



Etalonnage simultané de systèmes stéréoscopiques hybrides

Guillaume Caron, Damien Eynard

► To cite this version:

Guillaume Caron, Damien Eynard. Etalonnage simultané de systèmes stéréoscopiques hybrides. ORA-SIS - Congrès des jeunes chercheurs en vision par ordinateur, INRIA Grenoble Rhône-Alpes, Jun 2011, Praz-sur-Arly, France. inria-00595254

HAL Id: inria-00595254

<https://inria.hal.science/inria-00595254>

Submitted on 24 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Étalonnage simultané de systèmes stéréoscopiques hybrides

Guillaume Caron

Damien Eynard

Université de Picardie Jules Verne, laboratoire MIS
33 rue Saint-Leu, 80039 Amiens

{guillaume.caron, damien.eynard}@u-picardie.fr

Résumé

L'étalonnage est un problème classique en vision artificielle mais nécessaire pour retrouver l'information 3D à partir de mesures réalisées dans l'image. Ce travail présente une méthode d'étalonnage d'un banc stéréoscopique hybride composé de caméras de modèles différents (perspective, fisheye, catadioptrique). Cet article propose d'étendre la méthode d'étalonnage de caméra perspective utilisant l'asservissement visuel virtuel (AVV) à l'étalonnage simultané intrinsèque et extrinsèque d'un banc stéréo de caméras de modèles hybrides. L'approche présentée est valide pour l'étalonnage de bancs stéréo composés de N caméras modélisées par N modèles différents. Les résultats d'étalonnage, comparés aux approches de l'état de l'art, et utilisés pour une application d'estimation de plan 3D, montrent l'efficacité de l'approche. Enfin, un logiciel multi-plateforme implémentant cette méthode est mis à disposition de la communauté.¹

Mots Clef

Stereovision, étalonnage, vision omnidirectionnelle, mixte, hybride

Abstract

Calibration is a classical issue in computer vision needed to retrieve 3D information from image measurements. This work presents a calibration approach for hybrid stereo rig involving multiple central camera types (perspective, omnidirectional). The paper extends the method of monocular perspective camera calibration using virtual visual servoing. The simultaneous intrinsic and extrinsic calibration of central cameras rig, using different models for each camera, is developed. The presented approach is suitable for the calibration of rigs composed by N cameras modelled by N different models. Calibration results, compared with state of the art approaches, and a 3D plane estimation application, allowed by the calibration, show the effectiveness of the approach. A cross-platform software implementing this method is available.¹

Keywords

Stereovision, calibration, omnidirectional vision, mixed, hybrid

1 Introduction

Les systèmes stéréoscopiques sont généralement composés de deux caméras du même type. Deux caméras perspectives présentent l'inconvénient d'un champ de vue limité tandis que deux caméras omnidirectionnelles ont une résolution spatiale limitée et non uniforme. Combiner différents types de caméras en un même banc permet de fusionner le très large champ de vue de la caméra omnidirectionnelle et la précision de la caméra perspective [1, 2].

Les caméras perspective et omnidirectionnelle sont dites centrales, *i.e.* elles possèdent un point de vue unique, sous certaines conditions [3]. Les caméras perspectives sont généralement modélisées par le modèle perspective. Les caméras omnidirectionnelles centrales peuvent être représentées par le modèle de projection centrale unifié, sphérique [4]. Ce modèle est aussi valide pour une caméra perspective même s'il en complexifie la projection. De plus, le calcul de flot optique par la méthode de Lucas et Kanade à partir du plan image est mieux adapté au modèle perspective qu'aux images omnidirectionnelles [5]. En effet, la projection perspective planaire préserve un voisinage carré et un échantillonnage régulier autour d'un point, ce qui n'est pas le cas du modèle sphérique [6]. Ainsi, les méthodes basées intensité n'étant pas liées au modèle de projection peuvent être adaptées à des modèles hybrides.

Le champ d'application des systèmes mixtes se retrouve principalement dans la vidéo-surveillance et la navigation autonome de robot (mobile, aérien, etc). En utilisant un banc stéréo hybride fisheye et perspective, mixant les modèles sphériques et perspective, Eynard *et al.* [2] ont adapté l'algorithme de plane-sweeping pour estimer l'altitude d'un drone et segmenter le sol. L'algorithme estime la profondeur du plan du sol en minimisant les différences d'intensités sur l'image plane. Cette méthode nécessite néanmoins la connaissance des paramètres intrinsèques et extrinsèques du banc stéréo ainsi que de l'attitude estimée dans la vue fisheye. Cette coopération entre les capteurs est un autre exemple de l'intérêt de l'utilisation d'un

1. <http://damien.eyn.free.fr/HYSCAS/>

banc stéréo mixte.

Pour retrouver l'information 3D, ces bancs stéréo mixtes doivent être étalonnés, l'étalonnage étant un processus numérique permettant de trouver les paramètres d'un modèle. Cet article propose une nouvelle approche pour estimer simultanément les paramètres de projection et les poses relatives de N caméras, modélisées différemment, composant un banc stéréo.

L'étalonnage de caméra est un problème classique aussi bien en vision perspective [7, 8] qu'en vision omnidirectionnelle [10–12]. Marchand *et al.* [7] ont proposé un étalonnage de caméra perspective basé Asservissement Visuel Virtuel (AVV). L'AVV [13, 14] est une technique d'optimisation non linéaire de calcul de pose, extensible à l'étalonnage, tout en bénéficiant des connaissances nombreuses de l'asservissement visuel.

