An innovative hand-held vision-based digitizing system for 3D modelling

Benjamin Coudrin, Michel Devy, Jean-José Orteu, Ludovic Brèthes

An innovative hand-held vision-based digitizing system for 3D modelling

Benjamin Coudrin[a,b,c,d,*], Michel Devy[b,c], Jean-José Orteu[d], Ludovic Brèthes[a]

[a] *NOOMEO ; rue Galilée, BP 57267, 31672 Labège CEDEX, France*
[b] *CNRS ; LAAS ; 7 avenue du colonel Roche, F-31077 Toulouse Cedex 4, France*
[c] *Université de Toulouse ; UPS, INSA, INP, ISAE ; UT1, UTM, LAAS ; F-31077 Toulouse Cedex 4, France*
[d] *Université de Toulouse ; Mines Albi ; Institut Clément Ader (ICA) ; Campus Jarlard, F-81013 Albi, France*

Abstract

We describe a new hand-held 3D modelling device using vision and inertial measurements. Our system allows fast and accurate acquisition of the geometry and appearance information of any 3D object. We focused our work towards an easy manipulation and operating condition.

Our methods allow automatic registration with no preparation of the scene (i.e. no markers) even when the object is moved between two acquisitions.

In this paper, the design of the system and the developed methods for its use are presented. The system has been evaluated, qualitatively and quantitatively, using reference measurements provided by commercial scanning devices. The results show that this new hand-held scanning device is really

*Tel : (+33)5.61.00.77.18 ; Fax : (+33)9.74.76.02.52
Email addresses: benjamin.coudrin@noomeo.eu (Benjamin Coudrin), michel.devy@laas.fr (Michel Devy), jean-jose.orteu@mines-albi.fr (Jean-José Orteu), ludovic.brethes@noomeo.eu (Ludovic Brèthes)

competitive for modelling any 3D object.

Keywords: 3D digitizing; 3D modelling; hand-held; registration

1. Introduction

3D Object Modelling has generated many research works in the last decades, especially for Non Destructive Testing or Quality Control applications. Several systems can be found off-the-shelf, either based on vision [1] or on laser ranging [2]. New applications where 3D modelling must be performed by non expert people are arising, for instance for archeologists, curators or artists who want to model art objects. Moreover, several potential mass market applications are emerging, e.g. 3D modelling from smart devices either embedded on assistance robots or moved by hand around an object.

These new applications involve several constraints. In particular, the 3D modelling task is often executed in a human environment, typically at home, where it is better to avoid laser-based technologies. Moreover the system must be user-friendly, lightweight to be hand-held (i.e. compact), operated with a minimal preparation of the environment (i.e. no marker), and cheaper than existing systems. In that context, Vision, generally associated with illuminators [3, 4], is the better way to acquire sensory data. Some hand-held systems exist in the industrial community. Mostly these systems require to place a reference in the scene. Some use magnetic references or photometric targets. This leads to a more complex digitization process and can forbid to move the object during the process.

This paper presents a new Vision-based hand-held 3D scanner allowing acquisition of geometry and texture from an object without any scene prepa-

ration. The object is first represented by clouds of 3D points and then with a textured mesh. Digitizing 3D objects requires a sequential process of geometric modelling from visual data acquired from camera(s) moved by a robot or by an operator around the object. 3D reconstruction gives a 3D point cloud from each acquisition. Clouds of 3D points are fragmented representations of space that need to be merged. This imposes to know the shooting positions of the scanner. Our scanner being hand-held, we have no *a priori* model for this motion, like we would have if the scanner was held by a robot or if we had placed targets in the scene.

Several 3D registration algorithms have been proposed since a long time in order to merge 3D data obtained from several viewpoints. Besl *et al.* introduced the main method used to estimate the rigid transformation between two clouds of 3D points, the so-called ICP algorithm (*Iterative Closest Point*) [5]. This method has since been adapted with numerous variations [6] in order to speed up the convergence and decrease the complexity. Sandhu *et al.* [7] recently combine ICP with a particular filtering approach. Li *et al.* [8] improved the robustness, providing a guaranteed registration. Our registration method combines several operations, based on IMU (Inertial Motion Unit) measurements, ICP-based algorithms and visual odometry from tracked interest image-points.

The aim of this new 3D modelling system is to provide an easy to use device. This means that we focused our work on providing reliable and mostly automatic operation. Being hand-held, and lightweight, the device is easily moved and can access a large range of view points, allowing to scan complex shape objects in details or in difficult access conditions. Since no preparation

is required, the overall modelling process has been made shorter and the skill level required for the operator has been reduced. Moreover, the system only requires to be plugged to a computer, which allows to use it with a laptop in outdoor scanning tasks for example.

Next section is devoted to the sensor description. Then section 3 presents the main modelling steps: image acquisition, 3D reconstruction and 3D registration. In section 4, several models built from our approach are evaluated, from comparisons either with the actual models for known objects, or with models built with commercial scanning systems. Finally, section 5 presents some outlooks for our future works.

2. Description of the system

This new vision-based digitizing system is principally composed of two cameras used as a stereovision pair. 3D data is obtained by triangulating image points between a pair of images. Matching pixels in areas of homogeneous intensity in images is impossible. To tackle this problem, we use a projection of a speckle pattern to give the observed scene a suitable random texture.

At a given instant, the system can only acquire a partial view of the scene due to the field of view of the cameras. Modelling a scene from a hand-held vision-based system requires to aggregate these partial views acquired at several instants. This is the problem of estimating the pose or the motion of the digitizing system throughout time. To help us in this task, the digitizing system has been equipped with an inertial motion unit (IMU) able to measure the dynamic or the attitude of the digitizing system.

The system weights 1.8 kg, its dimensions are $220 \times 240 \times 90$ mm. Figure 1 describes our system and the frames associated with the sensors.

[Figure 1 about here.]

2.1. Cameras

The digitizing system is mainly built around two cameras. The cameras are equipped with a global shutter CCD sensors with a 1024×768 resolution. They can operate up to a 39 Hz maximum frame rate. However, using a frame rate close to the maximum decreases image quality.

The first camera, placed at the bottom of the system, is called the *coaxial camera* because it shares its frame with the absolute frame chosen for the system. The *coaxial frame* is noted $\mathcal{C}_0$.

