A Signal Processing Approach to Modeling Low-level Vision, and Applications

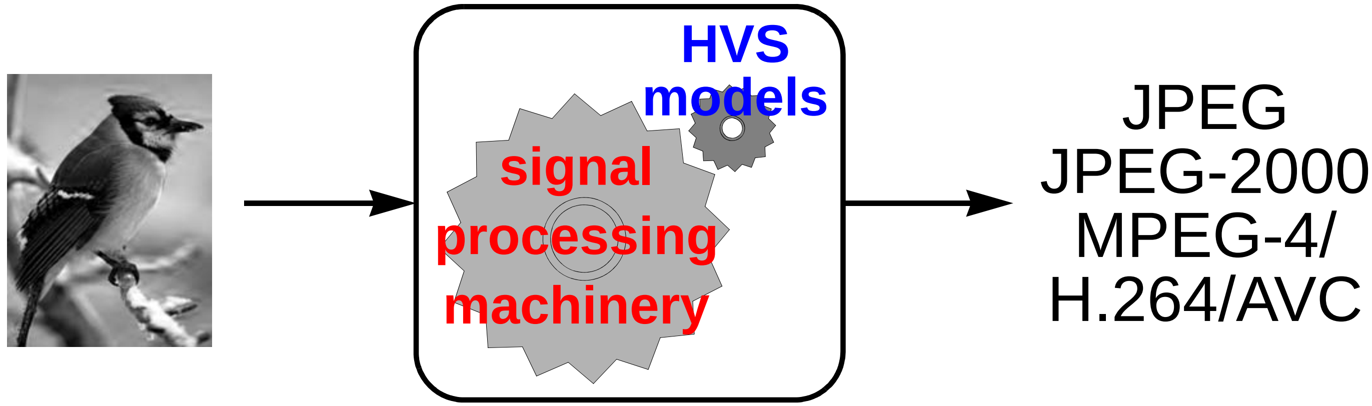
Sheila S. Hemami

Department of Electrical & Computer Engineering
Northeastern University

Visual Communications Lab

Representation & Transmission of Visual Information

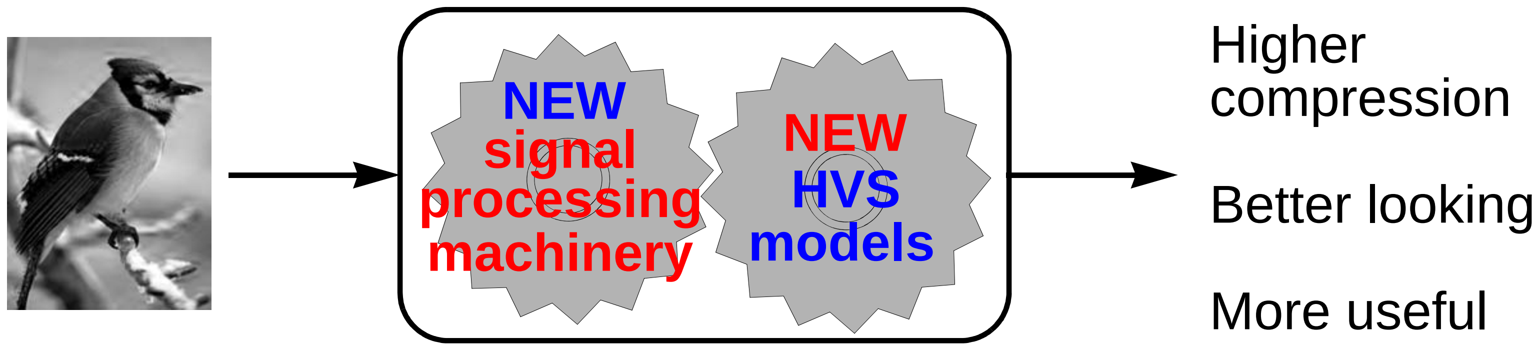
- A model for current state-of-the-art techniques:



- This model works well when *transmission resources are not limited* (bandwidth, QoS, etc.).

- When resources become scarce, *every bit counts*. The SP machinery starts to break at low rates.

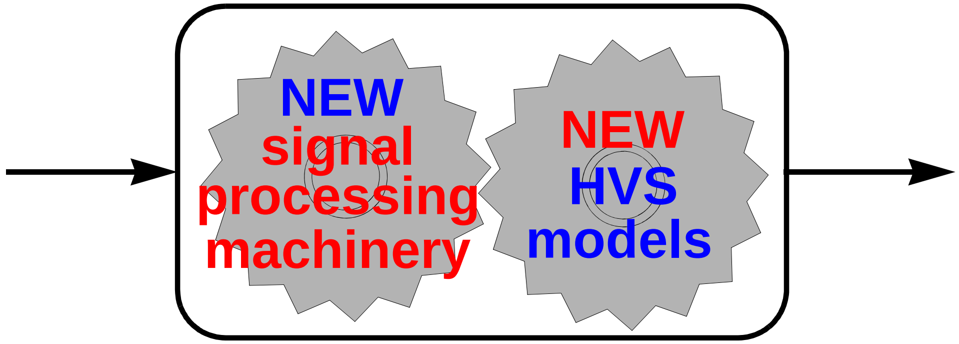
Our Dual Approach: Understand How We See, and Develop SP to Exploit this Understanding



- Develop *more appropriate HVS models* suitable for image applications via strategic psychophysical experimentation.

- Develop signal processing theory and practice to exploit this HVS characterization.

Our Dual Approach: Understand How We See, and Develop SP to Exploit this Understanding



Higher compression

Better looking

More useful

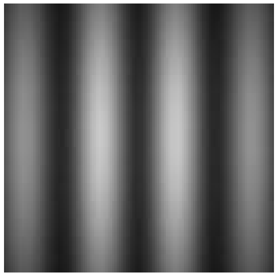
- Develop *more appropriate HVS models* suitable for image applications via strategic psychophysical experimentation.
- Develop signal processing theory and practice to exploit this HVS characterization.
- End goal: an image processing system incorporating a model which exhibits better performance than if the model is not used.

Outline

- Three “classical” psychophysics results/HVS characterizations.
- Wavelets, the multichannel model, and images.
- Characterizing the HVS using natural images.
- Some SP strategies and applications to compression which exploit our characterization.
- Current work: image utility.

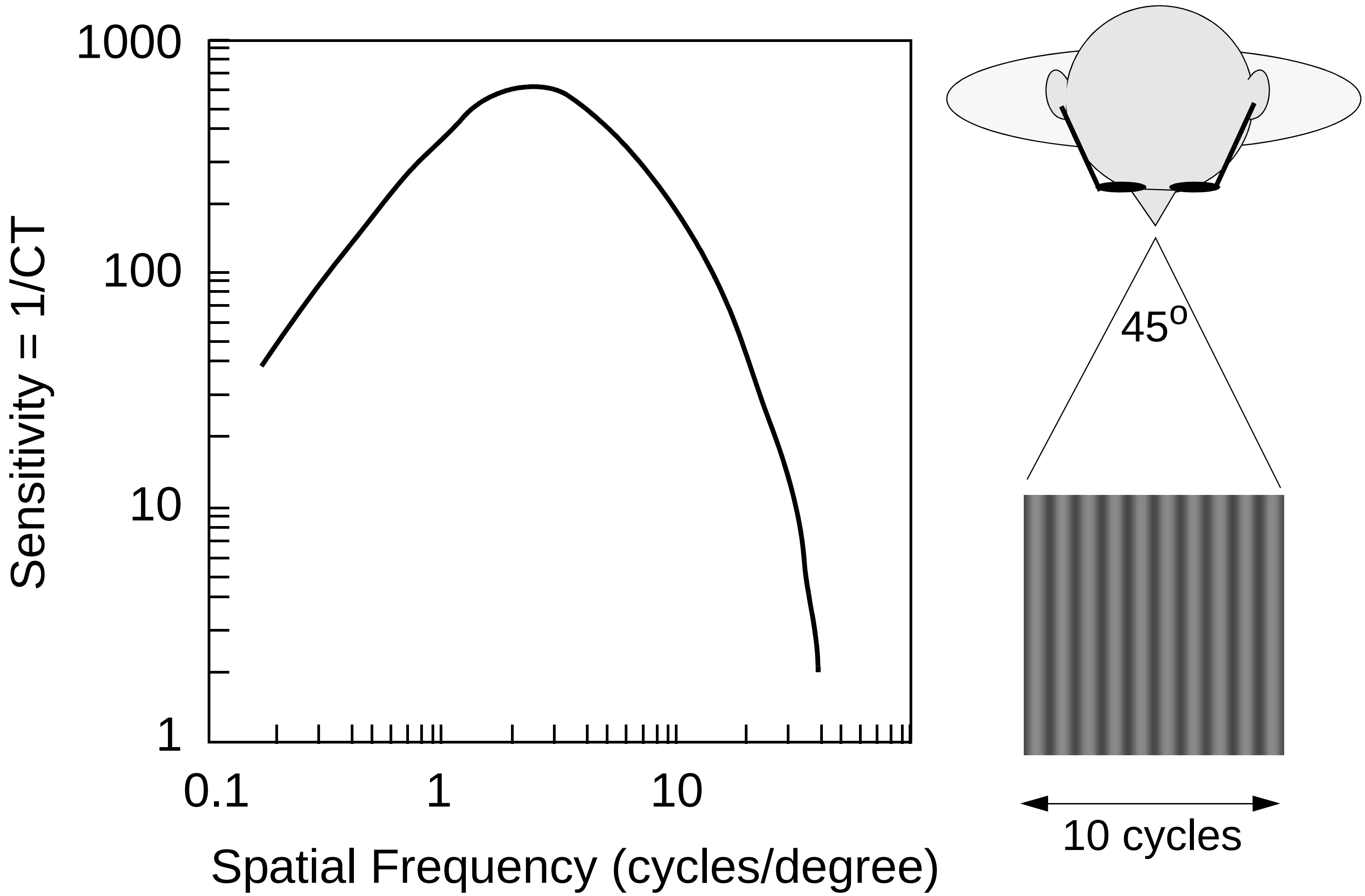
Three Classical Psychophysical Results (V1)

Experiments with sinusoidal gratings yield the following:

