Bundle adjustment: a fast method with weak initialisation.

Sébastien Cornou[1,2], Michel Dhome[1] and Patrick Sayd[2].

[1]LASMEA, UMR 6602 du CNRS
Blaise Pascal University
F-63000 Clermont-Ferrand
[cornou, dhome]@lasmea.univ-bpclermont.fr

[2]DRT/LIST/DTSI/SLA/LCEI
CEA Saclay
F-91191 Gif-sur-Yvette cedex
[sebastien.cornou, patrick.sayd]@cea.fr

Abstract

Bundle adjustment is one of the cornerstone to recover the scene structure from a sequence of images. The main drawback of this technique, due to nonlinear optimisation, is the need of initial conditions for intrinsic and extrinsic camera parameters and for the 3D structure that we want to reconstruct. In this paper, we demonstrate that we can reduce, comparatively to the standard approach, the number of degrees of freedom. Synthetic and real experiments show that our method is as accurate as the classical one while reducing the computation time and increasing widely the convergence domain on 3D shape and intrinsic parameters.

1 Introduction

Bundle adjustement is a nonlinear method to recover structure of a scene from camera motion. This technique is used in camera calibration, autocalibrated reconstruction and in many other folders of computer vision. Usually, bundle adjustment is considered as the estimation of the 3D structure of the scene and the estimation of camera parameters with the help of a nonlinear optimisation process. Here, we suggest that the main information we want to recover is the structure of the 3D model. We consider that camera extrinsic parameters are not real unknowns because they can often be computed from the knowledge of 3D structure and its projection in images.

In the first part of this article, we explain the general approach of this method. In a second part, we suggest a solution for the reconstruction problem of 3D structure observed with pinhole cameras. Then, we show the efficiency and the large convergence domain of this approach in the case of multiple view reconstruction of a cloud of synthetic points. Last, we give an example of reconstruction from a real sequence of images taken with an uncalibrated camera.

2 General approach

In this paper, we deal with the problem of autocalibration. The a priori knowledges are the existence of 3D features and their projections in images. Bundle adjustment is a classical solution to this problem. As expressed in [1], [3] and [5] this technique is able to

BMVC 2002 doi:10.5244/C.16.20

resist of 3D features occultations in some images and optimal solutions were proposed. The main problem is the need of initialisation of extrinsic parameters. This highlights the Achilles'heel of this approach: importance of initialisation to stay in the convergence domain. Here, we suggest that bundle adjustment parameters can be decomposed in parameters that we can evaluate with a monocular approach and in parameters that need multiple views to be estimated.

We now introduce some notations. Let us consider that the 3D shape can be represented by a vector of parameters denoted Φ_{3D}. A function, called f, links Φ_{3D} and the 3D features (for example, if we consider a cube, the function f expresses the summit positions for a given edge length). The k^{th} 3D feature is $f^k(\Phi_{3D})$. Cameras are usually described by intrinsic and extrinsic parameters. Here, we use these parameters for the i^{th} camera but we build two vectors denoted Φ^i_{Cmo} and Φ_{Cml} (we denote Φ_{Cmo} the vector that contains all the Φ^i_{Cmo}). Φ^i_{Cmo} contains parameters that can be evaluated from one image, the estimation of the parameter vector Φ_{Cml} needs multiple views. We denote p^i_k the projection of the k^{th} 3D feature in the i^{th} camera. We write ρ_i the set of 2D projection p^i_k in the i^{th} camera. The function P projects 3D features in images such as $\tilde{p}^i_k = P(\Phi^i_{Cmo}, \Phi_{Cml}, f^k(\Phi_{3D}))$. Finally, we define a g function verifying $g(\Phi_{3D}, \rho_i, \Phi_{Cml}) = \Phi^i_{Cmo}$.

The classical bundle adjustment consists in an iterative minimization of the distance between the detected p^i_k and the feature estimated $\tilde{p}^i_k$ for the set of nonlinear parameters $(\Phi_{3D}, \Phi_{Cmo}, \Phi_{Cml})$. This criterion is:

$$\mathcal{C} = \sum_{i,k} \|\tilde{p}^i_k - p^i_k\|^2$$

where $\|.\|$ is a distance between 2D features.