The second camera, placed at the top of the system, is the *lateral camera*. Its frame is noted $\mathcal{C}_1$. We suppose that the transformation between the *lateral* and the *coaxial* frames is known[1] and noted $\mathbf{H}^{\mathcal{C}_1\mathcal{C}_0}$, such that :

$$\begin{bmatrix} p^{\mathcal{C}_1} \\ 1 \end{bmatrix} = \mathbf{H}^{\mathcal{C}_1\mathcal{C}_0} \begin{bmatrix} p^{\mathcal{C}_0} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} p^{\mathcal{C}_0} \\ 1 \end{bmatrix} \tag{1}$$

where $\mathbf{R}$ and $\mathbf{t}$ are respectively the rotation matrix and translation vector relating the *coaxial frame* and the *lateral frame*.

It should be noted that, to guarantee the best precision in matching and to avoid noise, the cameras need to be perfectly synchronized during the image acquisition. This is achieved by using a hardware trigger signal.

[1] The stereo-vision system has been calibrated using a chessboard calibration target, following the method described in [9]. The chessboard has been modified to allow automatic initialization.

2.2. Inertial Motion Unit

The IMU is composed of three accelerometers, three gyrometers and three magnetometers. It can operate up to approximately 120 Hz and gives information about the dynamics of our hand-held system :

- acceleration $\mathbf{a} = [a_x \ a_y \ a_z]^T$
- angular rate $\omega = [\omega_x \ \omega_y \ \omega_z]^T$
- surrounding magnetic field direction $\mathbf{b} = [b_x \ b_y \ b_z]^T$

The integration of acceleration and angular rate measurements can give static information – position, attitude – to a certain extent. Indeed, due to important noise level and biases, position and attitude can drift quickly. It is yet possible to compose all the dynamics measurements to obtain a relatively accurate estimation of the attitude of the *IMU* using a Kalman filter [10, 11].

Magnetic field is static and consequently its measurement does not drift incrementally. Accelerometers measure all accelerations applied to them, including gravity acceleration. Thus, we are able to express current measurements $\mathbf{g}_k$ and $\mathbf{l}_k$ – respectively current gravity acceleration vector and current magnetic field direction – with respect to the current estimated attitude.

$$\mathbf{g}_k = R\left(\mathbf{e}_k\right)\mathbf{g}_0 \tag{2}$$

$$\mathbf{l}_k = R\left(\mathbf{e}_k\right)\mathbf{l}_0 \tag{3}$$

where $\mathbf{g}_0 = [0 \ 0 \ g]^T$ is the gravity acceleration vector expressed in world frame, $\mathbf{l}_0$ is the measurement of the magnetic field when the inertial sensor

is aligned with world frame, $\mathbf{e}_k$ is the current attitude expressed in Euler angles, and $R(\mathbf{e})$ is the rotation matrix composed from euler angles vector $\mathbf{e}$.

Angular rate measurements $\omega_\mathbf{k}$ are combined at the prediction step of the Kalman Filter by integration at high rate. Indeed, locally the estimation does not drift much and the system is considered linear.

$$\dot{\mathbf{e}}_k = \omega_\mathbf{k} \tag{4}$$

Inertial motion unit frame is noted $\mathcal{S}$. Transformation between the *coaxial frame* $\mathcal{C}_0$ and the *inertial frame* is noted $\mathbf{H}^{\mathcal{S}\mathcal{C}_0}$:

$$\begin{bmatrix} p^{\mathcal{S}} \\ 1 \end{bmatrix} = \mathbf{H}^{\mathcal{S}\mathcal{C}_0} \begin{bmatrix} p^{\mathcal{C}_0} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}^{\mathcal{S}\mathcal{C}_0} & \mathbf{t}^{\mathcal{S}\mathcal{C}_0} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} p^{\mathcal{C}_0} \\ 1 \end{bmatrix} \tag{5}$$

with $\mathbf{R}^{\mathcal{S}\mathcal{C}_0}$ and $\mathbf{t}^{\mathcal{S}\mathcal{C}_0}$ being, respectively, the rotation matrix and translation vector of the frame change. This transformation is calibrated by comparing attitude measurements and camera poses from the observation of a known chessboard target. The problem to be solved is then similar to a Hand-Eye calibration [12].

2.3. Light pattern

The aim of the system is to allow 3D digitizing in the widest possible conditions. Consequently we focused on mechanisms to gain independence to environmental disturbances, lighting being the most important. This is the reason why the system is equipped with a pattern projector to texture the scene and facilitate the matching process between images. This projector

is a LED matrix coupled with a slide image. This slide is a speckle pattern printed on a glass tile.

Moreover, LED rings are added to each camera to be used as secondary light sources in low-lighted environments.

Since our system is vision-based, some conditions may still be limitating. Specular reflections will not allow good 3D reconstruction for example. The system tries to adapt its exposure parameters to avoid saturation on images but in the case of intense or large specular reflections, since the system is blinded by light, it will fail to find a good exposure time, and will fail to measure properly. This happens when one scans a specular surface with a frontal point of view : all the light from the projector is reflected towards the system. In this case, the best solution is to scan with a less frontal point of view to direct the reflection away from the cameras. Moreover, since we merge several scans acquired from multiple points of view, a hole created by a specular reflection in an acquisition is filled by the modelling from another acquisition without specular reflection.

3. Operationg conditions

In the sequel, strategies and operating modes for 3D modelling using our particular system design are described. An overview is provided and some details about the main methods are given.

3.1. Description

Our digitizing system is an incremental modelling device. The construction of the 3D model is done incrementaly using several raw data acquired sequentially.

The aim of our project was to design the system as simply as possible. The user controls the digitizing system using a trigger. A single pression on it switches on the device. Prior to doing anything, when starting, the system launches a sequence for calibration of exposure time and light intensity parameters to fit the environment. This step lasts a couple of seconds.

The system is then operational and runs in a preview mode. Three operating modes are available. Figure 2 shows how actions on the trigger drive transitions between each operating mode.

[Figure 2 about here.]

The starting state is when the digitizing system is off. As we stated above, a short pression on the trigger powers the device on to enter a preview state. In this operating mode, raw data are acquired, processed and displayed, and then erased. Pressing the trigger for a longer time allows the actual modelling, that is to say raw data are still acquired and processed but also stored and results of the processing are merged to the results of previous acquisitions.

The methods used for incremental modelling were chosen in order to facilitate the overall process. A common difficulty in modelling systems is, for instance, the acquisition of the support side of an object. The object is placed on a table and the support side is not visible unless the object is moved. Most of existing systems require to digitize this side separately and merge it manually during post-processing. The pipeline we propose (Figure 3) is designed to allow a fully automatic modelling during acquisition.