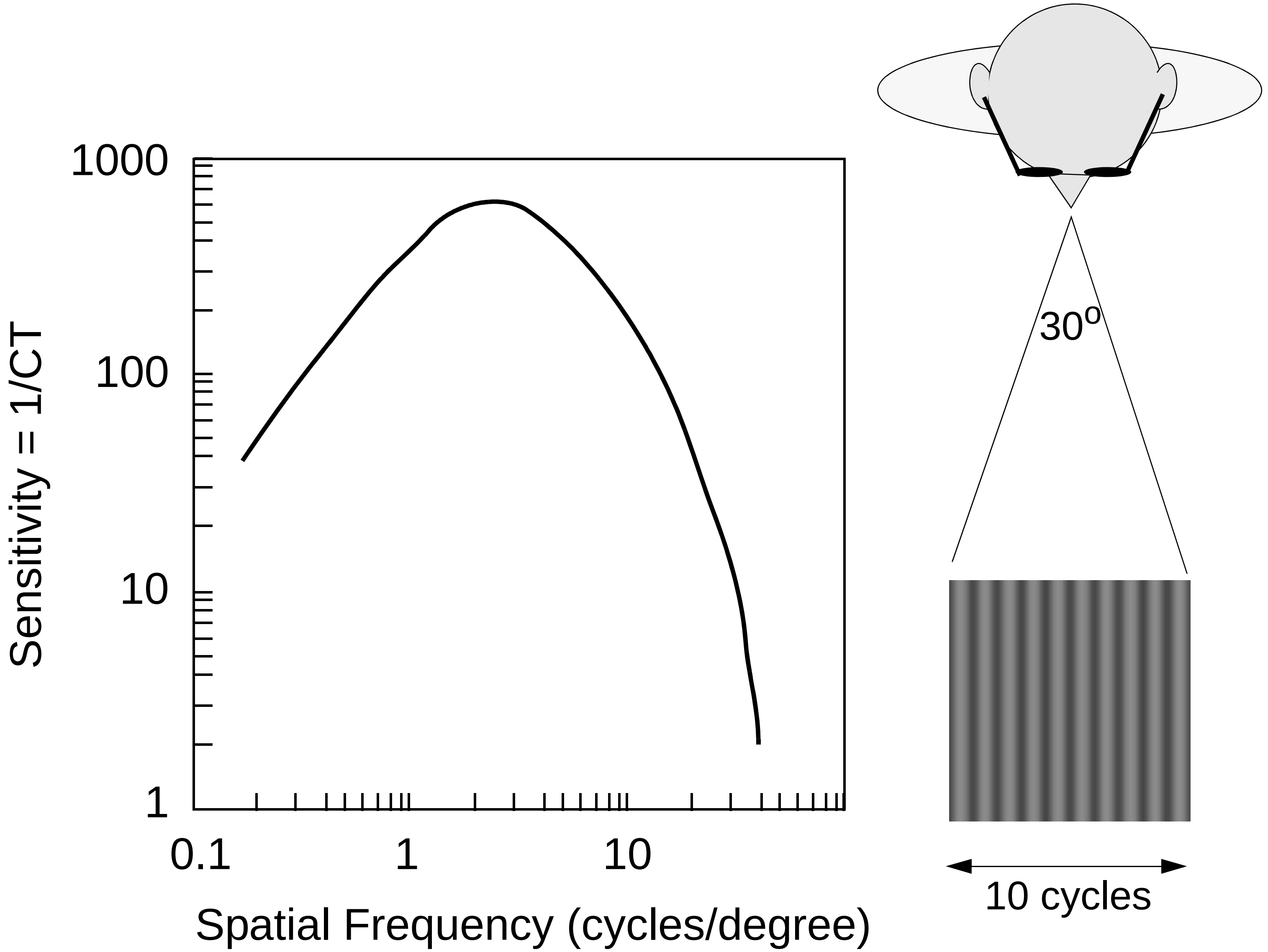


1. The human *contrast sensitivity function* (CSF)

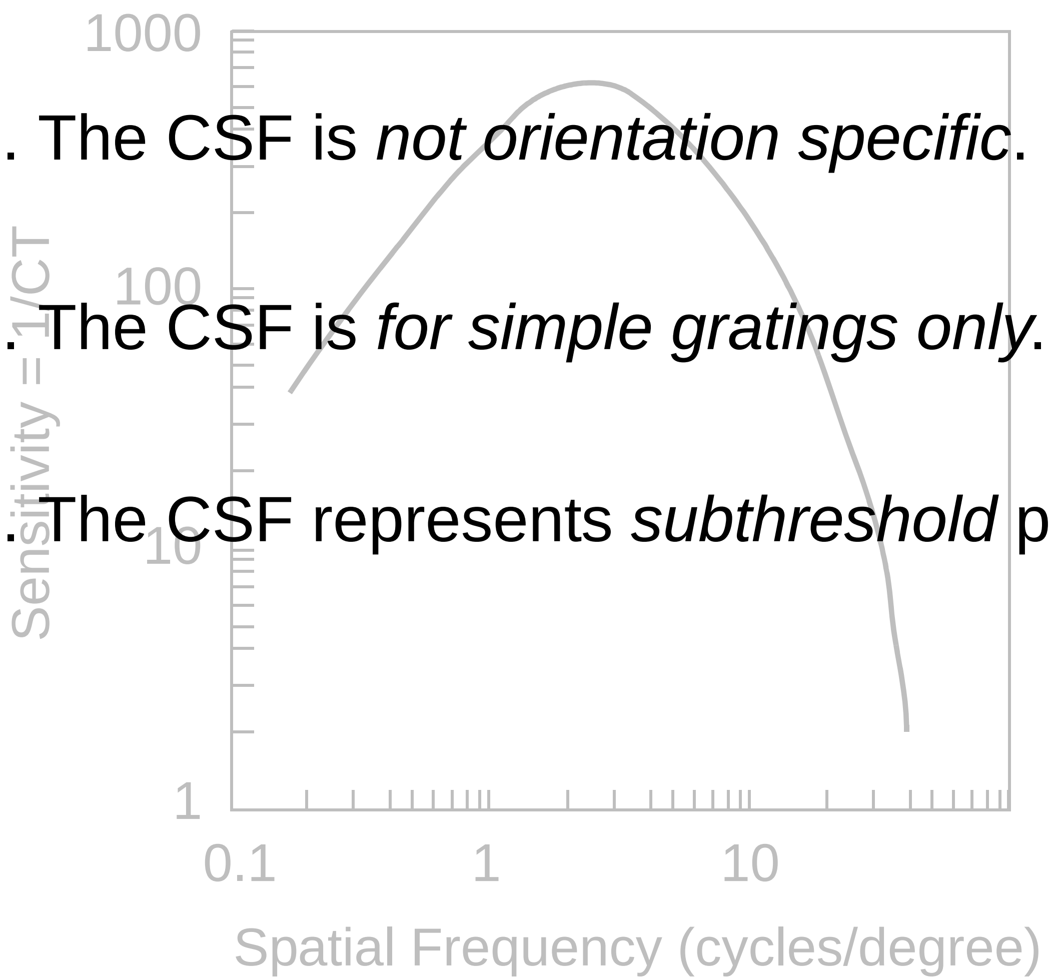
Human Contrast Sensitivity Function (CSF)



Human Contrast Sensitivity Function (CSF)



Some Comments on the CSF



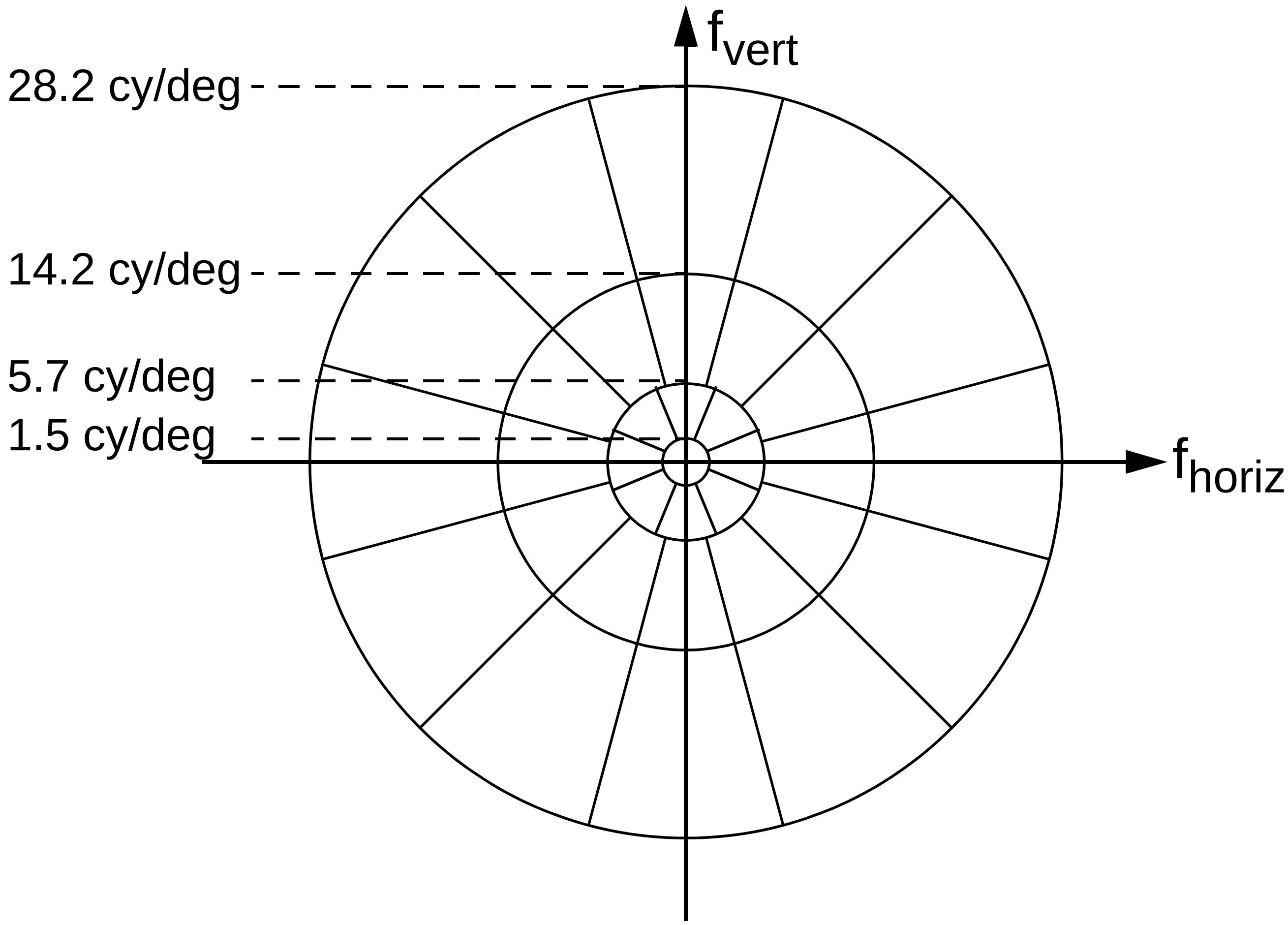
1. The CSF is *not orientation specific*.

2. The CSF is *for simple gratings only*.

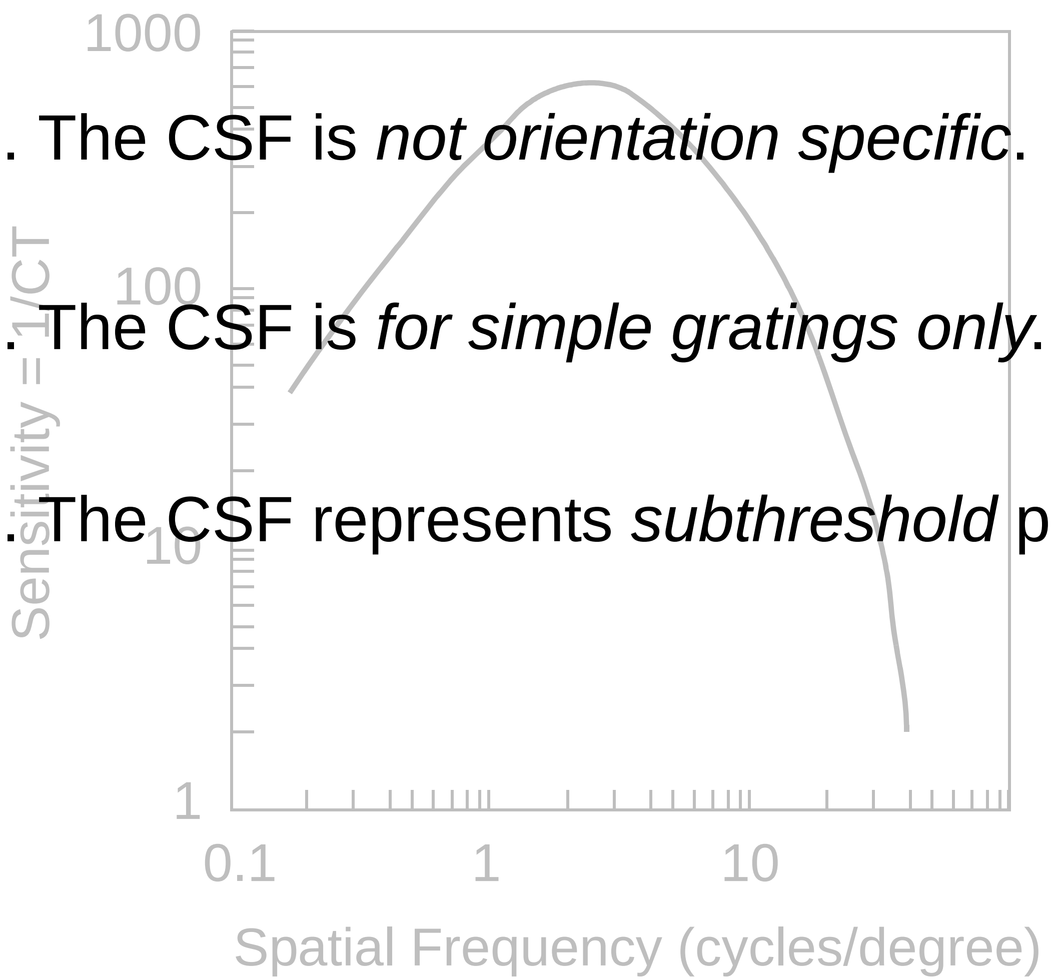
3. The CSF represents *subthreshold* perception.

Multi-Channel Model of the HVS

The HVS consists of *channels*, each tuned to range of spatial frequencies and orientations.



Some Comments on the CSF

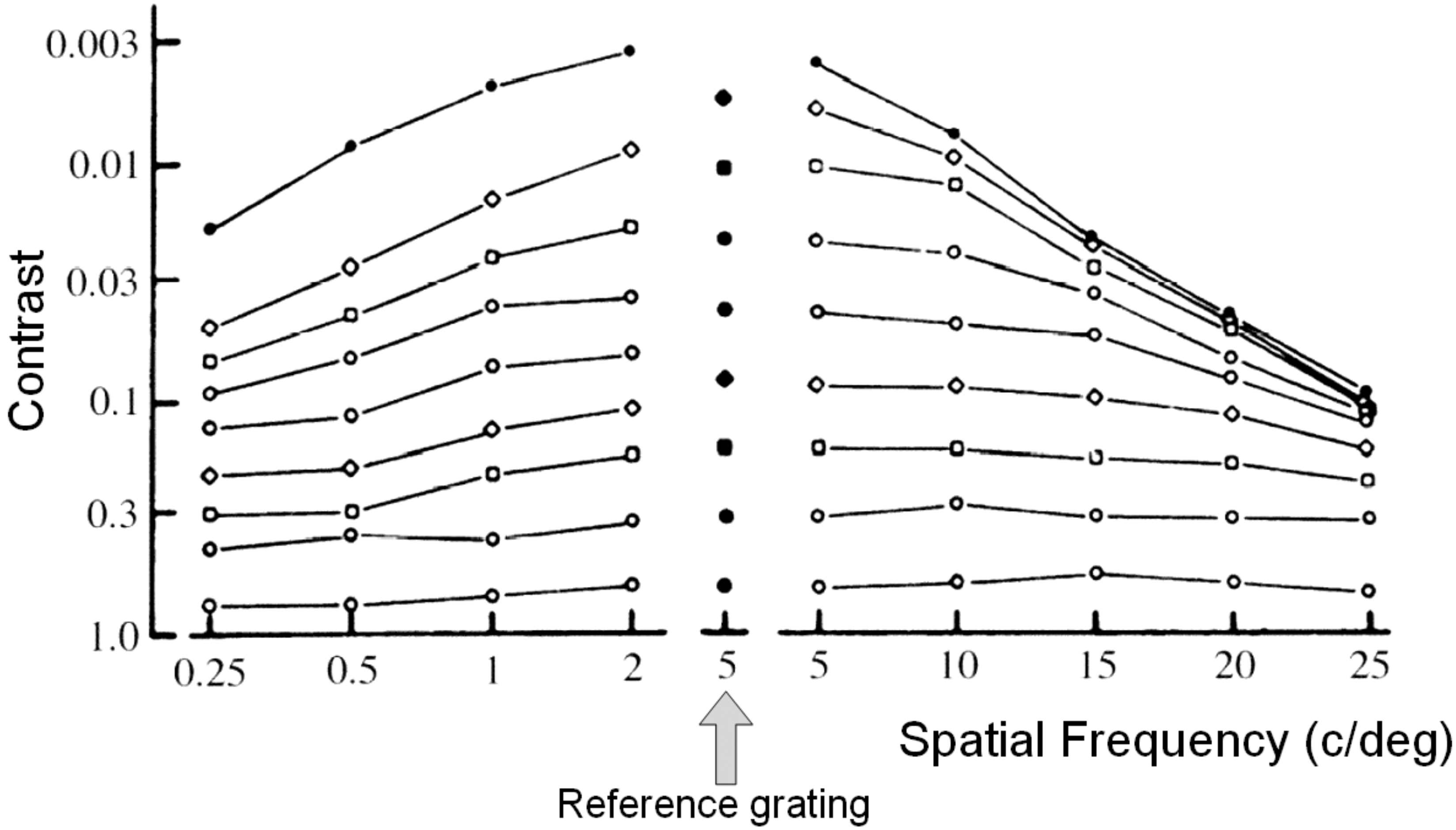


1. The CSF is *not orientation specific*.

2. The CSF is *for simple gratings only*.

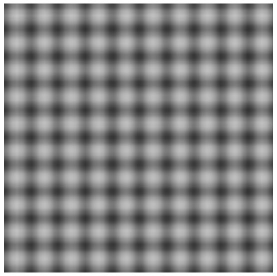
3. The CSF represents *subthreshold* perception.

