

ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

SCUOLA DI INGEGNERIA  
Corso di Laurea in Ingegneria Informatica

CALIBRAZIONE MEDIANTE  
CONICA ASSOLUTA

Relatore:

Prof. MASSIMO FERRI

Candidato:

DAVIDE PERINTI

III Sessione

Anno Accademico 2019-2020

# Indice

Introduzione	4
Capitolo 1 Calibration Camera	5
1.1 Modello della fotocamera Pinhole	5
1.1.1 Modello	5
1.1.2 Cenni storici	6
1.1.3 Descrizione	7
1.1.4 Assunti aggiuntivi	8
1.2 Come avviene la trasformazione di coordinate	8
1.3 Camera Matrix	10
1.3.1 Parametri estrinseci	10
1.3.2 Parametri intrinseci	11
1.4 Camera calibration	15
Capitolo 2 Strumenti matematici	17
2.1 Ampliamento proiettivo	17
2.2 Conica assoluta	20
2.3 Geometria epipolare e matrice fondamentale	24
2.3.1 Geometria epipolare	24
2.3.2 Matrice fondamentale	27
2.3.3 Matrice essenziale	30
2.4 Equazione di Kruppa	31
Capitolo 3 Metodi di risoluzione	35
3.1 Metodo delle 3 palle	35
3.2 Metodi self-calibration	40
3.2.1 Metodo senza equazioni di Kruppa	40
3.2.2 Metodo self-calibration con equazioni di Kruppa	43
Conclusione	44
Bibliografia	45
Sitografia	46



# Introduzione

In questa tesi voglio presentare il problema della camera calibration, che è una questione centrale quando si necessita di determinare la metrica di una scena attraverso delle fotografie. Questo problema consiste nel determinare alcuni parametri che riguardano proprietà interne alla fotocamera e alcuni parametri che riguardano la posa della fotocamera nello spazio; questi parametri permettono di determinare come un punto nella scena viene rappresentato nella fotografia, permettendo così attraverso la fotografia di inferire le caratteristiche della scena. Risolvere questo problema è un prerequisito per ogni applicazione dove si ha bisogno di collegare le immagini 2D con il mondo 3D.

Esistono svariati metodi per risolvere questo problema in letteratura, in questo documento ho deciso di analizzarne alcuni, fra quelli proposti, che utilizzano la conica assoluta; questa scelta è motivata principalmente dallo stupore nel vedere come un concetto così astratto come quello della conica assoluta, la quale “risiede” non solo “all’infinito” ma “all’infinito immaginario”, sia così legato al “concreto”, a cose che noi tutti possiamo facilmente vedere mentre l’unico modo per “vedere” la conica assoluta è attraverso l’equazione che rappresenta la sua immagine. Io trovo molto intrigante questo entanglement fra due “mondi” che se guardati con occhio non attento sembrano non avere a che fare l’uno con l’altro e invece si rivelano molto legati.

Il seguente documento è diviso in 3 parti: la prima in cui andrò a presentare il problema e come il problema viene formulato in matematica, nella seconda parte illustrerò le fondamentali matematiche necessarie per affrontare la spiegazione di alcuni metodi ripresi da articoli accademici che andò ad analizzare nella terza parte del documento.

# Capitolo 1 Calibration Camera

In questo capitolo andrò a presentare la taratura di una fotocamera o camera calibration. Inizio presentando la traduzione matematica utilizzata per analizzare questo problema, spiegando poi in cosa consiste il problema di taratura e concludo presentando alcune motivazioni che spingono allo studio di metodi per la risoluzione del suddetto.

## 1.1 Modello della fotocamera Pinhole

Procedo innanzitutto con la presentazione del modello di fotocamera Pinhole per affrontare l'analisi della taratura.

### 1.1.1 Modello

Il modello di fotocamera Pinhole (in italiano stenopeico), descrive la relazione matematica tra le coordinate di un punto nello spazio tridimensionale e la sua proiezione sul piano dell'immagine di una fotocamera ideale, dove l'apertura della fotocamera è descritta come un punto e non vengono utilizzati obiettivi per mettere a fuoco la luce.

Il modello non include, ad esempio, distorsioni geometriche o sfocature di oggetti sfocati causati da lenti e aperture di dimensioni finite. Inoltre, non tiene conto del fatto che la maggior parte delle fotocamere pratiche ha solo coordinate di immagine discrete. Ciò significa che il modello di fotocamera stenopeica può essere utilizzato solo come approssimazione del primo ordine della mappatura da una scena 3D a un'immagine 2D. La sua validità dipende dalla qualità della fotocamera e in generale, diminuisce dal centro dell'immagine ai bordi all'aumentare degli effetti di distorsione dell'obiettivo.

Alcuni degli effetti che il modello di camera stenopeica non tiene in considerazione possono essere compensati, ad esempio applicando opportune trasformazioni di coordinate sulle coordinate dell'immagine; altri effetti sono sufficientemente piccoli da essere trascurati se si utilizza una fotocamera di alta qualità. Ciò significa che il modello di fotocamera stenopeica spesso può essere utilizzato come una descrizione ragionevole

di come una fotocamera rappresenta una scena 3D, ad esempio nella visione artificiale e nella computer grafica.

## 1.1.2 Cenni storici

Il concetto di fotografia stenopeica deriva dalla camera oscura, una stanza completamente buia con un piccolo foro su un lato e un pittore all'interno, il quale traccia e copia l'immagine proiettata. La camera oscura era più una scoperta che un'invenzione. Le prime idee risalgono al V e IV secolo a.C., quando sia i cinesi sia Aristotele cominciarono a parlare di immagini invertite, dovute alla luce passante attraverso un foro. La prima camera oscura nacque nel IX secolo d.C., quando lo scienziato persiano Ibn al-Haytham pubblicò il libro "Deli Aspecti"(questa è la traduzione d'epoca), scrivendo di questa nuova invenzione. Lui, infatti, aveva pensato di creare un foro che portasse in una stanza buia e di piazzare delle candele al di fuori della stanza per creare un'immagine. Quest'invenzione divenne un ottimo strumento per esperimenti scientifici di altri scienziati nei secoli a seguire. Per esempio, Isaac Newton utilizzò la tecnica della camera oscura per l'esperimento della scomposizione dei colori della luce solare con il prisma e lo scienziato olandese Gemma Frisius usò la camera oscura per osservare le eclissi solari.

Il concetto fu poi ripreso da Leonardo da Vinci, il quale nei suoi scritti specificava la necessità di sigillare la stanza per evitare l'entrata di luce non voluta e il fatto che la camera oscura e l'occhio avessero molte somiglianze, in termini di funzionamento. Il concetto fu ulteriormente spiegato da altri scienziati nei secoli a seguire, come Giovanni Battista della Porta nel suo libro "Magia Naturalis" nel 1558. Le prime fotografie vere e proprie, catturate con la camera oscura risalgono agli inizi del XIX secolo, quando il francese Joseph Nicéphore Niépce utilizzò del materiale fotosensibile per la prima volta per catturare la luce e creare un'immagine. Tuttavia, la foto non era visibile perché doveva essere sviluppata, una tecnica che fu inventata nel 1839 da un astronomo. Infatti, il primo a creare la prima foto vera e propria fu lo scienziato Sir David Brewster, il quale diede il nome alla fotografia stenopeica nel suo libro, pubblicato nel 1850. Le sue fotografie sono esposte in alcuni musei di Londra. Mentre nel XX e XXI secolo, ci fu

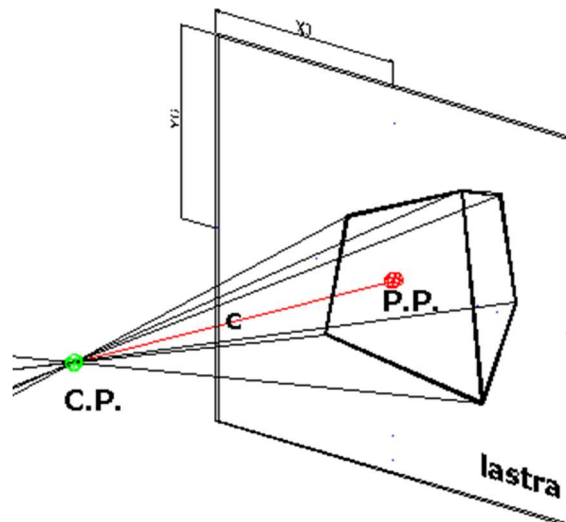
l'avvento della fotografia digitale, il concetto della fotografia stenopeica fu di grande aiuto nel sviluppare oggetti di tecnologia avanzata da parte delle grandi industrie, dove le lenti non erano necessarie.

### 1.1.3 Descrizione

Il contenuto del seguente paragrafo è tratto da [s1]

Il nome "pin-hole" (letteralmente buco di spillo) nasce proprio dal fatto che si suppone che la luce che impressiona la lastra o i sensori di una fotocamera digitale, attraverso un foro di piccole dimensioni, tanto piccole da potersi considerare un punto nell'accezione geometrica del termine.

Il punto che s'individua tracciando una retta ortogonale al piano della lastra o del sensore e passante per il centro di proiezione CP viene detto Punto Principale, ed indicato con PP. La posizione di tale punto rispetto ad un sistema di riferimento solidale con la lastra, sistema di riferimento normalmente con origine nel centro geometrico della stessa lastra, viene indicata con  $(X_c, Y_c)$ .



La distanza tra i punti PP e CP è la focale (o distanza principale) della camera e si indica generalmente con  $c$ .

L'assunto che tutti i raggi proiettivi passino per il punto CP è chiaramente solo un'approssimazione. Nella realtà tale luogo anziché essere un punto, è una chiazza di

grandezza non nulla. Questo produce una deformazione nell'immagine, che prende il nome di distorsione. Sperimentalmente si è visto che tale deformazione si compone di una componente tangenziale ed una radiale, di cui solo quest'ultima è rilevante. Esisterà allora un punto di simmetria radiale per tale funzione, che in generale non coincide con il CP e che prende il nome di Punto Principale di Simmetria, PPS.

### 1.1.4 Assunti aggiuntivi

Per concludere questa trattazione sul modello adottato voglio fare un elenco di assunzioni aggiuntive per poter procedere con la trattazione visto che considererò nella seguente parte di avere a che fare con fotocamere digitali:

- Le lenti sono perfette, quindi non presentano distorsioni della luce;
- Tutti i raggi luminosi sono considerati linee rette che passano dal foro stenopeico;
- Il punto focale e il punto principale giacciono nell'asse ottico;
- La distanza fra il piano immagine e il foro stenopeico è costante nel tempo.

## 1.2 Come avviene la trasformazione di coordinate

In questo sottocapitolo espongo come avviene la trasformazione di coordinate dal sistema di coordinate del mondo a quello della fotografia considerando una macchina fotografica digitale approssimata al modello pinhole. I contenuti sono tratti da [s2]

Inizio con scrivere la relazione che lega le coordinate pixel della fotografia con le coordinate di un punto nella scena:

$$x = PX$$

Dove  $x$  sono le coordinate nella fotografia 2-dimensionali  $P$  è detta camera matrix e  $X$  sono le coordinate nel mondo 3-dimensionali. Per i punti utilizzerò spesso le coordinate omogenee, (questa scelta sarà giustificata meglio in seguito nel capitolo 2 che indicherò con:



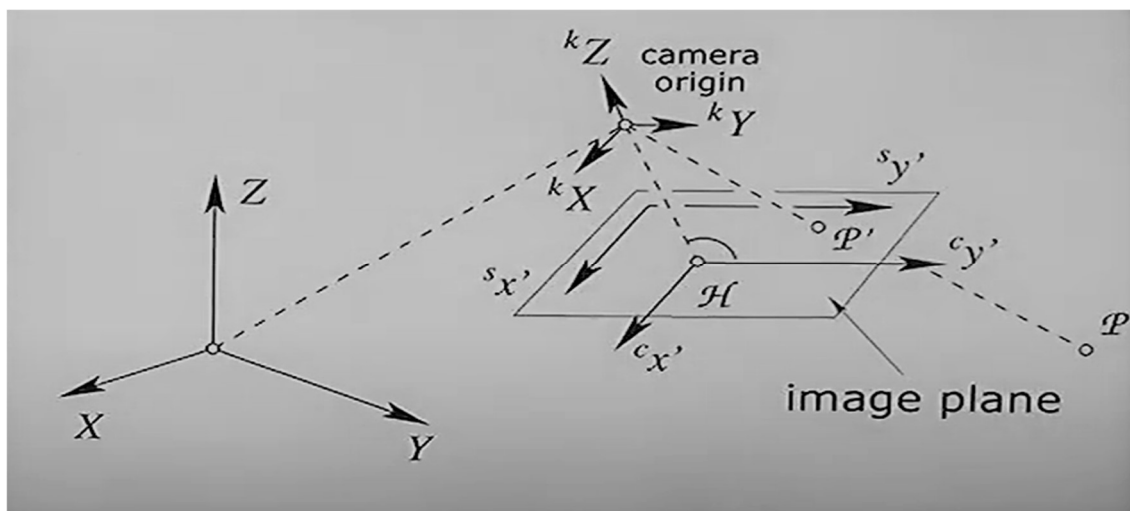
$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Il fatto che la coordinata aggiuntiva in entrambi li spazi sia uguale a 1 significa che consideriamo per adesso solo punti propri, dell'ampliamento proiettivo dei due spazi affini (vedi capitolo 3 per maggiori informazioni).

