Hardware-accelerated ambient occlusion computation

Mirko Sattler, Ralf Sarlette, Gabriel Zachmann, Reinhard Klein

University of Bonn
Institute for Computer Science II, Bonn, Germany
Email: {sattler,sarlette,zach,rk}@cs.uni-bonn.de

Abstract

In this paper, we present a novel, hardware-accelerated approach to compute the visibility between surface points and directional light sources. Thus, our method provides a first-order approximation of the rendering equation in graphics hardware. This is done by accumulating depth tests of vertex fragments as seen from a number of light directions. Our method does not need any preprocessing of the scene elements and introduces no memory overhead. Besides of the handling of large polygonal models, it is suitable for deformable or animated objects under time-varying high-dynamic range illumination at interactive frame rates.

1 Introduction

Shadows are one of the most important visual clues about the spatial structure of an object. E.g. for virtual reality applications, such as life-sized cloth visualization, or medical surgery planning, they are important because they increase the presence of the virtual objects and the overall realism. However, the shadows need to be computed at interactive frame rates, otherwise usability and presence will break down. While local illumination models are the strength of modern graphics hardware, more advanced techniques, which include e.g. soft shadows and ambient occlusion, are notoriously hard to perform efficiently. On the other hand, recent methods like pre-computed radiance transfer [1] enable complex illumination effects at fast frame-rates, but are limited to static objects or pre-defined animations due to their long precomputation times.

In this paper, we present a method to efficiently compute the visibility for many light directions for each vertex on the GPU at interactive frame rates, even for large models. This includes self-occlusion as well as occlusion caused by other objects. This is done by using hardware occlusion query results from vertex fragments as seen from a number of light directions. All geometry in the scene can deform and move, and the illumination can change at no extra cost. Thus, we can compute and render a first-order approximation of the rendering equation [2] for opaque, polygonal objects on the graphics hardware.

Our algorithm features the following advantages:

- no precomputation, no complex data structures or special preprocessing is needed
- handles arbitrary deforming geometry
- fast hardware-accelerated occlusion calculation.
- high-dynamic range image based illumination
- easy to implement.
2 Previous work

A vast amount of work has been done on shadow algorithms. Shadow maps [3] or shadow volumes [4] and all derivates are the classical approaches for point-like light sources. A good overview of algorithms which produce soft shadows can be found in Hasenfratz et al. [5]. All these algorithms cannot efficiently be used for arbitrary lighting environments. On the other hand, environment mapping as introduced by Blinn et al. [6] is able to render reflections of incident lighting, but without shadows. Ray tracing is capable of handling globally illuminated scenes, but is naturally limited to the current camera position. Interactive rates are only achieved in a massive parallel environment with optimal acceleration structures [7, 8], which take several seconds to build. Other theoretical work [9] on GPU-based raytracing is not yet available in hardware. Very recently there has also been approaches to solve radiosity on graphics hardware [10], with interactive rates for small scenes.

To evaluate the illumination received by a point on the surface, there exist mainly two approaches. The first category gathers the incoming radiance at each surface point or vertex and scales with that count. The hemisphere defined by the vertex normal is sampled in different ways. Either rays are shot into predefined directions using standard ray tracing methods, or hemicube [11] sides are rendered using the standard pipeline [12]. With ray casting, intensive intersection tests have to be calculated and acceleration structures have to be maintained (e.g. space partitioning), whereas hemicubes cannot yet be evaluated efficiently on the graphics hardware. Other methods compute visibility cones [13], blocker maps [14], obscurance maps [15, 16], visibility maps [17] or radiance transfer [1] which also includes inter-reflections. All these methods require certain amounts of pre-computation time, and are therefore not suitable for dynamic objects. Very recent work on the topic discussed in this paper was presented by Kautz et. al [18]. It is mainly based on fast hemicube rasterization in order to detect blocker triangles. Downsampling of the visibility mask and a coarser blocker mesh are used for speed up. Interactive frame rates are achieved for small animated models. In contrast to their work, our method does not need a mesh hierarchy nor any additional graphics memory during run-time.

