



**HAL**  
open science

# Multiple camera types simultaneous stereo calibration

Guillaume Caron, Damien Eynard

► **To cite this version:**

Guillaume Caron, Damien Eynard. Multiple camera types simultaneous stereo calibration. IEEE Int. Conf. on Robotics and Automation, ICRA'11, May 2011, Shanghai, China. pp.2933-2938, 10.1109/ICRA.2011.5979975 . hal-01270416

**HAL Id: hal-01270416**

**<https://hal.science/hal-01270416v1>**

Submitted on 7 May 2024

**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.

# Multiple Camera Types Simultaneous Stereo Calibration

Guillaume Caron and Damien Eynard

**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, fisheye, catadioptric). 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<sup>1</sup>.

## I. INTRODUCTION

Stereoscopic sensors are generally composed by two cameras of the same type. Considering two perspective cameras leads to a limited field of view whereas two omnidirectional cameras have a non uniform and limited spatial resolution. Mixing different types of camera on a same rig can merge the wide field of view of an omnidirectional camera and the precision of a perspective camera [1], [2].

Perspective and omnidirectional cameras are central but only under certain conditions for the latter type [3]. Perspective cameras are generally modelled by the pinhole model. All central omnidirectional cameras can be modelled by the unified spherical projection model [4]. This model is also valid for perspective cameras but more complex than the pinhole model. For instance, the Lucas-Kanade optical flow computation was designed for perspective model and offers a better efficiency on perspective images than on omnidirectional ones [5]. This is due to the fact that the planar model preserves a square and regularly sampled neighbourhood around a point whereas it is not the case in the spherical model [6]. So for intensity based methods, there is no need to model a perspective camera by a sphere. Hence, using different models for central cameras is still useful with the main advantage of applying algorithms adapted for specific models.

The main application field of mixed systems are video-surveillance and autonomous navigation. Using a hybrid fisheye and perspective stereo rig, mixing spherical and planar perspective models, Eynard *et al.* [2] extend the plane-sweeping algorithm to estimate the altitude of a UAV.

Guillaume Caron and Damien Eynard are with Université de Picardie Jules Verne, MIS laboratory, Amiens, FRANCE; e-mail: {guillaume.caron, damien.eynard}@u-picardie.fr

<sup>1</sup><http://damien.eynard.free.fr/HYSCAS/>

The algorithm estimates the deepness of the ground plane minimising intensity differences on the planar image, but needs the knowledge of intrinsic and extrinsic parameters of the stereo rig. For this technique, attitude has to be computed before the plane estimation and it can be estimated using the fisheye view. This multi-task collaboration between two camera types is another example of the interest of mixing cameras on a same rig.

So to retrieve 3D information, such mixed stereo rigs have to be calibrated. The calibration of parametric models is a numerical process allowing to find parameter values of the model. This paper proposes a new approach to estimate simultaneously the projection parameters and relative poses of  $N$  different cameras composing a stereo rig.

Camera calibration is a common problem as well in perspective vision [7], [8] as in omnidirectional vision [10]–[12]. Marchand *et al.* [8] proposed a virtual visual servoing (VVS) based perspective camera calibration. VVS [13], [14] is a non-linear optimisation technique useful for pose computation and extendable to calibration, receiving the wide knowledge of visual servoing.

Works about calibration of stereo rigs generally tackle perspective cameras [9] (based on [7]) or only estimates extrinsic parameters in hybrid systems [15]. This paper generalises the VVS based central stereoscopic pose estimation method [16] to the calibration of  $N$  cameras defined by  $N$  different models, using points. The generic formulation of the problem allows to deal with any projection model.

After a recall of projection models, VVS based single camera calibration is tackled, followed by the calibration of stereoscopic system composed by different types of camera. Finally, calibration results are evaluated and a 3D plane estimation method is applied, using the calibration obtained by the method presented in this paper, in order to show the parameters accuracy.

