Visual Attention From a Graphics Point of View:

Computer Graphics Applications

Sumanta Pattanaik

Associate Professor

Department of Computer Science

University of Central Florida, Orlando, FL, USA

sumant@cs.ucf.edu

Visual Attention

Visual attention is the process of selectively concentrating on a discrete aspect of visual information while ignoring others.

Source: The Quest for Consciousness: : A Neurobiological Approach. Koch 2004

Visual Attention

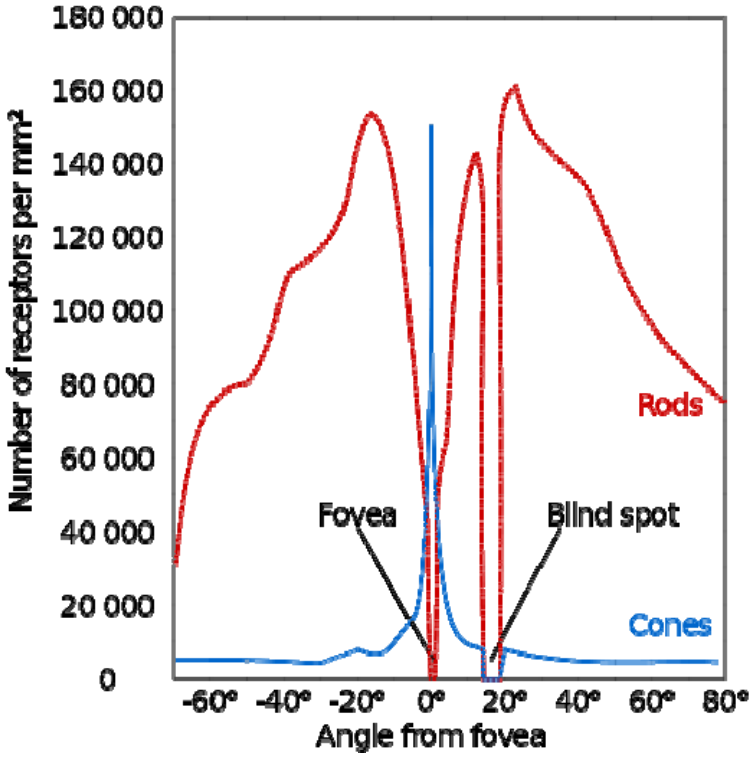
Visual attention is the process of selectively concentrating on a discrete aspect of visual information while ignoring others.

"…. is a mechanism that reduces the effect of processing bottleneck in visual cortex, limiting the computation to the smaller subregion highlighted by the attentional focus."

Christof Koch

Source: The Quest for Consciousness: : A Neurobiological Approach. Koch 2004

Distribution of Receptor Cells in Retina

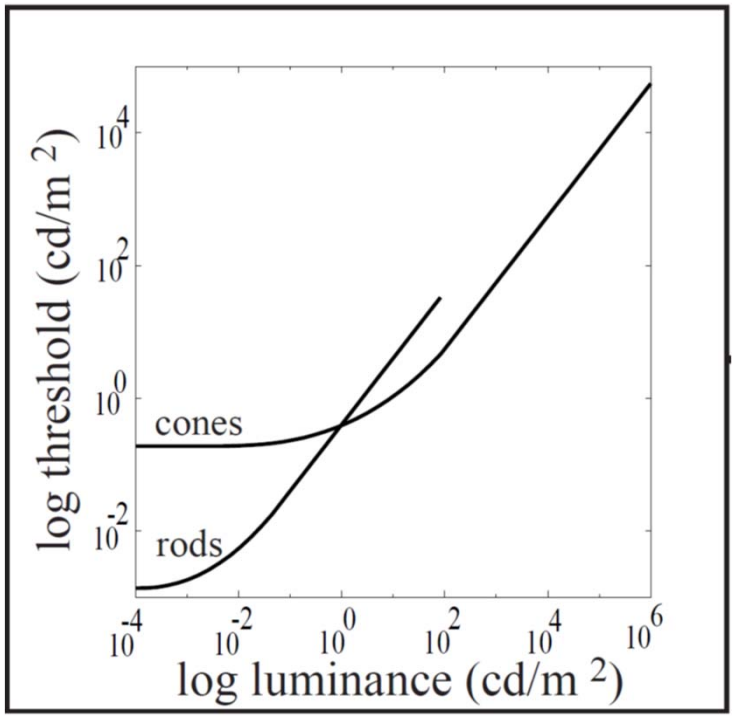


Highly nonuniform distribution of photoreceptors on the human retina: Only a small region of 2 – 5 degres of visual angle (the fovea) around the center of gaze is captured at high resolution, with logarithmic resolution falloff with eccentricity

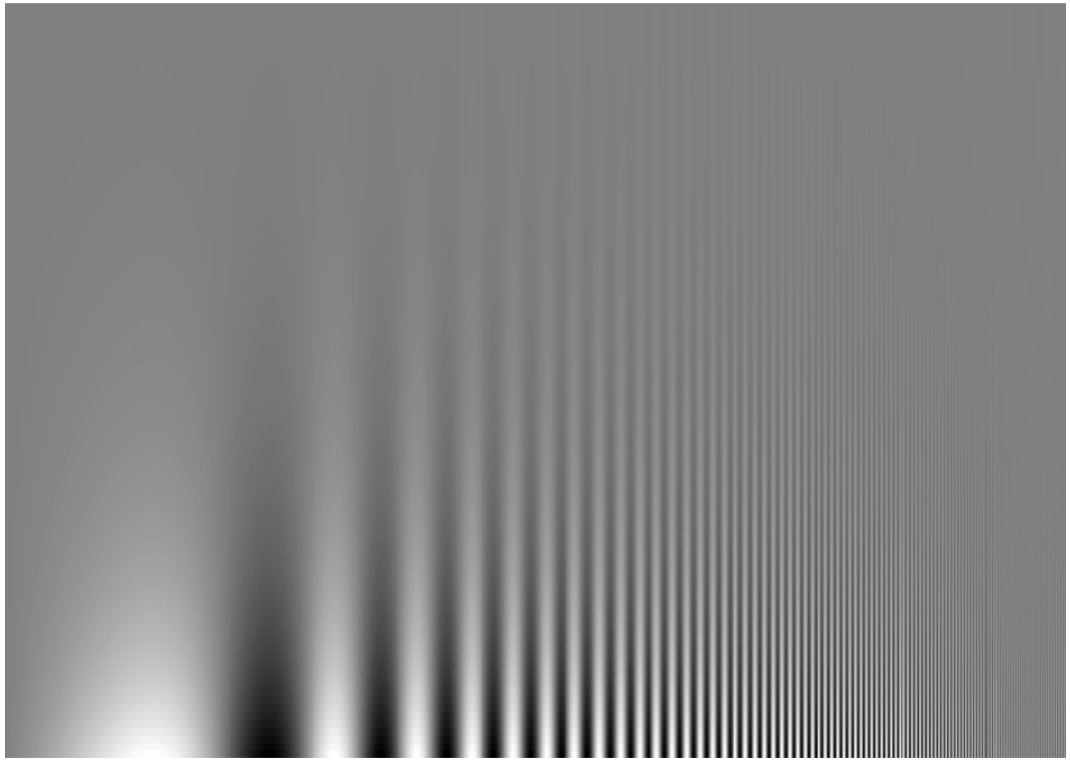
Human eye has a Wide Field of View (114°), but the clarity of the vision is not the same.

- There is a strong fall-off of visual acuity from the center of the fovea to the periphery.

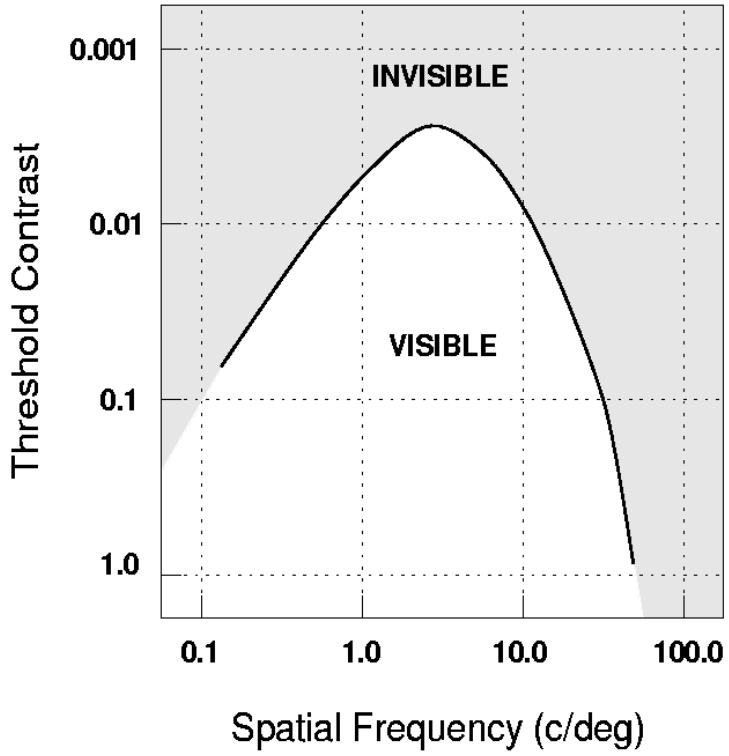
Visual Limits: Weber Law



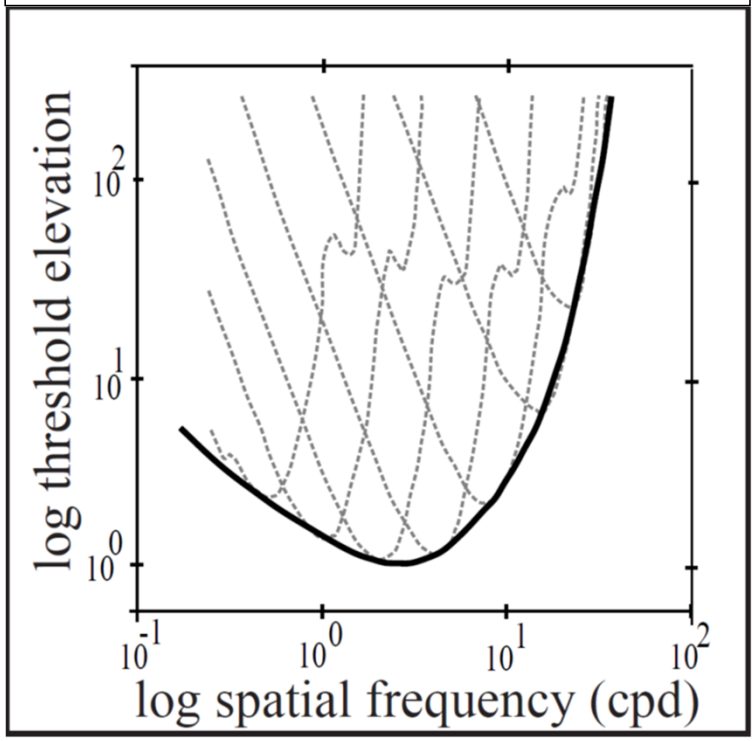
Visual Limits: Contrast sensitivity function



