Towards a Versatile Handheld 3D Laser Scanner

F. Arnaoutoglou[1], A. Koutsoudis[1], G. Pavlidis[1], V. Tsioukas[2] and C. Chamzas[3]

[1] Cultural and Educational Technology Institute, Research Centre Athena, Xanthi, Greece
[2] Dept. of Architectural Engineering, Democritus University of Thrace, Xanthi, Greece
[3] Dept. of Electrical and Computer Engineering, Democritus University of Thrace, Xanthi, Greece

Abstract

Over the past years, the significance of digital recording of cultural heritage has been realized as a major factor for preservation and dissemination. Due to a continuous growing interest, a numerous of digital recording techniques have been devised to meet the requirements of the cultural heritage sector. 3D digitization is one of the most important aspects in digital recording. Various techniques have been proposed for 3D digitization but still there is not an all-in-one solution due to limitation in technology. In this paper, we are discussing versatility as a key factor of a successful handheld 3D laser scanning system that is applicable to the recording of cultural heritage. The proposed system is based on laser triangulation and 3D camera tracking from a sequence of images. We are considering the case where only minimal information is available to the system prior to its usage.

I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture

1. Introduction

The significance of digital recording of cultural heritage has been realized, over the past years, as a major factor for the preservation and dissemination of culture. Nowadays, 3D digitization has already established its foundations on archiving cultural heritage as great advances in 3D technologies offer new opportunities to record every detail of cultural heritage in high precision, and to present it in a more attractive ways [TAK*03].

A numerous of digital recording techniques and methodologies have been devised and proposed in order to meet the requirements of the cultural sector. In fact, three-dimensional (3D) digitization is considered as one of the most important aspects in digital recording. It is the first stage in every digital recording project and is the process that produces the first version of the digital content. One of the biggest recognized advantages of digitally archiving artefacts is the production of un-durable works of art. Although various techniques have been proposed for 3D digitization, there isn't an all-in-one solution – that is, a solution that meets the requirements for every digitization project – due to limitation in technology. Laser scanning is one of the most successful techniques that have been developed during the past years in order to tackle with the problems of reverse engineering in industry and accurate 3D digitization for every possible application, within its technological limitations.

Present 3D acquisition systems are usually pushed to their limits when used for the recording of cultural heritage, as challenges uprise due to the physical characteristics of the artefacts. The raw materials that have been used to construct them, in combination with the morphologic complexity contribute on producing shadowy texture areas, subsurface scattering of laser light and major occlusion problems. On the other hand, the delicate and fragile nature of such treasures, prohibit their physical contact and their moving. An immovable artefact is hard to be fully digitized. Thus, we consider versatility of a non-contact scanning system as a key factor for a successful archiving project. Nonetheless, it should be able to derive accurate geometry and texture while being sensitive to the artefact's surface. Laser scanning devices have proved to be applicable in such cases as they have the ability to maintain a narrow beam over long distances [LPC*00]. An accidental collision between the scanning device and the artefact is a non-accepted situation. In fact, there is no silver bullet for safety issues and on this account laser scanning allows at least an acceptable standoff from the artefact to avoid undesirable situations.

Summarizing, the digitization of cultural heritage objects, of relatively small size, is a problem that can be successfully tackled with laser scanning techniques. As these techniques evolve and are being applied and used in real life cases, new requirements become even more evident:

easiness of usage and portability of the scanning system. A handheld 3D scanning device introduces unique flexibility at high accuracy levels and it can, thus, be considered as a highly applicable device on the delicate area of heritage archiving. Working in this light, we propose a technological framework towards a versatile handheld 3D laser scanning system that can be efficiently be used for the digitization of cultural heritage artefacts.

2. Handheld 3D Scanning Systems

Many commercial and experimental handheld systems have already been proposed. Most of them share the idea of manually sweeping the laser beam over the scene or the object. This is a great advantage as it allows the complete scanning of complex geometry from different views without constrains on motion imposed by a mechanical translation or rotation system. Handheld systems can overcome the size range limitation of static systems while keeping the cost in low levels as no mechanical structures are required. A portable handheld scanner can reduce data collection and modelling time while providing flexibility, which is a necessity. Nevertheless, a handheld scanner is not a panacea [Heb01]. In some cases, depending on the size of the object, a handheld sensor can be used as a complementary device for all those places which is hard to be reached by other static systems [LPC*00]. Building a handheld scanning device presupposes that the laser light integration time should be short enough in respect to the displacement of the sensor. It is only then possible to avoid motion blur within a single image frame [Heb01]. Laser line scanning systems are intrinsically faster but finding the correspondence of the points on the line does pose some problems [BFB*98].

