Euclidean Reconstruction Independent on Camera Intrinsic Parameters

Ezio MALIS
I.N.R.I.A.
Sophia-Antipolis, FRANCE

Adrien BARTOLI
INRIA
Rhone-Alpes, FRANCE

***Abstract*— Standard bundle adjustment techniques for Euclidean reconstruction consider camera intrinsic parameters as unknowns in the optimization process. Obviously, the speed of an optimization process is directly related to the number of unknowns and to the form of the cost function. The scheme proposed in this paper differs from previous standard techniques since unknown camera intrinsic parameters are not considered in the optimization process. Considering less unknowns in the optimization process produces a faster algorithm which is more adapted to time-dependent applications as in robotics. Computationally expensive metric reconstruction, using for example several zooming cameras, considerably benefits from an intrinsics-free bundle adjustment.**

I. INTRODUCTION

Bundle adjustment is the final refinement step of a reconstruction process. Standard metric reconstruction methods starts with a projective reconstruction of the scene obtained from several images [3] [5] [1]. The projective reconstruction is then updated to an Euclidean reconstruction after camera self-calibration [11] [7] [12]. Finally, the visual reconstruction is refined using bundle adjustment in the Euclidean space [10]. Standard bundle adjustment [13] [6] 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 paper is to investigate a new fast bundle adjustment method to be used in real-time robotic 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 [9]. 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 paper, 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 paper 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 the 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.

II. THEORETICAL BACKGROUND

A. 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. 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:

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

where ζ_i is a 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.

B. Camera model

The information given by a pinhole camera is an image point $\mathbf{p}_i$. The point $\mathbf{p}_i = (u_i, v_i, 1)$ depends on the camera internal parameters and on $\mathbf{m}_i$:

$$\mathbf{p}_i = \mathbf{K}\mathbf{m}_i \tag{2}$$

where:

$$\mathbf{K} = \begin{bmatrix} f_p & s & u_0 \\ 0 & rf_p & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

u_0 and v_0 are the coordinates of the principal point (pixels), f_p is the focal length measured in pixel, s is the skew and r is the aspect ratio.

C. Invariance by projective transformation

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

$$\mathbf{S}_p = \frac{1}{n}\sum_{i=1}^{n} \mathbf{p}_i \mathbf{p}_i^\top$$

This matrix, computed using all observed points, depends on camera intrinsic parameters. Similarly, let $\mathbf{S}_m$ be the following positive symmetric (3×3) matrix:

$$\mathbf{S}_m = \frac{1}{n}\sum_{i=1}^{n} \mathbf{m}_i \mathbf{m}_i^\top$$

This matrix only depends on the structure of the object and on its position with respect to the current frame. 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 $\mathbf{K}$:

$$\mathbf{S}_p = \mathbf{K}\,\mathbf{S}_m\,\mathbf{K}^\top \ . \tag{4}$$

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

$$\mathbf{S}_m = \mathbf{T}_m\,\mathbf{T}_m^\top \quad \text{and} \quad \mathbf{S}_p = \mathbf{T}_p\,\mathbf{T}_p^\top \tag{5}$$

where $\mathbf{T}_m$ and $\mathbf{T}_p$ are upper triangular matrices. Thus, from equations (4) and (5) we obtain:

$$\mathbf{T}_p = \mathbf{K}\mathbf{T}_m \tag{6}$$

Note that $\mathbf{T}_m$ does not depend on $\mathbf{K}$. Using $\mathbf{T}_p$ we define a projective space independent on camera intrinsic parameters:

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

The computation of $\mathbf{T}_p$ from image points is stable, as shown in [8] since it is based on averaging measurements.

D. Estimating camera parameters "a posteriori"

From equation (6), the camera parameters can be computed "a posteriori' as a function of the known matrix $\mathbf{T}_p$ and of the estimated matrix $\mathbf{T}_m$:

$$\mathbf{K} = \mathbf{T}_p\mathbf{T}_m^{-1} \tag{7}$$

Since the matrix $\mathbf{T}_p$ is measured from image points corrupted by noise, only a measure $\tilde{\mathbf{T}}_p$ can be used to compute an estimated matrix $\widehat{\mathbf{K}} = \tilde{\mathbf{T}}_p\widehat{\mathbf{T}}_m^{-1}$, where $\widehat{\mathbf{T}}_m$ is an estimation of the unknown matrix $\mathbf{T}_m$ obtained from the bundle adjustment.

