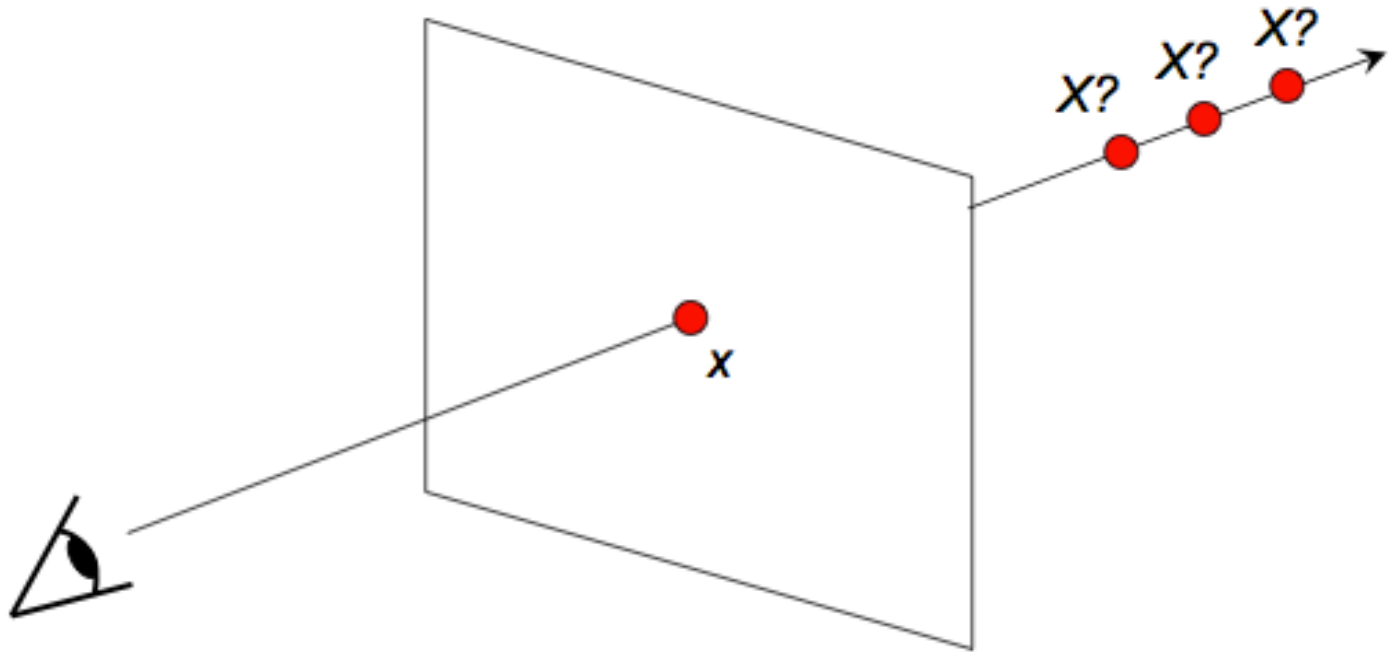
Visão Computacional
Aula 04

Calibração de Câmeras

Computer Vision - A Modern Approach
Set: Introduction to Vision
Slides by D.A. Forsyth
Adapted by Flávio Vidal

O problema?



