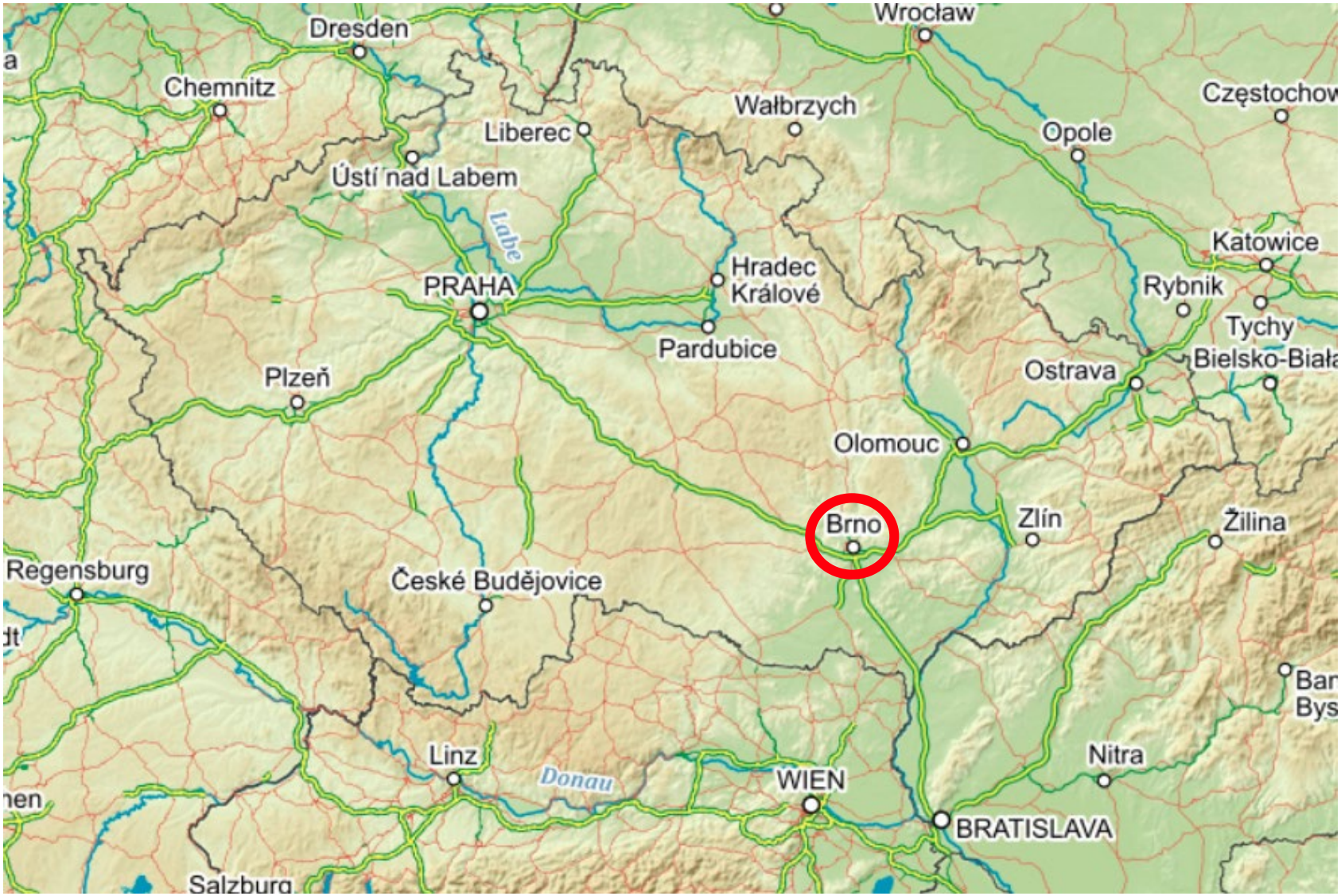
Introduction to Image and Video Quality Assessment

Martin Čadík

Brno University of Technology, Faculty of Information Technology, CPhoto@FIT
Božetěchova 1/2, 612 66 Brno, Czech Republic
cadik@fit.vutbr.cz, http://cadik.posvete.cz, @cadikm

BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

Brno University of Technology, Czech Republic



CPhoto@FIT

- CPhoto@FIT http://cphoto.fit.vutbr.cz/
  B.Sc., M.Sc., Ph.D. and other projects
- Computational Photography Course
  http://www.fit.vutbr.cz/study/course-l.php?id=12352
- VGS-IT - Invited Talks on Vision, Graphics, and Speech@FIT
  http://vgs-it.fit.vutbr.cz/

- HiVisualComputing meeting in the mountains
  http://www.hiviscomp.cz/

2014 HiVisComp 2015 HiVisComp 2016 HiVisComp 2017 HiVisComp 2018 HiVisComp 2019 HiVisComp 2020 HiVisComp


HEROBUS

[Zvěřina 18]

[Klusoň 17]

[Ostroukh 17]

[Jelen 19]

[Jedlička 19]

[Ostroukh 17]

Interactive Walkthrough FIT

- http://cphoto.fit.vutbr.cz/fit/

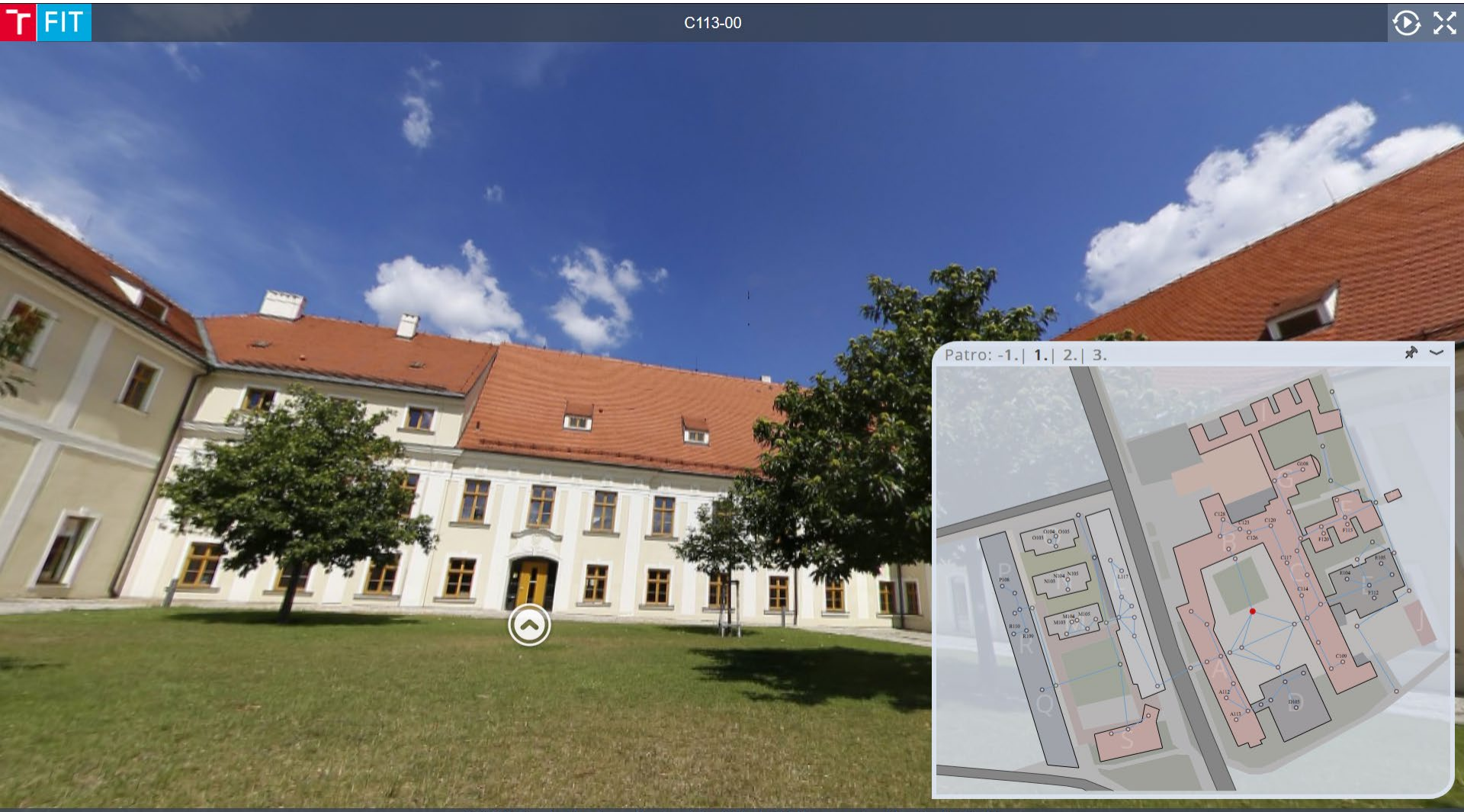


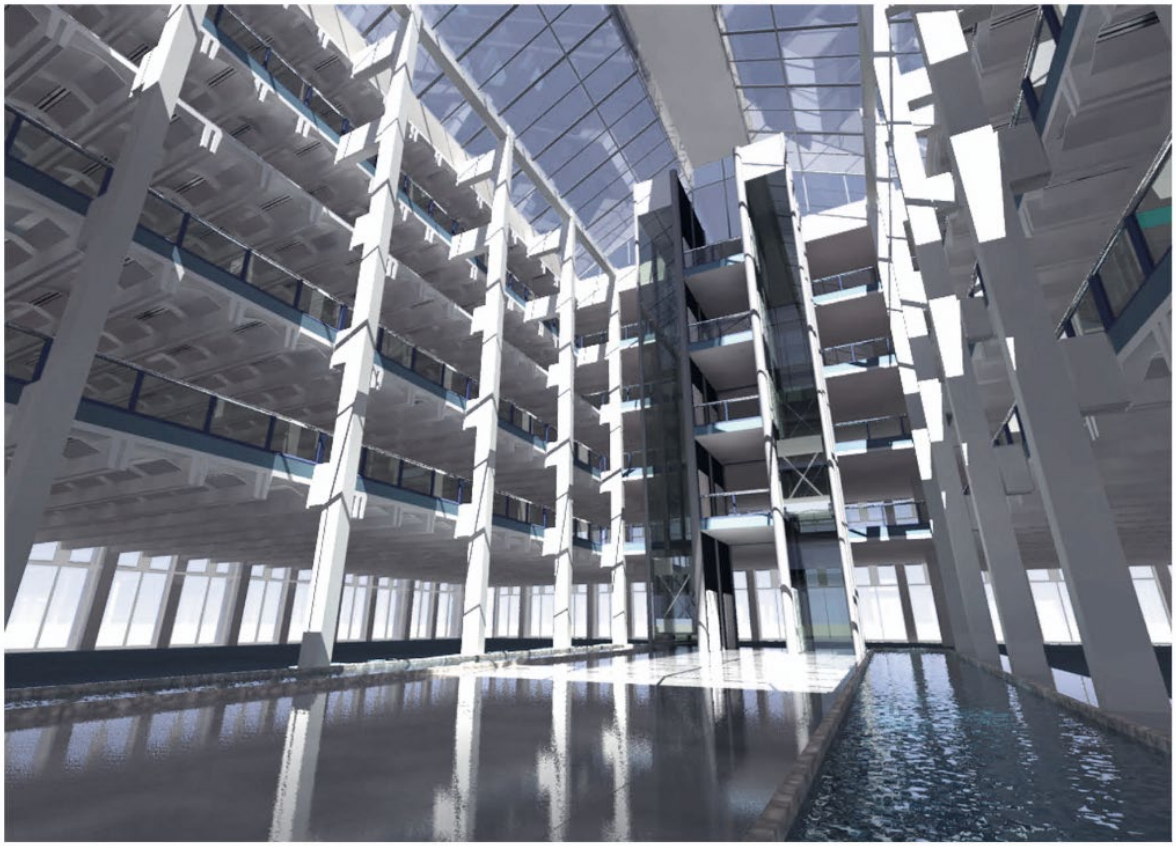
IMAGE AND VIDEO QUALITY ASSESSMENT

Motivation

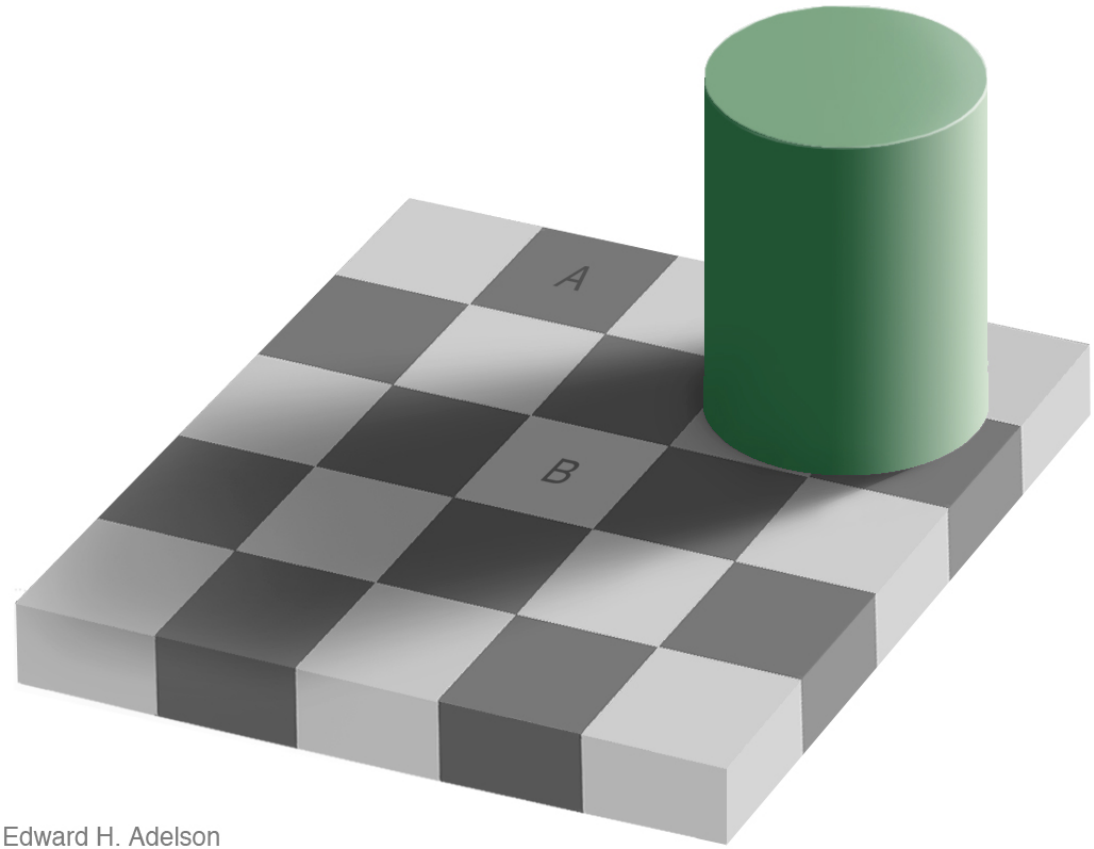
- Physical fidelity
  - Measure illumination
  - RMSE, PSNR

Motivation

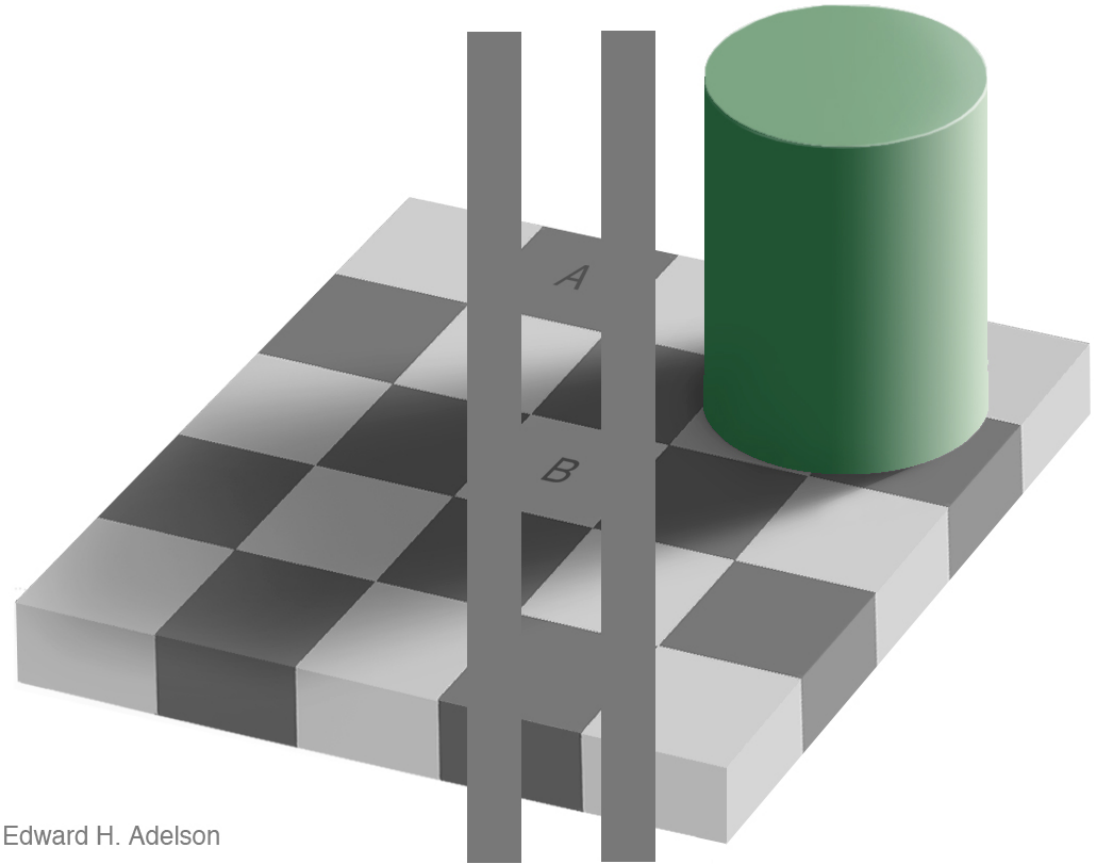
- Perceptual/cognitive fidelity
  - Specialized image/video quality metrics



Human Visual Perception



Edward H. Adelson



Edward H. Adelson

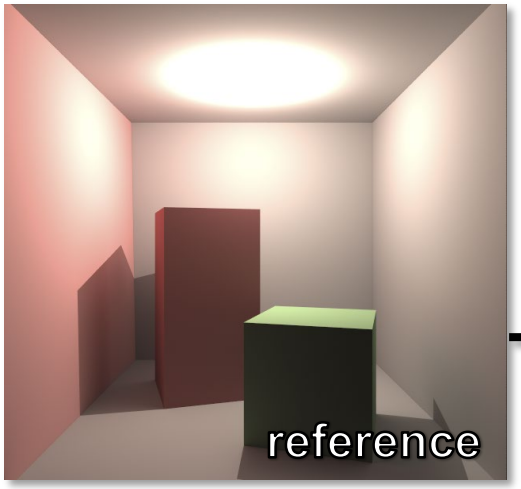
Overview

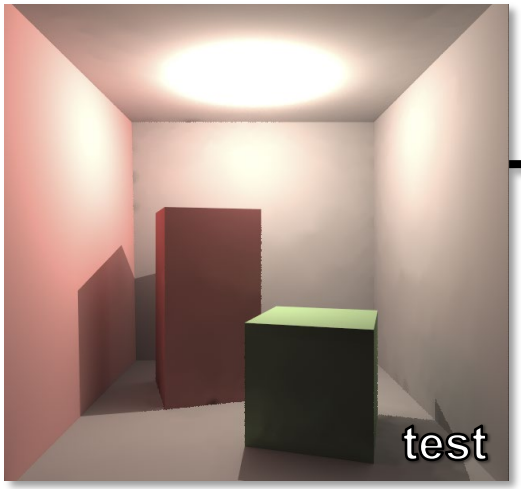
- Introduction to Objective Quality Assessment
- Image Quality Assessment
  - HVS-based Metrics (bottom-up)
  - Structural Similarity (top-down)
  - Data-driven Approaches (top-down)
- Video Quality Assessment

FR Quality Assessment (IQA, VQA)

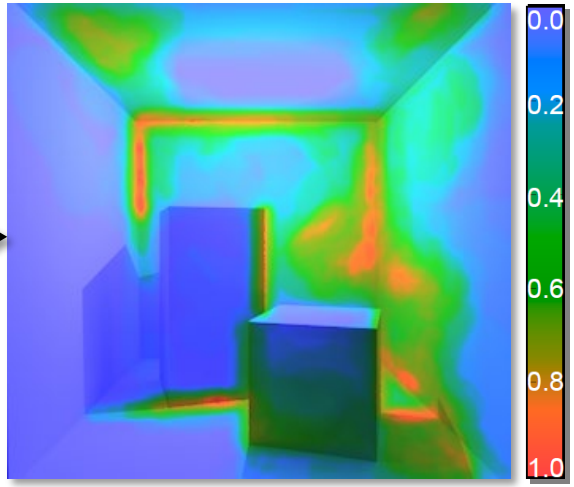
+ Reliable **- High cost**

Full-Reference Image Quality Metrics





FR IQM



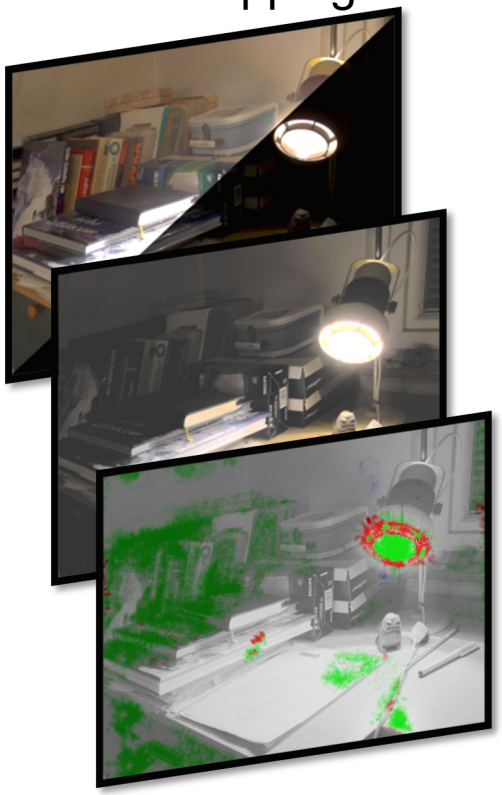
localized distortion map

Image Quality Metrics

- What are they good for?
  - Algorithm analysis/validation/evaluation
  - Quality monitoring for compression/transmission
  - Guiding/parameter optimization of renderers
  - Stopping criterions
  - Speed/quality enhancements
  - Machine learning (loss)

