

Sommario

- Introduzione
- **Percezione**
- Formazione delle immagini
- Elaborazione delle immagini a basso livello
- Estrazione di informazione 3D da un'immagine
- Riconoscimento di oggetti
- Manipolazione e navigazione
- Conclusioni

Percezione

- Acquisizione da parte dell'*agente* di informazioni relative al mondo che lo circonda
- *Sensori*: registrano particolari caratteristiche dell'ambiente e le danno in input al programma agente
- Percezione *visiva*
- Gli stimoli percettivi possono essere impiegati secondo due differenti approcci
 - Estrazione di caratteristiche (l'agente rileva e usa poche caratteristiche presenti nell'input percettivo)
 - Approccio basato sul modello (l'agente ricostruisce un modello del mondo dagli stimoli percettivi)

Percezione

- Approccio basato sul modello

$$S = f(W)$$

con f funzione che lega lo stato del mondo W allo stimolo S

- La funzione f è compresa bene dalla Fisica/Ottica
- *Grafica Computerizzata*: ottenere S da f e W
- *Visione Artificiale*: ottenere W da f e S

$$W = f^{-1}(S)$$

Problema complesso: la funzione f non ha una vera e propria funzione inversa!

Ambiguità \Rightarrow serve informazione aggiuntiva!

Problema troppo oneroso da un punto di vista computazionale

Percezione

- Approcci migliori
 - si costruisce una distribuzione di probabilità sui mondi, anziché cercare di trovarne uno singolo
$$P(W) = P(W/S) P(S)$$
 - estrarre solo l'informazione necessaria per il particolare task
 - navigazione
 - manipolazione
 - riconoscimento / identificazione
 - ...

Sommario

- Introduzione
- Percezione
- **Formazione delle immagini**
- Elaborazione delle immagini a basso livello
- Estrazione di informazione 3D da un'immagine
- Riconoscimento di oggetti
- Manipolazione e navigazione
- Conclusioni

Formazione delle immagini

- La risoluzione di molte problematiche di Computer Vision parte dall'analisi del processo di *formazione dell'immagine* di una scena
- Alcune tecniche di *calibrazione e ricostruzione* seguono una sorta di *reverse engineering* di questo processo
- Analizziamo come si forma l'immagine di una scena

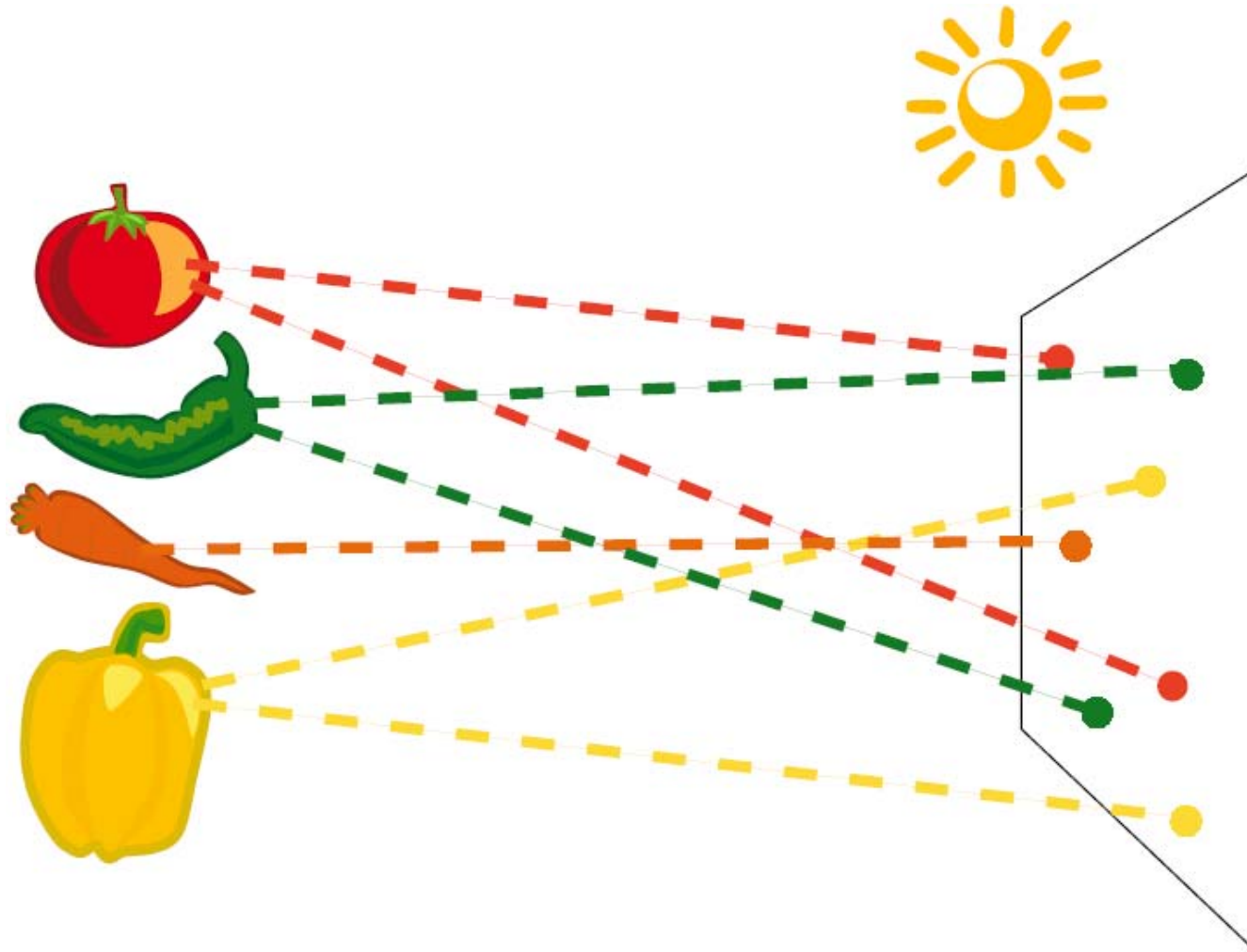
Formazione delle immagini

- Principio della visione: gli oggetti che compongono una *scena 3D* riflettono la luce, creando una *immagine 2D* su un piano di immagine, ricoperto da materiale fotosensibile
 - rodopsina nella retina
 - sali d'argento su una pellicola fotografica
 - sensore *CCD* in una fotocamera digitale
- Per definire la funzione $f(W)$ occorre studiare
 - geometria (cioè il modo in cui la luce si riflette rimbalzando sugli oggetti del *mondo* fino a un punto sul *piano d'immagine*)
 - fotometria (cioè il modo in cui la luce nella scena determina la luminosità dei punti sull'immagine)
- La scena è grande, il piano d'immagine piccolo: occorre focalizzare la luce su di esso (con o senza una lente)

Formazione delle immagini

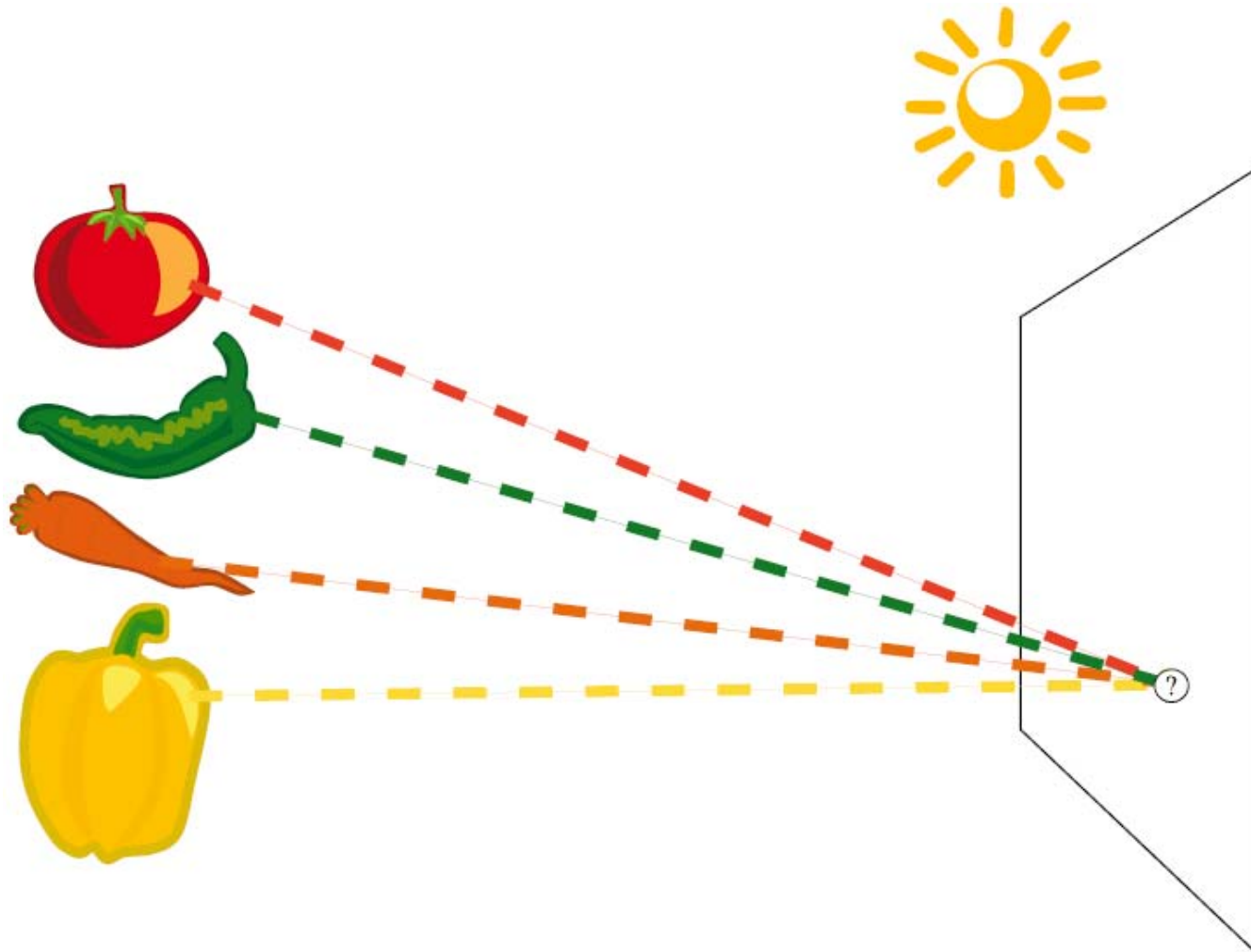
- Abbiamo una o più fonti di luce, degli oggetti non completamente trasparenti e una superficie sensibile ai raggi di luce (pellicola, sensore digitale, retina, ...)
- I raggi di luce vengono rifratti dagli oggetti, presenti nella scena, in modo “caotico”

Formazione delle immagini



Ogni punto della scena “influisce” sui diversi punti del piano immagine ...

Formazione delle immagini

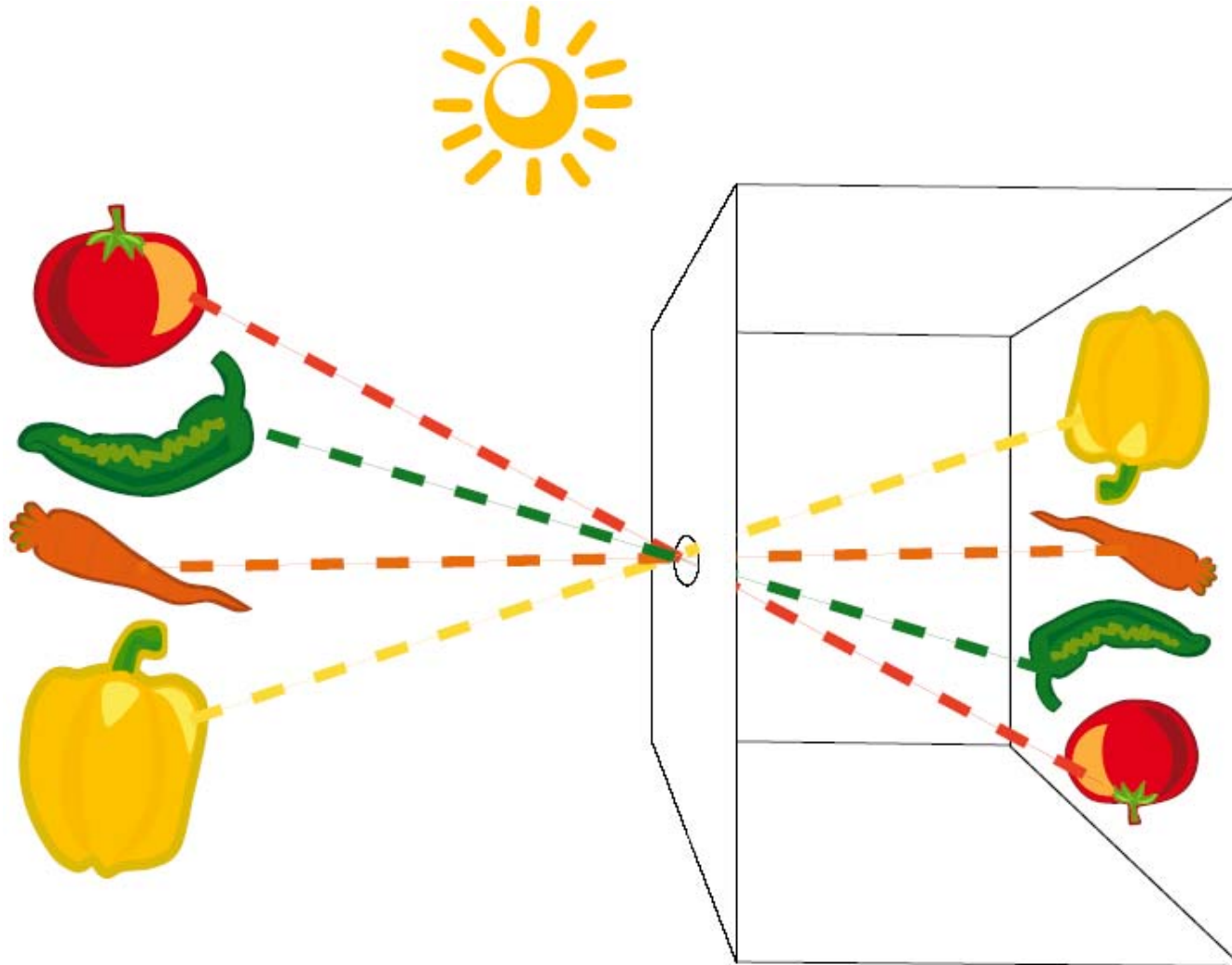


... e ogni punto del piano immagine è colpito da raggi provenienti da punti della scena differenti

Formazione delle immagini

- Vogliamo fare in modo che ogni punto della scena “influisca” su di un solo punto del piano immagine
- Una possibilità è quella di costringere tutti i raggi a passare per un foro molto piccolo