## II. PROJECTION MODELS

### A. Perspective Model

The perspective projection (Fig. 1(a)) models pinhole cameras. Thus,  $\mathbf{X} = (X \ Y \ Z \ 1)^T$ , a 3D point expressed in the camera frame, is projected on the image plane as  $\mathbf{x} = (x \ y \ 1)^T$ :

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

$\mathbf{x}$  is the point on the normalised image plane and  $\mathbf{u} = (u \ v \ 1)^T$ , the pixelic point, is obtained by the rela-

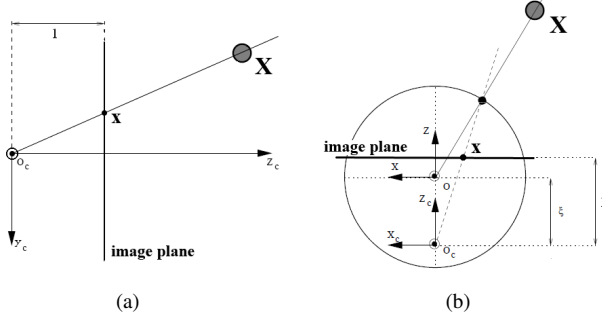


Fig. 1. (a) Perspective model used for pinhole cameras. (b) Spherical model used for omnidirectional cameras.

tion  $\mathbf{u} = \mathbf{K}\mathbf{x}$ .  $\mathbf{K}$  is the intrinsic parameters matrix, knowing parameters  $\gamma_1 = \{p_x, p_y, u_0, v_0\}$ :

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

So the full perspective projection of a 3D point to the pixelic image plane is  $pr_{\gamma_1}(\mathbf{X}) = \mathbf{K}pr(\mathbf{X})$ .

### B. Spherical Model

The unified spherical projection model [4] is suitable for central catadioptric cameras and is also valid for some fish-eye cameras [17]. Following the spherical model (Fig. 1(b)), a 3D point  $\mathbf{X}$  is first projected onto a unitary sphere, centred at  $(0 \ 0 \ \xi)^T$ . The obtained point is then perspectively projected on the image plane as  $\mathbf{x}$ , knowing intrinsic parameters  $\gamma_2 = \{p_x, p_y, u_0, v_0, \xi\}$ :

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

and  $\rho = \sqrt{X^2 + Y^2 + Z^2}$ . This is also known as a stereographic projection and a pixelic image point is obtained from a 3D point using  $pr_{\gamma_2}(\mathbf{X}) = \mathbf{K}pr_{\xi}(\mathbf{X})$  (eq. 2).

### C. Stereovision

For a rig of  $N$  cameras, each camera is modelled by its more suitable model: perspective model for perspective cameras, spherical model for omnidirectional cameras (catadioptric, fisheye). Poses of  $N - 1$  cameras  $c_j$  of the rig, modelled by homogeneous matrices  ${}^{c_j}\mathbf{M}_{c_1}$ , are defined w.r.t. the reference one,  $c_1$ .

## III. SINGLE CAMERA CALIBRATION

The pose computation problem using points can be defined as a VVS issue. Usually, image based visual servoing aims to move a camera to a desired pose minimising errors between current image features and features of the image acquired at the desired pose. The VVS virtualises the camera and starting from an initial pose, moves the virtual camera to make a perfect correspondence between the object forward projection, for its virtual pose, and the object in the real image. Camera model parameters can be simultaneously

estimated since the projection function is involved in the features motion in the pixelic image.

The virtual camera  $c$  is defined by its projection function  $pr_{\gamma_m}()$  and its position  ${}^c\mathbf{M}_o$ , w.r.t. a reference object or the world. With  $\mathbf{r}$ , the pose vector, the method estimates the real pose minimising the error  $\Delta$  between detected points  $\mathbf{u}_k^*$  and current points obtained by forward projection for the current pose  $\mathbf{u}_k(\mathbf{r})$ :

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

${}^o\mathbf{X}_k$  is the  $k$ -th 3D point expressed in the object coordinate system, as the calibration target, for instance ( $\mathbf{X} = {}^c\mathbf{M}_o \circ \mathbf{X}$ ).

