Ray-Traced Reflections in Real-Time Using Heuristic Based Hybrid Rendering

Paulo Andrade       Thales Sabino       Esteban Clua      Paulo Pagliosa
UFF             Tecgraf/PUC-RJ           UFF                    UFMS
IC, Medialab                                       IC, Medialab           FACOM

Abstract

Game engines typically use cube maps or screen space local reflections for simulating reflections effects at the real-time renderer pipeline. While these techniques work in many situations, they cannot deal well with reflections of reflections, reflections from covered spaces, reflections from elements outside the screen space or reflections from elements that change its shape or position in real-time. In this paper, we present a method capable of creating true real-time ray-traced reflections by using a heuristic based method and a hybrid renderer. Our solution maintain real-time frame rates by selecting, for each frame, the best reflexive surfaces to ray trace, based on the user’s point of view.

Keywords: ray-tracing, real-time rendering, hybrid rendering, specular reflection

Authors’ contact: paulo@andrade.com; tsabino@tecgraf.puc-rio.br; esteban@ic.uff.br; pagliosa@facom.ufms.br

1. Introduction

Reflections are a typical component of real environments, making it an important factor in the creation of realistic digital images. Reflections can be specular, producing a mirror like effect in the surface, and can be diffuse, generating blurred reflections. The surface can also have both characteristics, with different intensities for each reflection.

Current game engines implement one or more techniques to produce reflections. No current technique can offer true real reflections for all the possible situations. Some techniques cannot render reflection of elements outside the screen space or elements occluded by others, when considering the camera field of view.

Other methods cannot deal with moving elements. Most of these limitations are due to the deferred rendering and lighting process, commonly used by these real-time rendering engines, where only fragments of the elements inside the screen space are shaded.

This work presents a strategy to offer true ray-traced specular reflections, in real-time, by using a heuristic that can select, also in real-time, the best elements to receive reflections, considering the first-person view of the virtual environment. These elements are inside the camera’s field of view and are the elements that best contribute to the user’s visual experience. This selective rendering approach renders only the most relevant elements, within specific time constraints. This time constraints are used to maintain real-time frame rates. Tests show that, when focusing on the most relevant elements in the virtual environment, it is possible to offer superior visual experience, while maintaining real-time frame rates.

Our heuristic based hybrid rendering strategy can overcome some rendering limitations imposed by methods used in current game engines. Some of these methods are discussed in Section 2. The heuristic approach can also accommodate other ray-traced effects, like refractions and shadows.

2. Related Work

In order to generate reflections, most game engines support an old method called cube map [Greene 1986], a variant of the environment mapping method, proposed by Blinn and Newell [1976]. This planar reflection method maps all possible reflections onto six cube faces. A texture covering these faces represents the surrounding environment. This texture must be updated to reflect changes in the environment. Cube maps have three shortcomings. First, it does not deal with motion...
parallax [Yu et al. 2005]. Second, constant updating the cube map affects the overall performance. Third, the reflector element is omitted from the environment map, making reflections from reflections impossible.

First mentioned in [Sousa et al. 2011], Screen Space Reflections, also called Real-Time Local Reflections, allow curved surfaces to reflect nearby surroundings in real-time. This method also allows self-reflections but cannot reflect occluded surfaces or surfaces outside the camera’s field of view.

Reflection Capture Actors, also called Environment Probes [Sinha et al. 2012], are invisible objects that are placed throughout the environment, to capture virtual environment visual information. These objects, usually a cube or sphere, capture visual information from the virtual environment and feed reflection data into the Reflection Environment, which allows image-based lighting and reflections. Multiple probes are used to reduce parallax problems related to image-based methods. Strategically placed probes can feed reflection data from near elements. Near probes have priority against distant probes, reducing the parallax effect. Environment Probes can produce reflections from static elements outside the screen space. Environment Probes main limitation is that the probe placement is static and happens during the level design process. Environment Probes are inspired by Paul Debevec work with physical light probes [Debevec 2001, 1998] to capture HDRI data.

In order to design the heuristic, observations from the selective rendering [Cater et al. 2003, 2002; Chalmers et al. 2007; Debattista et al. 2005; Hasic and Chalmers 2009; Lever et al. 2005; Mania et al. 2008], and visual attention [Cater 2004; Desimone and Duncan 1995; Green and Bavelier 2003; Parker et al. 1999; Sundstedt et al. 2005; Yan and Seif El-Nasr 2006] fields of study were considered. Both field of studies supports the idea that it is not necessary to render every detail of moving images, in order to offer a rich visual experience.

