3D Res (2015) 6:11
DOI 10.1007/s13319-015-0041-z

3DR EXPRESS

Incremental Multi-view 3D Reconstruction Starting from Two Images Taken by a Stereo Pair of Cameras

Soulaiman El hazzat · Abderrahim Saaidi · Antoine Karam · Khalid Satori

Received: 30 November 2014 / Revised: 19 February 2015 / Accepted: 21 February 2015 / Published online: 13 March 2015

Abstract In this paper, we present a new method for multi-view 3D reconstruction based on the use of a binocular stereo vision system constituted of two unattached cameras to initialize the reconstruction process. Afterwards, the second camera of stereo vision system (characterized by varying parameters) moves to capture more images at different times which are used to obtain an almost complete 3D reconstruction. The first two projection matrices are estimated by using a 3D pattern with known properties. After that, 3D scene points are recovered by triangulation of the matched interest points between these two images. The proposed approach is incremental. At each insertion of a new image, the camera projection matrix is estimated using the 3D information already calculated and new 3D points are recovered by triangulation from the result of the matching of interest points between the inserted image and the previous image. For the refinement of the new projection matrix and the new 3D points, a local bundle adjustment is performed. At first, all projection matrices are estimated, the matches between consecutive images are detected and Euclidean sparse 3D reconstruction is obtained. So, to increase the number of matches and have a more dense reconstruction, the Match propagation algorithm, more suitable for interesting movement of the camera, was applied on the pairs of consecutive images. The experimental results show the power and robustness of the proposed approach.

Keywords Multi-view 3D reconstruction · Incremental approach · Local bundle adjustment · Sparse 3D reconstruction · Dense 3D reconstruction

1 Introduction

3D reconstruction from images is important and widely studied in computer vision field, it has many applications: robotics, monitoring, measurement, quality control, virtual reality and others.

S. El hazzat (✉) · A. Saaidi · K. Satori
LIIAN, Department Of Mathematics and Informatics, Faculty of Sciences Dhar-Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco
e-mail: soulaiman.elhazzat@yahoo.fr

A. Saaidi
e-mail: saaidi.abde@yahoo.fr

K. Satori
e-mail: khalidsatorim3i@yahoo.fr

A. Saaidi
LSI, Department of Mathematics, Physics and Informatics, Polydisciplinary Faculty of Taza, Sidi Mohamed Ben Abdellah University, Taza, Morocco

A. Karam
Faculty of Science II, Lebanese University, Fanar Campus, Jadeite El Metn., P.O. Box 90656, Beirut, Lebanon
e-mail: amkaram@ul.edu.lb

Several approaches provide a solution to this problem. They can be divided into two major categories: active approaches [26, 27] and passive approaches [4–11]. Active approaches use the laser or a structured light to find the three-dimensional coordinates. On the other side, passive approaches use only the 2D images taken by one or more cameras for three-dimensional reconstruction of the captured scene.

In this work, we are interested in the passive approaches that make 3D reconstruction from a set of images taken from different viewpoints. But, to recover the depth and render the scene in 3D. The intrinsic and extrinsic camera parameters must be recovered, either by a known pattern (calibration) [1] or from images without any a priori recognition about the scene (self-calibration) [2]. After camera parameters estimation, the three-dimensional geometry of the object (or scene) can be recovered by solving the matching problem between the images [4–7, 11] which is to find projections of the same 3D scene points in images. From these matches and camera parameters, the 3D coordinates of the scene points are estimated by triangulation. Currently, there are approaches based on the stereo vision that allow, from a sequence of calibrated stereo images, to obtain dense 3D models with high accuracy [23]. Also, there are other approaches called volumetric [9] which are based on a discretization of space into basic elements called voxels. Starting from a volume containing the initial object (or the scene) discretized into voxels. The treatment is to remove the voxels not complying with certain criteria, to finally find a volume representation of the object or scene (remaining voxels). The structure from motion approach [13, 14, 21, 22] allows you to automatically recover, both the 3D scene structure and the camera positions. It is based on the detection and matching of interest points between different images.

In this paper, we proposed an incremental approach for multi-view 3D reconstruction based on the use of a binocular stereo vision system, consisting of two unattached cameras, for reliable initialization of the reconstruction process. Afterwards, to have an almost complete 3D reconstruction, the second camera of stereo vision system moves around the object or scene to capture more images (Fig. 1). The first two projection matrices are well estimated using a 3D pattern with known properties. The matching of interest points, detected by Harris [12], is made by the NCC correlation [28] and the coordinates of 3D points are recovered by triangulation from the obtained matches and projection matrices. At each insertion of a new image, the camera projection matrix is estimated by the use of the 3D information recovered previously. The camera projection matrix P_k for the image I_k $(k \geq 3)$ is estimated from N points $(N \geq 6)$ of 3D points already recovered (by triangulation of interest points matched between I_{k-2} and I_{k-1}) and their projections localized in the image I_k. After that, the new 3D points are recovered from interest points detected and matched between the current image I_k and the previous image I_{k-1}. For the refinement of the estimated projection matrix and the new 3D points, a local bundle adjustment is performed. After insertion of the last image, we have all projection matrices and a sparse 3D reconstruction of the all interest points matched between consecutive images.

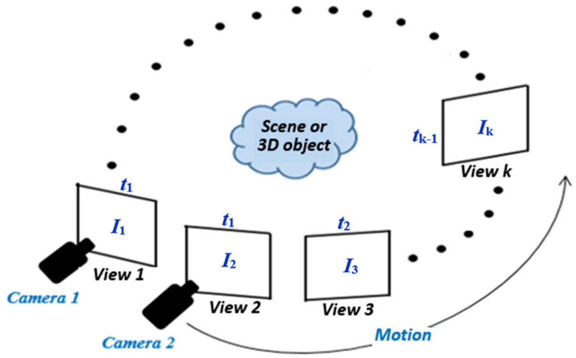


Fig. 1 Multi-view 3D reconstruction system used. The first two images $\{I_1, I_2\}$ are taken by a system of binocular stereovision (two unattached cameras). After, the second camera moves to capture the other images

The estimation of a limited number of 3D coordinates of the points is insufficient to define the shape of the object (or scene). So, to increase the number of reconstructed points and have a more dense 3D reconstruction, the Match Propagation algorithm is used [3]. Starting from the points already matched (seeds) between consecutive images and searching in the vicinity of the new matches.

The proposed method has many advantages: the use of a binocular stereo vision system consisting of two unattached cameras offers more flexibility to properly choose the distance between the two cameras (baseline), a fact that offers more robustness for 3D reconstruction of objects or scenes of different sizes (small, medium and large). The effective initialization of the reconstruction system by the use of a binocular stereo vision system and the local bundle adjustment after the insertion of a new image allowed an automatic and reliable estimation of the camera projection matrices for the other images and to have an Euclidean 3D reconstruction without passing by a projective 3D reconstruction as in the 3D reconstruction methods from uncalibrated images based on structure from motion approach. In addition, it allowed to have satisfying results (3D models) by avoiding a bad initialization of the reconstruction process. The estimation of the camera projection matrix for each inserted image is made by a linear system resolution. The obtained solution can be refined by minimizing a nonlinear criterion. As opposed to other methods based on the self-calibration that require the formulation of nonlinear equations that require more time to be solved.

This paper is organized as follows. The Sect. 2 presents related works. The Sect. 3 describes the notations used. In the Sect. 4, the presentation and description of the proposed method. Experimentation and comparison of the proposed method with other methods is presented in Sect. 5. Finally, the conclusion is presented in the Sect. 6.

2 Related Work

The methods of 3D reconstruction from images taken from different viewpoints by one or many cameras can be classified as follows: methods that use calibrated images and other methods that start from uncalibrated images to automatically find the camera parameters and make the 3D reconstruction of the scene at the same time.

