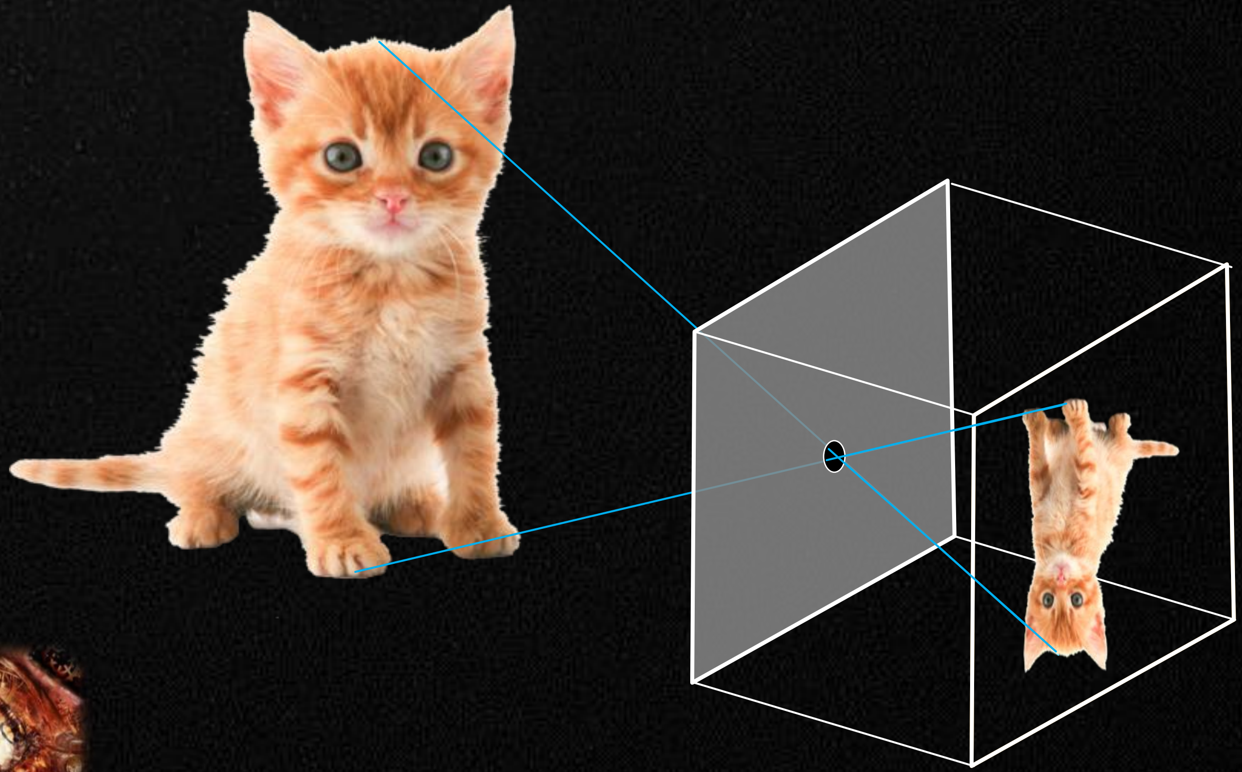
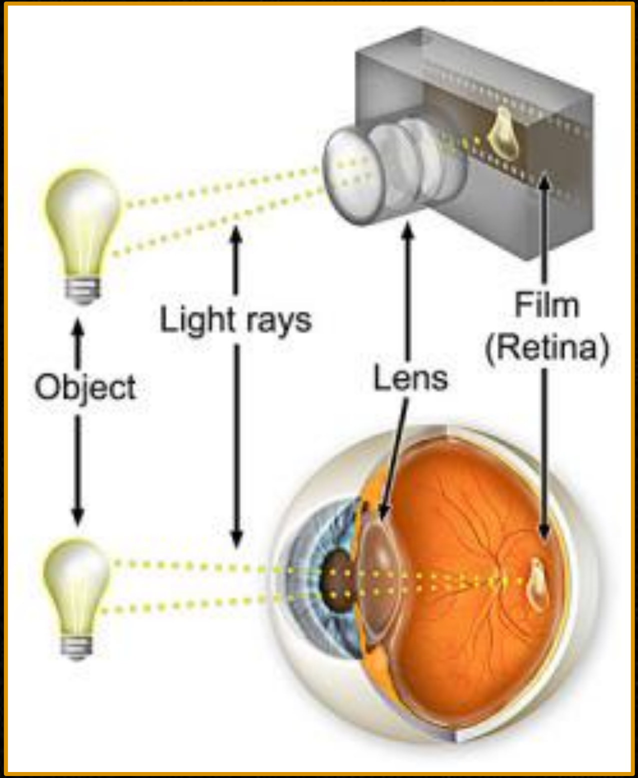
Formazione immagini e calibrazione telecamera

17634 | VISIONE ARTIFICIALE

Formazione delle immagini

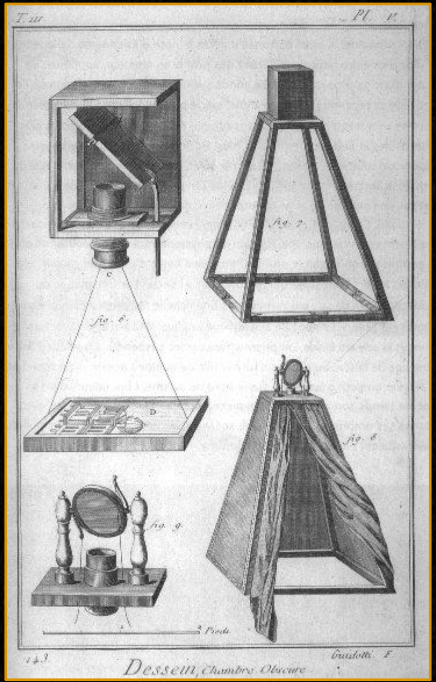
- Il processo di formazione delle immagini è dovuto a raggi di luce che entrano in un sistema ottico e colpiscono un sensore (piano immagine):
  - Nell'occhio umano il sistema ottico è il cristallino, il piano immagine la retina;
  - In una fotocamera il sistema ottico è la lente, il piano immagine il sensore digitale.

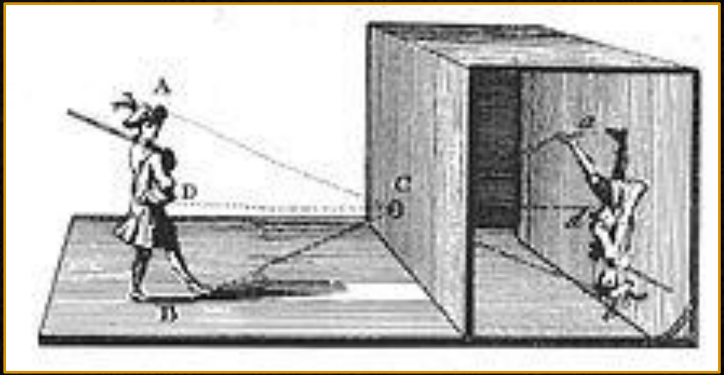


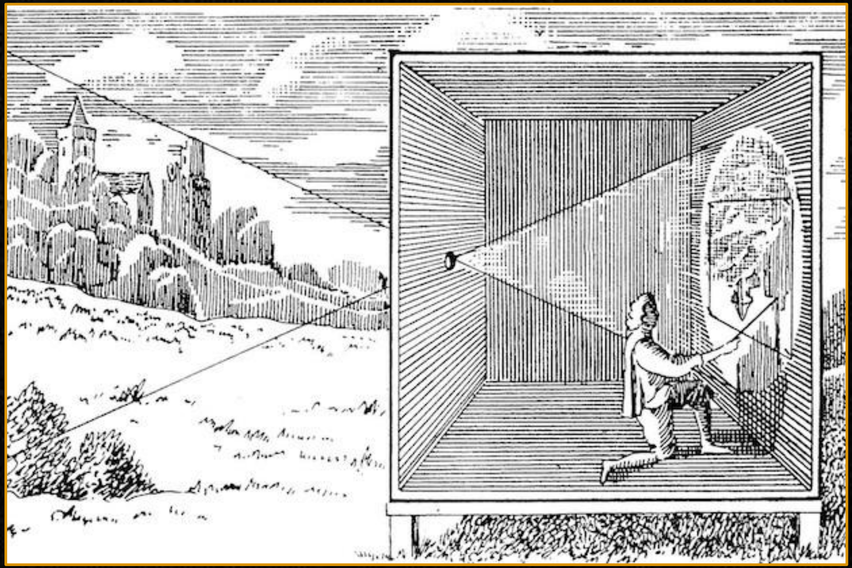


Stenoscopia e camera oscura

- Utilizzo di un foro stenopeico (dal greco *stenos opaios*, stretto foro) capace di proiettare la luce creando un'immagine.
  - Camera oscura: una stanza completamente buia con un foro su un lato e un pittore all'interno che copia l'immagine proiettata.
  - La scoperta della camera oscura risale almeno al V secolo a.C., sulla base di descrizioni del fenomeno presenti in testi cinesi e greci dell'epoca.
- Fotocamere odierne: evoluzione di questa antica tecnica (il nome inglese "camera" deriva dal latino "camera obscura").







