Realistic Image Synthesis

- Perception-based Rendering -

Philipp Slusallek
Karol Myszkowski
Gurprit Singh
Making Rendering Efficient

• **Realistic image synthesis goal**
  – Generate an image that evokes from the visual perception system a response that is indistinguishable from that evoked by the original environment
  – Global illumination important component of realism

• **The solution of the global illumination problem is computationally hard:**
  – Take into account characteristics of the Human Visual System to concentrate the computation exclusively on the visible scene details
Outline

• Perceptually based adaptive sampling algorithm
• Steering Monte Carlo ray (path) tracing using perception inspired image quality metrics
• Image-based rendering for animations
• Eye tracking driven rendering
A Perceptually Based Adaptive Sampling Algorithm

by Mark Bolin & Gary Meyer

SIGGRAPH 1998

- Uses a multi-scale visual model (the Sarnoff Visual Discrimination Model) to guide the sampling pattern in MC Ray Tracing
  - Optimized for speed
    - Haar wavelets are used at the cortex filtering stage instead of costly Laplacian pyramid originally used in the VDM
  - Correct color handling
    - CIE XYZ transformed to SML space modeling retinal cone sensitivity
    - Opponent contrast space: a single achromatic (A) and two opponent color channels (C₁ and C₂)
    - Independent contrast sensitivity processing for AC₁C₂ channels
Chromatic CSF

Independent contrast sensitivity processing for $AC_1C_2$ channels

Band-pass filter

Low-pass filter

Luminance

Red-Green Opponent

Blue-Yellow Opponent
Visual Masking

- Achromatic and chromatic CSFs with noise (left), and perceptual metric response in the comparison with noiseless CSF (right).
- Brighter shades denote better noise visibility (less masking).
A chapel image *without* (left) and *with* imposed sinusoidal distortion (center). Visual difference metric results (right): brighter shades of grey denote less masking and better visibility of the sinusoidal distortion pattern.
Perception-based Adaptive Sampling

- **Step I:** compute an estimate of the image using lesser number of samples per pixel
  - A Haar wavelet image approximation is generated and then refined

- **Step II:** from MC variance in samples of each pixel estimate the pixel error bounds.
  - The error expressed in terms of the variance of the detail terms in the Haar image representation
Perception-based Adaptive Sampling

- **Step III:** from an **Estimated Image** and error-bounds compute a **Lower Bound Image** and an **Upper Bound Image**.
Perception-based Adaptive Sampling

- **Step IV**: Compute oriented band-pass images.
Perception-based Adaptive Sampling

• **Step V:** For each band compute threshold from TVI, CSF and Masking functions. Normalize the band pass images with the computed threshold.
Perception-based Adaptive Sampling

- **Step VI**: Find the difference between each band of the two images.

![Image 1](image1.png) ![Image 2](image2.png)

- **Step VII**: Refine the area with maximum difference.
Perception-based Adaptive Sampling

- Algorithm summary

1. Sample scene
2. Cone fundamentals
3. Refine cortex representation
4. Refine error estimate
5. Construct boundary images
6. Local visual difference prediction
7. Update maximum error tree
8. Determine next sample location
Perception-based Adaptive Sampling

Image

Sample Density
A Perceptually Based Physical Error Metric for Realistic Image Synthesis

by Mahesh Ramasubramanian, Sumanta N. Pattanaik, and Donald P. Greenberg
Siggraph 1999

Aims for perceptual accuracy

• Limitations of the human visual system...
  perceptual accuracy < physical accuracy
• Perceptual accuracy guides rendering, not physical accuracy.
A Perceptually Based Physical Error Metric for Realistic Image Synthesis

by Mahesh Ramasubramanian, Sumanta N. Pattanaik, and Donald P. Greenberg

Siggraph 1999

Aims for perceptual accuracy

• Limitations of the human visual system...
  perceptual accuracy < physical accuracy.
• Perceptual accuracy guides rendering, not physical accuracy.
Realistic Image Synthesis SS19 – Perception-based Rendering

Preview

effort distribution (darker regions - less effort)

physically accurate

perceptually accurate

6% effort
Preview

physically accurate

perceptually accurate

effort distribution (darker regions - less effort)

6% effort
Perceptually Based Rendering

Traditional approach:
Pair of images to compare at each time step

(a) intermediate images at consecutive time steps.