The methods based on the stereo vision take in input stereo calibrated images and allow a dense reconstruction of the object or the scene. Da and Sui [15] have proposed a method based on binocular stereo vision (using only two images) for dense 3D reconstruction of face. First, they began with the calibration and rectification of two images. Then, the matching is performed in two steps: the sparse matching of interest points detected by Harris [12] and dense matching by using piecewise dynamic programming. Finally, the 3D information is recovered by triangulation. However, to increase the reconstructed area and have complete 3D models, the reconstruction process needs more than two images, this is called multi-view stereo. Furukawa and Ponce [6] have proposed a new method for multi-view stereo. It's based on the reconstruction of a set of oriented points (patches) covering the surface of the object or the scene. So, they began with the matching based on the epipolar constraint between key points detected by the DoG and Harris operators. After that, for each pair of matched points they constructed a patch candidate defined by the center, the normal and the reference image. For the elimination of false matches, they used the visibility constraints. They also proposed simple methods to transform the resulting patch model into a polygonal mesh which can finally be refined by application of the photometric consistency and regularization constraints. Other methods are based on the depth map estimation. Goesele et al. [10] proposed an algorithm to solve the problem of multi-view stereo, it consists in a first step to reconstruct a depth map for each input view by the use of window-matching with a small number of neighboring views which allows to have good matches. The second step is to merge the resulting depth maps into a mesh model using a volumetric approach. Yuan and Lu [25] presented an incremental method to solve the problem of multi-view stereo using Bayesian learning. First, they reconstructed an initial 3D model, from uniformly distributed key images on a view sphere. Then, when a new calibrated image is inserted, the initial 3D model is updated automatically by the use of Bayesian learning with the photometric consistency and geometric constraints. Mouragnon et al. [21] presented a real-time method for estimating the motion of a calibrated camera and the 3D structure from video, it is applied both for the perspective camera model and the generic camera model. The proposed method is based on local bundle adjustment to refine the positions of the camera and the 3D points. Other methods called volumetric start from an initial volume, containing the object, discretized into basic elements called voxels and uses the 2D information to restore the shape of the object. The treatment is to keep only the voxels representing the object and eliminate other. The voxel coloring method [17], generalized voxel coloring [18] and space carving [19] are volumetric methods, which depart from a bounding volume discretized into voxels and use the photo-consistency test and visibility constraint for volumetric 3D reconstruction. Slabaugh et al. [16] presented improvements to calculate the visibility and photo-consistency for volumetric 3D reconstruction from calibrated images taken by multiple cameras placed at arbitrary viewpoints. Mulayim et al. [20] presented a complete system of 3D reconstruction of real objects in a controlled environment (turn-table). It is based on the multi-image calibration, the use of silhouettes image, photo-consistency and visibility of voxels to finally have the textured 3D models.

All methods already cited are based on the calibration or the use of already calibrated images. Sometimes it is necessary to completely automate the 3D reconstruction process. In this case, there is a need to cameras' self-calibration.

Pollefeys et al. [13] presented a complete system of 3D reconstruction from uncalibrated image taken with a hand-held camera. It is based on the structure from motion approach. First, they started by the detection and matching of interest points between images to calculate the relationship between the different views and recover the projective structure of the scene and camera motion. After, to have an affine structure, they have passed through a phase of the camera self-calibration. Finally, a multi-view stereo matching algorithm is used to obtain a dense 3D reconstruction. In [14], they presented a method based on perspective factorization and self-calibration for recovering the Euclidean 3D structure and camera motion from video sequence. They proposed to initialize the projective depths via a projective structure reconstructed from two views with large camera movement, which is very useful during the optimization phase (the solution converges rapidly). To recover the Euclidean structure, they proposed a self-calibration method based on Kruppa constraint. Lhuillier and Quan [5] proposed a method of quasi-dense reconstruction from uncalibrated images, this approach is based on the use of match propagation [3] to have the quasi-dense matching and then the robust and accurate geometry estimation and have a more adequate surface representation.

Each method cited above has its advantages and disadvantages. First, we begin with the first class of methods. Methods of reconstruction based on binocular stereo vision [15] used to obtain reliable results because of a priori knowledge of the stereo camera pair configuration which facilitates the calibration and matching between images. But from two images only, one cannot have a complete reconstruction. Currently, there are so-called stereo multi-view methods (MVS) [6, 10, 25] that allow to have accurate results from a set of calibrated stereo images. The volumetric methods [16–20] require knowledge of a 'bounding box'. Incremental 3D reconstruction methods [21, 25] are fast and allow to have satisfactory results after the refinement of result of initial 3D reconstruction. However, this class of methods requires the use of calibrated stereo images. The second class of methods [5, 13, 14] start from uncalibrated images for find both the projective 3D structure and the camera motion and requires a self-calibration phase to recover 3D affine structure. But, this class of methods allows to have 3D reconstruction results up to a scale factor and self-calibration problem often requires to impose constraints on the camera parameters.

3 Notation and Background

In this work, the pinhole camera model is used. A scene point $M_j = (X_j, Y_j, Z_j, 1)^T$ is projected onto the image plane at a point $m_{ij} = (u_{ij}, v_{ij}, 1)^T$. This projection is represented by the following formula:

$$\lambda_{ij} m_{ij} = P_i M_j$$

With: λ_{ij} is a nonzero scale factor and P_i is the perspective projection matrix.

The following notations are used:

I_k is the kth image.

$P_k = (p^k_{xy})^{y=1..4}_{x=1..3}$ is the camera projection matrix corresponding to the image I_k.

$A_{i,j} = \{(m_{ik}, m_{jk})/k = 1, \ldots, n_{i,j}\}$ is the set of pairs of interest points matched between the images I_i and I_j, with $n_{i,j}$ is the number of matches.

$A_i^j = \{m_{ik} = (u_{ik}, v_{ik})/k = 1, \ldots, n_{i,j}\}$ is the set of interest points in the image I_i matched with interest points in the image I_j.

$S_{k,k+1}$ is the set of 3D points obtained by triangulation from the set of matches $A_{k,k+1}$ and the projection matrices P_k and P_{k+1}.

4 Proposed Method

The proposed method is based on the reliable initialization of 3D reconstruction process by using a binocular stereo vision system (two unattached cameras properly installed) that allows to avoid any bad initialization so as not to affect the 3D reconstruction result as in the case of the 3D reconstruction system from uncalibrated images based on structure from motion approach [13] and allows also to offer more flexibility to the 3D reconstruction of objects or scenes of various sizes by the suitable choice of the distance between the two cameras (baseline). The major drawback of the binocular stereo vision is that the reconstructed area is limited because of the use of two images only. To avoid it, new images are captured around the object or scene by the second camera (characterized by varying parameters) at different moments and gradually inserted to get an almost complete 3D reconstruction (it is easier to move a single camera than the displacement of binocular stereo vision system because the two cameras are unattached).

Our approach for 3D reconstruction is performed in four essential steps outlined below:

1. The use of a binocular stereo vision system for reliable initialization of the 3D reconstruction process (the reliable estimation of the coordinates of 3D points that correspond to interest points matched between the first two stereo images).
2. Automatic estimation of the camera projection matrix P_k $(k \geq 3)$ after the insertion of the image I_k and 3D reconstruction of interest points matched between the images I_{k-1} and I_k (the set $S_{k-1,k}$).
3. Refinement of the camera projection matrix P_k and coordinates of new reconstructed 3D points (the set $S_{k-1,k}$).
4. Matching and dense 3D reconstruction by fusion of results obtained (after the application of the match propagation algorithm [3]) between consecutive images pairs.

The step 2 and the step 3 are repeated for each new inserted image I_k ($3 \leq k \leq n$, n is the total number of images).

The diagram presented in the Fig. 2 describes the enchainment of our approach.

In this work, interest points are detected by Harris algorithm [12]. The normalized cross-correlation (NCC) [28] was used for the matching of these points between consecutive images.

For an interest point of the image I_k, its correspondent in the image I_{k+1} (if it exists) is the interest point of maximum NCC value and greater than a threshold. The matches obtained are not all correct, RANSAC algorithm [24] was used to eliminate false matches.

4.1 Initialization

Our 3D reconstruction approach requires an essential initialization phase which can be realized by the following five steps:

1. Installation and calibration of our binocular stereo vision system using a 3D pattern with known properties (Fig. 3).
   Let P_1 and P_2 be the projection matrices of the two cameras.
2. Acquisition of two images, I_1 and I_2.
3. Detection of interest points with Harris algorithm [12] and matching of these points by the NCC [28] and epipolar constraint [11]. Let $A_{1,2}$ be the set of obtained matches, $A_{1,2}$ can be decomposed as follows:

$$A_{1,2} = A_1^2 \times A_2^1 \tag{1}$$

4. Sparse 3D reconstruction by triangulation from the obtained matches and the projection matrices. Let $S_{1,2} = \{M_j = (X_j, Y_j, Z_j)^T / j = 1, \ldots, n_{1,2}\}$ be the set of 3D points obtained by triangulation from pairs of matched points $(m_{1j}, m_{2j}) \in A_{1,2}$ ($n_{1,2}$ is the number of matches).
   The coordinates of the 3D point M_j are calculated by the following system of equations:

$$\begin{cases} \tilde{m}_{1j} \sim P_1 \tilde{M}_j \\ \tilde{m}_{2j} \sim P_2 \tilde{M}_j \end{cases} \tag{2}$$