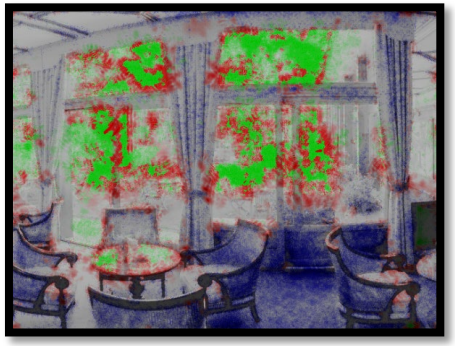
Applications in HDR

Tone Mapping



Inverse
Tone Mapping

Display
Analysis

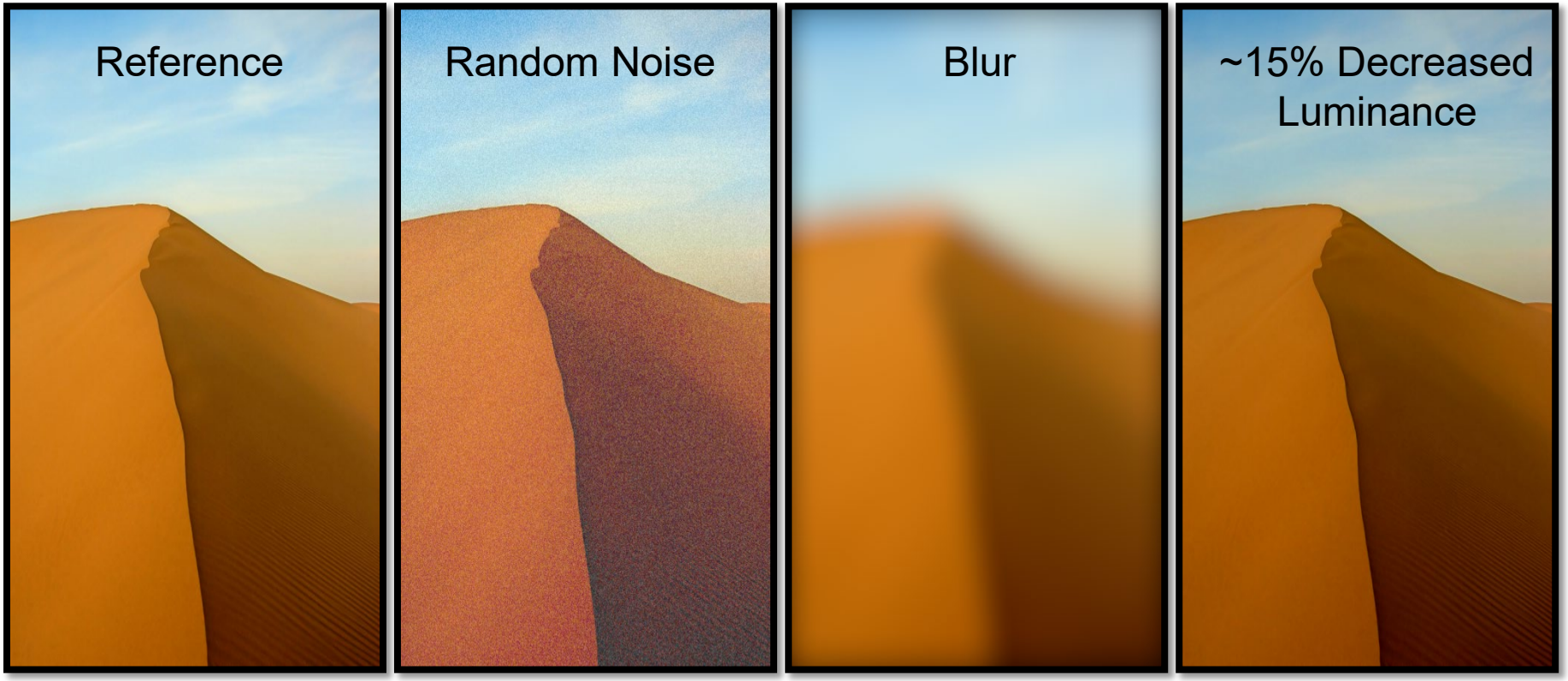


Math-based FR Metrics

- AD $M = |ref - test|$
- (R)MSE $M = (ref - test)^2$ $MSE = \frac{1}{n}\sum_{i=1}^{n}(ref_i - test_i)^2$
- PSNR $PSNR = 10\log_{10}\frac{MAX^2}{MSE}$
- sCORREL $M = SRCC(ref,\ test)$

  (Spearman's rank correlation coefficient)

Simple Full-reference Metrics

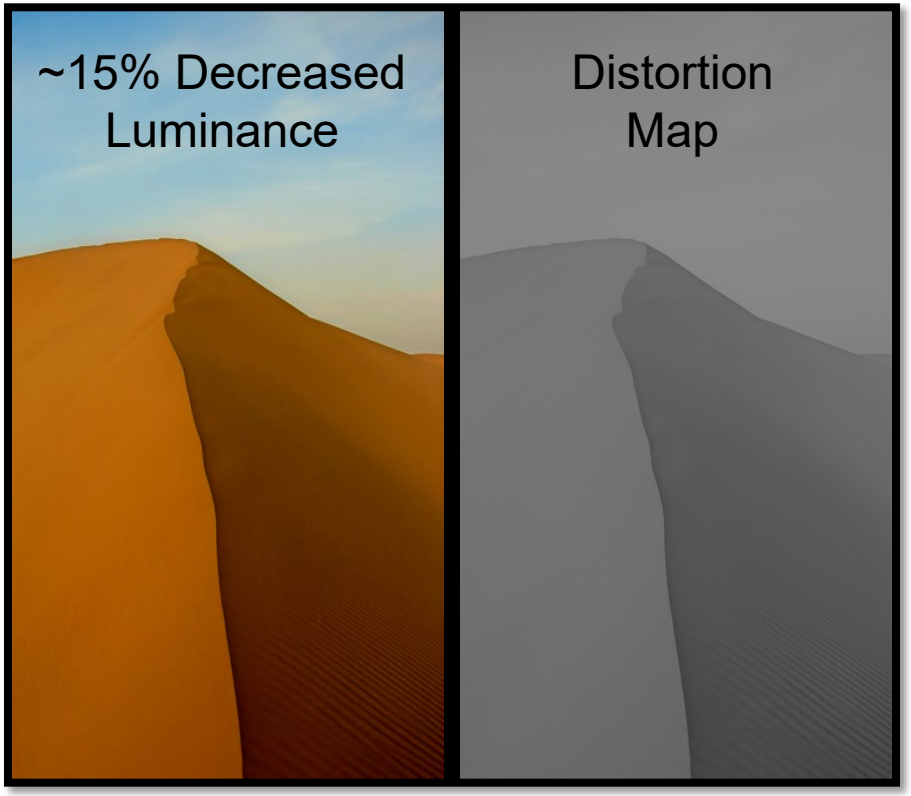


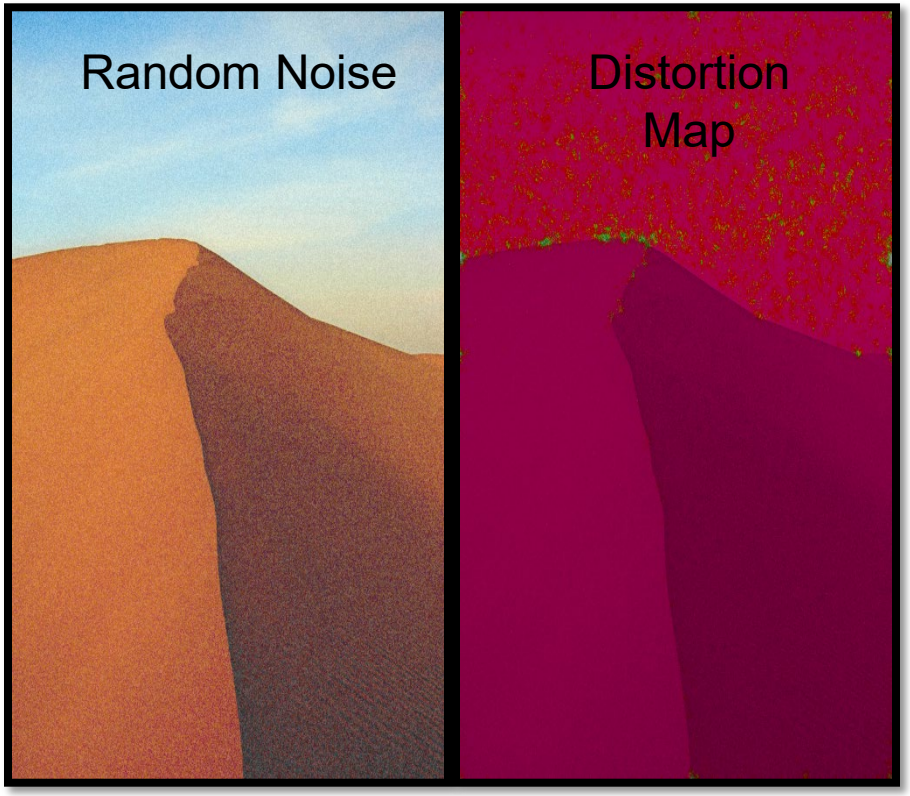
MSE = 280 MSE = 280 MSE = 280 !

Overview

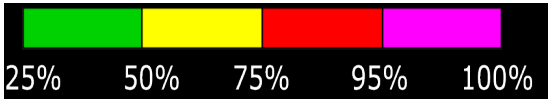
- Introduction to Objective Quality Assessment
- Image Quality Assessment
  - HVS-based Metrics (bottom-up)
  - Structural Similarity (top-down)
  - Data-driven Approaches (top-down)
- Video Quality Assessment

HVS Based Metrics



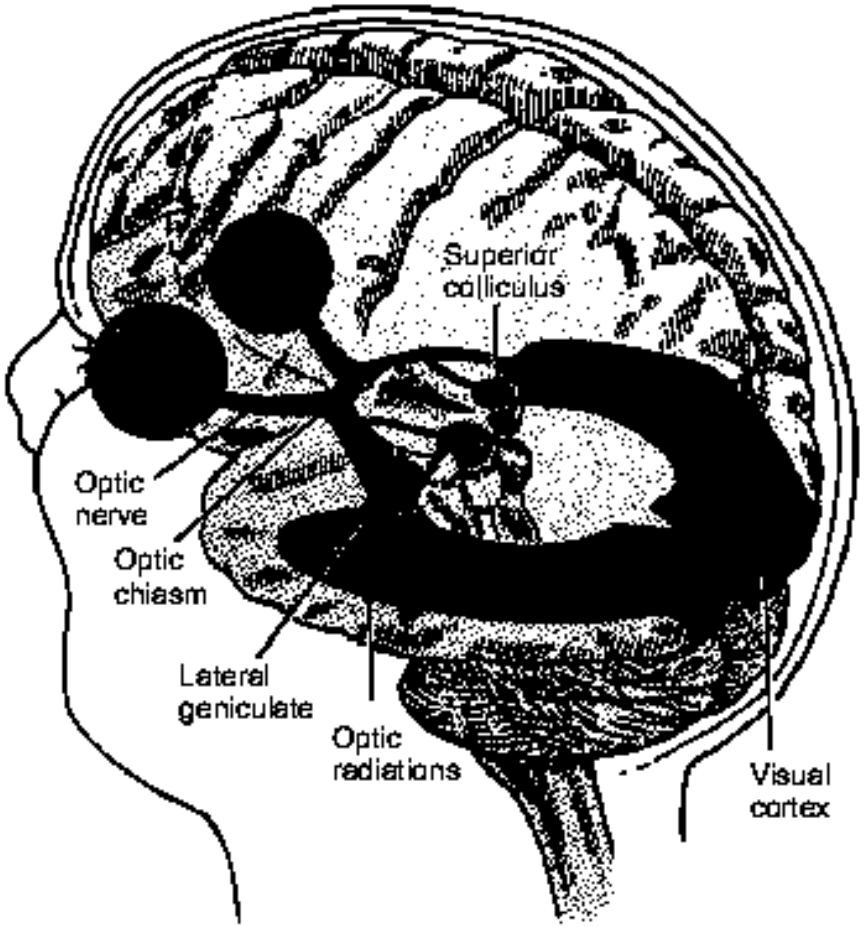


Probability of Detection:



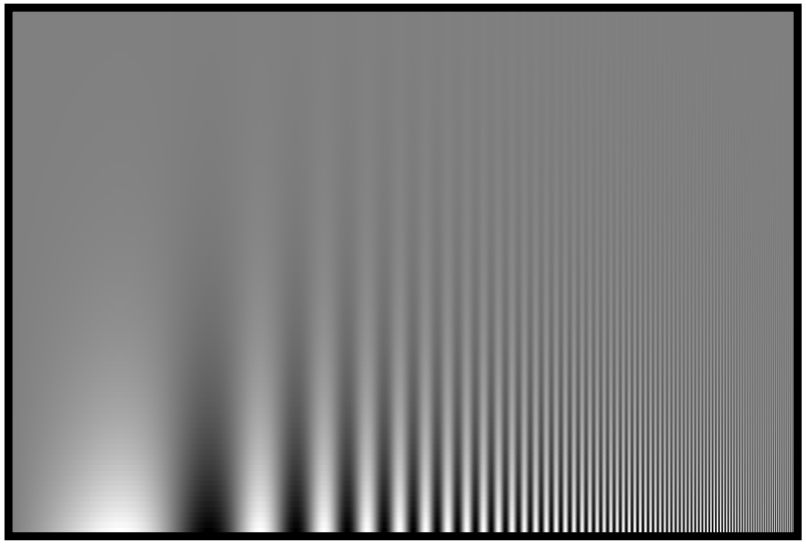
Human Visual System

- Physical structure well established (early vision)
- High-level vision still not fully understood

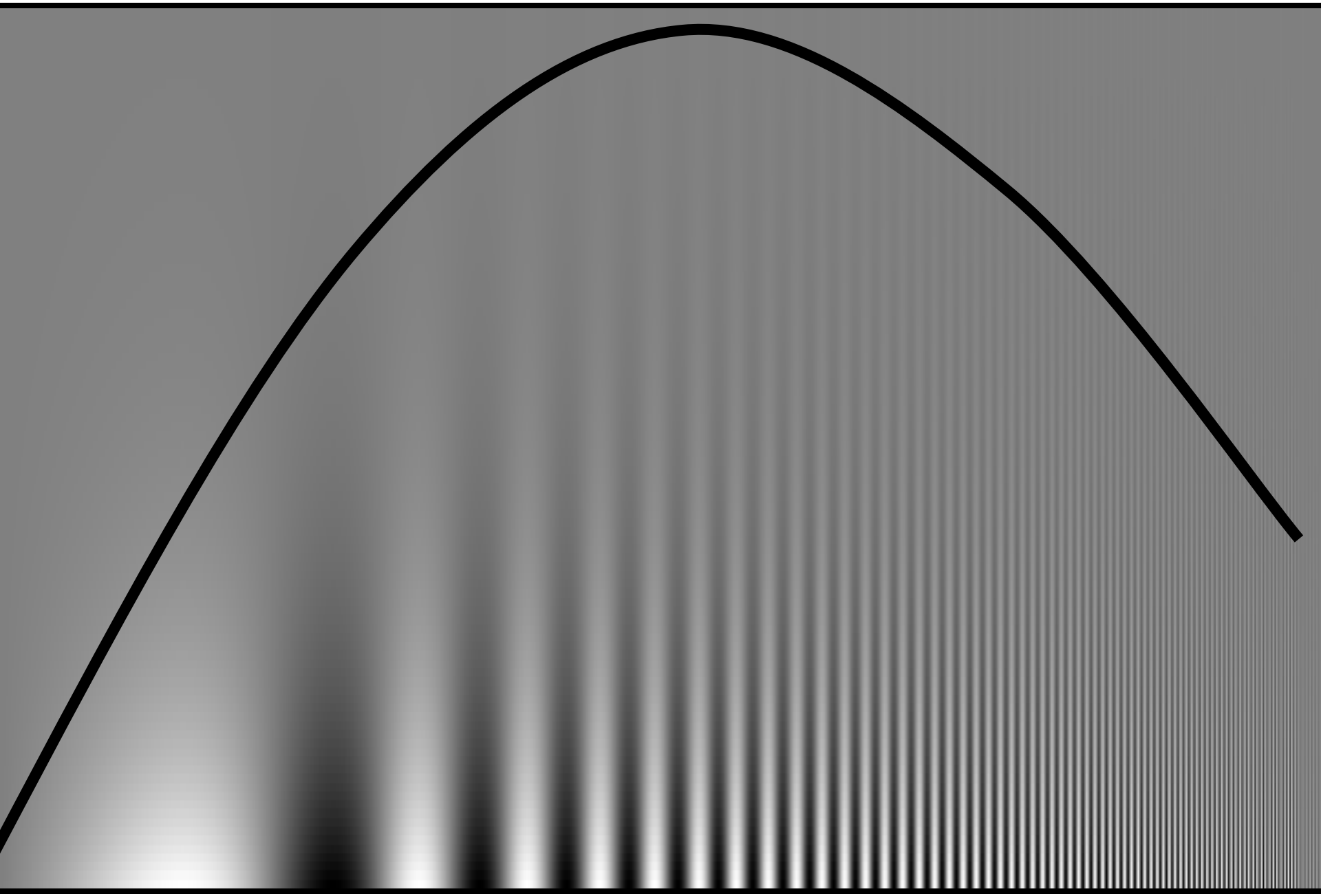


Human Visual System (cont.)

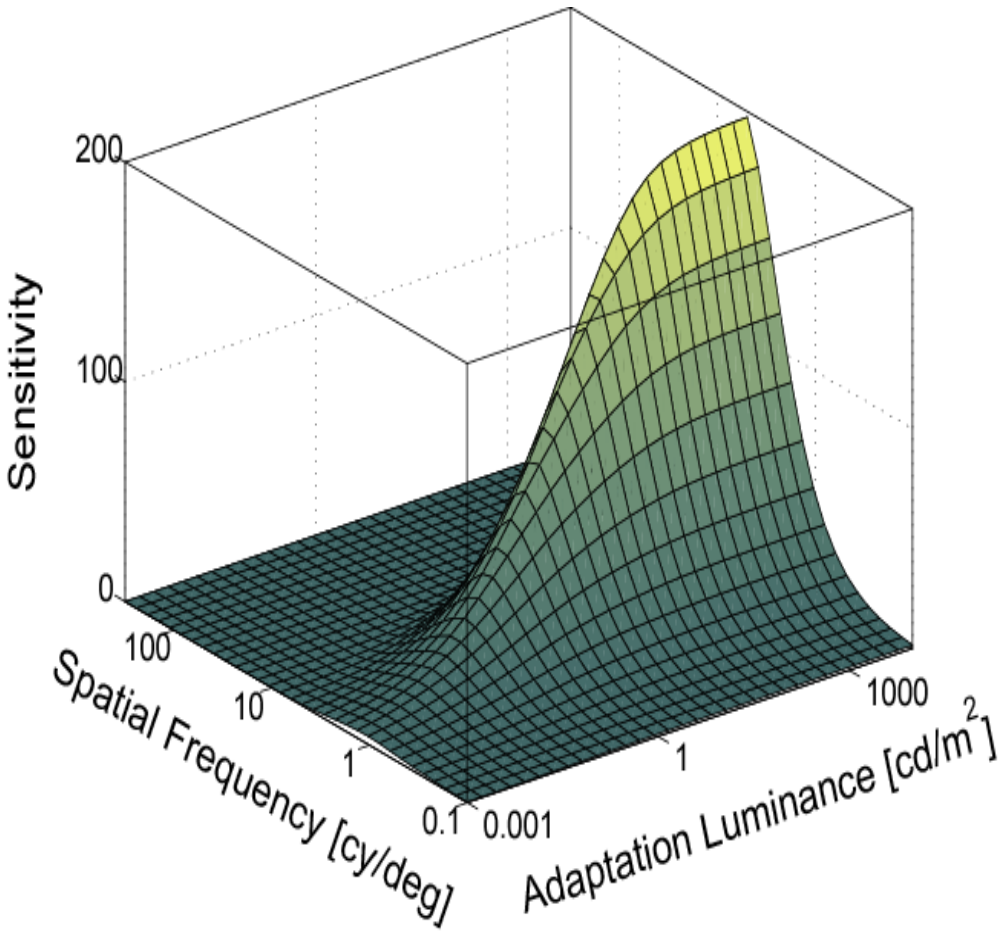
- CSF
  - sensitivity (1/detection threshold) as a function of the spatial frequency
  - depends on
    - spatial frequency
    - adaptation level
    - temporal freq.
    - orientation
    - viewing dist
    - eccentricity, …



[Campbell and Robson 1968]

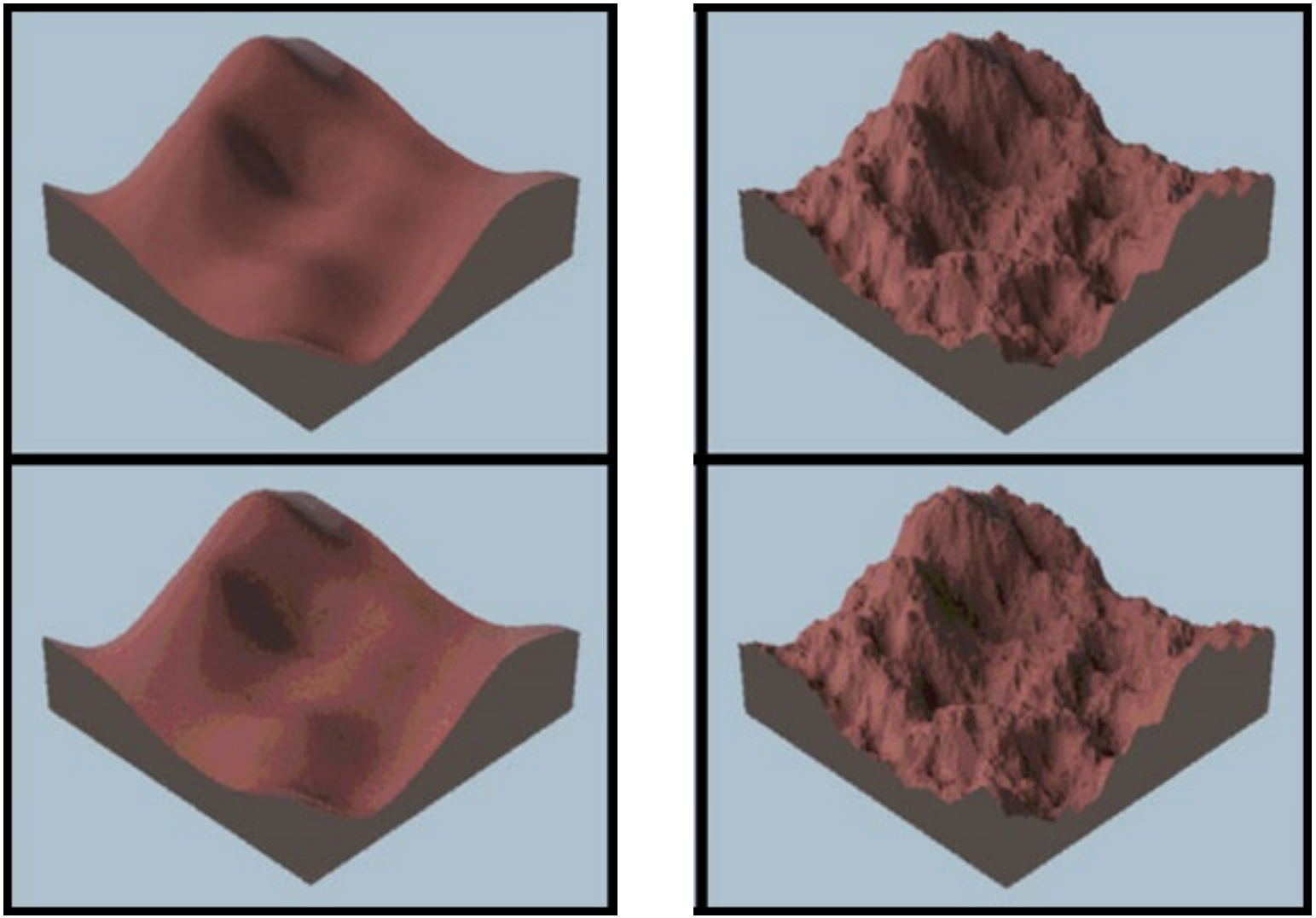


Contrast Sensitivity Function (CSF)



- Steady-state CSF[S]
  - incl. adaptation luminance

HVS – Visual Masking



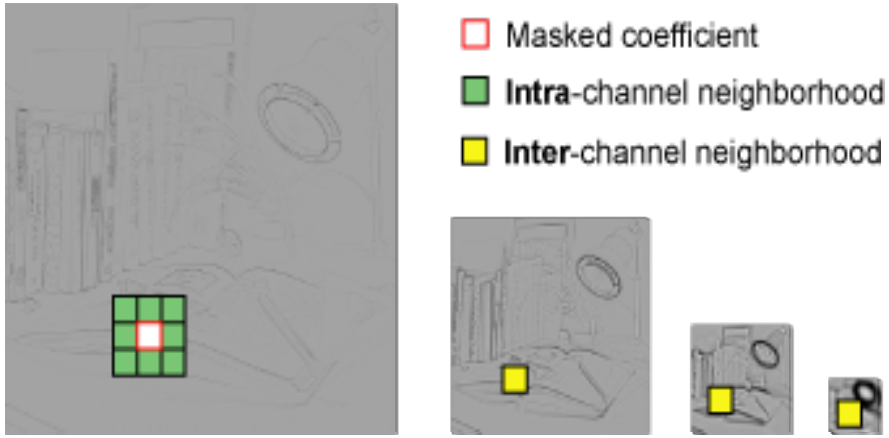
[Ferwerda et al. 97]

Visual Masking

Loss of sensitivity to a signal
with the presence of a “similar” signal “nearby”.