The error to be regulated is, hence,  $\mathbf{e} = \mathbf{u}(\mathbf{r}) - \mathbf{u}^*$ . Imposing an exponential decoupled decrease of the error, the features motion is linked to the virtual camera by  $\dot{\mathbf{e}} = -\lambda\mathbf{e}$ , only depending on  $\dot{\mathbf{u}}$  [8]:

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

$\mathbf{L}_{\mathbf{x}} = \frac{\partial \mathbf{x}}{\partial \mathbf{r}}$  is known as the pose interaction matrix related to a perspective normalised image plane point [18] as well as to a normalised omnidirectional image plane point [19]. Actually,  $\frac{\partial \mathbf{x}}{\partial \mathbf{r}} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} \frac{\partial \mathbf{X}}{\partial \mathbf{r}}$  and it means that to compute a camera pose with another model, one has to express an image point as a function of a 3D point to compute the Jacobian  $\frac{\partial \mathbf{x}}{\partial \mathbf{X}}$ . Jacobian  $\frac{\partial \mathbf{X}}{\partial \mathbf{r}}$  [20] does not depend on any camera or projection model.

Intrinsic parameters Jacobians are detailed below.

The VVS control law for calibration using  $p$  images is then [8]:

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

with  $\mathbf{v}^i$ , the camera pose velocity vector associated to image  $i$ ,  $\dot{\gamma}_m$  the time variation of intrinsic parameters (unique for the images set) and  $\mathbf{u}^{*i}$ , the points set detected in image  $i$ .  $\mathbf{H}^+$  is the left pseudo inverse of  $\mathbf{H}$  which is:

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

with  $\mathbf{r}^i$ , the pose vector linked to image  $i$ . For brevity, there is no particular notation to show the used projection model to compute  $\mathbf{L}_{\mathbf{u}}^i$ .  $\frac{\partial \mathbf{u}}{\partial \gamma_m}$  depends on the projection model and is different if the perspective model is used with  $\gamma_1$  [8] or the spherical model:

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

Poses are then updated using the exponential map of  $se(3)$  [21] using  ${}^c\mathbf{M}_o^{t+1} = {}^c\mathbf{M}_o^t e^{[\mathbf{v}]}$  and intrinsic parameters are updated by  $\gamma_m^{t+1} = \gamma_m^t + \dot{\gamma}_m$ .

Non-linear optimisation methods need an initial guess. Each camera parameters set is coarsely initialised with  $u_0$ ,

$v_0$  at the image centre,  $p_x, p_y$  at the half vertical size of the image (mirror radius, for catadioptric cameras).  $\xi$  is initialised with its known theoretical values [4], depending on the omnidirectional camera type. The pose of the calibration grid w.r.t. a camera is initialised with a linear perspective camera pose method [22] designed for pinhole cameras, and the adaptation we did of this method for spherical cameras. VVS has proven to be robust to bad initial guesses for perspective intrinsic calibration [8], that is why these coarse initialisations can be used.

#### IV. STEREOSCOPIC CALIBRATION

Extending the mono camera approach, the optimisation criterion becomes [16], for a rig of  $N$  cameras:

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

with  $pr_{\gamma_{m_j}}$ , the projection function of camera  $j$  ( $m_j = 1$  or  $m_j = 2$ , depending on the used projection model). Considering  $\mathbf{L}_j$ , the camera  $j$  pose interaction matrix, the stereo pose interaction matrix is then [16]:

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

${}^{c_j}\mathbf{V}_{c_1}$  is the twist transformation matrix between velocity vectors of camera 1  ${}^{c_1}\mathbf{v}$  and camera  $j$   ${}^{c_j}\mathbf{v}$ .  ${}^{c_j}\mathbf{R}_{c_1}$  and  ${}^{c_j}\mathbf{t}_{c_1}$  are the rotation and translation blocs of  ${}^{c_j}\mathbf{M}_{c_1}$  and  $[{}^{c_j}\mathbf{t}_{c_1}]_{\times}$ , the skew-symmetric matrix of  ${}^{c_j}\mathbf{t}_{c_1}$ . So, with  ${}^{c_j}\dot{\mathbf{u}}_i$ , the set of time variation of camera  $j$  features in image  $i$ , the stereoscopic calibration virtual control law for  $p$  shots is:

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

where  $\mathbf{v}_1^i$  is the stereo rig pose velocity vector, for each set of  $N$  images and  $\mathbf{v}_{1,j}$ , the relative pose velocity vector between camera  $j$  and camera 1. Considering  $\mathbf{L}^i$ , the interaction matrix of the stereo rig pose  $i$  and  $\mathbf{L}_{1,j}^i$ , the interaction matrix of relative pose between camera 1 and camera  $j$  for pose  $i$ , the expression of  $\mathbf{H}_S$  is:

$$\mathbf{H}_S = \begin{pmatrix} \begin{array}{ccc|ccc} 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \mathbf{L}_{1,2}^1 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & \mathbf{L}_{1,N}^1 \end{array} & \begin{array}{ccc} \frac{\partial {}^{c_1}\mathbf{u}_1}{\partial \gamma_{m_1}} & 0 & \dots & 0 \\ 0 & \frac{\partial {}^{c_2}\mathbf{u}_1}{\partial \gamma_{m_2}} & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & \dots & 0 & \frac{\partial {}^{c_N}\mathbf{u}_1}{\partial \gamma_{m_N}} \end{array} \\ \hline \begin{array}{ccc|ccc} \vdots & \dots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \dots & 0 & \mathbf{L}^p & \mathbf{L}_{1,2}^p & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & \mathbf{L}_{1,N}^p \end{array} & \begin{array}{ccc} \frac{\partial {}^{c_1}\mathbf{u}_p}{\partial \gamma_{m_1}} & 0 & \dots & 0 \\ 0 & \frac{\partial {}^{c_2}\mathbf{u}_p}{\partial \gamma_{m_2}} & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & \dots & 0 & \frac{\partial {}^{c_N}\mathbf{u}_p}{\partial \gamma_{m_N}} \end{array} \end{pmatrix}.$$

This generic formulation of simultaneous hybrid stereo calibration allows to deal with any projection model  $m$  if a pixelic point can be related to a 3D point and Jacobians  $\frac{\partial \mathbf{u}}{\partial \mathbf{X}}$  and  $\frac{\partial \mathbf{u}}{\partial \gamma_m}$  expressed.

#### V. RESULTS

This section evaluates the presented calibration approach, comparing results obtained with our software (available as an installer and a demo video: <http://damien.eyn.free.fr/HYSCAS/>) to existing toolboxes, for omnidirectional camera [10] and perspective stereo rig [9]. The evaluation of VVS based single perspective camera calibration is done in [8].

##### A. Single omnidirectional camera calibration

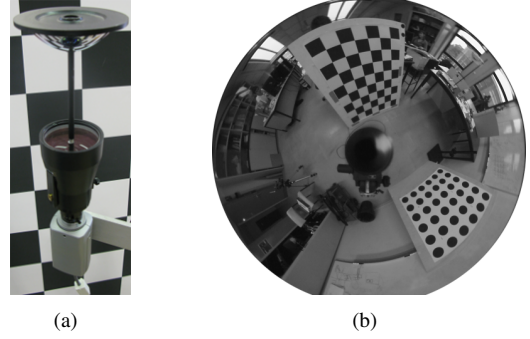


Fig. 2. (a) Catadioptric sensor. (b) An image used for calibration (two calibration pattern types).

In this section, calibration results are evaluated for a single camera modelled by a sphere. Comparison is made between results obtained using our calibration software, implementing the method proposed in this paper, and the Mei's toolbox [10].

The calibration grid is a 82 cm  $\times$  105 cm chessboard of 7  $\times$  9 squares. Another calibration pattern, available in our software, is made of 36 dots placed in a 80 cm  $\times$  80 cm square.

This experiment shows the calibration of a catadioptric camera composed by a camera (Sony DFW-SX910) and a Remote Reality (Fig. 2(a)) paraboloid optic. Six images of 1280  $\times$  960 pixels resolution are used for a total of 288 points.

Calibration is done without taking into account distortions and results are reported in the table I. For a fair results comparison, the same set of subpixelic image points, extracted from the corners of the chessboard target, is used. The last column of this table shows results obtained using the second target type composed by dots. For the comparison, the two pattern types appear in each calibration image (Fig. 2(b)). Results show higher accuracy with corners than with dots for which their centre of gravity in the image is considered as their real centre. This is only an approximation in perspective images as well as in omnidirectional images. However, their detection is easier and more robust to low image quality or tough conditions than corners.