Camera oscura involontaria

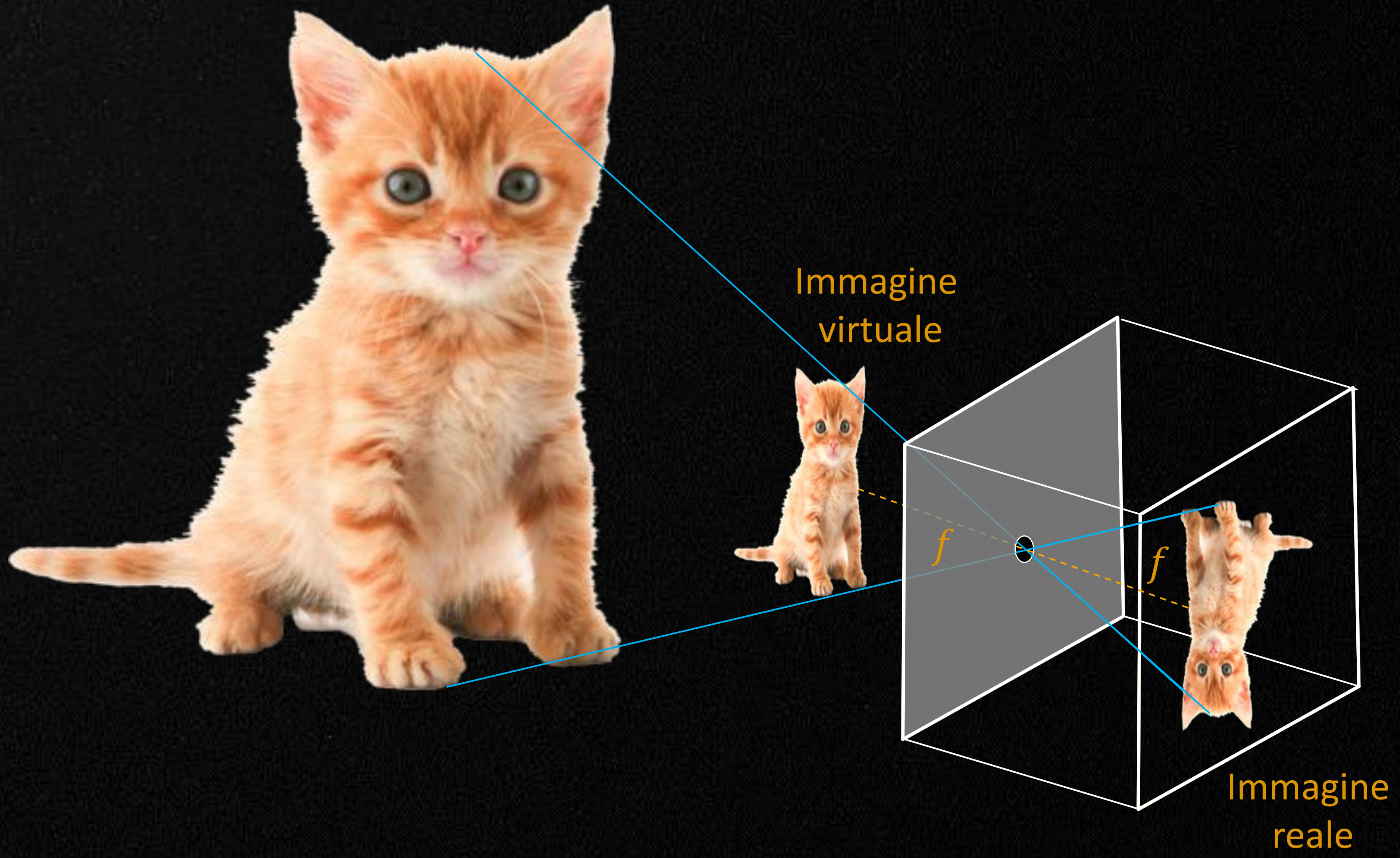
- Nella foto (Castel Grande di Bellinzona, Svizzera): effetto "camera oscura" sulla parete opposta ad alcune feritoie
- Si possono intravedere i tetti rossi delle case e le chiome verdi degli alberi

Appartamento trasformato in camera oscura

Fonte: http://stenop.es

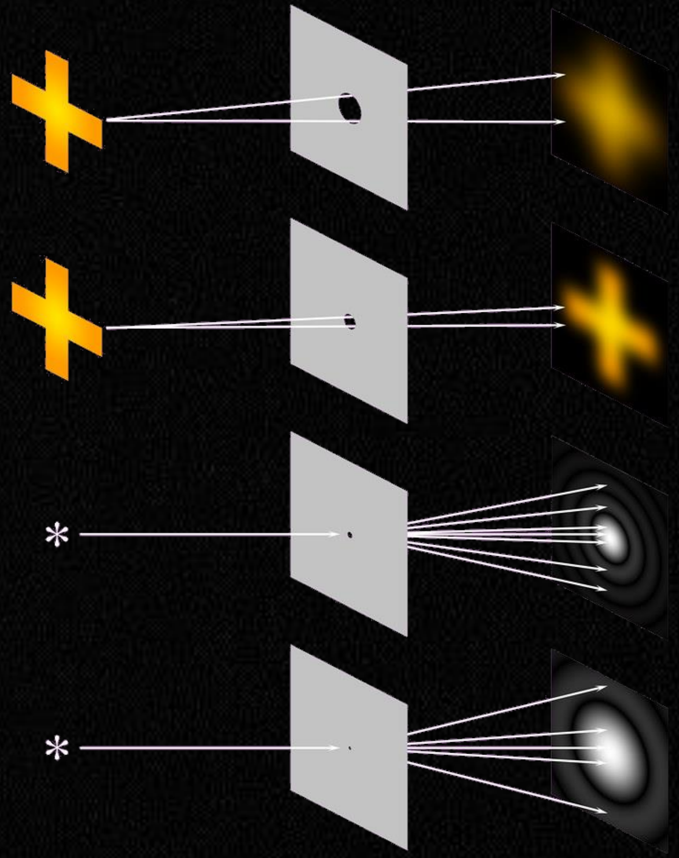
Immagine virtuale

- È equivalente (e può essere più comodo) considerare l'immagine (non rovesciata), simmetrica rispetto al foro.



