

## Approximation du changement de focale et de mise au point par une transformation affine à trois paramètres

---

### *A Three-Parameter Affine Approximation of Focus and Zoom Variations*

par R. ENCISO, T. VIÉVILLE, O. FAUGERAS

Projet Robotvis, INRIA Sophia, 06902 Sophia-Antipolis, email : renciso@sophia.inria.fr

#### Résumé

Dans ce travail, on approxime dynamiquement le changement de focale et de la mise au point en suivant deux points quelconques entre deux images prises par la même caméra.

Plus précisément, on étudie la variation des paramètres *intrinsèques* lors d'un changement de la mise au point et du zoom. On démontre grâce à cette analyse, qu'un modèle de transformation affine à 3 paramètres est tout à fait suffisant, et qu'un modèle de transformation affine général ne se justifie pas, car la transformation à utiliser n'a – mathématiquement – que 5 paramètres et non 6 tandis que l'analyse physique du système montre que 3 paramètres suffisent. Expérimentalement, le modèle est justifié, car la précision est meilleure que 1.5 pixel pour des variations de la mise au point et dans tous les cas meilleure que si l'on utilise le modèle général.

Par une méthode des moindres carrés on accède à des équations très simples qui nous permettent d'obtenir une précision dépendant du nombre des points suivis.

**Mots clés :** Calibration des paramètres intrinsèques, Zoom, Mise au Point

#### Abstract

*This work aims to develop a dynamic approximation to the variation in a lens' intrinsic parameters when the zoom and focus parameters are modified. This approximation is built by tracking two generic points in a monocular image sequence.*

*Our preliminary analysis demonstrates that a particular three-parameter affine model is sufficient to describe these modifications. The general affine model is not acceptable on a mathematical or physical level : the mathematical transformation to be used has only five parameters, instead of six, and an analysis of the physics reveals that three parameters are sufficient. Experimentally, this approximation is entirely valid, with the precision being better than 1.5 pixels in almost every case. Using a least-squares method, we obtain very simple equations in which the precision of the estimate increases with the number of available correspondences.*

**Key words :** Intrinsic Parameters Calibration, Zoom, Focus

## 1. Introduction

Le problème de la calibration d'un capteur visuel statique est un sujet bien étudié [13, 14]. Une revue de ces méthodes est donnée dans [15] et une présentation générale dans [4], nous n'y reviendrons pas.

Deux écoles principales se sont développées :

- celle de la **calibration "forte"**, où l'on calcule à partir d'un objet bien connu les paramètres *intrinsèques* (l'angle entre les deux axes du repère caméra  $\theta$ , les facteurs d'échelle  $\alpha_u, \alpha_v$ , et le point principal intersection de l'axe optique et le plan

rétinien  $(u_0, v_0)$ ), et les paramètres *extrinsèques* (la translation et rotation du capteur par rapport au système robotique),

- celle de l'**auto-calibration** d'un système robotique où la même tâche est accomplie sans aucune connaissance "a priori" des objets observés et sans une connaissance complète du mouvement du système [5, 11, 16].

Dans le domaine de la Vision Active la calibration "forte" n'est pas envisageable, car les paramètres du système visuel varient en permanence et qu'il n'est pas facile de maintenir toujours un objet connu dans le champ visuel.

En se basant sur un modèle projectif de la géométrie du système il est possible de considérer une notion de calibration "faible", construite à partir de la connaissance de la géométrie épipolaire

et de sa représentation sous forme de *matrice fondamentale* [5, 11, 10] qui nous permet de reconstruire un monde projectif.

De manière plus simple, dans le cas d'un système mécanique au comportement reproductible, on peut essayer de maintenir à jour une **matrice de projection** de la caméra quels que soient les paramètres du système. C'est le but de ce travail, où l'on va approximer dynamiquement le changement de mise au point et de focale en suivant deux points quelconques entre deux images prises par la même caméra.

Plus précisément, nous allons étudier la variation des paramètres **intrinsèques** lors d'un changement de mise au point ou de focale. Nous démontrons grâce à cette analyse qu'un modèle de transformation affine à 3 paramètres est tout à fait suffisant et qu'un modèle de transformation affine général à 6 paramètres ne se justifie pas, car la précision est meilleure que 1.5 pixels dans le cas de la mise au point et meilleure que 5 pixels pour la focale.

Par une méthode des moindres carrés on déduit des équations très simples qui nous permettent d'obtenir une précision augmentant avec le nombre des points suivis.

## 2. Modèle sténopé de la caméra.

### 2.1. VALIDITÉ DU MODÈLE STÉNOPE LORS D'UNE VARIATION DE FOCALE OU DE MISE AU POINT

Dans le domaine de la Vision par ordinateur le modèle sténopé de la caméra, décrit dans la subsection suivante, a été largement utilisé car c'est un modèle très simple, et donc adapté aux besoins des tâches comme la Vision Active ou la planification de trajectoires en mouvement.

Dans le cas d'une caméra munie d'un zoom et d'un mécanisme de mise au point on ne modélise pas chaque lentille de l'objectif, tâche qui peut-être extrêmement compliqué, sinon le tout comme une "boîte noire" qui nous rend en sortie une distance focale lié au zoom, et une distance de mise au point. Il est tout à fait clair que les quantités liées à la mise au point et à la focale d'un objectif sont parfaitement mesurables. Néanmoins, dans le cas d'un système robotique, les méthodes proposées par les physiciens ne sont pas faciles à mettre en place sur un système "embarqué". D'où cette volonté de fournir des modèles de calibration qui agissent comme des "boîtes noires".

On note que le modèle sténopé peut être utilisé dans le cas d'un objectif à lentille munie d'un zoom ([6]), (au lieu du modèle épais), si la distance de l'objet à l'image peut être considérée comme *étant variable et si l'on reste dans un voisinage restreint de la rétine*. C'est le cas de notre étude, car les points utilisés pour la calibration appartiennent à la mire qui n'occupe qu'une partie de l'image, donc la métrique estimée sera locale à une région de la rétine, et locale à un objet puisqu'on calcule les coordonnées

du centre optique dans le repère du monde attaché à la mire, donc adaptée aux différentes positions de celle-là.

On conseille fortement la lecture de l'article [7] apparu dans cette même revue sur l'utilisation du modèle sténopé.

### 2.2. MODÈLE STÉNOPE

On utilise le *modèle sténopé* pour la caméra. On suppose que la caméra réalise une transformation perspective parfaite de centre **C** (le centre optique de la caméra) à une distance  $f$  (distance focale) du plan rétinien.

Les éléments du modèle, représentés Figure 1 [10], sont :

- un plan  $\mathcal{R}$  : la *rétine*
- un point **C** de l'espace n'appartenant pas à  $\mathcal{R}$  : la *centre optique*
- une *distance focale*  $f$ , distance entre le point **C** et le plan  $\mathcal{R}$ .

La projection **m** d'un point de l'espace **M** est la trace du *rayon optique* **CM** sur la rétine.

Le *plan focal*  $\mathcal{F}$  est le plan parallèle au plan rétinien et contenant le centre optique.

On définit le *repère de la caméra* comme il suit : il a pour origine le centre optique **C**, et pour axes  $(X_c, Y_c, Z_c)$ . Les deux premiers vecteurs  $(X_c, Y_c)$  constituent une base orthonormée du *plan focal* et le troisième vecteur  $Z_c$ , orthogonal au plan rétinien, définit l'*axe optique*.

