Image Processing and Visual Communications

Perceptual Image Quality Assessment and Processing

Zhou Wang

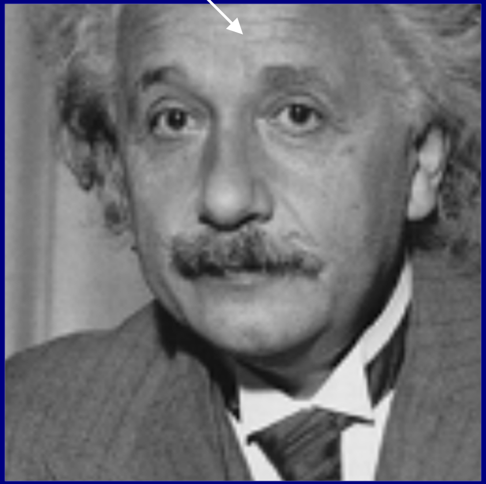
Dept. of Electrical and Computer Engineering
University of Waterloo

Outline

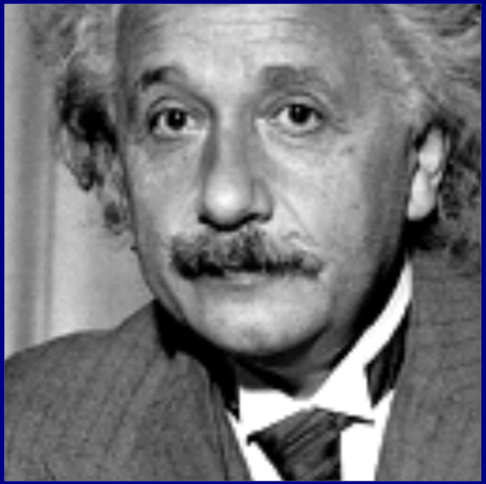
- **Why "Perceptual"?**
  - Motivation
  - Overview
- **Perceptual Image Quality Assessment**
  - What's wrong with mean squared error?
  - Error visibility methods
  - Structural Similarity methods
- **Perceptual Image Processing**
  - Image compression
  - Image denoising
- **Real-World Challenges**

Motivation

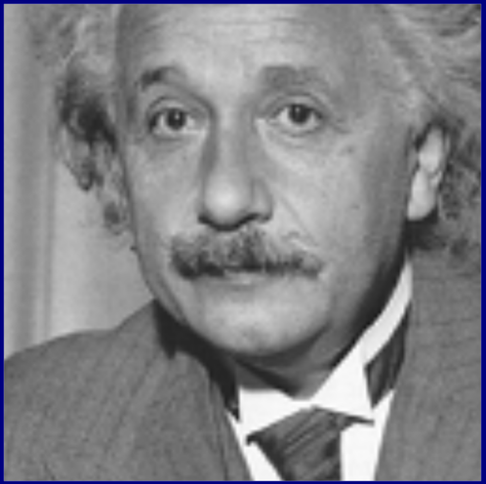
original Image



MSE=0, MSSIM=1



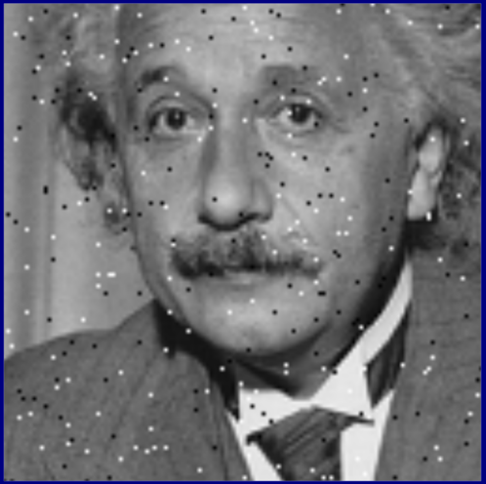
MSE=309, MSSIM=0.928



MSE=309, MSSIM=0.987

MSE=309, MSSIM=0.580

MSE=308, MSSIM=0.641



MSE=309, MSSIM=0.730

Perceptual Image Processing Overview

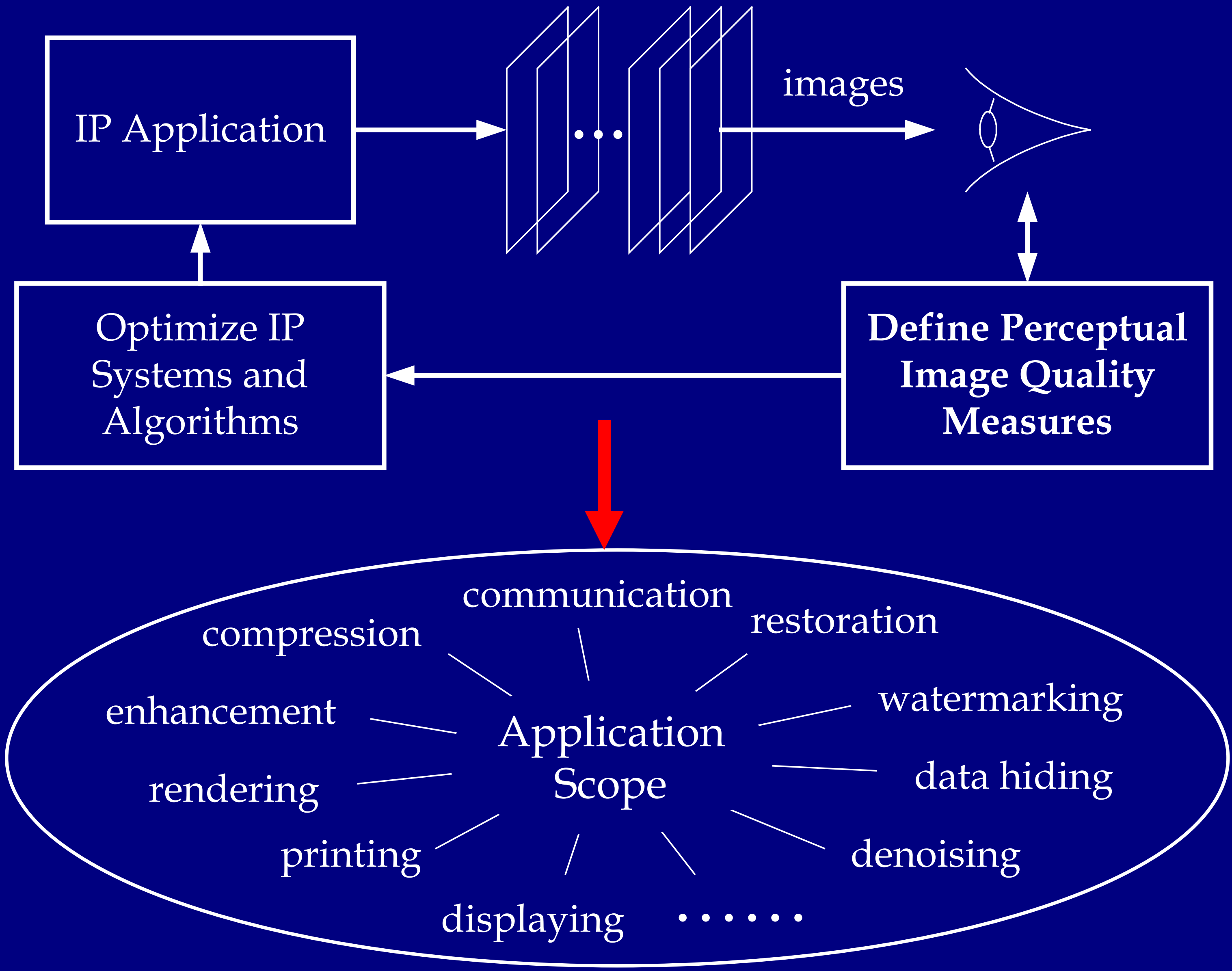


Image Quality Assessment: Classifications

	Full-reference (FR)	Reduced-reference (RR)	No-reference (NR)
Application-specific	Many	Some	Many
General-purpose	Many	Some	Many

Widely used: MSE and PSNR

- **Availability of Reference:**
  - Full-Reference (FR): reference (original) image available
  - No-Reference (NR): reference image not available
  - Reduced-Reference (RR): reference image partially available
- **Application Scope**
  - General-purpose vs. application-specific

What's Wrong with Mean Squared Error? (1)

- Minkowski Metric (MSE as a special case)

image $\boldsymbol{x}$

image $\boldsymbol{y}$

$$\left[\sum_i \left|x_i - y_i\right|^p\right]^{1/p} \longrightarrow E$$

- Advantages
  - Easy to compute
  - Easy to optimize, especially when $p = 2$
  - Clear physical meaning: energy

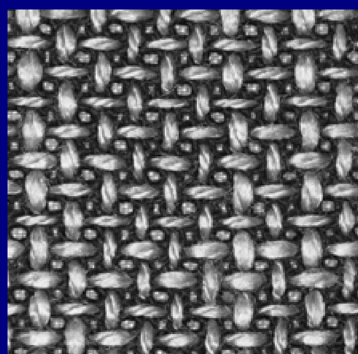
- What's the problem?

What's Wrong with Mean Squared Error? (2)

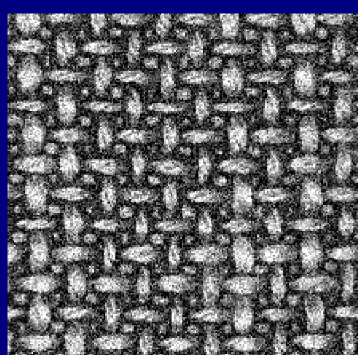
$$E = \left[\sum_i \left| x_i - y_i \right|^p \right]^{1/p}$$

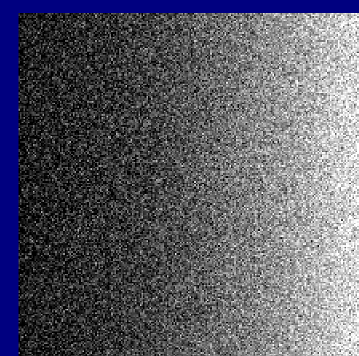
Don't care about ordering

Same Minkowski Metric !



reordering





MSE = 1600, MSSIM = 0.6373

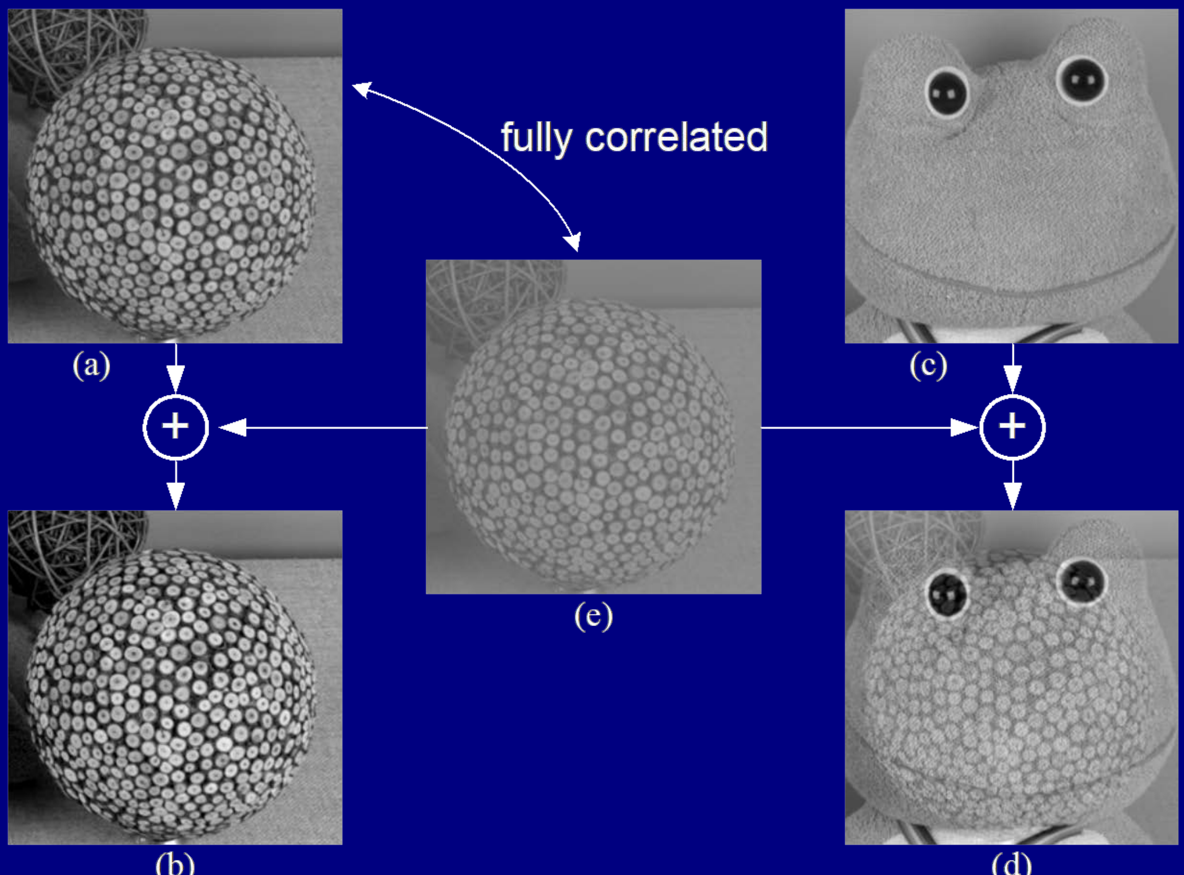
MSE = 1600, MSSIM = 0.0420

What's Wrong with Mean Squared Error? (3)

$$E = \left[\sum_i \left| x_i - y_i \right|^p \right]^{1/p}$$

Care about signal difference **only**, not the underlying signals

Same Minkowski Metric !



What's Wrong with Mean Squared Error? (4)

$$E = \left[\sum_i \left| x_i - y_i \right|^p \right]^{1/p}$$

Don't care about the sign

Same Minkowski Metric !

