# Adaptive Space Carving

Anselmo Antunes Montenegro
Instituto de Matemática Pura e Aplicada
ansmont@superig.com.br

Marcelo Gattass
Departamento de Informática / PUC-Rio
mgattass@tecgraf.puc-rio.br

Paulo Cezar Pinto Carvalho
Instituto de Matemática Pura e Aplicada
pcezar@impa.br

Luiz Carlos Pacheco Rodrigues Velho
Instituto de Matemática Pura e Aplicada
lvelho@impa.br

## Abstract

*In this work, we present an adaptive space carving method for scene reconstruction from a set of images obtained from low cost calibrated webcams. Our method uses a combination of silhouette and photometric information to efficiently carve the shape of the observed scene out of a volumetric space represented by an octree data structure. In this method different resolutions are considered both in object space and in image space. This led us to adopt a strategy in which the information used by the photo-consistency test is registered in scene space by projective texture mapping. Another important question addressed in this work is the high level of noise present in low cost webcams. To deal with this problem we devised a statistical photo-consistency test that uses statistical estimators for the noise introduced by the sensors of the cameras.*

## 1. Introduction

*Space carving* methods are considered one of the best approaches to practical small scale scene reconstruction from a set of images obtained from commodity cameras disposed around the object.

In these methods, the scene of interest is assumed to be bounded by a *reconstruction volume*, which is typically divided in a set of voxels. Then, a sequence of queries are applied to the voxels to determine which of them are able to reproduce, within an error tolerance, the input images of the scene when rendered from the original viewpoints.

Although several contributions were made to the state of art of scene reconstruction by space carving, some important aspects still remain not completely explored. For instance, the problem of efficient model reconstruction and economical representation of the reconstruction volume was only investigated in [8, 9].

In this work we present an adaptive space carving method that works on an octree representation of the reconstruction space and uses a combination of silhouette and photometric constraints. We show that an octree representation is not only an efficient way to represent the volume of occupancy of the scene, as claimed in [9], but also enables fast model reconstruction. This is made possible by combining the adaptive nature of octree data structures with a photo-consistency test that uses both photometric and silhouette information in increasing levels of resolution. To deal adequately with the reconstruction problems that arise as different levels of resolution are considered, we adopt a simple mipmapping based strategy combined with supersampling. Another contribution of this work is a new photo-consistent test that accounts for the high levels of noise introduced by the sensors of the cameras.

This article is organized as follows. In section 2, we summarize the main concepts related to the space carving methods. In section 4, we introduce the adaptive space carving approach. In section 5 we describe the implemented acquisition system. In section 6, we present some results and finally, in section 7, we make some considerations about the work presented here.

## 2. Space Carving Background

Space carving is a class of volumetric based methods used to solve the problem of reconstructing the shape and the color information of a scene from a set of input images.

**Definition 1 (Image based reconstruction problem)** *Let $S$ be a tridimensional scene in a closed subset $U \subset \mathbb{R}^3$ and let $I$ be a set of images of $S$ taken from a set of cameras $C$. Determine the shape $V$, composed by a subset of adequately colored points of $U$, that reproduces the input images when rendered from the original viewpoints of the cameras $C$.*

The problem of scene reconstruction from a set of images is an ill-posed problem. This occurs because, in general, there is more than one shape that is able to reproduce the set of input images of the scene when rendered from the original viewpoints [5].

In order to pose this problem appropriately, some researchers fomulate it as an optimization problem with *regularization constraints* [13]. On the other hand, Seitz and Dyer [10], and later, Kutulakos and Seitz [5] proposed a completely different approach based on the concept of point photo-consistency.

**Definition 2 (Point photo-consistency)** *An infinitesimally small point $p$ on a lambertian surface $S$ is photo-consistent with a set of images $I$ if the projections of $p$ in the images in which $p$ is visible are of the same color.*

Seitz, Dyer and Kutulakos, based on a photo-consistency analysis of subsets of points of the 3D space, identified the existence of a shape with unique properties. This shape consists of the maximal shape that subsumes all photo-consistent shapes and was denominated *Photo Hull*. It was also proved that the *Photo Hull* can be obtained by the systematic removal of all non-photo-consistent points of the scene reconstruction space. This strategy used to obtain the $PhotoHull$ is the key idea behind all space carving methods.