Modeling Visual Masking

- **Example:** JPEG's pointwise extended masking:

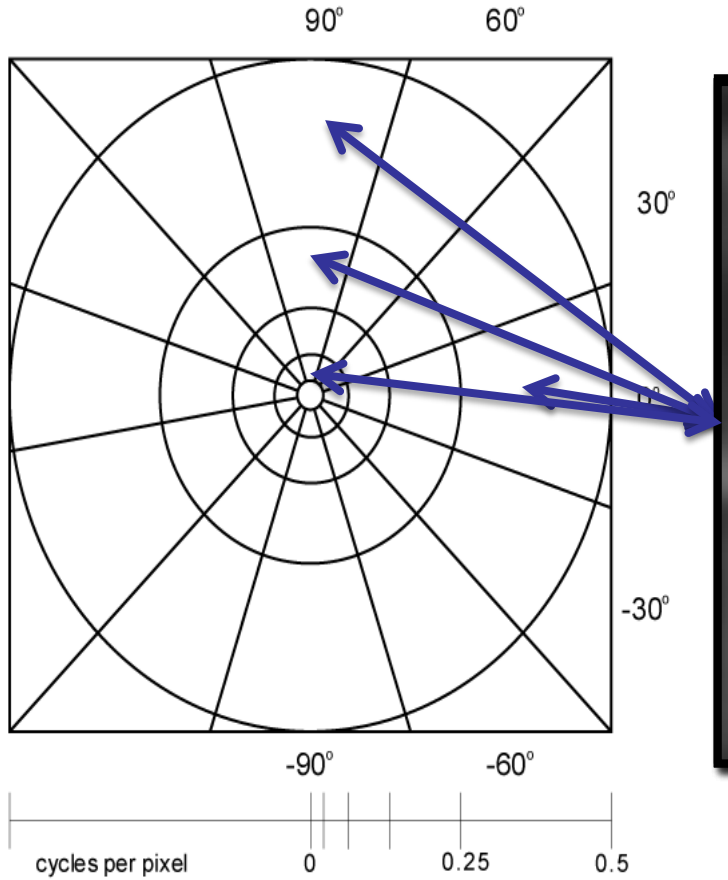


$$R = \frac{sign(C')|C'|^{0.5}}{(1 + \sum_K |C'_k|^{0.2})}$$

C': Normalized Contrast

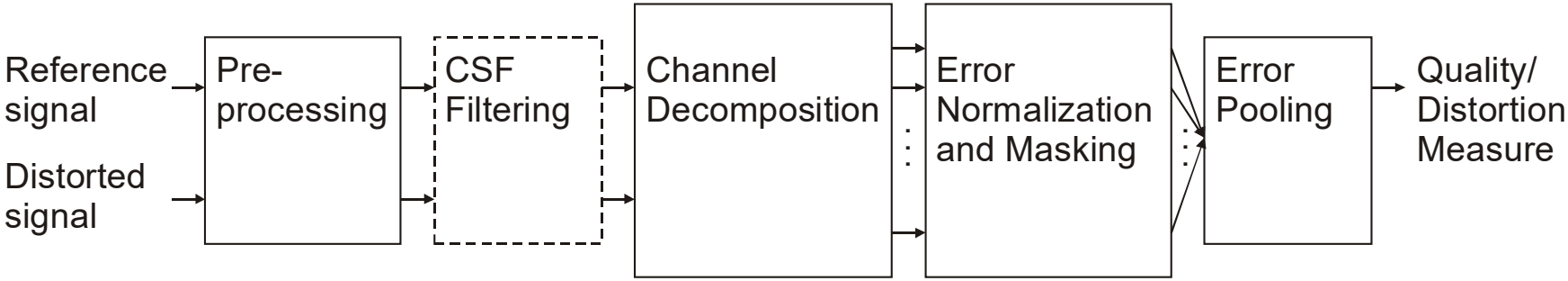
Modeling Visual Masking - Visual Channels

Cortex Transform



Error Sensitivity Based Approaches

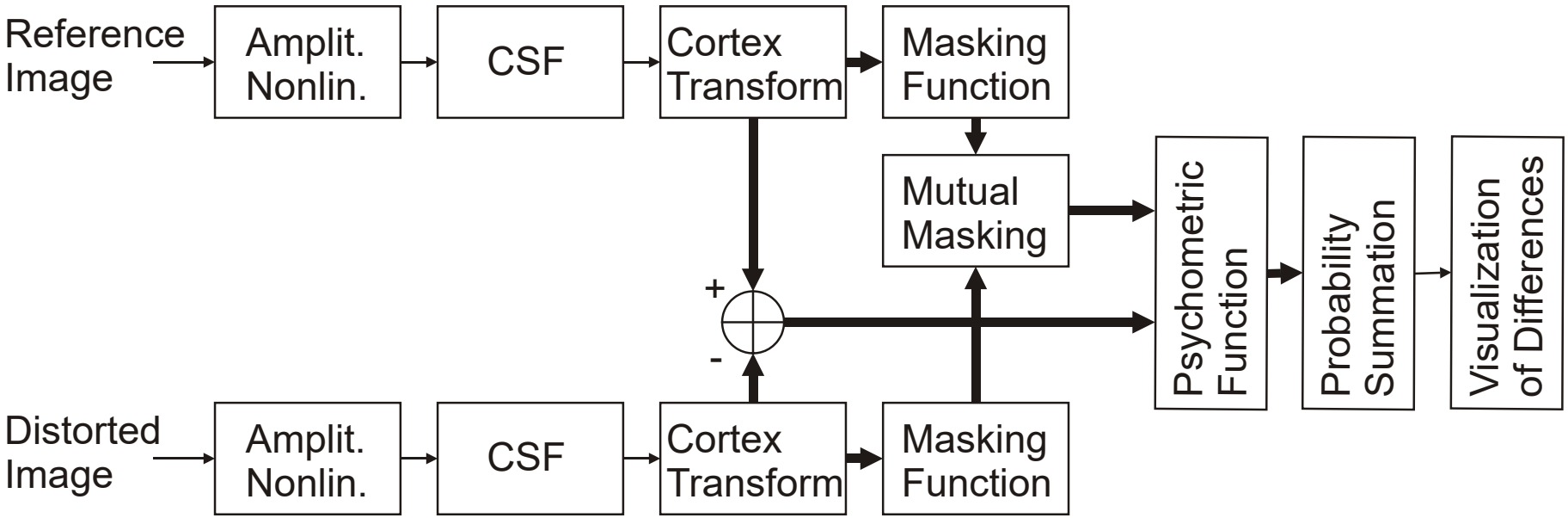
- General framework



- Visible Differences Predictor [Daly93]
- Perceptual Image Distortion [Teo, Heeger 94]
- Visual Discrimination Model [Lubin 95]
- Gabor pyramid model [Taylor et al. 97]
- WVDP [Bradley 99]
- HDRVDP3 [Mantiuk et al. 05, Mantiuk et al. 11]

Error Sensitivity Based Approach

- Visible Differences Predictor (VDP) [Daly 93]

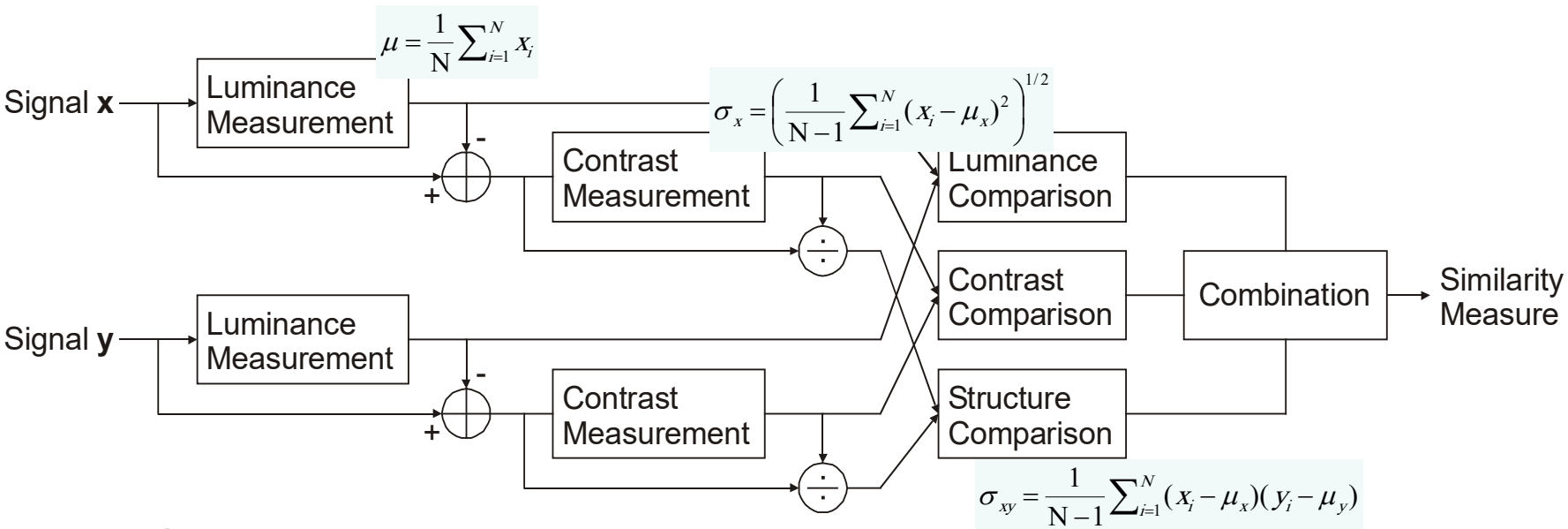


- **Threshold** sensitivity
- Early vision modeling
- Visual Masking

Overview

- Introduction to Objective Quality Assessment
- Image Quality Assessment
  - HVS-based Metrics (bottom-up)
  - Structural Similarity (top-down)
  - Data-driven Approaches (top-down)
- Video Quality Assessment

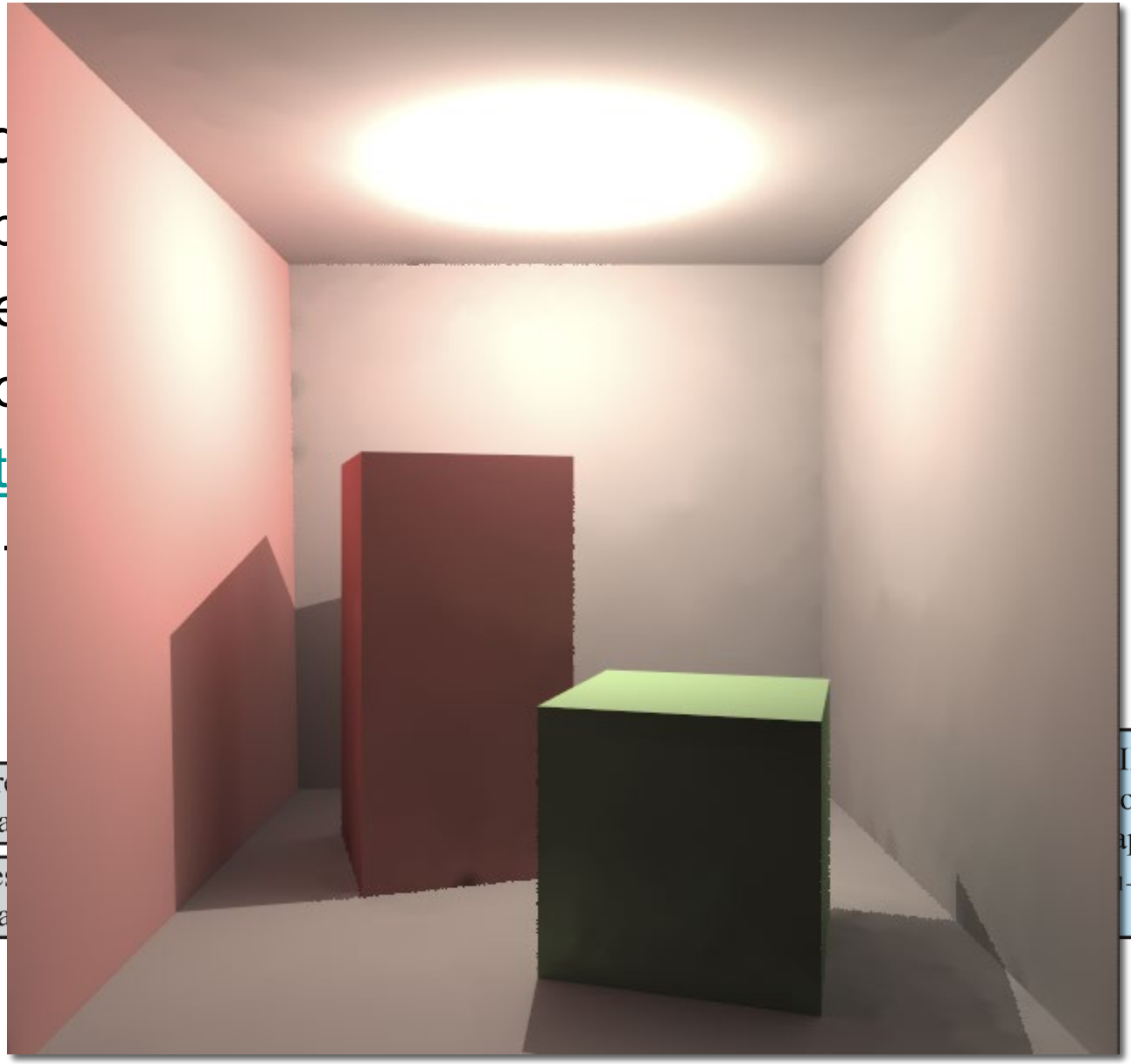
Structural Similarity-Based Approaches



- UQI [Wang 02]
- SSIM [Wang 04]
- M-SSIM [Wang et al. 04]
- Multidimensional Quality Measure Using SVD [Shnayderman 04]
- FLIP [Andersson et al. 20]

FLIP

- [Anc
  - sp
  - pe
  - no
  - ht
  - C



refer
ima

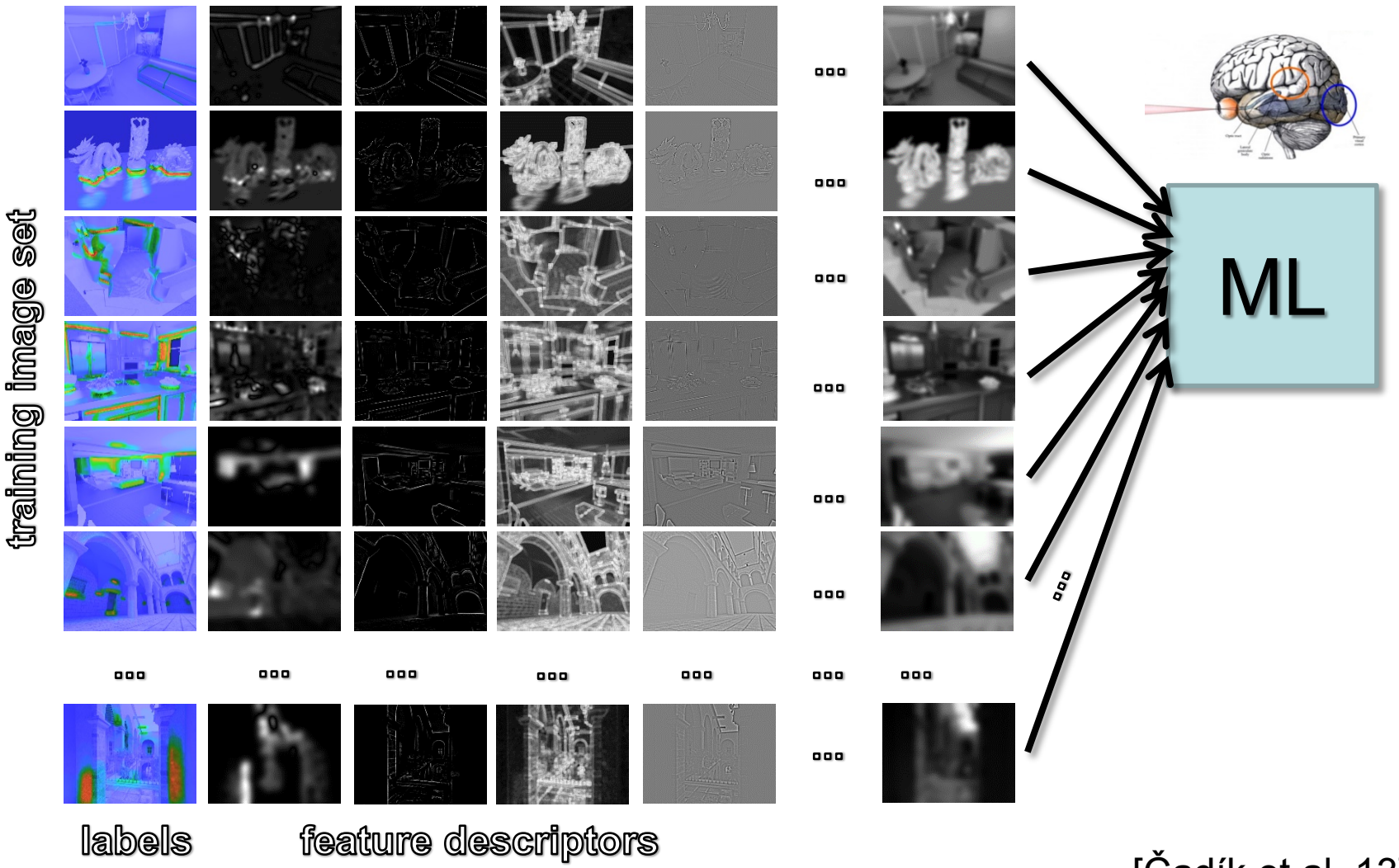
te
ima

IP
or
ap
$-\Delta E_f$

Overview

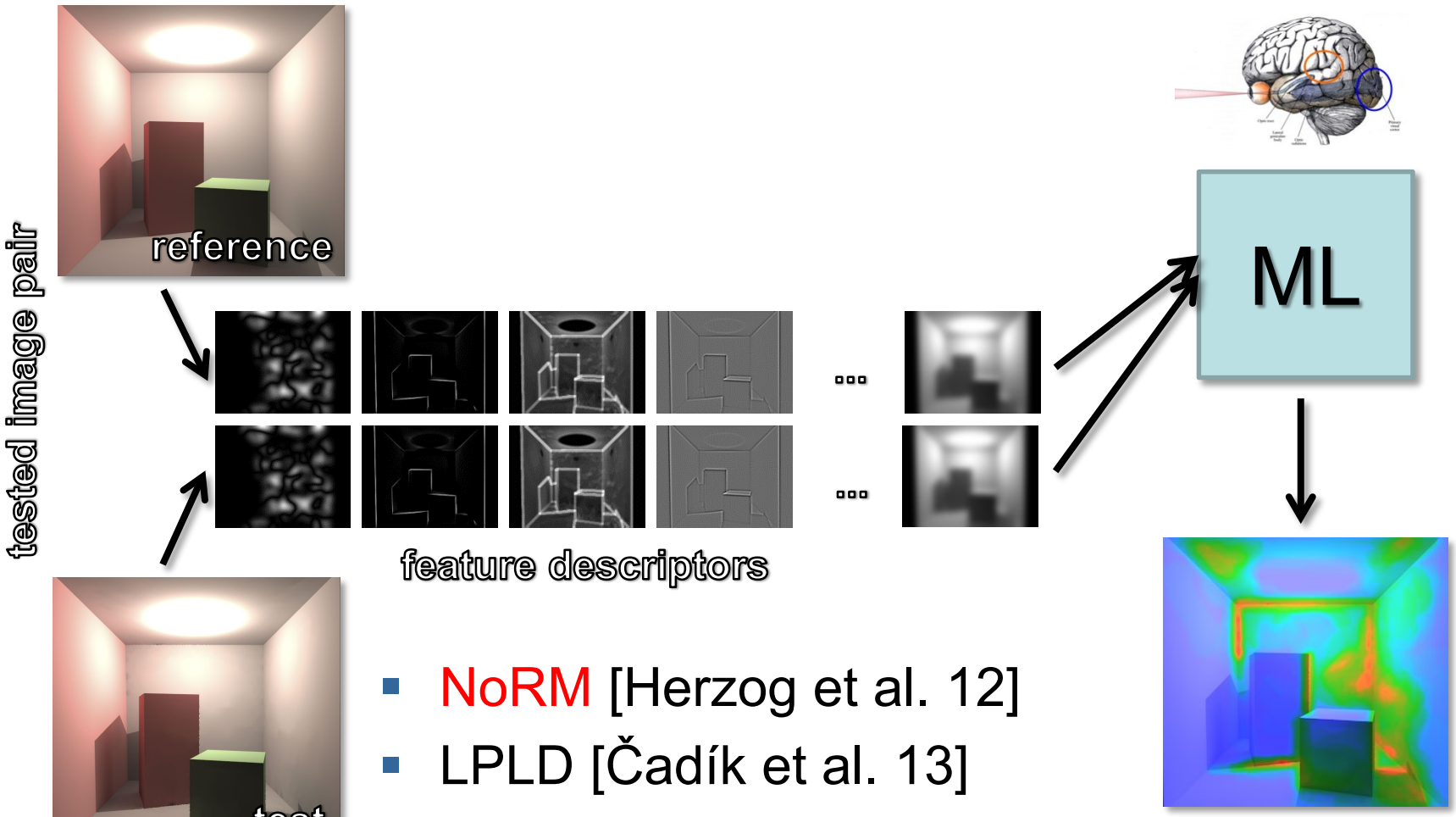
- Introduction to Objective Quality Assessment
- Image Quality Assessment
  - HVS-based Metrics (bottom-up)
  - Structural Similarity (top-down)
  - Data-driven Approaches (top-down)
- Video Quality Assessment

