Calibration of Non-Overlapping Cameras - Application to Vision-Based Robotics

Pierre Lébraly[13]
Pierre.LEBRALY@lasmea.univ-bpclermont.fr
Omar Ait-Aider[13]
Omar.AIT-AIDER@lasmea.univ-bpclermont.fr
Eric Royer[23]
Eric.ROYER@lasmea.univ-bpclermont.fr
Michel Dhome[13]
Michel.Dhome@lasmea.univ-bpclermont.fr

[1] Clermont Université, Université Blaise Pascal, LASMEA, BP 10448, F-63000 CLERMONT-FERRAND
[2] Clermont Université, LASMEA, BP 10448, F-63000 CLERMONT-FERRAND
[3] CNRS, UMR 6602, LASMEA, F-63177 AUBIERE

Abstract

Multi-camera systems are more and more used in vision-based robotics. An accurate extrinsic calibration is usually required. In most of cases, this task is done by matching features through different views of the same scene. However, if the cameras fields of view do not overlap, such a matching procedure is not feasible anymore.

This article deals with a simple and flexible extrinsic calibration method, for non-overlapping camera rig. The aim is the calibration of non-overlapping cameras embedded on a vehicle, for visual navigation purpose in urban environment. The cameras do not see the same area at the same time. The calibration procedure consists in manoeuvring the vehicle while each camera observes a static scene. The main contributions are a study of the singular motions and a specific bundle adjustment which both reconstructs the scene and calibrates the cameras. Solutions to handle the singular configurations, such as planar motions, are exposed. The proposed approach has been validated with synthetic and real data.

1 Introduction

Recently, some navigation methods have been proposed for mobile robots using a single camera and natural landmarks. However, a single camera does not offer sufficient robustness against outdoor illumination problems, *e.g.*, overexposure. To alleviate this problem, a solution consists in mounting several cameras on the vehicle, for example, one looking at the front and an other at the back. If the sun is in front of one camera, an other gives some useful information. Then, the localization process requires to calibrate the multi-camera system. But few approaches have been proposed in order to calibrate a set of cameras with non-overlapping fields of view. It is possible to use an additional camera moving around the static cameras. Calibration is then obtained with a 3D reconstruction up to a scale factor.

Alternatively, a planar mirror can be used to create an overlap between views. The extrinsic calibration of a multi-camera system is done using a known [15] or unknown [17]

BMVC 2010 doi:10.5244/C.24.10

geometry calibration pattern. By contrast, non-overlapping static cameras can be calibrated by tracking a mobile object, the missing trajectory information is estimated in the unobserved areas [2, 22]. In the same way, Lamprecht *et al*. [16] calibrate a moving multi-camera system. A static calibration object is tracked and a prior knowledge about the system speed is required. An other approach proposed by Esquivel *et al*. [7] consists in using the rigidity constraint between the coupled cameras. First, each camera trajectory is computed. Second, the relative poses are deduced since they do not change over the time. A additional work on rotation averaging is done by Dai *et al*. [4].

Once this rigidity constraint is estimated, minimal pose problems [18] are solved to locate the multi-camera system [3, 9, 12, 13, 23], with an Extended Kalman Filter [11, 21] or with a structure from motion algorithm [23, 24, 25]. Some of these articles claim to use a manual extrinsic calibration. However, approximate angles or distance measurements (using a ruler) might be too inaccurate. In this article, we propose a flexible strategy to achieve an accurate extrinsic calibration of a mobile set of rigidly linked cameras.

This article can be seen either as an improvement of the algorithm proposed by Esquivel *et al*. [7], or as an extension of the classical bundle adjustment proposed by Triggs *et al*. [26] for rigidly linked cameras (with possible totally non-overlapping fields of view) for a calibration purpose. After a system overview (Section 2), we introduce a linear initialization of the extrinsic parameters for general or singular motions (Section 3). Section 4 is dedicated to a non-linear refinement, which optimizes the extrinsic parameters, the scene and the trajectory of the multi-cameras rig. Finally (Section 5), the results validate our approach with both synthetic and real data.