Les travaux sur l'étalonnage de bancs stéréo traitent généralement des caméras perspectives [9] (se basant sur [8]) ou estiment uniquement les paramètres extrinsèques de systèmes hybrides [15]. Cet article étend la méthode d'estimation de pose stéréoscopique, en vision omnidirectionnelle, basée AVV [16] à l'étalonnage de N caméras définies par N modèles différents en utilisant des points. La formulation générique du problème permet de traiter tout modèle de projection, s'il est possible d'exprimer un point image à partir de point 3D correspondant.

Après un rappel des modèles de projection, nous traiterons de l'étalonnage d'une caméra basé AVV. Ensuite, nous présenterons l'étalonnage de systèmes stéréoscopiques composés de différents types de caméras. Enfin, dans une dernière partie, nous évaluerons les résultats d'étalonnage puis les appliquerons à une méthode d'estimation de plan 3D, afin de montrer la précision d'estimation des paramètres des modèles de projection.

2 Modèles de projection

2.1 Projection perspective

La projection perspective (Fig. 1(a)) modélise la caméra perspective. Soit, $\mathbf{X} = (X \ Y \ Z \ 1)^T$, un point 3D exprimé dans le repère caméra, est projeté dans le plan image en $\mathbf{x} = (x \ y \ 1)^T$:

$$\mathbf{x} = pr(\mathbf{X}) \quad \text{avec} \quad \begin{cases} x = \frac{X}{Z} \\ y = \frac{Y}{Z} \end{cases} . \quad (1)$$

$\mathbf{x}$ est le point sur le plan image normalisé. $\mathbf{u} = (u \ v \ 1)^T$, le point pixelique à coordonnées décimales, est obtenu par la relation $\mathbf{u} = \mathbf{K}\mathbf{x}$. $\mathbf{K}$ est la matrice des paramètres intrinsèques, connaissant les paramètres $\gamma_1 = \{p_x, p_y, u_0, v_0\}$:

$$\mathbf{K} = \begin{pmatrix} p_x & 0 & u_0 \\ 0 & p_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

où p_x et p_y sont respectivement le rapport de la focale en mètres sur la largeur et la hauteur d'un pixel en mètre, et u_0

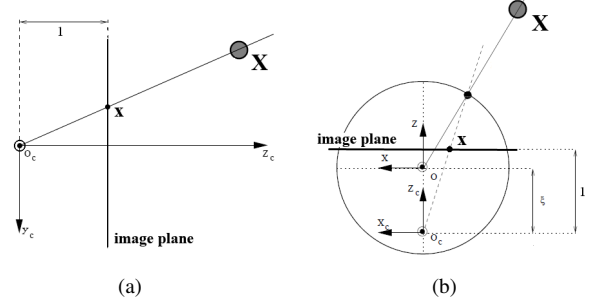


FIGURE 1 – (a) Modèle perspective utilisé pour les caméras perspectives. (b) Modèle sphérique utilisé pour les caméras omnidirectionnelles.

et v_0 , les coordonnées du point principal dans l'image. La fonction complète de projection perspective d'un point 3D dans le plan image pixelique est alors $pr_{\gamma_1}(\mathbf{X}) = \mathbf{K}pr(\mathbf{X})$.

2.2 Projection centrale unifiée

Le modèle de projection unifié [4], représenté par une sphère, est valide pour certaines caméras catadioptriques [3] et aussi pour certaines caméras fisheye [17]. En suivant le modèle sphérique (Fig. 1(b)), un point 3D $\mathbf{X}$ est d'abord projeté sur une sphère unitaire, centrée en $(0 \ 0 \ \xi)^T$, puis projeté dans le plan image en $\mathbf{x}$, connaissant les paramètres intrinsèques $\gamma_2 = \{p_x, p_y, u_0, v_0, \xi\}$:

$$\mathbf{x} = pr_{\xi}(\mathbf{X}) \quad \text{avec} \quad \begin{cases} x = \frac{X}{Z + \xi\rho} \\ y = \frac{Y}{Z + \xi\rho} \end{cases} , \quad (3)$$

et $\rho = \sqrt{X^2 + Y^2 + Z^2}$. On parle aussi de projection stéréographique. Enfin, un point image pixelique est obtenu à partir d'un point 3D par $pr_{\gamma_2}(\mathbf{X}) = \mathbf{K}pr_{\xi}(\mathbf{X})$ (eq. 2).

2.3 Stereovision

Pour un banc stéréo de N caméras, chacune est modélisée par son modèle le plus adapté : les caméras perspectives par le modèle perspective, les caméras omnidirectionnelles (catadioptrique, fisheye) par le modèle sphérique. Les poses des $N - 1$ caméras c_j du banc, modélisées par des matrices homogènes ${}^{c_j}\mathbf{M}_{c_1}$, sont définies par rapport à la caméra de référence c_1 .

3 Etalonnage d'une caméra

Le problème de calcul de pose à partir de points peut être défini comme un problème d'asservissement visuel virtuel. Habituellement, l'asservissement visuel basé image vise à déplacer une caméra à une position et une orientation désirées en minimisant l'erreur entre les primitives (points, droites, etc) de l'image courante et les primitives de l'image acquise à la pose désirée. L'AVV virtualise la caméra en partant d'une pose initiale, puis déplace la caméra virtuelle pour obtenir une correspondance parfaite entre la projection de l'objet, pour sa pose virtuelle, et l'objet dans l'image réelle. La fonction de projection présente

dans l'équation du mouvement des primitives permet d'estimer simultanément les paramètres du modèle de caméra. La caméra virtuelle c est définie par sa fonction de projection $pr_{\gamma_m}()$ ($m = 1$ ou $m = 2$, suivant le modèle de projection utilisé) et sa position ${}^c\mathbf{M}_o$, par rapport à un objet de référence ou au monde. Avec $\mathbf{r}$, une représentation vectorielle de la pose, la méthode estime la pose réelle en minimisant l'erreur Δ entre l'ensemble des points détectés $\mathbf{u}^*$ et l'ensemble courant de ces points, obtenus par projection de la pose courante $\mathbf{u}(\mathbf{r})$:

$$\Delta = \sum_k (pr_{\gamma_m}({}^c\mathbf{M}_o \circ \mathbf{X}_k) - \mathbf{u}_k^*)^2. \quad (4)$$

${}^o\mathbf{X}_k$ est le k -ième point 3D exprimé dans le repère de l'objet, comme la mire d'étalonnage, par exemple ($\mathbf{X} = {}^c\mathbf{M}_o \circ \mathbf{X}$).

