

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

---

SCUOLA DI SCIENZE  
Corso di Laurea in Matematica

# GEOMETRIA DELLA VISIONE

Tesi di Laurea in Geometria

Relatore:  
Chiar.ma Prof.ssa  
Rita Fioresi

Presentata da:  
Luca Grementieri

I Sessione  
Anno Accademico 2014-2015

*A Silvia*

# Indice

<b>Introduzione</b>	<b>5</b>
<b>1 Modellizzazione del moto di una telecamera</b>	<b>7</b>
1.1 Gruppi matriciali e algebre di Lie . . . . .	7
1.2 Esponenziale di matrici . . . . .	9
1.3 Il gruppo ortogonale speciale . . . . .	11
1.4 $SO(3)$ e $\mathfrak{so}(3)$ . . . . .	14
1.5 Rappresentazione del moto di corpo rigido . . . . .	16
<b>2 Modello di telecamera</b>	<b>17</b>
2.1 Formazione dell'immagine . . . . .	17
2.2 Parametri intrinseci e estrinseci . . . . .	19
2.3 Rappresentazione di punti e rette . . . . .	21
2.4 Punti corrispondenti . . . . .	23
<b>3 Geometria epipolare</b>	<b>25</b>
3.1 Matrice essenziale . . . . .	25
3.2 Lo spazio essenziale . . . . .	28
3.3 Matrice fondamentale . . . . .	31
3.4 Ricostruzione proiettiva . . . . .	33
<b>Bibliografia</b>	<b>35</b>
<b>Ringraziamenti</b>	<b>37</b>



# Introduzione

La visione é il processo cerebrale mediante il quale l'organismo umano riesce a estrarre informazioni dal dato visivo proveniente dalla retina. É un processo molto complesso e di cui abbiamo attualmente solo una minima conoscenza. La capacità di generalizzazione e apprendimento del cervello umano é fondamentale per la visione e permette il riconoscimento di oggetti anche in situazioni di luminosità molto diverse rispetto a quelle in cui si trovava l'oggetto quando lo si è visto la prima volta.

Tentare di imitare questo comportamento mediante un elaboratore elettronico, il cosiddetto problema della visione, é una delle maggiori sfide del XXI secolo. Per comprendere quanto possa essere difficile risolvere questo problema, basti pensare che metà del cervello umano é continuamente occupato ad analizzare le immagini provenienti dai nostri occhi. Un elaboratore é capace di eseguire solo brevi precisi comandi, non ha quindi una naturale capacità di generalizzazione. Questa deve essere algoritmicamente programmata, ma é un compito tutt'altro che facile: infatti, poiché la conoscenza umana del cervello é tuttora molto ridotta, non é possibile procedere imitando le strutture biologiche adibite allo stesso scopo.

In questo contesto questa tesi tratta di una delle pochissime parti matematicamente formalizzata del processo della visione. La teoria qui sviluppata discende da studi che risalgono già all'inizio XX secolo. Lo scopo della tesi é dare una descrizione degli strumenti matematici che permettono di modellizzare la visione stereoscopica ed esporre le condizioni sotto le quali sia possibile effettuare una ricostruzione 3D ambientale a partire da due riprese della stessa scena.

Il cervello é capace di percepire la profondità degli oggetti e di collocarli nello spazio tridimensionale a partire dalle due sole immagini provenienti dagli occhi. In particolare il cervello non ha difficoltà a capire quali siano i punti delle immagini corrispondenti allo stesso punto nello spazio. Bisogna però notare che il problema della ricostruzione 3D é più generale rispetto a quello risolto dal cervello: infatti la posizione reciproca degli occhi é sempre fissa e l'acutezza visiva assicurata dalla retina é incomparabile rispetto a quella delle più moderne telecamere.

Nella tesi si vedrà come sia possibile trovare un algoritmo in grado di ottenere una ricostruzione 3D della scena a partire da due immagini nell'ipotesi di assenza di errore. Questo problema si riduce a trovare la posizione reciproca delle due telecamere che hanno

ripreso le immagini perché, una volta noto questo dato, è possibile calcolare la posizione di un qualsiasi punto nello spazio mediante l'algoritmo di triangolazione. La ricostruzione ottenuta risulta euclidea, cioè fedele alla scena reale, nel caso si abbia conoscenza della matrice di calibrazione della telecamera che ha ripreso le due immagini. Nell'eventualità in cui non si abbia questo dato aggiuntivo, la ricostruzione migliore ottenibile è solo proiettivamente equivalente alla scena reale, esiste quindi una proiettività che permette di ottenere la ricostruzione euclidea.

La tesi è strutturata in tre capitoli.

Nel Capitolo 1 si analizza il moto del corpo rigido. Questo studio preliminare è necessario per capire come descrivere la posizione reciproca delle telecamere. Per ottenere una parametrizzazione del gruppo delle rotazioni si introduce lo strumento delle algebre di Lie e se ne illustrano le principali proprietà.

Nel Capitolo 2 si formalizza il concetto di immagine e si descrive il modello di telecamera scelto per descrivere la formazione dell'immagine. Questa descrizione permette di capire quali siano i parametri sufficienti a descrivere una telecamera. Nel caso in cui questi siano noti si parla di telecamera calibrata. Inoltre si introduce una rappresentazione di punti e rette dell'immagine che permette di descrivere le relazioni di incidenza e di appartenenza algebricamente. Infine viene spiegato come sia possibile selezionare un insieme ridotto di punti di cui sia possibile individuare la posizione in entrambe le immagini. Questo è un punto fondamentale: infatti si vedrà che otto coppie di punti corrispondenti sono sufficienti a ricavare la posizione reciproca di due telecamere calibrate.

Infine nel Capitolo 3 viene descritta la geometria epolare e gli strumenti principali utilizzati per la sua descrizione: la matrice essenziale nel caso calibrato e la matrice fondamentale nel caso non calibrato. L'importanza di queste matrici deriva dal fatto che entrambe possono essere stimate a partire da otto corrispondenze di punti e inoltre la loro scomposizione permette di ottenere proprio la posizione reciproca delle telecamere.

In questo lavoro di tesi viene ignorato il problema dato dagli errori di misurazione sulla posizione dei punti nelle immagini. Questo errore è inevitabile perché, anche se matematicamente è comodo modellizzare un'immagine come una funzione definita su un compatto di  $\mathbb{R}^2$ , nella realtà le immagini sono griglie discrete di pixel. Ciò nonostante esistono algoritmi basati sulla stessa teoria capaci di gestire anche il caso reale con errore sulle misurazioni. Questi algoritmi producono una soluzione approssimata minimizzando un'opportuna funzione e restituiscono il risultato finale proiettando la soluzione approssimata sullo spazio a cui, secondo la teoria, deve appartenere la soluzione. La scelta di non dilungarsi su queste tematiche deriva dalla volontà di dare un taglio profondamente matematico all'esposizione.

Inoltre è necessario citare un'ulteriore ipotesi implicita fatta sulla ripresa della scena: nel caso analizzato si presuppone che tutti gli oggetti non cambino posizione nelle due riprese. In questo caso, di complessità molto maggiore, il problema è notevolmente difficile anche nell'ipotesi di assenza di errore.

# Capitolo 1

## Modellizzazione del moto di una telecamera

Per gli scopi di questa tesi si possono modellizzare le telecamere come corpi rigidi ed è quindi necessario studiare le trasformazioni che descrivono il moto di un corpo rigido.

La caratteristica di un corpo rigido è che se  $X(t)$  e  $Y(t)$  sono le coordinate di due qualsiasi punti  $P$  e  $Q$  del corpo rigido in dipendenza dal tempo, allora la loro distanza è costante:

$$\|X(t) - Y(t)\| \equiv \text{costante} \quad \forall t \in \mathbb{R}.$$

Per questo motivo, per descrivere la posizione nel tempo del corpo rigido è sufficiente conoscere il moto di un unico punto  $P$  e il movimento di tre assi coordinati con origine in  $P$  solidali con il corpo relativamente a un sistema di riferimento fisso, scelto arbitrariamente. Dunque la configurazione della telecamera è determinata da due componenti:

- la parte *traslazionale*: il vettore  $T$  tra l'origine del sistema di riferimento solidale alla telecamera e quello fisso;
- la parte *rotazionale*: la rotazione del sistema di riferimento solidale alla telecamera rispetto a quello fisso.

### 1.1 Gruppi matriciali e algebre di Lie

Si inizia ora ad approfondire lo studio della parte rotazionale. Le rotazioni nello spazio 3D sono un gruppo con la composizione isomorfo ad un'importante gruppo matriciale; in questa sezione si definisce questo concetto e si illustrano le principali proprietà. Inoltre si illustra il concetto di algebra di Lie, un potente strumento che permette di trattare spazi non lineari in modo lineare.

Per tutto il capitolo si lavorerà sul campo dei numeri reali  $\mathbb{R}$ , bisogna però notare che tutte le definizioni e molti dei teoremi qui esposti valgono anche per un generico

campo  $\mathbb{K}$ . Si consideri  $M_n(\mathbb{R})$  lo spazio vettoriale delle matrici quadrate  $n \times n$  dotato della norma

$$\|A\| = \sup \left\{ \frac{\|Ax\|}{\|x\|} \mid x \in \mathbb{R}^n \setminus \{0\} \right\}.$$

Vale il seguente risultato per la dimostrazione si veda [1] (Theorem 1.12 pag. 10).

**Teorema 1.1.** *Lo spazio metrico  $(M_n(\mathbb{R}), \|\cdot\|)$  è completo.*

Si definisce il gruppo delle matrici  $n \times n$  invertibili

$$GL_n(\mathbb{R}) = \{A \in M_n(\mathbb{R}) \mid \det A \neq 0\}.$$

**Definizione 1.2.** Un sottogruppo  $G \subseteq GL_n(\mathbb{R})$  si dice *gruppo matriciale* se è un sottospazio chiuso rispetto alla topologia indotta da  $M_n(\mathbb{R})$  su  $GL_n(\mathbb{R})$ . Un sottogruppo chiuso  $H \subseteq G$  si dice *sottogruppo matriciale* di  $G$ .

**Definizione 1.3.** Si dice *algebra di Lie* una coppia  $(\mathfrak{g}, [\cdot, \cdot])$  dove  $\mathfrak{g}$  è uno spazio vettoriale e  $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$  è una funzione bilineare detta *parentesi di Lie* che soddisfa le seguenti proprietà:

- $[x, y] = -[y, x] \quad \forall x, y \in V$  (proprietà antisimmetrica)
- $[x, [y, z]] + [y, [x, z]] + [z, [x, y]] = 0 \quad \forall x, y, z \in V$  (identità di Jacobi)

Sia  $\mathfrak{a}$  una algebra di Lie, uno spazio vettoriale  $\mathfrak{b} \subseteq \mathfrak{a}$  è una *sottoalgebra di Lie* di  $\mathfrak{a}$  se  $\forall x, y \in \mathfrak{b}, [x, y] \in \mathfrak{b}$ .

La funzione bilineare *commutatore*

$$[\cdot, \cdot] : M_n(\mathbb{R}) \times M_n(\mathbb{R}) \longrightarrow M_n(\mathbb{R}) \quad (A, B) \mapsto [A, B] = AB - BA$$

soddisfa le condizioni della Definizione 1.3, quindi  $M_n(\mathbb{R})$  è un'algebra di Lie.

**Definizione 1.4.** Sia  $I$  intervallo con  $0 \in I$  e  $G \subseteq GL_n(\mathbb{R})$  un gruppo matriciale, lo *spazio tangente* a  $G$  in  $X \in G$  è

$$T_X G = \{\gamma'(0) \in M_n(\mathbb{R}) \mid \gamma : I \longrightarrow G \text{ è una curva differenziabile in } G \text{ con } \gamma(0) = X\}.$$

**Proposizione 1.5.**  $T_X G$  è un spazio vettoriale di  $M_n(\mathbb{R})$ .

*Dimostrazione.* Siano  $\alpha, \beta$  curve differenziabili in  $G$  tali che  $\alpha(0) = \beta(0) = X$ , si definisce

$$\gamma : I \longrightarrow G \quad \gamma(t) = \alpha(t)X^{-1}\beta(t)$$

curva differenziabile in  $G$  con  $\gamma(0) = X$ . Derivando si ha  $\gamma'(t) = \alpha'(t)X^{-1}\beta(t) + \alpha(t)X^{-1}\beta'(t)$  quindi  $\gamma'(0) = \alpha'(0) + \beta'(0)$  che mostra che  $T_X G$  è chiuso per addizione. Analogamente, sia  $r \in \mathbb{R}$ , si definisce  $\eta(t) = \alpha(rt)$ , curva differenziabile in  $G$  con  $\eta(0) = X$ . Poiché  $\eta'(0) = r\alpha'(0)$ ,  $T_X G$  è chiuso per moltiplicazione per uno scalare.  $\square$

Si indica con  $\mathfrak{g} = \text{Lie}(G) = T_I G$  lo spazio tangente a  $G$  nell'identità,  $\mathfrak{g}$  è dotato della struttura di algebra di Lie e perciò è un potente strumento per l'analisi di  $G$ .

