Vol. 9(32), Apr. 2019, PP. 4123-4132

Contents list available at IJMEC

International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)

Journal Homepage: www.aeuso.org

PISSN: 2411-6173 - EISSN: 2305-0543

Bundle Adjustment Algorithm with Respect to Synthetic Scene

Abdulkadir Iyyaka Audu[1*] and Abdul Hamid Sadka[2]

[1] *Computer Engineering Department, Faculty of Engineering, University of Maiduguri*

[2] *Department Electronic and Computer, Engineering, Brunel University, London, United Kingdom*

Phone Number: *+234-0709252528*

****Corresponding Author's E-mail:*** *lm324fairchild@yahoo.com*

Abstract

Bundle adjustment Algorithm is an optimization technique which shows potential to improve the reconstruction of three-dimensional scene starting with multiple two-dimensional representation of different parts of the same scene. Since the objects of interest in the world are usually a set of points, this paper presents performance evaluation of bundle adjustment algorithm in the construction of three-dimensional scene from randomly generated two-dimensional points. This is in contrast to the ego-motion techniques used in the generation of a cloud of three-dimensional points in the study of structure and motion.

Keywords: *Perspective projection, scene reconstruction, synthesized scene, bundle adjustment, structure and motion.*

1. Introduction

Structure and motion (SAM) recovery as evidenced by its application in movie production, architecture design and advertisement, is an exceptional problem. It involves the use of two or three view epipolar constraints [1], in the algebraic initialization of the projective structure and motion. The projective reconstruction is further enhanced to a metric one by some self-calibration techniques [2]. Bundle adjustment (BA) algorithm is widely employed in the optimization of the solution obtained through these techniques [3, 4]. BA algorithm is a computational technique in which the geometry of the scene, camera intrinsic parameters, positions and orientation are solved based on a database of features automatically extracted from a set of multiple overlapping images [5]. These images must capture full three-dimensional (3D) structure of the scene viewed from a wide array of positions. The analysis and generation of point cloud which is sufficiently representative of objects in 3D scenes and the corresponding numerical simulations is an interesting and challenging problem in 3D reconstruction. In the work of [6], in which five simple cues are motivated to model specific patterns of motion and structure, a cloud of 3D points are automatically generated from video sequences using ego-motion techniques. The contribution of this paper, from a practical point of view, is to evaluate the performance of BA in the reconstruction of 3D scene from multiple randomly generated two-dimensional (2D) points. The paper is organized as follows: In Section 2, BA is described. Section 3 presents the concept of synthetic scene. The obtained results are presented in Section 4. The conclusion of this work is drawn in Section 5.

2. Bundle Adjustment Problem Formulation

A camera can be modeled in several different ways. Affine and orthographic projections are sometimes useful for distant cameras, and more exotic models such as push-broom and rational

Article History:
Received Date: Oct. 08, 2018
Accepted Date: Dec. 06, 2018
Available Online: Jan. 04, 2019

polynomial cameras are needed for certain applications [3]. Other camera models can be derived from it. But in addition to pose (position and orientation), and simple internal parameters such as focal length and principal point, real cameras also require various types of additional parameters to model internal aberrations such as radial distortion. However, perspective projection, [7], is the most prominent.

Perspective projection is the linear mapping between the extended coordinates of any world point *M* and its corresponding image point *m* such that collinearity property exists between *M, m*, and *C* (center of projection). This can be expressed as

$$\Lambda_m = QM = P \quad (1)$$

where *M* is 4 × 1 vector and *m* is 3 × 1 vector. Λ is an arbitrary scale factor. Q is a 3 × 4vector referred to as the projection matrix. Therefore, in reality, an object or structure consists of several *M* points and image of such an object or structure will consist of a corresponding number of *m* points. An important characteristic which can be exhibited by a perspective camera is for the first three left columns of *M* to be non-singular. It therefore means *Q* can be further decomposed such that

$$\Lambda m = K \left[R| \quad t \right] M \quad (2)$$

K is a 3 × 3 upper triangular matrix. It is called camera or calibration matrix of the camera. It comprises of the optical properties of the camera namely: Focal length, principal point and aspect ratio. *R* is an orthogonal 3 × 3 matrix and *t* a 3 × 1 vector. R and t are collectively referred to as the camera's extrinsic orientation and correspond, respectively, to the rotation and translation that make up the rigid transformation from the world to the camera coordinate frame [4]. The coordinate system *C* attached to the camera is related to the world coordinate system through a rotation R followed by a translation t.

