Registration of surfaces minimizing error propagation for a one-shot multi-slit hand-held scanner[☆]

C. Matabosch[a,*], D. Fofi[b], J. Salvi[a], E. Batlle[a]

[a]Institute of Informatics and Applications of the University of Girona, Spain
[b]University of Burgundy, France

Abstract

We propose an algorithm for the on-line automatic registration of multiple 3D surfaces acquired in a sequence by a new hand-held laser scanner. The laser emitter is coupled with an optical lens that spreads the light forming 19 parallel slits that are projected to the scene and acquired with subpixel accuracy by a camera. Splines are used to interpolate the acquired profiles to increase the sample of points and Delaunay triangulation is used to obtain the normal vectors at every point. A point-to-plane pair-wise registration method is proposed to align the surfaces in pairs while they are acquired, conforming paths and eventually cycles that are minimized once detected. The algorithm is specially designed for on-line applications and can be classified as a closing-the-loop technique, where there are not that many competing methods, though it has been compared to the literature. Experiments providing qualitative and quantitative evaluation are shown by means of synthetic and real data and we demonstrated the reliability of our technique.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Computer vision; 3D registration; Multi-view; Closing-the-loop

1. Introduction

The complete 3D acquisition of a given object, surface or even scene has many research challenges and it is definitely in the research interests of the computer vision community. Besides, there are several applications which may benefit by such complete acquisition such as reverse engineering, modeling, metrology, visual inspection and even robot navigation.

Overall, there are diverse 3D acquisition systems which are basically based on laser triangulation [1,2] and pattern projection [3], especially when dense images are required. These systems gather range images from which 3D information can be extracted. Besides, other acquisition systems are based on processing one or many 2D images captured by cameras. One of

these techniques is *Shape from Silhouettes*, in which several images of the measuring object are acquired and the complete 3D model of the object is computed by means of its silhouette at every image. Although some authors tend to use this technique without knowing the camera poses from where images were acquired [4], such positions are usually required and only a rough accuracy is obtained. Other techniques are based on directly processing the set of images by means of solving the matching problem [5]. In this case, features such as points/lines/regions of interest must be determined within the images reducing drastically the resolution and hence obtaining a sparse acquisition. In conclusion, laser triangulation is considered one of the most reliable techniques of acquiring 3D data, so that data is dense and accurate and the correspondence problem alleviated.

In general, laser triangulation techniques are based on the use of a laser emitter coupled to a cylindrical lens that spreads the light forming a plane that is projected to the measured surface. The projection of a laser plane only lets us acquire a profile of the measuring surface. Hence, in most cases a mechanical system is added to permit the scanning, so that: (a) the laser plane is projected onto a rotating mirror and reflected towards the

surface; (b) the laser beam is attached to a moving worm gear; (c) the laser beam keeps motionless while the object is placed on a rotating table, obtaining dense acquisitions. However, the accuracy of the 3D acquisition depends on the mechanical system as potential vibrations are likely to produce misalignments. Furthermore, the sequence of images that are captured in the scanning process forces the object to be motion controlled reducing the number of potential applications. Finally, an uncompleted acquisition of the object is usually obtained due to object occlusions and the limited field of view of the sensor.

Summarizing, laser triangulation is a reliable technique to acquire dense and accurate 3D data. However, existing commercial sensors usually only acquire a partial view of the object, others are constraint to mechanical structures such as rotating tables and moving gears. There are few sensors that can compute their ego-pose in a free space. For instance, the ZScanner700 of ZCorporation [6] can compute such ego-pose, but a set of reflective targets must be stuck on the measuring surface. The set of views are directly aligned once the ego-pose is computed. Registration is a technique that can permit to perform such alignment without adding markers or any other reference to the object or to the measuring scenario.

Pair-wise registration is a well-studied problem in the literature. It is known that the published techniques can be classified in coarse and fine registration techniques depending on the way of solving the registration: (a) in a closed-form solution from a reduced set of points of interest obtained from surface features and solving a global matching; (b) in an iterative minimization solution using a sample of points and solving the matching locally. Multi-view registration is a more difficult problem and there are two strategies to solve the alignment: local (sequential) and global (multi-view). The sequential registration of views does not give an optimal solution due to the accumulation and propagation of errors. Besides, global registration tries to distribute the registration errors evenly among all the views but it is not suitable for on-line registration since all the views are first required to initiate the registration. In this paper, we propose a closing-the-loop technique for the on-line registration of sequence of views in which cycles are minimized once detected preserving an optimal solution. The pair-wise registration is solved by using a variant of the point-to-plane technique and cycles are minimized considering only the views involved constricting the computing time.

The remaining of the paper is structured as follows. First, a brief overview of registration techniques is presented in Section 2, discussing the pros and cons of the existing methods with the aim of justifying our proposal. Then, Section 3 presents our proposal including pair-wise registration, cycle detection and cycle minimization. Experimental results provided by both synthetic and real data are presented in Section 4. The article ends with conclusions.

2. Overview of range image registration techniques

According to a recent study of the state-of-art concerning surface registration techniques [7], all the existing methods are classified in two main groups of techniques: (a) Coarse Registration and (b) Fine Registration. In the following part, pros and cons of these techniques are described and every technique is summarized.

Coarse registration techniques obtain a rough alignment by means of finding correspondences between two surfaces. Most coarse registration techniques are based on searching for points (curves) in the second surface that are similar to points (curves) in the first surface. Overall, there are only two different methods to choose the interest points: (a) feature-to-points and (b) point-to-features. The first selects points in the first surface that are similar to a predefined feature [8,9]. In the second, some points in the first surface are arbitrarily selected and characterized considering the position of its neighbors and searched in the second surface [10–12]. In both techniques, all points in the second surface must be compared with the selected points to establish correspondences. Once the correspondence problem is solved, the Euclidean motion that aligns both surfaces is computed in a closed-form solution.

Coarse registration techniques have two main drawbacks. The long time required to solve the matching among points and the poor quality of the registration. Hence, a fine registration technique is usually applied *a posteriori* to improve the results by minimization.

Fine registration techniques search for an accurate alignment of two acquired surfaces by minimizing the distance between both surfaces iteratively. Such distance is computed in each iteration by means of temporal matching among points from both surfaces. Overall, the existing techniques solve the temporal matching using one of the following three approaches: (1) point-to-point, in which correspondences are established by searching for the points in the second surface that are closest to a set of points in the first [13]; (2) point-to-plane, in which the points in the second surface are computed at every step by the intersection of a plane and a line. The line is defined by a point and a vector, so that the point is a given point in the first surface; and the vector is orthogonal to the surface defined by a neighborhood around the given point. The plane is located where that line intersects with the second surface and it is oriented in such a way that the plane is tangent to that surface [14]; and finally (3) point-to-projection, in which correspondences are established between points in the first surface and the points obtained by projecting the first points onto the second surface using the point of view of the second surface [15].

Although point-to-projection is the fastest technique because searching is avoided, results obtained are not very satisfactory. Besides, point-to-plane provides the best results because it is not influenced by local minima [12,16]. Despite the difficulty in determining the intersection of a line with a cloud of points in 3D, several authors have presented proposals to facilitate this computation [17,18].

One-to-one alignment of views in a sequence causes a drift that is propagated throughout the sequence. Hence, some techniques have been proposed to reduce the propagating error benefiting from the existence of cycles and re-visited regions and considering the uncertainty in the alignment.

In order to minimize the propagating error, some authors have improved their algorithms by adding a final step that aligns all

the acquired views at the same time. This approach spreads one-to-one pair-wise registration errors throughout the sequence of views, known as multi-view registration [19].

