J. Vis. Commun. Image R. 39 (2016) 120–131

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Simultaneous reconstruction and calibration for multi-view structured light scanning ☆

S. Garrido-Jurado, R. Muñoz-Salinas *, F.J. Madrid-Cuevas, M.J. Marín-Jiménez

Computing and Numerical Analysis Department, and Maimonides Institute for Biomedical Research (IMIBIC), Edificio Einstein, Campus de Rabanales, Córdoba University, 14071 Córdoba, Spain

ARTICLE INFO

Article history:
Received 9 December 2015
Revised 22 March 2016
Accepted 23 May 2016
Available online 26 May 2016

Keywords:
Structured light
Calibration
Multi-view
3D reconstruction

ABSTRACT

Structured light 3D scanning from a single viewpoint requires multiple scans and a registration process for a complete description of the scanned object. Instead, using multiple cameras and projectors simultaneously can reduce the scanning time and increase the visibility. However, a precise estimation of the extrinsic parameters of all the components is a time consuming process prone to errors. This paper proposes a method to automatically reconstruct and self-calibrate multi-view structured light systems with an arbitrary number of devices. The experimentation shows that the proposed method is precise and robust, surpassing other current state of the art approaches. The achieved calibration accuracy is similar to that obtained by a traditional chessboard pattern calibration, but being able to adapt to a wider range of situations.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Reconstruction of three-dimensional objects is an important topic in many computer vision applications such as industrial inspection [1,2], cultural heritage recording [3,4] or anthropometric analysis [5,6]. Among the different techniques to obtain depth information from a scene, structured light is very popular since it allows obtaining precise measures within a reduced cost, i.e. only conventional cameras and projectors are required [7,8].

The simplest structured light systems are the single-view ones, which are comprised by a camera-projector pair. Structured light patterns are projected over the scene and identified in the camera image, producing a set of camera-projector correspondences that are triangulated to reconstruct the scene. The range of the scanned scene is limited to the intersection of the camera and projector field of views, thus, points not illuminated by the projector nor observed by the camera cannot be reconstructed.

In order to measure a complete object (e.g. a 360 degrees scan) single-view systems require to move the object (or the system) so that it is visible from many view-points and then apply a registration technique to join the scans [9,10]. On the contrary, the approach explored in this paper, Multi-View Structured Light Systems (MVSLS), consists in employing several cameras and projectors. This helps to simplify the scanning process avoiding to move the system or the object, which is of great importance in applications such as human body scanning (Fig. 1).

One of the main downsides of MVSLS is the calibration of its devices. While calibration of single-view systems is relatively simple, the process becomes more complex and tedious as the number of devices increases.

In the recent years, there has been an effort to develop multi-view calibration methods for MVSLS, however, in most cases, they require manual intervention [11], special equipment [12] or are constrained to specific conditions [13–15] such as specific device arrangements or limitations in the number of devices. Some of these proposals aim to calibrate both intrinsic and extrinsic parameters simultaneously. Intrinsic parameters are inherent to the camera and projector devices and, thus, in many applications they only need to be estimated once, unless the lens focus is modified. In practice, extrinsic parameters or poses are more likely to be modified during the day-to-day manipulation, either intentionally (e.g. device displacements to cover a specific area) or unintentionally (e.g. accidental hits or weak device supports).

This paper proposes a method to automatically reconstruct and self-calibrate the extrinsic parameters of MVSLS devices (both cameras and projectors) without being restricted to any special arrangement and without requiring any special equipment, thus avoiding a tedious calibration step each time a device is moved. The proposed method starts by sequentially projecting the light

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail addresses: i52gajus@uco.es (S. Garrido-Jurado), rmsalinas@uco.es (R. Muñoz-Salinas), fjmadrid@uco.es (F.J. Madrid-Cuevas), mjmarin@uco.es (M.J. Marín-Jiménez).

http://dx.doi.org/10.1016/j.jvcir.2016.05.014
1047-3203/© 2016 Elsevier Inc. All rights reserved.

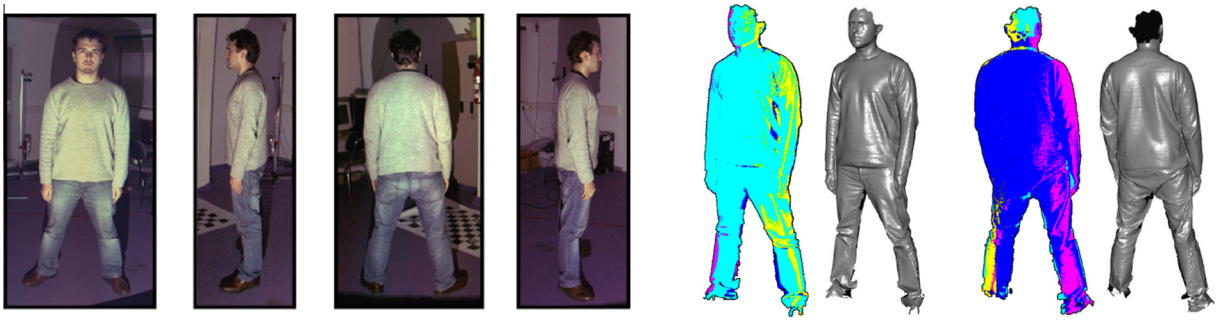

Fig. 1. Example of 360 degrees reconstruction of a person employing a MVSLS composed by 4 cameras and 4 projectors. Figure shows real camera images, colored set of point clouds of each camera-projector pair and the final complete mesh reconstruction. The MVSLS approach allows to perform the scan without moving the system or the scanned object. (Best viewed in color.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

patterns (one projector after another) while all the cameras capture them simultaneously. Then, the intersecting field of views between the devices are automatically determined in order to find the optimal path connecting them. To that end, a Mixed Integer Linear Programming (MILP) model has been proposed which finds the optimal path based on the image correspondences between devices. MILP is an optimization technique which allows finding the optimum of a mathematical model represented by linear relationships and where some unknowns are constrained to be integers. Then, our method starts the process of determining the extrinsic parameters of all devices by following the estimated path. Finally, the three-dimensional points and extrinsic estimations are further refined using Sparse Bundle Adjustment (SBA) [16].

The rest of the paper is structured as follows. Section 2 reviews related work. Section 3 presents the basics behind structured light techniques. Section 4 details the proposed calibration method. Finally, Section 5 presents the experimentation carried out, and Section 6 draws some conclusions.

2. Related work

In general, a single-view structured light scanning cannot scan a whole scene or object since it can only provide one viewpoint. Some typical examples of these situations are 360 degrees scans or large scene reconstructions, such a whole room. To address these situations, some authors have opted for keeping a single-view scanner but employing some special equipment that permits the complete scan. An usual solution is a computer-controlled turntable [9,10] which rotates the object while a single-view scanner takes several scans on different angles, allowing a 360 degrees scan. The main problems are the extra cost of the equipment, the size limitation of the scanned object and the additional scanning time. Furthermore, the system is limited to 360 degrees scans, not allowing other types of multi-view scans. An alternative to turntables is the use of computer-controlled robotic devices [17] that can move the single-view scanner to different viewpoints while, at the same time, tracking the scanner positions. This approach can adapt to different ranges of multi-view scans, nevertheless, it involves a high cost of the precise equipment and it is still limited to the device scope.

Some approaches obtain various single-view reconstructions from different viewpoints and then integrate them. In general, this is a complicated task which usually requires manual intervention to provide an initial registration. An alternative is the use of synthetic markers to easily perform an automatic estimation of the scanner positions [18–20]. The main problems of these approaches are that the scene needs to be manipulated to attach the markers and, furthermore, the single-view scanner has still to be moved to obtain the reconstructions from different viewpoints.

In [21], a mirror system is proposed to obtain complete reconstructions of an object using a single-view scanner. The projected patterns are reflected by the set of mirrors, simulating additional virtual scanners that compose a multi-view scanner. The main benefit is that the complete reconstruction is performed using a single scan, reducing the scanning time. However, the method applicability is very restrictive, since the object size is limited to the working area surrounded by the set of mirrors, which also needs to be calibrated a priori. Furthermore, the method requires Fresnel lens which are considerably expensive.

As an alternative to the previous methods, some proposals employ a higher number of cameras and projectors, i.e. a MVSLS. These systems can also be expensive if using a high amount of devices but, contrary to previous alternatives, they can adapt to different types of scenes or objects by simply reorienting the devices to the desired areas. However, they also present some drawbacks. First, scanning with multiple projectors simultaneously is complex because of the interferences generated on the overlapped surfaces. As a consequence, many authors employ the projectors sequentially. Second, the calibration process is more complex due to the higher number of devices.

The simplest solution for extrinsic calibration is to use a calibration pattern, e.g. a chessboard pattern, which is visible by all devices, and perform the cameras and projectors calibration based on that pattern, such as [22]. However, this solution is not feasible in most cases, since each device view is usually covering different scene areas and, thus, a pattern cannot be observed by all devices simultaneously.