Formazione delle immagini

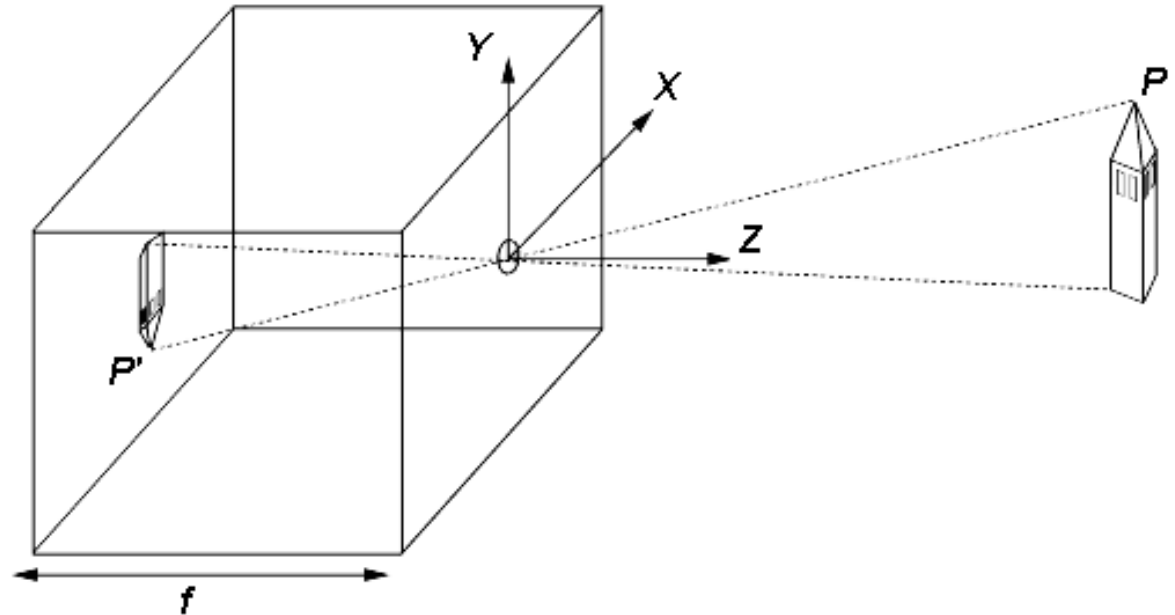


La “scatola” entro cui si forma l’immagine è detta *camera pinhole* (*stemoscopio*)

Formazione delle immagini

Immagine senza lenti:
la scatola nera

f : distanza del forellino O
dal piano d'immagine



P : punto della scena, di coordinate (X, Y, Z) in un sistema di coordinate tridimensionali con l'origine in O (forellino)

P' : immagine di P sul piano d'immagine, di coordinate (x, y) nello stesso sistema di coordinate tridimensionali con l'origine in O (forellino)

equazioni fondamentali di proiezione prospettica

$$\frac{-x}{f} = \frac{X}{Z}, \frac{-y}{f} = \frac{Y}{Z} \Rightarrow x = \frac{-fX}{Z}, y = \frac{-fY}{Z}$$

N.B. (pointing to the minus sign in the equation)
N.B. (pointing to the Z in the denominator)

Formazione delle immagini

Nella *Proiezione Prospettica*, le linee parallele convergono in un punto sull'orizzonte.

Dim: nella scena, una linea che attraversa il punto (X_0, Y_0, Z_0) in direzione (U, V, W) può essere definita come il seguente insieme di punti

$$X_0 + \lambda U, Y_0 + \lambda V, Z_0 + \lambda W \quad \text{con } \lambda \in (-\infty, +\infty)$$

la proiezione di un punto P_λ da questa linea sul piano d'immagine è data da

$$\left(f \frac{X_0 + \lambda U}{Z_0 + \lambda W}, f \frac{Y_0 + \lambda V}{Z_0 + \lambda W} \right) \quad \text{ottenuta sostituendo le coordinate sopra nelle Equazioni della Proiezione Prospettica}$$

per $\lambda \rightarrow +\infty$ o $\lambda \rightarrow -\infty$ si ottiene

$$p_\infty = \left(f U/W, f V/W \right) \quad \text{se } W \neq 0$$

che prende il nome di *punto di fuga* associato alla famiglia di rette con direzione $(U, V, W) \implies$ le linee che hanno la stessa direzione condividono il punto di fuga (c.d.d)

Formazione delle immagini

Se l'oggetto è piccolo rispetto alla sua distanza dalla scatola nera è possibile approssimare la *Proiezione Prospettica* con una *Proiezione Ortografica Scalata*.

Cioè se la profondità Z di un oggetto varia in un intervallo

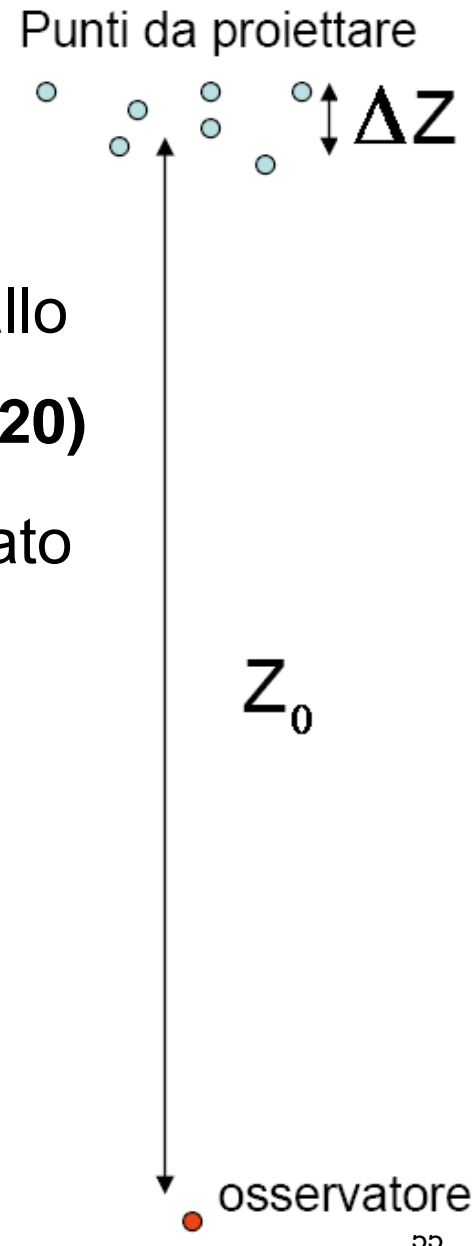
$$Z_0 \pm \Delta Z, \text{ con } \Delta Z \ll Z_0 \text{ (in pratica: } \Delta Z \leq Z_0/20)$$

il *fattore prospettico di scala* f/Z può essere approssimato da una costante $s=f/Z_0$

Le equazioni per la proiezione delle coordinate della scena (X, Y, Z) sul piano d'immagine diventano quindi

$$x = sX \text{ e } y = sY$$

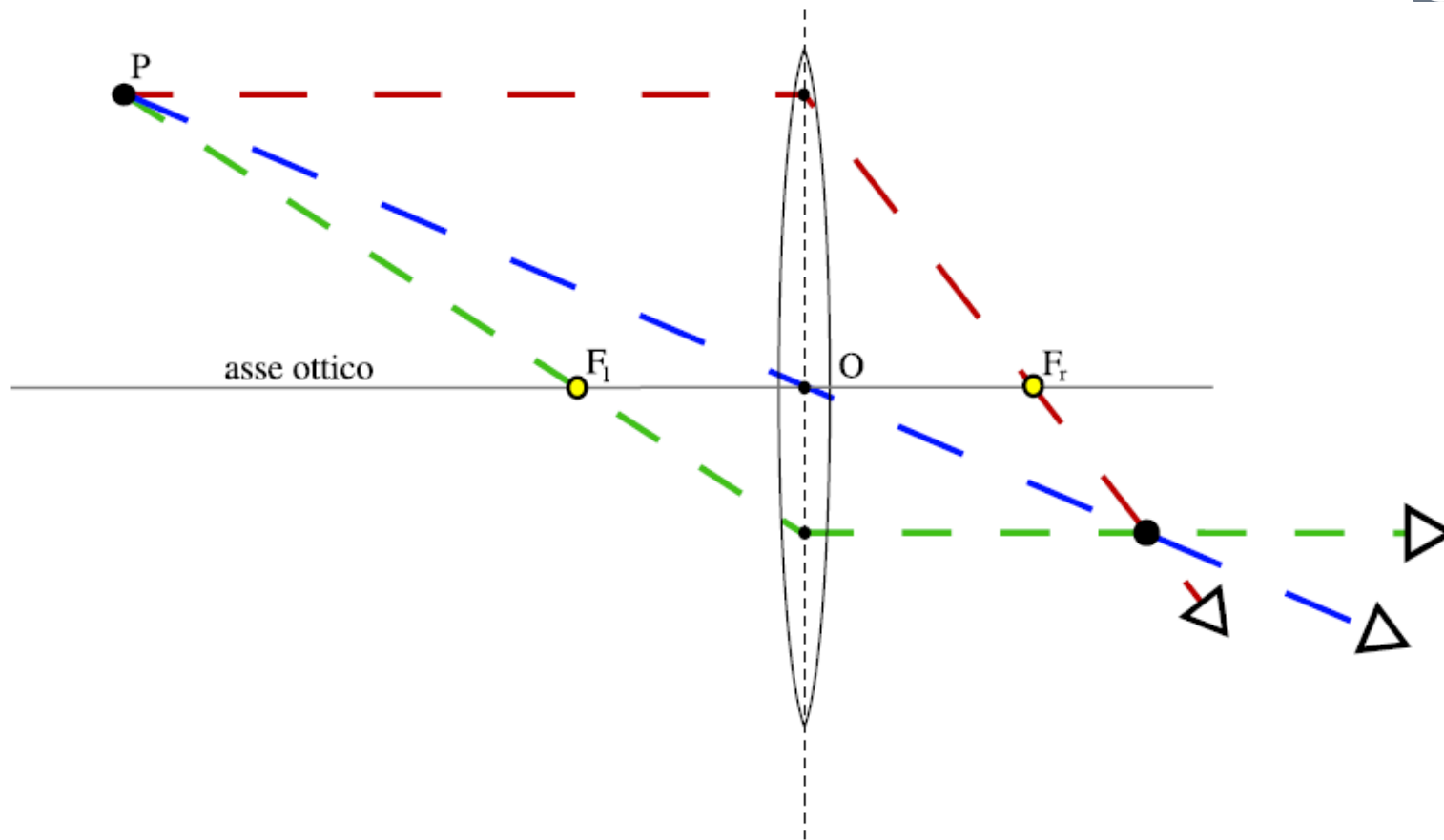
N.B. tale approssimazione vale solo per le parti della scena che variano poco in profondità e non per l'immagine "in grande"!



Formazione delle immagini

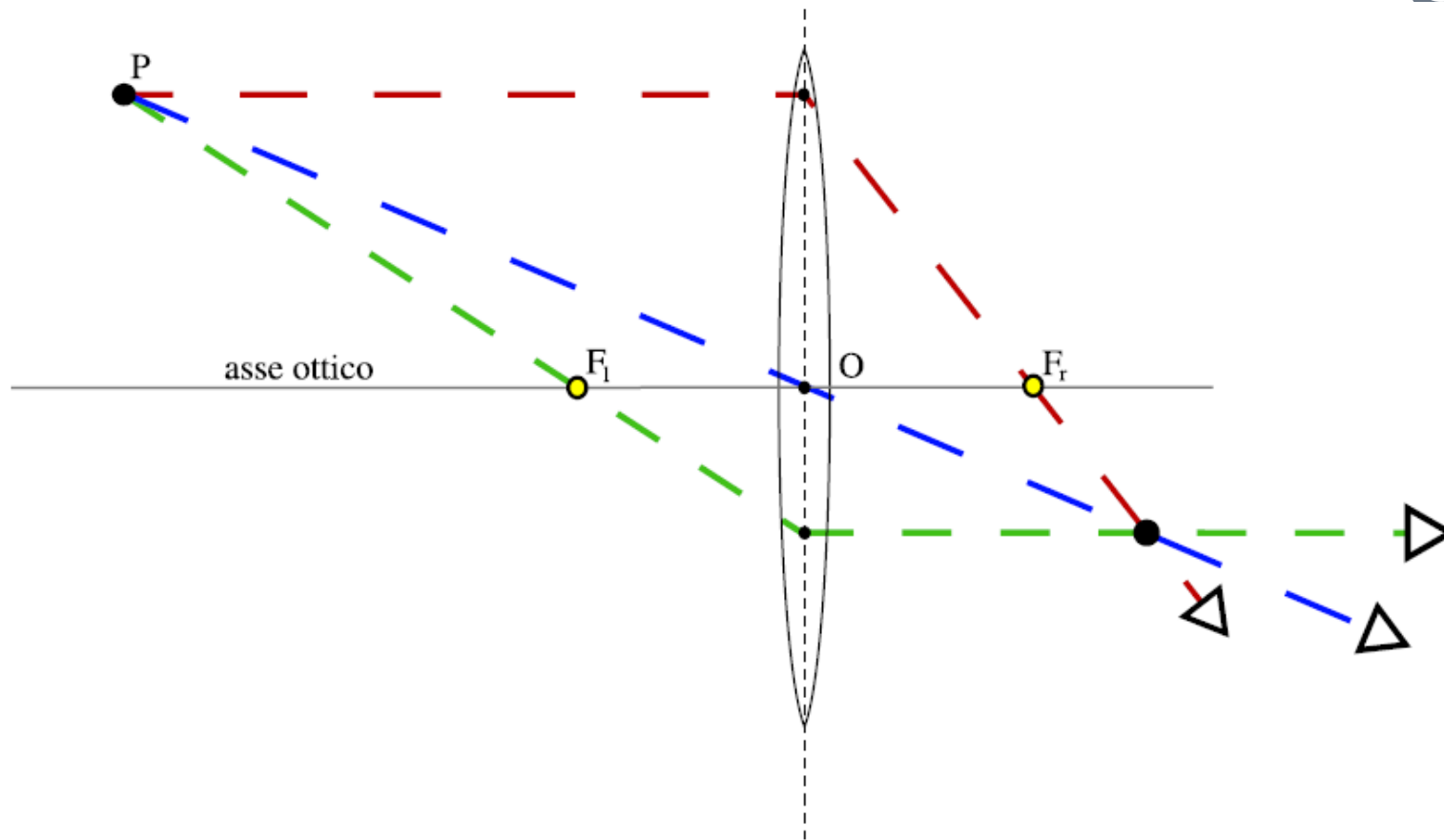
- La *camera pinhole* richiede una superficie con un range di sensibilità assai elevato
- Non è pratica da utilizzare (impossibile *zoomare*)
- Soluzione migliore: le *lenti sottili*
- Le *lenti sottili* sono un dispositivo ottico più complesso e flessibile per mettere a fuoco l'immagine di una scena
- Possiamo immaginarle come un sottile disco di vetro di una materiale trasparente in cui sono definiti un *asse ottico* e due *fuochi*, ovvero sia due punti particolari dell'asse ottico, esterni alla lente stessa