Calibration results using VVS or Mei’s Toolbox are the same, up to  $10^{-4}$  of a pixel for standard deviation of backprojection error in favour of VVS.

pattern	corners	corners	dots
method	VVS	Mei	VVS
$p_x$	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
$\xi$	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 iterations	21	60	22

TABLE I

ESTIMATED PARAMETERS AND COMPARISON.  $\mu_{\mathbf{u}}$  IS THE MEAN BACKPROJECTION ERROR AND  $\sigma_{\mathbf{u}}$  ITS STANDARD DEVIATION.

To evaluate the robustness of the calibration using VVS w.r.t. initialisation, table II shows convergence results under various initial values. The technique is robust to important initialisation errors on all parameters and converges to the optimal values even starting from a huge initial residual error: more than a thousand pixels in a mean.

$p_x$	$p_y$	$u_0$	$v_0$	$\xi$	$\mu_{\mathbf{u}}$ init	iter
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 II

INTRINSIC PARAMETERS ROBUSTNESS TO INITIALISATION. WHEN CONVERGING, ESTIMATED PARAMETERS ARE THE ONE OF THE FIRST COLUMN OF THE TABLE I, WITH THE SAME RESIDUAL ERROR. “ $\mu_{\mathbf{u}}$  INIT” IS THE INITIAL MEAN REPROJECTION ERROR AND THE LAST COLUMN IS THE NUMBER OF ITERATIONS TO CONVERGE (NC: NO CONVERGENCE).

## B. Stereoscopic calibration

1) *Perspective stereo rig calibration:* In this experiment we use two perspective cameras (IDS  $\mu Eye$ ) with different focal length (8 mm and 6 mm) but same resolution ( $752 \times 480$  pixels). The stereo rig has a baseline of 30 cm and we assume that both camera principal axes are parallel. Table III shows intrinsic and extrinsic calibration results obtained with our method and compares them with the ones estimated using the Bouguet’s toolbox [9]. Calibration is done using five pairs of chessboard images (36 points by image) without estimating distortions. The same set of image points is used for both methods for a fair comparison.

	$c_1$		$c_2$	
	VVS	Bouguet	VVS	Bouguet
$p_x$	1118.27	1122.55	1412.82	1418.39
$p_y$	1118.01	1122.14	1408.10	1413.70
$u_0$	414.91	414.23	342.99	342.43
$v_0$	216.24	212.31	253.14	254.70
$t_x$	/	/	-0.304	-0.305
$t_y$	/	/	0.007	-0.006
$t_z$	/	/	0.054	0.056
$\theta_x$	/	/	2.64	2.64
$\theta_y$	/	/	0.64	0.85
$\theta_z$	/	/	2.69	2.70

TABLE III

PARAMETERS COMPARISON FOR PERSPECTIVE STEREO CALIBRATION. TRANSLATIONS ARE IN METERS AND ROTATIONS IN DEGREES.

The standard deviation of backprojection error with Bouguet’s toolbox for this calibration is 0.150 pixel whereas with our method, it is 0.119 pixel.

Extrinsic calibration results are clearly similar with a difference in translation estimation of around 1 mm and validate our approach.

### 2) Mixed stereo rig calibration:

a) *First configuration of the hybrid stereo rig:* The calibration of a mixed stereo rig (Fig. 3), composed by a fisheye camera modelled by a sphere and a perspective camera modelled by a plane, is now presented. The calibration pattern is composed of dots with 4.75 cm radius and 4.6 cm spacing (Fig. 4(a) and 4(b)). Six pairs of calibration images are used with 36 points in each image.



Fig. 3. First mixed stereo sensor.

