The Vacuum Buffer

Voicu Popescu, Anselmo Lastra
University of North Carolina at Chapel Hill

ABSTRACT

Image-based rendering (IBR) techniques have the potential of alleviating some of the bottlenecks of traditional geometry-based rendering such as modeling difficulty and prohibitive cost of photorealism. One of the most appealing IBR approaches uses as primitives images enhanced with per-pixel depth and creates new views by 3D warping (IBRW). Modeling a scene with depth images lets one automatically capture intricate details hard to model conventionally. Also rendering from such representations has the potential of being efficient since it seems that the number of samples that need to be warped is independent of the scene complexity and is just a fraction above the number of samples in the final image. However, selecting the subset of reference-image samples that need to be warped to generate the new view is a very difficult task.

We present a sample-selection method that uses the *vacuum buffer* algorithm. Like other techniques, our method proceeds by considering samples of reference images that were acquired from locations close to the current camera position. Unlike other techniques however, our method offers a conservative estimate on whether samples of visible surfaces were potentially missed and it also points to the scene locations where such surfaces might be. The vacuum buffer is essentially a generalized z-buffer and it measures what sub-volumes of the current view-frustum have been determined.

Another important difference is that our method uses the current view, which allows it to reduce the number of chosen samples more than other methods that offer a sample-selection solution to be used for several desired camera views. The tradeoff for using the current view is having to solve the sample selection at each frame. By exploiting the coherence in the reference images, groups of nearby samples become the actual primitive, which massively reduces the total cost. Since our method does not build an intermediate scene representation, it avoids the undesirable additional resampling and is also better suited for use in conjunction with IBRW hardware.

KEYWORDS: image-based modeling, image-based rendering.

1 INTRODUCTION

The illusive goal of interactive photorealism and the enormous difficulty of modeling complex scenes with polygons and material-and-light descriptions made 3D-computer-graphics researchers consider images as the rendering primitive. Images capture automatically intricate geometric detail and complex light / material interactions and there is hope that the quality of the photographs can be conveyed to the rendered images. Numerous image-based rendering (IBR) techniques were developed.

Chen [Chen95] proposes panoramas that even though they offer a correct 3D view only from one location, they are a simple way of modeling and rendering complex natural scenes. Other methods approximate the 3D transformations of scenes with simpler 2D transforms ([Seitz96], [Torborg96]). The Lumigraph and Lightfield ([Gortler96], [Levoy96]) approaches create a database of rays that is queried at rendering time. Unfortunately the databases grow to impractical sizes for complex scenes.

McMillan suggested enhancing the photographs with per pixel depth [McMillan95]. The depth, the pixel coordinates and the camera pose uniquely locate the color sample in 3D space so samples can be reprojected (warped) at will on new images. The IBR by warping method (IBRW) is essentially reusing the original rays, under the lambertian-reflectance-model assumption. The method is general and it has reasonable storage / memory bandwidth requirements.

The method also has the potential of being very efficient if only one can determine a small set of reference-image samples that suffice to reconstruct the current view. Ideally the number of samples needed would not depend on the scene complexity but only on the size of the output image. Finding such a set however is a challenging problem. When an input (reference) image is warped, surfaces that were not originally visible can become disoccluded due to motion parallax and gaps form in the warped image (disocclusion errors). The gaps need to be covered with samples from other reference images. Also the sampling of surfaces from reference to output (desired) image changes differently for every surface.

This paper presents a new technique for selecting reference-image samples based on the vacuum buffer algorithm. Before we present the method in detail, we discuss previous sample-selection techniques for IBRW.

1.1 Related work

One simple solution is to just warp all samples of reference images that were taken from locations nearby the desired camera position. Complicated scenes require dense sampling with reference images and this approach generates too many samples.

Layered Depth Images (LDIs) bring substantial improvement [Shade98]; they generalize the concept of a depth image by allowing for more than one sample along a ray. Consequently an LDI can store samples of surfaces that are hidden from the view of the LDI. As the view changes, the originally hidden samples become visible avoiding disocclusion errors. Since there are only few samples in the deeper layers, the total number of samples in an LDI is only marginally larger than the number of samples in an equivalent depth image. The LDIs are constructed as a preprocess by warping reference images to the view of the LDI and discarding samples that warp at the same location and at the same depth.

An important question is what reference images need to be combined in an LDI in order to completely eliminate disocclusion errors? One approach is to combine as many as possible regularly spaced reference images, hoping that all potentially visible surfaces are sampled in at least one of the reference images ([Shade98], [Popescu98]). Such an approach can evidently miss surfaces.