In practice, some problems must be taken into consideration in the specification of a computational method based on this idea. First, we cannot work with an infinite subset of points. Second, we cannot rely on a photo-consistency criteria that is based on the color equality of infinitesimal points and that assumes that the scenes are perfectly lambertian. Consequently, in space carving we represent the reconstruction space discretely as an array of voxels and use photo-consistency criteria based on statistic measures.

The statistic measures are used to verify if the difference observed among the colors in the projections of a voxel $v$ are explained by a statistical model of error that takes into consideration the noise introduced by the camera sensors and the non-lambertian components in the radiance of the scene surface.

Several statistic measures can be used. Seitz, for instance, proposed a likelihood ratio test, based on a $\chi^2$ statistic. More recently, Broadhurst proposed the use of the F-statistic measures in his multiple threshold voxel coloring [2]. Other approaches, based on more complex probabilistic models, were proposed by Bhotika [1] and also by Broadhurst [3].

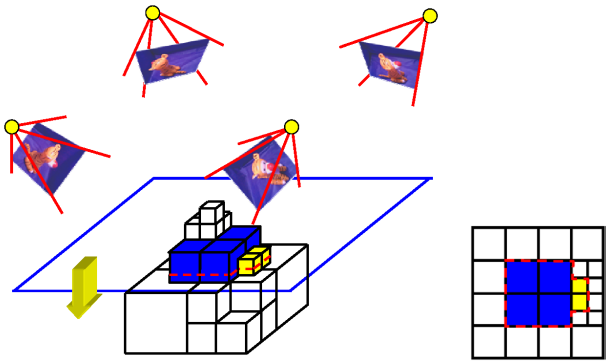
Another crucial problem in space carving is visibility determination. Seitz and Dyer's Voxel Coloring assumes that all points in the reconstruction space can be globally ordered in relation to the calibrated camera centers. This assumption simplifies considerably the determination of visibility of the partial scene, nevertheless, at the expense of greater generality. The existance of such ordering is also assumed in this work.

The most simple mechanism to deal with visibility questions is to associate visibility maps to each of the cameras. Initially all elements in the visibility maps are set to zero. Then, when a consistent voxel is found, all visible pixels in its projections in the input images are set to 1. This strategy is used in [10] and [5]. Other more intricate data structures can also be used, such as the *layered depth images* and the *item buffers* described in [4].

## 3. Adaptive Space Carving

Voxel based space carving is a simple and powerful way to reconstruct the shape and the color information of a scene. Nevertheless, due to its intrinsic structure, voxel carving spends too much effort evaluating a large number of small elements that do not belong to the scene surface which is the real target of 3D reconstruction process.

In order to avoid this, we propose here an adaptive space carving algorithm by progressive refinement. Our idea is to use the partial knowledge of the structure of the reconstruction space which is determined concurrently with the classification process to determine more efficiently which regions are non-photo-consistent. This can be implemented by applying a sequence of space carving iterations on an octree data structure that represents the space in which the problem is defined (Figure 1(a)).



(a) Adaptive Space Carving (b) Selection

**Figure 1. (a) - Adaptive space carving on a octree. The cells are processed in the order shown by the arrow. (b) - selection of cells in a layer at a give level of refinement (yellow cells are the ones to be processed)**

The use of octree data structures to represent volumetric reconstruction spaces was overlooked when voxel based space carving methods were introduced. This can

be partially understood if we consider that Voxel Coloring/Space Carving methods are inherently associated with point photo-consistency, which cannot be easily generalized to octree cells.

Prock, for example, proposed a space carving method that, although adaptive, does not use any hierarchical spatial data structure [8]. Sainz, on the other hand, re-introduced the use of octree representations in space carving, but only to store photo-consistent regions efficiently, which were, in fact, obtained by voxel based carving [9].

In a certain way, the basic structure of the method proposed here is similar to the *Octree Carving* proposed by Szeliski [11]. The difference in Szeliski's Octree Carving is that it uses only silhouette constraints whereas the adaptive space carving proposed here uses both photometric and silhouette constraints.

