Sommario

- Introduzione
- Percezione
- Formazione delle immagini
- Elaborazione delle immagini a basso livello
- Estrazione di informazione 3D da un’immagine
- Riconoscimento di oggetti
- Manipolazione e navigazione
- Conclusioni

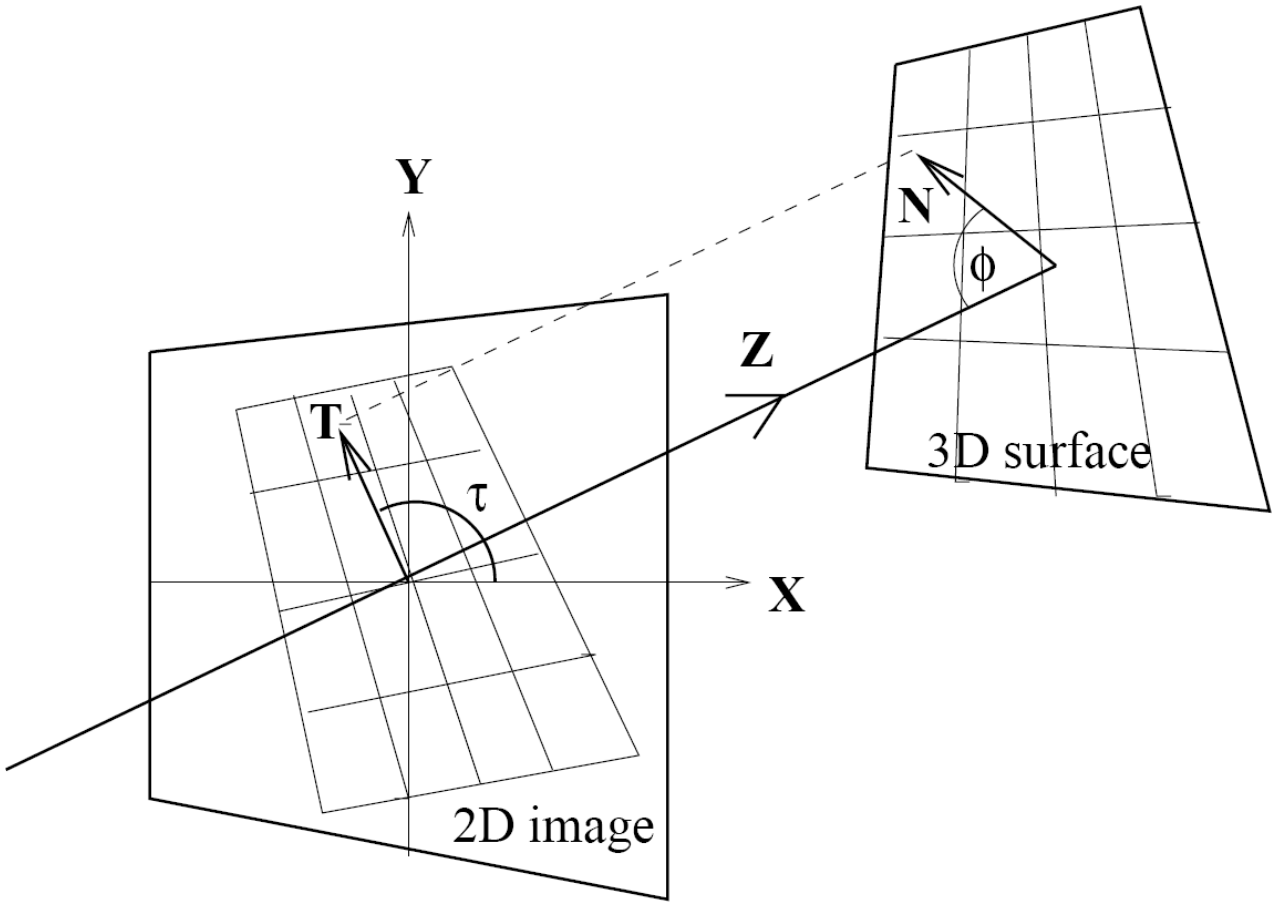
Estrazione di informazione 3D da un'immagine

- Vediamo ora come sia possibile passare da un'immagine 2D a una rappresentazione 3D della scena
- La maggior parte degli agenti non necessita di tutti i dettagli della scena, ma solo di una rappresentazione astratta limitata di alcuni dei suoi aspetti
- *Riconoscimento di oggetti:* processo di conversione delle caratteristiche dell'immagine (e.g., i bordi) in un modello di oggetti noti (e.g., la cucitrice); si articola in tre fasi
  - *segmentazione* della scena in oggetti distinti
  - determinazione della *posizione* e *orientazione* di ogni oggetto rispetto all'osservatore *(posa)* $\Rightarrow$ fondamentale per la manipolazione e la navigazione *(feedback)*
  - determinazione della *forma* di ogni oggetto

Estrazione di informazione 3D da un'immagine

- Occorre specificare *posizione* e *orientazione* in termini matematici
  - conosciamo la *proiezione prospettica (x,y)* del punto P(X,Y,Z) della scena 3D sul piano dell'immagine, non conosciamo la *distanza*
  - il termine *orientazione* può avere due significati
    - *orientazione dell'oggetto nella sua interezza* (può essere specificato come una sua rotazione 3D rispetto alla telecamera)
    - *orientazione della superficie dell'oggetto nel punto P* (può essere specificato con **n** (versore perpendicolare alla superficie), o con le variabili *angolazione (slant)* e *inclinazione (tilt)*)

Estrazione di informazione 3D da un'immagine



Slant φ : angolo fra la normale alla superficie N e l'asse Z *(line of sight)*

Tilt τ : angolo fra la proiezione di N sul piano d'immagine e l'asse X

Estrazione di informazione 3D da un'immagine

- Quando la telecamera si muove rispetto a un oggetto cambiano *distanza* e *orientazione*, non la *forma*
- Difficoltà nel definire matematicamente il concetto di *forma* partendo dall'invarianza rispetto ad alcune trasformazioni
- Rappresentazione *globale* di forme generali: impossibile!
- Caratterizzazione della forma *locale:* possibile (*e.g., curvatura:* modo in cui la normale cambia lungo una superficie); se ne occupa la Geometria Differenziale
- *Forma*, *colore*, *texture*: indizi più utili per riconoscere un oggetto
- Questione fondamentale: con la *proiezione prospettica* tutti i punti della scena 3D sullo stesso raggio sono proiettati nello stesso punto immagine $\Rightarrow$ come recuperare l'informazione 3D?

Estrazione di informazione 3D da un'immagine

Lo stimolo visivo contiene molti *indizi* che possono essere utilizzati per recuperare l'informazione 3D

- moto *(motion)*
- stereoscopia binoculare *(binocular stereopsis)*
- gradiente di texture *(texture)*
- ombreggiatura *(shading)*
- contorni *(contour)*

Estrazione di informazione 3D da un'immagine

- *Moto*

  - Le *differenze* fra fotogrammi consecutivi possono essere un'importante fonte di informazione

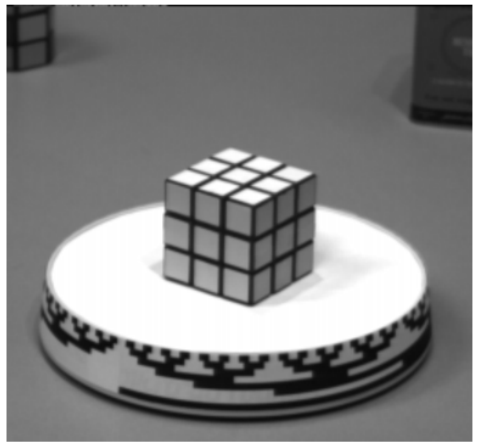
  - Se la telecamera si muove rispetto alla scena 3D, il moto apparente che ne risulta prende il nome di *flusso ottico*

  - La direzione e la velocità del movimento degli elementi *nell'immagine* sono una conseguenza del *moto relativo* fra osservatore e scena

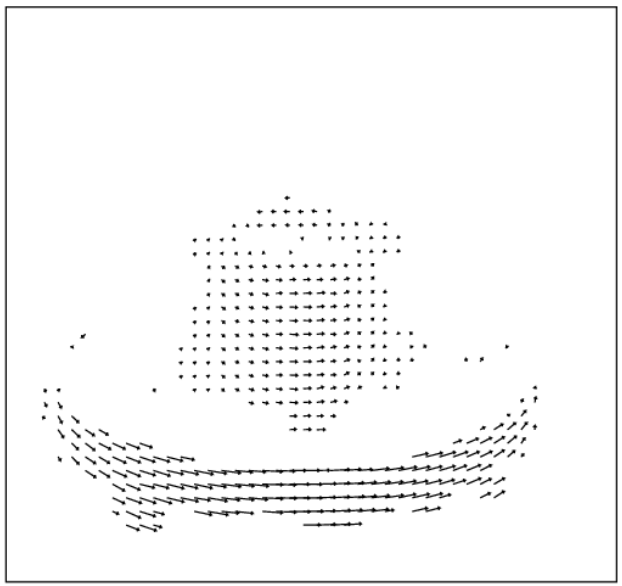
Estrazione di informazione 3D da un'immagine

- *Moto*

Due fotogrammi del video di un *cubo di Rubik* su una piattaforma rotante (il secondo fotogramma è relativo a 19/30 di secondo dopo)



In basso sono riportati i *vettori di flusso ottico* calcolati confrontando le due immagini sopra



Estrazione di informazione 3D da un'immagine

- *Moto*
  - Il *flusso ottico* codifica informazione utile sulla struttura della scena (e.g., oggetti distanti mostrano un moto apparente più lento di quelli vicini $\Rightarrow$ è possibile stimare la distanza)

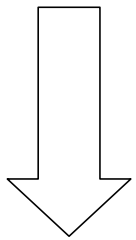
  - Il *campo dei vettori di flusso ottico* è rappresentato dai suoi componenti $v_x(x,y)$ in direzione x e $v_y(x,y)$ in direzione y

  - Per misurare il *flusso* occorre trovare i punti corrispondenti fra un fotogramma e il successivo: aree della immagine centrate su punti corrispondenti hanno schemi di intensità simili $\Rightarrow$ il blocco di pixel centrato intorno al pixel $p(x_0,y_0)$ nell'istante t_0 va confrontato con altri blocchi centrati su diversi pixel candidati q_i in (x_0+D_x, y_0+D_y) nell'istante t_0+D_t

Estrazione di informazione 3D da un'immagine

- *Moto*
  - Per misurare la somiglianza si può minimizzare la *Somma delle Differenze Quadrate (SSD)*

$$SSD(D_x, D_y) = \sum_{(x,y)} (I(x, y, t) - I(x + D_x, y + D_y, t + D_t))^2$$



Flusso Ottico in (x_0, y_0): $(v_x, v_y) = (D_x / D_t, D_y / D_t)$

dove (D_x, D_y) è il punto che minimizza la *SSD*

Estrazione di informazione 3D da un'immagine

- *Moto*
  - Per misurare la somiglianza si può massimizzare la *Cross-Correlazione*

$$Correlazione(D_x, D_y) = \sum_{(x,y)} I(x, y, t) I(x + D_x, y + D_y, t + D_t)$$

Risultati migliori in presenza di *texture* nella scena

Estrazione di informazione 3D da un'immagine

- *Moto*

  E' possibile ricavare l'equazione che lega fra loro le *velocità dell'osservatore*, il *flusso ottico* e le *posizioni degli oggetti nella scena* (hp: *lunghezza focale f* = 1)

$$v_x(x,y) = \left[-\frac{T_x}{Z(x,y)} - \omega_y + \omega_z y \right] - x\left[-\frac{T_z}{Z(x,y)} - \omega_x y + \omega_y x \right]$$

$$v_y(x,y) = \left[-\frac{T_y}{Z(x,y)} - \omega_z x + \omega_x \right] - y\left[-\frac{T_z}{Z(x,y)} - \omega_x y + \omega_y x \right]$$

  con *T velocità traslazionale* e ω *velocità angolare* dell'osservatore *(egomovimento)* e *Z(x,y)* coordinata *z* del punto nella scena 3D corrispondente al punto in *(x,y)* nell'immagine 2D.