Another concern is that an LDI offers only one sampling rate for a particular surface[1], which has to be adapted to the desired-image sampling rate at rendering time. Chang addresses this

---

[1] Moreover it is a *resampling* of the surface, which always introduces additional errors.
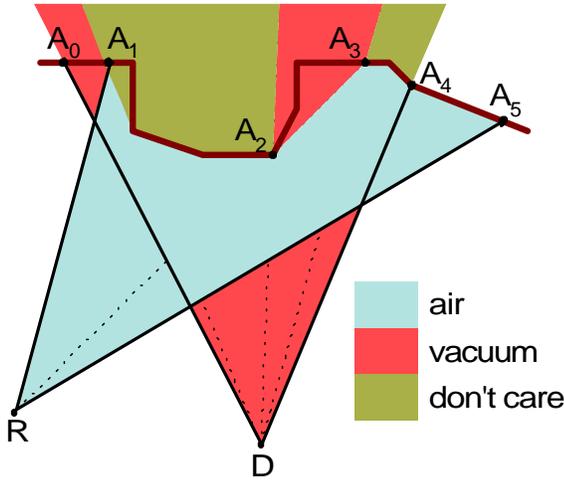
Figure 1. The reference image R is used to determine the volume of the view frustum of the desired image D

problem by using a tree of LDIs [Chang99] and choosing the appropriate resolution level according to the desired-image sampling requirement for each surface.

Also since LDIs are used to recreate several views, an LDI will inherently contain samples that are hidden for a particular view, and thus are unnecessarily warped.

Another important motivation in looking for a new sample-selection technique is our interest in designing and building hardware for accelerating IBRW ([Popescu00]). LDIs are complicated structures that cannot be easily warped in hardware.

Our sample-selection technique is based on the vacuum buffer algorithm that conservatively decides whether a set of reference-image samples is sufficient for a particular desired view. If not, the algorithm will indicate where in the scene surfaces might have been missed. The next section presents the vacuum buffer algorithm in detail and section 3 describes its use at choosing reference-image samples to adequately reconstruct the desired view.

2   THE VACUUM BUFFER ALGORITHM

The main idea behind the vacuum buffer algorithm is to determine the subvolumes of the desired view frustum that could contain visible surfaces that were missed by the current set of reference images. We call these undetermined subvolumes *vacuum*.

The vacuum buffer algorithm makes use of information contained in depth images that is usually ignored. The depth to the sample has been used to warp (reproject) the sample to the output image but it also tells us the distance to the *first surface* in the reference image. Consequently we know that there are no other occluders between the center of projection of the reference image and the surfaces sampled.

Figure 1 shows a reference image with center of projection R used to reconstruct the desired image with center of projection D. The scene is shown by the line $A_0A_1\ldots A_5$. The area (volume in 3D) shown in light gray (blue) is determined as being free of occluders by the reference image R. For simple referencing we call it *air*, since in most scenes it corresponds indeed to air. The

desired image D sees part of the air seen by the reference image R and that subvolume of the desired image is determined as empty.

The A_1A_2 segment of the scene is a connected opaque surface (occluder) and although other surfaces might be located behind it as seen from D, such surfaces cannot affect the desired image since they are hidden. The "shadow" that is cast in the desired-image view-frustum by occluders define a *don t care* subvolume. For the purpose at hand, don't care subvolumes are equivalent to air volumes.

The *vacuum* subvolumes shown in dark gray (red) could contain surfaces that are not sampled in the reference image R and are visible in the desired image D. Vacuum is not a guarantee that visible surfaces were missed. In figure 1 for example, the vacuum zone close to D might resolve to air when an appropriate reference image that encompasses it is used. However vacuum zones *might* contain visible surfaces and one has to resolve them.

Considering the case presented in figure 1, one could naively consider that the problem of missing samples could be detected by searching the frame buffer for uninstantiated pixels. Indeed the framebuffer will contain a gap between the new positions of A_2 and A_3 but missing samples can occur even when the framebuffer is instantiated as seen in the simple case presented in figure 2.

The next subsection presents the algorithm in more details.

