Martin Čadík (cadik@fit.vutbr.cz)

HDR IMAGE AND VIDEO QUALITY ASSESSMENT

MAX-PLANCK-GESELLSCHAFT
BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY
MARIE CURIE ACTIONS
SoMoPro

Outline

1) Introduction to HDR
2) Quality Assessment
    1) Introduction to IQA, HVS
    2) Images
    3) Videos
3) Experimental Evaluations

- 2x1h

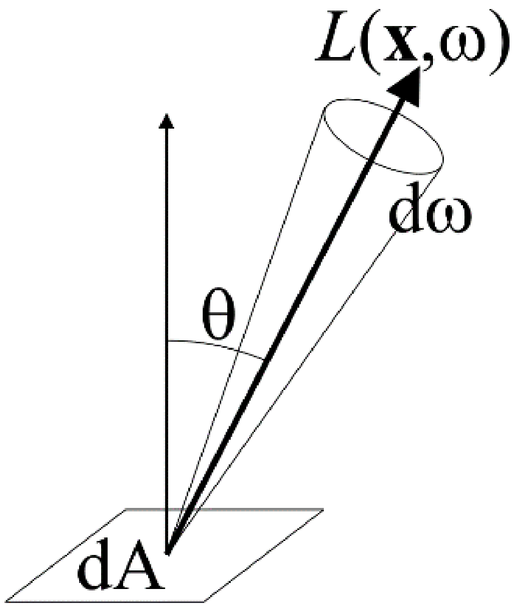
1) INTRODUCTION TO HDR

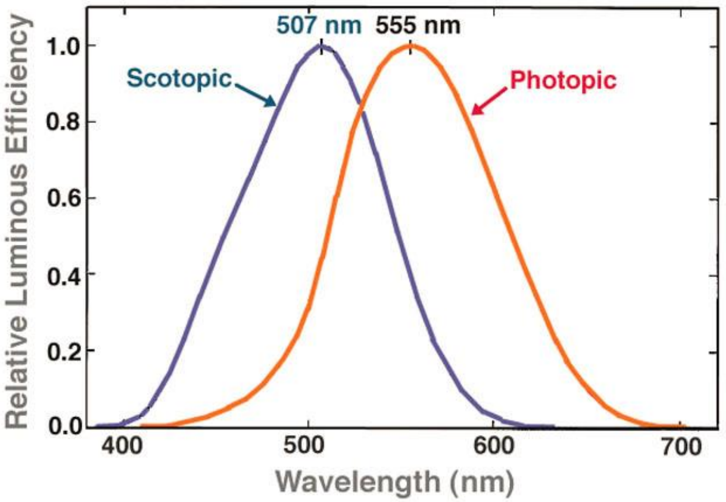
Luminance (1)

- Physically:

$$L_v = \frac{d^2\Phi_v}{dA\,d\Omega\cos\theta} \quad [\text{nit}=\text{cd/m}^2]$$

  - luminous power [lm] per unit solid angle per unit area
  - analogous to “what we see with our eyes”
  - photometric analog of radiance (weighted by luminous efficiency function)

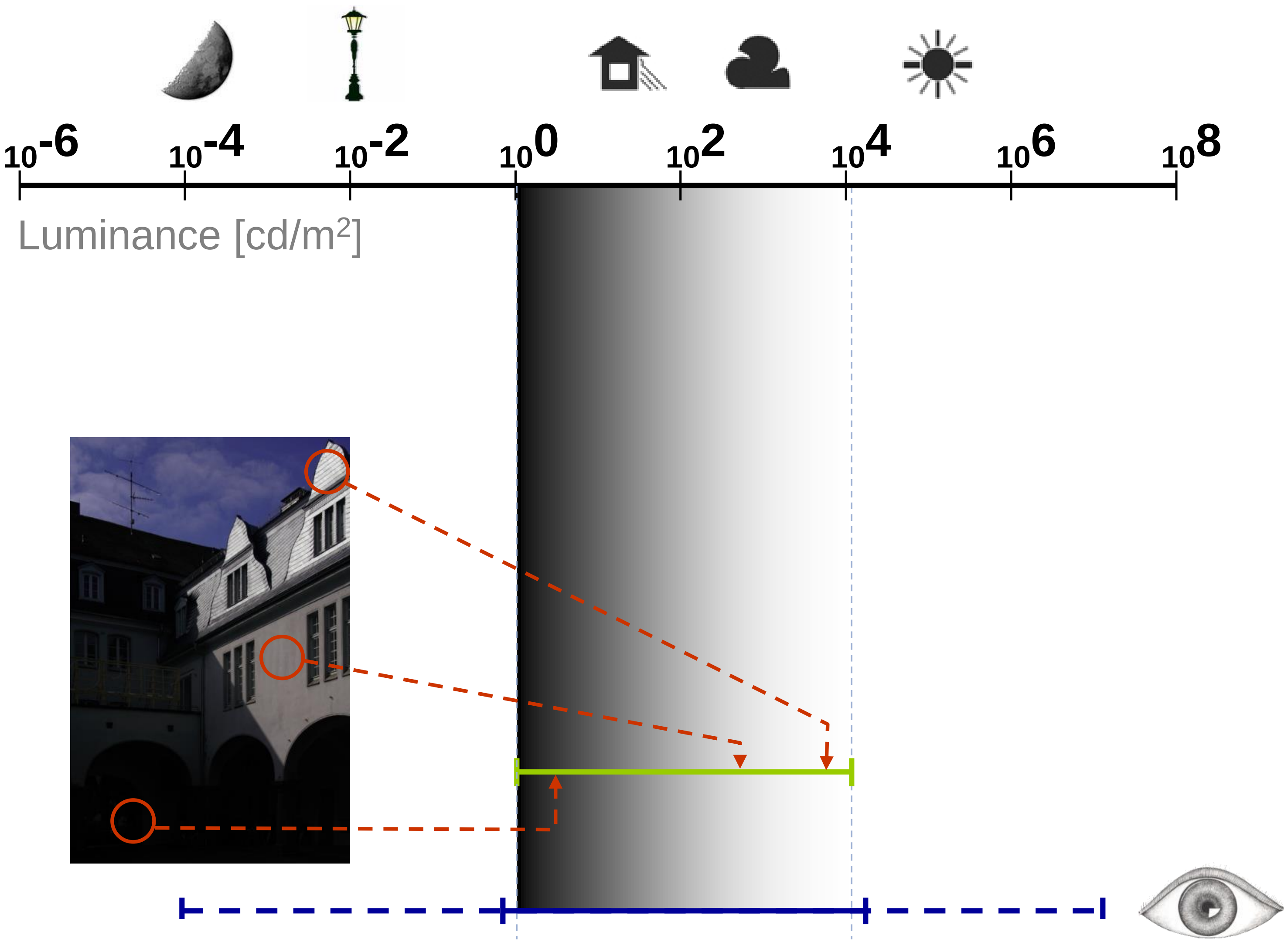




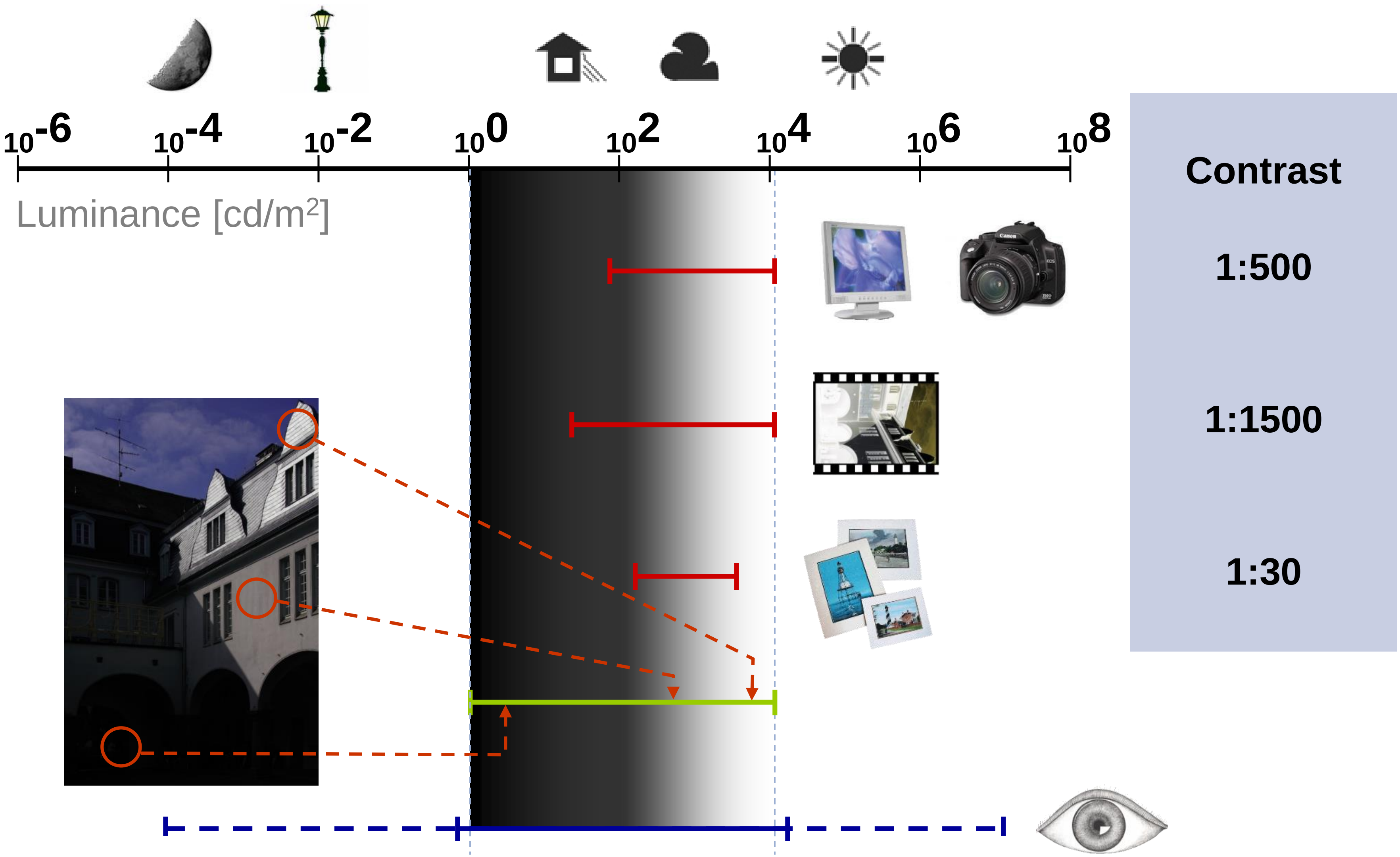
Luminance (2)

- Luminance
  - In color science: weighted sum of **linear** RGB
    - $Y = 0.2126\ R + 0.7152\ G + 0.0722\ B$
- Luma
  - Weighted sum of **gamma corrected** (nonlinear) RGB
    - $Y' = 0.2126\ R' + 0.7152\ G' + 0.0722\ B'$

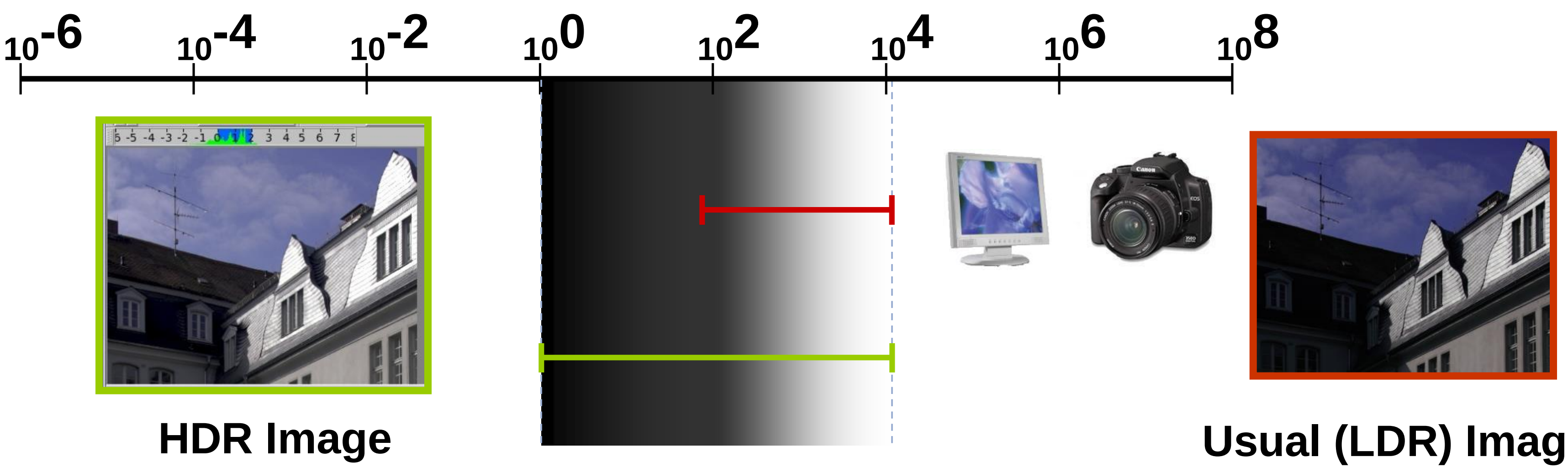
Various Dynamic Ranges (1)



Various Dynamic Ranges (2)



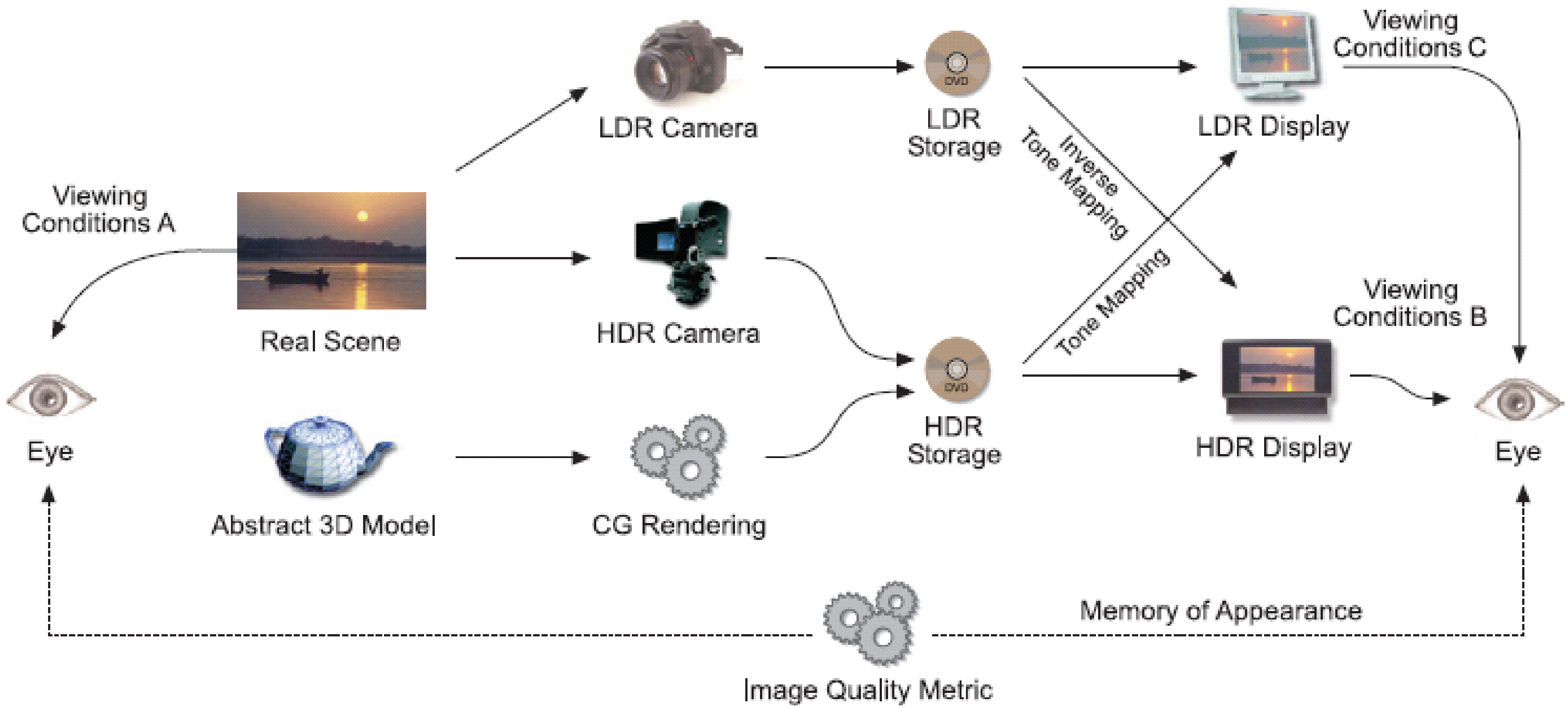
High Dynamic Range



Measures of Dynamic Range

Contrast ratio	$\mathbf{CR = 1 : (Y_{peak}/Y_{noise})}$	displays (e.g., 1:500)
Orders of magnitude	$\mathbf{M = \log_{10}(Y_{peak})-\log_{10}(Y_{noise})}$	HDR imaging (= 2.7 orders)
Exposure latitude (f-stops)	$\mathbf{L = \log_{2}(Y_{peak})-\log_{2}(Y_{noise})}$	photography (= 9 f-stops)
Signal to noise ratio (SNR)	$\mathbf{SNR = 20*\log_{10}(A_{peak}/A_{noise})}$	digital cameras (= 53 [dB])

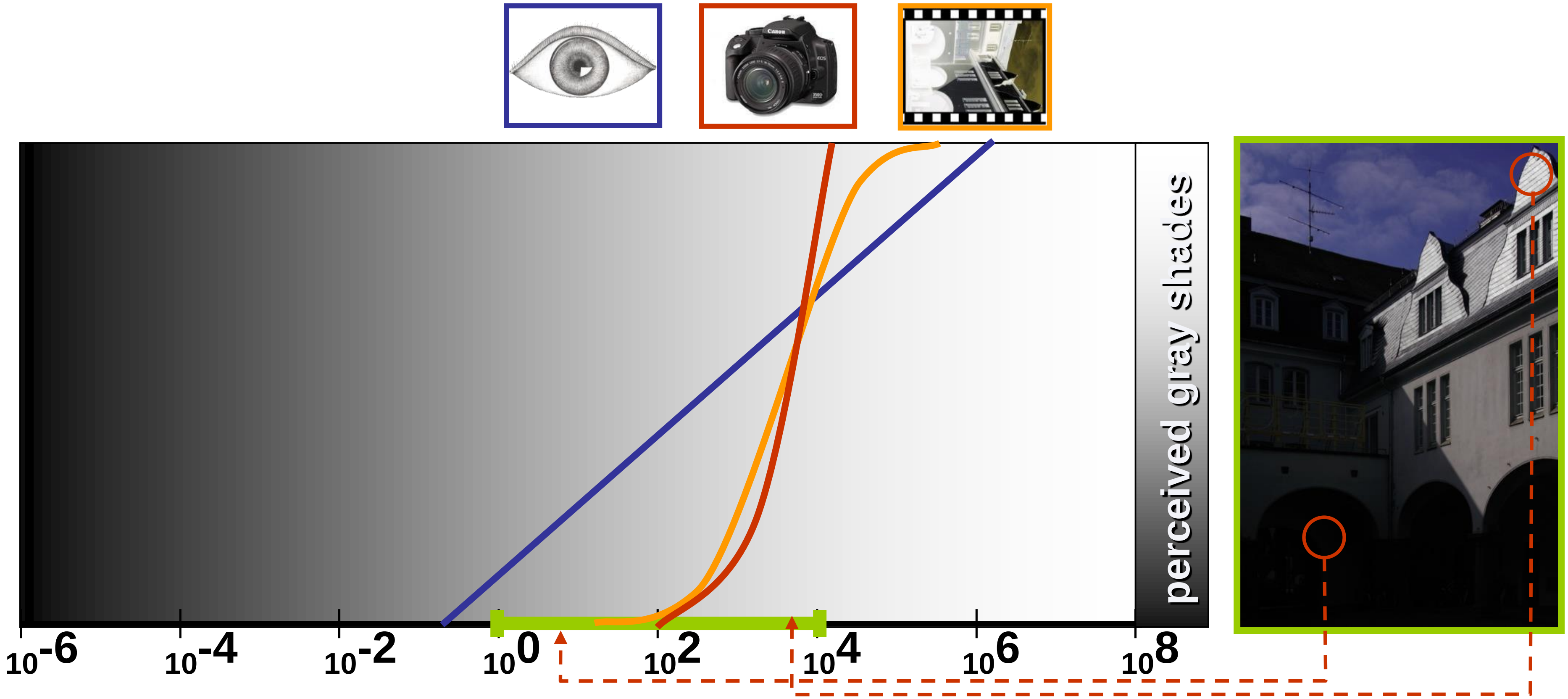
HDR Pipeline



1.1 Overview

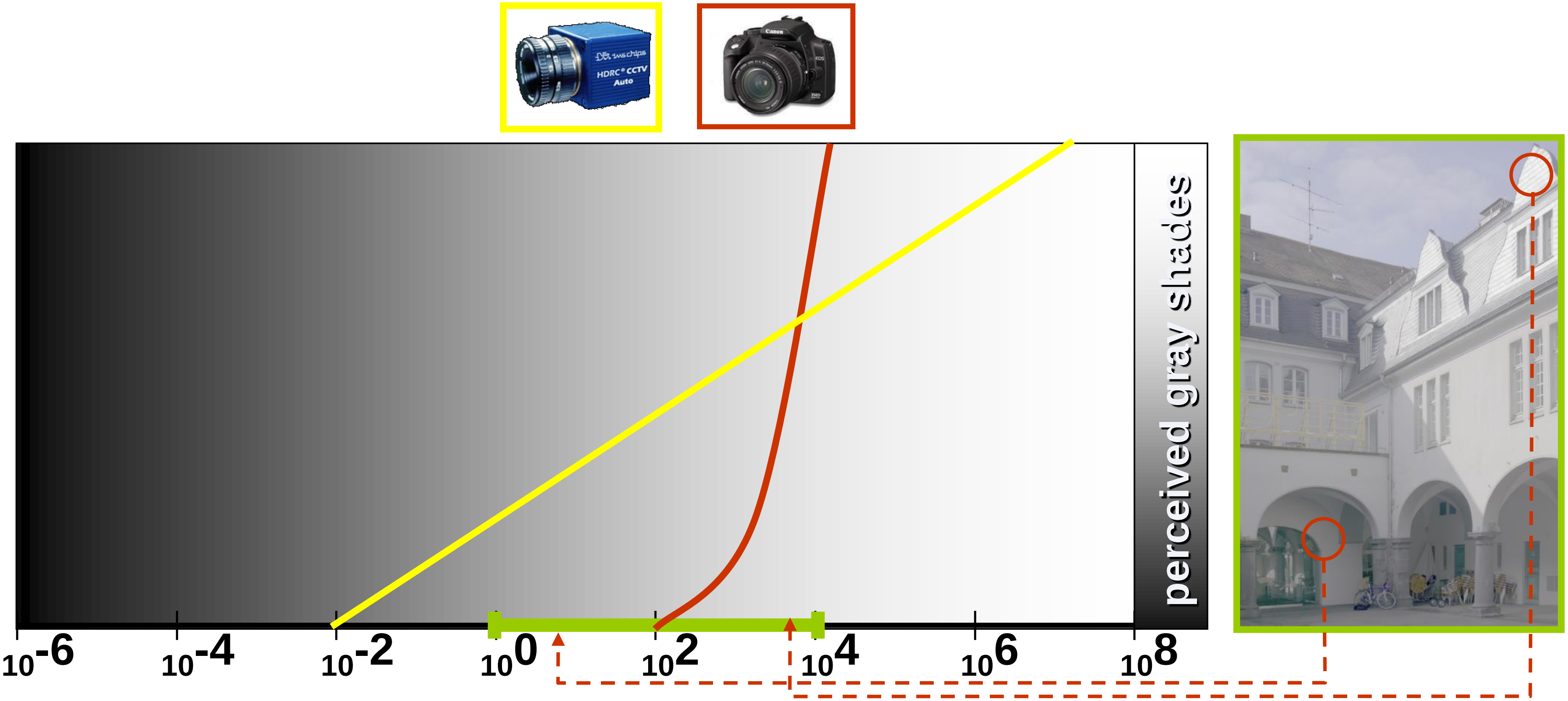
- Capture of HDR images and video
  - HDR sensors
  - Multi-exposure techniques
  - Photometric calibration
- Tone Mapping of HDR images and video
  - Early ideas for reducing contrast range
  - Image processing – fixing problems
  - Alternative approaches
  - Perceptual effects in tone mapping

HDR: a normal camera can't…



- linearity of the CCD sensor
- bound to 8-14bit processors
- saved in an 8bit gamma corrected image

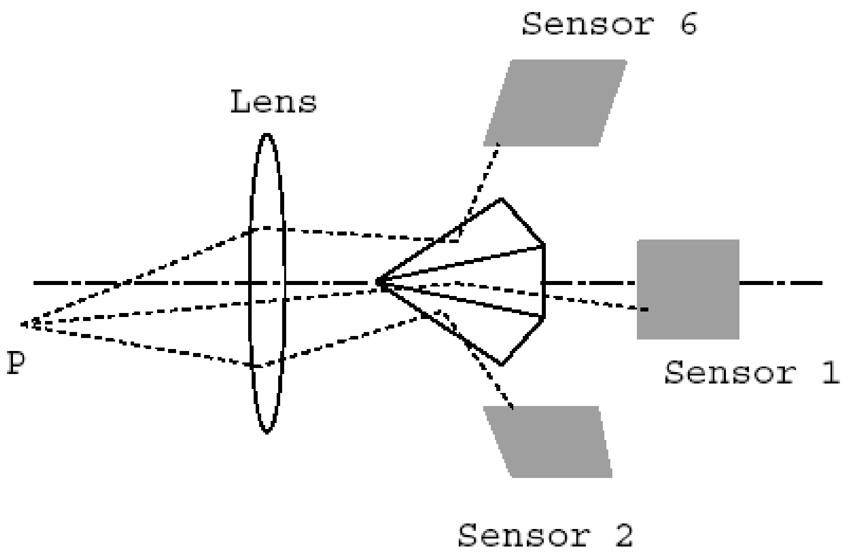
HDR Sensors



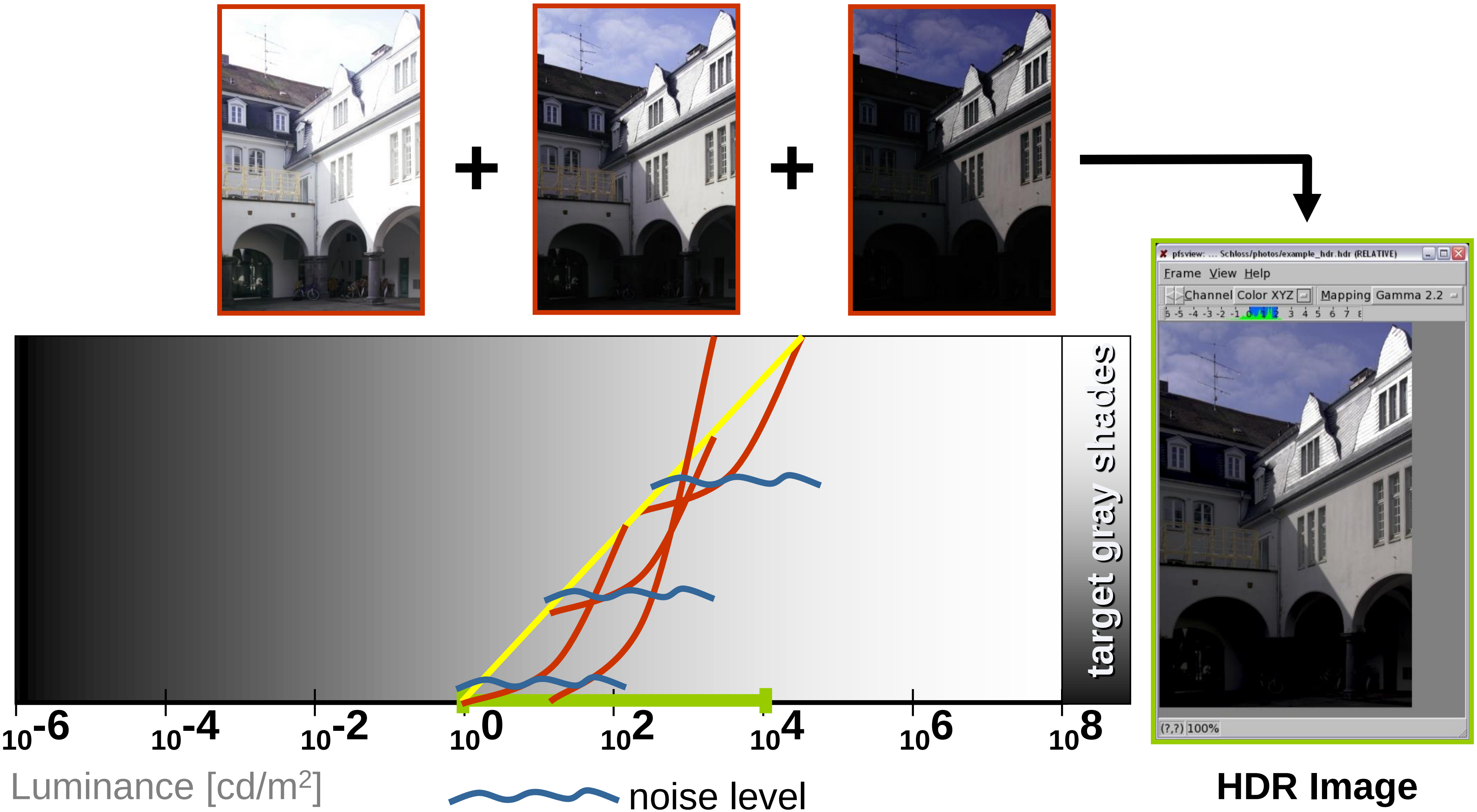
- logarithmic response
- locally auto-adaptive
- hybrid sensors (linear-logarithmic)

HDR Using Multiple Sensors

- semi-transparent mirror /prism
- multimple sensors with different sensitivity
- Panoscan Mark3, SpheronVR (scanning panoramic cameras), HDR video, HDR-Cam, etc.



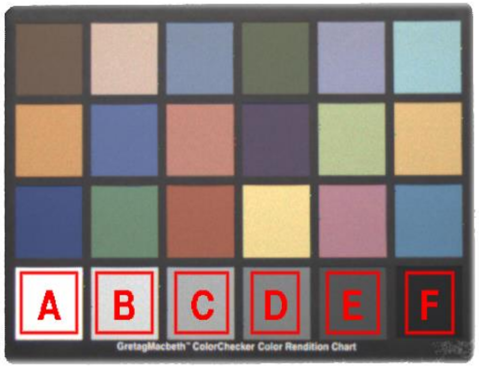
Multi-exposure Technique (1)



Photometric Calibration

- Converts camera output to luminance
  - requires camera response,
  - and a reference measurement for known exposure settings
- Applications
  - predictive rendering
  - simulation of human vision response to light
  - common output in systems combining different cameras

Photometric Calibration (cntd.)

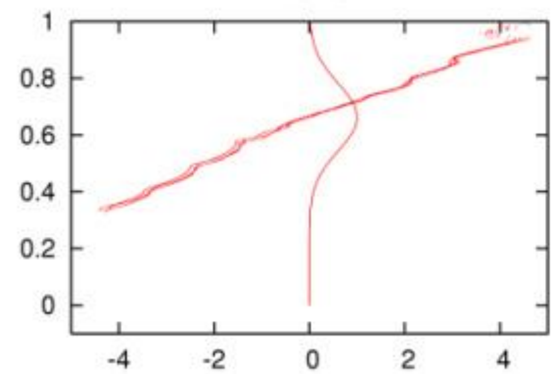


acquire target

camera output values

measure luminance

luminance values

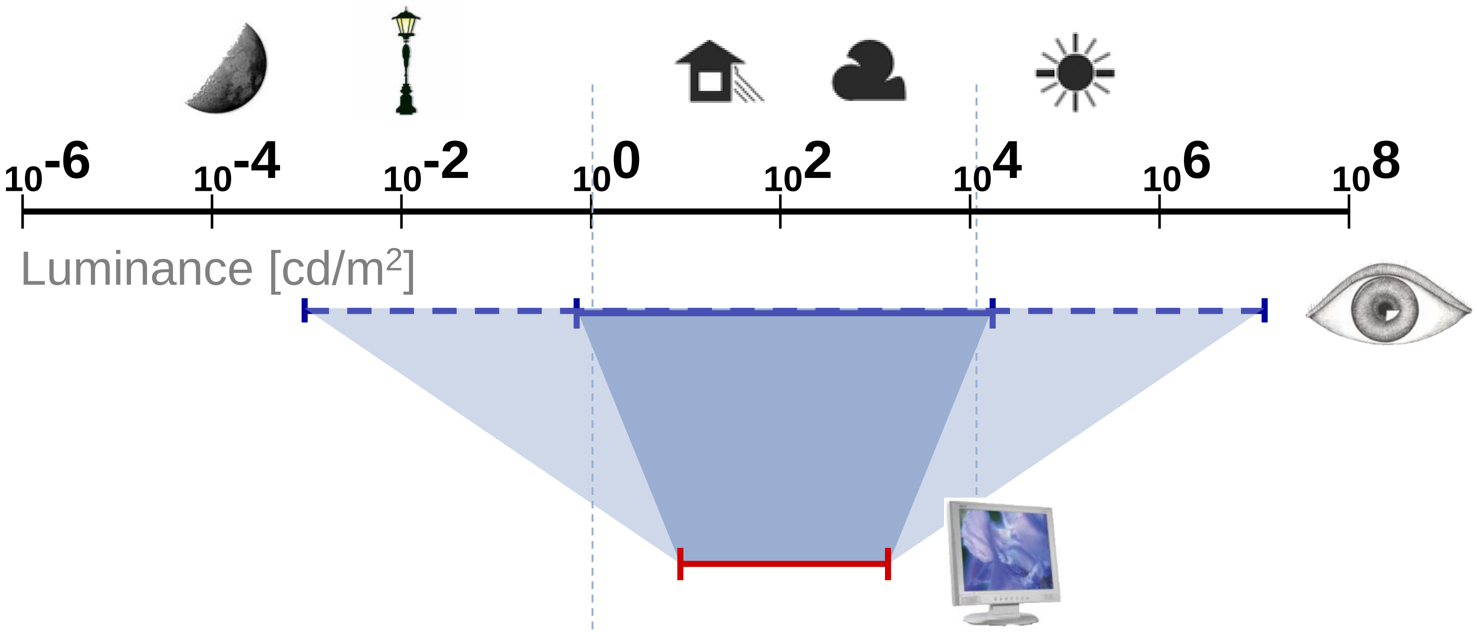


camera response

1.2 Overview

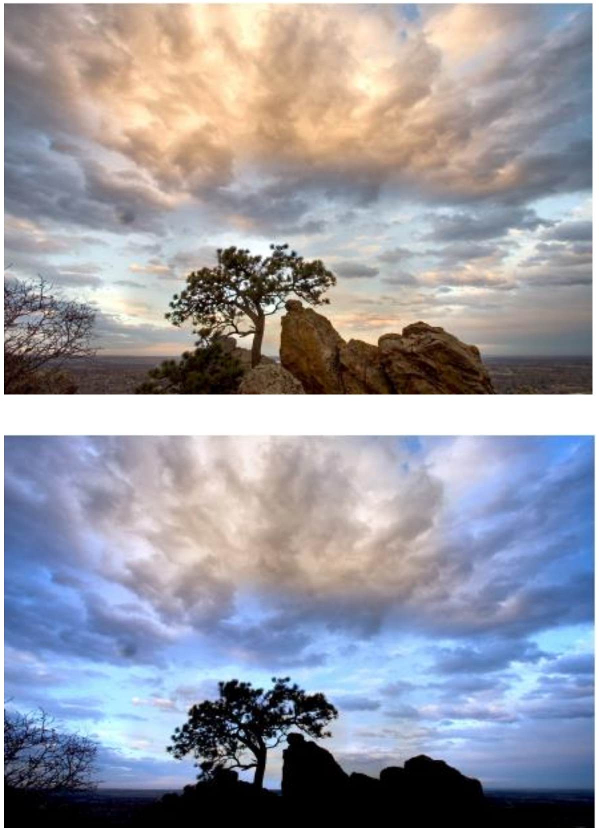
- Capture of HDR images and video
  - HDR sensors
  - Multi-exposure techniques
  - Photometric calibration
- Tone Mapping of HDR images and video
  - Early ideas for reducing contrast range
  - Image processing – fixing problems
  - Alternative approaches
  - Perceptual effects in tone mapping

HDR Tone Mapping

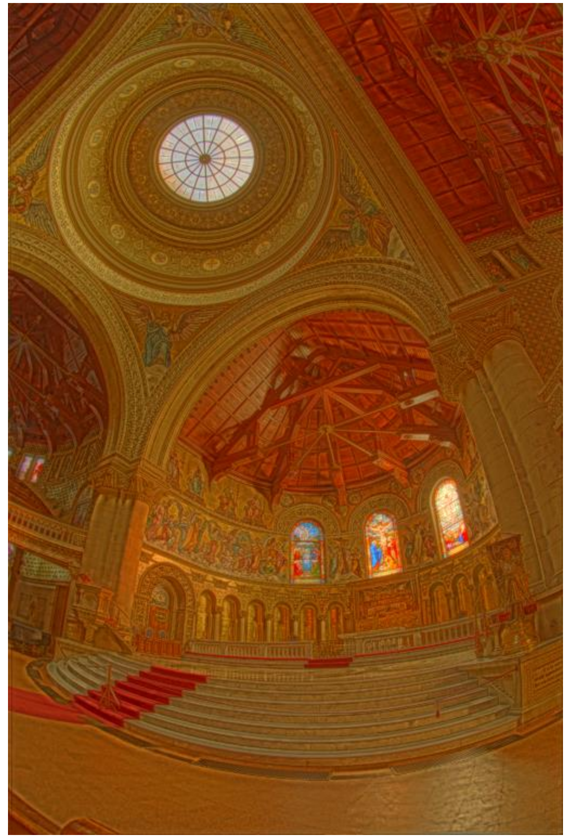


- Objectives of tone mapping
  - really application dependent…

Objective of Tone Mapping



Aesthetical

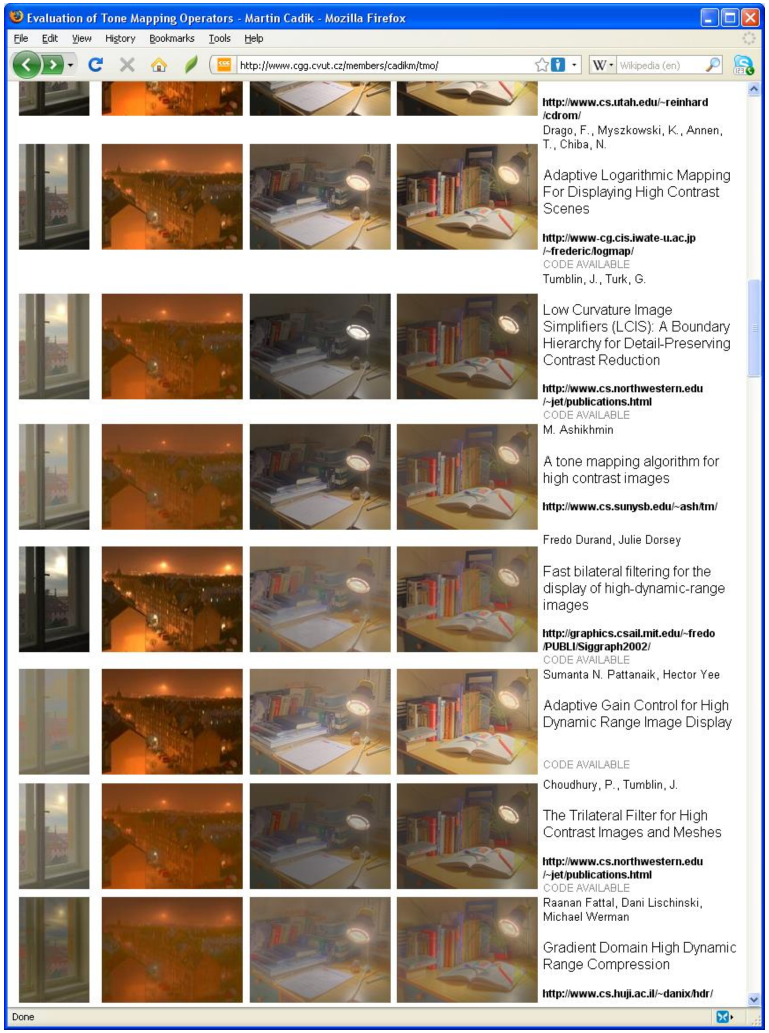


Cognitive

Perceptual

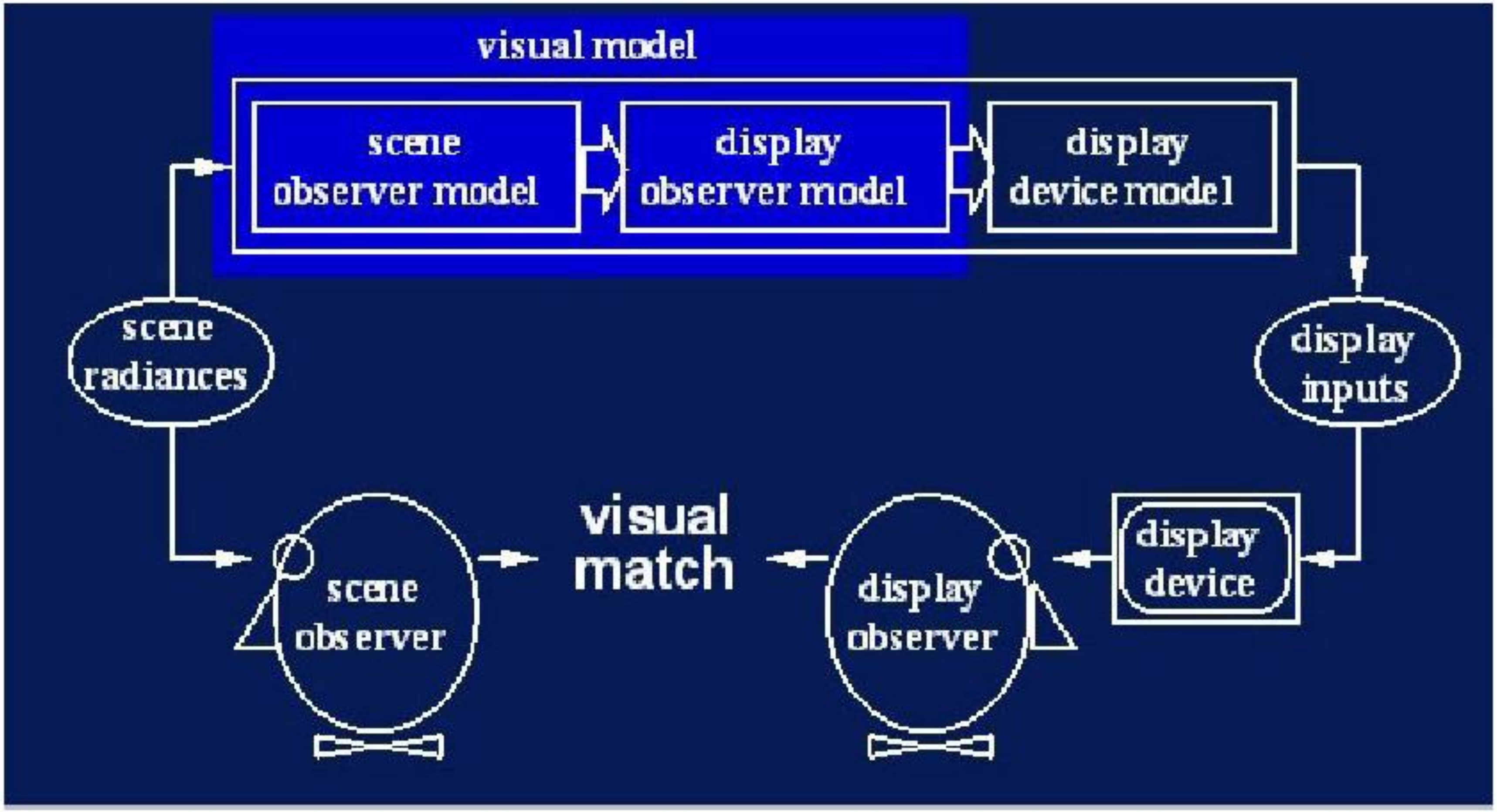
[Čadík et al. 06]

