

IFT 6145
Vision tridimensionnelle
Calibration I

Sébastien Roy
Département d'informatique et de recherche opérationnelle
Université de Montréal

hiver 2005

En bref...

Chapitre 6 de *Trucco & Verri*.

- Le problème de la calibration de caméra 6.1
- Estimation directe 6.2
- Estimation indirecte 6.3

Calibration de caméra

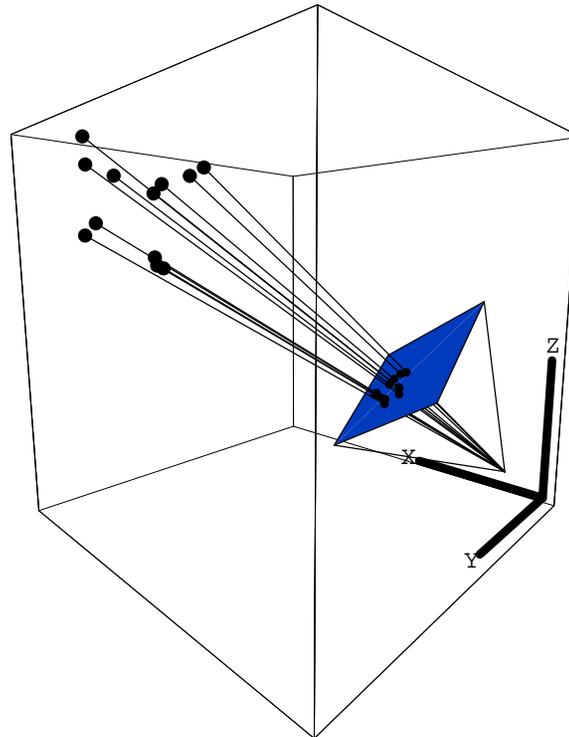
Problème :

Calculer les paramètres d'une caméra en utilisant uniquement une ou plusieurs images d'un objet de calibration.

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix}' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \mathbf{M}_{4 \times 4} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}'$$

- Quels paramètres ?
- Quel objet de calibration ?

Exemple de géométrie de caméra



Types de calibration

Modèle de caméra

- perspective
- perspective faible
- orthographique
- affine

Paramètres à déterminer

- paramètres internes et externes
- seulement les paramètres internes
- seulement les paramètres externes (paramètres internes connus)

Types de calibration

Type d'objet de calibration

- Objet de calibration complètement connu (cube)
- Objet connu en partie (lignes parallèles, surface plane, ...)
- Objet inconnu (scène arbitraire)

Les étapes de la calibration

- Connaître la position de points 3D (objet de calibration).
- Retrouver (automatiquement) la position de ces points dans l'image.
- Résoudre le système d'équation à partir des correspondances.

Modèles de caméra

Perspective

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix}' = \mathbf{JART} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}'$$

Affine

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix}' = \mathbf{J} \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}'$$

Modèles de caméra

Perspective faible ($\alpha = \bar{Z}$)

Orthographique ($\alpha = 1$)

$$\begin{bmatrix} u \\ v \\ h \end{bmatrix}' = \mathbf{JA} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \alpha \\ 0 & 0 & 1 & 0 \end{pmatrix} \mathbf{RT} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}'$$

Ces modèles se calibrent différemment....

Paramètres internes \mathbf{A}

Paramètres : Centre de l'image (o_x, o_y) , taille d'un pixel (s_x, s_y) et distance focale f .

$$\mathbf{A} = \begin{pmatrix} -\frac{f}{s_x} & 0 & o_x & 0 \\ 0 & -\frac{f}{s_y} & o_y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -f_x & 0 & o_x & 0 \\ 0 & -f_y & o_y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Plus généralement, simple transformation affine de l'image :

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Paramètres internes A

Si on décide de réduire la taille de l'image d'un facteur 2 (en x et en y), par quoi doit-on multiplier la matrice A pour obtenir la nouvelle matrice pour l'image réduite ?

$$\mathbf{B} = \begin{pmatrix} ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \\ ? & ? & ? & ? \end{pmatrix} \mathbf{A}$$

$$\mathbf{B} = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathbf{A} \Rightarrow \mathbf{A} = \begin{pmatrix} -\frac{1}{2}f_x & 0 & \frac{1}{2}O_x & 0 \\ 0 & -\frac{1}{2}f_y & \frac{1}{2}O_y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Paramètres externes \mathbf{R} et \mathbf{T}

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

avec la propriété d'orthonormalité $\mathbf{R}^\top \mathbf{R} = \mathbf{I}$
et déterminant $|\mathbf{R}| = 1$.