2.1 Algorithm Overview

Given a set of reference images and a desired view, the algorithm computes the amount and location of vacuum that remains after all reference images are used. Initially the entire view frustum of the desired image is undetermined thus filled with vacuum. The air and the occluders of each reference image are used to resolve vacuum, by intersecting the vacuum with the air and with the shadows of the occluders. The volume intersections are computed efficiently using a generalized z-
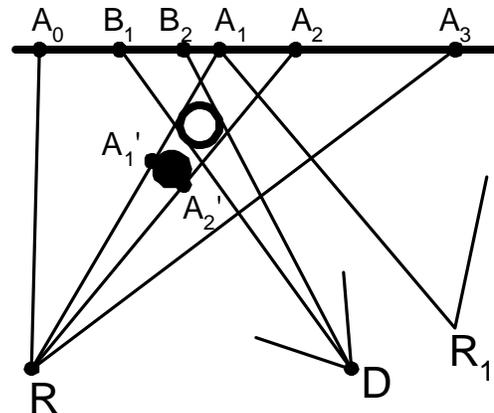

Figure 2. The reference image R samples the surfaces $A_0A_1A_1'A_2'A_2A_3$. It does not sample the hollow sphere. When used to reconstruct the desired image D, the hollow sphere projects to B_1B_2 and the corresponding pixels are already instantiated with a fragment of A_0A_1. Choosing another reference image to fill the gap between A_1 and A_2 might not suffice to reveal the missing object as is the case of reference image R_1.
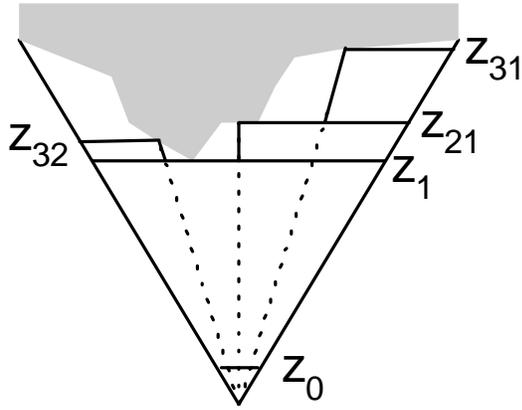
Figure 3. 2D view of the quadtree subdivision of the reference image frustum. Only three levels are shown. At each subdivision one of the children will have the same closest z as its parent so three new frusta are created with each subdivision (one in the 2D view shown).

buffer, which we call the *vacuum buffer*. The vacuum buffer is similar to the z-buffer but it stores z intervals, or *spans*, along each ray, whereas the classic z buffer stores a single value[2]. The list of z spans at one vacuum buffer location corresponds to the vacuum remaining in the view frustum along that particular ray.

The algorithm first processes the air of the current reference image. As a preprocess, the reference image is recursively subdivided in quadtree fashion and the closest z values are precomputed for each subregion. This subdivides the reference image frustum in subfrusta as seen in figure 3. The first frustum is defined by the hither plane and the plane of closest z. The next level frusta are defined by the parent-region's closest z plane and each subregion's closest z plane.

The vacuum buffer algorithm processes the frusta recursively, starting from the root of the quadtree. The frustum composed of six quadrilateral faces is transformed, projected to the desired image plane and each of the six faces are scan-converted. Since the frustum is a convex polyhedron, a vacuum buffer location will be either hit twice or not hit at all. When a vacuum buffer location is hit by two samples of the faces of the air frustum, the list of intervals is updated by eliminating the vacuum according to the span of air between the two samples (see figure 4).

The next important question is when to stop the recursive subdivision of the reference image frustum. The lower the level of the recursion, the smaller the progress (amount of eliminated vacuum) while the cost increases exponentially. On the other hand, stopping the recursion early doesn't use all the information available in the reference image; some of the air information is wasted.

There is another consideration crucial for deciding what the smallest subregion (tile) of the reference image should be. Remember that the second step of the algorithm is to ignore all vacuum behind occluders. If the tiles are small enough they are, in general, part of the same occluder and one can ignore everything behind the desired image projection of the tile. However computing the exact desired image projection of the
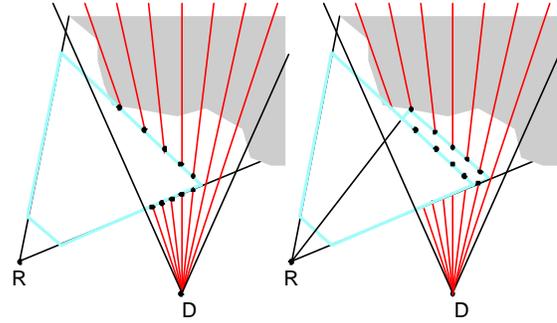
---

[2] The method of intersecting volumes using rasterization buffers was first used in constructive solid geometry ([vanHook86], [Salesin90] and others).

Figure 4. Initially, each vacuum buffer location contains the span (*hither, yon*) since nothing has been determined. The figure on the left shows the vacuum buffer after the first level frustum of the reference image R has been processed. Several locations contain two spans since part of the vacuum has been determined. The right figure shows the vacuum buffer after the second level frustum has been processed.

tiles is expensive and it is equivalent to warping the tile. This obviously defeats the purpose when the vacuum buffer algorithm is used to choose the tiles needed for the current frame. In order to compute the desired image projection of tiles efficiently we replace the height field corresponding to the tile with a single quad[3]. All vacuum behind the projection of the quad is eliminated (see figure 5).

