Stereo vision
(Visione stereoscopica)

- Geometria della visione binoculare
  - disparità
  - vincolo epipolare: matrice essenziale e fondamentale
  - image rectification
- Ricostruzione
- Tecniche di matching
  - Approccio locale: area-based, edge-based
  - Approccio globale: graph cut

(Forsyth/Ponce: Capitolo 7)

Slide credits:
materiale rielaborato a partire da slide di Luigi Cinque (Univ. La Sapienza, Roma) e altre sorgenti (citate)

Visione stereoscopica (stereo vision)

Stereo vision: definizione del problema

Data una coppia di immagini stereo calibrate (***stereo pair***)

➔ ottenere l'informazione di profondità di ogni punto osservato (***depth map***)

left image

right image

Depth map

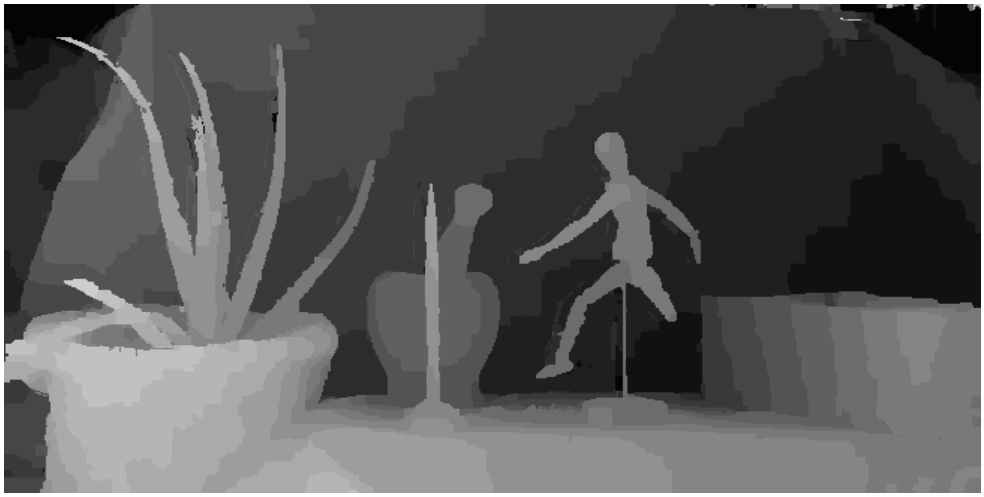
Application: View Interpolation

immagine sinistra

Application: View Interpolation

immagine destra

Application: View Interpolation



disparità (→ distanza?)

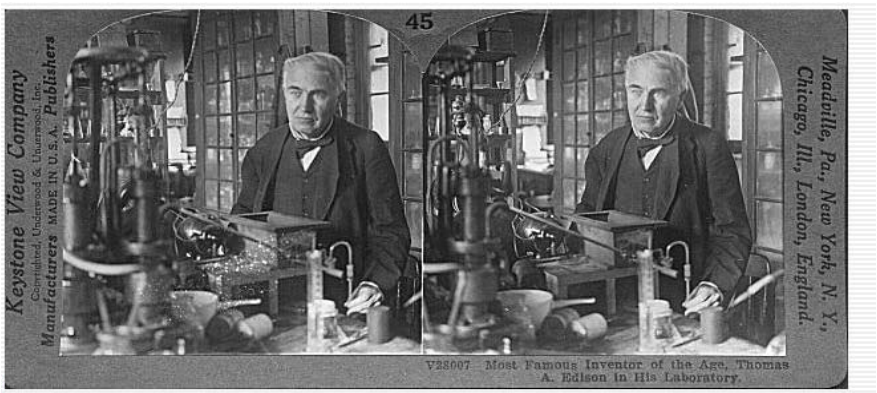
Stereopsi umana

La nostra "stereo vision": **stereopsis**

data una coppia di immagini stereo calibrate (***stereo pair***)

➔ ottenere l'informazione di profondità di ogni punto osservato (***depth map***)

- Il **sistema visivo umano** lo sa fare: **stereopsi**



Stereogrammi
(Sir Charles Wheatstone, 1838)

Stereopsi umana

L'occhio, muovendosi, annulla la disparità su ciò che fissa (immagine → fovea)

- Cerchio di **Vieth-Müller**: luogo dei punti a **zero disparità → stessa distanza**

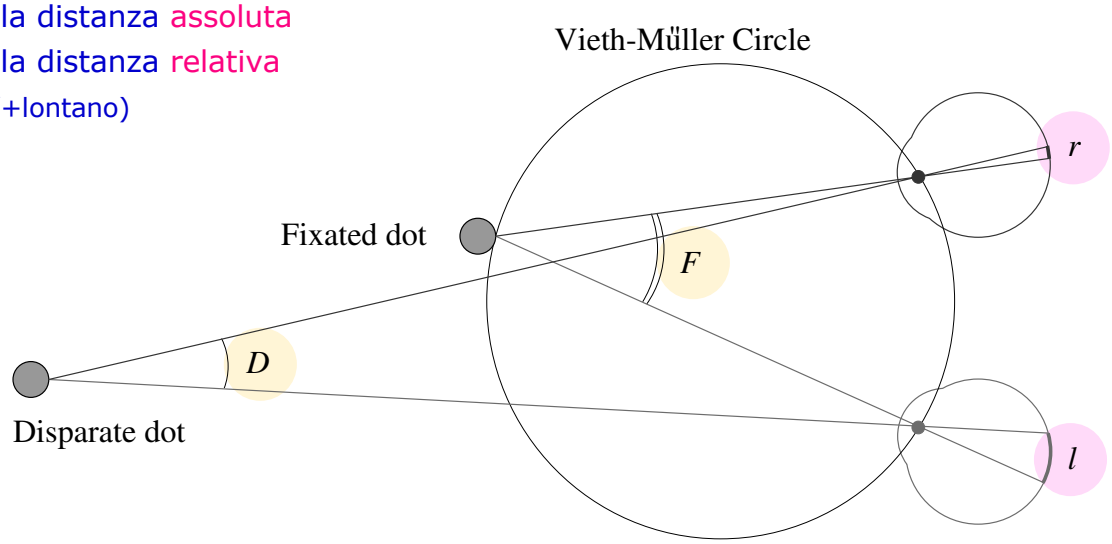
$$disparity: \quad d = r - l = D - F$$

- Tali angoli non vengono misurati accuratamente in valore assoluto, ma le differenze tra disparità diverse osservate sì (Helmholtz, 1909)

➔ stima male la distanza assoluta

➔ stima bene la distanza relativa

  - (+vicino/+lontano)



Stereo vision (visione binoculare): problema generale

Stereo vision: data una coppia di immagini stereo **calibrate** (*stereo pair*)

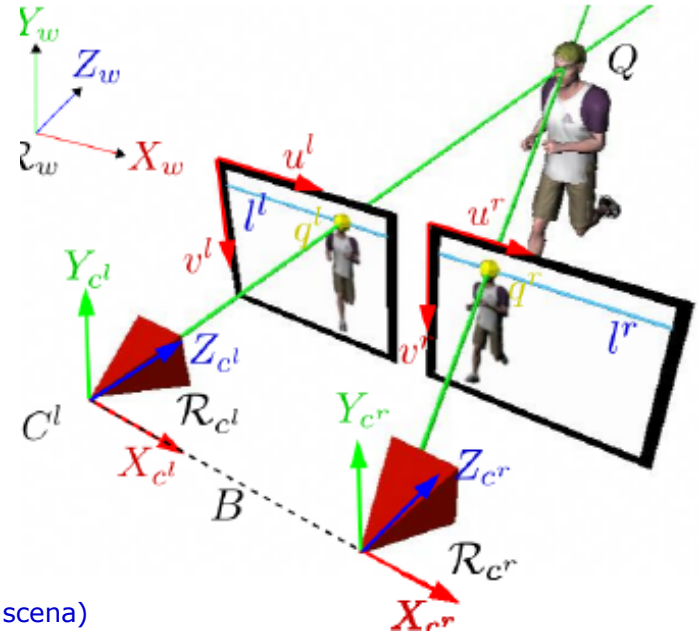
➔ ottenere l'informazione di profondità di ogni punto osservato (***depth map***)

Problemi da risolvere:

- **Calibrazione**:
  Date 2 camere calibrate: C^l, C^r

- **Corrispondenza (matching)**:
  Scelto un punto q^l in un'immagine,
  trovare il corrispondente q^r nell'altra

- **Ricostruzione (triangolazione)**:
  data la coppia di punti (q^l , q^r),
  trovare la posizione di Q (*pre-image*)



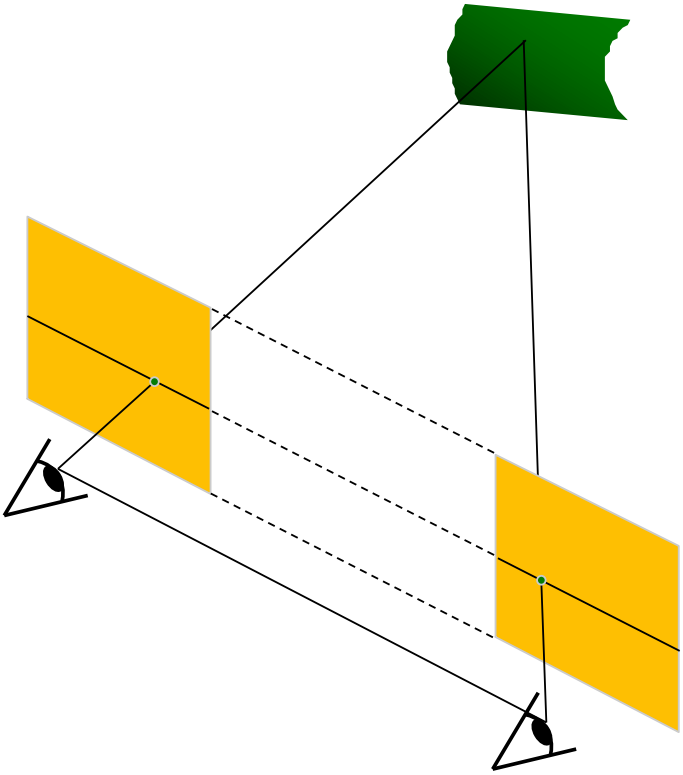
Problema alternativo:

Camera geometry / camera motion:

data le coppie di punti (q^l , q^r) (per più punti di scena)
➔ trovare la posizione delle camere: C^l, C^r

© Javier Ibanez-Guzman

Visione binoculare: immagini parallele



Caso più semplice:

immagini parallele

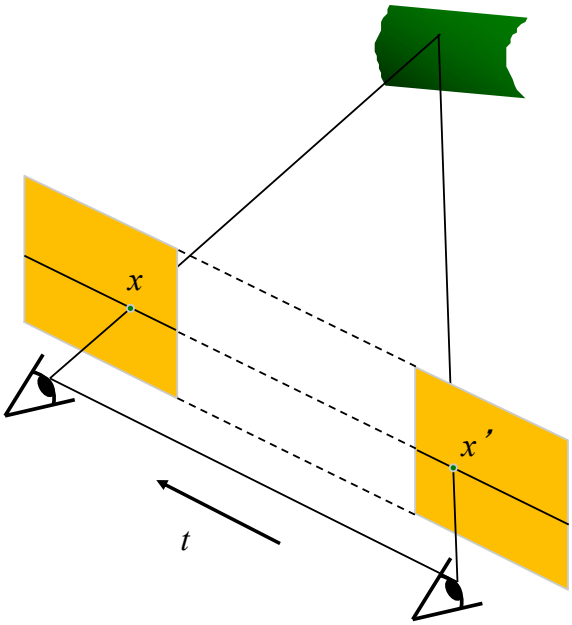
- Piani immagine delle camere paralleli tra loro e alla "baseline" **OO'**
- Centri ottici (punti principali) alla stessa altezza nelle 2 immagini
- Stessa lunghezza focale f

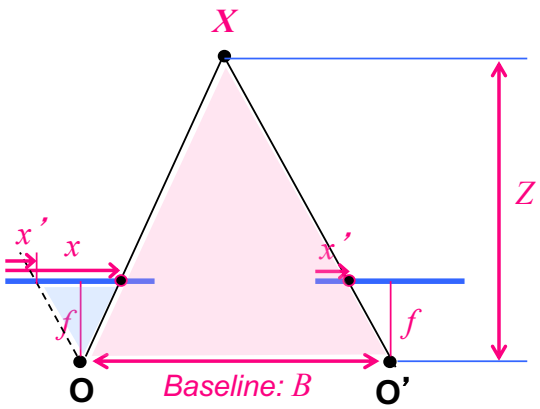
⇩

- piani immagine coplanari

Dalla disparità alla distanza – caso più semplice

Caso più semplice: **immagini parallele**





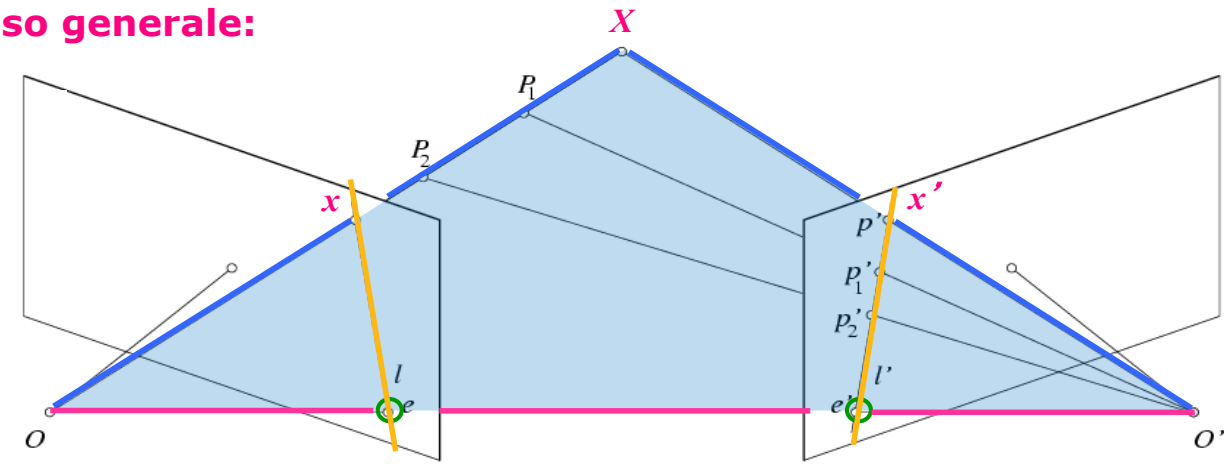
$$disparity: x - x'$$

$$\frac{x - x'}{f} = \frac{B}{Z} \rightarrow \boxed{Z = \frac{Bf}{x - x'}}$$

➔ **disparità** inversamente proporzionale alla **distanza *Z***

Geometria binoculare – caso generale

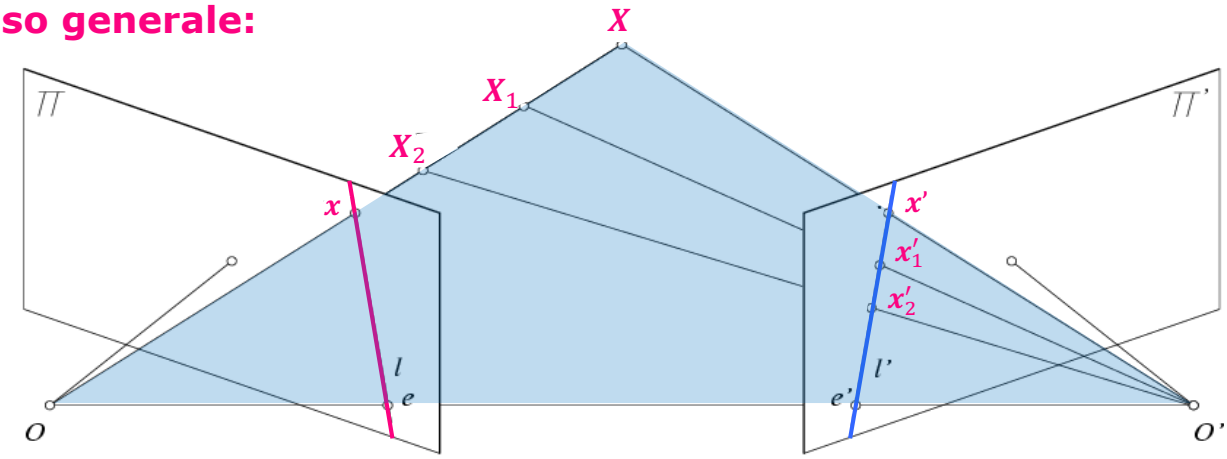
Caso generale:



- **Baseline** – la linea che passa per i due centri ottici **O** e **O'**
- **Piano epipolare** – piano che contiene la baseline e il punto ***X*** (fascio)
- **Epipoli:** punti di intersezione della baseline con i piani immagine:
  ➔ proiezioni di ogni centro ottico sull'altra immagine
- **Linee epipolari:** intersezioni del piano epipolare con i due piani immagine (esistono sempre in coppie corrispondenti)

Geometria binoculare – vincolo epipolare

Caso generale:

