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

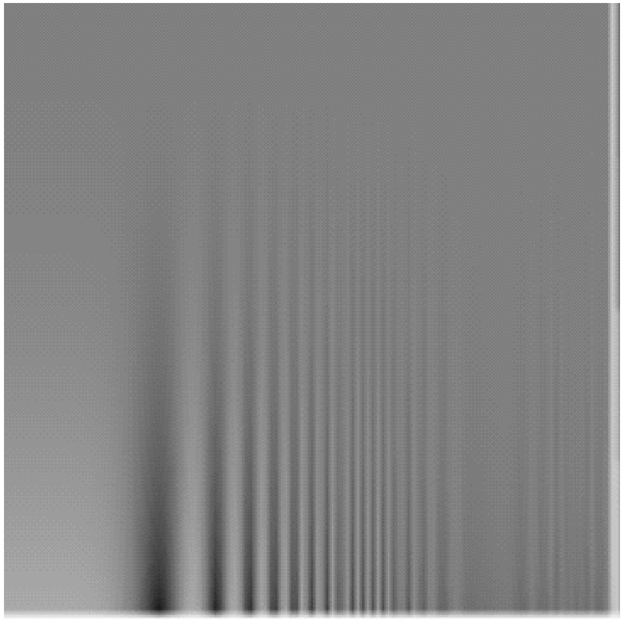
- **Uses a multi-scale visual model (the Sarnoff Visual Discrimination Model) to guide the sampling pattern in MC Ray Tracing**
  - Optimized for speed
    - Haar wavelets are used at the cortex filtering stage instead of costly Laplacian pyramid originally used in the VDM
  - Correct color handling
    - CIE XYZ transformed to SML space modeling retinal cone sensitivity
    - Opponent contrast space: a single achromatic (A) and two opponent color channels (C_1 and C_2)
    - Independent contrast sensitivity processing for AC_1C_2 channels

Chromatic CSF

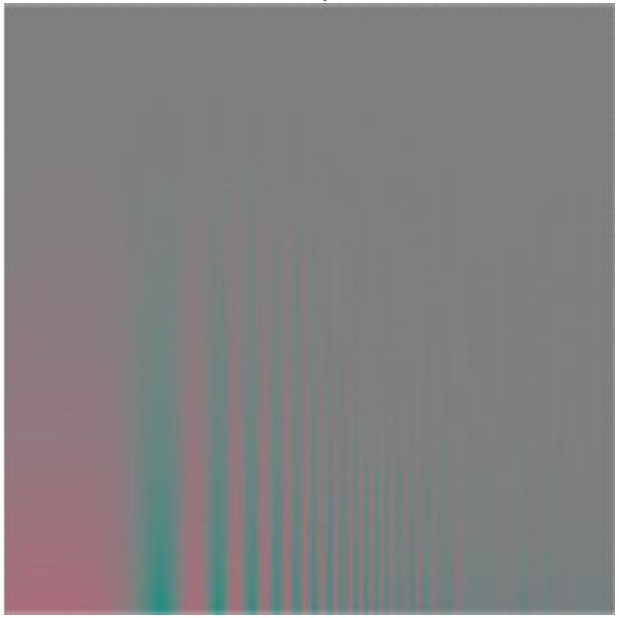
Independent contrast sensitivity processing for AC_1C_2 channels

Band-pass filter

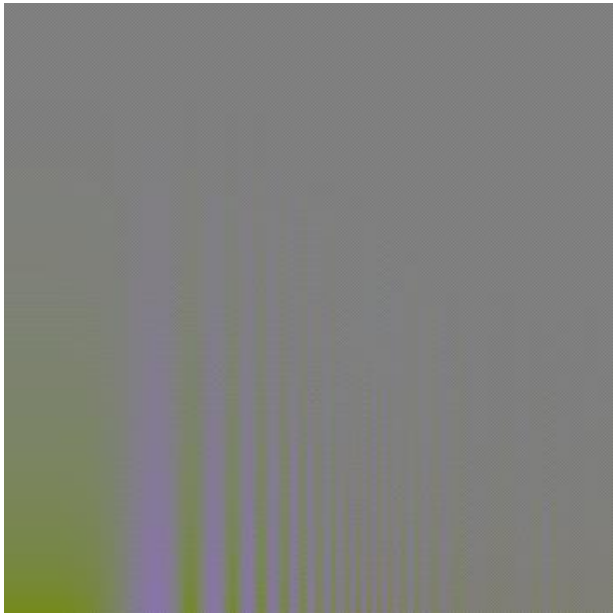
Low-pass filter



Luminance

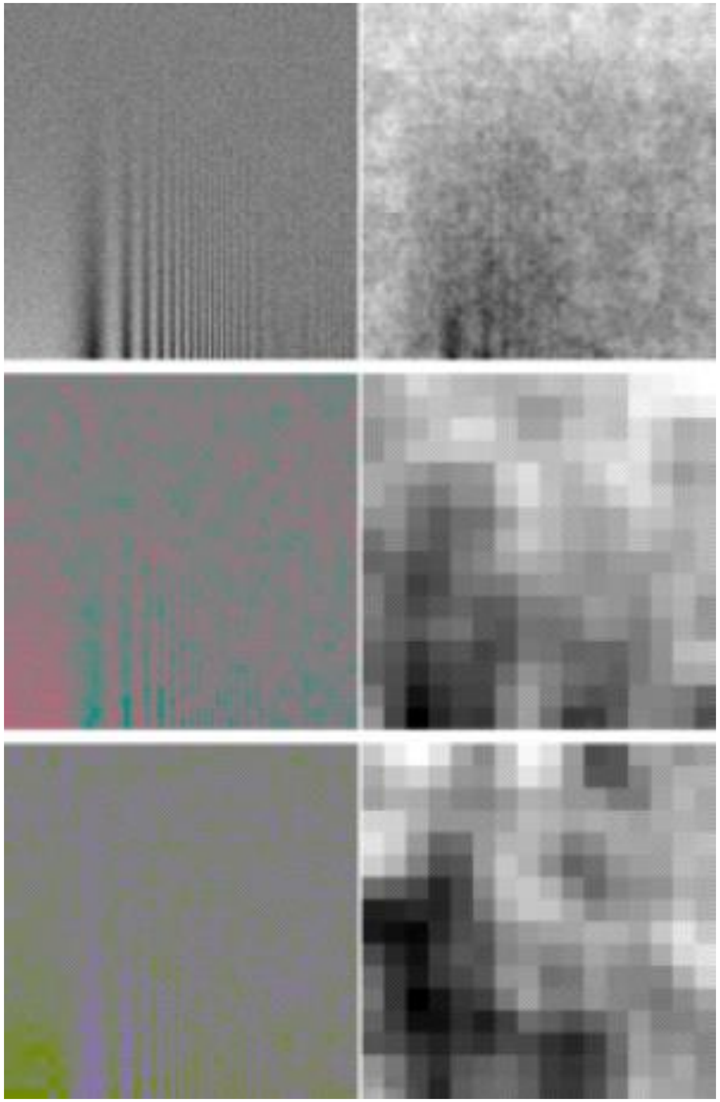


Red-Green Opponent



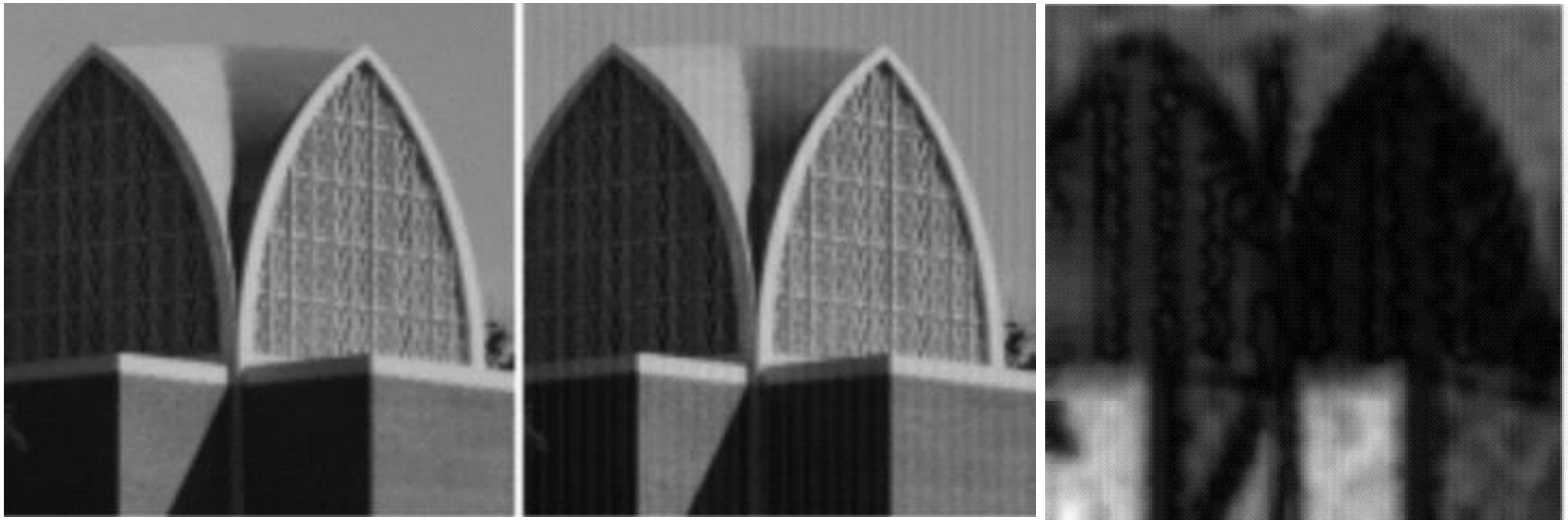
Blue-Yellow Opponent

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).**

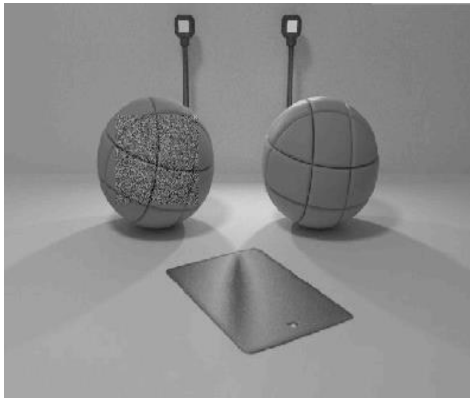
Visual Masking



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

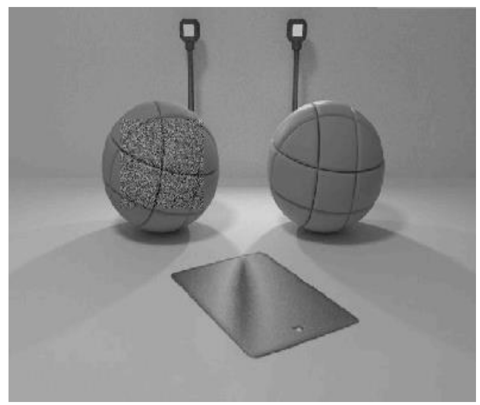
- **Step I:** compute an estimate of the image using lesser number of samples per pixel
  - **A Haar wavelet image approximation is generated and then refined**

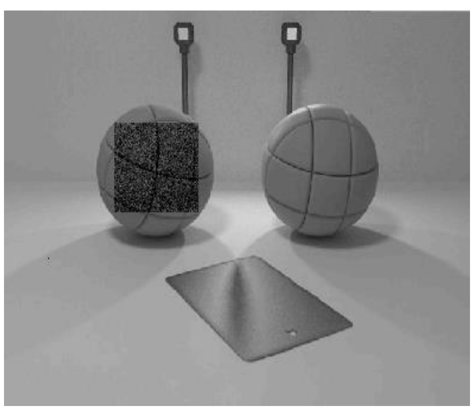


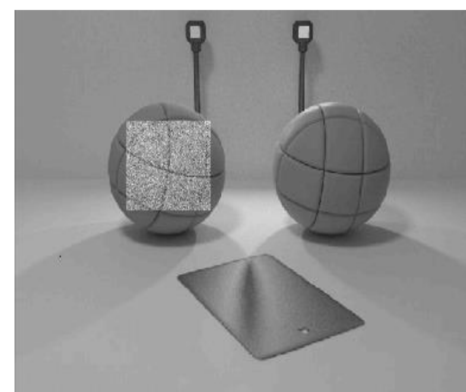
- **Step II:** from MC variance in samples of each pixel estimate the pixel error bounds.
  - The error expressed in terms of the variance of the detail terms in the Haar image representation

Perception-based Adaptive Sampling

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

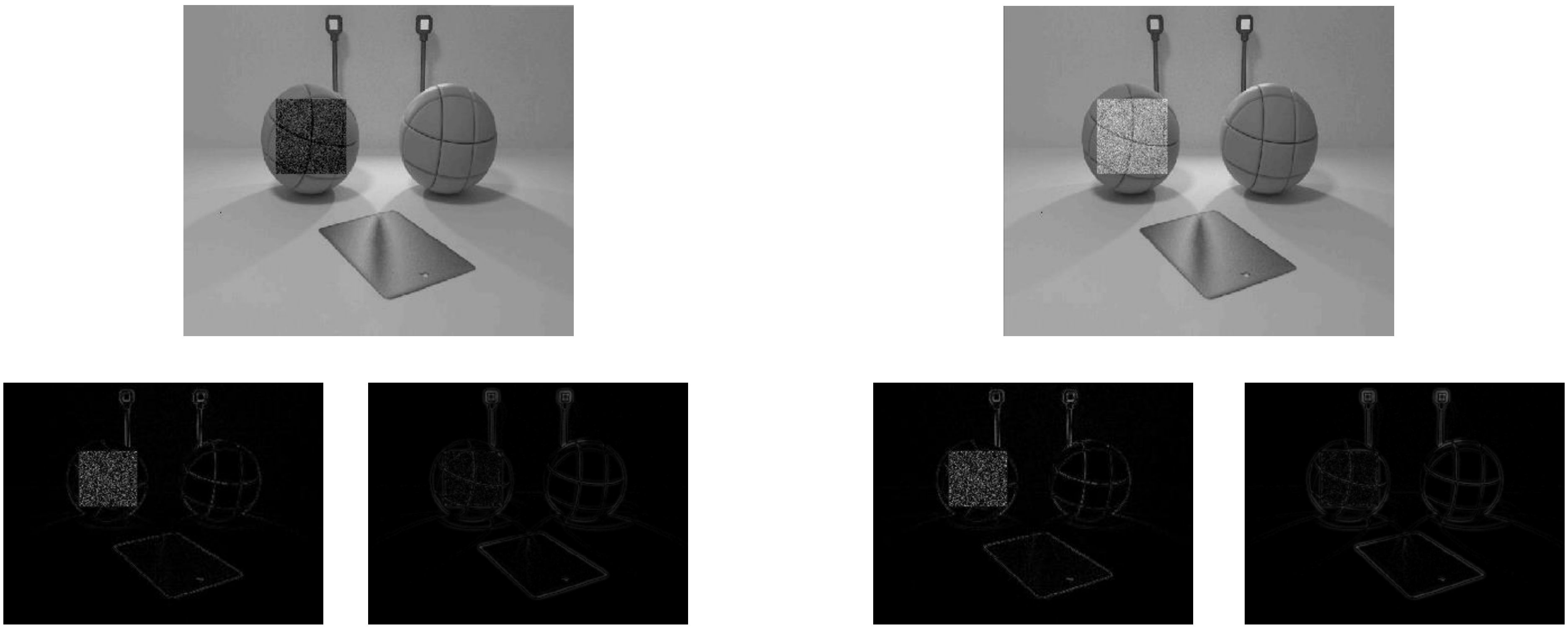






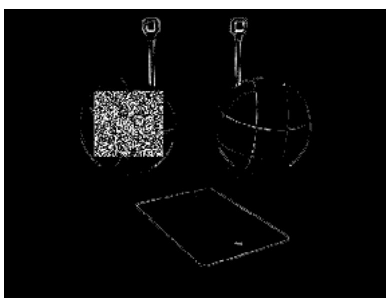
Perception-based Adaptive Sampling

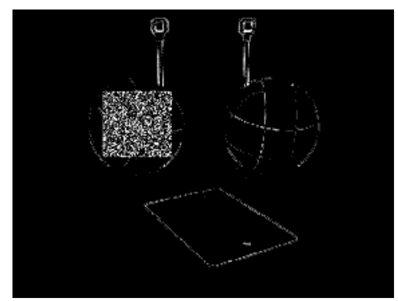
- **Step IV:** Compute oriented band-pass images.



Perception-based Adaptive Sampling

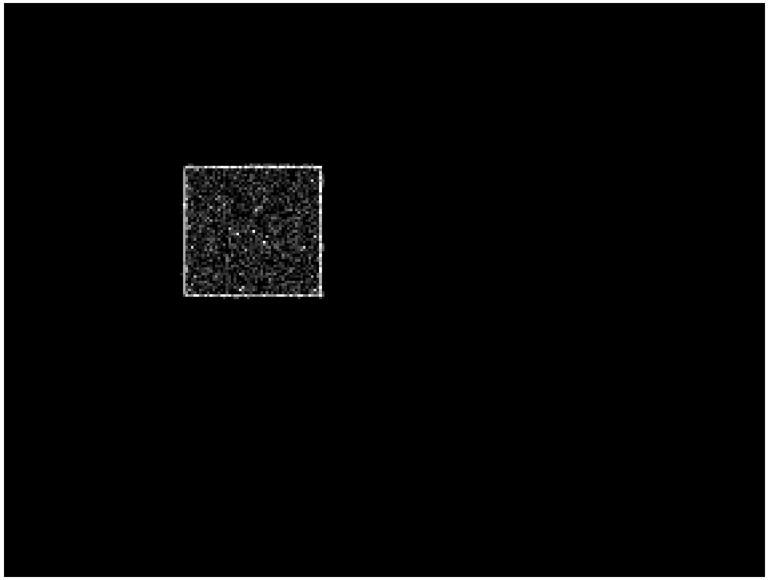
- **Step V:** For each band compute threshold from TVI, CSF and Masking functions. Normalize the band pass images with the computed threshold.






Perception-based Adaptive Sampling

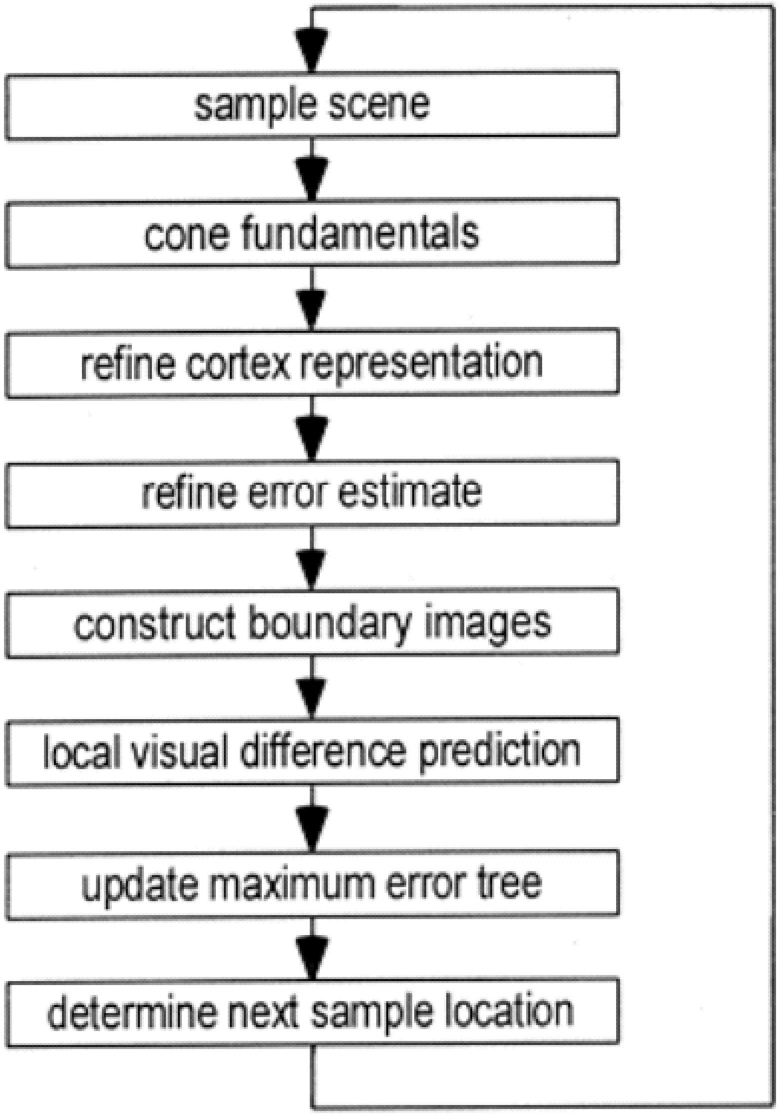
- **Step VI:** Find the difference between each band of the two images.



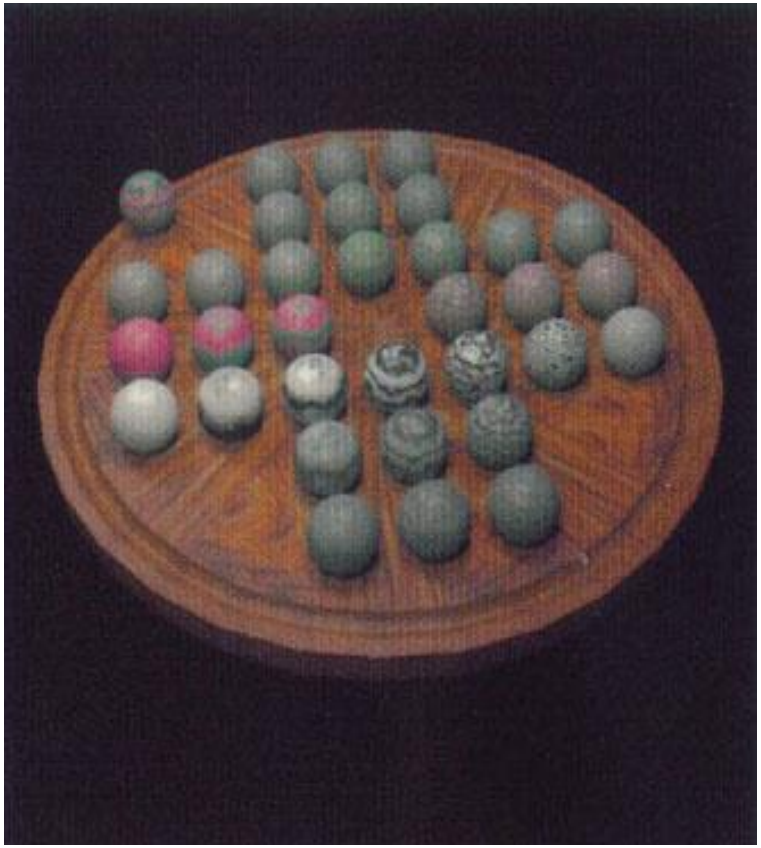
- **Step VII** : Refine the area with maximum difference.

Perception-based Adaptive Sampling

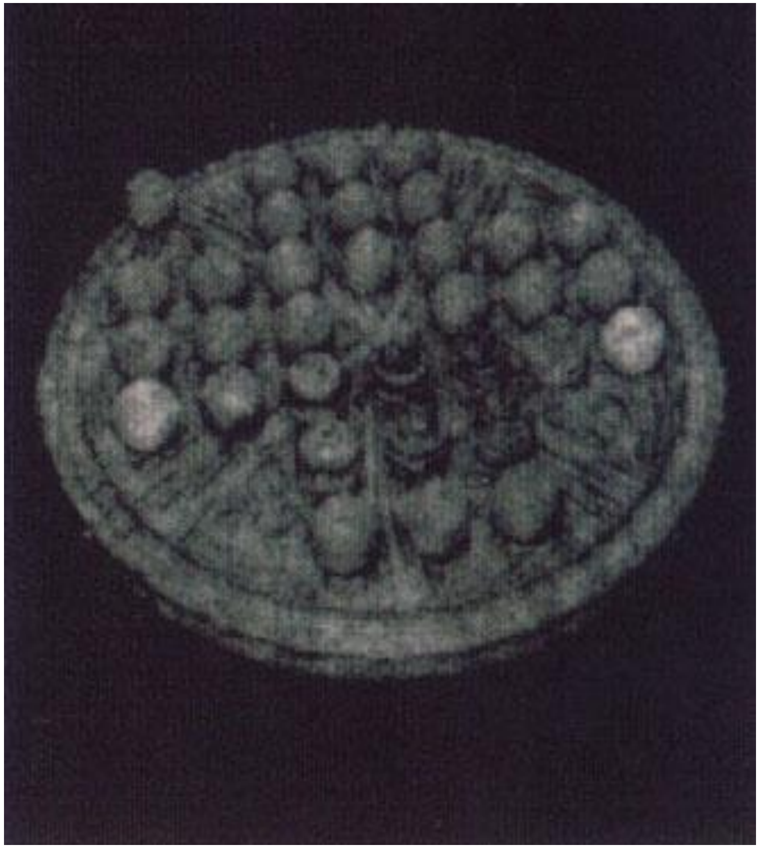
- **Algorithm summary**



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

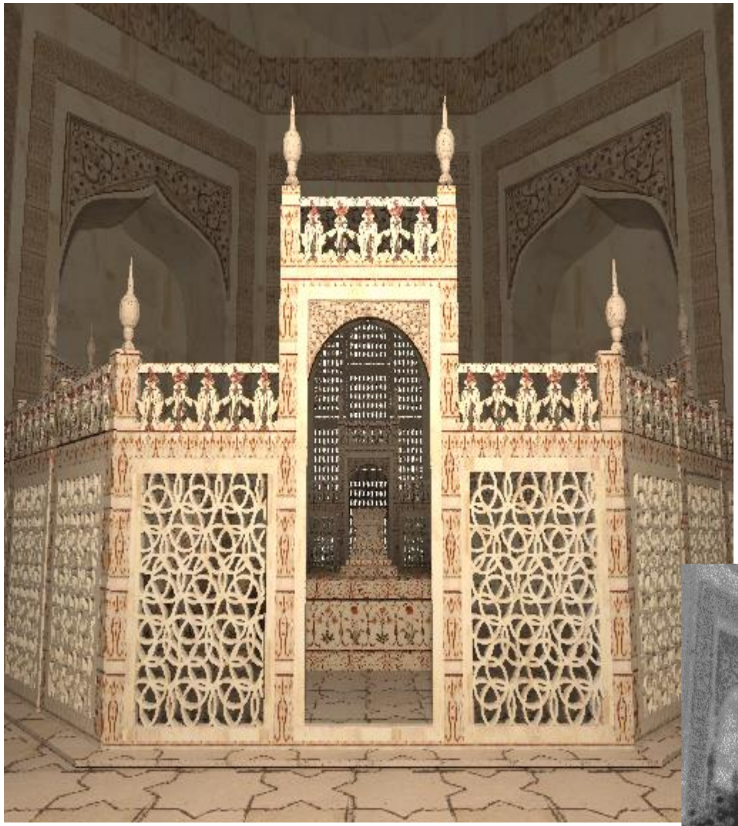
Aims for perceptual accuracy

- Limitations of the human visual system...

  perceptual accuracy < physical accuracy.
- Perceptual accuracy guides rendering, not physical accuracy.