III. Metric reconstruction from images

Suppose to have m 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 paper we propose to use a different model of $\mathbf{p}_{ij}$ in the bundle adjustment. Suppose that we have a good estimation of the matrix $\widehat{\mathbf{T}}_m$ (i.e. $\widehat{\mathbf{T}}_m \approx \mathbf{T}_m$). In this case, we have shown in [8] that the error $\widehat{\mathbf{K}} - \mathbf{K}$ can be neglected. As a consequence, using $\widehat{\mathbf{K}}_j$ instead of $\mathbf{K}_j$ in the bundle adjustment has a little effect on the optimization process preserving the optimality of the solution. Note that the estimated matrix $\widehat{\mathbf{T}}_m$ is continuously updated since it is computed from the estimated 3D points. Thus, errors on the estimation of the matrix $\mathbf{T}_m$ will produce errors in the reprojection of image points. Since the goal of bundle adjustment is to minimize the error between the reprojected and measured image points, the result of the minimization will be $\widehat{\mathbf{T}}_m \approx \mathbf{T}_m$ even starting with a bad estimation of $\mathbf{T}_m$. Using $\widehat{\mathbf{K}}_j = \tilde{\mathbf{T}}_{pj}\widehat{\mathbf{T}}_{mj}^{-1}$ instead of $\mathbf{K}$ we have:

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

where $\tilde{\mathbf{T}}_{pj}$ 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 IV, 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 problem:

$$\min_{\mathbf{x}} \mathcal{C}(\mathbf{x}) = \min_{\mathbf{x}} \sum_{i=1}^{n}\sum_{j=1}^{m} \|\tilde{\mathbf{T}}_{pj}\,\widehat{\mathbf{q}}_{ij}(\boldsymbol{\xi}_j, \boldsymbol{\mathcal{X}}) - \tilde{\mathbf{p}}_{ij}\|^2 \ . \tag{8}$$

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 $\tilde{\mathbf{T}}_{pj} \approx \mathbf{T}_{pj}$ (see the experiments in [8]) then our bundle adjustment is close to be a maximum likelihood estimator. By setting $\tilde{\mathbf{q}}_{ij} = \tilde{\mathbf{T}}_{pj}^{-1} \; \tilde{\mathbf{p}}_{ij}$, 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 paper we will use the cost function defined in (8) 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.

A. Computational cost of the non-linear optimization

The optimization problem (8) 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 series of the cost function at each iteration k of the optimization. Setting $\Delta\mathbf{x} = \mathbf{x}_{k+1} - \mathbf{x}_k$ we have:

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

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 $O^3(\Delta\mathbf{x})$ 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, for a standard Newton minimization method, one needs to compute or approximate the gradient vector and the Hessian matrix at each iteration. 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 paper, we use the Levenberg-Marquardt method [2] with numerical differentiation. 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 [13].

B. 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 [3] between the two farthest images and triangulating point features in projective space using [4]. 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. The self-calibration is obtained from the projective reconstruction by using the linear method proposed in [7] 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.

C. Imposing geometric constraints

For most of pinhole cameras it is a reasonable approximation to suppose a zero skew (i.e. $s = 0$) 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_{ij}^p and τ_{ij}^m be the ij entries of matrices $\mathbf{T}_p$ and $\mathbf{T}_m(\mathbf{x})$ respectively. It is easy to obtain from equation (7) the equivalence between the following constraints. Zero skew: $s = 0 \;\Leftrightarrow\; \tau_{12}^p \tau_{11}^m(\mathbf{x}) = \tau_{11}^p \tau_{12}^m(\mathbf{x})$. Unit aspect ratio: $r = 1 \;\Leftrightarrow\; \tau_{11}^p \tau_{22}^m(\mathbf{x}) = \tau_{22}^p \tau_{11}^m(\mathbf{x})$. 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.