+ 30

MSE = 900

MSSIM = 0.9329

+ (rand sign)* 30

MSE = 900

MSSIM = 0.2470

What's Wrong with Mean Squared Error? (5)

- Minkowski Metric

$$E = \left[\sum_i \left| x_i - y_i \right|^p \right]^{1/p}$$

- Natural Images

highly structured

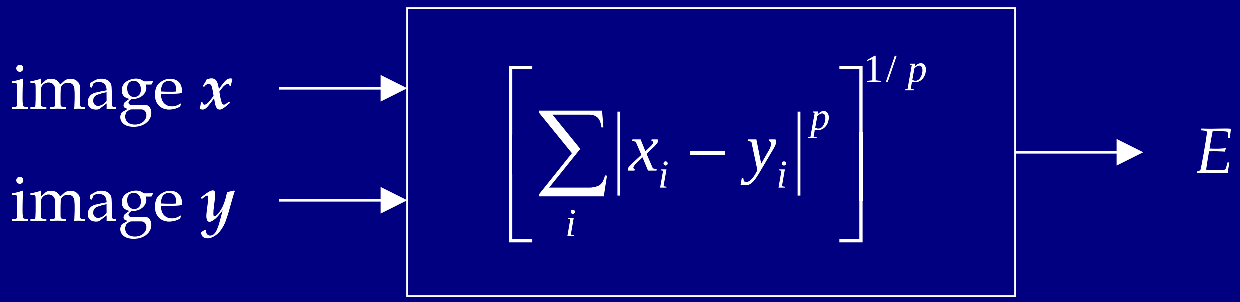
Conflict

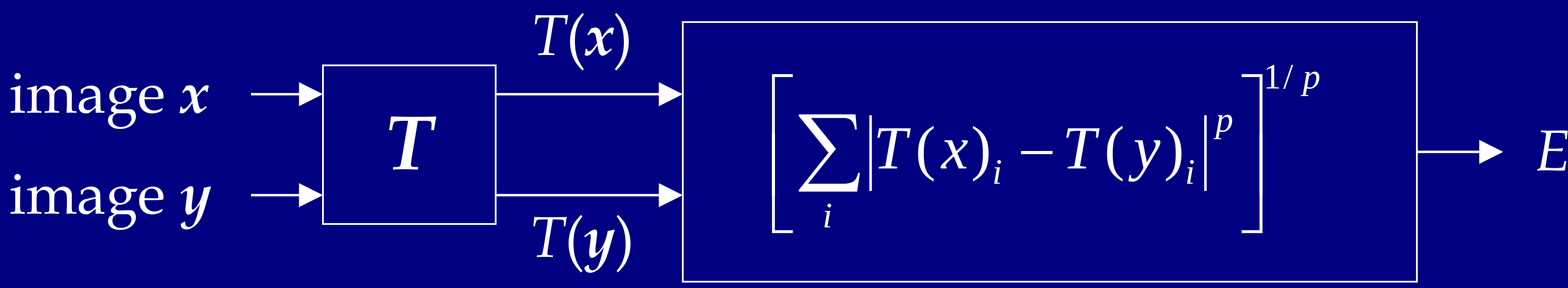
signal samples are independent

signal samples highly correlated

What's Wrong with Mean Squared Error? (6)

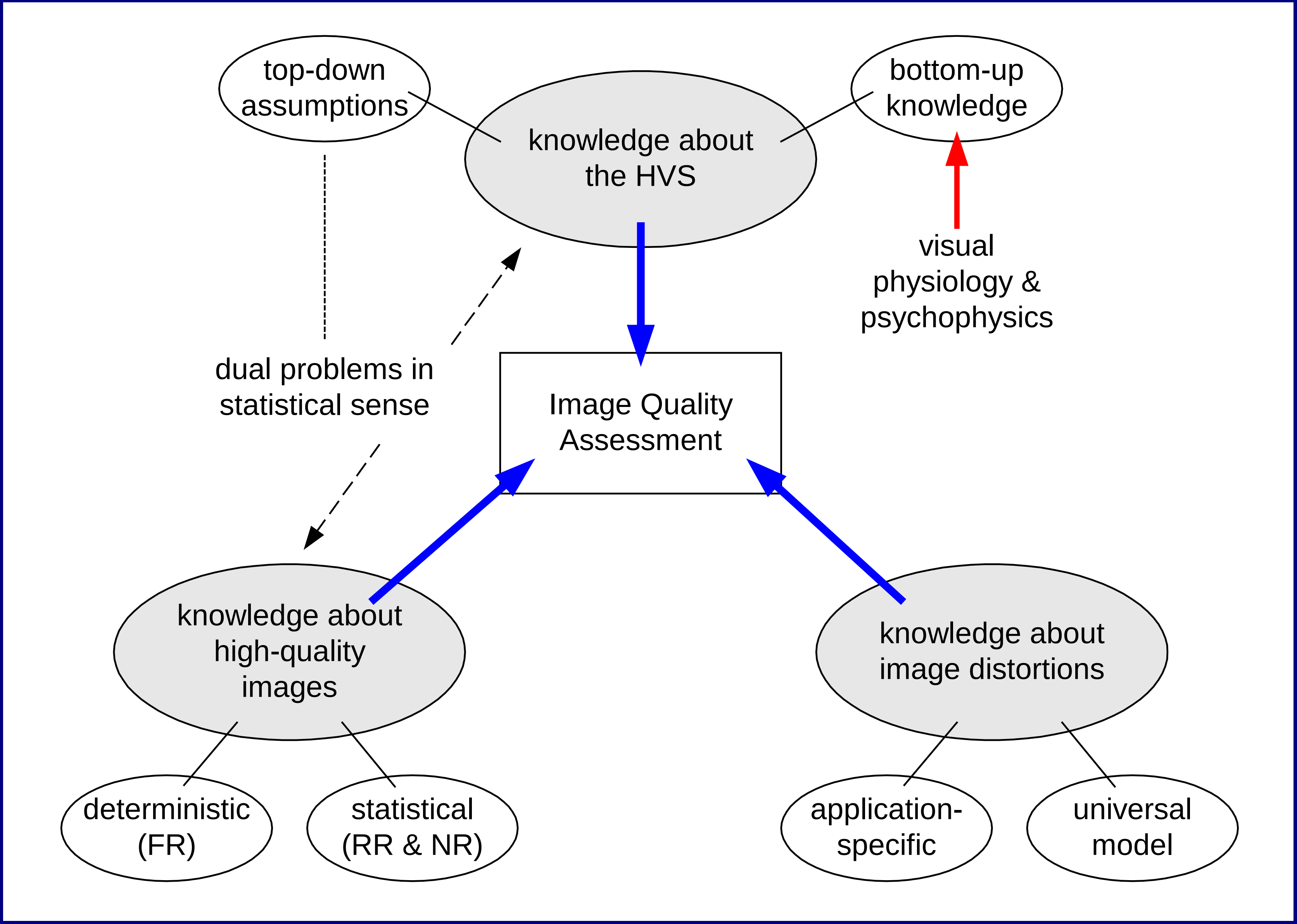
- Possible solution:





Make $T(\boldsymbol{x})$ and $T(\boldsymbol{y})$ independent, at least decorrelated

Overview: Source of Knowledge for IQA



[Wang & Bovik, *Modern Image Quality Assessment*, 2006]

Error Visibility Method: Idea

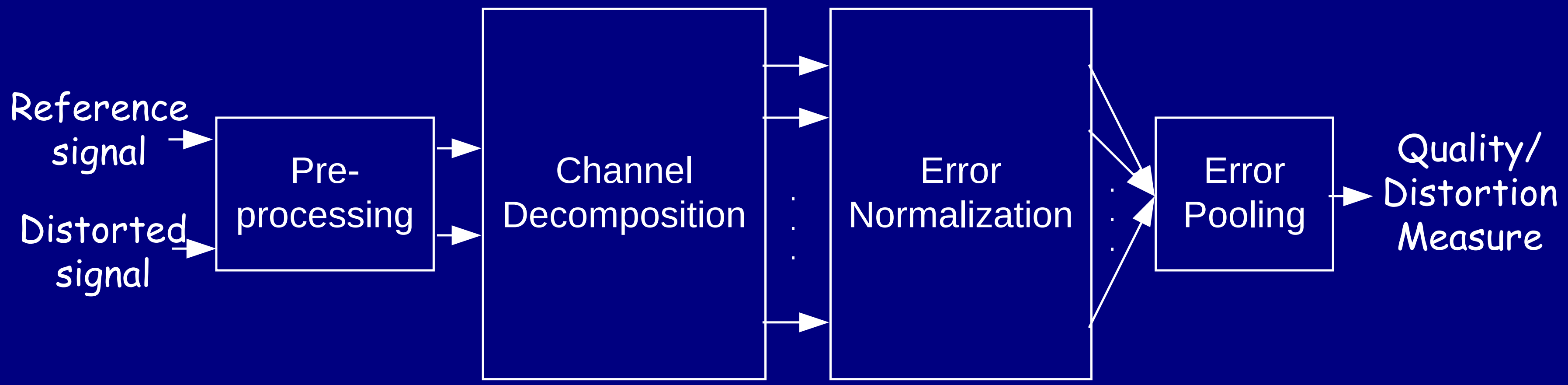
distorted signal = reference signal + error signal

↓

Quantify **error signal** perceptually

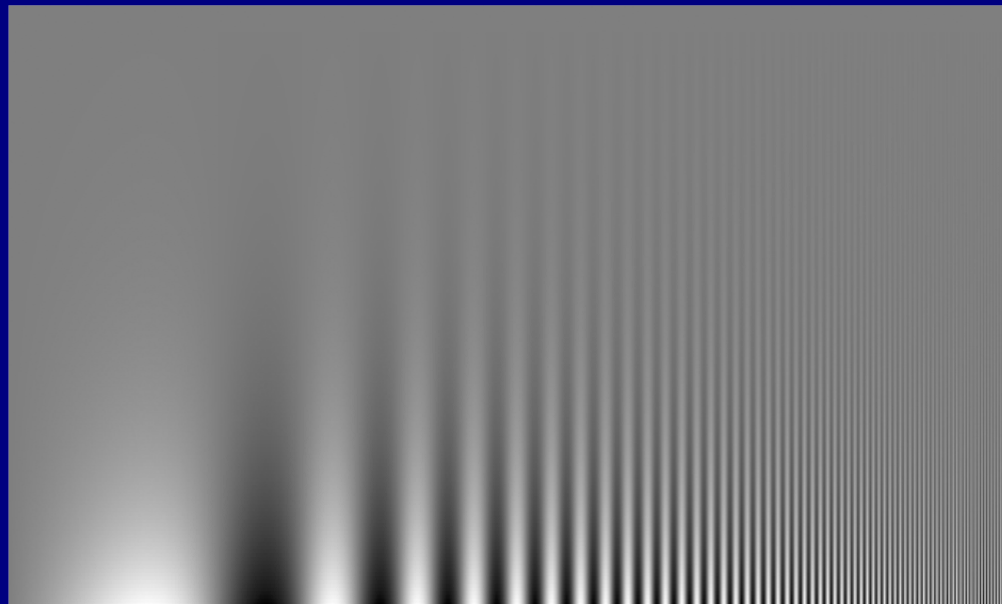
- **Representative work**
  - Frequency weighting (pioneering work) [Mannos & Sakrison '74]
  - Sarnoff model [Lubin '93]
  - Visible difference predictor [Daly '93]
  - Perceptual image distortion [Teo & Heeger '94]
  - DCT-based method [Watson '93]
  - Wavelet-based method [Safranek '89, Watson *et al.* '97]
  - ……

Error Visibility Method: Framework



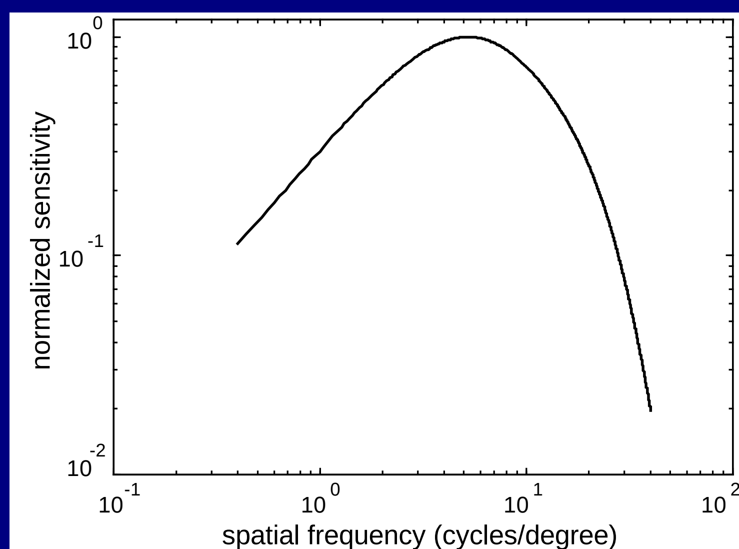
- Goal: simulate relevant early HVS components
  - Structures motivated by **physiology**
  - Parameters determined by **psychophysics**