Preview

6% effort



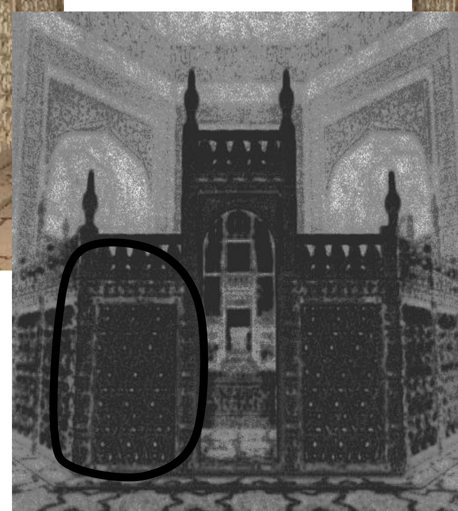
effort distribution (darker regions - less effort)

physically accurate

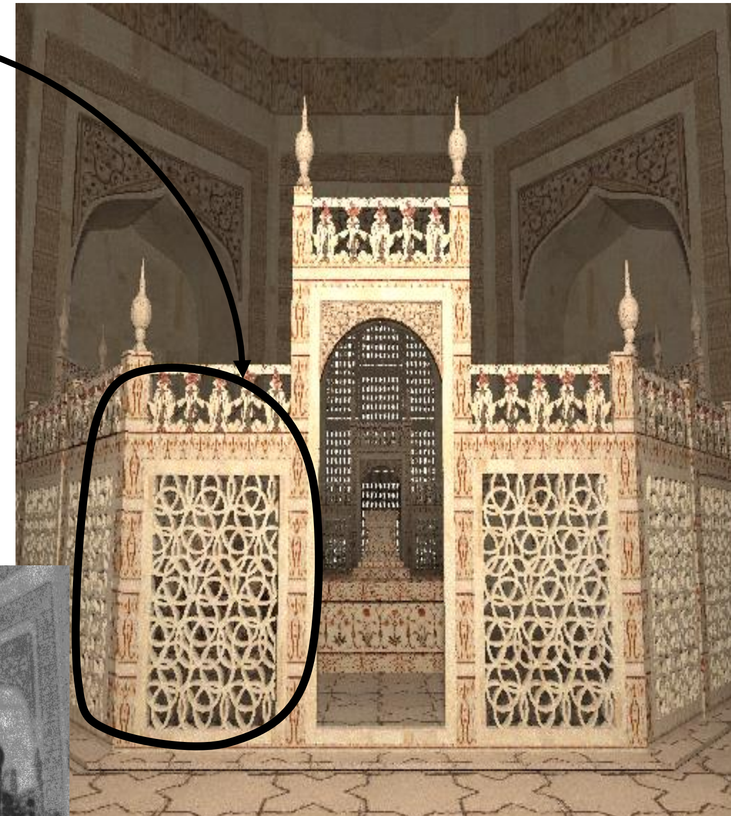
perceptually accurate

Preview

effort distribution (darker regions - less effort)



6% effort



physically accurate

perceptually accurate

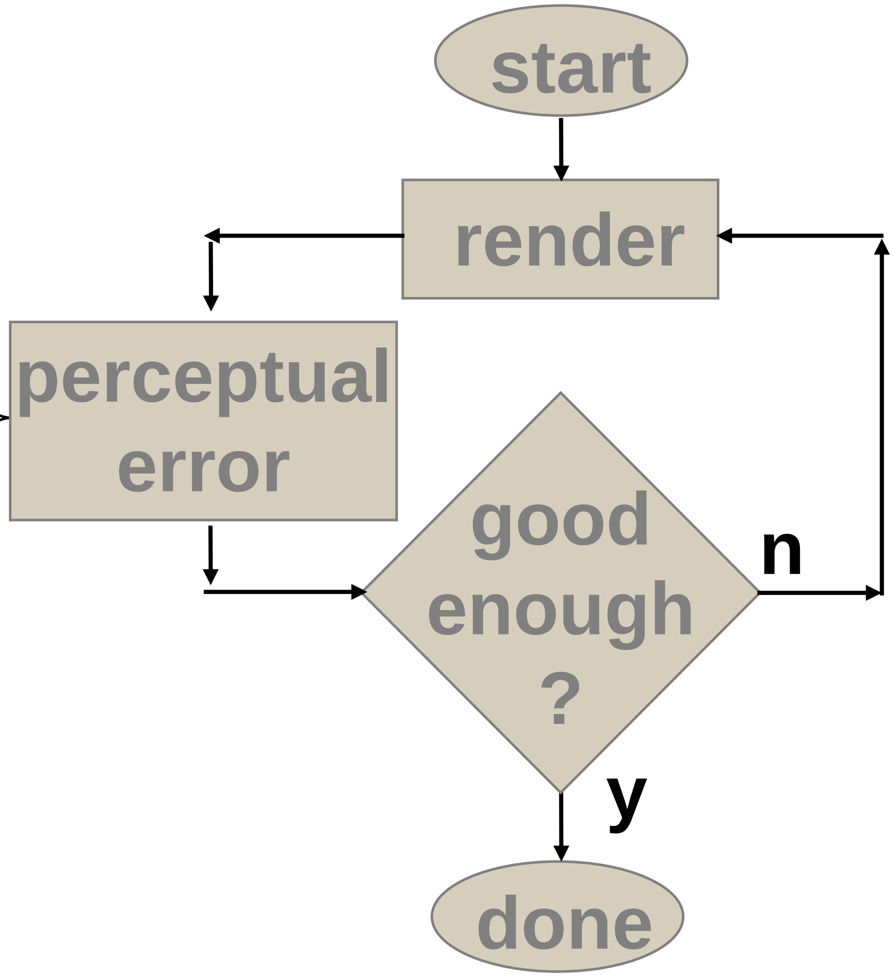
Perceptually Based Rendering

Traditional approach:

Pair of images to compare at each time step

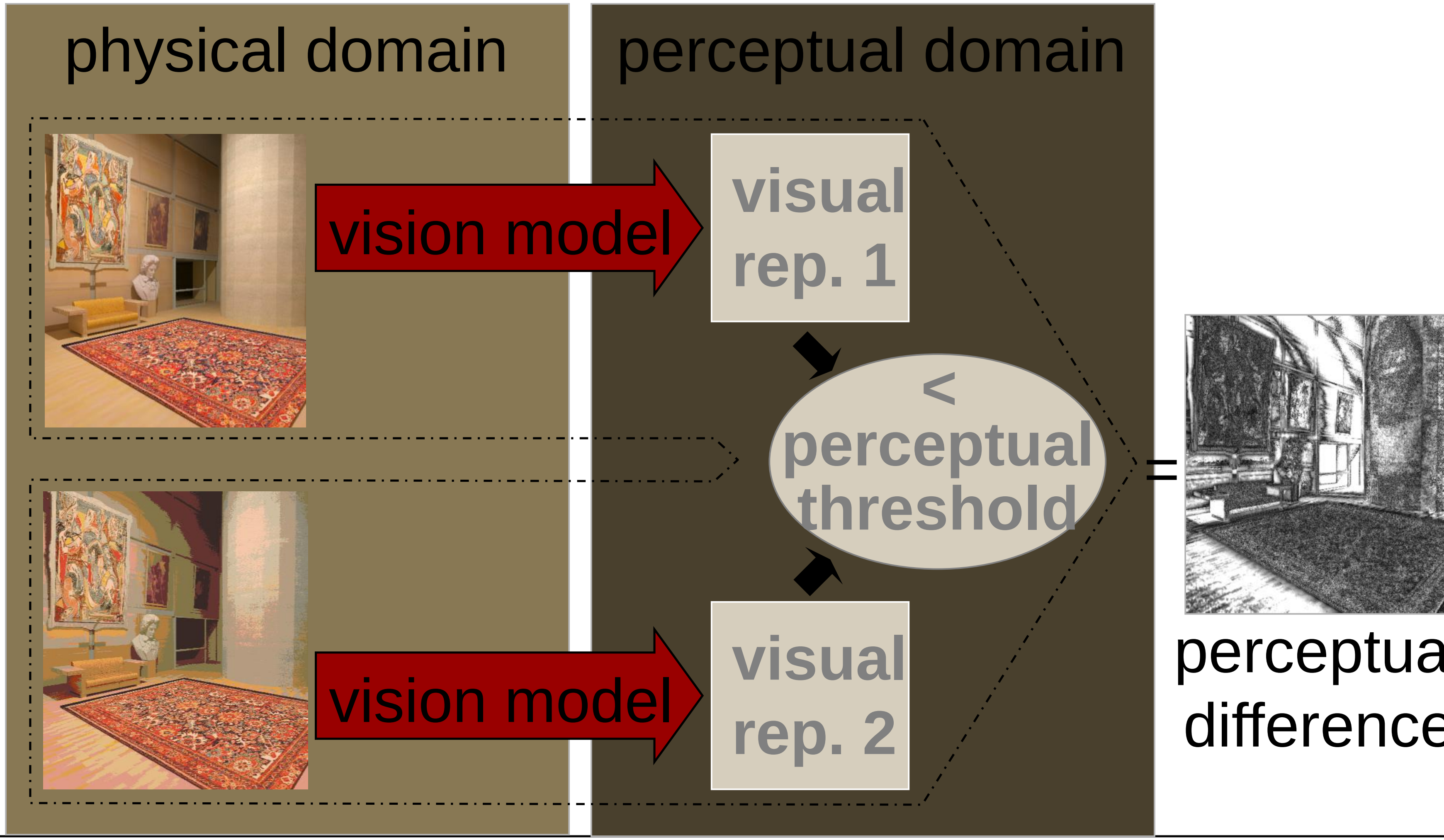
(a) intermediate images at consecutive time steps.

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

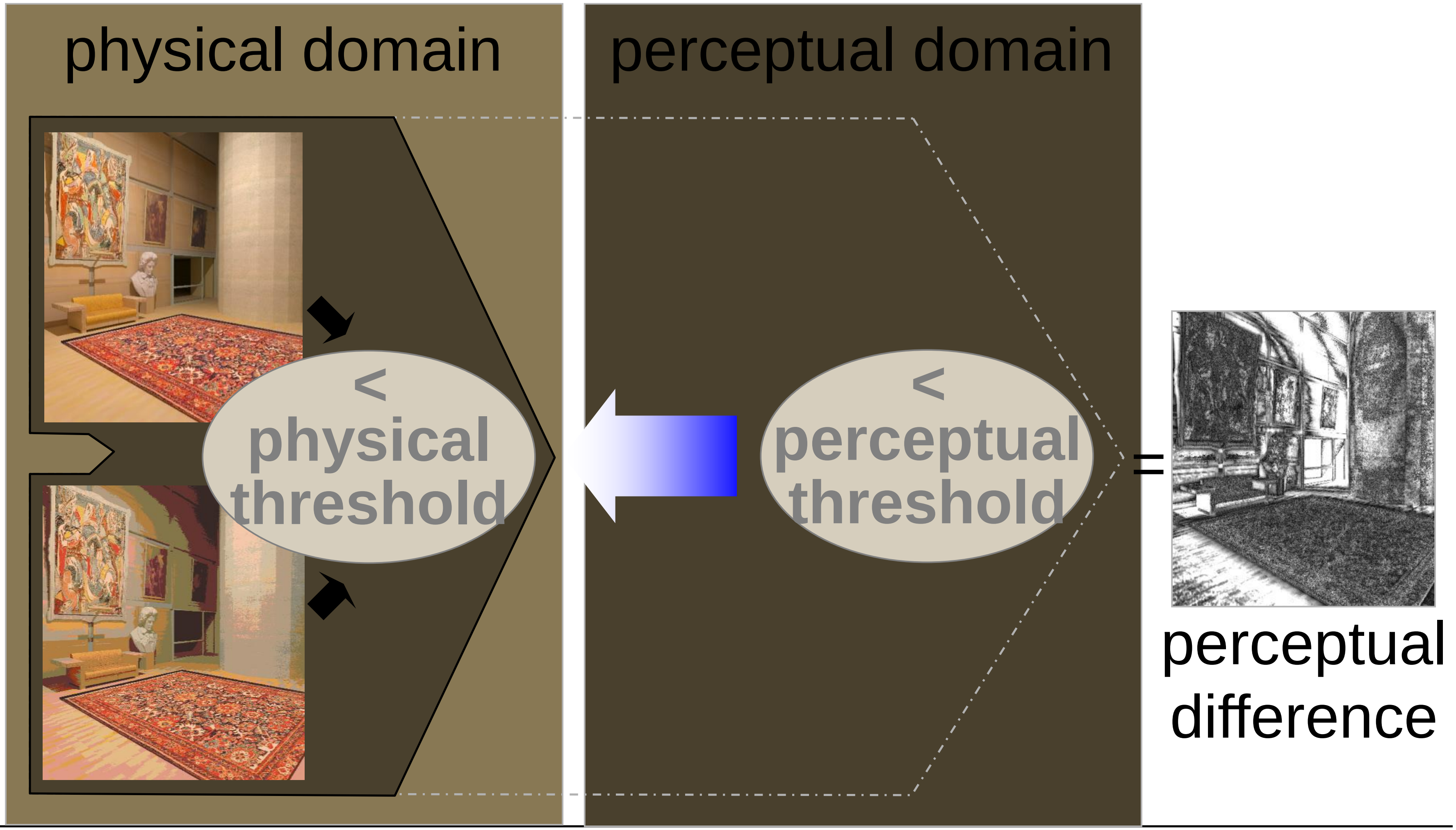


Perceptual Error Metric

Vision model - expensive



Perceptually Based
Physical Error Metric

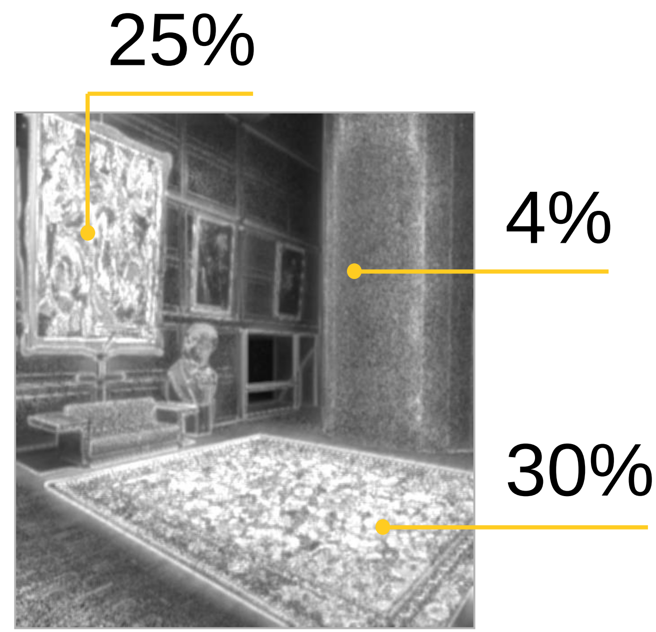


Physical Threshold Map

Predicted bounds of permissible luminance error

input image

threshold model



physical threshold (brighter regions - higher thresholds)

Threshold Model

Components

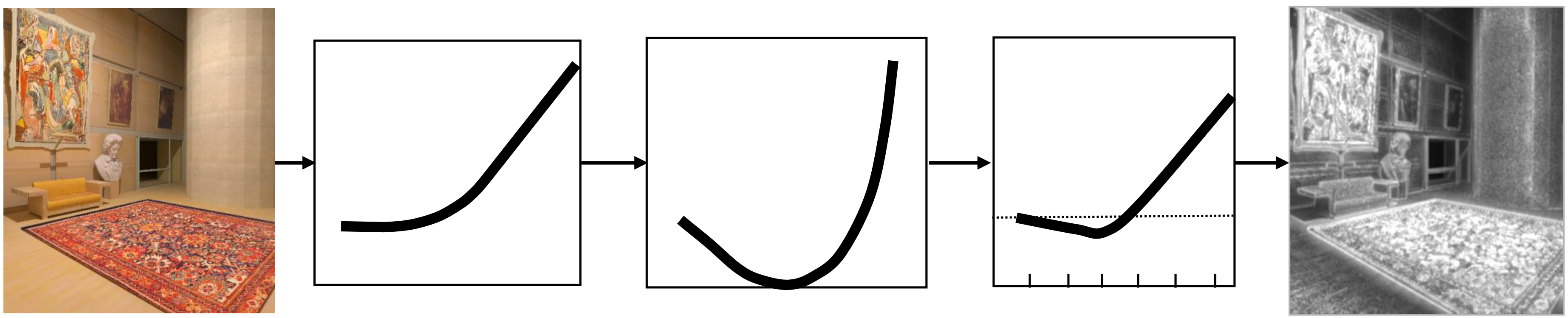
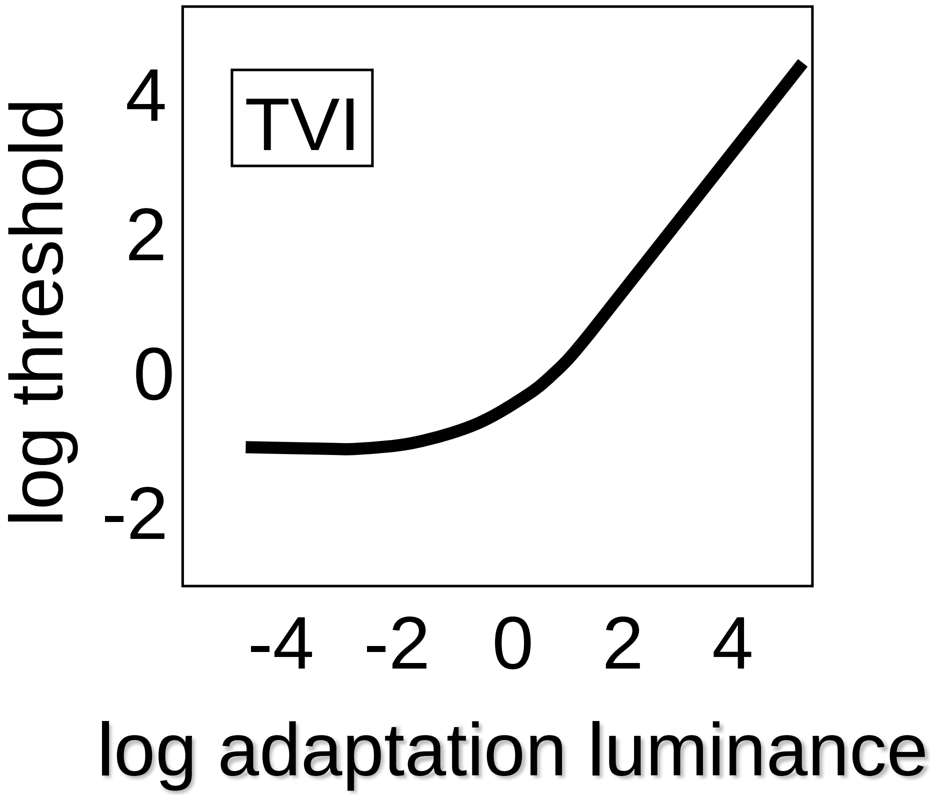
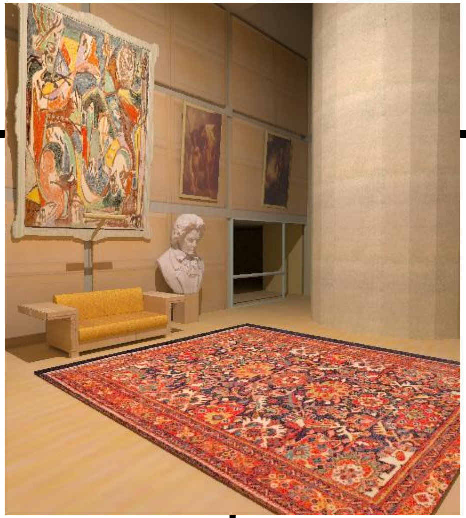


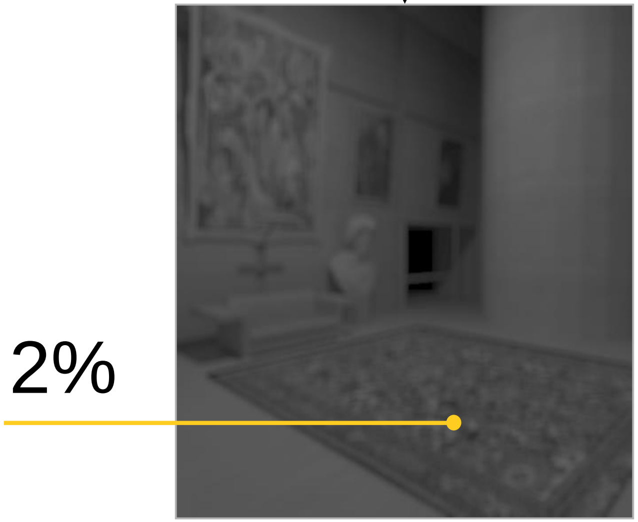
image | luminance component | frequency component | contrast component | threshold map

