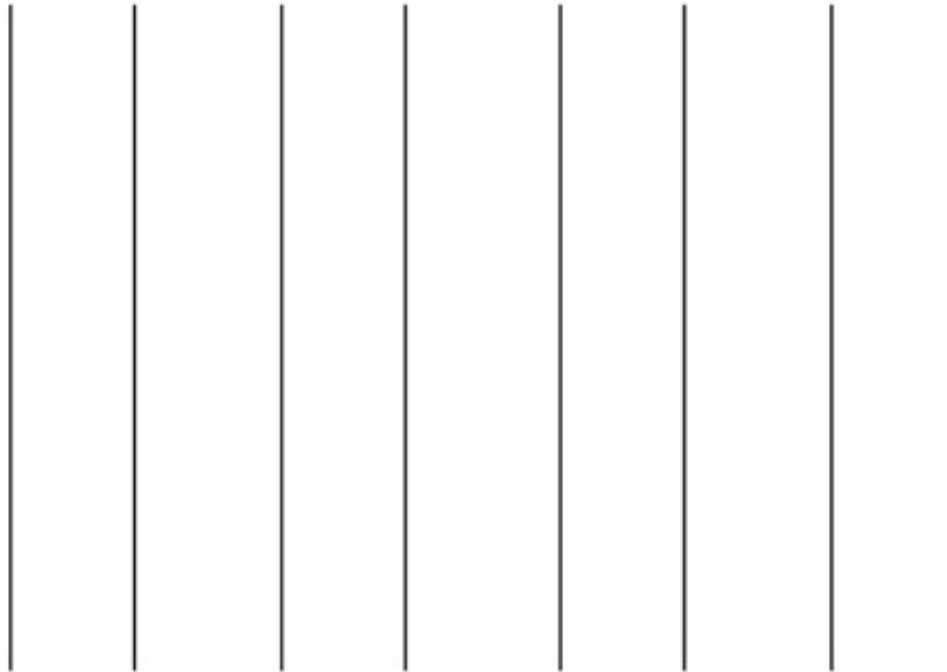
Visione lineare

Indeterminazione radiale: propagazione rettilinea della luce e asse dello sguardo



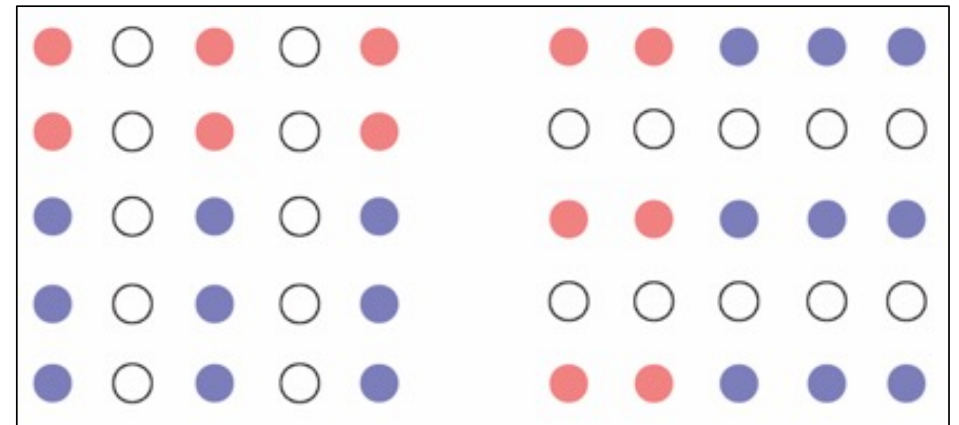
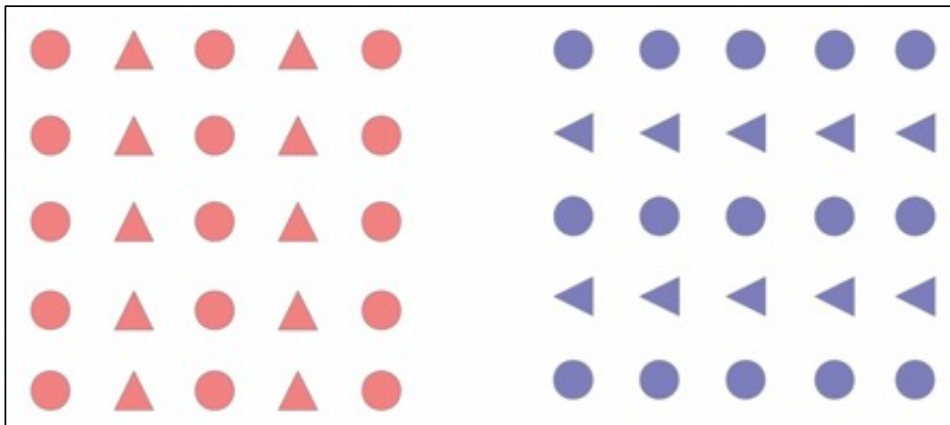
# Raggruppamento di tipo percettivo: Gestalt

Legge della vicinanza: si accorpano elementi vicini



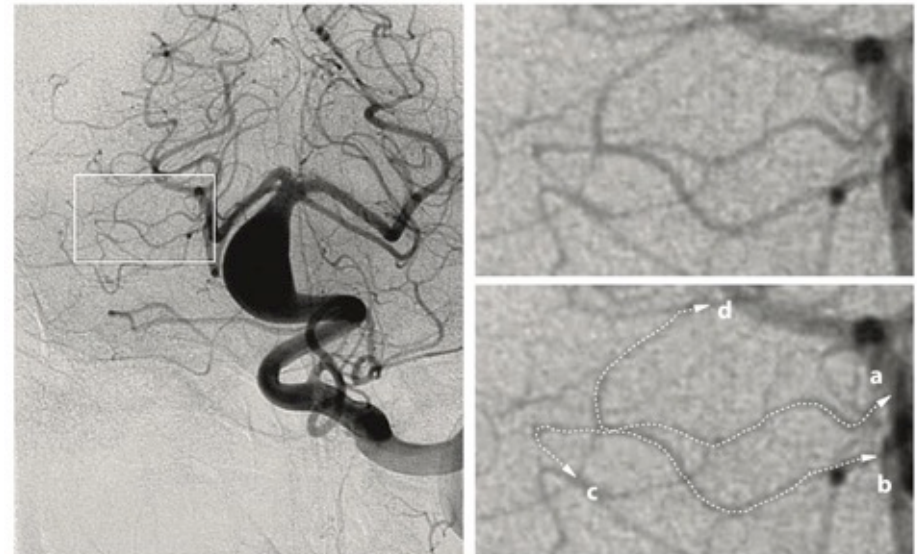
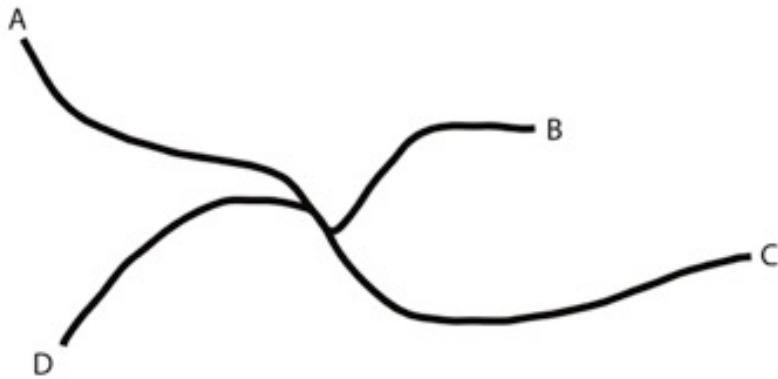
# Raggruppamento di tipo percettivo: Gestalt

Legge della somiglianza: si accorpano elementi simili in forma, grandezza e colore



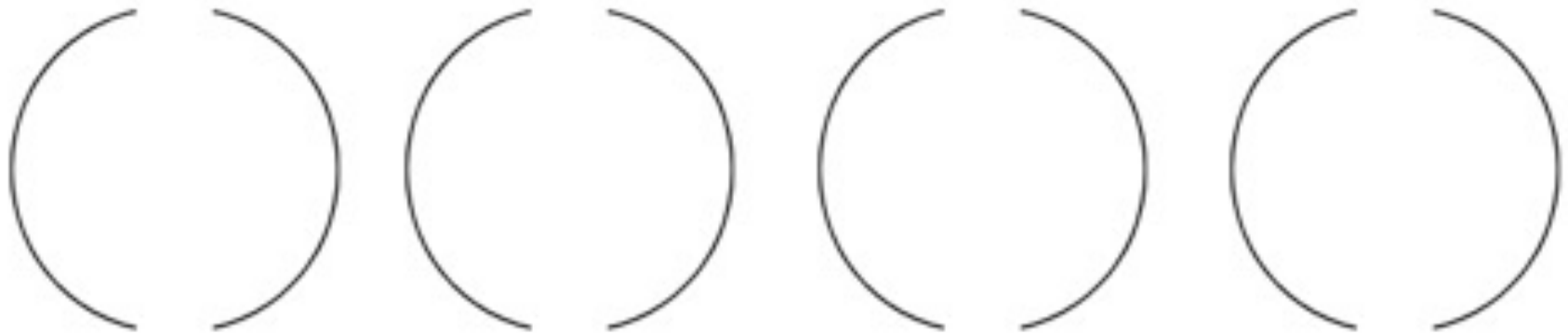
# Raggruppamento di tipo percettivo: Gestalt

Legge della continuità di direzione: accorpamento definito dalla distribuzione lungo un cammino



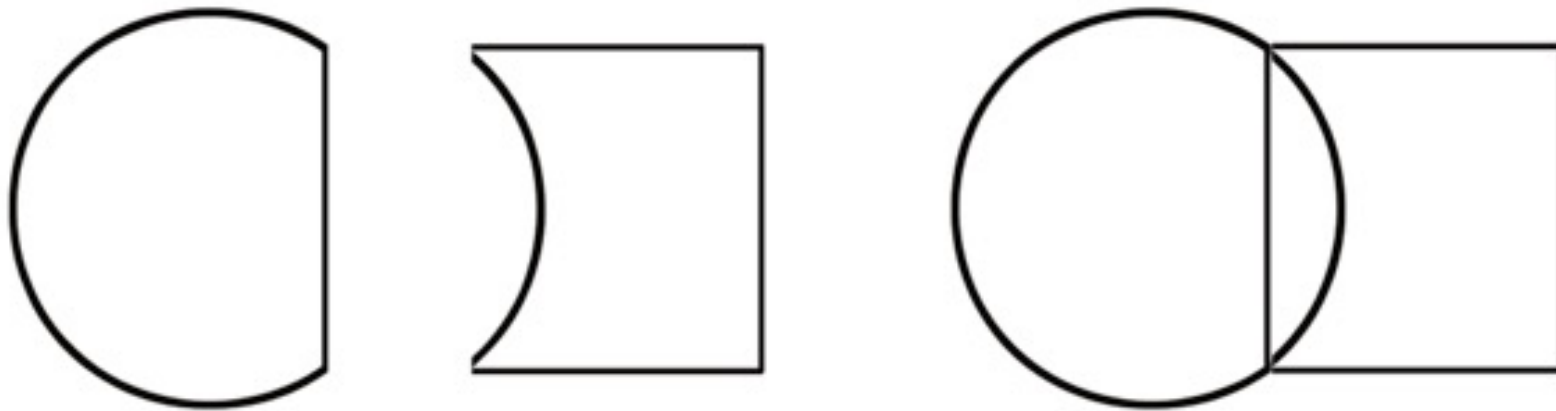
# Raggruppamento di tipo percettivo: Gestalt

Legge della chiusura: tendenza a chiudere i contorni di una figura  
(figura chiusa favorita rispetto a figura aperta)



# Raggruppamento di tipo percettivo: Gestalt

Legge della buona forma o pregnanza: quando figure diverse si uniscono, ciascuna conserva la propria forma

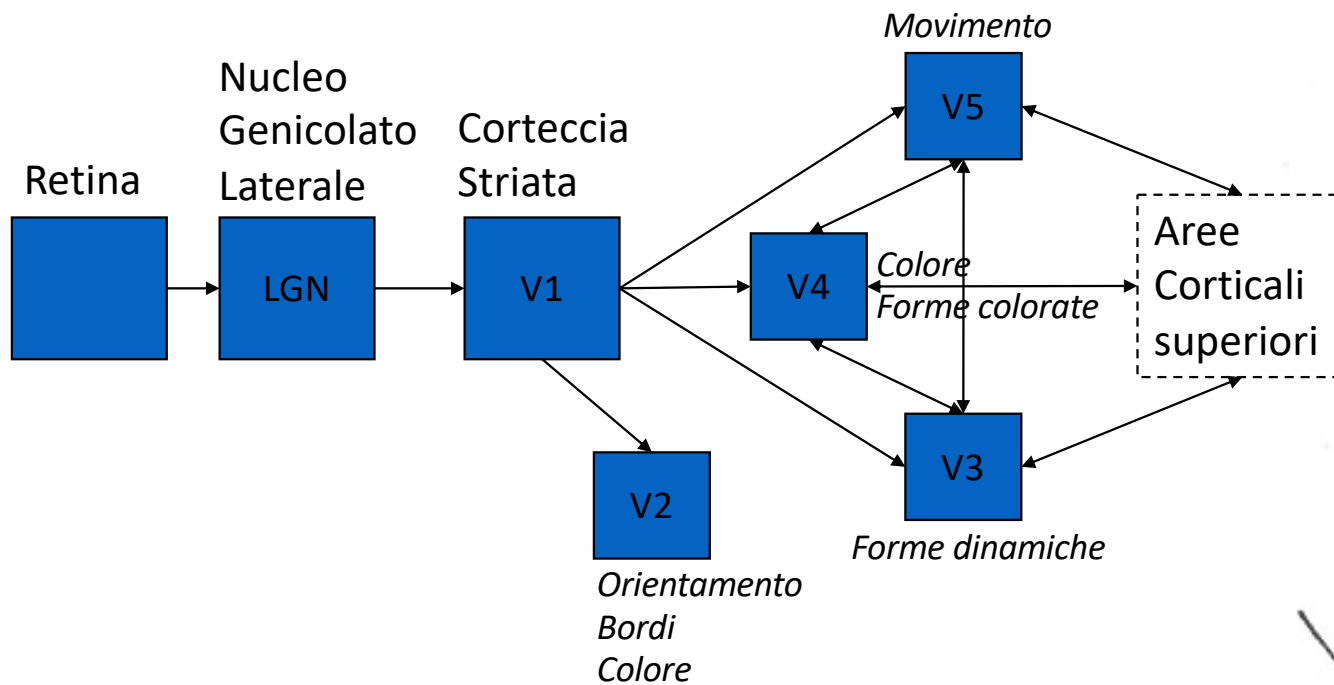


# Raggruppamento di tipo percettivo: Gestalt

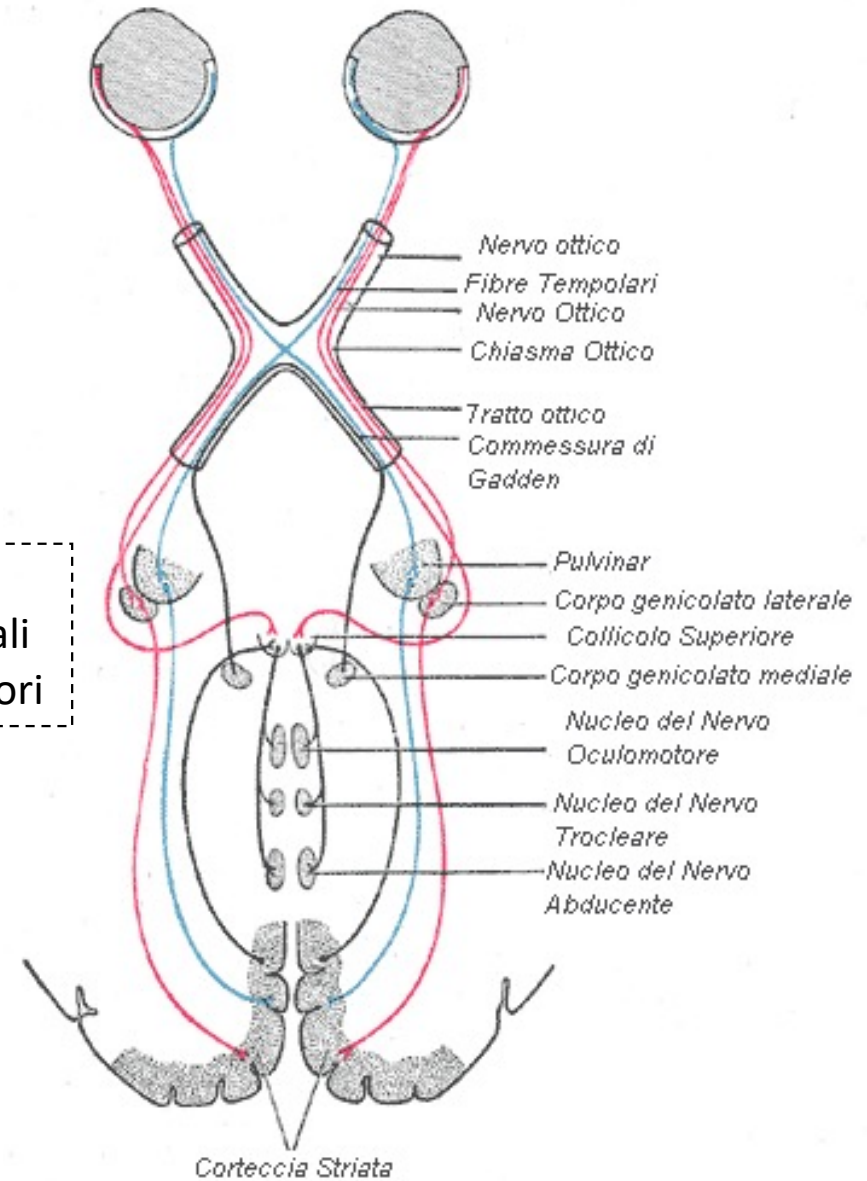
Legge dell'esperienza passata: segmentazione influenzata da esperienze passate; favorita la costruzione di forme familiari



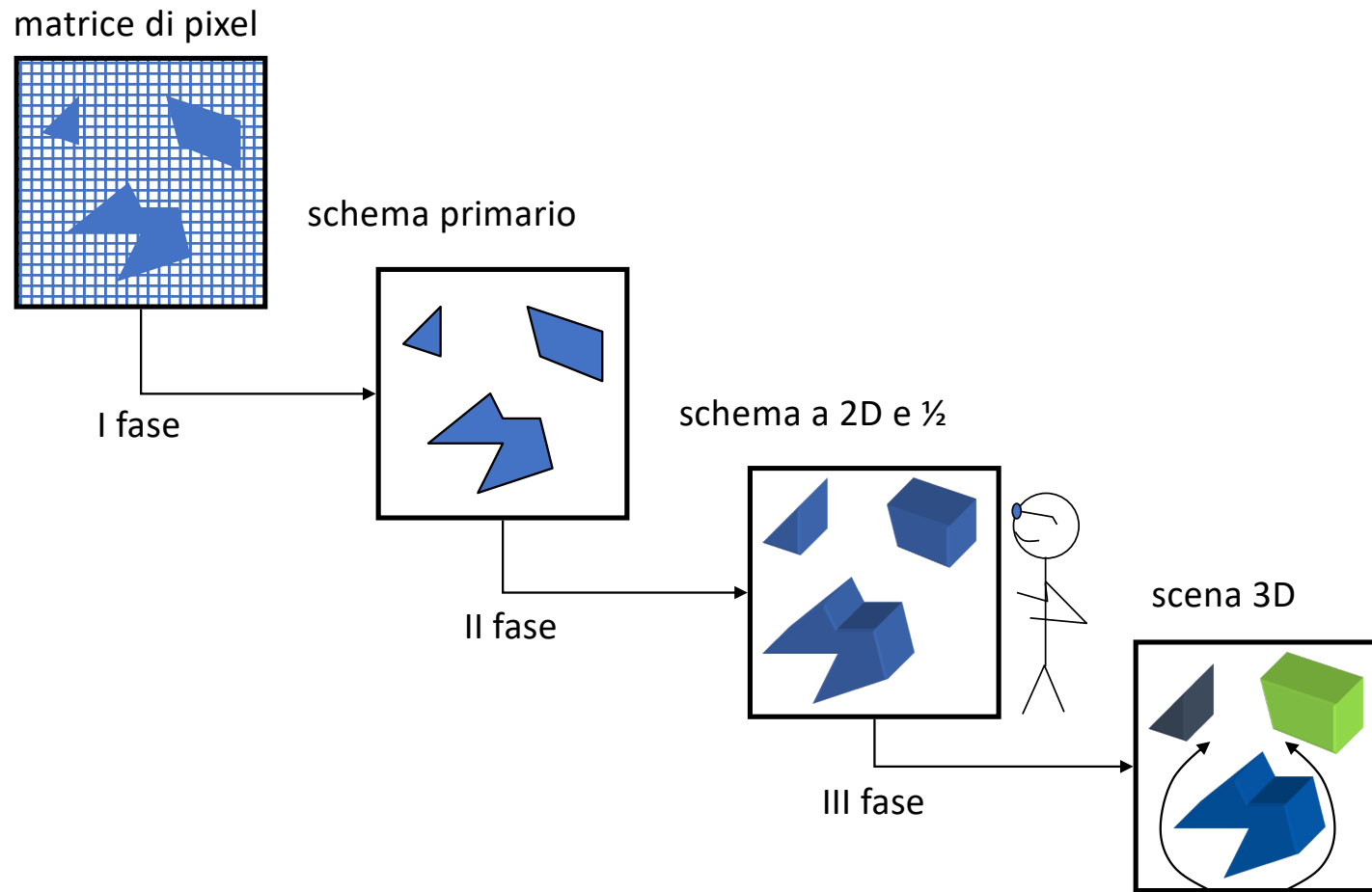
# La fisiologia della visione



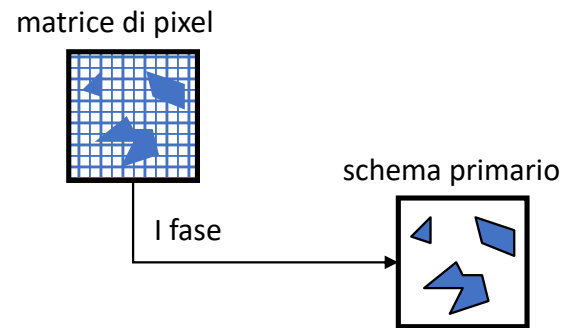
[Wikipedia – Sistema visivo]



# Il passaggio da 2D a 3D







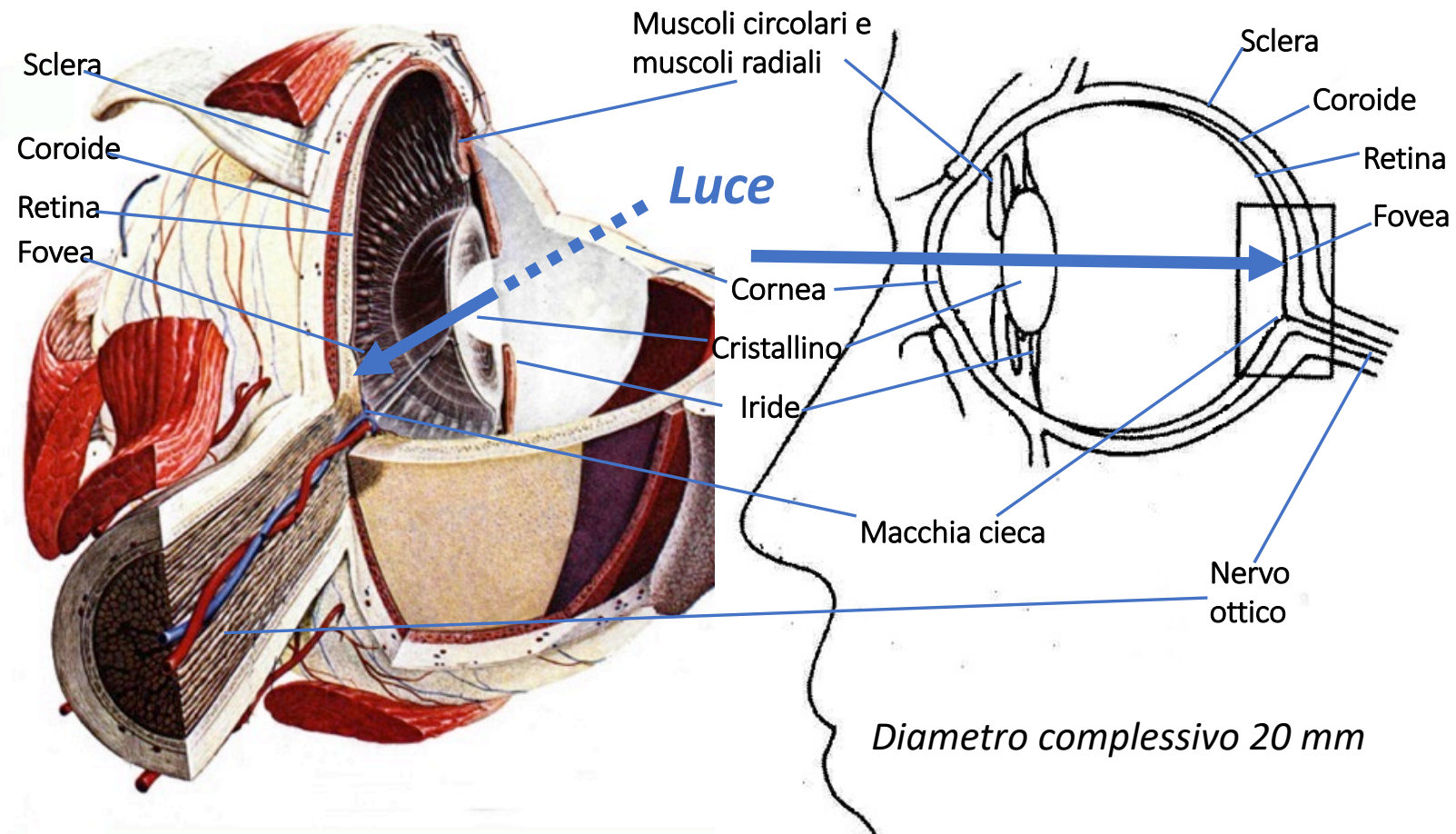
# Prima fase

Campionamento

Localizzazione dei contorni

Schema primario

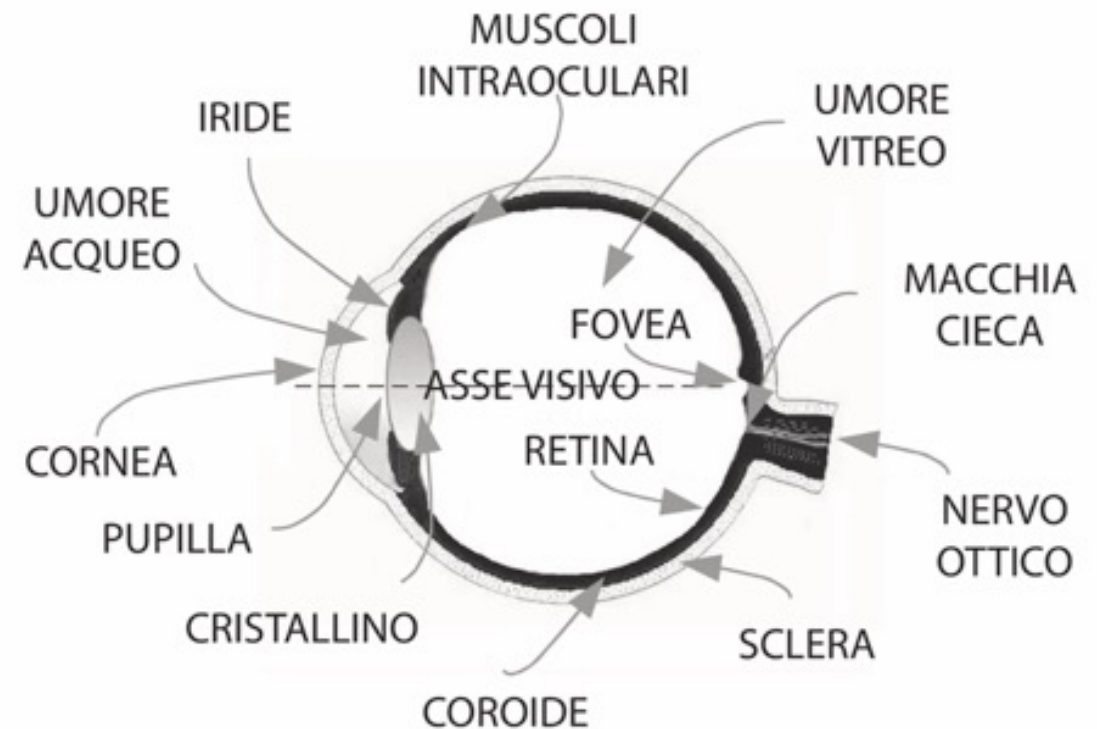
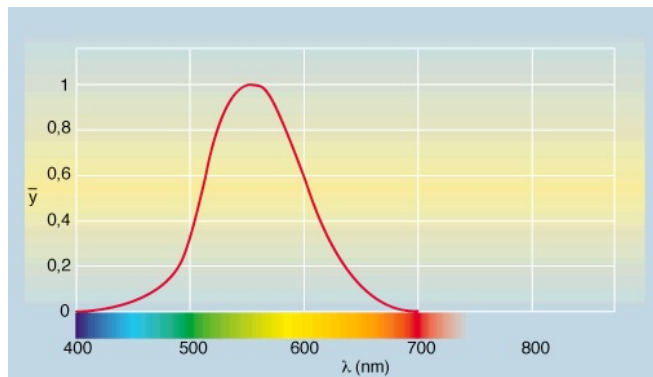
# L'occhio: sistema ottico + sistema sensoriale



