

Localizzatori ottici

I *localizzatori ottici* utilizzati in sala operatoria sono basati su camere ed opportuni software per l'identificazione di oggetti (*marker*) nelle tre dimensioni. Per questo motivo i clinici spesso identificano il localizzatore ottico anche come “la/le telecamere”.

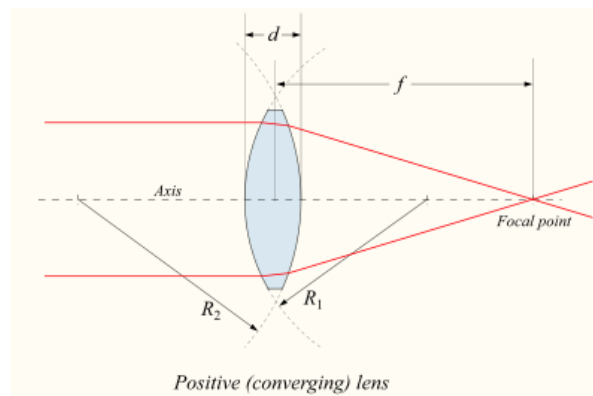
Accenni sull'architettura delle camere

Una camera è costituita da:

- un *sensore*, composto da elementi sensibili alla radiazione luminosa (*pixel*);
- un'*ottica* che convoglia i raggi di luce (nell'approssimazione dell'ottica geometrica) sul sensore.

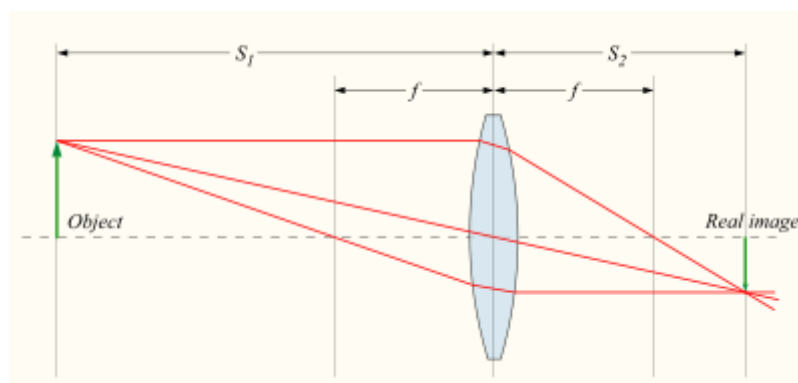
I sensori comunemente più utilizzati nelle camere sono *bidimensionali* (con pixel quadrati e loro disposizione in formato 4:3 o 16:9) e le ottiche sono frequentemente composte da *lenti sferiche* (<https://it.wikipedia.org/wiki/Lente>).

Una lente sferica focalizza un fascio di raggi paralleli all'asse della lente in un *punto focale* f :



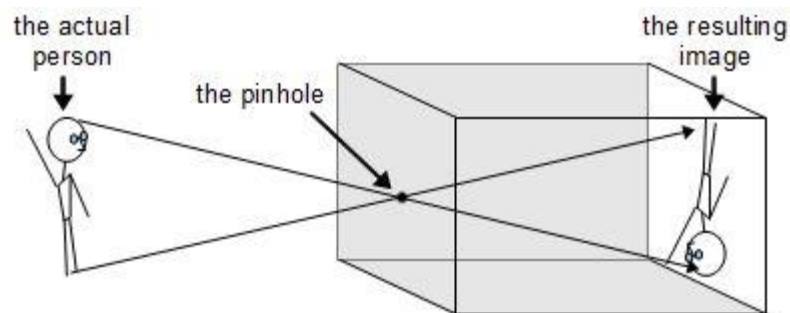
Nell'esempio precedente è rappresentata una lente positiva biconvessa. Nel caso di lenti negative si deve considerare il punto focale anteriore.