Formazione delle immagini



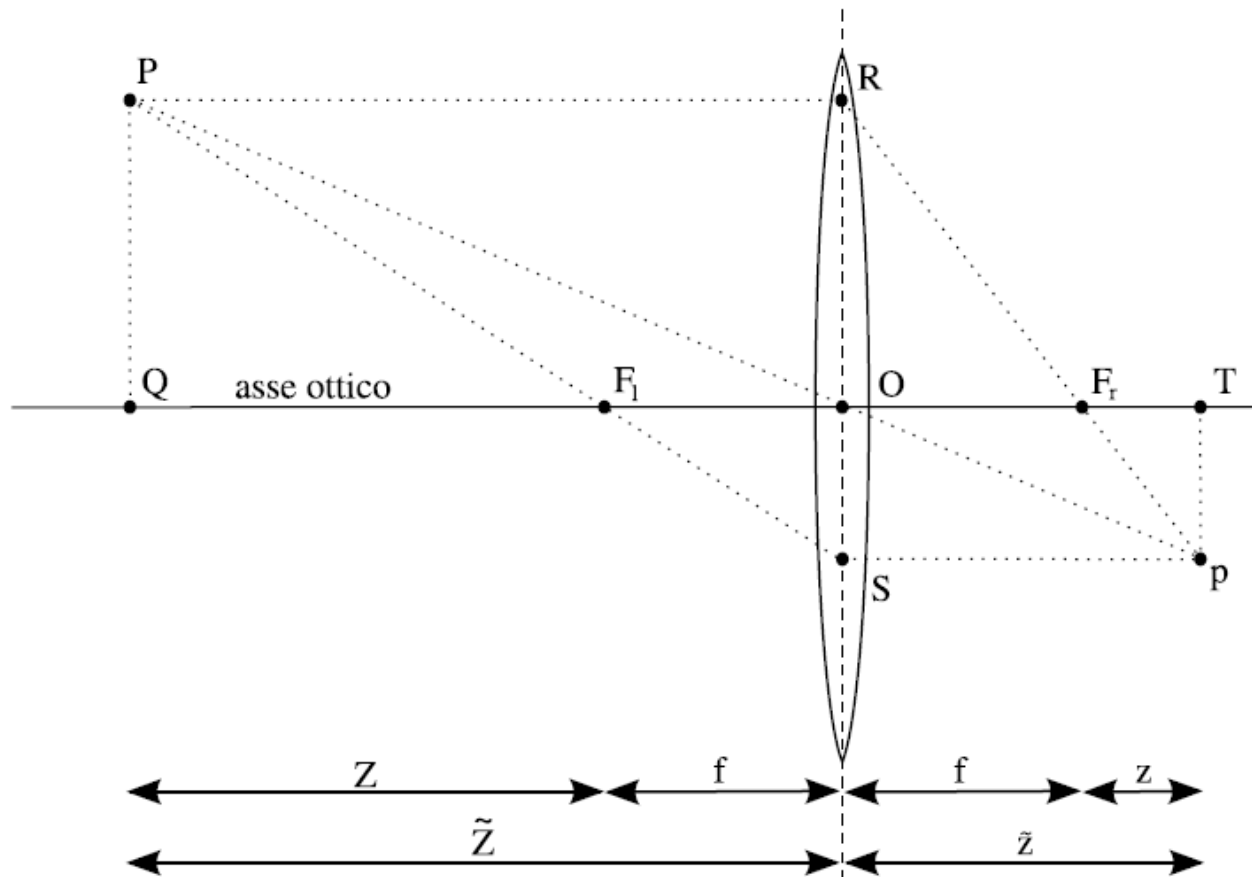
- Le lenti deviano i raggi di luce seguendo due regole
- ogni **raggio di luce** che entra da un lato della lente parallelamente all'asse ottico viene deviato verso il fuoco che si trova dall'altro lato

Formazione delle immagini



- Le lenti deviano i raggi di luce seguendo due regole
- ogni **raggio di luce** che entra da un lato della lente passando per il fuoco, esce dall'altro lato parallelamente all'asse ottico

Formazione delle immagini



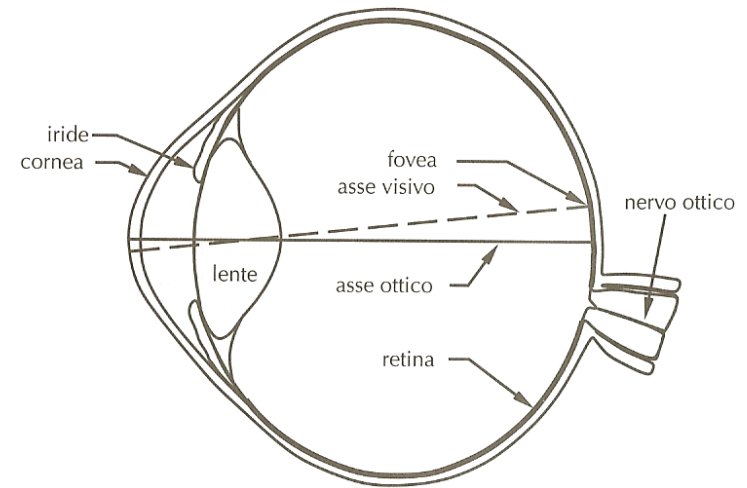
Tramite considerazioni geometriche sulle coppie di triangoli PQF_1 , SOF_1 e ROF_2 , TpF_2 , otteniamo la relazione

Equazione fondamentale delle lenti sottili

$$\frac{1}{\tilde{Z}} + \frac{1}{\tilde{z}} = \frac{1}{f}$$

Formazione delle immagini

Nell'occhio umano, il sistema ottico è il *cristallino* e il piano d'immagine la *retina*
In una fotocamera, il sistema ottico è la *lente*, il piano d'immagine la *pellicola* o il *sensore digitale*



L'immagine di un oggetto che si trova nella scena a distanza \tilde{Z} è prodotta a una distanza fissa dalla lente \tilde{z} e la relazione fra \tilde{Z} e \tilde{z} è data da

Equazione fondamentale delle lenti sottili

$$\frac{1}{\tilde{Z}} + \frac{1}{\tilde{z}} = \frac{1}{f}$$

\tilde{Z} : distanza dell'oggetto
 \tilde{z} : distanza d'immagine
 f : lunghezza focale della lente

Dato che solitamente \tilde{Z} è molto maggiore di \tilde{z} o di f , vale l'approssimazione

$$\frac{1}{\tilde{Z}} + \frac{1}{\tilde{z}} \approx \frac{1}{\tilde{z}} \Rightarrow \frac{1}{\tilde{z}} \approx \frac{1}{f} \Rightarrow \tilde{z} \approx f$$

Formazione delle immagini

- L'*equazione fondamentale delle lenti sottili* ha una conseguenza importante: affinché l'immagine sia a fuoco, a parità di *lunghezza focale* (i.e., distanza del fuoco dal centro) i punti devono trovarsi alla stessa distanza dalla lente
- In altri termini, una lente è in grado di mettere a fuoco solo una sezione della scena parallela al piano immagine

Richiami di Algebra

- Una **Matrice di Rotazione** in uno spazio a n dimensioni è una matrice R quadrata, di ordine n , che gode di determinate proprietà, tra cui
 - $\det(R)=1$
 - $R^T \Rightarrow$ rotazione inversa

Dalle proprietà precedenti segue che R deve essere *ortonormale*, ossia deve valere $RR^T=R^TR=I_n$

Tale vincolo garantisce che, se applichiamo ad un vettore V la rotazione R , applicando la rotazione inversa otteniamo nuovamente il vettore originario.

Dunque per una matrice *ortonormale*, la trasposta coincide con l'inversa.

Inoltre segue dalla definizione che se R è *ortonormale*, le sue colonne (o righe) sono vettori *unitari mutuamente ortogonali* (ossia *ortonormali*).

Richiami di Algebra

- Ci chiediamo quanti siano i *gradi di libertà* di una *matrice di rotazione* nello spazio a $n=3$ dimensioni

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$$

- Gli elementi della matrice sono $n \cdot n = 9$, ma essi non corrispondono ad altrettanti *gradi di libertà*: infatti, dalla condizione di *ortonormalità* di cui sopra, segue che
 - il prodotto scalare di ogni riga per se stessa deve essere 1 (3 vincoli)
 - il prodotto scalare di ogni riga per le altre deve essere 0 (3 vincoli)
- Aggiunti sei *vincoli*, rimangono tre *gradi di libertà*
- Intuitivamente, questi tre *gradi di libertà* corrispondono agli angoli di rotazione sui tre assi

Richiami di Algebra

- Le rotazioni su assi differenti non godono della proprietà commutativa \Rightarrow applicare a un vettore lo stesso insieme di rotazioni, ma in ordine diverso, cambia il risultato.

Conviene, pertanto, esprimere le rotazioni sui tre assi tramite un'unica operazione matriciale, piuttosto che memorizzare separatamente i tre angoli di rotazione

- Per convenzione, premultiplicare per R equivale ad applicare una rotazione antioraria rispetto al senso dell'asse di rotazione; postmultiplicare per R o premultiplicare per R^T equivale ad applicare la rotazione inversa

Richiami di Algebra

- La ***Singular Value Decomposition (SVD)*** è una importante *fattorizzazione* per matrici a valori reali o complessi
- Ogni matrice $M \in \mathbb{C}^{m \times n}$ può essere fattorizzata in
$$M = U \Sigma V^*$$
dove
 - U è una matrice $m \times m$ unitaria (cioè $U U^T = I_m$)
 - Σ è una matrice $m \times n$ diagonale con soli elementi reali non negativi
 - V^* è la coniugata trasposta della matrice $n \times n$ unitaria V
- Gli elementi della diagonale di Σ sono detti *valori singolari* di M

Richiami di Algebra

- SVD e *sistemi omogenei*: sia $\mathbf{Ax}=\mathbf{0}$ è un sistema omogeneo con *rango* inferiore di una unità al numero di incognite (ammette dunque soluzioni oltre a quella banale); una *base* dello spazio delle soluzioni è data proprio dalla colonna di V^* corrispondente all'unico valore singolare nullo in Σ ; ogni altra soluzione è legata a questa da un *fattore di proporzionalità* reale
- Significato geometrico: decomporre una Matrice di Trasformazione M tramite SVD può essere visto come una separazione tra lo *scaling* e le altre operazioni effettuate dalla matrice originaria. In particolare
 - U rappresenta una *trasformazione affine* nello spazio C^m
 - Σ corrisponde al ridimensionamento che M effettua sulle varie dimensioni
 - V^* rappresenta una rotazione in C^n

Richiami di Algebra

- La **Geometria Proiettiva (GP)** modella i concetti di *prospettiva e orizzonte*
- Si occupa delle proprietà geometriche *invarianti* alle trasformazioni proiettive
- Lo spazio considerato è quello *euclideo* con l'aggiunta dei *punti all'infinito*, che insieme formano la *retta proiettiva* e il *piano proiettivo*
- Per descrivere i punti nella *GP* sono utilizzate le **Coordinate Omogenee**: per individuare un punto in uno spazio a n dimensioni si utilizzano $n+1$ valori: un punto di una retta è individuato da due coordinate, un punto sul piano da tre, un punto nello spazio da quattro,

...

Richiami di Algebra

- La *coordinata aggiuntiva* può essere immaginata come un *fattore di scala* che definisce una *classe di equivalenza* tra i punti
 - se uguale a zero, si tratta di un punto all'infinito
 - se diversa da zero, possiamo tradurre le coordinate omogenee in cartesiane dividendo ogni coordinata per il *fattore di scala* (e.g., se $P_h=(x_h, y_h, t_h)$ è un punto in coordinate omogenee e $P_c=(x_c, y_c)$ l'equivalente cartesiano, valgono le relazioni $x_c=x_h/t_h$, $y_c=y_h/t_h$, con $t_h \neq 0$)
- Per verificare se due punti in coordinate omogenee $P_1=(x_1, y_1, t_1)$ e $P_2=(x_2, y_2, t_2)$ appartengono alla stessa classe di equivalenza (i.e., hanno lo stesso equivalente cartesiano), occorre verificare se sono l'uno multiplo dell'altro

Richiami di Algebra

- In termini formali, occorre verificare se vale una qualsiasi delle seguenti condizioni

- $\exists \lambda: P_1 = \lambda P_2$

- $x_1/x_2 = y_1/y_2 = t_1/t_2$

- $x_1/t_1 = x_2/t_2 \wedge y_1/t_1 = y_2/t_2$

In altri termini, i due vettori non devono essere *linearmente indipendenti*

- Viceversa, per esprimere un punto *euclideo* in *coordinate omogenee*, basta aggiungere alla n -pla delle coordinate un valore reale diverso da zero

Richiami di Algebra

- Vantaggio delle coordinate omogenee: un punto cartesiano in 2D può essere moltiplicato solo per una matrice $2 \times 2 \Rightarrow$ non è possibile includere nell'operazione anche una traslazione; se invece esprimiamo lo stesso punto in coordinate omogenee \Rightarrow è possibile esprimere in una sola operazione matriciale una trasformazione più complessa (e.g., una *rototraslazione* o una più generica *trasformazione affine* (i.e., proiettività che trasforma punti effettivi in punti effettivi e punti ideali in punti ideali))
- Per tale motivo le *coordinate omogenee* sono largamente utilizzate, oltre che nella Computer Vision, anche nella Computer Grafica

Richiami di Algebra

- *Trasformazioni nello Spazio Proiettivo*: in tale spazio, le principali trasformazioni geometriche (e.g., rotazione, traslazione, ...) sono invariate rispetto alle corrispondenti nello *spazio euclideo*
- Tuttavia, utilizzare $n+1$ coordinate ci consente di gestire una *rototraslazione* con un'unica matrice
- Se R è una *matrice di rotazione* nello spazio 3D e T è il *vettore di traslazione* sui tre assi, la *rototraslazione* di punti in coord. cartesiane è espressa come $P_r = R(P-T)$.
N.B. si è eseguita prima la *traslazione* e poi la *rotazione*; volendo invertire l'ordine delle due operazioni, occorre traslare con un vettore differente $T' = -RT$ (il segno è invertito per comodità): $P_r = R(P-T) = RP - RT = RP + T'$

Richiami di Algebra

- Se esprimiamo P in coordinate omogenee, ponendo $P_h = (x, y, z, 1)$ la *rototraslazione* può essere espressa tramite la matrice A (4x4)

$$A = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t'_x \\ r_{21} & r_{22} & r_{23} & t'_y \\ r_{31} & r_{32} & r_{33} & t'_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- Si può verificare che moltiplicare A per P_h equivale a eseguire la *traslazione dopo* la *rotazione*; perciò nella matrice compaiono i valori di traslazione di T' e non di T
- Nell'esempio il *vettore* T è sottratto; talvolta può essere invece sommato (a seconda dell'applicazione, della definizione e della trasformazione (se $W \rightarrow C$ o viceversa)). Poiché abbiamo invertito il segno nella definizione di $T' \Rightarrow$ non occorre negare i valori di traslazione della matrice

