Geometria per la ricostruzione tridimensionale da immagini

Quando i matematici non entrano in aula

Cristina Turrini

UNIMI

Trento, 9 aprile 2017

- C. T. - Geometria per la ricostruzione tridimensionale da immagini
- Luca Magri - Ricostruzione tridimensionale da immagini

PROBLEMA: data una serie di immagini della stessa scena




ricostruire la posizione dei diversi oggetti nello spazio.

index

- insieme di punti dello spazio:

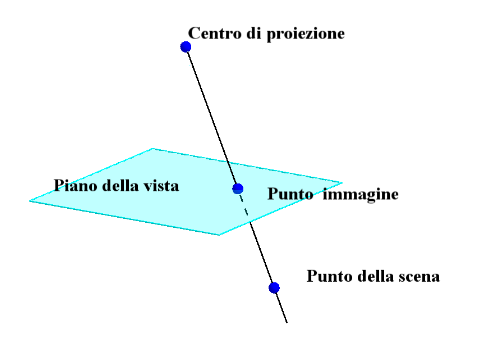
  la scena

- insieme di macchine fotografiche "pinhole":

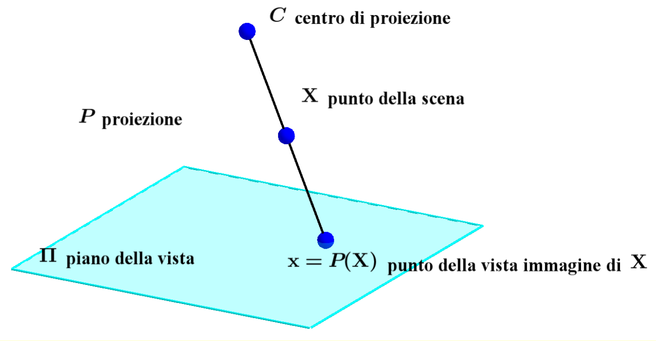
  le camere

- piani delle fotografie:

  le viste



Il caso di una vista



- *la scena*
- *la camera di centro* **C**
- *la vista*

La retta per **C** e **X** viene detta *raggio* per **X**.

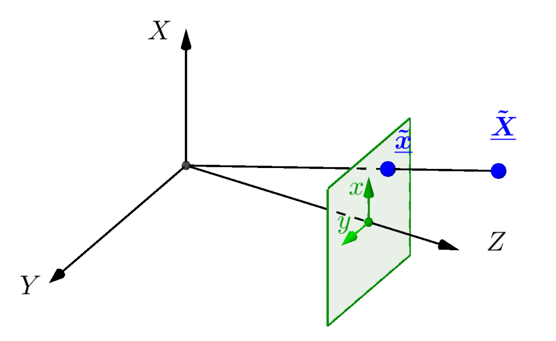
Proiezione standard

$\widetilde{\mathbf{X}} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ punto della scena

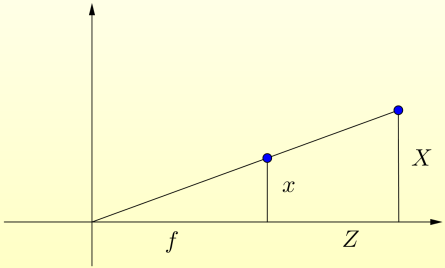
P proiezione di centro $\widetilde{\mathbf{C}}$ (origine del sistema di riferimento)

$\widetilde{\mathbf{x}} = P(\widetilde{\mathbf{X}}) = \begin{pmatrix} x \\ y \end{pmatrix}$ punto immagine

f distanza focale



$$X : x = Z : f, \qquad x = \frac{fX}{Z}$$



$$\widetilde{\mathbf{X}} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad \longrightarrow \quad \widetilde{\mathbf{x}} = \begin{pmatrix} \frac{fX}{Z} \\ \\ \frac{fY}{Z} \end{pmatrix}$$