For instance, “Autoscan” [BFB*98] is a portable 3D scanner that consists of a laser pointer, a pair of video cameras and a real-time processor that detects the circular spots of the laser in the scene. Its overall weight is 15 kg and the video cameras angle is at least 60 degrees to guarantee high accuracy (0.1mm at a standoff distance of 1.5 m and a baseline distance of 1m). The scanning time is a drawback of the system as it uses one laser pointer that corresponds to approximately 200 triangles per second.

A variant of “Autoscan”, the “ModelCamera”, proposed in [PSB03], involves the usage of sixteen laser pointers fixed with respect to an ordinary video camera. While the user scans the scene, the laser beams produce blobs in the video frame where they hit the objects’ surfaces. Their actual positions in the 3D space are being derived by triangulation in every frame. The registration of the frames results in an evolving model. The user can vary the sampling rate by zooming in and out. This system requires an improved blob detection algorithm when complex surface properties like colour, texture and specularity are introduced [PSB03].

Takatsuka et al. [TWV*99] used a fixed calibrated camera in combination with a handheld laser pointer on which three green light emitting diodes are always locatable along the optical axis of the laser. The positions of those LEDs in space are computed from their projections on the image plane and then they are being used to determine the optical axis of the laser. The 3D point is derived as the intersection of the viewing direction of the camera and laser axis.

Hebert [Heb01] presented a handheld system based on structured light projection that integrates both shape measurements and self-referencing. The configuration proposed consists of two synchronized cameras and a laser diode projecting two perpendicular light planes. “HandyScan 3D” [Han06] is a commercial product which is based on similar principles. Its weight is almost a 1 kg and its accuracy is 0.25 mm on a distance up to 500 mm. Another good example is the “FastSCAN” [Pol06] series by Polhemus. The “FastSCAN” series is designed to scan non-metallic, opaque objects using 1mw lasers and either a single or double camera configuration.

Bahmutov et al. [BPM06] describe an efficient and interactive system for modelling large scale building interiors. The system is based on the structured light technique following a custom approach of projecting a matrix of 11 x 11 laser spots in the field of view of a digital camera. The depth is calculated using multiple dense colour and sparse depth frames which share the same centre of projection. As a result the resolution of the obtained geometry is not enough for the description of objects with high complexity.

Marc Pollefeys et al. proposed in [PVV*03] a handheld 3D model acquisition system that at its first step of operation is quite similar to the system proposed in this paper. The system initially estimates the motion of the camera and sparsely approximates the 3D scene. These data are used to produce a dense estimation of the reconstructed geometry using a flexible multi-view stereo matching scheme. The similarity between Pollefeys approach and the one proposed in this paper lies in the camera motion estimation part. However, Pollefeys’ approach for dense mesh generation using stereo matching appears to be computational expensive and inadequate for ill conditioned image sequences, like sequences where not enough significant points to match are available, or sequences where the transitional motion of the camera does not provide the appropriate pixel disparity between a stereo pair.

Rusinkiewicz et al. proposed in [RHL02] a real-time handheld 3D model acquisition system that permits the user to rotate an object by hand and see a continuously-updated model as the object is scanned. The advantage of this system is that the user can find and fill holes in the model in real time and determine when the object is completely covered. The disadvantage of this system (as presented) is that it requires physical contact with the subject and specific and synchronized hardware.