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & T_x \\ 0 & 1 & 0 & T_y \\ 0 & 0 & 1 & T_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Estimation directe

Points 2D et 3D \rightarrow modèle perspective

À l'aide d'une liste de points correspondants entre l'image $\mathbf{p}_i = (x_i, y_i)$ et la scène $\mathbf{P}_i = (X_i, Y_i, Z_i)$, estimer les paramètres internes et externes directement à partir de l'équation de projection de \mathbf{P} vers \mathbf{p} .

Pour une caméra perspective, on a

$$\begin{bmatrix} x_i h_i \\ y_i h_i \\ h_i \end{bmatrix}' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \mathbf{ART} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}'$$

qui devient, après expansion...

$$x_i - o_x = -f_x \frac{r_{11}X_i + r_{12}Y_i + r_{13}Z_i + T_x}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + T_z} \quad (1)$$

$$y_i - o_y = -f_y \frac{r_{21}X_i + r_{22}Y_i + r_{23}Z_i + T_y}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + T_z} \quad (2)$$

Paramètres connus

- points images (x_i, y_i)
- points de la scènes (X_i, Y_i, Z_i)
- centre de l'image (o_x, o_y) (!!!)

Paramètres à estimer

- distance focale f_x, f_y
- rotation $r_{11}, r_{12}, \dots, r_{33}$
- translation T_x, T_y, T_z

Pour la suite, x_i et y_i représenteront $(x_i - o_x)$ et $(y_i - o_y)$.

Solution détaillée

On remarque que (1) et (2) ont le même dénominateur, donc on peut dire

$$x_i f_y (r_{21} X_i + r_{22} Y_i + r_{23} Z_i + T_y) = y_i f_x (r_{11} X_i + r_{12} Y_i + r_{13} Z_i + T_x)$$

En posant $\alpha = f_x / f_y$, ceci correspond à une équation linéaire à 8 inconnues (v_1, \dots, v_8) de la forme

$$\mathbf{l}_i^\top \mathbf{v} = 0$$

avec

$$\begin{aligned} \mathbf{l}_i^\top &= \left[x_i X_i \quad x_i Y_i \quad x_i Z_i \quad x_i \quad -y_i X_i \quad -y_i Y_i \quad -y_i Z_i \quad -y_i \right] \\ \mathbf{v}^\top &= \left[v_1 \quad v_2 \quad v_3 \quad v_4 \quad v_5 \quad v_6 \quad v_7 \quad v_8 \right] \end{aligned}$$

où on définit

$$\begin{aligned}v_1 &= r_{21} & v_5 &= \alpha r_{11} \\v_2 &= r_{22} & v_6 &= \alpha r_{12} \\v_3 &= r_{23} & v_7 &= \alpha r_{13} \\v_4 &= T_y & v_8 &= \alpha T_x\end{aligned}$$

En prenant N points, on obtient le système linéaire

$$\begin{bmatrix} \mathbf{l}_1^\top \\ \mathbf{l}_2^\top \\ \dots \\ \mathbf{l}_N^\top \end{bmatrix} \mathbf{v} = 0$$

- Si $N \geq 7$ et les points ne sont pas coplanaires
→ solution unique (à un facteur d'échelle près)
- Résolu par la décomposition SVD

(Système linéaire homogène)

Comment résoudre le système

$$\mathbf{A}\mathbf{x} = 0$$

1. Calculer la **décomposition en valeurs singulières** (*SVD*) de \mathbf{A} telle que

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\top}$$

où \mathbf{D} est une matrice diagonale (contenant les valeurs singulières), et \mathbf{U} , \mathbf{V} sont des matrices orthonormales.

2. La solution est le vecteur colonne de \mathbf{V} correspondant à la plus petite valeur singulière de \mathbf{D} .

Idéalement, cette valeur est l'unique zéro.

Avec du bruit, c'est plus grand que zéro.

Solution détaillée (suite)

Soit notre solution obtenue $\bar{\mathbf{v}}$

$$\bar{\mathbf{v}} = \gamma(r_{21}, r_{22}, r_{23}, T_y, \alpha r_{11}, \alpha r_{12}, \alpha r_{13}, \alpha T_x)$$

Puisque

$$r_{11}^2 + r_{12}^2 + r_{13}^2 = 1 \quad \text{et} \quad r_{21}^2 + r_{22}^2 + r_{23}^2 = 1$$

on peut isoler les facteurs $|\gamma|$ et α .