In the beginning of the process, we start with a coarse representation of the reconstruction space which is defined by a bounding box of the scene associated to the cell in the root of the octree. Then, we try to classify this initial cell according to some photo-consistent criterion. If we succeed to classify the cell, either as photo-consistent or as non-photo-consistent(transparent), then we can finish processing that cell and attribute the adequate color to it. On the other hand, if we fail to classify it, that is, if we can not decide precisely whether the cell is photo-consistent or not, then we subdivide it into eight new cells and repeat the same classification procedure for each of the new cells created. This process is repeated successively until all regions are classified or a maximum resolution is obtained.

In this work, we assume as in [10] that there is a partial ordering of the points in the reconstruction space according to their distances to the centers of projection of the cameras. Hence, at a given level of refinement we classify the octree cells in the order defined by their distance to the cameras.

In a octree representation we cannot explicitly define layers as in voxel arrays. Nevertheless, layers can also be used to specify what cells are supposed to be classified in a given iteration of a certain level of refinement.

For a certain level of refinement of the octree, we can identify a voxel array structure whose elements have resolution compatible with the cells in the leaves of the octree. The layers of the corresponding voxel array structure induce virtual layers in the octree structure that define the partial order in which the cells are evaluated.

On the other hand, differently from voxel arrays, only the octree cells that belong to the current level of refinement need to be evaluated. The cells that belong to coarser levels of refinement either have been classified or subdivided at the previous level. To determine the active octree cells in a given layer we can simply intersect a plane that approximates such layer with the whole octree. Once the intersection is determined, we select only the cells that belong to the current level of refinement. This strategy is shown in figure 1(b).

Although simple, the application of this scheme requires the solution of the following subproblems: *how to classify an octree cell and when to refine it*, *how to evaluate the photo-consistency of the octree cells* and *how to attribute color to a cell*.

### 3.1. Octree cell classification and refinement

As we described before, original space carving classifies points, in a bounded 3D space, in two groups: *photo-consistent* and *non-photo-consistent* points. Similarly, our adaptive approach must be able to classify all octree cells in groups of photo-consistent and non-photo-consistent cells.

First of all, we must decide whether the concepts of photo-consistency and non-photo-consistency can be generalized to large volumetric regions. In our adaptive method we want to investigate only a subset of the elements inside the region to be classified, otherwise we would fall into the same exhaustive approach as before.

In this case, two classes are not sufficient to convey all circumstances that may arise in a adaptive space carving process. For example, let $oc$ be an octree cell. If we classify $oc$ considering only a subset of its elements, for instance, a planar region inside $oc$, then we cannot guarantee that there is no photo-consistent surface fragment inside $oc$. Another problem is that a large cell will probably contain a combination of background and foreground projecting points.

To the best of our knowledge, only Kutulakos proposed a photo-consistency criteria, based on shuffle transforms, that is able to classify large regions adequately [6]. Neverthless, his concept of photo-consistency is too conservative since it requires that the projections of a photo-consistent octree cell in the input images do not contain any pixels with similar colors in a given neighborhood around their centroids. Furthermore, defining the correct neighborhood for each octree cell projection, in order to take into account adaptive reconstruction, is not a trivial task and is still an open problem.

Instead of searching for other complex criteria, we propose a simpler mechanism that can still keep the conservative property of space carving. We create another classification group, the *undefined cells*, to contemplate those cells that can be neither classified as photo-consistent nor classified as non-photo-consistent at a certain level of refinement.

Hence, we classify octree cells according to their photo-consistency in three groups: *consistent*, *inconsistent*, and *undefined*. In our classification scheme, consistent cells consist of those cells that were classified as consistent according to some photo-consistency evaluation criteria. Thus, the concept of consistent cell is very similar to the concept of consistent point.

On the other hand, we restrict the group of inconsistent

cells to those cells in the highest octree resolution level that failed the photo-consistency test and to those large cells in intermediary levels of resolution that project completely on the background.

The group of undefined cells contemplate those cells that, at a certain stage of the process, cannot be classified, neither as consistent nor as inconsistent. This may occur when a certain cell contains points that belong to empty space together with points that belong to the scene surface, or when the photo-consistency criteria cannot decide, precisely, whether the cell is consistent or inconsistent. The later case happens when the sampled elements used in the photo-consistency checking procedure are not representative enough.

Whenever a cell is considered undefined, it must be subdivided into eight new octree cells that are evaluated in the next space carving iteration, which works in higher resolution.