(b) upper and lower bound images at each time step.
Perceptual Error Metric

Vision model - expensive

physical domain

perceptual domain

vision model

visual rep. 1

<

perceptual threshold

visual rep. 2

perceptual difference
Perceptually Based Physical Error Metric

Physical domain

Perceptual threshold < physical threshold

Perceptual domain

Perceptual threshold < perceptual threshold

Perceptual difference
Physical Threshold Map

Predicted bounds of permissible luminance error

input image

threshold model

physical threshold (brighter regions - higher thresholds)
Threshold Model

Components

image → luminance component → frequency component → contrast component → threshold map
Threshold Model

1. Luminance component

![Graph showing the relationship between log adaptation luminance and log threshold. The graph indicates a threshold value of 2% due to luminance.]
Threshold Model

2. Frequency component

- Frequency component
- Threshold due to luminance + freq.

log threshold factor

log Spatial Frequency (cpd)
Threshold Model

3. Contrast component (visual masking)

threshold due to luminance + freq. + contrast

masking function

log threshold factor

log contrast

15%

30%
Validation

image + noise = image + noise
Threshold Model

image → luminance component → frequency component → contrast component → threshold map
Global Illumination Revisited

global illumination = direct illumination (fast) + indirect illumination (slow)
Threshold Model Revisited

Spatially-dependent processing

Luminance-dependent processing

Partial global illum.

Direct illum.

12 sec

1 time precompute

0.1 s

N times iterate

Realistic Image Synthesis SS19 – Perception-based Rendering
Adaptive Rendering Algorithm

1. start
2. precompute
3. spatial info.
4. direct illumination
5. refine global illumination
6. perceptual error
7. n
8. good enough?
9. y: done
10. iterate
Results

5% effort

effort distribution (darker regions - less effort)

reference solution

adaptive solution
Results: Masking by Textures

effort distribution (darker regions - less effort)

reference solution

adaptive solution

5% effort
Results

5% effort

direct illumination + adaptive indirect illumination = adaptive global illumination

noisy

masked
Results: Masking by Geometry

5% effort

effort distribution (darker regions - less effort)

reference solution

adaptive solution
Results: Masking by Shadows

effort distribution
(darker regions - less effort)

6% effort

reference solution

adaptive solution
New and efficient perceptually based global illumination technique.

Advantage:
Exploits spatial information in scene, but computes it only once.

Limitation:
Only for view-dependent rendering.
Incorporating temporal sensitivity.
Handling Moving Patterns: Spatiovelocity CSF

- Contrast sensitivity data for traveling gratings of various spatial frequencies were derived in Kelly’s psychophysical experiments (1960).

- Daly (1998) extended Kelly’s model to account for target tracking by the eye movements.
Deriving Pixel Flow Using Image-Based Rendering Techniques

Realistic Image Synthesis SS19 – Perception-based Rendering
Animation Quality Metric (AQM)

- Perception-based visible differences predictor for still images was extended.
- Pixel Flow derived via 3D Warping provides velocity data as required by Kelly’s SV-CSF model.
Image-based Rendering for Animations

• Use ray tracing to compute all key frames and selected glossy and transparent objects.
• For inbetween frames, derive as many pixels as possible using computationally inexpensive Image Based Rendering techniques.
• The animation quality as perceived by the human observer must not be affected.
Keyframe Placement

- The selection of keyframes should be considered in the context of the inbetween frame computation technique.
- In IBR techniques reference frames are usually placed:
  - uniformly in space at the nodes of 2D or 3D grid (Chen95),
  - uniformly along the animation path (Mark97),
  - at manually selected locations (Darsa97).
- A notable exception is work done by Nimeroff et al. 1996, who used a simple quality criterion.
Keyframe Placement

• Our goal is to find inexpensive and automatic solution, which reduces animation artifacts which can be perceived by the human observer.
• Our solution consists of two stages:
  – initial keyframe placement which reduces the number of pixels which cannot be properly derived using IBR techniques due to occlusion problems,
  – further refinement of keyframe placement which takes into account perceptual considerations, and is guided by AQM predictions.
Keyframe Placement

Animation Quality Predictor

Are the differences acceptable?