Large choice of Tone Mapping Methods



http://cadik.posvete.cz/tmo/

General Principle

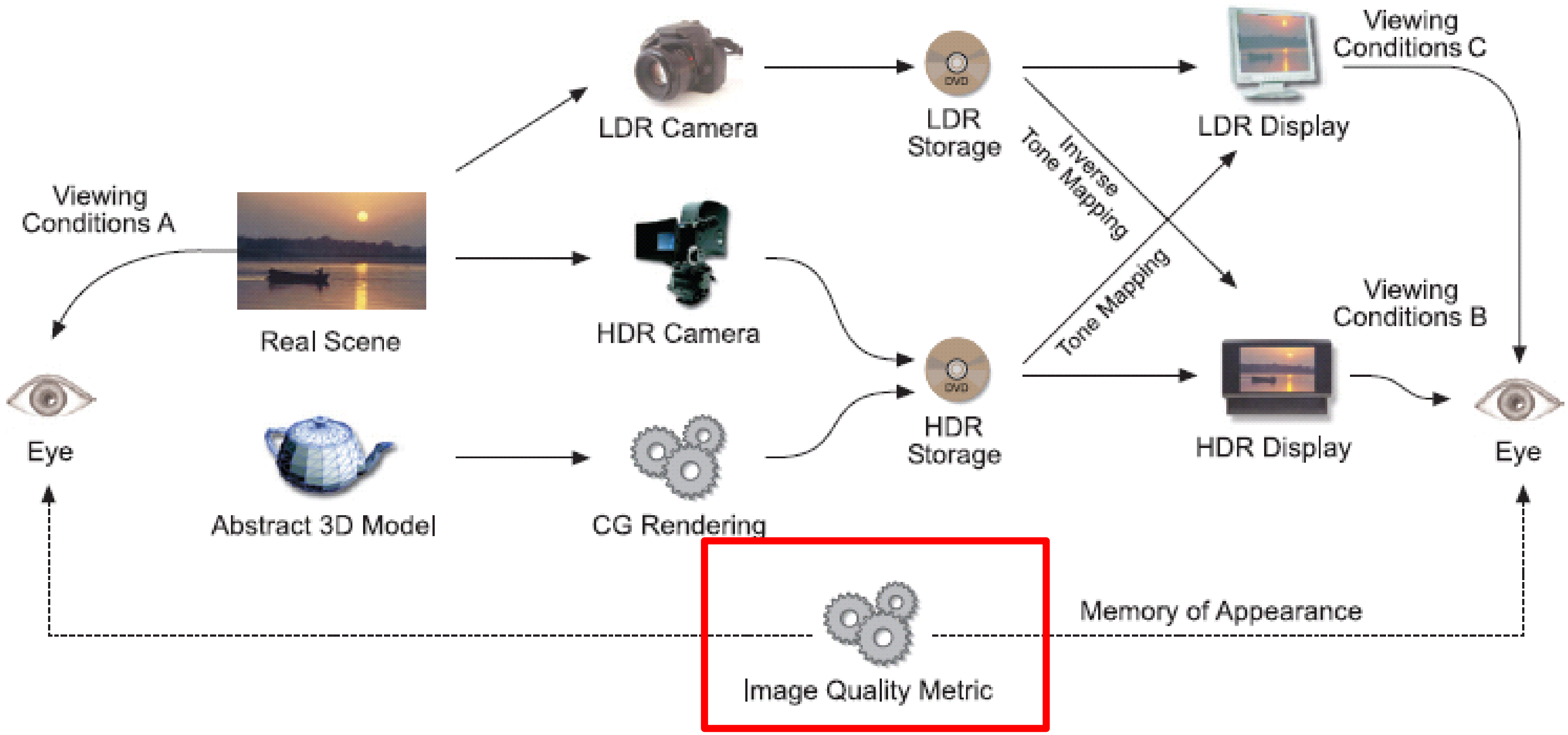


General Ideas

- Luminance as an input
  - absolute luminance
  - relative luminance (luminance factor)
- Transfer function
  - maps luminance to a certain pixel intensity
  - may be the same for all pixels (**global operators**)
  - may depend on spatially local neighbors (**local operators**)
  - dynamic range is reduced to a specified range
- Pixel intensity as output
  - often requires gamma correction
- Colors
  - most algorithms work on luminance
    - use RGB to Yxy color space transform
    - inverse transform using tone mapped luminance
  - otherwise each RGB channel processed independently

2) HDR IMAGE AND VIDEO QUALITY ASSESSMENT

HDR Pipeline



2.1 Overview

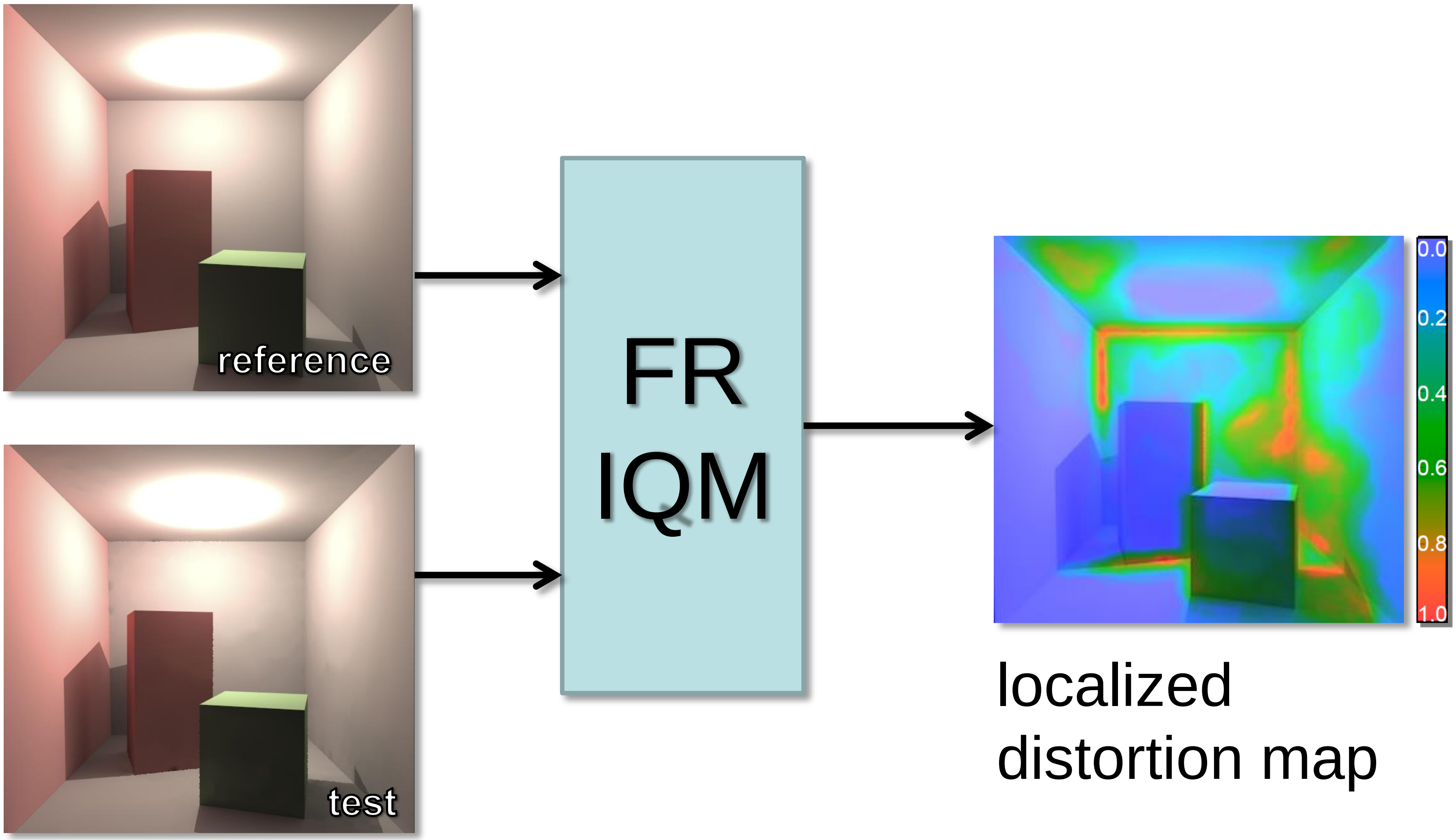
- Introduction to Objective Quality Assessment
- Image Quality Assessment
  - HVS-based Metrics (bottom-up)
  - Structural Similarity (top-down)
  - Data-driven Approaches (top-down)
- Video Quality Assessment

FR Quality Assessment (IQA, VQA)

Rate
the
Quality

+ Reliable **- High cost**

Full-Reference Image Quality Metrics



localized distortion map

Full-Reference Metrics

- What are they good for?
  - Quality assessment scenarios in compression/transmission, etc.
  - Algorithm analysis/validation/evaluation

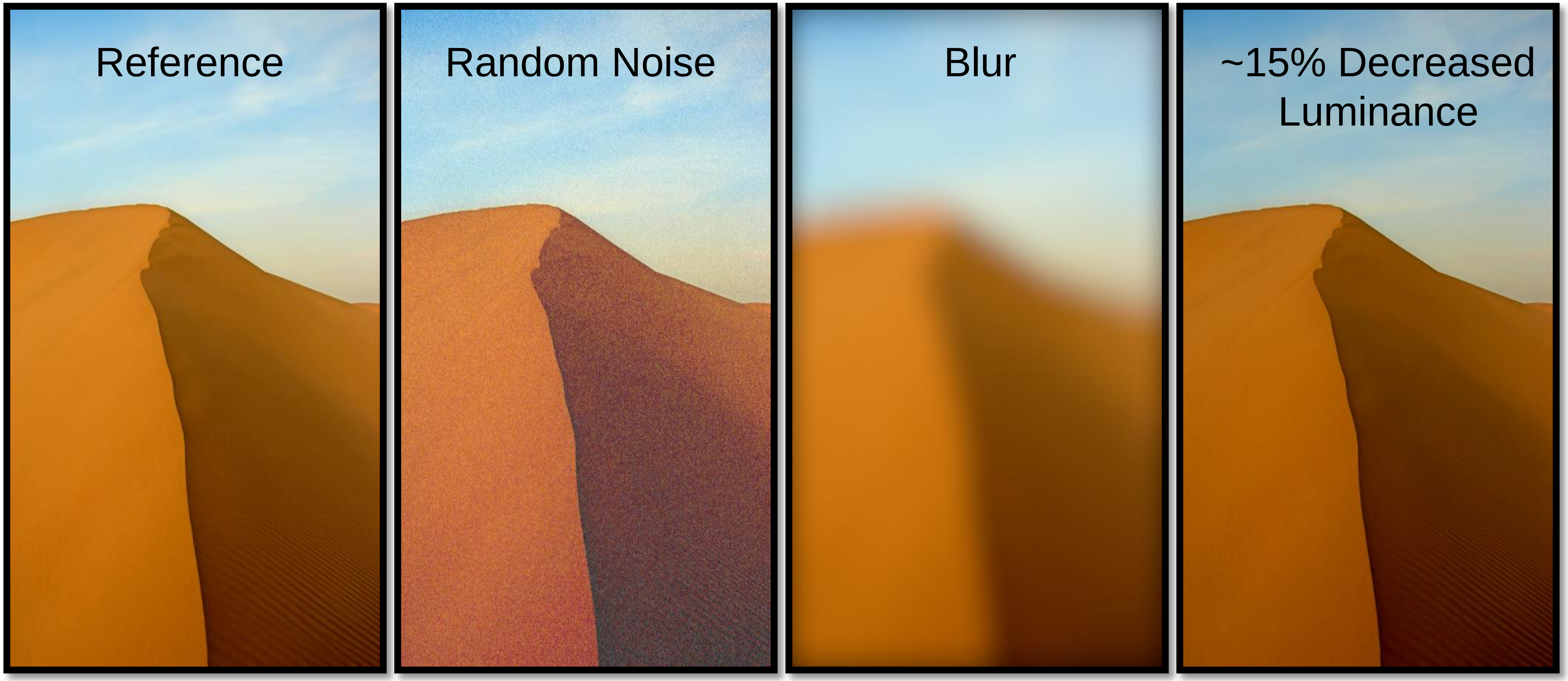
  - Guiding/ parameter estimation of renderers
  - Stopping criterions
  - Speed/ quality enhancements

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 per block)

Simple Full-reference Metrics

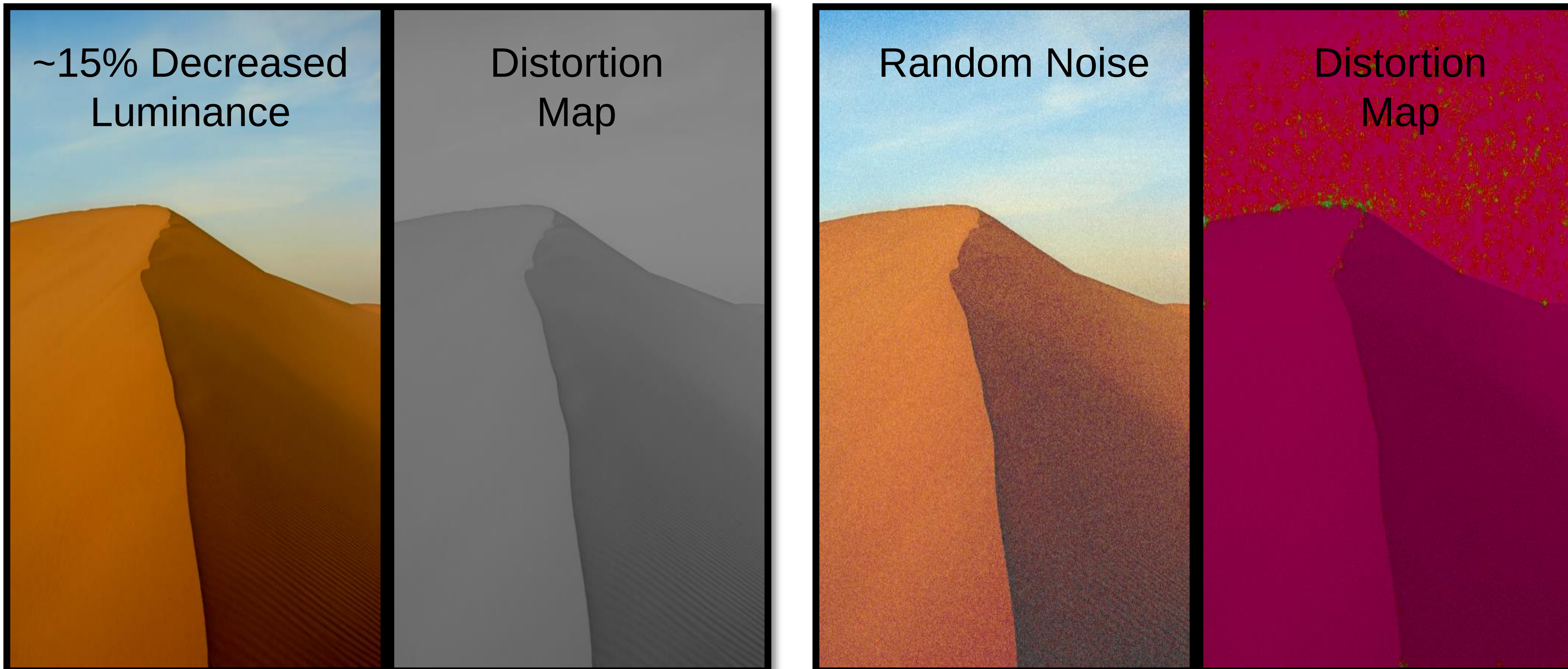


MSE = 280 MSE = 280 MSE = 280 !

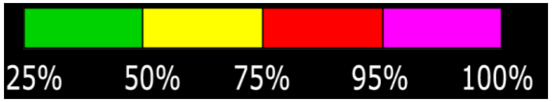
2.2 Overview

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

HVS Based Metrics



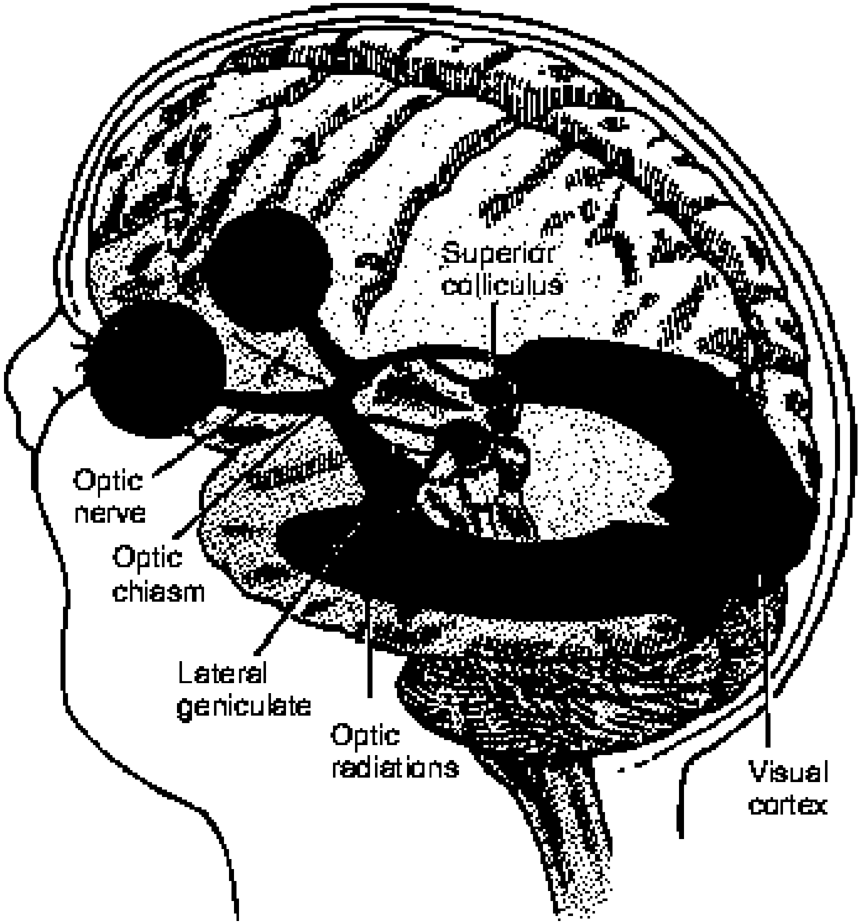
Probability of Detection:



Human Visual System

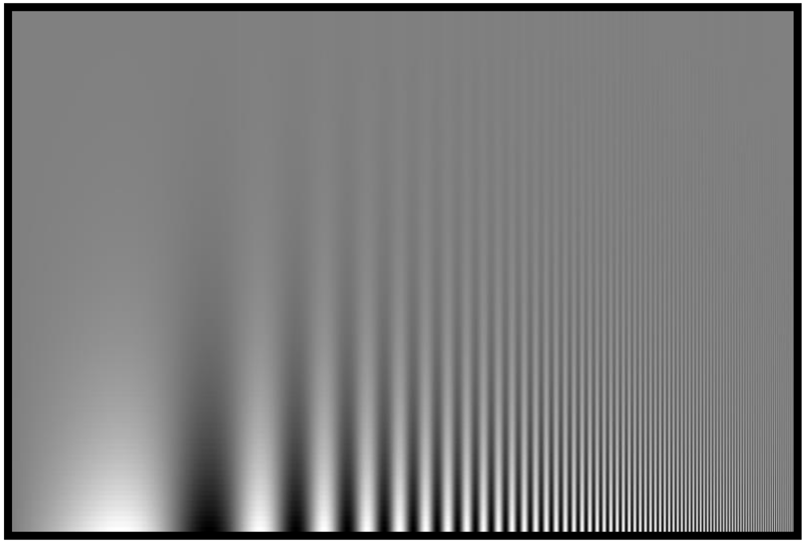
- Physical structure
  well established
  (early vision)

- High-level vision
  still not fully understood

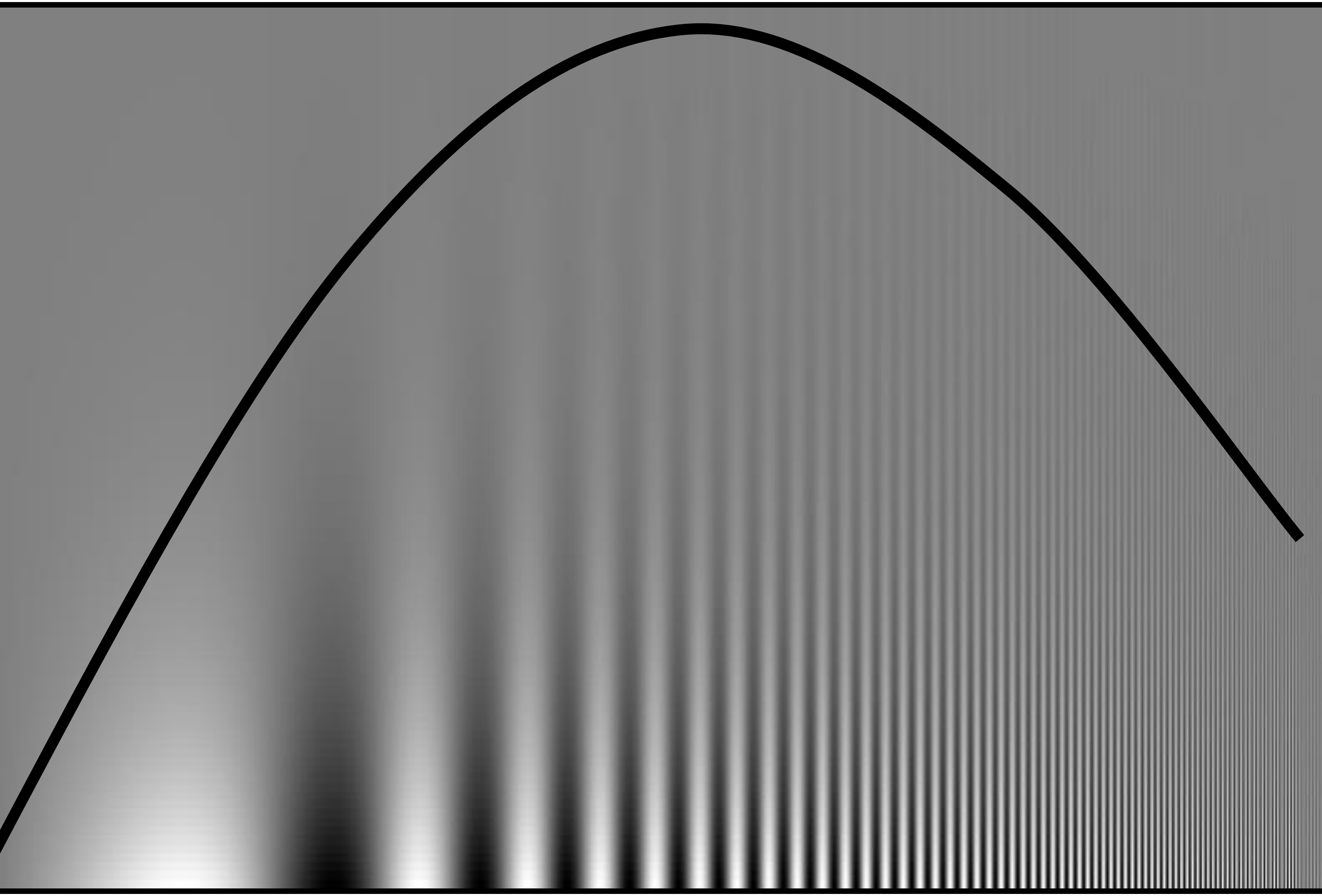


Human Visual System (cont.)

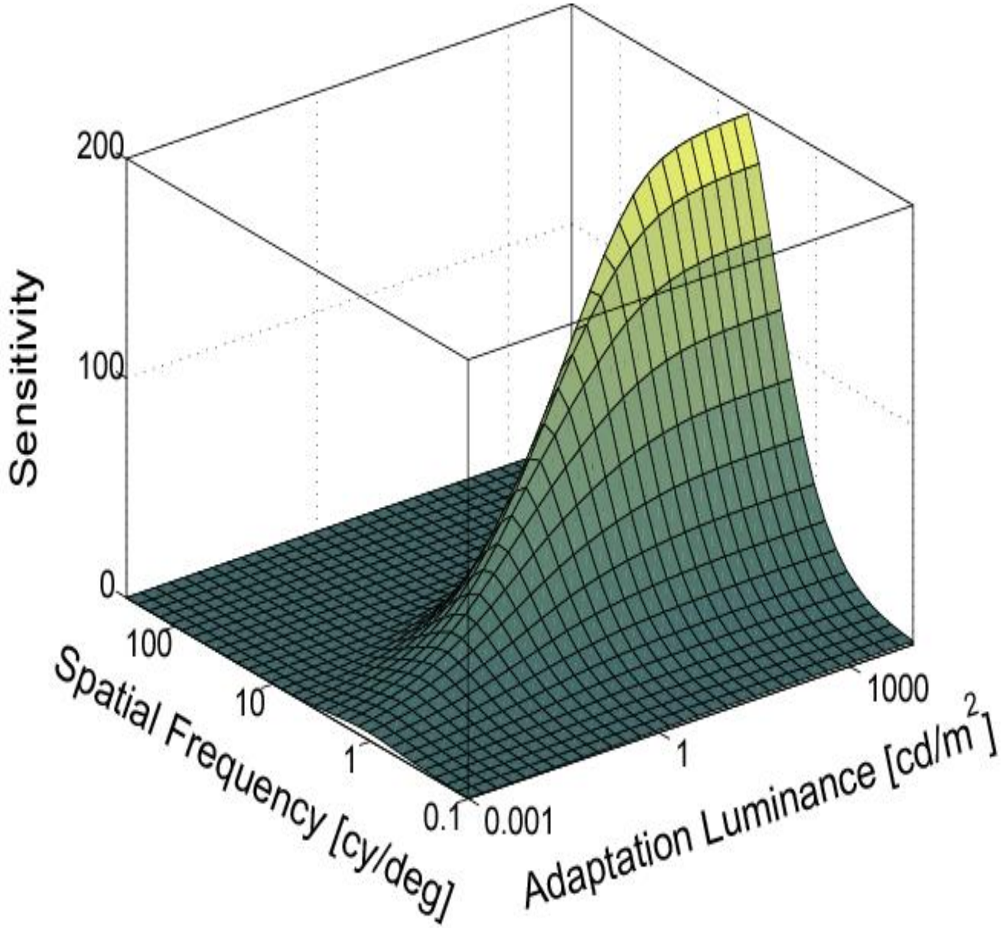
- CSF
  - specifies the sensitivity (1/detection threshold) as a function of the spatial frequency
  - depends on
    - spatial frequency
    - adaptation level
    - temporal freq.
    - orientation
    - viewing dist
    - eccentricity, …



[Campbell and Robson 1968]

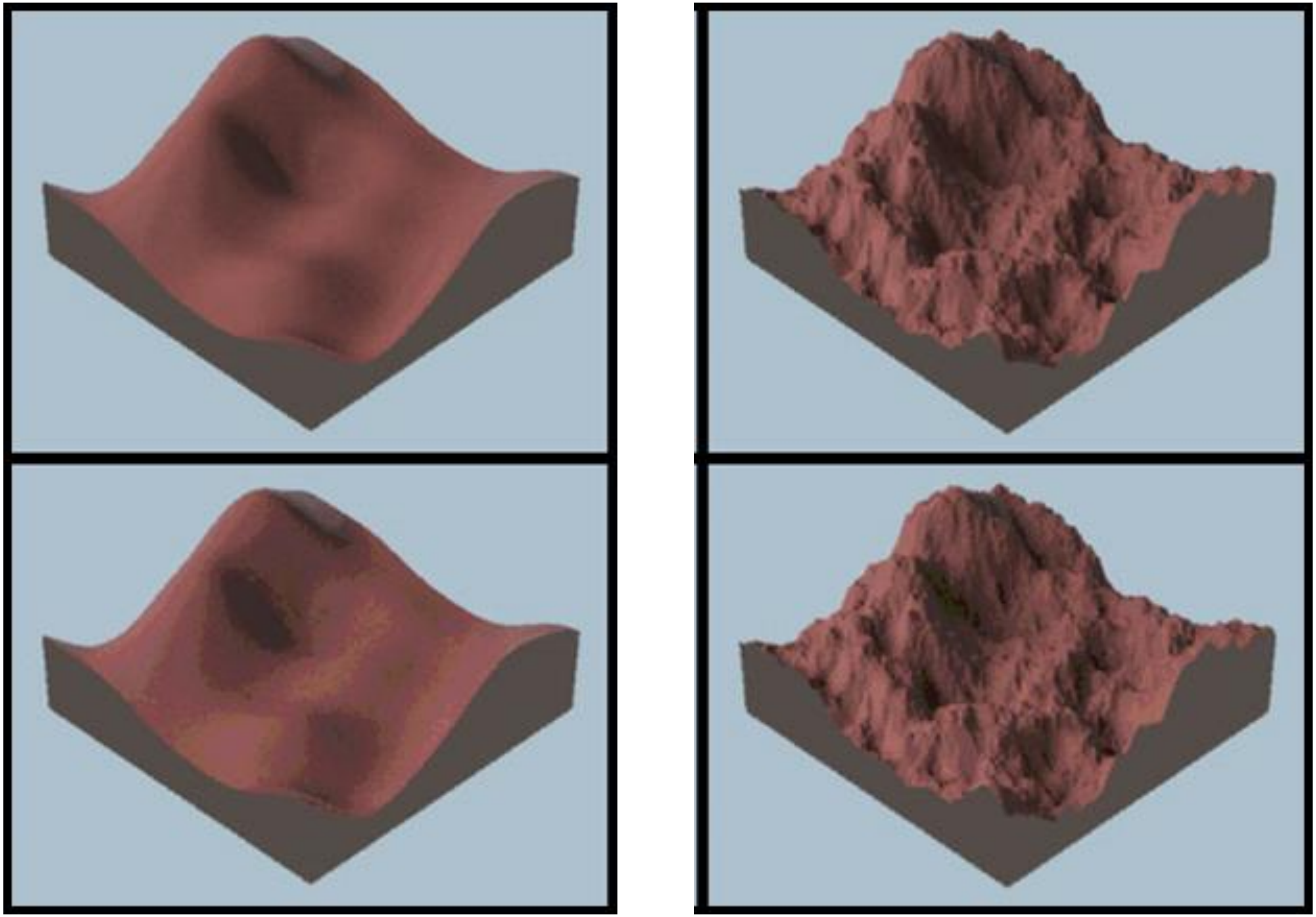


Contrast Sensitivity Function (CSF)



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

HVS – Visual Masking

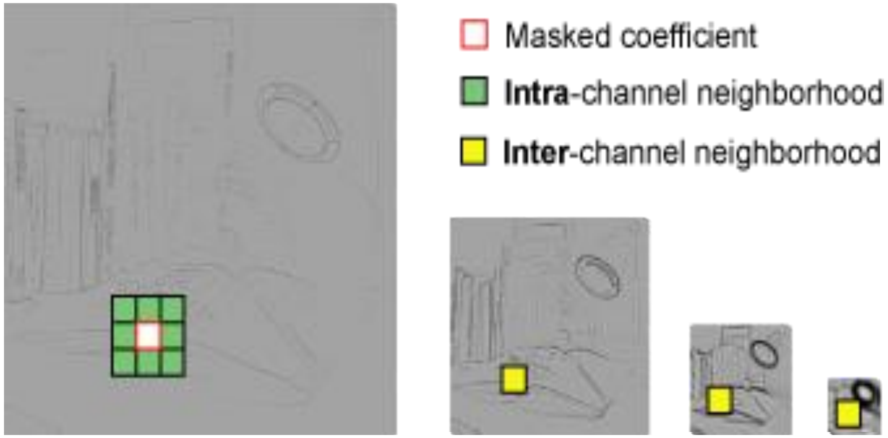


Visual Masking

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

Modeling Visual Masking

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

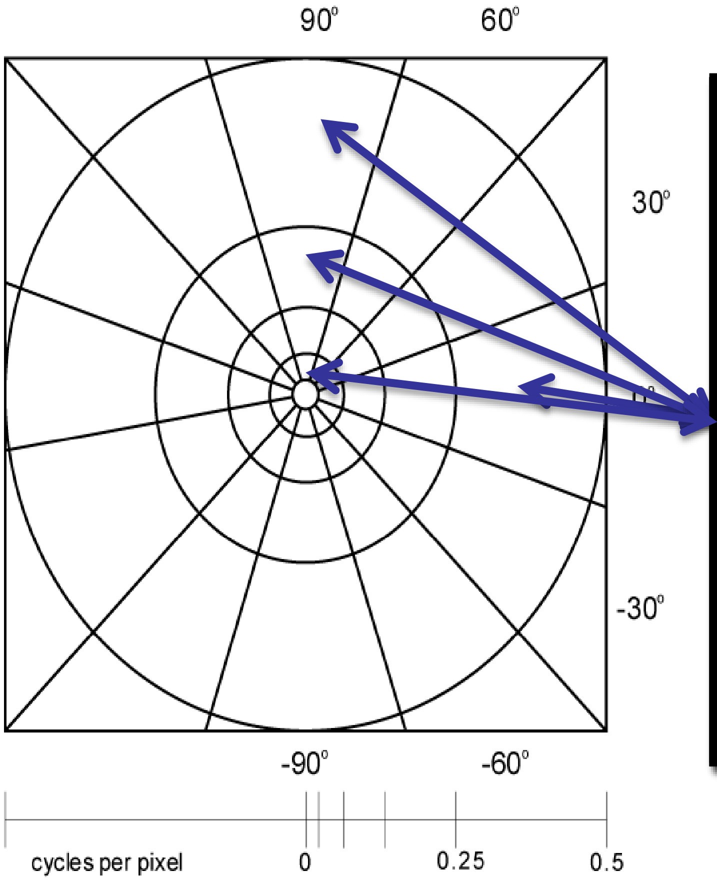


$$R = \frac{sign(C')|C'|^{0.5}}{(1 + \sum_K |C'_k|^{0.2})}$$

C': Normalized Contrast

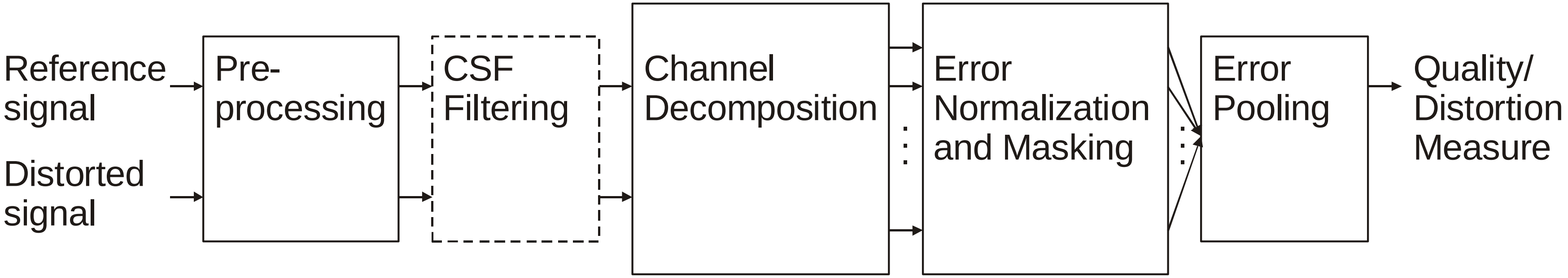
Modeling Visual Masking - Visual Channels

Cortex Transform



Error Sensitivity Based Approaches

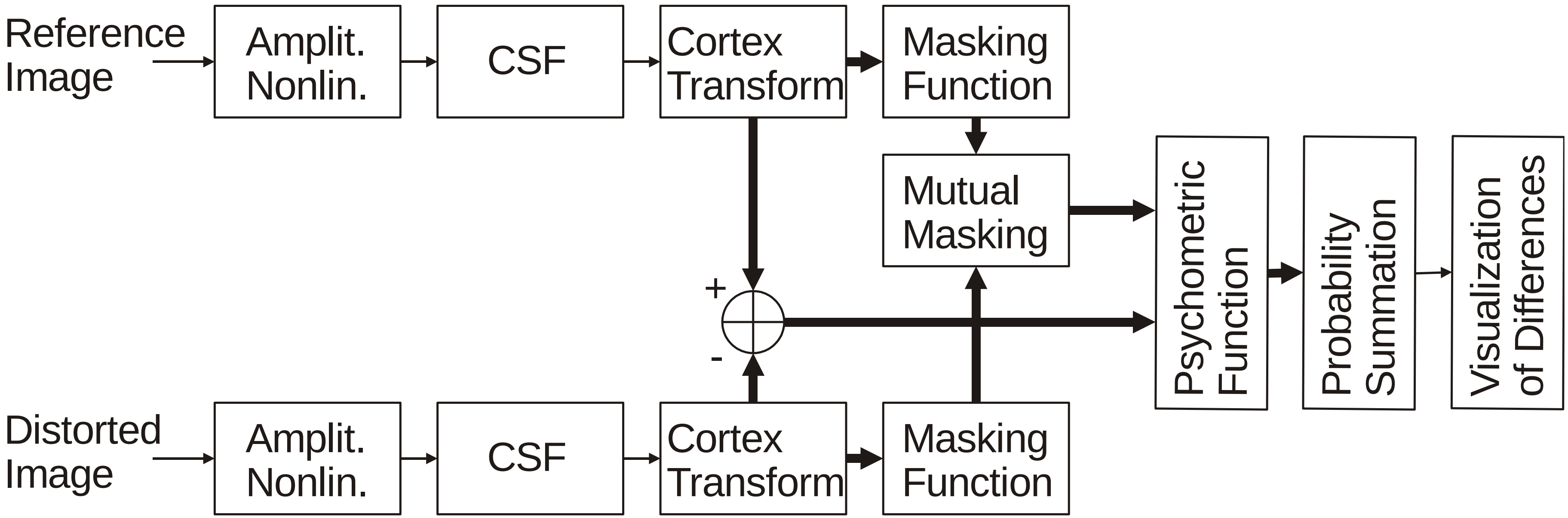
- General framework



- Visible Differences Predictor [Daly93]
- Perceptual Distortion Measure [Teo, Heeger 94]
- Visual Discrimination Model [Lubin 95]
- Gabor pyramid model [Taylor et al. 97]
- WVDP [Bradley 99]
- HDRVDP2 [Mantiuk et al. 05, Mantiuk et al. 11]

Error Sensitivity Based Approach

- Visible Differences Predictor (VDP) [Daly 93]

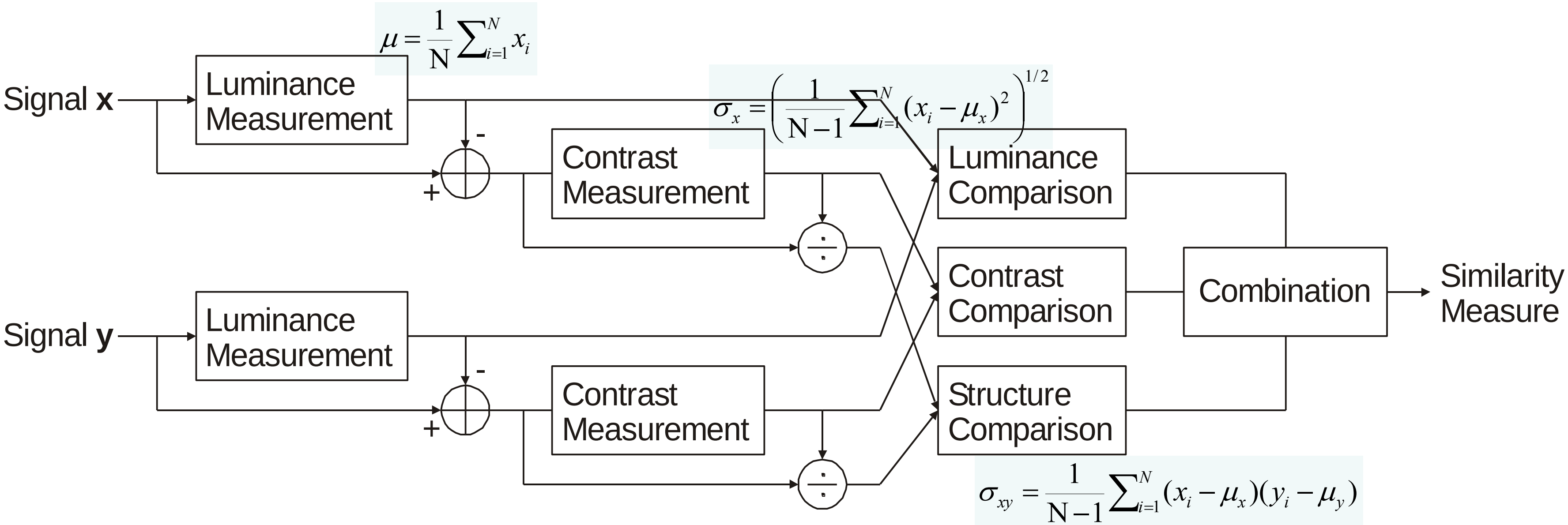


- Threshold sensitivity
- Early vision modeling
- Visual Masking

2.2 Overview

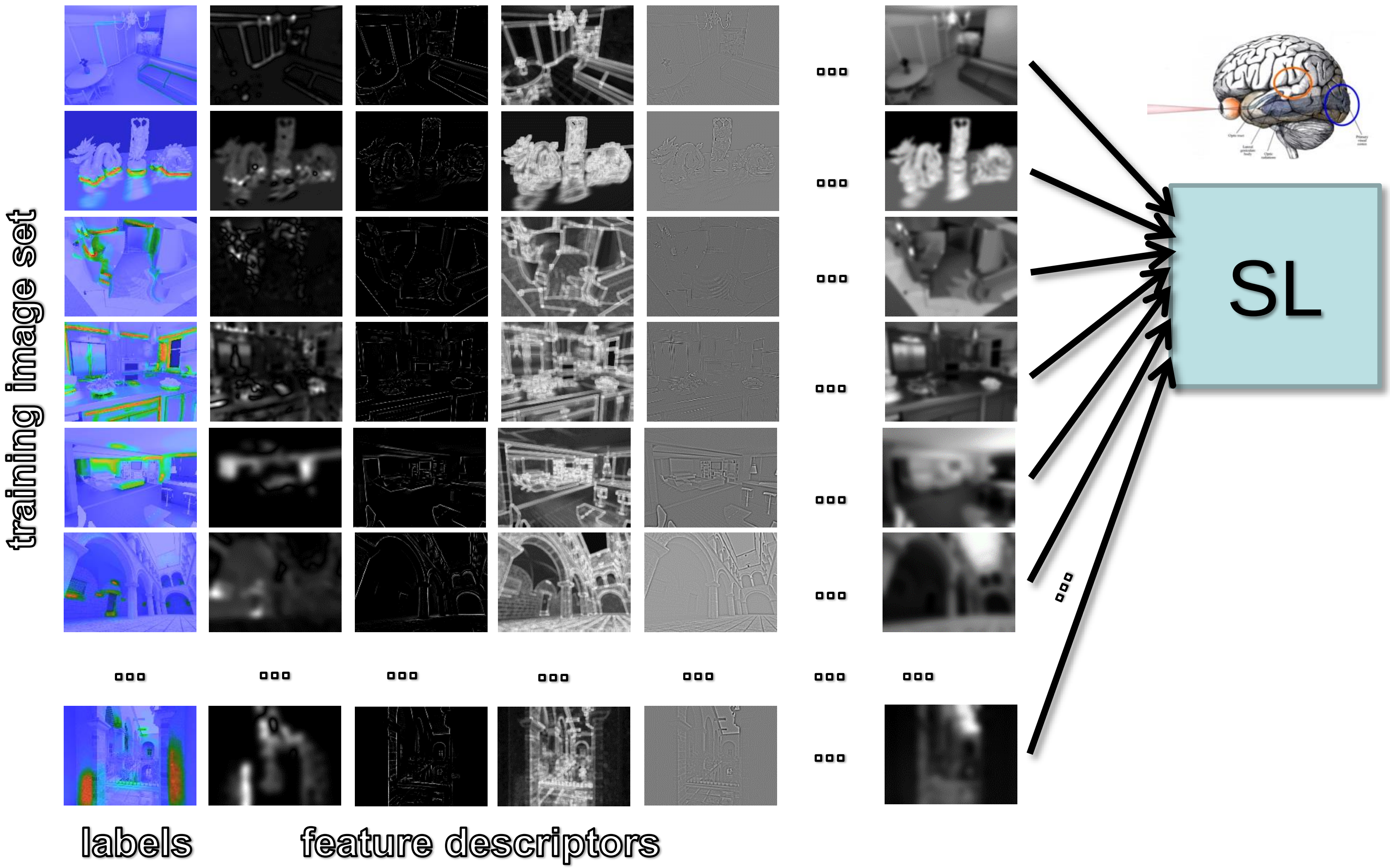
- Introduction to Objective Quality Assessment
- Image Quality Assessment
  - HVS-based Metrics (bottom-up)
  - Structural Similarity (top-down)
  - Data-driven Approaches (top-down)
  - HDR IQM
- Video Quality Assessment