2.1. The handheld laser scanning system (or Versatility is the key)

The proposed system is based on simple and well established notions in order to deduce the geometry of an object using a sequence of images taken from a video camera (or a photographic camera). Since we are considering the uncalibrated case (only minimal knowledge for the camera is available and a calibration process is not applied), these notions involve camera tracking techniques in order to acquire knowledge of the position and orientation of a camera in a 3D space. The principle of triangulation is employed in order to resolve relative positions of points of the scanned object in the 3D space. In such a system, the coordinates of an imaged point of the object at a given time can be computed by typical matrix multiplications that reflect both the camera model and the camera position and orientation:

$$\overline{C} = I \cdot E \cdot \overline{c} \tag{1}$$

where $\overline{C}$ is the vector of coordinates to be computed, $\overline{c}$ is the vector of the point coordinates in the image plain and I and E are the intrinsic and extrinsic camera parameters in the form of matrices. The matrix of the intrinsic parameters (in an augmented form) is usually defined as:

$$I = \begin{bmatrix} fc_x & \varphi \times fc_x & cc_x & 0 \\ 0 & fc_y & cc_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{2}$$

where fc_x, fx_y represent the focal length in units of horizontal and vertical pixels, φ is the angle between x and y sensor axes (typically $\varphi \times fc_x = 0$) and cc_x, cc_y are the coordinates of the principal point (ideally the centre of the image sensor).

These parameters are called intrinsic because they are specific to the type of camera used and are constant for a given camera. Matrix I is estimated once for every digitization project. On the other hand, the extrinsic parameters refer to the orientation and position of the camera relative to a reference world coordinate system (the coordinate system of the scanned object). These parameters' values can vary significantly throughout the process of digitization. The extrinsic parameters matrix is defined (in an augmented form) as:

$$E = \begin{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} & \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \\ 0 \quad 0 \quad 0 & 1 \end{bmatrix} \tag{3}$$

where the 3x3 upper-left matrix with the r_{ij} elements is the rotation matrix and the 1x3 upper-right vector with the t_i elements is the translation matrix.

Figure 1 depicts a possible configuration of the proposed system, as well as a graphical representation of the process of triangulation for a point in space.

In order to keep the cost of the proposed system low, a common arrangement of both the camera and the laser was followed. They are positioned in such way that they form an imaginary triangle with the target point. The baseline distance between the camera and laser diode is denoted by d, while φ is the angle of the camera and θ the angle of the laser both with the axis vertical to the line that connects them. In total, the variables that are the known parameters of this arrangement are:

- The camera field of view (angle, *FOV*)
- The camera frame resolution, i.e. the frame width w and the frame height h
- The relative topology of the arrangement, i.e. the camera and laser angles φ and θ and the distance between them (d)
- The key value of the camera focal length can be deduced from the known parameters:

$$\left.\begin{aligned} \tan\left(\frac{FOV}{2}\right) &= \frac{r}{f_c} \\ (2r)^2 &= w^2 + h^2 \end{aligned}\right\} \Rightarrow fc = \frac{\frac{\sqrt{w^2+h^2}}{2}}{\tan\left(\frac{FOV}{2}\right)} \tag{4}$$

It should be noted here that all these parameters and the deduced geometries are based on the pinhole camera model.

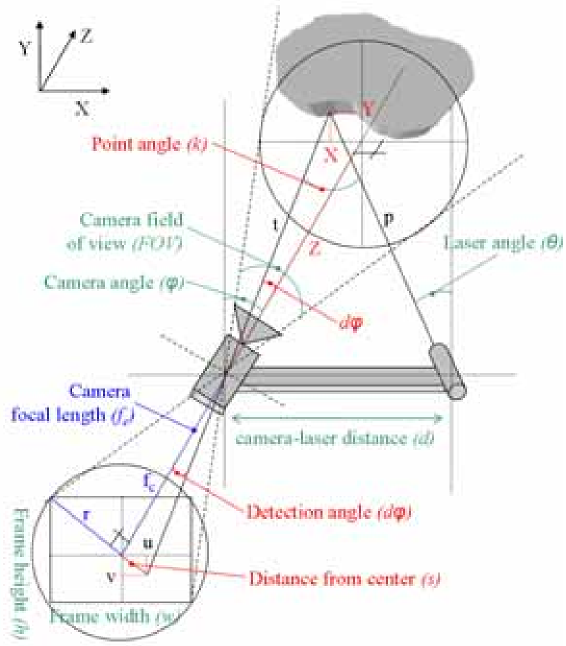

Figure 1: *The proposed system and the process of triangulation for a point in space*

As stated, one of the two stages of the proposed system is the triangulation for the estimation of the position of an unknown point of the scanned object in space. Triangulation is based on the law of sines, which states that if the sides of an arbitrary triangle are a, b and c and the angles opposite those sides are A, B and C:

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} = 2R \tag{5}$$

where R is the radius of the triangle's circumcircle.

