IMPROVING THE PERFORMANCE OF WEB STREAMING BY SUPER-RESOLUTION UPSAMPLING

Nabajeet Barman[§], Yuriy Reznik[‡] and Patrick Wagstrom[‡]
[§]Kingston University London, UK
[‡]Brightcove Inc, Seattle, US

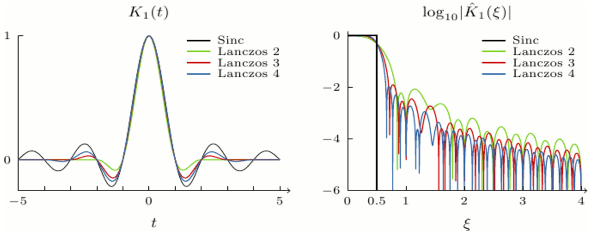
ACM Mile High Video
May 7-10, Denver

mhv/2023

Context & Objectives

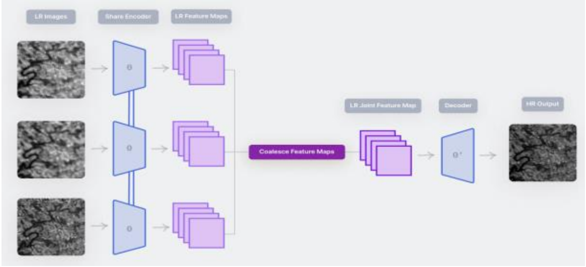
Traditional image / video scaling in web browsers

- Until recently – classic signal processing techniques:
  - Bi-cubic interpolation
  - Sinc, Lanczos, Mitchell-Netravali – type filters, etc.
- Typically implemented by GPUs / graphics drivers + OS layers



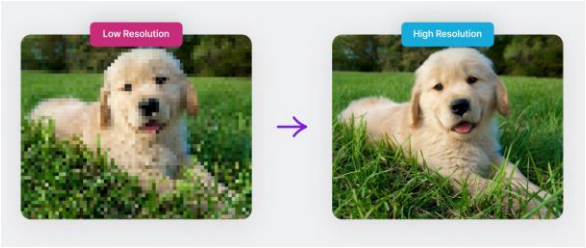
Super-resolution or "AI"-powered scaling

- Relatively new trend (2015+)
- Supported by many new GPUs (NVIDIA, AMD, etc.) and SDKs
- Proprietary APIs. Varying performance. No consistently across browsers/platforms.



Questions

- What are the advantages of SR over traditional scaling?
- How to model/quantify super-resolution scaling capability?
- How to use SR for improved image/video delivery?
- How significant could be the gains achieved by using SR?



Talk objectives:

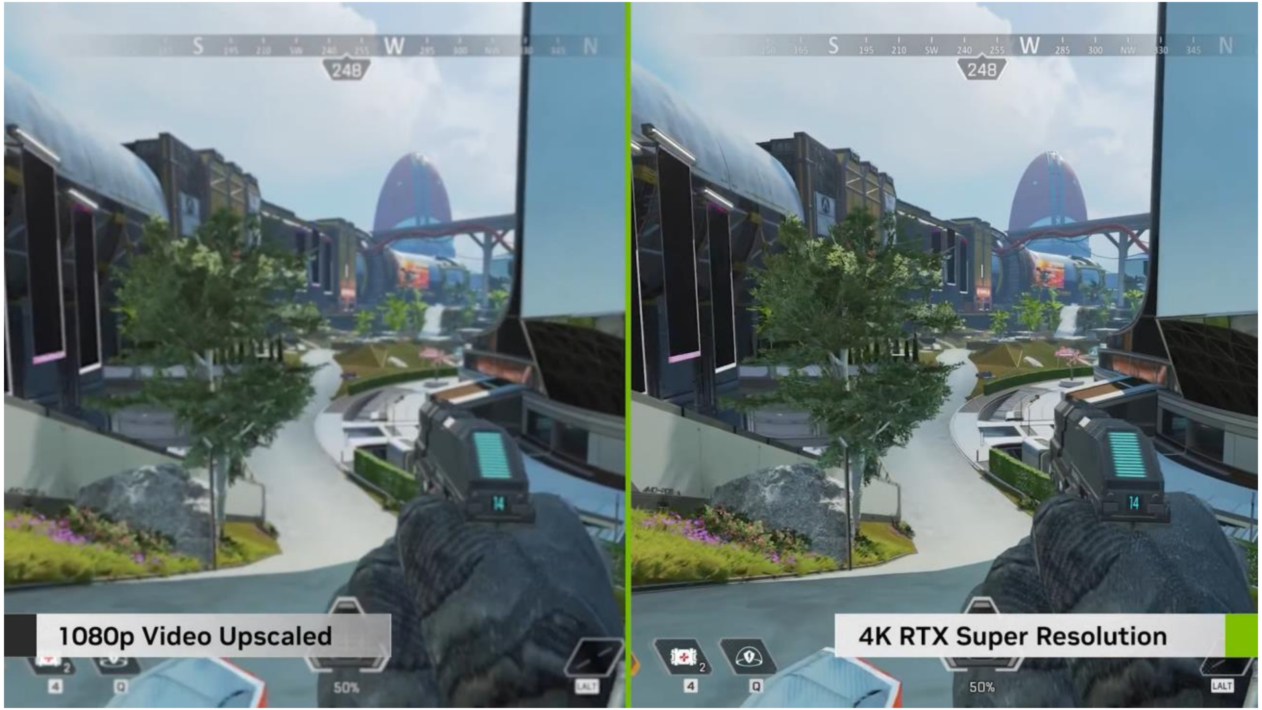
- Try to answer above questions.
- Bring some relevant results