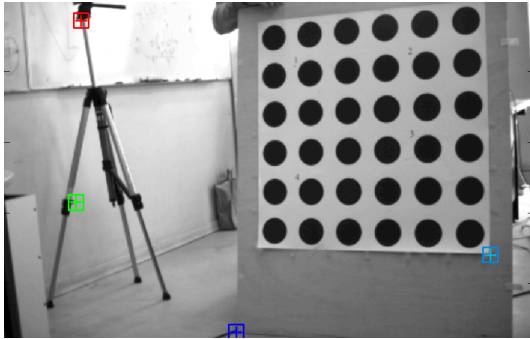
For this experiment, backprojection mean error is 0.170 pixel and 0.187 of standard deviation. Both cameras are fixed on the same straight stand and in order to have their principal axes parallel, which is estimated (Tab. IV(a)). The stereo rig has a baseline of manually measured 30 cm along the  $X$ -axes of both cameras. So the estimated translation error is about 6 mm (2%).

	fisheye	perspective		fisheye	perspective
$p_x$	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
$\xi$	1.22	/	$\xi$	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
$\theta_x$	/	1.80	$\theta_x$	/	0.27
$\theta_y$	/	-0.69	$\theta_y$	/	3.84
$\theta_z$	/	1.89	$\theta_z$	/	90.28

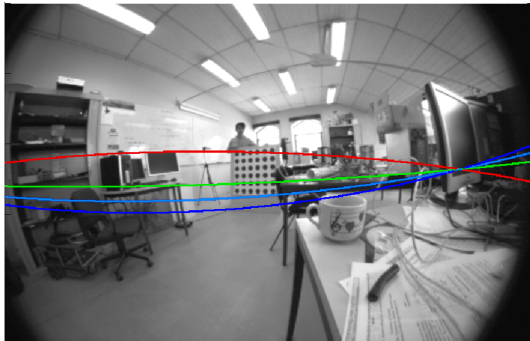
TABLE IV

CALIBRATION RESULTS OF TWO SPHERICAL / PERSPECTIVE STEREO RIG. TRANSLATIONS ARE IN METERS AND ROTATIONS IN DEGREES.

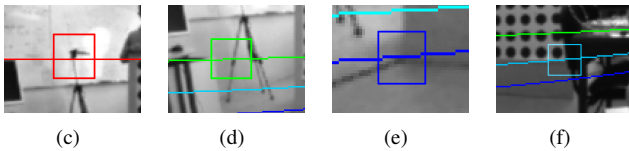
To evaluate the calibration precision directly in images, points of the environment are selected in the perspective image (Fig. 4(a)). Their corresponding epipolar conics are obtained with the essential matrix computed thanks to calibration results and are plotted on the fisheye image



(a)



(b)



(c)

(d)

(e)

(f)

Fig. 4. Interest points are selected in a perspective image (a) and corresponding epipolar lines are computed in the fisheye image (b). (c), (d), (e) and (f) show more precisely the part of epipolar conics corresponding to the point selected in the perspective image.

(Fig. 4(b)). When zooming the fisheye image on interest points corresponding to the selected ones in the perspective image, it is clear that they are on their epipolar conic (Fig. 4). This qualitative observation shows the precision and likelihood of the calibrated parameters.

b) *Second configuration of the hybrid stereo rig:* A second configuration of a fisheye and perspective stereo rig is used with a rotation of  $90^\circ$  between two Sony XCD-V50 cameras (Fig. 5).



Fig. 5. Mixed stereo sensor in a second configuration with the fisheye camera on the right side of the image and the perspective on the left side.

This mixed stereo rig is calibrated using 7 pairs of images ( $640 \times 480$  pixels) with 36 points for each image leading to 504 points. The mean final reprojection error is 0.337 pixels with 0.353 pixels of standard deviation. Table IV(b) shows the calibration results with a  $90.28^\circ$  estimated rotation around the  $Z$ -axis, the optical axis of the reference camera, leading to a 0.3% of estimation error for this parameter.

### C. Plane-sweeping

An application to validate the VVS based mixed stereo calibration is the estimation of altitude using plane-sweeping [2], *i.e.* the distance between the fisheye camera of the hybrid stereo rig and the ground plane. This algorithm consists of estimating the ground plane minimising a global error between a reference image on the perspective view and the projected fisheye view on this reference. Cameras intrinsic parameters and relative pose between both cameras is known from our VVS based simultaneous calibration and the ground plane normal vector is deduced from an IMU. Figure 6 shows the altitude estimation accuracy w.r.t. a laser telemeter. Altitude estimation mean error over 20 measurements and between 55 cm and 215 cm is about 3.2 cm, *i.e.* an error ratio over real altitudes of about 2.4 %.