Visual Limits: Contrast Sensitivity Function (CSF)

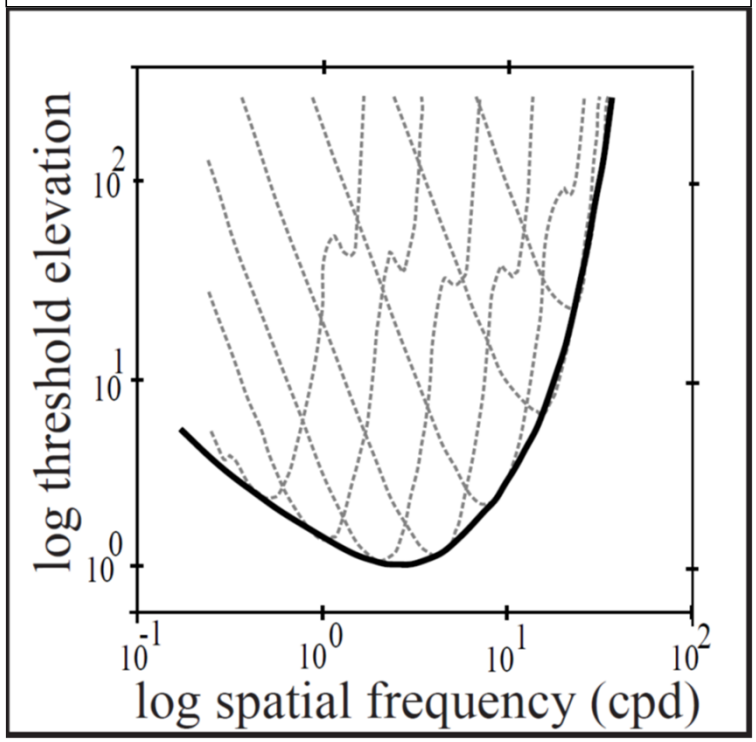


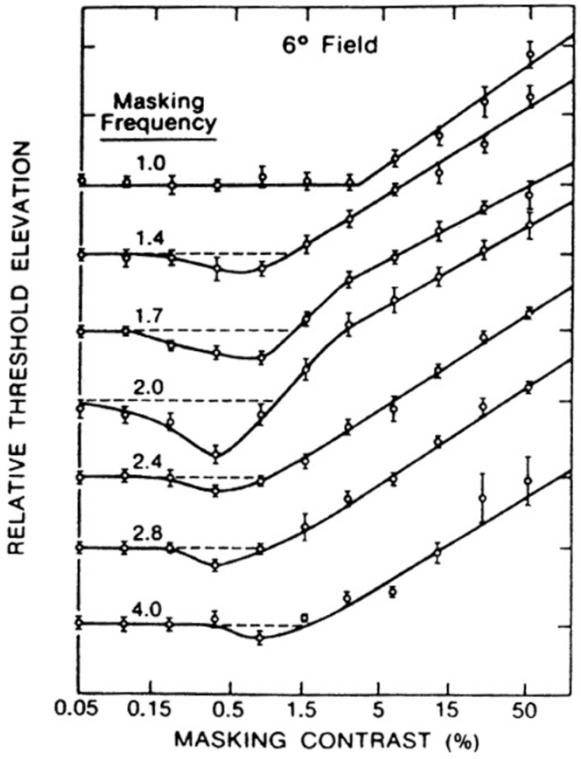
Frequency Dependent Threshold Elevation



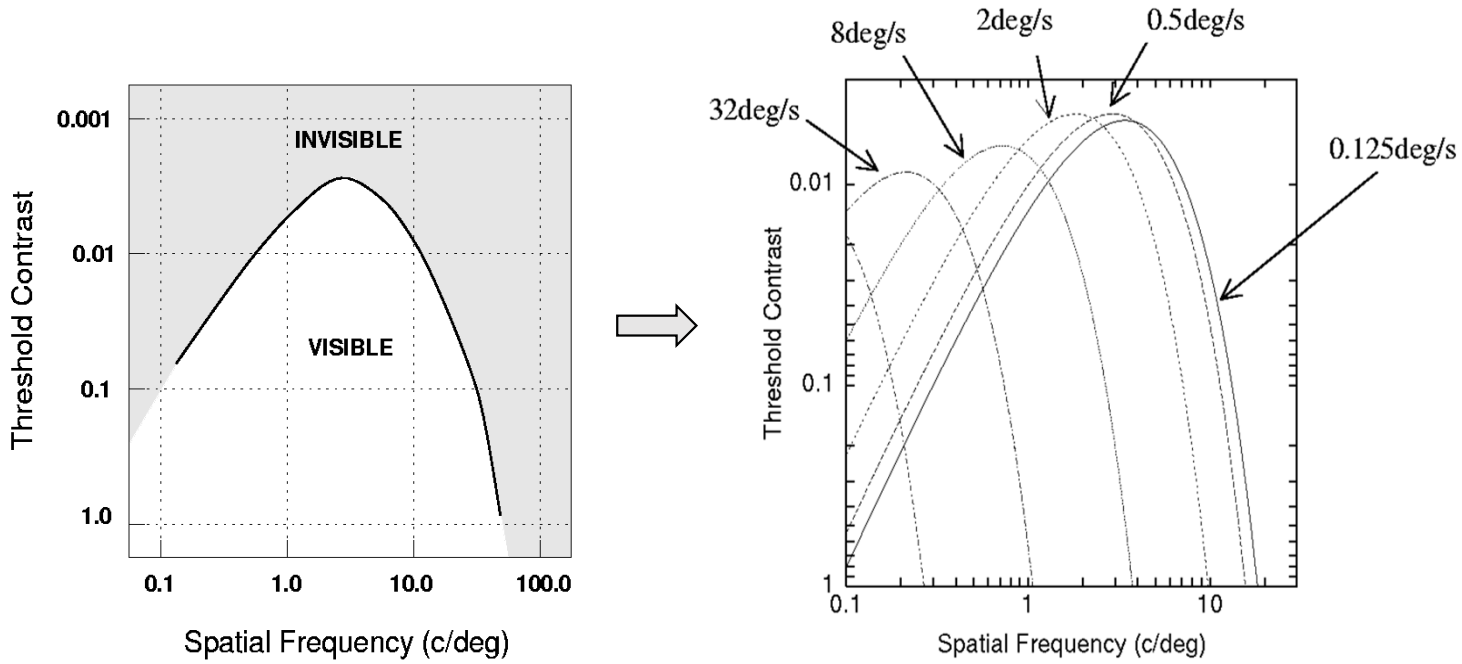
Visual Limits: Masking

Frequency Dependent Threshold Elevation





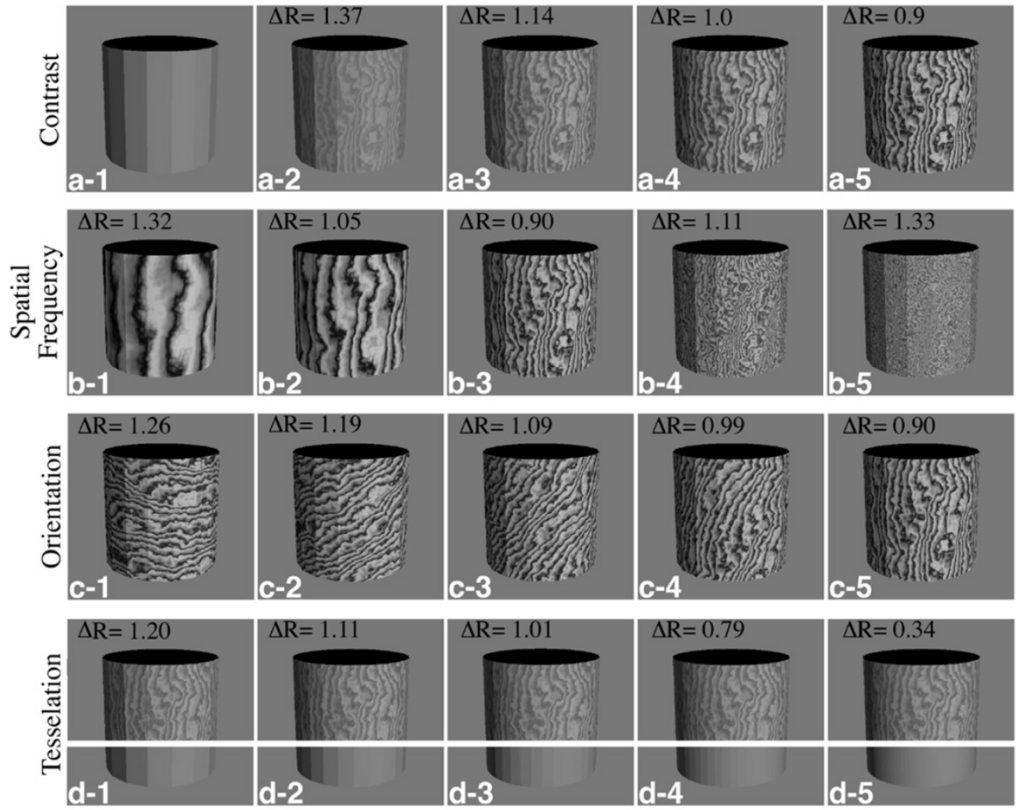
Temporal loss of Sensitivity



Attention Aware CG Application

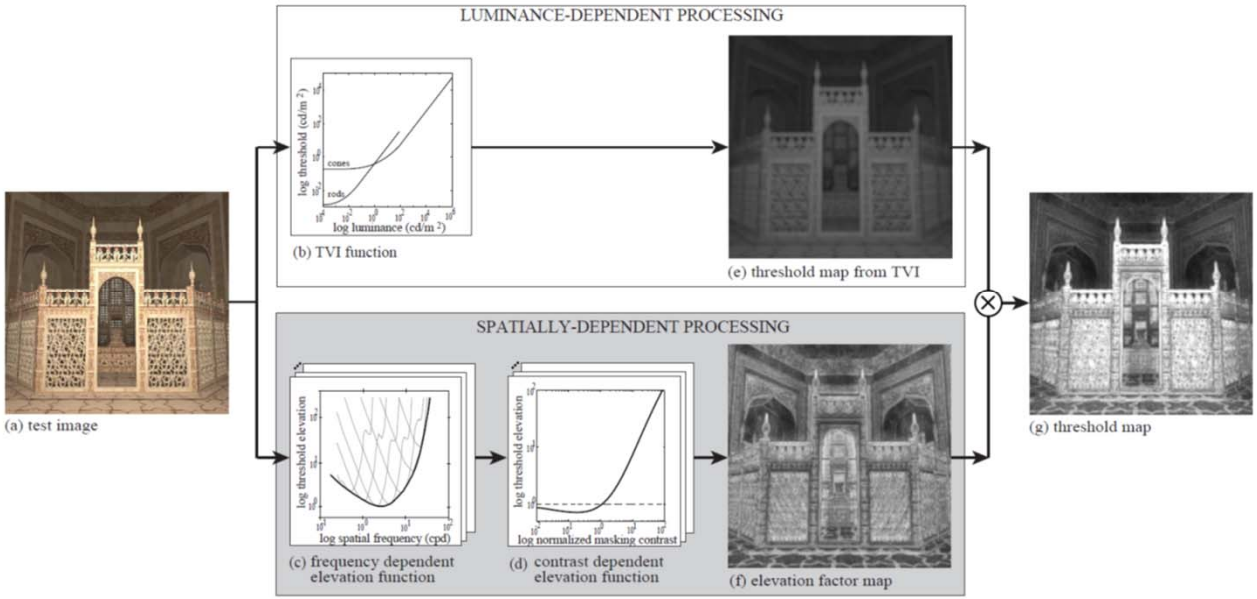
- Attention Based Rendering Optimization
  - by rendering unattended image regions less accurately
  - by using less accurate representations of the geometry (LOD)
- Depth-of-field simulations.
- Image/video Compression
- Game level setting
- Gaze as pointing device
- Attention directing

Applying Masking Model



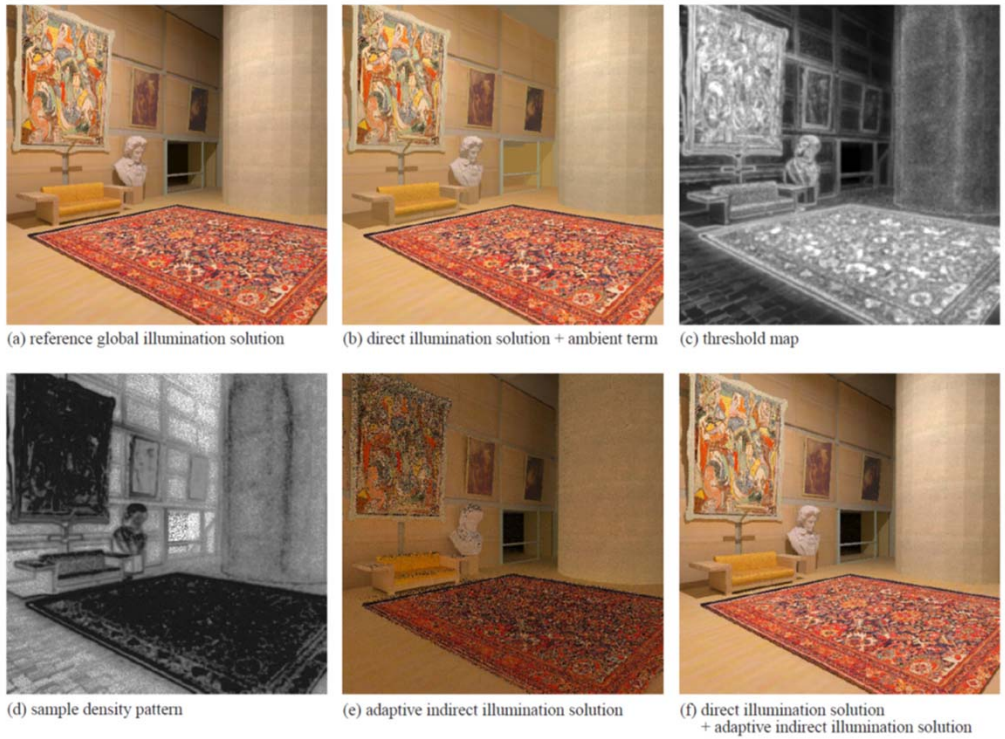
Source: “A Model of Visual Masking for Computer Graphics”, Ferwerda, 1997.

Perceptually Based Physical Error Metric for Realistic Image Synthesis



Source: “A Perceptually Based Physical Error Metric for Realistic Image Synthesis”. Ramasubramanian, 1999.