L'erreur à minimiser est, par conséquent, $\mathbf{e} = \mathbf{u}(\mathbf{r}) - \mathbf{u}^*$. En imposant une décroissance exponentielle de l'erreur, le mouvement des primitives est lié à la vitesse de la caméra virtuelle par $\dot{\mathbf{e}} = -\lambda \mathbf{e}$, qui dépend de $\dot{\mathbf{u}}$ [7] :

$$\dot{\mathbf{u}} = \frac{\partial \mathbf{u}}{\partial \mathbf{r}} \frac{d\mathbf{r}}{dt} + \frac{\partial \mathbf{u}}{\partial \gamma_m} \frac{d\gamma_m}{dt} = \frac{\partial \mathbf{u}}{\partial \mathbf{X}} \frac{\partial \mathbf{X}}{\partial \mathbf{r}} \frac{d\mathbf{r}}{dt} + \frac{\partial \mathbf{u}}{\partial \gamma_m} \frac{d\gamma_m}{dt}. \quad (5)$$

$\mathbf{L}_\mathbf{X} = \frac{\partial \mathbf{X}}{\partial \mathbf{r}}$ est la matrice d'interaction de la pose, liée à un point du plan image perspective normalisé [18], également exprimée pour un point du plan image normalisé de l'image omnidirectionnelle [19]. De plus, $\frac{\partial \mathbf{X}}{\partial \mathbf{r}} = \frac{\partial \mathbf{X}}{\partial \mathbf{X}} \frac{\partial \mathbf{X}}{\partial \mathbf{r}}$, c'est-à-dire que pour calculer une pose de caméra avec un autre modèle de projection, il faut exprimer le point image en fonction du point 3D pour exprimer le jacobien $\frac{\partial \mathbf{X}}{\partial \mathbf{X}}$. Le jacobien $\frac{\partial \mathbf{X}}{\partial \mathbf{r}}$ [20] est indépendant des caméras et du modèle de projection. Les jacobiens des paramètres intrinsèques sont détaillés ci-après.

La loi de commande de l'AVV utilisant p images est alors [7] :

$$(\mathbf{v}^1 \quad \mathbf{v}^2 \quad \dots \quad \mathbf{v}^p \quad \dot{\gamma}_m)^T = -\lambda \mathbf{H}^+ \begin{pmatrix} \mathbf{u}^1 - \mathbf{u}^{*1} \\ \mathbf{u}^2 - \mathbf{u}^{*2} \\ \vdots \\ \mathbf{u}^p - \mathbf{u}^{*p} \end{pmatrix} \quad (6)$$

avec $\mathbf{v}^i$, le torseur cinématique de la pose de caméra associée à l'image i , $\dot{\gamma}_m$ la variation temporelle des paramètres intrinsèques (unique pour l'ensemble des images) et $\mathbf{u}^{*i}$, l'ensemble des points détectés dans l'image i . $\mathbf{H}^+$ est la pseudo inverse de $\mathbf{H}$ qui est la jacobienne suivante :

$$\mathbf{H} = \begin{pmatrix} \mathbf{L}_\mathbf{u}^1 & 0 & \dots & 0 & \frac{\partial \mathbf{u}^1}{\partial \gamma_m} \\ 0 & \mathbf{L}_\mathbf{u}^2 & \dots & 0 & \frac{\partial \mathbf{u}^2}{\partial \gamma_m} \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \dots & \dots & \mathbf{L}_\mathbf{u}^p & \frac{\partial \mathbf{u}^p}{\partial \gamma_m} \end{pmatrix} \quad \text{avec} \quad \mathbf{L}_\mathbf{u}^i = \frac{\partial \mathbf{u}^i}{\partial \mathbf{r}^i} \quad (7)$$

et avec $\mathbf{r}^i$, le vecteur de pose lié à l'image i . Par souci de clarté, aucune notation particulière ne définit le modèle de

projection utilisé pour calculer $\mathbf{L}_\mathbf{u}^i$. $\frac{\partial \mathbf{u}}{\partial \gamma_m}$ dépend du modèle de projection. Il est différent si le modèle perspective est utilisé avec γ_1 [7] ou avec le modèle sphérique :

$$\frac{\partial \mathbf{u}}{\partial \gamma_2} = \begin{pmatrix} x & 0 & 1 & 0 & -\frac{p_x \rho_x}{Z + \xi \rho} \\ 0 & y & 0 & 1 & -\frac{p_y \rho_y}{Z + \xi \rho} \end{pmatrix} \quad (8)$$

Les poses sont ensuite mises à jour en utilisant l'application exponentielle de $se(3)$ [21] par ${}^c\mathbf{M}_o^{t+1} = {}^c\mathbf{M}_o^t e^{[\mathbf{v}]}$ et les paramètres intrinsèques mis à jour par $\gamma_m^{t+1} = \gamma_m^t + \dot{\gamma}_m$.