Also related to this work are the general principles of good level design for first-person games [Brown and Chen 2001; Dormans 2010; Feil and Scatteredgood 2005; Hullet 2010]. Good level design is important not only for entertainment, but for overall game performance. A poorly designed level can cause game slowdowns that degrade user’s experience.

3. PHRT Renderer

PHRT is the name of the heuristic based, real-time hybrid rendering engine used in this work. PHRT is an improved version of the engine presented in [Sabino et al. 2012; Sabino 2012] and is capable of ray-trace specific elements in the virtual environment, according to a heuristic input. The decision if which elements should be rendered using ray tracing happens for every frame. An early version of PHRT was used in [Andrade et al. 2014a, 2014b]. PHRT’s current version can better deal with reflections and have many performance improvements. In addition, both the proposed heuristic and PHRT’s current version were specifically designed to deal with recursive reflections, where one reflexive surface reflects another reflexive surface.

PHRT use deferred shading [Deering et al. 1988; Pritchard et al. 2004; Saito and Takahashi 1990; Whitted and Weimer 1982] for rasterization and OptiX [Ludvigsen and Elster 2010; Parker 2009; Parker et al. 2010] for ray tracing effects. PHRT has the following features:

- Render in real-time reflections, refractions, ray-traced and shadow mapping shadows;
- Render the virtual environment using only OpenGL, only OptiX ray tracing or a heuristic based hybrid mode;
- Render pre-defined camera animation, in order to test the heuristic and collect data for statistics.

In order to have a steady real-time frame rate, the ray-traced reflections phase have a time limit. The heuristic takes into account this time constraint to decide which elements can be ray-traced.

Figure 2 is an overview of PHRT rendering pipeline. The pipeline is described in Sections 3.2 to 3.6.

![Figure 2: PHRT rendering pipeline.](image-url)
3.1 Pre-Production

The Pre-Production phase is an off-line phase and happens before PhRT’s execution. In this phase, the environment or level designer assigns a priority index $K$ for each reflexive surface. This information is stored in an XML file that is read in the initialization phase. $K$ value identifies how important one element is when compared to the other elements in the virtual environment. An element with a high $K$ value is considered more relevant than an element with a lower $K$ value. With $K$, the environment designer helps the heuristic to do good choices, by informing which elements are more important to render during the ray-traced reflections phase.

3.2 Initialization

The Initialization phase starts before the proper rendering stage. PhRT loads the 3D meshes, textures, rendering parameters and heuristic parameters. The virtual environment is built in the computer’s memory, and both the deferred and OptiX rendering processes are initialized.

The Selection Graph is built during the Initialization phase. The Heuristic Decision phase uses the Selection Graph to select which elements in the virtual environment must be rendered using ray tracing. Section 4 present the Selection Graph construction and update procedures.

3.3 Deferred Shading and Primary Ray Resolution

The Deferred Shading and Primary Ray Resolution phase uses the virtual environment data to fill the G-buffer. Visibility tests are done using the G-buffer data, and basic light computation is done using OpenGL. OptiX uses the Z-Buffer data stored in the G-buffer as the equivalent of the primary ray hit phase of the ray tracing algorithm. In PhRT, OptiX is used to trace secondary rays for shadows, reflections and refractions.

3.4 Heuristic Decision

In general terms, the heuristic must select the best candidates for the Ray-Traced Reflections phase. The heuristic goal can be defined as “for a given set of $X$ elements, select the $Y$ most relevant elements that can be ray-traced given a time constraint $T$.” $Y$ are the more relevant elements at that moment, according to the heuristic. $T$ is the time available for the ray tracing phase. $X$, $Y$ and $T$ can change for every frame, according to the camera position in the virtual environment. $X$ does not represent all the reflexive elements in the virtual environment, just the elements involved in the in the current image generation.

3.5 Ray-Traced Reflections

With the virtual environment information stored in the G-buffer, secondary rays are traced from each selected surface into the virtual environment. If a secondary ray hit one element in the virtual environment, the element color and light information is retrieved for the compositing phase. The environment map information is used if there is no element in the ray’s path. If the secondary ray hit a reflexive element, this element is included in the Selection Graph if it is not already in. This procedure allows reflections from reflexive surfaces outside the camera’s field of view.