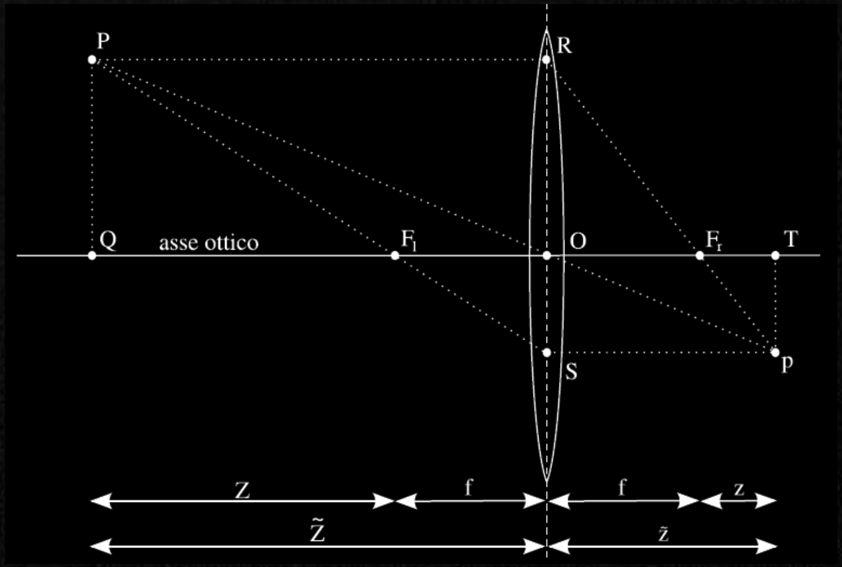
Dal foro alla lente

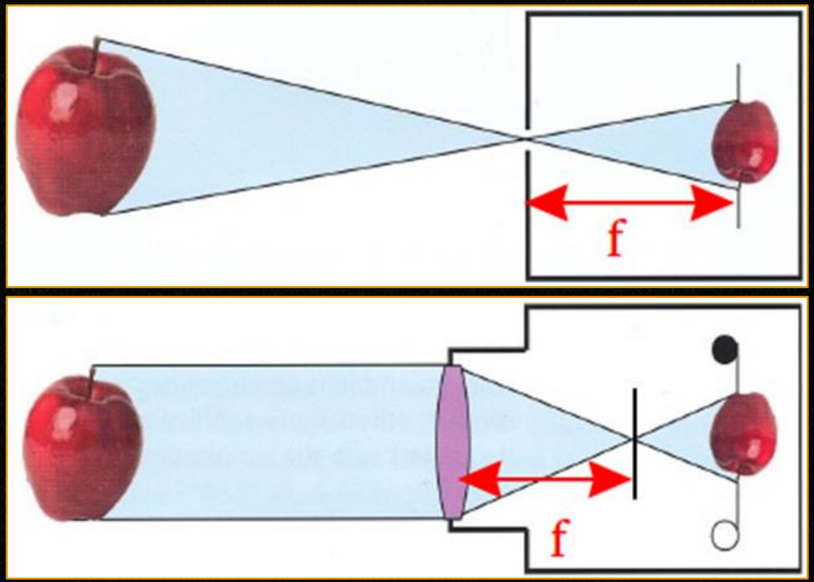
- Per ottenere un'immagine ben a fuoco, il foro dovrebbe essere molto piccolo, affinché ogni punto della scena possa influire su un solo punto del piano immagine.
- D'altra parte, per "memorizzare" l'immagine su una superficie sensibile alla luce sarebbe necessario un foro non troppo piccolo.
- Rimpicciolendo il foro:
  - Immagine più a fuoco ma 1) quantità di luce insufficiente, 2) fenomeni di diffrazione
- Allargando il foro:
  - Maggiore quantità di luce ma immagine sfocata.
- Come permettere l’ingresso di più luce, mantenendo il vincolo secondo cui ogni punto della scena deve influire solo su di un punto del piano immagine?
- La soluzione più semplice consiste nell'utilizzare una "lente sottile" (di spessore trascurabile)



Lenti sottili

- Fuochi F_l e F_r:
  - Due punti esterni al corpo della lente situati sull'asse ottico a una certa distanza dal centro O.
  - Assumiamo per semplicità che F_l e F_r abbiano la stessa distanza f da O (lunghezza focale).
- Proprietà:
  - ogni raggio di luce che entra parallelamente all'asse ottico è deviato verso l'altro fuoco;
  - ogni raggio che entra da un lato passando per il fuoco esce parallelamente all'asse ottico;
  - ogni raggio che passa dal centro della lente mantiene la sua direzione.



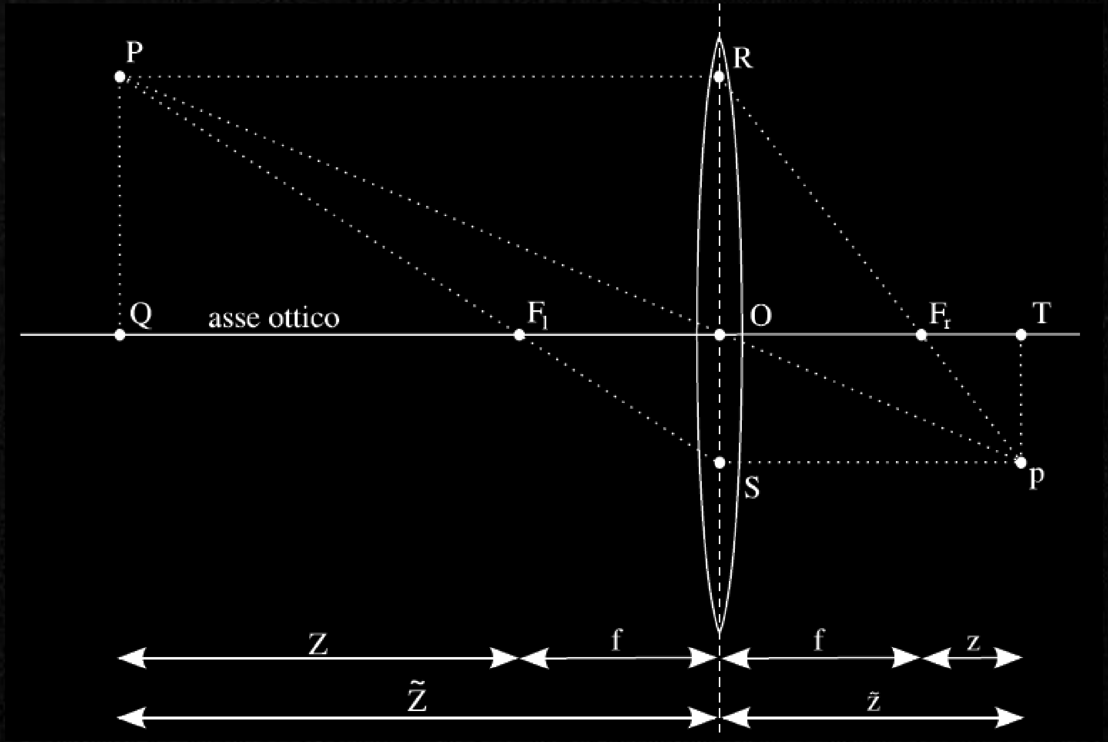


Equazione fondamentale delle lenti sottili

- Date le distanze $\tilde{Z}$ fra lente e oggetto e $\tilde{z}$ fra lente e immagine, si ha:

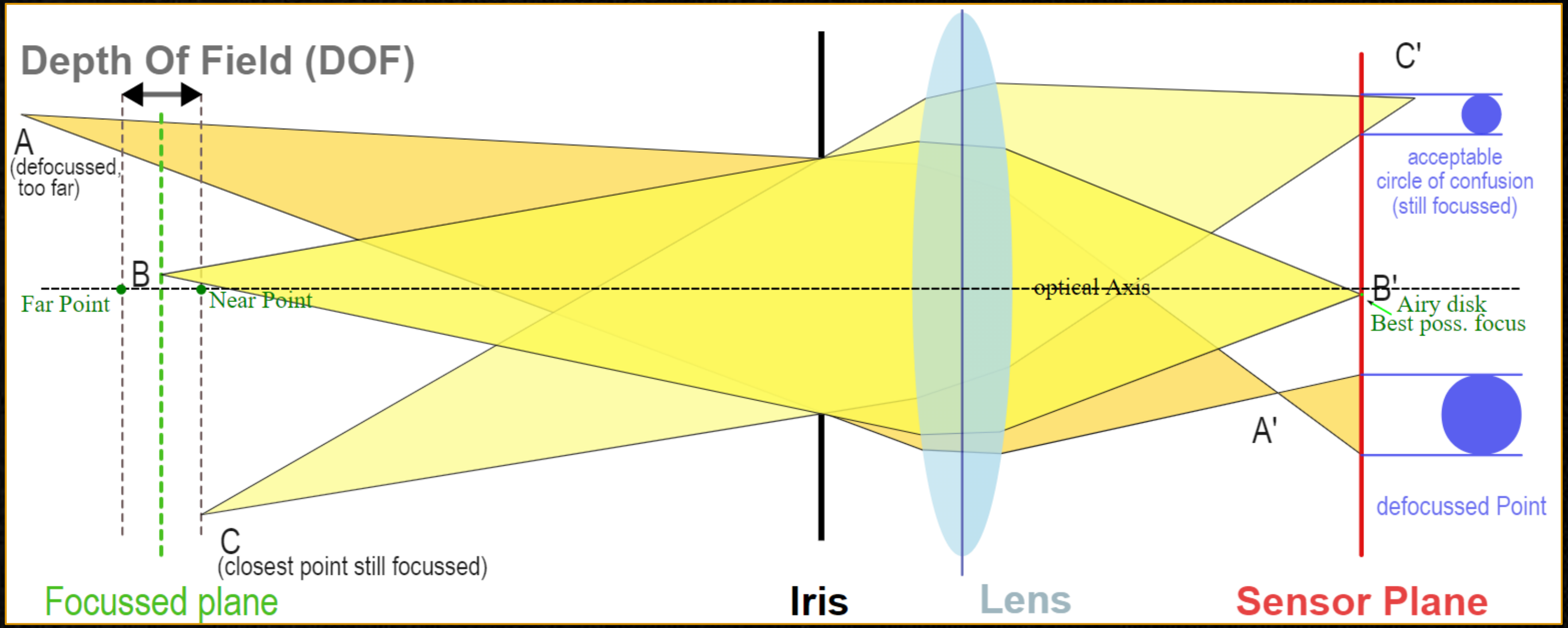
$$\frac{1}{\tilde{Z}} + \frac{1}{\tilde{z}} = \frac{1}{f}$$

- Da questo deriva che: se un oggetto è posto a distanza $\tilde{Z}$ sull'asse di una lente con lunghezza focale f, su uno schermo posto a distanza $\tilde{z}$ si formerà l'immagine dell'oggetto.



