Realistic Image Synthesis

- Perception: Image Quality Metrics -

Philipp Slusallek
Karol Myszkowski
Gurprit Singh

Making Rendering Efficient

- **The solution of the global illumination problem is computationally hard**
- **New global illumination and rendering algorithms:**
  - deal well with the scene complexity, in terms of both storage and computation time requirements
  - are general and practical: reliable (fail-safe), user-friendly, automatic, easy to implement and to validate
  - take into account characteristics of the Human Visual System to concentrate the computation exclusively on the visible scene details

Outline

- **Questions of Appearance Preservation**
- **Basic characteristics of Human Visual System in image perception**
- **Daly's Visible Differences Predictor (VDP)**
- **Metric for rendering artifacts**
  - No-reference SVM-based metric
  - Full-reference CNN-based metric

Image Quality Metrics

- **Application examples which require metrics of the image quality as perceived by the human observer**
  - Lossy image compression and broadcasting
  - Design of image input/output devices
    - scanners, cameras, monitors, printers, and so on
  - Watermarking
  - Computer graphics, medical visualization

Questions of Appearance Preservation

- **The concern is not whether images are the same**
- **Rather the concern is whether images appear the same.**

How much computation is enough?

How much reduction is too much?

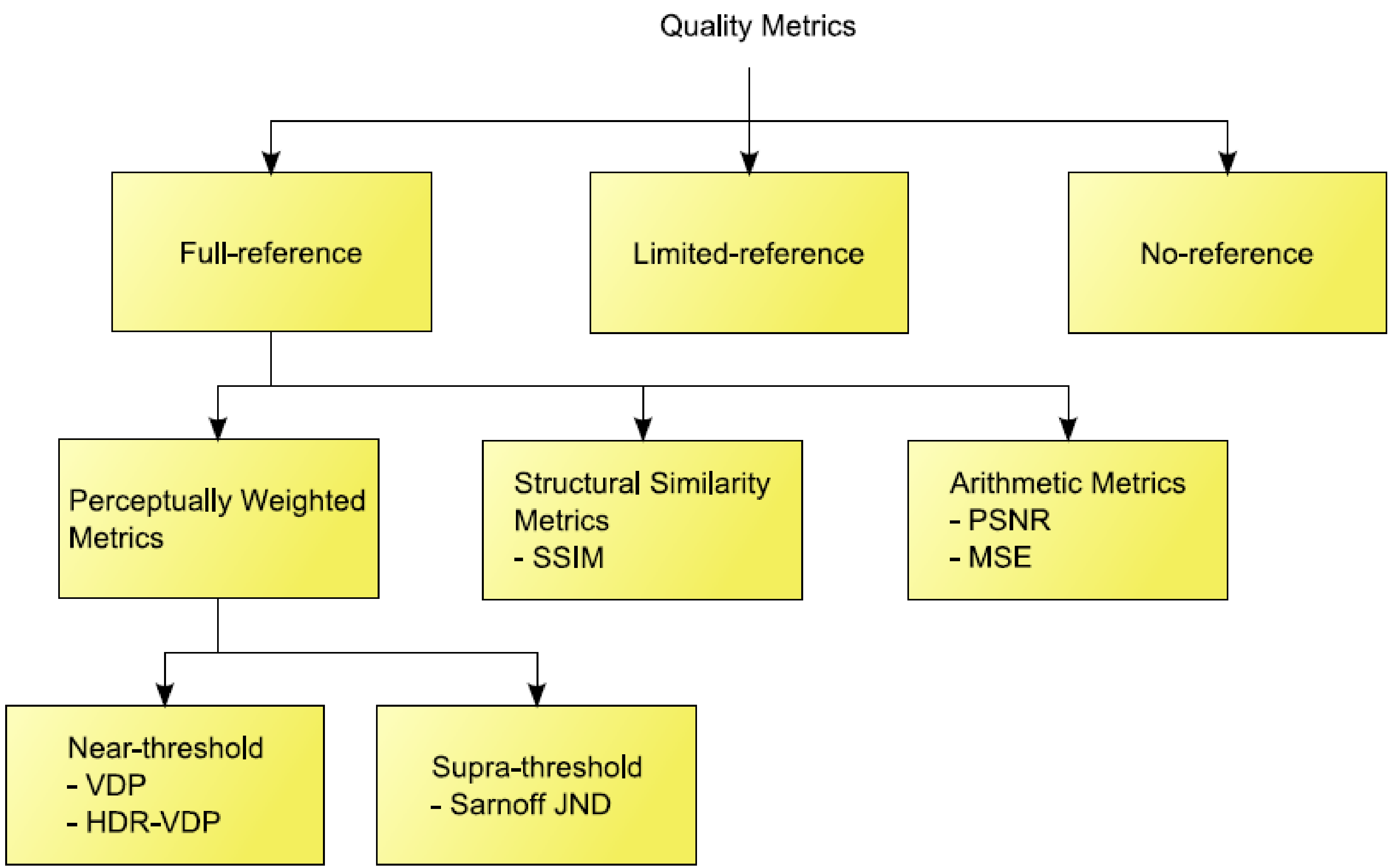
Subjective Methods

- **The best results can be obtained when human observers are involved**
  - Carefully controlled observation conditions
  - Representative number of participants
    - Averaging individual visual characteristics
    - Limiting the influence of emotional reactions
- **Very costly**
- **Limited use in practical routine applications**

Objective Methods

- **Usually rely on the comparison of images against the reference image**
  - – Measure perceivable differences between images, but an absolute measure of the image quality is difficult to obtain
  - – Not always in good agreement with the subjective measures
  - \+ Good repeatability of results
  - \+ Easy to use
  - \+ Low costs

Classification of Objective Quality Metrics



Classification of Objective Quality Metrics

- **Full-reference (FR)** where the reference image is available as it is typical in image compression, restoration, enhancement and reproduction applications.
- **Limited-reference (RR)** where a certain number of features characteristic for the image is extracted and made available as reference through a back-channel with reduced distortion. To avoid the back-channel transmission, known in advance and low magnitude signals, such that their visibility is prevented (as in watermarking), are directly encoded into an image and then the distortion of these signals is measured after the image transmission on the client side.
- **No-reference (NR)** which are focused mostly on detecting distortions which are application specific and predefined in advance such as blockiness (typical for DCT encoding in JPEG and MPEG), and ringing and blurring (typical for wavelet encoding in JPEG2000).

Full-reference Quality Metrics (1)

- **Pixel-based Metrics** with the mean square error (MSE) and the peak signal-to-noise ratio (PSNR) difference metrics as the prominent examples. In such a simple framework the HVS considerations are usually limited to the choice of a perceptually uniform color space such as CIELAB and CIELUV, which is used to represent the reference and distorted image pixels.
- **Structure-based Metrics** with the *Structural SIMilarity (SSIM) index* one of the most popular and influential quality metric in recent years. Since the HVS is strongly specialized in learning about the scenes through extracting structural information, it can be expected that the perceived image quality can be well approximated by measuring structural similarity between images.

Full-reference Quality Metrics (2)

- **Perception-based Fidelity Metrics** the *visible difference predictor* (VDP) and the *Sarnoff visual discrimination model* (VDM) as the prominent examples. These contrast-based metrics are based on advanced models of early vision in the HVS and are capable of capturing just visible (near threshold) differences or even measuring the magnitude of such (supra-threshold) differences and scale them in JND (just noticeable difference) units.

Pixel–based Metrics: Mean Square Error

$$RMSE = \sqrt{MSE} = \frac{1}{n}\sum_{i,j}(P_{ij} - Q_{ij})^2$$

$$PSNR = 20\log_{10}\frac{Pixel_{Max}}{MSE}$$

Reference image (*P*)

Compared images (*Q*)

Pixel–based Metrics: Mean Square Error

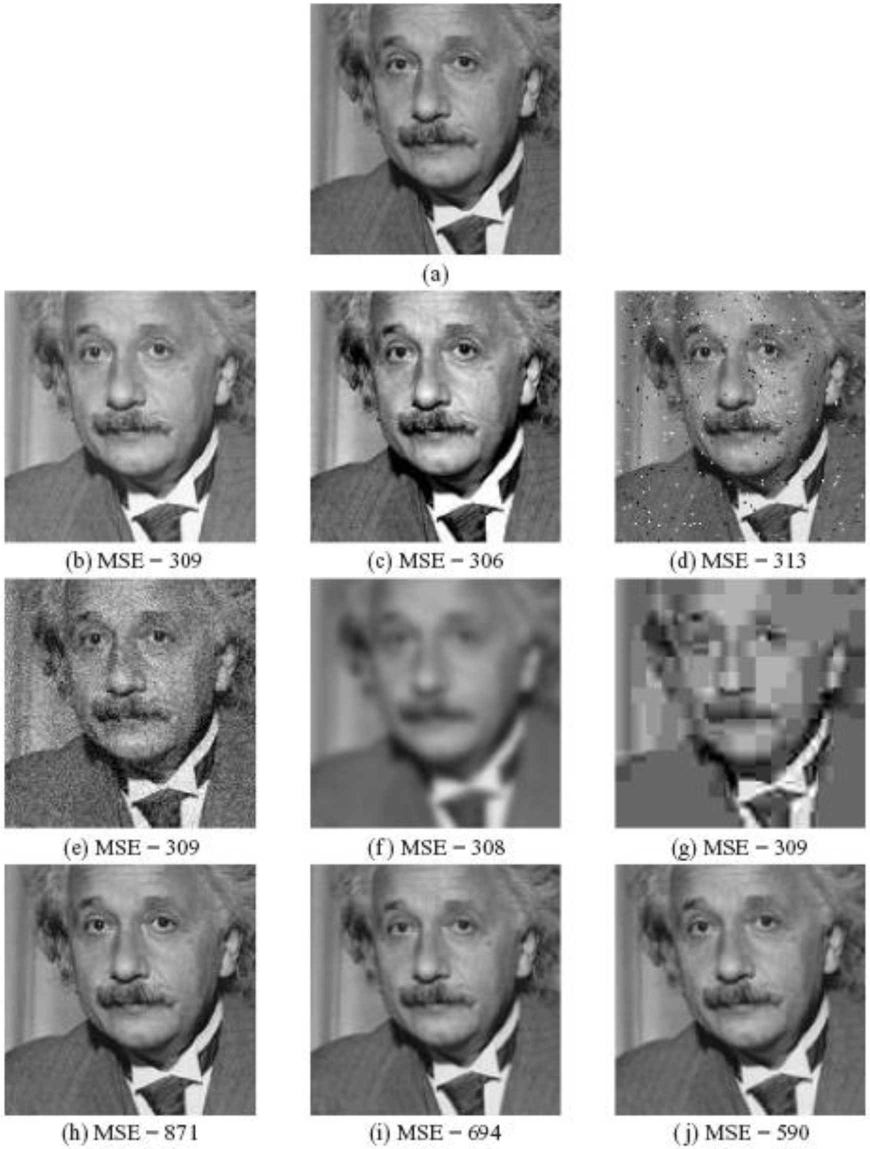
$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \frac{1}{n}\sum_{i,j}(P_{ij} - Q_{ij})^2$$

$$\mathrm{PSNR} = 20\log_{10}\frac{\mathrm{Pixel}_{Max}}{\mathrm{MSE}}$$

Reference image (*P*)

Compared images (*Q*)

Pixel–based Metrics: Mean Square Error



Einstein image altered with different types of distortions:
(a) “original image”;
(b) mean luminance shift;
(c) a contrast stretch;
(d) impulsive noise contamination;
(e) white Gaussian noise contamination;
(f) blurring;
(g) JPEG compression;
(h) a spatial shift (to the left);
(i) spatial scaling (zooming out);
(j) a rotation.

Note that images (b)–(g) have almost the same MSE values but drastically different visual quality. Also, note that the MSE is highly sensitive to spatial translation, scaling, and rotation [Images (h)–(j)].

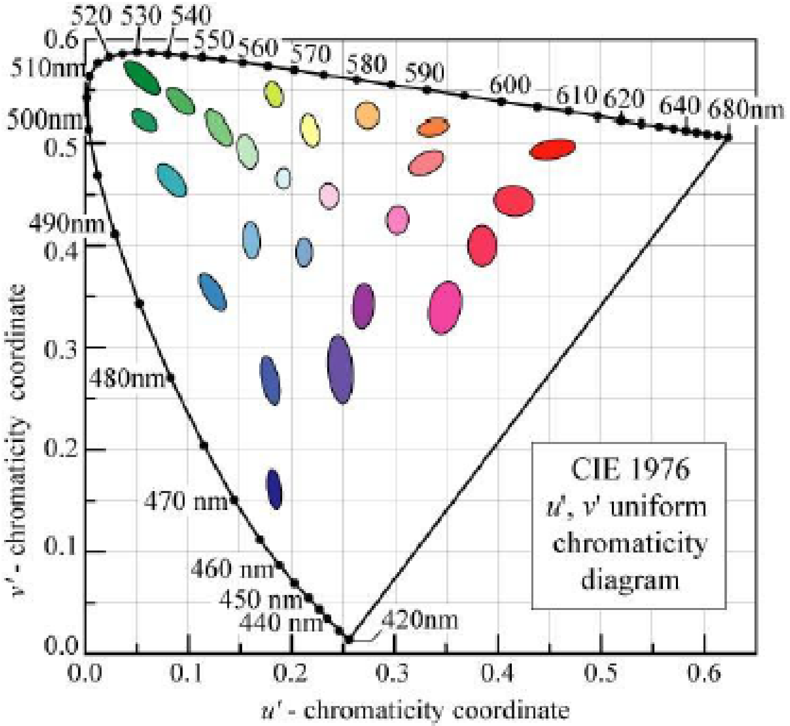
Wang & Bovik

Color Appearance Spaces

- CIE 1976 L*u*v* and L*a*b*
  - Color (X, Y, Z) reflected by a surface under known illuminant (X_n, Y_n, Z_n) ("white point")
  - $f(r) = \begin{cases} r^{1/3} & \text{if } r > 0.008856 \\ 7.787r + 16/116 & \text{otherwise} \end{cases}$ (log-like)
  - $L^* = 116\, f(Y/Y_n) - 16$
  - $u' = 4X \,/\, (X+15Y+3Z)$
    $v' = 9Y \,/\, (X+15Y+3Z)$
  - $u^* = 13\, L^*\, (u'-u_n')$
    $v^* = 13\, L^*\, (v'-v_n')$
  - $a^* = 500\, [f(X/X_n) - f(Y/Y_n)]$
    $b^* = 200\, [f(Y/Y_n) - f(Z/Z_n)]$
  - Euclidean distances ΔE^*_{uv} and ΔE^*_{ab}

Color Appearance Spaces

- u'v' chromaticity diagram
  - Deformed ellipses
- CIELUV and CIELAB
  - Close to uniform
  - Useful for practical color differences
  - Not perfect



Full-reference Quality Metrics

- **Structure-based Metrics** with the *Structural SIMilarity (SSIM) index* one of the most popular and influential quality metric in recent years.
- Since the HVS is strongly specialized in learning about the scenes through extracting structural information, it can be expected that the perceived image quality can be well approximated by measuring structural similarity between images.

Structural SIMilarity (SSIM) index

- The SSIM index decomposes similarity estimation into three independent comparison functions: **luminance**, **contrast**, and **structure**.
- The **luminance** comparison function l(x, y) for an image pair x and y is specified as:

$$l(x,y) = l(\mu_x, \mu_y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad \text{where} \quad \mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$$

- The **contrast** comparison function c(x, y) is specified as:

$$c(x,y) = c(\sigma_x, \sigma_y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad \text{where} \quad \sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2}$$

- The **structure** comparison function s(x, y) is specified as:

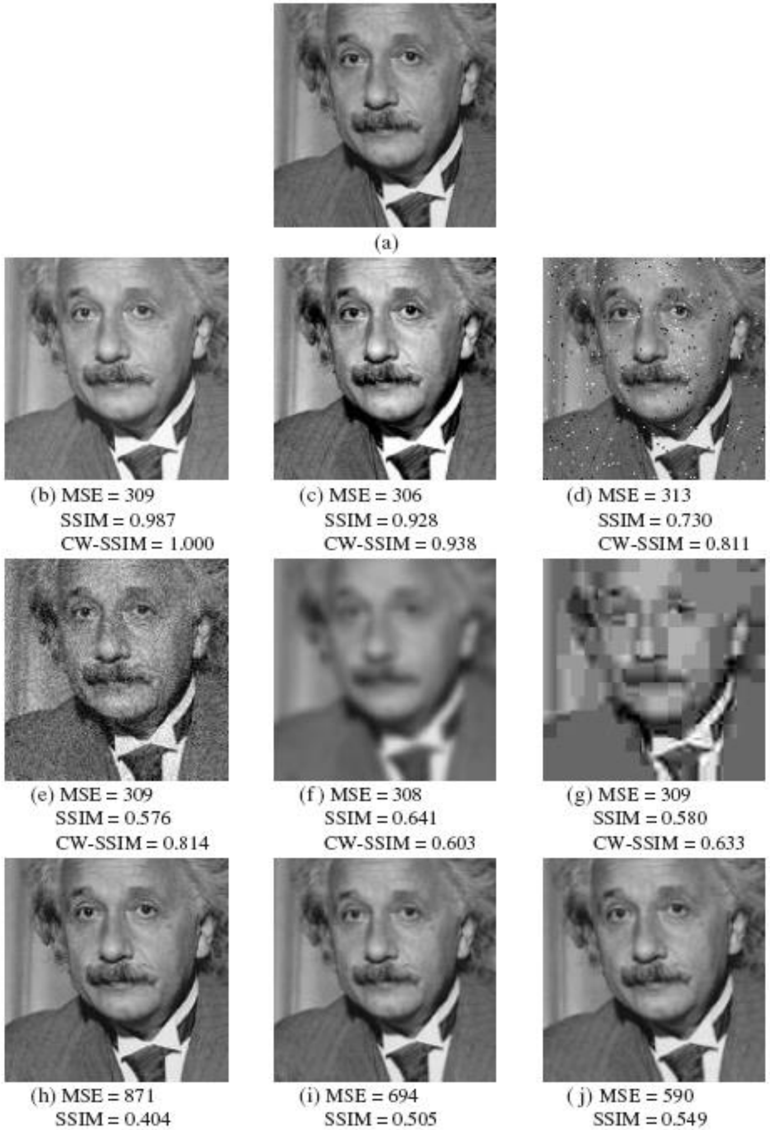
$$s(x,y) = s\left(\frac{x-\mu_x}{\sigma_x}, \frac{y-\mu_y}{\sigma_y}\right) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad \text{where} \quad \sigma_{xy} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}$$

- The three comparison functions are combined in the SSIM index:

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma}$$

- To obtain a local measure of structure similarity all statistics μ, σ are computed within a local 8 × 8 window which slides over the whole image.

Structural SIMilarity (SSIM) index



Einstein image altered with different types of distortions:
(a) “original image”;
(b) mean luminance shift;
(c) a contrast stretch;
(d) impulsive noise contamination;
(e) white Gaussian noise contamination;
(f) blurring;
(g) JPEG compression;
(h) a spatial shift (to the left);
(i) spatial scaling (zooming out);
(j) a rotation.

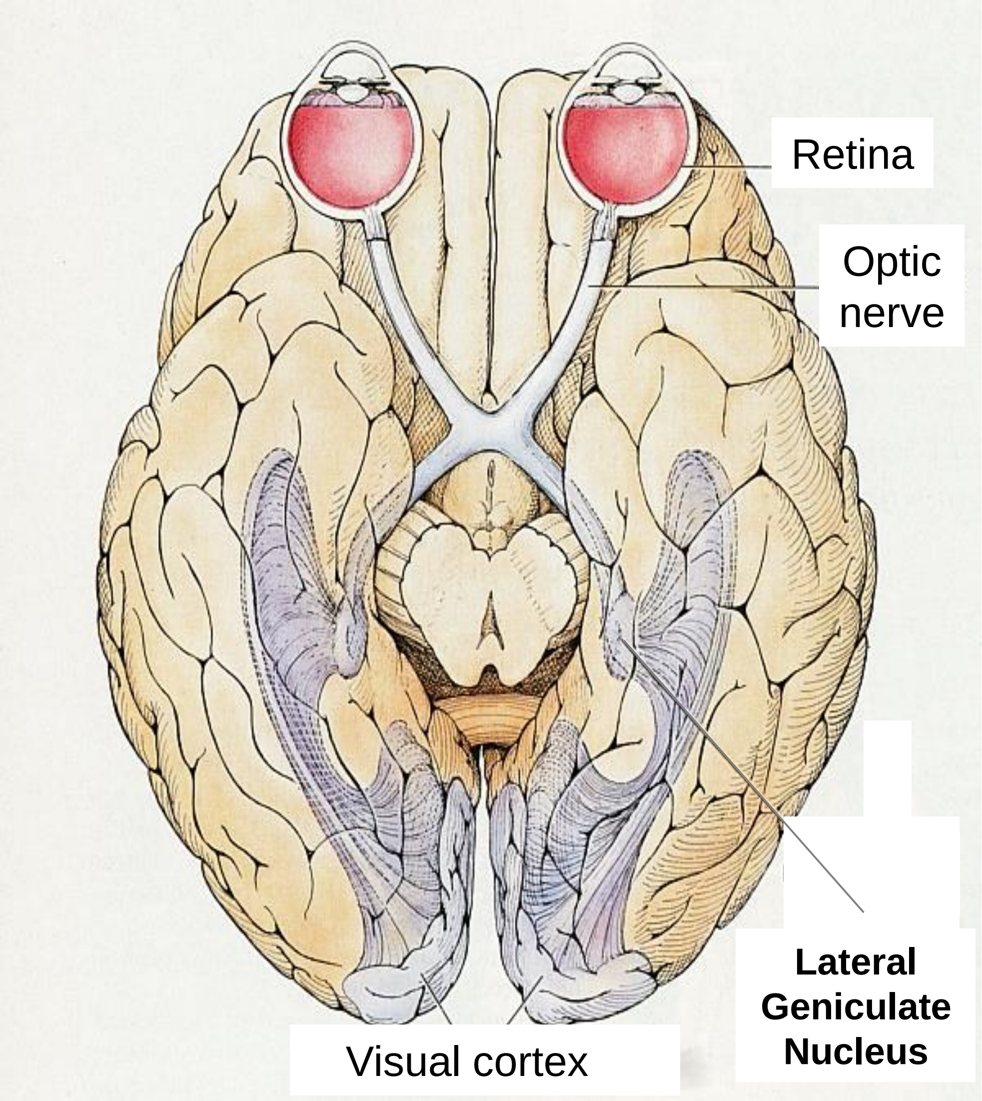
Images (b)–(g) drastically different visual quality and SSIM captures well such quality degradation. Also, note that the SSIM is highly sensitive to spatial translation, scaling, and rotation [Images (h)–(j)].

Human Visual System (HVS) vs. Image Quality Metrics

- **Anatomy and physiology of visual pathway determine its sensitivity on various image elements.**
- **Basic HVS characteristics must be taken into account to estimate perceivable differences between images.**
- **Complete model of image perception has not been elaborated so far.**

Visual Pathway

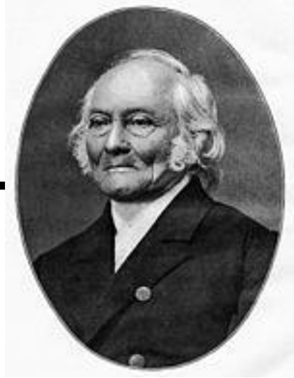
- Functionality of visual pathway from retina to the visual cortex are relatively well understood.
- Modeling on the physiological level too complex.
- Behavioral models acquired through psychophysical experiments are easy to use.