**Teorema 1.6.** *Sia  $G \subseteq GL_n(\mathbb{R})$  un gruppo matriciale, allora  $\mathfrak{g}$  è una sottoalgebra di Lie di  $M_n(\mathbb{R})$ .*

*Dimostrazione.* Siano  $\alpha, \beta$  curve differenziabili in  $G$  tali che  $\alpha(0) = \beta(0) = I_n$ , matrice identità di ordine  $n$ , si considerino  $I, J$  intervalli contenenti 0 e la funzione

$$F : I \times J \longrightarrow G \quad F(s, t) = \alpha(s)\beta(t)\alpha(s)^{-1}.$$

$F$  è continua e differenziabile inoltre fissato  $s \in I$ ,  $F(s, 0)$  è una curva differenziabile in  $G$  tale che  $F(s, 0) = I_n$ . Derivando si ha

$$\left. \frac{dF(s, t)}{dt} \right|_{t=0} = \alpha(s)\beta'(0)\alpha(s)^{-1} \in T_I G = \mathfrak{g}.$$

Poiché  $\mathfrak{g}$  è un sottospazio chiuso si ha che  $\lim_{s \rightarrow 0} \frac{1}{s}(\alpha(s)\beta'(0)\alpha(s)^{-1} - \beta'(0)) \in \mathfrak{g}$ .

$$\begin{aligned} \lim_{s \rightarrow 0} \frac{1}{s}(\alpha(s)\beta'(0)\alpha(s)^{-1} - \beta'(0)) &= \left. \frac{d}{ds} \alpha(s)\beta'(0)\alpha(s)^{-1} \right|_{s=0} \\ &= \alpha'(0)\beta'(0)\alpha(0)^{-1} - \alpha(0)\beta'(0)\alpha(0)^{-1}\alpha'(0)\alpha(0)^{-1} \\ &= \alpha'(0)\beta'(0) - \beta'(0)\alpha'(0) = [\alpha'(0), \beta'(0)] \end{aligned}$$

Dunque  $[\alpha'(0), \beta'(0)] \in \mathfrak{g}$  quindi  $\mathfrak{g}$  è una sottoalgebra di Lie di  $M_n(\mathbb{R})$ . □

Si noti che se  $\mathfrak{g}$  è sottoalgebra di Lie di  $M_n(\mathbb{R})$  non è detto esista  $G \subseteq GL_n(\mathbb{R})$  gruppo matriciale tale che  $\mathfrak{g} = T_I G$ ; però per il Secondo Teorema di Lie questa corrispondenza esiste nel caso non si richieda che il gruppo di matrici sia chiuso. Per maggiori dettagli si veda [2] (Chapter 2).

## 1.2 Esponenziale di matrici

Per poter trovare il legame tra un gruppo matriciale  $G$  e l'algebra di Lie  $\mathfrak{g}$  associata è necessario introdurre la funzione esponenziale, che rappresenta lo strumento fondamentale per trasferire le proprietà dell'algebra di Lie al gruppo corrispondente.

Sia  $A \in M_n(\mathbb{R})$ , si definisce la serie esponenziale

$$\exp A = \sum_{k \geq 0} \frac{1}{k!} A^k = I + A + \frac{1}{2!} A^2 + \frac{1}{3!} A^3 + \dots \quad (1.1)$$

**Proposizione 1.7.**  *$\exp A$  è ben definito  $\forall A \in M_n(\mathbb{R})$ .*

*Dimostrazione.* È noto che  $e^{\|A\|}$  converge  $\forall A \in M_n(\mathbb{R})$ , quindi

$$\forall \varepsilon \exists n, m \text{ con } n > m \text{ tali che } \left| \sum_{k=0}^n \frac{1}{k!} \|A\|^k - \sum_{k=0}^m \frac{1}{k!} \|A\|^k \right| = \sum_{k=m}^n \frac{1}{k!} \|A\|^k < \varepsilon.$$

Poiché  $\left\| \sum_{k=m}^n \frac{1}{k!} A^k \right\| < \sum_{k=m}^n \frac{1}{k!} \|A\|^k$  si ha che

$$\forall \varepsilon \exists n, m \text{ con } n > m \text{ tali che } \left\| \sum_{k=0}^n \frac{1}{k!} A^k - \sum_{k=0}^m \frac{1}{k!} A^k \right\| = \left\| \sum_{k=m}^n \frac{1}{k!} A^k \right\| < \varepsilon$$

quindi la successione delle somme parziali è una successione di Cauchy. Per il Teorema 1.1, la serie esponenziale converge.  $\square$

La curva  $\alpha(t) : \mathbb{R} \rightarrow M_n(\mathbb{R})$   $\alpha(t) = \exp tA$  è ben definita e differenziabile con

$$\alpha'(t) = \sum_{k \geq 1} \frac{t^{k-1}}{(k-1)!} A^k = A \exp tA.$$

Dunque la curva esponenziale è soluzione dell'equazione differenziale di primo ordine  $\alpha'(t) = \alpha(t)A$ , per il Teorema di Cauchy-Peano questa è unica fissato il dato iniziale. La curva  $\alpha$  è fondamentale nella teoria dei gruppi di Lie infatti come si vedrà nella Proposizione 1.8, l'immagine di  $\alpha$  è sottogruppo di  $G$  e si dice *sottogruppo a un parametro*.

Si enunciano ora alcune formule e proprietà necessarie per lo sviluppo della teoria, per la loro dimostrazione si veda [1] (Proposition 2.1 pag. 46, Theorem 7.26 pag. 196).

**Proposizione 1.8.** *Sia  $A \in GL_n(\mathbb{R})$  per ogni  $t, s \in \mathbb{R}$*

$$\exp((t+s)A) = \exp(tA) \exp(sA), \quad (1.2)$$

*in particolare se  $\exp(A) \in GL_n(\mathbb{R})$  allora  $\exp(A)^{-1} = \exp(-A)$ .*

**Teorema 1.9.** *Siano  $U, V \in M_n(\mathbb{R})$  allora valgono le seguenti formule*

$$\exp(U+V) = \lim_{r \rightarrow \infty} \left( \exp\left(\frac{1}{r}U\right) \exp\left(\frac{1}{r}V\right) \right)^r \quad (1.3)$$

$$\exp([U, V]) = \lim_{r \rightarrow \infty} \left( \exp\left(\frac{1}{r}U\right) \exp\left(\frac{1}{r}V\right) \exp\left(-\frac{1}{r}U\right) \exp\left(-\frac{1}{r}V\right) \right)^{r^2} \quad (1.4)$$

Sia  $G \subseteq GL_n(\mathbb{R})$  un sottogruppo matriciale, si definisce

$$\tilde{\mathfrak{g}} = \{A \in M_n(\mathbb{R}) \mid \forall t \in \mathbb{R}, \exp(tA) \in G\}.$$

**Teorema 1.10.**  $\tilde{\mathfrak{g}}$  è una sottoalgebra di Lie di  $M_n(\mathbb{R})$ .

*Dimostrazione.* Per definizione  $\tilde{\mathfrak{g}}$  è chiusa per moltiplicazione per scalare inoltre se  $U, V \in \tilde{\mathfrak{g}}$  e  $r \in \mathbb{N} \setminus \{0\}$  le seguenti matrici sono elementi di  $G$ :

$$\left( \exp\left(\frac{1}{r}U\right) \exp\left(\frac{1}{r}V\right) \right)^r = \left( \exp\left(\frac{1}{r}U\right) \exp\left(\frac{1}{r}V\right) \exp\left(-\frac{1}{r}U\right) \exp\left(-\frac{1}{r}V\right) \right)^{r^2}.$$

Poiché  $G$  è un sottospazio chiuso di  $GL_n(\mathbb{R})$  per la formula (1.3)

$$\exp(tU + tV) = \lim_{r \rightarrow \infty} \left( \exp\left(\frac{1}{r}tU\right) \exp\left(\frac{1}{r}tV\right) \right)^r \in G$$

e analogamente, ricordando che  $t[U, V] = [tU, tV]$  e usando la formula (1.4)

$$\exp(t[U, V]) = \lim_{r \rightarrow \infty} \left( \exp\left(\frac{1}{r}tU\right) \exp\left(\frac{1}{r}tV\right) \exp\left(-\frac{1}{r}tU\right) \exp\left(-\frac{1}{r}tV\right) \right)^{r^2} \in G$$

□

**Proposizione 1.11.** Sia  $G \subseteq GL_n(\mathbb{R})$  un sottogruppo matriciale,  $\tilde{\mathfrak{g}}$  è una sottoalgebra di Lie di  $\mathfrak{g}$ , inoltre  $\tilde{\mathfrak{g}} = \mathfrak{g}$ .

*Dimostrazione.* Si dimostra solo la prima affermazione, per la dimostrazione della seconda parte si veda [1] (Theorem 7.31 pag. 202).

Sia  $U \in \tilde{\mathfrak{g}}$ , allora la curva  $\gamma(t) = \exp(tU) \in G$  ha  $\gamma(0) = I$  e  $\gamma'(0) = U$  quindi  $U \in G$ . □

Sviluppando la teoria dei gruppi di Lie è possibile dimostrare il seguente risultato, si veda [1] (Theorem 10.16 pag. 258).

**Teorema 1.12.** Sia  $G \subseteq GL_n(\mathbb{R})$  un gruppo matriciale e  $\exp : \mathfrak{g} \rightarrow M_n(\mathbb{R})$  la funzione esponenziale, allora  $\text{Im } \exp \subseteq G$ ; inoltre se  $G$  è compatto e connesso  $\exp$  è suriettiva.

### 1.3 Il gruppo ortogonale speciale

Il gruppo delle rotazioni e più in generale il gruppo ortogonale, che comprende trasformazioni più generali come le simmetrie assiali, è di fondamentale importanza nello studio del moto del corpo rigido. In questa sezione lo studio di questo gruppo permette di trovarne una comoda parametrizzazione.

Si definisce  $O(n) = \{A \in GL_n(\mathbb{R}) \mid A^\top A = I_n\}$  il *gruppo ortogonale*; è facile verificare che  $O(n)$  è un gruppo matriciale inoltre si ha

$$(\det A)^2 = \det A^\top \det A = \det(A^\top A) = \det I_n = 1 \quad \forall A \in O(n)$$

da cui si ricava che  $\det A = 1$  o  $\det A = -1$ . È quindi possibile definire il sottogruppo matriciale detto *gruppo ortogonale speciale*

$$SO(n) = \{A \in O(n) \mid \det A = 1\} \subseteq O(n).$$

**Lemma 1.13.** *Sia  $I$  intervallo con  $0 \in I$  e sia  $\alpha : I \rightarrow GL_n(\mathbb{R})$  una curva differenziabile tale che  $\alpha(0) = I_n$ , matrice identità  $n \times n$ , allora*

$$\left. \frac{d(\det \alpha(t))}{dt} \right|_{t=0} = \text{tr } \alpha'(0). \quad (1.5)$$

*Dimostrazione.* Detta  $C_{ij}(t)$  la matrice ottenuta da  $\alpha(t)$  eliminando la  $i$ -esima riga e la  $j$ -esima colonna, per lo sviluppo di Laplace del determinante si ha  $\det \alpha(t) = \sum_{j=1}^n (-1)^{n+j} a_{nj} \det C_{nj}$ . Sia  $a_{ij}(t) = \alpha(t)_{ij}$ , poiché  $\alpha(0) = I_n$  si ha  $a_{nj}(0) = \delta_{nj}$  e  $\det C_{nj}(0) = \delta_{nj}$  quindi

$$\begin{aligned} \left. \frac{d(\det \alpha(t))}{dt} \right|_{t=0} &= \sum_{j=1}^n (-1)^{n+j} \left. \frac{da_{nj}}{dt} \right|_{t=0} \det C_{nj}(0) + a_{nj}(0) \left. \frac{d \det C_{nj}}{dt} \right|_{t=0} \\ &= \sum_{j=1}^n (-1)^{n+j} \left. \frac{da_{nj}}{dt} \right|_{t=0} \det C_{nj}(0) + \left. \frac{d \det C_{nn}}{dt} \right|_{t=0} \\ &= \left. \frac{da_{nn}}{dt} \right|_{t=0} + \left. \frac{d \det C_{nn}}{dt} \right|_{t=0} \end{aligned}$$