In our case, (5) becomes:

$$\frac{d}{\sin k} = \frac{p}{\sin\left(\frac{\pi}{2} - \varphi + d\varphi\right)} = \frac{t}{\sin\left(\frac{\pi}{2} - \vartheta\right)} \tag{6}$$

In this topology the problem is a typical geometric problem that can be easily solved, as:

$$k + \left(\frac{\pi}{2} - \varphi + d\varphi\right) + \left(\frac{\pi}{2} - \vartheta\right) = 2\pi \Rightarrow \tag{7}$$

$$k = 2\pi - \left(\frac{\pi}{2} - \varphi + d\varphi\right) - \left(\frac{\pi}{2} - \vartheta\right) = \pi + \varphi + \vartheta - d\varphi$$

and, as of this, the unknown distance from the camera is:

$$\frac{d}{\sin k} = \frac{t}{\sin\left(\frac{\pi}{2} - \vartheta\right)} \Rightarrow t = d\,\frac{\sin\left(\frac{\pi}{2} - \vartheta\right)}{\sin(\pi + \varphi + \vartheta - d\varphi)} \tag{8}$$

The only unknown variable here is the angle $d\varphi$, which can easily be estimated trigonometrically:

$$\tan(d\varphi) = \frac{s}{f_c} \Rightarrow d\varphi = \arctan\left(\frac{s}{f_c}\right) \Rightarrow \tag{9}$$

$$d\varphi = \arctan\left(2\frac{\sqrt{u^2+v^2}}{\sqrt{w^2+h^2}}\tan\left(\frac{FOV}{2}\right)\right)$$

In practice, it is more convenient to estimate the *X, Y, Z* coordinates of the detected point instead of its distance from the camera *t*. This can be achieved if instead of working with the distance s we estimate angles in *X* and *Y* axis separately. The notion is depicted graphically in Figure 2, where the detection angle $d\varphi$ is represented by two angles that are relative to one axis, i.e. $d\varphi_X$ for *X* and $d\varphi_Y$ for *Y*.

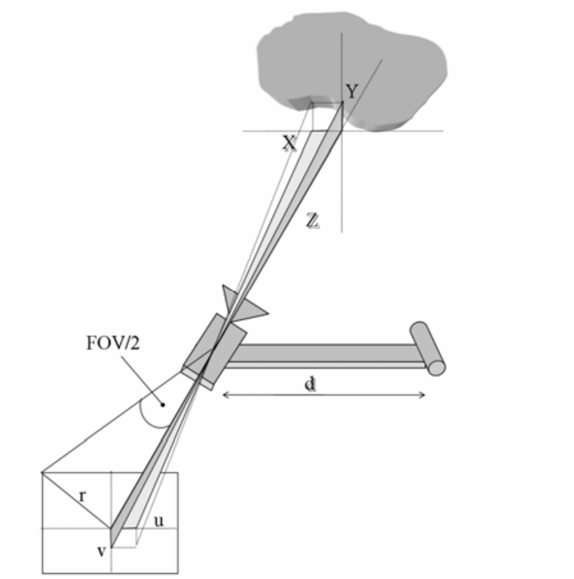

Figure 2: *A more practical approach to the estimation of the unknown coordinates of a detected point*

The estimation of these angles is equivalent to the estimation of $d\varphi$ in (6). The final equations are:

$$d\varphi_X = \arctan\left(\frac{u}{r}\tan\left(\frac{FOV}{2}\right)\right) \tag{10}$$

$$d\varphi_Y = \arctan\left(\frac{v}{r}\tan\left(\frac{FOV}{2}\right)\right)$$

$$L = 2\tan\left(\frac{FOV}{2}\right)\Big/\sqrt{w^2+h^2}$$

$$Z = d\,\frac{\cos(\arctan(uL))\sin(\pi - \vartheta)}{\sin((\pi - \vartheta) + \varphi + \arctan(uL))}$$

$$Y = ZvL$$

$$X = ZuL$$

Another basic stage of the proposed system is the estimation of the camera position and orientation in every frame of the sequence. This is usually referenced as the *camera tracking* problem and is a typical photogrammetric procedure. Consecutive frames, coming out of a single camera moving around a 3D object (often referenced as rigid motion), are processed in a way that emulates the stereoscopic vision of humans. Further processing using photogrammetric algorithms and the colinearity and coplanarity equations may lead to the creation of the 3D shape (but not the exact size) of the object space. If additional information of the scale of the 3D objects is also provided, the exact size of them could be acquired. The main positive consequences from the determination of the relative orientation of two camera frames are:

- the stereoscopic viewing ability (produced through the epipolar geometry and usage of special stereo viewing hardware configuration [Pom99])
- the restriction of the matching algorithms (from 2D to 1D) for the determination of conjugate points [TSP00] and the further processing using space intersection algorithms for the determination of the imaged points' 3D coordinates [Gru01]

Several algorithms have been proposed for the determination of image points' conjugates of two consecutive camera frames. We have used and tested two algorithms for the extraction of points of interest and their matching in two consecutive camera frames:

- Kanade-Lucas-Tomasi (KLT) tracking [ST94].
- Scale Invariant Feature Transform (SIFT) [Low99].

KLT is based on the selection of regions of interest and their tracking in a sequence of images according to a dissimilarity metric that is used to quantify the change of appearance of a feature between frames. This dissimilarity metric is defined as:

$$\varepsilon = \iint_W [J(Ax+d) - I(x)]^2 w(x)dx \tag{11}$$

where *J* and *I* are two consecutive images, *A* is a deformation matrix, *d* is the translation of the feature window's centre, *W* is a given feature window and *w(x)* a weighting function (taken either as 1 or as a gaussian function). Thus the problem of determining the motion parameters is that of finding the A and d that minimize the dissimilarity in (11).

SIFT is similar in notion to KLT. Its goal is to select scale-invariant features by employing a staged filtering approach that results in multiple SIFT keys. These keys are used to identify candidate object models. The main advantage of this method is the improvement expected by using SIFT features that are largely invariant to changes in scale, illumination and local affine distortions. The SIFT detector appeared to be the most effective algorithm in this approach. The implementation of the SIFT detector we used has been created by Alexandre Jenny [Jen04] and has been embedded in our implementation in order to extract conjugate points between two consecutive camera frames. The relative orientation algorithm accepts as input the conjugate image points' coordinates in two camera frames and produces the 5 relative orientation parameters:

- βy, βz translation parameters along the Y and Z axes relative to the translation along X axis and
- $\delta\omega$, $\delta\varphi$, $\delta\kappa$ rotation of the camera axes of the second image relative to the first.

These estimated parameters are used to determine the camera position and orientation in order to be used after the triangulation process so that the relative estimated coordinates can be transformed to the world coordinate system.

In order to achieve high accuracy on the determination of the rotational parameters of the camera a large number of points should be identified between consecutive image frames. In order to extract a large number of interest points from one image to be matched in the next one, the texture of the images should be high enough while their relative rotation could either be low or high. The SIFT detector has the ability to match two camera frames no matter how great their relative rotation is. In our case the simulated images have been enriched in texture and the rotation between camera frames is relatively low. This is why, in most cases, the algorithm succeeded to provide the correct relative estimates of the camera frames.

2.2. Experimental results

The proposed system has been tested using synthetic data, i.e. image sequences of primitive 3D objects exported by 3dStudioMax. The sequences were produced using a simulated 28mm video camera with an active sensor frame size that corresponds to 640x480 pixels resolution. The produced sequences correspond to videos of a 25 frames per second. Thus, the test data correspond to data equivalent to the PAL system. Several sequences have been produced, with the camera forced to perform multiple translations and rotations simultaneously and individually. The laser that has been simulated was a monochromatic green line projected on the surface of the scene objects. Figure 3 depicts a sample sequence of the test data set in gray-scale format.

