INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Euclidean Bundle Adjustment Independent on Camera Intrinsic Parameters

Ezio Malis — Adrien Bartoli

N° 4377

Décembre 2001

THÈMES 4 et 3

ISSN 0249-6399 ISRN INRIA/RR--4377--FR+ENG

INRIA
SOPHIA ANTIPOLIS

Euclidean Bundle Adjustment Independent on Camera Intrinsic Parameters

Ezio Malis *, Adrien Bartoli †

Thèmes 4 et 3 — Simulation et optimisation
de systèmes complexes — Interaction homme-machine,
images, données, connaissances
Projets Icare et Movi

Rapport de recherche n° 4377 — Décembre 2001 — 27 pages

Abstract: This research report presents a bundle adjustment technique which is independent on camera intrinsic parameters. Standard bundle adjustment techniques consider camera intrinsic parameters as unknowns in the optimization process. The scheme proposed in this report differs from previous standard techniques since unknown camera intrinsic parameters are computed as a function of the 3D structure and the camera pose. Considering less unknowns in the optimization process produces a faster algorithm which is more adapted to real-time applications. Computationally expensive metric reconstruction, using for example several zooming cameras, considerably benefits from an intrinsics-free bundle adjustment. Indeed, experimental results, obtained using both simulated data and real images, show that the algorithm we propose not only considerably reduces the computational cost but also improves the accuracy of the metric reconstruction.

Key-words: structure from motion, Euclidean reconstruction, bundle adjustment, projective invariance.

* Ezio Malis is with the ICARE Project at INRIA Sophia-Antipolis
† Adrien Bartoli is with the MOVI Project at INRIA Rhône-Alpes

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Ajustement euclidien de faisceaux indépendant des paramètres intrinsèques des caméras

Résumé : Ce rapport de recherche présente une technique d'ajustement de faisceaux indépendante des paramètres intrinsèques des caméras. Les techniques standards d'ajustement de faisceaux considèrent ces paramètres comme des inconnues dans le processus d'optimisation. Le principe proposé dans ce rapport est différent. En effet, les paramètres intrinsèques des caméras sont calculés à partir de la structure 3D et la position des caméras. Un nombre inférieur d'inconnues est alors impliqué dans le processus d'estimation, lequel converge alors plus rapidement. Les reconstructions métriques coûteuses, par exemple à partir de plusieurs caméras à zoom variable, tirent un bénéfice non négligeable de cet ajustement de faisceaux indépendant des données de calibrage. En effet, les résultats expérimentaux sur données simulées et réelles montrent qu'une réduction notable du temps de calcul ainsi qu'une légère amélioration de la précision de la reconstruction métrique sont ls obtenues.

Mots-clés : invariance projective, ajustement de faisceaux, reconstruction métrique.

Contents

1 Introduction **4**

2 Theoretical background **5**
2.1 Perspective projection 5
2.2 Camera model 6

3 Projective spaces invariant to camera parameters **6**
3.1 Invariance by projective transformation 7
3.2 Computing camera parameters "a posteriori" 9

4 Metric reconstruction from several images **13**
4.1 Computational cost of the non-linear optimization 14
4.2 Initialization of the non-linear optimization 14
4.3 Imposing geometric constraints 15
4.4 Expected improvements over standard bundle adjustment 15
4.4.1 Fixed intrinsic parameters 15
4.4.2 Varying focal length and principal point. 16
4.4.3 Time-varying structure and motion tracked by several zooming cameras. 16

5 Experimental Results **16**
5.1 Simulated Data 16
5.1.1 Fixed intrinsic parameters 17
5.1.2 Varying focal length and principal point. 19
5.1.3 Time-varying structure and motion tracked by several zooming cameras. 22
5.2 Real Images 24
5.2.1 Fixed intrinsic parameters 24
5.2.2 Varying focal length and principal point. 25

6 Conclusion **26**

1 Introduction

Bundle adjustment is the final refinement step of a reconstruction process. Standard metric reconstruction methods start with a projective reconstruction of the scene obtained from several images [4] [6] [2]. The projective reconstruction is then updated to an Euclidean reconstruction after camera self-calibration [11] [8] [12]. Finally, the visual reconstruction is refined using bundle adjustment in the Euclidean space [10]. Standard bundle adjustment [13] [7] is based on the joint estimation of the 3D structure, of the camera extrinsic parameters (i.e. the rotation and the translation of the cameras with respect to the absolute frame) and of the camera intrinsic parameters (i.e. the focal length, the principal point, the skew and the aspect ratio). Therefore, bundle adjustment is a high dimensional optimization problem which involves a large number of unknowns. The objective is achieved by minimizing a cost function which contains the error between the features measured from the images and the features predicted from a model. The cost function is generally non-linear and it can eventually be subject to non-linear constraints. Obviously, the speed of an optimization process is directly related to the number of unknowns and to the form of the cost function. Many applications of metric reconstruction, and in particular robotic vision applications, need a computational speed close to real-time. Thus, it is extremely important to have a fast algorithm despite the large number of images used for the reconstruction. The primary objective of this research report is to investigate a new fast bundle adjustment method to be used in real-time applications. As an example of robotic application needing a huge amount of computation, consider the measurement and the tracking over time of the 3D model of a moving non-rigid object using several zooming cameras. In order to accomplish such difficult applications, the bundle adjustment that we propose is based on parameterizing the reconstruction problem by using the 3D structure and the camera extrinsic parameters but without considering the camera intrinsic parameters despite the fact that they are unknowns. The key idea is to partially work in a projective space which is invariant to camera intrinsic parameters. A general approach to invariance theory for systematically extracting invariants is given by Lie group theory [14]. Obviously, there are various ways in which invariance to camera parameters can be obtained. For example, simple invariants to focal length and principal point have been proposed in [15] for robotics applications. Contrarily to previous work, in this report we do not limit the invariance to focal length and principal point but we exploit invariance to all the camera intrinsic parameters, following an idea proposed in [9], where invariance is obtained by a change of projective coordinates. Thus, the number of unknowns in the bundle adjustment optimization problem can be considerably reduced. Even if the computational cost in terms of CPU time is the key issue of this report, a second objective of our work is to have a fast convergence of the optimization process without loosing the accuracy of the solution. Obviously, the accuracy of 3D measurements increases with the number of cameras and/or the number of images. If the number of cameras increases, so does the number of unknown camera parameters in a standard bundle adjustment. On the contrary, the bundle adjustment proposed in this report does not add any supplementary unknown camera parameter in that case. Consequently, we eliminate the possibility to find, in the non-linear optimization, local minima which may be due to correlation between

intrinsic and extrinsic parameters of the camera. Experimental results, obtained using simulated data and real images, show that our algorithm not only considerably reduces the computational cost but also improves the accuracy of the reconstruction.

2 Theoretical background

In this section, we give the necessary theoretical background and we introduce the mathematical notations used in this report.

2.1 Perspective projection

Let $\mathcal{F}_0$ be the absolute frame attached to a rigid object. Let the 3D point $\mathcal{C}$ be the origin of frame $\mathcal{F}$ and the center of projection (see Figure 1). Let the plane of projection π be parallel to the plane $(\overrightarrow{x}, \overrightarrow{y})$. A 3D point with homogeneous coordinates $\boldsymbol{\mathcal{X}}_i = (X_i, Y_i, Z_i, 1)$ (with respect to frame $\mathcal{F}_0$) is projected to the point $\mathbf{m}_i = (x_i, y_i, 1)$ on π:

$$\zeta_i \mathbf{m}_i = \left[\begin{array}{cc} \mathbf{R} & \mathbf{t} \end{array} \right] \boldsymbol{\mathcal{X}}_i \,, \tag{1}$$

where ζ_i is a positive scalar factor, $\mathbf{R}$ and $\mathbf{t}$ are respectively the rotation and the translation between frame $\mathcal{F}_0$ and $\mathcal{F}$. Let $\mathbf{r} = \theta \mathbf{u}$ be the (3×1) vector containing the axis of rotation $\mathbf{u}$ and the angle of rotation θ $(0 \leq \theta < 2\pi)$. If $[\mathbf{r}]_\times$ is the skew symmetric matrix associated to vector $\mathbf{r}$, then $\mathbf{R} = \exp([\mathbf{r}]_\times)$. Let $\boldsymbol{\xi} = (\mathbf{t}, \mathbf{r})$ be the (6×1) vector containing global coordinates of an open subset $\mathcal{S} \subset \mathbb{R}^3 \times SO(3)$. The vector $\mathbf{m}_i(\boldsymbol{\xi}, \boldsymbol{\mathcal{X}}_i)$ depends on the position $\boldsymbol{\xi}$ of the camera and on the 3D coordinates $\boldsymbol{\mathcal{X}}_i$ of the point.