D'un point de vue général, les méthodes d'optimisation ont besoin d'être initialisées. Les paramètres sont grossièrement initialisés : u_0, v_0 au centre de l'image, p_x, p_y à la moitié de la verticale d'image (le "rayon" du miroir, pour les caméras catadioptriques). ξ , initialisée par ses valeurs théoriques connues [4], dépend du type de caméra omnidirectionnelle. La pose de la mire d'étalonnage par rapport à une caméra est initialisée par une méthode de calcul linéaire de pose de caméra perspective [22]. Nous avons adapté cette méthode pour les caméras sphériques. L'AVV se montre robuste aux mauvaises initialisations d'étalonnage intrinsèque de caméra perspective [7], c'est pourquoi ces initialisations grossières peuvent être utilisées.

4 Etalonnage stéréoscopique

En étendant l'étalonnage de caméra unique, le critère d'optimisation devient [16], pour un banc de N caméras :

$$\Delta_S = \sum_{j=1}^N \sum_{k=1}^{k_j} (pr_{\gamma_{m_j}}({}^{c_j}\mathbf{M}_{c_1} {}^{c_1}\mathbf{M}_o \circ \mathbf{X}_k) - {}^{c_j}\mathbf{u}_k^*)^2, \quad (9)$$

avec $pr_{\gamma_{m_j}}$, la fonction de projection de la caméra j ($m_j = 1$ ou $m_j = 2$, suivant le modèle de projection utilisé). En considérant $\mathbf{L}_j$, la matrice d'interaction de la pose de la caméra j , la matrice d'interaction de pose stéréoscopique est alors [16] :

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 {}^{c_2}\mathbf{V}_{c_1} \\ \vdots \\ \mathbf{L}_N {}^{c_N}\mathbf{V}_{c_1} \end{bmatrix} \quad \text{avec} \quad {}^{c_j}\mathbf{V}_{c_1} = \begin{bmatrix} {}^{c_j}\mathbf{R}_{c_1} & [{}^{c_j}\mathbf{t}_{c_1}]_\times \\ 0 & {}^{c_j}\mathbf{R}_{c_1} \end{bmatrix}. \quad (10)$$

Les ${}^{c_j}\mathbf{V}_{c_1}$ sont les matrices de changement de repère entre le torseur cinématique de la caméra 1 ${}^{c_1}\mathbf{v}$ et ceux des caméras j ${}^{c_j}\mathbf{v}$. ${}^{c_j}\mathbf{R}_{c_1}$ et ${}^{c_j}\mathbf{t}_{c_1}$ sont les blocs de rotation et de translation de ${}^{c_j}\mathbf{M}_{c_1}$ et $[{}^{c_j}\mathbf{t}_{c_1}]_\times$, la matrice de pré-produit vectoriel de ${}^{c_j}\mathbf{t}_{c_1}$. Donc, avec ${}^{c_j}\dot{\mathbf{u}}_i$, l'ensemble des variations des primitives de la caméra j dans l'image i , la "loi de commande" de l'étalonnage stéréoscopique pour p prises de vue est :

$$\begin{pmatrix} \mathbf{v}_1^1 & \dots & \mathbf{v}_1^p & \mathbf{v}_{1,2} & \dots & \mathbf{v}_{1,N} & \dot{\gamma}_{m_1} & \dots & \dot{\gamma}_{m_N} \end{pmatrix}^T = \mathbf{H}_S^+ \begin{pmatrix} {}^{c_1}\dot{\mathbf{u}}_1 & \dots & {}^{c_N}\dot{\mathbf{u}}_1 & {}^{c_1}\dot{\mathbf{u}}_2 & \dots & {}^{c_N}\dot{\mathbf{u}}_2 & \dots & {}^{c_N}\dot{\mathbf{u}}_p \end{pmatrix}^T \quad (11)$$

où $\mathbf{v}_1^i$ est le torseur de la pose du banc stéréo, pour chaque ensemble de N images et $\mathbf{v}_{1,j}$, le torseur de pose relative

entre la caméra j et la caméra 1. Considérant $\mathbf{L}^i$, la matrice d'interaction de la pose i du banc stéréo et $\mathbf{L}_{1,j}^i$, la matrice d'interaction de la pose relative entre la caméra 1 et la caméra j pour la pose i , l'expression de $\mathbf{H}_S$ est :

$$\mathbf{H}_S = \begin{pmatrix} \begin{array}{ccc|ccc} 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & \mathbf{L}_{1,N}^1 \end{array} & \begin{array}{ccc|ccc} 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & \mathbf{L}_{1,N}^1 \end{array} & \begin{array}{ccc|ccc} \frac{\partial {}^{c1}\mathbf{u}_1}{\partial \gamma_{m_1}} & 0 & \dots & 0 \\ 0 & \frac{\partial {}^{c2}\mathbf{u}_1}{\partial \gamma_{m_2}} & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & \dots & 0 & \frac{\partial {}^{cN}\mathbf{u}_1}{\partial \gamma_{m_N}} \end{array} \\ \hline \begin{array}{ccc|ccc} \vdots & \dots & \ddots & \vdots & & \vdots \\ 0 & \dots & 0 & \mathbf{L}^p & \vdots & \vdots \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & \mathbf{L}_{1,N}^p \end{array} & \begin{array}{ccc|ccc} 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & \mathbf{L}_{1,N}^p \end{array} & \begin{array}{ccc|ccc} \frac{\partial {}^{c1}\mathbf{u}_p}{\partial \gamma_{m_1}} & 0 & \dots & 0 \\ 0 & \frac{\partial {}^{c2}\mathbf{u}_p}{\partial \gamma_{m_2}} & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & \dots & 0 & \frac{\partial {}^{cN}\mathbf{u}_p}{\partial \gamma_{m_N}} \end{array} \end{pmatrix}.$$

Cette formulation générique de l'étalonnage stéréoscopique hybride simultané permet de prendre en compte tout modèle de projection. Cette formulation est valide si un point pixelique peut être relié à un point 3D et si les jacobiens $\frac{\partial \mathbf{u}}{\partial \mathbf{X}}$ et $\frac{\partial \mathbf{u}}{\partial \gamma_{m_j}}$ sont exprimés.