Early approaches proposed the aggregation of subsequent views in a single metaview which is progressively enlarged each time another view is registered [14]. Here, the main constraint is the lack of flexibility to re-register views already merged in the metaview due to the greedy approach of the technique. In 1999, Pulli proposed an ICP[1] relaxation method based on the previous metaview approach but considering all the potential alignments between views before proceeding with the multi-view registration. In addition this method takes into account the information of all the overlapping areas and the already registered regions can be analyzed again for further transformations [20]. Later on, Nüchter proposed a global relaxation method based on Pulli's approach with the main difference that no iterative pair-wise alignment is required. However the success of this method drastically depends on the disposal of an accurate initial estimation of the pose [21].

A different approach was proposed by Bergevin [22], who presented a multi-view registration technique based on the graph theory: views are associated to nodes and transformations to edges. Authors consider all views as a whole and align all of them simultaneously. The same idea was proposed later on by Silva [23], Huber [24] and Krishnan [25]. Besides, Masuda presented a multi-view registration algorithm based on the Matching Signed Distance Fields in which outliers are automatically removed obtaining a more robust method [26]. Lu's technique is based on cycle minimization, though the relationship (edges) among views (nodes) are established prior to minimization [27]. Overall, multi-view techniques suffer two main drawbacks: (a) the whole set of 3D views has to be acquired before the algorithm starts; (b) an accurate estimation of the motion between views is needed as initial guesses to ensure convergence. Thus, multi-view techniques are not considered for on-line applications.

Few authors have faced the challenge of registering 3D views in a sequence while they are acquired avoiding or at least controlling error propagation. For instance, Sharp [28] proposed the registration of pairs of consecutive views until a cycle is found. Since only pair-wise registration is required, the method becomes very fast. Here, the interest is in the way of distributing the motion (and hence the propagation error) among the different views. The author proposed to use weights directly related to the residue obtained in the pair-wise registration. Actually, this is not very accurate especially in the presence of misalignments between end views in the cycle as a matter of noise and object occlusions. In this case, the whole motion of such a cycle is also distributed to all the views increasing the error in the registration.

Finally, in the last few years, a photogrammetric technique called Bundle Adjustment has increased popularity in the computer vision community and it is growing in interest in robotics. Bundle adjustment is the problem of refining a visual reconstruction to produce jointly optimal 3D structures and viewing parameter (camera pose and/or calibration) estimates [29]. Therefore, bundle adjustment techniques can be used in both robot/camera localization and 3D mapping in many fields such as camera calibration, robot navigation, and scene reconstruction. Since bundle adjustment is a non-linear minimization problem, it is solved by means of iterative non-linear least squares or total squares methods such as Levenberg–Marquardt or M-estimator techniques [7,30]. Although bundle adjustment is commonly classified as a multi-view technique, some authors have used it in consecutive pair-wise alignment as a technique to reduce error propagation [31].

In summary, we conclude that methods based on the metaview approaches present good results when initial guesses are accurate and the surface to be registered does not have a large scale. Otherwise, the method suffers a large propagation error producing drift and misalignments and its greedy approach usually falls in local minima. The use of methods based on graphs has the advantage of minimizing the error in all the views simultaneously but these techniques usually require a previous pair-wise registration step, whose accuracy can be determinant in the global minimization process. Besides, closing the loop strategies provide trustworthy constraints for error minimization but require a huge amount of memory and usually involve a high computational cost. Bundle adjustment techniques provide good results in the presence of outliers, but need a good enough initial guess and it is hardly used in large robot missions or large scale objects.

All these pros and cons of the existing methods have been considered to present a new surface registration technique which is presented and discussed in the rest of the paper.

3. Registering a sequence of surfaces

This section describes the proposing method for continuously registering a sequence of 3D views while they are acquired. The method first aligns the consecutive views by means of point-to-plane pair-wise registration. When a cycle is detected, a multi-view technique is applied only in the views conforming to the cycle leading to fast and accurate results and preserving the on-line registration for many and varied applications (see Fig. 1).

3.1. Pair-wise registration

Pair-wise registration is divided into a first coarse registration to estimate an initial alignment, followed by a fine registration computed by means of minimization techniques. In our case, views are acquired consecutively and a slight movement between views is assumed. Slight movement is defined like the movement that guarantees at least a 60% of overlapping with consecutive views. We initialize fine registration considering motionless views, avoiding the expensive computation required to compute initial guesses and preserving a high accuracy, as demonstrated in the following paragraphs and shown in the experimental results.

Point-to-plane has been chosen as the most suitable fine registration technique as discussed in the previous section. The technique we propose is based on the fast variant proposed
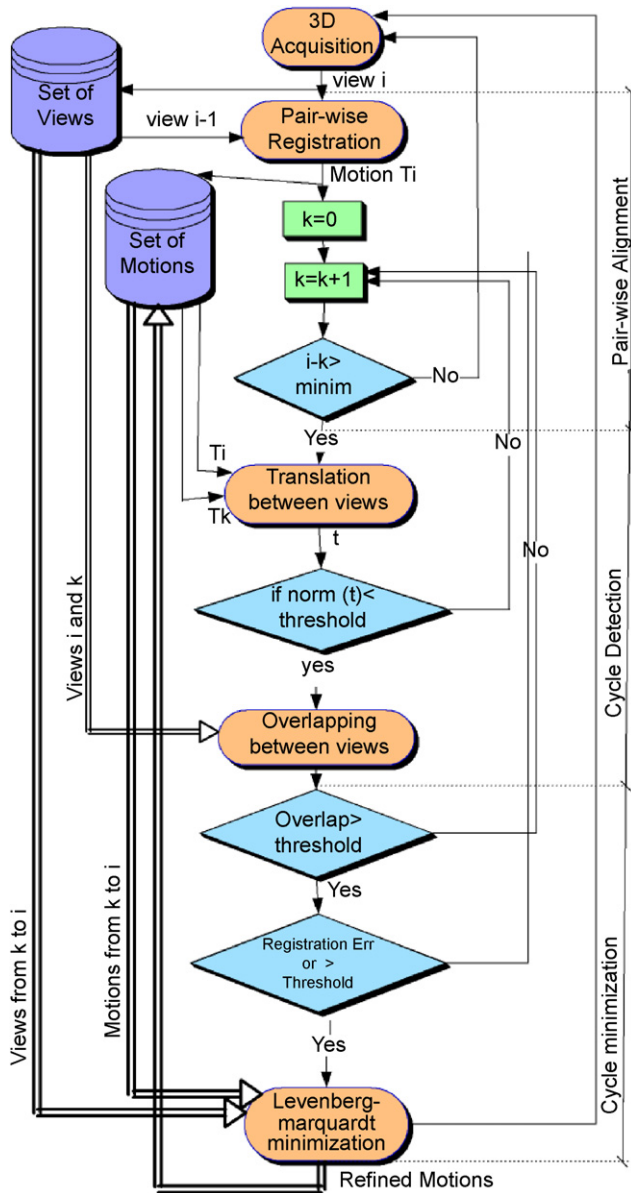
---

[1] Iterative closest point.

Fig. 1. Flow diagram of the proposed method.

points to remove becomes very easy. In the image plane, the bounding box is computed. A rectangle whose dimensions are 80% of the bounding box is centered to the image projection and all points out of this rectangle are not taken into account in the registration step.

Second, only a sample of the remaining points of the present view is preserved for the fine registration. There are several types of sampling: uniform sampling [26,32], random sampling [33], and normal sampling [16], among others. Although sampling is normally used to speed up the algorithm by selecting a reduced set of points, sampling can be also used to increase accuracy by selecting also the most appropriate points. Note that, in smooth surfaces with even shape registration becomes difficult. In this situation, only a small percentage of points give useful shape information. For instance, consider a flat surface with two perpendicular cuts. If all the points are considered in the registration, results are not accurate because of the low influence of points in cuts with respect to the rest of the points. However, if the registration is done with a high percentage of points on the uneven area, accuracy increases (see Fig. 2).