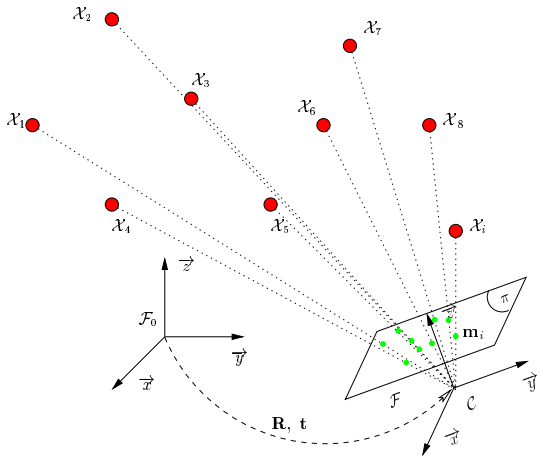


Figure 1: Perspective projection of 3D points in the frame $\mathcal{F}$.

2.2 Camera model

Pinhole cameras perform a perspective projection of a 3D point. However, the point $\mathbf{m}_i$ is not directly measured by the cameras. The information given by a pinhole camera is an image point $\mathbf{p}_i$. The point $\mathbf{p}_i = (u_i, v_i, 1)$ observed in the image $\mathcal{I}$, taken at the position $\mathcal{F}$, depends on the camera internal parameters:

$$\mathbf{p}_i = \mathbf{K}\mathbf{m}_i \ , \qquad \mathbf{p}_i \in \mathcal{I}(\boldsymbol{\xi}, \mathbf{K}) \ , \tag{2}$$

where:

$$\mathbf{K} = \begin{bmatrix} f_m k_u & -f_m k_u \cot(\phi) & u_0 \\ 0 & f_m k_v / \sin(\phi) & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

u_0 and v_0 are the coordinates of the principal point (pixels), f_m is the metric focal length (meters), k_u et k_v are the magnifications respectively in the $\overrightarrow{u}$ and $\overrightarrow{v}$ directions (pixels/meters), and ϕ is the angle between these axes. The image $\mathcal{I}(\boldsymbol{\xi}, \mathbf{K})$ depends on the intrinsic parameters of the camera. For the sake of simplicity, we will also use the following equivalent notation of the camera parameters:

$$\mathbf{K} = \begin{bmatrix} f_p & s & u_0 \\ 0 & r f_p & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where $f_p = f_m k_u$ is the focal length measured in pixel, $s = -f_m k_u \cot(\phi)$ is the skew and $r = k_v / (k_u \sin(\phi))$ is the aspect ratio.

3 Projective spaces invariant to camera parameters

Invariance to camera intrinsic parameters can be obtained if there exist a non-singular (3×3) matrix $\mathbf{Q}$ measured from image features, such that:

$$\mathbf{Q} = \mathbf{K}\mathbf{M} \ ,$$

where $\mathbf{M}$ is a matrix only depending on the 3D coordinates of the object with respect to frame $\mathcal{F}$. The matrix $\mathbf{Q}$ can be used to define a $\mathcal{P}^2$ projective space $\mathcal{Q}$. The transformed point $\mathbf{q}_i \in \mathcal{Q}$ is:

$$\mathbf{q}_i = \mathbf{Q}^{-1}\mathbf{p}_i \ , \qquad \mathbf{q}_i \in \mathcal{Q}(\boldsymbol{\xi}) \ . \tag{5}$$

From now on we will refer to the transformed space $\mathcal{Q}(\boldsymbol{\xi})$ as an "invariant" space where the invariance is related to camera intrinsic parameters. It is easy to show that the transformed image point $\mathbf{q}_i$ does not depend on the intrinsic parameters of the camera. Indeed, from equation (3) we obtain:

$$\mathbf{q}_i = \mathbf{Q}^{-1}\mathbf{p}_i = \mathbf{M}^{-1}\mathbf{K}^{-1}\mathbf{K}\mathbf{m}_i = \mathbf{M}^{-1}\mathbf{m}_i \ .$$

Therefore, $\mathbf{q}_i$ only depends on the position of the camera with respect to the observed object and on its 3D structure. Matrix $\mathbf{Q}$ can be built by arbitrarily choosing three non collinear image points as proposed in [9]. However, the transformation proposed in [9] depends on the choice of the three points and the transformed coordinates $\mathbf{q}_i$ are also invariant to a rotation r_z around the $\overrightarrow{z}$ axis of frame $\mathcal{F}$. In order to avoid the selection of the points, we propose a different method to compute the projective transformation $\mathbf{Q}$. This method uses all available features in the image.

3.1 Invariance by projective transformation

Suppose that n points, not all collinear, are available in image $\mathcal{I}$. Let $\mathbf{S}_p$ be the following symmetric (3×3) matrix:

$$\mathbf{S}_p = \frac{1}{n}\sum_{i=1}^{n}\mathbf{p}_i\mathbf{p}_i^T = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n}u_i^2 & \sum_{i=1}^{n}u_iv_i & \sum_{i=1}^{n}u_i \\ \sum_{i=1}^{n}u_iv_i & \sum_{i=1}^{n}v_i^2 & \sum_{i=1}^{n}v_i \\ \sum_{i=1}^{n}u_i & \sum_{i=1}^{n}v_i & \sum_{i=1}^{n}1 \end{bmatrix} .$$

This matrix, computed using all observed points, depends on camera intrinsic parameters and it has a simple geometric meaning. Indeed, if:

$$\mathbf{S}_p = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & 1 \end{bmatrix} ,$$

then σ_{13} and σ_{23} are the coordinates of the centroid of the n points, while σ_{11}, σ_{12} and σ_{22} are the moments of inertia of the points about the origin axes. Similarly, let $\mathbf{S}_m$ be the following positive symmetric (3×3) matrix:

$$\mathbf{S}_m = \frac{1}{n}\sum_{i=1}^{n}\mathbf{m}_i\mathbf{m}_i^T = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n}x_i^2 & \sum_{i=1}^{n}x_iy_i & \sum_{i=1}^{n}x_i \\ \sum_{i=1}^{n}x_iy_i & \sum_{i=1}^{n}y_i^2 & \sum_{i=1}^{n}y_i \\ \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}y_i & \sum_{i=1}^{n}1 \end{bmatrix} .$$

This matrix only depends on the structure of the object and on its position with respect to frame $\mathcal{F}$. Since $\mathbf{p}_i = \mathbf{K}\mathbf{m}_i$, $\mathbf{S}_p$ can be written as a function of $\mathbf{S}_m$ and of the camera intrinsic parameters $\mathbf{K}$:

$$\mathbf{S}_p = \frac{1}{n}\sum_{i=1}^{n}\mathbf{p}_i\mathbf{p}_i^T = \frac{1}{n}\sum_{i=1}^{n}\mathbf{K}\mathbf{m}_i\mathbf{m}_i^T\mathbf{K}^T = \mathbf{K}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{m}_i\mathbf{m}_i^T\right)\mathbf{K}^T = \mathbf{K}\,\mathbf{S}_m\,\mathbf{K}^T . \qquad (6)$$

It is possible to decompose $\mathbf{S}_p$ and $\mathbf{S}_m$ using a Cholesky decomposition as:

$$\mathbf{S}_m = \mathbf{T}_m \, \mathbf{T}_m^T \quad (7)$$
$$\mathbf{S}_p = \mathbf{T}_p \, \mathbf{T}_p^T \quad (8)$$

where $\mathbf{T}_m$ and $\mathbf{T}_p$ are upper triangular matrices. Thus, from equations (6), (7) and (8) we obtain:

$$\mathbf{T}_p = \mathbf{K}\mathbf{T}_m \quad (9)$$

Note that $\mathbf{T}_m$ does not depend on $\mathbf{K}$, thus choosing $\mathbf{Q} = \mathbf{T}_p$ the transformed space defined by equation (5) is independent on camera intrinsic parameters:

$$\mathbf{q} = \mathbf{T}_p^{-1}\mathbf{p} = \mathbf{T}_m^{-1}\mathbf{K}^{-1}\mathbf{p} = \mathbf{T}_m^{-1}\mathbf{K}^{-1}\mathbf{K}\mathbf{m} = \mathbf{T}_m^{-1}\mathbf{m}$$

Note that $\mathbf{T}_p$ is easily obtained from the Cholesky decomposition of matrix $\mathbf{S}_p$:

$$\mathbf{Q} = \mathbf{T}_p = \begin{bmatrix} \sqrt{\sigma_{11} - \sigma_{13}^2 - \frac{(\sigma_{12} - \sigma_{13}\sigma_{23})^2}{\sigma_{22} - \sigma_{23}^2}} & \sqrt{\frac{(\sigma_{12} - \sigma_{13}\sigma_{23})^2}{\sigma_{22} - \sigma_{23}^2}} & \sigma_{13} \\ 0 & \sqrt{\sigma_{22} - \sigma_{23}^2} & \sigma_{23} \\ 0 & 0 & 1 \end{bmatrix}.$$

The computation of $\mathbf{Q}$ from image points is extremely stable since it is based on averaging measurements. This is proven by simulation results obtained with $n = 100$ points randomly distributed in a sphere with $\rho = 10$ cm radius. The image coordinates of each point are corrupted with an additive Gaussian noise with zero mean and increasing standard deviation. For each value of σ, we compute the mean and the standard deviation, over $N = 10000$ samples, of the relative errors on the estimated matrix $\tilde{\mathbf{Q}}$. Let $\tilde{q}_{ij}$ and q_{ij} be respectively the ij entries of estimated matrix $\tilde{\mathbf{Q}}$ and of the true matrix $\mathbf{Q}$. Since $\tilde{q}_{21} = q_{21} = 0$, $\tilde{q}_{31} = q_{31} = 0$, $\tilde{q}_{32} = q_{32} = 0$ and $\tilde{q}_{33} = q_{33} = 1$, we compute the following five relative errors:

$$\frac{\tilde{q}_{ij} - q_{ij}}{q_{ij}} \qquad (i,j) \in \{(1,1),(1,2),(1,3),(2,2),(2,3)\}$$

The results plotted in Figure 2 show that the average relative error on $\tilde{\mathbf{Q}} - \mathbf{Q}$ remains small despite σ increases. Obviously the standard deviation of the error increases but it is small enough to consider $\tilde{\mathbf{Q}}$ as a very good approximation of $\mathbf{Q}$. This is very important since matrix $\tilde{\mathbf{Q}}$ will be used in the cost function minimized in the bundle adjustment. We will see in the next subsection that having a good approximation of the true matrix $\mathbf{Q}$ allows us to correctly estimate "a posteriori" the camera intrinsic parameters.