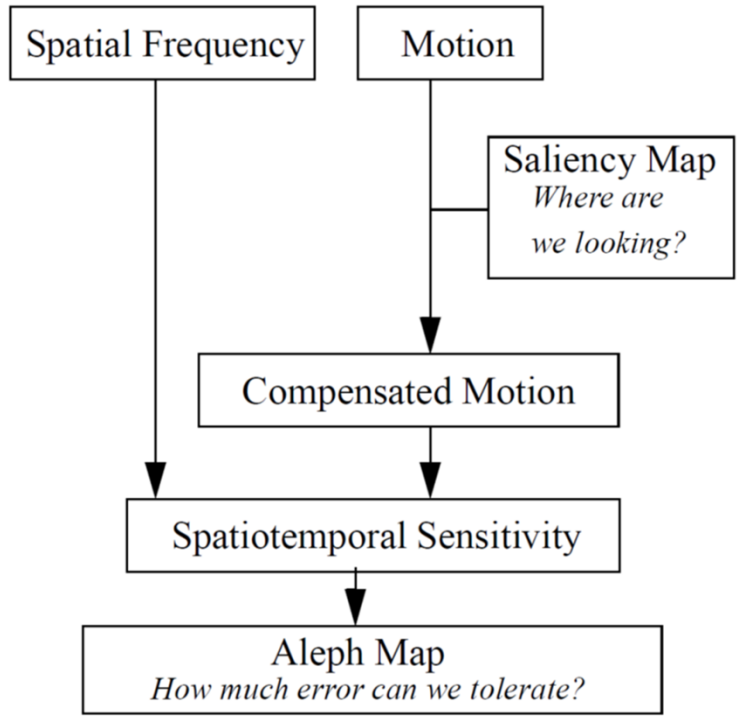
Perception Driven Adaptive Rendering



(a) reference global illumination solution
(b) direct illumination solution + ambient term
(c) threshold map
(d) sample density pattern
(e) adaptive indirect illumination solution
(f) direct illumination solution + adaptive indirect illumination solution

Source: “A Perceptually Based Physical Error Metric for Realistic Image Synthesis”. Ramasubramanian, 1999.

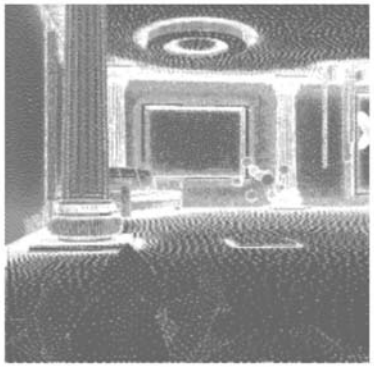
Saliency Driven Error Tolerance Measure



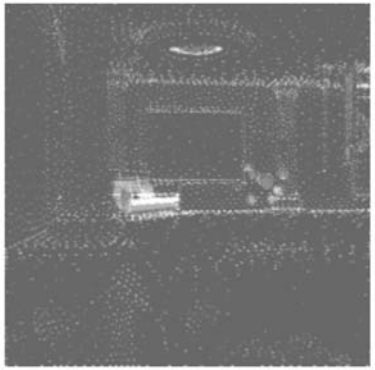
Source: “Spatio-Temporal Sensitivity and Visual Attention in Dynamic Environments”. Yee, 2001.

Saliency Driven Adaptive Rendering

Reference Sampling Density



Aleph Map Assisted Sampling Density

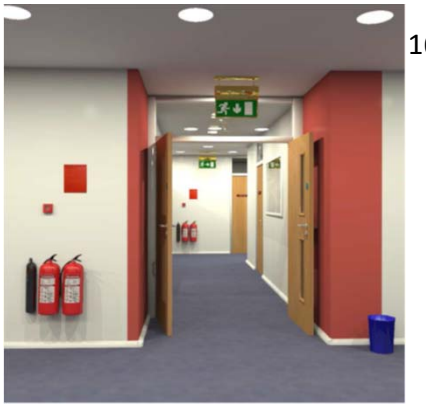


Reference

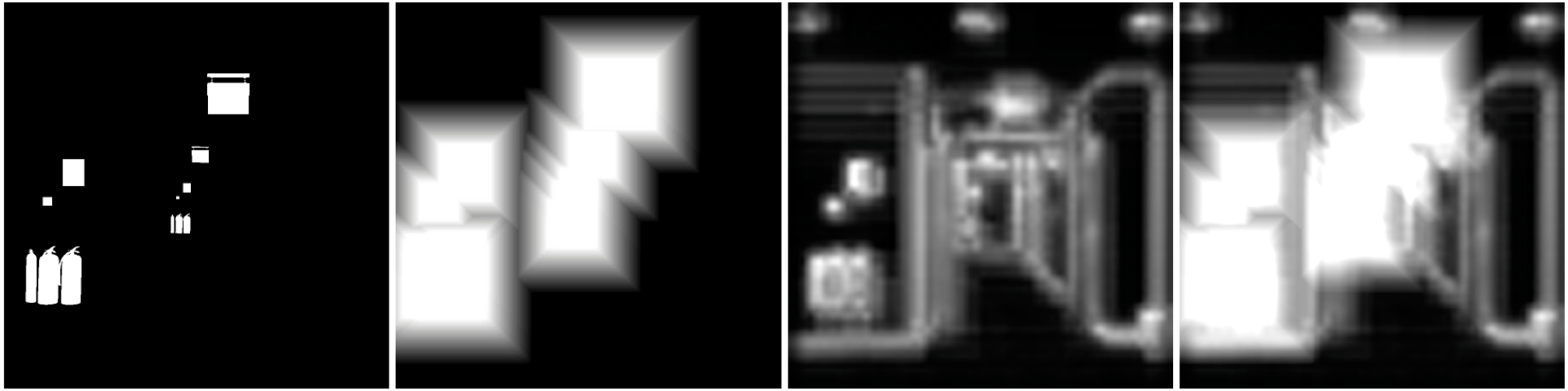
Aleph Map Accelerated Solution

Source: “Spatio-Temporal Sensitivity and Visual Attention in Dynamic Environments”. Yee, 2001.