5 Résultats

Cette section évalue l'approche présentée pour l'étalonnage, en comparant les résultats obtenus avec notre logiciel (accessible en vidéo et en exécutable : <http://damien.eyn.free.fr/HYSCAS/>) à des toolbox existantes pour caméra omnidirectionnelle [10] et banc stéréo de caméras perspectives [9]. L'évaluation de l'étalonnage de caméra perspective basé AVV est réalisé dans [7].

5.1 Caméra omnidirectionnelle

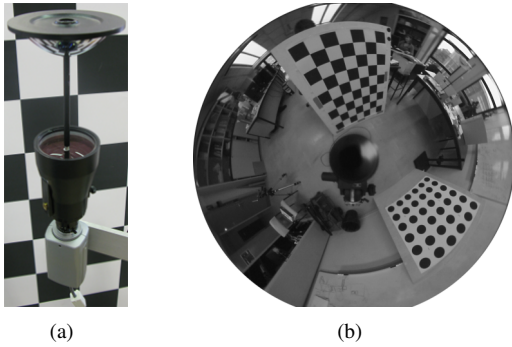


FIGURE 2 – (a) Capteur catadioptrique. (b) Une image utilisée pour l'étalonnage (deux types de mire d'étalonnage).

Dans cette section, les résultats d'étalonnage sont évalués pour une caméra unique modélisée par une sphère. Les résultats obtenus par notre logiciel d'étalonnage, implantant

les idées présentées dans cet article, sont comparés à la toolbox de Mei [10].

La mire d'étalonnage est un damier de 82 cm × 105 cm comptant 7 × 9 carrés. Une autre mire, utilisable avec notre

logiciel, est constituée de 36 disques placés dans un carré de 80 cm × 80 cm. Cette expérimentation montre l'étalonnage de caméra catadioptrique composée d'une caméra (Sony DFW-SX910) et d'une optique parabolique Remote Reality (Fig. 2(a)). Six images d'étalonnage d'une résolution de 1280 × 960 pixels sont utilisées pour un total de 288 points.

L'étalonnage est réalisé sans estimation explicite des distorsions. Les résultats sont reportés dans le tableau 1. Pour une comparaison équitable, nous utilisons les mêmes points. Ces points sont issus d'une détection subpixelique des coins du damier dans les images. La dernière colonne de ce tableau montre les résultats obtenus en utilisant le second type de mire composé de disques. Pour comparer, les deux types de mire apparaissent dans chaque image d'étalonnage (Fig. 2(b)). Les résultats montrent une meilleure précision avec le damier qu'avec les disques. En effet, avec les disques c'est leur centre de gravité dans l'image qui est considéré comme centre réel. Ceci n'est qu'une approximation dans les images perspectives comme dans les images omnidirectionnelles. Cependant, leur détection est plus simple et plus robuste à une mauvaise qualité d'image ou à des conditions difficiles qu'un damier.

mire	damier	damier	disques
méthode	AVV	Mei	AVV
p_x	460,33	460,33	459,24
p_y	459,65	459,65	458,11
u_0	635,99	635,99	632,94
v_0	490,19	490,19	489,90
ξ	1,14	1,14	1,14
$\mu_{\mathbf{u}}$	0,168	0,168	0,197
$\sigma_{\mathbf{u}}$	0,137	0,137	0,177
nb itérations	21	60	22

TABLE 1 – Paramètres estimés et comparaison. $\mu_{\mathbf{u}}$ est l'erreur de reprojection moyenne et $\sigma_{\mathbf{u}}$ son écart-type.

Les résultats d'étalonnage en utilisant l'AVV ou la toolbox de Mei sont identiques, à 10^{-4} près pour l'écart-type de l'erreur de reprojection en faveur de l'AVV.

Pour évaluer la robustesse de l'étalonnage par AVV par rapport à l'initialisation, le tableau 2 montre les résultats de convergence pour différentes valeurs initiales. La technique est robuste à d'importantes erreurs d'initialisation sur tous les paramètres. De plus, elle converge vers les valeurs optimales même en partant d'une erreur résiduelle initiale très importante : plus de mille pixels en moyenne.

p_x	p_y	u_0	v_0	ξ	μ_u init	itér
480	480	640	480	1	95	23
0	0	640	480	1	174	21
2500	2500	640	480	1	815	24
2500	0	640	480	1	473	29
480	480	0	0	1	601	22
0	0	0	0	1	610	23
2500	2500	0	0	1	1030	24
0	2500	0	0	1	1050	36
480	480	640	480	0	4069	NC
480	480	640	480	0,5	178	22
480	480	640	480	2,0	98	27

TABLE 2 – Robustesse à l'initialisation des paramètres intrinsèques. Quand l'estimation converge, les paramètres estimés sont ceux de la première colonne du tableau 1, avec la même erreur résiduelle. “ μ_u init” est l'erreur moyenne de reprojection initiale et la dernière colonne est le nombre d'itérations effectuées pour converger (NC : Non Convergence).