Coordinate omogenee

nello spazio $\quad \mathbf{X} = \begin{pmatrix} \widetilde{\mathbf{X}} \\ \\ 1 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} \rho X \\ \rho Y \\ \rho Z \\ \rho 1 \end{pmatrix}$

e nel piano $\quad \mathbf{x} = \begin{pmatrix} \widetilde{\mathbf{x}} \\ \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} \mu x \\ \mu y \\ \mu 1 \end{pmatrix}.$

La proiezione in coordinate omogenee

$$P(\mathbf{X}) = P \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} =$$

$$= \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{pmatrix} fX/Z \\ fY/Z \\ 1 \end{pmatrix} = \mathbf{x}$$

Posto

$$K = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

e

$$[I|\mathbf{0}] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

la proiezione diventa

$$P(\mathbf{X}) = K \cdot [I|\mathbf{0}] \cdot \mathbf{X},$$

ove K è detta *matrice di calibrazione*.

Se il sistema della vista non è centrato nel *punto principale*:

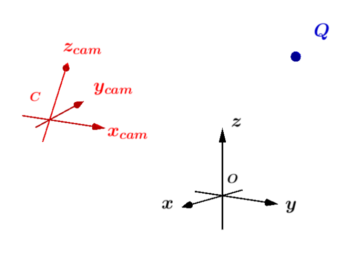
$$K = \begin{pmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{pmatrix}.$$

Se il sistema di coordinate non è monometrico e ortogonale:

$$K = \begin{pmatrix} f_x & s & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{pmatrix}.$$

Sistemi di riferimento della camera e del mondo

I punti della scena, in generale, hanno coordinate assegnate rispetto a un sistema di riferimento indipendente dalla camera.



Il legame tra il sistema di riferimento fissato nello spazio ($\widetilde{\mathbf{X}}_{mondo}$) e il sistema di riferimento della cammera ($\widetilde{\mathbf{X}}_{cam}$) è dato da una rototraslazione

$$\widetilde{\mathbf{X}}_{cam} = R\widetilde{\mathbf{X}}_{mondo} + \mathbf{t},$$

ove R è la matrice di una rotazione (l'orientamento della camera rispetto al mondo), e $\mathbf{t}$ è il vettore di traslazione (coordinate dell'origine del mondo nel sistema di riferimento della camera).

In coordinate omogenee

$$\mathbf{X}_{cam} = \begin{pmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \cdot \mathbf{X}_{mondo}.$$

Combinando la trasfomazione di coordinate "mondo" $\rightarrow$ "camera" con la proiezione prima descritta:

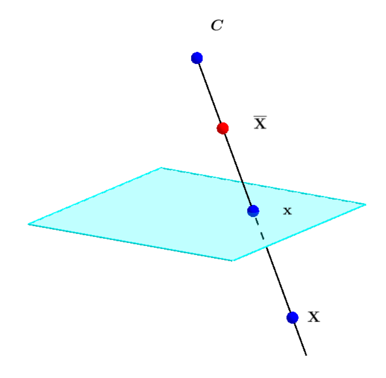
$$P(\mathbf{X})_{cam} = K \cdot [I|\mathbf{0}] \cdot \mathbf{X}_{cam}$$

si arriva alla forma più generale di proiezione

$$P(\mathbf{X}_{mondo}) = K \cdot [R|\mathbf{t}] \cdot \mathbf{X}_{mondo}$$

in cui la matrice della proiezione $P = K \cdot [R|\mathbf{t}]$ è prodotto di K (parametri interni) per $[R|\mathbf{t}]$ (posizionamento della camera).

Nota la matrice di proiezione P si ottengono:



- il centro: è l'annullatore destro della matrice

$$P \cdot C = 0;$$

- un punto della scena $\overline{\mathbf{X}}$ che si proietta su $\mathbf{x}$

$$\overline{\mathbf{X}} = P^{+} \cdot \mathbf{x}$$

(ove P^{+} è una quasi-inversa di P, ossia $P \cdot P^{+} = I$);

- il raggio che si proietta su $\mathbf{x}$.

