A Methodology for 3D Registration of Range Images For Object Visualization

DANIEL CÂMARA, MÁRIO FERNANDO MONTENEGRO CAMPOS

DCC/UFMG, Departamento de Ciência da Computação, Universidade Federal de Minas Gerais
Caixa Postal 702 - Cep 30123-970, Belo Horizonte, MG, Brazil
danielc,mario@dcc.ufmg.br

Abstract. The main goal of this paper is to describe a method of image description in VRML, using a device for tridimensional image acquisition. The information obtained from the capture of each object face is organized and a description of the scanned object is made in VRML format. This paper presents a study of the minimum number of captures necessary to get the complete description of the object, shows the problem of face coherence, superposition of points and occluded faces, and discusses the solutions for these problems.

Keywords: 3D visualization, image coherence, depth map, 3D scanner.

1 Introduction

Image Registration is a fundamental task in image processing. It is used to match two or more pictures taken, for example, at different times, from different sensors or from different viewpoints. A broad range of techniques have been developed to solve this problem. This paper presents a methodology for matching images obtained from a three-dimensional scanner, that was developed in the Robotics, Computational Vision and Active Perception Laboratory at DCC/UFMG [13]. The main goal of this work is to provide a mechanism that helps the visualization and manipulation of digitalized objects (Figure 1).

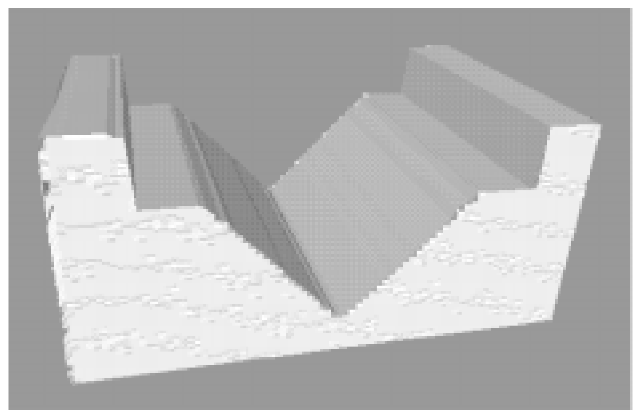


Figure 1: Three-dimensional visualization of a scanned object.

Many applications need to capture and visualize three-dimensional images. The applications for 3D scanners vary from assembling lines, where the objective is to increase the quality control of the process, to the project of new products. Working with CAD (Computer Aided Design), it is possible to project and test products before they are constructed, but this task is not simple with complex objects. The digitalization of a sculpture, for example, in a CAD is a complex process, because its details, shapes and proportions must be maintained. With the method described here, this digitalization becomes relatively simple. A better solution is to work with the two systems, using the image obtained from the scanner and the image generated by CAD. In this way, it is possible to compensate the difficulty of the CAD's modeling, because the scanner provides a good object approximation. On the other hand, some problems with the objects that are inherent to the scanner capture, can be solved by CAD.

The 3D scanner generates the depth map of the object, like the one shown in Figure 2. In the depth map the object is described in a discrete way, in terms of its depths diferences (Z axis) (Figure 3).

Unfortunately, only the depth map does not satisfy all the user's needs. Many times, the goal of the object digitalization is to allow its manipulation and visualization in the three-dimensional space, and the tests of its properties separately or interacting with other objects. With a three-dimensional visualization, it is possible to simulate complex objects that were not constructed in the real world yet. In the case of new products, project failures in its projects may be detected in an early step, saving time and avoiding future problems.

In the next section several algorithms and techniques related to image registration will be discussed. In section 3, the proposed methodology will be presented, showing its main features and problems. Section 4 explains the geometric fusion and, in section 5, the object visualization in VRML will be discussed. In section 6 the conclusions of this work will be presented and finally, in section 7, some future work will be suggested.

2 Background

The goal of registration is to transform sets of surface measurements into a common coordinate system. This is necessary, in this case, because most objects self occlude, no single range image suffices to describe the entire object.

Besl and McKay presented in [2] the Iterative Closest Point (ICP) algorithm to register two sets of points on a free-form surface. ICP is a general-purpose, representation-independent method for an accurate and computationally efficient registration of 3-D shapes including free-form curves and surfaces. Extensions of this algorithm are now widely used for registration of multiple sets of surface data. The original algorithm registers two point sets provided one is a subset of the other and that the transform between the sets is approximately known.

Modifications to the original ICP algorithm have been made to improve the rate of convergence and to register partially overlapping sets of points. Chen and Medioni in [5] demonstrated the registration of partially overlapping range image data. In their work, the rotation imposed to the object between successive captures vary from 15 to 20 degrees, so it is necessary 18 to 24 captures to registrate a object completely.