Structural Similarity-Based Approaches

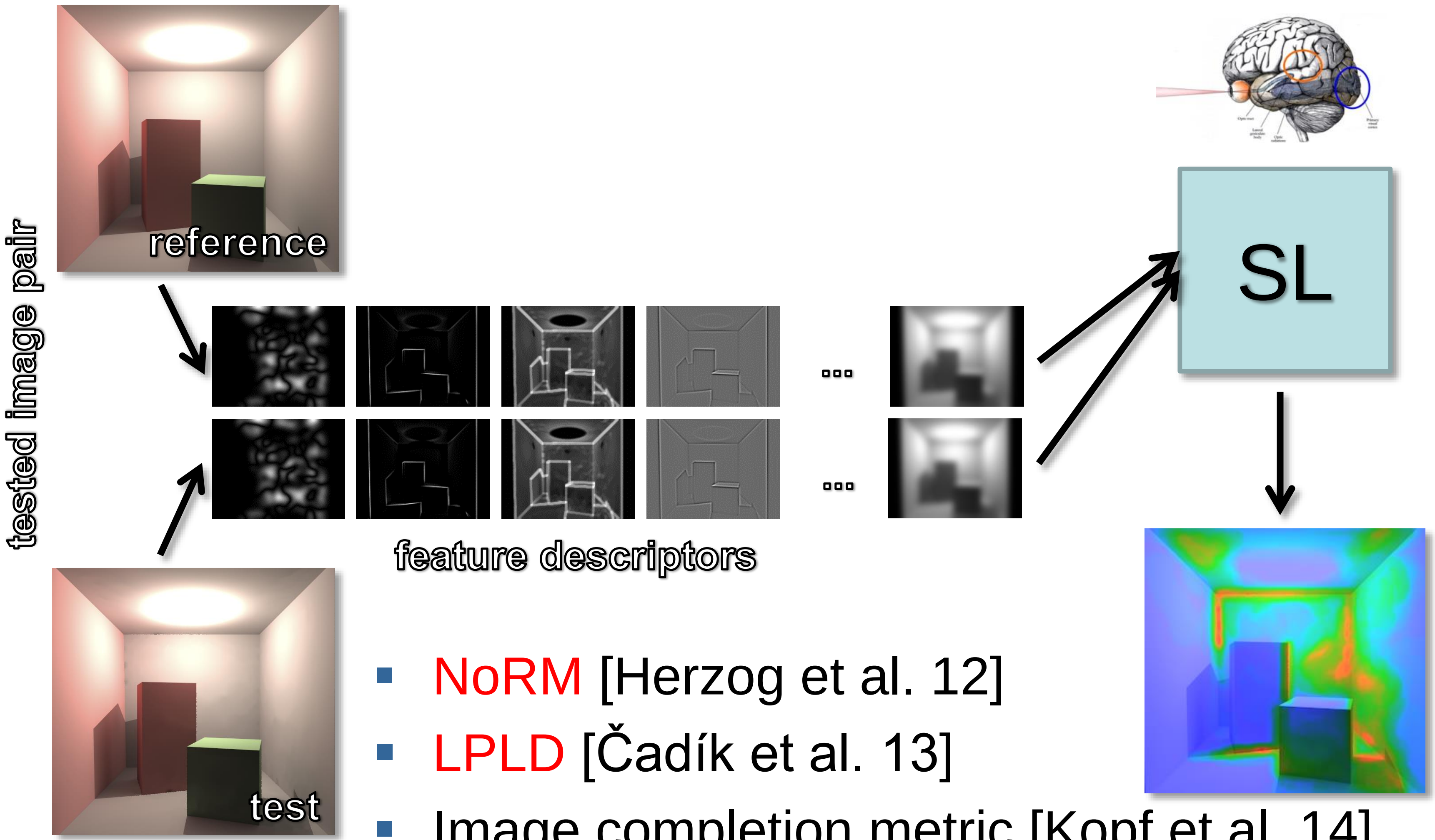


- UQI [Wang 02]
- SSIM [Wang 04]
- M-SSIM [Wang et al. 04]
- Multidimensional Quality Measure Using SVD [Shnayderman 04]

Supervised Learning – Training Phase



Supervised Learning – Prediction



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

No-Reference Image Quality Metrics

- Detecting blockiness in JPEG/MPEG
  - [Wang & Bovik ’06, Wu & Rao ‘05]
- Blurriness measure
  - [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]

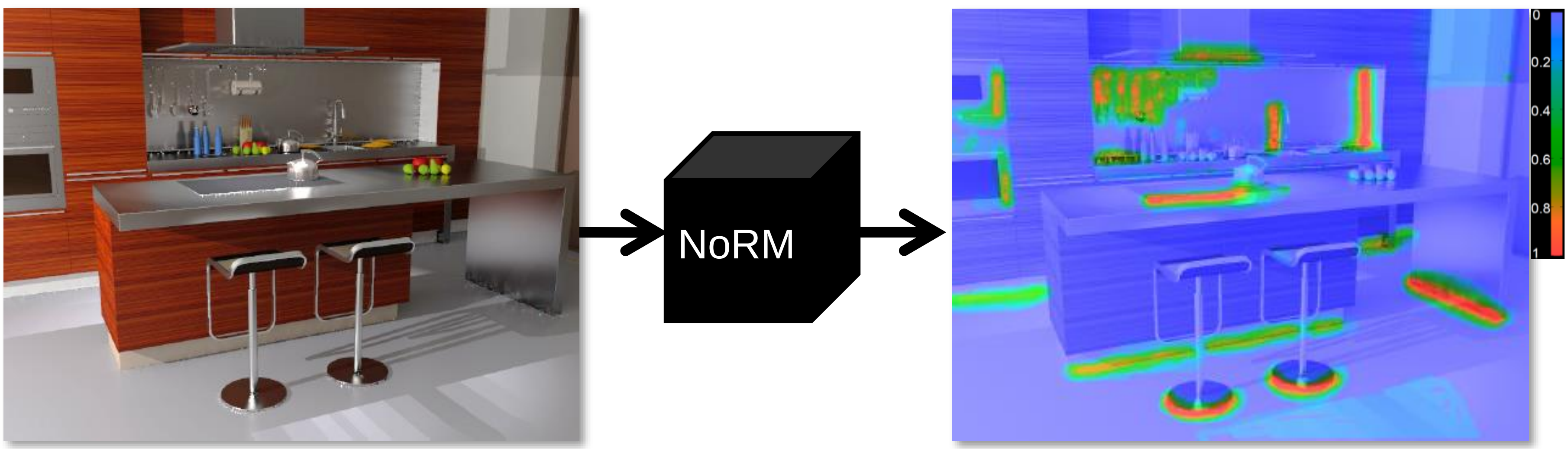
4.2

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

NoRM: No-Reference Metric

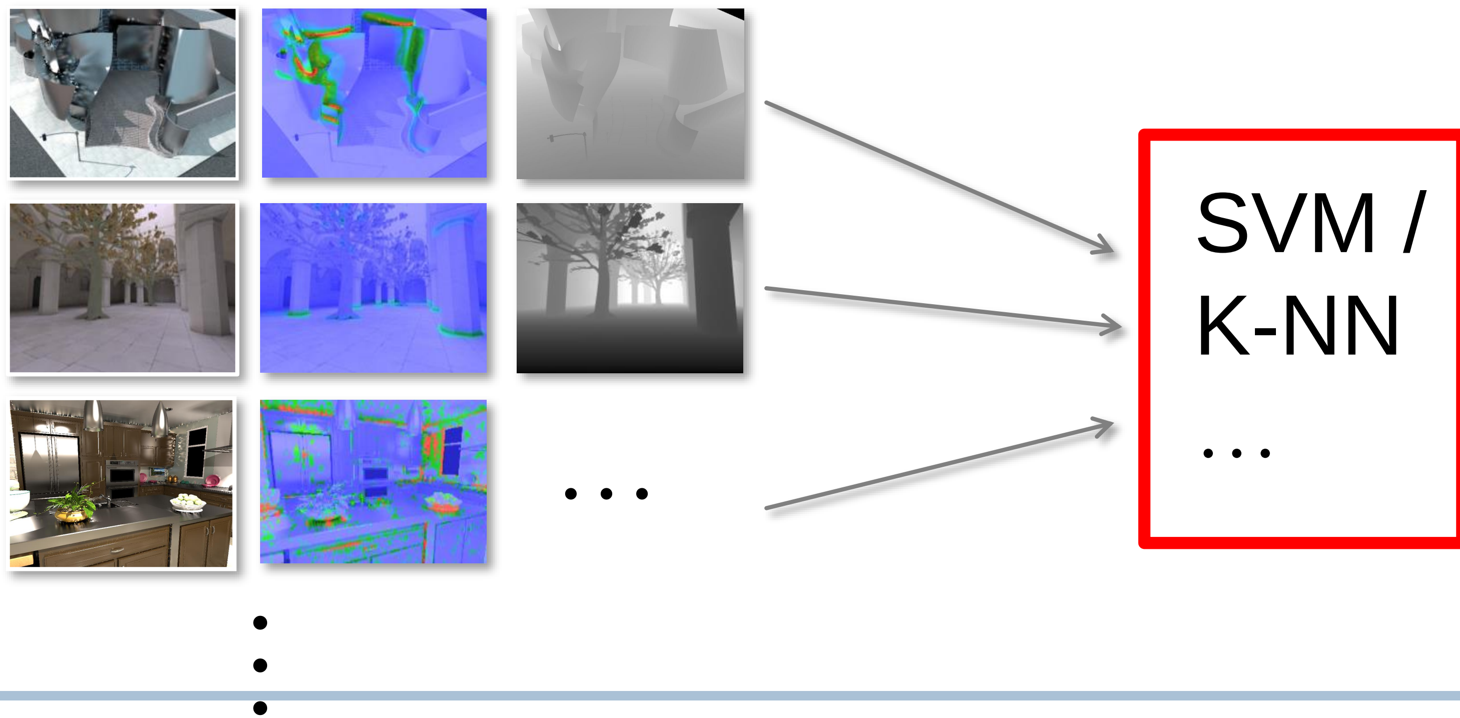
[Herzog et al. 2012]

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



Data-Driven No-Reference IQM

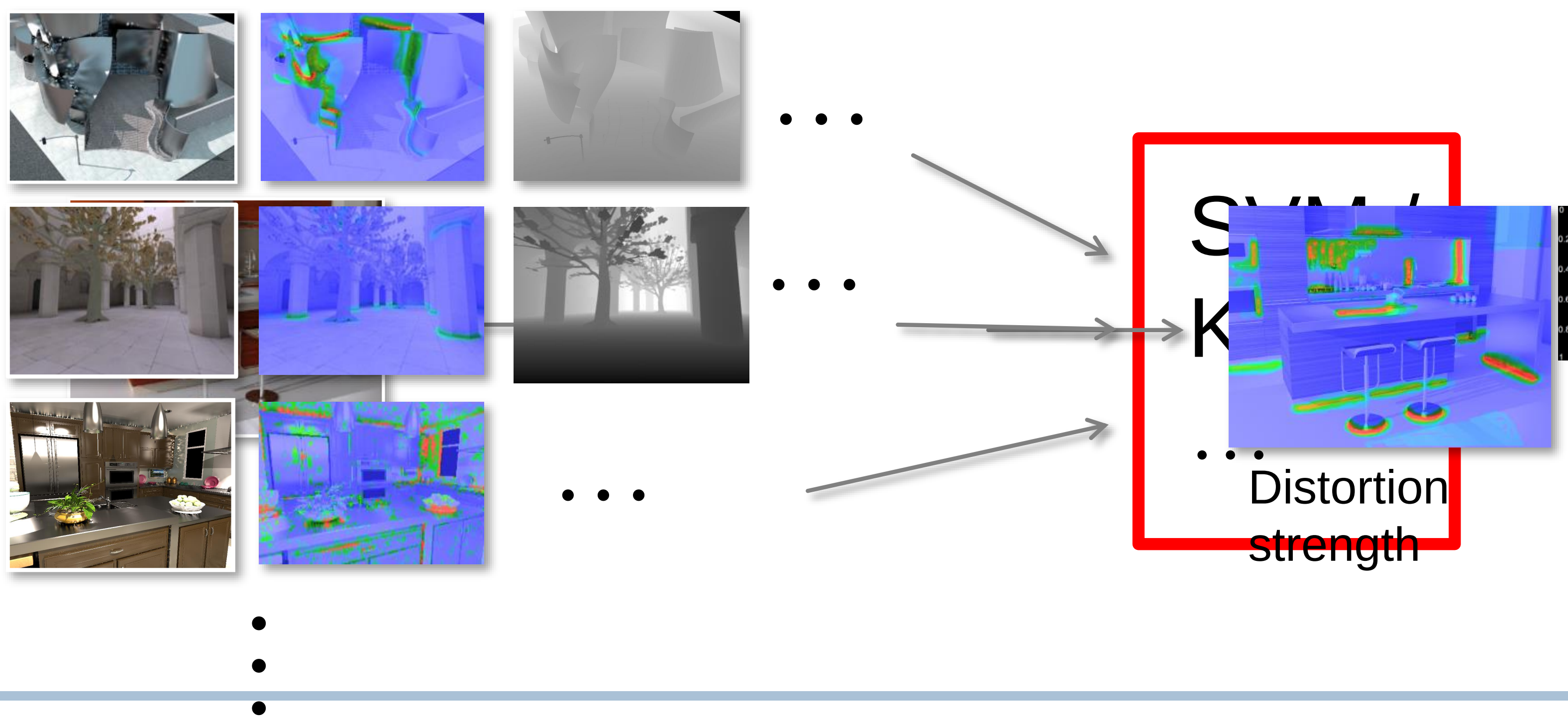
- Feature descriptors (various information available)
- Distortion maps (possibly real subjective data)
- Depth + 3D related information



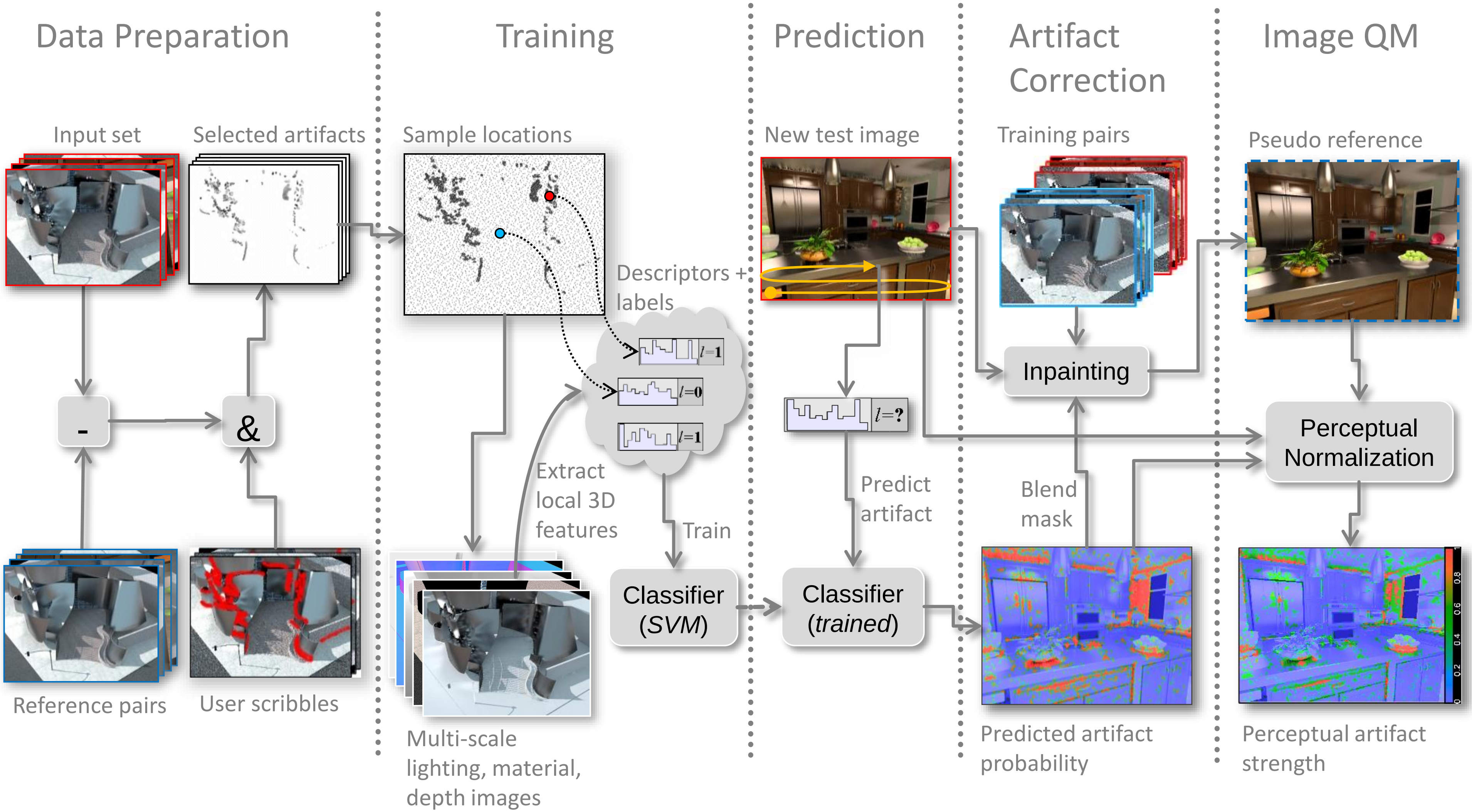
Data-Driven No-Ref. IQM

- Distorted (rendered) image → prediction
  - Traditional metrics: just a number on scale 1-5
- We want a distortion map per pixel
  - Much harder problem
  - But ... we have 3D data!!!

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

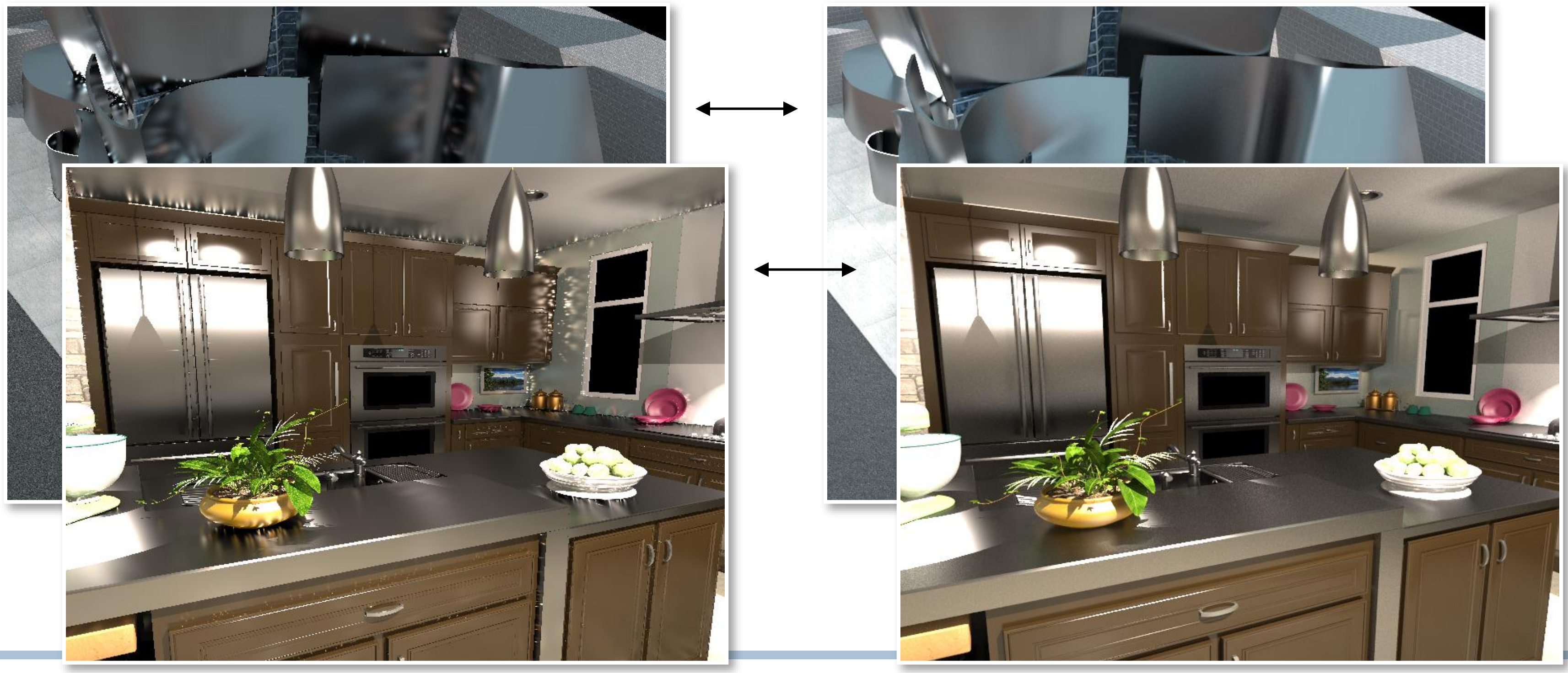


System Pipeline NoRM

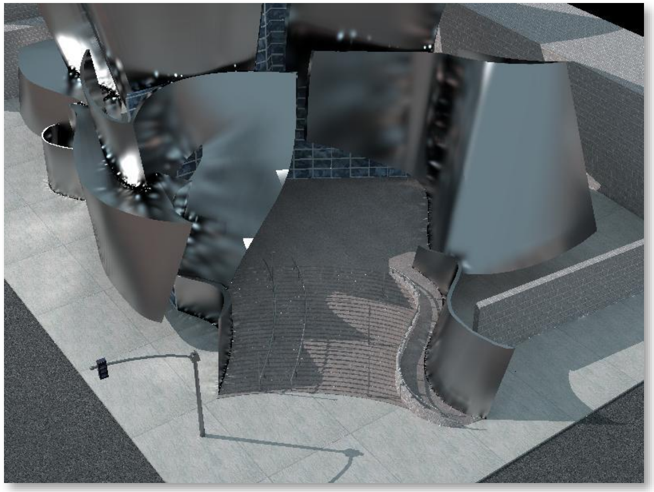


Training Classifier

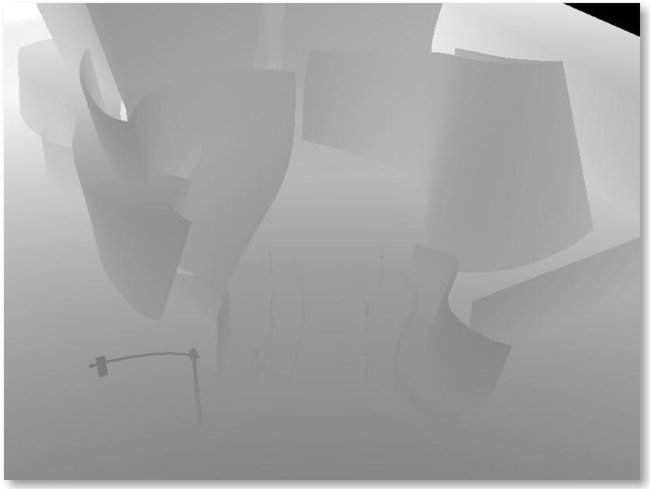
- Given input data:
  - color, depth, material for one artifact type
  - user scribbled artifact mask
  - reference image without artifacts



Rendering Output – Classification Input



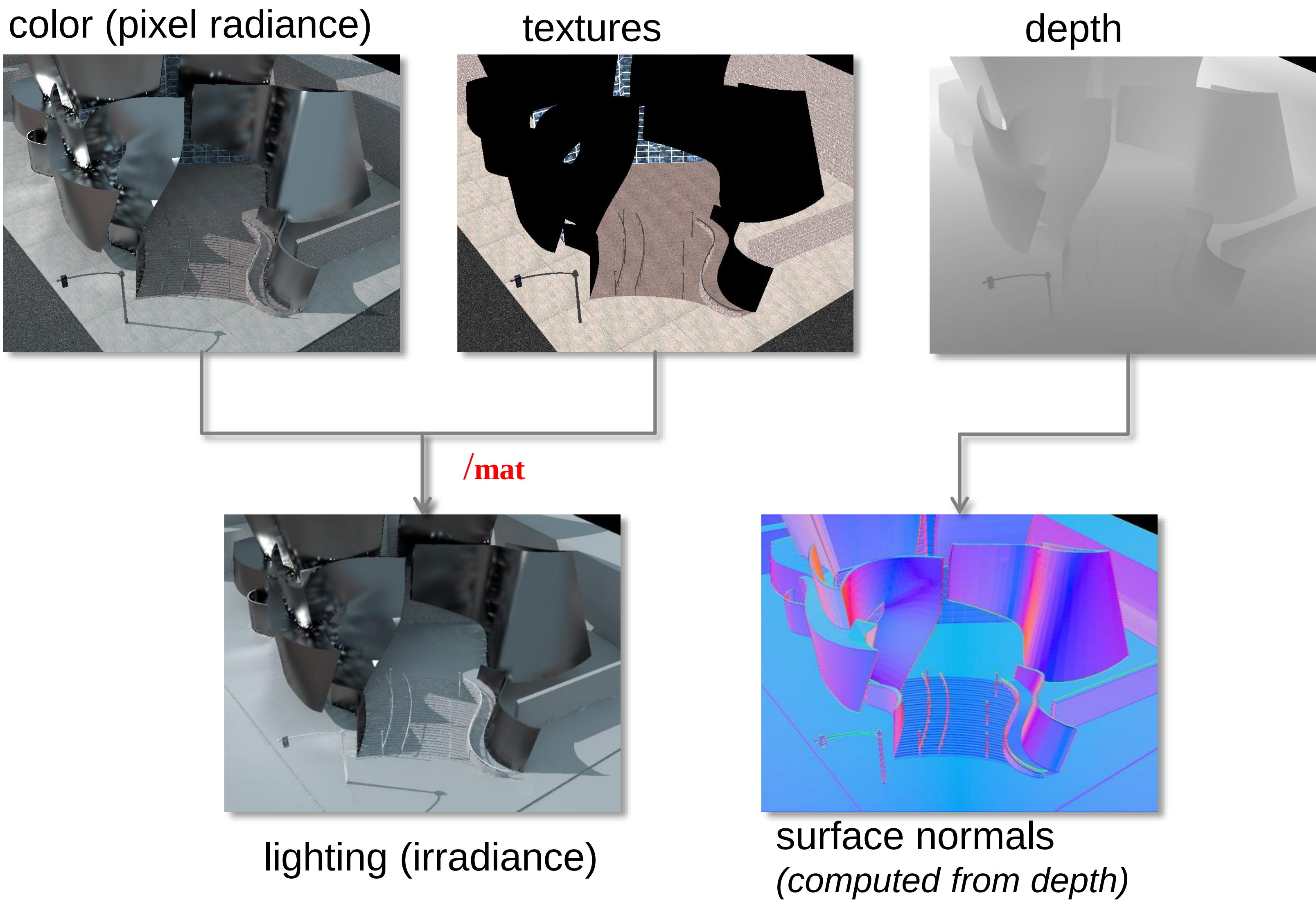
HDR (LDR) color image
(may contain noise)



depth buffer
(in high precision, no noise)

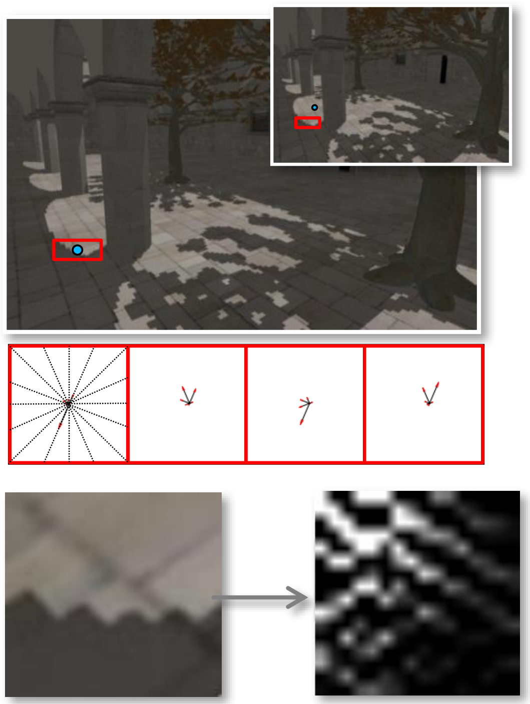
diffuse texture buffer

Computation of Additional Input Data



Feature Descriptors

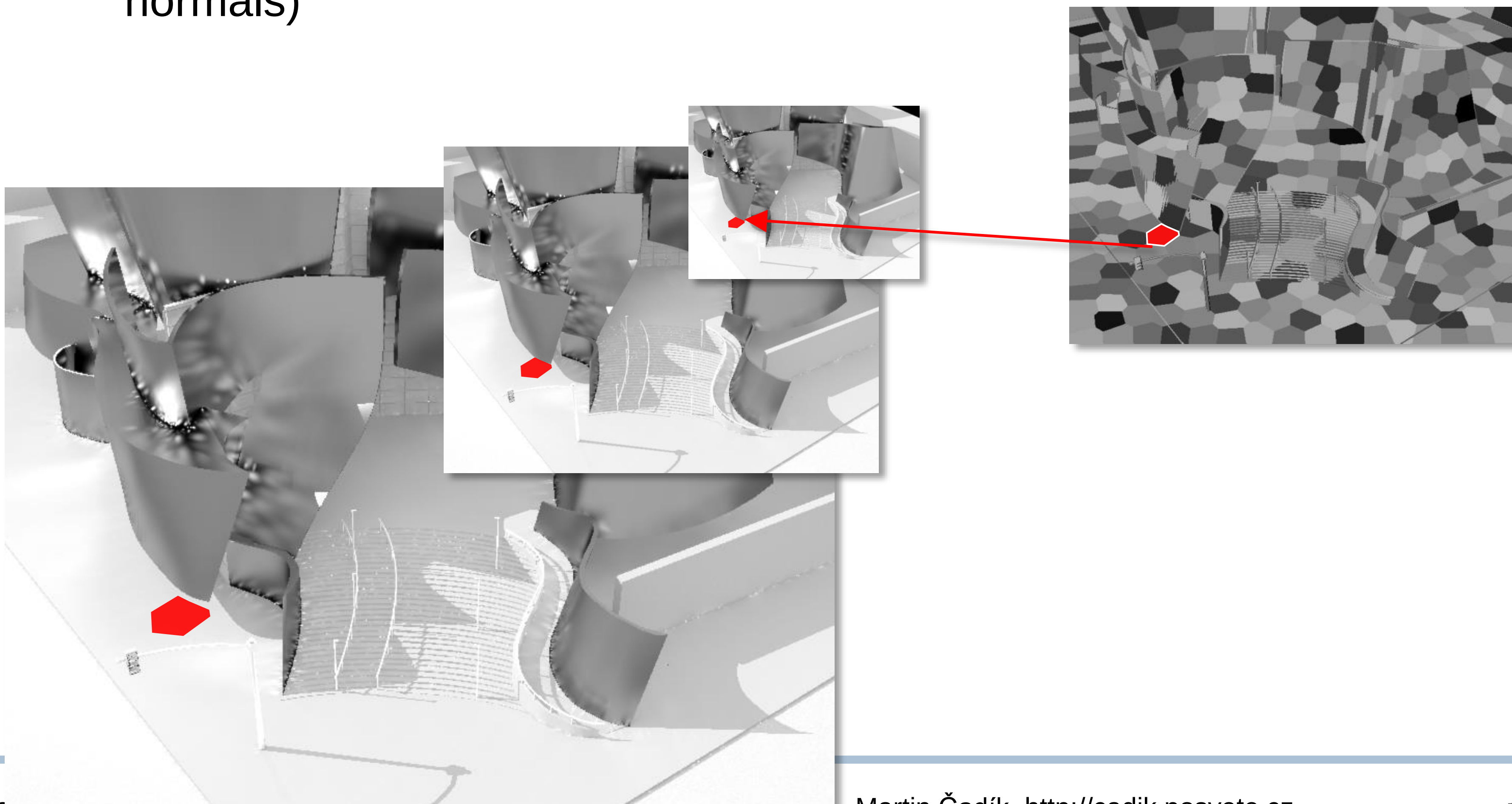
- Tested several “standard” features
- Color-features from computer vision
  - Histogram of oriented Gradients (HoG)
  - Frequency domain features (DCT)
  - Difference of Gaussians (DoG)
  - Local first-order statistics
- Plus 3D features given depth



Cropped block DCT coeff.

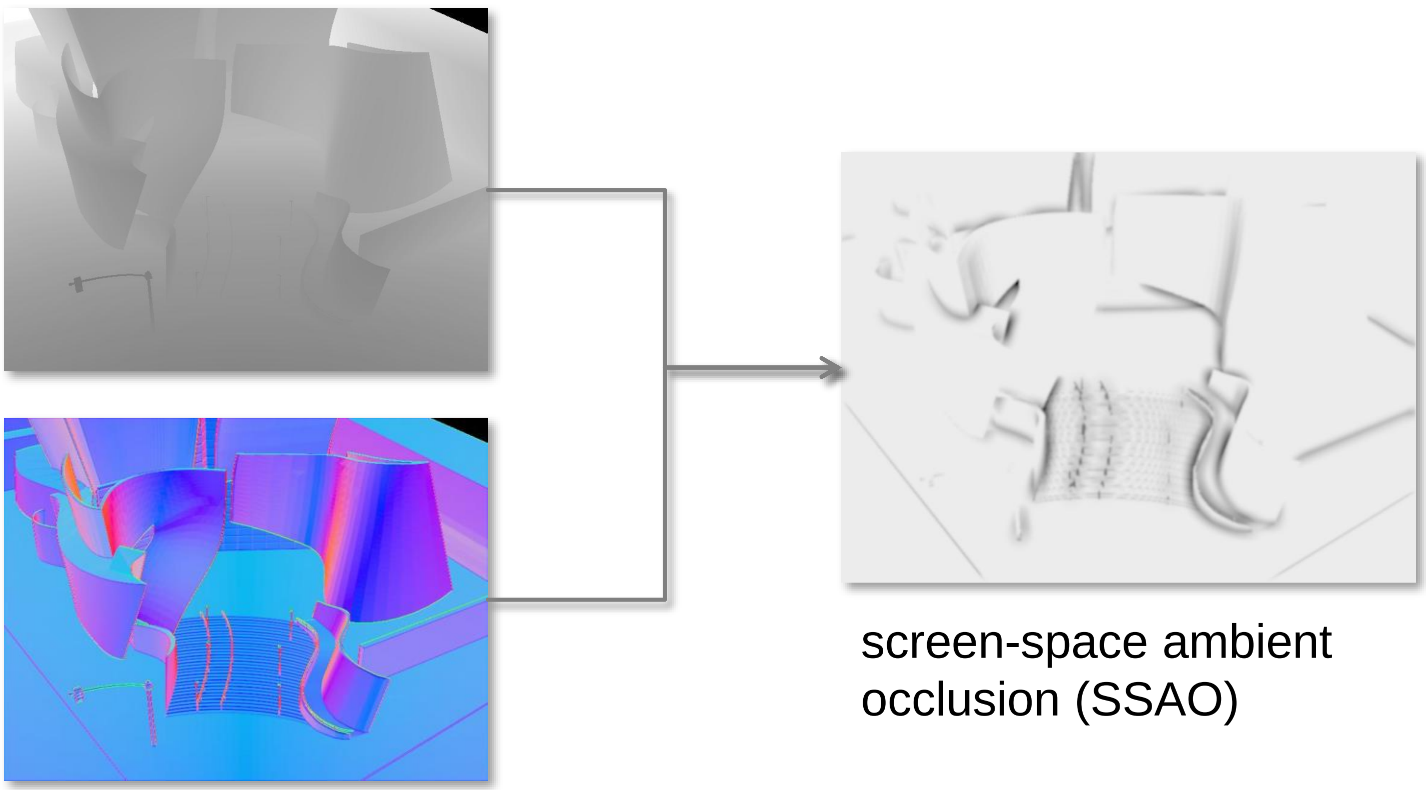
3D Features: Local First-order Statistics

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

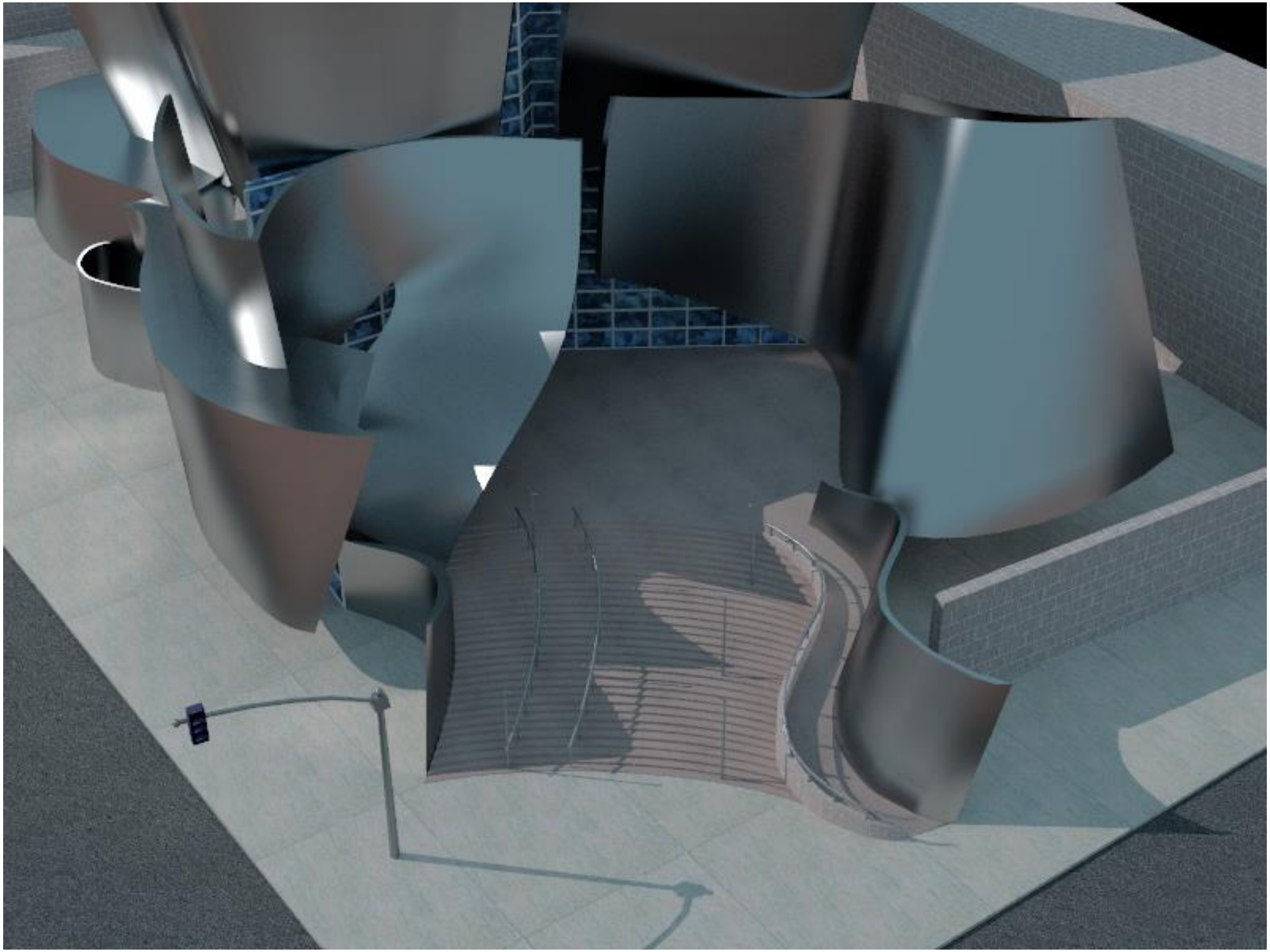


3D Features: Ambient Occlusion

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



screen-space ambient occlusion (SSAO)



REFERENCE IMAGE

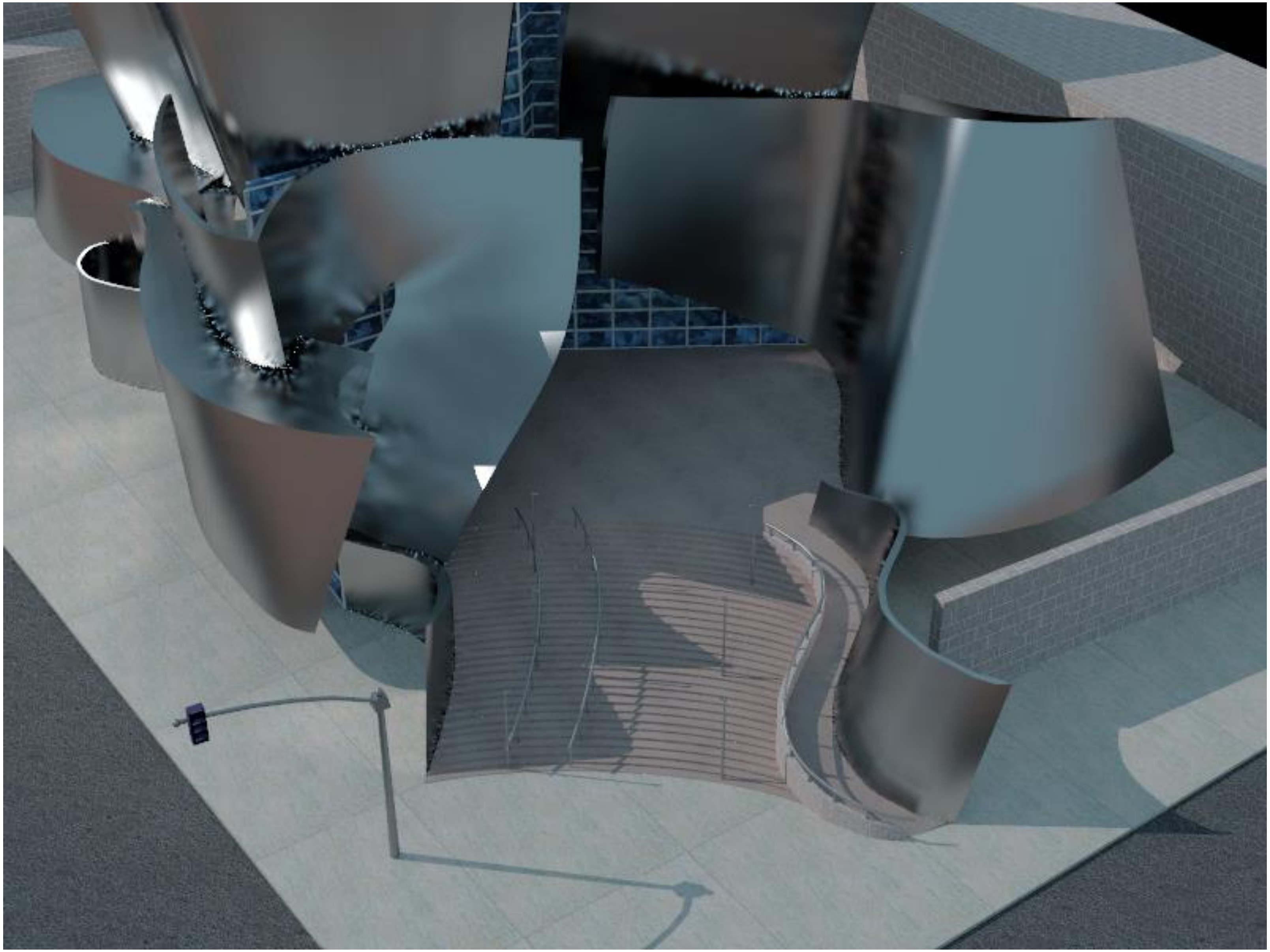
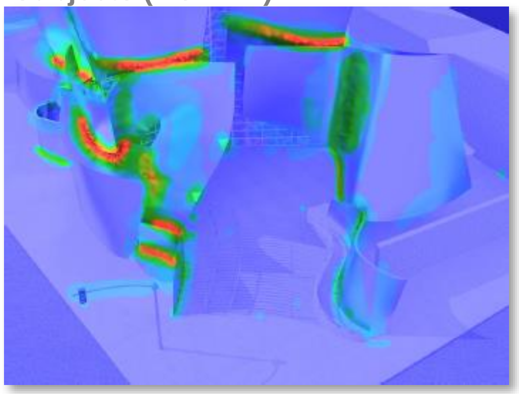


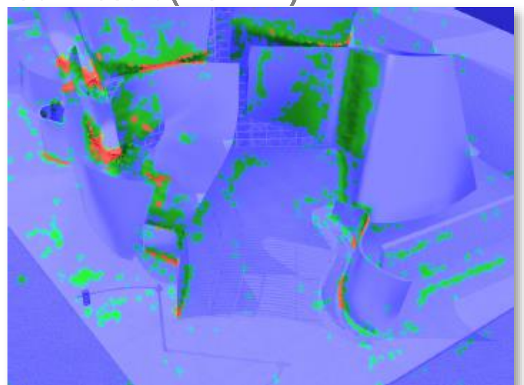
IMAGE WITH ARTIFACTS

Results (VPL noise)

Subjects (NO REF)

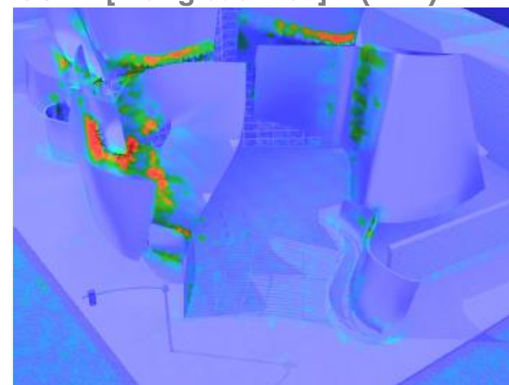


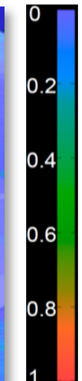
Our Result (NO REF)



corr = 0.662 (0.628)

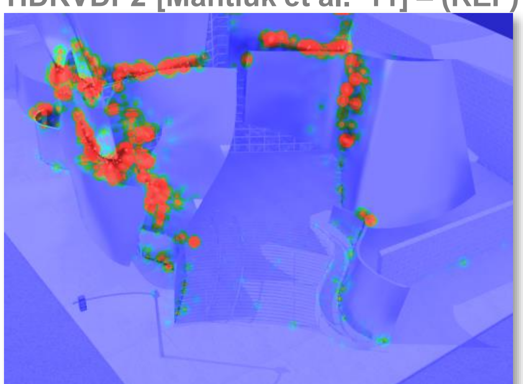
SSIM [Wang et al. '04] – (REF)