Supervised Learning – Training Phase



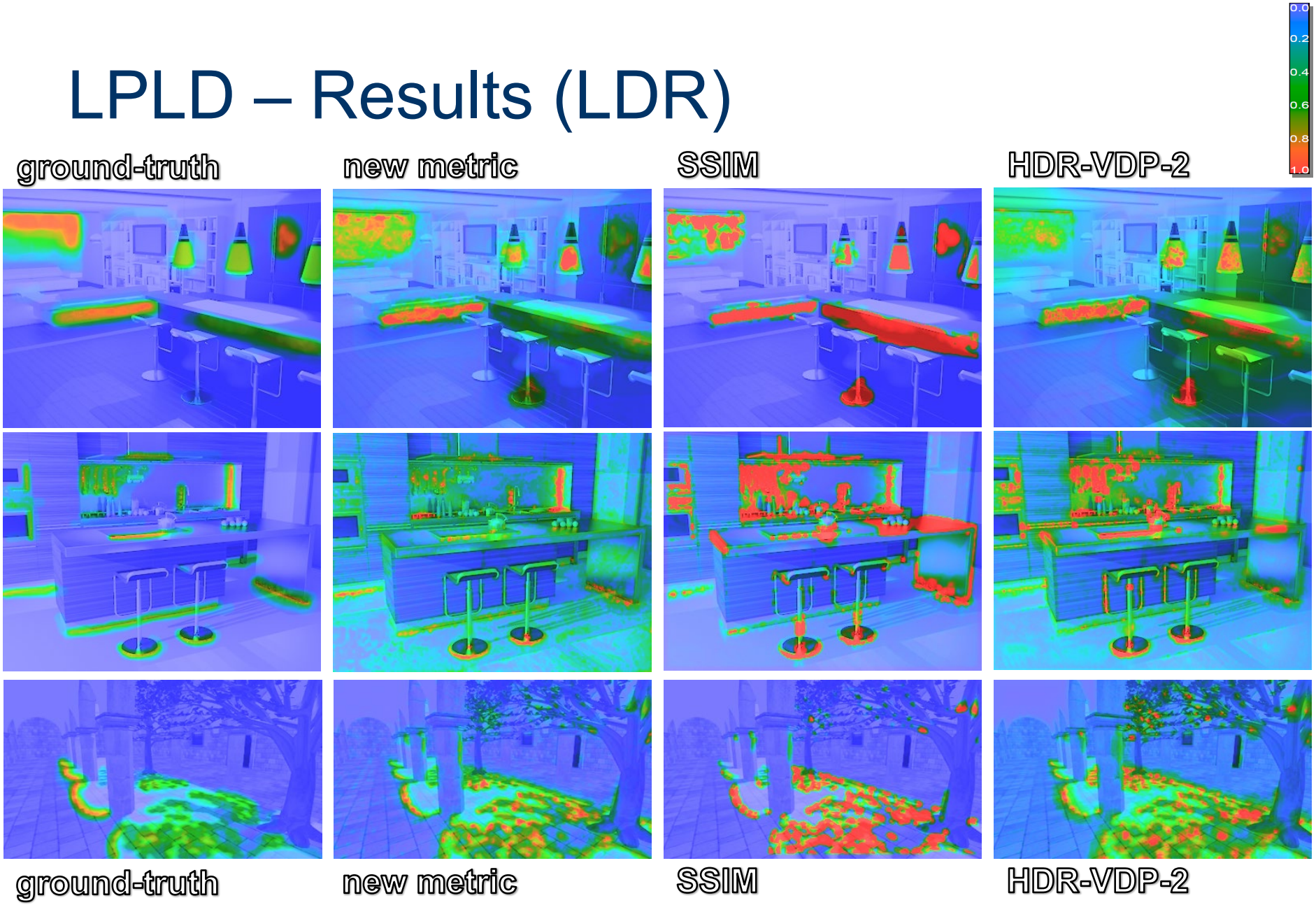
[Čadík et al. 13]

Supervised Learning – Prediction



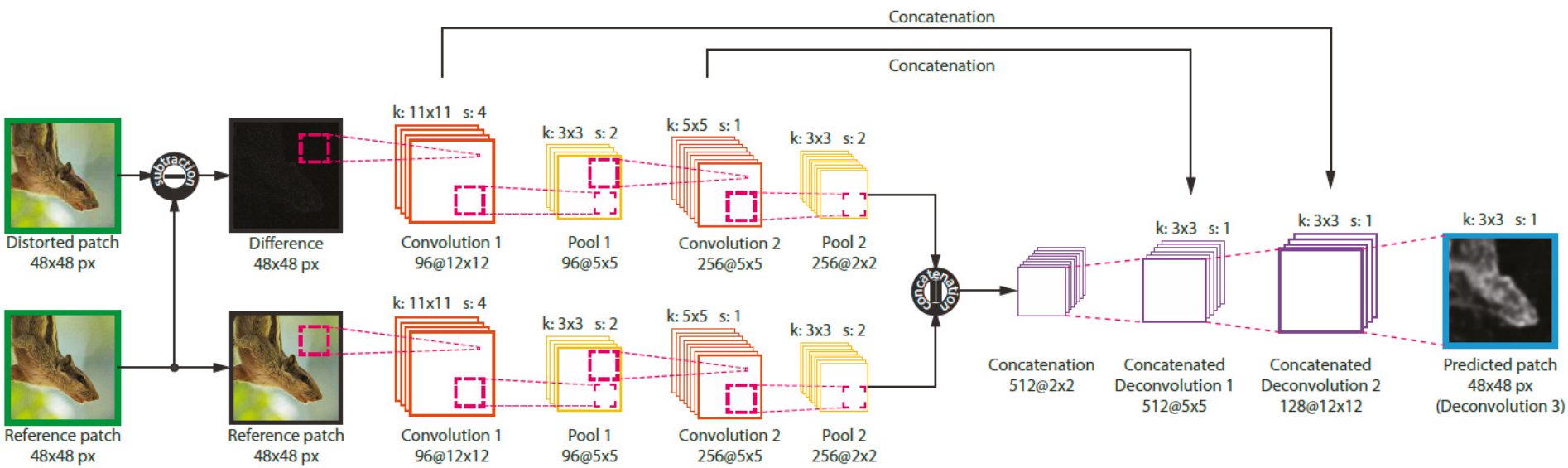
- NoRM [Herzog et al. 12]
- LPLD [Čadík et al. 13]
- Image completion metric [Kopf et al. 14]
- CNN_visibility_metric [Wolski et al. 18]

LPLD – Results (LDR)



CNN Visibility Metric

- [Wolski et al. 18]
  - two-branch fully convoluted architecture
  - separate weights for each branch
  - fine tuning (AlexNet weights)
  - viewing distance, luminance level not considered
  - https://github.com/Chuudy/CNN_visibility_metric



No-Reference Image Quality Metrics

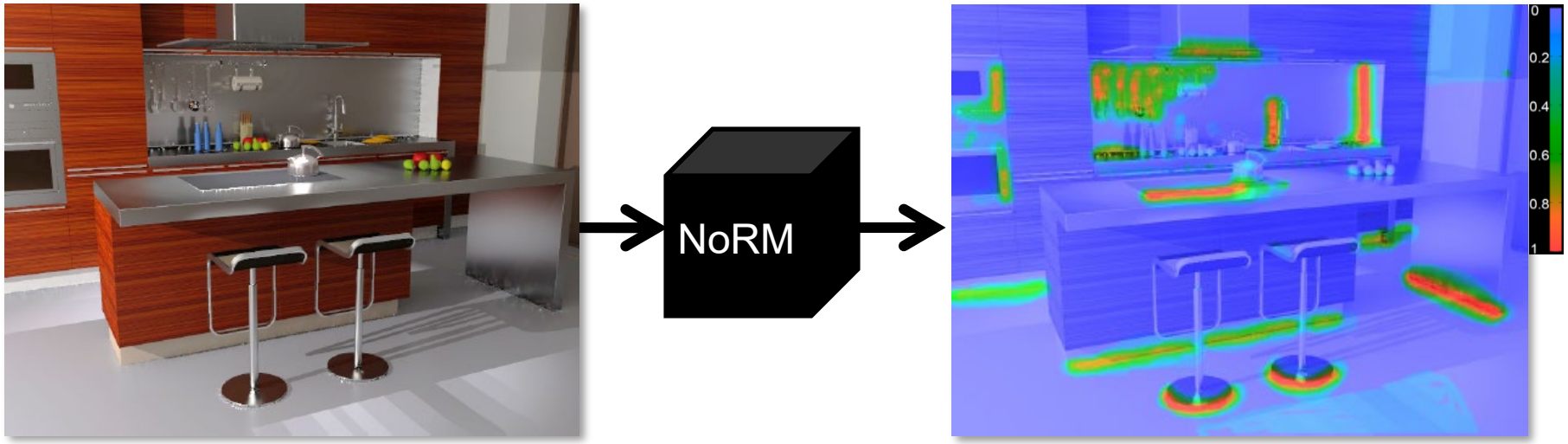
- Detecting blockiness in JPEG/MPEG
  - [Wang & Bovik ’06, Wu & Rao ‘05]
- Blurriness measures
  - [Liu et al. SPIE ’11, Chen et al. SPIE ’11]
- Detection / removal of false contours (color quantization)
  - [Daly S. & Feng SPIE ’04]
- Natural image statistics no-ref. QA
  - [Sheikh et al. ‘05, Jpeg2000]

- “Real” no-ref metric
  - **NoRM [Herzog et al. 12]**

NoRM: No-Reference Metric

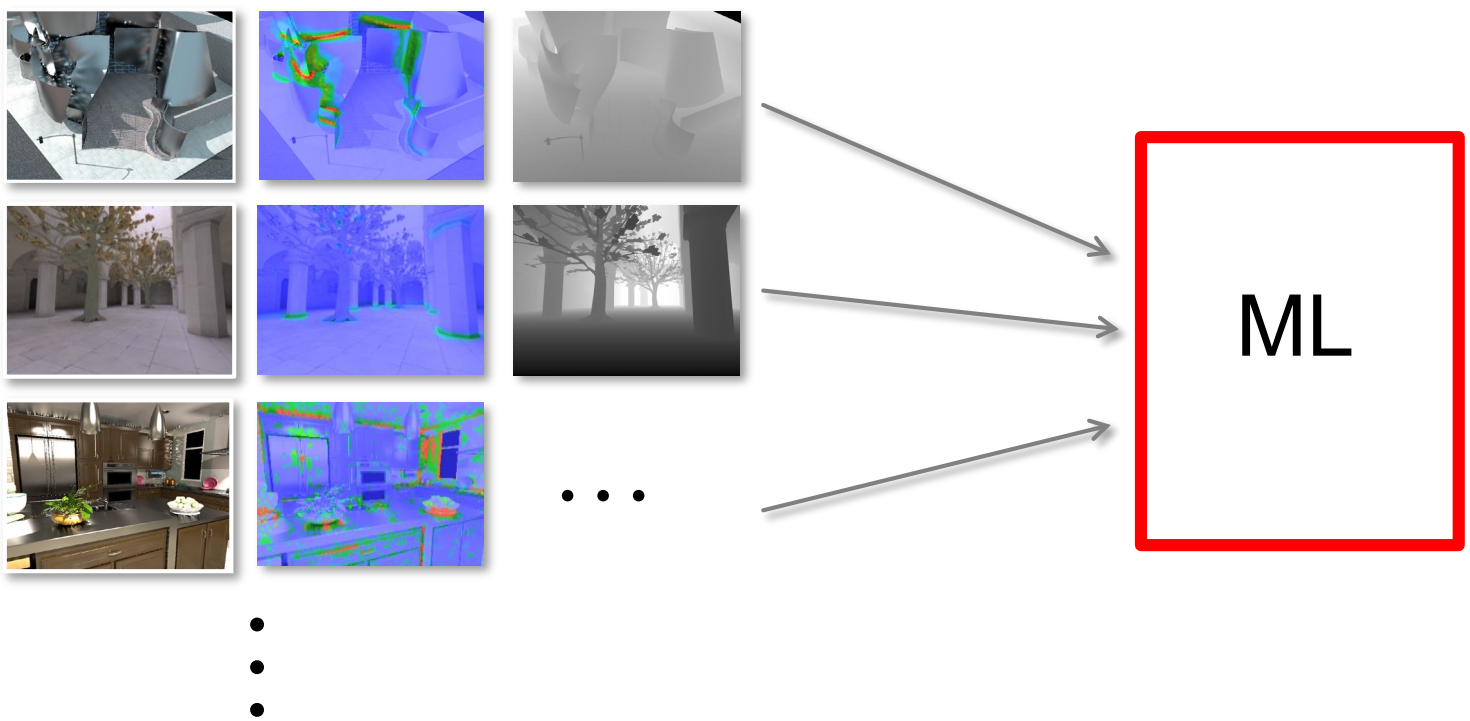
[Herzog et al. 2012]

- Input: distorted image/video frame (no reference)
- Output: map of distortions (possibly perceptually weighted)

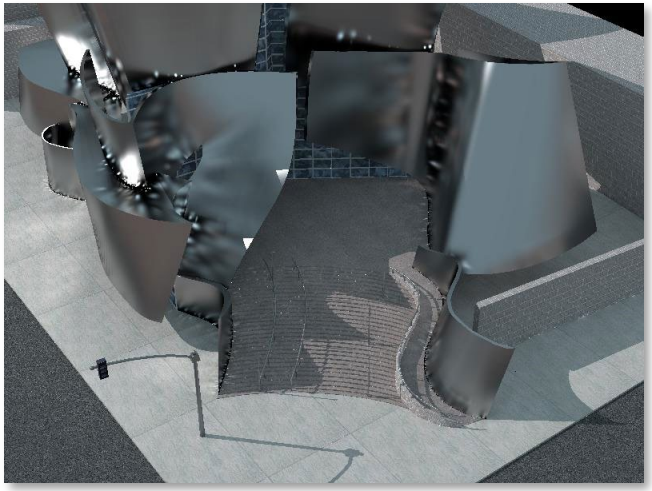


Data-Driven No-Reference IQM

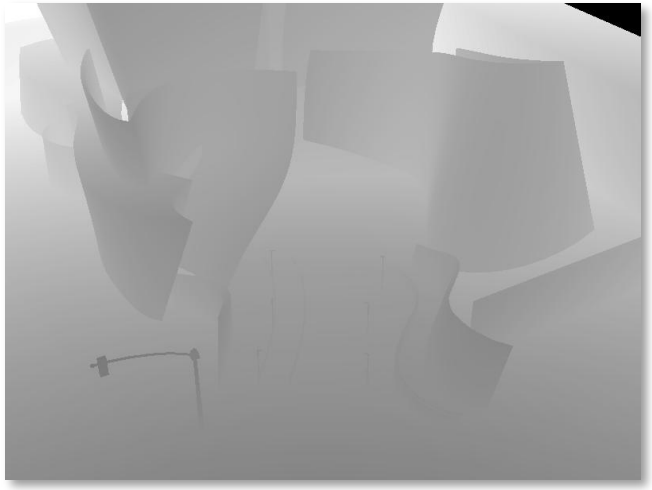
- Feature descriptors (various information available)
- Distortion maps (possibly real subjective data)
- Depth + 3D related information



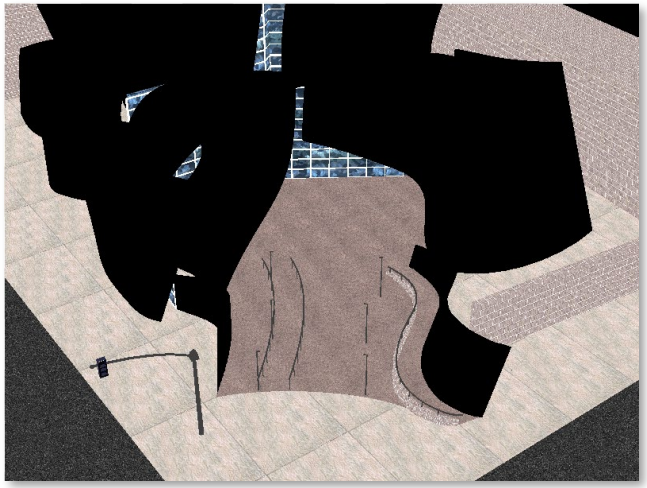
Rendering Output – Classification Input



HDR (LDR) color image
(may contain noise)



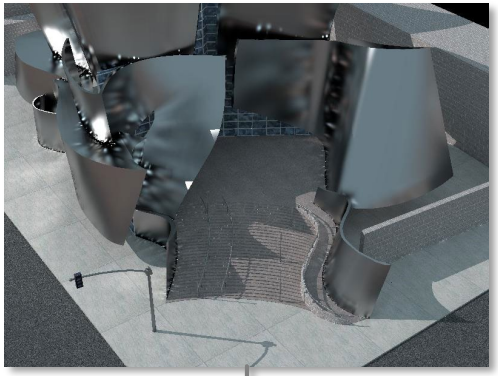
depth buffer
(in high precision, no noise)



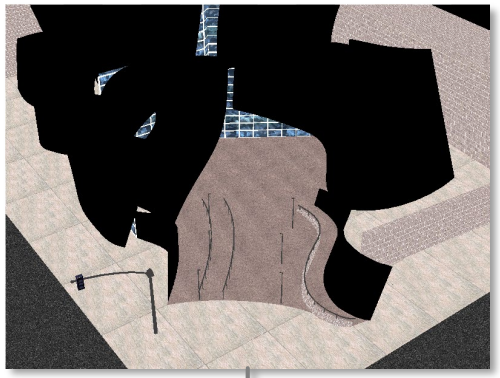
diffuse texture buffer

Computation of Additional Input Data

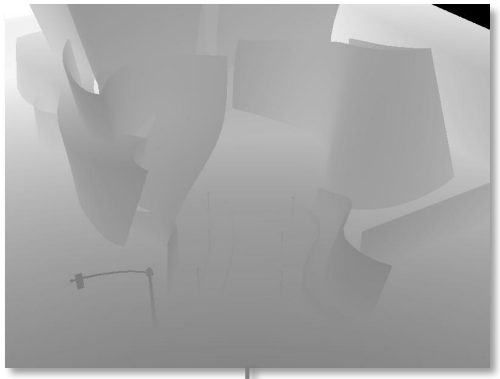
color (pixel radiance)



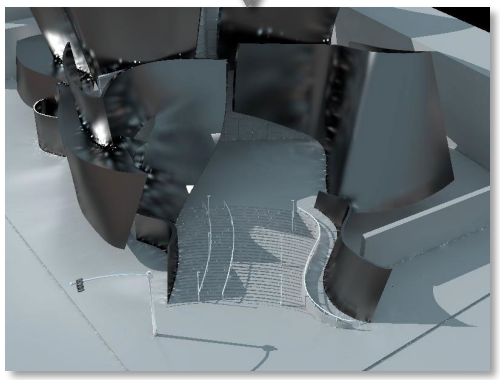
textures



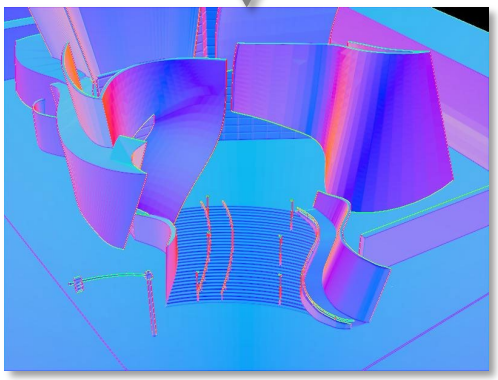
depth



/mat



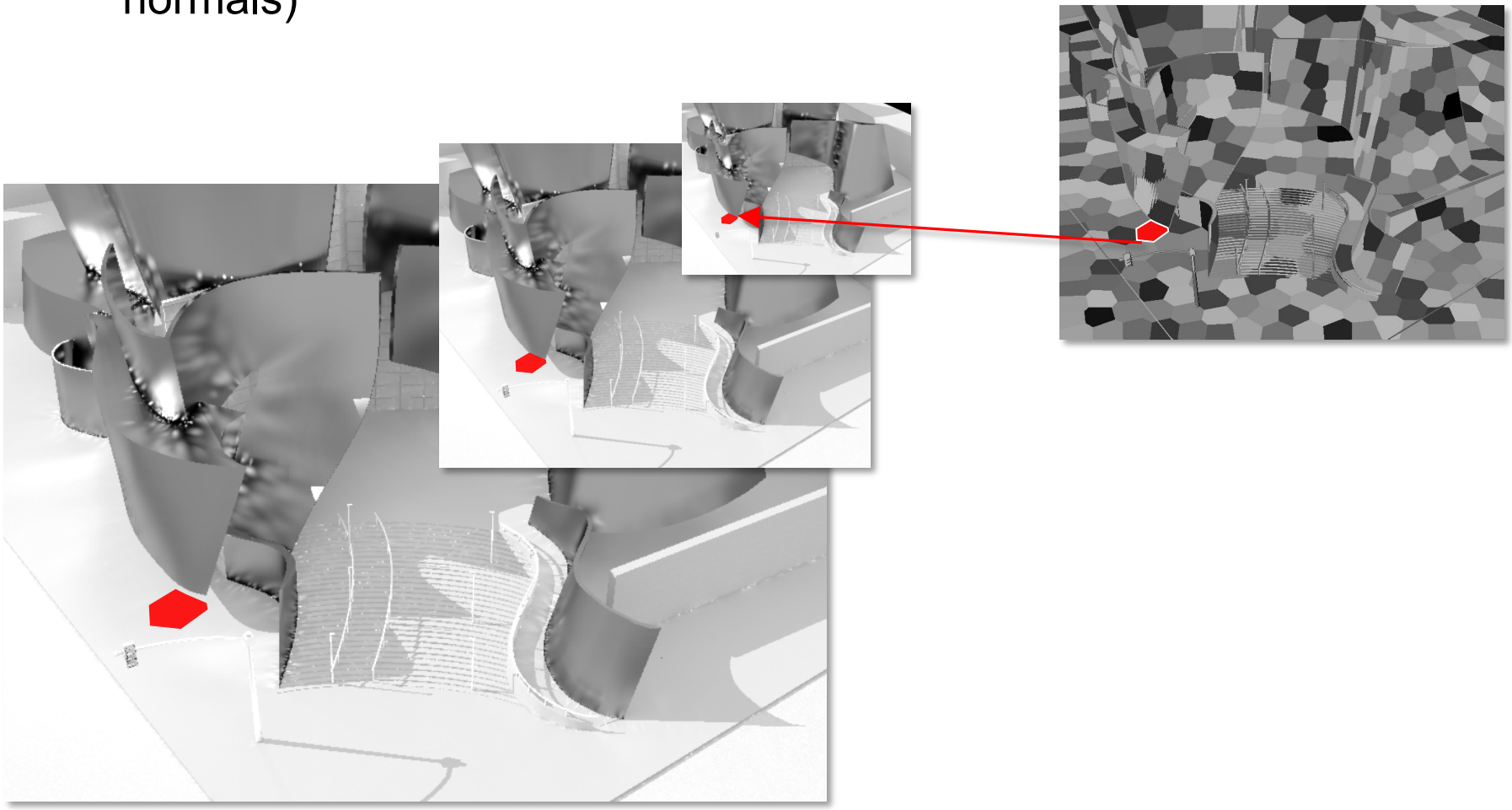
lighting (irradiance)



surface normals
(computed from depth)