Saliency Driven Adaptive Rendering



- Selective Guidance: a combination of saliency and task relevance to direct the rendering computation

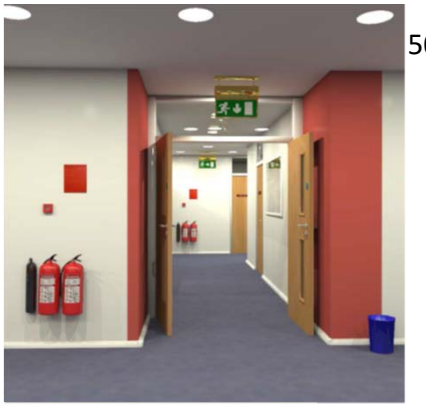


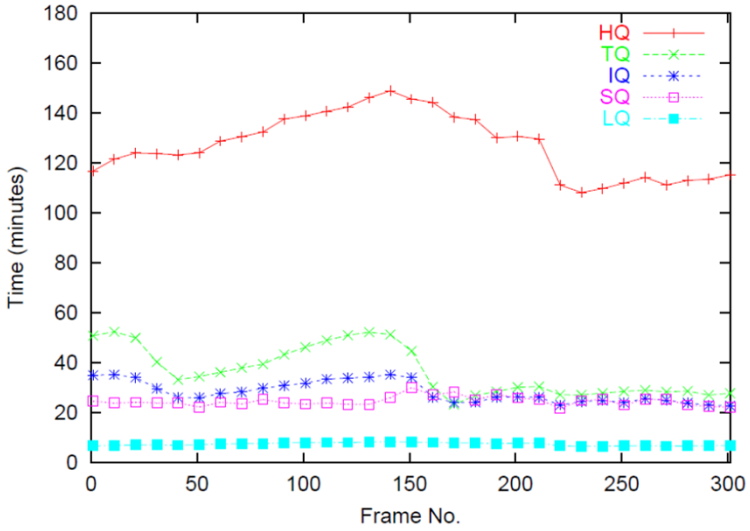
Importance map for the corridor scene : (a) task objects, (b) task map(c) saliency map and (d) combined task and saliency map

Task: Counting the number of fire safety items in the scene.

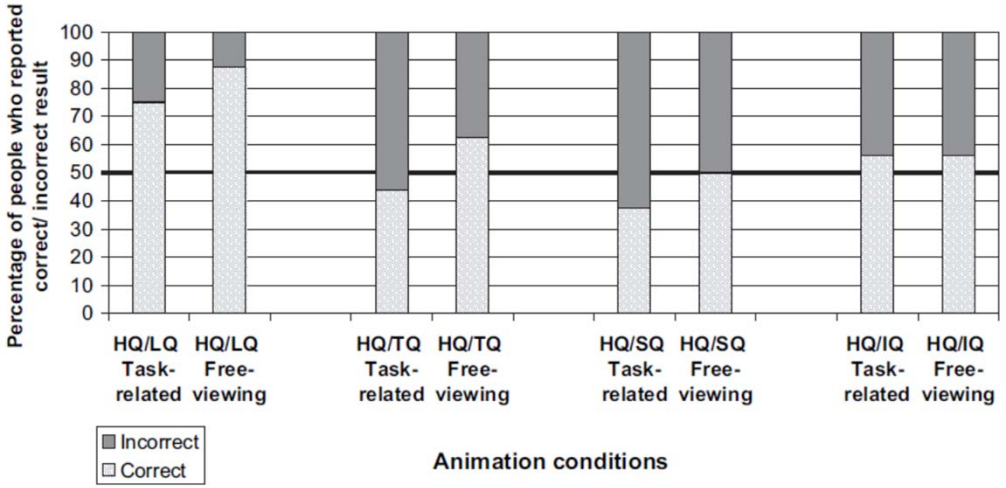
Source: “Visual Attention for Efficient High-Fidelity Graphics”. Sundstedt, 2005.

Saliency Driven Adaptive Rendering





HQ	High Quality
LQ	Low Quality
TQ	Task Quality
IQ	Importance Quality
SQ	Saliency Quality

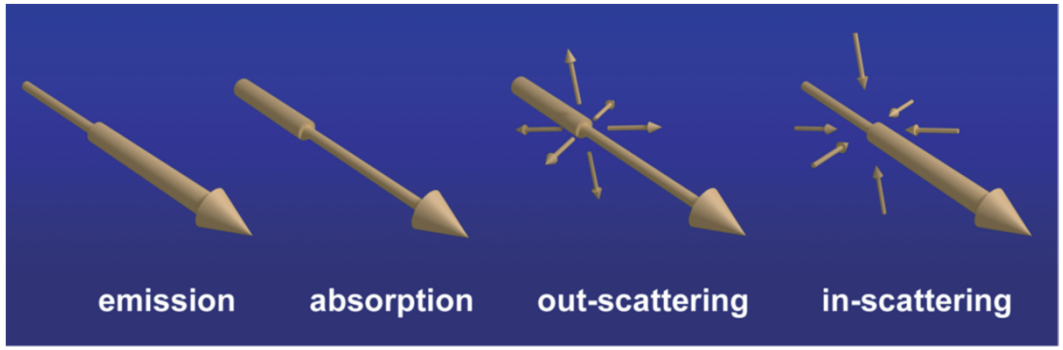


Q: which of the two animations the participants thought contained the worse rendering quality. The participants performed a two-alternative forced-choice (2AFC) task.

Source: “Visual Attention for Efficient High-Fidelity Graphics”. Sundstedt, 2005.

Saliency Driven Adaptive Rendering

- Efficient Rendering for Participating Media
  - computational requirements for calculating the interaction of light with such participating media are substantial.



$$L(x,\vec{\omega}) = e^{-\tau(x_0,x)}L(x_0,\vec{\omega}) + \int_{x_0}^{x} e^{-\tau(x',x)}\alpha(x')L_e(x',\vec{\omega})dx' + \int_{x_0}^{x} e^{-\tau(x',x)}\sigma(x')\int_{\Omega} p(x',\vec{\omega}',\vec{\omega})L(x',\vec{\omega}')d\vec{\omega}'dx'$$

Source: National Geographic

Where optical pathlength $\tau(x',x) = \int_{x'}^{x} \kappa(s)ds$

Source: “Perceptual Rendering of Participating Media”, Sundstedt 2007

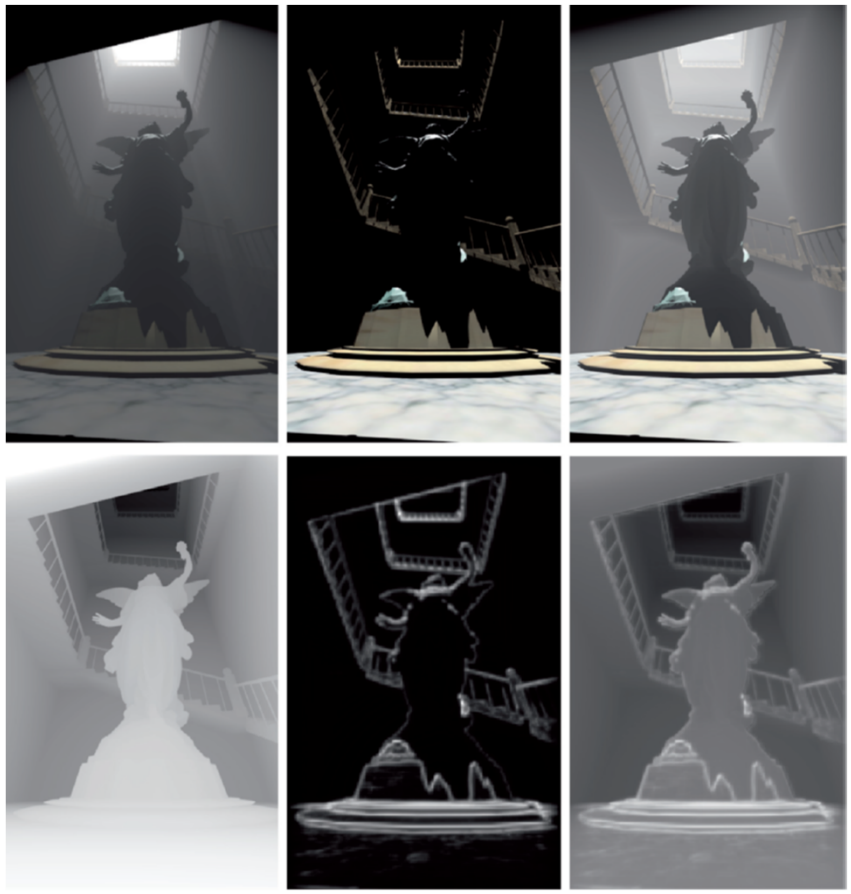
Saliency Driven Adaptive Rendering

- Efficient Rendering for Participating Media
  - computational requirements for calculating the interaction of light with such participating media are substantial.
  - Extinction (X) Map: Stores in image-space the exponential decay of light in the medium. ($e^{-\tau(x',x)}$)
  - XS map: X map in combination with Saliency (S) map is used to guide a perception-based renderer for inhomogeneous participating media.
  - The maximum number of rays per pixel is input as a user-defined parameter. For each pixel the number of rays is calculated by modulating the maximum value with XS map value at that pixel.