D. Expected improvements over standard methods

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 IV.

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. There are $s = 3n + 6m + 5$ unknowns for the standard method while there are $p = 3n + 6m$ unknowns for the intrinsics-free method. 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.

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 of 4 images are needed. There are $s = 3n + 6m + 3m$ unknowns for the standard method while there are $p = 3n + 6m$ unknowns for the intrinsics-free method. 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 considerably reduced.

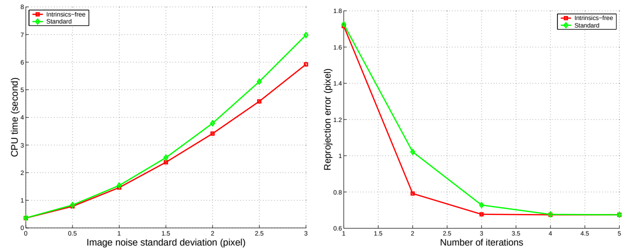
3) Time-varying structure and motion tracked by several zooming cameras.: The third setup is the most general one. The structure, the camera intrinsic and extrinsic parameters are continuously changing. It is an interesting setup for modern robotic vision since it allows the achievement of new applications 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 faster than the standard bundle adjustment, especially when the number of cameras involved in the reconstruction is large.

IV. Experimental Results

A. Simulated Data

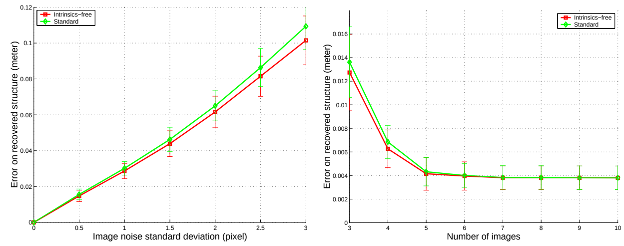
We consider the metric reconstruction of $n = 25$ points lying inside a cube of 1 meter side. The 3D coordinates of the points are reconstructed using m images ($m \geq 3$), taken with one or more cameras depending on the setup. Images are taken around the cube from different positions between 5 and 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. Since the reconstruction is made up to a scalar factor, we use the ground truth just to fix the scale. In order to have an error independent on the frame in which the reconstruction is done, we consider 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 $\epsilon_f = \frac{\widehat{f}_p - f_p}{f_p}$.

1) Fixed intrinsic parameters: In the first set of experiments, we consider a single moving camera with fixed intrinsics parameters. The focal length is $f_p = 1000$, the principal point is at the image center, the skew is zero and the aspect ratio is one. Figure 1 presents the comparison between the properties of convergence and computational cost of the algorithms. As expected, the CPU time is slightly higher for the standard bundle adjustment algorithm especially when the noise standard deviation σ increases (see Figure 1(a)). This result is confirmed by figure 1(b) which shows that the convergence to the minimum of the cost function requires less iterations for the intrinsics-free method. Note that the algorithms converge almost to the same minimum of the cost function. This has been the case in most of our experiments. Figure 2 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 2(a)) and when only few m images are used (Figure 2(b)).



(a) CPU time versus σ (b) Reprojection error vs. N

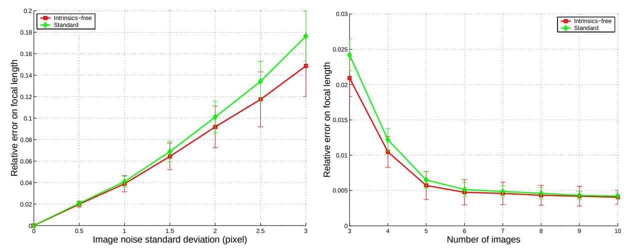
Fig. 1. Comparison of the computational cost between the standard (green) and the intrinsics-free (red) bundle adjustment.



(a) Error $\epsilon_{\mathcal{X}}$ versus σ (b) Error $\epsilon_{\mathcal{X}}$ versus m

Fig. 2. Comparison of the accuracy of the 3D reconstruction between the standard (green) and of the intrinsics-free (red) bundle adjustment.

