Introduzione alla Visione Artificiale

Giuseppe Sansonetti
gsansone@dia.uniroma3.it

Dipartimento di Informatica e Automazione

Sommario

- Percezione
- Formazione delle immagini
- Elaborazione delle immagini a basso livello
- Estrazione di informazione 3D da un'immagine
- Riconoscimento di oggetti
- Conclusioni

Percezione

- L'agente acquisisce informazioni sul mondo che lo circonda tramite i *sensori*
- Percezione *visiva*
- Gli stimoli percettivi sono utilizzati secondo due diversi approcci
  - Estrazione di caratteristiche (l'agente estrae dallo stimolo un numero limitato di feature)
  - Approccio basato sul modello (l'agente utilizza lo stimolo per ricostruire un modello del mondo)

Percezione

- Approccio basato sul modello:

$$S = f(W)$$

  dove W è il mondo e S è lo stimolo prodotto
- La funzione f è definita e ben compresa dalla fisica
- *Grafica computerizzata:* ottenere S da f e W
- *Visione artificiale:* ottenere W da f e S

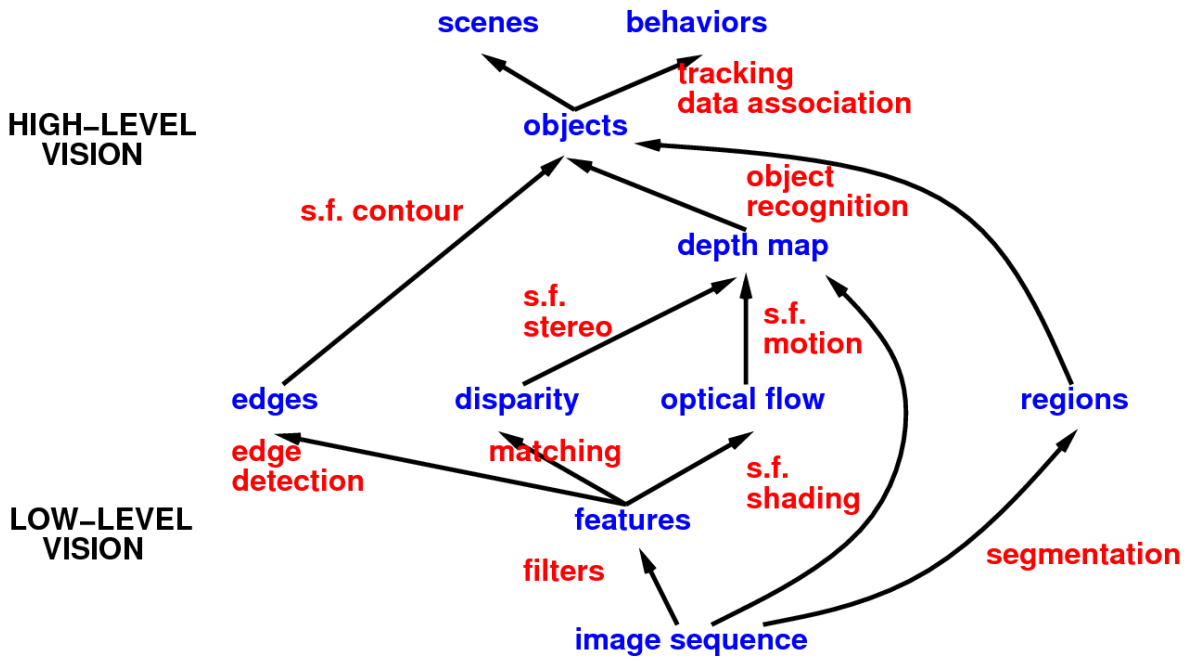
$$W = f^{-1}(S)$$

  Problema difficile per vari motivi
  - funzione f^{-1} indefinita
  - ambiguità
  - alta complessità computazionale