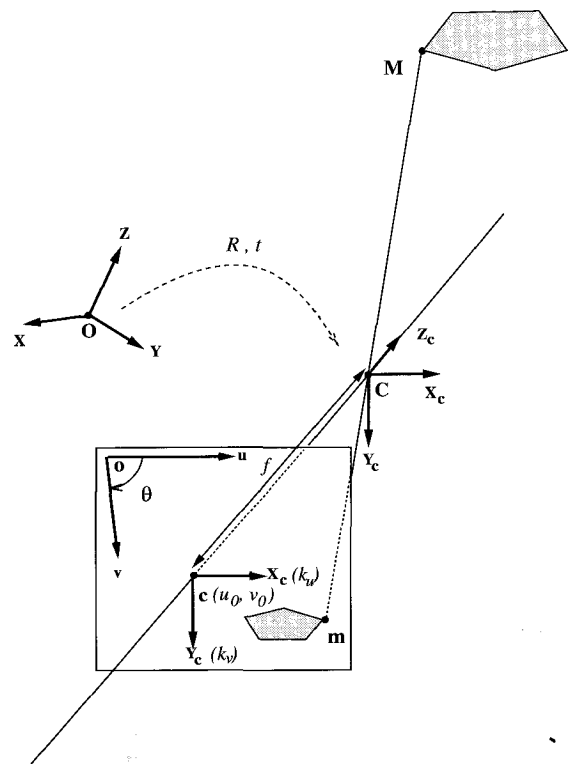


Figure 1. – Le modèle sténopé de la caméra.

Si de plus, les unités du repère de la caméra sont choisies de sorte que la distance focale  $f$  soit égale à 1 dans ce repère, nous dirons que ce dernier est *normalisé*.

En adoptant le *repère normalisé du plan rétinien*, qui est le repère centré au *point principal*  $\mathbf{c}$ , intersection de l'axe optique avec le plan rétinien, et ayant pour axes  $(X_c, Y_c)$ , la relation entre les coordonnées tridimensionnelles  $X, Y, Z$  d'un point  $\mathbf{M}$  et les coordonnées bidimensionnelles  $u, v$  de son image  $\mathbf{m}$  par la projection perspective est :

$$\frac{1}{Z} = \frac{u}{X} = \frac{v}{Y} \quad (1)$$

En utilisant les coordonnées projectives pour l'espace et pour le plan rétinien, l'équation (1) peut se réécrire :

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\mathbf{P}} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (2)$$

La matrice  $\mathbf{P}$  est dite *matrice de projection perspective*. La relation entre les coordonnées projectives de l'espace et celles du plan image est ici linéaire. Elle apparaît ici sous sa forme la plus simple.

Il est simple de montrer que cette propriété demeure vérifiée quels que soient les repères rétinien et repères du monde choisis.

### 2.3. LES PARAMÈTRES INTRINSÈQUES ET EXTRINSÈQUES DE LA CAMÉRA

Dans le cas le plus général la **matrice de projection perspective**  $\mathbf{P}$   $3 \times 4$ , qui lie les coordonnées 3D aux coordonnées pixelliques, peut se décomposer de la manière suivante [4] (à condition de satisfaire la contrainte  $||[p_{31}p_{32}p_{33}]|| \neq 0$ ), de manière unique sous la forme :

$$\mathbf{P} = \mathbf{A} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \mathbf{D}$$

où  $\mathbf{A}$  est la matrice des **paramètres intrinsèques** et  $\mathbf{D}$  la matrice des **paramètres extrinsèques** de la forme :

$$\mathbf{D} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix}$$

Ici,  $\mathbf{t}$  est un vecteur  $3 \times 1$ ,  $\mathbf{0}$  est le vecteur nul  $1 \times 3$ , et  $\mathbf{R}$  est une matrice  $3 \times 3$  telle que  $\mathbf{R} \cdot \mathbf{R}^t = \mathbf{I}$ .

Dans cette décomposition,  $\mathbf{A}$  est une matrice  $3 \times 3$  triangulaire supérieure de la forme :

$$\mathbf{A} = \begin{pmatrix} \alpha_u & -\alpha_u \cot \theta & u_0 \\ 0 & \frac{\alpha_v}{\sin \theta} & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

faisant passer du repère rétinien normalisé (coordonnées normalisées  $\mathbf{m}$ ) au repère image (coordonnées pixelliques  $\mathbf{p}$ ), à travers la relation :

$$\mathbf{p} = \mathbf{A} \cdot \mathbf{m} \quad (3)$$

On peut décrire ainsi les 5 **paramètres intrinsèques** :

- $\alpha_u, \alpha_v$  sont les facteurs d'échelle sur les axes.

Ils sont reliés à la distance focale en mètres  $f$  par un facteur de proportionnalité :

$$\begin{aligned} \alpha_u &= k_u \cdot f \\ \alpha_v &= k_v \cdot f \end{aligned}$$

- $(u_0, v_0)$  est le point principal intersection de l'axe optique et du plan image.

- $\theta$  est l'angle formé par les deux axes du repère image  $(u, v)$ .

## 3. Influence du changement de focale et de mise au point sur la matrice de projection

On choisit d'étudier le problème dans un cas très particulier, en faisant les 2 hypothèses suivantes :

- **Le changement de mise au point ou de focale revient à une transformation affine** effectuée dans le plan image.
- On s'intéresse seulement à la variation des paramètres **intrinsèques**, sachant que le centre optique suit un déplacement au long de l'axe optique lors de la variation de focale [8].

La première hypothèse nous amène à introduire une **matrice de transformation affine** en coordonnées projectives :  $\mathbf{A}_{aff}$ .

La nouvelle **matrice de projection perspective**  $\mathbf{P}'$  s'écrit alors :

$$\mathbf{P}' = \mathbf{A}_{aff} \cdot \mathbf{P} + \mathbf{t}_c$$

où  $\mathbf{t}_c$  est le vecteur translation qui représente la translation suivi par le centre optique le long de l'axe optique lors d'un changement de focale (selon [7] pour un zoom de haute qualité le changement de focale est réalisé par une translation pure de deux blocs mobiles de lentilles. Pour les zoom standard cette translation est couplée à une rotation). Donc :

$$\mathbf{P}' = \mathbf{A}' \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \mathbf{D} + \mathbf{t}_c$$

Dans la mesure où nous ne nous intéressons qu'à la variation des paramètres **intrinsèques**, nous n'analyserons pas les variations de  $\mathbf{t}_c$ . On pose :

$$\mathbf{A}' = \mathbf{A}_{aff} \mathbf{A}$$

(où  $\mathbf{A}'$  représente la nouvelle matrice des paramètres intrinsèques).

Il vient alors la relation suivante :

$$\mathbf{A}' = \mathbf{A}_{aff} \mathbf{A} = \begin{pmatrix} m_{11} & m_{12} & a_0 \\ m_{21} & m_{22} & b_0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \alpha_u & -\alpha_u \cot \theta & u_0 \\ 0 & \frac{\alpha_v}{\sin \theta} & v_0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \alpha'_u & -\alpha'_u \cot \theta' & u'_0 \\ 0 & \frac{\alpha'_v}{\sin \theta'} & v'_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

L'identification des coefficients et quelques calculs simples conduisent au système suivant **dans le cas général** :

$$\begin{cases} m_{11} & = & \frac{f'}{f} \cdot \frac{k'_u}{k_u} \\ m_{12} & = & \frac{f'}{f} \cdot \frac{k'_u}{k_v} \\ (cot \theta - cot \theta') \cdot \sin \theta & = & 0 \\ m_{21} & = & 0 \\ m_{22} & = & \frac{f'}{f} \cdot \frac{k'_v}{k_v} \cdot \frac{\sin \theta}{\sin \theta'} \\ m_{11} \cdot u_0 + m_{12} \cdot v_0 + a_0 & = & u'_0 \\ m_{22} \cdot v_0 + b_0 & = & v'_0 \end{cases}$$

où l'on voit que  $m_{21} = 0$ , et que l'on peut calculer les nouveaux paramètres intrinsèques en connaissant  $\mathbf{A}_{aff}$  et les anciens paramètres.

Nos transformations auront donc forcément la forme suivante :

$$\mathbf{A}_{aff} = \begin{pmatrix} m_{11} & m_{12} & a_0 \\ 0 & m_{22} & b_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Elles engendrent un **sous-groupe des matrices de transformation affine**, qui est l'intersection des matrices triangulaires supérieures et des matrices de transformation affine. Ce sous-groupe préserve la direction de l'axe des  $X$ .