## 3.2. Photo-consistency evaluation

Photo-consistency evaluation is the kernel of space carving methods. In our approach, the photo-consistency evaluation of a cell $c$ is done at a planar region defined by its intersection with a layer plane(Figure 1). Each layer plane here is called a *registration plane*. The evaluation itself is based on a hypothesis test that relies on information about the scene, stored in the color and opacity channels of image maps. The information corresponding to a cell is registered in scene space via OpenGL [14], by projecting the image maps as texture maps onto the registration planes. In the next subsections, we address in detail the issues involved in the photo-consistency evaluation of an octree cell.

### 3.2.1 Information used by the photo-consistency test

Original space carving uses two kinds of information in the photo-consistency test: the *photometric information* that comes from the input images and the *visibility information* computed as the process is executed. Although many researchers consider it too restrictive for an space carving approach, it is also possible to use *segmentation information*. In our method, segmentation information is used, as well as a new kind of information, namely, the *noise maps of the cameras*. Each noise map consists of an estimate of the noise in each color component of the sensors of a camera given by the standard deviations of the observed values in the components $rgb$.

For each image, we use two $rgba$ maps to store all the information used by the photo-consistency test. The first one, denominated *Photometric and Segmentation Map*($PSM$) stores the photometric information in the channels $r$, $g$ and $b$ , and the segmentation information in the channel $a$. The second map, denominated *Visibility and Noise Map*($VNM$) stores the estimated noise of each color component of the sensors in the components $r$,$g$ and $b$, and the visibility information in the channel $a$.

### 3.2.2 Registration

To evaluate the photo-consistency of a given volumetric element $v$, we must sample the information that will be used in the photo-consistent test by re-projecting $v$ onto the projection planes of the cameras . As the projections of $v$ may contain several elements, these elements must be registered in some way.

In Seitz and Kutulakos' space carving approach, the registration of the information associated with each voxel is trivial because only one element of information is obtained from each camera. When large regions must be evaluated, this procedure cannot be used anymore. In this case, the registration of these elements is very complex, as the polygonal regions associated to the projections of $v$ may have arbitrary shapes and contain several elements in its interior.

To deal with these problem, Prock proposed the registration of all relevant information in scene space (Figure 2). In this strategy each information map is represented as a texture map that is projected on some surface in the scene.

The surfaces used by Prock in the registration process are determined by planes that approximate the voxel layers. In our method, we register all the information on planes that approximate the virtual layers induced by the corresponding voxel array structure that also defines the order of classification of the cells. We call theses planes *registration planes*.

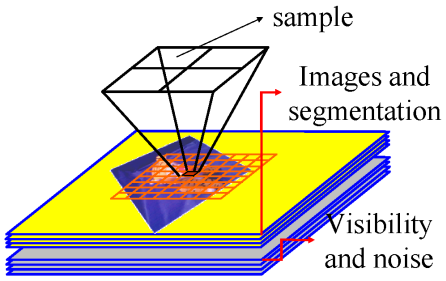



**Figure 2. Registration**

When testing consistency for a given virtual layer, we project the Photometric Segmentation Map and the Visibility and Noise Map for each image onto the registration plane, obtaining the corresponding *Projected Photometric and Segmentation Map* and the *Projected Visibility and Noise Map*.

Our strategy automatically solves the problem of registration and also provided an adequate treatment of aliasing

problems by projecting the textures as mipmaps. Moreover, fast photo-consistency evaluation is also a possibility if graphical hardware programming is used.

### 3.2.3 Sampling

The intersection of a registration plane with an octree cell $c$ defines a planar region $pr$ inside $c$. The number of elements inside $pr$ is determined by the resolution of the projected texture map which must be proportional to the spatial resolution of the current level of resolution of the octree. To make anti-aliasing more robust, we generate the projected information maps with a resolution that is greater than the spatial resolution of the current octree. For instance, in our implementation we produce the maps with resolution equal to four times the spatial resolution of the octree.

Each element inside $pr$ is addressed by indices $s$ and $t$, in a coordinate system associated to the grid induced by the intersection of the registration plane with the supersampled voxels. For each element $e_{st}$ we will take a sample $PPSM^i_{st}$ from the $ith$ *Projected Photometric and Segmentation Map*(PPSM) at the position $(s,t)$. Similarly, we will take a sample $PNVM^i_{st}$ from the $ith$ *Projected Noise and Visibility Map* at the position $(s,t)$.