5.2 Etalonnage stéréoscopique perspective

Dans cette expérimentation, deux caméras perspectives (IDS μEye) sont utilisées avec différentes focales (8 mm et 6 mm) et avec la même résolution (752×480 pixels). Le banc stéréo possède un empattement (distance entre les deux caméras) de 30 cm. Il est admis que les axes principaux des deux caméras sont parallèles. Le tableau 3 montre les résultats d'étalonnage intrinsèque et extrinsèque obtenus avec notre méthode. Il compare également ces résultats avec ceux estimés par la toolbox de Bouguet [9]. L'étalonnage est réalisé en utilisant cinq paires d'images de damier (36 points par image) sans estimer les distorsions. Le même ensemble de points est utilisé pour les deux méthodes pour une comparaison équitable.

Les résultats d'étalonnage extrinsèque, meilleur moyen de comparaison entre les deux méthodes, sont clairement similaires avec une différence dans l'estimation de la translation quasiment nulle (inférieure au centième de millimètre) et valident notre approche.

	c_1		c_2	
	AVV	Bouguet	AVV	Bouguet
p_x	1122,57	1122,55	1418,43	1418,39
p_y	1122,17	1122,14	1413,74	1413,70
u_0	414,23	414,23	342,43	342,43
v_0	212,31	212,31	254,70	254,70
t_x	/	/	-0,305	-0,305
t_y	/	/	0,006	-0,006
t_z	/	/	0,056	0,056
θ_x	/	/	0,82	0,82
θ_y	/	/	2,61	2,61
θ_z	/	/	-0,24	-0,24

TABLE 3 – Comparaison des étalonnages pour un banc stéréo perspective. Les translations sont en mètres et les rotations en degrés.

5.3 Etalonnage stéréoscopique hybride : caméras fisheye et perspective

Première configuration. Nous présentons maintenant l'étalonnage d'un banc stéréo mixte (Fig. 3), composé d'une caméra fisheye modélisée par une sphère et une caméra perspective modélisée par un plan. La mire d'étalonnage est composée de disques de 4,75 cm de rayon et 4,6 cm d'espacement (Fig. 4(a) et 4(b)). Six paires d'images d'étalonnage sont utilisées avec 36 points pour chaque image.



FIGURE 3 – Premier capteur stéréo mixte.

Pour cette expérimentation, l'erreur moyenne de reprojection est de 0,170 pixels et 0,187 d'écart-type. Les deux caméras sont fixées sur un même support droit de telle sorte que leurs axes principaux soient parallèles (Tab. 4(a)). Le banc stéréo possède un empattement de 30 cm (mesurée manuellement) selon les axes X des deux caméras. L'erreur d'estimation de la translation est d'environ 6 mm (2%).

Pour évaluer la précision de l'étalonnage directement dans les images, des points de l'environnement sont sélectionnés dans l'image perspective (Fig. 4(a)). Leurs coniques épipolaires sont obtenues à partir de la matrice essentielle calculée grâce aux résultats d'étalonnage et sont tracées sur l'image fisheye (Fig. 4(b)). En grossissant l'image fisheye aux points d'intérêt correspondant à ceux de l'image perspective, il est clair qu'ils se trouvent sur leur conique épipolaire (Fig. 4). Cette observation qualitative montre la précision et la vraisemblance des paramètres d'étalonnage.

	fisheye	persp		fisheye	persp
p_x	482,11	1164,57	p_x	512,08	673,12
p_y	484,15	1170,25	p_y	507,87	679,59
u_0	344,92	385,70	u_0	325,94	337,34
v_0	242,97	218,47	v_0	241,15	212,02
ξ	1,22	/	ξ	1,05	/
t_x	/	-0,293	t_x	/	-0,075
t_y	/	0,006	t_y	/	-0,075
t_z	/	-0,010	t_z	/	0,029
θ_x	/	1,80	θ_x	/	0,27
θ_y	/	-0,69	θ_y	/	3,84
θ_z	/	1,89	θ_z	/	90,28

(a) (b)

TABLE 4 – Résultats d'étalonnage de deux bancs stéréo sphérique/perspective. Les translations sont en mètres et les rotations en degrés.

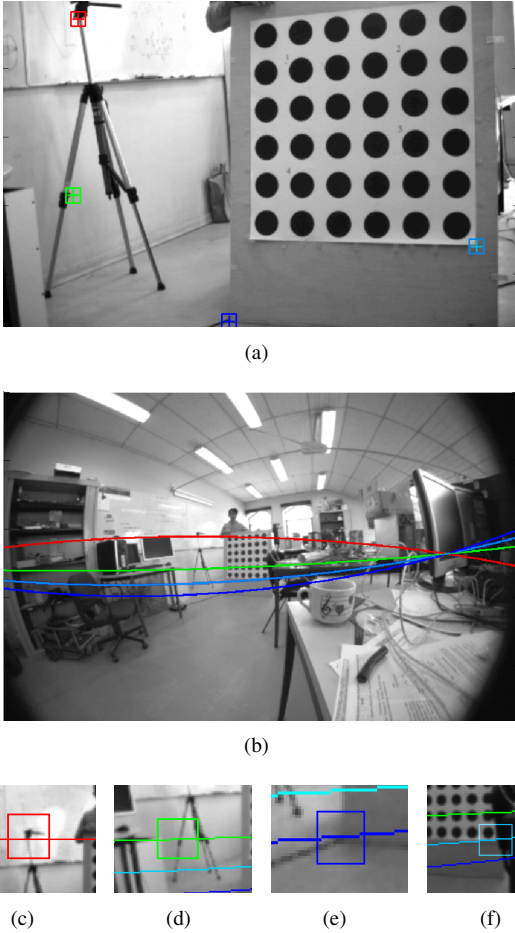


FIGURE 4 – Des points d'intérêt sont sélectionnés dans une image perspective (a) et les droites épipolaires correspondantes sont calculées et tracées dans l'image fisheye (b). (c), (d), (e) et (f) montrent plus précisément la partie de conique épipolaire correspondant au point sélectionnés dans l'image perspective.

