Vis Comput (2018) 34:1443–1460
https://doi.org/10.1007/s00371-017-1451-0

ORIGINAL ARTICLE

3D reconstruction system based on incremental structure from motion using a camera with varying parameters

Soulaiman El Hazzat[1] · **Mostafa Merras**[1] · **Nabil El Akkad**[1,3] · **Abderrahim Saaidi**[1,2] · **Khalid Satori**[1]

Published online: 20 October 2017

Abstract In this paper, we present a flexible and fast system for multi-scale objects/scenes 3D reconstruction from uncalibrated images/video taken by a moving camera characterized by variable parameters. The proposed system is based on incremental structure from motion and good exploitation of bundle adjustment. At first, from two selected images, our system allows to recover, in a well-chosen reference, coordinates of a set of 3D points. In this context, we have proposed a new method of self-calibration based on the use of two unknown scene points with their image projections. After that, new images are inserted progressively using 3D information already obtained. Local bundle adjustment is used to adjust the new estimated entities. At some time, we introduce a global bundle adjustment to adjust as best as possible all estimated entities and to have an initial 3D model of quality covering an interesting part of the object/scene. This model will be used as reference for the insertion of the rest of images. The proposed system allows to obtain satisfactory results within a reasonable time.

Keywords 3D reconstruction · Self-calibration · Incremental structure from motion · Bundle adjustment

✉ Soulaiman El Hazzat
soulaiman.elhazzat@yahoo.fr;
soulaiman.elhazzat@usmba.ac.ma

Mostafa Merras
merras.mostafa@gmail.com

Nabil El Akkad
nabil.elakkad@usmba.ac.ma

Abderrahim Saaidi
abderrahim.saaidi@usmba.ac.ma

Khalid Satori
khalidsatorim3i@yahoo.fr

[1] LIIAN, Department of Mathematics and Informatics, Faculty of Sciences Dhar-Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco

[2] LSI, Department of Mathematics, Physics and Informatics, Polydisciplinary Faculty of Taza, Sidi Mohamed Ben Abdellah University, Taza, Morocco

[3] Department of Mathematics and Computer Science, National School of Applied Sciences (ENSA) of Al-Hoceima, Mohamed First University, Oujda, Morocco

1 Introduction

3D reconstruction from images/video is an important and widely studied subject in recent decades. It is to recover 3D information from 2D images taken from different viewpoints or from video.

Several approaches [1–15] have been proposed to solve this problem. There are approaches that are based on points matching between different images. Structure from motion approach [5,11,12,15] allows automatic recovery of 3D structure and camera motion. It is based on the detection and matching of interest points between different images. The matched points with other estimated geometric entities (epipolar geometry) will be used to recover a projective (uncalibrated images) and sparse representation of the scene. To move to a metric/Euclidean representation, a step of camera self-calibration to recover the intrinsic parameters is necessary. The reconstructed sparse 3D point cloud does not allow to properly define the shape of objects. So, to have dense results closer to the reality, dense matching methods should be used [16]. Delaunay triangulation, the Crust method [17] and the Poisson surface [18] are methods used to convert the obtained 3D point cloud into triangulated surface model. The approaches based on multi-view stereo [7–10] are often used to get high-quality dense 3D reconstruction results, but they require in input calibrated stereo images as

well as a long computation time. In robotics, simultaneous localization and mapping (SLAM) [19,20] simultaneously allows robot location and environment map construction using data retrieved from the sensors which may include cameras.

In this work, we propose a complete 3D reconstruction system from uncalibrated image/video sequences. Our 3D reconstruction system is able to produce very realistic three-dimensional models using a single camera. The camera intrinsic parameters are variable, the displacements of the camera are free, and the reconstruction environment is uncontrollable. All these factors offer more flexibility and generality for the 3D reconstruction of any objects/scenes (multi-scale objects/scenes). Our 3D reconstruction system is based on the incremental structure from motion. It is initialized from two images with a sufficient number of matches and a large camera motion [15]. In this context, we have proposed a new method for automatic recovery of intrinsic and extrinsic camera parameters that correspond to these two images. The 3D structure of the object/scene is initiated by the triangulation of matched/tracked interest points between these two images. The quality of initialization affects the whole system. Therefore, bundle adjustment is applied to adjust the estimated entities. The projection matrix of each new inserted image is estimated after locating projections of 3D reconstructed points in the inserted image. This estimate is based on the use of RANSAC algorithm [21] by solving a linear system using 3D points already reconstructed and their projections located in the inserted image. After, new 3D points are recovered from the interest point matching result between the inserted image and the image that precedes, and a local bundle adjustment is performed to adjust the new estimated entities. The local optimization does not guarantee the accuracy of the obtained solutions. So, in our system we integrate a global bundle adjustment after the insertion of M_0 images (in our experiments $10 \leq M_0 \leq 20$) to have an initial 3D model that will be used as a reference to insert the rest of the images in order to obtain a more complete final 3D model. To have a surface model, the Poisson surface algorithm [18] or the 3D Crust method [17] can be applied to the obtained 3D point cloud. Finally, the texture mapping provides realistic results.

The good exploitation of the existing (incremental structure from motion [12], bundle adjustment [5,12,22,23], …) and our own vision to solve the problem (camera self-calibration from only two images, the use of camera with varying intrinsic parameters, local and global vision of the problem, …) allow us to propose a system that is fully automatic, flexible and able to reconstruct multi-scale three-dimensional models (small, medium and large) within a reasonable calculation time. On the other hand, there are systems [5,24] that are based on the global bundle adjustment of all estimated entities, which require a fairly important computation time and demand a very important step of parameter initialization to avoid falling into local optima. Other systems [11,12] require prior information on camera parameters to make a 3D reconstruction of the scene.

This paper is organized as follows. Section 2 presents related work. Section 3 describes the notations and background. The proposed method is described in Sect. 4. The experiments and the comparison of our method with other methods are presented in Sect. 5. Finally, the conclusion is presented in Sect. 6.

2 Related work

Several methods have been proposed to solve the problem of 3D reconstruction from images/video. The methods based on multi-view stereo [8,10,25] allow to have satisfactory results with a high density. Tran and Davis [25] presented the graph cut method to recover the 3D object surface by the use of silhouettes and foreground color information. In [8], the authors presented a new method for large-scale multi-view stereo based on dense matching between very high-resolution images. It allows to obtain a 3D point cloud that is very dense and of high quality at a relatively low computational cost. However, it requires the use of rich texture images to avoid making use of costly optimization algorithms. Furukawa and Ponce [10] also proposed a method to solve the problem of multi-view stereo. It consists in retrieving an initial set of patches covering the surface of the object/scene from the matching result of key points detected by Harris and DoG operators. The final patches are obtained by iteration between an expansion step, to obtain a dense set of patches, and a filtering step based on the visibility constraint to eliminate false matches. Finally, the resulting patch model is converted into a polygonal mesh, which can be refined by applying the photometric consistency and regularization constraints. All these methods provide dense three-dimensional models of good quality, but they require stereo cameras of known parameters (methods that start from stereo calibrated images) and they are expensive in terms of computation time.

Structure from motion methods [5,11,12,26] allows to automatically recover both the three-dimensional structure and camera motion from uncalibrated image sequences. These methods are based on the detection and matching of interest points between different images. Pollefeys et al. [5] presented a complete system for three-dimensional modeling from uncalibrated images. First, structure from motion approach was used for the recovery of projective 3D structure and camera motion. Then, the different estimated entities will be refined by global bundle adjustment. To pass to a metric 3D reconstruction, they have gone through a camera self-calibration phase based on the use of the absolute conic. Finally, pairs of images are rectified and multi-view

stereo matching is used to obtain a dense 3D reconstruction. However, the global bundle adjustment requires a very long calculation time, especially with the use of a large number of images and can converge to a local solution due to a bad initialization. Snavely et al. [11,26] presented a system for 3D reconstruction from large photo collections. It is based on the incremental structure from motion approach to simultaneously recover the camera motion and a sparse 3D reconstruction of the scene. However, it requires prior information to initialize camera parameters (the use of EXIF tags) and requires a long calculation time, especially with the increase of images number, dominated by the bundle adjustment after the insertion of each new image. In [27], the authors presented a new incremental Structure from Motion technique based on geometric verification strategy, next best view selection and robust triangulation method. Fuhrmann et al. [13] presented a 3D reconstruction system of multi-scale scenes from images. It is based on structure from motion, multi-view stereo depth maps and surface reconstruction. Mouragnon et al. [12] proposed a method for real-time estimation of motion and 3D structure from video captured by a calibrated camera. The proposed method is based on the local bundle adjustment to refine the camera poses and 3D structure. However, this method requires the use of camera with known and unchanged intrinsic parameters during the acquisition of images. Thus, the quality of the 3D reconstruction is not assured because of the accumulation of errors when increasing the images number.

3 Notation and background

3.1 Pinhole camera model

The pinhole camera model is used to project a scene 3D point $A_j = (X_j, Y_j, Z_j, 1)^{\mathrm{T}}$ in the image point $a_{ij} = (u_{ij}, v_{ij}, 1)^{\mathrm{T}}$. This projection is represented by the following formula:

$$\lambda_{ij} a_{ij} = P_i A_j \tag{1}$$

where λ_{ij} is a nonzero scale factor, $P_i = K_i \left[R_i \; t_i \right]$ is a 3×4 projection matrix, t_i is a translation vector, R_i is 3×3 a rotation matrix defined by:

$$R_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha_i & -\sin\alpha_i \\ 0 & \sin\alpha_i & \cos\alpha_i \end{pmatrix} \begin{pmatrix} \cos\beta_i & 0 & \sin\beta_i \\ 0 & 1 & 0 \\ -\sin\beta_i & 0 & \cos\beta_i \end{pmatrix} \times \begin{pmatrix} \cos\gamma_i & -\sin\gamma_i & 0 \\ \sin\gamma_i & \cos\gamma_i & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

as α_i, β_i and γ_i represent the three Euler angles.