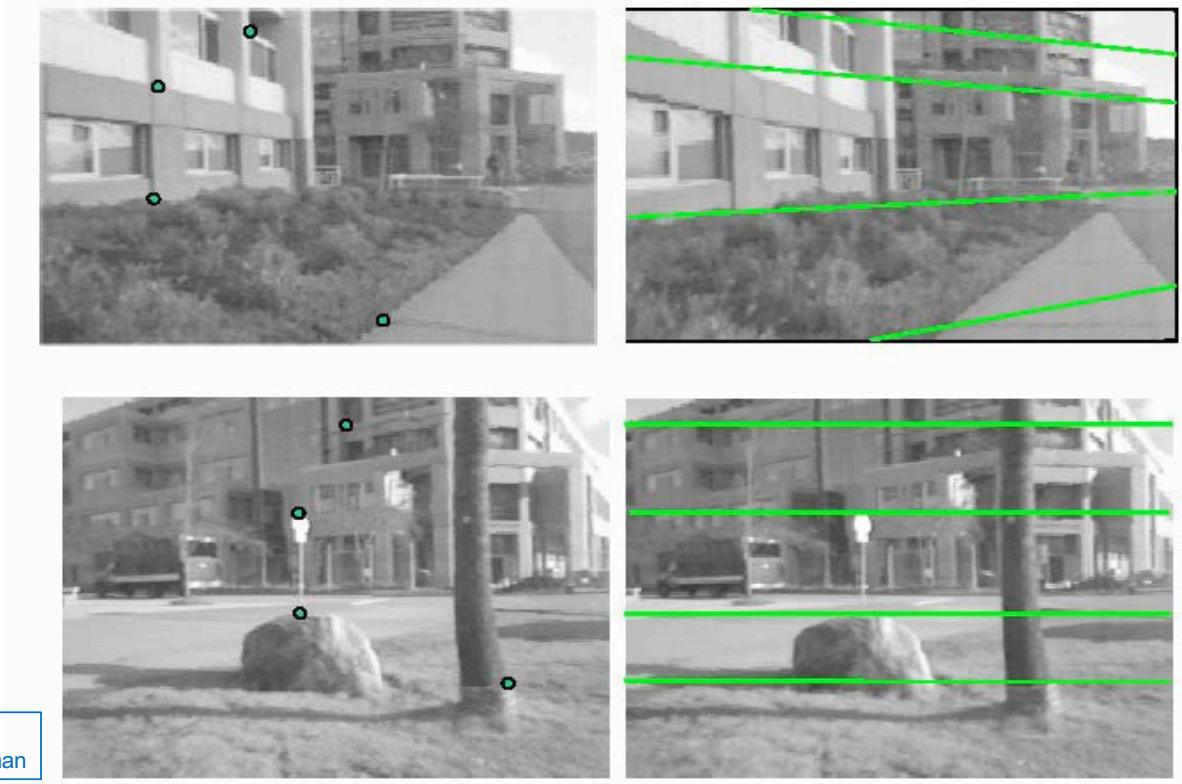


Vincolo epipolare:

in un sistema binoculare i due centri ottici O e O', il punto osservato X e le sue immagini x e x', giacciono sullo stesso piano, detto **piano epipolare**

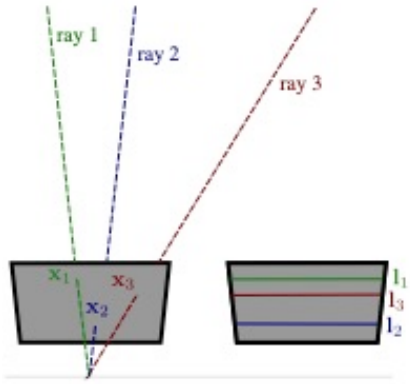
- il punto x', corrispondente di x in Π', deve giacere sulla linea epipolare l'
- il punto x, corrispondente di x' in Π, deve giacere sulla linea epipolare l

Esempio di geometria epipolare

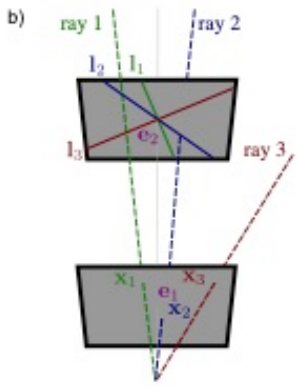


Source:
K. Grauman

Esempi di geometria epipolare

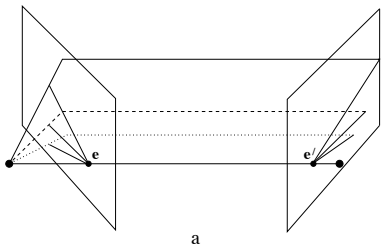


Coppia rettificata



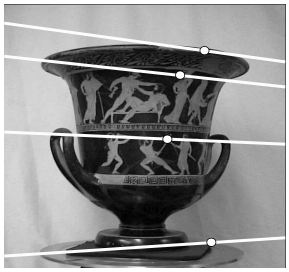
© Simon Prince (2012)

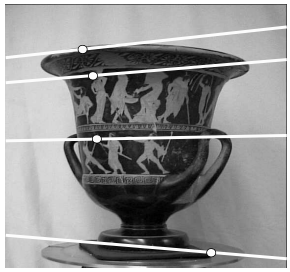
Coppia coassiale



a

Assi ottici convergenti





source:
Hartley, Zisserman

Coordinate normalizzate

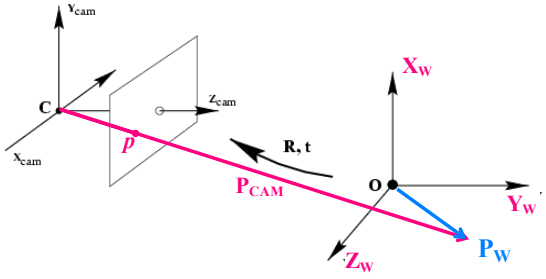
Coordinate-immagine e coordinate **normalizzate**
combinando tutte le equazioni del modello:

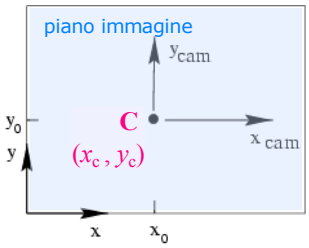
$$\tilde{\mathbf{P}}_{cam} = \begin{bmatrix} x_{cam} \\ y_{cam} \\ z_{cam} \\ 1 \end{bmatrix} = \begin{bmatrix} & \mathbf{R} & & \mathbf{t} \\ [0 & 0 & 0] & 1 \end{bmatrix} \tilde{\mathbf{P}}_{\mathrm{W}} = \mathbf{K}_E \, \tilde{\mathbf{P}}_{\mathrm{W}}$$

Matrice di ***calibrazione estrinseca*** [4x4]

$$\tilde{\mathbf{p}}_{im} = \begin{bmatrix} x_{im} \\ y_{im} \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & x_C & 0 \\ 0 & f & y_C & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tilde{\mathbf{P}}_{cam} = \mathbf{K}_I \, \tilde{\mathbf{P}}_{cam}$$

$\mathbf{K}_I$*: Matrice di* ***calibrazione intrinseca*** [3x4]





Coordinata normalizzata $\mathbf{p}_n$:

$\mathbf{p}_n$: $\mathbf{P}_{cam}$ diviso per la sua distanza $\boldsymbol{z}_{cam}$

→ proiezione prospettica normalizzata rispetto a f ($f = 1$)

$$\mathbf{p}_n = \frac{1}{z_{cam}} \mathbf{P}_{cam} = \begin{bmatrix} x_n = x_{cam}/z_{cam} \\ y_n = y_{cam}/z_{cam} \\ 1 \end{bmatrix}$$

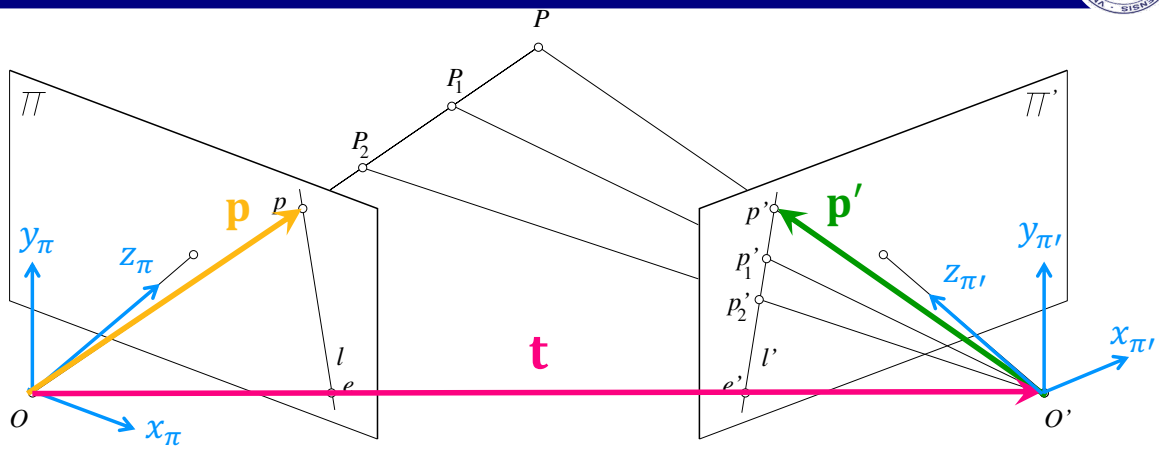
$$\tilde{\mathbf{p}}_{im} = \begin{bmatrix} x_{im} \\ y_{im} \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & x_C \\ 0 & f & y_C \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p}_n = \mathbf{K} \, \mathbf{p}_n$$

K: *Matrice di* ***calibrazione intrinseca*** [3x3]

$$\mathbf{p}_n = \frac{1}{z_{cam}} \mathbf{P}_{cam} = \begin{bmatrix} x_{cam}/z_{cam} \\ y_{cam}/z_{cam} \\ 1 \end{bmatrix}$$

$\mathbf{p}_n$: coordinata **normalizzata**
($\mathbf{P}_{cam}$ scalata: $f = 1$)

Vincolo epipolare in coordinate normalizzate



Op , ***O'p'*** e ***OO'*** sono **coplanari**
→ il **doppio prodotto misto** è **nullo**:

$$\overrightarrow{Op} \cdot \left(\overrightarrow{OO'} \times \overrightarrow{O'p'} \right) = 0$$

- *Op*, *O'p'* sono $\boldsymbol{p}$ e $\boldsymbol{p}'$ in coord. normalizzate:
- *OO'* è il vettore traslazione **t** (nel rif. di π):
- *O'p'* nel sistema di riferimento di π:

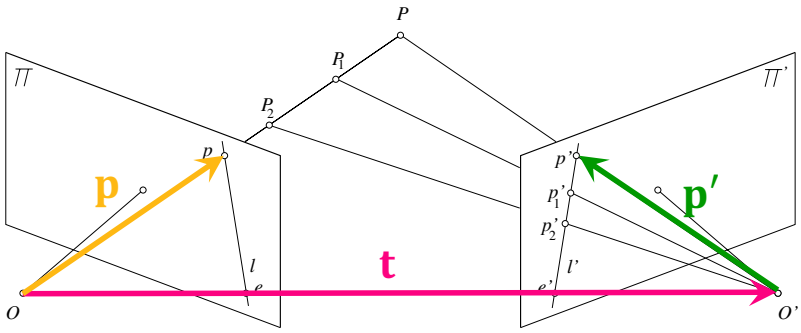
$$\overrightarrow{Op} = \mathbf{p} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}; \quad \mathbf{p}' = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}$$

$$\overrightarrow{O'p'} = \mathbf{R} \, \mathbf{p}' \; ; \quad \overrightarrow{OO'} = \mathbf{t}$$

Vincolo epipolare: $\mathbf{p} \cdot (\mathbf{t} \times \mathbf{R} \, \mathbf{p}') = 0$

Vincolo epipolare: matrice essenziale E

Vincolo epipolare in coordinate normalizzate (segue):



Posso riscrivere il doppio prodotto misto in forma matriciale:

$$\mathbf{p} \cdot (\mathbf{t} \times \mathbf{R}\,\mathbf{p}') = 0 \;\rightarrow\; \mathbf{p}^T(\mathbf{t}_\times\,\mathbf{R})\;\mathbf{p}' = 0$$

dove $\boldsymbol{t}_x$ è la matrice *"skew-symmetric"* di $\boldsymbol{t}$:

$$\mathbf{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \rightarrow \mathbf{t}_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \;\Rightarrow\; \mathbf{t} \times \mathbf{p} = \mathbf{t}_\times\,\mathbf{p}$$

Ottenendo:

$$\mathbf{p}^T\mathbf{E}\,\mathbf{p}' = 0\,; \quad \mathbf{E} = \mathbf{t}_\times\,\mathbf{R}$$

E: matrice **essenziale**
Longuet-Higgins (1981)

Vincolo epipolare: matrice essenziale E

Relazione tra **matrice Essenziale** e **parametri di camera**

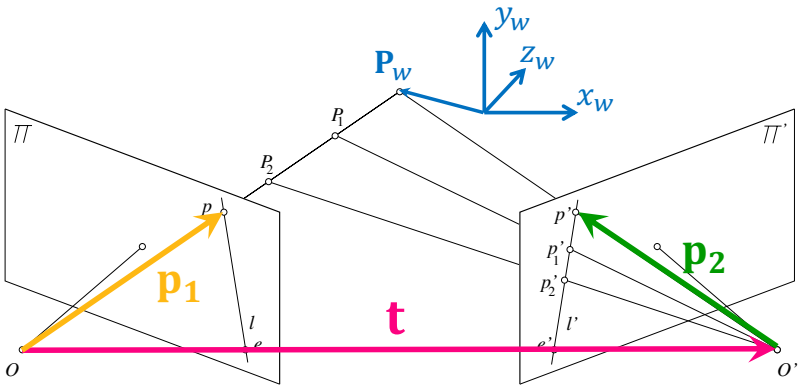
consideriamo un sistema binoculare:

- Camera 1 (riferimento)

$$\mathbf{P}_{cam1} = \mathbf{R}_1\mathbf{P}_W + \mathbf{T}_1$$

- Camera 2

$$\mathbf{P}_{cam2} = \mathbf{R}_2\mathbf{P}_w + \mathbf{T}_2$$
$$\rightarrow\; \mathbf{P}_w = \mathbf{R}_2^T\,(\mathbf{P}_{cam2} - \mathbf{T}_2)$$



$$\mathbf{p_1}\mathbf{E}\,\mathbf{p_2} = 0\,, \qquad \mathbf{E} = \mathbf{t}_\times \cdot \mathbf{R}\,, \qquad dove\colon\; \mathbf{R},\mathbf{t}:\; \mathbf{P}_{cam1} = \mathbf{R}\,\mathbf{P}_{cam2} + \mathbf{t}$$

$$\mathbf{p}_1 = \frac{\mathbf{P}_{cam1}}{Z_{cam1}} = \alpha_1\mathbf{P}_{cam1}\;; \quad \mathbf{p}_2 = \frac{\mathbf{P}_{cam2}}{Z_{cam2}} = \alpha_2\mathbf{P}_{cam2}$$

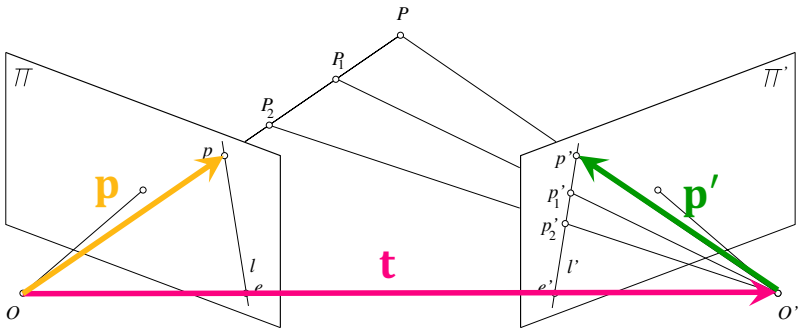
$$\mathbf{P}_{cam1} = \mathbf{R}_1\mathbf{P}_W + \mathbf{T}_1 = \mathbf{R}_1\mathbf{R}_2^T\,(\mathbf{P}_{cam2} - \mathbf{T}_2) + \mathbf{T}_1 = \mathbf{R}_1\mathbf{R}_2^T\,\mathbf{P}_{cam2} + \mathbf{T}_1 - \mathbf{R}_1\mathbf{R}_2^T\mathbf{T}_2$$

$$\Rightarrow \quad \mathbf{R} = \mathbf{R}_1\mathbf{R}_2^T\,, \qquad \mathbf{t} = \mathbf{T}_1 - \mathbf{R}_1\mathbf{R}_2^T\,\mathbf{T}_2 \quad \Rightarrow \quad \mathbf{E} = \mathbf{t}_\times \cdot \mathbf{R}$$

Vincolo epipolare: matrice essenziale E

Matrice essenziale E:

definisce in modo univoco la geometria binoculare di una coppia di immagini normalizzate ($f = 1$)



Proprietà di E:

- Definita a meno di un fattore di scala ($\det \mathbf{E} = 0$):
- **5 gradi** di libertà
  $3(\mathbf{R}) + 3(\mathbf{t}) - 1(|\mathbf{E}| = 0)$
- **SVD:** 2 valori singolari uguali e 1 nullo

$$\mathbf{p}^T \mathbf{E}\, \mathbf{p}' = 0 \quad \rightarrow \quad \begin{matrix} k\,\mathbf{E} \equiv \mathbf{E} \\ \det \mathbf{E} = \mathbf{0} \end{matrix}$$