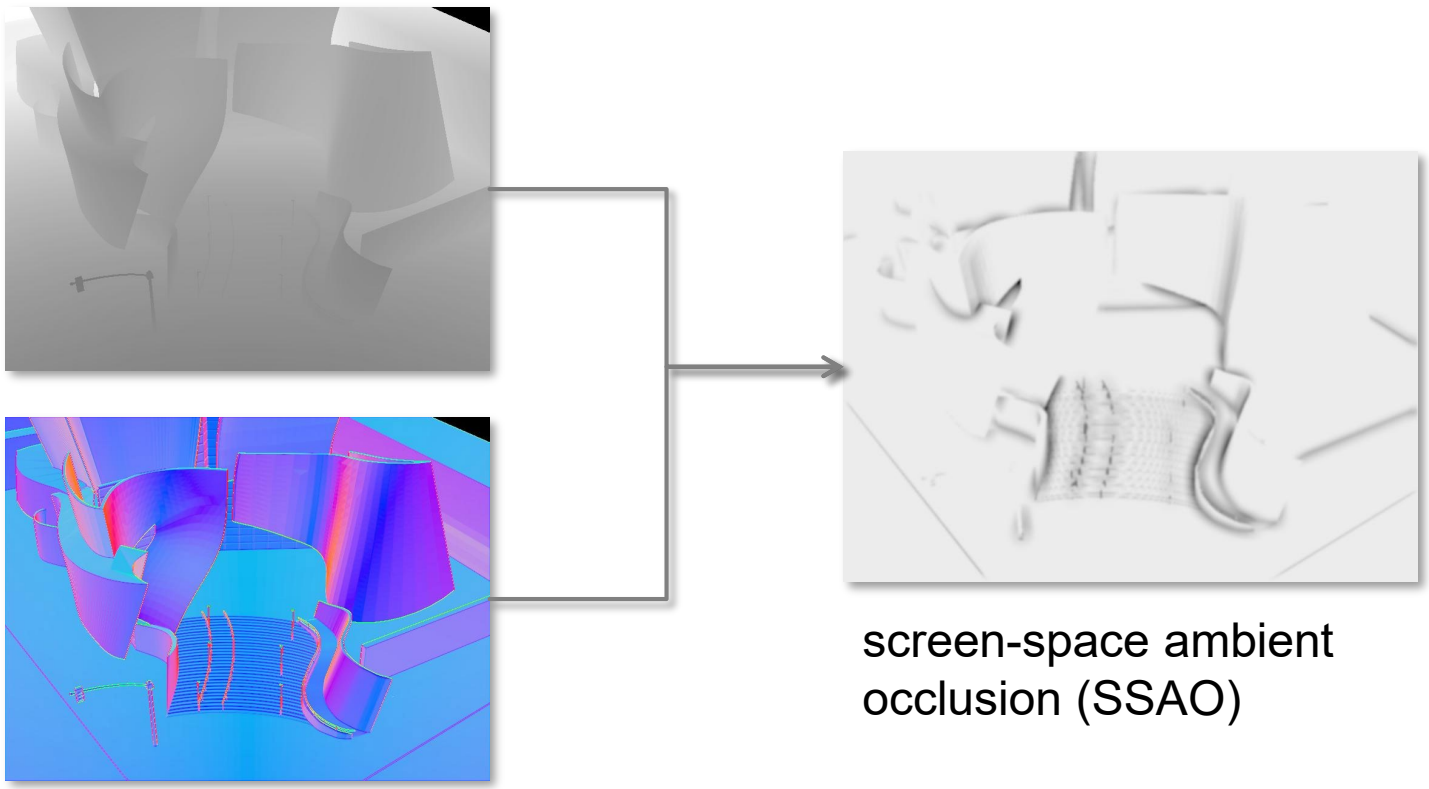
3D Features: Local First-order Statistics

- compute mean, variance, skewness, kurtosis in each segment at different scales of the grey-scale image pyramid (also for depth, normals)

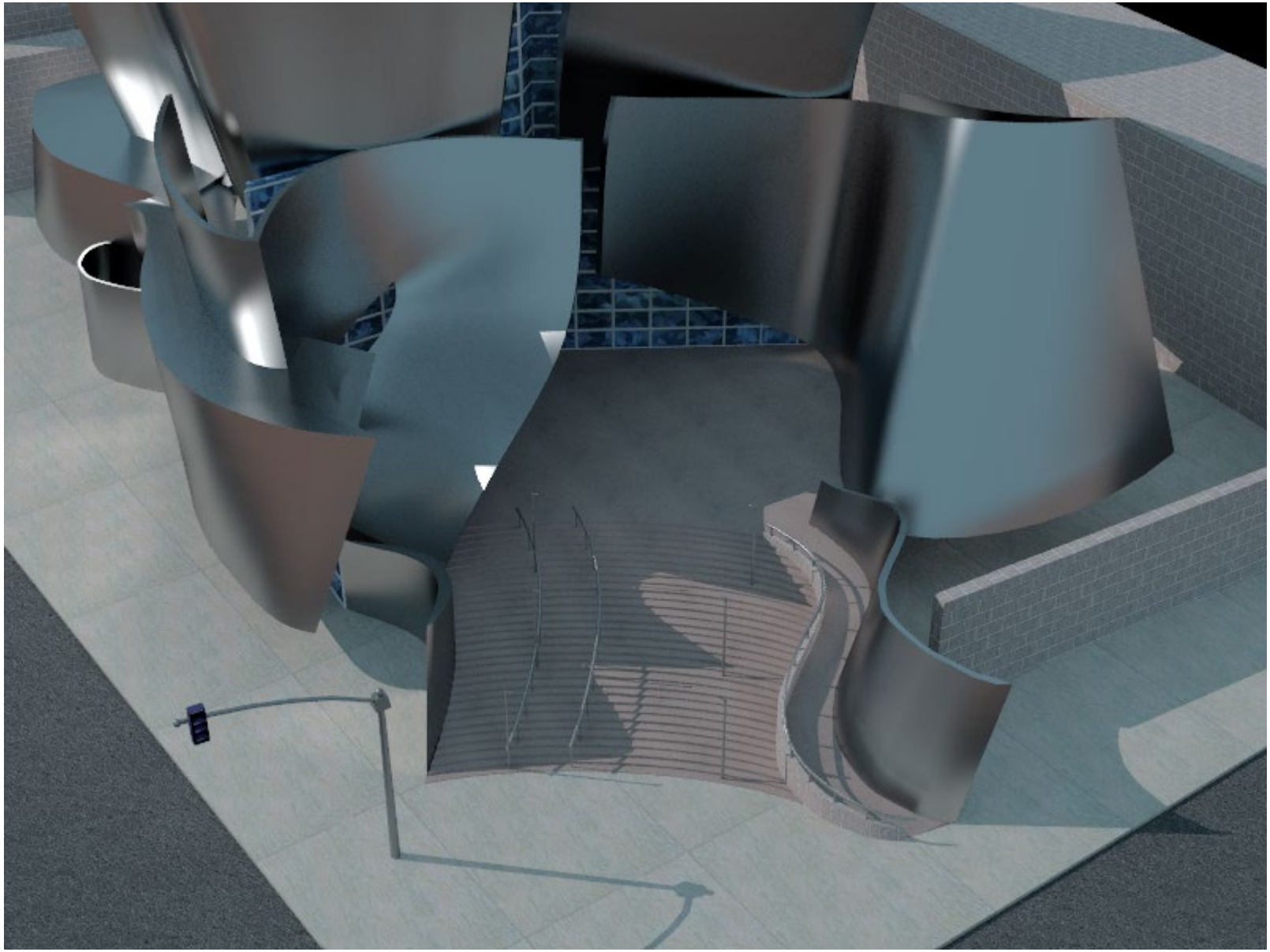


3D Features: Ambient Occlusion

- given depth extract approx. ambient occlusion per pixel (distance to nearest occluder)



screen-space ambient occlusion (SSAO)



REFERENCE IMAGE

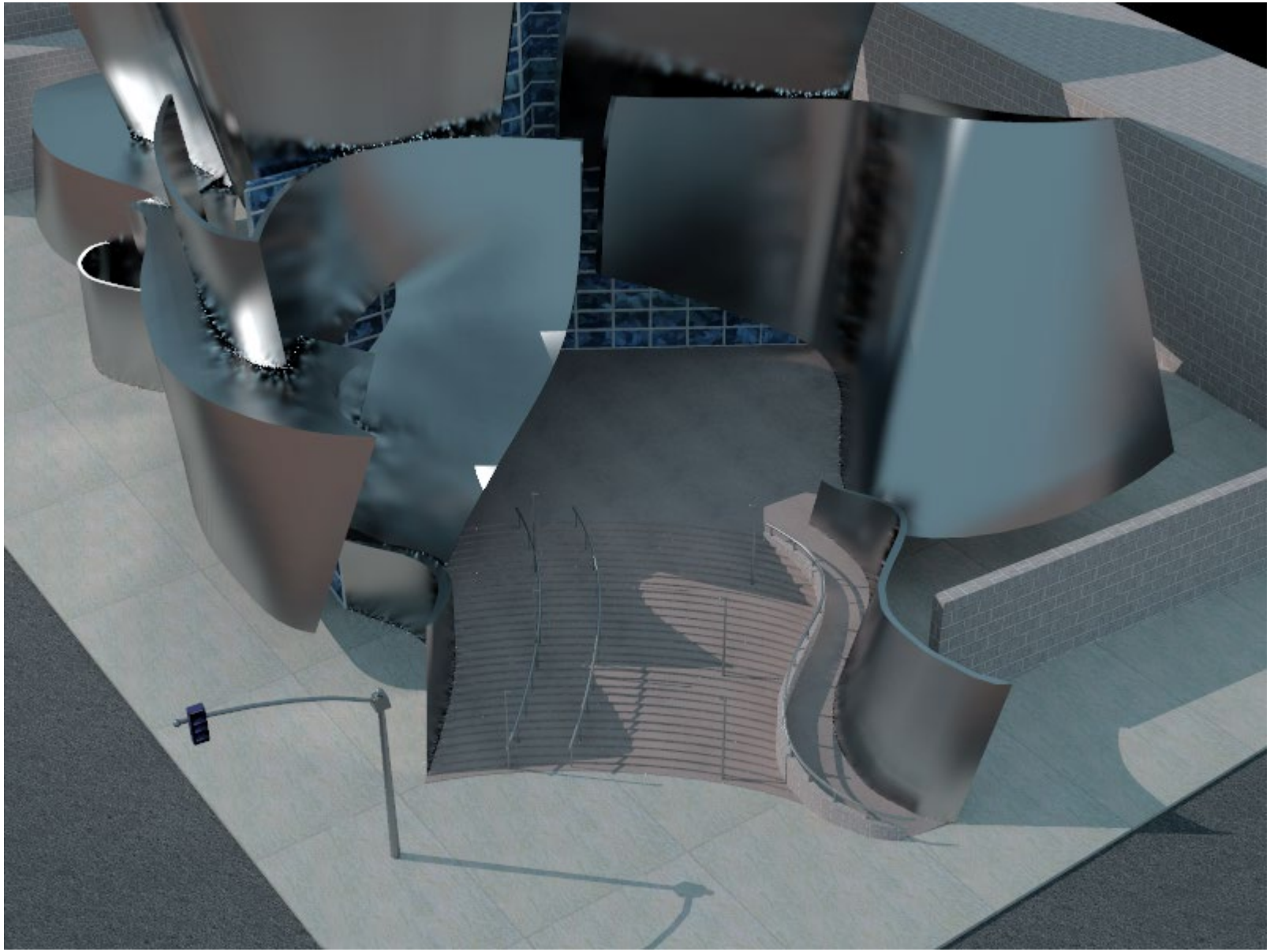
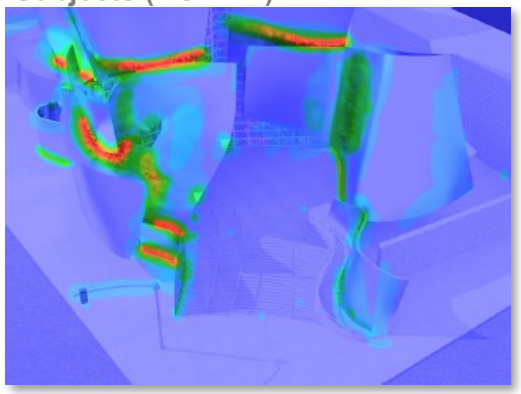


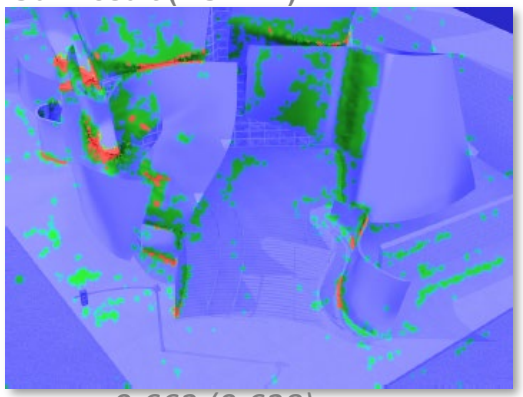
IMAGE WITH ARTIFACTS

Results (VPL noise)

Subjects (NO REF)

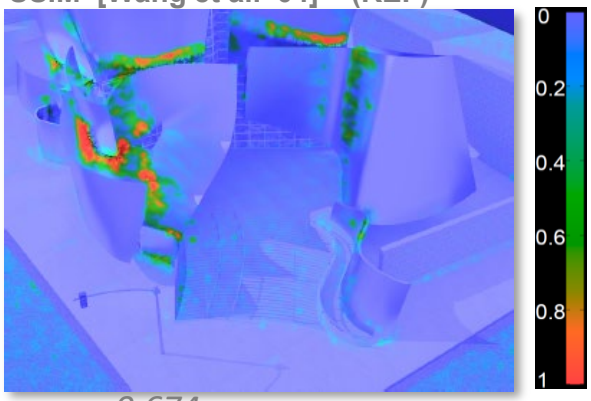


Our Result (NO REF)



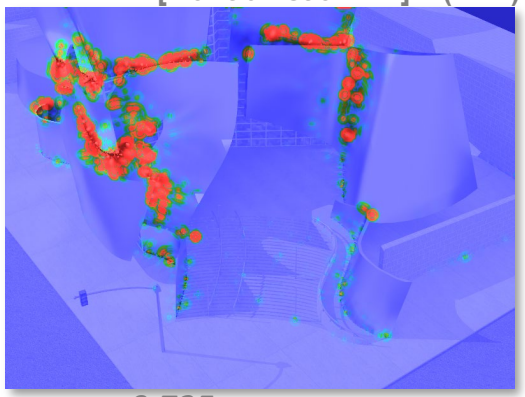
corr = 0.662 (0.628)

SSIM [Wang et al. '04] – (REF)



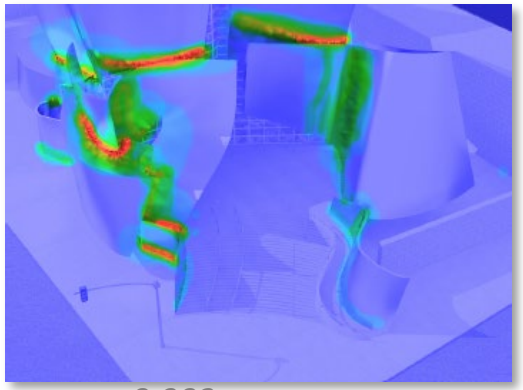
corr = 0.674

HDRVDP2 [Mantiuk et al. '11] – (REF)



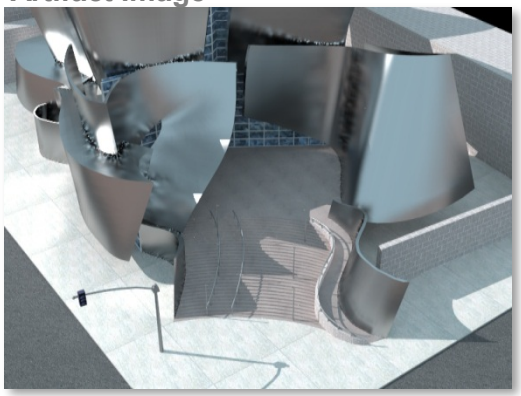
corr = 0.725

Subjects (REF)



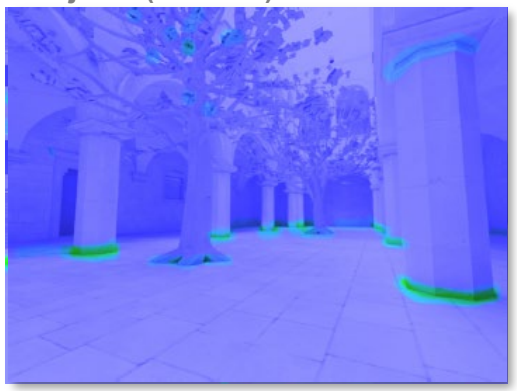
corr = 0.903

Artifact Image

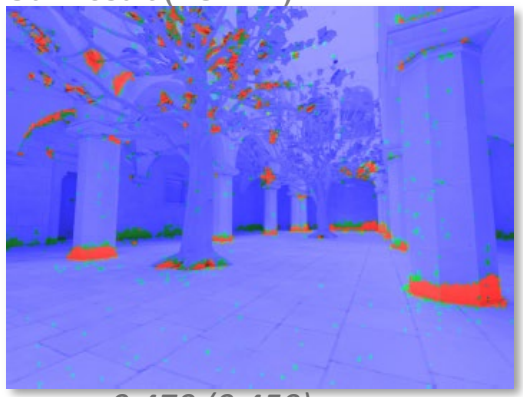


Results (VPL clamping)

Subjects (NO REF)

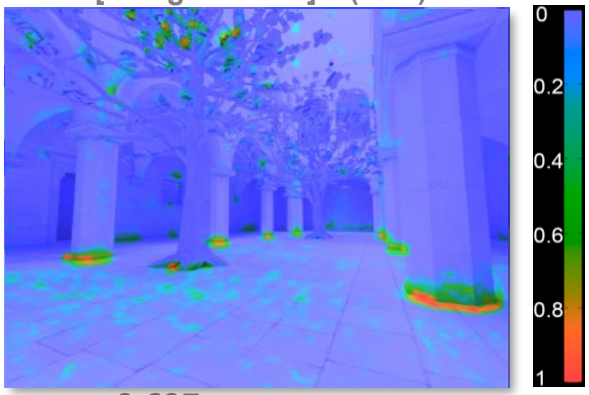


Our Result (NO REF)



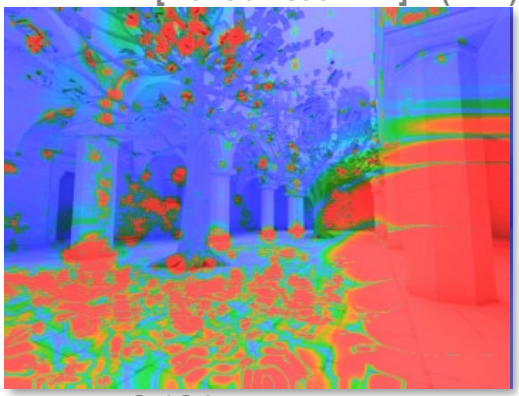
corr = 0.470 (0.450)

SSIM [Wang et al. '04] – (REF)



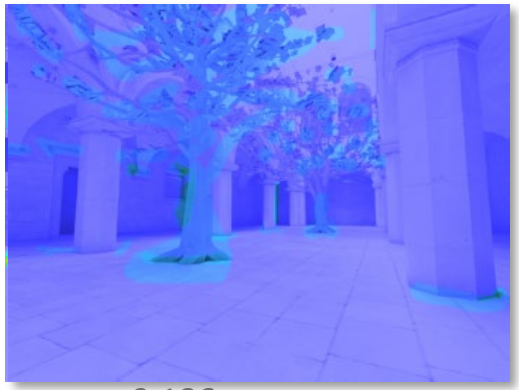
corr = 0.637

HDRVDP2 [Mantiuk et al. '11] – (REF)



corr = 0.134

Subjects (REF)

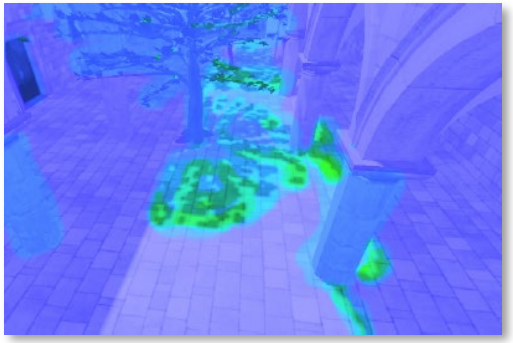


corr = 0.186

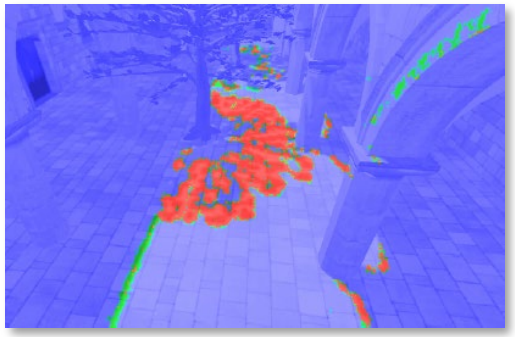
Artifact Image

Results (Shadow aliasing)

Subjects (NO REF)

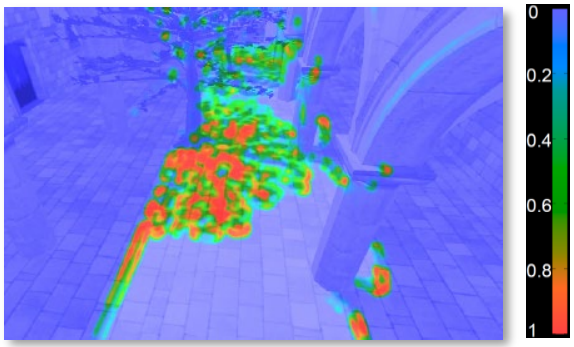


Our Result (NO REF)



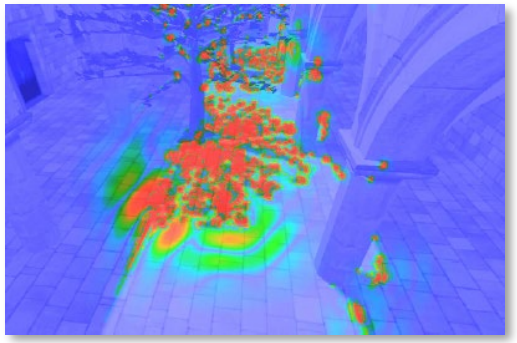
corr = 0.767 (0.638)