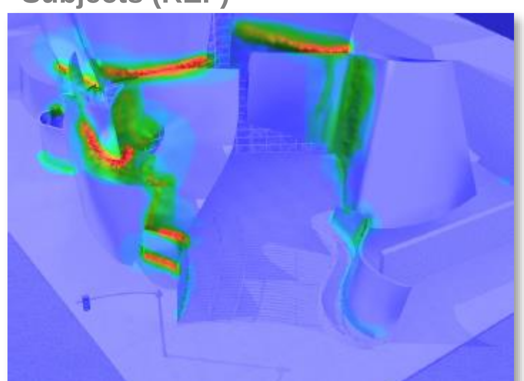
corr = 0.674

HDRVDP2 [Mantiuk et al. '11] – (REF)



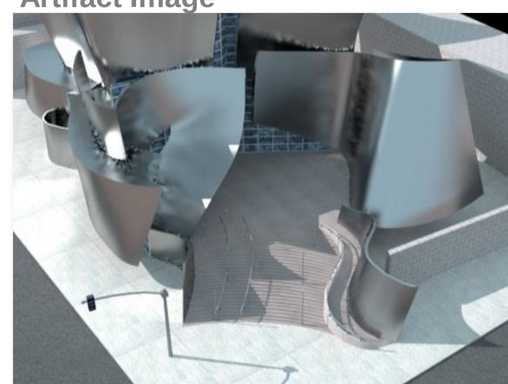
corr = 0.725

Subjects (REF)



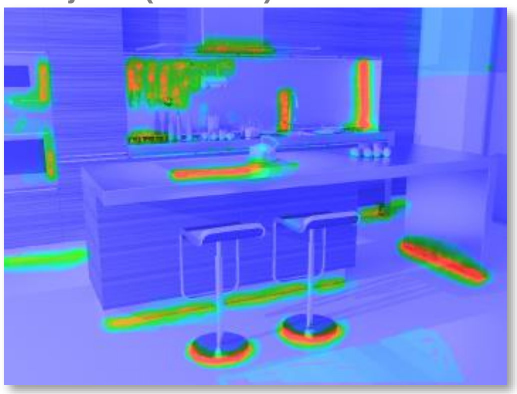
corr = 0.903

Artifact Image

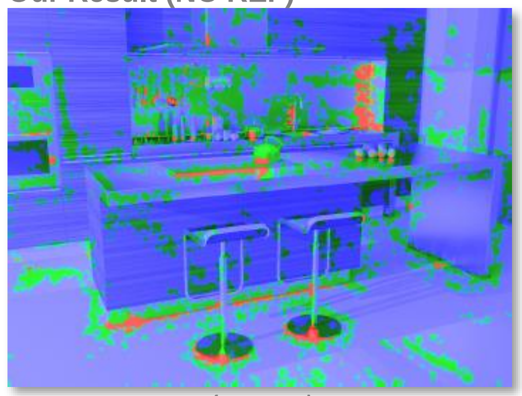


Results (VPL noise)

Subjects (NO REF)

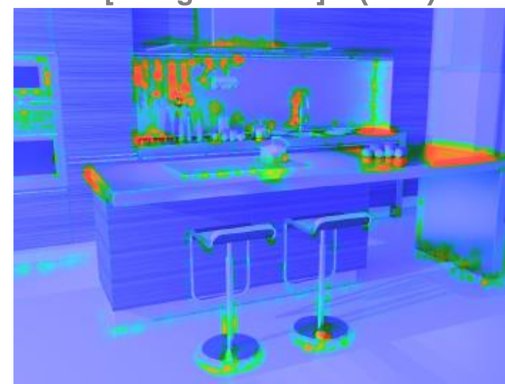


Our Result (NO REF)

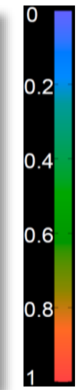


corr = 0.436 (0.298)

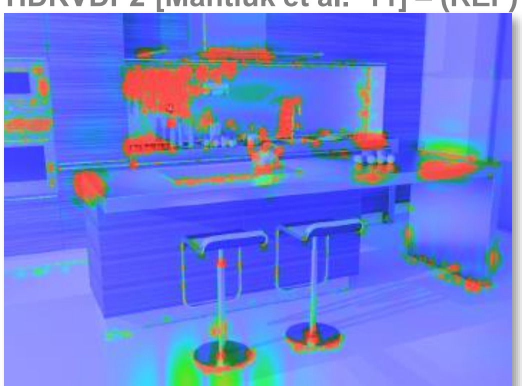
SSIM [Wang et al. ‘04] – (REF)



corr = 0.469

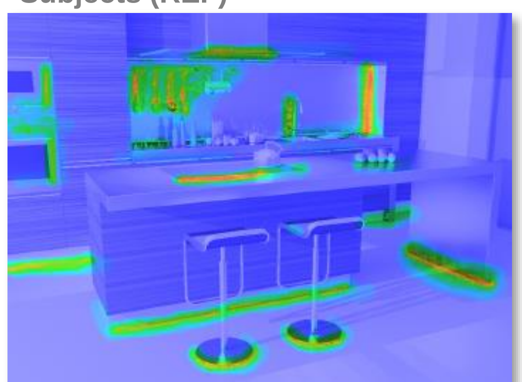


HDRVDP2 [Mantiuk et al. ‘11] – (REF)



corr = 0.495

Subjects (REF)

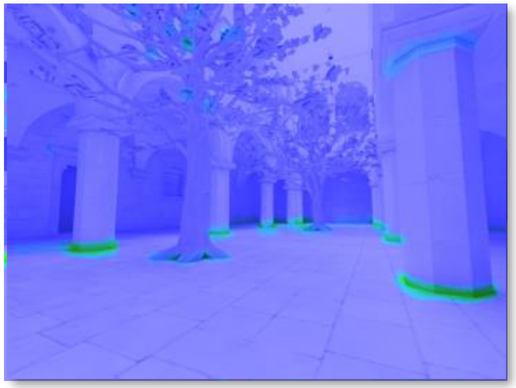


corr = 0.913

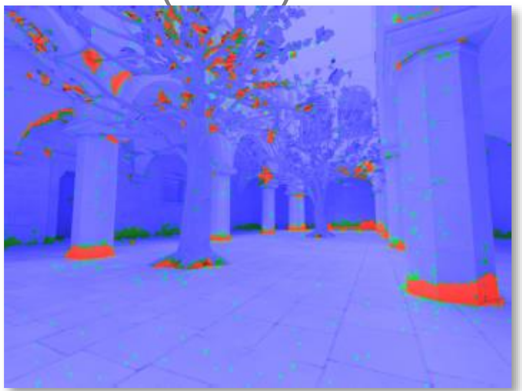
Artifact Image

Results (VPL clamping)

Subjects (NO REF)

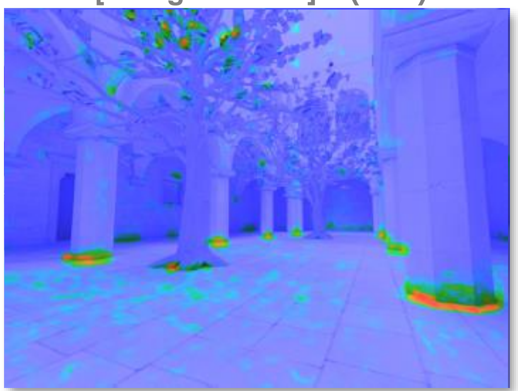


Our Result (NO REF)



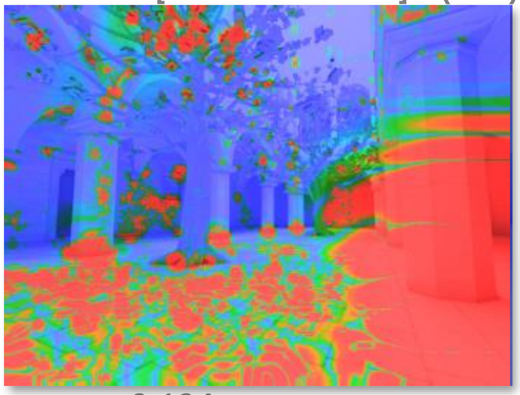
corr = 0.470 (0.450)

SSIM [Wang et al. '04] – (REF)



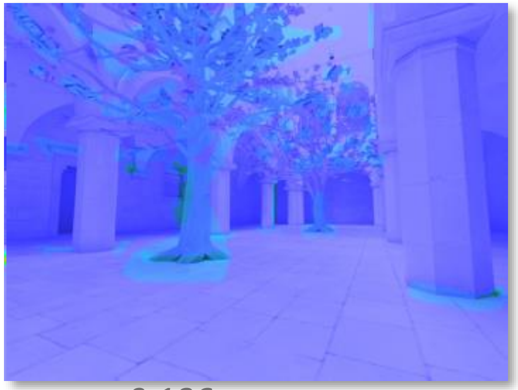
corr = 0.637

HDRVDP2 [Mantiuk et al. '11] – (REF)



corr = 0.134

Subjects (REF)

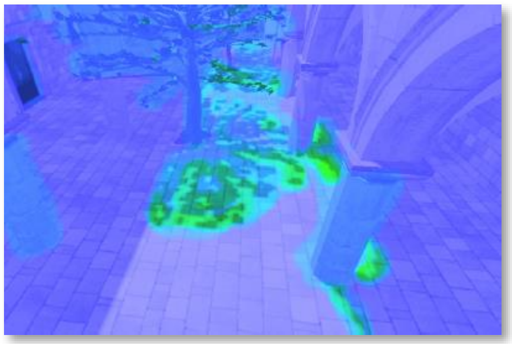


corr = 0.186

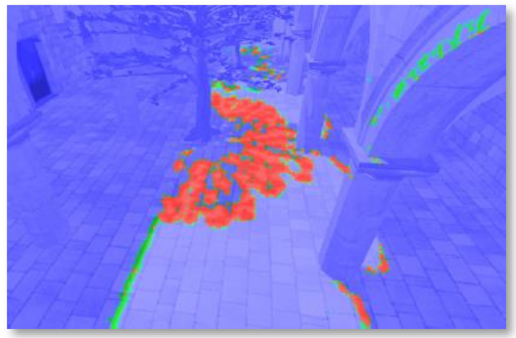
Artifact Image

Results (Shadow aliasing)

Subjects (NO REF)

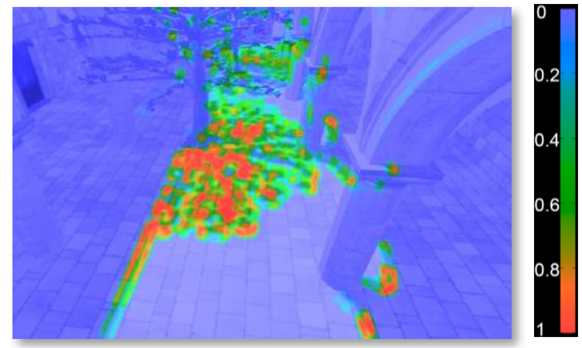


Our Result (NO REF)



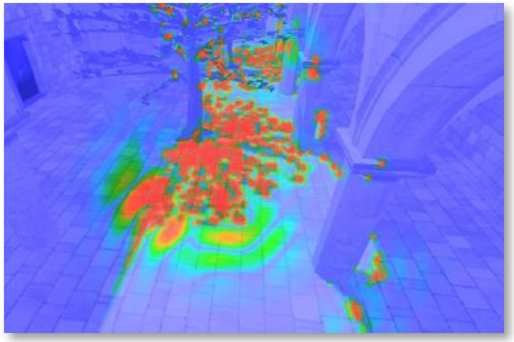
corr = 0.767 (0.638)

SSIM [Wang et al. '04] – (REF)



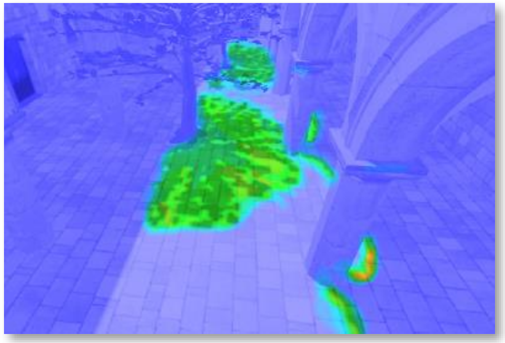
corr = 0.742

HDRVDP2 [Mantiuk et al. '11] – (REF)



corr = 0.669

Subjects (REF)



corr = 0.772

Artifact Image

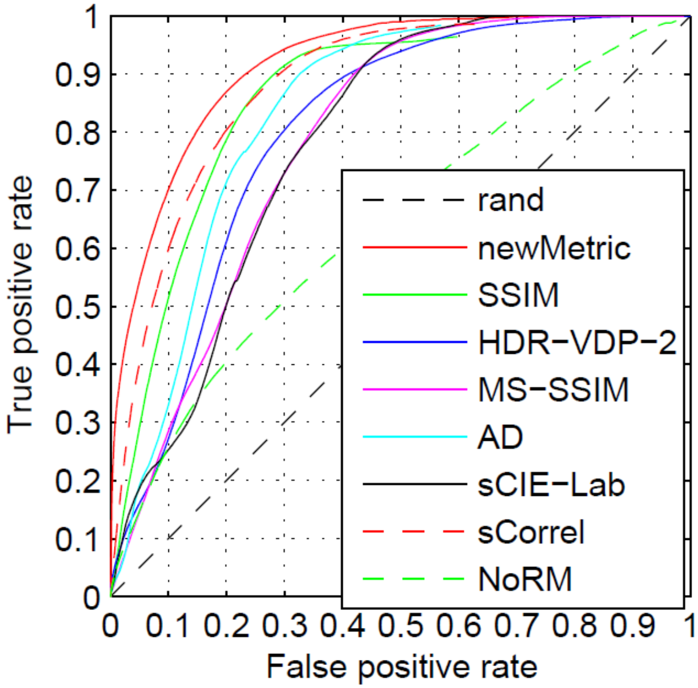
FR Data-driven IQM (LPLD)

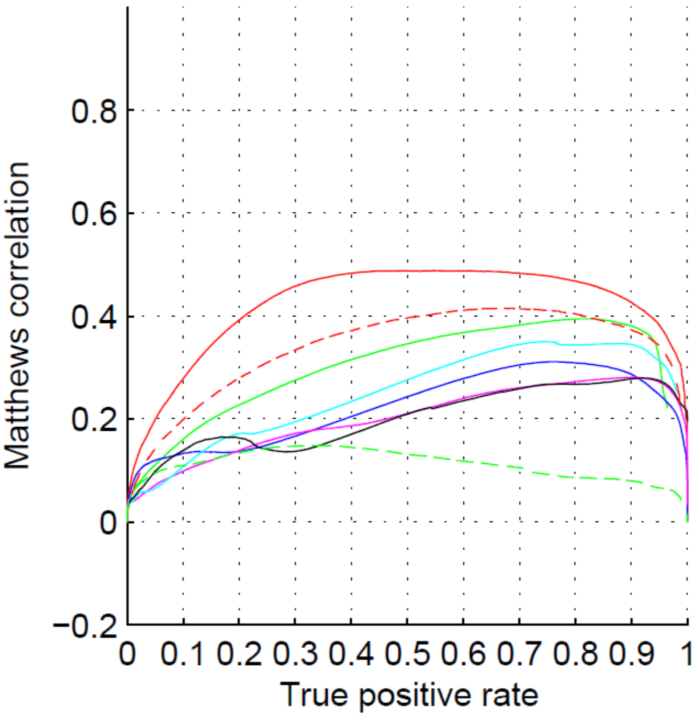
- [Čadík et al. 13]
- SL=ensembles of bagged decision trees
  - t=20 trees, avg. depth=10
- 10 best features ranked by feature selection
- LOCCG dataset for training

- Advantages
  - Computer graphics content
  - Many distortion types
  - Superposition of distortions

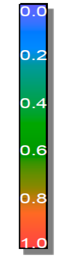
LPLD – Performance

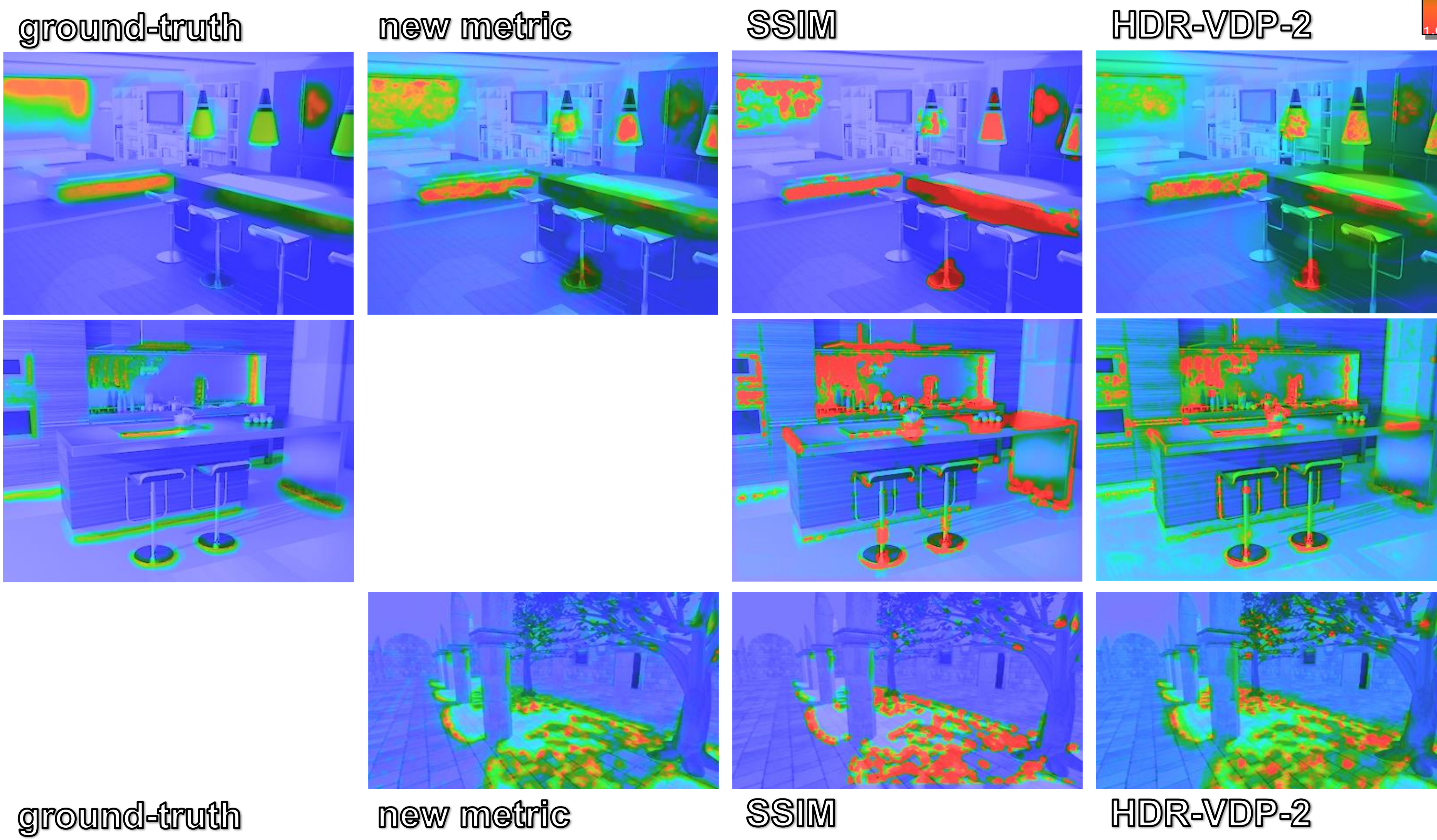
- Metric performance – ROC analysis
  - LOCCG dataset – leave one out cross validation
  - Compared to 7 state-of-the-art IQM





LPLD – Results (LDR)

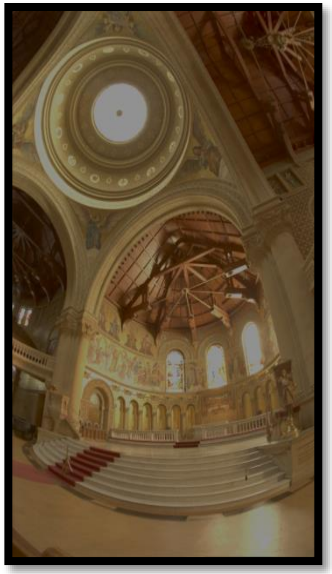




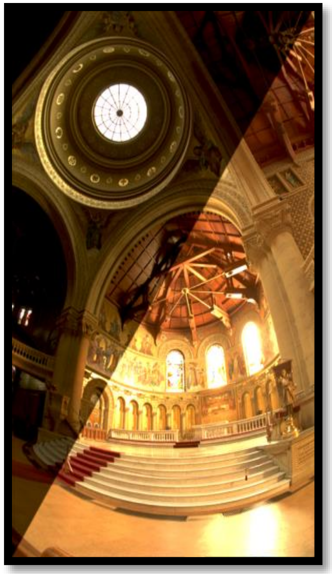
2.2 Overview

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

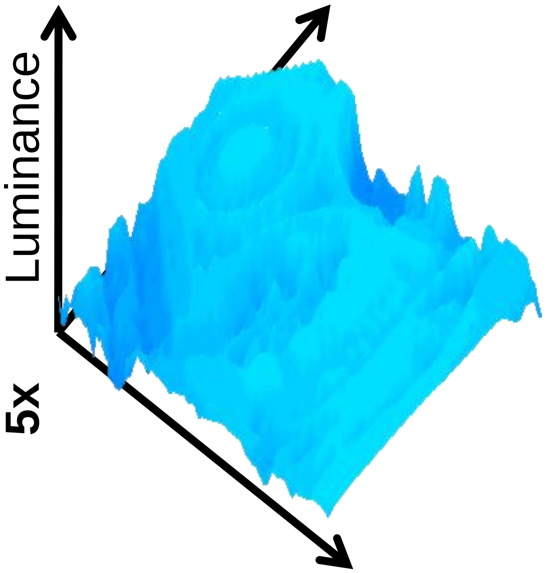
HDR vs. LDR



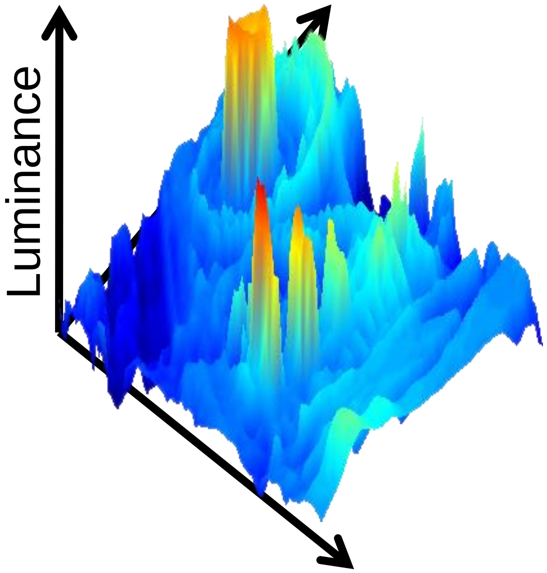
LDR



HDR

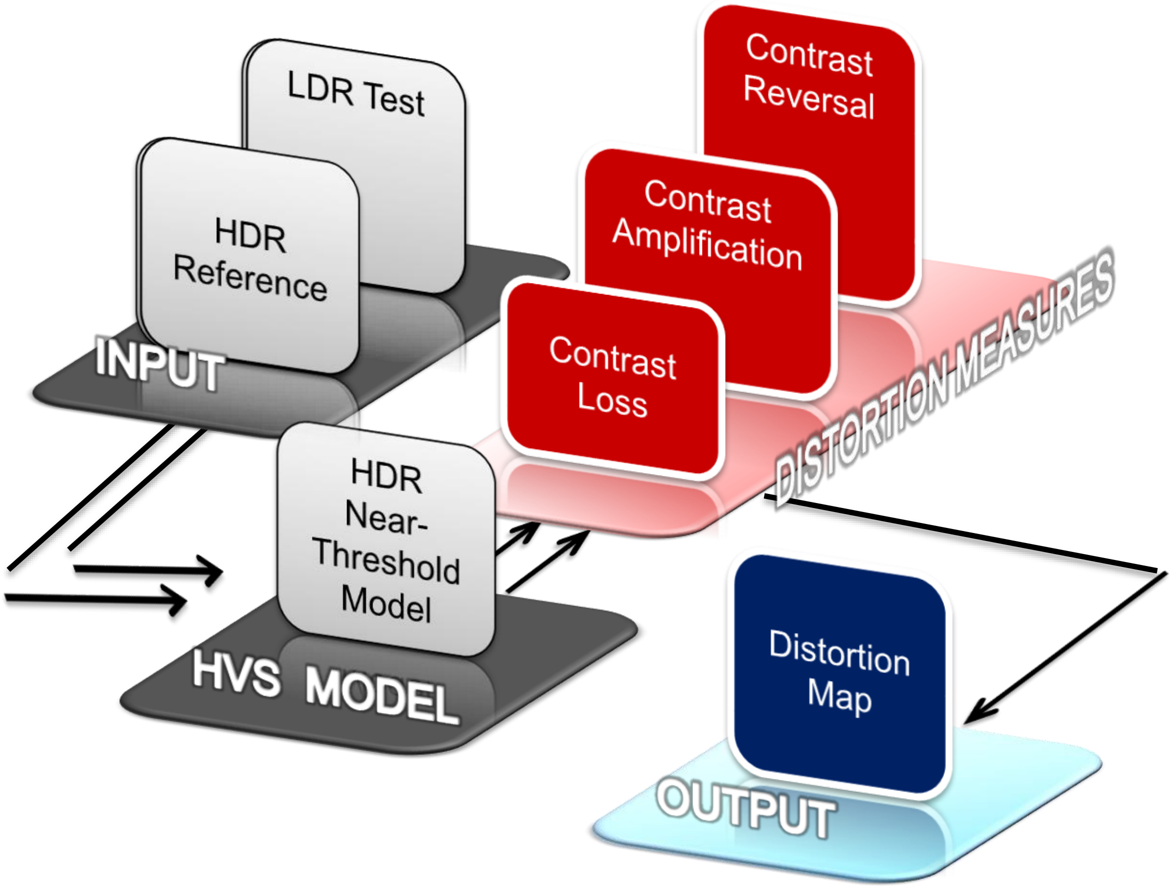


LDR



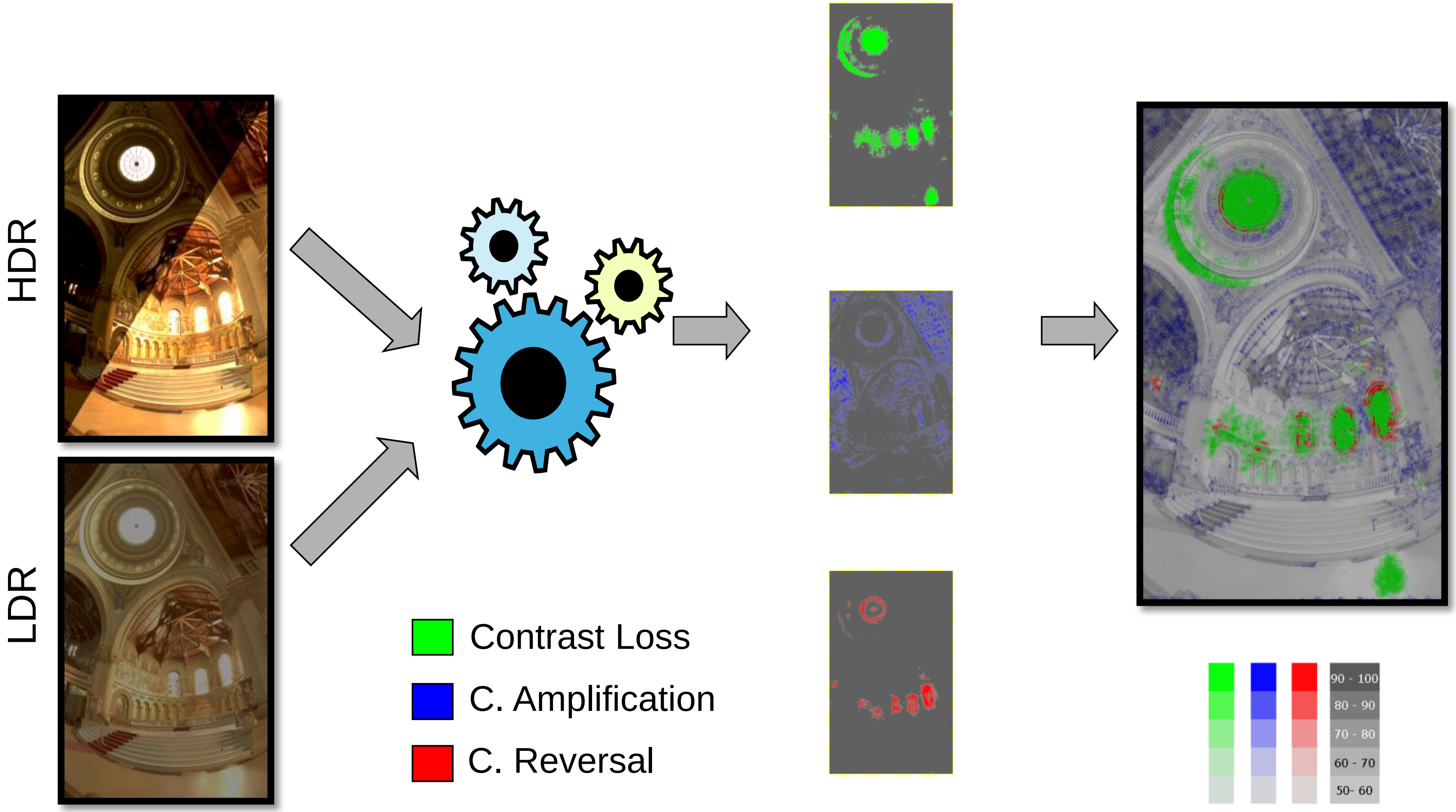
HDR

Dynamic Range Independent IQM (DRIM)

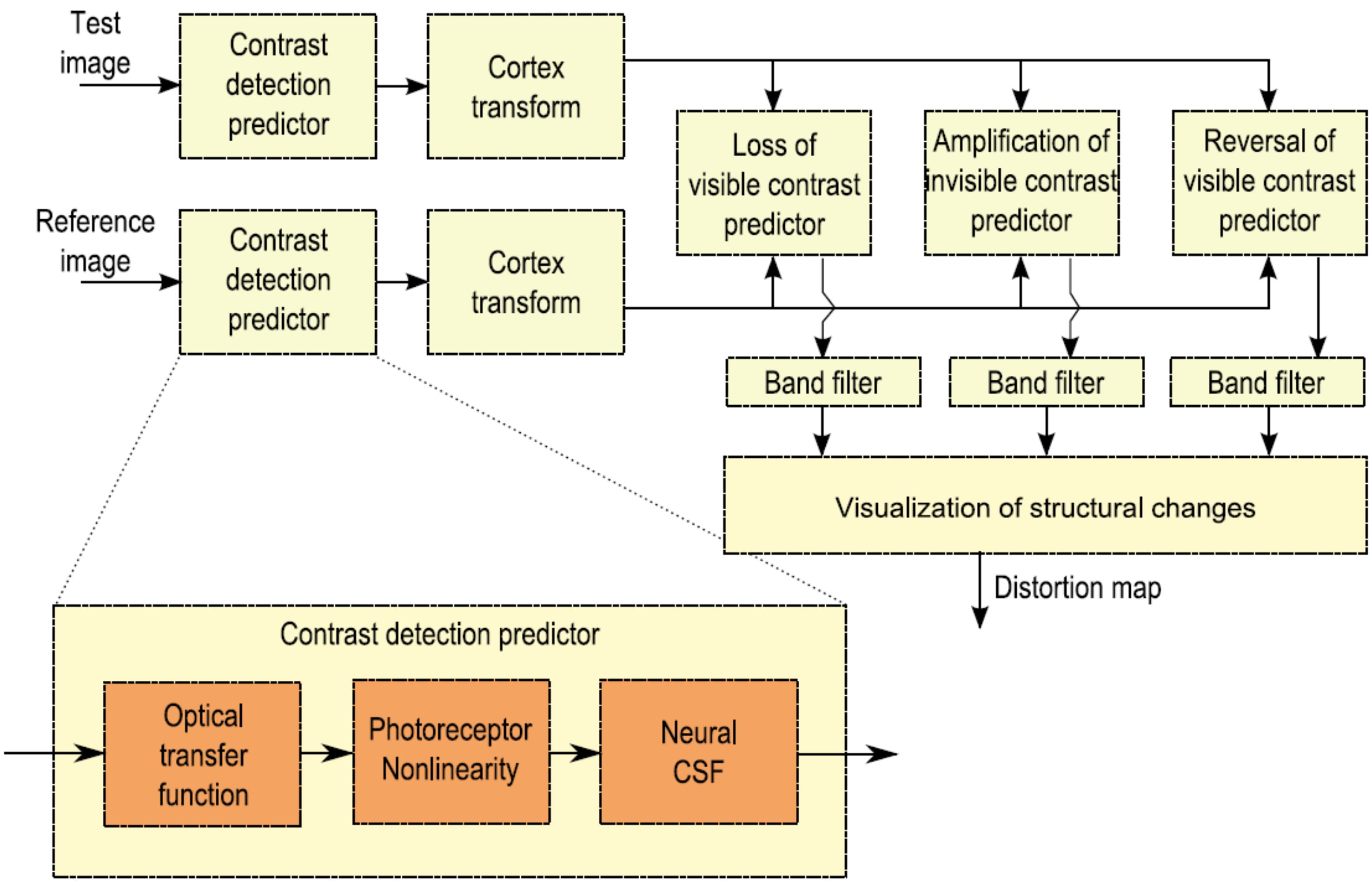


- [Aydın et al. 2008]
- **Key Idea:** Instead of the traditional contrast difference, use distortion measures agnostic to dynamic range difference
- **Result:** An IQM that can meaningfully compare an LDR test image with an HDR reference image, and vice versa
- Enables evaluation of tone mapping operators

Distortion Measures



DRIM



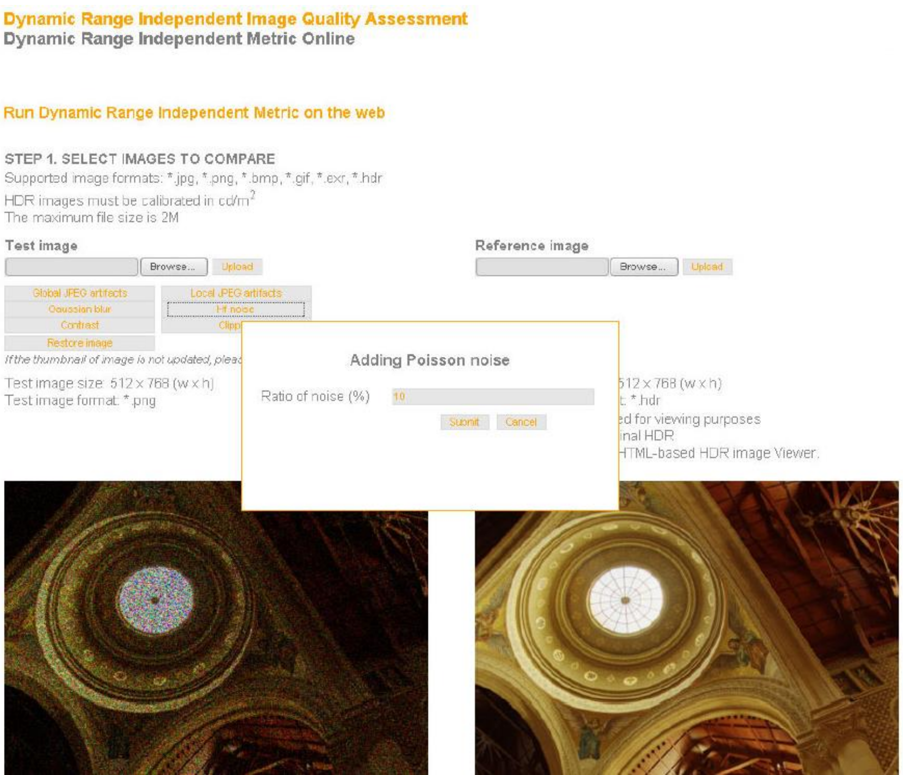
http://metrics.mpi-inf.mpg.de/

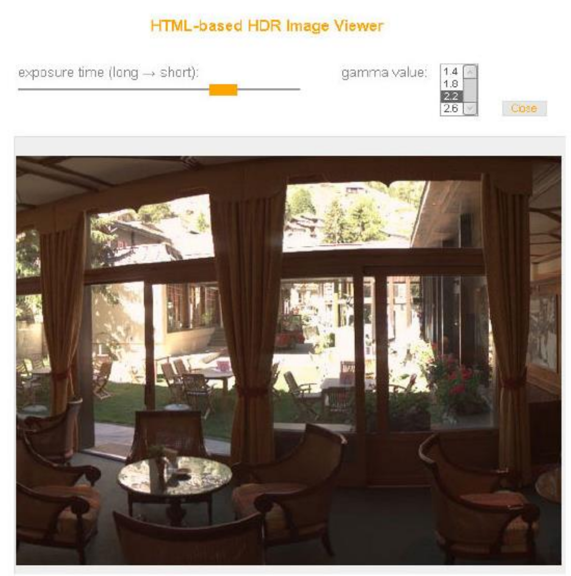
DRIM – Results

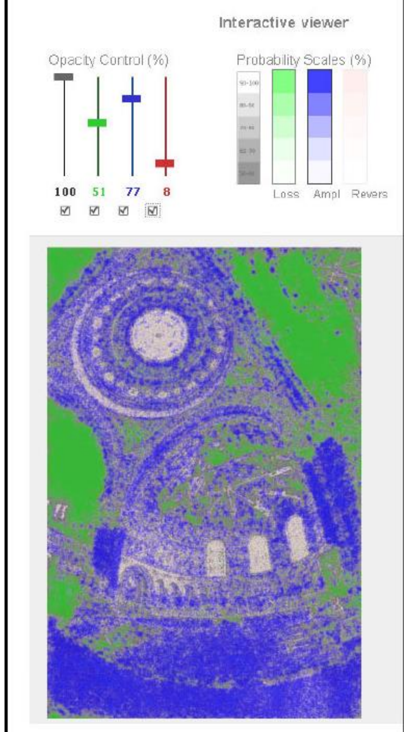
HDR Reference LDR Test DRIQM

DRIM – Limitations

- Grayscale
- not very accurate
- +/- user interface (http://metrics.mpi-inf.mpg.de/)





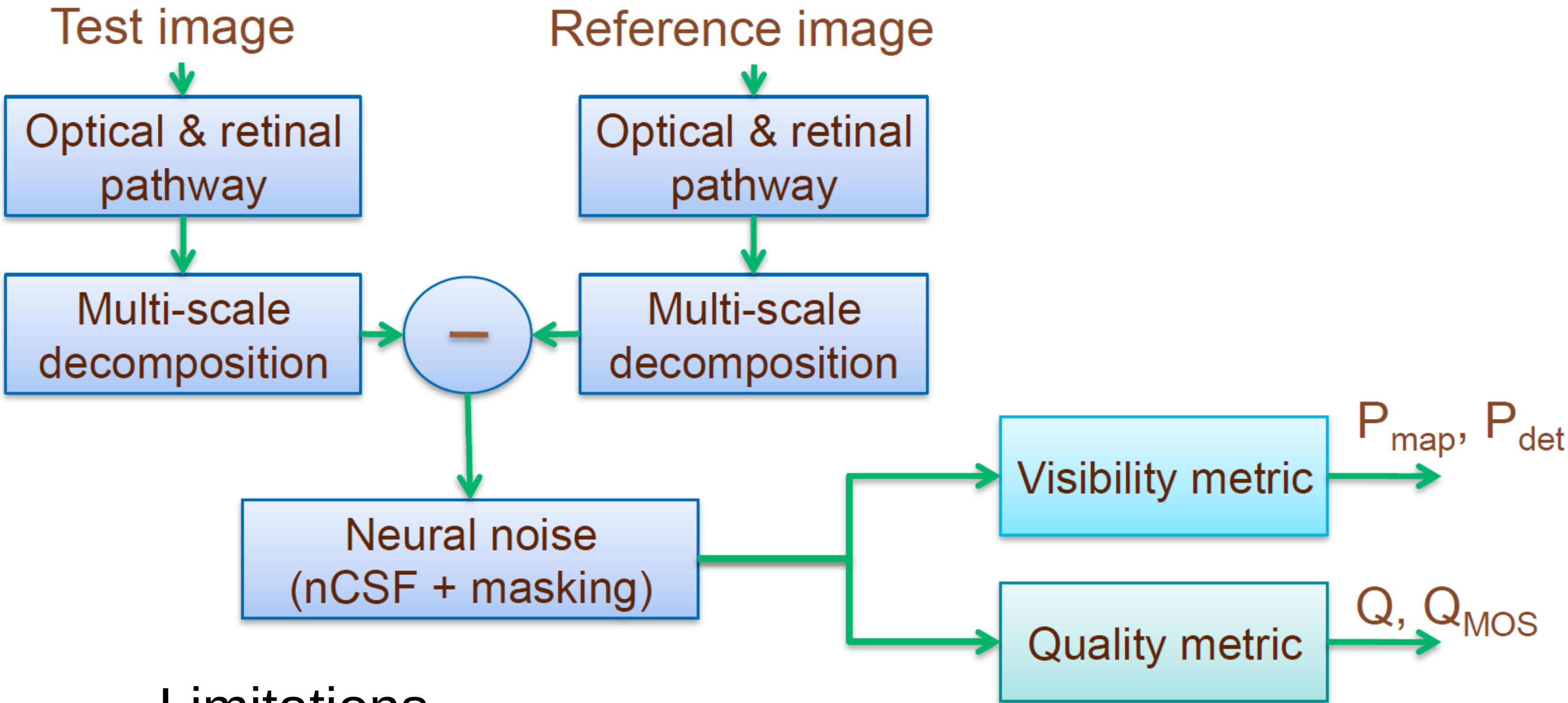


HDRVDP2

- [Mantiuk et al. 11]
  - Matlab code available http://hdrvdp.sourceforge.net/
  - Online version: http://metrics.mpi-inf.mpg.de/
  - opencl GPU implementation coming soon

  - Carefully calibrated with experimental data
    - New CSF measurements
    - LIVE, TID2008
  - Chromatic CSF
  - Steerable pyramid

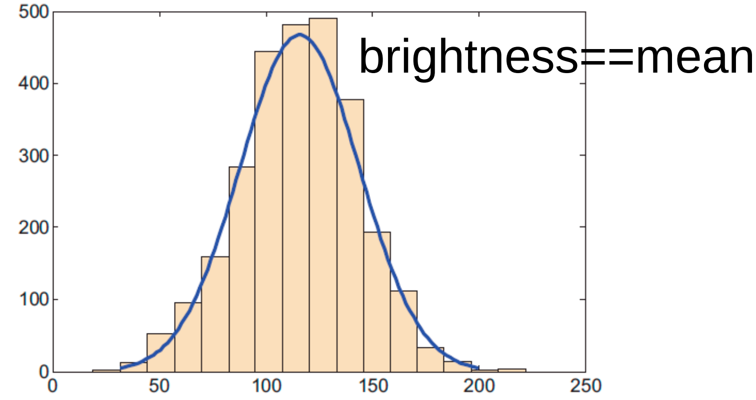
HDRVDP2

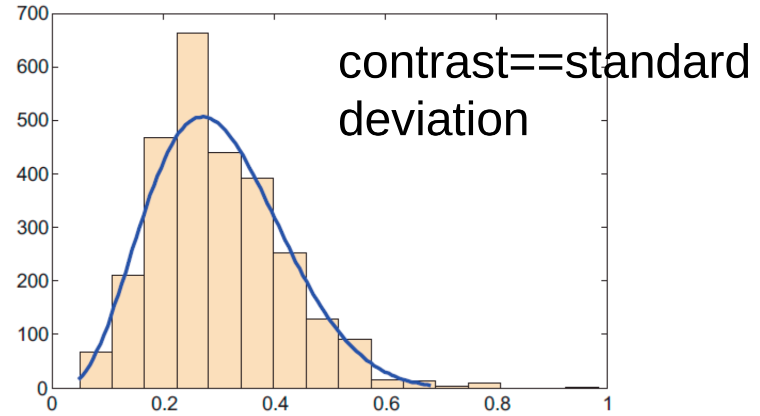


- Limitations
  - grayscale
  - slower than SSIM

Tone Mapped image Quality Index (TMQI)

- [Yeganeth and Wang 13]
  - Matlab code available https://ece.uwaterloo.ca/~z70wang/research/tmqi/
  - FR IQM for tone mapped images
  - very simple





1. multiscale **SSIM**
2. measure of **naturalness**

based on statistics of natural images [Čadík, Slavík 05]:
brightness, contrast most important → statistical model of naturalness