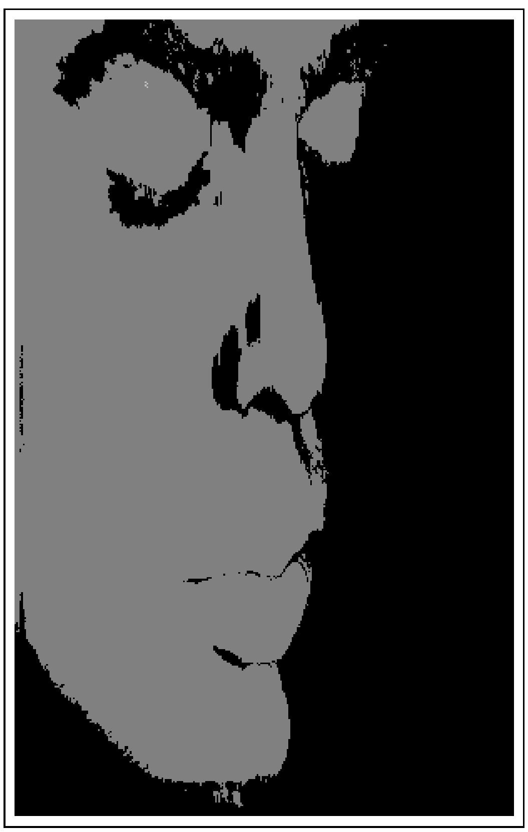


Figure 2: Depth Map generated by the scanner.

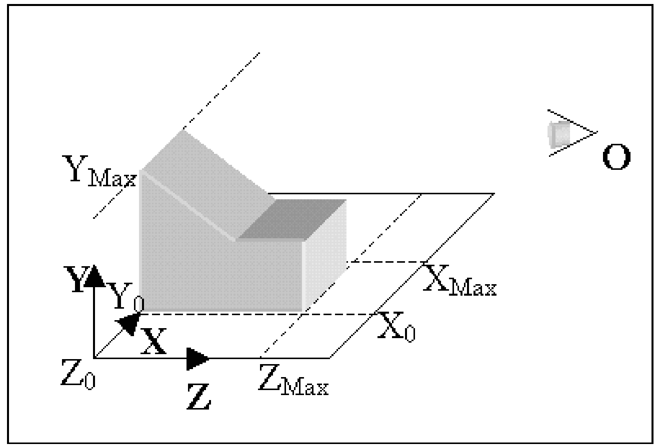


Figure 3: Object dimensions. The point O represents the viewer's position and it is where the camera is positioned. X_0 and X_{max} are, respectively, the first and last columns of the object captured by scanner. Z_0 is the system origin, where the depth Z is 0. Z_{max} is the greatest distance Z from origin Z_0. Y_0 is the first line where the object is captured, and Y_{max} is the greatest one.

Boissonnat [4] and Rutishauser et al. [12] use a graph-based approach to determine correspondences between overlapping meshes. Retriangulation of two overlapping meshes to form a single mesh is then performed using local 2D constraints on triangle shape. Turk and Levoy [14] integrate overlapping triangulated meshes using a 'zippering' approach. Overlapping meshes are eroded and the boundary correspondences found by operations in 3D space. A local 2D constrained triangulation is then used to join the overlapping mesh boundaries. This approach may fail for complex geometries if incorrect boundary correspondences are found. This can occur in regions of high surface curvature and thin object sections.

Curless and Levoy in [6] proposed an volumetric representation that consists in a cumulative weighted signed distance function, working with one range image at a time. The method first scan-convert the image to a distance function, then combine it with the data already acquired using a simple additive scheme. In order to achieve space efficiency, it employs a run-length encoding of the volume, and to achieve time efficiency, the method resamples the range image to align with a voxel grid and traverses the range and voxel scanlines synchronously. The manifold is generated by extracting an isosurface from the volumetric grid. The method has the great ability to fill gaps in the reconstruction, and it attains a great level of detail. But the method is mathematically expensive, and need a large amount of images to describe a object.

Hoppe et al. [8] introduced the use of an intermediate implicit surface representation for reconstruction from unstructured sets of 3D point. A Euclidean minimum spanning tree is constructed to estimate the local surface topology based on the point-to-point distance. A local planar fit to a point neighbourhood is used to estimate the surface orientation. An implicit surface representation is then constructed based on the zero-set of a signed field function which is the distance to the nearest point on the surface. Marching Cubes implicit surface polygonisation is then used to reconstruct an explicit triangulated surface model. Lorenson and Cline [9] developed the Marching Cubes algorithm for triangulation of iso-surfaces in discrete volumetric field function representation such as 3D medical images. Implementations of Marching Cubes are widely available such as Bloomenthal [3]. There are two major limitations on this approach due to the computational cost of constructing the minimum spanning tree and restrictions on the object shape. Estimation of surface topology based on point-to-point distance limits the complexity of the surface geometry that is correctly reconstructed. This approach will fail for regions of high curvature or thin surface sections. To overcome this problem techniques have been developed to construct an intermediate implicit surface representation from multiple 2.5D range images using the image structure to estimate the local surface topology.