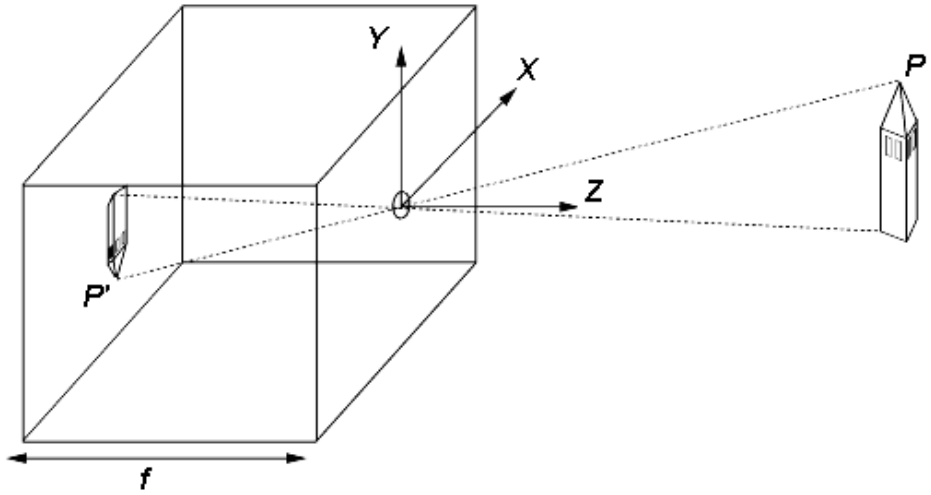
Percezione

- Approcci migliori
  - $P(W) = P(W/S)\ P(S)$
  - estrarre la sola informazione utile per il task
    - ✓ navigazione
    - ✓ manipolazione
    - ✓ classificazione / identificazione
    - ✓ ...

Percezione



Formazione delle immagini



P : punto nella scena, di coordinate *(X,Y,Z)*

P' : immagine di *P* sul piano d'immagine, di coordinate *(x,y)*

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

Proiezione Prospettica

Formazione delle immagini

Formazione delle 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

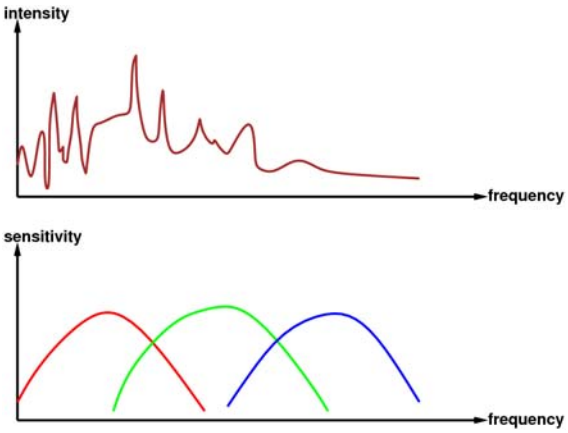
I(x,y,t): intensità in *(x,y)* all'istante *t*

Sensore *CCD* ≈ 5.000.000 pixel

Occhio umano ≈ 240.000.000 pixel (i.e., 0.25 terabit/s)

Formazione delle immagini

L'intensità varia con la frequenza ⇒ segnale a infinite dimensioni

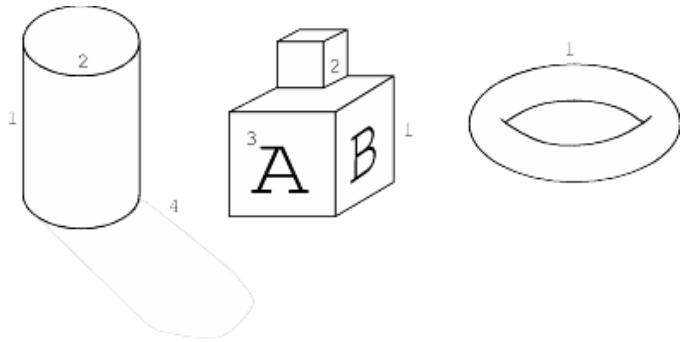


L'occhio umano ha tre tipi di cellule sensibili al colore.

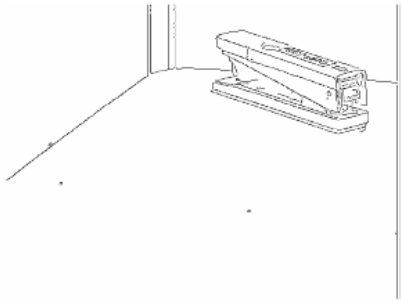
Ciascuna integra il segnale ⇒ un'immagine può essere rappresentata mediante un vettore di tre soli valori di intensità per pixel, uno per ciascuna lunghezza d'onda primaria

Elaborazione delle immagini a basso livello

- *Rilevamento dei bordi (edge detection)*
  Discontinuità nella scena ⇒ bordi nell'immagine
  Esistono diversi tipi di bordi, relativi a
  - discontinuità di profondità (1)
  - discontinuità di orientamento delle superfici (2)
  - discontinuità di riflettanza (3)
  - discontinuità di illuminazione (ombre) (4)