The second category is based on the approximation of the ambient environment by point or directional light sources, which amounts to reversing the first approach from “inside-out” to “outside-in”. That is, the visibility computation is originated at the light sources. Lately, NVIDIA proposed a hardware-accelerated 2-pass method, using accumulated shadow maps [19, 20], which is also used in many shaders in commercial rendering software packages [21]. To minimize sampling artifacts, jittering of the depth maps was introduced. This approach involves common shadow mapping projection problems [22]. To achieve usable results, several seconds per frame are needed. Changing to a new viewpoint requires new shadow mapping passes or an unwrapping process to obtain an occlusion texture. Superimposing images was also done by Keller et al. [23] to compute instant radiosity. This also requires space-partitioning structures.

3 Ambient occlusion calculation

In this section, we will present the core of our new method. In the following, a triangle mesh with vertex normals is given. To compute the outgoing radiance $L_r$ at a surface point $x$ into direction $v$ we have to compute:

$$L_r(x,v) = \int_{\Omega} f_r(x,v,l) L_i(x,l) V(x,l)(n_x \cdot l)dl$$

where $\Omega$ is the hemisphere domain over $x$, $f_r$ the BRDF, $L_i$ the incident radiance from direction $l$, $V(x,l)$ the visibility from $x$ to direction $l$ and $n_x$ the vertex normal at $x$. We discretize the integral by $k$ light directions $l_j$, $j \in [1, \ldots, k]$, which leads to

$$L_r(x,v) \approx \sum_{j=1}^{k} f_r(x,v,l_j) L_i(x,l_j) V(x,l_j)(n_x \cdot l_j)$$

We now concentrate on the efficient computation of the term $L_i(x,l_j) V(x,l_j)(n_x \cdot l_j)$.

3.1 Overview

Our method relies on depth test results from renderings of the desired object. We approach the problem of light direction visibility for each vertex by considering a set of $k$ directional light sources $l_j$ and determine the visibility of all $N$ vertices at
once as seen from each of the light source directions. Because the number of light source directions \( k \) is much smaller than the number of vertices \( N \) for large models, the "outside-in" approach is much more efficient, than the "inside-out" approach, that is, a much smaller number of render passes, \( k \) instead of \( N \), of the object into the depth buffer is needed.

More precisely, we render all geometry as seen from a light source direction into the depth buffer. Then, all vertices are rendered again as a point set. An individual occlusion query per vertex allows us to retrieve those vertices that passed the depth test. For these, the currently considered light source direction is marked as visible and stored in a matrix \( M \), which we call visibility matrix. We repeat this process for each light source, updating the appropriate entries in \( M \).

### 3.2 Visibility matrix computation

Our method makes heavy use of the OpenGL extension \texttt{ARB\_OCCLUSION\_QUERY} [24]. The purpose of this extension is to deliver the number of fragments that passed both depth and stencil test. In contrast to its predecessor \texttt{HP\_OCCLUSION\_QUERY} it is asynchronous, i.e. it does not use a "stop-and-wait" execution model for using multiple queries. This allows applications to issue many occlusion queries before asking for the result of any one. As mentioned above, in the first pass, the unlit scene is rendered into the depth buffer from one of the light source directions. We do this in orthographic projection mode. In the second pass, all vertices are rendered as \texttt{glPoints} with size 1, without updating the depth buffer. An offset (\texttt{glPolygonOffset}) with default values \((1.0, 1.0)\) is used to avoid rounding issues. Each single vertex \( i \in N \) is handled by an individual occlusion query and the visibility matrix \( M \), which stores visibility information for each vertex \( i \) to the light direction \( l_j \), is updated for all \( j \). For later performance, we do not store a boolean visibility bit, but instead the dot product computed from the vertex normal \( n_i \) and the vector defined by the light source direction \( l_j \) if the vertex is visible from that direction. \( l_j \) is computed once for each virtual light source direction. Therefore we obtain:

\[
M_{ij} = \begin{cases} 
  n_i \cdot l_j : & \text{vertex visible} \\
  0 : & \text{vertex invisible} 
\end{cases}
\]