3 Methodology

3.1 Scanner 3D

The scanner developed in the Robotics, Computational Vision and Active Perception Laboratory at DCC/UFMG captures, in a CCD camera (Charged Couple Device) the plans formed by a light laser generator (Figure 4). The camera and the laser generator are dislocated linearly throughout the object that will be scanned. The depth of the points in the plan is determined by a geometric model. This model calculates the depth by simple triangulation, assuming perspective projection, or by the cross-ratio model [7]. The device produces a depth map of the object (Figure 2). In [13], a more complete description of the methods and problems of the device can be found.

The depth map generated by the 3D scanner is related to the object's visible face. The camera and the laser generator are linearly dislocated while the object remains fixed. So, the scanner can only get one face description each time (Figure 4). From now, each capture will be called a face.

It is necessary, therefore, more than one capture to digitalize the entire object. The next step of the process is to determine what is the minimum number of captures necessary to obtain a complete description of the object.

3.2 Occlusion

The minimum number of captures necessary to obtain a complete description of the whole object is related to the occlusion's problem. This problem, typical to this kind of dispositive, occurs in a scene when a object is occluded by another, or the object has concave or convex extremes that occlude other parts of the same object [10]. When this problem occurs, it becomes impossible to determine precisely the shape of what is in the shadow area.

The scanner has some other limitations in relation to the face's capture. As shown in Figure 4, the camera and the laser generator are not in the same line. Because of this fact, in some object's points, depending on its shape, the laser light reflex may not be captured by the camera. Some examples are holes in the object and concave objects.

According to [11], to find a point $p_i(x, y)$ in the plan being digitalized, is necessary to find all the directions from where the point $p_i(x, y)$ is visible. To determinate if the point $p_i(x, y)$ is visible from the direction j, is necessary to determine if the points $p_k(x, y)$ of the scene throughout this direction j occlude the point $p_i(x, y)$. The relation between the distance l (distance between the two points of the plan being digitalized) and the occlusions length l', caused by $p_k(x, y)$ in direction j, is defined by:

$$l' = \frac{h(p_k) - h(p_i)}{\tan(h(p_k))}$$

where $h(p_i)$ and $h(p_k)$ are the distances between the scene and the points $p_i(x, y)$ and $p_k(x, y)$, respectively. If the distance between $p_i(x, y)$ and $p_k(x, y)$ is greater than l', so the point $p_i(x, y)$ is visible in direction j (Figure 5). Each point $p_i(x, y)$ is visible from several different directions.

Once characterized the occlusion problem, a specific case can be analyzed. The registration of a sphere will be analyzed, because it is one of the worst cases relative to convex objects for this device. Only $2/3$ of the sphere's visible face is shown in the depth map (Figure 6) when the sphere is digitalized. The lighter colors indicate the deeper areas, the darker colors indicate the flat ones. The black color shows not captured points. The sphere's part that is not shown in the depth map was occluded by the part nearest to the camera.

3.3 Overlapping

With this device, it is impossible to completely map one sphere's face. However, if the capture is repeated rotating the object, a registration between the faces can be done finding regions that where digitalized twice. With these regions, it is possible to minimize the occlusion problem. The Chen and Medioni's method [5] obtain good results, but its complexity and the number and precision of captures required are inadequate for the purpose of this work.

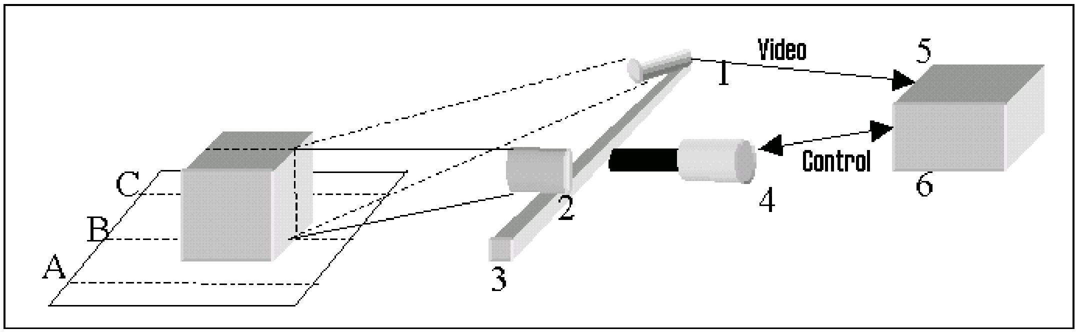


Figure 4: Scanner description: camera (1), laser generator (2), rigid base (3), horizontal translation system (4), digitalizer sheet (5), computer (6). B is the plan being digitalized, A and C are the visualization volume limits.