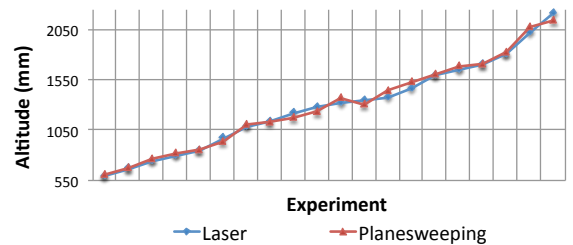


Fig. 6. Comparison between estimated altitude by plane-sweeping and measured altitude using a laser telemeter. Plane-sweeping is done with a fisheye and perspective cameras stereo rig calibrated using the VVS based calibration method of this paper.

## VI. MULTI-MODEL STEREO CALIBRATION SOFTWARE

We have developed a software implementing the hybrid stereo calibration method presented in this paper. It is cross-platform (Mac OS X 10.6, Windows XP, Vista, Seven and Linux) as most of available calibration toolboxes developed using Matlab. Our software is based on the Qt library from Nokia and the ViSP library [23]. Figure 7 shows the user interface of this software. It is possible to select different camera models and for each selected camera to import image files of calibration pattern. A tab displays calibration results.

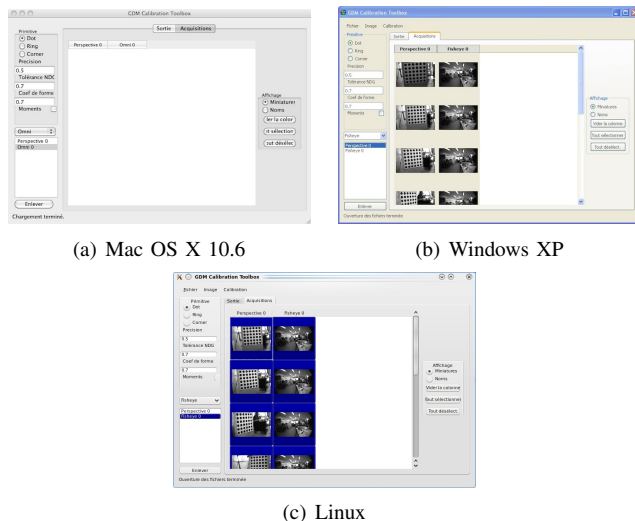


Fig. 7. Software screenshots.

Processing time clearly depends on the used number of cameras, calibration images and points. Several processing time measures has been made using the windows version of our software, for a perspective camera, a fisheye camera and a hybrid stereo rig composed by both of them (36 points for each image). These measures are the elapsed time between the last selection click and the display of results (Tab. V).

number of poses	perspective	fisheye	hybrid
4	109 ms	115 ms	681 ms
6	244 ms	253 ms	1467 ms
8	469 ms	492 ms	2786 ms

TABLE V

CALIBRATION PROCESSING TIME OF OUR SOFTWARE. EACH POSE IS AN IMAGE FOR ONE CAMERA, TWO IMAGES FOR TWO CAMERAS

## VII. CONCLUSION

A new generic multi model calibration method of stereoscopic hybrid sensor has been presented. The method allows the simultaneous intrinsic and extrinsic calibration of a  $N$  cameras rig. Results show the achievement of calibration leading to high accuracy of 3D plane reconstruction in a plane-sweeping technique. Finally, our generic hybrid stereo calibration method only needs to formulate two Jacobians to take into account a new projection model. This work has

been implemented in a cross-platform software which allows to easily calibrate a camera or a stereo rig, hybrid or not.

## REFERENCES

- [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*, Taipeh, 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] Z. ZHANG, "A flexible new technique for camera calibration", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, 2000.
- [8] E. MARCHAND, F. CHAUMETTE, "A New Formulation for Non-Linear Camera Calibration Using Virtual Visual Servoing", *INRIA Research Report, no 4096*, january 2001.
- [9] J-Y. BOUGUET, "Camera Calibration Toolbox for Matlab", [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](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.