## 4. Étude de l'invariance de $k_u, k_v$ , et $\theta$

Les paramètres  $k_u, k_v$  étant des paramètres liés uniquement à la mécanique de la lentille, il paraît raisonnable de les considérer comme invariants lors d'un changement de mise au point. De plus,  $\theta$  est en pratique très proche de  $\frac{\pi}{2}$ , donc on peut négliger ses variations ([10]). On étudie dans cette section la validité de ces hypothèses.

### 4.1. INVARIANCE DE L'ANGLE $\theta$

La matrice  $\mathbf{A}_{aff}$  **dans le cas général** est :

$$\mathbf{A}_{aff} = \begin{pmatrix} \frac{\alpha'_u}{\alpha_u} & \frac{\alpha'_u}{\alpha_v} \cdot (cot \theta - cot \theta') \cdot \sin \theta & a_0 \\ 0 & \frac{\alpha'_v}{\alpha_v} \cdot \frac{\sin \theta}{\sin \theta'} & b_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Étudions pour une petite variation  $d\theta$  de l'angle relatif entre les deux axes, comment se transforment les équations.

Soit l'angle entre les deux axes sur la première image  $\theta$ . On note  $\theta = \frac{\pi}{2} + \epsilon$ .

Et soit l'angle sur la deuxième image  $\theta' = \frac{\pi}{2} + \epsilon + d\epsilon$ . Donc :

$$\frac{\sin \theta}{\sin \theta'} = \frac{\sin(\frac{\pi}{2} + \epsilon)}{\sin(\frac{\pi}{2} + \epsilon + d\epsilon)} = \frac{\cos \epsilon}{\cos(\epsilon + d\epsilon)} = \frac{1}{\cos(d\epsilon) - \tan(\epsilon) \cdot \sin(d\epsilon)}$$

Par développements de Taylor on trouve :

$$\frac{\sin \theta}{\sin \theta'} \approx 1 + \frac{d\epsilon}{2} + \epsilon \cdot d\epsilon + \Phi$$

où  $\Phi$  est d'ordre 3 en  $d\epsilon$  et  $\epsilon$ .

D'un autre côté,

$$(cot \theta - cot \theta') \cdot \sin \theta \approx -\epsilon \cdot d\epsilon - \epsilon \cdot d\epsilon + \Phi$$

avec  $\Phi$  d'ordre 3 en  $d\epsilon$  et  $\epsilon$ .

Il paraît raisonnable de conclure  $m_{22} = \frac{\alpha'_v}{\alpha_v}$  et  $m_{12} = 0$  dès que  $d\epsilon$  ou  $\epsilon$  est petit.

Si l'on fait donc l'hypothèse de l'invariance de  $\theta$  on obtient :

$$\mathbf{A}_{aff} = \begin{pmatrix} \frac{\alpha'_u}{\alpha_u} & 0 & a_0 \\ 0 & \frac{\alpha'_v}{\alpha_v} & b_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

avec 4 inconnues.

### 4.2. INVARIANCE DE $k_u, k_v$

Faisons maintenant une nouvelle hypothèse, à savoir que  $k_u, k_v$  (paramètres liés à la mécanique de la lentille et au système d'acquisition) ne sont pas modifiés lors du changement de mise au point.

Puis finalement, si l'on appelle  $C_0 = \frac{f'}{f}$  on obtient :

$$\mathbf{A}_{aff} = \begin{pmatrix} C_0 & C_0 \cdot \frac{k_u}{k_v} \cdot (cot \theta - cot \theta') \cdot \sin \theta & a_0 \\ 0 & C_0 \cdot \frac{\sin \theta}{\sin \theta'} & b_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

### 4.3. INVARIANCE DE $k_u, k_v$ ET L'ANGLE $\theta$

Si l'on se place dans le cas particulier où ces paramètres sont invariants :

$$\mathbf{A}_{aff} = \begin{pmatrix} C_0 & 0 & a_0 \\ 0 & C_0 & b_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Deux remarques :

- Il s'agit d'une homothétie plus une translation, donc elle conserve les angles.
- On peut voir que l'on avait au départ 13 inconnues, et 6 équations. Le fait de supposer l'invariance de  $k_u, k_v$  et  $\theta$  a fourni 5 relations supplémentaires, puisque  $\theta = \theta'$  nous donne  $\cot\theta = \cot\theta'$  et  $\sin\theta = \sin\theta'$ . Il nous reste donc  $13-6-5 = 2$  indéterminées, que l'on retrouve bien exprimées sous la forme des 2 relations suivantes :

$$\begin{cases} C_0 \cdot u_0 + a_0 = u'_0 \\ C_0 \cdot v_0 + b_0 = v'_0 \end{cases}$$

### 4.4. CALCUL DE LA TRANSFORMATION AVEC HYPOTHÈSES D'INVARIANCE EN EFFECTUANT LE SUIVI DE 2 POINTS

Puisqu'il ne reste plus que 3 inconnues :  $C_0, a_0, b_0$ , et que chaque point fournit 2 équations, il suffit de suivre deux points.

On considère deux points en coordonnées projectives  $P_1 = (u_1, v_1, \xi)$  et  $P_2 = (u_2, v_2, \xi)$  où  $\xi \in \{0, 1\}$ . On a  $\xi = 1$  pour les points à distance finie et  $\xi = 0$  pour les points à l'infini, c'est à dire les directions de droite. Notre calcul sera donc valable pour des points et des points à l'infini.

Soient  $P'_1, P'_2$  leurs correspondants sur la deuxième image. Donc :

$$\begin{aligned} \mathbf{A}_{aff} \cdot P_1 &= P'_1, \\ \mathbf{A}_{aff} \cdot P_2 &= P'_2 \end{aligned}$$

nous donne les 4 équations suivantes :

$$\begin{cases} u'_1 = C_0 \cdot u_1 + a_0 \cdot \xi \\ v'_1 = C_0 \cdot v_1 + b_0 \cdot \xi \\ u'_2 = C_0 \cdot u_2 + a_0 \cdot \xi \\ v'_2 = C_0 \cdot v_2 + b_0 \cdot \xi \end{cases}$$

En soustrayant la troisième équation à la première, ainsi que la quatrième à la deuxième, il vient :

$$C_0 = \frac{u'_1 - u'_2}{u_1 - u_2} = \frac{v'_1 - v'_2}{v_1 - v_2} \Rightarrow \frac{v'_1 - v'_2}{u'_1 - u'_2} = \frac{v_1 - v_2}{u_1 - u_2}$$

où l'on voit bien que cette implication correspond au fait que les angles sont conservés par cette homothétie.

Et dans le cas des points à l'infini on obtient :

$$C_0 = \frac{u'_1}{v'_1} = \frac{u_1}{v_1} = \frac{u'_2}{v'_2} = \frac{u_2}{v_2}.$$

On doit donc suivre **au moins** un point à distance finie et une direction, ou deux points à distance finie, pour pouvoir calculer notre transformation.

## 5. Étude expérimentale de l'influence de la mise au point et de la focale sur les paramètres intrinsèques

Avant d'étudier expérimentalement notre modèle examinons la validité de nos hypothèses.

Sur la "tête" de l'INRIA (voir Figure 2) la caméra droite dispose d'un mécanisme de zoom et de mise au point. On n'accède qu'à des mesures en volts pour la distance de mise au point et la focale. On note *focus* et *zoom* ces mesures qu'on a normalisé en un premier temps entre 0 et 1.

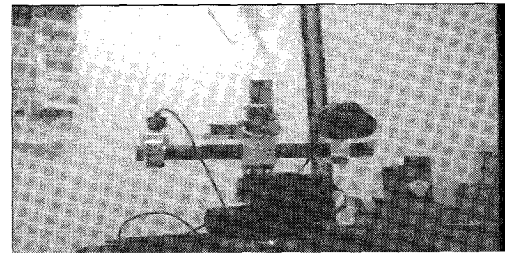


Figure 2. - La "tête" de l'INRIA.