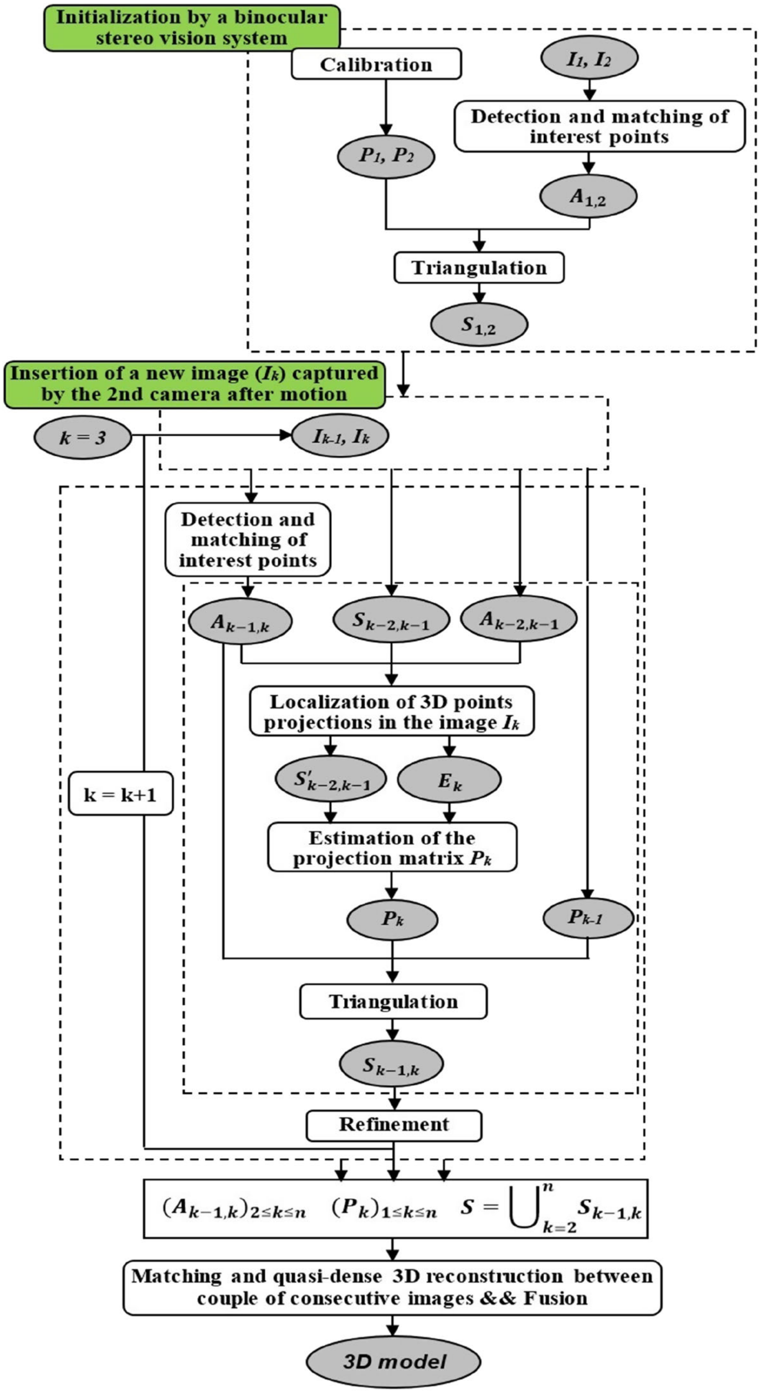
In developing these equations, we obtain a system form:

$$A\tilde{M}_j = 0 \tag{3}$$

With:

$$A = \begin{pmatrix} p_{11}^1 - p_{31}^1 u_{1j} & p_{12}^1 - p_{32}^1 u_{1j} & p_{13}^1 - p_{33}^1 u_{1j} & p_{14}^1 - p_{34}^1 u_{1j} \\ p_{21}^1 - p_{31}^1 v_{1j} & p_{22}^1 - p_{32}^1 v_{1j} & p_{23}^1 - p_{33}^1 v_{1j} & p_{24}^1 - p_{34}^1 v_{1j} \\ p_{11}^2 - p_{31}^2 u_{2j} & p_{12}^2 - p_{32}^2 u_{2j} & p_{13}^2 - p_{33}^2 u_{2j} & p_{14}^2 - p_{34}^2 u_{2j} \\ p_{21}^2 - p_{31}^2 v_{2j} & p_{22}^2 - p_{32}^2 v_{2j} & p_{23}^2 - p_{33}^2 v_{2j} & p_{24}^2 - p_{34}^2 v_{2j} \end{pmatrix} \tag{4}$$

Fig. 2 Descriptive Scheme of the proposed approach


Initialization by a binocular stereo vision system
Calibration
I1, I2
Detection and matching of interest points
P1, P2
A1,2
Triangulation
S1,2
Insertion of a new image (Ik) captured by the 2nd camera after motion
k = 3
Ik-1, Ik
Detection and matching of interest points
Ak−1,k
Sk−2,k−1
Ak−2,k−1
Localization of 3D points projections in the image Ik
S′k−2,k−1
Ek
k = k+1
Estimation of the projection matrix Pk
Pk
Pk-1
Triangulation
Sk−1,k
Refinement
(Ak−1,k)2≤k≤n (Pk)1≤k≤n S = ⋃nk=2 Sk−1,k
Matching and quasi-dense 3D reconstruction between couple of consecutive images && Fusion
3D model

The coordinates *(X*$_j$*, Y*$_j$*, Z*$_j$*)* of the point M_j are obtained by a singular value decomposition (SVD) of the matrix A.

5. Refinement of the coordinates of 3D points by minimizing the criterion (5) by the Levenberg–Marquard algorithm [29]:

$$C(\{M_j\}_{j=1}^{n_{1,2}}) = \sum_{i=1}^{2}\sum_{j=1}^{n_{1,2}}\left\|m_{ij} - \varphi(P_i, M_j)\right\|^2 \qquad (5)$$

$n_{1,2}$ is the number of matches (this is also the number of reconstructed 3D points).

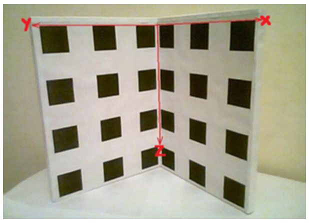


Fig. 3 3D pattern used for the calibration of our binocular stereo vision system

$\varphi(P_i, M_j)$ is the projection of the point M_j in the image $I_{i.}$

4.2 Insertion of a New Image

The second camera of the stereo vision system, characterized by varying parameters, moves around the object or scene to capture more images $\{I_k\}_{3 \le k \le n}$ at different times. For each new inserted image I_k, the estimation of the camera projection matrix and the new 3D points is performed by the following steps:

1. Detection of interest points of the inserted image I_k $(k \ge 3)$.
2. Matching of interest points between the inserted image I_k and the previous image I_{k-1} (already inserted). Let $A_{k-1,k}$ be the set of obtained matches. $A_{k-1,k}$ can be decomposed as follows:

$$A_{k-1,k} = A^k_{k-1} \times A^{k-1}_k. \tag{6}$$

3. Localization of the projections of 3D points $M_j \in S_{k-2,k-1}$ in the image I_k (use of matches between the images I_{k-1} and I_k) (see Figs. 4, 5).

   Let B_{k-1} denote the set of interest points of the image I_{k-1} matched with the interest points of the image I_{k-2} as well as with the interest points of the image I_k. This set can be expressed by the following formula:

$$B_{k-1} = A^{k-2}_{k-1} \cap A^k_{k-1} = \{b^{k-1}_1, b^{k-1}_2, \ldots, b^{k-1}_{m_{k-1}}\} \tag{7}$$

   Such as: $\dim(B_{k-1}) = m_{k-1} \le \min(n_{k-2,k-1}, n_{k-1,k})$.
   And denote by:

$$S'_{k-2,k-1} = \{N^{k-1}_i = (X^{k-1}_i, Y^{k-1}_i, Z^{k-1}_i)^T / i = 1, \ldots, m_{k-1}\} \tag{8}$$

   The set of 3D points, with $b^{k-1}_i \in B_{k-1}$ is the projection of N^{k-1}_i in the image I_{k-1} $(S'_{k-2,k-1} C S_{k-2,k-1})$.

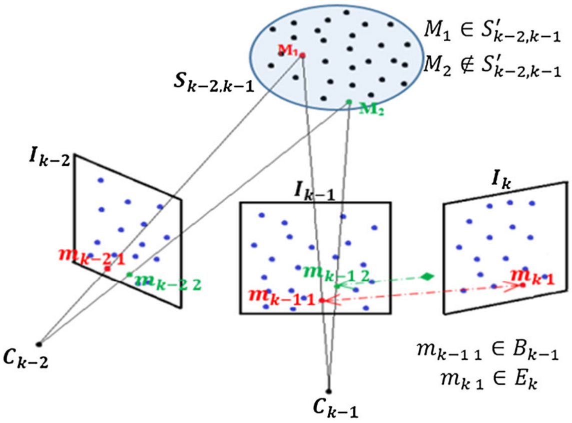


Fig. 4 Localization of the projections of 3D points $M_j \in S_{k-2,k-1}$ in the image I_k (m_{k1} is the projection of M_1 localized in I_k by matching between I_{k-1} and I_k. On the other hand, M_2 has not been located)