Source: National Geographic

XS Map

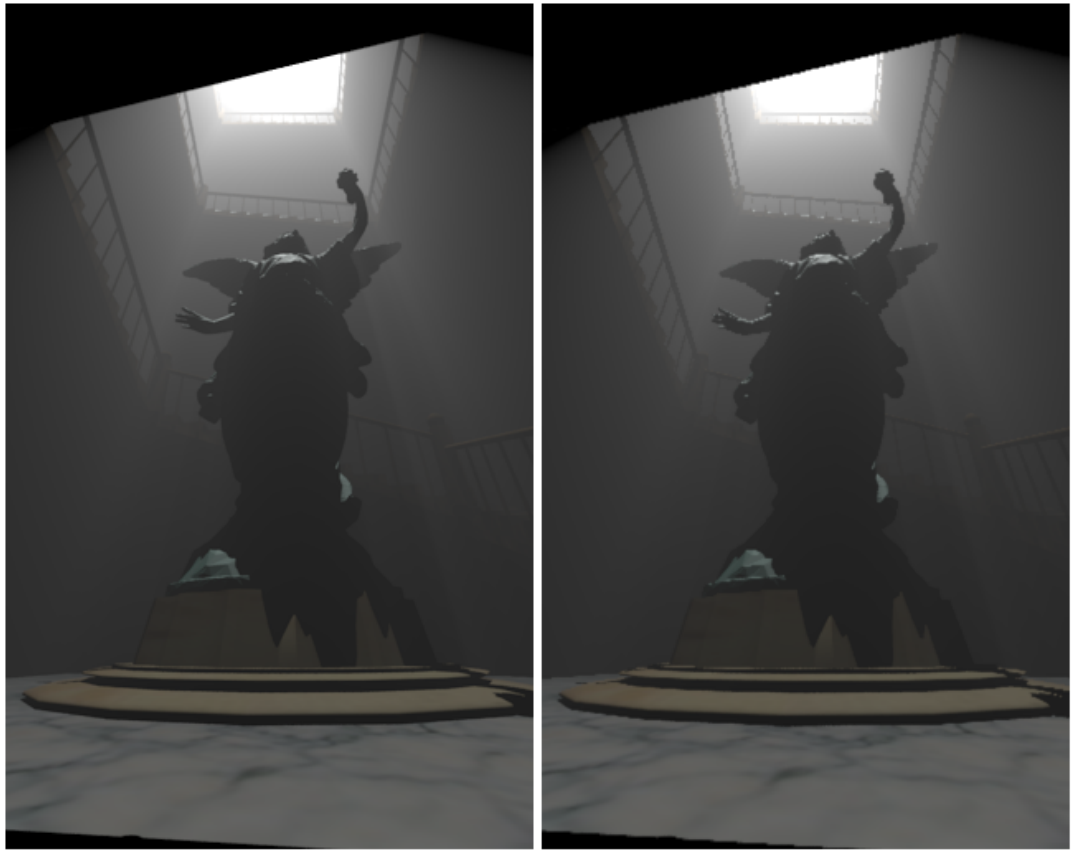
- Top: (left) Full Solution, (middle) initial fast ray-traced snapshot, and (right) a false estimate of the medium
- Bottom: (left) the X map, (middle) the S map, and (right) the map.



Source: “Efficient Selective Rendering of Participating Media”. Anson 2006.

Efficient Rendering for Participating Media

- Speedup is scene dependent.
- 10x for the example scene.

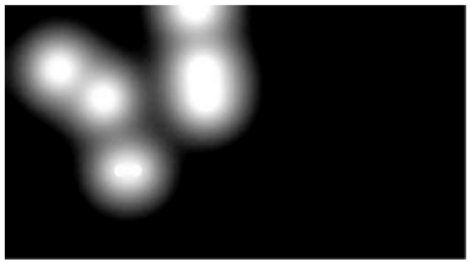


Source: “Efficient Selective Rendering of Participating Media”. Anson 2006.

Efficient Rendering for Participating Media

- Light Map: Apply point spread functions to the light source points (multi-Gaussian filtering).

$$L_n = \begin{cases} L_0 & i = 0 \\ L_{i-1} + \left(L_{i-1} * G(\sigma_{base} i)\right) & i > 0 \end{cases}$$



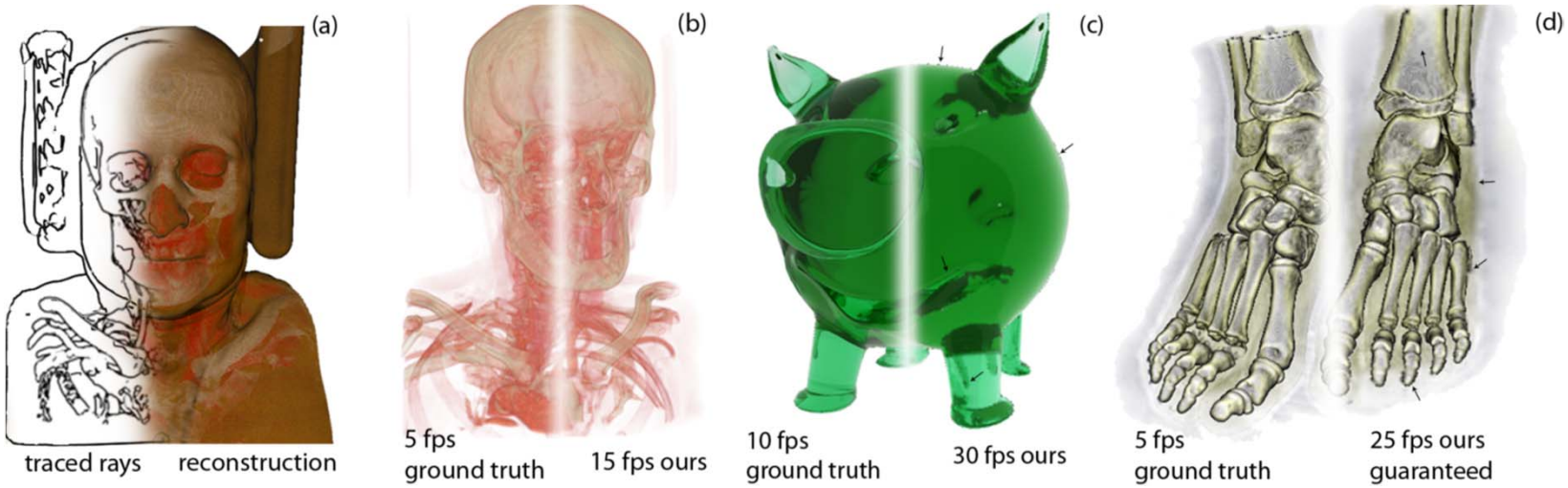
Source: “Perceptual Rendering of Participating Media”, Sundstedt 2007

Efficient Rendering for Participating Media

Source: “Perceptual Rendering of Participating Media”, Sundstedt 2007

Saliency Driven Adaptive Rendering

- Saliency based Ray Prioritization for guaranteed frame rate:



Source: “Ray prioritization using stylization and visual saliency” Steinberger, 2012.

Saliency Driven Adaptive Rendering

Ray Prioritization

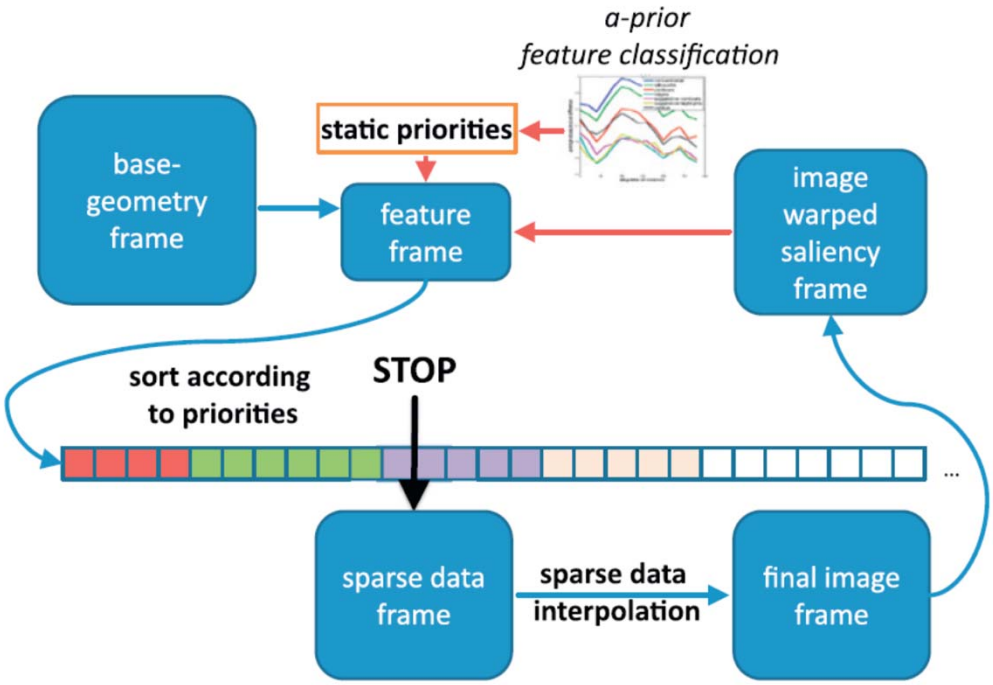
Overview:

- Every ray's priority is computed using feature priority map
- Rays are chosen from a 1D ray priority queue and traced till time budget is reached.

More rays are sent into the scene in feature-rich areas and their vicinity, while the sampling frequency for feature-poor areas is strongly decreased.

Feature priority from the feature buffer:

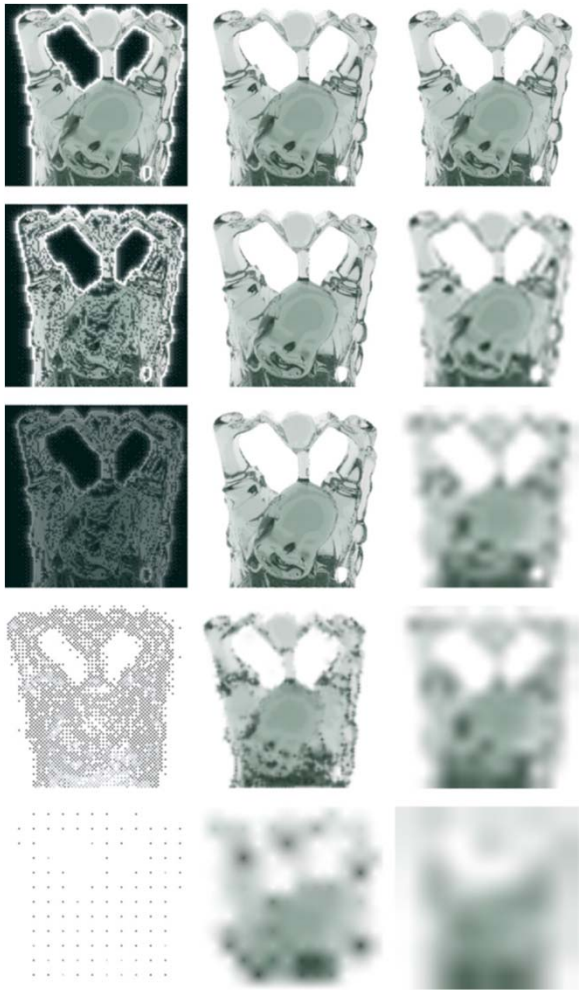
- Object space features: such as silhouettes, suggestive contours, ridges and valleys, are extracted from the data set's corresponding meshes, projected to screen space.
- Image space features: are deduced by performing a saliency analysis of the previously rendered frame, warping the saliency to the current frame, and analyzing the result for disocclusions.



Source: "Ray prioritization using stylization and visual saliency" Steinberger, 2012.