For a multi-view setting, consider having M_j scene points captured by several cameras described by Q^i. Assuming the projection of M_j point due to Q^i camera is $m^i{}_j$. Starting from multi-view image samples, multi-view 3D reconstruction involves the determination of M_j and Q^i such that (3) is satisfied as expressed in (3)

$$\Lambda m_j^i = Q^i M_j \quad (3)$$

A significant challenge in SBA is that (3) is not exactly satisfied. m_{ij} has inherent noise superimposed during the measurement process. Therefore, for every image point m_j^i, a predictive model $\Lambda m_j^i = m(Q^i, M_j)$, is required such that

$$\Delta \hat{m}_j^i \equiv m_j^i - m\left(Q^i, M_j\right) \quad (4)$$

$$f\left(Q, M\right) = \sum_{i=1}^{c} \sum_{j=1}^{d} V_{ij} || m_j^i - m\left(Q^i, M_j\right) ||^2 \quad (5)$$

Equation (4) is referred to as re-projection error. Hence the problem of scene reconstruction and camera parameter estimation boils down to the minimization of the re-projection error between the image locations of observed and predicted image points, which is expressed as the sum of squares of a large number of nonlinear, real-valued functions. The objective function for the minimization problem defined in the context of bundle adjustment is expressed in (5), where *c* is the number of scene points and *d* is the number of cameras. V_{ij} is a visibility weight which equals 1 if a scene point *j* can be seen in camera *i*, otherwise it equals 0. If the unexpected variation in pixel coordinates is modeled as Gaussian noise with zero mean, then (5) becomes a statistical nonlinear model. Using the condition of linear independence of the columns of Q^i, (5) can be expressed as a set of linear equations as in (6).

$$\left(Q^i\right)^T Q^i \hat{f} = \left(Q^i\right)^T m_j^i \quad (6)$$

The parameters of (6) can be estimated using least-squares algorithm like Levenberg-Marquardt (LMA) [8]. However, for a large set of object points and camera parameters which, constitute the unknown contributing to the minimized re-projection error, the system represented by (6) becomes overdetermined. The computational cost will then have cubic complexity [9]. Therefore, a specialized LMA known as SBA is required in order to seek a minimal solution.

3. Synthetic Scene

Here, Here, synthetic scene implies using a set of randomly generated points to represent a scene. This idea emanates from the fact that any worldly object or structure is comprised of a set of points. With regards to this work, 3D reconstruction can be defined as the problem of using 2D measurements arising from a set of randomly generated points, depicting the same scene from different viewpoints, aiming to derive information related to the 3D scene geometry as well as the relative motion and the optical characteristics of the camera(s) employed to acquire these images.

To demonstrate the idea of scene reconstruction, three fundamental steps are outlined. 1) These are creation of scene points and camera pose, 2) superposition of error on the scene and pose data, 3) and data recovery through bundle adjustment technique.

The implementation of the first step starts with the assumption of a certain number of cameras of known intrinsic parameters and image resolution. The poses of the assumed cameras are generated in Matlab based on normally distributed pseudorandom numbers. Mersenne twister (default), Multiplicative congruential generator, Multiplicative lagged Fibonacci generator, Multiplicative lagged Fibonacci generator, Shift-register generator summed with linear congruential generator, and Modified subtract with borrow generator are the random number generators available in Matlab. Mersenne twister has been used in this work.

Mersenne Twister provides a super astronomical period of 219937 − 1 and 623-dimensional equi-distribution up to 32-bit accuracy, while using a working area of only 624 words [10]. This is a new variant of the previously proposed generators, Twisted Generalized Feedback Shift Register (TGFSR), modified so as to admit a Mersenne-prime period. Two new ideas have been added to the existing twisted Generalized Feedback Shift Register (GFSR). One is the realization of Mersenne-prime period using incomplete array. Second, the primitivity of the characteristic polynomial of a linear recurrence can be tested with a fast algorithm. This is called the inversive-decimation method. The characteristic polynomial has many terms.

450 2D Normally Distributed Pseudorandom Pixels Locations (NDPPL) with corresponding depths are generated on the image plane of the reference camera. 450 is the expected maximum number of features to be tracked on each frame of 640 × 480 resolution. Using the chosen reference camera (first camera), the pixel locations are re-projected so as to determine the 3D points. Only pixel locations with positive depth are re-projected and registered in the Number of scene points × number of cameras, visibility map. The generated 3D points are tested to see if they can be seen in the second

camera. Since the number of pixels seen by the second camera is less than the maximum expected, another set of NDPPL is generated based on the difference between the maximum expected to be tracked and number already tracked. The new set of NDPPL is again tested for positivity of depth and the pixels with positive depth are re-projected in order to generate their corresponding 3D points. This process is repeated until all the cameras have been considered. The total number of 3D points generated through re-projection of pixel locations in the process just described constitute the scene points. This is a reliable method which is widely used and also documented in literature.

Next, the scene points, camera pose, camera matrix, and pixel locations are corrupted with Normally Distributed Pseudorandom Error (NDPE). A statistical error is usually given by the standard deviation ρ, equal to the square root of variance (expectation value of the square of the difference to the mean).