[Figure 3 about here.]

Our methods rely on the use of image processing algorithms and inertial sensing. The main drawback of most methods combining vision and inertial data is that all the processing is based on a fixed global frame. Moving the object is then not possible because it would mean changing the absolute reference. Our pipeline is designed to detect and correct these displacements.

For each acquisition time, the digitizing system acquires sequentially, at high rate, two pairs of images – one with the projected pattern, one with no projection – and two attitude estimations synchronised with each pair. A 3D point cloud is created, expressed in the coaxial camera frame. Incremental modelling requires to find the transformation between this camera frame and the global frame of the current model already created. Using attitude measurements and images, a best candidate for alignment process can be quickly chosen. After the alignment step, the 3D points can be merged in a voxel-map-based structure. This is used mainly for visualization, allowing fast and simple sampling of geometric information. Raw data are stored to be used in post-process refinements.

If the object has been moved or if the overlap of the current view with the current model is insufficient, no good candidate for registration is found. In this case, a second best candidate search method is used. Of course, if the overlap is too small, this method will fail and the user will be notified. If the object has moved, the inertial measurements become inconsistent with the model frame. To align the inertial frame to the new world frame, a correction transformation related to the object motion has to be estimated. The second method used is based on image indexation to retrieve a good registration candidate from an interest points base. This method is slower

than combination of vision and inertial data but is not dependent on a global reference. When a correspondant is found, the registration process gives us the actual transformation between the coaxial camera frame and the current model, allowing us to apply a correction to the attitude estimation from the inertial sensor.

When an acquisition is resumed after an interruption, the system checks if the object has moved. Matching is performed using inertial sensing and image detection. If no consistent pose can be found with this method, then the system suggests that the object has moved and the inertial measurement is not usable. The image-only method is then applied.

With such a pipeline, the acquisition process is made more user-friendly. Indeed, the modelling can be stopped and resumed easily, and the entire object can be scanned in the same automatic process by pausing the acquisition, moving the object, and resuming the operations.

3.2. 3D generation

Matching points between stereo images allows an inverse projection to retrieve 3D information from each pair of points. It is the so-called triangulation process [13]. When the matching of points or camera parameters are not perfect, optical rays linking camera centers to these points does not intersect. Optimisation or selection heuristic need to be chosen.

Considering epipolar geometry, a point in an image and each point on the corresponding epipolar line in the other image are coplanars with optical centers of the cameras. Using stereo rectification [14, 15], matching points lie on the same row in rectified image space, so the intersection between two optical rays always exists. Moreover, epipolar geometry reduces the search

space for stereo correspondent to a line of points.

Matching is then a two step process : coarse dense pairing, and refinement. First step is a global process, trying to match all pixels in an image. The search space in the other image is discretized – to 1 pixel for instance – and is consequently imprecise. This process uses the projected pattern to identify corresponding points without ambiguity.

Increasing the search space to be continuous allows the refinement of coarse pairs. Moreover, fine correlation techniques [16] optimize a transformation of the matching pattern to approximate the projective distortion.

3.3. Registration

The online registration process is divided into two main actions :

- finding the best previous view for a pairwise alignment
- estimating the transformation for the best fit of both views

These methods are only described briefly here and will be discussed more precisely in a forthcoming paper.

Finding the best candidate for a pairwise alignment in real time needs, in a way or another, a kind of mapping or classification process. It consists in finding a previously acquired view that maximizes a proximity criterion.

In our approach, we use the inertial sensing in first intent to fastly find a candidate in the rotation space. The rotation shift between the two views is then immediatly available. The translation part is found by matching interest points between images. A score is used to decide whether the candidate is good enough or not.

It should be noted that 3D data is obtained thanks to the projection of a speckle-pattern onto the scene that is moving with our digitizing system. Matching interest points between images [17] with a moving pattern projection is impossible. This is the reason why each acquisition is composed of a pair of images with projection and a pair without projection. Image points can be matched in images without illumination between two acquisition instant n and $n+1$, allowing to compute pose transformation. It is obvious then that, due to the hand motion, the pose is not exactly the same when acquiring images with projection – from which 3D points are generated – and when acquiring images without projection – from which pose estimation can be found. This introduces an imprecision, acceptable to have a good initial approach, but not accurate enough for an acceptable alignment. A fast registration method – a few iterations of ICP [18, 5] – needs to be run to finish the process.

If the search in rotation space gives no result or a result where matching does not pass the score test, then the inertial sensing may have been biased by the motion of the object. The rotation measurement is then forgotten and the candidate is chosen by image indexation, trying to find a previous acquision whose image interest points match well with those of the current acquisition. This process is not as fast as the first one, which is acceptable for ponctual events like moving the object, but would lead to an important loss in acquisition frequency if used in first intent.

In the latter approach, the fast final registration gives an absolute rotation shift between the previous – known – acquisition and the current – biased – one. This shift is used as a corrector for next acquisitions, giving the

possibility to use the inertial method again.

3.4. Finalization

The model constructed incrementaly in real time is still quite imprecise and can be noisy. To finalize the process, this model needs to be refined and cleaned. This means that the poses and computed 3D points are reestimated together in a kind of bundle adjustment process.

The representation of the model is important regarding to the target application. Meshing is required most of the time. Higher level of representation like surfacing can be needed or other informations like surface texturing.

4. Evaluations

This section aims at showing the potentialities of the scanner within two applications, a qualitative one and a more quantitative one, comparing results to some reference scanning systems of the community.

4.1. Experimental protocol

For our evaluations we scanned several objects with three systems : our hand-held digitizing system and two commercial scanning devices widely used in the community, which will be called A and B in the sequel. The two commercial devices are fixed scanning devices using optical technology for 3D reconstruction. These systems are not clearly identified because, being system designers ourselves, we don't want to bias the purpose of this article to a commercial discussion.

In the evaluations we used the Geomagic Qualify v12 software for comparisons between scans and CAD surfaces. The system A does not provide a registration solution so we used Geomagic Studio v12 to align the scans.

Scanner A is a fixed device, using laser triangulation for 3D measurement. It is an heavy system not easy to handle. System B is also a fixed device, lighter than system A. It uses fringe projection for 3D measurement and is coupled to a photogrammetry system for localization, requiring that several targets be placed in the scene.

Both systems have switchable optics for focal length modification.

Operating conditions and specifications of both scanners A and B are not really comparable to our scanner. However they have been chosen for the announced accuracy and because they are widely used in scientific and industrial communities. Technical informations about these devices are provided in table 1.