- Split segment
- Recurse

YES

Generate inbetween images

NO
Atrium: final keyframe placement

Animation path with marked keyframe locations
In-between frame generation
Visualization of the AQM Responses

No eye tracking. PF x 1. P(>0.75)=10.5%

No eye tracking. PF x 3. P(>0.75)=3.0%

Probability of detecting the differences
Examples of final frames

Supersampled frame used in traditional animations

Corresponding frame derived using spatiotemporal filtering

In both cases the perceived quality of animation appears to be similar!
Eye Tracking - Motivation

1. **Improving computational efficiency**
   - There is a trend towards higher resolution displays
     → Higher computational requirement for 3D rendering
   - Only a fraction of pixels is consciously attended and perceived in the full-resolution

2. **Improving realism**
   - Eye is always focused on the screen plane; nevertheless, it is possible to simulate Depth-of-Field (DoF) effect by artificially blurring out-of-focus regions according to the gaze location

3. **Improve perceived quality**
   - Human Visual System (HVS) has local adaptation property
   - Perception of luminance, contrast and color are not absolute and highly dependent on both spatial and temporal neighborhood of the gaze location

Images adapted from [https://www.nngroup.com/articles/computer-screens-getting-bigger/](https://www.nngroup.com/articles/computer-screens-getting-bigger/)
Eye Tracking - Outline

- Basic Technology
- Types of Eye Motion
- Level-of-Detail (LoD) Rendering
- Foveated 3D Graphics
  - Latency
  - Noise
- Depth-of-Field (DoF) Rendering
- Gaze-contingent Stereo
- Local Adaptation
- Subtle Gaze Direction
- Saliency
Eye Tracking

• **Basic Technology:**

  Corneal Reflection (also known as “glint” or “1st Purkinje Reflection”)

  ![Corneal Reflection Image](Image)

  ![Eye Tracker Image](Image)

  ![User Image](Image)

• **Eye trackers mostly operate using infrared imaging technology**

• **Once the pupil is detected the vector between the center of the pupil and the corneal reflection of the infrared light source is translated into the gaze location on screen coordinates**

• **Requires calibration at the beginning**

Images adapted from [http://twiki.cis.rit.edu/twiki/bin/view/MVRL/QuadTracker](http://twiki.cis.rit.edu/twiki/bin/view/MVRL/QuadTracker) and [http://psy.sabanciuniv.edu](http://psy.sabanciuniv.edu)
**Eye Tracking**

- **Individual calibration is necessary for each observer**
- **Relative location of the corneal reflection and the pupil is different among the population due to**
  - Difference in eye ball radius and shape
  - Eye-glasses

Images adapted from [http://wiki.cogain.org](http://wiki.cogain.org)
Eye Tracking

• Some of the other types of setups are used only for specific applications since they may be highly intrusive (e.g. chin-rest eye trackers) and not comfortable for the end-users in practice

• Head-mounted displays (HMD) offer 3D stereo and augmented reality capabilities in addition to eye tracking
Types of Eye Motion

<table>
<thead>
<tr>
<th>Type</th>
<th>Duration (ms)</th>
<th>Amplitude (1° = 60’)</th>
<th>Velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixation</td>
<td>200-300</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><em>Microsaccade</em></td>
<td>10-30</td>
<td>10-40’</td>
<td>15-50°/s</td>
</tr>
<tr>
<td><em>Tremor</em></td>
<td>-</td>
<td>&lt;1’</td>
<td>20’/sec</td>
</tr>
<tr>
<td><em>Drift</em></td>
<td>200-1000</td>
<td>1-60’</td>
<td>6-25°/s</td>
</tr>
<tr>
<td>Saccade</td>
<td>30-80</td>
<td>4-20°</td>
<td>30-500°/s</td>
</tr>
<tr>
<td><em>Glissade</em></td>
<td>10-40</td>
<td>0.5-2°</td>
<td>20-140°/s</td>
</tr>
<tr>
<td>Smooth Pursuit</td>
<td>variable</td>
<td>variable</td>
<td>10-30°/s</td>
</tr>
</tbody>
</table>

- While the mechanisms are not exactly known, it is thought that the brain performs visual suppression and compensation during *saccades* and smooth pursuits against motion blur on the retina.