Contrast Sensitivity Function



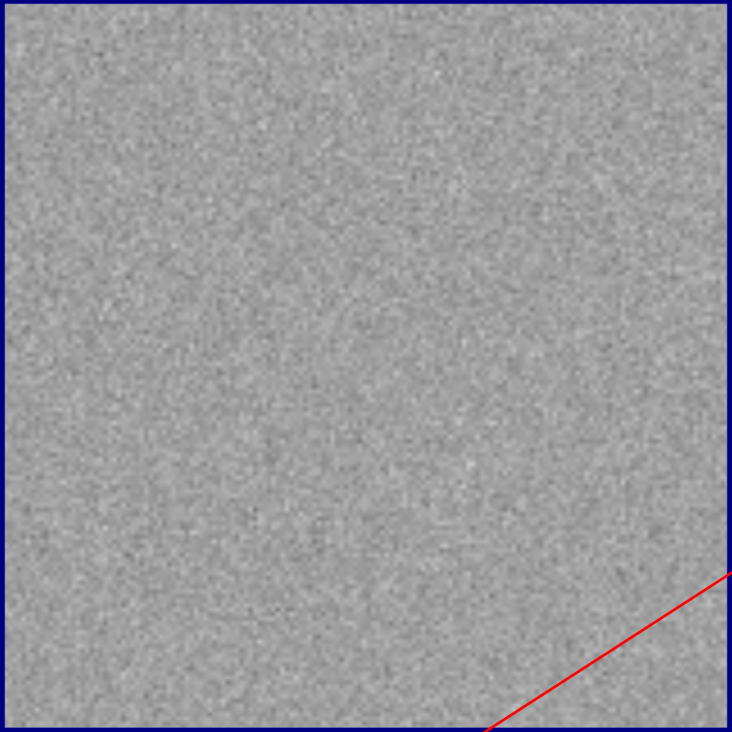
contrast

Campbell-Robson
CSF chart



spatial frequency

Masking

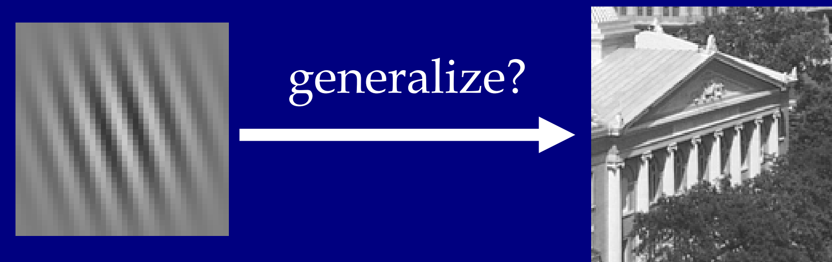


highly visible
weak masking

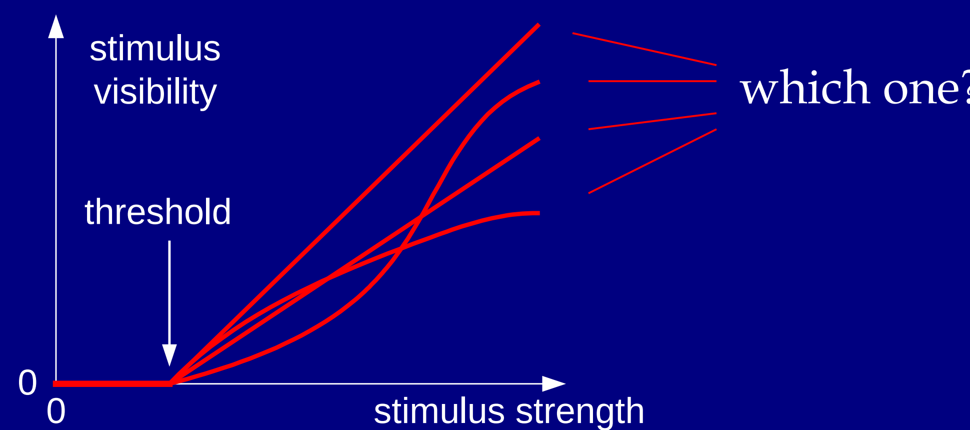
hardly visible
strong masking

Error Visibility Method: Difficulties

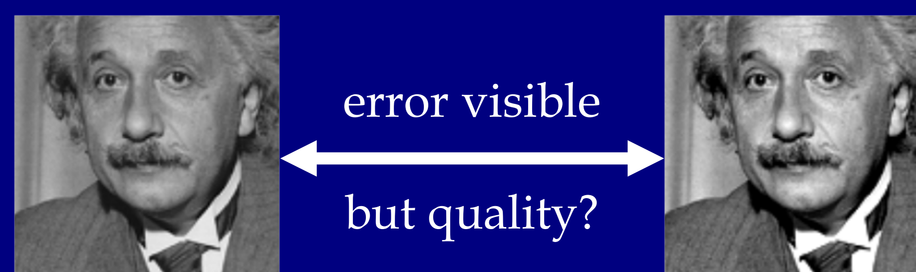
- **Natural image complexity problem**
  - Based on simple-pattern psychophysics



- **Suprathreshold problem**
  - Based on threshold psychophysics



- **Quality definition problem**
  - Error visibility = quality ?



[Wang *et al.*, *IEEE Trans. Image Processing*, ’04]

Structural Similarity Methods: Idea

Purpose of vision: extract structural information

↓

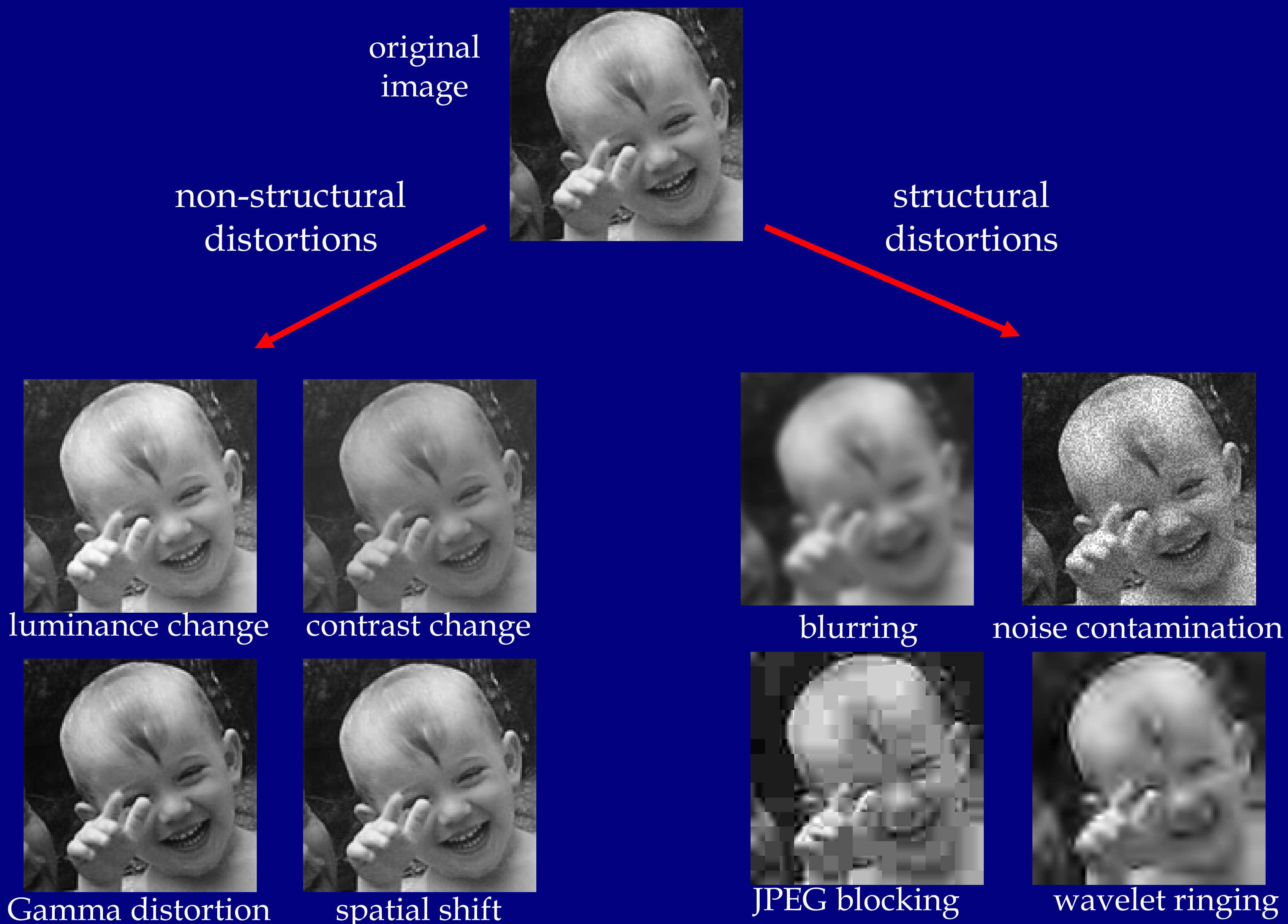
Quantify **structural distortion**

- **Questions:**

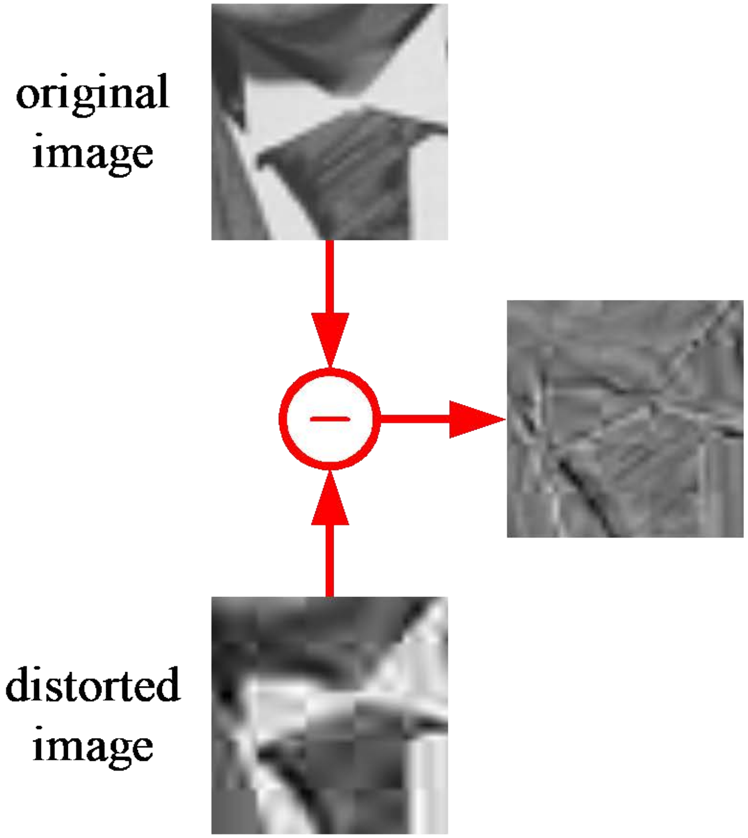
  How to define structural/nonstructural distortions?

  How to separate structural/nonstructural distortions?

What are Structural/Non-Structural Distortions?