Vediamo ora che più raggi emessi o riflessi lungo direzioni diverse dallo stesso punto nella scena (la punta della freccia) attraversano una lente (positiva) con le modalità descritte nella seguente figura:



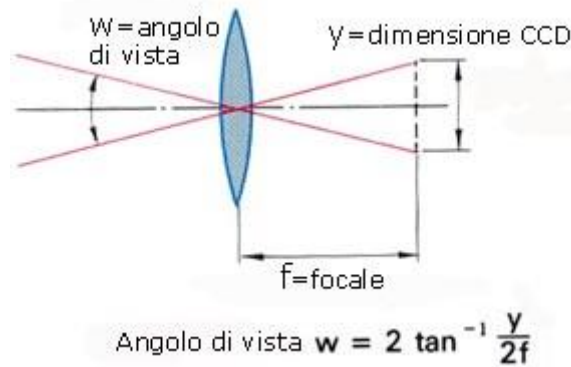
I raggi paralleli sono focalizzati nel punto di *fuoco posteriore* (come si vede nell'immagine ne esiste anche uno *anteriore*). I raggi lungo le direzioni passanti per il *punto nodale* (centro della lente in questo caso) la attraversano senza subire deviazioni. Analizzando queste due direzioni è possibile ricavare a quale distanza S_2 un oggetto posto a distanza S_1 dalla lente sia correttamente focalizzato. Una lente permette di avere immagini a fuoco per oggetti a distanza S_1 quando il suo sensore si trova a distanza S_2 . E' comunque possibile avere a fuoco oggetti in un intorno di S_1 detto *lunghezza di campo*. E' bene evidenziare che i raggi non attraversano la lente solo lungo le due direzioni che solitamente si utilizzano per i suddetti ragionamenti sulla messa a fuoco, ma lungo infinite direzioni. Si è soliti considerare le suddette particolari direzioni poiché permettono analisi e calcoli molto semplici.

Nel caso di camera con sensore alla corretta distanza di fuoco possiamo considerare solo i raggi passanti per il punto nodale della lente e approssimare la camera con un modello detto pinhole. Nel *modello pinhole* della camera si ipotizza di avere un piccolo forellino in corrispondenza del punto nodale della lente attraverso il quale possono passare solo alcuni raggi (riflessi o emessi dagli oggetti presenti nella scena) lungo direzioni passanti per detto punto:



Da notare che nel modello pinhole ideale della camera tutti gli oggetti sono a fuoco, indipendentemente dalla loro distanza rispetto alla camera, mentre in una camera con lenti sferiche gli oggetti possono essere a fuoco o meno. Notare inoltre che l'implementazione di un modello fisico pinhole (con un forellino al posto della lente) comporterebbe bassissima luminosità sul sensore poiché passerebbero, per un dato punto della scena, solo i raggi lungo la direzione passante per il forellino stesso; di conseguenza solo una bassissima quota di energia luminosa raggiungerebbe il sensore (avendo così la necessità di disporre di sensori estremamente sensibili).

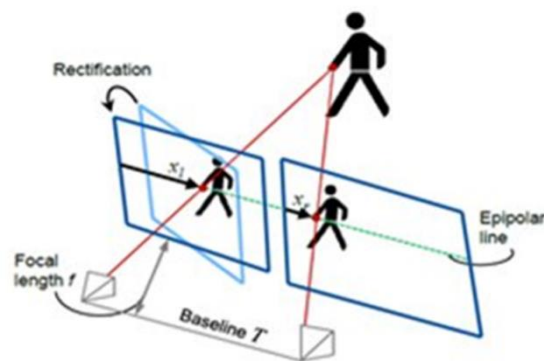
Notare infine che date le dimensioni di un sensore e data la sua distanza dal pinhole, sarà possibile ottenere la proiezione solo per gli oggetti che si trovano all'interno degli angoli di vista della camera (solitamente i sensori sono rettangolari e pertanto l'angolo di vista verticale è diverso da quello orizzontale):