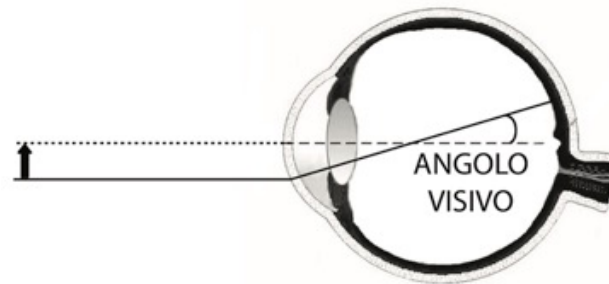
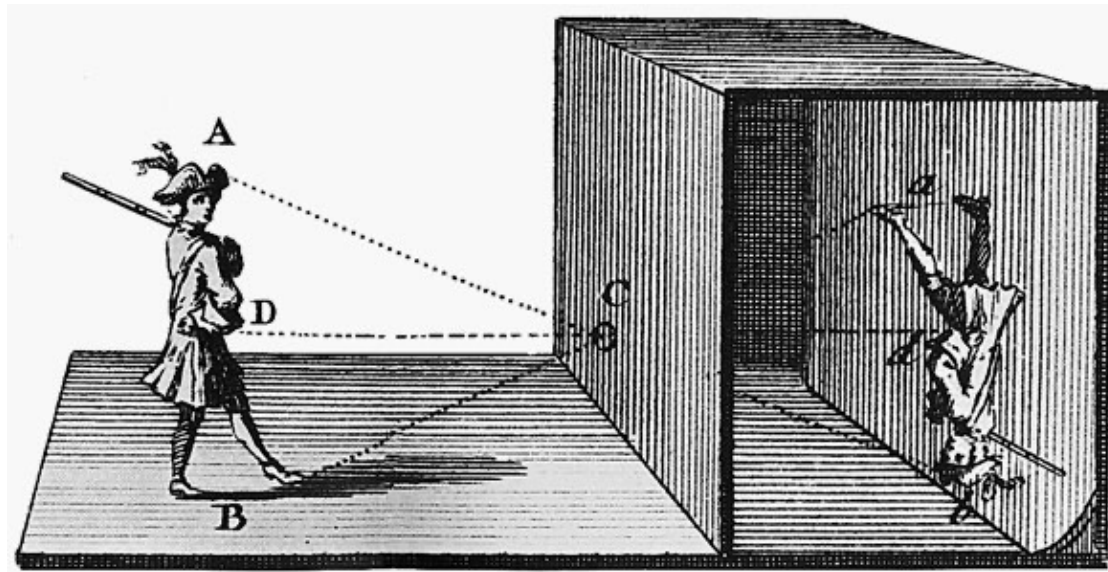
# Sistema ottico: il cristallino

60/70 % acqua

Assorbe circa 8% radiazione  
nella luce visibile,  
soprattutto  $\lambda$  minori

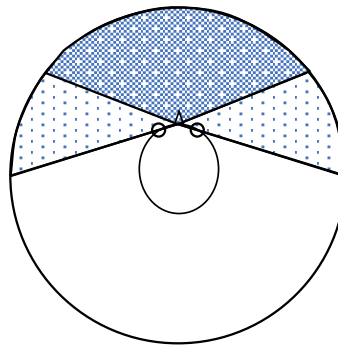


# Il sistema ottico: la camera oscura

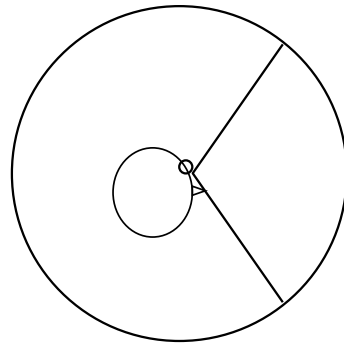


# Campo visivo

- Laterale 208°



- Verticale 120°



Aspect ratio = 1.73

$$16/9 = 1,778$$

# Il sistema sensoriale: la retina

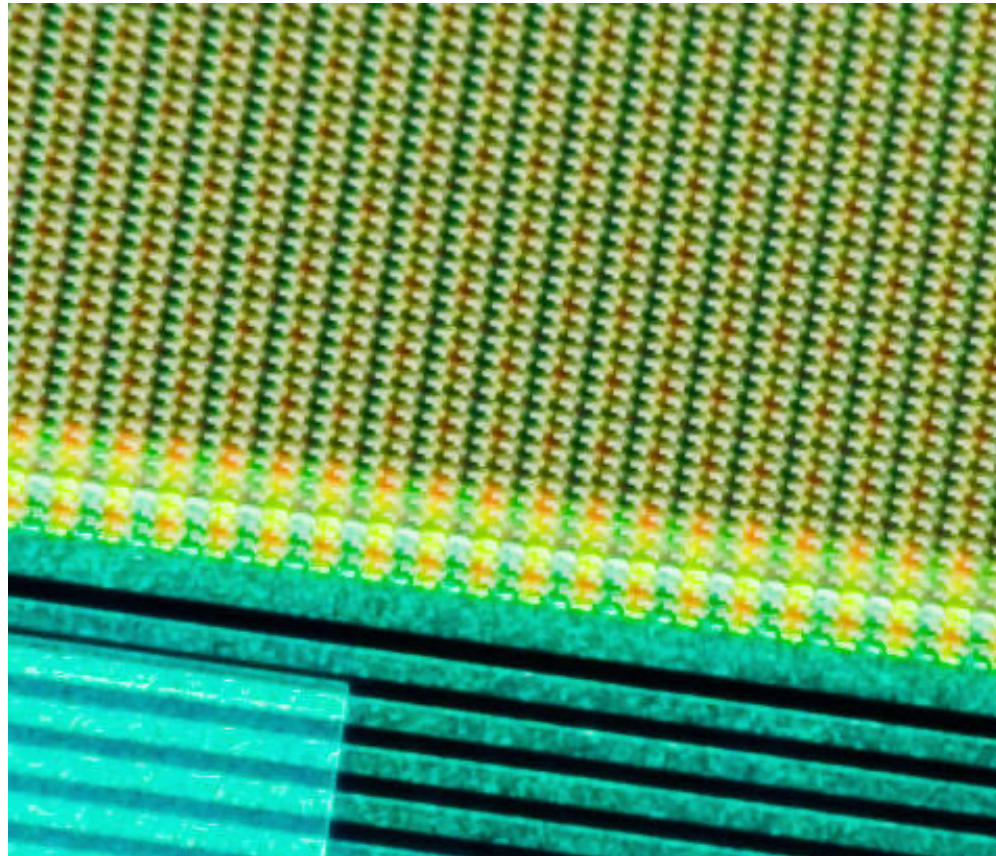
- il piano dell'immagine
- coperto di materiale fotosensibile
  - alogenuro d'argento (pellicole)
  - circuiti di silicio (telecamera CCD)
  - rodopsina e varianti (retina)

# Il sistema sensoriale: i numeri

- immagine suddivisa in pixel
- CCD: alcuni Megapixel, griglia rettangolare
  - The **Canon EOS 5DS and EOS 5DS R cameras** offer the highest resolution capture in the history of EOS: a Canon designed and manufactured, full-frame **50.6 Megapixel CMOS sensor**. It captures 8712 x 5813 effective pixels ...
  - **Sony A7R II Type:** Mirrorless | **Sensor size:** full frame, 42.4Mp | **Monitor:** 3-inch tilting, 1,229k dots | **Continuous shooting:** 5 fps | **Movies:** 4K | **User level:** Expert
- retina: da 75 a 150M bastoncelli, 6 o 7 M coni, mosaico esagonale



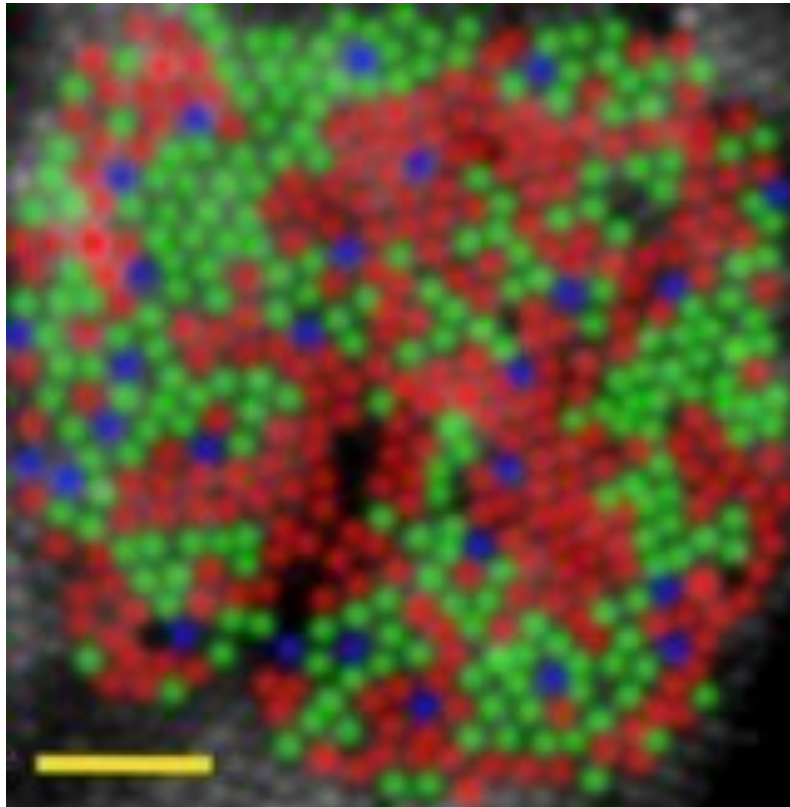
# Griglia rettangolare CCD



<http://emdma.com/technology-explained-how-does-a-digital-camera-work/>



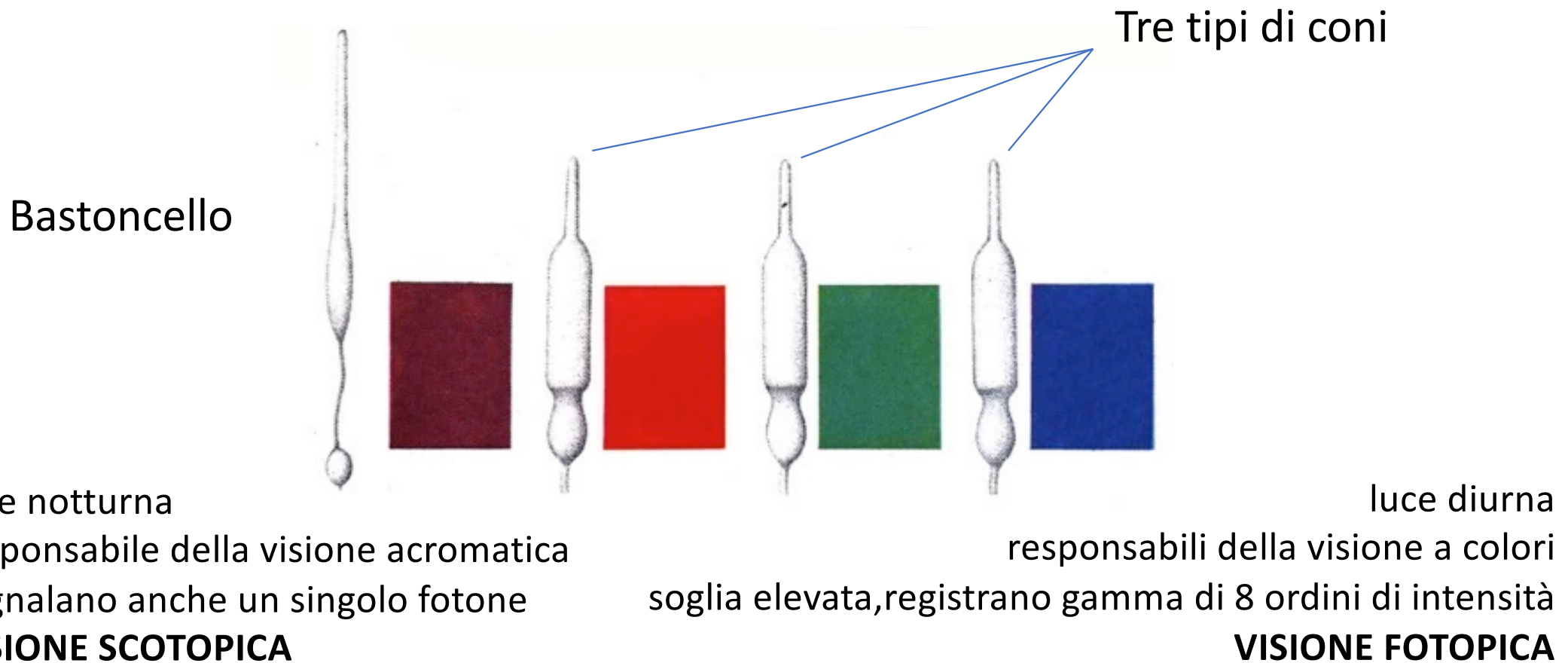
# Mosaico esagonale della retina



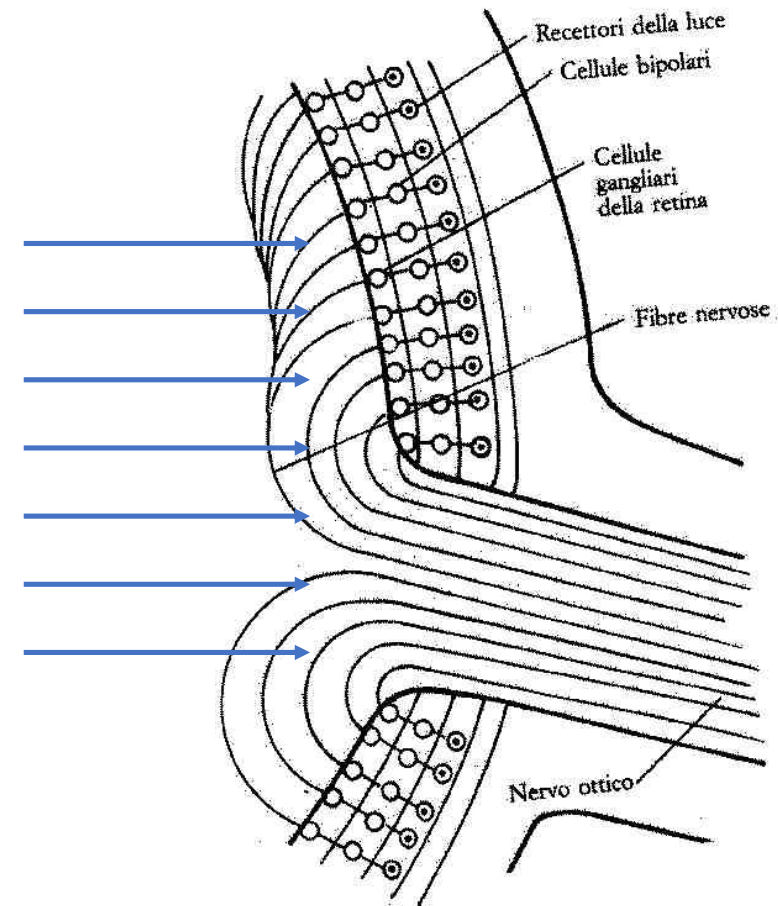
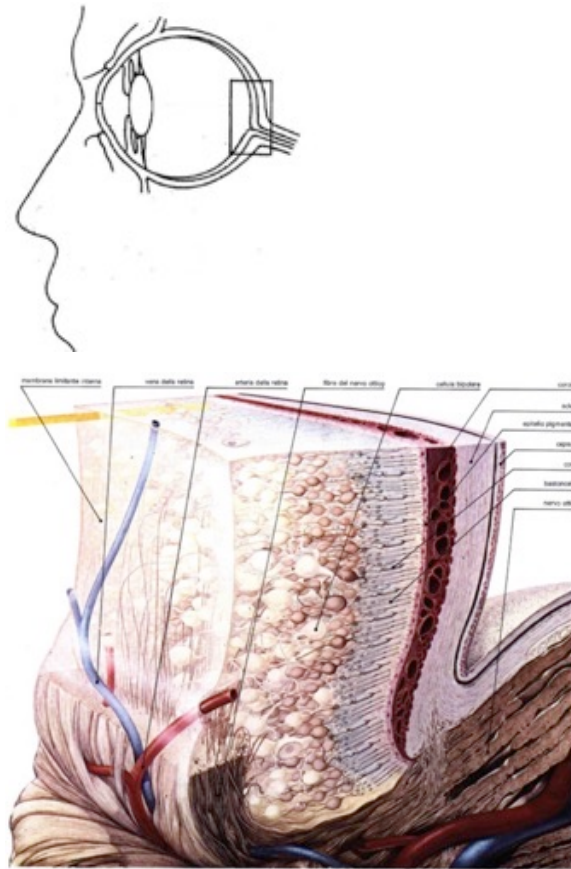
Chromatic Mosaic of the Human Retina: False color image showing the arrangement of the 'red', 'green' and 'blue' cones in the human retina at a location 1 deg nasal from the central fovea. The scale bar represents 5 arcmin of visual angle. Each cone can be as small as 28 arcsec. Adapted from [Roorda Lab](http://www.roorda.nl/).

<http://www.kybervision.com/Blog/files/AppleRetinaDisplay.html>

# Elementi fotosensibili: bastoncelli e coni

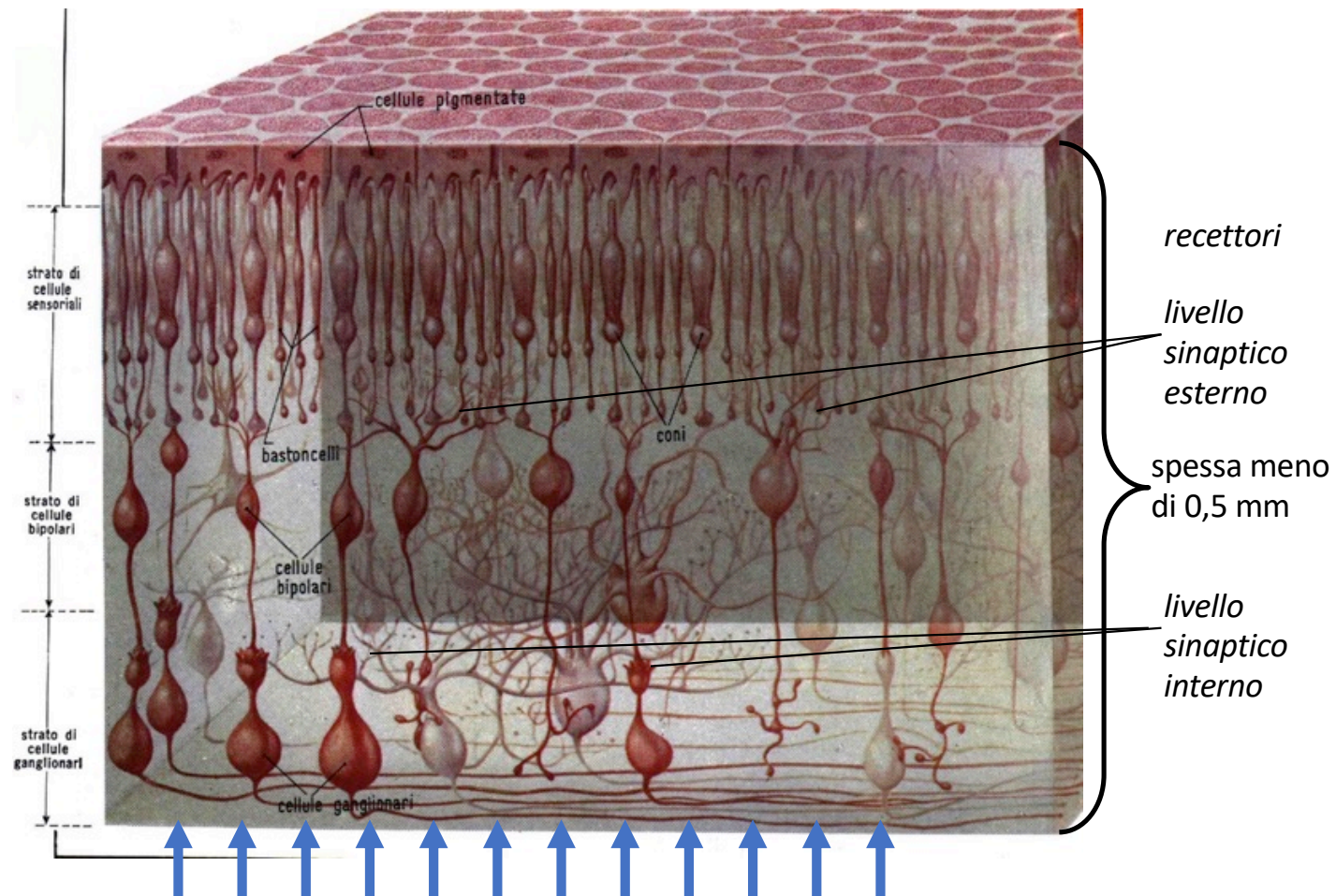


# La retina

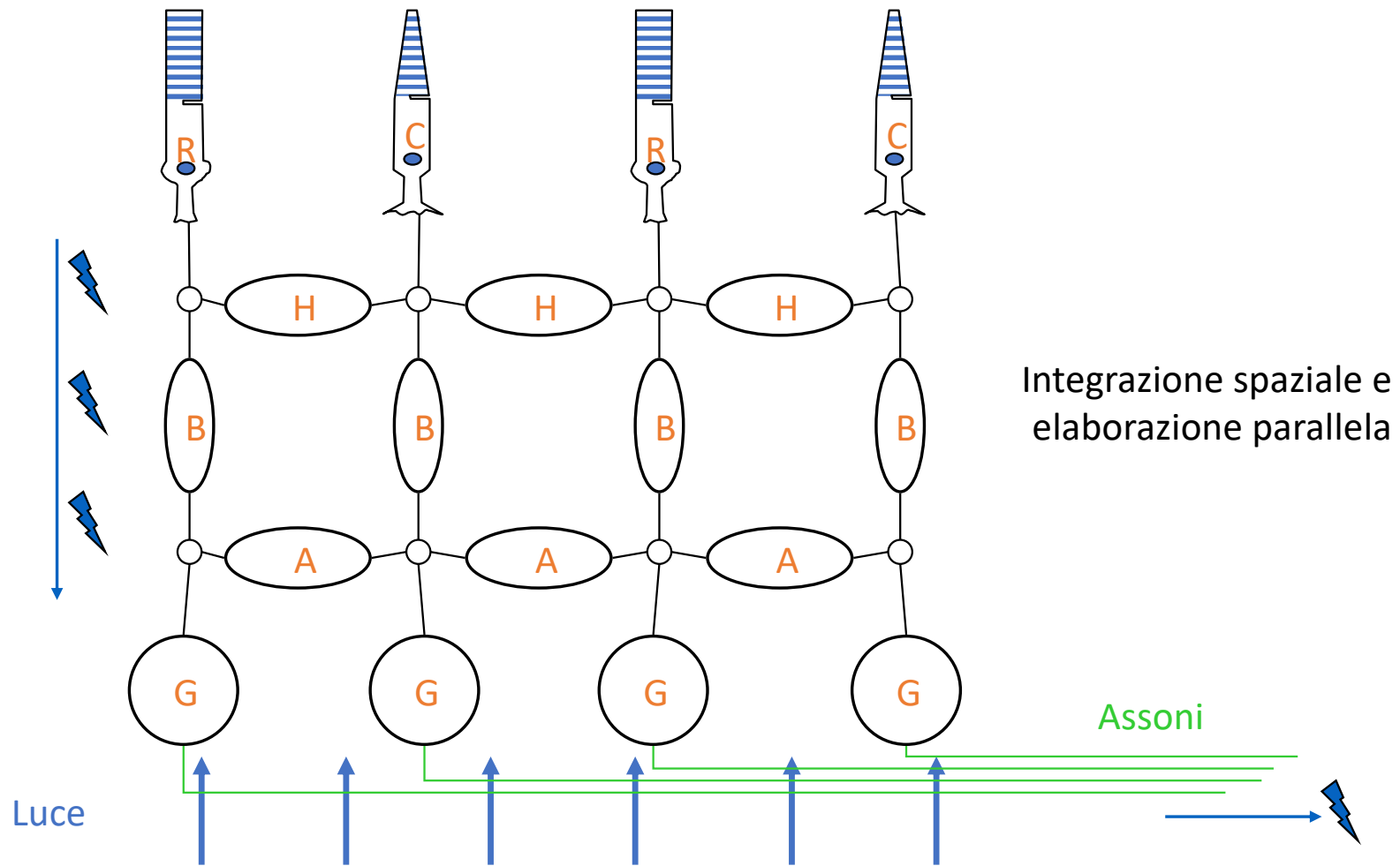


1.2 M  
di fibre nervose

# Strati retinici



# Le connessioni principali

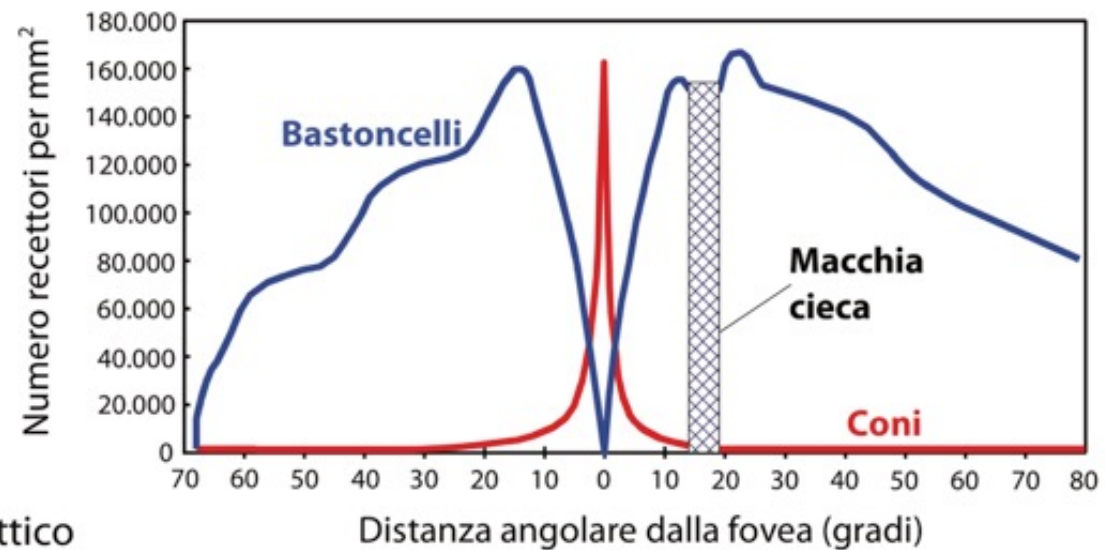
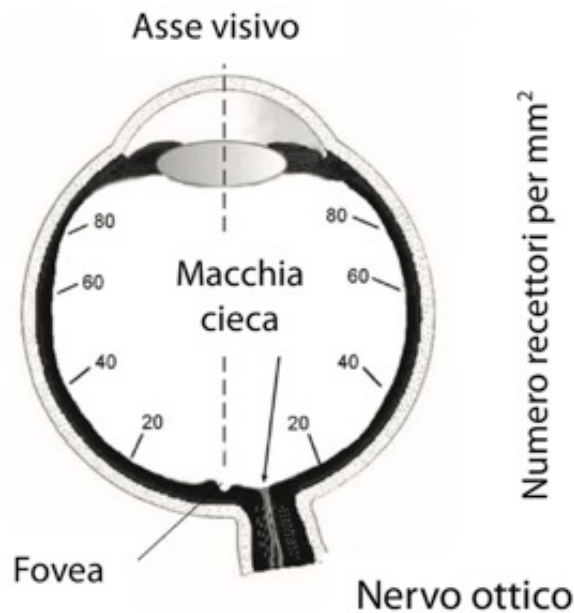




# Distribuzione dei recettori

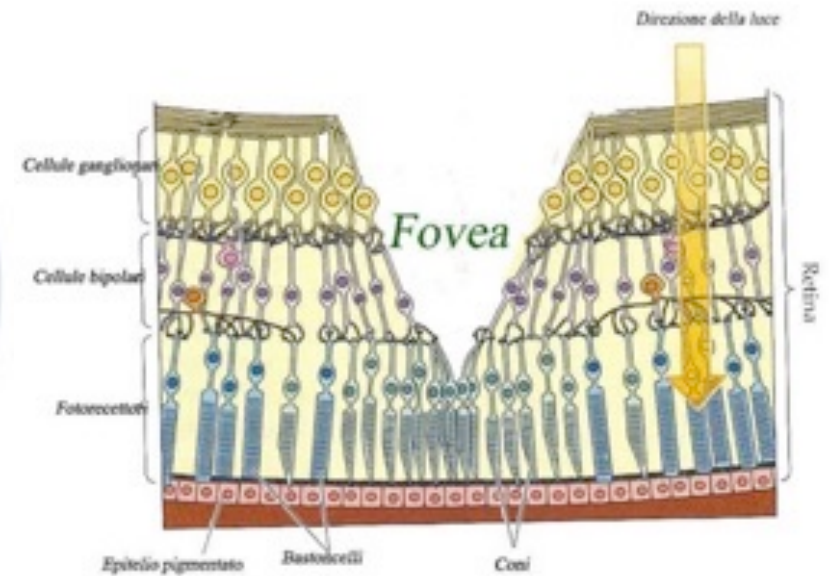
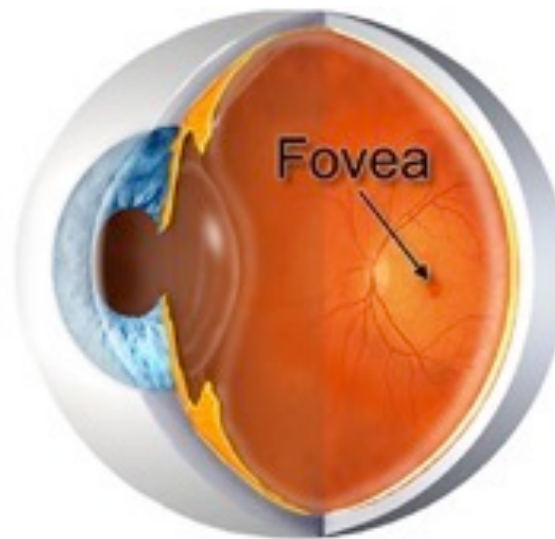
Coni ammassati nella e vicino alla fovea (max acuità visiva)