   And:

$$E_k = \{e^k_1, e^k_2, \ldots, e^k_{m_{k-1}}\} CA^{k-1}_k \tag{9}$$

   The set of the points in the image I_k that corresponds to the set of points B_{k-1}.
   Then, the points $e^k_i \in E_k$ are projections of points $N^{k-1}_i \in S'_{k-2,k-1}$ in the image I_k, with:

$$\dim(E_k) = \dim(B_{k-1}) = \dim(S'_{k-2,k-1}) = m_{k-1}$$

4. Estimation of the camera projection matrix P_k $(k \ge 3)$.

   The initial projection matrix will be estimated from N points $(N \ge 6)$ selected from the set of points E_k (the choice is based on the correlation score NCC and the distribution of points in the image) and their corresponding 3D points of the set $S'_{k-2,k-1}$.
   For an image point $e^k_j \in E_k$ and their corresponding 3D point $N^{k-1}_j \in S'_{k-2,k-1}$. The formula of perspective projection is represented by:

$$\mu \tilde{e}^k_j = P_k \tilde{N}^{k-1}_j \tag{10}$$

   With: $\tilde{e}^k_j = (u_j, v_j, 1)^T$ and $\tilde{N}^{k-1}_j = (X_j, Y_j, Z_j, 1)^T$.
   µ is a nonzero scale factor.
   By developing the Eq. (10) and by replacing the nonzero scale factor, we find:

$$\begin{cases} X_jp_{11}^k + Y_jp_{12}^k + Z_jp_{13}^k + p_{14}^k - u_jX_jp_{31}^k - u_jY_jp_{32}^k - u_jZ_jp_{33}^k = u_jp_{34}^k \\ X_jp_{21}^k + Y_jp_{22}^k + Z_jp_{23}^k + p_{24}^k - v_jX_jp_{31}^k - v_jY_jp_{32}^k - v_jZ_jp_{33}^k = v_jp_{34}^k \end{cases}$$

Then, to determine the coefficients of the projection matrix. The linear system (11) needs to be solved.

$$AQ_k = 0 \tag{11}$$

With: $Q_k = (p_{11}^k, \ldots, p_{34}^k)^T$ and.

$$A = \begin{pmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 & -u_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 & -v_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -u_NX_N & -u_NY_N & -u_NZ_N & -u_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -v_NX_N & -v_NY_N & -v_NZ_N & -v_N \end{pmatrix}$$

N is the number of points used for the initial estimation of the projection matrix ($N \geq 6$).
The coefficients of the matrix P_k are obtained by a SVD of the matrix A.

5. Estimation of new 3D points ($S_{k-1,k}$) by triangulation of the matched points between I_{k-1} and I_k ($A_{k-1,k}$).

4.3 Local Bundle Adjustment

At each insertion of a new image I_k, the elements of the projection matrix P_k and new reconstructed 3D points ($M_j \in S_{k-1,k}$) are refined. The refinement is performed by a local bundle adjustment, faster approach than the global bundle adjustment, the last three images I_{k-2}, I_{k-1} and I_k were used (the 3D points reconstructed from the images I_{k-2} and I_{k-1} and the projection matrices P_{k-2} and P_{k-1} are already estimated and refined, and they will be used in this step to refine the estimated entities). Then the criterion (12) can be minimized by the Levenberg-Marquard algorithm [29]:

$$C(P_k, \{M_j\}_{j=1}^{n_{k-1,k}}) = \sum_{i=k-2}^{k} \sum_{j=1}^{n_{k-1,k}} \left\| m_{ij} - \varphi(P_i, M_j) \right\|^2 \tag{12}$$

With: $M_j \in S_{k-1,k}$, $n_{k-1,k} = \dim \ (S_{k-1,k}) = \dim (A_{k-1,k})$,

m_{ij} is the jth point in the image $I_{i.}$

$\varphi(P_i, M_j)$ is the projection of the point M_j in the image $I_{i.}$

4.4 Matching and Quasi-Dense 3D Reconstruction

The result of the sparse 3D reconstruction of all interest points matched between consecutive images I_{k-1} and I_k ($S = \cup_{k=2}^{n} S_{k-1,k}$) is insufficient to define the shape of the object or the scene. To increase the number of matches and then the number of reconstructed 3D points, we used the match propagation method [3] that is more suitable for interesting movement of the camera and also more practical for unrectified images. This method takes as input, for two consecutive images I_{k-1} and I_k, the set of initial matches $A_{k-1,k}$ (seeds). The treatment consists to seek in each time a new matches in the vicinity of the others.

5 Experiments

In all these experiments, the two first images of the sequence are taken by a pair of unattached stereo cameras. The other images are taken by the second camera of the stereo vision system, characterized by varying parameters that make displacements around the object or scene, of angles between ten and twenty degrees, in order to have an almost complete tridimensional reconstruction.

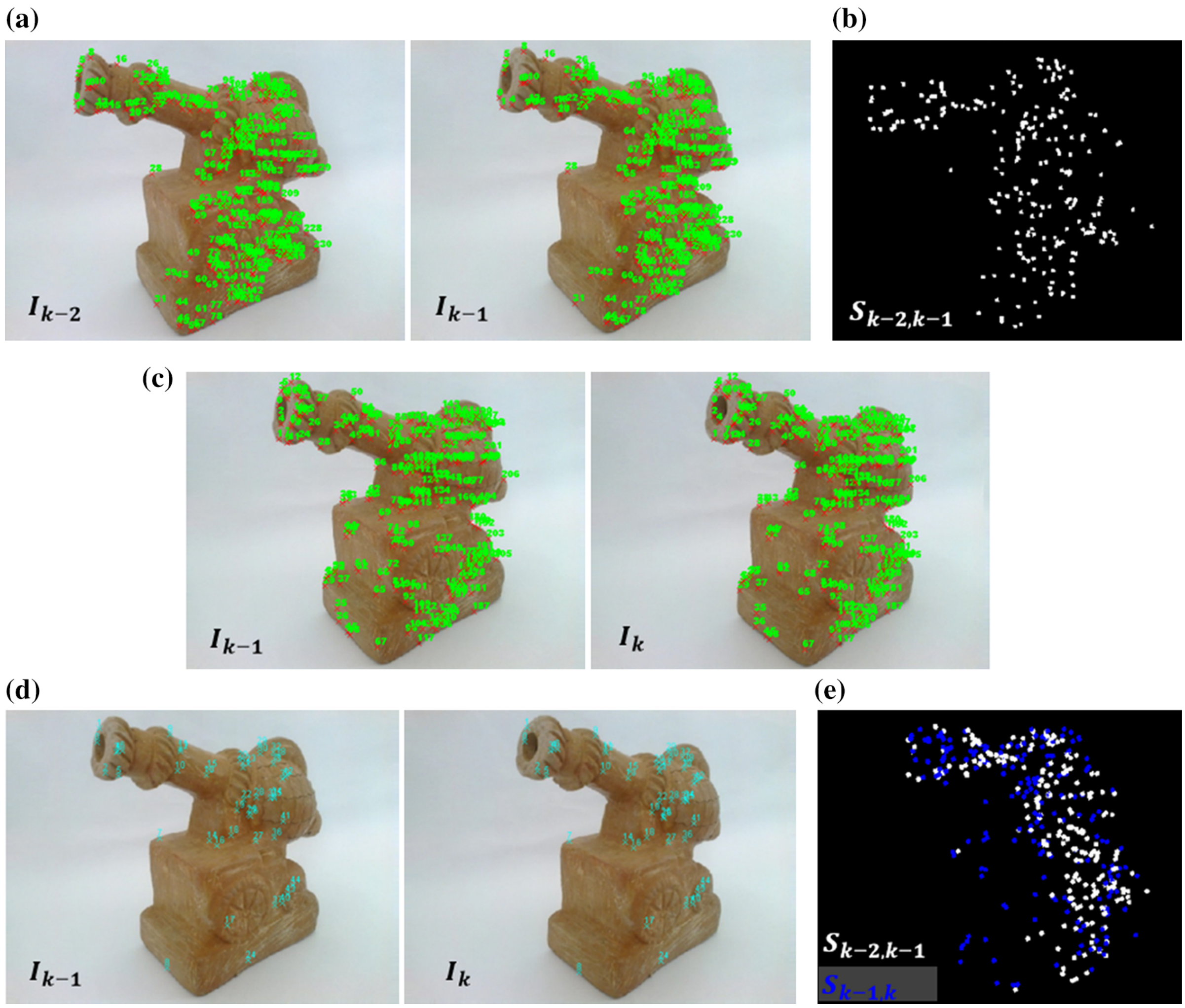


Fig. 5 Steps of the localization of the projections of 3D points $M_j \in S_{k-2,k-1}$ in the image I_k : **a** Matching between the images I_{k-2} and I_{k-1}. **b** Set of 3D points $S_{k-2,k-1}$ estimated by triangulation from the obtained matches between the images I_{k-2} and I_{k-1}. **c** Matching between the imagess I_{k-1} and I_k. **d** projection of points of $S_{k-2,k-1}$ localized in the image I_k. **e** The Set of 3D points obtained $S_{k-1,k}$ (in *blue* color) after the estimation of the projection matrix P_k. (Color figure online)

The different algorithms have been implemented in Java under Eclipse. JAMA Library has been used for matrix computations and Java 3D API for 3D visualization.

5.1 Simulations

A sequence of fourteen images (our approach is operational for $n \geq 2$ images, when $n = 2$ we talk about a binocular stereo vision system) of a 3D box (Fig. 6) of dimension 34 × 34 × 44 cm with 57 squares (every square is of dimension 6 × 6 cm) that is to say 57 × 4 = 228 corners, have been taken from different viewpoints.