Observing the depth maps of the sphere, it can be seen that with 90^o rotation and four captures, a description of the entire object can be obtained (Figure 7 (a)). No points are lost due the occlusion, because there is an overlap between the adjacent captured faces, in relation to occluded points. In this manner, the sphere can be completely described (Figure 7 (b)).

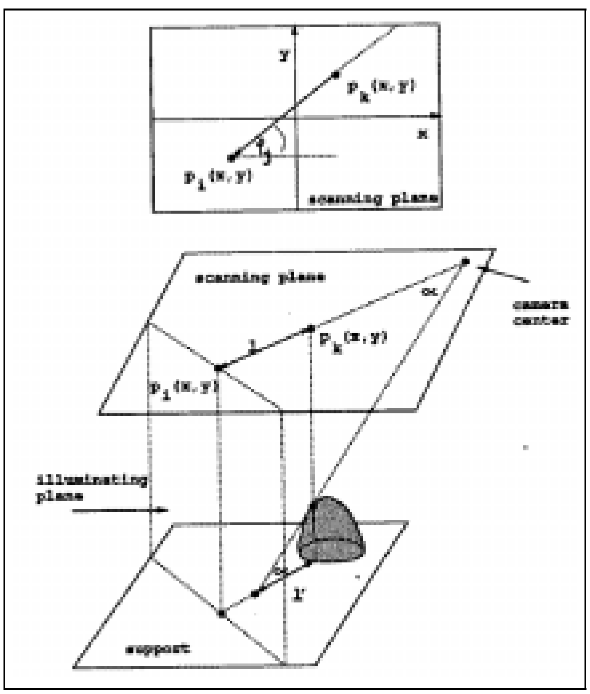
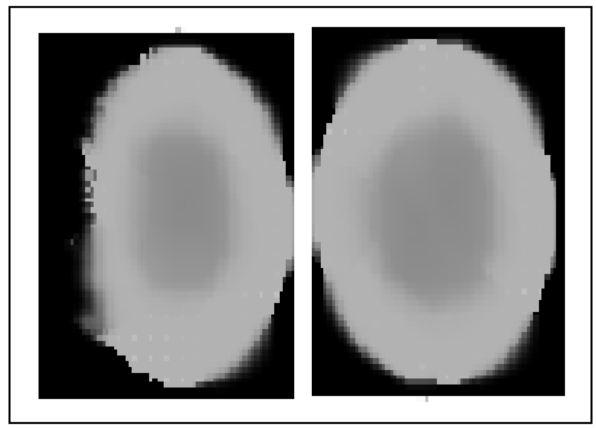


Figure 5: Description of a shadow area [11].

3.4 Referential's normalization

In order to make the union of the captured faces, the first step to be done is the referential normalization. The first captured face, face 1, has the referential considered correct. All the other faces will be moved in relation to the first face



(a) (b)

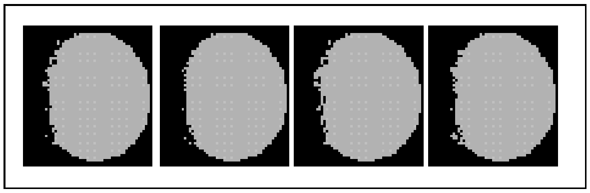
Figure 6: Sphere's depth map. With the occlusion problem (a) and without the occlusion problem (b).

referential. It will be considered that X_0 and Z_0 are the scene's origin, and its values are equal 0. The second face is the object's right face, and its depth (Z) must be relative to the width (X) of the face 1. The depth (Z) captured from face 2 is converted to width related to face 1 (X), and its width is translated to depth in relation to face 1. When face 3 is normalized, it is made an inversion on its values, the greatest $Z(Z_{max})$ becomes the smallest $Z(Z_{min})$, and the same occurs with X. Table 1 shows the normalizations.

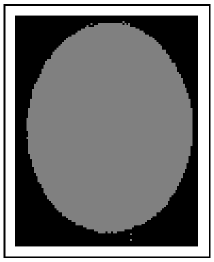
With all the captures in the same referential (Figure 8 and Figure 10(a)), the next step is to make the face registration. The assembled sphere is shown in Figure 9.

3.5 Registration

The next step is to know if the faces have common points, and discover such points if they exist. In this step, it is especially important to find common points that can recover some occluded region. The traditional methods of image registration try to find correspondence between dif-



(a) Sphere rotated 0^o, 90^o, 180^o and 270^o.



(b) After the intersection of the overlap areas from four captures.

