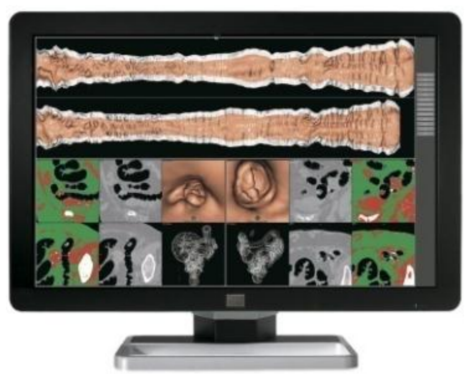
Temporal Image Retargeting

Karol Myszkowski
Max-Planck-Institut für Informatik

Elmar Eisemann
Telecom ParisTech / CNRS-LTCI

Observations: New Displays

Bigger & brighter



More resolution

Higher refresh rates

3D

Observations: Bigger & Brighter

- Increased role of peripheral vision
  - Higher sensitivity to flickering
  - Lower acuity for high eccentricity

Panasonic 150" Plasma

Observations: Bigger & Higher resolution

- More pixels to render
- SHD = 2 x HD
- People move closer
  - Higher angular and pixel velocity
  - More perceived **blur** due to smooth pursuit eye motion

Barco Coronis Fusion 6MP DL (MDCC-6130)

Observations: Faster refresh

- **120 Hz** displays (3D stereo applications)
  - LCD displays for gamers: *Samsung*, … (~ $300)
  - DMD projectors: *DepthQ* , … (> $2000)

Observations: 3D is a hot topic

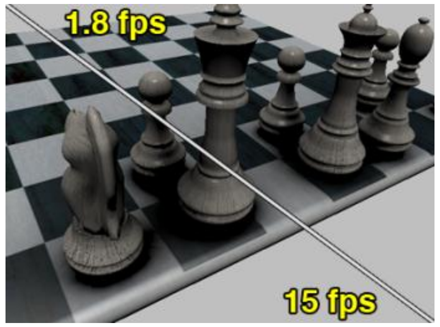
Standard stereo

Backward-compatible stereo

Observations: GPU

- More powerful, multi-core
- More than 50 fps not unusual
- For uncontrolled #fps judder effect
- Advanced per-pixel shaders costly

Super-resolution
[*Yang* et al. EGSR 2008]



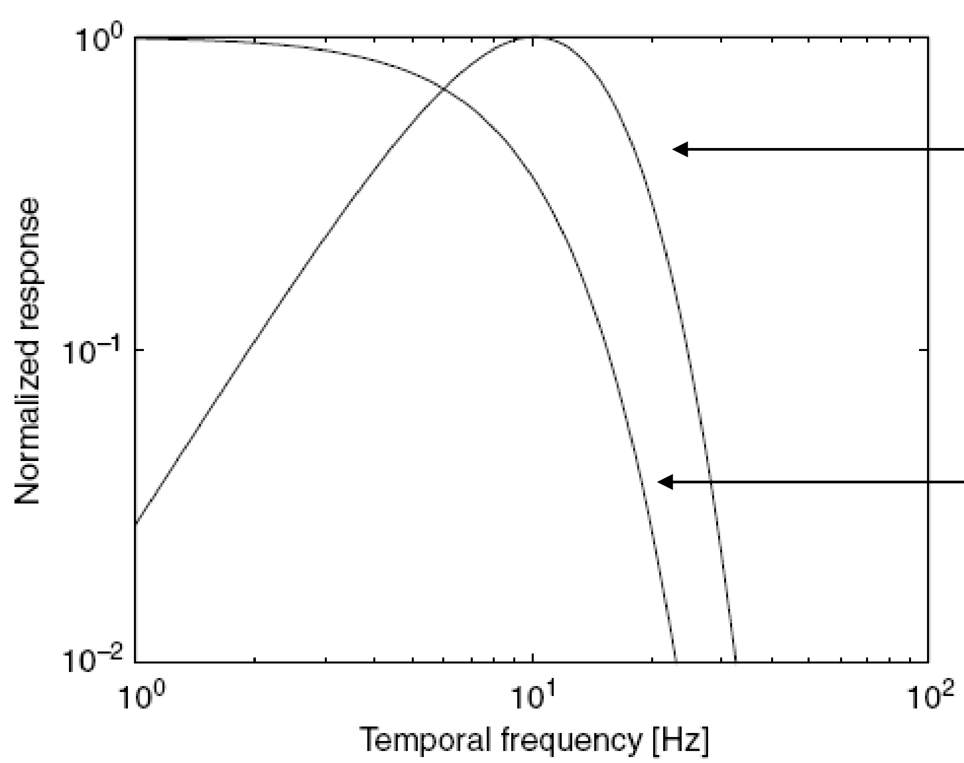
Shader decomposition and caching
[*Sitthi-Amorn* et al.,Siggraph Asia 2008]

Motivation

- More fps help in blur and flicker reduction
  - Adding extra frames in time domain easy
    - TV makers do this using relatively imprecise optical flow computation (100Hz and 200Hz TV sets)
    - In rendering motion flow simulation cheap and precise
    - New opportunities in the design of sharpening filters
      - Take into account perception, image content and display characteristics for rendered frame enhancement
      - So far rendering & enhancement usually separate steps
- Through super-resolution algorithms spatial resolution can be extended
  - Many people in graphics tried this

Basic Psychophysics

- Temporal integration of signal performed by HVS to improve the signal to noise ratio
  - Integration duration up to 120ms
  - Temporal summation faster for low spatial frequencies



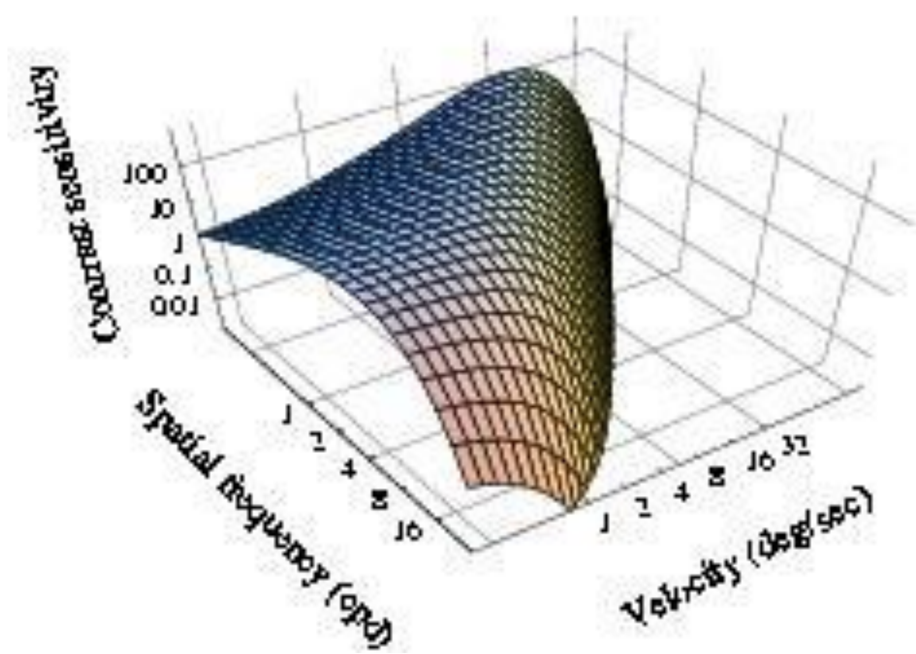
- Temporal frequency responses
  - Band-pass: Fast visual channels tuned to low spatial and high temporal frequencies (***transient*** response) – motion detection
  - Low-pass: Slow visual channels tuned to high spatial and low temporal frequencies (***sustained*** response) – object identification

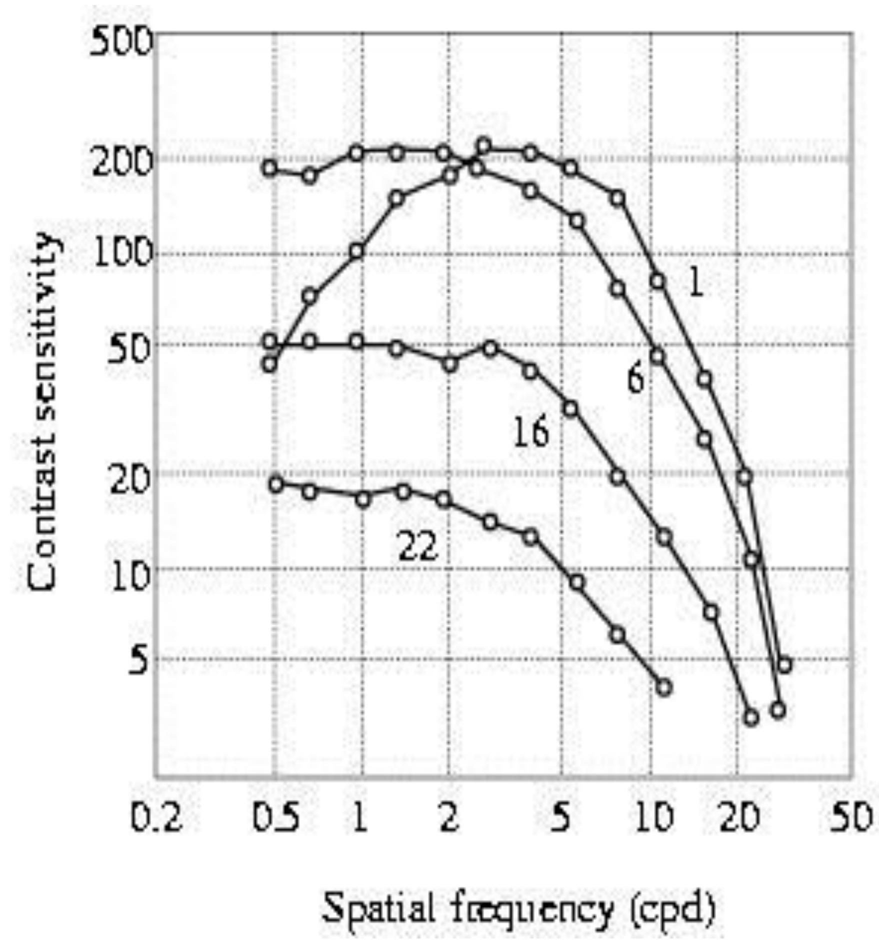
Winkler, Video Quality Metrics, Wiley 2005

Spatio-temporal Contrast Sensitivity Function

- Low sensitivity of HVS to temporal change of high spatial frequencies and high sensitivity to low spatial frequencies

  → high spatial frequencies can be sampled in temporal domain more sparsely





Perception: Flickering

- Critical Flicker Frequency (CFF)
  - Increases with display brightness
    - The Ferry-Porter law:

$$CFF \approx a \cdot \log(luminance) + b$$

  - For bright adaptation conditions and patterns of wide spatial extent the highest flicker sensitivity at the periphery
  - Otherwise, the highest flicker sensitivity at the fovea

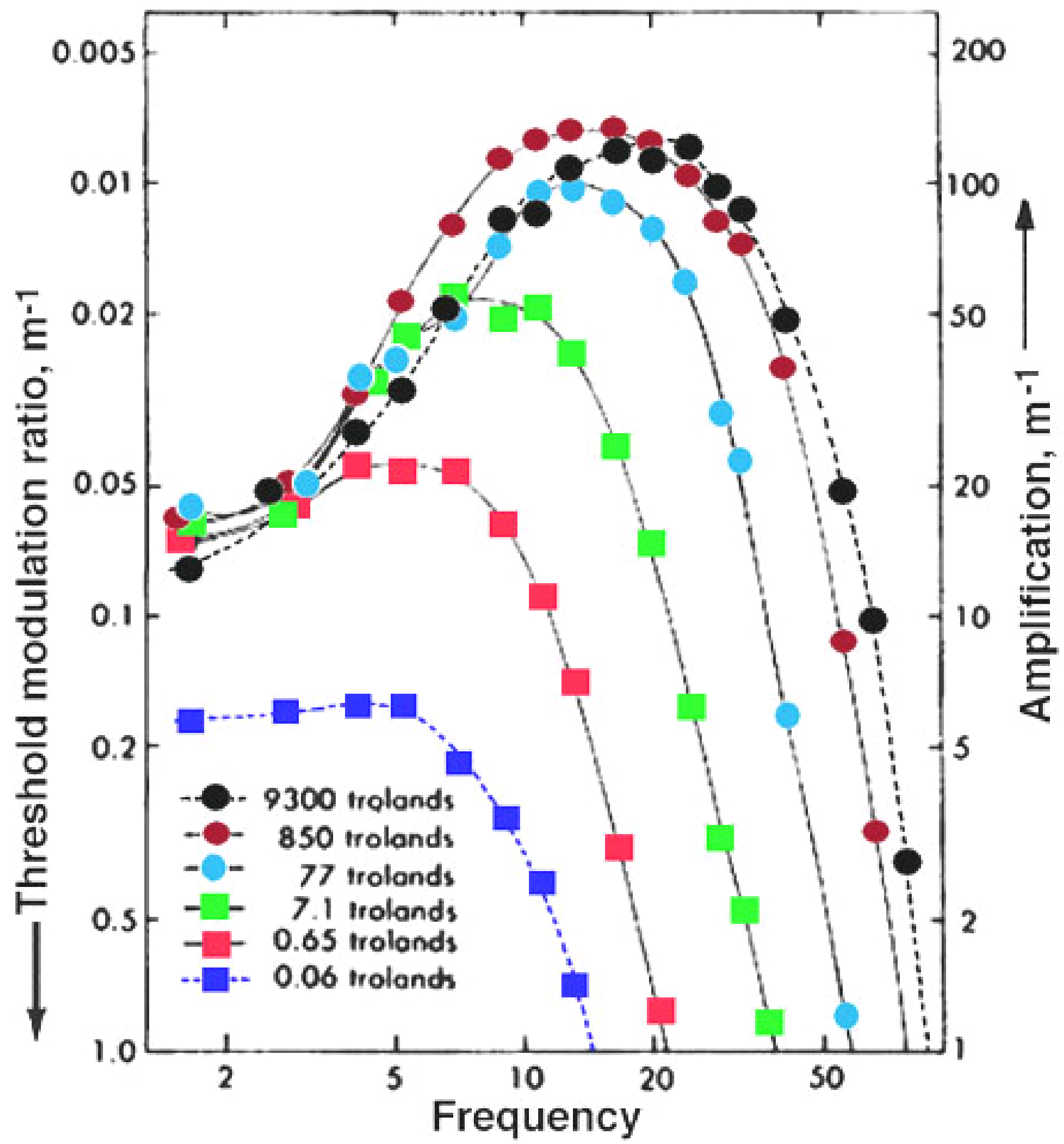
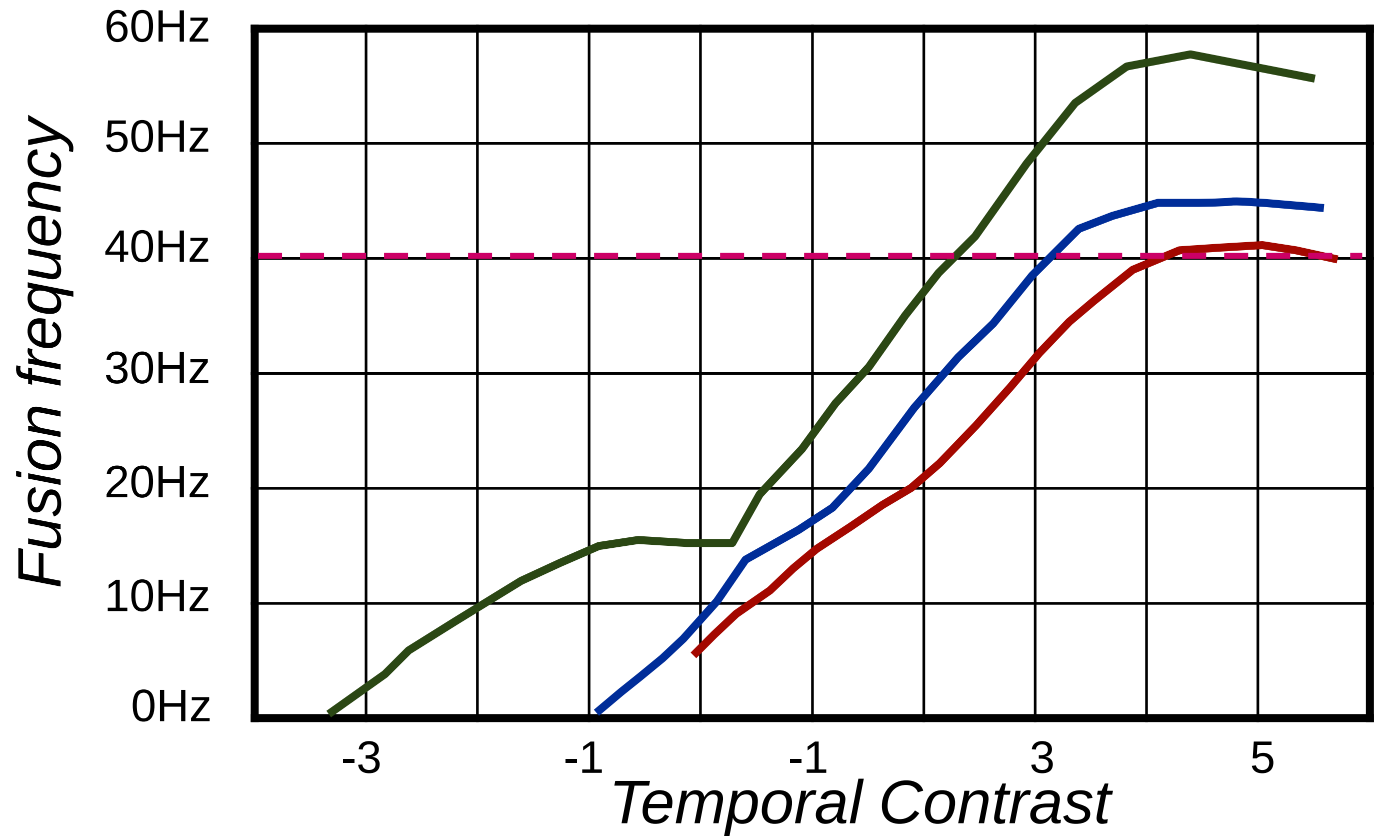


Fig. 11. Temporal Contrast Sensitivity Function (TSF) for various adapting fields. Kelly's data from Hart Jr, W. M., The temporal responsiveness of vision. In: Moses, R. A. and Hart, W. M. (ed) Adler's Physiology of the eye, Clinical Application. St. Louis: The C. V. Mosby Company, 1987.

Perception: Flickering

Fusion frequency vs. temporal contrast & pattern spatial extent



Critical Flicker Frequency - Hecht and Smith's data from Brown J. L. *Flicker and Intermittent Simulation*

19 deg

2 deg

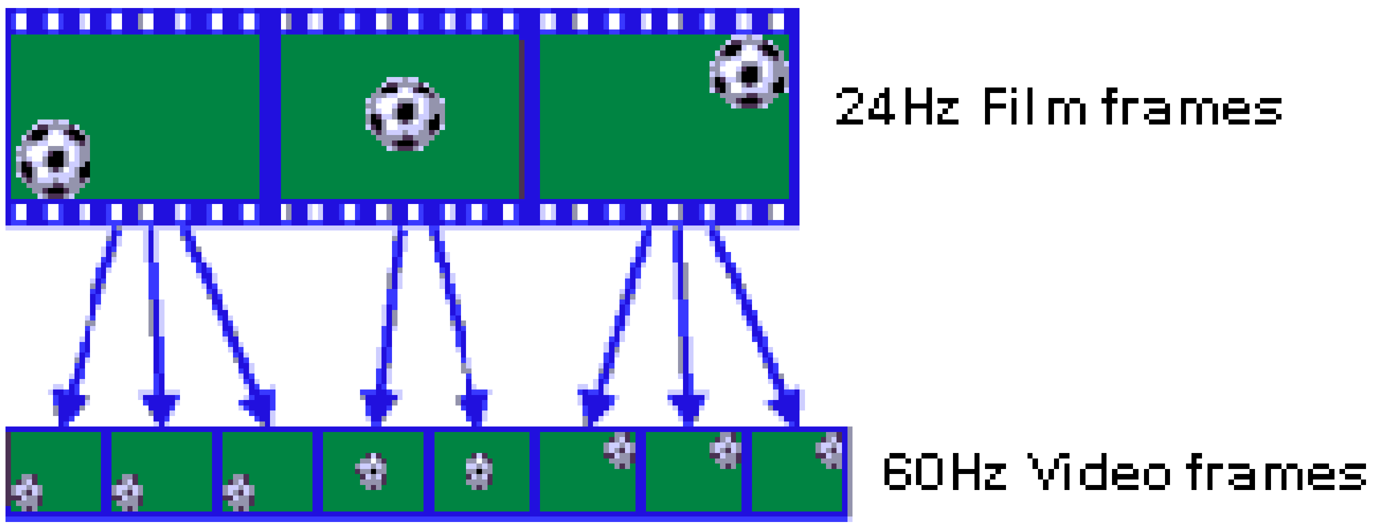
0.3 deg → **40Hz**

Perception: Smooth Pursuit Eye Motion (SPEM)

- Enables to maintain the object of interest in the fovea
- Blur due to object motion is eliminated
- Eye tracking experiment [Laird et al. 2006]
  - Almost perfect tracking for steady linear motion with velocities of 0.625 – 7 deg/s
  - Still possible up to 80 deg/s
- SPEM initialization very fast
  - Good tracking possible in 100ms after switching gaze between objects moving in different directions
- Other fixational eye movements during SPEM: tremors, drifts, and microsaccades similar to static fixations
  - Compensated by HVS contribute little to blur

Perception: Judder

- Repeating the previous frame while the eye is smoothly tracking moving object
- Most noticeable for camera pans, scrolling text, and so on
- 8Hz difference between rendered and displayed frames the most critical, i.e. 42 fps on 50 Hz display
- 3:2 pulldown judder: Converting 24Hz film material to 60Hz

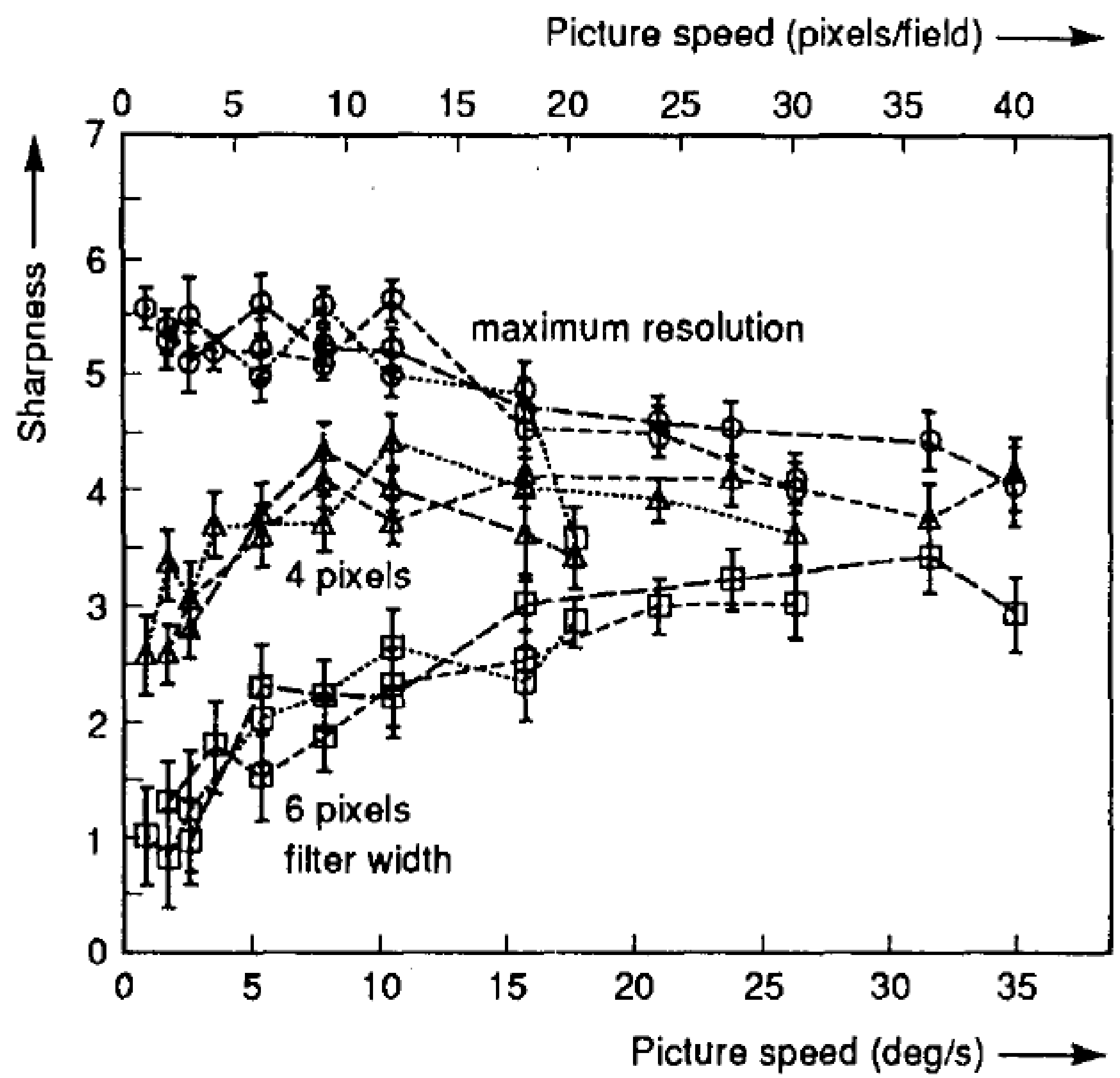


http://msdn.microsoft.com/en-us/windows/hardware/gg463407.aspx

Perception: Blur

- Sharp edges suffer blurring during motion
  - Perceived blur increases with velocity
- Blurred edges appear sharper [Westerink&Teunissen 1994]
  - Apparent sharpening increases with velocity
- Shortly shown blurred edge (7-40ms) appears sharper than the same edge shown for a longer time
- Higher contrast looks sharper
- Adding noise to texture may increase apparent sharpness [*Fairchild and Johnson* 2000, 2005]

Perception: Perceived Sharpness vs. Velocity



J. Westerink, K. Teunissen, Perceived sharpness in complex moving images, Displays 1994

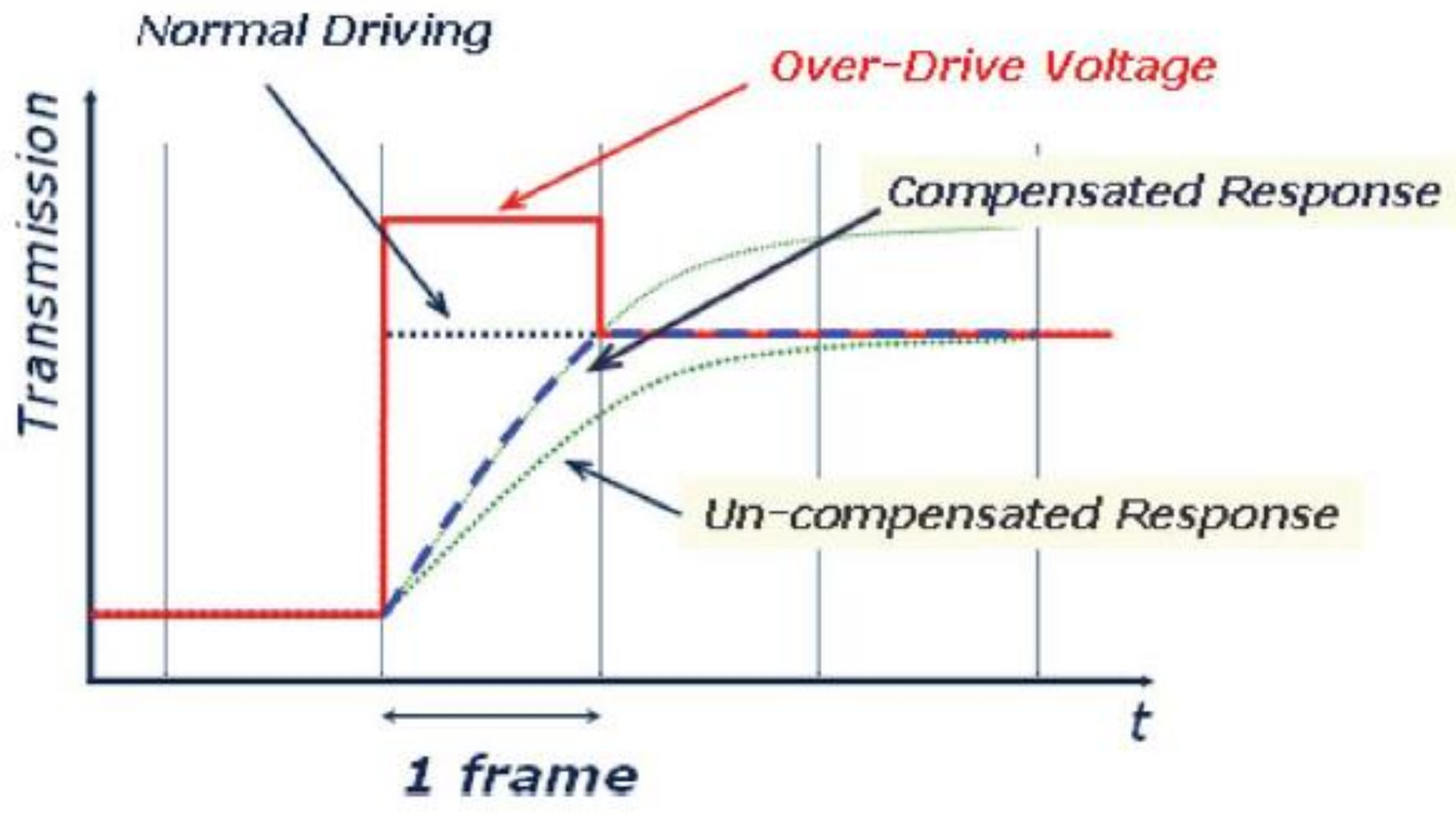
Blur in Hold-type Displays (LCD)

Two main reasons:

- Slow-response of LC
  - 16ms display responsible for only 30% of blur effect
  - Now for 2ms displays mostly negligible
- Image is held while the eye is tracking moving object (*smooth pursuit eye motion* SPEM), which causes blur in the retina image
  - Purely perceptual effect
  - Can be modeled as a box function in temporal domain

Overdriving in LCD TV

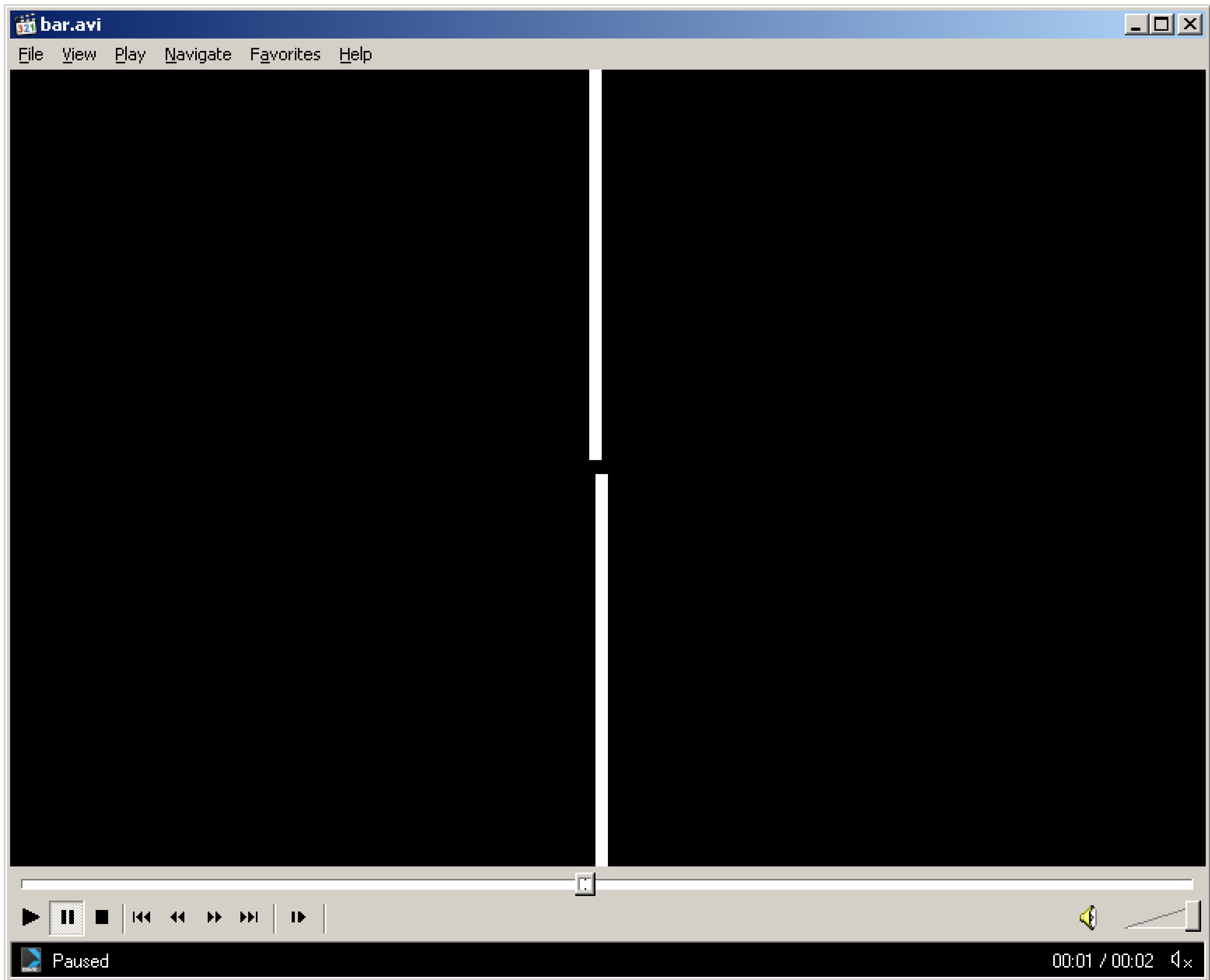
Combating slow response of LC



J.H. Souk, J. Lee, Recent Picture Quality Enhancement Technology Based on Human Visual Perception in LCD TVs, 2007