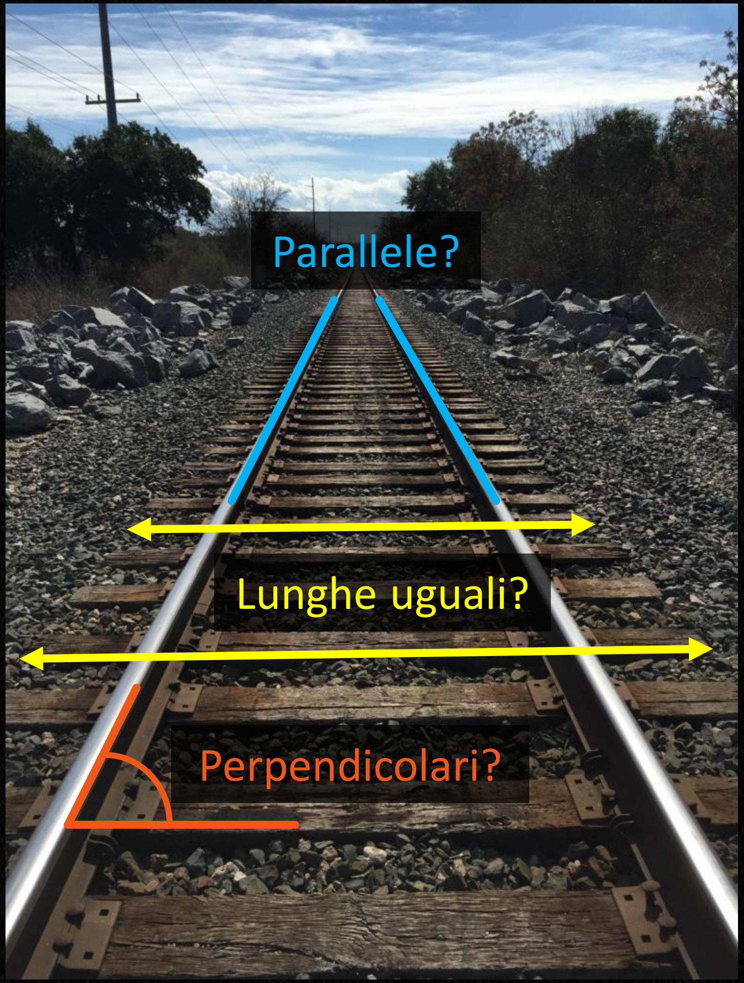
Profondità di campo (Depth Of Field)

- Con una lente non tutto è a fuoco
- Profondità di campo:
  - L'intervallo di distanze considerabili "a fuoco" nell'immagine
  - Minore è l'apertura, maggiore è la profondità di campo



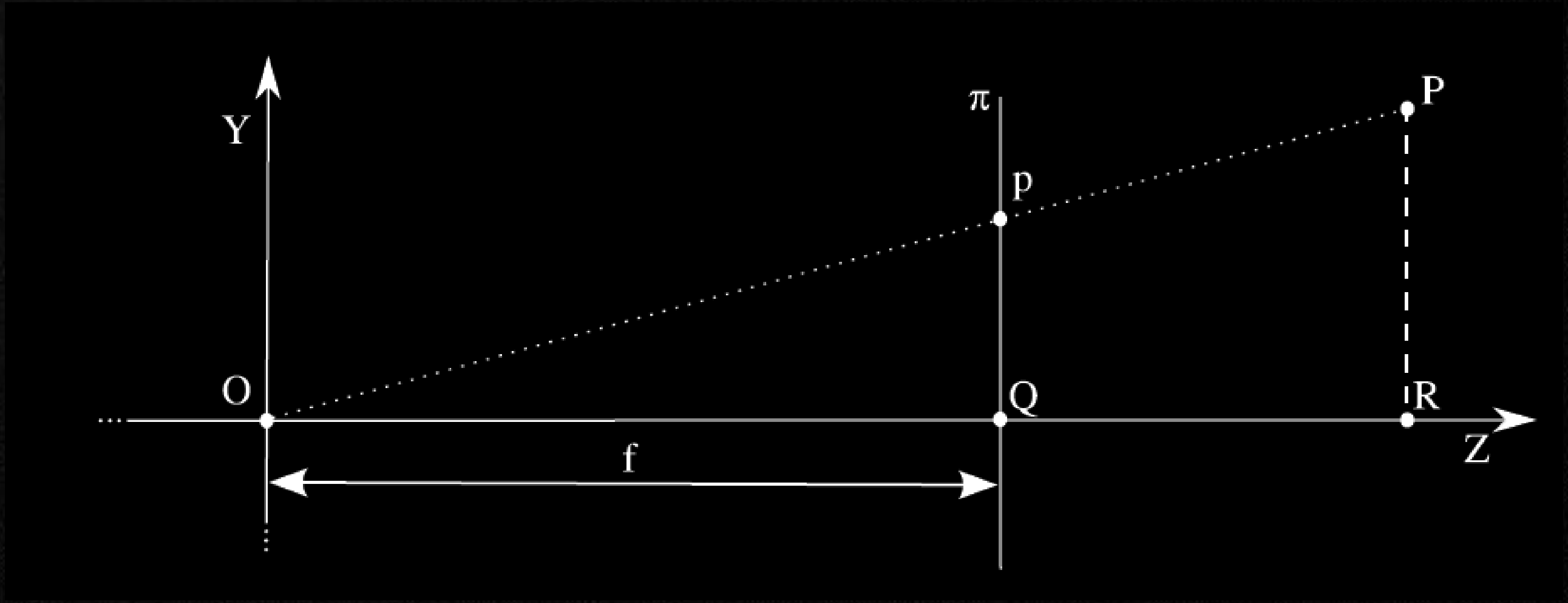
Proiezione di una scena 3D su un'immagine 2D

- Qualsiasi sia il sistema di acquisizione (camera oscura, occhio umano, fotocamera con lente sottile), l'immagine si forma attraverso la proiezione geometrica di una scena 3D su un piano bidimensionale.
- Cosa succede in questo tipo di proiezione geometrica?
  - Le dimensioni degli oggetti sono inversamente proporzionali alla distanza
  - Le linee rette rimangono tali
  - Le linee parallele divengono linee convergenti
  - Le lunghezze e gli angoli non sono preservati



Modello semplificato: proiezione prospettica

- Il sistema di riferimento della camera coincide con quello del mondo 3D:
  - il centro di proiezione o fuoco della camera coincide con l'origine O degli assi,
  - il piano immagine o piano di proiezione π è perpendicolare all'asse Z,
  - la distanza tra O e π è la lunghezza focale f.
- La figura mostra il modello in sezione laterale (lungo il piano YZ):