Applicando la formula a  $C_{nn}$ , matrice quadrata di dimensione  $(n-1)$ , e procedendo a ritroso si ottiene

$$\left. \frac{d(\det \alpha(t))}{dt} \right|_{t=0} = \sum_{i=1}^n \left. \frac{da_{ii}}{dt} \right|_{t=0} = \text{tr } \alpha'(0).$$

□

**Proposizione 1.14.** *Sia  $A \in M_n(\mathbb{R})$  allora  $\det \exp A = e^{\text{tr } A}$ .*

*Dimostrazione.* Sia  $\gamma : \mathbb{R} \rightarrow GL_1(\mathbb{R}) = \mathbb{R} \setminus \{0\}$   $\gamma(t) = \det \exp tA$ , applicando l'equazione (1.5) si ottiene

$$\begin{aligned} \gamma'(t) &= \lim_{h \rightarrow 0} \frac{1}{h} (\det \exp((t+h)A) - \det \exp(tA)) = \det \exp(tA) \lim_{h \rightarrow 0} \frac{1}{h} (\det \exp(hA) - 1) \\ &= \det \exp(tA) \left. \frac{d(\det \exp(hA))}{dh} \right|_{h=0} \\ &= \det \exp(tA) \text{tr } A \\ &= \gamma \text{tr } A. \end{aligned}$$

$\gamma$  soddisfa la stessa equazione differenziale e le stesse condizioni iniziali della curva  $t \mapsto e^{t \operatorname{tr} A}$ , quindi per il teorema di Cauchy  $\det \exp(tA) = e^{t \operatorname{tr} A}$  da cui segue la tesi per  $t = 0$ .  $\square$

Si definisce  $\operatorname{Antisym}_n(\mathbb{R}) = \{A \in M_n(\mathbb{R}) \mid A^\top = -A\}$ ;  $\operatorname{Antisym}_n(\mathbb{R})$  è uno spazio vettoriale e soprattutto è sottoalgebra di Lie di  $M_n(\mathbb{R})$ .

**Teorema 1.15.**  $\mathfrak{so}(n) = T_I SO(n) = \operatorname{Antisym}_n(\mathbb{R})$ .

*Dimostrazione.* Sia  $\alpha : [a, b] \rightarrow SO(n)$  una curva differenziabile tale che  $\alpha(0) = I_n$ , si ha

$$\frac{d}{dt} \alpha(t)^\top \alpha(t) = 0 \Rightarrow \alpha'(0)^\top \alpha(t) + \alpha(t)^\top \alpha'(t) = 0.$$

Valutando in 0 si ottiene

$$\alpha'(0)^\top + \alpha'(0) = 0 \Rightarrow \alpha'(0)^\top = -\alpha'(0) \Rightarrow \alpha'(0) \in \operatorname{Antisym}_n(\mathbb{R}).$$

Quindi  $\mathfrak{so}(n) = T_I SO(n) \subseteq \operatorname{Antisym}_n(\mathbb{R})$ . D'altra parte se  $A \in \operatorname{Antisym}_n(\mathbb{R})$ , fissato  $\varepsilon > 0$  si consideri la curva

$$\alpha : [-\varepsilon, \varepsilon] \rightarrow GL_n(\mathbb{R}) \quad \alpha(t) = \exp(tA),$$

allora

$$\alpha(t)^\top \alpha(t) = \exp(tA)^\top \exp(tA) = \exp(tA^\top) \exp(tA) = \exp(-tA) \exp(tA) = I$$

Inoltre per  $A \in \operatorname{Antisym}_n(\mathbb{R})$  si ha  $\operatorname{tr} A = \operatorname{tr}(A^\top) = \operatorname{tr}(-A) = -\operatorname{tr} A \Rightarrow \operatorname{tr} A = 0$ , quindi per la Proposizione 1.14  $\det \exp(tA) = 1$ . Ne consegue che  $\alpha$  è una curva di  $SO(n)$  e poiché  $\alpha'(0) = A$ ,  $\operatorname{Antisym}_n(\mathbb{R}) \subseteq \mathfrak{so}(n) = T_I SO(n)$  da cui la tesi.  $\square$

Si noti che dalla dimostrazione discende anche il fatto che se  $A \in \operatorname{Antisym}_n(\mathbb{R})$  allora  $\exp(tA) \in SO(n) \forall t \in \mathbb{R}$ , come previsto dal Teorema 1.12. Inoltre vale il seguente Teorema per la cui dimostrazione si rimanda a [3] (Theorem 3.67 pag. 130).

**Teorema 1.16.**  $SO(n)$  è compatto e connesso.

**Teorema 1.17.** La funzione  $\exp : \operatorname{Antisym}_n(\mathbb{R}) \rightarrow SO(n)$  è suriettiva.

*Dimostrazione.* Per il Teorema 1.16  $SO(n)$  è compatto e connesso, dunque dal Teorema 1.12 discende la tesi.  $\square$

Questo ultimo risultato è fondamentale per gli scopi di questa tesi in quanto mostra che è possibile parametrizzare le rotazioni, cioè gli elementi del gruppo  $SO(3)$ , come esponenziali di matrici antisimmetriche.

## 1.4 $SO(3)$ e $\mathfrak{so}(3)$

Le matrici antisimmetriche di  $\mathfrak{so}(3)$  hanno proprietà aggiuntive rispetto agli elementi dei generici  $\mathfrak{so}(n)$ , questo perché in  $\mathbb{R}^3$  è possibile definire un prodotto vettoriale.

**Definizione 1.18.** Siano  $u, v \in \mathbb{R}^3$  il loro *prodotto vettoriale* è il vettore

$$u \times v = \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix} \in \mathbb{R}^3.$$

Fissato  $u$  la funzione  $v \mapsto u \times v$  è lineare in  $v$  e può quindi essere rappresentata da una matrice denotata  $\hat{u} \in M_3(\mathbb{R})$ . È facile vedere che

$$\hat{u} = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix} \in M_3(\mathbb{R}).$$

Si noti che  $u \times u = 0$ , dunque in forma matriciale si ha

$$\hat{u}u = 0 \quad \forall u \in \mathbb{R}^3. \quad (1.6)$$

Si osservi che  $M \in M_3(\mathbb{R})$  è antisimmetrica se e solo se esiste  $u \in \mathbb{R}^3$  tale che  $M = \hat{u}$ , dunque è possibile definire  $\mathfrak{so}(3) = \text{Antisym}_3(\mathbb{R}) = \{\hat{\omega} \in M_3(\mathbb{R}) \mid \omega \in \mathbb{R}^3\}$ . Per il Teorema 1.17 possiamo parametrizzare  $SO(3)$  utilizzando la funzione esponenziale suriettiva

$$\exp : \mathfrak{so}(3) \longrightarrow SO(3) \quad \hat{\omega} \mapsto e^{\hat{\omega}}.$$

Per rendere veramente utile la parametrizzazione scelta è necessario avere un modo facile per calcolare l'esponenziale di una matrice antisimmetrica, questo è fornito dalla *Formula di Rodriguez*.

**Teorema 1.19** (Formula di Rodriguez). *Sia  $\omega \in \mathbb{R}^3$  allora*

$$e^{\hat{\omega}} = I + \frac{\hat{\omega}}{\|\omega\|} \sin(\|\omega\|) + \frac{\hat{\omega}^2}{\|\omega\|^2} (1 - \cos(\|\omega\|)). \quad (1.7)$$

*Dimostrazione.* Siano  $u = \frac{\omega}{\|\omega\|}$  e  $t = \|\omega\|$ , è facile verificare le seguenti uguaglianze per le potenze di  $\hat{u}$

$$\hat{u}^2 = uu^\top - I \quad \hat{u}^3 = -\hat{u}. \quad (1.8)$$

Dunque la serie esponenziale (1.1) diventa

$$\begin{aligned} e^{\hat{\omega}} = e^{\hat{u}t} &= I + \left( t - \frac{t^3}{3!} + \frac{t^5}{5!} - \dots \right) \hat{u} + \left( \frac{t^2}{2!} - \frac{t^4}{4!} + \frac{t^6}{6!} - \dots \right) \hat{u}^2 \\ &= I + \sin(t)\hat{u} + (1 - \cos(t))\hat{u}^2. \end{aligned}$$

□

La formula precedente rende possibile un'analisi dettagliata di  $SO(3)$  e permette di approfondire le proprietà del legame con  $\mathfrak{so}(3)$ .

**Proposizione 1.20.** *Sia  $\omega \in \mathbb{R}^3$  allora  $R = e^{\widehat{\omega}}$  è la rotazione di asse  $\omega$  e angolo  $\|\omega\|$  radianti.*

*Dimostrazione.* Poiché l'asse di una rotazione è l'autospazio di dimensione 1 relativo all'autovalore 1 della matrice associata si procede verificando che  $(e^{\widehat{\omega}} - I)\omega = 0$ . Sfruttando la formula di Rodriguez, le equazioni (1.6) e (1.8) si ha

$$\begin{aligned} (e^{\widehat{\omega}} - I)\omega &= \sin(\|\omega\|) \frac{\widehat{\omega}}{\|\omega\|} \omega + (1 - \cos(\|\omega\|)) \frac{\widehat{\omega}^2}{\|\omega\|^2} \omega \\ &= (1 - \cos(\|\omega\|)) \left( \frac{\omega}{\|\omega\|} \frac{\omega^\top}{\|\omega\|} - I \right) \omega \\ &= (1 - \cos(\|\omega\|)) \left( \omega \frac{\omega^\top \omega}{\|\omega\|^2} - \omega \right) = (1 - \cos(\|\omega\|))(\omega - \omega) = 0. \end{aligned}$$

Sia ora  $\theta$  l'angolo di rotazione di  $R$  allora si ha  $\text{tr } R = 1 + 2 \cos \theta$ : infatti la traccia di una matrice è invariante per similitudine e ogni rotazione è simile a una matrice della forma

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix} \in SO(3).$$

Si verifica che

$$\begin{aligned} \text{tr } e^{\widehat{\omega}} &= \text{tr } I + \frac{\sin(\|\omega\|)}{\|\omega\|} \text{tr } \widehat{\omega} + \frac{1 - \cos(\|\omega\|)}{\|\omega\|^2} \text{tr } \widehat{\omega}^2 \\ &= 3 + \frac{1 - \cos(\|\omega\|)}{\|\omega\|^2} (-2 \|\omega\|^2) = 3 - 2(1 - \cos(\|\omega\|)) = 1 + 2 \cos(\|\omega\|) \end{aligned}$$

quindi  $|\theta| = \|\omega\|$ . □

Da questo si deduce che la funzione esponenziale  $\exp : \mathfrak{so}(3) \rightarrow SO(3)$  non è iniettiva in quanto se  $\|\omega\| = 2k\pi$  con  $k \in \mathbb{Z}$  allora  $e^{\widehat{\omega}} = I$ .

**Proposizione 1.21.** *Sia  $\omega \in \mathbb{R}^3$  allora  $e^{\widehat{\omega}}$  e  $\widehat{\omega}$  commutano, cioè*

$$e^{\widehat{\omega}} \widehat{\omega} = \widehat{\omega} e^{\widehat{\omega}}.$$

*Dimostrazione.* Poiché per la formula di Rodriguez  $e^{\widehat{\omega}}$  è un polinomio di  $\widehat{\omega}$ , le due matrici commutano. □

## 1.5 Rappresentazione del moto di corpo rigido

La teoria sviluppata precedentemente e le sue applicazioni sono necessarie per parametrizzare al meglio la parte rotazionale del moto di un corpo rigido, ora si procederà a collegare questa parte con quella traslazionale.

Si consideri un corpo  $\overline{C}$  e un suo punto  $P$ , sia inoltre  $W$  un sistema di riferimento fisso e sia  $C$  il sistema di riferimento solidale a  $\overline{C}$ . Si indica con  $X_w$  il vettore che descrive la posizione di  $P$  rispetto a  $W$  e analogamente con  $X_c$  il vettore che denota la posizione di  $P$  relativa a  $C$ .  $X_w$  si può vedere come somma della traslazione  $T_{wc}$  dell'origine di  $C$  rispetto a  $W$  e del vettore  $X_c$  espresso relativamente a un sistema di riferimento  $W'$  con gli assi paralleli a quelli di  $W$  e con origine l'origine di  $C$ . Detta  $R_{wc} \in SO(3)$  la rotazione che porta  $C$  in  $W'$  si ha

$$X_w = R_{wc}X_c + T_{wc}. \quad (1.9)$$

Bisogna notare che  $g_{wc} = (R_{wc}, T_{wc})$  non solo descrive la configurazione del corpo rigido  $\overline{C}$  ma rappresenta anche la trasformazione di coordinate tra due sistemi di riferimento. Proprio questa caratteristica è fondamentale per poter parlare dello stesso punto visto da due telecamere e quindi da due sistemi di riferimento diversi.