- On a donc les deux premières rangées de \mathbf{R} (à un signe près).
- La troisième rangée de \mathbf{R} est le produit vectoriel des deux premières.
- Les deux premières rangées de \mathbf{R} ne sont pas vraiment orthogonales.
→ Il faut faire subir un *massage orthonormal* à \mathbf{R} .

(Massage orthonormal)

Comment trouver la matrice $\hat{\mathbf{R}}$ orthonormale la plus *proche* d'une matrice \mathbf{R} qui n'est pas orthonormale ?

1. Décomposer \mathbf{R} par *SVD* tel que

$$\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{V}^{\top}$$

2. Remplacer \mathbf{D} par la matrice identité \mathbf{I}

$$\hat{\mathbf{R}} = \mathbf{U}\mathbf{I}\mathbf{V}^{\top}$$

Solution : γ

On veut déterminer le signe de $|\gamma|$.

Puisque les points dans la scène sont situés *devant* la caméra, on doit avoir

$$x_i(r_{11}X_i + r_{12}Y_i + r_{13}Z_i + T_x) < 0$$

Si ce n'est pas le cas, on inverse le signe des deux premières rangées de $\hat{\mathbf{R}}$ ainsi que les signes de T_x et T_y .

Solution : T_z et f_x

On veut déterminer T_z et f_x

En partant des équations (1) et (2), on peut obtenir le système

$$x_i(r_{31}X_i + r_{32}Y_i + r_{33}Z_i + T_z) = -f_x(r_{11}X_i + r_{12}Y_i + r_{13}Z_i + T_x)$$

ou encore

$$\mathbf{A} \begin{bmatrix} T_z \\ f_x \end{bmatrix} = \mathbf{b}$$

avec

$$\mathbf{A} = \begin{bmatrix} x_1 & (r_{11}X_1 + r_{12}Y_1 + r_{13}Z_1 + T_x) \\ x_2 & (r_{11}X_2 + r_{12}Y_2 + r_{13}Z_2 + T_x) \\ \vdots & \vdots \\ x_N & (r_{11}X_N + r_{12}Y_N + r_{13}Z_N + T_x) \end{bmatrix}$$

et

$$\mathbf{b} = \begin{bmatrix} -x_1(r_{31}X_1 + r_{32}Y_1 + r_{33}Z_1) \\ -x_2(r_{31}X_2 + r_{32}Y_2 + r_{33}Z_2) \\ \vdots \\ -x_N(r_{31}X_N + r_{32}Y_N + r_{33}Z_N) \end{bmatrix}$$

qui est un système surdéterminé résolu par moindres carrés

$$\begin{bmatrix} \hat{T}_z \\ \hat{f}_x \end{bmatrix} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$$

ou avec SVD bien sûr.

Et voilà !

Solution détaillée (suite)

Il faut estimer le centre de l'image (o_x, o_y) ...

Trucco & Verri suggère d'utiliser les points de fuite (*vanishing lines*) de lignes parallèles. Voir algorithme IMAGE_CENTER_CAL.

En pratique, on utilise une valeur *raisonnable* de (o_x, o_y) .
Exemple : Si la taille d'image est (x_s, y_s) , utiliser $(o_x, o_y) = (\frac{x_s}{2}, \frac{y_s}{2})$.

Estimation indirecte

Points images et 3D \rightarrow modèle affine \rightarrow modèle perspective

1. Soit une liste de points correspondants entre l'image $\mathbf{p}_i = (x_i, y_i)$ et la scène $\mathbf{P}_i = (X_i, Y_i, Z_i)$.
2. Estimer les paramètres d'un modèle plus général que celui qui nous intéresse.
3. Convertir les paramètres du modèle plus général vers ceux du modèle qui nous intéresse.

Typiquement, résoudre le modèle **affine**, puis convertir en modèle **perspective**.

Avantage par rapport à l'estimation directe ? Hummm...