Formazione delle immagini

- La telecamera non è il centro del mondo: non abbiamo più un unico sistema di riferimento!
- Le coordinate dei punti sono state finora considerate in un sistema di riferimento solidale con la camera nei suoi movimenti, detto sistema di riferimento della camera stessa (*camera reference*)
- In generale, i punti sono forniti in un sistema di riferimento “assoluto”, detto sistema di riferimento del mondo (*world reference*), la cui relazione col sistema della camera è spesso sconosciuta e quindi deve essere ricostruita sulla base delle immagini disponibili
- Le leggi della prospettiva ci dicono come si trasformano i punti della scena nei riferimenti solidali alla camera, le trasformazioni tra *camera reference* e *world reference* completano il quadro

Formazione delle immagini

- Per passare da un sistema di riferimento all'altro ci serve una *rototraslazione* nello spazio
- Tale *rototraslazione* è descrivibile con
 - T , vettore che esprime gli offset di *traslazione* (3 valori)
 - R , matrice che esprime la *rotazione* (3 gradi di libertà)

$$P_C = R(P_W - T)$$

relazione fra un punto espresso nel *sistema di riferimento del mondo* (P_W) e l'equivalente nel *sistema di riferimento della camera* (P_C)

- I sei parametri (tre coordinate e tre angoli di rotazione) che descrivono questa trasformazione, sono i ***parametri estrinseci*** della camera e consentono di passare dal sistema di riferimento del mondo a quello della camera

Formazione delle immagini

- Per definire completamente la proiezione cui ogni punto è sottoposto, servono altri parametri
 - la *lunghezza focale* f (in mm), correlata con lo *zoom ottico* nelle camere reali
 - il *piano d'immagine* ha un proprio sistema di riferimento, esprimibile in *pixel*, la cui origine è data dagli offset \mathbf{o}_x e \mathbf{o}_y , che rappresentano le coordinate (in pixel) del *centro ottico*
 - l'unità di misura del mondo 3D non è la stessa del piano immagine \Rightarrow definiamo \mathbf{s}_x e \mathbf{s}_y (in mm) come le dimensioni orizzontali e verticale di un *pixel* del sensore (a volte basta conoscere il loro rapporto $\alpha = s_y / s_x$, detto *aspect ratio*)
 - le lenti reali introducono una *distorsione radiale* (o *lenticolare*) esprimibile con i due parametri k_1 e k_2 (numeri puri)
- Queste sette grandezze sono dette **parametri intrinseci** ed esprimono la *mappatura* fra le coordinate *geometriche* e le coordinate *in pixel* nell'immagine digitale prodotta

Formazione delle immagini

(x, y) : coordinate geometriche di un punto sullo schermo della telecamera

(x_p, y_p) : coordinate del medesimo punto quando è riguardato come un pixel nella immagine digitale che si forma

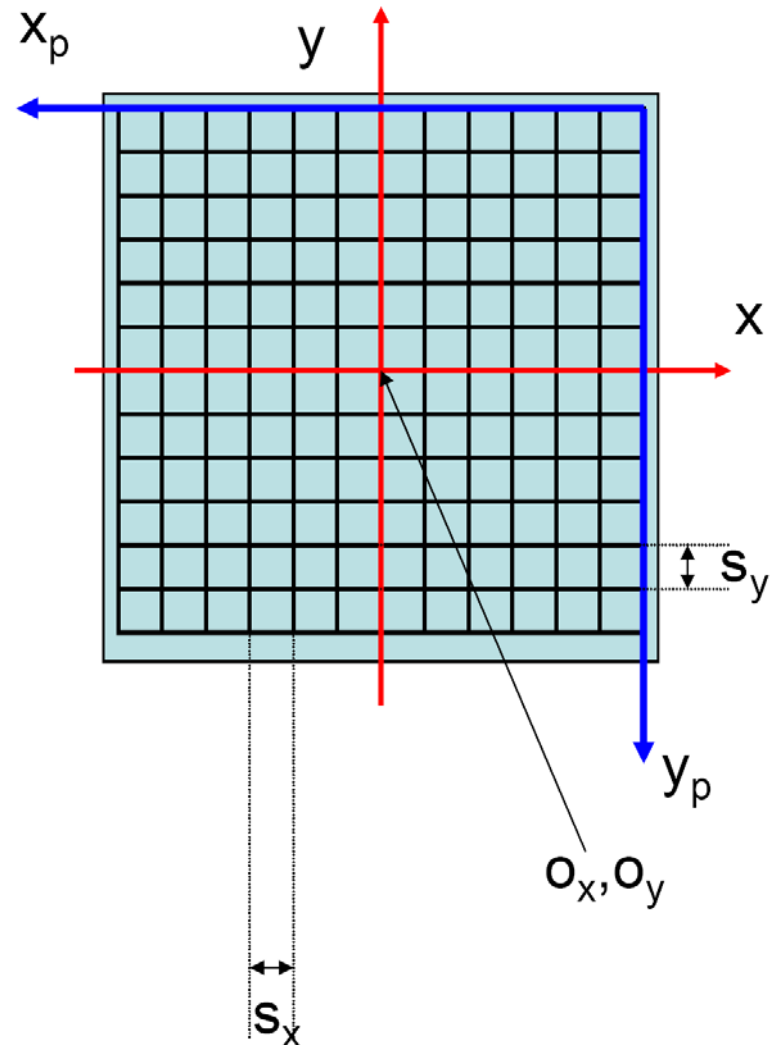
(s_x, s_y) : dimensioni effettive di un pixel

(o_x, o_y) : coordinate del pixel al centro dell'immagine

La relazione fra le coordinate *fisiche* e quelle *in pixel* è quindi

$$x = -(x_p - o_x)s_x$$

$$y = -(y_p - o_y)s_y$$



Formazione delle immagini

- Riassumendo, una telecamera reale è caratterizzata da due gruppi di parametri
 - ***parametri estrinseci***: sono i sei parametri che definiscono la *rototraslazione* del sistema di riferimento solidale alla telecamera rispetto al sistema di riferimento fisso
 - ***parametri intrinseci***: sono i sette parametri che dipendono da proprietà fisiche della camera fotografica
- La determinazione con accuratezza accettabile di tutti i parametri di una camera è detta ***calibrazione*** della camera.
- E' una esigenza fondamentale nella maggior parte di applicazioni di Computer Vision

Formazione delle immagini

- Le lenti reali distorcono l'immagine planare che si forma
- Tale distorsione è tipicamente piccola vicino al centro dell'immagine, ma diventa significativa nella periferia
- Un modello matematico “semplice” per tener conto di tale distorsione è detto *radiale*
- Se (x,y) sono le coordinate geometriche e (x_d,y_d) sono le coordinate dopo la *distorsione*, si ha

$$\begin{aligned}x &= x_d (1 + k_1 r^2 + k_2 r^4) & r^2 &= x_d^2 + y_d^2 \\y &= y_d (1 + k_1 r^2 + k_2 r^4)\end{aligned}$$

dove r è la distanza del punto dal centro dell'immagine

- Se si trascurano i pixel periferici, si può porre $k_2=0$, ma k_1 va comunque stimato per *calibrare* la camera

Formazione delle immagini

Esempi di *distorsione radiale* e loro correzione

La correzione richiede la conoscenza dei parametri intrinseci

Non affronteremo tale tema nel corso di queste lezioni



Formazione delle immagini

- Definiti i *parametri estrinseci* della camera, siamo in grado di tradurre il sistema di riferimento del mondo in quello della camera
- Per proiettare un punto della scena P_W sul piano immagine di una camera arbitrariamente orientata e calibrata, occorre
 1. tradurre le coordinate di P_W nel sistema di riferimento della camera tramite i *parametri estrinseci* [$P_C=R(P_W-T)$]
 2. proiettare P_C con le equazioni fondamentali della *Proiezione Prospettica*
 3. tradurre le coordinate geometriche del punto proiettato in pixel tramite i *parametri intrinseci*

Formazione delle immagini

- Con qualche semplice calcolo si ottiene

$$\begin{aligned} -(x_p - o_x)s_x &= f \frac{R_1(P_W - T)^T}{R_3(P_W - T)^T} \\ -(y_p - o_y)s_y &= f \frac{R_2(P_W - T)^T}{R_3(P_W - T)^T} \end{aligned}$$

versione lineare delle equazioni fondamentali di proiezione prospettica

ove R_i è la riga i -esima della Matrice di Rotazione

- Abbiamo trascurato la *distorsione*

Formazione delle immagini

- E' possibile riscrivere le precedenti equazioni lineari in una notazione più compatta ed espressiva, definendo

$$M_{\text{int}} = \begin{pmatrix} -f/s_x & 0 & o_x \\ 0 & -f/s_y & o_y \\ 0 & 0 & 1 \end{pmatrix} \quad M_{\text{ext}} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & -R_1 T^T \\ r_{21} & r_{22} & r_{23} & -R_2 T^T \\ r_{31} & r_{32} & r_{33} & -R_3 T^T \end{pmatrix}$$

Si ha così

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = M_{\text{int}} M_{\text{ext}} \begin{pmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{pmatrix} \quad \text{versione matriciale delle equazioni fondamentali di proiezione prospettica}$$

dove $x_p = x_1/x_3$ e $y_p = x_2/x_3$, essendo il punto proiettato in coordinate omogenee

Formazione delle immagini

- Le due matrici M_{int} e M_{ext} possono essere a loro volta compattate in una sola matrice $M = M_{int} M_{ext}$
- Per semplicità, poniamo $o_x = o_y = 0$ e $s_x = s_y = 1$, ossia ipotizziamo che la camera e il mondo abbiano la stessa unità di misura. Si ottiene

$$M = \begin{pmatrix} -fr_{11} & -fr_{12} & -fr_{13} & fR_1T^T \\ -fr_{21} & -fr_{22} & -fr_{23} & fR_2T^T \\ r_{31} & r_{32} & r_{33} & -R_3T^T \end{pmatrix}$$

- In assenza di ulteriori vincoli, M è detta *Matrice di Proiezione*
- La matrice M descrive una trasformazione tra un punto proiettivo nello spazio e un punto proiettivo nel piano

Calibrazione

- Abbiamo visto come si forma l'immagine di una scena: se conosciamo i dati 3D della scena e i parametri di una camera, sappiamo calcolare ciò che “vede” la camera
- Ma l'obiettivo della Computer Vision è risolvere il problema opposto: risalire da una o più immagini alle informazioni sulla scena
- Se conosciamo già i parametri della camera, questa operazione è relativamente semplice (*triangolazione*).
- Problema: come si procede se non conosciamo nulla della camera?

Calibrazione

- Soluzione: è possibile risalire ai parametri (intrinseci ed estrinseci) di una camera a partire da una sola immagine, purché conosciamo *a priori* (e.g., tramite *widget*) un certo numero n di corrispondenze tra punti della scena e punti dell'immagine
- Dati noti (input)
 - n punti nel sistema di riferimento del mondo
 - i corrispondenti n punti proiettati sul piano immagine
- Incognite (output)
 - parametri *estrinseci* della camera (posizione e orientazione)
 - parametri *intrinseci* della camera (mappatura tra punti e pixel) [hp. semplificativa: trascuriamo k_1 e k_2]
- Quanto vale n ?

Calibrazione

- *Calibrazione*: la determinazione con accuratezza accettabile dei parametri intrinseci ed estrinseci di una camera
- Esistono due metodi di *calibrazione* simili fra loro
 - *calibrazione diretta*
 - *calibrazione indiretta*

Calibrazione diretta

- Osservazione: dalle equazioni

$$-(x_p - o_x)s_x = f \frac{R_1(P_W - T)^T}{R_3(P_W - T)^T} \quad \text{versione lineare delle equazioni fondamentali di proiezione prospettica}$$
$$-(y_p - o_y)s_y = f \frac{R_2(P_W - T)^T}{R_3(P_W - T)^T}$$

si nota come l'effetto delle coordinate in pixel (o_x, o_y) del centro ottico sulle coordinate finali dei punti proiettati consiste in una traslazione: è come se spostassimo la camera lungo X e $Y \Rightarrow$ per semplicità, possiamo assorbire (o_x, o_y) in T , ponendoli pari a zero

- Troveremo successivamente i valori reali tramite il *metodo dell'ortocentro*

Calibrazione diretta

Se poniamo $o_x = o_y = 0$ nelle precedenti, otteniamo

$$-x_p s_x = f \frac{R_1 (P_W - T)^T}{R_3 (P_W - T)^T}$$

$$-y_p s_y = f \frac{R_2 (P_W - T)^T}{R_3 (P_W - T)^T}$$

Assumiamo, inoltre, che la traslazione venga effettuata dopo la rotazione e non prima

$$-x_p s_x = f \frac{R_1 P_W^T + T_x}{R_3 P_W^T + T_z}$$

$$-y_p s_y = f \frac{R_2 P_W^T + T_y}{R_3 P_W^T + T_z}$$

N.B. (T_x, T_y, T_z) sono vettori diversi da T , in quanto già ruotati!