[Table 1 about here.]

4.2. Cylindrical Gauge Block

To get a more defined idea of the relative accuracy of the systems, we started with scanning a cylindrical gauge block used for micrometer calibration. The inner diameter of the cylinder is known (70.004 mm) with an accuracy of 1 μm.

Evaluation is done using a single acquisition. For each evaluation, the 3D point cloud is registered to a theoretical surface of a 70.004 mm diameter

[1]Needs to change the focal length to cover all the distances.

[2]To the Z reference plane.

cylinder using Geomagic Qualify. Points are projected orthogonaly on the surface and the projection distance – the error – is measured. The analyses are based on the mean and standard deviation of the error distances.

Table 2 shows the results of this test.

[Table 2 about here.]

This first test was not intended to make a comparison and ranking of the scanners but to be used as a basis for the evaluation. It tends to give us an absolute reference of the 3D reconstruction process for each device. One can note that our device is in the same accuracy range than system A whereas system B provides more accurate results.

4.3. Statue

The first evaluation consists in scanning a statue shown in Figure 4. This object is composed of large smooth parts (face, support side, ...) and complex shape areas (cloak mostly).

[Figure 4 about here.]

At the beginning of the process, the statue is disposed on its support side. We began the acquisition with the face, and then moving to our left. The object has been turned around the vertical axis several times during the process to acquire the cylinder of the head. After a complete turn, we scanned the top of the head. Then the object has been moved to be placed face up. By scanning from the bottom of the face we acquired the support base, which has been completed by placing the object face down and doing the same from the back of the head to the support side.

The result of the scan is shown in Figure 5.

[Figure 5 about here.]

The model is noisy but already recognizable. It still needs to be finalized. We refined the model, cleaned outliers and then meshed. The final result is shown in Figure 6. The texture provided by the images has also been applied to the final model.

[Figure 6 about here.]

The operation of our device allows us an easy handling which is of great interest in this type of object. The statue has been scanned using the commercial scanning devices but handling fixed systems adds difficulties when scanning slighltly hidden details like those on the cloak. Figure 7 illustrates a detail of the cloak which suffers information loss in a scan from one of the commercial devices.

[Figure 7 about here.]

4.4. Mechanical test piece

For a more quantitative analysis, we used a stamped sheet metal part. To avoid specular effect, the object was mattified.

The object is scanned only in its upper face. The sheet being thin, the interest of scanning the opposite face and the sides is questionable considering the resolution and the accuracy of most scanners.

[Figure 8 about here.]

4.4.1. Comparison of the three scans with the CAD model

In our test benchmark, we first compared our scan and the scans provided by systems A and B with the CAD model. Due to mechanical efforts, particularly on release of blank holders, the object has been deformed and it does not correspond anymore to its theoretical CAD model. We registered our test scans to the CAD reference along a same interest zone, favouring the central stamping, supposed to be less deformed. Results are shown in Figure 9.

[Figure 9 about here.]

The metric used is based on a direct orthogonal distance from scanned points to the reference surface. Figure 10 provides the error distributions for each scan.

[Figure 10 about here.]

This test allows us to check if there is a bias between several measurements of the same object. Comparing the results, our first observation is the similarity of the error maps. With this test, we can see also that the comparison method is stable. Similar data leads to similar registration and projection, considering we used a common alignment reference. This allows us to use this comparison method for the next tests.

4.4.2. Comparison with scan B

The next evaluation focuses on the impact of our finalization step on accuracy. The CAD surface being a bad ground truth we decided to use scan

B provided by the more accurate device – according to the manufacturers' specification and the cylinder test – as a reference.

[Figure 11 about here.]

Figure 11 compares the 3D shape measured with our digitizing system before the post-processing step and after this finalization. Point clouds are projected onto the reference – from system B – model.

In the raw cloud, a lot of noise appears. This noise is an oscillating phenomenon from side to side of the mean surface. Large deviation is observable at the extremities of the model. A boundary effect appears, leading to largely biased measurements in these areas.

After the finalization step, noise has been largely reduced, at the point that it is not visible with our colour span. Observing the histogram, we can see that deviation caused by the fast registration method used has been largely reduced by using more accurate registration. The model is more largely comparable to scan B. One can note that boudaries effect, if reduced, still appears.

Summary of errors in both steps is given in Table 3.

[Table 3 about here.]

5. Conclusion

In this article we presented our new hand-held 3D scanning device based on vision technologies. We focused the design of this scanner on an easy-to-use scanner. No equipment or markers need to be put in the scene and, being hand-held, the scanner allows a more dexterous manipulation. The operation

has been made simple, real-time and mostly automatic. A finalization step allows the creation of a more accurate and usable model.

We evaluated the performances of our scanner on several test objects to cover more widely the 3D digitizing applications. We based our evaluations on results obtained by scanning the same objects with well known and widely used commercial devices. Our tests tend to prove that the proposed system matches the requirements of common applications in qualitative terms, accuracy and usability but with an eased operation allowing to fasten the overall modelling process.

Our future works will focus on improving the texturing operation and qualifying more precisely the accuracy and performances of our system. We will also focus our work on trying to reduce noise in the digitizing process. Another improvement can be made by considering ROI-based reconstruction [19]. With such an approach, reconstruction can be processed faster and with reduced noise level due to filtering ill-observed zones outside the central object in images.

Acknowledgments

We thank the IRIT (Institut de Recherche en Informatique de Toulouse) laboratory from Toulouse – France, and the ENIT LGP (Ecole Nationale d'Ingénieurs de Tarbes – Laboratoire Génie de Production) laboratory from Tarbes – France, for the help they provided by scanning our test objects with their commercial scanners (scanners A and B in this paper).

References

[1] Pan Q. , Reitmayr G. , Drummond T.W. . Interactive model reconstruction with user guidance. In: Mixed and Augmented Reality, IEEE / ACM International Symposium on. 2009, p. 209–10.

[2] Curless B. . From range scans to 3D models. SIGGRAPH Computer Graphics 2000;33(4):38–41.

[3] Matabosch C. , Fofi D. , Salvi J. , Batlle E. . Registration of surfaces minimizing error propagation for a one-shot multi-slit hand-held scanner. Pattern recognition 2008;41(6):2055–67.

[4] Strobl E. , Mair E. , Bodenmüller T. , Kielhöfer S. , Sepp W. , Suppa M. , et al. The self-referenced DLR 3D-Modeler. In: International Conference on Intelligent Robots and Systems. St. Louis, MO, USA; 2009, p. 21–8.