Important Characteristics of the HVS

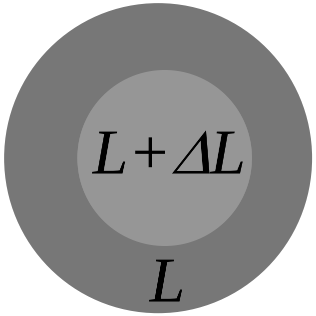
- **Visual adaptation**
- **Temporal and spatial mechanisms (channels) which are used to represent the visual information at various scales and orientations as it is believed that primary visual cortex does.**
- **Contrast Sensitivity Function which specifies the detection threshold for a stimulus as a function of its spatial and temporal frequencies.**
- **Visual masking affecting the detection threshold of a stimulus as a function of the interfering background stimulus which is closely coupled in space and time.**

Visual Adaptation



Ernst Heinrich Weber
[From wikipedia]

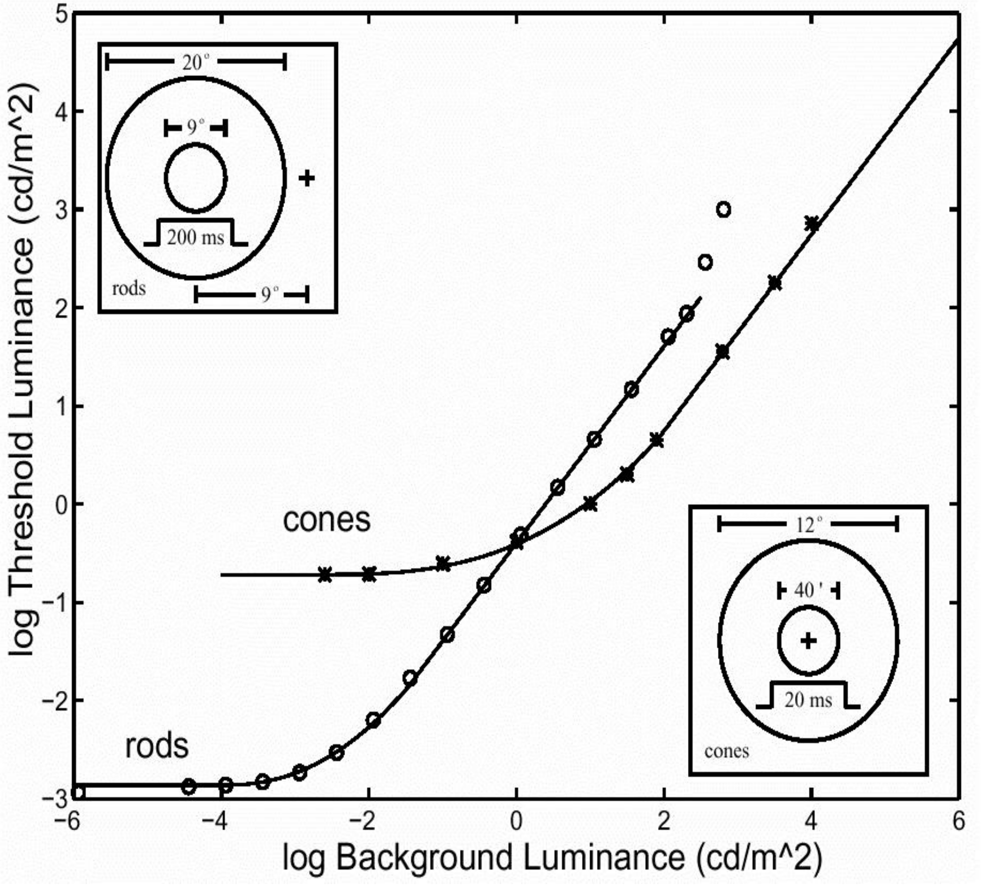
- Adaptation of visual system to various levels of background luminance



- Weber's law:

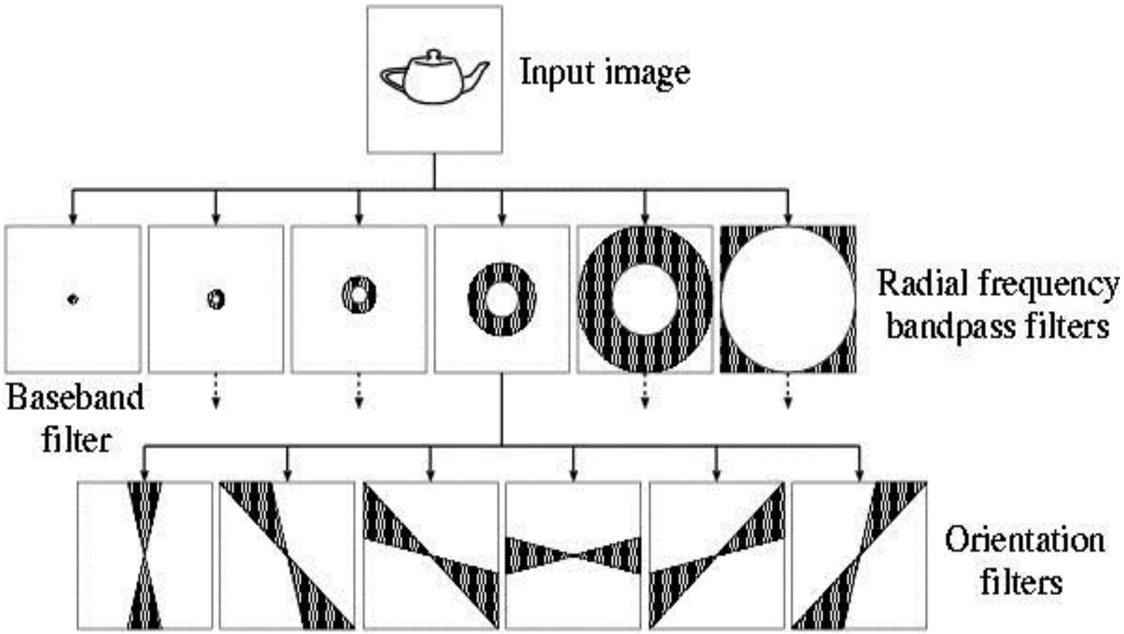
$$\frac{\Delta L}{L} = const$$

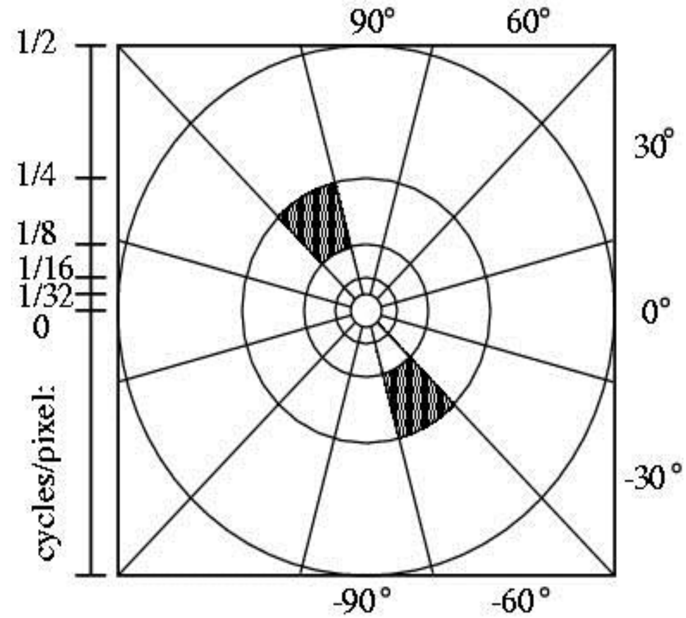
TVI – Threshold *versus* Intensity function

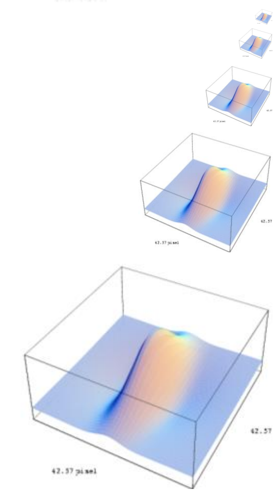


Ferwerda et al.

Cortex Transform: Filter Bank

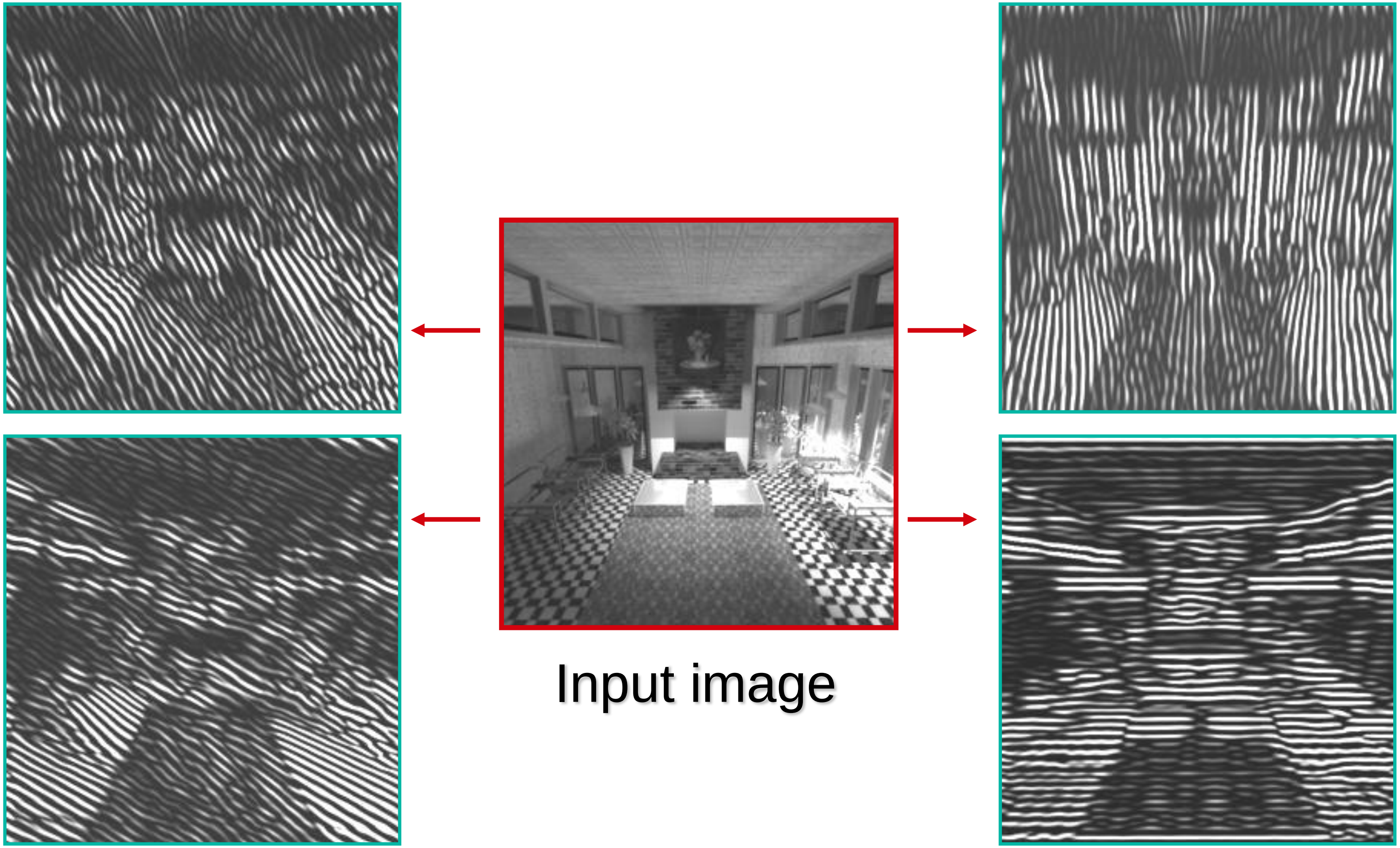






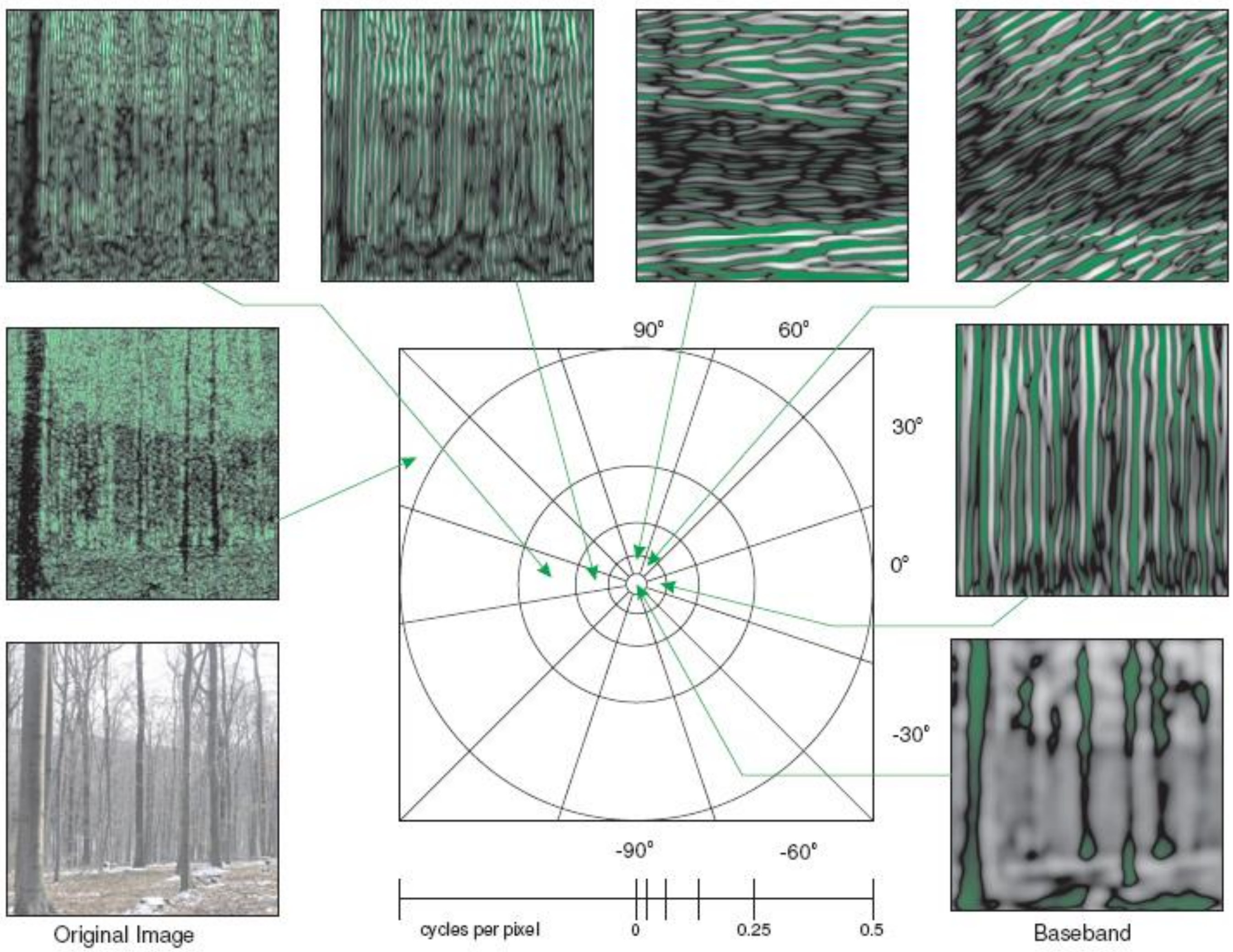
Filter bank examples: Gabor functions (Marcelja80), steerable pyramid   transform (Simoncelli92), Discrete Cosine Transform (DCT), difference of Gaussians (Laplacian) pyramids (Burt83, Wilson91), Cortex transform (Watson87, Daly93).

Cortex Transform: Orientation Bands

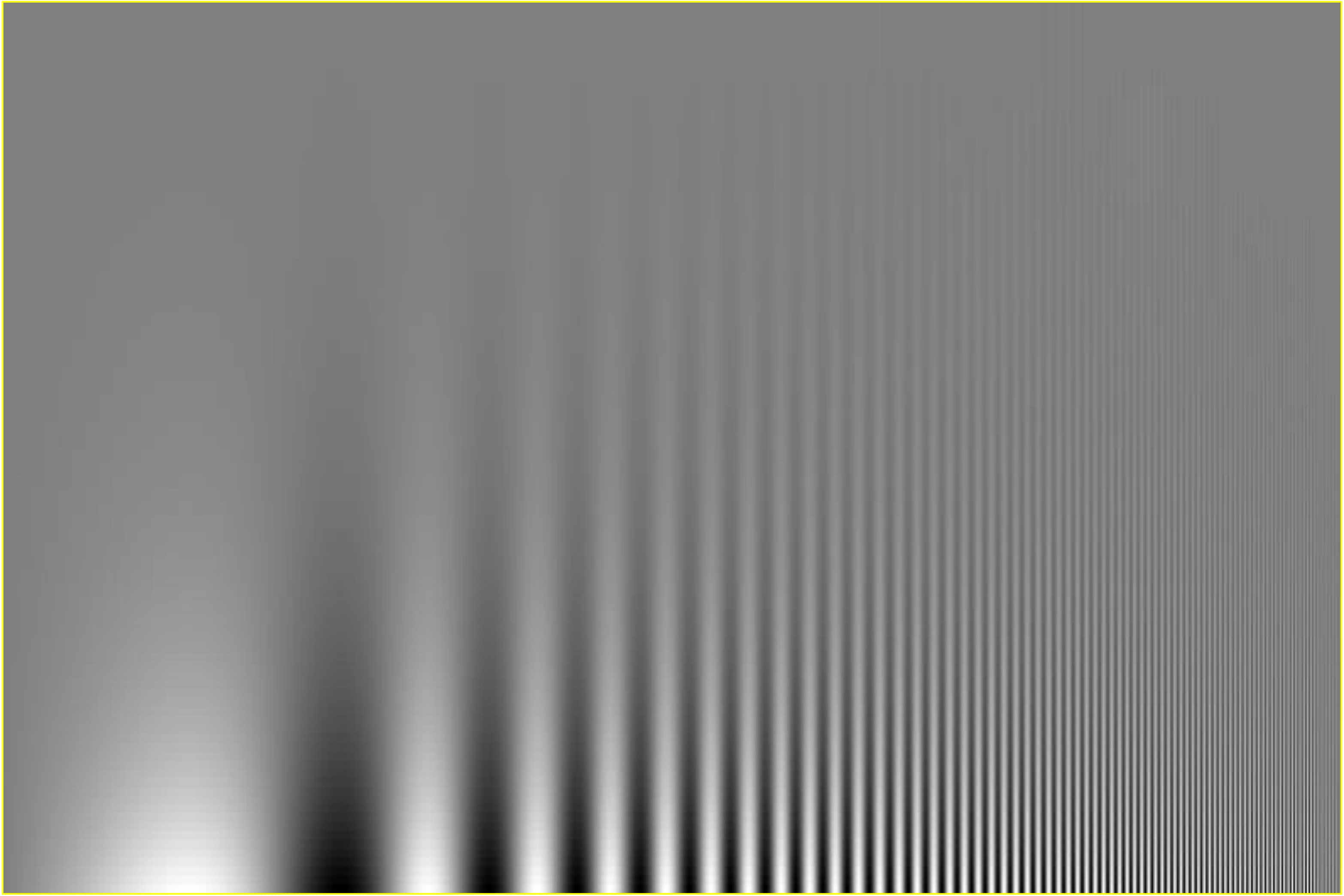


Input image

Cortex Transform: Frequency and Orientation Bands

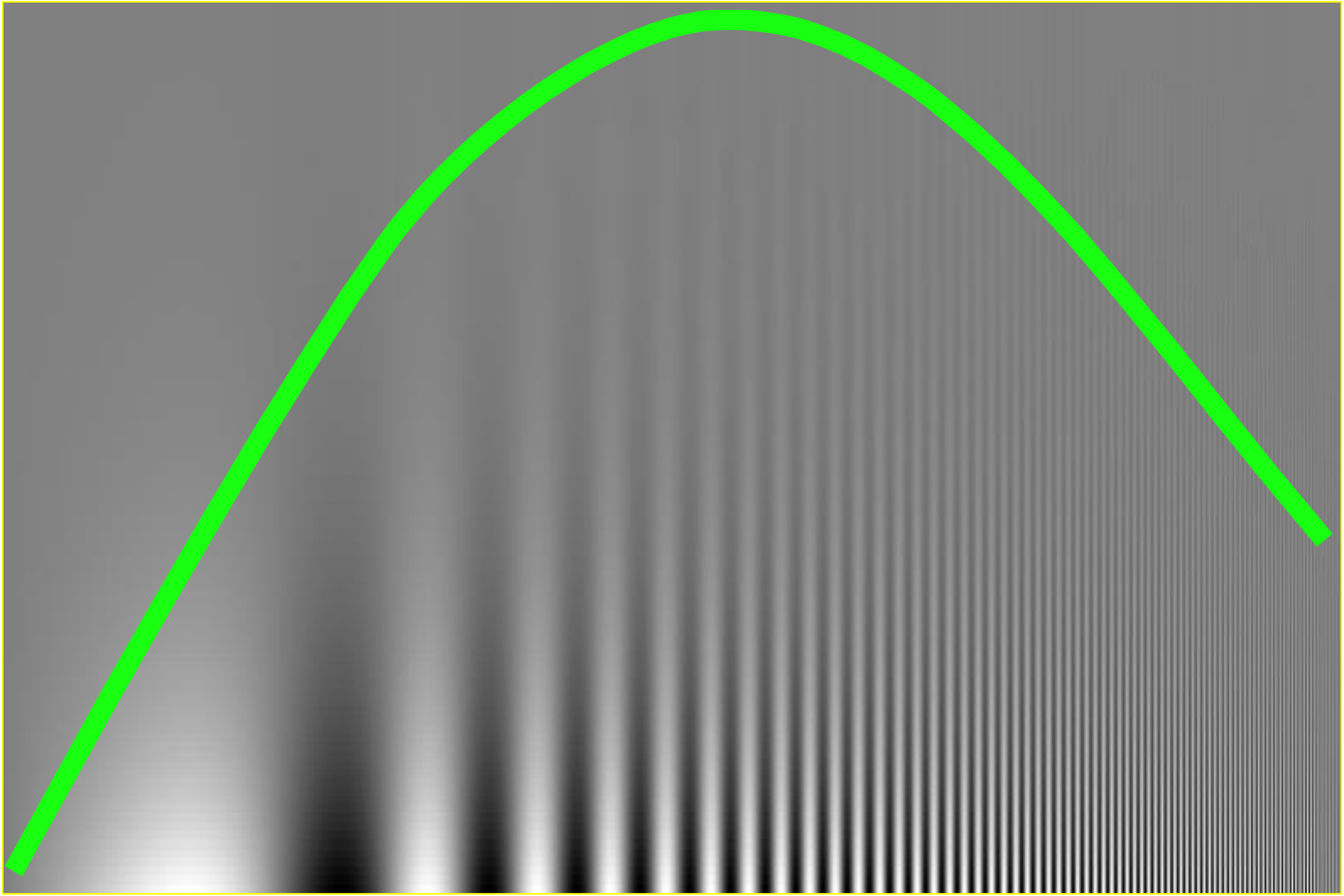


Contrast Sensitivity Function



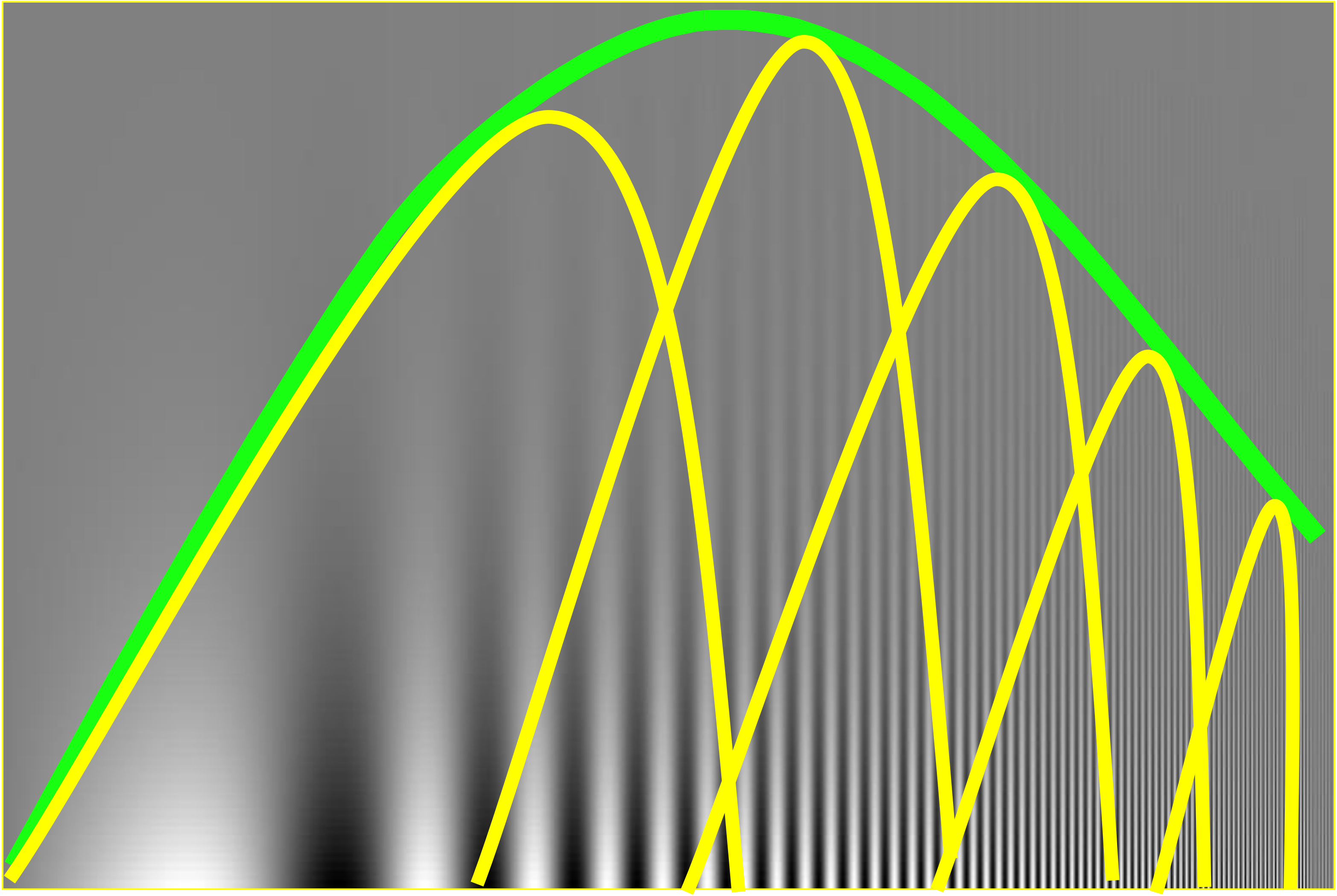
J.G. Robson CSF chart

Contrast Sensitivity Function



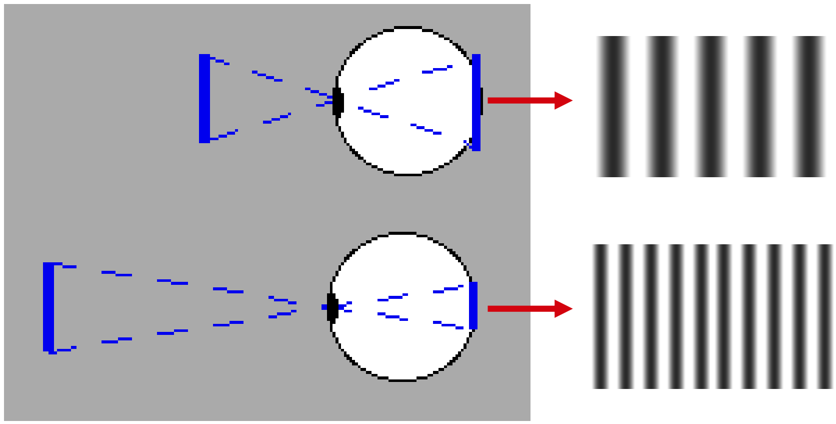
J.G. Robson CSF chart

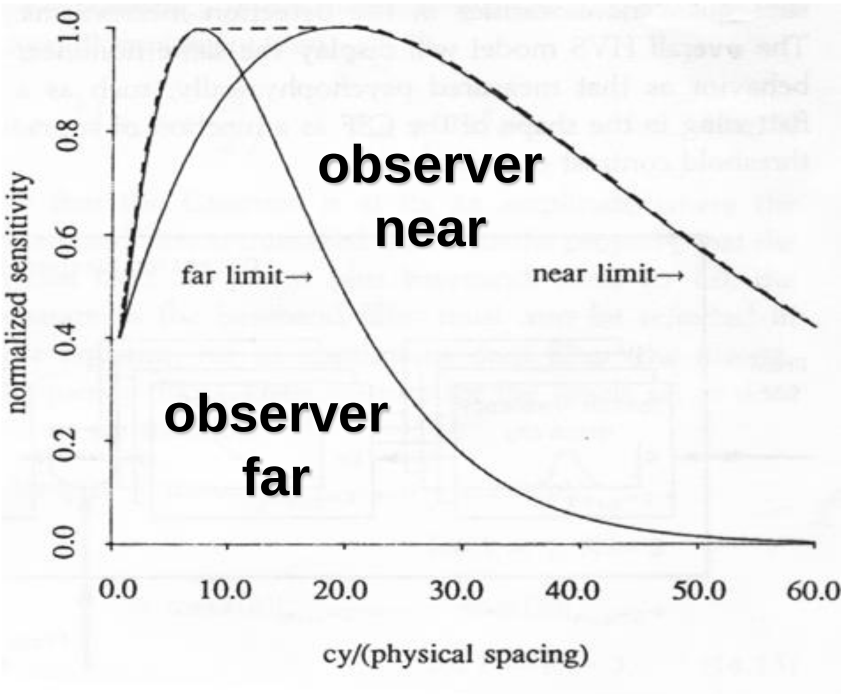
Contrast Sensitivity Function (CSF)