Tutti i punti del raggio per $\mathbf{X}$ hanno la stessa immagine di $\mathbf{X}$:

con una sola vista la ricostruzione non è possibile !

index

Punti corrispondenti

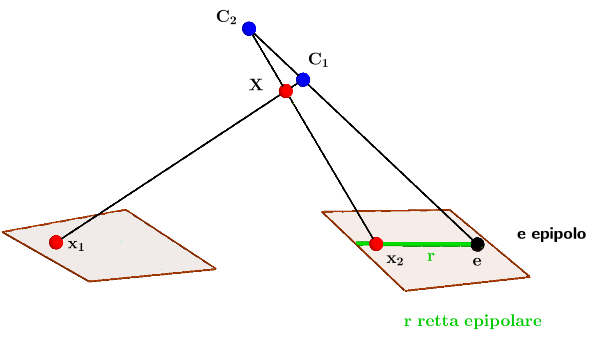
Due viste della stessa scena.

Punti delle due viste che sono immagini dello stesso punto della scena vengono detti *corrispondenti*.

La geometria delle due viste, o *geometria epipolare*, descrive le relazioni che sussistono tra i punti corrispondenti nelle due viste.

la retta epipolare

Il punto $\mathbf{X}$ viene proiettato dalla camera di centro $\mathbf{C}_1$ sul punto $\mathbf{x}_1$ della prima vista e dalla camera di centro $\mathbf{C}_2$ sul punto $\mathbf{x}_2$ della seconda vista.
Il punto $\mathbf{e}$ è l'immagine nella seconda vista del centro della camera $\mathbf{C}_1$.



Tutti i punti della retta per $\mathbf{C}_1$ e $\mathbf{X}$ hanno la stessa immagine $\mathbf{x}_1$ nella prima vista.
Tutti i punti della retta per $\mathbf{e}$ e $\mathbf{x}_2$ (*retta epipolare*) sono corrispondenti di $\mathbf{x}_1$.

La matrice fondamentale

$\mathbf{x}_1 = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$ e $\mathbf{x}_2 = \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix}$ punti corrispondenti.

Esiste $\mathbf{X}$ tale che $P_1 \cdot \mathbf{X} = \lambda \mathbf{x}_1$ e $P_2 \cdot \mathbf{X} = \mu \mathbf{x}_2$.

Il sistema lineare omogeneo (di 6 equazioni in 6 incognite)

$$\begin{pmatrix} P_1 & -\mathbf{x}_1 & \mathbf{0} \\ P_2 & \mathbf{0} & -\mathbf{x}_2 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{X} \\ \lambda \\ \mu \end{pmatrix} = \mathbf{0}$$

ha una soluzione non banale.

Pertanto si ha

$$(*) \qquad \det \begin{pmatrix} P_1 & -\mathbf{x}_1 & \mathbf{0} \\ P_2 & \mathbf{0} & -\mathbf{x}_2 \end{pmatrix} = 0.$$

Il primo membro di $(*)$ è un'espressione lineare sia nelle coordinate di $\mathbf{x}_1$ che in quelle di $\mathbf{x}_2$; quindi $(*)$ si può scrivere nella forma

$$(x', y', 1) \cdot F \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0,$$

dove F è una matrice quadrata 3×3 (definita a meno di una costante moltiplicativa), detta *matrice fondamentale*.

Fissato $\bar{\mathbf{x}}_1 = \begin{pmatrix} a \\ b \\ 1 \end{pmatrix}$, $\quad (x', y', 1) \cdot F \cdot \begin{pmatrix} a \\ b \\ 1 \end{pmatrix} = 0,$

è l'equazione della retta epipolare corrispondente a $\bar{\mathbf{x}}_1$ nella seconda vista e,

Fissato $\bar{\mathbf{x}}_2 = \begin{pmatrix} a' \\ b' \\ 1 \end{pmatrix}$, $\quad (a', b', 1) \cdot F \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0,$

è l'equazione della retta epipolare corrispondente a $\bar{\mathbf{x}}_2$ nella prima vista.