Threshold Model

1. Luminance component



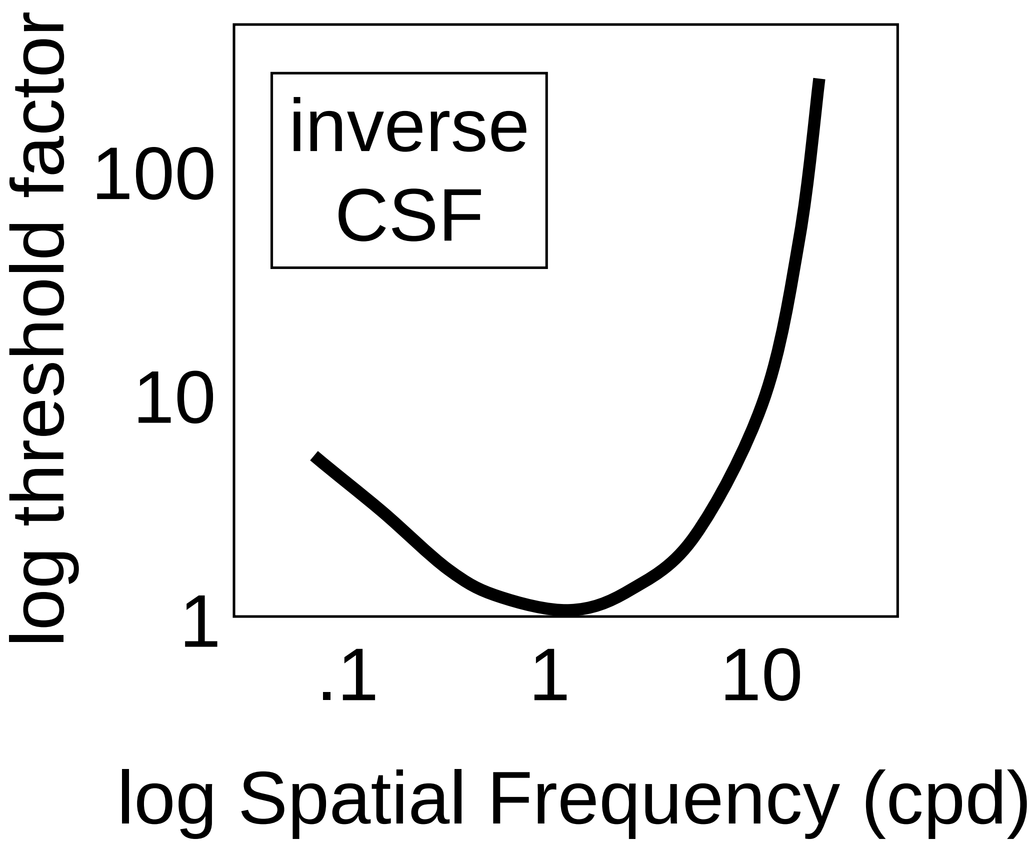


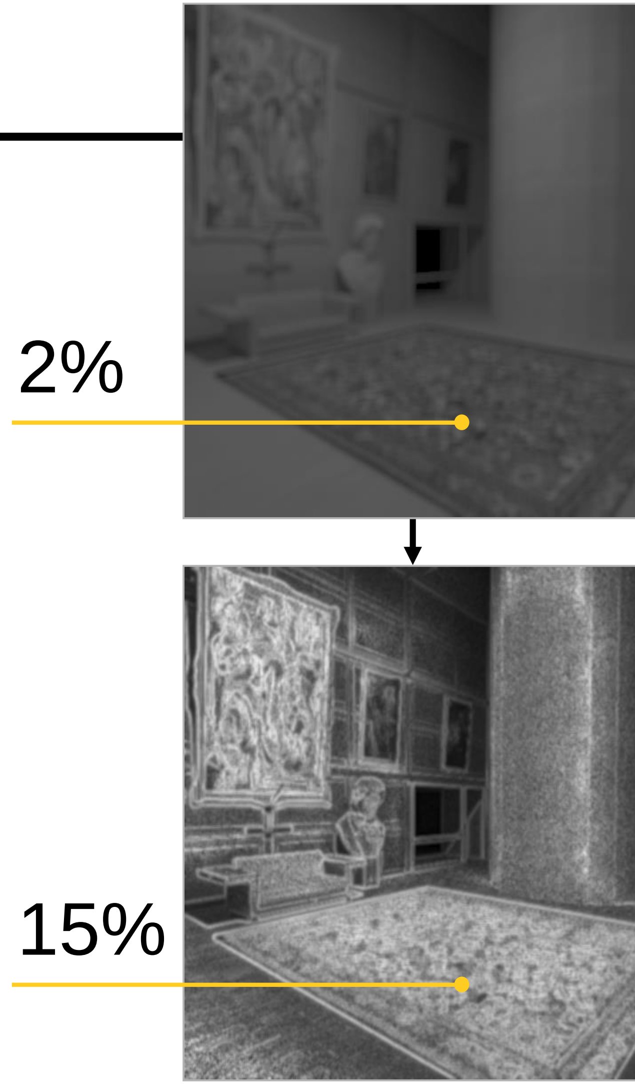


threshold due to luminance

Threshold Model

2. Frequency component

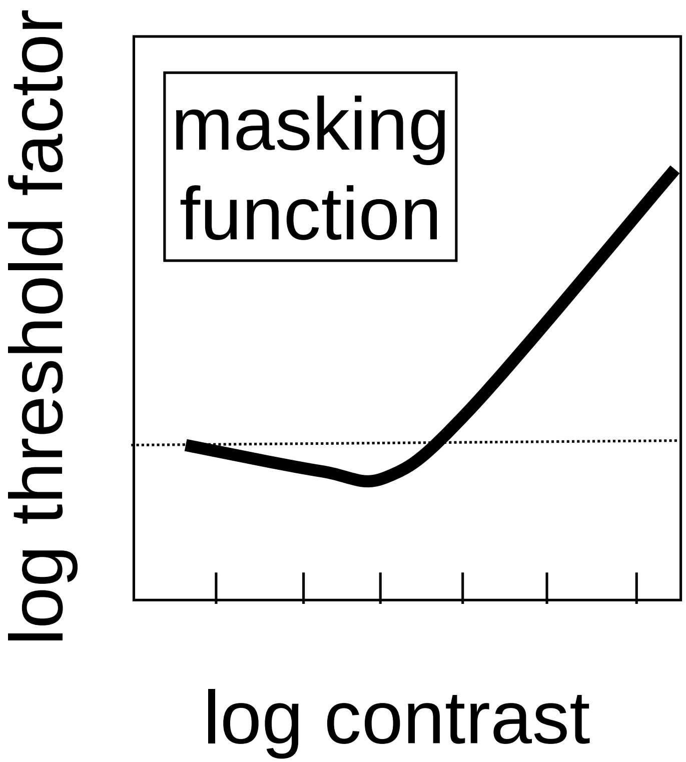


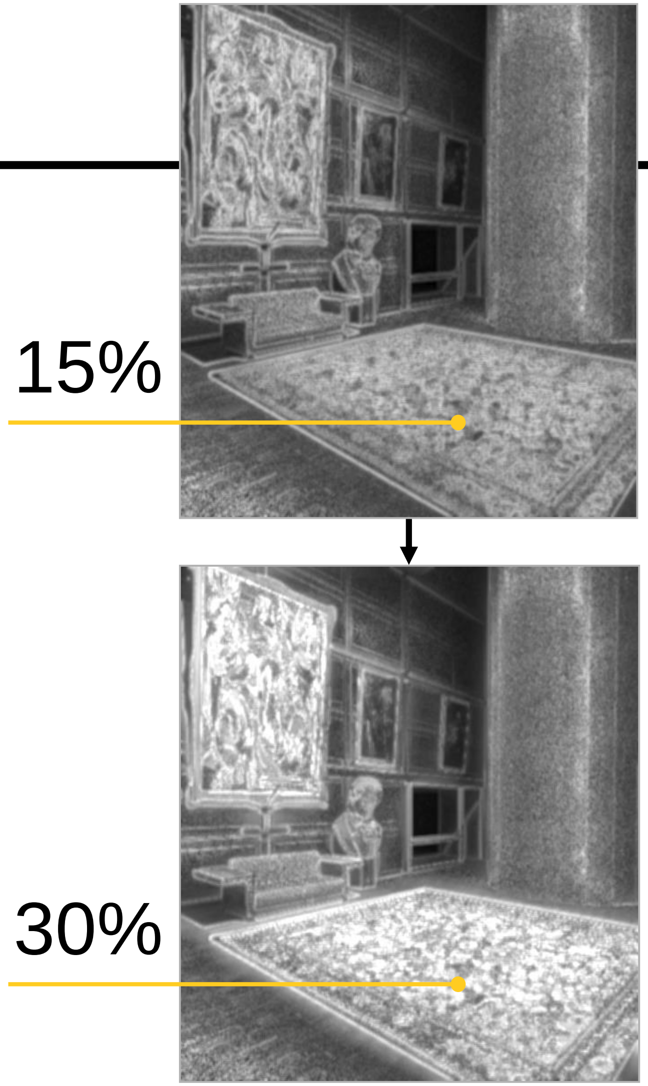


threshold due to luminance + freq.

Threshold Model

3. Contrast component (visual masking)





threshold due to luminance + freq. + contrast

Validation

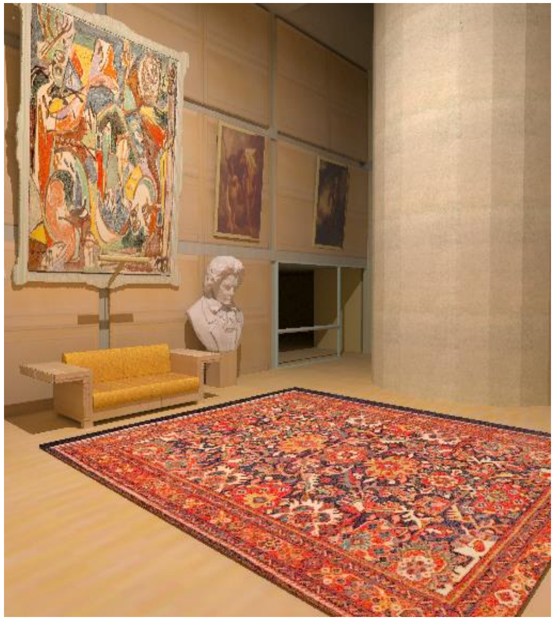
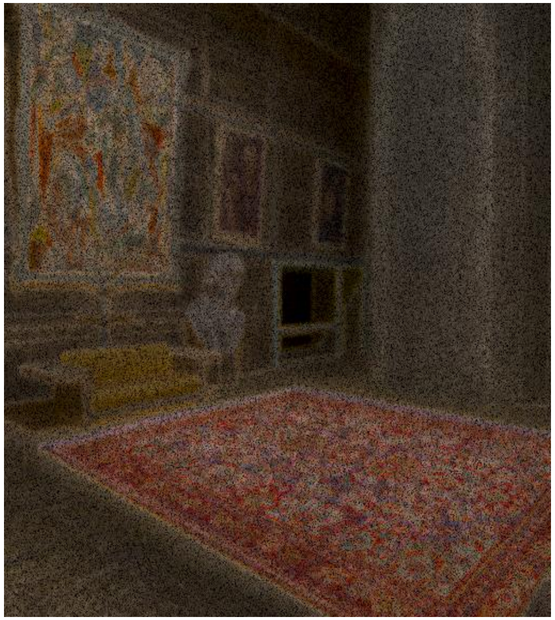
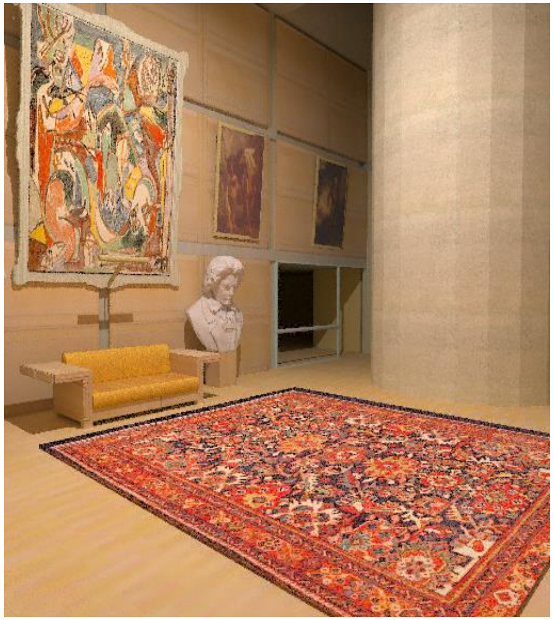
 +  = 

image noise image + noise

Threshold Model

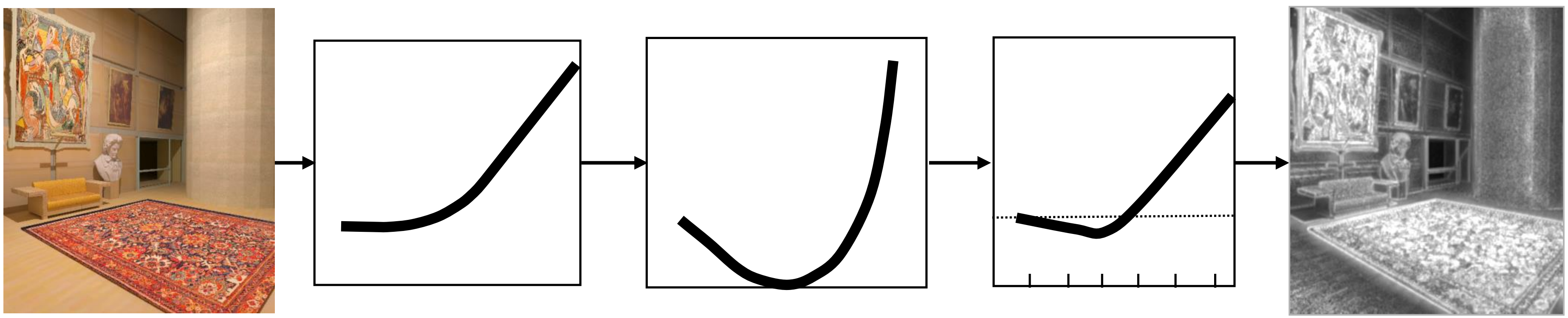
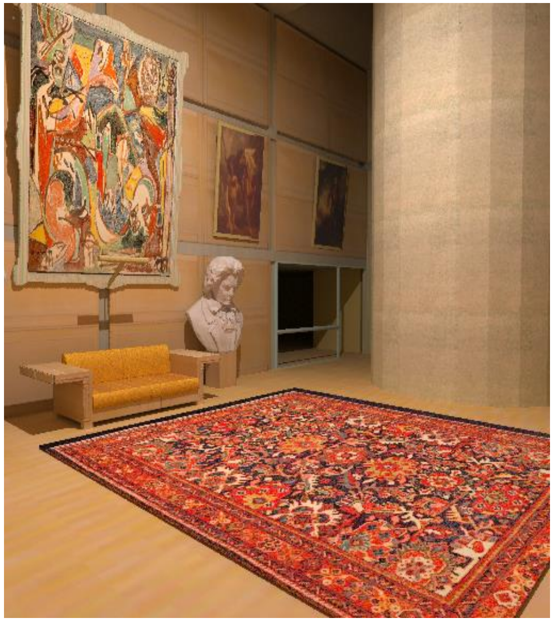
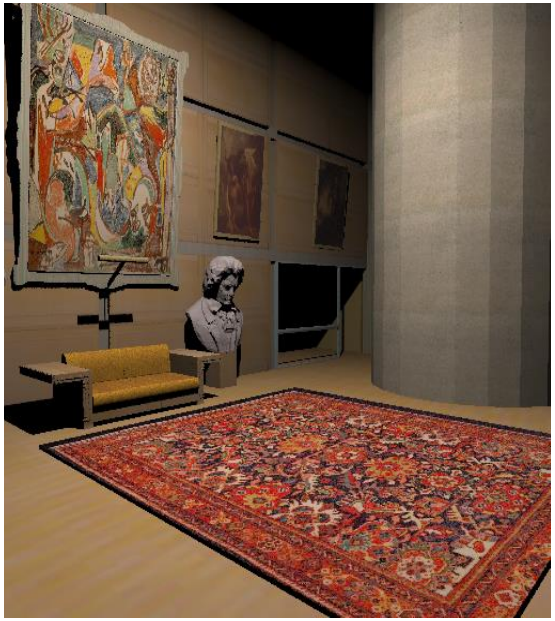


image | luminance component | frequency component | contrast component | threshold map

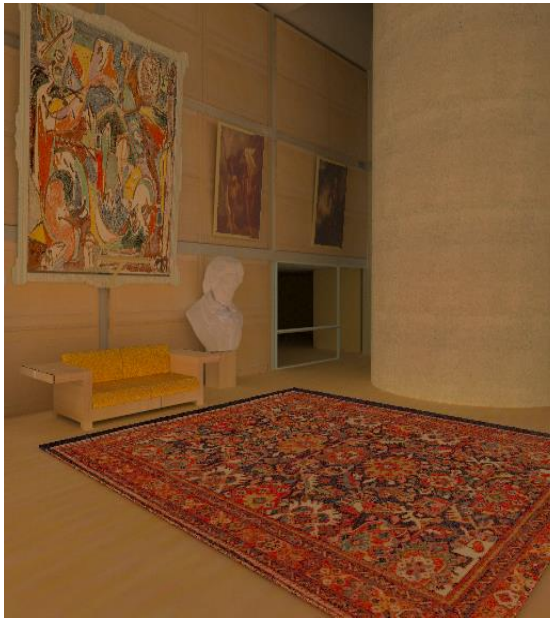
Global Illumination Revisited



=



+



global illumination

direct illumination (fast)

indirect illumination (slow)

Threshold Model Revisited

direct illum.

spatially-dependent processing

12 sec

1 time
precompute

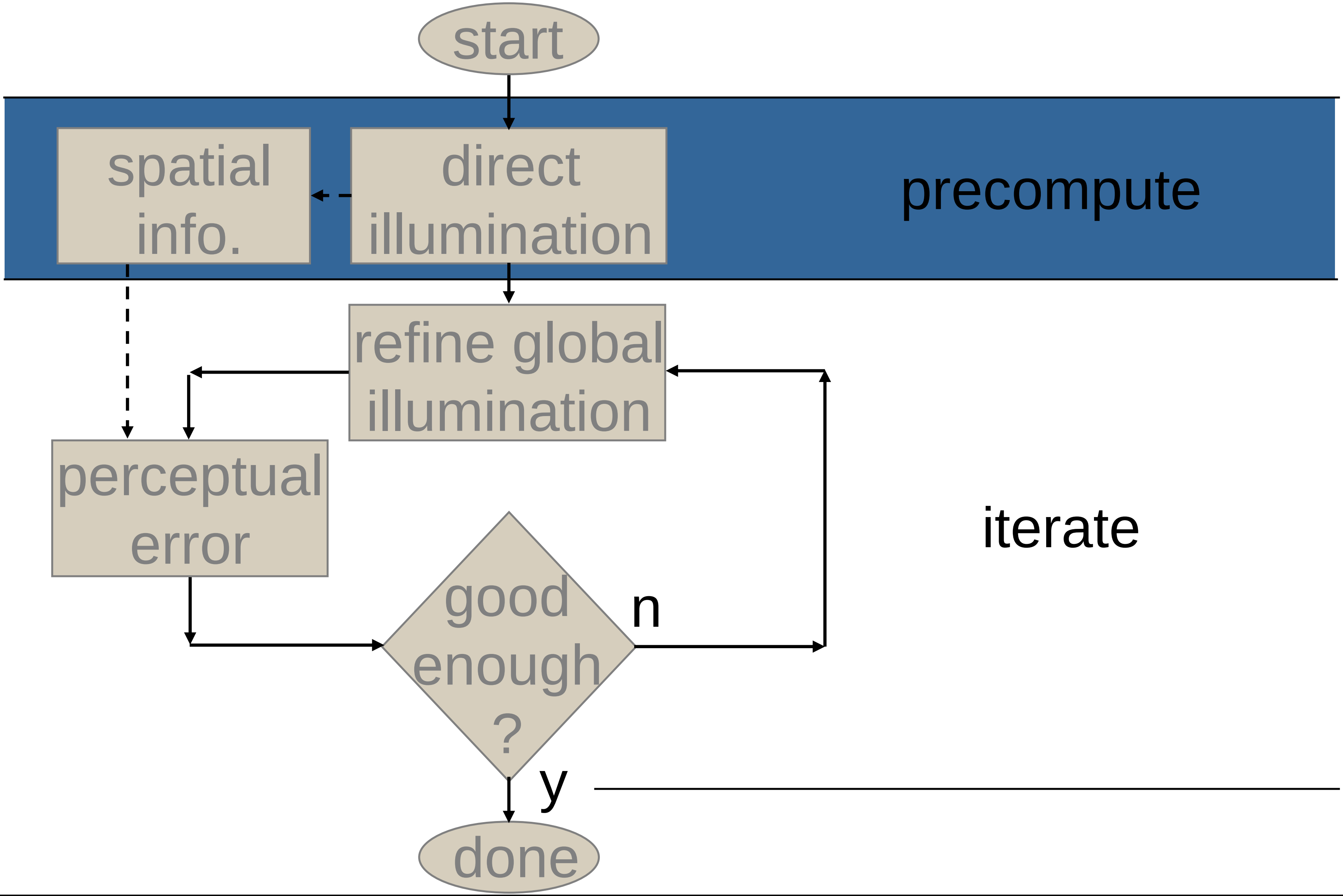
partial global illum.

luminance-dependent processing

0.1 s

N times
iterate

Adaptive Rendering Algorithm

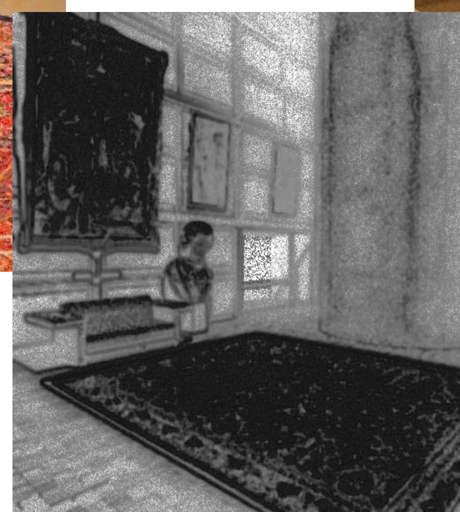


Results

5% effort

reference solution

effort distribution (darker regions - less effort)



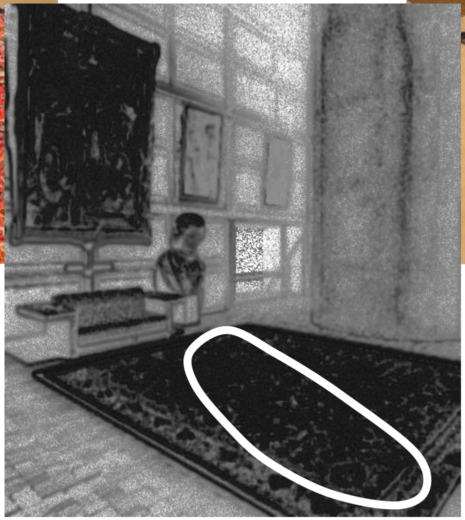
adaptive solution

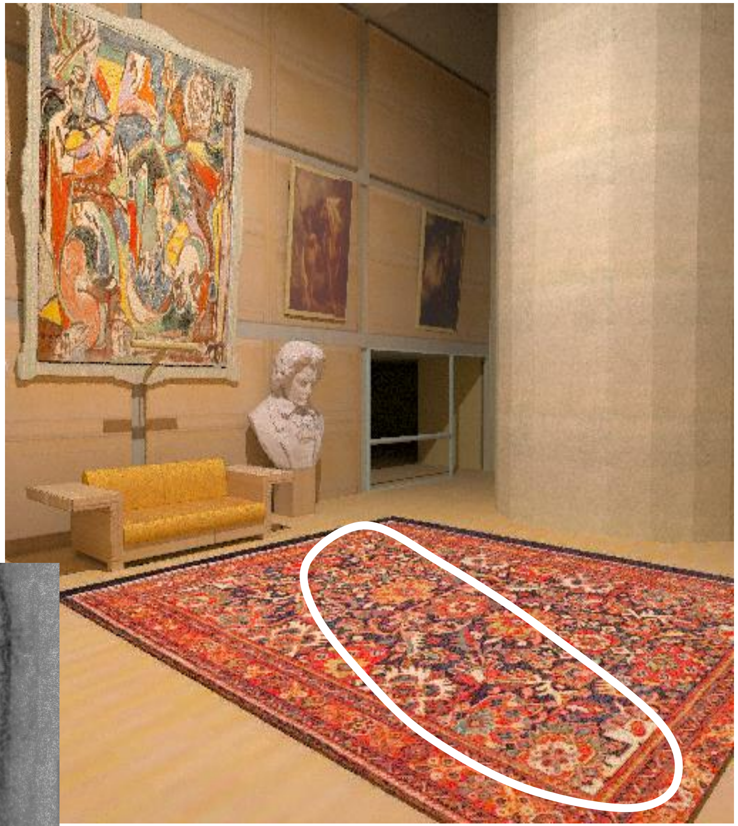
Results: Masking by Textures

5% effort

reference solution

effort distribution (darker regions - less effort)

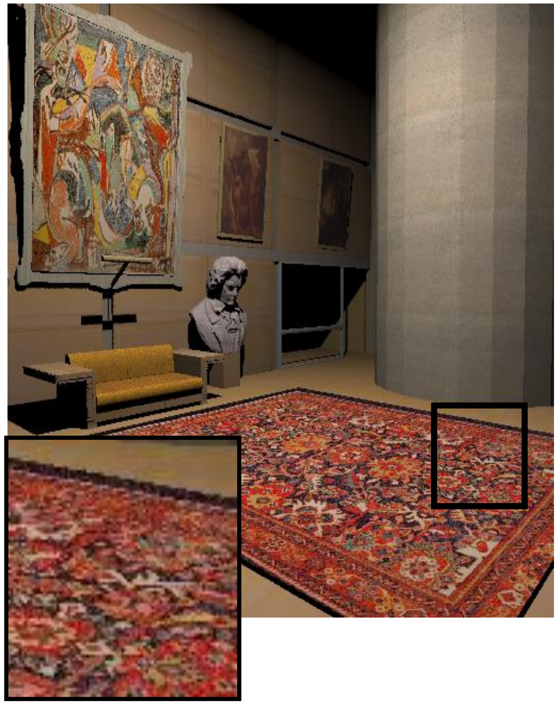




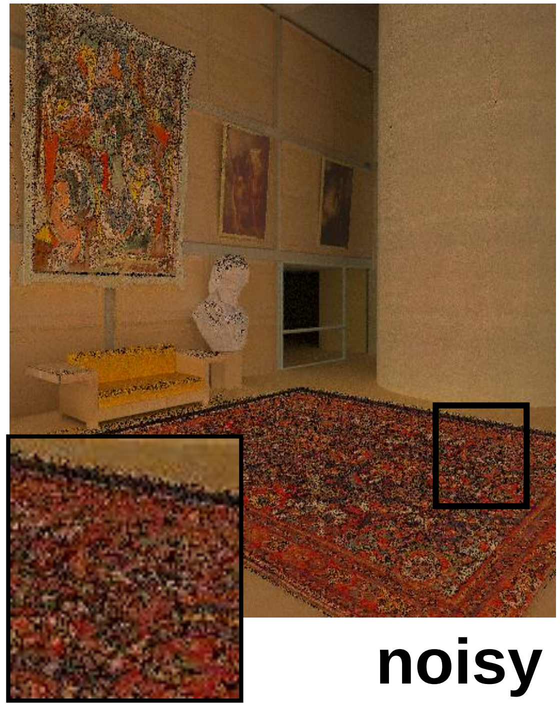
adaptive solution

Results

5% effort



+



noisy

=

masked

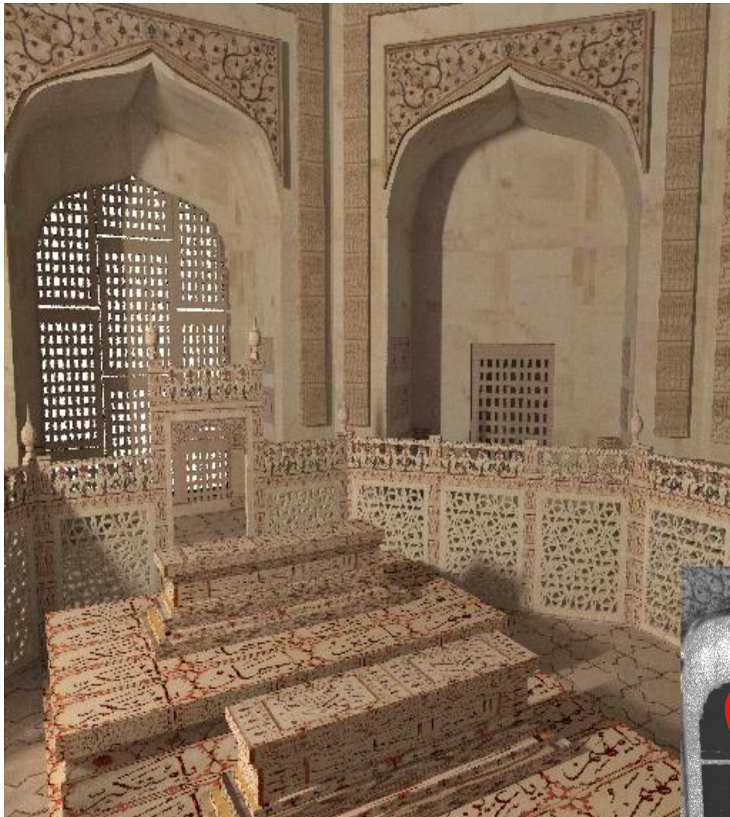
direct illumination

adaptive indirect illumination

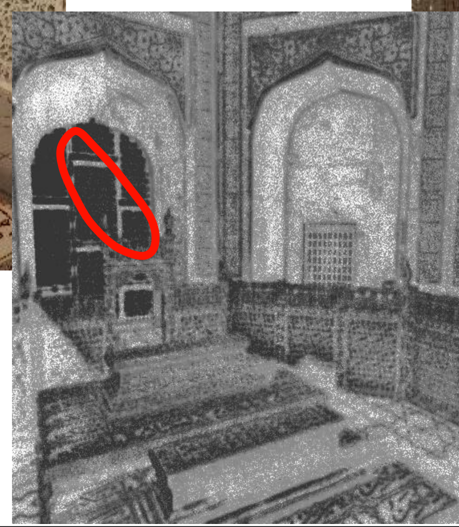
adaptive global illumination

Results: Masking by Geometry

5% effort



effort distribution (darker regions - less effort)