Bastoncelli assenti nella fovea; concentrati nella periferia



# Fovea

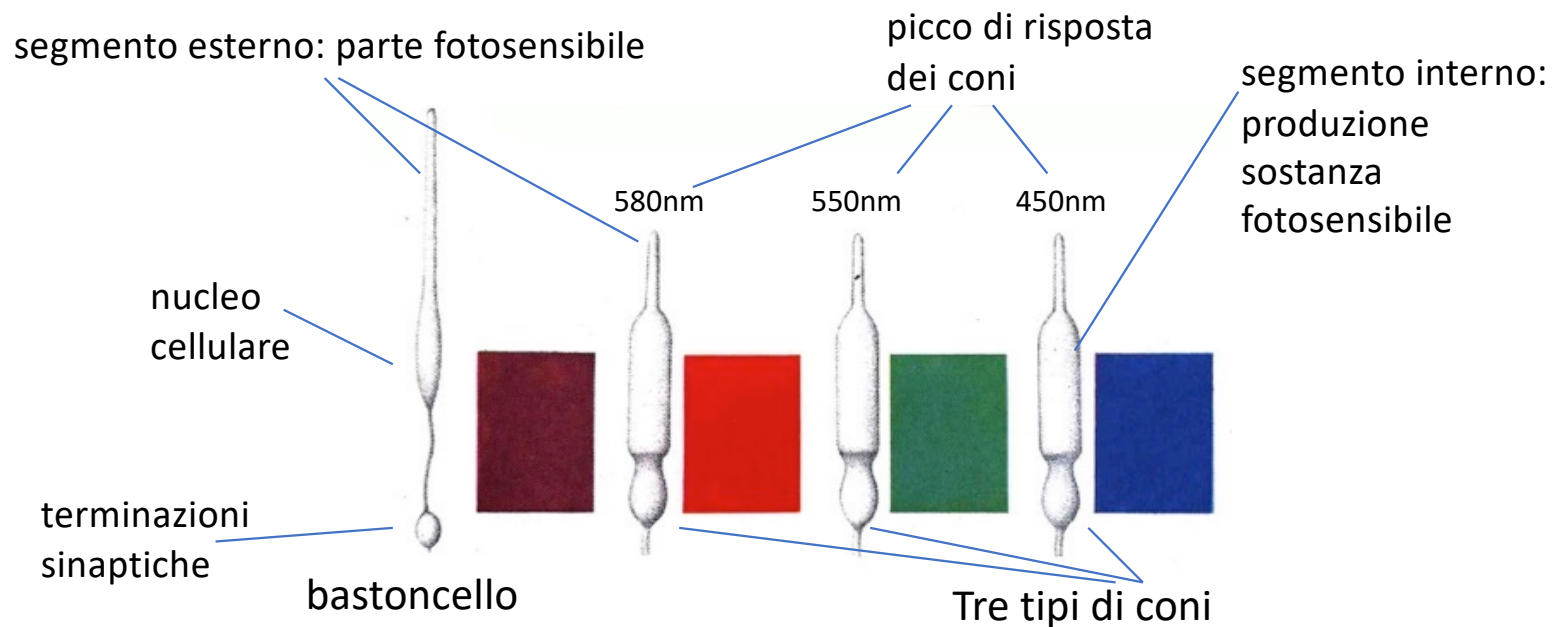
- Rientranza circolare (d 1,5 mm)
- Circa 337000 coni



<http://scienzaonline.forumativo.com/t243-cosa-occorre-sapere-sui-nostri-occhi>

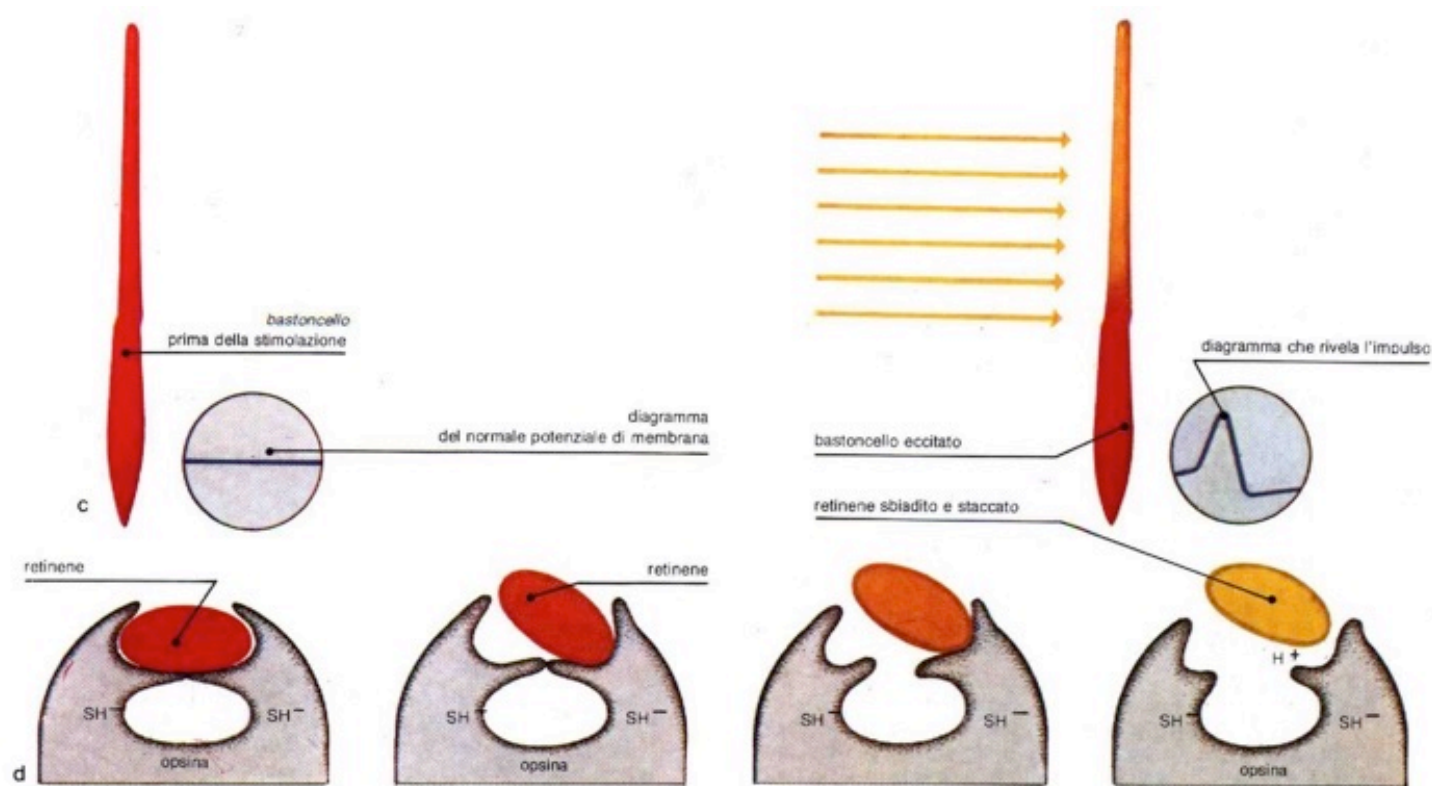
# Ipotesi del tristimolo

Qualsiasi colore reale come somma pesata delle risposte dei tre pigmenti allo stimolo di colore





# La trasduzione

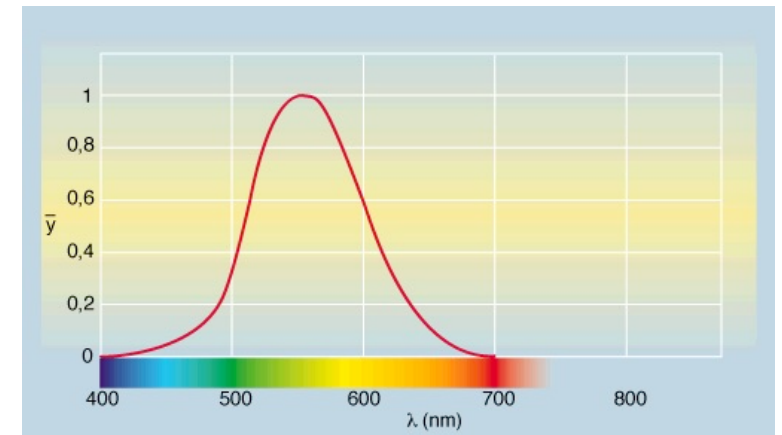
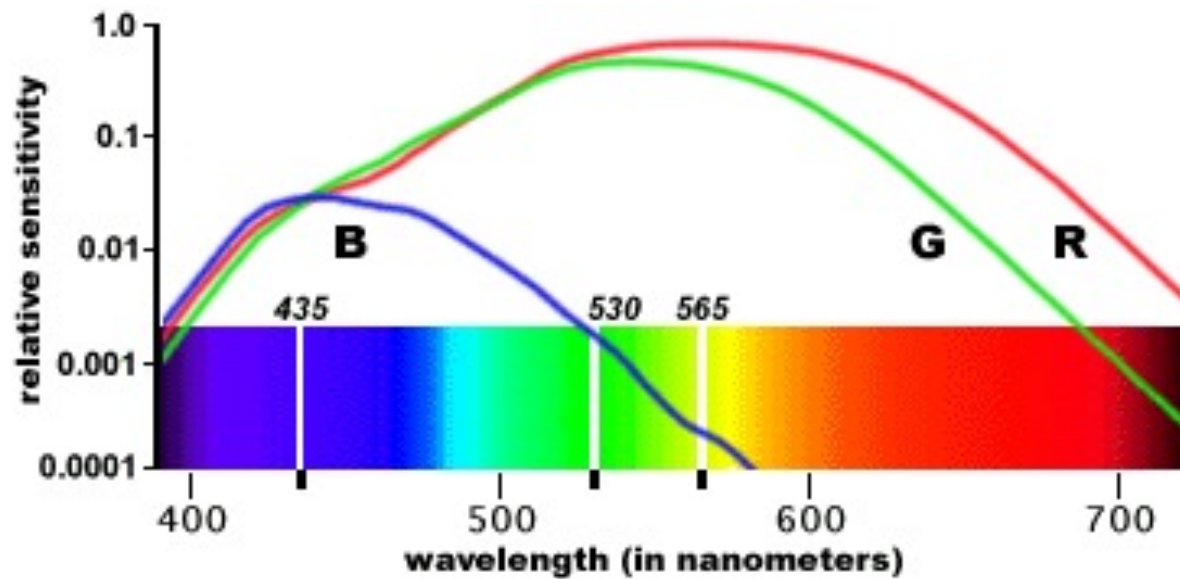


- i bastoncelli rimpiazzano i dischi continuamente
- i coni rimpiazzano il pigmento dei dischi

# Sensibilità dei fotorecettori

- Il pigmento liberato dipende proporzionalmente dalla luce assorbita
- Si riduce quindi la sensibilità alla luce
- Adattamento: variando la luce nell'ambiente, nuovo equilibrio sulla quantità di pigmento

# Efficienza luminosa dei fotorecettori



<http://www.physics.utoronto.ca/~jharlow/cones.html>

# Conclusione campionamento: la matrice a livelli di grigio

- interazione fisica tra luce focalizzata sulla retina e pigmento visivo delle cellule retiniche
- risposta sia delle cellule retiniche sia di una camera = matrice 2D dei valori di intensità

# Matrice a livelli di grigio



225	221	216	219	219	214	207	218	219	220	207	155	136	135	130	131	125
213	206	213	223	208	217	223	221	223	216	195	156	141	130	128	138	123
206	217	210	216	224	223	228	230	234	216	207	157	136	132	137	130	128
211	213	221	223	220	222	237	216	219	220	176	149	137	132	125	136	121
216	210	231	227	224	228	231	210	195	227	181	141	131	133	131	124	122
223	229	218	230	228	214	213	209	198	224	161	140	133	127	133	122	133
220	219	224	220	219	215	215	206	206	221	159	143	133	131	129	127	127
221	215	211	214	220	218	221	212	218	204	148	141	131	130	128	129	118
214	211	211	218	214	220	226	216	223	209	143	141	141	124	121	132	125
211	208	223	213	216	226	231	230	241	199	153	141	136	125	131	125	136
200	224	219	215	217	224	232	241	240	211	150	139	128	132	129	124	132
204	206	208	205	233	241	241	252	242	192	151	141	133	130	127	129	129
200	205	201	216	232	248	255	246	231	210	149	141	132	126	134	128	139
191	194	209	238	245	255	249	235	238	197	146	139	130	132	129	132	123
189	199	200	227	239	237	235	236	247	192	145	142	124	133	125	138	128
198	196	209	211	210	215	236	240	232	177	142	137	135	124	129	132	128
198	203	205	208	211	224	226	240	210	160	139	132	129	130	122	124	131
216	209	214	220	210	231	245	219	169	143	148	129	128	136	124	128	123
211	210	217	218	214	227	244	221	162	140	139	129	133	131	122	126	128
215	210	216	216	209	220	248	200	156	139	131	129	139	128	123	130	128
219	220	211	208	205	209	240	217	154	141	127	130	124	142	134	128	129
229	224	212	214	220	229	234	208	151	145	128	128	142	122	126	132	124
252	224	222	224	233	244	228	213	143	141	135	128	131	129	128	124	131
255	235	230	249	253	240	228	193	147	139	132	128	136	125	125	128	119
250	245	238	245	246	235	235	190	139	136	134	135	126	130	126	137	132
240	238	233	232	235	255	246	168	156	141	129	127	136	134	135	130	126
241	242	225	219	225	255	255	183	139	141	126	139	128	137	128	128	130
234	218	221	217	211	252	242	166	144	139	132	130	128	129	127	121	132
231	221	219	214	218	225	238	171	145	141	124	134	131	134	131	126	131
228	212	214	214	213	208	209	159	134	136	139	134	126	127	127	124	122
219	213	215	215	205	215	222	161	135	141	128	129	131	128	125	128	127

# Primo livello

Campionamento

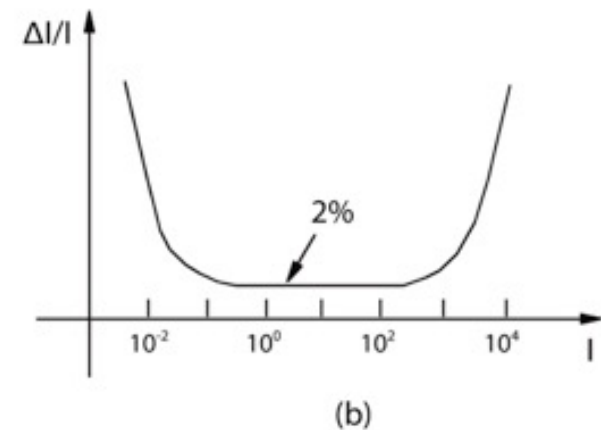
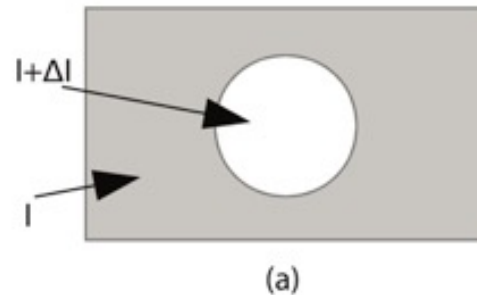
Localizzazione dei contorni

Schema primario

## Risposta alle variazioni di intensità Sensibilità al contrasto

- Non è lineare

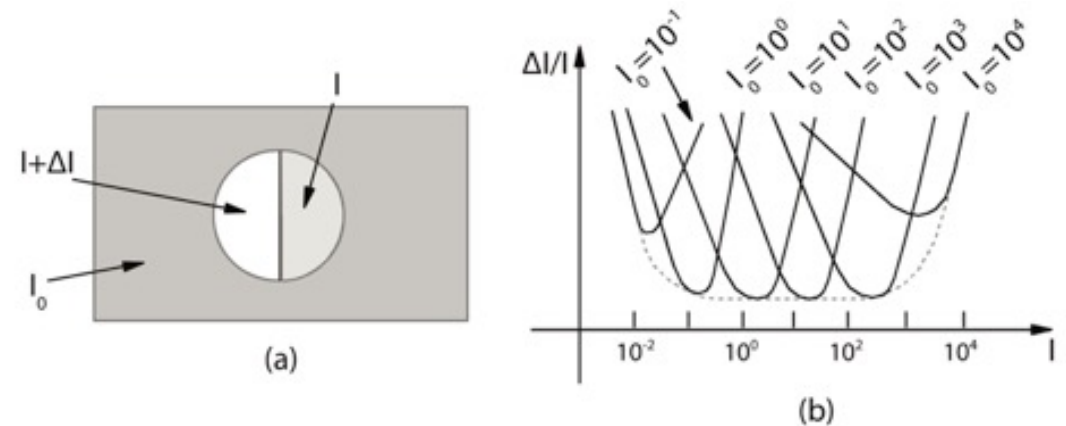
*Situazione  
sperimentale*



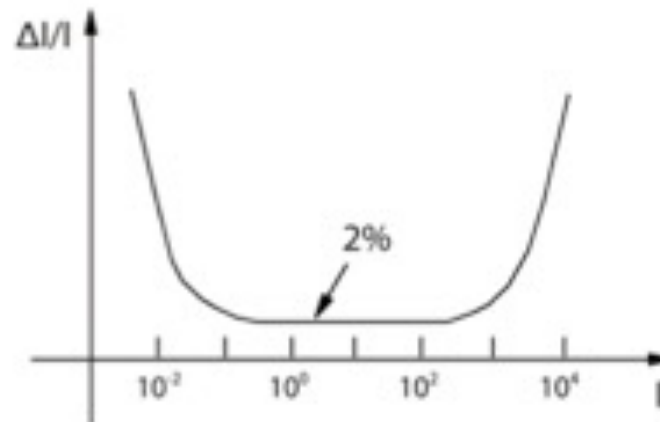
- Si aumenta  $\Delta I$  fino a una differenza percepibile
- $\Delta I / I$  (*frazione di Weber*) è quasi sempre costante (circa 2%)
- $\Delta I$ , dipendente da  $I$ , è detta *soglia di contrasto*

# Il background

- Frazione di Weber e gamma di valori in cui è costante dipendono da luminosità del bg

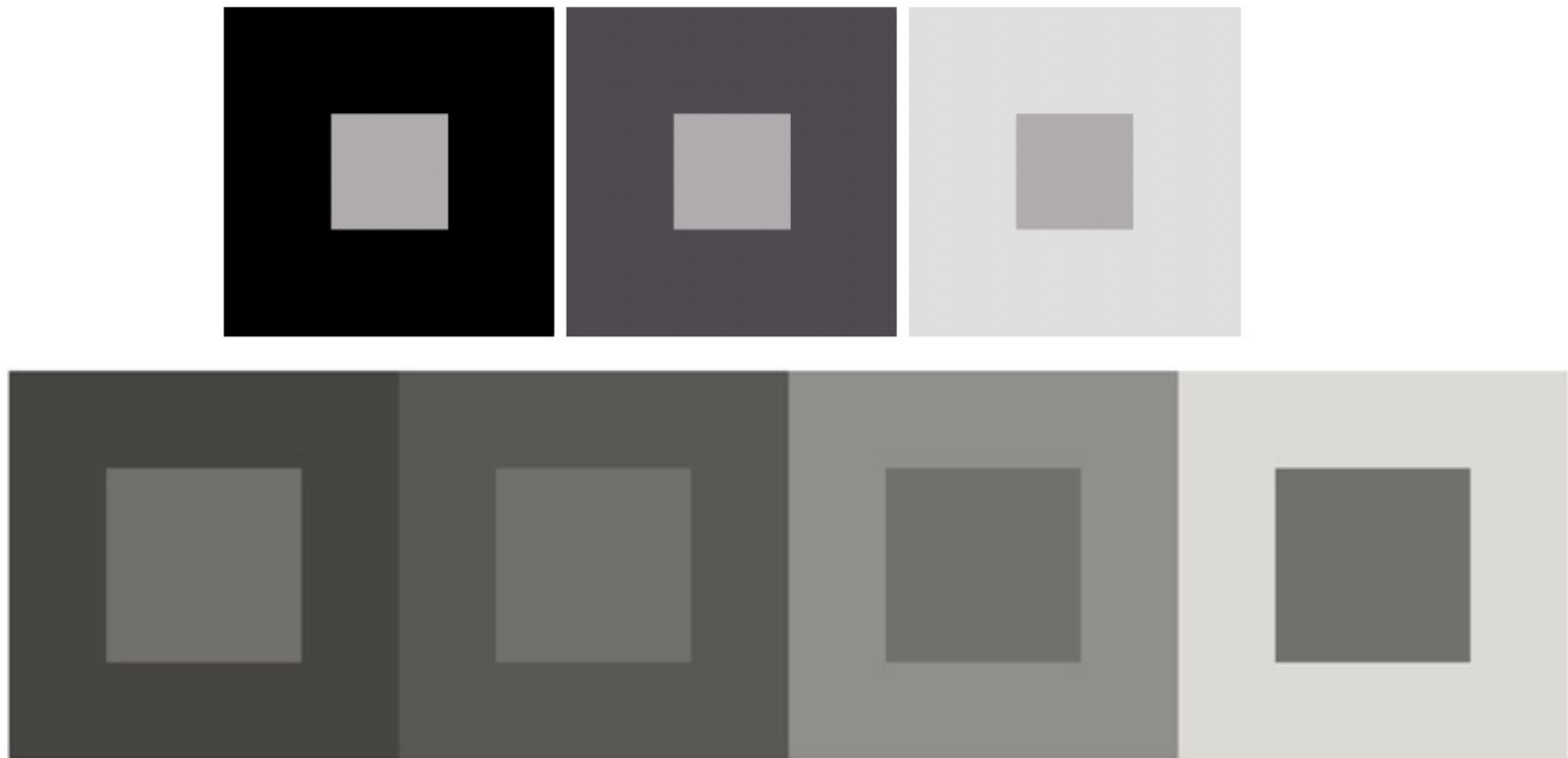


- L'inviluppo delle varie risposte corrisponde alla situazione precedente



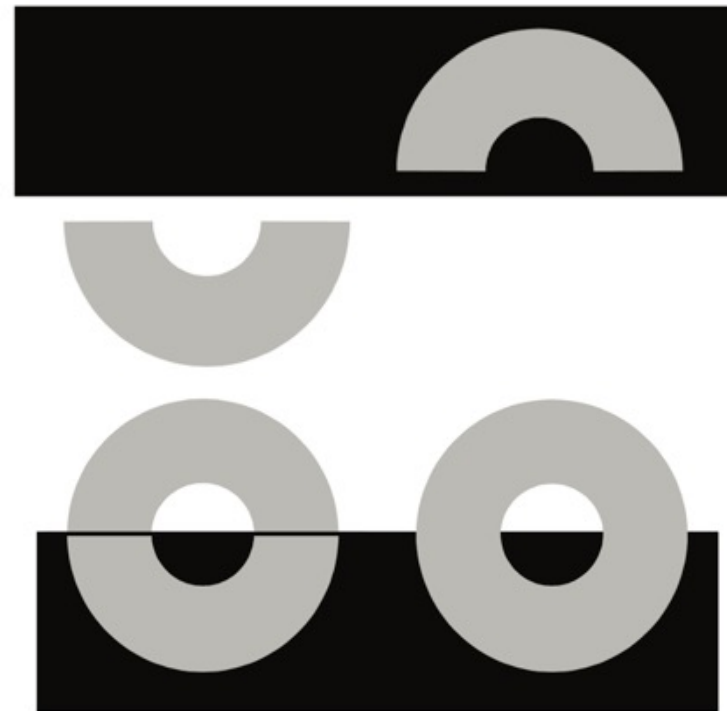
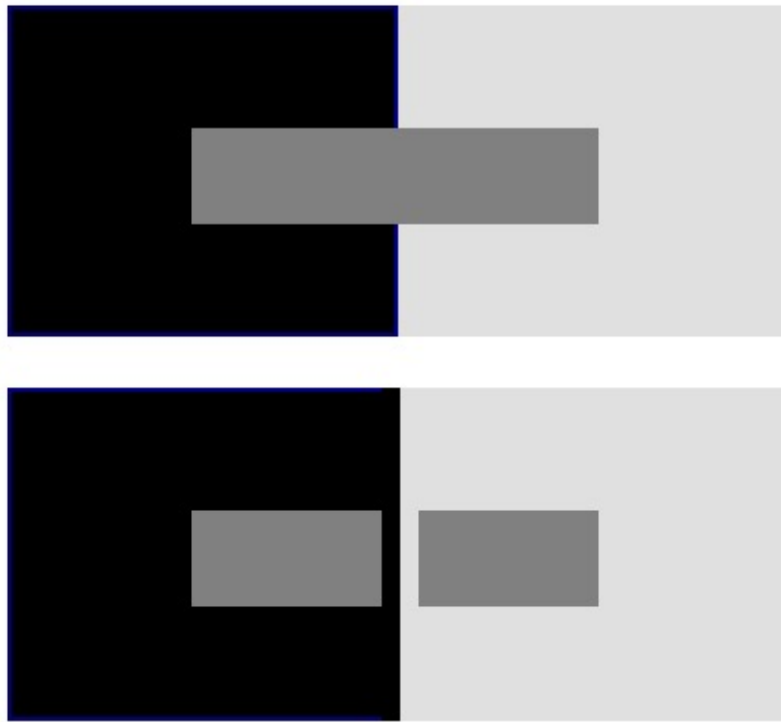