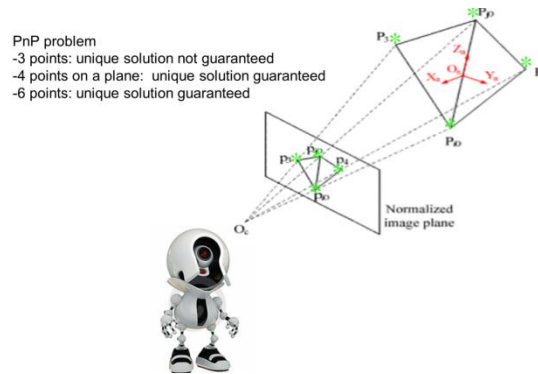
Nel modello pinhole, così come in una camera dotata di lenti sferiche, la proiezione degli oggetti sul sensore è di tipo *proiettivo prospettico* e di conseguenza le dimensioni degli oggetti proiettati sul sensore dipendono dalla loro distanza dalla camera oltre che dalle loro dimensioni reali (e in generale dal loro orientamento rispetto alla telecamera). Esistono anche altre tipologie di ottiche, come quelle *telecentriche* nelle quali la proiezione è di tipo *parallelo* e non prospettica. Tali ottiche trovano impiego ad esempio nelle misurazioni di precisione di parti meccaniche nei processi produttivi (per il controllo qualità) poiché queste neutralizzano gli effetti prospettici dovuti a un posizionamento impreciso degli oggetti su una linea di produzione.

Quando si lavora con un sensore rettangolare (teoricamente anche curvo) e un'ottica proiettiva, dato un punto sul sensore è possibile determinare a quale direzione appartiene il punto nella scena:

[https://en.wikipedia.org/wiki/Triangulation_\(computer_vision\)](https://en.wikipedia.org/wiki/Triangulation_(computer_vision))



Con una sola camera non è possibile determinare la posizione di un punto nello spazio. Con una sola camera è comunque possibile ricavare la posizione e l'orientamento di oggetti nello spazio nei quali siano distinguibili più punti (almeno 4 coplanari oppure almeno 6 nel caso generale).



Nell'esercitazione in aula si vedrà come localizzare "scacchiere" con una sola telecamera, ricorrendo alle funzioni del Computer Vision Toolbox di Matlab.

Esistono anche localizzatori basati su sensori lineari ed ottiche con particolari lenti (cilindriche). Con questo tipo di a sensori lineari si può determinare su quale piano giace un punto nello spazio proiettato sul sensore. In questo caso pertanto servono almeno tre camere per determinare dove giace un punto nello spazio.

Utilizzo di luce nell'infrarosso per ottenere una localizzazione robusta ed accurata.

Nel paragrafo precedente è stato introdotto che tramite camere è possibile localizzare punti o oggetti in una scena. Il punto di partenza per localizzare punti o in generale oggetti nello spazio è la loro identificazione nelle immagini proiettate sul sensore. Esistono molte soluzioni per localizzare oggetti o punti in una scena illuminata con luce tradizionale nello spettro del visibile artificiale (o anche naturale per scenari all'aperto). Le variazioni delle condizioni d'illuminazione, dovute a vari fenomeni come le ombre create dai soggetti e gli oggetti presenti nella scena, possono compromettere la corretta identificazione degli oggetti d'interesse sull'immagine e rendere di conseguenza la loro localizzazione poco robusta e inaccurata. Per questo motivo si può ricorrere all'illuminazione all'infrarosso.

Localizzatori con marker attivi

Una prima classe di localizzatori impiega *marker attivi*, solitamente implementati con led IR (infrarosso). Inserendo di fronte all'ottica un filtro in grado di far passare solo le componenti IR è possibile distinguere in modo molto netto i marker accesi dal resto della scena (nella quale solitamente la componente IR è decisamente inferiore). L'accensione in sequenza nota dei marker permette la loro identificazione e localizzazione multipla. Il limite della tecnologia con marker attivi

è la necessità di alimentare a batteria o con un cavo elettrico ciascun marker.

Localizzatori con marker passivi

La localizzazione con *marker passivi* prevede un illuminatore IR posizionato pressoché coassialmente con le camere e marker costituiti da oggetti riflettenti solitamente sferici. I riflessi generati sugli elementi riflettenti saturano alcuni gruppi di pixel che diventano facilmente distinguibili (segmentabili) nelle immagini:



In questo caso, quando nella scena sono presenti più marker, la loro discriminazione (e conseguente localizzazione) è possibile solo conoscendo vincoli geometrici imposti tra di loro.