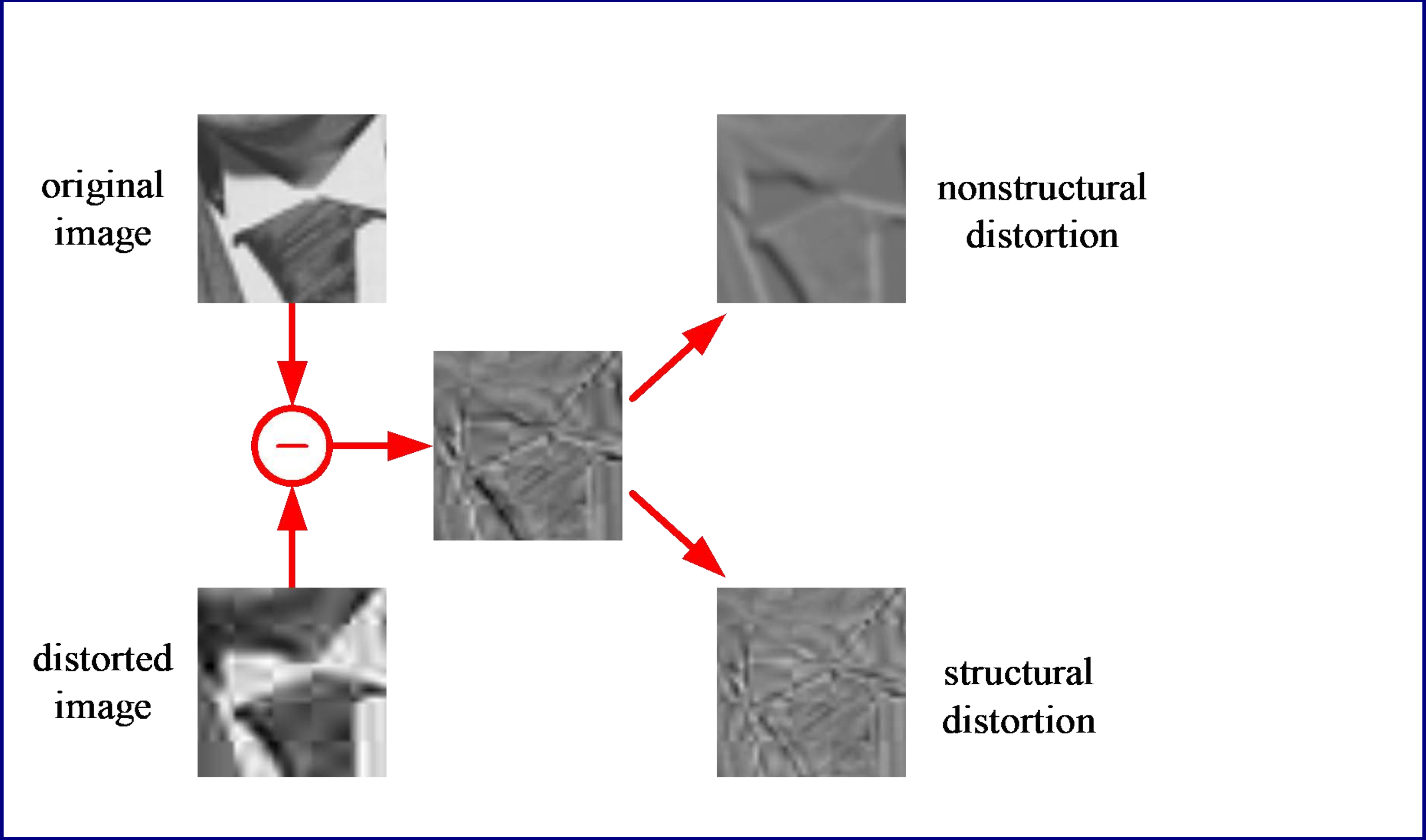


Separation of Structural/Non-Structural Distortions



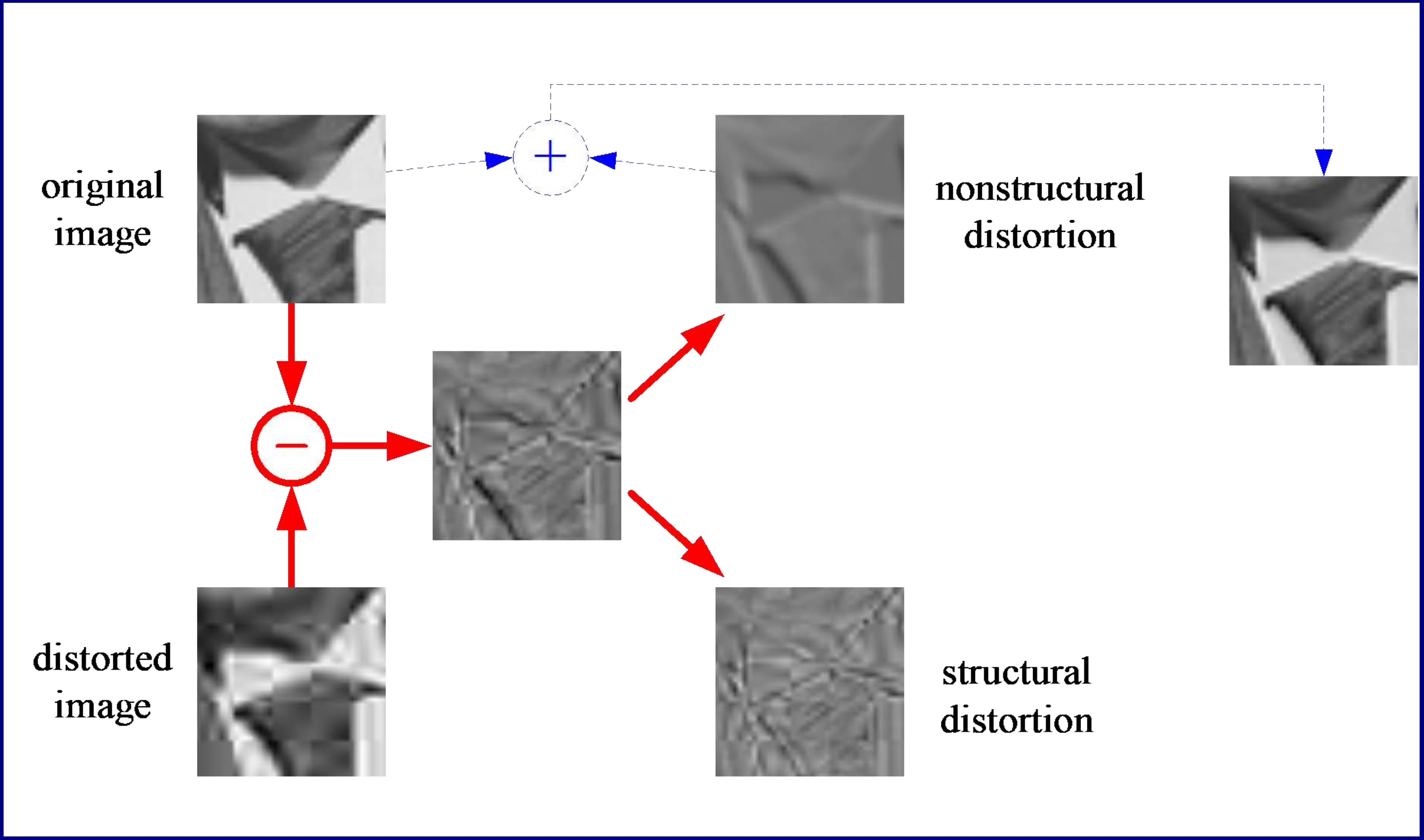
[Wang & Simoncelli, *ICIP* ’05]

Separation of Structural/Non-Structural Distortions



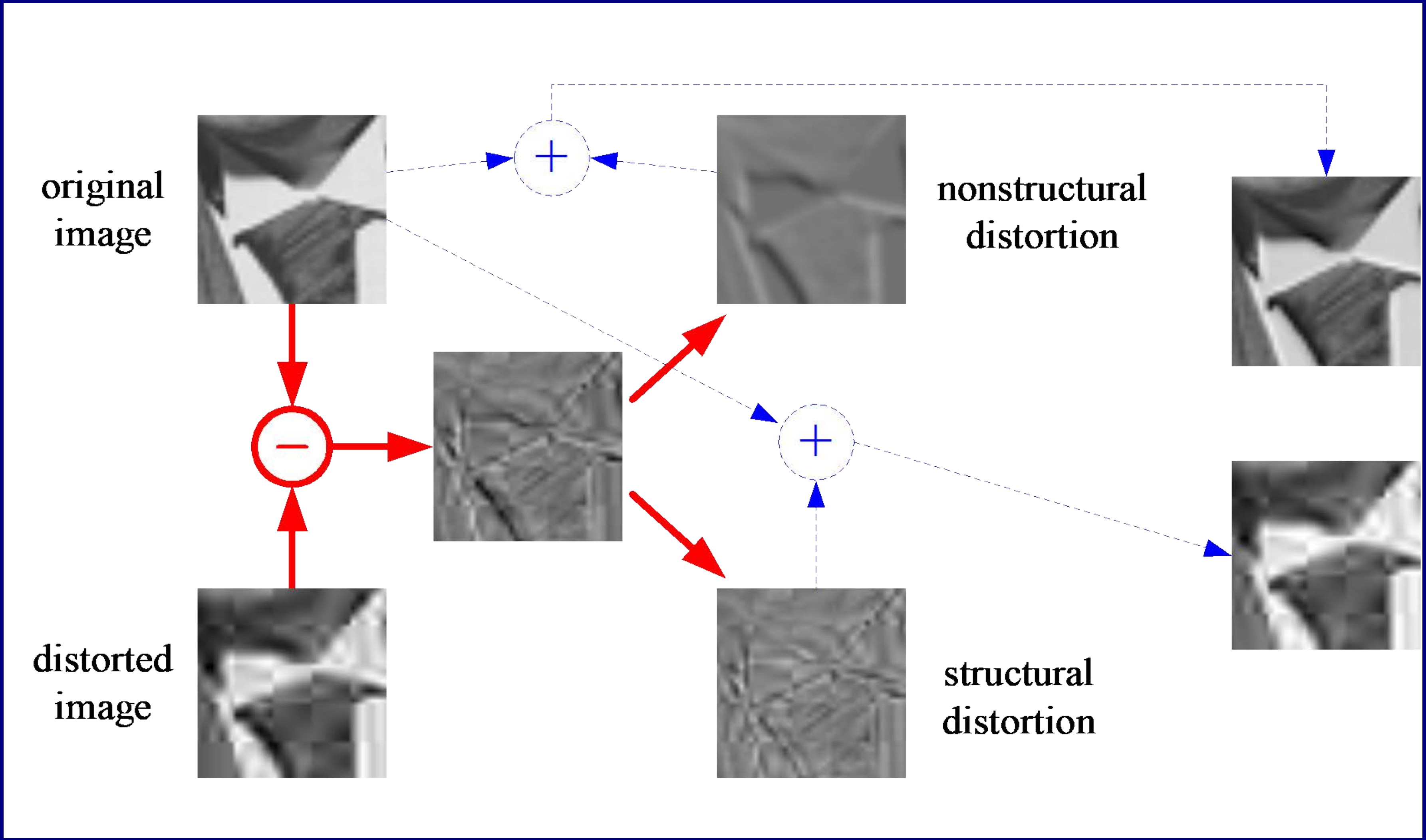
[Wang & Simoncelli, *ICIP* ’05]

Separation of Structural/Non-Structural Distortions



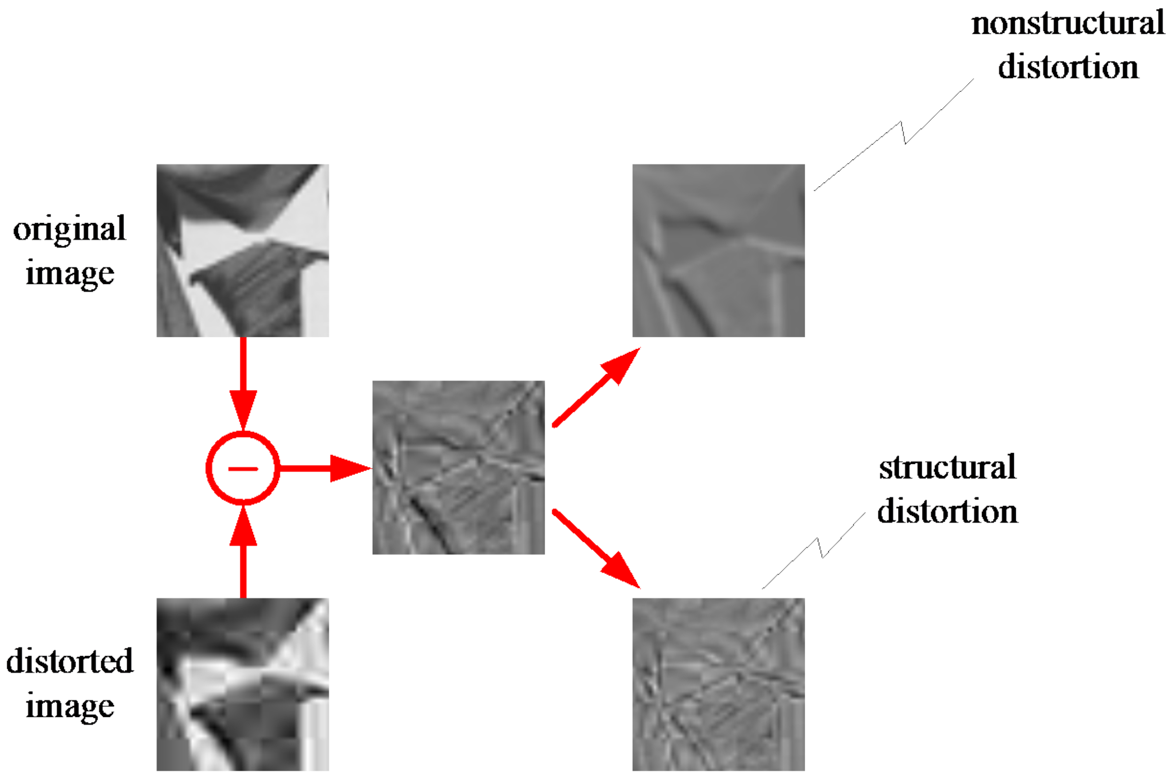
[Wang & Simoncelli, *ICIP* '05]

Separation of Structural/Non-Structural Distortions



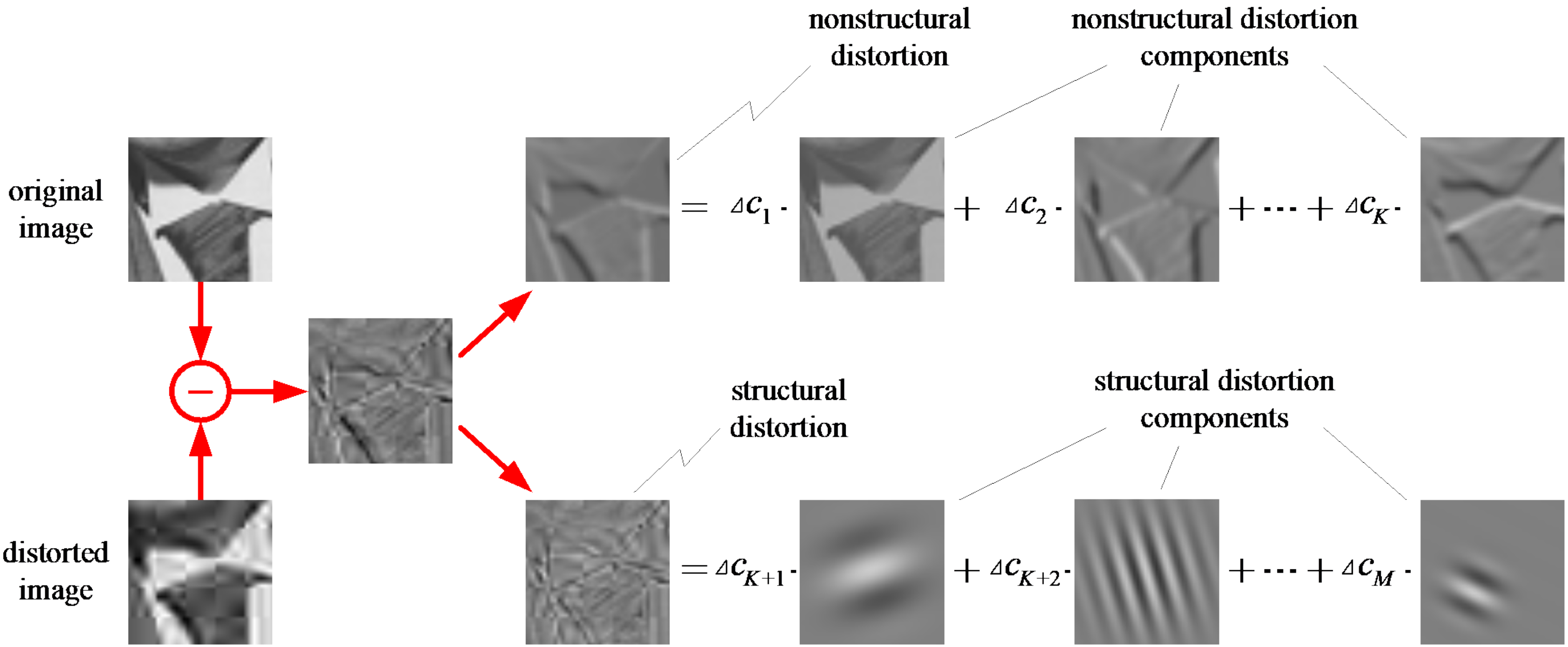
[Wang & Simoncelli, *ICIP* ’05]