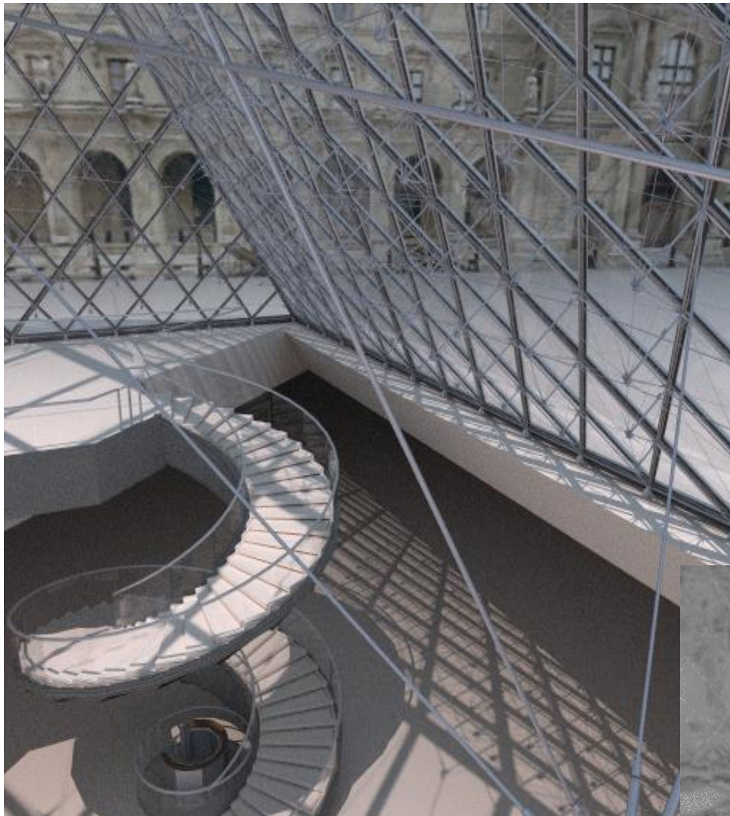


reference solution

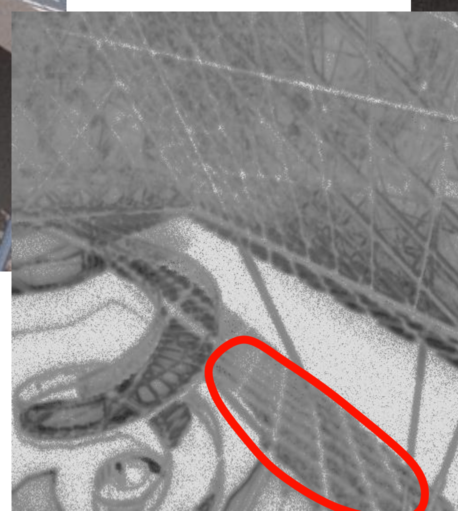
adaptive solution

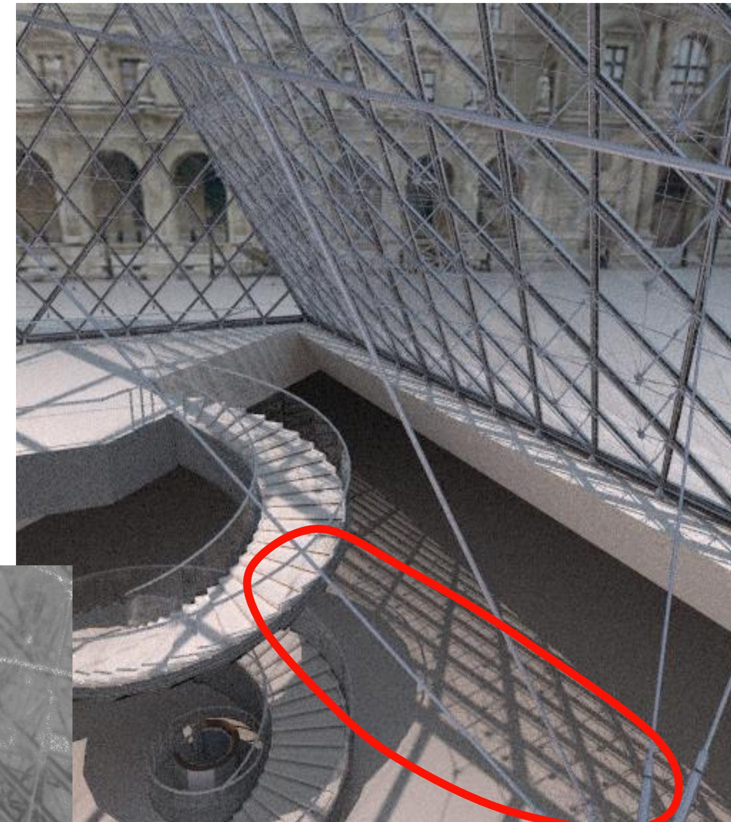
Results: Masking by Shadows

6% effort



effort distribution (darker regions - less effort)





reference solution

adaptive solution

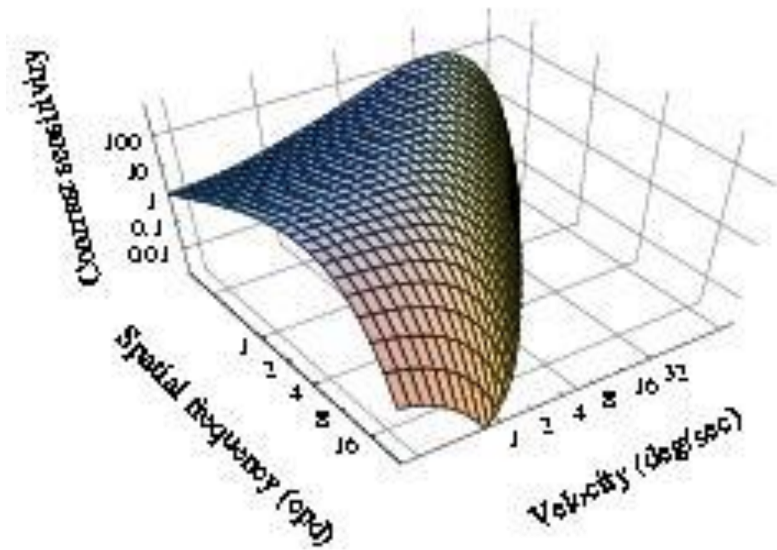
Discussion

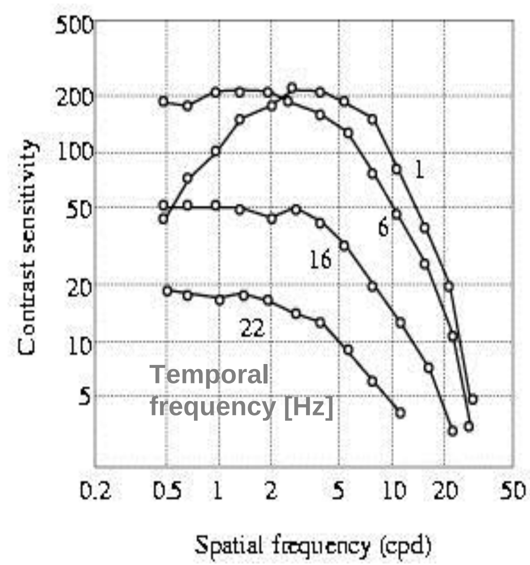
New and efficient perceptually based global illumination technique.

- 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).**

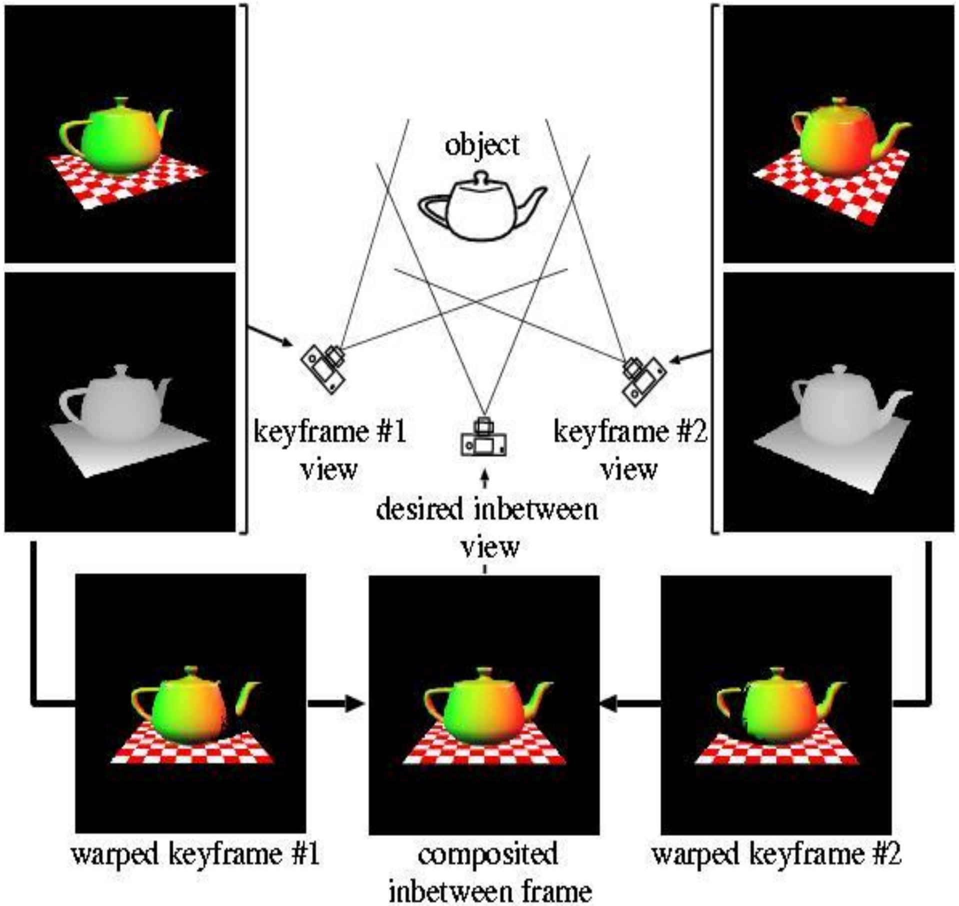


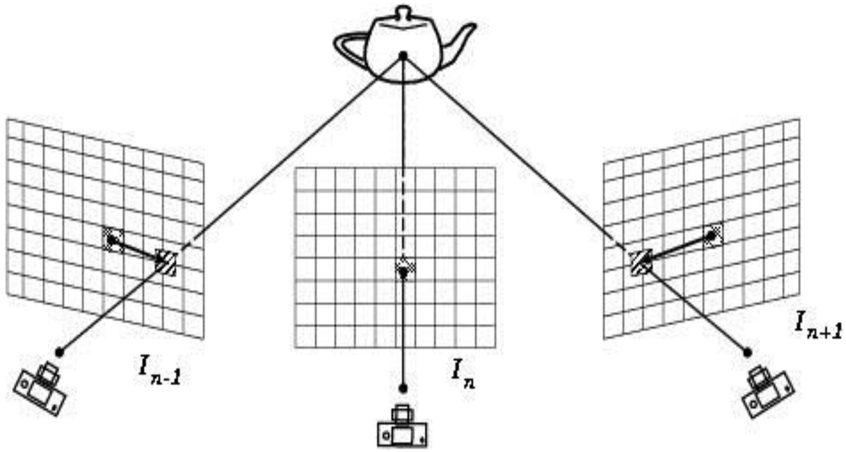


- **Daly (1998) extended Kelly's model to account for target tracking by the eye movements.**

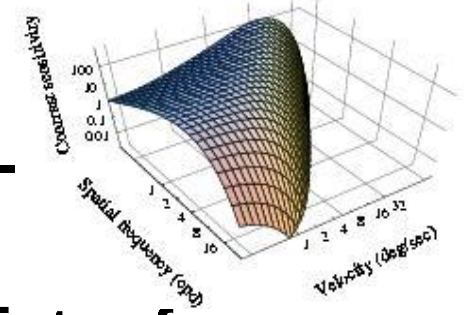
Deriving Pixel Flow Using

Image-Based Rendering Techniques





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.**

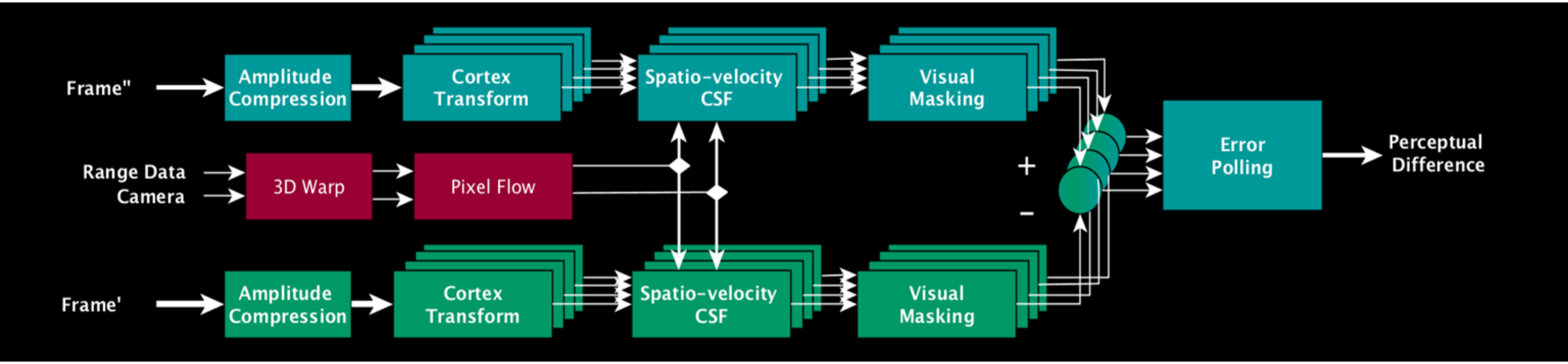


Image-based Rendering for Animations

- **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.**

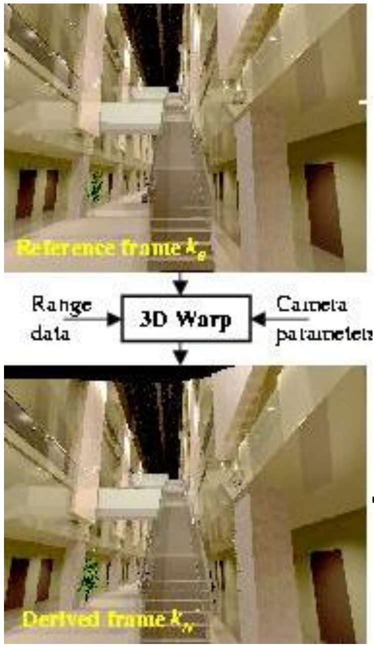
Keyframe Placement

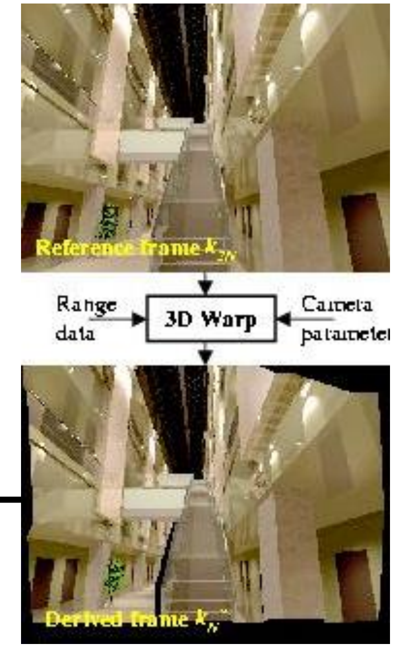
- **The selection of keyframes should be considered in the context of the inbeteween frame computation technique.**
- **In IBR techniques reference frames are usually placed:**
  - uniformly in space at the nodes of 2D or 3D grid (Chen95),
  - uniformly along the animation path (Mark97),
  - at manually 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 ?

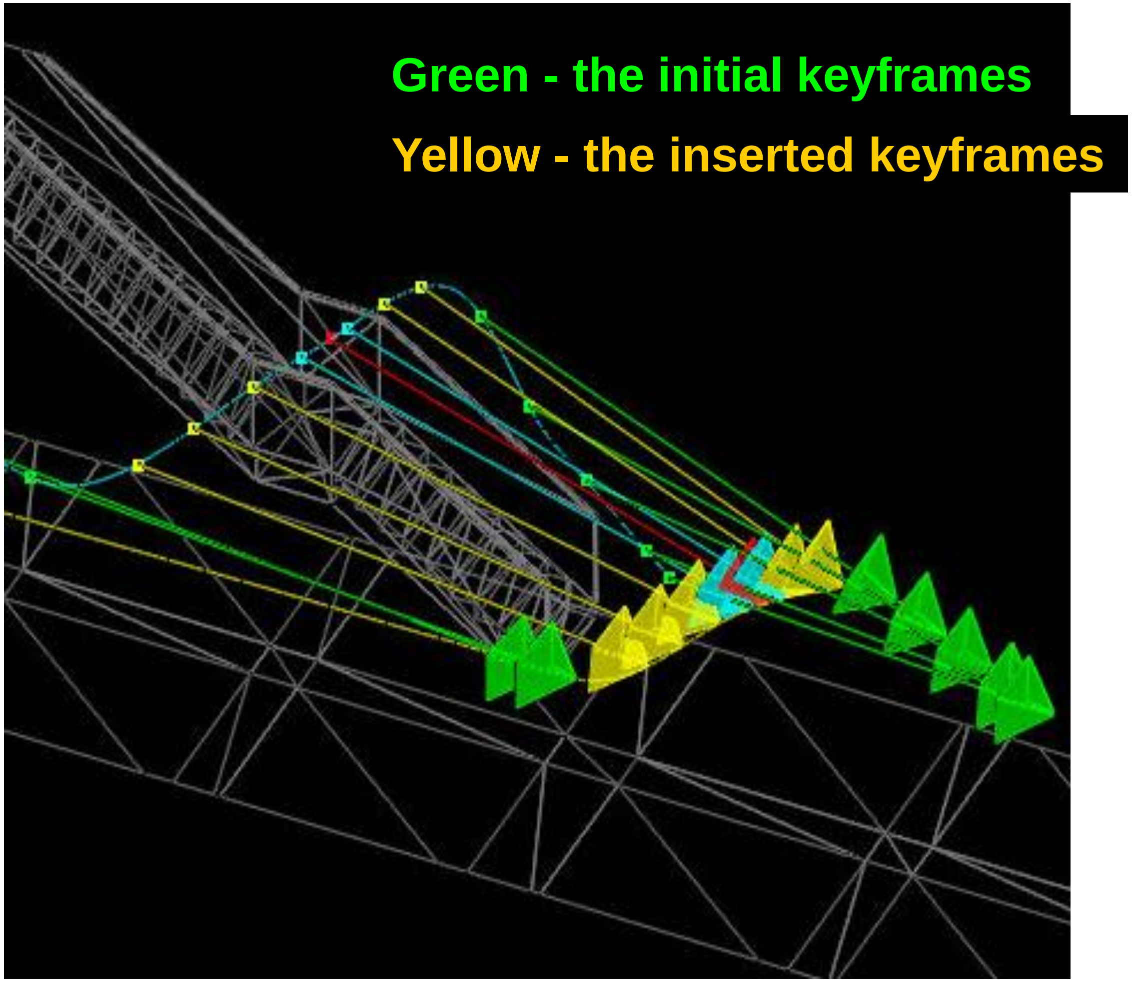
NO

- **Split segment**
- **Recurse**

YES

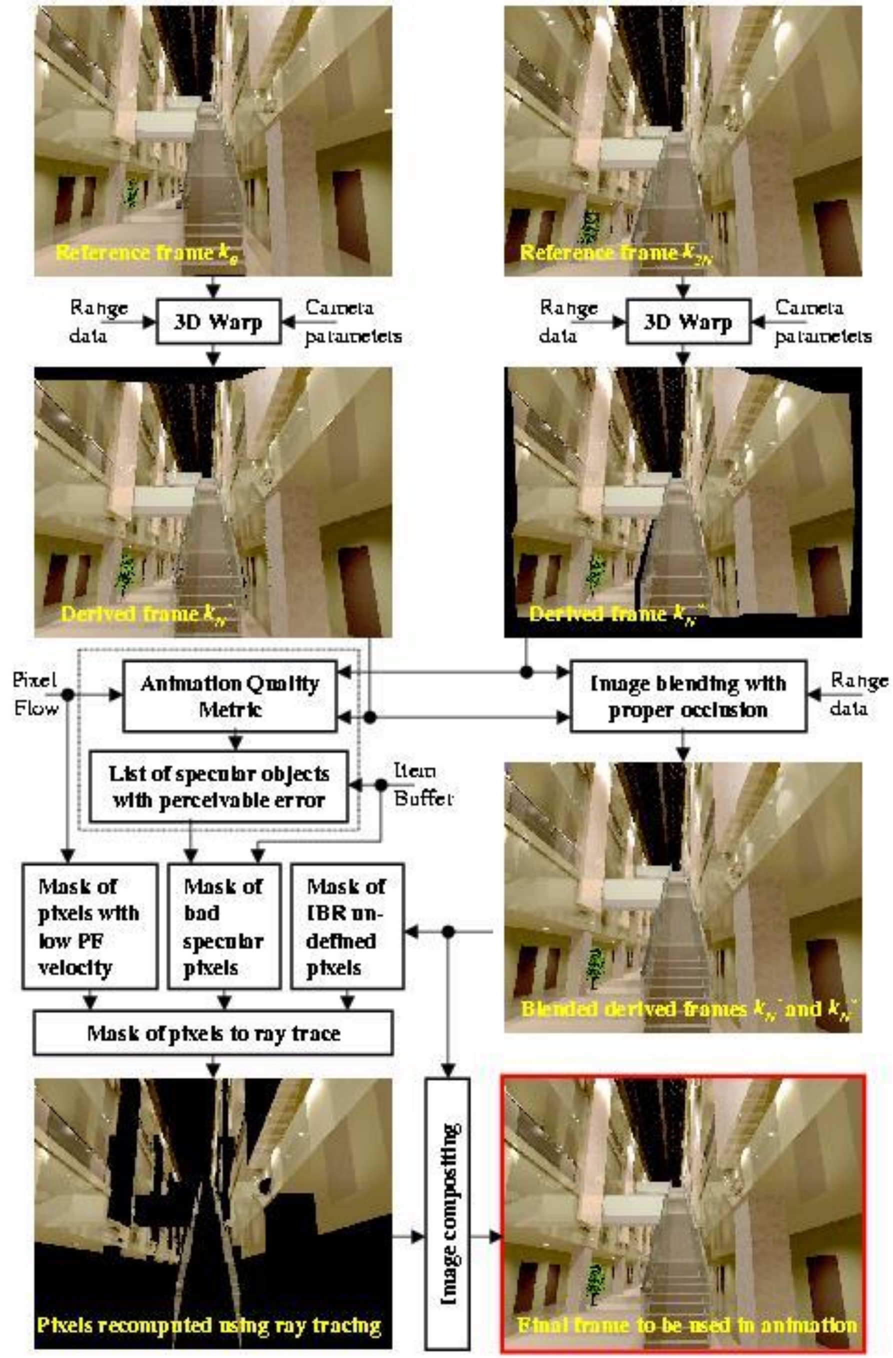
Generate inbetween images

Atrium: final keyframe placement

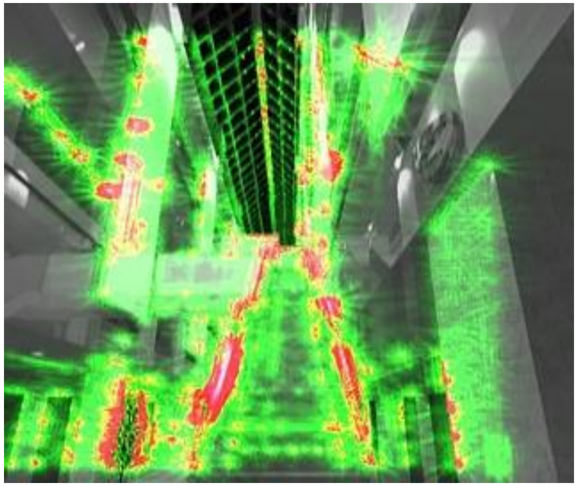


Animation path with marked keyframe locations

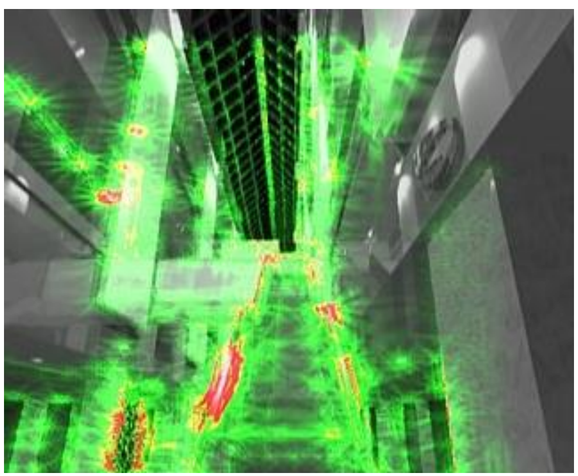
In-between frame generation



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

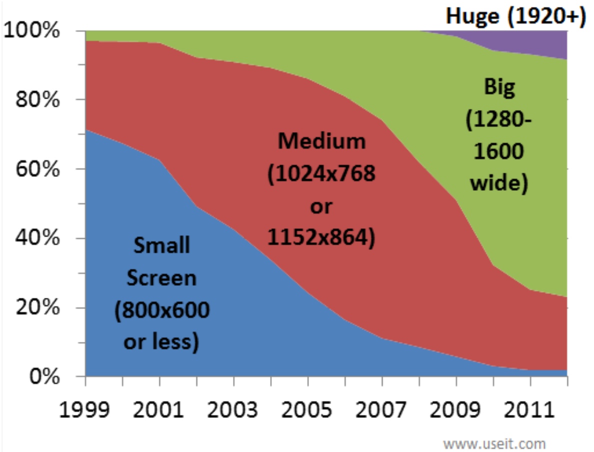
Supersampled frame used in traditional animations

Corresponding frame derived using spatiotemporal filtering

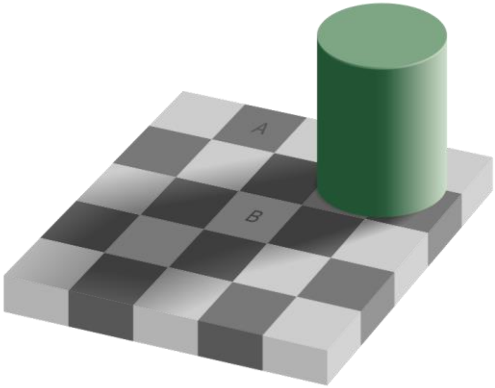
In both cases the perceived quality of animation appears to be similar!