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f\,(x) dx \qquad (7)$$

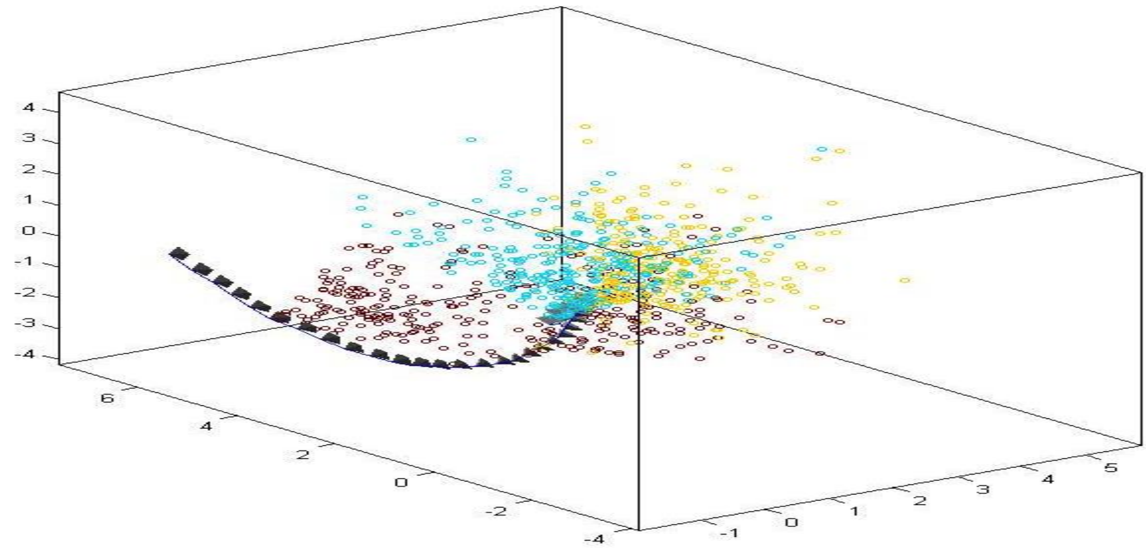
where x is a measure of data. μ is the mean of the data set.

In the third step, an attempt is then made to recover scene points and cam- era parameters using the bundle adjustment. In the ordinary bundle adjustment, advantage is not taken of the existence sparsity. The implementation of ordinary bundle adjustment involves error minimization by way of optimization of parameters purely performed based on Matlab coding.

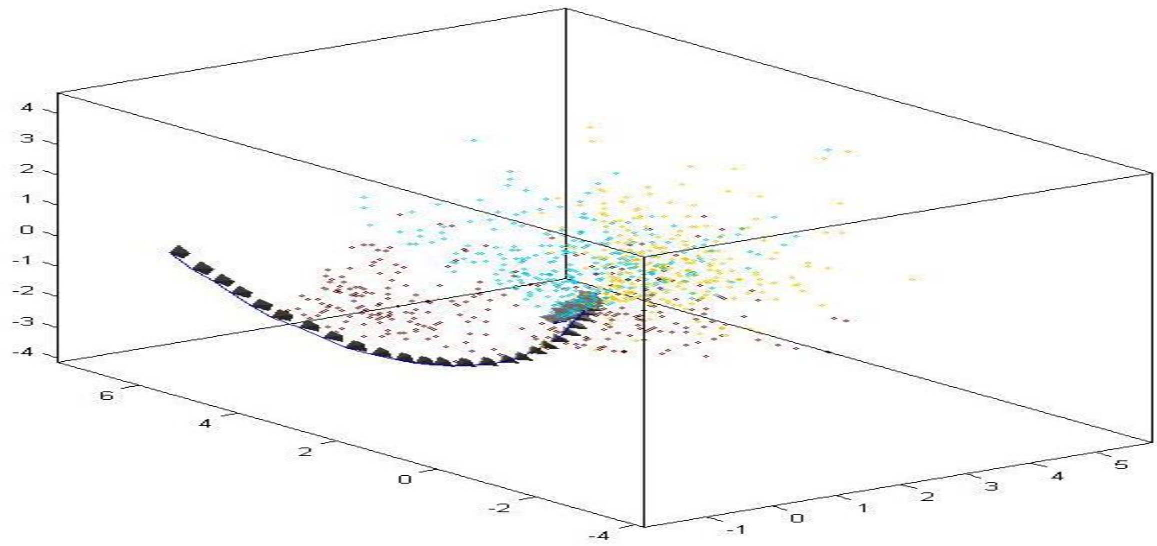
4. Experimental Result

A simulated A simulated experiment based on the earlier mentioned steps in section 3 was conducted in Matlab environment. A set of scene points is created as shown in Figure 1 (a). The points are represented with a circle in red, blue, and yellow colors with respect to some cameras symbolically represented with triangles. Fifty cameras of known calibration were used to capture these scene points from different observation points. Each frame has a resolution of 640 × 480. In order to demonstrate the concept of scene reconstruction, the coordinates of the scene point, camera calibration matrix, rotation and translation were corrupted with some randomly generated error. Bundle adjustment was then used to reconstruct the scene points and camera parameters (intrinsic and extrinsic). Figure 1 (d) provides information regarding the number of scene points that is captured in each camera frame.