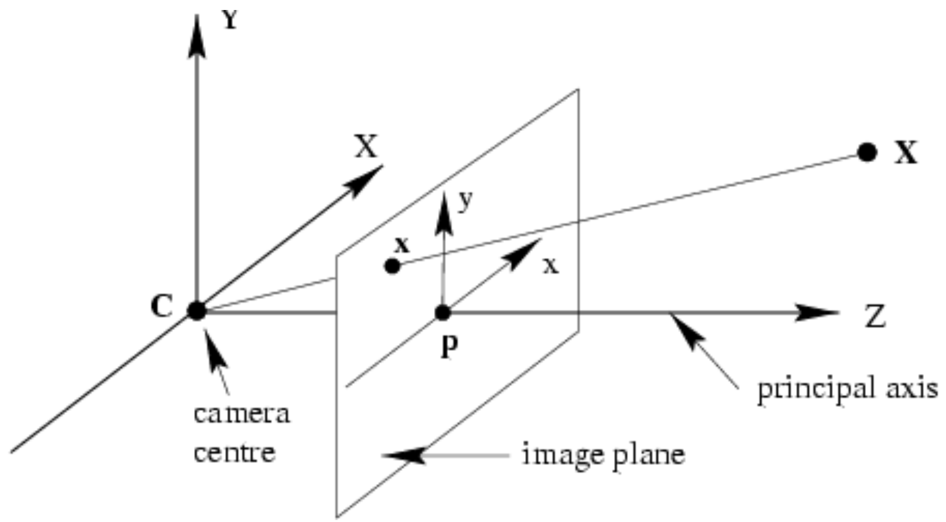
Parâmetros de câmera

- Reconstrução 3D ou cálculo da posição de objetos no espaço necessitam definir relações entre coordenadas de pontos 3D com as coordenadas 2D de imagens dos mesmos
- Alguns pressupostos devem ser assumidos:

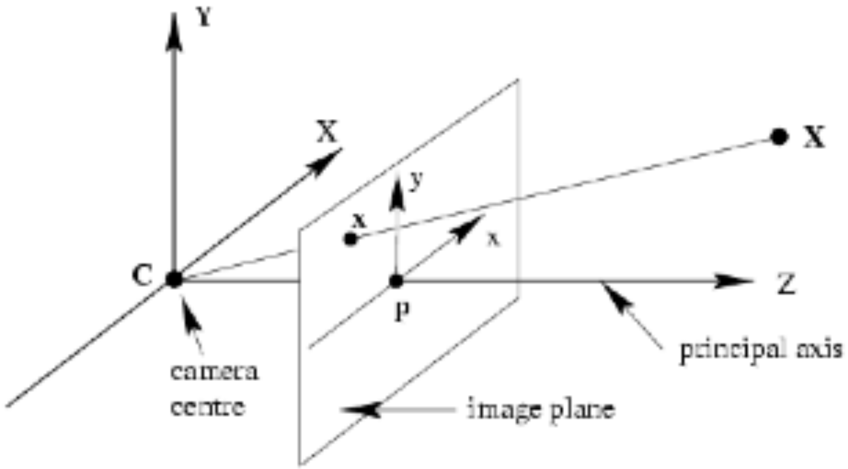
Pressupostos

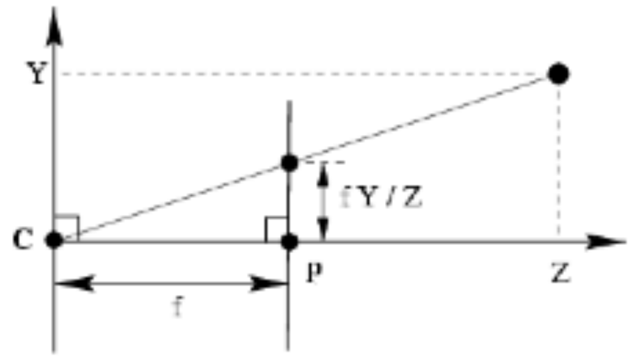
- Frame é o “Sistema de referência”
- O frame da câmera pode ser localizado em relação a algum outro frame bem conhecido (frame de mundo) - referencial assumido.
- Coordenadas das imagens de pontos no frame de câmera podem ser obtidas das coordenadas de pixels (únicas disponíveis a partir da imagem), pelo menos x e y.