Adaptive Linear System



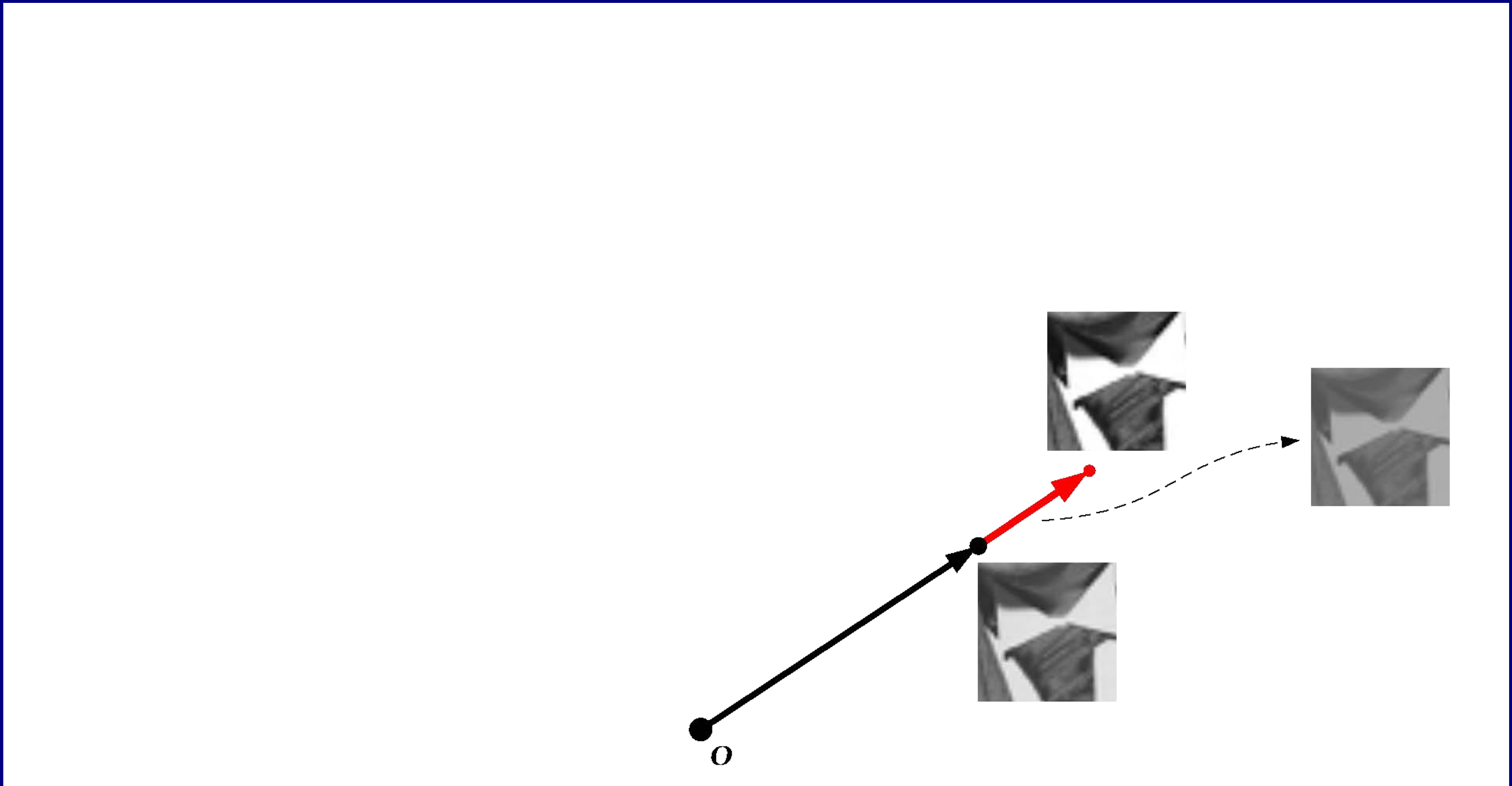
[Wang & Simoncelli, *ICIP* ’05]

Adaptive Linear System



[Wang & Simoncelli, *ICIP* ’05]

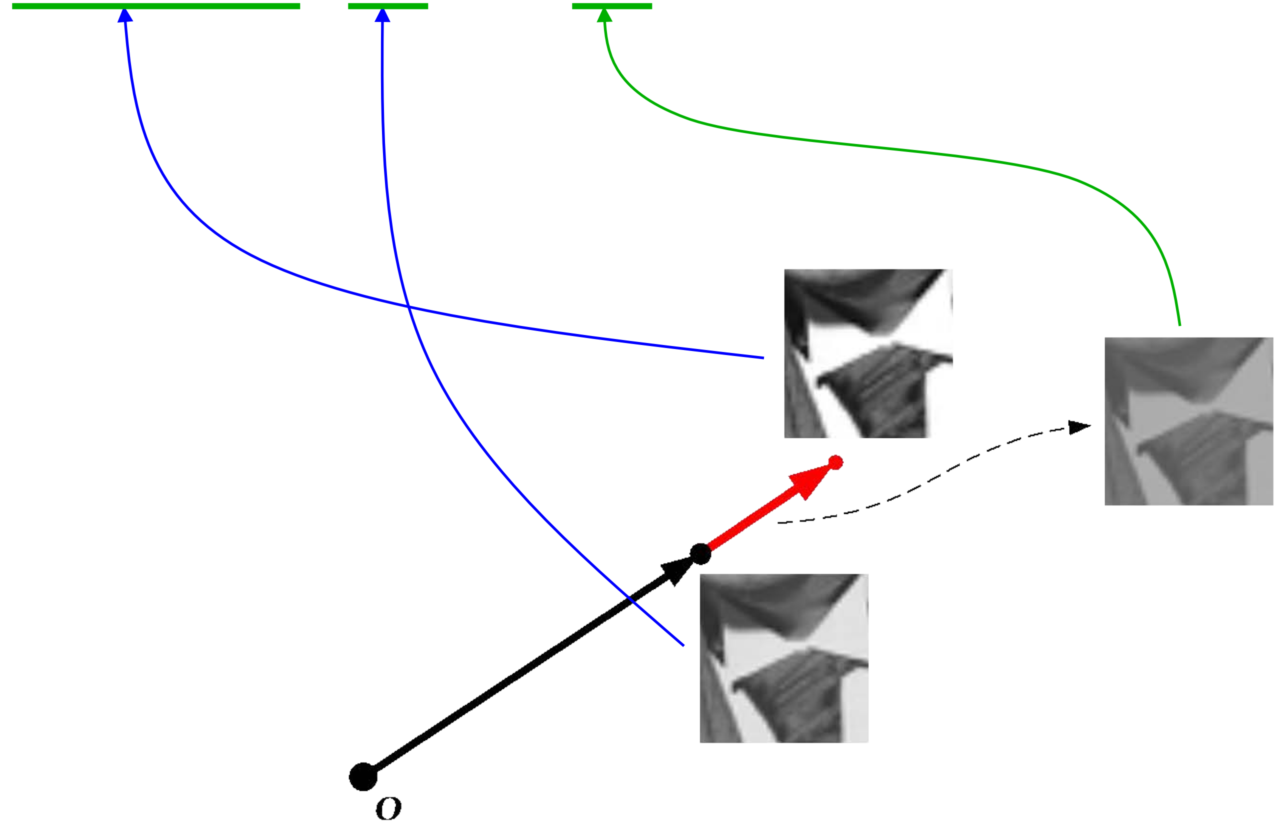
Intuition: Geometry in the Image Space



[Wang & Simoncelli, *ICIP* ’05]

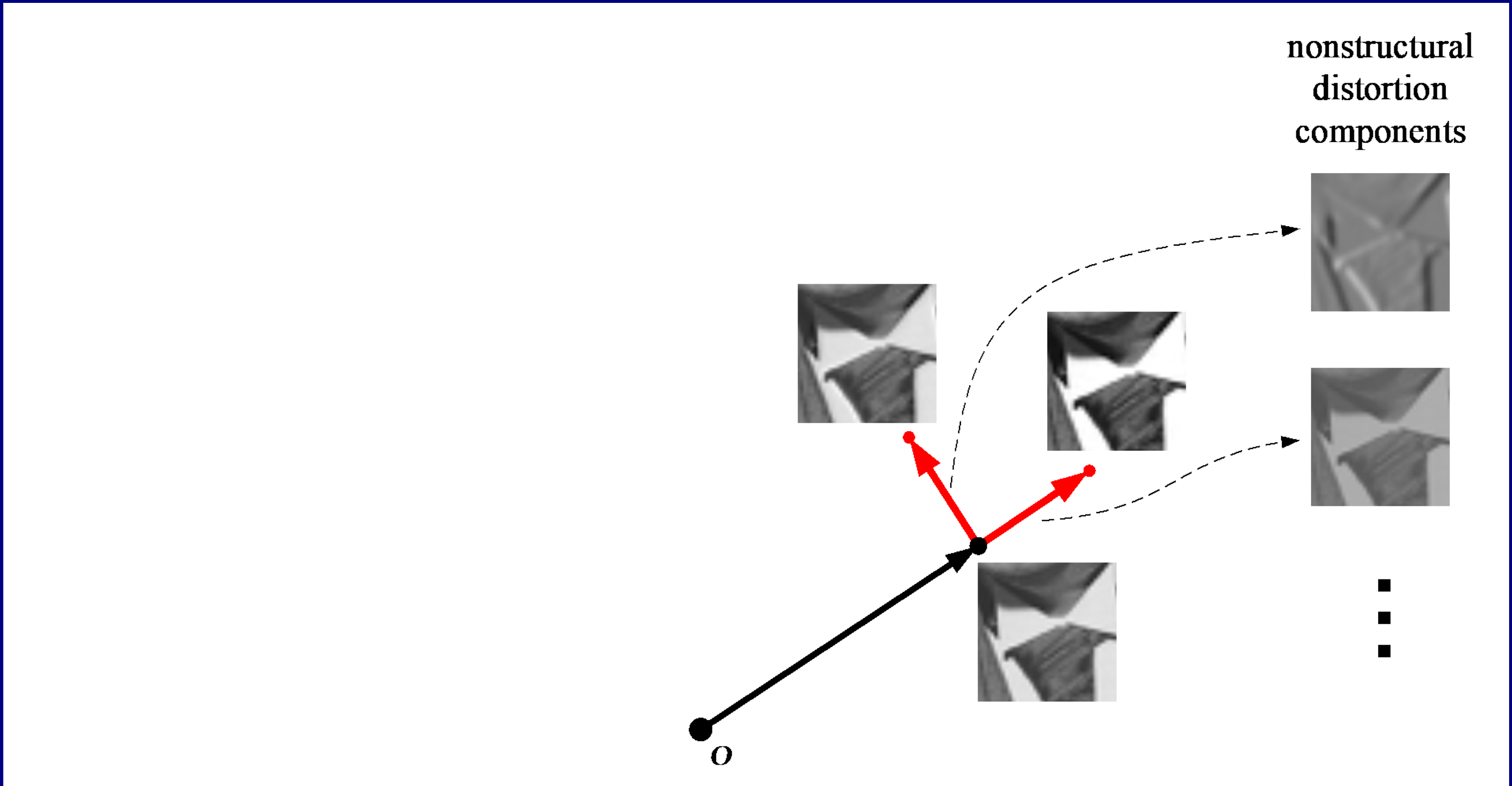
Intuition: Geometry in the Image Space

$$(1+\Delta C)\cdot\mathbf{x} = \mathbf{x} + \Delta C\cdot\mathbf{x}$$



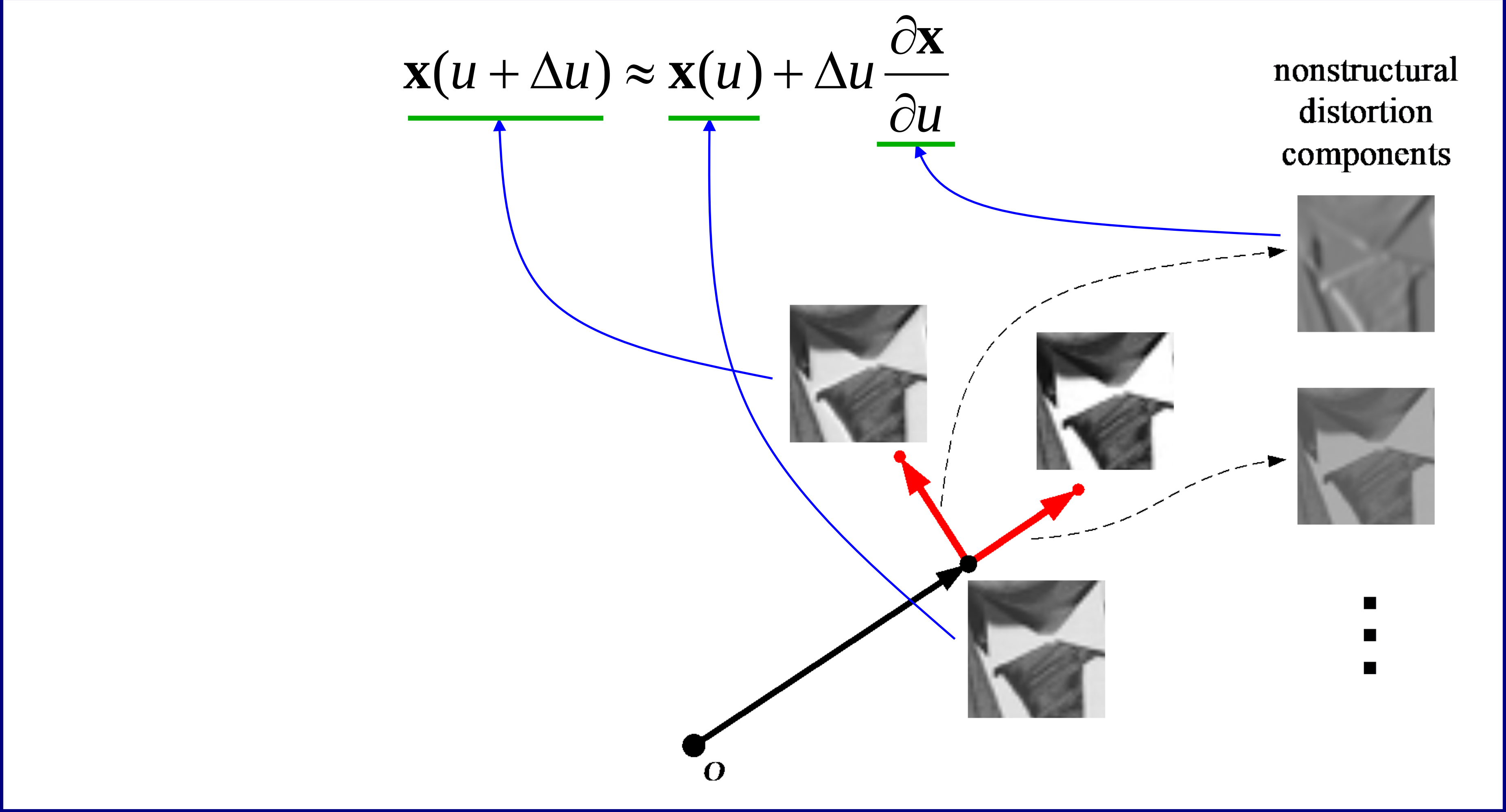
[Wang & Simoncelli, *ICIP* ’05]