Suprathreshold VTs — Contrast Constancy



Two gratings at different frequencies have equal perceived contrast at equal physical contrast as they become increasingly suprathreshold.

Three Classical Psychophysical Results



Experiments with sinusoidal gratings

yield the following:

1. The human *contrast sensitivity function* (CSF) — the HVS has a low-pass response at the detection threshold, becoming flat as gratings become more visible.

2. *Summation*

Summation — How we see multiple components

If this stimulus has contrast threshold

CT_A

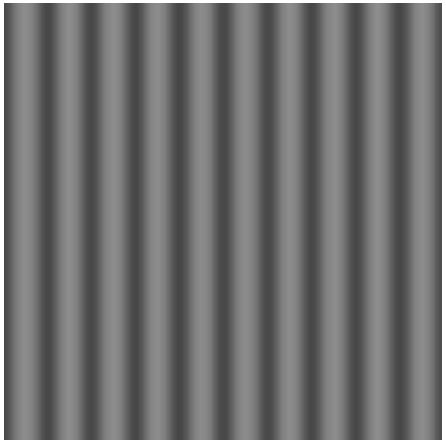
...and this stimulus has contrast threshold

CT_B

Then what is the contrast threshold of this stimulus?

$CT_{A+B} = ?$

Summation — How we see multiple components

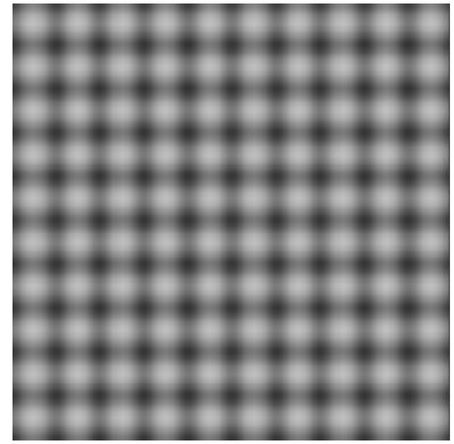


If this stimulus has contrast threshold

$$CT_A$$

...and this stimulus has contrast threshold

$$CT_B$$



Then what is the contrast threshold of this stimulus?

$$CT_{A+B} = ?$$

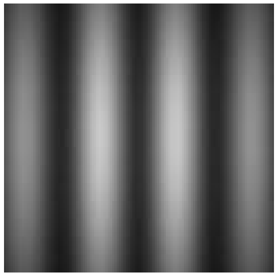
- For the compound stimuli to be as detectable as either of the individual components,

$$(C_A / CT_A)^{\beta} + (C_B / CT_B)^{\beta} = 1$$

- For sinusoidal components, $\beta \in [2, 4]$.

Three Classical Psychophysical Results

Experiments with sinusoidal gratings yield the following:



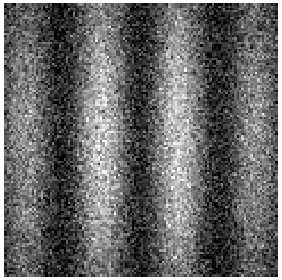
1. The human *contrast sensitivity function* (CSF) — the HVS has a low-pass response at the detection threshold, becoming flat as gratings become more visible.

2. *Summation* — The contrast threshold for a given sinusoid is 40% lower when it is shown simultaneously with another, different sinusoid.

Three Classical Psychophysical Results

Experiments with sinusoidal gratings

yield the following:

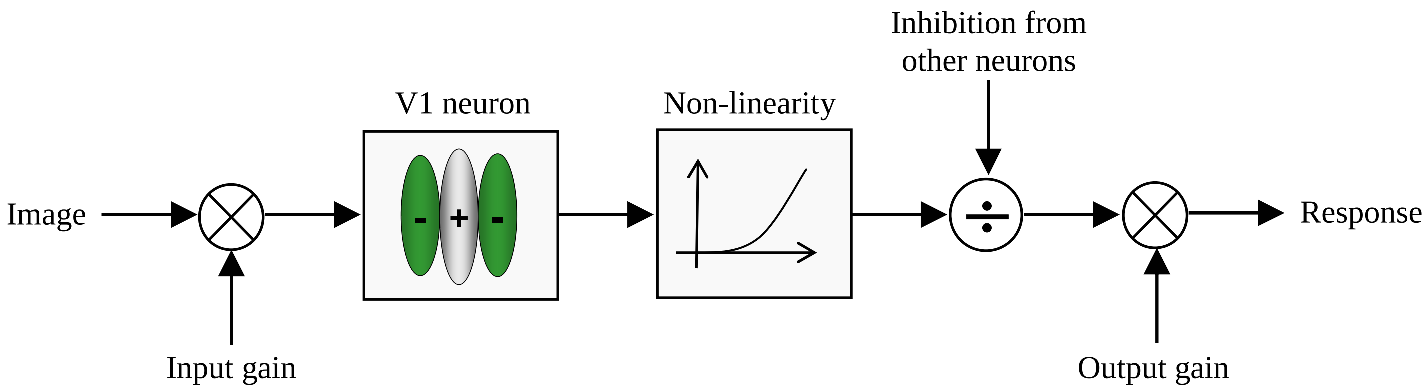


1. The human *contrast sensitivity function* (CSF) — the HVS has a low-pass response at the detection threshold, becoming flat as gratings become more visible.

2. *Summation* — The contrast threshold for a given sinusoid is 40% lower when it is shown simultaneously with another, different sinusoid.

3. The standard gain control model for *masking* describes how thresholds are impacted based on surrounding image content.

Standard Gain Control Model (Masking)



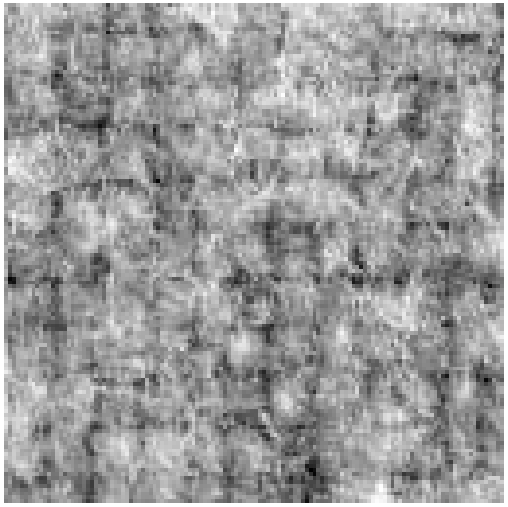
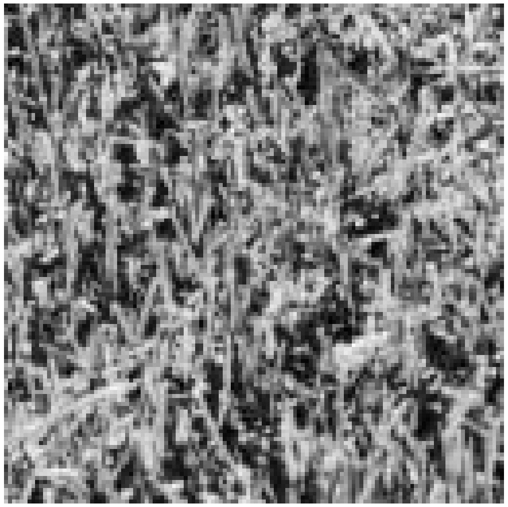
Neural response:

$$r(x, f, \theta) = \frac{w(x, f, \theta)^p}{b^q + \sum_{(x, f, \theta) \in S} w(x, f, \theta)^q}$$

w = e.g., wavelet coefficient at location x, frequency f, orientation θ

Usually $p \approx 2$, $q \approx 2$ — effectively variance!

Standard Gain Control Applied to *Textures*

The standard visual masking model predicts the masking elevations well *for homogeneous textures*.

The Signal Processor’s Question

Should the 3 classical psychophysical results, based on sinusoidal gratings be directly applied to processing images?

- Images are the superposition of many sinusoidal components.
- Images provide a very sophisticated “mask” to any distortions introduced by compression.
- Arbitrary image patches are not necessarily homogeneous textures.
- [Images have higher-level meaning to observers.]

The Short Answer

Should the 3 classical psychophysical results, based on sinusoidal gratings be directly applied to processing images?

NO

Questions

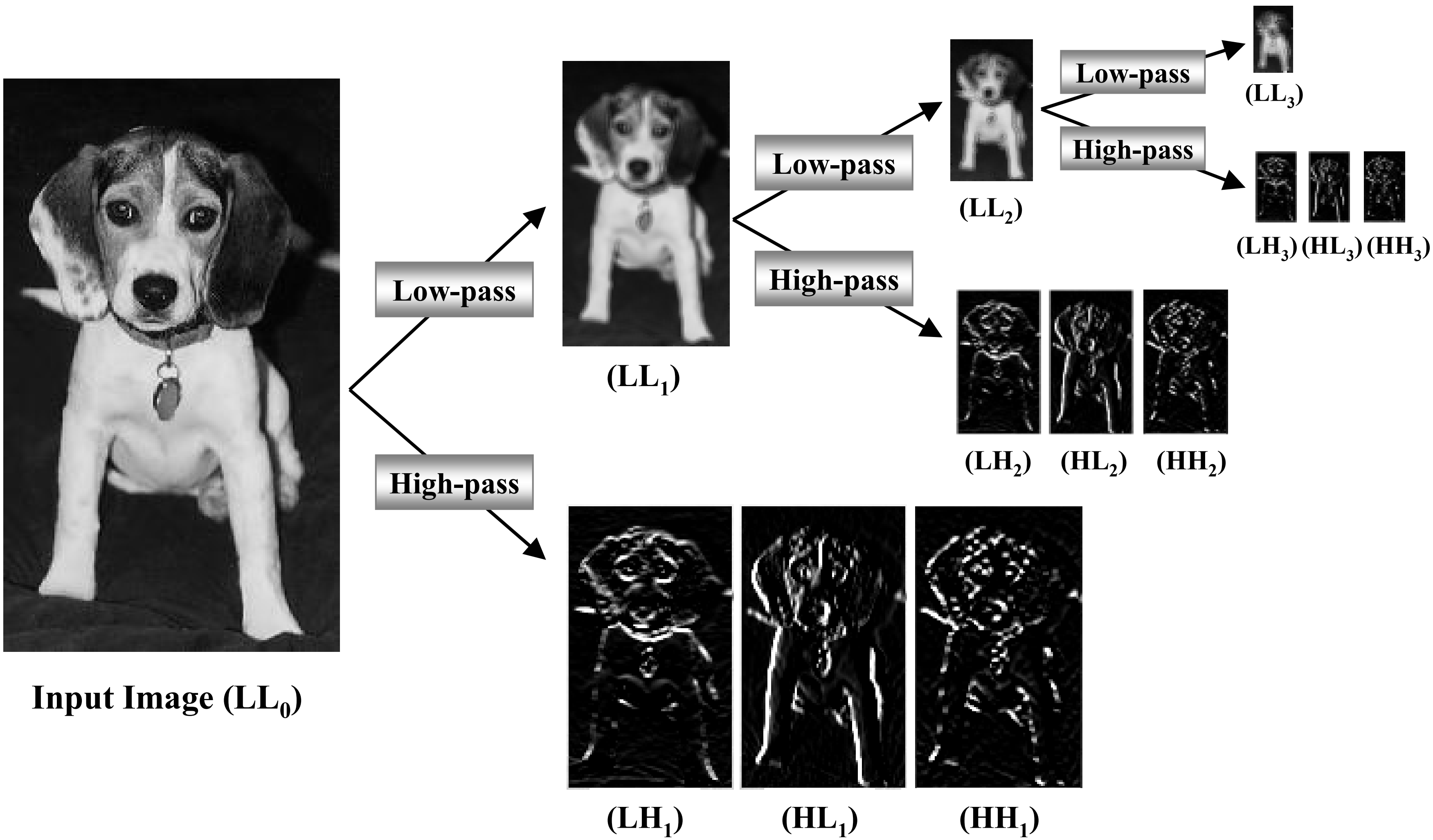
Using realistic maskers (images) and realistic stimuli (bandlimited, correlated quantization noise)...