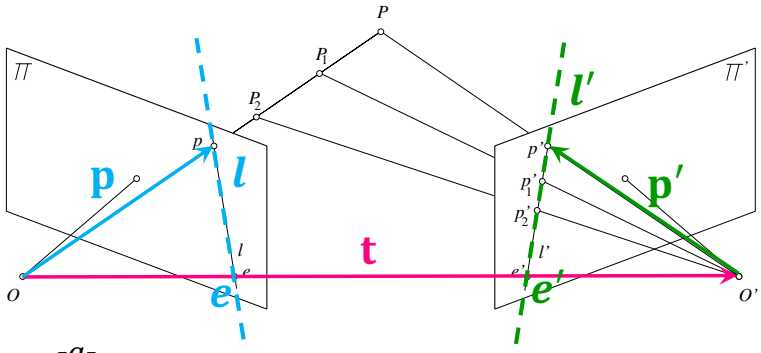
$$\mathbf{E} \equiv k\mathbf{E} = k(\mathbf{t}_\times \cdot \mathbf{R}), \qquad \forall k \neq 0$$

$$\mathrm{svd}(\mathbf{E}) = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^T = \mathbf{U} \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{V}^T$$

Vincolo epipolare: matrice essenziale E

Proprietà della matrice essenziale E:

$$\mathbf{p}^T \mathbf{E}\, \mathbf{p}' = 0$$



Linee epipolari

$$\mathbf{E}\,\mathbf{p}' = \boldsymbol{l} \;\rightarrow\; \mathbf{p}^T \cdot \boldsymbol{l} = 0 \;\rightarrow\; [x \quad y \quad 1] \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = ax + by + c = 0$$ *linea epipolare* $\boldsymbol{l}$

$$\mathbf{E}^T\mathbf{p} = \boldsymbol{l}' \;\rightarrow\; \mathbf{p}'^T \cdot \boldsymbol{l}' = 0 \;\rightarrow\; [x' \quad y' \quad 1] \cdot \begin{bmatrix} a' \\ b' \\ c' \end{bmatrix} = a'x' + b'y' + c' = 0$$ *linea epipolare* $\boldsymbol{l}'$

Epipoli

I vettori $\boldsymbol{e}$ ed $\boldsymbol{e}'$ giacciono su $\mathbf{t}$

- $\mathbf{O}\mathbf{O}' \times \mathbf{O}'\boldsymbol{e}' = 0$ (riferim: Π) $\qquad \mathbf{t} \times \mathbf{R}\boldsymbol{e}' = (\mathbf{t}_\times \mathbf{R})\boldsymbol{e}' = \mathbf{E}\,\boldsymbol{e}' = 0$
- $\boldsymbol{O}'\mathbf{O} \times \boldsymbol{O}'\boldsymbol{e} = 0$ (riferim: Π') $\qquad -\mathbf{t} \times \mathbf{R}^T\boldsymbol{e} = (\mathbf{t}_\times^T \mathbf{R}^T)\boldsymbol{e} = \mathbf{E}^T\boldsymbol{e} = 0$

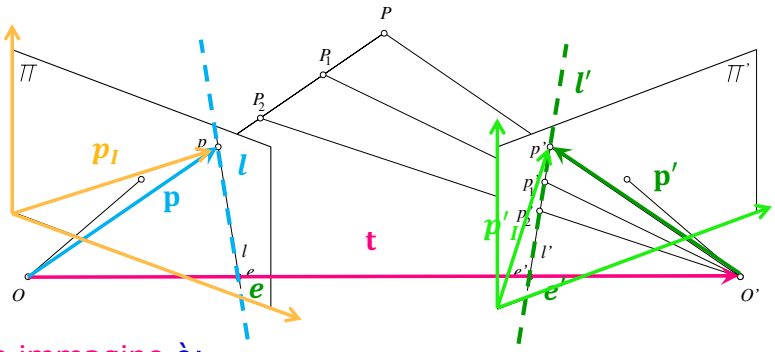
➔ epipoli $\boldsymbol{e}$, $\boldsymbol{e}'$: **spazi nulli** di $\mathbf{E}$, $\mathbf{E}^T$

Vincolo epipolare: matrice fondamentale F

La matrice essenziale **E** è definita sulle coordinate normalizzate: $\mathbf{p}$

Considero le **coordinate immagine:** $\mathbf{p}_I$

$$\mathbf{p} = \begin{bmatrix} x_{cam}/z_{cam} \\ y_{cam}/z_{cam} \\ 1 \end{bmatrix}; \quad \mathbf{p}_I = \begin{bmatrix} x_I \\ y_I \\ 1 \end{bmatrix}$$



- La relazione tra coordinate normalizzate e coordinate immagine è:

$$\mathbf{p}_I = \boldsymbol{K}\,\mathbf{p}\,, \qquad \boldsymbol{K} = \begin{bmatrix} f & 0 & c_X \\ 0 & f & c_Y \\ 0 & 0 & 1 \end{bmatrix} \quad \rightarrow \quad \mathbf{p} = \boldsymbol{K}^{-1}\,\mathbf{p}_I$$

K : matrice di calibrazione intrinseca

- Sostituisco $\mathbf{p}$ e $\mathbf{p}'$ nell'espressione del vincolo epipolare:

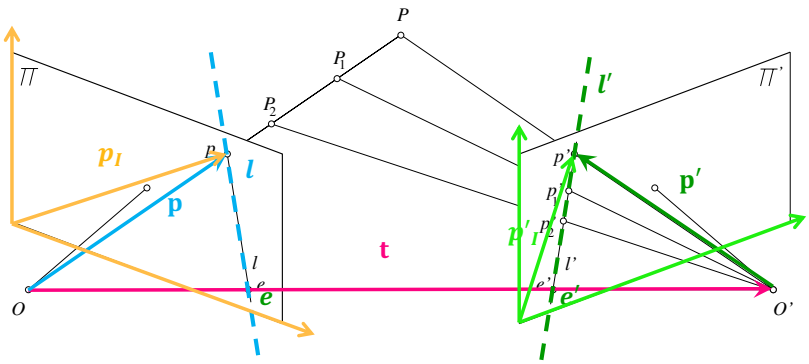
$$\mathbf{p}^T\,\mathbf{E}\,\mathbf{p}' = 0 \quad \rightarrow \quad \left(\boldsymbol{K}^{-1}\mathbf{p}_I\right)^T \mathbf{E}\left(\boldsymbol{K}'^{-1}\mathbf{p}'_I\right) = \mathbf{p}_I^T\left(\boldsymbol{K}^{-T}\mathbf{E}\,\boldsymbol{K}'^{-1}\right)\mathbf{p}'_I = 0$$

$$\rightarrow \quad \mathbf{p}_I^T\,\mathbf{F}\,\mathbf{p}'_I = 0\,, \qquad \mathbf{F} = \boldsymbol{K}^{-T}\mathbf{E}\boldsymbol{K}'^{-1}$$

F : matrice **fondamentale**
Faugeras and Luong (1992)

Vincolo epipolare: matrice fondamentale F

Matrice fondamentale F:
definisce in modo univoco la geometria binoculare di una coppia stereo, **inclusi i parametri intrinseci** delle camere



Proprietà di F:

- Definita a meno di un fattore di scala:
- **8 gradi** di libertà $(|\mathbf{F}| = 0)$

$$\mathbf{p}_I^T\mathbf{F}\,\mathbf{p}'_I = 0 \quad \rightarrow \quad \begin{matrix} k\,\mathbf{F} \equiv \mathbf{F} \\ \det \mathbf{F} = 0 \end{matrix}$$

- Linee epipolari: $\mathbf{F}\boldsymbol{p}_I'$ / $\mathbf{F}\boldsymbol{p}_I$

$$\begin{cases} \mathbf{F}\,\mathbf{p}'_I = \boldsymbol{l} & \rightarrow \quad \mathbf{p}_I \cdot \boldsymbol{l} = 0 \\ \mathbf{F}^T\mathbf{p}_I = \boldsymbol{l}' & \rightarrow \quad \mathbf{p}'_I \cdot \boldsymbol{l}' = 0 \end{cases}$$

- Epipoli $\boldsymbol{e}_I$ / $\boldsymbol{e}_I'$: spazi nulli di **F** / $\mathbf{F}^T$

$$\mathbf{F}\,\mathbf{e}'_I = 0\,; \quad \mathbf{F}^T\mathbf{e}_I = 0$$

Camera Calibration – lab exercise

Esercitazione MATLAB®:

– calcolo matrici essenziale e fondamentale

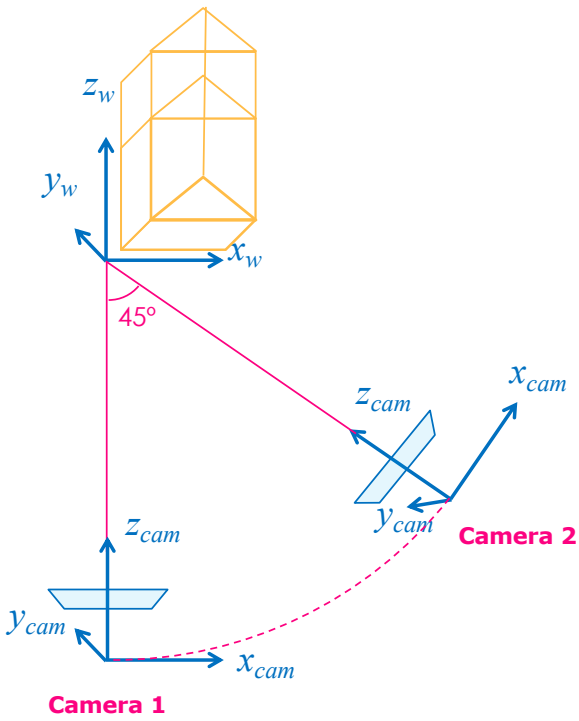
– verifica vincolo epipolare

- MATLAB script: **`Epipolar.m`**

```
>> edit Epipolar
```

Esercitazione:

- *esecuzione dello script*
- *ri-esecuzione, con parametri differenti*

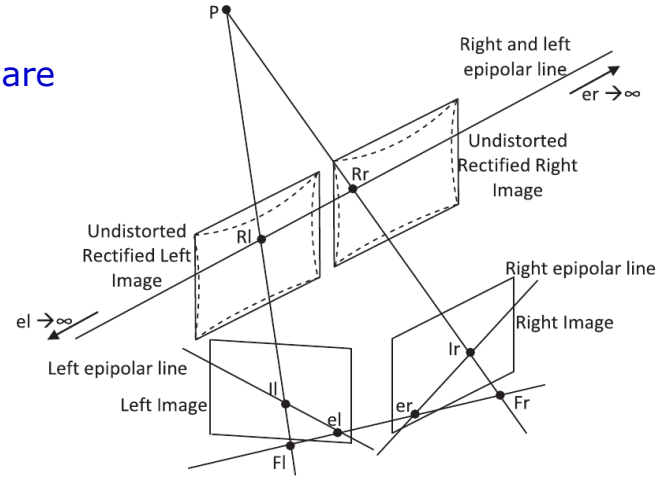


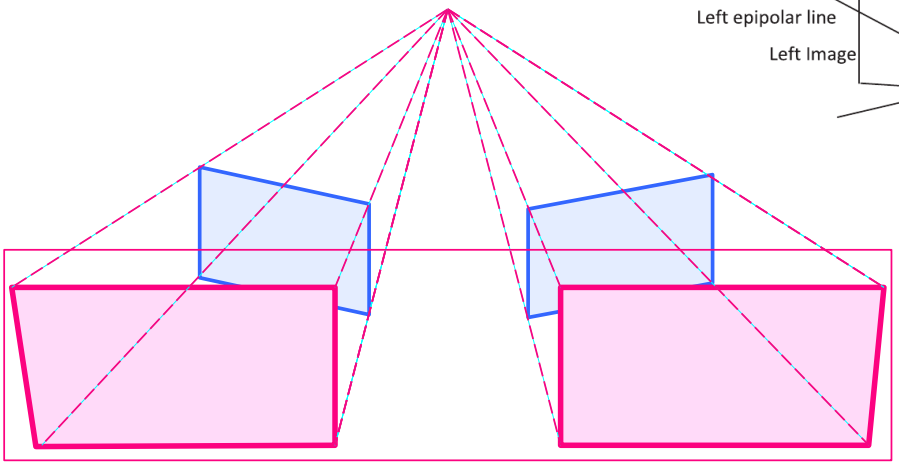
Visione binoculare: immagini parallele

Image rectification:

trasformazione di una geometria binoculare generica in una con immagini parallele

- piani immagine coplanari
- linee epipolari coincidono con linee orizzontali corrispondenti (scanlines)





Stereo image rectification

Image rectification

- I piani immagine vengono riproiettati su un piano comune parallelo alla baseline

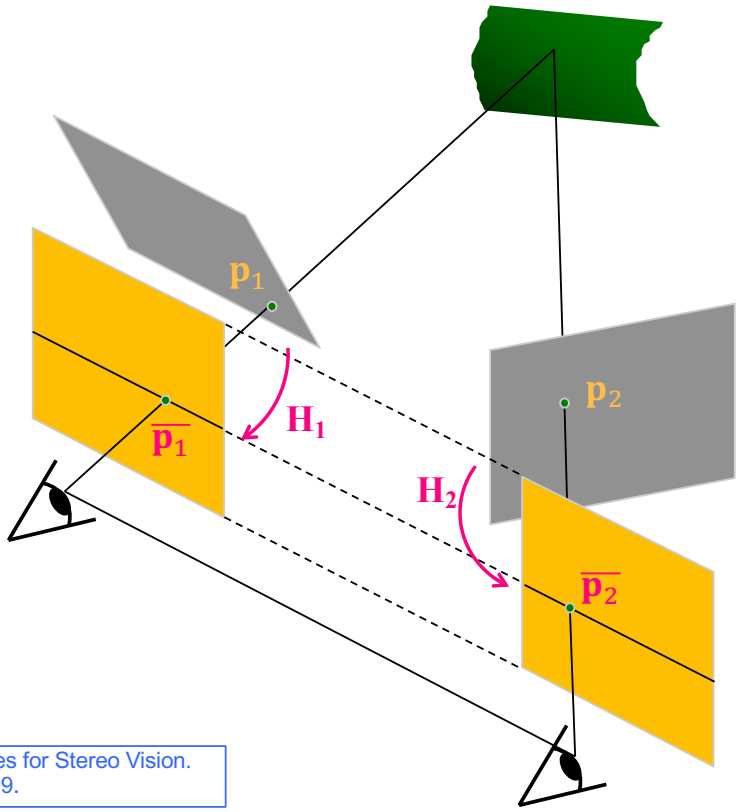
Legge di trasformazione: **Omografia**

(matrice di trasformazione $\boldsymbol{H}_{3x3}$)
delle coordinate-immagine

- una per ogni immagine

$$\bar{\mathbf{p}} = \mathbf{H}\mathbf{p} \qquad \mathbf{p} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix};\ \bar{\mathbf{p}} = \begin{bmatrix} \bar{x} \\ \bar{y} \\ 1 \end{bmatrix}$$

Conoscendo le omografie $\boldsymbol{H}_1$ e $\boldsymbol{H}_2$ posso rimappare ogni pixel $\mathbf{p}$ dell'immagine originale in $\bar{\mathbf{p}}$ nell'immagine rettificata



C. Loop and Z. Zhang. Computing Rectifying Homographies for Stereo Vision. *IEEE Conf. Computer Vision and Pattern Recognition*, 1999.