Intuition: Geometry in the Image Space



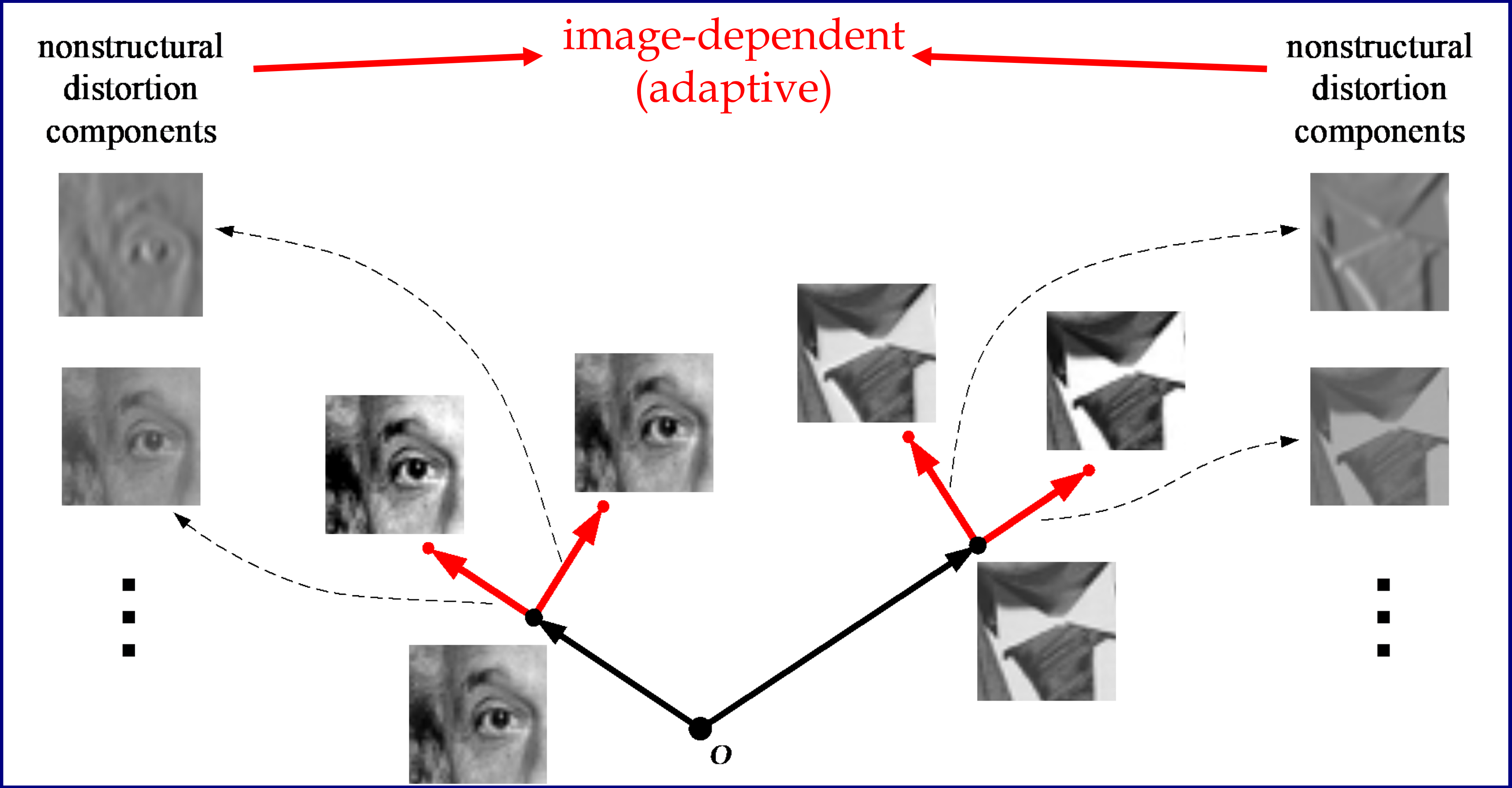
[Wang & Simoncelli, *ICIP* ’05]

Intuition: Geometry in the Image Space



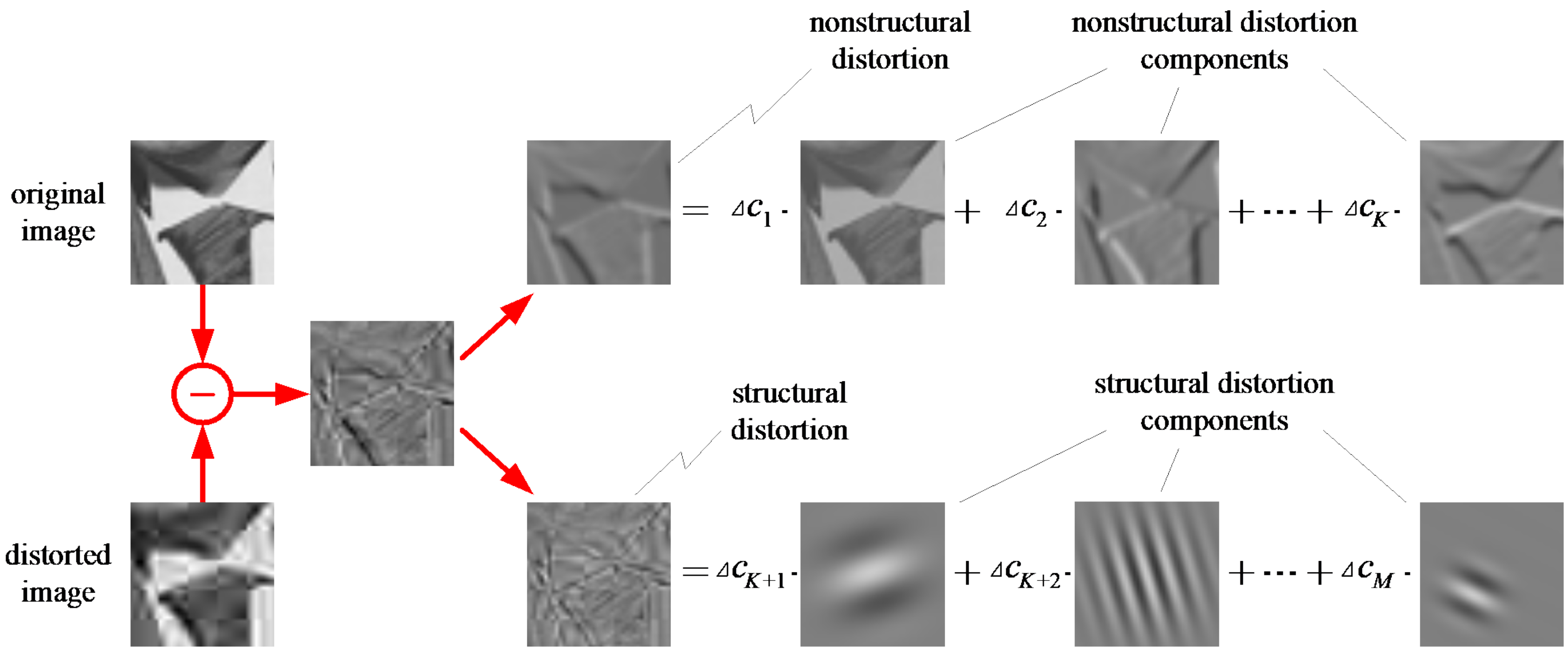
[Wang & Simoncelli, *ICIP* ’05]

Intuition: Geometry in the Image Space



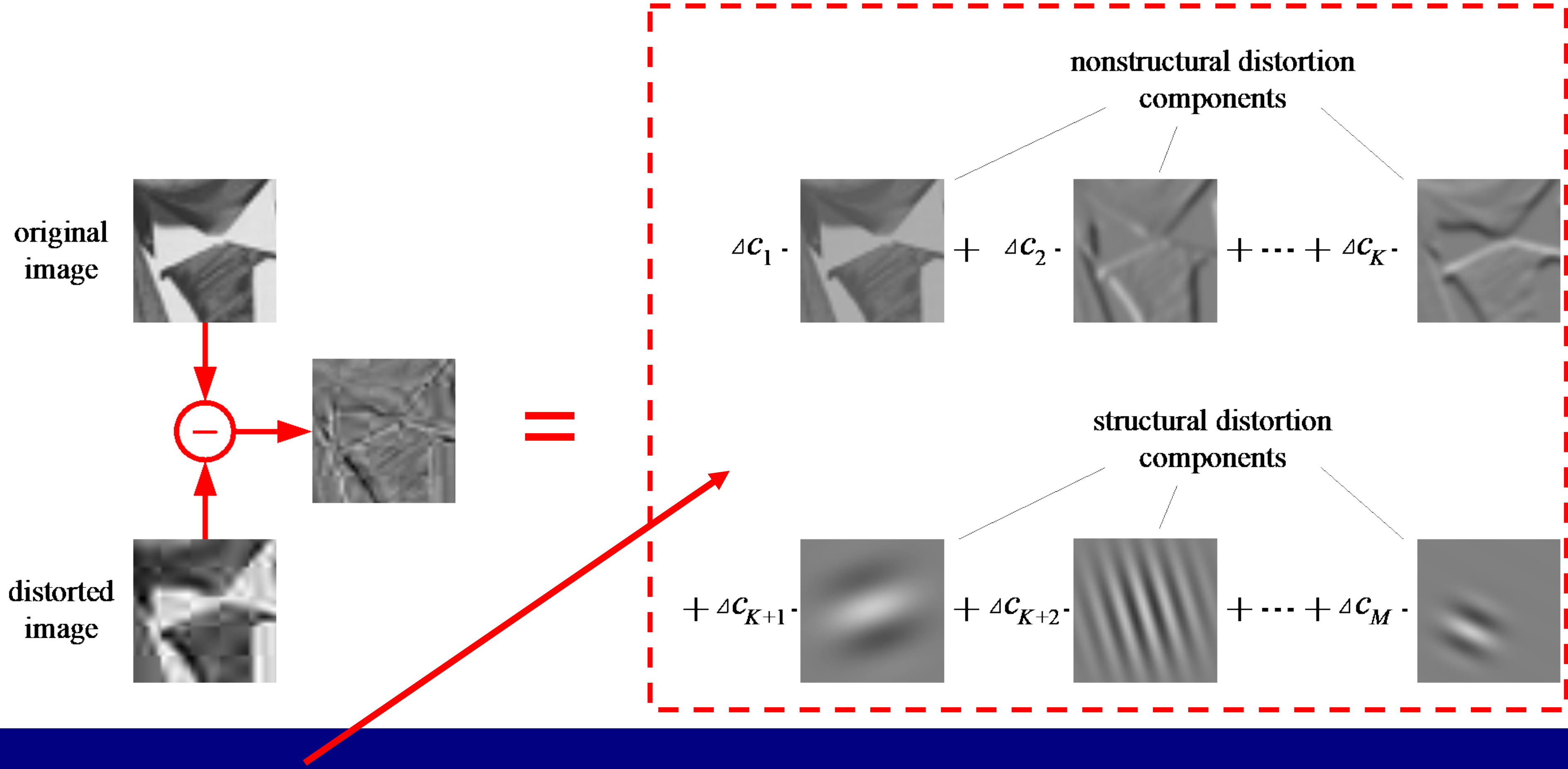
[Wang & Simoncelli, *ICIP* ’05]

Adaptive Linear System



[Wang & Simoncelli, *ICIP* ’05]