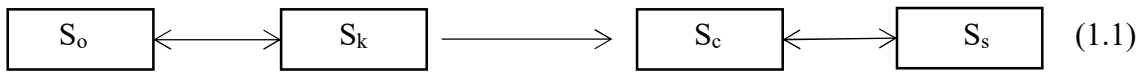
In questa descrizione sono presenti 4 sistemi di coordinate:

- Sistema mondo che indicherò con  $S_o$  : sistema 3-dimensionale, su cui mappare gli oggetti del mondo a cui voglio scattare la foto.
- Sistema della fotocamera che indicherò con  $S_k$ : sistema 3-dimensionale, su cui mappare gli oggetti del mondo a cui voglio scattare la foto e che ha come origine il foro stenopeico.
- Sistema dell'immagine che indicherò con  $S_c$ : sistema 2-dimensionale, per mappare le informazioni nel piano immagine.
- Sistema dei sensori che indicherò con  $S_s$ : sistema 2-dimensionale, il quale dipende da valori interni della fotocamera e serve per mappare le informazione del piano immagine alle coordinate dei pixel.

Di seguito riporto una schematizzazione di queste trasformazioni.



Adesso mostro un diagramma a blocchi che descrive tutti i passaggi della trasformazione di coordinate.



Le frecce bidirezionali sono presenti dove la trasformazione è invertibile, infatti la trasformazione fra  $S_k$  e  $S_c$  è un passaggio da 3 dimensioni a 2 dimensioni che non può essere invertito. Uno spazio 3-dimensionale contiene un numero di informazioni maggiori rispetto ad uno 2-dimensionale di conseguenza ci sono punti diversi nello spazio mondo che sono trasformati in uno stesso punto nel piano immagine.

## 1.3 Camera Matrix

Adesso vado ad analizzare nel dettaglio la matrice  $P$  introdotta nel precedente capitolo e per farlo ripercorro il diagramma a blocchi della trasformazione di coordinate precedentemente indicato. Il contenuto è tratto da [s2]

Nella seguente trattazione il pedice riferito ai punti sta ad indicare il sistema di riferimento in cui esprimo le sue coordinate.

### 1.3.1 Parametri estrinseci

Con parametri estrinseci si intende quei parametri che si riferiscono a proprietà esterne alla fotocamera.

La prima trasformazione dello schema (1.1) che analizzo è quella che sancisce il passaggio da un generico sistema di riferimento nello spazio al sistema di riferimento della fotocamera. Questa trasformazione dipende da due fattori: orientamento e posizione della fotocamera.

I parametri estrinseci quindi devono descrivere una trasformazione di traslazione e rotazione (cioè un movimento rigido) del sistema di riferimento  $S_o$  per ottenere il sistema di riferimento  $S_k$ ; da qui la necessità di 6 parametri estrinseci per descrivere questa trasformazione: 3 per lo spostamento nelle tre direzioni e 3 per la rotazione rispetto ai 3 assi coordinati.

Prendiamo un generico punto  $Q$  nello spazio espresso in coordinate rispetto al sistema  $S_o$ , per semplificare la trattazione adesso utilizzo coordinate euclidee:

$$Q_o = [Q_x, Q_y, Q_z]^T$$

Considero adesso  $O_o$  come l'origine del sistema di coordinate solidale con la fotocamera (il quale corrisponde con il foro stenopeico) che risulta nel sistema  $S_o$ :

$$O_o = [O_x, O_y, O_z]^T$$

Il vettore  $O_o$  corrisponde al vettore traslazione  $t$  che trasforma l'origine di  $S_o$  nell'origine di  $S_k$ . Per esprimere la rotazione degli assi è sufficiente una matrice ortogonale che indicheremo con la lettera  $R$ . Lo spazio delle matrici ortogonali  $3 \times 3$  è uno spazio di dimensione 3 perciò per definire una matrice ortogonale  $3 \times 3$  sono sufficienti 3 parametri.

È possibile ora esprimere il punto  $Q$  rispetto al sistema di coordinate  $S_k$  come:

$$Q_k = R(Q_o - O_o)$$

Vengono espresse ora le coordinate omogenee :

$$\begin{bmatrix} Q_k \\ 1 \end{bmatrix} = \begin{bmatrix} R & 0 \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} I_3 & -O_o \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} Q_o \\ 1 \end{bmatrix} = \begin{bmatrix} R & -RO_o \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} Q_o \\ 1 \end{bmatrix}$$

Dove  $0^T$  indica un vettore riga di 3 componenti tutte uguali a zero,  $0$  indica un vettore colonna di 3 componenti tutte uguali a zero e  $I_3$  rappresenta la matrice identità  $3 \times 3$ .

La matrice  $4 \times 4$  ottenuta risulta rappresentare la generica trasformazione di un punto in coordinate omogenee fra  $S_o$  ed  $S_k$ .

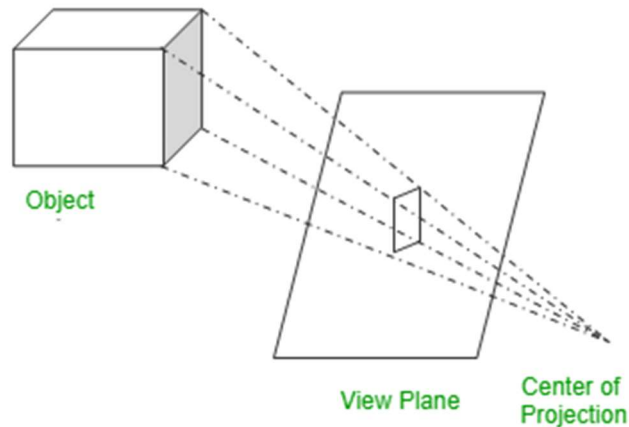
### 1.3.2 Parametri intrinseci

Con parametri intrinseci si intende quei parametri relativi a valori interni alla fotocamera.

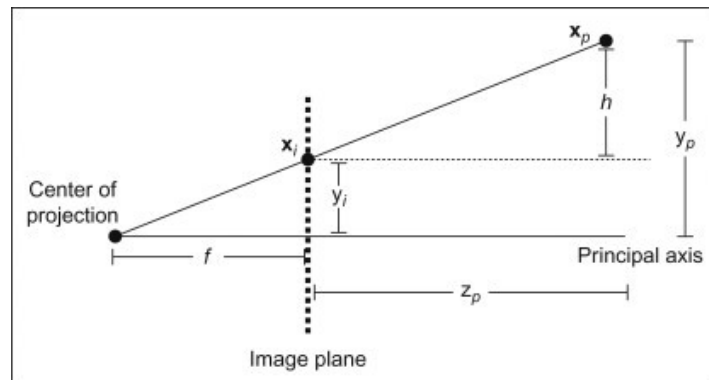
Adesso vado ad analizzare la trasformazione non invertibile fra  $S_k$  e  $S_c$ , essa consiste in una proiezione dallo spazio mondo al piano immagine. Considero questa proiezione ideale, infatti devono valere le assunzioni espresse nel paragrafo 1.1.4..

Chiamo  $f$  la lunghezza focale cioè la distanza fra il piano immagine e il foro stenopeico.

Vado a considerare per semplicità che il piano sia posto fra lo spazio e il foro stenopeico; nonostante questo sia fisicamente irrealizzabile è del tutto equivalente a considerare il foro in mezzo al piano immagine e allo spazio.



Semplificando il disegno soprastante ho:



Notando che ho due triangoli simili, quello piccolo a sinistra del piano immagine e quello che unisce il piccolo triangolo a sinistra del piano immagine con il trapezio rettangolo a destra del piano immagine ottengo:

$$\frac{h + y_i}{f + z_p} = \frac{y_i}{f} \quad \longrightarrow \quad y_i = f * \frac{y_p}{z_p + f}$$

Notando che  $y_p$ ,  $z_p+f$  sono le coordinate sugli assi  $y$  e  $z$  del punto  $x_p$  rispetto al sistema di riferimento  $S_k$  e per l'asse  $x$  vale una formula analoga a quella sopra riportata per l'asse  $y$ , posso affermare perciò che moltiplicando per la lunghezza focale  $f$  il rapporto fra una coordinata rispetto ad una direzione generica e la coordinata perpendicolare al

piano immagine ottengo la coordinata del punto  $x_p$  nel piano immagine rispetto a quella direzione. Ho trovato qual è l'informazione perduta: calcolando con la formula appena definita la coordinata nella direzione perpendicolare al piano immagine, ho che essa per tutti i punti vale  $f$ ; questo fatto non sorprende dato che il piano immagine rispetto al sistema  $S_k$  ha equazione:

(considero  $z$  l'asse perpendicolare al piano e per semplicità scrivo l'equazione in coordinate euclidee)

$$P_i: z=f$$

Per essere più precisi dobbiamo dire che l'informazione persa riguarda il fatto che nel piano immagine non possiamo distinguere fra due punti che si trovano entrambi sulla retta che congiunge l'origine di  $S_c$  (cioè il foro stenopeico) e uno dei due punti.

Adesso vado ad esprimere la trasformazione in questione tramite coordinate omogenee, tenendo in considerazione il risultato ottenuto.

$$\begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_k \\ Y_k \\ Z_k \\ T_k \end{bmatrix}$$

Andando a svolgere i calcoli:

$$\begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} = \begin{bmatrix} f \frac{X_k}{Z_k} \\ f \frac{Y_k}{Z_k} \\ \frac{Z_k}{Z_k} \end{bmatrix} \longrightarrow \begin{bmatrix} fX_k \\ fY_k \\ Z_k \end{bmatrix} = \begin{bmatrix} f \frac{X_k}{Z_k} \\ f \frac{Y_k}{Z_k} \\ 1 \end{bmatrix}$$

Questo corrisponde esattamente a quello che ho trovato prima; ciò evidenzia che gli unici punti che da  $S_k$  si trasformano in punti impropri di  $S_c$  sono quelli che hanno  $Z_k$  uguale a zero.

Adesso vado ad analizzare la trasformazione fra  $S_c$  ed  $S_s$ ; in questa trasformazione avvengono 3 passaggi:

1. Posizionare l'origine sul punto principale;
2. Effettuare variazioni di scala nelle direzioni dei due assi;
3. Effettuare una trasformazione causata dal fatto che l'angolo compreso fra i due assi nel piano del sistema  $S_s$ , può non essere di  $90^\circ$ .

Per quanto riguarda il punto 1. è sufficiente sommare la componente  $x_p$  e  $y_p$  alle coordinate dei punti; posso tradurre questa trasformazione con la seguente matrice:

$$\begin{bmatrix} 1 & 0 & x_p \\ 0 & 1 & y_p \\ 0 & 0 & 1 \end{bmatrix}$$

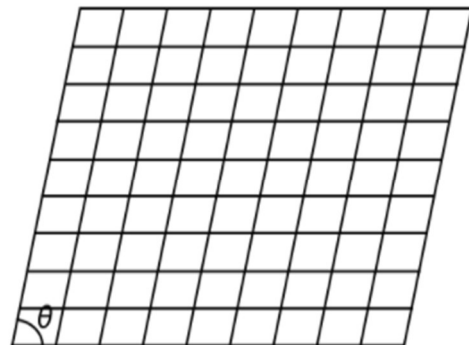
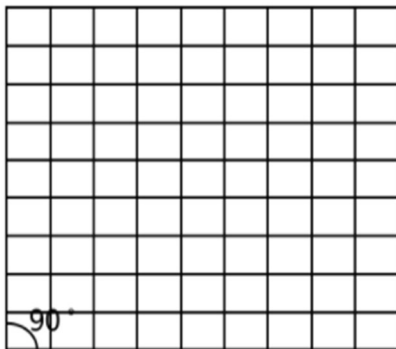
Questa trasformazione consiste semplicemente in una traslazione.

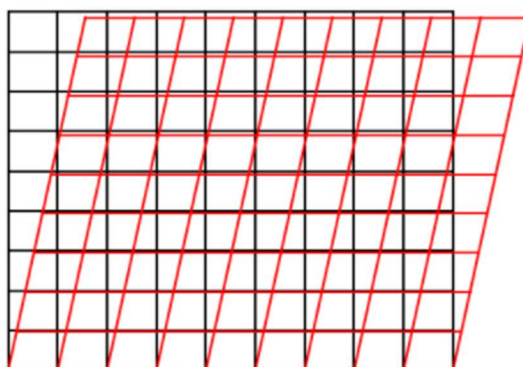
Per quanto riguarda la trasformazione al punto 2. essa consiste in una trasformazione in scala degli assi x e y dovuta alla lunghezza focale ottica del sensore nelle due direzioni.

Riassumo in una sola matrice queste due trasformazioni chiamando  $k_x$  e  $k_y$  i fattori di scala rispettivamente dell'asse x e y.

$$\begin{bmatrix} k_x & 0 & x_p \\ 0 & k_y & y_p \\ 0 & 0 & 1 \end{bmatrix}$$

Per il punto 3. è necessario aggiungere un quinto parametro, il quale può essere aggiunto in due forme: come un angolo che rappresenta proprio l'angolo di inclinazione fra i due assi oppure come un parametro chiamato skew, il quale è chiaramente proporzionale al precedente.





Mostrerò di seguito solo la rappresentazione con il parametro skew  $s$  ponendolo uguale a  $-fk_x \cot(\theta)$ . Per maggiori dettagli [1]

Unendo tutti i parametri intrinseci possiamo chiamare  $\mathbf{K}$  la matrice:

$$\begin{bmatrix} fk_x & s & x_p \\ 0 & fk_y & y_p \\ 0 & 0 & 1 \end{bmatrix}$$

Aggiungo a questo anche la matrice  $\mathbf{R}$  e il vettore traslazione  $\mathbf{t}$  e ottengo la camera matrix  $\mathbf{P}$ , che mi consente di trasformare punti dello spazio in punti nella foto.

$$\mathbf{P} = \mathbf{KR}[\mathbf{I}_3 | \mathbf{t}] \quad (2)$$

## 1.4 Camera calibration

La calibrazione della fotocamera è un task importante nella visione artificiale. L'obiettivo è di stabilire la proiezione dalle coordinate del mondo 3D alle coordinate dell'immagine 2D. Una volta che questa proiezione è nota, le informazioni 3D possono essere dedotte dal 2D e viceversa. Pertanto, la calibrazione della fotocamera è un prerequisito per qualsiasi applicazione dove è necessaria la relazione tra le immagini 2D e il mondo 3D.