After OptiX computes all the reflections for the selected elements, PhRT verifies if there is still time available in the Ray-Traced Reflections phase. If there are, the Selection Graph is updated, and the Heuristic Decision phase is executed again, while there is time for rendering.

3.6 Compositing

Compositing is the final phase, when both the raster and ray traced information is combined to generate the final image. Figure 3 shows PhRT’s full rendering pipeline, also showing ray traced shadows and refractions.

Figure 3: Visual representation of the rendering pipeline.

4. Heuristic Mechanics

The heuristic core is the selection of the best candidate elements to ray trace, giving a specific time constraint $T$. The heuristic was designed considering first-person point of view.

The three basic observations considered, in order to design the heuristic are:

1. During the virtual environment exploration, first-person game players and virtual reality users tend to center the focus of the camera at the most relevant elements in the virtual environment;

XIII SBGames - Porto Alegre - RS - Brazil, November 12th - 14th, 2014 923
2. First-person game players and virtual reality users pay more attention to elements near and in front of the virtual camera.
3. Elements that frequently change its appearance attract more attention than elements that maintain its overall appearance.

These considerations were used to design the visibility metric \( V \), presented in Section 4.1. Visibility is computed for each candidate element to ray trace and is used to calculate the processing cost \( C \) and relevance \( R \) of each candidate element. \( C \) and \( R \) are stored in the Selection Graph.

With \( C \) and \( R \) values for each candidate element, PHRT constructs and maintain a directed graph as the Selection Graph for the Heuristic Decision phase. The Selection Graph is constructed during the Initialization phase and is updated during the Heuristic Decision phase. Later, when there is no more time available for rendering, the Selection Graph is rebuild.

This heuristic have some similarities with the heuristics presented in [Andrade et al. 2014b], but was adapted to deal specifically with reflections. One of the key differences is that this heuristic can deal dynamically with reflections from reflections.

### 4.1 Selection Graph Construction

The Selection Graph is first constructed during the Initialization phase and considers all the candidate (reflective) elements in the virtual environment. Every node in the Selection Graph corresponds to an element. In addition, every node has an estimated processing cost \( C \), and a relevance \( R \). \( C \) is estimated by taking into account the element visibility \( V \) and its ray tracing cost \( Q \). The ray tracing cost \( Q \) initial value is 1 for every element. \( Q \) value is increased by 1 when a secondary ray from its surface hit another reflexive surface in the virtual environment, and so on. \( Q \) is the equivalent of the number of ray bounces between reflexive surfaces, and it is limited to 5, representing 5 bounces after the primary ray hit in the reflexive surface. This limitation is only due to performance constraints of the target hardware used in the tests. When two reflexive surfaces are involved, \( C \) is the sum of the costs of each surface.

Equation (1) present visibility \( V \). In Equation (1), \( A \) is the visible area of the element projection in the view plane of the camera. \( P \) is the normalized distance between the geometric center of the element bounding box and the center of the view plane. \( P \) equals 1 means that the element is in the view plane center, and \( P \) equals 0 means that the element is outside the view plane. \( D \) is the distance between the element and the view plane in the virtual environment. The higher is distance from the view plane, the smaller is visibility \( V \), making big and distant elements less “visible” than near elements with the same area \( A \) in the view plane. Equation (2) present the element cost \( C \), \( O \) is only used when the selected element also reflect a reflexive surface. \( O \) equals 1 when the distance between the two reflexive surfaces is equal or less than \( D \). \( O \) increases in 1 every time the distance between the two reflexive surfaces doubles, when compared to \( D \).

\[
V = \frac{(A \times P)}{D}
\]

\[
C = \sum (V \times Q / O)
\]

While cost \( C \) is a strategy to measure how much work is necessary to render a particular element, relevance \( R \) represents the element contribution to the visual experience. In (3), selection \( S \) is a binary value where 1 means that the element was previously selected for ray tracing and the value 0 means it was not. Selection weight \( I \) is a constant that defines how important is select a previously selected element.

\[
R = (S \times I + V) \times K
\]

The Selection Graph has two types of edges. The first type is a linked list, where one node connects to the other node where the cost \( C \) is smaller than its own cost but bigger than the cost \( C \) of all the other nodes. The second type, also a linked list, connects each node to all the other nodes where relevance \( R \) is bigger than its own relevance but with cost \( C \) equal or less than its own cost. The edges related to cost \( C \) identify the elements that can be rendered, given the time constraint \( T \), and the edges related to relevance \( R \) are used to identify the element that best contribute to the visual experience.