- What are the visibility thresholds for *quantization distortions as occur in natural images*? (CSF without and with masking)
- How are distortions from multiple quantized subbands perceived? (Summation)
- Can we predict visibility thresholds from *local natural image characteristics*? (Masking)
- [How should higher-level processing (i.e., the task) impact any necessary signal processing?]

Outline

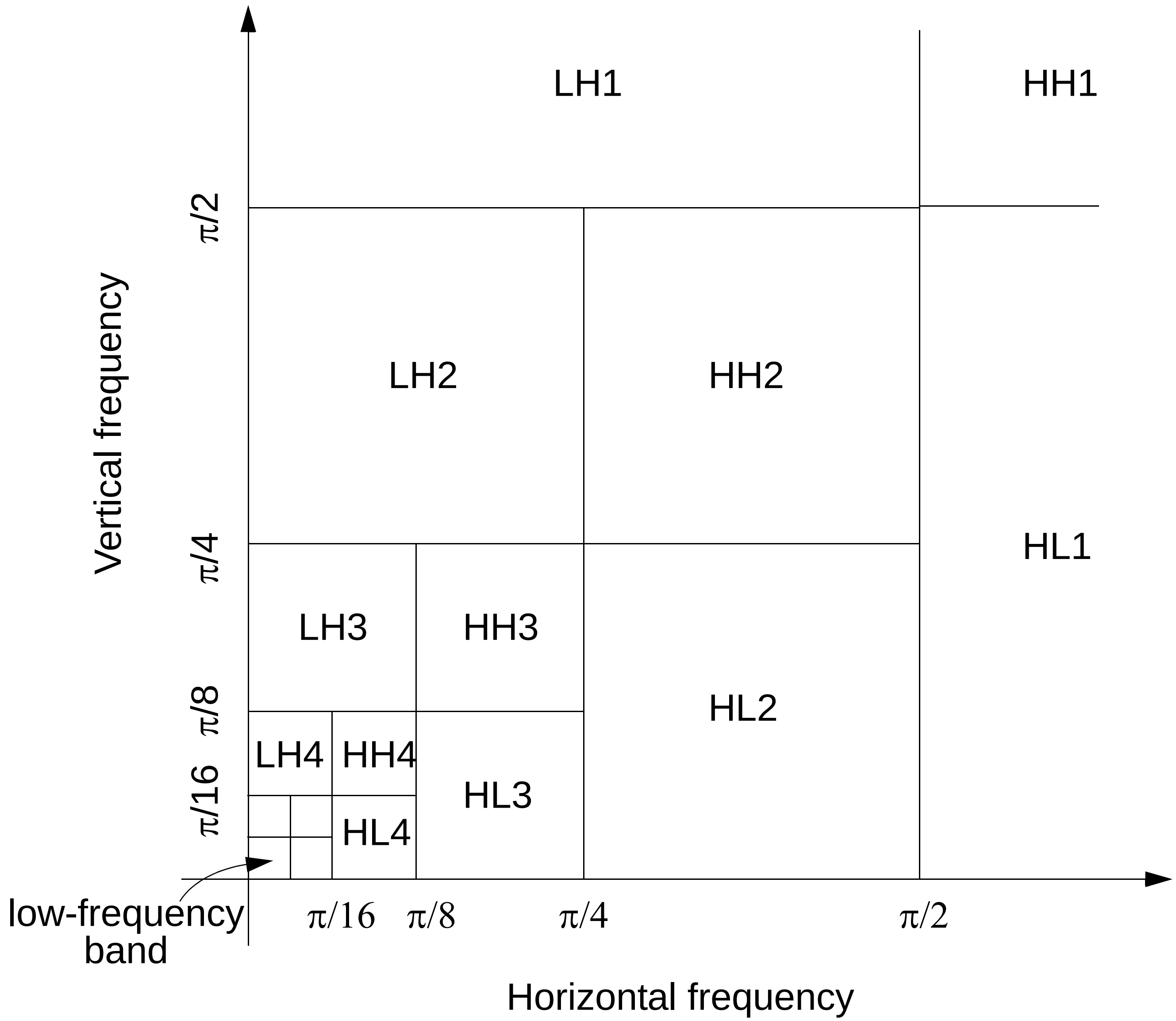
- Three “classical” psychophysics results/HVS characterizations.
- Our image coding framework: wavelets, the multichannel model, and digital images.
- Characterizing the HVS using natural images.
- Some SP strategies and applications to compression which exploit our characterization.

2-D Wavelet Transform



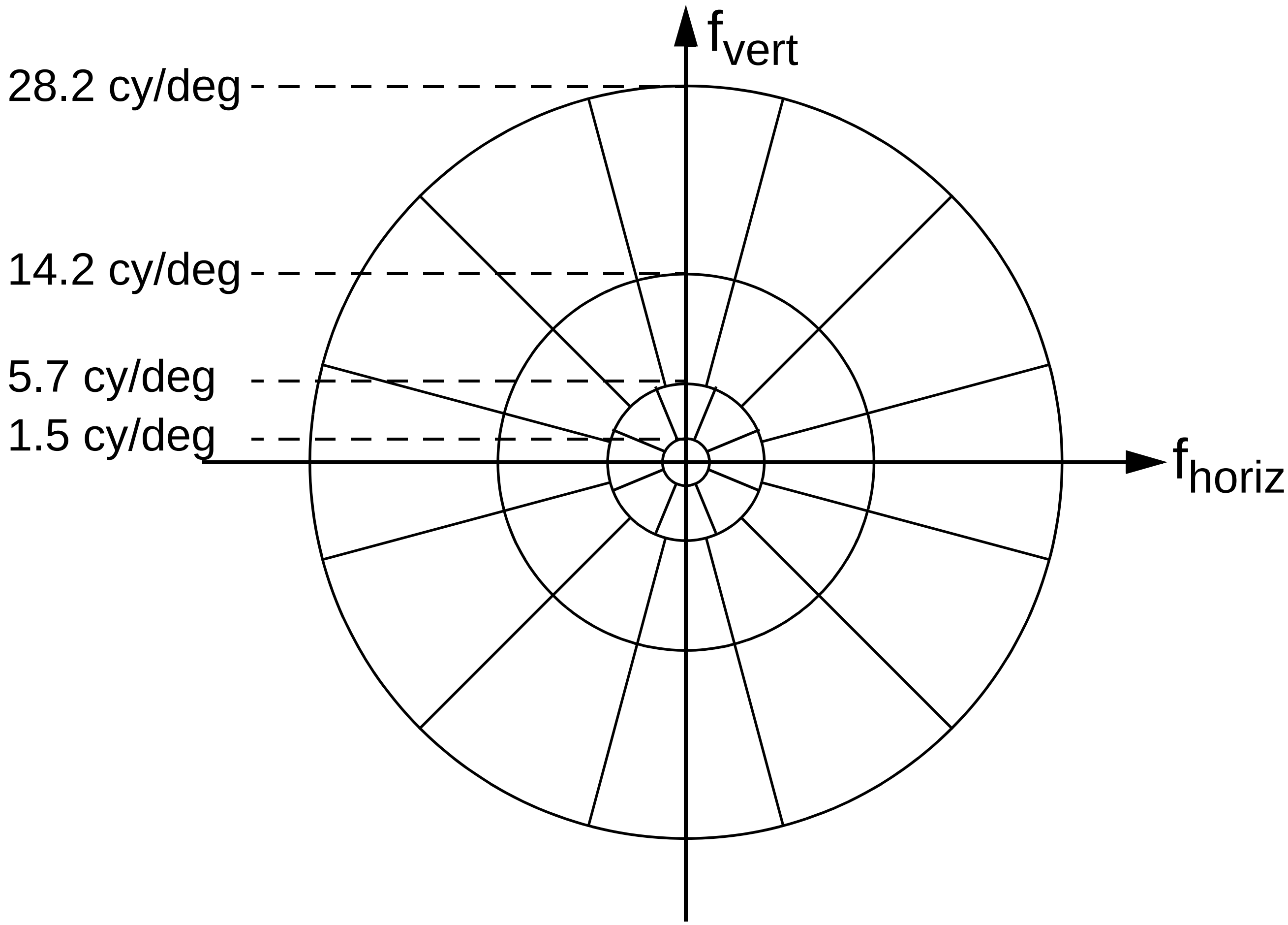
Input Image (LL$_0$)

Wavelet Decomposition in Frequency Space



Multi-Channel Model of the HVS

The HVS consists of *channels*, each tuned to range of spatial frequencies and orientations.



The Digital Signal vs. What We See

- Pixel values vs. display luminance

$$L = (b + k \times p)^{\gamma}$$

displayed luminance

black-level offset

voltage-to-pixel scaling factor

pixel value

monitor's luminance-to-voltage response curve exponent

$$\text{Contrast} = \frac{\text{luminance change}}{\text{mean background luminance}}$$

- We describe stimuli in terms of *contrast*.
- For complex images, we'll use *RMS contrast*.

Contrast for Complex Images

- *RMS Contrast* defined using RMS deviation from mean background luminance L

$$C_{rms} = \frac{1}{\bar{L}}\sqrt{\frac{1}{N}\sum(L_i - \bar{L})^2}$$

- Recall $L = (b + k \times p)^{\gamma}$. For typical values of b, k, γ, this can be linearized via a Taylor series, and

$$C_{rms}^2 = \xi^2 D$$

where D is the variance of the pixels, and

$$\xi = \frac{L}{k\gamma}(b + k\bar{p})^{1-\gamma}$$

Contrast of Distorted Images

original

-

quantized

=

quantization noise

For the quantization noise, $contrast = \frac{1}{L}\sqrt{\frac{1}{N}\sum L_i^2}$

Note that we achieve C_{max} at band discard.

Detection & Masked Detection, Simple Targets

 =  + 

quantized | original | quantization noise

Detection: stimuli = target

Masked detection:

stimuli = mask + target

Summation Stimulus, Unmasked Target

Unmasked uniform quantization noise in the HL5, LH5, and HL5 + LH5 subbands.

If this stimulus has contrast threshold

CT_{HL5}

...and this stimulus has contrast threshold

CT_{LH5}

Then what is the contrast threshold of this stimulus?

$CT_{HL5+LH5} = ?$

Summation Stimulus, Masked Target

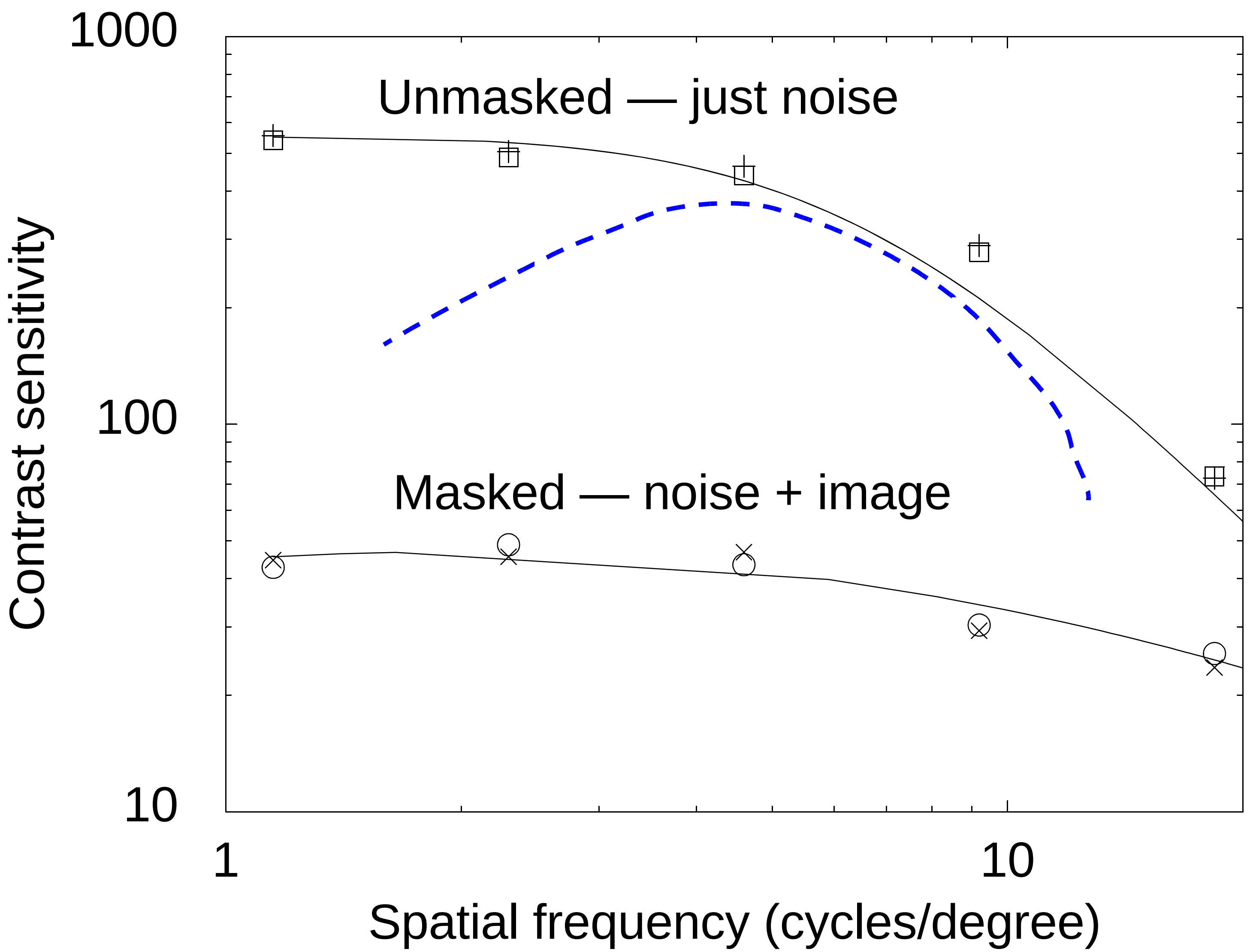
Masked uniform quantization noise in the HL5, LH5, and HL5 + LH5 subbands.