Alcuni localizzatori ottici sia attivi sia passivi (che ibridi) più comuni possono essere visionati sul sito della ditta produttrice Northen Digital: <http://www.ndigital.com/>

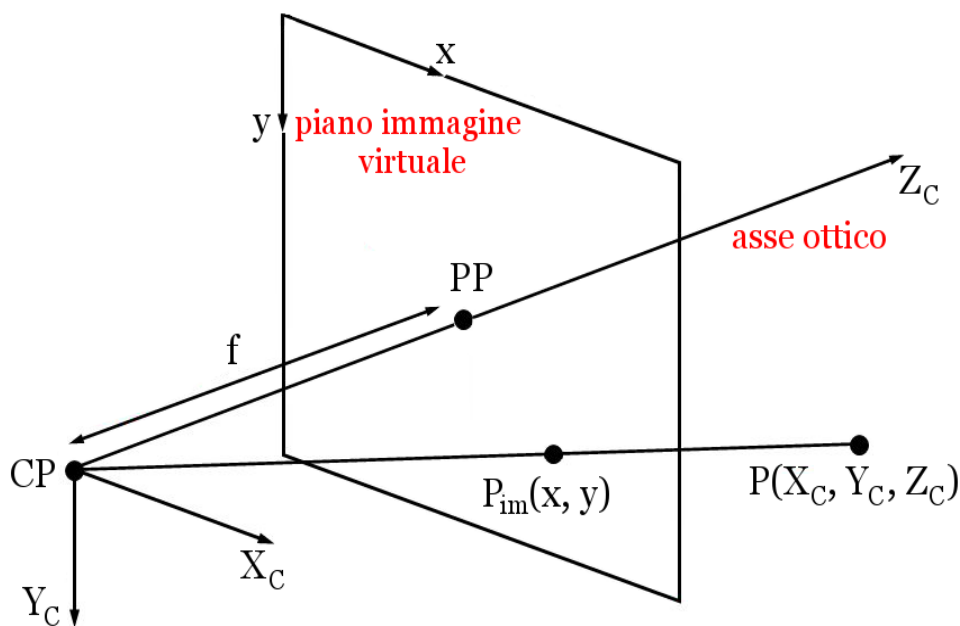
Parametri del modello pinhole di una camera

La localizzazione con telecamere richiede di conoscere in modo accurato come i raggi sono proiettati sul sensore. Tramite un *processo di calibrazione* (che proveremo durante l'esercitazione) è possibile ottenere i parametri di due sotto modelli: il primo che caratterizza come un punto della scena è proiettato sul sensore tenendo conto di un modello pinhole ideale; il secondo che caratterizza la distorsione introdotta dalla lente (che fa sì che i raggi non procedano in modo perfettamente rettilineo attraverso il pinhole).

Quando si utilizza un modello pinhole, i punti delle strutture 3D nella scena inquadrata sono mappate sulla superficie bidimensionale del sensore in base alle regole della proiezione prospettica: si suppone che la luce viaggi lungo un percorso rettilineo attraverso un foro di piccole dimensioni che costituisce il centro di proiezione CP. La luce viene quindi focalizzata sul piano del sensore,

producendo un'immagine capovolta.

Consideriamo adesso il centro di proiezione CP come l'origine del sistema di riferimento della camera. E' possibile considerare l'immagine come se non fosse capovolta scegliendo un piano di proiezione virtuale in posizione frontale anziché posteriore rispetto al centro di proiezione. L'asse Z è scelto perpendicolare al piano immagine e lo interseca in un punto detto punto principale PP. La distanza tra il centro di proiezione e il piano focale è chiamata lunghezza focale f.



E' possibile mettere in relazione le coordinate di un punto nello spazio P con la sua proiezione P_{im} sul piano immagine sfruttando le proprietà dei triangoli simili. Consideriamo per semplicità la proiezione sul piano immagine virtuale, che si trova traslando il piano immagine affinché intersechi l'asse Z in $+f$, piuttosto che $-f$. Dati i sistemi di riferimento in figura, si ricava:

$$\frac{f}{Z_C} = \frac{x}{X_C} = \frac{y}{Y_C}$$

Dove X_C , Y_C e Z_C rappresentano le coordinate nel sistema di riferimento 3D della camera e x ed y le coordinate nel sistema di riferimento del sensore.