L'insieme delle coppie  $(R, T)$  che descrivono il moto di corpo rigido si chiama spazio delle *trasformazioni Euclidee speciali*

$$SE(3) = \{(R, T) \mid R \in SO(3), T \in \mathbb{R}^3\}.$$

$SE(3)$  può essere rappresentato come un gruppo matriciale  $SE(3) \subseteq GL_4(\mathbb{R})$  sfruttando le *coordinate omogenee*. È possibile infatti considerare  $\mathbb{R}^3$  come un sottoinsieme di  $\mathbb{R}^4$  mediante l'immersione

$$i : \mathbb{R}^3 \longrightarrow \mathbb{R}^4 \quad X \mapsto \begin{bmatrix} X \\ 1 \end{bmatrix} = \overline{X}.$$

Sfruttando queste nuove coordinate è possibile scrivere la relazione (1.9) in forma matriciale

$$\overline{X}_w = \begin{bmatrix} X_w \\ 1 \end{bmatrix} = \begin{bmatrix} R_{wc} & T_{wc} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ 1 \end{bmatrix} = \overline{g}_{wc} \overline{X}_c. \quad (1.10)$$

Dunque è possibile ridefinire

$$SE(3) = \left\{ \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \mid R \in SO(3), T \in \mathbb{R}^3 \right\} \subset GL_4(\mathbb{R}).$$

Inoltre è facile vedere che è un sottogruppo chiuso di  $GL_4(\mathbb{R})$ , cioè un gruppo matriciale.

## Capitolo 2

# Modello di telecamera

Dopo la trattazione matematica che permette di descrivere la posizione tra due diverse telecamere e il cambio di sistema di riferimento associato, questo capitolo sposta l'attenzione sul modello di telecamera che verrà considerato per semplificare il complesso sistema della visione ad una situazione matematicamente trattabile. Il modello di telecamera trattato è detto *ideal pinhole camera*. La lente di questo tipo di telecamera è approssimata ad un punto, il centro ottico  $o$ . Dunque i punti che contribuiscono alla luminosità di un punto dell'immagine sono disposti esclusivamente su una linea passante per il centro ottico. Il problema della ricostruzione 3D è molto difficile perché un'immagine memorizza il dato visivo in un modo molto vicino alla percezione umana del mondo e di conseguenza molto lontano dall'interpretazione di questo dato. Solo isolando caratteristiche di più alto livello è possibile effettuare un'analisi più profonda sull'immagine. Per iniziare si procede quindi con la definizione matematica dell'immagine e le sue caratteristiche, poi nel corso del capitolo si mostreranno quali possano essere le caratteristiche di alto livello utili alla ricostruzione 3D.

### 2.1 Formazione dell'immagine

La mente umana è abituata a pensare ad un'immagine come ad un accostamento di macchie di colori delimitate da contorni e modificate al variare delle condizioni di luminosità. Il dato visivo inoltre appare agli occhi umani come qualcosa di profondamente continuo, addirittura la mente, sulla base di questa ipotesi di continuità, riesce a percepire interamente un oggetto anche vedendone solo una parte. Per una telecamera invece un'immagine (in scala di grigi) è una matrice di numeri interi, quindi discreti. Il sensore della telecamera è una griglia di microsensori, detti *pixel*, che raccoglie la radiazione luminosa proveniente da una determinata area dell'ambiente circostante durante tutto il tempo di esposizione. I valori della matrice rappresentano la quantità di luce incidente nel pixel e si ottengono discretizzando integrale dell'energia luminosa rispetto al tempo

e alla dimensione del pixel. Nonostante questo, è comodo definire l'immagine come un oggetto su uno spazio continuo.

**Definizione 2.1.** Sia  $\Omega \subset \mathbb{R}^2$  compatto; si definisce *immagine* una funzione  $I : \Omega \rightarrow \mathbb{R}^+$  che associa ad ogni punto di  $(x, y) \in \Omega$  l'intensità luminosa incidente nel punto.  $I(x, y)$  è detta *luminosità del punto*.

$I$  dipende da molteplici fattori, non controllabili a priori, come le forme della scena, le proprietà di riflettanza dei materiali e la distribuzione delle fonti di luce. Il caso più semplice da trattare si ha quando le superfici riprese non cambiano luminosità al variare della posizione dell'osservatore, un materiale che gode di questa proprietà si dice *Lambertiano*.

Prima di proseguire, è preferibile introdurre qualche terminologia specifica relativa a una telecamera. Come si vede nella Figura 2.1, si dice *piano dell'immagine* il piano su cui si forma l'immagine, l'*asse ottico* è la retta ortogonale a questo passante per il centro ottico. Il punto di intersezione tra il piano dell'immagine e l'asse ottico si chiama *punto principale*. Dalle leggi ottiche ricavabili per le lenti sottili, è possibile ottenere la descrizione matematica di una *ideal pinhole camera*, modello ottenibile quando il diametro della lente tende a 0.

Sia  $P$  un punto nello spazio di coordinate  $\mathbf{X} = [X, Y, Z]^T$  rispetto a un sistema di riferimento di origine il centro ottico  $o$  e con l'asse  $z$  coincidente con l'asse ottico, allora è facile stabilire le seguenti relazioni tra  $P$  e il corrispondente punto dell'immagine  $\mathbf{x}$

$$x = -f \frac{X}{Z}, \quad y = -f \frac{Y}{Z} \quad (2.1)$$

dove  $f$  è la lunghezza focale della lente. La corrispondenza risultante detta *proiezione prospettica* è la funzione  $\pi : \mathbb{R}^3 \rightarrow \mathbb{R}^2 \quad \mathbf{X} \mapsto \mathbf{x}$ . Come anticipato per questo modello, si può dedurre dall'equazione (2.1) che ogni punto sulla retta passante per  $P$  e  $o$  è proiettato sulle stesse coordinate  $\mathbf{x} = [x, y]^T$ .

Si noti inoltre che l'immagine risultante usando l'equazione (2.1) risulta capovolta a causa dei segni negativi, quindi si adotterà la seguente proiezione

$$x = f \frac{X}{Z}, \quad y = f \frac{Y}{Z} \quad (2.2)$$

che corrisponde al posizionare il piano dell'immagine in  $z = f$  invece che  $z = -f$ . È più comodo esprimere il modello descritto in coordinate omogenee, come sono state espresse anche le trasformazioni di corpo rigido in  $SE(3)$ . Utilizzando le coordinate omogenee, l'equazione (2.2) si può scrivere come

$$Z \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (2.3)$$

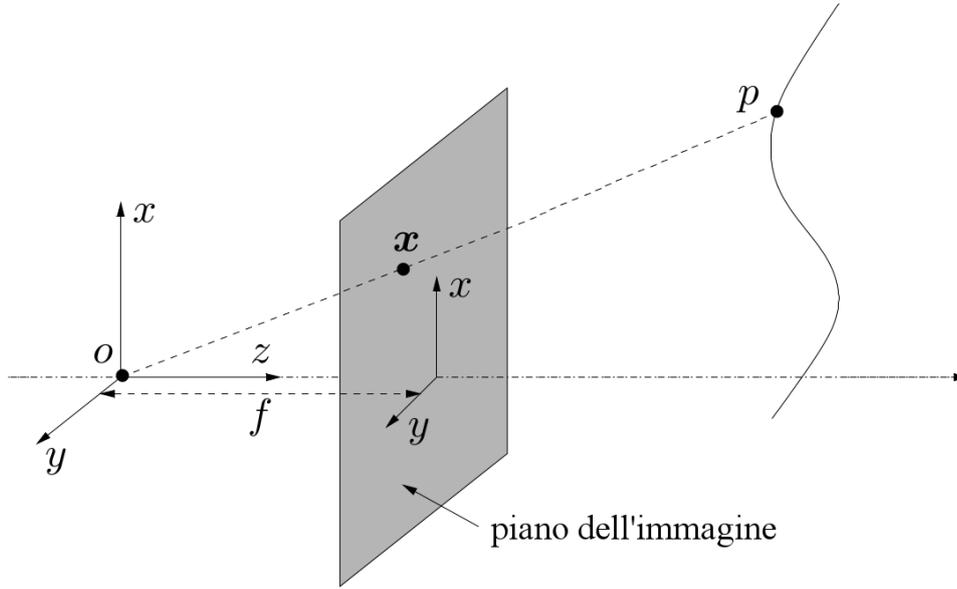


Figura 2.1: Modello di *ideal pinhole camera*

Si noti che è possibile la seguente fattorizzazione

$$\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Si definiscono quindi le seguenti matrici

$$K_f = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \in M_3(\mathbb{R}) \quad \Pi_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \in M_{3,4}(\mathbb{R}). \quad (2.4)$$

La matrice  $\Pi_0$  è detta *matrice della proiezione canonica*. Detta  $\lambda \in \mathbb{R}^+$  la profondità ignota di  $P$ , il modello di camera ideale si può sintetizzare attraverso l'equazione

$$\lambda \mathbf{x} = K_f \Pi_0 \mathbf{X} \quad (2.5)$$

## 2.2 Parametri intrinseci e estrinseci

Il modello finora descritto si applica solo a particolari scelte del sistema di riferimento interno alla telecamera e del sistema di riferimento fisso esterno alla telecamera, quindi si procederà ora a estenderlo al caso generale.

Fino ad ora si è visto solo il caso in cui l'origine del sistema di riferimento dell'immagine coincide con il punto principale, ma solitamente in un'immagine digitale i pixel

sono indicati rispetto a un sistema di riferimento di origine l'angolo in alto a sinistra dell'inquadratura. Inoltre la posizione del punto nell'immagine è espressa rispetto alla posizione del pixel, quindi in coordinate intere. Per convertirla in coordinate reali è necessario conoscere anche le dimensioni dei pixel. Siano  $\mathbf{x}' = [x', y']^\top$  le coordinate reali dell'immagine si ha

$$x' = s_x x + o_x \quad y' = s_y y + o_y$$

dove  $s_x, s_y$  sono le dimensioni dei pixel e  $(o_x, o_y)$  sono le coordinate del punto principale in pixel. Utilizzando le coordinate omogenee si ha

$$\mathbf{x}' = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \quad (2.6)$$

Nel caso generale, in cui i pixel non siano rettangolari, è presente anche il parametro  $s_\theta$  proporzionale a  $\cot(\theta)$  dove  $\theta$  è l'angolo tra gli assi dell'immagine. Di conseguenza la matrice di trasformazione assume la forma

$$K_s = \begin{bmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \in M_3(\mathbb{R}) \quad (2.7)$$

Combinando le equazioni (2.5) e (2.7) si ottiene

$$\lambda \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.8)$$

È possibile definire quindi la *matrice di calibrazione* o *matrice dei parametri intrinseci*

$$K = K_s K_f = \begin{bmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} f s_x & f s_\theta & o_x \\ 0 & f s_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2.9)$$

Spesso, quando ad esempio non si ha accesso fisico alla telecamera,  $K$  è ignota e quindi bisogna approssimare i parametri intrinseci con appositi algoritmi a partire dalle immagini a disposizione.