### 3.2.4 Statistical Photo-consistency Test

First of all, to evaluate the photo-consistency of a given planar region $pr$ associated to a cell $c$ we must determine the set of visible elements $VE$ in $pr$ composed by elements $ve_{st}$ that are visible in some camera. If $VE$ is empty then $pr$ is not seen by any cameras and the cell is trivially classified as *consistent*.

In the next step we verify whether there is a visible element $ve_{st}$ in $VE$ that projects in the foreground of some image. When this is true, the cell is classified as *undefined* if the process is not in the maximum resolution. If the process is already in the maximum resolution, we classify it as *inconsistent*.

When it is not possible to take any decision based on the segmentation and visibility information, it is necessary to check the consistency of the photometric information.

As we mentioned before, due to noise and other sources of error, we cannot test the photo-consistency of the photometric information by checking the exact equality of the samples. Instead we adopt an statistical approach based on a hypothesis test. To simplify the notation used, we will consider now only one color component of the photometric information.

In the statistical approach proposed here, we consider the photometric information obtained from each camera for an element $ve_{st}$ as the realization of a random variable $X_i$, with normal distribution, and mean $\mu$ equal to the corresponding surface color and variance $\sigma_i^2$, corresponding to the level of noise introduced by the sensor. We consider that the variances are possibly different for each image due to the difference of the sensors. In this model, the value of $\mu$ is unknown but the values of $\sigma_i^2$, from each sample $i$, are obtained from the estimates provided by the noise maps.

To check the consistency of the photometric information associated to a give element, we must test the hypothesis of equality of the means corresponding to the different photometric samples. Thus, our problem is to test the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_n$ given samples $X_1, X_2, \cdots, X_n$, independent, with distribution $N(\mu_i, \sigma_i^2)$.

In order to derive a likelihood test, we must obtain the maximum likelihood estimator (MLE), under the hypothesis $H_0$, of the common mean of all distributions. Due to the different values of $\sigma_i^2$, this is not simply the mean of $X_1, X_2, \cdots, X_n$. Rather, one can show (see [7]) that the MLE is given by

$$X' = \frac{\sum_{i=1}^{n} X_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2} \tag{3.1}$$

The corresponding test statistics is given by $\sum_{i=1}^{n} \left(\frac{X_i - X'}{\sigma_i}\right)^2$, whose distribution is given by the following theorem:

**Theorem 3.1** *Let $X_1, X_2, \cdots, X_n$ be independent and $X_i \sim N(\mu, \sigma_i^2)$. Then, $\sum_{i=1}^{n} \left(\frac{X_i - X'}{\sigma_i}\right)^2 \sim \chi^2_{n-1}$.*

The proof of this theorem will not be shown here but is presented in its totality in [7].

In order to test the hypothesis $H_0 : \mu_0 = \mu_1 = \cdots = \mu_n$ we check if $Z = \sum \left(\frac{X_i - X'}{\sigma_i}\right)^2$ belongs to critical region $RC = [0, \chi^2_0]$, such that $P(Z \in RC) = P(Z < \chi^2_0) = \alpha$, where $\alpha$ is a significance value. If $\chi^2_0 \in RC$, we accept $H_0$, otherwise we reject it.

Now, to verify the photo-consistency of a planar region, with $c \times d$ visible elements $ve_{st}$, we can rely on a hypothesis test based on the statistic given by

$$\mathbf{Z} = \frac{1}{c.d} \sum_{s=0}^{c} \sum_{t=0}^{d} \sum_{i=0}^{n} \left(\frac{X_{ist} - X'_{st}}{\sigma^2_{ist}}\right)^2 \tag{3.2}$$

As the sum of $\chi^2$ distributions is another $\chi^2$ distribution we have $\mathbf{Z} \sim \chi^2_{cd(n-1)}$

The filtering operation applied in the registration of the images at intermediary levels of resolution causes some attenuation to the deviations observed in corresponding elements. To compensate this attenuation we propose an adjustment that is based on the fact that the mean $\overline{X}$ of a set $X$ of independent samples $X_1, X_2, \cdots, X_n$ with the same variance $\sigma^2$, has variance $\sigma^2/n$.

