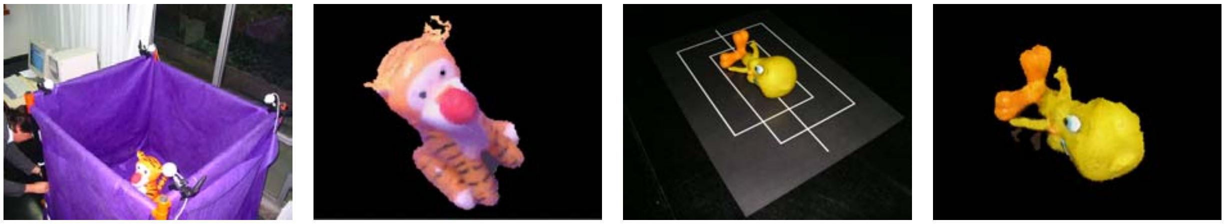
Towards realtime space carving with graphics hardware

Anselmo A. Montenegro * IMPA

Luiz Velho † IMPA

Paulo C. P. Carvalho ‡ IMPA

Marcelo Gattasss § Puc-Rio



(a) fixed cameras setup (b) result 1 (c) hand-held setup (d) result 2

Figure 1: Adaptive space carving systems with some results

The problem of 3D object reconstruction is one of the most investigated topics in computer graphics and vision. Because of its relevance, several different approaches have been proposed, ranging from laser scanning techniques, to the image-based methods. The latter, in particular, commence to be considered one of the most promising, as high quality digital cameras are becoming a commodity hardware.

Among the image-based methods for object reconstruction, those based on volumetric carving are very popular. Basically, these methods can be grouped in two distinct classes: the silhouette-based methods, also known as *visual-hull* oriented methods, and the photo-consistency based methods, usually known as *space carving* or *voxel coloring*. Visual-hull oriented methods have been successfully used in the implementation of realtime systems, as for example, the MIT system of Matusic. On the other hand, the development of such systems based on photo-consistency is still a challenge that must be overcome.

The main problem of using space carving methods in realtime reconstruction is that such methods are based on the classification of thousands of discrete elements (voxels) in scene space according to photo-consistency within scene images. Such classification process involves two main subprocess: the determination of image pixels corresponding to the re-projection of the voxels and their classification based on statistic measures computed on the set of corresponding pixels. As many elements must be evaluated, efficient strategies have to be used for realtime processing.

The determination of the pixels corresponding to a given voxel by re-projection is not a satisfactory solution. It yields several aliasing problems and is also extremely computationally intensive. A better solution is to register all images in planes embedded in scene space so that all voxels in the corresponding layer can be evaluated in the same iteration. This can be done efficiently by appropriately projecting the input images onto such planes as texture maps, an operation that is hardware accelerated in all modern graphics boards.

After the registration process is concluded, the registered information is taken to main memory so that the statistic measures, necessary for photo-consistency evaluation, can be finally computed. Unfortunately this process introduces significant overhead to the overall process, as the communication between the graphics card and main memory is still very slow in the current hardware. Because such overheads make realtime processing unfeasible, we must avoid these memory movements at any cost. Fragment programming is the ideal solution. Statistic measures associated to a voxel can be computed cumulatively as the corresponding pixels in the registered images are rendered in the registration plane.

As the number of elements to be processed can be excessively high, all these hardware based optimizations may not be sufficient for realtime processing. In order to deal with the computational complexity of space carving, we propose an adaptive approach.

We start with a coarse representation of the reconstruction space which is defined by a bounding box of the scene associated to the cell at the root of the octree. Then, we try to classify this initial cell according to some photo-consistent test based on the registered information with compatible resolution. If we succeed to classify the cell, either as photo-consistent or as non-photo-consistent, then we can finish processing that cell and attribute an adequate color to it. If we cannot 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 using registered images in higher resolution. This process is repeated successively until all regions are classified or a maximum resolution is obtained.

Based on the aforementioned ideas, we implemented two 3D reconstruction systems: the first one consists of a fixed multi-camera setup based on low cost webcams and the other one requires only one hand-held camera. Both systems use GPU basic operations for image registration by projective texture mapping. Photo-consistency statistics computation in GPU was not implemented yet and is being investigated at this moment. The fixed-camera setup is a prototype for a closed environment in which the cameras can be previously calibrated and the background estimated (Figure 1(a)). The hand-held space carving uses a calibration process that relies on the placement of a special pattern into scene space (Figure 1(c)). Such method is not affected by partial occlusions of the pattern and can be done in realtime. Moreover, we also estimate the background for each input image from a set of images of the scene without the object to be reconstructed. Such model is obtained from a set of background images warped to the desired input image viewpoint.

The implementation of the systems described above yielded promising results (Figure 1 (b,d)). This has led us to believe that space-carving in realtime is feasible in the near future by combining adaptive space carving, using multiresolution photometric and silhouette information, with GPU hardware processing. A video demonstrating the system is at the following URL: `<http://w3.impa.br/~lvelho/outgoing/gp2/>`.

*e-mail: ansmont@superig.com.br

†e-mail:lvelho@impa.br

‡e-mail:pcezar@impa.br

§e-mail:mgattas@tecgraf.puc-rio.br