## Il contrasto simultaneo



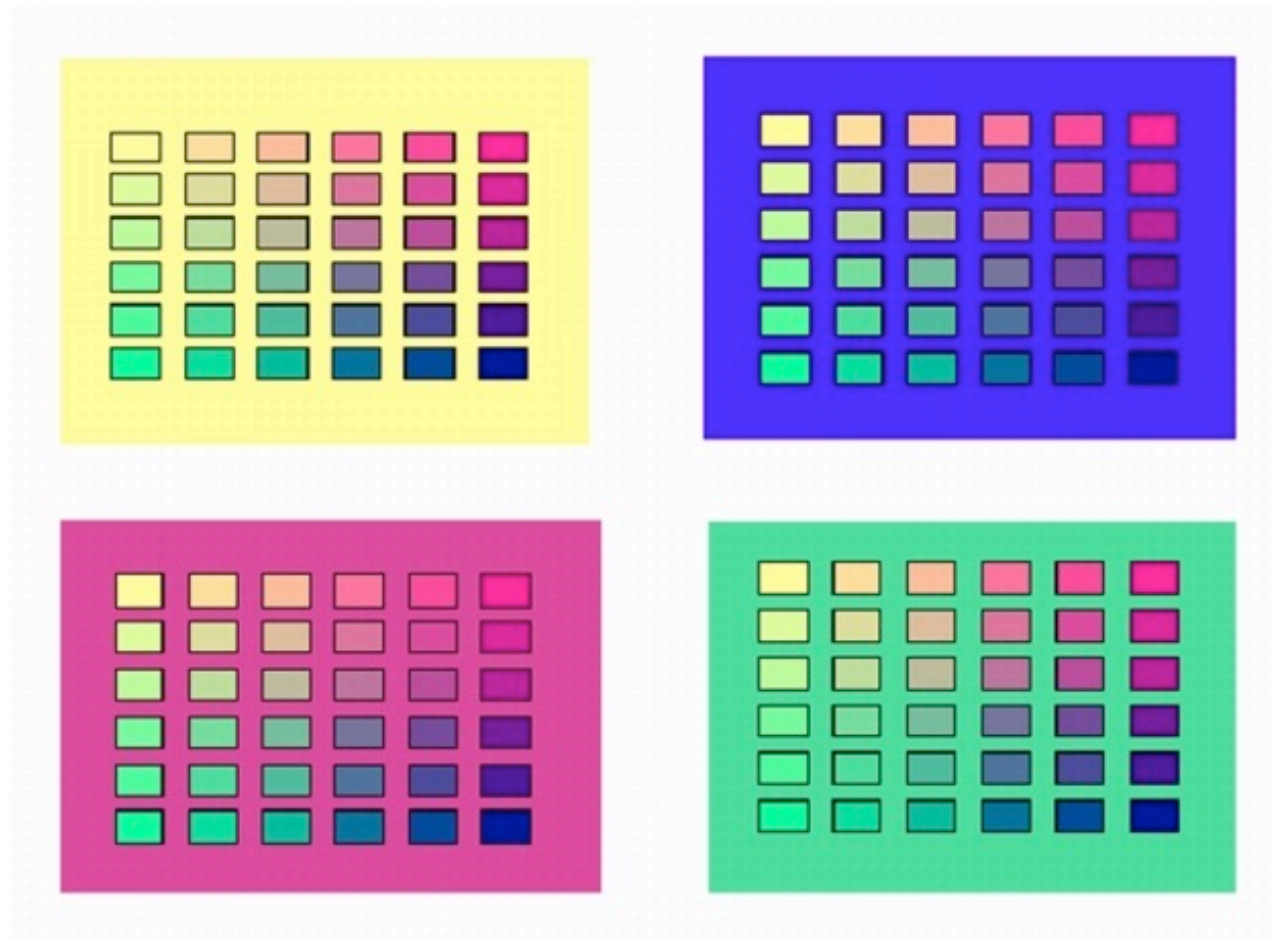
La sensibilità dipende dal contesto

## Il contrasto simultaneo

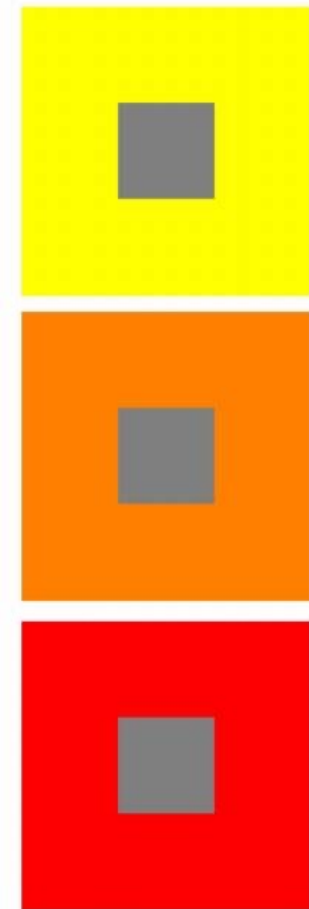
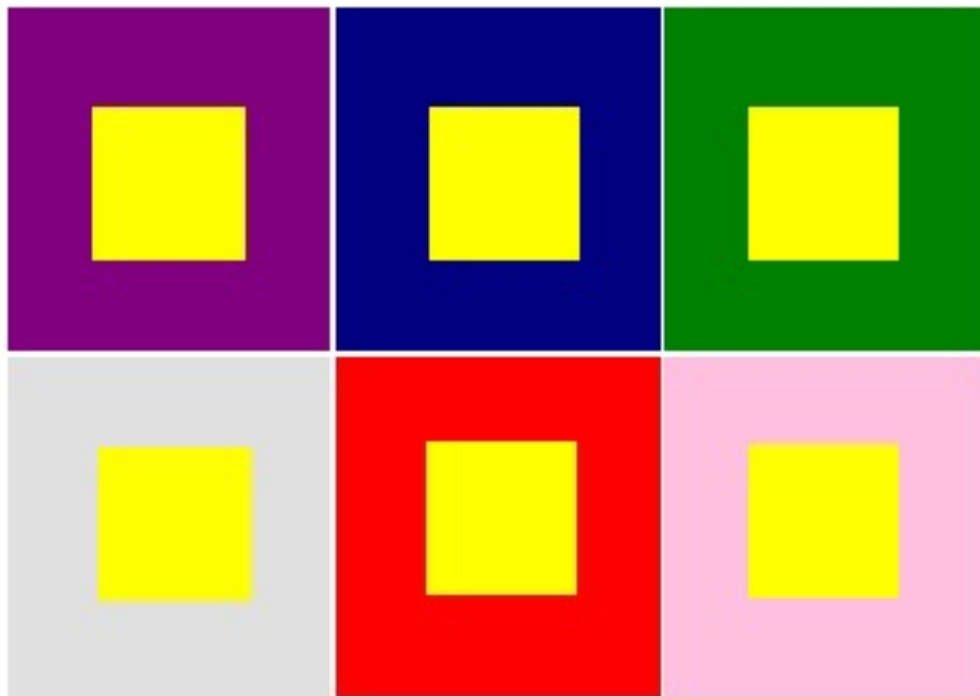


Propagazione e omogeneizzazione degli effetti di contrasto (filling-in) entro zone delimitate da bordi netti

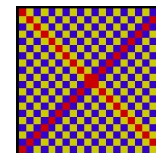
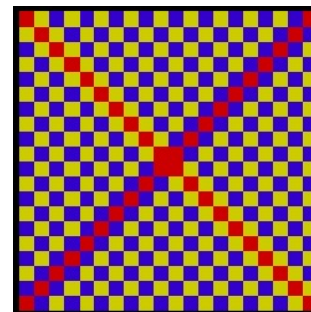
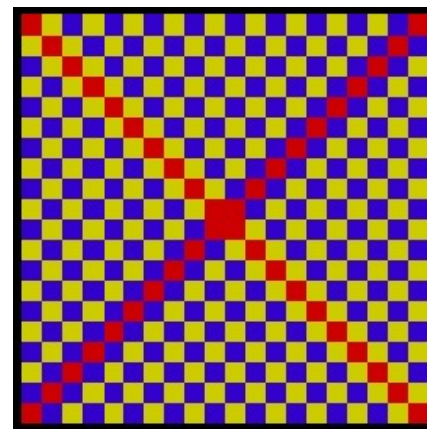
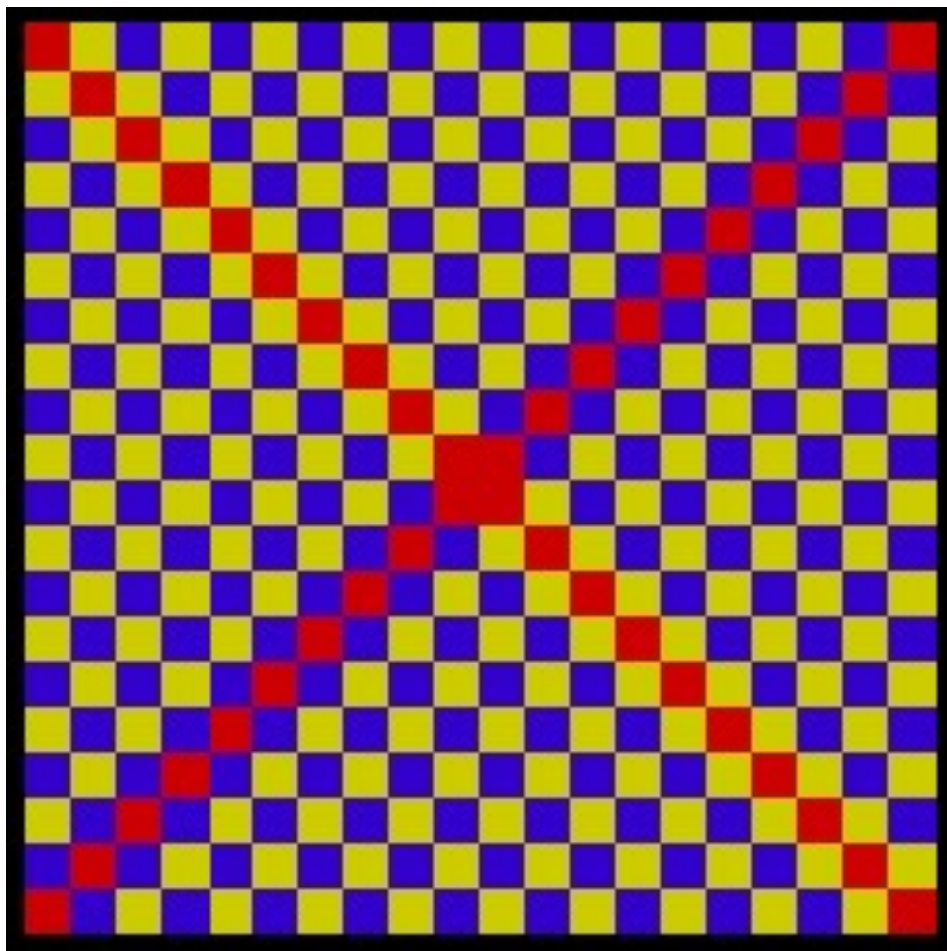
# Contrasto simultaneo a colori



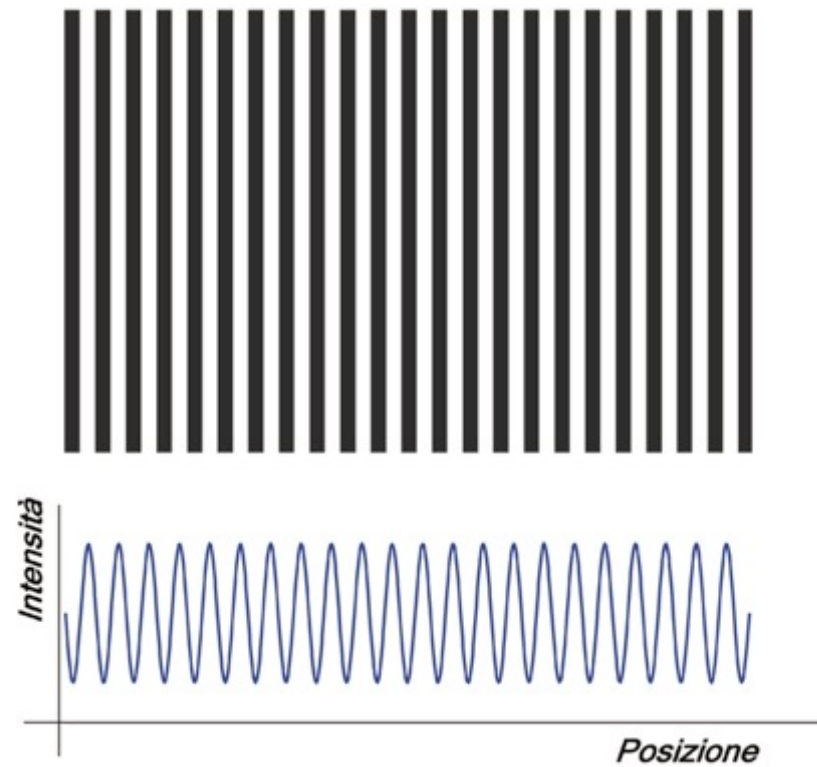
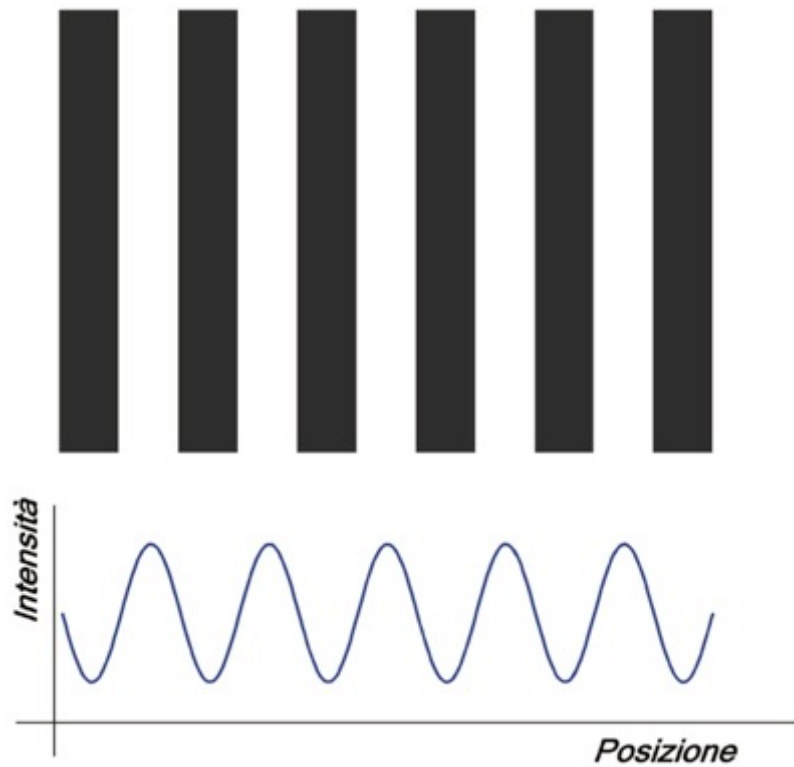
# Contrasto cromatico di Itten



# Assimilazione dei colori

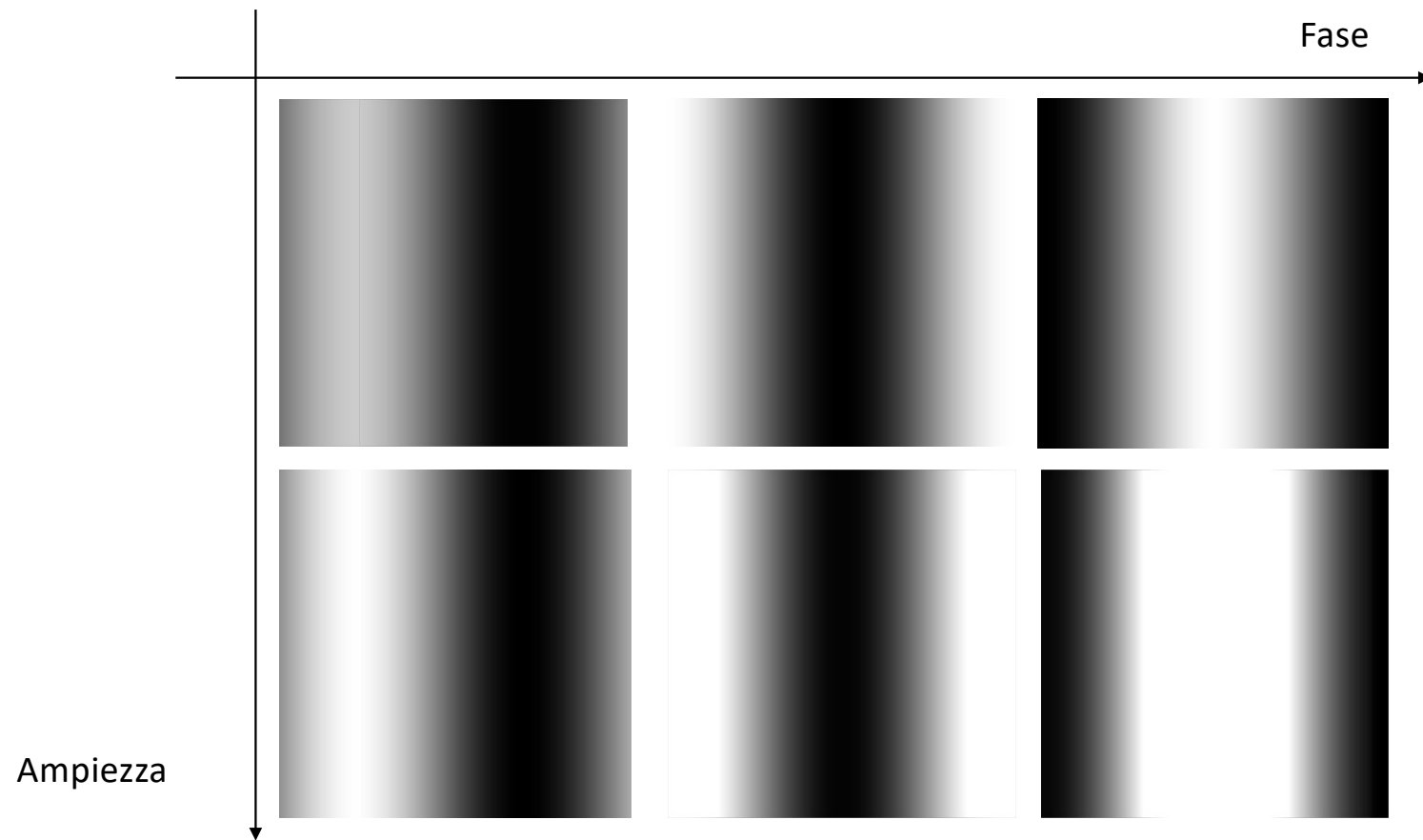


# Frequenza spaziale

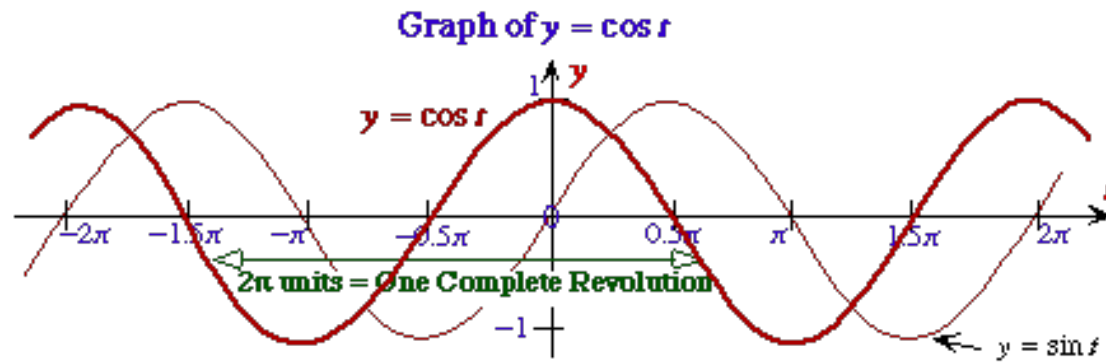
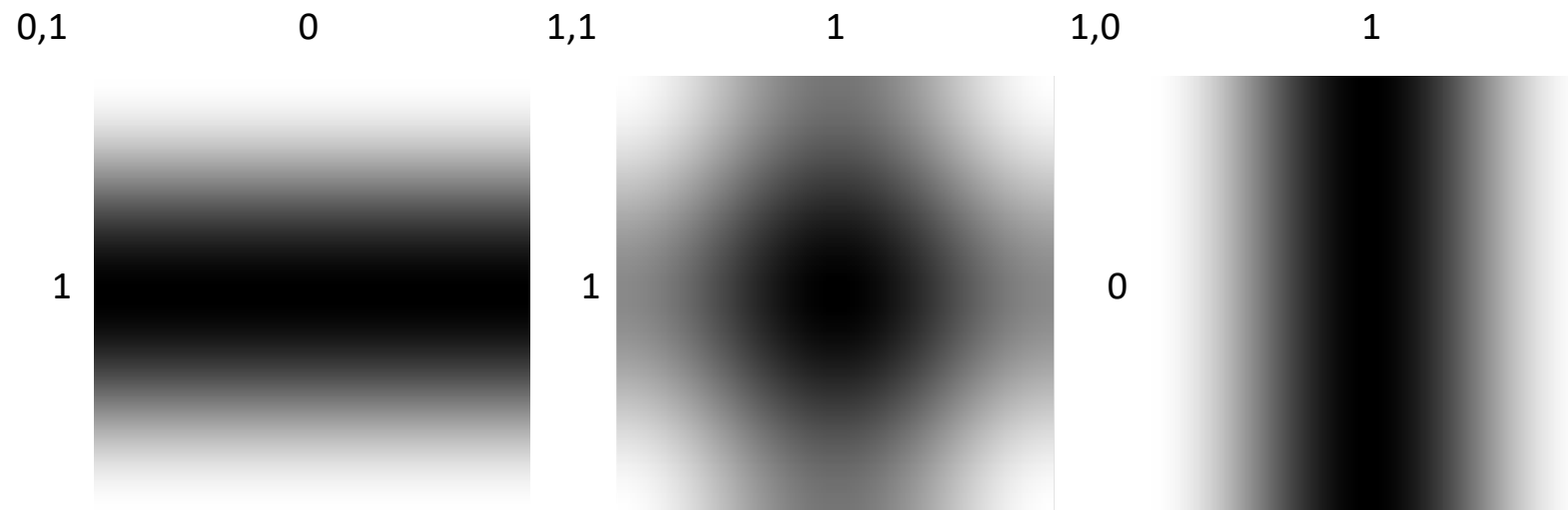


Reticoli orizzontali

# Reticoli (Teorema di Fourier)

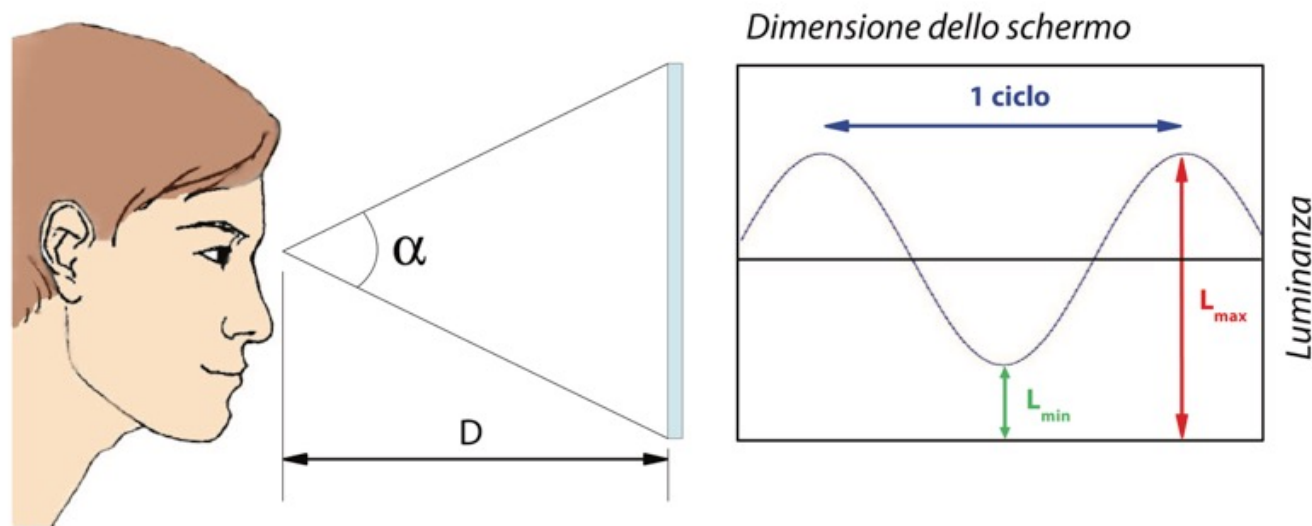


# Reticolo (frequenza 1)





# Misura della frequenza spaziale



Numero di sinusoidi che cadono in un grado di campo visivo: *Cicli/grado*

Grado di campo visivo: angolo che sottende uno schermo di 1 cm di lato alla distanza  $D = 57$  cm (allontanandosi, frequenza aumenta)

# I contorni: variazioni di intensità

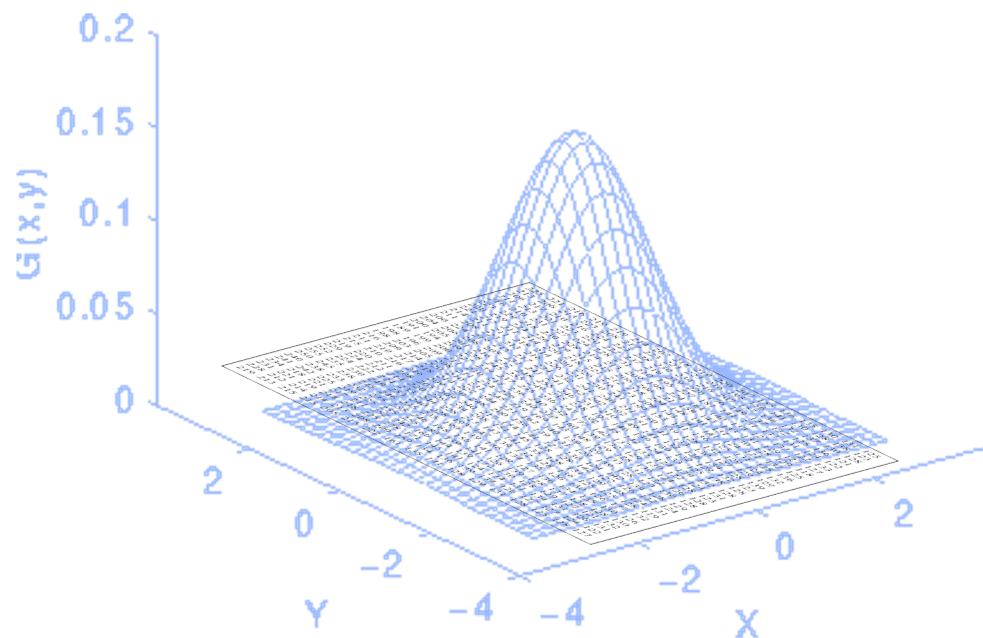
- regioni con diverse intensità luminose
  - chiazze brillanti e opache (direzione della luce)
  - l'intensità tende a cambiare in prossimità degli spigoli
  - potenza dei disegni senza sfumature
- localizzare i **contorni**
  - passaggi tra regioni di diversa intensità
  - problema del “rumore”

# Come si localizzano i contorni

- Si filtra l'immagine in modo da ...
  - attenuare le piccole differenze
  - esaltare le differenze significative
- Filtraggio mediante più scale di valori
  - rappresentazione struttura grossolana dei contorni
  - mantenimento dettagli sulle superfici

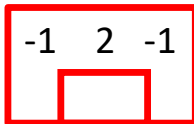
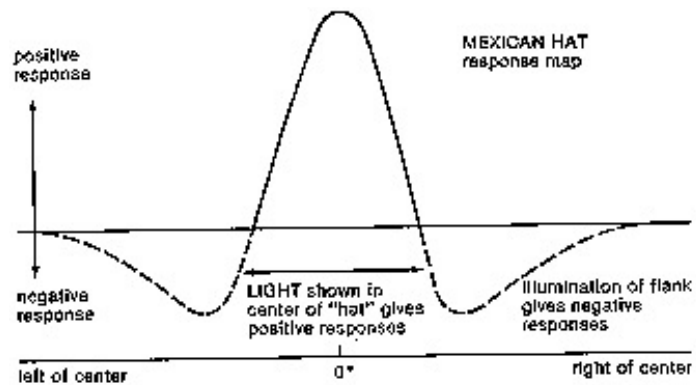
# La gaussiana

- tiene conto di una più ampia gamma di valori vicini
- attribuisce un peso in funzione della distanza



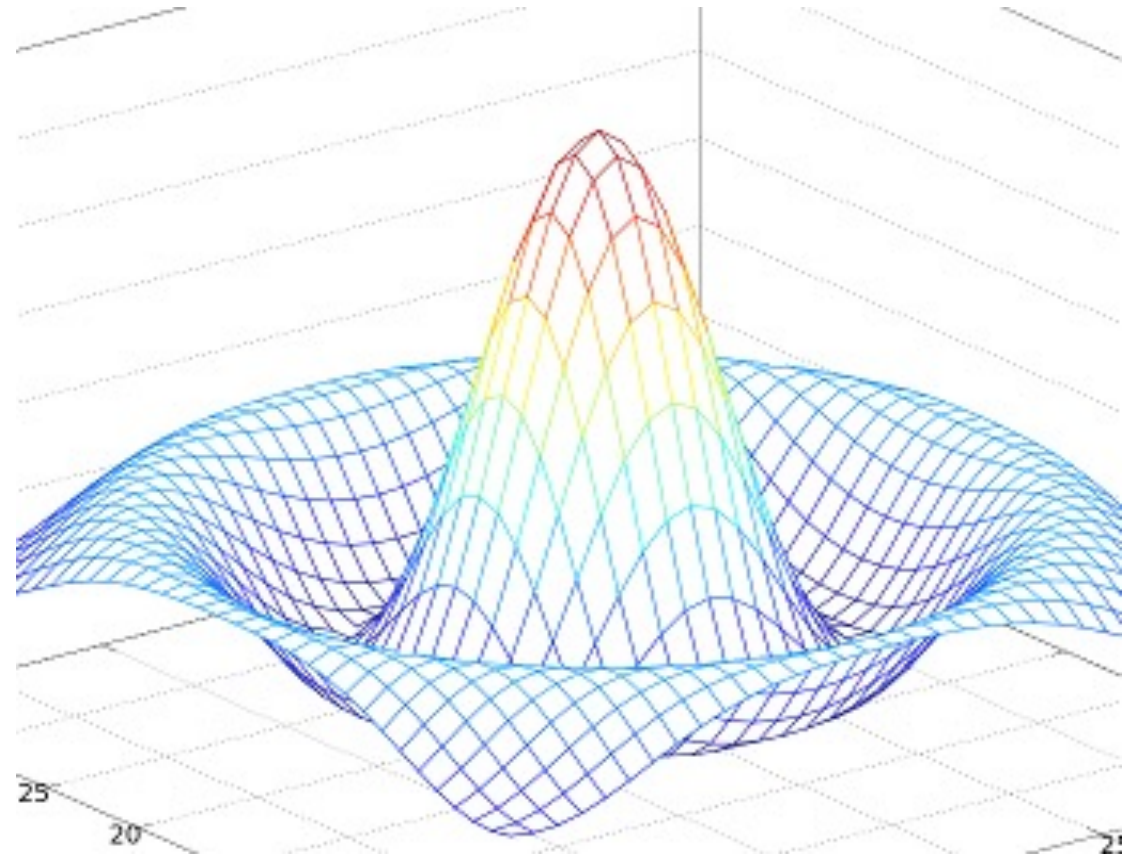
225	221	216	219	219	214	207	218	219	220	207	155	136	135	130	131	125
213	206	213	223	208	217	223	221	223	216	195	156	141	130	128	138	123
206	217	210	216	224	223	228	230	234	216	207	157	136	132	137	130	128
211	213	221	223	220	222	237	216	219	220	176	149	137	132	125	136	121
216	210	231	227	224	228	231	210	195	227	181	141	131	133	131	124	122
223	229	218	230	228	214	213	209	198	224	161	140	133	127	133	122	133
220	219	224	220	219	215	215	206	206	221	159	143	133	131	129	127	127
221	215	211	214	220	218	221	212	218	204	148	141	131	130	128	129	118
214	211	211	218	214	220	226	216	223	209	143	141	141	124	121	132	125
211	208	223	213	216	226	231	230	241	199	153	141	136	125	131	125	136
200	224	219	215	217	224	232	241	240	211	150	139	128	132	129	124	132
204	206	208	205	233	241	241	252	242	192	151	141	133	130	127	129	129
200	205	201	216	232	248	255	246	231	210	149	141	132	126	134	128	139
191	194	209	238	245	255	249	235	238	197	146	139	130	132	129	132	123
189	199	200	227	239	237	235	236	247	192	145	142	124	133	125	138	128
198	196	209	211	210	215	236	240	232	177	142	137	135	124	129	132	128
198	203	205	208	211	224	226	240	210	160	139	132	129	130	122	124	131
216	209	214	220	210	231	245	219	169	143	148	129	128	136	124	128	123
211	210	217	218	214	227	244	221	162	140	139	129	133	131	122	126	128
215	210	216	216	209	220	248	200	156	139	131	129	139	128	123	130	128
219	220	211	208	205	209	240	217	154	141	127	130	124	142	134	128	129
229	224	212	214	220	229	234	208	151	145	128	128	142	122	126	132	124
252	224	222	224	233	244	228	213	143	141	135	128	131	129	128	124	131
255	235	230	249	253	240	228	193	147	139	132	128	136	125	125	128	119
250	245	238	245	246	235	235	190	139	136	134	135	126	130	126	137	132
240	238	233	232	235	255	246	168	156	141	129	127	136	134	135	130	126
241	242	225	219	225	255	255	183	139	141	126	139	128	137	128	128	130
234	218	221	217	211	252	242	166	144	139	132	130	128	129	127	121	132
231	221	219	214	218	225	238	171	145	141	124	134	131	134	131	126	131
228	212	214	214	213	208	209	159	134	136	139	134	126	127	127	124	122
219	213	215	215	205	215	222	161	135	141	128	129	131	128	125	128	127

# Il sombrero messicano



... 50 50 50 100 50 50 50 ...