Per raggiungere il suddetto scopo usando il modello precedentemente esposto di fotocamera è sufficiente determinare la camera matrix  $\mathbf{P}$ .

Esistono due categorie principali di metodi per effettuare questa operazione:

1. Metodi standard;
2. Metodi self-calibration.

La categoria 1. consiste in metodi che utilizzano un oggetto speciale il quale viene posto nel campo visivo della fotocamera. La forma 3D dell'oggetto di calibrazione è nota, in altre parole le coordinate di alcuni punti di riferimento su di esso sono note in un sistema di coordinate solidale coll'oggetto. Solitamente l'oggetto di calibrazione è una piastra piatta con un motivo regolare contrassegnato su di essa (comunemente viene utilizzata una scacchiera). Il modello è scelto in modo tale che le coordinate dell'immagine dei punti di riferimento proiettati possano essere misurate con grande precisione.

Questo metodo è ampiamente utilizzato. Fornisce una determinazione molto accurata dei parametri della fotocamera.

Lo svantaggio che presenta è che in molte applicazioni non può essere utilizzato dato che non sempre è possibile posizionare certi oggetti nel modo adeguato. Un altro svantaggio è che non è possibile calibrare in linea quando la fotocamera è già coinvolta in un'attività visiva.

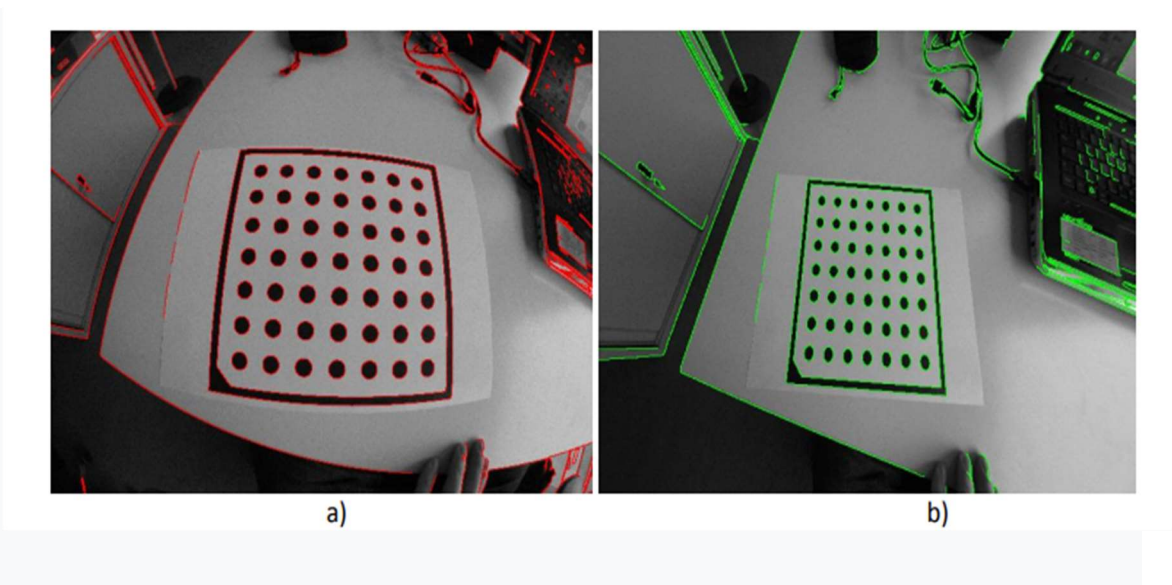
Mentre i metodi di categoria 2. rimediano a questi svantaggi nonostante forniscano una stima peggiore dei parametri di interesse; infatti questi metodi possono essere elaborati utilizzando le stesse immagini richieste per eseguire l'attività visiva e risultano molto utili quando la fotocamera subisce una serie di spostamenti in una scena rigida. L'unico requisito è che la visione artificiale sia in grado di stabilire corrispondenze tra i punti in immagini diverse; in altre parole può identificare coppie di punti, uno da ciascuna immagine, che sono proiezioni dello stesso punto della scena.

Sono descritti molti metodi per ottenere coppie di punti corrispondenti in due immagini nella letteratura. La corrispondenza viene quindi eseguita mediante tecniche di correlazione o un metodo di tracciamento come quello descritto in [2].

Questa seconda classe di metodi quindi non si basa su oggetti speciali posizionati nella scena ma sul confronto tra diverse foto di una stessa scena prese ad angolazioni diverse, utilizzando strumenti di stereovisione. [mettere articolo sulla stereovisione]

La camera calibration risulta indispensabile quando si vuole effettuare delle misure di una scena attraverso la sua fotografia. Di seguito mostro il confronto fra due foto prima e dopo aver realizzato la calibraizone sulla fotocamera.





## Capitolo 2 Strumenti matematici

In questo capitolo espongo gli strumenti matematici necessari per poter presentare 3 metodi per la camera calibration i quali utilizzano la conica assoluta.

### 2.1 Ampliamento proiettivo

Nella seguente trattazione utilizzerò il simbolo  $\sim$  per indicare la proporzionalità fra 2 elementi.

Definizione spazio proiettivo sul campo  $A$ :

Ogni terna  $(V,P,f)$ , dove  $V$  è uno spazio vettoriale su  $K$ ,  $P$  è un insieme e

$f: V - \{0_v\} \longrightarrow P$  è un'applicazione suriettiva tale che:

$$f(v) = f(w) \Leftrightarrow w \sim v$$

Gli elementi dell'insieme  $P$  si dicono punti dello spazio proiettivo.

Definizione proiettività:

Siano  $(V, P, f), (V', P', f')$  spazi proiettivi su  $A$ . Sia  $T : V \longrightarrow V'$  una trasformazione lineare iniettiva. Allora  $T$  induce un'applicazione  $T^* : V - \{0_V\} \longrightarrow V' - \{0_{V'}\}$  compatibile con la proporzionalità, cioè tale che

$$u \sim v \Leftrightarrow T^*(u) \sim T^*(v)$$

Dunque  $T$  induce anche un'applicazione  $q_T : P \longrightarrow P'$  che verrà associata a  $T$ .

Dunque chiamerò proiettività da  $P$  a  $P'$  ogni applicazione  $q : P \longrightarrow P'$  tale che esista una trasformazione lineare iniettiva  $T$  per cui sia  $q = q_T$ .

Definizione di riferimento proiettivo:

Chiamerò riferimento proiettivo di uno spazio proiettivo  $(V, P, f)$  di dimensione  $n$  ogni  $(n+2)$ -pla  $S = (H_0, \dots, H_n, U)$  di punti a  $n+1$  a  $n+1$  linearmente indipendenti.

Esiste un'applicazione biettiva  $g : \{\text{Basi di } V\} \longrightarrow \{\text{Riferimenti di } (V, P, f)\}$ , che associa ogni base un riferimento nel seguente modo:

$$[(v_0, \dots, v_n)] \rightarrow ([v_0], \dots, [v_n], [v_0 + \dots + v_n])$$

Dove le parentesi quadre indicano che possiamo prendere il suo contenuto moltiplicato per un qualsiasi coefficiente appartenente al campo dove è stato definito lo spazio proiettivo.

Definizione base normalizzata rispetto ad  $S$

Dato un riferimento  $S$ , una qualunque base  $B$  rappresentante di  $g^{-1}(S)$  si dice normalizzata rispetto ad  $S$ .

Definizione coordinate omogenee (o proiettive)

Dato un punto  $P$ , le sue coordinate omogenee rispetto ad un riferimento  $S$  sono le componenti  $(x_0, \dots, x_n)$  di un suo qualunque rappresentante rispetto ad una base normalizzata rispetto ad  $S$ .

Le coordinate omogenee di un punto sono non tutte nulle e determinate a meno di un fattore moltiplicativo.

Dati gli spazi proiettivi  $P, P'$  di dimensioni  $m, n$ , loro riferimenti  $S, S'$  rispettivamente e una proiettività  $q : P \rightarrow P'$  esiste una matrice  $M$  dimensione  $(n+1) \times (m+1)$  per cui, se  $Q$  è rappresentato in coordinate omogenee come  $(x)$  e  $g(Q)$  è  $(y)$  in coordinate omogenee allora:

$$a(y) = M(x)$$

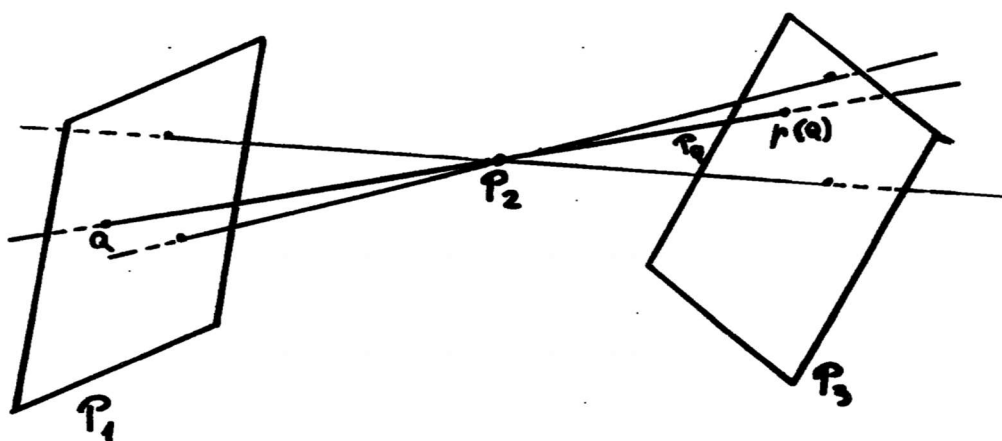
Per  $a \in A$  opportuno.

#### Definizione sottospazio congiungente

Dati sottospazi  $P', P''$  di uno spazio proiettivo  $P$ , chiamiamo sottospazio congiungente il sottospazio proiettivo associato al sottospazio vettoriale congiungente dei sottospazi vettoriali ad essi associati.

#### Definizione di proiezione

Siano  $P_1, P_2, P_3$  sottospazi di uno spazio proiettivo  $P$ . Per ogni punto  $Q \in P_1$  sia  $P_Q$  il sottospazio congiungente di  $Q$  e  $P_2$ . Se per ogni  $Q \in P_1$  vale che  $P_Q \cap P_3$  è costituito da un solo punto, l'applicazione  $p: P_1 \rightarrow P_3$  che a  $Q$  associa tale punto viene detta proiezione di  $P_1$  da  $P_2$  su  $P_3$ .



Nel caso della fotocamera nel modello pinhole precedentemente analizzato abbiamo che  $P_1$  è la scena 3-dimensionale,  $P_2$  è il foro stenopeico e  $P_3$  è il piano immagine.

#### Definizione ampliamento proiettivo:

Sia  $Z^n$  uno spazio affine  $n$ -dimensionale sul campo  $A$ ; sia inoltre  $Z^{*n}$  il suo spazio dei

vettori liberi. Sia inoltre  $P^{*n-1}$  lo spazio proiettivo canonicamente associato ad  $Z^{*n}$ , quindi abbiamo che i punti di  $P^{*n-1}$  sono le direzioni di  $Z^n$ . Fisso adesso un riferimento affine  $S = (M, B)$  su  $Z^n$ .

Definisco un'applicazione  $f : A^{n+1} - \{(0)\} \longrightarrow Z^n \cup P^{*n-1}$  così:

$$\begin{aligned} (X_0, \dots, X_{n-1}, X_n) &\rightarrow \left( \frac{X_0}{X_n}, \dots, \frac{X_{n-1}}{X_n} \right) \in Z^n && \text{con } X_n \neq 0 \\ (X_0, \dots, X_{n-1}, X_n) &\rightarrow (X_0, \dots, X_{n-1}) \in P^{*n-1} && \text{con } X_n = 0 \end{aligned}$$

Ho che  $(A^{n+1}, Z^n \cup P^{*n-1}, f)$  è uno spazio proiettivo, tale spazio è detto ampliamento proiettivo di  $Z^n$ .

**Definizione punti impropri**

Dato l'ampliamento proiettivo  $(A^{n+1}, Z^n \cup P^{*n-1}, f)$ , si dice improprio ogni punto che appartiene al sottoinsieme  $P^{*n-1}$ , che chiamo iperpiano improprio.

Di conseguenza abbiamo che tutti i punti propri sono tutti quelli che non sono impropri.

## 2.2 Conica assoluta

In questo paragrafo vado ad introdurre il concetto di conica assoluta, presentando le basi necessarie.

Per prima cosa fisso uno spazio proiettivo  $(V^{n+1}, P^n, f)$  sul campo  $A$  e un suo riferimento proiettivo  $S$ , ai quali nel seguito della trattazione mi riferirò.

**Definizione iperquadrica**

Un'iperquadrica di  $(V^{n+1}, P^n, f)$  è una classe di proporzionalità di forme quadratiche non nulle su  $V^{n+1}$ .

**Definizione discriminante di una iperquadrica**

Il discriminante di una iperquadrica  $[q]$  rispetto ad  $S$  è la matrice simmetrica associata a  $q$  o a qualunque altra rappresentazione rispetto alla base normalizzata.

Definizione immagine di iperquadrica

L'immagine di un'iperquadrica  $[q]$ , indicato come  $\text{Im}[q]$ , è l'insieme:

$$\text{Im}[q] = \{P=[u] \in P^n \mid q(u)=0\}$$

Quindi dato un discriminante  $D$  associato a  $[q]$  abbiamo che (chiamando  $(X)$  il vettore colonna con generiche coordinate omogenee e con  $(X)^T$  il suo trasposto):

$$\text{Im}[q]: (X)^T D (X) = 0$$

Le iperquadriche di uno spazio proiettivo 2-dimensionale si dicono coniche.