Some of the most basic proposals for multi-view structured light calibration require manual intervention or special equipment. In [11], a manual method for multi-view calibration is presented. First, the camera parameters are calibrated independently using the method in [23], which employs a led or laser pointer projected over the scene. Then, the calibration of the projectors is performed by projecting squared fiducial markers [24] while the user moves a white surface in front of them. The previously calibrated cameras triangulate the marker corners to obtain their 3D position which are finally used to calibrate the projectors. The main drawback is that the system requires manual operation, which in many cases can be tedious and inconvenient. Furthermore, if any of the devices is moved, the calibration process needs to be repeated.

In [12], a complete MVSLS is calibrated using a special translucent sheet. The system is limited to devices that surround the scene, i.e. 360 degrees scans. The patterns are projected over the sheet and, thanks to the translucent material, they are observed by all the cameras, independently of their position, allowing establishing correspondences between all the devices. The main drawback is that the translucent sheet needs to be moved manually

during the calibration process. Moreover, the system is limited to 360 degrees scans.

In practice, automatic calibration methods are more convenient. However, whereas there have been proposed numerous techniques that address the automatic calibration of single-view structured light systems [25–30], the number of proposals for MVSLS is limited and most of them are restricted to some specific conditions, such as specific number of devices or device arrangement constraints.

In [13] an automatic self-calibration for MVSLS is proposed. However, it is restricted to trinocular systems, i.e. systems composed by only two cameras and one projector, and assumes there is not camera distortion, which is a strong constraint. In [31], a 3D reconstruction method for trinocular systems is proposed, including the automatic calibration of the projector. Besides being restricted to trinocular systems, the cameras need to be calibrated a priori.

In [14], an iterative method to automatically scan and recalibrate pose parameters is proposed. On each iteration, the method allows one device (camera or projector) being moved to a new position, so that after several iterations the complete scene can be reconstructed while the pose on each new position is estimated automatically. In [32], the same method in [14] is redefined applying a modified version of Bundle Adjustment as the last step, obtaining a complete system calibration, including intrinsic and extrinsic parameters. However, it assumes that initial values for the intrinsic parameters are known a priori with some level of precision and only the focal lengths are redefined during the process.

In [33,15], an automatic calibration method for MVSLS is presented. The authors employ photometric techniques (which obtain the surface normals from different lighting conditions) to determine an initial pose of the structured light system. Thus, the calibration depends significantly on the results of the photometric technique used, being constrained to the limitations of these types of techniques.

In this work, a generic method for automatic self-calibration of MVSLS is presented. Our method determines the extrinsic parameters of the system elements without imposing restrictions on the number of devices or their arrangement and without requiring any special equipment or manual intervention. Since the proposed method is only based on the correspondences between different devices, it can perform calibration and reconstruction simultaneously. The method consists in an iterative process which is optimized using a proposed MILP model. Finally, Bundle Adjustment is applied to refine the obtained extrinsic parameters. The results show that the method is precise and robust and, due its flexibility, it is suitable for measuring a wide range of scenes and objects.

3. Structured light fundamentals

Structured light is an active stereo method, in which one or several patterns are projected onto the scene in order to label its surface. Usually, a projector is employed to illuminate the scene while a camera captures the sequence. Binary codes provide the highest robustness to noise and precision by using two illumination levels (black and white) that defines 2^m stripes encoded by a sequence of m images. Each pattern of the sequence subdivides the scene into successively smaller regions. Encoding in a single direction (vertical or horizontal) allows plane-to-ray triangulation, however, both directions are required to uniquely label the scene points as our method requires. Eventually, every projector pixel has a unique codeword formed by a sequence of zeros and ones. Inokuchi et al. [34] improved this codification scheme with the Gray code, in which consecutive codewords have a Hamming distance of one. In Fig. 2, the Gray coded patterns can be observed.

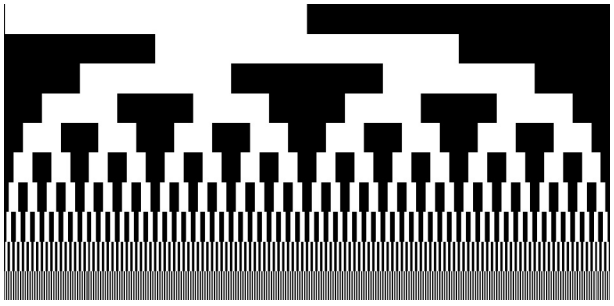

Fig. 2. Gray code pattern sequence employed in structured light systems.

By decoding the sequence, it is possible to determine for each illuminated image pixel $\mathbf{x_i} \in \mathbb{R}^2$, the corresponding projector pixel $\mathbf{u_j} \in \mathbb{R}^2$ it was projected from. In practice, camera resolution is higher than the projector resolution, so a projector pixel corresponds to multiple camera pixels. Nevertheless, it is possible to reduce this ambiguity by analyzing only the pattern stripe transitions instead of the whole stripe. In this manner, it is possible to determine the camera pixel with subpixel accuracy [35].

One of the problems of MVSLS is the possible interference between the patterns projected by different projectors. The simplest solution consists in projecting the patterns sequentially, so that the interferences are avoided at the expense of a slower scanning. There also exist some alternative codifications to simultaneously perform multi-view scanning, [36,37], but in general, the results are not as precise as these of standard binary codification techniques.

It must be remarked that the proposed method is suitable for different codifications methods provided that they obtain correspondences between the different devices. Nevertheless, binary codification has been employed in our experiments due to its robustness and high precision.

3.1. Pinhole camera model

The pinhole camera model [38] defines the relationship between the 3D location of a scene point $\mathbf{p} \in \mathbb{R}^3$ and its projection $\mathbf{x}$ in a image as:

$$\lambda \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = K \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} T \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix}, \tag{1}$$

where λ is a scale factor, K is known as the intrinsic camera matrix and T is the extrinsic camera matrix. The T matrix defines the 3D rotation and translation required to move a scene point from a global coordinate system to the camera coordinate system.

Camera calibration is the process of determining the intrinsic and extrinsic matrices, which is easily performed using a calibration pattern. A projector can also be though as an inverse camera, in which pixels are projected out of the camera sensor instead of on the camera sensor. However, calibration is a more complicated task since requires the use of an external camera. In this work, the calibration method proposed in [22] has been employed.

While the intrinsic parameters remain fixed as long as the optical lenses are not modified (since they are inherent to the camera), extrinsic parameters vary when camera or projector are moved.

3.2. Three-dimensional reconstruction

Given a pair of calibrated cameras or camera-projector, recovering the three-dimensional location of a point $\mathbf{p}_k \in \mathbb{R}^3$ projecting in

locations $\mathbf{x}_i$ and $\mathbf{u}_j$ is performed by triangulation. In theory, the rays passing through the optical centers of the sensor and the pixels should intersect in space. However, due to errors in the calibration process and in the decoding step, these rays may not intersect in practice. So, the triangulation can be seen as an optimization problem aiming to determine the location $\hat{\mathbf{p}}_k$ that minimizes the reprojection error to the observed locations. The simplest solution to that problem is the mid-point method, consisting in finding the central point of the common perpendicular joining the two rays. A more advanced solution is the iterative linear triangulation method proposed by [39] which performs an iterative minimization of the reprojection error.

4. Multi-view self-calibration and reconstruction

This section explains the basis of the proposed method. Let us denote by

$$\mathcal{C} = \{C_p |\ p = 1, \ldots, N\},$$

the set of cameras in our system and by

$$\mathcal{P} = \{P_q |\ q = 1, \ldots, M\},$$

the set of projectors. Additionally, we shall denote by $\mathbf{x}_i^p$ a pixel in camera C_p and by $\mathbf{u}_j^q$ a pixel in projector P_q. Also, we assume that the camera and projector intrinsic parameters have been determined in a previous calibration step.

Our method can be divided in four stages: acquisition and decoding, initial pose estimation, pose refinement, and reconstruction. In the first stage, the projectors illuminate the scene while all the cameras capture them. Then the camera-projector correspondences are obtained by decoding the patterns. Afterwards, the initial extrinsic components are estimated based on the previously obtained correspondences and employing an iterative process. Thereafter, the initial estimates are refined by a global minimization of the reprojection error with subpixel accuracy, using the SBA algorithm. As a result, a refined estimation of the extrinsic components are obtained and employed to perform the final triangulation. The rest of this section details these processes.

The complete calibration procedure is summarized in Algorithm 1.

Algorithm 1. Automatic calibration process

```
Acquire and decode projected patterns (Section 4.1)
Find correspondences for each device pair and trio
  (Section 4.2.1)
Determine optimal Pose Estimation Sequence (Section 4.2.2).
for each step in Pose Estimation Sequence do
  if first iteration then
    Estimate first device pair poses from Essential Matrix
    Refine new pose with SBA (Section 4.3)
  else
    Find trio of correspondences between the new device and
   the previously calibrated devices
    Triangulate correspondences with the calibrated devices
    Estimate new pose from 3D points using PNP algorithm
    Reproject points to the new device and remove outliers
    Refine new pose with SBA (Section 4.3)
  end if
end for
Refine full calibration with SBA (Section 4.3)
Reconstruct 3D scene (Section 4.4)
```

4.1. Acquisition and decoding

In the first stage, the projectors are commanded to illuminate the scene with the sequence of Gray codes. When a projector is illuminating the scene, the rest of them are set to black to avoid interferences. At the same time, all cameras are capturing the sequence simultaneously. So, a total of $M \times N$ Gray code sequences are obtained, one for each camera-projector pair. When all projectors have finished, the Gray code sequences are decoded. We shall denote by

$$\mathbb{P}(\mathbf{u}_j^q) = \{\mathbf{x}_i^p\}, \tag{2}$$

to the set of camera pixels $\mathbf{x}_i^p$ in which the pixel $\mathbf{u}_j^q$ has been projected. It must be remarked here that the set $\mathbb{P}(\mathbf{u}_j^q)$ contains no duplicated elements, i.e., a projector pixel appears at most once in a camera.