Elaborazione delle immagini a basso livello

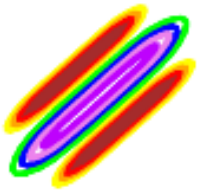


Elaborazione delle immagini a basso livello

- *Rilevamento dei bordi*

1. Si calcola la convoluzione dell'immagine *I(x,y)* con i filtri spazialmente orientati (possibilmente multi-scala) $f_V(x,y)$ e $f_H(x,y)$, ottenendo rispettivamente $R_V(x,y)$ e $R_H(x,y)$. Si definisce

$$R(x,y) = R_V^2(x,y) + R_H^2(x,y)$$

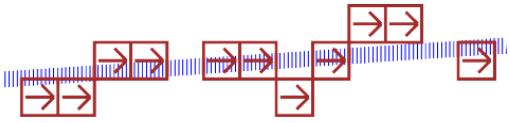


Elaborazione delle immagini a basso livello

- *Rilevamento dei bordi*

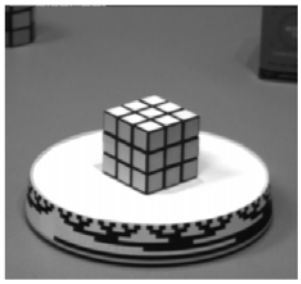
2. Si marcano i picchi in $\|R(x,y)\|$ che superano una soglia *T* specificata

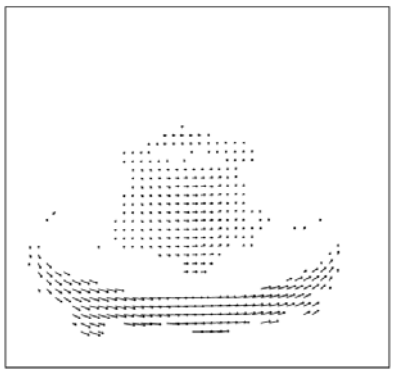
3. Si collegano fra loro i pixel di bordo che appartengono alle stesse curve (due pixel adiacenti che fanno parte di un bordo con lo stesso orientamento appartengono alla stessa curva)