Calibrazione diretta

Sviluppiamo la moltiplicazione tra vettori

$$-x_p s_x = f \frac{(r_{11}X + r_{12}Y + r_{13}Z + T_x)}{(r_{31}X + r_{32}Y + r_{33}Z + T_z)}$$

$$-y_p s_y = f \frac{(r_{21}X + r_{22}Y + r_{23}Z + T_y)}{(r_{31}X + r_{32}Y + r_{33}Z + T_x)}$$

Invertiamo il segno e portiamo i fattori di scala al secondo membro

$$x_p = -\frac{f}{s_x} \frac{(r_{11}X + r_{12}Y + r_{13}Z + T_x)}{(r_{31}X + r_{32}Y + r_{33}Z + T_z)}$$

$$y_p = -\frac{f}{s_y} \frac{(r_{21}X + r_{22}Y + r_{23}Z + T_y)}{(r_{31}X + r_{32}Y + r_{33}Z + T_x)}$$

Calibrazione diretta

Le precedenti hanno la seguente forma

$$x_p = -\frac{f (r_{11}X + r_{12}Y + r_{13}Z + T_x)}{s_x (r_{31}X + r_{32}Y + r_{33}Z + T_z)} \Rightarrow x_p = -A/D$$

$$y_p = -\frac{f (r_{21}X + r_{22}Y + r_{23}Z + T_y)}{s_y (r_{31}X + r_{32}Y + r_{33}Z + T_x)} \Rightarrow y_p = -B/D$$

da cui

$$x_p = -A/D, y_p = -B/D \Rightarrow -D = A/x_p, -D = B/y_p$$

$$\Rightarrow A/x_p = B/y_p \Rightarrow y_p A = x_p B$$

Si ottiene, pertanto

Calibrazione diretta

$$y_p \frac{f}{s_x} (r_{11}X + r_{12}Y + r_{13}Z + T_x) = x_p \frac{f}{s_y} (r_{21}X + r_{22}Y + r_{23}Z + T_y)$$

Poniamo $f_x=f/s_x$ e $f_y=f/s_y$, che rappresentano la lunghezza focale espressa in pixel (f è espressa in mm)

$$y_p f_x (r_{11}X + r_{12}Y + r_{13}Z + T_x) = x_p f_y (r_{21}X + r_{22}Y + r_{23}Z + T_y)$$

In precedenza abbiamo definito $\alpha=s_y/s_x$, per cui si ha

$$\frac{f_x}{f_y} = \frac{f/s_x}{f/s_y} = \frac{f}{s_x} \frac{s_y}{f} = \frac{s_y}{s_x} = \alpha$$

Nel seguito, essendo f_x e α legate alle incognite s_x e s_y , possiamo cercare le prime in luogo delle seconde

Calibrazione diretta

Dividendo ambo i membri per f_y , si ottiene

$$y_p \alpha (r_{11}X + r_{12}Y + r_{13}Z + T_x) = x_p (r_{21}X + r_{22}Y + r_{23}Z + T_y)$$

che possiamo riscrivere come

$$x_p r_{21}X + x_p r_{22}Y + x_p r_{23}Z + x_p T_y - y_p \alpha r_{11}X - y_p \alpha r_{12}Y - y_p \alpha r_{13}Z - y_p \alpha T_x = 0$$

Se poniamo

$$(v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8) = (r_{21}, r_{22}, r_{23}, T_y, \alpha r_{11}, \alpha r_{12}, \alpha r_{13}, \alpha T_x)$$

$$(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8) = (x_p X, x_p Y, x_p Z, x_p, -y_p X, -y_p Y, -y_p Z, -y_p)$$

la precedente diventa

$$a_1 v_1 + a_2 v_2 + a_3 v_3 + a_4 v_4 + a_5 v_5 + a_6 v_6 + a_7 v_7 + a_8 v_8 = 0$$

Calibrazione diretta

- Siamo dunque in presenza di una equazione lineare *omogenea* (ossia con termine noto uguale a zero), che possiamo riscrivere nella forma

$$Av=0$$

in cui abbiamo raccolto i coefficienti in un'unica matrice A e indicato con v il vettore delle incognite

- In generale, i sistemi omogenei possono avere
 - a) una sola soluzione, quella nulla;
 - b) infinite soluzioni; nel nostro caso, se il rango di A è 7, tutte queste infinite soluzioni sono legate fra loro da un fattore di proporzionalità

Calibrazione diretta

- Abbiamo otto incognite e il termine noto è nullo \Rightarrow occorre un sistema omogeneo di almeno sette equazioni, ossia almeno sette corrispondenze fra punti
- Per ogni coppia di punti di cui conosciamo le coordinate $(\text{punto dello spazio}, \text{punto proiettato}) = (X \ Y \ Z \ x \ y)$ otteniamo una equazione come la precedente
- In particolare, il punto i -esimo $(X_i \ Y_i \ Z_i \ x_i \ y_i)$ dà origine a un vettore di coefficienti $(a_{i1}, a_{i2}, a_{i3}, a_{i4}, a_{i5}, a_{i6}, a_{i7}, a_{i8})$
- In realtà, possono esserci anche un numero maggiore di corrispondenze (in tal caso la matrice A ha dimensione $N \times 8$). Si può infatti dimostrare che “se i punti considerati non giacciono sullo stesso piano \Rightarrow la matrice ha rango 7” e si può dunque procedere al calcolo della soluzione

Calibrazione diretta

- La soluzione non banale è ottenuta scomponendo la matrice A in $U\Sigma V^*$ tramite la *Singular Value Decomposition (SVD)*
- La colonna di V^* corrispondente al valore singolare nullo è la nostra soluzione
- Piccoli errori, rumore sperimentale e altro potrebbero portare, anche in presenza di punti non complanari, a una matrice A di rango massimo (cioè 8)
- In tal caso, si deve prendere come soluzione la colonna di V^* che corrisponde al più piccolo valore singolare

Calibrazione diretta

- Risolvendo il sistema tramite SVD, si ottiene un vettore soluzione v' , ossia il vettore soluzione v a meno di un fattore di proporzionalità c

$$v' = cv = c(r_{21}, r_{22}, r_{23}, T_y, \alpha r_{11}, \alpha r_{12}, \alpha r_{13}, \alpha T_x)$$

- Essendo R una *matrice di rotazione*, deve valere

$$r_{11}^2 + r_{12}^2 + r_{13}^2 = 1$$

$$r_{21}^2 + r_{22}^2 + r_{23}^2 = 1$$

da cui, ricordando come abbiamo definito gli elementi di v , deve essere

$$\sqrt{v_1'^2 + v_2'^2 + v_3'^2} = \sqrt{c^2 r_{21}^2 + c^2 r_{22}^2 + c^2 r_{23}^2} = |c|$$

- Dalle prime tre componenti di v' abbiamo dunque ricavato il fattore di proporzionalità c (a meno del segno)

Calibrazione diretta

- Analogamente, da v'_5, v'_6, v'_7 ricaviamo α (positivo per definizione)

$$\sqrt{v'^2_5 + v'^2_6 + v'^2_7} = \sqrt{c^2 \alpha^2 (r_{11}^2 + r_{12}^2 + r_{13}^2)} = \alpha |c|$$

da cui, conoscendo $|c|$ e il segno di α , otteniamo α

- Conosciamo $|c|$ e $\alpha \Rightarrow$ è banale trovare tutti gli elementi di v
- Poiché R è *ortonormale*, il prodotto vettoriale delle prime due righe ci dà la terza: conosciamo dunque la terza riga di R (incluso il segno) senza ancora conoscere il segno delle prime due righe
- Resta da chiarire il segno di T_x, T_y e delle prime due righe di R

Calibrazione diretta

- Per determinare il segno di T_x , T_y e delle prime due righe di R si considera una delle coppie di punti note e si verifica la discordanza tra i membri di

$$x_p = -\frac{f (r_{11}X + r_{12}Y + r_{13}Z + T_x)}{s_x (r_{31}X + r_{32}Y + r_{33}Z + T_z)}$$

$$y_p = -\frac{f (r_{21}X + r_{22}Y + r_{23}Z + T_y)}{s_y (r_{31}X + r_{32}Y + r_{33}Z + T_x)}$$

i due membri di ciascuna uguaglianza sono discordi

- In particolare, consideriamo le coordinate del punto prima e dopo la proiezione e verifichiamo che tale condizione sia vera (con $f, s_x, s_y, Z_C > 0$)
- In particolare, devono essere discordi il primo membro e il numeratore del secondo
- Se ciò non accade, significa che il segno di c è negativo; in questo caso invertiamo il segno

Calibrazione diretta

- Abbiamo determinato R , α , T_x e T_y ; rimangono da calcolare T_z e f_x . Per tale scopo utilizziamo una delle due equazioni iniziali

$$x_p = -\frac{f_x (r_{11}X + r_{12}Y + r_{13}Z + T_x)}{s_x (r_{31}X + r_{32}Y + r_{33}Z + T_z)} \Rightarrow$$

$$\Rightarrow x_p = -f_x \frac{(r_{11}X + r_{12}Y + r_{13}Z + T_x)}{(r_{31}X + r_{32}Y + r_{33}Z + T_z)} \Rightarrow$$

$$\Rightarrow x_p (r_{31}X + r_{32}Y + r_{33}Z + T_z) = -f_x (r_{11}X + r_{12}Y + r_{13}Z + T_x) \Rightarrow$$

$$\Rightarrow x_p (r_{31}X + r_{32}Y + r_{33}Z) + x_p T_z = -f_x (r_{11}X + r_{12}Y + r_{13}Z + T_x) \Rightarrow$$

$$\Rightarrow x_p T_z + (r_{11}X + r_{12}Y + r_{13}Z + T_x) f_x = -x_p (r_{31}X + r_{32}Y + r_{33}Z)$$

Calibrazione diretta

$$x_p T_z + (r_{11}X + r_{12}Y + r_{13}Z + T_x) f_x = -x_p (r_{31}X + r_{32}Y + r_{33}Z)$$

Al secondo membro figura il termine noto. Da un insieme di equazioni in questa forma è semplice ricavare un sistema lineare

$$A \begin{pmatrix} T_z \\ f_x \end{pmatrix} = b$$

la cui soluzione ottimale (nel senso dei minimi quadrati) è data da

$$\begin{pmatrix} \tilde{T}_z \\ \tilde{f}_x \end{pmatrix} = (A^T A)^{-1} A^T b$$

Calibrazione diretta

Vale infatti il seguente richiamo di Algebra Lineare: se ho $Ax=b$ e voglio trovare la soluzione che dà il minimo errore nel senso dei minimi quadrati debbo semplicemente fare

$$x = (A^T A)^{-1} A^T b$$

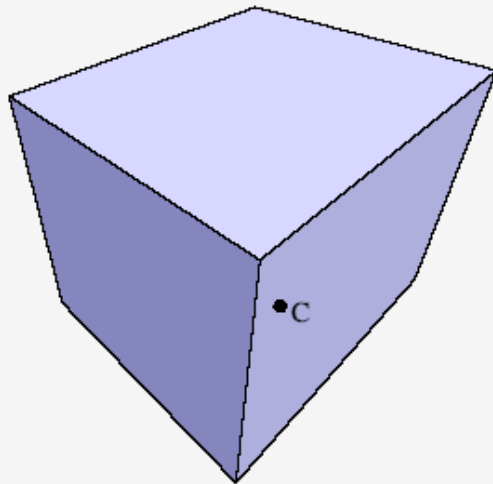
Calibrazione diretta

- Il metodo di *calibrazione* presentato è tratto dal libro di testo [Trucco&Verri]
- E' una semplificazione dell'algoritmo di Tsai, che include anche la correzione della *distorsione*, cioè la stima dei parametri k_1 e k_2 (che noi abbiamo trascurato per semplicità)
- E' assai diffuso e impiegato nei laboratori di tutto il mondo

Metodo dell'Ortocentro

- All'inizio della trattazione relativa alla *calibrazione diretta*, abbiamo posto per semplicità $o_x = o_y = 0$, facendo in tal modo coincidere il *centro ottico* con il *centro geometrico* dell'immagine
- Determiniamo ora dove si trova il *centro ottico* effettivo
- Un metodo visuale molto semplice per individuarlo è quello proposto da Caprile e Torre nel 1990, noto come *Metodo dell'Ortocentro*
- Esso si basa su una proprietà geometrica (dimostrata) dei *punti di fuga*: “se l'immagine è a tre *punti di fuga*, il *centro ottico* del piano immagine coincide con l'*ortocentro* (intersezione delle altezze) del triangolo i cui vertici sono proprio i tre punti di fuga dell'immagine”

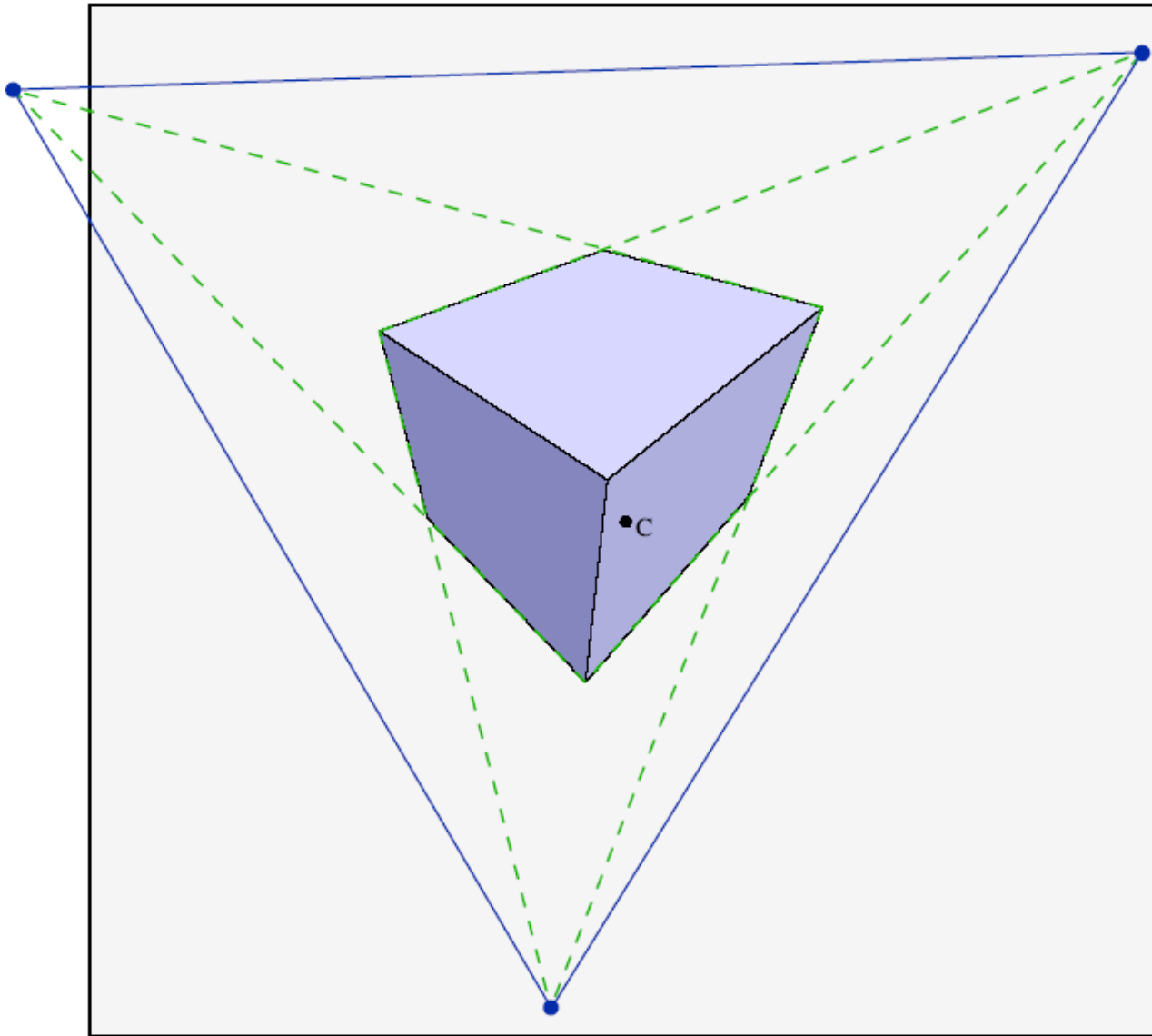
Metodo dell'Ortocentro



esempio di applicazione
del *Metodo dell'Ortocentro*
a partire dall'immagine di
un cubo a tre punti di fuga