4.2. Initial pose estimation

In this step, the extrinsic parameters of all the devices (cameras and projectors) are estimated respect to a common global reference system. The method is based on an iterative process which can be summarized as follow:

- In the first iteration, the poses of an initial pair of devices are estimated simultaneously employing the essential matrix.
- In subsequent iterations, the pose of an uncalibrated device is obtained *wrt* a pair of previously calibrated devices by means of triangulation.

It must be noted that, in this context, the allusions to calibrated or uncalibrated devices refer only to the extrinsic parameters since intrinsic ones are known a priori. Also, that term *correspondence* refers to the set of observations of a real scene point in several devices (see Section 4.2.1). Fig. 3 shows an example of the order followed to estimate the poses for a system composed by three cameras and two projectors.

As indicated, the poses of two devices are estimated (up to a scale factor) using the essential matrix [38] in the first iteration. First, fundamental matrix is obtained from a set of undistorted point correspondences [40]. In order to avoid instability caused by noisy correspondences, the RANSAC algorithm is employed. Then, the essential matrix is estimated from the fundamental matrix considering also that the intrinsic matrices of both sensors are known. The essential matrix contains information about the rotation and translation between the two sensors, which can be obtained by singular values decomposition. In our work, the obtained solution is further refined by a minimization of the reprojection error using the Levenberg–Marquardt algorithm (LMA) [41,42], an iterative method to solve non-linear least squares problems.

In subsequent iterations, two calibrated devices are employed to calculate the pose of an uncalibrated one. To that end, first, the common correspondences between the three devices are extracted from the projected patterns (see Section 4.2.1). These correspondences are triangulated using the two calibrated devices and the iterative linear triangulation method [39]. Then, from the 3D points and their projections in the uncalibrated device, a non-linear system is obtained from Eq. (1), where only the matrix T is unknown, which represents the pose of the new device. This problem can be solved employing a PNP algorithm [38] which typically consists in applying an optimization method to reduce the reprojection error, such as the Levenberg–Marquardt algorithm (LMA) [41,42].

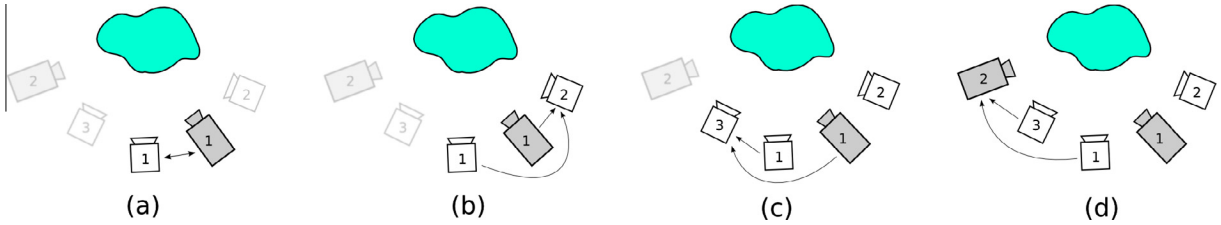

Fig. 3. Example of sequence followed to estimate the devices' poses in a system composed by three cameras (white devices) and two projectors (gray devices). Devices whose pose has not been estimated yet are translucent. (a) Initial step composed by only two devices: camera 1 and projector 1. The poses are obtained from the essential matrix. (b) The pose of camera 2 is obtained by triangulation using camera 1 and projector 1. This process is detailed in Fig. 4. (c) The pose of camera 3 is also obtained from camera 1 and projector 1. (d) Finally, the pose of projector 2 is obtained from cameras 1 and 3.

After the new device pose has been estimated, the 3D points are then projected back to the uncalibrated device using the obtained pose and Eq. (1). Those points with a high reprojection error are considered noise and removed from further processing. Finally, the estimated pose is refined before the next iteration by applying a Bundle Adjustment [43] process to the trinocular subsystem (see Section 4.3).

Fig. 4 illustrates the calibration process of a new device from two previously calibrated ones.

The main benefits of using triangulation (instead of essential matrices as in the first iteration) are two. First, since we are filtering out those correspondences that exceed a minimum epipolar distance and a minimum reprojection error when reprojecting triangulated points, correspondences shared by a higher number of devices have had to accomplish a higher number of requirements and, thus, they are less likely to be erroneous. Second, the new poses are relative to the same scale factor of the initial step, avoiding the necessity of rescaling each new pose to fit the same scale factor.

An important aspect of the proposed method is selecting the order in which devices are processed. Since there are multiple possible paths, it is necessary to establish a criterion to choose one. Our approach considers the number of correspondences between the devices as a key parameter to guarantee the robustness of the estimated poses and finds the optimal path that maximizes this criterion (see Section 4.2.2).

4.2.1. Correspondences estimation

A correspondence between two devices is defined as a pair of pixels, one per device, that correspond to the projection of the same scene point. The set of correspondences between a camera C_p and a projector P_q, let us denote it by $\mathbb{C}(C_p, P_q)$, can be obtained directly from the decodification of the projected code, as it is exposed in Section 4.1,

$$\mathbb{C}(C_p, P_q) = \{(\mathbf{x}_i^p, \mathbf{u}_j^q) \mid \mathbf{x}_i^p \in \mathbb{P}(\mathbf{u}_j^q)\}. \quad (3)$$

The correspondences between two cameras C_p and C_r can also be obtained from the projected codes as the set of projector points observed in both cameras:

$$\mathbb{C}(C_p, C_r) = \{(\mathbf{x}_i^p, \mathbf{x}_j^r) \mid \exists \mathbf{u}_k^q \; s.t. \; \mathbf{x}_i^p, \mathbf{x}_j^r \in \mathbb{P}(\mathbf{u}_k^q)\}. \quad (4)$$

Finally, the correspondences between two projectors can also be obtained from the observation of the cameras. If a camera observes a projection from two different projectors on the same image pixel, a correspondence between the two projectors, P_q and P_s, can be established,

$$\mathbb{C}(P_q, P_s) = \{(\mathbf{u}_i^q, \mathbf{u}_j^s) \mid \exists \mathbf{x}_k^p \; s.t. \; \mathbf{x}_k^p \in \mathbb{P}(\mathbf{u}_i^q), \mathbf{x}_k^p \in \mathbb{P}(\mathbf{u}_j^s)\}. \quad (5)$$

In practice, since subpixel refinement is employed in the decoding process, it is highly unlikely that two projections coincide in the same camera subpixel. For this reason, subpixel values are truncated to their nearest discrete values to increase the number of projector–projector correspondences. It must be noted that truncation is only done at this step and that subpixel information is employed in the calibration and reconstruction processes in order to achieve a high precision.

Once the correspondences have been estimated, they are filtered out based on the epipolar geometry using the RANSAC algorithm [44]. Correspondences farther from its corresponding epipolar lines than a certain threshold are removed from further processing.

Estimating the pose of a new device requires the common correspondences of a trinocular subsystem (composed by the uncalibrated device and the two calibrated ones). Let us define the set of correspondences between three devices, X, Y and Z, as the set of trios composed by the correspondent pixels on each device,

$$\mathbb{C}(X, Y, Z) = \{(\mathbf{a}, \mathbf{b}, \mathbf{c}) | (\mathbf{a}, \mathbf{b}) \in \mathbb{C}(X, Y), (\mathbf{a}, \mathbf{c}) \in \mathbb{C}(X, Z)\}. \quad (6)$$

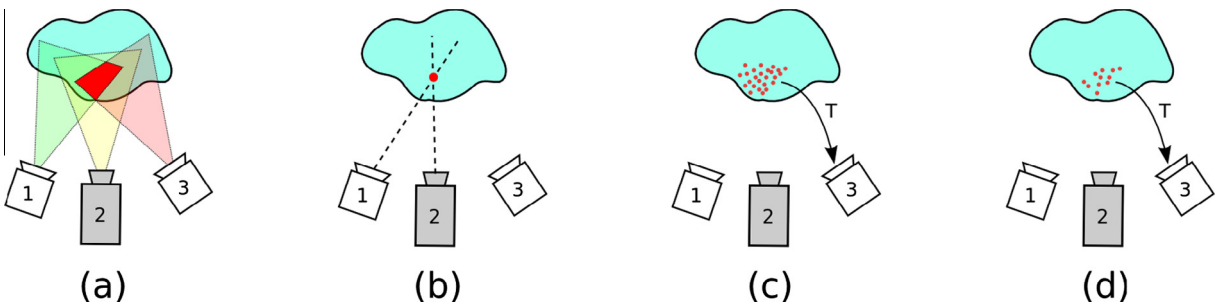

Fig. 4. Pose estimation of a new device (device 3) using the calibrated devices 1 and 2. (a) Common correspondences to the three devices (overlapping area colored in red) are extracted from the projected pattern. (b) Correspondences are triangulated using the calibrated devices. (c) The pose of device 3 (T in Eq. (1)) is obtained by minimization of the reprojection error (PNP algorithm). (d) Correspondences with a high reprojection are removed from further processing and the final pose is refined applying a Bundle Adjustment process. (Best seen in color.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2.2. Pose estimation sequence

In order to apply the iterative method for pose estimation, it is necessary to determine the order in which devices are processed. We have proposed a MILP model to automatically determine the optimal device path and which is based on the number of employed correspondences.