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



Evolution of computer screen sizes



Checker shadow illusion

Images adapted from https://www.nngroup.com/articles/computer-screens-getting-bigger/

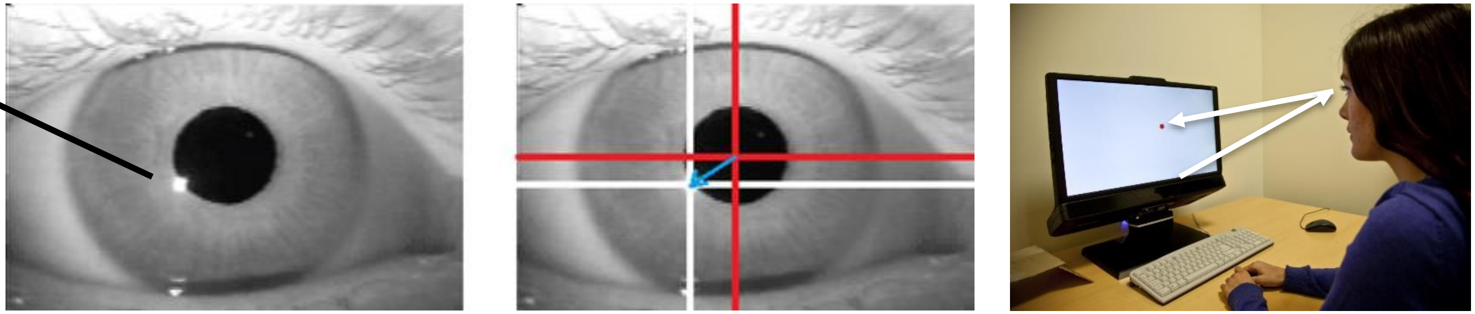
Eye Tracking - Outline

- **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:**

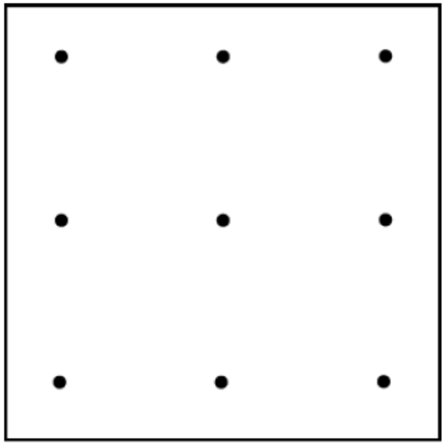
Corneal Reflection (also known as "glint" or "1st Purkinje Reflection")



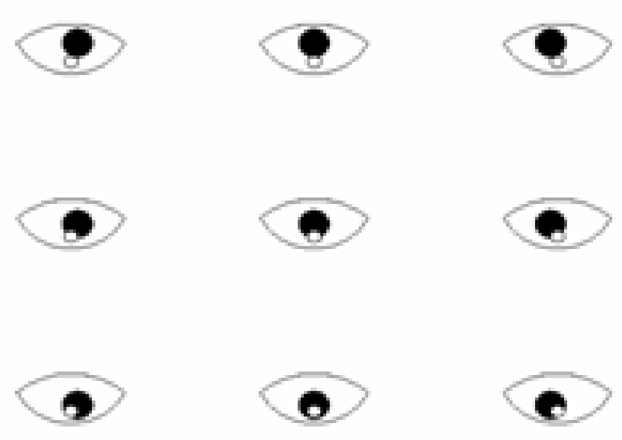
- **Eye trackers mostly operate using infrared imaging technology**
- **Once the pupil is detected the vector between the center of the pupil and the corneal reflection of the infrared light source is translated into the gaze location on screen coordinates**
- **Requires calibration at the beginning**

Images adapted from http://twiki.cis.rit.edu/twiki/bin/view/MVRL/QuadTracker and http://psy.sabanciuniv.edu

Eye Tracking



Sample 9-point calibration grid



Relative positions of the pupil and the corneal reflection

- **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

Eye Tracking

Chin-rest (EyeLink 1000/2000)

Glasses (SMI Eye Tracking Glasses)

Head-mounted displays (Oculus Rift)

- **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**

Images adapted from http://web.ntnu.edu.tw, http://youtube.com and http://techinsider.io

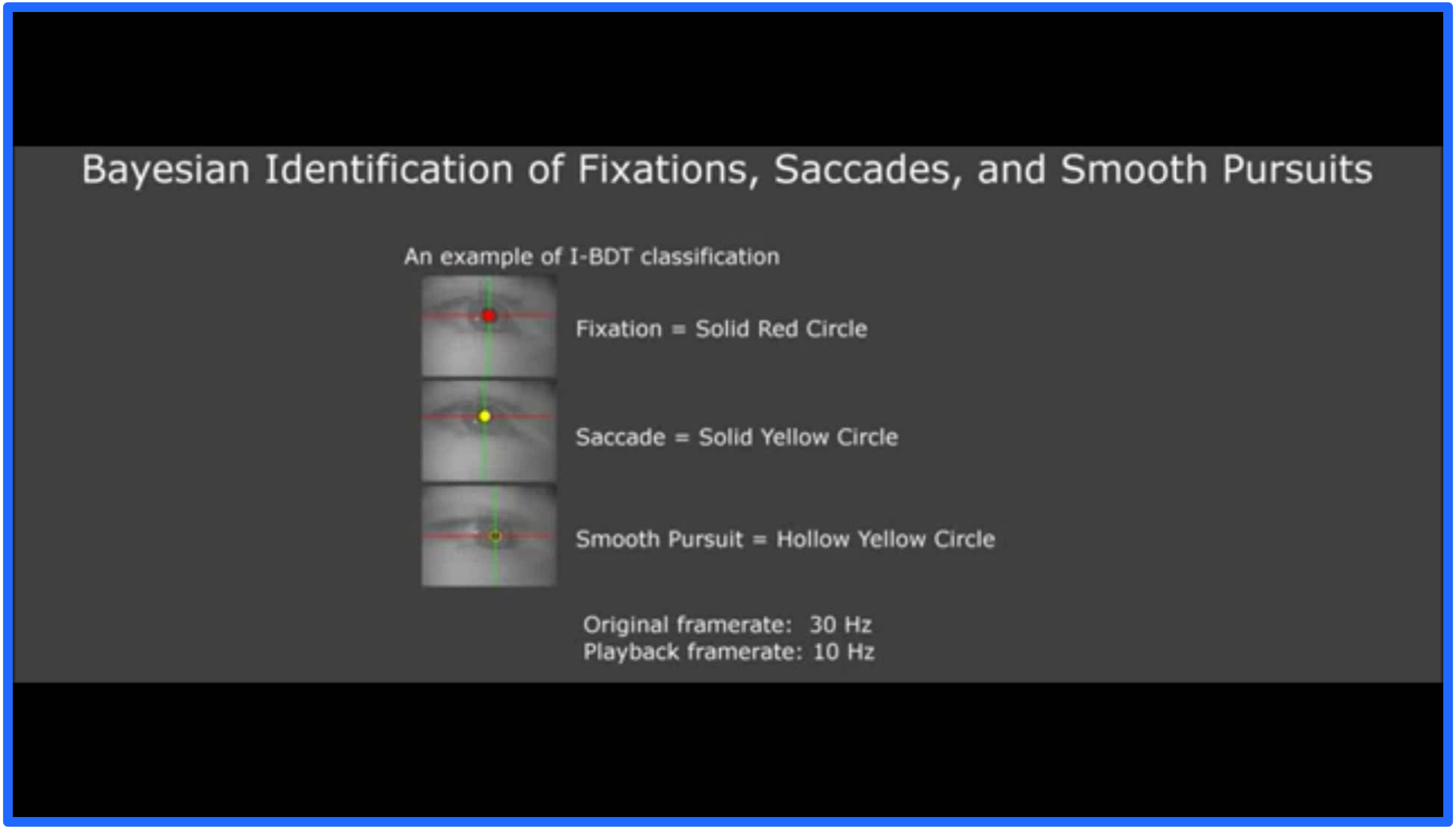
Types of Eye Motion

Type	Duration (ms)	Amplitude ($1^{\circ} = 60'$)	Velocity
Fixation	200-300	-	-
Microsaccade	10-30	10-40’	15-50°/s
Tremor	-	<1’	20’/sec
Drift	200-1000	1-60’	6-25’/s
Saccade	30-80	4-20°	30-500°/s
Glissade	10-40	0.5-2°	20-140°/s
Smooth Pursuit	variable	variable	10-30°/s

- 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.

Reference: Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. OUP Oxford.

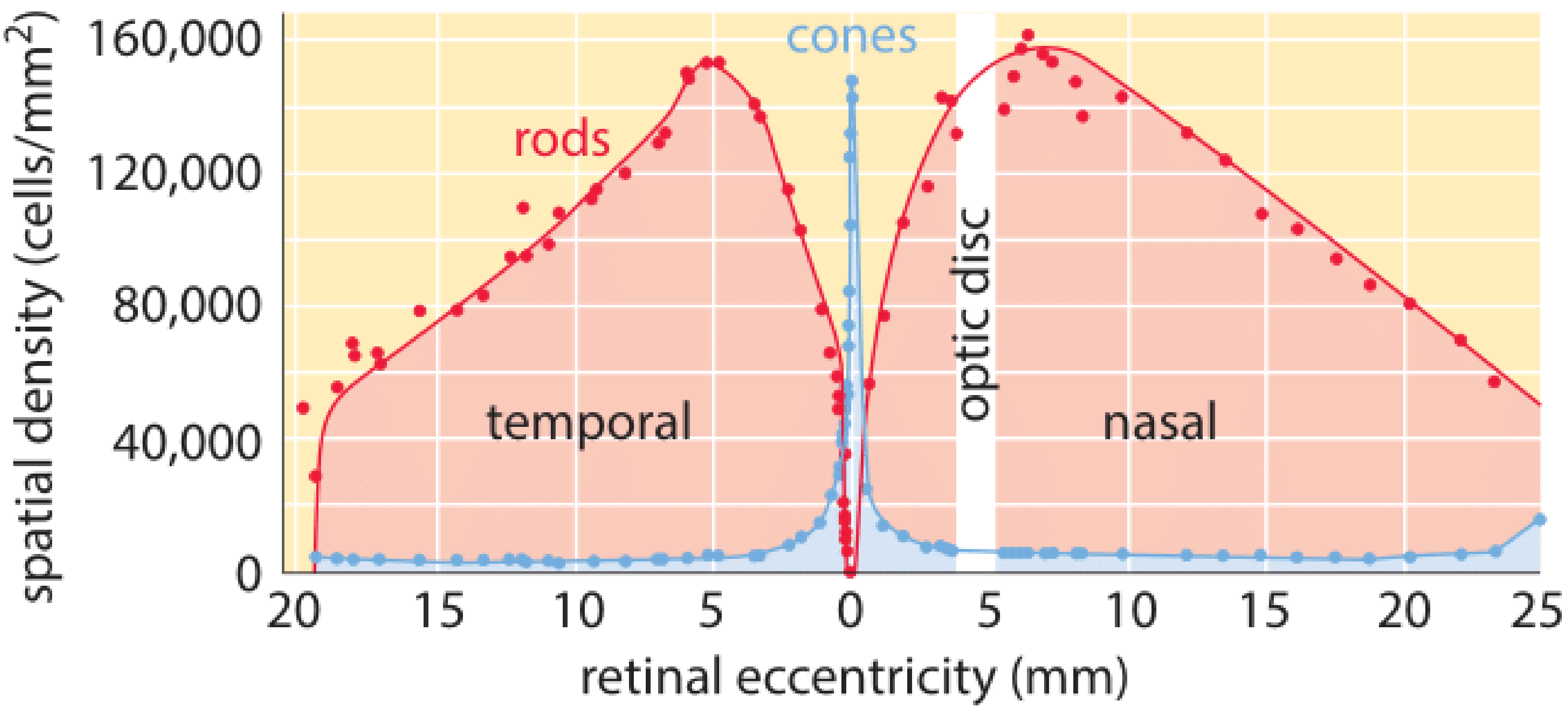
Eye Tracking in Action



Adapted from T. Santini, W. Fuhl, T. Kübler, and E. Kasneci. Bayesian Identification of Fixations, Saccades, and Smooth Pursuits ACM Symposium on Eye Tracking Research & Applications, ETRA 2016.

Visual Acuity

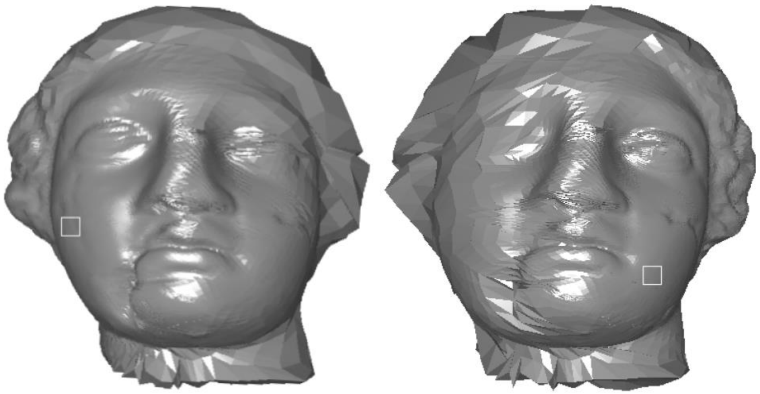
- **Distribution of photoreceptor cells in the retina**



Adapted from R. W. Rodieck, The First Steps of Seeing, Sinauer Associates, 1998.

Level-of-Detail Rendering

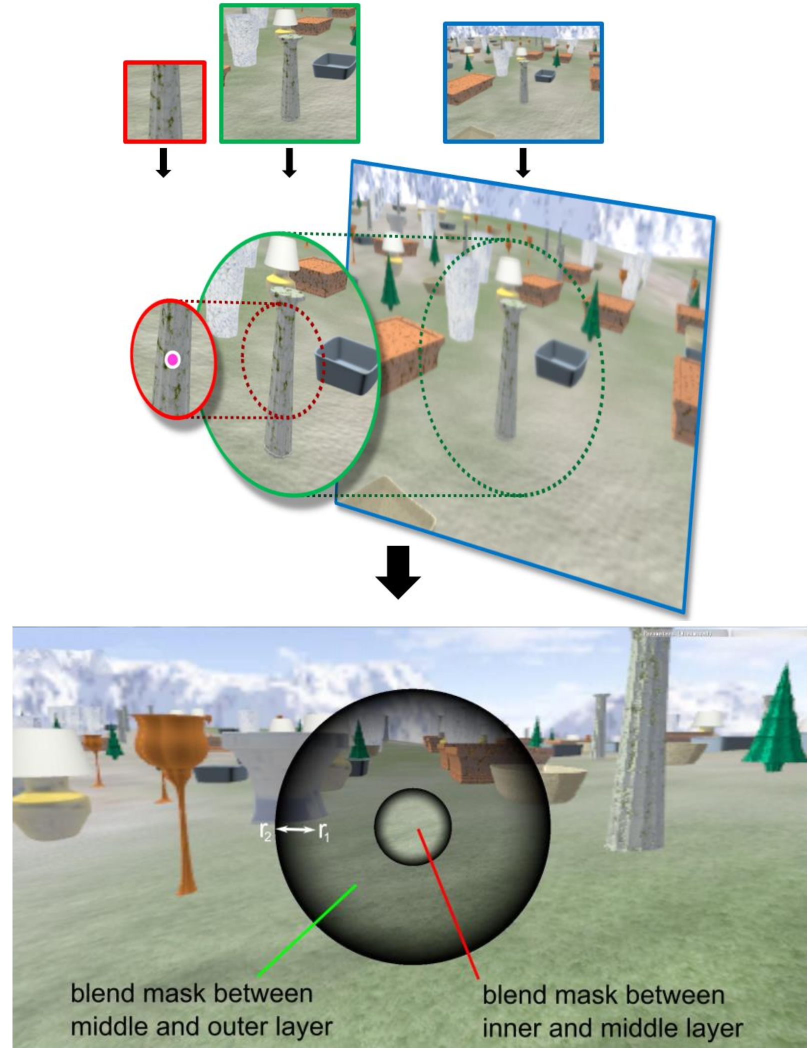
- **The model resolution may be degraded according to the visual angle and the acuity of HVS at the given angle**
  - Mesh structure of the model is partitioned into tiles using Voronoi diagram
  - Tiles are mapped to planar polygons
  - Remeshing into multiresolution form



Adapted from Murphy, Hunter, and Andrew T. Duchowski. "Gaze-contingent level of detail rendering." EuroGraphics 2001 (2001).

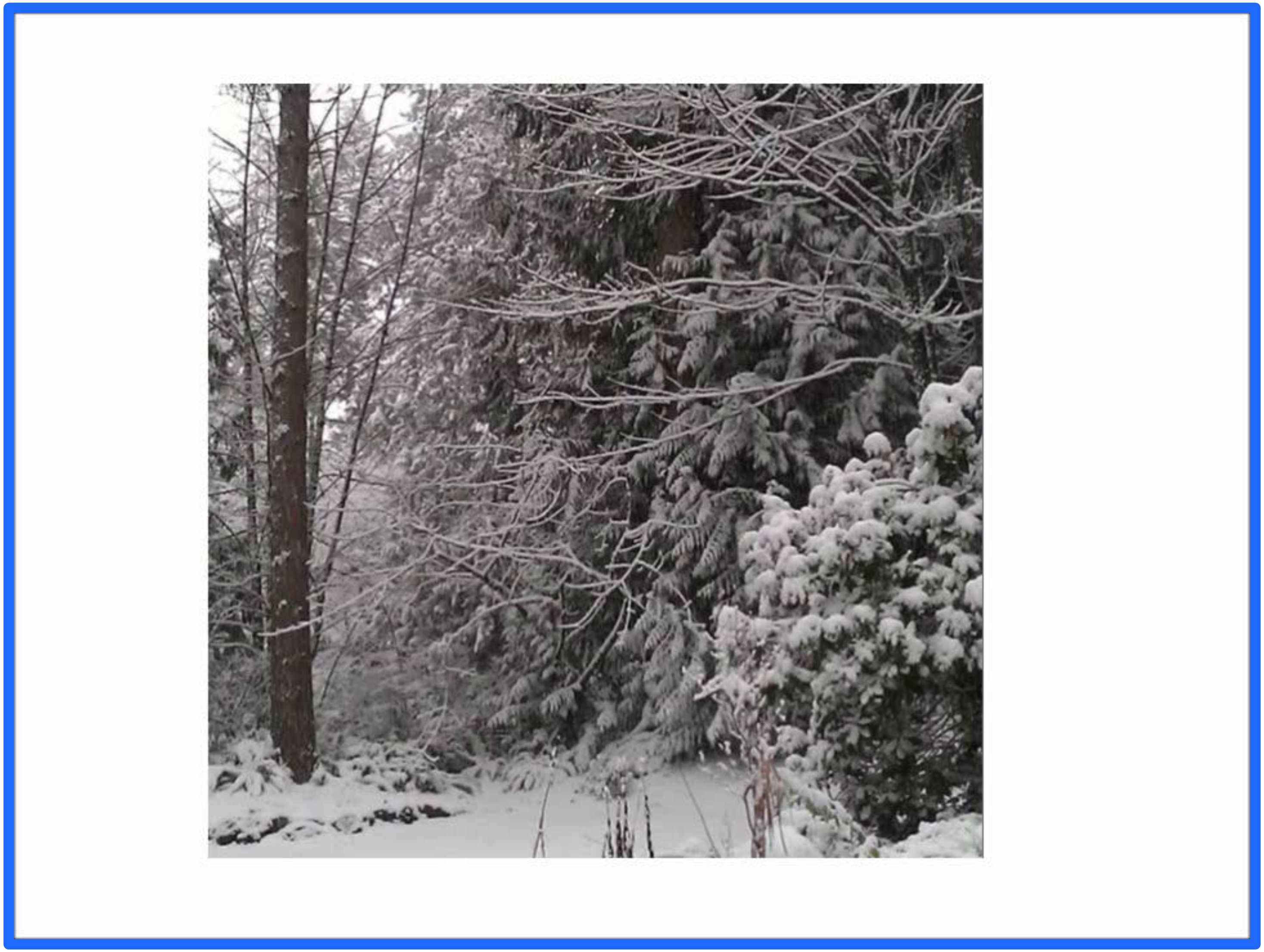
Foveated 3D Graphics

- **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 <u>caused artifacts</u> for the observer**
- **Results using** 120Hz **monitor and Tobii TX300 with** 300Hz **sampling frequency and** 10ms **latency were <u>tolerable</u>**



Images adapted from Guenter, B., Finch, M., Drucker, S., Tan, D., & Snyder, J. (2012). Foveated 3D graphics. ACM Transactions on Graphics (TOG), 31(6), 164.

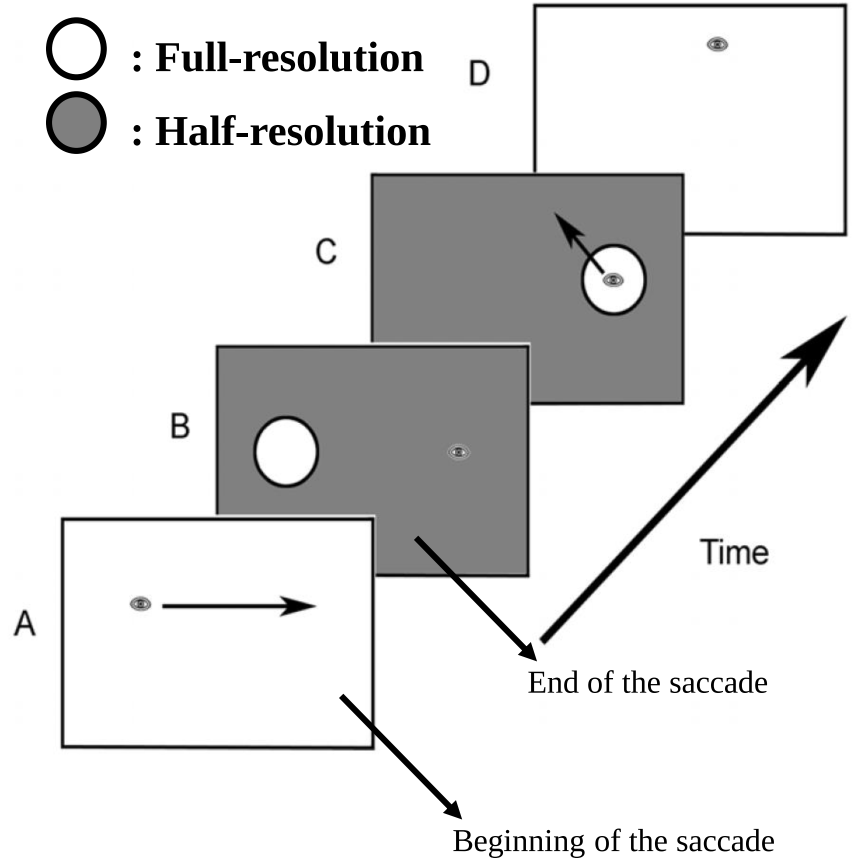
Foveated 3D Graphics



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

Latency Measurement

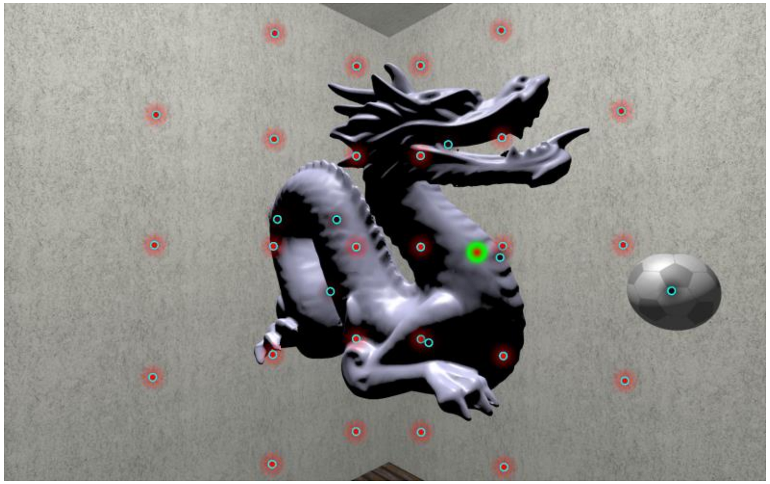
- **Transition from B** (end of the saccade) **to C** (switching from half to full-resolution in the gaze location)**:**
  - 5, 20, 40, 60 or 80 ms are tested
  - Viewers never detected a change up to a delay of 5 ms after the saccade is completed
- **E2:** the retinal eccentricity where resolution drops 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



Images adapted from Loschky, L. C., & Wolverton, G. S. (2007). How late can you update gaze-contingent multiresolutional displays without detection?. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 3(4), 7.