#### Figure 2: Outline of the core algorithm for visibility matrix calculation.

#### 3.3 Rendering

After \( M_{ij} \) has been computed, the final color \( c_i \) for each vertex \( i \) can be computed as:

\[
c_i = \sum_{j=1}^{k} M_{ij} I_j
\]

where \( I_j \) is the 3-component (RGB) color of the light coming from direction \( l_j \). Figure 3 shows a simplified data flow of our approach. Note, that neither \( M_{ij} \) nor \( I_j \) change, if the viewpoint is changed.

#### Figure 3: Simplified data flow of our approach.
3.4 Creating the lightsphere

To approximate ambient occlusion with single directional light sources, we have to distribute \( k \) light directions. Each light direction is represented by a point on the unit sphere. We seek a distribution of points on the sphere satisfying the following conditions. The visibility computation should be done only once and the illumination environment should be changeable without doing new queries (see Section 5). Furthermore, to allow an easy increase of the number of light directions, already computed parts of \( M \) should be re-usable. Therefore, the distribution should equally sample the environment in all directions. Recent approaches like [25, 26] which do an efficient sampling of the environment map are not easily adaptable for our method. They need seconds to minutes of preprocessing time to reduce the number of light directions, therefore would not allow interactive change of the lighting environment. For equal distribution of points on a sphere, several methods exist [27–29]. We propose the following preprocessing procedure, based on subdivision of a regular solid. We start with the vertices of a unit octahedron, i.e. \( k = 6 \) light directions at subdivision level \( s = 0 \). To generate level \( s + 1 \), we do midpoint subdivision of the edges on level \( s \) and project the new vertices on the unit sphere, thus creating a polyhedron with \( 2 \cdot 4^{s+1} \) faces and \( 2 + 4^{s+1} \) vertices. This structure allows us to add new sets of well distributed points, while using the occlusion queries of all coarser subdivision levels. Figure 5 shows several increasing configurations.

4 Optimizations

Whenever the object is moved or deformed, that is when vertex positions change, we need a complete re-computation of \( M \). In the following, we will present several methods that significantly reduce that effort, so that we can maintain interactive frame rates even under these circumstances. Even for static geometry, depending on the viewpoint, not all vertices must be evaluated.

4.1 Vertex filtering using temporal coherence

The first optimization exploits temporal coherence by observing that during a viewpoint change only a small number of polygons become visible for the first time (at most those that cross the object silhouette). Consequently, we can compute the visibility matrix \( M \) lazily, thereby distributing the computation effort over several frames. More precisely, we maintain two lists of triangles: a list of “unseen” triangles, \( U \), containing all triangles that have not yet been visible and a list \( T \) (“todo”), containing triangles that will be seen in the next frame for the first time. \( U \) is initialized with all vertices once. Vertices which belong to triangles stored in \( U \) have not yet been visible from the camera so far. Thus, we do not need their occlusion with respect to the light directions. Vertices which belong to triangles which are stored in \( T \) have to be processed for the next frame, because these triangles will then be visible. Using the two lists, in each frame we need to determine the occlusion only for a small fraction of vertices (see also Figure 3). In order to compute list \( T \), we perform an algorithm similar to the one in Section 3, except that here we render triangles in the second pass and use perspective projection. More precisely, in the first pass, we render the unlit mesh into the depth buffer as seen from the new camera position. In the second pass, we render only the triangles still in list \( U \), each with its own occlusion query (again with offset and without depth buffer update). Obviously, when \( U \) contains less triangles than a certain threshold, we do not gain any performance any more. In that case, we just add all triangles from \( U \) to \( T \) and skip this optimization in the remaining frames. Note that for this optimization we need to consider triangles, not vertices. Otherwise, artifacts could occur, because a triangle might be visible although one of its vertices is not, and thus its corresponding value in \( M \) has not yet been computed. This is illustrated in Figure 4. In Section 7, we discuss the performance gains introduced by this optimization.
4.2 Changing the Lightsphere configuration