Video Super Resolution (VSR) Support in Browsers

Traditional Videos

[CES 23: Nvidia outs RTX 4070 Ti, new RTX Video Super Resolution for Microsoft Edge & Chrome](https://www.neowin.net/news/ces-23-nvidia-outs-rtx-4070-ti-new-rtx-video-super-resolution-for-microsoft-edge--chrome/)
https://www.neowin.net/news/ces-23-nvidia-outs-rtx-4070-ti-new-rtx-video-super-resolution-for-microsoft-edge--chrome/
https://blogs.windows.com/msedgedev/2023/03/08/video-super-resolution-in-microsoft-edge/
https://www.amd.com/en/technologies/vsr

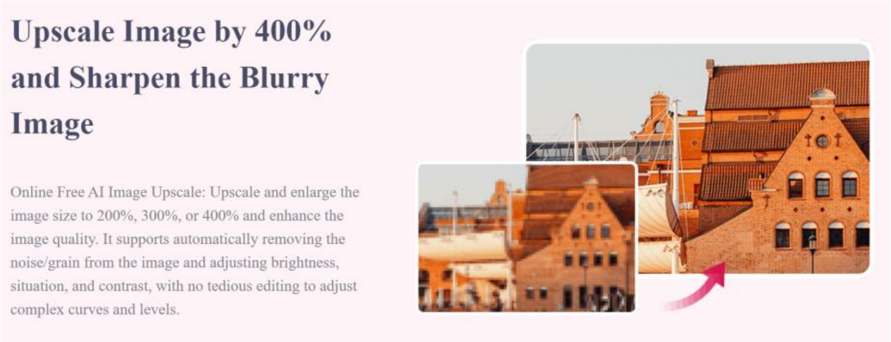
Video Super Resolution (VSR) Support in Browsers



Gaming!!

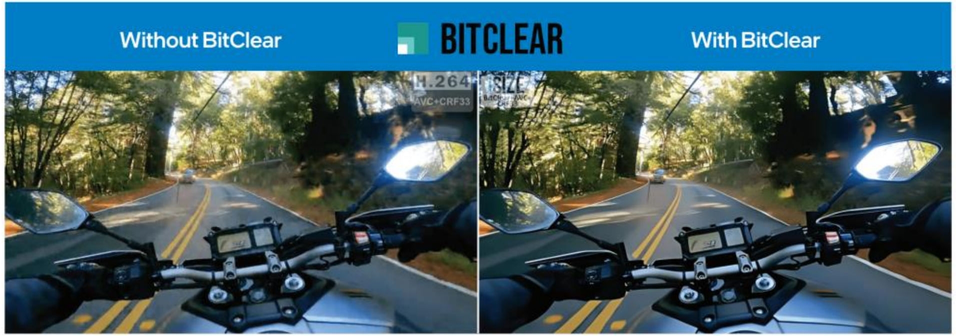
[CES 23: Nvidia outs RTX 4070 Ti, new RTX Video Super Resolution for Microsoft Edge & Chrome](https://www.neowin.net/news/ces-23-nvidia-outs-rtx-4070-ti-new-rtx-video-super-resolution-for-microsoft-edge--chrome/)
https://www.neowin.net/news/ces-23-nvidia-outs-rtx-4070-ti-new-rtx-video-super-resolution-for-microsoft-edge--chrome/
https://blogs.windows.com/msedgedev/2023/03/08/video-super-resolution-in-microsoft-edge/
https://www.amd.com/en/technologies/vsr

Examples of Proprietary Solutions



https://avc.ai/upscale-image/

Espresso Media



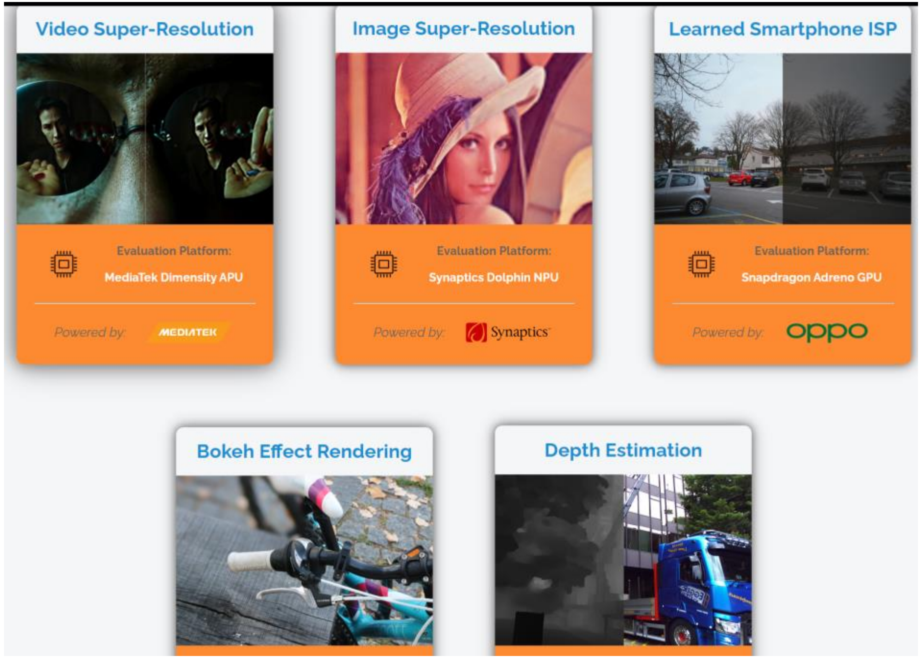
https://builders.intel.com/docs/networkbuilders/isize-bitclear-deep-perceptual-denoising-and-upscaling-with-intel-advanced-matrix-extensions-1674513550.pdf

MediaKind (MHV'23 Presentation)

Research Works

- HR Non-Homogeneous Dehazing started!
- Night Photography Rendering started!
- Real-Time Image Super-Resolution - Track 1 started!
- Real-Time Image Super-Resolution - Track 2 started!
- Bokeh Effect Transformation started!
- 360° Omnidirectional Super-Resolution (X4) - Track 1 Image started!
- 360° Omnidirectional Super-Resolution (X4) - Track 2 Video started!
- Single Image Super-Resolution (X4) Bicubic started!
- Light Field Image Super-Resolution Challenge started!
- Stereo Image Super-Resolution - Track 1 Fidelity & Bicubic started!
- Stereo Image Super-Resolution - Track 2 Perceptual & Bicubic started!
- Stereo Image Super-Resolution - Track 3 Fidelity & Realistic started!
- Quality Assessment for Video Enhancement started!
- Image Shadow Removal started!
- Video Colorization - Track 1 FID Optimization started!
- Video Colorization - Track 2 CDC Optimization started!
- Image Denoising started!
- Efficient Image Super-Resolution started!
- HR Depth from Images of Specular and Transparent Surfaces - Track 1 Stereo started!
- HR Depth from Images of Specular and Transparent Surfaces - Track 2 Mono started!