J.G. Robson CSF chart

CSF *versus* Observation Distance

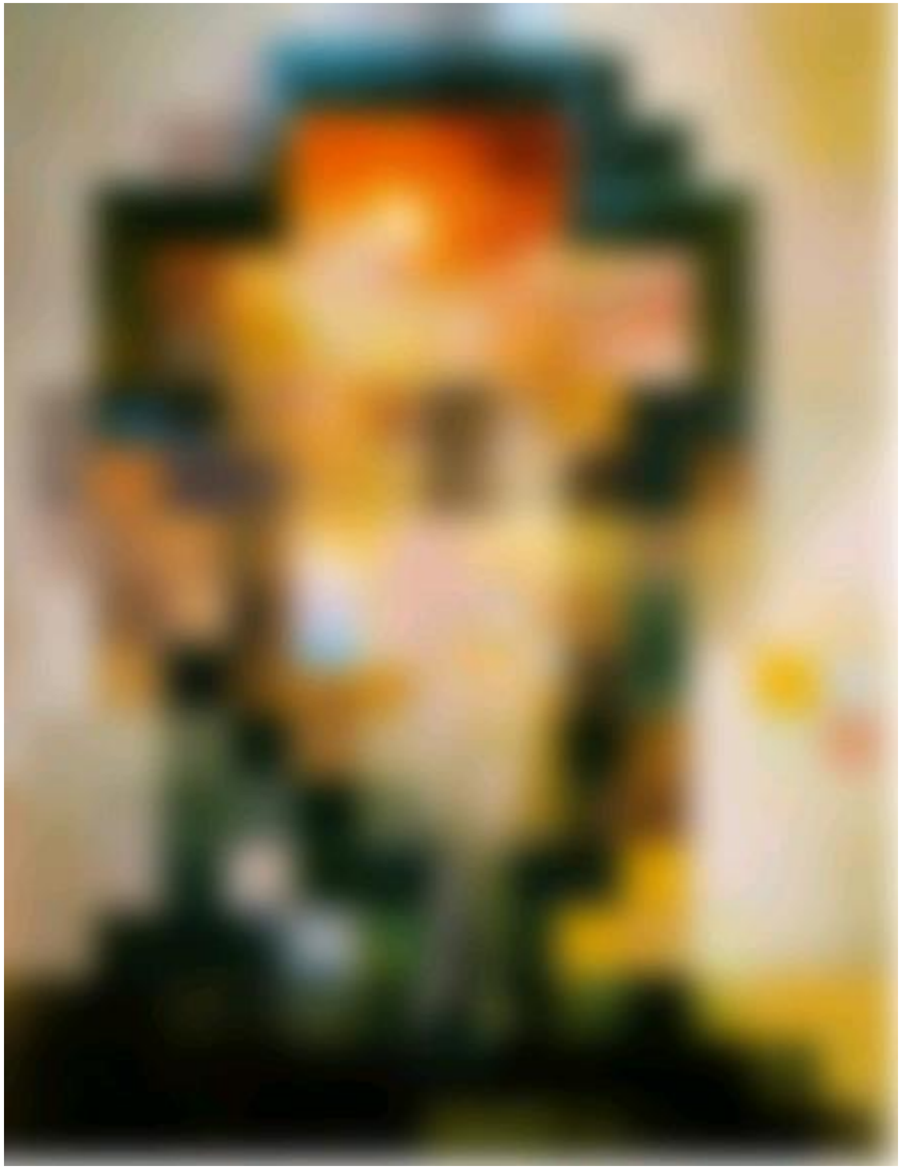




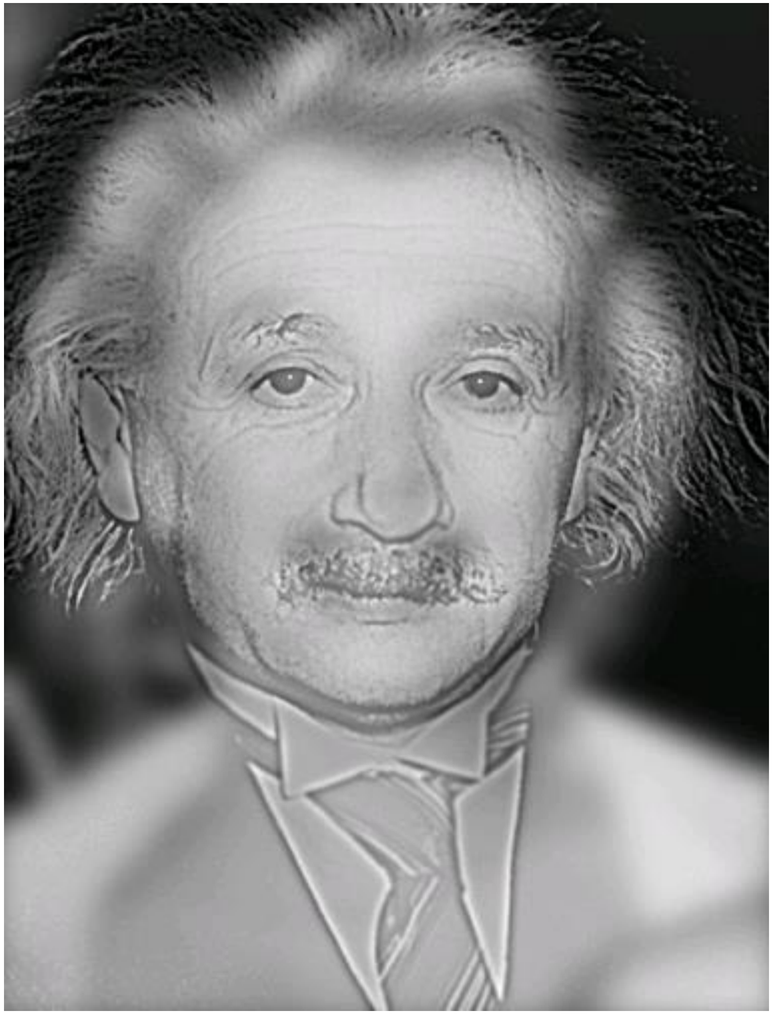
- **Spatial frequencies projected on the retina increase proportionally to the observation distance.**
- **Image elements represented by low (high) spatial frequencies might become visible (invisible) with the increase of the observation distance.**

To estimate conservatively the image quality for variable observer positions the envelope of CSFs for the extreme observer locations can be used.

Lincoln illusion



Hybrid Images



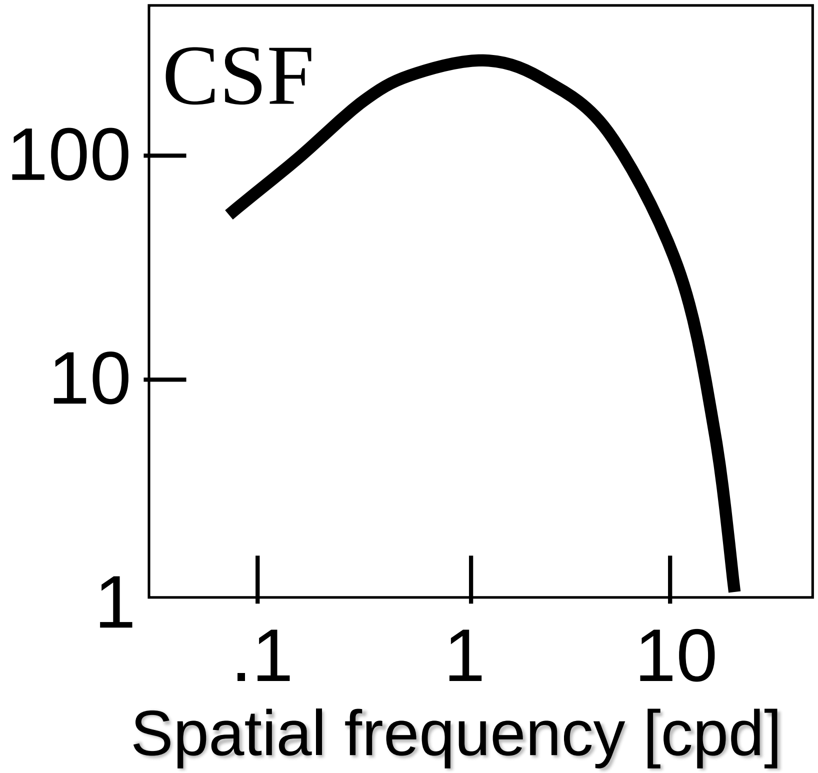
Hybrid Images

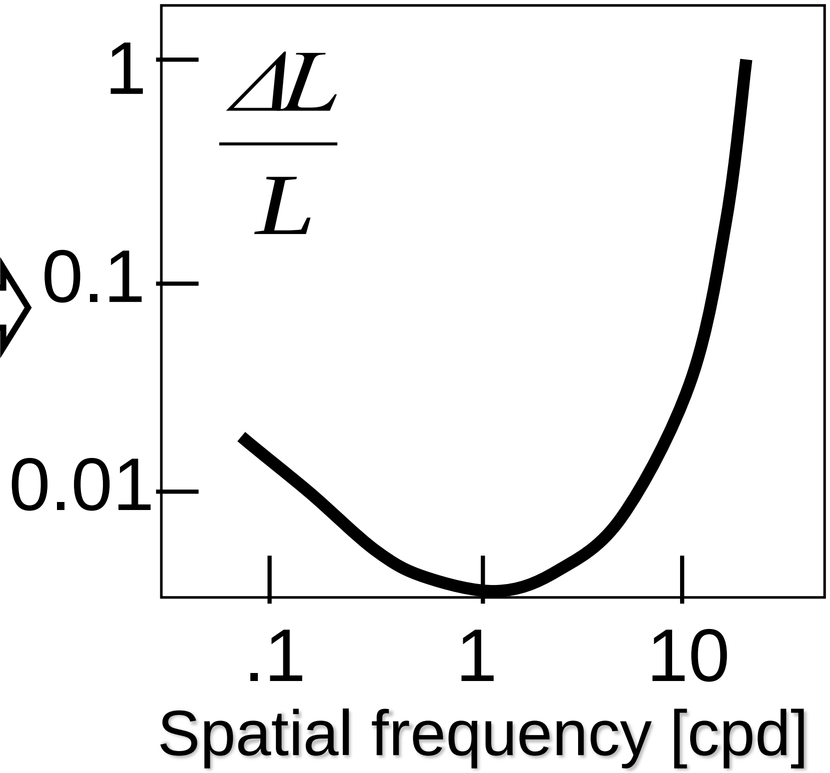
Hybrid Images

Contrast Senstivity ***versus***

Detection Threshold ***ΔL/L***

$$\mathrm{CSF} = \frac{L}{\Delta L} \Rightarrow \frac{\Delta L}{L} = \frac{1}{\mathrm{CSF}}$$

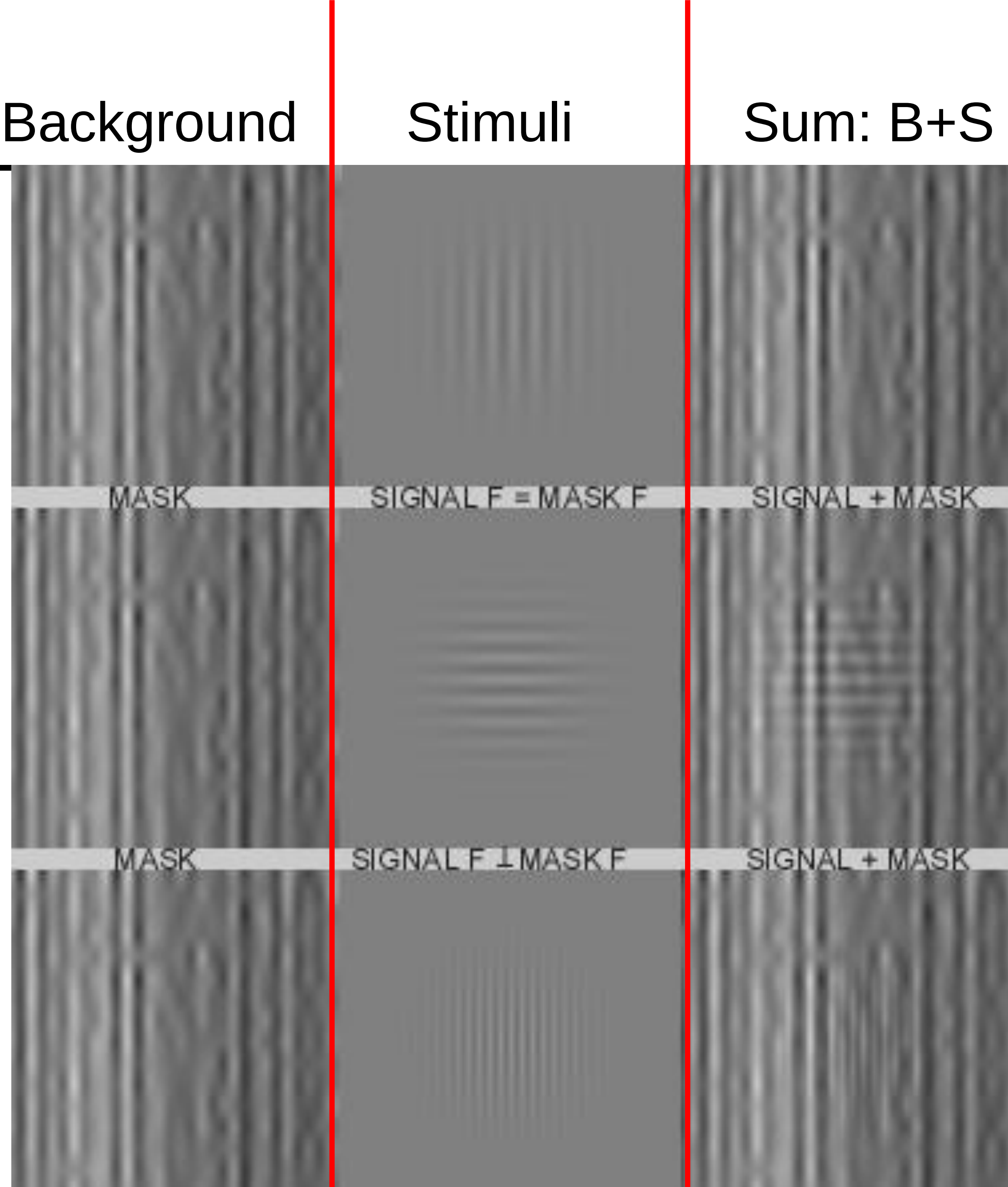




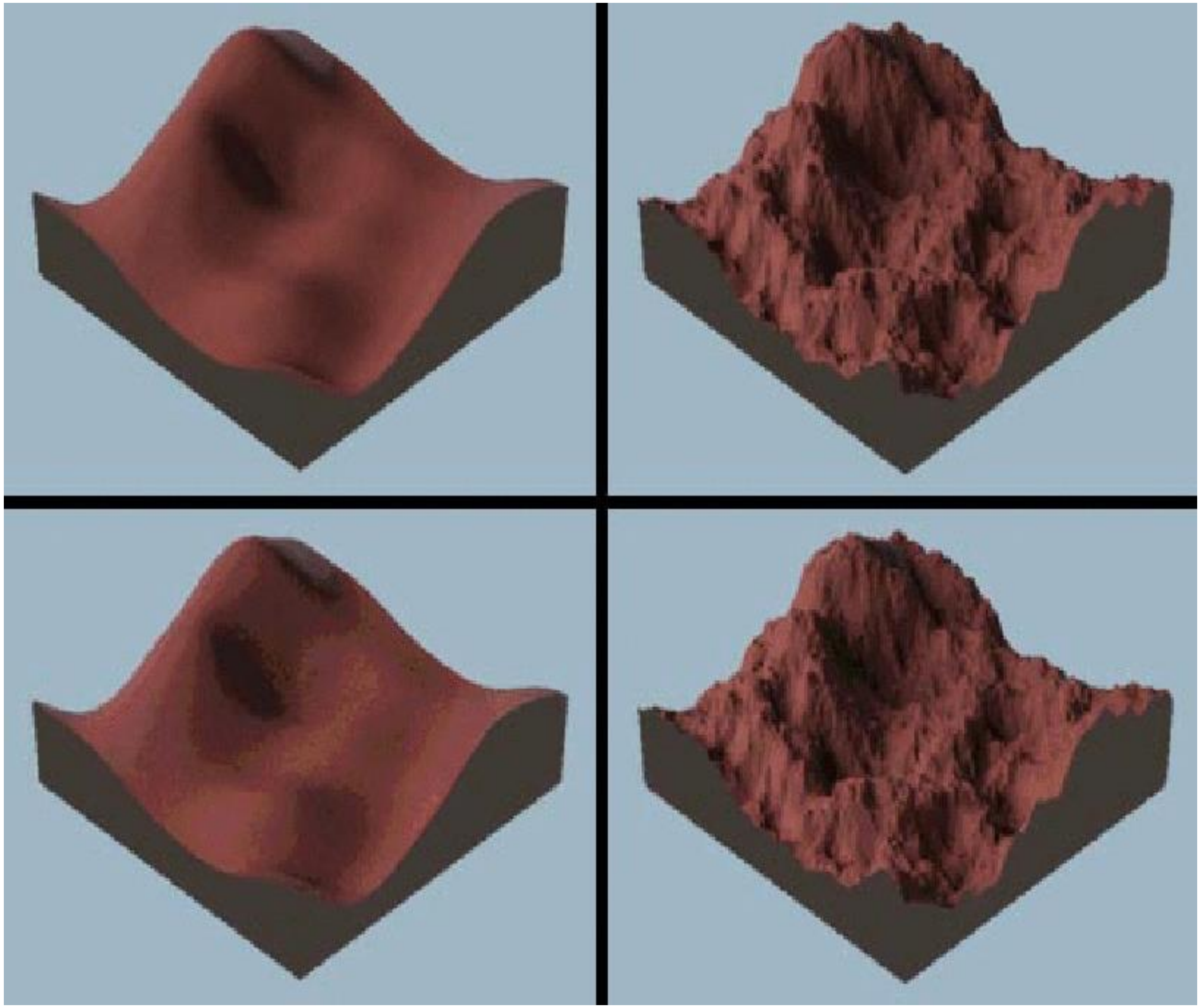
Pattanaik et al.