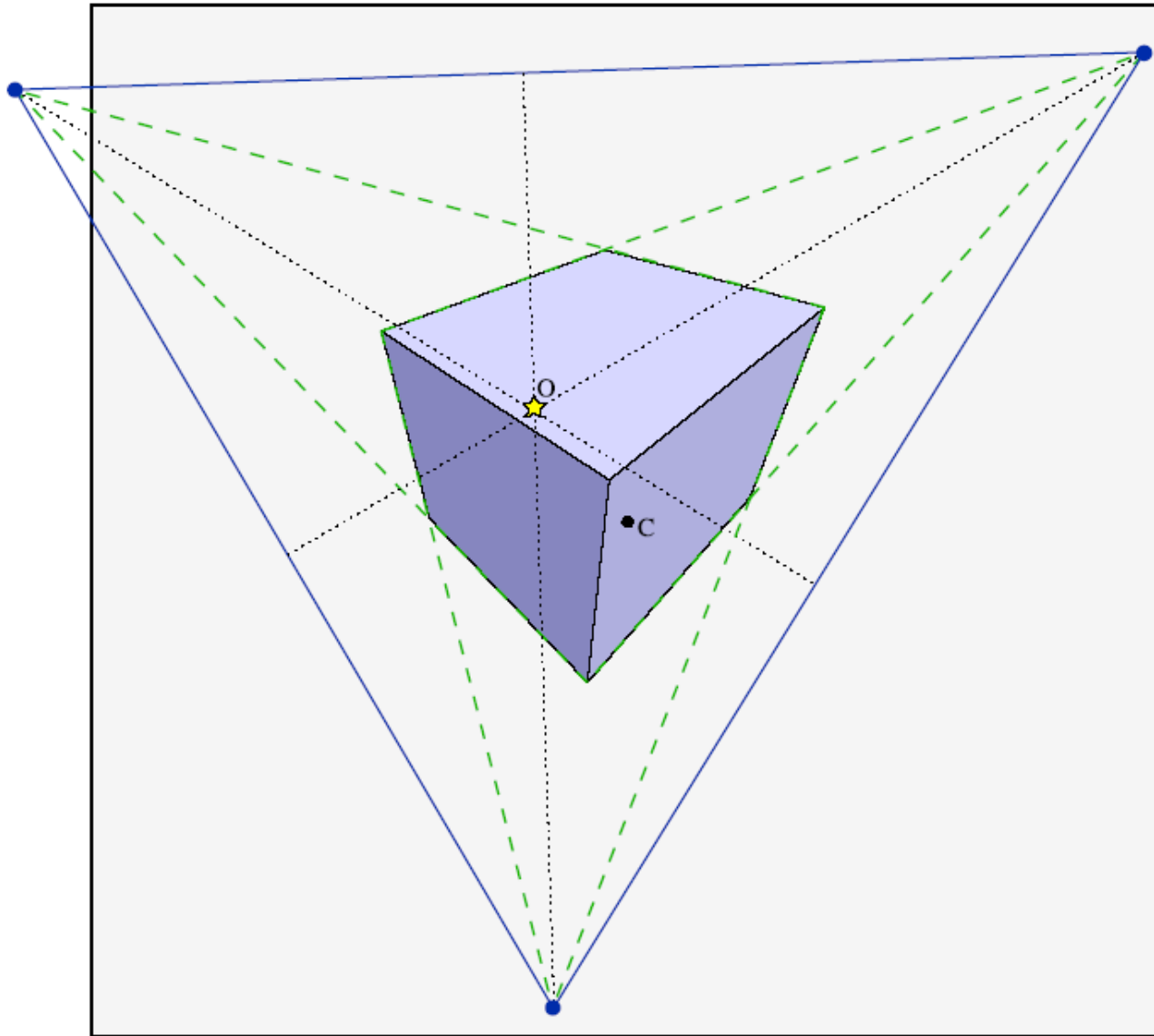
C: centro geometrico
dell'immagine

Metodo dell'Ortocentro



uno dei punti di fuga
ricade al di fuori
dell'immagine

Metodo dell'Ortocentro



il centro ottico, O , in questo caso, si discosta dal centro geometrico C dell'immagine

Metodo dell'Ortocentro

Problemi

- Per applicare il *Metodo dell'Ortocentro* occorre che l'immagine sia a tre *punti di fuga*: non tutte lo sono (*Condizione sufficiente affinché un'immagine sia a tre punti di fuga è che siano visibili almeno tre coppie di linee parallele alla scena tali che ogni coppia sia perpendicolare alle altre due e non perpendicolare all'asse ottico della camera*)
- Se i tre *punti di fuga*, tipicamente esterni all'immagine visibile, sono lontani dall'immagine stessa (ossia, se le rette che convergono in quel punto sono “quasi parallele”), l'errore di stima che si commette può risultare non trascurabile
- Per stimare con precisione le rette, solitamente si procede con *rilevazione bordi + fitting lineare*, ma anche con immagini artificiali l'errore può risultare significativo

Metodo dell'Ortocentro

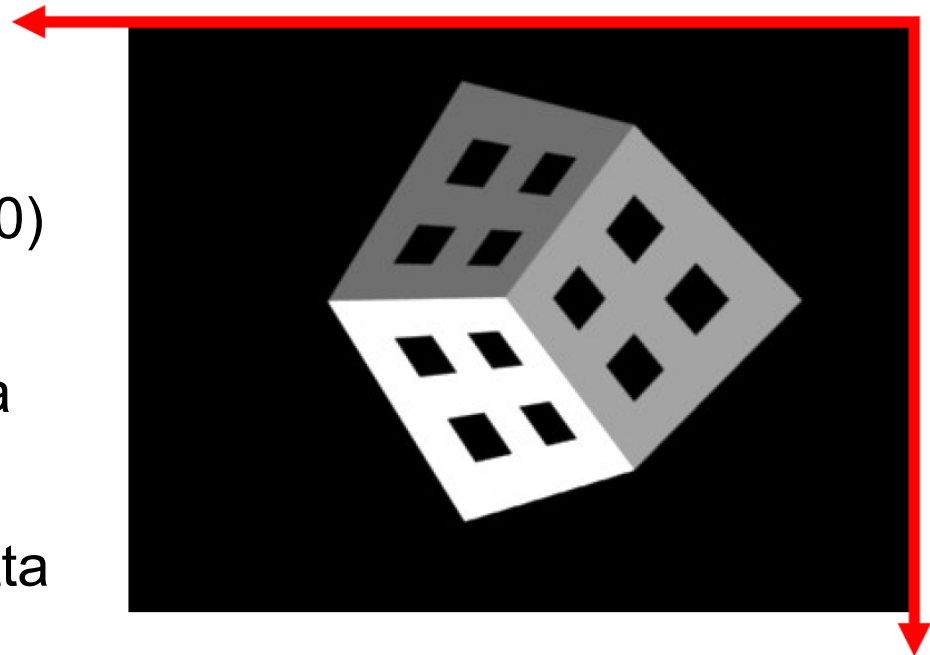
Caso pratico: calcolo del centro di proiezione, ossia dei parametri intrinseci o_x e o_y

- Operazione puramente geometrica
- Si parte da una foto di un oggetto di calibrazione
- Operazione da eseguire *off-line* una volta per tutte se non si cambia focale o lente

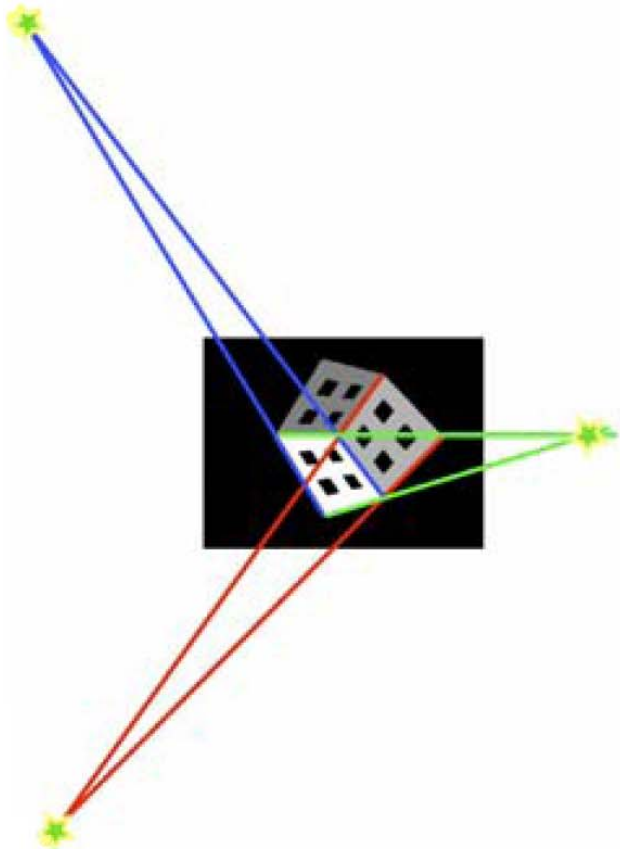
Non è affatto detto che il *centro ottico* coincida con il centro dell'immagine digitale (che in questo caso cade in (200,150) essendo l'immagine 400 x 300)

Spesso il sensore è messo fuori asse di qualche mm per ragioni di progetto della camera

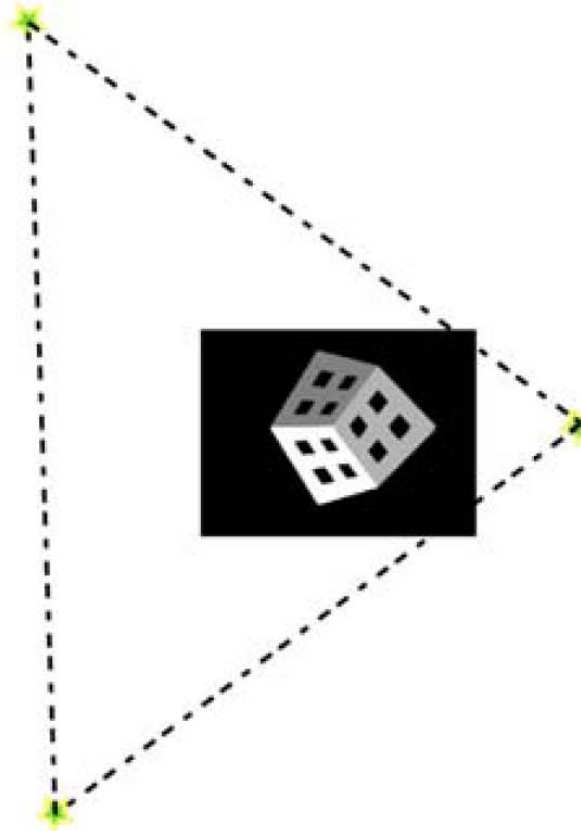
Nell'esempio: immagine sintetica prodotta con la camera "ideale" di *Blender*