2 System overview

We consider at least $N_{cam} \geq 2$ rigidly linked cameras, with known intrinsic parameters and non-overlapping fields of view. The cameras C_i are assumed to be synchronized. K motions are performed. Thus, $K+1$ is the number of poses of the multi-camera rig over the time. C_i^k represents the i^{th} camera at time k. Each pose of the camera C_i^k is expressed relative to its first pose at time $k=0$. Let T_i^k be the homogeneous transformation of C_i coordinate systems from time 0 to time k, for $k \in [\![0..K]\!]$ and $i \in [\![1..N_{cam}]\!]$ (see Figure 1). In the same way, ΔT_i is the unknown homogeneous transformation from C_1 to C_i coordinate system. Each homogeneous transformation T is represented with a rotation R and a translation $\mathbf{t}$ such that:

$$T = \begin{pmatrix} R & \mathbf{t} \\ 0_{1\times 3} & 1 \end{pmatrix} \tag{1}$$

While the camera rig is moving, each camera C_i observes a static scene S_i of 3D points. s_i^k are the projected points of the scene S_i, in the image plane of C_i^k.

The calibration (see Figure 2) consists first in computing each camera trajectory T_i^k (see below). Second, the extrinsic parameters ΔT_i are linearly initialized (Section 3), and finally refined as well as the scenes thanks to a specific bundle adjustment (Section 4).

Trajectory estimation: First, the trajectory of each camera is independently computed. To do so, if the scene geometry is totally unknown, a structure and motion algorithm (as described in [10]) is used. In contrast, if the scene geometry is approximately known, each camera pose is first initialized by the method described by Dementhon [5]. Then, for each camera C_i, both the camera's trajectory T_i^k and the scene S_i are refined by a standard bundle

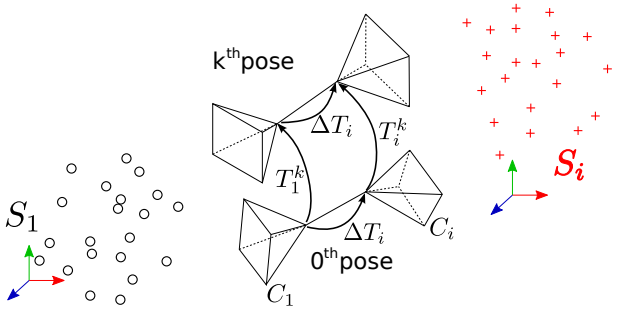

Figure 1: Non-overlapping cameras rig moving along a static scene.

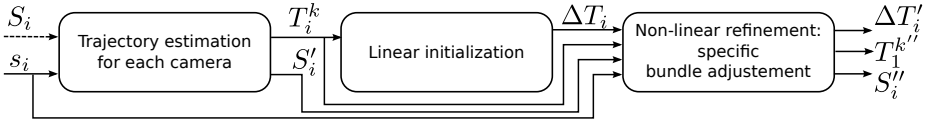

Figure 2: Extrinsic calibration scheme of linked cameras. The successive estimations are designated by the prime symbol.

adjustment algorithm [26]. In the following, we suppose that for each scene S_i, a distance measure is available between two points. Thus, the whole system has the same scale factor and the relative scale δ_i between the scene S_1 and the scene S_i is set to one. Notice that if these measures are not available, δ_i could be added to the unknown parameters to be recovered.

3 Linear initialization

3.1 General motions

An estimate of the relative pose ΔT_i can be obtained linearly from the trajectory T_i^k of each camera. The rigidity assumption and simple changes of basis lead to:

$$\forall i \in [\![1..N_{cam}]\!], \forall k \in [\![1..K]\!],\ T_1^k \Delta T_i = \Delta T_i T_i^k \tag{2}$$

These are equations of the form $AX = XB$, where X is the unknown matrix. The same equations are expressed for the hand-eye calibration problem [6], whose solutions represent the rotations by unit quaternions [7] or by a 3×3 orthogonal matrix [1]. Using the definition (1) and the property (3), the equation (2) splits into two parts:

$$vec(ABC) = (A \otimes C^\top) vec(B) \tag{3}$$

$$\text{Using (1), (2)} \Leftrightarrow \forall i \in [\![1..N_{cam}]\!], \forall k \in [\![1..K]\!], \begin{cases} R_1^k \Delta R_i = \Delta R_i R_i^k & \text{(4a)} \\ R_1^k \Delta \mathbf{t}_i + \mathbf{t}_1^k = \Delta R_i \mathbf{t}_i^k + \Delta \mathbf{t}_i & \text{(4b)} \end{cases}$$

$$\Leftrightarrow \forall i \in [\![1..N_{cam}]\!], \forall k \in [\![1..K]\!], \begin{cases} (I_9 - R_1^k \otimes R_i^k) vec(\Delta R_i) = 0_{9\times 1} & \text{(5a)} \\ (I_3 - R_1^k) \Delta \mathbf{t}_i = \mathbf{t}_1^k - \Delta R_i \mathbf{t}_i^k & \text{(5b)} \end{cases}$$

Where $\otimes$ denotes the matrix Kronecker product, *vec* vectorizes a matrix into a column vector by stacking the transposed rows of the matrix and I_n is the identity matrix of size n.