Visual Masking

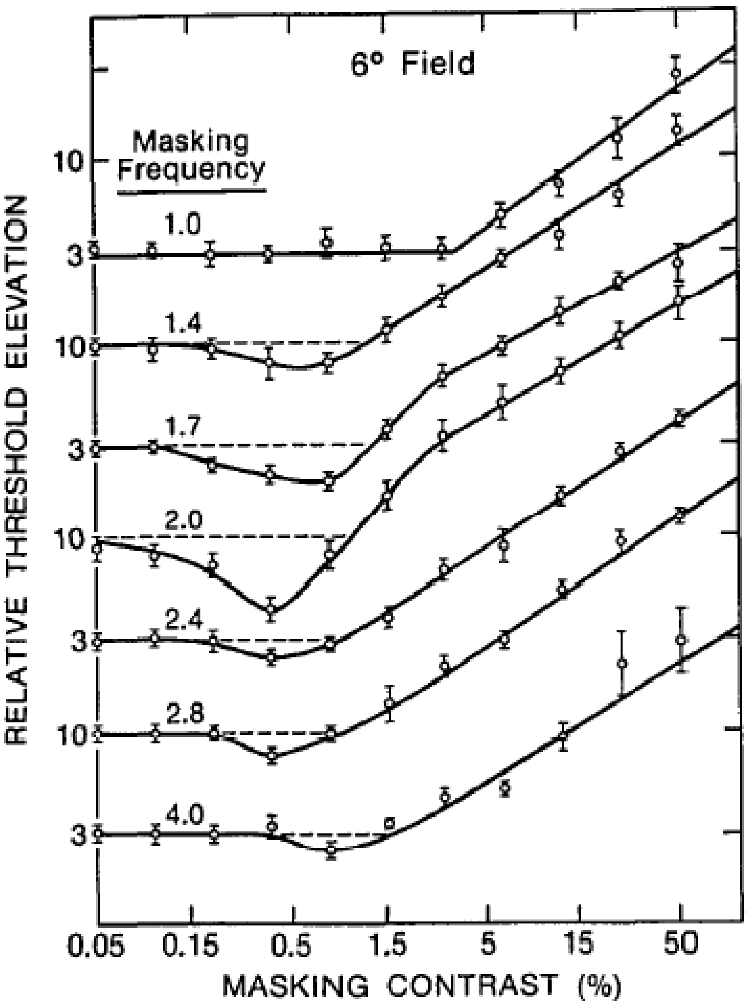
- **Strong masking: similar spatial frequencies**
- **Weak masking: different orientations**
- **Weak masking: different spatial frequencies**



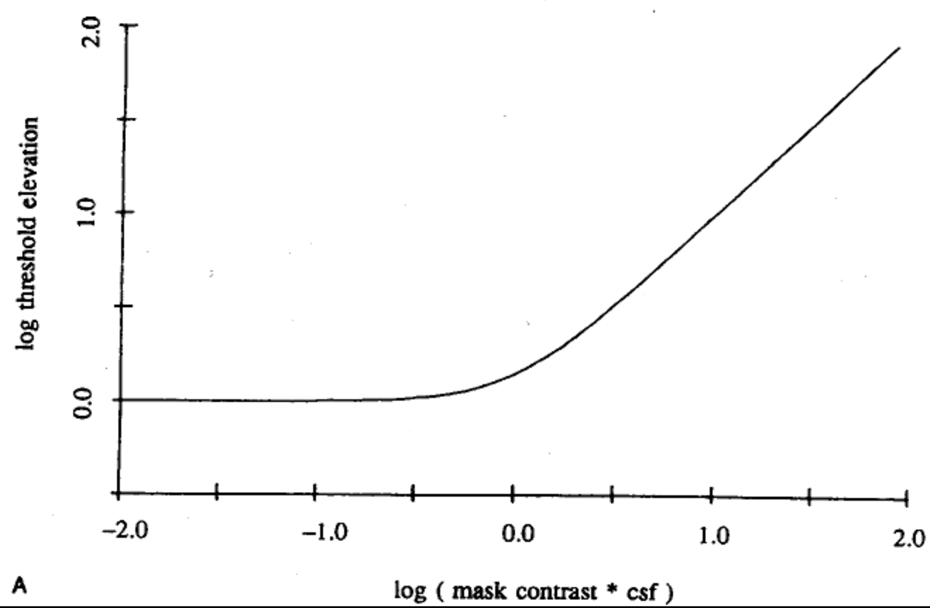
Visual Masking Example



Visual Masking Model

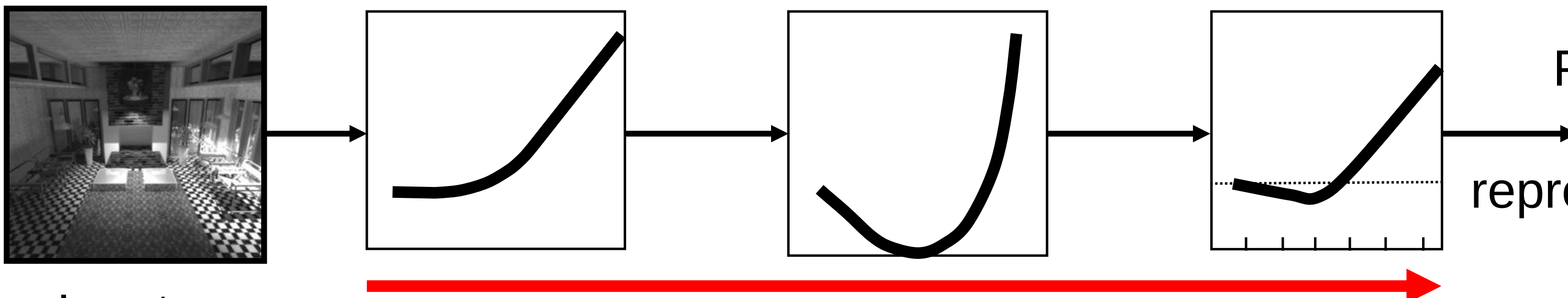


- **Masking is strongest between stimuli located in the same perceptual channel, and many vision models are limited to this intra-channel masking.**
- **The following threshold elevation model is commonly**



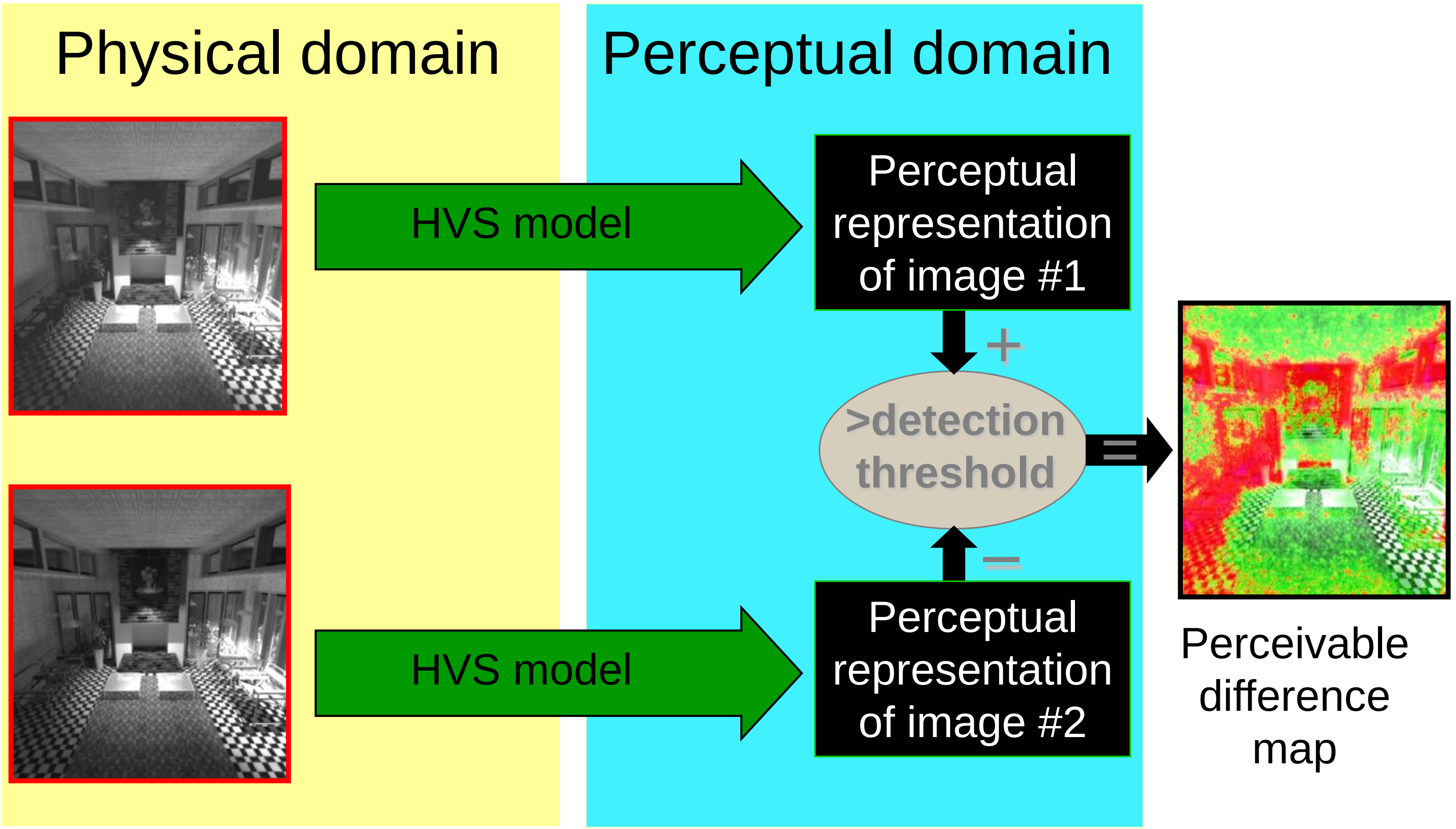
Typical HVS Model

Detection of perceivable differences between images strongly depends on the following characteristics of the human visual system:



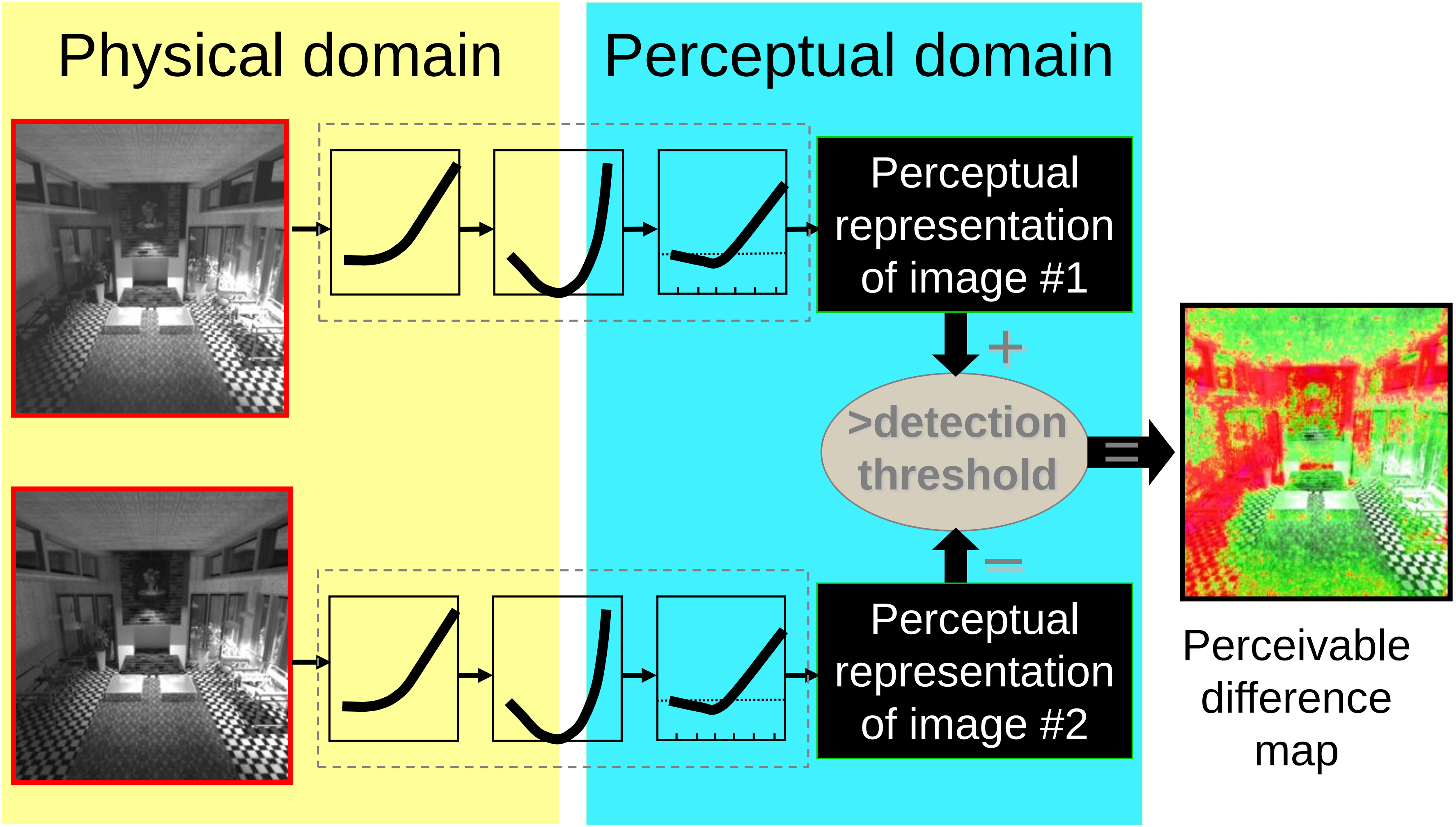
Pattanaik et al.

Perceivable Differences Predictor



Pattanaik et al.

Perceivable Differences Predictor

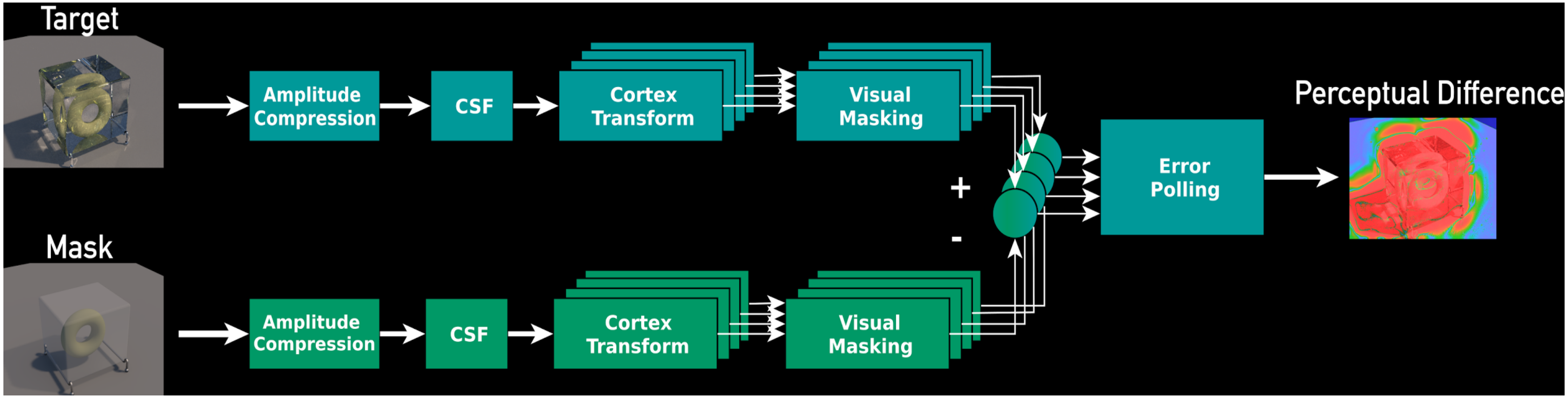


Pattanaik et al.

Color Problem

- **Contrast sensitivity for the color contrast is significantly lower than for the luminance contrast.**
- **HVS model for chromatic channels is similar as for the achromatic (luminance) channel .**
- **Two chromatic channels must be considered which leads to tripling the computation cost**

Daly’s Visible Differences Predictor

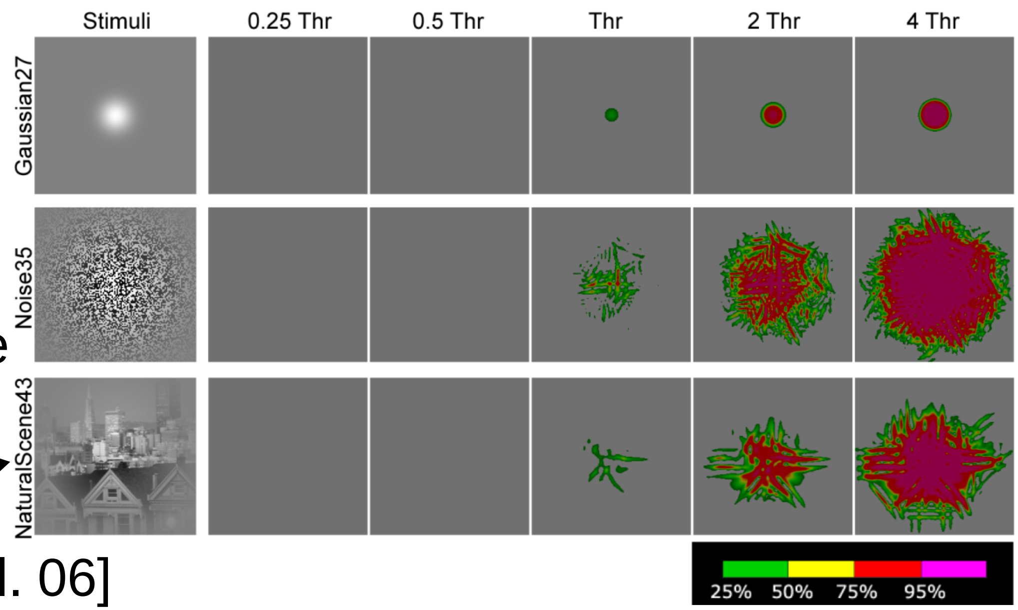


VDP: Outstanding Features

- **Predicts local differences between images**
- **Takes into account important visual characteristics:**
  - a Weber’s law-like amplitude compression,
  - advanced CSF model,
  - masking (mutual or unidirectional)
- **Uses the Cortex transform, which is a pyramid-style, invertible, and computationally efficient image representation**

Evaluation of Image Quality Metrics

- **Input images + Subjective responses = dataset**
- **Datasets**
  - Simpler evaluations
  - Reproducible evaluations
  - Should comprise typical artifacts
  - Should be publicly available



- **IMAGES**
  - Modelfest [Watson 99]
  - LIVE image db [Sheikh et al. 06]
  - TID (Tampere Image Database) [Ponomarenko et al. 09]
- **VIDEOS**
  - VQEG FRTV Phase 1 [VQEG '00]
  - LIVE video db [Seshadrinathan et al. 09]

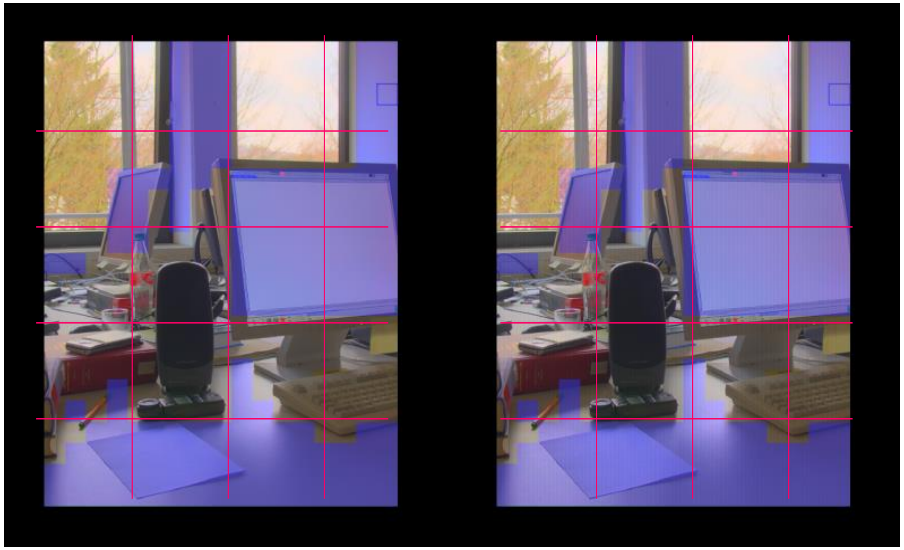
Evaluation of Image Quality Metrics

- Mostly only photos/real videos
- Focus on compression/transmission related artifacts
- Subjective responses: only overall quality (MOS)

Mean Opinion Score (MOS)		
MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Calibration: Experiment

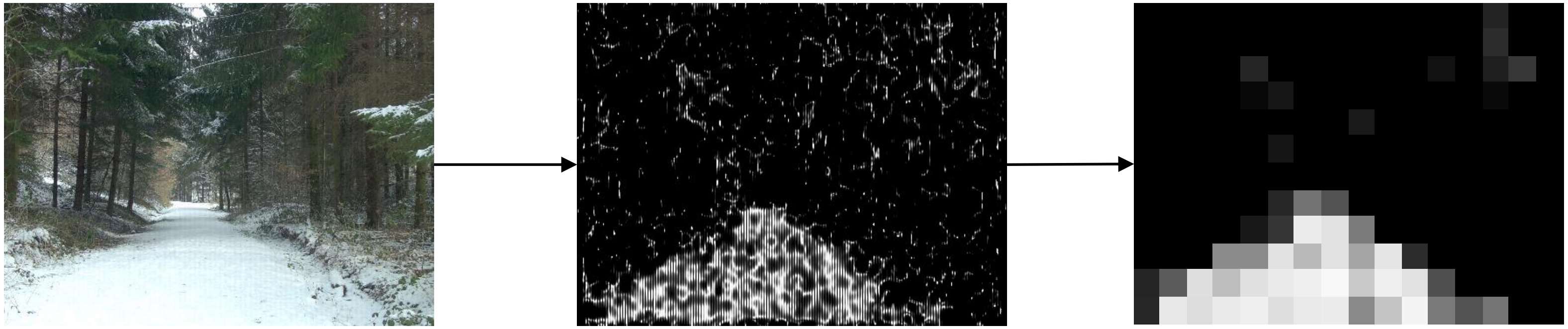
- **Subjects were to mark visible differences using rectangular blocks**



- **Results averaged across subjects**
  - fuzzy detection probability map

Calibration: Data Fitting

- **HDR VDP response converted to format of the subjective data**

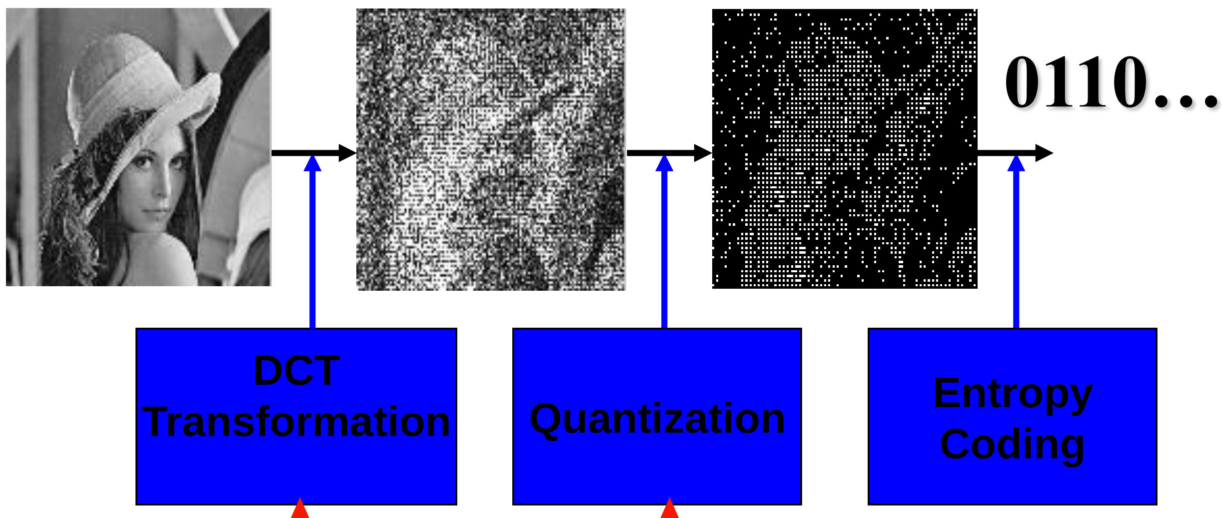


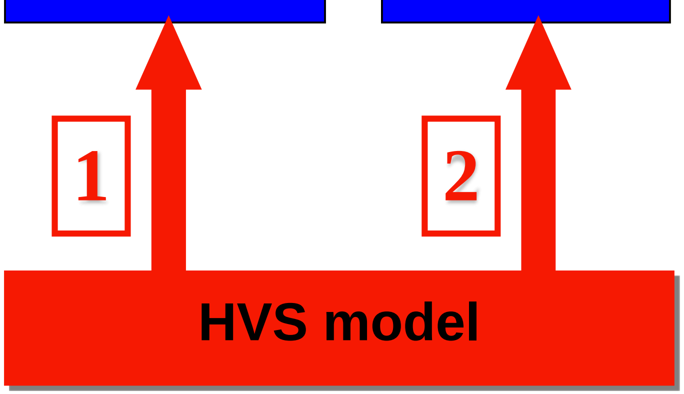
Distorted Image   VDP Response   Integrated Resp.

- **Found the best fit for peak threshold contrast and masking function slope**

Application Example –

Lossy Image Compression





16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Quantization matrix in JPEG [Annex K]

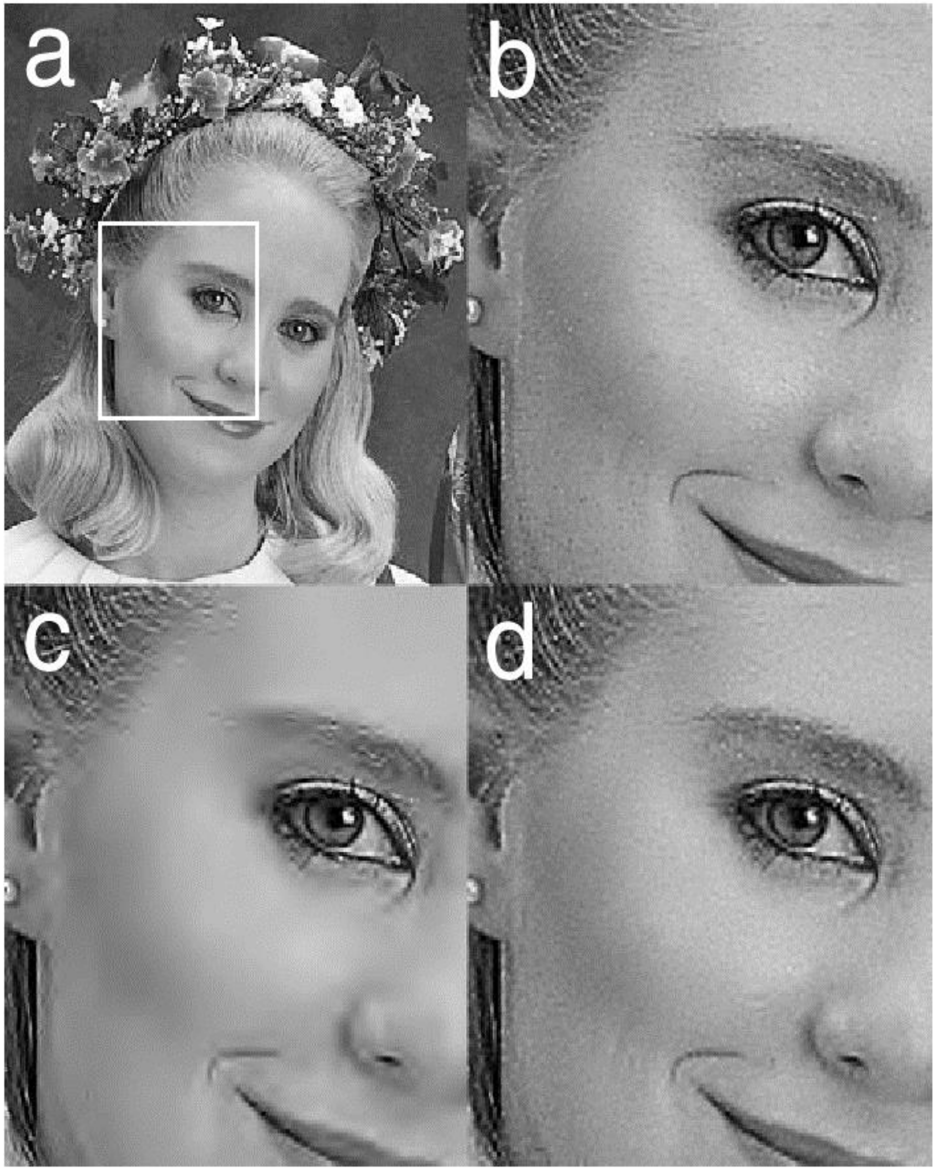
1. Image representation obtained as the result of DCT transformation should approximate the image representation in the Visual Cortex.
2. Perceivability of image distortions resulting from the quantization should be measured and controlled by a perceptual error metric.