If the tiles are too large then the approximation is too coarse and one can incorrectly eliminate vacuum where unsampled surfaces
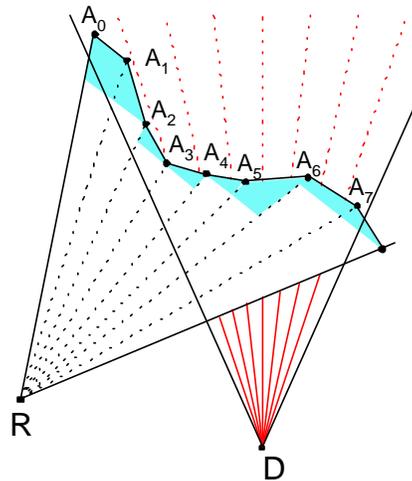

Figure 5. The reference image is split in eight tiles. The leafs of the quadtree are frusta defined by the closest-z plane of the previous level and the quads (here A_kA_{k+1} segments) that approximate the height field of the tile. The leaf frusta, shown in gray (blue), are processed similarly to the other frusta. The occluding faces of the frusta are used to eliminate all vacuum behind them (the dotted rays in D's frustum indicate the eliminated vacuum).

---

[3] For surface continuity one chooses the four corner samples that are connected by two triangles; note that in general the four samples are not coplanar but there is always a way of selecting the two triangles such that the resulting polyhedron is convex.

could potentially hide. We found that 16x16 sample tiles are adequate for a variety of test scenes. Moreover, 16x16 tiles are rendering primitives favorable to efficient IBRW architectures [Popescu00].

No matter how small the rectangular tiles are there will be some tiles stretch from one object to the other. Such silhouette tiles that are not patches of a single occluder cannot be assumed to be a continuous surface and the vacuum behind them cannot be eliminated.

We detect these tiles by detecting the depth discontinuities in the reference image [Popescu00]. Tiles with depth discontinuities are still useful for the first phase of the algorithm when the air of reference images is used to resolve vacuum from the vacuum buffer. At the second step, depth discontinuity tiles can be ignored relying on tiles from other reference images to sample the two occluders independently. However for small features like the light poles in Color Figure 2, it is very likely that no reference image will have tiles that map entirely on the light pole. Thus one should segment the tile according to the surfaces sampled (see Color Figure 2). The resulting tile segments will each model one object now and they can be treated as regular tiles. Note that tile segmentation does not depend on the desired view so it can be done as a preprocess. The next subsection summarizes the algorithm and presents implementation guidelines.

2.2 Algorithm Implementation

1. For each reference image
   1.1. Compute depth discontinuities
   1.2. Segment tiles
   1.3. Build quadtree of frusta
2. Initialize vacuum buffer
3. Initialize vacuum accounting tree (VAT)
4. Clear item buffer
5. For each reference image
   5.1. ProcessFrustum(F_0)
6. Done: VAT measures and locates the amount of vacuum left in the view frustum

The recursive ProcessFrustum(Frustum *F) routine is summarized next:

1. if (F == null) return
2. if F not leaf and F->closestZ is F->parent->closestZ
   2.1. go to 6
3. Transform, clip, and project frustum
4. Scan-convert faces in item buffer
   4.1. if item buffer location hit first time
      4.1.1.   UpdateZBuffer(z_0)
      4.1.2.   UpdateItemBuffer(F)
   4.2. if item buffer location hit second time
      4.2.1.   dv = UpdateVacuumBuffer(z_0, z_1)
      4.2.2.   UpdateVAT(dv)
   4.3. if current face is occluder
      4.3.1.   dv =UpdateVacuumBuffer(z_0, yon)
      4.3.2.   UpdateVAT(dv)
5. if F is leaf
   5.1. ProcessFrustum(F->next)
6. else
   6.1. for (i = 0; i < 4; i++)
      6.1.1.   ProcessFrustum(F->child[i])

The reader will recognize the algorithm described previously. The resolution at which vacuum is determined, in other words, the number of rays in the vacuum buffer, does not have to be the

resolution at which desired images will ultimately be produced. A lower vacuum-buffer resolution can be used (we used 320x240 for VGA resolution output). However, the resolution cannot be lowered indefinitely since the projection of the quads that approximate the tile height-field need to have meaningful sizes and shapes.