La relation qui lie la distance focale en mètres  $f$  du modèle sténopé et ces mesures *focus* et *zoom* est supposée affine :

$$f = \delta \cdot \text{zoom} + \beta \cdot \text{focus} + \gamma \quad (4)$$

La deuxième hypothèse est que la distance de mise au point est supposée constante quand le zoom varie, et vice-versa.

Les données proposées sont le résultat de 10 calibrations successives où l'on fait varier le *focus* mesuré dans l'intervalle  $\{0.1, 0.2, \dots, 1.0\}$ .

Nous avons aussi complétés ces résultats par un mesure de 5 calibrations successives en faisant varier la focale mesurée (*zoom*) entre  $\{0.2, 0.4, 0.6, 0.8, 1.0\}$ .

A chaque prise d'image on a calibré grâce à la mire :

- la valeur des paramètres intrinsèques :  $\alpha_u, \alpha_v, u_0, v_0$ ,
- l'angle entre les deux axes  $\theta$ ,
- la position du centre optique dans le repère attaché à la mire, et l'origine du repère de la mire dans le repère caméra, ceci en utilisant la méthode développée par [12].

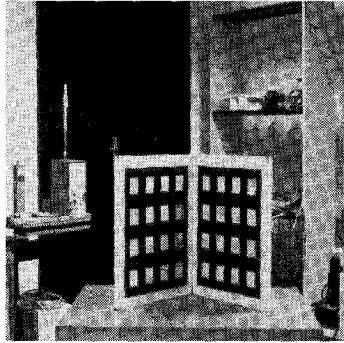


Figure 3. – Exemple d'image d'une mire utilisée pour la calibration.

### 5.1. STABILITÉ DES MESURES DES PARAMÈTRES PAR RAPPORT À LA LUMIÈRE PRÉSENTE DANS LA SALLE

Notre premier essai avait pour but de vérifier la stabilité des paramètres intrinsèques résultat de la calibration par rapport à la lumière présente dans la salle.

On peut noter dans la plupart des graphiques suivants que malgré la disparité des résultats, le comportement global des paramètres est le même en présence d'un fort éclairage (*avec*) que d'un plus faible (*sans*). Cette stabilité est liée à la méthode utilisée.

On peut remarquer les résultats suivants :

- On note  $\theta = \frac{\pi}{2} + \epsilon$ . On a représenté  $\epsilon$  (Figure 4), et on peut noter que dans le cas avec lumière, on obtient des valeurs pour  $\epsilon$  plus centrés autour de 0 et avec une plus petite variance.

Dans la section suivante on a vérifié cette hypothèse en approximant aux moindres carrés.

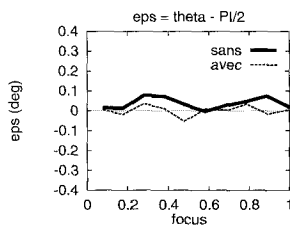


Figure 4. – Evolution de  $\epsilon$  par rapport à la présence ou absence de lumière.

- $\alpha_u$  et  $\alpha_v$  décroissent linéairement (Figure 5).

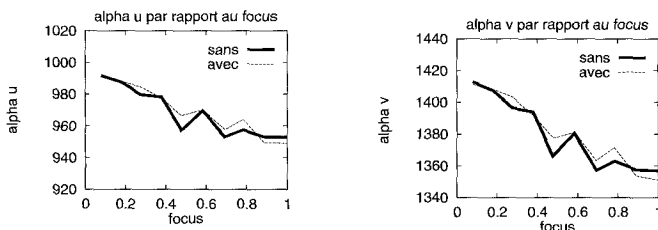


Figure 5. – Evolution de  $\alpha_u$  et  $\alpha_v$  par rapport à la présence ou absence de lumière.

- On a utilisé la même échelle pour représenter  $u_0$  et  $v_0$ . Le deuxième croît deux fois plus vite que le premier. (Figure 6).

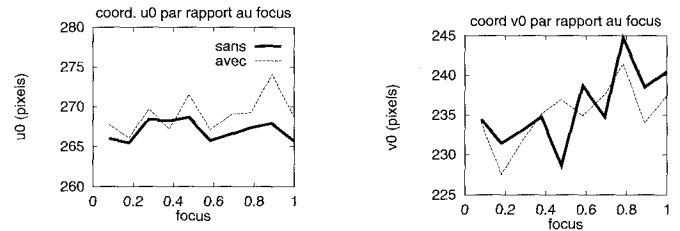


Figure 6. – Evolution du point principal par rapport à la présence ou absence de lumière.

### 5.2. PRÉ-CALIBRATION DU ZOOM ET DE LA MISE AU POINT

Comme la lentille n'est calibrée ni au niveau du zoom, ni au niveau de la mise au point, il faut calculer les gains et les offsets de notre modèle :

$$f = \delta \cdot \text{zoom} + \beta \cdot \text{focus} + \gamma$$

On va donc pour cela mesurer  $\alpha_u$  et  $\alpha_v$ , en faisant varier le *focus* d'abord, puis dans une seconde série de mesures en faisant varier le *zoom*. On a représenté :

$\alpha_u$  par rapport au *focus* mesuré en supposant le *zoom* constant,

$\alpha_u$  par rapport au *zoom* mesuré en supposant le *focus* constant

De même pour  $\alpha_v$ .

Après on a approximé par une droite au sens des moindres carrés  $y = a * x + b$ . On n'a pas pu faire un test de  $\chi^2$  puisque l'on ne connaît pas l'erreur de mesure de chaque point  $\sigma_i$ , (reprendre plusieurs mesures pour chaque focale et mise au point aurait été une solution, mais le procès de calibration a déjà été assez long et tedieux).

Les conclusions les plus intéressantes sont :

- **Le rapport  $\frac{\alpha_u}{\alpha_v}$  est invariant.** L'écart-type de  $\frac{\alpha_u}{\alpha_v}$  est  $\sigma = 4.10^{-4}$  pour le *focus* et  $\sigma = 10^{-3}$  pour le *zoom*. Un test de Student sur ces données montre que le modèle  $\frac{\alpha_u}{\alpha_v} = \text{constant}$ , ne peut être rejeté même avec une probabilité  $P = 0.002$  pour le *focus* et  $P = 0.01$  pour le *zoom*.

Le rapport est 0.702 pour le *focus* et 0.706 pour le *zoom*. Voir Figure 7.

De plus, les données constructeur étant de 0.7, nous sommes fort proche de cette valeur théorique. Notons que ce résultat est valable dans le cas d'une variation soit de mise au point soit de zoom.

- On a **calibré la mise au point** (Figure 8).

Par interpolation linéaire on peut écrire :

$$\alpha_u = -47.9 \text{ focus} + 995.43$$

$$\alpha_v = -68.97 \text{ focus} + 1418.21$$

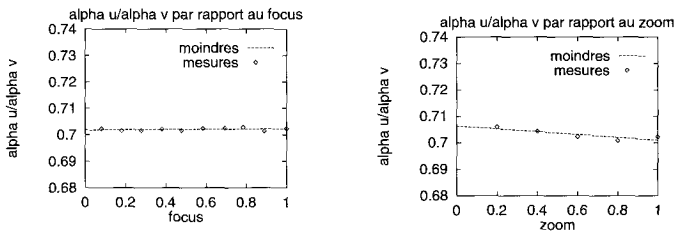


Figure 7. – Rapport  $\alpha_u/\alpha_v$  en fonction de la variation de la mise au point et du zoom

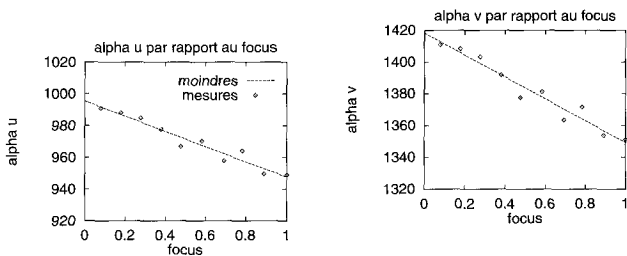


Figure 8. –  $\alpha_u$  et  $\alpha_v$  par rapport à la variation de la mise au point.

avec un coefficient de corrélation égal à -0.97 pour les deux graphiques.