K_i is intrinsic parameter matrix defined by:

$$K_i = \begin{pmatrix} f_i & s_i & u_{0i} \\ 0 & \varepsilon_i f_i & v_{0i} \\ 0 & 0 & 1 \end{pmatrix}$$

where f_i is the focal length, ε_i is the scale factor, s_i is the skew factor and (u_{0i}, v_{0i}) are the coordinates of the principal point.

3.2 Estimation of distortion coefficients

The distortion effect affects the quality of the 3D reconstruction [28]. In this work, we consider the first two coefficients of radial distortion k_1 and k_2 in order to obtain more accurate results.
The relationship between the distorted image points (u_d, v_d) and the undistorted image points (u, v) is defined by [29]:

$$\begin{cases} u_d = u + (u - u_{0i}) \left(k_1 \left(x^2 + y^2 \right) + k_2 \left(x^2 + y^2 \right)^2 \right) \\ v_d = v + (v - v_{0i}) \left(k_1 \left(x^2 + y^2 \right) + k_2 \left(x^2 + y^2 \right)^2 \right) \end{cases}$$

where (u_{0i}, v_{0i}) are the coordinates of the principal point that correspond to the ith image and $(x, y, 1)^{\mathrm{T}} = K_i^{-1} (u, v, 1)^{\mathrm{T}}$. So, for each image point we have the following formula :

$$\begin{bmatrix} (u - u_0)\left(x^2 + y^2\right) & (u - u_0)\left(x^2 + y^2\right)^2 \\ (v - v_0)\left(x^2 + y^2\right) & (v - v_0)\left(x^2 + y^2\right)^2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} u_d - u \\ v_d - v \end{bmatrix}$$

3.3 Homography between two images

The homography between two images I_i and I_j is represented by a 3×3 matrix denoted H_{ij}. For each point a_{ik} of the image I_i and its corresponding a_{jk} in the image I_j, we have the following relationship:

$$a_{jk} \sim H_{ij} a_{ik}$$

Four non-aligned matches are sufficient for the estimation of this matrix. The use of the RANSAC algorithm [21] provides a reliable solution.

3.4 Selection of two images with a large displacement

In this work, we used the criteria already presented in [15] to select two images with a large displacement.

Let I_r be the reference image, and the disparity matrix is defined as follows:

$$D = \begin{bmatrix} \|a_{11} - a_{r1}\|_F & \cdots & \|a_{1n} - a_{rn}\|_F \\ \vdots & \ddots & \vdots \\ \|a_{m1} - a_{r1}\|_F & \cdots & \|a_{mn} - a_{rn}\|_F \end{bmatrix}$$

where m is the number of images, n is the number of matches, $\|a_{ij} - a_{rj}\|_F$ is the disparity between the two points and (a_{ij}, a_{rj}) is the jth matched point between I_r and I_i.

The image $I_{r'}$ that corresponds to a large camera motion relative to the reference image I_r is obtained by the use of the following formula:

$$r' = \max\left(\frac{D_M \odot D_S}{\|D_M\|_F \|D_S\|_F}\right)$$

where $D_M = \begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix}$ is a vector which represents the mean of each row of D; $D_S = \begin{bmatrix} s_1 \\ \vdots \\ s_m \end{bmatrix}$ is a vector which represents the standard deviation of each row of D; $\odot$ denotes the element-by-element multiplication.

4 Proposed method

We present an incremental 3D reconstruction system based on structure from motion approach and the good exploitation of bundle adjustment. It takes as input uncalibrated images/videos captured by a camera with variable parameters. As output, it determines the camera parameters (intrinsic and extrinsic) and the three-dimensional structure. Our system offers more flexibility to adapt and to reconstruct multi-scale objects/scenes (small, medium and large) in a reasonable time compared to methods applied in real time [12].

As already known, structure from motion approach using uncalibrated images allows to recover only a 3D projective reconstruction. To get a 3D metric/Euclidean reconstruction, it must pass through a camera self-calibration phase. In this work, we proposed a system that can directly retrieve the metric structure of the 3D scene. First, it is to initialize our system from two images with a sufficient number of matched interest points and a large movement of the camera [15]. In this context, we have proposed a new method of camera self-calibration from two images, which allows us retrieving coordinates of a set of 3D points in the scene corresponding to the matched image points. After each insertion of a new uncalibrated image, the camera parameters are retrieved based on the previously estimated 3D structure and new 3D points are recovered from interest point matching between the inserted image and the image that precedes. Our 3D reconstruction system is realized in three main steps: detection and matching/tracking of interest points between different images, initialization of the reconstruction system from two images with a large displacement [15] and incremental insertion of new images. The global algorithm of our 3D reconstruction system is presented as follows:

Algorithm 1 : Camera parameters and 3D structure Recovery

Input : m number of images, $m_0 = 3$ and M_0 integer selected between 10 and 20
Output : camera parameters (intrinsic and extrinsic) and the 3D structure

Begin

1. Interest point detection and matching/tracking
2. Initialization from two images
   - 2.1. Searching two images $\{I_1, I_2\}$ with a large movement and a sufficient number of matched points
   - 2.2. Definition of scene global reference and estimation of camera parameters that correspond to these two images
   - 2.3. Recovery of 3D point coordinates from the camera parameters and the interest point matching result between these two images
   - 2.4. Bundle adjustment taking into account the radial distortion
3. Incremental insertion of new images $\{I_k\}_{3 \le k \le m}$
   - 3.1. **For** $k = 3$ **to** m
     - Projections' localization of already reconstructed 3D points in the image I_k.
     - Estimation of projection matrix P_k using RANSAC Algorithm.
     - Projection matrix decomposition to obtain the intrinsic and extrinsic parameters.
     - Recovering new 3D points from interest point matching result between the images I_{k-1} and I_k.
     - Local bundle adjustment between the last m_0 images taking into account the radial distortion.
     - **If** $k == M_0$ **Then**
       Global bundle adjustment between the M_0 images.
       End If

     End For
   - 3.2. **If** $m < M_0$ **Then**
     Global bundle adjustment between the m images.

     End If

End

Algorithm 1 allows, from m ($m \geq 2$) input images, to gradually recover the metric 3D structure and the camera parameters (intrinsic and extrinsic) that corresponds to each image. The value of m_0 is initialized to 3 because the local bundle adjustment is applied to the three last images. This allows to provide a reliable initial solution for the global bundle adjustment (GBA) which will be applied after the insertion of M_0 images to have an initial 3D model of quality. The value of M_0 is selected between 10 and 20 because the application of the GBA on a large number of images requires much calculation time. Also, the use of a small number of images does not allow to have a reliable initial 3D model. When inserting the remain images $\{I_k\}_{M_0 < k \le m}$, the value of m_0 can be taken greater than 3 to increase the system reliability. The initialization of our system from two images is a very interesting phase. So, to ensure the stability and reliability, we have selected two images with a sufficient number of matched points and a large movement of the camera [15]. The selected images allow to ensure the stability of epipolar geometry calculation (estimation of the fundamental matrix). This matrix will be used later to estimate the camera parameters.

4.1 Interest point detection and matching/tracking

In this work, we have chosen to use the SIFT algorithm [30] for interest point matching between different images because of its robustness to scale changes compared to other methods [31]. For the elimination of false matches and the estimation of fundamental matrix, the RANSAC algorithm was used [21].

4.2 Initialization of 3D reconstruction system from two images

The initialization of our 3D reconstruction system is performed from two selected images with a sufficient number of matched points and a large displacement of the camera [15] in order to stabilize the calculations. It consists in:

1. Camera self-calibration: estimation of intrinsic parameters from two images.
2. Estimation of extrinsic parameters.
3. Retrieving a set of 3D points from matched interest points.
4. Bundle adjustment taking into account the radial distortion.

4.2.1 Self-calibration

In this step, we present a new formulation of the self-calibration problem based on the good choice of global reference and the use of planar calibration/self-calibration concepts [29]. This formulation allows to obtain a linear system which leads, assuming that the principal point is in the center of the image and the skew factor is equal to zero, to determine the scale factor and the focal length. Thus, this formulation allows us to automatically estimate the camera extrinsic parameters.

Let A_1 and A_2 two unknown points of the 3D scene as a_{i1} and a_{i2} are, respectively, their projections in the image I_i with $1 \leq i \leq 2$. We define an Euclidean reference (O, X, Y, Z) as O is the midpoint of segment $[A_1A_2]$, and the two points A_1 and A_2 belong to the plane Z = 0 (plane OXY) (see Fig. 1).

In this reference:

$$A_1 = (d\cos\theta, d\sin\theta, 0, 1)^{\mathrm{T}}$$
$$A_2 = (-d\cos\theta, -d\sin\theta, 0, 1)^{\mathrm{T}}$$

where $d = A_1A_2/2$ and θ is the angle between the line (A_1A_2) and the X-axis of the reference.

To simplify the calculations, we can choose the global reference such as $\theta = \pi/3$ (other values can be selected). So, we obtain:

$$A_1 = \left(d/2, \sqrt{3}d/2, 0, 1\right)^{\mathrm{T}} \quad \text{and}$$
$$A_2 = \left(-d/2, -\sqrt{3}d/2, 0, 1\right)^{\mathrm{T}}$$

We consider the Euclidean reference (O, X, Y).

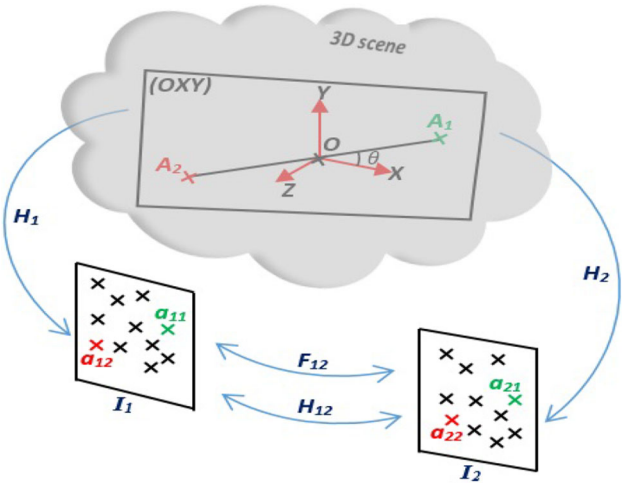

Fig. 1 Different entities used for automatic estimation of camera parameters

In this reference:

$$A_1 = \left(d/2, \sqrt{3}d/2, 1\right)^{\mathrm{T}} \tag{2}$$
$$A_2 = \left(-d/2, -\sqrt{3}d/2, 1\right)^{\mathrm{T}} \tag{3}$$

The projection of the plane $Z = 0$ in the image plane I_i is defined by the homography H_i as:

$$H_i \sim K_i R_i \begin{pmatrix} 1 & 0 & \\ 0 & 1 & R_i^{\mathrm{T}} t_i \\ 0 & 0 & \end{pmatrix}, \quad i = 1, 2 \tag{4}$$

As $A_1 \in (O, X, Y)$ and $A_2 \in (O, X, Y)$.