Pressupostos



Modelo da Câmera Pinhole



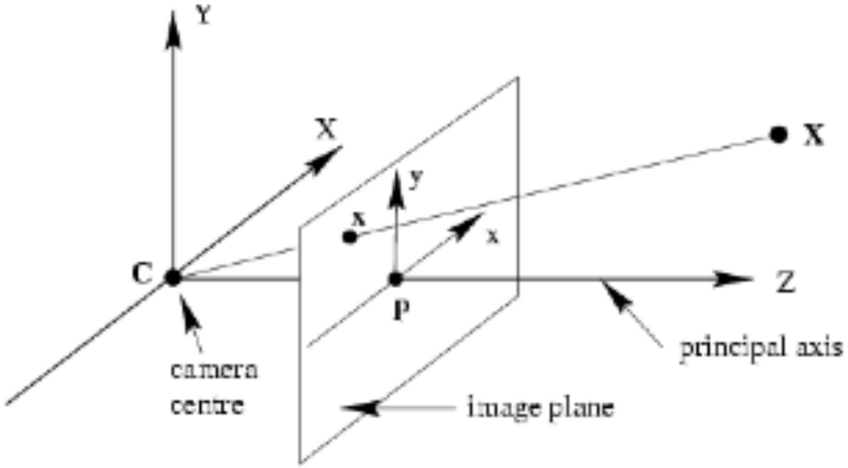


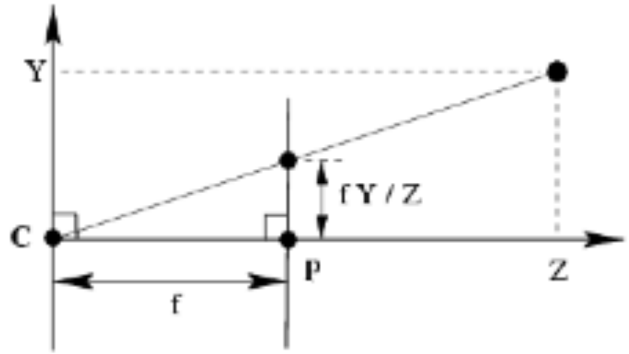
$$(X,Y,Z)\mapsto(fX/Z,fY/Z)$$

$$\begin{pmatrix}X\\Y\\Z\\1\end{pmatrix}\mapsto\begin{pmatrix}fX\\fY\\Z\end{pmatrix}=\begin{bmatrix}f & & & 0\\ & f & & 0\\ & & 1 & 0\end{bmatrix}\begin{pmatrix}X\\Y\\Z\\1\end{pmatrix}$$

$$\mathbf{x}=\mathbf{P}\mathbf{X}$$

Modelo da Câmera Pinhole





$$\begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & 0 \\ & 1 & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

$$\mathbf{x} = \mathrm{P}\mathbf{X} \qquad \mathrm{P} = \mathrm{diag}(f, f, 1)[\mathrm{I} \mid 0]$$

Parâmetros intrínsecos e extrínsecos (internos e externos)

- Parâmetros intrínsecos são os necessários para ligar as coordenadas de pixel de um ponto na imagem com as respectivas coordenadas no frame de câmera.
- Parâmetros extrínsecos são os que definem a localização e orientação do frame de câmera com relação a um frame de mundo conhecido

Parâmetros intrínsecos

- Caracterizam as propriedades óticas, geométricas e digitais da câmera visualizadora. Para pinhole, 3 conjuntos:
  - projeção perspectiva (único parâmetro é *f*)
  - transformação entre frames de câmera e píxel
  - distorção geométrica introduzida pelo sistema ótico

De câmera para pixels

- Devemos ligar (x_{im}, y_{im}), em pixels, com as coordenadas (x,y) do mesmo ponto no frame de câmera
- Neglicenciando distorções e assumindo que o CCD é uma matriz retangular:

$$x = -(x_{im} - o_x)s_x$$
$$y = -(y_{im} - o_y)s_y$$

sendo (o_x, o_y) as coordenadas em pixel do centro da imagem (ponto principal) e (s_x, s_y) o tamanho efetivo do pixel (em milímetros) horizontal e verticalmente.