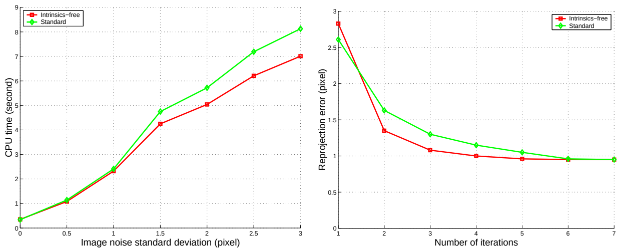
The intrinsic parameters are not directly considered as unknown in our algorithm but they are estimated "a posteriori". Since the 3D structure estimated with the intrinsics-free bundle adjustment is more accurate, it is not surprising, as shown in Figure 3, that the relative error ϵ_f on the focal length is also smaller. Figure 3(a) shows the results obtained when the standard deviation of the noise 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 noise the intrinsics-free method performs better. Figure 3(b) shows the results obtained when varying the number of images used in the reconstruction from 3 to 10. For a small number of images, in this case less than 5, the intrinsics-free method performs better than the standard one.



(a) Error ϵ_f vs. σ (b) Error ϵ_f vs. N

Fig. 3. Comparison of the relative error on the recovered focal length between the standard (green) and of the intrinsics-free (red) methods.

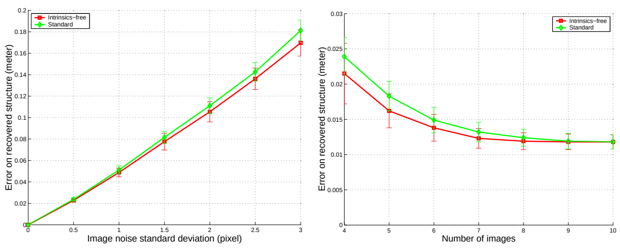
2) Varying focal length and principal point.: In this set of experiments we consider a moving and zooming camera. The focal length f_p changes between 800 and 1300 pixels while moving the camera around the 3D structure. We suppose that $s = 0$, $r = 1$ and that the principal point is at the image center. 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 4. Again, the estimation processes is more stable.



(a) CPU time vs. σ (b) Reprojection error vs. N

Fig. 4. Comparison of the computational cost between the standard (green) and the intrinsics-free (red) bundle adjustment.

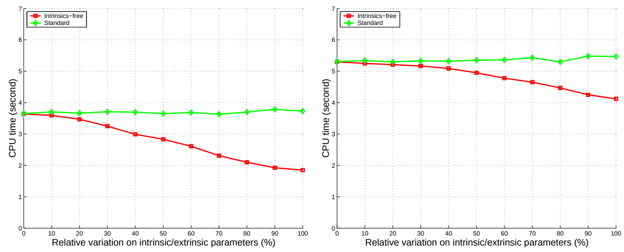
Thus, the reconstruction provided by the intrinsics-free bundle adjustment is more accurate. Figure 5 confirms the results obtained in the previous simulation. With a small number m of images and for higher noise σ the intrinsics-free bundle adjustment performs better. Otherwise, the results are undistinguishable. Concerning the errors on the estimation of the different focal lengths, we have observed, as in the previous case, that 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 (complete results are in [8]).



(a) Error $\epsilon_{\mathcal{X}}$ versus σ (b) Error $\epsilon_{\mathcal{X}}$ versus m

Fig. 5. Comparison of the accuracy of the 3D reconstruction between the standard (green) and of the intrinsics-free (red) bundle adjustment.

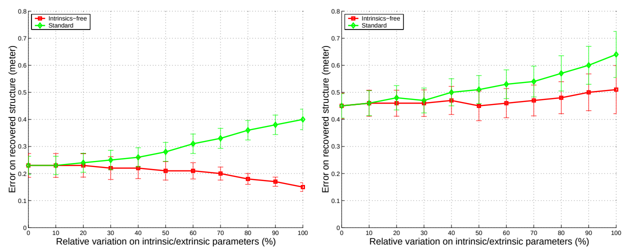
3) Time-varying structure and motion tracked by several zooming cameras.: This experiment consists in tracking time-varying structure and motion. At each iteration, structure and motion are estimated using bundle adjustment. The estimate of iteration k is the starting point of iteration $k+1$. We simulate 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 change. The deformation of the structure through time 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 6 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 6(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 6(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) Rigid structure (b) Non rigid structure

Fig. 6. 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 7. We observe in Figure 7(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 7(b)), the difference between the two algorithm is reduced but still the intrinsics-free bundle adjustment is more accurate.