The VAT is a fast way of knowing how much vacuum is left in the desired view frustum and where it is located. It is a quadtree subdivision of a buffer of vacuum buffer resolution. A leaf node stores for every vacuum buffer location the sum of the vacuum spans in that linked list. In other words a leaf stores the amount of vacuum along a certain ray of the vacuum buffer. A higher level node stores the sum of the vacuum stored at its four children. The root stores the amount of vacuum left in the entire vacuum buffer. The VAT is initialized to full once per desired view: vacuum from hither to yon along all rays in the vacuum buffer. It is then updated as frusta are processed, using the amount of vacuum determined by the current air-faces sample pair or current occluder sample.

The item buffer is also at the vacuum buffer resolution and it stores unique frusta identifiers. It is used to detect second hits of samples from the faces of the same frustum. Since the frusta identifiers are unique, the item buffer does not need to be cleared from one frustum to the next. The z-buffer, also at the vacuum buffer resolution, is used to record the z of the first hit. When the second hit occurs, the value in the z-buffer and the new z value are used to update the vacuum buffer. Updating the vacuum buffer is equivalent to subtracting the new determined interval from the vacuum intervals stored in the list at that location. The amount of vacuum determined is returned and used to update the VAT. The z-buffer does not need to be cleared since it is always used in conjunction with the item buffer.

The tiles that are segmented generate two or more frusta at the leaf level of the reference image quadtree. In this case the the additional frusta need also to be processed (ProcessFrustum(Frustum *F), step 5.1. and 1.). The frusta are processed recursively. If the current frustum has the same closest z as its parent, it doesn't need to be processed since no progress can be made (ProcessFrustum(Frustum *F), step 2).

A case that needs to be treated carefully is the case of frusta clipped by the hither plane. New faces need to be introduced to make sure that the faces that are scan-converted form indeed a convex polyhedron; scan-converting a frustum without the top base incorrectly misses second hits since one can "see inside" the frustum.

One might wonder what happens when only one hit occurs or several hits occur, due to potential scan-conversion precision limitations. If only one hit occurs, typical for the desired-view silhouette of the frustum, there will be no air span generated to update the vacuum. This is the correct degenerate-case behavior. If more than two hits occur, typical for edges of the frustum that project over another face, the vacuum buffer is updated using tiny air spans. The results are correct and the sole penalty is in efficiency. We avoid such cases by setting a threshold below which air spans are discarded.

From figure 4 one can see that the typical vacuum buffer update is done with adjacent air spans. To insure the adjacency in the presence of numerical error, the vacuum-buffer update routine starts by increasing the air span with epsilon at each end.

Adjacency is of course important since it keeps the vacuum-span lists short, which translates into efficient update times.

The next section presents the application of the vacuum buffer algorithm to reference-image sample choosing, called tile choosing since tiles are the level at which samples are selected.

3 Tile choosing

In the introduction we discussed the requirements for a sample-selection method. Here they are again, phrased more concisely. An ideal set of samples satisfies the following conditions:

- *completeness*: all surfaces visible in the desired view should be represented

- *good quality*: the resolution at which the surfaces are sampled in the reference images should be as close as possible to the resolution at which the surfaces are sampled in the desired view; *oversampling* leads to aliasing artifacts and high cost and *undersampling* leads to poor quality (blurriness)

- *non-redundancy*: invariably some surfaces are sampled in more than one reference image; assuming that the surfaces are close to diffuse, the multiple samples are equivalent and one can / should discard the redundant copies[4].

- *low depth complexity*: for efficiency, the set should contain very few or no samples that are hidden in the desired image.

For the reasons discussed previously we split the reference images in tiles and the set of samples that has to be determined is actually a set of tiles. It is obvious that the vacuum buffer measures the completeness of a set of tiles but more needs to be said in order to explain how it can help with satisfying the other three conditions.

We measure the quality of the samples in a tile by analyzing how much it stretches or shrinks when projected in the desired image. Small changes indicate similarity between the sampling rates in the reference and output images thus the samples of such a tile are preferred. From an implementation point of view, the quality is efficiently derived from the size of the bounding box of the projection of the quad corresponding to the tile. Having defined a quality metric we need to explain next how to determine whether two samples originating from different reference images sample the same surface.

In order to do this we utilize an additional z-buffer and an additional item buffer at the resolution of the vacuum buffer that simply store at each location the z of the closest occluder and a pointer to the tile it originates from. When an occluding face is processed, the occluder z-buffer is consulted. If the current occluder sample is within epsilon of the closest occluder z, it is assumed to sample the same surface. If the quality of the current tile is better, the occluder item buffer location is overwritten, otherwise the current sample is simply discarded. This eliminates redundant samples of lesser quality.

---

[4] Even in the case of specular surfaces, ensuring that each surface is sampled only by one reference image helps avoiding inconsistencies that produce disturbing speckling and frame to frame flickering. The shading is of course not correct but the static highlights are by far more esthetically pleasing than perfectly diffuse surfaces.