NTIRE Workshop 2023

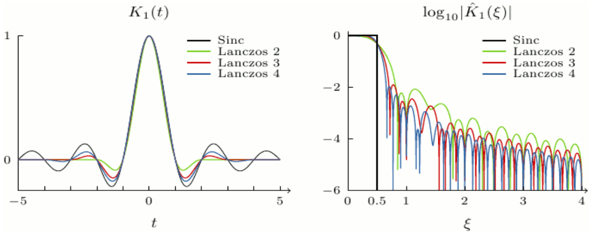


Mobile AI Workshop 2022

Context & Objectives

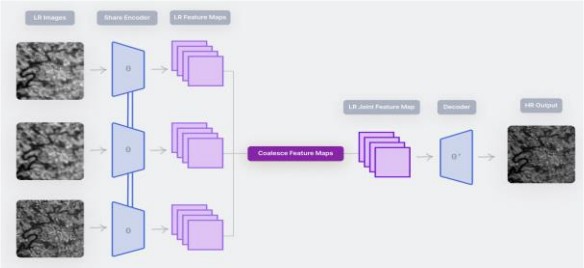
Traditional image / video scaling in web browsers

- Until recently – classic signal processing techniques:
  - Bi-cubic interpolation
  - Sinc, Lanczos, Mitchell-Netravali – type filters, etc.
- Typically implemented by GPUs / graphics drivers + OS layers



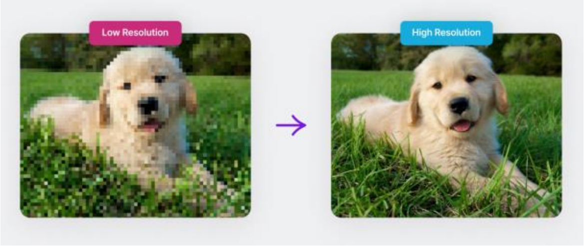
Super-resolution or “AI”-powered scaling

- Relatively new trend (2015+)
- Supported by many new GPUs (NVIDIA, AMD, etc.) and SDKs
- Proprietary APIs. Varying performance. No consistently across browsers/platforms.



Questions

- What are the advantages of SR over traditional scaling?
- **How to model/quantify super-resolution scaling capability?**
- How to use SR for improved image/video delivery?
- How significant could be the gains achieved by using SR?

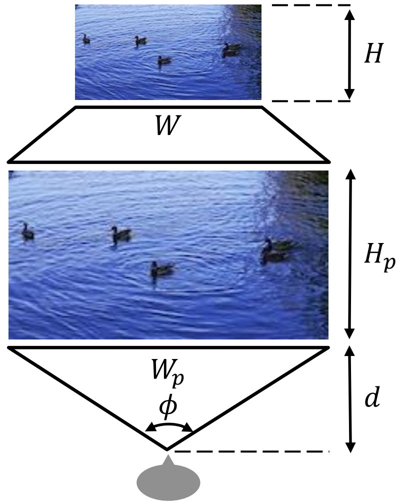


Understanding the Impacts of Scaling on Perceived Quality

Angular Metrics

Video reproduction chain

- **Encoded video:**
- **Scaled:**
- **Displayed:**
- **Perceived:**



Main parameters involved

Parameters	**Meaning**	**Unit**
$W,\ H$	encoded video width, height	pixels
$W_p,\ H_p$	display/player width, height	pixels
d	viewing distance	inches
ρ	display pixel density	dots per inch
$\phi = 2\arctan\left(\frac{W_p}{2d\rho}\right)$	viewing angle	degrees
$\phi_c = 2\arctan\left(\frac{W_p/W}{d\rho}\right)$	angle to 2 pixels (1 cycle)	degrees
$u = \frac{1}{\phi_c}$	angular resolution of video	cycles per degree (cpd)

Relevant for human perception

- **viewing angle ϕ** ➔ angular span of video frame, as visible on screen
- **angular resolution u** ➔ inverse of angular span of 2 pixels (length of smallest "cycle") in encoded video

Note: Another way to describe angular resolution is to say that it is a Nyquist frequency of video, expressed in angular units, reflecting projection the screen.

Scaling and Perceived Quality

Westerink-Roufs Model

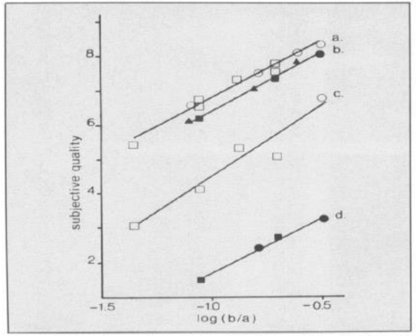


Figure 4. Subjective image quality as a function of the picture angle. The subjective quality values on the vertical axis are plotted as a function of log (b/a), which differs from the picture angle ϕ only by a constant. The different sets correspond to different resolutions: (a) greater than 33 ~/°, (b) between 23 and 28 ~/°, (c) between 8.6 and 8.7 ~/°, and (d) between 2.6 and 2.7 ~/°. Different symbols represent different viewing distances: ○ = 2.9 m, △ = 3.9 m, □ = 5.4 m. Every point is the result of 80 judgments.