SSIM [Wang et al. '04] – (REF)



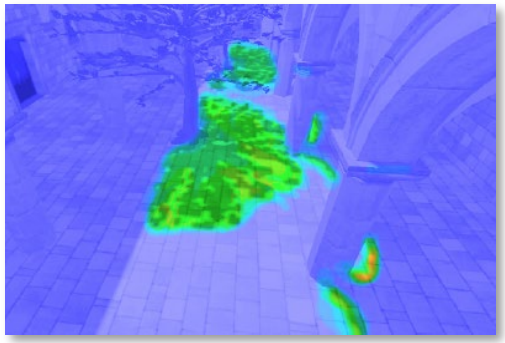
corr = 0.742

HDRVDP2 [Mantiuk et al. '11] – (REF)



corr = 0.669

Subjects (REF)



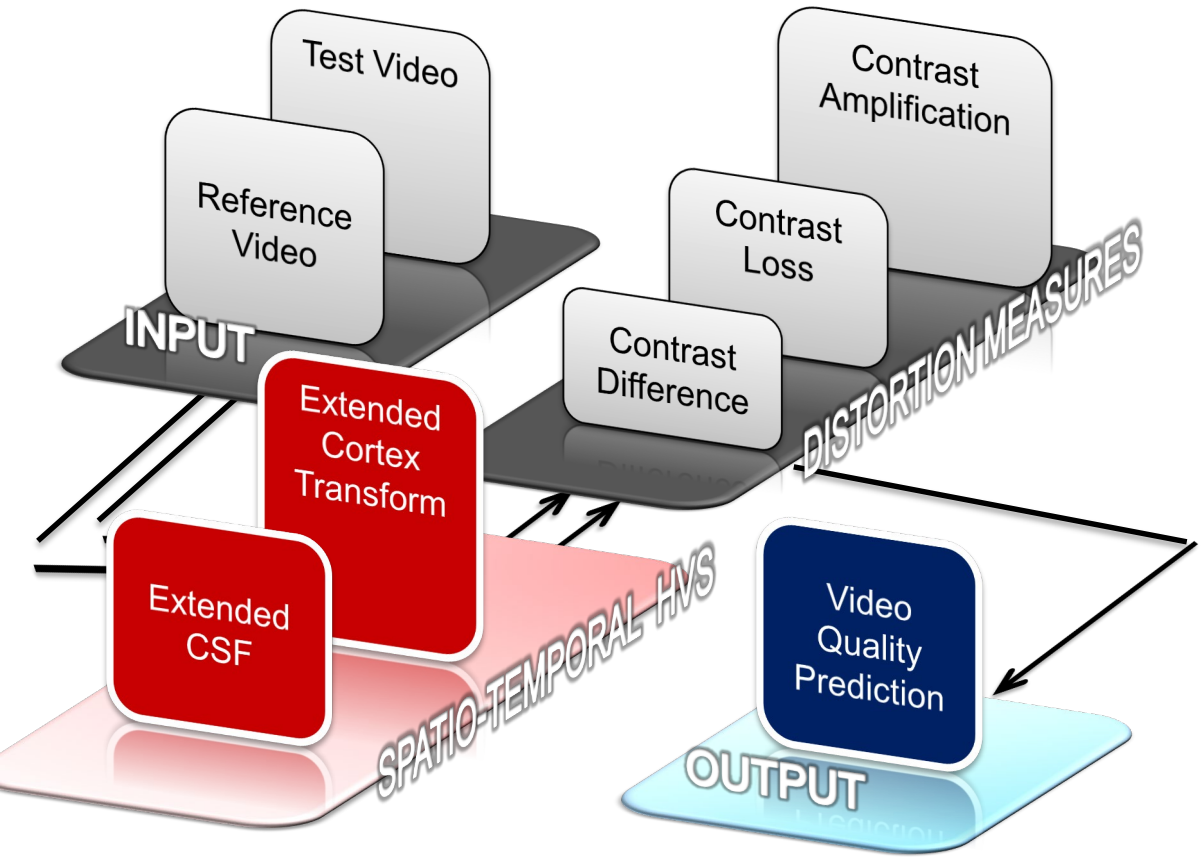
corr = 0.772

Artifact Image

Overview

- Introduction to Objective Quality Assessment
- Image Quality Assessment
  - HVS-based Metrics (bottom-up)
  - Structural Similarity (top-down)
  - Data-driven Approaches (top-down)
- Video Quality Assessment

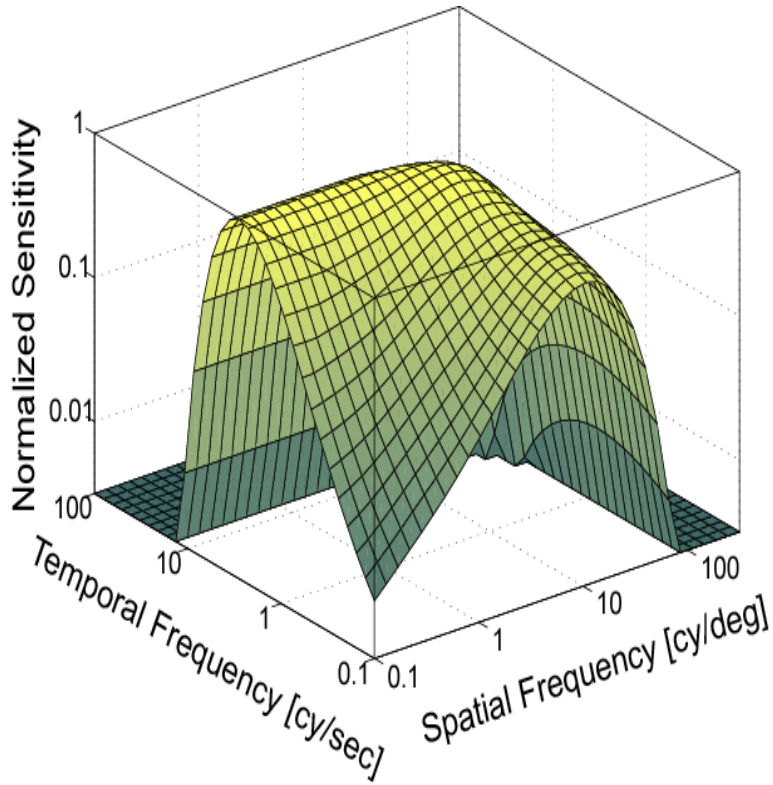
Dynamic Range Independent VQA



- [Aydın et al. 2010]
- **Key Idea:** Extend the Dynamic Range Independent pipeline with temporal aspects to evaluate video sequences
- **Result:** An objective VQM that evaluates rendering quality, temporal tone mapping and HDR compression

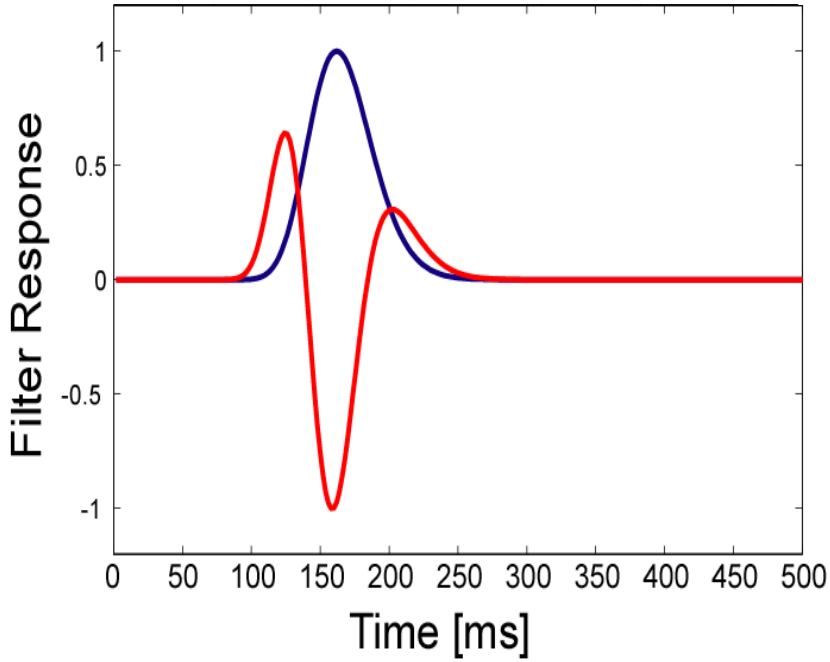
Contrast Sensitivity Function

- *CSF:* $\omega, \rho, L_a \rightarrow S$
  - ω: temporal frequency,
  - ρ: spatial frequency,
  - L_a: adaptation level,
  - S: sensitivity.

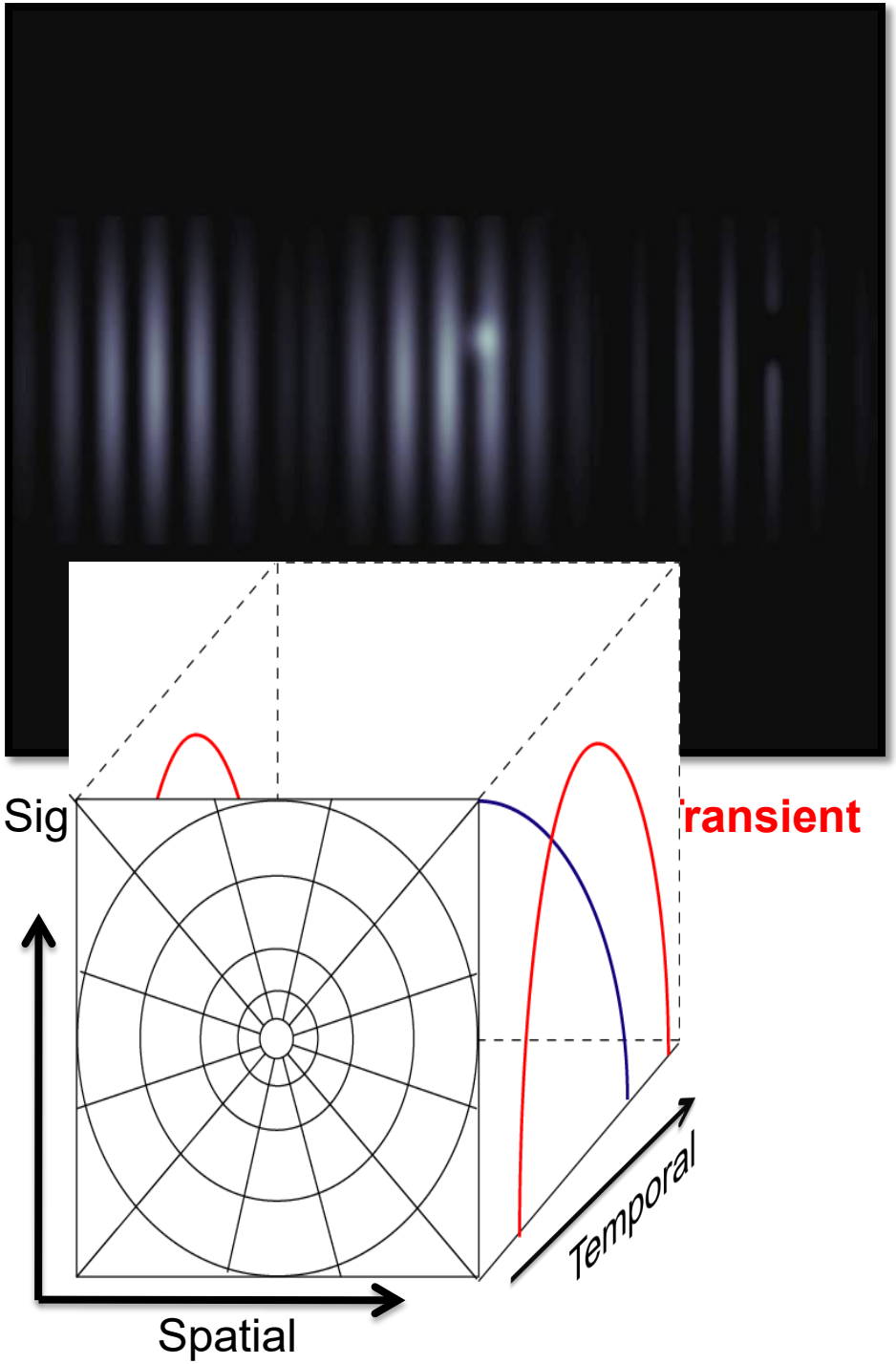


Spatio-temporal CSF

Extended Cortex Transform



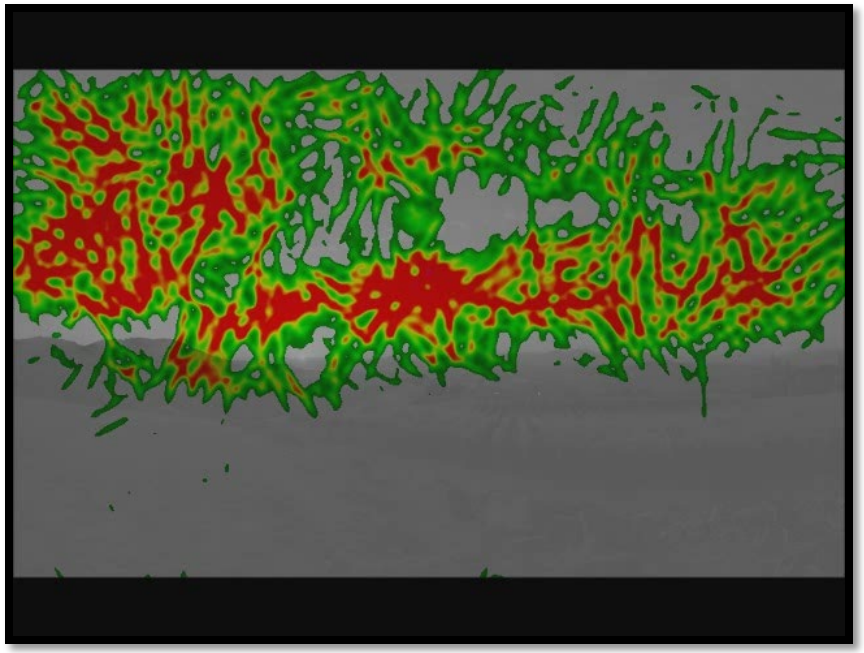
Sustained and **Transient**
Temporal Channels [Winkler 2005]



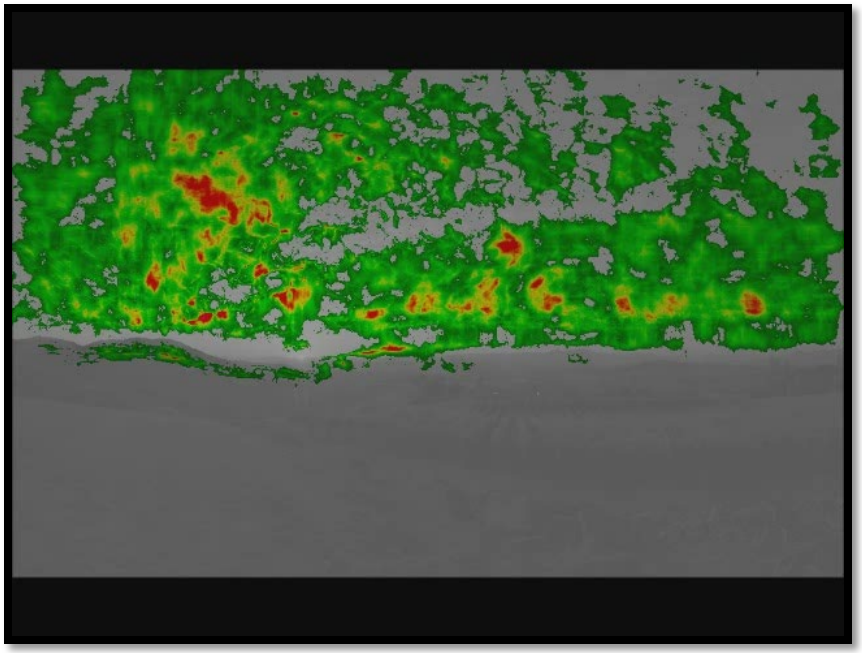
Comparison: Test Scene

- HDR Scene tone mapped with [Pattanaik 2000]
- Spatio-temporal distortion
  - Random pixel noise filtered with a Gaussian.

Metric Comparison LDR-LDR



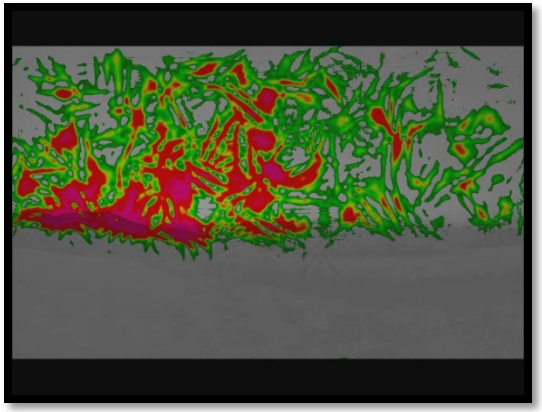
Our Metric

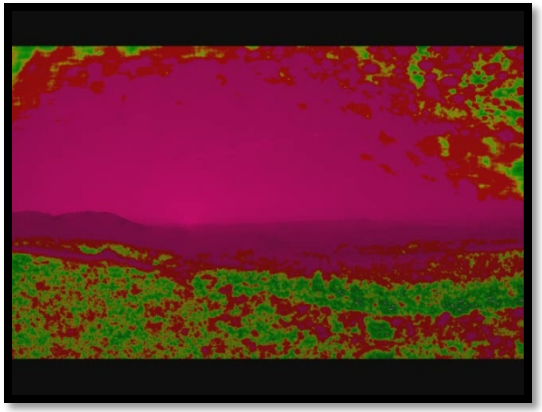


PDM [Winkler 2005]

Metric Comparison HDR-LDR

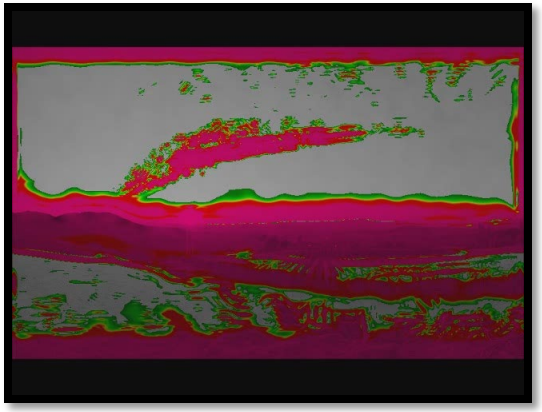
Our Metric





PDM
[Winkler 2005]

HDRVDP
[Mantiuk
et al 2005]

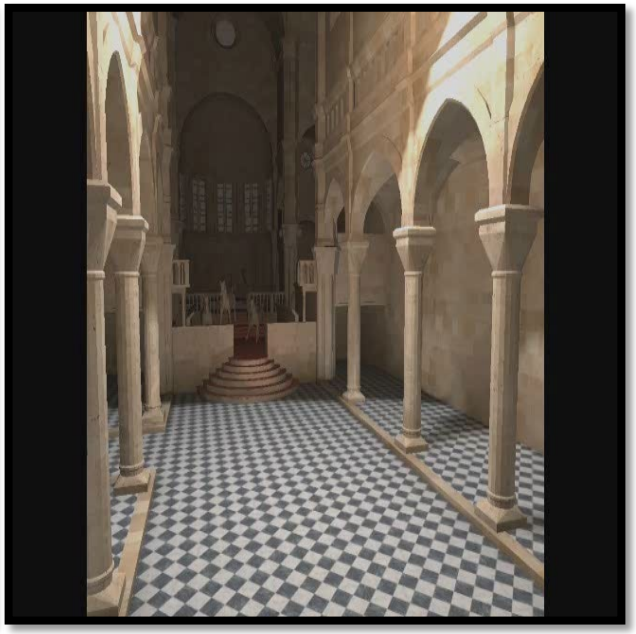


DRIM
[Aydin et al. 2008]

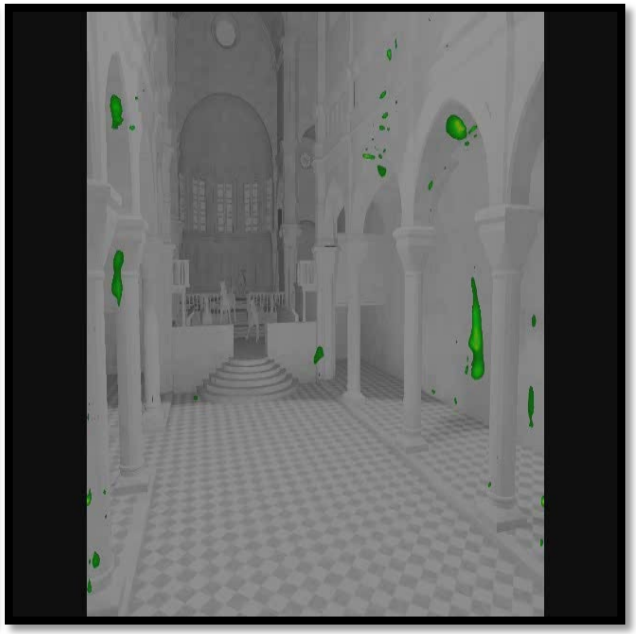
Evaluation of Rendering Methods

- http://drim.mpi-inf.mpg.de/

With temporal filtering
[Herzog et al. 2010]