The real 3D coordinates of corners in a well-chosen landmarks (Fig. 6) are stocked in a text file.

The proposed method has been used for 3D reconstruction of all corners matched between the consecutive images in the same landmarks. We began with the 3D reconstruction of the matched corners (detected by Harris algorithm [12]) between the first two stereo image (for the calibration, a classic method [11] based on knowledge of 3D points and their projection in the image was used). Then, each

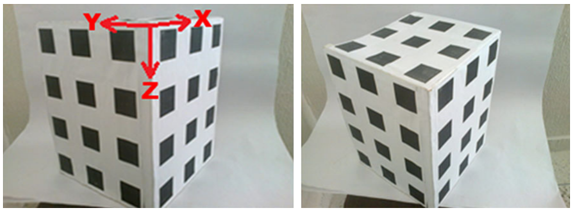


Fig. 6 3D box used to test the performance of the proposed method

insertion of a new image, the projection matrix is initially estimated by the use of the 3D information already calculated and new 3D points are recovered by triangulation from the result of the matching of interest points (corners) between the inserted image and the previous image. The refinement of the new projection matrix and the new 3D points is performed by a local bundle adjustment [21, 22]. The obtained results are compared with those obtained by the approach of Pollefeys [13] and Fitzgibbon [30], which are methods that allow to recover the 3D structure from a sequence of uncalibrated images.

The Fig. 7 presents 3D root mean square (RMS) error defined by the formula (13) as a function the number of images. The value of the error ≪RMS≫ is small and almost stable even if with the increase of the image number a fact that shows the precision of the 3D structure estimation by our incremental 3D reconstruction method. The obtained results indicate also the power of the proposed method compared with the two other methods. The reliability of the obtained results is justified by the good initialization of the 3D reconstruction process based on the use of a binocular stereo vision system and the refinement of new entities estimated by the local bundle adjustment during the insertion of a new image to avoid as much as possible, the error accumulation.

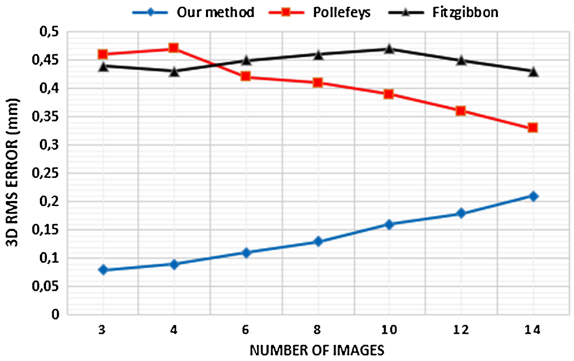


Fig. 7 3D RMS error corresponding to the number of images

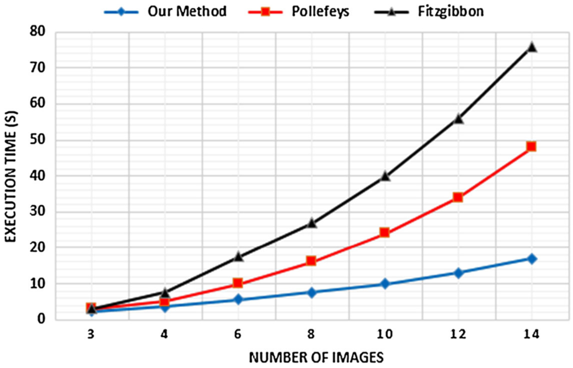


Fig. 8 Execution time corresponding to the number of images

The 3D RMS error is defined by:

$$RMS_{3D} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n} \|M_i - M_i^c\|^2} \qquad (13)$$

n is the number of reconstructed 3D points.

$M_i = (X_i, Y_i, Z_i)$ is the 3D point of known coordinates.

$M_i^c = (X_i^c, Y_i^c, Z_i^c)$ is the reconstructed 3D point.

Figure 8 shows the rapidity of the proposed approach compared with the methods of Pollefeys [13] and Fitzgibbon [30]. Our method is based on a series of local bundle adjustment. On the other hand, the two other methods are based on a global bundle adjustment, which requires the optimization of a large number of parameters, particularly with the increased number of images used, and then requires more computing time.

To test the accuracy and the robustness of the proposed method in presence of noise. A Gaussian noise has been added to all images pixels. The Fig. 9 shows the quality of 3D reconstruction, presented by the 3D RMS error defined by the formula (13), in terms of the Gaussian noise value that varies between 0 and 3. The 3D error value increases with increasing of Gaussian noise. However, for our method the error remains weak if compared to other methods, a fact that shows the robustness of the proposed approach, because of the reliable initialization by the binocular stereo vision system that offers more reliability, especially during the matching step. In addition, at each insertion of a new image, the local bundle adjustment reduces the noise's influence. On the other

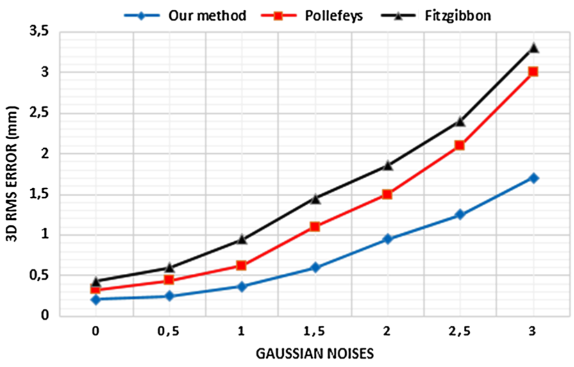


Fig. 9 3D RMS error corresponding to Gaussian noises

side, the two other methods are more sensitive to noise (they use uncalibrated images taken by a moving camera and global bundle adjustment).

5.2 Real Data

To test and validate the robustness of our approach, four real image sequences of chosen objects of different natures are used. The first, a sequence of eighteen images of resolution 640 × 480 of an object of complex shape. The second, a sequence of ten images of resolution 800 × 600 of face. The third is a sequence of eight images of resolution 960 × 1280 of a traditional door and the last is a sequence of seven images of resolution 960 × 1280 of a house.

Table 1 shows the result of sparse 3D reconstruction of the different image sequences. The small value of the reprojection error defined by the formula (14) shows the accuracy of the proposed approach. Which confirms the results of the simulation.

Table 2 and Fig. 11 show the progression of the sparse 3D reconstruction of the different sequences of images. The number of reconstructed 3D points increases with increasing of the images (at each insertion of a new image, new 3D points are reconstructed). The value of the reprojection error also increases but it remains low because of the use of the local bundle adjustment between the last triplets of images, after each insertion of a new image, to maintain the reliability of the system.

The reprojection error is defined by:

$$e = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\varepsilon_{ij}\|m_{ij} - \varphi(P_i, M_j)\|^2 \qquad (14)$$

Table 1 Results of sparse 3D reconstruction of different image sequence

	Number of reconstructed points	Reprojection error
Sequence 1	1013	0.097
Sequence 2	415	0.087
Sequence 3	1087	0.085
Sequence 4	805	0.095

With:

Table 2 Progression of the sparse 3D reconstruction of the different sequences of images

Number of images	Number of reconstructed points	Reprojection error
Sequence 1		
4	351	0.05
8	587	0.07
12	823	0.08
18	1013	0.097
Sequence 2		
3	89	0.054
6	221	0.067
8	332	0.079
10	415	0.087
Sequence 3		
2	412	0.056
4	689	0.068
6	913	0.076
8	1087	0.085
Sequence 4		
2	245	0.053
4	493	0.073
6	710	0.091
7	805	0.095

ε_{ij} is a binary visibility factor.
m is the number of images.
n is the number of 3D points.
P_i is the camera projection matrix for the image I_i.
M_j is the jth reconstructed 3D point.
m_{ij} is the *j*th point in the image $I_{i.}$
$\varphi(P_i, M_j)$: the projection of the 3D point M_j in image I_i.