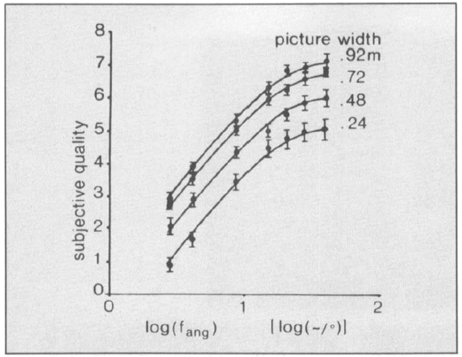


Figure 3. Subjective quality as a function of resolution. Every point is the result of 100 judgments, and the error indicated is plus or minus the standard error of the mean.

Observed phenomena:

- Perceived quality grows approximately as logarithm of viewing angle ($\boldsymbol{\phi}$)
- Perceived quality also grows with angular resolution ($\boldsymbol{u}$), but saturates at around 25-40 cycles/degree

Model describing these effects (*)

$$Q_{WR}(\phi, u) = 3.6\log(\phi) + 2.9 + 4.6\log(u) + 2.7\log(u)^2 - 1.7\log(u)^3$$

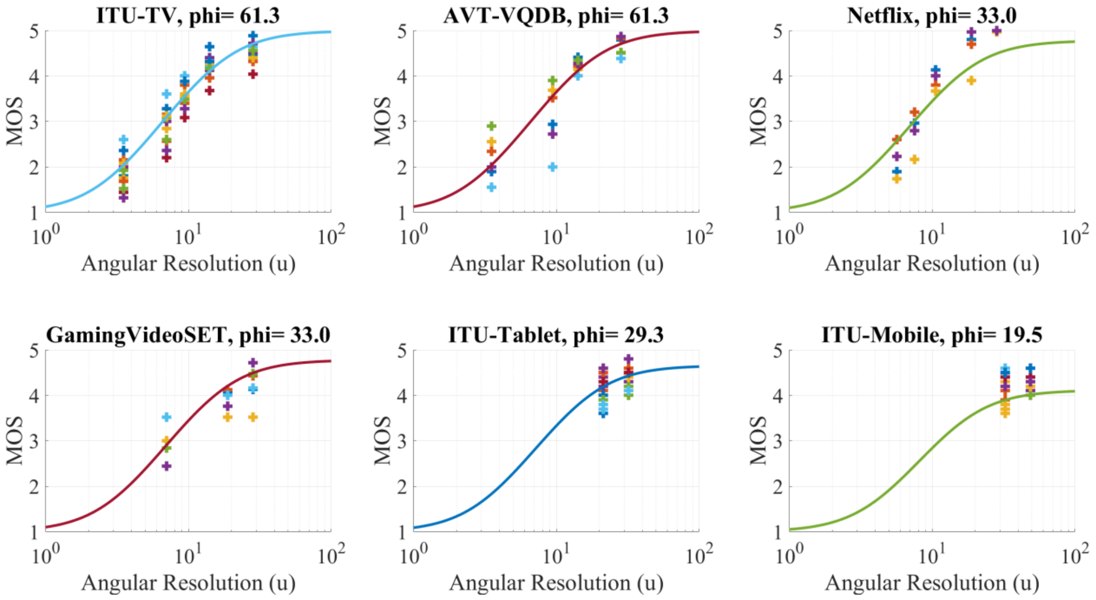
(*) J. Westerink and J. Roufs, "Subjective image quality as a function of viewing distance resolution and picture size," SMPTE Journal, vol. 98, 1989, pp. 113-19.

Generalized Westerink-Roufs Model

Generalized model (*)

$$Q_{WR}(\phi,u) = \log\left(\alpha + \beta \cdot \left(1+\left(\frac{\phi}{\phi_s}\right)^{-k}\right)^{-\frac{\gamma}{k}} \cdot \left(1+\left(\frac{u}{u_s}\right)^{-l}\right)^{-\frac{\delta}{l}}\right)$$

- ϕ – viewing angle, u – angular resolution
- $\alpha, \beta, \gamma, \delta, \phi_s, k, u_s, l$ – model parameters

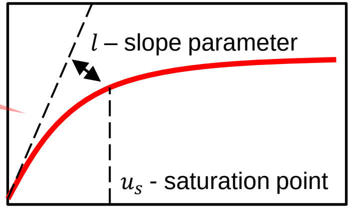


(*) N. Barman, et al, "Generalized Westerink-Roufs Model for Predicting Quality of Scaled Video," *QoMEX, 2022*

Modeling the Effects of Different Upscaling Algorithms

Generalized WR model (*)

$$Q_{WR}(\phi,u)=\log\left(\alpha+\beta\cdot\left(1+\left(\frac{\phi}{\phi_s}\right)^{-k}\right)^{-\frac{\gamma}{k}}\cdot\left(1+\left(\frac{u}{u_s}\right)^{-l}\right)^{-\frac{\delta}{l}}\right)$$



- Key parameters:
  - $\boldsymbol{u_s, l}$ – saturation point and slope for angular resolution
  - these are the main parameters that may be affected by the upscaling techniques

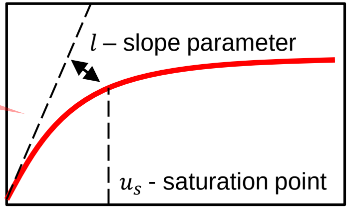
(*) N. Barman, et al, "Generalized Westerink-Roufs Model for Predicting Quality of Scaled Video," *QoMEX, 2022*

(**) A. Mackin, et al, “A Study of Subjective Video Quality at Various Spatial Resolutions,” *ICIP* 2018.