[5] Besl P.J. , McKay N.D. . A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 1992;14(2):239–56.

[6] Rusinkiewicz S. , Levoy M. . Efficient variants of the ICP algorithm. In: Third International Conference on 3D Digital Imaging and Modeling (3DIM). 2001,.

[7] Sandhu R. , Dambreville S. , Tannenbaum A. . Particle filtering for registration of 2D and 3D point sets with stochastic dynamics. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2008,.

[8] Li H. , Hartley R. . The 3D-3D registration problem revisited. In: IEEE International Conference on Computer Vision. 2007, p. 1–8.

[9] Zhang Z. . A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 2000;22:1330–4.

[10] Marins J.L. , Yun X. , Bachmann E.R. , McGhee R.B. , Zyda M.J. . An Extended Kalman Filter for Quaternion-Based Orientation Estimation Using MARG Sensors. In: International Conference on Intelligent Robots and Systems. 2001,.

[11] Lefferts E.J. , Markley F.L. , Shuster M.D. . Kalman filtering for spacecraft attitude estimation. Journal of Guidance, Control and Dynamics 1982;5(5):417–29.

[12] Park F.C. , Martin B.J. . Robot sensor calibration: solving AX=XB on the euclidean group. IEEE Transactions on Robotics and Automation 1994;10(5):717–21.

[13] Hartley R.I. , Sturm P. . Triangulation. Computer Vision and Image Understanding 1997;68(2):146–57.

[14] Fusiello A. , Trucco E. , Verri A. . A compact algorithm for rectification of stereo pairs. Machine Vision Application 2000;12(1):16–22.

[15] Bugarin F. , Henrion D. , Sentenac T. , Lasserre J.B. , Orteu J.J. . Optimisation globale polynomiale appliquée à la rectification projective d'images non calibrées (in french). In: Conférence en Reconnaissance des Formes et Intelligence Artificielle. 2010,.

[16] Garcia D. . Mesure de formes et de champs de déplacements tridimensionnels par stéréocorrélation d'images (in french). Ph.D. thesis; Institut National Polytechnique de Toulouse; 2001.

[17] Bay H. , Tuytelaars T. , Van Gool L. . Surf: Speeded up robust features. In: European Conference on Computer Vision. 2006, p. 404–17.

[18] Chen Y. , Medioni G. . Object modelling by registration of multiple range images. Image and Vision Computing 1992;10(3):145–55.

[19] Li W. , Schütze R. , Böhler M. , Boochs F. , Marzani F.S. , Voisin Y. . Preprocessing of region of interest localization based on local surface curvature analysis for three-dimensional reconstruction with multiresolution. Optical Engineering 2009;48(6).

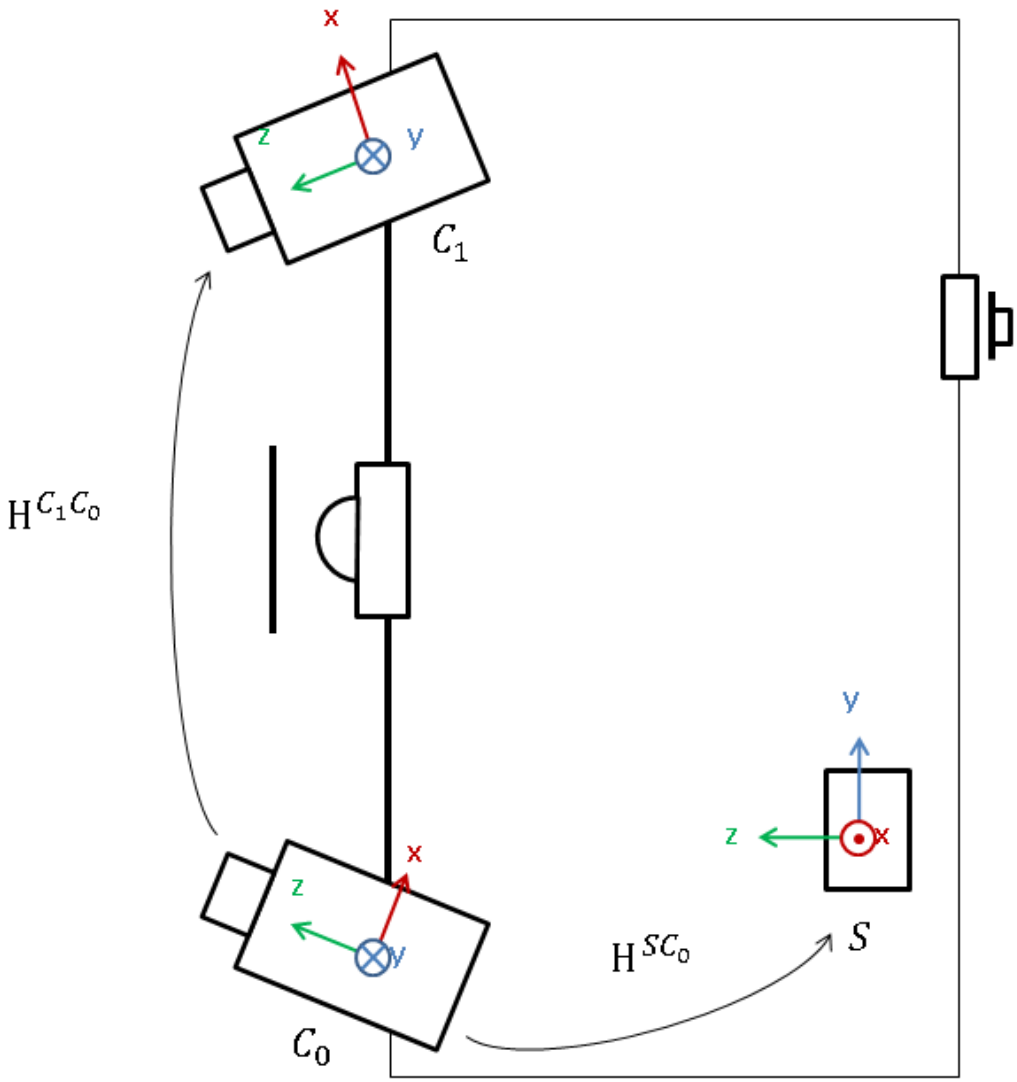


Figure 1: Design of the system with sensor frames and transformations.