Saliency Driven Adaptive Rendering

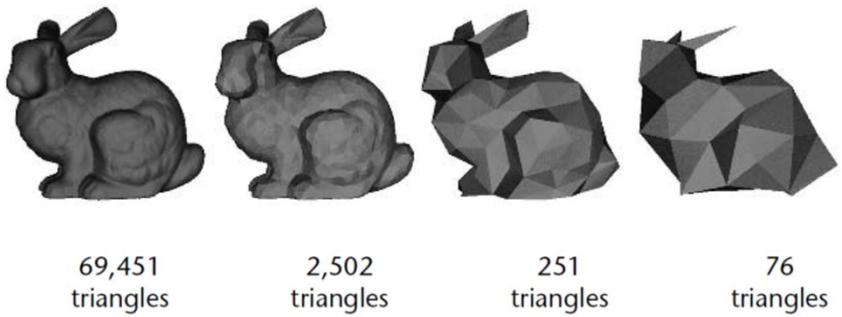
Frame reconstruction from Sparse Samples



Source: “Ray prioritization using stylization and visual saliency” Steinberger, 2012.

Computational Acceleration: Attention Guided Level of Detail

- Geometric Simplification (LOD):
  - Complexity of our 3D models can be higher than the ability of our hardware to render them
  - generate and render less detailed representations for small, distant, or unimportant portions of the scene.



Source: Level of details for 3D Graphics. Leubke, 2003.

Choice of LOD

- Eccentricity:
  - An object's LOD is chosen based on its angular distance from the center of the user's gaze.
- Velocity:
  - (if we ignored smooth pursuit) simplifying objects moving quickly across the user's gaze more aggressively than slow-moving or still objects.
- Depth of Field:
  - simplifying more aggressively those objects that are visually blurred because the user's gaze is focused at a different depth.

Source: Level of details for 3D Graphics. Leubke, 2003.

LOD Selection as an Optimization problem

- Framework for LOD Selection: (Funkhouser, Sequin 1993)

  Given object tuples defined as (O; L; R)

  where O: Object, L: Level of Detail, R: Rendering algorithm,

  and S the set of object tuples rendered in each frame,

  choose a level of detail and rendering algorithm for each potentially visible object so as to:

  Maximize: $\sum_S benefit(\text{O}; L; R)$

  Subject to: $\sum_S cost(\text{O}; L; R) \leq$ TargetFrameTime

Source: “Adaptive Display Algorithm for Interactive Frame Rates During Visualization of Complex Virtual Environments”. Funkhouser 1993.

LOD Selection as an Optimization problem

- Cost Heuristics:

$$\text{cost}(\text{O};\text{L};\text{R}) = \max\begin{cases} c_1 Poly(\text{O};L) + c_2 vert(\text{O};L) \\ c_3 pix(\text{O}) \end{cases}$$

Source: “Adaptive Display Algorithm for Interactive Frame Rates During Visualization of Complex Virtual Environments”. Funkhouser 1993.

LOD Selection as an Optimization problem

- Cost Heuristics:

$$\text{cost}(\text{O};\text{L};\text{R}) = \max\begin{cases} c_1 Poly(\text{O};L) + c_2 vert(\text{O};L) \\ c_3 pix(\text{O}) \end{cases}$$

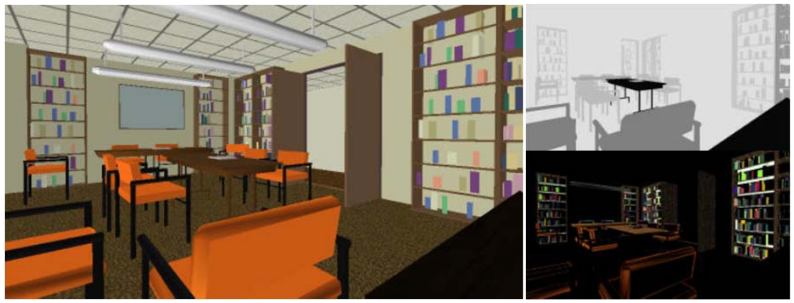
- Benefit Heuristics: an estimate of the 'contribution to model perception' of rendering object O with level of detail L and rendering algorithm R
  - $\text{Benefit}(\text{O};\text{L};\text{R}) = \text{Size(O)} \times \text{Accuracy(O; L;R)} \times \text{Importance(O)} \times \text{Focus(O)} \times \text{Motion (O)} \times \text{Hysteresis (O; L;R)}$

Source: “Adaptive Display Algorithm for Interactive Frame Rates During Visualization of Complex Virtual Environments”. Funkhouser 1993.

LOD Selection Examples

No detail elision (~20K polygon)

0.15 Second (4K polygon)



0.10 Second (< 1.5K polygon)

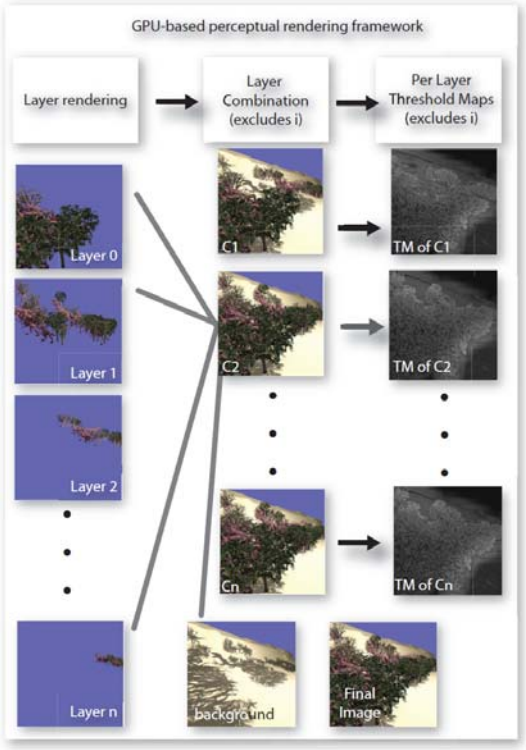
Optimization detail elision algorithm

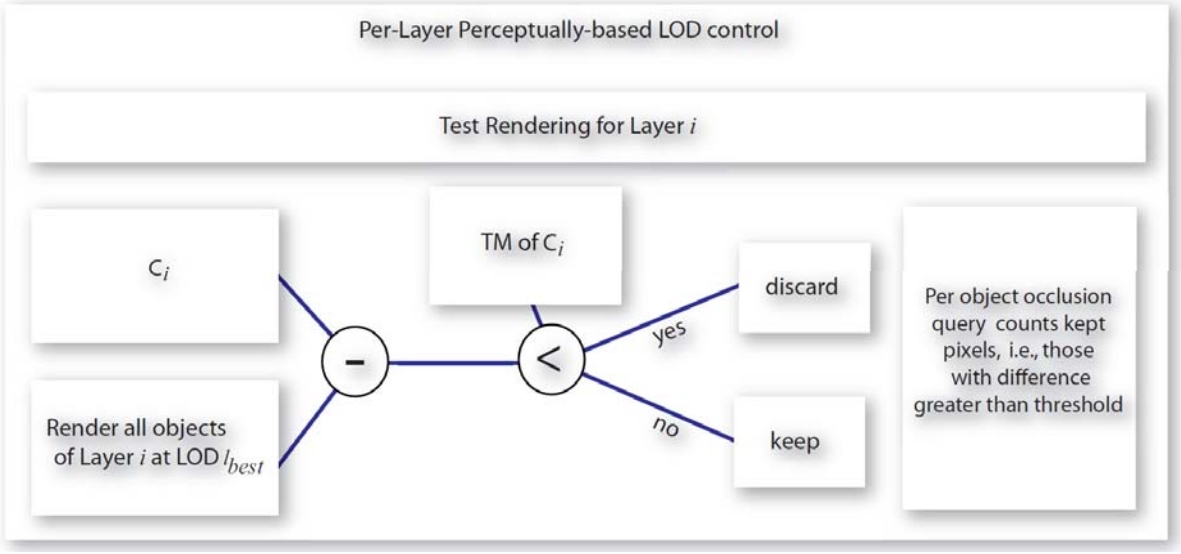
Source: “Adaptive Display Algorithm for Interactive Frame Rates During Visualization of Complex Virtual Environments”. Funkhouser 1993.

Interactive Perceptual Layered Rendering

Source: “An Interactive Perceptual Rendering Pipeline using Contrast and Spatial Masking”, Drettakis, 2007.

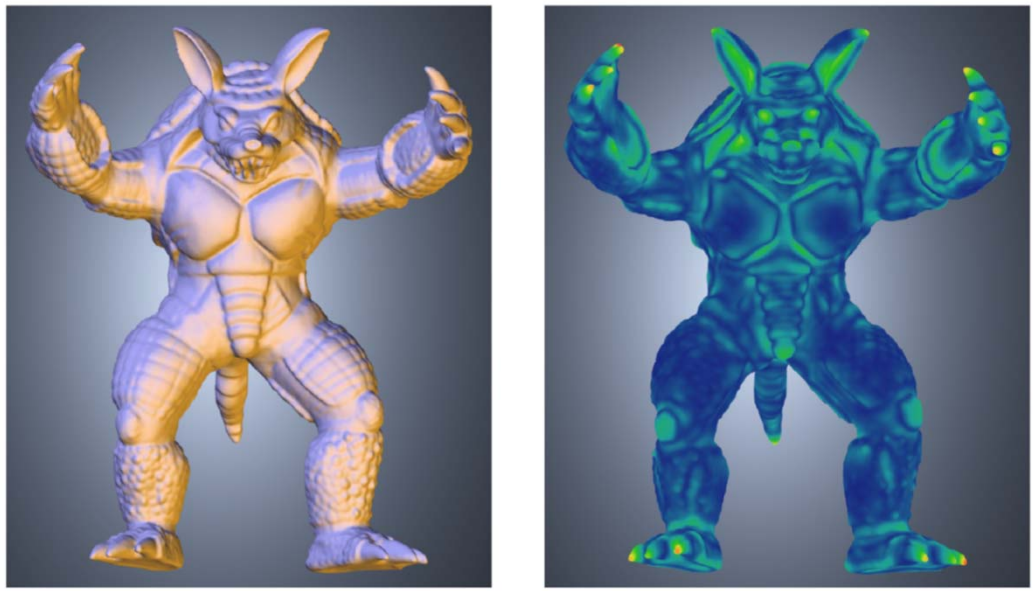
Interactive Perceptual Layered Rendering





Source: "An Interactive Perceptual Rendering Pipeline using Contrast and Spatial Masking", Drettakis, 2007.

Geometry Simplification



Source: “Mesh Saliency”, Lee 2005.