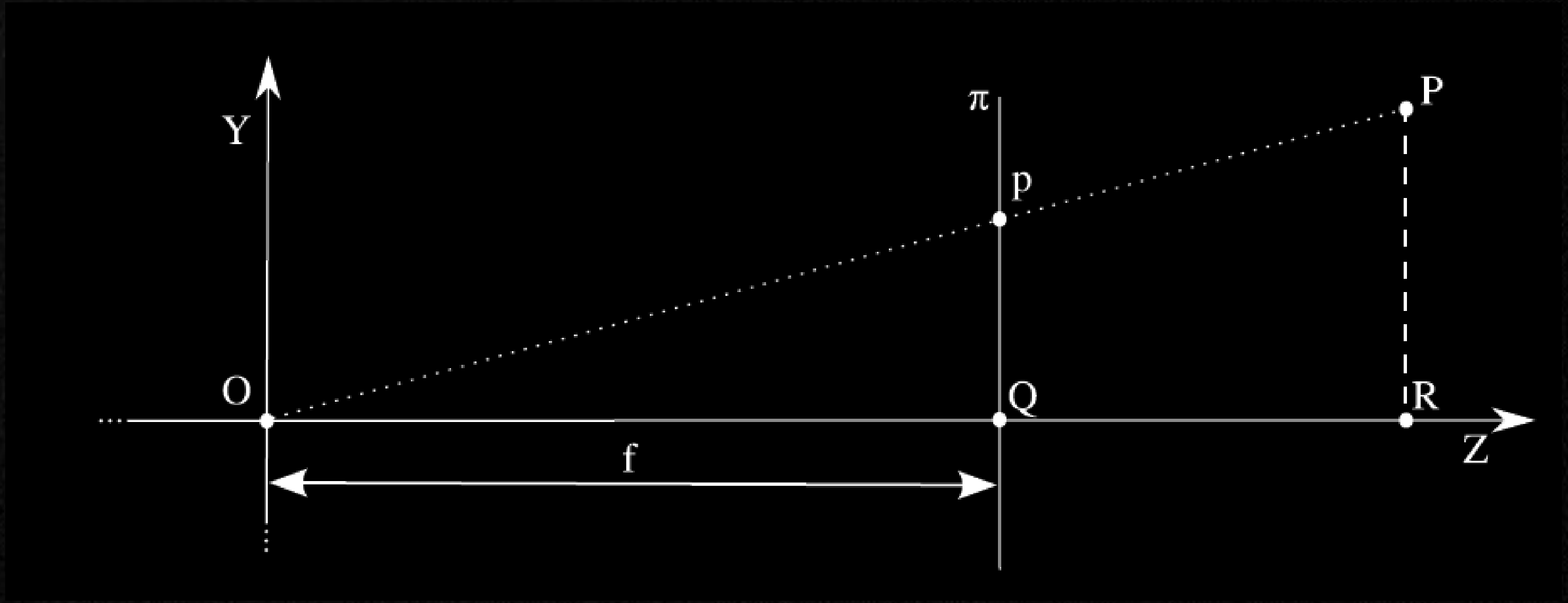
Proiezione prospettica: equazioni fondamentali

- Quale relazione lega le coordinate 3D di un punto $P = (X, Y, Z)$ nello spazio alle coordinate della sua proiezione $p = (x, y)$ sul piano immagine π?
  - Dato che i triangoli pOQ e POR sono simili, si ha:

$$\overline{pQ} : \overline{PR} = \overline{OQ} : \overline{OR} \Rightarrow y : Y = f : Z$$

  - da cui (e procedendo analogamente per x), si ottengono le due equazioni fondamentali:

$$x = f \cdot \frac{X}{Z} \qquad y = f \cdot \frac{Y}{Z}$$



Parametri estrinseci

- Il modello precedente assume che il sistema di riferimento della camera sia lo stesso della scena 3D.
- Nelle applicazioni pratiche questo di solito non avviene: è necessario trasformare le coordinate della scena 3D in quelle della camera prima di applicare le equazioni del lucido precedente.
- Sia $P_w = (X_w, Y_w, Z_w)$ un punto nelle coordinate della scena: per trasformarlo nelle coordinate $P_c = (X_c, Y_c, Z_c)$ del sistema di riferimento della camera sono necessari:
  - una matrice di rotazione R (ottenuta a partire da 3 angoli);
  - un vettore di traslazione t (3 componenti).

$$P_c = R \cdot P_w + t$$

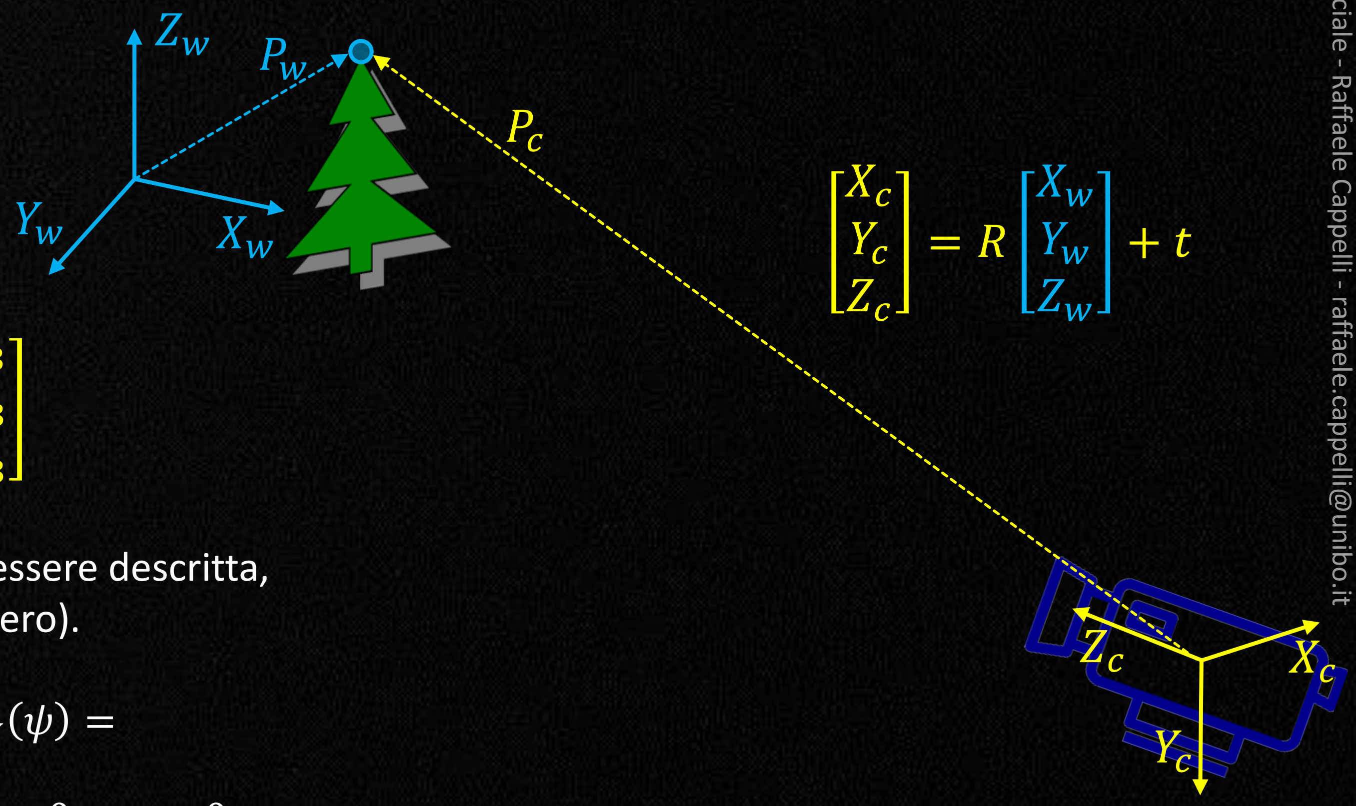
- Questi 6 parametri sono chiamati parametri estrinseci della camera:
  - Possono essere utilizzati per descrivere il movimento di una camera attorno a una scena statica, oppure il movimento di un oggetto rispetto a una camera ferma.

Modello generale: parametri estrinseci

Coordinate 3D "mondo" $P_w = (X_w, Y_w, Z_w)$

$$P_c = R \cdot P_w + t$$

Coordinate 3D "camera" $P_c = (X_c, Y_c, Z_c)$



$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + t$$

Matrice di rotazione 3D: $R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$

R ha 9 elementi, ma i gradi di libertà sono 3: può essere descritta, ad esempio con 3 angoli di rotazione (angoli di Eulero).

$$R = R(\phi, \theta, \psi) = R_Z(\phi) \cdot R_Y(\theta) \cdot R_X(\psi) =$$

$$\begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{bmatrix} =$$

Parametri intrinseci

- Quando le coordinate sono espresse nel sistema di riferimento della camera, si possono applicare le equazioni fondamentali della proiezione prospettica:

$$x = f \cdot \frac{X_c}{Z_c} \qquad y = f \cdot \frac{Y_c}{Z_c}$$

- Per determinare esattamente la trasformazione dalla scena all'immagine, sono necessari anche altri parametri, specifici della camera:
  - Lunghezza focale f (nelle telecamere è una caratteristica dell'obiettivo);
  - Coordinate del punto principale (c_x, c_y): non è detto infatti che il centro del piano immagine risieda sull'asse ottico;
  - La relazione tra la dimensione dei pixel e l'unità di misura della scena 3D (s_x, s_y), in altre parole quanto misura un pixel.
- Questi parametri, specifici della camera, sono chiamati parametri intrinseci.