Eye Tracking in Action

Bayesian Identification of Fixations, Saccades, and Smooth Pursuits

An example of I-BDT classification

Fixation = Solid Red Circle
Saccade = Solid Yellow Circle
Smooth Pursuit = Hollow Yellow Circle

Original framerate: 30 Hz
Playback framerate: 10 Hz

Visual Acuity

- Distribution of photoreceptor cells in the retina

Level-of-Detail Rendering

- The model resolution may be degraded according to the visual angle and the acuity of HVS at the given angle
  - Mesh structure of the model is partitioned into tiles using Voronoi diagram
  - Tiles are mapped to planar polygons
  - Remeshing into multiresolution form

Foveated 3D Graphics

- Screen-based (in contrast to model-based methods)
- Human eye has full acuity in around 5° foveal region
- The efficiency of image generation can be improved by maintaining high image resolution only around the gaze location
- Using 60Hz monitor and Tobii X50 eye tracker with 50Hz sampling frequency and 35ms latency caused artifacts for the observer
- Results using 120Hz monitor and Tobii TX300 with 300Hz sampling frequency and 10ms latency were tolerable

Foveated 3D Graphics

Video adapted from http://research.microsoft.com
Latency Measurement

- **Transition from B** (end of the saccade) to **C** (switching from half to full-resolution in the gaze location):
  - 5, 20, 40, 60 or 80 ms are tested
  - Viewers never detected a change up to a delay of 5 ms after the saccade is completed
- **E2**: the retinal eccentricity where resolution drops to half-maximum
  - Viewers never detected a change for E2 > 6.22°
  - For E2 = 3.11°, the detection rate is <10% for 5, 20, 40, 60 ms delays

Overcoming Eye Tracker Noise

• Accuracy of existing eye trackers is insufficient for gaze-driven Depth-of-Field (DoF) applications
  – P-CR RED250 tracker
    Claimed: 0.5°
    Measured: 1.83° std: 1.07°

• Gaze accuracy is improved by “snapping” the gaze location to the nearest potential focus-point using the information from tracker and 3D scene (including focus-point position and velocity)

Overcoming Noise

Effect of Depth-of-Field

- Improves the rendering realism and enhances the depth perception

**Depth-of-Field Rendering**

- **Circle of Confusion:**

  \[
  CoC = a \cdot \left| \frac{f}{d_0 - f} \right| \cdot \left| 1 - \frac{d_0}{d_p} \right|
  \]

  - \(a\) - diameter of the lens aperture
  - \(f\) - focal length of the lens
  - \(d_0\) - distance between the focal plane and lens
  - \(d_p\) - distance from an object to the lens

- \(d_p\) is obtained from reverse mapping of the z-buffer

- **Addresses the artifacts due to the depth discontinuity near object boundaries by spreading the blur outside the object boundary**

Vergence-accommodation Conflict

Stereo 3D: Binocular Disparity

- Vergence
- Accommodation (focal plane)
- View discomfort
- Depth
- Object perceived in 3D
- Comfort zone
- Pixel disparity
- Object in right eye
- Object in left eye

Screen
Vergence-accommodation Conflict

Depth Manipulation

Viewing discomfort  Viewing comfort

Scene manipulation

Comfort zone
Vergence-accommodation Conflict

Depth Manipulation

Pixel disparity map

Modified pixel disparity

Input pixel disparity

Output pixel disparity

Mapping function

Similar to tone mapping

Function:
- Linear
- Logarithmic
- Content dependent

Other possibilities:
- Gradient domain
- Local operators

“Nonlinear Disparity Mapping for Stereoscopic 3D” [Lang et al. 2010]
Disparity Perception (Stereo 3D)

Sensitivity to disparity corrugations in peripheral vision, Vision Research, Volume 38, Issue 17, September 1998

Replotted from Figure 3 of Simon J.D Prince, Brian J Rogers

Realistic Image Synthesis SS19 – Perception-based Rendering
Vergence-accommodation Conflict