As suggested in [1], we opt for a matrix representation (which allows to express the required additional equations in section 3.2) and a two-step solution: solve first the rotations ΔR_i, then the translations $\Delta \mathbf{t}_i$. With a general motion assumption, the equation (2) admits a unique solution. First, for each camera C_i, the rotation ΔR_i is estimated thanks to equation (6):

$$(5a) \Rightarrow \forall i \in [\![2..N_{cam}]\!], \underbrace{\begin{pmatrix} I_9 - R_1^1 \otimes R_i^1 \\ \vdots \\ I_9 - R_1^k \otimes R_i^k \\ \vdots \\ I_9 - R_1^K \otimes R_i^K \end{pmatrix}}_{=L_i} vec(\Delta R_i) = 0_{9K\times 1} \tag{6}$$

For general motions, L_i has rank 8. Let v_i be a vector of the null space of L_i and let V_i be the 3×3 matrix such that $v_i = vec(V_i)$, then:

$$\Delta R_i = V_i \; sign(det(V_i)) |det(V_i)|^{-\frac{1}{3}} \tag{7}$$

Second, the translation is estimated thanks to the full-rank equation (8), obtained from (5b) for each motion k:

$$(5b) \Rightarrow \forall i \in [\![2..N_{cam}]\!], \begin{pmatrix} I_3 - R_1^1 \\ \vdots \\ I_3 - R_1^k \\ \vdots \\ I_3 - R_1^K \end{pmatrix} \Delta \mathbf{t}_i = \begin{pmatrix} \mathbf{t}_1^1 - \Delta R_i \mathbf{t}_i^1 \\ \vdots \\ \mathbf{t}_1^k - \Delta R_i \mathbf{t}_i^k \\ \vdots \\ \mathbf{t}_1^K - \Delta R_i \mathbf{t}_i^K \end{pmatrix} \tag{8}$$

3.2 Singular motions

3.2.1 Overview

This section outlines the critical motions where the equations (6) and (8) become singular and cannot be used anymore. First of all, the pure translations are singular motions, and have already been studied by Esquivel [7]. However, some other singular cases should be outlined. As this linear extrinsic calibration problem can be formulated with the same equations as the hand eye calibration problem, we can deduce singular cases from the previous studies [1, 8]. The degenerate cases are also analysed by Kim and Chung [11], for a similar topic: motion and structure from stereo images without stereo correspondence. As a result, the singular cases raise when the axes $\mathbf{n}_1^k$ of the rotations R_1^k are parallel for every $k \in [\![1..K]\!]$. In other words, the singular motions are the orbit of the first pose (for $k = 0$) under the action of the group G, where G is composed of the screw motions with collinear axis.

The singular motions are rotations and screw motions about an axis (when the axes $\mathbf{n}_1^k$ are the same), or pure translations, planar motions and screw motions with parallel axes (when the axes $\mathbf{n}_1^k$ are different). For these singular motions, the calibration is partial: only some parameters can be recovered. Table 1 counts the number of observable degrees of freedom of the extrinsic parameters.

Planar motions can be seen as several screw motions about parallel axes $\mathbf{n}_1^k$ with zero translations along these axes. As illustrated by Table 1 and Figure 3, one dimension of the extrinsic translation $\Delta\mathbf{t}_i$ is not observable along the plane normal $\mathbf{n}$. All we can say is that $\Delta\mathbf{t}_i$ is of the form:

$$\Delta\mathbf{t}_i = \mathbf{t}_{i\perp} + \alpha_i\mathbf{n} \tag{9}$$

where α_i is any scalar value and $\mathbf{t}_{i\perp}$ is the projection of $\Delta\mathbf{t}_i$ in the plane of the motion.

Motions	Axes of rotation	ΔR_i	$\Delta\mathbf{t}_i$
1) Rotations and screw motions about an axis $\mathbf{a}$	same ($\mathbf{a} = \mathbf{n}_1^k$)	2	2
2) Pure translations along an axis $\mathbf{a}$	no rotation	2	0
3) Planar and screw motions about several axes	unequal and parallel	3	2
4) 3D (general case)	unequal	3	3

Table 1: Observability of extrinsic parameters with respect to K motions of the non-overlapping cameras. Notice that for the motion 1), if the axis $\mathbf{a}$ go through all camera optical centers, $\Delta\mathbf{t}_i$ isn't observable. For the motions 1) and 2), if a screw motion about $\mathbf{a}$ is composed with the extrinsic parameters and applied to S_i, the observations are not changed.