  Tale equazione prende il nome di ***Equazione del Flusso Ottico***

Estrazione di informazione 3D da un'immagine

- *Moto*

Nel caso di *traslazione pura* ($\omega_x = \omega_y = \omega_z = 0$) le componenti del campo di flusso diventano

$$v_x(x,y) = \left[-\frac{T_x}{Z(x,y)}\right] - x\left[-\frac{T_z}{Z(x,y)}\right] = +\frac{-T_x + xT_z}{Z(x,y)}$$

$$v_y(x,y) = \left[-\frac{T_y}{Z(x,y)}\right] - y\left[-\frac{T_z}{Z(x,y)}\right] == +\frac{-T_y + yT_z}{Z(x,y)}$$

Tali componenti valgono zero nel punto

$$x = T_x / T_z$$

$$y = T_y / T_z$$

che prende il nome di ***fuoco di espansione*** del campo di flusso

Estrazione di informazione 3D da un'immagine

- *Moto*

Supponiamo di porre l'origine del piano *x-y* nel *fuoco di espansione* eseguendo il seguente cambiamento di coordinate

$$x' = x - T_x / T_z$$

$$y' = y - T_y / T_z$$

in seguito al quale si ottiene

$$v_x(x', y') = + \frac{x' T_z}{Z(x', y')}$$

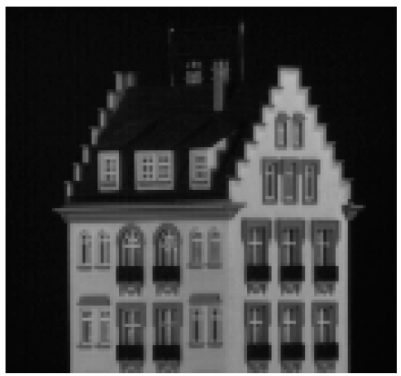
$$v_y(x', y') = + \frac{y' T_z}{Z(x', y')}$$

Il campo di flusso ottico istantaneo non fornisce quindi né la distanza *Z*, né la componente di velocità T_z ,ma il rapporto fra le due, il che può risultare utile in talune applicazioni

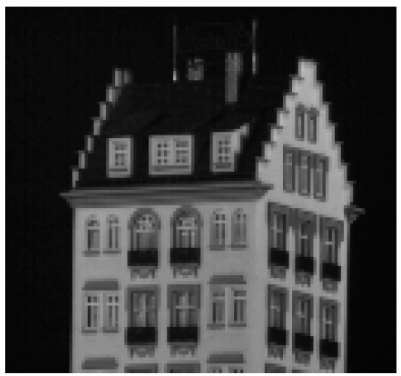
Estrazione di informazione 3D da un'immagine

- *Moto*

  Tramite più fotogrammi successivi è possibile ottenere informazioni riguardanti la *profondità*

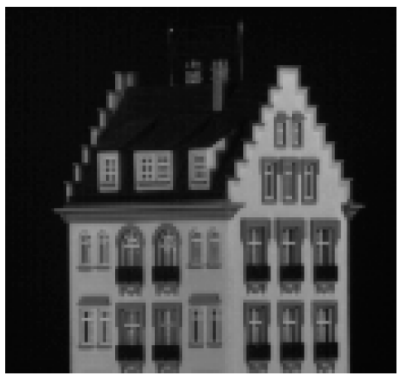


1

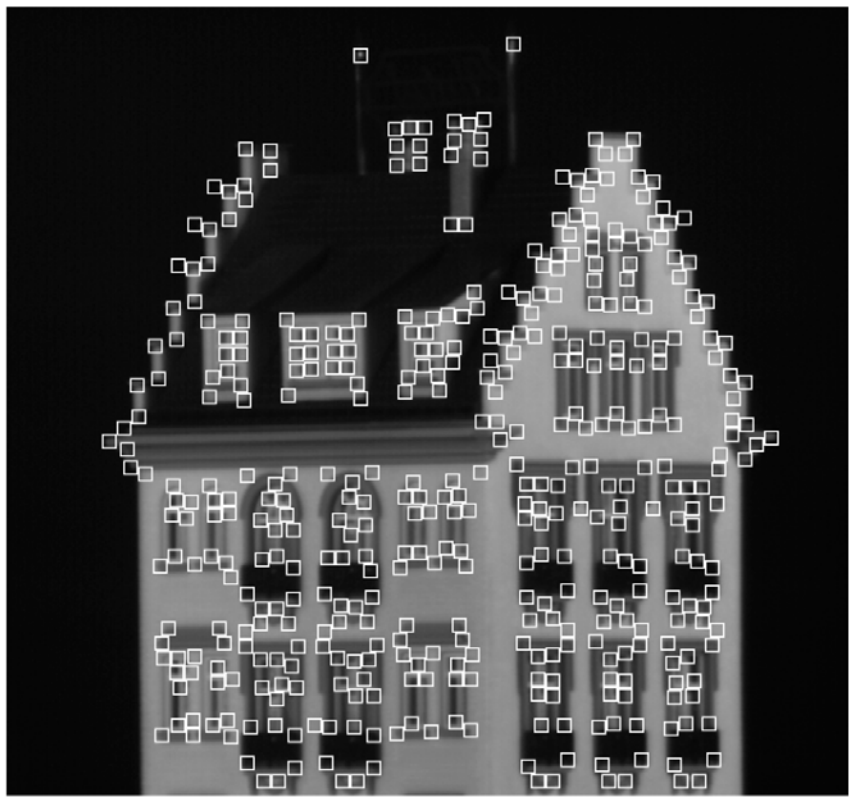


60

120



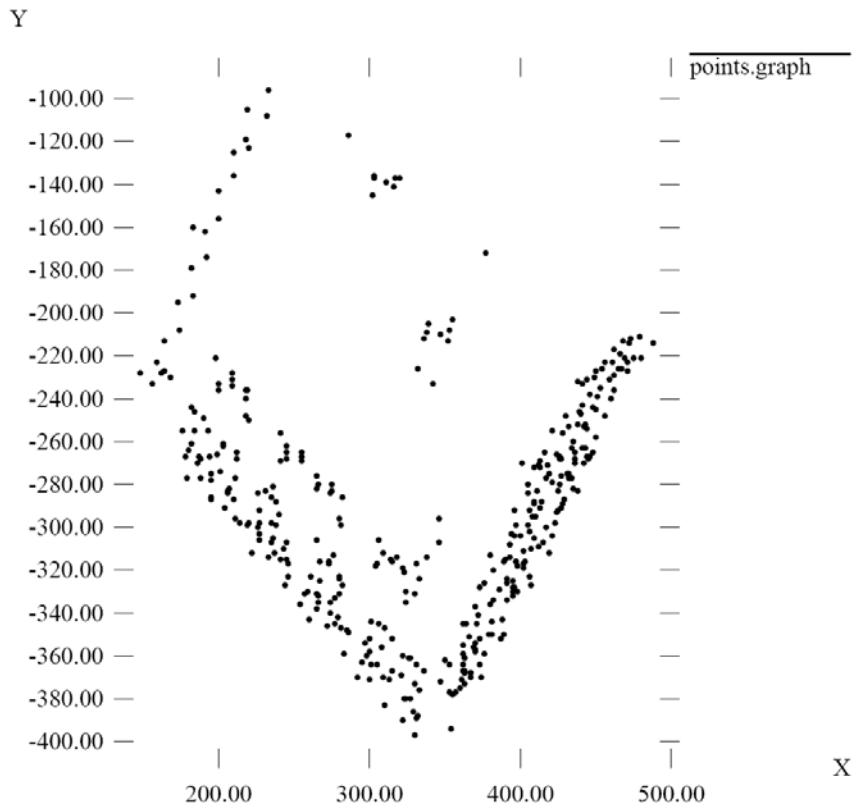
150

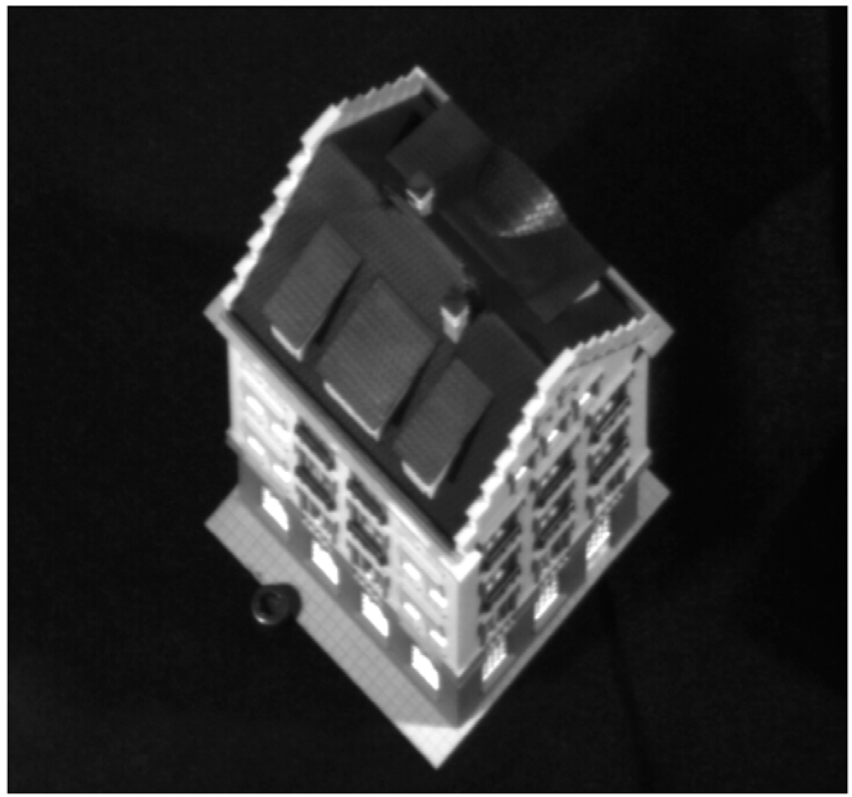


Estrazione di informazione 3D da un'immagine

- *Moto*

  Tramite più fotogrammi successivi è possibile ottenere informazioni riguardanti la *profondità*





Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - Dal greco στερεο′ζ, "solido": *stereovisione* ≈ "visione solida"
  - Una "visione solida" presuppone un meccanismo di percezione della *profondità:* il più comune è fondere due proiezioni della medesima scena ottenuta da due punti di vista differenti
  - La "fusione" non è un compito banale, ma il nostro cervello è in grado di svolgerlo egregiamente

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - Esperimento 1: allineare la punta di due penne tenute una con la destra e l'altra con la sinistra, prima tenendo chiuso un occhio (difficile), poi con tutti e due gli occhi aperti (facile)
  - Esperimento 2:
    - Fissare la faccina
    - Posizionare il pollice in modo che la faccina appaia a destra di esso se visto con l'occhio destro, a sinistra se visto con l'occhio sinistro
    - Aprire tutti e due gli occhi e mettere a fuoco il dito: si vedono due faccine?
    - Mettere a fuoco la faccina: si vedono due dita?