Modeling the Effects of Super Resolution

Generalized WR model (*)

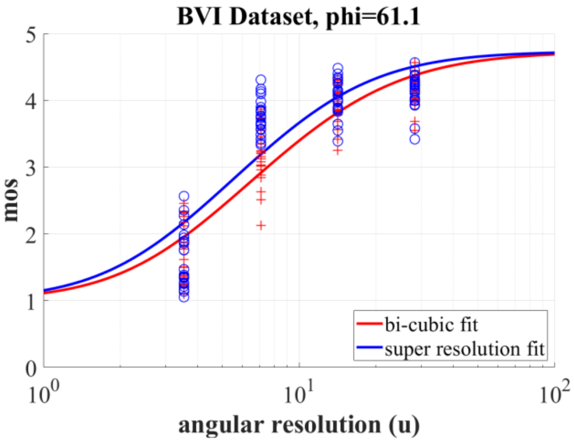
$$Q_{WR}(\phi,u) = \log\left(\alpha + \beta \cdot \left(1+\left(\frac{\phi}{\phi_s}\right)^{-k}\right)^{-\frac{\gamma}{k}} \cdot \left(1+\left(\frac{u}{u_s}\right)^{-l}\right)^{-\frac{\delta}{l}}\right)$$



- Key parameters:
  - u_s, l – saturation point and slope for angular resolution
  - these are the main parameters that may be affected by the upscaling techniques

Fit to different up-sampling methods:

- BVI dataset(**)
- **The use of SR lowers the saturation u_s and increases the slope parameter l in the generalized WR model.**



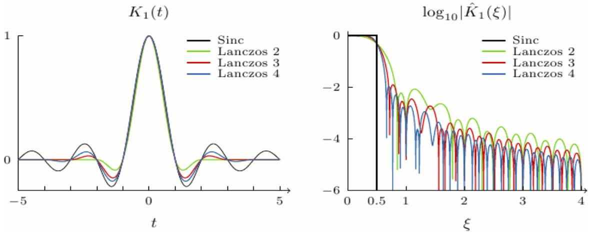
(*) N. Barman, et al, "Generalized Westerink-Roufs Model for Predicting Quality of Scaled Video," *QoMEX, 2022*

(**) A. Mackin, et al, "A Study of Subjective Video Quality at Various Spatial Resolutions," *ICIP* 2018.

Context & Objectives

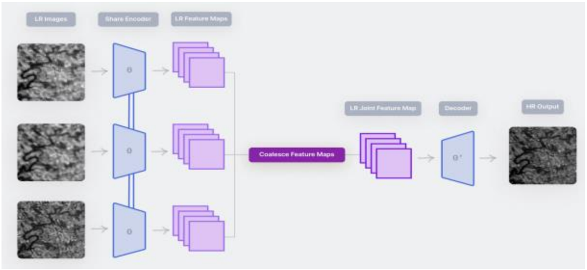
Traditional image / video scaling in web browsers

- Until recently – classic signal processing techniques:
  - Bi-cubic interpolation
  - Sinc, Lanczos, Mitchell-Netravali – type filters, etc.
- Typically implemented by GPUs / graphics drivers + OS layers



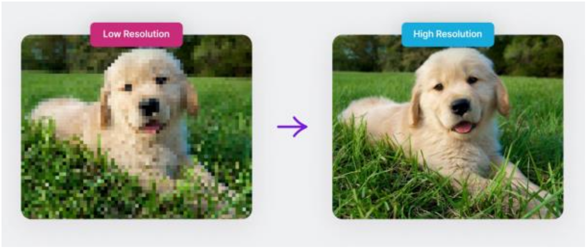
Super-resolution or "AI"-powered scaling

- Relatively new trend (2015+)
- Supported by many new GPUs (NVIDIA, AMD, etc.) and SDKs
- Proprietary APIs. Varying performance. No consistently across browsers/platforms.



Questions

- What are the advantages of SR over traditional scaling?
- How to model / quantify / declare super-resolution scaling capability?
- **How to use SR for improved image/video delivery?**
- **How significant could be the gains achieved by using SR?**

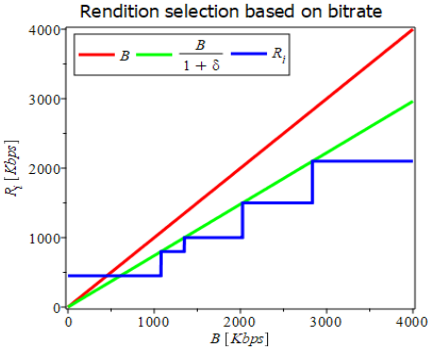


Resolution Selection for ABR Streaming

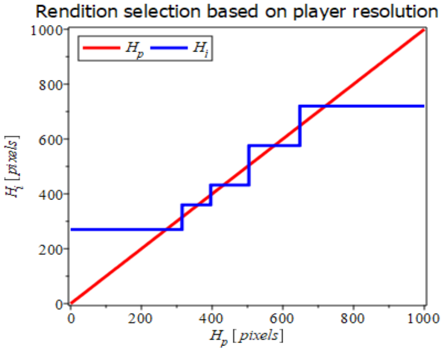
Adaptation to Player Size

Conceptual model of adaptation logic in streaming clients (*)

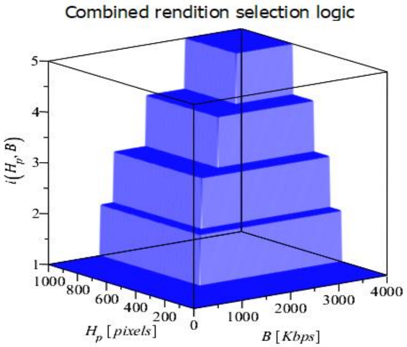
- Adaptation to network bandwidth



Adaptation to player size



Combined selection logic



(*) Y. Reznik, K. Lillevold, A. Jagannath, and X. Li, "Towards Understanding of the Behavior of Web Streaming," *PCS'21*, Bristol, UK, June 29 - July 2, 2021

Optimal Device-aware Resolution Selection for ABR Streaming

Optimal resolution-based selection algorithm (*):