- On a calibré le zoom par la même méthode. Voir Figure 9.

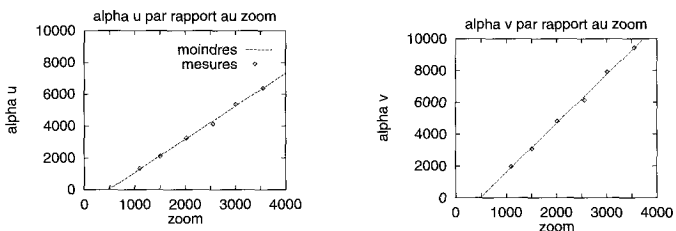


Figure 9. –  $\alpha_u$  et  $\alpha_v$  par rapport à la variation du zoom

On a obtenu un coefficient de corrélation de 0.99 pour les deux graphiques.

- On constate une **variation linéaire** du point principal  $(u_0, v_0)$  (Figure 10), par rapport à la mise au point, avec un coefficient de corrélation de 0.48 pour  $u_0$  et de 0.63 pour  $v_0$ . L'écart-type est  $\sigma = 2.2$  pour  $u_0$  et  $\sigma = 3.6$  pour  $v_0$ .

$$u_0 = 3.68 * focus + 267$$

$$v_0 = 7.72 * focus + 230.9$$

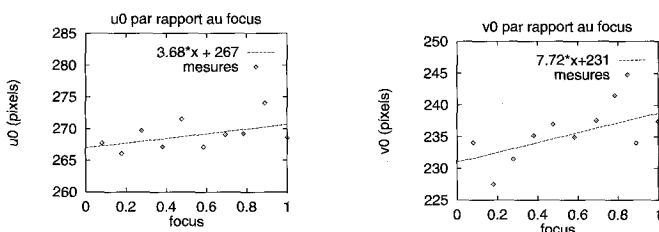


Figure 10. – Evolution du point principal  $(u_0, v_0)$  en fonction de la mise au point.

A partir ces données un test de Student a montré que le modèle ne pouvait être rejeté même avec une probabilité  $P = 0.01$ .

Ce phénomène a été très précisément étudié par [20]. Il y est établi qu'un modèle linéaire de variation des coordonnées du point principal en fonction de la mise au point est valide. Les variations de  $(u_0, v_0)$  en fonction du zoom sont plus complexes et nécessitent un modèle quadratique, mais nous n'aborderons pas ce point déjà étudié [20].

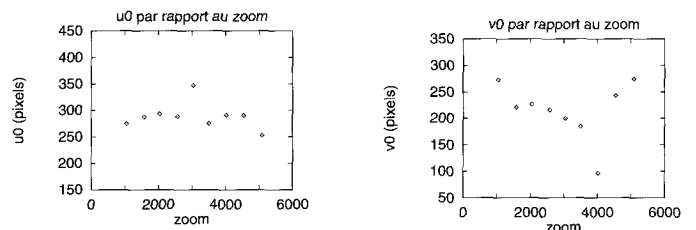


Figure 11. – Evolution du point principal  $(u_0, v_0)$  en fonction du zoom.

- L'angle  $\theta$  est égal à  $\frac{\pi}{2}$  à moins de 0,05 degrés près, dans le cas des mesures avec fort éclairage.

On note  $\theta = \frac{\pi}{2} + \epsilon$ . On a obtenu les droites d'approximation suivantes :

$$\epsilon = 0.0025 * focus - 0.0004 \text{ avec forte lumière, et}$$

$$\epsilon = 0.0074 * focus + 0.033 \text{ sans lumière.}$$

L'écart-type est  $\sigma = 0.024$  et  $\sigma = 0.027$  pour  $\epsilon$  avec forte et faible lumière, et le modèle n'a pas été à nouveau statistiquement rejeté.

Ceci est encore tout à fait valide pour les variations de focale.

Par conséquent cela valide notre hypothèse d'invariance de l'angle dans les deux cas (Figure 12).

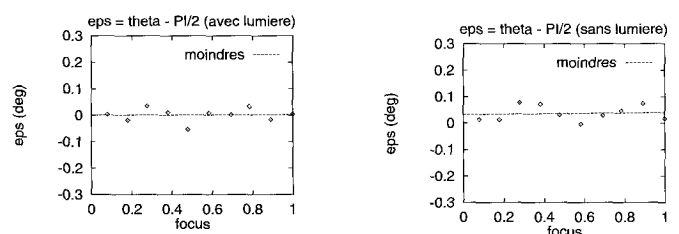


Figure 12. –  $\epsilon$  par rapport à la variation de la mise au point.

- On a représenté les coordonnées du centre optique, par rapport à la variation de la mise au point, dans le repère du monde attaché à la cible, puisque celle-là n'a pas changé de position. Les coordonnées sont en millimètres sur les trois axes  $X_{co}, Y_{co}, Z_{co}$ , et on a approximé par une droite aux moindres carrés les mesures par rapport à la mise au point.

Les écart-type sont  $\sigma = 2.85, 1.48$  et  $5.51$ .

Les coefficients de corrélation sont : 0.57, 0.46 et 0.73.

On pourrait conclure que **le centre optique se déplace surtout au long de l'axe Z au cours d'un changement de mise au point**, bien que la variation en X et Y ne soit pas négligeable.

Par conséquent la mise au point doit vraisemblablement décaler le centre optique lors de sa variation, ce qui explique aussi la translation de  $u_0$  et  $v_0$ , comme attendu [20].

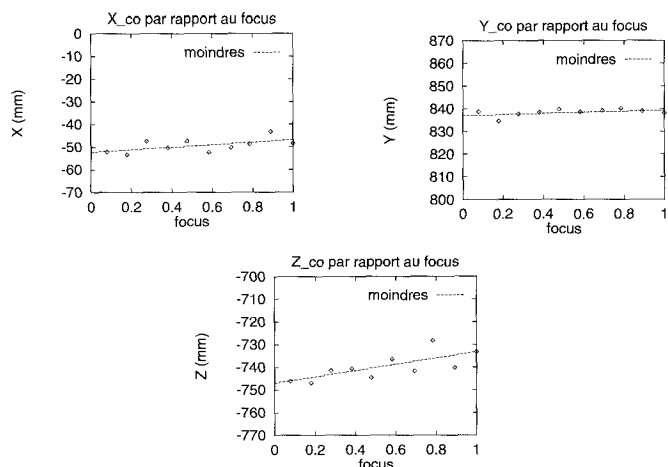


Figure 13. – Centre optique par rapport à la mise au point.

- Pour pouvoir étudier l'évolution du centre optique par rapport aux variations de focale on a pris une nouvelle séquence d'images. L'idée est la suivante : pour pouvoir représenter la position du centre optique toujours dans le même repère il a fallu calculer le déplacement de la mire nécessaire pour pouvoir l'avoir toujours dans le champ de vue. Pour calculer ce déplacement on a calibré en deux étapes. D'abord on calibre pour une focale donnée, puis on déplace la mire et on recalibre.

Le déplacement du centre optique n'est plus linéaire comme prévu [20] puisque notre caméra est de qualité standard et le déplacement du centre optique est couplé avec une rotation.

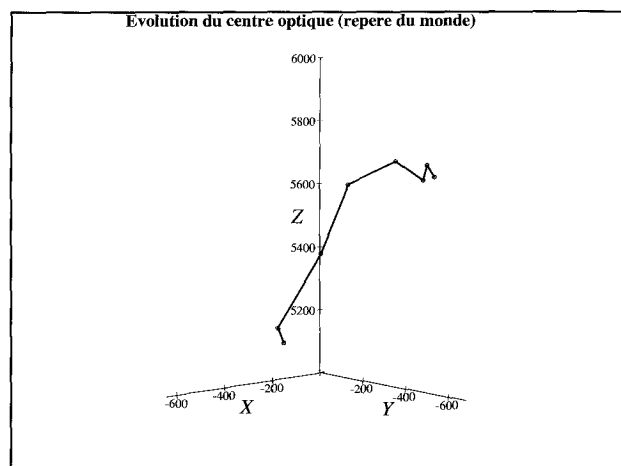


Figure 14. – Centre optique par rapport à la variation du zoom.