Résoudre le modèle affine

On doit résoudre le modèle

$$\begin{bmatrix} x_i h_i \\ y_i h_i \\ h_i \end{bmatrix}' = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}'$$

qui s'exprime aussi sous la forme

$$x_i = \frac{m_{11}X_i + m_{12}Y_i + m_{13}Z_i + m_{14}}{m_{31}X_i + m_{32}Y_i + m_{33}Z_i + m_{34}}$$
$$y_i = \frac{m_{21}X_i + m_{22}Y_i + m_{23}Z_i + m_{24}}{m_{31}X_i + m_{32}Y_i + m_{33}Z_i + m_{34}}$$

On définit le système linéaire homogène

$$\mathbf{A}\mathbf{m} = 0, \quad \mathbf{A} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{v}_1 \\ \mathbf{u}_2 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{u}_N \\ \mathbf{v}_N \end{bmatrix} \quad \mathbf{m} = \begin{bmatrix} m_{11} \\ m_{12} \\ \vdots \\ m_{34} \end{bmatrix}$$

avec

$$\begin{aligned} \mathbf{u}_i &= \begin{bmatrix} X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -x_i X_i & -x_i Y_i & -x_i Z_i & -x_i \end{bmatrix} \\ \mathbf{v}_i &= \begin{bmatrix} 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -y_i X_i & -y_i Y_i & -y_i Z_i & -y_i \end{bmatrix} \end{aligned}$$

On résoud par la décomposition *SVD* de \mathbf{A} telle que $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ et en choisissant la colonne de \mathbf{V} correspondant à la plus petite valeur singulière.

Conversion Affine \rightarrow Perspective

Dans le modèle perspective, la matrice affine \mathbf{M} s'exprime comme

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

Soit $\hat{\mathbf{M}}$ la matrice affine obtenue. On définit les vecteurs

$$\begin{aligned} \mathbf{q}_1 &= \begin{bmatrix} \hat{m}_{11} & \hat{m}_{12} & \hat{m}_{13} \end{bmatrix}^\top \\ \mathbf{q}_2 &= \begin{bmatrix} \hat{m}_{21} & \hat{m}_{22} & \hat{m}_{23} \end{bmatrix}^\top \\ \mathbf{q}_3 &= \begin{bmatrix} \hat{m}_{31} & \hat{m}_{32} & \hat{m}_{33} \end{bmatrix}^\top \\ \mathbf{q}_4 &= \begin{bmatrix} \hat{m}_{14} & \hat{m}_{24} & \hat{m}_{34} \end{bmatrix}^\top \end{aligned} \quad \text{donc} \quad \hat{\mathbf{M}} = \left(\begin{array}{ccc|c} \mathbf{q}_1^\top & & & \\ \mathbf{q}_2^\top & & & \\ \mathbf{q}_3^\top & & & \\ \hline & & & \mathbf{q}_4 \end{array} \right)$$

Puisque $\hat{\mathbf{M}}$ est définie à un facteur d'échelle près, nous avons

$$\hat{\mathbf{M}} = \gamma \mathbf{M}$$

On peut obtenir γ en remarquant que

$$\sqrt{m_{31}^2 + m_{32}^2 + m_{33}^2} = |\gamma| \sqrt{r_{31}^2 + r_{32}^2 + r_{33}^2} = |\gamma|$$

À partir d'ici, on suppose que $\hat{\mathbf{M}}$ est normalisée, donc

$$\hat{\mathbf{M}} = \sigma \mathbf{M}$$

où $\sigma = \pm 1$. Ce signe sera déterminé plus tard.

On peut maintenant extraire la plupart des paramètres

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

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

$$o_x = \mathbf{q}_1^\top \mathbf{q}_3$$

$$o_y = \mathbf{q}_2^\top \mathbf{q}_3$$

$$f_x = \sqrt{\mathbf{q}_1^\top \mathbf{q}_1 - o_x^2}$$

$$f_y = \sqrt{\mathbf{q}_2^\top \mathbf{q}_2 - o_y^2}$$

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

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

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

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

Note

La rotation $\hat{\mathbf{R}}$ extraite de $\hat{\mathbf{M}}$ est approximative.
Elle requiert un *massage orthonormal*.

Le signe σ

On détermine σ en utilisant le fait que les points (X_i, Y_i, Z_i) sont toujours *devant* la caméra.

Et voilà !

Autres types de calibration

Plan + parallaxe

Points images sur un même plan 3D.

Points de fuite (*vanishing points*)

Points images à l'intersection de droites 3D parallèles.

Géométrie épipolaire (matrices essentielles et fondamentales)

Points correspondants entre deux images.

Point d'expansion (*focus of expansion*)

Point de convergence du déplacement entre deux images.

Méthode du triangle

Solution analytique lorsque les paramètres internes sont connus.