... 0 0 100 0 0 ...



<http://williamcalvin.com/bk4/bk4ch6.htm>, <http://mattoneal.com/>

# Filtri convolutivi

- Ogni “vetro”, maschera, contiene dei pesi
- Pesi = coefficienti dei valori dei pixel nell’intorno
- Prodotti peso-valore sommati

$$N = \begin{pmatrix} \begin{matrix} L_{1,1} & L_{1,2} & L_{1,3} \\ L_{2,1} & L_{2,2} & L_{2,3} \\ L_{3,1} & L_{3,2} & L_{3,3} \end{matrix} * \begin{matrix} P_{1,1} & P_{1,2} & P_{1,3} \\ P_{2,1} & P_{2,2} & P_{2,3} \\ P_{3,1} & P_{3,2} & P_{3,3} \end{matrix} \end{pmatrix} \cdot K \cdot G$$

N = Nuovo valore calcolato del pixel

$L_{ij}$  = valori di luminanza dei pixel originali

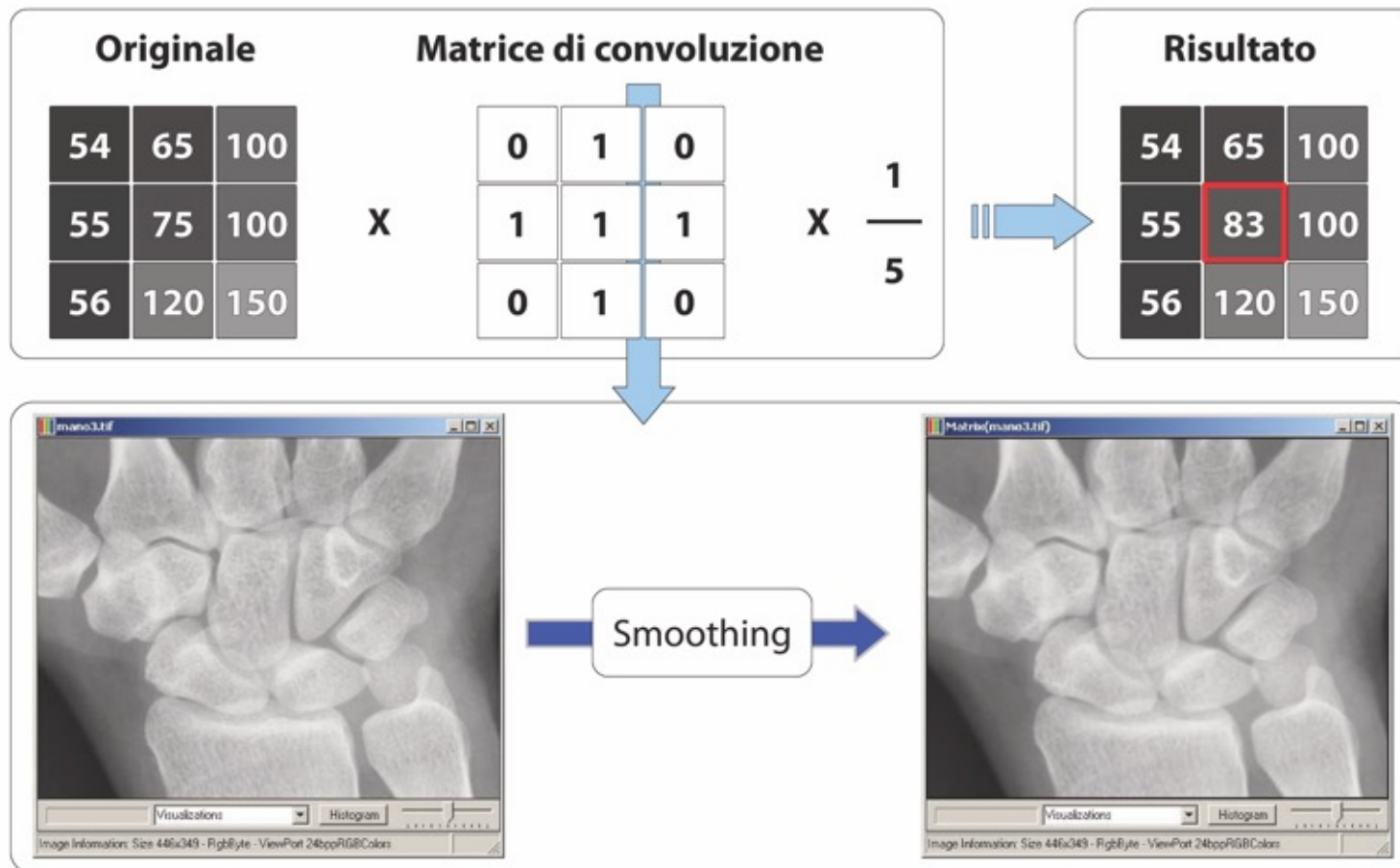
$*$  =  $(L_{1,1} \times P_{1,1}) + (L_{1,2} \times P_{1,2}) + \dots + (L_{3,3} \times P_{3,3})$

$P_{ij}$  = pesi della matrice di convoluzione

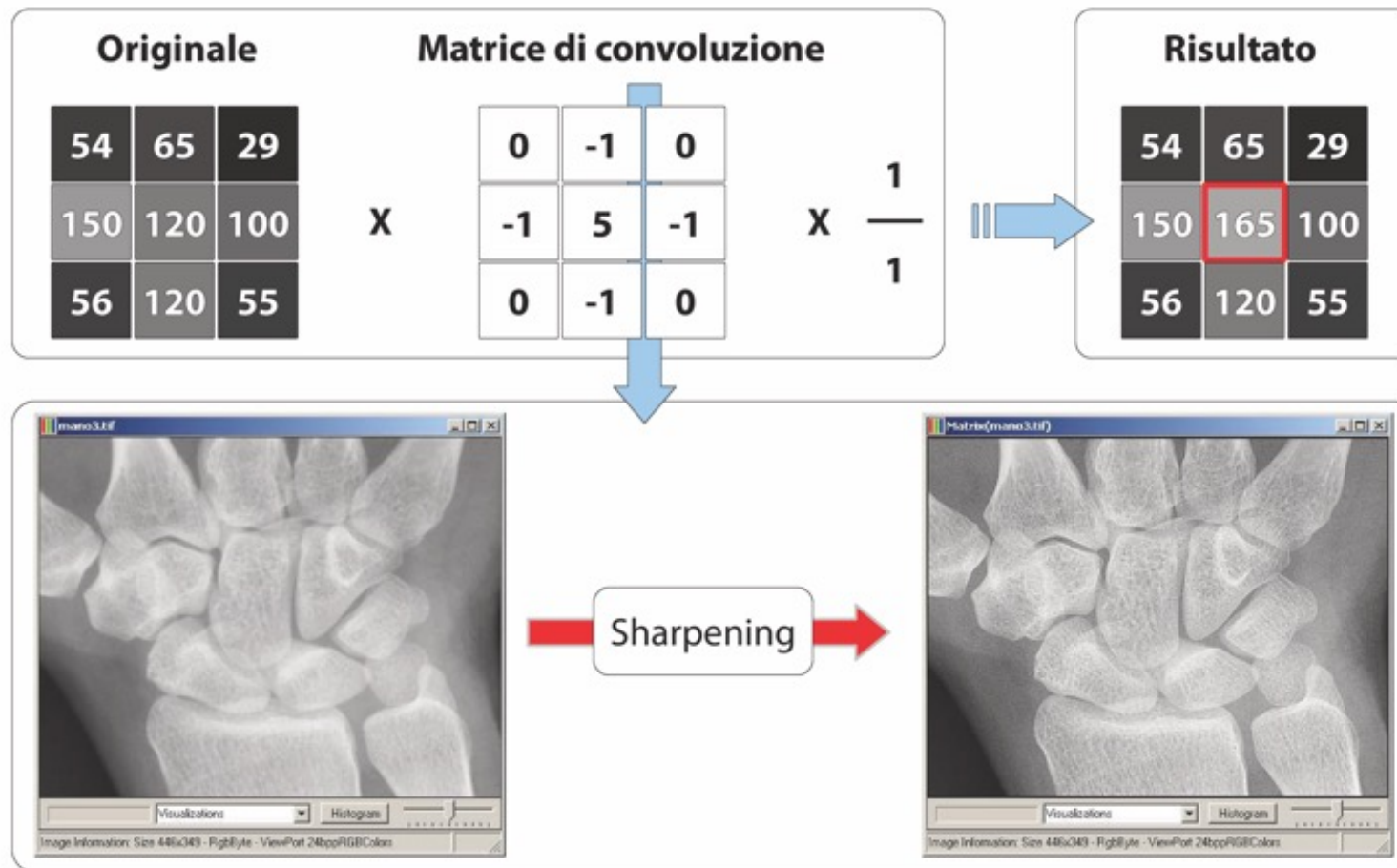
G = guadagno globale

$K = 1 / (P_{1,1} + \dots + P_{ij})$

# Smoothing

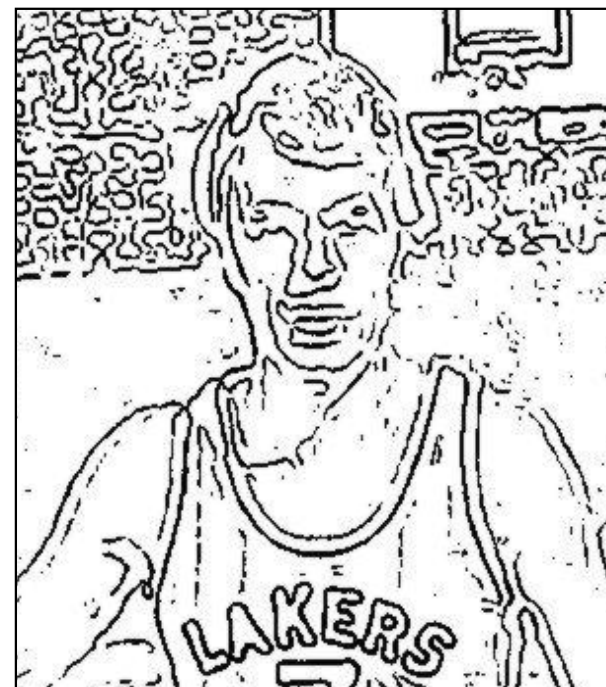
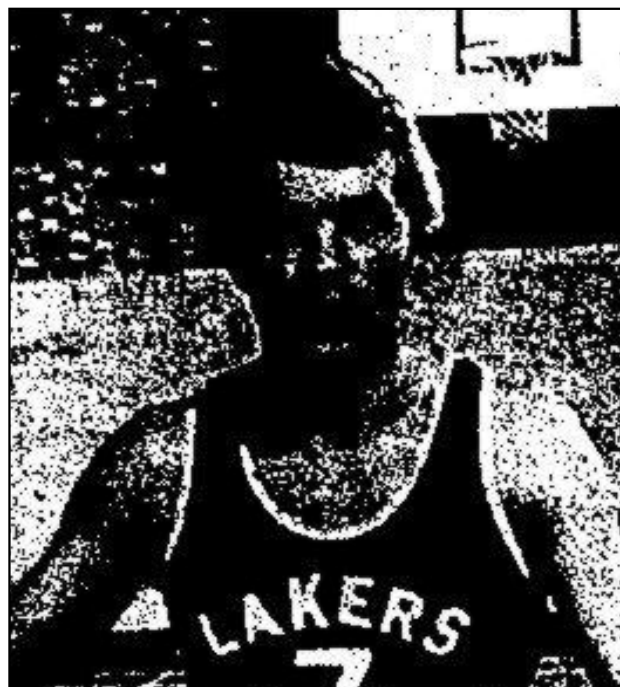
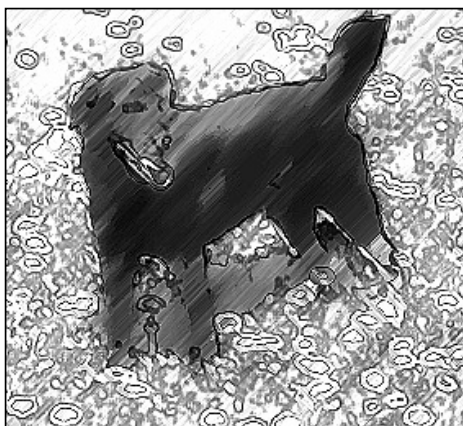
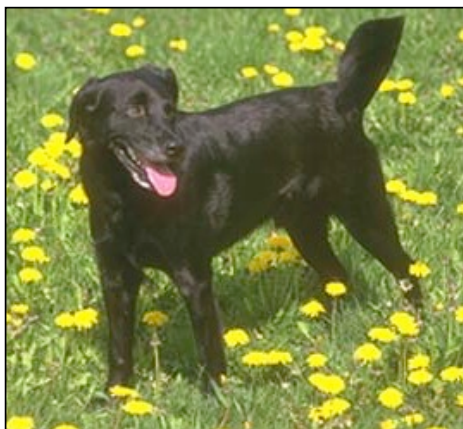


# Sharpening





# Estrazione di contorni



# Il filtraggio della matrice

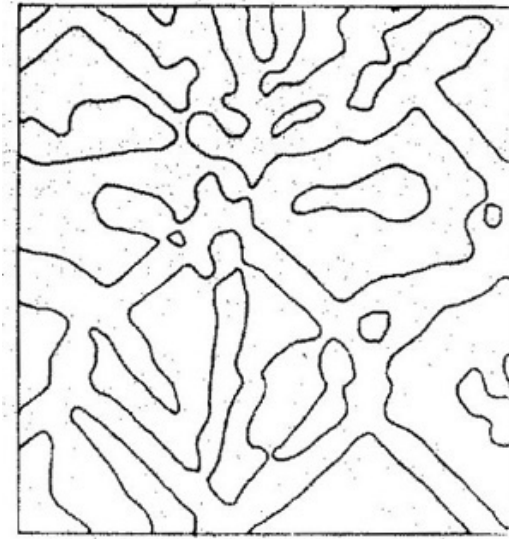
- dimensione ottimale del sombrero messicano
  - sombrero largo: variazioni graduali su area vasta
  - sombrero stretto: molte piccole variazioni in dettaglio
- esseri umani sensibili a un'ampia gamma di variazioni: sombrero flessibile?

# Confronto tra i filtri



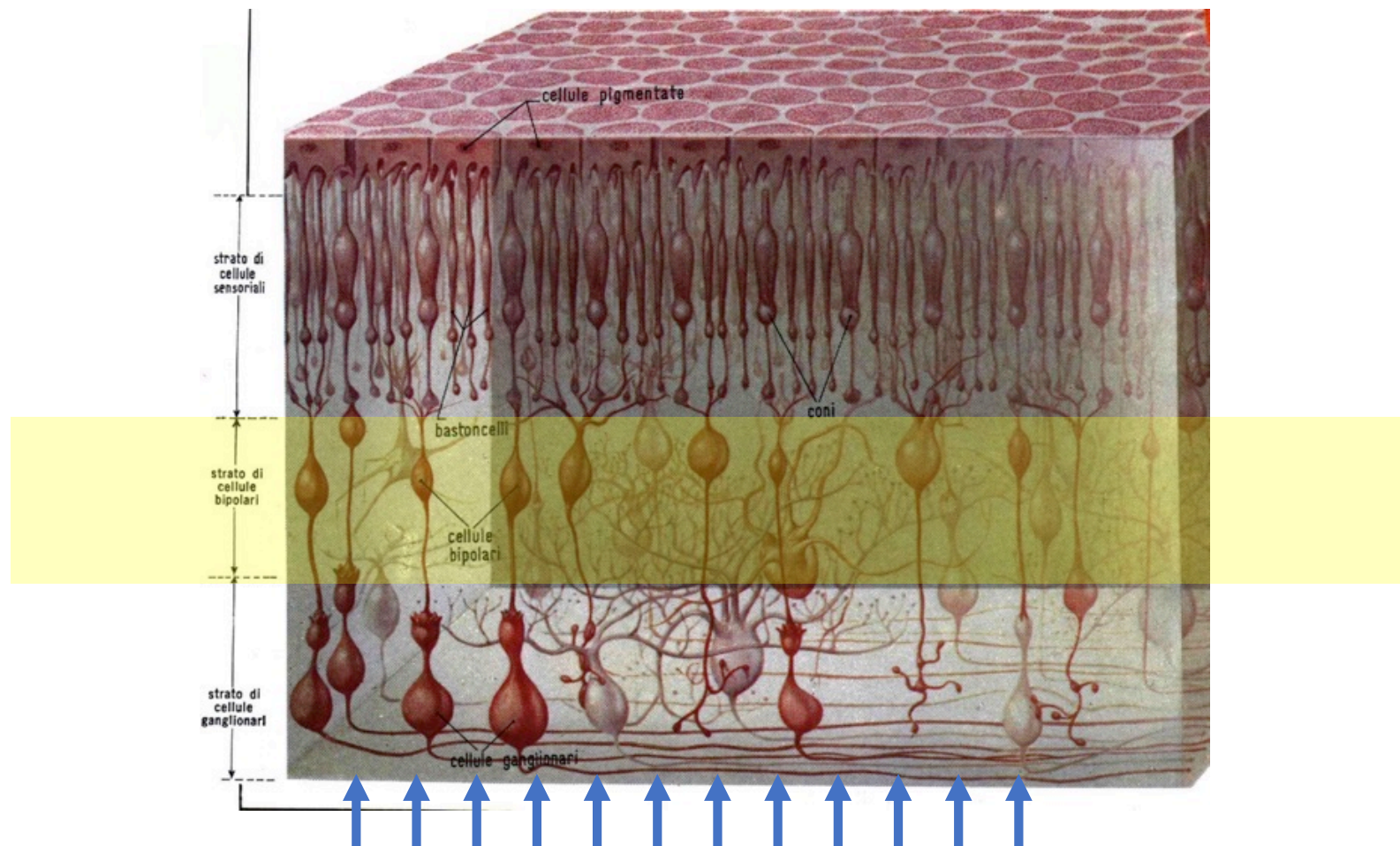
(a) filtro piccolo  
9 pixel

Immagine a livelli  
di grigio (320x320)



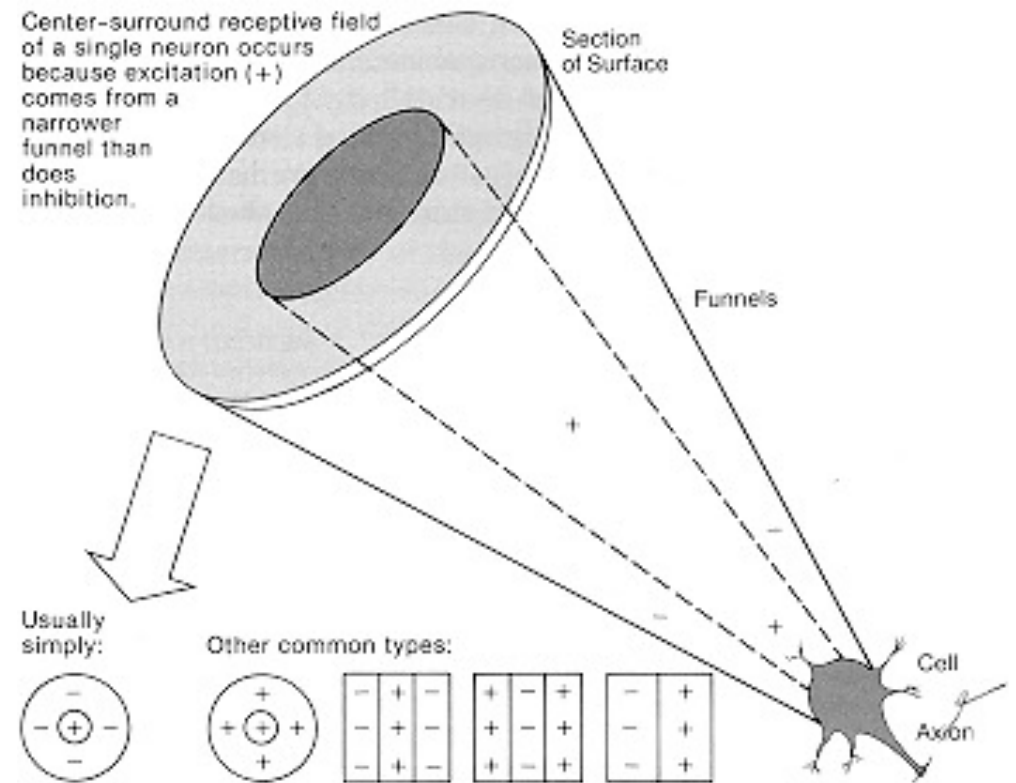
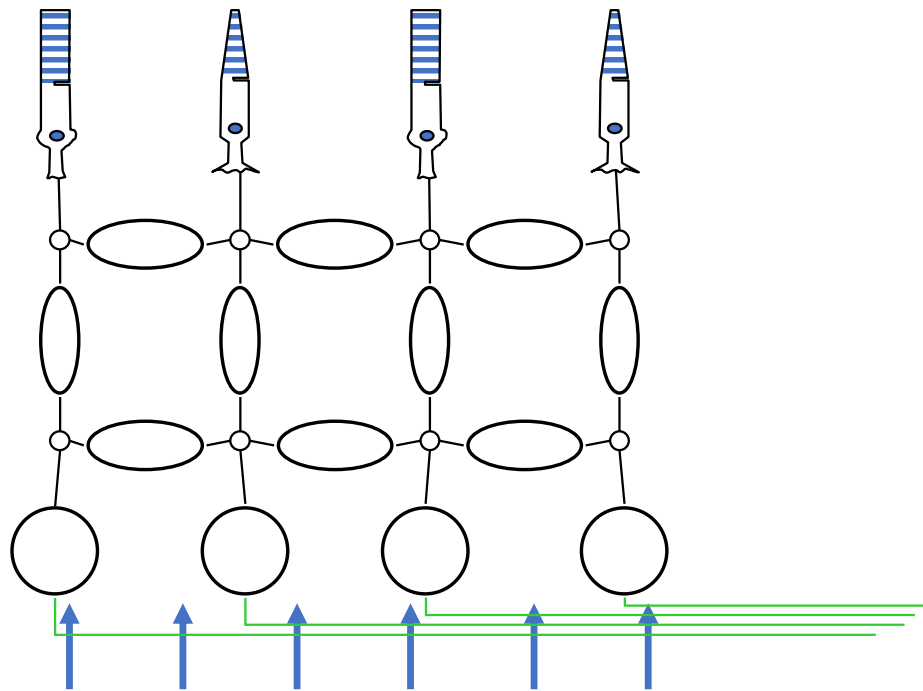
(b) filtro grande  
18 pixel

# Strati retinici





# Il funzionamento della retina

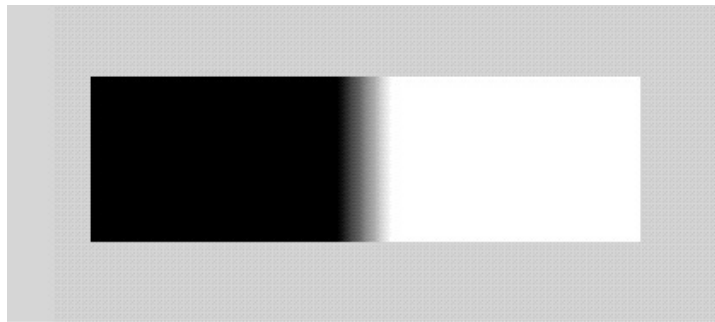


<http://williamcalvin.com/bk4/bk4ch6.htm>

Conferma percezione: effetto Mach

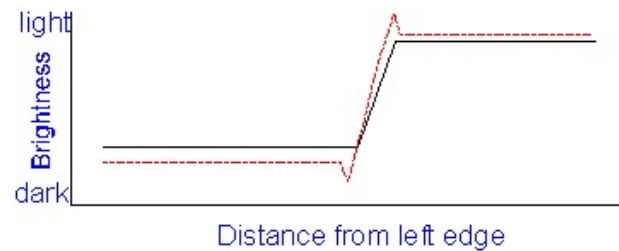


# La risposta del sistema visivo

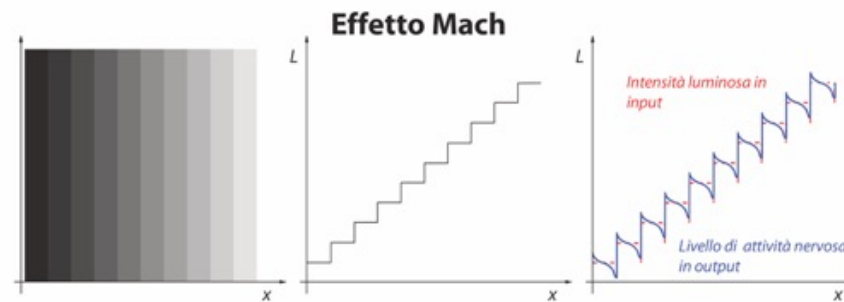
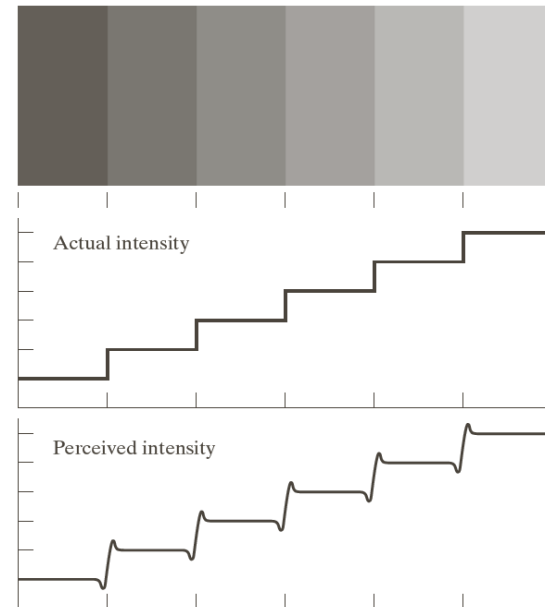


Bande di Mach

Variazione di intensità



Risposta del sistema visivo



# La spiegazione

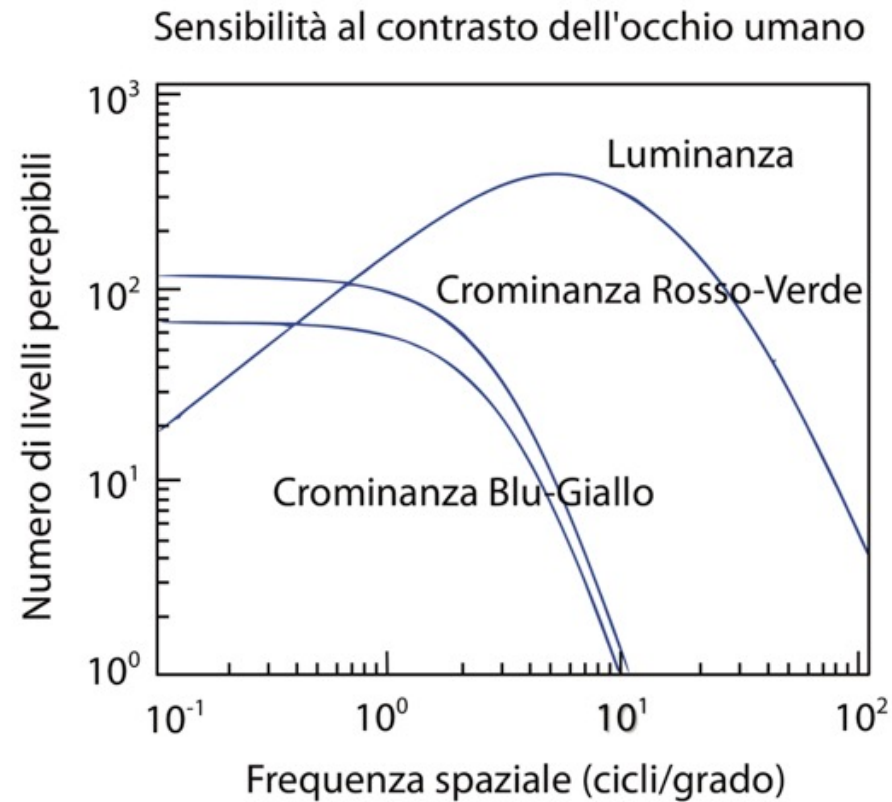
- Vantaggio evolutivo dalla localizzazione dei contorni degli oggetti
- Contorni individuati anche in condizioni minime di variazione tra un oggetto e il contorno
- Il sistema enfatizza la variazione per identificare i contorni