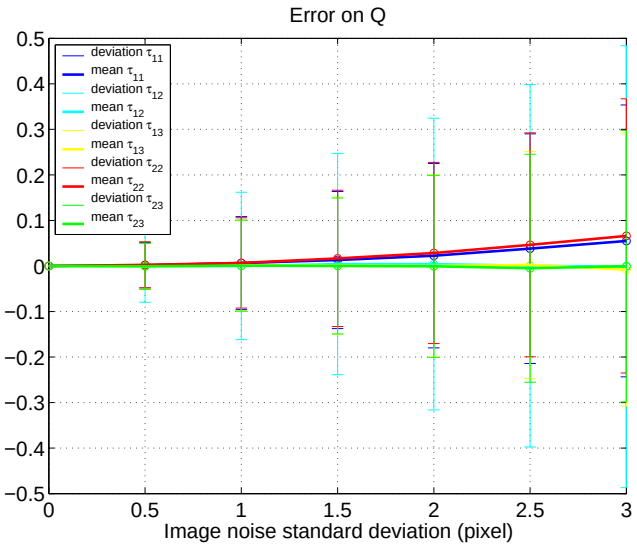


Figure 2: Relative error on the entries of the matrix $\mathbf{Q}$ estimated with an increasing standard deviation of noise on pixel coordinates. The mean and the standard deviation of the error are sufficiently small to consider $\tilde{\mathbf{Q}}$ as a good approximation of the true matrix $\mathbf{Q}$.

3.2 Computing camera parameters "a posteriori"

From equation (9), the camera parameters can be computed "a posteriori' as a function of the known matrix $\mathbf{T}_p$ and of the unknown matrix $\mathbf{T}_m$:

$$\mathbf{K} = \mathbf{T}_p \mathbf{T}_m^{-1} \tag{10}$$

Since the matrix $\mathbf{T}_p$ is measured from image points corrupted by noise, only an estimation $\tilde{\mathbf{T}}_p$ can be used to compute an estimated matrix $\widehat{\mathbf{K}}$:

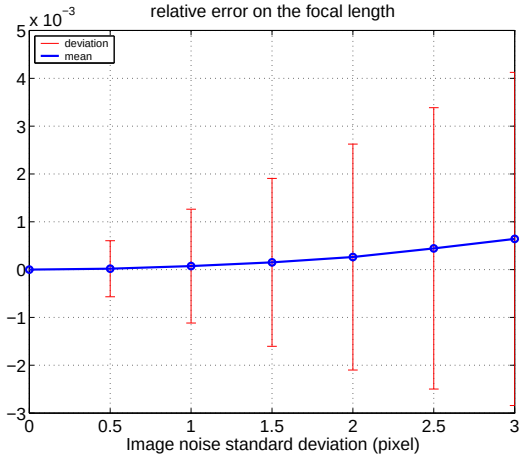
$$\widehat{\mathbf{K}} = \tilde{\mathbf{T}}_p \widehat{\mathbf{T}}_m^{-1} \tag{11}$$

where $\widehat{\mathbf{T}}_m$ is an estimation of the unknown matrix $\mathbf{T}_m$. During the bundle adjustment, we will change the estimation $\widehat{\mathbf{T}}_m$ trying to find the true matrix. Suppose that the matrix $\widehat{\mathbf{T}}_m$ is perfectly estimated (i.e. we have $\widehat{\mathbf{T}}_m = \mathbf{T}_m$). In this case, we can show that the error $\tilde{\mathbf{K}} - \mathbf{K}$ can be neglected. As a consequence, using $\tilde{\mathbf{K}}$ instead of $\mathbf{K}$ in the bundle adjustment has a little effect on the optimization process preserving the optimality of the solution. In order to have a ground truth, simulations have been carried out using a pinhole camera with $f = 1000$ pixels, $r = 1$, $\phi = \frac{\pi}{2}$ and $u_0 = v_0 = 250$ pixels. For each simulation, the results are evaluated by computing the following relative errors:

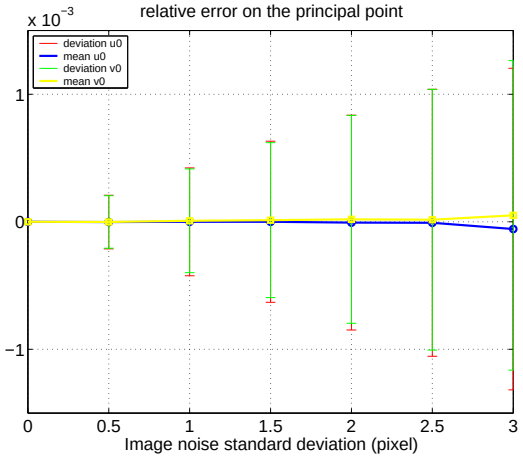
- the relative error on the focal length: $\dfrac{\widehat{f}_p - f_p}{f_p}$
- the relative error on the aspect ratio: $\dfrac{\widehat{r} - r}{r}$
- the relative error on the skew: $\dfrac{\widehat{\phi} - \phi}{\phi}$
- the relative errors on the principal point: $\dfrac{\widehat{u}_0 - u_0}{u_0}$ and $\dfrac{\widehat{v}_0 - v_0}{v_0}$

In these simulations, the distribution of the 3D points is the same one used in the previous subsection. In the first simulation, the image coordinates of each point are corrupted with an additive Gaussian noise with zero mean and increasing standard deviation ($0 \leq \sigma \leq 3$ pixels). For each value of σ, we compute the mean and the standard deviation, over $N = 10000$ samples, of the relative errors. The results given in Figure 3(a) show that the error on the camera intrinsic parameters is considerably smaller than the error usually obtained with standard self-calibration methods (see the results in Section 5).

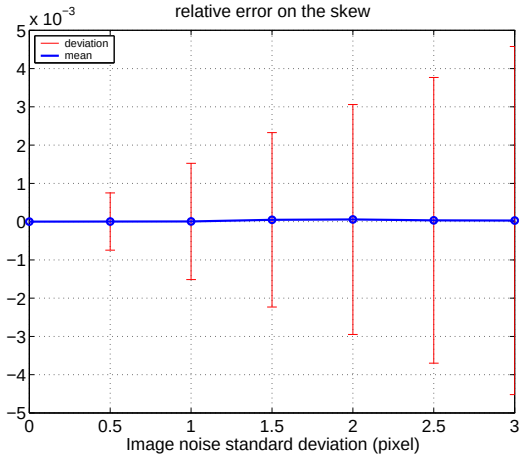
Figure 4(b) show the results obtained with a fixed standard deviation ($\sigma = 1$) of the Gaussian noise. The number of points increases ($10 \leq n \leq 100$) at each iteration. For each value of n, we compute the mean and the standard deviation, over $N = 10000$ samples, of the relative errors. This means that the error is very small and practically constant while the standard deviation slowly decreases as the number of points increases. This result confirms that the error on the intrinsic camera parameters is negligible for any number of points.