JPEG 2000

a,b – original image,

c – standard JPEG 2000 algorithm controlled by a metric minimizing the MSE. The missing skin texture appears blurred and unnatural to the human observer. Exact reproduction of spatial detail, e.g., hair of the woman is less important due to visual masking by strong textures.

d – JPEG 2000 controlled by a perceptual image quality metric.



Nedenau et al.

Prediction of Shadow Masking

- Visualization of the contrast threshold elevation due to masking.
- Stronger masking occurs when the target image contains a texture (top row).
- Bright green denotes more masking.

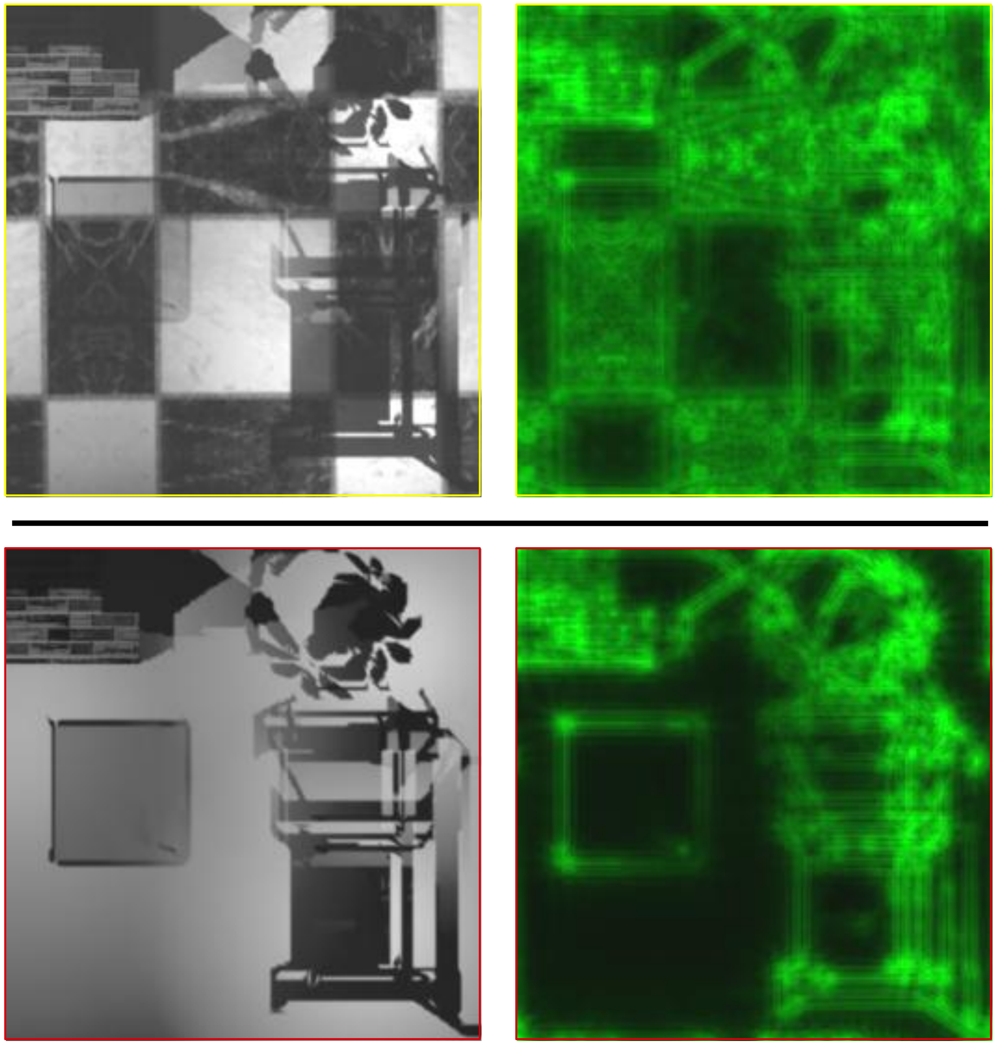


Image Quality Metrics

- Common quality metrics were designed for predicting visibility of typical distortions in photographs:
  blur, sharpending, noise, JPEG/ MPEG compression,...

Blur Sharpening

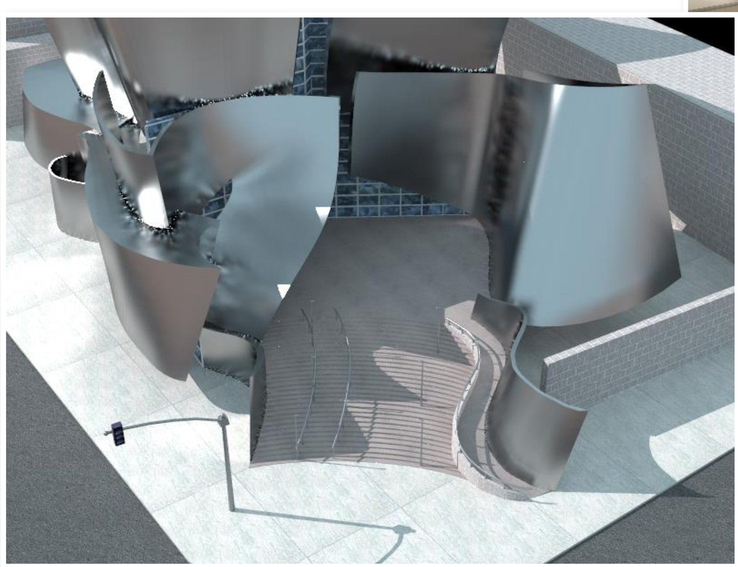
What about synthetic CG-images?

JPEG/ MPEG distortions

Contouring, banding

Rendering Artifacts

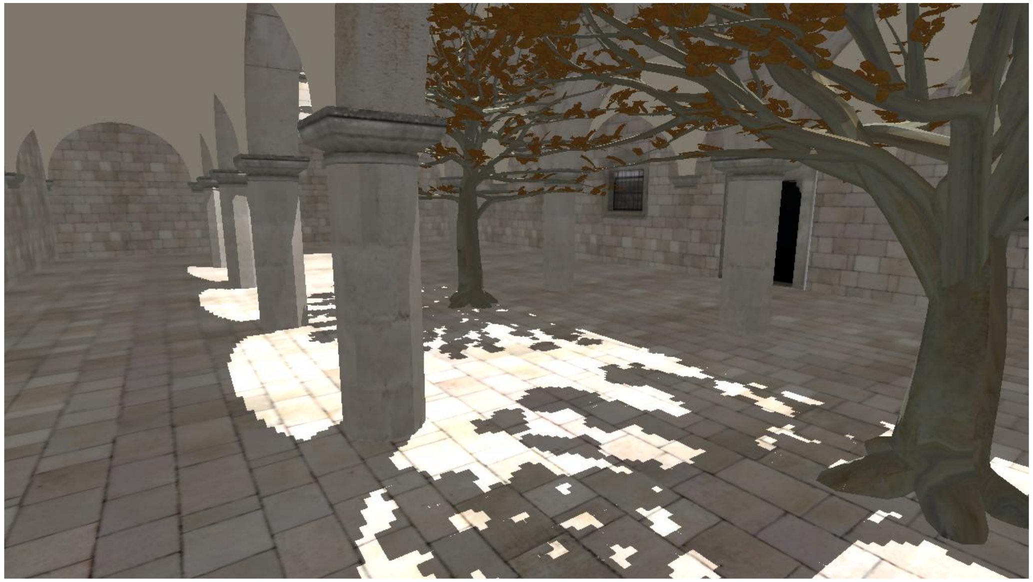
- **e.g., low-freq. noise from glossy instant radiosity or photon density estimation**



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 iterations


8 iterations



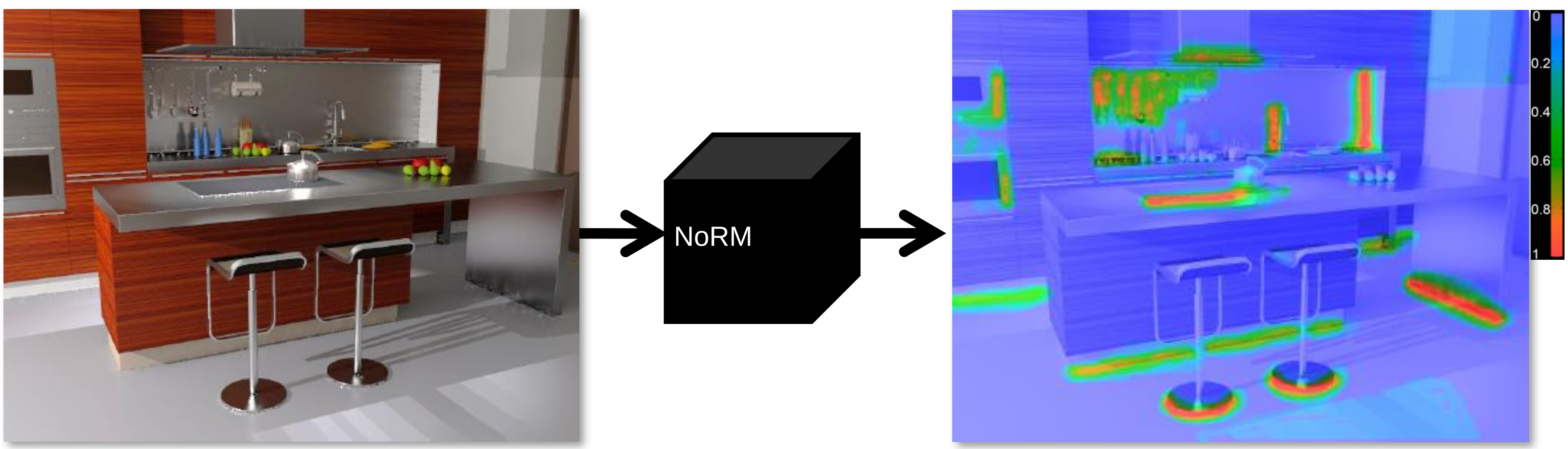
60 iterations


150 iterations


1500 iterations

No-Reference Metric of Image Quality

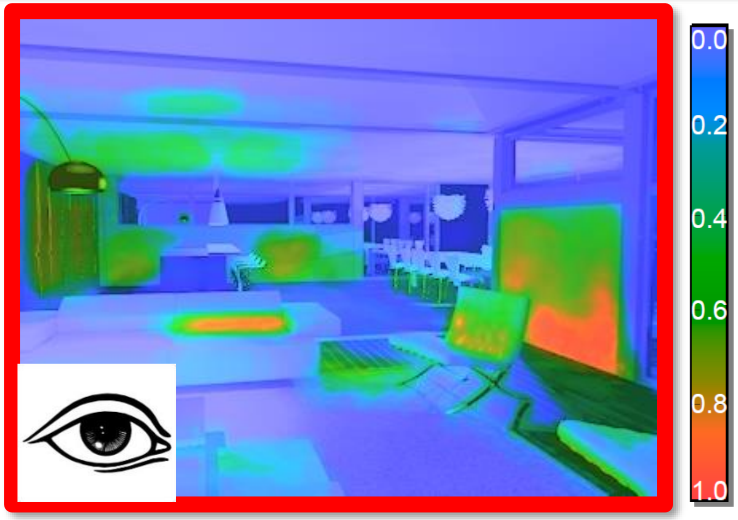
- **NoRM**
  - Input: distorted image/video frame (no reference)
  - Output: map of distortions (possibly perceptually weighted)



Experiment - Mean Distortion Maps

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

User Experiment – with Reference

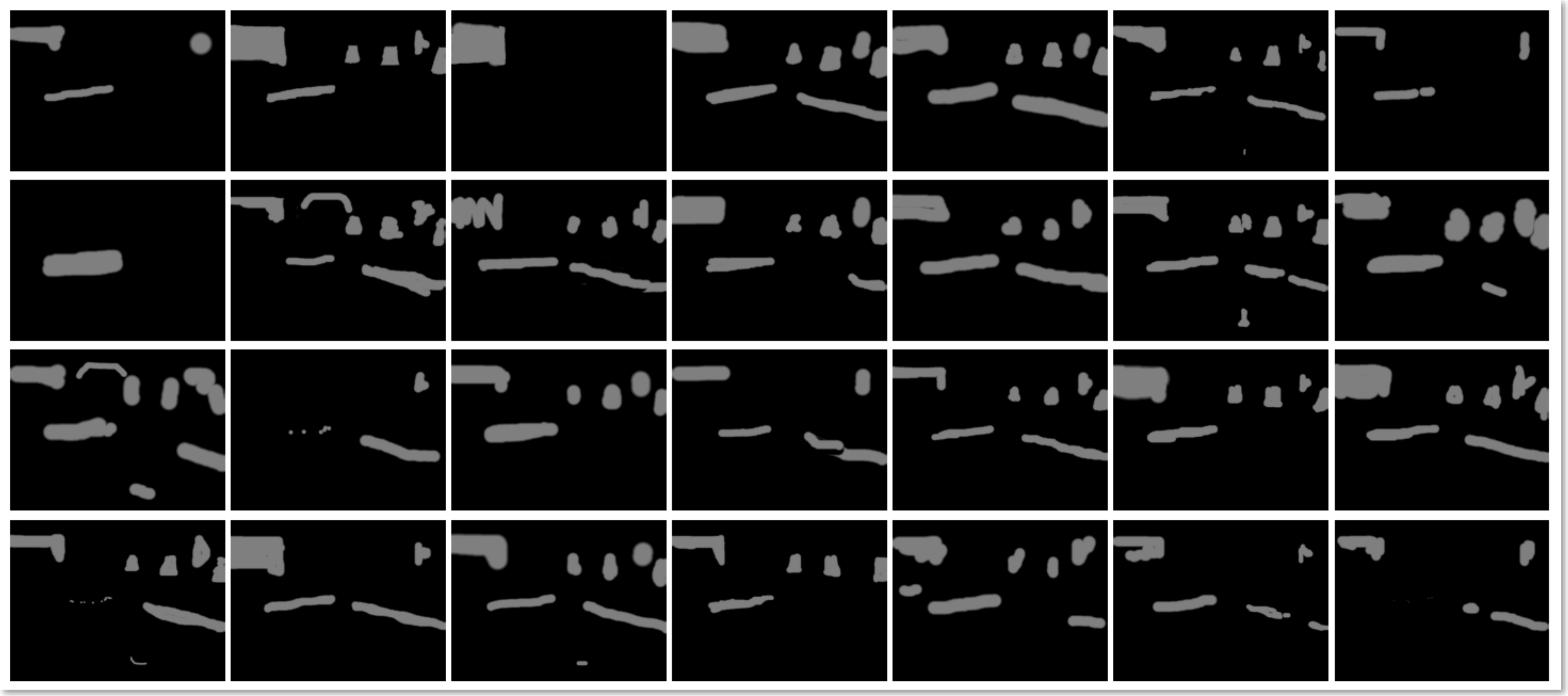


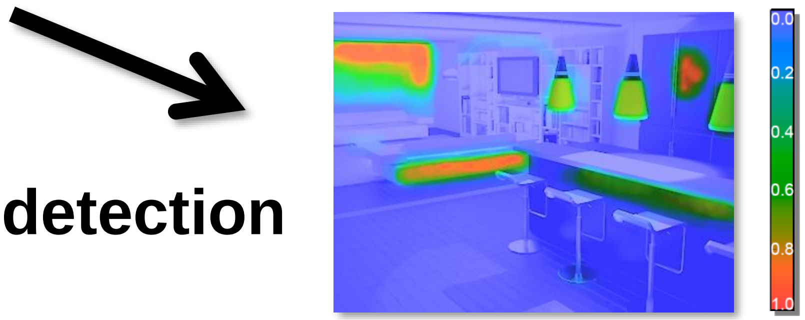
- **Noticeable distortions: Mean Distortion Map**

User Experiment – No Reference

- **Objectionable distortions: Mean Distortion Map**

Example User Responses

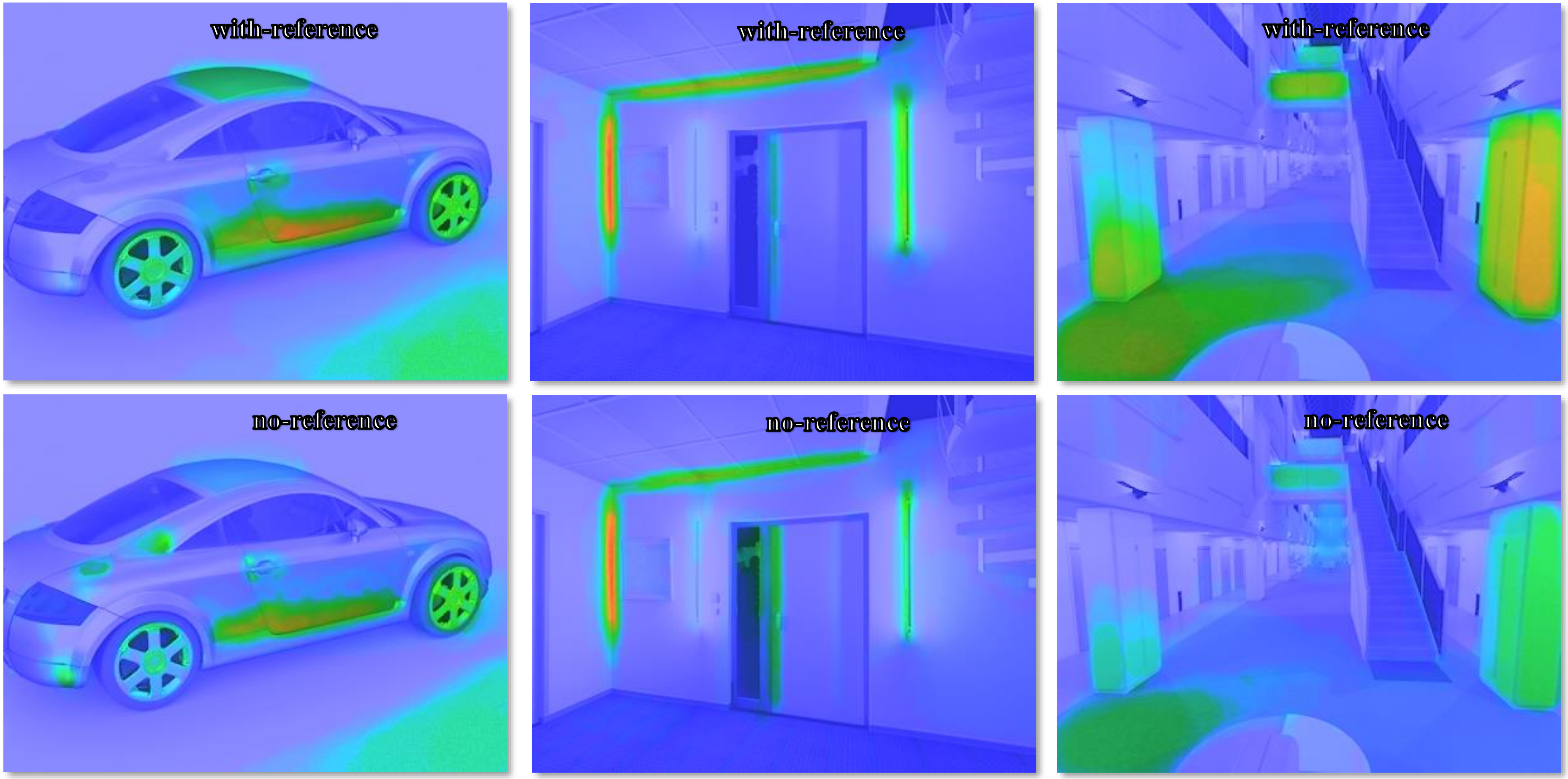




- **Probability of detection**

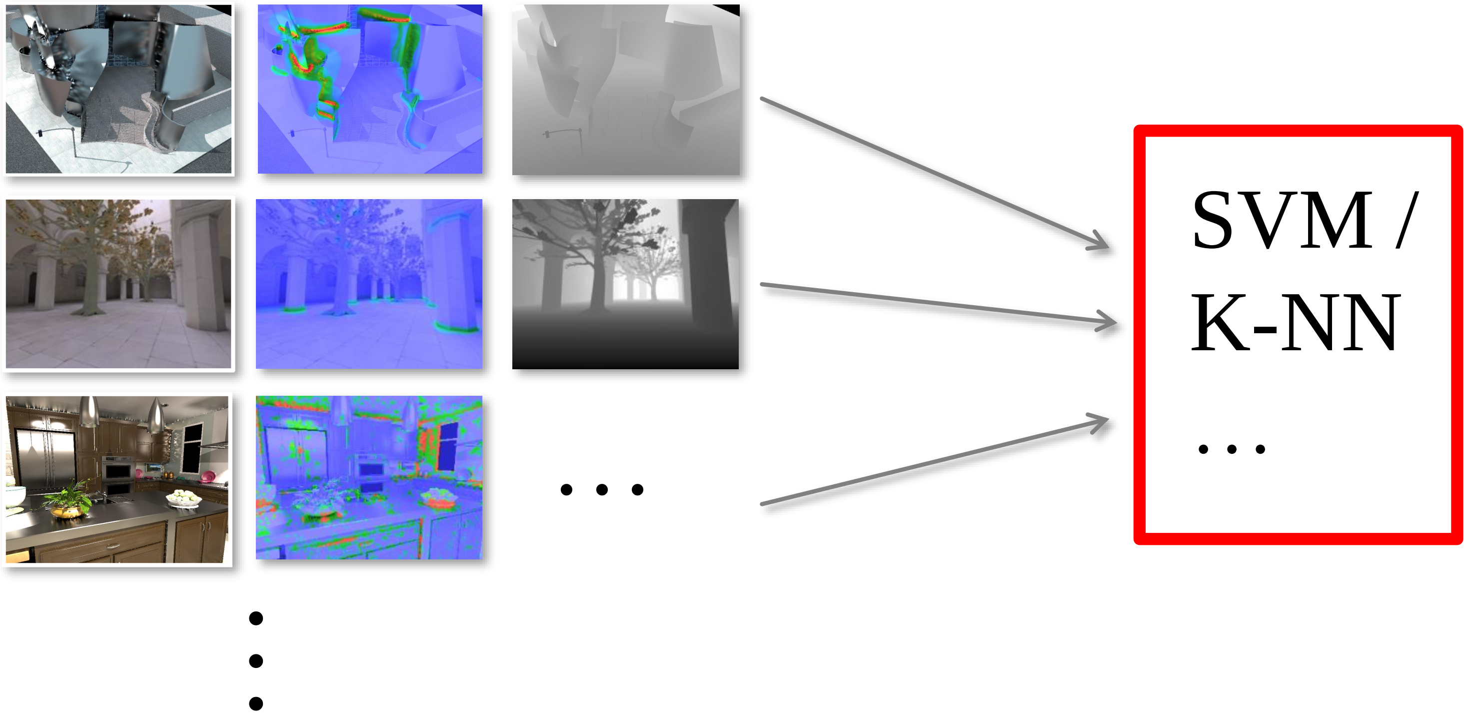
With-reference vs. No-reference

- **Results rather similar**



Data-Driven No-Ref. IQM

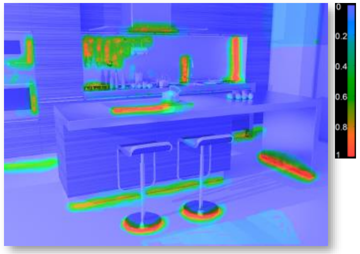
- Feature descriptors (various information available)
- Distortion maps (possibly real subjective data)
- Depth + 3D related information



Data-Driven No-Ref. IQM

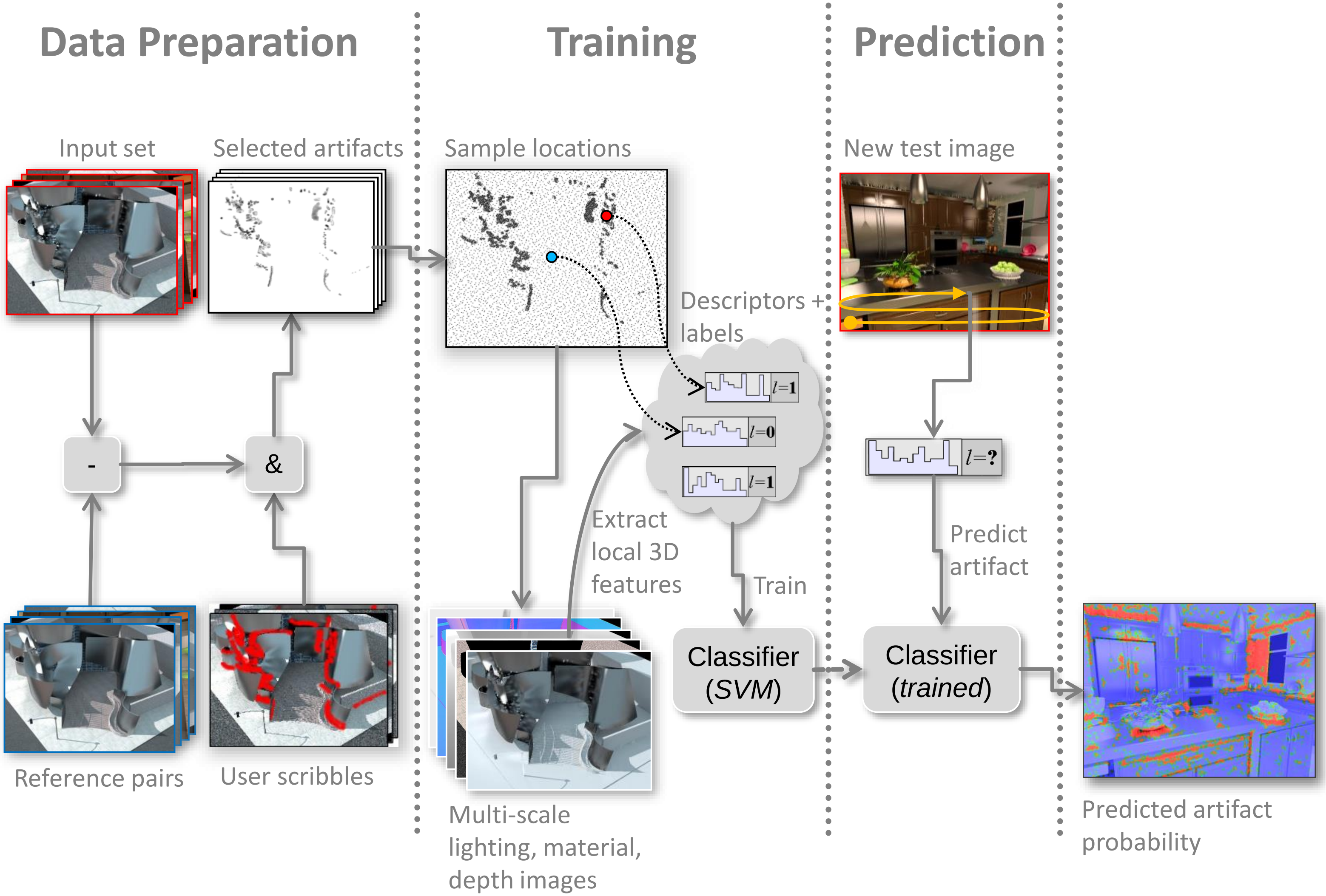
- **Distorted (rendered) image → prediction**
  - Traditional metrics: just a number on scale 1-5
- **We want a distortion map per pixel**
  - Much harder problem
  - But ... we have 3D data!!!