If the current occluder sample is clearly behind the closest occluder sample, the current sample is again discarded. This reduces the depth complexity.

The tiles that could not be segmented in order to avoid internal depth discontinuities cannot be processed by the algorithm, and, conservatively, have to be chosen.

The tile choosing algorithm starts with the reference images that were acquired from a location closest to the desired camera location. More and more distant reference images are processed until the amount of vacuum remaining is below a certain threshold. The chosen tiles are the tiles that have at least one sample present in the occluder item buffer: they were not completely occluded nor were they completely replaced by better tiles. In order to avoid scanning through the occluder item buffer, we maintain presence counters for each tile.

Color Figure 1 illustrates tile choosing using the vacuum buffer algorithm.

3.1 Results

We tested the tile choosing technique described on a complex model of a town (seen in the accompanying videotape). The reference images used as input were rendered using the 3D Studio Max modeling package. The reference images were rendered from locations that form a regular 3D grid; the grid subdivides the viewing volume in equal cells (boxes).

Our first version of the tile-choosing algorithm presented with the WarpEngine architecture [Popescu00] did not use the vacuum buffer. It just considered the 8 sampling locations defining the cell of the current camera position and rendered the tiles in the occluder item buffer and the occluder z-buffer. In order to minimize the chances of missing a visible surface the cell size was small (3mx3mx3m for eurotown). Also no tile segmentation was attempted.

With the vacuum buffer we were able to double the sides of the cell. This equates to 8 times fewer reference-images, which is obviously of great importance when real world data is acquired. Like in the previous case the tile choosing starts by considering the images of the current cell. Tile choosing stops when the total amount of vacuum decreases below a threshold. We used $1/z$ for the vacuum buffer and the initial vacuum span was $(100.0, 0)$. Using $1/z$ is convenient since it gives more importance to vacuum spans that are close to the camera. Objects that are close have a large screen area and the artifact resulting from missing them is more noticeable. In our particular case, the threshold below which no more disocclusion errors were noticeable was 1500 (the average per location amount of vacuum is about 0.02). Increasing the size of the cell means that occasionally one had to consider images from neighboring cells since the vacuum would not decrease below the set threshold only by using the reference images of the current cell.

Tile segmentation was done allowing a maximum of 6 segments per tile. If a tile could not be successfully segmented it was conservatively chosen. The following table presents the results of tile choosing for both the old and the current method with output at VGA resolution.

	old	current
Visible tiles	15K	13K
Selected tiles & tile segments	7K	3.3K

Selected unsegmented tiles	2.5K	1.8K
Selected tiles that are segmented	0	1.1K
Selected tile segments	0	1.5K
Unsegmentable tiles	4.5K	0.4K

First, there are fewer visible tiles since the reference images are further apart. The number of selected tiles (including segments) is much lower mainly because the number of unsegmentable tiles decreased substantially. The number of selected tiles that did not have to be segmented (no depth discontinuities) also decreased since the tile segments eliminated some of the full tiles.

In a VGA resolution image there are 1.2K 16x16 tiles. So the ratio between the number of reference-image samples selected and the number of output image samples is 2.75. We counted 256 samples per tile segments since in the case of the WarpEngine architecture the samples of a tile are processed in SIMD fashion and one cannot save if some samples are "off". Other factors that make this figure deviate from the ideal value of 1.0 are:

- The closest reference-image sampling rate for a surface is not *exactly* the desired image sampling rate

- There still are a few tiles that are not segmented, which cause undetected redundancy.

The next section analyzes the complexity of the algorithm and investigates possible hardware acceleration.

4 HARDWARE ACCELERATION FOR VACUUM BUFFER ALGORITHM

As stated earlier, the resolution of the buffers used in the algorithm can be lower than the resolution of the final image. Thus clearing the buffers (once per frame) is not a substantial burden. The bulk of the work is done at step 4 of the `ProcessFrustum` routine.

If at most eight sampling locations are considered at each frame, and if cube depth-panoramas are used centered at the sampling location, there could be as many as 6 x 8 = 48 reference images that need to be considered.

Of course not all reference images have samples that project in the view frustum. At step 3, if the current sub-region of the reference image is completely outside the view frustum it is culled and the recursion is stopped early. In order to determine whether the sub-region is needed, one transforms, clips and projects a second bottom face of the frustum, defined by the *farthest* z. If the larger frustum defined by the parent's closest z plane and the farthest z is outside of the view frustum, the subregion is not useful for the current view and can be safely ignored (see figure 6). In our test scenes, using 1k x 1k reference images[5] on average 13K 16x16 tiles passed the view-frustum-culling test. This is equivalent to 3-4 full reference images.