Matrice essenziale di immagini rettificate

Vincolo epipolare in immagini rettificate

(verifichiamo che **epipolar lines = scanlines**)

Vincolo epipolare:

$$x^T E x' = 0, \quad E = [t_\times]R$$

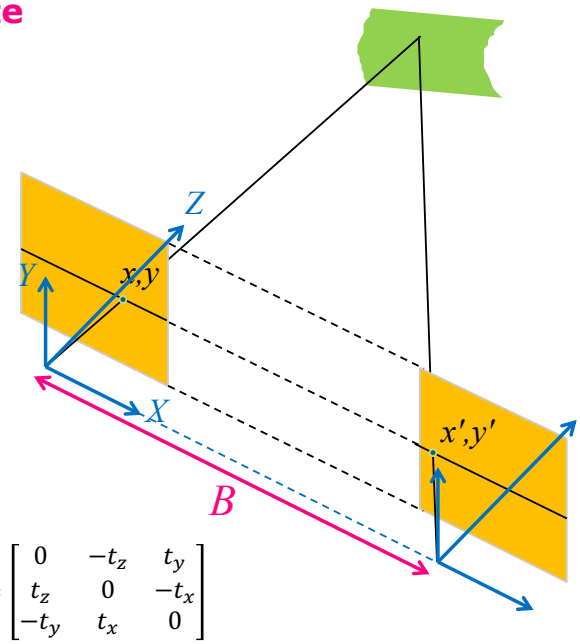
Nel caso di immagini rettificate:

$$\mathbf{R} = \mathbf{I}_3, \quad t = \begin{bmatrix} B & 0 & 0 \end{bmatrix}$$

Quindi la **E** di immagini rettificate è:

$$\mathbf{E} = \mathbf{t}_\times \mathbf{R} = \mathbf{t}_\times = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -B \\ 0 & B & 0 \end{bmatrix}$$

$$\mathbf{t}_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$



Infatti:

$$\begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -B \\ 0 & B & 0 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -B \\ By' \end{bmatrix} = 0 \quad \rightarrow \quad \cancel{By} = \cancel{By'}$$

→ **scanlines = epipolari**
(righe orizzontali omologhe)

Stereo image rectification

Rettifica di coppie di stereo-immagini

Vogliamo determinare le **omografie H** e **H'** che mappano ogni pixel $\mathbf{p_i}$ dell'immagine originale in quella rettificata $\bar{\mathbf{p}}_i$

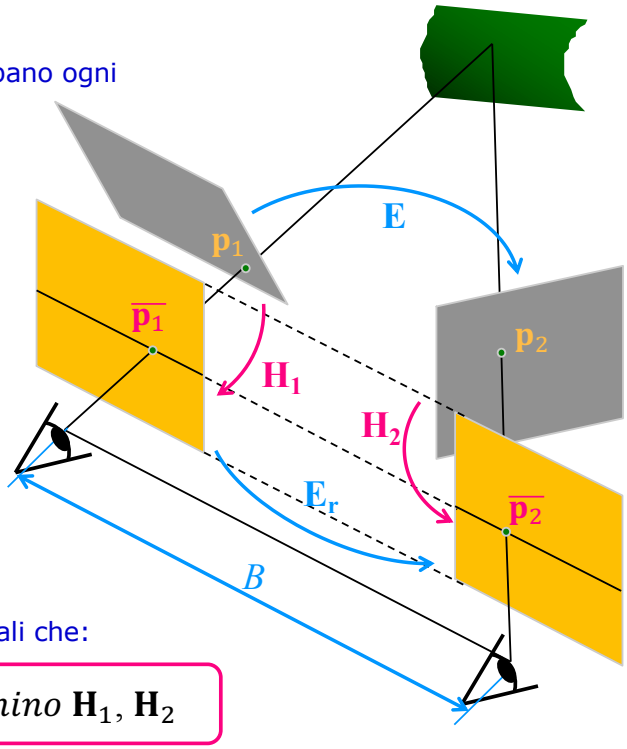
$$\bar{\mathbf{p}}_1 = \mathbf{H}_1\mathbf{p}_1, \quad \bar{\mathbf{p}}_2 = \mathbf{H}_2\mathbf{p}_2$$

Il vincolo essenziale sulle immagini rettificate:

$$\bar{\mathbf{p}}_2^T \, \mathbf{E}_\mathrm{r} \, \bar{\mathbf{p}}_1 = \mathbf{p}_2^T \mathbf{H}_2^T \, \mathbf{E}_\mathrm{r} \, \mathbf{H}_1 \mathbf{p}_1 = 0$$

dove: $\mathbf{E}_\mathrm{r} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -B \\ 0 & B & 0 \end{bmatrix}$

Ma: $\mathbf{p}_2^T \, \mathbf{E} \, \mathbf{p}_1 = 0 \rightarrow \mathbf{E} = \mathbf{H}_2^T \, \mathbf{E}_\mathrm{r} \, \mathbf{H}_1$

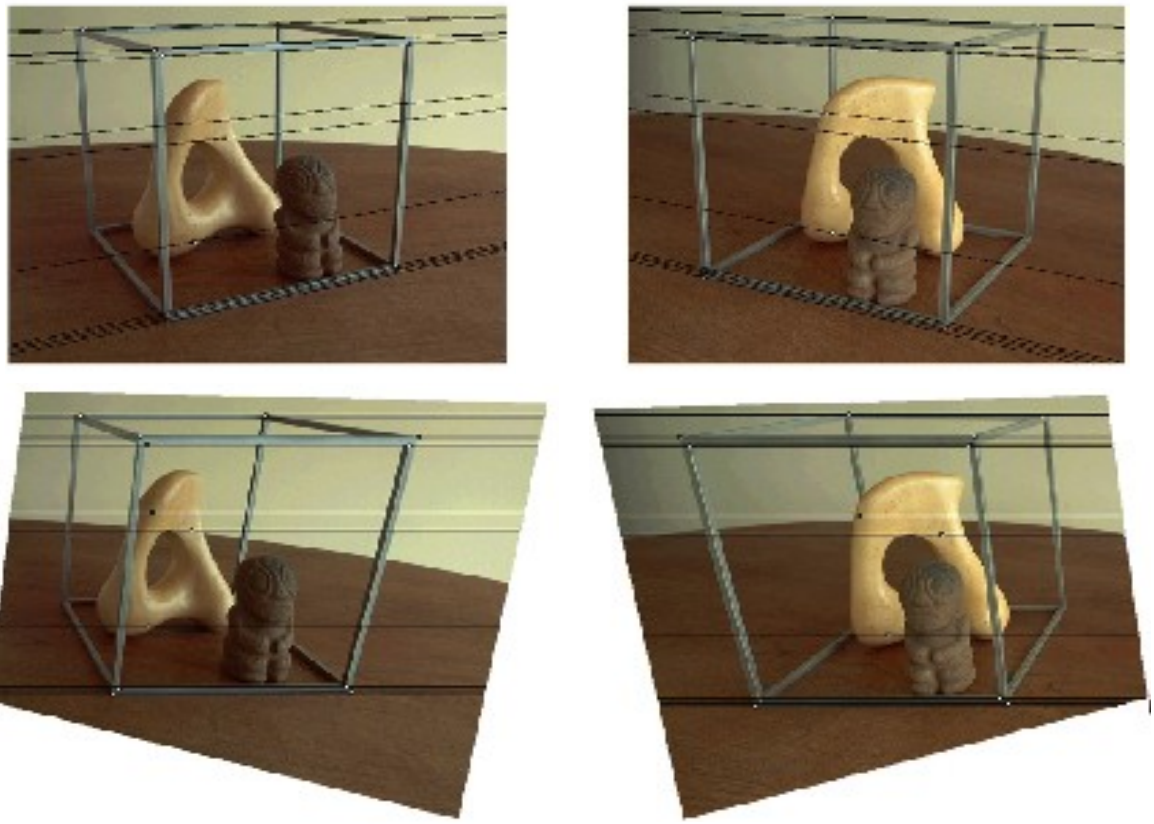


Quindi, data **E** e scelto *B*, posso determinare **H** e **H'** tali che:

$$\mathbf{E} = \mathbf{H}_2^T \, \mathbf{E}_\mathrm{r} \, \mathbf{H}_1 : \ \text{note } \mathbf{E}, \mathbf{E}_r \rightarrow \ \textit{determino } \mathbf{H}_1, \mathbf{H}_2$$

Soluzione non univoca! → $\textit{soluzione ottima} : \mathbf{H}_1, \mathbf{H}_2 \ \textit{a minima distorsione}$

Immagini rettificate: esempio



Visione stereoscopica (*Stereo Vision*)

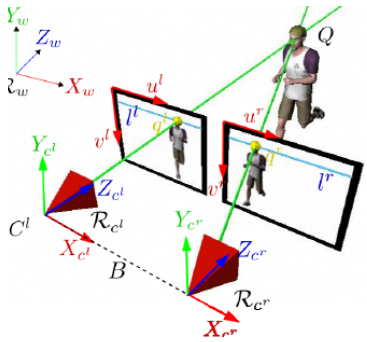
- Geometria della visione binoculare
  - disparità
  - vincolo epipolare: matrice essenziale e fondamentale
  - image rectification
- **Ricostruzione**
- Tecniche di matching
  - Approccio locale: area-based, edge-based
  - Approccio globale: graph cut

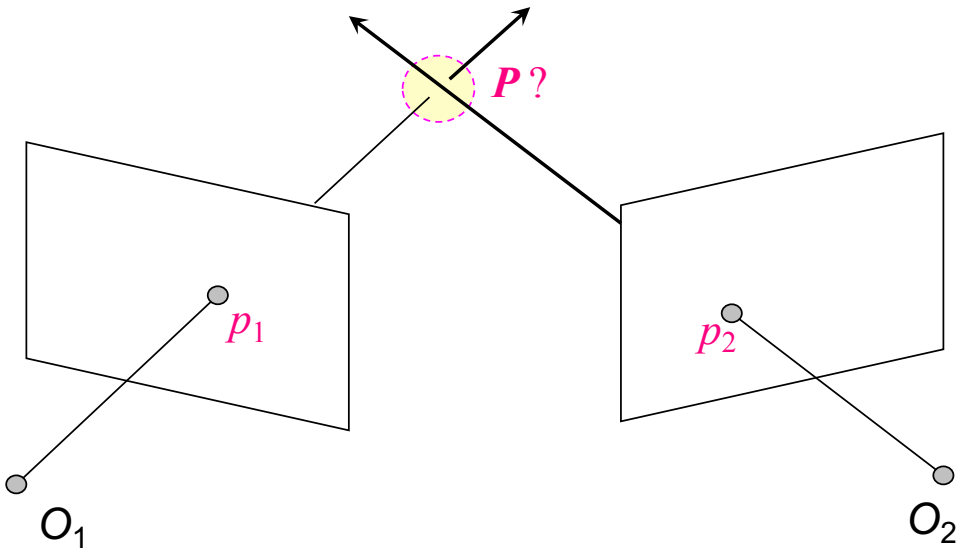
Ricostruzione: triangolazione

Triangolazione: definizione del problema

- data una coppia di immagini stereo calibrate (*stereo pair*)
- data la coppia di punti immagine (p_1, p_2) dello stesso punto di scena (matched points)

➔ determinare le coordinate 3D del punto ***P*** (pre-image)

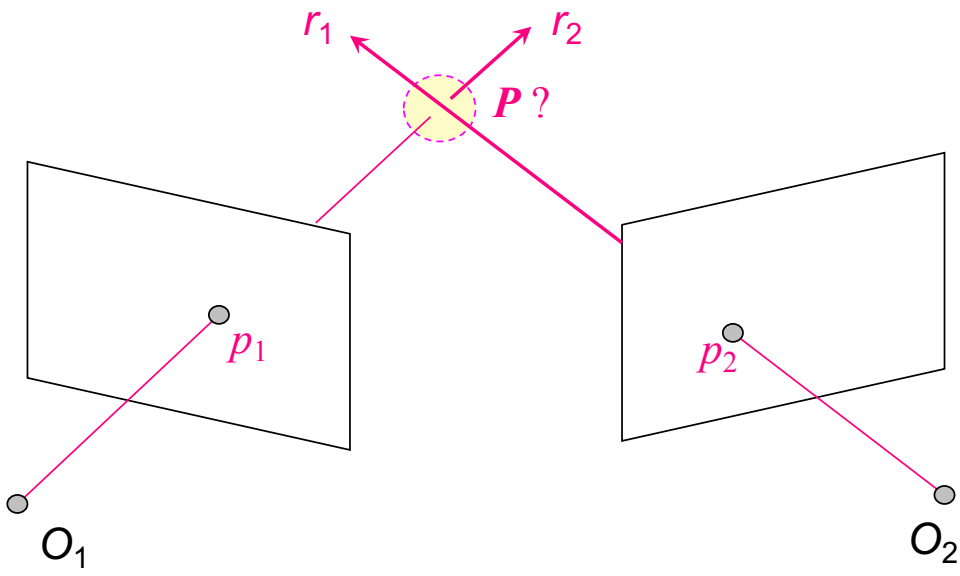




Triangolazione: il problema **in pratica**

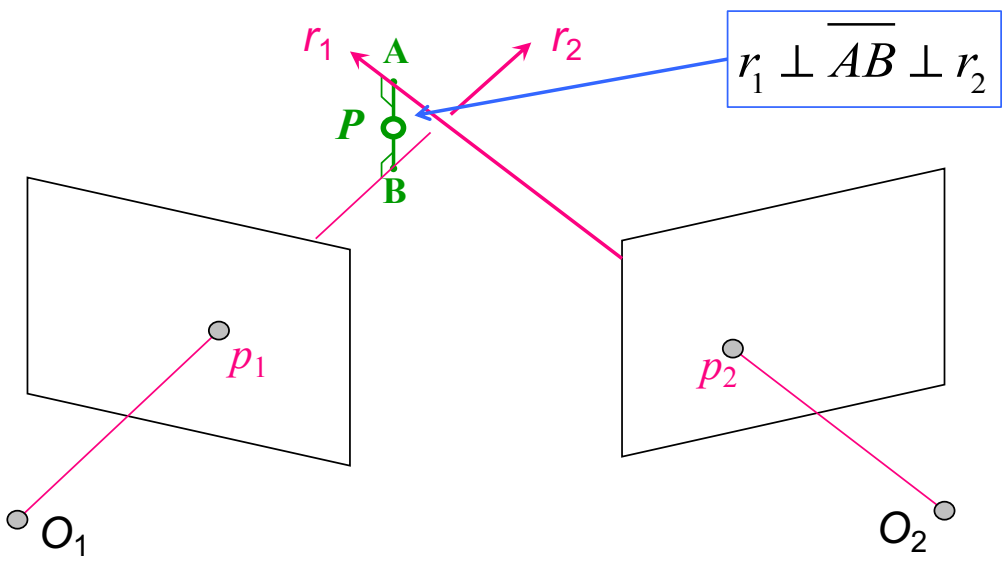
i due raggi r_1 e r_2 di norma non si intersecano! (rette sghembe)

- a causa di errori di localizzazione/calibrazione/...



Triangolazione – soluzione geometrica

- ***P*** – punto medio del segmento **AB** congiungente i due raggi r_1, r_2 nel punto di loro massima vicinanza
- Il segmento congiungente 2 rette sghembe nel punto di massima vicinanza è perpendicolare a entrambe



Triangolazione – soluzione algebrica lineare

Principio: in coordinate omogenee, i punti immagine $\mathbf{p}_1$ e $\mathbf{p}_2$ coincidono con la proiezione di $\mathbf{P}$ in ciascuna immagine: $p_1 \approx M_1 P$ e $p_2 \approx M_2 P$

$\rightarrow \lambda\, \mathbf{p}_i = \mathbf{M}_i\mathbf{P},\ i = 1,2$ ➔ **prodotto vettore nullo:** $\mathbf{p}_i \times \mathbf{M}_i\mathbf{P} = \mathbf{0},\ i = 1,2$

$$\begin{cases} \lambda_1 \mathbf{p}_1 = M_1\mathbf{P} \\ \lambda_2 \mathbf{p}_2 = M_2\mathbf{P} \end{cases} \rightarrow \begin{cases} \mathbf{p}_1 \times M_1\mathbf{P} = 0 \\ \mathbf{p}_2 \times M_2\mathbf{P} = 0 \end{cases} \rightarrow \begin{bmatrix} [\mathbf{p}_{1\times}]M_1 \\ [\mathbf{p}_{2\times}]M_2 \end{bmatrix} \mathbf{P} = 0$$

Dati: $\mathbf{p}_{1,2} = [x_{1,2} \quad y_{1,2} \quad 1]^T$ e $\mathbf{M}_{1,2}$:

ho 6 equazioni in 3 incognite ($\mathbf{P}$) ➔ sistema lineare sovradeterminato (non omogeneo)

$$\begin{bmatrix} \mathbf{p}_1 \times \boldsymbol{M}_1\mathbf{P} \\ \mathbf{p}_2 \times \boldsymbol{M}_2\mathbf{P} \end{bmatrix} = \begin{bmatrix} \mathbf{p}_{1\times}\, \boldsymbol{M}_1\, \mathbf{P} \\ \mathbf{p}_{2\times}\, \boldsymbol{M}_2\, \mathbf{P} \end{bmatrix} = \begin{bmatrix} \mathbf{p}_{1\times}\, \boldsymbol{M}_1 \\ \mathbf{p}_{2\times}\, \boldsymbol{M}_2 \end{bmatrix} \mathbf{P} = \mathbf{A}'_{[6\times4]} \begin{bmatrix} x_P \\ y_P \\ z_P \\ 1 \end{bmatrix} = \mathbf{0}_{[6\times1]}$$

definiti: $\mathbf{A}_{[6\times3]} = \mathbf{A}'(:, 1:3)\ ;\ \mathbf{b}_{[6\times1]} = -\mathbf{A}'(:, 4)$

$$\rightarrow \quad \mathbf{A}_{[6\times3]} \begin{bmatrix} \boldsymbol{x_P} \\ \boldsymbol{y_P} \\ \boldsymbol{z_P} \end{bmatrix} = \mathbf{b}_{[6\times1]} \quad \rightarrow \quad \boldsymbol{P} = \begin{bmatrix} \boldsymbol{x_P} \\ \boldsymbol{y_P} \\ \boldsymbol{z_P} \end{bmatrix} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

soluzione ottima (minimi quadrati)

Triangolazione – soluzione algebrica lineare

Dati: $\mathbf{p}_{1,2} = [p_x \quad p_y \quad 1]^T$ e $\boldsymbol{M_1}, \boldsymbol{M_2}$:

6 equazioni in 3 incognite ➔ sistema lineare sovradeterminato (non omogeneo)

$$\mathbf{p}_i \times M\,\mathbf{P} = [\mathbf{p}_{1\times}]M\,\mathbf{P} = \begin{bmatrix} 0 & -p_z & p_y \\ p_z & 0 & -p_x \\ -p_y & p_x & 0 \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} x_P \\ y_P \\ z_P \\ 1 \end{bmatrix} = 0$$

$$\mathbf{p}_{\boldsymbol{i}} \times \begin{bmatrix} m^i_{11} & m^i_{12} & m^i_{13} \\ m^i_{21} & m^i_{22} & m^i_{23} \\ m^i_{31} & m^i_{32} & m^i_{33} \end{bmatrix} \begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix} = -\mathbf{p}_{\boldsymbol{i}} \times \begin{bmatrix} m^i_{14} \\ m^i_{24} \\ m^i_{34} \end{bmatrix} \qquad i = 1,2$$

$$[\mathbf{p}_{1\times} \quad \mathbf{p}_{2\times}] \begin{bmatrix} m^1_{11} & m^1_{12} & m^1_{13} \\ m^1_{21} & m^1_{22} & m^1_{23} \\ m^1_{31} & m^1_{32} & m^1_{33} \\ m^2_{11} & m^2_{12} & m^2_{13} \\ m^2_{21} & m^2_{22} & m^2_{23} \\ m^2_{31} & m^2_{32} & m^2_{33} \end{bmatrix} \begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix} = -[\mathbf{p}_{1\times} \quad \mathbf{p}_{2\times}] \begin{bmatrix} m^1_{14} \\ m^1_{24} \\ m^1_{34} \\ m^2_{14} \\ m^2_{24} \\ m^2_{34} \end{bmatrix}$$

$$\mathbf{A}_{[6\times3]} \begin{bmatrix} \boldsymbol{x_P} \\ \boldsymbol{y_P} \\ \boldsymbol{z_P} \end{bmatrix} = \mathbf{b}_{[6\times1]} \quad \rightarrow \quad \boldsymbol{P} = \begin{bmatrix} \boldsymbol{x_P} \\ \boldsymbol{y_P} \\ \boldsymbol{z_P} \end{bmatrix} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

soluzione ottima (minimi quadrati)