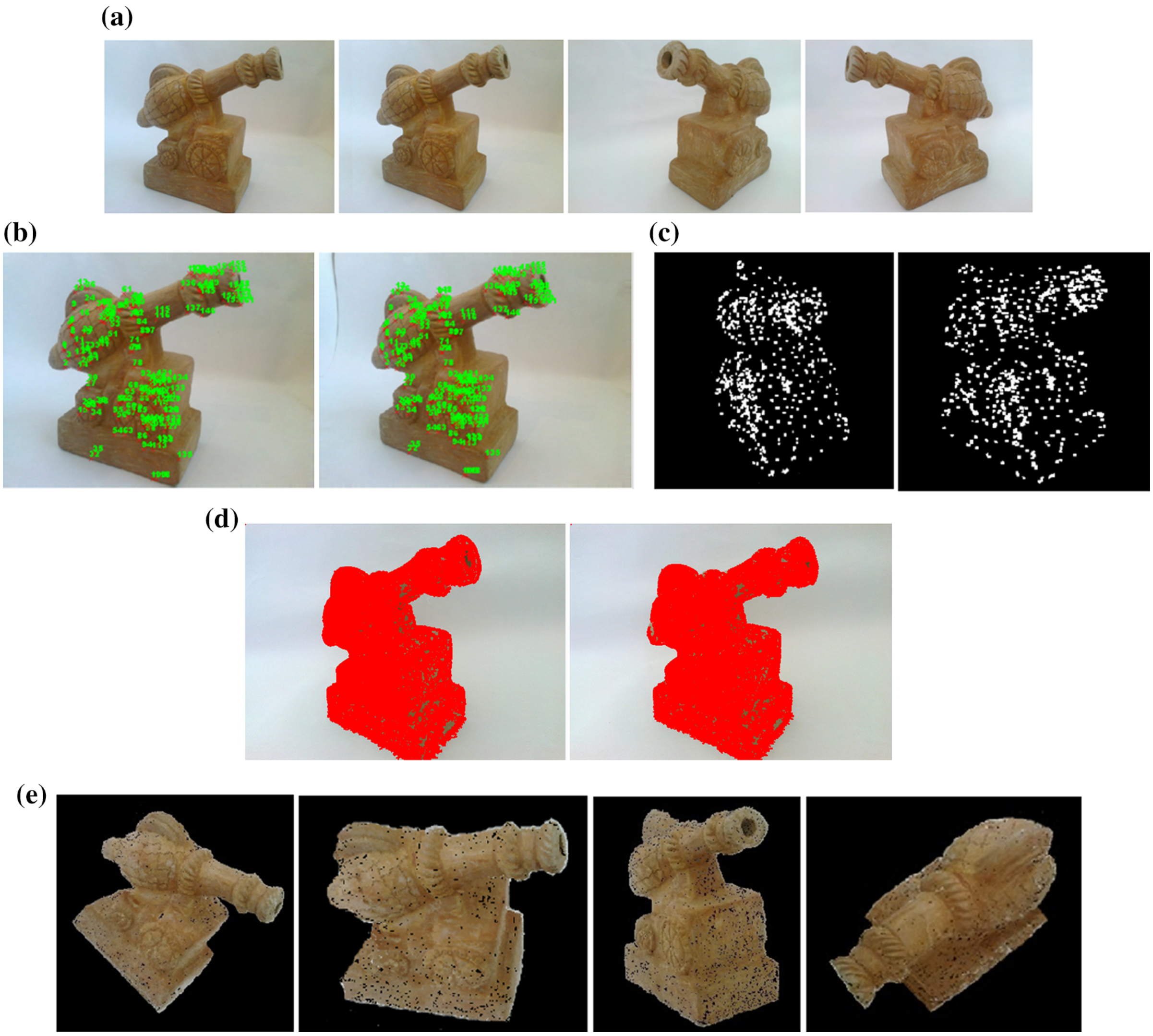


Fig. 10 **a** Four images of the sequence. **b** Interest points matching. **c** Sparse 3D reconstruction result.**d** Matching result after the application of the match propagation method. **e** Four views of the obtained 3D model

5.2.1 Real Sequence 1

In this first experiment we used a sequence of eighteen images of an object of complex shape, the two first are taken by the binocular stereo vision system and the others are captured by the second camera which undergoes displacements of about twenty degrees to capture every time a new image and finally make a complete turn around the object, in order to obtain an almost complete 3D reconstruction. Three images of the sequence are presented in Fig. 10. The matching of points of interest (detected by Harris) by NCC correlation measure is presented in the Fig. 10b (the elimination of false matches is made by the RANSAC algorithm [24]). The result of sparse 3D reconstruction of interest points matched between consecutive images ($S = \cup_{k=2}^{n} S_{k-1,k}$) is presented in Fig. 10c.

After the estimation of all projection matrices and the recovery of the matching results of interest points between consecutive images. The match propagation method [3] was applied to increase the density of matches. The result of the execution of this algorithm for a couple of consecutive images is presented in Fig. 10d. From 139 initial matches (seeds) 144,822 matches are detected.

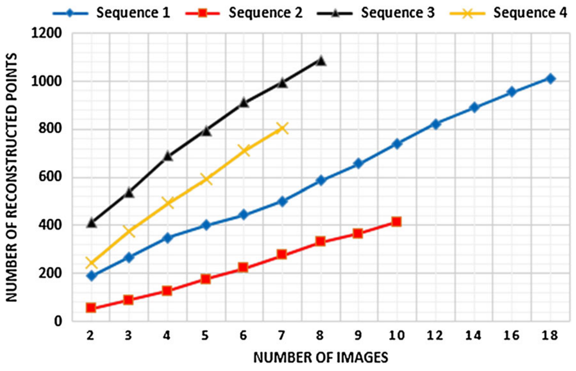


Fig. 11 Progression of sparse 3D reconstruction depending on the number of images for the four sequences

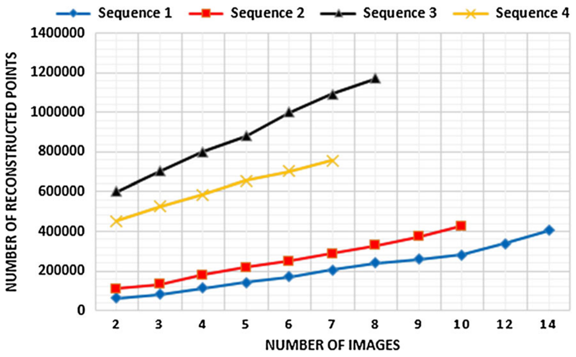


Fig. 12 Progression of dense 3D reconstruction depending on the number of images for the four sequences

The complete 3D model is obtained by the fusion of 3D reconstruction results between the consecutive images. The Fig. 10e presents four views of the 3D model obtained.

Visually, the result of the reconstruction presented in Fig. 10e shows the quality and high-density of the reconstructed 3D points, a fact that shows the power of the proposed approach for 3D reconstruction of objects from a small number of images taken from suitable viewpoints. The use of match propagation algorithm for all couples of consecutive images (calibrated images), allowed us to increase the density of matches (to avoid the false matches, epipolar constraint was used). The fusion of 3D reconstruction results between pairs of consecutive images allowed us to have dense 3D models.

The Fig. 12 shows the progression of dense 3D reconstruction of the different image sequences. Table 3 presents the result of dense 3D reconstruction of the different image sequences. The small value of the reconstruction error shows the quality of obtained reconstruction by the proposed approach (the match propagation algorithm is applied to reliable initial matches and the use of epipolar geometry). The big number of reconstructed points indicates the high density of obtained 3D models (because of the use of match propagation algorithm and the fusion of results between pairs of consecutive images).

Table 3 Results of dense 3D reconstruction of different sequence after the application of the match propagation method and the fusion between different results

Resolution	Number of reconstructed points	Reprojection error
Sequence 1		
640 × 480	512,087	0.41
Sequence 2		
800 × 600	426,123	0.63
Sequence 3		
960 × 1280	1,172,822	0.57
Sequence 4		
960 × 1280	755,468	0.73

5.2.2 *Real Sequence 2*

In this second experiment, the power of our approach is tested for 3D face reconstruction. A problem that attracts a lot of interest in itself.

A sequence of ten images taken from suitable viewpoints was used. Four images of the sequence are presented in Fig. 13a. The matching of interest points (detected by Harris) by NCC correlation measure is presented in Fig. 13b. For the elimination of false matches, RANSAC algorithm [24] is used. The result of sparse 3D reconstruction of interest points matched between consecutive images is presented in Fig. 13c. The quasi-dense matching result obtained by Match propagation method is presented in Fig. 13d. Figure 13e presents four views of the 3D reconstruction result.

To make the 3D reconstruction and/or the 3D face recognition, many works are based on the use of a binocular stereo vision system [15, 31]. However, as it is indicated in the Fig. 14, the 3D reconstruction result from two stereo images presents multiple areas not reconstructed (the presence of a lot of holes) because of occlusions. The proposed method doesn't use a binocular stereo vision system unless to initialize the