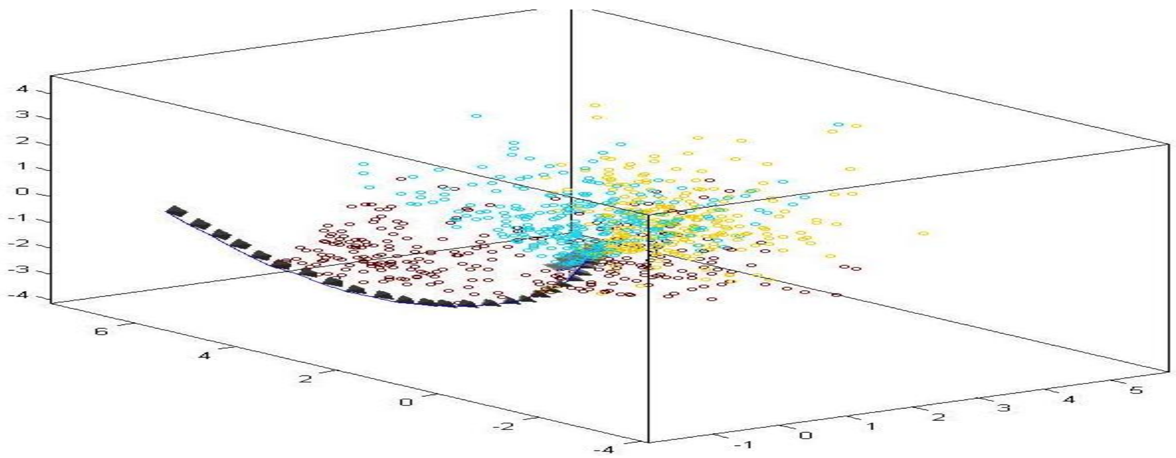
The motion of the cameras from one point to another usually comprises of rotation and translation. Therefore, Figure 2 and 3 present the recovered rotation and translation of the cameras. Rotation is about the x, y and z axis. Translation of the camera is in the x, y, and z directions. The recovered intrinsic camera parameters are presented in Figure 4. The intrinsic parameters appear as straight lines since they have the same values for all the cameras (cameras of the same model).


4
3
2
1
0
-1
-2
-3
-4
6
4
2
0
-2
-4
-1
0
1
2
3
4
5

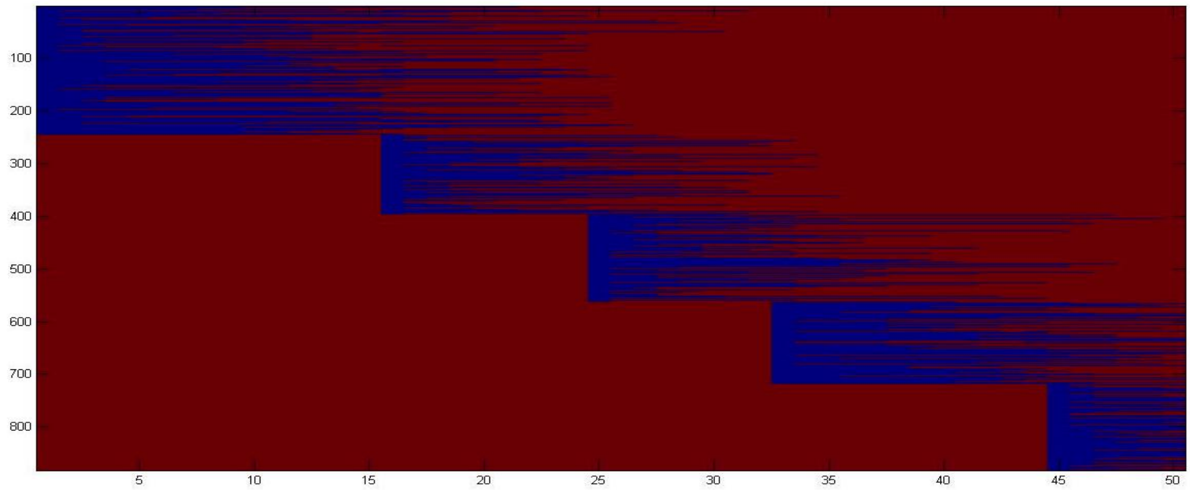
(a)


4
3
2
1
0
-1
-2
-3
-4
6
4
2
0
-2
-4
-1
0
1
2
3
4
5

(b)