Algorithm 1: Optimal Rendition Resolution Selection Based on Player Size

Data:

- Viewing angle ϕ
- Angular resolution μ
- Available video rendition heights, $H_{renditions} = H_1, ...H_n$, such that $H_1 \leq .. \leq H_n$
- Player Window Height H_p
- Distance from the display d
- Effective pixel density of the screen, ρ

Result: Best rendition height, H_{best}

```
MOS_best = 0;
best_rendition-index = 0;
for i ← 1 to n do
    Calculate Viewing angle φ
    Calculate Angular resolution μ
    Calculate MOS Q(φ, μ) ;
    if MOS is ≤ best_mos then
        MOS_best = MOS ;
        best_rendition-index = i ;
    end
end
H_best = H_renditions(best_rendition-index)
```

Computed by using the generalized Westerink-Roufs model, calibrated to specifics of viewing setup

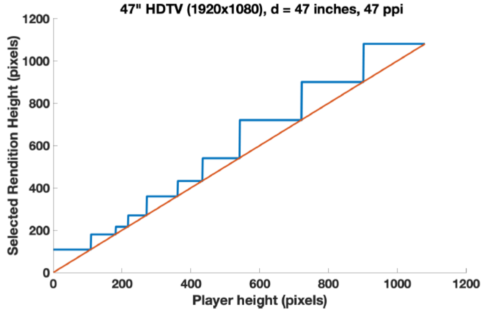
Notes:

- Effectively, this is a search for a rendition delivering best MOS (as predicted by Westerink-Roufs model) for a given rendition resolution and other reproduction setup parameters.

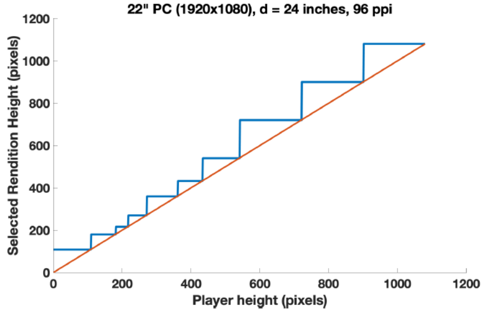
(*) Y. Reznik, et al, "Optimal Rendition Resolution Selection Algorithm for Web Streaming Players," *SPIE ADIP 2022.*

Optimal Adaptation: Different Screens

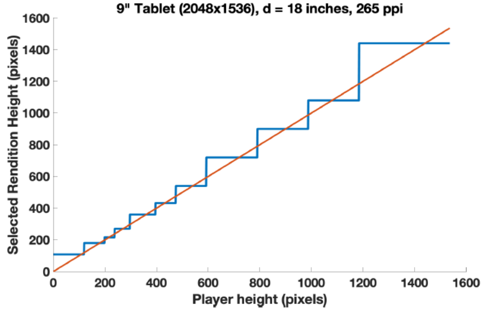
Observed selection behavior with different devices/screens:



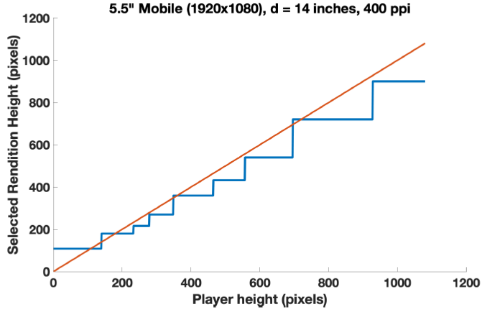
(a) 47″ HDTV (1920x1080), d = 69.12″(3H), 47 ppi.



(b) 24″ PC (1920x1080), d = 22″, 96 ppi.



(c) 9″ Tablet (2048x1536), d = 18″, 265 ppi.



(d) 5.5″ Mobile (1920x1080), d = 14″, 400 ppi.

Main Observation: Optimal selection behavior is different for different devices/screens!

(*) Y. Reznik, et al, "Optimal Rendition Resolution Selection Algorithm for Web Streaming Players," *SPIE ADIP 2022.*

Optimal SR-aware Resolution Selection for ABR Streaming

Optimal SR-aware Adaptation

Modified algorithms, accounting for type of scaling:

Algorithm 1: Optimal Rendition Resolution Selection Based on Player Size and BC Upsampling Algorithm

Data:

Viewing angle, ϕ
Angular resolution, μ
Number of available renditions, n
Available video rendition heights, $H_{renditions} = H_1, ...H_n$
Player Window Height, H_p
Distance from the display, d
Effective pixel density of the screen, ρ
Client Side Upsampling Algorithm, UP_{algo} = BC
Model fit values, α=2.72, β=106.91, ϵ=1.08, γ=1.55ϵ, δ=2.12ϵ.

Result: Best rendition height (BC upsampling), $H_{best_{BC}}$, and Best MOS (BC upsampling), $MOS_{best_{BC}}$

$best_{mos} = 0$;
$best_{rendition-index} = 1$;
for $i \leftarrow 1$ **to** n **do**
  Calculate Viewing angle ϕ
  Calculate Angular resolution μ
  $\mu_s = 13.93; l = 1.76$; /* BC upsampli
  Calculate MOS, $Q(\phi, \mu)$;
  if MOS is $\geq$ $best\_mos$ **then**
    best_mos = MOS ;
    $best_{rendition-index} = i$;
  end
end
$H_{best_{BC}} = H_{renditions}(best_{rendition-index})$
$MOS_{best_{BC}} = best_{mos}$

Generalized Westerink-Roufs model for Bicubic scaling

Algorithm 2: Optimal Rendition Resolution Selection Based on Player Size and SR Upsampling Algorithm

Data:

Viewing angle, ϕ
Angular resolution, μ
Number of available renditions, n
Available video rendition heights, $H_{renditions} = H_1, ...H_n$
Player Window Height, H_p
Distance from the display, d
Effective pixel density of the screen, ρ
Model fit values, $\alpha = 2.72$, $\beta = 106.91$, $\epsilon = 1.08$, $\gamma = 1.55\epsilon$, $\delta = 2.12\epsilon$.
MOS values from Algorithm 1, $MOS_{best_{BC}}$