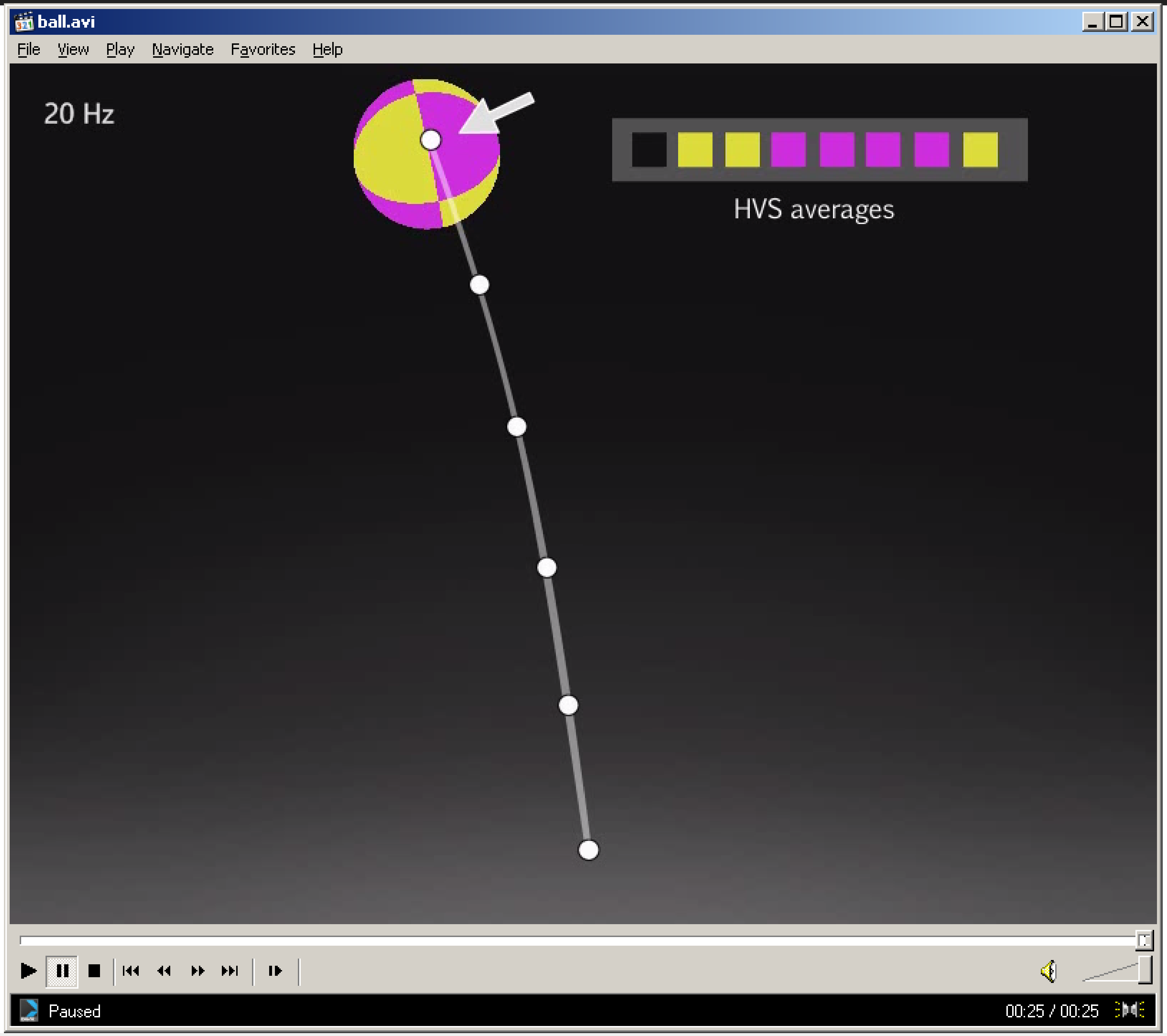
Hold-type Blur Demo: 30Hz vs. 60Hz

30 Hz

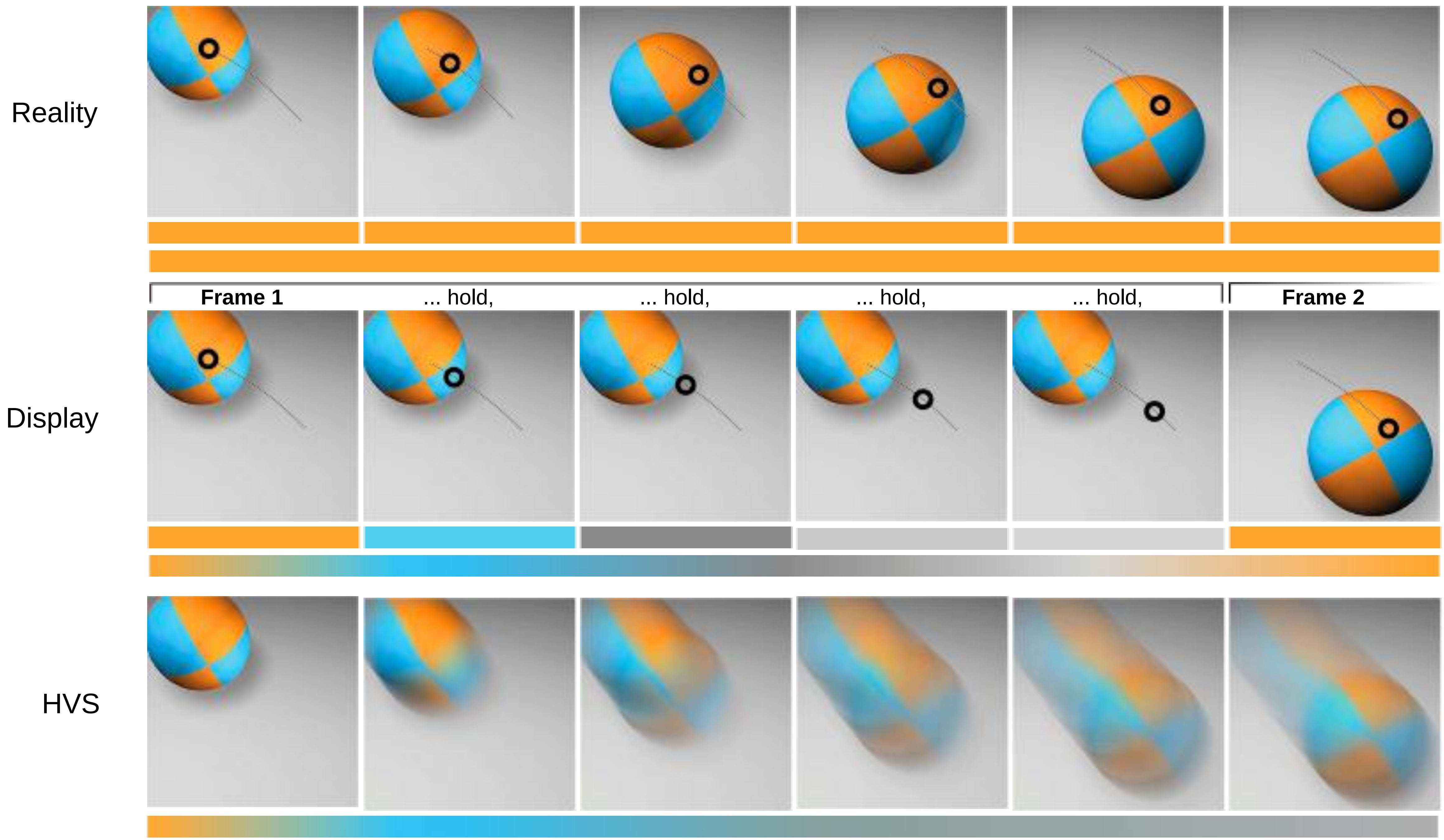


60 Hz

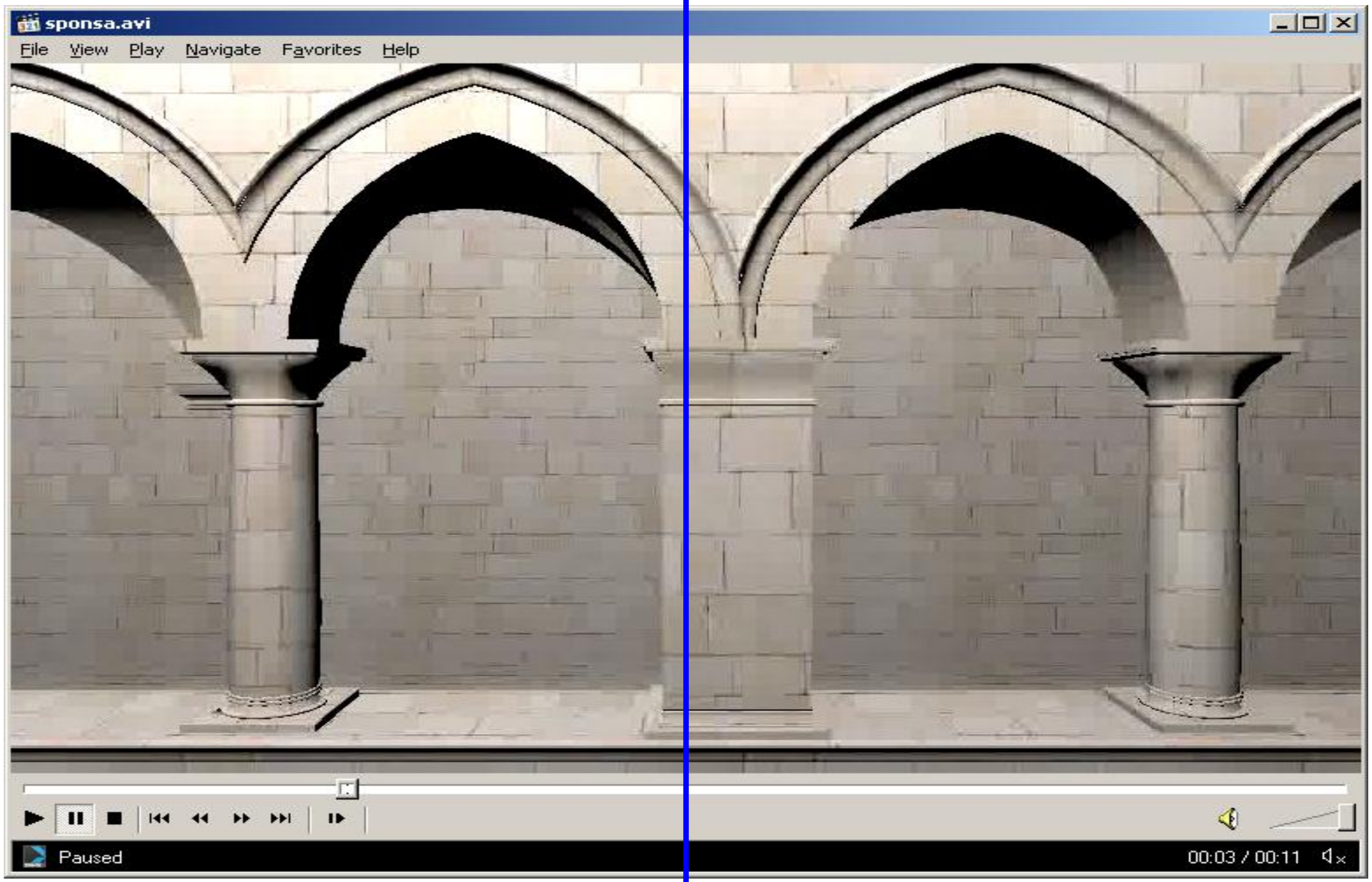
Hold-type Blur Demo: Ball



Hold-type Blur Explanation

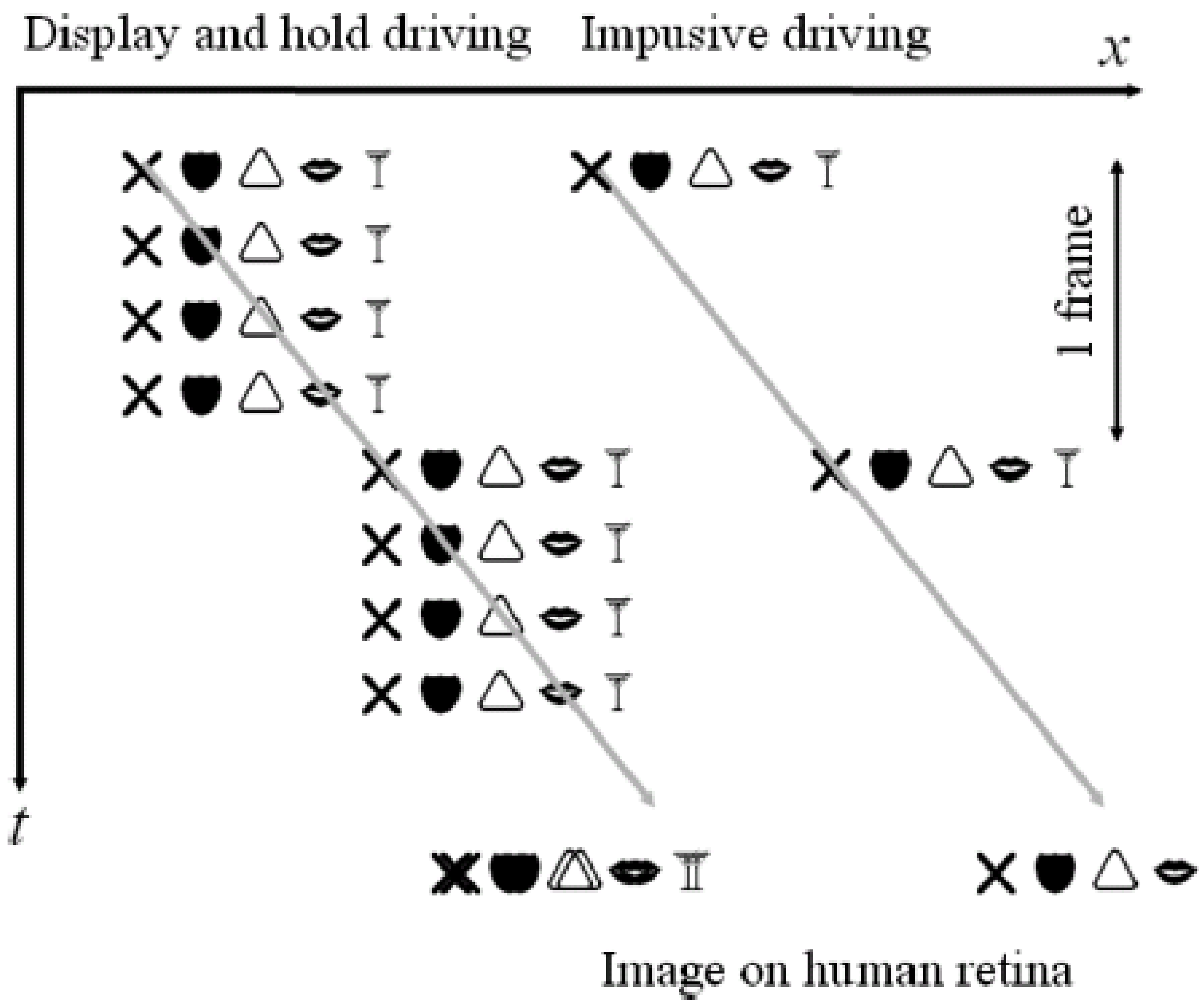


Demo: Gaze Fixing vs. Dynamic Object Tracking



SIGGRAPH ASIA 2011 HONG KONG

Hold Effect: LCD vs. CRT Displays

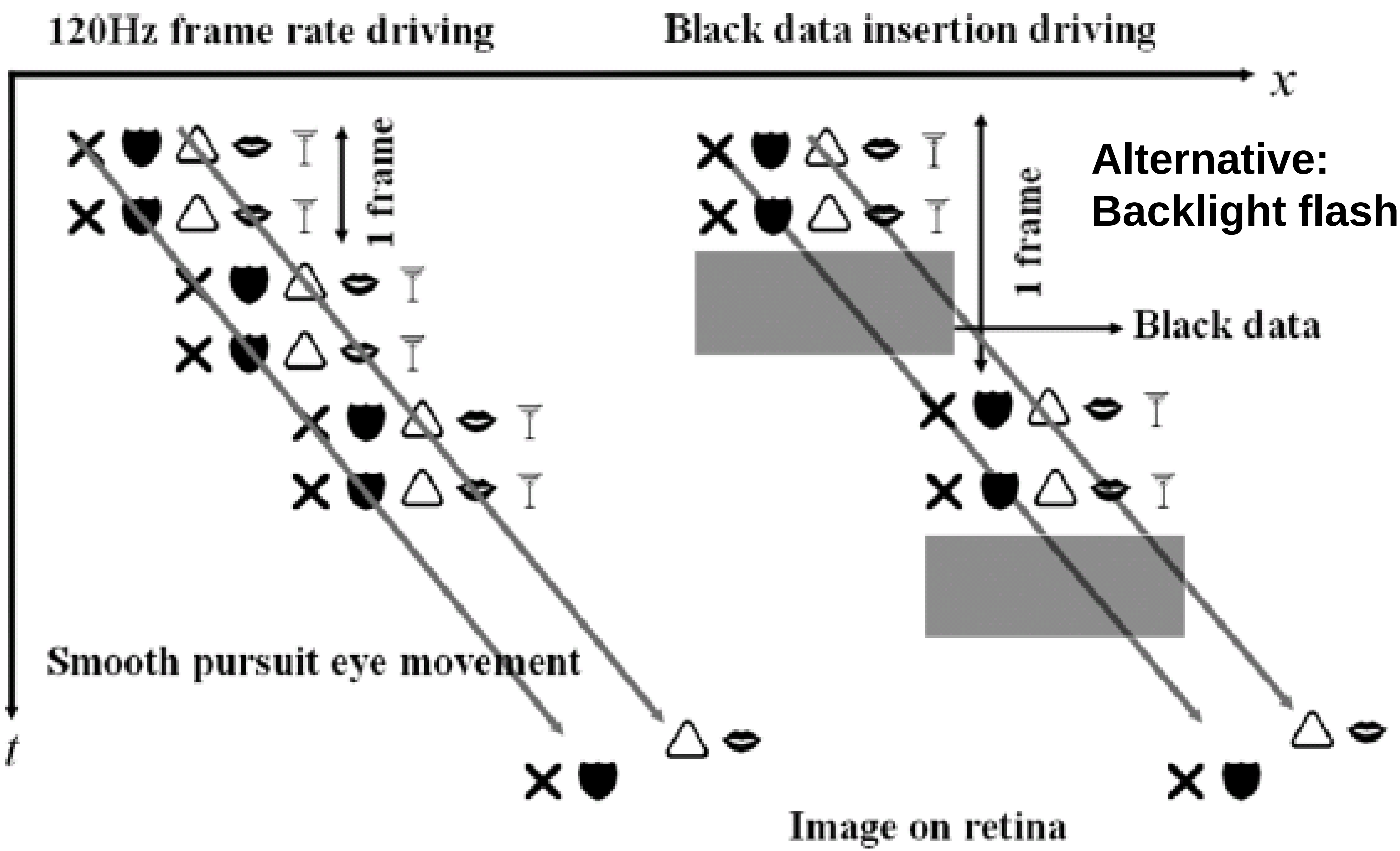


J.H. Souk, J. Lee, Recent Picture Quality Enhancement Technology Based on Human Visual Perception in LCD TVs, 2007

Combating Hold-type Blur in TV Sets

- **Black data insertion (BDI)**
  - Black frames interleaved with the original frames
  - Mimics CRT behavior
- **Frame rate doubling (FRD)**
  - Additional frames are obtained by interpolating pairs of original frames along their optical-flow trajectories
  - Requires introducing latency of one keyframe, which is not a problem in broadcasting applications, but is not suitable for gaming
  - The final effect depends on optical flow accuracy
- **Blurred frame insertion (BFI)**
  - Cheap version of FRD
  - Original frames are replicated and blurred
  - Ghosting for dynamic objects due to lack of motion compensation

Combating Hold-type Blur in TV Sets



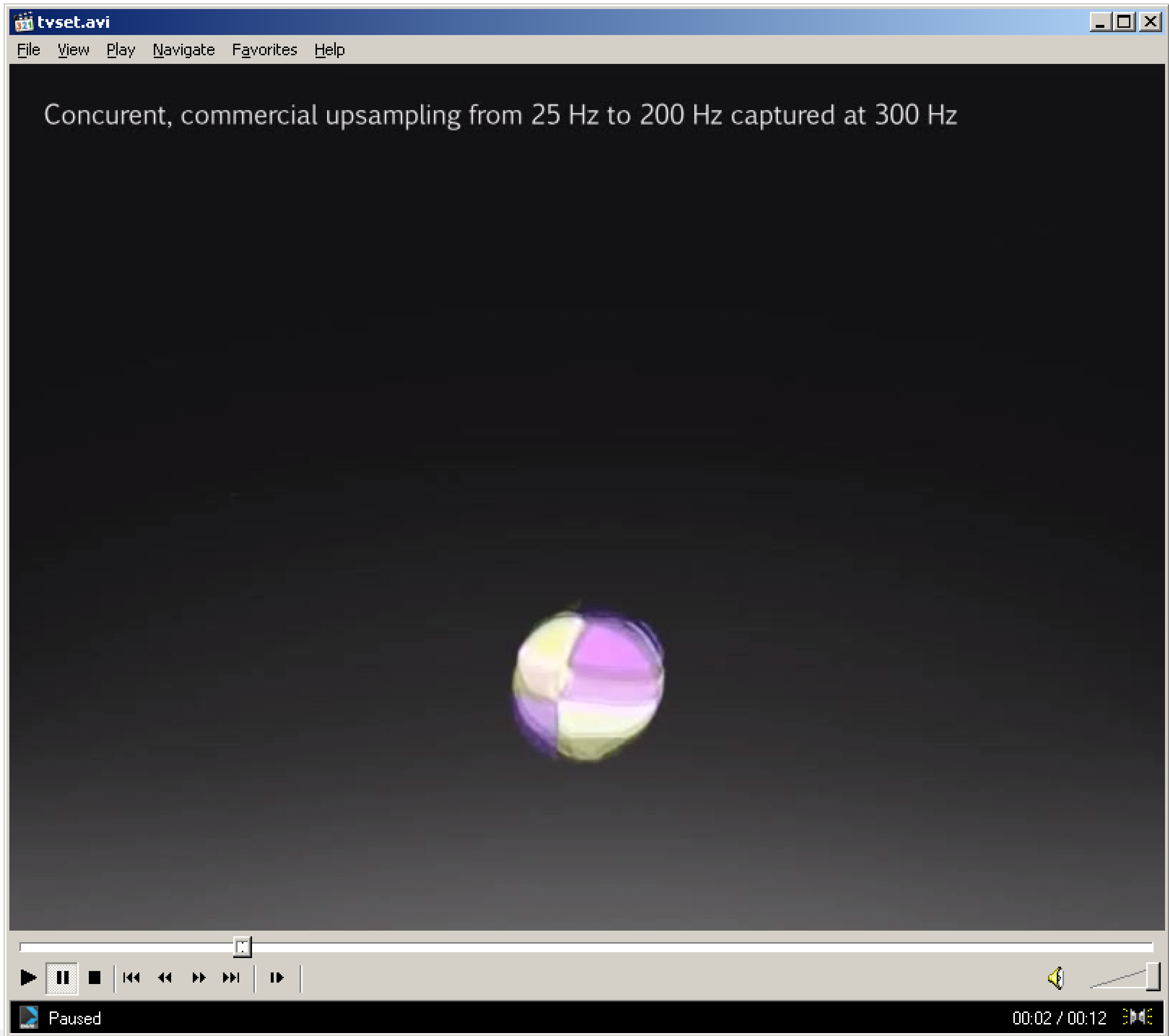
Alternative: Backlight flashing

J.H. Souk, J. Lee, Recent Picture Quality Enhancement Technology Based on Human Visual Perception in LCD TVs, 2007

Combating Hold-type Blur in TV Sets

- **Backlight flashing (BF)**
  - Turning the backlight of an LCD panel on and off
  - LED response is very fast, so flashing 500 Hz and more is possible
  - Flashing on can be synchronized with steady states of LC (reduces ghosting)
- **Motion compensated inverse filtering (MCIF)**
  - Filtering an input image, which aims at inverting hold-type blur
  - Effectively local 1D sharpening filtering, which is computed along the optical flow trajectories
  - Cannot restore frequencies that are completely removed by hold-type blur, but may magnify frequencies that are attenuated
  - Image saturation may cause problems
- **Hybrid Methods**
  - FRD + BF

High-speed Camera Recording: TV-Set



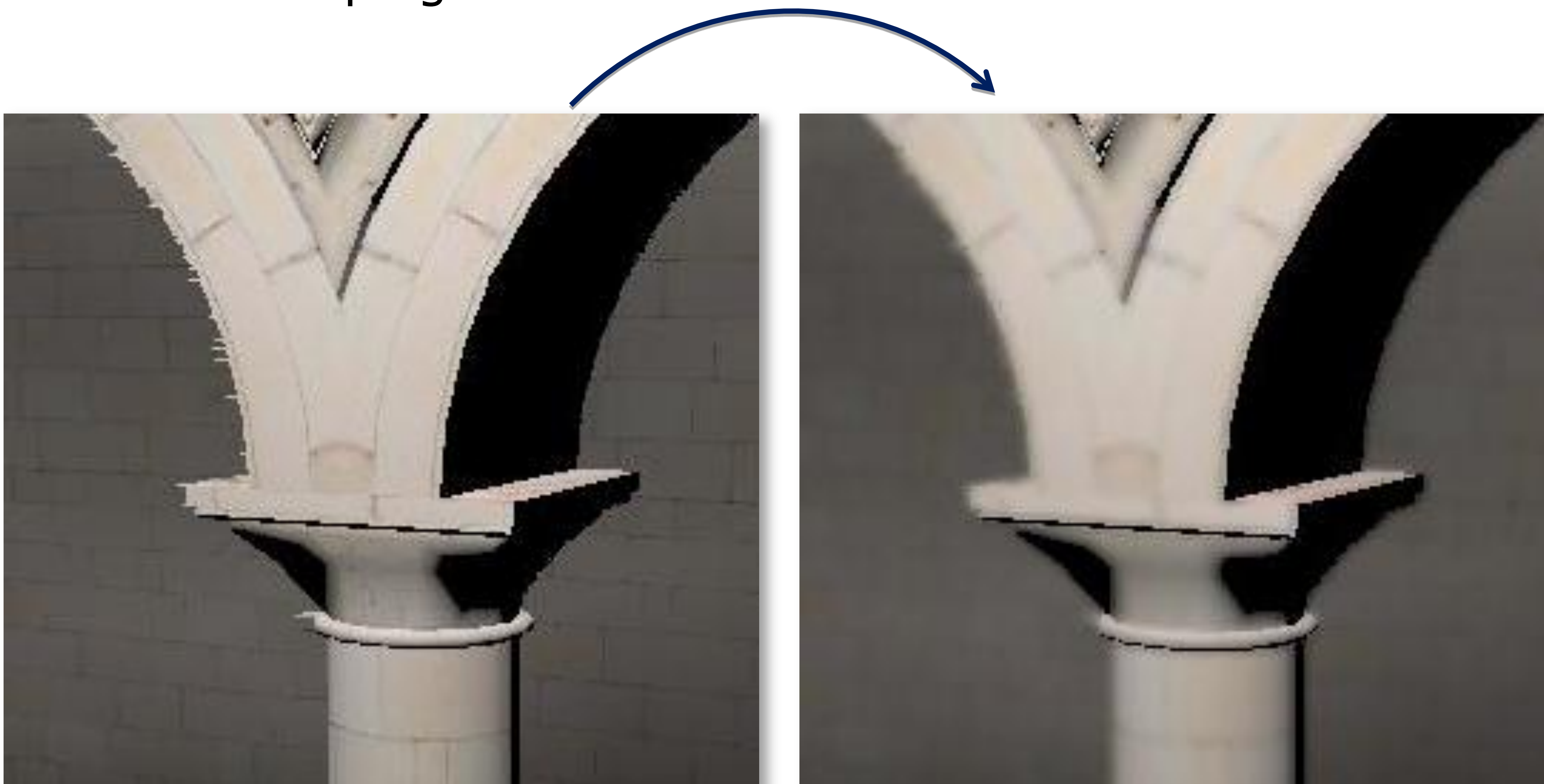
Combating Hold-type Blur in Rendering

Frame warping up to 120Hz

40 Hz rendering

[Didyk et al. 2010]

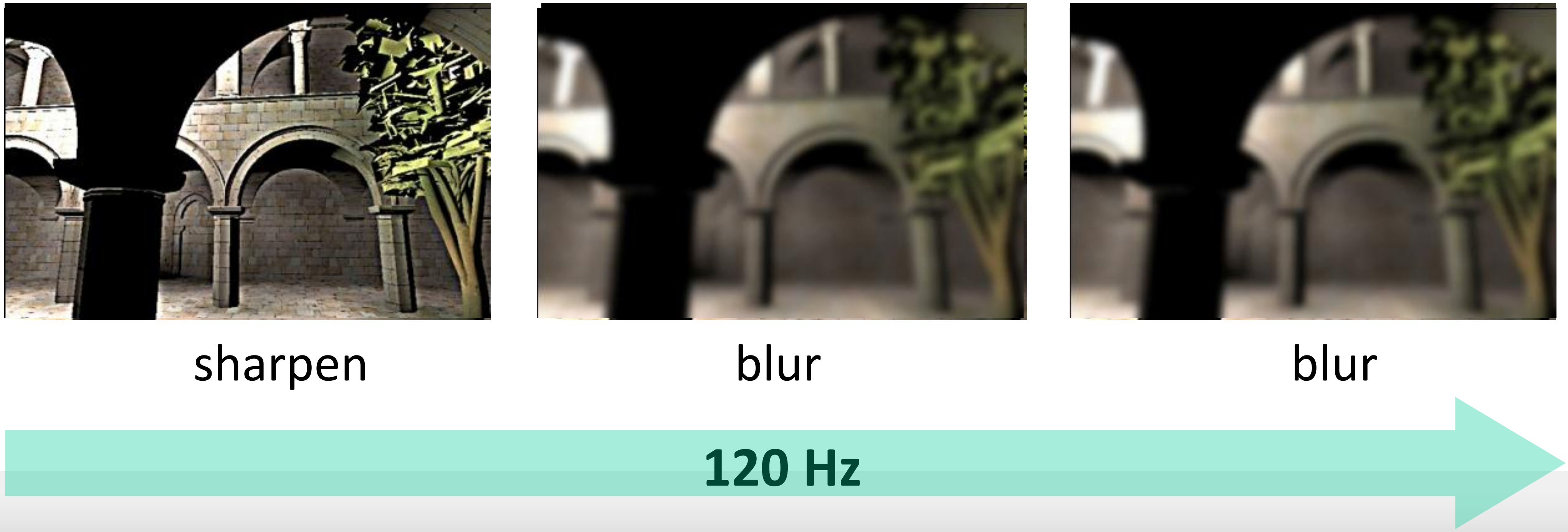
Blur out warping artifacts



[Didyk et al. 2010]

Combating Hold-type Blur in Rendering

- Interleave blurred and sharp (with doubled high-pass frequencies) frames
  - Energy-wise (brightness) equivalent
  - Blur filter size as a function of retinal velocity
  - Hold effect reduced as high frequencies displayed shorter and low frequencies do not matter for blur



[Didyk et al. 2010]

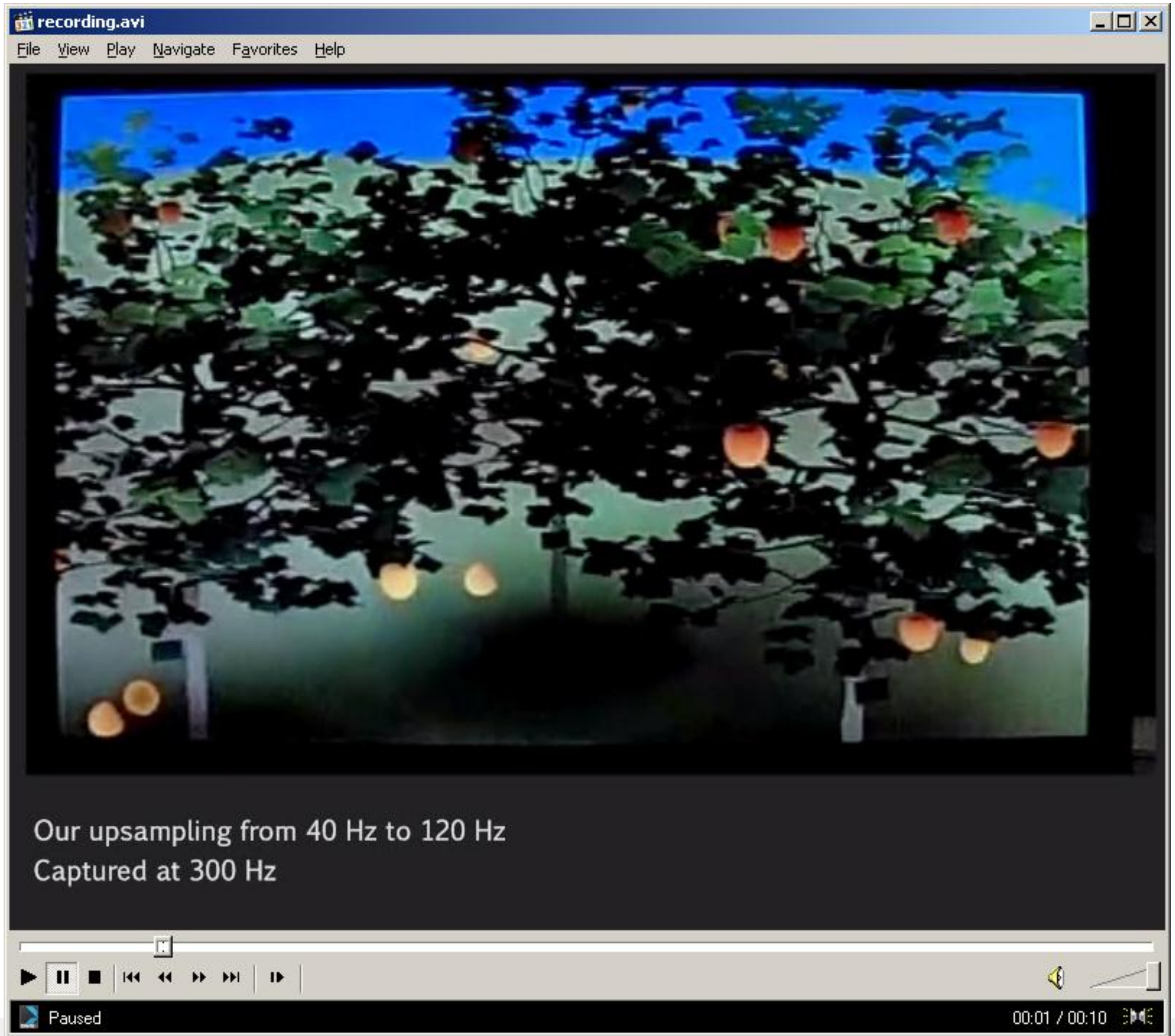
Perceived Blur Reduction Magnitude Estimation

40 Hz

120 Hz

[Didyk et al. 2010]