Com distorção

- Com introdução de distorção:

$$x = x_d(1+k_1r^2+k_2r^4)$$
$$y = y_d(1+k_1r^2+k_2r^4)$$

sendo (x_d, y_d) as coordenadas dos pontos distorcidos e $r^2 = x_d^2+y_d^2$.

- Veja que a distorção é um deslocamento radial dos pontos na imagem. Deslocamento é zero no centro da imagem, crescendo para as bordas

Parâmetros intrínsecos - resumo

- f = distância focal
- (o_x, o_y) = localização do centro da imagem, em coordenadas de pixel
- (s_x, s_y) = tamanho efetivo horizontal e vertical do pixel
- (k_1, k_2) = coeficientes de distorção, se forem requeridos

Parâmetros extrínsecos

- Frame de câmera permite escrever equações de projeção perspectiva de uma forma simples, mas o sistema de câmera é geralmente desconhecido
- Determinar a localização e orientação do frame de câmera em relação a algum frame de referência, usando apenas informação da imagem.

Parâmetros extrínsecos

- Qualquer conjunto de parâmetros que permitem identificar unicamente a transformação entre o frame desconhecido de câmera e um frame conhecido, normalmente denominado frame de mundo.

Descrevendo a transformação

- Vetor 3D de translação, ***T***, que descreve as posições relativas das origens dos dois frames
- Uma matriz 3x3, de rotação, ***R***, a princípio ortogonal ($\boldsymbol{R}^t\boldsymbol{R}=\boldsymbol{R}\boldsymbol{R}^t$), desejado ortonormal, que traz os eixos correspondentes dos dois frames um no outro
- Ortogonalidade reduz o número de graus de liberdade para 3

Notação

- A relação entre as coordenadas de um ponto P em frame de mundo (P_w) e câmera (P_c) é dada por:

$$P_c = \boldsymbol{R}(P_w - \boldsymbol{T})$$

$$\boldsymbol{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \qquad \boldsymbol{T} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}$$

Parâmetros extrínsecos - resumo

- ***T*** = vetor de translação
- ***R*** = matriz de rotação (ou os seus parâmetros livres)
- Especificam a transformação entre o frame de câmera e o frame de mundo

Melhorando o modelo de câmera

$$\left.\begin{aligned} &P_c = \boldsymbol{R}(P_w - \boldsymbol{T}) \\ &x = -(x_{im} - o_x)s_x \\ &y = -(y_{im} - o_y)s_y \end{aligned}\right\} \longrightarrow \left\{\begin{aligned} x &= f\,(X/Z) \\ y &= f\,(Y/Z) \end{aligned}\right.$$

$$-(x_{im} - o_x)s_x = f\,[(\boldsymbol{R_1}^t(P_w - \boldsymbol{T}))/(\boldsymbol{R_3}^t(P_w - \boldsymbol{T}))]$$

$$-(y_{im} - o_y)s_y = f\,[(\boldsymbol{R_2}^t(P_w - \boldsymbol{T}))/(\boldsymbol{R_3}^t(P_w - \boldsymbol{T}))]$$

- $\boldsymbol{R_i}$, $i=1,2,3$ é um vetor 3D formado pela i-ésima coluna da matriz $\boldsymbol{R}$. Relaciona coordenadas de mundo às de imagem, usando parâmetros intrínsecos e extrínsecos

Reescrevendo como multiplicação de matrizes

- Sejam as matrizes:

$$M_{int} = \begin{pmatrix} -f/s_x & 0 & o_x \\ 0 & -f/s_y & o_y \\ 0 & 0 & 1 \end{pmatrix}$$

$$M_{ext} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \mathbf{-R_1^t T} \\ r_{21} & r_{22} & r_{23} & \mathbf{-R_2^t T} \\ r_{31} & r_{32} & r_{33} & \mathbf{-R_3^t T} \end{pmatrix}$$

Equação matricial

- M_{int} depende apenas dos parâmetros internos e M_{ext} apenas dos externos.
- Negligenciando distorção radial e expressando P_w em coordenadas homogêneas:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = M_{int} M_{ext} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