La stereoscopia è talmente radicata nel nostro cervello che esso "inventa" la realtà pur di non rinunciare alle sue informazioni

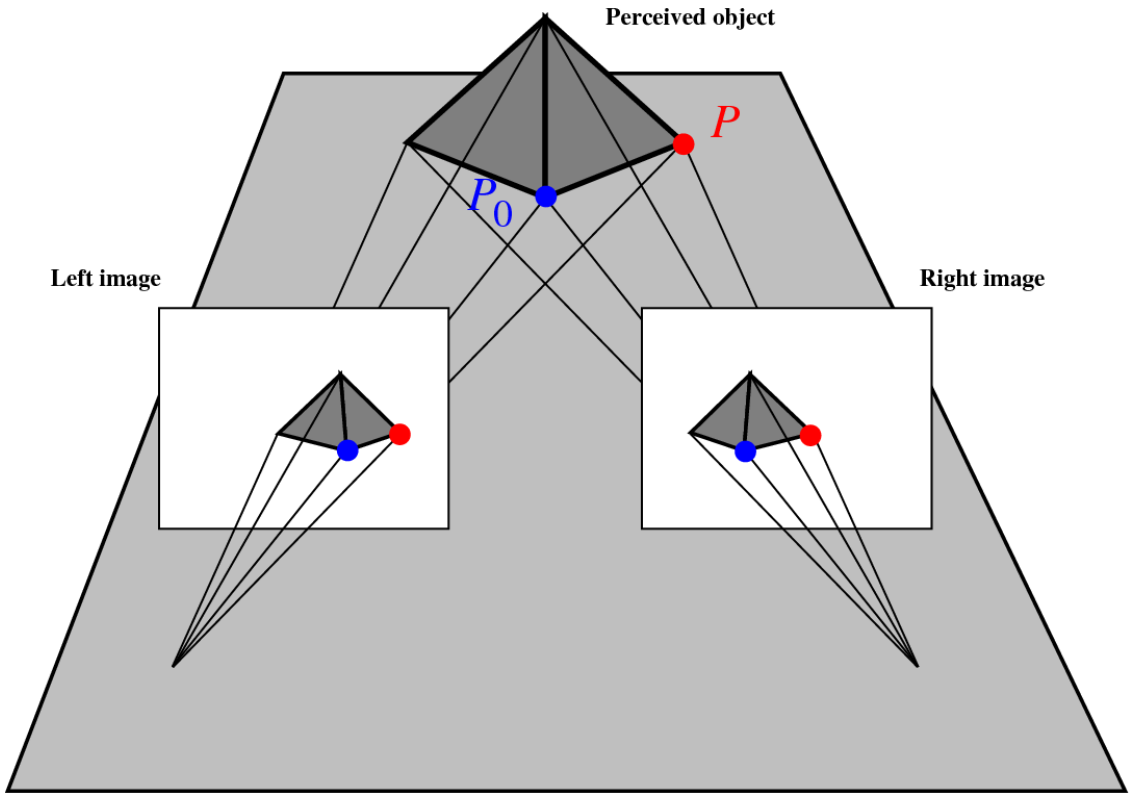
Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - La maggior parte degli animali ha due occhi $\Rightarrow$ i predatori sfruttano la *sterescopia binoculare*
  - L'idea della *stereoscopia binoculare* è simile a quella della *parallasse del movimento*, ma con due (o più) immagini separate nello spazio (e.g., quelle prodotte dai due occhi degli esseri umani), anziché consecutive nel tempo
  - In due immagini separate nello spazio, un elemento della scena si trova in posizione differente rispetto all'asse *z* di ogni piano d'immagine $\Rightarrow$ sovrapponendo le due immagini si ha una *disparità*

Estrazione di informazione 3D da un'immagine

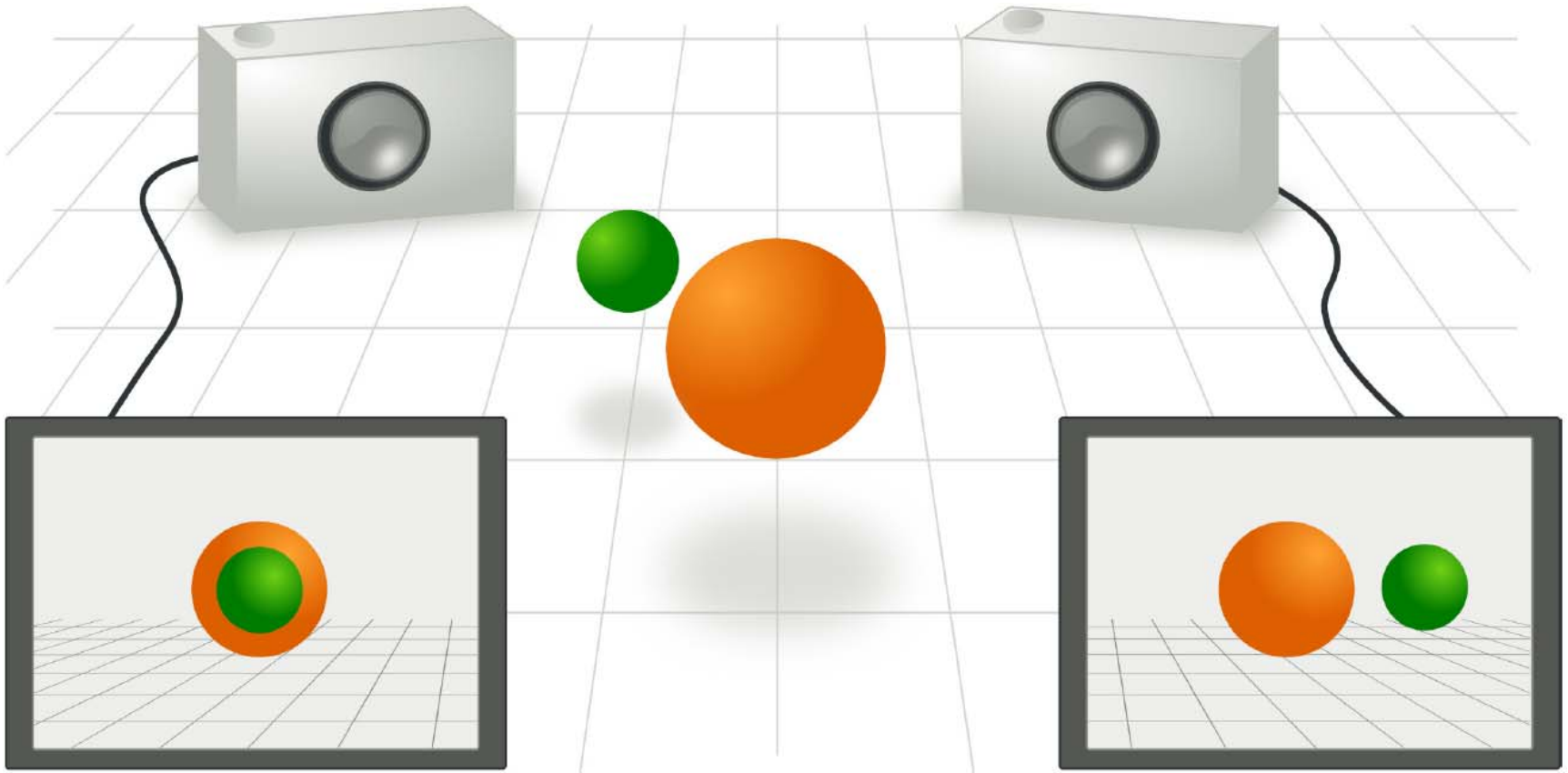
- *Stereoscopia binoculare*

L'idea della stereoscopia: differenti posizioni delle telecamere danno come risultato viste 2D leggermente diverse fra loro della stessa scena 3D



Estrazione di informazione 3D da un’immagine

- *Stereoscopia binoculare*



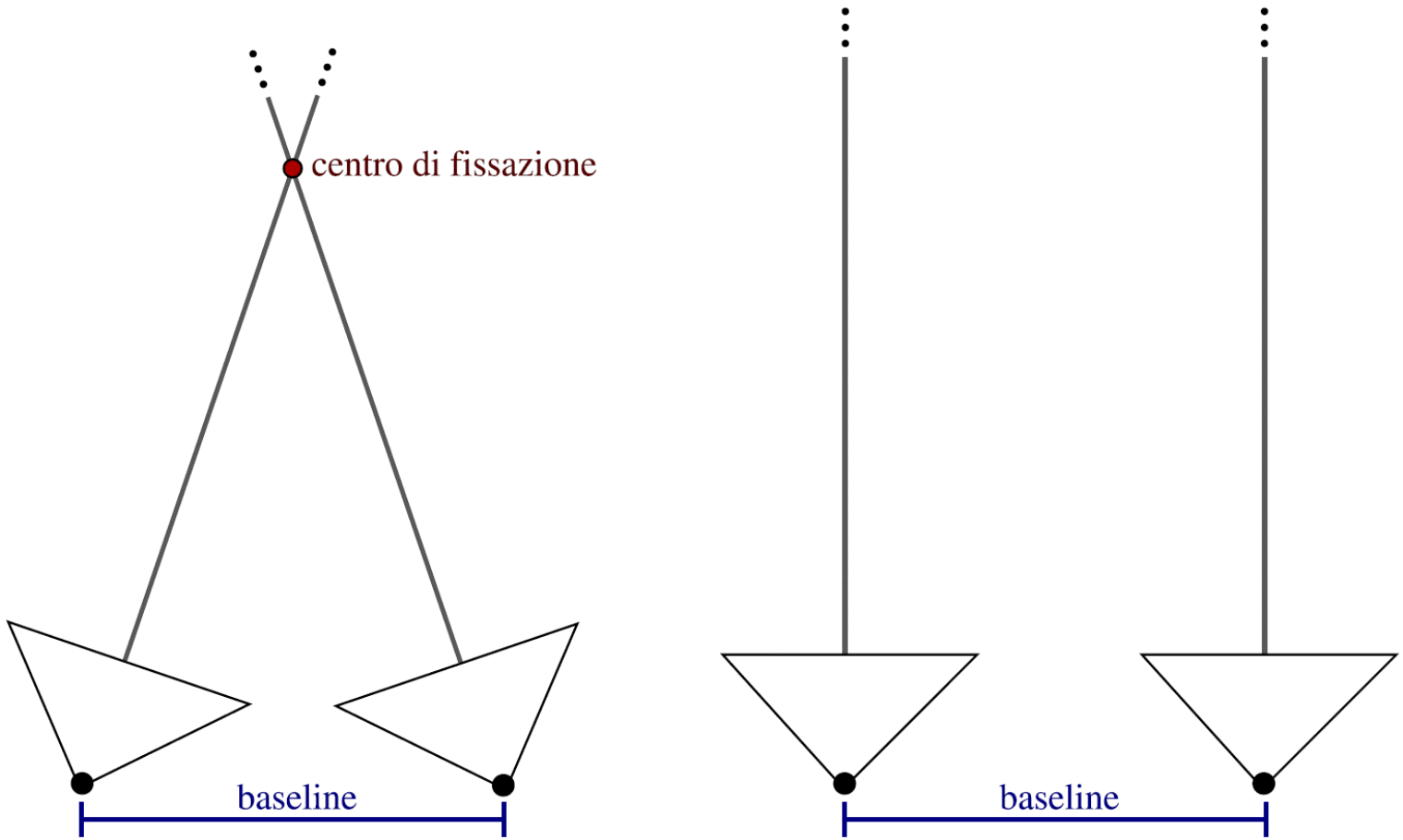
Ogni camera “vede” la scena in modo lievemente diverso dall’altra

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - Un *sistema stereo* è un sistema di visione consistente in (almeno) due camere che osservano la medesima scena; solitamente ci si riferisce alle due camere come *camera sinistra* e *camera destra*
  - Gli *assi focali* delle camere possono essere fra loro *paralleli* o *incidenti;* nel primo caso il *centro di fissazione,* ossia il punto di intersezione tra i due assi focali, è all'infinito; nel secondo caso, il più comune, il *centro di fissazione* è un punto nello spazio ad una distanza finita dai fuochi
  - La distanza fra i fuochi delle due camere è detta *linea di base (baseline)* o *distanza inter-ottica*
  - Nel sistema visivo umano gli occhi hanno assi focali incidenti e sono in grado di ruotare, potendo così variare nel tempo la posizione del *centro di fissazione*

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*



Due semplici sistemi stereo con *centro di fissazione* finito e infinito

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - Vogliamo ricavare i *parametri* di un sistema stereo, in particolare
    - ***Parametri intrinseci:*** i *parametri intrinseci* di entrambe le camere, come li abbiamo definiti in precedenza
    - ***Parametri estrinseci:*** posizione e orientazione *relativa* delle camere