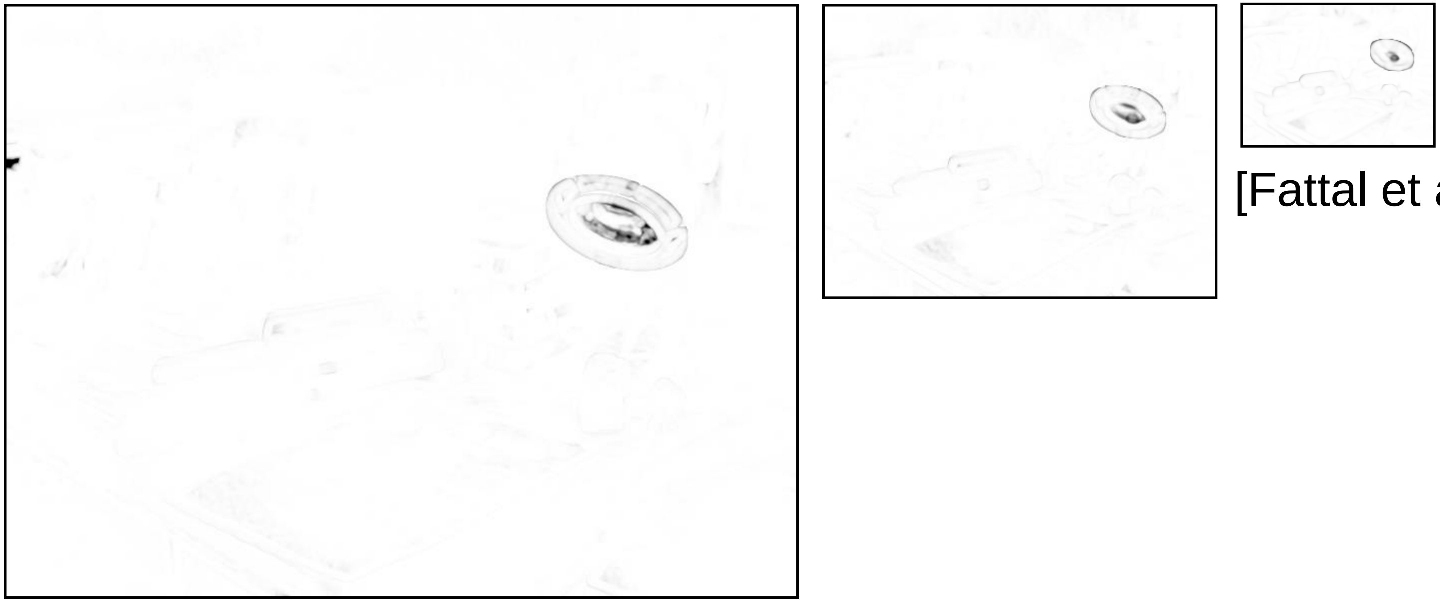
TMQI - results

- overall quality, distortion maps for each scale
- limitations
  - Grayscale, +/- dubiously simple ☺

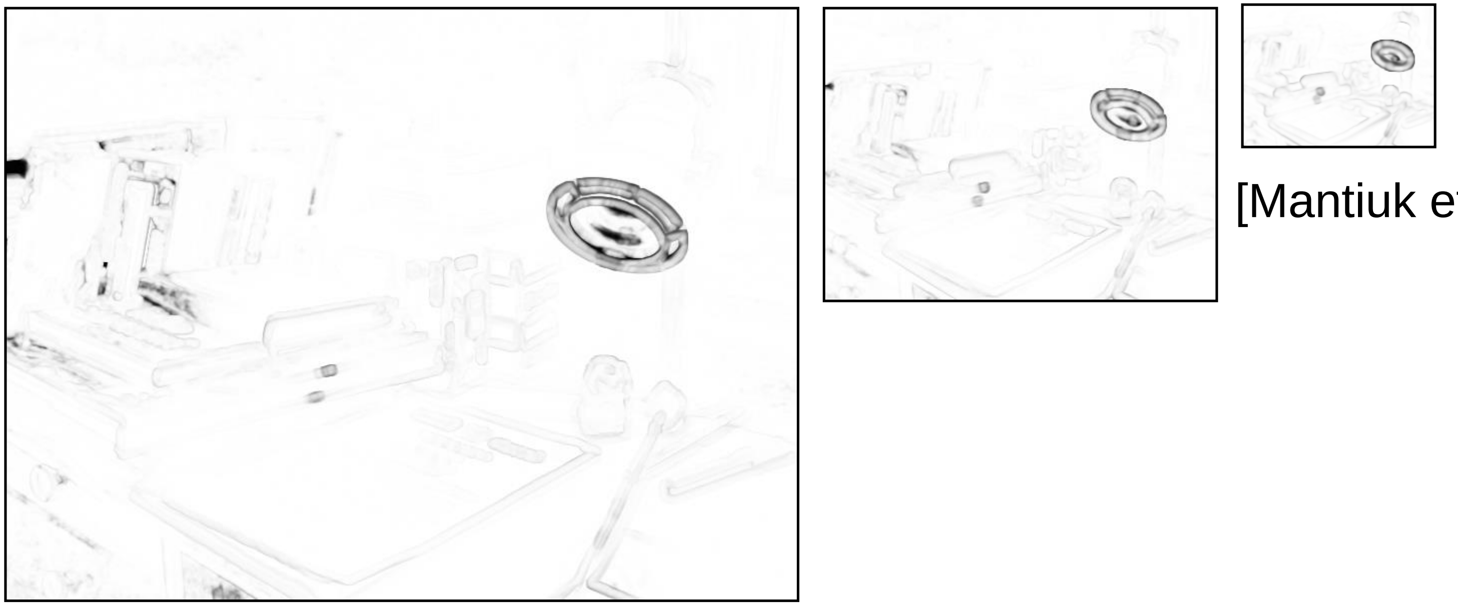
[Fattal et al. 02]

[Mantiuk et al. 08]

TMQI - results



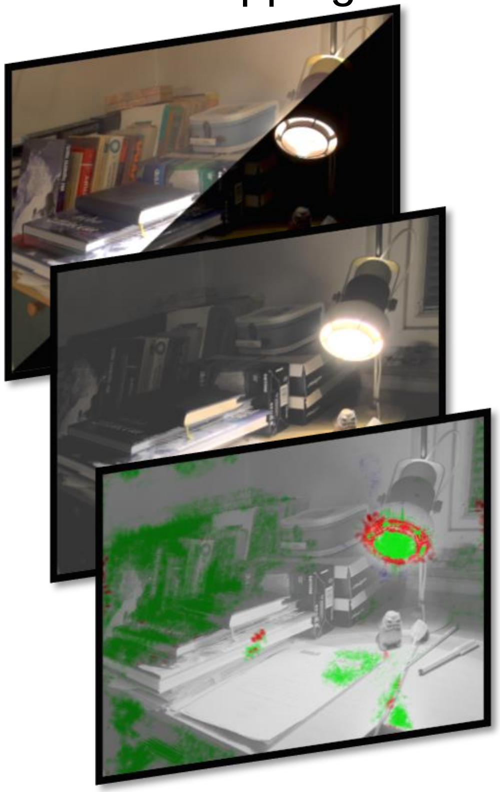
[Fattal et al. 02]



[Mantiuk et al. 08]

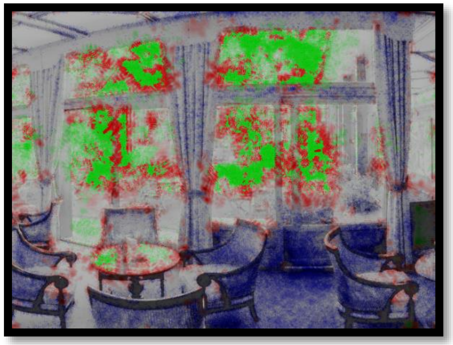
Applications in HDR

Tone Mapping



Inverse
Tone Mapping

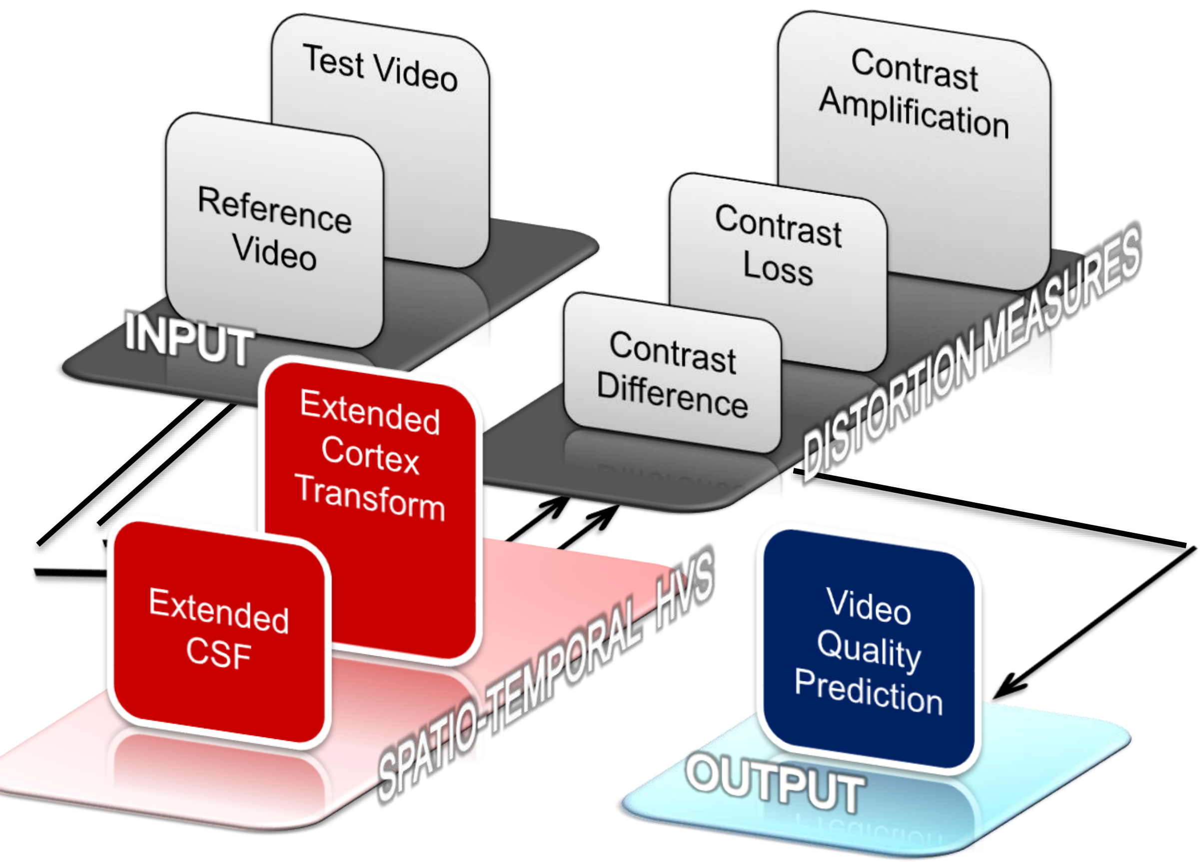
Display
Analysis



2.3 Overview

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

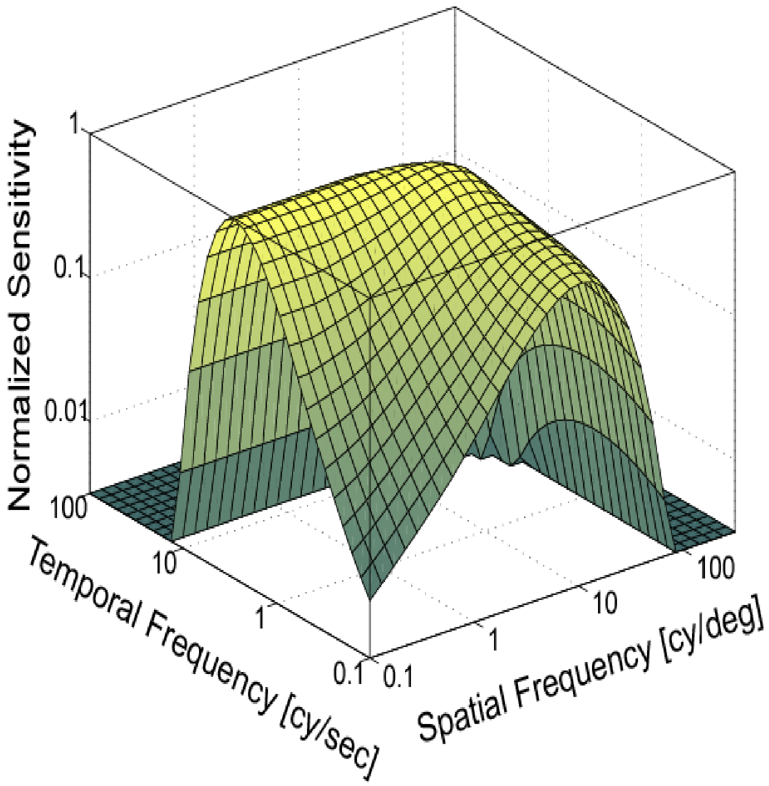
Dynamic Range Independent VQA



- [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

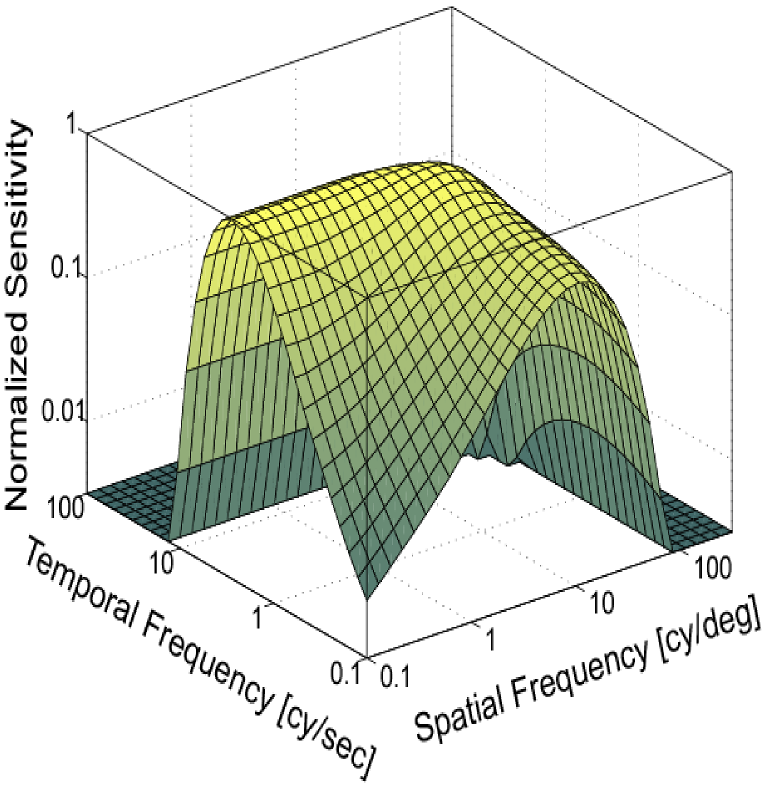
- *CSF:* $\omega,\rho,L_a \rightarrow S$
  - *ω:* temporal frequency,
  - *ρ:* spatial frequency,
  - L_a: adaptation level,
  - *S:* sensitivity.



Spatio-temporal CSF

Contrast Sensitivity Function

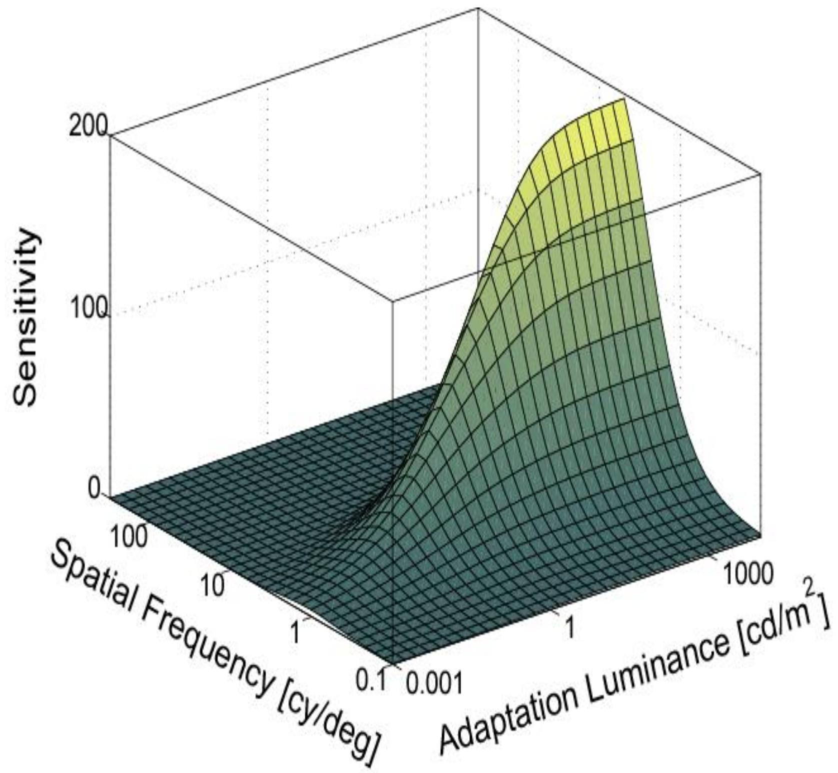
- *CSF:* $\omega, \rho, L_a \rightarrow S$
  - ω: temporal frequency,
  - ρ: spatial frequency,
  - L_a: adaptation level,
  - S: sensitivity.



Spatio-temporal CSF[T]

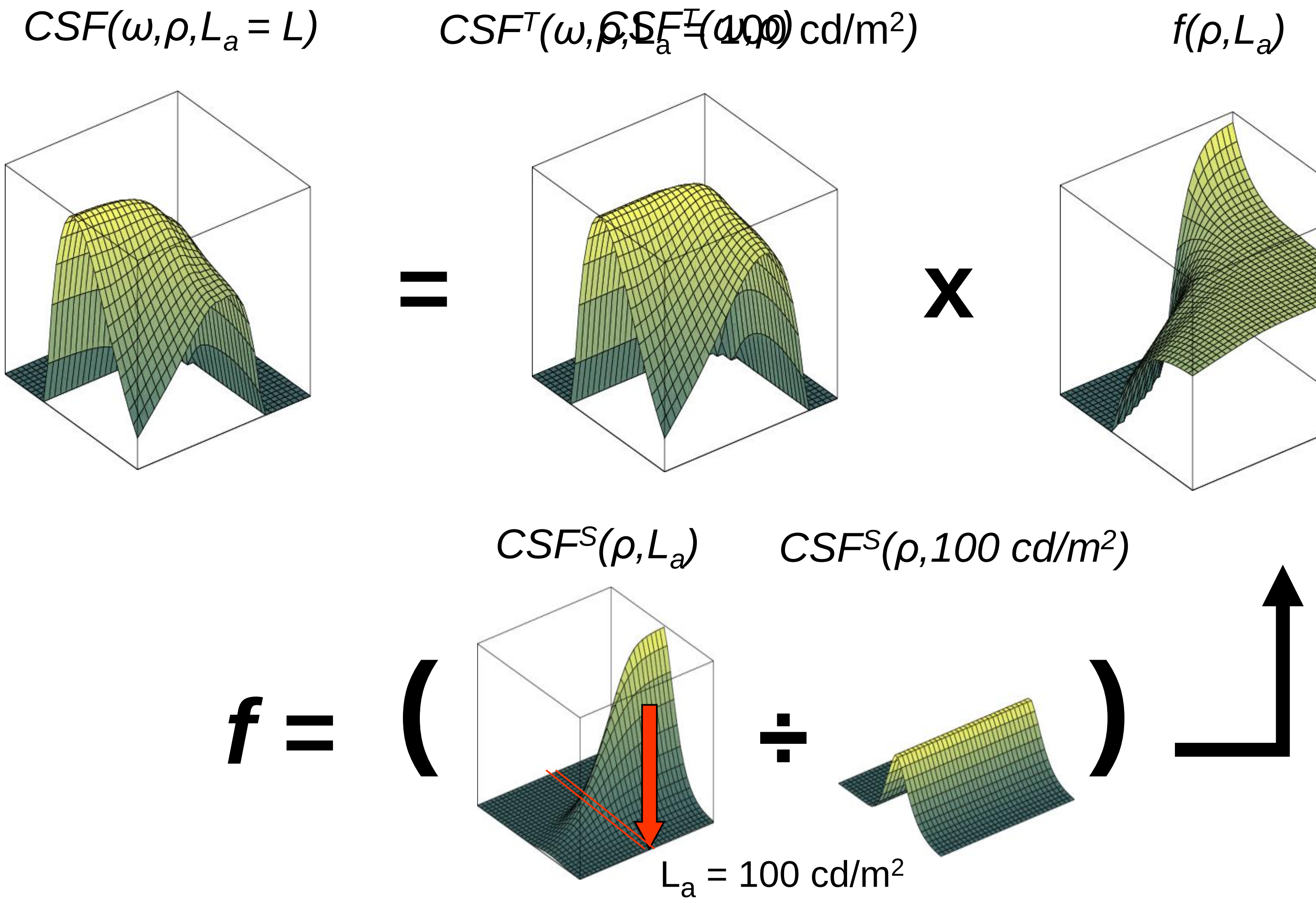
Contrast Sensitivity Function

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

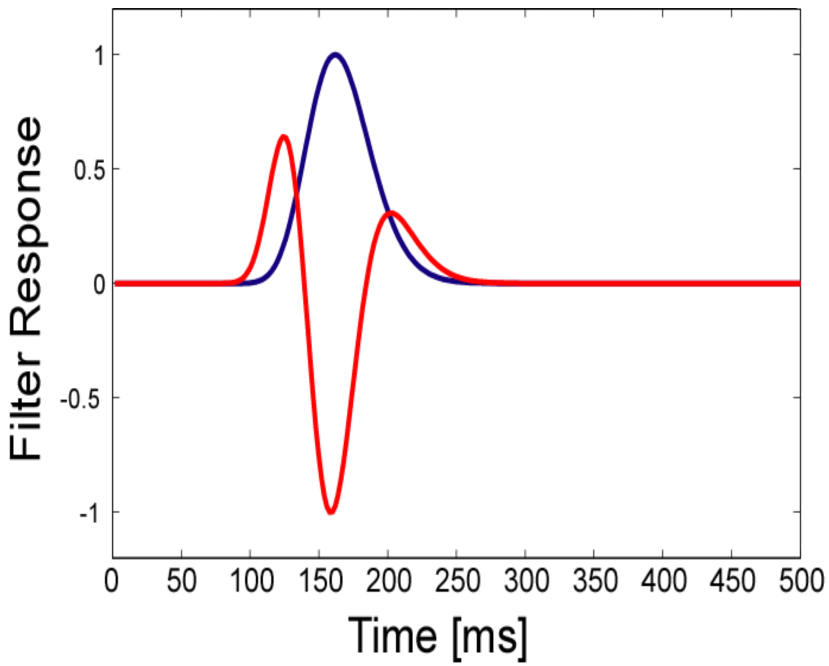


Steady-state CSF^S

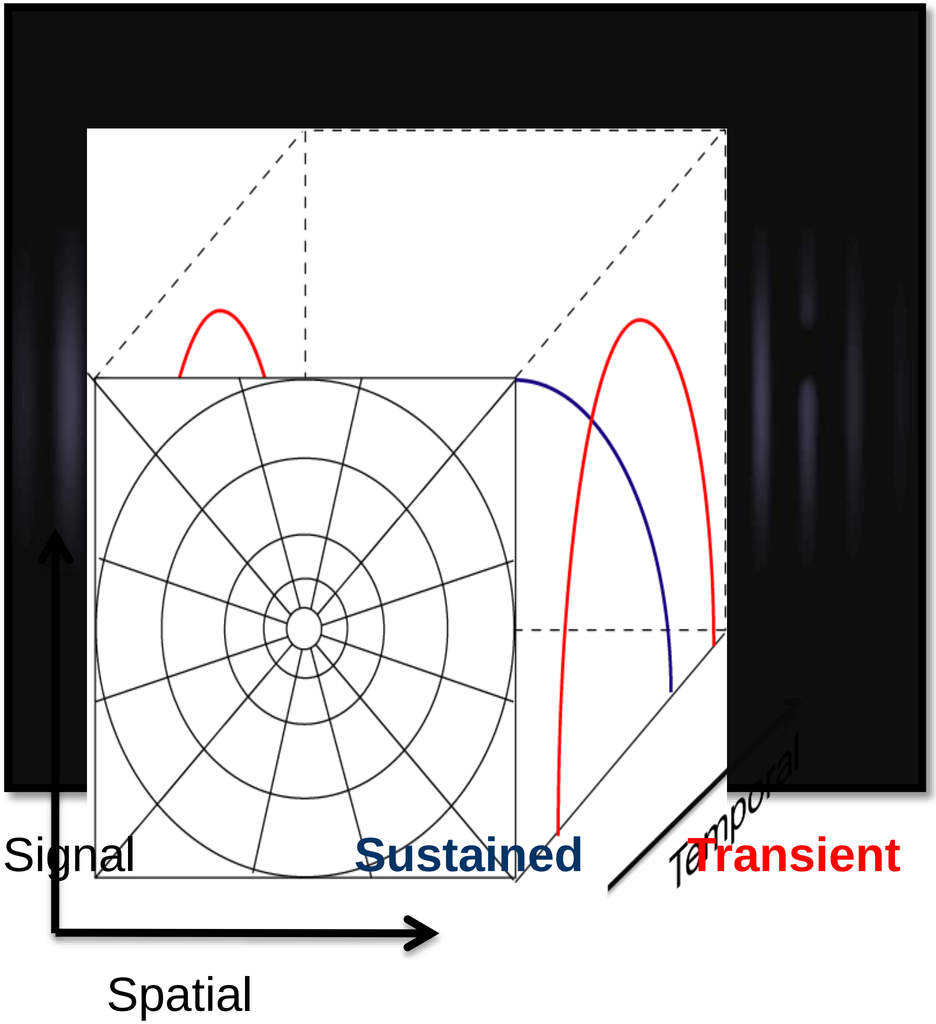
Contrast Sensitivity Function



Extended Cortex Transform



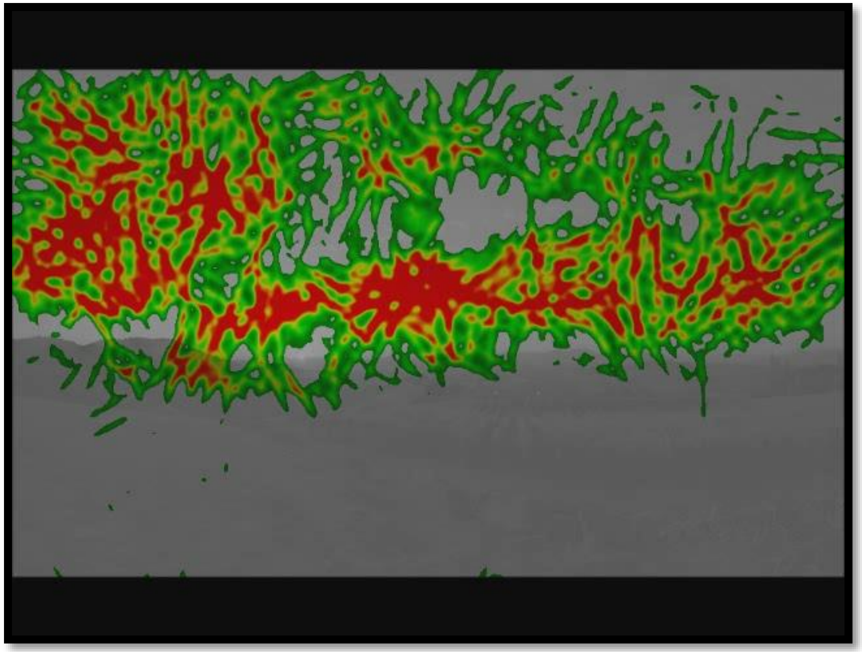
Sustained and **Transient**
Temporal Channels [Winkler 2005]



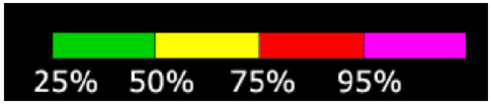
Comparison: Test Scene

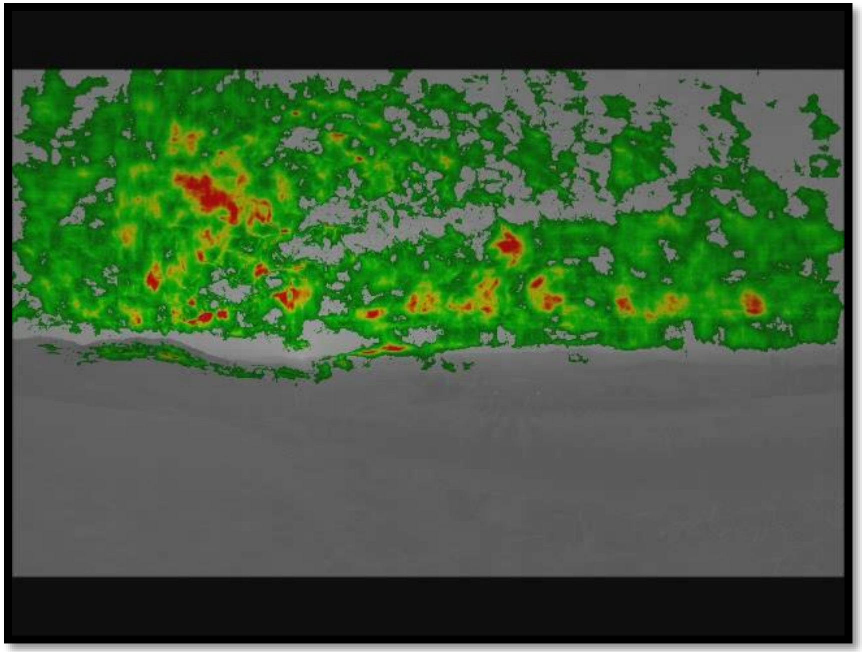
- HDR Scene tone mapped with [Pattanaik 2000]
- Spatio-temporal distortion
  - Random pixel noise filtered with a Gaussian.

Metric Comparison LDR-LDR



Our Metric

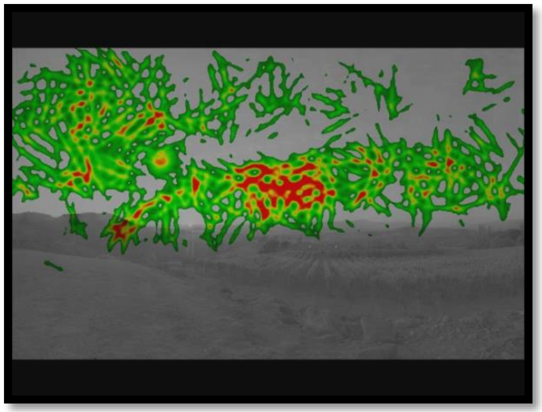




PDM [Winkler 2005]

Metric Comparison HDR-HDR

Our Metric



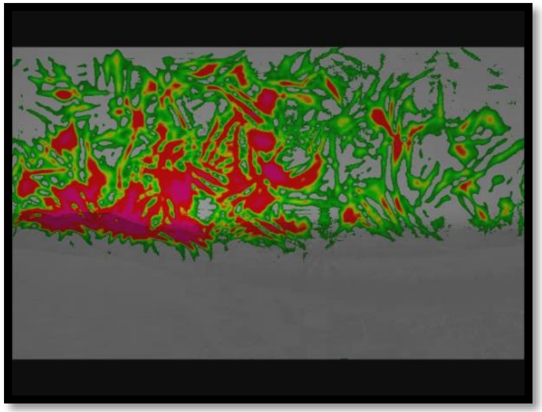
PDM
[Winkler 2005]

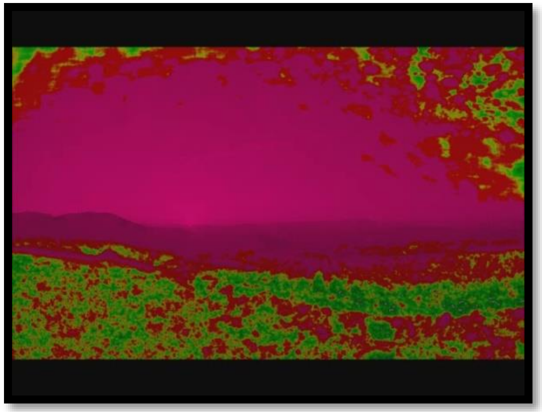
HDRVDP
[Mantiuk
et al 2005]

DRIM
[Aydin et al. 2008]

Metric Comparison HDR-LDR

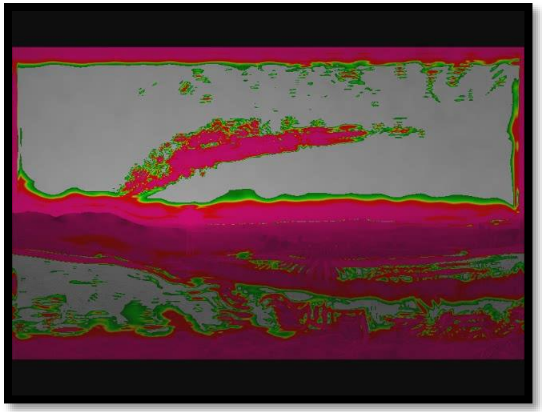
Our Metric





PDM
[Winkler 2005]

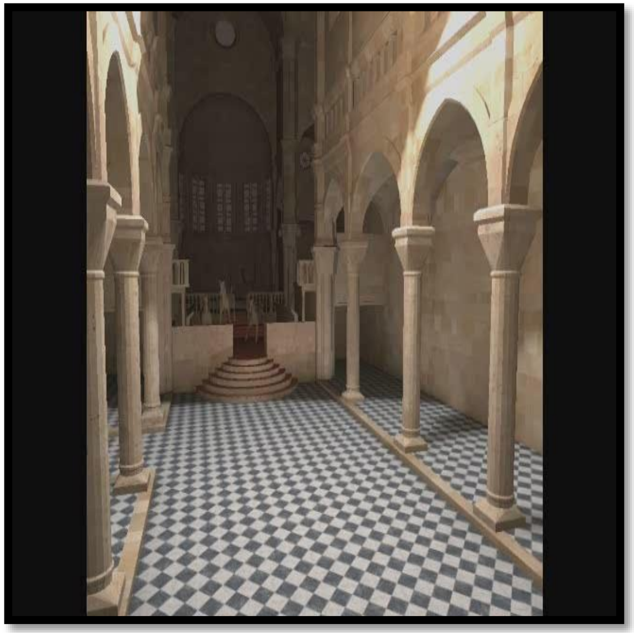
HDRVDP
[Mantiuk
et al 2005]



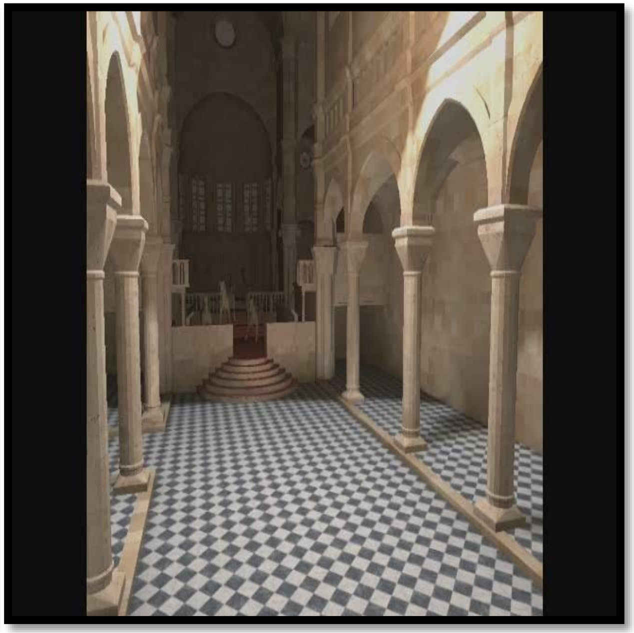
DRIM
[Aydin et al. 2008]

Evaluation of Rendering Methods

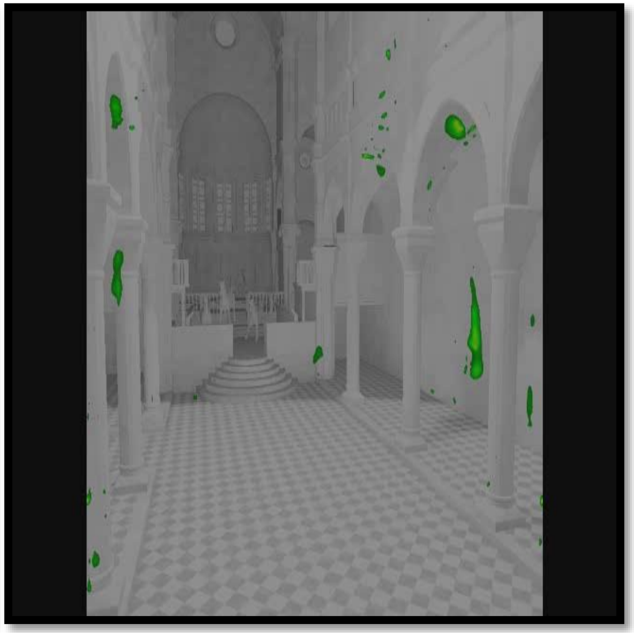
- http://drim.mpi-inf.mpg.de/



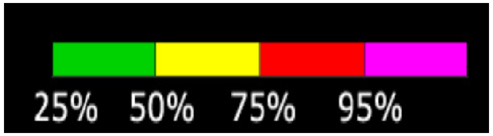
With temporal filtering [Herzog et al. 2010]



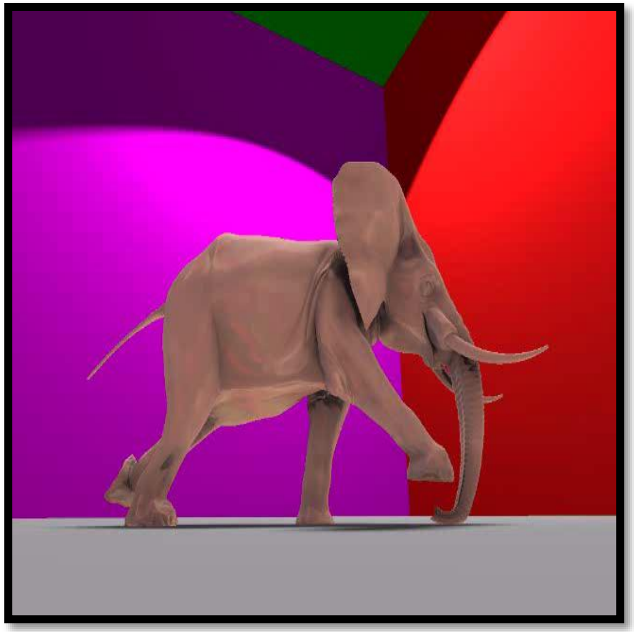
No temporal filtering



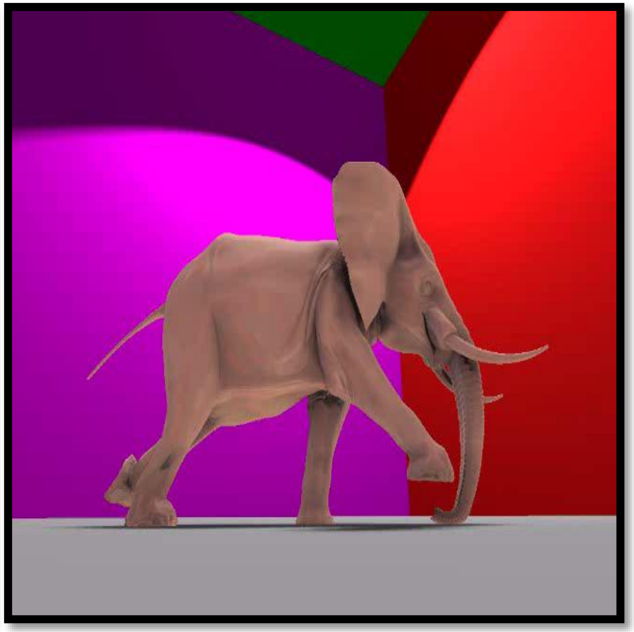
Predicted distortion map



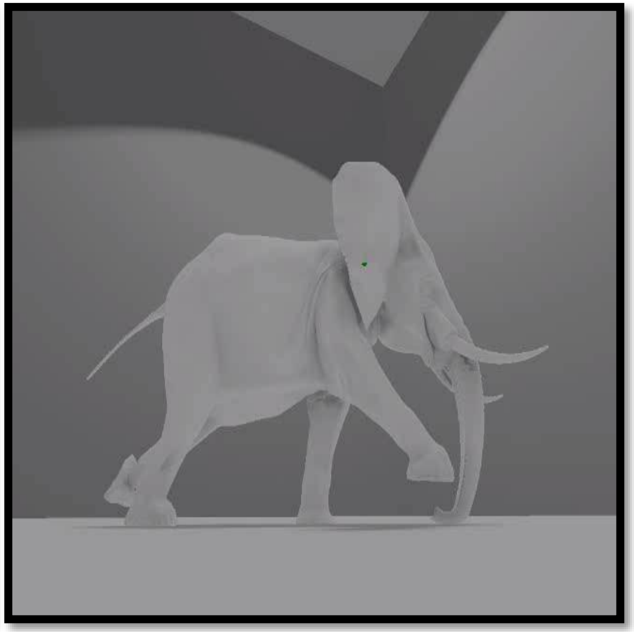
Evaluation of Rendering Qualities



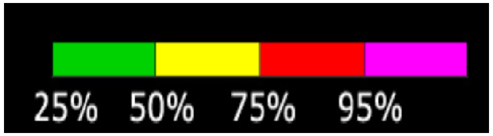
High quality



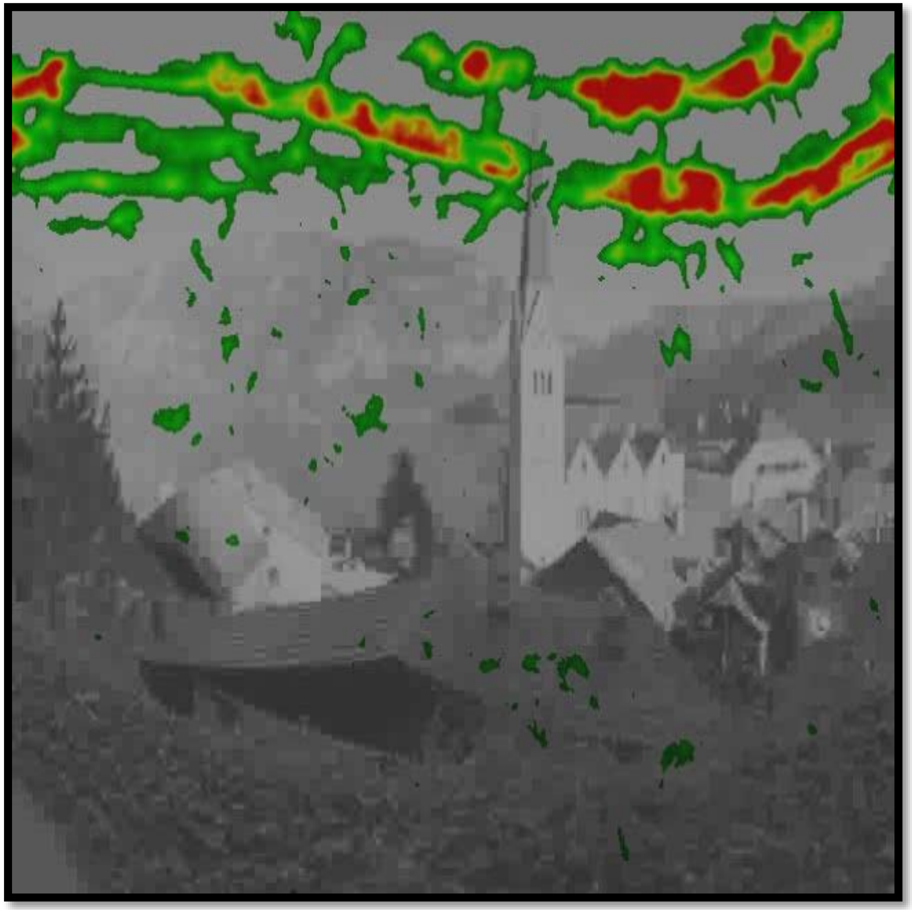
Low quality



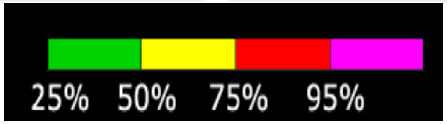
Predicted distortion map



Evaluation of HDR Compression



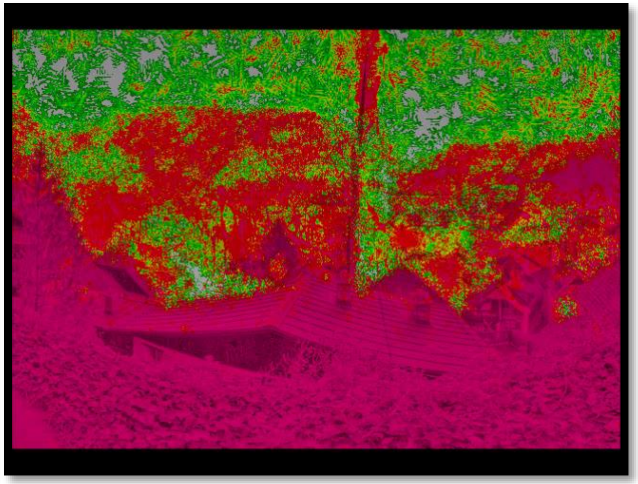
Medium Compression



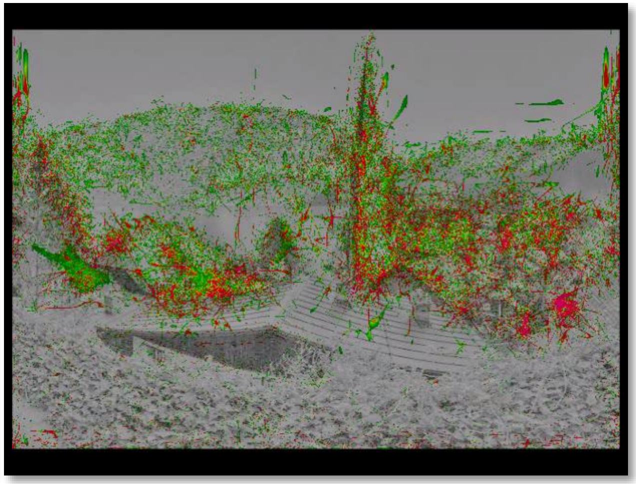
High Compression

Evaluation of Video Tone Mapping

[Fattal et al. 2002]



Detail Amplification

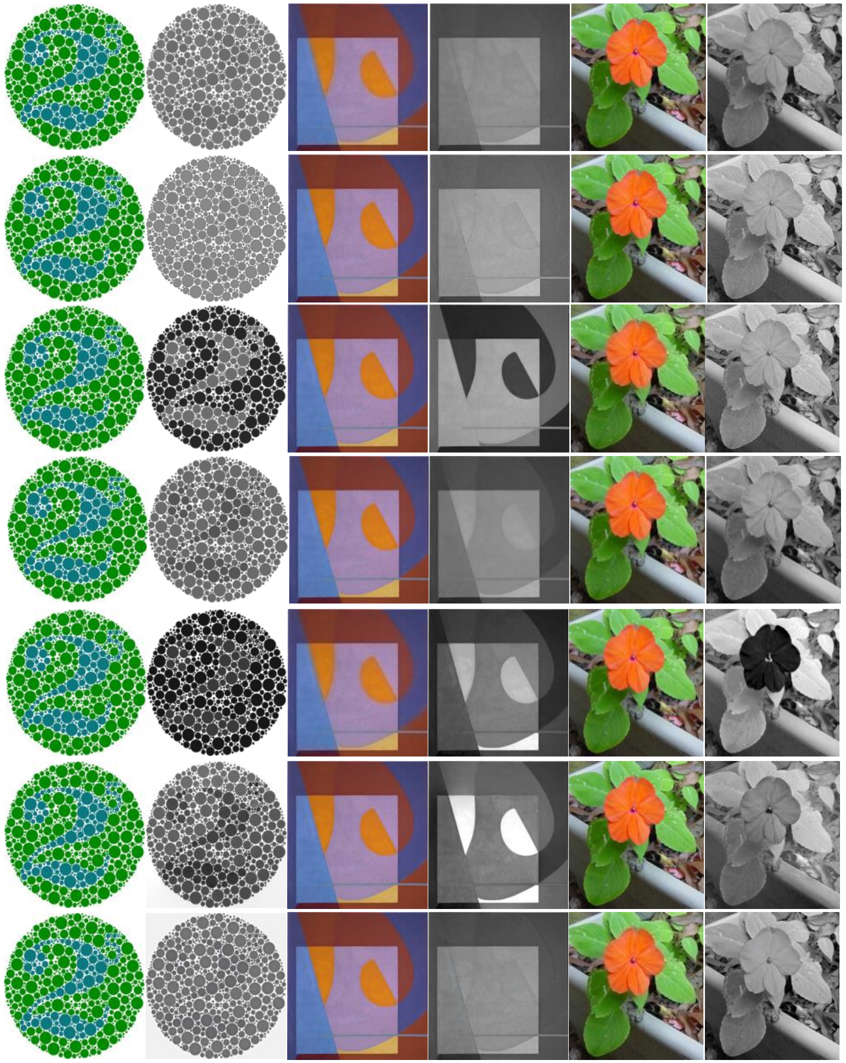


Detail Loss

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.