In our implementation, the value in a sample that is not in the maximum level of resolution is the mean of the values of four samples in the immediately superior level. This means that at a certain level $li$, the value of each sample is given by the mean of the values of $4^{(limax-li)}$ samples, where $limax$ is the highest level of resolution. Therefore, the variance calculated for $X$, in a level $li$ must be equal to $\sigma^2/4^{(limax-li)}$, where $\sigma$ is computed from the observed values.

To compensate the attenuation effect, the thresholds associated to the hypothesis test in a level $li$ must be proportional to the standard deviation corresponding to that level. Hence, it must be equal to $thresmax/2^{(limax-li)}$ where $thresmax$ is the threshold in the highest level of resolution.

This operation produces thresholds that make the photo-consistency criteria more rigid in the beginning of the process and more tolerant in the later stages.

### 3.3. Color attribution

In our approach, the color attributed to a photo-consistent cell is given by a weighted mean of the colors observed in its projection on the input images. This strategy may produce inadequate results when very few cameras observe some region of the scene with simple geometry but with highly variable texture. In this case such region will be classified as photo-consistent and will be colored in a way that does not adequately represent the variability observed in its projection in the images. Although this problem may happen in some cases, we believe that in practice this will rarely occur. This can be explained by the intrinsic properties of the adaptive method proposed.

The adaptive refinement strategy, combined with the use of segmentation information in increasing levels of resolution, favors the subdivision of cells near the surface of the scene to be reconstructed. In low and intermediary resolutions, the projection of the segmentation images in the planar regions associated to such cells contain values that are between the values that correspond exactly to foreground and background. This implies that these cells will be classified as undefined.

Moreover, the thresholds increase as the resolution is reduced, to compensate the reduction in the observed deviations due to the anti-aliasing operation. Therefore, the probability that a large heterogeneous regions be classified as photo-consistent is very low.

### 3.4. Algorithm

Algorithm 1 summarizes the entire procedure. For more implementation details see [7].

**Algorithm 1** Adaptive Space Carving

$node \leftarrow octree$
Initialize node.cell with the initial volume.
**repeat**
  $level \leftarrow 0$
  Clear visibility maps.
  **for all** sweep planes $\pi_k$, $k = 2^{level}..0$ **do**
    Determine the projected maps $PPSM$ and $PVNM$.
    **for all** octree node $nd$ such that $(nd.cell \cap \pi_k \neq \emptyset)$ and $(nd.cell.class$ =NOT_EVALUATED) **do**
      $pr \leftarrow nd.cell \cap \pi_k$
      $nd.cell.class \leftarrow$ PhotoConsistCheck($level$,$pr$,$PPSM$,$PVNM$)
      **if** $nd.cell.class =$ CONSISTENT **then**
        Attribute adequate colors to $nd.cell$.
      **else if** $nd.cell.class =$ UNDEFINED **then**
        Create eight child nodes of $nd$.
        Subdivide $nd.cell$ in eight new cells.
        Label each new cell as NOT_EVALUATED.
        Attribute the new cells to the children of $nd$.
      **end if**
    **end for**
    Update the visibility maps.
  **end for**
  $level \leftarrow level + 1$
**until** no cell has been subdivided or $level <$ MAX_LEVEL

## 4. Implemented system

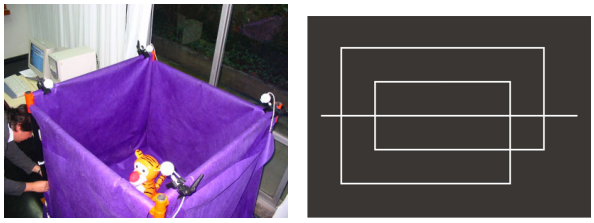
We implemented a simple image based 3D photography system to investigate the results obtained by the application of the proposed algorithm to real scenes. This system is composed by four low cost CMOS webcams placed at the corners of a cubic structure and connected to an *USB* hub (Figure 3(a)).

The calibration of the cameras was done using the method in [12] which is based on geometric model recognition . The pattern used in such calibration is shown in figure 3(b).

Object segmentation was done by comparing the input images with a statistic model of the crominance and rbg values of the background. To aid such process we covered the cubic structure with a purple cloth.

## 5. Results and analysis

The tests were executed in a 2.4 Ghz Intel Pentium IV with 1Gb of Ram memory equipped with a GeForce 4 graphic board. We used as data test a set of four 352x488 input images obtained from the system cameras. The im-



(a) System (b) Pattern

**Figure 3. System and Pattern**

ages show an small scale scene composed of a box covered with a textured cloth, a teddy bear and a plastic bird in four different points of view (Figure 4). The dimensions of the reconstruction space are 80x40x80 cm.