We introduce P in place of $\tilde{p}^i_k$:

$$\mathcal{C} = \sum_{i,k} \|P^k(\Phi^i_{Cmo}, \Phi_{Cml}, f(\Phi_{3D})) - p^i_k\|^2$$

then, using the g function:

$$\mathcal{C} = \sum_{i,k} \|P(g_{(f(\Phi_{3D}), \rho_i, \Phi_{Cml})}, \Phi_{Cml}, f^k(\Phi_{3D})) - p^i_k\|^2$$

This criterion does not depend on Φ_{Cmo}. Our approach of the bundle adjustment consists in estimating parameters (Φ_{3D}, Φ_{Cml}) and computing $g(f(\Phi_{3D}), \rho_i, \Phi_{Cml})$ in the core of the criterion $\mathcal{C}$ estimation.

The g function has a really specific position in this algorithm. Obviously, the set of 2D features is not perfect. In fact, our criterion measures the sum of errors produced by, on one hand, the inexact value of parameters (Φ_{3D}, Φ_{Cml}) and, on the other hand, the noise in the 2D features detection. The error on (Φ_{3D}, Φ_{Cml}) and the noise on 2D features are not separable and the error on the estimation of P depends on these factors and their combinations. It means that the part of influence of the noise varies during the estimation. Whatever, our goal is to reconstruct scene with very weak initial conditions. When the process starts with really rough informations the influence of the errors on (Φ_{3D}, Φ_{Cml}) is highly superior to the 2D noise one. When the process is near the solution, the 2D noise is the most influence factor and, if high accuracy is researched, an accurate g function is required.

3 Case of projective cameras

Now, the 3D structure is observed with a set of projective cameras. Available information are approximate 3D structure parameters and projections of features in images. We do not expect all features to be seen in all images.

We showed on the previous section that the number of parameters in bundle adjustment can be reduced. We present an application of this approach of bundle adjustment in the case of pinhole cameras. These cameras are modeled by a set of intrinsic parameters (focal length, principal point, skew parameters...) and with a set of extrinsic parameters which describes the position and the orientation of the sensor.

Now, intrinsic parameters matrix of the camera number i is denoted K^i. The extrinsic parameters matrix is denoted A^i. Other notations are the same as before.

We can thus write the new bundle adjustment criterion. Everything is known except the g function which compute the pose of the camera for a set of 2D/3D relations. Many solutions exist to solve this problem with projective cameras. Thus, if we consider that g exists, we can write:

$$K^i \Longleftrightarrow \Phi^i_{Cml} \text{ and } A^i \Longleftrightarrow g(\Phi_{3D}, \rho_i, K^i)$$

and the criterion is:

$$\mathcal{C} = \sum_{i,k} \|K^i.A^i(f^k(\Phi_{3D})) - p^i_k\|^2.$$

We use Levenberg Marquard algorithm to minimise this nonlinear criterion $\mathcal{C}$. Each time this iterative algorithm computes derivatives it calculates values of $\mathcal{C}$. The outline of the criterion $\mathcal{C}$ computation is:

- **Inputs:** the set of K^i, Φ_{3D} and 2D features detected in images.
- For each view, estimation of A^i by executing the g function (monocular estimation).
- Back-projection of 3D features in images in respect of the set of A^i, K^i and Φ_{3D}.
- The distance measurements between back-projection of 3D points and 2D features.
- **Outputs:** the sum of distances.

The reduction of the number of parameters leads to a shorter computation time. Last, we can expect that this method, taking a better account of the physical situation of the problem, will have a larger convergence area. Our wish is to obtain a fast and accurate reconstruction algorithm. The choice of g modifies results. The size of the convergence domain and the noise sensitivity depend on the g function. Experiments led in the next section demonstrate this.

4 Synthetic experiments

In order to evaluate performance of this new method, we lead first experiments on synthetic data. We choose to evaluate this in the case of a metric reconstruction. We build a virtual sphere of one meter radius. Twenty 3D points are randomly defined on the surface of the sphere. This sphere is opaque and so all 3D points are not seen in all images.