High-speed Camera Recording: Rendering

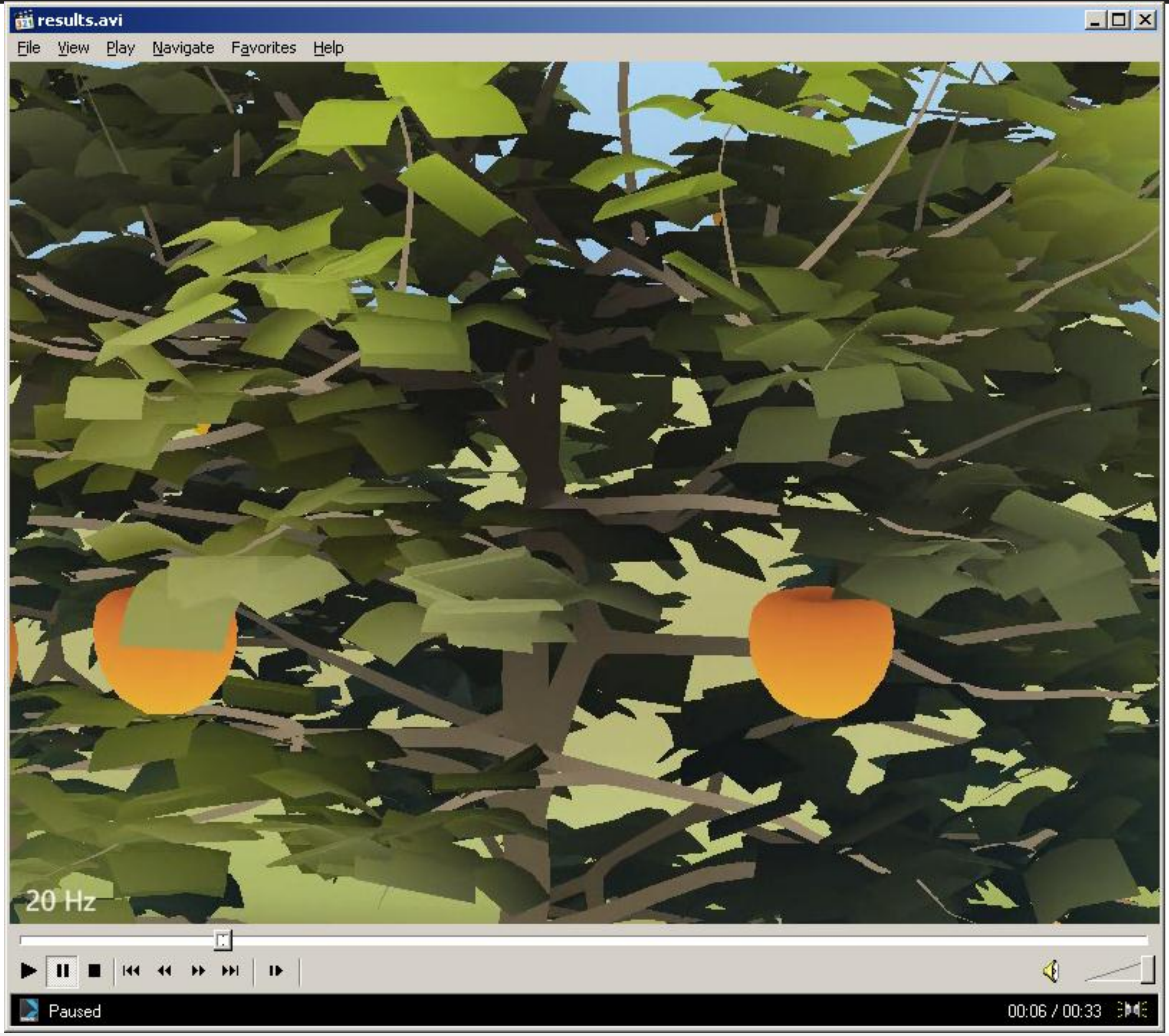


[Didyk et al. 2010]

Comparison

	BDI	BF	BFI	FRD	MCIF	Didyk et al.
LCD response required	High	Moderate	High	High	No	High
Backlight response required	No	High	No	No	No	No
Optical flow quality	No	No	No	High	Moderate	High
Ghosting artifacts	Possible	Possible	Yes	No	No	No
Flickering artifacts	Yes	Yes	No	No	No	No
Luminance reduction	Yes	Yes	No	No	No	No
Limitation of blur reduction	Flickering	Flickering	No	No	Freq. cut-off	No
Other possible artifacts	No	No	No	Fast motion	Oversaturation	No

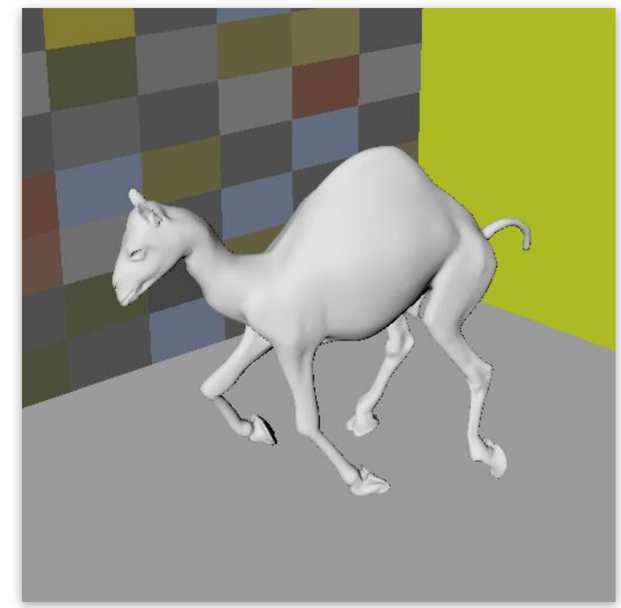
Rendering Comparison: Animation Examples

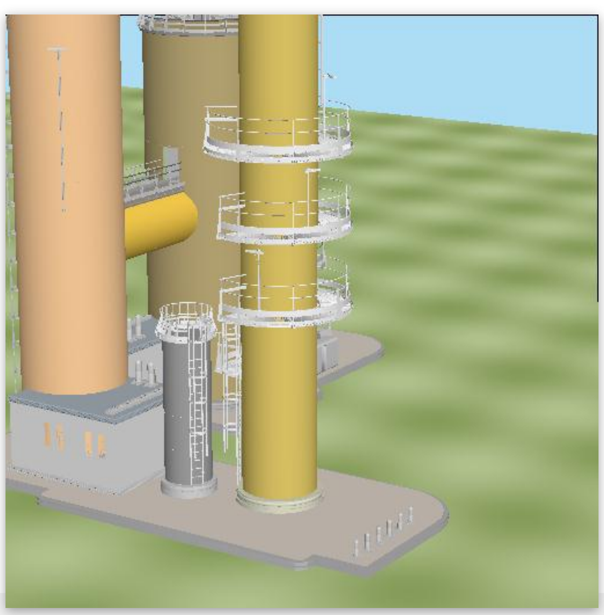


User Study

Pair-wise comparison

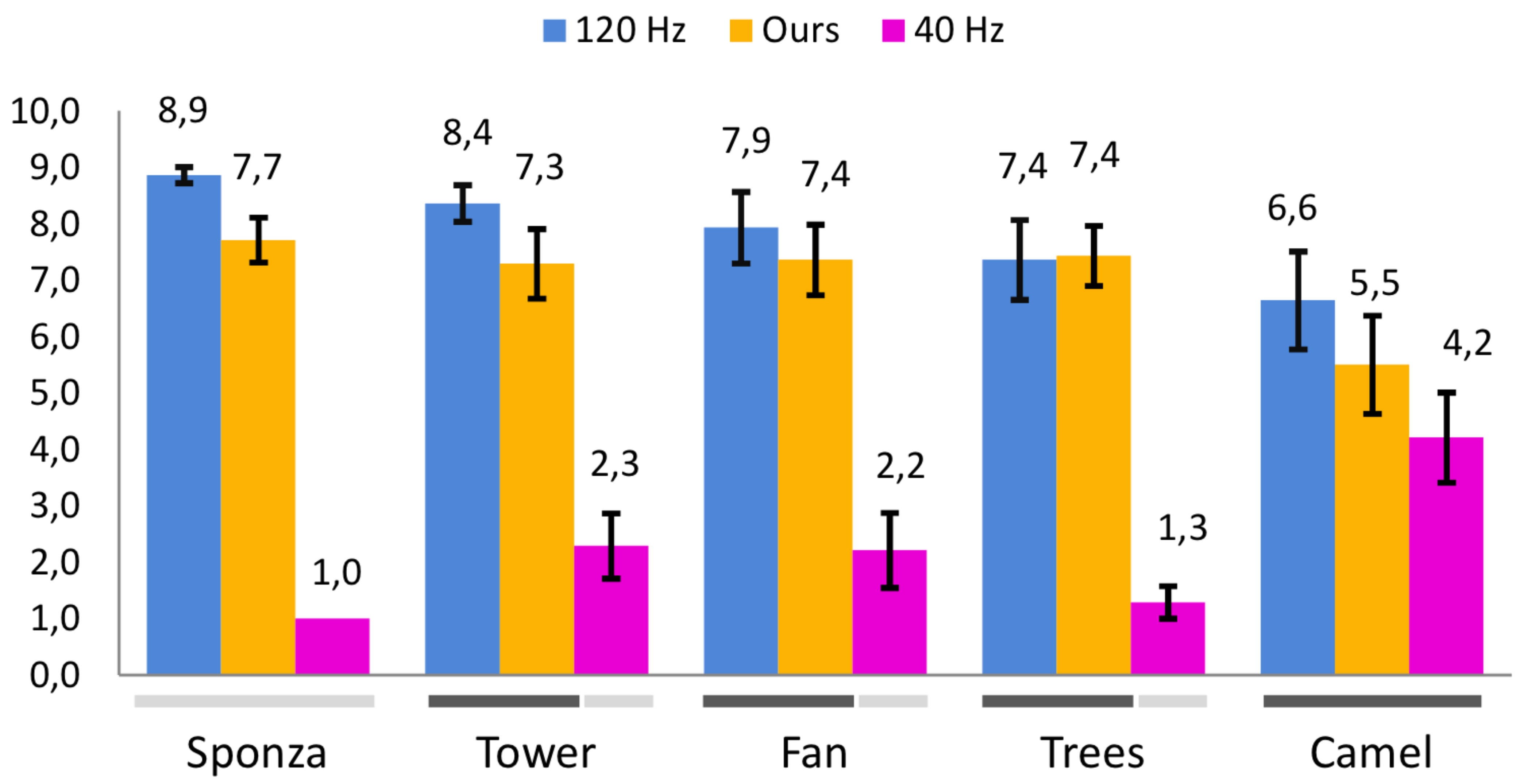
- 5 different sequences
- True 40Hz, 120Hz, Our 120Hz
- Blur judgment and artifacts





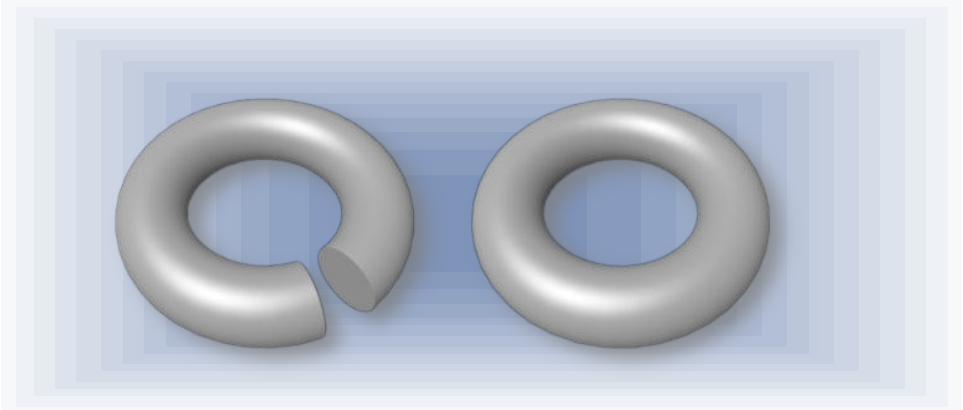
User Study

5 scenes (mean quality score + SEM)



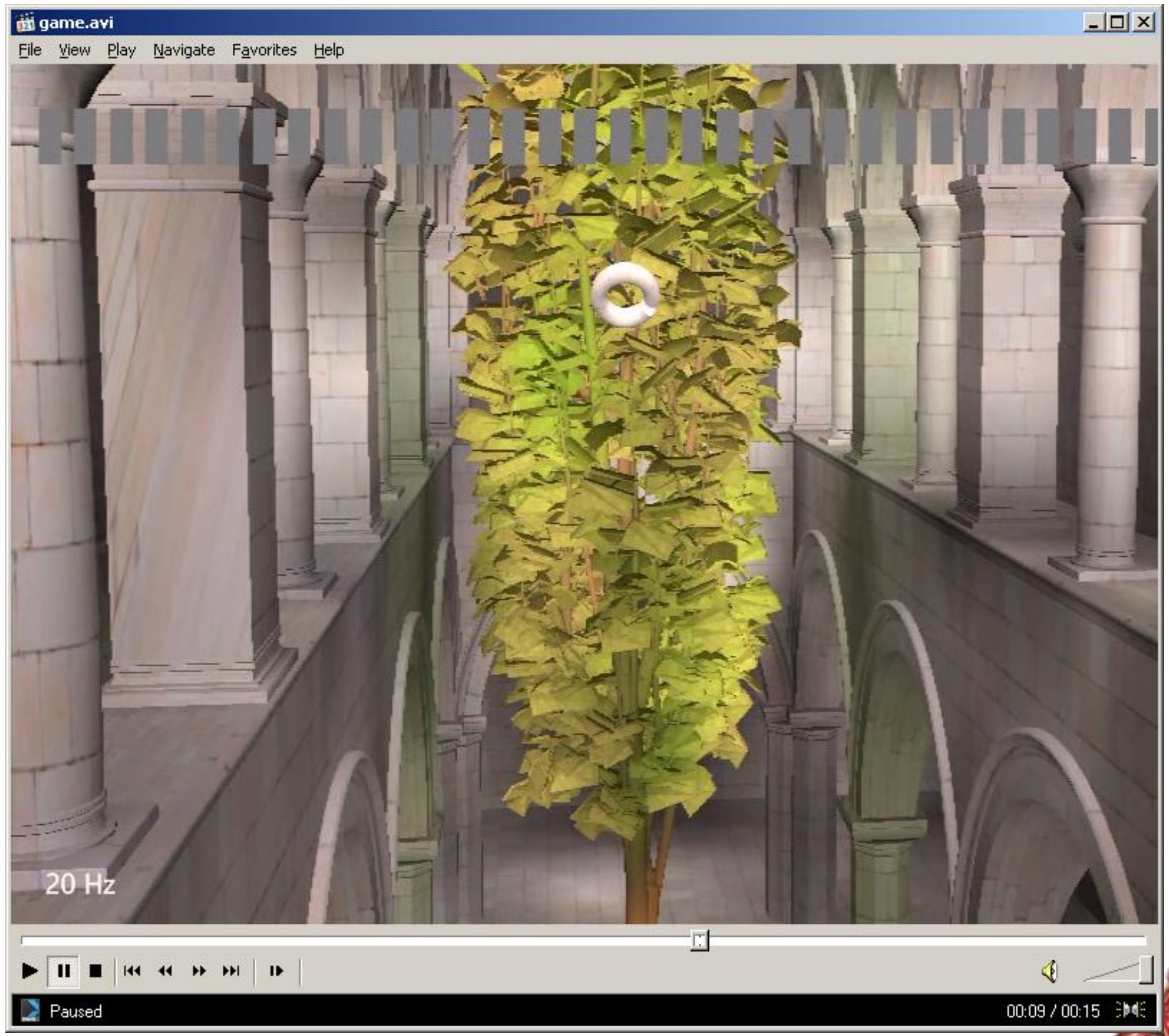
User Study: Game scenario

Targets:

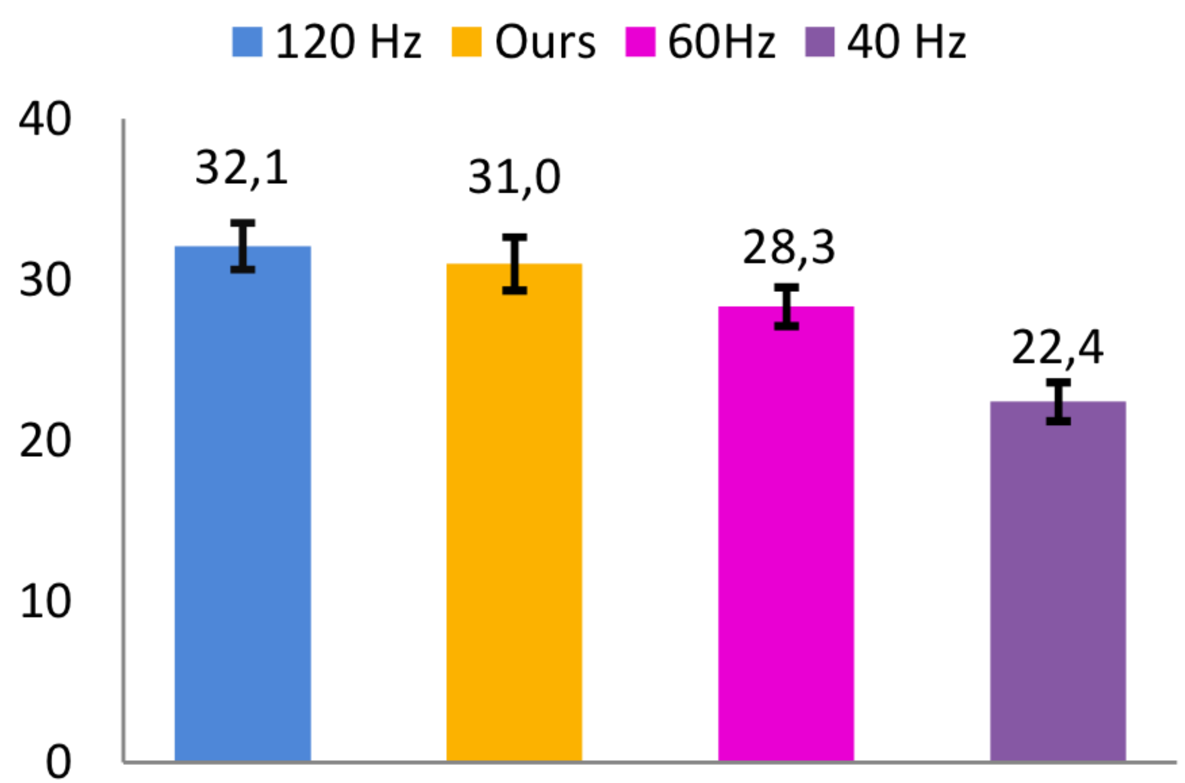


Task:

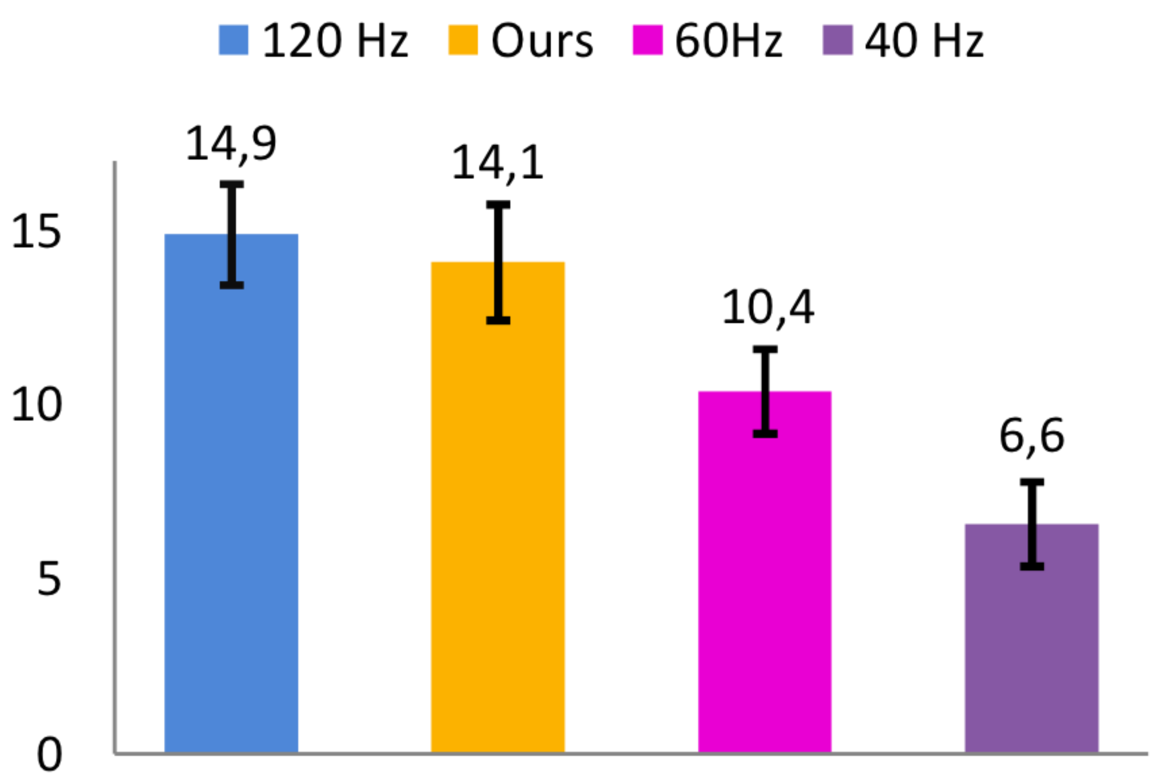
Detect open Landolt shape



User Study: Game scenario



Score



Succeeded Landolt detections

Changing Update Granularity

Why limit to the full frames if the eye can integrate signal @120Hz?

- Possible scenarios: update only 1 color channel, while the other two motion compensated
  - Does it pay off in terms of rendering costs?
- Local dimming behind fast moving and high contrast edges
  - Reduces hold effect
  - Flickering should not be a problem, but lost luminance should be compensated
- For HDR displays we could also control individually time/intensity of local LED backlight:
  - Fast moving objects **shorter**, but **brighter** LED impulses

3D Rendering vs. TV Solutions

- 3D rendering provides a lot of information, which is so difficult to recover based on images only (TV)
  - Precise motion flow, silhouette edges, textures,….
  - This should enable more sophisticated enhancement techniques integrated into rendering
- Perception + display device characteristics can be accounted for at rendering stage
  - Reducing hold effects

Rendering @120Hz

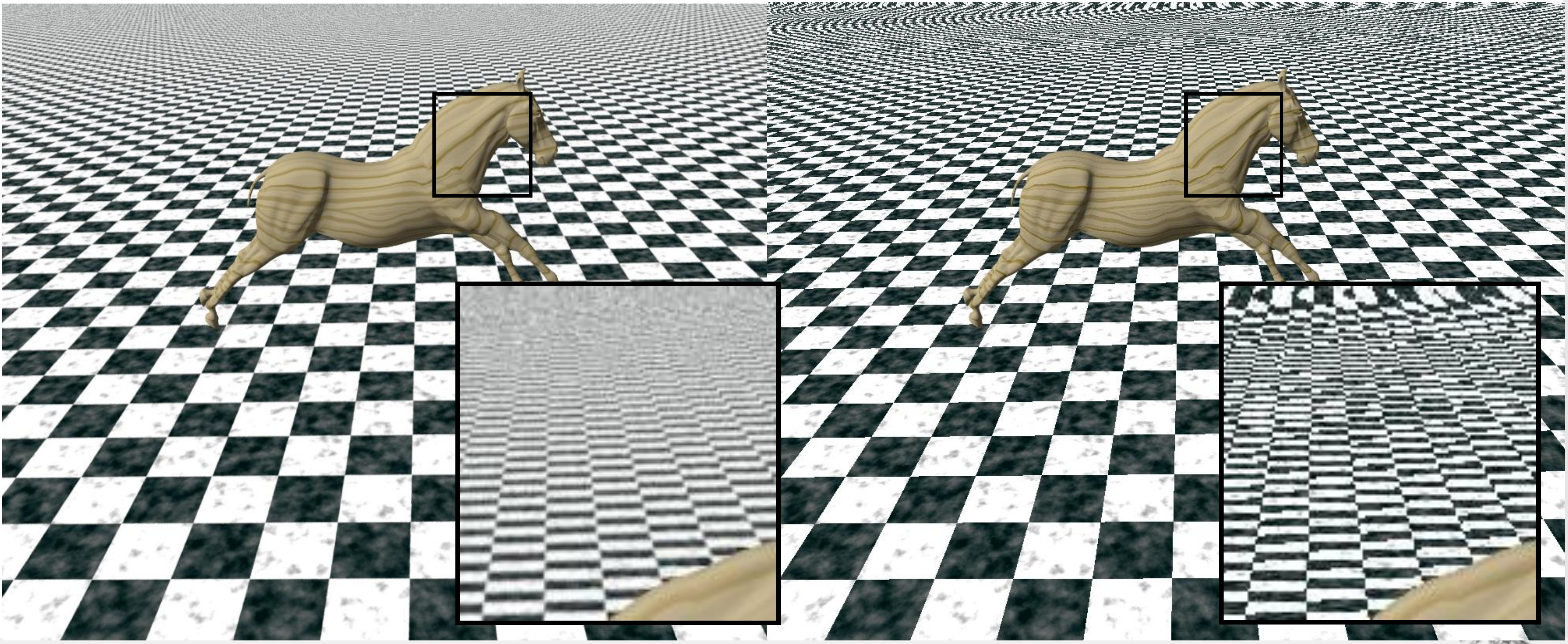
- We hope that the availability of 120 Hz displays can shift accents in rendering
  - More frames of much lower quality
  - Relying on integration in the eye
    - Interleaving such low quality frames at current display frequencies cause flickering, which should be much less visible at 120Hz
  - Extra frames over 60 Hz not wasted anymore

Reuse information

- Speed up: distribute workload over several frames

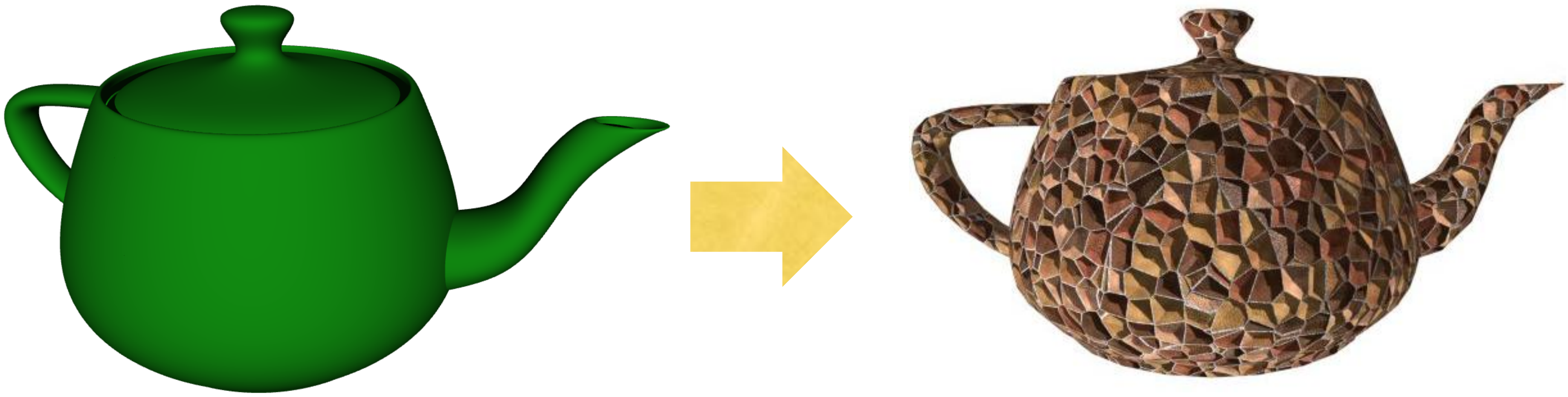
Reuse Information

- Increase in quality
  - Incorporate calculations from previous frames



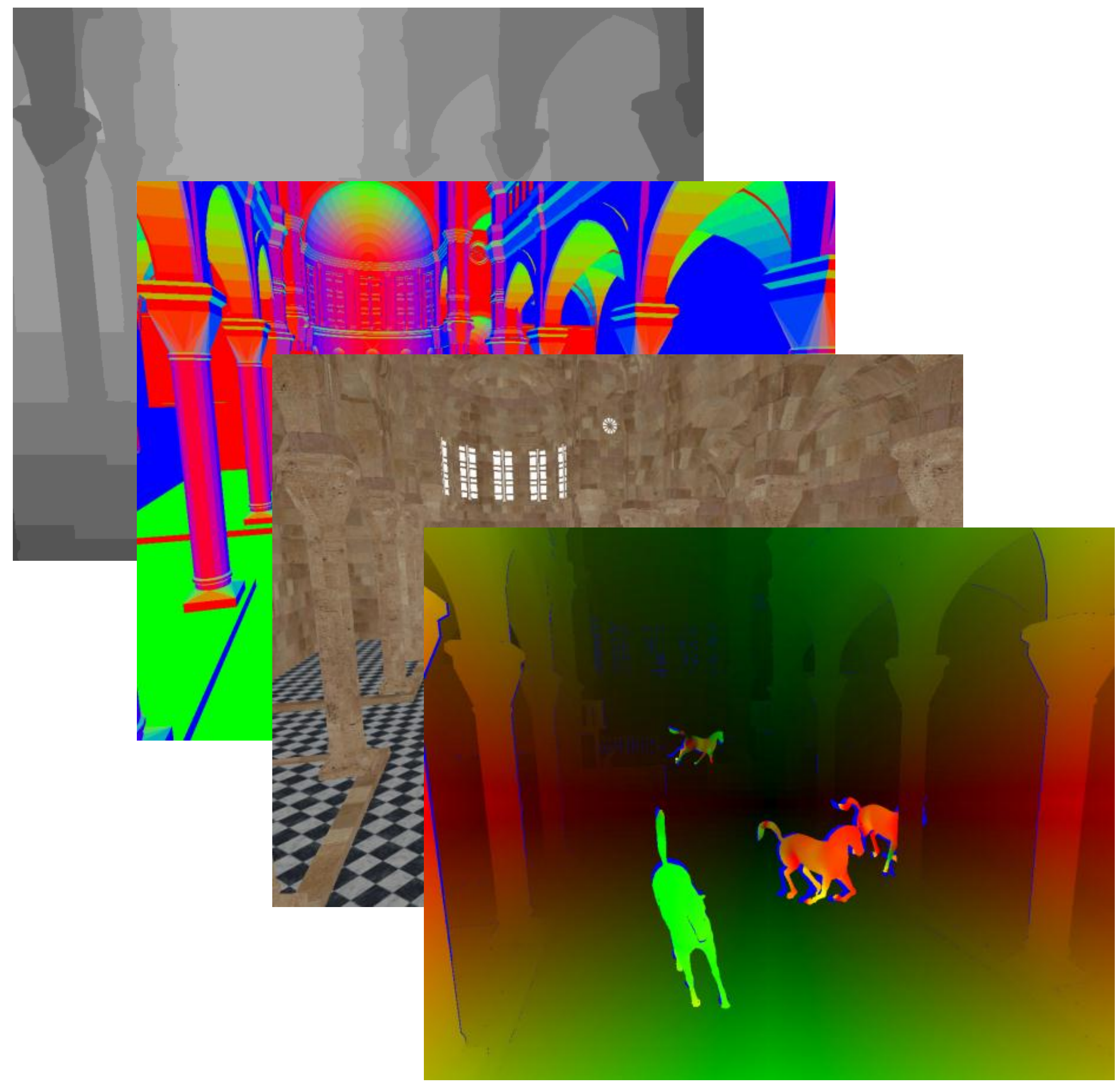
What is actually costly?

- Today's main cost is **shading**



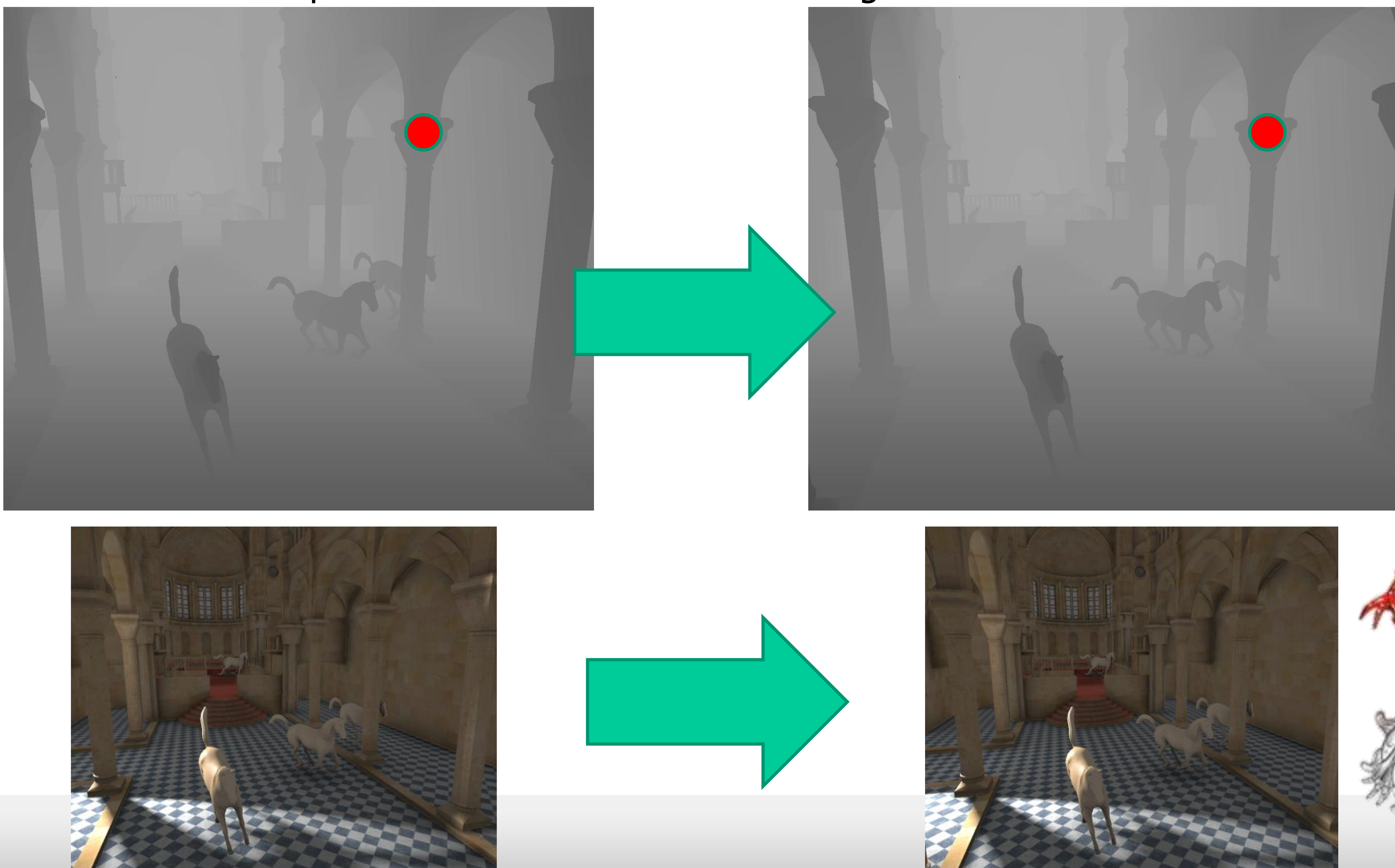
How to reduce shading cost?