Tramite le relazioni ottenute si può esprimere in forma matriciale la trasformazione dal sistema di riferimento della camera a quello in coordinate omogenee¹ dell'immagine:

¹ In geometria proiettiva si utilizzano le coordinate omogenee poiché consentono di rappresentare le trasformazioni in forma matriciale. Date le coordinate di un punto nell'immagine $(x_1, x_2)^T$ per ottenere le sue coordinate omogenee basta assegnare un valore unitario alla terza componente: $(x_1, x_2, 1)^T$. E' inoltre necessario normalizzare i calcoli affinché si abbia l'unità sul terzo componente. A volte si utilizza il simbolo \sim per esplicitare la necessità di questa

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{Z_C} \begin{bmatrix} f & 0 & x_{PP} \\ 0 & f & y_{PP} \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} \sim \begin{bmatrix} f & 0 & x_{PP} \\ 0 & f & y_{PP} \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix}$$

dove f è la lunghezza focale ovvero la distanza tra il piano immagine e il centro di proiezione. Le variabili x_{PP} e y_{PP} sono state introdotte poiché l'origine del sistema di riferimento sull'immagine è in genere decentrata rispetto al sensore. Esse rappresentano le coordinate del punto principale nel sistema di riferimento solidale al piano immagine e idealmente corrispondono alla metà delle dimensioni del sensore:

$$x = \frac{f \cdot X_C}{Z_C} + x_{PP}$$

$$y = \frac{f \cdot Y_C}{Z_C} + y_{PP}$$

Nella pratica, a causa di piccoli disallineamenti, il centro geometrico del sensore non coincide mai perfettamente con il punto principale ed è possibile risalire al valore esatto di questi parametri solo tramite il processo di calibrazione.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim \begin{bmatrix} f/s_x & 0 & x_{PP}/s_x \\ 0 & f/s_y & y_{PP}/s_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = \begin{bmatrix} p_x & 0 & c_x \\ 0 & p_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix}$$

Poiché stiamo trattando un'immagine digitale, le unità di misura in gioco sono i pixel. Scaliamone quindi i valori in funzione della dimensione dei pixel lungo le due direzioni s_x ed s_y :

Generalmente, da progetto, le dimensioni dei pixel lungo le due direzioni s_x ed s_y sono uguali, ma non sempre i pixel sono quadrati. I parametri $p_x = f/s_x$ e $p_y = f/s_y$ costituiscono la lunghezza focale della camera nella direzione x e y e si misurano in pixel. I parametri $c_x = x_{PP}/s_x$ e $c_y = y_{PP}/s_y$ corrispondono alle coordinate del punto principale, misurate anch'esse in pixel.

E' inoltre possibile che a seguito del processo di digitalizzazione, i pixel risultino inclinati. Quest'ultima eventualità è valutata attraverso il parametro skew definito come:

$$k = f_y \cdot \tan \alpha$$

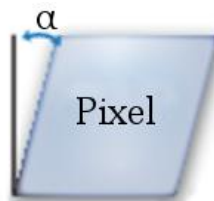


Illustrazione del concetto di skew. L'angolo α definisce di quanto il pixel si allontana dalla forma quadrata.

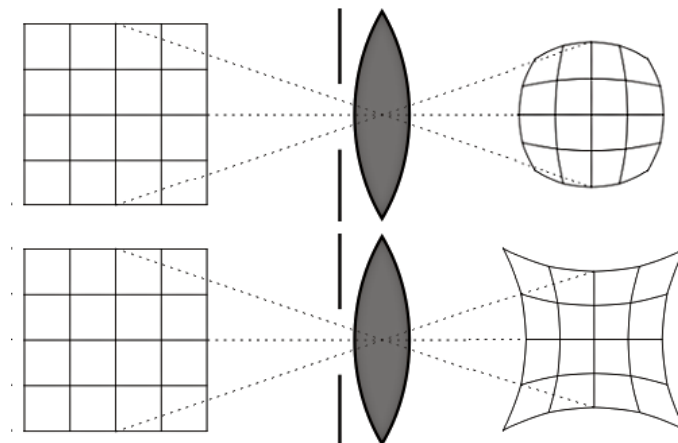
normalizzazione che nella pratica spesso viene omissa (il simbolo ma non la normalizzazione!).