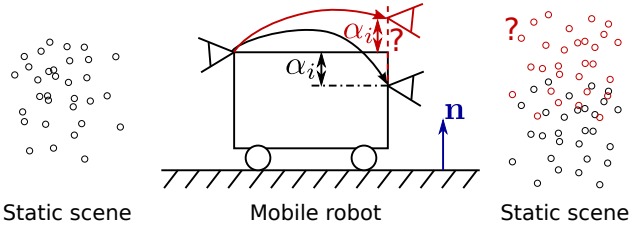

Figure 3: Singular planar motion: the camera's relative height α_i (with respect to the plane of motion) is not observable. The observations do not differ between the real and the estimated systems (respectively in black and in red).

3.2.2 Solutions to handle the singular motions

This section focuses on planar motions — the most usual singular case for a mobile robot. An initial estimate of the extrinsic rotations is provided, then we show a practical way to initialize the camera relative heights. Lastly, the specific bundle adjustment is applied (§4).

Rotation estimation: This case can neither be solved by the method of Esquivel *et al.* for general motion [7, §4.1] nor by our linear formulation (§3.1), where the first step of the estimation of the rotation (6) fails because of a rank deficiency of the linear equations system. Indeed, as R_1^k and R_i^k have the same eigenvalues, the maximum rank value of matrix L_i is 6.

Thus, for a planar motion, new constraints must be added to previous initialization (6) in order to calculate the rotations ΔR_i. Without loss of generality, a non-zero rotation and non-zero translation motion of the camera C_1 is defined as the first motion T_1^1. Then, noting $\mathbf{t}_i'^k = (I_3 - R_i^k)\mathbf{t}_i^1 - (I_3 - R_i^1)\mathbf{t}_i^k$ and assuming $\mathbf{t}_i'^k \neq 0_{3\times1}$, we get the following full rank system:

$$\forall i \in [\![2..N_{cam}]\!], \forall k \in [\![2..K]\!], \begin{pmatrix} I_3 \otimes \mathbf{t}_i'^k \\ I_3 \otimes (\mathbf{n}_i^k)^\top \\ I_3 \otimes (\mathbf{t}_i'^k \times \mathbf{n}_i^k)^\top \end{pmatrix} vec(\Delta R_i) = \begin{pmatrix} \mathbf{t}_1'^k \\ \mathbf{n}_1^k \\ \mathbf{t}_1'^k \times \mathbf{n}_1^k \end{pmatrix} \tag{10}$$

Proof: First, as R_1^1 and R_1^k commute because their axes of rotation are parallel, we get:

$$\forall k \in [\![1..K]\!],\ (I_3 - R_1^k)(I_3 - R_1^1) - (I_3 - R_1^1)(I_3 - R_1^k) = 0_{3\times 3} \tag{11}$$

Second, using the equation (11), and starting from equation (5b) for $k = 1$ and $k \neq 1$, the expression $(I_3 - R_1^k)(5b)_{k=1} - (I_3 - R_1^1)(5b)_{k\neq 1}$ simplifies to (12):

$$(5b)\&(11) \Rightarrow \quad (I_3 - R_1^k)(\mathbf{t}_1^1 - \Delta R_i \mathbf{t}_i^1) - (I_3 - R_1^1)(\mathbf{t}_1^k - \Delta R_i \mathbf{t}_i^k) = 0_{9\times 1} \tag{12}$$

$$(12) \Leftrightarrow \quad (I_3 - R_1^k)\Delta R_i \mathbf{t}_i^1 - (I_3 - R_1^1)\Delta R_i \mathbf{t}_i^k = (I_3 - R_1^k)\mathbf{t}_1^1 - (I_3 - R_1^1)\mathbf{t}_1^k \tag{13}$$

$$(13)\&(4a) \Leftrightarrow \quad \Delta R_i (I_3 - R_i^k)\mathbf{t}_i^1 - \Delta R_i (I_3 - R_i^1)\mathbf{t}_i^k = (I_3 - R_1^k)\mathbf{t}_1^1 - (I_3 - R_1^1)\mathbf{t}_1^k \tag{14}$$

$$\Leftrightarrow \quad \Delta R_i \mathbf{t}_i'^k = \mathbf{t}_1'^k \tag{15}$$

Moreover, the axes $\mathbf{n}_i^k$ of the rotations R_i^k verify:

$$\forall i \in [\![1..N_{cam}]\!], \forall k \in [\![1..K]\!],\ \Delta R_i \mathbf{n}_i^k = \mathbf{n}_1^k \tag{16}$$