The goal of normal sampling is to select the most representative points to increase the quality of the registration. Hence, all points are first transformed to a 2D normal space defined by α and β as follows:

$$\alpha = a \tan 2 \left(n_x, \sqrt{n_z^2 + n_y^2} \right),$$

$$\beta = a \tan 2(n_y, n_z), \tag{1}$$

where α and β are the coordinates in the normal space, and n_x, n_y and n_z are the three components of the normal vector of each point. Then, every point is placed in a 2D grid. Finally only one point from every grid cell is randomly selected, so that a single point is chosen among all points with similar normal vectors. These selected points actually conform to the reduced set of points used to register the present surface. The percentage of points that remains after the sampling step depends a lot on the unevenness of such surface but tends to be between a 20% and a 30% of the total points.

As stated before, the fine registration technique we propose is based on the fast variant proposed by Park [18] from the original point-to-plane registration proposed by Chen [14]. Here we use a recursive method to compute the intersection between lines and surfaces which is actually the main difficulty of the method. Hence, initially the selected points of the previous view are projected orthographically onto the *XY* plane of the camera. A grid composed of 50×50 square cells is scaled so that it contains the projection of all points. Second, a point p_0 of the current view is projected to such a grid, in whose cell we search for the closest point obtaining the point q_{p_0} in the previous surface. The projection of point q_{p_0} to the normal vector of p_0 defines a new point p_1, which is actually an approximation of the intersection. This approximation is refined recursively by projecting new points p_i until $norm(p_i - q_{p_i})$ is smaller than a threshold (see Fig. 3). Finally, the process is repeated for all the points conforming to the current view and a set of correspondences is obtained.

by Park [18] from the original point-to-plane registration proposed by Chen [14], although some modifications have been implemented to increase accuracy, which are explained in the following paragraph.

First, we remove the non-overlapping area of the present view before this view is registered with the former. In theory, this area is unknown because the movement is also unknown. However, as the views are taken in a sequence with slight movements between them, we can assume that points located in the center of the view are good candidates for the matching. Besides, most of the points located in the boundary of the surface might be hardly matched. In consequence, the boundary area of the present view is not considered in the fine registration step. In fact, the bounding area coincides with the boundary in the image formed by projecting the present view to the *XY* plane of the camera (orthogonal to the focal axis), so the selection of
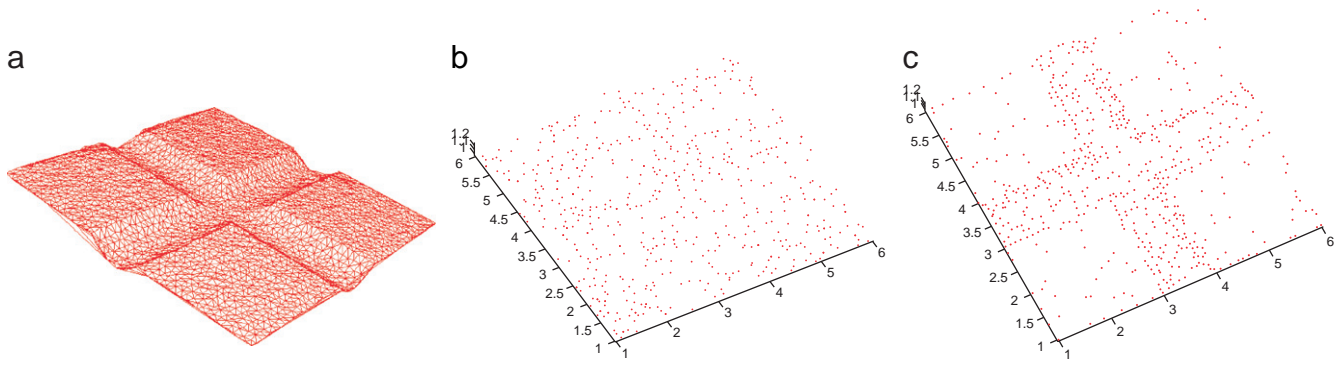
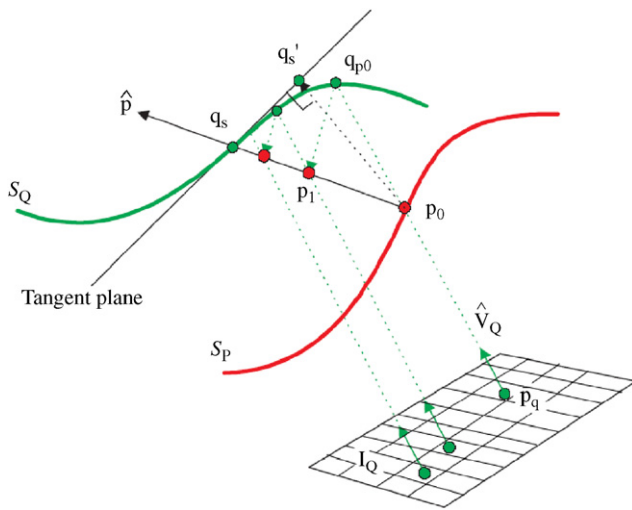Fig. 2. Effects of sampling: (a) Original surface; (b) Random sampling; (c) Normal sampling.

Fig. 3. Strategy used to compute the intersection between the tangent plane and the surface S_q along the orthogonal vector $\hat{p}$. See Park [18] for a extended review.

Once correspondences are established, minimization is applied to compute the motion between both surfaces (the previous and the current) as defined by Eq. (2).

$$f = \frac{1}{N_p} \sum_{i=1}^{N_p} \|m_i - Rp_i - t\|^2, \tag{2}$$

where N_p is the number of correspondences; m_i is the set of points selected in the former view that have a correspondence in the present view; p_i are the correspondences of m_i in the present view; and R and t are the rotation matrix and the translation vector that aligns both views, respectively.

Eq. (2) is minimized by means of quaternions [13] so that R and t are refined iteratively. In each iteration, the correspondences must be recomputed because initial correspondences are not usually correct. The algorithm stops when: the mean of the square errors (distances between correspondences) is smaller than a given threshold; or the mean of the square errors does not decrease.

Note that the views are registered consecutively, so that every registered view is referenced with respect to the first by

means of the product of all the consecutive Euclidean motions defined by the sequence of views. Hence, registration inaccuracies are propagated through the sequence. In the following sections, we aim to minimize the propagation error by detecting cycles and minimizing the views conforming the cycle all together.

3.2. Cycle detection

Now the interest is to detect every time the scanner re-visits the same object surface obtaining cycles of views that are used to reduce the propagation error significantly.

Cycle detection complexity varies depending on whether views are unorganized or views are acquired sequentially. In the former, the relationship among views is unknown and each view has to be registered to all the others to detect potential links [34]. The problem is simplified when such relationship is previously known [27]. However, in the latter, error propagation requires a robust cycle detection method. Nüchter proposed the registration of the last view to some of the already acquired views to detect such cycles [35], in which such views are selected based on some hypotheses concerning the laser range and the sensor pose.

Note that once any two views are registered, the Euclidean transformation between them is known and a link established. These links form paths through the views in which the motion of the scanner can be estimated from the product of the consecutive Euclidean transformations. Hence, the translation vector of such movement is considered, so that if this vector is smaller than a threshold and the views are not neighbors, a potential cycle is considered. The given threshold is computed dynamically considering the object scale and the number of views forming the potential cycle, so that the threshold increases proportional to the propagation error.

However, a sequence of views with an overall slight translation does not always conform a cycle, especially when rotation is relevant. Hence, the total rotation within the path is computed, so that such rotation has to be close to identity to consider the path like a cycle.