- Observation: shading correlates with geometry
- World information behind pixel is for "free"
  - Depth (position)
  - Normals
  - Materials, Textures
  - Geometric motion flow

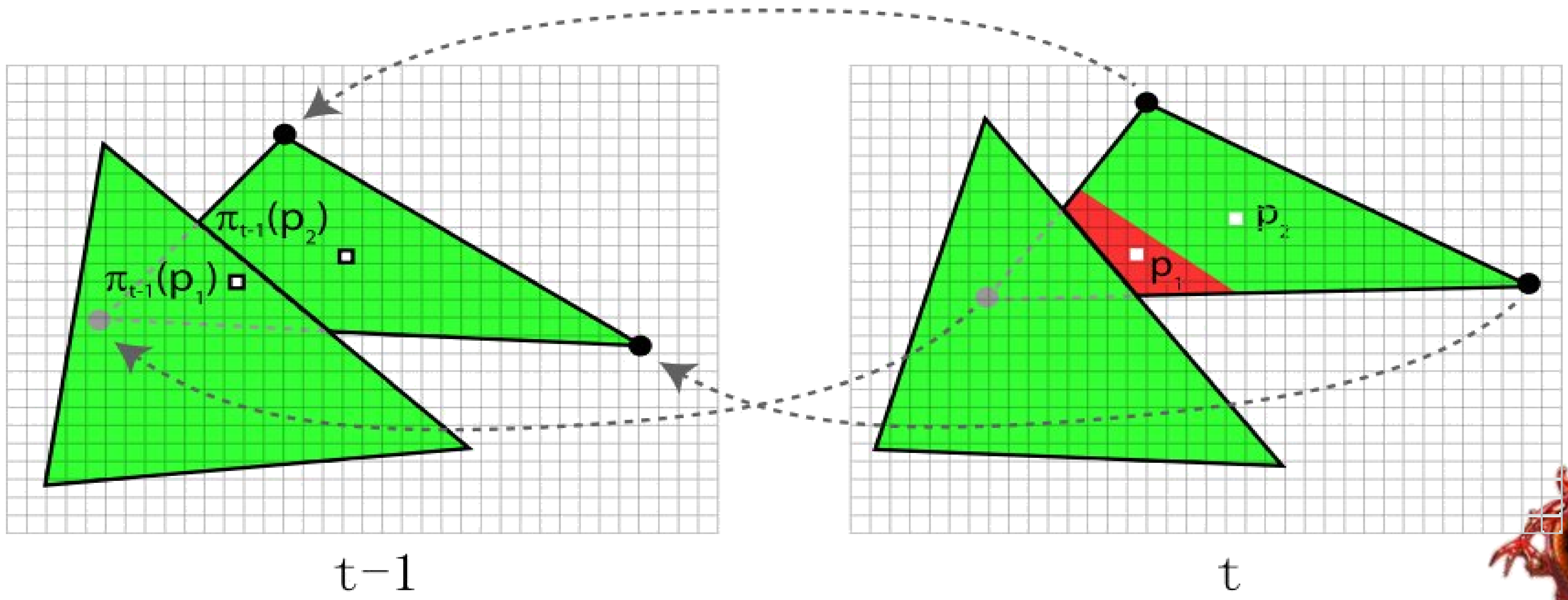


Why does rendering of depth & co. help?

- Find correspondences and transfer shading!



Not that simple...



Forward Reprojection

- Requires forward motion vectors
- Holes and gaps
- Difficult to implement with DX9/10

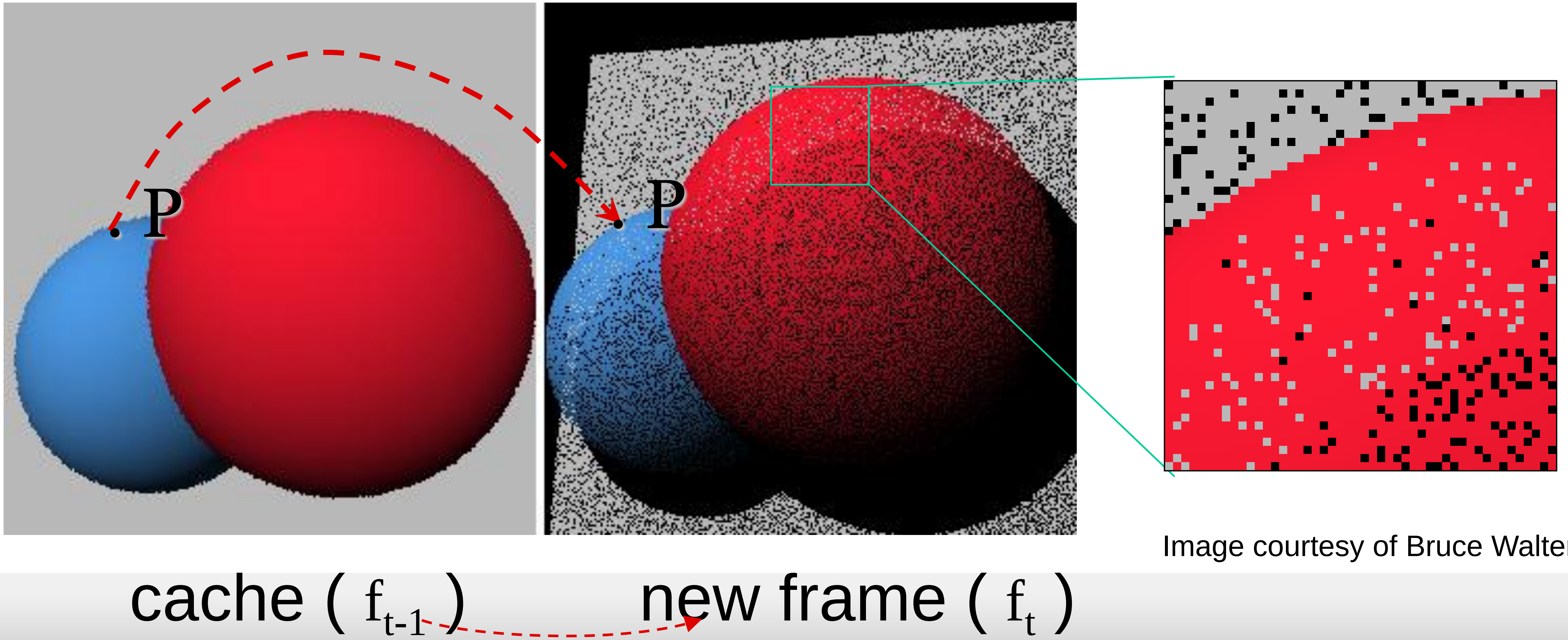
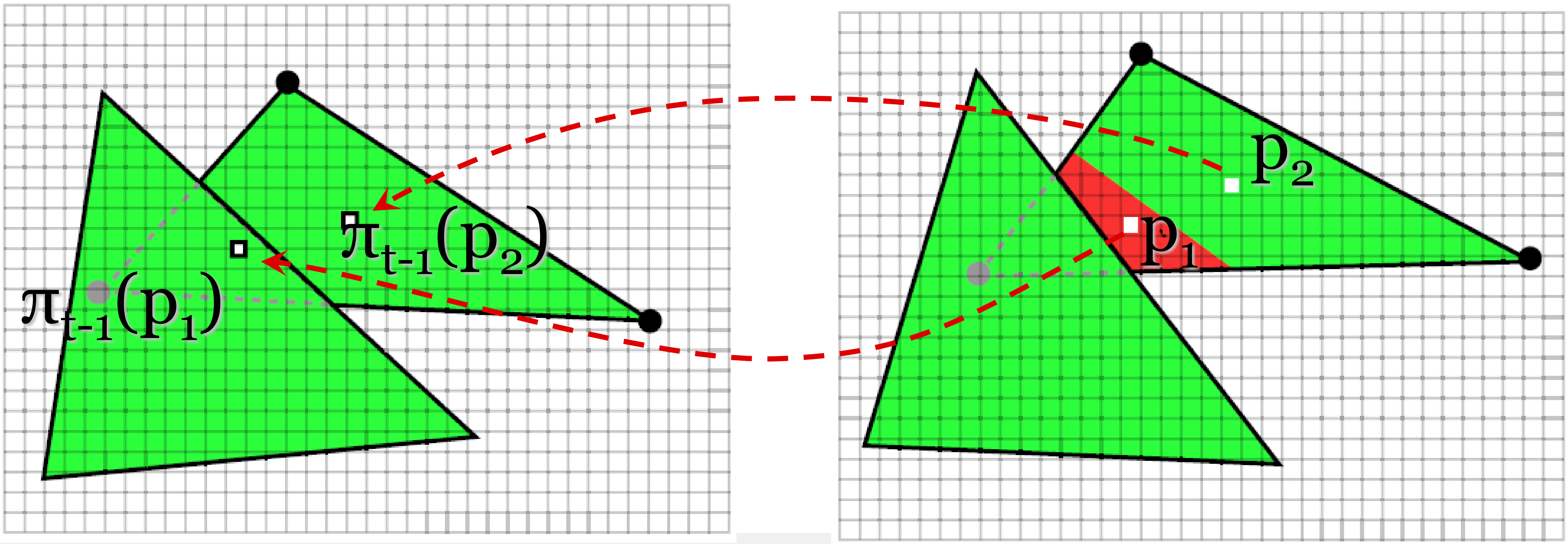


Image courtesy of Bruce Walter

Reverse Reprojection [Nehab 06/07, Scherzer 07]

- Reprojection operator $(x', y', z') = \pi_{t\text{-}1}(p)$
- Resolve occlusion: Test if $z' \approx d_{t\text{-}1}(x', y')$



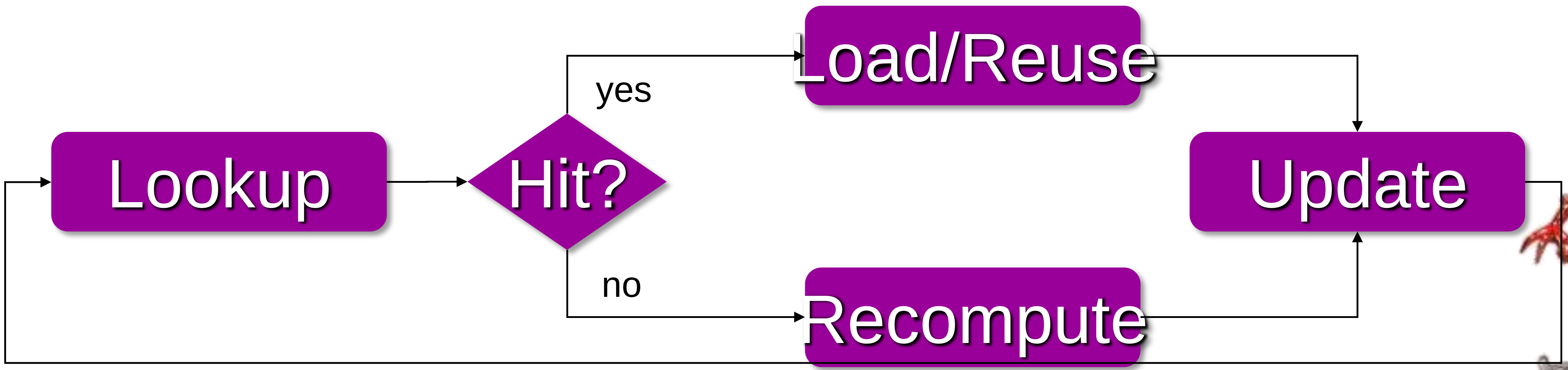
cache (f_{t-1}) new frame (f_t)

Reality Check

- Regular rendering loop (without using TC)
  - Recompute every pixel with original pixel shader

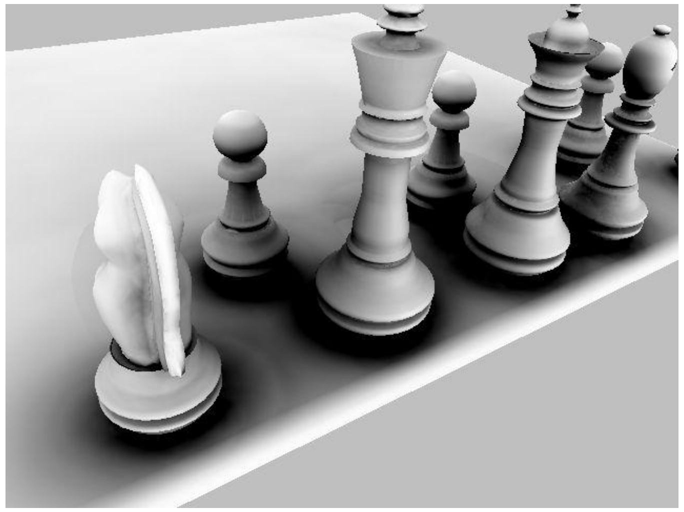
Reality Check

- Reuse previous results using the RRC
  - Reshade on demand
  - Cache reuse path must be cheaper for acceleration

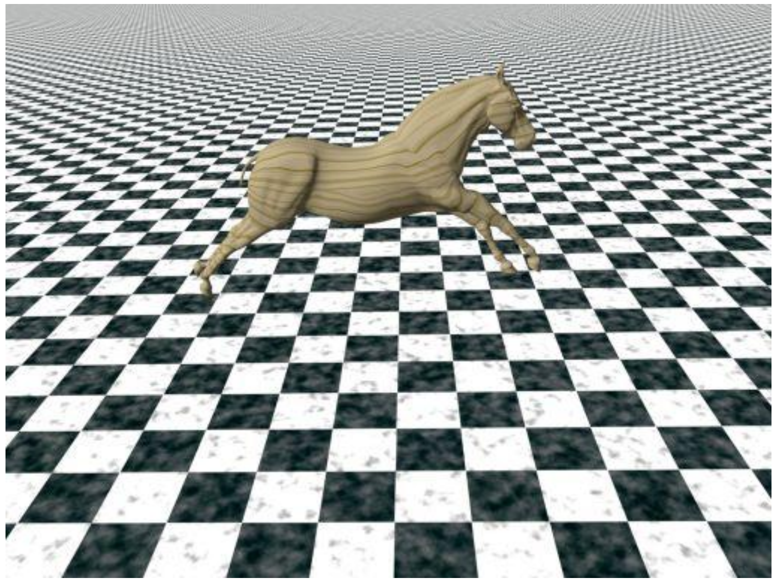


Good Examples to Cache

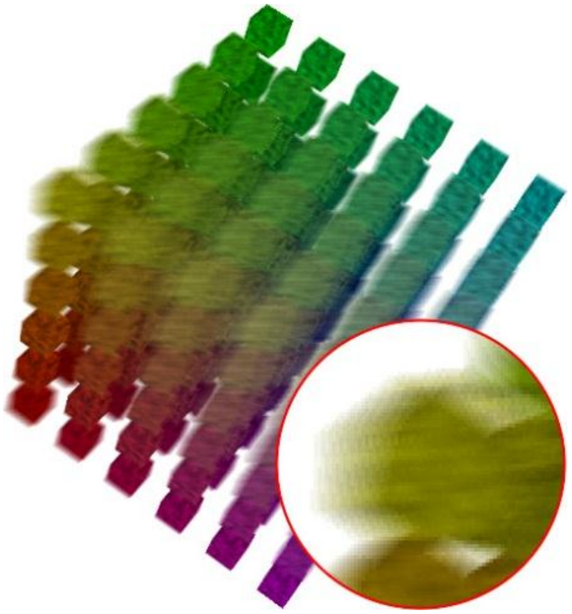
Static procedural texture



Global illumination

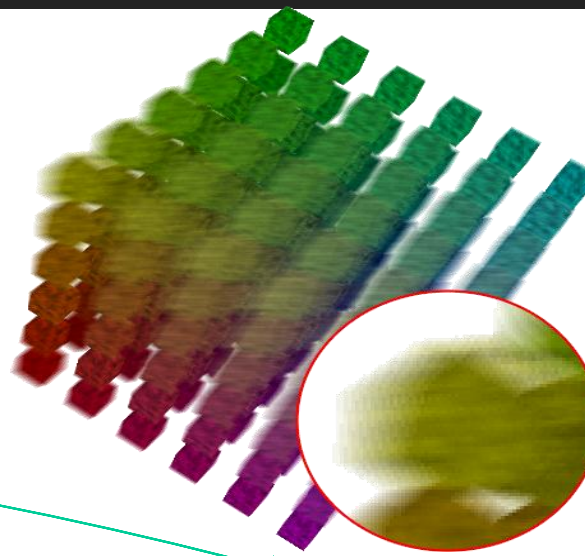


Numerical integral

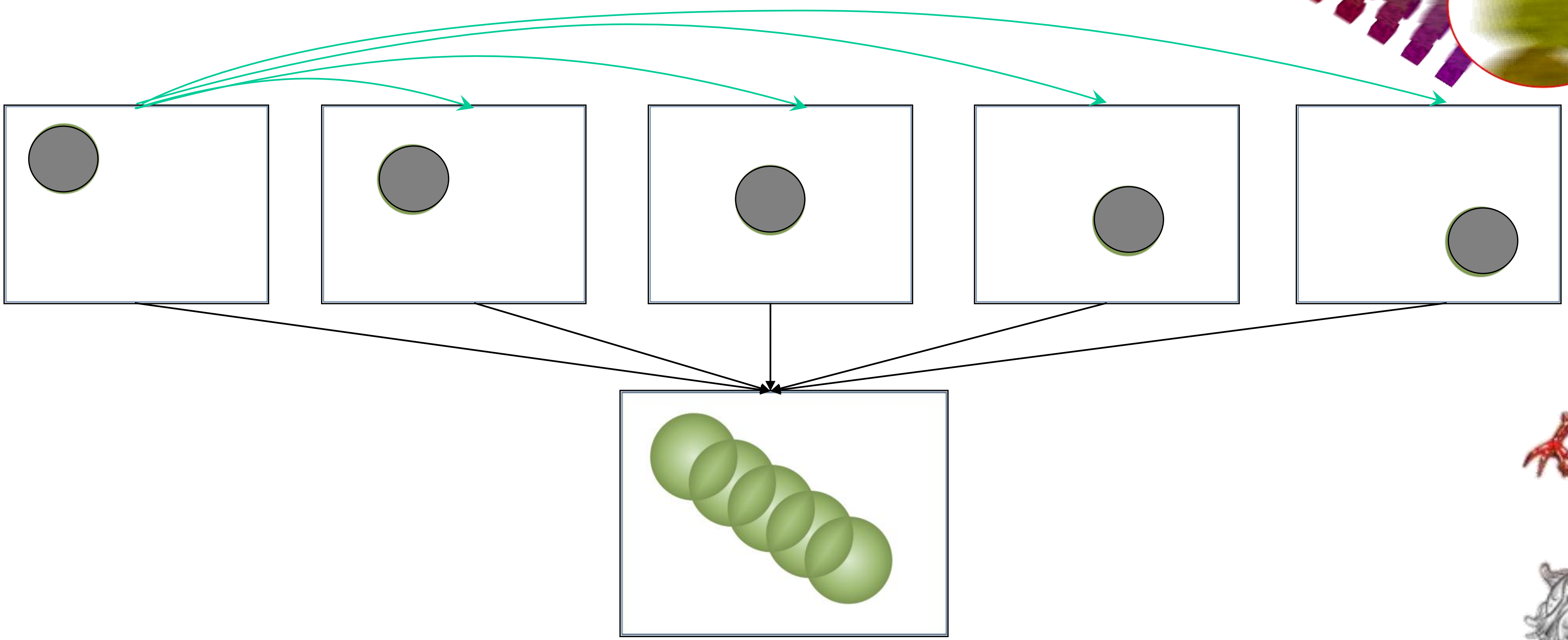


Multi-pass effects

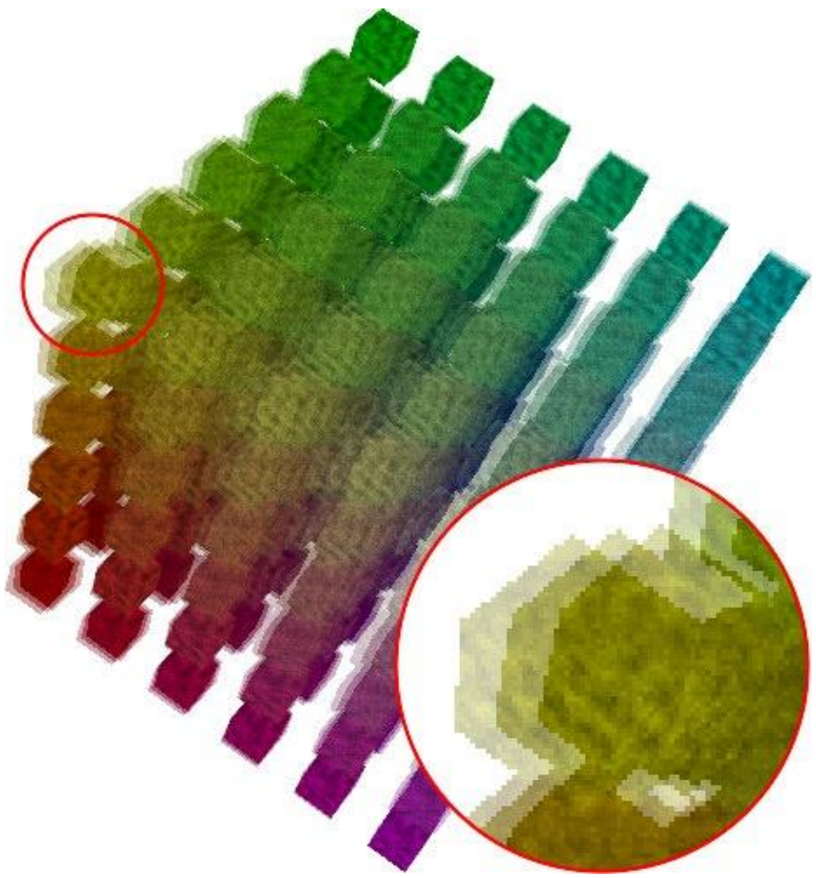
Multi-pass Rendering Effects



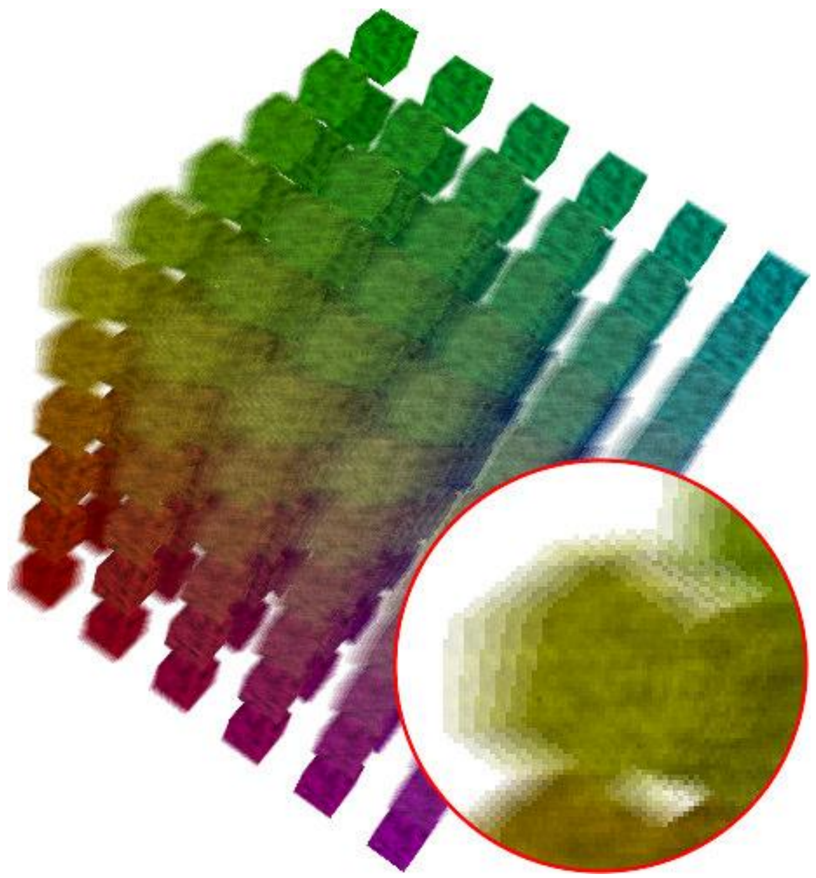
- Render a set of images with similar viewpoints
  - Shade one
  - Shade others via reprojection



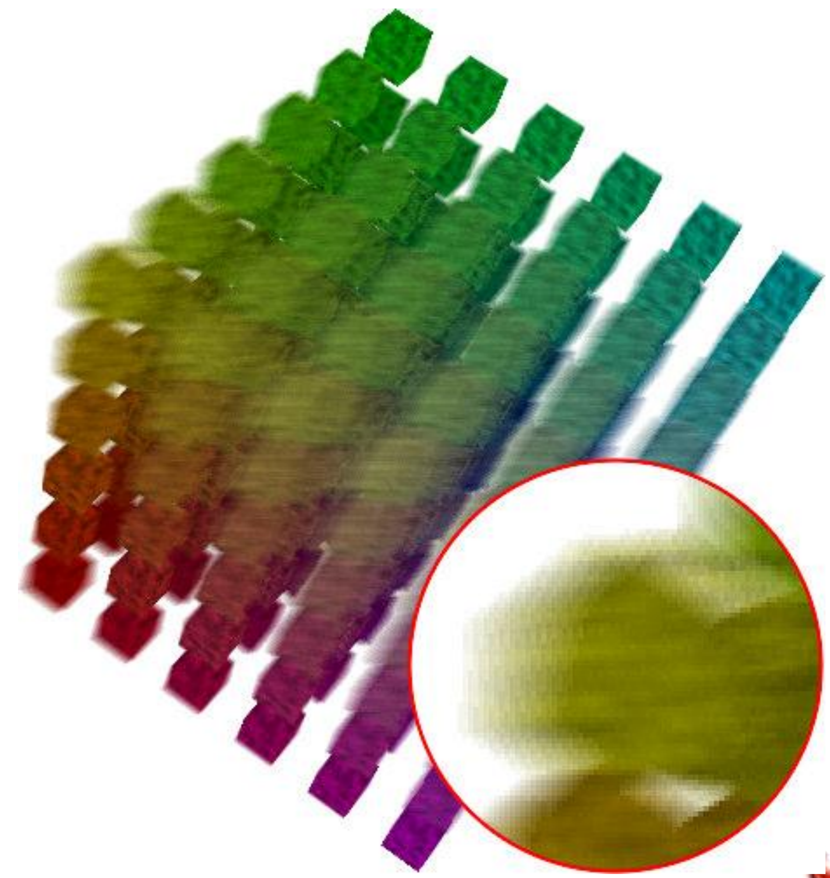
Motion Blur



3 time samples
60fps brute-force
60fps RRC

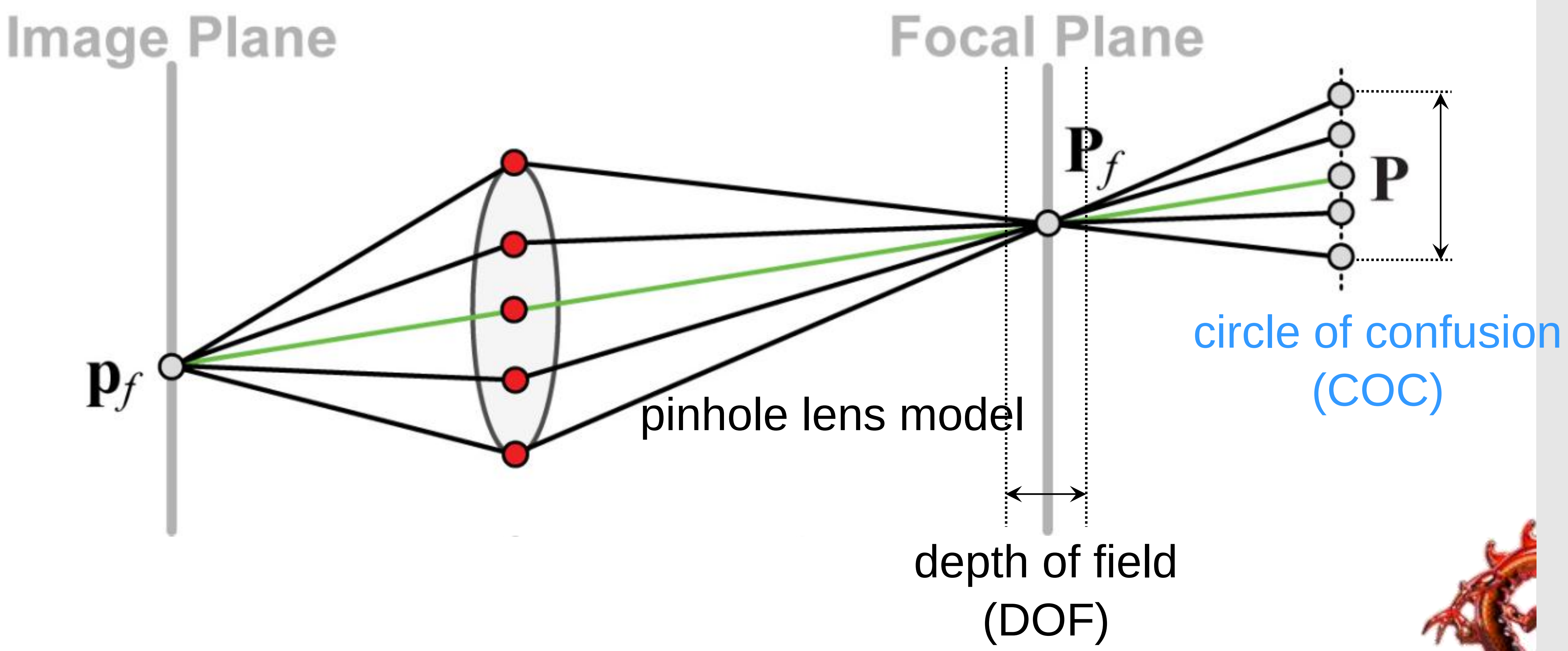


6 time samples
30fps brute-force
60fps RRC



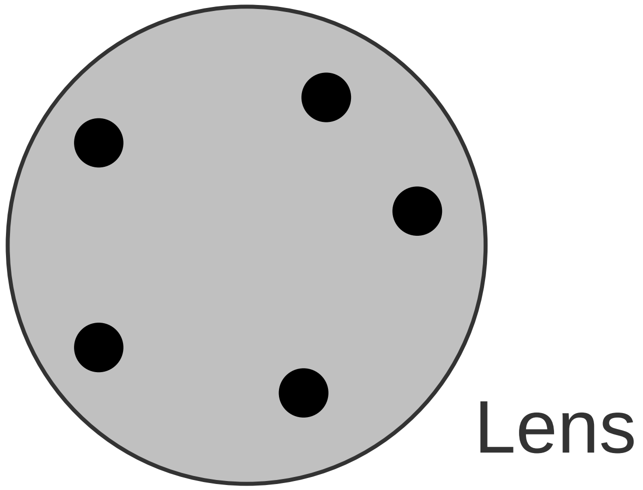
14 time samples
13fps brute-force
30fps RRC

Example: Depth of Field



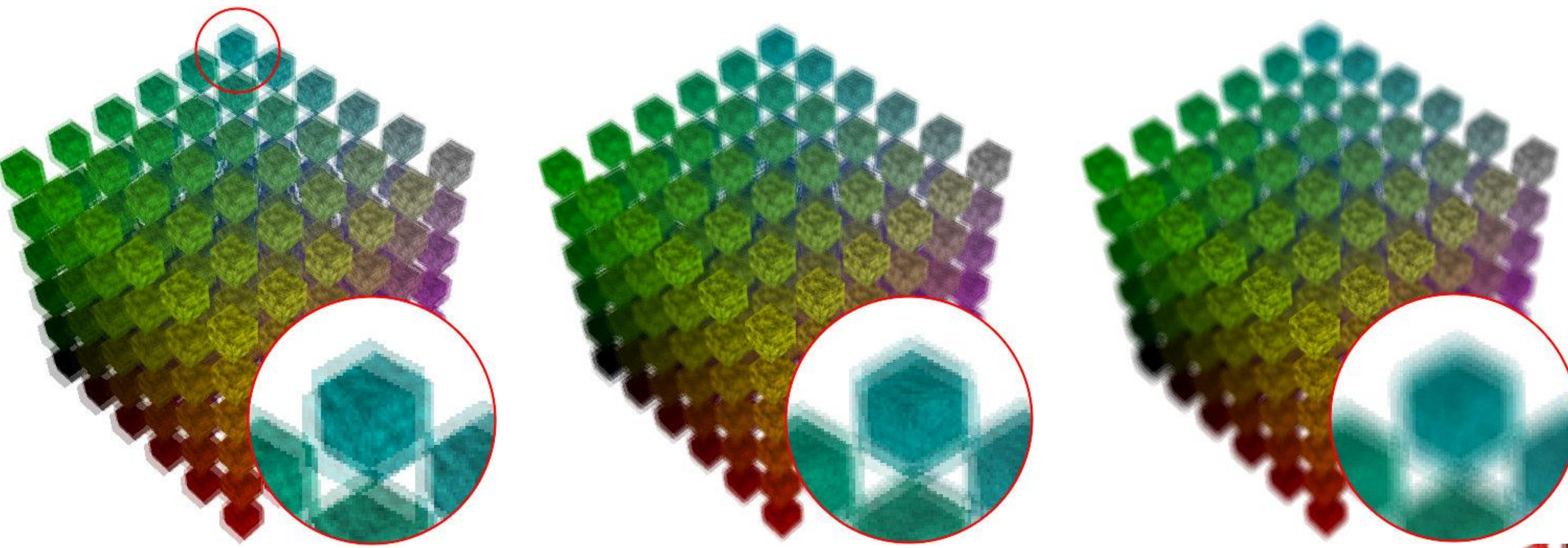
Our Algorithm

- View synthesis using image-based ray tracing



A few more “tricks” and you get…

Depth of Field



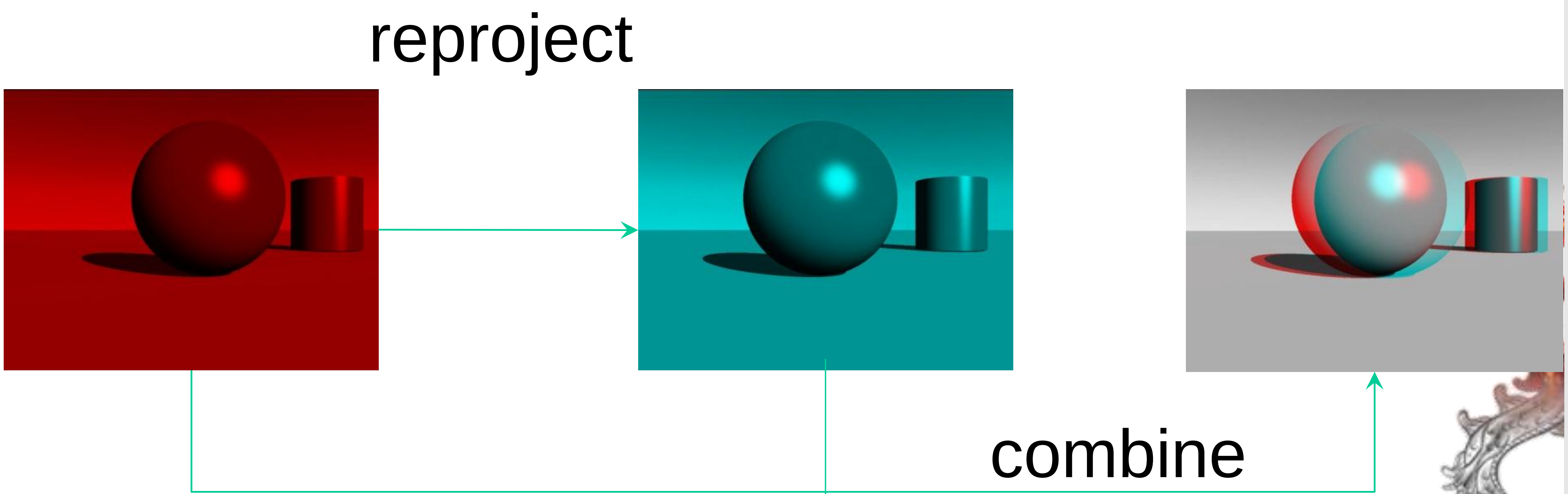
4 aperture samples
45fps brute-force
45fps RRC

9 aperture samples
20fps brute-force
45fps RRC

20 aperture samples
8fps brute-force
20fps RRC

Stereoscopic Rendering

- Generate images from two nearby views
  - Render the left eye normally
  - Render right eye with reprojection



This sounds amazing, but…

- So far: everything was static!
- Nothing moved… !
- How to deal with temporal changes?
  - Can we exploit spatial coherence?