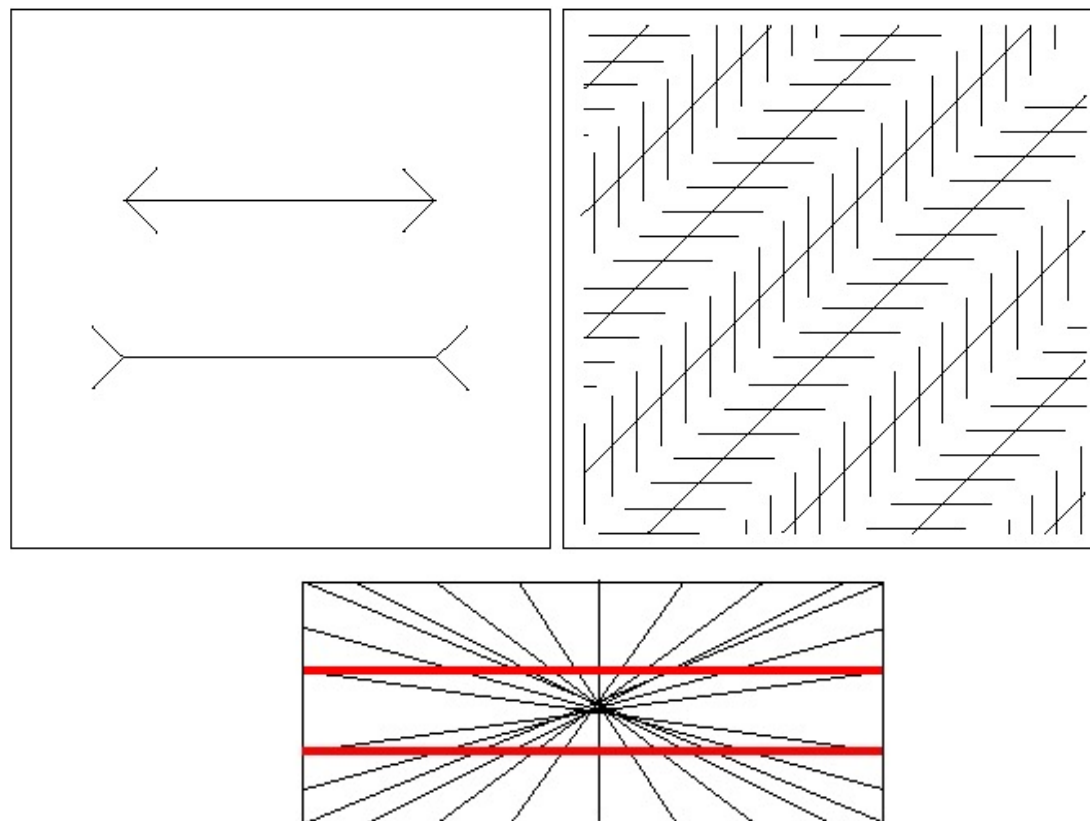
# Sensibilità funzione della frequenza

Minore alle frequenze  
alte e basse

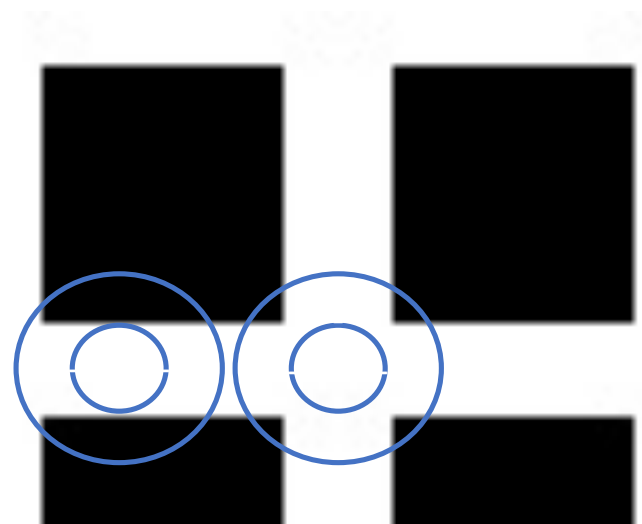
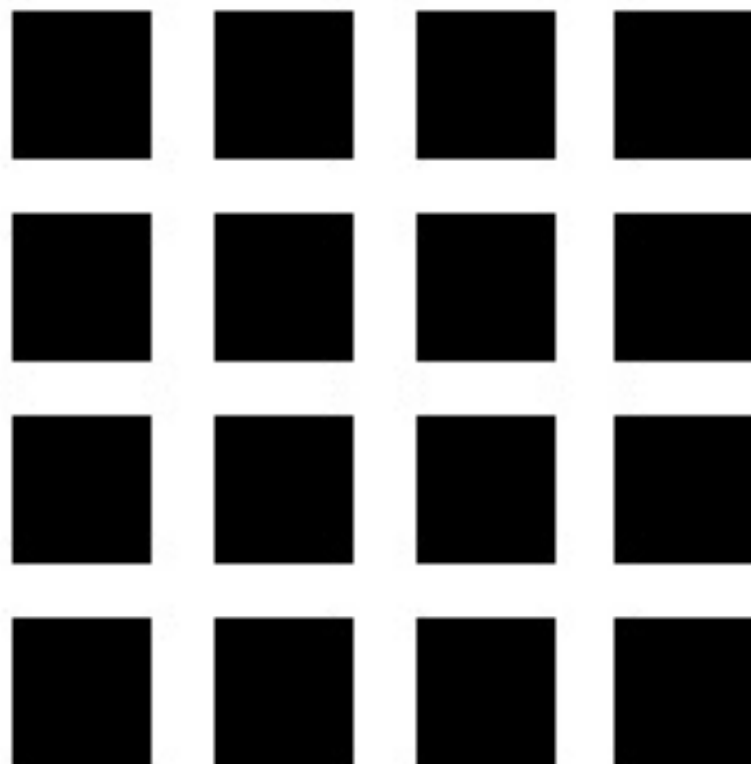
Maggiore alle frequenze  
intermedie



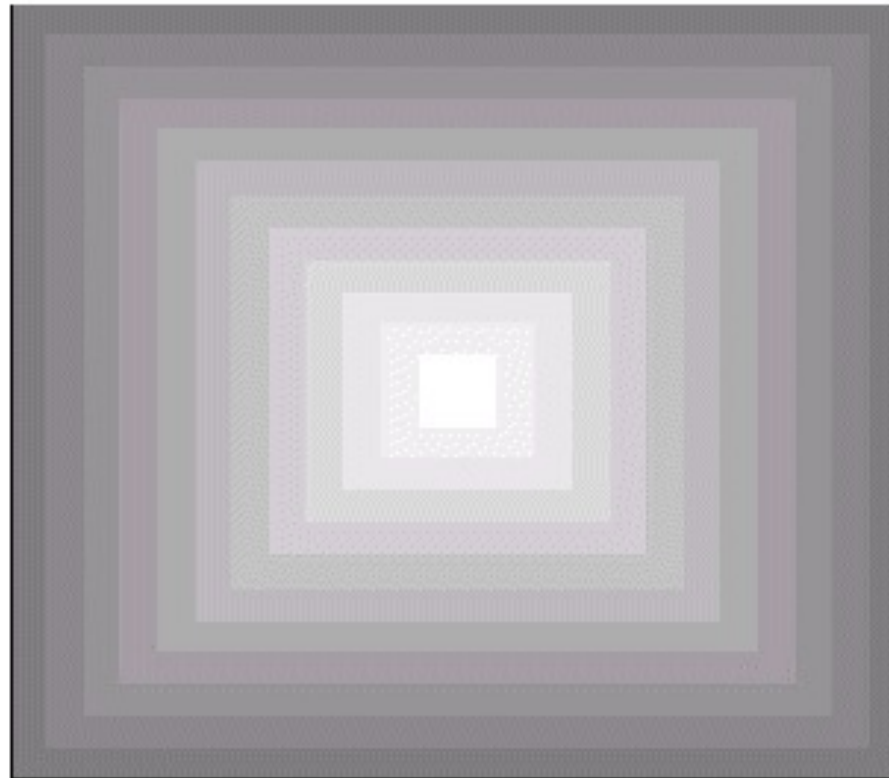
# Lunghezze e parallelismi



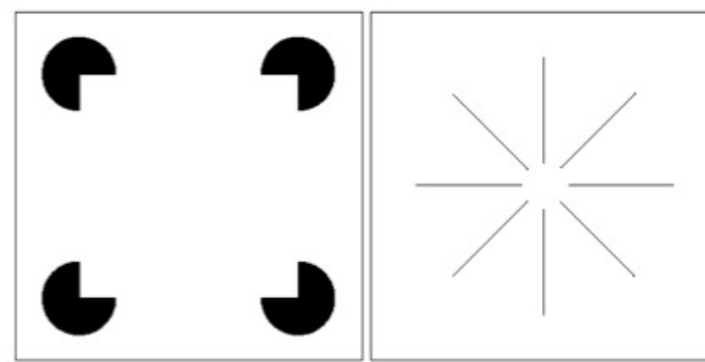
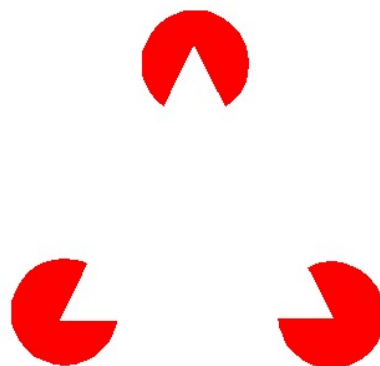
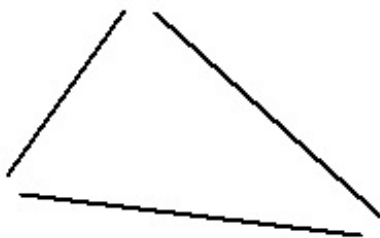
# Griglia di Hermann



Linee radianti



Completamento figure



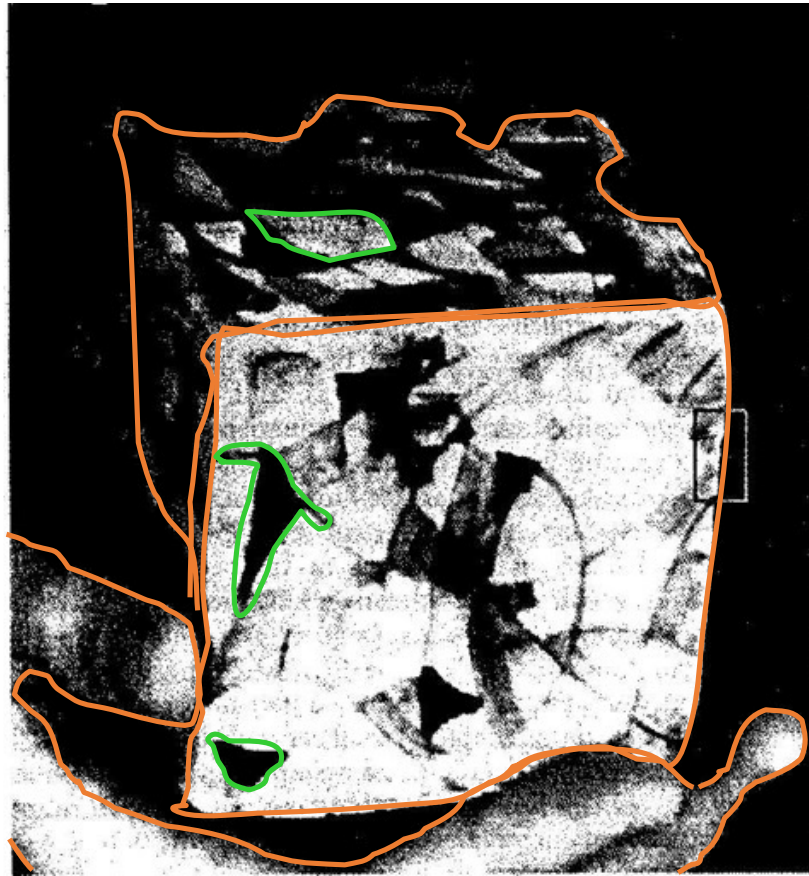
# Primo livello

Campionamento

Localizzazione dei contorni

Schema primario

Lo schema primario:  
barre, spigoli, macchie = regioni

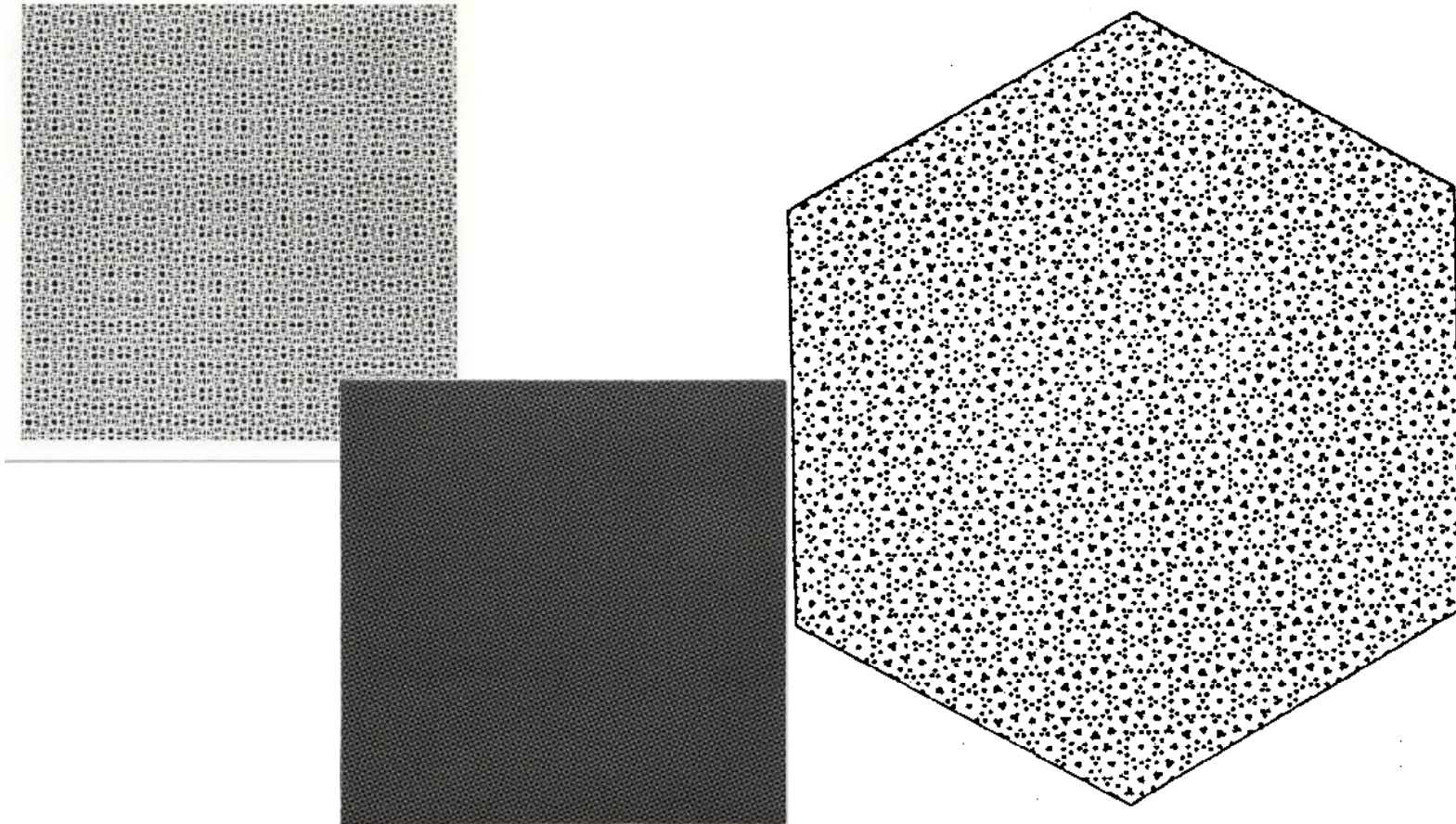


# Schema primario

- organizzazione completa dell'immagine
- guardare fuori fuoco con gli occhi socchiusi
- si costruisce raggruppando elementi simili in modo da formare
  - linee
  - macchie più grandi
  - gruppi strutturati



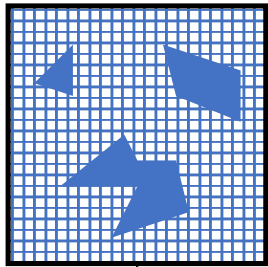
# Più organizzazioni potenziali



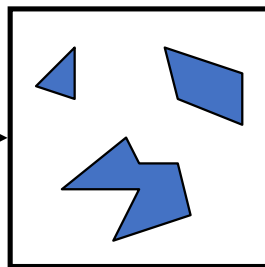
<http://doraballa-ommo.blogspot.com/2009/11/francois-morellet-1971.html>

# Il passaggio da 2D a 3D

matrice di pixel

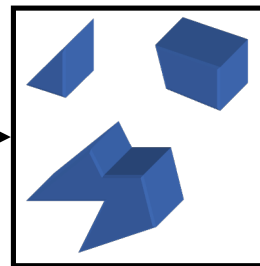


schema primario



I fase

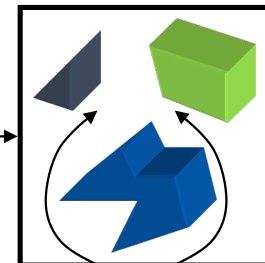
schema a 2D e ½

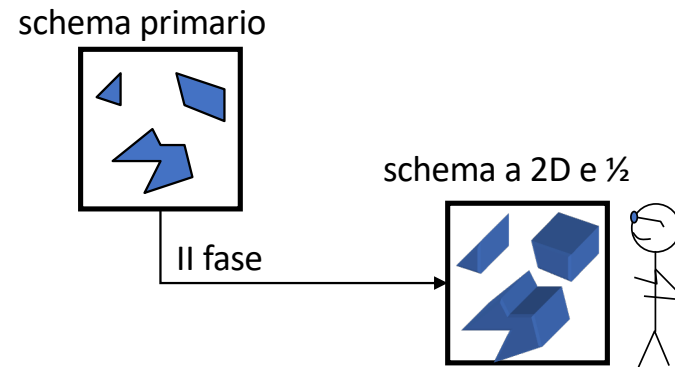


II fase

III fase

scena 3D





# Secondo livello: La rappresentazione intermedia

Da 2D a 2D e 1/2

# Esempio

*Cubo di un  
qualche  
materiale  
(cristallo?)*

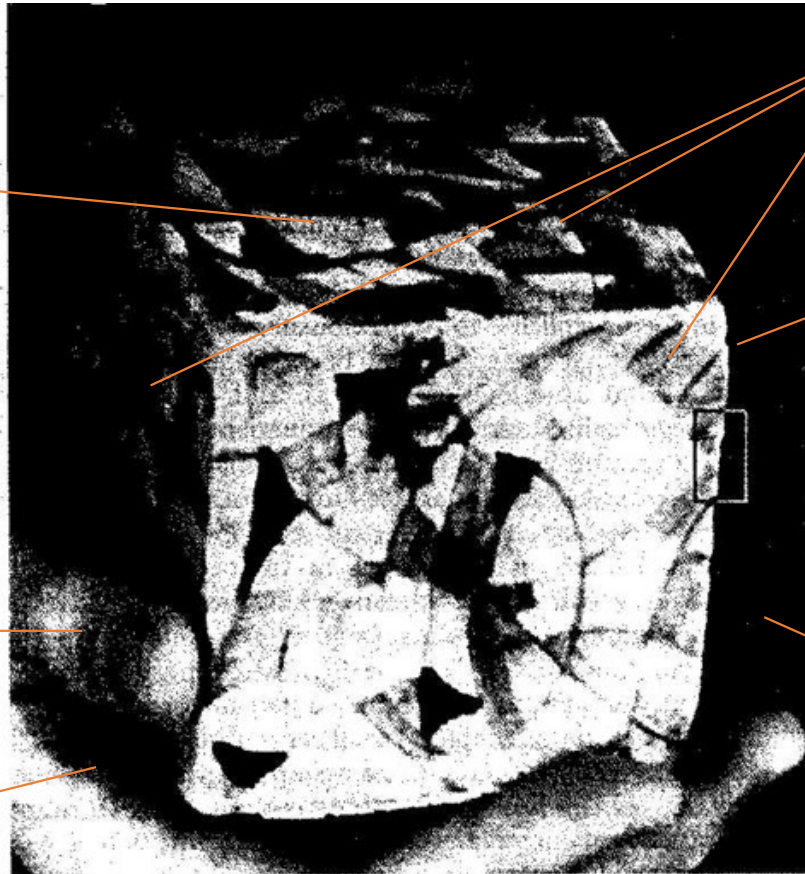
*Superfici  
perpendicolari  
tra loro*

*Superficie  
nascosta*

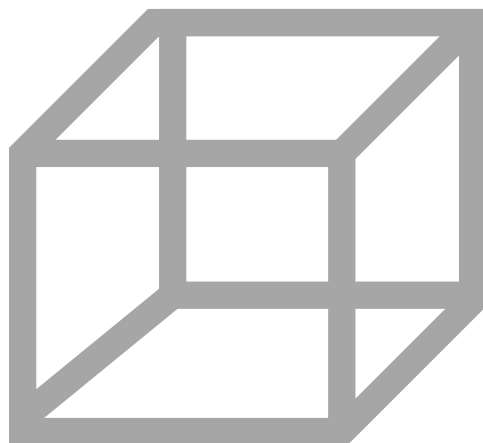
*Pollice*

*La mano  
sos(tiene)  
il cubo*

*Mano (dx)*

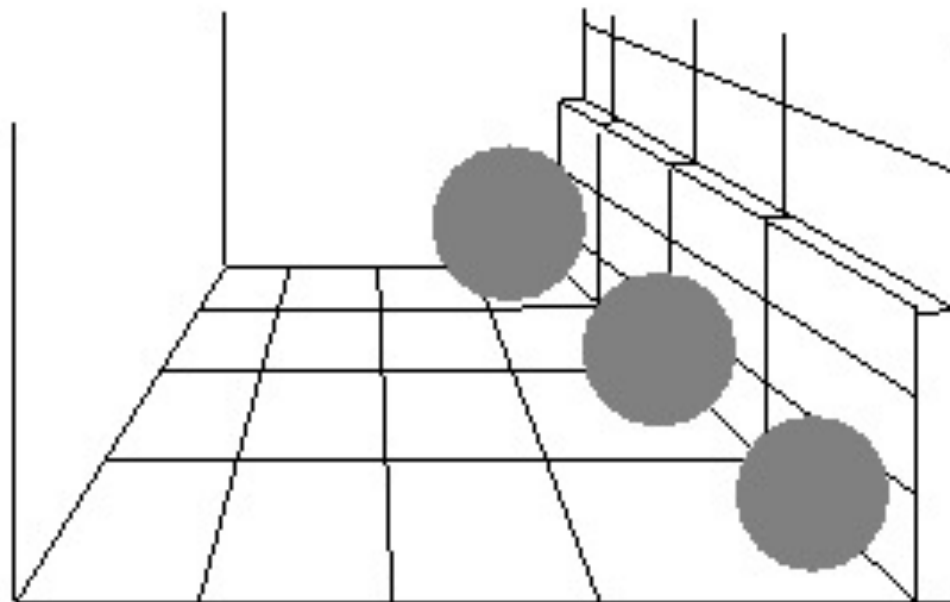


# Illusioni 3D



Il cubo di Necker

# Illusioni 3D



Prospettiva e percezione dimensioni

# Estrazione dell'informazione oltre 2-D

- segmentazione della scena in oggetti distinti
- posizione e orientamento di ogni oggetto in relazione all'osservatore
- forma di ogni oggetto

# Indizi da 2-D a 3-D

## 1. Segmentazione in oggetti

- Regioni corrispondono a entità nella scena semanticamente significative
- Assemblare gli elementi di un singolo oggetto per riconoscimento o presa

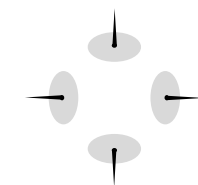
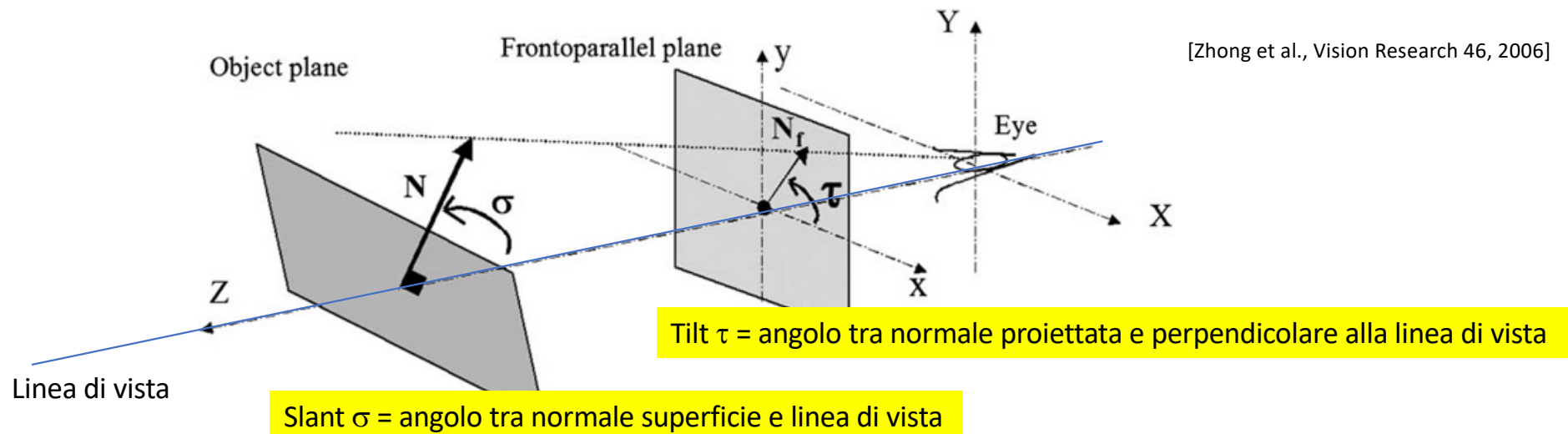


# Indizi da 2-D a 3-D

## 2. Posa/orientamento superfici

- Posa/orientamento rispetto all'osservatore
- Utile per manipolazione e navigazione
- Caso di un piano (descritto da slant e tilt)
  - Slant quantifica la pendenza della superficie relativa alla linea di vista (angolo tra normale e linea di vista)
  - Tilt: direzione della proiezione della normale nel piano fronte-parallelo

## 2. Slant e tilt di una superficie piana

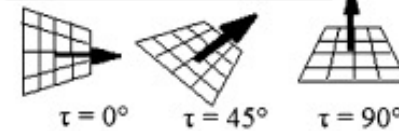


stesso *slant*,  
diverso *tilt*

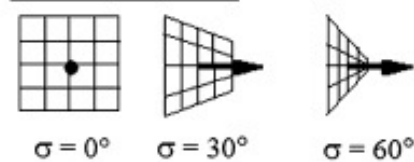


stesso *tilt*,  
diverso *slant*

### A Varving the tilt direction



### B Varving the slant



### 3. Forma

- Invariante rispetto a distanza e orientamento di un oggetto
- Forma di un oggetto e forma di una superficie rispetto al movimento
- Fondamentale per presa e riconoscimento



# Indizi per passare da 2D a 2D e 1/2

- colore
- profondità (stereopsi binoculare)
- tessitura
- movimento
- ombre
- contorni

# Il colore

Un fenomeno fisico e psicologico



# Capacità del sistema visivo umano

- Distingue approssimativamente 350 colori
- Differenze distinguibili sulla TINTA
  - $\lambda$  dominante distante circa 1 nm
  - Agli estremi dello spettro circa 10 nm
  - Circa 128 tinte
- Differenze distinguibili sulla SATURAZIONE
  - Capacità di distinzione più limitate

# Utilità dei colori nella visione

- Distinzione degli oggetti di interesse dallo sfondo
  - Es.: Frutta matura in contrasto con le foglie
- Proprietà degli oggetti
  - Es.: Colori diversi dal verde indicano la frutta matura
- Identificazione degli oggetti
  - Es.: Trovare la frutta matura in mezzo alle foglie