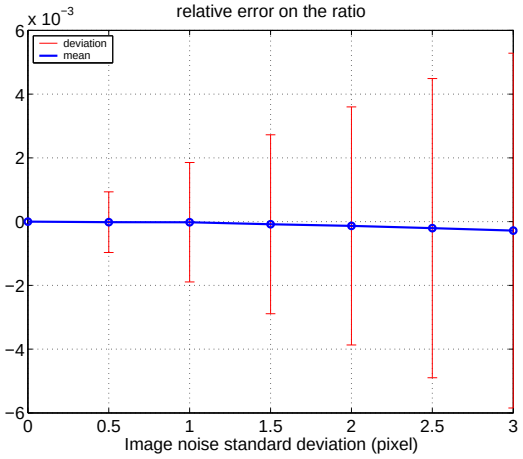
(a) Relative error on the focal length f



(b) Relative error on the principal point (u_0, v_0)

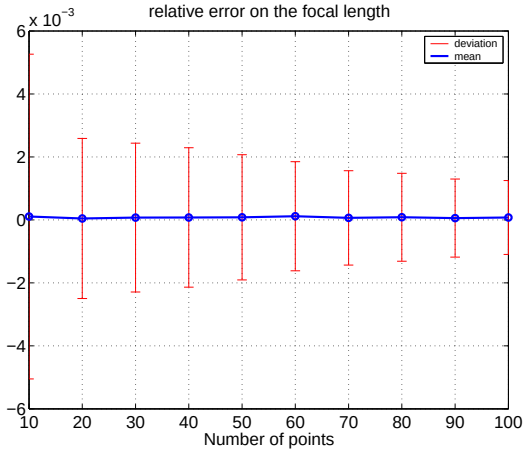


(c) Relative error on the angle ϕ

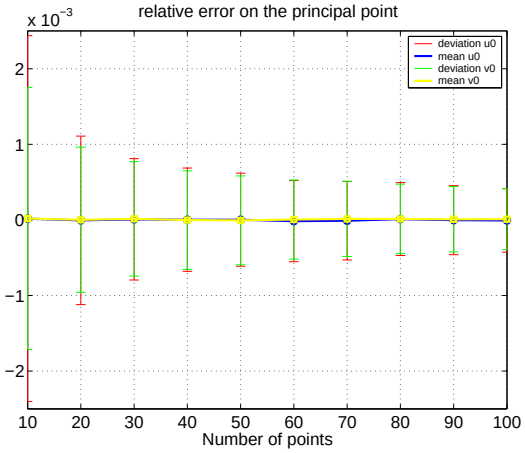


(d) Relative error on the ratio r

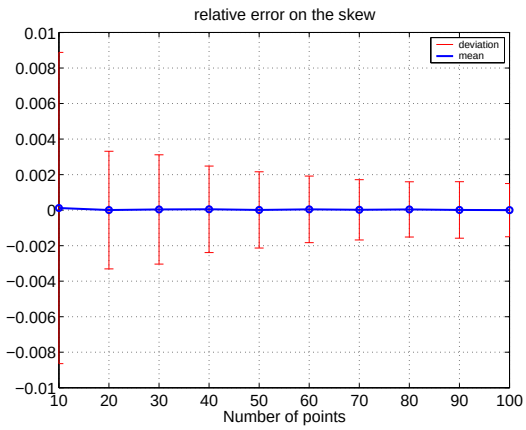
Figure 3: Results with an increasing standard deviation of noise on pixel coordinates (with $n = 100$ points). The mean error slightly increases for all the camera intrinsic parameters. The standard deviation increases faster but it remains small.



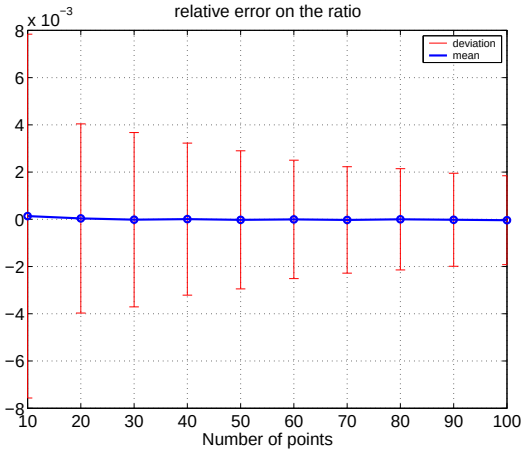
(a) Relative error on the focal length f



(b) Relative error on the principal point (u_0, v_0)



(c) Relative error on the angle ϕ



(d) Relative error on the ratio r

Figure 4: Simulation results with an increasing number of points and with fixed standard deviation $\sigma = 1$. The mean error is almost constant for all the camera intrinsic parameters. The standard deviation of the error decreases when the number of points used in the computation increases.

4 Metric reconstruction from several images

Suppose to have m different images (obtained from several zooming cameras) of the same n points. Bundle adjustment is achieved by minimizing the following cost function:

$$\mathcal{C} = \sum_{i=1}^{n} \sum_{j=1}^{m} \|\widehat{\mathbf{p}}_{ij} - \tilde{\mathbf{p}}_{ij}\|^2 ,$$

where $\tilde{\mathbf{p}}_{ij}$ is the known vector containing the pixel coordinates of point i measured in the image j and $\widehat{\mathbf{p}}_{ij}$ is the model estimated from unknown parameters. Standard bundle adjustment uses the following model:

$$\widehat{\mathbf{p}}_{ij}(\mathbf{K}_j, \boldsymbol{\xi}_j, \boldsymbol{\mathcal{X}}_i) = \mathbf{K}_j \widehat{\mathbf{m}}_{ij}(\boldsymbol{\xi}_j, \boldsymbol{\mathcal{X}}_i)$$

thus, for each image the unknowns are the camera parameters $\mathbf{K}_j$ (i.e. $5m$ unknowns for all images) and the camera pose $\boldsymbol{\xi}_j$ (i.e. $6m$ unknowns for all images), and for each point the unknowns are the 3D coordinates $\boldsymbol{\mathcal{X}}_i$ (i.e. $3n$ unknowns for all points). The total number of unknowns for a standard bundle adjustment is thus $s = 3n + 11m$.

In this report we propose to use a different model of $\mathbf{p}_{ij}$ in the bundle adjustment:

$$\widehat{\mathbf{p}}_{ij}(\boldsymbol{\xi}_j, \boldsymbol{\mathcal{X}}) = \tilde{\mathbf{T}}_{pj} \widehat{\mathbf{T}}_{mj}^{-1}(\boldsymbol{\xi}_j, \boldsymbol{\mathcal{X}}) \widehat{\mathbf{m}}_{ij}(\boldsymbol{\xi}_j, \boldsymbol{\mathcal{X}}_i) = \tilde{\mathbf{Q}}_j \, \widehat{\mathbf{q}}_{ij}(\boldsymbol{\xi}_j, \boldsymbol{\mathcal{X}})$$

where $\tilde{\mathbf{Q}}_j$ is measured from image features and $\widehat{\mathbf{q}}_{ij}(\boldsymbol{\xi}_j, \boldsymbol{\mathcal{X}})$ depends on the camera pose $\boldsymbol{\xi}_j$, on the 3D coordinates of all the points $\boldsymbol{\mathcal{X}} = (\boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{X}}_2, ..., \boldsymbol{\mathcal{X}}_n)$ but not on the camera intrinsic parameters. Consequently, the bundle adjustment we propose is intrinsics-free. The total number of unknowns of the intrinsics-free bundle adjustment is $p = 3n + 6m$. When compared to the standard bundle adjustment, the relative difference of unknowns is $(s-p)/s = 5m/(3n+11m))$. When $m >> n$ (for example when tracking a given object in a long sequence of images taken with several zooming cameras) we can obtain a reduction up to $(s-p)/s \approx 0.45$ (i.e. 45%) of the number of unknowns. Therefore, as it will be proven by the experimental results presented in Section 5, the computational cost of the optimization process can be considerably reduced.

Let $\mathbf{x} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_m, \boldsymbol{\mathcal{X}}_1, \boldsymbol{\mathcal{X}}_2, ..., \boldsymbol{\mathcal{X}}_n)$ be the (p×1) vector containing all the unknowns. The intrinsics-free bundle adjustment consists in solving the following optimization problem:

$$\min_{\mathbf{x}} \mathcal{C}(\mathbf{x}) = \min_{\mathbf{x}} \sum_{i=1}^{n} \sum_{j=1}^{m} \|\tilde{\mathbf{Q}}_j \, \widehat{\mathbf{q}}_{ij}(\boldsymbol{\xi}_j, \boldsymbol{\mathcal{X}}) - \tilde{\mathbf{p}}_{ij}\|^2 . \tag{12}$$

The cost function is thus defined as the sum of the residual errors of the reprojections in each image. If the error on the measured features $\tilde{\mathbf{p}}_{ij}$ has a Gaussian distribution and $\tilde{\mathbf{Q}}_j \approx \mathbf{Q}_j$ (see the experiment in Section 3) then our bundle adjustment is close to be a maximum likelihood estimator.

By setting $\tilde{\mathbf{q}}_{ij} = \tilde{\mathbf{Q}}_j^{-1} \, \tilde{\mathbf{p}}_{ij}$, the intrinsics-free bundle adjustment could also be defined as the minimization of a cost function completely defined in the invariant space $\mathcal{Q}$

$\min_{\mathbf{x}} \sum_{i=1}^{n} \sum_{j=1}^{m} \|\widehat{\mathbf{q}}_{ij}(\boldsymbol{\xi}_j, \boldsymbol{\mathcal{X}}) - \tilde{\mathbf{q}}_{ij}\|^2$. However, in the report we will use the cost function defined in (12) in order to compare the results with respect to the standard bundle adjustment which cost function is defined in the image space. Finally, note that, without loss of generality, we can set the absolute frame $\mathcal{F}_0$ to one of camera frame $\mathcal{F}_k$. Thus, the number of unknowns is reduced to $s = 3n + 11m - 6$ for the standard bundle adjustment and to $p = 3n + 6m - 6$ for the intrinsics-free bundle adjustment.