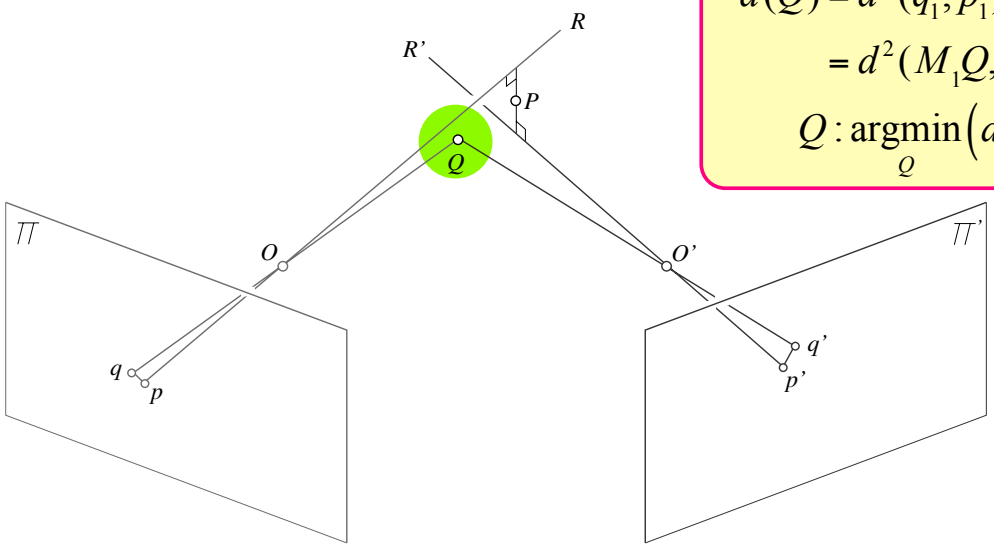
Triangolazione: approccio algebrico **non lineare**

Triangolazione – **soluzione algebrica non lineare (caso generale)**

- la soluzione lineare non è sempre ottima (ad es. se $f_1 \neq f_2$)

Soluzione ottima: *Q* – punto 3D la cui proiezione nelle due immagini q_1, q_2, è a **distanza quadratica minima dai punti immagine p_1, p_2**

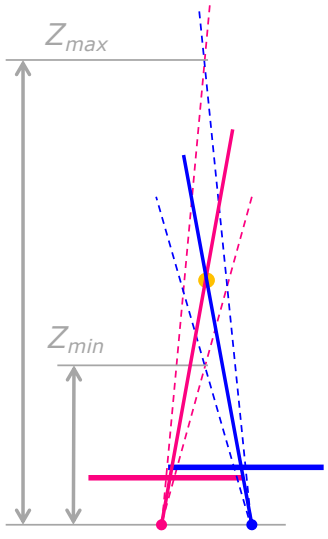
$$d(Q) = d^2(q_1, p_1) + d^2(q_2, p_2)$$
$$= d^2(M_1 Q, p_1) + d^2(M_2 Q, p_2)$$
$$Q : \underset{Q}{\operatorname{argmin}}\left(d(Q)\right)$$



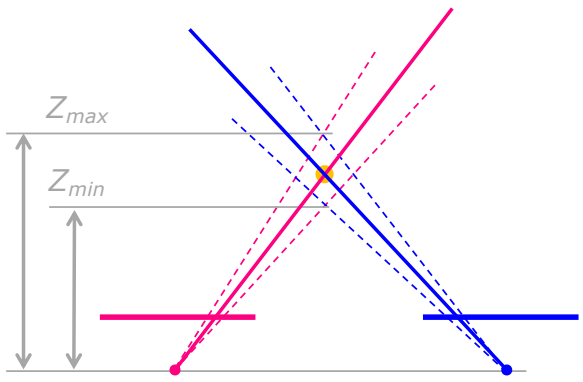
Ricostruzione: ruolo della **baseline**

Ruolo della ***baseline*** *nella qualità della ricostruzione*

- Small baseline: – maggiori errori in Z
  + *matching* più semplice (viste molto simili)
- Large baseline: + stima Z più accurata
  – *matching* più difficile (più occlusioni, viste diverse)



Small Baseline



Large Baseline

Source: S. Seitz

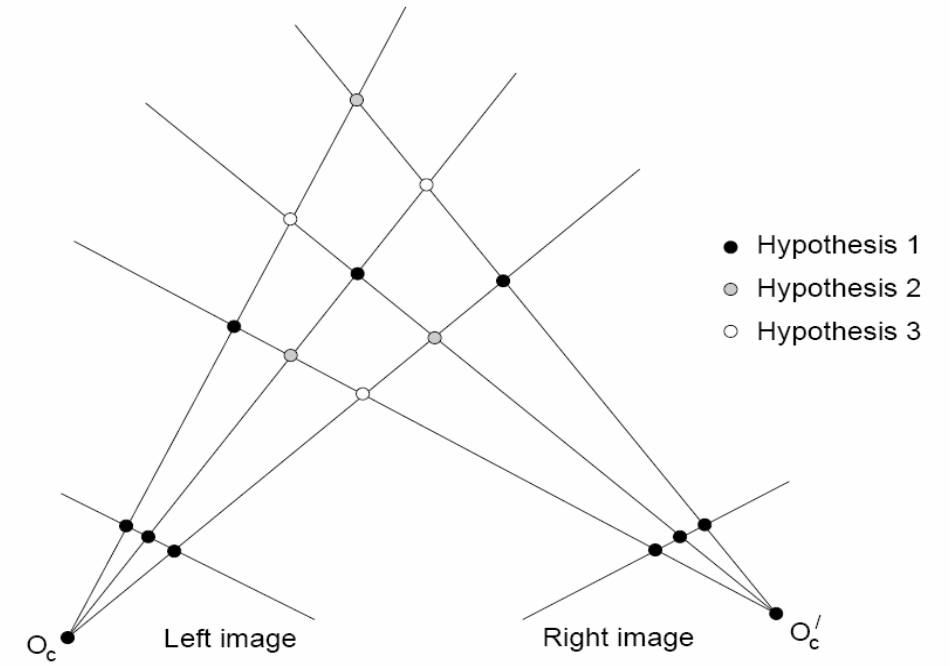
Visione stereoscopica (*Stereo Vision*)

- Geometria della visione binoculare
  - disparità
  - vincolo epipolare: matrice essenziale e fondamentale
  - image rectification
- Ricostruzione
- **Tecniche di matching**
  - Approccio **locale**: area-based, edge-based
  - Approccio **globale**: stereo scanline, graph cut
  - Approcci con **più di 2 viste**

Ambiguità della visione binoculare

Ambiguità:

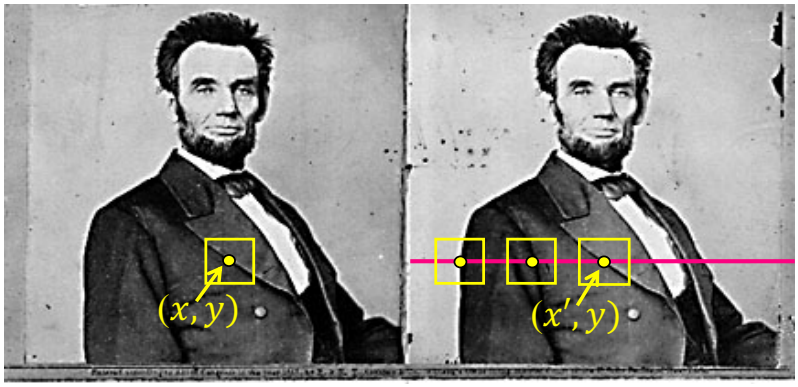
- *se ho più di un punto di cui determinare la corrispondenza, un errore di corrispondenza porta a un (grosso) errore di ricostruzione 3D*



Algoritmo di matching – approccio locale – correlazione

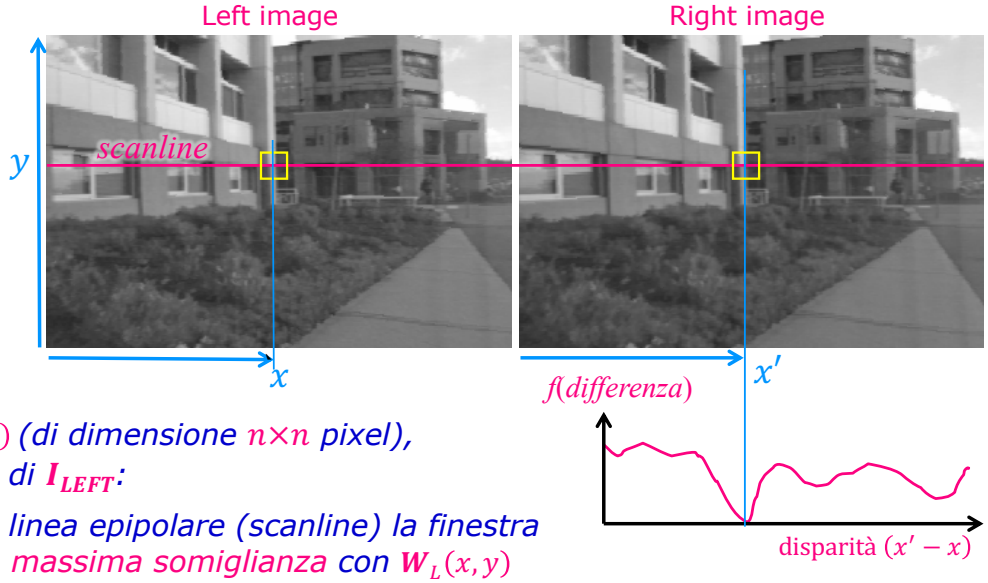
Algoritmo di matching – approccio classico (su immagini rettificate):

1. Rettifica delle stereo immagini (linee epipolari → scanlines)
2. Per ogni pixel $\mathbf{p} = (x, y)$ nell'immagine di partenza:
   - Trova la linea epipolare corrispondente (scanline) nell'immagine di destinazione
   - Esamina i pixel sull'epipolare e trova il punto x' **più somigliante: best match**
   - → Disparità: $x - x'$ → distanza (depth): $Z(\mathbf{p}) = \frac{Bf}{x - x'}$



- Ma… come misuro la "somiglianza"?
  → somiglianza dell'immagine "intorno al punto" → **regione intorno al punto**

Algoritmo di matching – approccio locale – correlazione



Algoritmo di matching:

Per ogni finestra $W_L(x, y)$ (di dimensione $n \times n$ pixel), centrata nel punto (x, y) di I_{LEFT}:

cerca in I_{RIGHT}, lungo la linea epipolare (scanline) la finestra $W_R(x', y)$ che presenta la massima somiglianza con $W_L(x, y)$

Metriche per misurare la somiglianza:

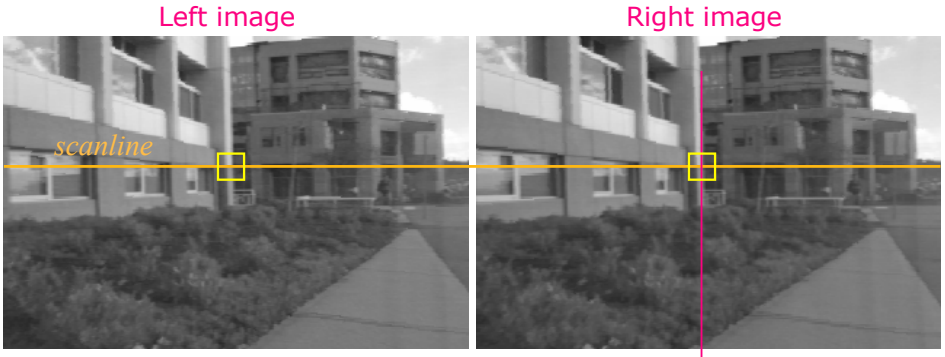
- **Correlazione** normalizzata
- **SSD** – Sum of Squared Differences
- **SAD** – Sum of Absolute Differences

Algoritmo di matching – approccio locale – correlazione

Correlazione normalizzata

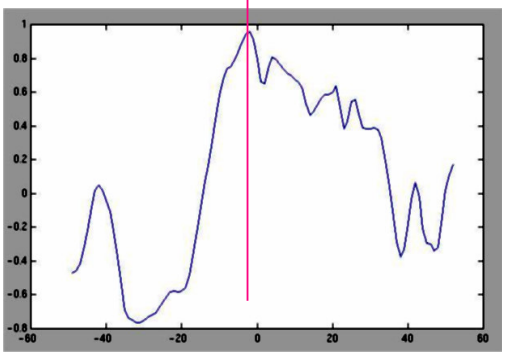
correlazione pixel-a-pixel di 2 finestre rettangolari

normalizzata: luminanze
– a media **= 0**
– a varianza **= 1**



$$I_L(x+i,y+j) \rightarrow w_L(k),\ k \in W$$
$$I_R(x'+i,y+j) \rightarrow w_R(k),\ k \in W$$

→ *finestra $I(x,y)$ → array $w(k)$*



$$C(x,y,d) = \frac{\sum_{k \in W}\left[\left(w_L(k)-\overline{w_L}\right)\cdot\left(w_R(k)-\overline{w_R}\right)\right]}{\left\|\mathbf{w}_L-\overline{w_L}\right\|\left\|\mathbf{w}_R-\overline{w_R}\right\|}$$

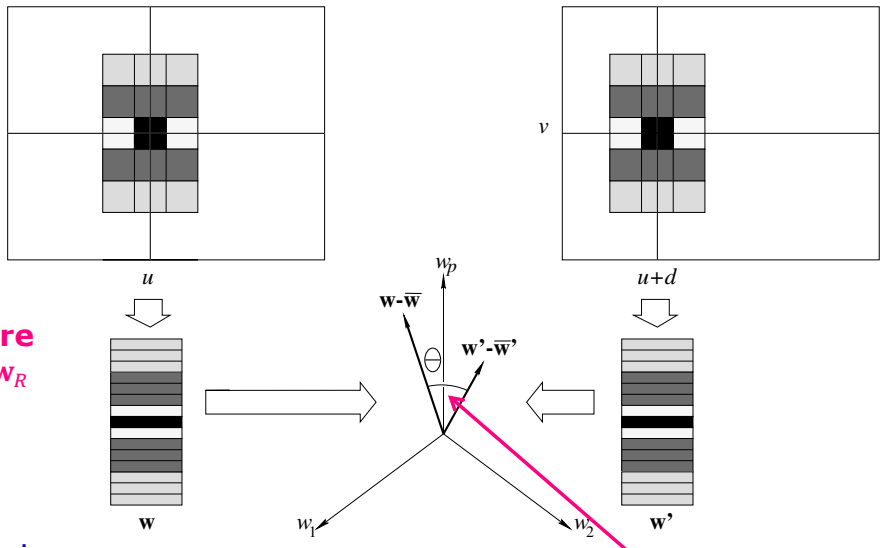
La normalizzazione rende la stima robusta rispetto a:

- variazioni di illuminazione / guadagno camera
- superfici non (o parzialmente) Lambertiane

Algoritmo di matching – approccio locale – correlazione

Correlazione normalizzata

considero la matrice di pixel delle 2 finestre come **vettori 1D:** $\mathbf{w}_L(i), \mathbf{w}_R(i)$



➔ la correlazione normalizzata coincide con il **prodotto scalare** normalizzato dei 2 vettori $\mathbf{w}_L \cdot \mathbf{w}_R$

$$\frac{w_L \cdot w_R}{\|w_L\|\|w_R\|} = \cos(\theta_{LR})$$

è **massimo** quando i due vettori:

$$\mathbf{w}_R = \mu + \lambda \mathbf{w}_L$$

➔ **insensibile** a gain/offset

$$C(x,y,d) = \frac{\sum_{k \in W}\left[\left(w_L(k)-\overline{w_L}\right)\cdot\left(w_R(k)-\overline{w_R}\right)\right]}{\left\|\mathbf{w}_L-\overline{w_L}\right\|\left\|\mathbf{w}_R-\overline{w_R}\right\|} = \cos\theta(d)$$

Approccio locale – **correlazione**

Metriche alternative alla correlazione

la correlazione normalizzata funziona bene, ma è pesante da calcolare ($\sim 5\, n^2$)

Metriche più efficienti:

SSD – *Sum of Squared Differences:*

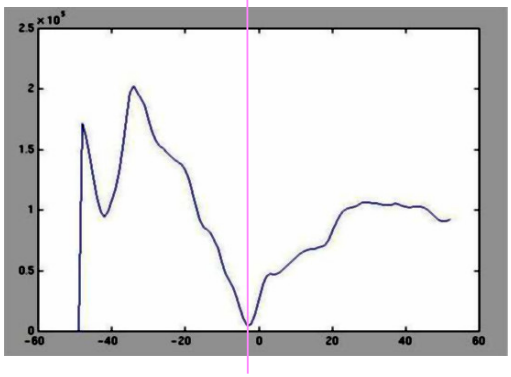
$$I_L(x+i,y+j) \rightarrow w_L(k),\ k \in W$$

$$I_R(x'+i,y+j) \rightarrow w_R(k),\ k \in W$$

$$C(x,y,d) = \sum_{k \in W} \left(w_L(k) - w_R(k)\right)^2, \quad d = x - x'$$

SAD – Sum of Absolute Differences:

$$C(x,y,d) = \sum_{k \in W} \left| w_L(k) - w_R(k) \right|$$



$\mathbf{SSD}(x - x')$

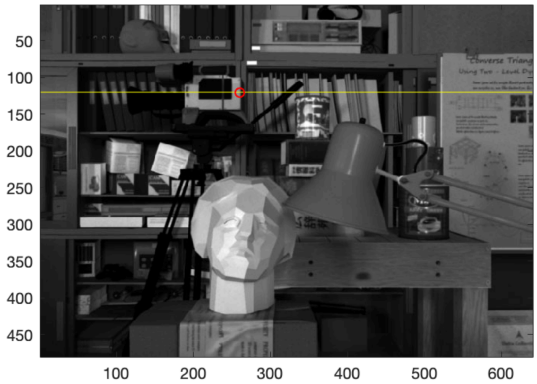
Esercitazione

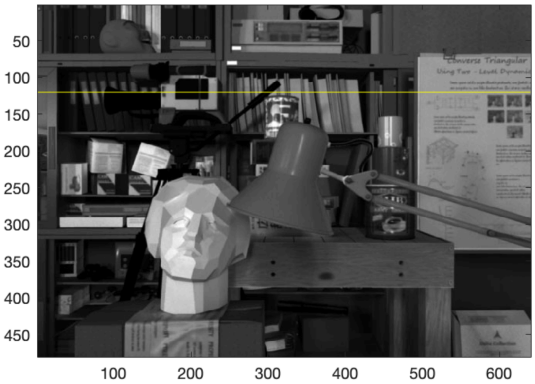
Esercitazione MATLAB®:

- Matching per correlazione:
  confronto metriche di somiglianza
  (correlazione normalizzata, SSD, SAD)

- MATLAB LIVE script: **`Correlation.mlx`**

```
>> edit Correlation.mlx
```

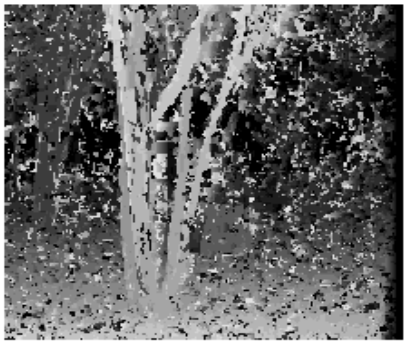




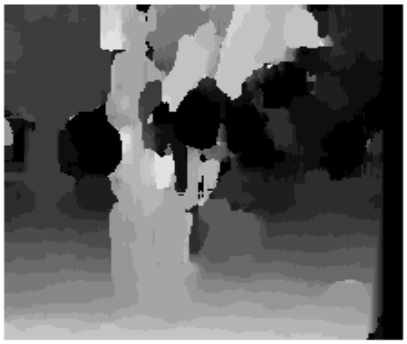
Effetti della dimensione di finestra

Effetto delle **dimensioni della finestra** ($w \times w$) sul risultato:

- Finestre **piccole**: + maggior dettaglio
  – risultati più rumorosi
- Finestre **grandi**: + risultati meno rumorosi, disparità "smooth"
  – perdita di dettaglio



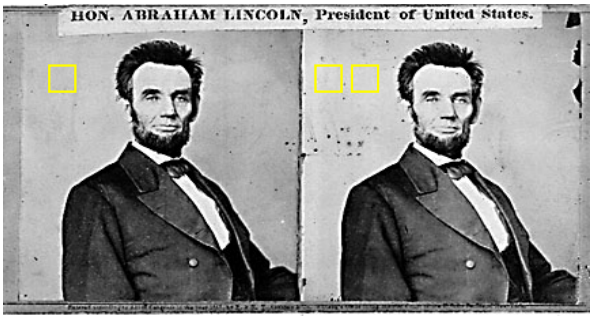
w = **3** pixel



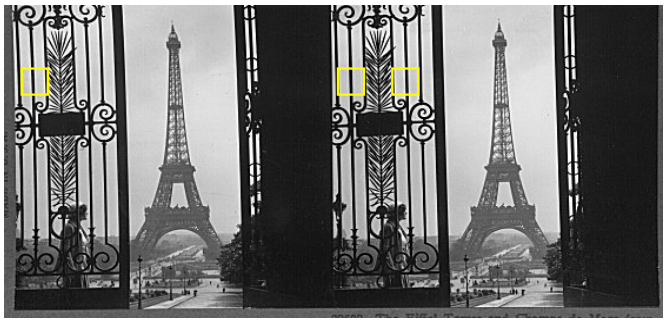
w = **20** pixel

Approccio a correlazione – **limiti**

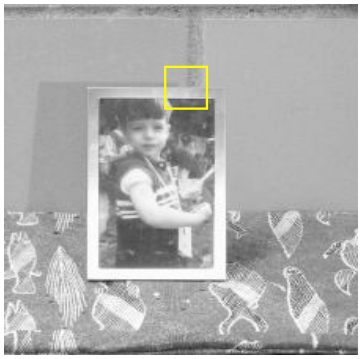
Limiti dell'approccio a correlazione:

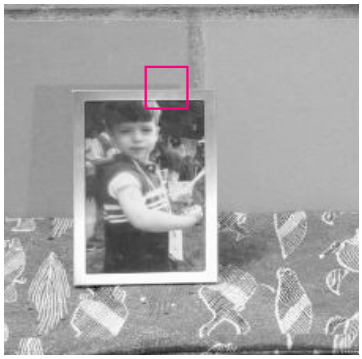


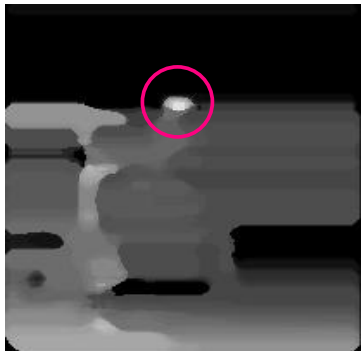
assenza di texture



pattern ripetuti







superfici non Lambertiane, specularità, occlusioni su bordi oggetti

Approccio a correlazione – **limiti**

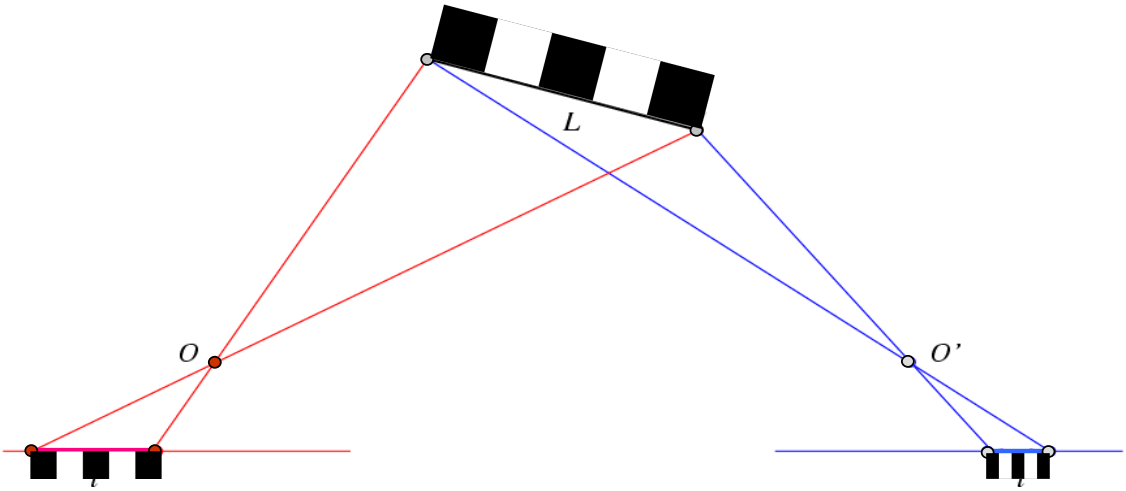
Limiti dell'approccio a correlazione

Foreshortening:

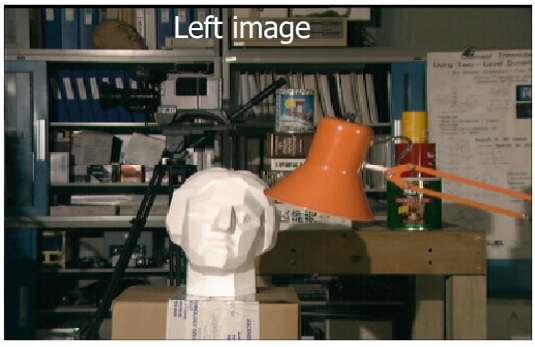
Superficie vista da angolazioni differenti, dalle 2 immagini

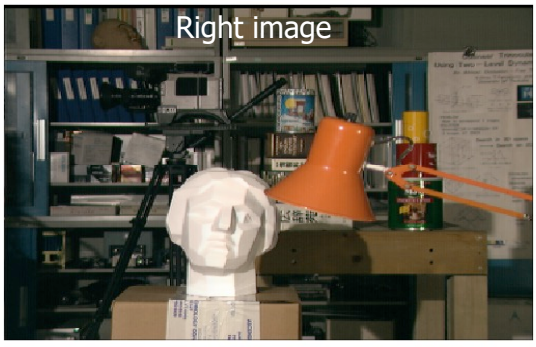
→ "stretching" (disorsione prospettica) di un pattern rispetto all'altro

→ correlazione con finestre a dim. fissa **non funziona**

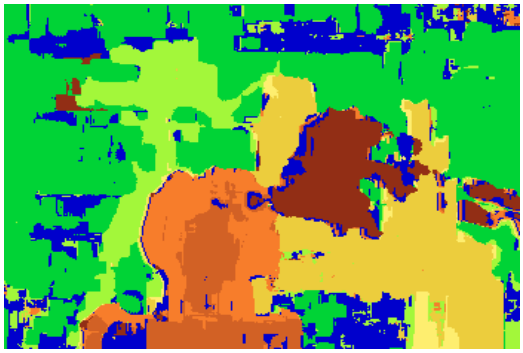


Approccio a correlazione – **esempio di risultato**





Window-based matching



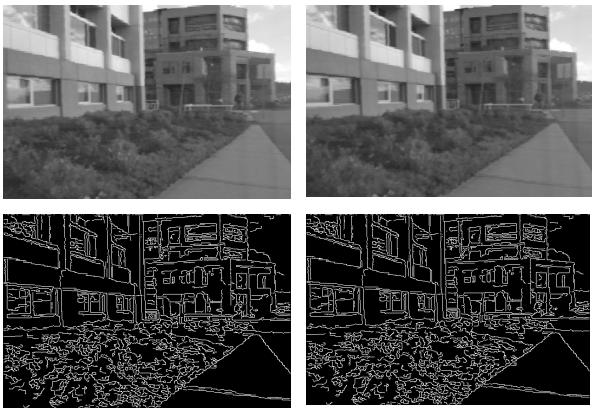
Ground truth

Approccio: edge matching

Approccio alternativo: **Edge Matching**

I contorni sono preziosi per la ricostruzione 3D:

- corrispondono alle caratteristiche dimensionali salienti di un oggetto (bordi)
- sono localizzabili con esattezza



Problema: edge detection "rumorosa"

➔ pre-filtraggio ("blurring") ➔ gli edge si spostano

Soluzione: edge detection piramidale

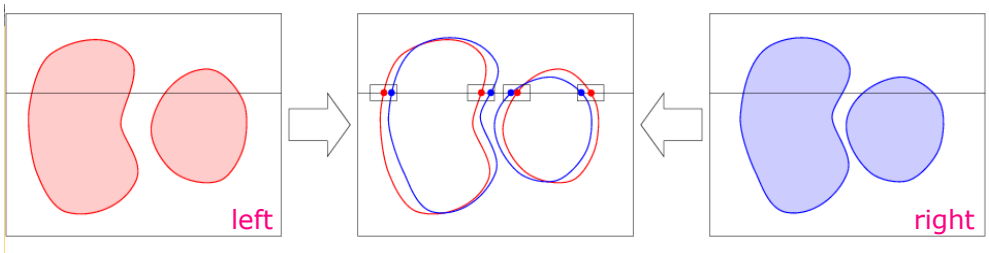
- edge **detection** con blurring **elevato**
- edge **tracking** con blurring **via via più leggero**

Approccio locale – **edge matching**

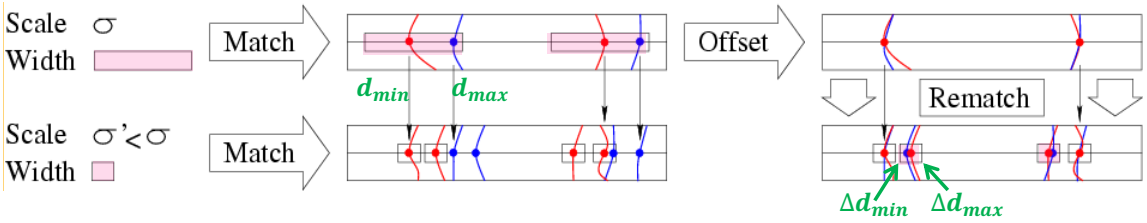
Approccio locale alternativo:

Edge Matching – approccio multi-risoluzione

1. edge **matching** a bassa risoluzione (blurring gaussiano elevato)
   - range disparità **ampio**: da d_{min} (Z_{max}) a d_{max} (Z_{min})



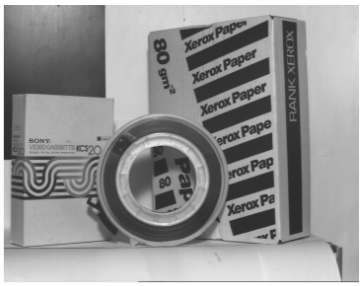
2. edge **tracking** cambiando via via risoluzione
   - range disparità **ristretto:** $\Delta d = f(\Delta\sigma)$ dovuto alla migrazione dei contorni per la variazione di σ.

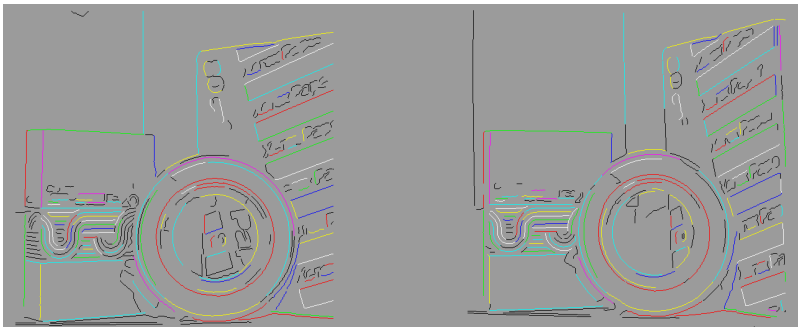


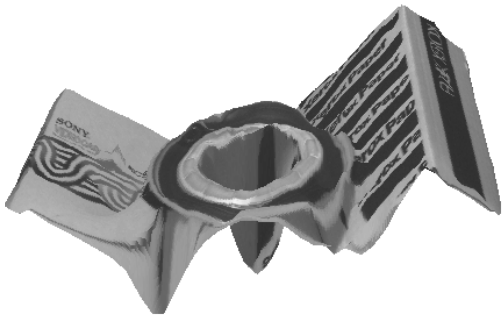
Approcio locale – **edge matching**

Edge Matching – **esempio**

- La ricostruzione 3D dei soli contorni fornisce una **depth map sparsa**, ma spesso le informazioni salienti di forma sono in corrispondenza dei contorni
  - mi basta **interpolare le superfici 3D** tra contorni vicini







Come migliorare ulteriormente il matching?

Entrambi gli approcci visti (window/edge matching) non danno risultati ideali

Come migliorare il matching?