Furthermore, the planar motion involves that $\mathbf{t}_i'^k$ and $\mathbf{n}_i^k$ are linearly independent. Hence, the cross product $\mathbf{t}_i'^k \times \mathbf{n}_i^k$ is linearly independent from them. As a result, using the property (3), we obtain the full rank system (10). □

In practice, the systems (6) and (10) are concatenated to estimate ΔR_i. However, as planarity conditions are not perfect with noisy data, the equation (11) is almost verified. Hence, it is advisable to orthogonalize the estimated rotation matrix thanks to a SVD.

Translation estimation: For a planar motion, the translation remains partially calculable. Thus, after the initialization, the undefined relative heights α_i of the cameras are set to prior rough value if present, or else 0 m (instead of using an wrong value). In practice, at least one non-coplanar pose can avoid the singularity. For example, if the camera rig is placed on a mobile robot, a bump could be located on the robot trajectory.

An other solution consists in moving the multi-camera system such that at least one point of the scene S can be seen by each camera at different times. During the acquisition process, the camera rig is moved such that the observed scenes are permuted. Hence, with known matchings between the views, the relative heights α_i are calculable. As a result, for a planar motion, both the rotations ΔR_i and translations $\Delta \mathbf{t}_i$ are fully calculable.

We will show how to complete the previous initialization (§3.1) with new constraints (provided by the scene permutation). They are valid either if the motions are planar or not. $T_{i\to\sigma(i)}^k$ is the homogeneous transformation between the camera C_i at time 0 to the camera $C_{\sigma(i)}$ at time k, with the permutation σ. For $N_{cam} > 2$, a system of coupled equations is obtained. The equations obtained for $N_{cam} = 2$, $\sigma(1) = 2$ and $\sigma(2) = 1$ follow:

$$T_{1\to 2}^k (\Delta T_2)^{-1} = \Delta T_2 T_{2\to 1}^k \tag{17}$$

$$\Leftrightarrow \begin{cases} R_{1\to 2}^k (\Delta R_2)^\top = \Delta R_2 R_{2\to 1}^k & \text{(18a)} \\ -\Delta R_2 R_{2\to 1}^k + \mathbf{t}_{1\to 2}^k = \Delta R_2 \mathbf{t}_{2\to 1}^k + \Delta \mathbf{t}_2 & \text{(18b)} \end{cases}$$

$$\Leftrightarrow \begin{cases} (I_9 - (R_{1\to 2}^k \otimes R_{2\to 1}^k)\Phi_9) vec(\Delta R_2) = 0_{9\times 1} & \text{(19a)} \\ (I_3 + \Delta R_2 R_{2\to 1}^k)\Delta \mathbf{t}_2 = \mathbf{t}_{1\to 2}^k - \Delta R_2 \mathbf{t}_{2\to 1}^k & \text{(19b)} \end{cases}$$

where Φ_9 is the permutation matrix such that $vec(M^\top) = \Phi_9 vec(M)$.

The equations (19a) can be added to equations (6), but it does not rise enough the rank of the system in case of planar motion contrary to equation (10). Finally, the equations (8) and (19b) are concatenated to linearly and fully estimate the translation $\Delta\mathbf{t}_2$.

4 Specific bundle adjustment

During the previous linear formulation, the trajectories of the cameras were supposed to be known, and were not optimized. In the opposite, this section is dedicated to an algorithm which optimizes the trajectory of the camera rig, the extrinsic parameters ΔT_i and the scene geometry. Notice that the similarity with the hand-eye calibration is not valid in this section.

Even if the N_{cam} previous bundle adjustments are the maximum-likelihood estimators for each camera, it is not the case for a multi-camera rig. Indeed, when the bundle adjustments are considered independently, the whole system is over-parametrized. For example, Triggs *et al*. [26] propose a bundle adjustment for one or several cameras. However, the optimized parameters are the whole cameras poses relative to a global coordinate system[1]. In our case, the cameras are rigidly linked. As proposed by King [14] for a stereo camera, a minimal parametrisation is advisable; only the *master camera* C_1 poses are expressed relative to a global coordinate system ($6(K+1)$ parameters for $\mathbf{t}_1^k$ and R_1^k). The other cameras are expressed relative to the master camera ($6(N_{cam}-1)$ parameters for $\Delta\mathbf{t}_i$ and ΔR_i, represented by local Euler angles).