## 6. Résultats expérimentaux

Dans cette section on étudie comment estimer notre modèle lors du suivi de points d'intérêt.

### 6.1. CALCUL DE LA MATRICE AFFINE AVEC LE NOMBRE MINIMUM DE POINTS

On utilise la méthode de [1] qui permet de mesurer la position de points d'intérêt avec une très grande précision, même dans le cas où l'image est floue. On extrait ces points sur deux images consécutives et on les apparie manuellement dans cette expérimentation.

A partir des hypothèses d'invariance et **en suivant deux points** on obtient les équations suivantes :

$$\begin{cases} C_0 = \frac{u'_1 - u'_2}{u_1 - u_2} = \frac{v'_1 - v'_2}{v_1 - v_2} \\ a_0 = u'_1 - C_0 \cdot u_1 = u'_2 - C_0 \cdot u_2 \\ b_0 = v'_1 - C_0 \cdot v_1 = v'_2 - C_0 \cdot v_2 \end{cases}$$

On a calculé la matrice de transformation à 3 paramètres noté *avec hypothèses* :

$$A_{hyp} = \begin{pmatrix} C_0 & 0 & a_0 \\ 0 & C_0 & b_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

D'un autre côté, on a calculé la matrice de transformation affine à 6 paramètres (noté *général*) en **suivant 3 points** :

$$A_{gen} = \begin{pmatrix} m_{11} & m_{12} & a_0 \\ m_{21} & m_{22} & b_0 \\ 0 & 0 & 1 \end{pmatrix}$$

Le graphique suivant (Figure 15) montre l'évolution de l'erreur dans la localisation des points image à partir des matrices de transformation calculées par les deux méthodes : *avec les hypothèses*  $m_{12} = 0$ ,  $m_{21} = 0$ , et  $m_{11} = m_{22} = C_0$ , et celle noté *général* qui calcule les 6 paramètres de la transformation. Ceci à partir de deux correspondances pour la première méthode et de 3 pour la deuxième. En abscisses on a représenté la distance du point localisé au baricentre des points utilisés pour le calcul de la matrice.

On peut noter que la méthode *avec hypothèses* est plus robuste avec un petit nombre des points car elle estime moins de paramètres, et reste toujours à moins de 2 pixels d'erreur pour la mise au point et à moins de 5 pour le zoom dans la plupart des cas.



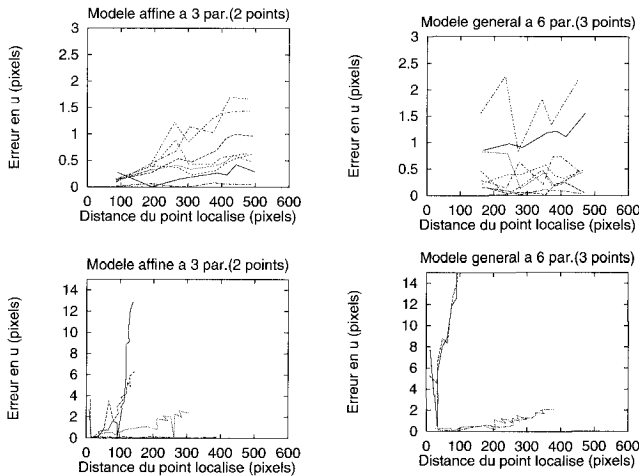


Figure 15. – Evolution de l'erreur dans la localisation des points image par les 2 méthodes : avec les hypothèses d'invariance à partir de 2 points et le modèle général à partir de 3 points de l'image. En haut les graphiques pour la variation de la mise au point et en bas pour le zoom. Les mesures au-delà de 15 pixels ont été rejetés.

## 6.2. MINIMISATION DE L'ERREUR AUX MOINDRES CARRÉS

On note  $nb$  le nombre total des points disponibles ayant été appariés entre deux images, et  $\sum = \sum_{i=1}^{nb}$  l'opérateur de sommation sur les quantités  $u, v, u', v'$ .

Par la première méthode ( $A_{hyp}$ ), en minimisant aux moindres carrés :

$$\sum_{i=1}^{nb} (u'_i - [C_0 \cdot u_i + a_0])^2 + (v'_i - [C_0 \cdot v_i + b_0])^2$$

on obtient en dérivant par rapport aux trois paramètres ( $a_0, b_0, C_0$ ) :

$$\begin{cases} C_0 = \frac{\sum u' \cdot u \cdot \sum 1 + \sum v' \cdot v \cdot \sum 1 - \sum u' \cdot \sum u - \sum v' \cdot \sum v}{\sum u^2 \cdot \sum 1 + \sum v^2 \cdot \sum 1 - (\sum u)^2 - (\sum v)^2} \\ a_0 = \frac{\sum u' - C_0 \cdot \sum u}{\sum 1} \\ b_0 = \frac{\sum v' - C_0 \cdot \sum v}{\sum 1} \end{cases}$$

En minimisant au sens des moindres carrés par la deuxième méthode ( $A_{gen}$ ) à 6 paramètres :

$$\sum_{i=1}^{max} (u'_i - [m_{11}u_i + m_{12}v_i + a_0])^2 + (v'_i - [m_{21}u_i + m_{22}v_i + b_0])^2$$

et en calculant la dérivée partielle par rapport à chacun de nos paramètres ( $m_{11}, m_{12}, a_0, m_{21}, m_{22}, b_0$ ) on obtient les équations linéaires suivantes :

$$\begin{cases} \sum u' \cdot u = m_{11} \cdot \sum uu + m_{12} \cdot \sum uv + a_0 \cdot \sum u \\ \sum u' \cdot v = m_{11} \cdot \sum uv + m_{12} \cdot \sum vv + a_0 \cdot \sum v \\ \sum u' = m_{11} \cdot \sum u + m_{12} \cdot \sum v + a_0 \cdot \sum 1 \\ \sum v' \cdot u = m_{21} \cdot \sum uu + m_{21} \cdot \sum uv + b_0 \cdot \sum u \\ \sum v' \cdot v = m_{21} \cdot \sum uv + m_{22} \cdot \sum vv + b_0 \cdot \sum v \\ \sum v' = m_{21} \cdot \sum u + m_{22} \cdot \sum v + b_0 \cdot \sum 1 \end{cases}$$

que l'on résout aisément de manière analytique.

Le graphique Figure 16 montre l'évolution de l'erreur dans la localisation des points image à partir des matrices de transformation calculées par les deux méthodes. Plus concrètement on a appliqué la matrice de transformation calculée à 20 points de l'image originale. L'erreur en  $u$  est la différence entre la coordonnée résultante et la position réelle qu'on a obtenue en utilisant la méthode sub-