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - Siano O_l e O_r le coordinate dei fuochi della camera sinistra e di quella destra, e siano R_l e R_r le matrici di rotazione che ruotano il sistema di riferimento del mondo in quello delle rispettive camere

$$P_r = R_r(P_W - T_r)$$

$$P_l = R_l(P_W - T_l)$$

  - Cerchiamo i vettori T e R del sistema stereo tali che la relazione fra le coordinate di un punto dello spazio nel sistema di riferimento della camera sinistra (P_l) e le coordinate dello stesso punto nel sistema di riferimento della camera destra (P_r) sia

$$P_r = R(P_l - T)$$

relazione fra le coordinate dello stesso punto nei due sistemi di riferimento

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - Il vettore di traslazione che "sposta" da un sistema di riferimento all'altro è

$$T = O_r - O_l$$

  - Per quanto riguarda *R*, notiamo che abbiamo le formule per tradurre $W \rightarrow C_r$ (1) e $W \rightarrow C_l$ (2); per trovare la relazione $C_l \rightarrow C_r$ possiamo invertire la (2), ottenendo $C_l \rightarrow W$, e successivamente applicare la (1).
    Dopo semplici calcoli si ottiene

$$P_r = RP_l - RT = \underbrace{R_r R_l^{-1}}_{R} P_l - \underbrace{R_r (T_r - T_l)}_{RT}$$

    da cui ricaviamo

$$R = R_r R_l^{-1} \quad \text{e} \quad T = R^{-1} R_r (T_r - T_l)$$

Estrazione di informazione 3D da un'immagine

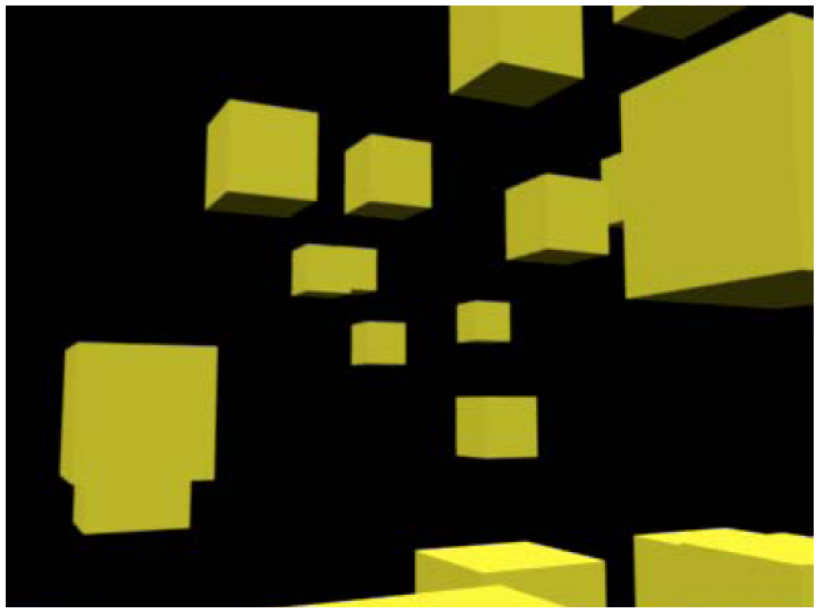
- *Stereoscopia binoculare*
  - Avevamo già calcolato *T* come differenza dei *centri ottici*, ossia
    $$T_o = O_r - O_l$$
    Ora abbiamo trovato un'altra relazione
    $$T = R^{-1} R_r (T_r - T_l)$$
    I due vettori *T* sono uguali in modulo, ma orientati in modo differente, in quanto il primo (che qui abbiamo chiamato T_o per distinguerlo dal secondo) è espresso nel sistema di riferimento del mondo e il secondo in quello di C_r; si ottiene
    $$T_o = T_r - T_l = R_l^T T$$
  - La conoscenza dei parametri intrinseci ed estrinseci di un sistema stereo è fondamentale ai fini della ricostruzione. Senza la conoscenza a priori (o tramite precedente calibrazione) dei parametri, non è possibile ricostruire l'intera scena, se non a meno di un *fattore di scala* o di una *trasformazione proiettiva*

Estrazione di informazione 3D da un'immagine

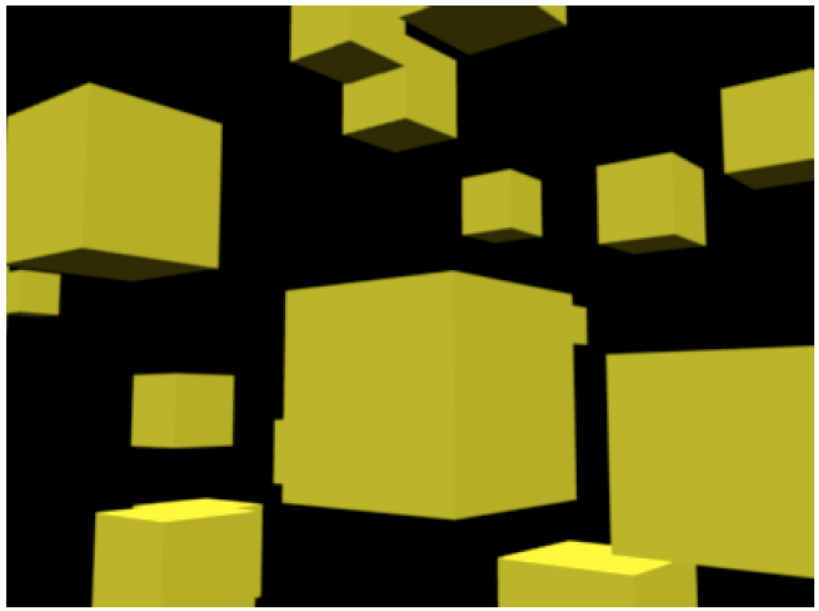
- *Stereoscopia binoculare*
  - I due problemi principali che un *sistema stereo* deve affrontare ai fini della Ricostruzione sono
    - ***Matching*** *(problema della corrispondenza):* individuare la corrispondenza fra i punti dell'immagine prodotta da una camera e i punti dell'immagine prodotta dall'altra camera
    - ***Depth Estimation*** *(problema della ricostruzione):* una volta risolta l'associazione fra punti, occorre ricostruire la posizione nello spazio di quello che vedo

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching***



Vista SX



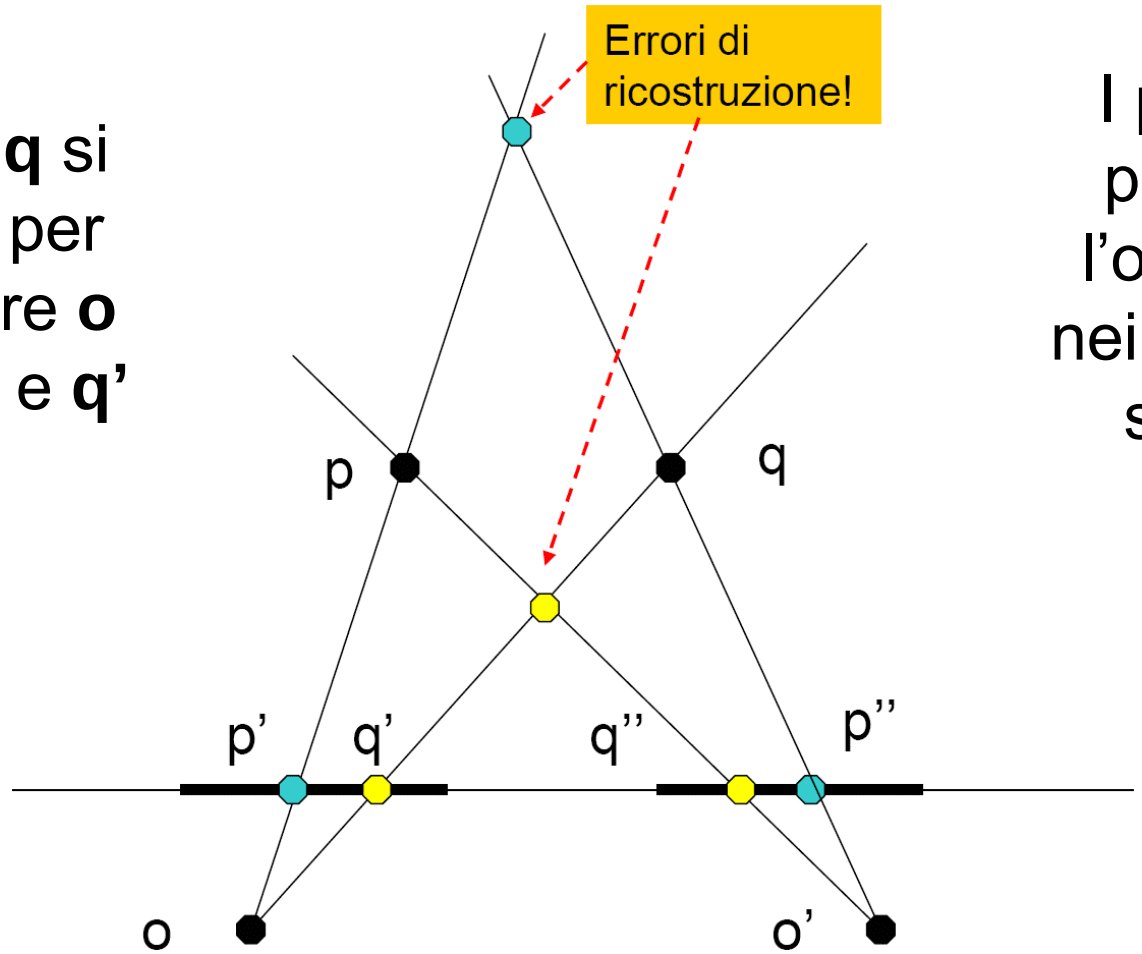
Vista DX

Quale cubo nella vista sinistra corrisponde a quale cubo nella vista destra?

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** errori nel *matching* portano a errori nella *stima di profondità*

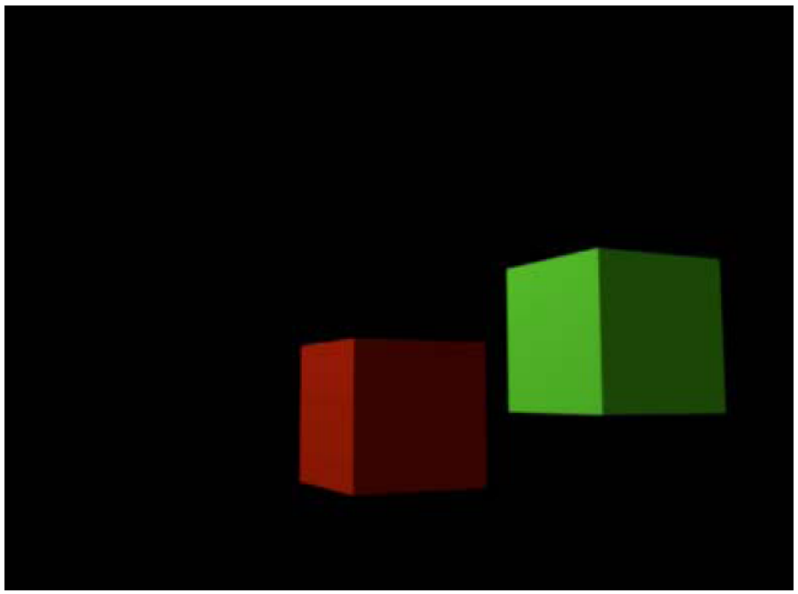
I punti **p** e **q** si proiettano per l'osservatore **o** nei punti **p'** e **q'**