Figure 5 show images of the reconstructed model taken from new viewpoints. This reconstruction was produced in a resolution corresponding to a voxel array of 256x256x256. Although we used only four cameras the shape and texture of the objects were captured correctly.

The behavior of the adaptive algorithm in different levels of refinement is shown in figure 6. The first five levels are not shown because in these stages of refinement no photo-consistent regions of the scene were detected. Figure 6(a) shows level 5 where, as expected, only planar homogeneous regions are classified as photo-consistent. Figure 6(b) shows level 6 where more information is capture. In level 7, shown by figure 6(c) we have an almost complete low resolution version of the model, except for some small regions in the cloth. Finally, figure 6(d) shows an image of the reconstructed model in level 8 where the fine details of the cloth texture can be observed.

In figure 7 we compare the results produced by the adaptive space carving and a non-adaptive spacecarving that also registers information in scene space. We can observe that the results are practically the same but are obtained faster in adaptive space carving as shown in table 1.

| Volume | Adap. Space Carving | Space Carving |
|---|---|---|
| $64^3$ | 4s | 4s |
| $128^3$ | 8s | 10s |
| $256^3$ | 16s | 36s |

**Table 1. Reconstruction time**

## 6. Conclusions

In this work we presented an adaptive space carving algorithm that works with an octree representation of the scene reconstruction space. This method uses photometric and silhouette information as well as an estimate of the noise produced by the sensors of the cameras.

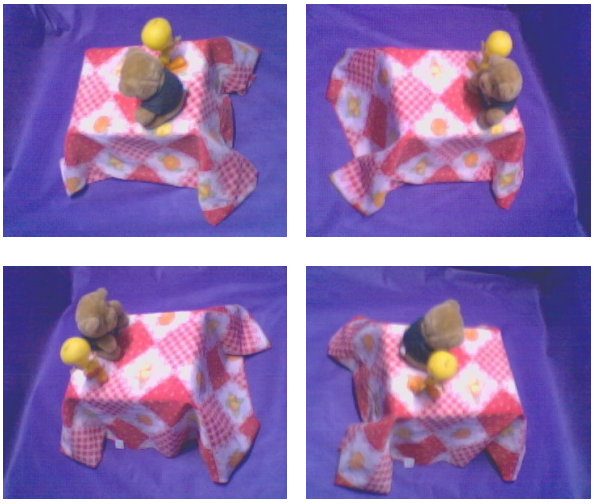


**Figure 4. Input Images**

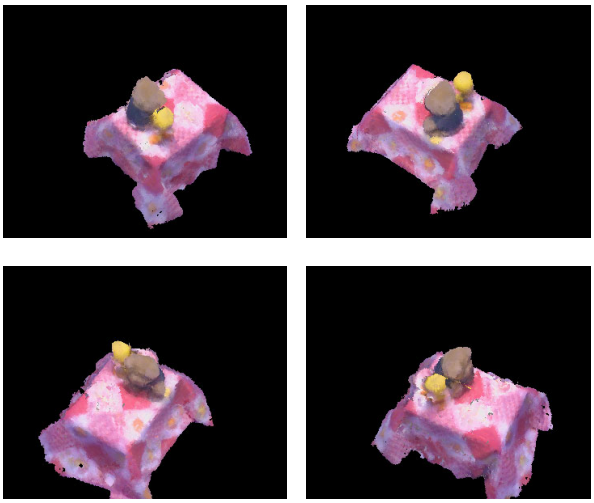


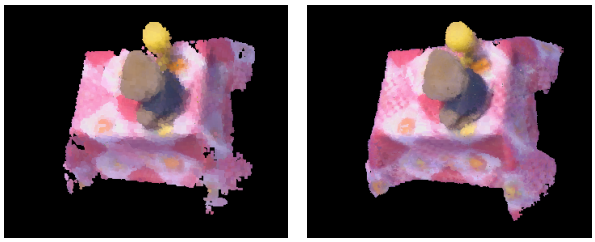
**Figure 5. Reconstructed scene**

Our method reconstructs a scene faster then original voxel-based space carving by combining adaptive carving and a photo-consistent test that uses constraint information in increasing levels of resolution.