SVM /
K-NN
…

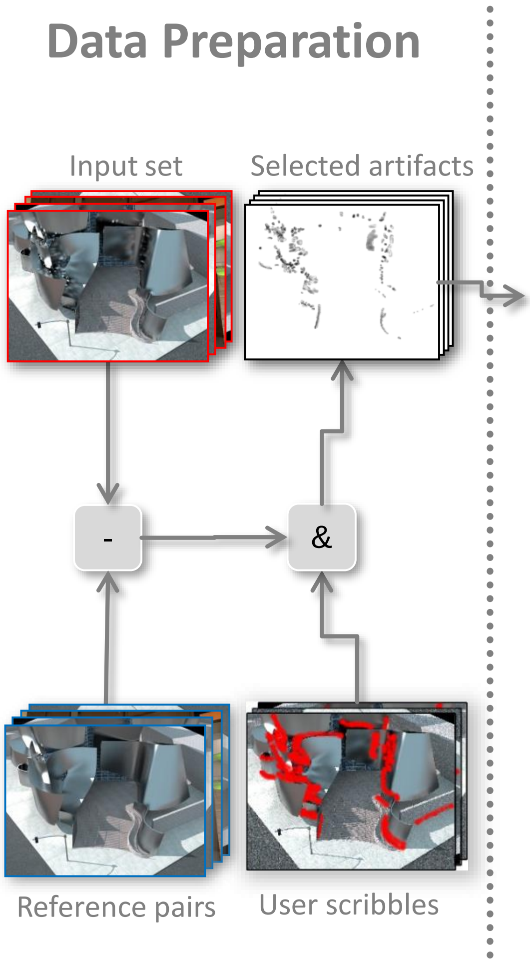


Distortion strength

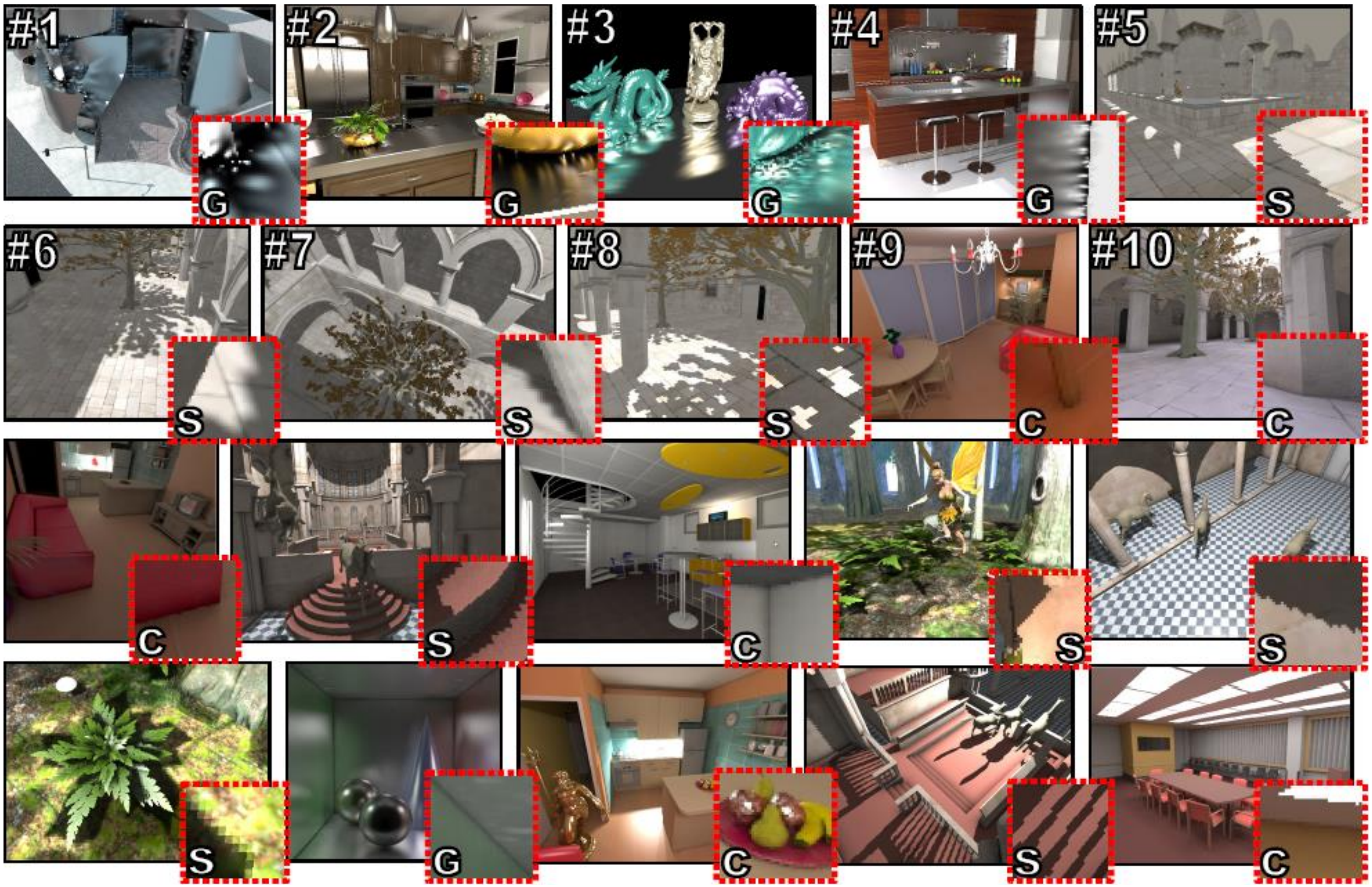
System Pipeline NoRM



System Pipeline NoRM



Rendering Artifact Data Sample

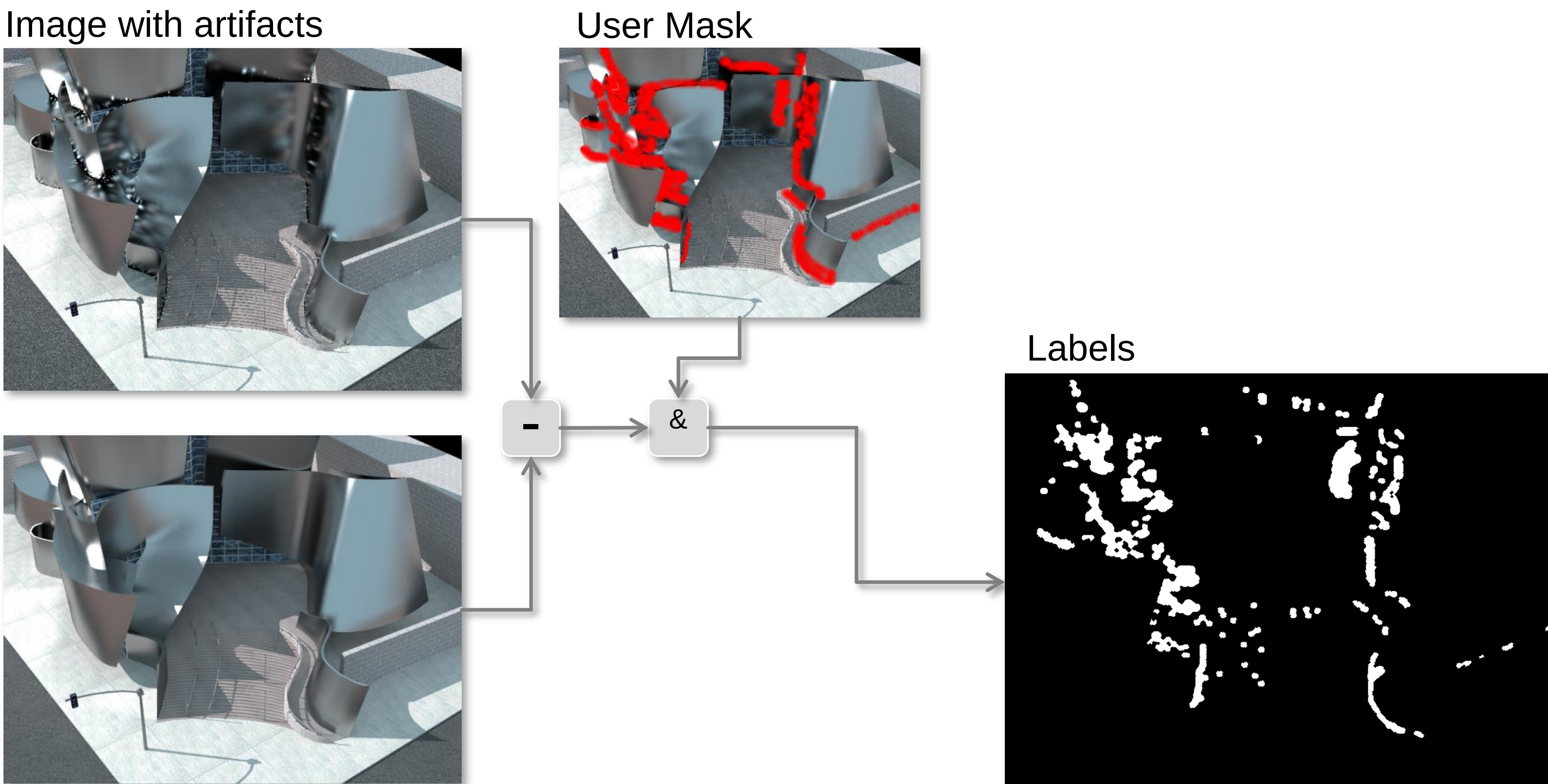


User Experiment

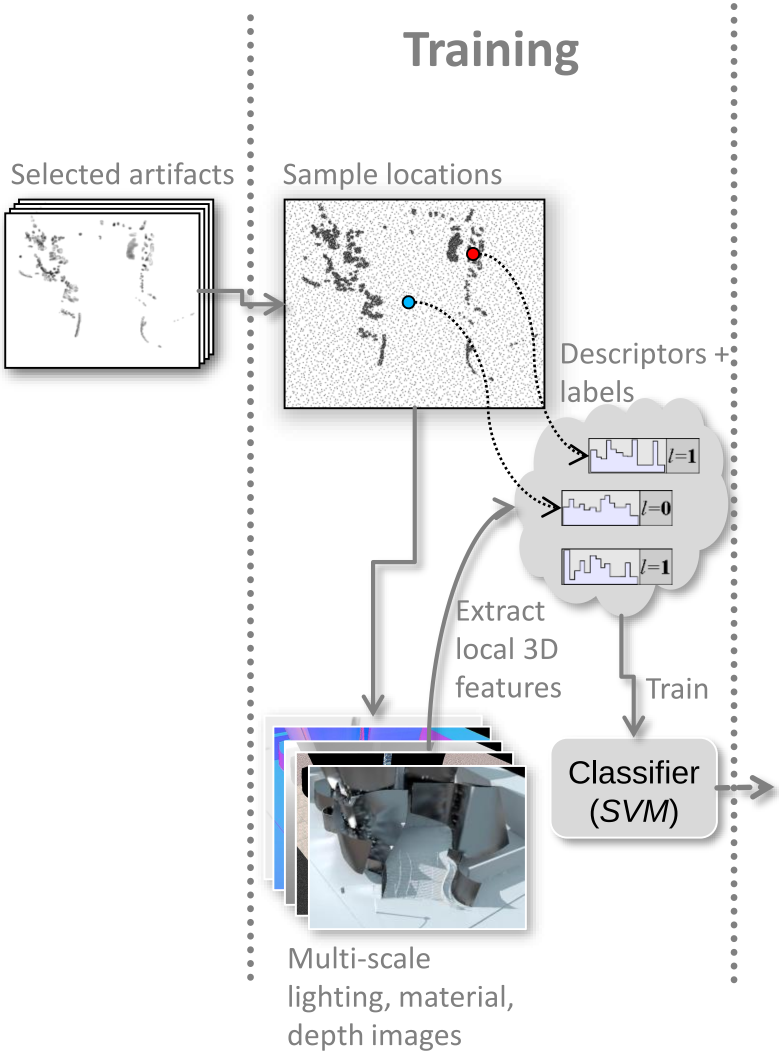
- **Which Pixels are Artifacts?**
  - → Asked 20 subjects
- **Scribbling application**
- **No-reference / With-reference**

Computing the Mask

- **Given artifact image + reference + user mask**
  - compute the error labels within the user mask

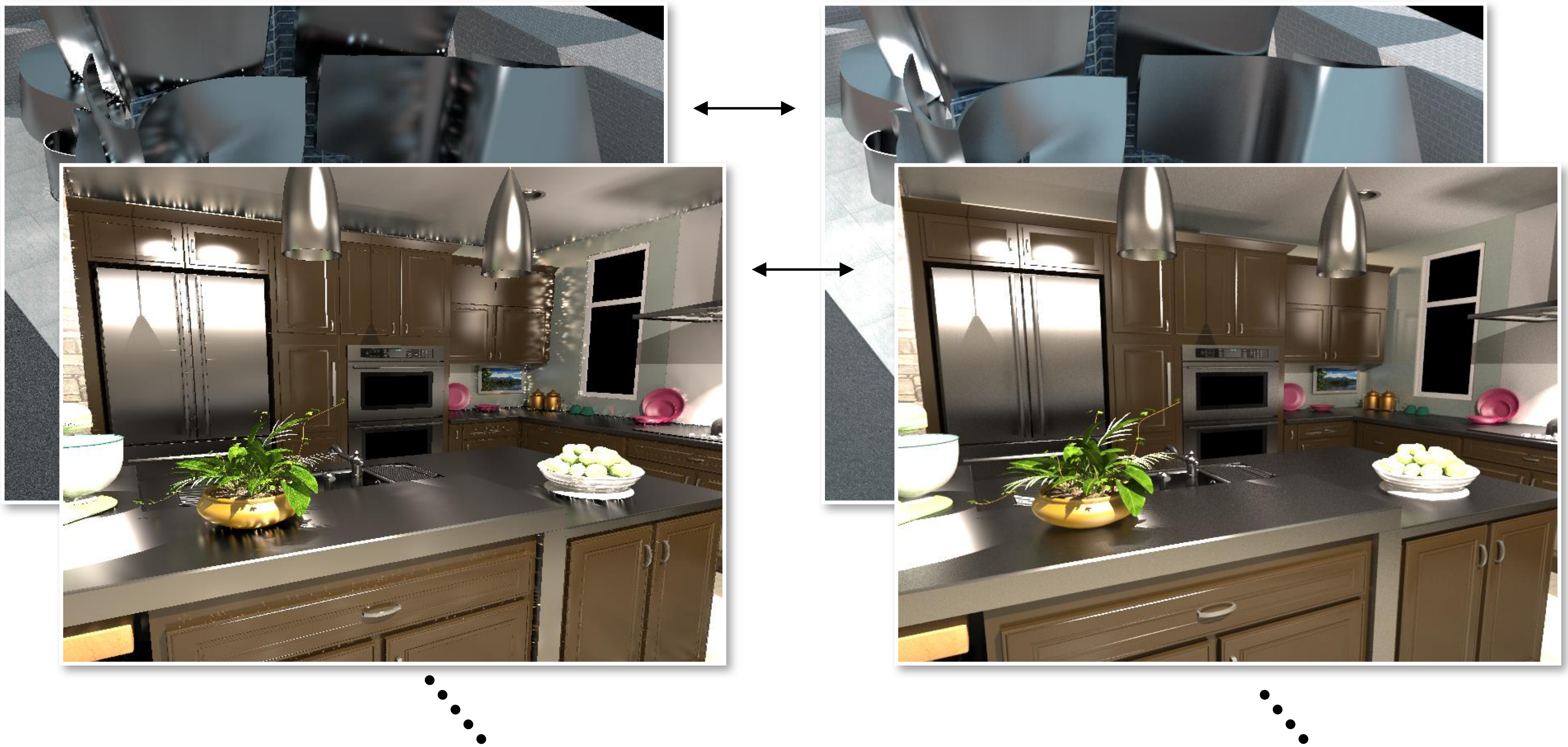


System Pipeline NoRM

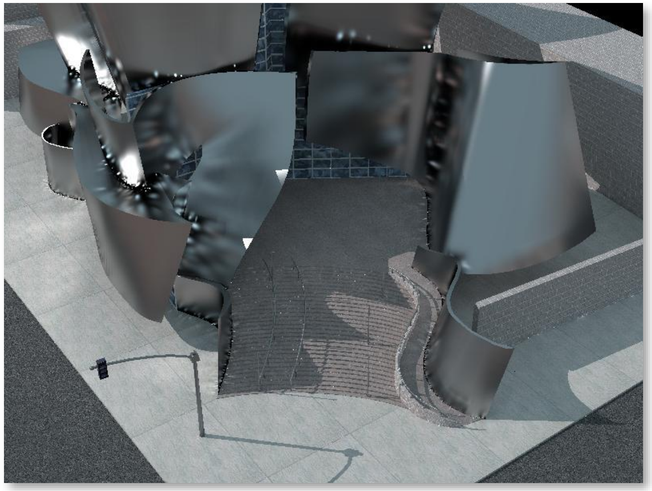


Training Classifier

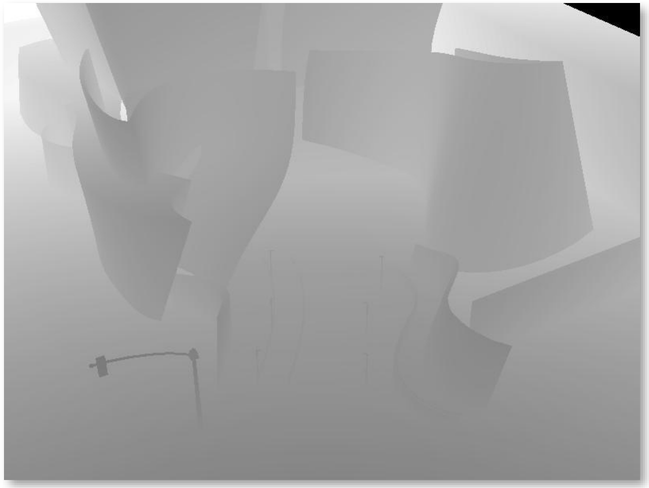
- **Given input data:**
  - color, depth, material for one artifact type
  - user scribbled artifact mask
  - reference image without artifacts



Rendering Output – Classification Input



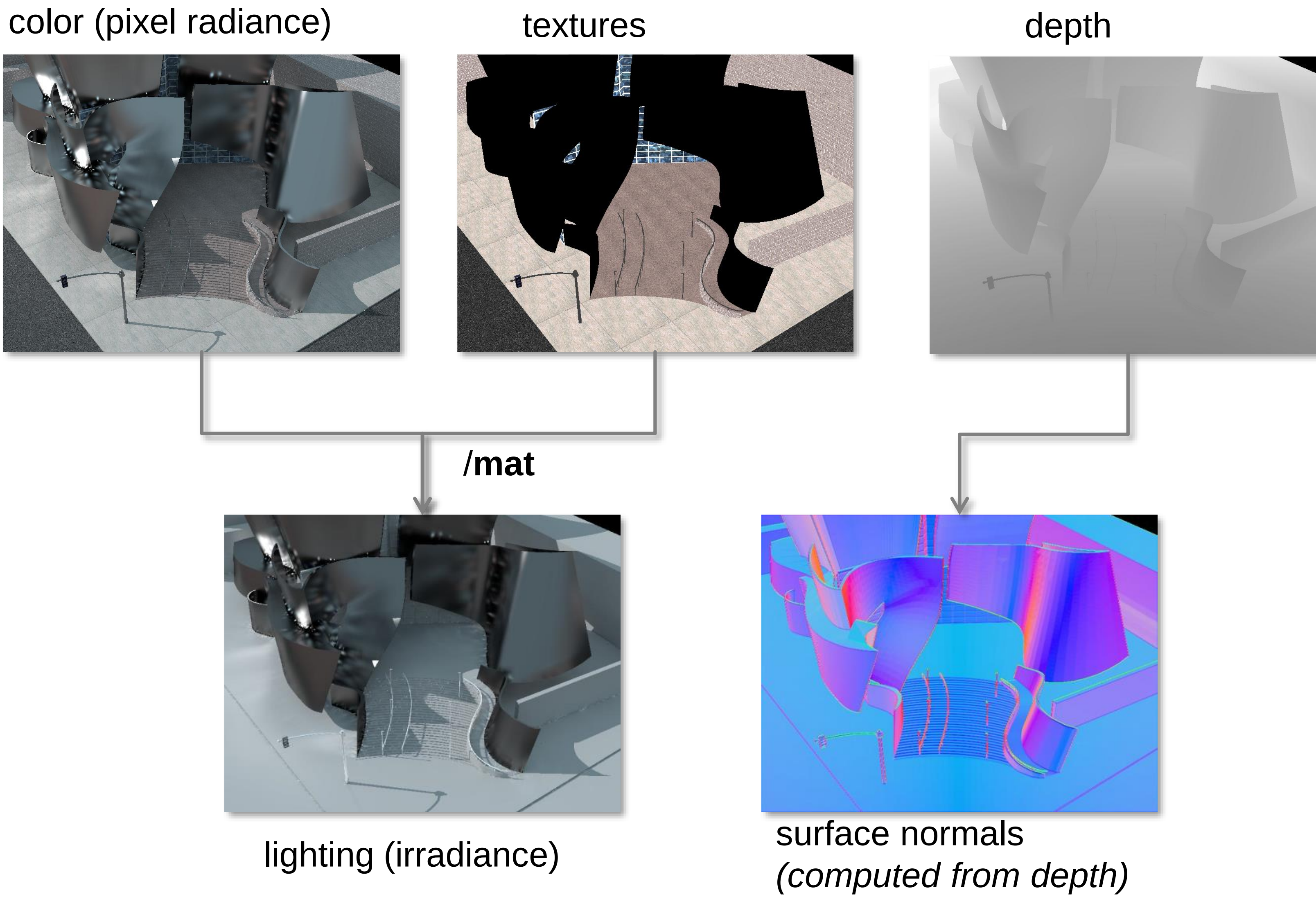
HDR (LDR) color image
(may contain noise)



depth buffer
(in high precision, no noise)