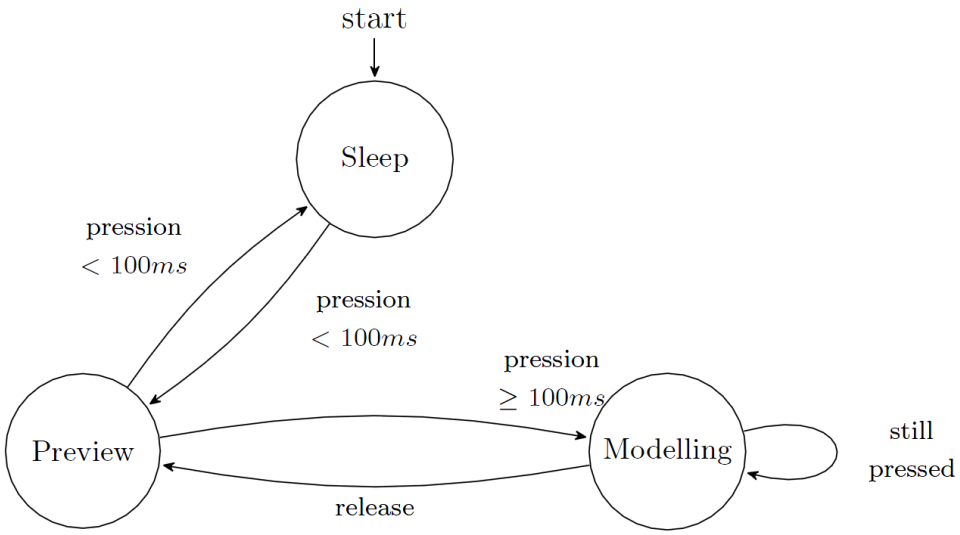


Figure 2: State diagram of trigger control

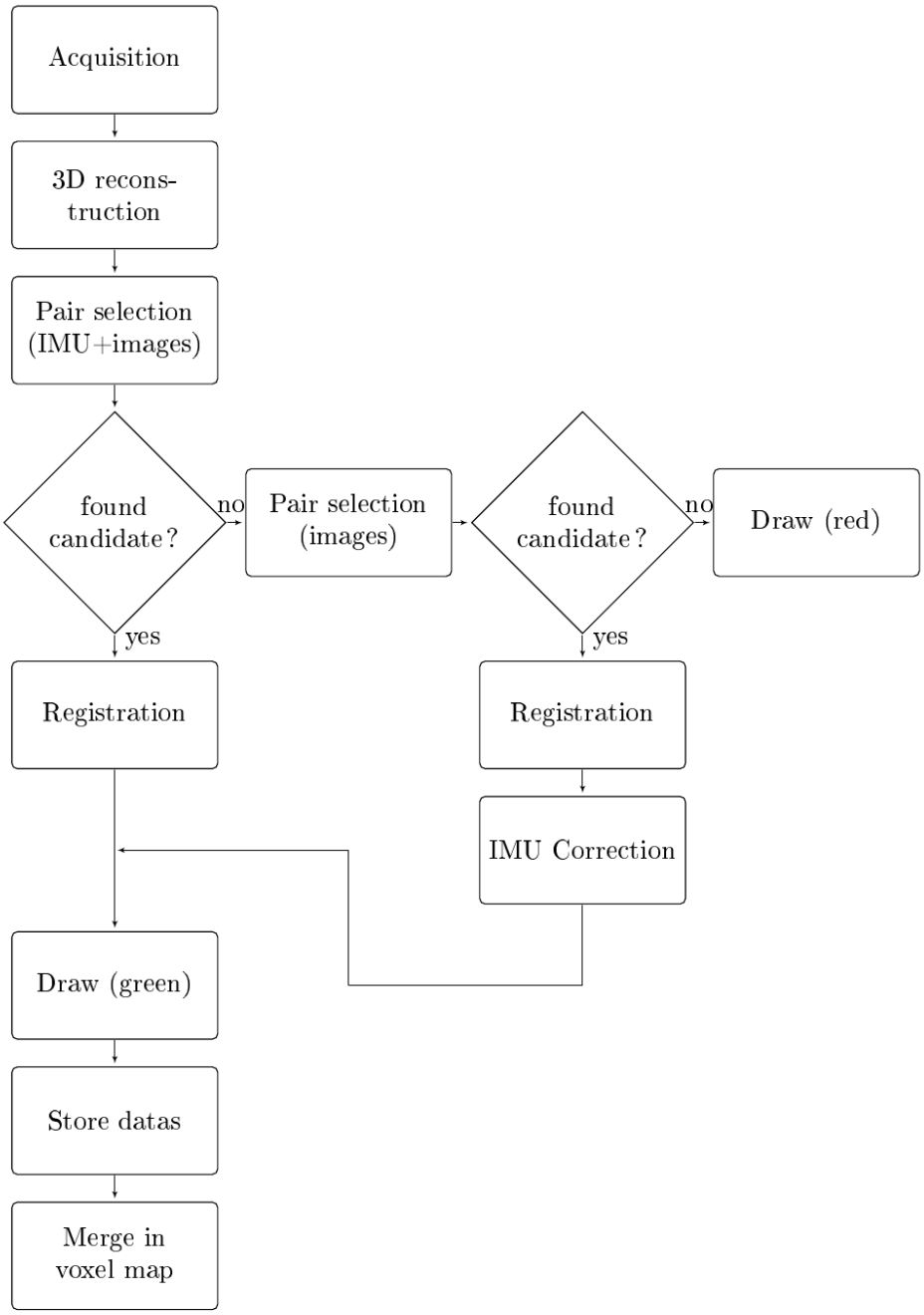


Figure 3: Operation diagram for single acquisition processing.

Figure 4: Test statue.