This 3D structure is observed with one pinhole camera. Seven views are taken while the camera is looking to the center of the sphere. The distance between the optical center

of the camera and the center of the sphere is 5 meters. The focal length is 1000 pixels. The image size is 640x480 and the principal point is fixed at (320,240) . The camera moves in the equatorial plane of the sphere. The seven points of view are regularly located around the sphere. The intrinsic camera parameters, except the focal length, are supposed to be known.

We evaluate several influence factors: the influence of the 3D points initialisation, the influence of the focal length initialisation, the influence of noise on 2D image points.

We test 3 versions of our method which differ by the choice of the g function and one version of the classical approach:

1- DementhonBA [1] : The g function used is the Dementhon algorithm [2]. Dementhon algorithm starts the pose estimation problem from a scale orthographic camera model and minimises iteratively a non geometric criterion to obtain a pose estimation in respect of the projective model. This algorithm is fast and does not need any initialisation.

2- DemLoweBA : The g function used is the Dementhon algorithm followed by the Lowe algorithm [4]. The Lowe algorithm consists in an iterative process (Newton-Raphson approach) which minimises a geometric criterion. Thus, it is slower but more accurate than the Dementhon algorithm.

3- DemLoweEndBA : We use first *DementhonBA* to obtain a reconstruction of the scene. This reconstruction is then used as initialisation for the *DemLoweBA* algorithm.

4- ClassicBA : This is the classical algorithm. It minimises the sum of distances between 2D points selected in images and back-projected 3D points. It needs an initialisation of the extrinsic parameters associated with each view. This initialisation is obtained by running the Dementhon algorithm, using the rough initialisation of 3D features, before the beginning of the estimation.

In all cases, Levenberg-Marquardt algorithm is used to minimise the nonlinear criterion $\mathcal{C}$ with a maximum of 25 iterations. In order to compare computation time, we use very similar implementations.

We denote m the number of cameras and n the number of 3D points.

ClassicBA admits as unknown parameters the focal length, orientation of each camera described with Euler angles, translation vector between the optical center and the center of the sphere and (X, Y, Z) coordinates of the n 3D points. So, there are $(6m + 3n + 1)$ degrees of freedom.

DementhonBA , *DemLoweBA* and *DemLoweEndBA* admit as unknown parameters the focal length and the (X, Y, Z) coordinates of 3D points. So, $(3n + 1)$ degrees of freedom are considered.

In next experiments, $n = 20$ and $m = 7$, we have 61 degrees of freedom with our method and 103 degrees of freedom with *ClassicBA* .

4.1 Influence of 3D point initialisation

In this test, a random noise is applied to 3D point initial positions. Initial focal length and the 2D points projections are exact. Results are presented on figure 1 [2].

[1] BA = Bundle Adjustment.

[2] Error bars show the standard deviation calculated on 20 trials. For the clearness of the drawing, the results of each method are presented with a little horizontal offset, the correct position is the *ClassicBA* one.