Ai punti del piano della prima vista corrispondono rette del fascio per l'epipolo: F non rappresenta un'applicazione biunivoca, $\det(F) = 0$.

index

Dati: "abbastanza" coppie di punti corrispondenti nelle due viste.

Passi della ricostruzione

1) Determinazione della matrice fondamentale
Ogni coppia $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2)$ di punti corrispondenti fornisce una relazione

$$\bar{\mathbf{x}}_2^T \cdot F \cdot \bar{\mathbf{x}}_1 = (a', b', 1) \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} a \\ b \\ 1 \end{pmatrix} = 0,$$

che può essere interpretata come equazione lineare nelle entrate incognite di F

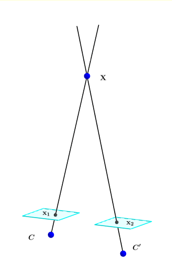
$$(a'a, a'b, a', b'a, b'b, b', a, b, 1) \begin{pmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{pmatrix} = 0.$$

Con un numero sufficiente di tali coppie, si determina F.

2) Ricostruzione delle camere
La matrice fondamentale F determina la coppia di camere (P_1, P_2) solo a meno un cambiamento nel sistema di riferimento proiettivo.

Nota F, si costruisce una coppia di matrici di proiezione (P, P') per cui F è matrice fondamentale
Si fissa una camera come $P = [I, \mathbf{0}]$ e si mostra che si può prendere come altra camera $P' = [[\mathbf{e}_2]_\times \cdot F \quad | \quad \mathbf{e}_2]$ ove $\mathbf{e}_2$ è l'epipolo della seconda vista e $[\mathbf{e}_2]_\times$ è un'opportuna matrice 3×3 che si costruisce a partire dalle coordinate di $\mathbf{e}_2$.

3) Ricostruzione della scena

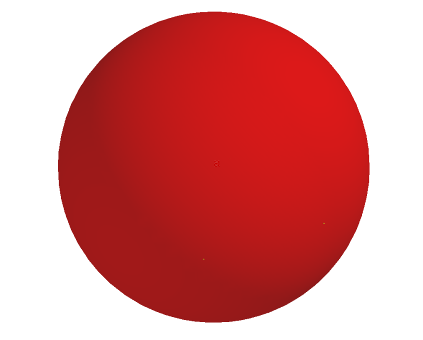


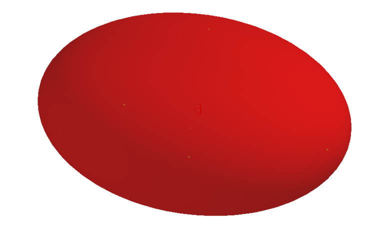
Da ciascuna matrice di proiezione si ricavano il centro di proiezione e i raggi proiettanti i punti della rispettiva vista. Data una coppia $(\mathbf{x}_1, \mathbf{x}_2)$ di punti corrispondenti nelle due viste, il punto $\mathbf{X}$ della scena di cui $\mathbf{x}_1$ e $\mathbf{x}_2$ sono immagini si ottiene intersecando i due raggi (*triangolazione*).

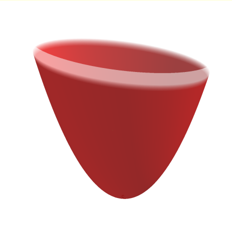
index

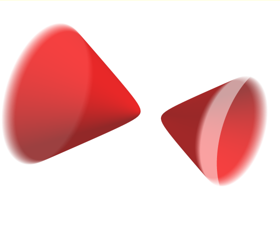
La matrici di proiezione P e P' ottenute nella ricostruzione non sono, in generale le vere camere di partenza, lo sono solo a meno di cambiamento nel sistema di riferimento proiettivo. Di conseguenza anche la ricostruzione della scena è tale solo meno di proiettività.