In the first iteration (based on the essential matrix), only the correspondences between pairs are considered. For the rest, the number of common correspondences between the trinocular subsystems (Eq. (6)) is considered.

A first approximation is selecting the path maximizing the total number of correspondences. However, this approach could produce unbalanced solutions, i.e., devices with a high number of correspondences and others with few. Although the sum of correspondences is high, those devices with few correspondences could produce unreliable poses, destabilizing the whole process. Instead, the criterion proposed in this work is to select the path that maximizes the minimum number of correspondences between all the devices in the path.

For a small system, an exhaustive search of all possible combinations could be feasible. However, when the amount of devices increases, the number of combinations grows exponentially. For a system composed by N devices, the number of pair combinations for the first iteration is $\binom{N}{2}$ and the number of combinations in any subsequent iteration is $(N-i)*\binom{i}{2}$, being i the number of devices already calibrated. Then, the total number of valid combinations for the complete sequence is $\binom{N}{2}\prod_{i=2}^{N-1}(N-i)*\binom{i}{2} = 2^{-(N-1)}N!(N-1)!(N-2)!$ For instance, using 12 devices, the total number of combinations is higher than $3\cdot 10^{19}$. For this reason, it is desirably to apply an optimization method that speeds up the convergence to the optimal solution.

One might think of modeling the problem as a graph where nodes are devices and edges are weights proportional to the number of correspondences. This approach, however, is not suitable (except for the first iteration), because the pose estimations are based on device trios, instead of pairs. A solution could be considering nodes as trios and edges weights to be proportional to number of correspondences. Nevertheless, it requires dealing with special constraints, e.g., a device pose should be estimated only once, a device can only be used for triangulation once its pose is already estimated. Because of these complex restrictions, there is not any graph optimization algorithm that fits to our problem, up to our knowledge. Instead, this work proposes a MILP formulation [45] to obtain the optimal path. MILP methods can achieve the optimal solution of a mathematical model represented by linear relationships, where some unknowns are constrained to be integers.

Let us assume that our system is composed by N generic devices (cameras or projectors) and that we have calculated the number of common correspondences for each device pair and device trio (see Section 4.2.1). We denote by D_i the i-th device in our system.

Each possible iteration of the solution is represented by a binary decision variable in the MILP model. $S_{i,j}$ binary variables indicate that the initial pair of devices is composed by the devices D_i and D_j, while $T_k^{i,j}$ binary variables indicate that device D_k pose is obtained from the trinocular subsystem composed by itself and the calibrated devices D_i and D_j. Furthermore, we define the variable $\mathbb{C}^{min}$ which represents the minimum number of correspondences for any step of the path and which is the value we want to maximize. Finally, we also define the auxiliary variables L_i, one per device, which indicate the order in which poses are estimated and which are used to avoid inner loops in the generated solution.

Then, the proposed MILP model aims at maximizing the minimum number of correspondences, $\mathbb{C}^{min}$ as:

$$
\begin{aligned}
&\text{maximize } \mathbb{C}^{min}\\
&\text{Subject to}\\
&\text{(I)} \quad \sum_{i=1}^{N}\sum_{j=1}^{N} S_{i,j} = 1.\\
&\text{(II)} \quad \sum_{i=1}^{N} S_{i,k} + \sum_{i=1}^{N}\sum_{j=1}^{N} T_k^{i,j} = 1 \quad \forall k \in [1,N].\\
&\text{(III)} \quad \sum_{i=1}^{N} \mathbb{C}(D_i,D_k) S_{i,k} + \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbb{C}(D_i,D_j,D_k) T_k^{i,j} \geqslant \mathbb{C}^{min} \quad \forall k \in [1,N].\\
&\text{(IV)} \quad L_k \geqslant L_i + 1 - N\left(1 - \sum_{j=1}^{N} T_k^{i,j}\right) \quad \forall i,k \in [1,N],\ i \neq k.\\
&\text{(V)} \quad 1 \leqslant L_i \leqslant N \quad \forall i \in [1,N].
\end{aligned}
\tag{7}
$$

Constraint I guarantees that only one initial step is selected while constraint II makes sure that each device pose is estimated only once, either by the initial step or by triangulation in a following step. Constraint III guarantees that the number of correspondences involved in the estimation of each device pose is equal to or higher than $\mathbb{C}^{min}$. Constraint IV avoids inner loops in the generated solution by making sure that if device D_k pose is obtained using device D_i, the value of L_k has to be higher than L_i. Finally, constraint V defines the bounds of the L_i variables.

The problem can have several optimal solutions. In that case, the one with the highest sum of correspondences in the whole process is selected. To that end, after the previous optimization, a second MILP optimization process is performed using a model that only differs from the previous one in two details. First, the variable $\mathbb{C}^{min}$ is replaced in constraint III by the value obtained in the earlier optimization. Second, the objective function is replaced by:

$$
\text{maximize} \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbb{C}(D_i,D_j) S_{i,j} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} \mathbb{C}(D_i,D_j,D_k) T_i^{j,k} \tag{8}
$$

so that the total number of correspondences is maximized. The solution of this second model is used to estimate the initial device poses as previously described.

Table 1 shows an example of pose estimation sequences for a system composed by 5 devices. The number of correspondences for each device pair and trio are presented in Table 1(a). The first proposed MILP model is applied based on these correspondences, obtaining the sequence in Table 1(b), which is composed by 4 iterations. For each iteration, the devices whose poses are obtained and the number of correspondences are shown. It can be observed that the minimum number of correspondences for any iteration is 110.

The sequence in Table 1(b) is one of several feasible solutions for the first proposed model, all of them having a minimum number of correspondences of 110. The goal of the second model is finding the sequence which maximizes the total number of correspondences among all the feasible solutions of the first model. The result is shown in Table 1(c). The minimum per iteration remains the same (110) but the total number of correspondences is higher (3580).

Finally, Table 1(d) shows the sequence that would be obtained by simply maximizing the total number of correspondences, ignoring the minimum number of correspondences per iteration. Although the total is higher than in the previous cases, the low number of correspondences in the second iteration (5) can lead to an erroneous pose estimation.

Table 1
Examples of pose estimation sequences for a system composed by 5 devices. (a) Number of correspondences for each device pair and trio. (b) Sequence obtained by the first proposed model. The minimum number of correspondences in any iteration is maximized (110). (c) Sequence obtained by the second proposed model. The total number of correspondences is maximized constrained to the minimum number of correspondences per iteration obtained from the first model (110). (d) Sequence obtained by maximizing the total number of correspondences, ignoring the minimum number of correspondences per iteration. Although the total is high, the low number of correspondences in the second iteration (5) can lead to an erroneous pose estimation.

Pair	# Corr.	Pair	# Corr.
(a) Number of correspondences for each pair and trio of devices			
$\mathbb{C}(D_1, D_2)$	3000	$\mathbb{C}(D_2, D_4)$	520
$\mathbb{C}(D_1, D_3)$	430	$\mathbb{C}(D_2, D_5)$	120
$\mathbb{C}(D_1, D_4)$	400	$\mathbb{C}(D_3, D_4)$	150
$\mathbb{C}(D_1, D_5)$	15	$\mathbb{C}(D_3, D_5)$	550
$\mathbb{C}(D_2, D_3)$	5000	$\mathbb{C}(D_4, D_5)$	1000
Trio	# Corr.	Trio	# Corr.
$\mathbb{C}(D_1, D_2, D_3)$	3	$\mathbb{C}(D_1, D_4, D_5)$	2
$\mathbb{C}(D_1, D_2, D_4)$	330	$\mathbb{C}(D_2, D_3, D_4)$	4
$\mathbb{C}(D_1, D_2, D_5)$	0	$\mathbb{C}(D_2, D_3, D_5)$	5
$\mathbb{C}(D_1, D_3, D_4)$	0	$\mathbb{C}(D_2, D_4, D_5)$	110
$\mathbb{C}(D_1, D_3, D_5)$	11	$\mathbb{C}(D_3, D_4, D_5)$	140

Iteration	New poses (triangulating devices)	# Correspondences
(b) Sequence obtained by the first proposed MILP model		
1	$D_3,\ D_4$	150
2	D_5 ($D_3,\ D_4$)	140
3	D_2 ($D_4,\ D_5$)	110
4	D_1 ($D_2,\ D_4$)	330
Total:		730
Iteration	New poses (triangulating devices)	# Correspondences
(c) Sequence obtained by the second proposed MILP model		
1	$D_1,\ D_2$	3000
2	D_4 ($D_1,\ D_2$)	330
3	D_5 ($D_2,\ D_4$)	110
4	D_3 ($D_4,\ D_5$)	140
Total:		3580
Iteration	New poses (triangulating devices)	# Correspondences
(d) Sequence obtained by maximizing total number of correspondences, ignoring number of correspondences per iteration		
1	$D_2,\ D_3$	5000
2	D_5 ($D_2,\ D_3$)	5
3	D_4 ($D_3,\ D_5$)	140
4	D_1 ($D_2,\ D_4$)	330
Total:		5475

4.3. Pose refinement

In the previous step, an initial pose is obtained for every device of the system. Now, these poses are refined further using the Bundle Adjustment (BA) algorithm [43]. BA simultaneously refines the 3D points describing the scene geometry as well as the pose parameters of the system devices by minimizing the reprojection error.