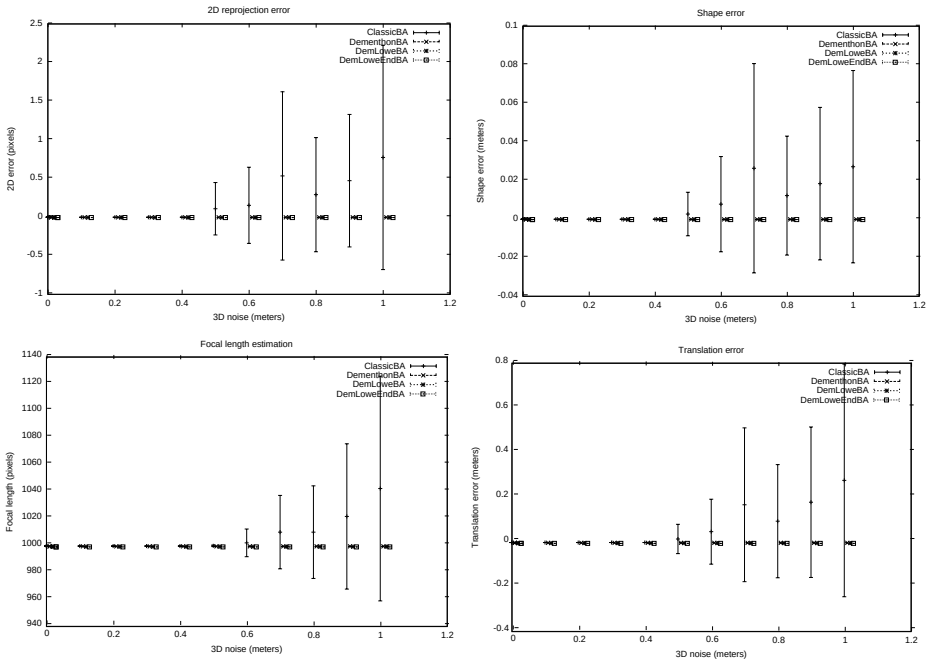

Figure 1: Performances of algorithms in function of the error on features detection.

We observe that *ClassicBA* is not able to reconstruct the scene while the other algorithms offer high quality results. This highlights the difference of criterion between the two approaches and the improvement given by reducing the number of parameters. It demonstrates that the new approach allows very rough initialisation of 3D feature parameters.

4.2 Influence of focal length initialisation

We showed in the previous section that our method is very tolerant on 3D shape initialisation. We now study the behaviour in case of bad initialisation of the focal length. Results are described on figure 2. On those graphs each point describes an experiment.

With our method, no knowledge of the focal length is required to perform estimation because of a very large convergence domain.

4.3 Influence of noise on 2D points

In this test, a gaussian noise is added to detected 2D point. Initial focal length and 3D point positions are exact. Results are summarised on figure 3^2.

First, we observe that *DementhonBA* has the highest sensitivity to 2D noise if we consider the 2D reprojection error. We note the really close comportement of *ClassicBA* and *DemenLoweEndBA* . *DemenLoweBA* seems to be as accurate as *ClassicBA* in this case. This test confirms the importance of the choice of the g function.

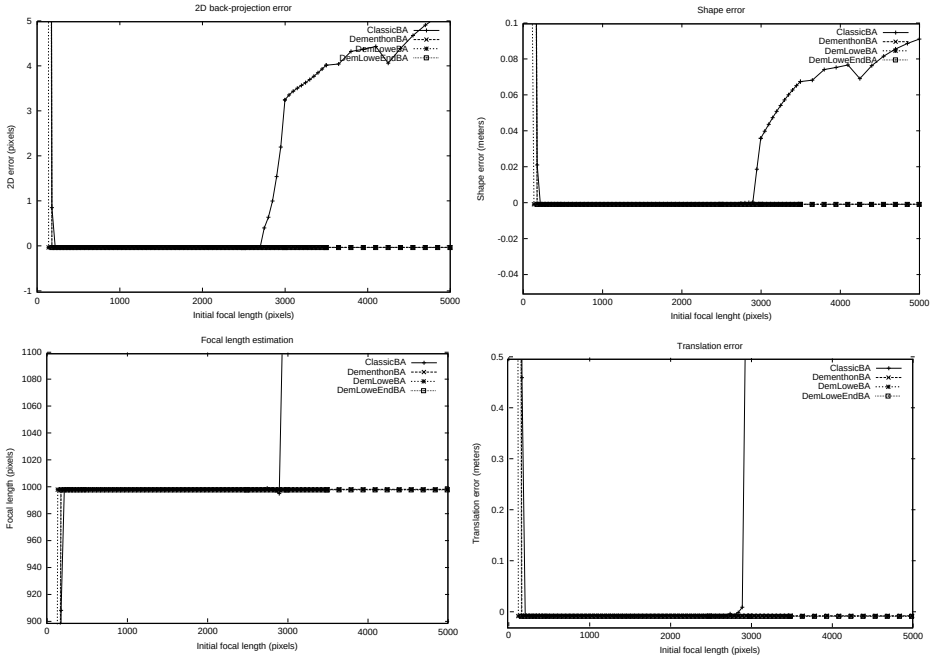

Figure 2: Influence of focal length initialisation.