Modello generale: parametri intrinseci

Coordinate 3D "mondo" $P_w = (X_w, Y_w, Z_w)$

$$P_c = R \cdot P_w + t$$

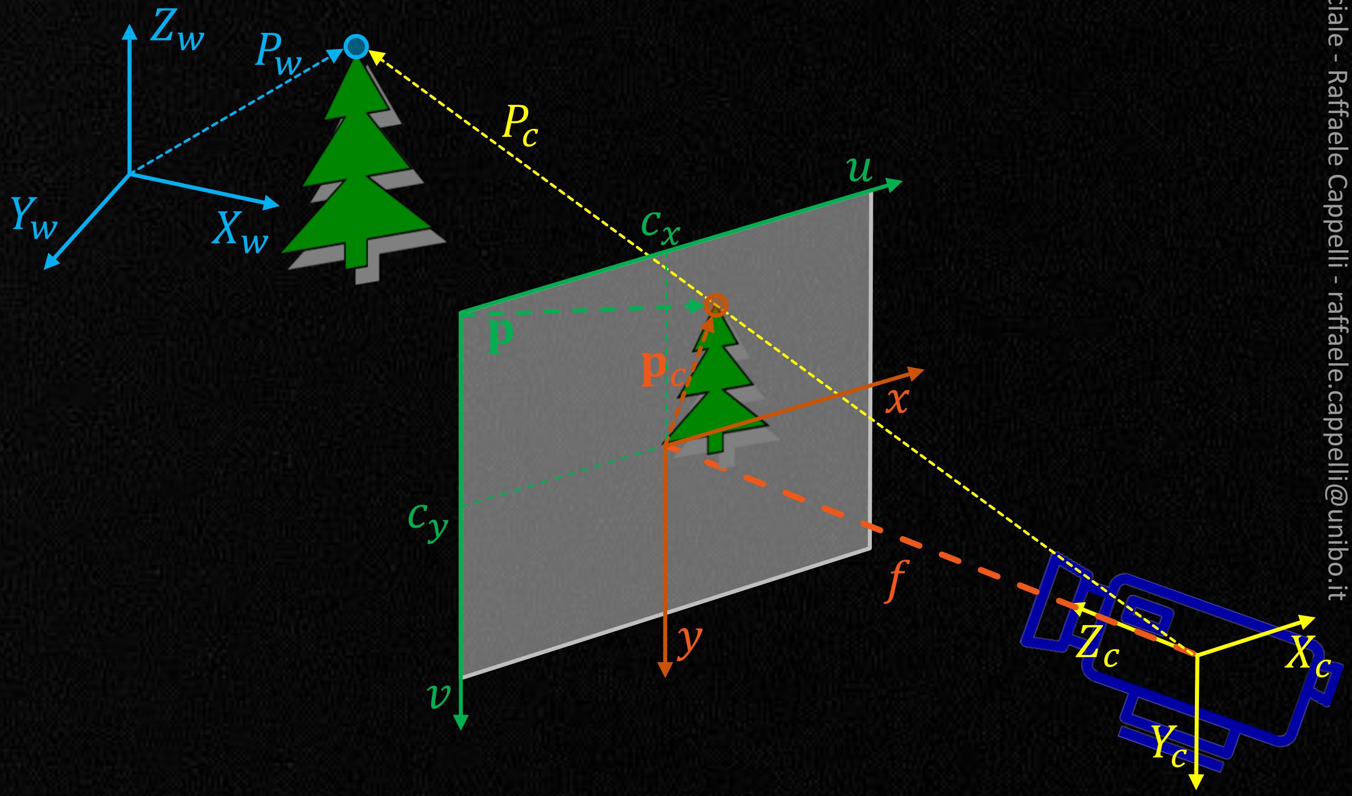
Coordinate 3D "camera" $P_c = (X_c, Y_c, Z_c)$

$$x = f \cdot {X_c}/{Z_c}$$
$$y = f \cdot {Y_c}/{Z_c}$$

Coordinate 2D "camera" $\mathbf{p}_c = (x, y)$

$$u = {x}/{s_x} + c_x$$
$$v = {y}/{s_y} + c_y$$

Coordinate 2D "immagine" $\mathbf{p} = (u, v)$



Coordinate omogenee e rototraslazione

- In coordinate omogenee, due vettori (punti nello spazio 3D) sono equivalenti se differiscono solo di un fattore si scala:

$$P = (X, Y, Z) \leftrightarrow \tilde{P} = (w \cdot X, w \cdot Y, w \cdot Z, w) = (X, Y, Z, 1), w \neq 0$$

- Rototraslazione da P_w a P_c in coordinate omogenee:

$$P_c = \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \cdot P_w + t = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ z_w \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$$

$$\begin{bmatrix} P_c \\ 1 \end{bmatrix} = \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ z_w \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} P_w \\ 1 \end{bmatrix}$$

Coordinate omogenee e proiezione prospettica

- Proiezione da $P_c = (X_c, Y_c, Z_c)$ a $\mathbf{p} = (u, v)$ in coordinate omogenee:

$$\mathbf{p} = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{f}{s_x} \cdot \frac{X_c}{Z_c} + c_x \\ \frac{f}{s_y} \cdot \frac{Y_c}{Z_c} + c_y \end{bmatrix}$$

N.B. In coordinate omogenee posso moltiplicare qualsiasi vettore per un fattore $Z_c \neq 0$

$$\begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} Z_c \cdot u \\ Z_c \cdot v \\ Z_c \end{bmatrix} = \begin{bmatrix} \frac{f}{s_x} & 0 & c_x & 0 \\ 0 & \frac{f}{s_y} & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

Modello generale

Coordinate 3D "mondo" $P_w = (X_w, Y_w, Z_w)$

$$P_c = R \cdot P_w + t$$

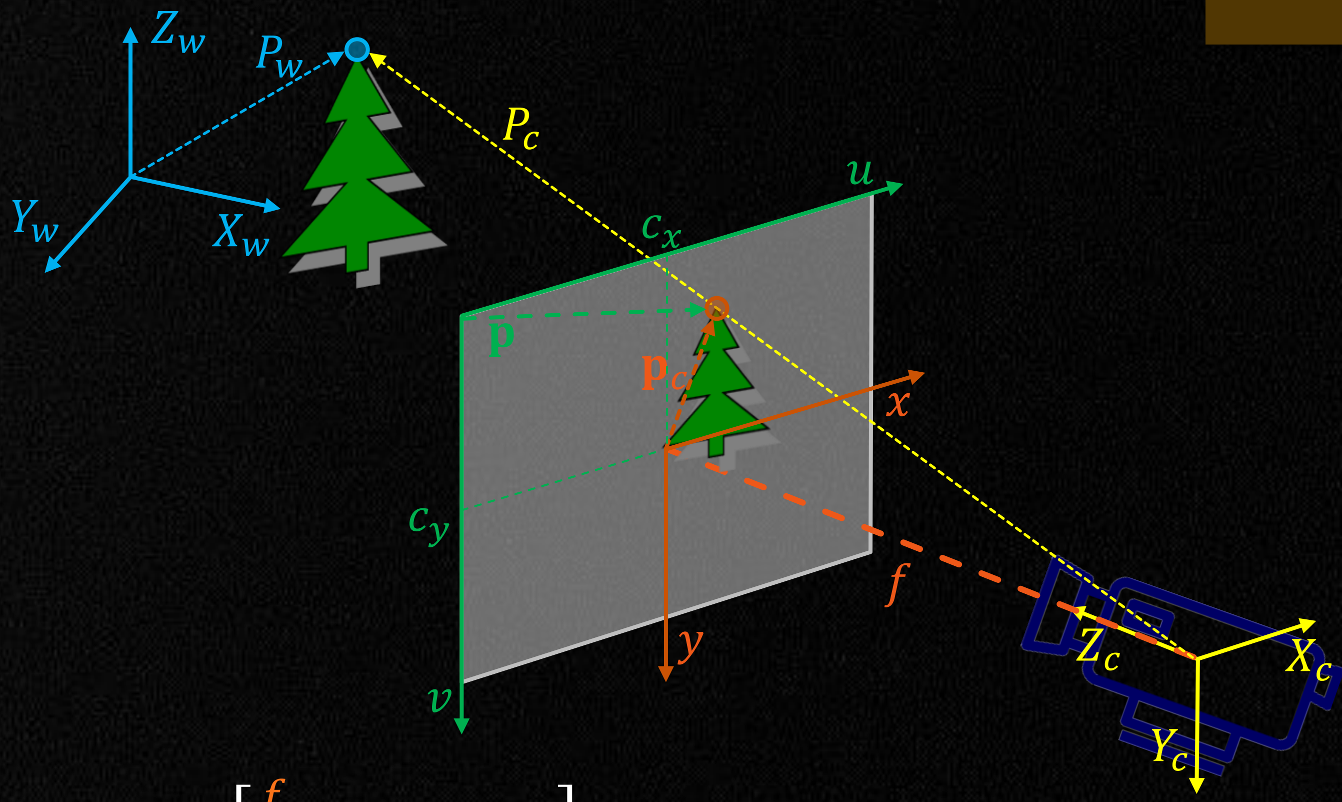
Coordinate 3D "camera" $P_c = (X_c, Y_c, Z_c)$

$$x = f \cdot {}^{X_c}/_{Z_c}$$
$$y = f \cdot {}^{Y_c}/_{Z_c}$$

Coordinate 2D "camera" $\mathbf{p}_c = (x, y)$

$$u = {}^{x}/_{s_x} + c_x$$
$$v = {}^{y}/_{s_y} + c_y$$

Coordinate 2D "immagine" $\mathbf{p} = (u, v)$



$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{s_x} & 0 & c_x & 0 \\ 0 & \frac{f}{s_y} & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