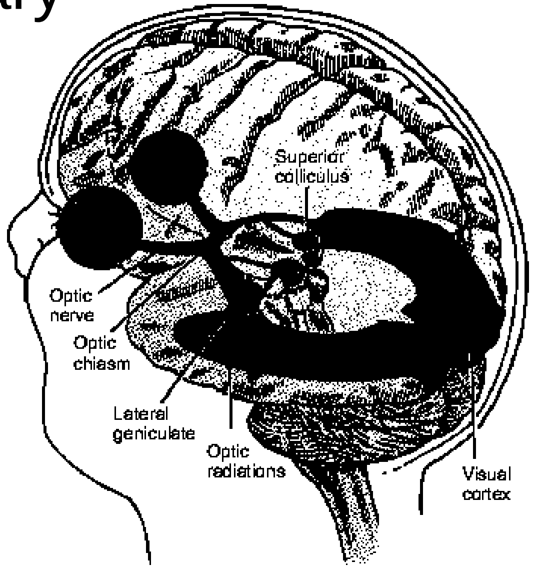
- DRIM absurdly used for comparing color-to-grayscale conversions



3) EXPERIMENTAL EVALUATIONS

Subjective Experiments

- Vision scientists, neurologists, physiologists, psychologists
  - push the knowledge about the HVS ahead
  - CG takes advantage of their results
- However, far from having computational model of HVS
- Experimental subjective analyses necessary
  - Validation and evaluation of methods
  - Deeper knowledge of examined field
  - Novel approaches

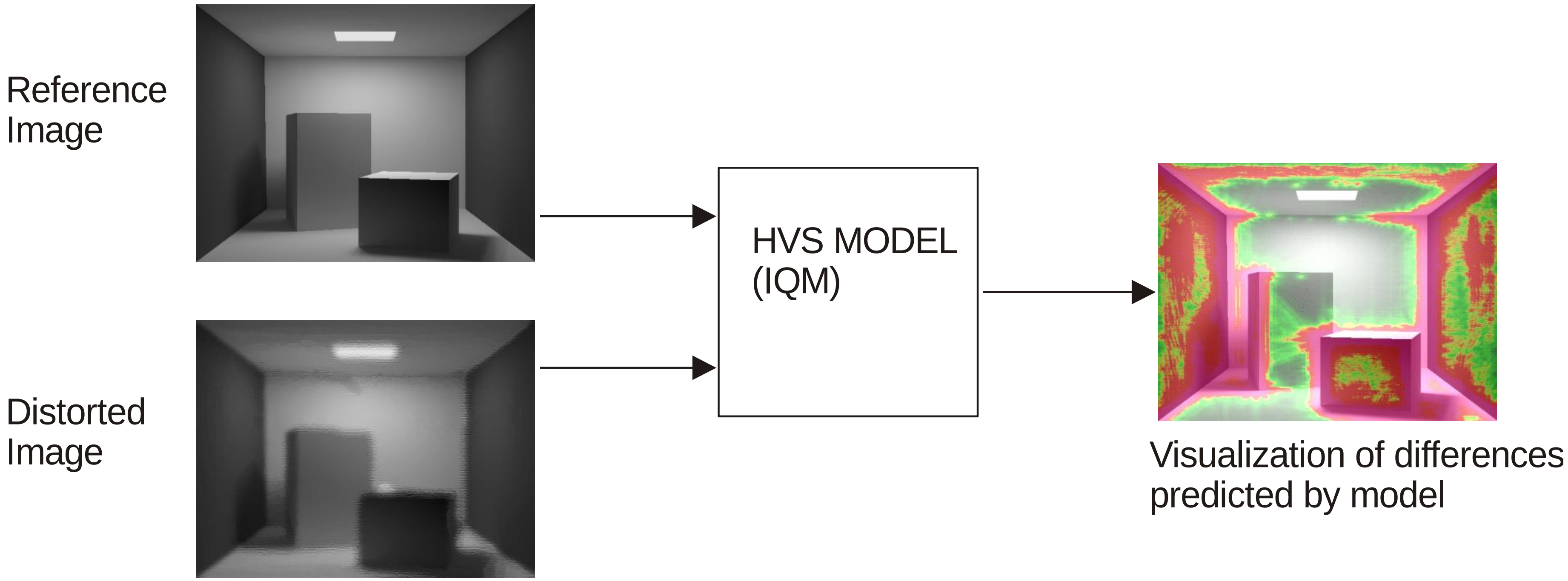


3.1 Overview – Evaluations of:

- Bottom-up vs. Top-down IQM
- STAR FR IQMs
- HDR Tone Mapping methods
- Color to Grayscale Conversions

3.1 Bottom-up vs. Top-down IQM

- automatic assessment of image quality
- image compression, global illumination, etc.

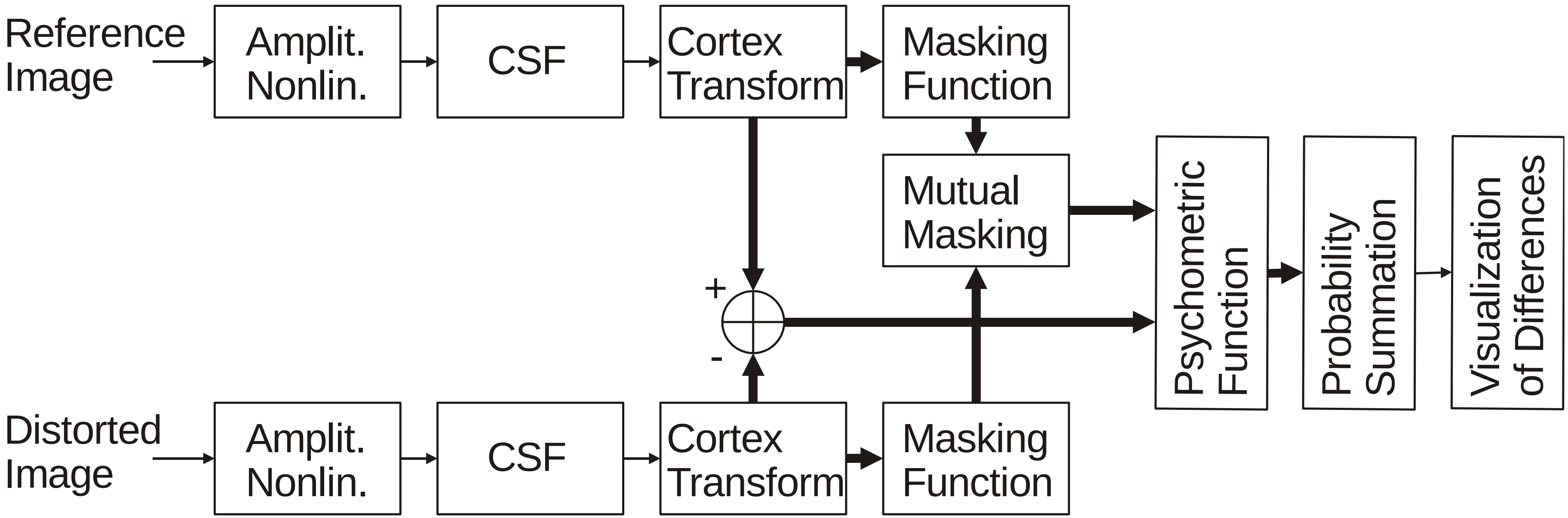


Traditional vs. Structural – Subjective Testing

- Traditional error sensitivity based approach
  - Bottom-up
  - VDP [Daly 93]
  - HDR-VDP [Mantiuk et al. 11]
- Structural similarity based approach
  - Top-down
  - SSIM [Wang et al. 04]

Error Sensitivity Based Approach

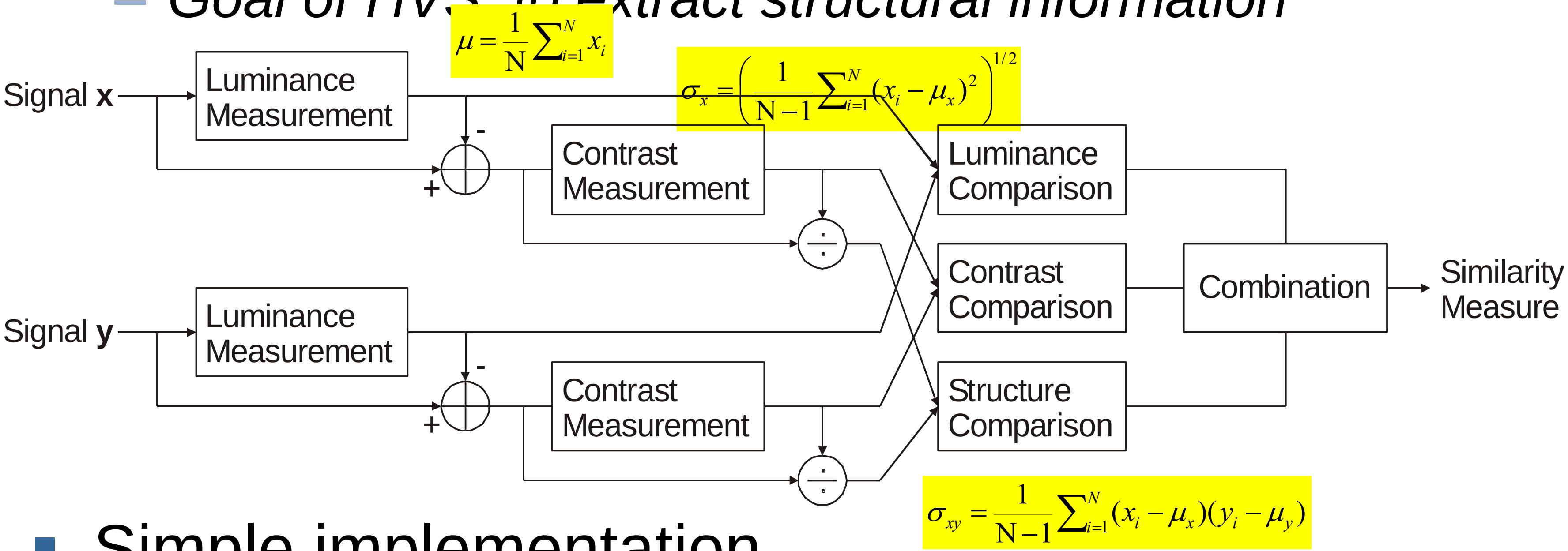
- Visible Differences Predictor (VDP) [Daly 93]



- Threshold sensitivity
- Early vision modeling
- Visual Masking

Structural Similarity Based Approach

- Structural SIMilarity Index [Wang et al. 04]
  - *Goal of HVS: to extract structural information*



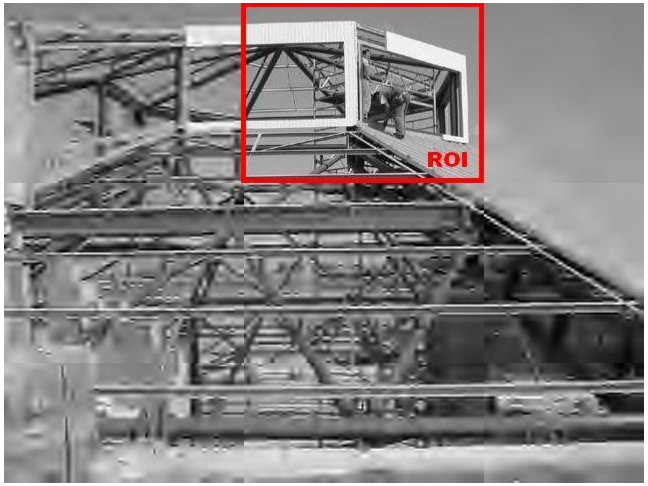
- Simple implementation
- Fast computation

Traditional vs. Structural – Subjective Testing

- Independent subjective tests
  - 32 subjects
  - 30 uniformly compressed images (JPEG2000)
  - 30 ROI compressed images
  - difference expressed by ratings
- Mean Opinion Scores

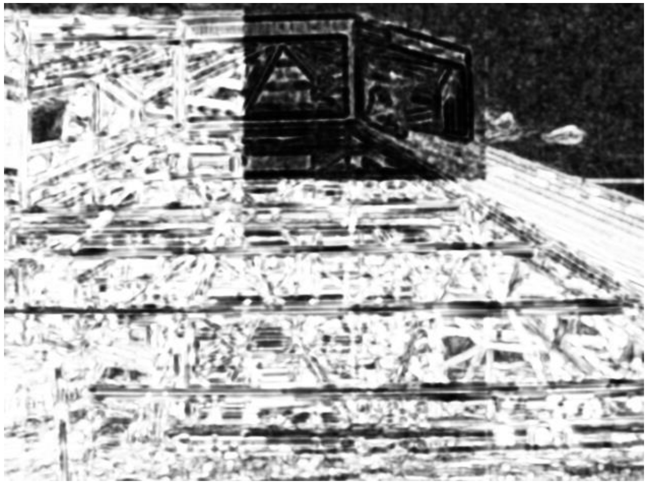
Traditional vs. Structural – Objective Testing

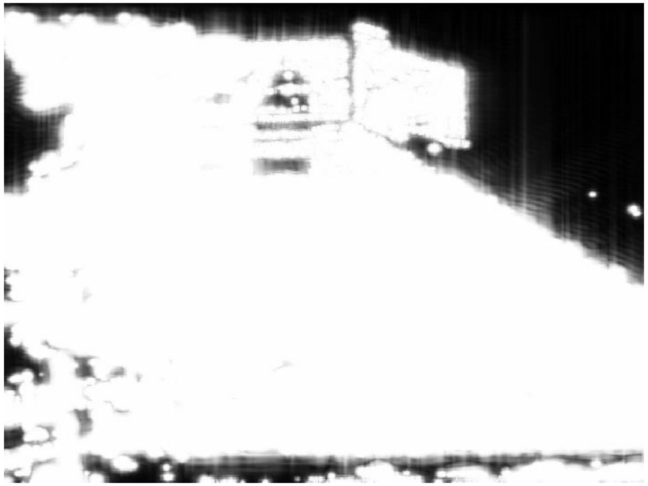
Original input image



ROI compressed input image

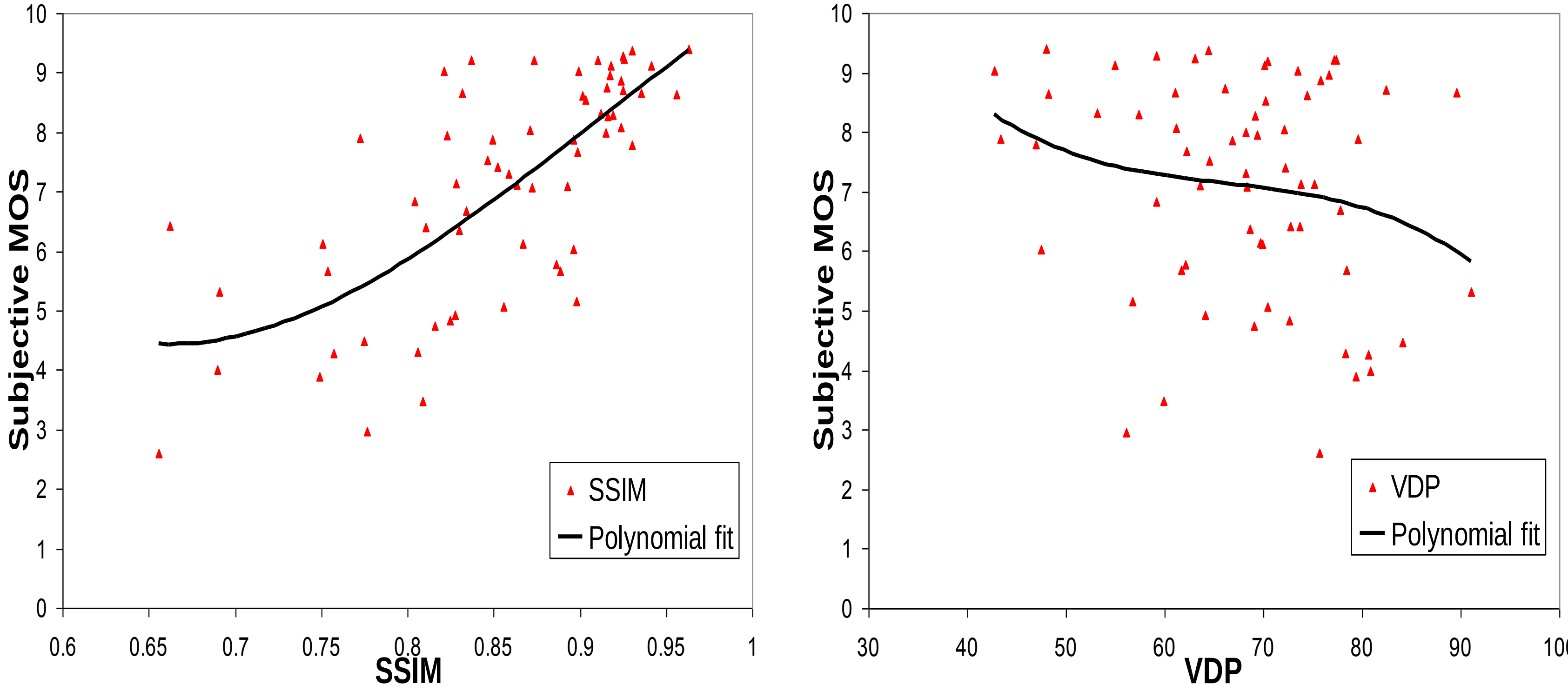
SSIM probability map





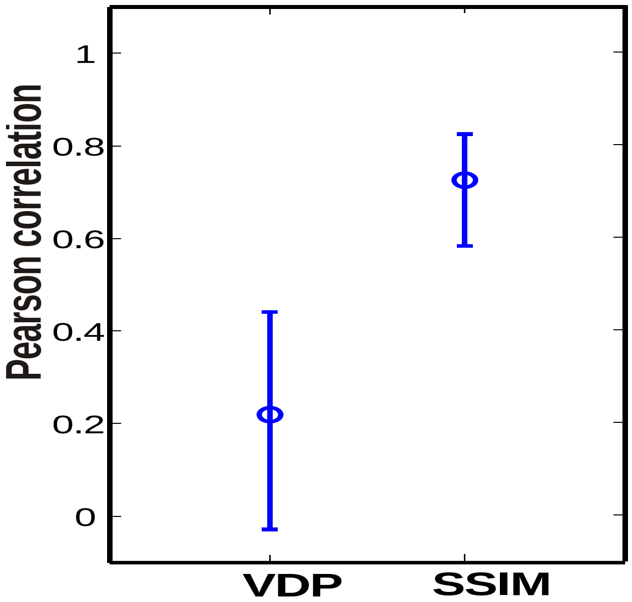
VDP probability map

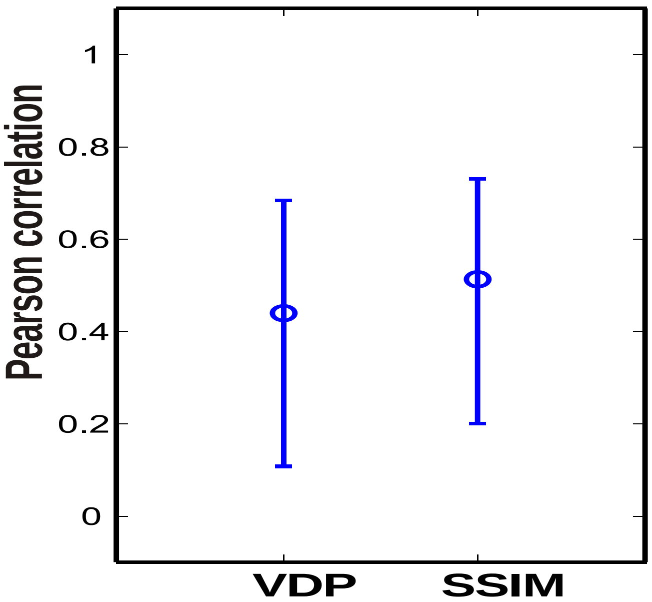
Traditional vs. Structural – Test Results



Quality predictions compared to subjective MOS for the SSIM (left) and for the VDP (right)

Traditional vs. Structural – Test Results (cont.)





	r		**r (ROI)**
VDP	0.22		0.44
SSIM	0.72		0.51

r – Pearson (parametric) correlation coefficient

Traditional vs. Structural – Conclusions

- Simple approach not necessarily worse
- Poor performance in ROI-compression task

3.2 Overview – Evaluations of:

- Bottom-up vs. Top-down IQM
- STAR FR IQMs
- HDR Tone Mapping methods
- Color to Grayscale Conversions

3.2 Evaluation of STAR FR-IQM

- 6 IQMs: AD (PSNR, MSE), sCIE-Lab, sCORREL, SSIM, MS-SSIM, HDRVDP-2

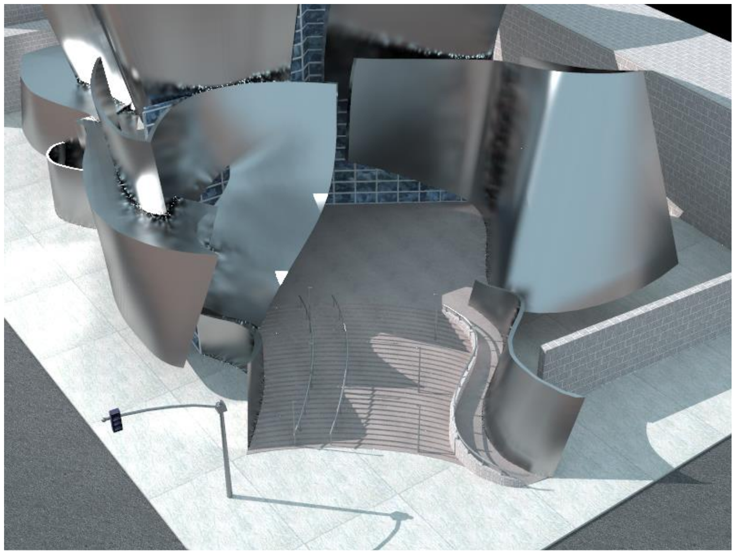
- How good are IQMs in **localizing** artifacts?
- Evaluation of distortion **maps** (not just mean-opinion-scores, i.e. one number per image)
- Computer graphics-generated contents and artifacts
- Two subjective tasks: given reference image and with no reference image

[Čadík et al. SIGGRAPH Asia, 2012]

http://www.mpii.de/resources/hdr/iqm-evaluation/

Our Dataset: Example Rendering Artifacts

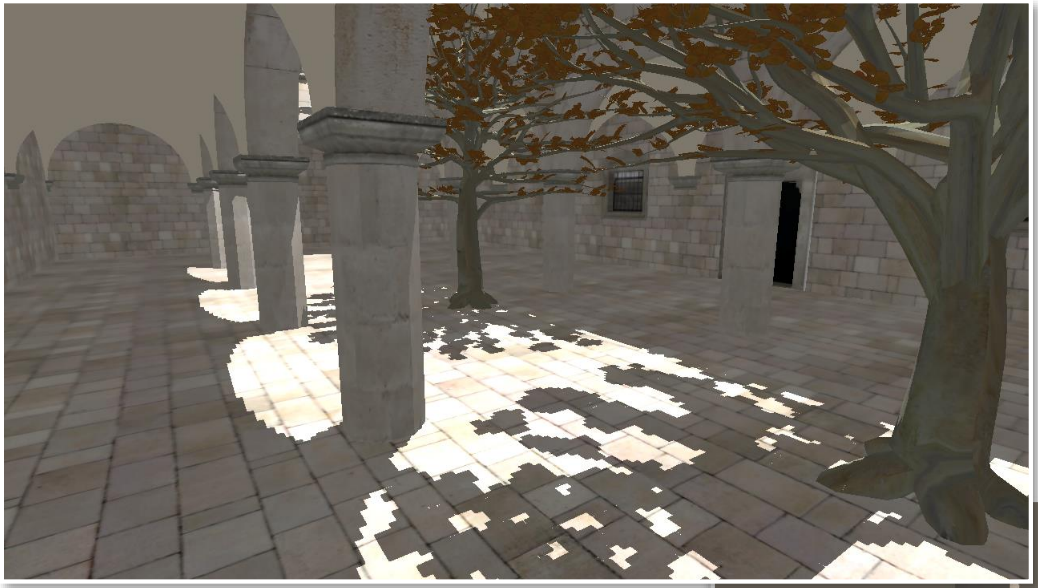
- e.g., low-freq. noise from glossy instant radiosity or photon density estimation



Rendering Artifacts

- Clamping Bias
  (darkening in corners)

Rendering Artifacts



- Shadow Mapping
  easy to generate large sample set

Rendering Artifacts

- Progressive photon mapping: when to stop iterating?

1 iteration

2 iteration

8 iteration

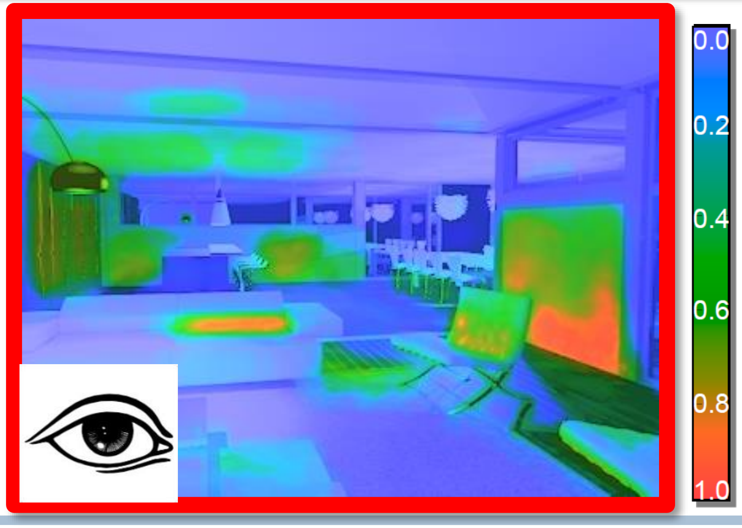
60 iteration

150 iteration

1500 iteration

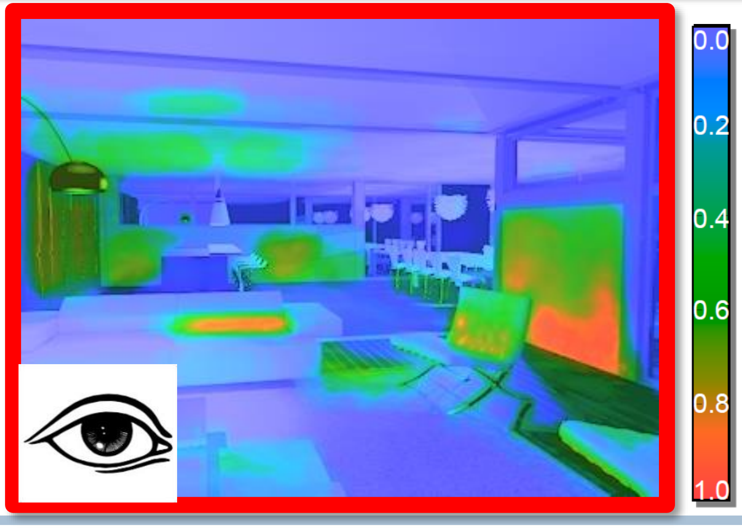
User Experiment - Mean Distortion Maps

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



User Experiment – With Reference

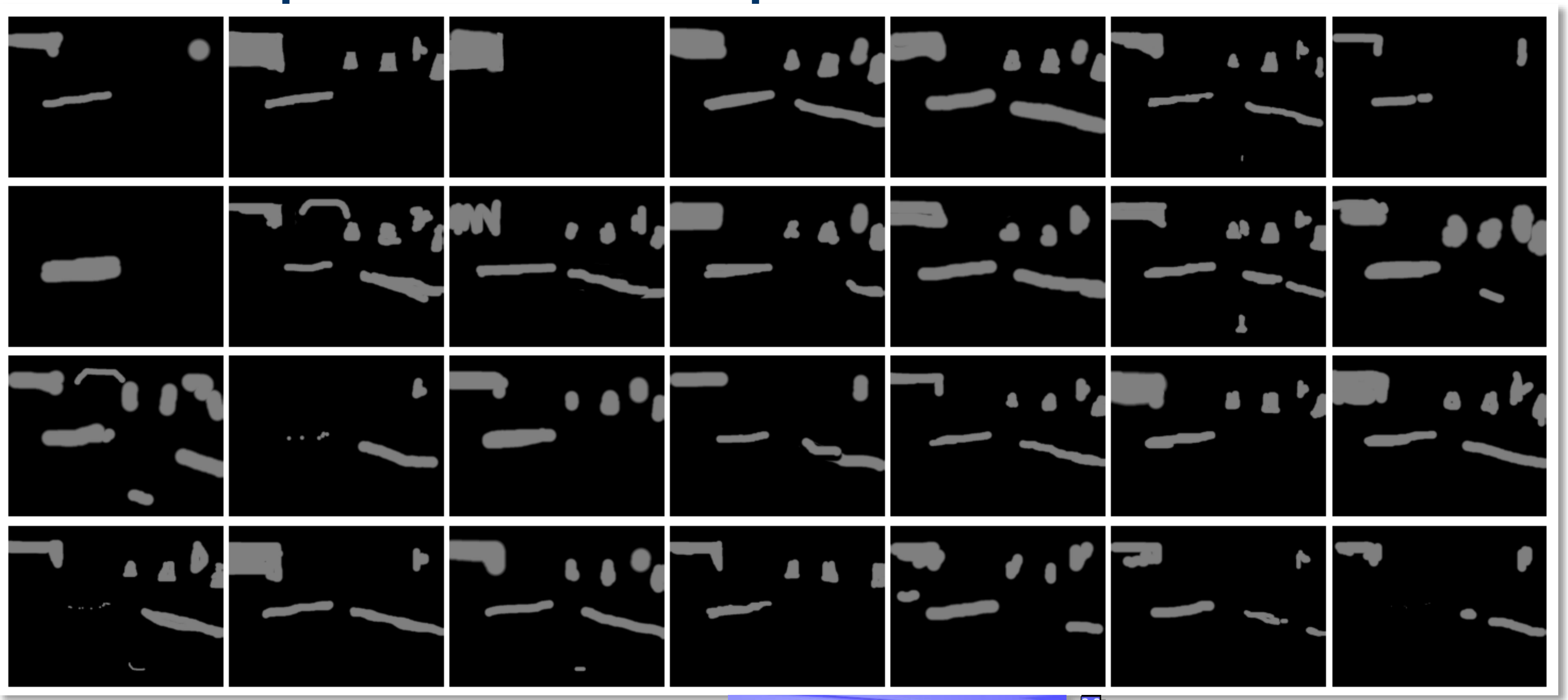
▪ Noticeable distortions



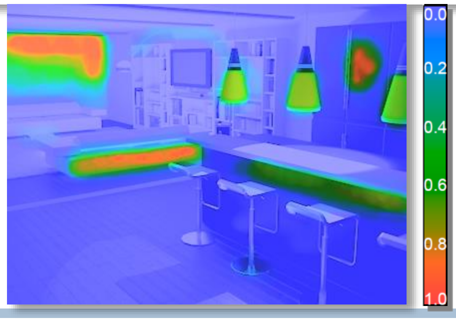
User Experiment – No Reference

- Objectionable distortions

Example User Responses

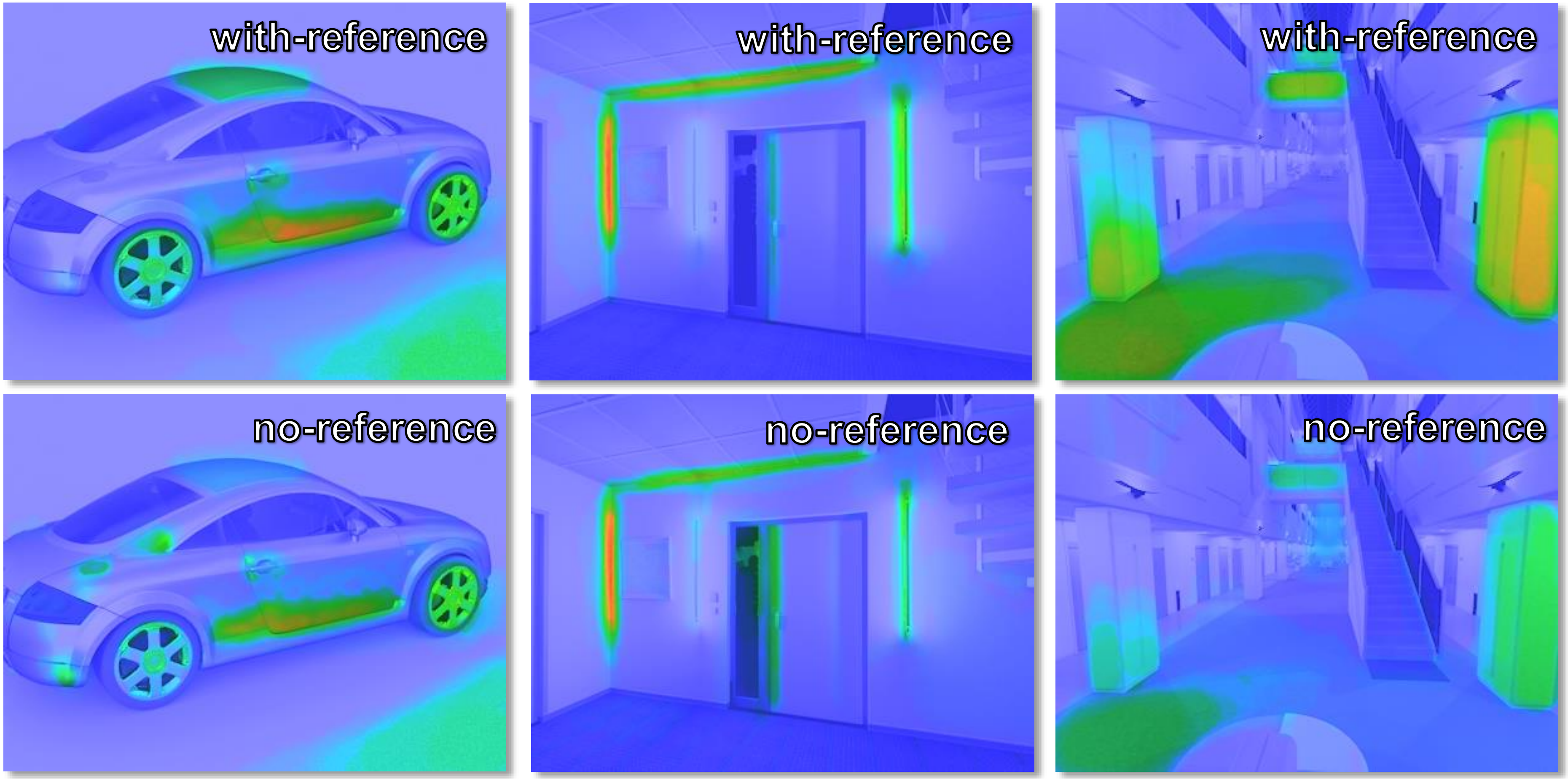


- Probability of detection



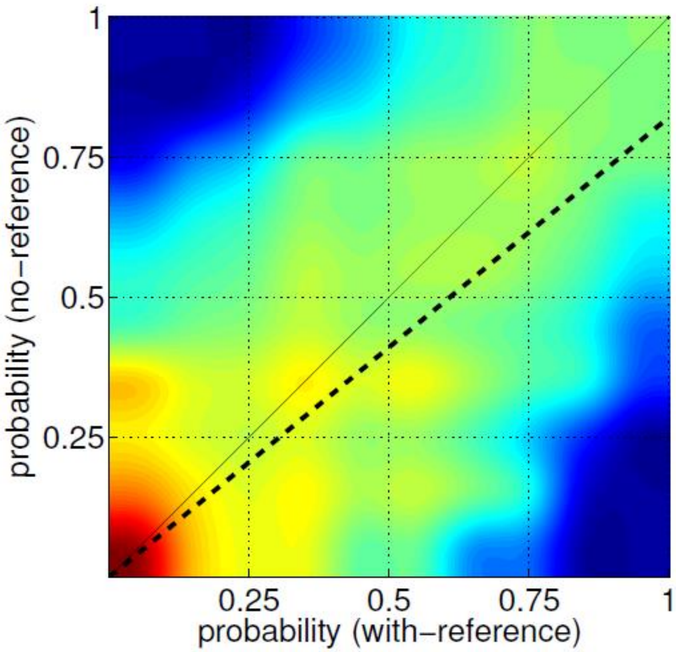
With-reference vs. No-reference

- Results rather similar

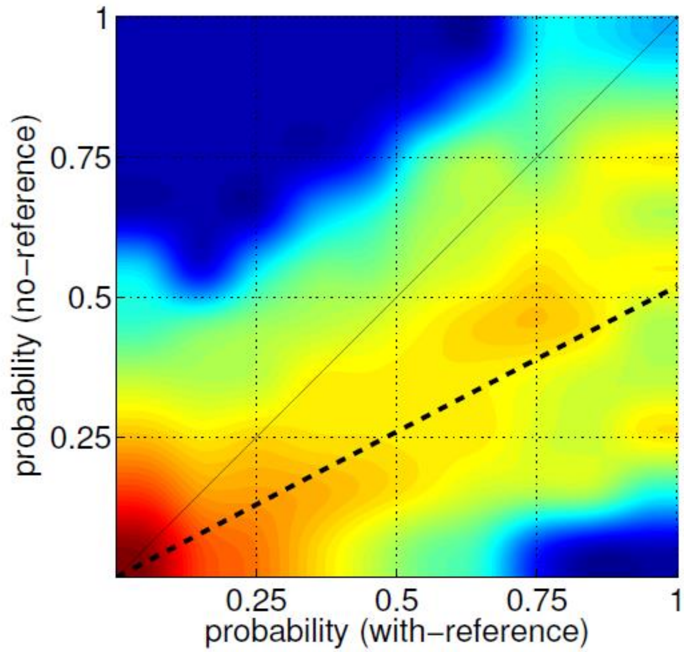


With-reference vs. No-reference (cont.)