Inserendo nell'analisi anche quest'ultimo parametro, si ottiene la matrice che rappresenta la trasformazione proiettiva dal sistema di riferimento tridimensionale della camera a quello 2-D dell'immagine in pixel:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim \begin{bmatrix} p_x & k & c_x \\ 0 & p_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix}$$

La trattazione ha tenuto conto della dimensione reale dei pixel e della distanza reale tra il sensore ed il centro di proiezione. Fino ad adesso l'unica alterazione rispetto alla realtà è stata considerare un piano dell'immagine virtuale anteriore anziché posteriore (com'è nella realtà). E' poi possibile considerare piani virtuali con distanza diversa rispetto a quella reale dal centro di proiezione. Questo spostamento si traduce in una scalatura uniforme lungo tutte le direzioni. In pratica, l'informazione proveniente dal mondo reale proiettata sui pixel appartenenti a piani immagine virtuali posti a distanza diversa non cambia se considero pixel scalati in funzione della variazione di detta distanza.

Il modello pinhole, pur consentendo una semplice formulazione matematica della relazione tra l'oggetto e la sua proiezione nell'immagine, trascura completamente il contributo delle lenti. Le lenti utilizzate nelle camere reali introducono delle distorsioni: a causa dei vincoli cui è soggetto il processo di fabbricazione è possibile che linee dritte della scena vengano mappate su linee curve del piano immagine. Ciascuna lente è generalmente radialmente simmetrica e per questo la distorsione risulta essere prevalentemente radiale (ma non solo). Questo effetto può presentarsi sotto forma di deformazione a barile quando l'immagine si gonfia procedendo verso l'interno, oppure come deformazione a cuscino se l'immagine si restringe nel centro:



Fortunatamente questa deformazione può essere compensata conoscendo la deformazione che si ottiene sul piano immagine e che risulta essere indipendente dalla scena inquadrata.

E' possibile correggere la distorsione radiale allontanando o avvicinando i punti dell'immagine in

$$\begin{aligned} u_{compensata} &= u(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ v_{compensata} &= v(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{aligned}$$

direzione radiale rispetto al centro della distorsione. Esistono vari modelli della distorsione e di conseguenza della loro compensazione (undistorsion in inglese). Ad esempio, indicando con r la distanza tra un punto dell'immagine e il centro della distorsione (che coincide col centro di proiezione), è possibile esprimere la funzione di correzione nella forma polinomiale:

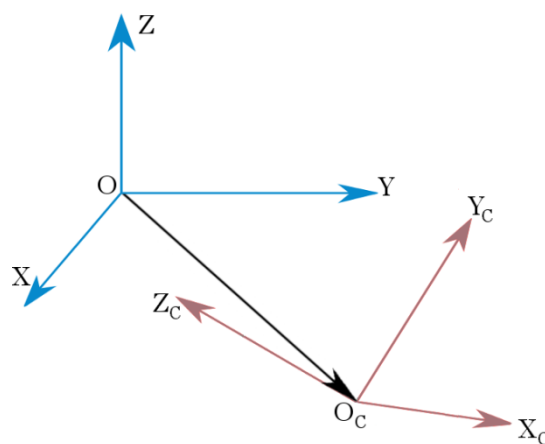
Il primo coefficiente è unitario cosicché l'immagine originale e quella corretta abbiano la stessa scala in corrispondenza del centro di distorsione. Gli altri coefficienti possono essere stimati misurando la curvatura di linee dell'immagine che sono dritte nella scena reale (solitamente in automatico con il processo di calibrazione della camera).

I parametri fisici di una camera sono comunemente divisi in due gruppi.

Quelli intrinseci, appena discussi, descrivono parametri interni della camera. I parametri intrinseci sono quindi l'insieme dei parametri del modello lineare pinhole e di quello non lineare per la compensazione della distorsione radiale.