Definizione estensione complessa di uno spazio

L'estensione complessa di uno spazio (vettoriale, proiettivo, affine o euclideo) reale è quello spazio a scalari complessi, di cui lo spazio di partenza si può considerare una parte selezionata prendendo tutti i coefficienti e coordinate a parte immaginaria nulla.

Nel caso dello spazio proiettivo consiste nel definire tale spazio rispetto a un campo  $A$  tale che  $A \subseteq \mathbb{C}$ .

Definizione punti e rette immaginarie

I punti e rette immaginari sono quelli non esprimibili con coordinate o coefficienti tutti reali o tutti completamente immaginari, ad esempio il punto  $(i, 2i, 3i)$  è un punto reale, mentre  $(i, 2, 3)$  è un punto immaginario.

Procedo con fare l'intersezione fra una circonferenza e la retta impropria

nell'ampliamento proiettivo complesso dello spazio euclideo 2-dimensionale. Chiamo con  $C$  l'equazione della circonferenza di raggio  $c$ , e con  $r$  la retta impropria.

$$C : x_1^2 + x_2^2 + a x_1 x_0 + b x_2 x_0 + c x_0^2 = 0 \quad r : x_0 = 0$$

andando a mettere a sistema ottengo due soluzioni a meno di un fattore moltiplicativo:

$$x_0 = 0, x_1^2 + x_2^2 = 0 \Rightarrow x_1 = 1; x_2 = \pm i; x_0 = 0$$

Da cui ricavo 2 punti:  $(0, 1, i)$ ,  $(0, 1, -i)$ ; Questi punti sono detti punti ciclici del piano e sono comuni a tutte le circonferenze.

Facendo l'intersezione fra una sfera generica e il piano improprio nell'ampliamento proiettivo complesso dello spazio euclideo 3-dimensionale ottengo, attraverso calcoli analoghi ai precedenti, quest'equazione:

$$C_a : x_1^2 + x_2^2 + x_3^2 = 0; \quad x_0 = 0$$

Questa è l'equazione della conica assoluta; essa appartiene al piano improprio nella sua parte immaginaria dell'ampliamento proiettivo complesso dello spazio euclideo 3-dimensionale, ha caratteristiche analoghe ai punti ciclici precedentemente definiti, infatti tutte le sfere appartenenti al suddetto ampliamento passano per la conica assoluta. Vado ad illustrare alcuni tipi di trasformazioni in modo da poter evidenziare alcune proprietà importanti della conica assoluta. Per farlo inizio con il dare alcune definizioni:

#### Definizione di matrice ortogonale

Una generica matrice  $M$  si dice ortogonale quando la sua trasposta coincide con la sua inversa.

#### Definizione di congruenza

La congruenza è una trasformazione lineare fra due spazi euclidei che compone rotazione e traslazione, quindi può essere rappresentata da una matrice e un vettore, tale matrice deve essere ortogonale.

Dato che una congruenza trasforma sfere in sfere e che tutte sfere contengono la conica assoluta si può affermare che la congruenza trasforma punti della conica assoluta in punti della conica assoluta.

#### Definizione di similitudine

La similitudine è una trasformazione composta da una congruenza e da una dilatazione/contrazione, quindi si può essere anch'essa rappresentata da una matrice e un vettore, tale matrice è una matrice ortogonale moltiplicata per un certo coefficiente, il quale determina la dilatazione/contrazione.

Anche la similitudine trasformando sfere in altre sfere anche in questo caso vale lo stesso ragionamento fatto in precedenza.

Questa proprietà della conica assoluta è uno dei motivi per cui risulta importante il suo utilizzo nella camera calibration: la conica assoluta è invariante rispetto alla trasformazione da  $S_0$  a  $S_k$ , si preserva al variare della posizione della fotocamera e

anche la sua immagine risulterà invariata allo scattare di diverse foto in posizioni differenti con la stessa fotocamera.

Risulta così un legame fra i parametri intrinseci e la conica assoluta che vado adesso ad approfondire. La seguente trattazione è ripresa da [3].

Per prima cosa definisco un vettore  $\mathbf{d} = [x_1, x_2, x_3]^T$ ; esprimo l'equazione della conica assoluta attraverso il vettore  $\mathbf{d}$  tralasciando  $x_0=0$  dato che è fissa e non porta alcun cambiamento alla seguente spiegazione; esprimo dunque l'immagine della conica assoluta ( $C_a^*$ ) nello spazio come

$$C_a^* : \mathbf{d}^T \mathbf{d} = 0$$

riprendo la formula (2) ottenuta nella trattazione sulla camera matrix  $P$

$$P = KR [ I_3 | \mathbf{t} ]$$

vado a trasformare  $[\mathbf{d} \ 0]$  (punto improprio della scena) in  $\mathbf{v}$

$$\mathbf{v} = P [\mathbf{d} \ 0]^T = K R \mathbf{d}$$

esprimo  $\mathbf{d}$  in funzione di  $\mathbf{v}$  il quale corrisponde alla proiezione di  $[\mathbf{d} \ 0]$  nella fotografia

$$\mathbf{d} = R^{-1} K^{-1} \mathbf{v}$$

quindi

$$\mathbf{d}^T = \mathbf{v}^T K^{-1T} R^{-1T}$$

Dato che  $R$  è una matrice che rappresenta una rotazione ed è una matrice ortogonale, la sua inversa corrisponde con la sua trasposta. Da ciò:

$$\mathbf{d}^T \mathbf{d} = \mathbf{v}^T K^{-1T} R R^{-1} K^{-1} \mathbf{v} = \mathbf{v}^T K^{-1T} K^{-1} \mathbf{v}$$

da questa espressione deduco che per far sì che  $\mathbf{v}$  sia un punto dell'immagine della conica assoluta devo avere

$$\mathbf{v}^T K^{-1T} K^{-1} \mathbf{v} = 0$$

questa espressione è la stessa considerata per rappresentare le immagini delle iperquadriche attraverso il loro discriminante; quindi risulta che l'immagine della conica assoluta proiettata nella fotografia è l'immagine della conica  $w$  tale che il suo discriminante sia  $K^{-1T} K^{-1}$ .

## 2.3 Geometria epipolare e matrice fondamentale

La seguente trattazione è tratta da [4].

La geometria epipolare è la geometria proiettiva intrinseca tra due viste. È indipendente dalla struttura della scena e dipende solo dai parametri interni delle fotocamere e dalla relativa posa (orientamento e posizione). La matrice fondamentale  $F$  incapsula questa geometria intrinseca. È una matrice  $3 \times 3$  di rango 2.

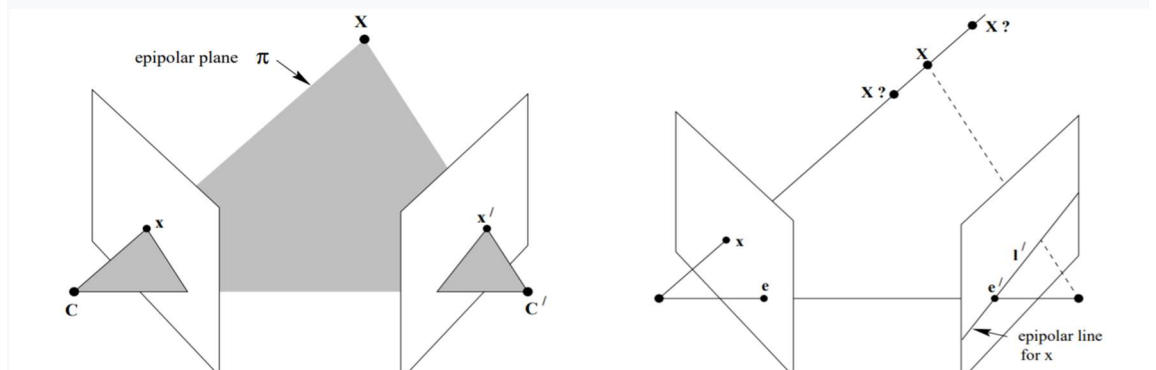
Se un punto nello spazio  $X$  è rappresentato come  $x$  nella prima vista e  $x'$  nel secondo, i punti dell'immagine soddisfano la relazione  $x'^T F x = 0$ , dove  $F$  è la matrice fondamentale. Se la calibrazione interna della fotocamera è nota, si mostra che il moto euclideo delle fotocamere tra le viste può essere calcolato dalla fondamentale matrice fino a un numero finito di ambiguità. La matrice fondamentale è indipendente dalla struttura della scena. Tuttavia, può essere calcolata solo dalle corrispondenze dei punti della scena acquisita, senza richiedere la conoscenza dei parametri interni delle fotocamere o della relativa posa.

### 2.3.1 Geometria epipolare

La geometria epipolare tra due viste è essenzialmente la geometria dell'intersezione dei piani dell'immagine con il fascio proprio dei piani aventi la linea di base come asse (la linea di base è la linea che unisce le diverse posizioni del centro della fotocamera).

Questa geometria è solitamente motivata considerando la ricerca di punti corrispondenti nella corrispondenza stereo.

Considero un punto  $X$  nello spazio, il quale sia rappresentato in due viste,  $x$  nella prima e  $x'$  nella seconda.



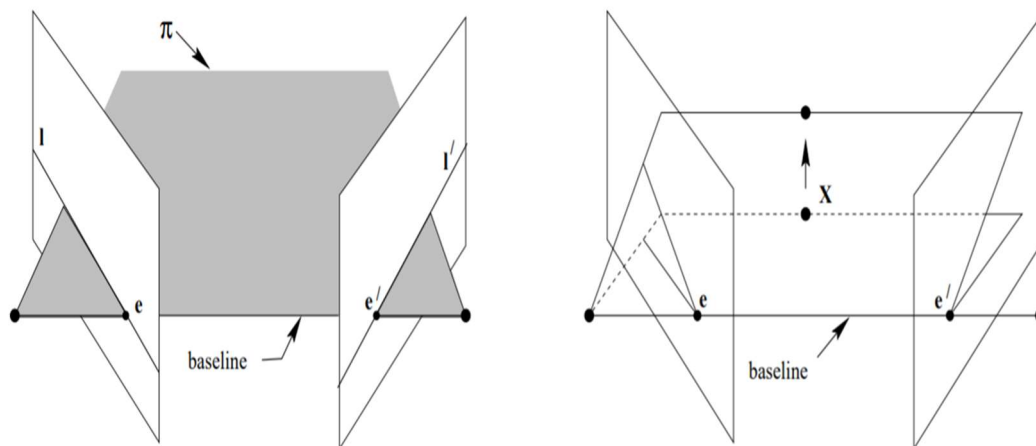


Come mostrato nella figura precedente i punti  $x$  e  $x'$ , il punto spaziale  $X$  e le diverse posizioni del centro della fotocamera sono complanari (il piano considerato lo definisco come  $\pi$ ).

Chiaramente, i raggi retroproiettati da  $x$  e  $x'$  si intersecano in  $X$ , e i raggi sono complanari, giacendo in  $\pi$ .

Supponendo ora di conoscere solo  $x$ , illustro come il punto corrispondente  $x'$  è vincolato. Il piano  $\pi$  è determinato dalla linea di base e dal raggio definito da  $x$ . A partire da quanto precedentemente riportato so che il raggio corrispondente al punto (sconosciuto)  $x'$  si trova in  $\pi$ , quindi il punto  $x'$  si trova sulla linea di intersezione  $l'$  fra  $\pi$  e il secondo piano immagine. Questa linea  $l'$  è l'immagine nella seconda vista del raggio retroproiettato da  $x$ : è la linea epipolare corrispondente a  $x$ . Risulta pertanto che il punto  $x'$  ha come vincolo quello di giacere sulla retta  $l'$ , perciò la ricerca del punto corrispondente a  $x$  non deve necessariamente coprire l'intero piano dell'immagine ma può essere limitato alla linea  $l'$ .

Le entità geometriche coinvolte nella geometria epipolare sono illustrate nella figura seguente.



### Definizione epipolo

L'epipolo è il punto di intersezione della linea che unisce i centri della fotocamera (la linea di base) con il piano dell'immagine. Allo stesso modo, l'epipolo è l'immagine in una vista del centro della fotocamera dell'altra vista. È anche il punto di fuga della linea di base.

### Definizione piano epipolare

Un piano epipolare è un piano contenente la linea di base.

Esiste un fascio proprio ad un parametro di piani epipolari.

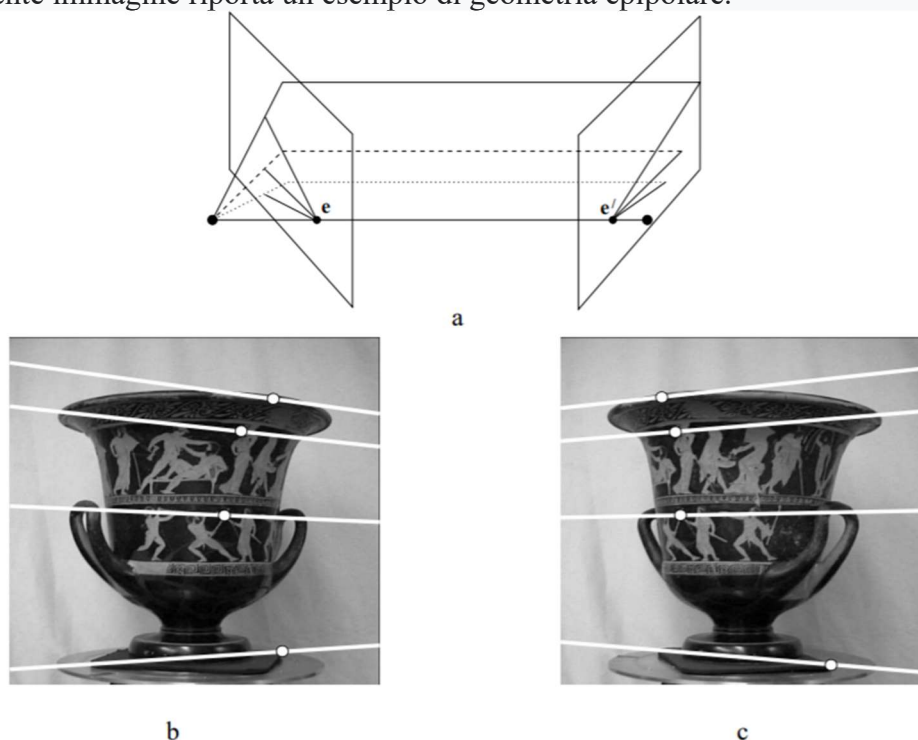
### Definizione linea epipolare

Una linea epipolare è l'intersezione di un piano epipolare con il piano dell'immagine.