- Il vincolo di somiglianza è **locale**
  - operazione di matching indipendente per ogni finestra
  - ➔ non si sfrutta l'informazione dei match vicini (in forte correlazione!)
  - ➔ si può incorrere in **minimi locali**

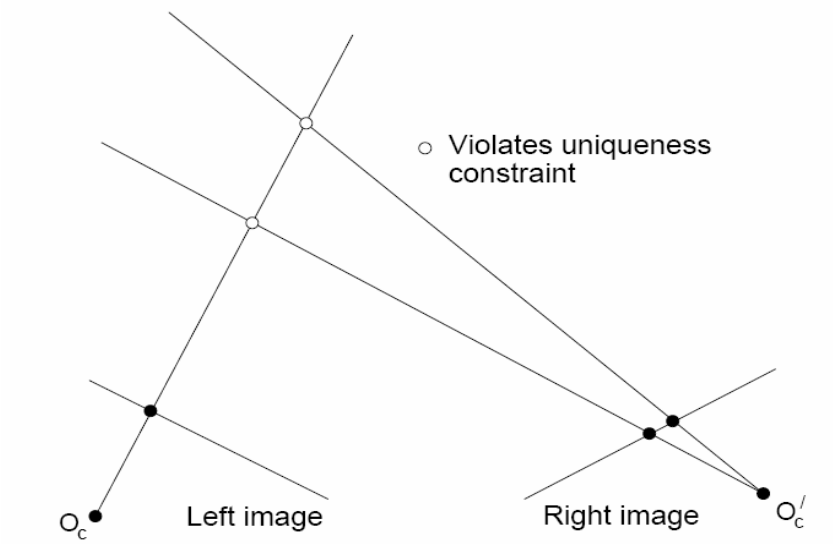
Idea:

- Superare l'approccio **locale:**
- Imporre tutti i vincoli di corrispondenza disponibili

➔ approccio **globale** su tutta la linea epipolare → **scanline**

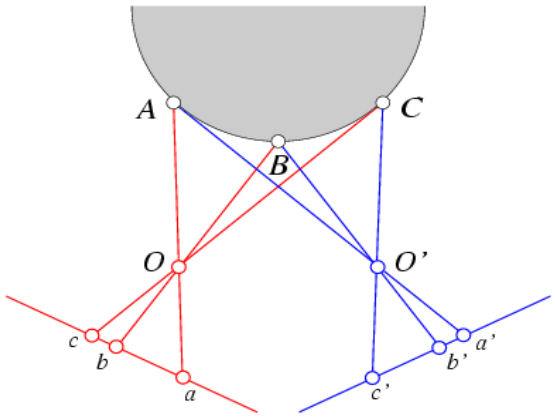
Vincoli di corrispondenza globali

- **Unicità**
  - *Per ogni punto in un'immagine, c'è **al più** un punto corrispondente nell'altra immagine*



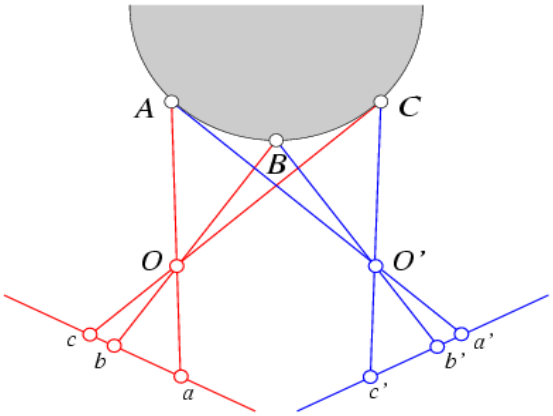
Vincoli di corrispondenza globali

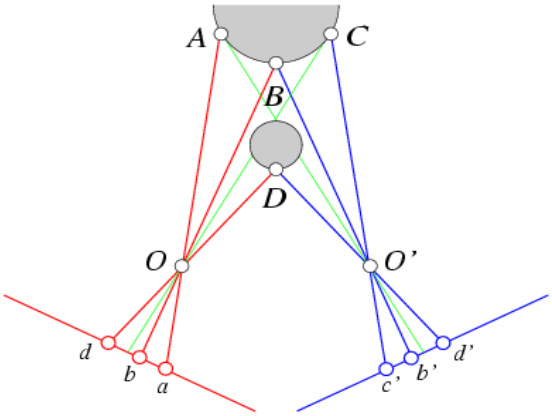
- **Unicità**
  - *Per ogni punto in un'immagine, c'è **al più** un punto corrispondente nell'altra immagine*
- **Ordinamento**
  - *Punti corrispondenti devono trovarsi nello **stesso ordine**, lungo la linea epipolare (scanline) in entrambe le immagini*



Vincoli di corrispondenza globali

- **Unicità**
  - *Per ogni punto in un'immagine, c'è **al più** un punto corrispondente nell'altra immagine*
- **Ordinamento**
  - *Punti corrispondenti devono trovarsi nello **stesso ordine**, lungo la linea epipolare (scanline) in entrambe le immagini*





Eccezione: ordinamento non rispettato

Vincoli globali di corrispondenza globali

- **Unicità**
  - *Per ogni punto in un'immagine, c'è **al più** un punto corrispondente nell'altra immagine*
- **Ordinamento**
  - *Punti corrispondenti devono trovarsi nello **stesso ordine**, lungo la linea epipolare (scanline) in entrambe le immagini*

- **Regolarità** (*smoothness*)

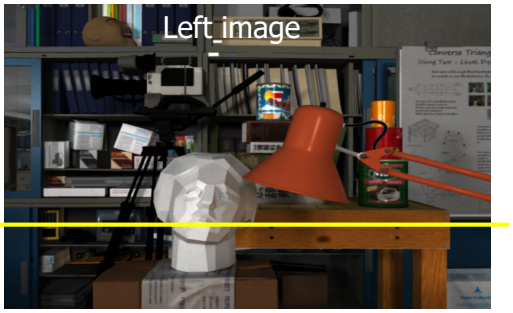
  *La funzione **disparità** deve:*
  - *variare **lentamente** muovendosi sulla linea epipolare (per la maggior parte del tempo)*

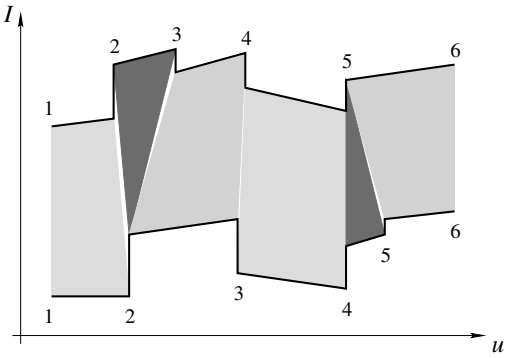
    oppure...
  - *avere una **discontinuità** (occlusioni: foreground → background)*

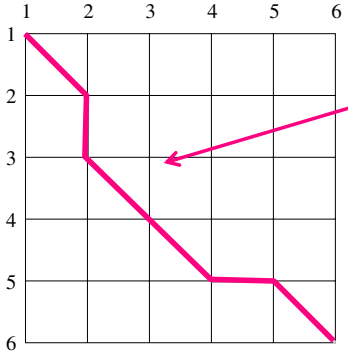
Approccio globale – *Scanline stereo*

Scanline stereo (Ohta & Kanade, 1985)

- Matching coerente di tutti i pixel di una linea epipolare (scanline)



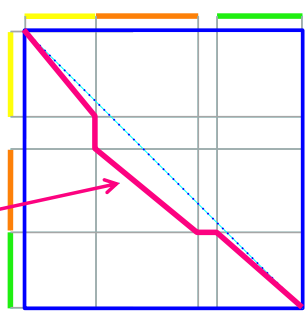




soluzione matching globale

La **disparità** è definita per **tutti i pixel** di ogni epipolare (scanline)

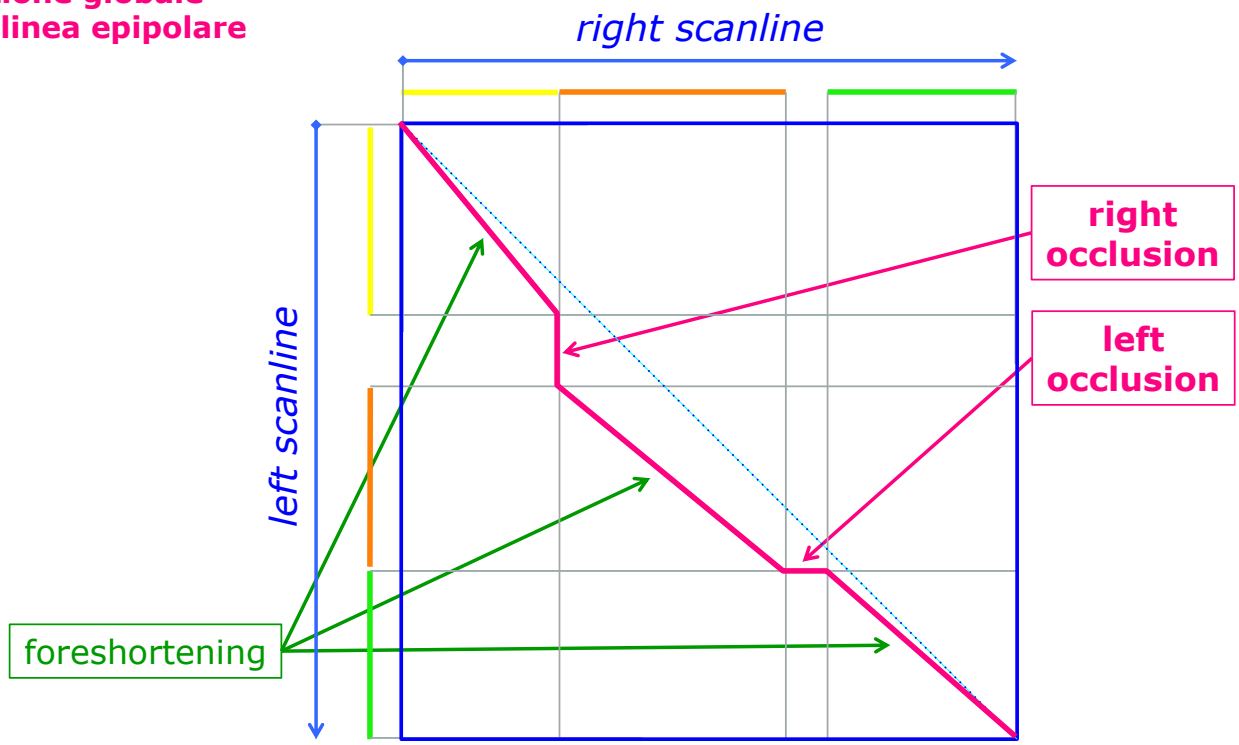
Scanline stereo



Soluzione globale sulla linea epipolare

Scanline stereo

Soluzione globale sulla linea epipolare

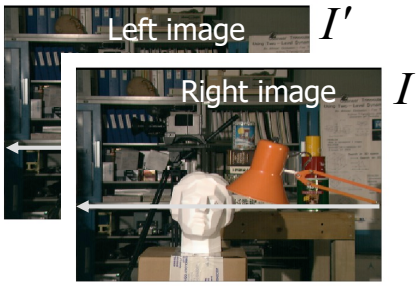


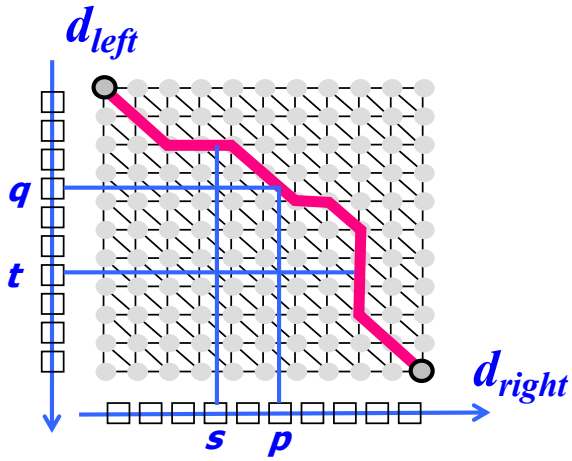
Scanline stereo: "shortest paths"

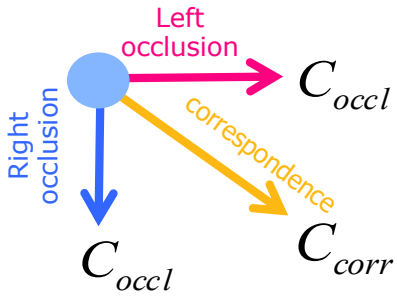
Scanline stereo

Matching definito come **cammino ottimo**,
calcolato mediante tecniche di programmazione dinamica

$$M = \{C_{OPT}(i)\} = \arg\min\left[\sum_i C(i)\right] \quad i = 0 \,..\, N-1$$







Ohta & Kanade (1985), Cox et al. (1996)

Slide credit: Y. Boykov

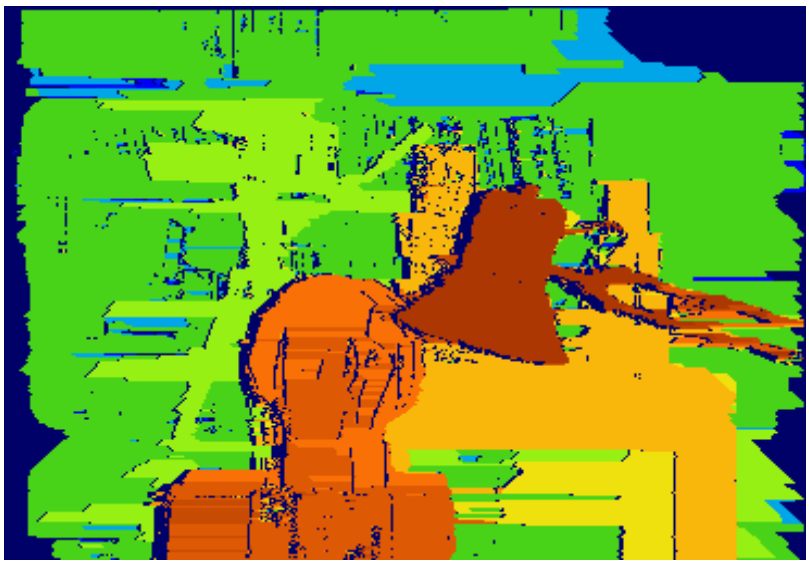
Scanline stereo

Scanline stereo genera *"streaking artifacts"*

- programmazione dinamica 1D → soluzione indipendente su ogni scanline
- non si riesce a sfruttare la programmazione dinamica su una griglia 2D

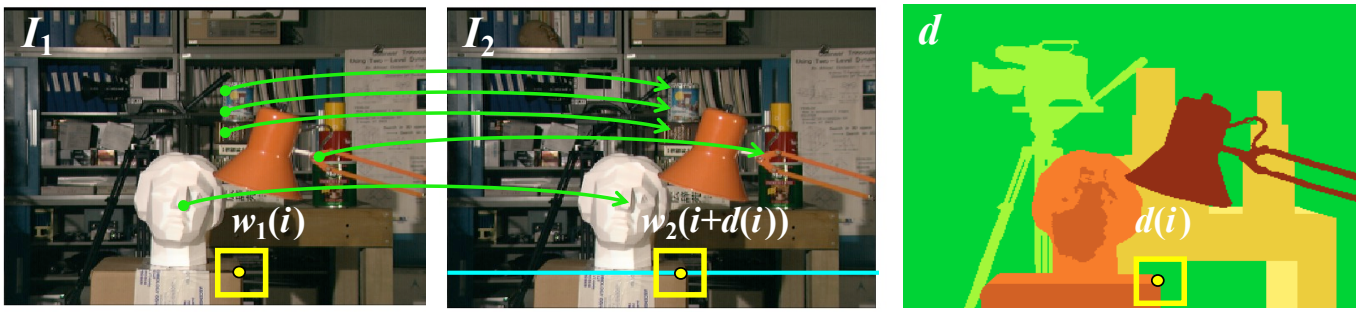
Soluzione (parziale):

- raffinamento risultati 1D sfruttando correlazione con scanline adiacenti



ground truth

Stereo matching as **energy minimization**



- Modello campo di disparità come un **grafo**:

$$G=(V,E),\quad \begin{cases} V: \text{pixel} \in I_1 \\ E: \text{archi verso pixel corrisp.} \in I_2 \\ \quad w_1(i) \rightarrow w_2(i+d(i)) \end{cases}$$

- Funzione **energia** del campo di disparità **d**:

$$E(G)=E(\mathbf{d})=\sum_{i=1}^{N}\underbrace{U\left(d(i)\right)}_{U(d):\text{ somiglianza}}+\sum_{i,j\in E}\underbrace{Eij\left(d(i),d(j)\right)}_{E_{ij}:\text{ regolarità}}$$

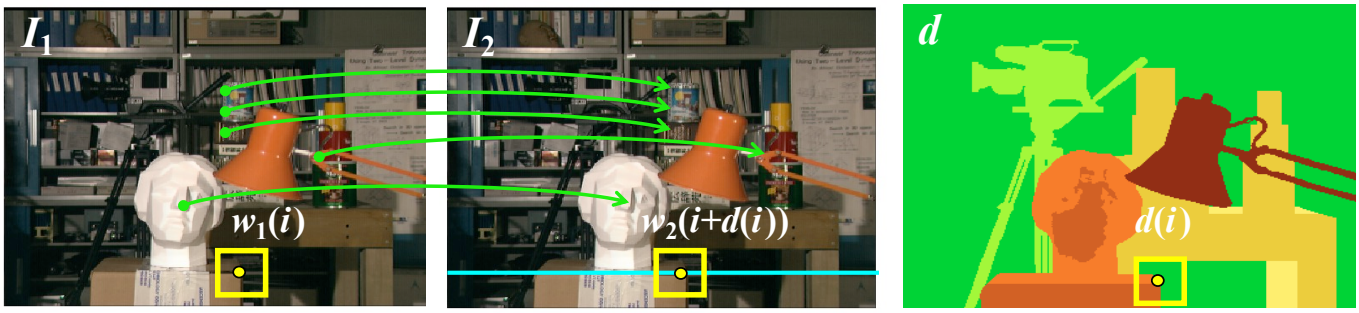
- termine di somiglianza:

$$U_i\left(d(i)\right)=\sum_{W}\left(w_1(i)-w_2(i+d(i))\right)^2$$

- regolarità della disparità:

$$E_{\text{ij}}=\sum_{\text{neighbors } i,j}\gamma\left|d(i)-d(j)\right|$$

Stereo matching as **energy minimization**



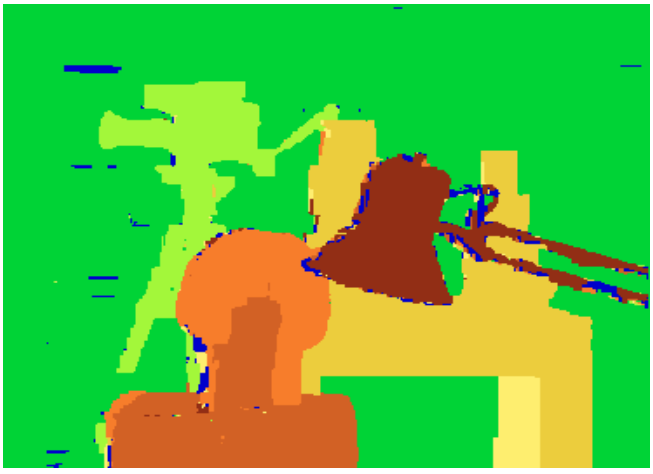
Soluzione: **minimo dell'energia *E*(d)**

$$\mathbf{d}_{OPT} = \underset{\mathbf{d}}{\operatorname{argmin}} \left\{ E(\mathbf{d}) = \sum_{i=1}^{N} U\left(d(i)\right) + \sum_{i,j \in E} Eij\left(d(i), d(j)\right) \right\}$$

$$U_i\left(d(i)\right) = \sum_{W} \left(w_1(i) - w_2(i + d(i))\right)^2 \qquad E_{\text{ij}} = \sum_{\text{neighbors } i,j} \gamma \left| d(i) - d(j) \right|$$