Let us consider a scene point, $\mathbf{p}_i$, and $\mathbf{x}_i^j$ its projection to the device D_j. After the initial pose estimation, if the point is visible on at least two devices, its 3D coordinates can be estimated by triangulation. Let us consider our system is composed by N devices and the number of triangulated points is K. Then, the reprojection error of the scene reconstruction is:

$$\varepsilon = \sum_{k=1}^{K} \sum_{n=1}^{N} \upsilon_{k,n} \|\mathbf{x}_k^n - \mathcal{H}(\mathbf{p}_k, D_n)\|^2, \tag{9}$$

where $\upsilon_{k,n}$ is a binary variable indicating if the point $\mathbf{p}_k$ is visible from device D_n, and $\mathcal{H}(\mathbf{p}, D)$ is a function that projects a scene point $\mathbf{p}$ in the device D (using the pin-hole model exposed in Section 3.1). The BA algorithm tries to minimize the reprojection error of the whole system using LMA. In our case, the parameters to be optimized are: the pose parameters (motion), and the 3D position of the correspondences previously obtained by triangulation (structure). Since they are based on the initial pose, they are prone to error and thus we also allow BA to modify them. In particular, the SBA implementation [16], a variation of the algorithm exploiting the sparse structure of the problem parameters to reduce the computational costs, has been employed.

4.4. Reconstruction

Finally, the scene reconstruction can be performed employing the refined extrinsic parameters of each device obtained from the Bundle Adjustment process and the same correspondences so that calibration and reconstruction are performed simultaneously.

Each 3D point which is visible in at least two devices can be triangulated to obtain its final 3D position. The set of produced points can be post-processed in order to obtain a final and unique mesh of the whole scene. Section 5 shows some of the reconstructions obtained after the proposed system calibration.

5. Experiments and results

This section explains the experimentation carried out to test our proposal. First, the accuracy of the automatic calibration method is evaluated by scanning a flat surface and comparing the results to a pattern-based calibration. Then, the proposed method is evaluated by scanning two complex scenes. The obtained results are checked against a ground truth obtained by a laser scanner and compared with the reconstructions obtained by KinectFusion [46] and VisualSFM [47]. Finally, the method versatility is tested by studying its adaptation capacity to a range of scenes which require different device arrangements.

The employed MVSLS is composed by 4 cameras and 4 projectors with resolutions of 2080×1552 and 1024×768 pixels respectively. The intrinsic and distortion parameters of all devices are obtained by calibration, while the extrinsic parameters are totally unknown. Each device is connected to a slave computer, while a master computer is responsible of coordinating the slaves to perform the calibration and scanning process. To solve the MILP models, the solver CBC 2.8 from the Coin-OR software package [48] has been used.

We have employed the following procedure to set up the devices in order to ensure that their fields of view overlap. First, all the projectors are set to project a white pattern, and placed in the best configuration. Since they are projecting light, we can easily see their overlapping areas as an even brighter region. Then, we turn the cameras on and place them where we consider appropriate by observing their images on the computer. Finally, if necessary, we make small adjustments to the device locations. In total, using 4 cameras and 4 projectors, the set up process takes no more than two minutes.

5.1. Accuracy evaluation in a planar surface

A flat surface of approximately 1 m^2 showing a chessboard calibration pattern has been employed to evaluate the accuracy of the proposed method. The surface has been scanned at several distances by the MVSLS. For each distance, the extrinsic parameters have been first estimated for the cameras using the corners of the chessboard (which are at known locations), and for the projectors using the method proposed in [22]. Then, the scene has been reconstructed using these extrinsic parameters. Additionally, our

Table 2
Mean deviations of the points reconstructed by different calibration methods to its ideal plane. Best result for each distance is set in bold type. When SBA refinement is applied, the proposed method achieves similar results to those obtained by a standard chessboard calibration method, otherwise our method outperforms the pattern-based calibration. Errors increase along with the distance to the plane.

	Distance to plane (m)			
	1.50	2.00	2.50	3.00
Proposed method	$8.13 \cdot 10^{-4}$	$\mathbf{9.47 \cdot 10^{-4}}$	$1.33 \cdot 10^{-3}$	$\mathbf{2.02 \cdot 10^{-3}}$
Proposed method without SBA	$1.15 \cdot 10^{-3}$	$1.34 \cdot 10^{-3}$	$1.48 \cdot 10^{-3}$	$2.16 \cdot 10^{-3}$
Pattern calibration	$1.47 \cdot 10^{-3}$	$1.59 \cdot 10^{-3}$	$1.70 \cdot 10^{-3}$	$2.98 \cdot 10^{-3}$
Pattern calibration with SBA	$\mathbf{8.09 \cdot 10^{-4}}$	$9.48 \cdot 10^{-4}$	$\mathbf{1.32 \cdot 10^{-3}}$	$\mathbf{2.02 \cdot 10^{-3}}$

method has been employed to reconstruct the scene. Then, for each case, the points obtained have been fitted to the ideal plane and their distance to it measured as an indication of the reconstruction error. The maximum epipolar distance and maximum reprojection error were set to 1 pixel.

Table 2 shows the results obtained. Four cases have been studied: (i) the fully automatic calibration method proposed in this paper, (ii) the proposed calibration without applying the final SBA optimization step, so that only the initial calibration is evaluated (see Section 4.2), (iii) the chessboard calibration employing [22] and (iv) the chessboard calibration with a final SBA optimization step.

The results show that the accuracy of the proposed calibration and the chessboard method are very similar. This is more noticeable in those cases where the final SBA refinement has been applied and the errors are almost identical. When final SBA refinement is not applied, the estimations obtained by our proposal outperform the pattern-based method in all cases. Finally, it can be noticed that errors increase along with the distance between the devices and the surface.

5.2. Accuracy evaluation in real complex scenes

This sections aims at evaluating the proposed method in real scenes using ground truth validation. To do so, a Leica laser ScanStation C10 has been employed to scan two different scenarios. Additionally, our method is compared against the KinectFusion approach [46] implemented in the official Windows SDK for the Kinect v2 sensor. In both methods, we manually selected a set of 3D correspondences to establish an initial alignment between the reconstruction and the gound truth point cloud. Then, the iteration to closest point algorithm (ICP) [49] was employed to refine the alignments. For our method, the correspondences were also employed to scale the point cloud before the alignment to the real scene scale.

The first scenario is a pile of boxes stored in the corner of a room, which was scanned with the three methods. Fig. 5(a) shows the original point cloud captured with the laser system. Next to it, Fig. 5(b), the reconstruction obtained by our method is shown. A total of 4 cameras and 4 projectors were used. The color of each point represents the error distance to the nearest point from the ground truth. The blue color represents the lowest error and red the highest. The maximum error of our reconstruction in this scenario is 5 mm. Fig. 5(c) shows the reconstruction obtained with the KinectFusion method. It can be noticed that, in this case, the maximum error is 50 mm. In order to properly evaluate the different methods, Fig. 5(d) shows the error distributions obtained. In our case, the average error was 1.26 mm, while for the KinectFusion method the average error was 9.84 mm.

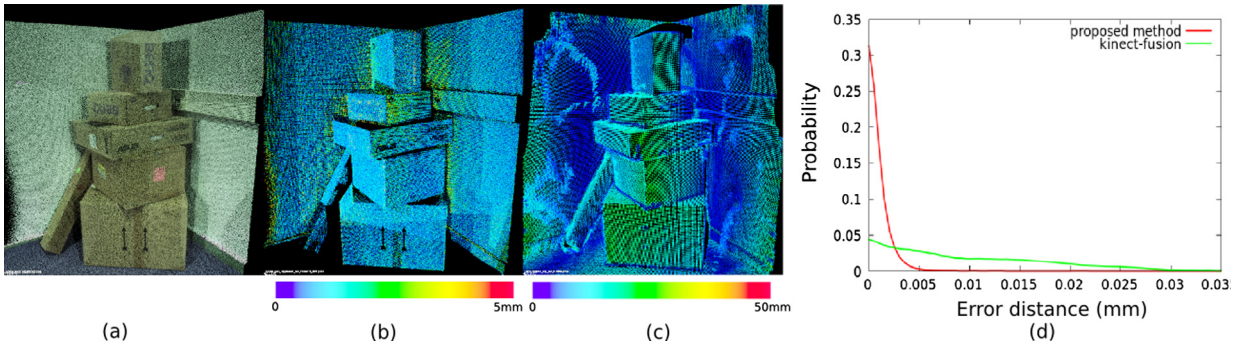

Fig. 5. Errors of the methods tested in the boxes scenario. (a) Ground truth point cloud obtained by laser. (b) Reconstruction errors obtained with our method and (c) with KinectFusion. (d) Point to plane error distribution of both methods.