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

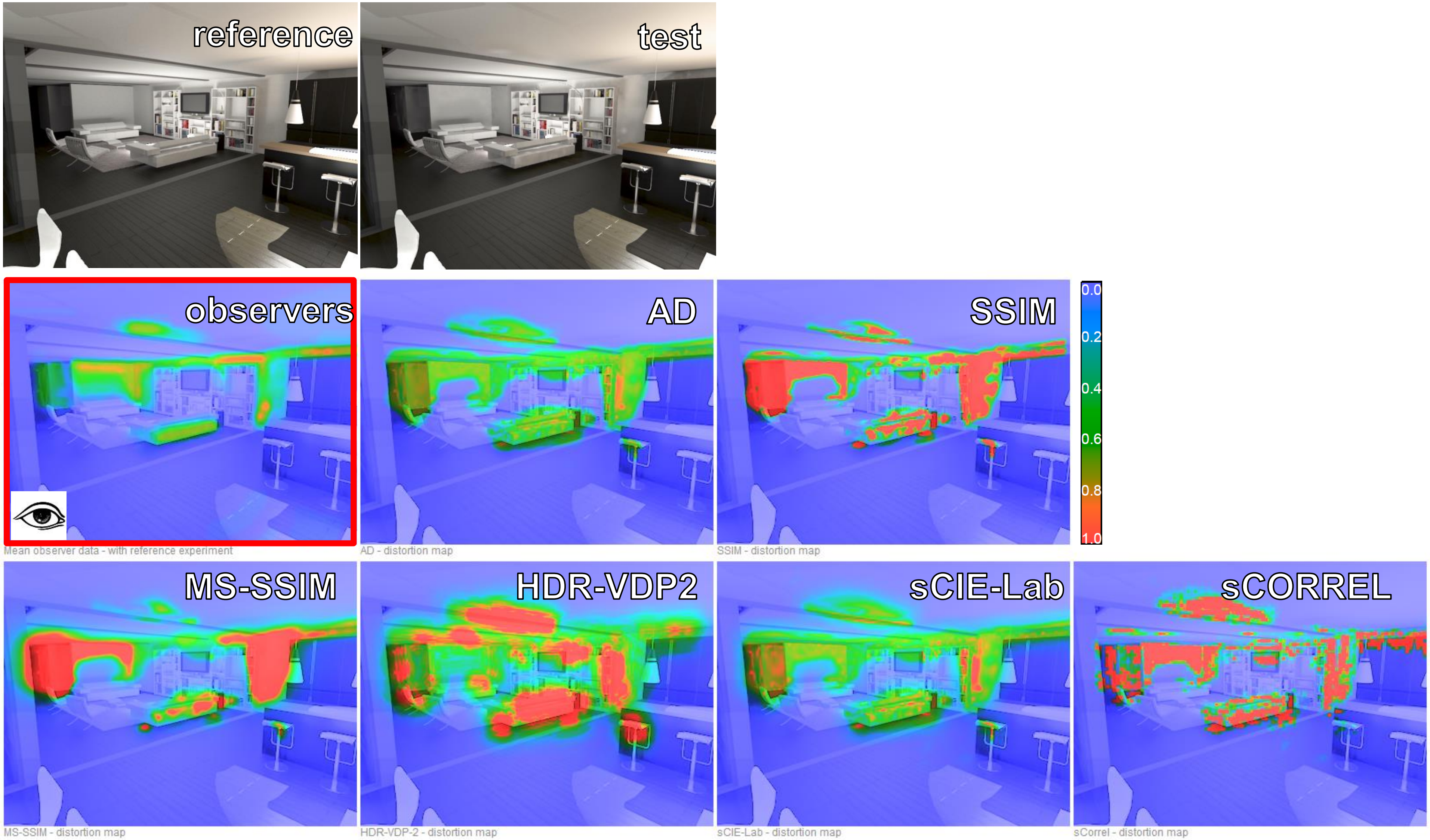


EG'12 dataset

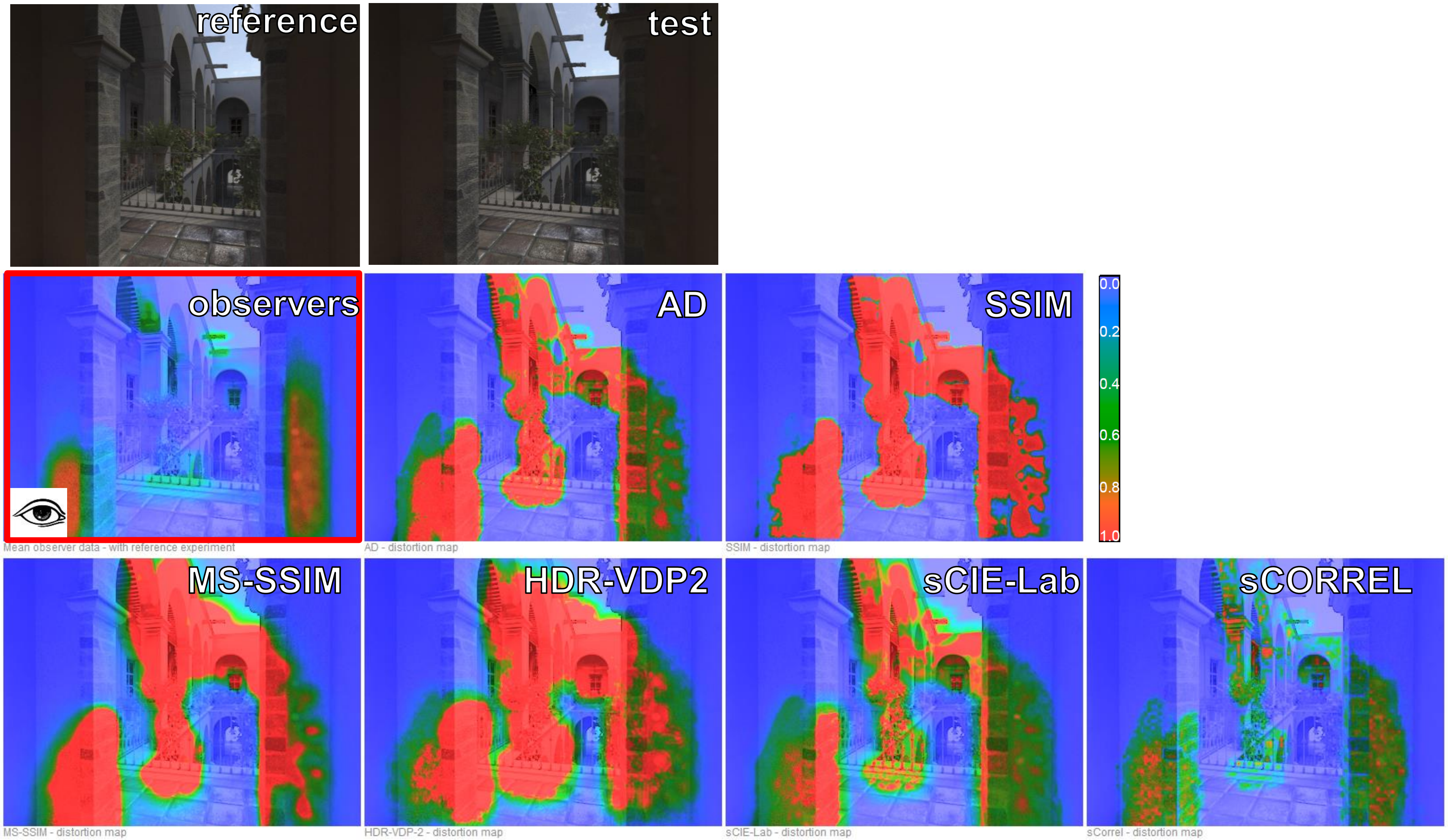


new dataset

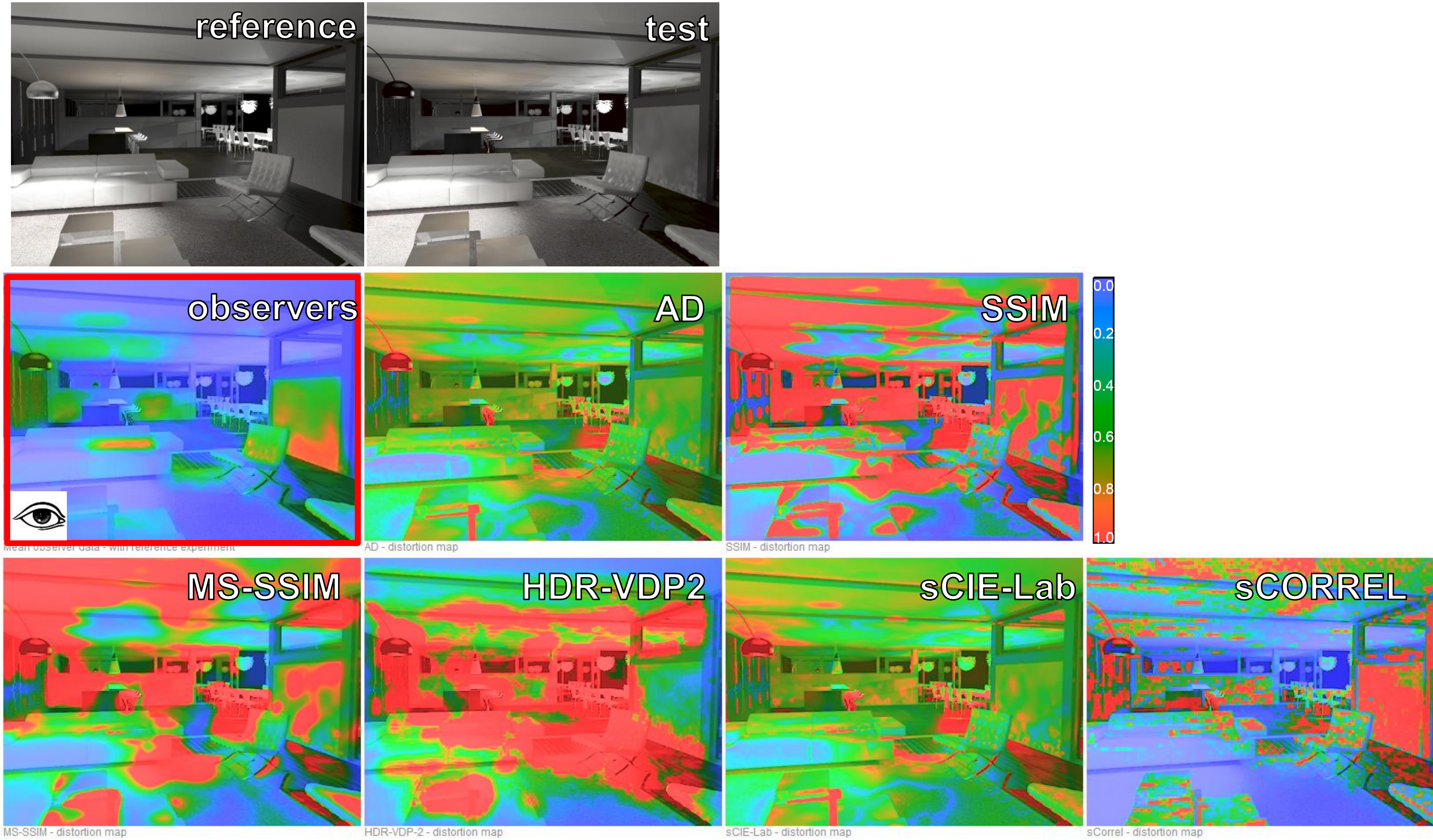
Results – Example of Metric Predictions



Results – Example of Metric Predictions

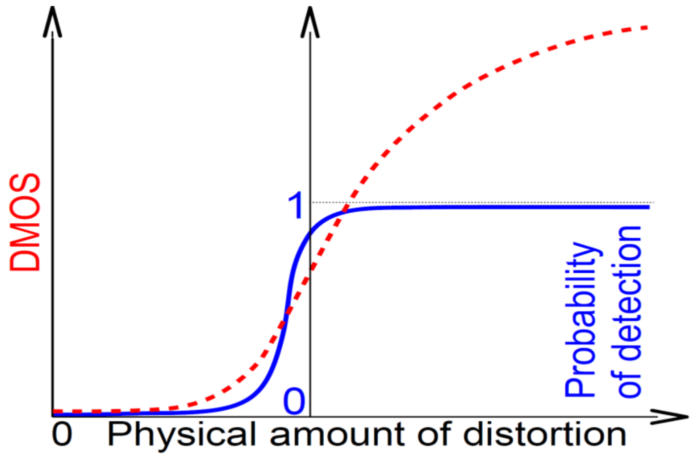


Results – Example of Metric Predictions

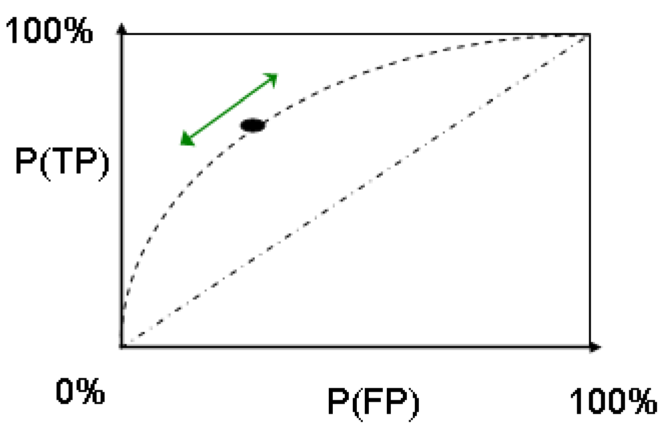


Measures of Metric Performance

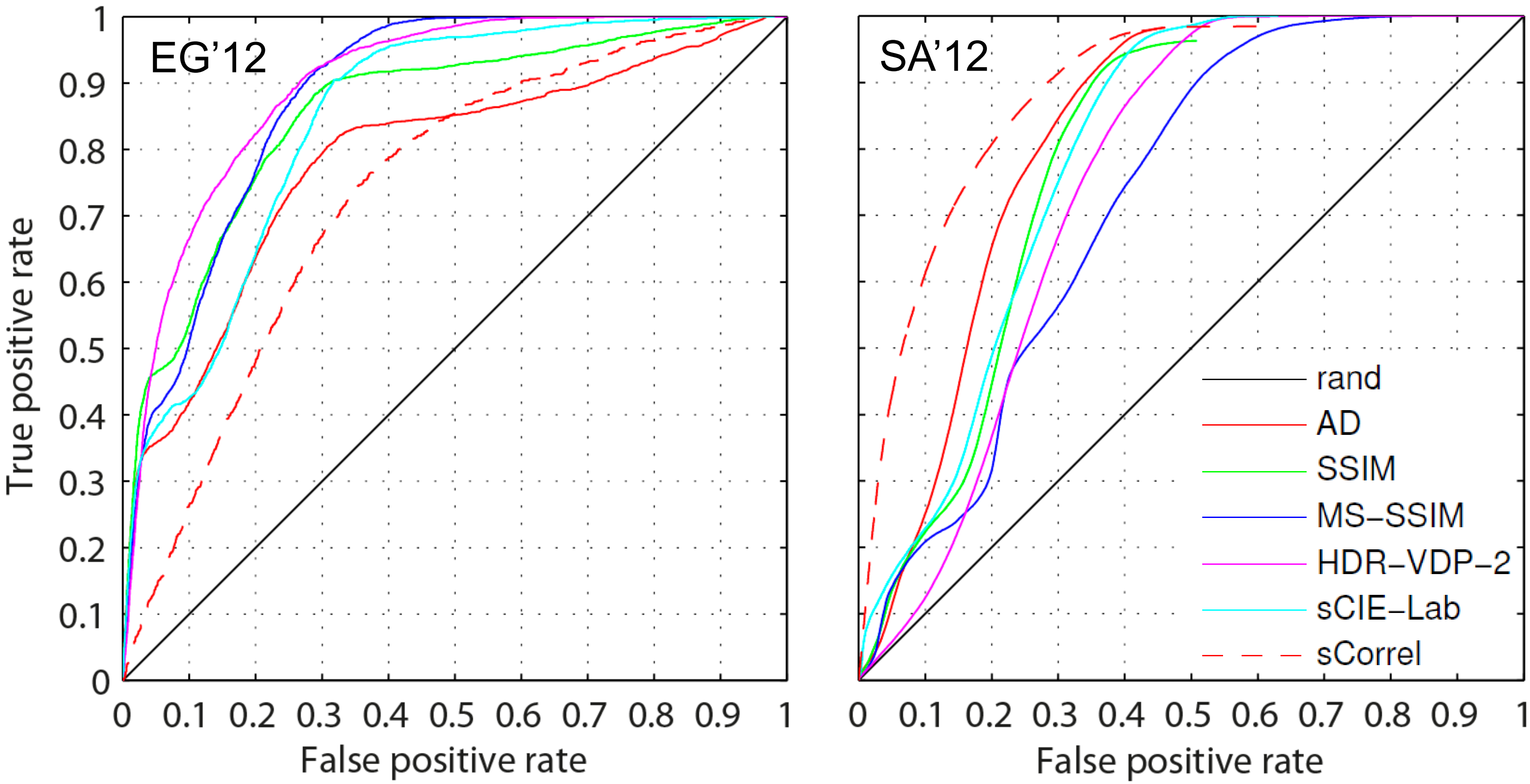
- Previous experiments
  - MOS/DMOS {1,2,3,4,5}
- No easy way to capture MOS locally
  - Probability of detection [0,1]



- Receiver operating characteristic (ROC)
  - Area under curve (AUC)
  - Thresholds (25%, 50%, 75%)



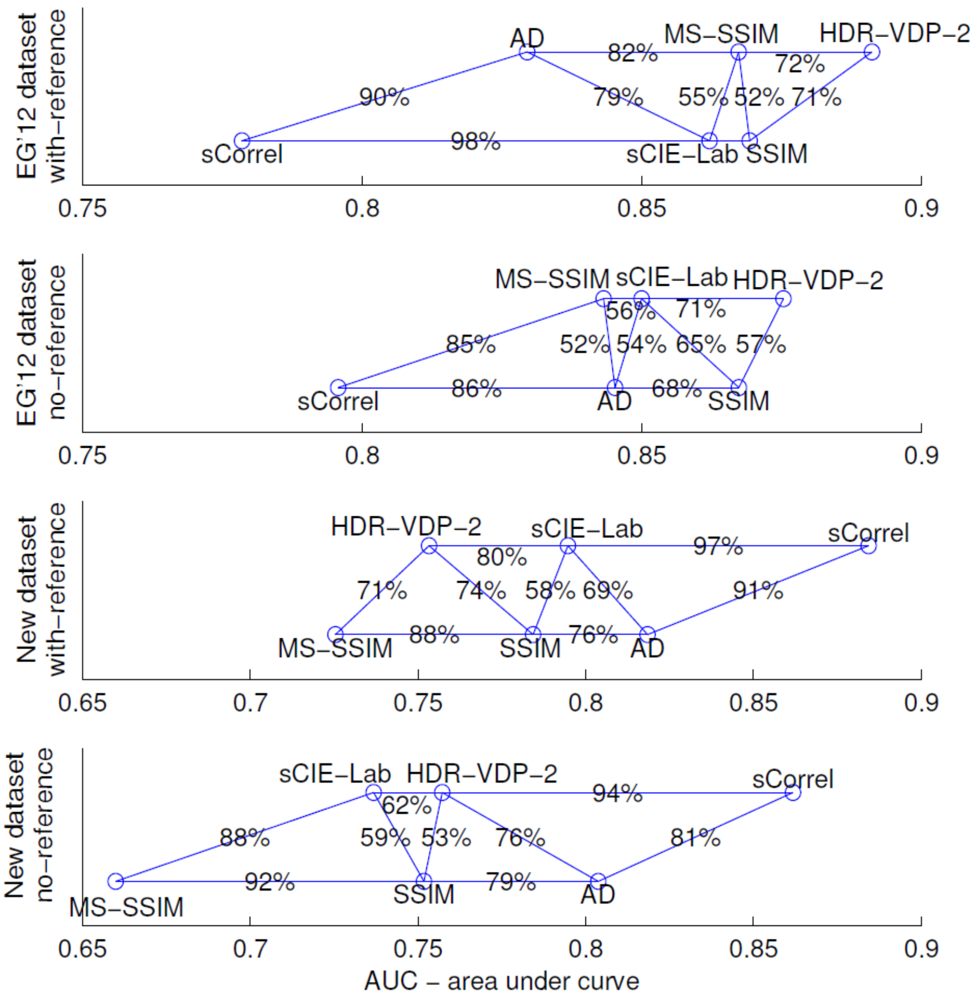
Metric Performance Comparison – ROC



- With-reference experiment results (see paper for no-ref.)

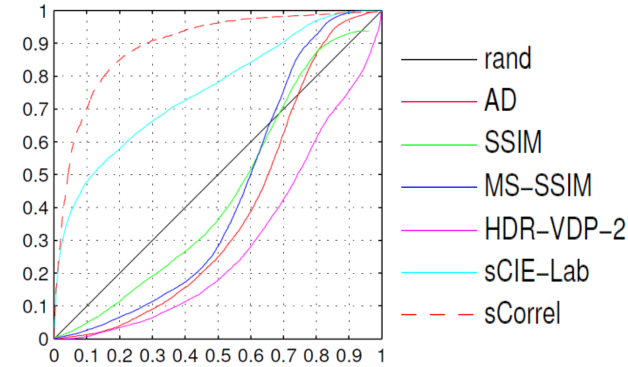
Metric Performance Comparison (cont.)

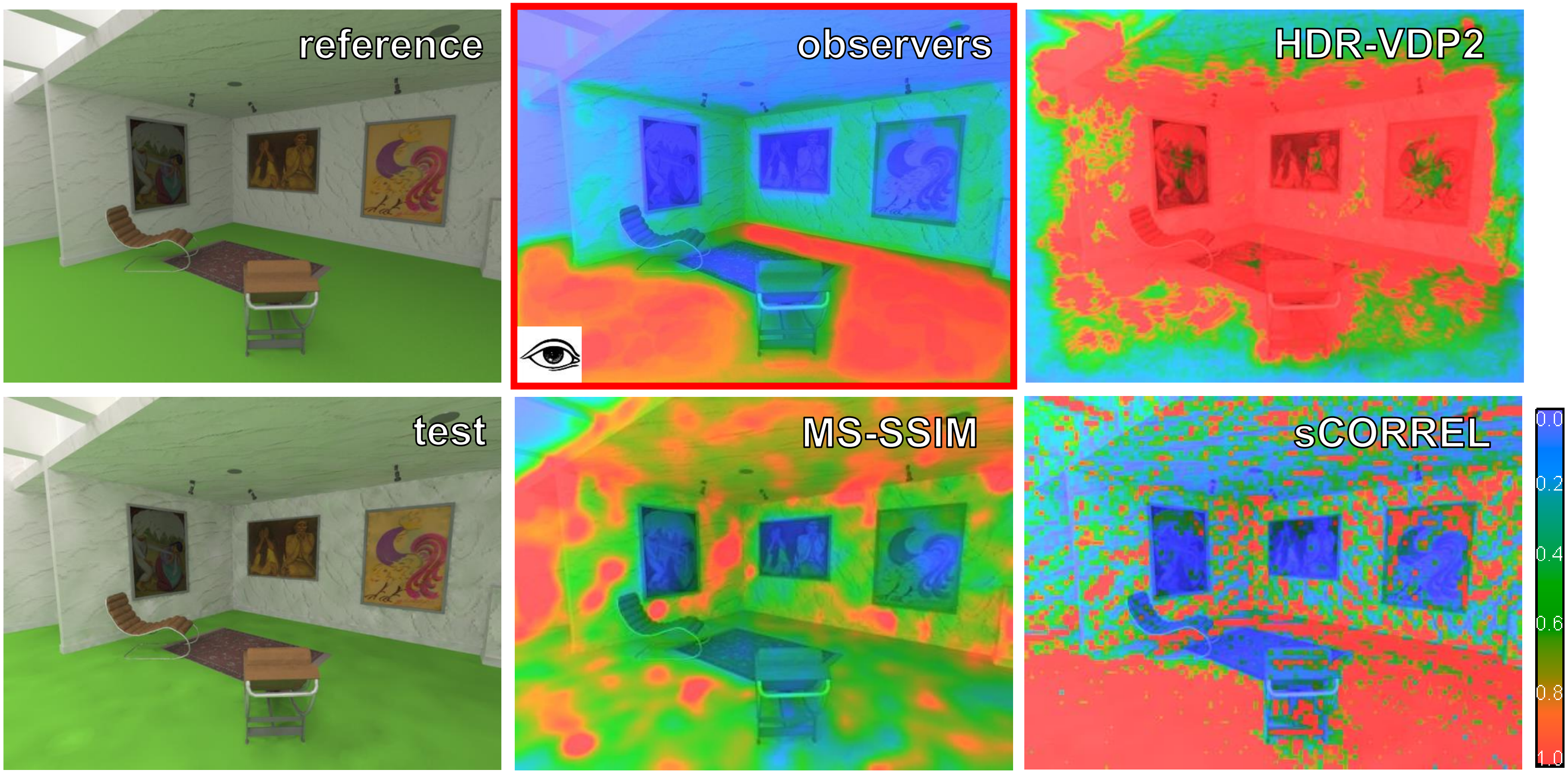
- Bootstrapping (randomization with repetitions 500x)
  - Bonferroni correction
- **No** statistically significant difference between IQMs
- Performance differs significantly per scene



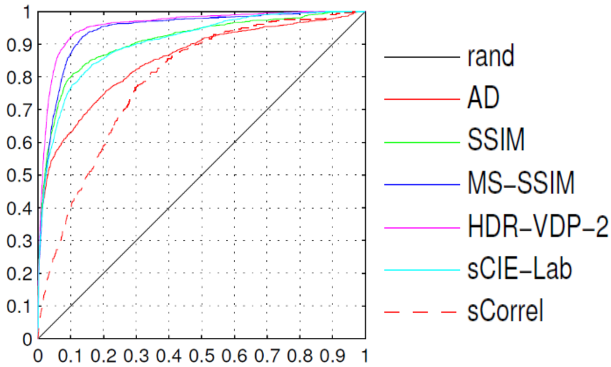
Analysis of Metric Failures

Brightness and contrast change

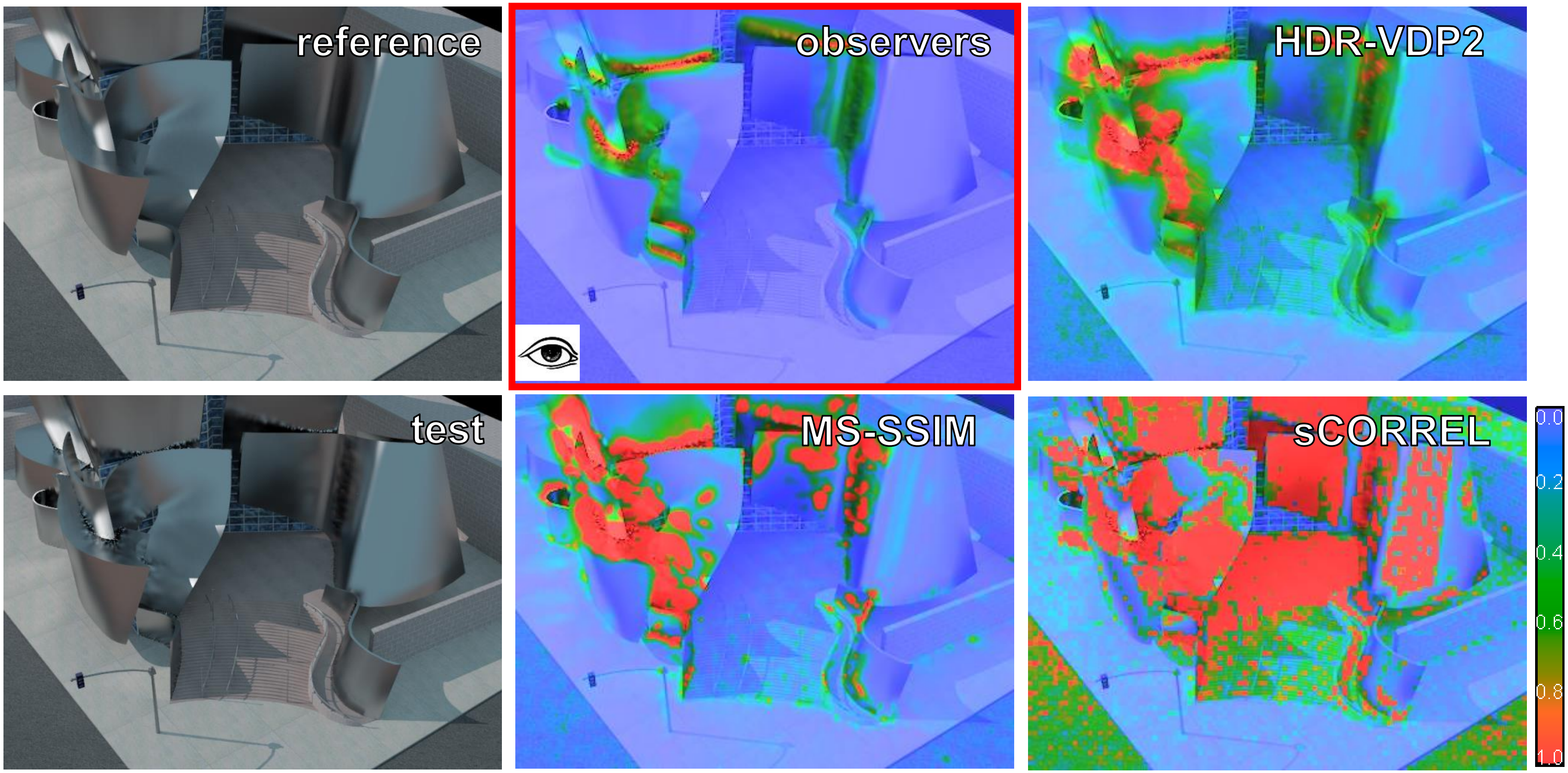




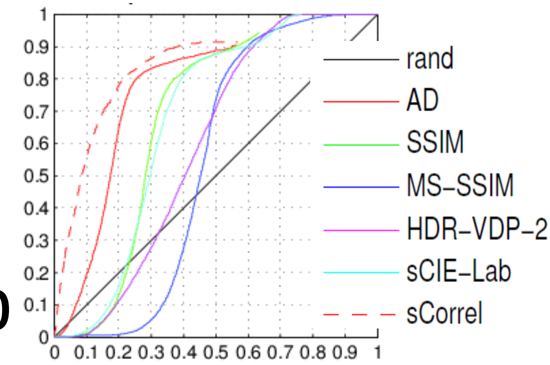
Analysis of Metric Failures



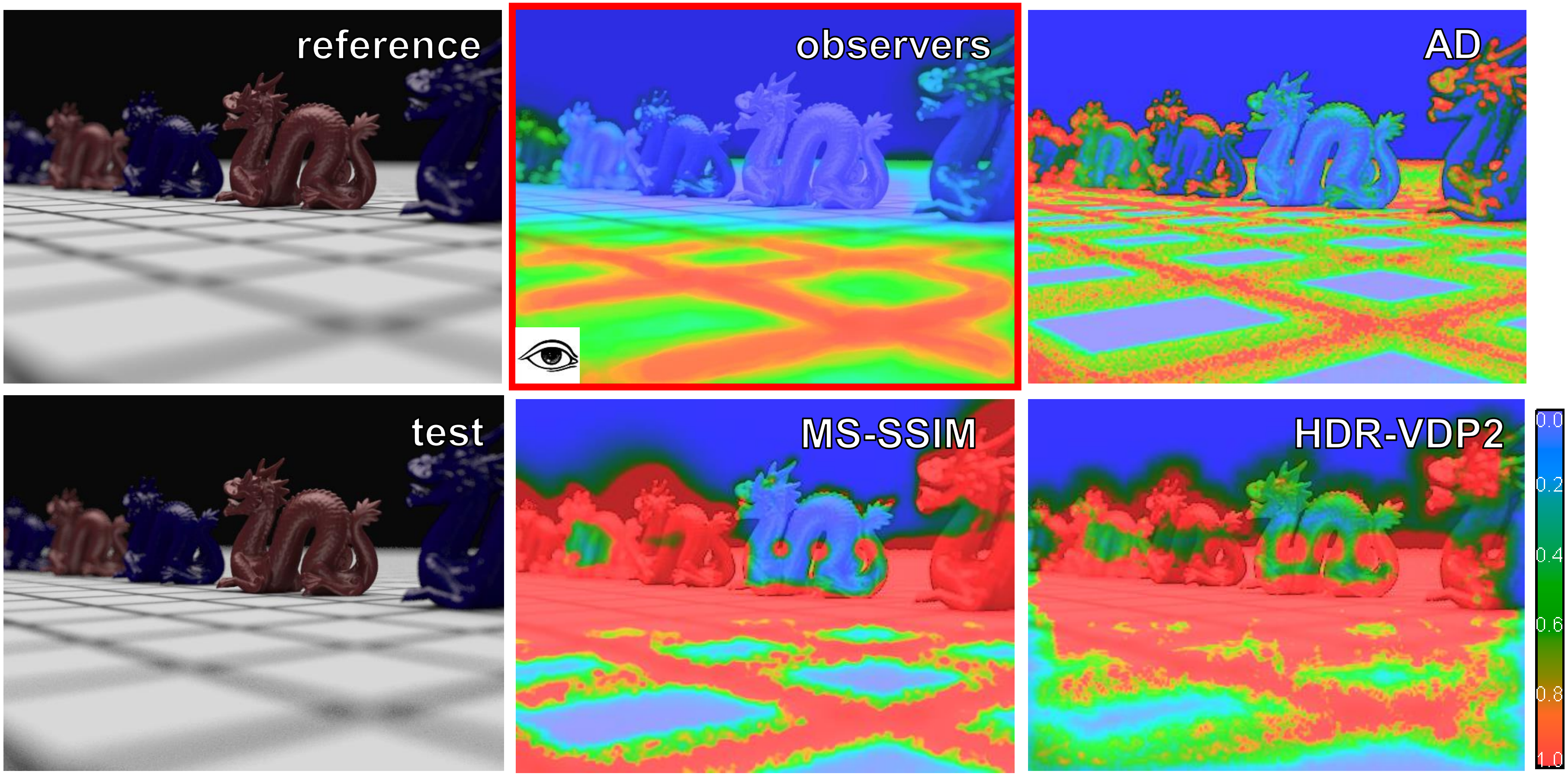
Visibility of low-contrast differences



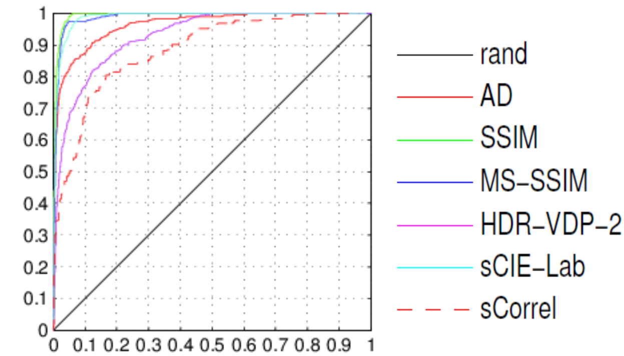
Analysis of Metric Failures



Spatial accuracy of the prediction map



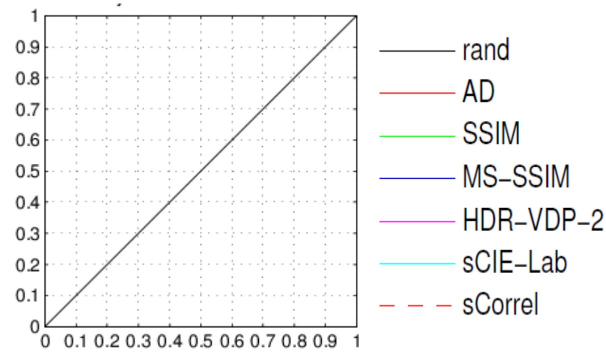
Analysis of Metric Failures

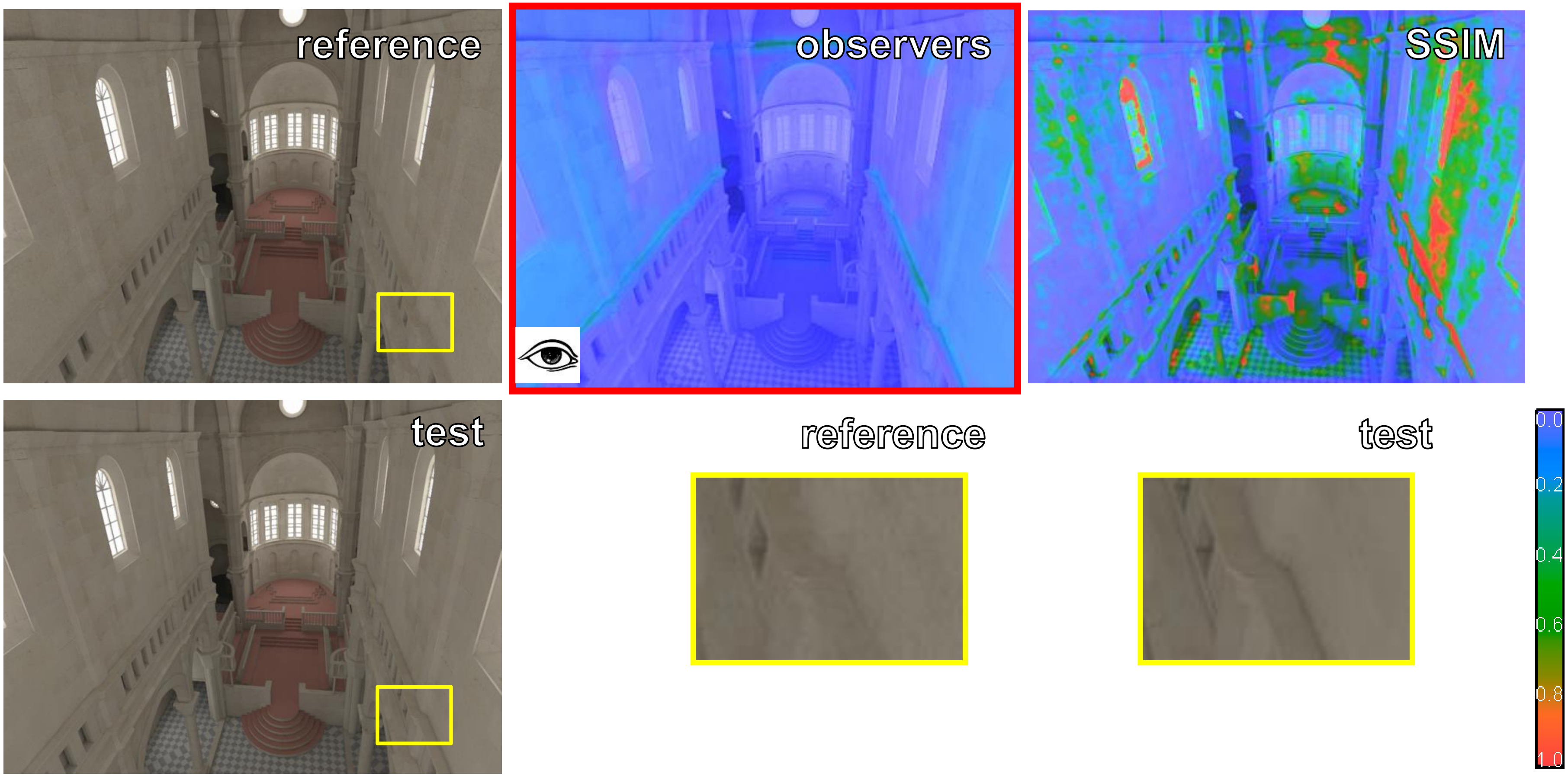


Plausibility of shading

Analysis of Metric Failures

Plausibility of shading (cont.)





Conclusions

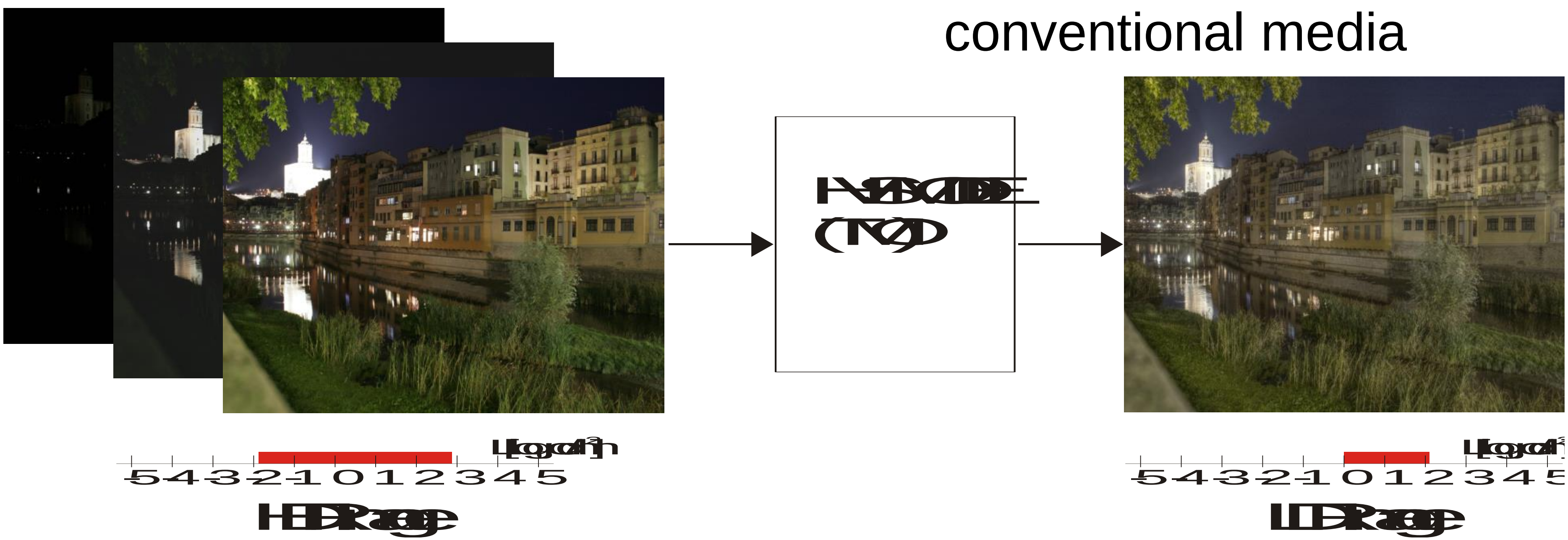
- Rendering datasets for IQM evaluation with subjective localized distortion maps
- With reference ≈ no-reference experiments
- State-of-the-art IQMs far from subjective ground-truths
- No universally reliable metric exists
- Large space for improvements

3.3 Overview – Evaluations of:

- Bottom-up vs. Top-down IQM
- STAR FR IQMs
- HDR Tone Mapping methods
- Color to Grayscale Conversions

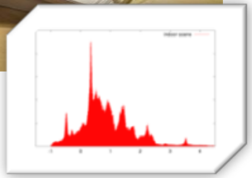
3.3 Evaluation of HDR Tone Mapping

- HDR images
  - several orders of magnitude
  - high precision
- LDR (ordinary) images
  - typically 8b per channel [0, 255]
  - low precision
  - displayable on conventional media

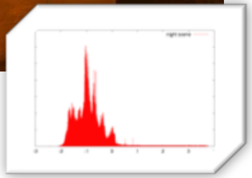


Evaluation of TMO

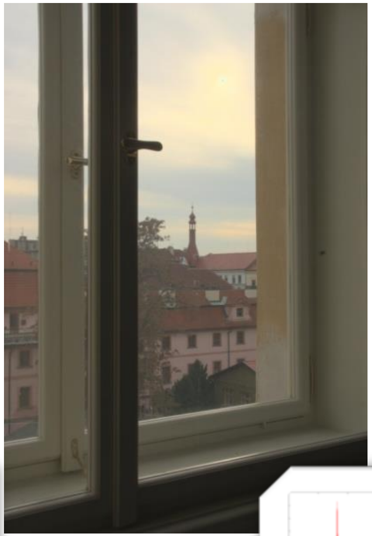
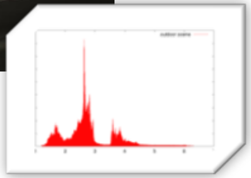
- 14 tone mapping operators
- 3 real-world scenes (natural)

indoor

night

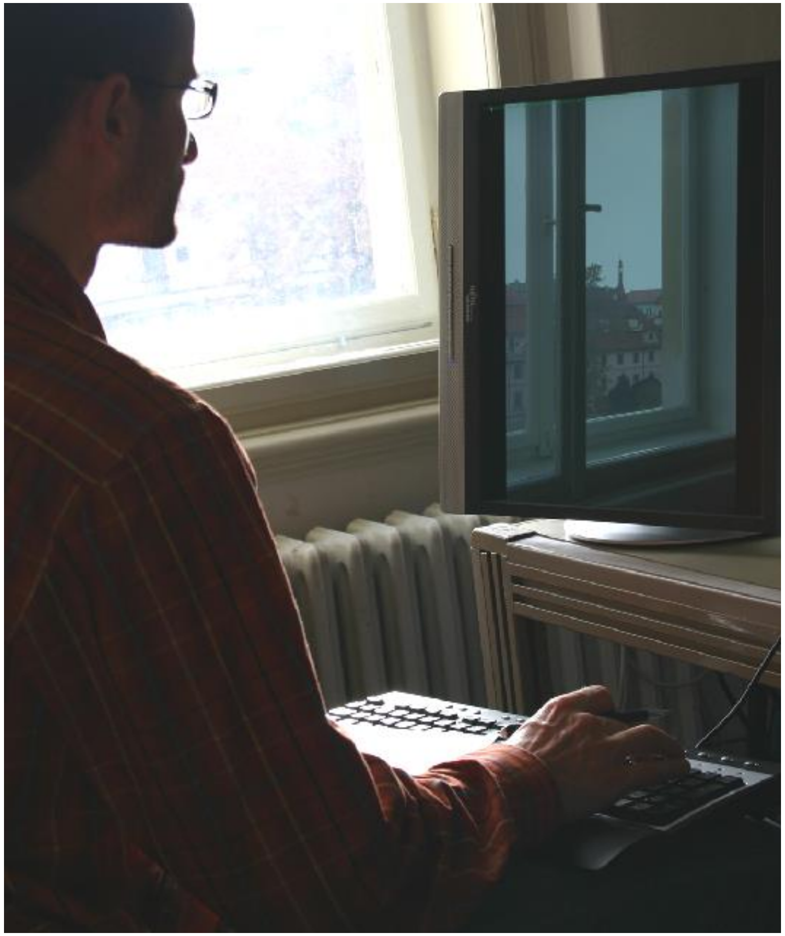
outdoor

- 6 basic attributtes
  - overall i.q., contrast, brightness, colors, details, artifacts
- 20 observers

[Čadík et al. 06], [Čadík et al. 08]

Perceptual Experiments

1) With reference study
   - original HDR scene and tone mapped image presented simultaneously
   - direct rating to the real world



Perceptual Experiments

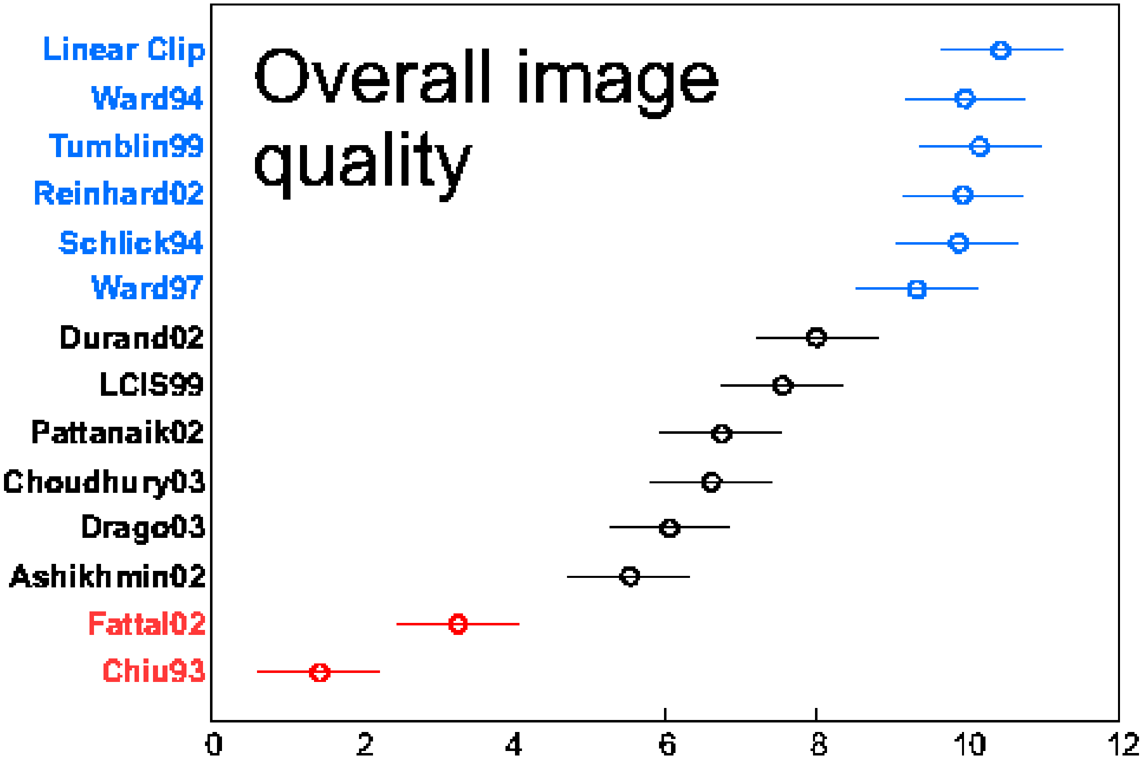
2) Without reference study
   - *another* observers (not aware of the original scene)
   - ranking
   - high-quality color printouts
   - mental model

Results

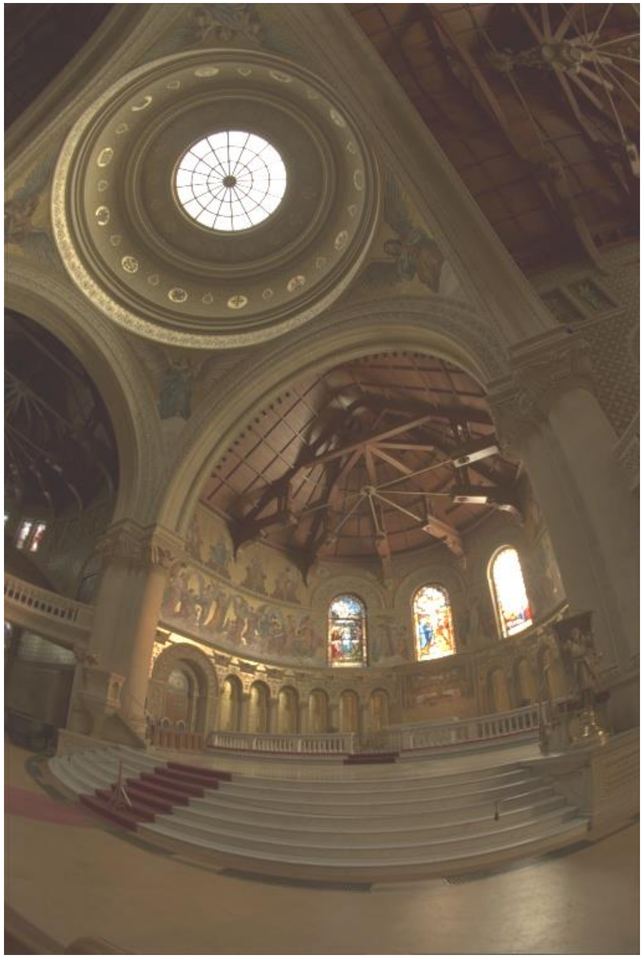
- In total: more than 5000 scores collected
- Parametric tests
  - ANOVA, MANOVA
  - Pearson correlations, etc.
- Nonparametric tests
  - Kruskal-Wallis, Friedman’s test, n-way nonparametric ANOVA
  - PERMANOVA (nonparametric MANOVA)
  - Spearman correlations, etc.