(a) Rigid structure (b) Non rigid structure

Fig. 7. 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.

B. *Real Images*

In this section, we validate our algorithm using real images of a calibration grid in order to have a ground truth.

1) Fixed intrinsic parameters: In the first experiment, we use 3 images of the calibration grid shown in Figure 8 observed by the same camera.

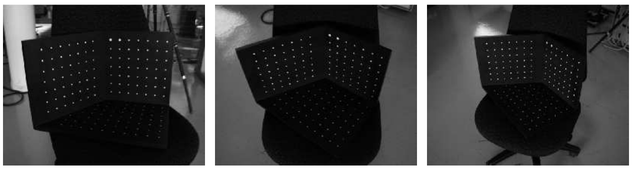


Fig. 8. Three images of a calibration grid used in the experiment.

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 I 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.

	linear	standard	intrinsics-free	ground truth
f_p	750	753	754	758
u_0	256	253	255	260
v_0	320	322	320	327

TABLE I

COMPARISON OF THE CALIBRATION RESULTS BETWEEN THE GROUND TRUTH, THE LINEAR SOLUTION AND THE DIFFERENT METHODS.

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 Figure 9). The model of the grid is used again as ground truth.

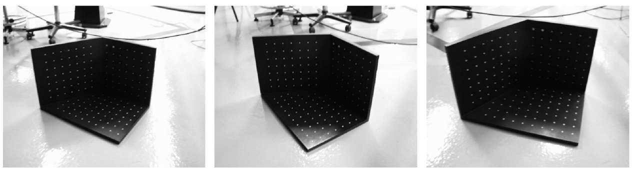


Fig. 9. Three images of a calibration grid used in the experiment.

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 II 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	752	754	757	766	1162	1993	1944
intrinsic-free	772	777	773	778	1202	2176	2100
standard	808	800	790	796	1240	2186	2229
	u_1	u_2	u_3	u_4	u_5	u_6	u_7
ground truth	324	303	295	341	316	321	237
intrinsic-free	335	318	298	350	328	319	317
standard	312	283	267	335	265	113	187
	v_1	v_2	v_3	v_4	v_5	v_6	v_7
ground truth	271	255	251	250	273	294	287
intrinsic-free	260	249	247	243	256	269	261
standard	270	246	238	249	237	110	348

TABLE II

COMPARISON OF THE CALIBRATION RESULTS BETWEEN THE GROUND TRUTH AND THE DIFFERENT BUNDLE ADJUSTMENT TECHNIQUES.

V. CONCLUSION

This paper has presented a new refinement technique for metric reconstruction which is particularly adapted to computer vision applications close to real-time as in robotics. 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] O. Faugeras, and Q-T. Luong. *Geometry of multiple images*. MIT Press, 2001.

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

[3] R. Hartley. In defence of the 8-point algorithm. In *ICCV, Cambridge, Massachusetts, USA*, pages 1064–1070, June 1995.

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

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

[6] F. Kahl and A. Heyden. Euclidean reconstruction and auto-calibration from continuous motion. In *IEEE ICCV*, vol. 2, Canada, July 2001.

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

[8] E. Malis and A. Bartoli. Euclidean Bundle Adjustment Independent on Camera Intrinsic Parameters. Technical Report 4377, Dec. 2001.

[9] E. Malis. A Unifi ed Approach to Model-based and Model-free Visual Servoing In *ECCV*, vol 4, pp. 433-447, Denmark, May 2002.

[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. *IJCV*, 8(2):123–151, 1992.

[12] B. Triggs. Autocalibration and the absolute quadric. In *Proceedings of the Conference on CVPR, 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 *Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Springer-Verlag, 2000.