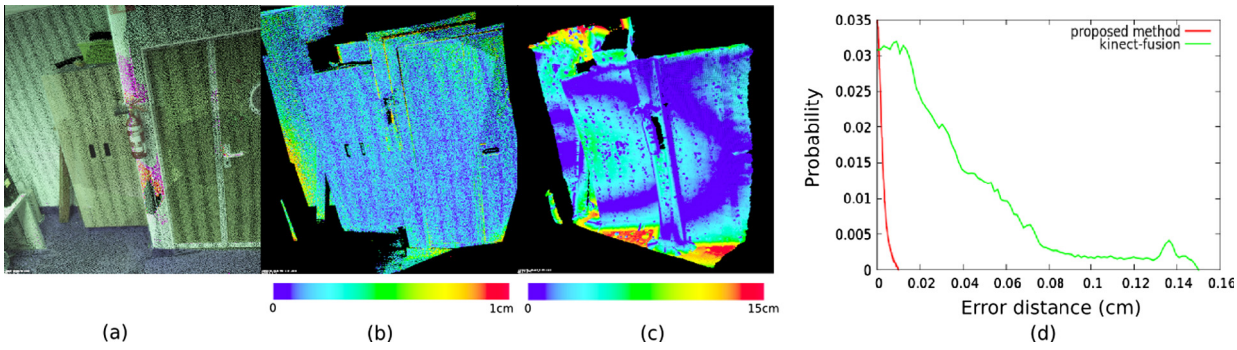

Fig. 6. Errors of the methods tested in the wardrobe scenario. (a) Ground truth point cloud obtained by laser. (b) Reconstruction errors obtained with our method and (c) with KinectFusion. (d) Point to plane error distribution of both methods.

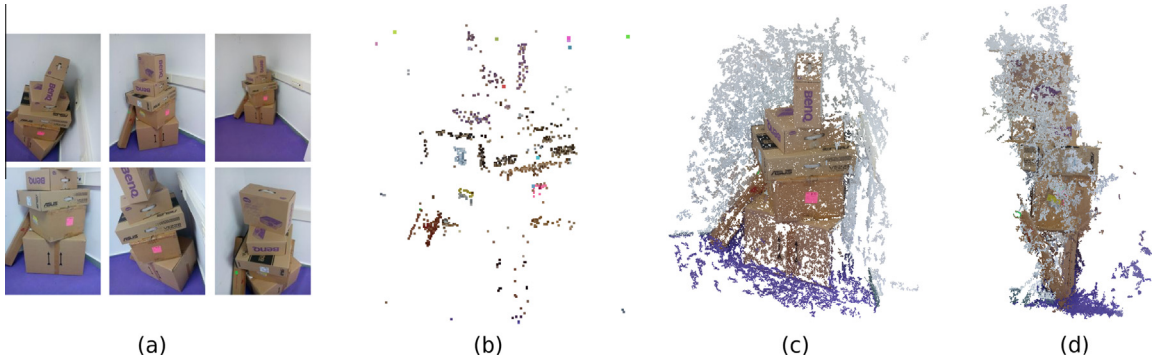

Fig. 7. Reconstruction obtained by VisualSFM for the Scene in Fig. 5. (a) Some of the employed images. (b) Sparse point cloud obtained after the Structure from Motion process. (c and d) Dense reconstruction obtained after applying PMVS. It can be observed the obtained reconstruction presents a huge amount of noise and holes.

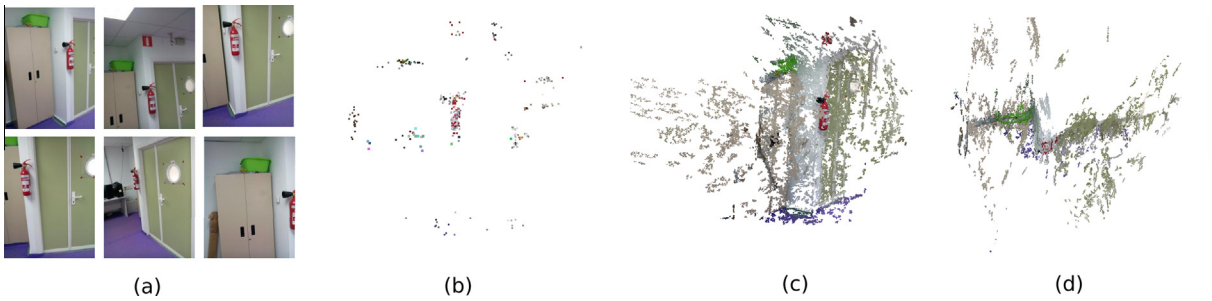

Fig. 8. Reconstruction obtained by VisualSFM for the Scene in Fig. 6. (a) Some of the employed images. (b) Sparse point cloud obtained after the Structure from Motion process. (c and d) Dense reconstruction obtained after applying PMVS. As in the scene in Fig. 7, the obtained reconstruction presents a huge amount of noise and holes.

The second scenario is a bit more challenging, since it is a larger area including a door and a wardrobe. In this case, we employed 3 cameras and 3 projectors. However, for the KinectFusion method, it took several attempts to obtain a reconstruction since the method lost track very easily. The results are shown in Fig. 6. In both cases, the errors are higher than in the previous scenario. This is because the measured area is much larger. However, the differences between our proposal and the Kinect one are even more pronunciated. In our case, the average error was 2.1 mm, while the Kinect method obtained an average error of 35 mm.

5.2.1. Qualitative comparison with Structure from Motion techniques

The scenes previously tested with our method have also been compared to those obtained by VisualSFM [47], one of the most popular libraries for Structure from Motion. Figs. 7 and 8 shows the reconstructions obtained by VisualSFM by providing a total of 25 images per scene with a resolution of 2560×1920 pixels. It can be noted at first glance that the results present significant lower quality in comparison to those obtained by KinectFusion and our approach. The errors are mainly generated during the patch-based dense reconstruction step, which produces a huge amount of noise and holes, leading to almost unusable results.

5.3. Processing time

The calibration and reconstruction processes required less than 2 min in analyzing the scenes of Figs. 7 and 8, employing an Intel Core2 Quad 2.40 GHz processor and a single thread. Like other multi-view reconstruction techniques, most of the processing time is consumed by the Bundle Adjustment step. However, it must be noted that the computing time might be significantly reduced using a parallelized implementation on CPU or GPU. The computation of the MILP model took less than 1 s for the proposed configuration. The processing time of VisualSFM, which also uses Bundle Adjustment, was similar to ours.

Finally, KinectFusion is based on a highly parallelized GPU implementation which allows real-time reconstruction. However, the scanning process is manual and, as previously stated, it often requires several attempts to obtain the reconstruction since the it loses track easily in scenes with large planar surfaces. Hence, the reconstructions of a scene like those in Figs. 7 and 8 can take several minutes.

5.4. Adaptation to different types of scenes

In this section, our method has been applied to different types of scenes. The results show the effectiveness of the method and its adaptation capacity to different environments. Figs. 9–11 show some of the selected scenes to evaluate the proposal. Each figure shows some real pictures of the scene and, from different viewpoints, the set of point clouds in different colors and the final generated mesh by the Poisson algorithm [50].

The results include some of the most representative cases. In Fig. 9, a wide area inside a room is reconstructed employing all the devices to cover the full scene. In Fig. 10, a detailed reconstruction of a set of objects (a pile of boxes and books) is performed using the different cameras and projectors to cover as many gaps as possible. Finally, Fig. 11 shows a 360 degrees reconstruction of two individuals, who have been surrounded by the set of cameras and projectors. A similar 360 degrees scan is shown in Fig. 1 for a single standing person.

Fig. 12 shows the different device arrangements that have been necessary to completely cover the scenes in 9–11 respectively.