So, we can write

$$a_{ij} \sim H_i A_j, \quad i = 1, 2 \quad \text{and} \quad j = 1, 2 \tag{5}$$

Expressions (2) and (3) can be written in the form:

$$A_1 = \begin{pmatrix} d & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \\ 1 \end{pmatrix} \tag{6}$$

$$A_2 = \begin{pmatrix} d & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1/2 \\ -\sqrt{3}/2 \\ 1 \end{pmatrix} \tag{7}$$

We put:

$$B = \begin{pmatrix} d & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad A_1' = \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \\ 1 \end{pmatrix} \quad \text{and}$$
$$A_2' = \begin{pmatrix} -1/2 \\ -\sqrt{3}/2 \\ 1 \end{pmatrix}$$

Expression (5) can be expressed as follows :

$$a_{ij} \sim H_i B A'_j, \quad i = 1,2 \quad \text{and} \quad j = 1,2 \tag{8}$$

We put:

$$Q_i = \begin{pmatrix} Q_{i00} & Q_{i01} & Q_{i02} \\ Q_{i10} & Q_{i11} & Q_{i12} \\ Q_{i20} & Q_{i21} & Q_{i22} \end{pmatrix} = H_i B, \quad i = 1,2 \tag{9}$$

Expression (8) can be written as follows:

$$a_{ij} \sim Q_i A'_j, \quad i = 1,2 \quad \text{and} \quad j = 1,2 \tag{10}$$

Q_i is a matrix 3×3 that allows to project the point A'_j in the image I_i.

At first, we want to recover the projection matrices $\{Q_i\}_{1 \le i \le 2}$.

From (9), we can write:

$$B = H_1^{-1} Q_1 \quad \text{and} \quad Q_2 = H_2 H_1^{-1} Q_1$$
$$Q_2 = H_{12} Q_1 \tag{11}$$

H_{12} is the homography between the images I_1 and I_2.

In (10), we replace Q_2 by its formula (11) and we obtain:

$$a_{2j} \sim H_{12} Q_1 A'_j, \quad j = 1,2 \tag{12}$$

From [32], we know that:

$$F_{12} \sim [e_2]_\times H_{12} \tag{13}$$

where $[e_2]_\times = \begin{pmatrix} 0 & -e_{2_z} & e_{2_y} \\ e_{2_z} & 0 & -e_{2_x} \\ -e_{2_y} & e_{2_x} & 0 \end{pmatrix}$ is the antisymmetric matrix associated with the epipole of the image I_2 $e_2 = \left(e_{2_x}, e_{2_y}, e_{2_z}\right)^{\mathrm{T}}$.

Then, formulas (12) and (13) give:

$$[e_2]_\times a_{2j} \sim F_{12} Q_1 A'_j, \quad j = 1,2 \tag{14}$$

From formulas (10) and (14), a linear system of eight linear equations (for the elements of Q_1) is obtained. The resolution of this system allows the estimation of matrix Q_1.

Expressions (11) and (13) give:

$$[e_2]_\times Q_2 \sim F_{12} Q_1 \tag{15}$$

The matrix Q_2 is estimated from Eq. (15).

Now, the estimated projection matrices will be used to recover the intrinsic parameters. Formulas (4) and (9) give:

$$K^{-1} Q_i \sim R_i \begin{pmatrix} 1 & 0 & \\ 0 & 1 & R_i^{\mathrm{T}} t_i \\ 0 & 0 & \end{pmatrix} B \tag{16}$$

The previous formula gives:

$$Q_i^{\mathrm{T}} \omega_i Q_i \sim \begin{pmatrix} d & 0 & 0 \\ 0 & d & 0 \\ & t_i^{\mathrm{T}} R_i & \end{pmatrix} \begin{pmatrix} d & 0 & \\ 0 & d & R_i^{\mathrm{T}} t_i \\ 0 & 0 & \end{pmatrix} \tag{17}$$

where $\omega_i = \begin{pmatrix} \omega_{i00} & \omega_{i01} & \omega_{i02} \\ \omega_{i10} & \omega_{i11} & \omega_{i12} \\ \omega_{i20} & \omega_{i21} & \omega_{i22} \end{pmatrix} = \left(K_i K_i^{\mathrm{T}}\right)^{-1}$ is the image of the absolute conic.

$$\omega_{i00} = \frac{1}{f_i^2}, \quad \omega_{i01} = \omega_{i10} = -\frac{s_i}{\varepsilon_i f_i^3},$$
$$\omega_{i02} = \omega_{i20} = \frac{u_{0i} s_i - \varepsilon_i u_{0i} f_i}{\varepsilon_i f_i^3},$$
$$\omega_{i11} = \frac{s_i^2}{\varepsilon_i^2 f_i^4} + \frac{1}{\varepsilon_i^2 f_i^2},$$
$$\omega_{i12} = \omega_{i21} = -\frac{s_i \left(v_{0i} s_i - u_{0i} \varepsilon_i f_i\right)}{\varepsilon_i^2 f_i^4} - \frac{v_{0i}}{\varepsilon_i^2 f_i^2},$$
$$\omega_{i22} = \frac{\left(v_{0i} s_i - u_{0i} \varepsilon_i f_i\right)^2}{\varepsilon_i^2 f_i^4} + \frac{v_{0i}^2}{\varepsilon_i^2 f_i^2} + 1.$$

We put $B' = \begin{pmatrix} d & 0 \\ 0 & d \\ 0 & 0 \end{pmatrix}$.

Formula (17) gives:

$$Q_i^{\mathrm{T}} \omega_i Q_i \sim \begin{pmatrix} B'^T B' & B'^T R_i^{\mathrm{T}} t_i \\ t_i^{\mathrm{T}} R_i B' & t_i^{\mathrm{T}} t_i \end{pmatrix} \tag{18}$$
$$B'^T B' = \begin{pmatrix} d^2 & 0 \\ 0 & d^2 \end{pmatrix}$$

Formula (18) gives:

$$\begin{pmatrix} \left(Q_i^{\mathrm{T}} \omega_i Q_i\right)_{00} & \left(Q_i^{\mathrm{T}} \omega_i Q_i\right)_{01} \\ \left(Q_i^{\mathrm{T}} \omega_i Q_i\right)_{10} & \left(Q_i^{\mathrm{T}} \omega_i Q_i\right)_{11} \end{pmatrix} \sim \begin{pmatrix} d^2 & 0 \\ 0 & d^2 \end{pmatrix}, \quad i = 1,2. \tag{19}$$

From (19), the following equation system is obtained:

$$\begin{cases} \left(Q_i^{\mathrm{T}} \omega_i Q_i\right)_{00} = \left(Q_i^{\mathrm{T}} \omega_i Q_i\right)_{11} \\ \left(Q_i^{\mathrm{T}} \omega_i Q_i\right)_{01} = \left(Q_i^{\mathrm{T}} \omega_i Q_i\right)_{10} = 0 \end{cases} \tag{20}$$

Assuming that the principal point (u_{0i}, v_{0i}) is in the center of the image and $s_i = 0$, ε_i and f_i will be determined.

Expression (20) gives:

$$\begin{cases} \alpha_1 \varepsilon_i^2 f_i^2 + \alpha_2 \varepsilon_i^2 + \alpha_3 = 0 \\ \beta_1 \varepsilon_i^2 f_i^2 + \beta_2 \varepsilon_i^2 + \beta_3 = 0 \end{cases} \tag{21}$$

where:

$$\begin{aligned}
\alpha_1 &= Q_{i20}^2 - Q_{i21}^2,\\
\alpha_2 &= Q_{i00}^2 - Q_{i01}^2 - 2\left(Q_{i00}Q_{i20} + Q_{i01}Q_{i21}\right)u_{0i}\\
&\quad + \left(Q_{i20}^2 - Q_{i21}^2\right)u_{0i}^2,\\
\alpha_3 &= Q_{i10}^2 - Q_{i11}^2 - 2(Q_{i10}Q_{i20} + Q_{i11}Q_{i21})v_{0i}\\
&\quad + \left(Q_{i20}^2 - Q_{i21}^2\right)v_{0i}^2,\\
\beta_1 &= Q_{i20}Q_{i21},\\
\beta_2 &= Q_{i00}Q_{i01} - Q_{i01}Q_{i20}u_{0i} - Q_{i00}Q_{i21}u_{0i}\\
&\quad + Q_{i20}Q_{i21}u_{0i}^2 \text{ and}\\
\beta_3 &= Q_{i10}Q_{i11} - Q_{i20}Q_{i11}v_{0i} - Q_{i10}Q_{i21}v_{0i}\\
&\quad + Q_{i20}Q_{i21}v_{0i}^2.
\end{aligned}$$

Expression (21) is a linear system of the form:

$$\begin{cases} a_0X_1 + a_1X_2 = b_0\\ a_2X_1 + a_3X_2 = b_1 \end{cases}$$

where $X_1 = \varepsilon_i^2 f_i^2$ and $X_2 = \varepsilon_i^2$.

Solving this linear system by substitution allows to estimate X_1 and X_2.

The values of ε_i and f_i are obtained from X_1 and X_2 (the two positive values).