Sufficiente per alcuni aspetti, ma molto insoddisfacente per altri ...

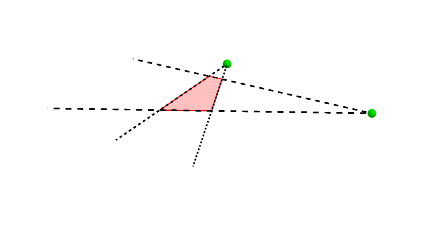






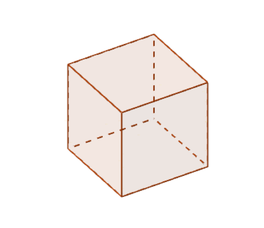


Per ricostruzioni più accurate occorre una migliore conoscenza della scena.



Per una ricostruzione a meno di affinità: 3 punti all'infinito (3 coppie di rette parallele) indipendenti individuano il piano all'infinito.

Per una ricostruzione a meno di similitudine (l'ambiguità di scala è ineliminabile!): 3 direzioni ortogonali e rapporti tra lunghezze.



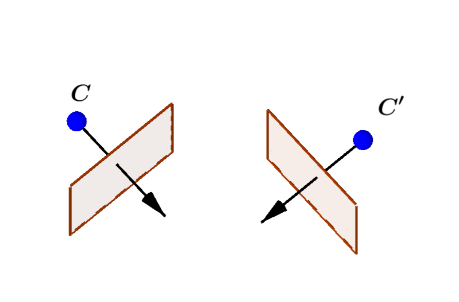
Se si conoscono i parametri interni delle camere (matrici di calibrazione K_1 e K_2) si può fare di meglio (anche a prescindere da ulteriori informazioni sulla scena):

si ricava F come visto sopra (da abbastanza coppie di punti corrispondenti)

e si considera la *matrice essenziale*

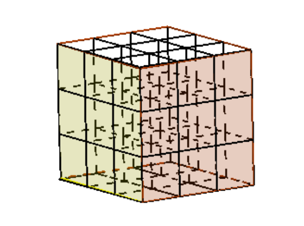
$$E = K_2^T \cdot F \cdot K_1.$$

E contiene tutte le informazioni sul mutuo posizionamento delle due camere.



index

Nel caso di più viste:
strumenti analoghi alla matrice fondamentale:
tensore trifocale, tensore quadrifocale ...



oppure si considerano le diverse coppie di viste.

index

Unici dati significativi, oltre alle posizione degli oggetti nella scena, sono i moduli delle loro velocità.

$$\mathbf{X}(t) = \begin{pmatrix} X + \lambda t dX \\ Y + \lambda t dX \\ Z + \lambda t dZ \\ 1 \end{pmatrix}$$

$$P \cdot \mathbf{X}(t) = P \cdot \begin{pmatrix} X + \lambda t dX \\ Y + \lambda t dX \\ Z + \lambda t dZ \\ 1 \end{pmatrix} = (\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4) \cdot \mathbf{X}(t) =$$

$$(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4, tdX\mathbf{p}_1 + tdY\mathbf{p}_2 + tdZ\mathbf{p}_3) \cdot \begin{pmatrix} X \\ Y \\ Z \\ 1 \\ \lambda \end{pmatrix}$$

Spazio a quattro dimensioni (5 coordinate omogenee): $\begin{pmatrix} X \\ Y \\ Z \\ 1 \\ \lambda \end{pmatrix} \in \mathbf{P}^4$.

Camera: matrice 3×5 che rappresenta una proiezione dallo spazio a 4 dimensioni sul piano della vista.

Ricostruzione nello spazio a 4 dimensioni: posizione dei punti della scena all'istante t_0 e rispettive velocità.