Estrazione di informazione 3D da un'immagine

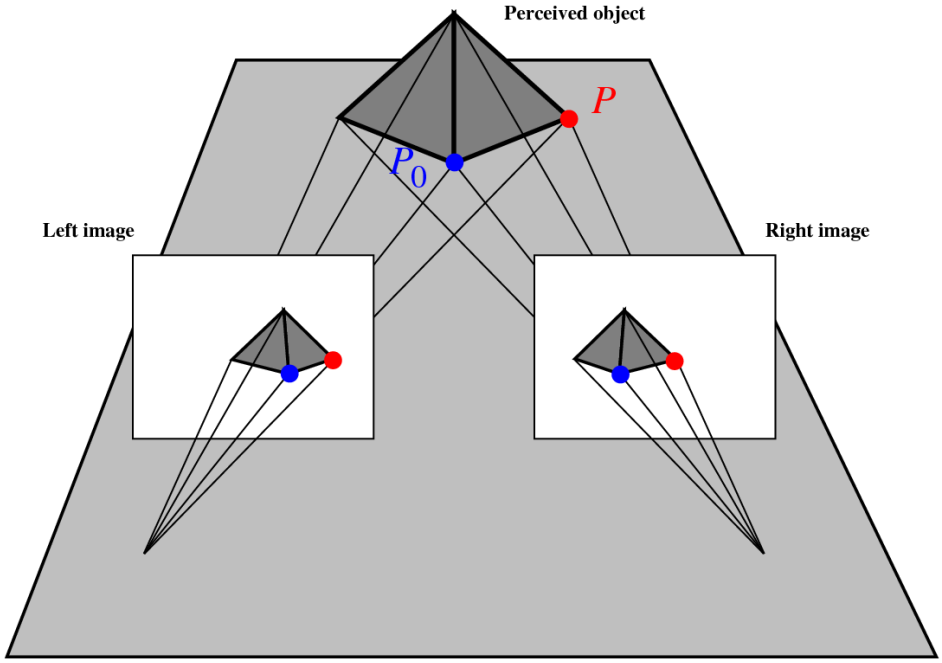
- *Moto*





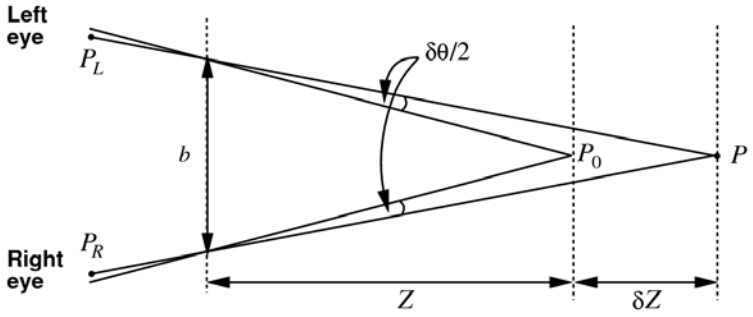
Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*



Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*

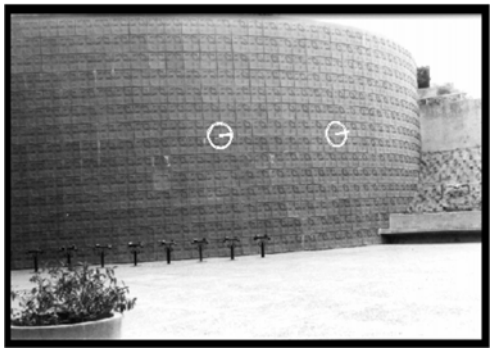


Geometria: $\delta Z = Z^2 \delta\theta / -b$
Fisiologia: $\delta\theta \geq 2.42 \times 10^{-5}$ radianti, b = 6 cm
Z = 30 cm $\Rightarrow \delta Z \approx 0.04$ mm,
Z = 30 m $\Rightarrow \delta Z \approx 40$ mm

Linea di base b maggiore $\Rightarrow$ Risoluzione maggiore

Estrazione di informazione 3D da un'immagine

- *Gradiente di texture*



Idea: presumendo che la texture reale sia uniforme, è possibile ricostruire l'orientamento della superficie

Estrazione di informazione 3D da un'immagine

- *Ombreggiatura (shading)*
  - Definizione: variazione di intensità luminosa su una superficie nella scena 3D dovuta alla geometria e alle proprietà di riflettanza
  - Problema non risolvibile, se non sotto hp semplificative
    - ✓ superficie *lambertiana*
    - ✓ fonte di luce puntiforme

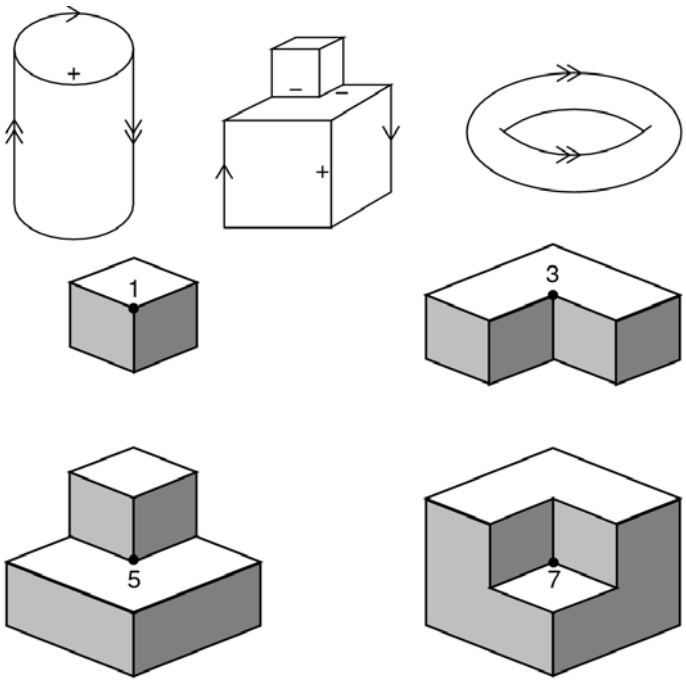
$$I(x,y) = k\boldsymbol{n}(x,y) \cdot \boldsymbol{s}$$

con k fattore di scala, $\boldsymbol{n}$ versore normale alla superficie e $\boldsymbol{s}$ versore della fonte di luce