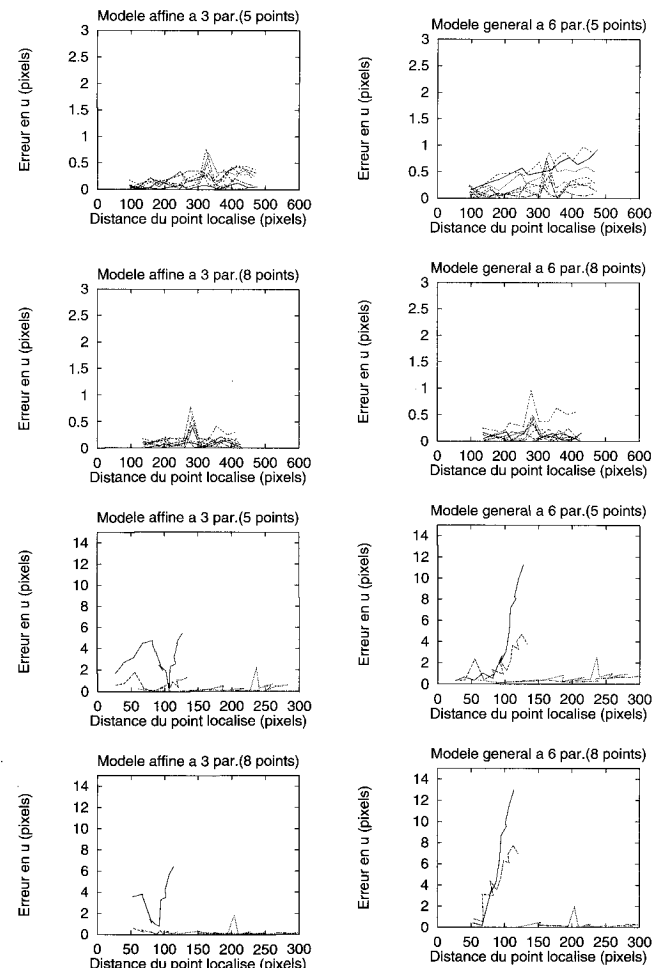


Figure 16. – Evolution de l'erreur  $u$  dans la localisation. En haut pour la mise au point et en bas pour le zoom.

pixélique décrite en [3] (Figure 16). La grandeur représentée sur abscisses est la distance du point localisé au baricentre des points utilisés (5 ou 8), pour le calcul de la matrice de transformation.

Sur les Figures 17 et 18, on a représenté les paramètres  $a_0$ ,  $m_{11}$ ,  $m_{12}$ ,  $m_{21}$  calculés par les deux méthodes à partir de 2 ou 3, 5 et 8 correspondances. Les paramètres  $b_0$  et  $m_{22}$  ont exactement le même comportement que  $a_0$  et  $m_{11}$  respectivement.

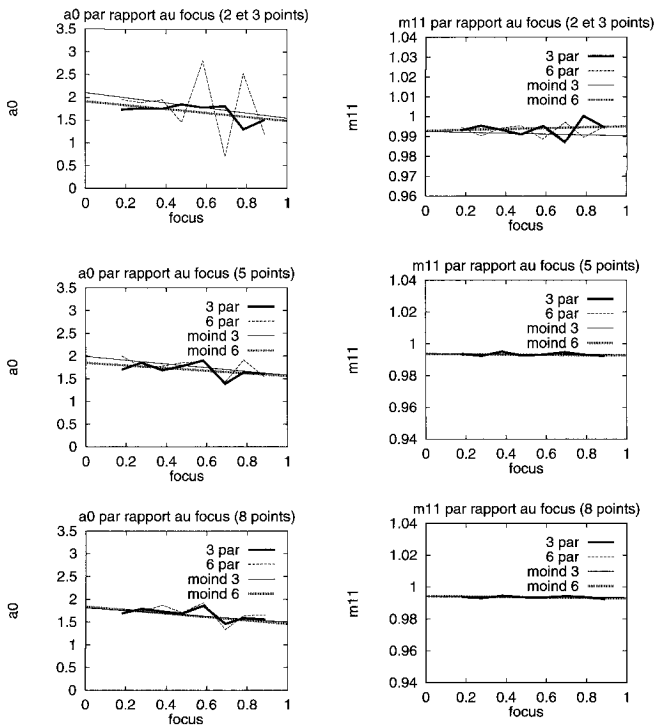


Figure 17. – Evolution de  $a_0$  et  $m_{11}$  avec 2 et 3, 5 et 8 points.

Les écart-type, par exemple, pour  $m_{21}$  sont 0.0056, 0.0011 et 0.0007 pour 3, 5 et 8 points. Le reste sont du même ordre.

Pour  $a_0$  on a pour le modèle général 0.64, 0.18 et 0.16 et pour notre modèle 0.17, 0.14 et 0.12.

On peut facilement conclure que nos hypothèses  $m_{12} = 0$ ,  $m_{21} = 0$ , et  $m_{11} = m_{22} = C_0$ , se vérifient, car avec le modèle à trois paramètres on obtient des résultats aussi précis, et quasiment égaux, à ceux obtenus par le modèle général (voir Figure 17 et 18).

## 7. Conclusion

Ce travail préliminaire à une étude dynamique de l'autocalibration permet de cerner la complexité du problème et de prévoir comment aborder la construction effective d'algorithmes d'autocalibration.

En effet, nous avons montré deux choses : tout d'abord il est possible de *pré-calibrer* avec une assez forte précision une lentille

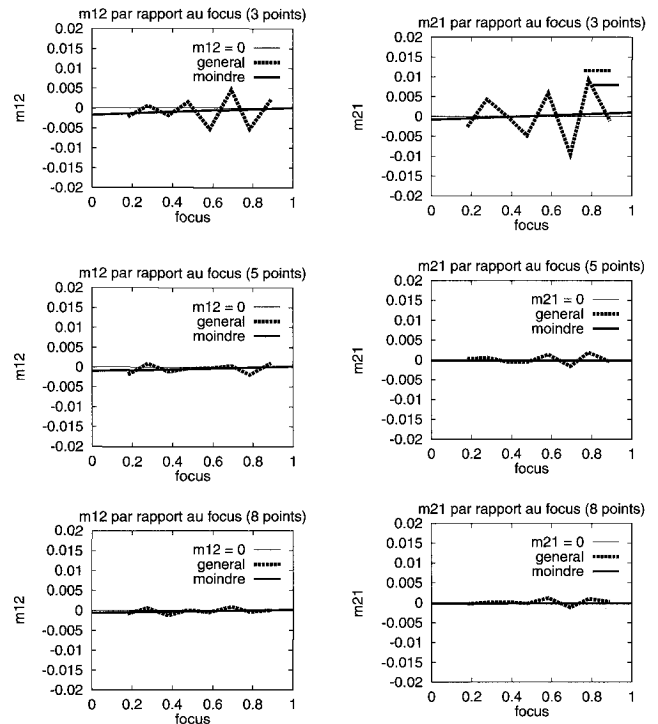


Figure 18. – Evolution de  $m_{12}$  et  $m_{21}$  avec 3, 5 et 8 points.

dotée d'un zoom et d'un mécanisme de mise au point; de plus il est possible de suivre les modifications de ces paramètres, de *recalibrer* à partir de variations simples du modèle initial.

Dans les travaux de l'auteur [20] dont nous donnons copie des résultats dans la figure 19, on trouve une variation assez complexe du point principal lors de changement de la mise au point et du zoom. Nos résultats montrent que pour la lentille considérée ici,

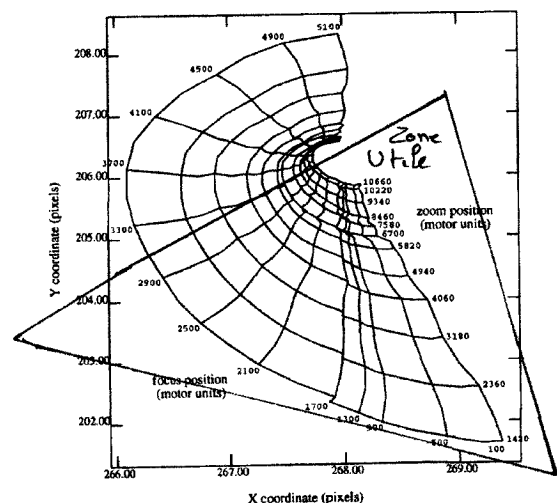


Figure 19. – Evolution de  $u_0$  et  $v_0$  en fonction du zoom et de la mise au point [20].

la variation est linéaire en fonction de la mise au point et plus complexe pour la variation du zoom.

Ceci est vraisemblablement relié à deux phénomènes : (1) la qualité mécanique de la lentille (un objectif Mechatronics type M10Z10), et (2) la plage de variation dont nous avons besoin.