Finally, a cycle is detected if both end views also share a common surface, that is a significant overlapping area.
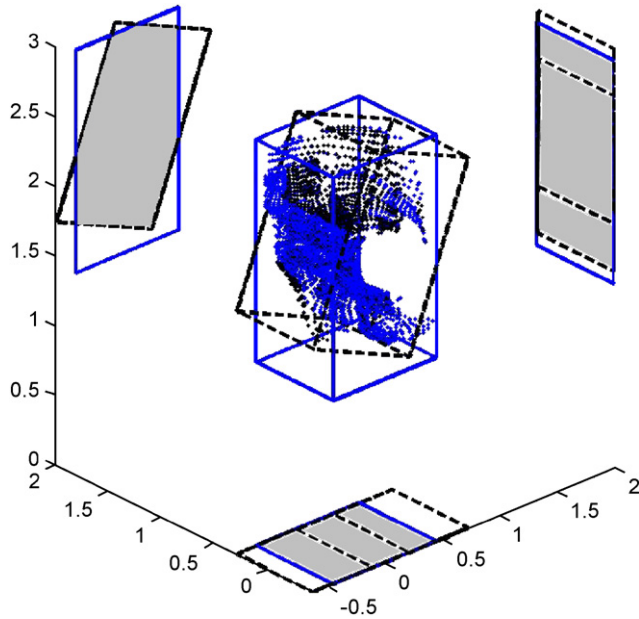
Fig. 4. Example of the projection of the bounding boxes of two different views in the *XY*, *XZ* and *YZ* planes. The grey area represents the overlapping.

The accurate computation of the percentage of overlapping would imply the fine registration between both end views and the computation of corresponding points. In order to avoid this expensive step, a fast technique is proposed based on the overlapping of bounding boxes, which is just an approximation of the convex hull of both surfaces, but accurate enough to detect cycles.

The bounding box of a given surface is defined as the minimum parallelepiped that contains all the points of the surface. The intersection of 3D bounding boxes is complex so that it is alleviated by projecting such boxes to the planes *XY*, *XZ* and *YZ* (see Fig. 4), defining two 2D bounding boxes in every plane and thus computing three overlapping areas. If the maximum of the three overlapping areas exceeds a given threshold of the total area and the distance between both bounding box centers is small enough, a cycle is considered.

The reason to choose the maximum overlapping value among the three planes instead of the product of overlapping values is in virtue of preserving the detection of potential cycles in the presence of almost flat surfaces. In this case, the bounding boxes in some of the three planes are usually not relevant.

3.3. Cycle minimization

Cycle minimization consists of a simultaneous minimization of all the correspondences between points of all the views that conform the cycle. In cycle minimization we assume that the overall motion in the cycle is null and hence the position of both end views coincides. This is actually impossible and that is the reason why a virtual view is added between both end views. This virtual view is nothing other than the first view of the cycle registered to the last one. We can assume that the overall motion in the cycle is null which means that the motion between both end views must be zero.

The significant points for every view are used to search for correspondences among all the other views in the cycle by using again our variant of the point-to-plane registration technique. This technique is based on the iterative minimization of the distances between temporal correspondences. However, at last iteration, temporal correspondences can be assumed to be accurate correspondences.

A threshold in the relative motion between views is used to ensure a significant overlapping area between views and hence many point correspondences. Obviously, this decision leads to a quite fast method without losing robustness. Otherwise, the algorithm wasted a lot of time searching for correspondences where it was known they are either not available or not significant.

Finally, a Levenberg–Marquardt minimization is applied to determine a more accurate registration among views in the cycle. The minimizing parameters are the rotation matrices (represented as quaternion vectors) and translation vectors of the Euclidean transformations between consecutive views. The minimizing function is the sum of distances between point correspondences because the distance between point correspondences should be minimized, and hopefully may reach zero, as shown in the following equation:

$$
\min \left\{ \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sum_{k=1}^{N_p} \| P_i(k) - T_j^i \times P_j(k) \| \right.
$$
$$
\left. + \| T_i^j \times P_i(k) - P_j(k) \| \right\},
\tag{3}
$$

where $P_i(k)$ and $P_j(k)$ are the points that configure the k correspondence between views i and j; N_p is the number of points correspondences; N is the number of views; and T_i^j and T_j^i are the Euclidean motions that transform points from i to j and from j to i, respectively, computed as follows:

$$
T_j^i = \prod_{k=i+1}^{j} T_k^{k-1},
\tag{4}
$$

and

$$
T_i^j = \left(\prod_{k=j}^{N-1} T_{k+1}^k \right) T_1^N \left(\prod_{k=2}^{i} T_k^{k-1} \right),
\tag{5}
$$

where $j > i$.

The closing-the-loop constraint ε_{cr} is added to the optimization function in Eq. (3), where

$$
\varepsilon_{cr} = \varepsilon_R + s_f \varepsilon_T,
\tag{6}
$$

where ε_R is the rotation constraint; ε_T is the translation constraint; and s_f is the scale factor that weights the translation constraint to be adequately compared to the rotation constraint.

The rotation constraint ε_R is

$$\varepsilon_R = sum(abs(R_{accum} - I_{3\times3})), \tag{7}$$

where R_{accum} is the product of all the rotation matrices conforming the cycle, and $I_{3\times3}$ is the identity matrix.

The translation constraint ε_T is

$$\varepsilon_T = norm(t), \tag{8}$$

where t is the translation vector between both initial and end views of the cycle, computed as follows:

$$\begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} = T_{cycle} = \left(\prod_{i=2}^{n} T_i^{i-1} \right) \cdot T_n^1. \tag{9}$$

The whole process leads to quite accurate results, but if they are not good enough, they can be repeatedly refined by selecting new significant point correspondences at the end of every refinement.

4. Experimental results

The proposed method has been implemented and compared to one of the most similar methods present in the literature, the method proposed by Sharp et al. in 2004 [28]. Both methods have been tested under the same conditions in order to evaluate their advantages and constraints. So, we have used the same point-to-plane method to determine the motion between consecutive views.

A low-cost multi-slit laser acquisition system has been used. The set-up is composed of an off-the-self camera, a 635 nm laser emitter and an optical lens that spreads the laser beam conforming 19 parallel planes. The depth field of the system is in the range of 100 to 300 mm in the Z-axis of the camera, which is limited due to system baseline, laser power, and camera focus. The camera and the laser emitter conform the one-shot hand-held 3D acquisition system especially developed to test our proposal. The 19 parallel planes are projected onto the measuring surface obtaining 19 profiles with subpixel accuracy. 3D profiles are obtained by triangulation and finally splines are used to interpolate a surface, increasing the sample of points used in the registration process. Once the 3D surface is acquired, the Delaunay triangulation is applied to obtain triangles and hence estimate the normal vector at every surface point which is used further on in the pair-wise alignment. Fig. 5 shows an acquired image and the process of spline interpolation. The reader is pointed to Matabosch [36] for more details about the acquisition sensor.

The performance of our method has been compared to the method of Sharp [28] both quantitatively and qualitatively. Our cycle detection method has been used in both methods, so they are tested in the presence of the same number of cycles. Note that in the method of Sharp cycles are detected manually. Experiments and results are presented in the following paragraphs.

4.1. Quantitative evaluation

Quantitative evaluation is analyzed from both synthetic and real data. Synthetic data is obtained from the 3D synthetic models courtesy of INRIA.[2] A synthetic scanner has been programmed to simulate the acquisition of a set of consecutive views acquired by the one-shot hand-held scanner (see Fig. 6). Here, the pose of the scanner for every acquisition is obviously given by the simulator. So, the accuracy of the registration can be precisely evaluated. The experiment is repeated adding some Gaussian noise to the 3D points. In addition, a real object has been placed on a motion-controlled table in which our one-shot hand-held scanner has been attached (see Fig. 7). The object is moved in three degrees of freedom (X, Y and Z) so that 29 consecutive views are acquired. In this case, the position of the sensor is given by the mechanics of the table.