4
3
2
1
0
-1
-2
-3
-4
6
4
2
0
-2
-4
-1
0
1
2
3
4
5

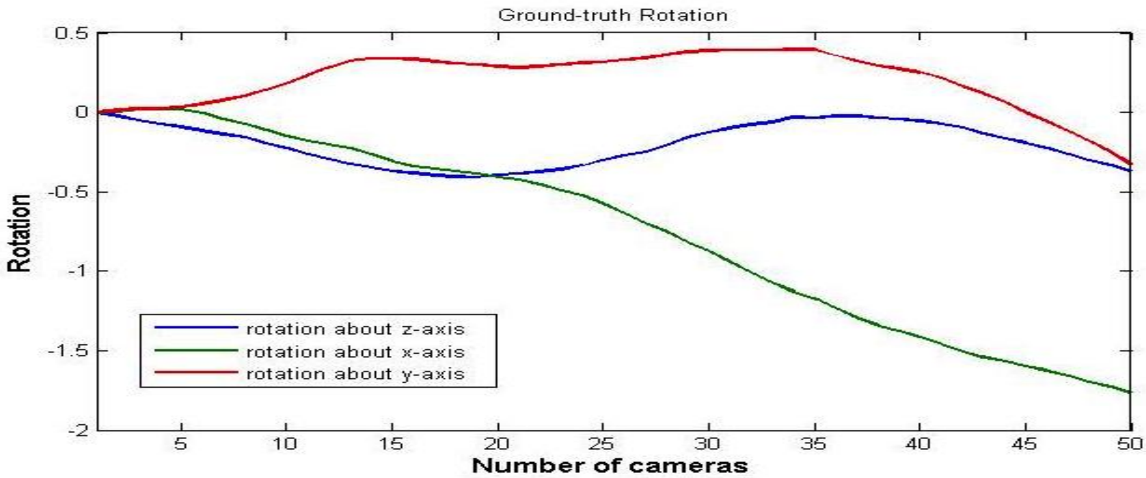
(c)



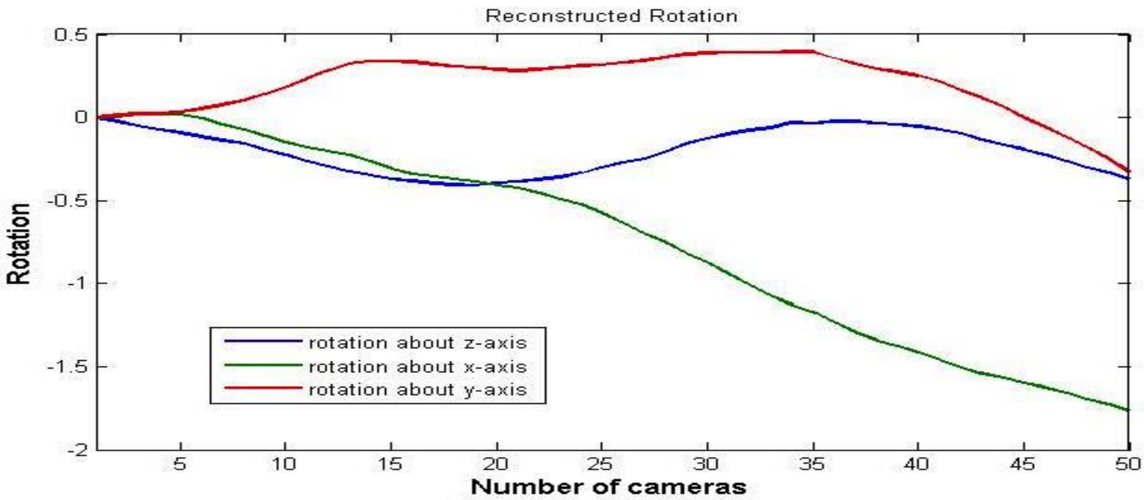
(d)

Figure 1: Demonstration of bundle adjustment. (a) shows a set of points representing a synthetic scene. (b) presents the corrupted scene. (c) shows the reconstructed scene. (d) shows the number of scene points captured by each camera.

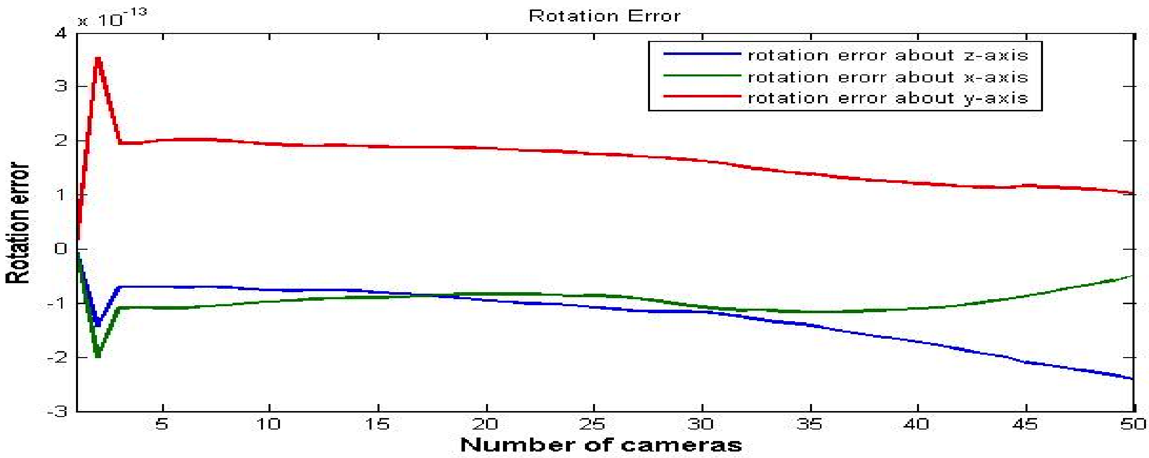
The motion of the cameras from one point to another usually comprises of rotation and translation. Therefore, Figure 2 and 3 present the recovered rotation and translation of the cameras. Rotation is about the x, y and z axis. Translation of the camera is in the x, y, and z directions. The recovered intrinsic camera parameters are presented in Figure 4. The intrinsic parameters appear as straight lines since they have the same values for all the cameras (cameras of the same model).