4.2.2 Estimation of extrinsic parameters

Formula (4) gives:

$$H_i \sim K_i \left[\, r_1^i \; r_2^i \; t_i \,\right], \quad i = 1, 2 \tag{22}$$

where r_k^i $(1 \le k \le 3)$ denotes the kth column of the rotation matrix R_i.

As already presented in [29], the formula (22) gives:

$$r_1^i = \mu_i K_i^{-1} h_1^i \tag{23}$$
$$r_2^i = \mu_i K_i^{-1} h_2^i \tag{24}$$
$$t_i = \mu_i K_i^{-1} h_3^i \tag{25}$$

where $\mu_i = \|K_i^{-1}h_1^i\|^{-1} = \|K_i^{-1}h_2^i\|^{-1}$.

In our situation, the homography $H_i = \left[\, h_1^i \; h_2^i \; h_3^i \,\right]$ is unknown because the scene is not planar.

Formula (9) gives:

$$H_i = Q_i B^{-1} \tag{26}$$

Expressions (23), (24), (25) and (26) give:

$$r_1^i = \mu_i' K_i^{-1} q_1^i \tag{27}$$
$$r_2^i = \mu_i' K_i^{-1} q_2^i \tag{28}$$
$$r_3^i = r_1^i \times r_2^i \tag{29}$$
$$t_i = d\mu_i' K_i^{-1} q_3^i \tag{30}$$

where $\mu_i' = \|K_i^{-1}q_1^i\|^{-1} = \|K_i^{-1}q_2^i\|^{-1}$ and q_k^i $(1 \le k \le 3)$ denote the kth column of the matrix Q_i.

The rotation matrix is obtained from (27), (28) and (29).

It remains to estimate the value of d to obtain the translation vector.

Expression (17) gives:

$$\begin{cases} \left(Q_i^{\mathrm{T}}\omega_i Q_i\right)_{00} = \nu_i d^2\\ \left(Q_i^{\mathrm{T}}\omega_i Q_i\right)_{02} = \nu_i d t_i^{\mathrm{T}} r_1^i \end{cases} \tag{31}$$

where ν_i is a nonzero scale factor.

So, from (31) we obtain.

$$d = \frac{\left(Q_i^{\mathrm{T}}\omega_i Q_i\right)_{00}}{\left(Q_i^{\mathrm{T}}\omega_i Q_i\right)_{02}} \left(t_i^{\mathrm{T}} r_1^i\right) \tag{32}$$

Then, substituting in (30) d by its formula (32), we obtain a linear system of the form:

$$At_i = 0 \tag{33}$$

where $A \in \mathbb{R}^{3\times 3}$.

The resolution of this system by the singular value decomposition (SVD) allows to estimate the translation vector.

4.2.3 Recovering 3D point coordinates

The coordinates of 3D points are recovered from the matching result between the two images $\{I_1, I_2\}$ and the projection matrices defined by:

$$P_1 = K_1\left[R_1 t_1\right] \quad \text{and} \quad P_2 = K_2\left[R_2 t_2\right]$$

4.2.4 Bundle adjustment

The optimization of different entities previously estimated (intrinsic and extrinsic parameters, radial distortion and the 3D point coordinates) is performed by minimizing the criterion (34) using the Levenberg–Marquardt algorithm [23,33]:

$$C(\theta) = \sum_{i=1}^{2}\sum_{j=1}^{n_{1,2}} \|a_{ij} - \mathcal{P}(K_i, k_{1i}, k_{2i}, R_i, t_i, A_j)\|^2 \tag{34}$$

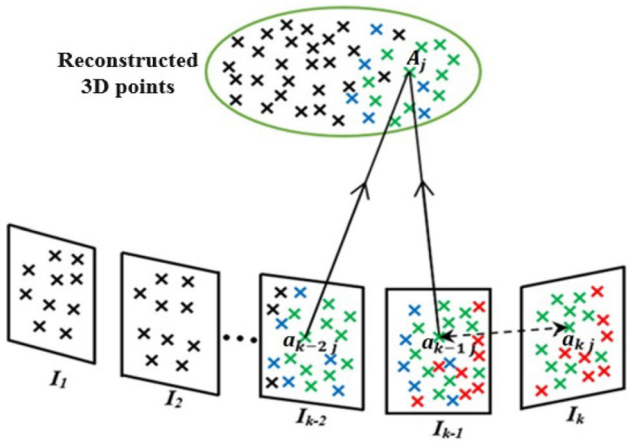

Fig. 2 Projections' localization of already reconstructed 3D points in the inserted image I_k using the interest point matching results. a_{kj} is the projection of the 3D point A_j, reconstructed from a_{k-2j} and a_{k-1j}, localized in the image I_k

where $n_{1,2}$ is the number of reconstructed 3D points and

$$\theta = \Big\{ f_1, \varepsilon_1, s_1, u_{01}, v_{01}, k_{11}, k_{21}, \alpha_1, \beta_1, \gamma_1, t_x^1, t_y^1, t_z^1, f_2, \varepsilon_2, s_2, u_{02}, v_{02}, k_{12}, k_{22}, \alpha_2, \beta_2, \gamma_2, t_x^2, t_y^2, t_z^2, X_1, Y_1, Z_1, \ldots, X_{n_{1,2}}, Y_{n_{1,2}}, Z_{n_{1,2}} \Big\}$$

4.3 Inserting a new image $I_k (3 \leq k \leq m)$

After inserting a new uncalibrated image, suitable projection matrix is estimated on the basis of the three-dimensional data already retrieved. The RANSAC algorithm [21] was used for the reliable recovery of this matrix by solving a linear system using previously reconstructed 3D points and their projections located in the inserted image [1] (Fig. 2). Then, new 3D points are retrieved from the interest point matching result between the inserted image and the previous image. So, the following steps are performed:

1. Projections' localization of already reconstructed 3D points in the image I_k.
2. Estimating the projection matrix P_k from n_0 ($n_0 \geq 6$) 3D points and their projections located in the image I_k by the use of RANSAC method [21]
3. Recovery of a set of 3D points from the interest point matching result between I_{k-1} and I_k.
4. Decomposition of the projection matrix P_k for the recovery of intrinsic and extrinsic parameters.
5. Local bundle Adjustment between the last m_0 images (in our experiments $m_0 = 3$) taking into account the radial distortion.
6. If $k = M_0$ ($10 \leq M_0 \leq 20$ for our experience) applied a global bundle adjustment between the M_0 inserted images.

4.3.1 *Projection matrix estimation and new 3D point recovery*

After the projections' localization of already reconstructed 3D points in the image I_k, P_k projection matrix is estimated from at least six already reconstructed 3D points and their projections localized in the image I_k [1] using RANSAC algorithm [21]. Then, the coordinates of new 3D points are estimated by triangulation from the interest point matching result between the images $\{I_{k-1}, I_k\}$ and the estimated projection matrices P_{k-1} and P_k.

Algorithm 2 : Estimating the projection matrix P_k

Input : Interest point matching/tracking results
S_{k-1} set of already reconstructed 3D points
Output : Projection matrix P_k

1) Find the set of interest points B_{k-1} in the image I_{k-1} matched at the same time with interest points in the images $\{I_{k-2}, I_k\}$. We denote by E_k the set of interest points in I_k that corresponds to the points of B_{k-1}
2) Find the set of already reconstructed 3D points S'_{k-1} which correspond to points of the set B_{k-1}. The points of S'_{k-1} also corresponds to points of E_k
3) Use of RANSAC algorithm for the estimation of the projection matrix P_k from the 3D points of the set S'_{k-1} and their projections localized in the image I_k (the points of the set E_k).

Algorithm 2 is based on the manipulation of the 3D information already estimated as well as on the interest point matching between the last 3 images $\{I_{k-2}, I_{k-1}, I_k\}$ to estimate the projection matrix P_k that corresponds to the inserted image I_k.

4.3.2 *Local bundle adjustment between the $m_0 = 3$ latest images*

The new estimated elements: the camera parameters that correspond to the image I_k, the radial distortion coefficients and the new reconstructed 3D points, are optimized by minimizing the criterion (35) [23,33].

$$C(\theta) = \sum_{i=k-m_0}^{k} \sum_{j=1}^{n_{k-1,k}} \| a_{ij} - \mathcal{P}\left(K_i, k_{1i}, k_{2i}, R_i, t_i, A_j\right) \|^2 \tag{35}$$

where $n_{k-1,k}$ is the number of new reconstructed 3D points from the interest point matching result between the couple image $\{I_{k-1}, I_k\}$ and