The number of frusta that have to be processed for each full reference image is:

$$N = F_{1024} + 3F_{512} + 4(3F_{256} + 4(...+4(3F_{32} + 4F_{16}))) = 5120,$$
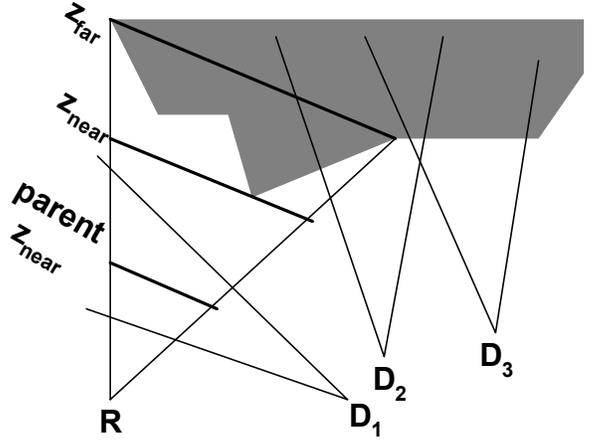

Figure 6. Only the current subregion of the reference image R is shown. Out of the possible desired view frusta D_1, D_2, and D_3, the subregion can be ignored only for D_3. The air information of the subregion is pertinent to D_1, which wouldn't have been detected if we used the frustum defined by the z_{near} and z_{far} planes. If we used the *parent* z_{near}- z_{near} frustum, as required by processing the subregion's air information, the subregion would have incorrectly been ignored for D_2. The only conservative approach is to use the *parent* z_{near}- z_{far} frustum.

where F_w is a frustum that corresponds to a w x w subregion. Analytically, the number of frusta can be computed with the formula

$$N = 4^k + 4^{k-1} = 5120$$

where k is defined as

$$k = \log_2 w_{image} - \log_2 w_{tile} = \log_2 1024 - \log_2 16 = 6$$

Each frustum has 6 quadrilateral faces, thus, the number of triangles that need to be rendered per second, assuming 30Hz update rate is[6]:

$$T = 30 * 4 * 6 * 2 * 5120 \approx 7 Mtris/s$$

Our software implementation takes about 20 seconds per frame, which is orders of magnitude too slow for interactive rendering. The average number of triangles rendered per frame for the eurotown scene was *263Ktris,* which represents *7.8Mtris/s*, confirming the estimate above.

The obvious way to accelerate the vacuum buffer is to use polygon-rendering hardware. Besides the fact that *7Mtris/s* is a sizeable task even for the most recent polygon-rendering hardware, there is the problem of updating the vacuum buffer.

Graphics hardware doesn't offer the possibility of storing several z values at each location. Having the CPU read back the z-buffer and item buffer after each face of a frusta is rendered is not a practical solution. First, the operation is very costly on most graphics architectures. Second, the CPU would have to traverse the bounding box of the projection of the face in order to use the z's of the current face, which most likely offsets the advantage of hardware rendering.

---

[5] The reference images had 90 degrees field of view and the output view frustum was 65 degrees horizontally.

[6] The tiles belong to more than four reference images so there are more higher level frusta than in the ideal case of full images.

We believe that the most promising solution is to design hardware for accelerating the vacuum buffer. The architecture would be similar to current triangle rendering architectures. However, the rendering task consists only of flat shaded triangles, so the architecture will be simpler in many ways.

The only additional complexity is a z-buffer that can store several samples at each location. Because of the incremental updates to the vacuum buffer, the number of z's that need to be stored is small, and it seems practical to provide storage for a fixed number of z values. The next table shows for our scene, how many times per frame, on average, a list grew longer than a certain value. For example on average, at each frame, there were 118 lists that grew from length 7 to length 8.

Length	4	5	6	7	8	9	10	11
Exceeded	2785	925	399	118	28	5	2	0

If the limit is exceeded, exceptions can be raised that discard the shortest interval or eliminate the smallest gap by merging nearly adjacent intervals (whichever is smaller). From our simulations, if the maximum hardware list length is set to 8, the amount of vacuum incorrectly eliminated to serve the exception was negligible[7] and the same set of tiles was chosen as in the case of unlimited list lengths. Since the resolution of the vacuum buffer is low, it is feasible to store a maximum of 8 spans at each location.

5 CONCLUSIONS

We presented a method of selecting reference samples to be warped to create the desired view. The method uses the vacuum buffer algorithm to conservatively estimate the subvolumes of the view frustum that have not been determined by the reference images considered so far. A software only implementation is too slow to be practical. The bulk of the work consists in rendering polygons and the number of polygons that need to be rendered per second is within the capabilities of today's hardware. The modification required is to extend the z-buffer to contain several spans of z. The number of spans at a vacuum buffer location is small. This enables a simple hardware implementation that supports vacuum-span lists of fixed maximum length.