Inoltre se  $\mathbf{X}_0 = [X_0, Y_0, Z_0]$  sono le coordinate di  $P$  rispetto al sistema di riferimento fisso, dall'equazione (1.10) si sa che le coordinate  $\mathbf{X} = [X, Y, Z]$  rispetto al sistema di riferimento solidale alla telecamera sono legate da una trasformazione di corpo rigido a queste:

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix}. \quad (2.10)$$

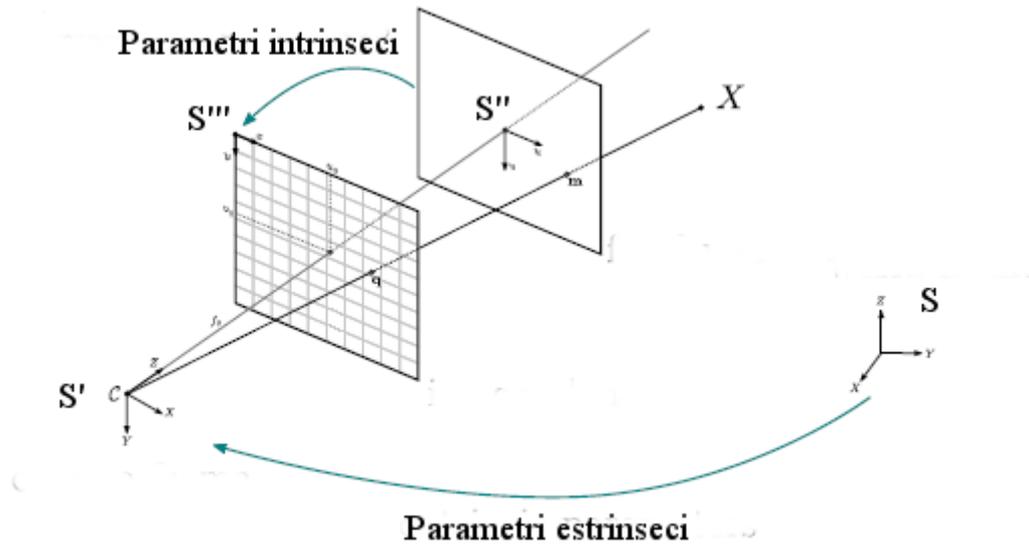


Figura 2.2: Parametri intrinseci ed estrinseci:  $S$  è il sistema di riferimento fisso, i parametri estrinseci codificano il cambiamento di coordinate al sistema di riferimento solidale alla telecamera  $S'$ .  $K_f$  proietta  $S'$  in  $S''$  e infine  $K_s$  porta  $S''$  in  $S'''$ , sistema di riferimento dell'immagine.

$(R, T)$  sono detti *parametri estrinseci*.

Sostituendo l'equazione (2.10) nell'equazione (2.8) si ottiene l'equazione finale che descrive il modello di telecamera scelto

$$\lambda \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} \quad (2.11)$$

o in forma compatta

$$\lambda \mathbf{x}' = K \Pi_0 \mathbf{g} \mathbf{X}_0 = [KR, KT] \mathbf{X}_0 = \Pi \mathbf{X}_0 \quad (2.12)$$

dove  $\Pi$  è detta *matrice di proiezione*.

## 2.3 Rappresentazione di punti e rette

Per poter esprimere matematicamente e in modo comodo le relazioni geometriche tra i punti di due diverse immagini è utile introdurre una opportuna rappresentazione per le rette in un'immagine.

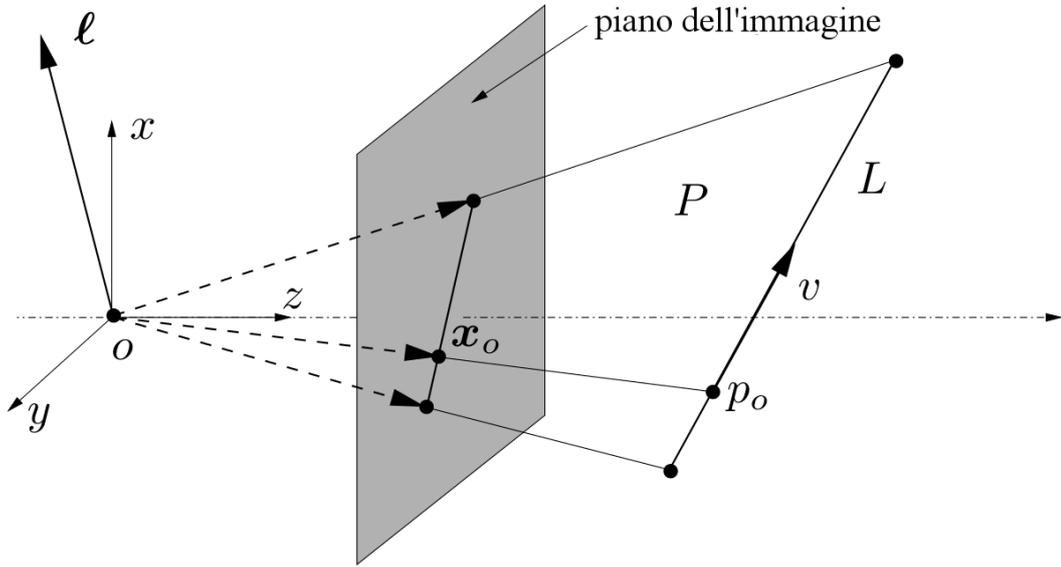


Figura 2.3: Preimmagini e coimmagini:  $x_o$  è un vettore della preimmagine di  $p_o$ ;  $P$  è la preimmagine della retta  $L$  e  $\ell$  è la sua coimmagine.

**Definizione 2.2.** Si dice *preimmagine* di un punto o di una retta nel piano dell'immagine l'insieme dei punti dello spazio appartenenti al più piccolo sottospazio vettoriale contenente il punto o la retta dati.

**Definizione 2.3.** Si dice *coimmagine* di un punto o di una retta nel piano dell'immagine il complemento ortogonale della preimmagine del punto o della retta dati.

La preimmagine di un punto  $P$  è la retta passante per  $P$  e per il centro ottico  $o$ , dunque è comodo identificare un punto con il vettore di direzione della sua preimmagine. Invece la preimmagine di una retta  $L$  è il piano passante per  $L$  e per  $p$ . Poiché la preimmagine di  $L$  è un sottospazio di  $\mathbb{R}^3$  di dimensione 2 la sua coimmagine è il sottospazio vettoriale generato dal vettore ortogonale a questo, e si indica  $\ell = [a, b, c]^\top$ . Una buona rappresentazione delle rette identifica ogni retta alla sua coimmagine. Sia quindi  $\mathbf{x}$  un punto  $P \in L$ , identificato con la sua preimmagine, allora

$$\ell^\top \mathbf{x} = 0.$$

Inoltre le colonne di  $\widehat{\mathbf{x}}$  sono un sistema di generatori della coimmagine di  $\mathbf{x}$  e analogamente le colonne di  $\widehat{\ell}$  sono un sistema di generatori della preimmagine di  $\ell$ , da cui le relazioni tra preimmagine e coimmagine

$$\widehat{\mathbf{x}}\mathbf{x} = 0 \quad \widehat{\ell}\ell = 0.$$

Questo tipo di rappresentazione di punti e rette come vettori in  $\mathbb{R}^3$  risulta molto comoda e si hanno i seguenti risultati.

Poiché i vettori scelti sono dei semplici rappresentanti dei sottospazi da essi generati si ha che è utile considerare la relazione di equivalenza

$$x \sim y \iff \exists \lambda \in \mathbb{R} \text{ tale che } y = \lambda x. \quad (2.13)$$

È chiaro che  $\mathbf{x}$  e  $\mathbf{x}'$  rappresentano lo stesso punto se e solo se  $\mathbf{x} \sim \mathbf{x}'$  e analogamente  $\ell$  e  $\ell'$  sono associati alla stessa retta se e solo se  $\ell \sim \ell'$ .

**Proposizione 2.4.** *L'intersezione di due rette  $\ell, \ell'$  è il punto  $\mathbf{x} = \ell \times \ell'$ .*

*Dimostrazione.* Per le proprietà del prodotto vettoriale si ha  $\ell^\top (\ell \times \ell') = \ell'^\top (\ell \times \ell') = 0$ , dunque  $\ell \times \ell'$  appartiene a entrambe le rette quindi alla loro intersezione.  $\square$

Con un ragionamento analogo si dimostra la seguente formula.

**Proposizione 2.5.** *La retta passante per due punti  $\mathbf{x}, \mathbf{x}'$  è  $\ell = \mathbf{x} \times \mathbf{x}'$ .*

## 2.4 Punti corrispondenti

Per poter operare la ricostruzione 3D è necessario mettere in relazione le due immagini a disposizione. Questo non è un compito facile, ma sono stati sviluppati efficaci algoritmi capaci di individuare coppie di punti, uno per ogni immagine, associati allo stesso punto nello spazio. La prima assunzione che deve essere fatta riguarda i materiali della scena ripresa, è indispensabile assumere che questi siano tutti Lambertiani. Infatti solo nel caso di materiali Lambertiani è ragionevole cercare coppie di punti corrispondenti solo tra i punti con luminosità molto simile. In tutti gli altri casi la diversa angolazione della ripresa modifica il valore della luminosità tra le due immagini anche per lo stesso punto nello spazio. Sotto queste ipotesi si potrebbero dunque cercare  $\mathbf{x}_1, \mathbf{x}_2$  immagini dello stesso punto  $P$ , tali che  $I_1(\mathbf{x}_1) = I_2(\mathbf{x}_2)$ . Questa condizione è però troppo stringente in quanto la misura dell'intensità luminosa, come ogni misurazione, è intrinsecamente soggetta a un errore statistico; è dunque preferibile considerare il vettore dei valori della luminosità in tutto un intorno del punto dell'immagine

$$l(\mathbf{x}) = \{I(\tilde{\mathbf{x}}) \mid \tilde{\mathbf{x}} \in W(\mathbf{x})\}$$

dove  $W(\mathbf{x})$  è un intorno di  $\mathbf{x}$ . Se la distanza tra due vettori  $l(\mathbf{x}_1), l(\mathbf{x}_2)$  è sufficientemente piccola, è possibile siano punti corrispondenti delle immagini.

Un problema molto difficile da risolvere è quello dell'occlusione: infatti esistono punti che non sono visibili da ogni angolazione perché coperti da altri oggetti nella scena. In questo caso è impossibile trovare la corrispondenza.

Un ulteriore problema sorge in caso di regioni dal colore uniforme: i pixel di queste regioni non sono univocamente associabili perché non sono sufficientemente peculiari. Questo problema viene però risolto cercando i *punti caratteristici* di ciascuna immagine e cercando le coppie tra questi. I punti caratteristici sono caratterizzati da un alto valore della derivata direzionale dell'intensità luminosa lungo due vettori linearmente indipendenti, sono quindi spesso angoli delle figure. Sono stati sviluppati molti algoritmi per l'individuazione dei punti caratteristici, il più diffuso è l'algoritmo di Harris, per una descrizione dettagliata del quale si rimanda a [5] (Sezione 5.3.1 pag. 149). I punti caratteristici individuati in entrambe le immagini sono poi accoppiati confrontando i valori della luminosità in un loro intorno, ottenendo coppie di punti corrispondenti.

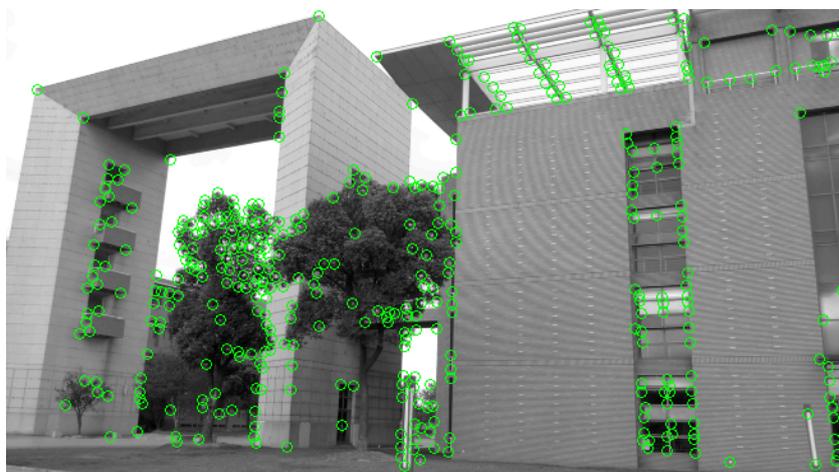


Figura 2.4: Immagine con evidenziati i punti caratteristici individuati dall'algoritmo di Harris

Quindi è necessario assumere che siano verificate alcune ipotesi ideali perché sia possibile estrarre i vincoli geometrici sulla scena dati da una coppia di immagini. Innanzitutto si esclude che ci siano oclusioni nella scena. Inoltre l'immagine non deve essere principalmente composta da campiture di colore uniforme o da schemi ripetuti. In caso contrario non sarebbe facile associare correttamente i punti corrispondenti e i possibili errori comprometterebbero irrimediabilmente la fedeltà della ricostruzione. Infine si assume che l'errore di misurazione sulle coordinate dei punti sia nullo. Quest'ultima assunzione, del tutto irrealistica, viene introdotta per semplificare la trattazione ma non è necessaria, poiché esistono algoritmi derivati dalla stessa teoria che verrà esposta e allo stesso tempo efficaci anche considerando gli errori sulle misure.

Nel seguito di questa tesi si presupporrà di avere a disposizione un insieme di coppie di punti corrispondenti con errore nullo sulle coordinate dei punti. A partire da tali corrispondenze si vedrà come possa essere possibile ricavare la posizione reciproca delle due telecamere che hanno scattato le immagini.