Ainsi, pour un autre système optique, nous avons pu relever les données suivantes, en utilisant un objectif Canon J8 × 6B48 de haute qualité avec une focale de 6mm à 48mm :

focal	u0	v0	au	av	au/av
06	262.66	248.05	128.63	190.10	0.6766
07	252.76	249.89	xxx	xxx	xxx
08	256.33	249.07	126.85	189.60	0.6690
09	253.77	241.16	462.17	660.38	0.6998
10	258.86	251.62	556.26	793.67	0.7009
12	252.85	242.48	668.74	953.09	0.7016
14	251.40	242.41	783.06	1114.55	0.7026
16	252.73	265.28	868.10	1235.35	0.7027
25	244.85	291.51	1427.20	2033.12	0.7020
30	247.81	299.21	1742.43	2482.23	0.7018

dont l'analyse a montré les mêmes phénomènes de variation linéaire.

De même, comme nous l'avons signalé en introduction, le modèle sténopé est valide parce que nous l'utilisons de manière locale comme recommandé par [8]. Ainsi dans le cas de forte focale avons nous dû bien limiter l'étendue de la grille de calibration sur la rétine et ne pas lui faire couvrir tout le champ mais seulement une zone centrale ("fovéale") de façon à ce que la calibration reste valide.

En fait cette étude montre que pour un objectif de vision robotique, i.e. bas de gamme, un modèle linéaire est bien suffisant pour décrire les variations des principaux paramètres intrinsèques, car les autres effets plus fins sont "noyés dans le bruit". Une telle information n'est pas évidente au départ, même si elle est cohérente avec d'autres études exhaustives du domaine telles que [19].

Ainsi ce mécanisme va permettre de maintenir de manière approchée mais néanmoins assez précise, les paramètres intrinsèques d'une lentille d'un système actif. Ce module, opérationnel, a été implémenté sur la tête de l'INRIA et a été utilisé dans des études ultérieures qui ont montré sa validité et on a pu vérifier qu'une reconstruction 3D à partir de notre étude était rendue possible [17, 2].

Par ailleurs, nous avons la possibilité *sans connaître explicitement les paramètres du système* de remettre à jour ce modèle de calibration étant donné : (a) un modèle initial, (b) un ensemble de correspondances. Bien que notre travail soit valide uniquement dans le cas où le système - immobile - observe une scène fixe et ne modifie que la focale ou la mise au point, il démontre néanmoins que certains cas particuliers valent la peine d'être exploités car ils correspondent à des algorithmes très simples et il est facile de ne déclencher ces algorithmes que dans les cas idoines.

Il va être très facile de généraliser ces travaux à des modèles plus complexes en considérant 2 ou 3 vues [10] où une séquence

de vues [18]. Dans ce cas, les algorithmes potentiels promettant d'être hautement non-linéaires [9] et dotés d'un rayon de convergence assez faible (on doit être proche de la solution finale pour qu'il y ait convergence) [18], il sera d'autant plus important de se baser sur une estimée initiale précise, comme obtenue ici.

## BIBLIOGRAPHIE

- [1] T. Blaszk. Modélisation et caractérisation des variations d'intensité dans une image. Technical report, Rapport de Stage de DEA, Université de Nice, Septembre 1992.
- [2] E. Clergue and T. Viéville. Methods for dense reconstruction in active vision. In *Proc. 17th European Conference on Visual Perception, Eindhoven, 1994*.
- [3] R. Deriche and T. Blaszk. Recovering and Characterizing Image Features Using An Efficient Model Based Approach. In *Computer Vision And Pattern Recognition, New-York, June 14-17 1993*.
- [4] O. Faugeras. *Three-dimensional Computer Vision : a geometric viewpoint*. MIT Press, Boston, 1993.
- [5] O. Faugeras, Q. T. Luong, and S. Maybank. Camera self-calibration : Theory and experiment. In *2nd ECCV, Genoa, 1992*.
- [6] J. Lavest. *Stereovision axiale par zoom pour la robotique*. PhD thesis, Université Blaise Pascal de Clermont-Ferrand, 1992.
- [7] J. Lavest, G. Rives, and M. Dhôme. Utilisation d'un objectif a focale variable en vision monoculaire en vue de la reconstruction 3d. *Traitement du Signal*, 9(6) :491-506, 1992.
- [8] J. Lavest, G. Rives, and M. Dhôme. 3d reconstruction by zooming. *IEEE Trans. on Robotics and Automation*, 9(2) :196-207, April 1993.
- [9] Q. Luong, R. Deriche, O. Faugeras, and T. Papadopoulos. On determining the fundamental matrix : analysis of different methods and experimental results. Technical Report RR-1894, INRIA, Sophia, France, 1993.
- [10] T. Luong. *Matrice Fondamentale et Calibration Visuelle sur l'Environnement*. PhD thesis, Université de Paris-Sud, Orsay, 1992. PhD thesis.
- [11] S. Maybank and O. Faugeras. A theory of self-calibration of a moving camera. *The International Journal of Computer Vision*, 8, 1992.
- [12] L. Robert. *Perception Stéréoscopique de Courbes et de Surfaces Tridimensionnelles, Application à la Robotique Mobile*. PhD thesis, Ecole Polytechnique, Palaiseau. France, 1992. PhD thesis.
- [13] G. Toscani and O. Faugeras. Camera calibration for 3D computer vision. In *Proceedings of the International Workshop on Machine Intelligence, Tokyo, February 1987*.
- [14] R. Y. Tsai. An efficient and accurate calibration technique for 3D machine vision. In *IEEE Proc CVPR'86, Miami Beach, Fl., June*, pages 364-374, 1986.
- [15] R. Y. Tsai. Synopsis of recent progress on camera calibration for 3D machine vision. *Robotics Review*, 1 :147-159, 1989.
- [16] T. Viéville. Autocalibration of visual sensor parameters on a robotic head. *Image and Vision Computing*, 12, 1994.
- [17] T. Viéville, E. Clergue, R. Enciso, and H. Mathieu. Experimenting 3d vision on a robotic head. In *The 12th Int. Conf. on Pattern Recognition*, 1994. 739-743.
- [18] T. Viéville, Q. Luong, and O. Faugeras. Motion of points and lines in the uncalibrated case. *International Journal of Computer Vision*, 1994. To appear.
- [19] R. Willson. *Modeling and Calibration of Automated Zoom Lenses*. PhD thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, 1994.
- [20] R. Willson and S. Shafer. What is the center of the image? In *IEEE Proc CVPR'93, New-York, June*, pages 670-671, 1993.

*Manuscrit reçu le 2 février 1994.*

## LES AUTEURS

### Reyes ENCISO



Reyes Enciso est étudiante en thèse dans le projet Robotvis à l'INRIA (Institut National de Recherche en Informatique et en Automatique) de Sophia-Antipolis. Elle a fait des études en Informatique à l'Université Polytechnique de Valence (Espagne), et puis a obtenu le DEA de l'Université d'Orsay (Paris XI). Son sujet de thèse est l'auto-calibration de capteurs visuels.

### Olivier FAUGERAS



Olivier Faugeras est Directeur de Recherche à l'INRIA-Sophia au sein du projet Robotvis. Il est éditeur associé de plusieurs revues internationales et en 1990 il a été Chairman de la première Conférence Européenne en Vision par Ordinateur (ECCV). Ses principaux domaines de recherche sont l'application des Mathématiques à la Vision par Ordinateur, la Robotique, les architectures des systèmes de Vision Artificielle, en plus des liens entre la Vision Artificielle et Biologique.

### Thierry VIEVILLE



Thierry Vieville est un Chercheur sur poste à l'INRIA-Sophia. Après son Diplôme à l'École Nationale Supérieure des Télécommunications (Bretagne et Paris), il a soutenu sa thèse dans le domaine de la neurobiologie. Habilité à diriger les recherches depuis 1993, ses principaux domaines de recherche sont l'application des Mathématiques à la Vision par Ordinateur, Analyse du Mouvement, en plus des liens entre la Vision Artificielle et Biologique.