Seconde configuration. Une seconde configuration du banc stéréo fisheye et perspective est utilisée avec une rotation de 90° entre deux caméras Sony XCD-V50 (Fig. 5).



FIGURE 5 – Capteur stéréo mixte dans une seconde configuration avec la caméra fisheye sur la droite de l'image et la caméra perspective sur la gauche.

Ce banc stéréo mixte est étalonné en utilisant sept paires d'images (640×480 pixels) avec trente-six points pour chaque image menant à 504 points. L'erreur moyenne de reprojection finale est de 0,337 pixel avec un écart-type de 0,353 pixel. Le tableau 4(b) montre les résultats d'étalonnage avec une rotation estimée de $90,28^\circ$ autour de l'axe Z , l'axe optique de la caméra de référence, menant à une erreur d'estimation de 0,3% pour ce paramètre.

5.4 Plane-sweeping

Un autre moyen de valider l'étalonnage stéréo mixte basé AVV est l'estimation d'altitude en utilisant le plane-sweeping [2], *i.e* la distance entre la caméra fisheye du banc stéréo hybride et le plan du sol. Cet algorithme consiste à estimer le plan du sol en minimisant une erreur globale entre les intensités d'une image de référence dans la vue perspective et la projection de la vue fisheye sur cette référence. Les paramètres intrinsèques des caméras et les poses relatives entre les deux caméras sont obtenues par notre méthode d'étalonnage simultané basé AVV. La normale au plan du sol est déduite d'une centrale inertielle. La figure 6 montre la précision de l'altitude estimée par rapport à un télémètre laser.

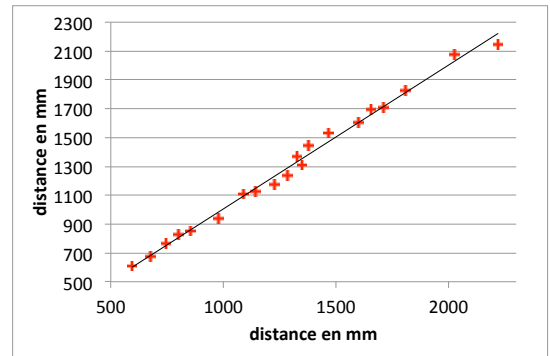


FIGURE 6 – Altitudes estimées par plane sweeping (croix) superposées à la droite des distances mesurées par laser en fonction d'elles-mêmes, c'est-à-dire les mesures idéales.

L'erreur moyenne d'estimation d'altitude sur 20 mesures, entre 55 cm et 215 cm est d'environ 3,2 cm, soit un ratio d'erreur sur les altitudes réelles de 2,4 %. Cette erreur d'estimation d'altitude a plusieurs sources comme le biais de la centrale inertielle, l'alignement de la centrale avec la caméra de référence qui est supposé parfait, le pas d'échantillonnage du plane-sweeping, ou encore l'étalonnage lui-même. La faiblesse de cette erreur montre donc clairement la précision de l'étalonnage.

5.5 Logiciel d'étalonnage stéréoscopique multi-modèles

Nous avons développé un logiciel implantant la méthode d'étalonnage stéréoscopique hybride présentée dans cet article. Il est multi-plateforme (Mac OS X 10.6, WindowsXp, Vista, Seven et Linux) puisque la plupart des toolbox d'étalonnage existantes sont développées avec Matlab. Notre logiciel est basé sur la librairie Qt de Nokia, OpenCV 2.x et la librairie [23]. La figure 7 montre l'interface utilisateur de ce logiciel. Il est possible de sélectionner différents modèles de caméra et pour chaque caméra sélectionnée, d'importer des fichiers d'images de mire d'étalonnage. Un onglet affiche les résultats d'étalonnage.

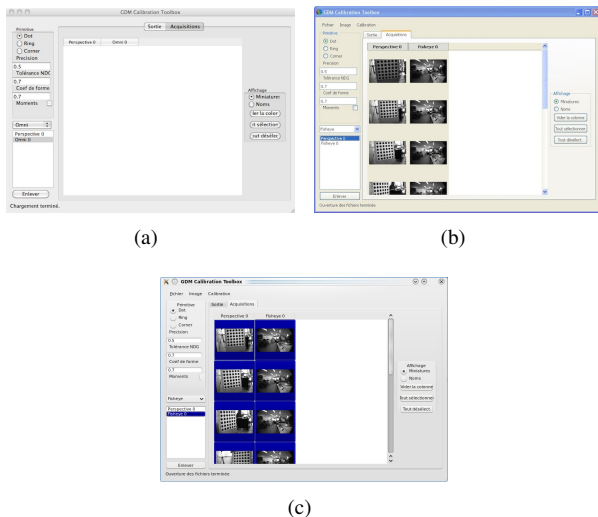


FIGURE 7 – Capture d'écran du logiciel Hyscas sous Mac OS X 10.6 (a), Windows Xp (b) et Linux (c).

Le temps de calcul dépend clairement du nombre de caméras utilisées, du nombre d'images d'étalonnages et du nombre de points. Plusieurs relevés de temps de calcul ont été réalisés en utilisant la version windows du logiciel, pour une caméra perspective, une caméra fisheye et un banc stéréo hybride composé des deux. Ces relevés correspondent au temps écoulé entre le dernier clic de sélection et l'affichage des résultats (Tab. 5).