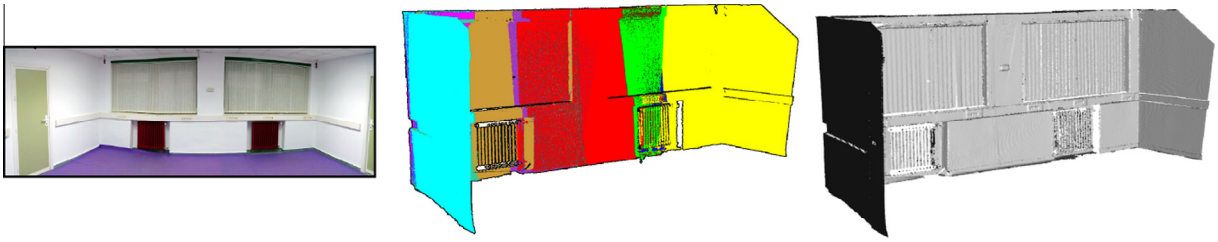

Fig. 9. Reconstruction of a wide area inside a room. Left figure shows a panoramic picture of the reconstructed scene. Employing multiple cameras and projectors allows to cover a wider area of the scene. For each viewpoint, the point clouds obtained for each camera-projector pair are displayed in different colors. Finally, the generated mesh from the Poisson algorithm is also shown. (Best viewed in color.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

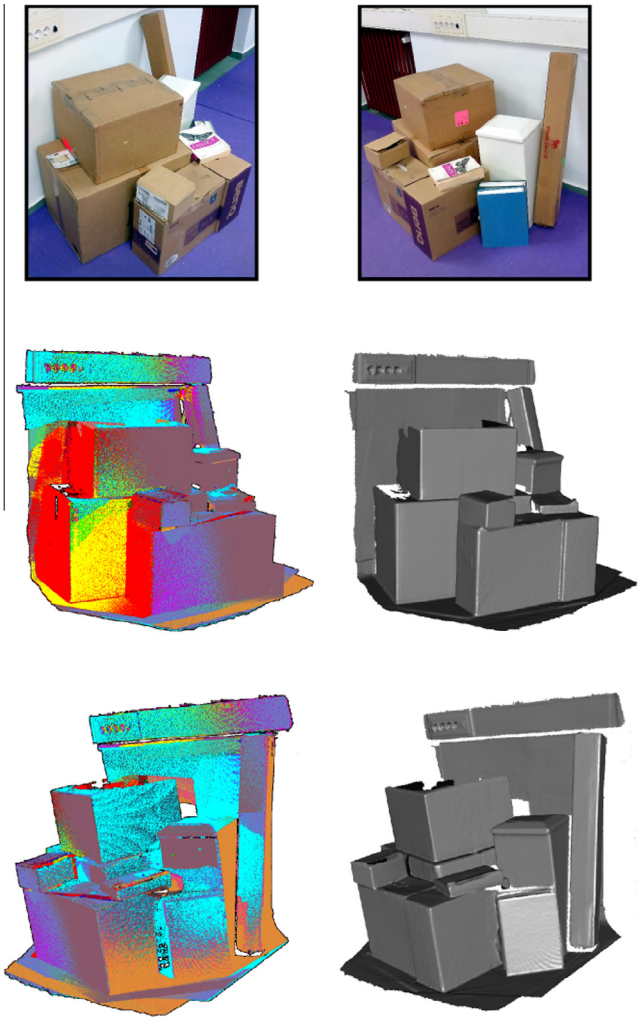

Fig. 10. Detailed reconstruction of a pile of boxes and books. Although the specific scene is not relatively large, employing multiple cameras and projectors allows to cover most of the nooks and gaps that could not be reconstructed otherwise. Two viewpoints of the obtained reconstruction are shown. For each viewpoint, the point clouds obtained for each camera-projector pair are displayed in different colors. Finally, the generated mesh from the Poisson algorithm is also shown. (Best viewed in color.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

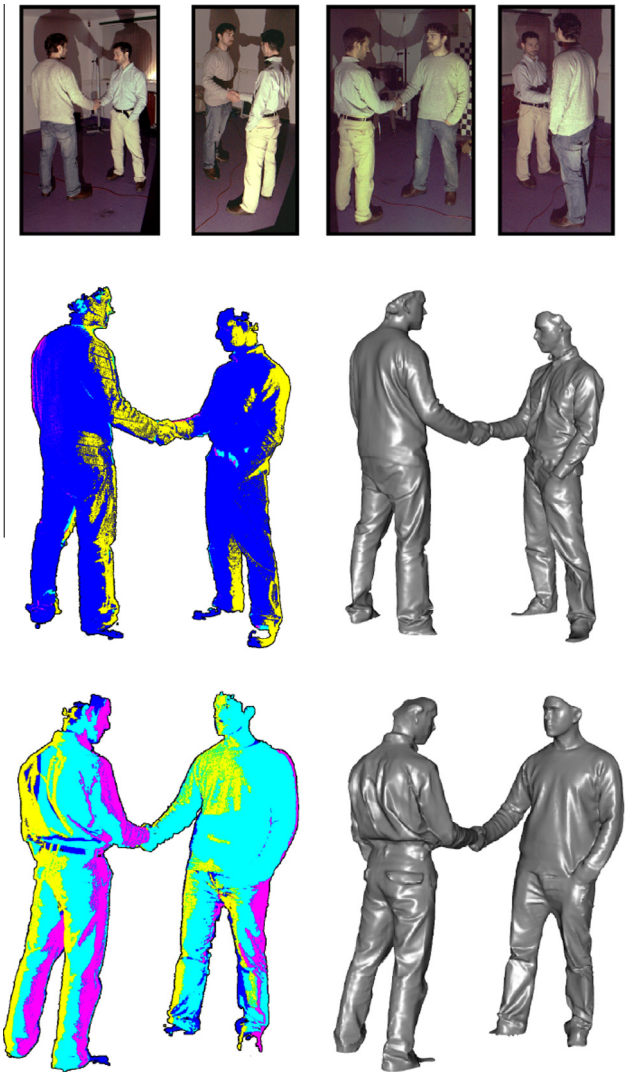

Fig. 11. Example of 360 degrees reconstruction consisting in two people shaking hands. Two viewpoints of the obtained reconstruction are shown. For each viewpoint, the point clouds obtained for each camera-projector pair are displayed in different colors. Finally, the generated mesh from the Poisson algorithm is also shown. A similar 360 degrees scan example can be seen in Fig. 1. (Best viewed in color.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

As it can be seen, the proposed method can perform a simultaneous self-calibration and reconstruction, obtaining a precise registration between the different viewpoints and easily adapting to different types of scenes. In all cases the reprojection error obtained during the self-calibration process was reduced below 0.5 pixels after the SBA refinement.

5.5. System limitations

The proposed method is a step forward in automatic MVSLS calibration. However, it still presents some particular drawbacks which are worth to be mentioned. First, projector's depths of field

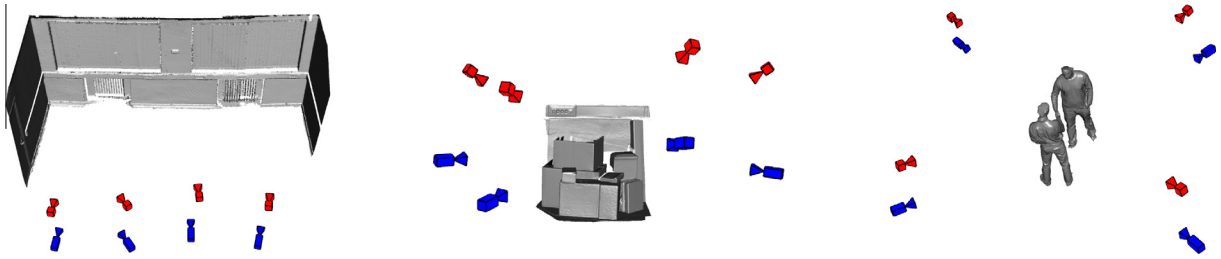

Fig. 12. Devices arrangements for scanned scenes in Figs. 9–11 respectively. Square red prisms represent cameras while rectangular blue prisms represent projectors. The proposed calibration method is suitable for different kinds of scenes and device arrangements. (Best viewed in color.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(DoF) are usually short. As a consequence, the projector positions which produce focused patterns over the scanned object are limited, unless projector focus is modified which would require the recalibration of its intrinsic parameters. Second, the calibration is estimated up to a scale factor since the method does not known any real distance reference in the scanned object. However, the distance reference can be provided manually or automatically by simply placing a square fiducial marker [51] on the scene which is visible by at least two devices. Finally, Bundle Adjustment can produce unstable results in the presence of a high number of outliers. Because of this, our method ensures to remove those correspondences which do not accomplish a minimum epipolar distance and reprojection error. However, we have experimented that an erroneous decodification and a high number of outliers can arise from an excessive light exposure, producing instabilities during the calibration.

6. Conclusions

In this paper, a method to simultaneously reconstruct and self-calibrate MVSLS is proposed. The method is based on device correspondences and a MILP optimization process, thus not being limited to any number or devices, arrangement or special equipment, and not requiring manual intervention.

The experimentation carried out shows that the proposed calibration method is precise and robust, achieving an accuracy similar to that obtained by a traditional chessboard pattern calibration. The results obtained when scanning real complex scenes are considerably close to the ground truth estimation and surpass other state of the art reconstruction techniques. Furthermore, due its flexibility, it is suitable for measuring a wide range of scenes and objects.

Acknowledgments

This work has been developed with the support of the Research Projects TIC-1692 (Junta de Andalucía) and TIC-0161 of the XXI Program for Research Support (University of Cordoba).