Result: Best rendition height (SR upsampling), $H_{best_{SR}}$, and Best MOS (SR upsampling), $MOS_{best_{SR}}$

$best_{mos} = MOS_{best_{BC}}$; /* MOS from Algorithm 1 */
$best_{rendition-index} = 1$;
for $i \leftarrow 1$ **to** n **do**
  Calculate Viewing angle ϕ
  Calculate Angular resolution μ
  $\mu_s = 12.24; l = 2.06$; /* SR Ups
  Calculate MOS, $Q(\phi, \mu)$;
  if MOS is $\geq$ $best\_mos$ **then**
    best_mos = MOS ;
    $best_{rendition-index} = i$;
    break; ; /* Minimum Rendition found, exit */
  end
end
$H_{best_{SR}} = H_{renditions}(best_{rendition-index})$
$MOS_{best_{SR}} = best_{mos}$

Generalized Westerink-Roufs model for SR scaling

Principle of operation:

- Algorithm 1 finds rendition delivering best possible quality by considering standard bicubic upscaling.
- Algorithm 2 find rendition matching the level of quality achievable with algorithm 1, but considering SR upscaling in rendering.

Example of SR-aware Adaptation

Ladder of resolutions (DVB DASH):

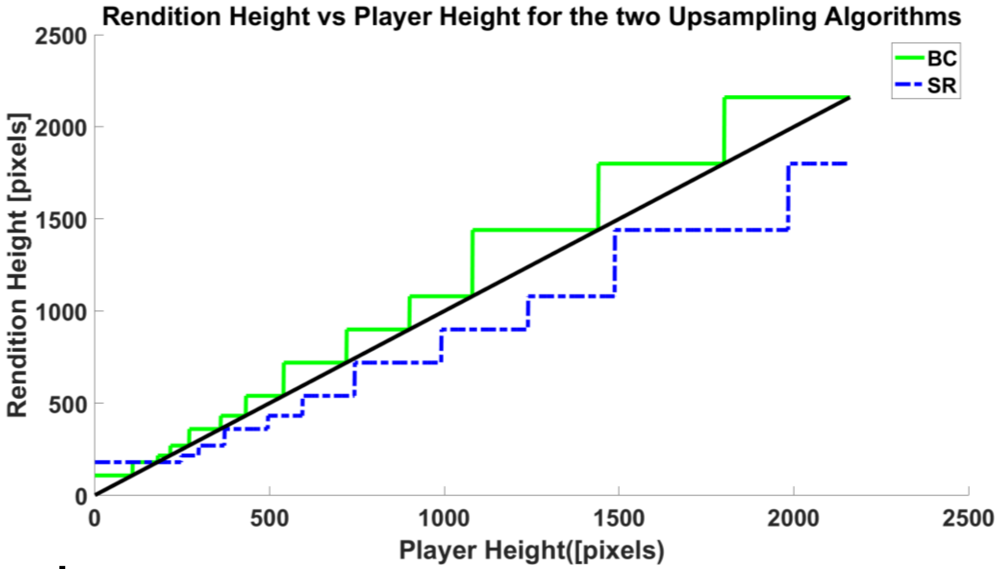
Horizontal @maxwidth	Vertical @maxheight
3 840	2 160
3 200	1 800
2 560	1 440
1 920	1 080
1 600	900
1 280	720
960	540
768	432
640	360
480	270
384	216
320	180
192	108

Example Encoding Ladder

Stream	Codec	Width [pixels]	Height [pixels]	Framerate [fps]	Bitrate [kbps]
1	HEVC	192	108	59.94	260
2	HEVC	320	180	59.94	500
3	HEVC	384	216	59.94	640
4	HEVC	480	270	59.94	930
5	HEVC	640	360	59.94	1350
6	HEVC	768	432	59.94	1960
7	HEVC	960	540	59.94	2550
8	HEVC	1280	720	59.94	3690
9	HEVC	1600	900	59.94	5350
10	HEVC	1920	1080	59.94	6950
11	HEVC	2560	1440	59.94	11130
12	HEVC	3200	1800	59.94	16140
13	HEVC	3840	2160	59.94	23400

Example of SR-aware Adaptation

Optimal Bicubic vs SR-based selection methods (TV screen):



Ladder of resolutions (DVB DASH):

Horizontal @maxwidth	Vertical @maxheight
3 840	2 160
3 200	1 800
2 560	1 440
1 920	1 080
1 600	900
1 280	720
960	540
768	432
640	360
480	270
384	216
320	180
192	108

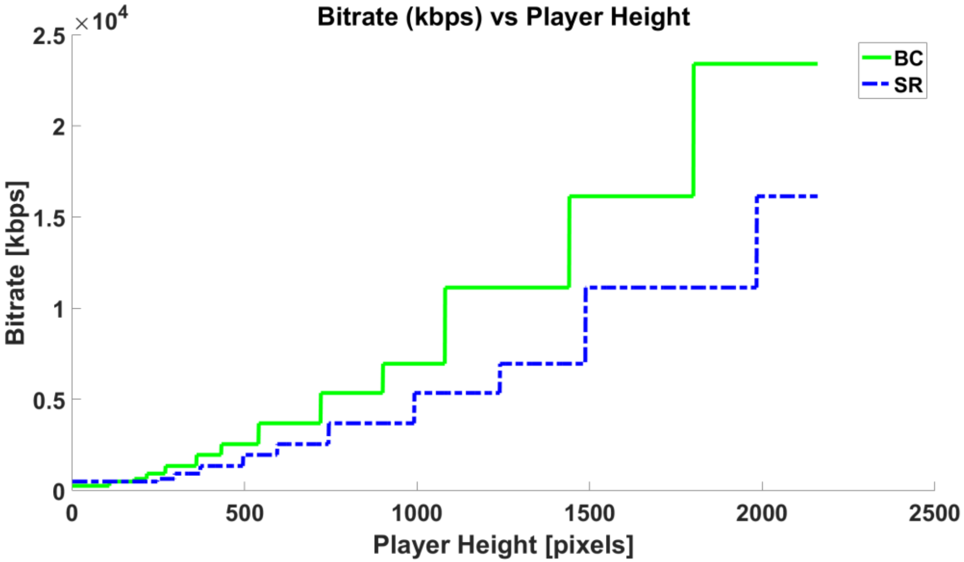
Observations:

- SR-based upsampling enables much more conservative choices of rendition resolutions
- In this example(*), we see about 16% reduction in frame height or 30% in pixel count in high-resolution regime.