6 Conclusion

Une nouvelle méthode générique d'étalonnage de capteur stéréoscopique hybride combinant différents modèles de

nombre de poses	perspective	fisheye	hybride
4	109 ms	115 ms	681 ms
6	244 ms	253 ms	1467 ms
8	469 ms	492 ms	2786 ms

TABLE 5 – Temps de calcul pour l'étalonnage en utilisant notre logiciel Hyscas. Chaque pose est une image pour une caméra, deux images pour deux caméras, etc. Trente-six points sont utilisés à chaque image.

projection a été présentée. Cette méthode permet l'étalonnage simultané intrinsèque et extrinsèque de bancs à N caméras de modèles différents. Les résultats montrent la réussite de l'étalonnage de divers types de bancs stéréo avec une grande précision, comme le montre clairement l'application de reconstruction 3D de plan par la technique du plane-sweeping.

Enfin, notre méthode d'étalonnage stéréo hybride nécessite uniquement la formulation de deux jacobiens pour prendre en compte un nouveau modèle de projection. Ce travail a mené à la réalisation d'un logiciel multi-plateforme permettant d'étalonner facilement une caméra ou un banc stéréo, hybride ou pas.

Références

- [1] P. STURM, "Mixing Catadioptric and Perspective Cameras", *In Proceedings of IEEE Omnidirectional Vision*, 2002.
- [2] D. EYNARD, P. VASSEUR, C. DEMONCEAUX, V. FREMONT, "UAV Altitude Estimation by Mixed Stereoscopic Vision", *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan, october 2010.
- [3] S. BAKER, S.K. NAYAR, "A theory of single-viewpoint catadioptric image formation", *Int. Journal of Computer Vision*, Vol. 35, No. 2, november 1999.
- [4] J.P. BARRETO, H. ARAUJO, "Issues on the Geometry of Central Catadioptric Image Formation", *Int. Conf. On Pattern Recognition*, Tampa, USA, december 2001.
- [5] A. RADGUI, C. DEMONCEAUX, E. MOUADDIB, D. ABOUTAJDINE, M. RZIZA, "An adapted Lucas-Kanade's method for optical flow estimation in catadioptric images", *The 8th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras - OMNIVIS*, Marseille, France, september 2008.
- [6] C. DEMONCEAUX, P. VASSEUR, "Omnidirectional Image Processing Using Geodesic Metric", *IEEE Int. Conf. on Image Processing*, Cairo, Egypt, november 2009.
- [7] E. MARCHAND, F. CHAUMETTE, "A New Formulation for Non-Linear Camera Calibration Using Virtual Visual Servoing", *INRIA Research Report, no 4096*, january 2001.

- [8] Z. ZHANG, "A flexible new technique for camera calibration", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, 2000.
- [9] J.-Y. BOUGUET, "Camera Calibration Toolbox for Matlab", http://www.vision.caltech.edu/bouguetj/calib_doc/, Intel, 2004.
- [10] C. MEI, P. RIVES, "Single View Point Omnidirectional Camera Calibration from Planar Grids", *IEEE Int. Conf. on Robotics and Automation*, Roma, Italy, april 2007.
- [11] J.P. BARRETO, H. ARAUJO, "Paracatadioptric Camera Calibration Using Lines", *IEEE Int. Conf. on Computer Vision*, Nice, France, october 2003.
- [12] D. SCARAMUZZA, R. SIEGWART "A New Method and Toolbox for Easily Calibrating Omnidirectional Cameras", *Workshop on Camera Calibration Methods for Computer Vision Systems*, Bielefeld, Germany, march 2007.
- [13] V. SUNDARESWARAN, R. BEHRINGER, "Visual servoing-based augmented reality", *IEEE Int. Workshop on Augmented Reality*, San Francisco, USA, november 1998.
- [14] E. MARCHAND, F. CHAUMETTE, "Virtual visual servoing : a framework for real-time augmented reality", *EUROGRAPHICS*, Saarbrücken, Germany, september 2002.
- [15] X. CHEN, J. YANG, A. WAIBEL, "Calibration of a Hybrid Camera Network", *IEEE Int Conference on Computer Vision*, Beijing, China, october 2003.
- [16] G. CARON, E. MARCHAND, E. MOUADDIB, "3D Model Based Pose Estimation For Omnidirectional Stereovision", *Int. Conf. on Intelligent Robots and Systems*, Saint-Louis, USA, october 2009.
- [17] X. YING, Z. HU, "Can We Consider Central Catadioptric Cameras and Fisheye Cameras within a Unified Imaging Model?" *European Conference on Computer Vision*, Vol. 1, Prague, Czech, 2004.
- [18] F. CHAUMETTE, S. HUTCHINSON, "Visual servo control, Part I : Basic approaches", *IEEE Robotics and Automation Magazine*, Vol. 13, no. 4, december 2006.
- [19] JOAO P. BARRETO, F. MARTIN, R. HORAUD, "Visual Servoing/Tracking Using Central Catadioptric Images", *Experimental Robotics VIII.*, B. Siciliano and P. Dario (Ed). Springer Verlag, 2003.
- [20] B. ESPIAU, F. CHAUMETTE, P. RIVES, "A new approach to visual servoing in robotics", *IEEE Trans. on Robotics and Automation*, Vol. 8, No. 3, juin 1992.
- [21] Y. MA, S. SOATTO, J. KOŠECKÁ, S. SASTRY, "An invitation to 3-D vision", *Springer*, 2004.
- [22] M.A. AMELLER, B. TRIGGS, L. QUAN, "Camera Pose Revisited - New Linear Algorithms", *European Conference on Computer Vision*, Dublin, Ireland, june 2000.
- [23] E. MARCHAND, F. SPINDLER, F. CHAUMETTE, "ViSP for visual servoing : a generic software platform with a wide class of robot control skills", *IEEE Robotics and Automation Magazine*, dec 2005.