Results – Evaluation of TMO

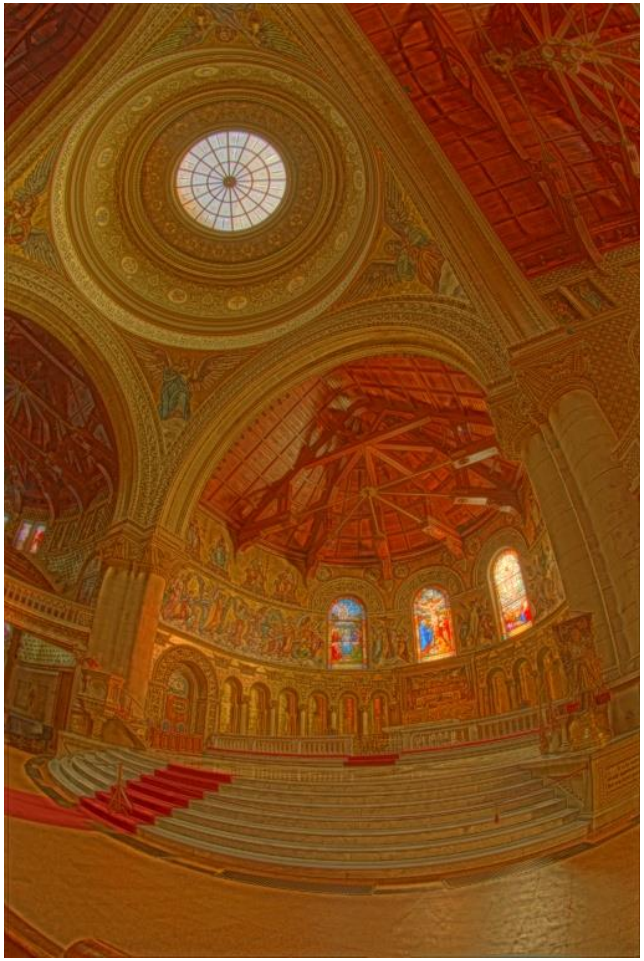
- Groups of methods



Global and Local Methods

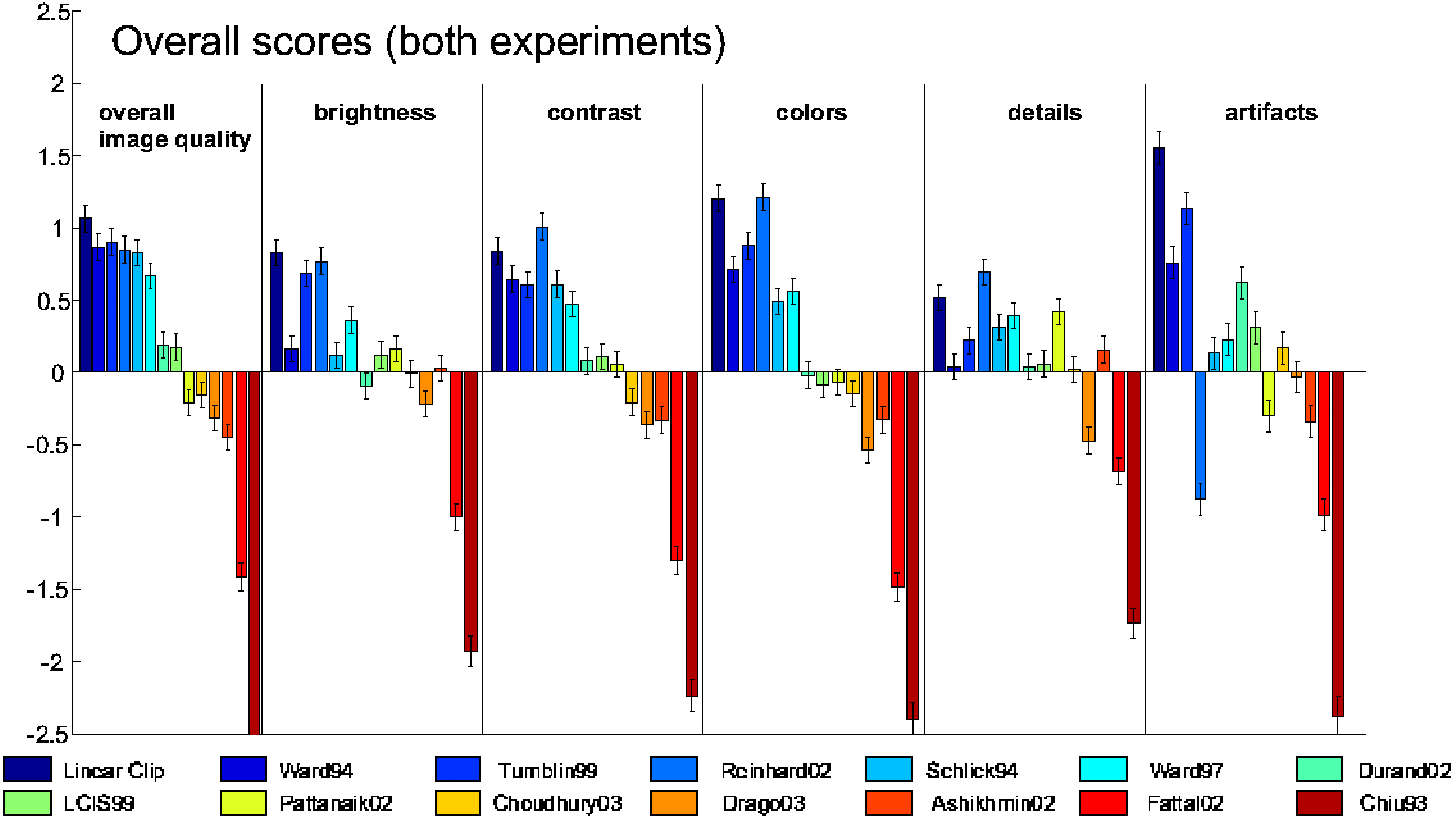


[Ward94]

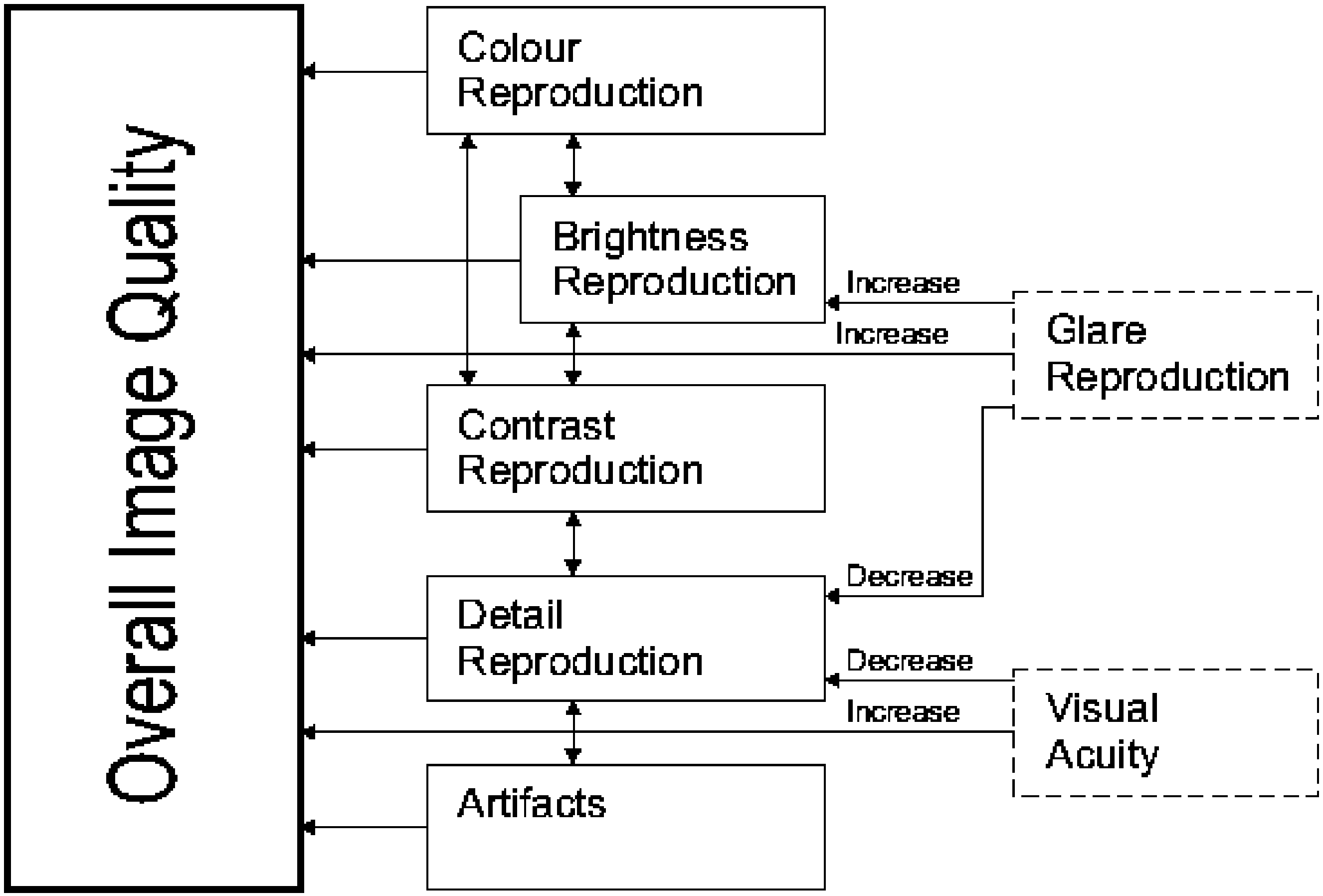


[LCIS99]

Results – Evaluation of TMO



Results – Image Attributes



Results – Image Attributes

- correlations between attributes verified
- highest importance of contrast

- regression results
  - various methods used

OIQ == 0.37 Contrast +

0.36 Colors +

0.21 Artifacts +

0.07 Brigtness +

0.06 Details

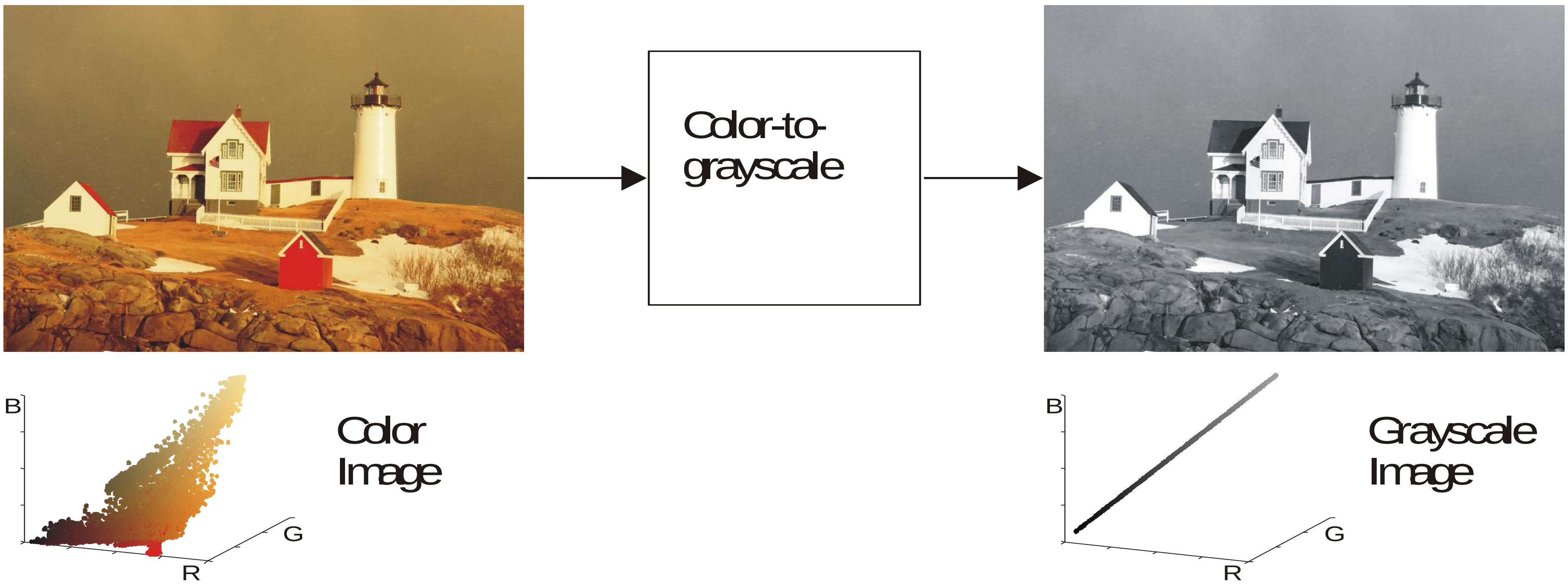
- ranking vs. rating
  - possible evaluation without reference

3.4 Overview – Evaluations of:

- Bottom-up vs. Top-down IQM
- STAR FR IQMs
- HDR Tone Mapping methods
- Color to Grayscale Conversions

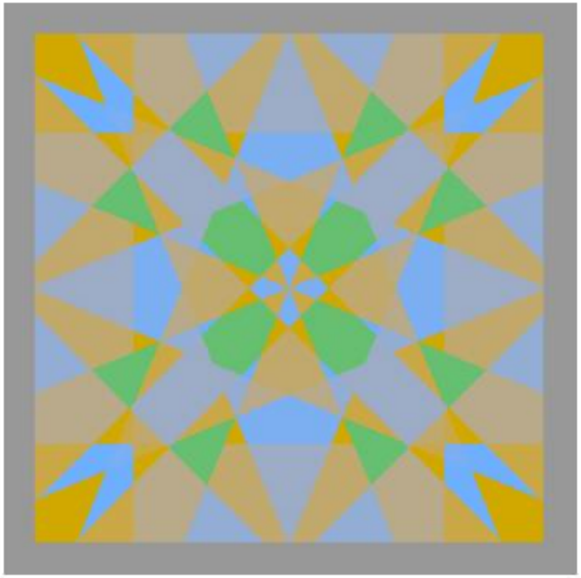
3.4 Evaluation of Color to Grayscale Conversions

- 3D data → 1D data



Color to Grayscale – Limit Case

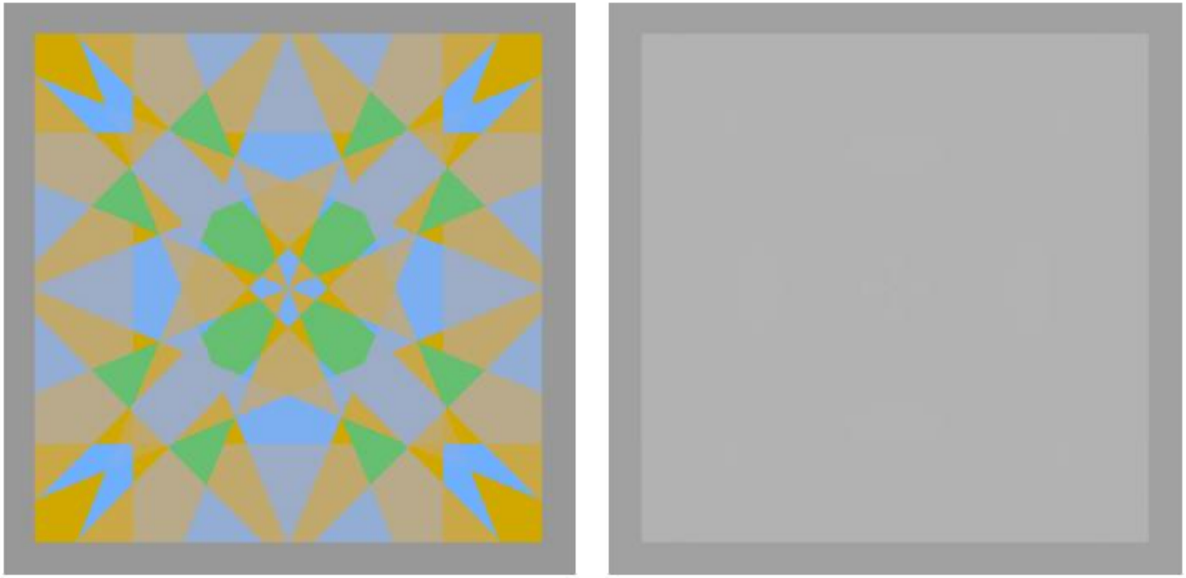
- Color image with constant luminance



Color to Grayscale – Limit Case

- Color image with constant luminance

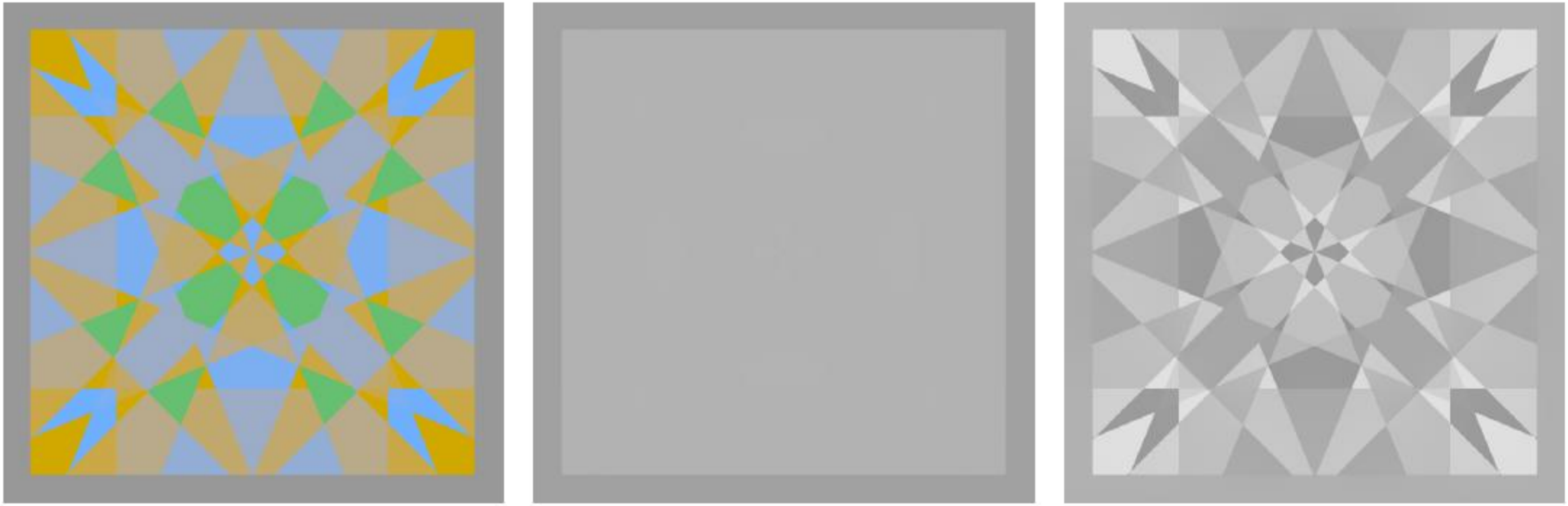
Widely used CIE-Y luminance conversion



Color to Grayscale – Limit Case

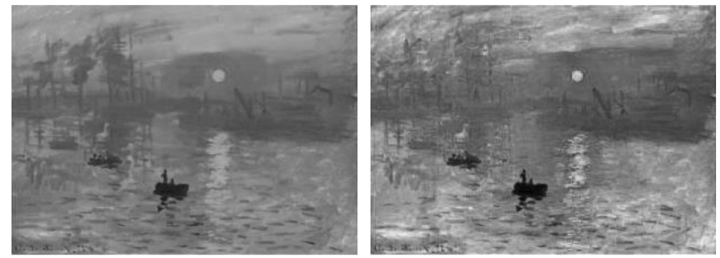
- Color image with constant luminance

[Neumann, Čadík 07]



C2G Experiments

- Accuracy
- Preference

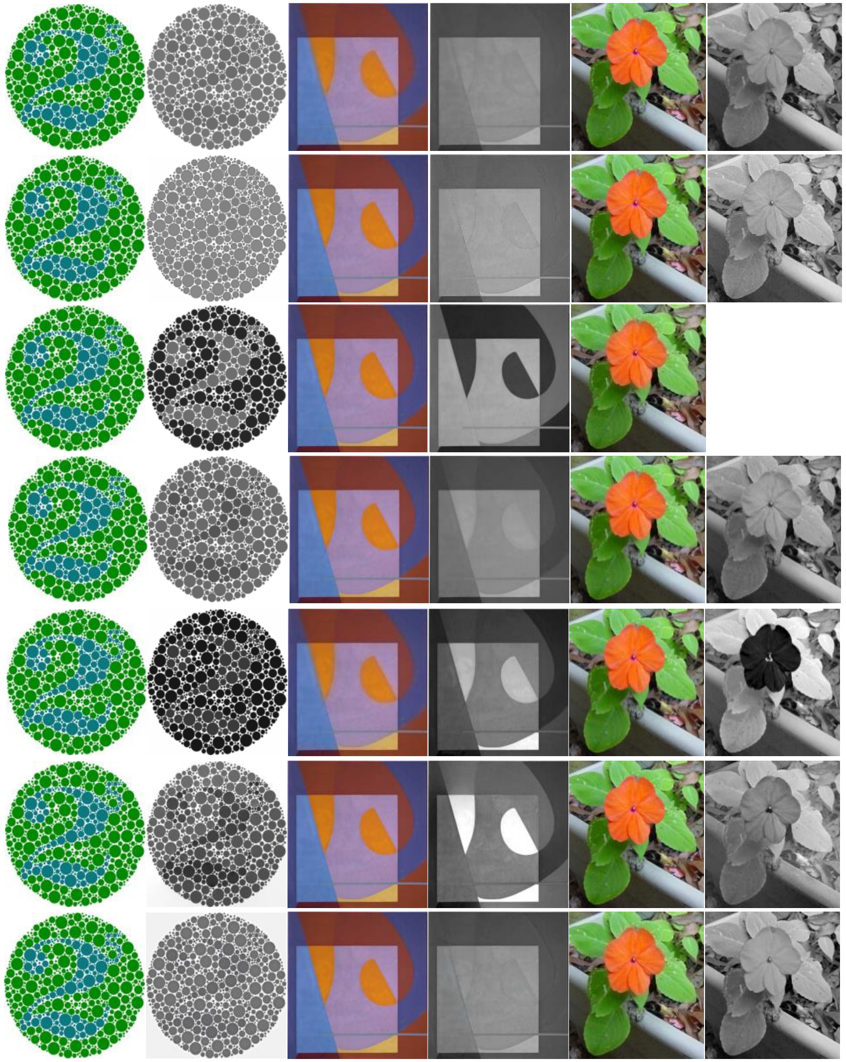


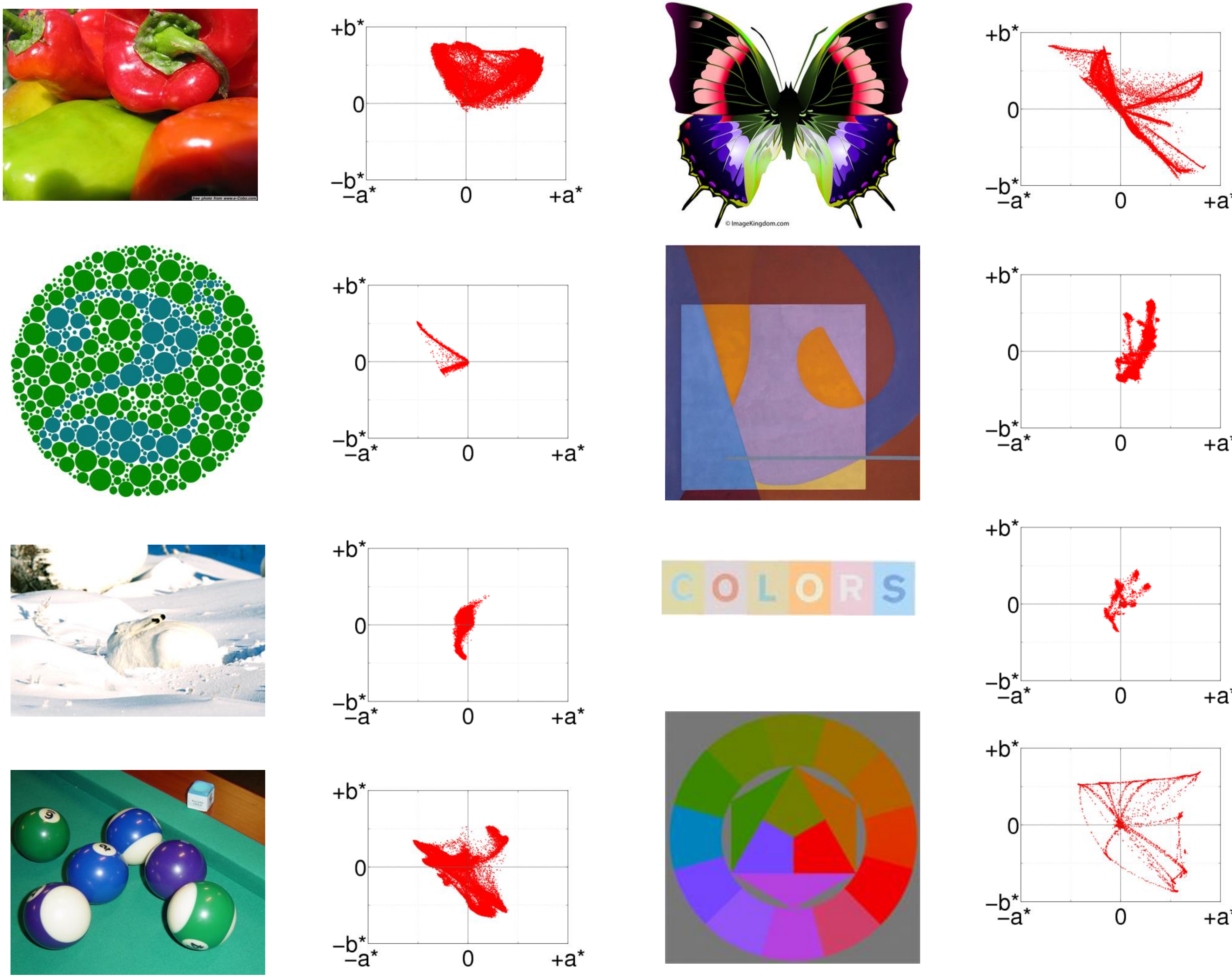
- 2AFC design
  - http://ranker.sourceforge.net
- 119 Participants
- 7 state-of-the art methods
  - default parameters to convert 24 input color images

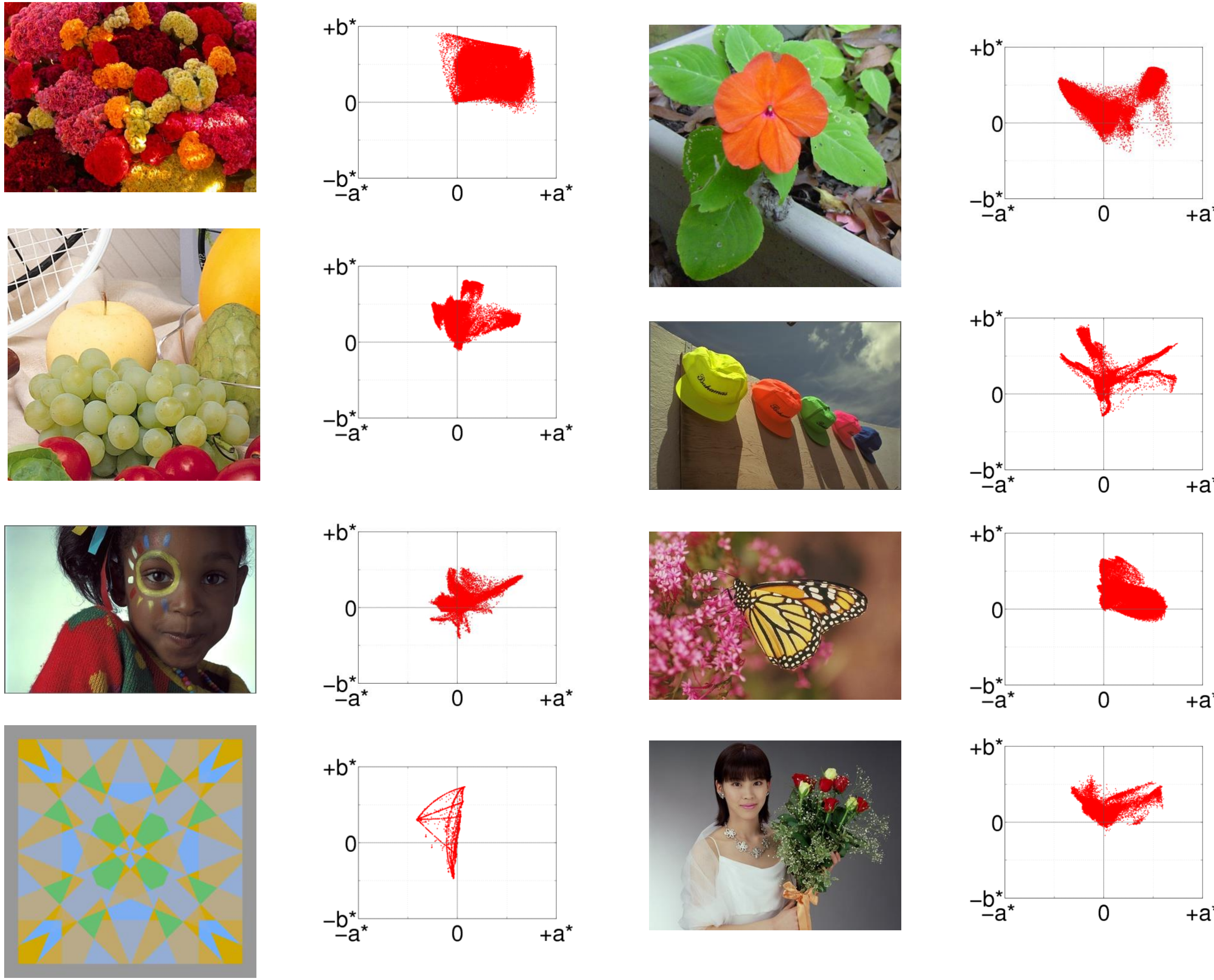
[Čadík 08]

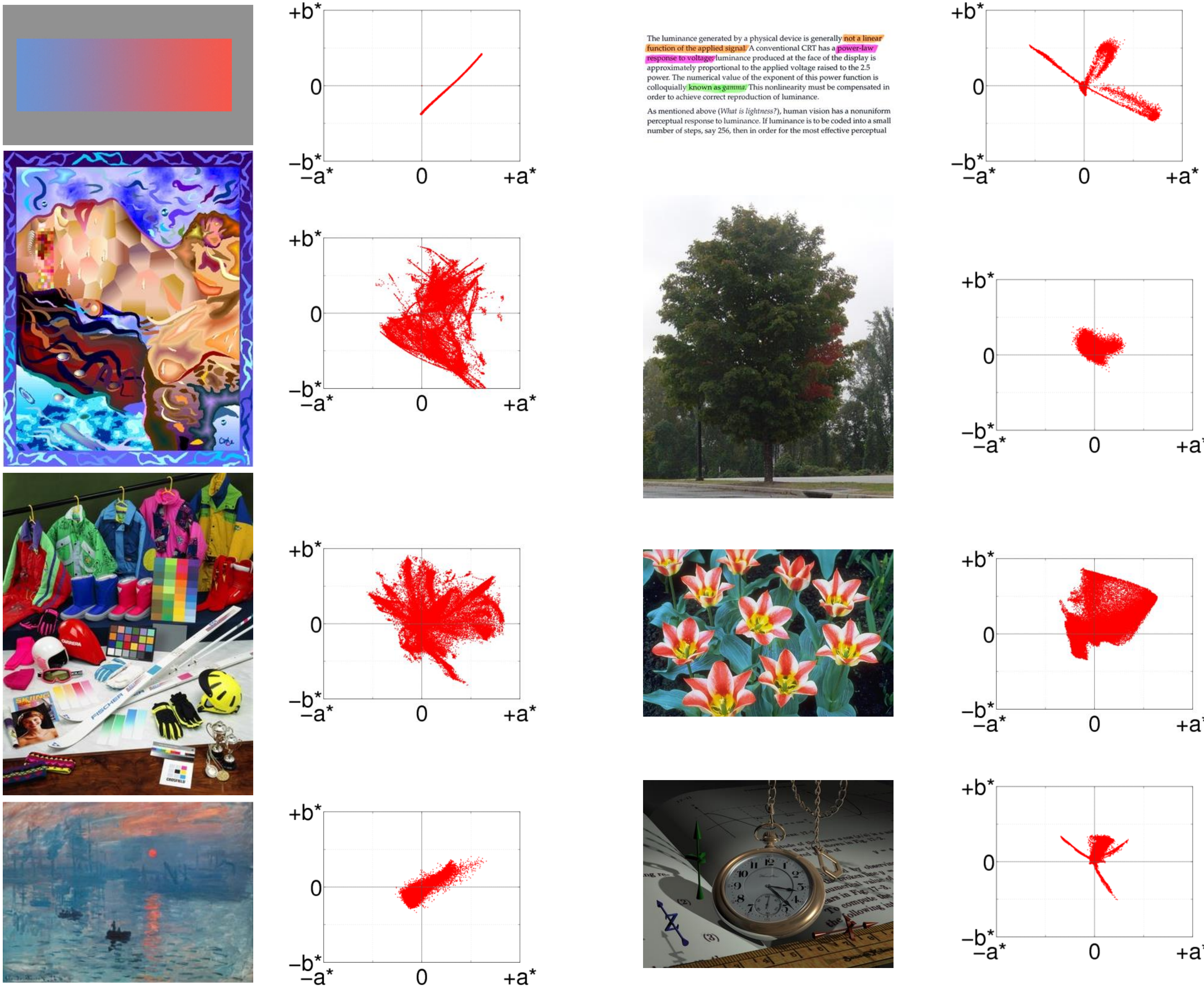
Evaluated Conversions

- CIE Y
  - Y channel of CIE XYZ model [1931]
- Bala04
  - [Bala & Eschbach 04]
- Decolorize
  - [Grundland & Dodgson 05]
- Color2Gray
  - [Gooch et al. 05]
- Rasche05
  - [Rasche et al. 05]
- Neumann07
  - [Neumann et al. 07]
- Smith08
  - [Smith et al. 08]




+b*
0
−b*
−a*
0
+a*
© ImageKingdom.com
COLORS


+b*
0
−b*
−a*
0
+a*


The luminance generated by a physical device is generally not a linear function of the applied signal. A conventional CRT has a power-law response to voltage: luminance produced at the face of the display is approximately proportional to the applied voltage raised to the 2.5 power. The numerical value of the exponent of this power function is colloquially known as *gamma*. This nonlinearity must be compensated in order to achieve correct reproduction of luminance.
As mentioned above (*What is lightness?*), human vision has a nonuniform perceptual response to luminance. If luminance is to be coded into a small number of steps, say 256, then in order for the most effective perceptual
+b*
0
−b*
−a*
0
+a*

Results

- Over 20 000 human responses collected ➔ Thurstone's Law of Comp. Judgments (case V) ➔

  z-scores (standard scores) ➔ statistics

- Multifactorial (n-way) ANOVA
  - Factors: *input images (24)*, *experiments (2)*, *conversions (7)*
  - Statistically significant main effect: *conversion* ➔ meaningful to proceed with the evaluation
  - Statistically significant interaction effects: *conversion x experiment*, *conversion x input image* ➔ meaningful to show results separately for each input image and each experiment

Results – Overall

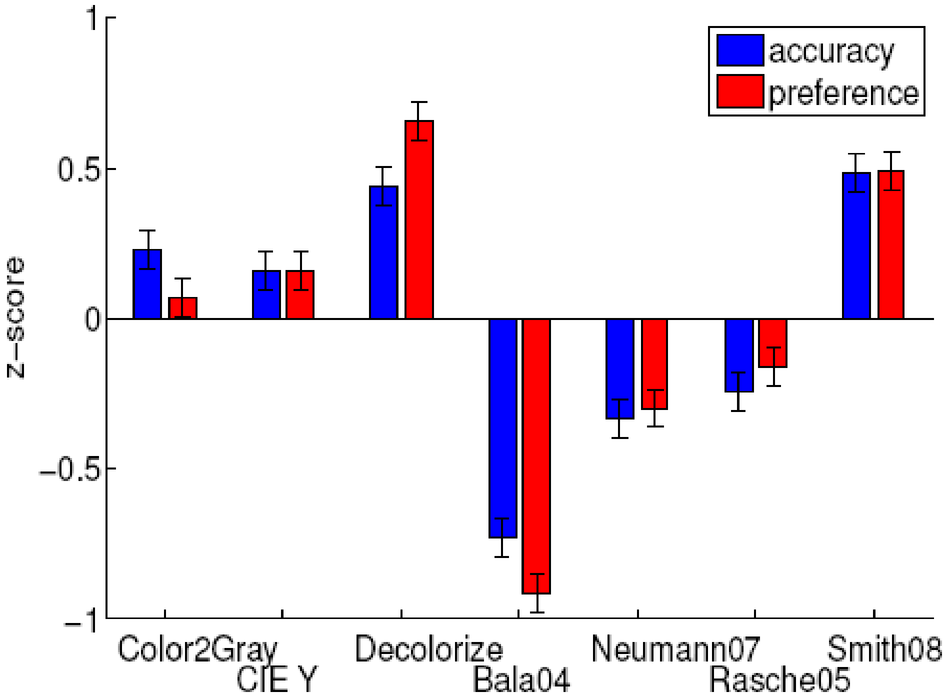
- Multiple comparison test [Tukey]
  - Overall ranking of conversions
  - Statistical significance of differences

Decolorize	Smith08	CIE Y	Color2Gray	Rasche05	Neumann07	Bala04
0.544	0.487	0.158	0.149	-0.203	-0.317	-0.819

Preference and Accuracy

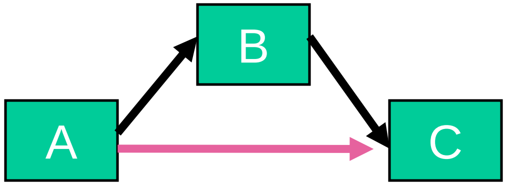
- Strong correlation between conversion **accuracy** and the grayscale image **preference**

  (r=0.97)
- PCA
  - 1st component: 96% of data variance
  - One dimension prevails
- CIE Y and Smith08 – consistent performance

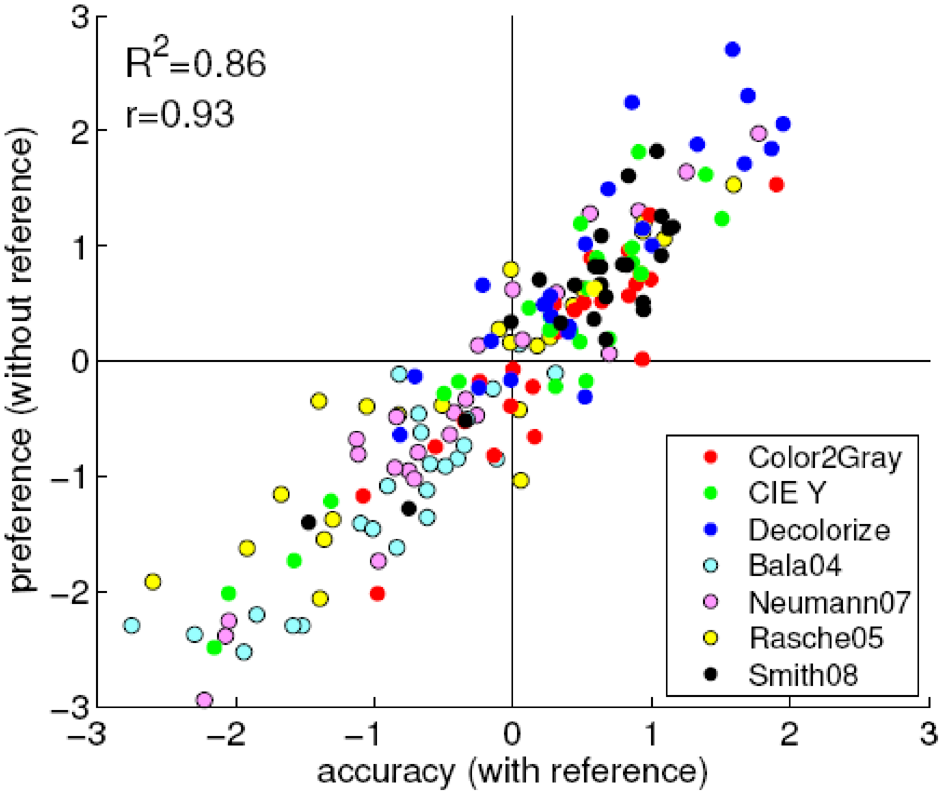


Individual Images

- z-scores independently for each image
- coef. of agreement
- coef. of consistency



- details tabulated in the paper



http://www.cgg.cvut.cz/~cadikm/color_to_gray_evaluation

Individual Images

- No conversion produces universally good results
- Each of inquired conversions ranked the worst for at least one input image
- Apart from Bala04, each conversion ranked the best for some input image

- Decolorize good for images with narrow gamuts
- Smith08 good for colorful images

➔ To improve robustness of current conversions over various inputs

Conclusions (1) – Methods to Use

- Images
  - HDR-HDR
    - **HDRVDP2** [Mantiuk et al. 11]
  - HDR-LDR
    - **HDRVDP2** [Mantiuk et al. 11]
    - **TMQI** [Yeganeh, Wang 13]
  - LDR-LDR
    - **LPLD** [Čadík et al. 13]
    - **SSIM** [Wang, Bovik 04]
- Videos
  - HDR-HDR
    - **DRIVQM** [Aydin et al. 10]
  - HDR-LDR
    - **DRIVQM** [Aydin et al. 10]
  - LDR-LDR
    - **PDM** [Winkler 05]

Try it out yourselves
http://metrics.mpi-inf.mpg.de/
http://resources.mpi-inf.mpg.de/hdr/metric/

Conclusions (2)

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

Interesting Directions in IQA

- Spectral image difference prediction
  - [Le Moan and Urban 14]
- Interestingness of images
  - interestingness = aesthetics, unusualness, general preferences [Gygli et al. 13]
- Pictorial quality of 3D models
  - [Váša et al.]
- Specific metrics
  - Visual popping [Schwarz, Stamminger 09]
  - Similarity measure for illustrations [Garces et al. 14]
  - Quality of image completion [Kopf et al. 14]
  - …

Thanks for your attention

- cadik@fit.vutbr.cz, http://cadik.posvete.cz/
- Many thanks to MPII Saarbrücken HDRI crowd