Idea: use low resolution, then upsample

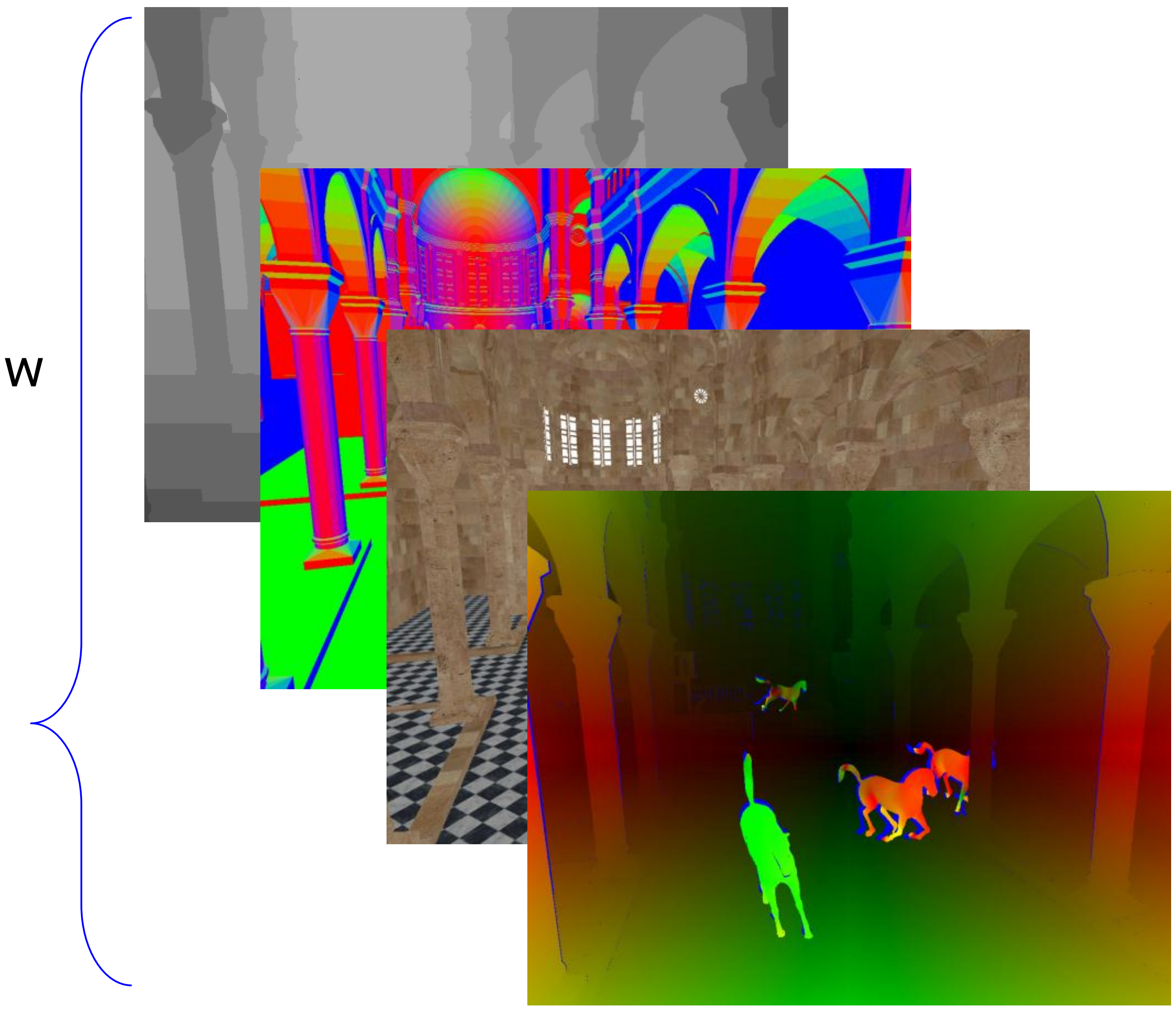
- Exploit spatial coherence:

Smart filter

? ≈

Remember?

- Observation: shading correlates with geometry
- World information behind pixel is for “free”
  - Depth (position)
  - Normals
  - Materials, Textures
  - Geometric motion flow



Joint-Bilateral Spatial Upsampling

Non-linear interpolation steered by geometry:

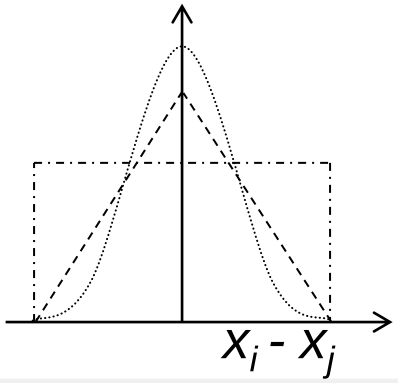
Low-res. shading input

$$h(i) = \frac{1}{\sum w_s} \sum_{j \in N\{i\}} w_s(i,j) \cdot l(\hat{j})$$

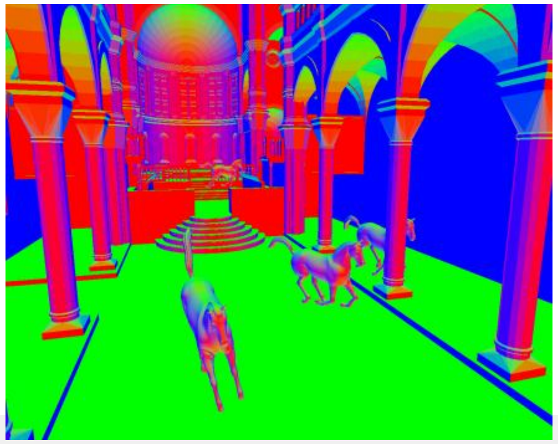
Image-space filter (e.g. hat/ box)

weights steered by geometry

$$w_s(i,j) = k(i,j) * d(z_i, z_j, \sigma_z^2) * n(\vec{n}_i, \vec{n}_j, \sigma_n^2)$$



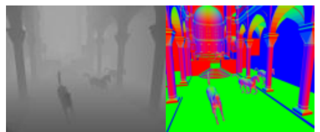
Depth (z)



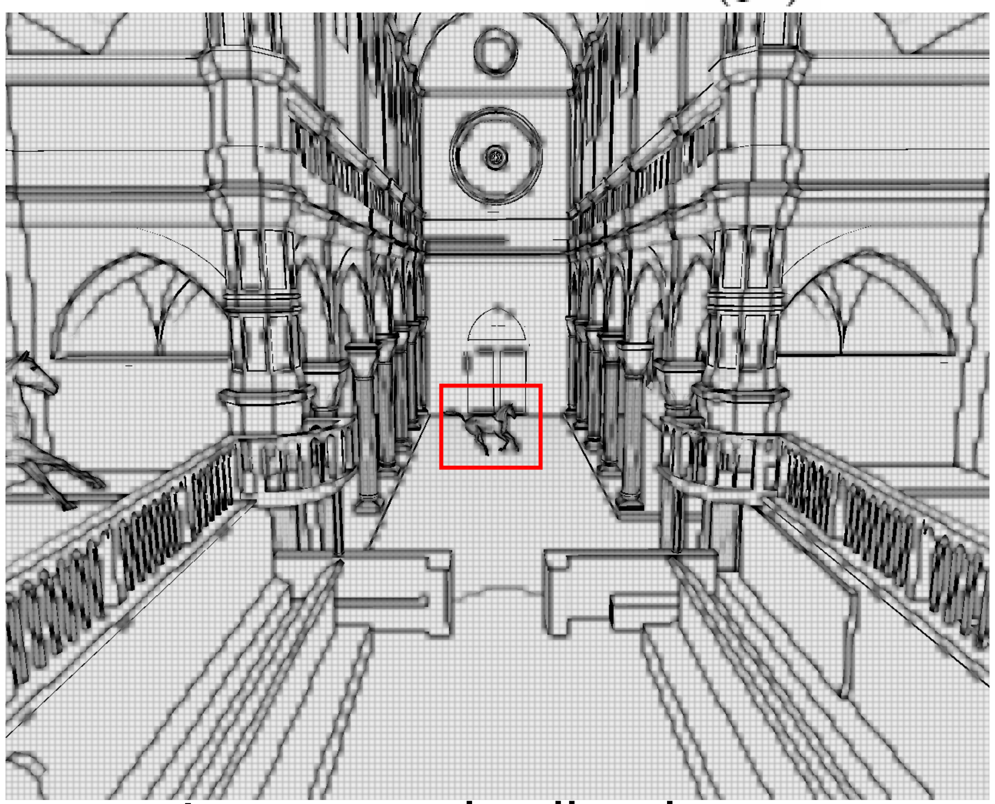
Normals

Joint/Cross-Bilateral Upsampling Revisited

$\frac{1}{\sum w_s}$

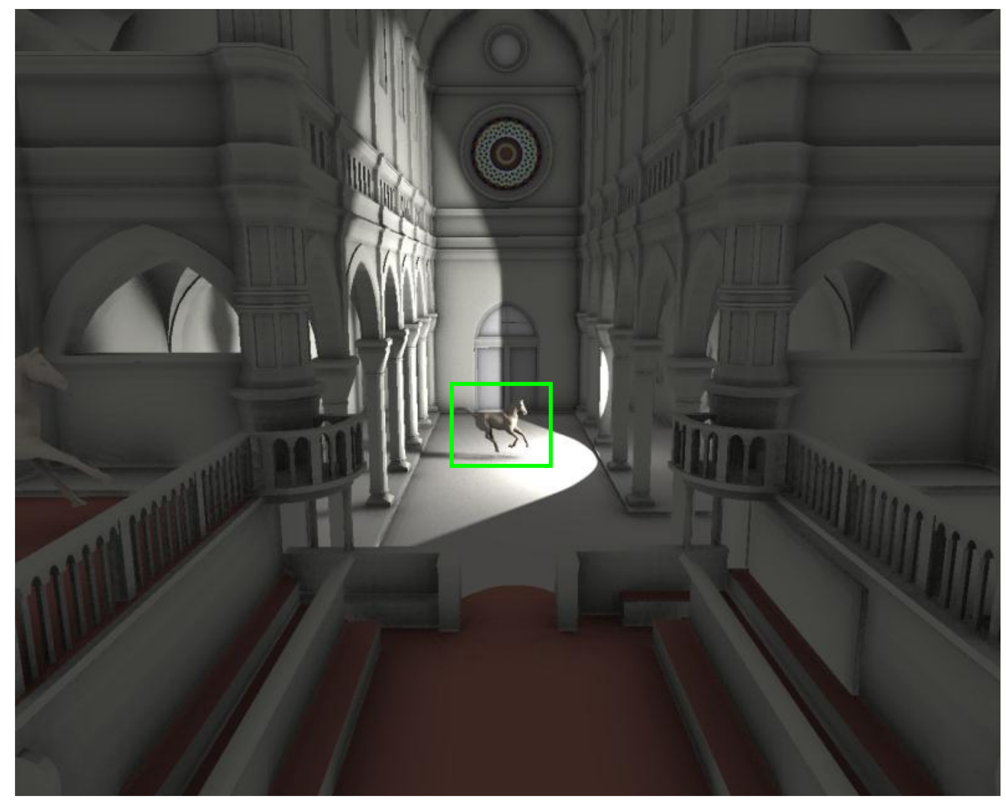


Reference:

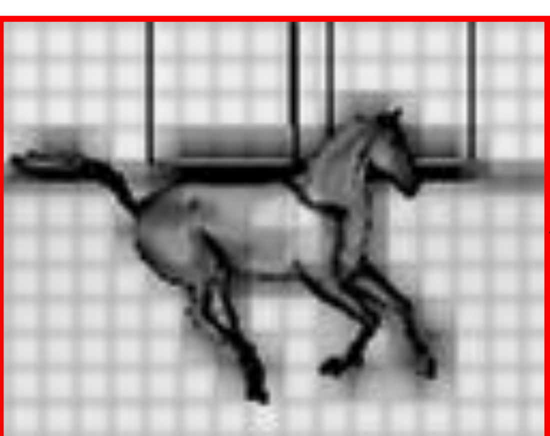


Low-res. shading input

$w_s(i, j)$ →



High-res. upsampled output



*

≈

Spatio-Temporal Upsampling

- Choose preferable method:

combine spatial upsampling

& temporal caching

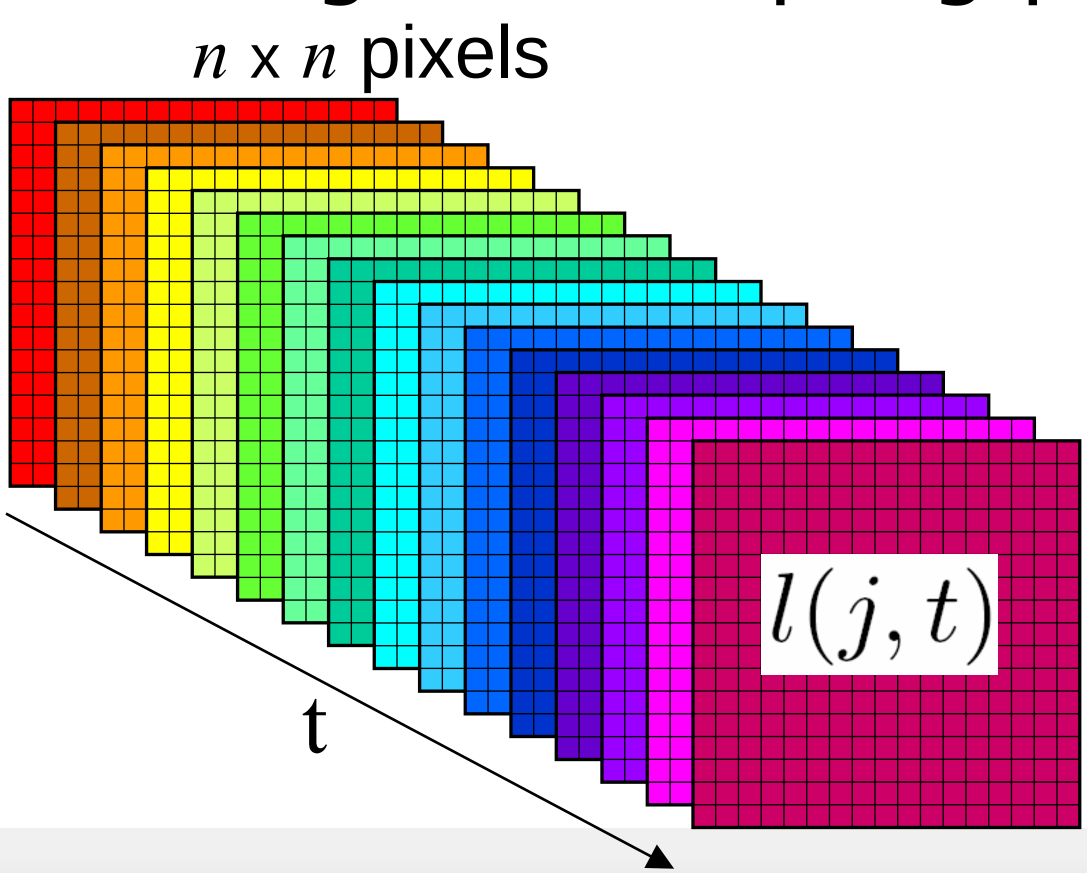
Gain information over time?

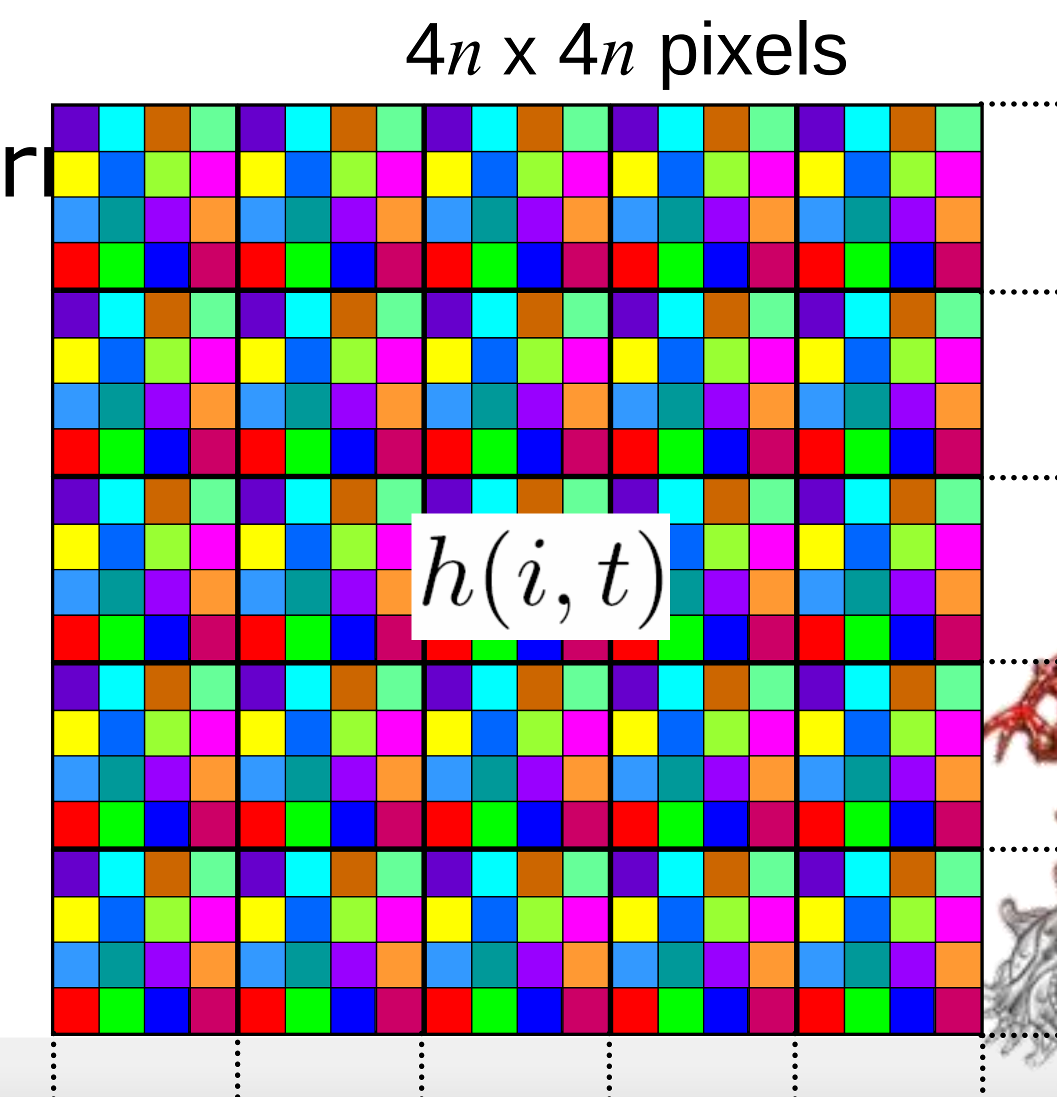
- The same low-res image gives the same information…

Temporally Interleaved Sampling

- Cache different pixel positions to upsample over time
  - → Refresh out-dated pixels (e.g. every **k** × **k** frames)

Regular sampling pattern





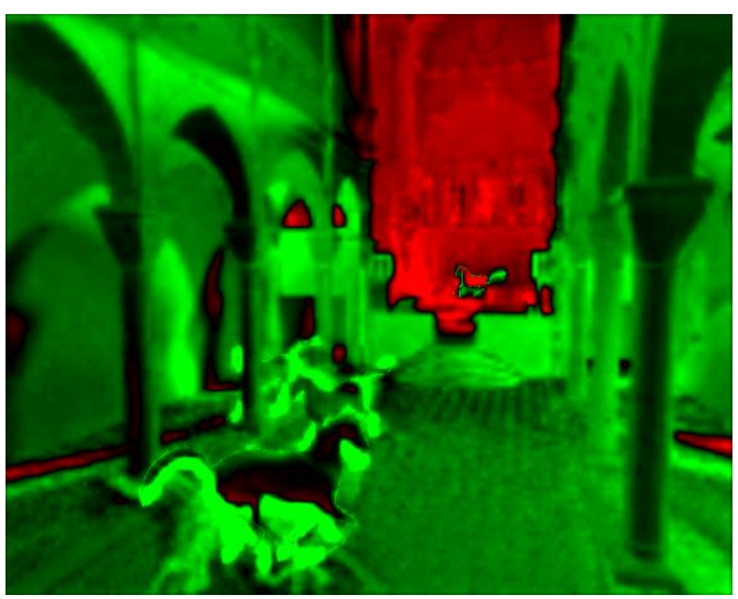
Putting things together:

- temporal

  Jittering -> more information for static over time

- Spatial

  Bilateral Upsampling (low2high) -> responsiveness



Choose according to change

4x4 upsampled result

Static Frame Convergence

Spatio-Temporal Upsampling [Herzog et al. 2010]

- Beneficial to use

  Spatial

  & temporal upsampling

  - Static frame convergence

  - Robustness with respect

    to changing lighting conditions

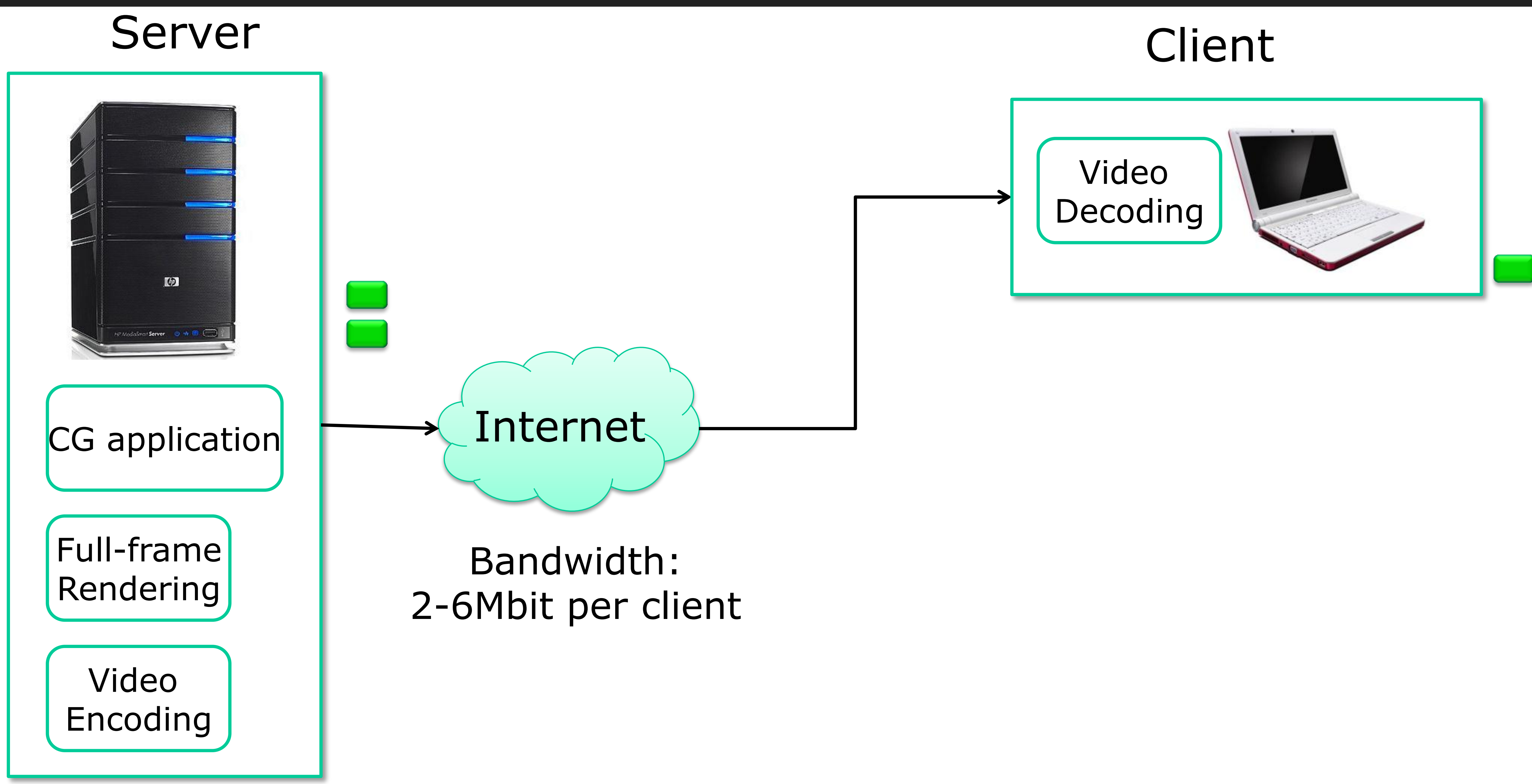
Extension: Remote Rendering

- OnLive, OToy, Gaikai rely on video encoding

  Naturally exploit coherence in video

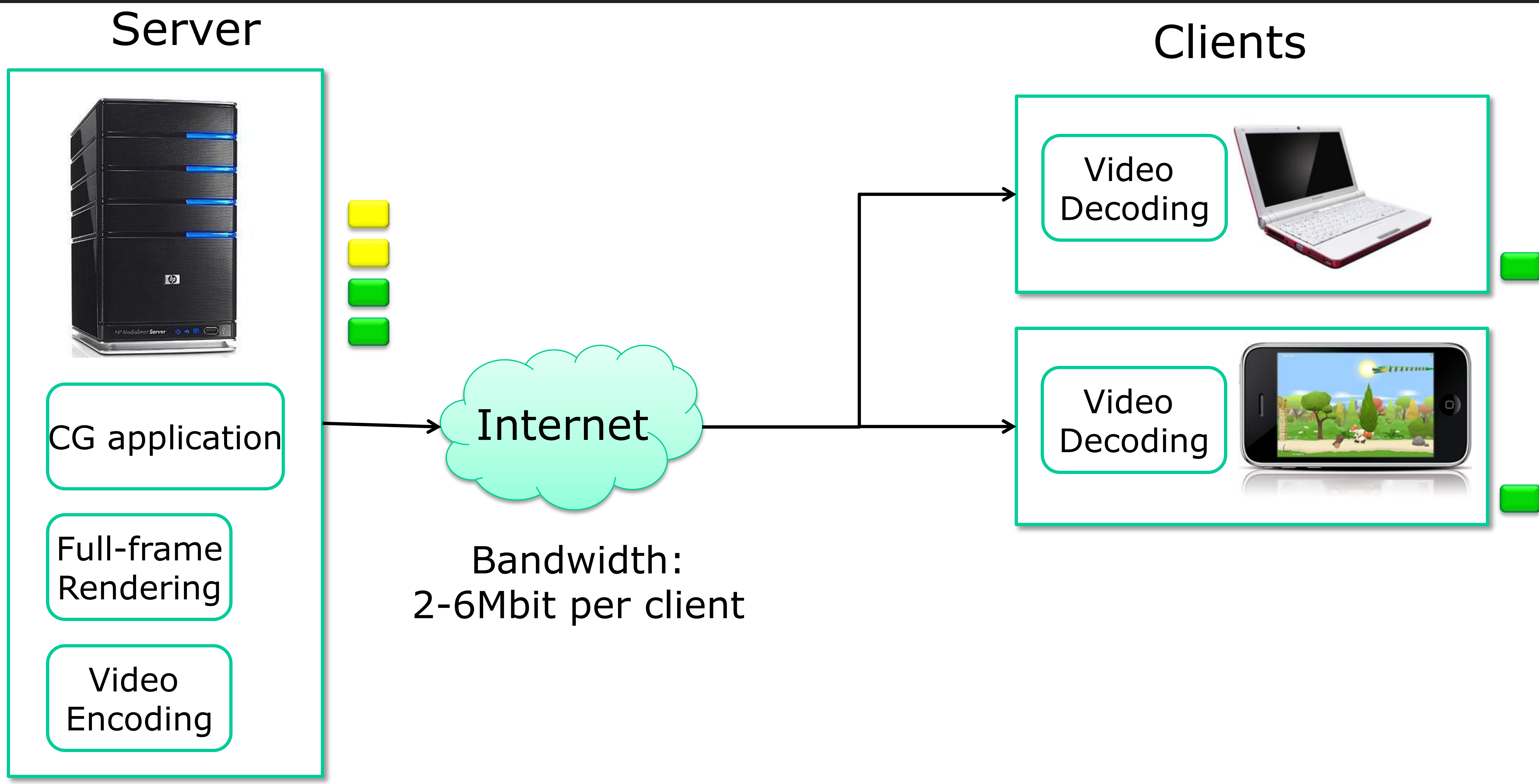
Streaming for Rendered Content

[Pajak et al. 11]



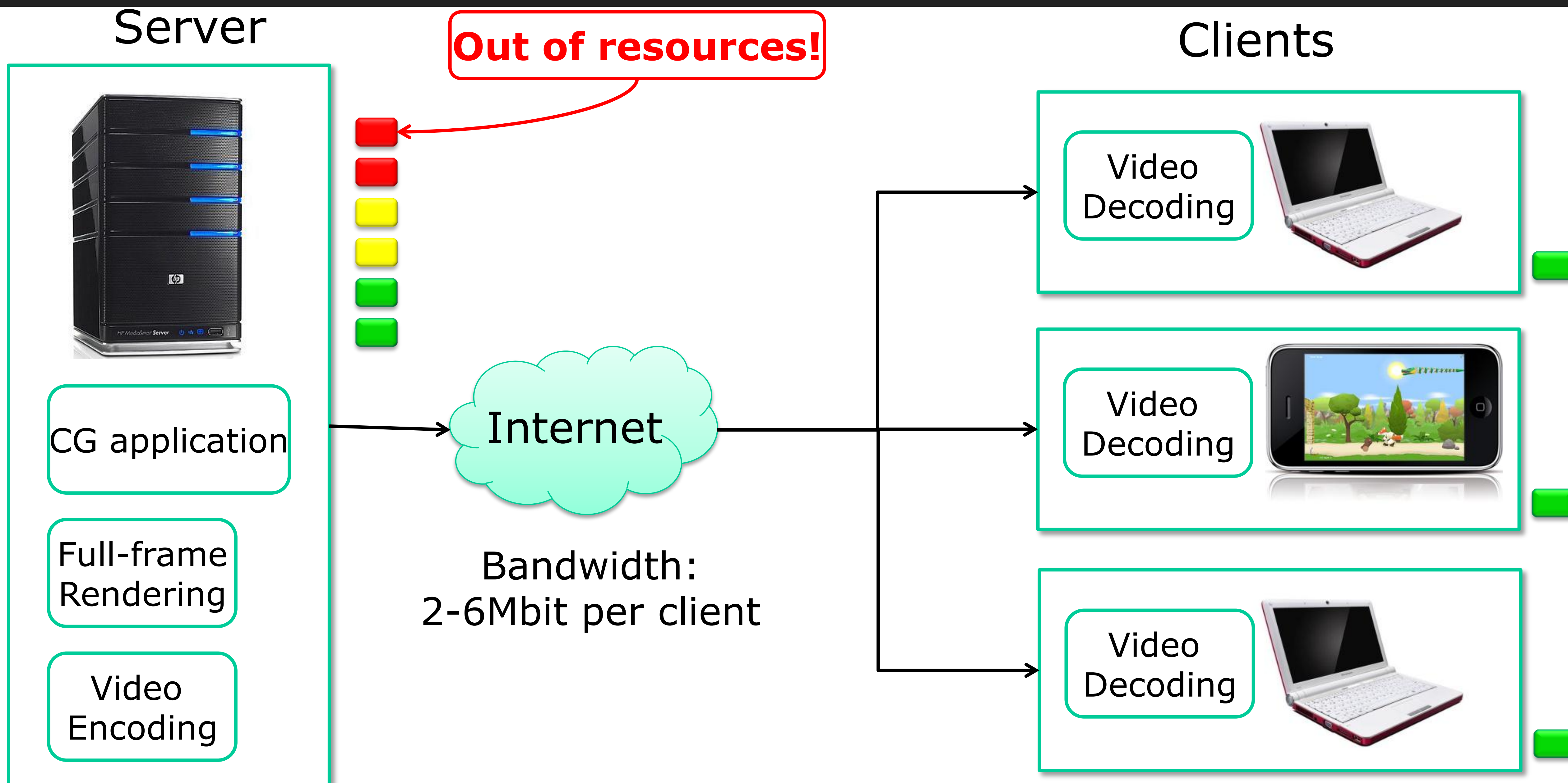
Streaming for Rendered Content

[Pajak et al. 11]



Streaming for Rendered Content

[Pajak et al. 11]