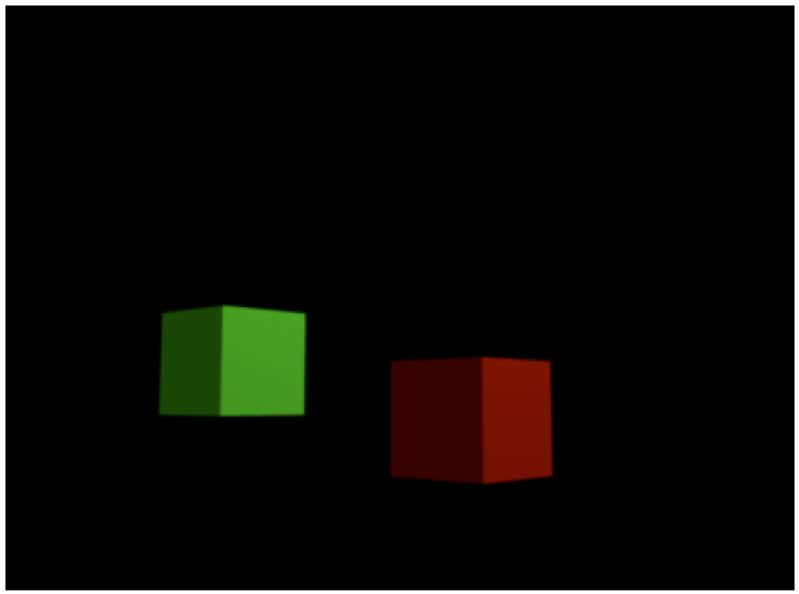
I punti **p** e **q** si proiettano per l'osservatore **o'** nei punti **p''** e **q''**, sbagliando il matching

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Depth Estimation***



Vista SX

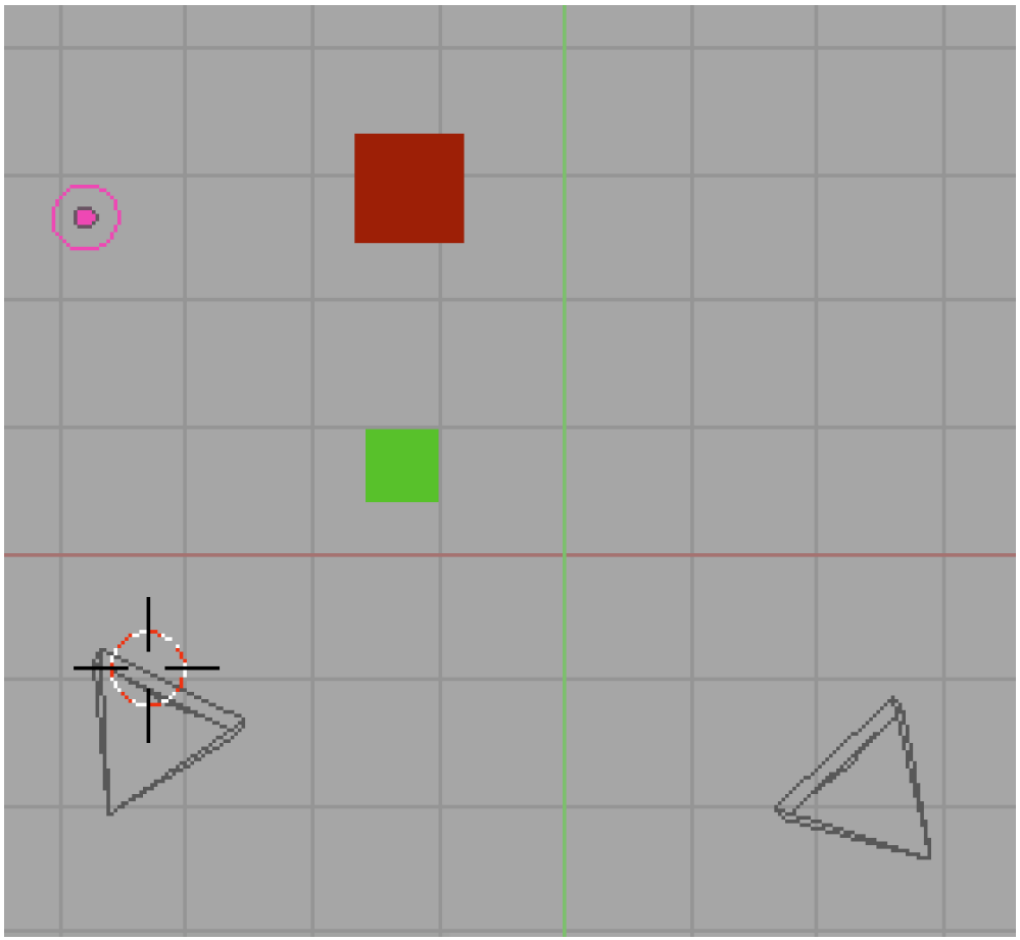


Vista DX

In questo caso il matching è facile, ma qual è la posizione reciproca dei due cubi rosso e verde?

Estrazione di informazione 3D da un'immagine

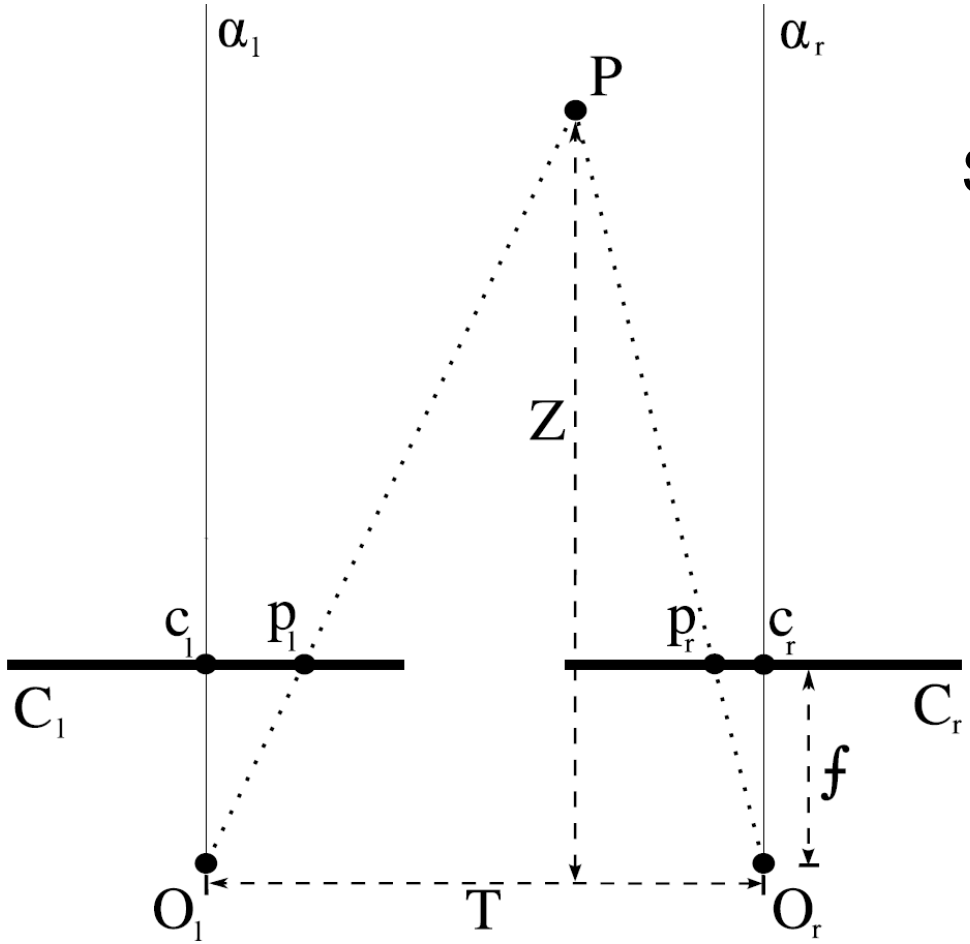
- *Stereoscopia binoculare*
  - ***Depth Estimation***



Vista ortografica dall'alto, corrispondente alla scena raffigurata nelle immagini precedenti

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*

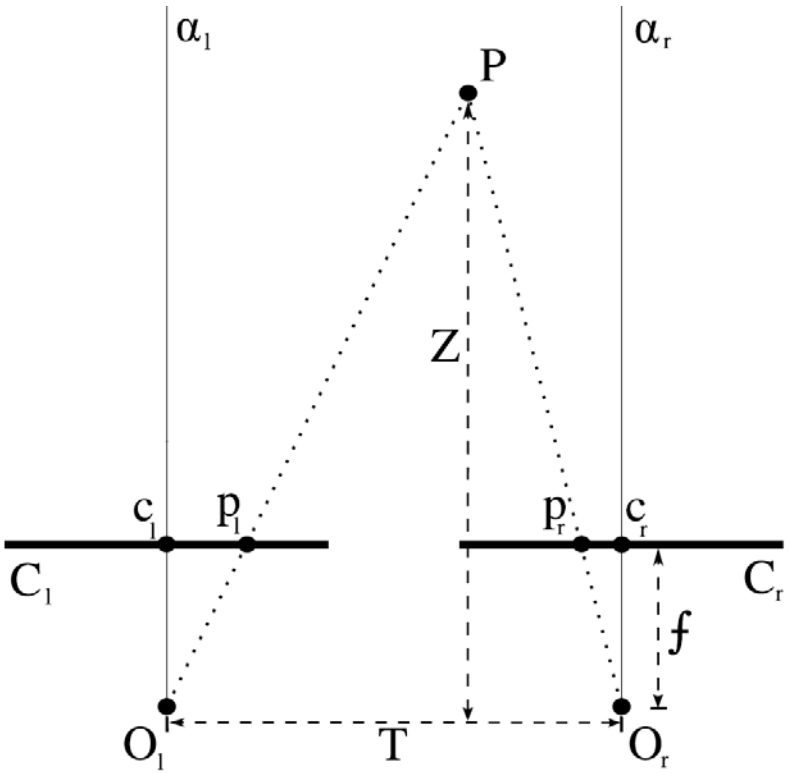


Sistema stereo composto di due camere con centro di fissazione all'infinito

P è proiettato in posizioni differenti sui due piani d'immagine: tale differenza è la ***disparità***

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*



Cerchiamo la relazione fra p_l, p_r e Z

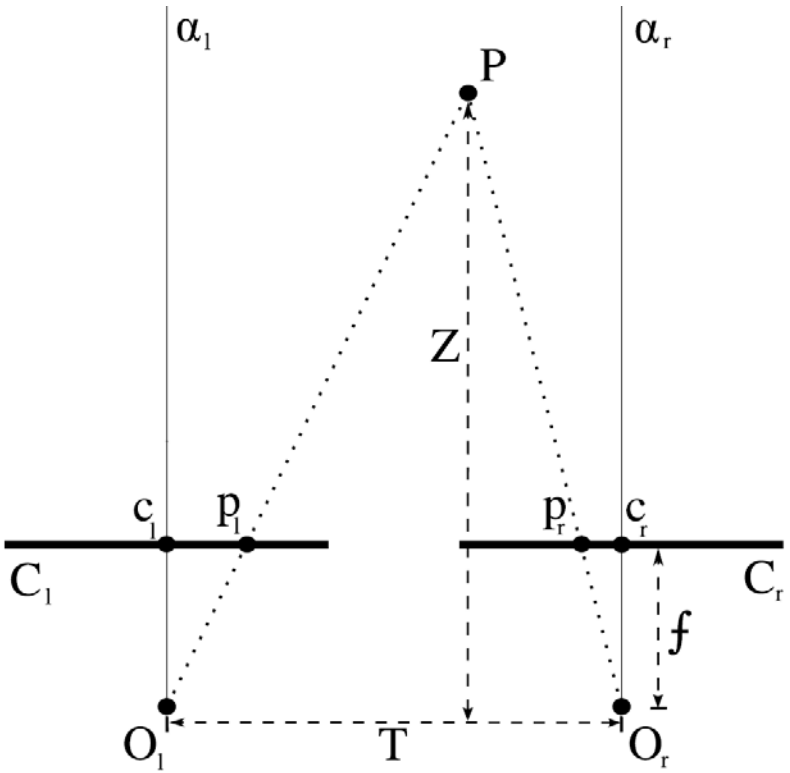
I triangoli $p_l P p_r$ e $O_l P O_r$ sono *simili* $\Rightarrow$ il loro rapporto base/altezza è uguale, quindi (gli assi X delle camere sono crescenti a sinistra)

$$\frac{T}{Z} = \frac{T + x_l - x_r}{Z - f}$$

Chiamiamo ***disparità*** la differenza $d=x_r-x_l$ e risolviamo per Z

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*



$$\frac{T}{Z} = \frac{T + x_l - x_r}{Z - f} \Rightarrow \frac{T}{Z} = \frac{T - d}{Z - f} \Rightarrow$$

$$\Rightarrow (Z - f)T = Z(T - d) \Rightarrow$$

$$\Rightarrow TZ - Tf = TZ - Zd \Rightarrow$$

$$\Rightarrow Z = \frac{Tf}{d}$$

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*

$$Z = \frac{Tf}{d}$$

Questa relazione ha delle importanti conseguenze

- La *profondità* di un punto è inversamente proporzionale alla *disparità*
- La relazione fra *Z* e *d* è non lineare
- Fissate la *lunghezza focale f* e la *linea di base (baseline) T,* la *profondità* di un punto dipende solo dalla *disparità*
- Errori nella stima della *disparità,* in particolare quando essa è molto piccola, si riflettono in grandi errori di stima della *profondità*