Figure 7: Depth maps of a sphere.

ferent images. This kind of correspondence requires a lot of mathematical processing, becoming, in many cases, computationally intensive. In this work, as the depth maps are in the same referential, the registration is performed trying to match the same points X, Y and Z in adjacent faces. The capture may be a little imprecise, so the points do not need to be the same, but must be within a near area. In this way, it is necessary to consider a error zone, that is, if two points have a difference less than the error zone, they can be considered the same point.

The error zone, however, must not undertake the confidence of neighbor points. If this value is chosen in an arbitrary way, the integrity of the neighbor points can be damaged. In this work this parameter is not automatically found, but it is possible, from a previous face analysis, to obtain an approximation for this error zone.

Face	X in Capture	Y in Capture
1	X_1	Z_1
2	Z_2	X_2
3	$X_{3Max} - X_3$	$Z_{3Max} - Z_3$
4	$Z_{4Max} - Z_4$	$X_{4Max} - X_4$

Table 1: Normalization of the captured face's points. X_1 is the coordinate of the column of the point captured in face1, Z_1 is the distance Z in relation to origin of face 1. Z_2 is the distance Z of the point in the capture of face 2 and X_2 is the X coordinate of the point in the face 2. X_{3max} is the greatest X captured in face 3 and X_3 is the point of the face 3 that is being translated. Z_{3max} is the greatest Z captured in face 3 and Z_3 is the point that is being analyzed. For the face 4, it is analogous.

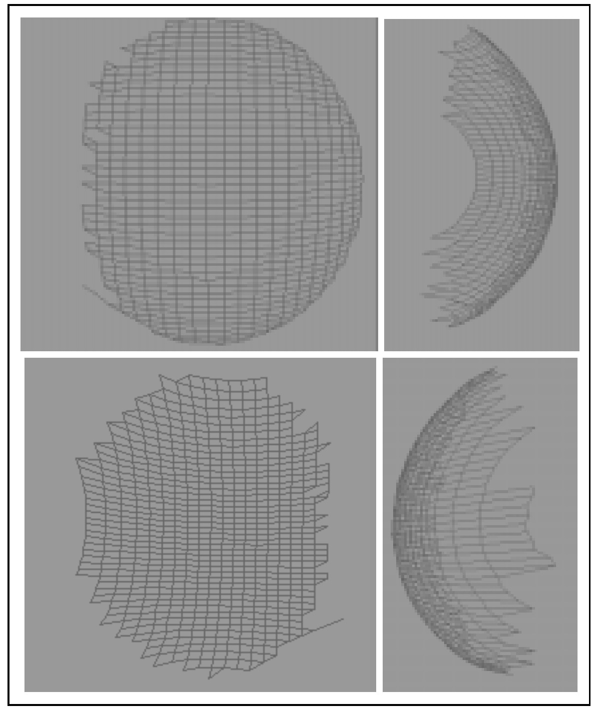


Figure 8: Sphere's faces placed in the same referential.

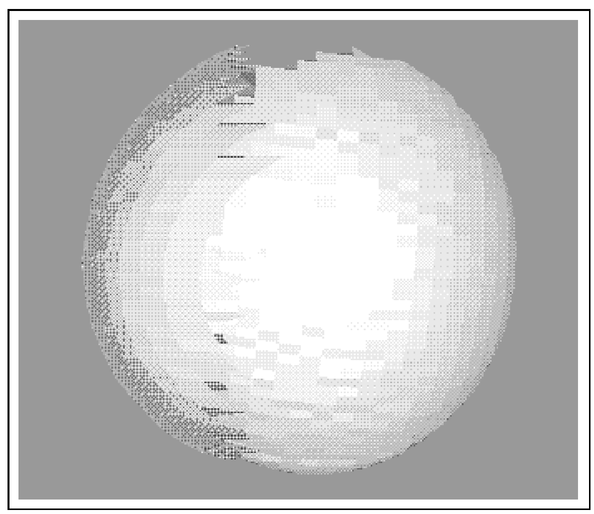


Figure 9: Digitalized Sphere.

This way of modeling the problem is very sensible to failures in the captures. If the capture was not initialized exactly in the same place or the object was not rotated exactly 90 degrees, the digitalization may suffer distortions. Points that the method considers the same can actually not be, causing mistakes in the final object representation.

3.6 Object's Superior and Inferior Face

There are parts of the object that are occluded in all captures, independently of the way the rotations are done or if the object's face are or not in the same referential. Because of the scanner's limitation, there are no information about the object's inferior and superior faces. Two more captures could be done in order to determine the shape of these faces. However, as explained before, the method is very sensible to mistakes in captures, and would be probably difficult to rotate the object exactly 90 degrees in these directions. But, with information about the four lateral faces, it is possible to determine the probable shape of the two faces not captured by the device.