Notice that if the world coordinate system were defined as the first pose of the master camera, the most minimal parametrization would be reached. But in practice, some local minima can occur due to this assumption (the reader can refer to [26] for more details about gauge freedom). Thus none camera pose is defined as the world coordinate system.

The algorithm is initialized with the linear estimate of the extrinsic parameters and the union of all scene points, expressed in the same coordinate system. Let M be the number of all the 3D points. Finally, $6(K+1)+6(N_{cam}-1)+3M$ parameters are optimized by a Levenberg Marquardt algorithm during the minimization of the reprojection errors.

Under the assumption of Gaussian pixel noise, the proposed algorithm is a maximum-likelihood estimator. Notice that the scene's occlusions are handled. Moreover, each 3D point can be seen by any camera. Thus the permutations of the scenes (see §3.2.2) are also handled.

5 Validation

5.1 Results with synthetic data

The experimental protocol is the following, with $N_{cam}=2$ cameras. First of all, 10 poses are synthesized for a $3D$ motion without U-turn. The scene points S_i are created in front of each camera C_i.This ground truth is used to calculate the image points s_i^k. Then, s_i^k are subject to an Additive White Gaussian Noise (AWGN) with a standard deviation σ. The distance measures — used to fix the scale factor — are also subject to an AWGN with a standard deviation of 0.5 mm. Finally, our calibration algorithm is applied on the noisy data.

During this experiment, the transformation ΔT_2 corresponds to the translation (0.1m 0.1m -2m) and the rotation from yaw, pitch and roll angles ($-179°$, $-4°$, $171°$).

[1] The intrinsic parameters can also be optimized for auto-calibration purpose.

For every σ value, 10 measures have been performed. The calibration accuracy is given by both the norm $\|dT\|$ of the translation error vector (and also $\|dT\|/\|\Delta T_2\|$ express as percentage of the magnitude of the camera baseline) and the angular error dR whose expression is given by:

$$dR = d(\hat{R}, R^*) = \arccos\left(\frac{trace(\hat{R}^\top R^*) - 1}{2}\right) \tag{20}$$

where R^* is the ground truth rotation and $\hat{R}$ is the estimated rotation. Figure 4 shows the performances of the algorithm, using a 11 points unknown 3D scene (about 1.5 m by 0.5 m) for S_1 and S_2 at distance around 1 m from each camera C_i. In case of $3D$ motion, the algorithm proposed in [7] gives the same results as our linear initialization. Both methods are outperformed by the specific bundle adjustment (the translation and the rotation estimation are about respectively four and three times more accurate).

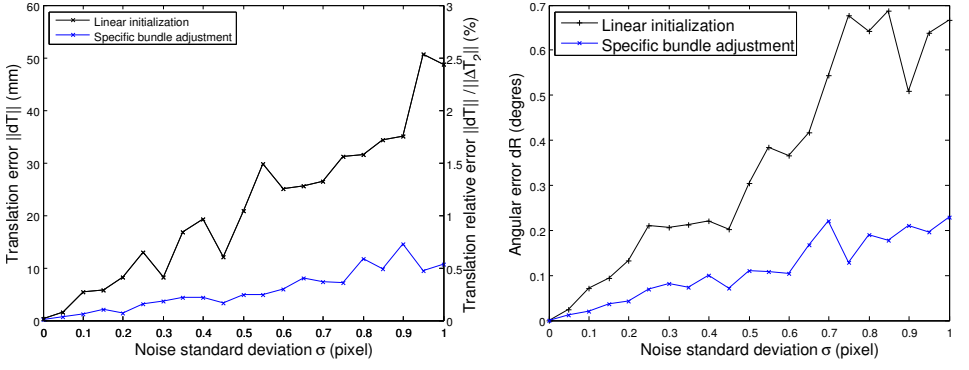

Figure 4: Results with synthetic data: extrinsic calibration error with respect to noise level. With real data, the noise standard deviation is below 0.1 pixel.

5.2 Results with real data

First, the accuracy of the proposed calibration method is demonstrated with an overlapping stereo camera system. Second, the feasibility of the approach is illustrated with non-overlapping cameras embedded on a vehicle. For scene features, we use circular landmarks similar to [19] and developed in [17, §VII], with a black bullseye and a circular code (see Figure 5). These landmarks guarantee a subpixel detection. An automatic feature detector allows to accurately estimate the landmarks' center and their label. Consequently the matchings between the views are automatically performed. The resolution of the images is 1600x1200. We use an 5.6 mm focal length for the cameras, with a pixel width of 4.4 μm.