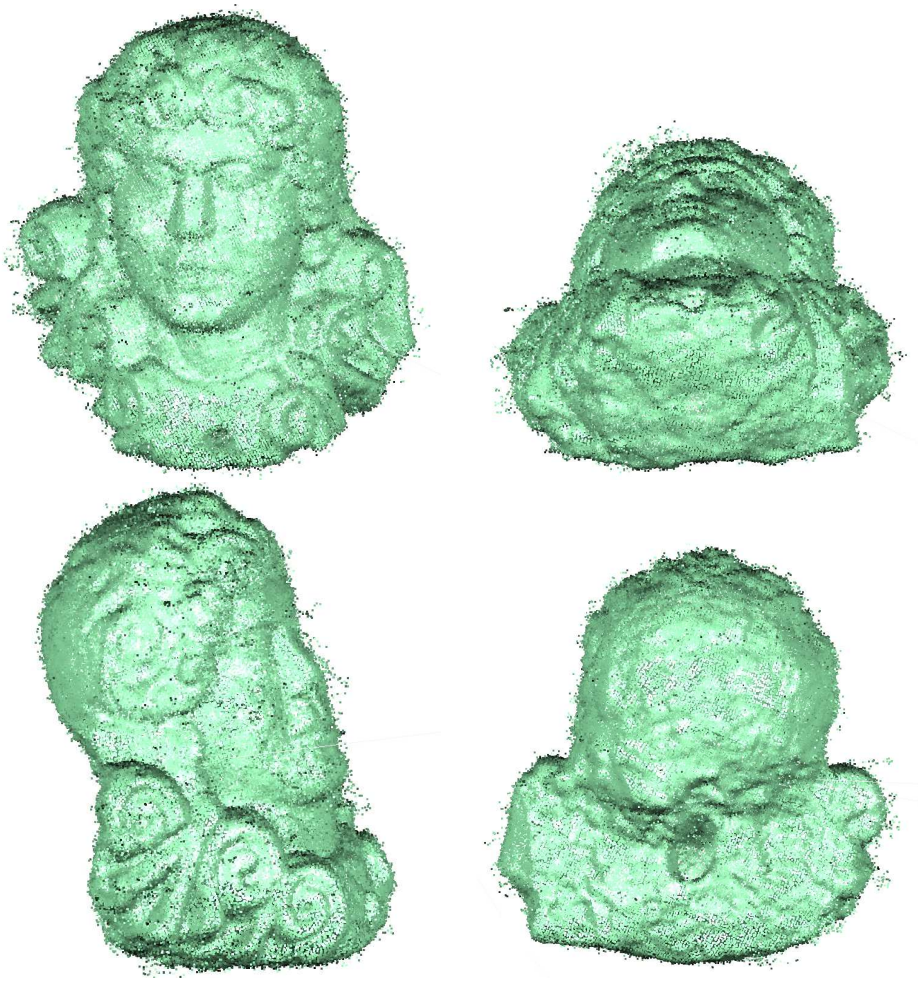
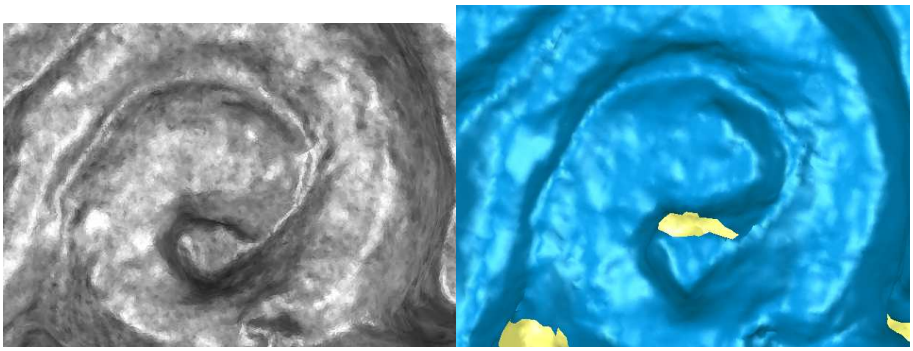


Figure 5: Result of the scan without finalization step.

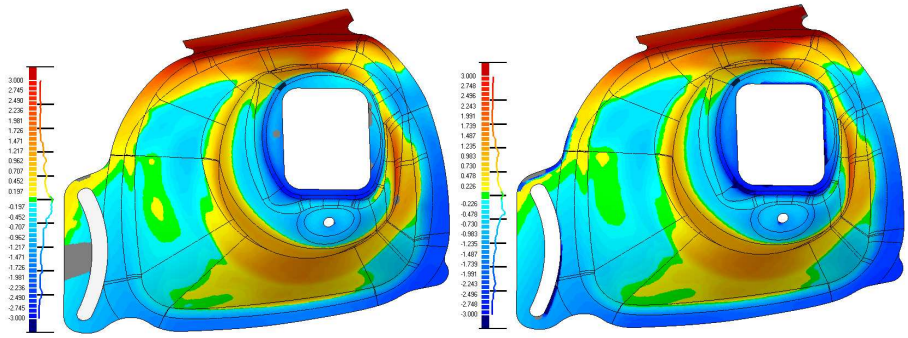
Figure 6: Finalized and textured scan of the statue using our system.



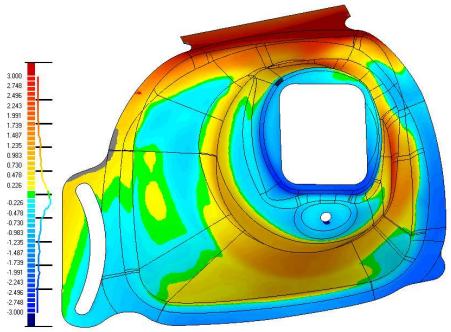
(a) (b)

Figure 7: Detail of the object. In complex areas of the object, some faces can be hard to observe due to occlusions. (a) Observing complex surfaces with our system is eased by its hand-held operation, allowing a more dexterous manipulation. (b) Scan from system A, with a less dexterous operation information, losses can occur.

Figure 8: Stamped sheet metal part used for our tests and resulting point cloud provided by our system.



(a) (b)



(c)

Figure 9: Comparison of scans with theoretical surface : (a) Scan using system A ; (b) Scan using system B ; (c) Scan using our hand-held scanner.