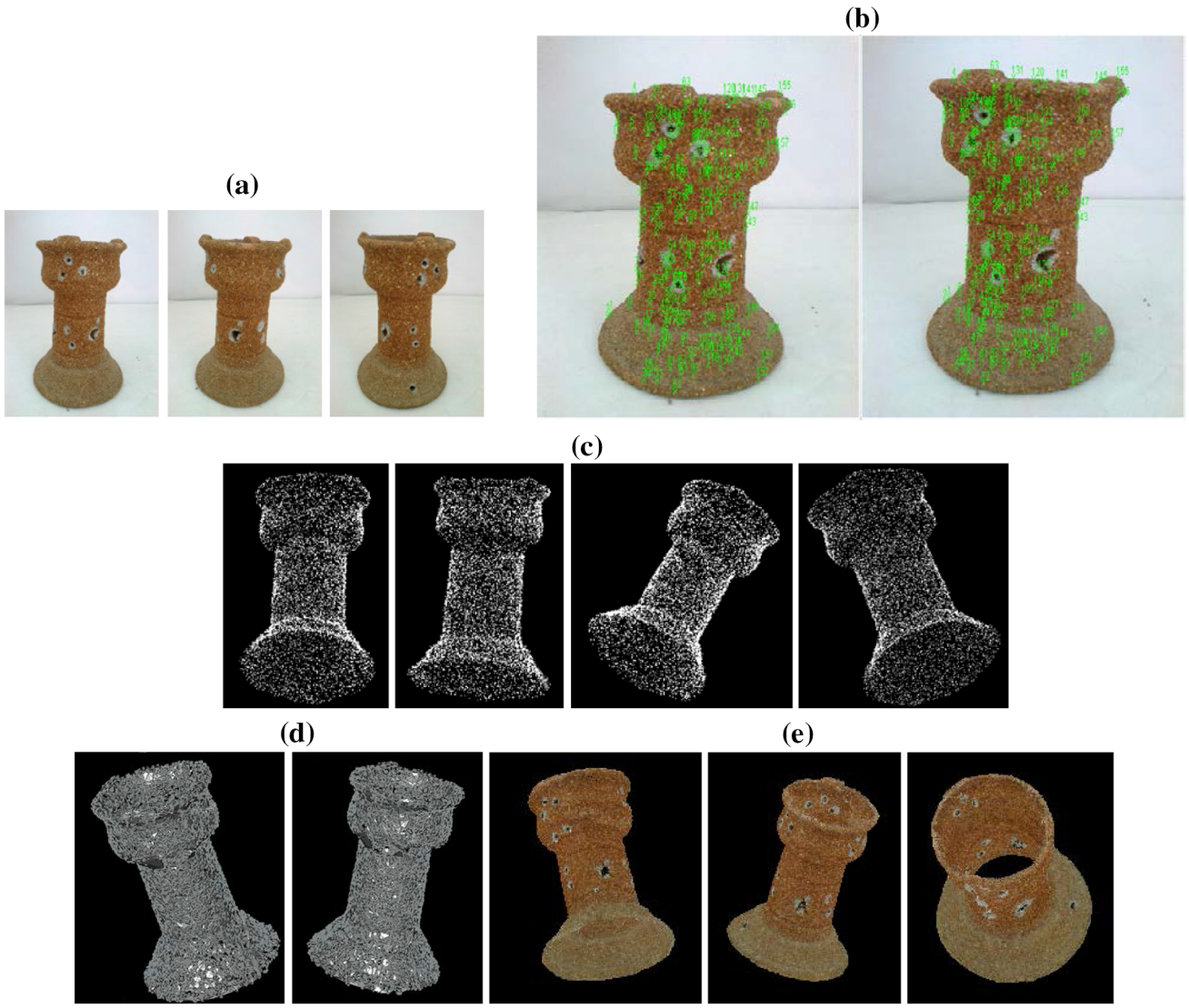

Fig. 3 **a** Three images of the sequence, **b** result of interest point matching between two images after removing false matches by RANSAC algorithm, **c** four views of the sparse 3D reconstruction, **d** two views of 3D surface model achieved using 3D Crust algorithm, **e** three views of textured 3D model

Table 1 Estimated camera intrinsic parameters that correspond to the first two images for the five sequences

Sequences	Images	f	ε	s	u_0	v_0	k_1	k_2
Vase	Image 1	1129	0.94	0.03	452	605	−0.053	0.023
	Image 2	1137	0.97	0.02	449	602	−0.045	0.041
Villa pot	Image 1	873	0.95	0.05	371	503	0.019	−0.121
	Image 2	889	0.93	0.04	367	501	−0.03	0.021
Medusa head	Image 1	1066	0.92	0.03	382	278	−0.34	0.028
	Image 2	1081	0.95	0.04	384	281	−0.12	0.03
Castle-P30	Image 1	910	0.96	0.02	1530	1019	−0.086	0.018
	Image 2	921	0.95	0.04	1541	1028	−0.099	0.023
Complex scene	Image 1	994	0.97	0.05	366	279	−0.105	0.035
	Image 2	1005	0.95	0.04	375	285	0.067	0.027

Table 2 The number of reconstructed 3D points (sparse 3D reconstruction) for the five sequences

Sequences	# Images	Resolution	M_0	# Reconstructed 3D points
Vase	32	900 × 1200	14	8547
Villa pot	28	750 × 1000	13	6435
Medusa head	26	765 × 560	12	7342
Castle-P30	30	1020 × 680	13	12,654
Complex scene	142	740 × 565	20	35,057

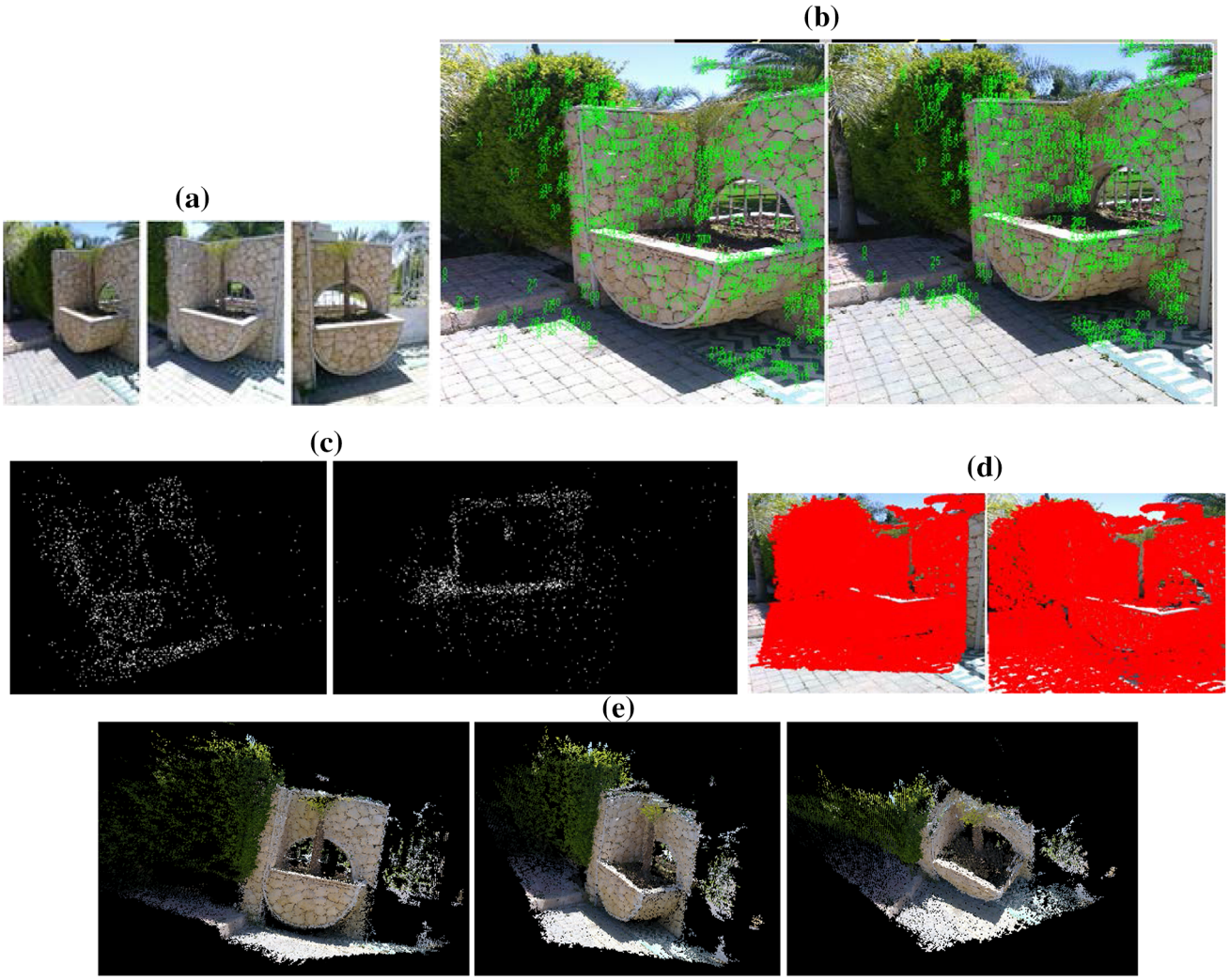

Fig. 4 **a** Three images of the sequence, **b** result of interest point matching between two images after removing false matches, **c** Two views of sparse 3D reconstruction, **d** matching result after the application of the match propagation algorithm, **e** three views of dense 3D reconstruction

$$\theta = \Big\{ f_k, \varepsilon_k, s_k, u_{0k}, v_{0k}, k_{1k}, k_{2k}, \alpha_k, \beta_k, \gamma_k, t_x^k, t_y^k, t_z^k, X_1, Y_1, Z_1, \ldots, X_{n_{k-1,k}}, Y_{n_{k-1,k}}, Z_{n_{k-1,k}} \Big\}$$

4.3.3 *Global bundle adjustment between the M_0 inserted images*

Global bundle adjustment applied to all images requires a very long calculation time, especially with the use of a large number of images. In this work, we have combined between the local bundle adjustment after the insertion of a new image and the global bundle adjustment after the insertion of M_0 images ($10 \leq M_0 \leq 20$) to accelerate the treatment maintaining system reliability. So, after the insertion of M_0 images, all estimated elements (already locally optimized) will be used as an initial solution to minimize the criterion (36) using the Levenberg–Marquardt algorithm [23,33].