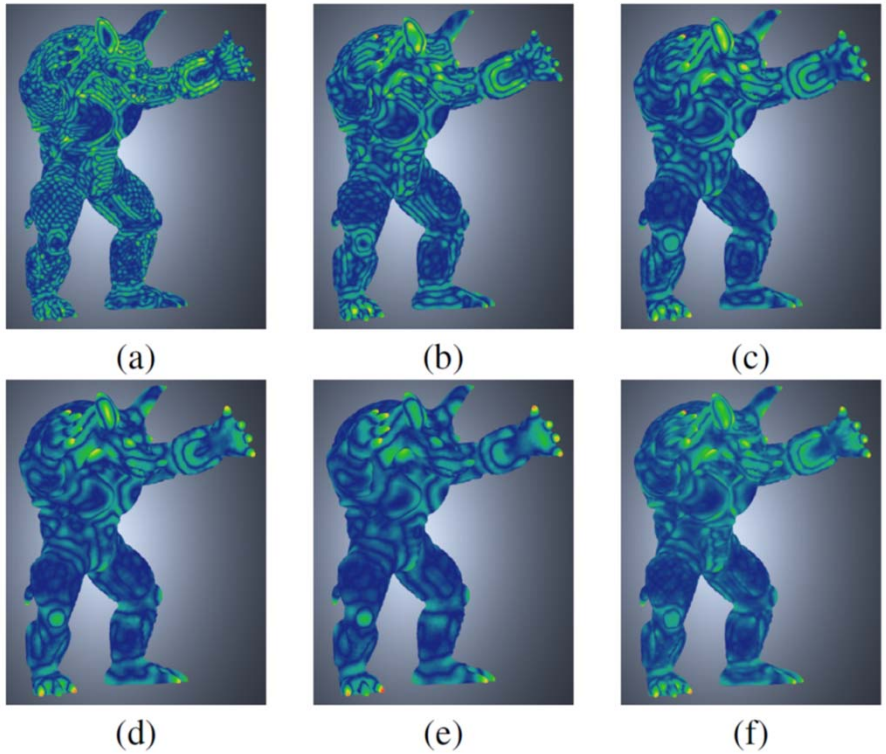
Mesh Saliency at Multiple Scales

- Mesh Saliency at each scale is computed as difference of Gaussian

$$G(\mathscr{C}(v),\sigma) = \frac{\sum_{x \in N(v,2\sigma)} \mathscr{C}(x) exp[-\|x-v\|^2/(2\sigma^2)]}{\sum_{x \in N(v,2\sigma)} exp[-\|x-v\|^2/(2\sigma^2)]}$$

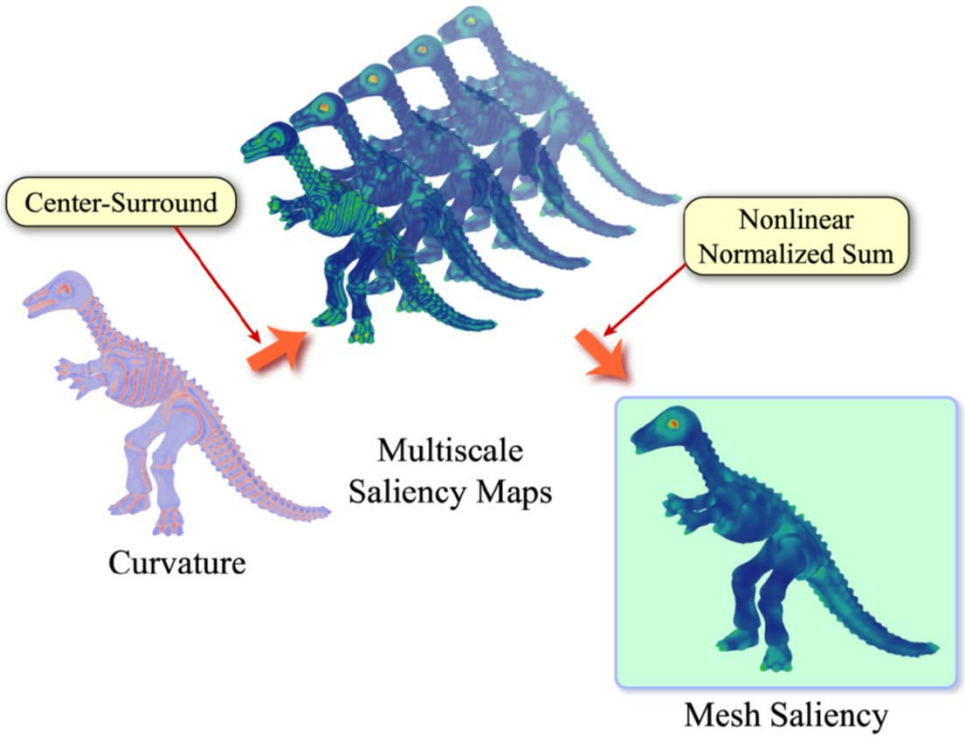
$$\mathscr{S}(v) = |G(\mathscr{C}(v),\sigma) - G(\mathscr{C}(v),2\sigma)|$$

- The saliency at scales of 2e, 3e, 4e, 5e, 6e, and aggregate saliency.
  - e is 0.3% of the length of the diagonal of the bounding box of the model.



Source: “Mesh Saliency”, Lee 2005.

Gaussian Weighted Average of Mean Curvature



Source: “Mesh Saliency”, Lee 2005.

Salient Simplification

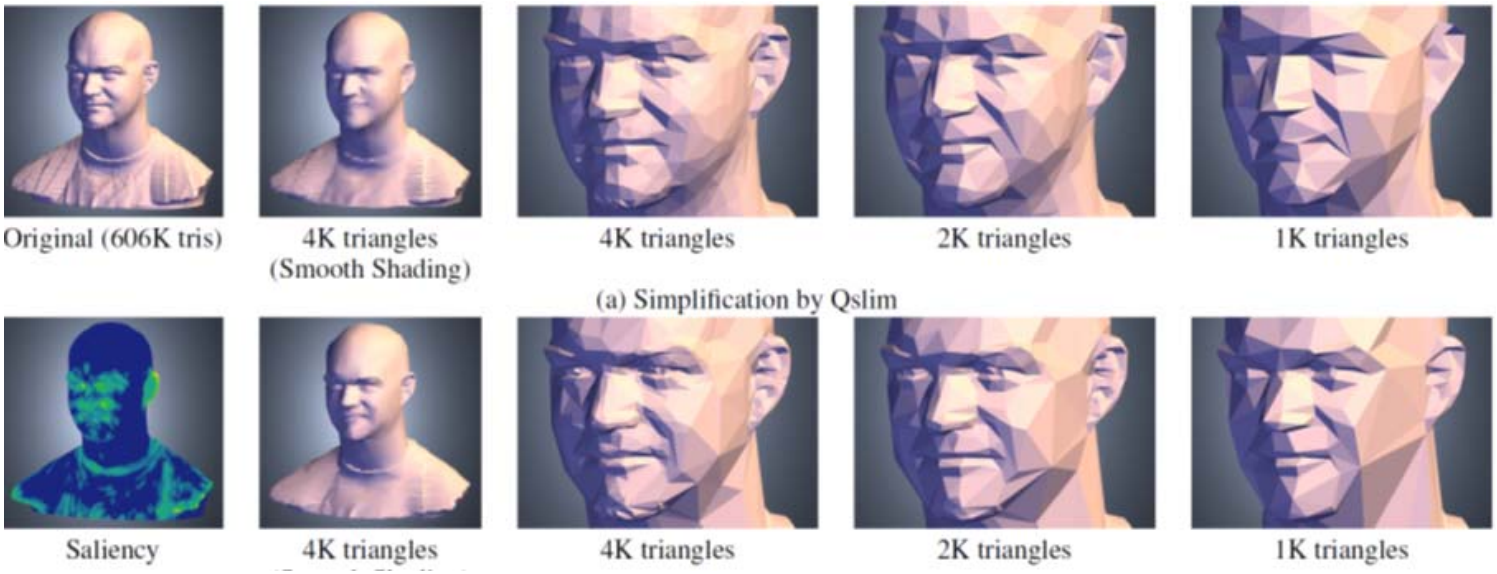
- Mesh saliency is used to modify an existing mesh simplification method.
  - Garland and Heckbert's (1997) QSlim method simplifies a mesh by repeatedly contracting vertex. Pairs ordered by increasing quadric errors.
    - Quadric of a plane P: $Q_P = pp^T$
    - Quadric of a mesh vertex v: $Q_v = \sum Q_P$
    - Optimal contraction $\bar{v}$ of vertex pair v_i , v_j minimizes :$\bar{v}^T(Q_{\bar{v}})\bar{v}$, and where $Q_{\bar{v}} = Q_{v_i} + Q_{v_j}$ is the quadric of $\bar{v}$.
  - In the modification Q_v is weighted by a saliency dependent weight W_v and the weight of contraction vertex $W_{\bar{v}} = W_{v_i} + W_{v_j}$

    Where $W_v = \begin{cases} \alpha S_v & if S_v > \beta \\ S_v & Otherwise \end{cases}$

    α = 100, β is 30th percentile Saliency

Source: “Mesh Saliency”, Lee 2005.

Salient Simplification



Source: “Mesh Saliency”, Lee 2005.

Computational Acceleration: Attention Guided Material Simplification

Subsurface light transport requires intense analytical calculations, making it expensive for mobile devices to render this effect. So are refraction and displacement mapping.

- Example material LODs:

Source: “C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics”. Koulieris, 2014.

Gaze-driven Depth-of-Field Blur

Focus range, or range of distances near the point of focus where objects appear sharp. Objects behind and in front of the point of focus are blurred.