CT_{HL5}

CT_{LH5}

$CT_{HL5+LH5} = ?$

Detection of Wavelet Quantization Noise in Images (Masked CSF)



Summation in Natural Images

- For 2 subbands simultaneously quantized in an image, $1.5 < \beta < 1.8$. Let's approximate $\beta \approx 1$.

$$(C_A / CT_A)^{\beta} + (C_B / CT_B)^{\beta} = 1$$

- *Linear* summation is consistent with summation observed in "object recognition tasks." (We are moving toward cognition...)

- *This suggests that observation is content-based rather than purely target-based* — and leads us to global precedence.

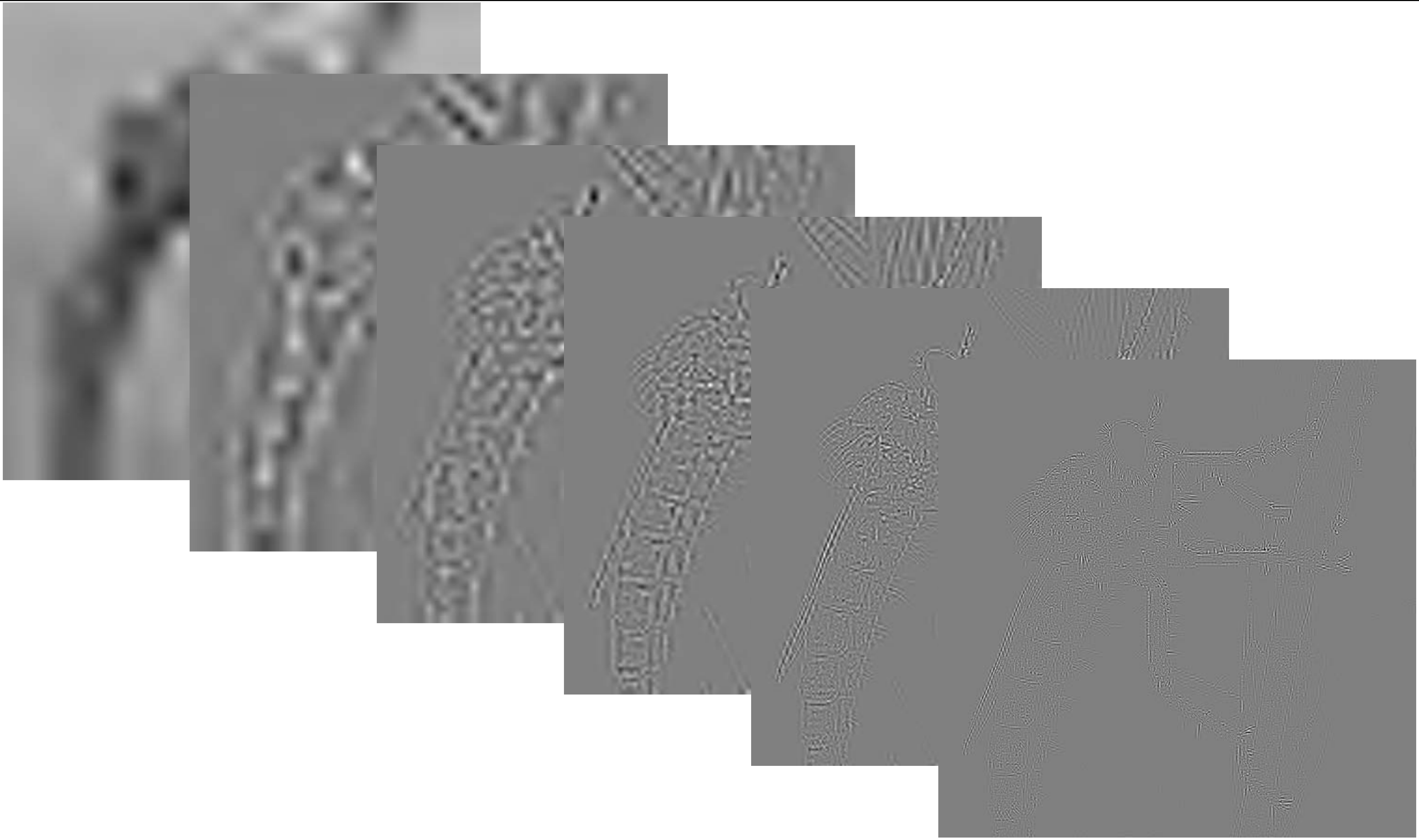
Global Precedence

Global Precedence

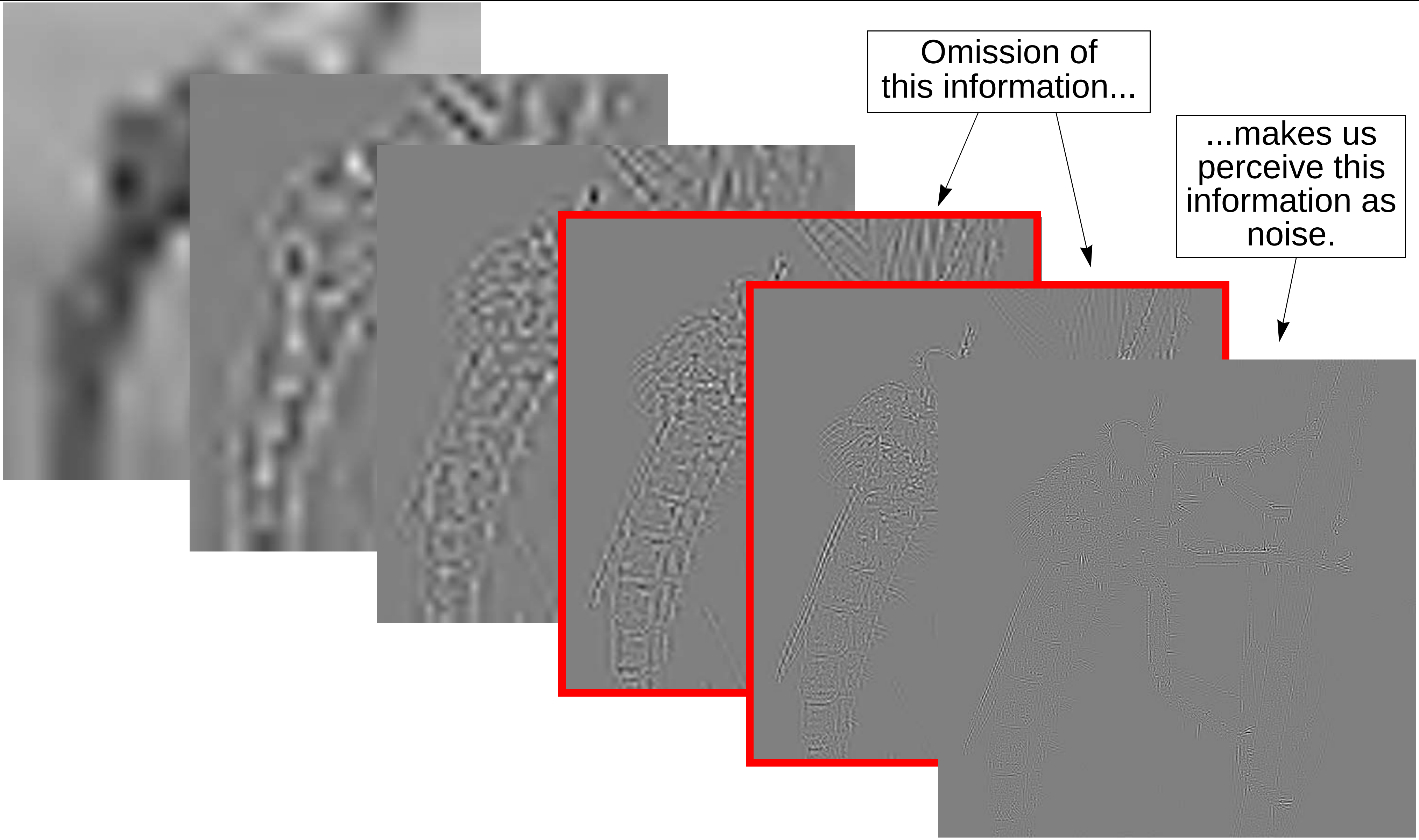
Global Precedence

Global Precedence

Global Precedence



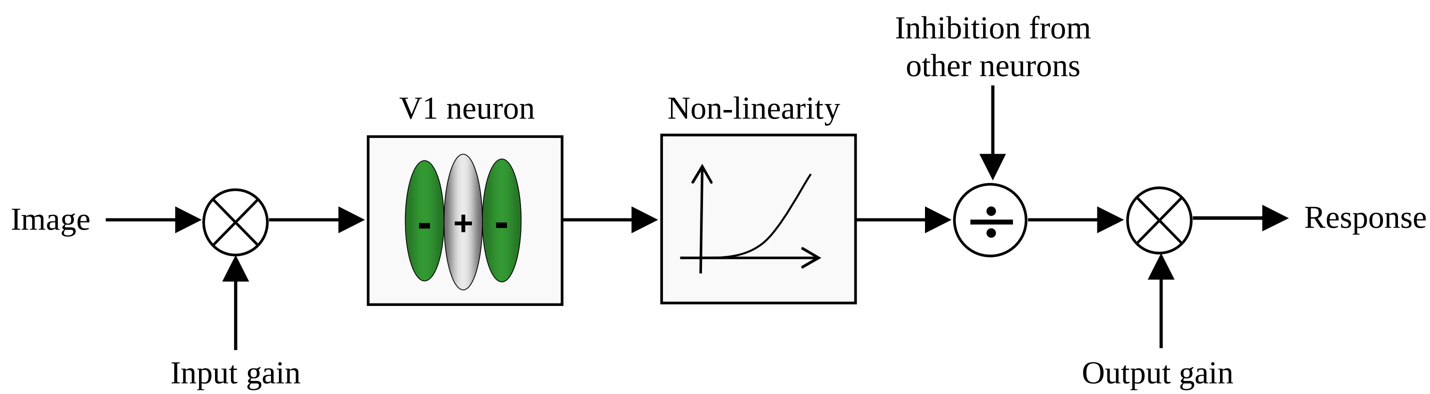
Global Precedence



Global Precedence

The *addition* of high-frequency content *visually degrades* the image.

Standard Gain Control Model (Masking)



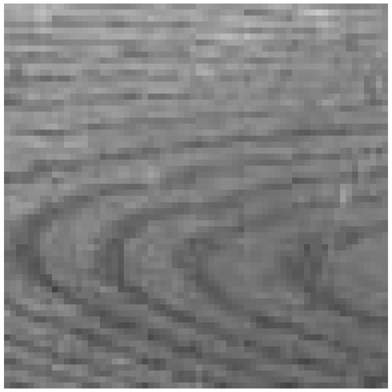
Neural response:

$$r(x, f, \theta) = \frac{w(x, f, \theta)^p}{b^q + \sum_{(x, f, \theta) \in S} w(x, f, \theta)^q}$$

w = e.g., wavelet coefficient at location x, frequency f, orientation θ

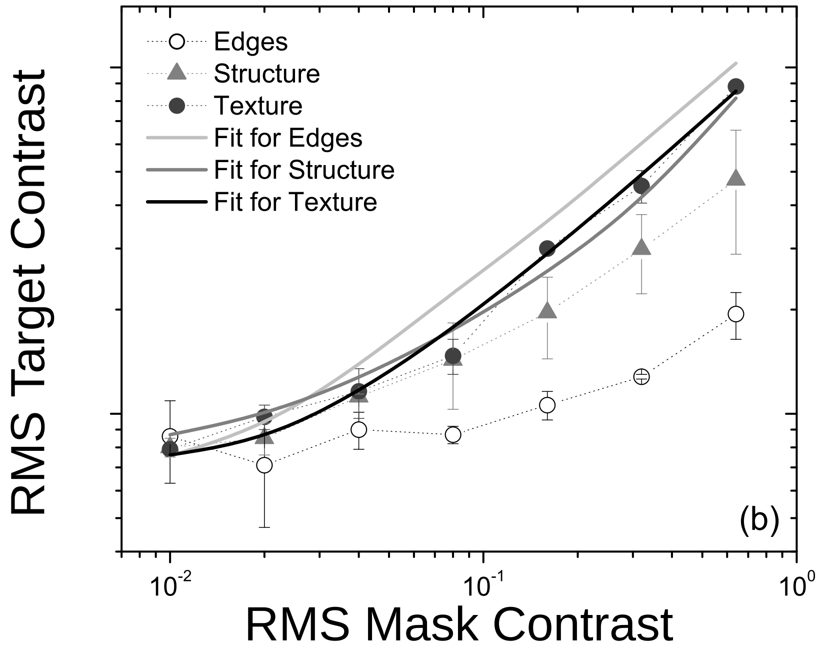
Usually $p \approx 2$, $q \approx 2$ — effectively variance!