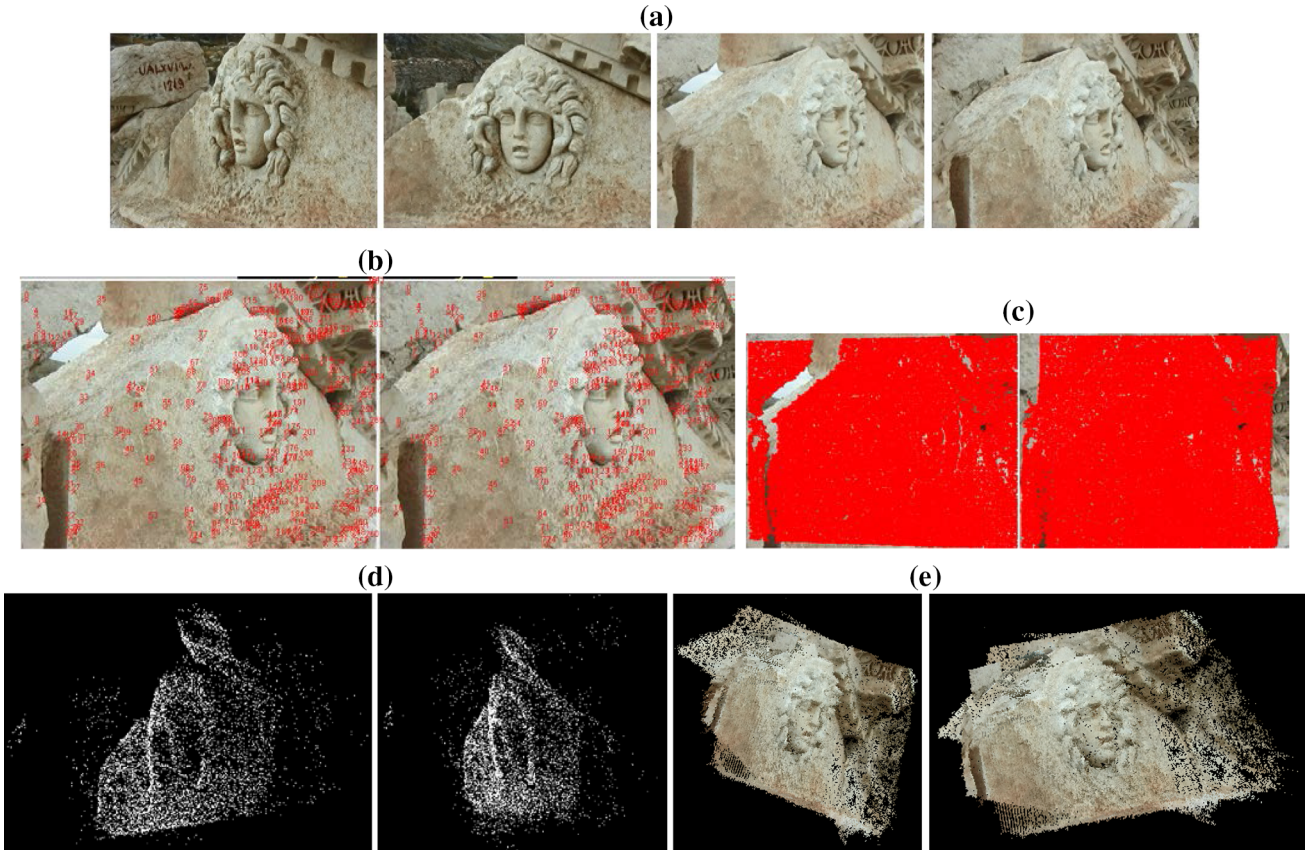

Fig. 5 **a** Four key frames, **b** example of interest point matching between two images, **c** matching result after the application of the match propagation algorithm, **d** two views of sparse 3D reconstruction, **e** two views of dense 3D reconstruction

$$C(\theta) = \sum_{i=1}^{M_0} \sum_{j=1}^{n} \| a_{ij} - \mathcal{P}\left(K_i, k_{1i}, k_{2i}, R_i, t_i, A_j\right) \|^2 \quad (36)$$

where n is the number of reconstructed 3D points and

$$\theta = \Big\{ f_1, \varepsilon_1, s_1, u_{01}, v_{01}, k_{11}, k_{21}, \alpha_1, \beta_1, \gamma_1, t_x^1, t_y^1, t_z^1, \ldots, f_{M_0}, \varepsilon_{M_0}, s_{M_0}, u_{0M_0}, v_{0M_0}, k_{1M_0}, k_{2M_0}, \alpha_{M_0}, \beta_{M_0}, \gamma_{M_0}, t_x^{M_0}, t_y^{M_0}, t_z^{M_0}, X_1, Y_1, Z_1, \ldots, X_n, Y_n, Z_n \Big\}$$

5 Experiments

To validate and test the robustness of the proposed approach, several images and video sequences are used. We present the results for five images/video sequences of scenes with different sizes (*vase, villa pot, Medusa head* [34], *castle-P30* [35] *and complex scene*). All experiments are executed on a machine HP 650 Intel Core i3, 2.30 GHz CPU and 4 GB RAM.

5.1 Vase sequence: small scale object

This first sequence consists of thirty-two images, with a resolution of 900 × 1200, taken around a small object. Three images of the sequence are presented in Fig. 3a. First, we begin by the detection and matching of interest points with SIFT method [30]. An example of interest point matching between two images is shown in Fig. 3b. Our reconstruction system is initialized from two images with a sufficient number of matches and a large camera motion [15]. After the estimation of camera parameters using our method, a set of 3D points is recovered from the result of interest point matching between these two images. A bundle adjustment, taking into account the radial distortion and using Levenberg–Marquardt algorithm [23,33], is applied to adjust as best as possible the estimated entities. Table 1 shows the estimated values of the camera intrinsic parameters and the first two radial distortion coefficients corresponding to the two first images for the five sequences. For each new inserted image, new 3D points are recovered and a local bundle adjustment between the last $m_0 = 3$ images is performed to adjust the new estimated entities. After the insertion of $M_0 = 14$ images, a global bundle adjustment is executed to adjust all parameters and to obtain a reliable initial 3D model that will be used with the local bundle adjustment for the insertion of new images. Four views of the sparse 3D reconstruction are shown in Fig. 3c. Figure 3d shows the obtained 3D surface model after applying 3D Crust algorithm [17]. The textured

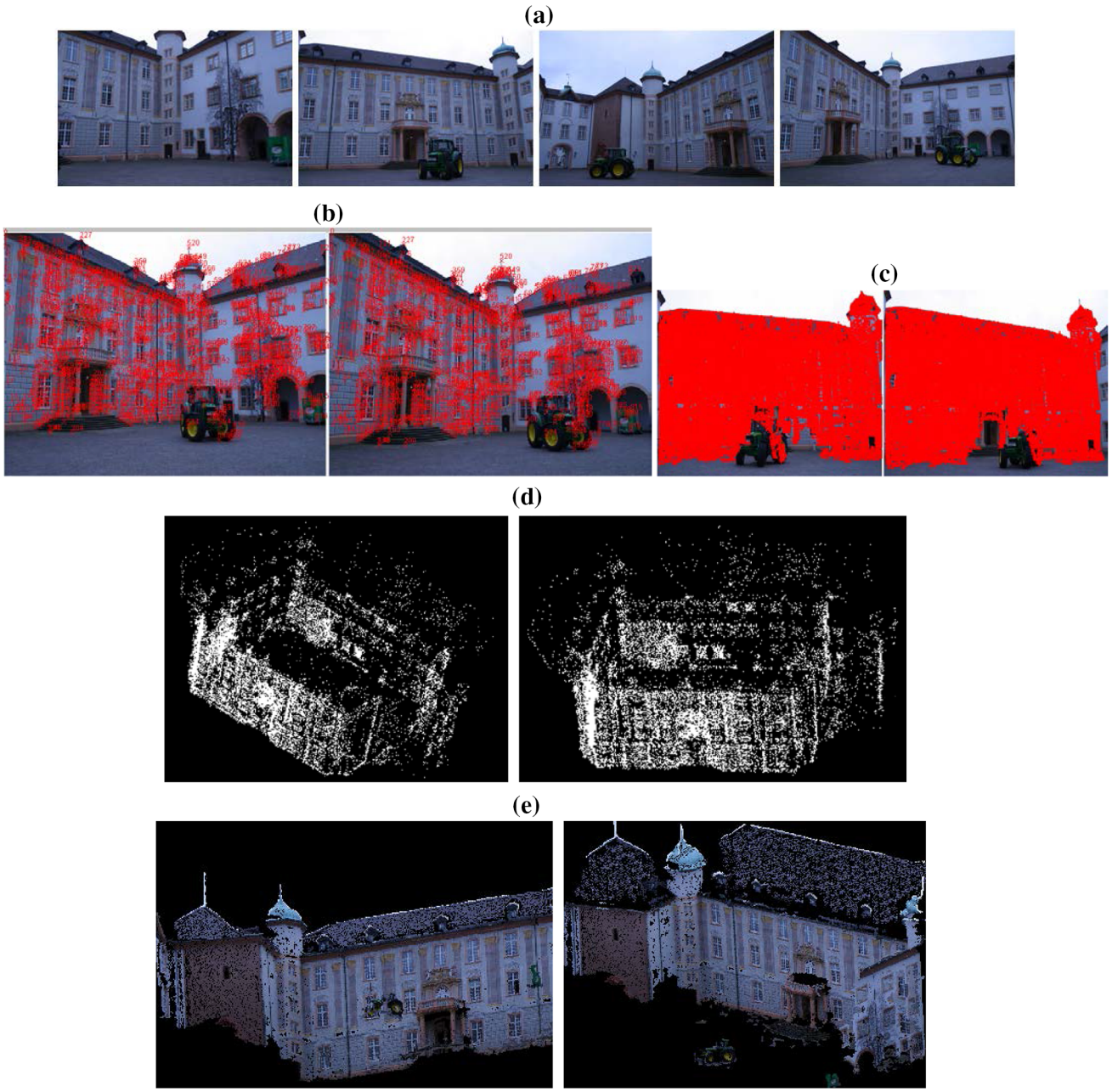

Fig. 6 **a** Four images of the sequence, **b** result of interest point matching between two images, **c** matching result after the application of the match propagation algorithm, **d** two views of the sparse 3D reconstruction, **e** two views of the dense 3D reconstruction

3D model is presented in Fig. 3e. Table 2 shows the values of M_0 and the number of reconstructed 3D points (sparse 3D reconstruction) for the five sequences.

5.2 Villa pot sequence: medium-scale scene

In this second experiment, a sequence of 28 images, with a resolution of 750 × 1000, was used. Three images of the sequence are shown in Fig. 4a. Figure 4b shows the result of interest point matching between two images after removing false matches using RANSAC algorithm [21] (353 matches are obtained). Two views of the sparse 3D reconstruction are presented in Fig. 4c. Figure 4d shows the almost dense matching result after applying the Match Propagation algorithm [36] (297694 matches are obtained). Three views of the dense 3D reconstruction are presented in Fig. 4e.

5.3 Medusa head sequence: medium-scale object

In this third experiment, we tested the power of our approach on the 'Medusa head' video downloaded from the Marc Pollefeys page [34]. Four key frames are shown in Fig. 5a.