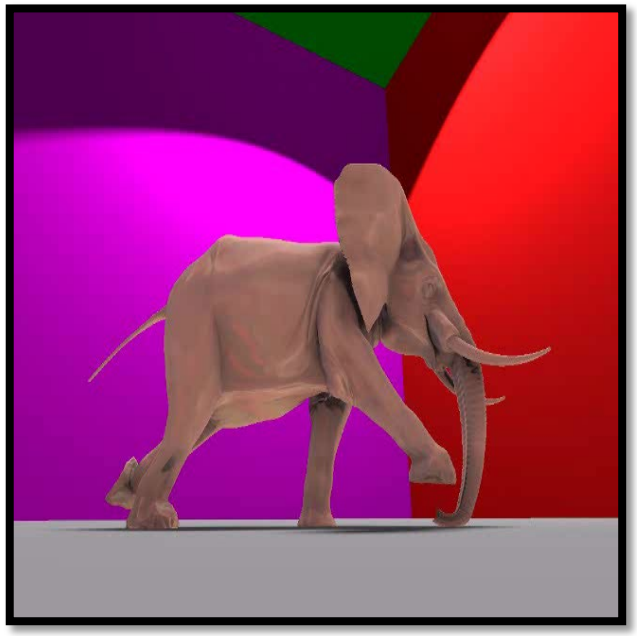


No temporal filtering

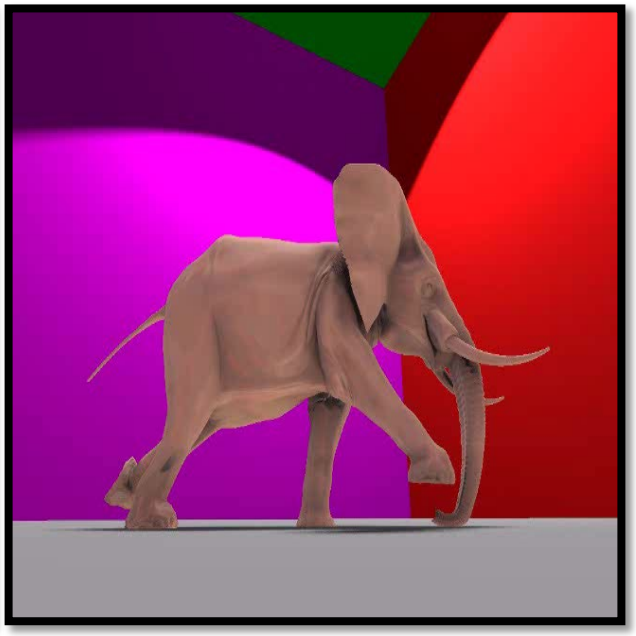


Predicted distortion map

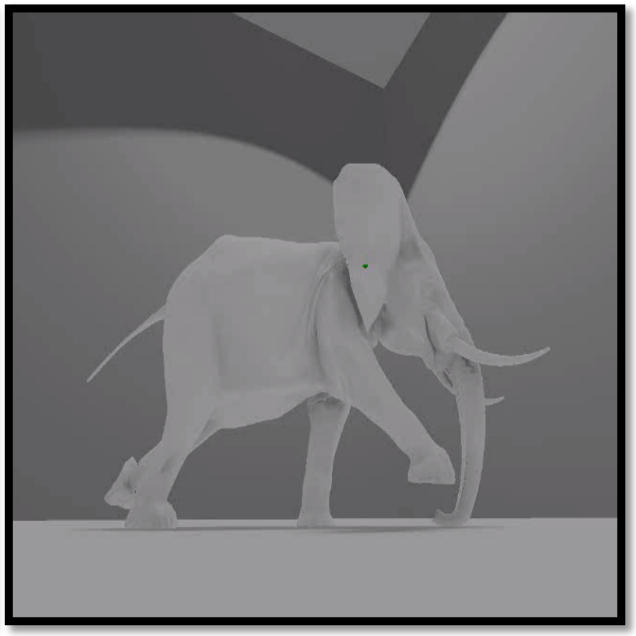
Evaluation of Rendering Qualities



High quality

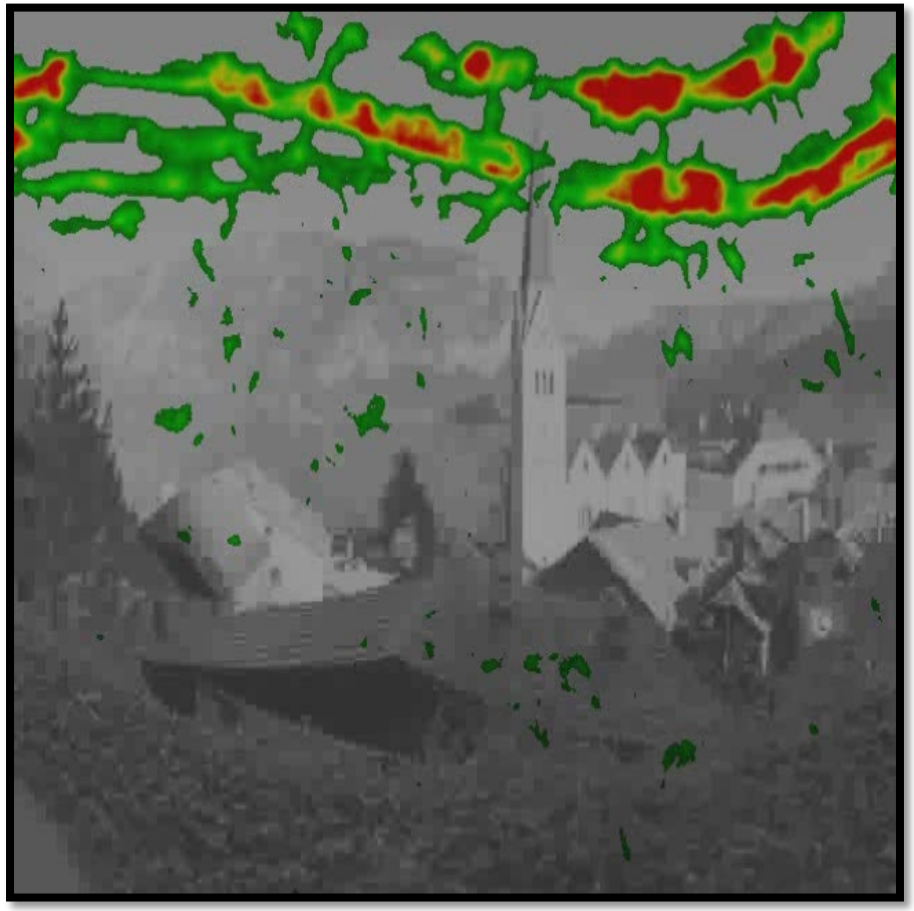


Low quality



Predicted distortion map

Evaluation of HDR Compression

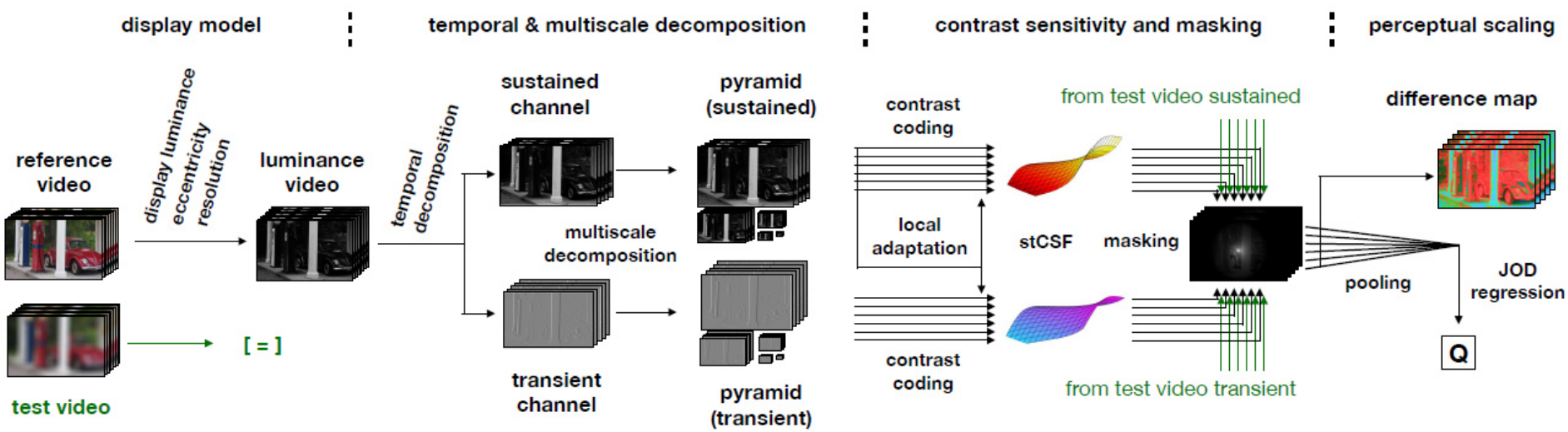


Medium Compression

High Compression

FovVideoVDP

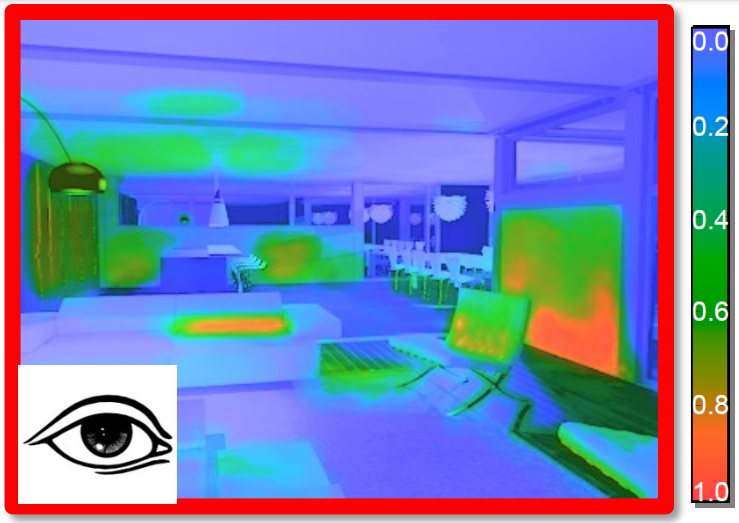
- [Mantiuk et al. 21]
  - spatial, temporal, foveated, peripheral perception
  - CSF, contrast masking
  - supra-threshold
  - no orientation-selective visual channels, no cross-channel masking, no color vision
  - Laplacian pyramid (no cortex transform) → 100x faster than HDRVDP3
  - Pytorch, matlab



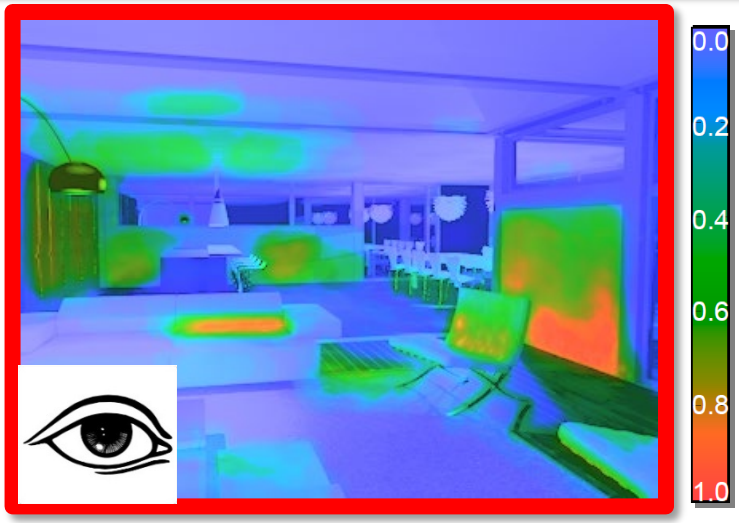
EXPERIMENTAL EVALUATIONS

User Experiment - Mean Distortion Maps

- 37 test images
- 35 subjects (expert and non experts)
- Localization of artifacts
- Scribbling interface

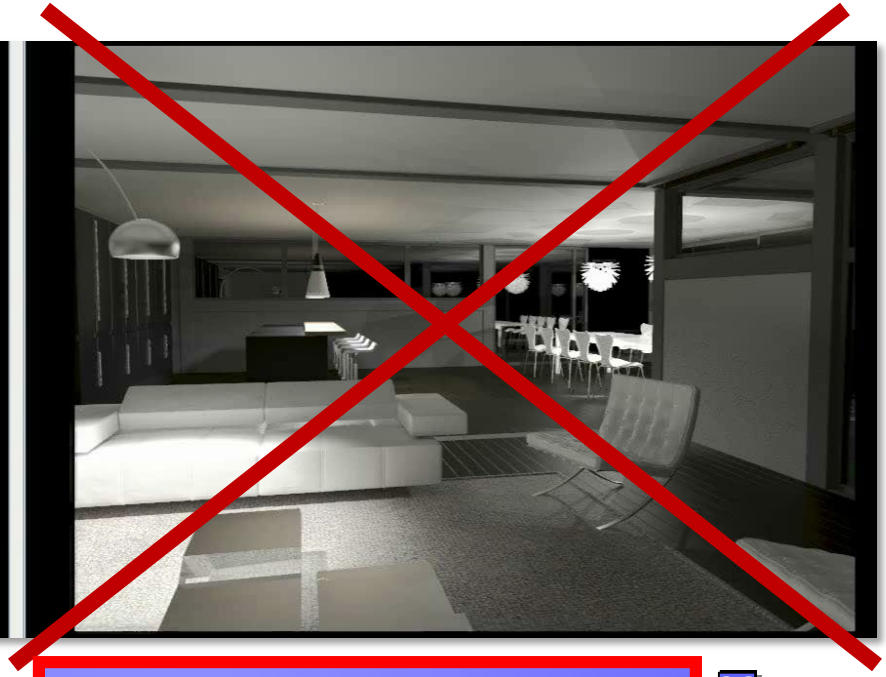


User Experiment – With Reference



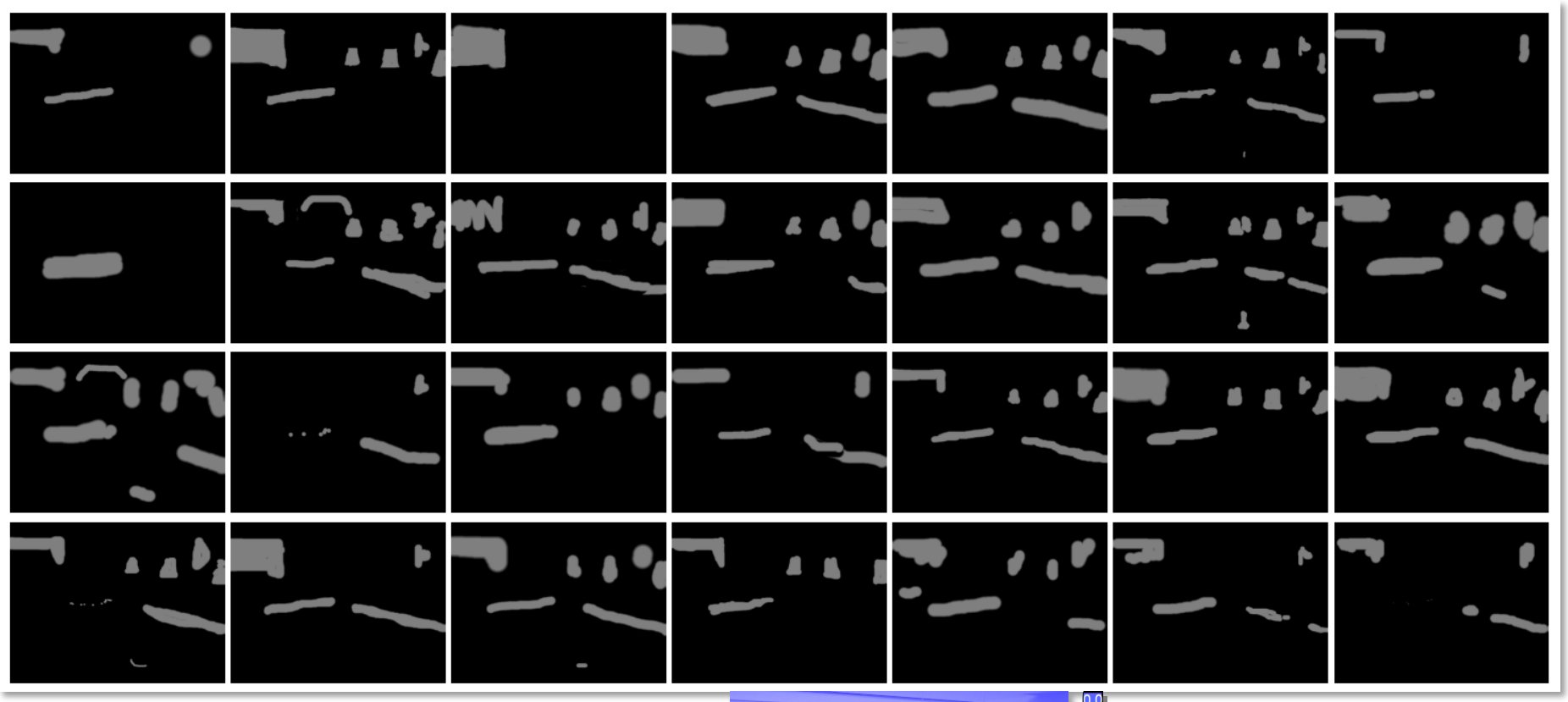
- Noticeable distortions

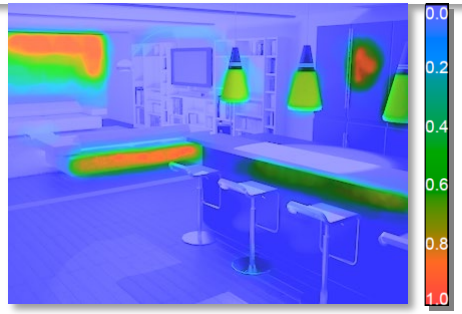
User Experiment – No Reference



- Objectionable distortions

Example User Responses

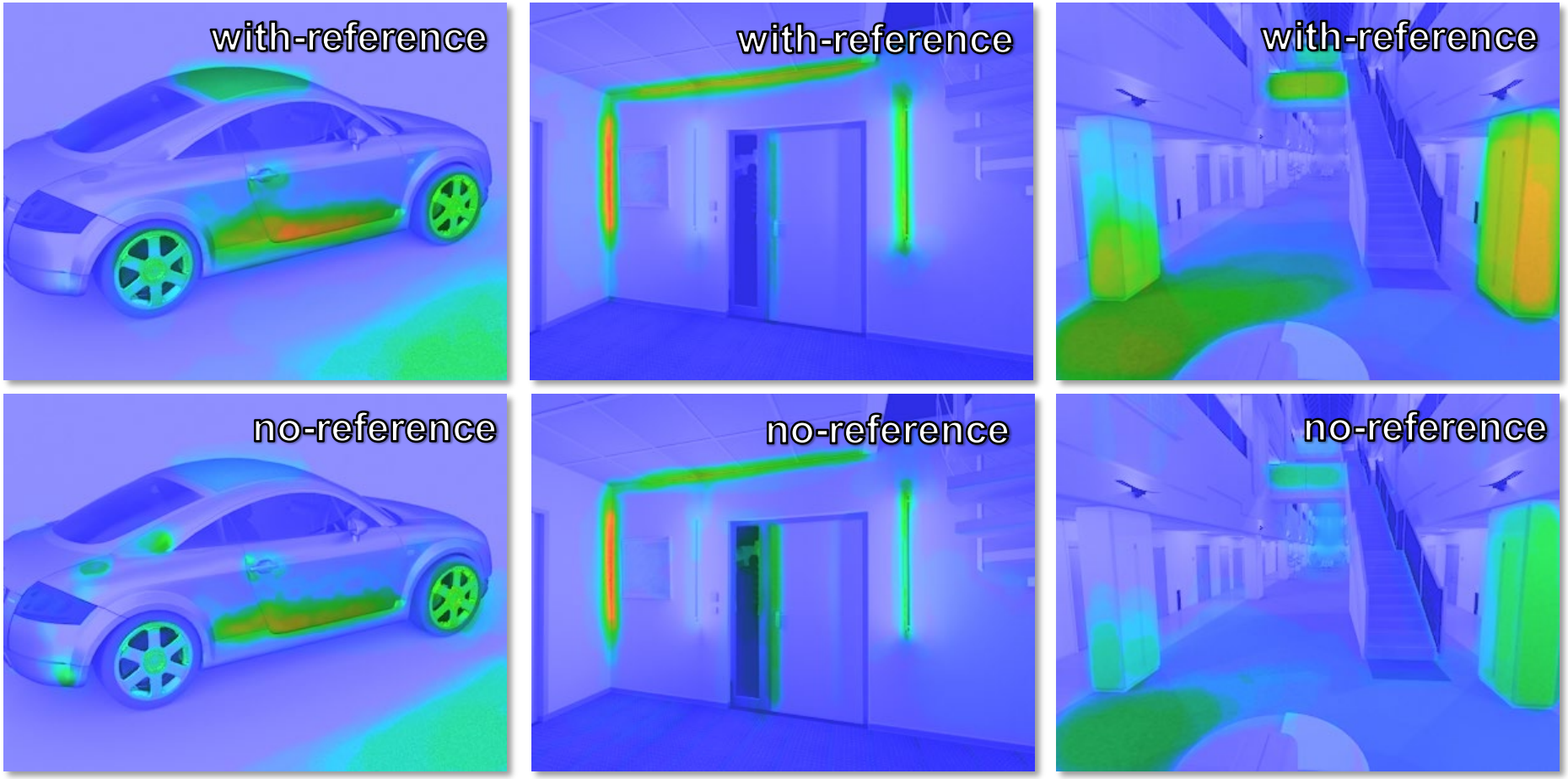




- Probability of detection

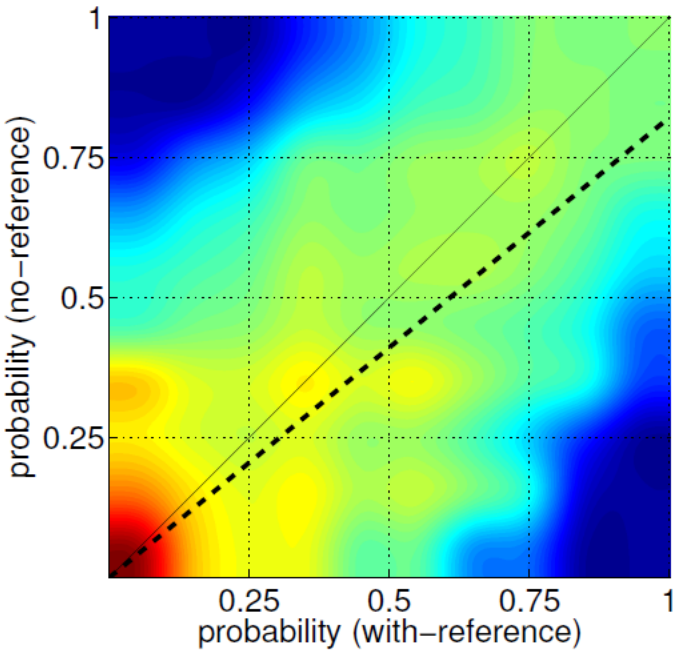
With-reference vs. No-reference

- Results rather similar

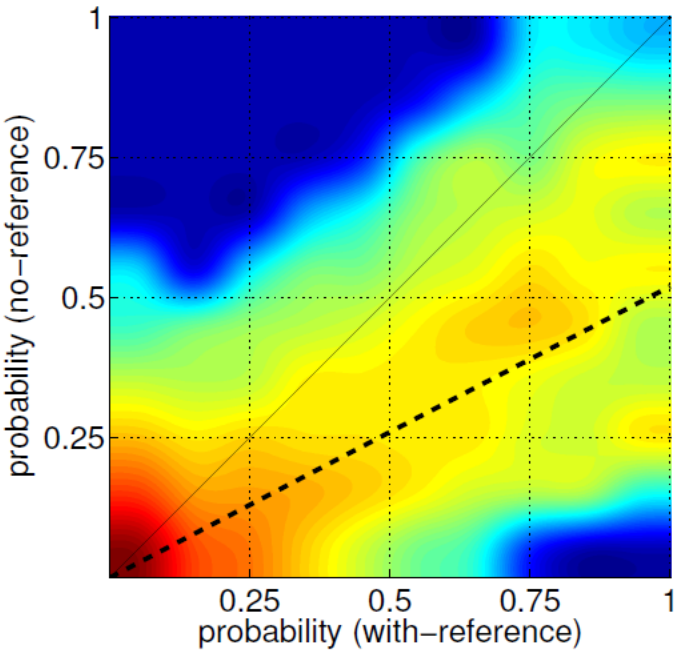


With-reference vs. No-reference (cont.)