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*

$$Z = \frac{Tf}{d}$$

La relazione è stata calcolata considerando un sistema con *punto di fissazione* all'infinito. Nel caso di un sistema con *punto di fissazione* "vicino", la *disparità* è inversamente proporzionale alla distanza dal punto di fissazione

La *disparità* è anche legata al cosiddetto *effetto parallasse:* oggetti a distanza differente dall'osservatore, che si muovono con la stessa velocità, appaiono tanto più lenti quanto più sono lontani, proprio perché la disparità tra punti (in questo caso, tra le proiezioni di uno stesso punto in due istanti di tempo separati) è inversamente proporzionale alla distanza

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - La stessa relazione fra *disparità* e *profondità* può essere ricavata utilizzando le equazioni del flusso ottico
  - *Hp:* assi ottici paralleli $\Rightarrow$ la relazione fra i due strumenti di acquisizione d'immagine è una traslazione lungo l'asse *x* pari al valore della *linea di base (baseline)*
  - Le *disparità orizzontale* e *verticale* sono definite come

$$H = v_x \Delta t,\ V = v_y \Delta t$$

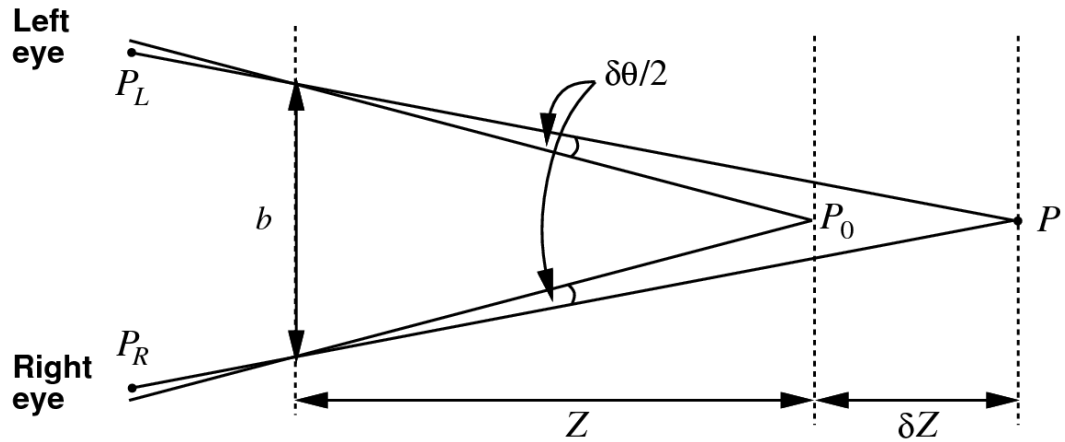
Sostituendo nelle precedenti le *equazioni del flusso ottico* e ponendo $T_x = b/\Delta t,\ T_y = T_z = 0$ e $\omega_x = \omega_y = \omega_z = 0$ (cioè parametri di rotazione nulli) si ottiene

Disparità Orizzontale $H = b/Z$ $\qquad$ $V = 0$ Disparità Verticale

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - Calcoliamo la *disparità angolare* in *radianti*
  - In condizioni normali di visione, gli esseri umani *fissano* $\Rightarrow$ gli assi ottici dei due occhi si intersecano in un punto della scena

Relazione tra *disparità angolare* e *profondità* nella stereoscopia

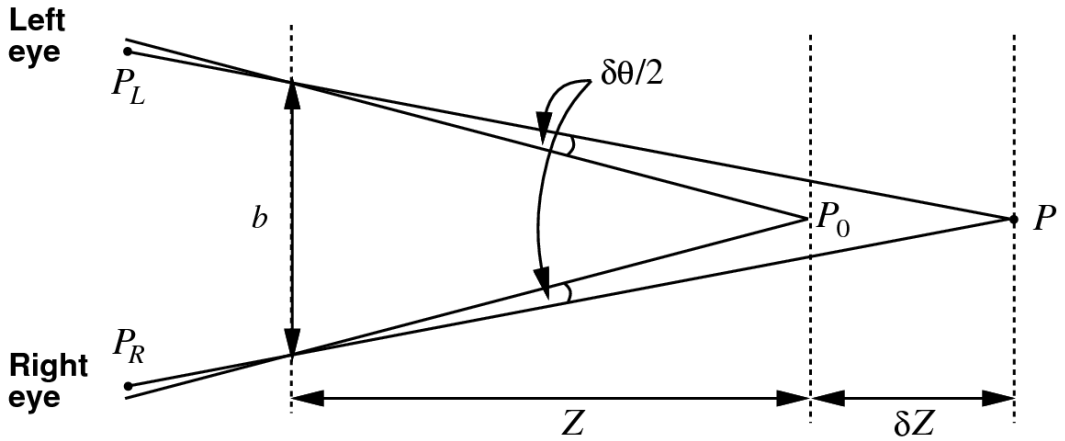


  - In figura: due occhi fissi su un punto P_0 a distanza Z dal punto mediano tra gli occhi
  - La *disparità* nel punto P_0 è nulla

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*

Relazione tra *disparità angolare* e *profondità* nella stereoscopia



- Consideriamo un altro punto *P* spostato di δZ e calcoliamo gli scostamenti angolari delle sue immagini dx e sx; se entrambe si discostano da P_0 di un angolo $\delta\theta/2 \Rightarrow$ lo scostamento fra esse, che è pari alla *disparita* di *P*, è $\delta\theta$

- Applicando la geometria elementare si ha $\dfrac{\delta\theta}{\delta Z} = \dfrac{-b}{Z^2}$

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*

$$\frac{\delta\theta}{\delta Z} = \frac{-b}{Z^2}$$

Fisiologia: per Z=1m il più piccolo $\delta\theta$ osservabile (≈1 pixel) è

$$\delta\theta \approx (5'')^\circ = 2{,}42 \times 10^{-5} rad$$

la linea di base è $b \approx 0{,}06m$

In tali condizioni risulta per $Z = 30cm \Rightarrow \delta Z \approx 0{,}036mm$

Linea di base *b* maggiore ⇒ Risoluzione maggiore

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching***
    - E' un'area di ricerca enormemente attiva
    - Non esistono soluzioni "assolute"
    - Fattori di complicanza

      - *Occlusioni:* punti che appaiono in una vista possono essere nascosti nell'altra

      - *Clipping:* punti che appaiono in una vista possono essere "fuori" dall'altra

      - *Variazioni di aspetto:* se si lavora con materiali molto *non-lambertiani* (quasi speculari) l'aspetto del medesimo punto geometrico può variare molto a seconda del punto di vista

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching***

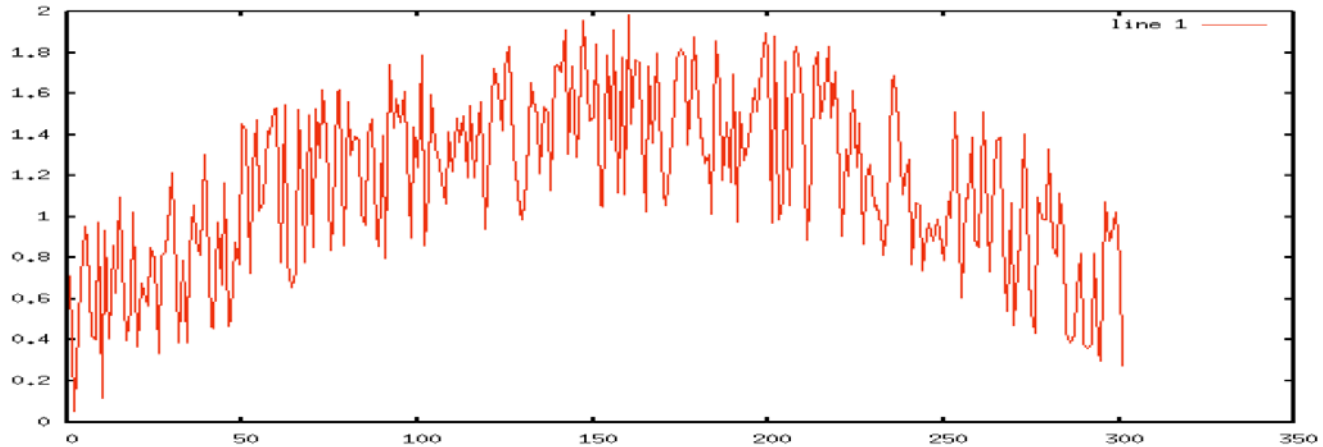
Idea base: individuare una caratteristica in una vista e cercarne la corrispondente nell'altra. A seconda delle caratteristiche ricercate si hanno

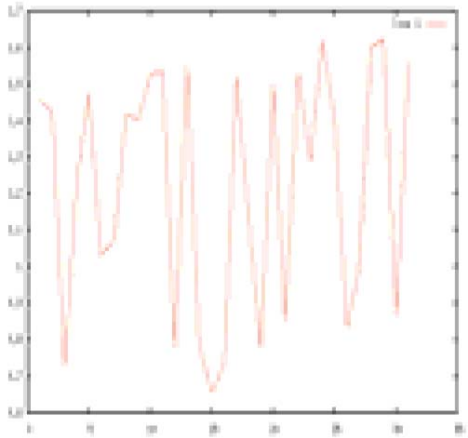
• Metodi basati sulla *correlazione:* operano nello spazio dei pixel dell'immagine; si cercano coppie di "intorni" che presentino caratteristiche simili di *luminanza, tessitura,* e *colore*

• Metodi basati su *feature:* operano nello spazio degli elementi geometrici *(lati, angoli, punti notevoli)* estratti dalla scena con opportuni operatori di basso livello; alla estrazione delle feature di ciascuna delle due scene segue il tentativo di stabilire una corrispondenza fra le feature estratte

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Metodi basati sulla correlazione*





Nel caso 1D la *correlazione* è l'operazione che consente di individuare nel grafico più lungo (a sinistra) l'intervallo dal quale è stato estratto il frammento di grafico a destra

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Metodi basati sulla correlazione*

Nel caso 1D, la *correlazione corr(d)* tra il vettore *B* e la finestra di *A* di *offset d* (i.e., la sottosequenza di *A* di lunghezza pari a quella di *B* che inizia alla posizione *d)* è data da

$$corr(d) = \sum_{i=0}^{size(B)} f(A[d+i], B[i])$$

dove *f* è una funzione opportunamente scelta.
Scelte comuni per *f* sono

- *f(x,y)=xy (correlazione incrociata* o *cross-correlation)*
- *f(x,y)=-(x-y)^2 (block matching)*

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Metodi basati sulla correlazione*

La funzione $f(x,y)=xy$ si rivela una buona scelta solo nel caso in cui il valore medio del "segmento" che vogliamo localizzare e il valore medio (su un segmento di analoga lunghezza) dell'intervallo di funzione su cui cercare sono all'incirca uguali e costanti (in pratica se A e B, nella formula precedente, hanno lo stesso valore medio).
Purtroppo tale ipotesi non è frequente

La funzione $f(x,y)=-(x-y)^2$, al contrario, fornisce buoni risultati anche in condizioni più generali

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Metodi basati sulla correlazione*

Nel caso di vettori 2D l'offset *d* è sdoppiato in due componenti (d_x e d_y) e la sommatoria è effettuata al variare di queste due. Ponendo per semplicità $A \in \Re^{MxM}$ e $B \in \Re^{NxN}$ si ha

$$corr(d_x, d_y) = \sum_{i=0}^{N} \sum_{j=0}^{N} f(A[d_x + i, d_y + j], B[i, j])$$