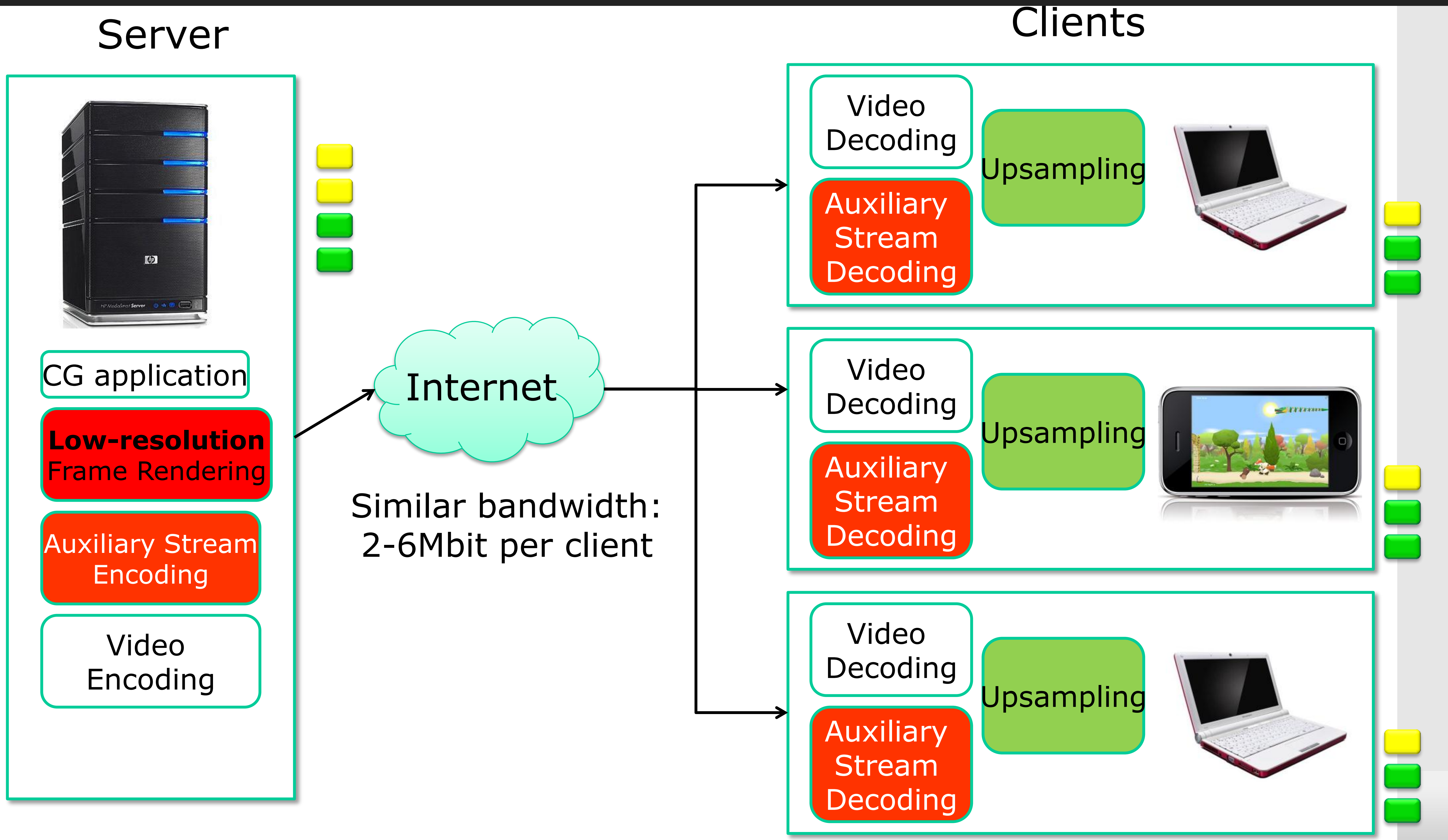
Design similar to current commercial solutions

Streaming for Rendered Content
[Pajak et al. 11]

- Rely on spatio-temporal upsampling strategies
  - Less bandwidth
  - Less server workload
- Specialized Encoding

Streaming for Rendered Content

[Pajak et al. 11]



Streaming for Rendered Content

[Pajak et al. 11]

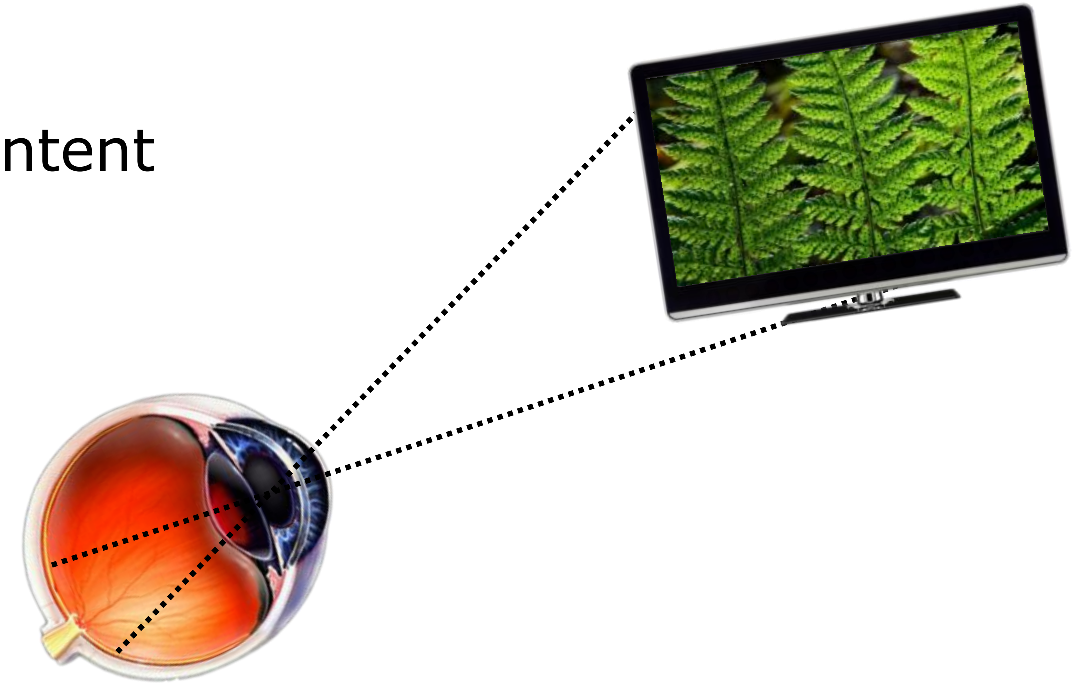
H264

Pajak et al. solution + more

Image-Space Coherence

- Very efficient
- Easy to implement
- Adapted to Graphics pipeline
- Important for streaming architectures

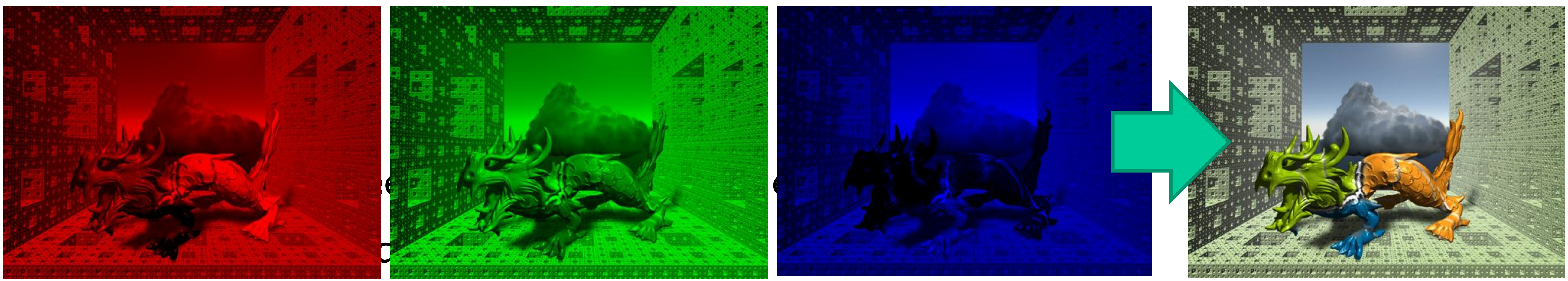
Exceed display limitations



- Idea: Temporal coherence to enrich content
  - Even beyond physical limits
- Examples:
  - Color bit depth: Frame Rate Control
  - Hold-type effect reduction: Temporal Upsampling
  - Resolution: Apparent Resolution Enhancement

Color Bit Depth: Frame Rate Control [Art04]

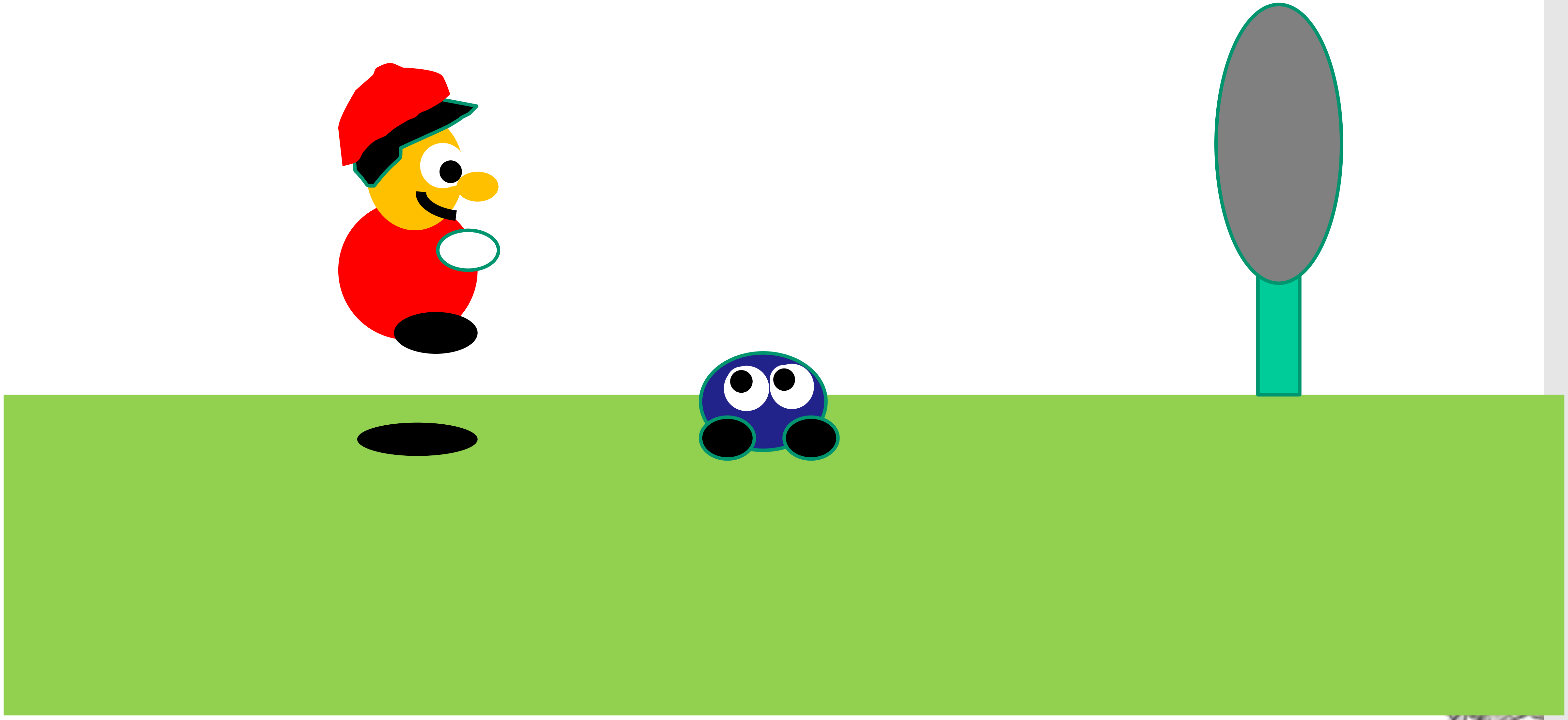
- Use eye latency to integrate color sequences
  - Similar principle as DLP projectors



-> Flicker different colors and have eye average them

Effect known from older video games

- Virtually augment the color palette



Flickering even works for >8 bit

- Fight mach banding artifacts
- Manually:
    - Switch last color bit
- Useful for HDR imagery,

  but very high refresh rates needed…

Display Improvement

1990's — 2000's — Today — Future

We are here

1990's

High refresh rate

more than 120Hz

Low brightness

Flicker for low rates

2000's

No flickering

Higher level of luminance

Low refresh rate - ~60Hz

Long response time

Today

Brighter

Better contrast

Low response time

High refresh rate

Exploit HVS to improve quality

Future

Small response time

Higher refresh rate

Better colors,

Better contrast

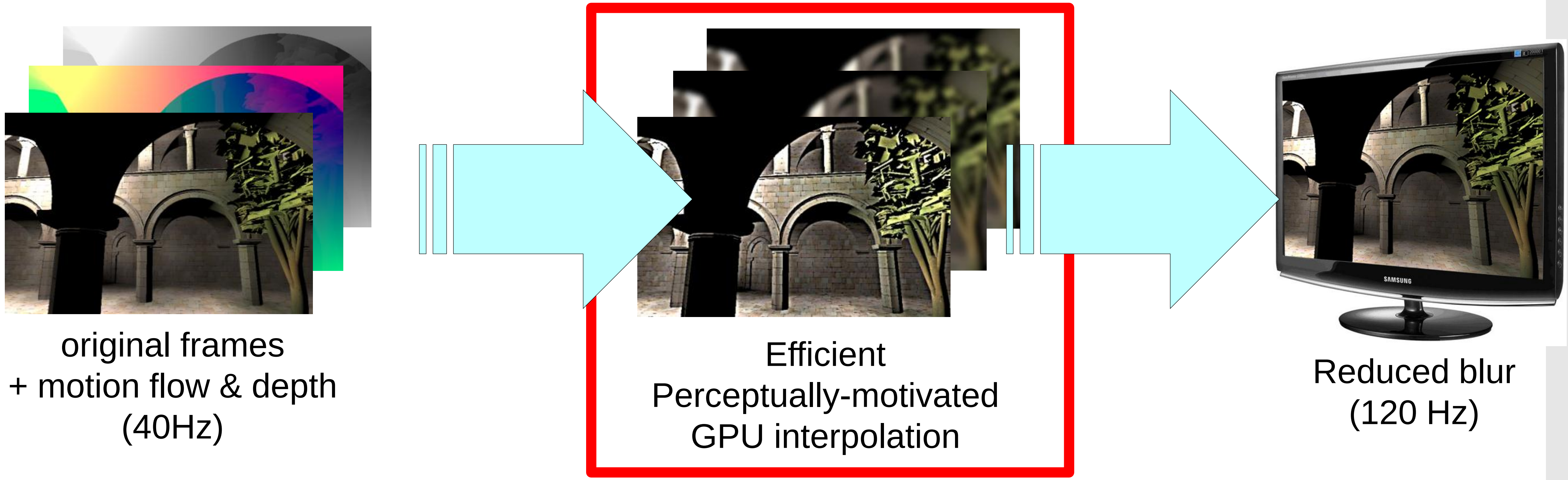
Better brightness

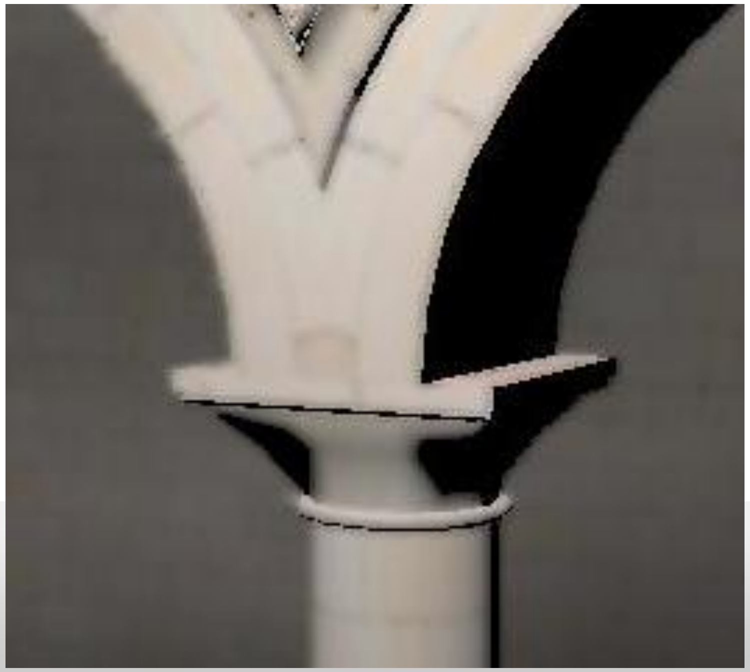
…

Less expensive ;)

Hold-Type Blur Reduction [Didyk10]

- Exploit limitations of the HVS





High-Frequency propagation

- High-frequency information

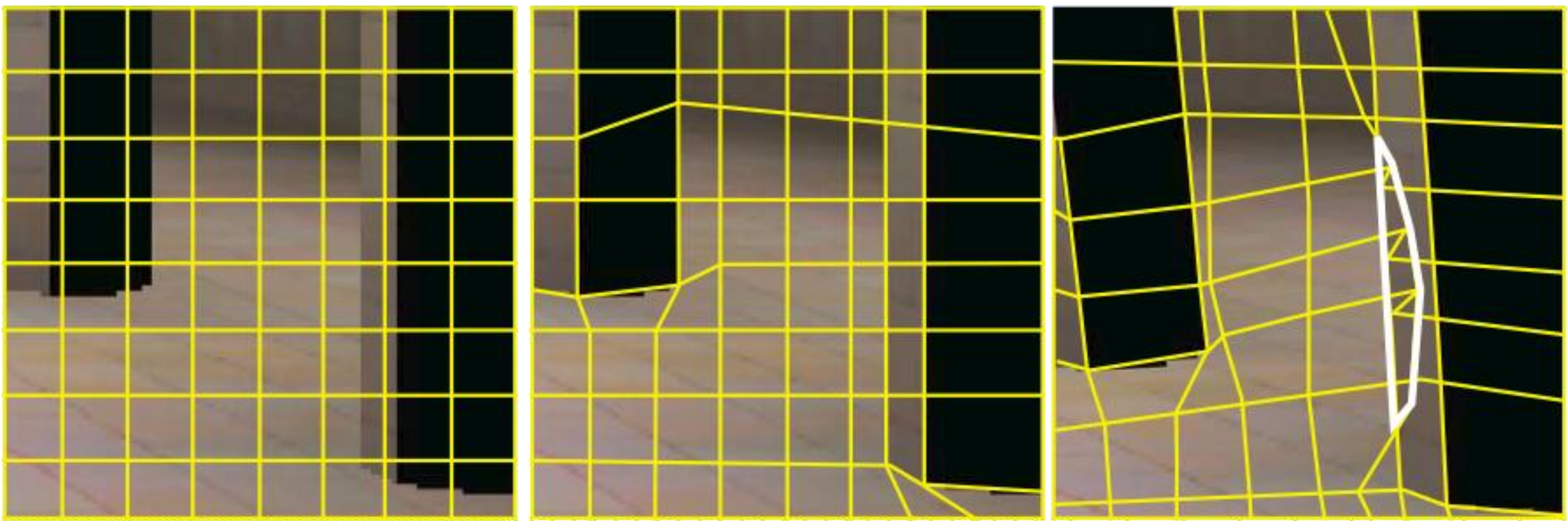
  is spread across time at 120Hz

-> Idea: Increase high-frequency in first frame

hide artifacts in extrapolation via blur

Use a cheap extrapolation technique

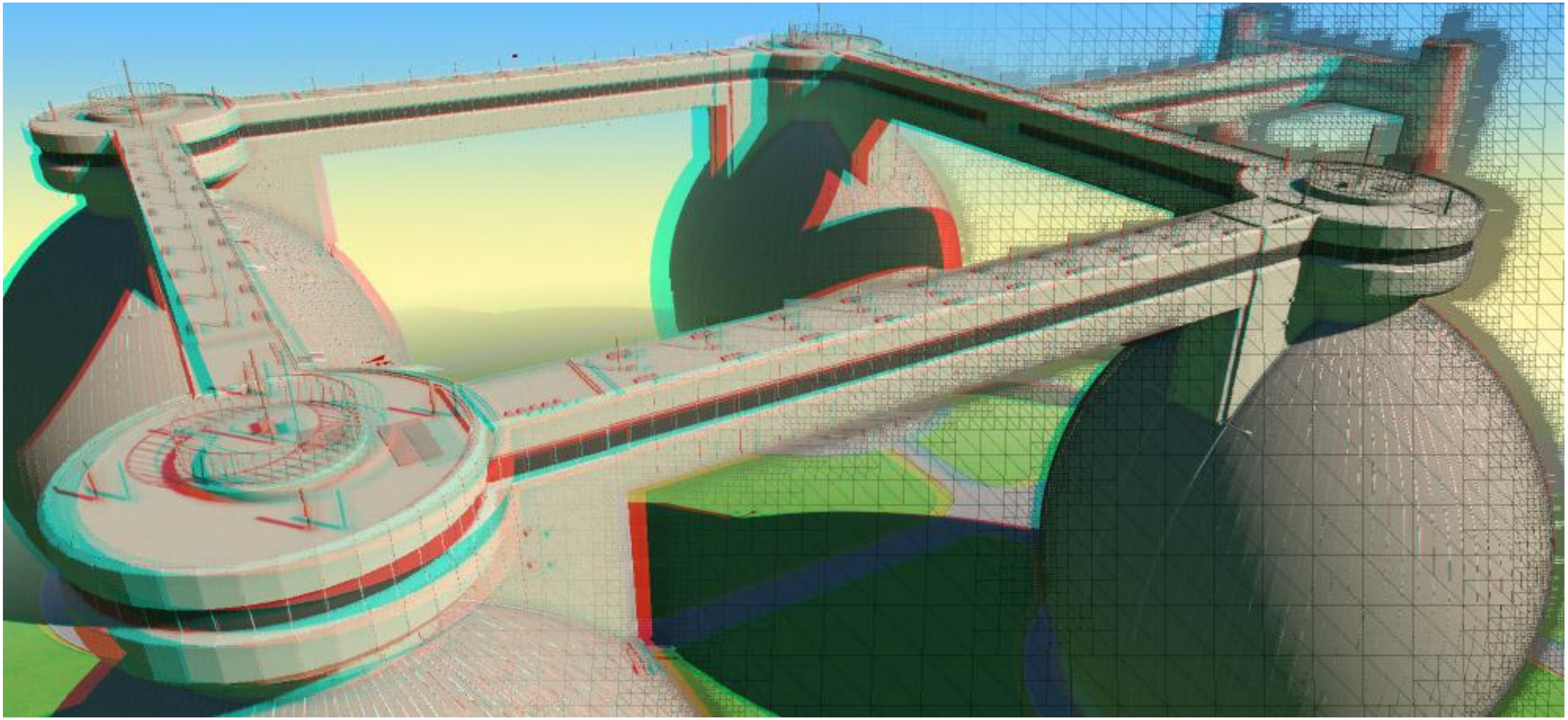
Regular Grid Snapped Morphed



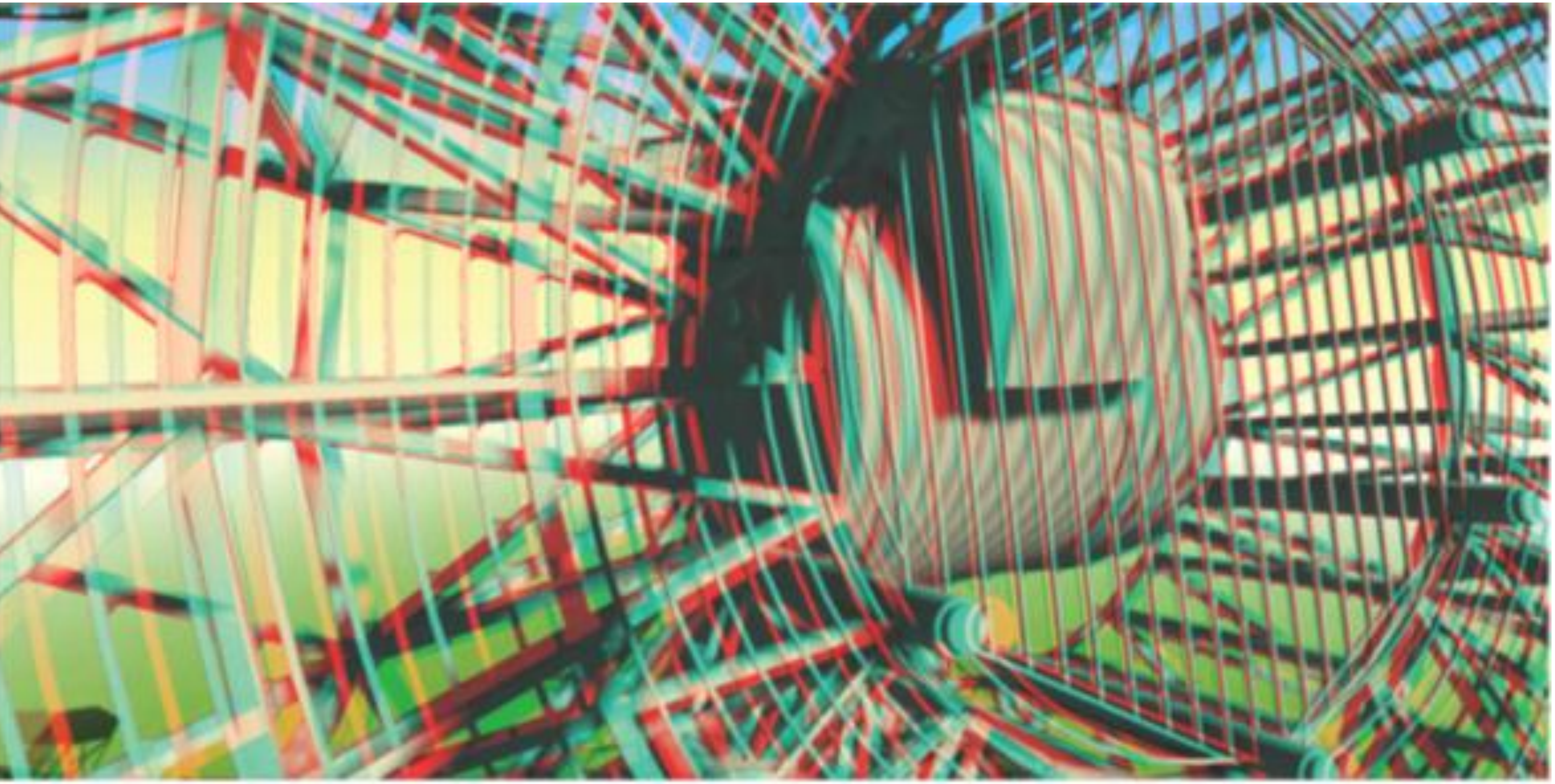
- Artifacts will be hidden by blur

Extension to Stereo

- Adaptive Image-space Stereo View Synthesis [Didyk et al. VMV`10]
- More sophisticated (adaptive) warping

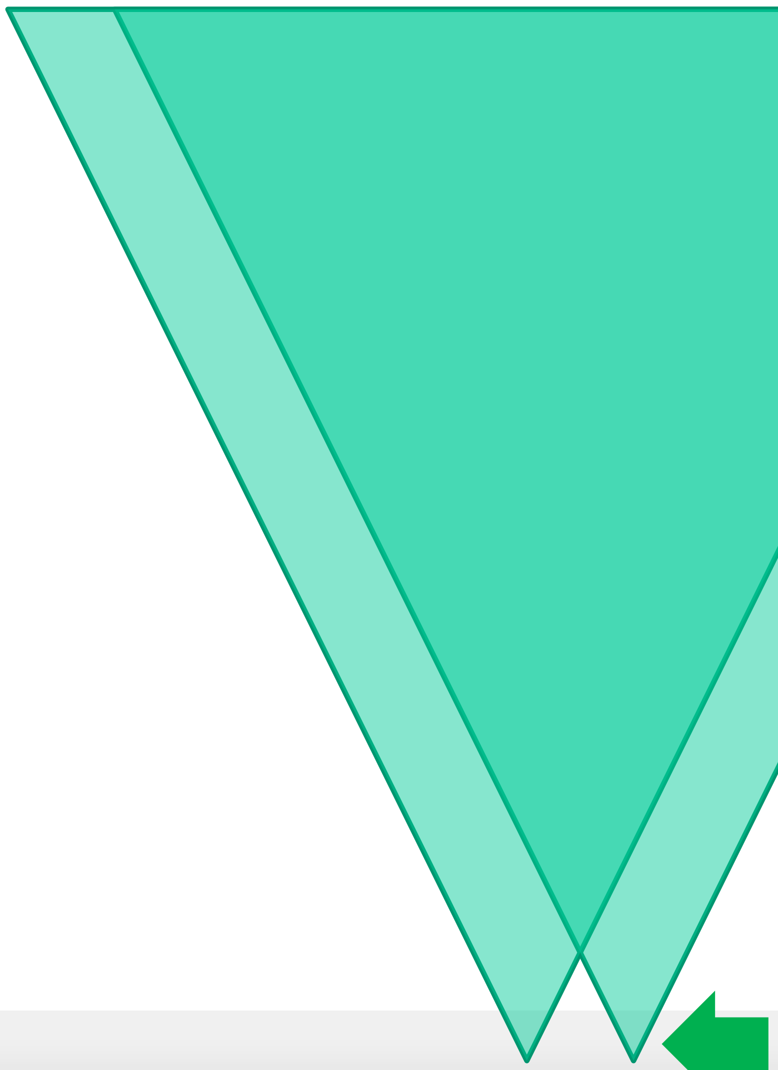


Extension to Stereo - Results



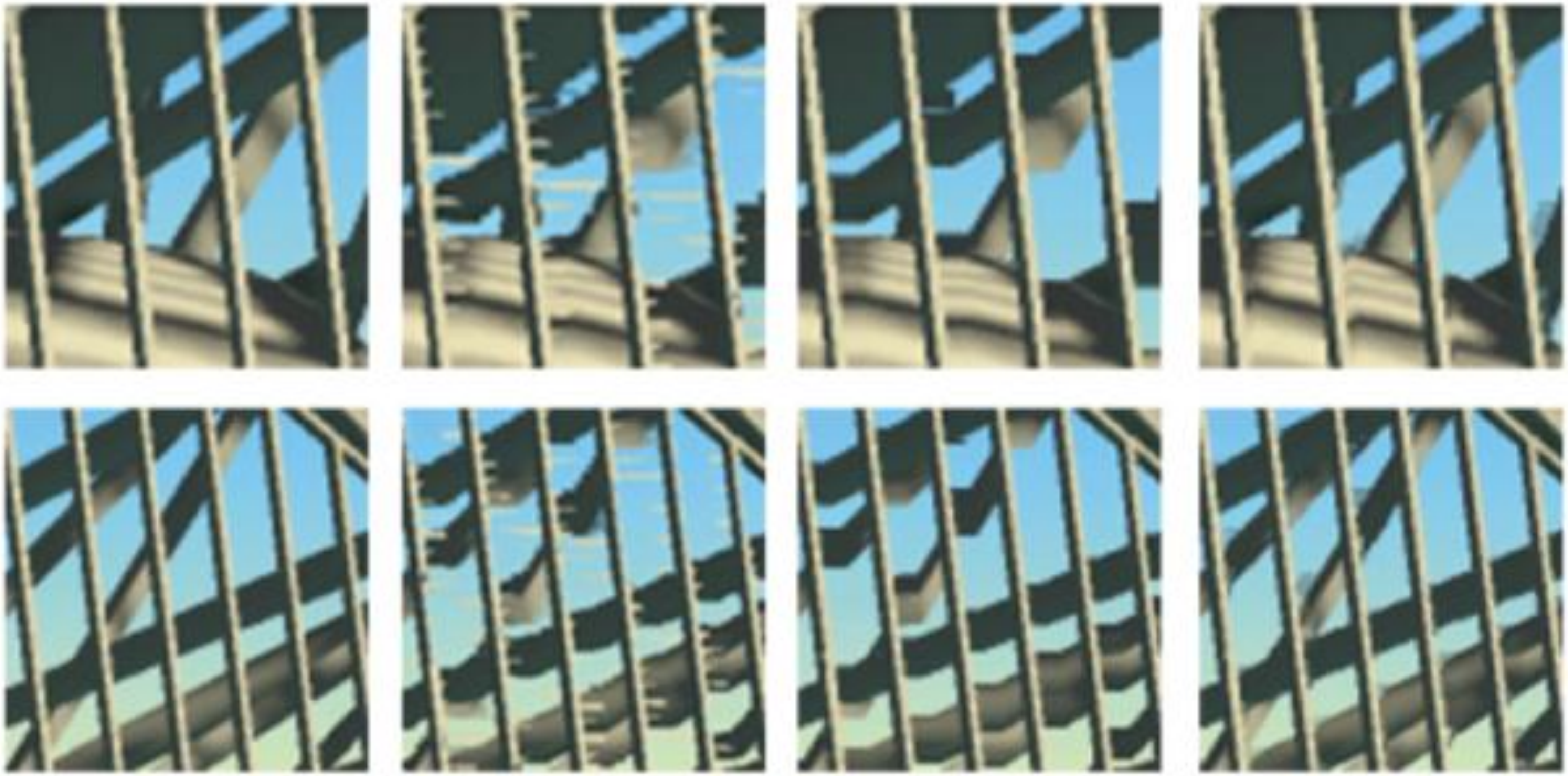
Extension to Stereo

- Temporal coherence of viewpoint
  - Reuse nearby view from previous frame
  - Only render one new view and rely on warping



Viewpoint at time t+1

Extension to Stereo - Results



Reference Previous work Warping Temp. Warping

Warping

- Very cheap alternative to complex methods
- Maps very well to GPU
- Executes in less than 4ms on a full-HD frame
  - NVIDIA GT 460
- Two applications, others exist
  - Hold-type blur reduction and Stereo

Combating Hold-type Blur [DER*10]

- Many advantages:
  - Crispness
  - Quality
  - Task-performance

  - Low overall cost

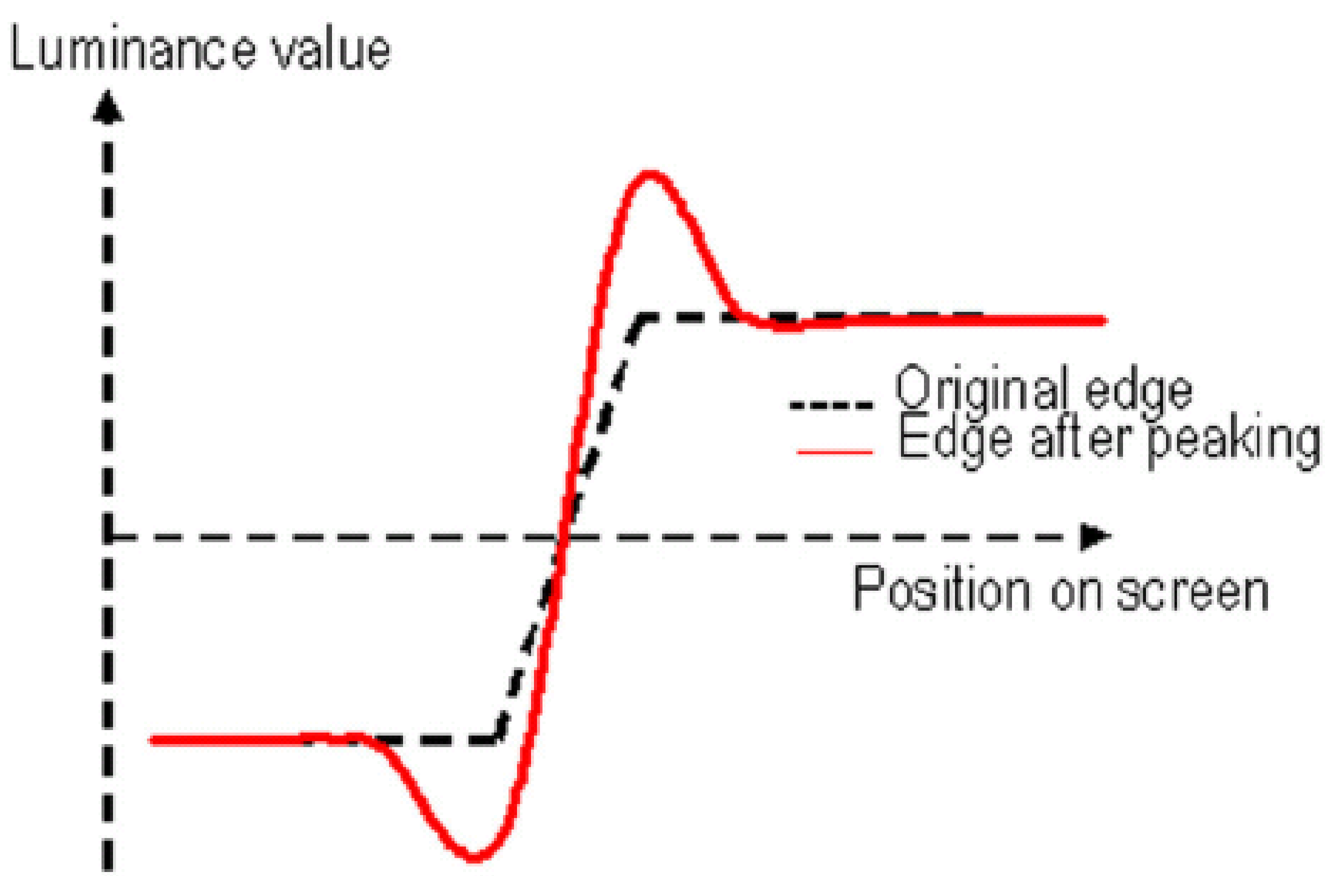
Can we push blur reduction even further?