(a)

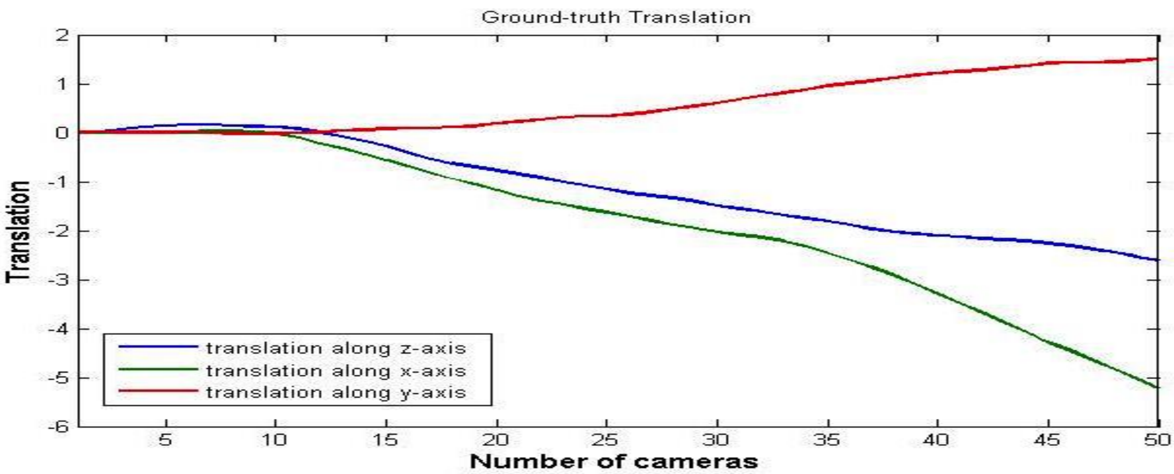


(b)

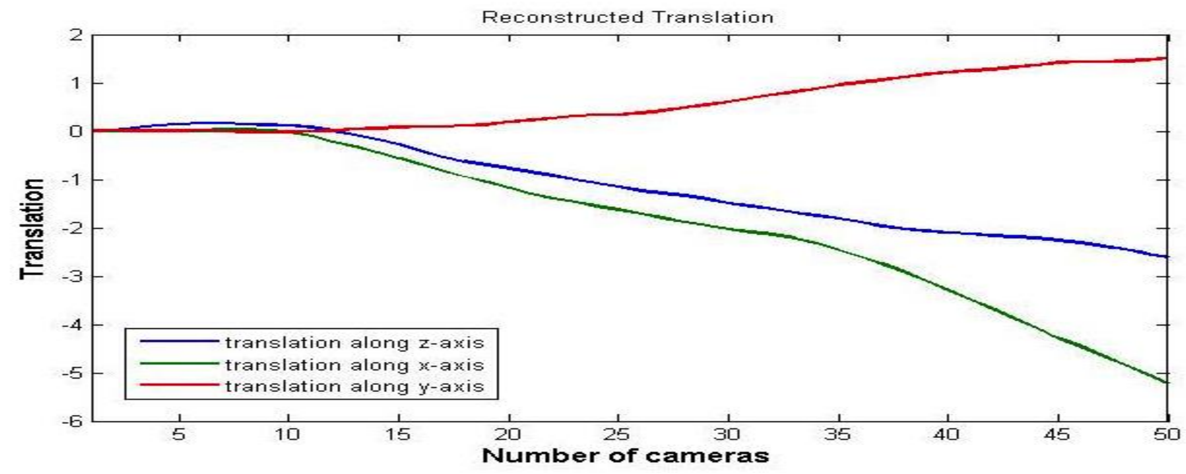


(c)