- Funzioni energia in questa forma possono essere minimizzate in modo esatto in tempo polinomiale, mediante algoritmi di taglio di grafo: **min-cut/max-flow**

Stereo matching as **energy minimization**



Approccio "*graph cuts*"

Ground truth

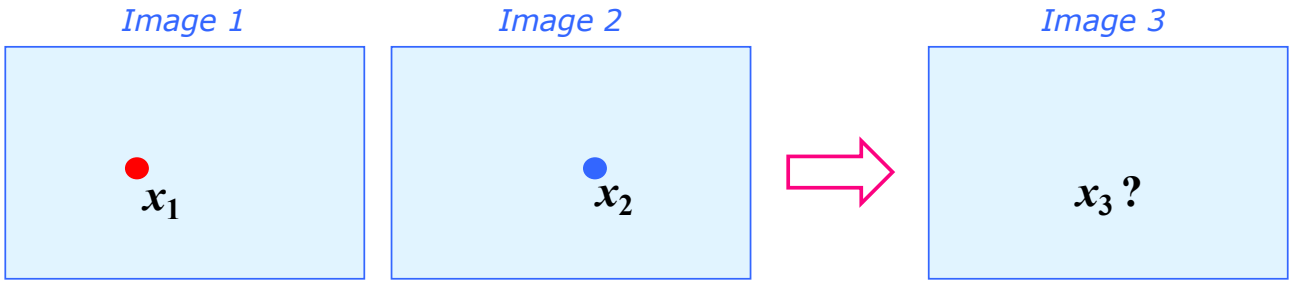
Y. Boykov, O. Veksler, and R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, PAMI 2001

Geometrie con **più di due immagini**

Geometrie stereo multi-oculari (più di 2 immagini)

- Date **3 camere calibrate**,
- date le coordinate-immagine corrispondenti (di uno stesso punto **X** di scena) in 2 immagini, x_1 e x_2

➔ cosa si può sapere delle coordinate immagine dello stesso punto nella terza?



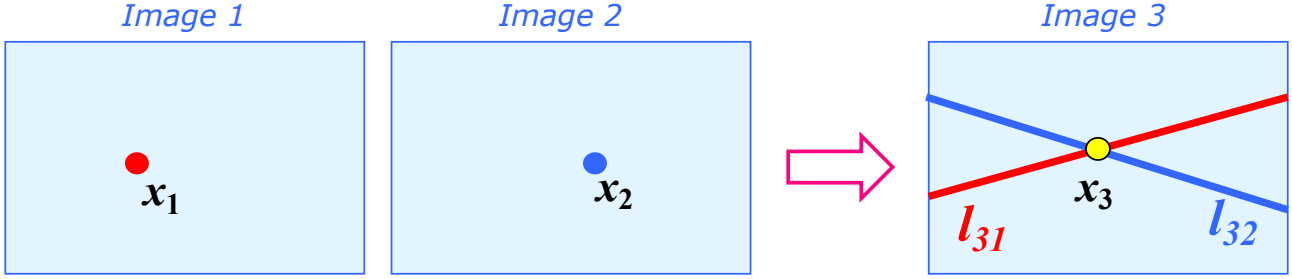
Geometrie con **più di due immagini**

Geometria trinoculare

- Date 3 camere calibrate,
- date le coordinate-immagine corrispondenti (di uno stesso punto X di scena) in 2 immagini, x_1 e x_2

➔ è possibile determinare le coordinate-immagine x_3 dello stesso punto nella terza.

Eccezione: epipolari **parallele** (come nelle immagini rettificate)



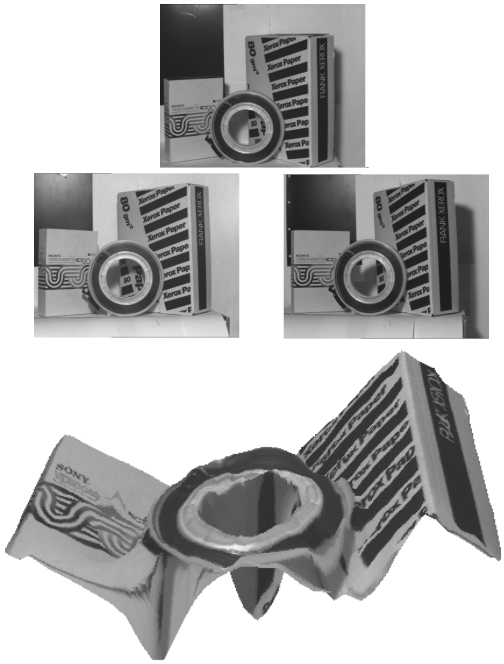
$$l_{31} = F_{13}^T x_1$$

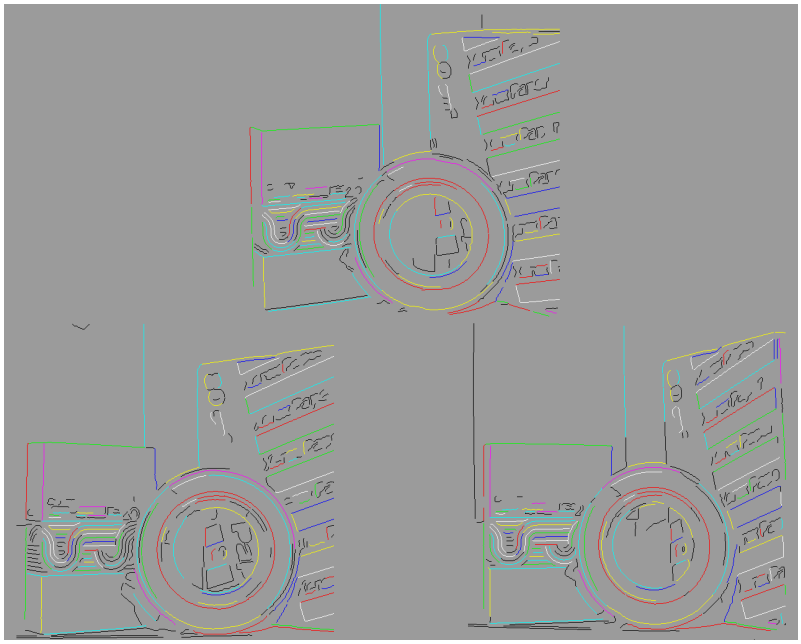
$$l_{32} = F_{23}^T x_2$$

Geometrie con **più di due immagini**

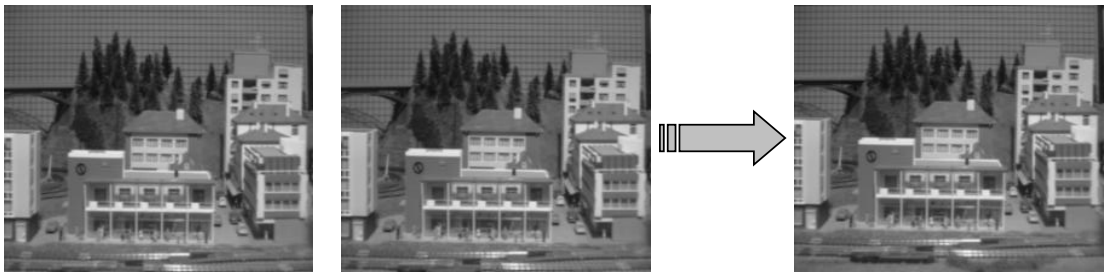
Edge Matching

- Accoppiamento di contorni – matching per vincoli di ordinamento
- Algoritmi efficienti O(n) di **matching tri-partito** di triplette di contorni $\langle l, m, r \rangle$





Geometrie con **più di due immagini**

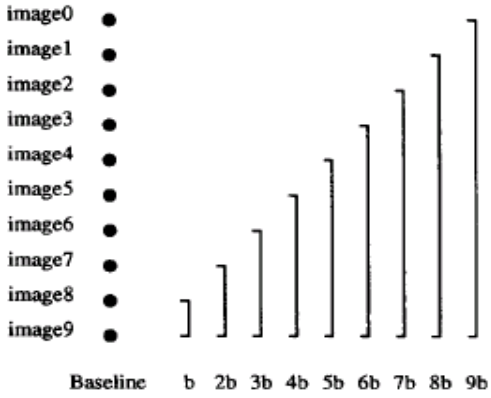


Multi-baseline stereo

- ***N*** immagini (più di 2), rettificate, relative alla stessa scena.

Le posso vedere come:

- ➔ ***N*–1** binocular stereo pairs rettificati
- ➔ ***N*–1** differenti baseline



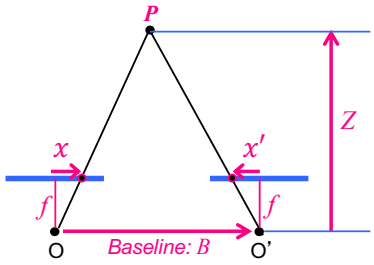
M. Okutomi and T. Kanade, "A Multiple-Baseline Stereo System," IEEE Trans. on Pattern Analysis and Machine Intelligence (1993)

Geometrie con **più di due immagini**

Multi-baseline stereo

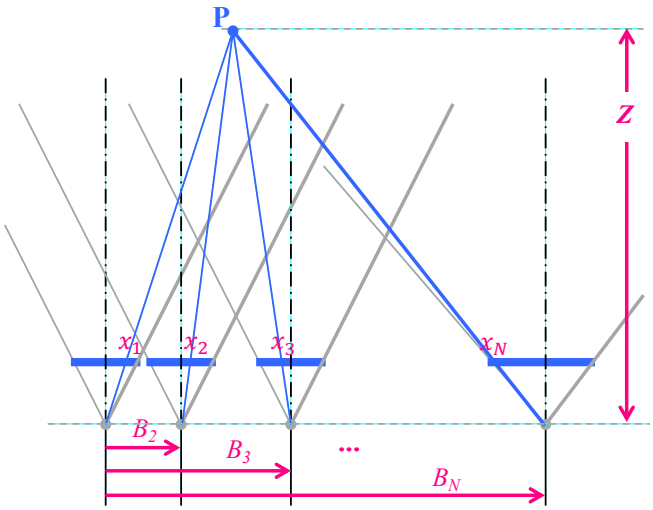
- Essendo immagini rettificate:
- Immagine 1: riferimento
  - N–1 baseline: B_i, $i = 2..N$
  - N–1 disparità: d_i, $i = 2..N$
- ***Z*** di **P** è costante → ***1/Z*** costante!

$$Z = \frac{B \cdot f}{x - x'} = \frac{B \cdot f}{d}$$



$$\frac{1}{Z_P} = \frac{1}{f}\frac{d_i}{B_i}, \quad i = 2..N$$

$$d_i = x_i - x_1, \quad B_i = |O_i - O_1|$$



Per ogni punto x_P nell'immagine 1 calcolo

SSD multi-baseline: $SSD\left(x_P, \frac{1}{Z}\right)$

$$SSD\left(x_P, \frac{1}{Z}\right) = \sum_{i=2}^{N} SSD\left(\mathbf{w}_1(x_P), \mathbf{w}_i(x_P + d_i)\right)$$

$$dove: \quad d_i = \frac{B_i}{f}\frac{1}{Z}$$

Multiple-baseline stereo

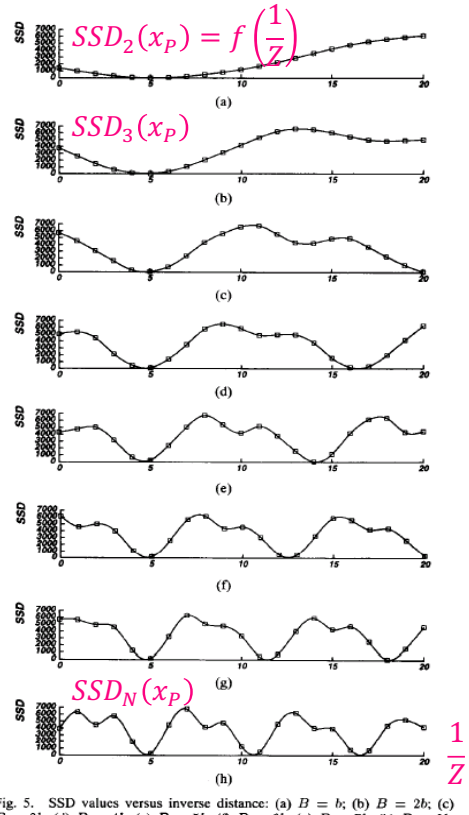


Fig. 5. SSD values versus inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

Sommando tutti i risultati a **tutte le baseline**, solo nel valore **1/Z** corretto hanno tutti un minimo!

$$SSD(x_P) = \sum_{i=2}^{N} SSD_i(x_P)$$

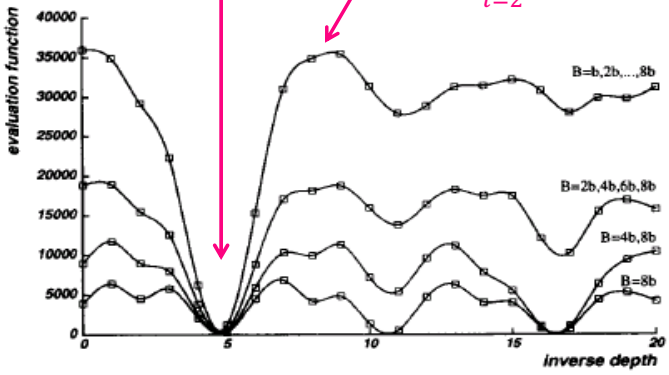
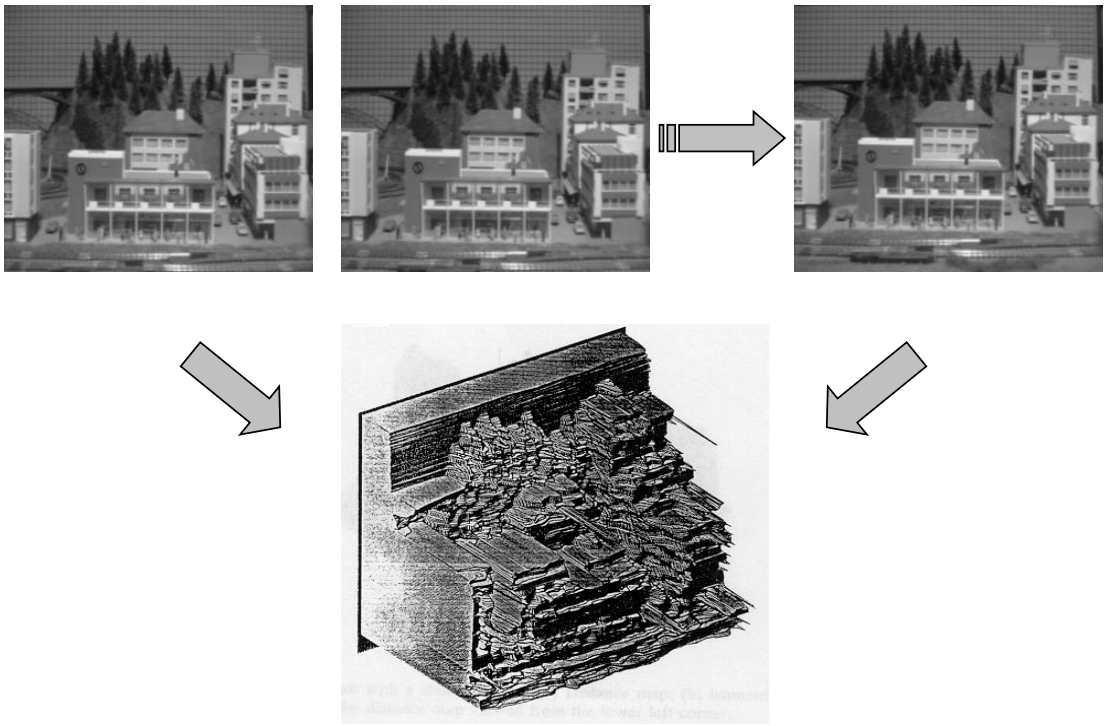


Fig. 7. Combining multiple baseline stereo pairs.

Multiple-baseline stereo



M. Okutomi and T. Kanade, "A Multiple-Baseline Stereo System", IEEE Trans. on Pattern Analysis and Machine Intelligence, 15(4):353-363 (1993).