- x_1/x_3 e x_2/x_3 são as coord. de imagem x_{im} e y_{im}

Modelo de câmera perspectiva

- Assumindo, por simplicidade, que $o_x=o_y$ e $s_x=s_y$, M pode ser re-escrita como:

$$M = \begin{pmatrix} -fr_{11} & -fr_{12} & -fr_{13} & f\mathbf{R_1}^t\mathbf{T} \\ -fr_{21} & -fr_{22} & -fr_{23} & f\mathbf{R_2}^t\mathbf{T} \\ r_{31} & r_{32} & r_{33} & \mathbf{R_3}^t\mathbf{T} \end{pmatrix}$$

Modelo com perspectiva fraca

- A imagem ***p*** de um ponto ***P*** é dada por:

$$p = M \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ \mathit{1} \end{pmatrix} = \begin{pmatrix} f\mathbf{R_1}^t(\mathbf{T}\text{-}\mathbf{P}) \\ f\mathbf{R_2}^t(\mathbf{T}\text{-}\mathbf{P}) \\ \mathbf{R_3}^t(\mathbf{P}\text{-}\mathbf{T}) \end{pmatrix}$$

- Mas $\left| \mathbf{R_3}^t (\mathbf{P}\text{-}\mathbf{T}) \right|$ é a distância de ***P*** ao centro de projeção ao longo do eixo ótico.

Modelo com perspectiva fraca

- Então, a equação que aproxima a perspectiva fraca pode ser escrita como:

$$\left| \left[\boldsymbol{R_3}^t(\boldsymbol{P_i}\text{-}\boldsymbol{P'}) \right] / \left[\boldsymbol{R_3}^t(\boldsymbol{P'}\text{-}\boldsymbol{T}) \right] \right| << 1$$

- onde $\boldsymbol{P_i}$ ($i=1,2$) são pontos no espaço e $\boldsymbol{P'}$ é o centróide deles

Modelo com perspectiva fraca

- Pode-se re-escrever a equação anterior:

$$p_i \approx \begin{pmatrix} fR_1^t(T-P_i) \\ fR_2^t(T-P_i) \\ R_3^t(P'-T) \end{pmatrix}$$

- A matriz de projeção se torna:

$$M_{wp} = \begin{pmatrix} -fr_{11} & -fr_{12} & -fr_{13} & fR_1^tT \\ -fr_{21} & -fr_{22} & -fr_{23} & fR_2^tT \\ 0 & 0 & 0 & R_3^t(P'-T) \end{pmatrix}$$

O problema de calibração

- Estabelecer equações lineares no parâmetro posição de um objeto (coordenadas de mundo) que deve ser determinado numa dada cena
- Coeficientes das equações são funções específicas da posição (conhecida) da projeção do objeto no plano imagem, da geometria da câmera (intrínsecos) e de sua ótica

Equacionando o problema de calibração

- Encontrar os parâmetros anteriores significa encontrar os coeficientes de equações lineares, dadas certas posições de objetos na cena em coordenadas de mundo e suas respectivas posições na imagem.
- Transformação de corpo rígido (translação mais rotação e projeção):

$$\boxed{(x_i, y_i, z_i)} - \boxed{R} - \boxed{(\tilde{x}'_i, \tilde{y}'_i, \tilde{z}'_i)} - \boxed{T} - \boxed{(x'_i, y'_i, z'_i)} - \boxed{P} - \boxed{(X_i\, Y_i)}$$

Equacionando o problema de calibração

- Ou ainda, entendendo que a translação T e a rotação R podem ser juntadas numa única matriz:

$$(x_i, y_i, z_i) \xrightarrow{RT} (x'_i, y'_i, z'_i) \xrightarrow{P} (X_i, Y_i)$$

- A partir dos parâmetros das transformações, pode-se determinar todos os parâmetros intrínsecos e extrínsecos, bem como o inverso também vale. Ou apenas um deles!