4.1 Computational cost of the non-linear optimization

The optimization problem (12) is efficiently solved using iterative second order methods. They are generally fast since they provide quadratic convergence to the solution if the starting point of the minimization $\mathbf{x}_0$ is in a neighborhood of the solution. Second order methods are based on the Taylor expansion of the cost function at each iteration k of the optimization process:

$$\mathcal{C}(\mathbf{x}_{k+1}) = \mathcal{C}(\mathbf{x}_k) + (\mathbf{x}_{k+1} - \mathbf{x}_k)^T \frac{\partial \mathcal{C}(\mathbf{x}_k)}{\partial \mathbf{x}} + \frac{1}{2}(\mathbf{x}_{k+1} - \mathbf{x}_k)^T \frac{\partial^2 \mathcal{C}(\mathbf{x}_k)}{\partial \mathbf{x}^2}(\mathbf{x}_{k+1} - \mathbf{x}_k)^T + O^3(\mathbf{x}_{k+1} - \mathbf{x}_k)$$

where $\frac{\partial \mathcal{C}(\mathbf{x}_k)}{\partial \mathbf{x}}$ is the (p×1) gradient vector, $\frac{\partial^2 \mathcal{C}(\mathbf{x}_k)}{\partial \mathbf{x}^2}$ is the (p×p) Hessian matrix and $O^3(\mathbf{x}_{k+1} - \mathbf{x}_k)$ is the residual error. Starting from this expansion, several methods have been proposed to find and update $\mathbf{x}_{k+1}$ of the unknowns. For example, a standard Newton method compute the update $\mathbf{x}_{k+1}$ of the unknowns by setting $O^3(\mathbf{x}_{k+1} - \mathbf{x}_k) = 0$:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \left(\frac{\partial^2 \mathcal{C}(\mathbf{x}_k)}{\partial \mathbf{x}^2} \right)^{-1} \frac{\partial \mathcal{C}(\mathbf{x}_k)}{\partial \mathbf{x}}$$

At each iteration, one needs to compute or approximate the gradient vector and the Hessian matrix. It is clear that the computational cost of the non-linear minimization is directly related to the number of unknowns p. The lower is p the faster is the computation. In this report, we use the non-linear second order optimizer Levenberg-Marquardt [3] with numerical differentiation (i.e. the gradient vector and the Hessian matrix are approximated). Note that, even if they are not currently used in our experimental results, the speed of any bundle adjustment algorithm can be increased by using optimization techniques which optimally exploit sparseness and data structure [13].

4.2 Initialization of the non-linear optimization

Non-linear optimization requires an initial value $\mathbf{x}_0$ of the unknowns. In our experiments, we first obtain a projective reconstruction and then use a rough self-calibration to upgrade the reconstruction to metric. The projective reconstruction is obtained by computing the fundamental matrix using [4] between the two farthest images, usually the first one and the last one, and triangulating point features in projective space using [5]. We then register each other image in turn by using 3D to 2D point correspondences. Finally, we refine the estimated structure and motion with a bundle adjustment in projective space [1]. The

self-calibration is obtained from the projective reconstruction by using the linear method proposed in [8] and inspired by [12]. We assume that each camera has zero skew, unity aspect ratio and its principal point lying at the image center. These assumptions allow to obtain a linear solution for the (eventually varying) focal length through the recovery of the absolute quadric.

4.3 Imposing geometric constraints

For most of pinhole cameras it is a reasonable approximation to suppose a zero skew (i.e. $\phi = \pi/2$) and unit aspect ratio (i.e. $r = 1$). Thus, there are some geometric constraints on camera internal parameters that can be imposed during the optimization. Let τ^p_{ij} and τ^m_{ij} be the ij entries of matrices $\mathbf{T}_p$ and $\mathbf{T}_m(\mathbf{x})$ respectively. It is easy to obtain from equation (11) the equivalence between the following constraints:

- zero skew: $\phi = \dfrac{\pi}{2} \;\Leftrightarrow\; \tau^p_{12}\tau^m_{11}(\mathbf{x}) = \tau^p_{11}\tau^m_{12}(\mathbf{x})$
- unit aspect ratio: $r = \dfrac{k_u}{k_v} = 1 \;\Leftrightarrow\; \tau^p_{11}\tau^m_{22}(\mathbf{x}) = \tau^p_{22}\tau^m_{11}(\mathbf{x})$

Imposing constraints on intrinsic parameters is thus equivalent to impose constraints on the unknowns $\mathbf{x}$ of the intrinsics-free bundle adjustment. Constraints on the intrinsics are modeled by using either Lagrange multipliers or heavily weighted artificial measurements added to the cost function. In the former case, each constraint is modeled by one additional parameter and in the latter case, not any parameter are necessary. We do not have observed significant differences during our experiments, in terms of accuracy, between these two means of modeling constraints.

4.4 Expected improvements over standard bundle adjustment

As mentioned before, the exact count of the unknowns depends on the considered setup (number of cameras, structure of the object,...). We present different setups used for metric reconstruction, supposing that singular configurations have been avoided and that a minimal number of data is available to achieve the reconstruction. For each case, we compare the expected performances of the intrinsics-free bundle adjustment with respect to the standard one. These considerations are confirmed by the experimental results presented in Section 5.

4.4.1 Fixed intrinsic parameters

In the first setup, we consider that all intrinsic parameters are fixed and unknown (all images are provided by the same camera). A minimum number of 3 images are necessary to self-calibrate. The unknowns of the standard bundle adjustment are $s = 3n + 6m + 5$ while the unknowns of the intrinsics-free bundle adjustment are $p = 3n + 6m$. Since $p \approx s$, the improvements of the new method on the computational cost are expected not to be impressive. However, it is an interesting setup since it is very common and it can be used

to prove that the accuracy of the solution found by the intrinsics-free bundle adjustment is as good as for standard bundle adjustment.

4.4.2 Varying focal length and principal point.

The second setup assumes that cameras have zero skew and unity aspect ratio. This case is often encountered in practice since these hypotheses are satisfied by most modern cameras. It corresponds to either a unique zooming camera in motion observing a static scene (note that not only the focal length but also the principal point position may change when zooming and/or focusing) or a set of cameras taking synchronized pictures of the same, eventually dynamic, scene. A minimum number of 4 images are then needed to self-calibrate. The unknowns of the standard bundle adjustment are $s = 3n + 6m + 3m$ while the unknowns of the intrinsics-free bundle adjustment are $p = 3n + 6m$. In this case, when $m >> n$ we can obtain a reduction up to $(s - p)/s \approx 0.33$ (i.e. 33%) of the number of unknowns. As a consequence, the computational cost is expected to be reduced.

4.4.3 Time-varying structure and motion tracked by several zooming cameras.

The third setup is the most general one. The structure, the camera intrinsic and extrinsic parameters are continuously changing. It is an interesting setup for modern robotic vision since it allows the achievement of new applications in visual servoing, object tracking or computer-aided surgery. The speed of the reconstruction algorithm is extremely important in order to perform the tracking. With this setup, the number of unknowns is the same as for the previous setup at each iteration. Thus, the intrinsics-free bundle adjustment is expected to work better than the standard bundle adjustment, especially when the number of cameras involved in the reconstruction is large.

5 Experimental Results

In this section, we validate our algorithm using both simulated data and real images and compare it to standard bundle adjustment methods.

5.1 Simulated Data

To test whether the theory presented in the report is a reasonable approximation we consider the metric reconstruction of $n = 25$ points lying inside a cube of 1 meter side. Several images (minimum is $m = 3$), taken with one or more cameras (the focal length and the principal point may also be changed) depending on the setup, are used to reconstruct the 3D coordinates of the points. Images are taken around the cube from different positions between 5 and 9 meters away from its center. The image coordinates of the points are corrupted by a centered Gaussian noise with varying standard deviation. The algorithms are repeated in order to compute the mean and the standard deviation over 1000 trials.

Both standard and intrinsics-free bundle adjustment are ran on the same computer. We measure the CPU time and the number N of iterations needed to the convergence of each algorithm to assess the corresponding computational cost. The accuracy of the solution is measured by evaluating the error on the metric reconstruction. Obviously the reconstruction is made up to a scalar factor, thus we use the ground truth just to fix the scale and make it possible to compare the results in a metric space. In order to have an error independent on the frame in which the reconstruction is done, we consider only differences between the estimated distance $\|\widehat{\boldsymbol{\mathcal{X}}}_i - \widehat{\boldsymbol{\mathcal{X}}}_j\|$ and the true distance $\|\boldsymbol{\mathcal{X}}_i - \boldsymbol{\mathcal{X}}_j\|$ between two points in the 3D space. Thus, the error $\epsilon_{\mathcal{X}}$ on the structure is obtained as:

$$\epsilon_{\mathcal{X}} = \sum_{i,j}^{n} \left(\|\widehat{\boldsymbol{\mathcal{X}}}_i - \widehat{\boldsymbol{\mathcal{X}}}_j\| - \|\boldsymbol{\mathcal{X}}_i - \boldsymbol{\mathcal{X}}_j\| \right)^2$$