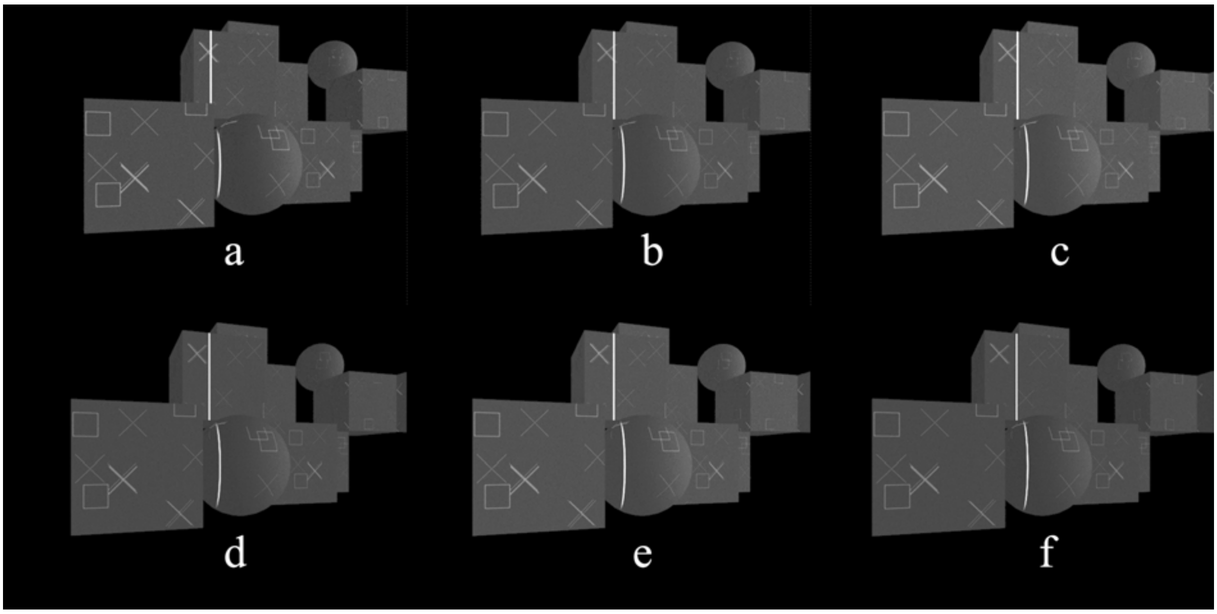

Figure 3: *Six (a to f) consecutive images (with a step of two images) from the test data set. The bright vertical line is the laser projected on the object.*

These preliminary experimental results verified that such a system is actually feasible. The accuracy as well as the resolution and productivity (i.e. time per scanning) of the system are a subject of our further work on this system. Extensive experimentations are also being planned in order to investigate possible system restrictions at extreme cases. Figure 4 depicts the results of the digitization process, at two different time instances as the triangulation algorithm operates on the data set.

3. Conclusions

In this work we attempt to combine the idea of single camera laser triangulation with the idea of 3D camera tracking in order to produce an operational friendly and safe 3D digitization device for both the user and the scanning subject. The proposed system is applicable to the digitization of cultural heritage artefacts and is aimed to be extremely simple and of low cost as well as able to support freeform handheld 3D scanning with no mechanical constraints while following a smooth video shooting procedure. The main advantage of the system is its simplicity and easiness of usage. Extensive experimentations are being planned in order to investigate any possible restrictions and to identify extreme cases. Additionally, error estimates are going to be conducted accompanied with accuracy, resolution and productivity measurements.

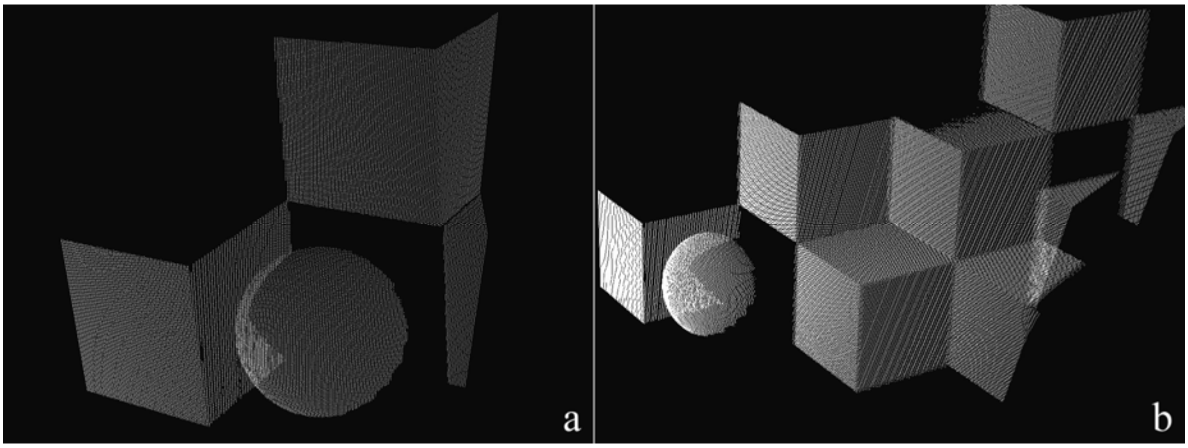

Figure 4: *The result of the digitization process (point cloud) at different time instances.*

3. Acknowledgements

This work is partially supported by the Greek GSRT grant PENED03#679.