In questo caso, possiamo immaginare che *A* sia un'immagine grande in cui cerchiamo la posizione della sottoimmagine più simile a *B.*
Otterremo il valore massimo, al variare di d_x e d_y tra 0 e *M-N*, in corrispondenza della posizione ottimale

Estrazione di informazione 3D da un'immagine

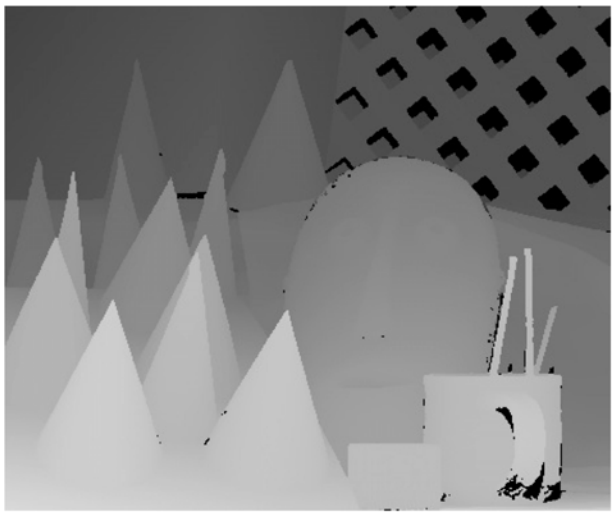
- *Stereoscopia binoculare*
  - ***Matching:*** *Metodi basati sulla correlazione*

Tramite la *correlazione*, le immagini risultano "dense" di corrispondenze $\Rightarrow$ è possibile calcolare la *disparità* praticamente in ogni punto, ottenendo una *Mappa di Disparità (Disparity Map)* densa di valori.
Se interpretiamo tale mappa come un'immagine in scala di grigi in base alla *disparità*, vedremo le zone più vicine alla camera più chiare e gli oggetti di una tonalità sempre più scura all'aumentare della distanza

Estrazione di informazione 3D da un’immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Metodi basati sulla correlazione*

Vista SX



Mappa di Disparità

Vista DX

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Metodi basati sulla correlazione*

Problemi

• Alto costo computazionale: il calcolo è troppo costoso per cercare la *correlazione* di una immagine (anche di dimensioni ridotte) all'interno di tutto lo spazio di una immagine grande

• Ambiguità: se l'immagine che si vuole correlare è piccola e troppo non "specializzata", i match significativi possono essere più di uno

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Metodi basati sulla correlazione*

Soluzioni

• non calcolare la *correlazione* sull'intera immagine, ma solo per piccoli valori di disparità

• cercare la *correlazione* tra finestre delle due immagini solo lungo la *retta epipolare* (vedi *Geometria Epipolare)* [soluzione migliore]

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Metodi basati su feature*

Si opera in un dominio differente, quello delle *caratteristiche* di un'immagine

Queste caratteristiche possono essere *bordi, angoli, forme geometriche, proprietà statistiche, ...*

L'algoritmo di rilevazione di una *feature* dipende dal tipo di *feature*

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Metodi basati su feature*

Esempio di *feature:* segmenti; per distinguere un segmento dagli altri si possono considerare

- Coordinate del punto medio $M=(m_x,m_y)$
- Lunghezza l
- Angolo di orientazione θ
- Stima del contrasto medio C lungo il segmento

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Metodi basati su feature*

Procedura

1. Si danno le immagini delle camere destra e sinistra in pasto a un *feature detector,* ossia a un algoritmo in grado di rilevare le *feature* e tutti i relativi parametri
2. Si cercano le corrispondenze tra i parametri (serve una metrica per misurarne la *similarità*)

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Metodi basati su feature*

Una metrica generica potrebbe essere

$$S = \frac{1}{w_0(l_l - l_r)^2 + w_1(\theta_l - \theta_r)^2 + w_2(M_l - M_r)^2 + w_3(C_l - C_r)^2}$$

dove w_i è il *peso* che si intende assegnare a ciascun parametro

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Metodi basati su feature*

Una *feature* particolarmente utilizzata in Computer Vision negli ultimi anni è la *Scale-Invariant Feature Transform (SIFT).*

In realtà, la *SIFT* è un algoritmo/tecnica proposta nel 2004 da David Lowe per l'estrazione di *feature (point-based)* da immagini.

Le *feature* estratte risultano essere

- invarianti ai cambiamenti di scala dell'immagine
- invarianti alle rotazioni
- invarianti alle traslazioni
- parzialmente invarianti ai cambiamenti del punto di vista
- parzialmente invarianti alle variazioni di illuminazione

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Correlazione vs Feature*

Correlazione

- Pregi
  - facile da implementare e *debuggare*
  - fornisce mappe "dense" di disparità
  - agevola la ricostruzione di superfici
- Difetti
  - computazionalmente costoso
  - sensibile alle distorsioni prospettiche
  - non robusto alla presenza di tessiture marcate

Estrazione di informazione 3D da un'immagine

- *Stereoscopia binoculare*
  - ***Matching:*** *Correlazione vs Feature*

Feature

- Pregi
  - robuste alle variazioni di illuminazione
  - computazionalmente efficienti
  - adatte alla navigazione in ambienti 3D (specie se ben strutturati)
- Difetti
  - non forniscono mappe dense
  - perdono dettagli
  - non adatte a tutte le applicazioni
  - più sensibili alle occlusioni

Estrazione di informazione 3D da un'immagine

- *Gradiente di texture*

  - *Texture:* un pattern ripetuto che può essere percepito visivamente (nella *scena*, talvolta la disposizione è perfettamente periodica, talvolta la regolarità è solo statistica)

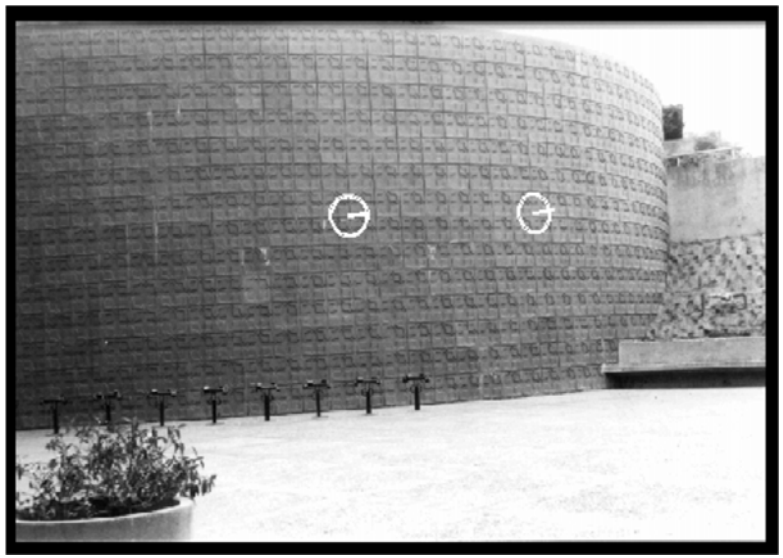
  - Nell'*immagine* le dimensioni apparenti, la forma, la disposizione spaziale, etc., degli elementi di texture *(texel)* variano per due ragioni
    - differenze nella distanza dei *texel* dalla telecamera
    - differenze nell'angolo visuale

Estrazione di informazione 3D da un'immagine

- *Gradiente di texture*
  - E' possibile calcolare analiticamente le espressioni del cambiamento di alcune caratteristiche dei *texel* nell'immagine, come *area*, *rimpicciolimento dovuto all'angolo di visione* e *densità*
  - Questi *gradienti di texture* sono funzioni della *forma* di una superficie e anche della sua *orientazione* (*slant* e *tilt*) rispetto all'osservatore
  - Processo in due passi per ricostruire la *forma* delle texture:
    1. si misurano i *gradienti di texture*
    2. si stimano la *forma*, lo *slant* e il *tilt* della superficie che possono dar luogo ai gradienti misurati

Estrazione di informazione 3D da un'immagine

- *Gradiente di texture*



Idea: presumendo che la texture reale sia uniforme, è possibile ricostruire la *forma* e l'*orientazione* della superficie

Estrazione di informazione 3D da un'immagine

- *Ombreggiatura*

  - Definizione: variazione di *intensità luminosa* su una superficie nella scena 3D dovuta alla *geometria* della scena stessa e alle proprietà di *riflettanza* delle superfici

  - Problema non risolvibile, se non sotto hp semplificative
    - superficie *lambertiana* (ossia *a diffusione perfetta*)
    - fonte di luce *puntiforme distante* (vale la *Proiezione Ortografica Scalata*)

    In tal caso la *luminosità* è data da

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

    con k fattore di scala, $\mathbf{n}$ versore normale alla superficie e $\mathbf{s}$ versore della fonte di luce (con k e $\mathbf{s}$ noti, si recupera $\mathbf{n}(x,y)$, da cui si deduce la forma della superficie)

Estrazione di informazione 3D da un'immagine

- *Ombreggiatura*
  - Sotto queste ipotesi è possibile ottenere un'equazione che dà **n** in funzione delle derivate parziali Z_x e Z_y della profondità $Z(x,y) \Rightarrow$ è possibile ottenere $Z(x,y)$
  - Generalizzazione: l'approccio precedente può essere applicato anche quando la superficie non è *lambertiana*, e la fonte di luce non *puntiforme*, purché sia però possibile calcolare la *Mappa di Riflettanza* $R(\mathbf{n})$, che specifica la *luminosità* di una porzione di superficie in funzione della sua *normale* ***n***
  - Problema: *riflessi interni*, cioè effetti di illuminazione reciproca $\Rightarrow$ $R(\mathbf{n})$ fallisce, in quanto la luminosità non dipende più solo dalla normale alla superficie, ma anche dalle complesse relazioni spaziali fra i diversi oggetti nella scena che diventano a tutti gli effetti *fonti secondarie*

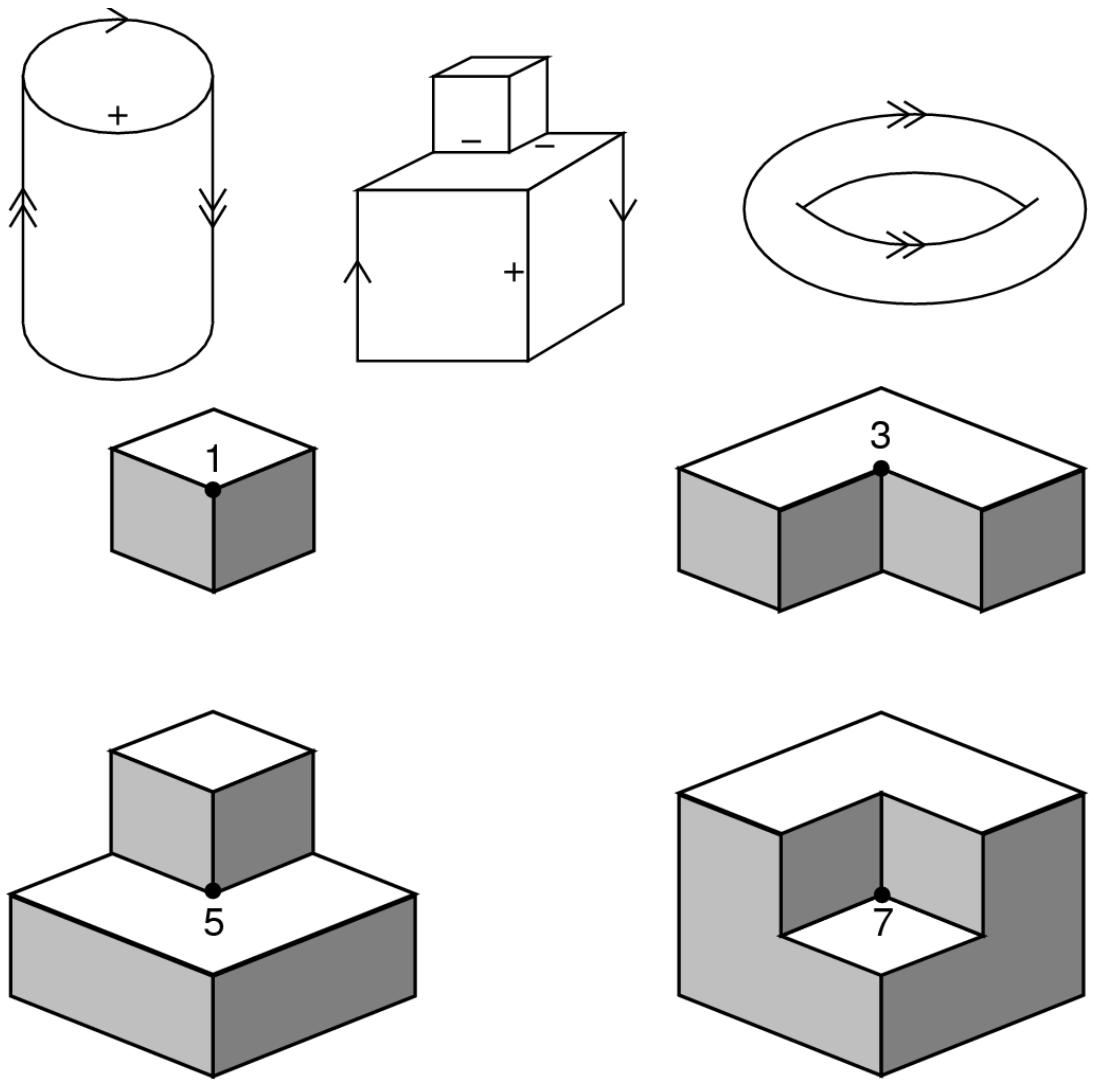
Estrazione di informazione 3D da un'immagine