6 FUTURE WORK

The current version considers the reference images of the current cell in the order defined by the increasing distance to the camera position. The residual amount of vacuum should decrease more rapidly if:

- First process the reference image that sees the desired view location and has the view direction the closest to the desired image view; this eliminates all the vacuum close to the desired view camera

- Use the VAT to localize the remaining vacuum and then select the reference images that see that vacuum

This reference image ordering might allow us to space the reference images even further apart. This brings us to what we believe is another important application of the vacuum buffer algorithm.

---

[7] The average was below .1 where the original vacuum span (hither, yon) is (100.0, 0)

6.1 Determining sampling locations

An obvious goal in computing the locations where the reference images should be placed is to acquire as few images as possible while minimizing the number of disocclusion errors. This is the well known *next best view* problem ([Connoly85], [Maver93] and many others).

We speculate on an iterative process based on the vacuum buffer algorithm that automatically produces a set of candidate sampling locations.

First, the scene is subdivided according to a regular 3D grid. A cell of this grid would be a cube of side 0.5m for a room or 20m for a town. Complete panoramas (6 faces of a cube) should be acquired at each grid node. Then, for each cell, the vacuum buffer algorithm is executed at locations that form a more refined grid, with nodes that are let's say .1 m apart. The reference images used are the reference images acquired at the corners of the current cell and the cells adjacent to the current cell. The node of the refined grid at which the most vacuum remained undetermined is a new sampling location. The residual vacuum at the refined grid locations is computed again. Again the location with the most undetermined vacuum becomes a new sampling location. The process stops when the budget of sampling locations / cell is reached or when the maximum residual vacuum value dips below a tolerable threshold.

The greedy approach described does not provide the optimal solution but it is fully automatic and the solution might be good enough.

7 ACKNOWLEDGEMENTS

REFERENCES

[Chang99] Shade J., Gortler S., He L., and Szeliski R., "LDI Tree: A Hierarchical Representation for Image-Based Rendering", *Proc. SIGGRAPH 99*, 291-298 (1999).

[Chen95] Chen S., "Quicktime VR - An Image-Based Approach to Virtual Environment Navigation", *Proc. SIGGRAPH 95*, 29-38 (1995).

[Connolly85] Connolly, "The detemination of next best views", *IEEE International Conference on Robotics and Automation*, pages 432-435, (1985).

[Gortler96] Gortler S., Grzeszczuk R., Szeliski R., and Cohen M., "The Lumigraph", *Proc. SIGGRAPH 96*, 43-54 (1996).

[Levoy96] Levoy M. and Hanrahan P., "Light Field Rendering", *Proc. SIGGRAPH 96*, 31-42 (1996).

[Maver93] Maver and Bajcsy, "Occlusions as a guide for planning the next view", *IEEE Transactions on Pattern Analysis and Mmachine Intelligence*, 15(5):417-433 (1993).

[McMillan95] McMillan L. and Bishop G., "Plenoptic Modeling: An Image-Based Rendering System", *Proc. SIGGRAPH 95*, 39-46 (1995).

[Popescu00] Popescu V., Eyles J., Lastra A., Steinhurst J., England N., and Nyland L., "The WarpEngine: An

Architecture for the Post-Polygonal Age", to appear *Proc. SIGGRAPH '00*, (2000). Note to the reviewers: this paper can be found online at http://www.cs.unc.edu/~popescu/warpengine.pdf

[Popescu98] Popescu V., Lastra A., Aliaga D., and Oliveira M., "Efficient Warping for Architectural Walkthroughs using Layered Depth Images", *Proc. VIS'98*, 211-216 (1998).

[Seitz96] Seitz S. and Dyer C., "View Morphing: Synthesizing 3D Metamorphoses Using Image Transforms", *Proc. SIGGRAPH 96*, 21-30 (1996).

[Shade98] Shade J., Gortler S., He L., and Szeliski R., "Layered Depth Images", *Proc. SIGGRAPH 98*, 231-242 (1998).

[Torborg96] Torborg J. and Kajiya J., "Talisman: Commodity Real-time 3D Graphics for the PC", *Proc. SIGGRAPH 96*, 353-364 (1996).

[vanHook] van Hook T., "Real-Time Shaded NC Milling Display", *Proc. SIGGRAPH 86*, 15-20 (1986).

[Salesin90] Salesin D., "Rendering CSG Models with a ZZ-Buffer", *Proc. SIGGRAPH 90*, 67-76 (1990).