The photo-consistent test proposed also contemplates the high levels of noise present in low cost cameras. Nevertheless, the results have shown that noise is not the only source of variation in the observed colors. Hence, it was necessary to use large levels of confidence so that no region of the scene would be lost.

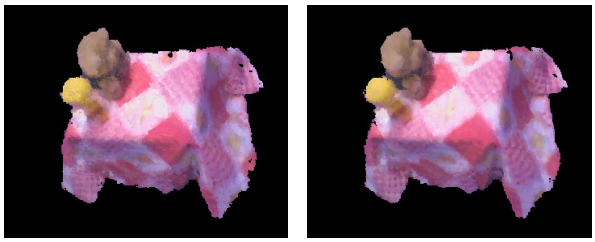
Topics that must be investigated in future work are the use of graphic hardware programming to accelerate the reconstruction process, a conservative strategy to remove large non-photo-consistent regions that do not project in the background regions of the images, and more sophisticate

(a) Level5 (b) Level6



(c) Level7 (d) Level8

**Figure 6. Images of different levels of refinement**



(a) Adaptive (b) Original

**Figure 7. Adaptive x Original Space Carving**

statistic measures that take into consideration sources of error other than camera noise.

## 7. Acknowledgements

This work was developed at TeCGraf/Puc-Rio and Visgraf/Impa, and was partially funded by CAPES and CNPq.

## References

[1] Rahul Bhotika, David J. Fleet and Kyriakos N. Kutulakos. A Probabilistic Theory of Occupancy and Emptiness. University of Rochester. Department of Computer Science. Technical Report 753. 2001

[2] A. Broadhurst and R. Cipolla. A statistical consistency check for the space carving algorithm. In M. Mirmehdi and B.Thomas, editors, Proc. 11th British Machine Vision Conference, volume I, pages 282-291, Bristol, September 2000.

[3] A. Broadhurst, T. Drummond and R. Cipolla. A probabilistic framework for the Space Carving Algorithm. In Proc. 8th International Conference of Computer Vision, pages 388-393, Vancouver, Canada, July 2001. IEEE Computer Society Press.

[4] W. B. Culbertson, T. Malzbender, and G. Slabaugh, Generalized Voxel Coloring, Proceedings of the ICCV Workshop, Vision Algorithms Theory and Practice, Springer-Verlag Lecture Notes in Computer Science 1883, September 1999, pp. 100-115.

[5] K. N. Kutulakos and S. M. Seitz, A Theory of Shape by Space Carving, International Journal of Computer Vision, Vol. 38, No. 3, July 2000, pp. 199-218.

[6] K. N. Kutulakos, Approximate N-View Stereo, Proceedings of the European Conference on Computer Vision, Springer Lecture Notes in Computer Science 1842, June/July 2000, Vol. 1, pp. 67-83.

[7] A. A. Montenegro, Reconstrução de Cenas a partir de Imagens através de Escultura do Espaço por Refinamento Adaptativo. Phd Thesis. Pontifícia Universidade Católica do Rio de Janeiro, September 2003.

[8] A. Prock and C. Dyer, Towards Real-Time Voxel Coloring, Proceedings of the DARPA Image Understanding Workshop, 1998, pp. 315-321.

[9] M. Sainz , N. Bagherzadeh and A. Susin, Hardware Accelerated Voxel Carving, Proceedings of the 1st Ibero-American Symposium on Computer Graphics, July 1-5, 2002, pp. 289-297, Guimarães, Portugal.

[10] S. Seitz and C. Dyer, Photorealistic Scene Reconstruction by Voxel Coloring, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 1997, pp. 1067- 1073.

[11] R. Szeliski, Rapid Octree Construction from Image Sequences, Computer Vision, Graphics and Image Processing: Image Understanding, Vol. 58, No. 1, July 1993, pp. 23-32.

[12] F. Szenberg, P. C. P. Carvalho, M. Gattass, Automatic Camera Calibration for Image Sequences of a Football Match, ICAPR 2001, March 11-14, 2001, p. 301.

[13] Vladimir Kolmogorov and Ramin Zabih. Multi-Camera Scene Reconstruction via Graph Cuts. In European Conference on Computer Vision, May 2002.

[14] http://www.opengl.org