This Model Does Not Work on Non-Homogeneous Patches

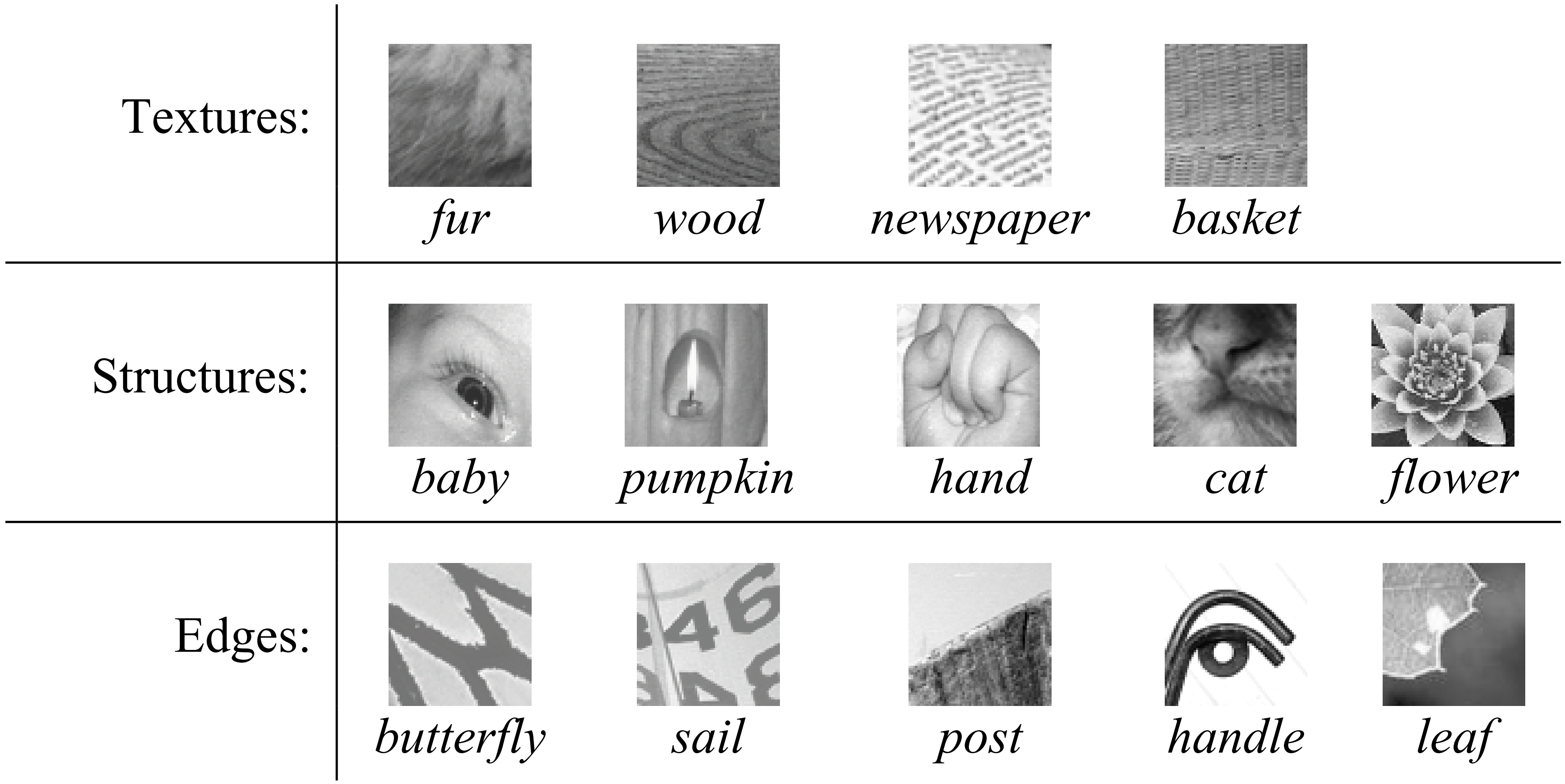

Texture


Structure


Edge

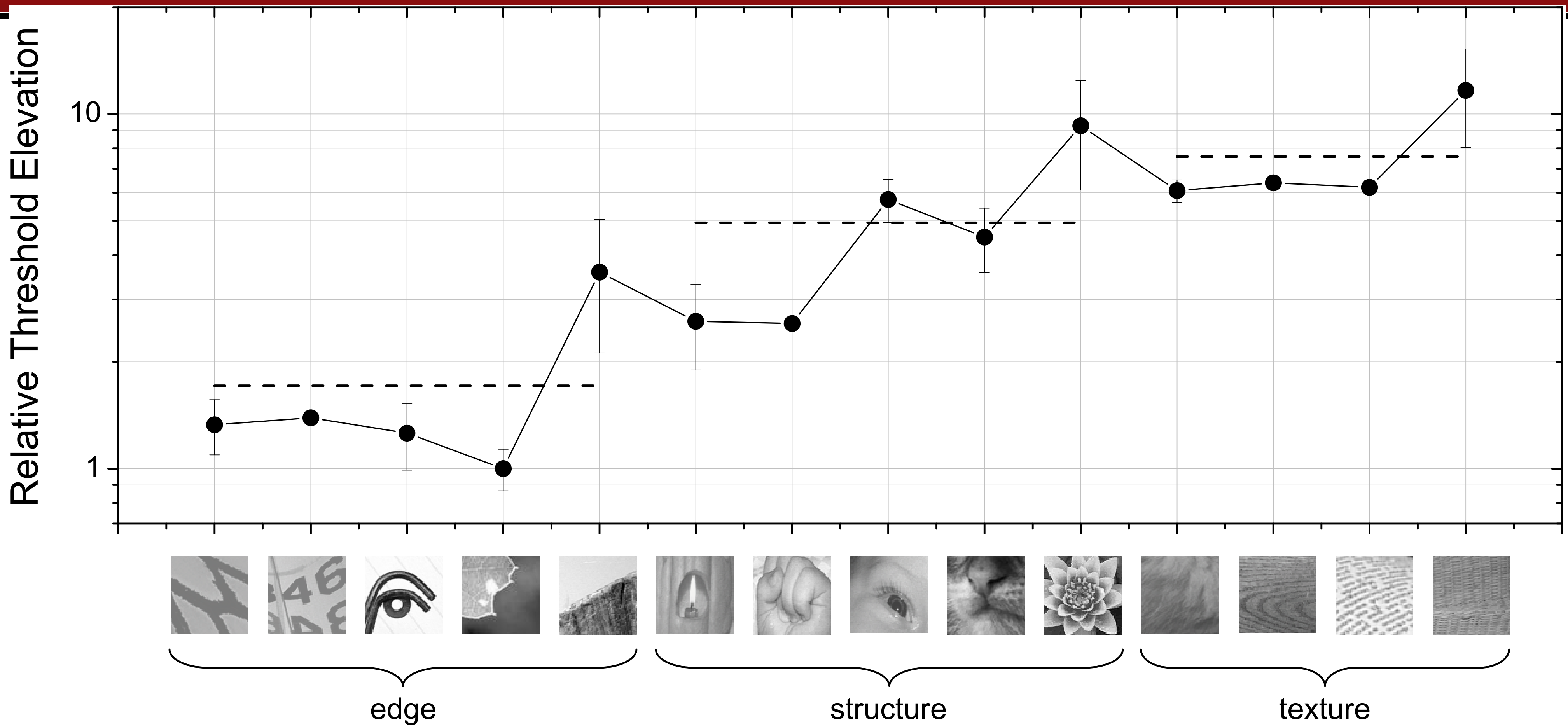


To Solve this Problem



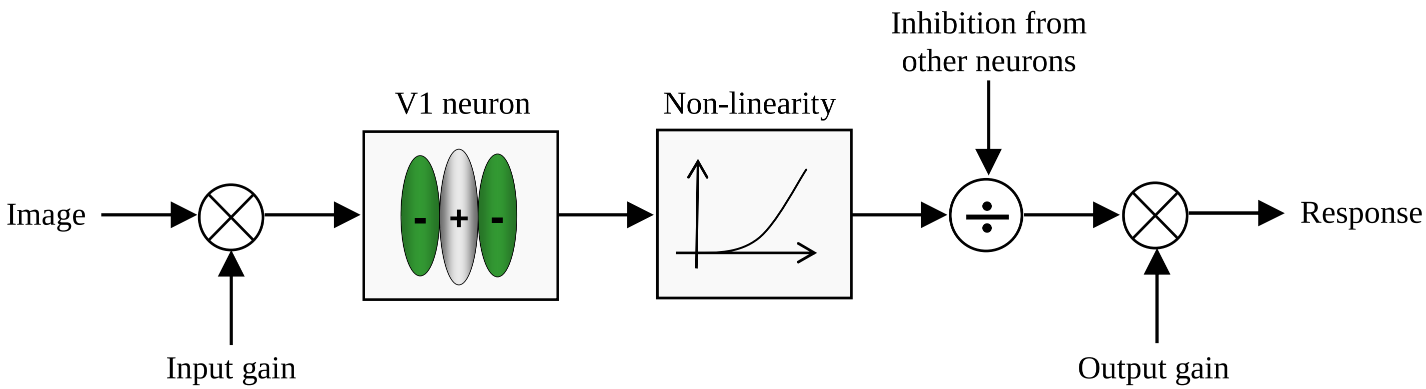
- Experimentally quantify masking of texture/structure/edge patches, and develop an appropriate gain control model.

Relative Threshold Elevations



Textures mask more than structures (2x), which mask more than edges (2.5x).

Improved Gain Control Model with V2 Feedback

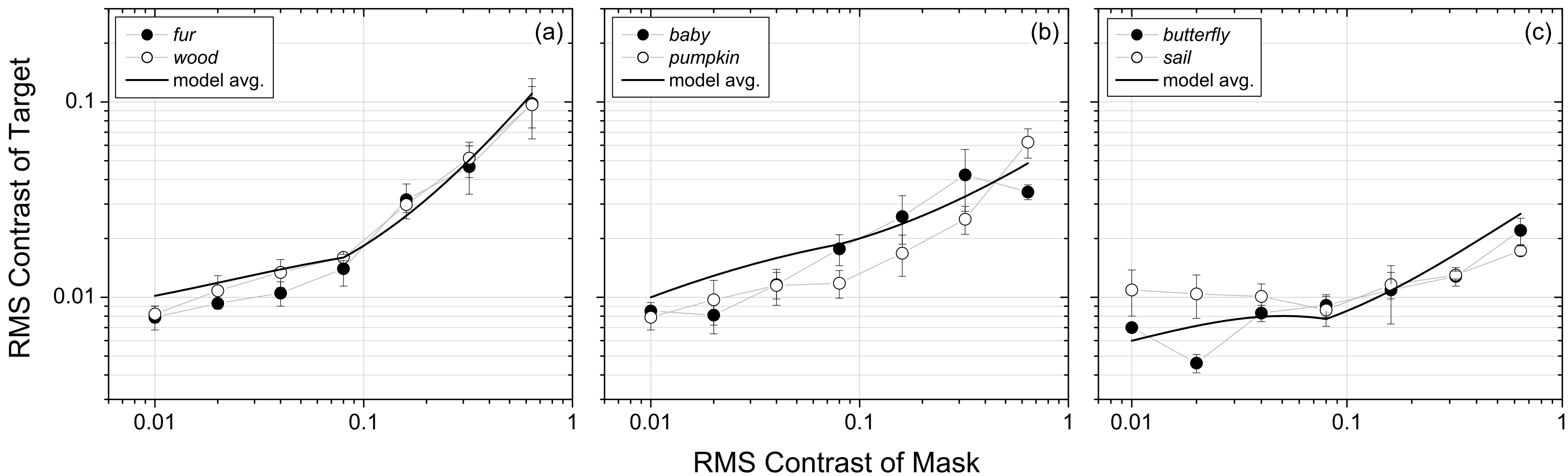


Neural response:

$$r(x, f, \theta) = \frac{w(x, f, \theta)^p}{b^q + g_m \sum_{(x, f, \theta) \in S} w(x, f, \theta)^q}$$

g_m is an inhibitory modulation term and varies based on patch type

...and the Resulting Model Fits



Texture Structure Edge

Applications of Our HVS Characterizations

- Masked CSF and summation/global precedence
  - Distortion-contrast quantization.
  - A new multiple description quantization strategy.
  - Visual signal-to-noise ratio (VSNR) — a quality metric.
- Masked CSF, summation/global precedence, and gain control model
  - Overhead-free optimal spatially localized quantization.

Distortion-Contrast Quantization

A quantization strategy for wavelet-coded natural images based on

1. Our masked detection results at and above threshold;
2. Linearity in summation;
3. Global precedence.

Result: a strategy which works seamlessly for all rates, producing better looking images at up to 30% lower rates.

JPEG-2000 compatible (but not necessary!)