# Capitolo 3

## Geometria epipolare

Questo capitolo è il nucleo del lavoro di tesi e illustra i vincoli geometrici che si possono ricavare su una scena a partire da due immagini nel caso di una telecamera calibrata. Per operare una ricostruzione 3D è sufficiente ricavare la trasformazione di  $SE(3)$  che porta una telecamera nella posizione dell'altra perchè con il metodo di triangolazione è immediato ricavare la posizione nello spazio di qualsiasi punto una volta che ciò sia noto. Dunque, le formule che si ricaveranno in questo capitolo riguardano  $R$  e  $T$ , parte rotazionale e parte traslazionale della trasformazione di corpo rigido. Si noti che le metodologie descritte non sono direttamente implementabili su calcolatore perché le ipotesi di precisione assoluta sulle corrispondenze dei punti di due immagini diverse non sono realistiche. Per elaborare algoritmi efficaci nella pratica è necessario tenere conto degli errori di misurazione nella posizione di un punto ed elaborare metodi che minimizzino l'errore dal valore reale, questo va oltre gli scopi di questa tesi, quindi non si approfondirà l'argomento.

### 3.1 Matrice essenziale

La matrice essenziale è lo strumento fondamentale capace di incorporare tutti i vincoli geometrici dati dalla ripresa di due immagini. Si procede ora a ottenerne la forma e a studiarne la struttura.

Come detto nel Capitolo 2, ad ogni telecamera è possibile associare un sistema di riferimento con origine  $o$  nel centro ottico e con l'asse  $z$  coincidente con l'asse ottico e la relazione di ognuno di questi riferimenti con il sistema di riferimento fisso è data da una trasformazione di corpo rigido. Senza ledere la generalità, è possibile assumere che il sistema di riferimento fisso coincida con quello associato alla prima telecamera; il sistema di riferimento associato alla seconda sarà legato a questo dalla trasformazione  $g = (R, T) \in SE(3)$ . Sia  $P$  un punto nello spazio di coordinate  $X_1, X_2 \in \mathbb{R}^3$  rispetto alle

due telecamere allora si ha

$$X_2 = RX_1 + T. \quad (3.1)$$

Siano ora  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^3$  i punti associati a  $P$  nelle due immagini, espressi in coordinate omogenee. Poiché  $X_i = \lambda_i \mathbf{x}_i$   $i = 1, 2$ , sostituendo nell'equazione (3.1) si ottiene

$$\lambda_2 \mathbf{x}_2 = R\lambda_1 \mathbf{x}_1 + T.$$

Per eliminare i parametri moltiplicativi  $\lambda_i$ , si moltiplica a sinistra per  $\widehat{T}$ . Ricordando l'equazione (1.6) si ottiene

$$\lambda_2 \widehat{T} \mathbf{x}_2 = \widehat{T} R \lambda_1 \mathbf{x}_1.$$

Per definizione  $\widehat{T} \mathbf{x}_2 = T \times \mathbf{x}_2$  è ortogonale a  $\mathbf{x}_2$  quindi  $\langle \mathbf{x}_2, \widehat{T} \mathbf{x}_2 \rangle = \mathbf{x}_2^\top \widehat{T} \mathbf{x}_2 = 0$  dunque si ottiene

$$\mathbf{x}_2^\top \widehat{T} R \lambda_1 \mathbf{x}_1 = 0.$$

Dato che  $\lambda_1 > 0$ , si è dimostrato il seguente teorema.

**Teorema 3.1.** *Siano  $\mathbf{x}_1, \mathbf{x}_2$  immagini dello stesso punto  $P$  secondo due telecamere con posizione relativa  $(R, T) \in SE(3)$  allora  $\mathbf{x}_1, \mathbf{x}_2$  soddisfano l'equazione*

$$\mathbf{x}_2^\top \widehat{T} R \mathbf{x}_1 = 0. \quad (3.2)$$

**Definizione 3.2.** La matrice  $E = \widehat{T} R \in M_3(\mathbb{R})$  è detta *matrice essenziale*; questa appartiene allo *spazio essenziale*

$$\mathcal{E} = \{\widehat{T} R \mid R \in SO(3), T \in \mathbb{R}^3\} \subset M_3(\mathbb{R}).$$

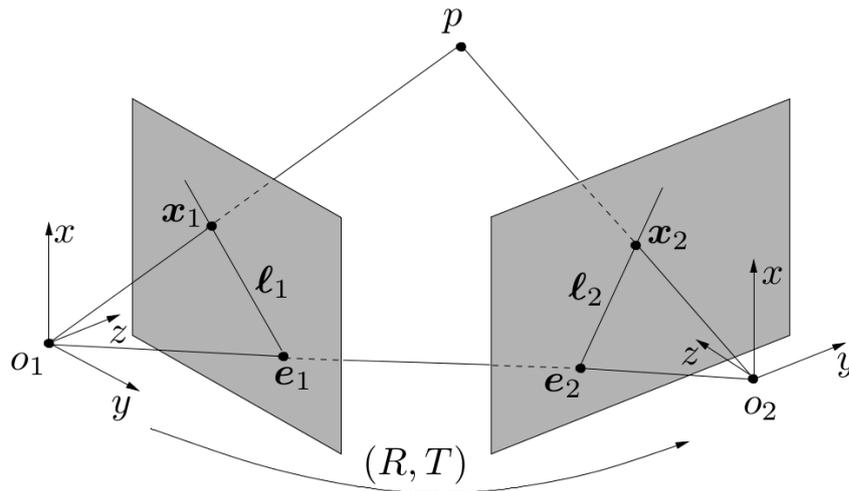


Figura 3.1: Proiezione del punto  $p$  sui piani dell'immagine delle due telecamere

L'equazione (3.2) può essere ricavata da considerazioni esclusivamente geometriche: infatti, come si vede nella Figura 3.1, il punto  $P$  e i centri ottici  $o_1$  e  $o_2$  devono formare un triangolo quindi i vettori che li collegano giacciono sullo stesso piano, ne consegue che il loro prodotto misto è nullo. Poichè rispetto alla seconda telecamera si ha

$$\overrightarrow{o_1P} = RX_1 = R\lambda_1\mathbf{x}_1 \quad \overrightarrow{o_2P} = X_2 = \lambda_2\mathbf{x}_2 \quad \overrightarrow{o_2o_1} = T$$

la condizione sul prodotto misto di questi vettori

$$\overrightarrow{o_2P} \cdot (\overrightarrow{o_2o_1} \times \overrightarrow{o_1P}) = 0$$

dà proprio l'equazione (3.2).

Si possono individuare notevoli entità geometriche associate a questa costruzione, se ne danno ora le definizioni e se ne illustrano le principali proprietà.

**Definizione 3.3.** Il piano passante per  $o_1, o_2, P$  si dice *piano epipolare* di  $P$ . La retta del piano epipolare passante per i due centri ottici è detta *linea base*. Inoltre le proiezioni  $e_1, e_2$  dei centri ottici  $o_1$  e  $o_2$  sul piano dell'immagine proprio dell'altra telecamera sono detti *epipoli*, mentre l'intersezione tra il piano epipolare e i due piani dell'immagine individua le rette  $\ell_1, \ell_2$  dette *rette epipolari* di  $P$ .

Gli epipoli e le rette epipolari risultano molto importanti anche perchè sono direttamente calcolabili a partire dalla matrice essenziale associata.

**Proposizione 3.4.** Siano  $e_1, e_2 \in \mathbb{R}^3$  i due epipoli allora

$$e_1 \sim R^\top T \quad e_2 \sim T$$

dove  $\sim$  è la relazione di equivalenza definita dalla condizione (2.13).

*Dimostrazione.* Per definizione  $e_2 \sim o_1$  e per l'equazione (3.1) che definisce il cambio di coordinate si ha che  $o_1$  ha coordinate  $T$  rispetto al sistema di riferimento solidale alla seconda telecamera. Analogamente  $e_1 \sim o_2$  e si verifica facilmente che  $o_2$  ha coordinate  $-R^\top T$  rispetto al sistema di riferimento solidale alla prima telecamera.  $\square$

**Proposizione 3.5.** Siano  $E = \widehat{T}R$  matrice essenziale e  $e_1, e_2 \in \mathbb{R}^3$  i due epipoli allora

$$Ee_1 = 0 \quad e_2^\top E = 0.$$

*Dimostrazione.* Per la Proposizione 3.4 e ricordando l'equazione (1.6)

$$Ee_1 = \mu_1 \widehat{T} R R^\top T = \mu_1 \widehat{T} T = 0 \quad e_2^\top E = \mu_2 T^\top \widehat{T} R = \mu_2 (R^\top \widehat{T}^\top T)^\top = -\mu_2 (R^\top \widehat{T} T)^\top = 0.$$

$\square$

**Proposizione 3.6.** Siano  $E = \widehat{T}R$  matrice essenziale e  $\ell_1, \ell_2 \in \mathbb{R}^3$  le linee epipolari associate ai punti  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^3$  allora

$$\ell_1 \sim E^\top \mathbf{x}_2 \quad \ell_2 \sim E \mathbf{x}_1.$$

*Dimostrazione.* Dalla Proposizione 2.5, ricordando che  $Rx \times Ry = R(x \times y)$  dove  $R \in SO(3)$  si ha che

$$\begin{aligned} \ell_1 &= e_1 \times \mathbf{x}_1 \sim R^\top T \times \mathbf{x}_1 = R^\top T \times R^\top R \mathbf{x}_1 = R^\top (T \times R \mathbf{x}_1) = \\ &= R^\top \widehat{T} R \mathbf{x}_1 \sim R^\top \widehat{T} (\lambda_2 \mathbf{x}_2 - T) = -\lambda_2 R^\top \widehat{T}^\top \mathbf{x}_2 \sim (\widehat{T}R)^\top \mathbf{x}_2 = E^\top \mathbf{x}_2. \end{aligned}$$

Analogamente

$$\ell_2 = e_2 \times \mathbf{x}_2 \sim T \times \mathbf{x}_2 = \widehat{T} \mathbf{x}_1 \sim \widehat{T} (R \lambda_1 \mathbf{x}_1 + T) \sim \widehat{T} R \mathbf{x}_1 = E \mathbf{x}_1.$$

□

## 3.2 Lo spazio essenziale

Si è visto come la matrice essenziale racchiuda in sé informazioni sulla posizione relativa di due telecamere. L'interesse applicativo nelle matrici essenziali deriva dal fatto che dalle corrispondenze di un sufficiente numero di punti è possibile calcolare la matrice essenziale. Dunque lo studio della particolare forma di queste matrici, cioè lo studio dello spazio essenziale, può permettere di ricavare la rotazione  $R$  e la traslazione  $T$  associate, dati fondamentali per effettuare la ricostruzione 3D.

**Proposizione 3.7.** Sia  $E = \widehat{T}R$  matrice essenziale, allora  $\|E\| = \|T\|$ .

*Dimostrazione.*  $\|E\| = \sup_{\|x\|=1} \|\widehat{T}Rx\| = \sup_{\|y\|=1} \|\widehat{T}y\| = \sup_{\|y\|=1} \|T \times y\| = \|T\|$ . □

È necessario introdurre un lemma dalla facile verifica, che verrà utilizzato per caratterizzare le matrici essenziali.

**Lemma 3.8.** Sia  $T \in \mathbb{R}^3$  e una matrice  $K \in M_3(\mathbb{R})$ , se  $\det(K) = 1$  e  $T' = KT$ , allora  $\widehat{T} = K^\top \widehat{T}' K$ .

Si ricordi che per ogni matrice  $A$  di dimensione  $m \times n$  esiste unica una particolare fattorizzazione detta *decomposizione ai valori singolari* o *SVD*, è cioè sempre possibile esprimere  $A$  con una scrittura del tipo

$$A = U \Sigma V^\top$$

dove  $U \in O(m)$ ,  $V \in O(n)$  e  $\Sigma$  è una matrice diagonale rettangolare di dimensioni  $m \times n$  a elementi non negativi. Il seguente teorema caratterizza completamente le matrici essenziali in base alla loro decomposizione ai valori singolari.

**Teorema 3.9.** Una matrice non nulla  $E \in M_3(\mathbb{R})$  è una matrice essenziale se e solo se  $E$  ha una decomposizione ai valori singolari  $E = U\Sigma V^\top$  tale che  $\Sigma = \text{diag}(\sigma, \sigma, 0)$  per un certo  $\sigma \in \mathbb{R}^+$  e  $U, V \in SO(3)$ .