Depth
Vergence-accommodation Conflict
Vergence-accommodation Conflict

More depth

Depth
Vergence-accommodation Conflict

- More depth
- More comfort
Vergence-accommodation Conflict

- More depth
- More comfort
- Seamless

Depth
Vergence-accommodation Conflict

- More depth
- More comfort
- Seamless
- Low cost
Gaze-contingent Stereo

- The region of attention may be predicted to manipulate disparity for comfortable viewing
- The online predictor uses Decision Forests (DF) to predict the object category that the viewer looks at
- A total of 13 game variables are used for prediction (e.g. Health, Hunger, Thirst, Ammo, Distance to the closest robot, ...) which are selected among 300 as the most “informative” ones (ignoring variables with little or no variability)
- The predicted objects in the current scene are placed as close to the plane of zero-disparity as possible

Local Adaptation

• Several physiologically-inspired artifacts may be introduced artificially into the video, depending on the gaze location to improve realism:
  – Adaptation to global lighting level
  – Retinal afterimages
  – Visual phenomena related to low-light (visual acuity loss in low light, Purkinje shift, mesopic hue shift)

Local Adaptation

- Adaptation to global lighting:

\[
A' \left\{ \begin{array}{l}
A + a_1 \Delta t \quad A < A_T \\
A - a_2 \Delta t \quad A > A_T
\end{array} \right.
\]

- Global photographic tone mapping based on Naka-Rushton Equation which predicts the response of photoreceptors after adaptation:

\[
R(I) = \frac{I^n}{I^n + \sigma^n}
\]

Local Adaptation

- **Afterimage**: Image of the stimuli which is still perceived after it ceases

- **May be in the form of:**
  - Bleaching afterimages
  - Local adaptation afterimages

- **Bleaching level** $B$ is given in the form of a differential equation (Baylor et al. 1974):

\[
\frac{dB}{dt} = b_1 (1 - B)I - b_2 B
\]

- $b_1$: bleaching sensitivity
- $b_2$: recovery rate of the photoreceptors
- $I$: incident luminance

Local Adaptation

- **Local adaptation afterimages:**
  - Attributed to the role of calcium ions in phototransduction (Matthews 1996)
  - Updated calcium concentrations after a timestep $\Delta t$:

$$C' \leftarrow (C - C_{\infty})e^{-c_2\Delta t} + C_{\infty}$$

$C'$: calcium concentration in the new timestep

$C_{\infty}$: equilibrium calcium concentration

$C$: calcium concentration in the previous timestep

$c_2$: controls the efflux of calcium

- $B$ and $C$ are used together to compute the pixel intensities in the presence of the afterimages.

Local Adaptation

- **Mesopic illumination range**: $10^{-3} - 10 \text{ cd/m}^2$
- **Mesopic hue shift**
  - As illumination decreased, the perceived color of neutral tones shift to the dull purple (Shin et al. 2004)
- **Purkinje shift**
  - As illumination decreased, the perceived relative intensities of the colors change
- **Visual acuity loss in low lighting**
  - Spatial acuity drops linearly with log-luminance (Riggs 1965)
  - Modeled as stochastic, time-varying loss of high frequency using band-pass filtering

Subtle Gaze Direction

- When viewing an image low-acuity peripheral vision detects areas of interest, then HVS directs gaze to those locations.
- HVS is very sensitive to changes in luminance (Spillmann et al. 1990) and opponent color channels (Hurvich and Jameson 1957).
- Introduces subtle image modulation to control the gaze direction of the observer.
- Luminance and warm-cool modulations are studied and both are found successful.

**Subtle Gaze Direction**

**F**: Fixation point, **A**: Predetermined Area of Interest  
**Goal**: To direct the user attention to from **F** to **A**  
Modulation is applied to **A** and $\theta$ is monitored real-time. When $\theta \leq 10^\circ$, the modulation is terminated immediately.

Subtle Gaze Direction

Top: Input image, Left: No modulation, Right: Modulation at white crosses

Visual Attention

- Shrink the amount of visual information reaching the eye to a manageable size
- Useful metaphor:
  - spotlight that enhances selected regions
- Two components of visual attention:
  - **bottom-up component**: fast; preattentive; primitive mechanism responding to color contrast, intensity contrast, orientation, ...
    - Itti saliency model – a popular choice
  - **top-down component**: slower; under cognitive control; task-driven

![Diagram](image.png)
Modeling Visual Attention


Realistic Image Synthesis SS19 – Perception-based Rendering
Saliency (Itti Model)

- Attention activity may be controlled in bottom-up (scene-dependent) and top-down (task-dependent) manner
- Model based on the bottom-up architecture proposed by Koch and Ullman:
  - Visual layer is decomposed into feature maps
  - The locations which stand out from their surround persist
  - All feature maps fed into a master saliency map