In order to analyze the superior face, it is assumed that the opposite faces have the same Y for the correspondent points of the faces. If this fact is not true, it is assumed that some error occurred in the object capture and the correspondence can not be guaranteed. With the opposite plan having the same high, the faces having the minimal Y are founded. From this Y value, plans are drawn between the opposite faces, creating the object's superior face (Figure 10(b)). The same methodology can be applied to the object's inferior face. This simple method can solve many cases, but unfortunately not all of them. For example, it is not possible to determine concave faces, as shown in Figure 11.

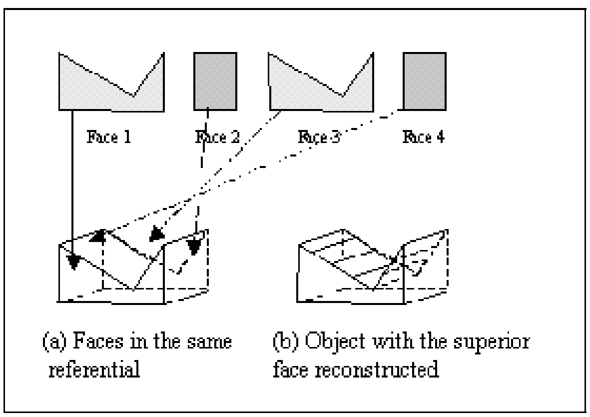


Figure 10: Captured faces and the object assemble.

In this work, the universe of objects that can be digitalized will be restrict to that ones without concave superior or inferior faces. With a little care in the object positioning for the capture, this restriction does not become strong. For other objects with wedge shape, like pyramids or cones for example, a perfect capture can be done, because the device can easily capture its different depths.

4 Geometric Fusion

With the faces in the same referential and the common points matched, the object is automatically recreated. The next step is the creation of the VRML description file. Each object's face is decomposed in several smaller plans, making possible to increase or decrease the digitalized image resolution (Figure 12). In some cases, a detailed object description is not necessary, however, in other cases it can be essential. A detailed VRML description file, can occupy several megabytes depending on the object size. For this reason, sometimes it is desired that the resolution can be decreased, also decreasing the size of the scene's description file.

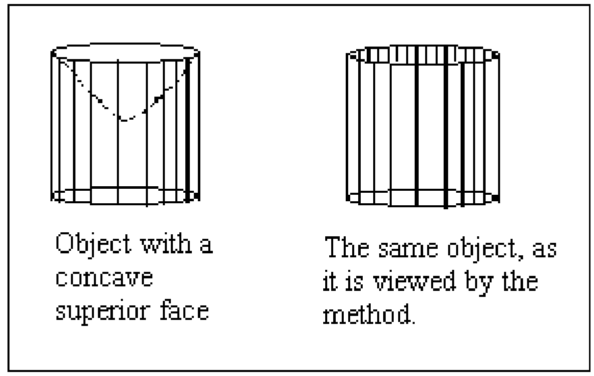


Figure 11: Example of an object that can not be described.

This resolution decrease is made automatically using different samplings. The sampling is made taking one point for each set of N points. For example, in a three-resolution, one in each three depth map points is used to form the object's face. In the greatest resolution (one-resolution) plans are drawn between all the points of the depth map.

5 Visualization

In this work, the VRML language (Virtual Reality Modeling Language) [15, 1] was chosen for object description, not only because it is becoming a standard, but also because it has compilers in several platforms. The language flexibility and portability is, in this case, one of its main features. The user does not become dependent to only one platform, he can work with the captured object in any platform without need of changes or updates in the scene.

In the depth map returned by the scanner, there is no information about color, texture or brightness of the object. All of these parameters can be inserted in the VRML description file, obtaining more realistic object visualization. Beside this, changing one simple parameter, the object can be shown in wireframe mode, facilitating its three-dimensional visualization.

The VRML language, besides its portability, allows the image generation with a high definition and resolution, as can be seen in Figure 13, the result of a human face digitalization or in Figure 14, a digitalized Greek statue.