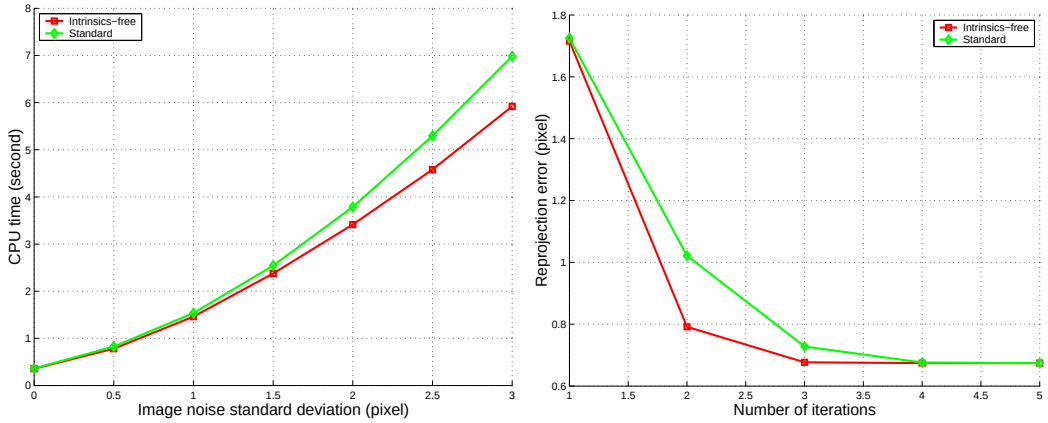
In order to verify that the camera parameters obtained "a posteriori" by our algorithm are a good approximation of the true parameters, we measure the relative error ϵ_f on the recovered focal length as:

$$\epsilon_f = \frac{\widehat{f}_p - f_p}{f_p} \ .$$

Measuring the error on the focal length is justified by the fact that this is usually the only parameter obtained by self-calibration and that may lie quite far from the optimal solution.

5.1.1 Fixed intrinsic parameters

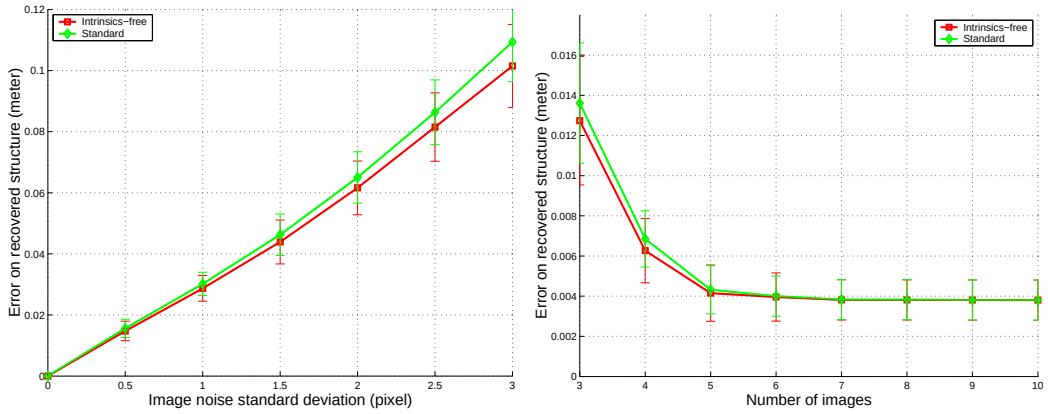
In the first set of experiments, we consider a single moving camera with fixed intrinsics parameters. The focal length of the camera is $f = 1000$ pixels, the principal point is at the image center, the skew is zero and the aspect ratio is one. Figure 5 presents the comparison between the properties of convergence and computational cost of the algorithms. As expected, the CPU time until convergence is slightly higher for the standard bundle adjustment algorithm especially when the standard deviation σ of the noise increases (see Figure 5(a)). This result is confirmed by figure 5(b) which shows that the convergence to the minimum of the cost function requires less iterations for the intrinsics-free method. It is worth noting that the algorithms converge almost to the same minimum of the cost function. This has been the case in most of our experiments.



(a) CPU time versus the level of added noise (b) Reprojection error versus iteration number

Figure 5: Comparison of the computational cost between the standard (green line) and the intrinsics-free (red line) bundle adjustment.

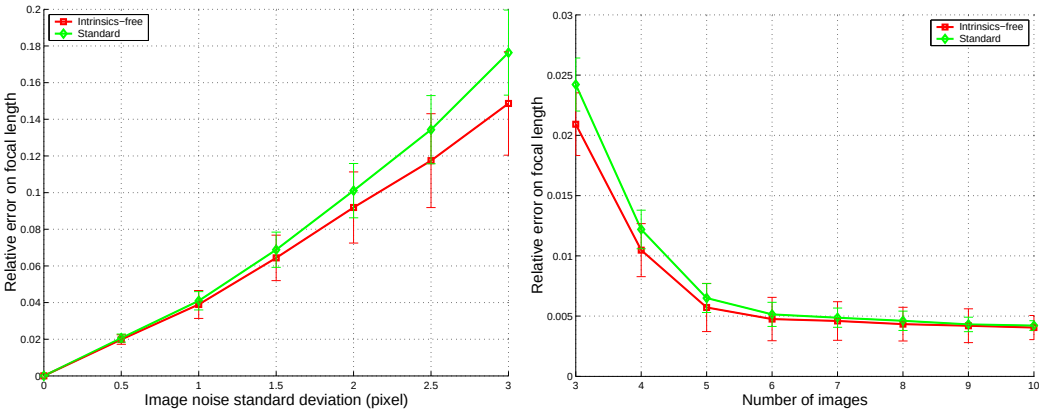
Figure 6 plots the error on the recovered structure. Even if the difference between the two algorithm is not impressive, the intrinsics-free method performs better when the noise σ is high (Figure 6(a)) and when only few m images are used (Figure 6(b)). As mentioned in Section 3, the camera intrinsic parameters are not directly considered as unknown in our algorithm but they can be estimated "a posteriori".



(a) Error $\epsilon_{\mathcal{X}}$ versus level of added image noise (b) Error $\epsilon_{\mathcal{X}}$ versus number of images

Figure 6: Comparison of the accuracy of the 3D reconstruction between the standard (green line) and of the intrinsics-free (red line) bundle adjustment.

Since the 3D structure estimated with the intrinsics-free bundle adjustment is more accurate, it is not surprising, as shown in Figure 7, that the relative error ϵ_f on the focal length is also smaller. Figure 7(a) shows the results obtained when the standard deviation of the noise in the image σ varies from 0 to 3 pixels. When the level of added noise is lower than 1 pixel, the results are undistinguishable. Beyond a certain level of added image noise the intrinsics-free bundle adjustment performs better. Figure 7(b) shows the results obtained when varying the number of images used in the reconstruction from 3 to 10. We observe that for a small number a images, in this case less than 5, the intrinsics-free method performs better than the standard one. When the number of images increases beyond 6 the results are undistinguishable



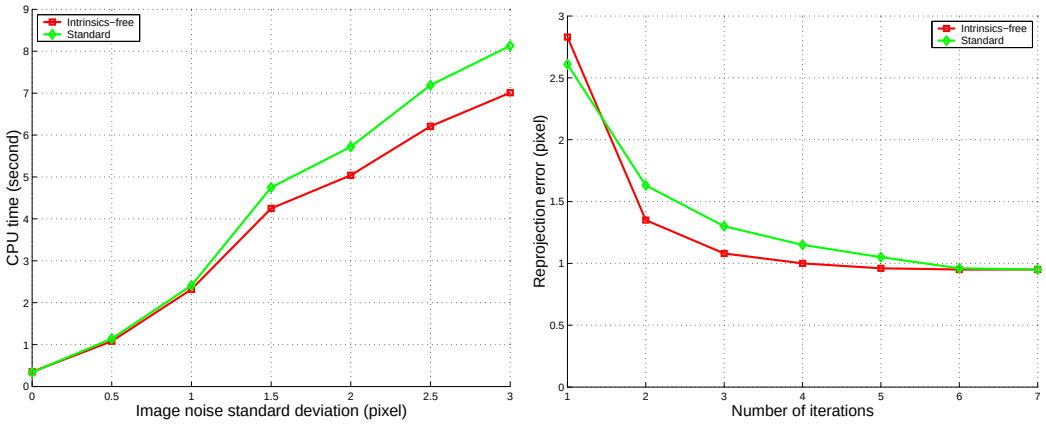
(a) Error ϵ_f versus level of added image noise

(b) Error ϵ_f versus number of images

Figure 7: Comparison of the relative error on the recovered focal length between the standard (green line) and of the intrinsics-free (red line) bundle adjustment.