I parametri estrinseci identificano in modo univoco la trasformazione rigida tra il sistema di riferimento della camera e quello solidale alla scena inquadrata. Nel caso di sistemi composti da più telecamere, i parametri estrinseci ne descrivono inoltre le relazioni reciproche. Tipicamente determinare questi parametri significa trovare:

- Il vettore di traslazione che porta a far coincidere le origini dei due sistemi di riferimento.
- La matrice di rotazione che allinea gli assi corrispondenti nei due sistemi di riferimento.



Si ricorda che, sebbene la matrice di rotazione consista di nove elementi, questa possiede solo tre gradi libertà. Il vettore di traslazione contribuisce con altri tre parametri per cui si ottiene un totale di sei parametri estrinseci, tramite i quali è possibile determinare le coordinate di un punto nel sistema di riferimento della telecamera a partire dalle coordinate nel sistema di riferimento noto:

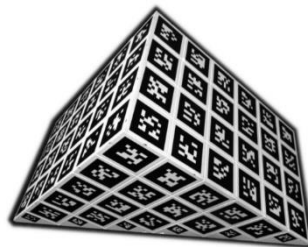
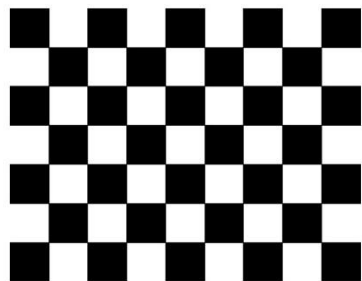
$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix}$$

Complessivamente il passaggio dal sistema di riferimento noto a quello dell'immagine può essere descritto ricorrendo alle coordinate omogenee²:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} p_x & k & c_x \\ 0 & p_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} m_{11} & m_{12} & m_{13} & X_o \\ m_{21} & m_{22} & m_{23} & Y_o \\ m_{31} & m_{32} & m_{33} & Z_o \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix}$$

Per effettuare la stima dei parametri intrinseci ed estrinseci è necessario mettere in relazione i punti della scena con i loro corrispondenti nell'immagine. Queste relazioni si ottengono tramite l'utilizzo di un oggetto di calibrazione noto. E' possibile classificare le tecniche di calibrazione in base alle dimensioni dell'oggetto utilizzato:

- Calibrazione tramite oggetti 3-D. La telecamera viene calibrata visualizzando un oggetto la cui geometria è nota con precisione: gli oggetti sono in genere costituiti da due o tre piani ortogonali.
- Calibrazione tramite pattern 2-D. Le tecniche di questa categoria richiedono l'utilizzo di più immagini relative a diversi orientamenti del pattern bidimensionale.
- Calibrazione tramite oggetti 1-D (a). L'oggetto è costituito da un insieme di punti collineari. La calibrazione viene effettuata tramite immagini dell'oggetto in rotazione attorno a un punto fisso.
- Auto-calibrazione. In questo caso non viene utilizzato alcun oggetto di calibrazione: i parametri vengono stimati a partire da più immagini di una scena statica non strutturata.



² Date le coordinate di un punto nello spazio tridimensionale, $(x_1, x_2, x_3)^T$, le sue coordinate omogenee si ottengono introducendo la quarta dimensione: $(x_1, x_2, x_3, 1)^T$. Rimane valida la proprietà che 2 vettori espressi in coordinate omogenee sono equivalenti quando l'uno si ottiene moltiplicando l'altro per uno scalare.

In generale la calibrazione viene effettuata con maggior precisione utilizzando oggetti tridimensionali. Tuttavia spesso si prediligono pattern 2-D poiché offrono una buona precisione nella calibrazione e possono essere prodotti ed utilizzati con facilità. Le tecniche che sfruttano oggetti “monodimensionali”, come piccole sfere, risultano particolarmente adatte quando è richiesto di individuare la posizione relativa di più telecamere oltre ai loro parametri interni. I punti dell’oggetto risultano infatti visualizzabili da molteplici angolazioni e costituiscono un buon sistema di calibrazione anche nel caso di telecamere montate l’una di fronte all’altra. Le procedure di auto-calibrazione non permettono di ottenere livelli di accuratezza comparabili con le altre categorie poiché necessitano la stima di un gran numero di parametri producendo un problema matematico più complesso.