Parameters and thresholds used during this experiment are reported in Table 1. Their value differs depending on the size of the measuring object. NSS_{grids} specifies the number of cells used in the normal space sampling grid. $Intersection_{error}$ is the threshold used in the point-to-plane registration and it corresponds to the maximum value of $norm(p_i - q_i)$, so that a small value increases registration accuracy but decreases the number of correspondences, specially in case of important misalignments. Concerning cycle detection, the first parameter fixes the minimum number of views to consider a potential cycle. Translation error specifies in millimeters the maximum distance between both end-views in a cycle to be considered a potential cycle. Overlapping area specifies the minimum overlapping required between both end-views of a potential cycle. Finally, a scale factor is introduced in the cycle minimization step to normalize translation and rotation errors.

In order to provide quantitative evaluation, the motion estimated by registration has to be compared to the real motion provided by the simulator or the 3D table in terms of both translation and rotation. Rotation is represented as a directional vector which can be easily extracted from any rotation matrix. Then, the estimated rotation is compared to the known one. So, the error is determined as the norm of the difference between both axes of rotation. The translation error is defined as the distance between the origin of the coordinate system estimated by registration with respect to the known origin, which is the norm of the difference between both translation vectors.

Additionally, the MSE (mean squared error) is computed. For each point of the registered acquisition, the nearest point in the set composed by the rest of acquisitions is found, determining a correspondence. The mean of all distances give us the estimation of the discrepancy between registered views.

Finally, our proposed method and Sharp's method are both compared to the precise alignment in terms of accuracy so that quantitative results are reported in Table 2 and Fig. 8. In Table 2 we have also included a fast variant of our approach. This fast variant differs only in the cycle minimization step, so that the minimization uses the correspondences previously obtained by the pair-wise registration avoiding the search for new

---

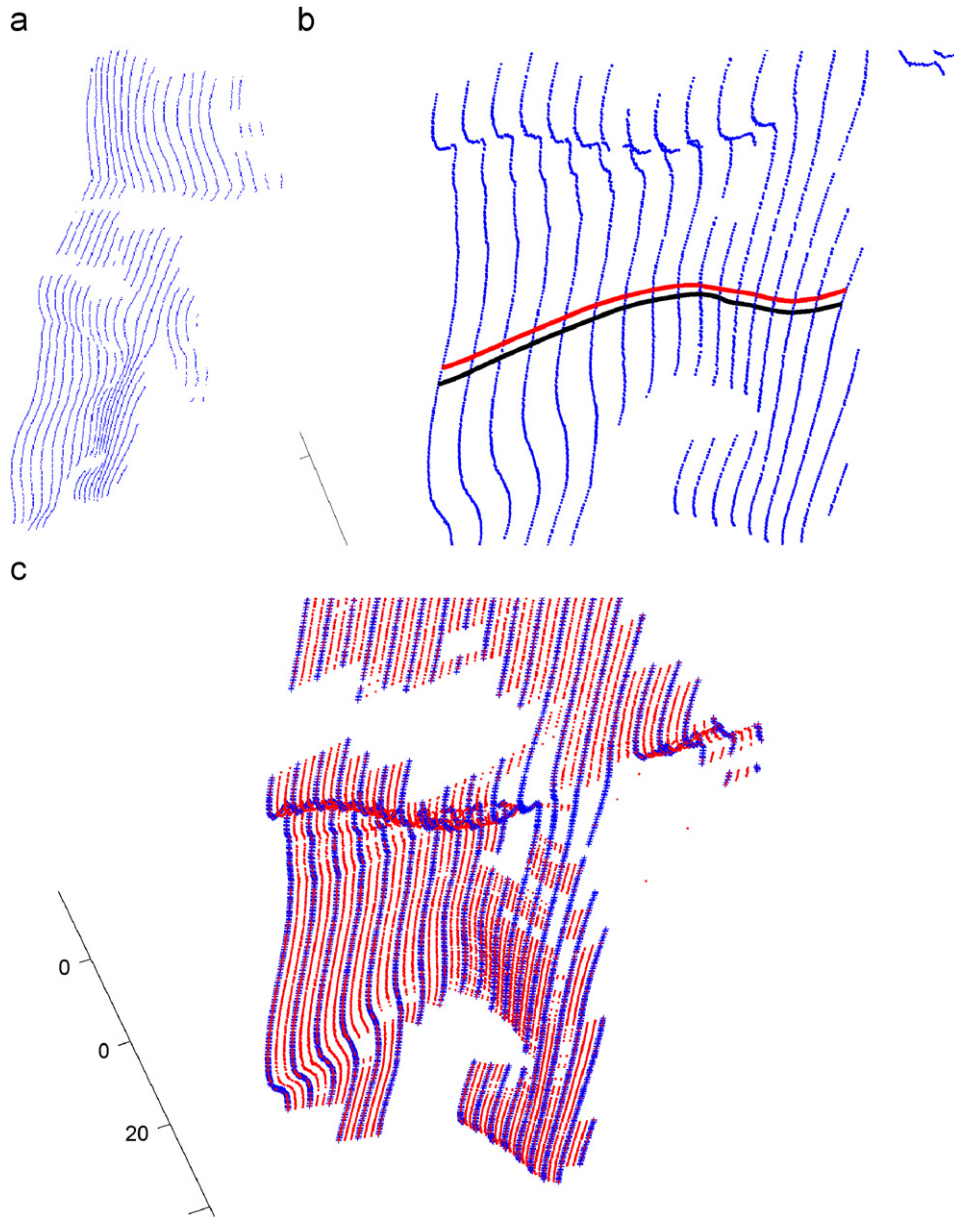[2] http://www-c.inria.fr/gamma/download/download.php

Fig. 5. Acquisition examples: (a) Acquired cloud of points from the 19 slits; (b) Spline curve computation (in blue the acquired profiles, in red and black two samples of splines); (c) Cloud of points obtained after spline sampling (in blue the original points, in red the new points computed).

matching among all the views in the cycle. Actually, matching is one of the most expensive steps. So, the fast variant approach consumes a computing time similar to Sharp's approach but preserves a good accuracy, as shown in Table 2.

Table 2 shows for every experiment the mean and the standard deviation computed from the set of rotation and translation errors. Note that both methods obtain similar results when acquisition noise is unimportant. When the acquisition noise becomes significant, pair-wise registration is not accurate enough and hence Sharp's approach distributes a large error in the cycle. Besides, it is shown that our proposal obtains better results in both synthetic and real data. A special attention requires the results obtained when noise was 1.25% and 3.75%. In both cases, pair wise registrations between both end-views of cycles

were not accurate enough. Such inaccuracies produce significant errors in the method of Sharp that are minimized by our method thanks to the multi-view registration performed inside every cycle.