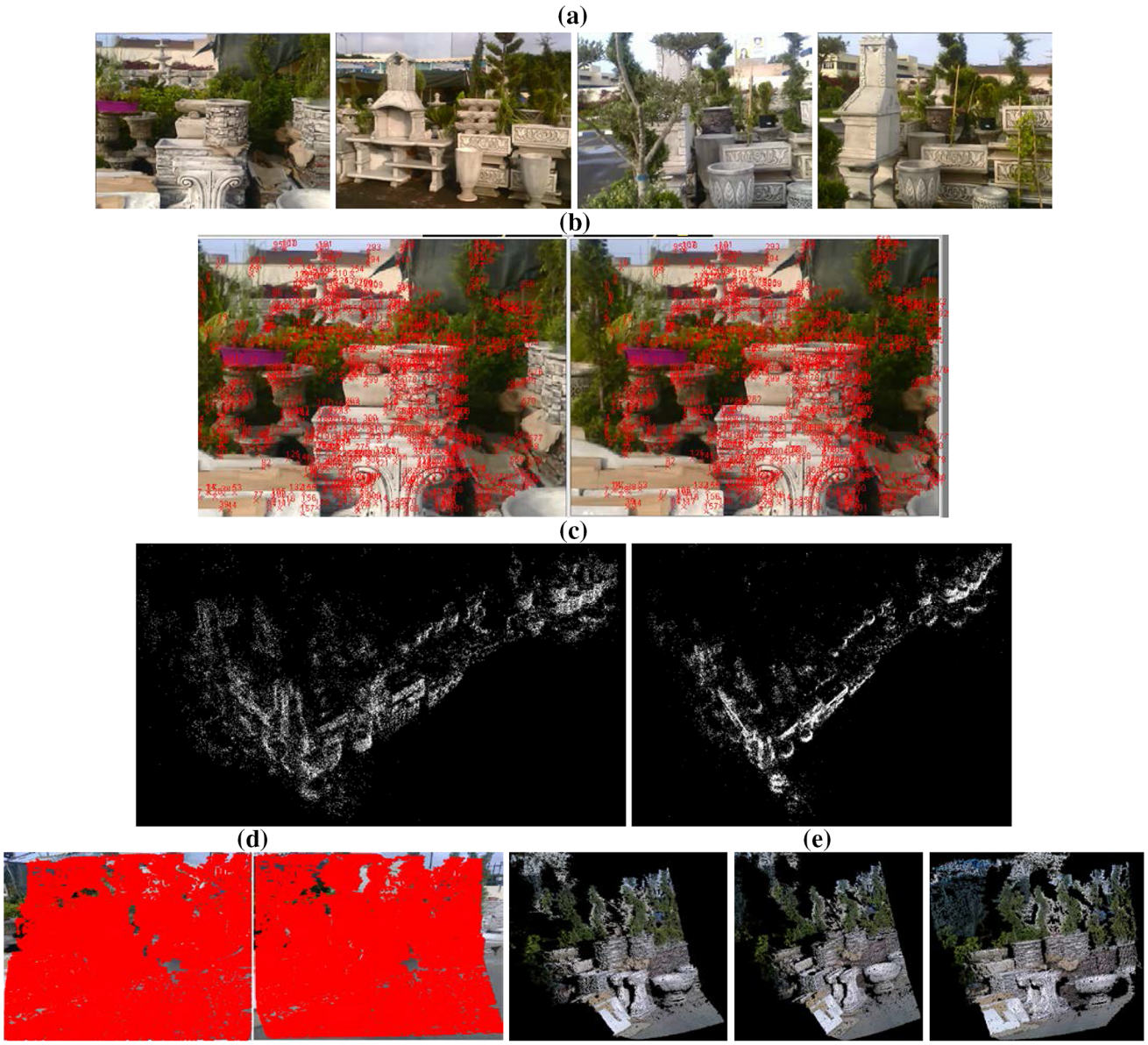

Fig. 7 **a** Four images of the sequence, **b** interest point matching between two images after removing false matches, **c** two views of the sparse 3D reconstruction, **d** matching result after the application of the match propagation algorithm, **e** three views of the dense 3D reconstruction that corresponds to the last matching result

Sparse and dense matching are shown in Fig. 5b, c respectively. Two views of the sparse 3D reconstruction (7342 3D points were reconstructed) are presented in Fig. 5d. Two views of the obtained dense 3D model (dense 3D point cloud) are presented in Fig. 5e.

5.4 Castle sequence: large-scale scene

In the previous experiments, we have tested our approach on small- and medium-scale scenes. In this part, we used a sequence of 30 images, with resolution of 1020 × 680, of a large-scale scene (castle-P30 [35]). Four images of the sequence are shown in Fig. 6a. We begin by the interest point matching between different images [30]. An example of result obtained between two images is shown in Fig. 6b. The initialization of our system is performed from two selected images [15]. Then, the rest of images is inserted progressively using the already estimated 3D structure and bundle adjustment. Two views of the obtained sparse 3D reconstruction, after the insertion of all images, are shown in Fig. 6d. To obtain a dense 3D reconstruction, we must pass through the dense matching between images. So, we used the match propagation method [36] which starts from the sparse matching result to search for new matches in the vicinities of the old.

Table 3 Comparison results

Approaches	Sequences	RMS error (pixel)	Time (s)
Our approach	*Vase*	0.21	16
	Villa pot	0.29	29.7
	Medusa head	0.23	25
	Castle-P30	0.51	39
	Complex scene	0.86	138
Mouragnon approach [12]	*Vase*	0.65	12
	Villa pot	0.80	25.6
	Medusa head	0.68	21
	Castle-P30	0.79	32
	Complex scene	1.32	117
Pollefeys approach [5]	*Vase*	0.47	108
	Villa pot	0.45	95
	Medusa head	0.52	92
	Castle-P30	0.71	115
	Complex scene	–	–
Schonberger approach [27]	*Vase*	0.18	17
	Villa pot	0.21	62.1
	Medusa head	0.22	43
	Castle-P30	0.39	64
	Complex scene	0.61	381
VisualSFM [37,38]	*Vase*	0.19	12
	Villa pot	0.23	47
	Medusa head	0.21	31
	Castle-P30	0.44	56
	Complex scene	0.65	357

Figure 6c shows the almost dense matching result obtained. The dense 3D model is presented in Fig. 6e.

5.5 Complex scene sequence

In this experiment, our approach was tested on a sequence of 142 images, with a resolution of 740×565, of complex scene composed of objects with different sizes. Four images of the sequence are shown in Fig. 7a. An example of interest point matching between two images, after removing false matches, is shown in Fig. 7b. Two views of sparse 3D reconstruction are presented in Fig. 7c. Figure 7d shows matching result after the application of the match propagation algorithm [36]. Three views of dense 3D reconstruction that corresponds to the last matching result are presented in Fig. 7e.

As presented in the experiments, our approach allows to obtain 3D reconstruction results of quality for different types of objects/scenes. These results also prove the reliability of our formulation for the estimation of intrinsic and extrinsic camera parameters.

To evaluate our approach, four state-of-the-art methods [5,12,27,37] are used. Pollefeys approach [5] uses structure from motion and global bundle adjustment for the recovery of 3D projective structure and camera motion. This approach requires a camera self-calibration step to pass from the projective 3D structure to metric 3D structure. Mouragnon approach [12] is based on incremental structure from motion and the local bundle adjustment for the estimation of 3D structure and camera motion from video captured by a calibrated camera. (To adapt this approach to our situation, we used our camera self-calibration method.) Schonberger approach [27] and VisualSFM [37,38] are two incremental Structure from Motion systems for 3D reconstruction from unordered image collections.

Table 3 shows that our method gives satisfactory results compared to the other methods [5,12,27,37]. This is due to the good initialization based on the proposition of a new approach for self-calibration from two images (taking into account the radial distortion) which allows us estimating a set of 3D points of the scene, and also to the incremental recovery of new 3D points (after inserting new images) based on the local bundle adjustment as well as to suitable integration of the global bundle adjustment. Concerning the computation time (not counting the matching time), our approach is close to that of Mouragnon approach [12] and it is more rapid

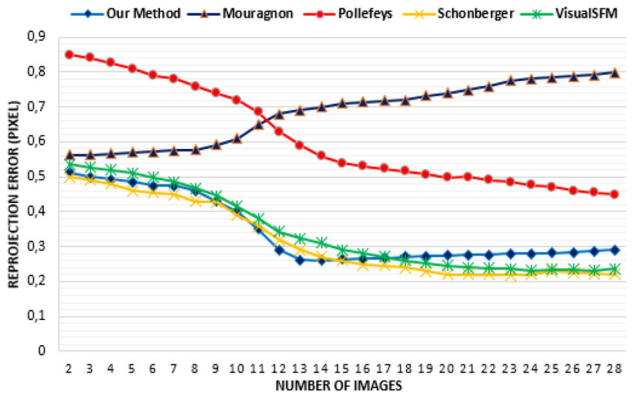

Fig. 8 Reprojection error in terms of images' number

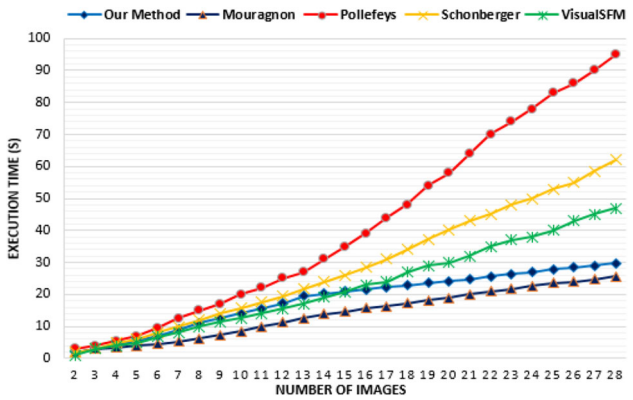

Fig. 9 Execution time needed for the sparse 3D reconstruction (not counting the matching time)

than Pollefeys approach [5], Schonberger approach [27] and VisualSFM [37,38].