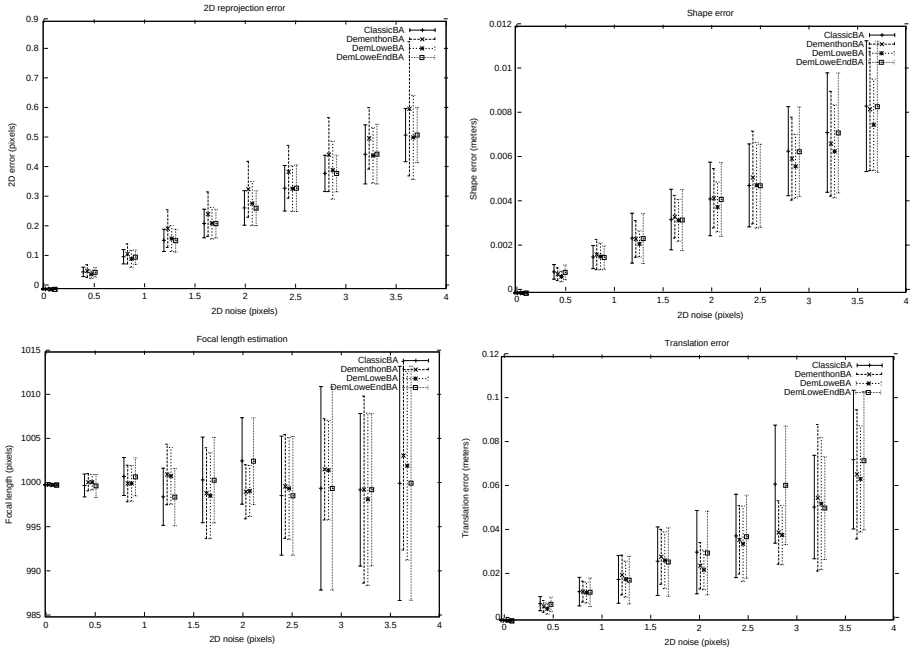

Figure 3: Algorithms evaluation in function of errors on 2D features detection.

4.4 Computation time

In the same test conditions, we present here the influence of several parameters on the computation time. To compare computation times we use very closed implementations of algorithms. Results are given on figure 4.

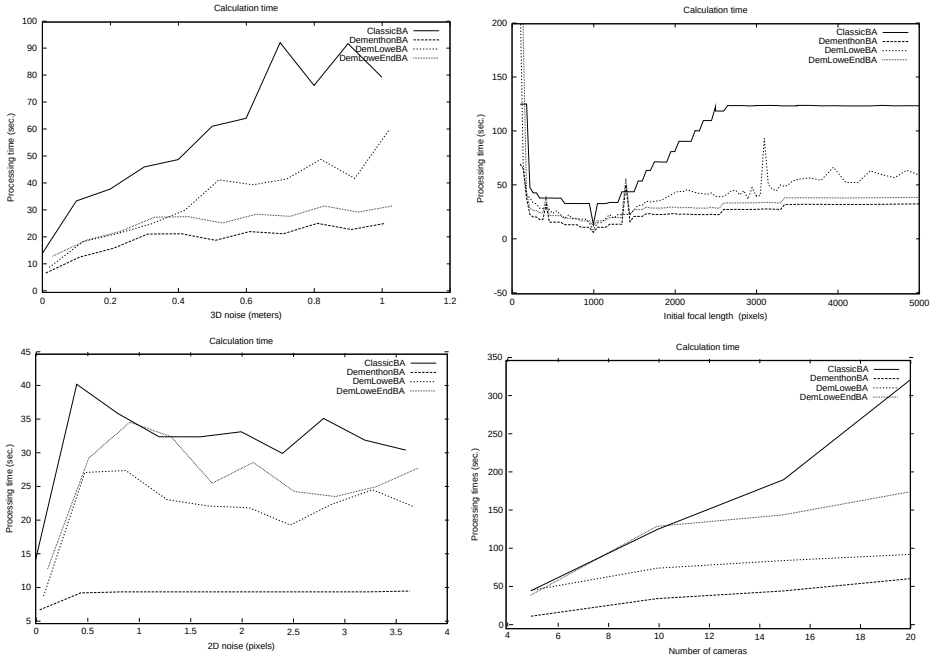

Figure 4: Computation time in front of several test.

We notice that all algorithms based on the new approach are faster than the classic one. It can be simply explained by the lower number of parameters in the nonlinear optimisation (61 parameters versus 103). We can also notice that the choice of the g function influences the computation time. *DementhonBA* is clearly the fastest one.