Fig. 8 presents the MSEs after register all views with our robust method and Sharp's method. Cycle detection algorithm determines a cycle between views 1 and 23. Our robust method obtained better results on these views. Sharp's method obtains good results results in the last view of the cycle, due to the closing-the-loop constrain. However, inside the cycle, propagation error is not always correctly distributed through the views. After view 23, all methods obtain the same results, because there is no cycle, and results are directly obtained from the pair-wise registration.
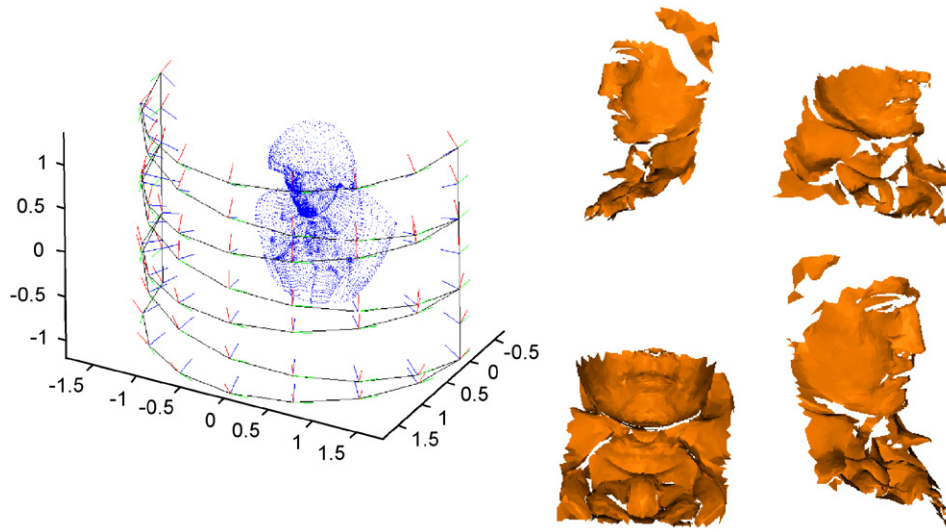
Fig. 6. Left: Path described by the simulator to scan a synthetic object (Beethoven). Right: Some of the acquired synthetic images.
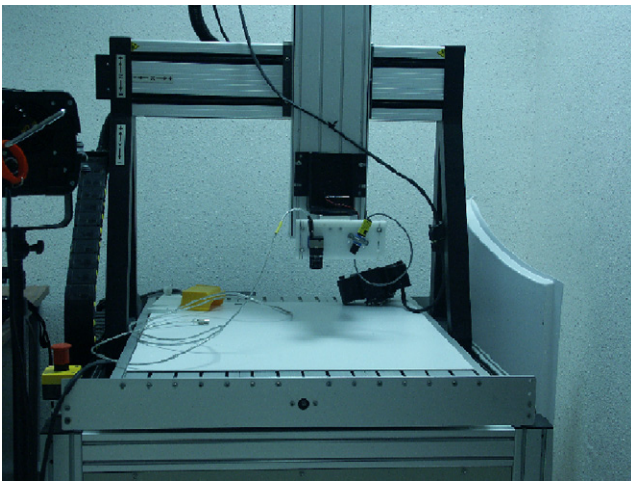

Fig. 7. Accurate motion-controlled table used in the quantitative evaluation.

Table 1
Quantitative experiment settings

Steps	Parameter	Beethoven	Sun/moon
Pair-wise	NSS_{grids}	2500	2500
	$Intersection_{error}$	0.3	6.0
Cycle detection	Minimum number views	8.0	8.0
	Translation error	7.0	15.0
	Overlapping area	50%	50%
Cycle minimization	Sf	1.0	0.1

4.2. Qualitative evaluation

In order to evaluate the performance of the methods, it is also useful to observe the registration of a real object and analyze it from a qualitative point of view. In this experiment,

the one-shot hand-held scanner is coupled to a FANUC industrial manipulator. The manipulator describes a trajectory so that a given object is scanned obtaining a sequence of views. As the kinematics of the manipulator is known, the views can be aligned without applying any registration and hence such raw alignment is provided for comparison.

Note that the kinematics of the manipulator provides the position of the robot hand H with respect to the coordinate frame of the robot base R (see Fig. 9). Besides, registration is referenced with respect to the frame S of the camera of the one-shot hand-held scanner. The rigid transformation between H and S is unknown and hence has to be first estimated.

The computation of $^{H}T_{S}$ is known as the eye-to-hand problem in the robotics community and it is based on solving equation $AX = XB$, where X is the matrix we are looking for. So, X transforms points from the coordinate frame of the scanner S to the coordinate frame of the hand H, A is the motion of the hand between two different positions of the robot given by the robot control system, and B is the motion computed by means of triangulating the movement in the image of the one-shot hand-held scanner.

There are several papers addressing the computation of $AX = XB$ [37,38]. In our case, we have acquired 10 views of a calibrating pattern and the X matrix is estimated by using the algorithm of Shiu [38]. First, the algorithm determines a set of A and B matrices from every view. Then, a system of equations with the form $AX - XB = 0$ is defined and solved. Theoretically X can be computed with only three views, though it is more accurate to solve the equation of an over-determined system by using singular value decomposition.

Once X is known, all views can be represented in the same reference using the following equation:

$$^{W}T_{S} = {}^{W}T_{R} \times {}^{R}T_{H} \times X, \tag{10}$$

Table 2
Quantitative results

Scene	Our method			Fast variant			Sharp's method		
	$error_R$	$error_T$	MSE	$error_R$	$error_T$	MSE	$error_R$	$error_T$	MSE
$\sigma = 0$	0.516	0.008	0.003	0.339	0.079	0.001	0.511	0.074	0.002
	1.120	0.004	0.003	0.867	0.191	0.006	1.006	0.034	0.002
$\sigma = 1.25\%$	0.675	0.154	0.004	1.177	0.459	0.006	2.225	4.403	0.026
	1.115	0.305	0.005	1.388	0.265	0.006	1.385	1.196	0.051
$\sigma = 2.5\%$	1.1286	0.4698	0.005	1.202	0.316	0.002	1.472	1.367	0.001
	1.1905	0.2149	0.005	1.410	0.217	0.006	1.202	0.704	0.002
$\sigma = 3.75\%$	0.246	0.056	0.003	1.552	0.875	0.007	2.601	3.485	0.026
	0.732	0.024	0.002	1.169	0.425	0.006	1.134	2.106	0.046
$\sigma = 5.0\%$	1.570	0.890	0.005	1.533	0.828	0.007	2.753	3.126	0.017
	1.284	0.682	0.005	1.144	0.497	0.006	1.212	2.225	0.020
	1.2804	5.063	0.334	1.3485	5.4103	0.389	1.3863	4.640	0.432
	0.303	2.459	0.335	0.285	2.425	0.529	0.291	2.308	0.380

Both our original method and its fast variant are compared to the method of Sharp: $error_R$ is the norm of the difference between both axes of rotation; $error_T$ is the norm of the difference between both translation vectors (distance between the points of origin of both coordinate systems); MSE is the mean squared error. Every table cell indicates the mean (up) and standard deviation (down) of the error for a set of synthetic experiments varying the Gaussian noise (σ) and one experiment with real data. The synthetic object (Beethoven) consists in 48 views composed of about 5000 points per view and with the presence of 6 detected cycles. The real object (sun/moon) consists in 27 views composed of about 8000 points per view and with the presence of 1 detected cycle between views 1 and 23.
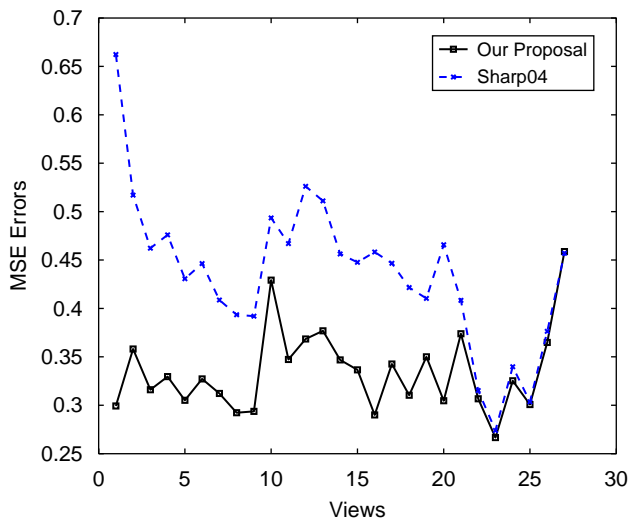

Fig. 8. Evolution of the MSE registration errors in the registration of the real object (sun/moon). Scale of the measured object: 180 mm (width) × 200 mm (height) × 56 mm (depth).
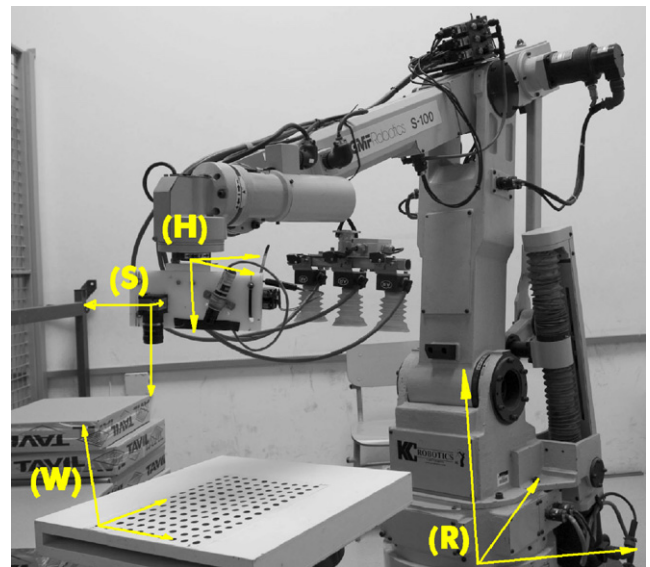