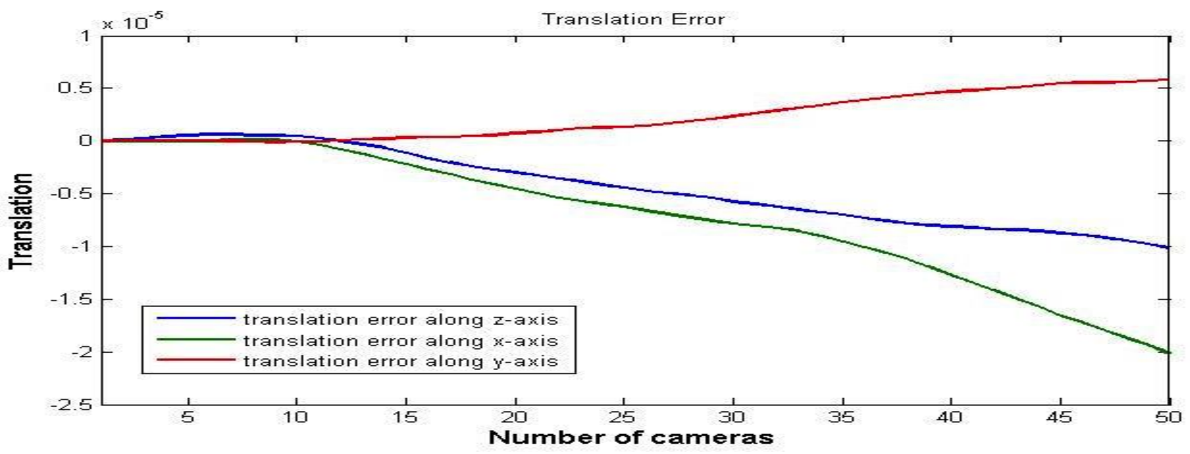
Figure 2: Demonstration of bundle adjustment. In this figure is shown the camera rotation about x, y, and z axis: (a) ground-truth rotation, (b) camera rotation required to cope with the corrupted of scene, (c) computed camera rotation error.



(a)

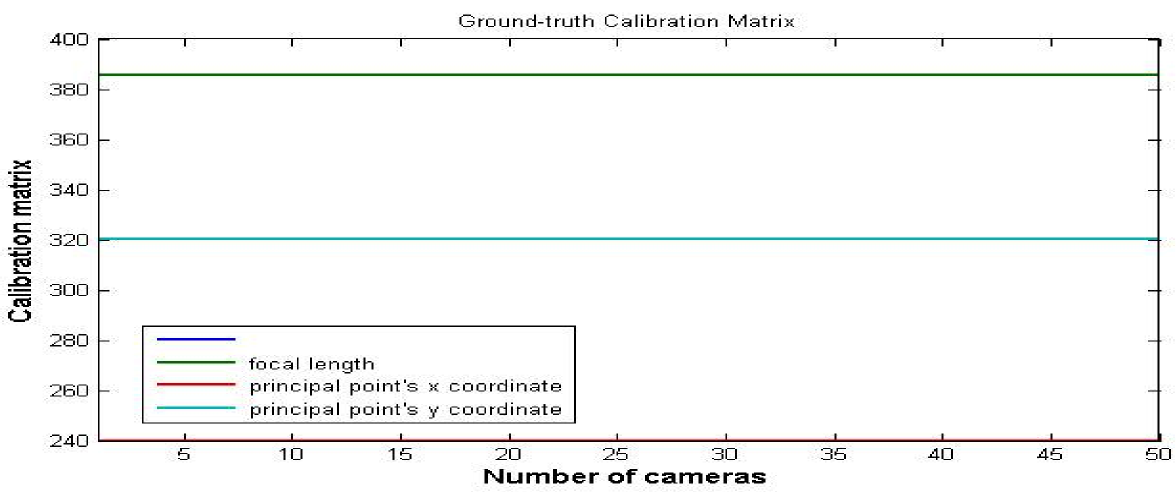


(b)

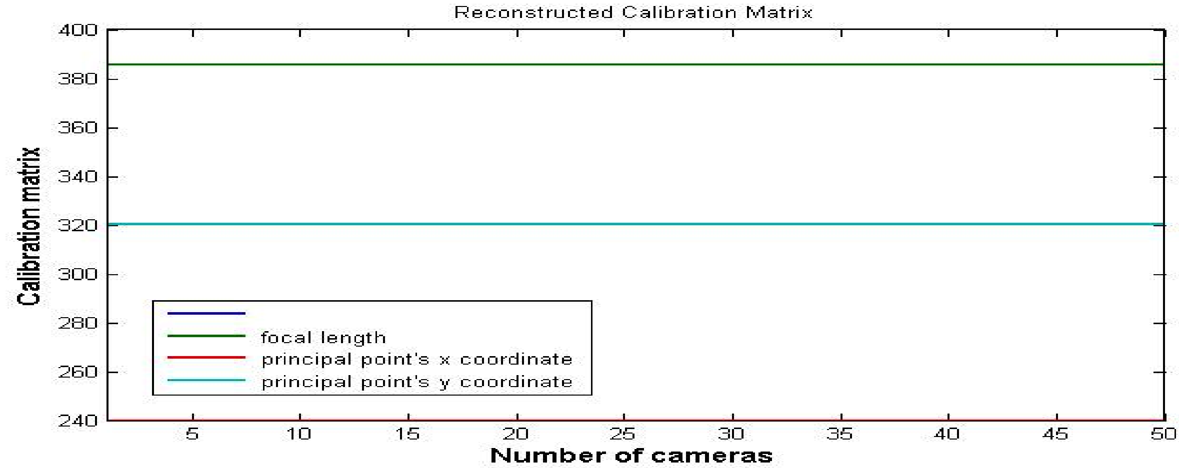


(c)