Adaptive Linear System

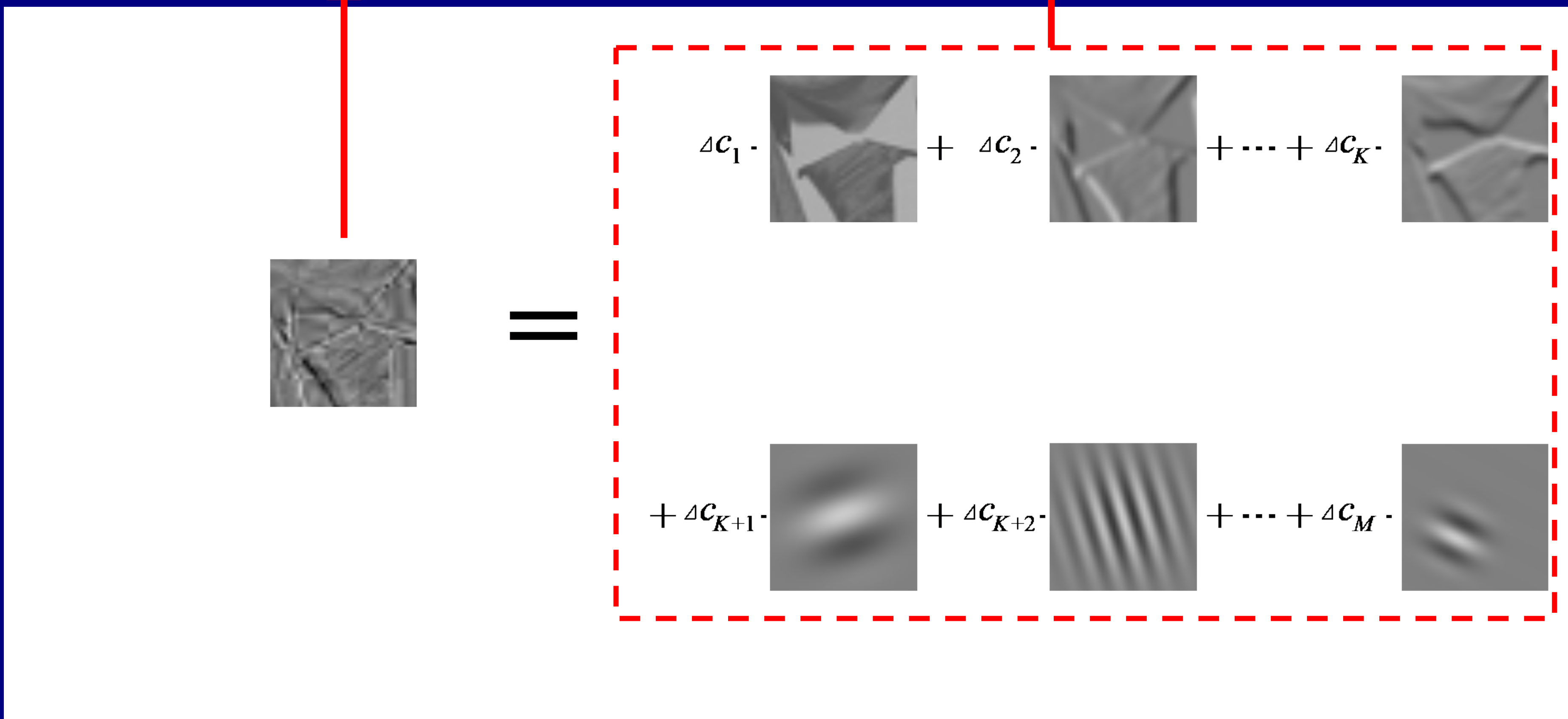


overcomplete, adaptive basis in the space of all images

[Wang & Simoncelli, *ICIP* '05]

Adaptive Linear System

$$\Delta\mathbf{x} = \Delta c_1 \mathbf{l}_1 + \Delta c_2 \mathbf{l}_2 + \cdots + \Delta c_K \mathbf{l}_K + \Delta c_{K+1} \mathbf{l}_{K+1} + \Delta c_{K+2} \mathbf{l}_{K+2} + \cdots \Delta c_M \mathbf{l}_M$$



[Wang & Simoncelli, *ICIP* ’05]

Adaptive Linear System

$$\Delta\mathbf{x} = \Delta c_1 \mathbf{l}_1 + \Delta c_2 \mathbf{l}_2 + \cdots + \Delta c_K \mathbf{l}_K + \Delta c_{K+1} \mathbf{l}_{K+1} + \Delta c_{K+2} \mathbf{l}_{K+2} + \cdots \Delta c_M \mathbf{l}_M$$

For each $\mathbf{l}_i$, define a weighting factor w_i

$$\mathbf{W} = diag\{w_1, w_2, \cdots w_M\}$$

$$\Delta\mathbf{x} = \mathbf{L}\Delta\mathbf{c}$$

[Wang & Simoncelli, *ICIP* '05]

Adaptive Linear System

$$\Delta\mathbf{x} = \Delta c_1\mathbf{l}_1 + \Delta c_2\mathbf{l}_2 + \cdots + \Delta c_K\mathbf{l}_K + \Delta c_{K+1}\mathbf{l}_{K+1} + \Delta c_{K+2}\mathbf{l}_{K+2} + \cdots \Delta c_M\mathbf{l}_M$$

For each $\mathbf{l}_i$, define a weighting factor w_i

$$\mathbf{W} = diag\{w_1, w_2, \cdots w_M\}$$

Find $\min_{\Delta\mathbf{c}} \left\|\mathbf{W}\Delta\mathbf{c}\right\|^2$ such that $\Delta\mathbf{x} = \mathbf{L}\Delta\mathbf{c}$

$$\boxed{D = \left\|\mathbf{W}^{-1}\mathbf{L}^T(\mathbf{L}\mathbf{W}^{-2}\mathbf{L}^T)^{-1}\Delta\mathbf{x}\right\|^2}$$

[Wang & Simoncelli, *ICIP* '05]

Adaptive Linear System

- **Direct solution:**

$$\boxed{D = \left\| \mathbf{W}^{-1}\mathbf{L}^T(\mathbf{L}\mathbf{W}^{-2}\mathbf{L}^T)^{-1}\Delta\mathbf{x} \right\|^2}$$

- **Simplified solution:**

Split into adaptive and non-adaptive parts:

$$\mathbf{L} = [\mathbf{A} \;\; \mathbf{B}] \qquad \Delta\mathbf{c} = [\Delta\mathbf{c}_A^T \;\; \Delta\mathbf{c}_B^T]^T \qquad \mathbf{W} = diag\{\mathbf{W}_A, \mathbf{W}_B\}$$

New solution:

$$\boxed{\begin{gathered} \mathbf{G} = \mathbf{B}^{-T}\mathbf{W}_B^2\mathbf{B}^{-1} \\ \Delta\mathbf{c}_A = (\mathbf{W}_A^2 + \mathbf{A}^T\mathbf{G}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{G}\Delta\mathbf{x} \\ D = \left\|\mathbf{W}_A\Delta\mathbf{c}_A\right\|^2 + \left\|\mathbf{W}_B\mathbf{B}^{-1}(\Delta\mathbf{x} - \mathbf{A}\Delta\mathbf{c}_A)\right\|^2 \end{gathered}}$$

- **Relationship to existing methods:**

differential optical flow estimation → special cases of the framework
tangent distance → special cases of the framework

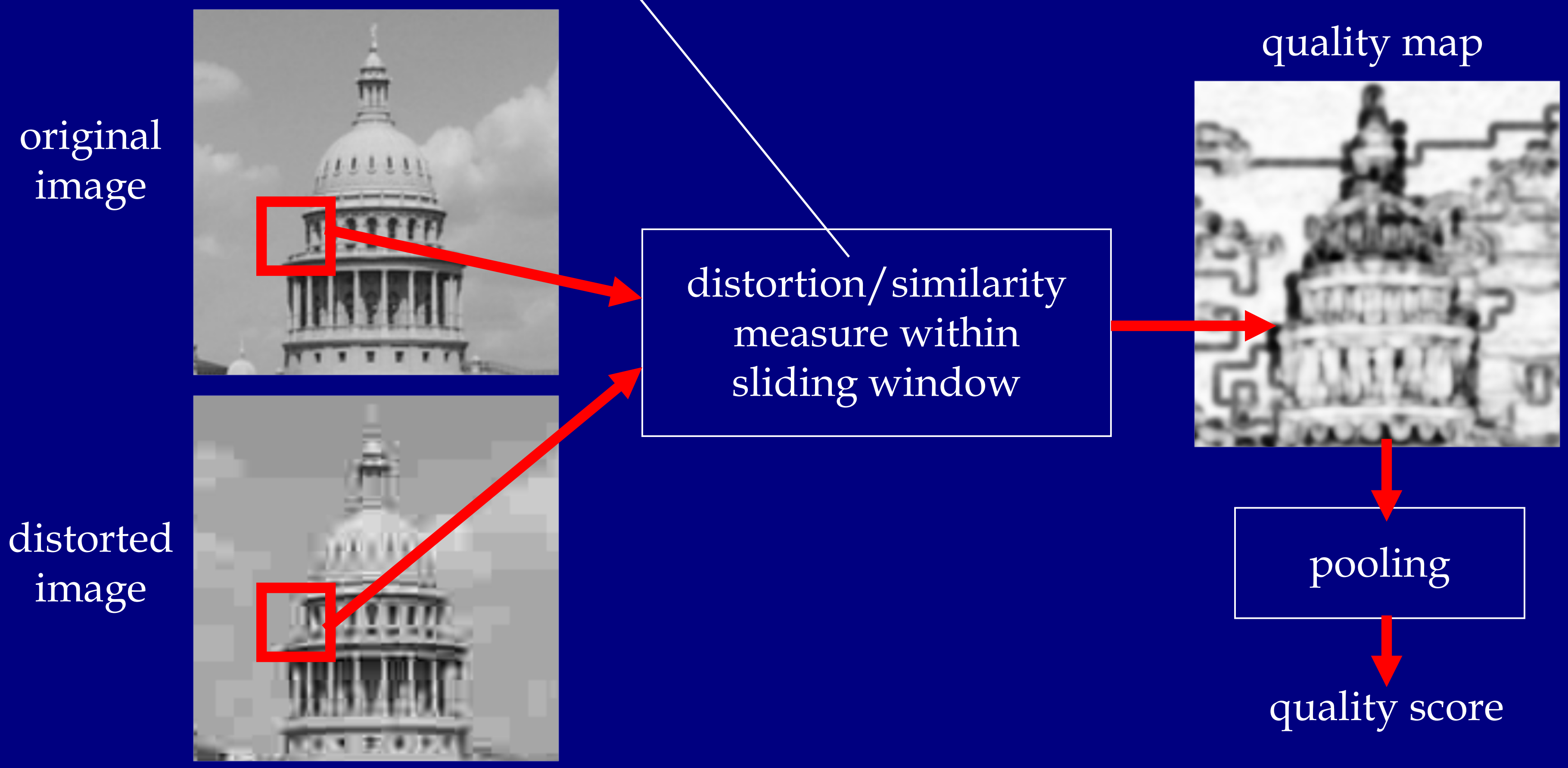
… …

[Wang & Simoncelli, *ICIP* ’05]

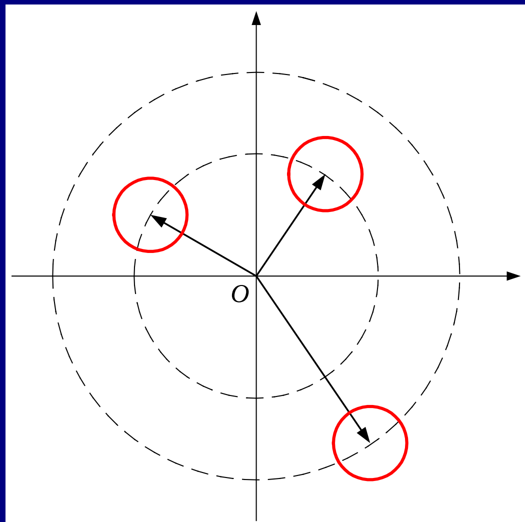
Structural Similarity (SSIM) Index

$$S(\mathbf{x},\mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

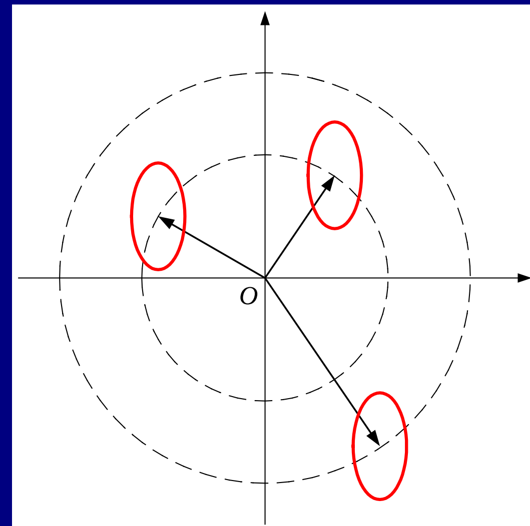
[Wang & Bovik, *IEEE Signal Proc. Letters*, ’02]
[Wang *et al.*, *IEEE Trans. Image Proc.*, ’04]



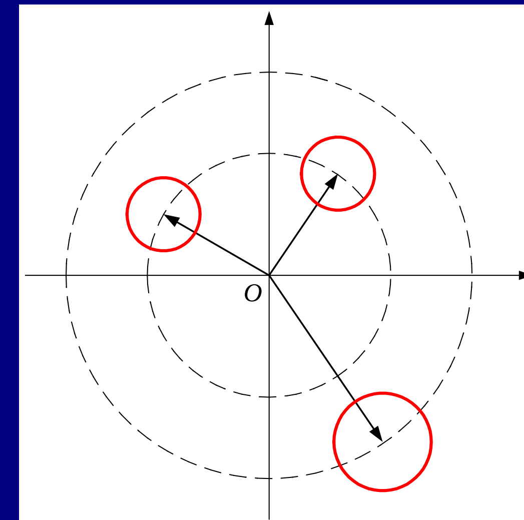
Model Comparison: Distortion Level Sets



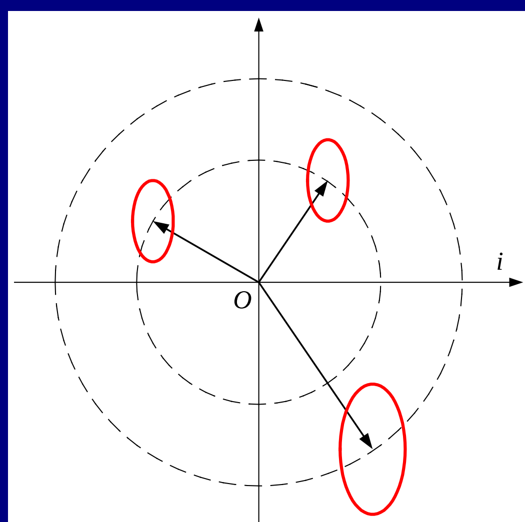
Minkowski (MSE)