Tutte le linee epipolari si intersecano nell'epipolo.

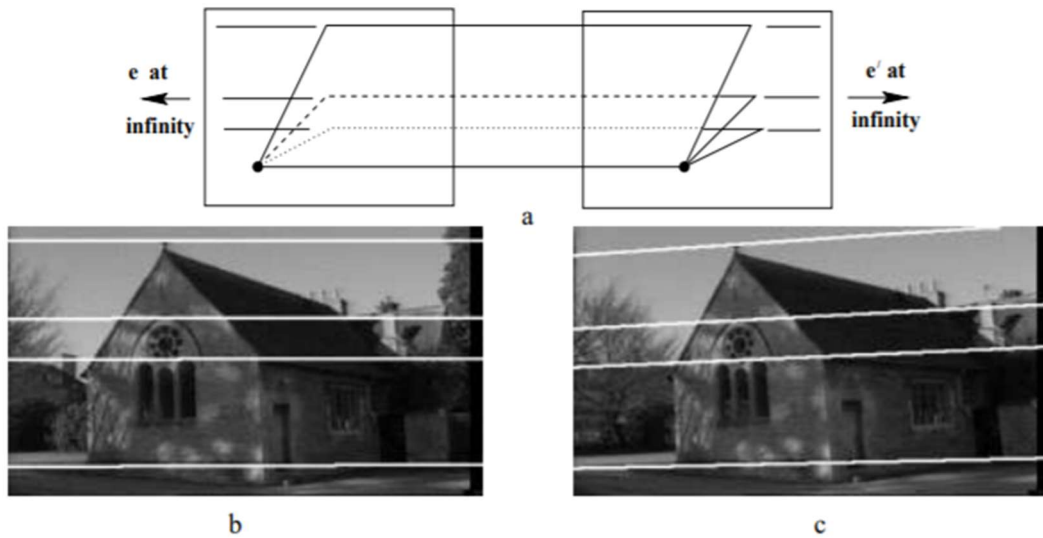
Un piano epipolare interseca entrambi i piani dell'immagine in linee epipolari, e definisce la corrispondenza tra le linee.

La seguente immagine riporta un esempio di geometria epipolare.



(a) Geometria epipolare per fotocamere convergenti. (b) e (c) Un paio di immagini con punti corrispondenti sovrapposti e le loro linee epipolari (in bianco). Il movimento tra le viste sono una traslazione e una rotazione. In ciascuna immagine si può dedurre la direzione dell'altra fotocamera dall'intersezione della matita di linee epipolari. In questo caso, entrambi gli epipoli si trovano al di fuori del visibile immagine.

Riporto in seguito un altro esempio.



Nel caso di un movimento speciale in cui si trova la direzione di traslazione parallela al piano dell'immagine e l'asse di rotazione è perpendicolare al piano dell'immagine, l'intersezione della linea di base con il piano dell'immagine è all'infinito. Di conseguenza gli epipoli sono all'infinito e le linee epipolari sono parallele. (a) Geometria epipolare per il movimento parallelo al piano dell'immagine. (b) e (c) un paio di immagini per le quali il movimento tra le viste è (approssimativamente) una traslazione parallela all'asse  $x$ , con nessuna rotazione. Quattro linee epipolari corrispondenti sono sovrapposte in bianco.

### 2.3.2 Matrice fondamentale

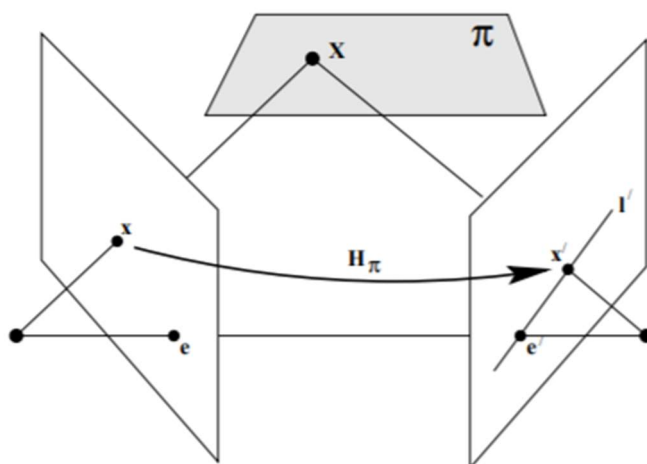
La matrice fondamentale è la rappresentazione algebrica della geometria epipolare. Data una coppia di immagini, per ogni punto  $x$  in un'immagine esiste una corrispondente linea epipolare  $l'$  nell'altra immagine. Qualsiasi punto  $x'$  nella seconda immagine, corrispondente al punto  $x$ , deve trovarsi sulla linea epipolare  $l'$ . La linea epipolare è la proiezione nella seconda immagine del raggio dal punto  $x$  attraverso il centro  $C$  della prima fotocamera. Quindi c'è un'associazione  $x \rightarrow l'$  da un punto in un'immagine alla sua linea epipolare corrispondente nell'altra immagine. Questa associazione è una correlazione (singolare), cioè una corrispondenza proiettiva da punti a linee, che è rappresentata da una matrice  $F$ , la matrice fondamentale.

Di seguito mostro una derivazione geometrica della matrice fondamentale. La

mappatura da un punto in un'immagine a una linea epipolare corrispondente nell'altra immagine può essere scomposto in due fasi. Nella prima fase, il punto  $x$  viene mappato su un punto  $x'$  nell'altra immagine che giace sulla linea epipolare  $l'$ . Questo punto  $x'$  è una potenziale corrispondenza per punto  $x$ . Nella seconda fase, la linea epipolare  $l'$  si ottiene come la linea che unisce  $x'$  all'epipolo  $e'$ .

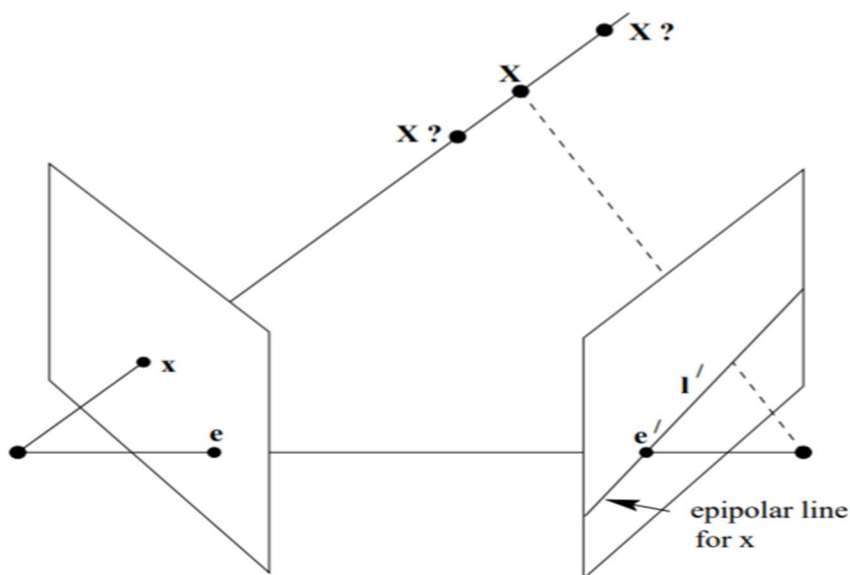
Passaggio 1: trasferimento del punto tramite un piano.

Nella seguente trattazione faccio riferimento alla figura sotto riportata.



Un punto  $x$  in un'immagine viene trasferito tramite il piano  $\pi$  a un punto di corrispondenza  $x'$  nella seconda immagine. La linea epipolare passante per  $x'$  si ottiene unendo  $x'$  all'epipolo  $e'$ . In simboli ho  $x' = H\pi x$  and  $l' = e' \times x' = e' \times H\pi x = Fx$  dove  $F = e' \times H\pi$  è la matrice fondamentale.

Considero un piano  $\pi$  nello spazio non passante per nessuno dei due centri delle fotocamere. Il raggio attraverso il centro della prima fotocamera corrispondente al punto  $x$  incontra il piano  $\pi$  in un punto  $X$ . Questo punto  $X$  è quindi proiettato in un punto  $x'$  nella seconda immagine. Questa procedura è nota come trasferimento tramite il piano  $\pi$ . Poiché  $X$  giace sul raggio corrispondente a  $x$ , il punto proiettato  $x'$  deve giacere sulla linea epipolare  $l'$  corrispondente all'immagine di questo raggio, come illustrato nella figura seguente.



I punti  $x$  e  $x'$  sono entrambi immagine del punto nello spazio  $X$ , il quale giace sul piano  $\pi$ . L'insieme di tutti questi punti  $x_i$  nella prima immagine e dei punti corrispondenti  $x_i'$  nella seconda immagine sono proiettivamente equivalenti, poiché sono proiettivamente equivalenti all'insieme di punti planari  $X_i$ . Quindi esiste un'omografia  $H_\pi$  che mappa ogni  $x_i$  a  $x_i'$ .

Passaggio 2: Costruire la linea epipolare.

Dato il punto  $x'$ , la linea epipolare  $l'$  di passaggio attraverso  $x'$  e l'epipolo  $e'$  può essere scritta come  $l' = e' \times x'$ . Poiché  $x'$  può essere scritto come  $x' = H_\pi x$  ho

$$l' = e' \times H_\pi x = Fx$$

Definisco  $F = e' \times H_\pi$  matrice fondamentale.

Dal punto di vista geometrico,  $F$  rappresenta una mappatura dal piano proiettivo bidimensionale della prima immagine al fascio di linee epipolari attraverso l'epipolo  $e'$ . Quindi, rappresenta una trasformazione da uno spazio proiettivo bidimensionale a uno monodimensionale, quindi deve avere il rango 2.

La matrice fondamentale è l'unica matrice omogenea di rango 2  $3 \times 3$  che supponendo di avere due immagini acquisite da fotocamere con centri non coincidenti, soddisfa

$$x'^T F x = 0$$

per tutti i punti corrispondenti  $x \leftrightarrow x'$ .

Adesso vado ad elencare le principali proprietà della matrice fondamentale:

1. Se  $\mathbf{F}$  è la matrice fondamentale della coppia di fotocamere  $(\mathbf{P}, \mathbf{P}')$ , allora  $\mathbf{F}^T$  è la matrice fondamentale della coppia nell'ordine opposto:  $(\mathbf{P}', \mathbf{P})$ . Dove  $\mathbf{P}$  e  $\mathbf{P}'$  sono le camera matrix delle due fotocamere.
2. Per ogni punto  $\mathbf{x}$  nella prima immagine, l'epipolare corrispondente la linea è  $\mathbf{l}' = \mathbf{F}\mathbf{x}$ . Allo stesso modo,  $\mathbf{l} = \mathbf{F}^T \mathbf{x}'$  rappresenta la linea epipolare corrispondente a  $\mathbf{x}'$  nella seconda immagine.
3. Per ogni punto  $\mathbf{x}$  (diverso da  $\mathbf{e}$ ) la linea epipolare  $\mathbf{l}' = \mathbf{F}\mathbf{x}$  contiene l'epipolo  $\mathbf{e}'$ . Così  $\mathbf{e}'$  soddisfa  $\mathbf{e}'^T (\mathbf{F}\mathbf{x}) = (\mathbf{e}'^T \mathbf{F}) \mathbf{x} = 0$  per ogni  $\mathbf{x}$ . Ne consegue che  $\mathbf{e}'^T \mathbf{F} = 0$ , cioè  $\mathbf{e}'$  è il vettore annullatore sinistro di  $\mathbf{F}$ . Allo stesso modo  $\mathbf{F}\mathbf{e} = 0$ , cioè  $\mathbf{e}$  è il destro vettore annullatore di  $\mathbf{F}$ .

### 2.3.3 Matrice essenziale

La matrice essenziale è la specializzazione della matrice fondamentale nel caso di coordinate dell'immagine normalizzate (le quali verranno spiegate in seguito). Storicamente, la matrice essenziale è stata introdotta (da Longuet-Higgins) prima della matrice fondamentale; la matrice fondamentale può essere pensata come la generalizzazione della matrice essenziale in cui l'ipotesi (non essenziale) di fotocamere calibrate viene rimossa.