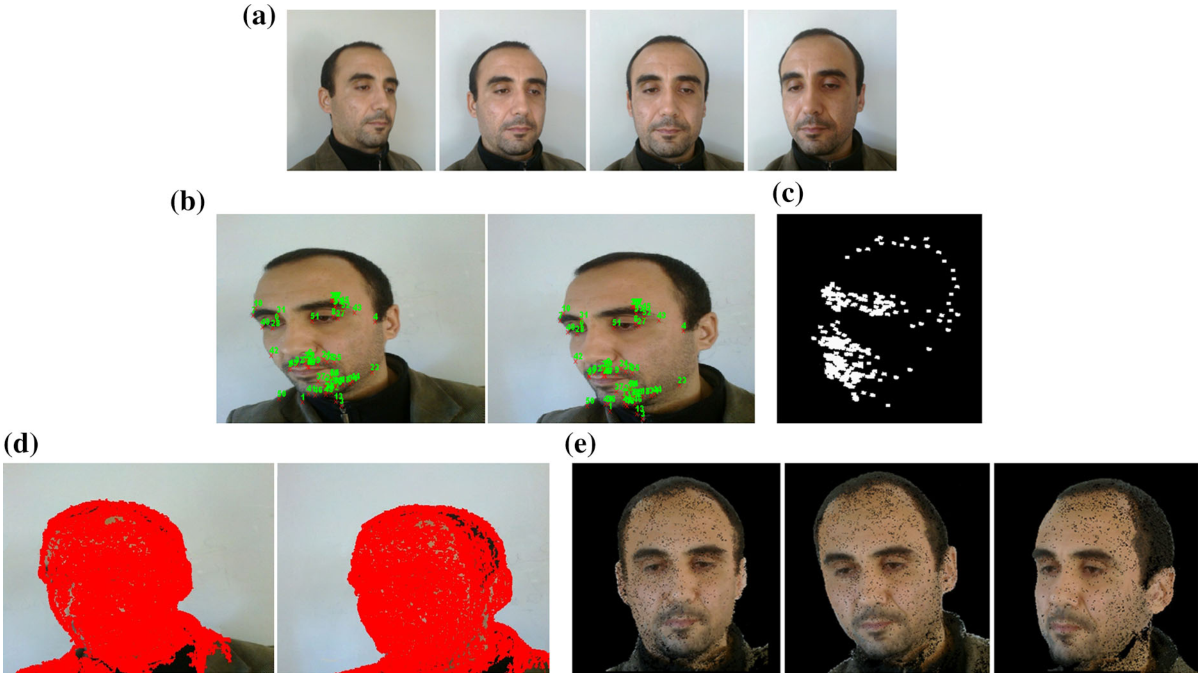


Fig. 13 **a** Four images of the sequence. **b** Example of interest points matching. **c** Result of sparse 3D reconstruction. **d** Matching result after applying the match propagation method. **e** Three views of the obtained 3D model

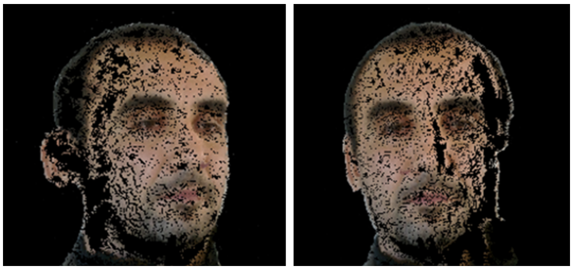


Fig. 14 Two views of the results of the 3D reconstruction by using a binocular stereovision system

3D reconstruction process and gradually reconstruct new areas from other images captured from suitable viewpoints in order to obtain an almost complete 3D reconstruction results. The Fig. 13e shows the obtained reconstruction result by our approach.

5.2.3 *Real Sequence 3*

In this third experiment, we would like to make the reconstruction of all the captured part of a medium-sized scene. A sequence of eight images of a traditional door taken from suitable viewpoints has been used. Four images of this sequence are presented in the Fig. 15a. The matching of interest points (detected by Harris) by NCC correlation measure is presented in the Fig. 15b (the elimination of false matches is made by RANSAC algorithm). The sparse reconstruction result is presented in the Fig. 15c. To get a dense 3D reconstruction, the match propagation method is applied to the consecutive image pairs, an example is presented in the Fig. 15d. In the Fig. 15e, three views of obtained 3D model after the fusion of the obtained results between consecutive images.

5.2.4 *Real Sequence 4*

In this forth experiment, we would like to make the reconstruction of large-sized scene. A sequence of seven images of a house taken from suitable viewpoints has been used. Four images of this sequence are presented in the Fig. 16a. The sparse matching of interest points is presented in the Fig. 16b. The sparse 3D reconstruction result is presented in the Fig. 16c. The quasi-dense matching result obtained after applying the match propagation method is presented in the Fig. 16d. In the Fig. 16e, three views of obtained 3D

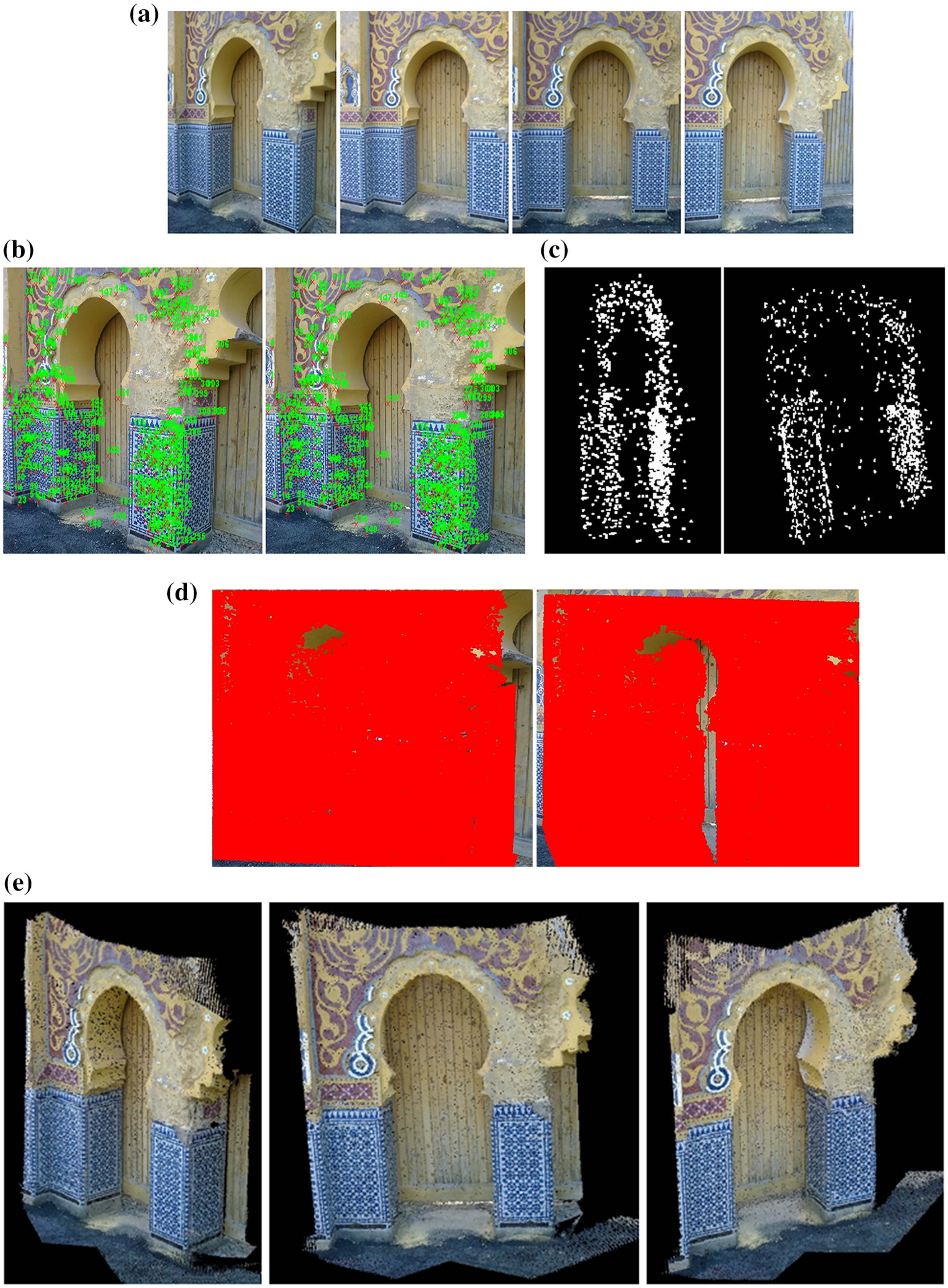


Fig. 15 **a** Four images of the sequence. **b** Example of interest points matching. **c** Result of sparse 3D reconstruction. **d** Matching result after applying the match propagation method. **e** Three views of the obtained 3D model