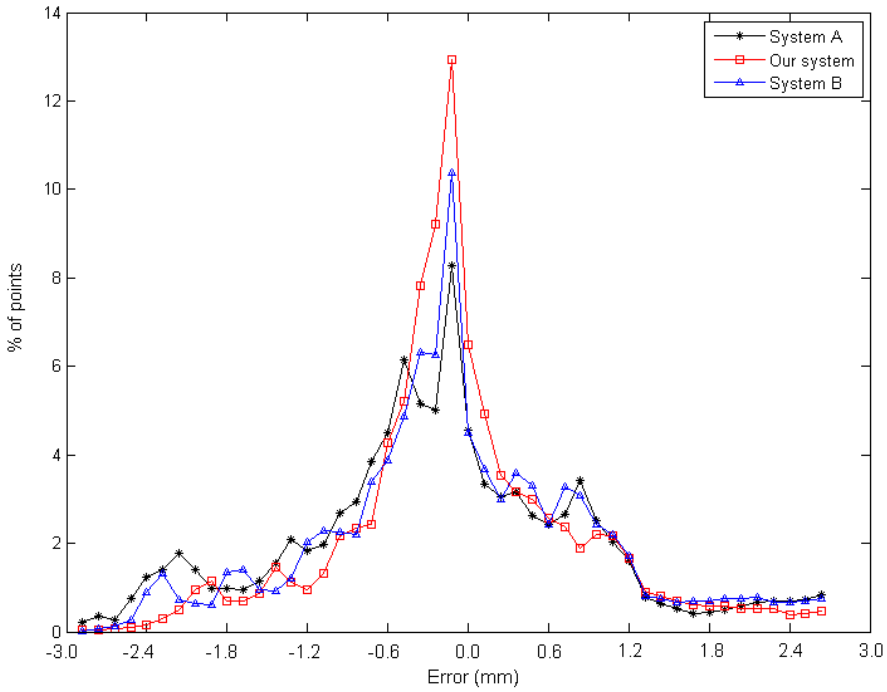
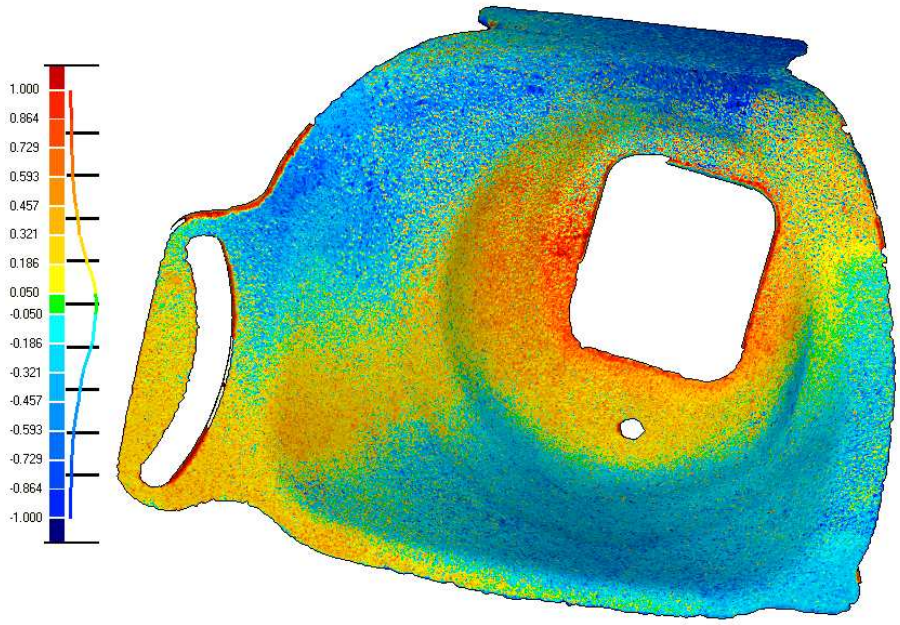
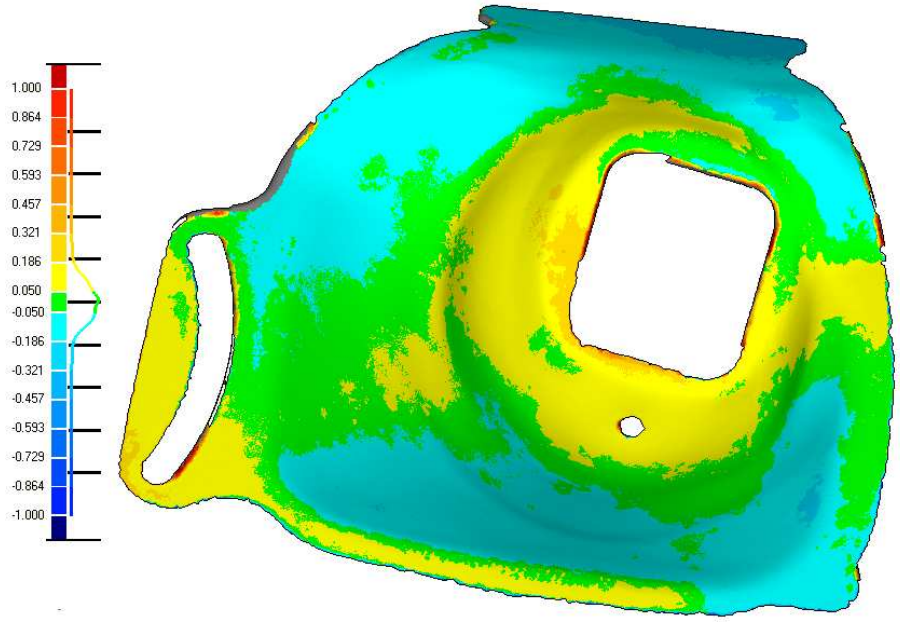


Figure 10: Distribution of error between 3D scans and the CAD surface.



(a)



(b)

Figure 11: Comparison between our scan and scan B : (a) before finalization ; (b) after finalization.

	A	**B**	**Our system**
Motion	Fixed	Fixed	Hand-held
Technique	Laser triangulation	Structured light	Structured light
Registration	Manual	Photogrammetry	Automatic
Weight (kg)	15	7.4	1.8
Dimensions (mm)	$221 \times 412 \times 282$	$490 \times 300 \times 170$	$220 \times 240 \times 90$
Focal length (mm)	14	16	8
Operating distance[1] (m)	0.6 – 1.2	0.7 – 2	0.35 – 0.5
Measurement volume (mm)	from $111 \times 83 \times 40$	from $135 \times 108 \times 108$	from $10cm^3$ to $1m^3$
	to $1196 \times 897 \times 800$	to $1700 \times 1360 \times 1360$	NC
Accuracy (X,Y,Z)[2] (mm)	$\pm\,(0.22, 0.16, 0.10)$	NC	NC
Point spacing (mm)	NC	0.08 – 1.0	0.3
Measuring noise (mm)	NC	0.002 – 0.02	0.1

Table 1: Systems specifications (Commercial systems specifications are provided according to manufacturers' data sheets).

	A	**B**	**Our system**
Mean (mm)	−0.00115	0.00003	−0.00115
Std. dev. (mm)	0.02764	0.00508	0.02480
Max. error (mm)	0.10238	0.02162	0.08649
Min. error(mm)	−0.09151	−0.02569	−0.09392

Table 2: Evaluation results scanning a known gauge block with the three systems.

	Before finalization	**After finalization**
Mean error (mm)	−0.004	−0.001
Std. dev. (mm)	0.373	0.125

Table 3: Errors between our scan and scan B