A further optimization is to dynamically change the lightsphere configuration. While the scene is animated or objects are deformed, we can decrease the number of light directions. In idle time, in order to converge to the exact solution, we increase that count as shown in the upper row in Figure 5. Note, that if the number $k$ is too small, under-sampling artifacts (left bunny) can occur depending on the used environment. In our experience a level of $s = 3$ and often even level $s = 2$ produces reasonable results. The algorithm allows user control through the choice of a desired frame rate or a certain configuration.

5 Image Based Illumination

In this section, the core algorithm is extended to incorporate image-based illumination [30–32], using high-dynamic range environment maps [33,34]. Similar to other approaches [35–37] we pre-filter the environment. Therefore, we use HDRshop [38] to generate a latitude-longitude lookup map out of a cube map and apply gaussian blur as a filter as shown in Figure 6. The black dots on the right image are the projected light directions. We now have the environment parameterized with $(\theta, \phi)$, with the angle $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$, which is also known as Mercator projection. Thus, we can easily look-up the intensity from the light coming out of the light direction $l_j$. When the environment map is rotated relative to the object, the light directions are multiplied by the same rotation matrix to obtain the new look-up positions. We apply standard gamma correction to all vertex color values.

6 Dynamic Geometry and Animations

It should be obvious by now, that in an environment with a single object, a rigid transformation can be handled quite efficiently: a translation can be ignored, while a rotation just amounts to an additional matrix multiplication before the look-up into the visibility matrix. If there are several objects changing positions relative to each other or an object changes shape, than the complete visibility matrix $M$ needs to be re-computed. However, this can still be done at interactive frame rates, because our algorithm does not need any spatial accelera-
tion structures. We only need to re-initialize list \( U \) (Section 4) and update the vertex arrays for the objects. Figure 7 shows sample key frames of an animation of a skeleton running at interactive frame rates, while changing the viewpoint and the lighting environment. In each frame, \( M \) is re-computed according to Figure 3.

### 7 Results and discussion

This section presents performance measurements of our new algorithm, which was run on an AMD Athlon64 3200+ (2.0 GHz) under Windows 2000 and an ATI Radeon 9800 XT using OpenGL 1.5. All images and videos were rendered at a resolution of 1280 \( \times \) 960 pixels. This paper is also accompanied by a video. As for the rendering results, we have chosen to render all our objects without textures to make the visual effects of our algorithm more evident. As a matter of course, our method can be combined very easily with texturing or standard shadow mapping techniques [3, 22], in order to handle point light sources as well. The left image on Figure 8 shows, that our method is feasible to render large objects for medical visualization. The \( \text{teeth} \) object consists of 116k vertices. Detailed surface structures are visible. The middle and right image in Figure 8 show several objects that occlude each other from light directions, which is handled correctly by our method. The bunny between the dragons is much darker, due to the occlusion. Figure 9 shows a comparison of our method with OpenGL Phong lighting and a ray-traced image, in a high-dynamic range illumination environment. To achieve the same visual quality 500 samples/ray were needed and it took over half an hour compared to under 1 minute using the new method. The \( \text{venus} \) object consists out of 134k vertices. The slight color shift is due to the slightly different exposures and environment orientations. Table 1 gives an overview over the computation time of the visibility matrix (time\(_1\)) in milliseconds for several different objects and light directions \( k \).