- Strong correlation
  - (perhaps people do not need the reference)
- SRCC=0.88 SRCC=0.85



EG'12 dataset

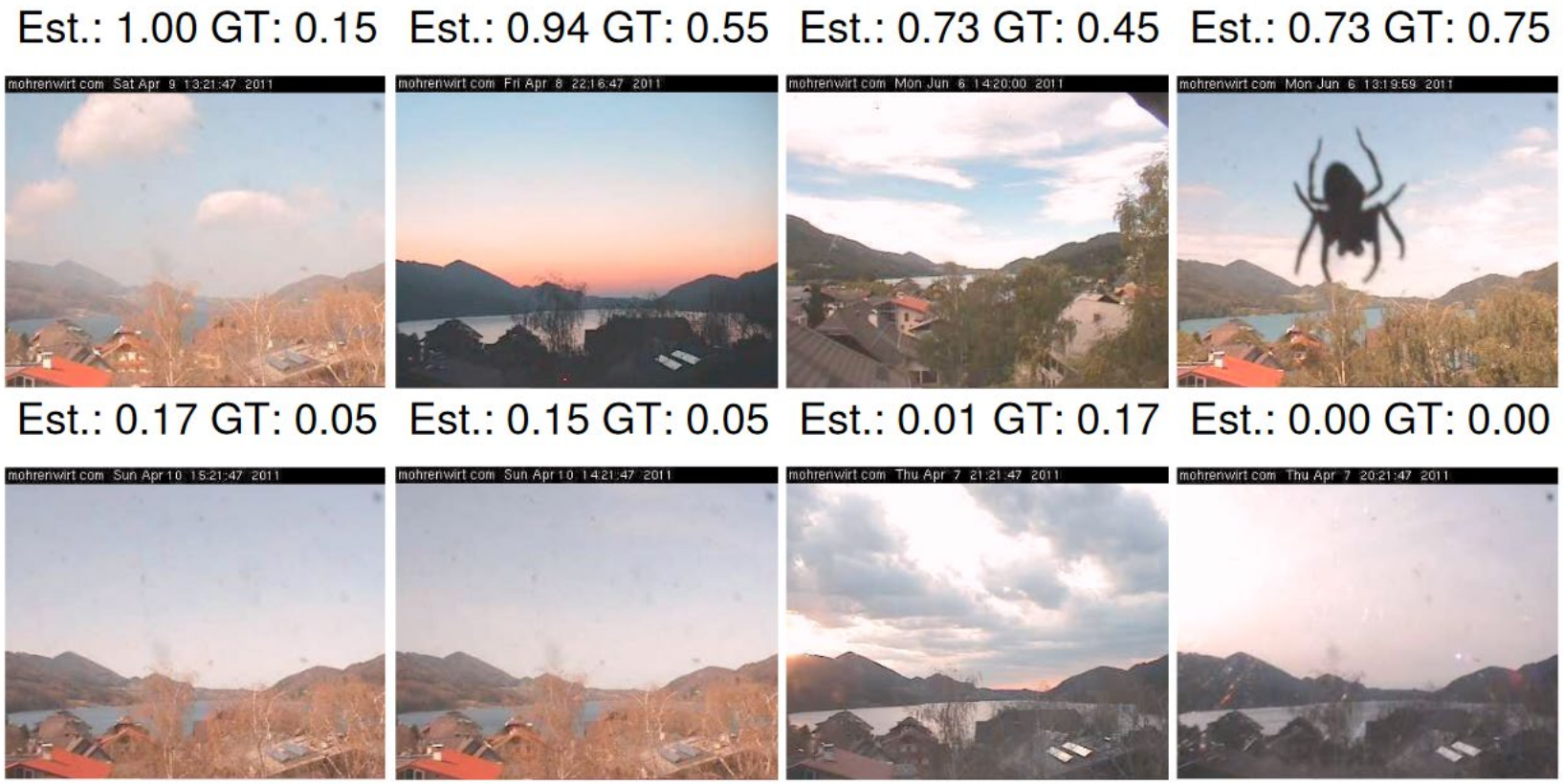


new dataset

OTHER METRICS

Interestingness of Images

- interestingness = aesthetics, unusualness, general preferences [Gygli et al. 13]



Automated aesthetic analysis

- [Aydin et al. 15]

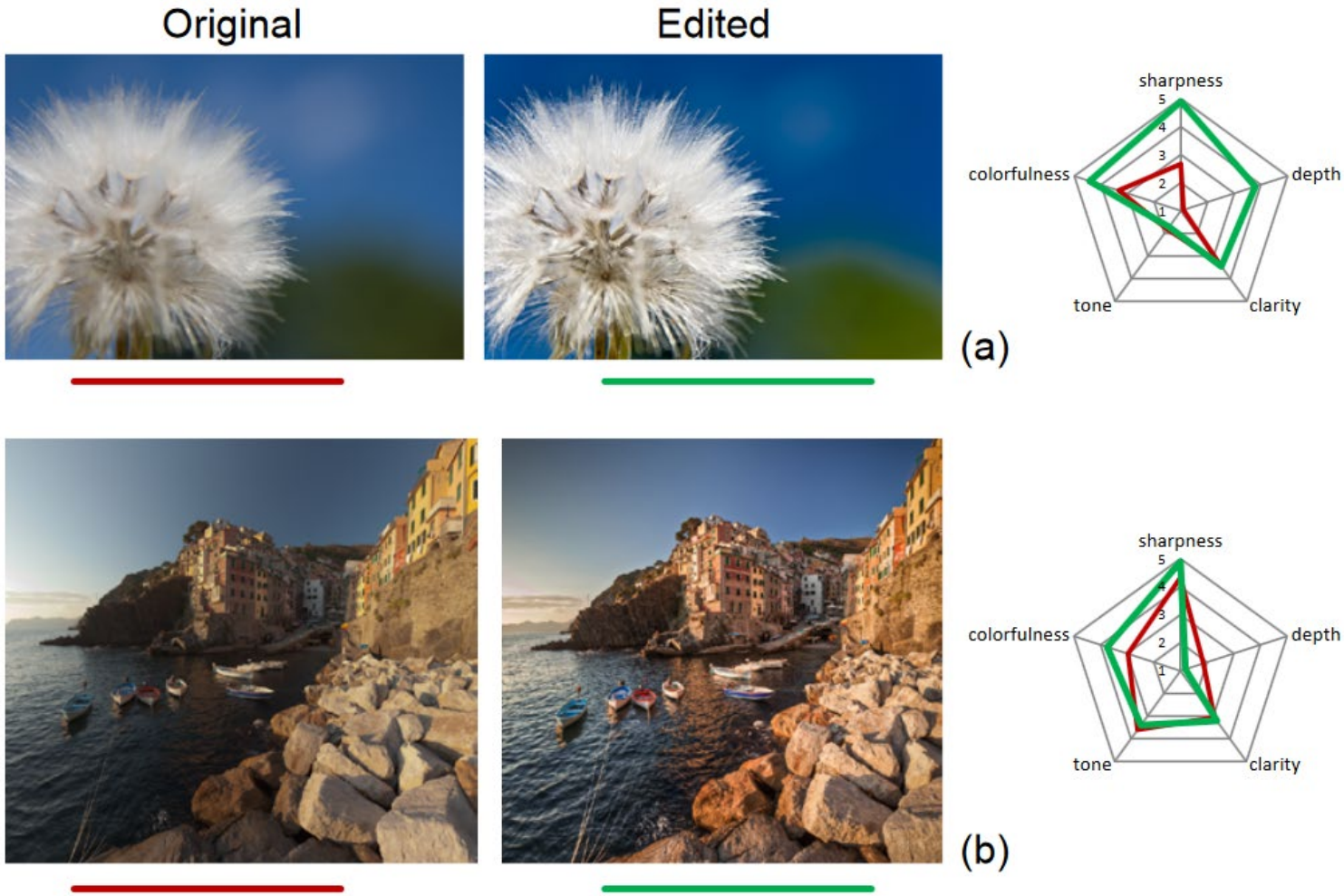


Image completion quality prediction

- [Kopf et al. 12]

Input panorama and quality prediction
(brighter is higher quality)

Full completion

Best crop using only “known” pixels (“conservative crop”)

Our optimized crop based on quality prediction

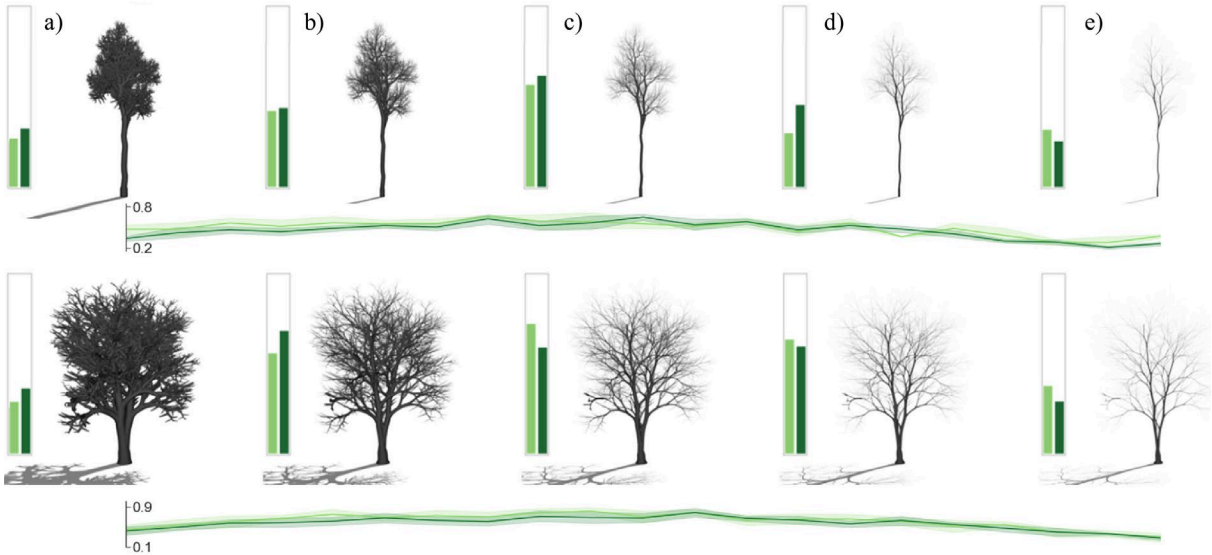
Specific Metrics

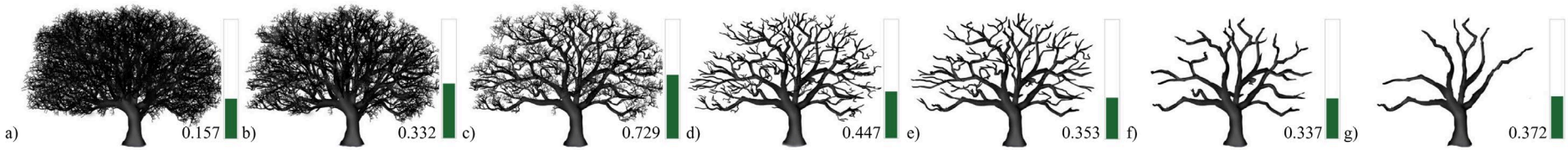
- Spectral metrics
  - [Le Moan and Urban 14]
- Lightfield metrics
- HDR metrics
- …

ICTree

- [Polasek, Hrusa, Benes, Čadík 21]

http://cphoto.fit.vutbr.cz/ictree/

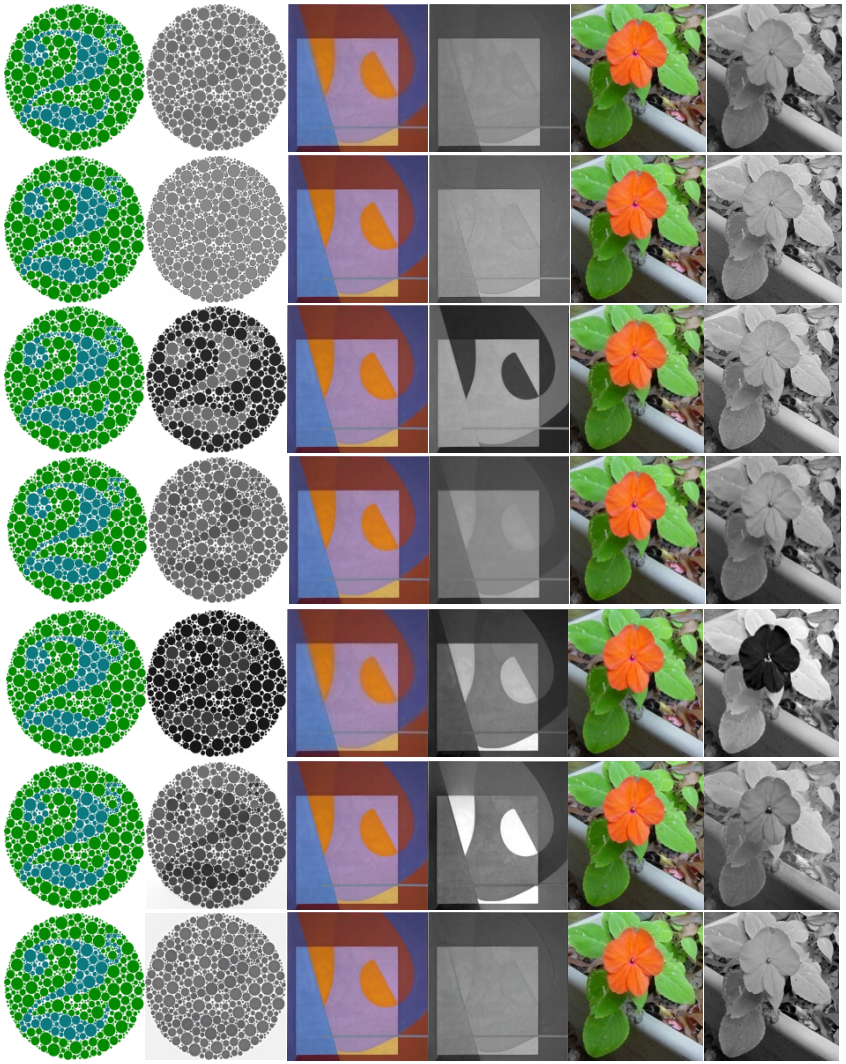




Wrong Usage of IQM/VQM

- Codruta O. Ancuti, Cosmin Ancuti and Philippe Bekaert, "Enhancing by Saliency-guided Decolorization", In Proc. IEEE Computer Vision and Pattern Recognition (IEEE CVPR), Colorado Springs, USA, 2011.

- DRIM used for comparing color-to-grayscale conversions



SUMMARY

Conclusions

- No universal method
- Simple, robust techniques score high
- Advanced fancy methods are nice, but need to improve on robustness
- With reference ~~ no-reference experiments

Methods to Try

- Images
  - HDR-HDR
    - **HDRVDP3** [Mantiuk et al. 11]
  - HDR-LDR
    - **HDRVDP3** [Mantiuk et al. 11]
    - **TMQI** [Yeganeh, Wang 13]
  - LDR-LDR
    - **CNN** [Wolski et al. 13]
    - **SSIM** [Wang, Bovik 04]
    - **FLIP** [Andersson et al. 20]
- Videos
  - HDR-HDR
    - **DRIVQM** [Aydin et al. 10]
  - HDR-LDR
    - **DRIVQM** [Aydin et al. 10]
  - LDR-LDR
    - **FovVideoVDP** [Mantiuk et al. 21]

Thanks for your attention

cadik@fit.vutbr.cz
http://cadik.posvete.cz

Many thanks to my colleagues, in particular to Tunç Ozan Aydın, Robert Herzog, Rafał Mantiuk, Radoslaw Mantiuk, László Neumann, Karol Myszkowski, Hans-Peter Seidel, Alessandro Artusi

References

[Daly93] Scott Daly. 1993. The visible differences predictor: an algorithm for the assessment of image fidelity. Digital images and human vision. MIT Press, Cambridge, MA, USA, 179–206.

[Teo, Heeger 94] Patrick C. Teo, David J. Heeger. Perceptual image distortion. Proceedings Volume 2179, Human Vision, Visual Processing, and Digital Display V; (1994), IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology, 1994, San Jose, CA, US.

[Lubin 95] JEFFREY LUBIN. A VISUAL DISCRIMINATION MODEL FOR IMAGING SYSTEM DESIGN AND EVALUATION, Vision Models for Target Detection and Recognition, pp. 245-283 (1995).

[Taylor et al. 97] Christopher C. Taylor, Zygmunt Pizlo, Jan P. Allebach, Charles A. Bouman. Image quality assessment with a Gabor pyramid model of the human visual system. Proceedings Volume 3016, Human Vision and Electronic Imaging II; (1997). Electronic Imaging '97, 1997, San Jose, CA, US.

[Bradley 99] A. P. Bradley. A wavelet visible difference predictor. IEEE Trans Image Process. 1999;8(5):717-30. doi: 10.1109/83.760338.

[HDRVDP, Mantiuk et al. 11] Rafał Mantiuk, Kil Joong Kim, Allan G. Rempel and Wolfgang Heidrich. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Transactions on Graphics (Proc. of SIGGRAPH'11), 30(4), article no. 40, 2011. http://hdrvdp.sourceforge.net/wiki/

[Wang 02] Zhou Wang and Alan C. Bovik. A Universal Image Quality Index. IEEE Signal Processing Letters, vol. 9, no. 3, pp. 81-84, March 2002.

[Wang 04] Zhou Wang; A.C. Bovik; H.R. Sheikh; E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing (Volume: 13, Issue: 4, April 2004).

[Wang et al. 04] Wang, Z.; Simoncelli, E.P.; Bovik, A.C. (2003-11-01). Multiscale structural similarity for image quality assessment. Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, 2004. 2. pp. 1398–1402 Vol.2. CiteSeerX 10.1.1.58.1939. doi:10.1109/ACSSC.2003.1292216. ISBN 978-0-7803-8104-9.

[Shnayderman 04] Shnayderman, Aleksandr & Gusev, Alexander & Eskicioglu, Ahmet. (2003). Multidimensional image quality measure using singular value decomposition. Proceedings of SPIE - The International Society for Optical Engineering. 5294. 10.1117/12.530554.

[Andersson et al. 20] Pontus Andersson, Jim Nilsson, Tomas Akenine-Möller, Magnus Oskarsson, Kalle Åström, and Mark D. Fairchild. 2020. FLIP: A Difference Evaluator for Alternating Images. Proc. ACM Comput. Graph. Interact. Tech. 3, 2, Article 15 (August 2020), 23 pages. DOI:https://doi.org/10.1145/3406183

[Čadík et al. 13] Čadík, M., Herzog, R., Mantiuk, R., Mantiuk, R., Myszkowski, K. and Seidel, H.-P. (2013), Learning to Predict Localized Distortions in Rendered Images. Computer Graphics Forum, 32: 401-410. https://doi.org/10.1111/cgf.12248

References

[Herzog et al. 12] Herzog, R., Čadík, M., Aydin, T. O., Kwang I. K., Myszkowski, K., Seidel, H.P. NoRM: No-Reference Image Quality Metric for Realistic Image Synthesis. Computer Graphics Forum, vol. 31, issue 2 (2012), (Proceedings of EUROGRAPHICS'12, Cagliari / Italy, 13–18 May), ISSN 0167-7055, 2012.

[Wolski et al. 18] Krzysztof Wolski, Daniele Giunchi, Nanyang Ye, Piotr Didyk, Karol Myszkowski, Radosław Mantiuk, Hans-Peter Seidel, Anthony Steed, and Rafał K. Mantiuk. 2018. Dataset and Metrics for Predicting Local Visible Differences. ACM Trans. Graph. 37, 5, Article 172 (November 2018), 14 pages. DOI:https://doi.org/10.1145/3196493

[Hasan et al. 11] Hasan M.M., Ahn K., Chae O. (2011) Measuring Blockiness of Videos Using Edge Enhancement Filtering. In: Kim T., Adeli H., Ramos C., Kang BH. (eds) Signal Processing, Image Processing and Pattern Recognition. SIP 2011. Communications in Computer and Information Science, vol 260. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-27183-0_2

[Chen et al. SPIE ’11] Chen, Chunhua & Chen, Wen & Bloom, Jeffrey. (2011). A Universal, Reference-Free Blurriness Measure. SPIE vol. 7867. 10.1117/12.872477.

[Daly S. & Feng SPIE ’04] Daly, Scott J. and Xiao-Fan Feng. “Decontouring: prevention and removal of false contour artifacts.” IS&T/SPIE Electronic Imaging (2004).

[Sheikh et al. 05] Sheikh, Hamid & Bovik, Alan & Cormack, Lawrence. (2005). No-reference quality assessment using natural scene statistics: JPEG2000. Image Processing, IEEE Transactions on. 14. 1918 - 1927. 10.1109/TIP.2005.854492.

[Aydın et al. 2010] Aydin, T.O., Čadík, M., Myszkowski, K., Seidel, H.P. Video Quality Assessment for Computer Graphics Applications. ACM Transactions on Graphics (SIGGRAPH Asia 2010), vol. 29, no. 5, ISSN 0730-0301, ACM, 2010.

[Pattanaik 2000] Pattanaik, Sumanta & Tumblin, Jack & Yee, Hector & Greenberg, Donald. (2000). Time-Dependent Visual Adaptation For Fast Realistic Image Display. Proceedings of the ACM SIGGRAPH Conference on Computer Graphics. 10.1145/344779.344810.

[Winkler 2005] WINKLER, S. 2005. Digital Video Quality: Vision Models and Metrics. Wiley.

[Aydin et al. 2008] Tunç Ozan Aydın, Rafał Mantiuk, Karol Myszkowski, Hans-Peter Seidel. Dynamic Range Independent Image Quality Assessment. ACM Transactions on Graphics (TOG) 2008 (SIGGRAPH 2008).

[Herzog et al. 2010] Robert Herzog, Elmar Eisemann, Karol Myszkowski, Hans-Peter Seidel. Spatio-Temporal Upsampling on the GPU. Proceedings of Symposium on Interactive 3D Graphics and Games (ACM), Washinton DC, US, Feburary 2010.

[Mantiuk et al. 21] Rafał K. Mantiuk, Gyorgy Denes, Alexandre Chapiro, Anton Kaplanyan, Gizem Rufo, Romain Bachy, Trisha Lian, and Anjul Patney. FovVideoVDP: A visible difference predictor for wide field-of-view video. ACM Transactions on Graphics (Proc. of SIGGRAPH 2021), 40(4), article no. 49, 2021.

References

[Gygli et al. 13] Gygli, Michael & Grabner, Helmut & Riemenschneider, Hayko & Nater, Fabian & Van Gool, Luc. (2013). The Interestingness of Images. Proceedings of the IEEE International Conference on Computer Vision. 1633-1640. 10.1109/ICCV.2013.205.

[Aydin et al. 15] Tunç Ozan Aydın, Aljoša Smolić, Markus Gross. Automated Aesthetic Analysis of Photographic Images, IEEE Transactions on Visualization and Computer Graphics (TVCG) 2015.

[Le Moan and Urban 14] S. Le Moan and P. Urban, "Image-Difference Prediction: From Color to Spectral," in IEEE Transactions on Image Processing, vol. 23, no. 5, pp. 2058-2068, May 2014, doi: 10.1109/TIP.2014.2311373.

[Polasek, Hrusa, Benes, Čadík 21] Tomas Polasek, David Hrusa, Bedrich Benes, Martin Čadík. ICTree: Automatic Perceptual Metrics for Tree Models. ACM Transactions on Graphics (Proceedings of SIGGRAPH ASIA 2021).