Considero una matrice di fotocamera come  $\mathbf{P} = \mathbf{K} [\mathbf{R} \mid \mathbf{t}]$ , e sia  $\mathbf{x} = \mathbf{P}\mathbf{X}$  un punto nell'immagine. Se la matrice di calibrazione  $\mathbf{K}$  è nota, allora posso applicare la sua inversa al punto  $\mathbf{x}$  per ottenere il punto  $\mathbf{x}^* = \mathbf{K}^{-1} \mathbf{x}$ . Allora  $\mathbf{x}^* = [\mathbf{R} \mid \mathbf{t}] \mathbf{X}$ , dove  $\mathbf{x}^*$  è il punto dell'immagine espresso in coordinate normalizzate. Può essere pensato come l'immagine del punto  $\mathbf{X}$  rispetto ad una fotocamera  $[\mathbf{R} \mid \mathbf{t}]$  avente la matrice identità  $\mathbf{I}$  come matrice di calibrazione. La matrice della fotocamera  $\mathbf{K}^{-1}\mathbf{P} = [\mathbf{R} \mid \mathbf{t}]$  è chiamata camera matrix normalizzata, essendo stato rimosso l'effetto della matrice di calibrazione nota. Consideriamo ora una coppia di matrici di fotocamere normalizzate  $\mathbf{P} = [\mathbf{I} \mid \mathbf{0}]$  e  $\mathbf{P}' = [\mathbf{R} \mid \mathbf{t}]$ . La matrice fondamentale corrispondente alla coppia di fotocamere normalizzate è abitualmente chiamata matrice essenziale, e ha la forma

$$\mathbf{E} = \mathbf{t} \times \mathbf{R}$$

In termini di coordinate dell'immagine normalizzate ho che

$$\mathbf{x}^{*\prime T} \mathbf{E} \mathbf{x}^* = 0$$

per punti corrispondenti  $\mathbf{x} \leftrightarrow \mathbf{x}'$ .

Sostituendo ottengo  $\mathbf{x}'^T \mathbf{K}'^{-1 T} \mathbf{E} \mathbf{K}^{-1} \mathbf{x} = 0$ , comparando questa relazione con la relazione che lega punti corrispondenti alla matrice fondamentale ho:

$$\mathbf{E} = \mathbf{K}'^T \mathbf{F} \mathbf{K}$$

Come la matrice fondamentale anche la matrice essenziale è un'entità omogenea, infatti entrambe sono definite a meno di un fattore moltiplicativo.

## 2.4 Equazione di Kruppa

In questo paragrafo andrò a mostrare due derivazioni dell'equazioni di Kruppa, iniziando con la derivazione dalla matrice fondamentale. Questa derivazione è tratta da [5].

Considero due camera matrix  $\mathbf{P}$  e  $\mathbf{P}'$  con la stessa calibrazione. Sia  $\mathbf{C}$  il duale dell'immagine della conica assoluta (abbreviata DIAC) ripresa da queste due fotocamere. L'immagine della conica assoluta (IAC) è indipendente dalla posa della fotocamera, e quindi è la stessa per le due fotocamere in questione.

Sia  $\mathbf{F}$  la matrice fondamentale per la coppia di fotocamere. Voglio applicare alle due trasformazioni proiettive rappresentate dalle matrici di trasformazione  $3 \times 3$   $\mathbf{A}$  e  $\mathbf{A}'$  immagini. Dopo la trasformazione, le matrici effettive della fotocamera saranno  $\mathbf{A}\mathbf{P}$  e  $\mathbf{A}'\mathbf{P}'$ , corrispondenti alla proiezione della fotocamera seguita dalla trasformazione proiettiva dell'immagine. Questo ovviamente cambierà il DIAC con nuove coniche, definite  $\mathbf{D}$  e  $\mathbf{D}'$ . Poiché  $\mathbf{A}$  e  $\mathbf{A}'$  possono essere diverse, non posso più assumere che  $\mathbf{D} = \mathbf{D}'$ . Suppongo che  $\mathbf{A}$  e  $\mathbf{A}'$  siano scelti in modo che la matrice fondamentale per le due nuove fotocamere matrici  $\mathbf{A}\mathbf{P}$  e  $\mathbf{A}'\mathbf{P}'$  abbia la forma speciale

$$E = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Questa ha la proprietà di avere i due epipoli all'origine e le corrispondenti linee epipolari sono identiche nelle due immagini. Considero ora un piano che passa per i due centri delle fotocamere, tangente alla conica assoluta. Un tale piano proietterà una coppia di linee epipolari corrispondenti nelle due immagini, e queste due linee saranno tangenti alla IAC. Poiché ci sono due di questi piani tangenti, ci sono due coppie di tangenti epipolari corrispondenti. Sapendo che le linee epipolari corrispondenti nelle due immagini sono identiche, sia  $(a, b, 0)^T$  tangente alla IAC, poiché  $\mathbf{D}$  è il DIAC nella prima immagine, questa relazione di tangenza può essere scritta come

$$(a,b,0)\mathbf{D}(a,b,0)^T=0$$

Stessa cosa vale per  $\mathbf{D}'$ ; vado ad esplicitare queste 2 espressioni

$$a^2 d_{11} + 2abd_{12} + b^2 d_{22} = 0 \quad \text{e} \quad a^2 d'_{11} + 2abd'_{12} + b^2 d'_{22} = 0$$

Dove  $\mathbf{D}=(d_{ij})$   $\mathbf{D}'=(d'_{ij})$ , entrambe sono matrici simmetriche.

Poiché le due linee tangenti alla IAC devono essere le stesse due linee nelle due immagini, queste due equazioni devono avere la stessa coppia di soluzioni per a e b. Ciò significa che devono essere equazioni identiche (a meno di un fattore di scala) e così ho

$$\frac{d_{11}}{d'_{11}} = \frac{d_{12}}{d'_{12}} = \frac{d_{22}}{d'_{22}}$$

Le due equazioni sopra definite sono le cosiddette equazioni di Kruppa.

Adesso voglio riproporre la derivazione delle equazioni di Kruppa illustrata nell'articolo [6].

Parto fornendo l'equazione della conica assoluta  $C_a$

$$\mathbf{x}_1^2 + \mathbf{x}_2^2 + \mathbf{x}_3^2 = \mathbf{x}_4 = \mathbf{0}$$

usando la proprietà invariante di  $C_a$ , per cui le immagini  $\omega$  e  $\omega'$ , delle rispettive fotocamere, dipendono solo dai parametri intrinseci. I piani tangenti di  $C_a$  sono anche tangenti a  $\omega$  e  $\omega'$  attraverso trasformazioni epipolari.

La trasformazione epipolare definisce un'omografia che chiamo  $\mathbf{H}$  dalle linee per e alle linee per e' (i due epipoli) tali che  $\mathbf{l}$  ed  $\mathbf{l}'$  corrispondano.



Supponendo che il punto di intersezione dell'epipolo  $\mathbf{e}$  con la linea all'infinito sia  $[y_1, y_2, 0]$ .  $\mathbf{l}$  è quindi una linea epipolare ( $\mathbf{e} \times \mathbf{y}$ ). Qualsiasi punto  $\mathbf{x}$  giace sulla linea se e solo se soddisfa

$(\mathbf{e} \times \mathbf{y})^T \mathbf{x} = 0$ . Se  $\mathbf{D}$  è il duale di  $\omega$  allora soddisfa

$$(\mathbf{e} \times \mathbf{y})^T \mathbf{D} (\mathbf{e} \times \mathbf{y}) = 0$$

e  $\mathbf{D}$  è rappresentato come una matrice simmetrica, come descritto da [Faugeras et al. (1992)], come segue

$$D = \begin{bmatrix} -g_{23} & g_3 & g_2 \\ g_3 & -g_{13} & g_1 \\ g_2 & g_1 & -g_{12} \end{bmatrix}$$

Questa matrice è individuata da 5 parametri dato che è definita a meno di un fattore moltiplicativo.

Adesso vado a sostituire  $\mathbf{D}$  nella formula precedente e trovo

$$d_{11} y_1^2 + 2d_{12} y_1 y_2 + d_{22} y_2^2 = 0$$

dove

$$d_{11} = -g_{13} e_3^2 - g_{12} e_2^2 - 2g_1 e_2 e_3$$

$$d_{12} = g_{12} e_1 e_2 - g_3 e_3^2 + g_2 e_2 e_3 + g_1 e_1 e_3$$

$$d_{22} = -g_{23} e_3^2 - g_{12} e_1^2 - 2g_2 e_1 e_3$$

Applicando lo stesso procedimento per l'altra vista

$$d'_{11} y'_1{}^2 + 2d'_{12} y'_1 y'_2 + d'_{22} y'_2{}^2 = 0$$

La coppia  $\mathbf{l}$  e  $\mathbf{l}'$  si corrisponde rispetto a l'omografia  $\mathbf{H}$ , se e solo se soddisfa  $\mathbf{y}' = \mathbf{N}\mathbf{y}$ .

Trasformazione dalla retta  $y_3 = 0$  a  $y'_3 = 0$  è data dalla trasformazione bilineare  $\mathbf{N}$ . Per

$\tau = \frac{y_2}{y_1}$ ,  $\tau' = \frac{y'_2}{y'_1}$  quindi ottengo che la trasformazione  $\mathbf{N}$  è equivalente a

$$\tau' = \frac{a\tau + b}{c\tau + d}$$

dove  $a, b, c, d$  possono essere calcolati fino a meno di un fattore di scala dai due epipoli  $\mathbf{e}, \mathbf{e}'$  e un insieme di punti che si corrispondono  $\mathbf{q}_i \leftrightarrow \mathbf{q}'_i$ ,  $1 \leq i \leq n$ . Attraverso una

procedura lineare dei minimi quadrati basata sull'equazione vista sopra  $d'_{11}y'_1{}^2 + 2d'_{12}y'_1y'_2 + d'_{22}y'_2{}^2 = 0$  con

$$\tau_i = \frac{e_3 q_{i2} - e_2 q_{i3}}{e_3 q_{i1} e_1 q_{i3}}; \tau_i' = \frac{e_3' q_{i2}' - e_2' q_{i3}'}{e_3' q_{i1}' e_1' q_{i3}'}$$

Dopo aver trovato a,b,c,d posso esprimere le equazioni precedenti come

$$d_{11} + 2d_{12}\tau + d_{22}\tau^2 = 0 \quad \text{e} \quad d'_{11}(b\tau + c)^2 + 2d'_{12}(b\tau + c)(\tau + a) + d'_{22}(\tau + a)^2 = 0$$

Queste due equazioni sono entrambe quadratiche rispetto a  $\tau$  e hanno le stesse radici, facendo il rapporto incrociato ottengo le equazioni di Kruppa

$$\frac{d_{11}}{d_{11}'(b\tau + c)^2} = \frac{2d_{12}\tau}{2d_{12}'(b\tau + c)(\tau + a)} = \frac{d_{22}\tau^2}{d_{22}'(\tau + a)^2}$$

Tramite le quali ottengo i parametri della matrice **D**.

Un punto  $X$  è sulla conica assoluta se soddisfa  $X^T C_a X = 0$ . A questo punto risulta che la conica assoluta  $C_a$  forma una matrice identità **I** sul piano all'infinito  $x_4 = 0$ .

L'equazione diventa  $X^T I X = 0$ , che è uguale a  $X^T X = 0$ .  $X$  viene proiettato sul piano dell'immagine come  $x = P X$ . Considerando solo la proiezione sul primo piano dell'immagine, dove la rotazione è la matrice identità e la fotocamera non è traslata, la matrice di proiezione è  $P = K [I \mid 0]$  e il punto  $X$  può essere scritto come  $(x, 0)$ . Da questa considerazione si nota che  $P = K$ . Ciò significa che  $P$  è invertibile, perché

$|P| \neq 0$  e quindi  $X = P^{-1}x$ . Per sostituzione, ottengo l'equazione

$(P^{-1}x)^T (P^{-1}x) = 0$  che è uguale a  $x^T P^{-1T} P^{-1} x$ . Sostituendo  $P$  con  $K$  si ottiene

$x^T K^{-1T} K^{-1} x = 0$ . Vorrei far notare l'equivalenza fra questa equazione e l'equazione

dei punti immagine della conica assoluta presentata nel capitolo 3.2 cioè  $x^T \omega x = 0$

con  $\omega = K^{-1T} K^{-1}$ . Ciò conferma che l'immagine della conica assoluta dipende solo

dalla parte intrinseca della fotocamera **K**. Questo fatto è cruciale per la spiegazione del metodo che andrò a presentare nel capitolo 4.2.

# Capitolo 3 Metodi di risoluzione

In questo capitolo andrò a presentare 3 metodi per effettuare una camera calibration, il primo è un metodo standard, cioè che utilizza degli oggetti con proprietà note, nello specifico questo metodo utilizza tre palle di taglia arbitraria; il secondo e il terzo metodo sono della tipologia self-calibration, non si avvalgono di nessun oggetto speciale ma utilizzano una sequenza di fotografie a diverse angolazioni. Il primo metodo è tratto da [7], il secondo è tratto da [8], il terzo è tratto da [6].