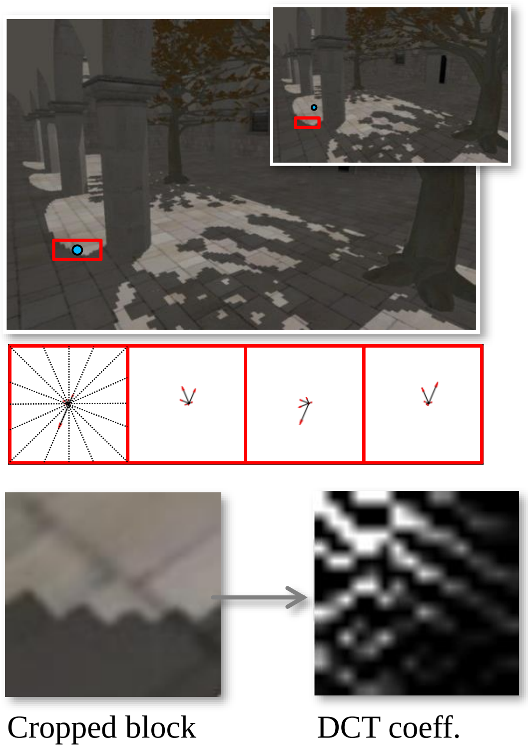
diffuse texture buffer

Computation of Additional Input Data

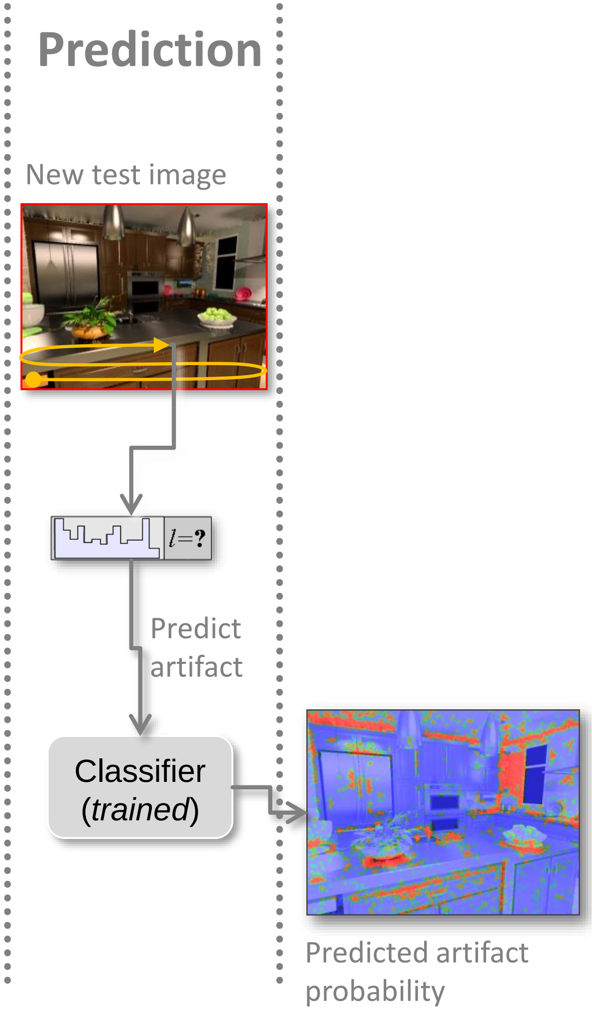


Feature Descriptors

- 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



System Pipeline NoRM

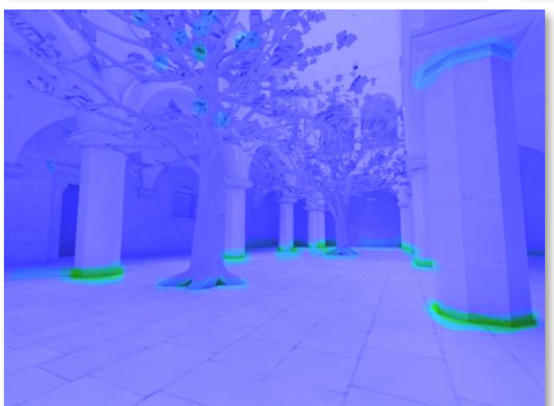


Performance 2D / 3D

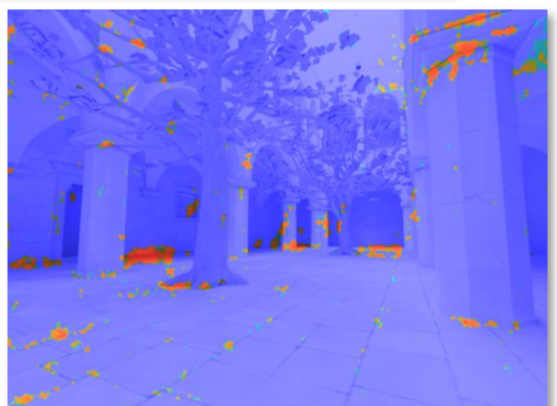
- HoG (lighting) versus HoG (lighting + depth)

Input

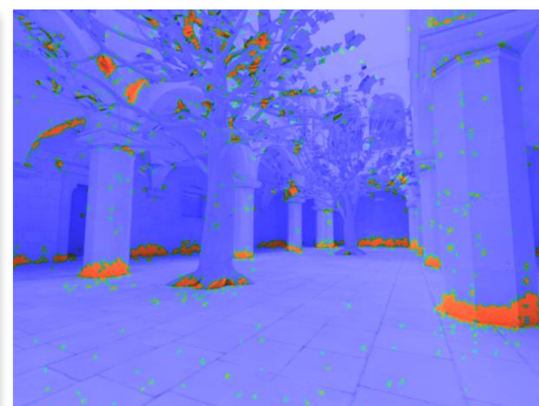
Color



Ground-truth
(User-masks)

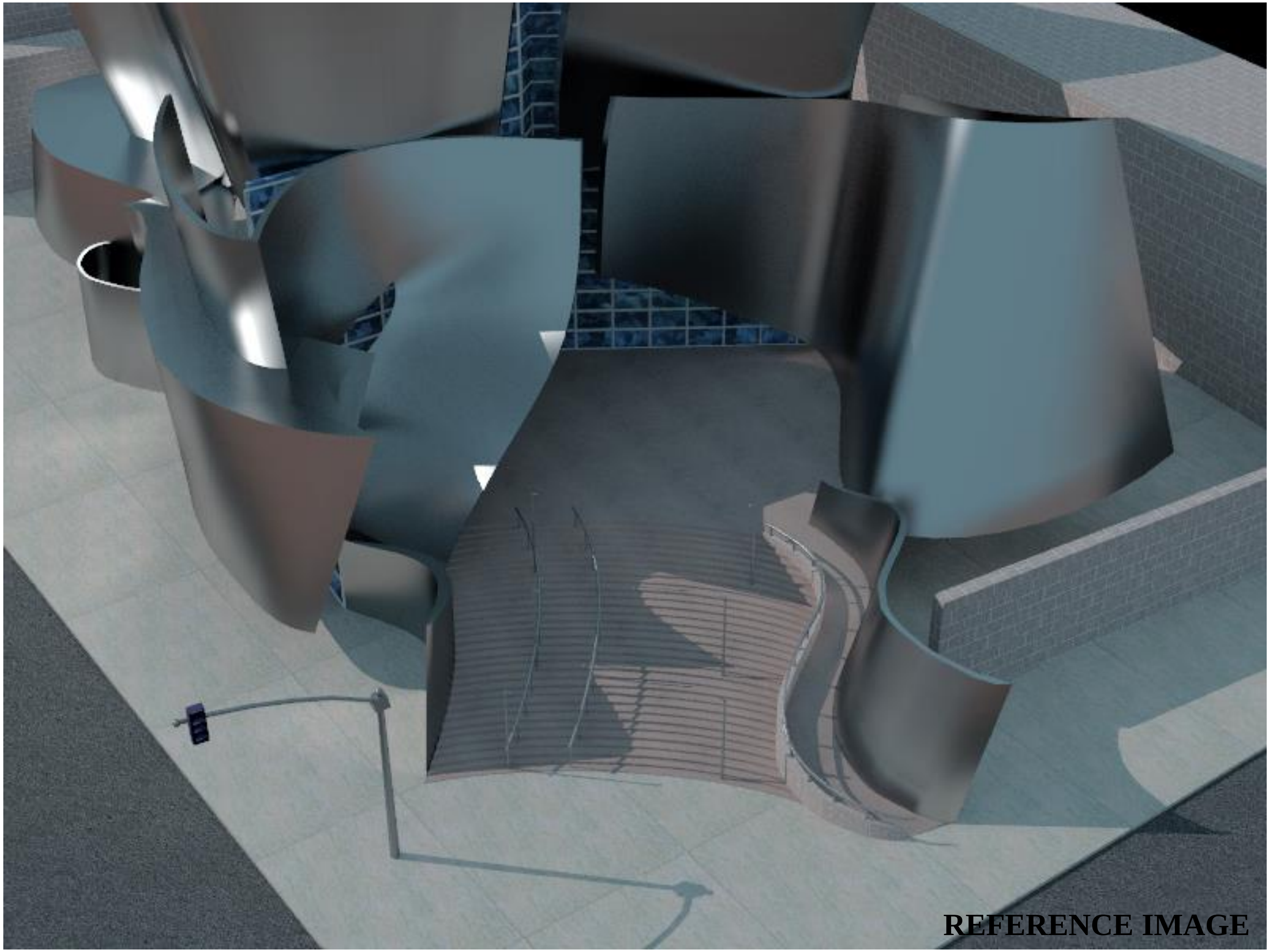


Color Descriptor

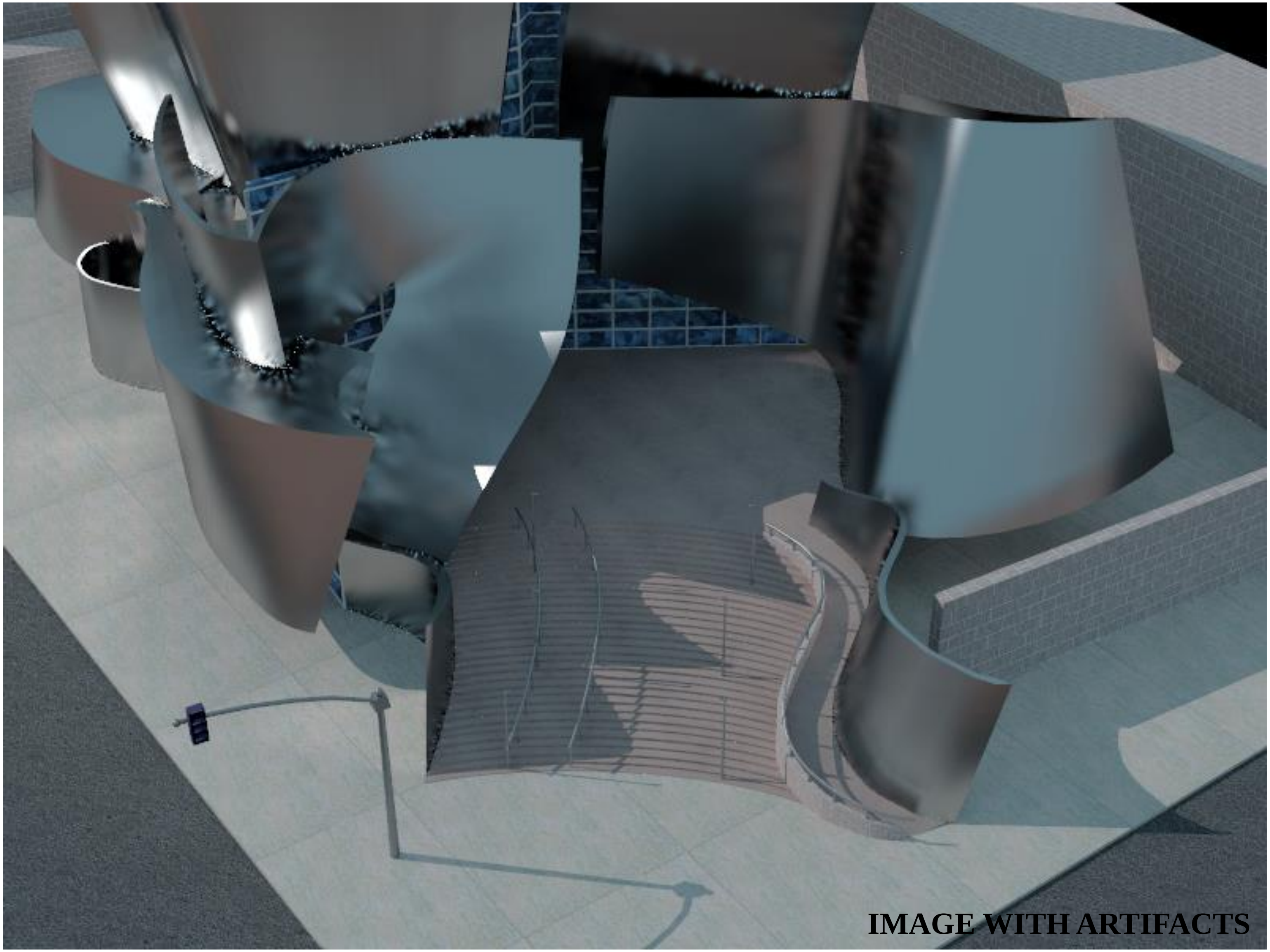


Color + Depth
Descriptor

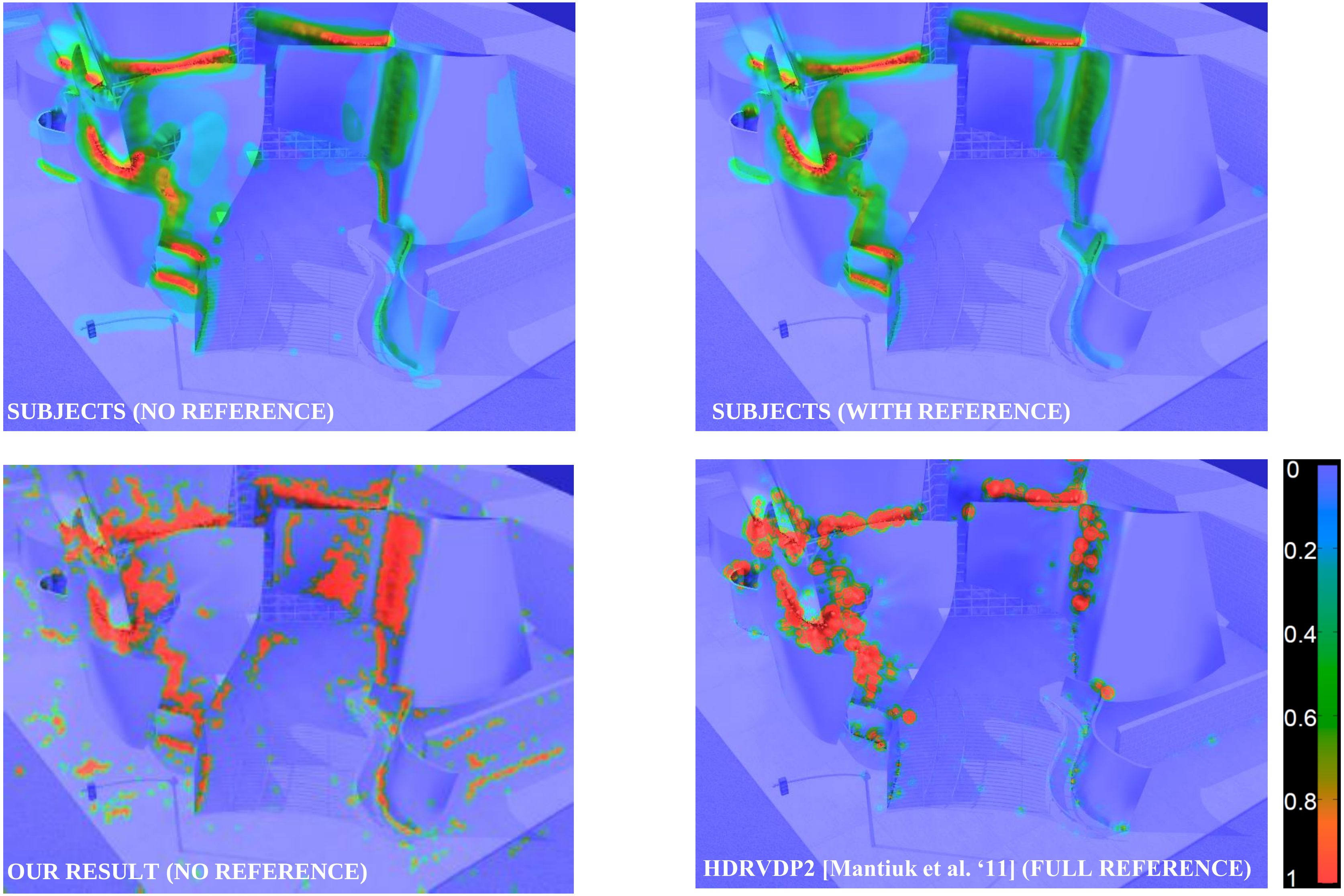
Comparison



Comparison

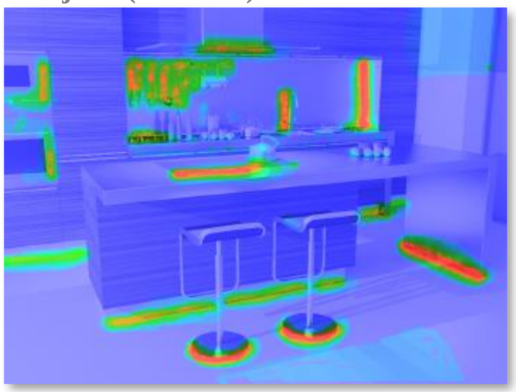


Classification Results

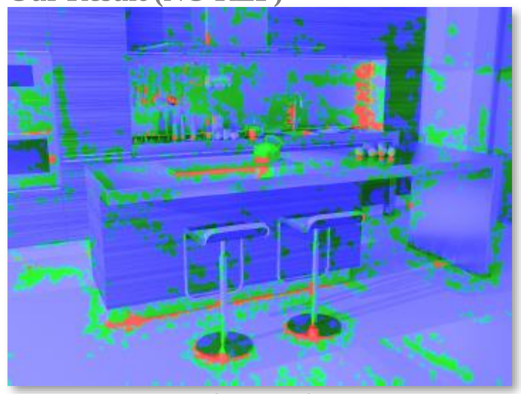


Results (VPL noise)

Subjects (NO REF)

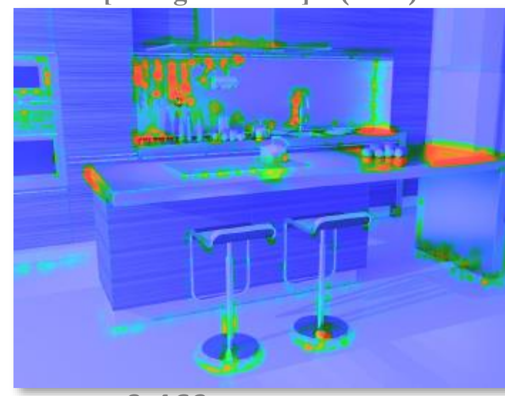


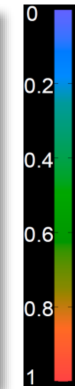
Our Result (NO REF)



corr = 0.436 (0.298)

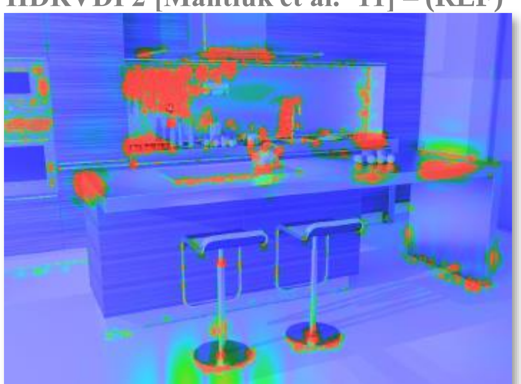
SSIM [Wang et al. '04] – (REF)





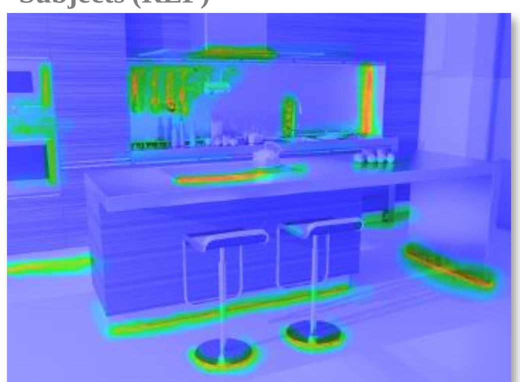
corr = 0.469

HDRVDP2 [Mantiuk et al. '11] – (REF)



corr = 0.495

Subjects (REF)

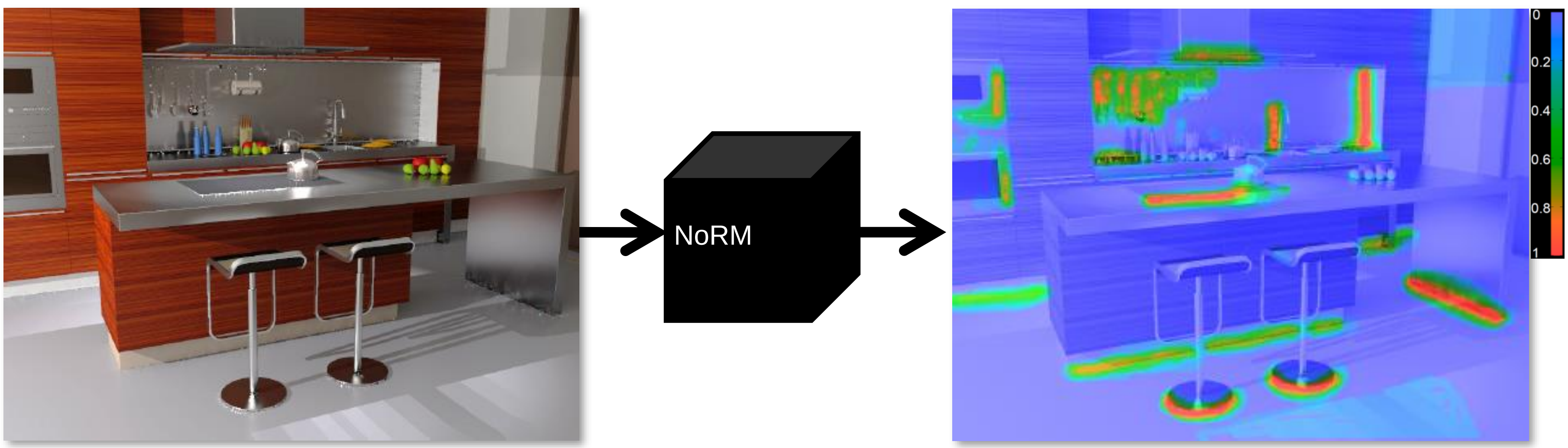


corr = 0.913

Artifact Image

No-Reference Data Driven Metrics

- **NoRM: No-Reference CG-image quality Metric**
- **Blind metric for local rendering artifacts is possible**
  - → If we know what we are looking for
  - → 3D and texture information is available

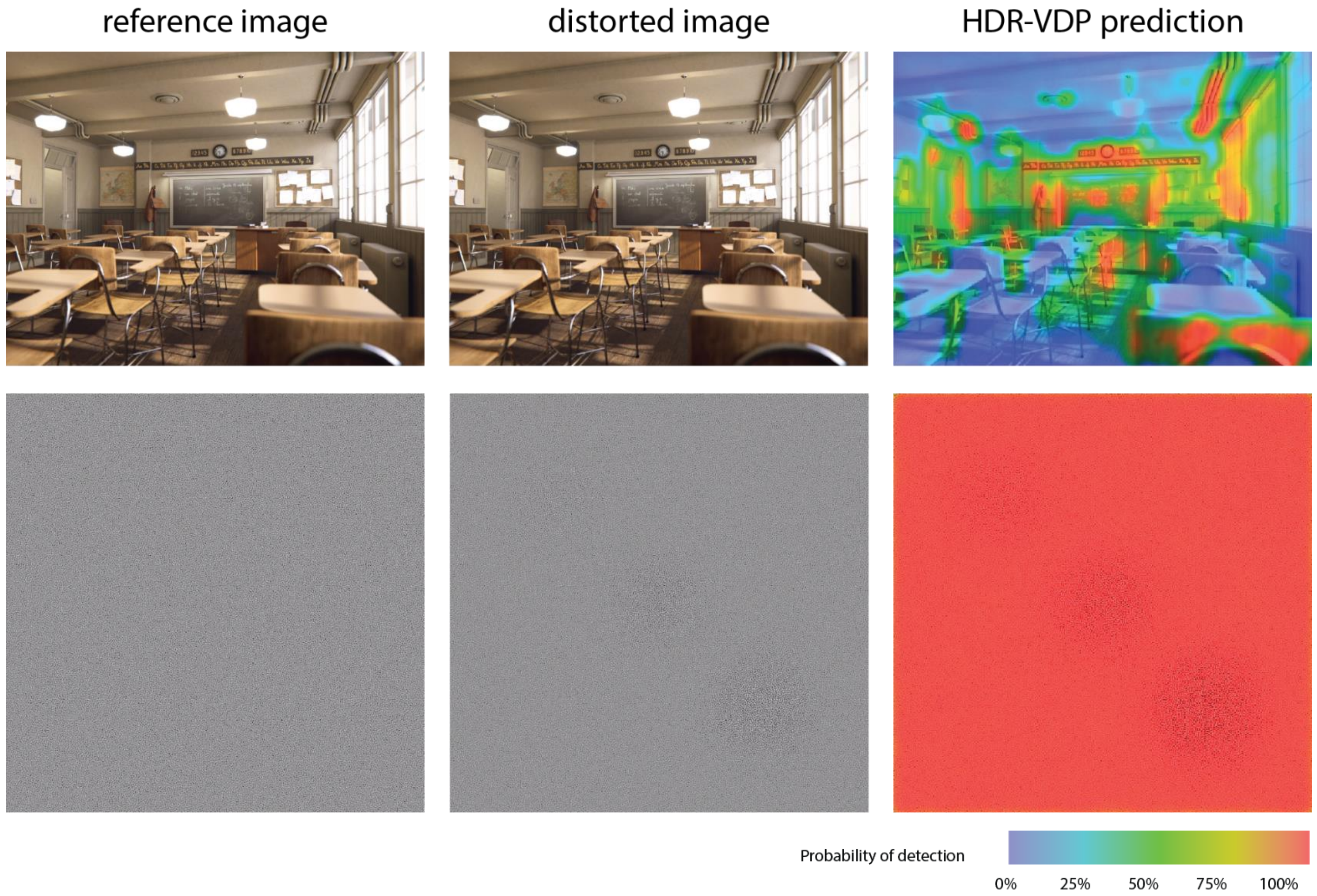


CNN-based FR local visibility metric

Motivation:

- No reference metrics typically work only for some particular distortion types.
- No reference metrics tend to mark non-distorted areas.
- As state-of-the-art research shows that learn-based methods outperform the hand-crafted ones.
- Existing visibility metrics (e.g. HDR-VDP) still have many flaws.
- Creating a versatile metric taking into account many type of distortions.