Super-resolution

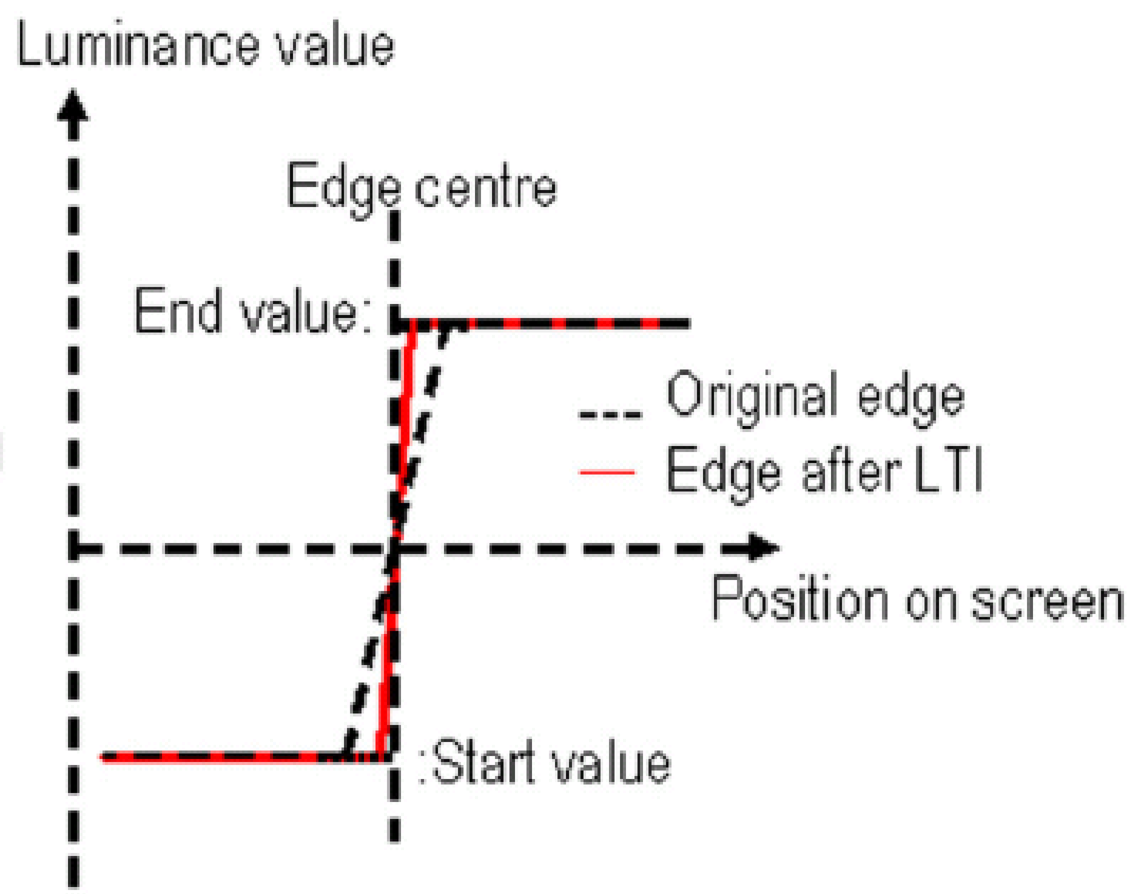
- Upscaling, solved problem, ICs at all PC
  - Does not add new frequencies
- Super-resolution goal: restore high frequencies
  - De-interlacing: images show alias
  - In graphics it is easy to get aliasing
- Typical sharpening algorithms used in TV sets
  - Peaking
  - Luminance Transient Improvement (LTI)
- Temporal domain can also be exploited

Sharpening Filters

Peaking



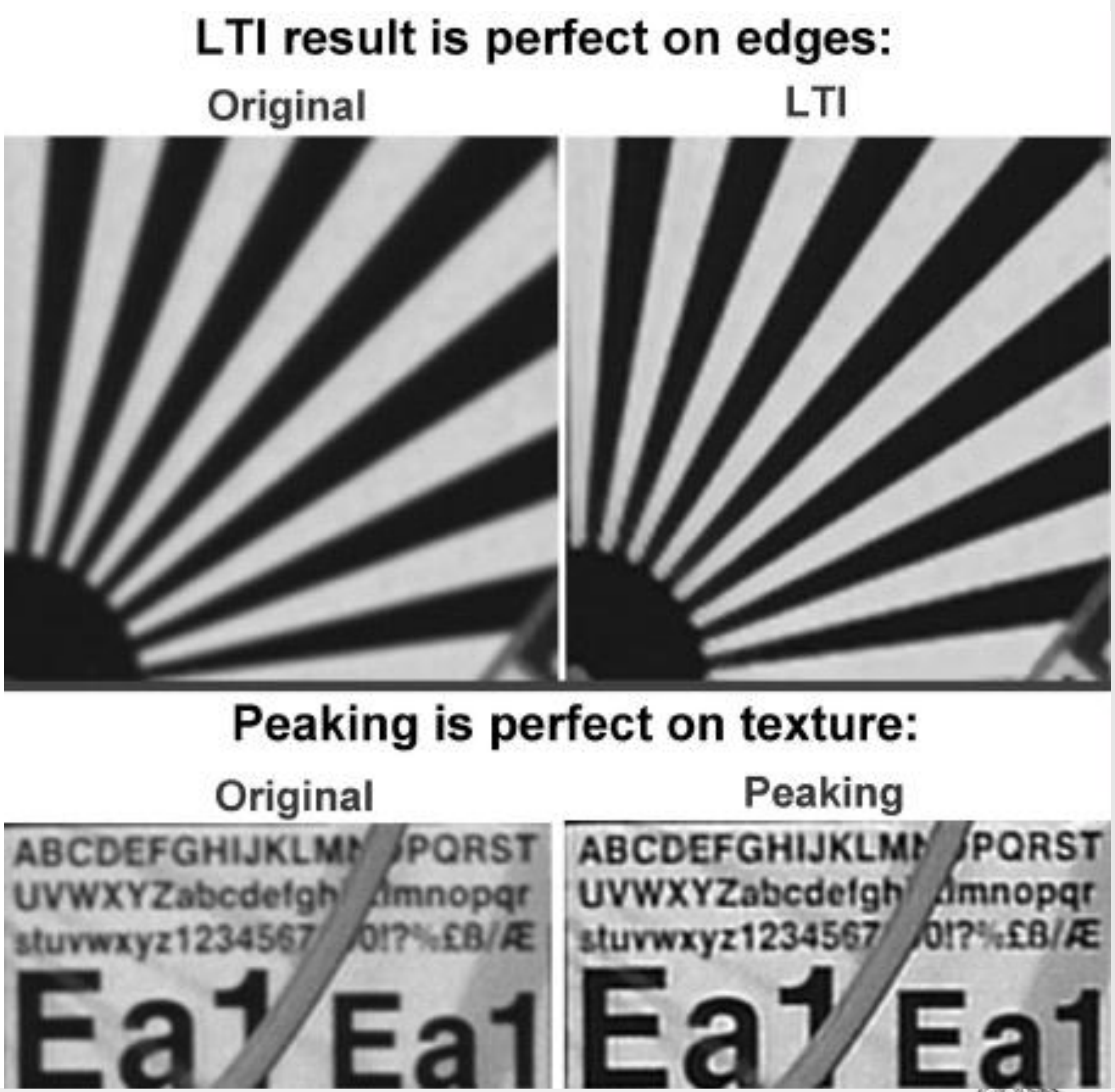
Luminance Transient Improvement (LTI)



M. Zhao, M. Bosma, G. de Haan, Making the best of legacy video on modern displays, 2006 Society for Information Display

Sharpening Filters: Results

- Peaking similar to unsharp masking
- In 3D rendering enhancement of noise signal is not a problem
- In 3D rendering we can better detect object silhouettes
- LTI ~ velocity



M. Zhao, M. Bosma, G. de Haan, Making the best of legacy video on modern displays, 2006 Society for Information Display

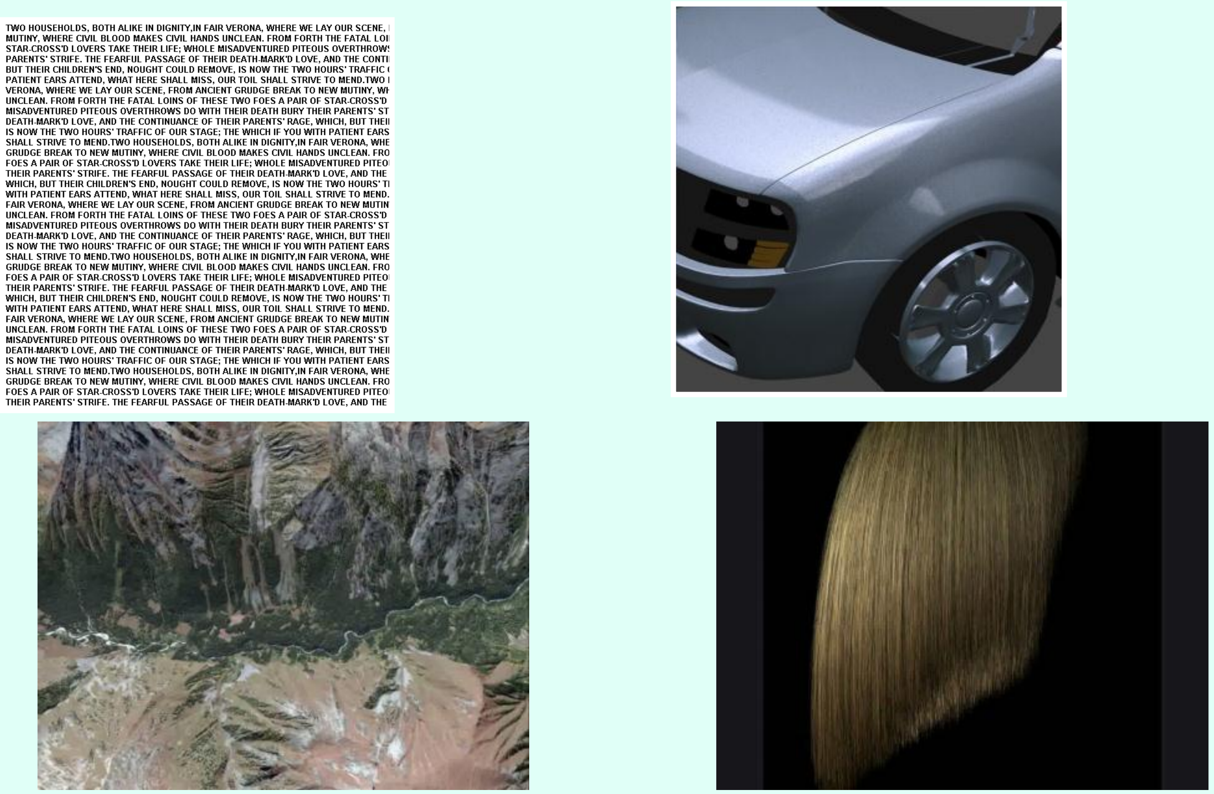
Many High-Resolution Sources

Photographs: > 10MPix

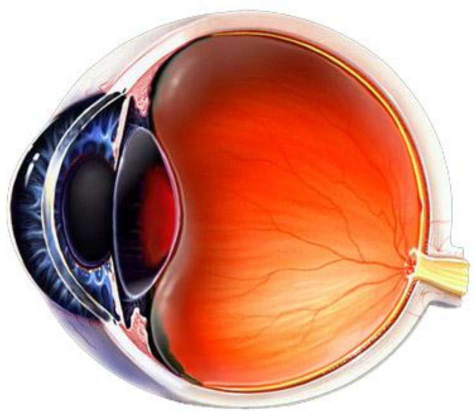
Panoramas: > 50MPix

Gigapixel Photography:

Computer generated: Unlimited



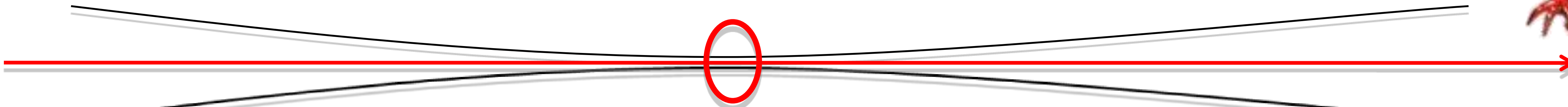
Motivation

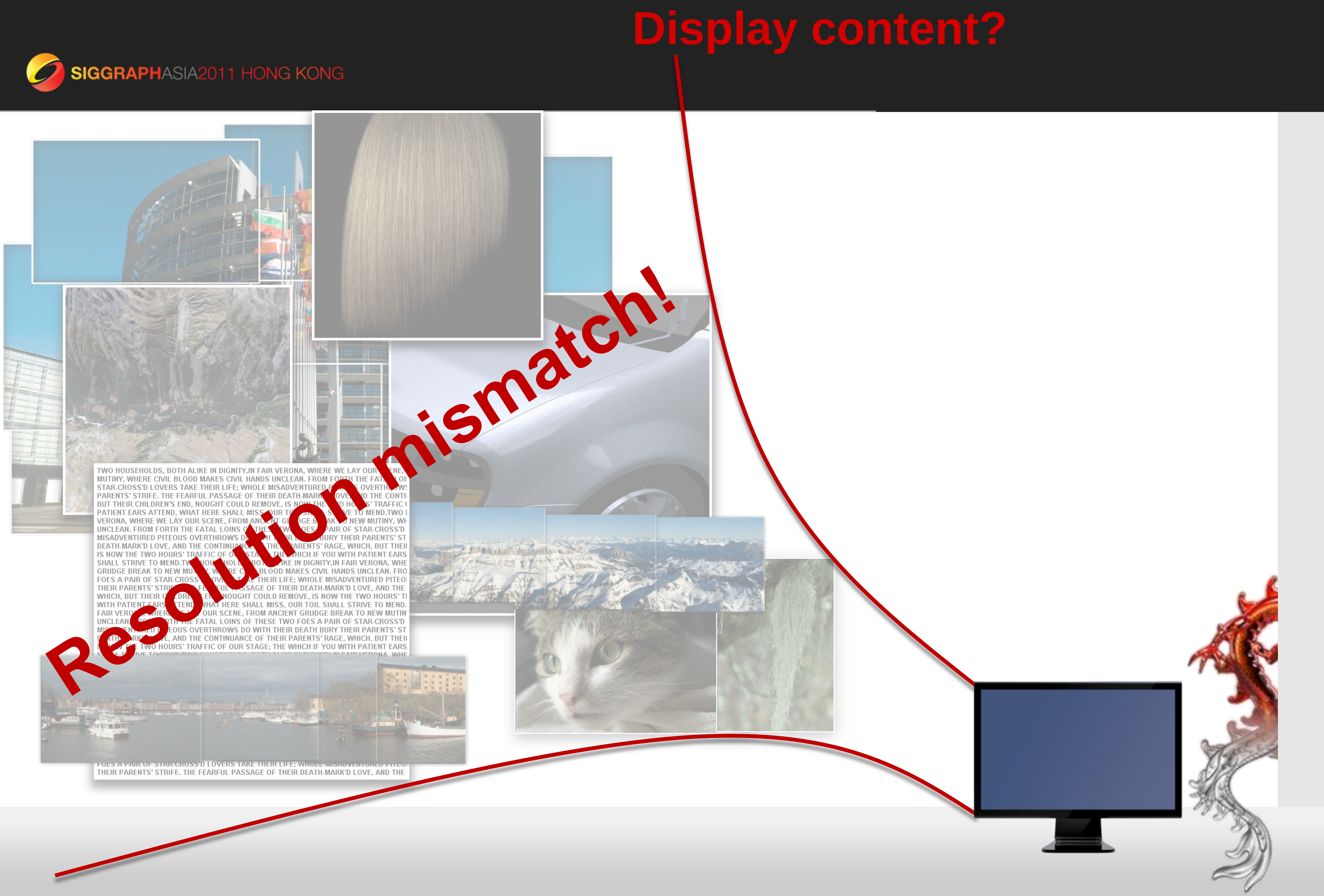


easily ~50 MPix

~ 2-8 MPix

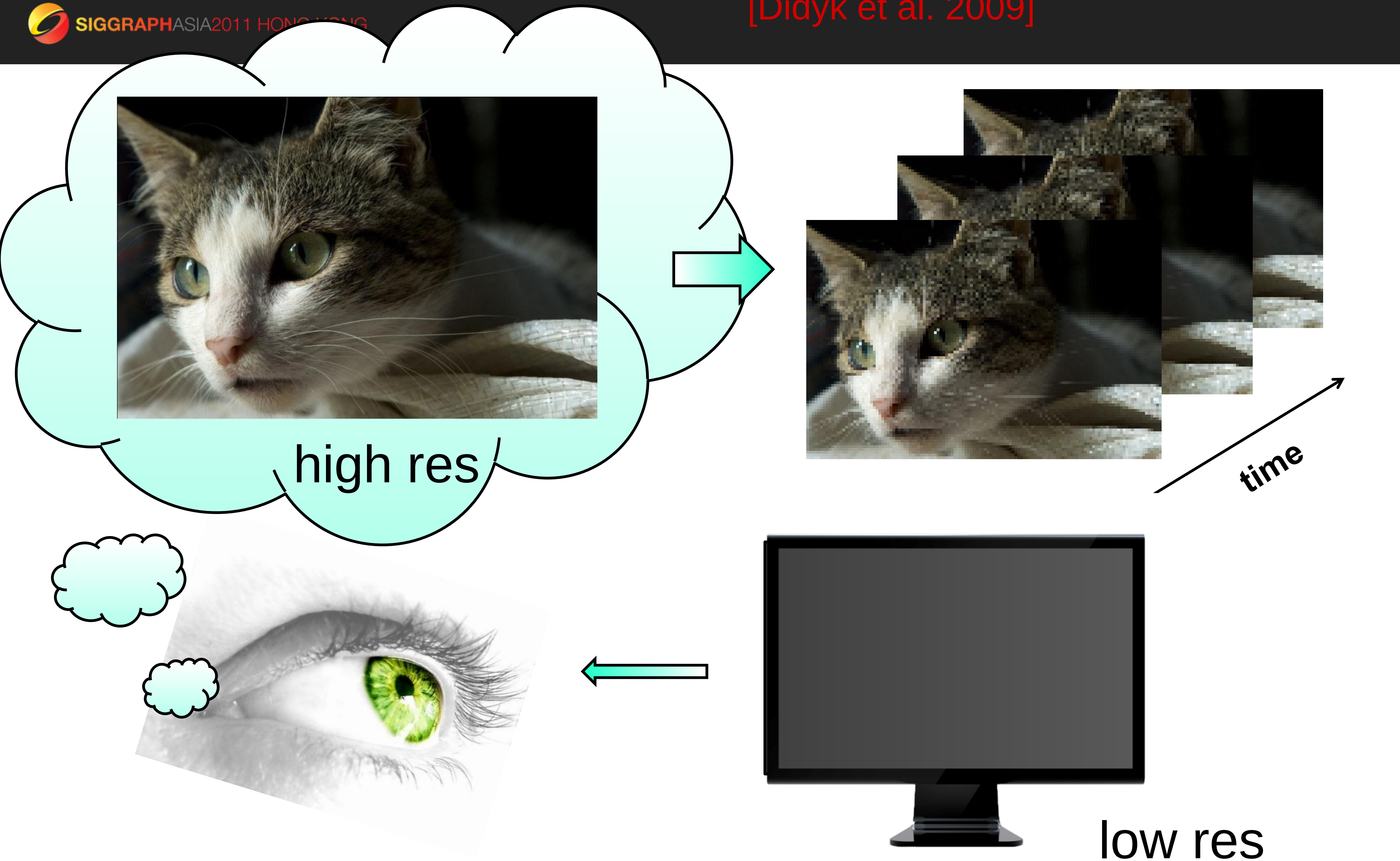
1px → > 9 receptors




Display content?
Resolution mismatch!

Apparent Resolution Enhancement

[Didyk et al. 2009]



increased apparent resolution

Perception: Spatial Visual Acuity

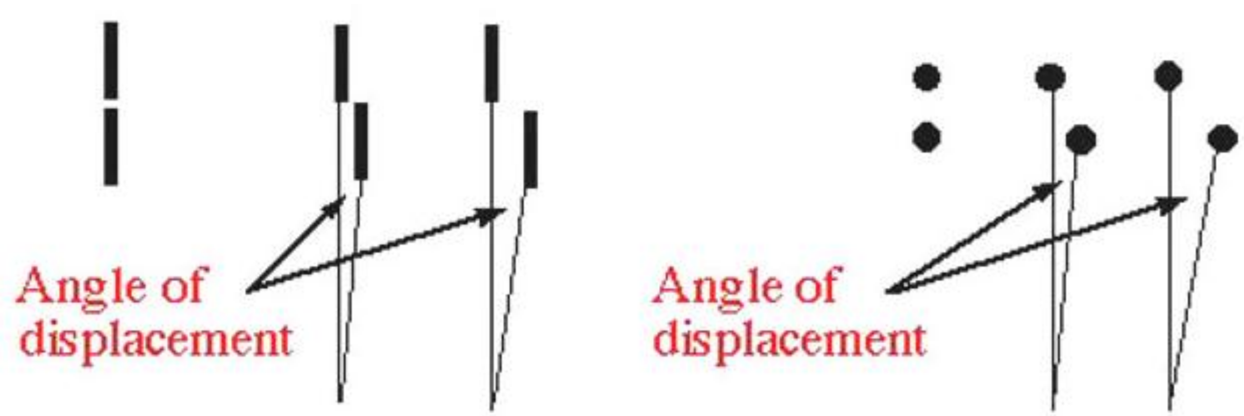
- Cone density in the fovea may reach 28″ (arc seconds)
  - Nyquist's theorem: then 1D sine gratings of 60cycles/deg can be resolved
  - Low-pass filtering in the eye optics removes higher frequencies causing aliasing
- Pixel size at a full-HD desktop display observed from 50cm distance spans 1.5′ (arc minutes)
  - In such observation conditions <u>1 pixel covers roughly 9 cones</u>
  - Estimation valid only for the central fovea region
- Visual *hyperacuity* enables to locate slightly shifted lines in the Vernier acuity task with precision higher than 5″ (arc seconds)
  - This more a *localization* task than a *resolution* task

SIGGRAPHASIA2011 HONG KONG

Perception: Spatial Visual Acuity

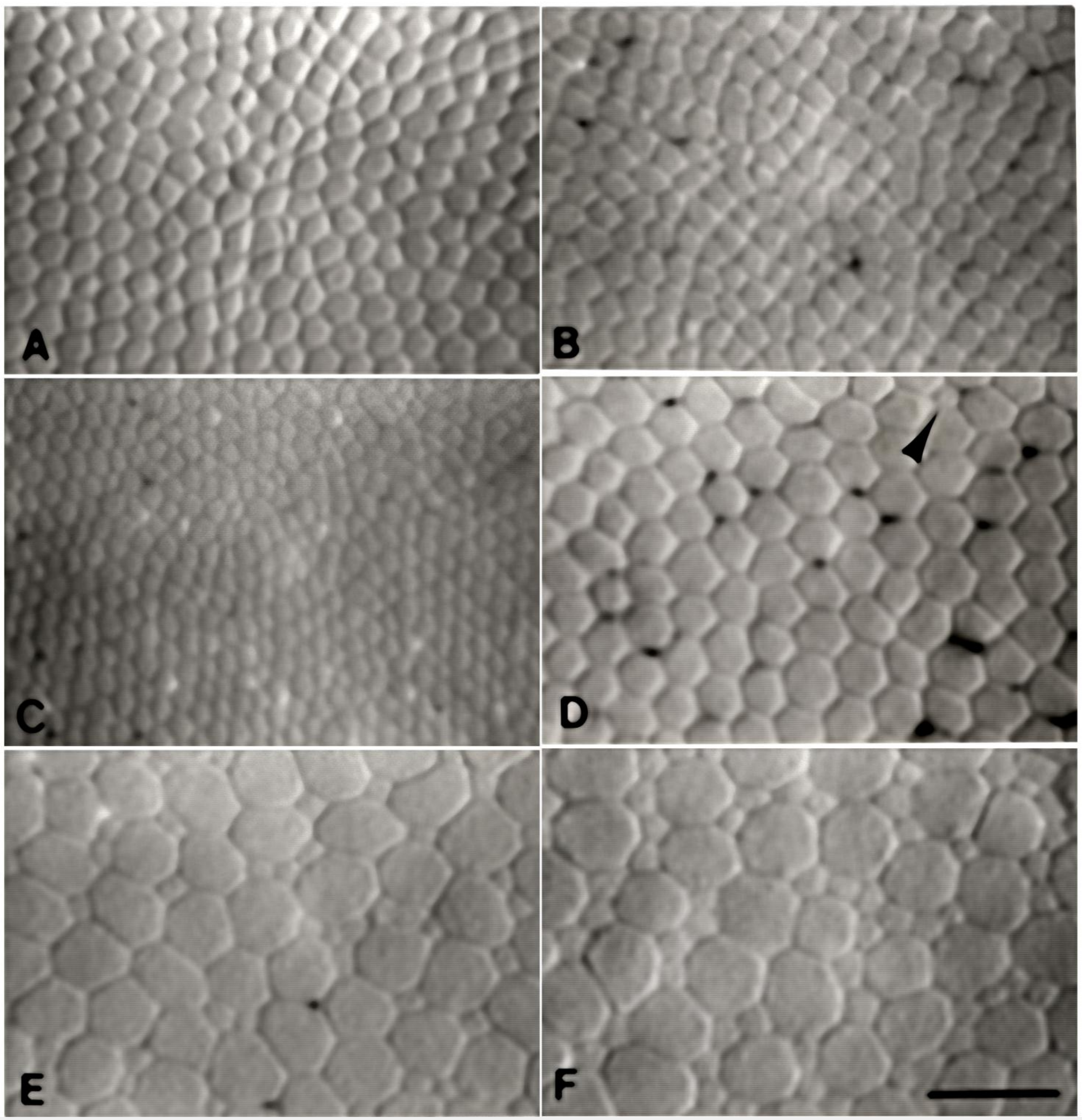
- Target *resolution* threshold: the smallest angular size at which subjects can discriminate

- Target *localization* threshold: the smallest difference in position which subjects can discriminate (Vernier hyperacuity)



http://webvision.med.utah.edu/book/part-viii-gabac-receptors/visual-acuity/

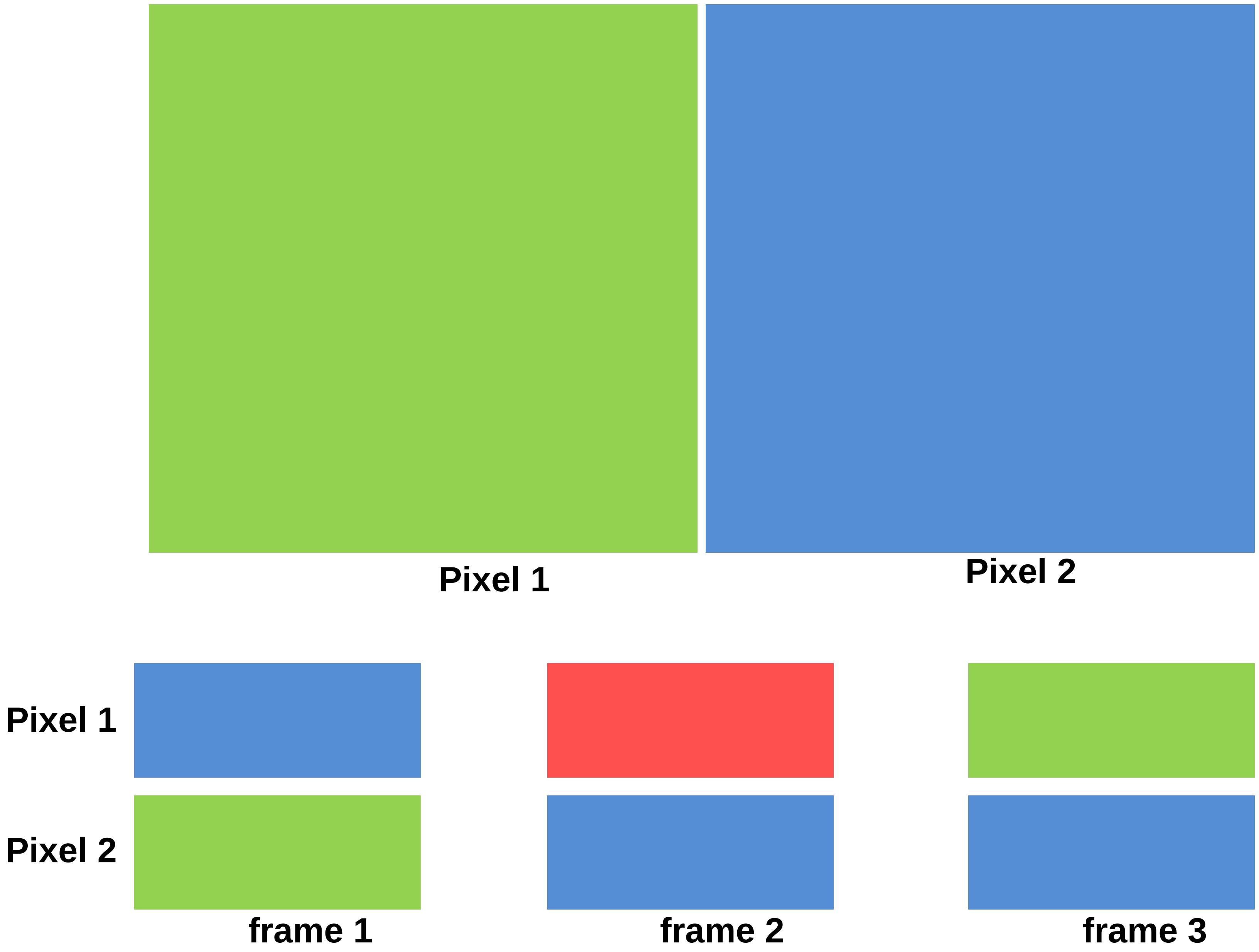
Foveal Photoreceptor Mosaic



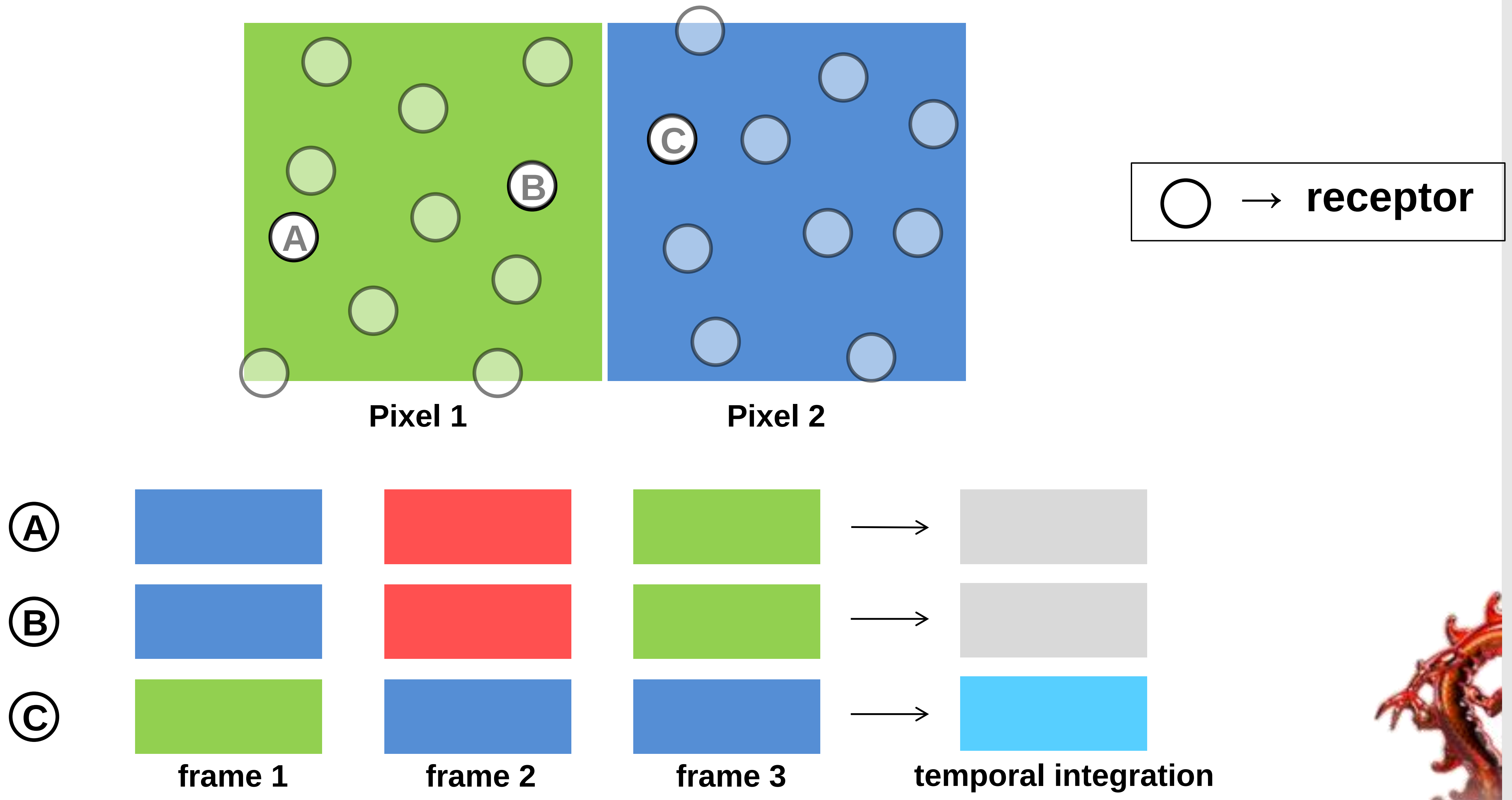
A-C fovea center - cones only
D rod-free region boundary,
the arrow shows rod
E cones-rods balanced
F rods outnumber cones