Parametri intrinseci

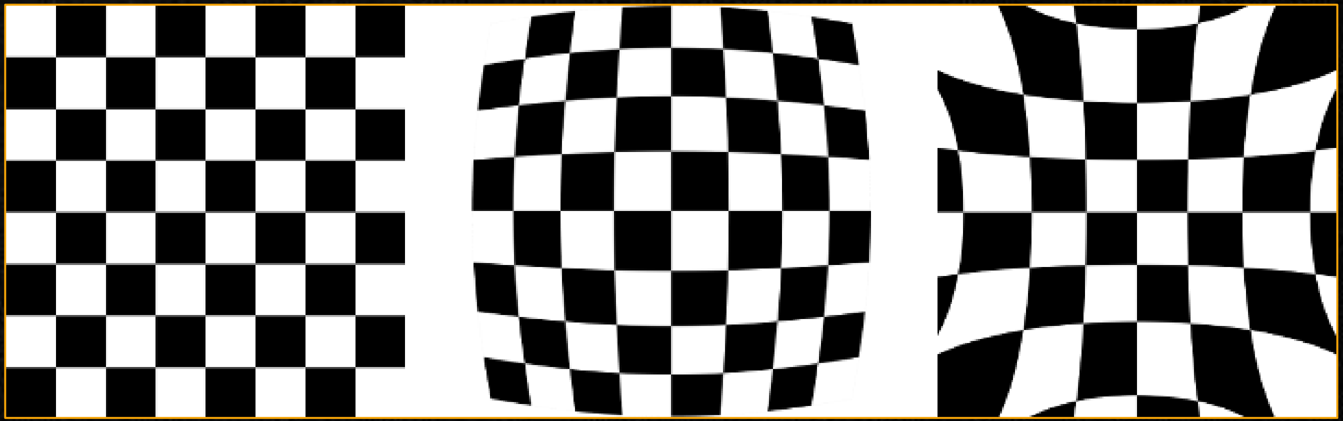
Parametri estrinseci

Distorsione radiale

- Le lenti generalmente causano una certa distorsione nell'immagine, ossia uno spostamento dei punti rispetto a quanto previsto dal modello geometrico teorico.
- Il tipo più comune di distorsione è quella radiale, che può essere modellata come segue.
  - Date le coordinate (non distorte) $\mathbf{p}_c = (x, y)$, le coordinate effettive (distorte) $\mathbf{p}'_c = (x', y')$ sono:

$$x' = x \cdot (1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$
$$y' = y \cdot (1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$

dove $r^2 = x^2 + y^2$



Nessuna distorsione $k_1 = k_2 = k_3 = 0$ | Distorsione a barilotto $k_1 > 0$ | Distorsione a cuscinetto $k_1 < 0$

Modello completo con distorsione

Coordinate 3D "mondo" $P_w = (X_w, Y_w, Z_w)$

$$P_c = R \cdot P_w + t$$

Coordinate 3D "camera" $P_c = (X_c, Y_c, Z_c)$

$$x = f \cdot {X_c}/{Z_c}$$
$$y = f \cdot {Y_c}/{Z_c}$$

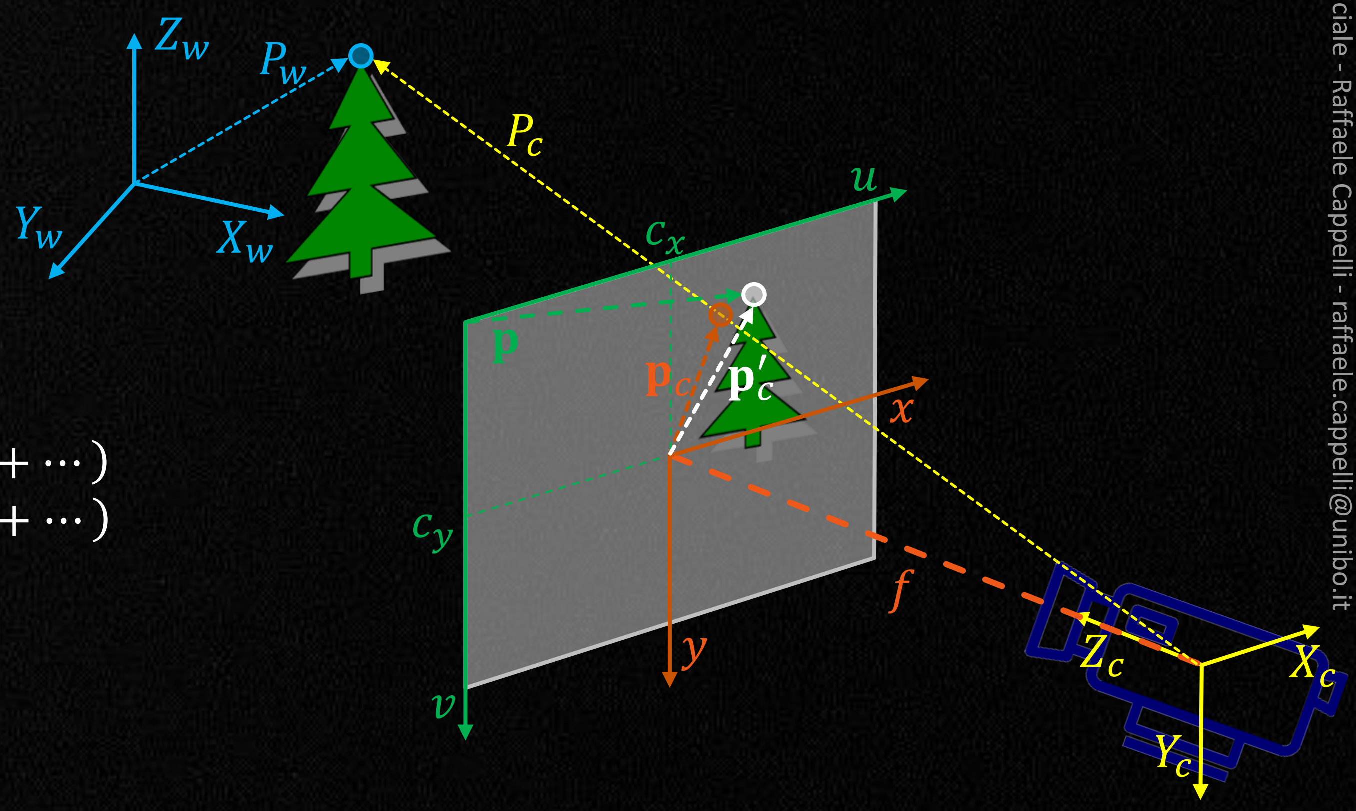
Coordinate 2D "camera" $\mathbf{p}_c = (x, y)$