Saliency

• **General computation principle in the retina, lateral geniculate nucleus and primary visual cortex:**
  - The stimuli in a small region at the **center** of the visual space promotes neuronal activity while a broader concentric region (**surround**) has inhibitory effect

• **Visual features of center-surround difference are extracted for color, intensity and orientation**

Saliency

- **Intensity:**
  - \( I = \frac{r + g + b}{3} \),

- **Color:**
  - \( R = r - \frac{g + b}{2} \)
  - \( G = g - \frac{r + b}{2} \)
  - \( B = b - \frac{r + g}{2} \)
  - \( Y = \frac{r + g}{2} - \frac{|r - g|}{2} - b \) (yellow)

- **Orientation:**
  - Oriented Gabor pyramids with 9 scales and 4 orientations (0°, 45°, 90° and 135°)

Saliency

- Center-surround difference is implemented in the model as subtraction between fine and coarse scales of Gaussian pyramid (9 scales) for each type of feature:
  - Center is in scale $c \in \{2, 3, 4\}$
  - Surround is in scale $c + \delta$, $\delta \in \{3, 4\}$

- The resulting maps are normalized and summed into final saliency map

Saliency

**C:** Color, **I:** Intensity, **O:** Orientation center-surround differences  
**S:** Final saliency map

Saliency

Input images (a) and corresponding saliency maps (b)

Visual Attention [YPG01]

- Shading artifacts in “unattended” image regions are likely to remain unnoticed.
  - Use the visual attention model to decide the local quality of indirect lighting computation in RADIANCE
  - Consider bottom-up component only
    - Saliency Map [Itti’98]
  - Consider early vision path modeling
    - Error Tolerance Map
  - Speedup of irradiance caching: 3-9 times
  - Further speedup by reusing the indirect lighting for up to 10 in-between frames

Visual Attention [HMYS01]

Interactive Scenario: Shading artifacts of “unattended” glossy objects are likely to remain unnoticed

– Use visual attention models to schedule corrective computations for glossy objects that are most likely to be “attended”:
  • Consider both the saliency- and task-driven selection of those objects
– Use progressive rendering approach:
  • Hierarchical sample splatting in the image space
  • Cache samples and re-use them for similar views
– Use multiple processors to increase the sample number

[HMYS01] Haber et al.: Perceptually guided corrective splatting. CGF 20, Eurographics ’01, pp. 142–153
Visual Attention Processing

Saliency map

Open GL rendering  Corrective splatting  Converged solution

[HMYS01] Haber et al.: *Perceptually guided corrective splatting*. CGF 20, Eurographics ’01, pp. 142–153
Adaptive Splatting

Zoom-in → Warped old samples → New samples:

- level 1: 58 samples
- level 2: 188 samples

[HMY01] Haber et al.: Perceptually guided corrective splatting. CGF 20, Eurographics ’01, pp. 142–153
Guided search theory: Attention can be biased toward targets of interest which contribute to the task. [Wo94]

- **spatial biases**
  - some region of space more likely to contain relevant information
  - example: searching for fire-extinguisher in a scene biases to red color

- **feature biases**
  - bias by visual features associated with object of interest
  - example: eyes more likely to look on the road while driving

- **object-based and cognitive biases**
  - law of physics (gravity, friction, etc.)
  - example: focus on the floating load due to resulting danger

- **bias very probably ‘overrides’ bottom-up saliency**

Attention Models: Summary

We don’t perceive the world as it is.

Foveal vision is most sensitive to spatial detail and static contrast.

Peripheral vision is most sensitive to motion.

Differences in visual performance across the visual field can often be compensated by scaling the stimulus with projected eccentricity.

Directing gaze is a strong hint for selective attention.

Attention is a limited resources that must be shared across tasks.

Attention may amplify or attenuate visibility of a stimulus.

Low-level features increase saliency but may be outperformed by cognitive features such as scene knowledge and observer’s task.

Blurred line between bottom-up and top-down strategies.
Acknowledgements

• I would like to thank Sumant Pattanaik, Okan Tursun, Martin Weier, Michael Stengel, and Steve Grogorick for sharing with me some of their slides.