About computation time, the main advantage of the new approach is to reduce the number of estimated parameters. Obviously, this can offer a tremendous advantage if the number of shots increases. This is demonstrated by the experimental results presented on figure 4 (botton-right). In this experiment, the same sphere is used but, this time, a gaussian noise with standard deviation of 0.5 pixel was added to 2D points. The initial focal length was fixed to 1100 pixels (the right one was 1000 pixels) and the 3D points were perturbed with a gaussian noise whose standard deviation was equal to 0.01 meter. Computation times are quite equal at the beginning but when the number of cameras increases our method is clearly faster than the classical one. *DementhonBA* is the fastest algorithm. It takes advantage of the high speed calculation due to the Dementhon pose estimation algorithm.

4.5 Success Rate

During previous tests we rejected outliers for which focal length estimation was up to

5000 pixels and 2D projection error was up to 5 pixels. We give on figure 5 the success rate for the 2D noise test and for the 3D initialisation test.

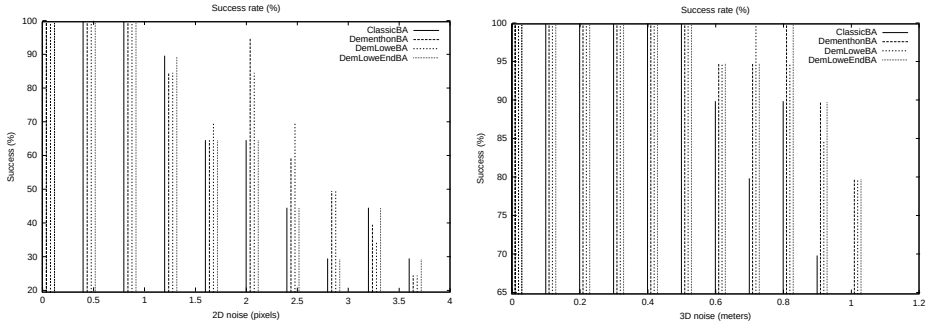

Figure 5: Computation time in front of several test.

We can observe that, in the case of 2D noise, *DementhonBA* and *DemLoweBA* obtain better results than the two other one when the noise level increases. In the case of 3D initialisation, the worst results are given by *ClassicBA* due to a shorter convergence domain.

4.6 Conclusion on synthetic experiments

In this section, we have evaluated performance of the new approach. These experiments show the influence of the g function choice on algorithm performance. The new approach offers a larger convergence domain in term of 3D features initialisation and is as accurate as the classic one. It supports a large number of images. Table 1 shows results obtained for conditions close to the real ones. *DementhonBA* seems to be the good choice for fast reconstruction while *DemLoweBA* seems to offer a lower reprojection errors. Having studied our approach on synthetic data, we will now evaluate it on real images.

Algorithm	Time (sec.)	reprojection errors (pix.)	Focal length (pix.)
ClassicBA	$36.04 - 108.9$	$0.0602 - 0.1332$	$1001.181 - 1001.136$
DementhonBA	$11.88 - 14.96$	$0.0805 - 0.0805$	$999.260 - 999.260$
DemLoweBA	$29.04 - 24.78$	$0.0607 - 0.0607$	$1001.179 - 1001,179$
DemLoweEndBA	$35.43 - 34.65$	$0.0697 - 0.0697$	$999.177 - 999.178$

Table 1: synthetic experiment. First number: 2D error=0.5 pixels, 3D initial error 0.01 meters and initial focal length 1100 pixels, second number: 2D error=0.5 pixels, 3D initial error 0.1 meters and initial focal length 1500 pixels.

5 Real data experiment

We test all the previous algorithms on a sequence of 9 images taken with a single camera. In each image we have manually selected visible corners of 5 rectangular parallelipipeds.

Each 3D element is modeled with 3 parameters (length,width, height). One of the elements is considered as the center of the world frame, the other are located in this frame using 6 parameters (rotation and translation with respect to the first element). These parameters are estimated in the reconstruction process except one length given by the user to fixe the scale factor. 101 parameters are estimated by *ClassicBA* while the other method just estimates 38 parameters. The 3D element are initialised as unit cubes located as shown on figure 6. The focal length is initially at 2500 pixels. For *ClassicBA* , the initialisation of extrinsic parameters has been obtained by the Dementhon algorithm (3D element initialisation is used).