(*) Uses SR algorithm from: J. Kim, J. K. Lee, and K. M. Lee. “Accurate Image Super-Resolution Using Very Deep Convolutional Networks”.
In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, pp. 1646–1654.

Example of SR-aware Adaptation

Optimal Bicubic vs SR-based selection methods (TV screen):



Ladder of resolutions (DVB DASH):

Horizontal @maxwidth	Vertical @maxheight
3 840	2 160
3 200	1 800
2 560	1 440
1 920	1 080
1 600	900
1 280	720
960	540
768	432
640	360
480	270
384	216
320	180
192	108

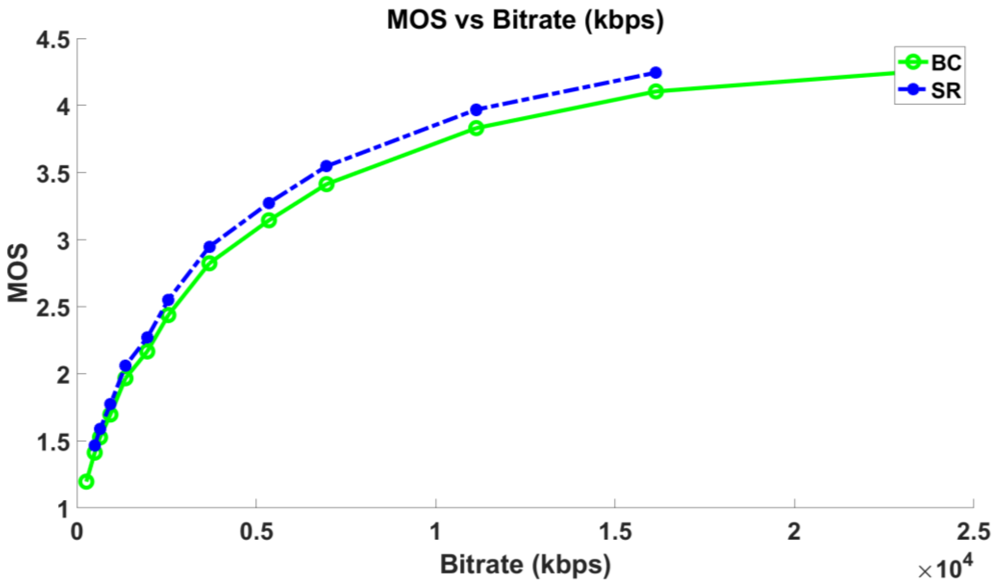
Observations:

- SR-based upsampling selects renditions of much lower bitrate, resulting in significant bandwidth savings!!
- In this example(*), we see about 38.9% bitrate savings!!!
- **NB: SR brings potential for significantly reducing the use of network bandwidth!**

(*) Uses SR algorithm from: J. Kim, J. K. Lee, and K. M. Lee. “Accurate Image Super-Resolution Using Very Deep Convolutional Networks”. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, pp. 1646–1654.

Example of SR-aware Adaptation

Optimal Bicubic vs SR-based selection methods (TV screen):



Ladder of resolutions (DVB DASH):

Horizontal @maxwidth	Vertical @maxheight
3 840	2 160
3 200	1 800
2 560	1 440
1 920	1 080
1 600	900
1 280	720
960	540
768	432
640	360
480	270
384	216
320	180
192	108

Observations:

- In this example(*), we see the use of SR upscaling can result in higher MOS scores at reduced bitrate values, especially for high-resolution playback

(*) Uses SR algorithm from: J. Kim, J. K. Lee, and K. M. Lee. “Accurate Image Super-Resolution Using Very Deep Convolutional Networks”. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, pp. 1646–1654.

Discussion and Next Steps

Discussion

SR techniques clearly show some promise

- They seem to work (very well for images, less well for videos, but getting better)
- Their performance can be characterized and modeled, and potential gains are pretty impressive

However to start using SR techniques we must have:

- Clearly defined APIs – supported by all browsers and platforms
- Clearly defined means for quantifying the effects on quality of scaling of different SR implementations
  - E.g., saturation point and gain parameters (u_s, l) in generalized WR model.

Discussion

SR techniques clearly show some promise

- They seem to work (very well for images, less well for videos, but getting better)
- Their performance can be characterized and modeled, and potential gains are pretty impressive

However to start using SR techniques we must have:

- Clearly defined APIs – supported by all browsers and platforms
- Clearly defined means for quantifying the effects on quality of scaling of different SR implementations
  - E.g., saturation point and gain parameters (u_s, l) in generalized WR model.

Possible steps forward:

- Standardize APIs: At browser and also possibly - OS levels.
- Standardize (or otherwise fully specify and fix) the implementations of SR algorithms
  - Similar to codecs: codecs have different performance, but everyone knows exactly what they are.
  - Pros: everything is transparent. Cons: possible vendors' reluctance to open algorithms.
- Standardize quality models / performance parameters for SR algorithms
  - Treat SR algorithms as black boxes, but rely on such models/parameters to make decisions
  - Where to define such models/metrics?

Lots of opportunities and challenges for the industry!!!

THANK YOU

https://www.linkedin.com/in/nabajeetbarman/