5.1.2 Varying focal length and principal point.

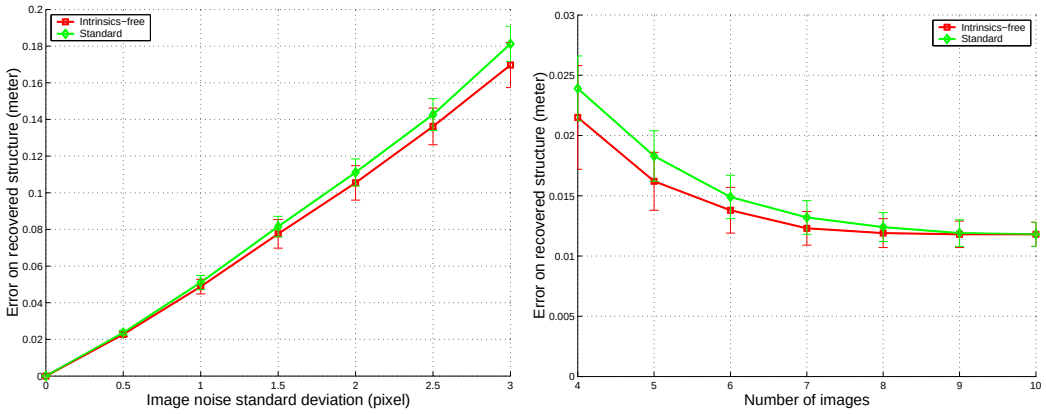
In this set of experiments we consider a moving and zooming camera. The focal length f_p changes between 800 and 1300 pixels while moving the camera around the 3D structure. We suppose that $s = 0$, $r = 1$ and that the principal point is at the image center. In this setup, the intrinsics-free bundle adjustment has less unknowns than the standard method. Thus, the CPU time and the number of iterations needed to converge are smaller, as it is shown in Figure 8.



(a) CPU time versus the level of added noise (b) Reprojection error versus iteration number

Figure 8: Comparison of the computational cost between the standard (green line) and the intrinsics-free (red line) bundle adjustment.

Again, the estimation processes is more stable. Thus, the reconstruction provided by the intrinsics-free bundle adjustment is more accurate. Figure 9 confirms the results obtained in the previous simulation. With a small number m of images and for higher noise σ the intrinsics-free bundle adjustment performs better. Otherwise, the results are undistinguishable.



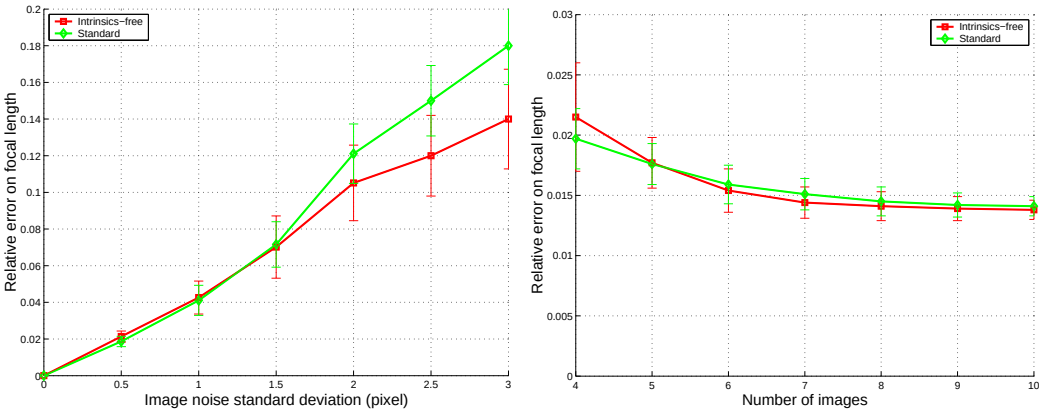
(a) Error $\epsilon_{\mathcal{X}}$ versus level of added image noise (b) Error $\epsilon_{\mathcal{X}}$ versus number of images

Figure 9: Comparison of the accuracy of the 3D reconstruction between the standard (green line) and of the intrinsics-free (red line) bundle adjustment.

In this experiment, multiple different focal lengths are recovered. Thus, we use the average error value over all images to compare the results:

$$\epsilon_f = \sum_{j=1}^{m} \frac{\widehat{f}_{pj} - f_{pj}}{f_{pj}} \ .$$

Figure 10(a) plots the relative error on the focal length when varying the level of noise on the image point coordinates, while the principal point is left constant at the center of the image. We observe in Figure 10(a) that, as in the previous case, the two methods perform equally well up to a certain level of noise ($\sigma = 1$ pixel), above which the intrinsics-free method begins to perform better than the standard one.

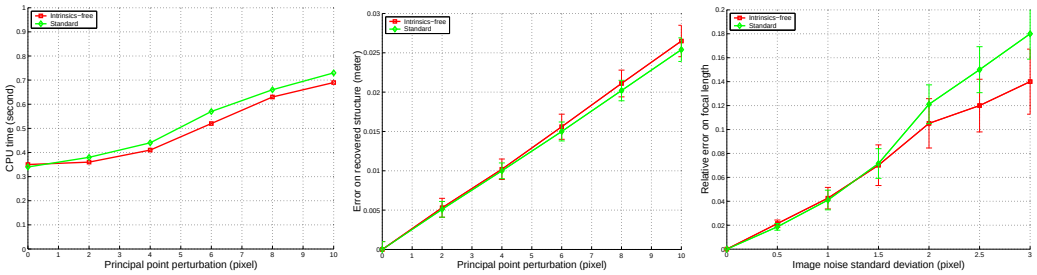


(a) Error ϵ_f versus level of added image noise

(b) Error ϵ_f versus number of images

Figure 10: Comparison of the relative error on the recovered focal lengths between the standard (green line) and of the intrinsics-free (red line) bundle adjustment.

Previous results are obtained supposing that the principal point is at the image center where effectively it was. On the other hand, what happen if we suppose that the principal point is at the image center but in reality it is not ? To answer this question, the principal point position is corrupted from the center of the camera with an additive centered Gaussian variable with varying standard deviation. Figure 11 shows that if the principal point is not at the image center the performance of both algorithm is reduced: the CPU time, the error on the recovered structure and focal length increase.

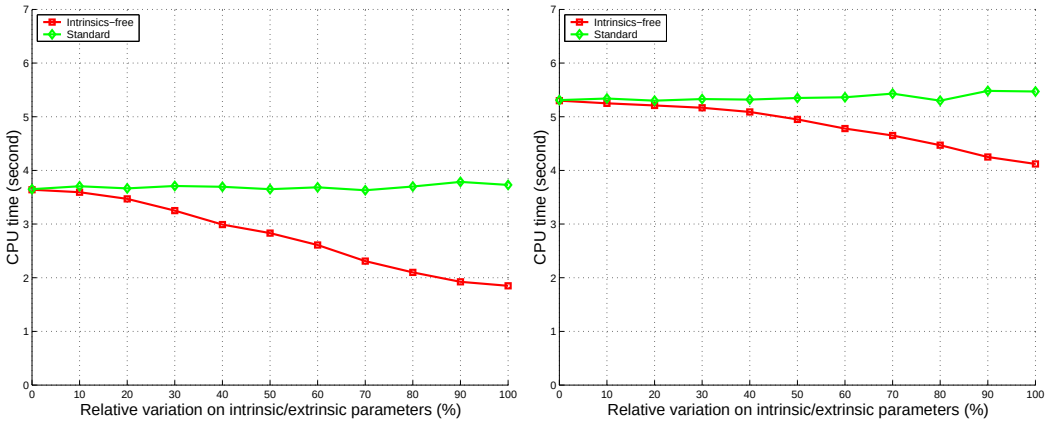


(a) CPU time (b) Error $\epsilon_{\mathcal{X}}$ on the structure (c) Error ϵ_f on the focal length

Figure 11: CPU time versus the perturbation of the principal point position.

5.1.3 Time-varying structure and motion tracked by several zooming cameras.

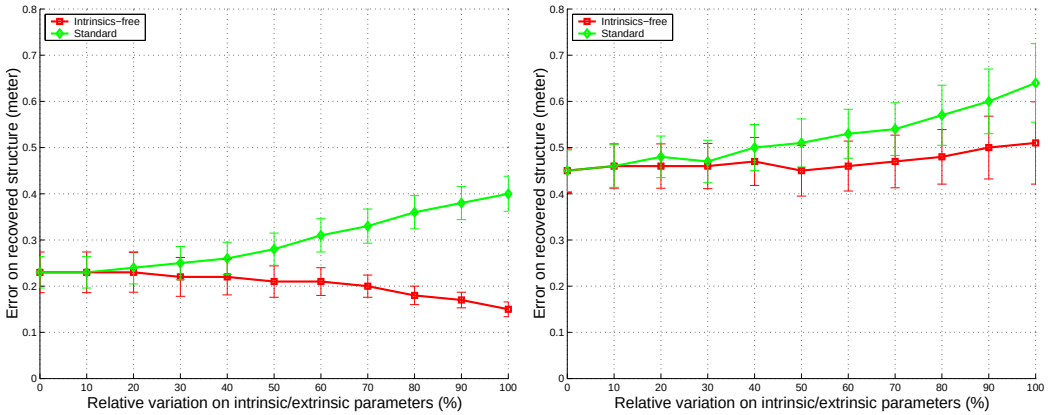
This experiment consists in tracking time-varying structure and motion using successive bundle adjustments. At each iteration, structure and motion are estimated using bundle adjustment. The estimate of iteration k is the starting point of iteration $k+1$. We simulate a moving object (eventually non-rigid) observed by a set of zooming cameras. At each iteration the object is roughly maintained at the center of the images using zooming and by changing camera orientations. Zooming affects camera focal length and principal point. Therefore, both camera intrinsic and extrinsic parameters might change. When we deform the structure through time, the deformation consists in offsetting each point from its position by a random value drawn from a normal distribution and with a variance of 10% of the object size. Figure 12 shows the needed CPU time when the object considered is either rigid (a) or non rigid (b). Object motion is a weighted combination between a circular motion which induces great changes in camera orientation and small changes in their intrinsic parameters, and a translation away from the cameras inducing the opposite effect. Figure 12(a) shows that when the change in camera orientation is dominant the behavior of the two bundle adjustments are very close, the intrinsics-free being slightly better. On the other hand, the more the object moves away from the cameras, the more their intrinsics vary which increases the difference between the two algorithms. The intrinsics-free method needs less CPU time to convergence than the standard bundle adjustment. Figure 12(b) shows that when the previous scenario is combined to an object deformation, the difference between the algorithms is slightly reduced but remains significant when the change in the intrinsics is dominant.