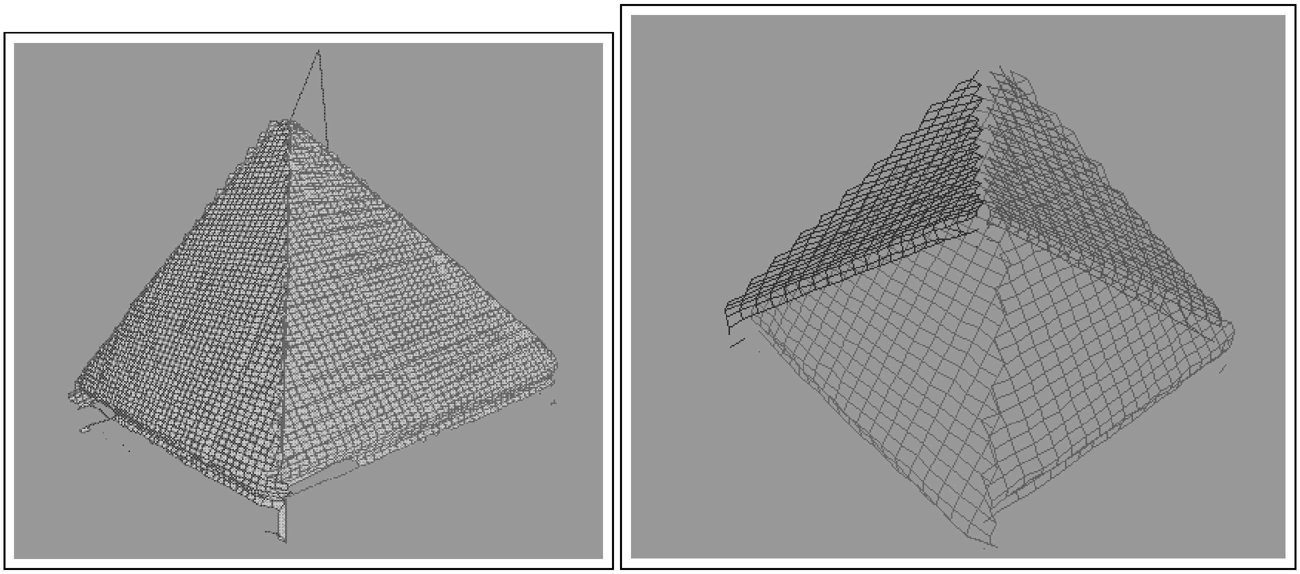


Figure 12: Two visons of the same digitalized pyramid, with distinct granularities.

6 Conclusion

This work presented a methodology for 3D registration of range images, with the purpose of transpose the digitalized object to a virtual world. In few words, the proposed methodology consists in four captures rotated 90 degrees of the same object. All the captures are placed in the same referential, and it is made the registration and the face geometric fusion.

The paper showed that the methodology is simple and has an easy implementation. Unfortunately, the model is very sensible to mistakes in the object captures and approximation errors caused by the object digitalization. Now, it is necessary to improve its robustness and decrease its failure

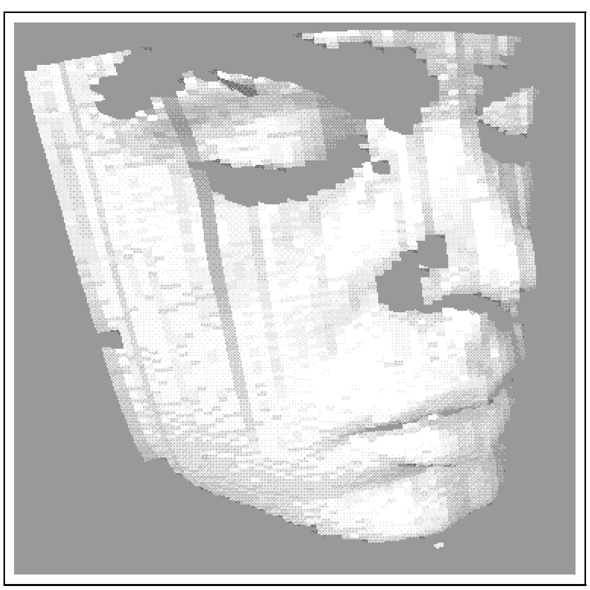


Figure 13: A digitalized human face using the depth map shown in Figure 2.

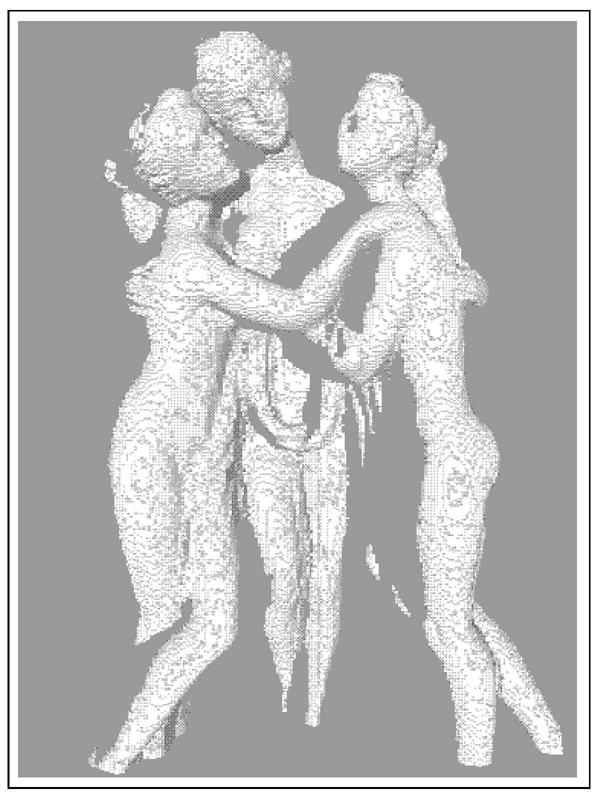


Figure 14: Face of a digitalized Greek statue.

sensibility, increasing the precision in description of scene details.