Figure 4 represents a Selection Graph with the edges related to relevance \( R \). The cost \( C \) for each node is in the green (left) part of the circles. The relevance \( R \) is in the blue (right) part of the circles. The arrows point to the nodes with cost smaller or equal than the actual node and relevance equal or bigger than the actual node.

![Figure 4: Visual representation of a Selection Graph.](image-url)
Table 1 summarizes all the parameters for Equation (1), (2) and (3). For each parameter, Const. and Var. indicate if the parameter value is constant or variable during PHRT execution.

Table 1: Heuristic parameters.

<table>
<thead>
<tr>
<th>Param</th>
<th>Definition</th>
<th>Const</th>
<th>Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>Element relevance among the others</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>V</td>
<td>Element visibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>Ray tracing cost</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Factor for the relative distances between 2 reflexive surfaces</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Processing cost</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Element area in the view plane</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>Distance from the center of the view plane</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Distance from the view plane (camera)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Relevancy</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>Element was previously selected or not</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>Weight of element being previously selected</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

At the end of this phase, the heuristic has a directed graph ready for use during the rendering phase.

4.2 Node Selection

Node Selection happens during the Heuristic Decision phase. The heuristic selects the $N$ most relevant $R$ nodes whose cost $C$ allows the nodes to be rendered and still maintains the expected frame rate. $N$ is the number of the stream processors of the target GPU. If there is still time to render other elements, the heuristic choose the most relevant nodes that have cost $C$ small enough to be rendered in the available time $T$.

4.3 Selection Graph Reconstruction

When there is no more time available to select another node, the Selection Graph is reconstructed and updated for the next rendering phase. New cost $C$ and relevance $R$ are calculated and the node edges are rebuilt. When the new graph is created, the renderer clears the G-Buffer and renders the new frame.

5 Tests and Results

PHRT, the real time hybrid renderer developed for this work is not a true game engine so, many common optimizations present in the game engines are not implemented, which limit our tests to small environments. In addition, at the moment of the tests, OptiX was not optimized for games, making the ray tracing part of the pipeline very demanding. However, these limitations do not impair the validation of heuristic use, on the contrary, with an increased rendering performance, the heuristic can offer even better results.

To test our heuristic, we created two virtual environments, one outdoor and one indoor. The outdoor environment is a cartoon like farm, where ponds, troughs and barrels have reflexive surfaces. The indoor environment is an underground sewer, where some elements, the water surface and some walls are reflexive. The water surface of the indoor environment is divided in segments, since the water surface covers huge areas.

For both environments, a fixed forward only camera path was created. These paths are used test how the heuristic performs considering element placement and surface sizes. In both paths, the camera “see” elements with different sizes and areas with big and small number of reflexive surfaces. Both camera paths runs for 10.000 frames in the animation and modeling software. For the tests, both paths were traversed 10 times. During the traversal, we measure the time to render each frame and which elements were selected for ray tracing, during the frame rendering.

The indoor environment was designed to offer many opportunities for recursive reflections, as shown in Figure 5.
and TaD – Sewer Kit [Forge 2014], respectively. The reflexive surfaces in both environments are colored in blue in Figures 5 and 6, but in a true virtual environment, a texture should be applied.

In order to test our heuristic (HEU1) qualitatively, we developed a distance heuristic (HEU2). HEU2 select elements that can be rendered in real time and are near the camera. The center of the bounding box around each element is used to measure the distance between the element and the camera.

All the tests were performed in a desktop computer with a 3.50 GHz Intel Core i7-4770K CPU and a NVIDIA GeForce GTX 760 GPU. This computer runs Windows 8.1.

The three metrics used to evaluate the proposed heuristic are described as follows. Since the heuristic maintain the minimum of 20 frames per second, the most important tests are qualitative.

5.1 Transitions Between Rendering Methods

When moving inside a virtual environment, the camera can change frequently between ray tracing and raster rendering. Frequent changes can cause poor visual experience. This metric considers the number of changes during each traversal, for each heuristic. Because of its naïve strategy, HEU2 produce many more rendering method changes than HEU1. These changes can easily be perceived when the camera moves forward and backward frequently. For HEU1, previously selected elements have priority to be rendered in ray tracing. Parameters $S$ and $I$ control this priority, reducing the number of changes.