5.2.1 Stereo camera

Here we calibrate an overlapping stereo camera with non-overlapping assumptions: each camera observes only one scene at the same time (illustrated on Figure 5a by the blue arrows or the green crossed arrows in case of permutation). S_1 and S_2 are supported by, respectively, the left and the right board (see Figure 5b). Our algorithm's results are compared to a classical stereo calibration algorithm, with overlapping fields of view and a 3D motion (see Table 2). For each case, 15 images are acquired. The camera baseline is about 22 cm. All the standard deviations of the reprojection errors are about 0.03 pix.

(a) (b) (c)

Figure 5: Calibration of stereo camera with non-overlapping assumption, for overlapping cameras (a) and (b) and embedded non-overlapping cameras (c).

Fields of view	Overlapping	Non-overlapping	
Motions	Planar	3D	Planar	
Scenes permutation	n/a	no	no	yes
Translation error $\|dT\|$ (mm)	0.14	0.41	60.39	0.08
Angular error $dR(^\circ)$	0.008	0.011	0.038	0.011

Table 2: Comparison of extrinsic calibration accuracy with respect to overlapping assumptions, motions and scene permutation. The reference is a standard stereo calibration, with overlapping fields of view and a 3D motion.

Results of Table 2 shows that our calibration with non-overlapping assumption is nearly as accurate as a classical algorithm with overlapping fields of view (see column 3D). In case of quasi-planar motion, if the scenes are not permuted (see 4^{th} column), then the singularity involves expected inaccurate results due to the unobservable translation. Therefore, the scene permutation (see last column) enables to have the same accuracy than a classical algorithm (below 0.5 *mm* for the translation — *i.e.* $\|dT\|/\|\Delta T_2\| < 0.2\%$ — and about 0.01° for the rotation).

5.2.2 Embedded multi-camera rig

Two cameras are now embedded on a mobile robot : one at the front, another at the back (the camera baseline is about 22 cm, see Figure 5c). According to Pless [20], this is the best design for two standard cameras for ego-motion purpose. The main advantages are the vehicle symmetry, and the robustness against outdoor illumination problems for the localization process.

The robot manoeuvres on a planar floor, between two scenes. A U-turn is applied during the acquisition to permute the observed scenes. Furthermore, the scene's occlusions are handled, therefore wide-based views are added to get an accurate scene reconstruction. The algorithm returns the extrinsic parameters. The standard deviation of the reprojection errors is about 0.09 *pix*. Notice that if the equation (19b) were not added to initialize the extrinsic translation, an error of about 1.6 *m* would occur along the normal of the motion's plane.

To validate the calibration, we consider one couple of images which was not used in the calibration process. Using the previous estimation of both the 3D points and the extrinsic parameters, the pose of the multi-camera system is computed. The standard deviation of the reprojection errors is about 0.05 *pix* (the same order of magnitude, with a zero mean value).

6 Conclusion

A flexible method for extrinsic calibration of mobile multi-camera system has been proposed and validated with both synthetic and real data. First, extrinsic parameters have been linearly initialized using the trajectory of each camera. Second, the specific bundle adjustment, which is a maximum-likelihood estimator, refined the scene, the trajectory of the multi-cameras rig and the extrinsic parameters. The singular motions have been outlined. Even if the motions are planar, the scene permutation solution allows to get a full calibration. Moreover, the results show that our calibration, with non-overlapping assumption, is as accurate as classical algorithm with overlapping fields of view.

In our future work, the circular landmarks will be replaced by interest points to compare the algorithms accuracy. To achieve an efficient implementation, we will analytically express the sparse Jacobian of the projection function with respect to the optimized parameters.

References

[1] N. Andreff, R. Horaud, and B. Espiau. Robot hand-eye calibration using structure-from-motion. *The International Journal of Robotics Research*, 20(3):228, 2001.

[2] N. Anjum, M. Taj, and A. Cavallaro. Relative position estimation of non-overlapping cameras. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*, volume 2, 2007.

[3] B. Clipp, J.H. Kim, J.M. Frahm, M. Pollefeys, and R.I. Hartley. Robust 6dof motion estimation for non-overlapping, multi-camera systems. In *IEEE Workshop on Applications of Computer Vision, 2008. WACV 2008*, pages 1–8, 2008.

[4] Y. Dai, J. Trumpf, H. Li, N. Barnes, and R. Hartley. Rotation Averaging with Application to Camera-Rig Calibration. *Computer Vision–ACCV 2009*, pages 335–346, 2010.