Independently of the first objective of this scanner, that was the acquisition of images of residues in train vagons or assembling lines [13], this work showed that is possible to use it to visualize a great number of objects.

With the possibility of object visualization and manipulation, it is possible to create, from simple objects, scenes more complex and interesting. Complex products can be projected and its interaction can be simulated in the virtual world, without the need to build models or prototypes.

7 Future Works

Several aspects can be improved, both in the face capture and in the treatment of the information obtained from the depth maps.

Today, with the four faces captured, the whole object creation process is automated. However, the error zone, as mentioned before, is not found automatically yet. This is another aspect that can be improved in this work. It can be also explored the local principle of points. For example, the greatest Z in face 1 may probably be the lowest X in face 2. From this fact, it can be concluded what is the error zone for the other image points.

A revolving platform is being constructed to minimize the capture mistakes. With this platform, other techniques for object creation from the depth maps can be evaluated. For example, the method proposed by Chen [5] can be implemented, and its performance can be compared with the one of this work.

With the new platform, another possibility is to change the scanner referential. Moving the object and maintaining the camera and the scanner fixed, a new depth map evaluation can be done. In this way, it is possible to evaluate which configuration is more adequate to the object digitalization. Another important point is that the scanner does several approximations and considerations about the captured object. One of these considerations, necessary to the perspective projection, is an adequation of the distance between the face being digitalized and the laser light. This adequation can cause mistakes in the capture, which can be observed in the left face in Figure 1. This point can be improved, both in the scanner and in this methodology.

It is also possible to improve the multiresolution of the method. The face can be analyzed, merging points that lay in the same plan. In this way, the number of points required to the object creation is decreased. Another improvement is try to work with other format of plans. Today, all the plans that form the image are rectangles. With triangles and squares, for example, it is possible to decrease the number of plans required to object creation and increase its resolution (Figure 12).

8 Acknowledgements

The authors would like to thank John Kennedy Schettino de Souza, by the support in the step of this project, helping with the scanner handling. The authors are also grateful to CNPq and FAPEMIG for the financial support of this work.

References

[1] G. Bell, R. Carry, and C. Marrin. VRML 2.0 final specification. 1996.

[2] Besl and McKay. A method for registration of 3d shapes. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992.

[3] Jules Bloomenthal. An implicit surface polygonizer. In Paul Heckbert, editor, *Graphics Gems IV*. Academic Press, Boston, 1994.

[4] Boissonnat. Geometric structures for three-dimensional shape representatio. *ACM Transactions on Graphics*, 1984.

[5] Yang Chen and Grard Medioni. Object modeling by registration of multiple range images. In *International Conference on Robotics and Automation*, 1991.

[6] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *Computer Graphics (SIGGRAPH '96 Proceedings)*, pages 303–312, 1996.

[7] Richard O. Duda and Peter E. Hart. Pattern classification and scene analysis. In *Wiley-Interscience Publication*, 1973.

[8] Hugues Hoppe, Tony DeRose, Tom Duchamp, John McDonald, and Werner Stuetzle. Surface reconstruction from unorganized points. volume 26, pages 71–78, July 1992.

[9] W. E. Lorensen and H. E. Cline. Marching cubes: a high resolution 3D surface construction algorithm. In M. C. Stone, editor, *SIGGRAPH '87 Conference Proceedings (Anaheim, CA, July 27–31, 1987)*, pages 163–170. Computer Graphics, Volume 21, Number 4, July 1987.

[10] A. D. Marchall and R. R Martin. Computer vision , models and inspection. In *World Scientific*, 1992.

[11] Jasna Maver and Ruzena Bajcsy. How to decide from the first view where to look next. Technical report, GRASP LAB, 1990.

[12] Stricker Rutishauser and Trobina. Merging range images of arbitrarily shaped objects. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1994.

[13] John Kennedy Souza Schettino and Mrio Fernando Montenegro Campos. Um dispositivo automtico para obteno da forma 3d de objetos. In *Sibgrapi 98*, 1998.

[14] Turk and Levoy. Zippered polygon meshes from range images. In *Computer Graphics Proceedings SIGGRAPH*, 1994.

[15] Vrml 97, international specification iso/iec is 14772-1, December 1997.