*Dimostrazione.* Si suppone  $E = \widehat{T}R$  con  $R \in SO(3)$  e  $T \in \mathbb{R}^3$ . Per ogni  $T$  esiste  $R_0 \in SO(3)$  tale che  $R_0T = (0, 0, \|T\|)^\top$ . Poiché  $\det(R_0) = 1$ , per il Lemma 3.8  $\widehat{T} = R_0^\top \widehat{R_0T} R_0$ . Allora

$$EE^\top = \widehat{T}R R^\top \widehat{T}^\top = \widehat{T}\widehat{T}^\top = R_0^\top \widehat{R_0T} \widehat{R_0T}^\top R_0.$$

Si ottiene con un semplice calcolo che

$$\widehat{R_0T} \widehat{R_0T}^\top = \begin{bmatrix} 0 & -\|T\| & 0 \\ \|T\| & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -\|T\| & 0 \\ \|T\| & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \|T\|^2 & 0 & 0 \\ 0 & \|T\|^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

e poiché  $EE^\top = U\Sigma V^\top V\Sigma^\top U^\top = U\Sigma^2 U^\top$ , i valori singolari di  $E$  sono  $(\|T\|, \|T\|, 0)$ . Rimane da dimostrare che  $U, V \in SO(3)$ , cioè che  $\det U = \det V = 1$ . Sia  $R_Z$  la rotazione di angolo  $\frac{\pi}{2}$  di asse l'asse  $Z$  in senso orario per cui vale

$$R_Z^\top \widehat{R_0T} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -\|T\| & 0 \\ \|T\| & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \|T\| & 0 & 0 \\ 0 & \|T\| & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

allora  $\widehat{R_0T} = R_Z R_Z^\top \widehat{R_0T} = R_Z \text{diag}(\|T\|, \|T\|, 0)$ .

Concludendo  $E = \widehat{T}R = R_0^\top \widehat{R_0T} R_0 R = R_0^\top R_Z \text{diag}(\|T\|, \|T\|, 0) R_0 R$ , quindi si ha

$$U = R_0^\top R_Z \in SO(3) \quad \Sigma = \text{diag}(\|T\|, \|T\|, 0) \quad V^\top = R_0 R \in SO(3).$$

Si dimostra ora l'implicazione inversa. Sapendo che la decomposizione ai valori singolari di  $E$  è  $E = U\Sigma V^\top$  con  $U, V \in SO(3)$  e  $\Sigma = \text{diag}(\sigma, \sigma, 0)$  è possibile definire

$$\widehat{T} = UR_Z \Sigma U^\top = -(UR_Z \Sigma U^\top)^\top \in \mathfrak{so}(3) \quad R = UR_Z^\top V^\top \in SO(3)$$

per cui vale  $\widehat{T}R = UR_Z \Sigma U^\top UR_Z^\top V^\top = U\Sigma V^\top = E$ , quindi  $E$  è matrice essenziale.  $\square$

Il calcolo della matrice essenziale a partire da  $R \in SO(3)$  e  $T \in \mathbb{R}^3$  consiste in una semplice moltiplicazione. Durante la dimostrazione del teorema di caratterizzazione si è trovata una soluzione al problema inverso, cioè il calcolo di  $R$  e  $T$  conoscendo  $E$ . Poiché dalla conoscenza di un sufficiente numero di corrispondenze di punti in due immagini è possibile calcolare  $E$ , ma per un'adeguata ricostruzione 3D è necessario risalire ai valori di  $R$  e  $T$ , si analizza ora il problema inverso nel dettaglio per conoscerne il numero di soluzioni e costruirle. Per fare ciò si dimostra il seguente lemma.

**Lemma 3.10.** *Siano  $T \in \mathbb{R}^3 \setminus \{0\}$ ,  $R \in SO(3)$  tali che  $\widehat{T}R \in \mathfrak{so}(3)$  allora*

$$R = I \text{ oppure } R = e^{\widehat{u}\pi} \text{ dove } u = \frac{T}{\|T\|}.$$

Inoltre  $\widehat{T}e^{\widehat{u}\pi} = -\widehat{T}$ .

*Dimostrazione.* Poiché  $\widehat{T}R \in \mathfrak{so}(3)$ ,  $(\widehat{T}R)^\top = -\widehat{T}R$  da cui  $R\widehat{T}R = \widehat{T}R$ . Per il Teorema 1.17 esistono  $\omega \in \mathbb{R}^3$   $\|\omega\| = 1$  e  $\theta \in \mathbb{R}$  tali che  $R = e^{\omega\theta}$ . Se  $\theta = 0$  allora  $R = I$ . Nel caso  $\theta \neq 0$  si ha

$$e^{\omega\theta}\widehat{T}e^{\omega\theta} = \widehat{T}. \quad (3.3)$$

Moltiplicando a sinistra per  $\omega$  si ottiene  $e^{\omega\theta}\widehat{T}e^{\omega\theta}\omega = \widehat{T}\omega$  e ricordando che  $e^{\omega\theta}\omega = \omega$  risulta

$$e^{\omega\theta}\widehat{T}\omega = \widehat{T}\omega.$$

$\widehat{T}\omega$  deve essere nullo perché altrimenti sarebbe autovettore relativo all'autovalore 1. Ciò è impossibile perché l'autospazio relativo all'autovalore 1 di  $e^{\omega\theta}$  ha dimensione 1 e contiene  $\omega$ , ma  $\widehat{T}\omega$  è ortogonale a  $\omega$  e in particolare linearmente indipendente. Dunque  $\omega = \frac{T}{\|T\|}$  oppure  $\omega = -\frac{T}{\|T\|}$ . Per la Proposizione 1.21  $e^{\widehat{\omega}\theta}\widehat{T} = \widehat{T}e^{\widehat{\omega}\theta}$ , dunque sostituendo in (3.3) si ottiene

$$e^{2\widehat{\omega}}\widehat{T} = \widehat{T}. \quad (3.4)$$

Per la Formula di Rodriguez (1.7)  $e^{2\widehat{\omega}} = I + \widehat{\omega} \sin(2\theta) + \widehat{\omega}^2 \cos(2\theta)$ , quindi da (3.4) risulta

$$\widehat{\omega} \sin(2\theta)\widehat{T} + \widehat{\omega}^2 \cos(2\theta)\widehat{T} = 0 \implies \widehat{\omega}^2 \|T\| \sin(2\theta) + \widehat{\omega}^3 \|T\| \cos(2\theta) = 0.$$

Poiché  $\widehat{\omega}^2$  ha sempre un elemento non nullo sulla diagonale, mentre  $\widehat{\omega}^3 = -\widehat{\omega}$  li ha tutti nulli, le due matrici sono linearmente indipendenti quindi  $\sin(2\theta) = 1 - \cos(2\theta) = 0$ . Da cui  $\theta = 2k\pi$  o  $\theta = 2k\pi + \pi$  e dunque  $R = I$  o  $R = e^{\widehat{\omega}\pi}$ . Sfruttando ancora la Formula di Rodriguez si ottiene  $e^{\widehat{\omega}\pi} = I + 2\widehat{\omega}^2$  e osservando che se  $T = k\omega$  allora  $\widehat{T} = k\widehat{\omega}$  si ha

$$e^{\widehat{\omega}\pi}\widehat{T} = \widehat{T} + 2\widehat{\omega}^2\widehat{T} = \widehat{T} + 2k\widehat{\omega}^3 = \widehat{T} - 2k\widehat{\omega} = -\widehat{T}.$$

□

**Teorema 3.11.** *Sia  $E \in \mathcal{E}$  una matrice essenziale non nulla allora esistono esattamente due configurazioni  $(R, T)$  con  $R \in SO(3)$  e  $T \in \mathbb{R}^3$  tali che  $E = \widehat{T}R$ .*

*Dimostrazione.* Siano  $(R_1, T_1), (R_2, T_2) \in SE(3)$  soluzioni dell'equazione  $E = \widehat{T}R$ . Si ha  $\widehat{T}_1R_1 = \widehat{T}_2R_2$  da cui

$$\widehat{T}_1 = \widehat{T}_2R_2R_1^\top \in \mathfrak{so}(3).$$

Poiché  $\widehat{T}_2 \in \mathfrak{so}(3)$  e  $R_2 R_1^\top \in SO(3)$ , sono verificate tutte le ipotesi del Lemma 3.10 quindi si hanno solo due possibili casi

$$(R_2, T_2) = (R_1, T_1) \text{ oppure } (R_2, T_2) = (e^{\widehat{u_1 \pi}} R_1, -T_1) \text{ con } u_1 = \frac{T_1}{\|T_1\|}.$$

□

Tra le due soluzioni trovate non vi può essere una vera ambiguità in quanto solo una di queste è compatibile con valori di profondità positivi, cioè con oggetti posti davanti alla telecamera e non dietro.

L'interesse per la matrice essenziale deriva dal fatto che questa è direttamente calcolabile a partire dalle corrispondenze di punti dati. Infatti, come si è visto, ogni corrispondenza di punti si traduce in un'equazione del tipo

$$\mathbf{x}_2^\top E \mathbf{x}_1 = 0 \tag{3.5}$$

che è lineare rispetto agli elementi di  $E$ . Dunque sono sufficienti otto corrispondenze di punti per determinare i nove elementi di  $E$  a meno di una costante moltiplicativa. Per risolvere l'ambiguità data dalla costante moltiplicativa si assume  $\|T\| = \|E\| = 1$  ottenendo così la *matrice essenziale normalizzata*. Anche normalizzando rimane l'ambiguità data dal segno di  $E$ : infatti se  $E$  è soluzione del sistema, anche  $-E$  lo è. Ogni matrice essenziale normalizzata ( $E$  o  $-E$ ) permette di ottenere due possibili coppie di valori  $(R, T)$ , dunque una volta calcolata  $E$  si possono ricavare quattro possibili soluzioni per la posizione relativa delle due telecamere. Di queste quattro però tre possono essere eliminate imponendo che i punti abbiano profondità positiva. È dunque possibile determinare univocamente la posizione relativa delle due telecamere. La conoscenza di  $(R, T)$  è sufficiente per ricavare la posizione nello spazio del punto ripreso dalle due immagini, mediante il metodo della triangolazione.

Nella discussione precedente si è implicitamente assunto che  $E$  sia una matrice non nulla, ipotesi necessaria per poter essere normalizzata. Poiché  $E \neq 0$  se e solo se  $T \neq 0$ , per applicare la teoria esposta è necessario i centri ottici delle due telecamere siano distinti. Questo permette di ricavare informazione sulla scena grazie alla conseguente parallasse.

### 3.3 Matrice fondamentale

Fino ad ora si è analizzato solamente il caso in cui si abbia a disposizione una telecamera calibrata, si è assunto cioè che la matrice di calibrazione  $K$  sia l'identità e dalle relazioni risultanti si è trovata la matrice essenziale capace di descrivere la configurazione geometrica degli oggetti ritratti. Tuttavia la matrice  $K$  è nota a priori solo in un ristretto numero di casi, in tutti gli altri deve essere anche questa stimata a partire dalle immagini

e si parla di telecamera non calibrata. In questo caso generale, per non appesantire le notazioni, si assumerà che le due immagini siano state riprese dalla stessa telecamera e che quindi  $K_1 = K_2 = K$ . Se  $K$  è incognita le corrispondenze di punti non permettono di stimare la matrice essenziale, ma è possibile ottenere la *matrice fondamentale*, che ricopre lo stesso ruolo nel caso non calibrato.