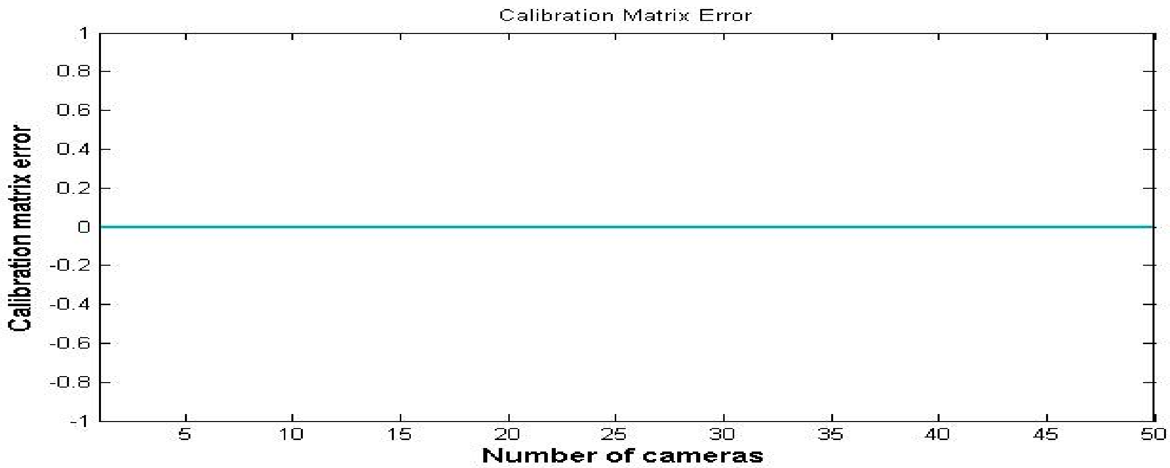
Figure 3: Demonstration of bundle adjustment. camera translation is presented: (a) camera ground-truth translation, (b) translation as a result of corruption of scene, and (c) translation error.



(a)



(b)



(c)

Figure 4: Demonstration of bundle adjustment. camera calibration matrix: (a) ground-truth calibration matrix, (b) calibration matrix as a result of corruption of scene, (c) calibration matrix error. A straight line plot means the intrinsic parameters are the same for all the cameras used.

Conclusion

This work presents a complete process of 3D reconstruction from a set of 2D synthetic scenes and computation of camera parameters primarily through BA. A preview of the complexity of the task of 3D reconstruction has been highlighted and the widely used tool (BA) to achieve accurate scene reconstruction has been described. Promising results in the simulation include recovery camera pose and intrinsic parameters. Information about camera tracks have equally been obtained.

References

[1] R. C. Bolles, H. H. Baker and D. H. Marimont, Epipolar-plane image analysis: An approach to determining structure from motion, International journal of computer vision 1 (1) (1987) 7-55.

[2] G. Zhang, X. Qin, W. Hua, T. T. Wong, P. A. Heng. and H. Bao, Robust metric reconstruction from challenging video sequences, In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference, (2007) 1-8.

[3] B. Triggs, P. F. McLauchlan, R. I. Hartley and A. W. Fitzgibbon, Bundle adjustment—a modern synthesis, In International workshop on vision algorithms (1999) 298-372.

[4] M. I. Lourakis and A. A. Argyros, SBA: A software package for generic sparse bundle adjustment, ACM Transactions on Mathematical Software (TOMS)3 (1) (2009) 2.

[5] M. J. Westoby, J. Brasington, N. F. Glasser, M. J. Hambrey and J. M. Reynolds, Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications, Geomorphology, 179 (2012) 300-314.

[6] G. J. Brostow, J. Shotton, J. Fauqueur and R. Cipolla, October. Segmentation and recognition using structure from motion point clouds, In European conference on computer vision (2008) 44-57.

[7] G. Medioni and S. B. Kang, , "Emerging topics in computer vision", Prentice Hall PTR.

[8] D. W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, Journal of the society for Industrial and Applied Mathematics 11 (2) (1963) 431-441.

[9] D. Sibley, C. Mei, I. D. Reid and P. Newman, Adaptive relative bundle adjustment", In Robotics: science and systems 32 (2009) 33.

[10] M. Matsumoto and T. Nishimura, Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator, ACM Transactions on Modeling and Computer Simulation (TOMACS) 8 (1) (1998) 3-30.