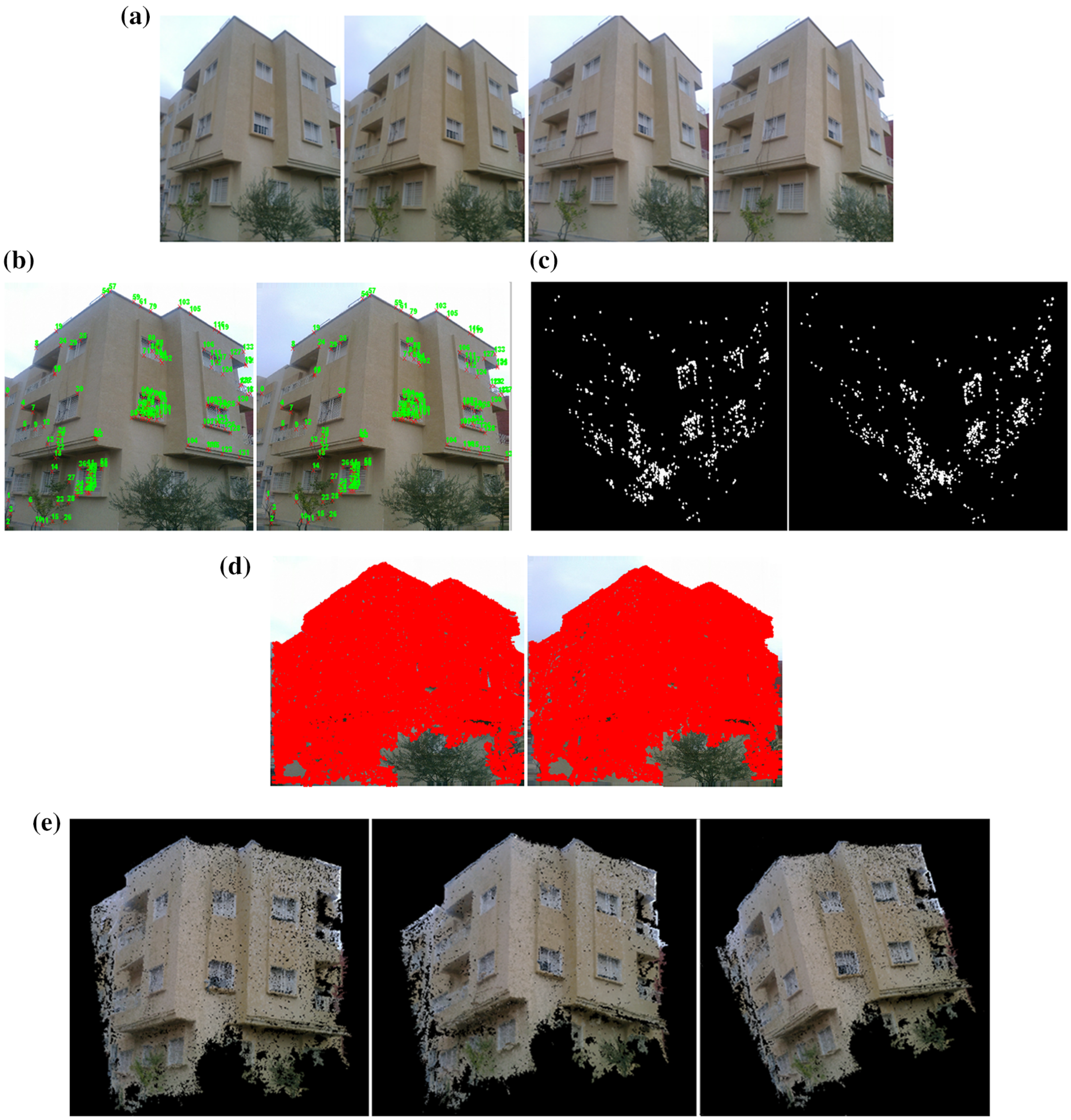


Fig. 16 **a** Four images of the sequence. **b** Example of interest points matching. **c** Result of sparse 3D reconstruction. **d** Matching result after applying the match propagation method. **e** Three views of the obtained 3D model

model after the fusion of the obtained results between consecutive images.

6 Conclusion

In this paper, we presented an incremental method for 3D reconstruction from multiple images taken from suitable viewpoints. First, we proposed the initialization of the reconstruction process by a binocular stereo vision system constituted of two unattached cameras in order to avoid the initialization errors that can affect the globality of the reconstruction system. Afterwards, for an almost complete 3D reconstruction and to release all the constraints, the second camera of the binocular stereo vision system (characterized by varying parameters) makes displacements around the object or scene to capture new images which are gradually inserted into our system. For each inserted image, the projection matrix is estimated from the 3D structure already calculated and new 3D points are recuperated. For the refinement of new estimated entities, a local bundle adjustment is performed so as to maintain the reliability and ensure the system rapidity. Finally, to obtain a dense 3D reconstruction, match propagation algorithm has been applied to consecutive images pairs and the final 3D model is obtained by fusion. The experimentation results show the power the robustness and the rapidity of the proposed approach.

References

1. Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(11), 1330–1334.
2. El akkad, N., Merras, M., & Satori, K. (2014). Camera self-calibration with varying intrinsic parameters by an unknown three-dimensional scene. *The Visual Computer, 30*(5), 519–530.
3. Lhuillier, M., & Quan, L. (2002). Match propagation for image-based modeling and rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(8), 1140–1146.
4. Pollefeys, M., Koch, R., Vergauwen, M., & Van Gool, L. (2000). Automated reconstruction of 3D scenes from sequences of images. *ISPRS Journal of Photogrammetry and Remote Sensing, 55*(4), 251–267.
5. Lhuillier, M., & Quan, L. (2005). A quasi-dense approach to surface reconstruction from uncalibrated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(3), 418–433.
6. Furukawa, Y., & Ponce, J. (2010). Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(8), 1362–1376.
7. El hazzat, S., Saaidi, A., Satori, K. (2014). Multi-view passive 3D reconstruction: comparison and evaluation of three techniques and a new method for 3D object reconstruction. In *Fifth International Conference on Next Generation Networks and Services* (NGNS) (pp. 194–201).
8. Liu, Y., Cao, X., Dai, Q., Xu, W. (2009). Continuous depth estimation for multi-view stereo. In *CVPR* (pp. 2121–2128).
9. Vogiatzis, G., Esteban, C. H., Torr, P. H. S., & Cipolla, R. (2007). Multi-view stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(12), 2241–2246.
10. Goesele, M., Curless, B., Seitz, SM. (2006). Multi-view stereo revisited. In *IEEE Proceedings of CVPR* (pp. 2402–2409).
11. El hazzat, S., Saaidi, A., & Satori, K. (2014). Euclidean 3D reconstruction of unknown objects from multiple images. *Journal of Emerging Technologies in Web Intelligence, 6*(1), 59–63.
12. Harris, C., Stephens, M. (1988). A combined corner and edge detector. In *Fourth Alvey vision Conference* (pp. 147–151).
13. Pollefeys, M., Gool, L. V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., & Koch, R. (2004). Visual modeling with a hand-held camera. *International Journal of Computer Vision, 59*(3), 207–232.
14. Wang, G., & Wu, Q. M. J. (2009). Perspective 3-D Euclidean reconstruction with varying camera parameters. *IEEE Transactions on Circuits and Systems for Video Technology, 19*(12), 1793–1803.
15. Da, F., & Sui, Y. (2011). 3D reconstruction of human facebased on an improved seeds-growing algorithm. *MachineVision and Applications, 22*(5), 879–887.
16. Slabaugh, G. G., Culbertson, W. B., Malzbender, T., Stevens, M. R., & Schafer, R. W. (2004). Methods for volumetric reconstruction of visual scenes. *International Journal of Computer Vision, 57*(3), 179–199.
17. Seitz, S., & Dyer, C. (1999). Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision, 35*(2), 151–173.
18. Culbertson, WB., Malzbender, T., Slabaugh, GG. (1999). Generalized voxel coloring. In *ICCV Proceedings of Workshop, Vision Algorithms Theory and Practice, Springer-Verlag Lecture Notes in Computer Science 1883* (pp. 100–115).
19. Kutulakos, K. N., & Seitz, S. M. (2000). A theory of shape by space carving. *International Journal of Computer Vision, 38*(3), 199–218.
20. Mulayim, A. Y., Yilmaz, U., & Atalay, V. (2003). Silhouette-based 3-D model reconstruction from multiple images. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics, 33*(4), 582–591.
21. Mouragnon, E., Lhuillier, M., Dhome, M., Dekeyser, F., & Sayd, P. (2009). Generic and real-time structure from motion using local bundle adjustment. *Image and Vision Computing, 27*(8), 1178–1193.
22. Zhang, Z., Shan, Y. (2003) Incremental motion estimation through modified bundle adjustment. In *IEEE Proceedings of international conference on image processing* 2 (pp. 343–346).
23. Seitz, SM., Curless, B., Diebel, J., Scharstein, D., Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *IEEE Proceedings of the conference on computer vision and pattern recognition* (pp. 519–528).
24. Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM, 24*(6), 381–395.
25. Yuan, Z. H., & Lu, T. (2013). Incremental 3d reconstruction using bayesian learning. *Applied Intelligence, 39*(4), 761–771.
26. Wang, Y., Liu, K., Hao, Q., Wang, X., Lau, D., & Hassebrook, L. (2012). Robust active stereo vision using kullback-leibler divergence. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(3), 548–563.

27. Li, Y. F., & Lu, R. S. (2004). Uncalibrated euclidean 3-D reconstruction using an active vision system. *IEEE Transactions on Robotics and Automation, 20*(1), 15–25.
28. Chambon, S., & Crouzil, A. (2011). Similarity measures for image matching despite occlusions in stereo vision. *Pattern Recognition, 44*(9), 2063–2075.
29. Moré, J. J. (1977). The Levenberg–Marquardt algorithm: Implementation and theory. In G. A. Watson (Ed.), *Numerical Analysis, Lecture Notes in Mathematics* (Vol. 630, pp. 105–116). Heidelberg: Springer.
30. Fitzgibbon, AW., Zisserman, A. (1998). Automatic camera recovery for closed or open image sequences. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 311–326).
31. Uchida, N., Shibahara, T., Aoki, T., Nakajima, H., Kobayashi, K. (2005). 3-D face recognition using passive stereo vision. In *Proceeding of IEEE Int Conf Image Process* 2 (pp. 950–953)