Fig. 9. Industrial manipulator used in experiments. The four coordinate frames are represented: W (world), R (robot), H (Hand) and S (Scanner).

where $^W T_S$ is the Euclidean motion that transforms points in S to the world coordinate system W (used by the one-shot hand-held scanner to refer 3D points), $^W T_R$ is the Euclidean motion that relates the world coordinate system W to the robot base R, $^R T_H$ is the motion given by the kinematics of the robot arm, and X is the Euclidean transformation between the camera of the one-shot hand-held scanner and the robot hand.

Now we can proceed with the experiment. The parameters of that experiment are presented in Table 3. The manipulator has been programmed so that an 8-shape trajectory is done over a ceramic object acquiring up to 41 images and hence 41 3D partial views of the object. Note that the trajectory ensures cycles which will be used in the registration. First, all the views are referenced with respect to the same frame by means of the X

Table 3
Qualitative experiment settings

Steps	Parameter	Value
Pair-wise	NSS_{grids}	2500
	$Intersection_{error}$	6.0
Cycle detection	Minimum number views	8.0
	Translation error	15.0
	Overlapping area	50%
Cycle minimization	Sf	0.1

Table 4
Registration time

Method	Time (s)
Robust	1192.47
Fast	900.69
Sharp	642.66

5. Conclusions

There are several techniques to register a set of views, though most of them are based on the multi-view approach. In general, multi-view techniques are constrained by the following drawbacks: (a) all the views must be first acquired before the aligning algorithm starts leading to off-line applications; (b) guesses to roughly align the views are needed to initialize the algorithm so that an expensive coarse registration technique is needed; and (c) matching is searched among all the views without considering neighborhood which is inefficient and computing intensive, especially in large data sets. Besides, multi-view techniques are not suitable for registering views that form sequences and loops because of the error propagation problem.

This paper presents a new multi-view registration technique which includes cycle minimization and it is updated in the measure that new views are acquired. Although the technique can be applied in short sequences of views, it is designed to deal with large data sets and with the presence of multiple cycles. First, a fast point-to-plane with normal space sampling and non-overlapping area removal is applied between consecutive views to obtain an accurate alignment. Second, in the measure that new views are acquired, the method searches for cycles considering neighborhood and overlapping percentage. Finally, once a cycle is detected it is minimized by means of a Levenberg–Marquardt approach, so that the system always ensures the most accurate global registration.

matrix. Second, a volumetric integration algorithm is applied to get a continuous surface [39]. Third, the sequence of views are aligned according to: (a) the registration algorithm proposed in this article; (b) the multi-view algorithm proposed by Sharp [28]; and (c) the kinematics of the robot. Finally, any surface smooth technique is applied to enhance the visualization. Qualitative results are shown in Fig. 10. Registration really improves the alignment provided by the kinematics of the robot. Note that the alignment directly obtained from the kinematics of the robot suffers not only from inaccuracies given by the mechanics but especially inaccuracies in the computation of *X*. Besides, the experiment also shows that our approach provides a surface with more details and less artefacts compared to the method proposed by Sharp. Registration time is presented in Table 4. These experiments are performed using Matlab 6.5 in a pentium IV 2.6 GHz.

The acquisition system obtains the structure of the surface with a single-shot and hence vibrations produced by a human operator do not affect the acquisition. Actually, there is no difference in the obtained results performing the experiments by a manipulator or by a human operator.
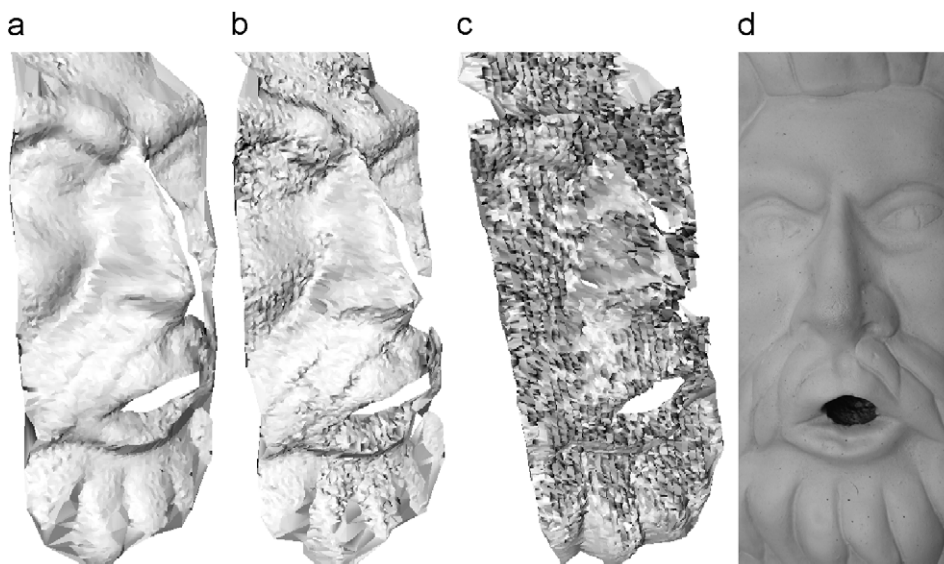

Fig. 10. Results of the registration: (a) Our method; (b) Sharp's method; (c) Mechanical alignment; (d) Real object.

Experiments with both synthetic and real data have been overcome. Synthetic data has been downloaded from well-known databases. A one-shot hand-held scanner composed of a camera and a multi-slit laser emitter has been developed to acquire real data. The scanner has been coupled to a commercial manipulator to acquire sequences of views. Our approach has been compared to: (a) the mechanical alignment provided by the kinematics of the manipulator; and (b) the multi-view alignment method proposed by Sharp [28], which from our point of view is one of the most similar. Results show from both a quantitative and a qualitative point of view that our approach provides a more accurate alignment.

References

[1] J. Forest, J. Salvi, An overview of laser slit 3D digitasers, in: Proceedings of the International Conference on Intelligent Robots and Systems, Lausanne, October 2002, pp. 73–78.

[2] J. Forest, J. Salvi, E. Cabruja, C. Pous, Laser stripe peak detector for 3D scanners. A FIR filter approach, in: International Conference on Pattern Recognition, Cambridge, August 2004, pp. 646–649.

[3] J. Salvi, J. Pagès, J. Batlle, Pattern codification strategies in structured light systems, Pattern Recognition 37 (4) (2004) 827–849.

[4] A. Bottino, A. Laurentini, Shape-from-silhouettes with unknown relative position of the viewpoints: investigating a new problem, in: Signal Processing, Pattern Recognition, and Applications (SPPRA 2002), June 2002.

[5] Z. Zhang, Q.T. Luong, O. Faugeras, Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction, Technical Report RR-2079, 1993.

[6] Zcorporation. zscanner700, in: ⟨http://www.zcorp.com/products/zscanner700.asp?ID = 1⟩, 2007.

[7] J. Salvi, C. Matabosch, D. Fofi, J. Forest, A review of recent range image registration methods with accuracy evaluation, Image Vision Comput. 25 (5) (2007) 578–596.

[8] I. Stamos, M. Leordeanu, Automated feature-based range registration of urban scenes of large scale, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, June 2003, pp. 555–561.