Indicando con  $\mathbf{x}'$  le coordinate non calibrate si ha  $\mathbf{x} = K^{-1}\mathbf{x}'$  e sostituendo questa relazione nel vincolo epipolare si ottiene

$$\mathbf{x}_2^\top \widehat{T} R \mathbf{x}_1 = 0 \iff \mathbf{x}_2'^\top (K^{-1})^\top \widehat{T} R K^{-1} \mathbf{x}_1' = 0. \quad (3.6)$$

Si definisce la *matrice fondamentale*  $F$  associata

$$F = (K^{-1})^\top \widehat{T} R K^{-1} \in M_3(\mathbb{R}). \quad (3.7)$$

Analogamente a come si è ricavata la matrice essenziale, cioè con operazione puramente algebriche, è possibile calcolare la matrice fondamentale e, posto  $T' = KT$ , si ottiene

$$F = \widehat{T}' R K^{-1}. \quad (3.8)$$

**Proposizione 3.12.**  *$F$  è ben definita a meno di costanti moltiplicative.*

*Dimostrazione.* Sia  $T' = KT$  si dimostra che  $(K^{-1})^\top \widehat{T}' R K^{-1} = \widehat{T}' R K^{-1}$  se  $\det K = 1$ , nel caso  $\det K \neq 1$  si ha comunque l'uguaglianza a meno di costanti moltiplicative. Posto  $\det K = 1$  si può applicare il Lemma 3.8 da cui

$$\widehat{T}' = \widehat{KT} = (K^{-1})^\top K^{-1} (\widehat{KT}) K^{-1} = (K^{-1})^\top \widehat{T} K^{-1}.$$

Pertanto  $F = (K^{-1})^\top \widehat{T}' R K^{-1} = (K^{-1})^\top \widehat{T} K^{-1} R K^{-1} = \widehat{T}' R K^{-1} = F$ .  $\square$

Tutte le formule dimostrate per il caso calibrato valgono sostituendo  $F$  ad  $E$  e  $\mathbf{x}'$  a  $\mathbf{x}$ . In particolare, come nel caso della matrice essenziale, conoscendo solo otto corrispondenze di punti è possibile stimare la matrice fondamentale a partire da due immagini. Però mentre nel caso calibrato la matrice essenziale è sufficiente per ricavare la posizione reciproca delle due telecamere, questo non può avvenire nel caso non calibrato dunque  $F$  non permette una ricostruzione 3D euclidea della scena. Questo si può dimostrare semplicemente confrontando il numero di gradi di libertà delle matrici in gioco.  $F$  ha otto gradi di libertà, poiché è definita da nove coefficienti determinati a meno di una costante moltiplicativa, ma dipende dalle matrici  $K$  (cinque gradi di libertà),  $R$  (tre gradi di libertà) e  $\widehat{T}$  (due gradi di libertà). Quindi è impossibile recuperare in modo univoco  $K$ ,  $R$  e  $T$ .

Per dare una prova più formale si ricava la struttura della matrice di proiezione nel caso non calibrato. Siano  $\mathbf{X}'$  le coordinate non calibrate, per cui vale quindi  $\mathbf{X}' = K\mathbf{X}$ , dall'equazione (2.12) si ottiene

$$\lambda \mathbf{x}' = [KR, KT] \mathbf{X}_0 = KR \mathbf{X}_0 + KT = KRK^{-1} \mathbf{X}'_0 + T' = [KRK^{-1}, T'] \mathbf{X}'_0 = \Pi \mathbf{X}'_0$$

dove  $T' = KT$  e  $\mathbf{X}'_0 = K\mathbf{X}_0$ . Nel caso calibrato è possibile estrarre da  $\Pi$  i valori  $(R, T)$ , questo non può avvenire nel caso non calibrato come si vede nella proposizione seguente.

**Proposizione 3.13.** *Sia  $\phi$  la funzione che associa a ogni matrice di proiezione la matrice fondamentale corrispondente*

$$\Pi = [K R K^{-1}, T'] \xrightarrow{\phi} F = \hat{T}' K R K^{-1}$$

allora  $\phi^{-1}(F)$  ha un numero non numerabile di elementi, quindi  $\phi$  non è iniettiva.

*Dimostrazione.* Siano  $t \in \mathbb{R}, v \in \mathbb{R}^3$  allora  $\phi([K R K^{-1} + T'v^\top, tT']) \sim \phi([K R K^{-1}, T'])$ .  $\square$

Nel caso calibrato non si verifica la stessa ambiguità perché la sottomatrice  $3 \times 3$  di  $\Pi$  deve essere una matrice di rotazione.

Si procede ora a scoprire come questa ambiguità influenzi il procedimento di ricostruzione 3D. A partire dall'equazione (2.12) si possono evidenziare tutte le possibili ambiguità

$$\lambda \mathbf{x}' = \Pi \mathbf{X} = K \Pi_0 g \mathbf{X} = \underbrace{K R_0^{-1} R_0 \Pi_0 H^{-1}}_{\tilde{\Pi}} \underbrace{H g g_w^{-1} g_w}_{\tilde{\mathbf{X}}} \mathbf{X} = \tilde{\Pi} \tilde{\mathbf{X}}$$

con  $R_0 \in SO(3)$ ,  $H \in GL_4(\mathbb{R})$  e  $g_w \in SE(3)$ .

In primo luogo si studia l'ambiguità introdotta da  $R_0$ . Dette  $\tilde{K} = K R_0^\top$ ,  $\tilde{R} = R_0 R$  e  $\tilde{g} = [R_0 R, R_0 T]$  si ha

$$\lambda \mathbf{x}' = K \Pi_0 g \mathbf{X} = K R_0^{-1} R_0 [R, T] \mathbf{X} = \tilde{K} [\tilde{R}, \tilde{T}] \mathbf{X} = \tilde{K} \Pi_0 \tilde{g} \mathbf{X}.$$

É quindi impossibile distinguere  $K$  da  $\tilde{K}$  a partire dalle misure di  $\mathbf{x}'$ , ma poiché  $K$  è una matrice triangolare superiore, utilizzando la fattorizzazione QR si può ottenere  $K$  conoscendo  $\tilde{K}$ . Infatti la scomposizione QR permette di calcolare  $Q \in SO(3)$  e  $R$  triangolare superiore univocamente determinate il cui prodotto sia una qualsiasi matrice assegnata. Anche l'ambiguità introdotta da  $g_w$  è rimovibile infatti poiché il sistema di riferimento fisso è arbitrario ogni trasformazione sulle coordinate  $\mathbf{X}$  misurate rispetto ad esso é irrilevante. Irrisolubile è invece l'ambiguità data dalla matrice  $H$ .

### 3.4 Ricostruzione proiettiva

Dato un insieme di corrispondenze di punti e una telecamera non calibrata, il meglio che si può ottenere è una ricostruzione proiettivamente equivalente alla scena reale. Come visto in precedenza, è possibile calcolare la matrice fondamentale  $F$  dalle coppie di punti corrispondenti  $\{(\mathbf{x}'_1, \mathbf{x}'_2)\}$ , si vede ora come ricavare la matrice di proiezione  $\Pi_2$  associata alla seconda telecamera. Questa è l'unica incognita, infatti l'altra matrice di

proiezione  $\Pi_1$  si può sempre porre nella forma  $[I, 0]$  con un'opportuna scelta del sistema di riferimento. Per quanto visto nella Proposizione 3.13, la scomposizione di una matrice fondamentale per ricavare la matrice di calibrazione e la posizione relativa delle due telecamere non è unica, ma dipende dalla particolare scelta dei parametri liberi  $t \in \mathbb{R}$  e  $v \in \mathbb{R}^3$ . Queste sono le uniche ambiguità possibili e ciò è garantito dal seguente teorema.

**Teorema 3.14.** *Due coppie di matrici di proiezione  $(\Pi_1, \Pi_2)$  e  $(\Pi_1, \tilde{\Pi}_2)$  sono associate alla stessa matrice fondamentale  $F$  se e solo se esistono  $t \in \mathbb{R}$ ,  $v \in \mathbb{R}^3$  e  $H \in GL_4(\mathbb{R})$  tali che*

$$\Pi_2 \sim \tilde{\Pi}_2 H \text{ con } H = \begin{bmatrix} I & 0 \\ v^\top & t \end{bmatrix}.$$

*Dimostrazione.* Si dimostra la prima implicazione. Si assume senza ledere la generalità  $\Pi_1 = [I, 0]$ . Siano  $A, B \in M_3(\mathbb{R})$  e  $a, b \in \mathbb{R}$  tali che  $\Pi_2 = [A, a]$  e  $\tilde{\Pi}_2 = [B, b]$ , allora per ipotesi  $\hat{a}A \sim \hat{b}B$ . Poiché  $a$  e  $b$  appartengono rispettivamente al nucleo sinistro di dimensione 1 della matrice a primo e a secondo membro, i due vettori sono linearmente dipendenti:  $a \sim b$ . Ne consegue  $A \sim B + bv^\top$  per un certo  $v \in \mathbb{R}^3$ . In forma matriciale questo si può scrivere come

$$[A, a] \sim [B, b] \begin{bmatrix} I & 0 \\ v^\top & t \end{bmatrix}$$

pertanto  $\Pi_2 \sim \tilde{\Pi}_2 H$  con  $H \in GL_4(\mathbb{R})$  della forma voluta. Con un facile calcolo si dimostra anche l'implicazione inversa.  $\square$

Dal Teorema 3.14 discende che, avendo a disposizione solo corrispondenze di punti provenienti da due telecamere non calibrate, non è possibile ottenere una ricostruzione 3D euclidea. Si può però individuare un modo canonico di scegliere la matrice di proiezione  $\Pi_2$  associata a una matrice fondamentale  $F$  data. Poiché se  $\|T'\| = 1$  vale l'identità  $\hat{T}'\hat{T}'^\top \hat{T}' = \hat{T}'$ , ponendo come  $T'$  l'unico vettore tale che  $(T')^\top F = 0$  e  $\|T'\| = 1$  si può scegliere la coppia di matrici di proiezione così fatta:

$$\Pi_1 = [I, 0] \quad \Pi_2 = [\hat{T}'^\top F, T']. \quad (3.9)$$

La *scomposizione canonica* della matrice fondamentale  $F$ , dipende solo da  $F$  in quanto  $T'$  è calcolato come vettore di norma unitaria appartenente al nucleo sinistro di  $F$ . La ricostruzione che si ottiene scegliendo le matrici di proiezione ottenute con la scomposizione canonica è proiettivamente equivalente alla scena reale.

# Bibliografia

- [1] A. Baker, *Matrix Groups - An Introduction To Lie Groups*, Springer, 2003
- [2] V. S. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations*, Springer, 1984
- [3] F. W. Warner, *Foundations of Differentiable Manifolds and Lie Groups*, Springer, 1983
- [4] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, Prentice Hall, 2007
- [5] D. A. Forsyth, J. Ponce, *Computer Vision: A Modern Approach*, Pearson, 2011
- [6] Y. Ma, S. Soatto, J. Kosecká, S. S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*, Springer, 2005
- [7] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2004



# Ringraziamenti

Questa tesi è l'atto conclusivo di un percorso di tre anni, un percorso che mi ha profondamente cambiato grazie a nuove persone che sono diventate importanti nella mia vita e a pilastri che sono rimasti tali per me. In primo luogo vorrei ringraziare la mia famiglia, che mi ha sostenuto in ogni situazione e mi ha spinto a sfruttare tutte le opportunità che mi sono state offerte. La più grande opportunità a cui mi riferisco è sicuramente quella offerta dal Collegio Superiore. In particolare vorrei ringraziare mio nonno perché sono convinto di aver superato la selezione per entrare al Collegio anche grazie a una risposta ispirata da un suo racconto di qualche giorno prima.

Parlando del Collegio Superiore non posso che dire che è stata l'esperienza più bella e formativa della mia vita. Ho conosciuto persone straordinarie condividendo momenti di gioia e studio. Tra queste voglio citare Matteo Monti che ha sempre stimolato e incuriosito la mia mente sottoponendole sempre problemi di non facile soluzione, ma soprattutto che ha sempre creduto in me e nelle mie potenzialità fin dal primo momento in cui ci siamo conosciuti. Credo che la cosa più importante che ho imparato da lui sia l'importanza delle persone sopra ogni cosa in questo mondo e ciò ha fortemente influenzato il mio modo di pormi nei confronti degli altri e del mondo stesso. Inoltre il Collegio ha reso possibile il più incredibile incontro della mia vita, quello con Silvia. Tuttavia credo che ognuno dei Collegiali mi abbia lasciato qualcosa e voglio citare tra questi Alberto, Alessandro, Angelo, Aurora, Caterina, Eleonora, Elvira, Francesco, Jacopo, Laura, Lisa, Lorena, Marco e Maria Elena. Da ognuno di questi ho imparato qualcosa sull'arte, sulla musica, sulla politica, sulle relazioni e su molto altro ancora.

Fuori dal Collegio, la matematica mi ha messo in contatto con i migliori compagni di studio con cui è stato possibile discutere di matematica con passione e continuità. In particolare voglio ringraziare Giovanni Ronchi che con la sua passione e la sua capacità di riflettere sulle piccole cose ha stimolato il mio interesse scientifico e filosofico in svariati momenti, mostrandomi che ovunque c'è meno banalità di quella che traspare.

Ma non voglio dimenticare che molte persone conosciute prima del percorso universitario sono costantemente rimaste al mio fianco anche durante questi ultimi tre anni. Il Team tutto (Cassa, Dens, Gios, Mats, Plaz e Toms) è stato il mio sostegno nei momenti difficili e ha organizzato le più spericolate avventure e serate per me. È per me una presenza certa che spero perduri a lungo nel tempo.