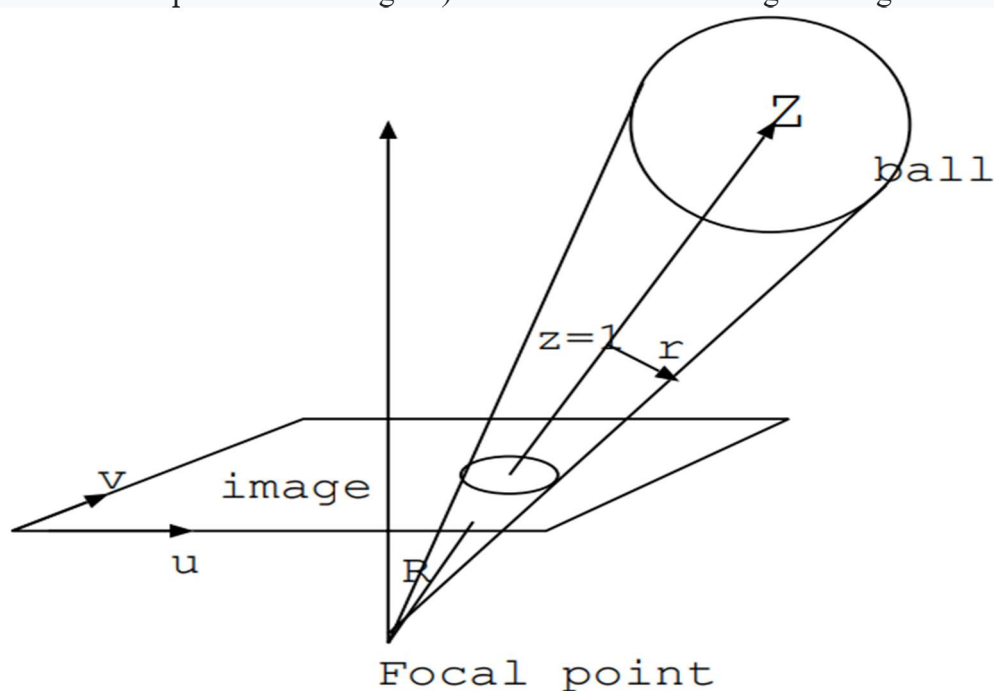
## 3.1 Metodo delle 3 palle

Questo metodo ha come obiettivo ricavare i parametri intrinseci della fotocamera da una singola immagine di palle. Le sfere vengono proiettate sull'immagine come ellissi, ciascuna di queste proiezioni fornisce due vincoli sui parametri intrinseci. Per stimare tutti e cinque i parametri intrinseci, abbiamo bisogno di tre palle. La dimensione della palla è arbitraria. È più flessibile di altri oggetti di calibrazione predefiniti. Viene utilizzata solo un'immagine e non ci sono problemi di corrispondenza coinvolti. Inoltre questo metodo è in grado di stimare oltre ai parametri intrinseci anche la dimensione e la posizione relative di ciascuna palla in modo ottimale e simultaneamente dai punti di confine di ogni palla.

In letteratura, le palle sono state utilizzate per la prima volta da Penna [9] per la camera calibration, il quale ha solo provato a calcolare le proporzioni dei due assi dell'immagine. Nel tentativo di calibrare fotocamere da linee di fuga, [10] dimostra che da superfici di rotazione che hanno linee parallele, le traiettorie dei punti di fuga sono ellissi [10] da ciò è stato trovato che gli assi maggiori delle ellissi si intersecano nel punto principale. Sulla base di questa osservazione, è stato proposto di determinare prima le proporzioni utilizzando tre ellissi, quindi determinare il punto principale e infine determinare la lunghezza focale. Parallelamente è stato proposto di utilizzare “sfere” per la camera calibration [11], tramite le quali sono stati ottenuti risultati simili a quelli di [10]. Rispetto a questi approcci, questo metodo non solo fornisce una

formulazione matematica generale del problema che collega le coniche con il concetto di conica assoluta, ma fornisce anche un algoritmo efficiente per trovare i parametri intrinseci e le relative posizioni e dimensioni delle palle, in maniera ottimale.

Le palle, viste da qualsiasi direzione, sono rotonde. Il bordo è sempre una circonferenza. La circonferenza e il punto focale della fotocamera formano un cono circolare, la cui intersezione con il piano dell'immagine diventa un'ellisse (diventa una parabola se la palla interseca il piano dell'immagine.) Come illustrato nella seguente figura



definisco l'origine del sistema di coordinate per il cono circolare in modo che sia lo stesso del punto focale della fotocamera e in modo che l'asse Z vada dall'origine al centro della circonferenza o, equivalentemente, il centro del palla. Tra il sistema di coordinate della fotocamera e quello del cono non c'è traslazione ma solo rotazione.

Le coordinate pixel  $\mathbf{m} = [u, v]^T$  e le coordinate dell'immagine normalizzata  $\mathbf{x} = [x, y]^T$  sono correlate dalla seguente equazione

$$\mathbf{m}^* = \mathbf{A}\mathbf{x}^*$$

dove  $\mathbf{m}^* = [u, v, 1]^T$  e  $\mathbf{x}^* = [x, y, 1]^T$  sono le coordinate omogenee del punto in coordinate pixel e coordinate dell'immagine normalizzata, rispettivamente, e  $\mathbf{A}$  è la matrice intrinseca, data da

$$A = \begin{bmatrix} a_u & b & u_0 \\ 0 & a_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

con  $(u_0, v_0)$  le coordinate del punto principale in sistema di coordinate pixel,  $a_u, a_v$  le lunghezze focali per gli assi  $u, v$  e  $b$  il parametro di inclinazione (skew) tra gli assi  $u$  e  $v$ . Generalmente, un punto  $\mathbf{M} = [X, Y, Z]^T$  nel sistema di coordinate universali viene proiettato sul punto dell'immagine  $\mathbf{m}$  da

$$\mathbf{sm}^* = \mathbf{A}[\mathbf{R} \ \mathbf{t}] \mathbf{M}^*$$

dove  $s$  è uno scalare,  $\mathbf{M}^* = [X, Y, Z, 1]^T$  sono le coordinate omogenee del punto  $\mathbf{M}$ ,  $\mathbf{R}$  e  $\mathbf{t}$  sono la matrice di rotazione e il vettore di traslazione, rispettivamente, tra il sistema di coordinate mondo e il sistema di coordinate solidale con la fotocamera.

Considero il vettore traslazione  $\mathbf{t}$  nullo. Riscrivo quindi l'uguaglianza precedente

$$\mathbf{sm}^* = \mathbf{A} \mathbf{R} \mathbf{M}$$

Elaborando quest'espressione ottengo

$$s \mathbf{R}^T \mathbf{A}^{-1} \mathbf{m}^* = \mathbf{M} \quad (4.1)$$

Affinché le terze componenti di entrambi i lati siano uguali, devo avere

$$s = \frac{Z}{\mathbf{r}_3^T \mathbf{A}^{-1} \mathbf{m}^*} \quad (4.2)$$

dove  $\mathbf{r}_3^T$  è il terzo vettore colonna di  $\mathbf{R}$ . Si può verificare facilmente che  $\mathbf{r}_3^T$  è l'asse centrale del cono circolare nel sistema di coordinate della fotocamera. Conoscere  $\mathbf{r}_3^T$  significa conoscere la posizione della palla rispetto alla profondità. Suppongo che il cono abbia raggio  $r$  in  $Z = 1$ . Moltiplicando la trasposta di ciascun lato di (4.1) da sinistra ho

$$\mathbf{m}^{*T} \mathbf{A}^{-1T} \mathbf{A}^{-1} \mathbf{m}^* = 1/s^2 \mathbf{M}^T \mathbf{M} = 1/s^2 (r^2 + 1) Z^2$$

sostituisco adesso con la (4.2)

$$\mathbf{m}^{*T} \mathbf{A}^{-1T} \mathbf{A}^{-1} \mathbf{m}^* = (r^2 + 1) \mathbf{m}^{*T} \mathbf{A}^{-1T} \mathbf{r}_3 \mathbf{r}_3^T \mathbf{A}^{-1} \mathbf{m}^*$$

portando il membro destro a sinistra abbiamo

$$\mathbf{m}^{*T} \mathbf{Q} \mathbf{m} = 0$$

dove

$$\mathbf{Q} \cong \mathbf{A}^{-1T} \mathbf{A}^{-1} - (r^2 + 1) \mathbf{A}^{-1T} \mathbf{r}_3 \mathbf{r}_3^T \mathbf{A}^{-1}$$

$\mathbf{Q}$  descrive una conica e  $\mathbf{A}^{-1T} \mathbf{A}^{-1}$  corrisponde all'immagine della conica assoluta nella fotografia, dalla quale posso ricavare i parametri intrinseci come mostrato in [12].

$$\begin{aligned} \mathbf{Q} &\cong \mathbf{A}^{-1T} (\mathbf{I}_3 - (r^2 + 1) \mathbf{r} \mathbf{r}^T) \mathbf{A}^{-1} \\ &= \mathbf{A}^{-1T} \mathbf{r}_1 \mathbf{r}_1^T + \mathbf{r}_2 \mathbf{r}_2^T - r^2 \mathbf{r} \mathbf{r}^T \mathbf{A}^{-1} \\ &= \mathbf{A}^{-1T} [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3] \text{diag}(1, 1, -r^2) [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]^T \mathbf{A}^{-1} \end{aligned}$$

si trova che  $\mathbf{Q}$  ha due autovalori positivi e uno negativo.

Dato  $\mathbf{r}_3$ , posso determinare la sua intersezione con il piano immagine. Chiamo le coordinate normalizzate dell'intersezione con l'immagine ( $x_c, y_c$ ); ho dunque

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} \frac{r_{13}}{r_{33}} \\ \frac{r_{23}}{r_{33}} \end{bmatrix}$$

Dove  $\mathbf{r}_3 = [r_{13}, r_{23}, r_{33}]^T$ . Tramite  $r, x_c$  e  $y_c$  posso determinare la dimensione e la posizione relative di una palla a meno della sua profondità.  $r$  e  $(x_c, y_c)$  hanno 3 gradi di libertà. Dal momento che ogni conica fornisce 5 vincoli, solo 2 di essi sono per parametri intrinseci. Dato che il modello della fotocamera presenta 5 parametri intrinseci, abbiamo bisogno di almeno 3 palle.

In tal caso ho 14 parametri indipendenti sconosciuti, 5 parametri intrinseci e 9 parametri,  $r(j), x_c(j), y_c(j), j = 1, 2, 3$ , per modellare la relativa dimensione e posizione delle palle. Come da convenzione, modello il rumore dell'immagine come gaussiana uniforme isotropa. La stima ottimale cerca di trovare i parametri sconosciuti in modo che la somma dei quadrati delle distanze euclidee tra i punti dell'immagine e le rispettive coniche siano ridotte al minimo.

Adesso andrò ad esporre tre difficoltà che si presentano nell'utilizzare questo metodo:

1. La distanza euclidea da un punto a una conica è molto complessa da esprimere esplicitamente;
2. La distanza è una funzione non lineare nei parametri sconosciuti, perciò sono necessari algoritmi di ottimizzazione non lineari, che coinvolgono i differenziali della distanza rispetto ai parametri sconosciuti. Nell'articolo [7] viene utilizzato l'algoritmo di

Levenberg-Marquardt che richiede solo il calcolo dei differenziali del primo ordine [13].

3. Gli algoritmi iterativi non lineari necessitano di buoni valori iniziali.

Supponendo che  $n$  punti vengano estratti dal contorno della palla nell'immagine, la funzione di costo è definita come

$$C = \frac{1}{2} \sum_{i=1}^n \{(u_i - U_j(u_i, v_i))^2 + (v_i - V_j(u_i, v_i))^2\} \quad (4.3)$$

dove  $u_i$  e  $v_i$  sono le coordinate pixel dell' $i$ -esimo punto nell'immagine,  $U_j(u_i, v_i)$  e  $V_j(u_i, v_i)$  sono le coordinate dei punti sull'ellisse della  $j$ -esima palla, a cui appartiene l' $i$ -esimo punto, più vicini a  $(u_i, v_i)$ .

L'immagine di un'ellisse generica può essere espressa in forma canonica dalla seguente equazione

$$\frac{U^2}{a^2} + \frac{V^2}{b^2} = 1$$

La distanza fra un generico punto  $(x, y)$  e l'ellisse sopra descritta può essere ottenuta minimizzando

$$(x-U)^2 + (y-V)^2 + \lambda \left( \frac{U^2}{a^2} + \frac{V^2}{b^2} - 1 \right)$$

dove  $\lambda$  è il moltiplicatore lagrangiano.

Impostando i differenziali dell'espressione precedente rispetto a  $U$  e  $V$  uguale a zero ottengo un'equazione polinomiale di 4° ordine su  $\lambda$ ,

$$\frac{a^2 x^2}{(a^2 + \lambda)^2} + \frac{b^2 y^2}{(b^2 + \lambda)^2} = 1 \quad (4.4)$$

Ci sono 4 soluzioni, 2 delle quali possono essere immaginarie, di cui si scelgono le soluzioni che forniscono la più breve distanza. Sebbene soluzioni analitiche per  $\lambda$  (e quindi per  $U, V$ ) siano disponibili, le espressioni sono lunghe e complesse, le quali diventano ancora più complesse con la trasformazione da un'ellisse generale alla sua forma canonica e differenziando la funzione costo (4.3) rispetto alle incognite.

La strategia adottata per risolvere questa difficoltà consiste nell'approssimare i differenziali mediante calcoli numerici. Utilizzando i valori correnti dei parametri

intrinseci e la dimensione e la posizione relative delle palle, è possibile determinare la trasformazione dal sistema di coordinate dell'immagine pixel al sistema di coordinate canoniche di ciascuna ellisse, quindi è possibile calcolare il punto sull'ellisse più vicino a un dato punto fuori dall'ellisse risolvendo (4.4) e trovando così la distanza minore fra le 4 soluzioni. Per calcoli più dettagliati consultare [7].

Il successo della convergenza al minimo globale dell'algoritmo di Levenberg-Marquardt dipende dalle condizioni iniziali, è stato scoperto che se si assume che il centro dell'immagine sia il punto principale, i pixel sono quadrati ( $a_u = a_v$ ) e l'inclinazione è zero ( $b = 0$ ), quindi gli altri parametri possono essere ottenuti recuperando prima l'ellisse e poi risolvendo alcune equazioni. Vedere [7] per maggiori dettagli. Una volta determinati i parametri intrinseci si può risalire facilmente alle relative posizioni e dimensioni delle palle, per i dettagli vedere [7].

## 3.2 Metodi self-calibration

In questo paragrafo vado a presentare due metodi della tipologia self-calibration, il primo che non utilizza le equazioni di Kruppa, mentre il secondo è uno dei primi metodi che le ha utilizzate.