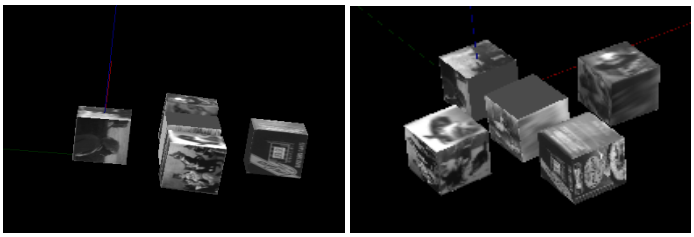

Figure 6: 2 camera points of view just after the initialisation step (*ClassicBA* algorithm).

Figure 7 shows pictures of the initial scene and the reconstruction. Figure 8 presents a detail of the reconstruction. With *ClassicBA* reconstruction 2 cubes intersect while with DementhonBA they are just in contact. This gives a visual evidence of errors existing in the *ClassicBA* reconstruction.

Figure 7: Results of reconstruction on a real sequence. An image used for the reconstruction (left), the textured 3D model given by *DemLowBA* (right).

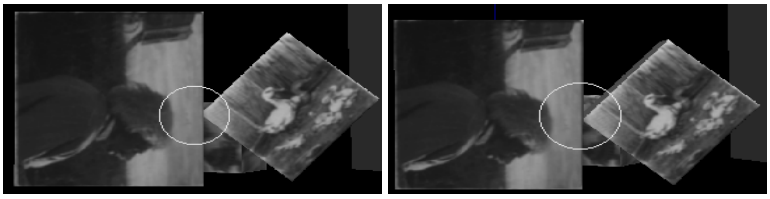

Figure 8: With *ClassicBA* (left), *DemLoweBA* (right).

Table 2 summarizes performances obtained with different methods. In this case, *DementhonBA* is the best solution. It offers a fast reconstruction with a good accuracy. We

	CBA	DBA	$DLBA$	$DLEBA$
Time (sec.)	136	49	75	86
Focal length (pixels)	3561	2416	2424	2424
σ estimated on f (pixels)	260	41	40	40
Mean of error on dimension (mm)	4.0	2.6	2.7	2.7
error/Dimension*100	5.25	3.40	3.49	3.49

Table 2: Results of the experimental evaluation of the algorithms (*CBA=ClassicBA, DBA=DementhonBA, DLBA=DemLoweBA and DLEBA=DemLoweEndBA).*

observe that *ClassicBA* is unable to provide a good estimation of the focal length (the real one is near 2400 pixels).

6 Conclusion

Furthermore, we presented a new approach for bundle adjustment. First, this method needs neither camera pose a priori knowledge nor accurate initialisation of other parameters. This simplifies the initialisation step which is critical for classical bundle adjustment.

Our optimisation process is focused on parameters that need multiple views to be estimated. The reduced set of parameters provides a faster method than the classical bundle adjustment. Nevertheless, the presented experiments showed that our method gives as accurate results as the classical one (when this latter converges!). Moreover, if great accuracy is not an issue, the computation time can be further lowered.

Such an approach can improve performance in many computer vision applications, for example accurate calibration or autocalibrated recontruction.

References

[1] H.A. Beyer. Geometric and radiometric analysis of a CCD-camera based photogrammetric close-range system. *PhD thesis, Institut fur Geodasie und Photogrammetry*, Nr 51, ETH, Zurich, May 1992.

[2] L.S.Davis D.F.Dementhon. Model-based object pose in 25 lines of code. *International Journal of Computer Vison*, 15(2):123–141, 1995.

[3] Richard Hartley and Andrew Zisserman. *Multiple View Geometry*. Cambridge university press, 2000.

[4] D.G. Lowe. In *Perceptual Organization and Visual Recognition, Kluwer, Boston*, page Chap 7, 1985.

[5] Bill Triggs, Philip McLauchlan, Richard Hartley, and Andrew Fitzgibbon. Bundle adjustment – A modern synthesis. In W. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, LNCS, pages 298–375. Springer Verlag, 2000.