- *Contorni*
  - *Etichettatura di linee:* valutazione dell'effettivo significato di ogni linea dell'immagine (uno dei primi argomenti di studio della CV) e determinazione degli assegnamenti. Inizialmente ci limitiamo a un *modello semplificato* (nessun segno sulle superfici degli oggetti, linee dovute a *discontinuità di illuminazione* rimosse), il che consente di considerare solo discontinuità di *profondità* o di *orientazione* $\Rightarrow$ ogni linea può quindi essere classificata come la proiezione di
    - *profilo (limb):* luogo dei punti di una superficie in cui la linea di visione è tangente alla superficie stessa
    - *bordo (edge):* discontinuità della normale alla superficie (può essere *convesso*, *concavo*, *occlusivo*)

Estrazione di informazione 3D da un'immagine

- *Contorni*
  - Etichette
    - + : bordo convesso
    - - : bordo concavo
    - frecce singole : bordo occlusivo convesso
    - doppie frecce: profilo

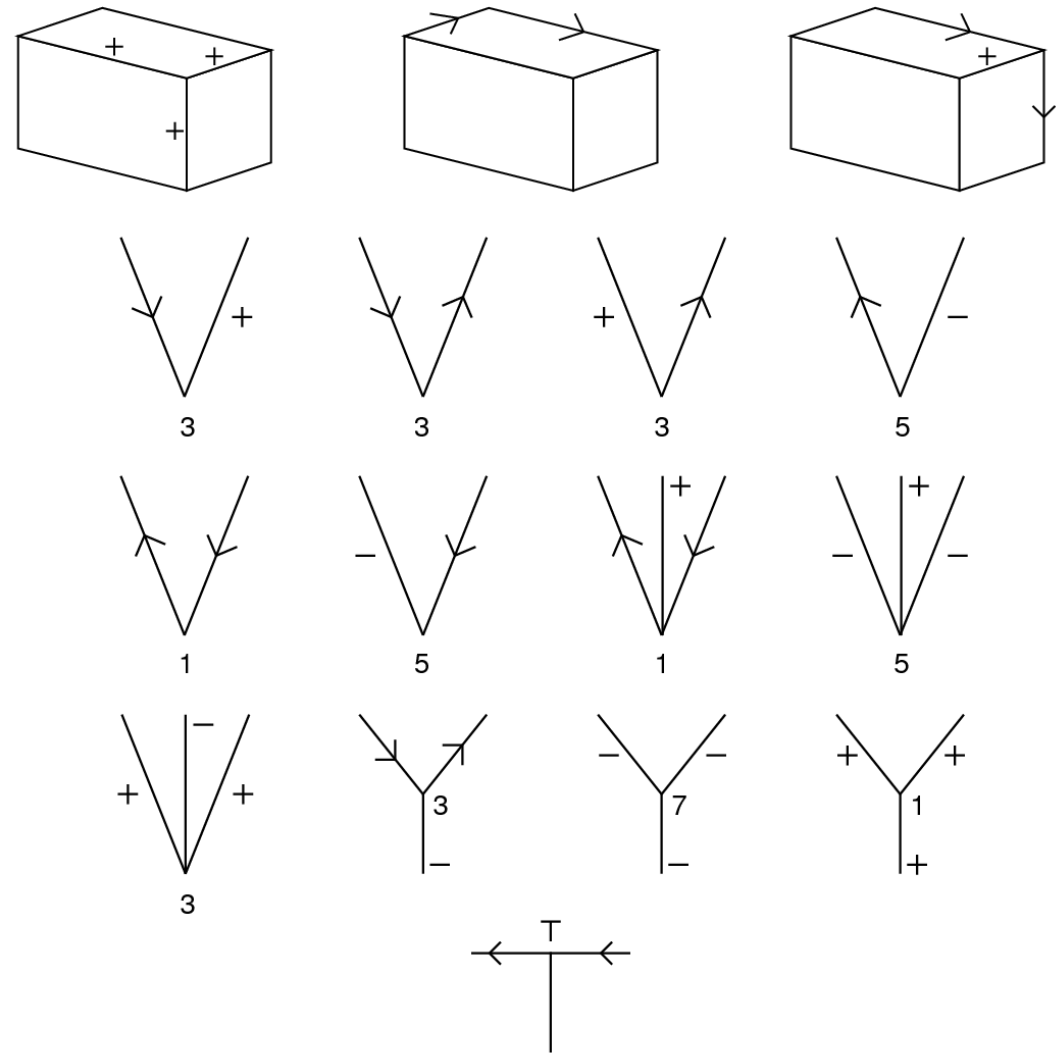
    Sono possibili 6^n assegnamenti a *n* linee
  - Huffman (1971) e Clowes (1971) hanno proposto indipendentemente un primo approccio sistematico all'analisi di scene poliedriche (sotto hp semplificative: solidi ***opachi poliedrici con vertici triedrici***, no rotture *(cracks)*, no variazioni di etichetta su una stessa linea)



4 modi in cui 3 superfici piane possono incontrarsi in un vertice (il numero è relativo agli ottanti)

Estrazione di informazione 3D da un'immagine

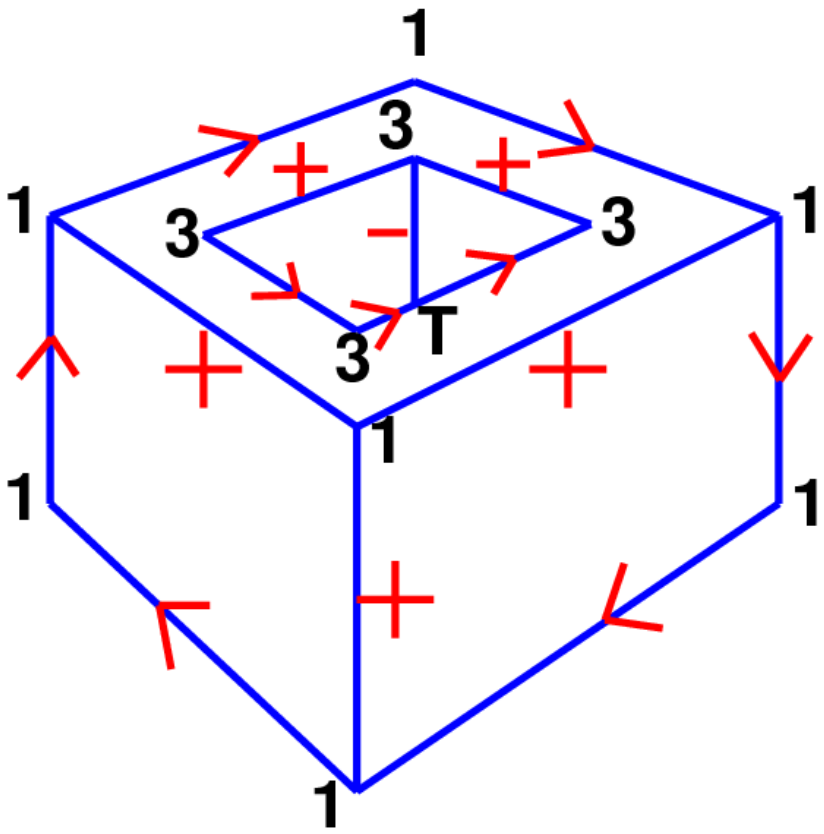
- *Contorni*
  - L'insieme delle etichette di Huffman-Clowes
  - Waltz (1975) ha proposto un algoritmo per risolvere questo problema (in una versione più complessa): una delle prime applicazioni *CSP* in IA
    - variabili: giunzioni
    - valori: etichette
    - vincoli: una linea = una etichetta

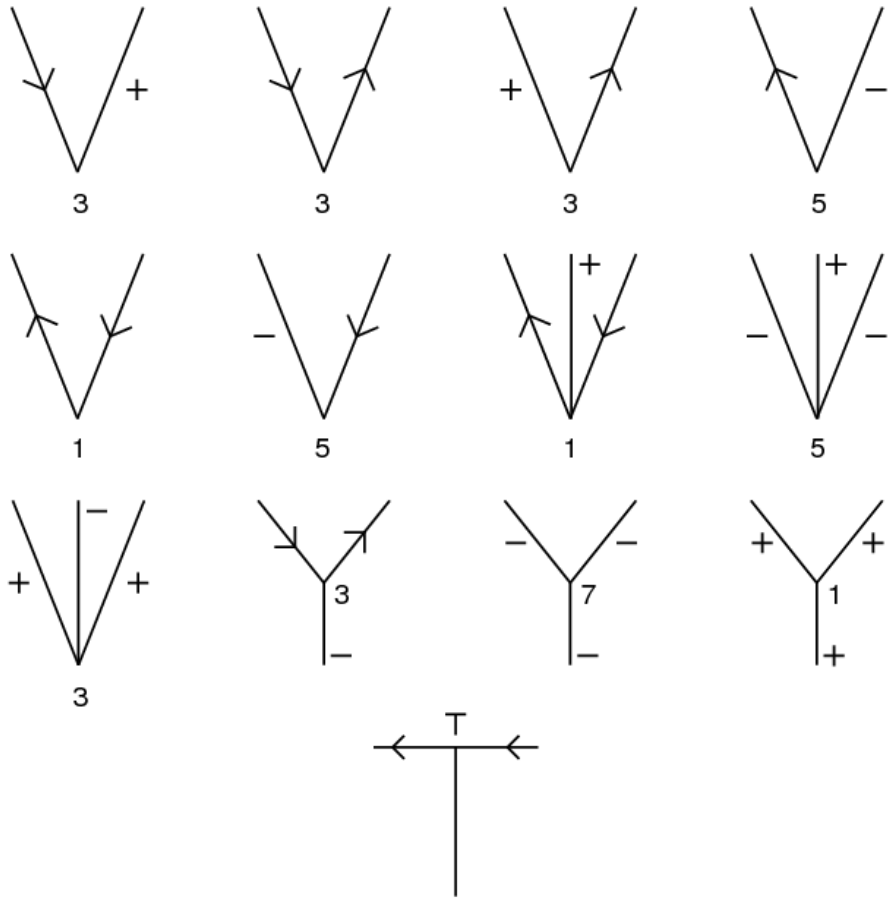


Diversi modi in cui si può osservare un vertice: *L, Y, freccia, T*

Estrazione di informazione 3D da un'immagine

- *Contorni*





Estrazione di informazione 3D da un'immagine

- *Contorni*