Resolvendo o problema

- Determinar um certo número de pontos na cena de coordenadas conhecidas
- Determinar suas projeções nas imagens (coordenadas de imagens conhecidas)
- Resolver as equações, encontrando os parâmetros procurados, geralmente usando mínimos quadrados ou outro método de otimização

Importância

- Reconhecimento e reconstrução 3D com conhecimento da geometria real do objeto pode ser muito mais eficiente
- Permite localização absoluta de sistemas em relação a um frame de mundo, somente a partir de imagens de objetos na cena

Uma forma simples de entender

$$(x_i, y_i, z_i) \xrightarrow{RT} (x'_i, y'_i, z'_i) \xrightarrow{P} (X_i, Y_i)$$

- Seja (x_i, y_i, z_i) a posição inicial de um ponto p_i numa cena em coordenadas de mundo.
- Após aplicar uma rotação R e uma translação T no ponto para referenciá-lo ao sistema de coordenadas da câmera temos a posição dada em coordenadas de câmera por (x'_i, y'_i, z'_i) .

Uma forma simples de entender

- É feita então uma projeção do ponto no plano imagem, resultando nas coordenadas de imagem (X_i, Y_i).
- O conjunto de equações a seguir representa as transformações, sendo que em (1) considera-se a rotação e translação como uma transformação homogênea e em (2) elas são separadas.

Derivando as equações

$$\left(x_i', y_i', z_i', 1\right) = \begin{bmatrix} R_{11} & R_{12} & R_{13} & D_1 \\ R_{21} & R_{22} & R_{23} & D_2 \\ R_{31} & R_{32} & R_{33} & D_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \qquad (1)$$

$$\left(x_i', y_i', z_i'\right) = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} + \begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} \qquad (2)$$

Derivando as equações

$$(X_i, Y_i) = \left(f\,\frac{x'_i}{z'_i}, f\,\frac{y'_i}{z'_i} \right) \Rightarrow$$

$$X_i = f\,\frac{R_{11}x_i + R_{12}y_i + R_{13}z_i + D_1}{R_{31}x_i + R_{32}y_i + R_{33}z_i + D_3}$$

$$Y_i = f\,\frac{R_{21}x_i + R_{22}y_i + R_{23}z_i + D_2}{R_{31}x_i + R_{32}y_i + R_{33}z_i + D_3}$$

12 parâmetros

Impondo restrições

- A matriz R representa uma transformação de rotação e isto permite estabelecer a restrição de que sua inversa seja igual a sua transposta ou (ortonormalidade).

$$R^t = R^{-1}$$

$$R_{11}^2 + R_{12}^2 + R_{13}^2 = 1 \qquad R_{11}R_{21} + R_{12}R_{22} + R_{13}R_{23} = 0$$

$$R_{21}^2 + R_{22}^2 + R_{23}^2 = 1 \quad \text{e} \quad R_{11}R_{31} + R_{12}R_{32} + R_{13}R_{33} = 0$$

$$R_{31}^2 + R_{32}^2 + R_{33}^2 = 1 \qquad R_{21}R_{31} + R_{22}R_{32} + R_{23}R_{33} = 0$$

Restrições de ortonormalidade

$$R_{11}^2 + R_{21}^2 + R_{31}^2 = 1 \qquad R_{11}R_{12} + R_{21}R_{22} + R_{31}R_{32} = 0$$

$$R_{12}^2 + R_{22}^2 + R_{32}^2 = 1 \quad \text{e} \quad R_{11}R_{13} + R_{21}R_{23} + R_{31}R_{33} = 0$$

$$R_{13}^2 + R_{23}^2 + R_{33}^2 = 1 \qquad R_{12}R_{13} + R_{22}R_{23} + R_{32}R_{33} = 0$$

Ou ainda:

Duas linhas são vetores unitários, ortogonais uma à outra, enquanto que a restante é o produto cruzado destas duas (válido também para as colunas).

Próxima aula...

- Métodos de Calibração (Cont.)