Overcoming Eye Tracker Noise

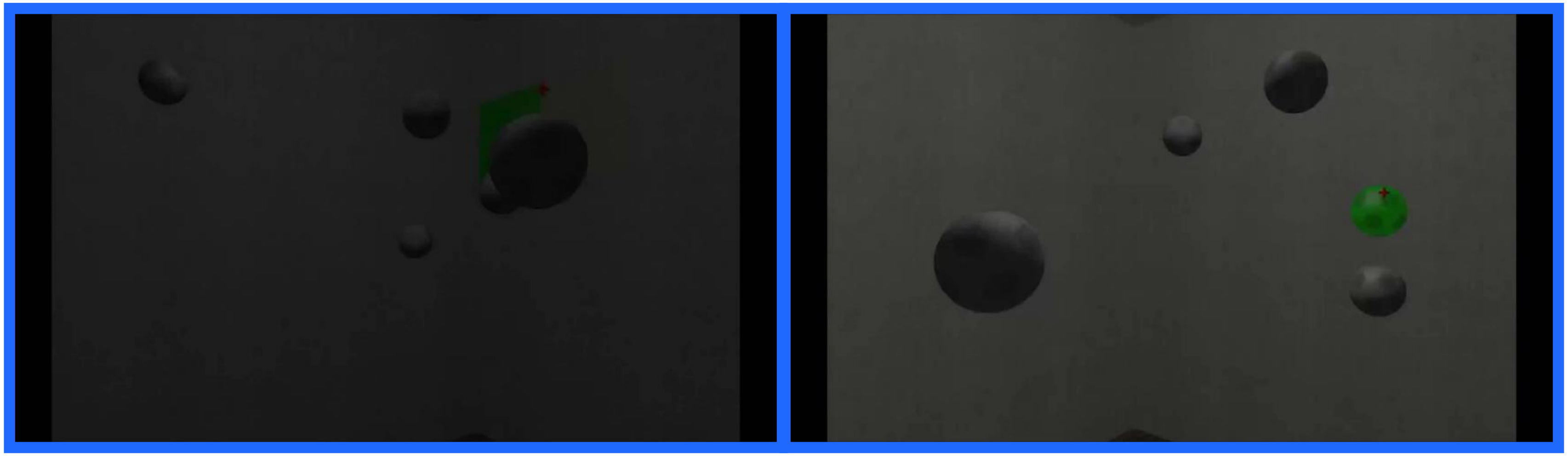
- Accuracy of existing eye trackers is insufficient for gaze-driven Depth-of-Field (DoF) applications
  - P-CR RED250 tracker
    Claimed: 0.5°
    Measured: 1.83° std: 1.07°
- Gaze accuracy is improved by "snapping" the gaze location to the nearest potential focus-point using the information from tracker and 3D scene (including **focus-point position** and **velocity**)



Potential focus-point markers

Images adapted from Mantiuk, Radoslaw, Bartosz Bazyluk, and Rafal K. Mantiuk. "Gaze-driven Object Tracking for Real Time Rendering." Computer Graphics Forum. Vol. 32. No. 2pt2. Blackwell Publishing Ltd, 2013.

Overcoming Noise



Videos adapted from Mantiuk, Radoslaw, Bartosz Bazyluk, and Rafal K. Mantiuk. "Gaze-driven Object Tracking for Real Time Rendering." Computer Graphics Forum. Vol. 32. No. 2pt2. Blackwell Publishing Ltd, 2013.

Effect of Depth-of-Field

- **Improves the rendering realism and enhances the depth perception**

(a) Image focused on objects at shallow depth (flower)

(b) Image focused on objects at large depth (Main Quad)

(c) Image with everything in focus

Images adapted from Gupta, Kushagr, and Suleman Kazi, “Gaze Contingent Depth of Field Display”, 2016. Video adapted from Mantiuk, Radoslaw, Bartosz Bazyluk, and Rafal K. Mantiuk. "Gaze-driven Object Tracking for Real Time Rendering." Computer Graphics Forum. Vol. 32. No. 2pt2. Blackwell Publishing Ltd, 2013.

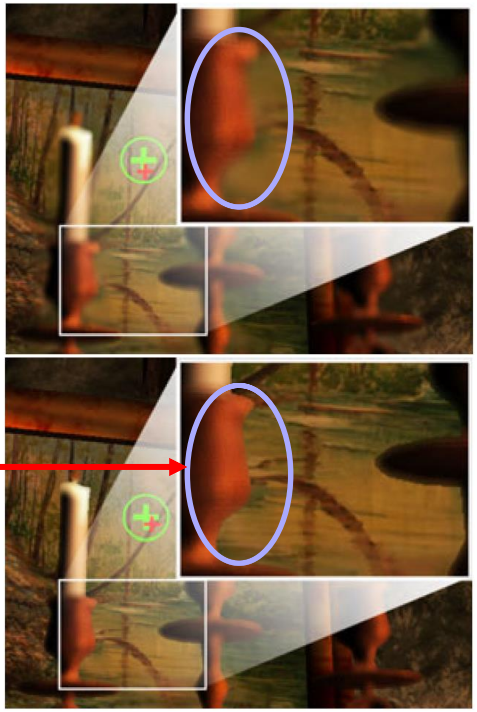
Depth-of-Field Rendering

- **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

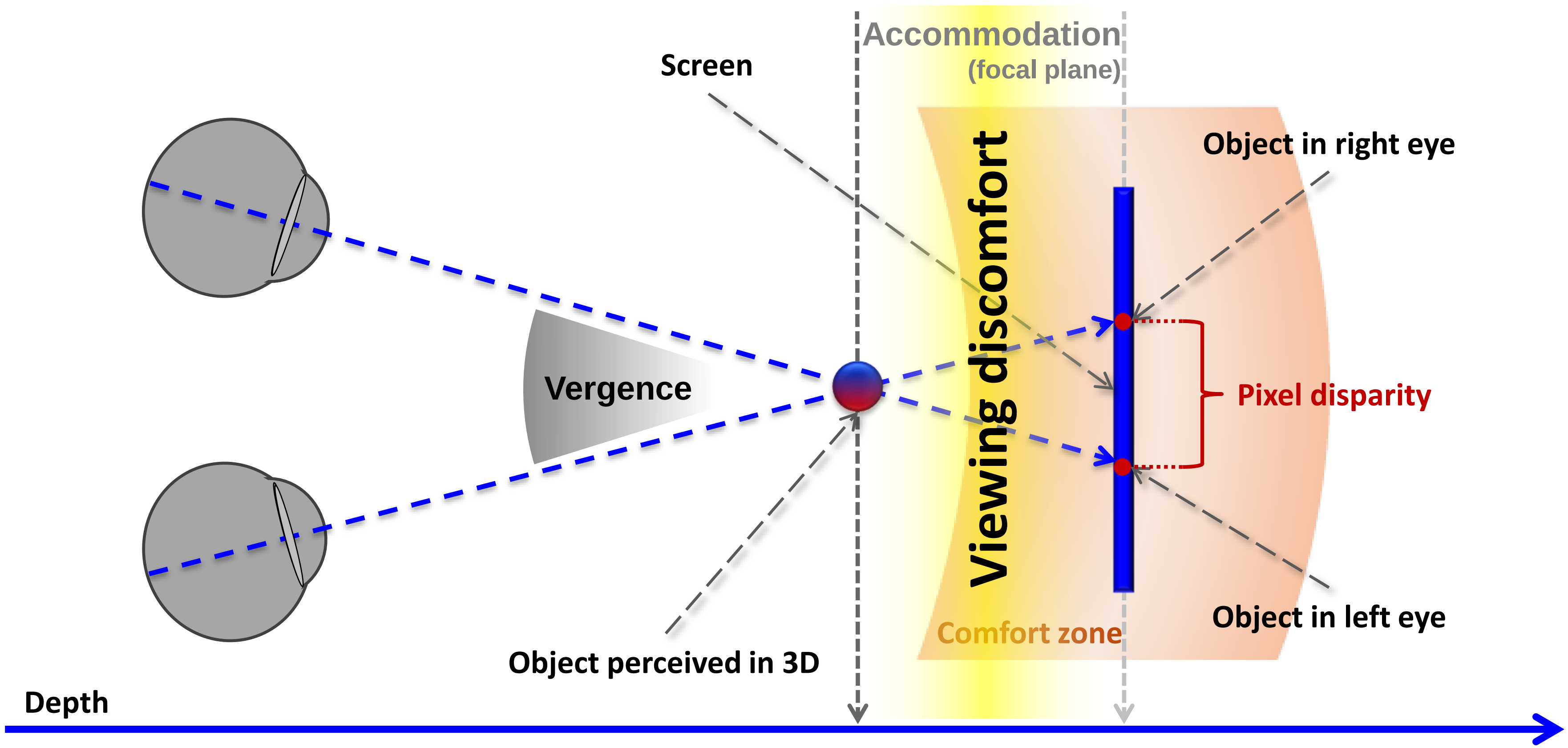
- d_p is obtained from reverse mapping of the z-buffer
- **Addresses the artifacts due to the depth discontinuity near object boundaries by spreading the blur outside the object boundary**



Images adapted from Mantiuk, R., Bazyluk, B., & Tomaszewska, A. (2011). Gaze-dependent depth-of-field effect rendering in virtual environments. In Serious Games Development and Applications (pp. 1-12). Springer Berlin Heidelberg.

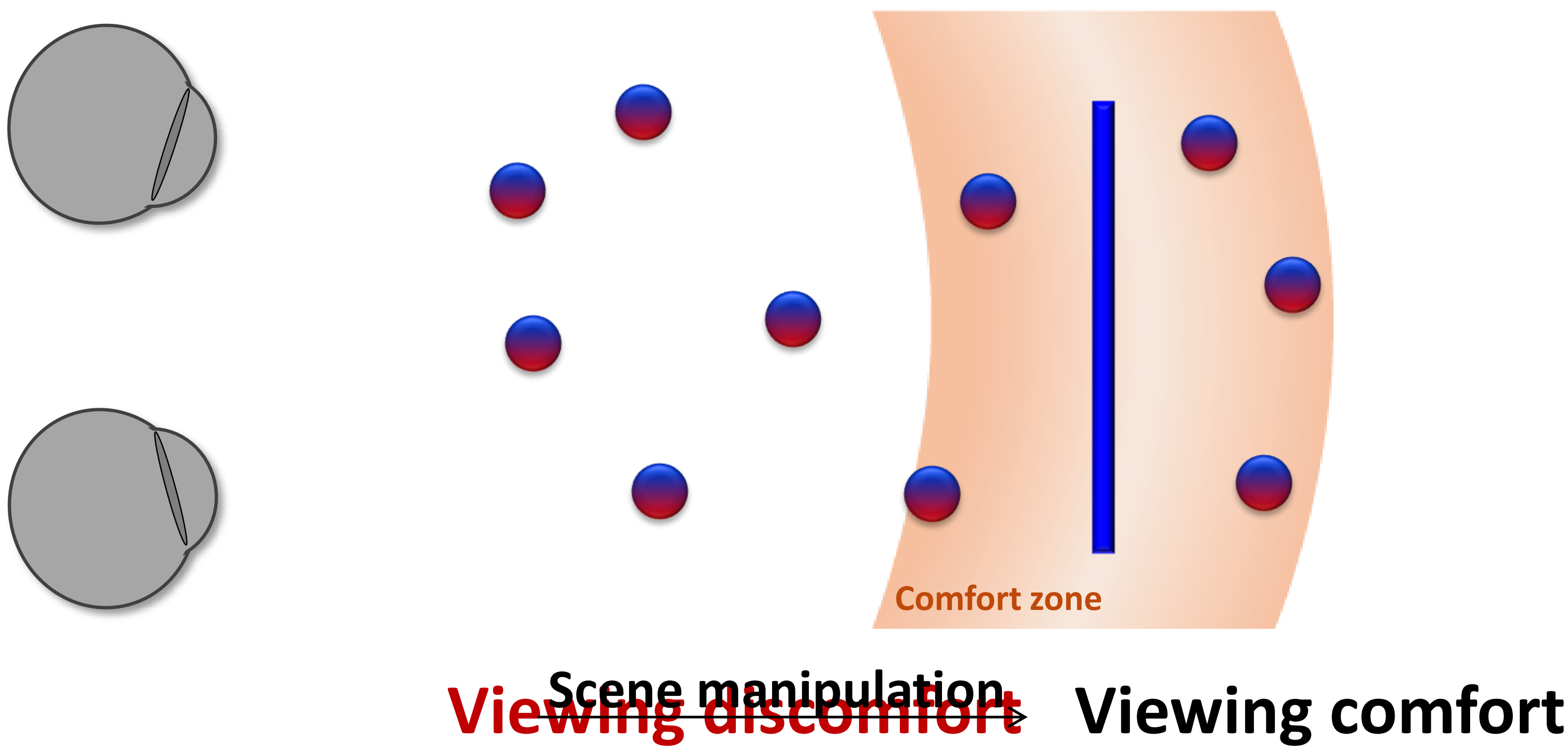
Vergence-accommodation Conflict

Stereo 3D: Binocular Disparity



Vergence-accommodation Conflict

Depth Manipulation

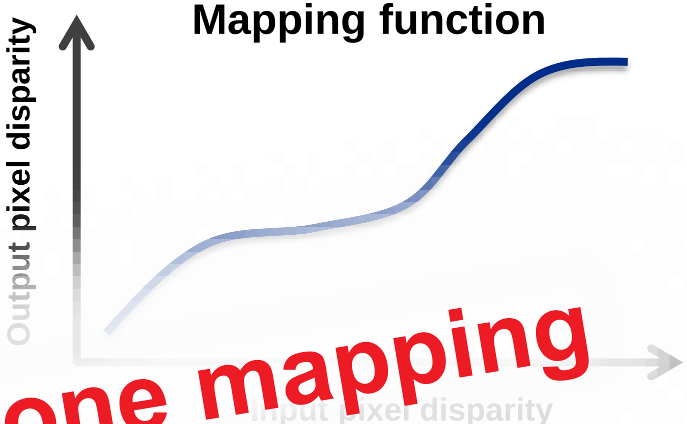


Vergence-accommodation Conflict

Depth Manipulation

Pixel disparity map

Modified pixel disparity



Similar to tone mapping

Function:

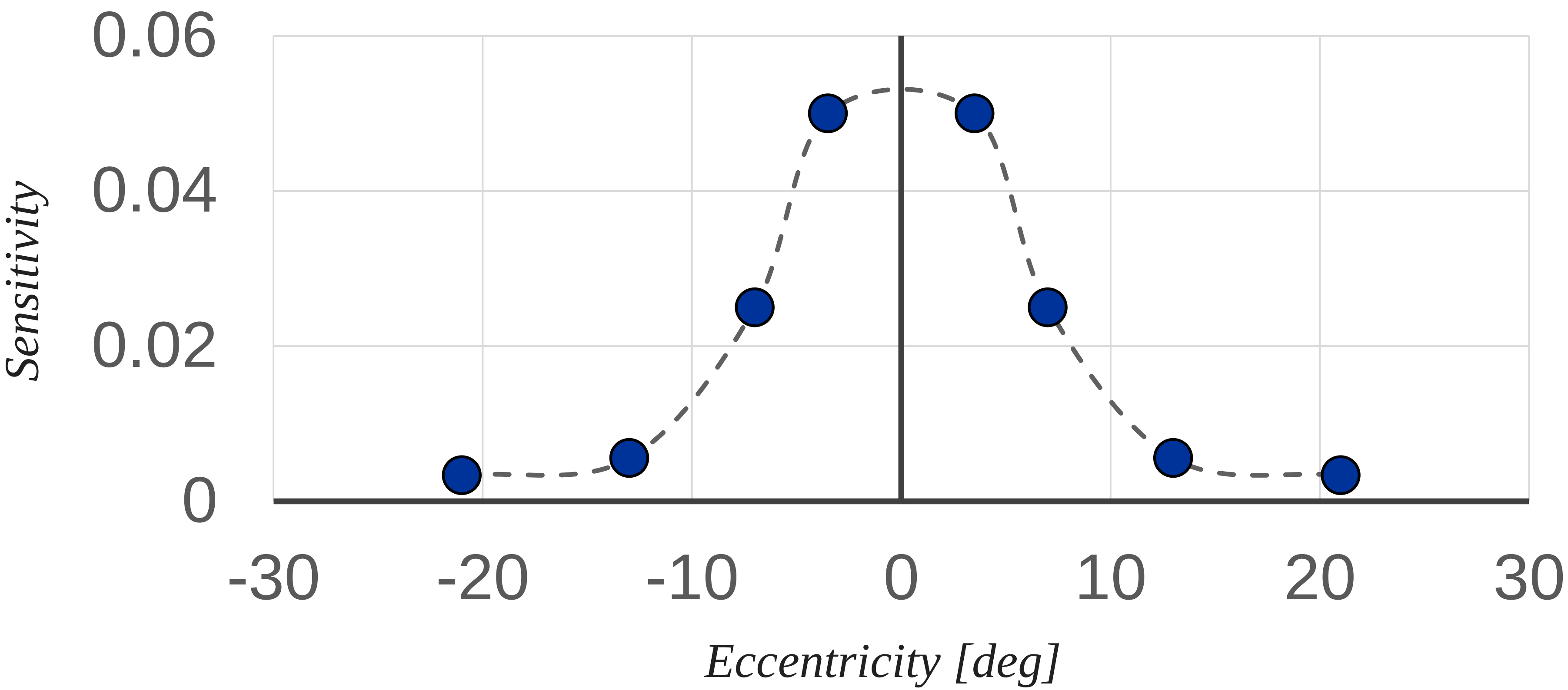
- Linear
- Logarithmic
- Content dependent

Other possibilities:

- Gradient domain
- Local operators

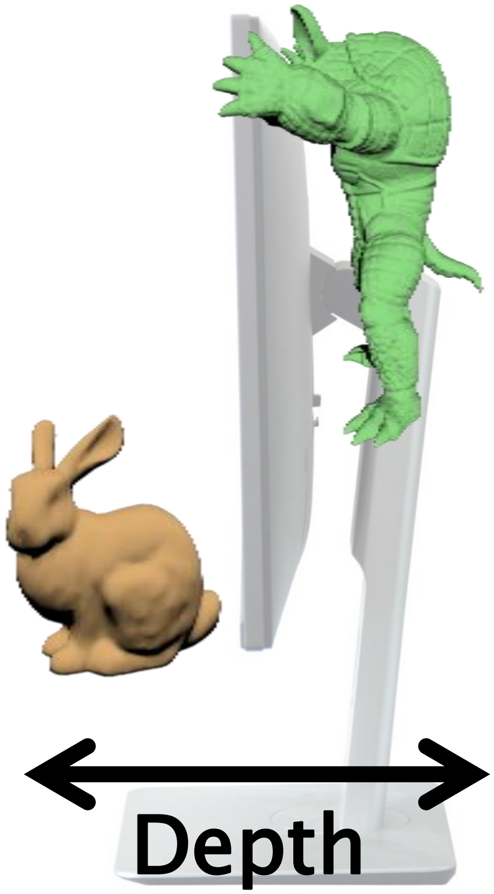
"Nonlinear Disparity Mapping for Stereoscopic 3D" [Lang et al. 2010]

Disparity Perception (Stereo 3D)

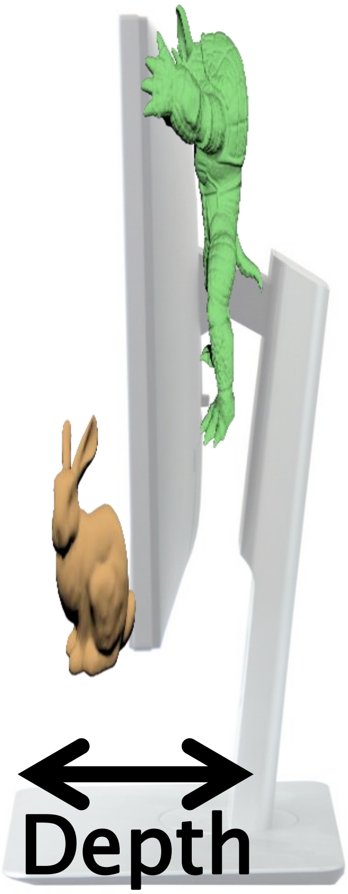


Replotted from Figure 3 of Simon J.D Prince, Brian J Rogers
Sensitivity to disparity corrugations in peripheral vision, Vision Research, Volume 38, Issue 17, September 1998