[5] Daniel F. Dementhon and Larry S. Davis. Model-based object pose in 25 lines of code. *International Journal of Computer Vision*, 15:123–141, 1995.

[6] M. Dhome. *Visual Perception Through Video Imagery*. Wiley-ISTE, 2009.

[7] S. Esquivel, F. Woelk, and R. Koch. Calibration of a Multi-camera Rig from Non-overlapping Views. In *Pattern Recognition: 29th DAGM Symposium, Heidelberg, Germany, September 12-14, 2007, Proceedings*, page 82. Springer-Verlag New York Inc, 2007.

[8] I. Fassi and G. Legnani. Hand to sensor calibration: A geometrical interpretation of the matrix equation AX= XB. *Journal of Robotic Systems*, 22(9):497, 2005.

[9] J.M. Frahm, K. Koser, and R. Koch. Pose Estimation for Multi-camera Systems. In *Pattern recognition: 26th DAGM Symposium, Tübingen, Germany, August 30-1 September 2004: proceedings*, page 286. Springer-Verlag New York Inc, 2004.

[10] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[11] J.H. Kim and M.J. Chung. Absolute motion and structure from stereo image sequences without stereo correspondence and analysis of degenerate cases. *Pattern Recognition*, 39(9):1649–1661, 2006.

[12] J.H. Kim, R. Hartley, J.M. Frahm, and M. Pollefeys. Visual Odometry for Non-Overlapping Views Using Second-Order Cone Programming. *Lecture Notes in Computer Science*, 4844:353, 2007.

[13] Jun-Sik Kim, Myung Hwangbo, and T. Kanade. Motion estimation using multiple non-overlapping cameras for small unmanned aerial vehicles. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 3076 –3081, 19-23 2008. doi: 10.1109/ROBOT.2008.4543678.

[14] BA King. Optimisation of Bundle Adjustments for Stereo Photography. *International Archives Of Photogrammetry And Remote Sensing*, 29:168–168, 1993.

[15] R.K. Kumar, A. Ilie, J.-M. Frahm, and M. Pollefeys. Simple calibration of non-overlapping cameras with a mirror. *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*, pages 1 –7.

[16] B. Lamprecht, S. Rass, S. Fuchs, and K. Kyamakya. Extrinsic camera calibration for an on-board two-camera system without overlapping field of view. In *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, pages 265 –270, sept. 2007. doi: 10.1109/ITSC.2007.4357679.

[17] Pierre Lébraly, Clément Deymier, Omar Ait-Aider, Eric Royer, and Michel Dhome. Flexible Extrinsic Calibration of Non-Overlapping Cameras Using a Planar Mirror: Application to Vision-Based Robotics. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, IROS 2010.

[18] Hongdong Li, Richard I. Hartley, and Jae-Hak Kim. A linear approach to motion estimation using generalized camera models. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2008*. IEEE Computer Society, 2008.

[19] Diego López de Ipi na, Paulo R. S. Mendonça, and Andy Hopper. Trip: A low-cost vision-based location system for ubiquitous computing. *Personal Ubiquitous Comput.*, 6(3):206–219, 2002. ISSN 1617-4909. doi: http://dx.doi.org/10.1007/s007790200020.

[20] R. Pless. Using many cameras as one. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2003. Proceedings*, pages 587–593, 2003.

[21] ME Ragab, KH Wong, JZ Chen, and MMY Chang. EKF Based Pose Estimation using Two Back-to-Back Stereo Pairs. In *IEEE International Conference on Image Processing, 2007. ICIP 2007*, volume 6, 2007.

[22] A. Rahimi, B. Dunagan, and T. Darrell. Simultaneous calibration and tracking with a network of non-overlapping sensors. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–187 – I–194 Vol.1, 27 2004. doi: 10.1109/CVPR.2004.1315031.

[23] T. Sato, S. Ikeda, and N. Yokoya. Extrinsic camera parameter recovery from multiple image sequences captured by an omni-directional multi-camera system. *Computer Vision-ECCV 2004*, pages 326–340, 2004.

[24] H. Stewenius and K. Åström. Structure and motion problems for multiple rigidly moving cameras. *European Conference on Computer Vision, ECCV 2004*, pages 252–263, 2004.

[25] Henrik Stewénius, Magnus Oskarsson, and Kalle Åström. Reconstruction from planar motion image sequences with applications for autonomous vehicles. In *Scandinavian Conf. on Image Analysis*, pages 609–618, 2005.

[26] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment-a modern synthesis. *Vision algorithms: theory and practice*, pages 153–177, 2000.