References

[1] Z. Bi, L. Wang, Advances in 3D data acquisition and processing for industrial applications, Robot. Comput.-Integr. Manuf. 26 (5) (2010) 403–413.
[2] J.L. Horijon, W.D. van Amstel, F.C.M. Couweleers, W.C. Ligthart, Optical system of an industrial 3D laser scanner for solder paste inspection, Proc. SPIE 2599 (1996) 162–170.
[3] G. Pavlidis, A. Koutsoudis, F. Arnaoutoglou, V. Tsioukas, C. Chamzas, Methods for 3D digitization of cultural heritage, J. Cultural Heritage 8 (1) (2007) 93–98.
[4] L. Schueremans, B.V. Genechten, The use of 3D-laser scanning in assessing the safety of masonry vaultsa case study on the church of Saint-Jacobs, Opt. Lasers Eng. 47 (34) (2009) 329–335.
[5] P.R. Jones, M. Rioux, Three-dimensional surface anthropometry: applications to the human body, Opt. Lasers Eng. 28 (2) (1997) 89–117.
[6] S. Bragança, P.M. Arezes, M. Carvalho, An overview of the current three-dimensional body scanners for anthropometric data collection, Occupational Safety and Hygiene III, CRC Press, 2015. pp. 149–153.
[7] M. Ribo, M. Brandner, State of the art on vision-based structured light systems for 3D measurements, in: International Workshop on Robotic Sensors: Robotic and Sensor Environments, 2005, 2005, pp. 2–6.
[8] J. Geng, Structured-light 3D surface imaging: a tutorial, Adv. Opt. Photon. 3 (2) (2011) 128–160.
[9] F. Sadlo, T. Weyrich, R. Peikert, M. Gross, A practical structured light acquisition system for point-based geometry and texture, in: Eurographics/IEEE VGTC Symposium Proceedings Point-Based Graphics, 2005, 2005, pp. 89–145.
[10] H.J. Chien, C.Y. Chen, C.F. Chen, Y.M. Su, Y.Y. Chang, Adaptive 3D reconstruction system with improved recovery of miscoded region to automatically adjust projected light patterns, in: 23rd International Conference Image and Vision Computing New Zealand, 2008, IVCNZ 2008, 2008, pp. 1–6.
[11] A. Griesser, L. Van Gool, Automatic interactive calibration of multi-projector-camera systems, in: 2006 Conference on Computer Vision and Pattern Recognition Workshop CVPRW06 00 (c), 2006, pp. 56–62.
[12] R. Garcia, A. Zakhor, Geometric calibration for a multi-camera-projector system, in: 2013 IEEE Workshop on Applications of Computer Vision (WACV), 2013, pp. 467–474.
[13] Y. Li, S. Lee, Stratified self-calibration and metric reconstruction of a trinocular structured light vision system, in: Sixth International Conference on 3-D Digital Imaging and Modeling, 2007, 3DIM '07, 2007, pp. 328–336.
[14] R. Furukawa, H. Kawasaki, Uncalibrated multiple image stereo system with arbitrarily movable camera and projector for wide range scanning, in: Fifth International Conference on 3-D Digital Imaging and Modeling, 2005, 3DIM 2005, 2005, pp. 302–309.
[15] D.G. Aliaga, Y. Xu, A self-calibrating method for photogeometric acquisition of 3D objects, IEEE Trans. Pattern Anal. Mach. Intell. 32 (4) (2010) 747–754.
[16] M.I. Lourakis, A.A. Argyros, Sba: a software package for generic sparse bundle adjustment, ACM Trans. Math. Software 36 (1) (2009).
[17] A. Paoli, A.V. Razionale, Large yacht hull measurement by integrating optical scanning with mechanical tracking-based methodologies, Robot. Comput.-Integr. Manuf. 28 (5) (2012) 592–601.
[18] S. Barone, A. Paoli, A.V. Razionale, Shape measurement by a multi-view methodology based on the remote tracking of a 3D optical scanner, Opt. Lasers Eng. 50 (3) (2012) 380–390.
[19] S. Barone, A. Paoli, A.V. Razionale, Multiple alignments of range maps by active stereo imaging and global marker framing, Opt. Lasers Eng. 51 (2) (2013) 116–127.
[20] J. Yuan, Q. Wang, X. Jiang, B. Li, A high-precision registration technology based on bundle adjustment in structured light scanning system, Math. Probl. Eng. 2014 (2014), http://dx.doi.org/10.1155/2014/897347.
[21] D. Lanman, D. Crispell, G. Taubin, Surround structured lighting for full object scanning, Sixth International Conference on 3D Digital Imaging and Modeling 3DIM 2007, vol. 1, 2007, pp. 107–116.
[22] D. Moreno, G. Taubin, Simple, accurate, and robust projector-camera calibration, in: Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission, 3DIMPVT'12, IEEE Computer Society, Washington, DC, USA, 2012, pp. 464–471.
[23] T. Svoboda, D. Martinec, T. Pajdla, A convenient multicamera self-calibration for virtual environments, Presence 14 (4) (2005) 407–422.
[24] M. Fiala, Designing highly reliable fiducial markers, IEEE Trans. Pattern Anal. Mach. Intell. 32 (7) (2010) 1317–1324.
[25] Y.F. Li, S. Member, S.Y. Chen, S. Member, Automatic recalibration of an active structured light vision system, IEEE Trans. Robot. Autom. 19 (2003).
[26] Z. Xie, X. Wang, S. Chi, Simultaneous calibration of the intrinsic and extrinsic parameters of structured-light sensors, Opt. Lasers Eng. 58 (0) (2014) 9–18.
[27] B. Zhang, Y.F. Li, Y.H. Wu, Self-recalibration of a structured light system via plane-based homography, Pattern Recogn. 40 (4) (2007) 1368–1377.
[28] H. Luo, J. Xu, N.H. Binh, S. Liu, C. Zhang, K. Chen, A simple calibration procedure for structured light system, Opt. Lasers Eng. 57 (0) (2014) 6–12.
[29] C. Yuan, Q. Guan, X. Wang, T. Fang, A multiple view self-calibration and metric reconstruction method for structured light system, in: 2010 Chinese Conference on Pattern Recognition (CCPR), 2010, pp. 1–5.

[30] Z. Liu, X. Li, F. Li, G. Zhang, Calibration method for line-structured light vision sensor based on a single ball target, Opt. Lasers Eng. 69 (0) (2015) 20–28.
[31] W. Jang, C. Je, Y. Seo, S.W. Lee, Structured-light stereo: comparative analysis and integration of structured-light and active stereo for measuring dynamic shape, Opt. Lasers Eng. 51 (11) (2013) 1255–1264.
[32] R. Furukawa, K. Inose, H. Kawasaki, Multi-view reconstruction for projector camera systems based on bundle adjustment, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2009, CVPR Workshops 2009, 2009, pp. 69–76.
[33] D.G. Aliaga, Y.X.Y. Xu, Photogeometric structured light: a self-calibrating and multi-viewpoint framework for accurate 3D modeling, in: IEEE Conference on Computer Vision and Pattern Recognition (2008), 2008, pp. 1–8.
[34] S. Inokuchi, K. Sato, F. Matsuda, Range imaging system for 3-D object recognition, in: ICPR84, 1984, pp. 806–808.
[35] R. Valkenburg, A. McIvor, Accurate 3D measurement using a structured light system, Image Vis. Comput. 16 (2) (1998) 99–110.
[36] Z. Yan, L. Yu, Y. Yang, Q. Liu, Beyond the interference problem: hierarchical patterns for multiple-projector structured light system, Appl. Opt. 53 (17) (2014) 3621–3632.
[37] C. Je, K.H. Lee, S.W. Lee, Multi-projector color structured-light vision, Signal Process.: Image Commun. 28 (9) (2013) 1046–1058.
[38] R.I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, second ed., Cambridge University Press, 2004. ISBN: 0521540518.
[39] R.I. Hartley, P.F. Sturm, Triangulation, Comput. Vis. Image Understand. 68 (2) (1997) 146–157.
[40] X. Armangué, J. Salvi, Overall view regarding fundamental matrix estimation, Image Vis. Comput. 21 (2) (2003) 205–220.
[41] K. Levenberg, A method for the solution of certain non-linear problems in least squares, Quart. J. Appl. Math. II (2) (1944) 164–168.
[42] D.W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, SIAM J. Appl. Math. 11 (2) (1963) 431–441.
[43] B. Triggs, P.F. McLauchlan, R.I. Hartley, A.W. Fitzgibbon, Bundle adjustment – a modern synthesis, in: Vision Algorithms: Theory and Practice, Springer, 2000, pp. 298–372.
[44] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.
[45] A. Schrijver, Theory of Linear and Integer Programming, John Wiley & Sons, Inc., New York, NY, USA, 1986.
[46] R.A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A.J. Davison, P. Kohli, J. Shotton, S. Hodges, A. Fitzgibbon, KinectFusion: real-time dense surface mapping and tracking, in: Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR '11, IEEE Computer Society, Washington, DC, USA, 2011, pp. 127–136.
[47] C. Wu, Towards linear-time incremental structure from motion, in: 2013 International Conference on 3D Vision – 3DV 2013, 2013, pp. 127–134.
[48] R. Lougee-Heimer, The common optimization interface for operations research: promoting open-source software in the operations research community, IBM J. Res. Dev. 47 (1) (2003) 57–66.
[49] P.J. Besl, N.D. McKay, A method for registration of 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 239–256.
[50] M. Kazhdan, M. Bolitho, H. Hoppe, Poisson surface reconstruction, in: Proceedings of the Fourth Eurographics Symposium on Geometry Processing, SGP '06, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 2006, pp. 61–70.
[51] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, M. Marín-Jiménez, Automatic generation and detection of highly reliable fiducial markers under occlusion, Pattern Recogn. 47 (6) (2014) 2280–2292.