[9] J. Vanden Wyngaerd, Automatic crude patch registration: toward automatic 3d model building, Comput. Vision Image Understand. 87 (2002) 8–26.

[10] C.S. Jarvis, R. Chua, Point signatures: a new representation for 3d object recognition, Int. J. Comput. Vision 25 (1) (1997) 63–85.

[11] A. Johnson, M. Hebert, Surface registration by matching oriented points, in: International Conference on Recent Advances in 3-D Digital Imaging and Modeling, May 1997, pp. 121–128.

[12] N. Gelfand, N.J. Mitra, L.J. Guibas, H. Pottmann, Robust global registration, in: Symposium on Geometry Processing, 2005, pp. 197–206.

[13] 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.

[14] G. Chen, Y. ad Medioni, Object modeling by registration of multiple range images, in: IEEE International Conference on Robotics and Automation, April 1991, pp. 2724–2729.

[15] P. Neugebauer, Geometrical cloning 3d objects via simultaneous registration of multiview range images, in: Shape Modeling and Applications, 1997, pp. 130–139.

[16] S. Rusinkiewicz, M. Levoy, Efficient variant of the icp algorithm, in: Third International Conference on 3-D Digital Imaging and Modeling, 2001, pp. 145–152.

[17] H. Gagnon, M. Soucy, R. Bergevin, D. Laurendeau, Registration of multiple range views for automatic 3-d model building, in: Computer Vision Pattern Recognition, June 1994, pp. 581–586.

[18] S.-Y. Park, M. Subbarao, A fast point-to-tangent plane technique for multi-view registration, in: 3DIM, Fourth International Caonference on 3D Digital Imaging and Modeling, 2003, pp. 276–284.

[19] E. Batlle, C. Matabosch, J. Salvi, Summarizing image/surface registration for 6dof robot/camera pose estimation, in: Third Iberian Conference on Pattern Recognition and Image Analysis, Girona, Spain, June 2007, pp. 105–112.

[20] K. Pulli, Multiview registration for large data sets, in: 3-D Digital Imaging and Modeling, Ottawa, Ont., Canada, 1999, pp. 160–168.

[21] A. Nüchter, H. Surmann, K. Lingemann, J. Hertzberg, S. Thrun, 6d slam with an application in autonomous mine mapping, in: IEEE International Conference on Robotics and Automation, ICRA '04, vol. 2, May 2004, pp. 1998–2003.

[22] R. Bergevin, M. Soucy, H. Gagnon, D. Laurendeau, Towards a general multiview registration technique, Pattern Anal. Mach. Intell. 18 (5) (1996) 540–547.

[23] L. Silva, O.R.P. Bellon, K.L. Boyer, Multiview range image registration using the surface interpenetration measure, Image Vision Comput. 25 (1) (2006) 114–125.

[24] D. Huber, M. Hebert, Fully automatic registration of multiple 3d data sets, Image Vision Comput. 21 (7) (2003) 637–650.

[25] S. Krishnan, P.Y. Lee, J.B. Moore, S. Venkatasubramanian, Global registration of multiple 3d point sets via optimization-on-a-manifold, in: Symposium on Geometry Processing, 2005, pp. 187–196.

[26] T. Masuda, Generation of geometric model by registration and integration of multiple range images, in: Third International Conference on 3-D Digital Imaging and Modeling, May 2001, pp. 254–261.

[27] F. Lu, E.E. Milios, Globally consistent range scan alignment for environment mapping, Auton. Robots 4 (4) (1997) 333–349.

[28] G.C. Sharp, S.W. Lee, D.K. Wehe, Multiview registration of 3d scenes by minimizing error between coordinate frames, IEEE Trans. Pattern Anal. Mach. Intell. 26 (8) (2004) 1037–1050.

[29] B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon, Bundle adjustment—a modern synthesis, in: International Workshop on Vision Algorithms, London, UK, 2000, pp. 298–372.

[30] A. Fitzgibbon, Robust registration of 2d and 3d point sets, in: The British Machine Vision Conference, 2001, pp. 662–670.

[31] M. Pollefeys, M.R. Koch, M. Vergauwen, L. Van Gool, Automated reconstruction of 3d scenes from sequences of images, Photogrammetry and Remote Sensing 55 (2000) 251–267.

[32] G. Turk, M. Levoy, Zippered polygon meshes from range images, in: SIGGRAPH '94: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, Orlando, FL, July 1996, pp. 311–318.

[33] T. Masuda, K. Sakaue, N. Yoyoka, Registration and intergration of multiple range images for 3-d model construction, in: ICPR '96: Proceedings of the 1996 International Conference on Pattern Recognition, vol. 1, Washington, DC, USA, August 1996, pp. 879–883.

[34] D.F. Huber, Automatic 3d modeling using range images obtained from unknown viewpoints, in: International Conference on 3D Digital Imaging and Modeling, May 2001, pp. 153–160.

[35] A. Nüchter, K. Lingemann, J. Hertzberg, H. Surmann, Heuristic-based laser scan matching for outdoor 6d slam, in: 28th German Conference on Artificial Intelligence, 2005, pp. 304–319.

[36] C. Matabosch, J. Salvi, D. Fofi, F. Meriaudeau, A refined range image registration technique for multi-stripe laser scanner, in: Machine Vision Applications in Industrial Inspection XIV, vol. 6070, SPIE, January 2006, pp. 1–8.

[37] I. Fassi, G. Legnani, Hand to sensor calibration: a geometrical interpretation of the matrix equation ax = xb, J. Robot. Systems 22 (9) (2005) 497–506.

[38] Y.C. Shiu, S. Ahmad, Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form $ax = xb$, IEEE Trans. Robot. Autom. 5 (1) (1989) 16–29.

[39] B. Curless, M. Levoy, A volumetric method for building complex models from range images, in: SIGGraph-96, 1996, pp. 303–312.

About the Author—CARLES MATABOSCH received the degree in industrial engineering from the University of Girona, Spain in 2002, and received the D.E.A. degree in computer science in October 2004. He is currently studying toward the Ph.D. degree in computer vision, also at the University of Girona, under the supervision of Dr. J. Salvi. His research interests include 3D reconstruction, laser projection and multi-view registration.

About the Author—DAVID FOFI studied electric and industrial informatics engineering as an undergraduate at the University of Picardie Jules Verne, Amiens and received the M.Sc. degree in Image & Signal Processing from the University of Cergy-Pontoise/ENSEA in 1997. He completed the Ph.D. degree in the area of coded structured light in 2001 at the University of Picardie Jules Verne. He is currently an Associate Professor at the University of Burgundy. His main research interests are in 3D vision, active vision, camera-projector systems, coded and structured light.

About the Author—JOAQUIM SALVI was graduated in computer science from the Technical University of Catalonia, Catalonia, Spain in 1993, and received the D.E.A. degree in computer science in July 1996, and the Ph.D. degree in industrial engineering in 1998, from the Computer Vision and Robotics Group, University of Girona, Girona, Spain. He is currently an associate professor with the Electronics, Computer Engineering and Automation Department, University of Girona. He is involved in some governmental projects and technology transfer to industrial environments. His current interests are in the field of computer vision and mobile robotics, focused on structured light, stereovision, and camera calibration. He is the leader of the 3-D Perception Lab. Dr. Salvi received the Best Thesis Award in engineering for his Ph.D. dissertation.

About the Author—ELISABET BATLLE received her B.S. degree in computer science from the University Autonoma of Barcelona in 2003 before joining the Computer Vision and Robotics Group at the University of Girona. She received her D.E.A. degree in computer science in June 2002 and she is currently studying toward her Ph.D. degree in computer vision, at the University of Girona. She is also an Assistant Professor with the Electronics, Computer Engineering and Automation Department, University of Girona. Her research interests include computer vision, 3D reconstruction, robotics, SLAM and multi-view registration.