Curcio, C. A., Sloan, K. R., Kalina, R. E., Hendrickson, A. E., 1990. Human photoreceptor topography. J Comp Neurol 292, 497-523

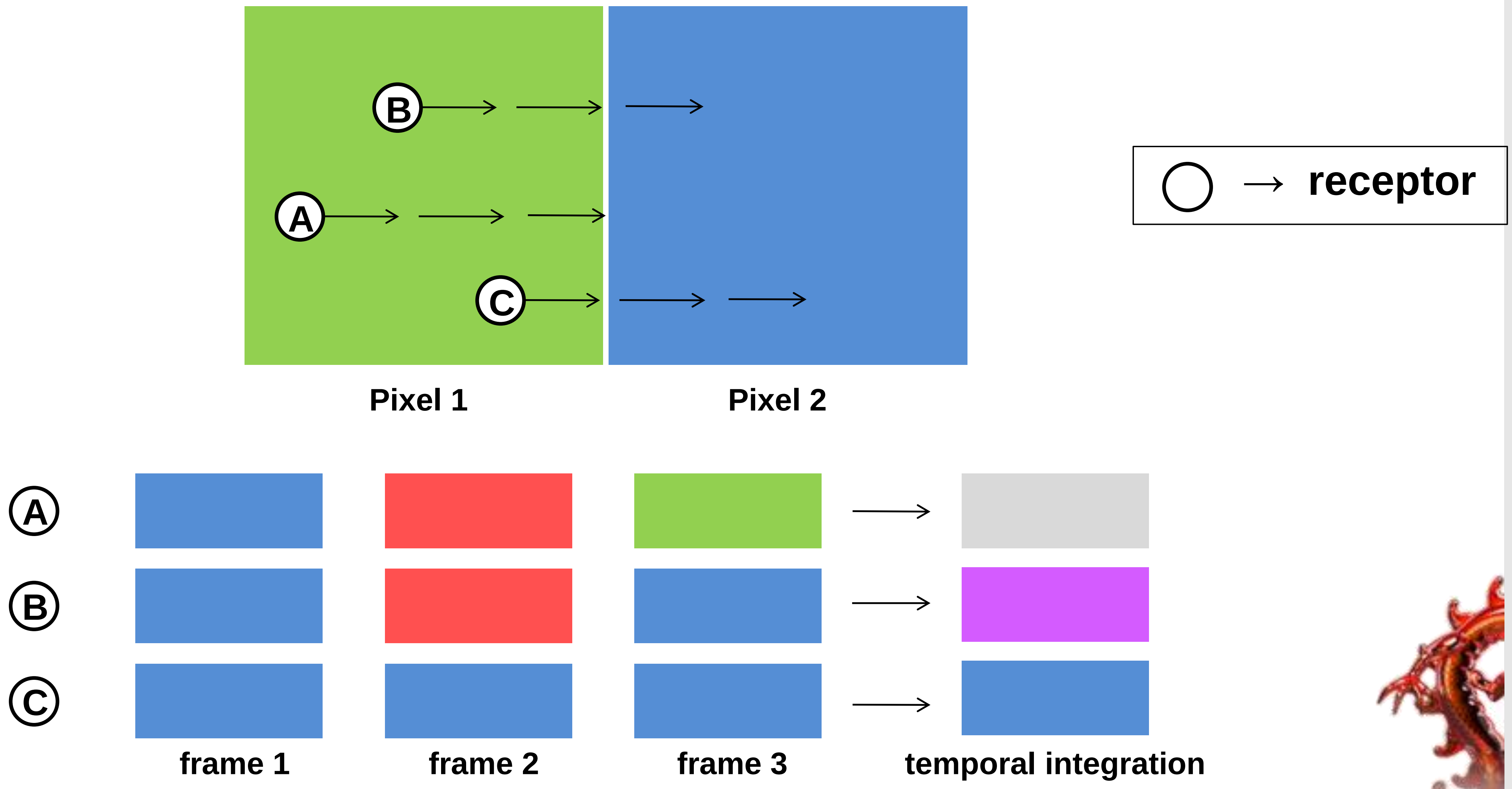
Temporal Domain



Temporal Domain – static case

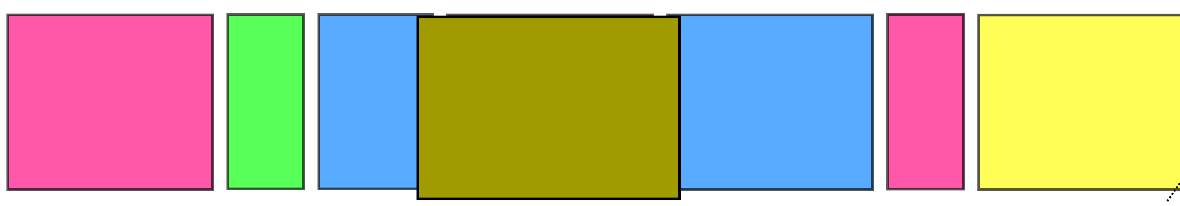


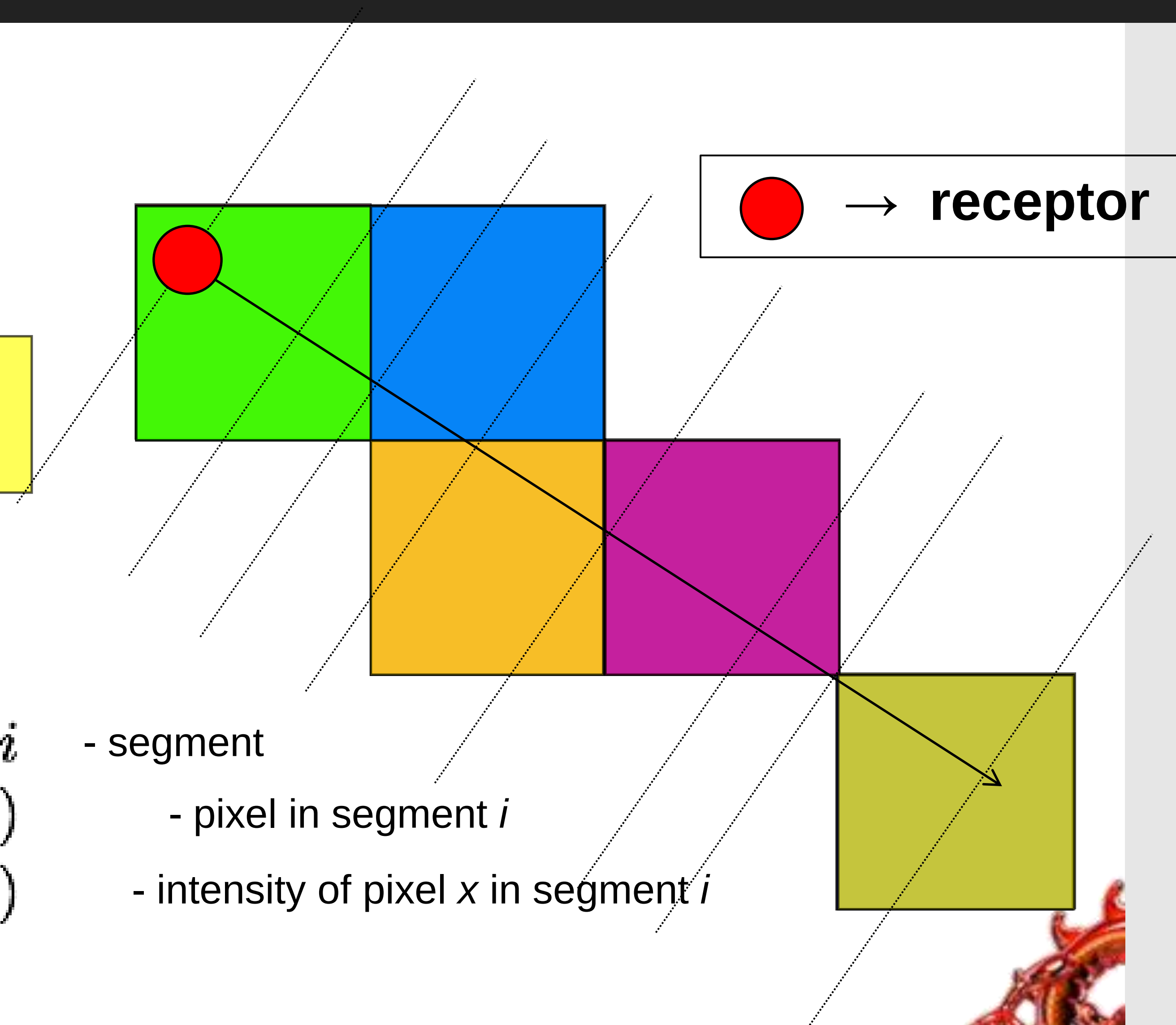
Temporal Domain – dynamic case



Temporal Integration Model

Receptor signal:





$$\sum_{i=0}^{N} w_i \, I(p(i), i)$$

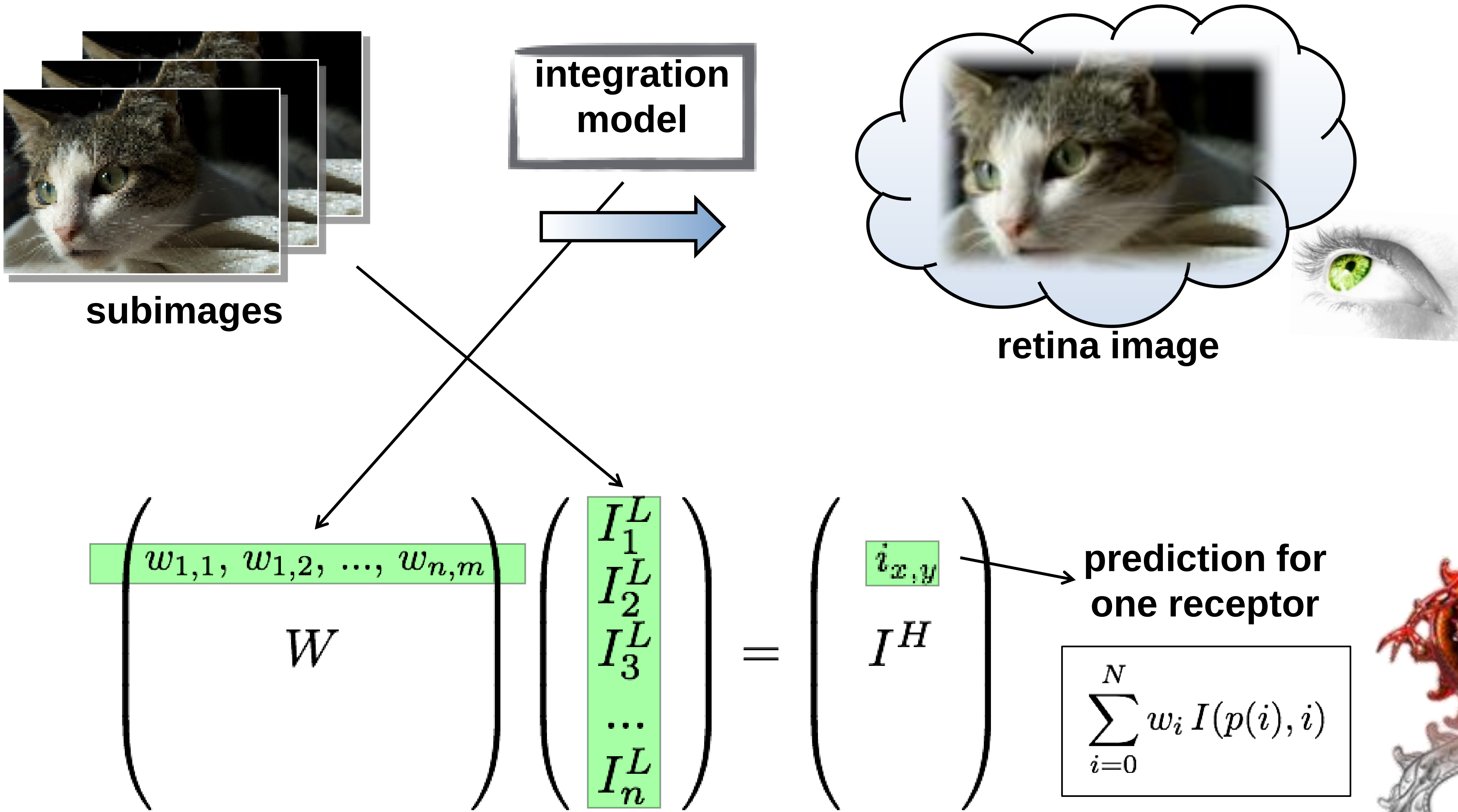
i - segment

$p(i)$ - pixel in segment *i*

$I(x, i)$ - intensity of pixel *x* in segment *i*

w_i - weights proportional to the length of the segment

Prediction in Equations

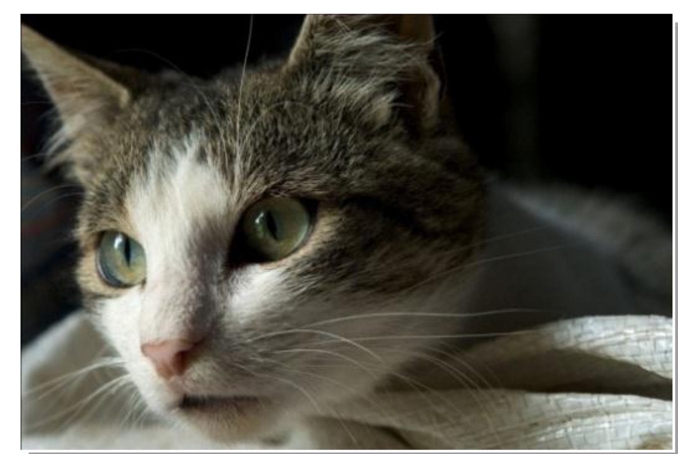


Optimization Problem

$$\begin{pmatrix} & & \\ & W & \\ & & \end{pmatrix} \begin{pmatrix} I_1^L \\ I_2^L \\ I_3^L \\ \cdots \\ I_n^L \end{pmatrix} = \begin{pmatrix} \\ I^H \\ \\ \end{pmatrix}$$

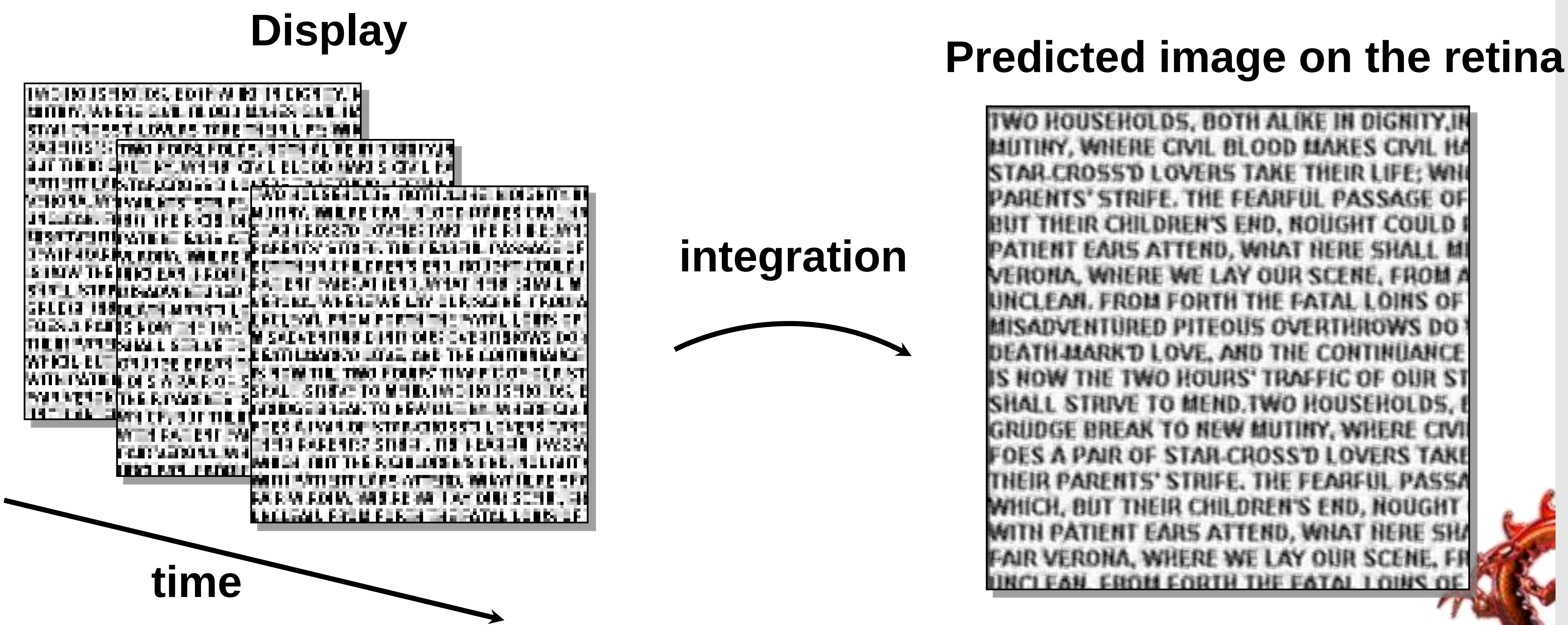
integration model

subimages



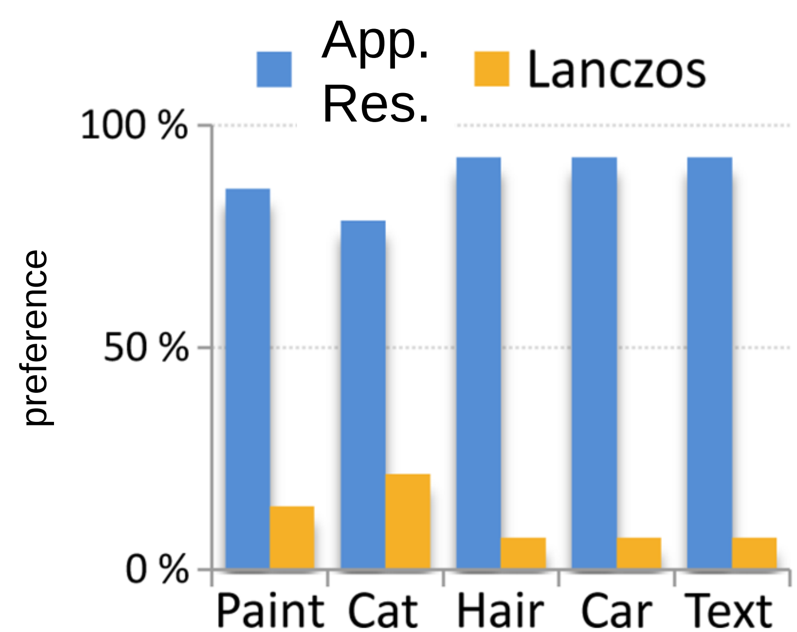
high resolution image

Optimization Result



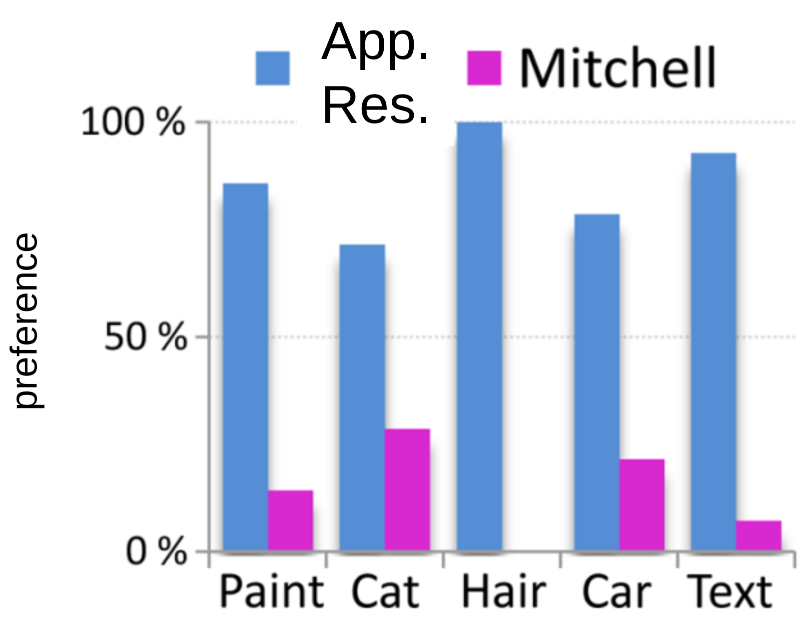
ARE vs. Lanczos

- compare each frame to moving image
  - downsample separately hence, slightly different information over time

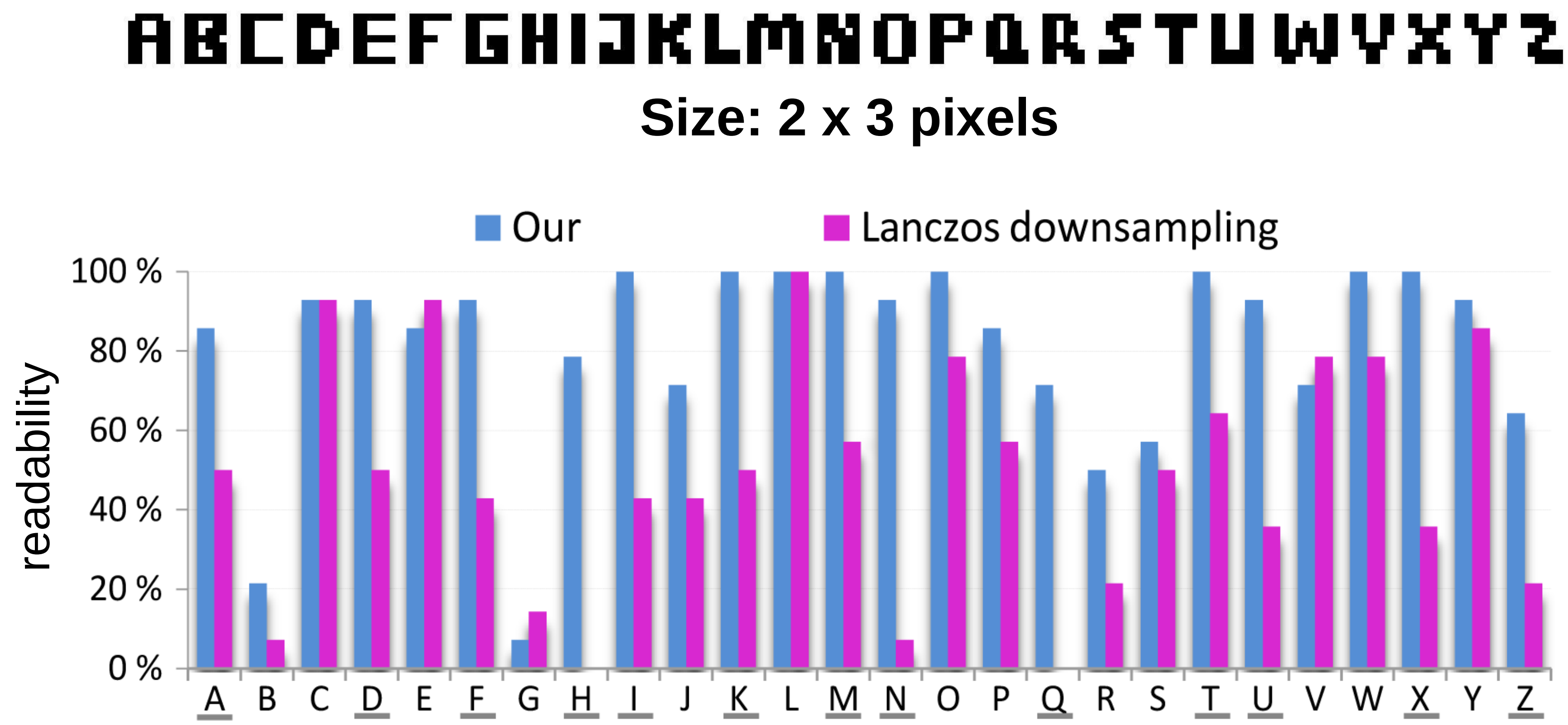


ARE vs. Mitchell

- Mitchell downsampling
  - participants adjusted parameters to match high resolution image



ARE - Alphabet



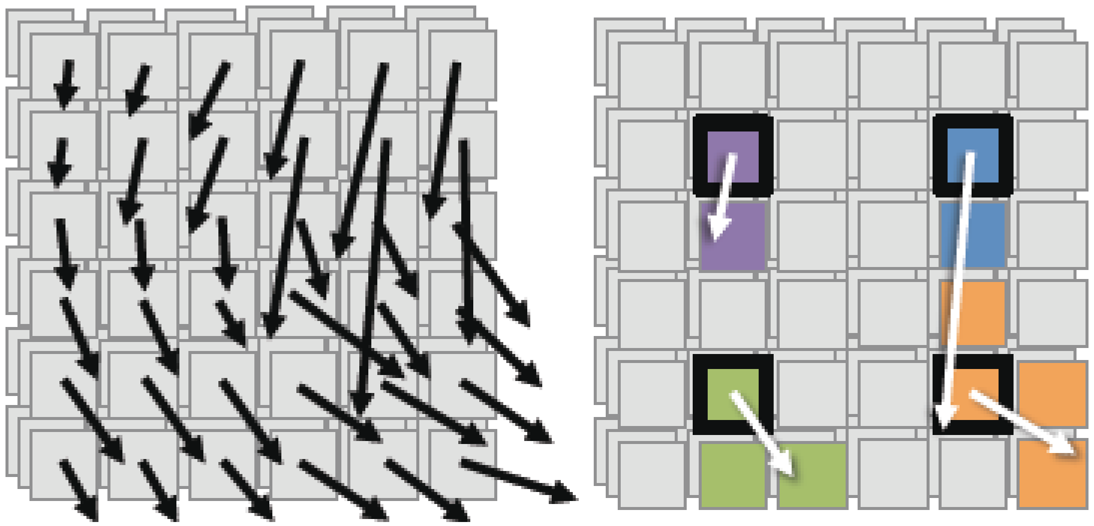
- Applications:
  - scrolling text or maps on low resolution devices
  - stock tickers, news headlines

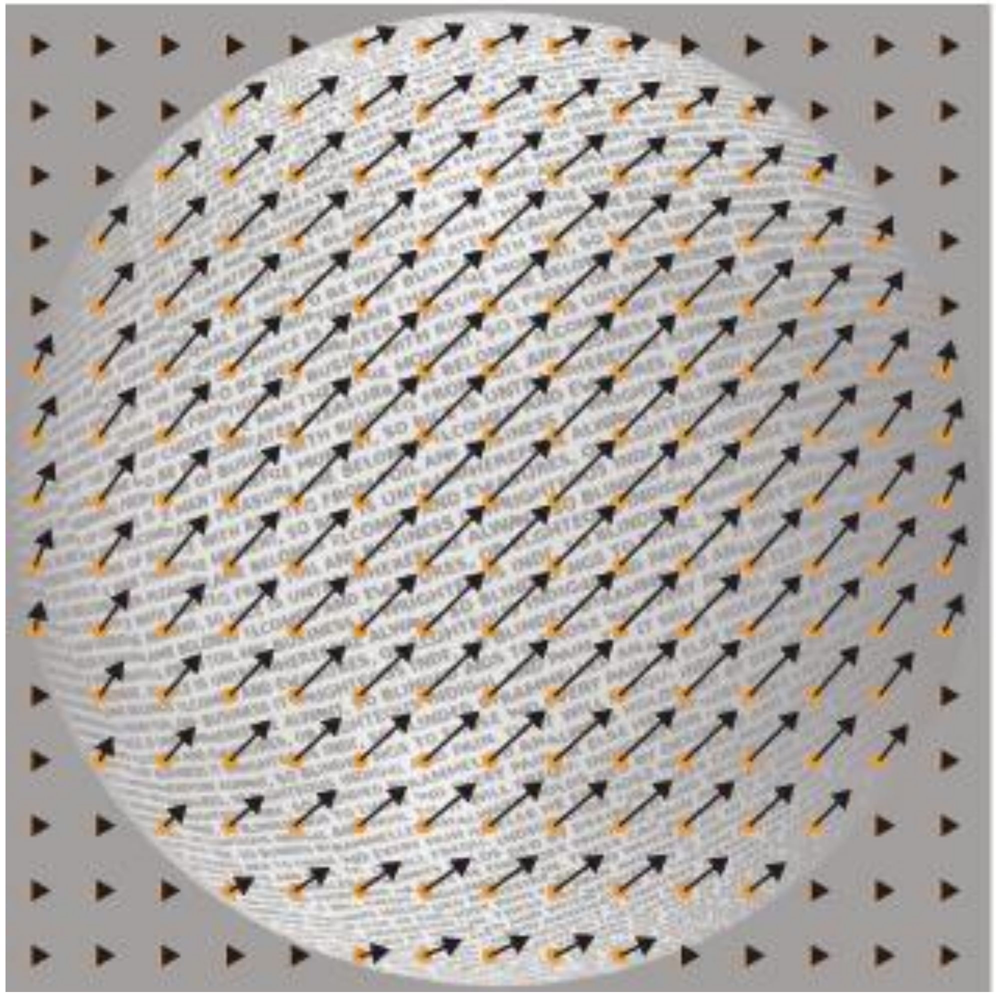
Recently: Extension to movies

SIGGRAPH ASIA 2011 HONG KONG

Apparent Resolution Enhancement for Animations

[Templin et al. SCCG 2011]





Conclusions

- Human perception is a crucial

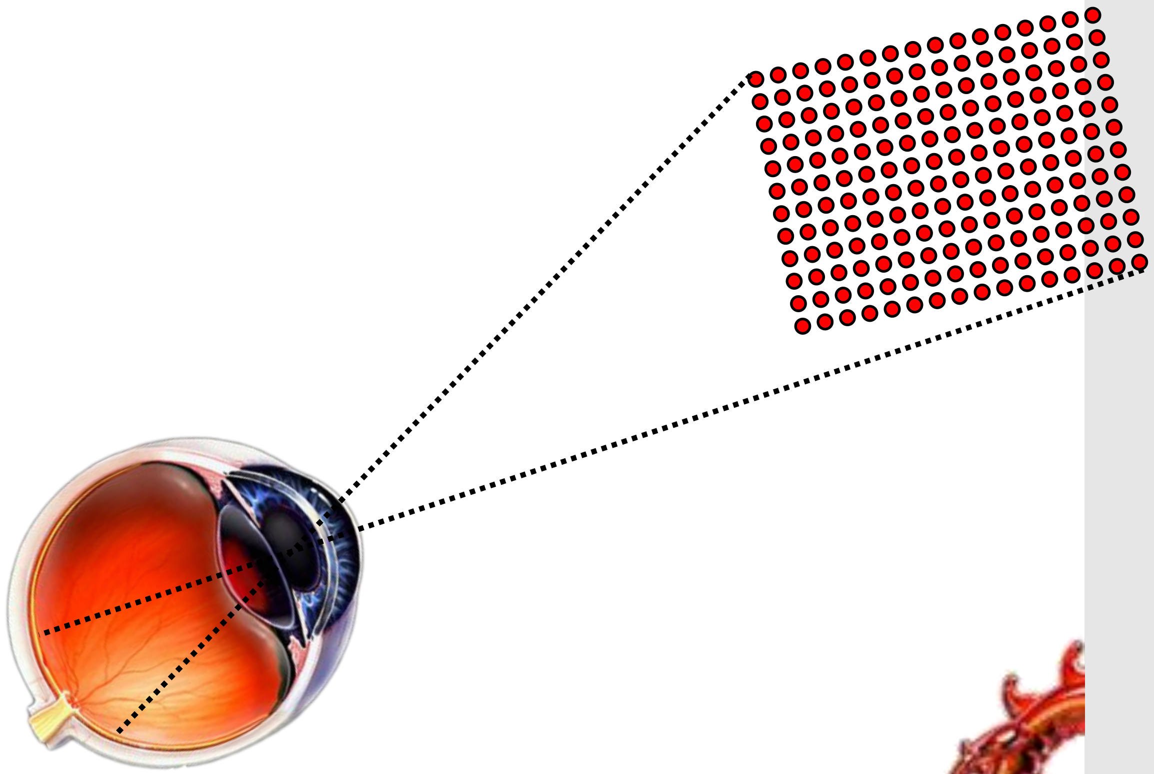
  component to high-quality imagery

- Resolution & Colors

  physical screen capabilities

- Works for large range

  of commonly used display devices



Future?

- Bigger,

  better,

  faster…
  - More realism
  - More details
  - More effects

- Higher quality beyond physical limitations
  - Only first steps in this direction
  - More to come…

Thank you very much for your attention!

Karol Myszkowski
karol@mpi-inf.mpg.de

Elmar Eisemann
eisemann@telecom-paristech.fr

Acknowledgment:
Daniel Scherzer, Robert Herzog and Dawid Pajak