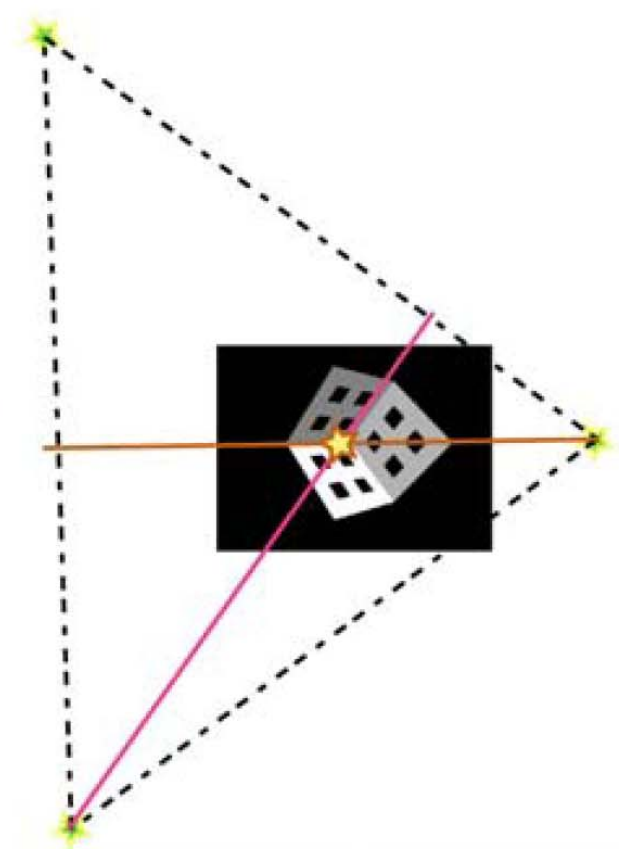
Metodo dell'Ortocentro



1. Calcolo dei punti di fuga

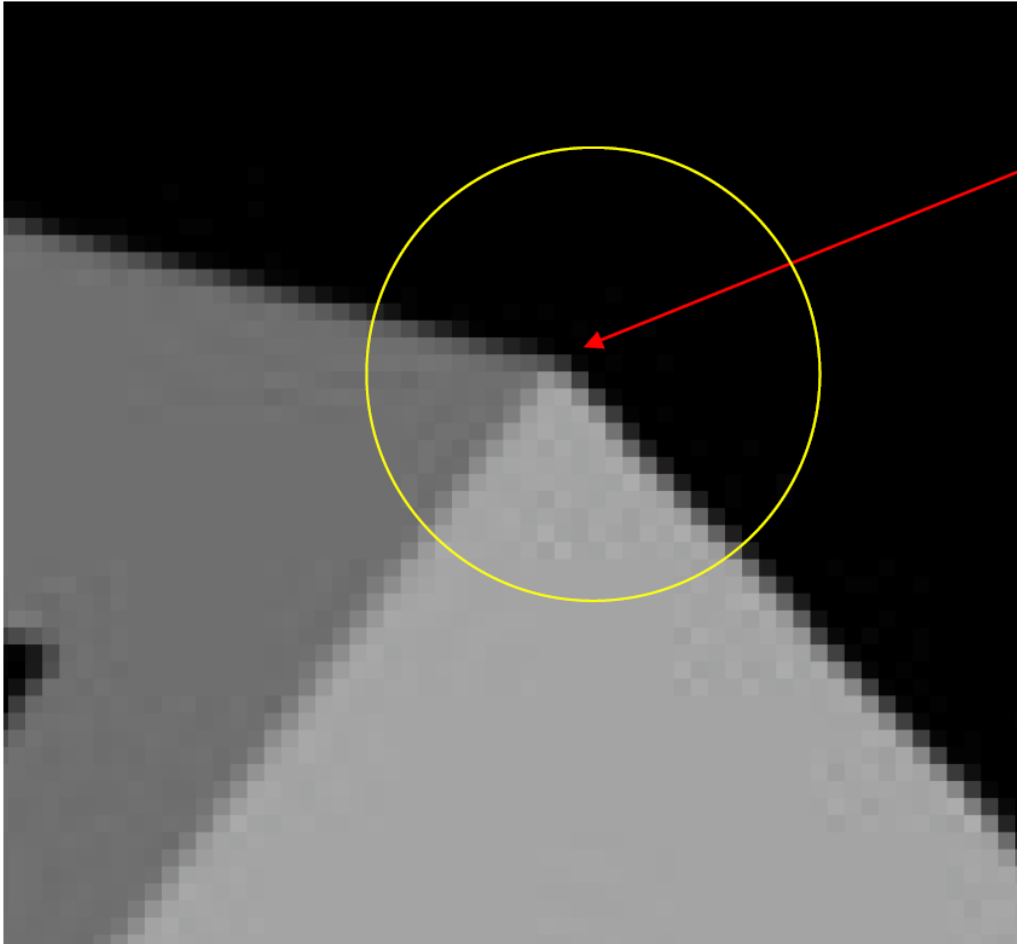


2. Calcolo del triangolo



3. Calcolo dell'ortocentro

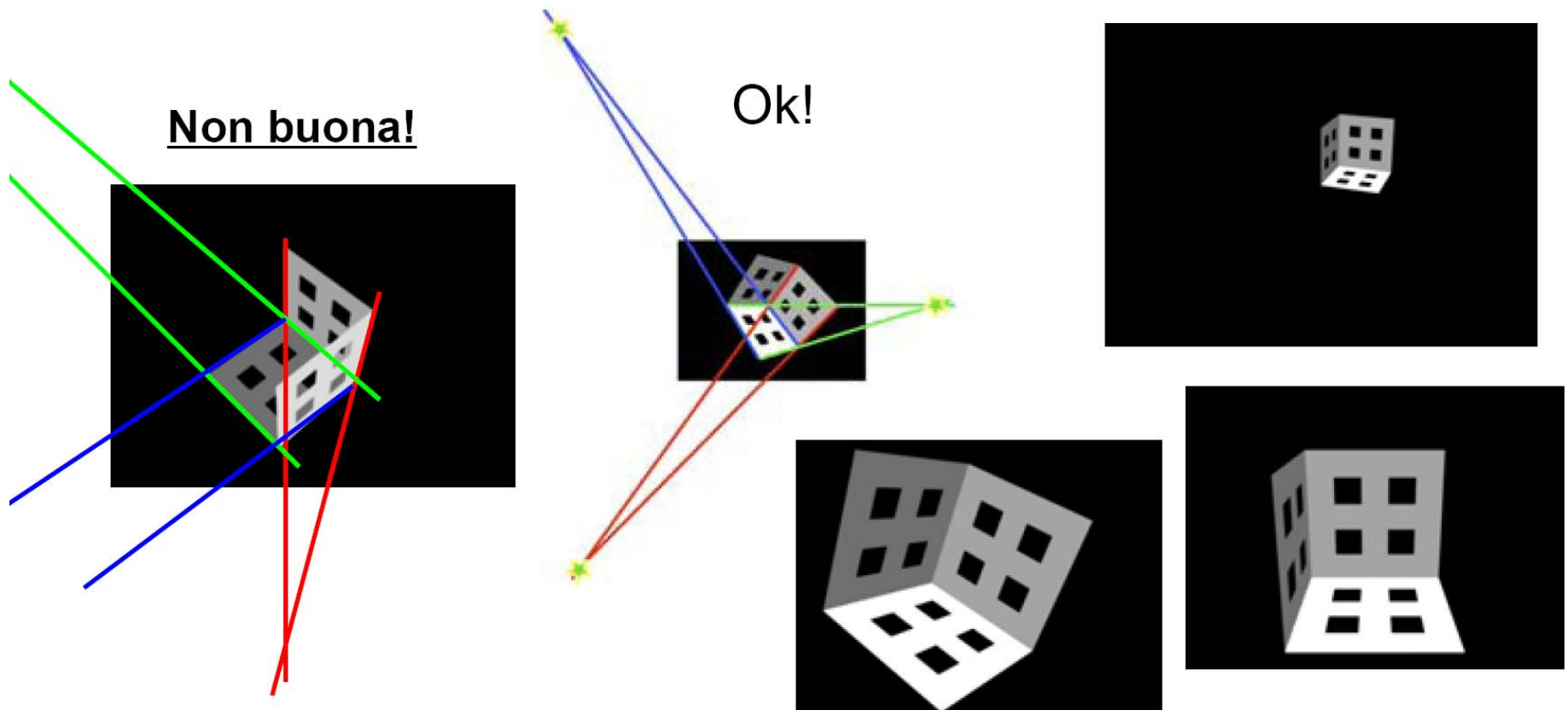
Metodo dell'Ortocentro



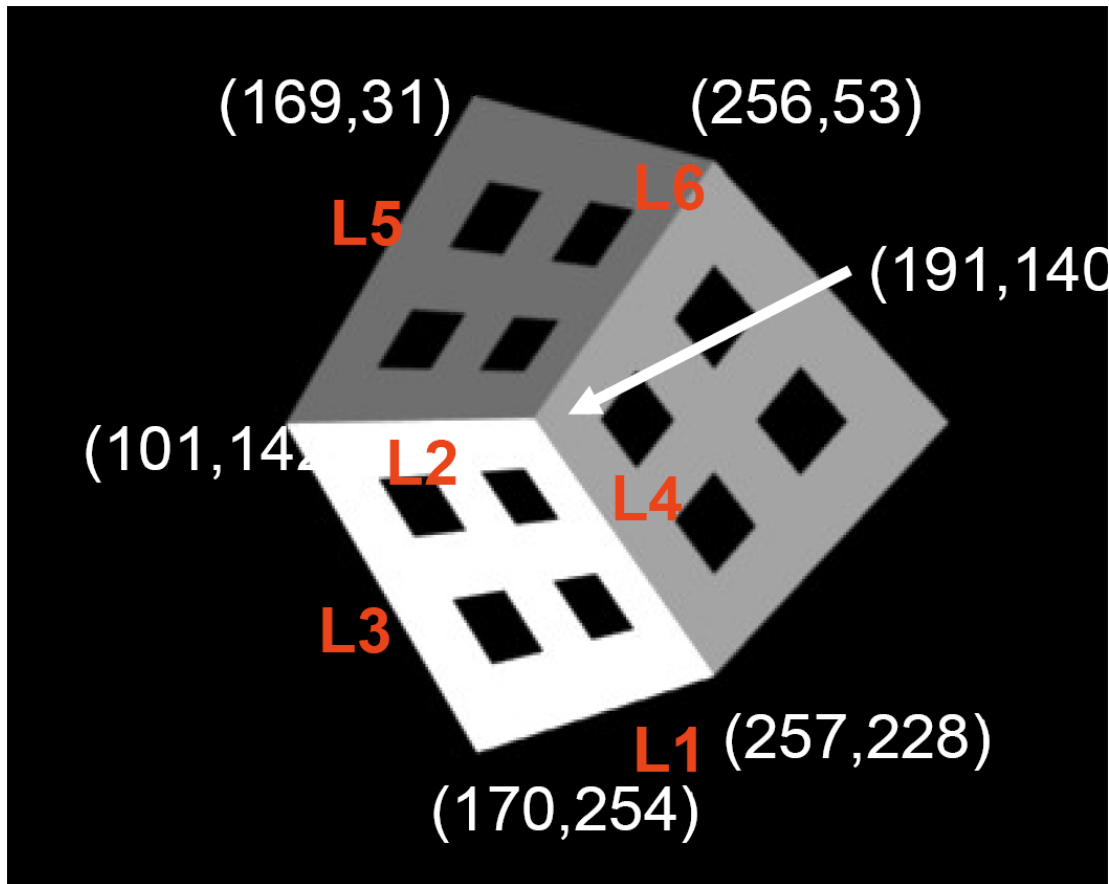
Problema n. 1: difficile conoscere esattamente le coordinate di questo spigolo (importante, in quanto attraverso di esso devono passare le rette usate per individuare il punto di fuga)

Metodo dell'Ortocentro

Problema n. 2: le tre coppie di rette convergenti scelte non si incontrano “abbastanza vicino” al centro dell'immagine, ma sono quasi parallele; in tal caso gli errori di stima e di approssimazione rendono il calcolo del centro ottico totalmente inaffidabile



Metodo dell'Ortocentro



Se partiamo dalle coordinate dei vari spigoli del *widget* di calibrazione, si trova che le intersezioni delle coppie di linee convergenti cadono nei punti (580,131), (-320,-541), (-302,800)

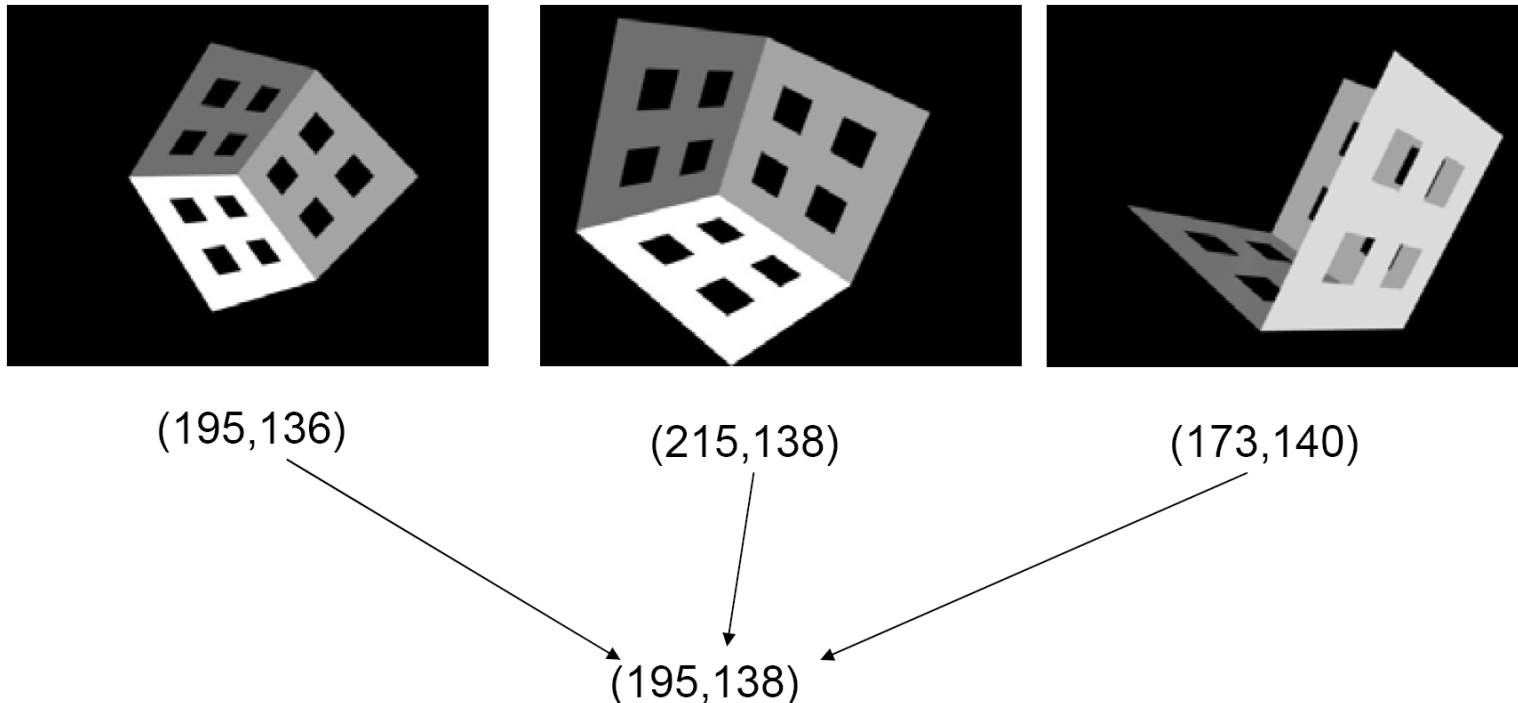
L'*ortocentro* del triangolo che ha questi tre punti come vertici è (195,136)

L'esperienza insegna che i numeri che vengono fuori sono assai sensibili anche a piccole variazioni delle coordinate dei punti in input

Variazioni dello 0,25% della dimensione dell'immagine producono errori fino al 1% nella locazione del centro ottico

Metodo dell'OrtoCentro

Soluzione n.1: esaminare più foto diverse dell'oggetto di calibrazione e fare la media delle locazioni del centro trovate. Le immagini devono essere *prospettive a tre punti*

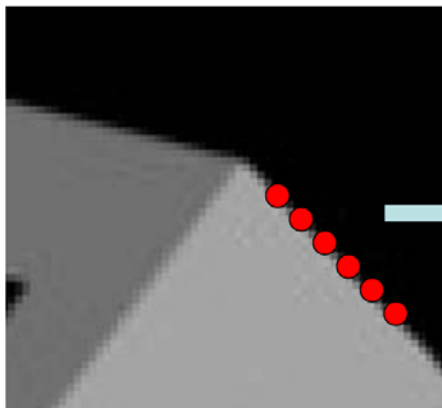


Questo è un interessante *caso di studio* che mostra come una normale precisione “a pixel” non garantisce soluzioni affidabili e con errore controllato \Rightarrow occorre scendere a livello “sub-pixel”

Metodo dell'Ortocentro

Soluzione n.2: finora abbiamo individuato gli spigoli e tracciato le rette che passano per ogni coppia di essi.