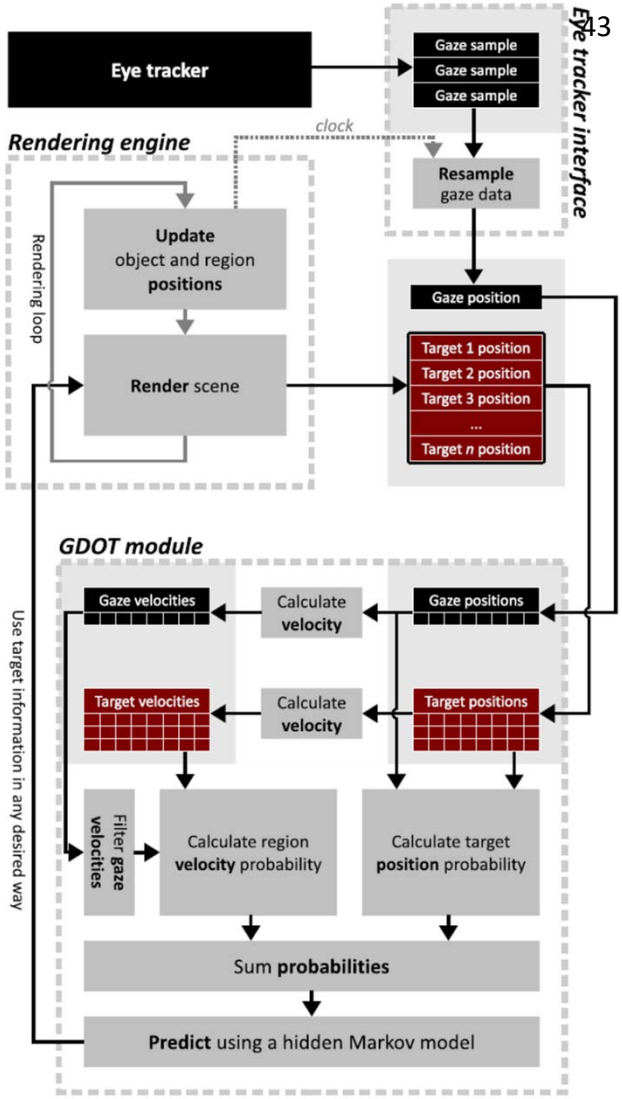
Source: Wiki

Gaze-driven Depth-of-Field Blur

- Availability of eye tracking devices make the problem seem simpler.
- Goal:
  - Eye tracking gives point(s) of focus on the screen (F_screen).
  - What we need is: F_d, the focal distance of the camera lens/eye.
- Problem:
  - Accuracy of the eye tracking device
  - Eye movement

Gaze-driven Depth-of-Field Blur

- [Mantiuk 2013] provide a probabilistic gaze-driven object racking algorithm that combines the information from the eye-tracker and the 3D scene to improve the accuracy of eye-tracking.
- Utilize the algorithm to simulate a dynamic and high quality DoF effect.



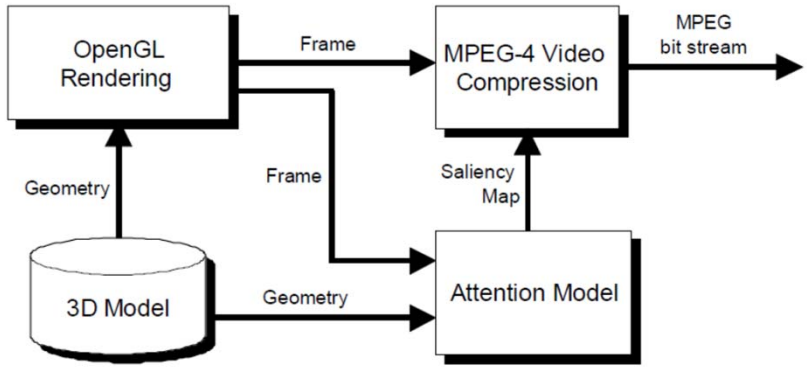
Source: “Gaze-driven Object Tracking for Real Time Rendering”. Mantiuk, 2013.

Reduce size of Compressed Video Streams

- Idea:
  - highly nonuniform distribution of photoreceptors on the human retina: Only a small region of 2 – 5 degres of visual angle (the fovea) around the center of gaze is captured at high resolution, with logarithmic resolution falloff with eccentricity
  - So, select a small number of interesting regions in each frame and to encode them in priority. Tolerate degradation outside priority region for increased compression It may decrease overall compressed file size.

Attention Guided Video Compression

- MPEG Compression of Animation:
- Saliency map is a bitmap of the resolution adjusted to the number of MPEG macroblocks
- saliency map – is used to choose the quantizer scale for each macroblock.
- Achieved a reduction of bit-stream size between 10% and 19%.



Source: “Attention Guided MPEG Compression for Computer Animations”, Mantiuk 2003.

Attention Guided Video Compression

- A dynamic saliency map is computed, over the entire duration of each video clip.
- Algorithms

  1: a snapshot of the saliency map at each frame directly determines the priority to be given to every spatial location in the frame, after normalization by a squashing function and temporal smoothing.

  2: a small number of discrete virtual foveas endowed with mass/spring /friction attempt to track a collection of most salient objects, with each fovea modeled as a unit mass at one end of a spring with zero rest length, and a salient location serving as an anchor point at the other end.
- A spatially-variable blur is applied (before standard video compression) to the input frames such that lower-priority regions are more strongly blurred.

Source: “Visual attention guided bit allocation in video compression”, Lee 2011.

Miscellaneous Applications

Attention Driven Game Design

- Attention plays a critical role during game playing.
- Based on a model of attention allocation, we can make a game harder by placing important game-relevant items in regions with less attention or we can make a game easier by placing important game-relevant items in regions with more attention.

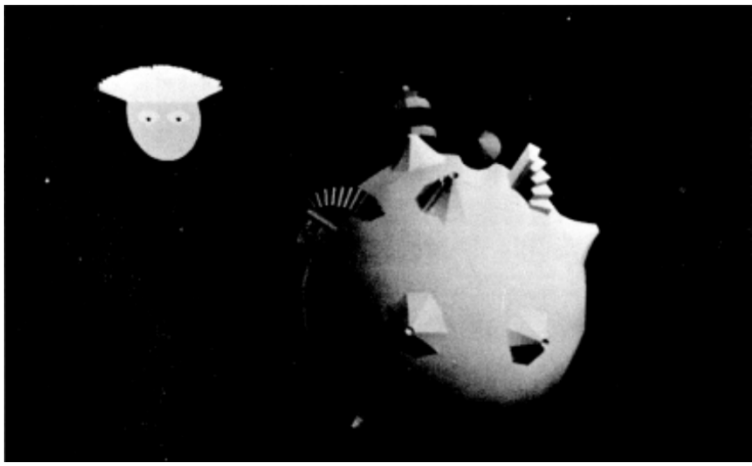
Attention Driven Game Design

The green figures are the three shooters. The blue figure represents players. The red circles show the fixation areas with low and high entropies (low entropy corresponding to easy level, high entropy corresponding to hard level.

Source: “Game design guided by visual attention”, Jie, 2007.

Miscellaneous Applications

- Context Dependent Information Delivery.
  - Example: Starker and Bold (1990)
    - Created a 3-D graphics world to be shown to the observer/users (a small planet slowly revolving against a background of twinkling stars).
    - Gaze was mapped to the object located on the planet to compute an interest level
    - The objects with the currently highest interest levels were then chosen to select a related topic to be narrated with synthesized speech.



Source: “A gaze-responsive self-disclosing display”. Starker, 1990.

Miscellaneous Applications

- Avatar Gaze Control in Immersive VE
  - Gaze controlled avatars provided improved subjective experience.

Source: “Eye Tracking for Avatar Eye Gaze Control During Object-Focused Multiparty Interaction in Immersive Collaborative Virtual Environments”. Steptoe, 2008.

References

- The Quest for Consciousness: Neurobiological Approach. Cristof Koch. Publisher: W. H. Freeman, 2004
- James A. Ferwerda, Peter Shirley, Sumanta N. Pattanaik, and Donald P. Greenberg. 1997. A model of visual masking for computer graphics. In Proceedings of the 24th annual conference on Computer graphics and interactive techniques (SIGGRAPH '97), pp 143-152.
- Mahesh Ramasubramanian, Sumanta N. Pattanaik, and Donald P. Greenberg. 1999. A perceptually based physical error metric for realistic image synthesis. In Proceedings of the 26th annual conference on Computer graphics and interactive techniques (SIGGRAPH '99), pp 73-82.
- Hector Yee, Sumanita Pattanaik, and Donald P. Greenberg. 2001. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. ACM Trans. Graph. 20, 1 (January 2001), 39-65.
- V. Sundstedt, K. Debattista, P. Longhurst, A. Chalmers, and T. Troscianko. 2005. Visual attention for efficient high-fidelity graphics. In Proceedings of the 21st Spring Conference on Computer Graphics (SCCG '05), pp 169-175.
- Veronica Sundstedt, Diego Gutierrez, Oscar Anson, Francesco Banterle, and Alan Chalmers. 2007. Perceptual rendering of participating media. ACM Trans. Appl. Percept. 4, 3, Article 15 (November 2007).
- Oscar Anson, Veronica Sundstedt, Diego Gutierrez, and Alan Chalmers. 2006. Efficient selective rendering of participating media. In Proceedings of the 3rd symposium on Applied perception in graphics and visualization (APGV '06), pp 135-142.
- Markus Steinberger, Bernhard Kainz, Stefan Hauswiesner, Rostislav Khlebnikov, Denis Kalkofen, and Dieter Schmalstieg. 2012. Special Section on CANS: Ray prioritization using stylization and visual saliency. Comput. Graph. 36, 6 (October 2012), 673-684.
- Level of Detail for 3D Graphics by D. Luebke, M. Reddy, J. Cohen, A. Varshney, B. Watson, and R. Huebner, published by Morgan Kaufmann, 2002.
- Thomas A. Funkhouser and Carlo H. Séquin. 1993. Adaptive display algorithm for interactive frame rates during visualization of complex virtual environments. In Proceedings of the 20th annual conference on Computer graphics and interactive techniques (SIGGRAPH '93), pp 247-254.
- George Drettakis, Nicolas Bonneel, Carsten Dachsbacher, Sylvain Lefebvre, Michael Schwarz, Isabelle Viaud-Delmon. 2007. An Interactive Perceptual Rendering Pipeline using Contrast and Spatial Masking, EGSR07: 18th Eurographics Symposium on Rendering.
- Chang Ha Lee, Amitabh Varshney, and David W. Jacobs. 2005. Mesh saliency. In ACM SIGGRAPH 2005 Papers (SIGGRAPH '05), PP. 659-666.
- G. A. Koulieris, G. Drettakis, D. Cunningham, and K. Mania. 2014. C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics. Comput. Graph. Forum 33, 4 (July 2014), 41-49.
- R. Mantiuk, B. Bazyluk, R. K. Mantiuk. Gaze-driven Object Tracking for Real Time Rendering.Computer Graphics Forum, Volume 32, Issue 2pt2, May 2013. PP 163–173
- Rafal Mantiuk, Karol Myszkowski, and Sumanta Pattanaik. 2003. Attention guided MPEG compression for computer animations. In Proceedings of the 19th Spring Conference on Computer Graphics (SCCG '03), pp 239-244.
- Zhicheng Li, Shiyin Qin, and Laurent Itti. 2011. Visual attention guided bit allocation in video compression. Image Vision Comput. 29, 1 (January 2011), 1-14.