<table>
<thead>
<tr>
<th>object</th>
<th>vertices</th>
<th>( k )</th>
<th>time(_1) (msec)</th>
<th>time(_2) (msec)</th>
<th>FPS(_1)</th>
<th>FPS(_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bunny static</td>
<td>35k</td>
<td>6</td>
<td>445</td>
<td>339</td>
<td>8.08</td>
<td>8.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>1082</td>
<td>797</td>
<td>3.66</td>
<td>4.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>66</td>
<td>3606</td>
<td>2666</td>
<td>1.14</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>258</td>
<td>13715</td>
<td>10373</td>
<td>0.30</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1026</td>
<td>60304</td>
<td>39765</td>
<td></td>
<td></td>
</tr>
<tr>
<td>teeth static</td>
<td>116k</td>
<td>6</td>
<td>1495</td>
<td>1257</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>3651</td>
<td>3041</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>66</td>
<td>12243</td>
<td>10290</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>258</td>
<td>46482</td>
<td>38976</td>
<td></td>
<td></td>
</tr>
<tr>
<td>skeleton animation</td>
<td>8325</td>
<td>6</td>
<td>133</td>
<td>113</td>
<td>18.85</td>
<td>22.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>337</td>
<td>245</td>
<td>9.23</td>
<td>12.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>66</td>
<td>937</td>
<td>740</td>
<td>6.63</td>
<td>8.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>258</td>
<td>3301</td>
<td>2472</td>
<td>3.03</td>
<td>4.63</td>
</tr>
<tr>
<td>trousers animation</td>
<td>3219</td>
<td>6</td>
<td>49</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>103</td>
<td>66</td>
<td>9.23</td>
<td>12.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>66</td>
<td>330</td>
<td>222</td>
<td>3.03</td>
<td>4.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>258</td>
<td>1217</td>
<td>811</td>
<td>0.82</td>
<td>1.35</td>
</tr>
</tbody>
</table>

Table 1: Computation times (\(time\)) for the visibility matrix for static objects and animations with different lightsphere configurations \( (k) \) and frame rates (FPS) when rendered. Times with \( 1 \) are without and \( 2 \) with vertex filtering enabled. See text for details.

### 8 Conclusions

We have presented a new method to calculate vertex-light direction visibility, thus providing a first-order approximation of the rendering equation. This is done by sampling the environment by several directional light sources and efficiently computing vertex visibility from these light directions using hardware-accelerated occlusion queries. In conjunction with vertex filtering optimization, we are able to handle deformable objects and animations at interactive frame rates. The proposed method also allows the usage of image based illumination stored in filtered high-dynamic environment maps. The algorithm drastically reduces visibility calculation times and might be incorporated in other visibility determination problems. Because the method is vertex based, artifacts may occur due to undersampling, which is, of course, true for all vertex-based approaches.

#### 8.1 Future Work

As a drawback, the proposed method has to store the visibility list \( M \) on the CPU, because the query result is always sent back from the graphics card. It would be a great performance increase, if this result...
would be available directly on the GPU on future hardware. Additional speed-up could be achieved by allowing parallel query updates through segmented result buffers. To achieve real-time frame rates for an animation or as a general speed-up, we want to implement our algorithm in a parallel environment.

8.2 Acknowledgements

Some of the used high-dynamic range images were made by Paul Debevec [32, 34]. The Stanford bunny and some other models are from the Stanford Computer Graphics Laboratory and Georgia Tech [39–41]. The teeth and venus sample were generated within the scope of the BMBF project "Virtual Try-On" with the cloth simulation engine TüTex [43–45]. The skeleton animation was made using POSER by Curious Labs.

The authors wish to thank Jutta Adelsberger and Mark J. Kilgard.

References


[34] http://www.debevec.org/Probes/.


[38] http://www.debevec.org/HDRShop/.


Figure 7: Sample key frames of the skeleton animation running at interactive frame rates. A lot of self-shadowing occurs among the bones.

Figure 8: Left image: Large object *teeth* (116k vertices) rendered using our new algorithm under homogeneous white illumination. Middle and right images: Scene with several objects, rendered under high-dynamic range or homogenous white illumination. Note the darker bunny between the dragons.

Figure 9: Comparison between different rendering methods of the venus object (134k vertices). From left to right: wireframe, OpenGL Lighting, under HDR illumination with our new algorithm and raytraced.