Una soluzione migliore è prendere più punti lungo ciascun lato dell'oggetto calibratore e calcolare con un *fitting* l'equazione della retta che passa per tale lato \Rightarrow la migliore stima di queste rette consente di ottenere un risultato migliore



(x_1, y_1)
 (x_2, y_2)
...
 (x_6, y_6)

Tali punti devono soddisfare tutti una equazione del tipo $Ax+By+C=0$, con A, B, C incogniti

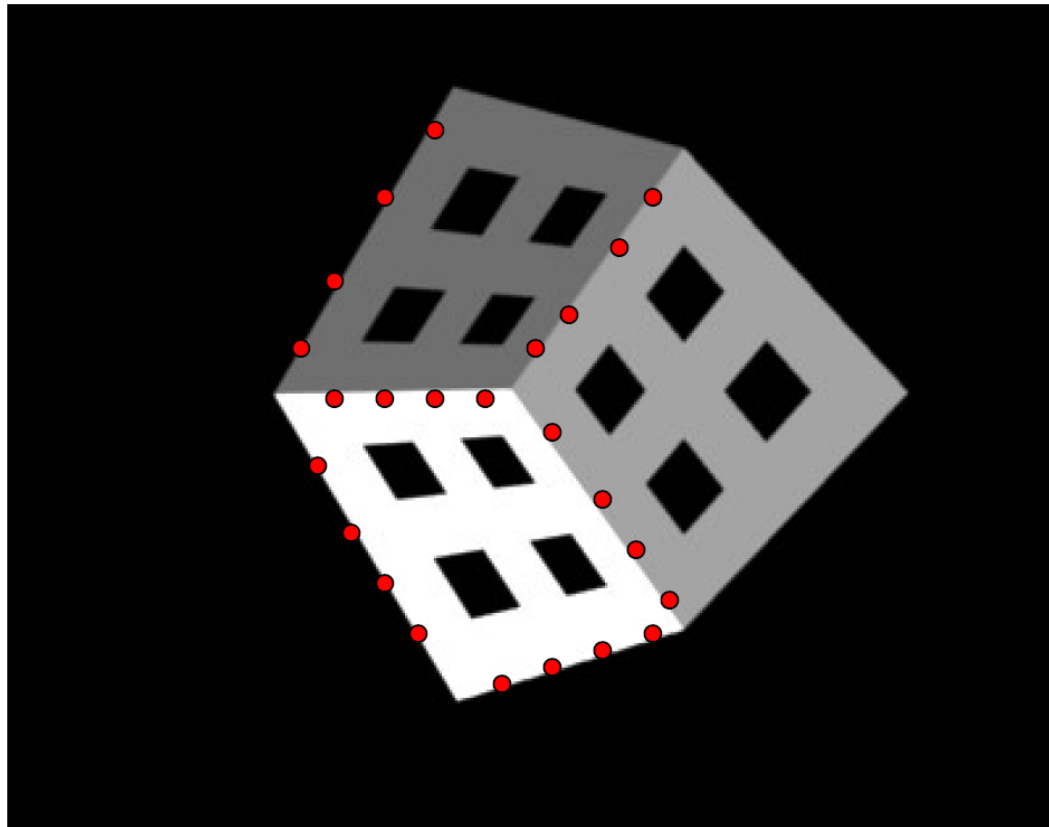
$x_1 \ y_1 \ 1$
 $x_2 \ y_2 \ 1$
...
 $x_6 \ y_6 \ 1$

$*$ $\begin{matrix} A \\ B \\ C \end{matrix}$

$=0$ La soluzione può essere trovata decomponendo la matrice con SVD e prendendo la colonna di V^* associata al valore singolare nullo

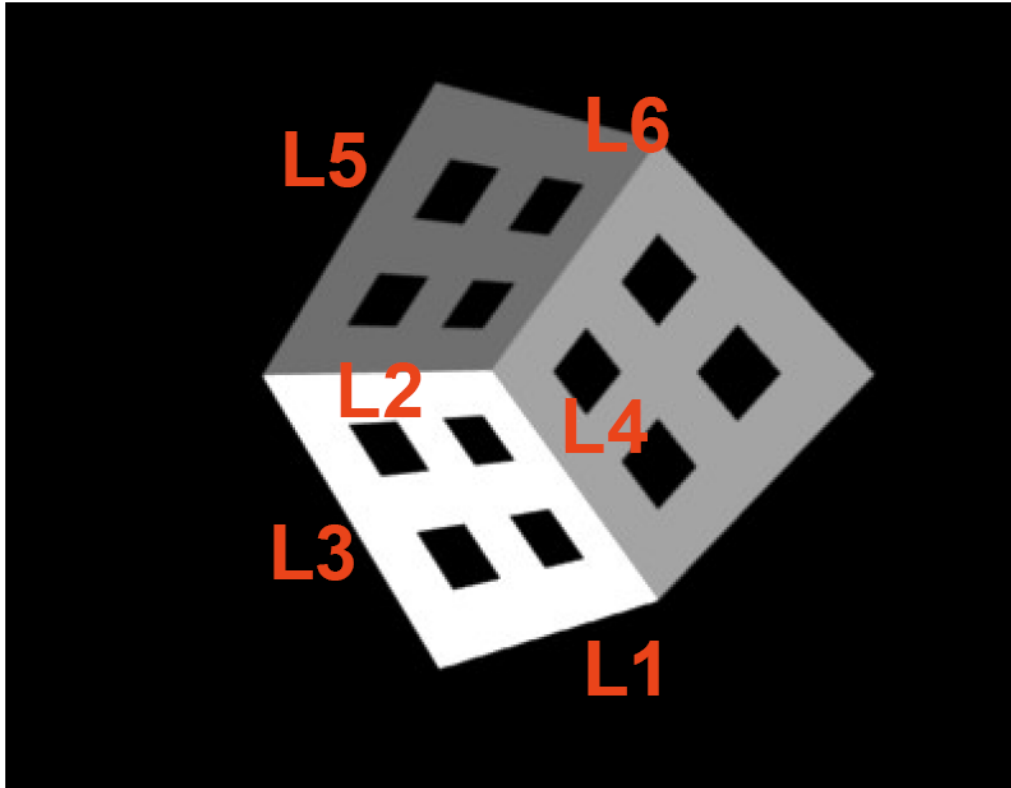
Metodo dell'Ortocentro

Una accortezza da seguire è quella di prendere molti punti



In CV esistono numerosi operatori per stimare *corner*, *linee* e altre *feature geometriche* in maniera automatica o quasi

Metodo dell'Ortocentro



Questo problema mette bene in evidenza la tipica struttura che un software di CV deve avere

Occorre integrare la possibilità di individuare punti nella immagine (in modalità manuale o automatica) con la possibilità di eseguire molti calcoli di Algebra Lineare

Prendendo sei punti per ciascun lato e svolgendo i calcoli di cui sopra, si ottiene una stima per il centro di (188,145) che, per come è fatta la camera di *Blender*, risulta più corretto

Calibrazione indiretta

- Un approccio alternativo alla *calibrazione diretta* dei parametri consiste nello stimare la *Matrice di Proiezione* M che traduce i punti della scena in punti dell'immagine
- E' poi possibile risalire ai parametri della camera tramite formule dirette da applicare agli elementi della matrice
- In alcune applicazioni non è necessario risalire ai singoli parametri e calcolare la sola *matrice di proiezione* è sufficiente agli scopi: in tal caso, questo approccio ha il vantaggio di essere più snello e veloce del precedente
- Trovare la *matrice di proiezione* significa stimare la matrice M (3x4) tale che

$$\begin{pmatrix} u_i \\ v_i \\ w_i \end{pmatrix} = M \begin{pmatrix} X_i^w \\ Y_i^w \\ Z_i^w \\ 1 \end{pmatrix}$$

Calibrazione indiretta

- Possiamo scrivere M in funzione dei parametri

$$M = \begin{bmatrix} -f_x r_{11} + o_x r_{31} & -f_x r_{12} + o_x r_{32} & -f_x r_{13} + o_x r_{33} & -f_x T_x + o_x T_z \\ -f_y r_{21} + o_y r_{31} & -f_y r_{22} + o_y r_{32} & -f_y r_{23} + o_y r_{33} & -f_y T_y + o_y T_z \\ r_{31} & r_{32} & r_{33} & T_z \end{bmatrix}$$

- M ha 12 elementi ma 11 gradi di libertà (perché definita a meno di un fattore di scala c) \Rightarrow occorrono almeno 11/12 equazioni per trovare tutti gli elementi m_{ij}
- Abbiamo bisogno di sei coppie di punti corrispondenti (proiettati e non proiettati). Le equazioni del sistema saranno del tipo

$$x_p = \frac{u_i}{w_i} = \frac{m_{11}X_i^w + m_{12}Y_i^w + m_{13}Z_i^w + m_{14}}{m_{31}X_i^w + m_{32}Y_i^w + m_{33}Z_i^w + m_{34}}$$

$$y_p = \frac{v_i}{w_i} = \frac{m_{21}X_i^w + m_{22}Y_i^w + m_{23}Z_i^w + m_{24}}{m_{31}X_i^w + m_{32}Y_i^w + m_{33}Z_i^w + m_{34}}$$

Calibrazione indiretta

- Analogamente a quanto fatto per la *calibrazione diretta*, impostiamo un sistema *omogeneo* del tipo $Am=0$ il cui vettore incognito m è formato dagli elementi di M

$$m = [m_{11}, m_{12}, \dots, m_{33}, m_{34}]^T$$

- Il *rango* di A è 11 \Rightarrow il sistema può essere risolto nuovamente tramite SVD: troviamo un vettore soluzione

$$\tilde{m} = cm$$

di cui non conosciamo il *fattore di scala* c

- Anche in questo caso, sfruttiamo le proprietà di R per determinare c

$$\sqrt{\tilde{m}_{31}^2 + \tilde{m}_{32}^2 + \tilde{m}_{33}^2} = \sqrt{c^2 (r_{21}^2 + r_{22}^2 + r_{23}^2)} = |c|$$

Calibrazione indiretta

- Definiamo per comodità i seguenti vettori

$$q_1 = [\tilde{m}_{11}, \tilde{m}_{12}, \tilde{m}_{13}]^T$$

$$q_2 = [\tilde{m}_{21}, \tilde{m}_{22}, \tilde{m}_{23}]^T$$

$$q_3 = [\tilde{m}_{31}, \tilde{m}_{32}, \tilde{m}_{33}]^T$$

$$q_4 = [\tilde{m}_{14}, \tilde{m}_{24}, \tilde{m}_{34}]^T$$

- I primi tre vettori corrispondono alle prime tre *righe* (meno l'ultimo elemento) di M , mentre q_4 rappresenta l'ultima *colonna* di M

Calibrazione indiretta

- Conosciamo c a meno del segno.
- Definiamo allora $\sigma = \pm 1$ e ricaviamo le formule per calcolare tutti i parametri di cui abbiamo bisogno

$$T_z = \sigma \tilde{m}_{34}$$

$$r_{3i} = \sigma \tilde{m}_{3i}, \quad i = 1, 2, 3$$

$$o_x = q_1 q_3^T$$

$$o_y = q_2 q_3^T$$

$$f_x = \sqrt{q_1 q_1^T - o_x^2}$$

$$f_y = \sqrt{q_2 q_2^T - o_y^2}$$

$$r_{1i} = \sigma(o_x \tilde{m}_{3i} - \tilde{m}_{1i}) / f_x, \quad i = 1, 2, 3$$

$$r_{2i} = \sigma(o_y \tilde{m}_{3i} - \tilde{m}_{2i}) / f_y, \quad i = 1, 2, 3$$

$$T_x = \sigma(o_x T_z - \tilde{m}_{14}) / f_x$$

$$T_y = \sigma(o_y T_z - \tilde{m}_{24}) / f_y$$

Calibrazione indiretta

- Rimane da determinare solo il segno di σ
- A tal fine, consideriamo l'uguaglianza

$$T_z = \sigma \tilde{m}_{34}$$

- Se l'origine del sistema di riferimento del mondo 3D si trova davanti alla camera, T_z deve essere positivo, dunque σ deve essere concorde con m_{34} ; altrimenti, deve essere negativo, da cui segue che σ deve risultare di segno opposto rispetto a m_{34}
- Stabilito il segno di σ , abbiamo determinato con esattezza tutti i parametri della camera

La luce: Fotometria della Formazione delle immagini



La luce: Fotometria della Formazione delle immagini



195	209	221	235	249	251	254	255	250	241	247	248
210	236	249	254	255	254	225	226	212	204	236	211
164	172	180	192	241	251	255	255	255	255	235	190
167	164	171	170	179	189	208	244	254	255	251	234
162	167	166	169	169	170	176	185	196	232	249	254
153	157	160	162	169	170	168	169	171	176	185	218
126	135	143	147	156	157	160	166	167	171	168	170
103	107	118	125	133	145	151	156	158	159	163	164
095	095	097	101	115	124	132	142	117	122	124	161
093	093	093	093	095	099	105	118	125	135	143	119
093	093	093	093	093	093	095	097	101	109	119	132
095	093	093	093	093	093	093	093	093	093	093	119

$I(x,y,t)$ è l'intensità della luce in (x,y) all'istante t

Sensore *CCD* $\approx 5.000.000$ pixel

Occhio umano $\approx 240.000.000$ pixel (i.e., 0.25 terabit/s)

La luce: Fotometria della Formazione delle immagini

- La luminosità di un pixel è proporzionale alla quantità di luce indirizzata verso la fotocamera dalla porzione di scena che si proietta su quel pixel; questa, a sua volta, dipende da
 - proprietà riflettenti delle superfici
 - posizione e distribuzione delle fonti di luce nella scena
 - proprietà riflettenti del resto della scena
- La riflessione può essere modellata in due modi
 - riflessione *speculare* (la luce è riflessa dalla superficie esterna di un oggetto; vincolo: angolo di riflessione = angolo di incidenza)
 - riflessione *diffusa* (la luce penetra nella superficie di un oggetto, è assorbita e riemessa; se la superficie è *lambertiana*, la luce si sparge in tutte le direzioni con uguale intensità)

La luce: Fotometria della Formazione delle immagini

- L'intensità dipende solo dall'angolo di incidenza della fonte di luce
- Vale la *Legge del Coseno di Lambert*

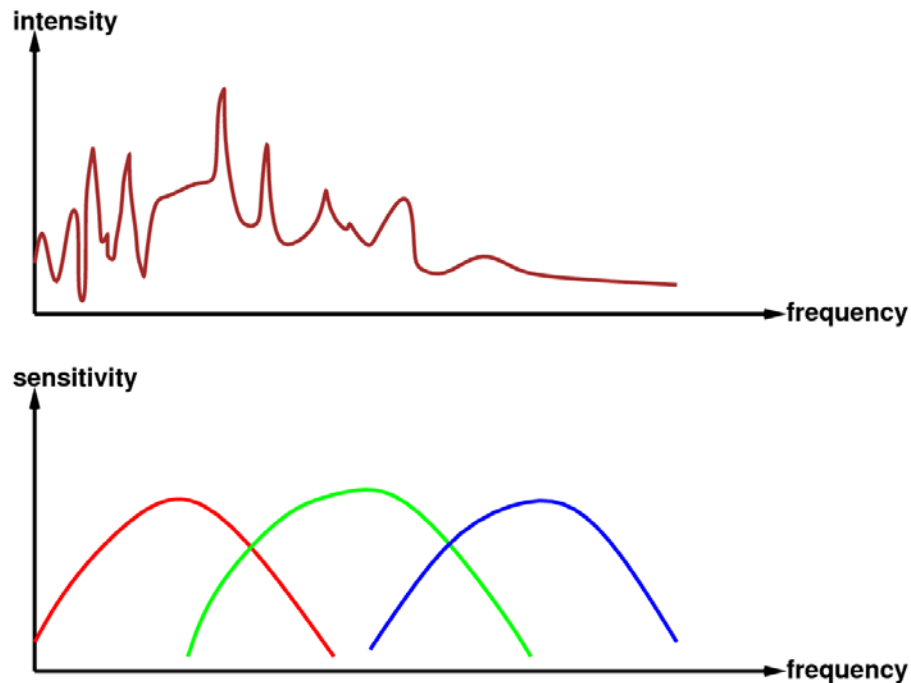
$$I = kI_0 \cos \theta$$

dove I_0 è l'intensità della sorgente luminosa, θ l'angolo tra essa e la normale alla superficie e k una costante detta *albedo*, che dipende dalle proprietà delle superfici ($k=0$ se nere, $k=1$ se bianche)

- Nel mondo reale le superfici mostrano una combinazione di caratteristiche speculari e diffuse
- *Ray Tracing* nella Grafica Computerizzata, tecnica che simula il processo in cui la luce, proveniente dalle fonti luminose, subisce riflessioni multiple

La luce: Fotometria della Formazione delle immagini

L'intensità varia con la frequenza \Rightarrow segnale a infinite dimensioni



L'occhio umano ha tre tipi di cellule sensibili al colore.

Ciascuna integra il segnale \Rightarrow un'immagine può essere rappresentata da un vettore di tre soli valori di intensità per pixel, uno per ciascuna lunghezza d'onda primaria