4. References

[BFB*98] BORGHESE N. A., FERRIGNO G., BARONI G., SAVAR R., FERRARI S., PEDOTTI A.: AUTOSCAN: A flexible and portable scanner of 3D surfaces, *IEEE Comput. Graphics I/O Devices* (1998), pp. 1-5.

[BPM06] BAHMUTOV, G., POPESCU, V., MUDURE, M.: Efficient large scale acquisition of building interiors, *Computer Graphics Forum*, Vol. 25 (3) (2006).

[Gru01] GRUEN A.: Least Squares matching: A Fundamental Measuring Algorithm, *Close Range Photogrammetry and Machine Vision* (2001), Edited by K. B. Atkinson, pp. 217-255.

[Han06] HANDYSCAN 3D, http://www.handyscan3d.com (2006).

[Heb01] HEBERT P.: A Self-Referenced Hand-Held Range Sensor, *Third International Conference on 3-D Digital Imaging and Modelling (3DIM '01)* (2001).

[Jen04] JENNY A.: Automatic Panorama Recognizing Tool for Panotools , http://autopano.kolor.com (2004).

[Low99] LOWE D.: Object recognition from local scale-invariant features, *in ICCV99* (1999), pp. 1150–1157.

[LPC*00] LEVOY M., PULLI K., CURLESS B., Rusinkiewicz S., Koller D., Pereira L., Ginzton M., Anderson S., Davis J., Ginsberg J., Shade J., Fulk D.: The Digital Michelangelo project: 3D scanning of large statues, *SIGGRAPH 2000*.

[Pol06] POLHEMUS FastSCAN, http://www.fastscan3d.com (2006).

[Pom99] POMASKA G.: Stereoscopic Images for Visualization of Cultural Heritage, *International Archives of Photogrammetry and Remote Sensing Vol. XXXII, Part 5W11* (1999), Thessaloniki, Greece, pp. 89-94.

[PSB03] POPESCU V., SACKS E., BAHMUTOV G.: The ModelCamera: a HandHeld Device for Interactive Modelling, *Fourth International Conference on 3-D Digital Imaging and Modelling (3DIM '03)* (2003).

[PVV*03] POLLEFEYS, M., VAN GOOL, L., M. VARGAUWEN, CORNELIS, K., VERBIEST, F., TOPS, J.: 3D capture of Archeology and Architecture with a hand-held camera, *Proc. Of the ISPRS Workshop on vision Techniques for Digital Architectural and Archaeological Archives*, July 1-3 2003, Ancona, Italy, *The International Archive of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 5/W12*, (2003) , pp. 262-267.

[RHL02] RUSINKIEWICZ, S., HALL-HOLT, V., LEVOY, M.: Real-Time 3D Model Acquisition, *SIGGRAPH 2002*.

[ST94] SHI J., TOMASI C.: Good Features to Track, *IEEE Conference on Computer Vision and Pattern Recognition* (1994), pp. 593-600.

[TAK*03] TSIRLIGANIS N., ARNAOUTOGLOU F., KOUTSOUDIS A., PAVLIDIS G., CHAMZAS C.: 3D-ArchGIS: Archiving Cultural Heritage in a 3D Multimedia Space, *DigiCULT.Info Issue 6*, (Dec. 2003).

[TSP00] TSIOUKAS V., STYLIANIDIS E., PATIAS P.: Epipolar Images for Close Range Applications, *International Archives of Photogrammetry and Remote Sensing Vol. XXXIII, Part B5* (2000), Amsterdam, The Netherlands.

[TWV*99] TAKATSUTA M., WEST G. A. W., VENKATESH S., CAELLI T. M.: Low-cost Interactive Active Monocular Range Finder, *in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* (Jun. 1999), Fort Collins, CO, USA, pp. 444-449.