  - Mappa di riflettanza R(**n**)
  - Problema dei riflessi interni

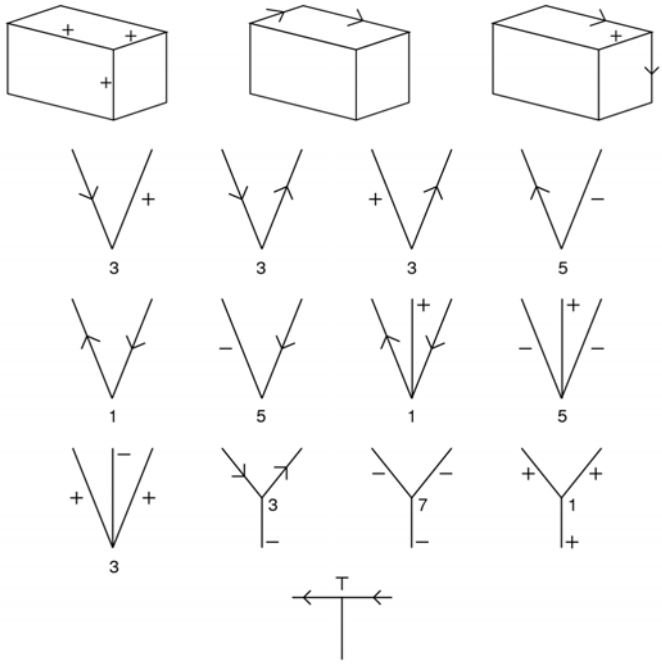
Estrazione di informazione 3D da un'immagine

- *Contorni*



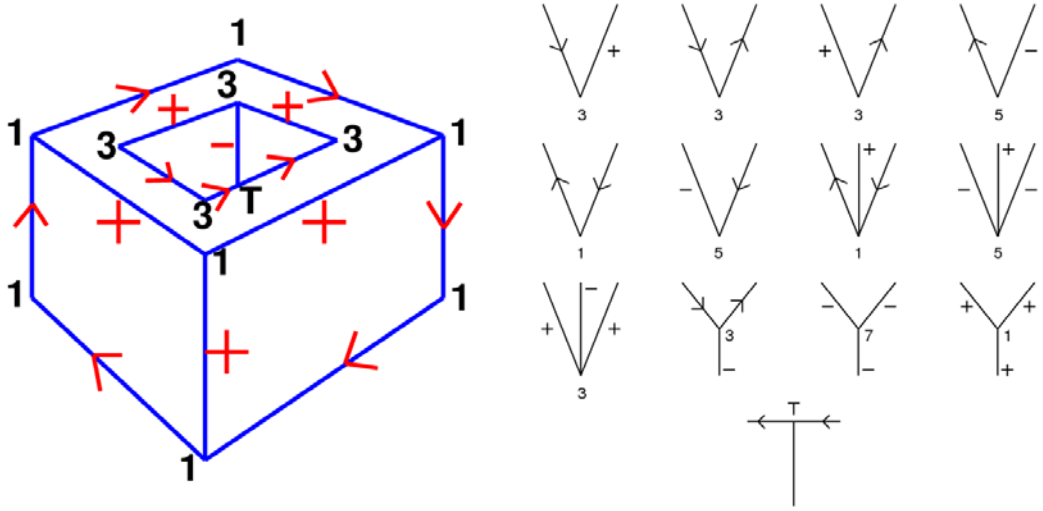
Estrazione di informazione 3D da un'immagine