To test the performance of our approach compared to other methods [5,12,27,37] in function of the number of used images. We used the sequence 2 (the other sequences lead to similar results). The results are shown in Figs. 8 and 9.

As shown in Fig. 8, concerning Mouragnon approach [12], the reprojection error rises with the increase of images number because of errors accumulation as it is an incremental approach based on local bundle adjustment. On the contrary, concerning Pollefeys approach [5], the reprojection error decreases when images' number increases because it is based on global bundle adjustment. So, with the use of a large number of images the Pollefeys approach becomes more stable. Our approach allows to have more accurate results than these two methods, and it is closer to those obtained by Schonberger approach [27] and VisualSFM [37,38]

When the images number is between 2 and 13 (in these experiments we took $M_0 = 13$), our approach performs as global structure from motion systems [5] with more precision, because it is based on global bundle adjustment (GBA) with good initialization of different parameters (those obtained by local bundle adjustment). When the number of images is greater than 13, we note that the reprojection error is almost stable with some augmentation because the new images are inserted on the basis of the obtained initial 3D structure (already optimized locally and globally), and on local bundle adjustment.

As shown in Fig. 9, our approach is faster compared to Schonberger approach [27] and VisualSFM [37,38], and it is much faster than Pollefeys approach [5]. This latter applies the global bundle adjustment on all estimated entities after the insertion of all images, which requires a long calculation time especially with the increase of images number and can even pose convergence problems (problem resolution by Levenberg–Marquardt algorithm [33] with a bad initialization). On the other hand, the proposed approach allows to have results of quality, in a time close to Mouragnon method [12] which is based on local bundle adjustment and applied in real time.

6 Conclusion

In this paper, we have proposed a complete system for 3D reconstruction from images/videos taken by a moving camera characterized by varying parameters. Our system allows to automatically recover camera parameters and to obtain metric 3D reconstruction results without gone through a 3D projective reconstruction. It is properly initialized from two images with a large camera motion. So, we have proposed a new method for automatic estimation of intrinsic and extrinsic camera parameters. Incremental 3D reconstruction systems are based on the local bundle adjustment to ensure the rapidity. But, the quality of reconstruction results can be affected by errors accumulation. Our system introduces a global bundle adjustment after the insertion of a suitable images number by avoiding the use of all images in order not to fall on optimization problems with a large parameters number to be optimized, which requires a lot of calculation time. The optimal 3D model obtained and the local bundle adjustment will be used for the insertion of the rest of images. Our 3D reconstruction system is completely automatic and provides more reliability in keeping rapidity.

References

1. El Hazzat, S., Saaidi, A., Karam, A., Satori, K.: Incremental multi-view 3D reconstruction starting from two images taken by a stereo pair of cameras. 3D Res. **6**, 11 (2015). doi:10.1007/s13319-015-0041-z
2. Lhuillier, M., Quan, L.: A quasi-dense approach to surface reconstruction from uncalibrated images. IEEE Trans. Pattern Anal. Mach. Intell. **27**(3), 418–433 (2005)
3. Wong, S.S., Chan, K.L.: 3D object model reconstruction from image sequence based on photometric consistency in volume space. Pattern Anal. Appl. **13**(4), 437–450 (2009)

4. Ding, L., Ding, X., Fang, C.: 3D face sparse reconstruction based on local linear fitting. Vis. Comput. **30**(2), 189–200 (2014)
5. Pollefeys, M., Gool, L.V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. Int. J. Comput. Vis. **59**(3), 207–232 (2004)
6. Merras, M., El Hazzat, S., Saaidi, A., Satori, K., Nazih, A.: 3D face reconstruction using images from cameras with varying parameters. Int. J. Autom. Comput. (2016). doi:10.1007/s11633-016-0999-x
7. Liu, J., Li, C., Mei, F., Wang, Z.: 3D entity-based stereo matching with ground control points and joint second order smoothness prior. Vis. Comput. **31**(9), 1253–1269 (2015)
8. Tola, E., Strecha, C., Fua, P.: Efficient large-scale multi-view stereo for ultra high-resolution image sets. Mach. Vis. Appl. **23**(5), 903–920 (2012)
9. Vu, H.H., Labatut, P., Pons, J.P., Keriven, R.: High accuracy and visibility-consistent dense multiview stereo. IEEE Trans. Pattern Anal. Mach. Intell. **34**(5), 889–901 (2012)
10. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. IEEE Trans. Pattern Anal. Mach. Intell. **32**(8), 1362–1376 (2010)
11. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3D. In: SIGGRAPH Conference Proceedings, pp. 835–846 (2006)
12. Mouragnon, E., Lhuillier, M., Dhome, M., Dekeyser, F., Sayd, P.: Generic and real-time structure from motion using local bundle adjustment. Image Vis. Comput. **27**(8), 1178–1193 (2009)
13. Fuhrmann, S., Langguth, F., Moehrle, N., Waechter, M., Goesele, M.: MVE—an image-based reconstruction environment. Comput. Gr. **53**, 44–53 (2015). Part A
14. El Akkad, N., El Hazzat, S., Saaidi, A., Satori, K.: Reconstruction of 3D scenes by camera self-calibration and using genetic algorithms. 3D Res. **7**, 6 (2016). 10.1007/s13319-016-0082-y
15. Wang, G., Wu, Q.M.J.: Perspective 3-D Euclidean reconstruction with varying camera parameters. IEEE Trans. Circuits Syst. Video Technol. **19**(12), 1793–1803 (2009)
16. Strecha, C., Tuytelaars, T., Van Gool, L.: Dense matching of multiple wide-baseline views. In: Proceedings of the International Conference on Computer Vision, pp. 1194–1201 (2003)
17. Amenta, N., Bern, M., Kamvysselis, M.: A new Voronoi-based surface reconstruction algorithm. In: Proceedings of SIGGRAPH'98, pp. 415–421 (1998)
18. Kazhdan, M., Bolithp, M., Hoppe, H.: Poisson surface reconstruction. In: Proceedings of Eurographics Symposium on Geometry Processing, pp. 61–70 (2006)
19. Lim, H., Lim, J., Kim, H. J.: Real-time 6-DOF monocular visual SLAM in a large-scale environment. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 1532–1539 (2014)
20. Kerl, C., Sturm, J., Cremers, D.: Dense visual slam for RGB-D cameras. In: International Conference on Intelligent Robots and Systems (IROS), pp. 2100–2106 (2013)
21. Fischler, M.A., Bolles, R.C.: Random sample consensus: aparadigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
22. Wu, C., Agarwal, S., Curless, B., Seitz. S. M.: Multicore bundle adjustment. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3057–3064 (2011)
23. Lourakis, M.A., Argyros, A.: SBA: a software package for generic sparse bundle adjustment. ACM Trans. Math. Softw. **36**(1), 1–30 (2009)
24. Brown, M., Lowe, D. G.: Unsupervised 3D object recognition and reconstruction in unordered datasets. In: Proceedings of the International Conference on 3D Digital Imaging and Modelling, pp. 56–63 (2005)
25. Tran, S., Davis, L.: 3D surface reconstruction using graph cuts with surface constraints. In: Proceedings of the European Conference on Computer Vision, pp. 219–231 (2006)
26. Snavely, N., Seitz, S., Szeliski, R.: Modeling the world from internet photo collections. Int. J. Comput. Vis. **80**(2), 189–210 (2008)
27. Schonberger, J. L., Frahm, J.-M.: Structure-from-motion revisited. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
28. Wu, C.: Critical configurations for radial distortion self-calibration. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 25–32 (2014)
29. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000)
30. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
31. Wu, J., Cui, Z., Sheng, V.S., Zhao, P., Su, D., Gong, S.: A comparative study of sift and its variants. Meas. Sci. Rev. **13**(3), 122–131 (2013)
32. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004)
33. Moré, J.J.: The Levenberg–Marquardt algorithm: implementation and theory. In: Watson, G.A. (ed.) Numerical Analysis, Lecture Notes in Mathematics, vol. 630, pp. 105–116. Springer, Berlin (1977)
34. http://www.cs.unc.edu/~marc/research.html
35. Strecha, C., Von Hansen, W., Van Gool, L., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: CVPR, pp. 1–8 (2008)
36. Lhuillier, M., Quan, L.: Match propagation for image-based modeling and rendering. IEEE Trans. Pattern Anal. Mach. Intell. **24**(8), 1140–1146 (2002)
37. Wu, C.: Towards linear-time incremental structure from motion. In: International Conference on 3D Vision (3DV), pp. 127–134 (2013)
38. http://ccwu.me/vsfm/

Soulaiman El Hazzat received the bachelor's and master's degrees from SMBA-Fez University in 2003 and 2012, respectively. He is currently working toward the PhD degree in the LIIAN Laboratory at SMBA-Fez University. His current research interests include camera self-calibration and 3D reconstruction.

Mostafa Merras received the bachelor's and master's degrees from SMBA-Fez University in 2006 and 2009, respectively. He is currently working toward the PhD degree in the LIIAN Laboratory at SMBA-Fez University. His current research interests include camera calibration and self-calibration by the genetic algorithms.

Nabil El Akkad received the PhD degree from SMBA-Fez University in 2014. He is currently a professor of computer science at National School of Applied Sciences (ENSA) of Al-Hoceima, University of Mohamed First, Oujda, Morocco. He is a member of the LIIAN Laboratory. His research interests include camera self-calibration, 3D reconstruction, genetic algorithms and real-time rendering.

Abderrahim Saaidi received the PhD degree from SMBA-Fez University in 2010. He is currently a professor of computer science at SMBA-Taza University. He is a member of the LIIAN and LIMAO Laboratories. His research interests include camera self-calibration, 3D reconstruction, genetic algorithms and real-time rendering.

Khalid Satori received the PhD degree from the National Institute for the Applied Sciences INSA at Lyon in 1993. He is currently a professor of computer science at SMBA-Fez University. He is the director of the LIIAN Laboratory. His research interests include real-time rendering, image-based rendering, virtual reality, biomedical signal, camera self-calibration, genetic algorithms and 3D reconstruction.