### 3.2.1 Metodo senza equazioni di Kruppa

In questo paragrafo andrò ad esporre un metodo per la camera calibration basato sulla conica assoluta e sul piano improprio, che utilizza una sequenza di immagini da diverse pose; in ognuna delle pose la camera matrix cambia.

Senza considerare i parametri intrinseci ogni immagine corrisponde ad una proiezione della scena nel suo piano immagine, queste proiezioni sono le camera matrix e possono essere rappresentate come

$$P_1 = [I_3 \mid \mathbf{0}] \text{ e } P_k = [H_{1k} \mid \mathbf{e}_{1k}]$$



dove  $\mathbf{H}_{1k}$  è l'omografia per un piano di riferimento (uguale per tutte le viste) e l'epipolo  $\mathbf{e}_{1k}$  la proiezione nella k-esima immagine rispetto alla prima posa della fotocamera.

Considerando come piano di riferimento il piano improprio posso esprimere la k-esima omografia così

$$\mathbf{H}_{\infty 1k} \sim \mathbf{H}_{1k} + \mathbf{e}_{1k} \mathbf{a}$$

per un generico  $\mathbf{a}$  (il quale è in generale sconosciuto), dato che tutte le omografie sono correlate a questa equazione. Andando a considerare adesso anche i parametri intrinseci posso scrivere le camera matrix come

$$\mathbf{P}_{E1} = \mathbf{K}[\mathbf{I}_3 | \mathbf{0}] \text{ e } \mathbf{P}_{Ek} = \mathbf{K}[\mathbf{R}_k | -\mathbf{R}_k \mathbf{t}_k] \quad (4.5)$$

con  $\mathbf{K}$  matrice dei parametri intrinseci (assunti uguali in tutte le immagini quindi costanti nel tempo), mentre  $\mathbf{R}_k$  e  $\mathbf{t}_k$  rappresentano rispettivamente l'orientamento e la posizione rispetto alla prima fotocamera.

Dall'equazione (4.5) posso scrivere

$$\mathbf{H}_{\infty 1k} \sim \mathbf{H}_{\infty k} \mathbf{H}_{\infty 1}^{-1} \sim \mathbf{K} \mathbf{R}_k \mathbf{K}^{-1}$$

dove  $\mathbf{H}_{\infty k}$  è l'omografia dal piano improprio al k-esimo piano immagine.

Le trasformazioni affini mantengono sempre fisso il piano improprio  $\pi_{\infty}$ , ciò non significa che ogni punto in  $\pi_{\infty}$  sia mappato sullo stesso punto, ma che tutti sono mappati sullo stesso piano  $\pi_{\infty}$ . Le trasformazioni euclidee oltre a  $\pi_{\infty}$  tengono fissa la conica assoluta  $\mathbf{C}_a$ ; perciò anche  $\mathbf{C}_{ak}$  dovrà essere fissa, dato che suppongo di utilizzare la stessa fotocamera e che essa ha  $\mathbf{K}$  costante nel tempo. Questo fatto vale anche per la conica duale a  $\mathbf{C}_{ak}$ ,  $\mathbf{C}_{ak}^{-1}$ , la quale verrà utilizzata per semplificare la trattazione.

Voglio adesso dimostrare che  $\mathbf{C}_a = \mathbf{I}_3$

$$\mathbf{C}_{a1}^{-1} \sim \mathbf{H}_{\infty 1} \mathbf{C}_a^{-1} \mathbf{H}_{\infty 1}^T \sim \mathbf{K} \mathbf{K}^T$$

$$\mathbf{C}_{ak}^{-1} \sim \mathbf{H}_{\infty k} \mathbf{C}_a^{-1} \mathbf{H}_{\infty k}^T \sim \mathbf{K} \mathbf{K}^T$$

Questo fatto può essere utilizzato per calcolare l'immagine del duale della conica assoluta.

Poiché  $\mathbf{C}_a$  appartiene a  $\pi_{\infty}$ , devono valere le seguenti equazioni

$$\mathbf{C}_{ak}^{-1} \sim \mathbf{H}_{\infty 1k} \mathbf{C}_{ak}^{-1} \mathbf{H}_{\infty 1k}^T \quad (4.6)$$

per ogni k. L'uguaglianza esatta (non con un fattore di scala) può essere ottenuta moltiplicando per un opportuno fattore  $\mathbf{H}_{\infty 1k}$  per ottenere  $\det(\mathbf{H}_{\infty 1k}) = 1$ . Nel caso in cui la calibrazione affine illustrata nell'articolo [14] fosse già stata ottenuta, avrei  $\mathbf{H}_{\infty 1k}$  nota

quindi l'equazione (4.6) produrrebbe un insieme di equazioni lineari per i coefficienti di  $\mathbf{C}_{ak}^{-1}$ . Una volta ottenuta  $\mathbf{C}_{ak}^{-1}$ ,  $\mathbf{K}$  può essere ottenuto da esso mediante fattorizzazione di Cholesky [15].

Se invece la calibrazione affine non è stata ottenuta il vettore  $\mathbf{a}$  deve essere aggiunto alle incognite. Nel caso generale ottengo le seguenti equazioni

$$\lambda \mathbf{K} \mathbf{K}^T = [\mathbf{H}_{1k} + \mathbf{e}_{1k} \mathbf{a}] \mathbf{K} \mathbf{K}^T [\mathbf{H}_{1k} + \mathbf{e}_{1k} \mathbf{a}]^T \quad (4.7)$$

Il problema di queste equazioni è che il fattore di scala  $\lambda$  è sconosciuto e quindi va aggiunta un'incognita per ogni equazione, questo schema risolutivo diventa quindi impraticabile per sequenze di immagini più lunghe poiché si ha un fattore di scala aggiuntivo per ogni immagine cioè un'altra incognita.

Per ovviare a questo problema è possibile calcolare questi fattori di scala in funzione dei 3 parametri affini (cioè le componenti del vettore  $\mathbf{a}$ ). Ciò si ottiene facendo il determinante del lato sinistro e destro dell'equazione (4.7):

$$\lambda \det(\mathbf{K} \mathbf{K}^T) = \det([\mathbf{H}_{1k} + \mathbf{e}_{1k} \mathbf{a}]) \det(\mathbf{K} \mathbf{K}^T) \det([\mathbf{H}_{1k} + \mathbf{e}_{1k} \mathbf{a}]^T)$$

da cui ottengo

$$\lambda = \det([\mathbf{H}_{1k} + \mathbf{e}_{1k} \mathbf{a}])^2.$$

che può essere fattorizzato così

$$\lambda = (|\mathbf{e}_{1k} \mathbf{h}_{21k} \mathbf{h}_{31k}| a_1 + |\mathbf{h}_{11k} \mathbf{e}_{1k} \mathbf{h}_{31k}| a_2 + |\mathbf{h}_{11k} \mathbf{h}_{21k} \mathbf{e}_{1k}| a_3 + |\mathbf{h}_{11k} \mathbf{h}_{21k} \mathbf{h}_{31k}|)^2$$

dove  $\det(\mathbf{H}_{1k}) = |\mathbf{h}_{11k} \mathbf{h}_{21k} \mathbf{h}_{31k}|$  e  $a_1, a_2, a_3$  sono i parametri affini cioè le componenti di  $\mathbf{a}$ .

Unendo tutto ottengo  $5(n-1)$  equazioni in 8 incognite dove  $n$  è il numero delle immagini, quindi sono necessarie almeno 3 immagini per determinare tutte le incognite.

Le equazioni possono essere risolte minimizzando la seguente espressione

$$\sum_{k=2}^n \left\| \frac{1}{\lambda} H_{\infty 1k} \mathbf{K} \mathbf{K}^T H_{\infty 1k}^T - \mathbf{K} \mathbf{K}^T \right\|_F$$

dove  $\| \cdot \|_F$  sta a indicare la norma di Frobenius.

### 3.2.2 Metodo self-calibration con equazioni di Kruppa

In questo capitolo andrò a presentare il metodo discusso in [6], il quale utilizza le equazioni di Kruppa presentate nel paragrafo 3.4.

Questo metodo è composto dalle seguenti azioni:

1. Fare 3 spostamenti, per ogni spostamento:
  - 1.1 Trovare le corrispondenze fra i punti delle due immagini;
  - 1.2 Calcolare gli epipoli. Per farlo sono disponibili vari metodi tra cui il metodo di Sturm e il metodo con la matrice fondamentale entrambi illustrati nell'articolo[6] ;
  - 1.3 Calcolare l'omografia delle linee epipolari;
  - 1.4 Calcolare le due equazioni di Kruppa.
2. Risolvere le sei equazioni di Kruppa usando il metodo della continuazione descritto in [DACAMBIARE9];
3. Calcolare i parametri intrinseci.

Due movimenti della telecamera producono due trasformazioni epipolari e quindi quattro vincoli sull'immagine della conica assoluta. La conica assoluta dipende da cinque parametri, quindi le coniche compatibili con i quattro vincoli formano una famiglia unidimensionale  $C$ . La famiglia  $c$  è una curva algebrica che parametrizza le calibrazioni della telecamera compatibili con le due trasformazioni epipolari.

La principale problematica di questo metodo è il rumore, il quale può essere attenuato attraverso una selezione accurata degli spostamenti, infatti si è visto [6] che ci sono trasformazioni epipolari che sono molto sensibili al rumore di pixelizzazione, mentre alcune producono trasformazioni più robuste. In particolare, è stato dimostrato che pure traslazioni portano a casi degeneri, dando così risultati molto sensibili al rumore.

# Conclusione

In questa tesi ho illustrato come attraverso il concetto matematico molto astratto di conica assoluta si possa risolvere un problema concreto che riguarda l'individuazione di parametri interni ed esterni alla fotocamera; da ciò è saltato fuori che la conica assoluta risulta strettamente legata ai parametri interni della fotocamera, per la precisione la sua immagine nella fotografia è completamente individuata dai parametri intrinseci della fotocamera, fatto molto curioso, inoltre conferma quanto concetti che sembrano lontani dalla nostra quotidianità siano nascosti in quello che percepiamo come realtà. Nello specifico ho scelto di evidenziare questo aspetto attraverso tre metodologie proposte in letteratura, molto diverse fra loro, una delle quali utilizza degli oggetti nella fotografia per risalire alla calibrazione mentre le altre si avvalgono solo di una sequenza di fotografie fatte da diverse angolazioni ad una stessa scena, nelle quali si possono trovare dei punti di corrispondenza fra le immagini. Il primo metodo risulta il più efficace dei tre ma nonostante utilizzi oggetti più comuni rispetto ad altri metodi dello stesso genere risulta comunque meno flessibile degli altri due metodi che possono essere utilizzati in molte più circostanze. Il metodo presentato che utilizza le equazioni di Kruppa è uno dei primi metodi che riuscì a svincolare la camera calibration dall'utilizzo di oggetti particolari; nonostante esso non risultasse molto efficace e robusto ha dato il via libera a numerosi metodi sviluppati in seguito che hanno raggiunto livelli molto alti di prestazioni. Il metodo che non utilizza le equazioni di Kruppa è un metodo che ha una buona accuratezza e pecca di robustezza ma risulta molto adatto per un approccio stratificato utile a raggiungere una robustezza ottima.

# Bibliografia

- [1] Teimouri, M., Kabudian, J., Hosseini, S.M., Nadarajah, S.: On the Skewness Parameter Estimation for Stable Distributions. *Markov Processes and Related Fields*. 21. 369-380. (2015)
- [2] Deriche, R., Faugeras O.D.: Tracking Line Segments. *Image and vision computing*, 8(4):261-270.(1990)
- [3] Pribanic, T., Sturm, P., Cifrek, M.: A Practical Way to Initialize Camera Parameters using the Absolute Conic. *IASTED International Conference on Visualization, Imaging, and Image Processing*. (2005)
- [4] Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision* (2nd. ed.). Cambridge University Press, USA (2003)
- [5] Hartley, R.: Kruppa's Equations Derived from the Fundamental Matrix G.E. CRD, Schenectady, NY, 12301 (1998)
- [6] Faugeras, O. D., Luong, Q.-T., Maybank, S. J.: Camera self-calibration: Theory and experiments, in 'Computer Vision-ECCV'92', Springer, pp. 321–334. (1992)
- [7] Teramoto, H. & Xu, G.: Camera calibration by a single image of balls: From conics to the absolute conic. *Proc. Asian Conf. Computer Vision*. (2002)
- [8] Pollefeys, M., Van Gool, L.: Self-calibration from the absolute conic on the plane at infinity.(2006)
- [9] Pennai, M.: Camera calibration: A quick and easy way to determine the scale factor. *IEEE Trans. PAMI*, 13(12):1240–1245. (1991)
- [10] Beardsley, P., Murray, D., Zisserman, A.: Camera calibration using multiple images. *Computer Vision --- ECCV'92 Springer Berlin Heidelberg Berlin, Heidelberg* 312--320. (1992)
- [11] Daucher, N., Dhome, M., Lapreste, J.T.: Camera calibration from spheres images. In *Proc. Third European Conf. Comput. Vision*, pages 449–454, Stockholm. (1994)
- [12] Maybank, S.J., Faugeras, O.D.: A theory of selfcalibration of a moving camera. *Int'l J. Comput. Vision*, 8(2):123–152. (1992)
- [13] Press W.H., Flannery B.P., Teukolsky, S.A., Vetterling, W.T.: *Numerical Recipes in C*. Cambridge University Press. (1988)
- [14] Sturm, P. Quan, L., Hlaváč, V., Šára, R.: Affine stereo calibration. *Computer Analysis of Images and Patterns*. Springer Berlin Heidelberg Berlin, Heidelberg 838-843. (1995)
- [15] Haddad, C.: Cholesky factorization Cholesky Factorization. (2008)

# Sitografia

[s1] <http://www.tridmetrix.com>

[s2] <https://www.ipb.uni-bonn.de>