- *Contorni*



Estrazione di informazione 3D da un'immagine

- *Contorni*

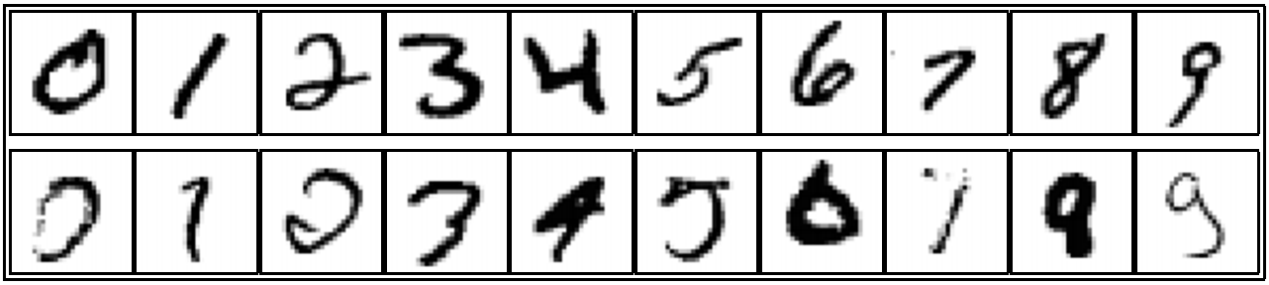


CSP: variabili = bordi, vincoli = possibili configurazioni nodi

Riconoscimento di oggetti

- Idea
  - estrarre forme 3D dalle immagini
  - operare il matching con una “libreria di forme”
- Problemi
  - estrarre superfici curve dall’immagine
  - rappresentare la forma dell’oggetto estratto
  - rappresentare la forma e la variabilità delle classi di oggetti
  - segmentazione errata
  - occlusioni
  - illuminazione indeterminata, ombre, segni, rumore, complessità, etc.
- Approcci
  - indicizzazione della libreria tramite misurazione delle proprietà invarianti degli oggetti
  - allineamento delle caratteristiche dell’immagine con quelle dell’oggetto della libreria
  - matching dell’immagine con visioni multiple dell’oggetto memorizzate nella libreria
  - metodi di *machine learning* basati sulle proprietà statistiche dell’immagine

Riconoscimento di oggetti



3–Nearest–Neighbor = 2.4% errore

400–300–10 unità MLP = 1.6% errore

LeNet: 768–192–30–10 unità MLP = 0.9% errore

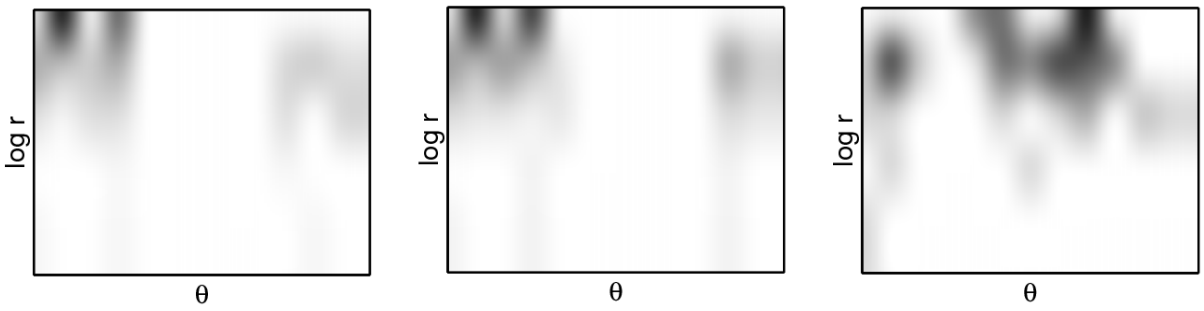
Riconoscimento di oggetti

- *Riconoscimento basato sulle caratteristiche*

Idea: convertire la *forma* (un concetto relazionale) in un set fisso di *attributi* tramite il *contesto spaziale* di ciascun punto di un set fisso sulla superficie della forma

Riconoscimento di oggetti

- *Riconoscimento basato sulle caratteristiche*



Ciascun punto è descritto tramite il suo istogramma di contesto locale (numero di punti che ricadono all'interno di ciascun *bin* del diagramma in coordinate polari logaritmiche)

Riconoscimento di oggetti

- *Riconoscimento basato sulle caratteristiche*

  Determinare la distanza totale fra forme dalla somma di distanze per punti corrispondenti sotto il miglior matching

  Un semplice classificatore Nearest-Neighbor dà 0.63% di tasso di errore sui dati del NIST

Conclusioni

- La visione è caratterizzata pesantemente da rumore, ambiguità, complessità
- La conoscenza a priori è necessaria per porre dei vincoli sul problema
- Necessità di combinare fra loro indizi multipli: moto, contorni, ombreggiatura, texture, stereoscopia
- Rappresentazione di oggetti della “libreria”: forma e aspetti
- Matching immagine/oggetto: feature, linee, regioni, etc.

Riferimenti bibliografici

- **Russell, S., and Norvig, P. *Artificial Intelligence: A Modern Approach (2nd Edition).* Prentice-Hall, 2003.**
- Forsyth, D.A., and Ponce, J. *Computer Vision: A Modern Approach.* Prentice-Hall, 2003.
- Shapiro, L., and Stockman, G. *Computer Vision.* Prentice-Hall, 2001.
- Ballard, D.H., and Brown, C.M. *Computer Vision.* Prentice-Hall, 1982.
- http://iris.usc.edu/Vision-Notes/bibliography/contents.html