# La profondità

Stereopsi binoculare

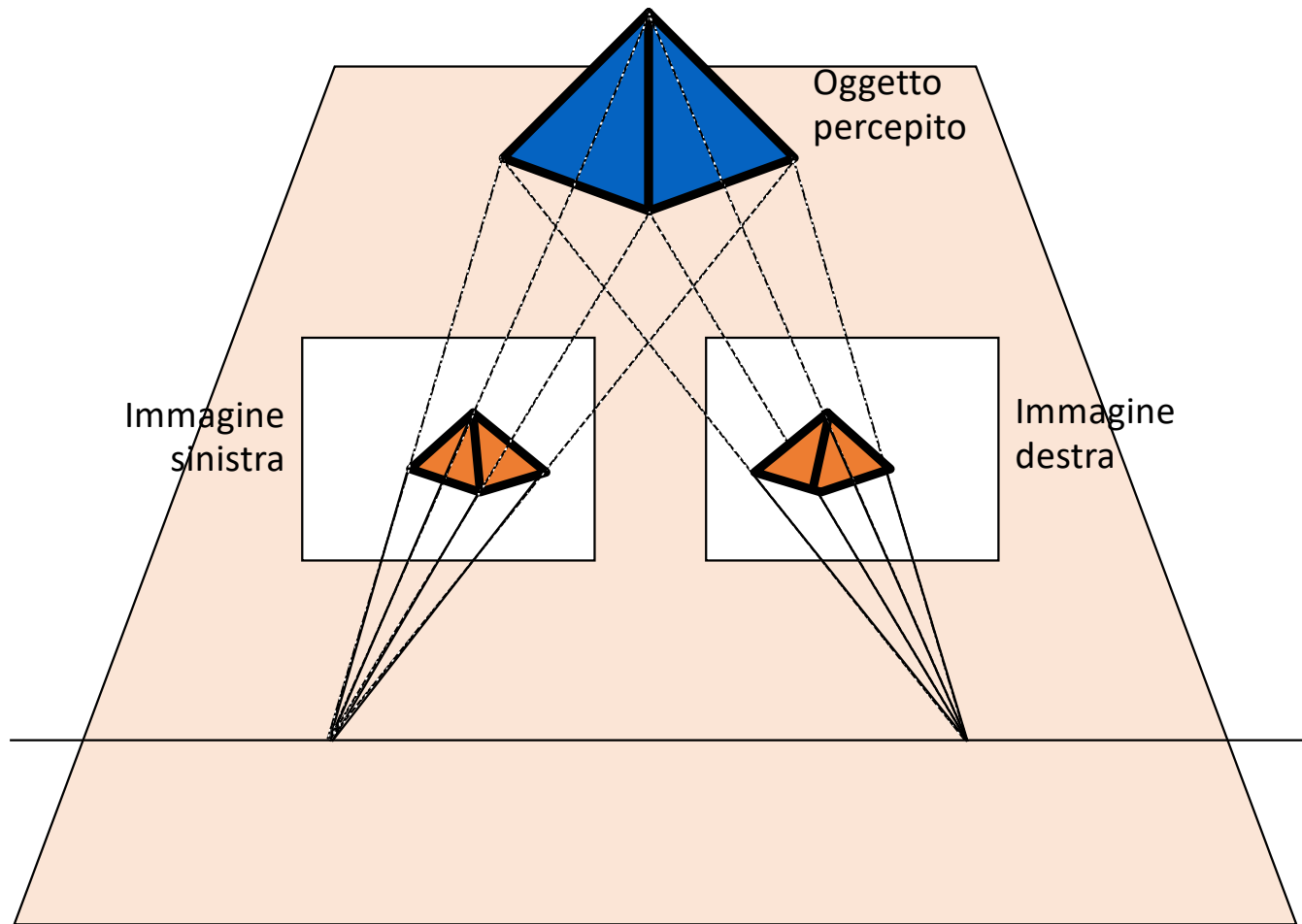


# La profondità

- Alcuni indizi di profondità richiedono un solo occhio (visione monoculare)
- Dipinti, fotografie, televisione
- Due occhi sulla stessa immagine
- Tecnica della prospettiva

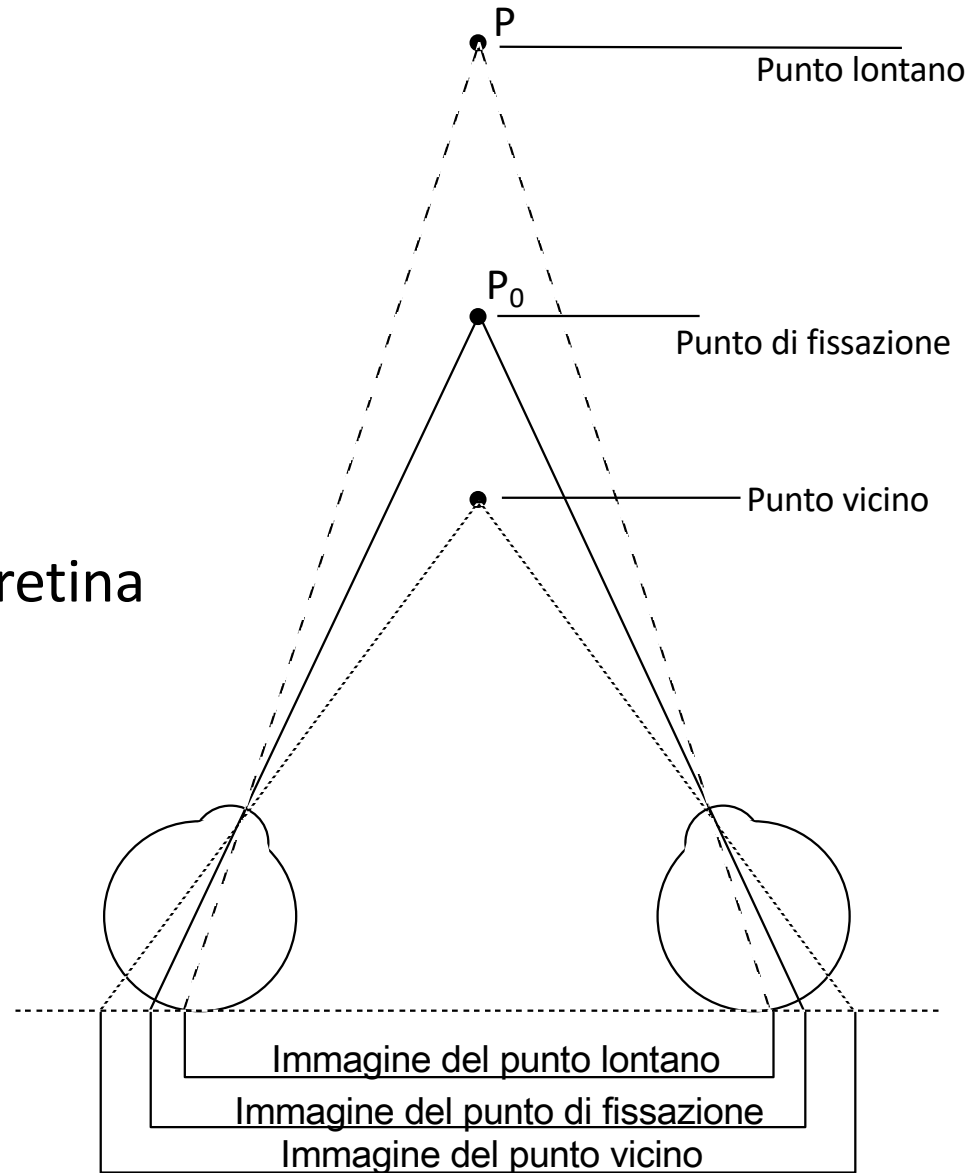


# La stereopsi (visione stereoscopica)



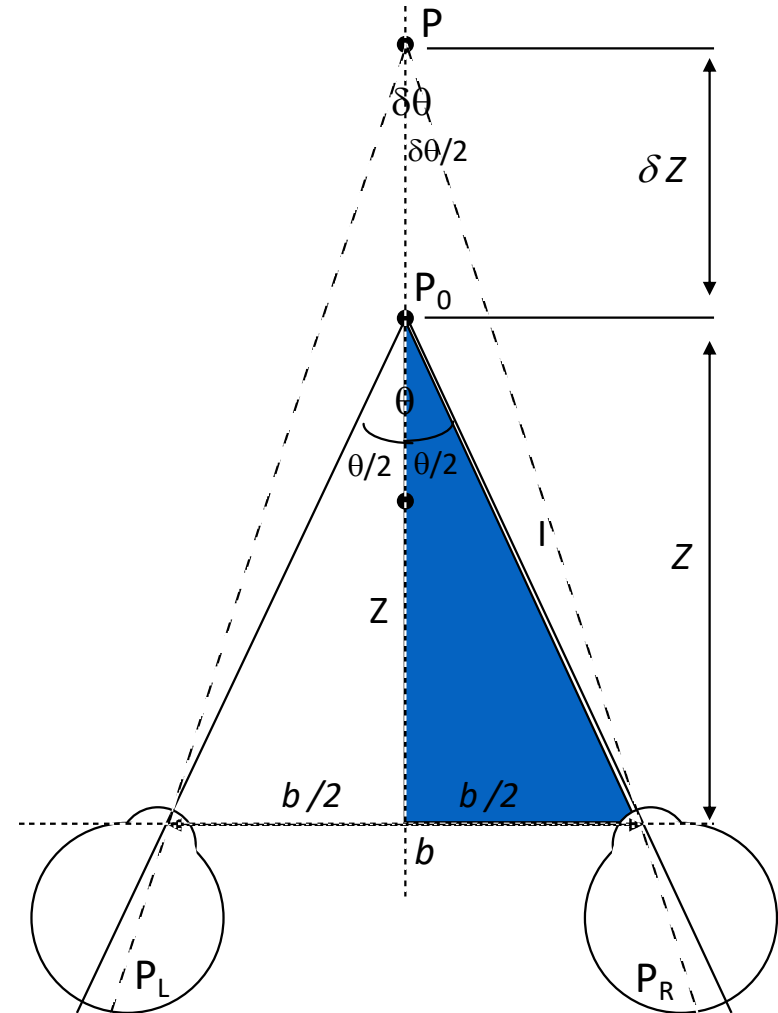
La disparità binoculare

**Il punto di fissazione**  
sempre al centro di ciascuna retina



## Dati numerici

- negli esseri umani,  $b=6$  cm
- Per  $Z=100$ cm
  - Minimo  $\delta\theta$  rilevabile  $0^{\circ},0014$
  - $\delta\theta = 2.42 \times 10^{-5}$  rad
  - $\delta Z$  è circa 0.4 mm
- Per  $Z=30$  cm
  - si arriva a un  $\delta Z$  di circa 0.036 mm



# Problema della corrispondenza

Come si fa a mettere in corrispondenza le proiezioni che nelle due immagini visive corrispondono a uno stesso punto della scena?

NB: le stesse posizioni sulle due retine (o immagini della camera) non corrispondono allo stesso punto sulla scena

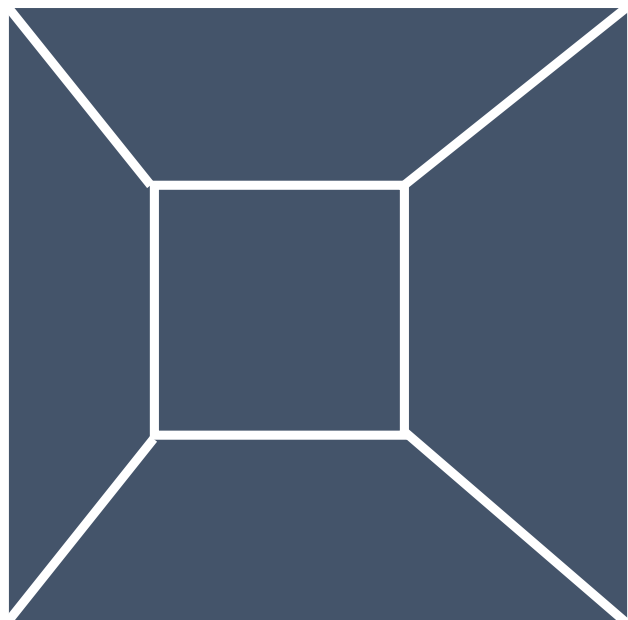
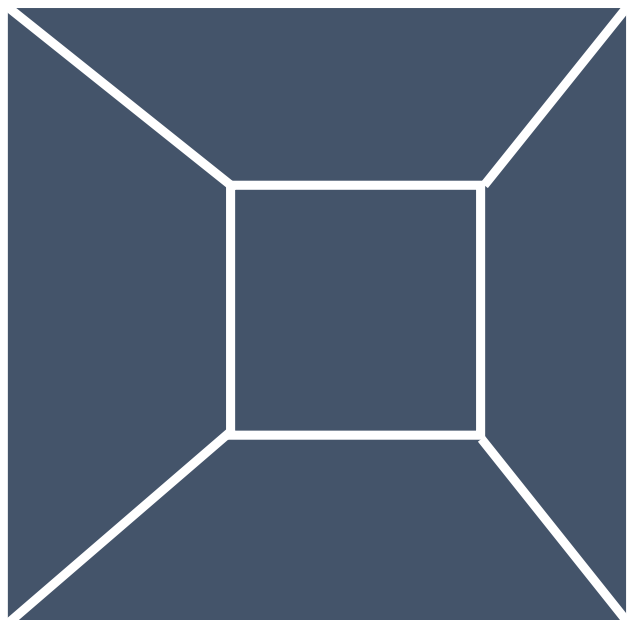
# Due ipotesi

- top-down: problema circolo vizioso
  - struttura degli oggetti predice disparità retiniche
  - si identificano gli oggetti nella scena e si fanno corrispondere le medesime parti
- bottom up: problema intensità differenti
  - valori intensità per predire le strutture a livello più alto
  - si fanno corrispondere valori di intensità delle due matrici

# Prove empiriche

- bottom-up: il sistema funziona senza conoscenze di alto livello
- top-down: il sistema funziona quando i dati sono degradati

# Stereogrammi con figure geometriche ('800)





# Stereogrammi a punti casuali di Julesz

(anni '60)

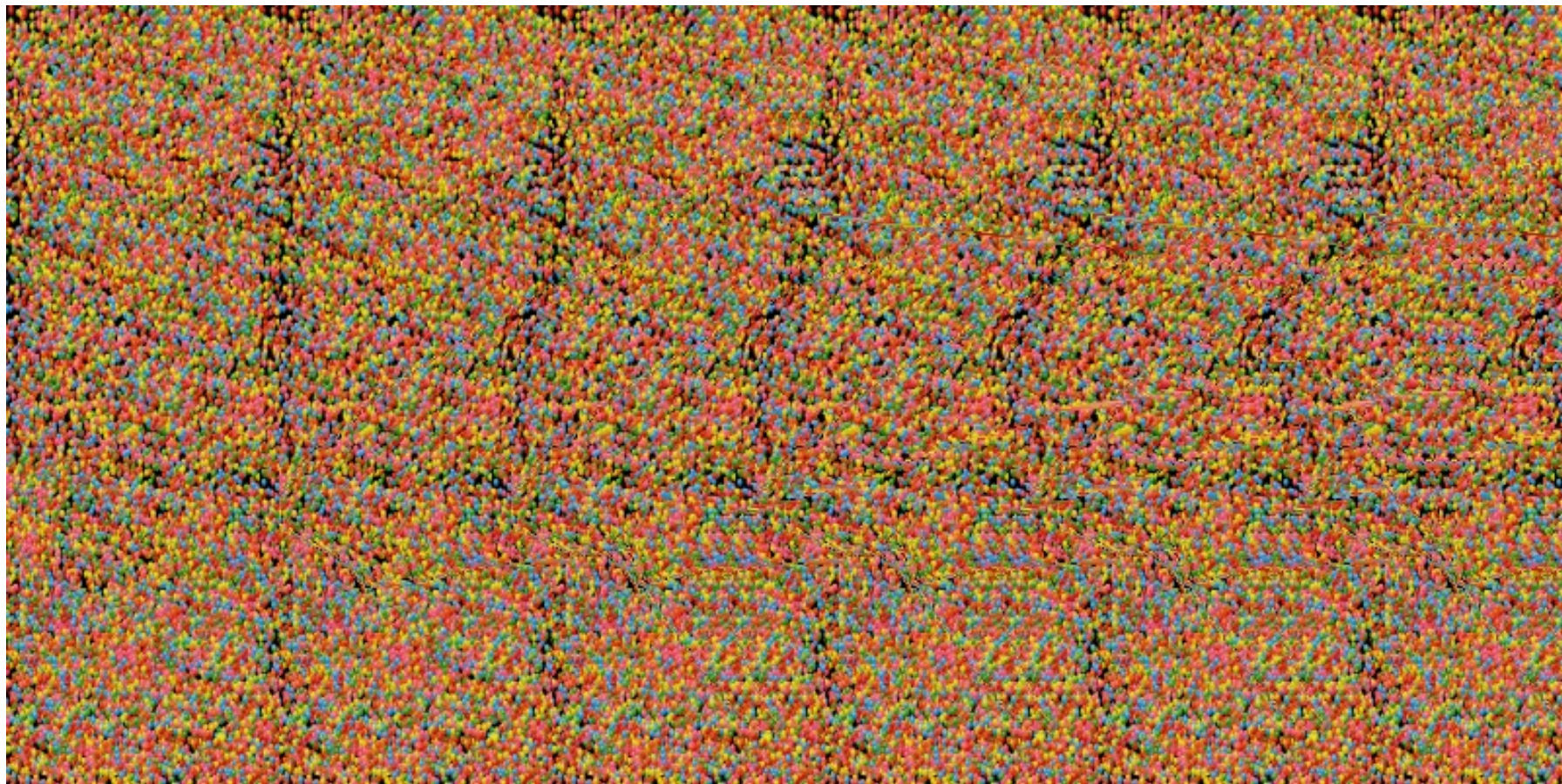


# Autostereogrammi (1979) (geometrico)





# Autostereogramma (zoologico)



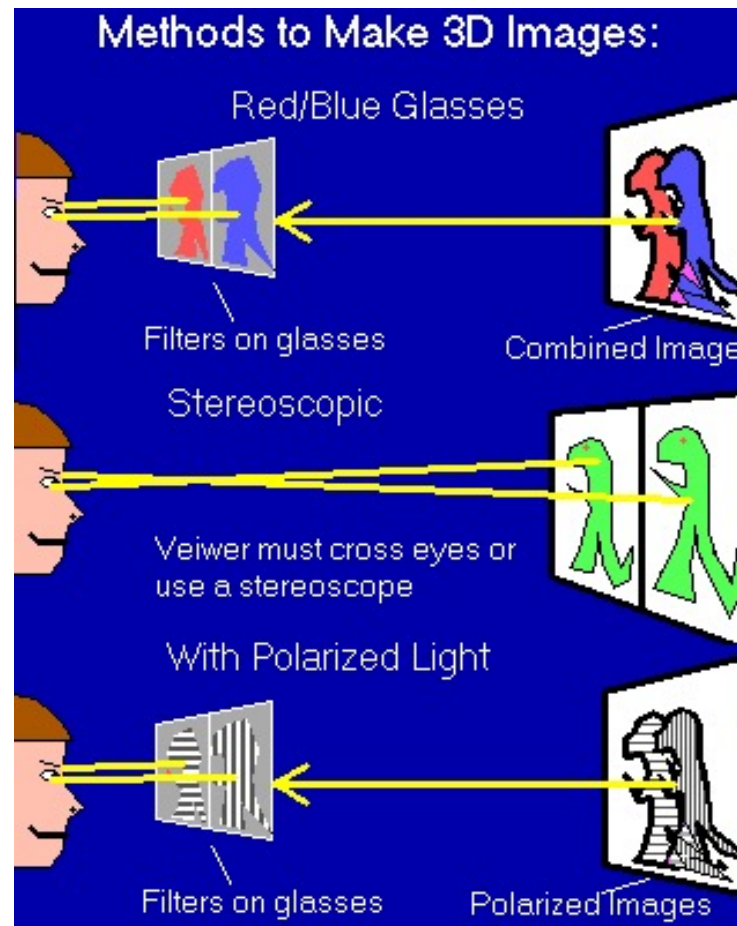
# Implicazioni di Julesz

- La stereopsi può funzionare bottom-up
  - non si percepisce il contenuto senza la stereopsi
  - le conoscenze di alto livello non aiutano la stereopsi
- Frisby, Clatworthy: dire alle persone cosa si dovrebbe vedere non facilita la percezione

# La proiezione stereoscopica



# Immagini stereoscopiche



# Conclusioni sulla profondità

Unici elementi certi della visione stereoscopica

- utilizza elementi di basso livello dello schema primario
- è guidata da vincoli innati propri del mondo fisico

La tessitura (texture)



# La tessitura

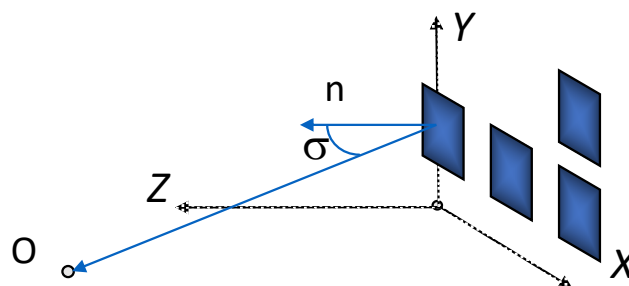
- proprietà delle superfici associate con qualità tattili che essi suggeriscono
- *pattern* ripetuto spazialmente su una superficie, periodici reali, o in senso statistico
- *gradiente di tessitura* creato dalla presenza di oggetti o segni più o meno della stessa forma o dimensione a intervalli regolari su una superficie

# Esempio: pattern periodici

- finestre e balconi su parete edificio
- macchie sulla pelle di un leopardo
- fili d'erba su un prato
- gente allo stadio



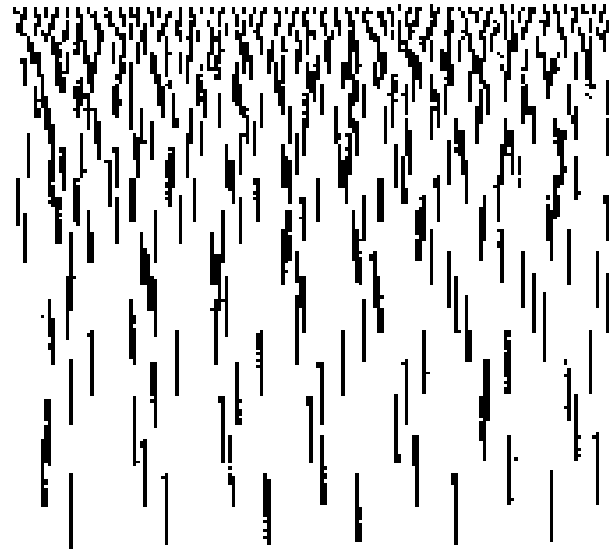
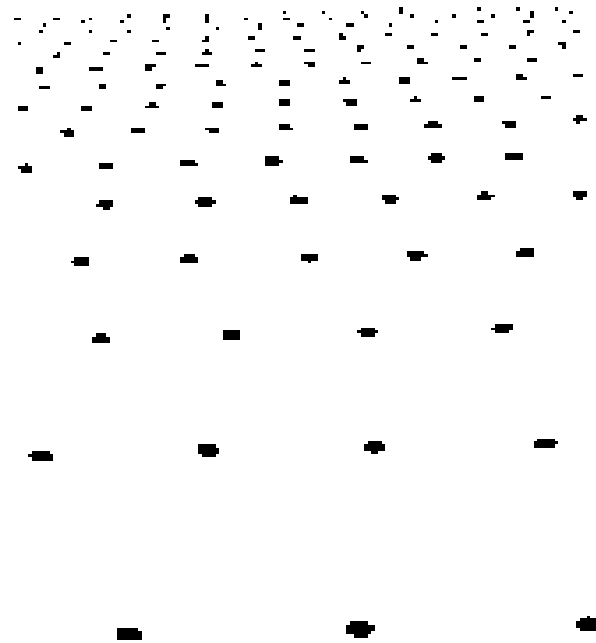
# Variazione dei tasselli



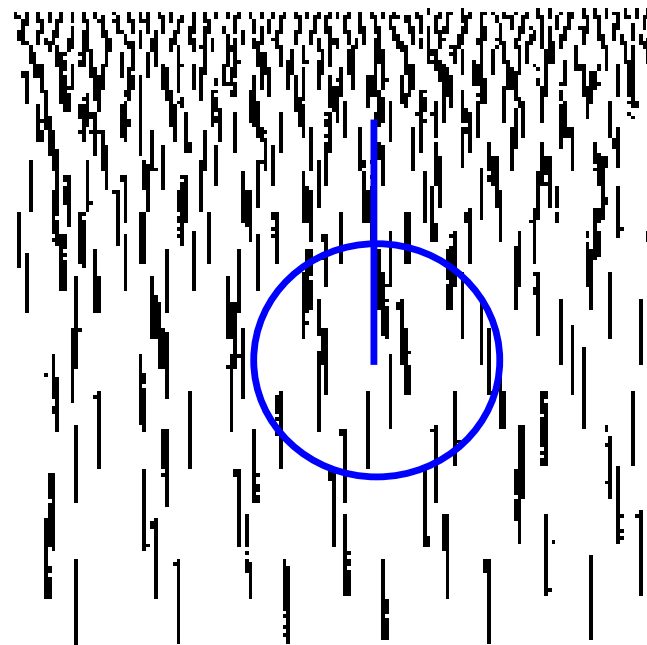
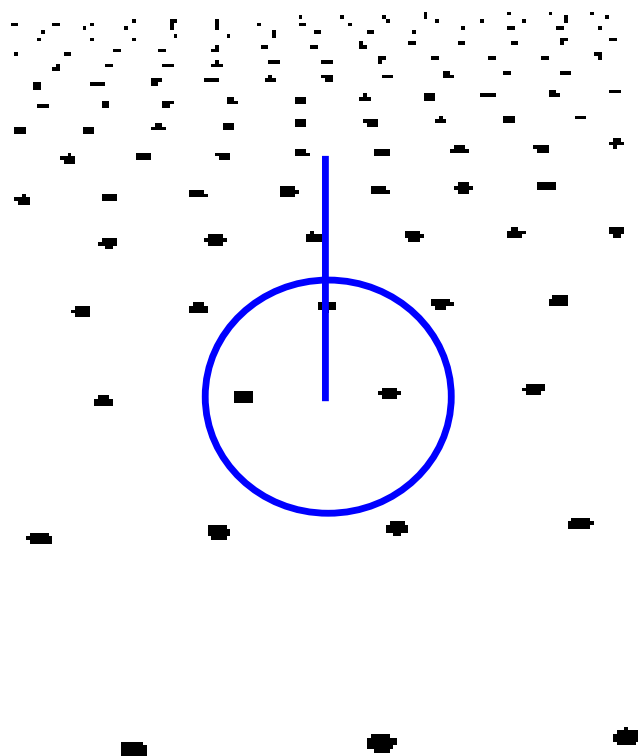
# Tessitura e superfici

- Da pattern regolari è possibile scoprire **orientamento** e **forma** delle superfici
- Elementi fondamentali: *tasselli*
  - **Dimensione, forma e spaziatura** “uniformi” nella scena 3-D
  - Invece **dimensione, forma e spaziatura** variano nell’immagine (una volta proiettati)