Imperfections of existing visibility metrics

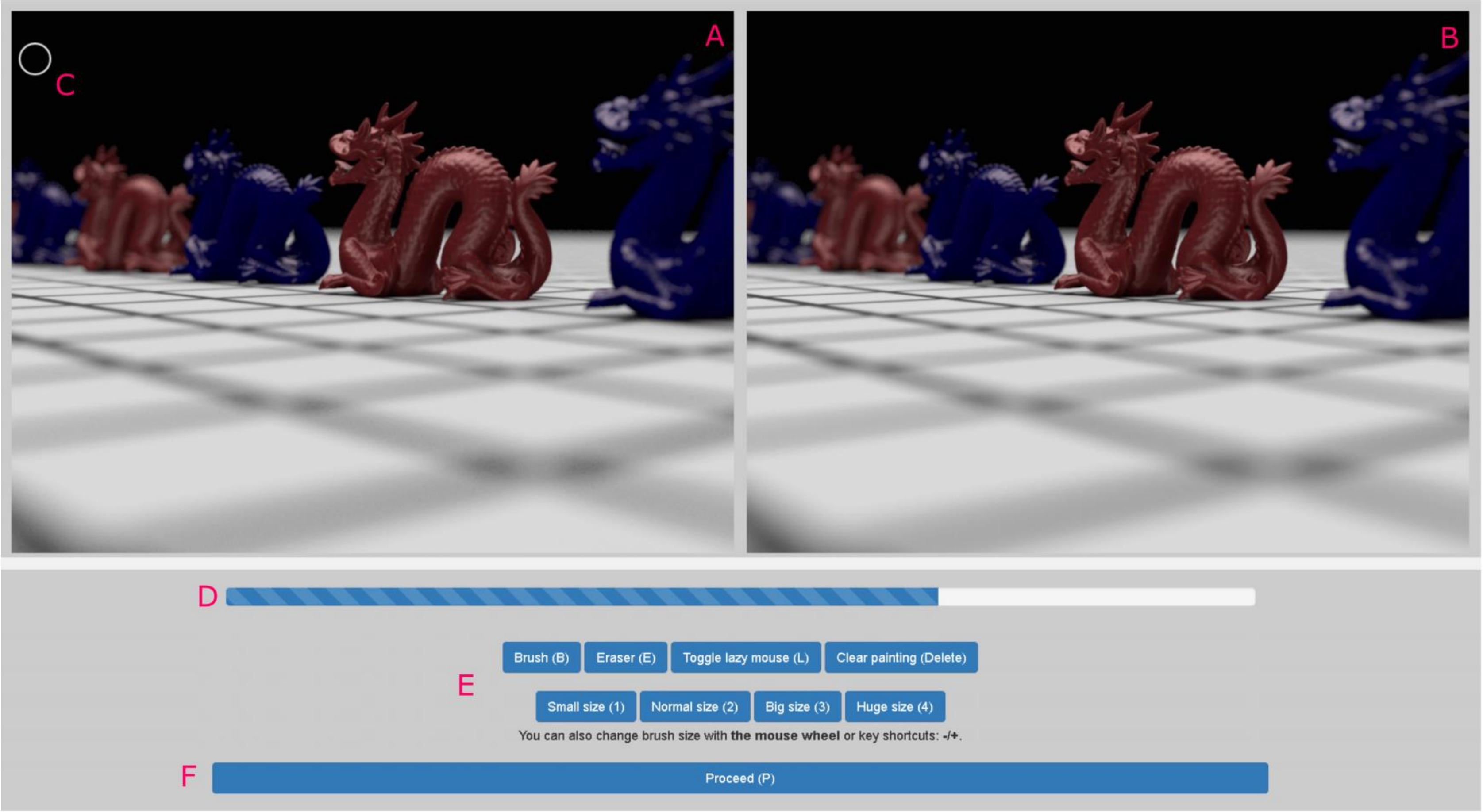


Dataset of visible distortions

Dataset covers some standard distortions (i.e. noise, blur, compression artifact) and specialized computer graphics artifacts (e.g. Peter panning, shadow acne, z-fighting, etc.).

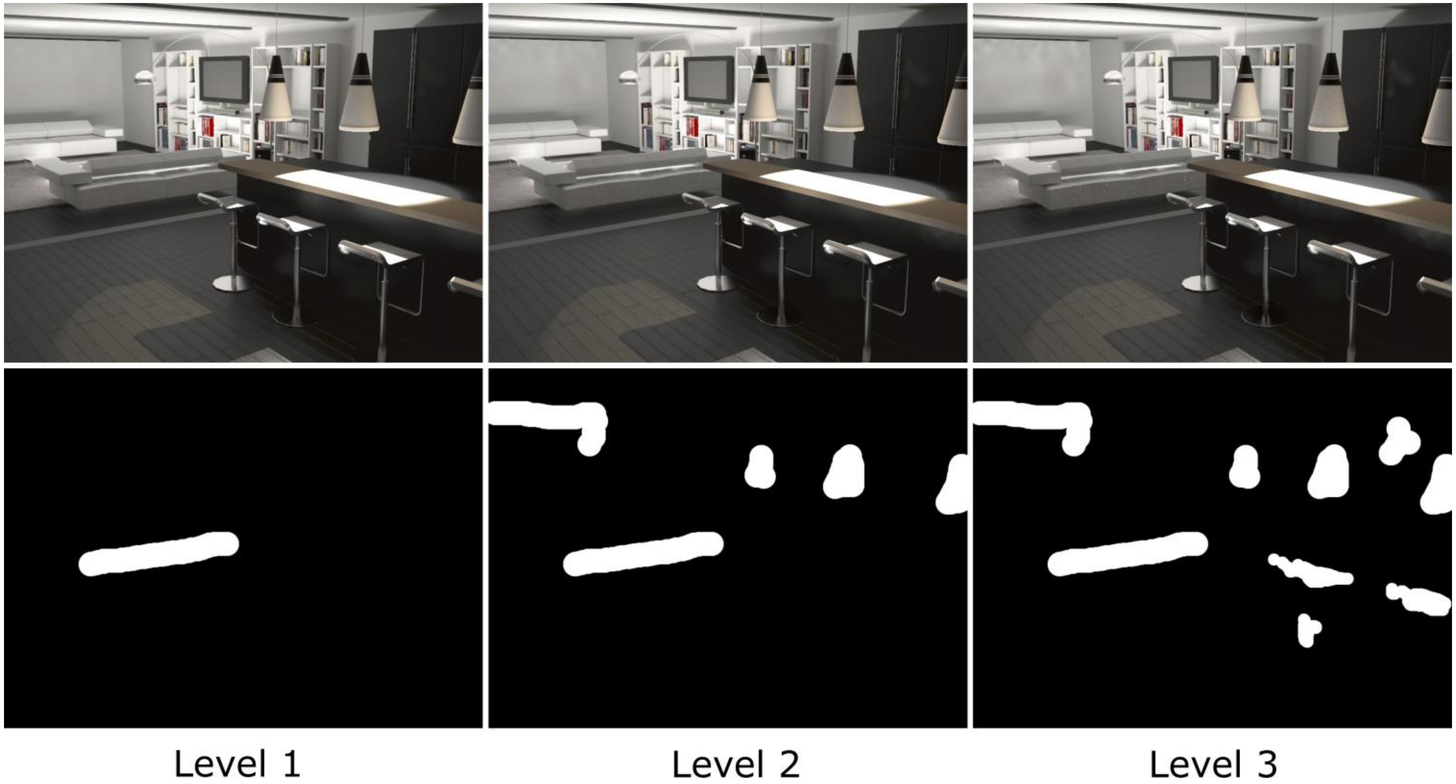
Data collection

For data collection purpose custom painting software was used. Approach is similar to the previous one, but...

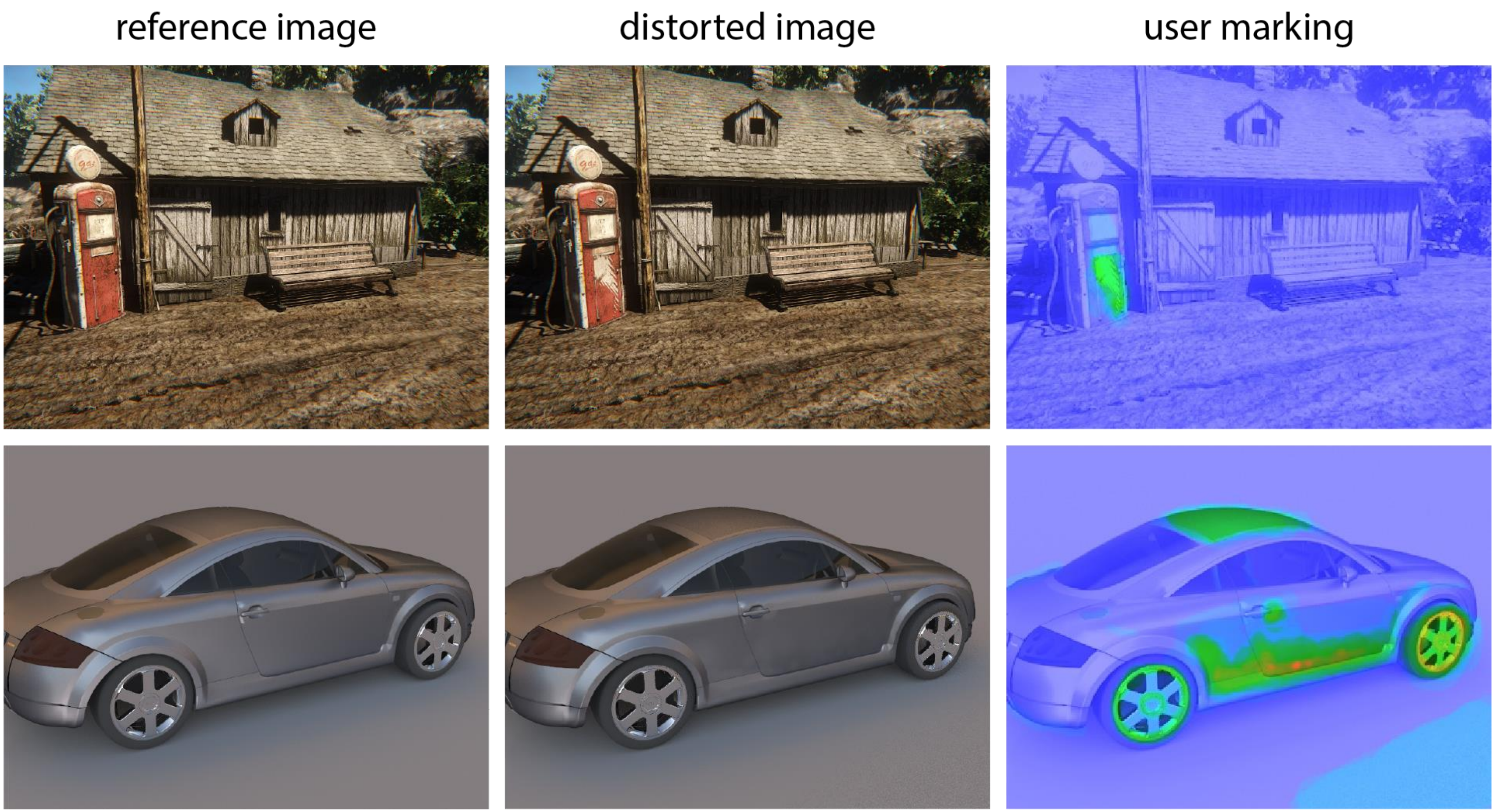


Data collection

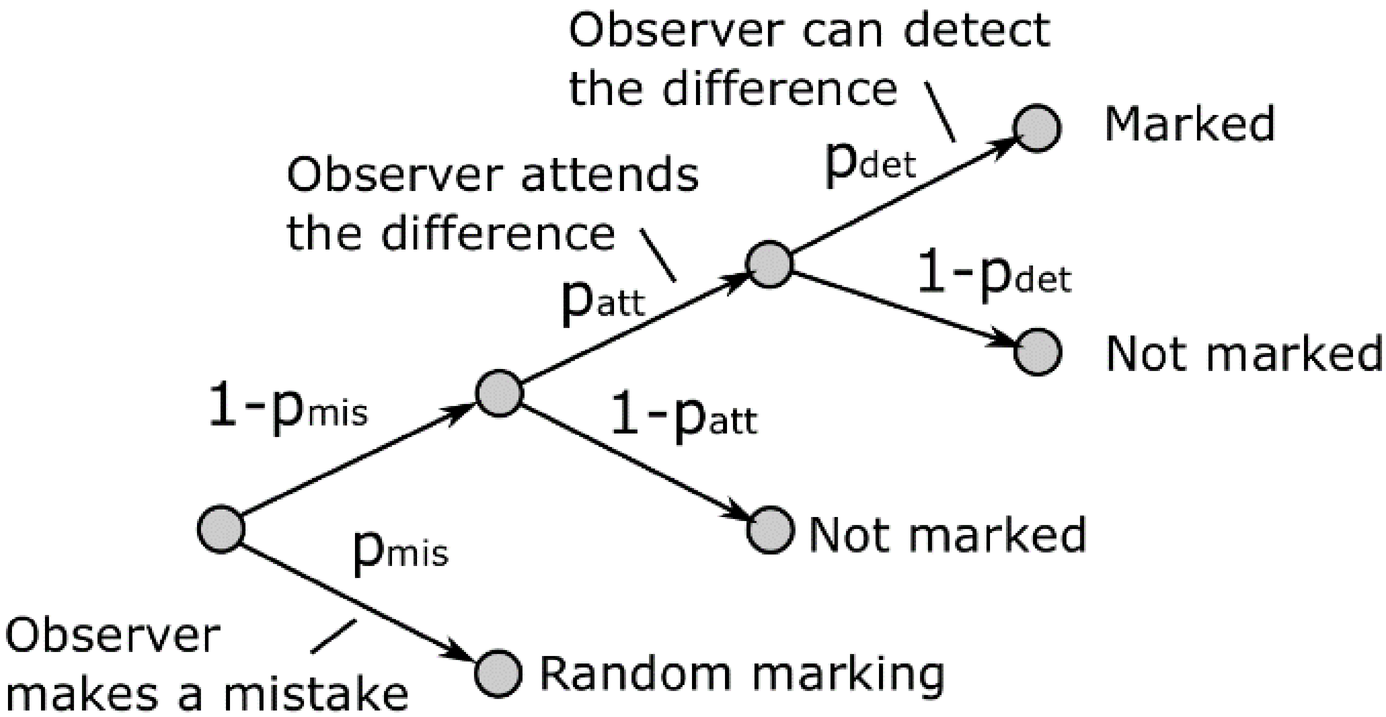
...more efficient way of gathering data was proposed.
For each scene from 1 to 3 levels of distortion magnitude were prepared. Each level had stronger distortions and for each level users painted only newly visible distortions.



Shall we trust the observers?

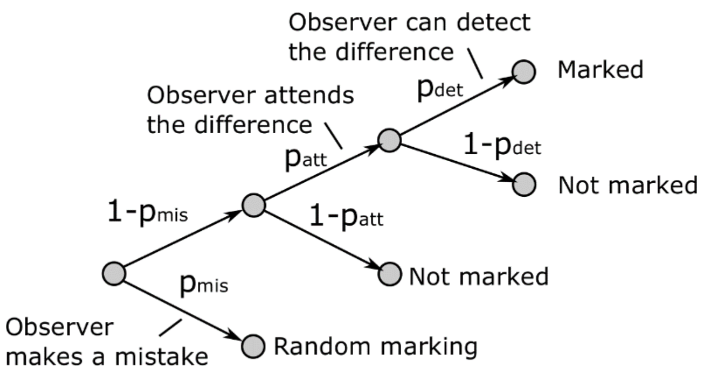


Modelling the data



$$P(data) = p_{mis} + (1 - p_{mis})\binom{N}{k}(p_{att} \cdot p_{det})^k (1 - p_{att} \cdot p_{det})^{n-k}$$
$$= p_{mis} + (1 - p_{mis})\, Binomial(k, N, p_{att} \cdot p_{det}) .$$

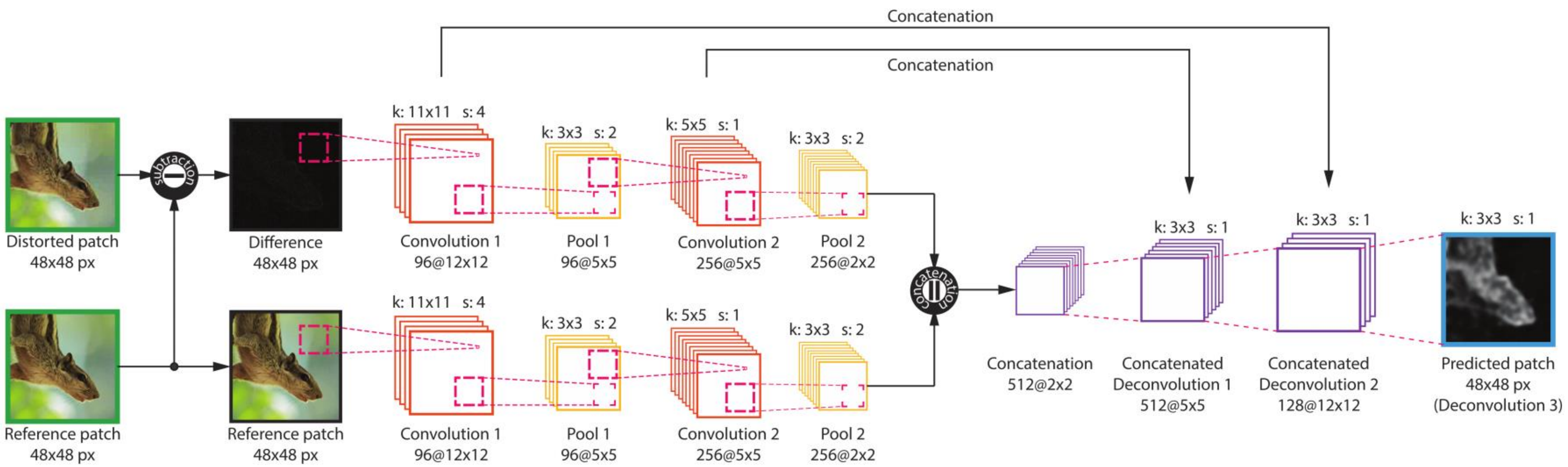
Likelihood loss function



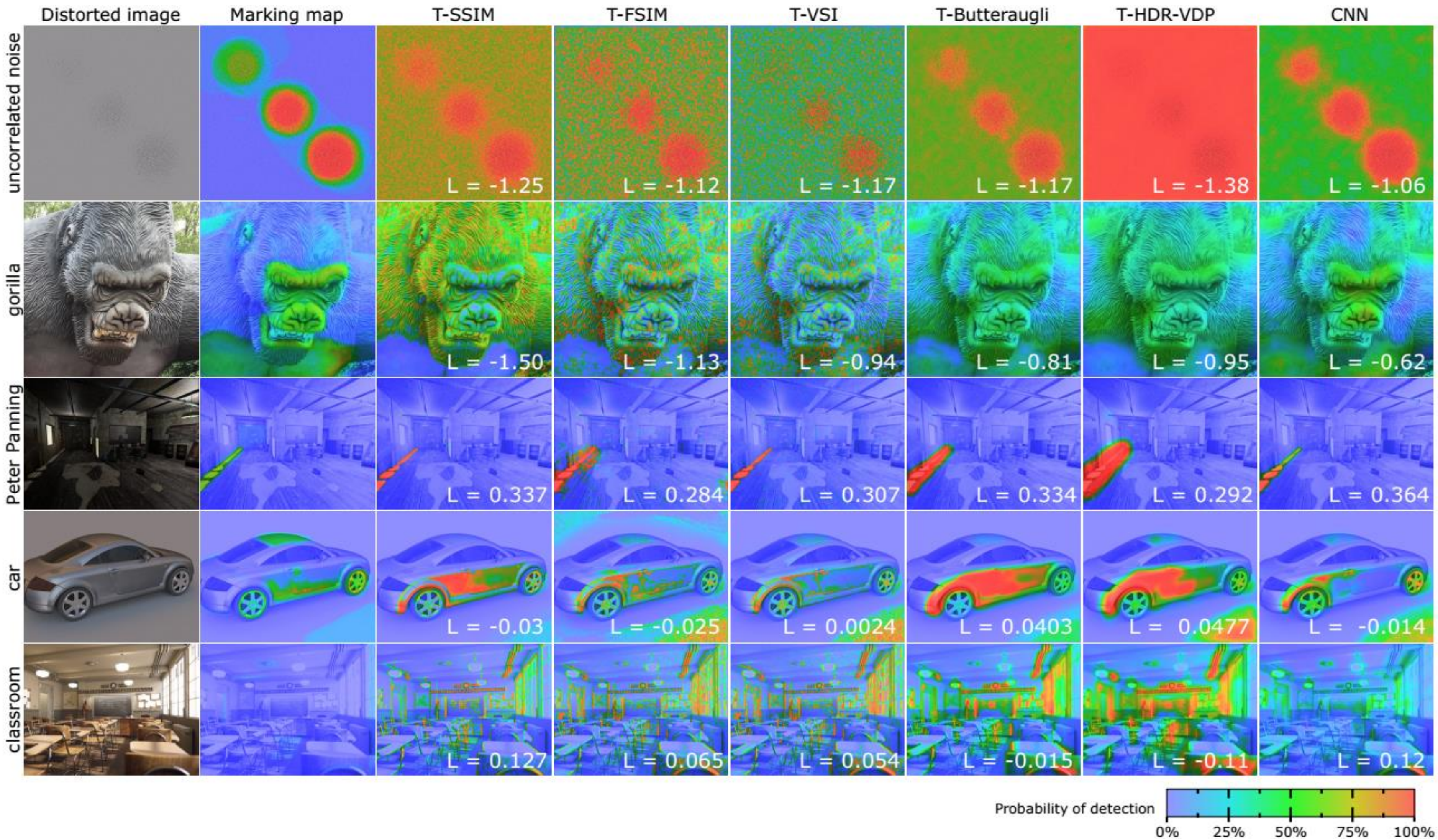
$$P(data) = p_{mis} + (1 - p_{mis})\binom{N}{k}(p_{att} \cdot p_{det})^k (1 - p_{att} \cdot p_{det})^{n-k}$$
$$= p_{mis} + (1 - p_{mis})\, Binomial(k, N, p_{att} \cdot p_{det}) \,.$$

$$L = \sum_{(x,y)\in\Theta} \log[p_{mis} + (1 - p_{mis}) \cdot \int_0^1 p_{att}(p) \cdot Binomial\,(k(x,y), N, p_{att}(p) \cdot p_{det}(x,y))\; dp]$$

Neural network architecture



Results comparison



Results comparison

Image Metric	Pear. correl	Spear. correl	RMSE	Likelihood
T-ABS	0.587	0.507	0.288	-0.26
T-CIEDE2000	0.609	0.499	0.283	-0.263
T-sCIELab	0.749	0.595	0.237	-0.196
T-SSIM	0.607	0.534	0.296	-0.261
T-FSIM	0.773	0.627	0.239	-0.158
T-VSI	0.782	0.627	0.231	-0.166
T-Butteraugli	0.799	0.653	0.227	-0.124
T-HDR-VDP	0.802	0.666	0.245	-0.111
CNN	0.92	0.755	0.145	-0.0566