Original Harbor Image

Harbor, 0.4 bits/pixel, JPEG-2000 Framework

JPEG-2000

Contrast-based JPEG-2000

Default

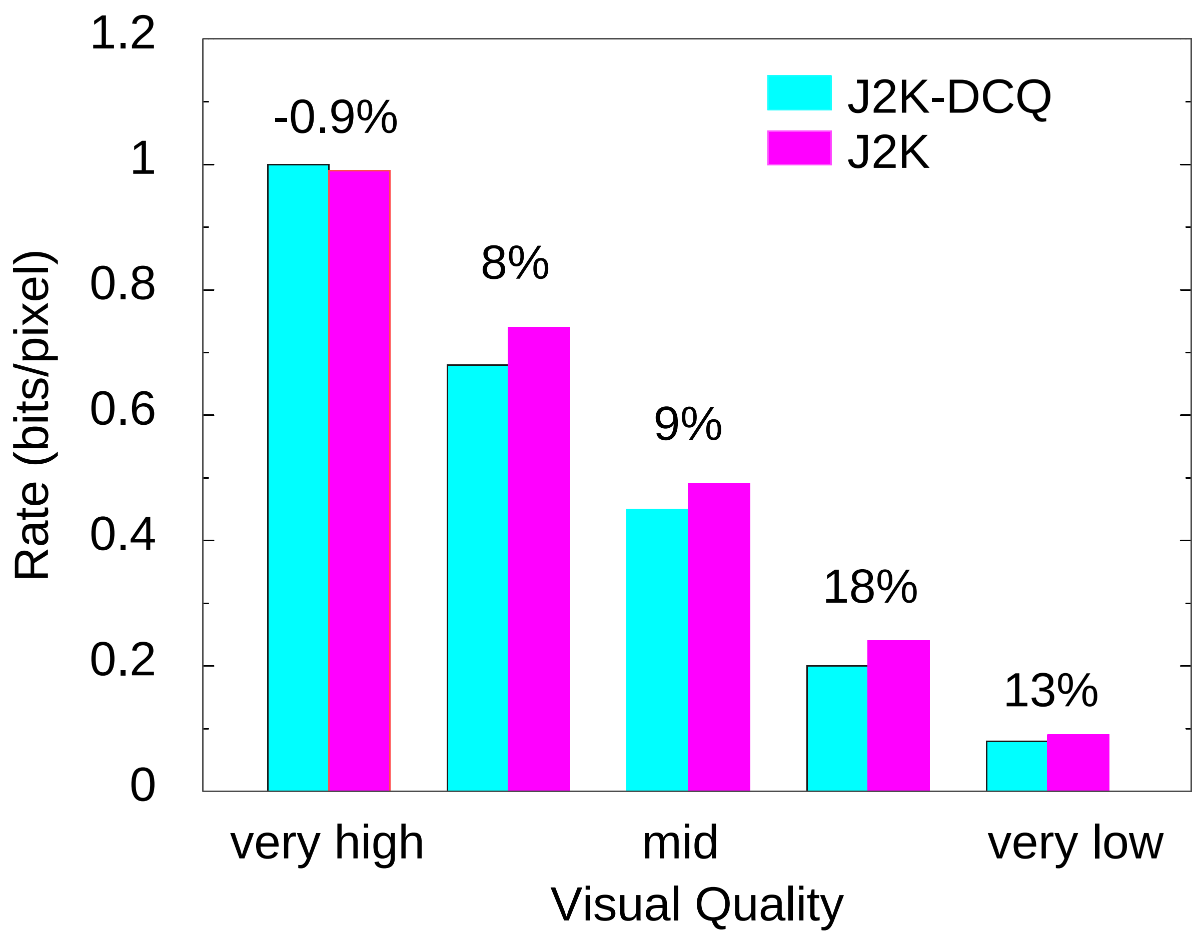
DCQ

So What's the Bit Rate Savings?

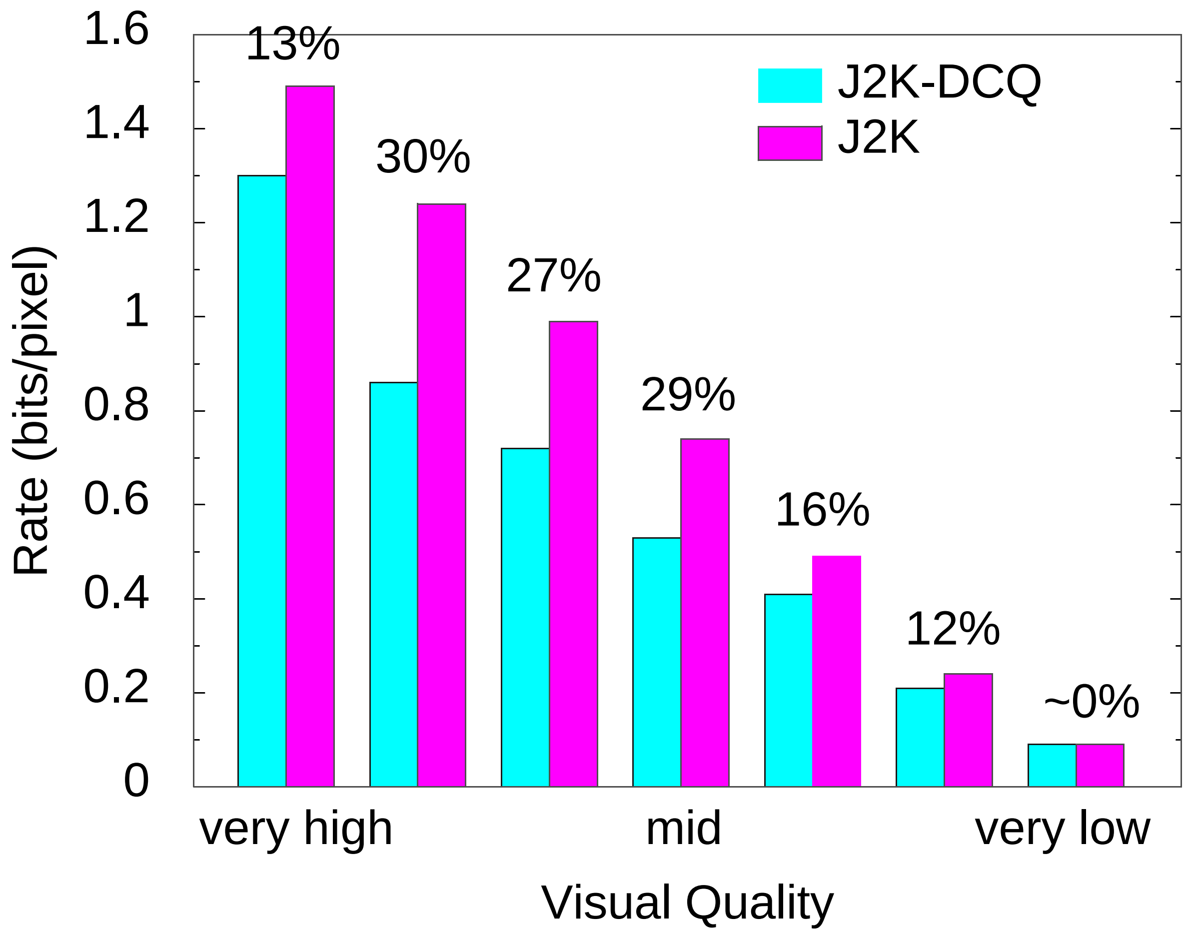
Cat

Rainriver

At Equal Quality: Rate Savings for Cat



At Equal Quality: Rate Savings for Rainriver



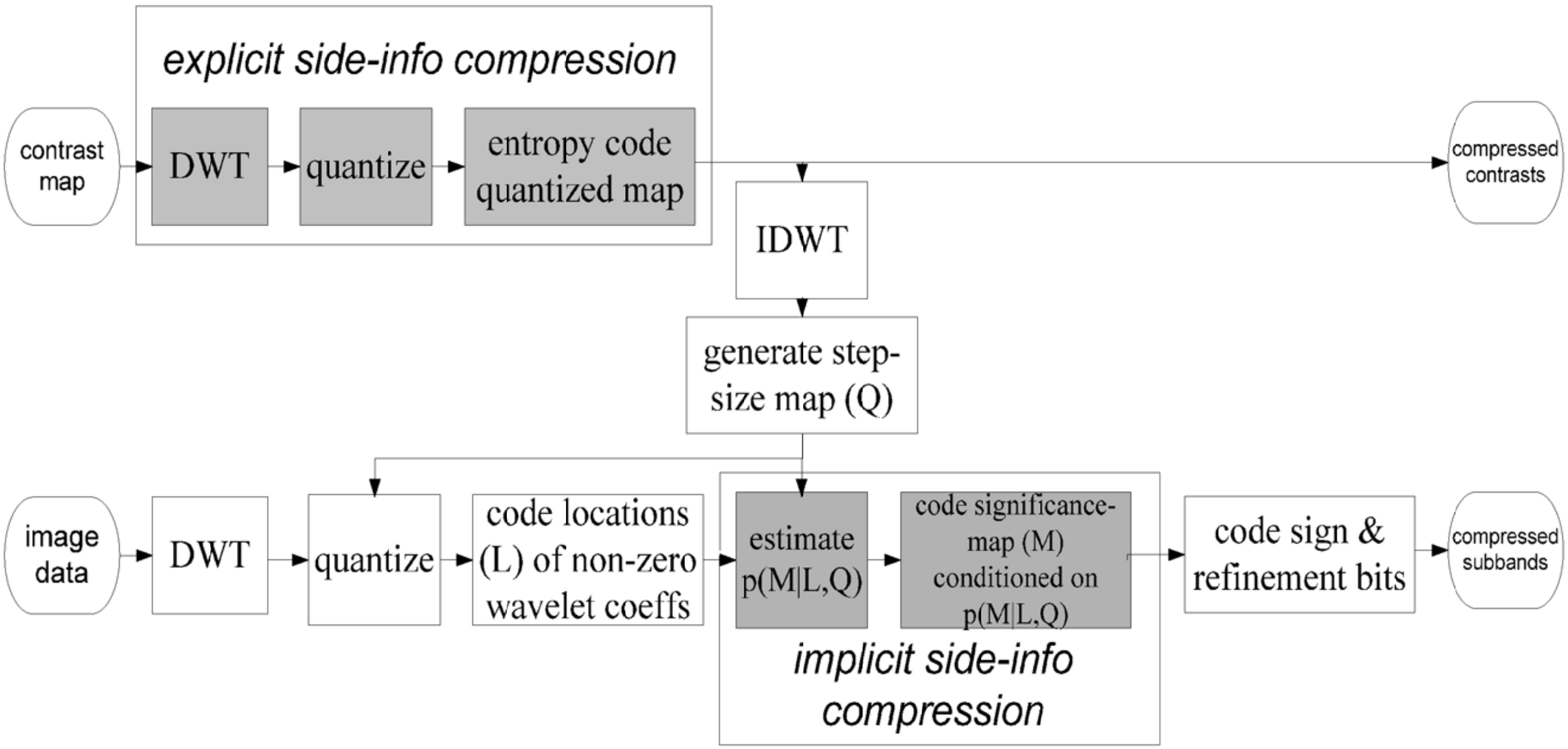
Overhead-free Optimal Spatially Localized Quantization

- Goal — set quantization step sizes locally within an image according to local masking thresholds.

- Problem — step sizes must then be transmitted along with the image. Until now, the overhead has proved to be prohibitive.

- Our solution — information used to produce the step sizes is used as side information to compress the image. This does NOT incur a rate penalty: conditioning reduces entropy.

Spatial Coder Block Diagram



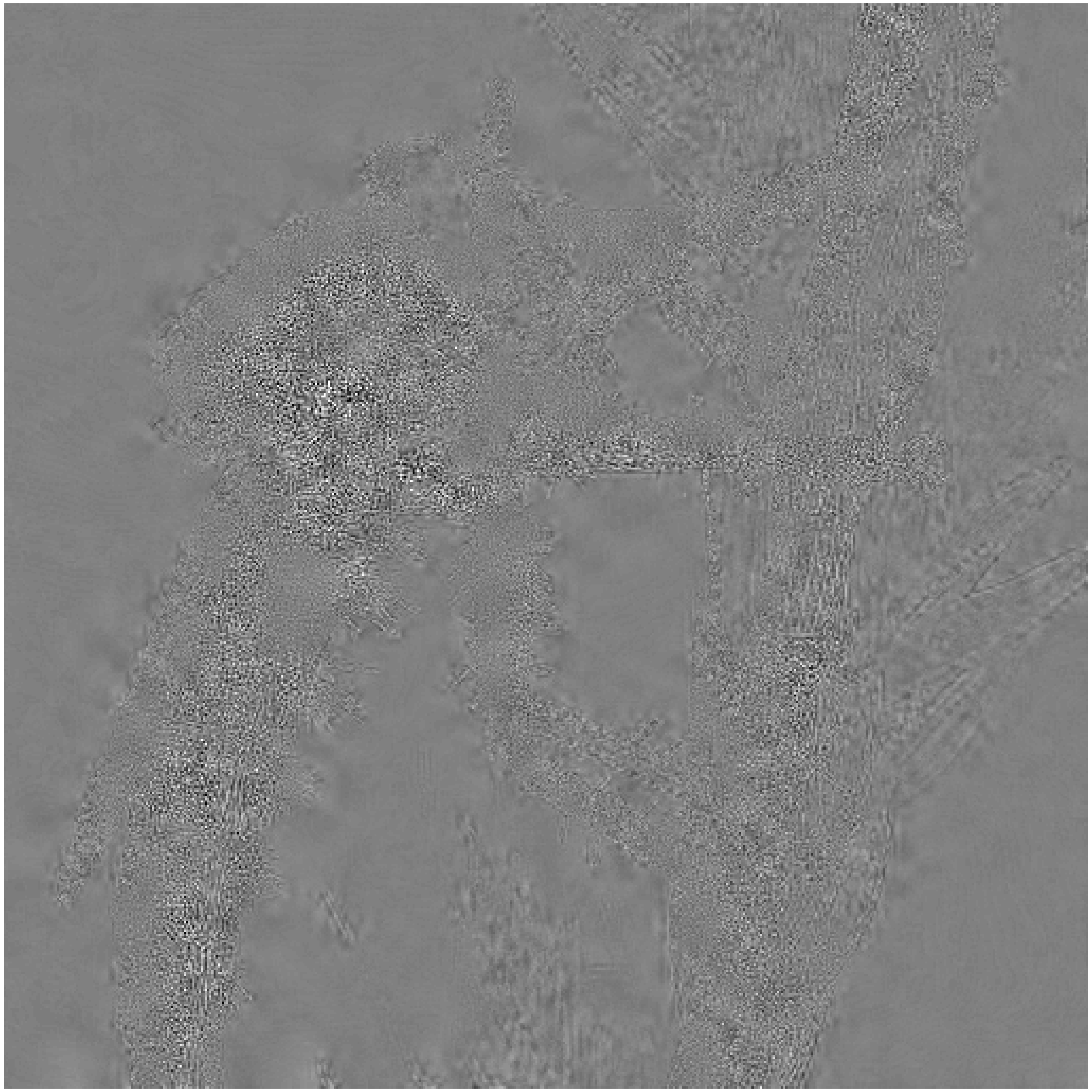
Visual & PSNR Results

Distortion visibility	Image	Number preferred		PSNR	
		Proposed	JPEG-2K	Proposed	JPEG-2K
Barely visible	horse (1.13 bpp)	8	0	30.0	32.1
	rhino (1.88 bpp)	7	1	24.3	29.0
Very visible	horse (0.64 bpp)	6	2	27.0	28.3
	rhino (1.24 bpp)	6	2	21.3	25.7