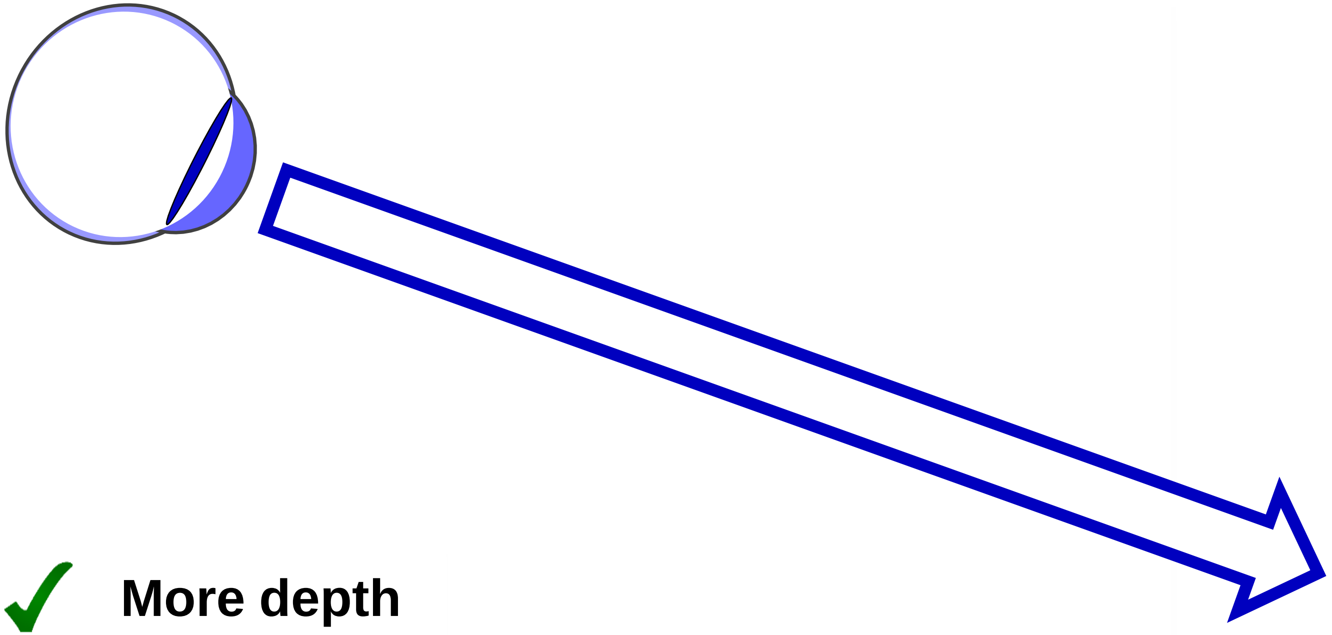
Vergence-accommodation Conflict



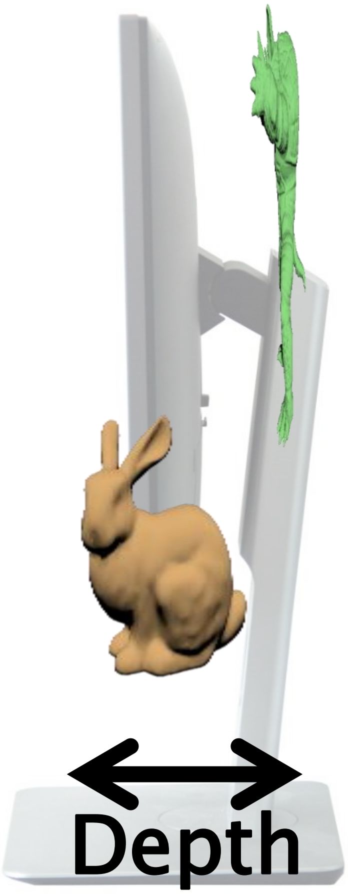
Vergence-accommodation Conflict



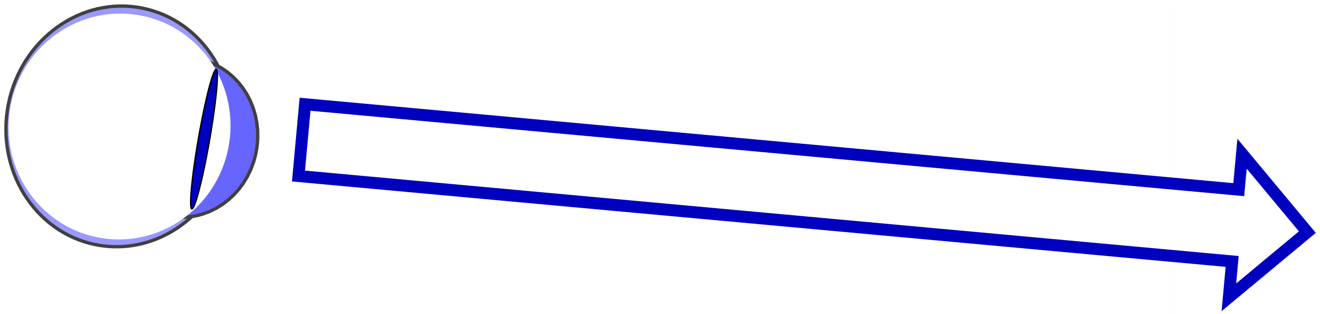
Vergence-accommodation Conflict

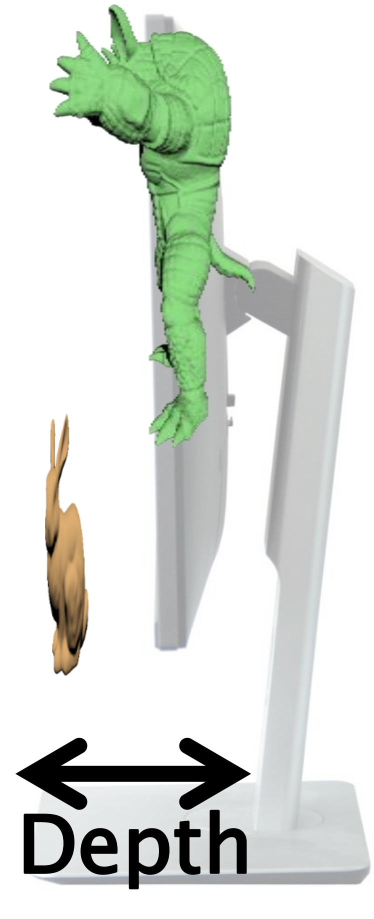


✓ **More depth**



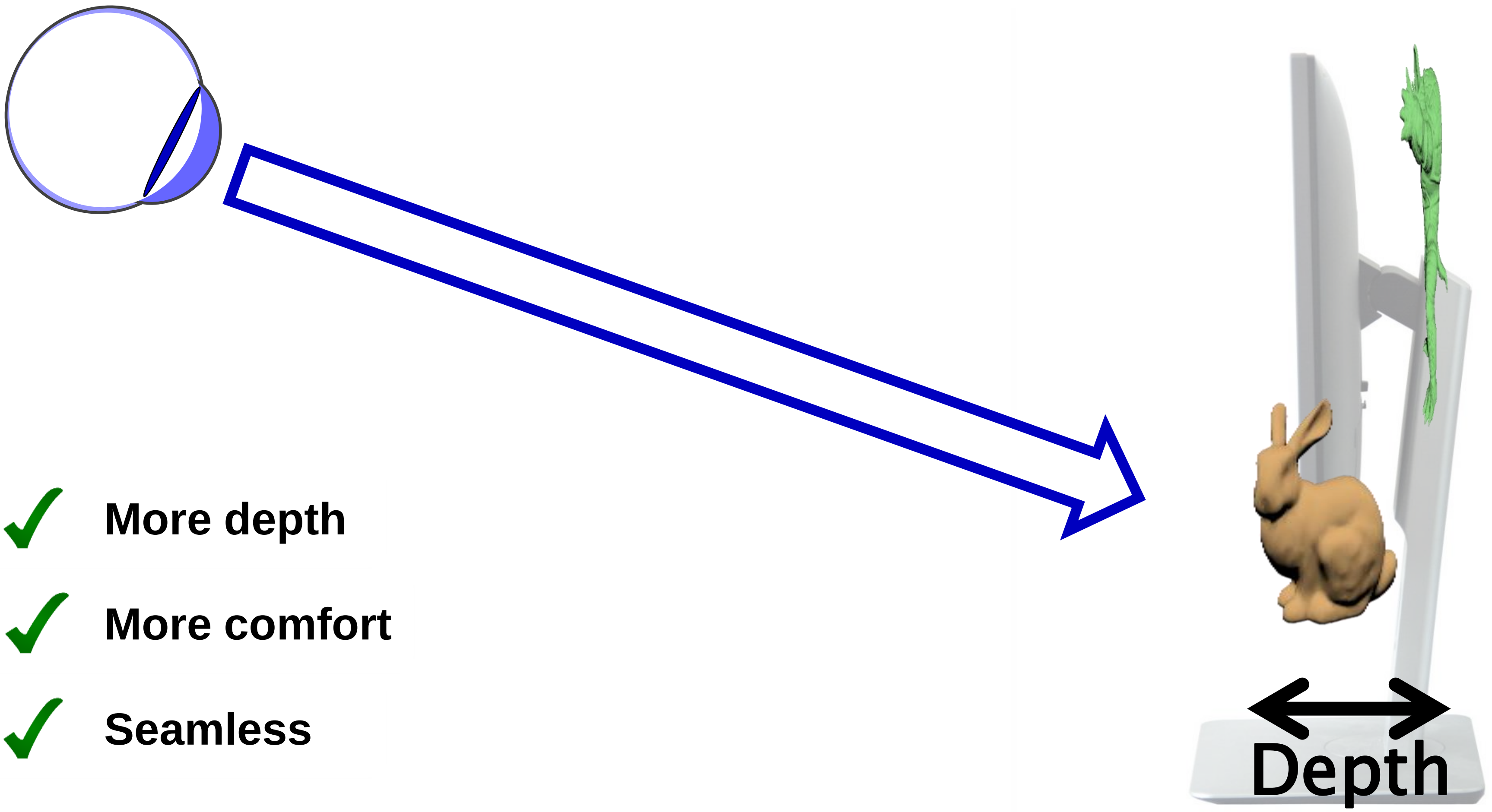
Vergence-accommodation Conflict





- ✓ **More depth**
- ✓ **More comfort**

Vergence-accommodation Conflict

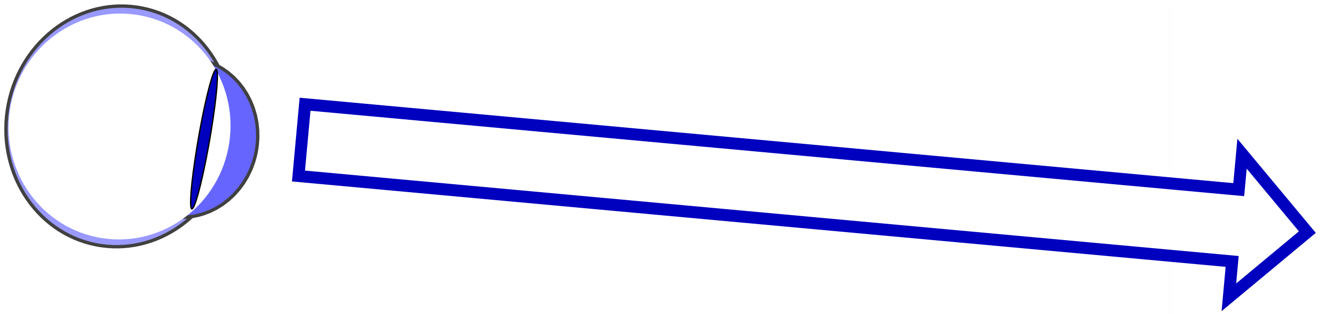


✓ **More depth**

✓ **More comfort**

✓ **Seamless**

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

Images adapted from Koulieris, George Alex, et al. "Gaze Prediction using Machine Learning for Dynamic Stereo Manipulation in Games." IEEE Virtual Reality. 2016.

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)

Images adapted from E Jacobs, D., Gallo, O., A Cooper, E., Pulli, K., & Levoy, M. (2015). Simulating the visual experience of very bright and very dark scenes. ACM Transactions on Graphics (TOG), 34(3), 25.

Local Adaptation

- **Adaptation to global lighting:**

$$A' \leftarrow \begin{cases} A + a_1 \Delta t & A < A_T \\ A - a_2 \Delta t & A > A_T \end{cases}$$

A: adaptation level in the previous time-step (in log-units)
A_T: luminance level of the target
a_1: adaptation rate if the target is brighter
a_2: adaptation rate if the target is darker

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

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

Global adaptation with respect to the gaze location (red arrow).

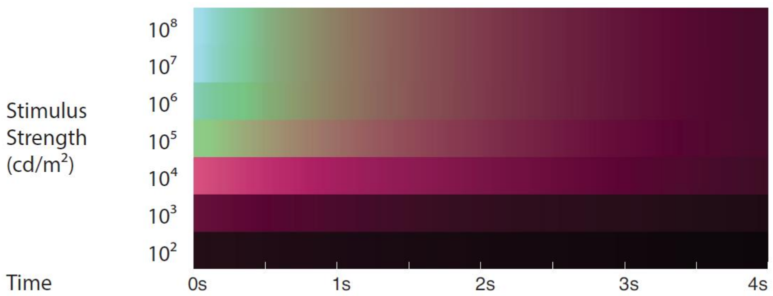
Images adapted from E Jacobs, D., Gallo, O., A Cooper, E., Pulli, K., & Levoy, M. (2015). Simulating the visual experience of very bright and very dark scenes. ACM Transactions on Graphics (TOG), 34(3), 25.

Local Adaptation

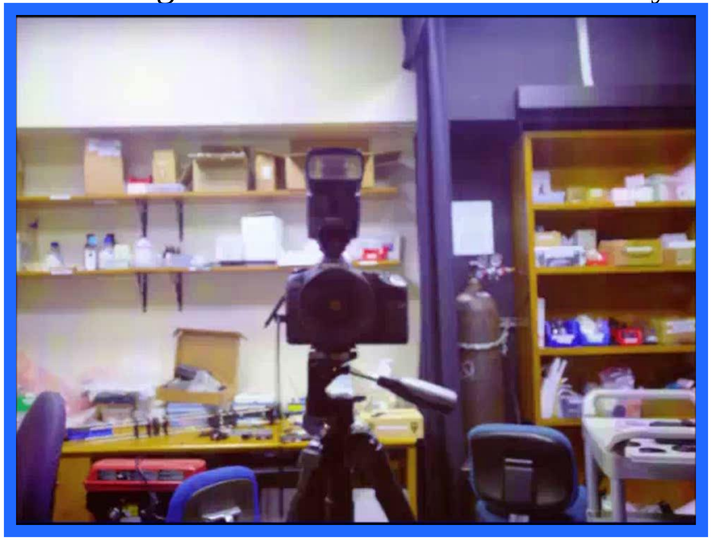
- **Afterimage:** Image of the stimuli which is still perceived after it ceases
- **May be in the form of:**
  - Bleaching afterimages
  - Local adaptation afterimages
- **Bleaching level B is given in the form of a differential equation (Baylor et al. 1974):**

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

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



Bleaching w.r.t. time and stimulus intensity.



Images adapted from E Jacobs, D., Gallo, O., A Cooper, E., Pulli, K., & Levoy, M. (2015). Simulating the visual experience of very bright and very dark scenes. ACM Transactions on Graphics (TOG), 34(3), 25.

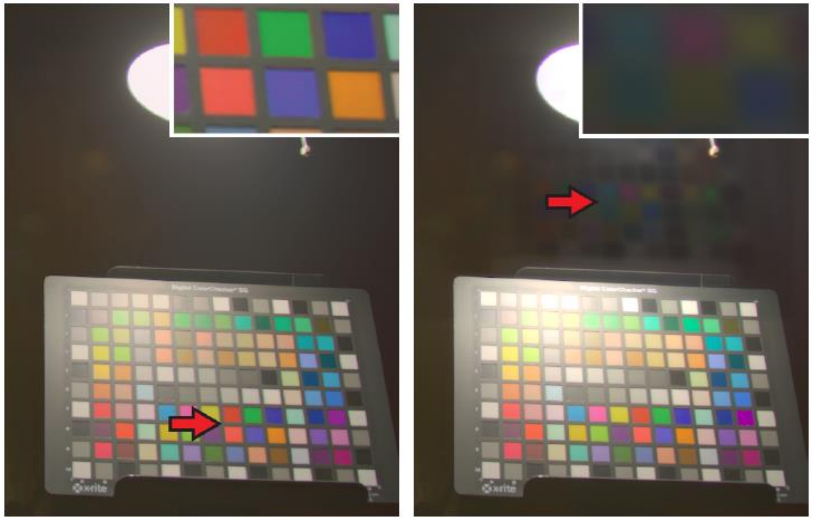
Local Adaptation

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

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

C'**:** calcium concentration in the new timestep
C_∞**:** 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.**

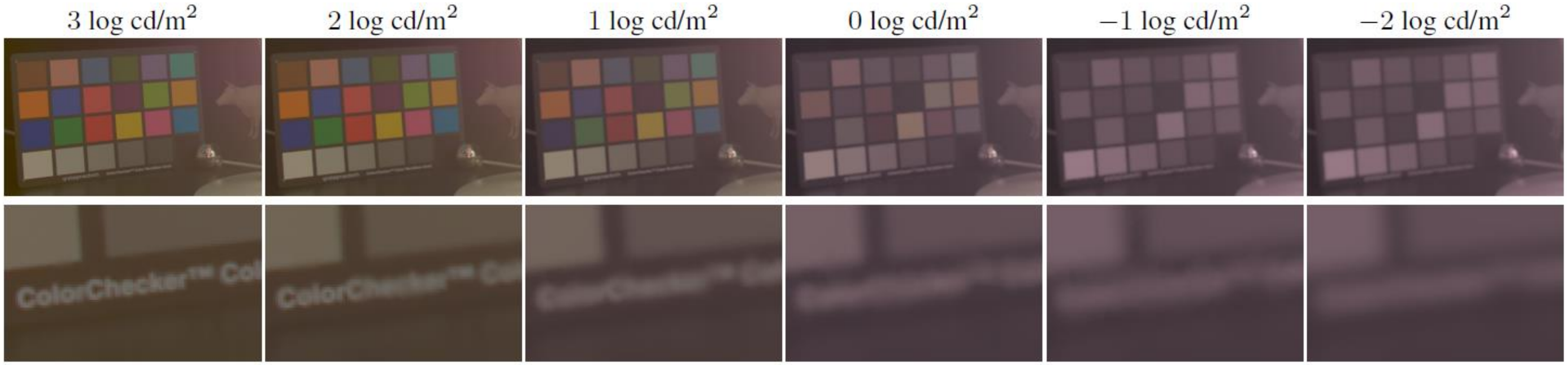


Local adaptation afterimage. Red arrows show the gaze position.

Images adapted from E Jacobs, D., Gallo, O., A Cooper, E., Pulli, K., & Levoy, M. (2015). Simulating the visual experience of very bright and very dark scenes. ACM Transactions on Graphics (TOG), 34(3), 25.

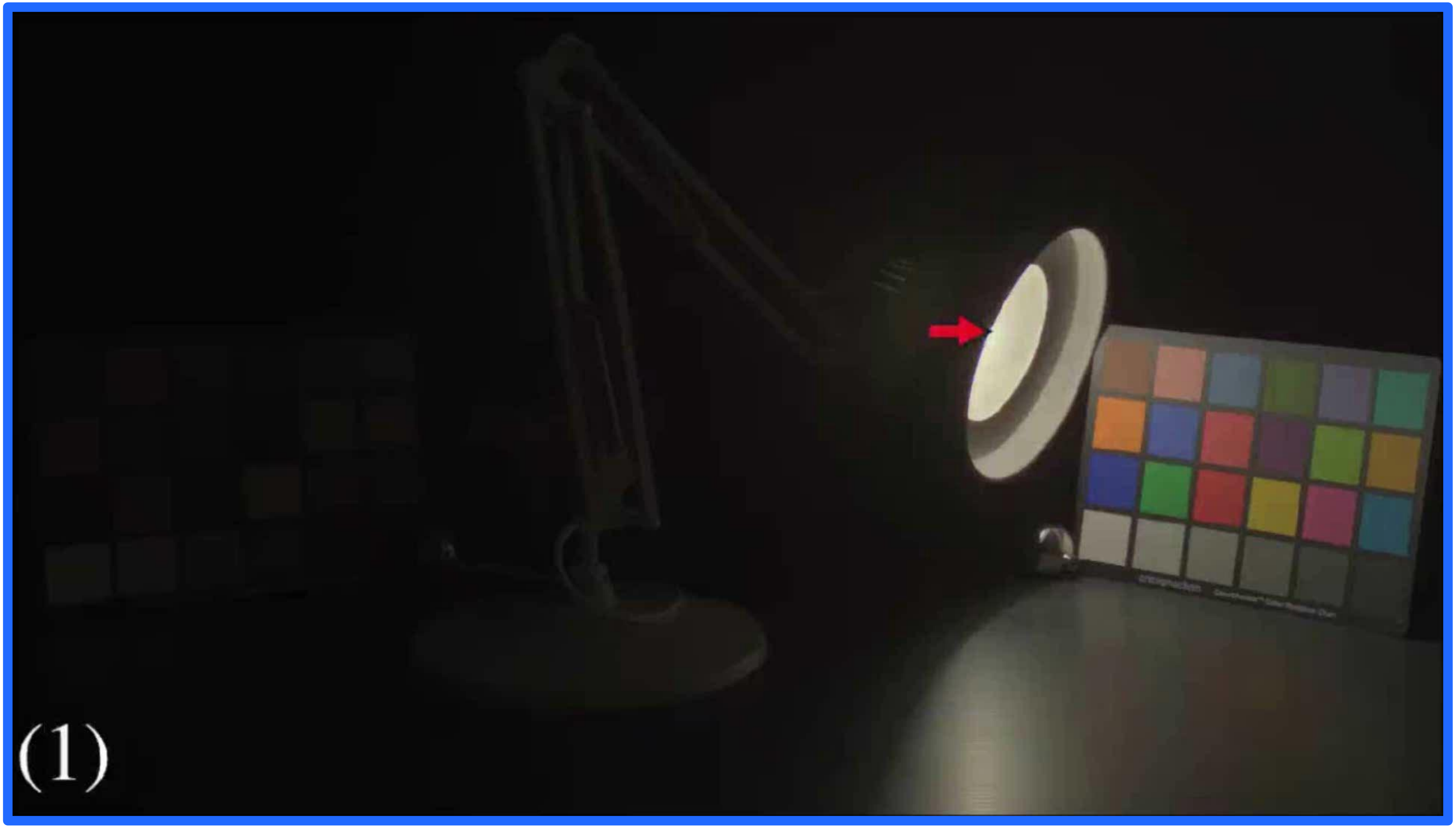
Local Adaptation

- **Mesopic illumination range:** 10^{-3} – 10 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



Images adapted from E Jacobs, D., Gallo, O., A Cooper, E., Pulli, K., & Levoy, M. (2015). Simulating the visual experience of very bright and very dark scenes. ACM Transactions on Graphics (TOG), 34(3), 25.

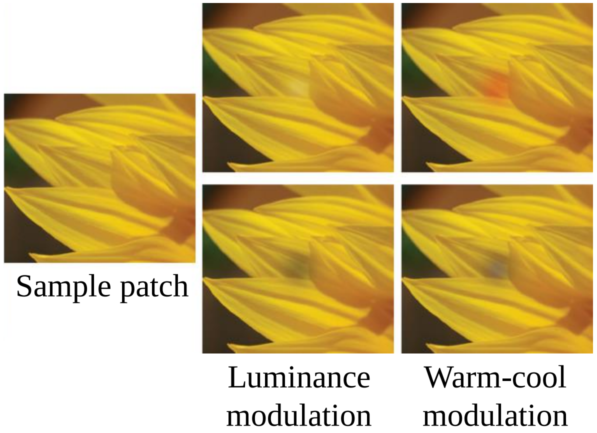
Local Adaptation



Images adapted from E Jacobs, D., Gallo, O., A Cooper, E., Pulli, K., & Levoy, M. (2015). Simulating the visual experience of very bright and very dark scenes. ACM Transactions on Graphics (TOG), 34(3), 25.

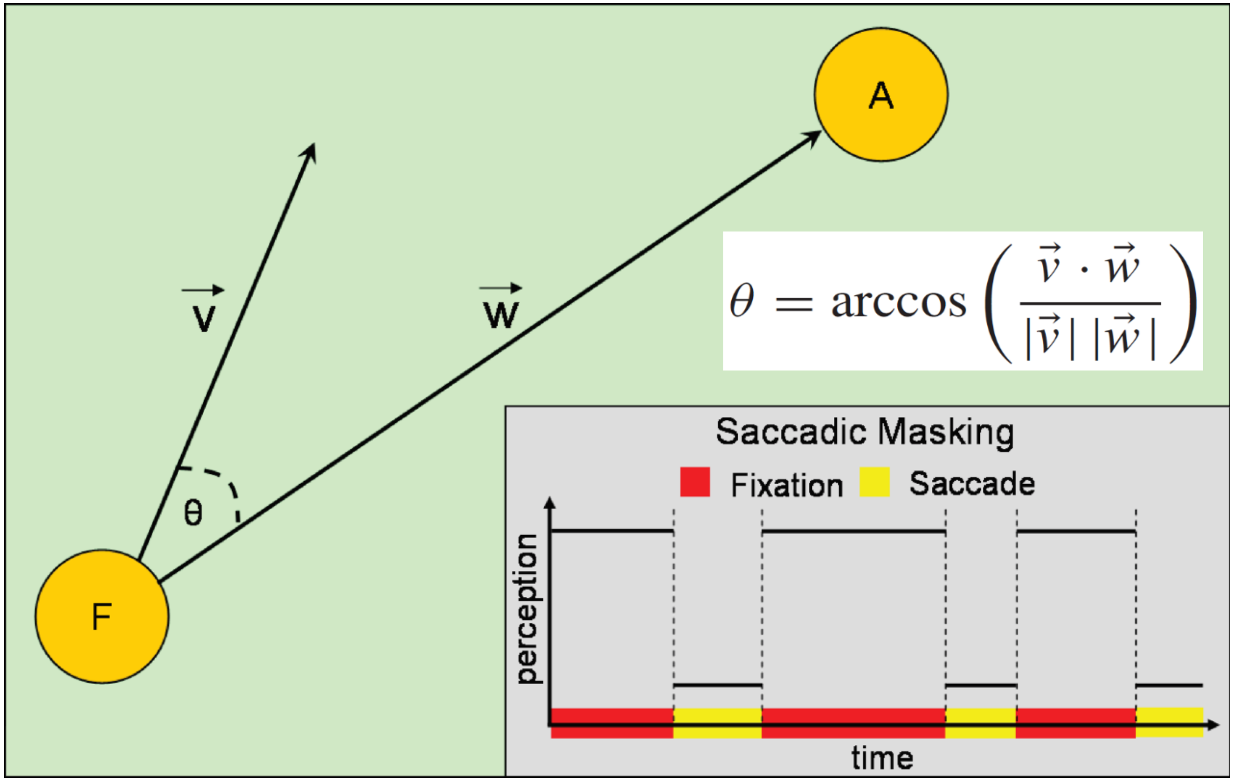
Subtle Gaze Direction

- 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



Images adapted from Bailey, R., McNamara, A., Sudarsanam, N., & Grimm, C. (2009). Subtle gaze direction. ACM Transactions on Graphics (TOG), 28(4), 100.

Subtle Gaze Direction



F: Fixation point**, A:** Predetermined Area of Interest

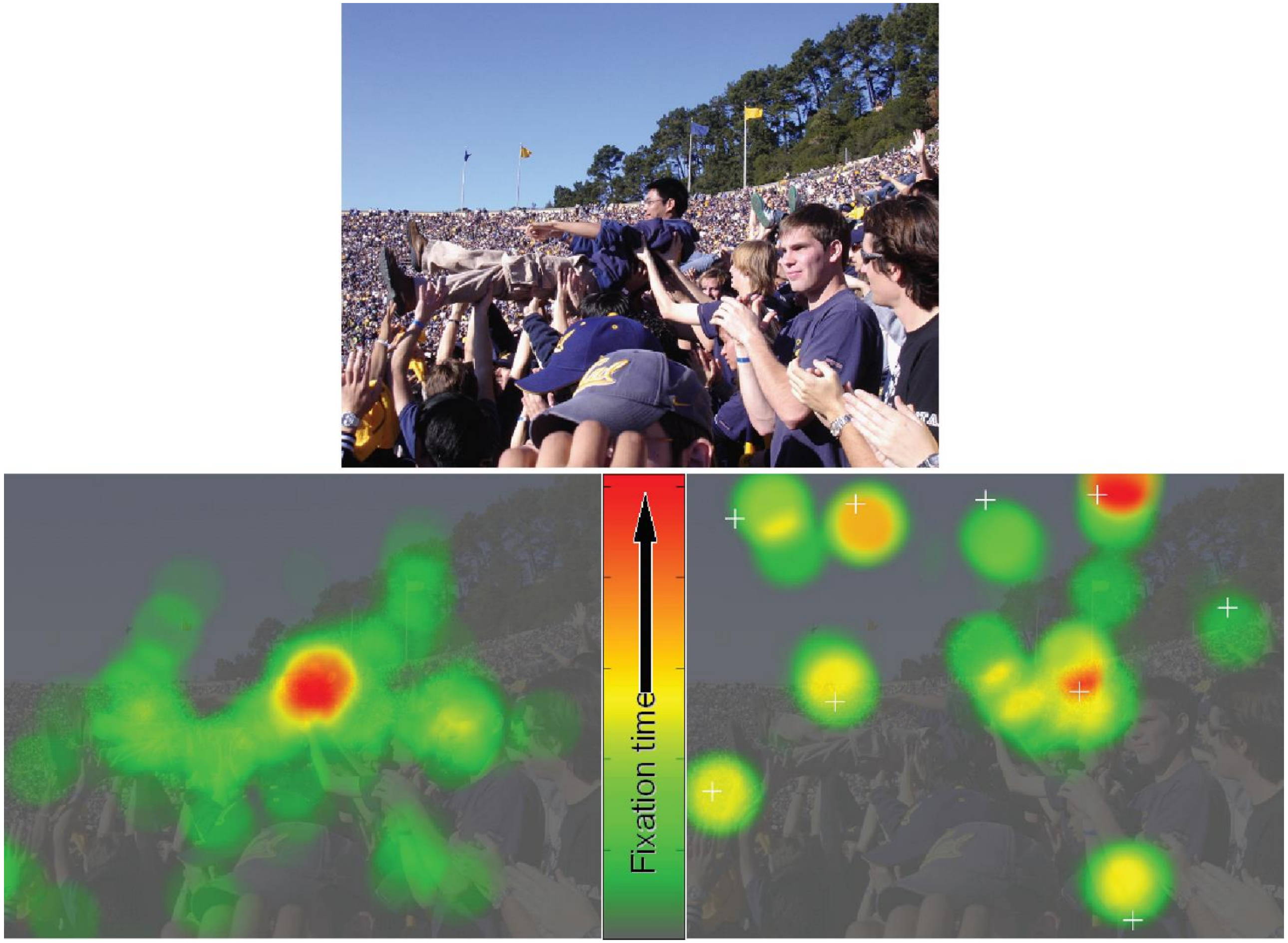
Goal: To direct the user attention to from **F** to **A**

Modulation is applied to A and θ is monitored real-time.

When $\theta \leq 10^{\circ}$, the modulation is terminated immediately.

Images adapted from Bailey, R., McNamara, A., Sudarsanam, N., & Grimm, C. (2009). Subtle gaze direction. ACM Transactions on Graphics (TOG), 28(4), 100.

Subtle Gaze Direction



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

Images adapted from Bailey, R., McNamara, A., Sudarsanam, N., & Grimm, C. (2009). Subtle gaze direction. ACM Transactions on Graphics (TOG), 28(4), 100.