# Esempi

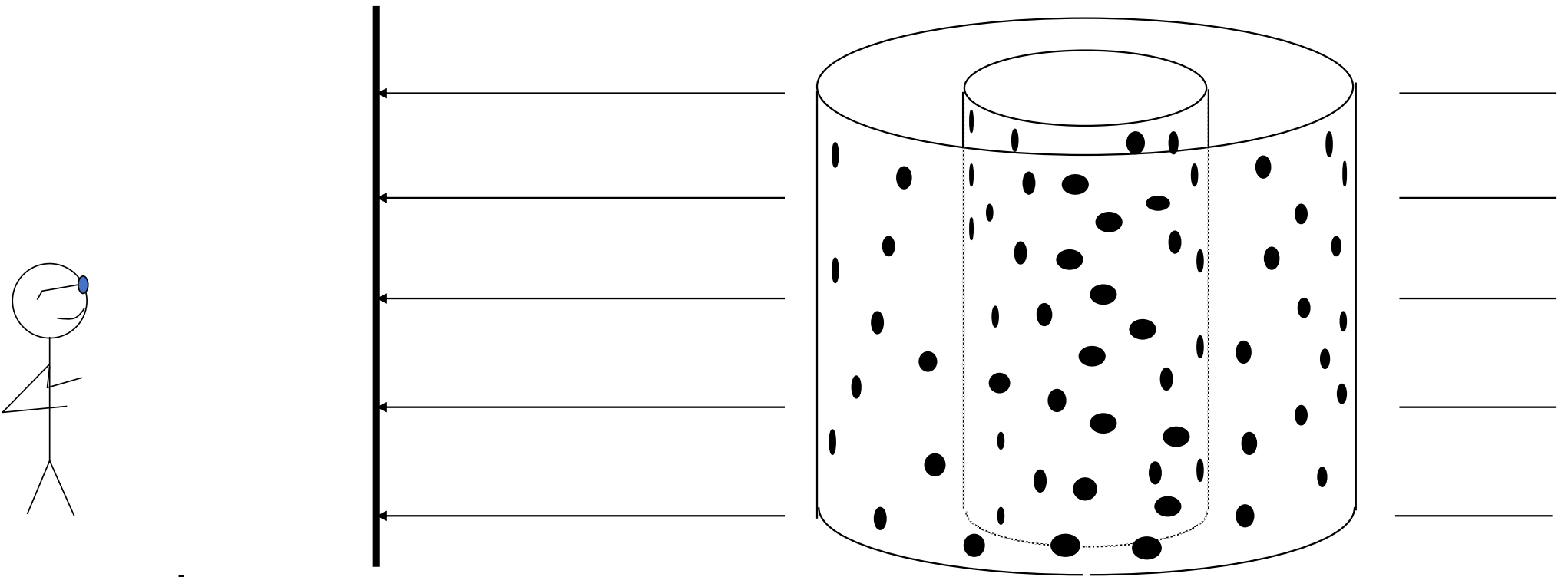


# Gradienti di tessitura



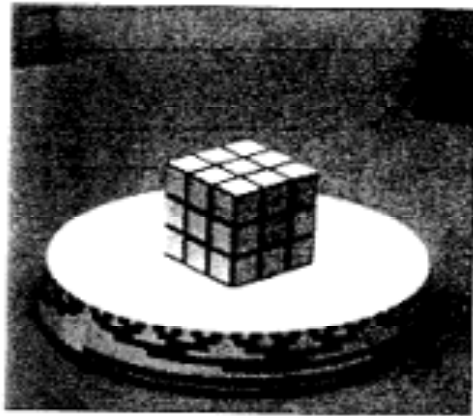
Il movimento

# Movimento e forma: dimostrazione di Ullman

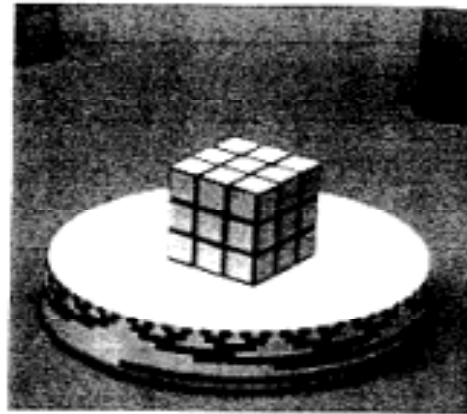




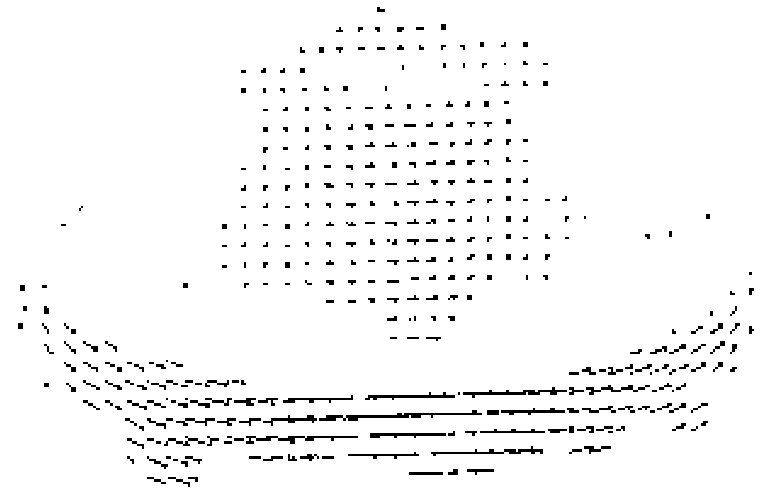
# Stereopsi e movimento



(a) un cubo di  
Rubik rotante



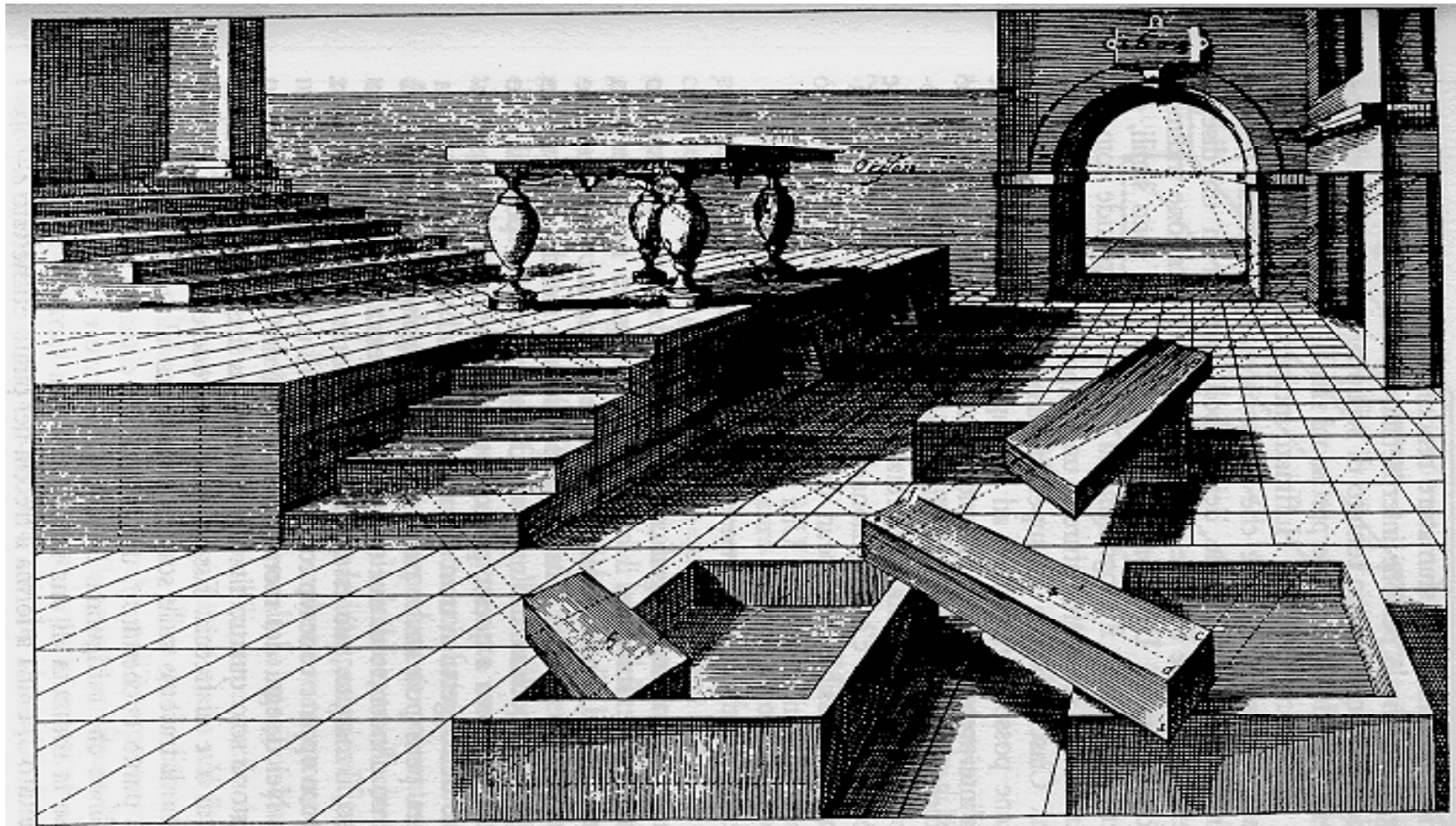
(b) 1/2 secondo  
più tardi



# La prospettiva e i contorni

Indizi dati dalle distorsioni della forma  
dovute alla proiezione

# Disegni senza sfumature e con ombre

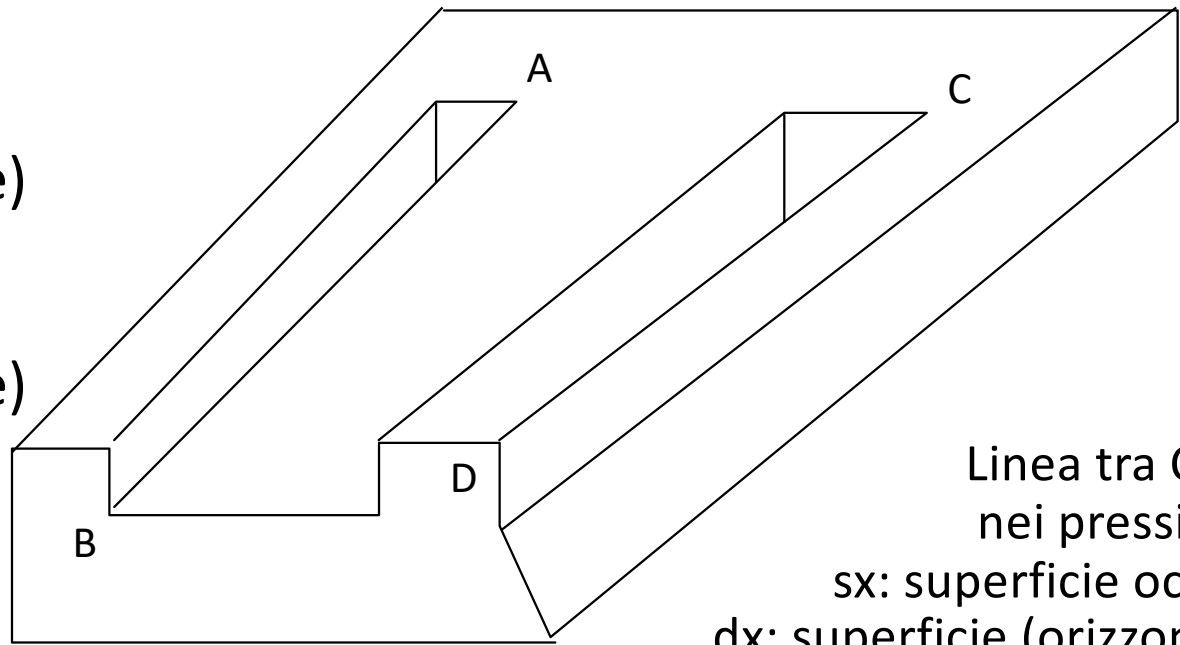


4 tipi di giunzione:  
L, T, Y, ↑

# Oggetti impossibili

Linea tra A e B.

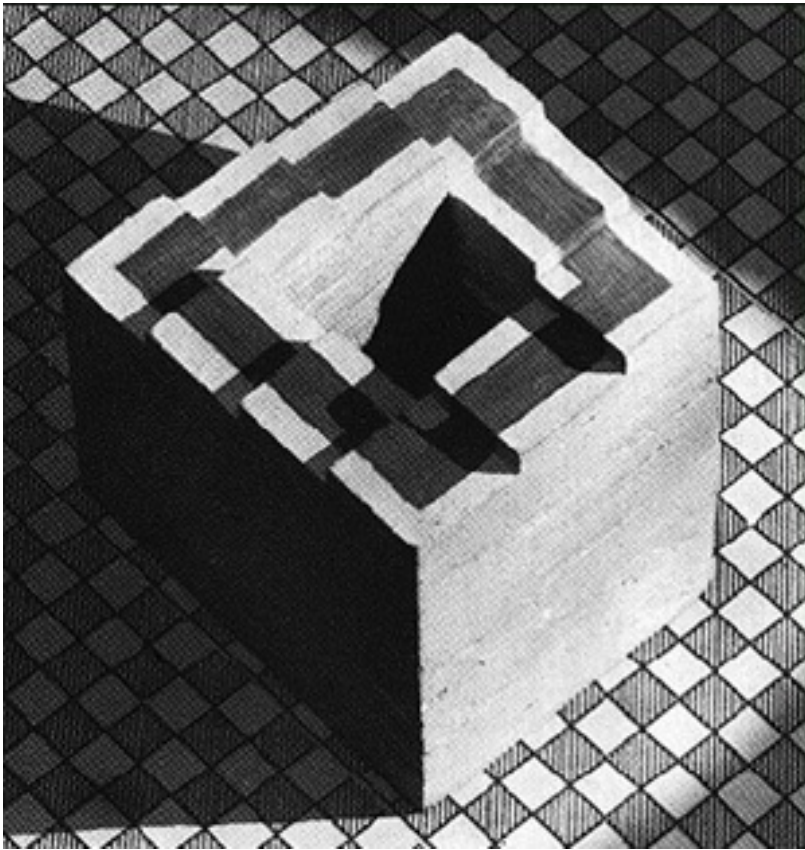
- nei pressi di A:  
sx: superficie occlusa  
dx: superficie (orizzontale)
- nei pressi di B:  
sx: superficie (verticale)  
dx: superficie (orizzontale)



Linea tra C e D.

- nei pressi di C:  
sx: superficie occlusa  
dx: superficie (orizzontale)
- nei pressi di B:  
sx: superficie (orizzontale)  
dx: superficie (verticale)

## Altri oggetti impossibili

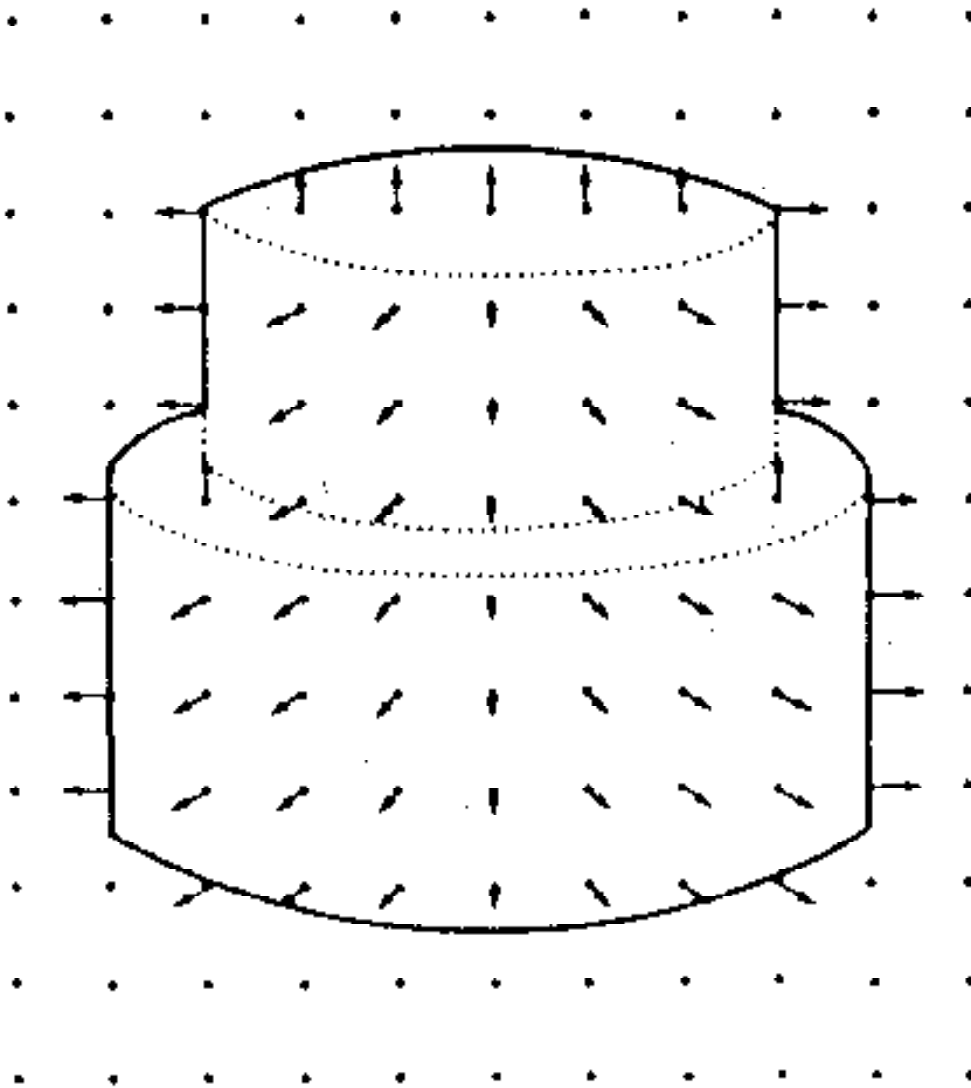


# Lo schema a 2-D e 1/2

Dove arriva la percezione pura

# Sketch a 2-D e 1/2

puntaspilli:  
ciascuno  
spillo  
rappresenta  
profondità e  
orientamento  
di una  
regione



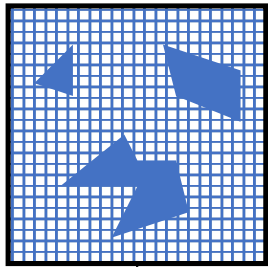
# Conclusioni sulla percezione pura

- Lo schema a 2-D e 1/2 non sufficiente a far muovere un robot nel mondo
- Rappresentazione di una scena:
  - deve rendere possibile l'identificazione degli oggetti
  - deve essere indipendente dal punto di vista particolare

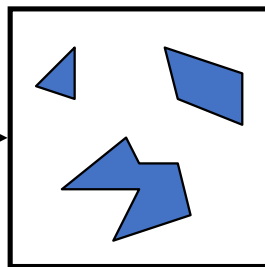


# Il passaggio da 2D a 3D

matrice di pixel

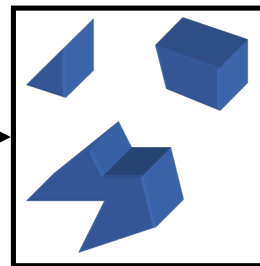


schema primario



I fase

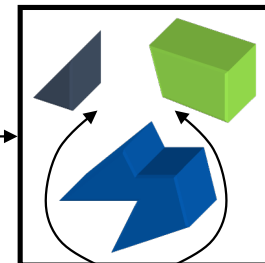
schema a 2D e ½

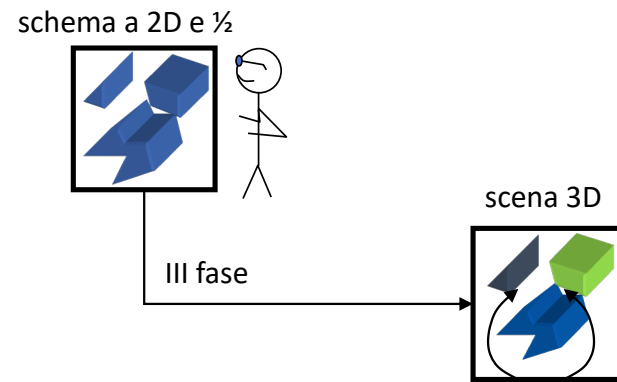


II fase

III fase

scena 3D





# Terzo livello

La struttura 3D indipendente  
dal punto di vista dell'osservatore

# Esempio

*Cubo di un  
qualche  
materiale  
(cristallo?)*

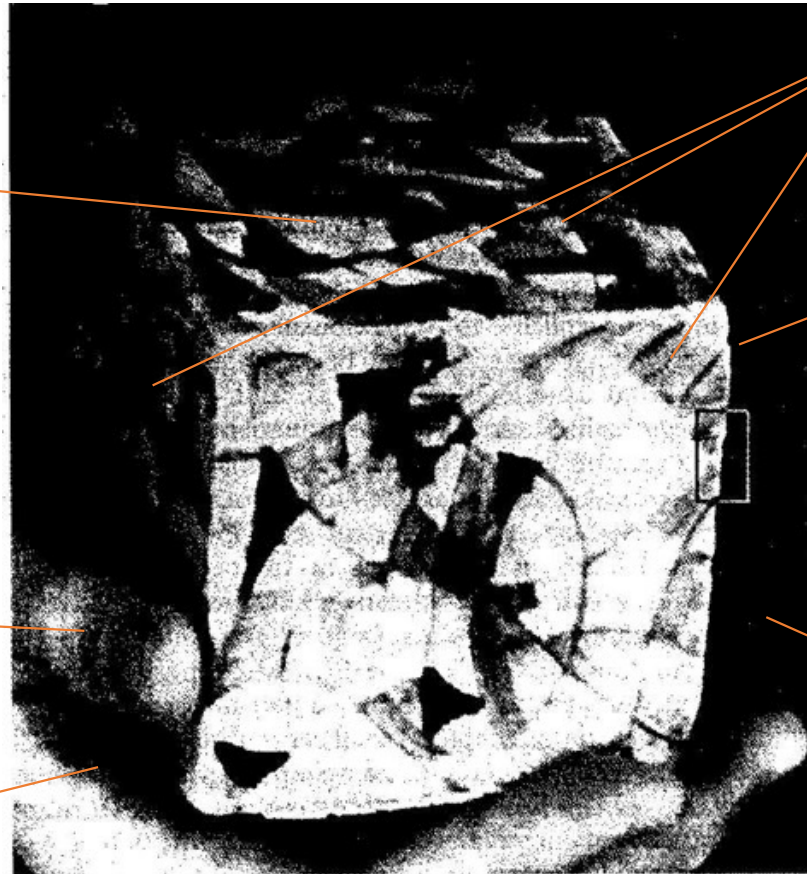
*Superfici  
perpendicolari  
tra loro*

*Superficie  
nascosta*

*Pollice*

*La mano  
sos(tiene)  
il cubo*

*Mano (dx)*



# Processi tipici di alto livello

- Riconoscimento di oggetti e volti
- Percezione di scene e effetti contestuali
- Effetti delle intenzioni e delle conoscenze sugli oggetti sulla percezione
- Strutture mentali in grado di integrare viste successive di oggetti e scene

# Input al terzo livello

- rappresentazione simbolica profondità e orientamento delle superfici nel campo visivo
- rappresentazione aggiornata in movimento
- rappresentazione più utile e stabile se rende esplicite forma 3D intrinseca degli oggetti + relazioni spaziali

# Output del terzo livello

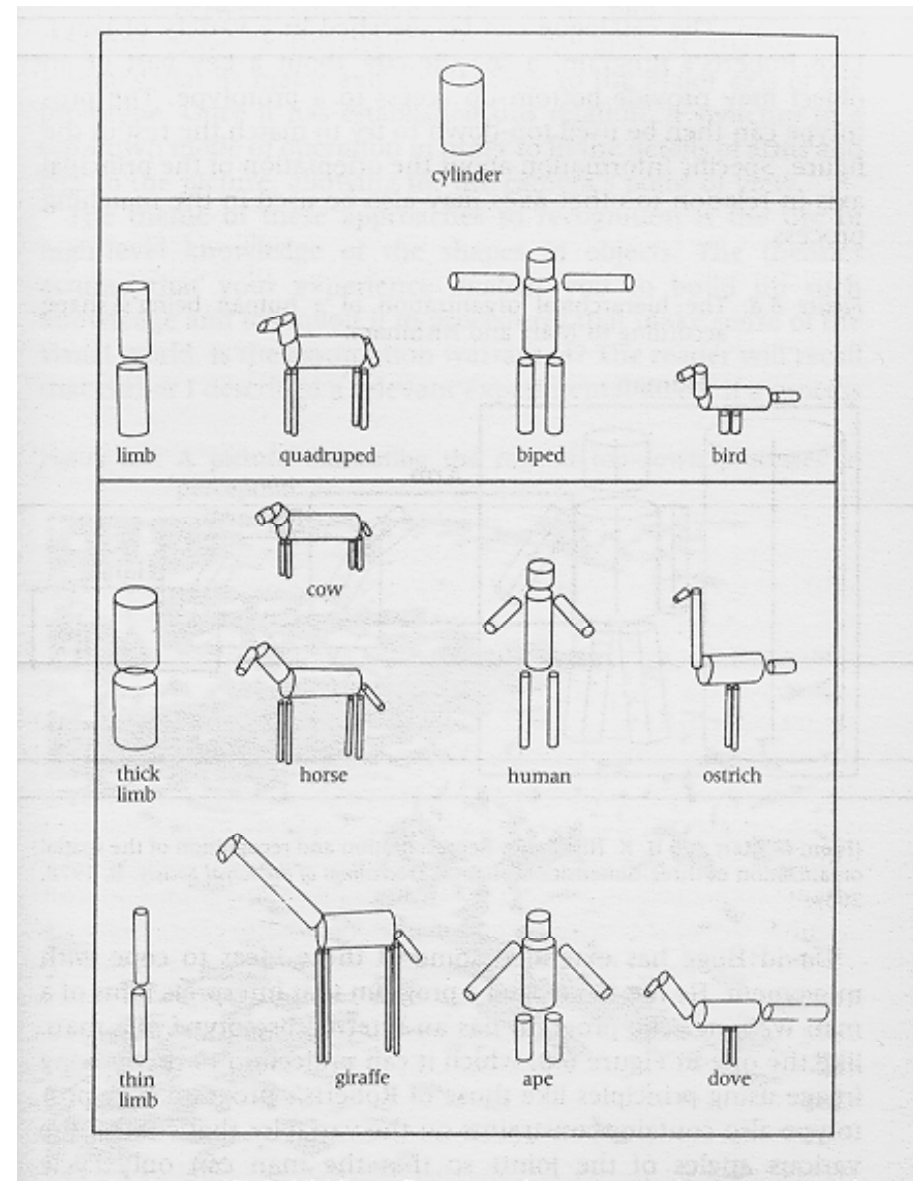
- forma esplicita di oggetti e spazi vuoti nella scena
- elementi riconosciuti, raggiungibili e manipolabili, specificando le 3 coordinate  $x,y,z$

# Catalogo di Marr e Nishihara

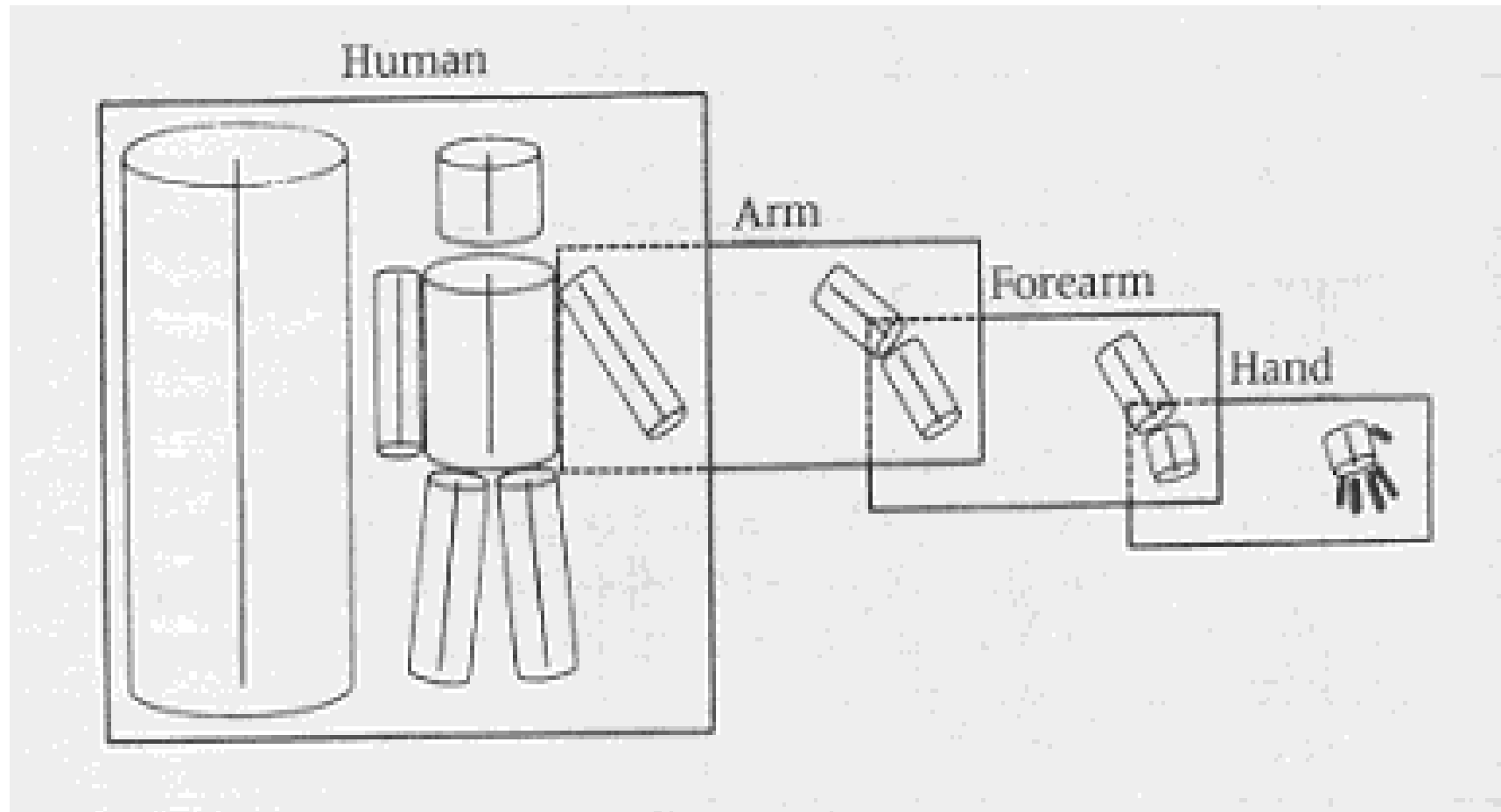
In realtà si usano i coni generalizzati: *geoni*

Esistono forme che non si possono rappresentare come coni generalizzati

Es.: giornali spiegazzati  
origami



# Organizzazione gerarchica degli oggetti complessi





# Conclusioni su approcci top-down

- caratteristica fondamentale: uso di informazioni di alto livello
- assunzione sperimentale: in presenza di dati degradati guida da conoscenze di alto livello



Pura illusione



# Conclusioni generali sulla visione

- Problema della visione: dall'immagine 2-D al mondo 3-D
- Elaborazione bottom-up (basso livello)
  - dall'evoluzione, incapsulata nel sistema nervoso
  - livello non influenzato dal controllo conscio
- Elaborazione top-down (alto livello)
  - acquisita dall'individuo, esplicita, accessibile
  - consapevolezza (e non) dell'uso di tale conoscenza

Grazie dell'attenzione