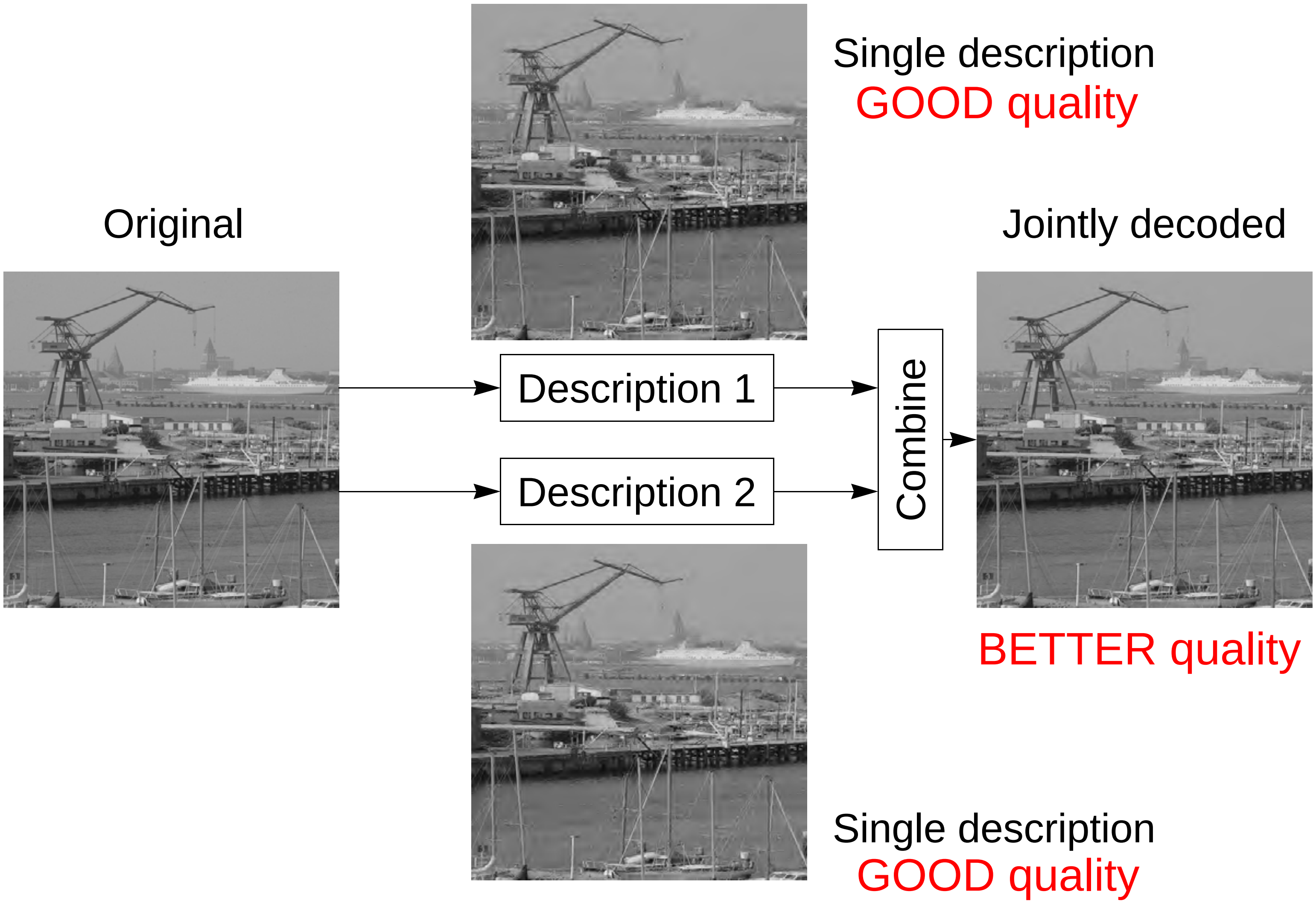
Spatially localized quantization hides much more error in the image for the same visual quality.

Example Image at Threshold

Residual Image



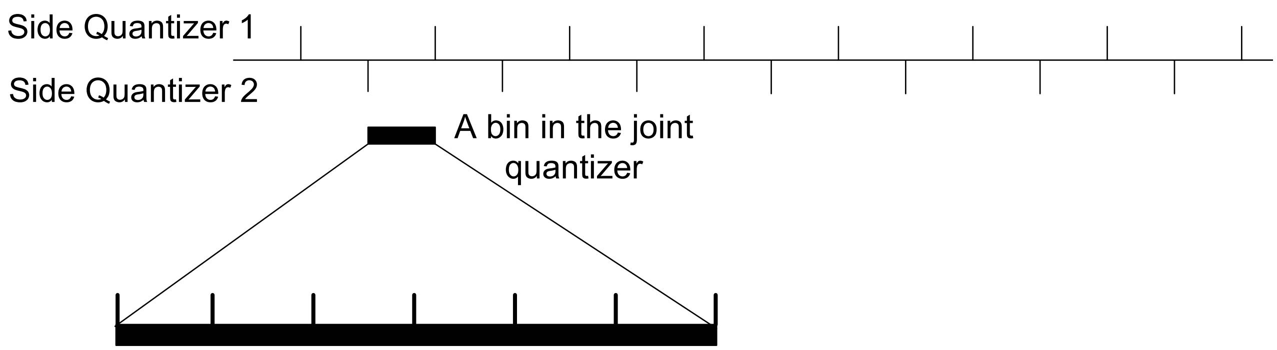
Multiple Description Image Coding



Visually Optimized MD Image Coding

- Problem: HVS results are for distortions caused by uniform (convex) quantization cells, BUT “standard” MD quantizers use non-convex cells.

- Our solution: design a new MD quantization strategy which has equivalent R-D performance to standard techniques but which uses convex cells.



Examples: Original Harbor Image

Harbor Images: 2 joint descriptions at equal quality...

MSE-optimized

Visually-optimized

...yield these 1 description images:

MSE-optimized

Visually-optimized

Concluding Comments

- Extensive psychophysical experiments have yielded more accurate HVS characterizations for image compression.
- These HVS characteristics have been used to drive signal processing algorithm development.
- The resulting algorithms outperform current state-of-the-art results.
- We have also applied this methodology to the design of a video quality measure.

Task-Based Imaging — Quantifying Image Usefulness and its Relationship to Image Quality

What is task-based imaging?

From the user/application perspective:

- Who is viewing it and why?
- How is the visual information to be used?

From the image processor’s perspective:

- What *must* be conveyed by the visual information?
- What is nice to have, but optional for the task?

What Makes an Image Recognizable?

Detection & Masked Detection, More Realistic Stimuli

=

+

quantized | original | quantization noise

Detection: stimuli = target

Masked detection:

stimuli = mask + target

Detection & Masked Detection, More Realistic Stimuli

The target is the *distortion*

=

+

quantized | original | quantization noise

Detection: stimuli = target

Masked detection:

stimuli = mask + target

Recognition/Utility — The Target is the Image Content

Questions

- How can we measure “usefulness” of an image?
- What distortions should we explore?
- How is “quality” related to usefulness (utility)?
- Can current quality estimators predict utility?
- Can we create a *utility estimator*?

A Framework for Measuring Image Utility

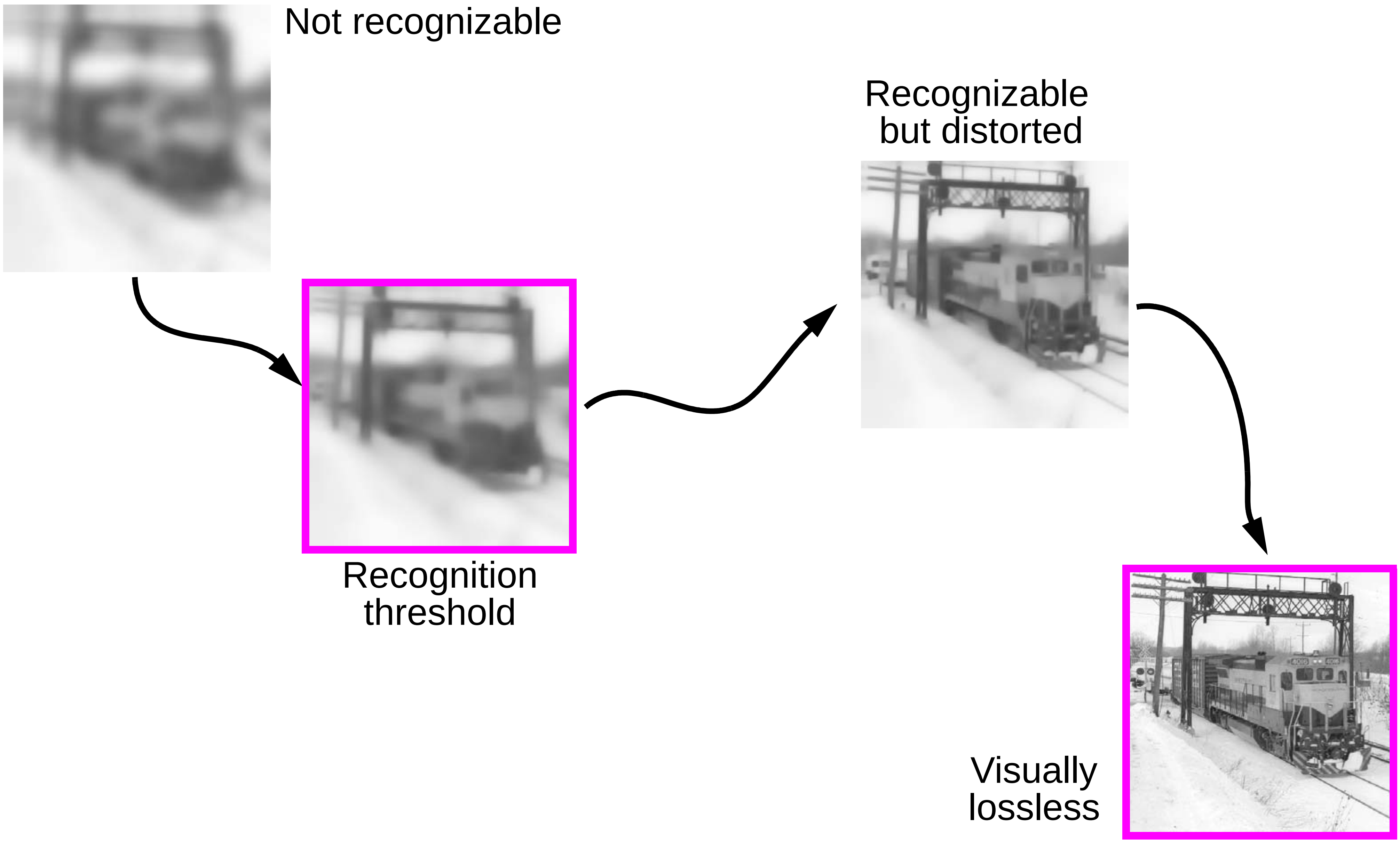
Not recognizable

Recognition threshold

Recognizable but distorted

Visually lossless

A Framework for Measuring Image Utility



A Framework for Measuring Image Utility

Not recognizable

Recognition threshold

Recognizable but distorted

Visually lossless

How “long” are these distances?

Three Experiments: Recognition, Utility Assessment, Quality Assessment

1. Recognition

   Single-image stimulus: “Do you recognize the image content?”

2. Utility assessment

   Image pair stimulus: “Which image tells you more about the content?”

3. Quality assessment

   SAMVIQ or ACR

Graduate Student Acknowledgements

- HVS/Perception: Prof. Damon Chandler, Dr. Marcia Ramos, Dr. Mark Masry, Bobbie Chern, Jeri Moses.
- Image Applications: Prof. Damon Chandler, Dr. Matt Gaubatz, Dr. Chao Tian.
- Utility work: Dr. David Rouse.

Papers on all these topics can be found at

http://foulard.ece.cornell.edu