$$x' = x \cdot (1 + k_1 r^2 + \cdots)$$
$$y' = y \cdot (1 + k_1 r^2 + \cdots)$$

Coordinate 2D "distorte" $\mathbf{p}'_c = (x', y')$

$$u = {x'}/{s_x} + c_x$$
$$v = {y'}/{s_y} + c_y$$

Coordinate 2D "immagine" $\mathbf{p} = (u, v)$

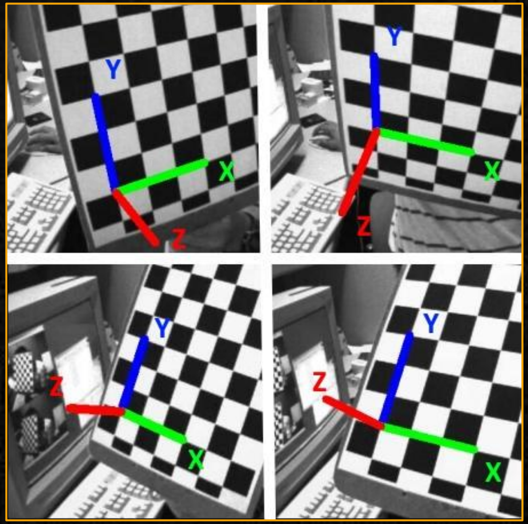


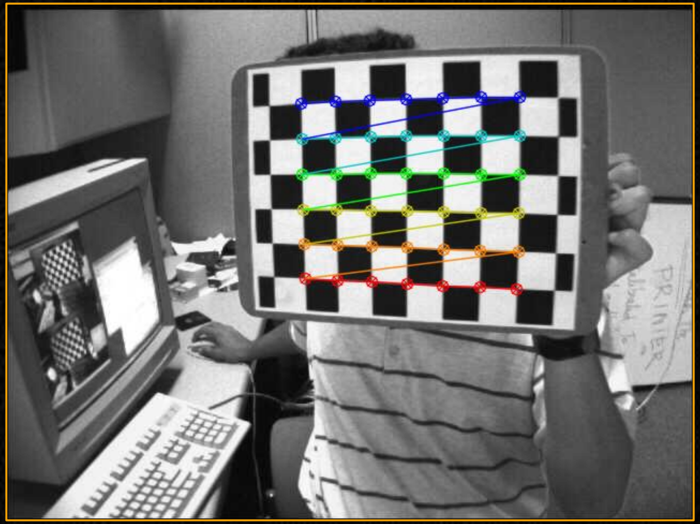
Calibrazione

- Consiste nel determinare i parametri estrinseci ed intrinseci di una camera
- Parametri estrinseci:
  - Matrice di rotazione R
  - Vettore di traslazione t
- Parametri intrinseci:
  - Lunghezza focale f;
  - Parametri del modello di deformazione: $k_1, k_2, k_3, \ldots$
  - Coordinate del punto principale (c_x, c_y);
  - Dimensioni dei pixel (s_x, s_y);
- Disponendo di un numero sufficiente di corrispondenze fra punti 3D della scena e punti 2D dell'immagine, è possibile stimare i vari parametri

Ricerca di punti corrispondenti

- Di solito avviene scegliendo un sistema di coordinate 3D per la scena e introducendo un oggetto rigido di aspetto noto su cui sia semplice individuare un insieme di m punti.
- Da una serie di n immagini con posizioni diverse (dell'oggetto o della camera), si ottengono quindi, per ogni immagine, una serie di corrispondenze $\langle P_w^i, \mathbf{p}^i \rangle$ fra coordinate nella scena 3D P_w^i e coordinate 2D $\mathbf{p}^i$.
- Il pattern più comunemente usato è una scacchiera, che consente di individuare facilmente l'insieme di punti corrispondenti ai vertici interni (i punti di contatto fra quattro caselle).

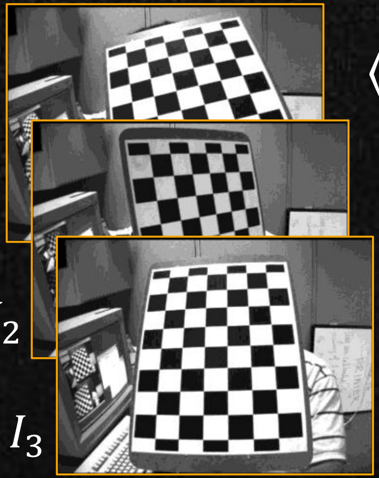




$\langle P_w^1, \mathbf{p}^1 \rangle, \langle P_w^2, \mathbf{p}^2 \rangle, \ldots, \langle P_w^m, \mathbf{p}^m \rangle$

Processo di calibrazione e risultati

I_1 I_2 I_3



$\langle P_W^1, \mathbf{p}^1\rangle, \langle P_W^2, \mathbf{p}^2\rangle, \dots, \langle P_W^m, \mathbf{p}^m\rangle$

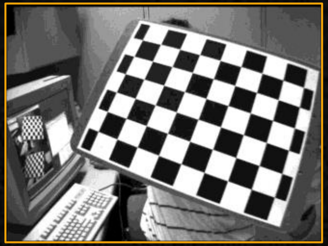
$\langle P_W^1, \mathbf{p}^1\rangle, \langle P_W^2, \mathbf{p}^2\rangle, \dots, \langle P_W^m, \mathbf{p}^m\rangle$

$\langle P_W^1, \mathbf{p}^1\rangle, \langle P_W^2, \mathbf{p}^2\rangle, \dots, \langle P_W^m, \mathbf{p}^m\rangle$

…

…

I_n



$\langle P_W^1, \mathbf{p}^1\rangle, \langle P_W^2, \mathbf{p}^2\rangle, \dots, \langle P_W^m, \mathbf{p}^m\rangle$

Calibrazione

$$\left\langle \left(\frac{f}{s_x}, \frac{f}{s_y}\right), (c_x, c_y), (k_1, k_2, k_3, \dots) \right\rangle$$

Parametri intrinseci della camera

$$\langle R^1, t^1\rangle, \langle R^2, t^2\rangle, \dots, \langle R^n, t^n\rangle$$

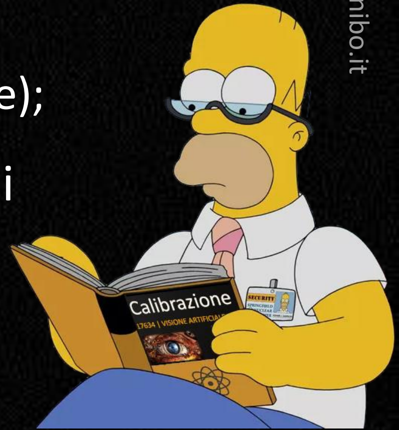
Parametri estrinseci
(per ogni immagine)

Calibrazione camera in OpenCV

- OpenCV implementa il modello geometrico introdotto nei lucidi precedenti (pinhole camera model) e offre molte funzionalità, fra cui:
  - Proiezione di punti 3D sull'immagine (conversione da P_w a $\mathbf{p}$) dati i parametri intrinseci ed estrinseci: cv.projectPoints();
  - Individuazione dei vertici interni di un pattern scacchiera su un'immagine, con possibilità di raffinare la ricerca con precisione sub-pixel e di disegnare i punti trovati: cv.findChessboardCorners(), cv.cornerSubPix(), cv.drawChessboardCorners();
  - Stima dei parametri intrinseci ed estrinseci da un insieme di corrispondenze $\langle P_w^i, \mathbf{p}^i \rangle$ individuate in una serie di immagini di un pattern di calibrazione: cv.calibrateCamera();
  - Rettifica della deformazione di un'immagine dati i parametri intrinseci della camera: cv.undistort(), cv.getOptimalNewCameraMatrix();
  - Funzioni per lavorare con matrici di rotazione 3D: cv.RQDecomp3x3(), cv.Rodrigues().

Riepilogo

- Processo di formazione delle immagini nei sistemi visivi naturali e artificiali:
  - Camera oscura e stenoscopia, immagine reale (rovesciata) e virtuale;
  - Lenti, caratteristiche di una lente sottile e sua equazione fondamentale, distorsione radiale;
  - Proiezione geometrica da scena 3D a immagine 2D: caratteristiche generali.
- Modello semplificato di proiezione prospettica:
  - Equazioni fondamentali per convertire coordinate da 3D a 2D.
- Modello generale:
  - Parametri estrinseci (rotazione e traslazione per trasformare le coordinate nel sistema di riferimento della camera);
  - Parametri intrinseci (lunghezza focale, coordinate del punto principale, misure del pixel, distorsione);
- Calibrazione a partire da corrispondenze punti 3D e 2D su una serie di viste di un pattern di calibrazione noto.



Per approfondire

- Nel libro [Szeliski, Computer Vision: Algorithms and Applications, 2022]:
  - Sezione 2.1: Geometric primitives and transformations
  - Sezione 2.2: Photometric image formation
  - Sezione 11.1: Geometric intrinsic calibration
- Documentazione OpenCV del modulo «Camera Calibration and 3D Reconstruction»:
  - https://docs.opencv.org/master/d9/d0c/group__calib3d.html