(a) The structure of the object is rigid (b) The structure of the object is non rigid

Figure 12: CPU time to convergence versus relative variation k on camera parameters. When $k = 0$ % only camera extrinsic parameters change, while when $k = 100$ % only camera intrinsic parameters change.

The reduction of the number of the unknowns has a beneficial effect on the accuracy of the reconstruction, as it is shown by the results plotted in Figure 13. We observe in Figure 13(a) that when the structure is rigid and camera intrinsic parameters change, the intrinsic-free approach performs much better than the standard bundle adjustment. When the structure is non-rigid (see Figure 13(b)), the difference between the two algorithm is reduced but still the intrinsics-free bundle adjustment is more accurate.



(a) The structure of the object is rigid (b) The structure of the object is non rigid

Figure 13: Error on recovered structure as a function of the relative variation k on camera parameters. When $k = 0$ % only camera extrinsic parameters change, while when $k = 100$ % only camera intrinsic parameters change.

5.2 Real Images

In this section, we validate our algorithm using real images of a calibration grid.Image point positions, calibration and pose of the camera can be estimated with a very high accuracy. This serves as ground truth.

5.2.1 Fixed intrinsic parameters

In the first experiment, we use 3 images of the calibration grid shown in Figure 14 observed by the same camera.

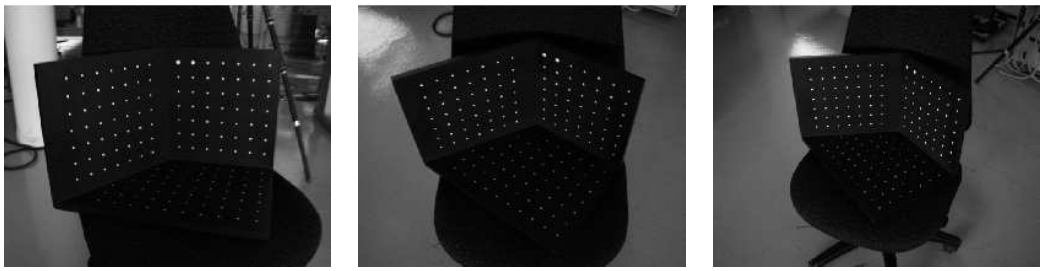


Figure 14: Three images of a calibration grid used in the experiment.

We run the bundle adjustments from an initial solution obtained by the linear method while enforcing the constraint that the intrinsics are identical for all the three images. The performance of the algorithms are close since the number of unknowns is almost the same (only 5 unknown less for the intrinsics-free method). However, the intrinsics-free bundle adjustment is slightly faster (7 %) than the standard bundle adjustment and the error on the recovered structure is almost the same. The results obtained with simulated data are confirmed by this experiment. Table 1 shows the results obtained for the focal length and the principal point position for the ground truth, the linear algorithm, the standard and the intrinsics-free bundle adjustment. We observe that the standard bundle adjustment performs globally worse than the intrinsics-free. However, the gap is negligible compared to the difference with the ground truth. The reprojection error of both algorithms is a hundred times lower than a pixel, which is normal since the accuracy of image point position is very high.

	ground truth	linear	intrinsics-free	standard
focal length	757.5	749.9	754.3	753.1
principal point, u_0	259.6	256.0	255.1	253.4
principal point, v_0	327.0	320.0	319.6	321.8

Table 1: Comparison of the static calibration results, taken as ground truth, the linear solution and the different bundle adjustment techniques.

5.2.2 Varying focal length and principal point.

In this experiments, we use 7 images of a calibration grid taken with a single zooming camera (see as an example the three images in Figure 15). The model of the grid is used as ground truth.

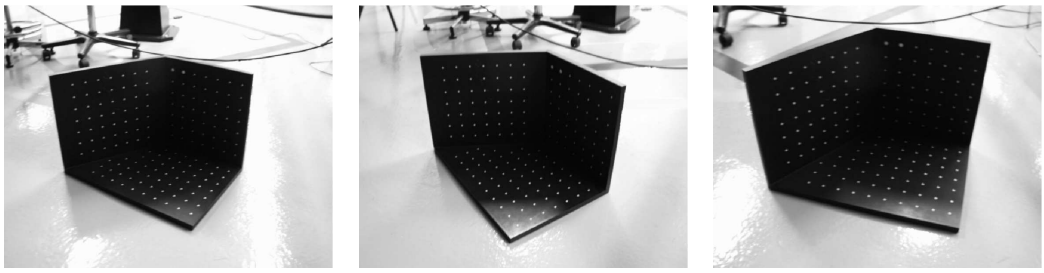


Figure 15: Three images of a calibration grid amongst the 7 used in the experiment.

With respect to the previous experiment, the intrinsics-free bundle adjustment need much less unknowns than the standard method since the camera is zooming. As a consequence,

the intrinsics-free method is 25 % faster and requires less iterations to converge (9 instead of 13). Again, the error on the recovered structure is almost the same. The results given in Table 2 prove that, for all the seven images, the intrinsics-free method is able to recover "a posteriori" the unknown intrinsic parameters of the camera better than the standard bundle adjustment.

	f_1	f_2	f_3	f_4	f_5	f_6	f_7
ground truth	751.9	754.1	757.2	765.5	1161.7	1992.5	1943.6
intrinsic-free	771.6	777.2	772.9	777.6	1201.9	2176.0	2100.5
standard	807.7	799.8	789.6	795.9	1240.3	2185.9	2229.5
	u_1	u_2	u_3	u_4	u_5	u_6	u_7
ground truth	324.3	302.9	294.9	341.3	316.2	321.1	237.2
intrinsic-free	334.7	311.7	298.2	350.4	328.1	318.9	317.1
standard	311.9	282.9	267.3	334.7	265.2	112.9	186.8
	v_1	v_2	v_3	v_4	v_5	v_6	v_7
ground truth	271.3	254.7	250.9	249.8	273.5	293.9	286.9
intrinsic-free	259.6	244.8	247.0	242.8	255.6	268.7	261.4
standard	269.5	246.4	238.5	249.1	237.4	110.1	348.0

Table 2: Comparison of the static calibration results, taken as ground truth, the linear solution and the different bundle adjustment techniques.

6 Conclusion

This report has presented a new refinement technique for metric reconstruction which is particularly adapted to computer vision applications close to real-time. Indeed, the proposed bundle adjustment uses much less unknowns in the optimization process allowing for a faster reconstruction algorithm. At the same time, the accuracy of the solution is slightly better than the accuracy obtained with standard bundle adjustment techniques. Future work will concern the test of our algorithm with several uncalibrated zooming cameras.

References

[1] A. Bartoli and P. Sturm. Three new algorithms for projective bundle adjustment with minimum parameters. Technical Report 4236, INRIA, August 2001.

[2] O. Faugeras, and Q-T. Luong. *Geometry of multiple images.* MIT Press, 2001.

[3] P.E. Gill, W. Murray, and M.H. Wright. *Practical Optimization.* Academic Press, 1981.

[4] R. Hartley. In defence of the 8-point algorithm. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 1064–1070, June 1995.

[5] R. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

[6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.

[7] F. Kahl and A. Heyden. Euclidean reconstruction and auto-calibration from continuous motion. In *IEEE International Conference on Computer Vision*, volume 2, Vancouver, Canada, July 2001.

[8] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *International Journal of Computer Vision*, 32(1):7–25, 1999.

[9] E. Malis. Visual servoing invariant to changes in camera intrinsic parameters. In *International Conference on Computer Vision*, volume 1, pages 704–709, Vancouver, Canada, July 2001.

[10] C.C. Slama, editor. *Manual of Photogrammetry, Fourth Edition*. American Society of Photogrammetry and Remote Sensing, Falls Church, Virginia, USA, 1980.

[11] S. Maybank and O. Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.

[12] B. Triggs. Autocalibration and the absolute quadric. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 609–614. IEEE Computer Society Press, June 1997.

[13] B. Triggs, P.F. McLauchlan, R.I. Hartley, and A. Fitzgibbon. Bundle ajustment — a modern synthesis. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Springer-Verlag, 2000.

[14] L. Van Gool, T. Moons, E. Pauwels, and A. Oosterlinck. Vision and lie's approach to invariance. *Image and vision computing*, 13(4):259–277, May 1995.

[15] M. Werman, S. Banerjee, S. Dutta Roy, and M. Qiu. Robot localization using uncalibrated camera invariants. In *IEEE International Conference on Computer Vision and Pattern Recognition*, volume II, pages 353–359, Fort Collins, CO, June 1999.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399