Figure 8 shows a transition between raster, ray tracing and raster again, in order to illustrate the effect of frequent changes in rendering methods.

<table>
<thead>
<tr>
<th>Environment</th>
<th>HEU1/HEU2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outdoor</td>
<td>3.8 %</td>
</tr>
<tr>
<td>Indoor</td>
<td>4.6 %</td>
</tr>
</tbody>
</table>

Table 2 present the percentage of transitions produced by the proposed heuristic (HEU1), when compared to the distance heuristic (HEU2), for both environments.

As shown in Table 2, the number of transitions produced by our heuristic is a fraction of the transitions produced by the distance heuristic. The large difference shows that a good heuristic must have a method to deal with constant changes between rendering methods.

The experiments also shows that HEU1 performs better, when the value of $I$ is, most of the time, bigger
than \( V \). However, if \( I \) is always much bigger than \( V \), elements near the camera are never selected.

### 5.2 Average Number of Elements Selected

The average number of elements selected for ray tracing is an important metric, since it shows how efficient the heuristic is in rendering elements. This metric is affected by the average size of the reflexive surfaces but, in general, the more the number of elements selected, the better the heuristic is.

Table 3 presents the average number of elements selected for each environment and heuristic, and the total number of reflexive surfaces.

<table>
<thead>
<tr>
<th>Environment</th>
<th>HEU1</th>
<th>HEU2</th>
<th>Surfaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outdoor</td>
<td>5</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>Indoor</td>
<td>15</td>
<td>12</td>
<td>33</td>
</tr>
</tbody>
</table>

The relative small number of selections for the outdoor environment is due to the camera path. In a good part of the path, just a small number of reflexive elements are inside the camera view. Also, since both ponds have high priority (high \( K \) value), the ponds are rendered most of the time, giving little processing resources for other reflexive surfaces. For the indoor environment, the opposite happens. The water surfaces have low priority, allowing more reflexive elements to be selected. As expected, this observation shows how important is a good level design, in order to offer improved visual experience.

### 5.3 Average Frame Rate

Table 4 presents the average frame rate for each environment and heuristic.

<table>
<thead>
<tr>
<th>Environment</th>
<th>HEU1</th>
<th>HEU2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outdoor</td>
<td>25.4</td>
<td>26.4</td>
</tr>
<tr>
<td>Indoor</td>
<td>21.8</td>
<td>23.3</td>
</tr>
</tbody>
</table>

HEU2 performs better than HEU1, but this is expected since, as presented in 5.2, HEU2, on average, selects fewer elements to render than HEU1. Also, the path has a key influence, as observed in Section 5.2. The indoor path, on average, has more reflexive elements visible to the camera than the Indoor environment.

### 6 Conclusions

This paper proposes a heuristic to render reflections in real time, using a hybrid renderer. The main advantage of this method is to offer reflections without screen space limitations and true recursive reflections.

The first major conclusion is that true real-time ray-traced reflections are achievable using a well-designed heuristic. The selective rendering approach has the drawback of creating frequent transitions between rendering methods. However, as observed in Section 5.1, the impact of this problem is minimized with a well-designed heuristic, that have a strategy to reduce transitions.

Another conclusion resulted from the visual observation of the real-time rendering of many tests is that the heuristic does good decisions considering elements to render. With the level designer input, the heuristic prioritizes the right elements to render. For the element without high \( K \), most of the time, the heuristic select elements near the camera and the center of the view plane. When moving in real time, it was observed the elements in the corners of the view plane are barely noticeable. Unfortunately, visual quality tests are much more complex to do than numerical tests, and the distance heuristic is too simple to be for visual comparisons. However, when compared to raster only rendering, reflections clearly improve the visual experience.

As future works, there is the possibility that both Unreal Engine 4 [Rege 2014a] and CryENGINE [Rege 2014b] will support OptiX by means of supporting NVIDIA GameWorks [Coombes 2014]. With commercial engines supporting OptiX, it will be possible to migrate the heuristic to these engines and test the heuristic in a real environment designed for games. With this possibility, new heuristics can be developed. In addition, it will be easier to evaluate the visual quality and overall performance impact of heuristic based rendering in the rendering pipeline.

### References


Parker, S., 2009. Interactive ray tracing with the NVIDIA®OptiX™engine. SIGGRAPH.


doi:10.1145/97880.97901