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*

Bottom-up Model

Top-Down Model

Saliency Map   Input Image   Task Map

Modeling Visual Attention

Saliency map ‘only’

Fixation Prediction

Bottom-Up Attention Models

Top-Down Attention Models

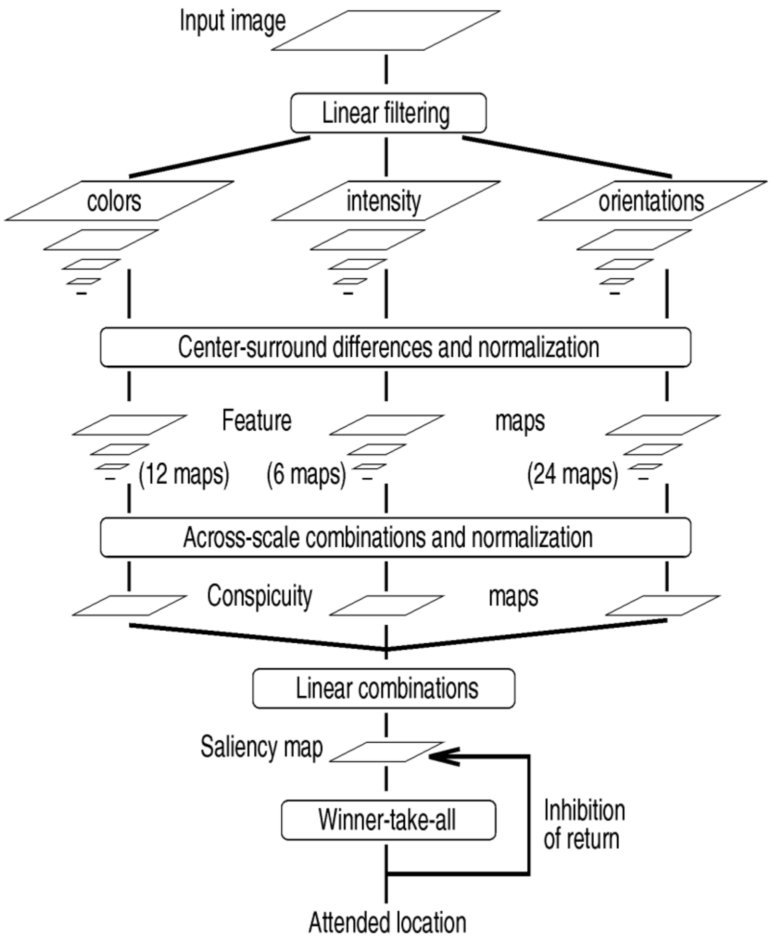
[BI13]

No	Model	Year	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13
Bottom-up (saliency models)															
1	Itti et al. [14]	1998	+	-	-	+	-	-	+	f	+	CIO	C	-	-
2	Privitera & Stark [127]	2000	+	-	-	+	-	-	+	f	+	-	O	-	Stark and Choi
3	Salah et al. [52]	2002	+	+	-	+	-	-	+	-	+	O	G	DR	Digit & Face
4	Itti et al. [119]	2003	+	-	+	+	+	+	+	f	+	CIOFM	C	-	-
5	Torralba [92]	2003	-	+	-	+	-	-	+	s	+	CI	B	DR	Torralba et al.
6	Sun & Fisher [117]	2003	+	-	-	+	-	-	+	-	-	CIO	G	-	-
7	Gao & Vasconcelos [146]	2004	-	+	-	+	-	-	+	s	-	DCT	D	DR	Brodatz, Caltech
8	Ouerhani et al. [210]	2004	+	-	-	+	-	-	+	f	+	CIO+Corner	C	CC	Ouerhani
9	Boccignone & Ferraro [175]	2004	+	-	+	-	+	-	+	f	-	Optical Flow	B	-	BEHAVE
10	Frintrop [50]	2005	+	+	+	+	+	+	+	f/s	+/-	CIOM	C	-	-
11	Itti & Baldi [145]	2005	+	-	+	+	+	+	-	f	+	CIOFM	B	KL, AUC	ORIG-MTV
12	Ma et al. [33]	2005	+	-	+	+	-	-	+	f	+	M*	O	-	-
13	Bruce & Tsotsos [144]	2006	+	-	-	+	-	+	+	f	+	DOG, ICA	I	KL, ROC	Bruce and Tsotsos
14	Navalpakkam & Itti [51]	2006	-	+	-	+	-	+	+	s	+	CIO	C	-	-
15	Zhai & Shah [103]	2006	+	-	+	+	+	-	+	f	+	SIFT	O	-	-
16	Harel et al. [121]	2006	+	-	-	+	-	-	+	f	+	IO	G	AUC	Bruce and Tsotsos
17	Le Meur et al. [41]	2006	+	-	-	+	-	-	+	f	+	LM*	C	CC, KL	Le Meur et al.
18	Walther & Koch [35]	2006	+	-	-	+	-	+	+	f	+/-	CIO	C	-	-
19	Peters & Itti [101]	2007	+	+	+	+	+	-	+	i	+	CIOFM	P	KL, NSS	Peters and Itti
20	Liu et al. [43]	2007	+	-	-	+	-	-	+	f	-	Liu*	G	F-measure	Regional
21	Shic & Scassellati [74]	2007	+	-	+	+	+	-	+	f	+	CIOM	C	ROC	Shic and Scassellati
22	Hou & Zhang [150]	2007	+	-	-	+	-	+	+	f	+	FFT, DCT	S	NSS	DB of Hou and Zhang, 2007
23	Cerf et al. [167]	2007	+	+	-	+	-	+	+	f/s	+	CIO :)	C	AUC	Cerf et al.
24	Le Meur et al. [138]	2007	+	-	+	+	+	-	+	f	+	LM*	C	CC, KL	Le Meur et al.
25	Mancas [152]	2007	+	-	+	+	+	+	+	f	+	CI	I	CC	Le Meur et al.
26	Guo et al. [156]	2008	+	-	-	+	-	-	+	f	+	CIO	D	CC	Self data
27	Zhang et al. [141]	2008	+	-	-	+	-	+	+	f	+	DOG, ICA	B	KL, AUC	Bruce and Tsotsos
28	Hou & Zhang [151]	2008	+	-	+	+	+	-	+	f	+	ICA	I	AUC, KL	Bruce and Tsotsos, ORIG
29	Pang et al. [102]	2008	+	+	+	+	+	-	+	f	+	CIOM	G	NSS	ORIG, Self data
30	Kootstra et al. [136]	2008	+	-	-	+	-	-	+	f	+	Symmetry	C	CC	Kootstra et al.
31	Ban et al. [172]	2008	+	-	+	+	+	-	+	f	+	CIO+SYM	I	-	-
32	Rajashekar et al. [174]	2008	+	-	-	+	-	-	+	f	+	R*	S	CC	Rajashekar et al.
33	Kienzle et al. [165]	2009	+	-	-	+	-	-	+	f	+	I	P	K*	Kienzle et al.
34	Marat et al. [49]	2009	+	-	+	+	+	-	+	f	+	SM*	C	NSS	Marat et al.
35	Judd et al. [166]	2009	+	-	-	+	-	-	+	f	+	J*	P	AUC	Judd et al.
36	Seo & Milanfar [108]	2009	+	-	+	+	+	+	+	f	+	LSK	I	AUC, KL	Bruce and Tsotsos, ORIG
37	Rosin [169]	2009	+	-	-	+	-	-	+	f	+	C+ Edge	O	PR, F-measure	DB of Liu et al, 2007
38	Yin Li et al. [171]	2009	-	+	+	+	+	+	+	s	+	RGB	S	DR	DB of Hou and Zhang, 2007
39	Bian & Zhang [159]	2009	+	-	+	+	+	+	+	f	+	FFT	S	AUC	Bruce and Tsotsos
40	Diaz et al. [160]	2009	+	-	-	+	-	+	+	f	+	CIO	O	AUC	Bruce and Tsotsos
41	Zhang et al. [142]	2009	+	-	+	-	+	-	+	f	+	DOG, ICA	B	KL, AUC	Bruce and Tsotsos
42	Achanta et al. [158]	2009	+	-	-	+	-	-	+	f	+	DOG	S	PR	DB of Liu et al, 2007
43	Gao et al. [147]	2009	+	-	+	+	+	+	+	f	+	CIO	D	AUC	Bruce and Tsotsos
44	Chikkerur et al. [154]	2010	+	+	-	+	-	+	+	f/s	+/-	CIO	B	AUC	Bruce and Tsotsos, Chikkerur
45	Mahadaven & Vasconcelos [106]	2010	+	-	+	-	+	-	+	-	+	I	D	DR, AUC	SVCL background data
46	Avraham & Lindenbaum [153]	2010	+	+	-	+	-	+	+	f/s	+/-	CIO	G	DR, CC	UWGT, Ouerhani et al.
47	Jia Li et al. [133]	2010	-	+	+	+	+	-	+	f	+	CIO	B	AUC	RSD, MTV, ORIG, Peters and Itti
48	Guo et al. [157]	2010	+	-	+	+	+	+	+	f/s	+/-	FFT	S	DR	Self data
49	Borji et al. [89]	2010	-	+	-	+	-	+	+	s	+/-	CIO	O	DR	-
50	Goeferman et al. [46]	2010	+	-	-	+	-	-	+	-	+	C :)	O	AUC	DB of Hou and Zhang, 2007
51	Murray et al. [200]	2011	+	-	-	+	-	-	+	f	+	CIO	C	AUC, KL	Bruce and Tsotsos, Judd et al.
52	Wang et al. [201]	2011	+	-	-	+	-	-	+	f	+	ICA	I	AUC	Self data
Top-down (general attention models)															
53	McCallum [163]	1995	-	+	-	+	-	+	-	i	+	-	R	-	Self data
54	Rao et al. [23]	1995	-	+	-	+	-	-	+	s	+	CIO	O	-	Self data
55	Ramstrom & Christiansen [168]	2002	-	+	-	+	-	-	+	-	+	CI	O	-	-
56	Sprague & Ballard [109]	2003	-	+	+	-	+	+	+	i	-	S*	R	-	-
57	Renninger et al. [94]	2004	-	+	-	+	-	+	-	s	-	Edgelet	I	DR	Self data
58	Navalpakkam & Itti [80]	2005	-	+	-	+	-	+	+	-	+	CIO	C	-	Self data
59	Paletta et al. [164]	2005	-	+	-	+	-	-	+	-	-	SIFT	R	DR	COIL-20, TSG-20
60	Jodogne & Piater [162]	2007	-	+	-	+	-	-	+	i	-	SIFT	R	-	-
61	Butko & Movellan [161]	2009	-	+	+	+	+	+	+	s	-	-	R	-	-
62	Verma & McOwan [214]	2009	+	-	-	+	-	+	-	s	-	CIO	O	-	-
63	Borji et al. [89]	2010	-	+	-	+	-	-	+	i	-	CIO	R	-	-

Borji, Itti: *State-of-the-art in visual attention modeling.* IEEE Transactions on Pathern Analysis and Machine Intelligence (2013)

Saliency (Itti Model)

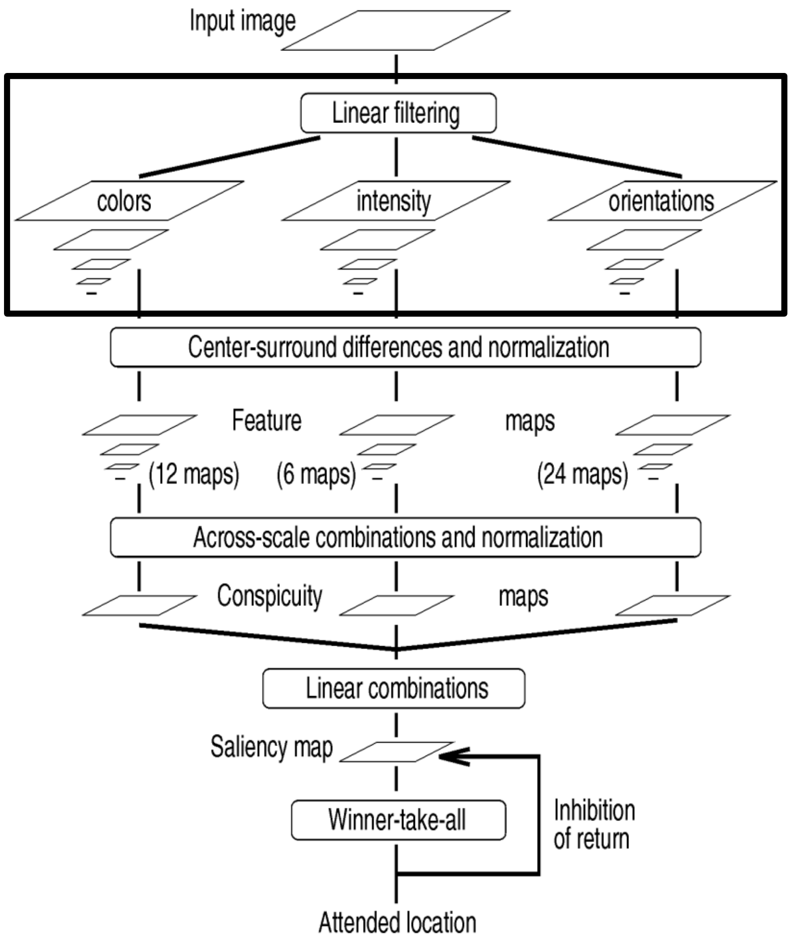
- **Attention activity may be controlled in bottom-up (scene-dependent) and top-down (task-dependent) manner**
- **Model based on the bottom-up architecture proposed by Koch and Ullman:**
  - Visual layer is decomposed into feature maps
  - The locations which stand out from their surround persist
 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



Images adapted from Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE PAMI, (11), 1254-1259.

Saliency

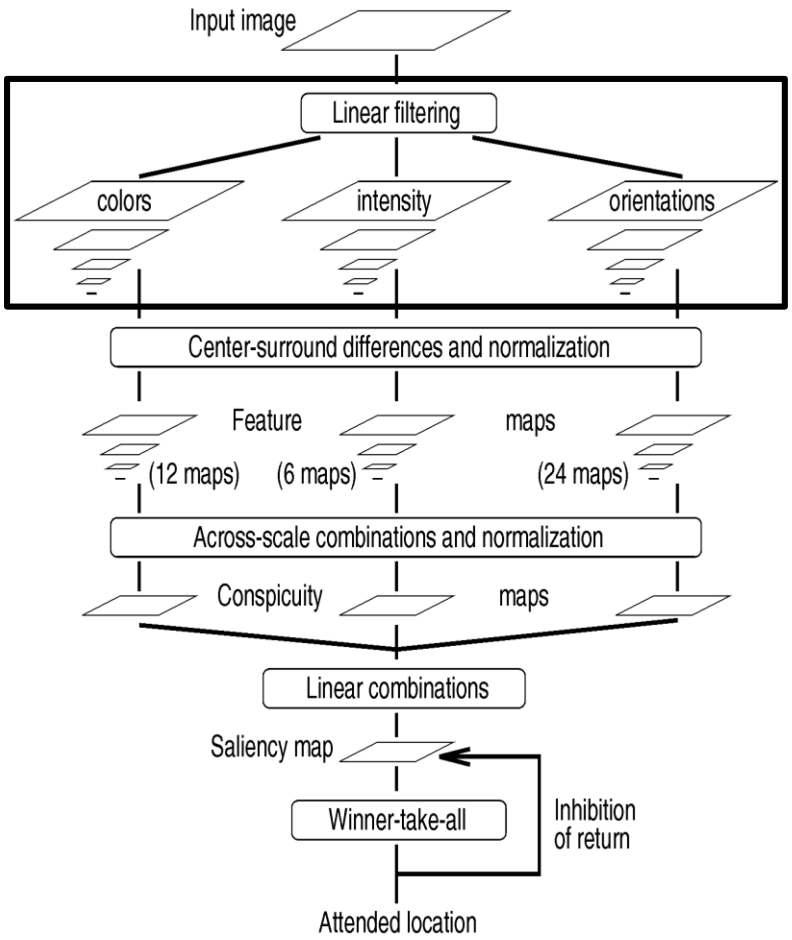
- **General computation principle in the retina, lateral geniculate nucleus and primary visual cortex:**
  - The stimuli in a small region at the **center** of the visual space promotes neuronal activity while a broader concentric region (**surround**) has inhibitory effect
- **Visual features of center-surround difference are extracted for color, intensity and orientation**



Images adapted from Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE PAMI, (11), 1254-1259.

Saliency

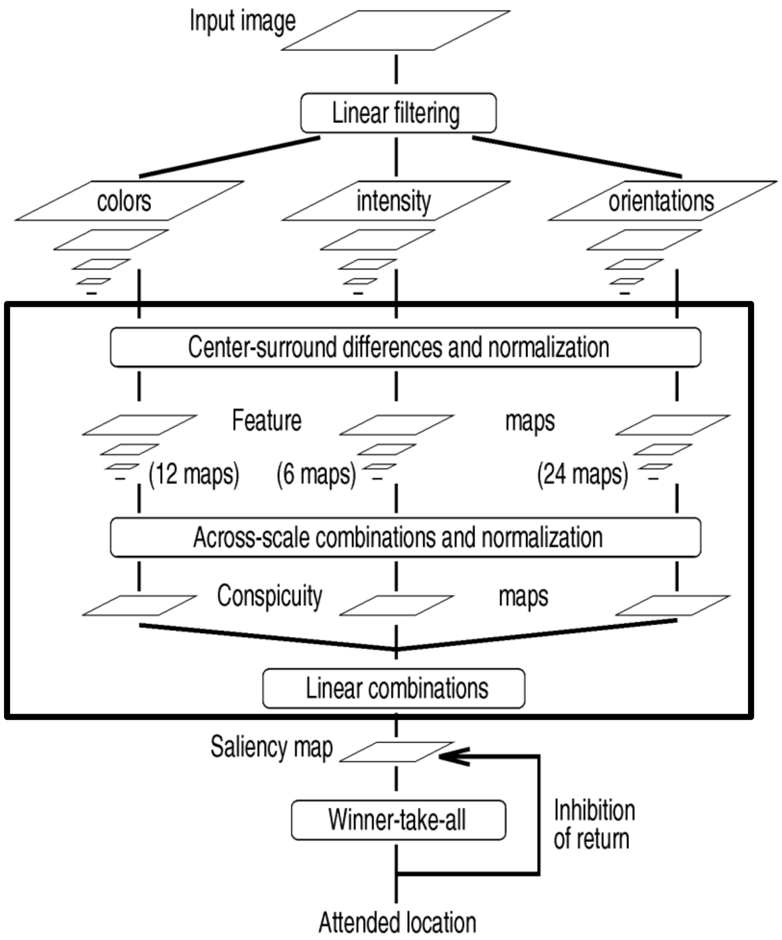
- **Intensity:**
  - I = (r + g + b) / 3,
- **Color:**
  - R = r – (g + b)/2
  - G = g – (r + b)/2
  - B = b – (r + g)/2
  - Y = (r + g)/2 – |r – g|/2 – b (yellow)
- **Orientation:**
  - Oriented Gabor pyramids with 9 scales and 4 orientations (0°, 45°, 90° and 135°)



Images adapted from Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE PAMI, (11), 1254-1259.

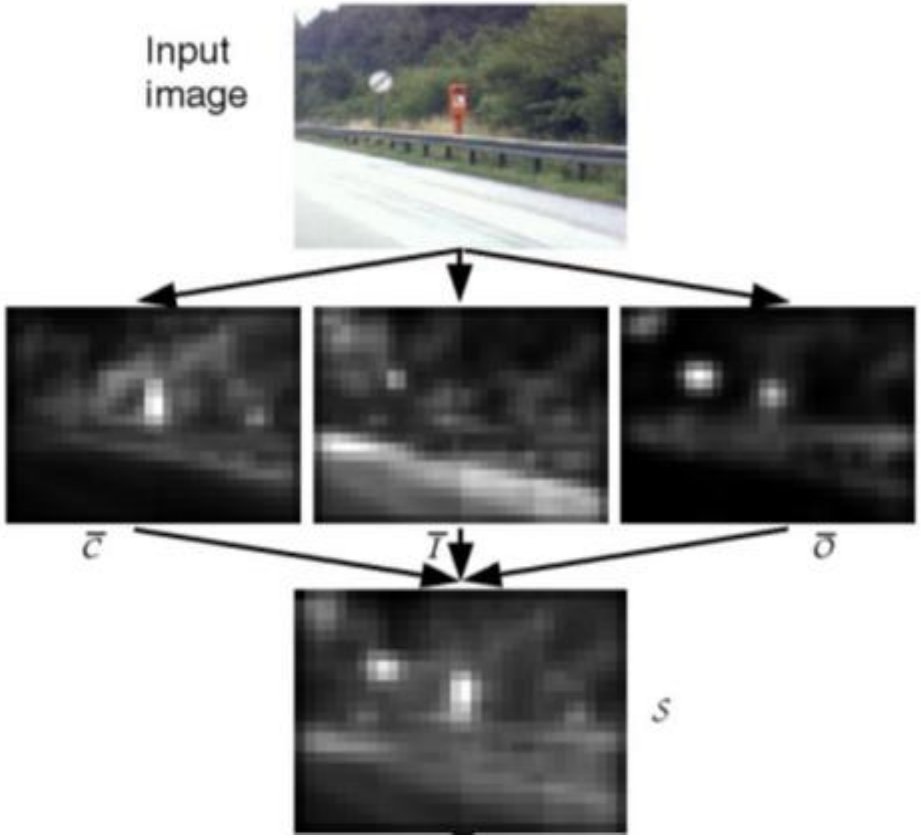
Saliency

- **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 ϵ {2, 3, 4}
  - Surround is in scale c + δ, δ ϵ {3, 4}
- **The resulting maps are normalized and summed into final saliency map**



Images adapted from Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE PAMI, (11), 1254-1259.

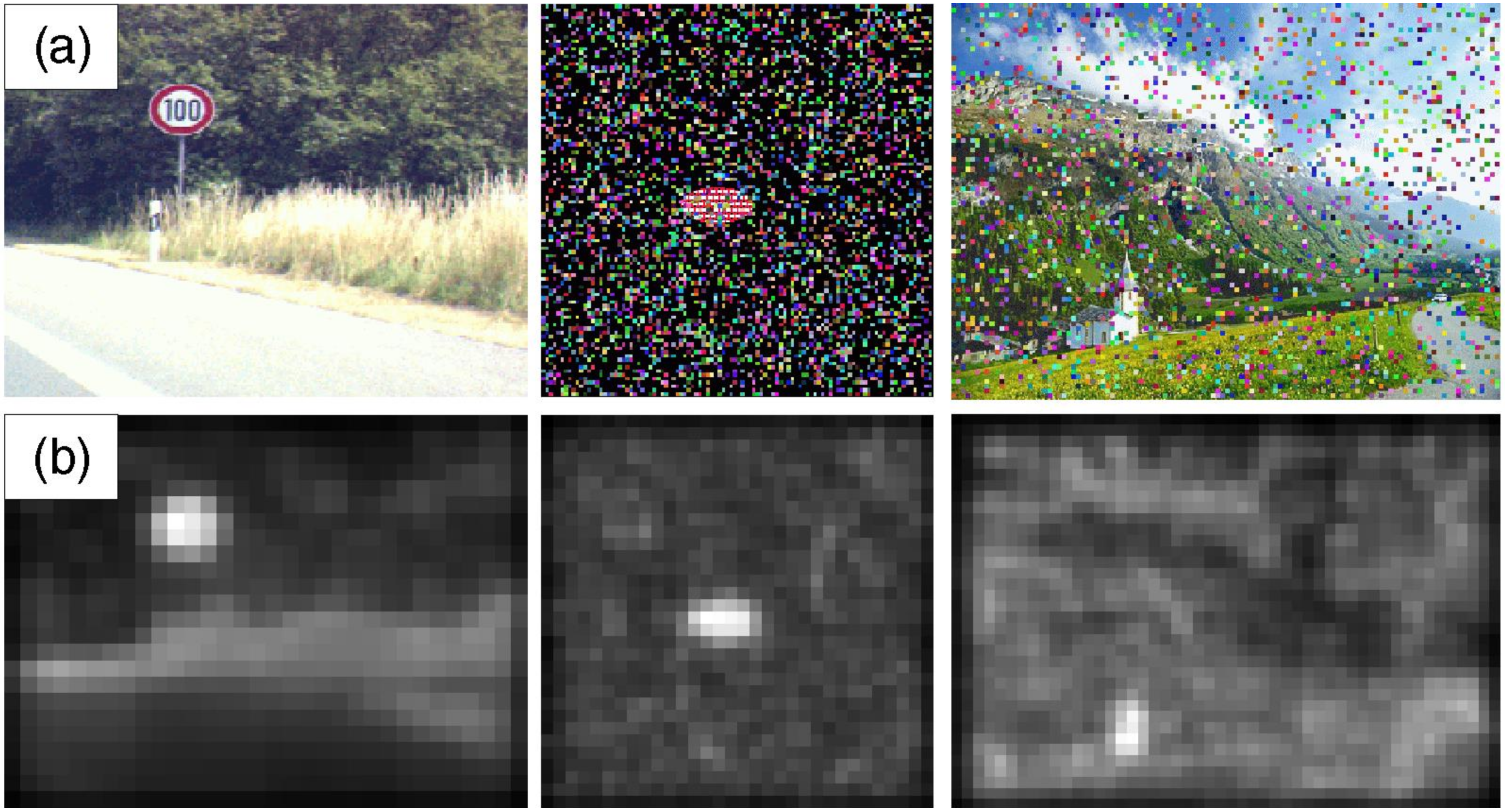
Saliency



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

Images adapted from Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE PAMI, (11), 1254-1259.

Saliency

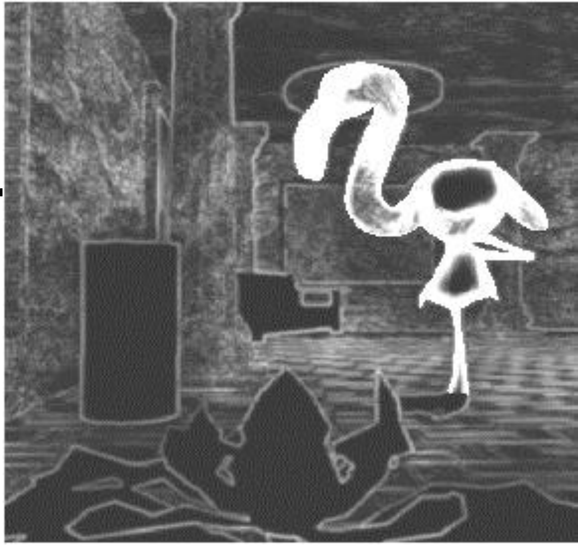


Input images **(a)** and corresponding saliency maps **(b)**

Images adapted from Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE PAMI, (11), 1254-1259.

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



Error Tolerance Map: higher tolerance in brighter regions

[YPG01] **Yee et al.:** *Spatiotemporal Sensitivity and Visual Attention for Efficient Rendering of Dynamic Environments.* ACM TOG 20, 1 (2001), pp. 39–65

Images: Yee *et al.*

Visual Attention [HMYS01]

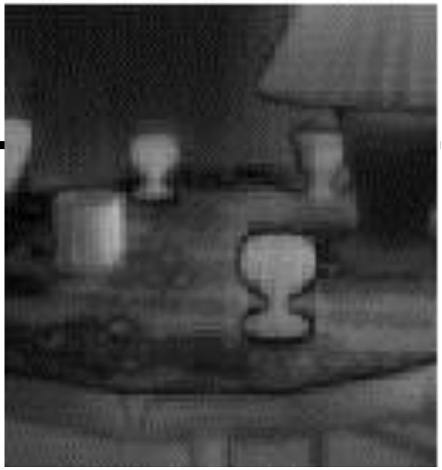
Interactive Scenario: Shading artifacts of "unattended" glossy objects are likely to remain unnoticed

- 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

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

Modeling High-level Attention

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**

[Wo94] **Wolfe:** *Guided search 2.0 a revised model of visual search.* Psychonomic Bulletin & Review (1994)

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.**