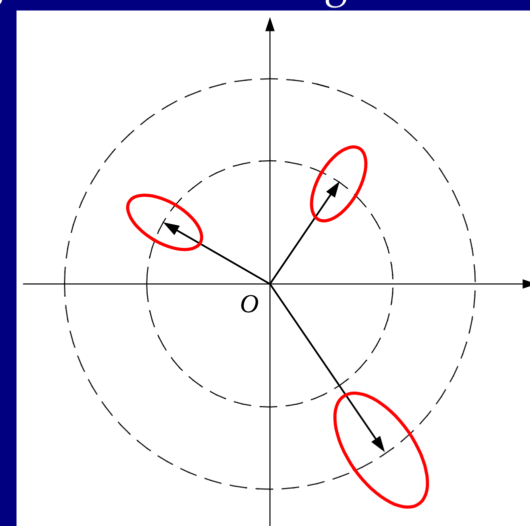
component-weighted



magnitude-weighted



magnitude and component-weighted



proposed adaptive system

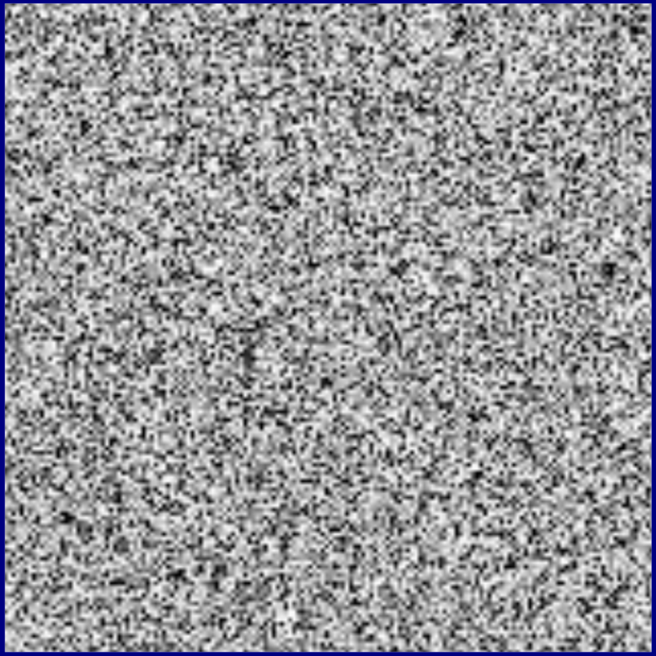
[Wang *et al.*, *IEEE Trans. Image Proc.*, ’04]

Gaussian noise corrupted image

original image

SSIM index map

absolute error map

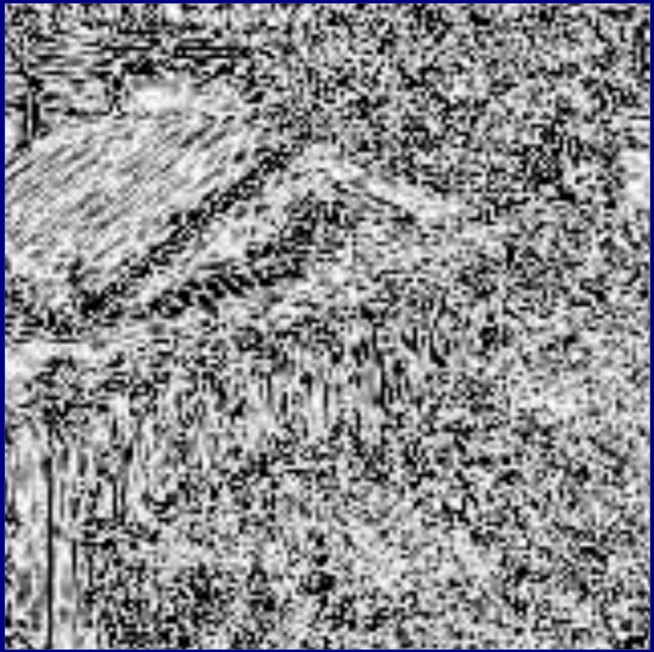


JPEG2000 compressed image

original image

SSIM index map

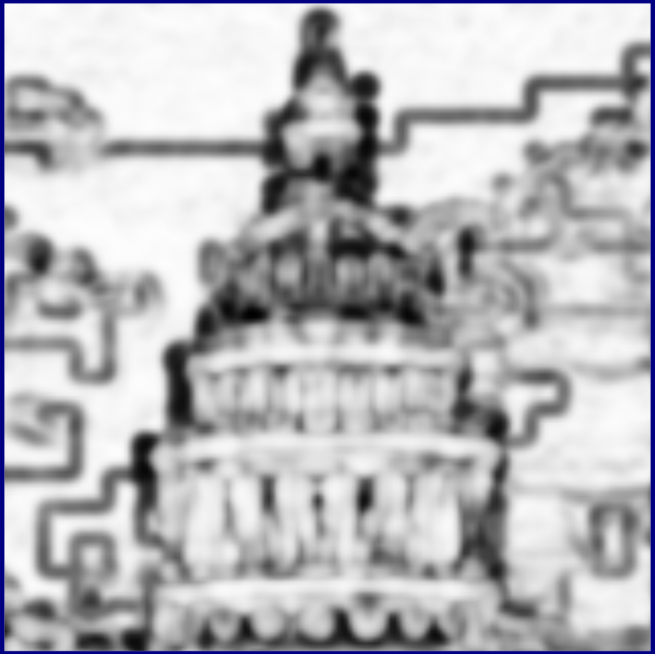
absolute error map



JPEG compressed image

original image

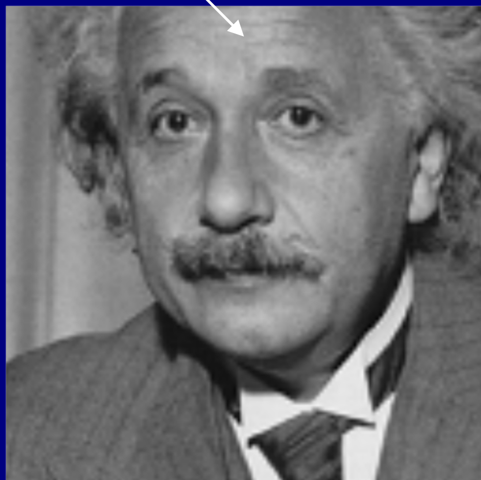
SSIM index map



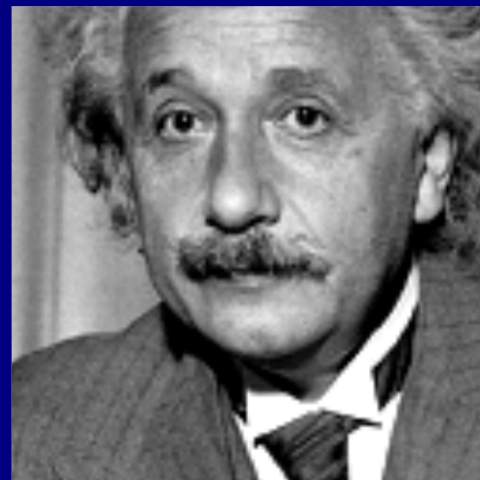
absolute error map

Motivation

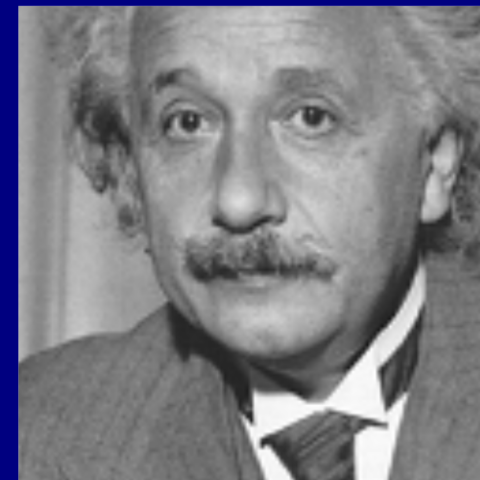
original Image



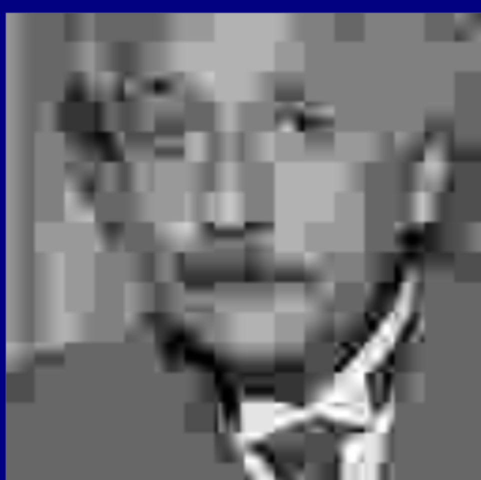
MSE=0, MSSIM=1



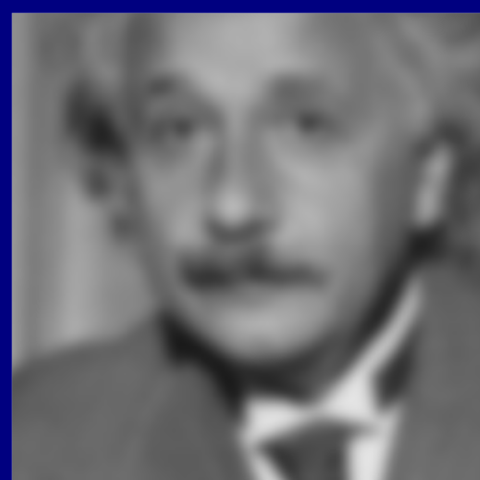
MSE=309, MSSIM=0.928



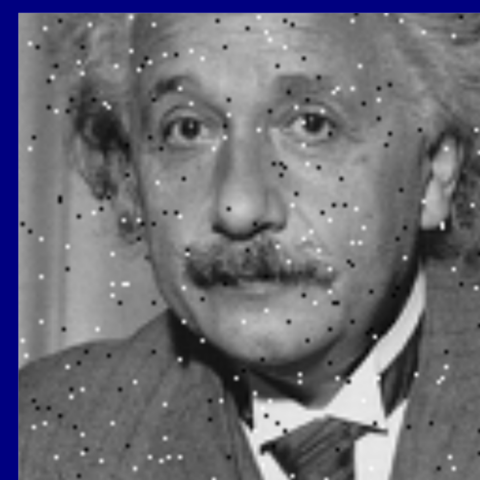
MSE=309, MSSIM=0.987



MSE=309, MSSIM=0.580

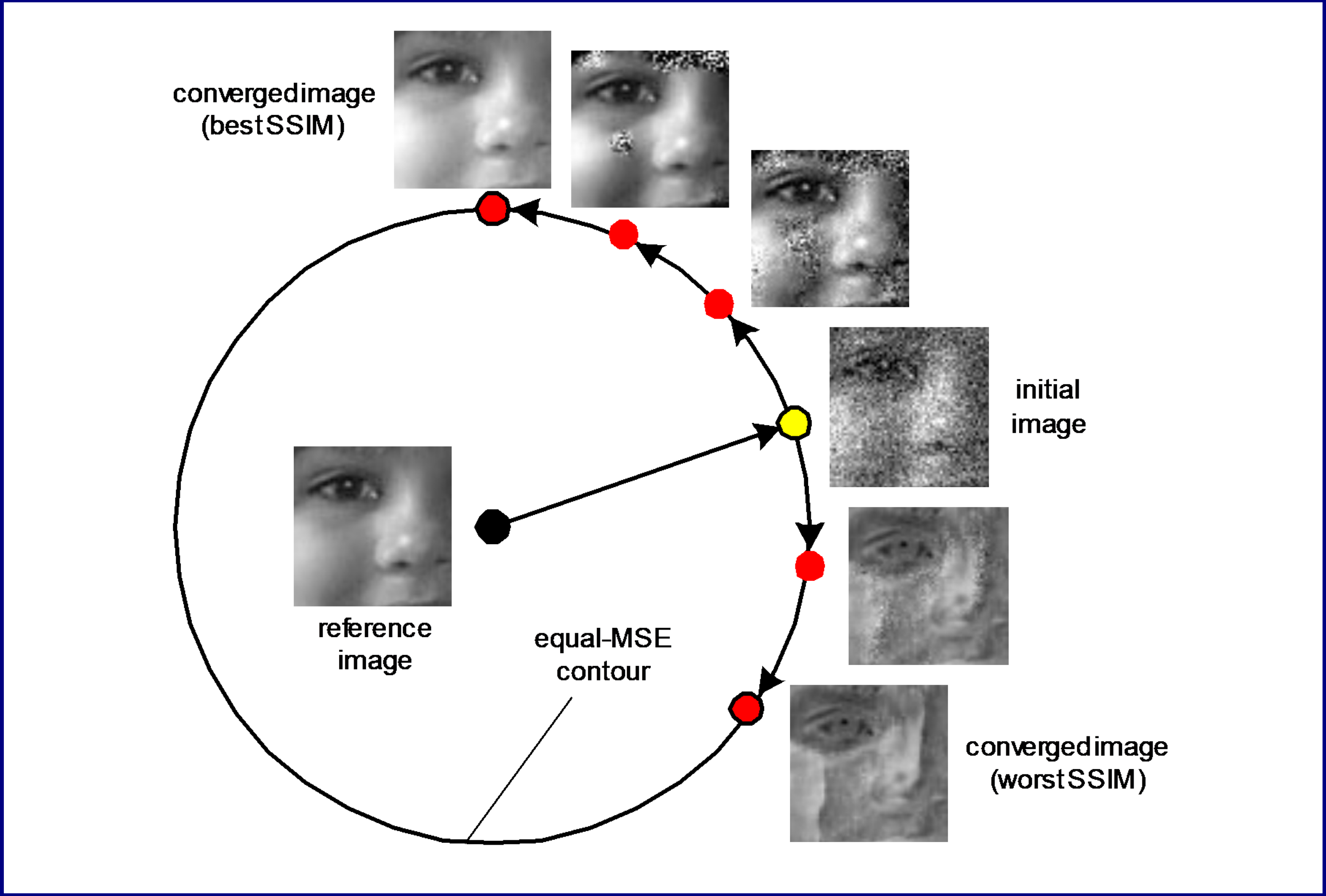


MSE=308, MSSIM=0.641



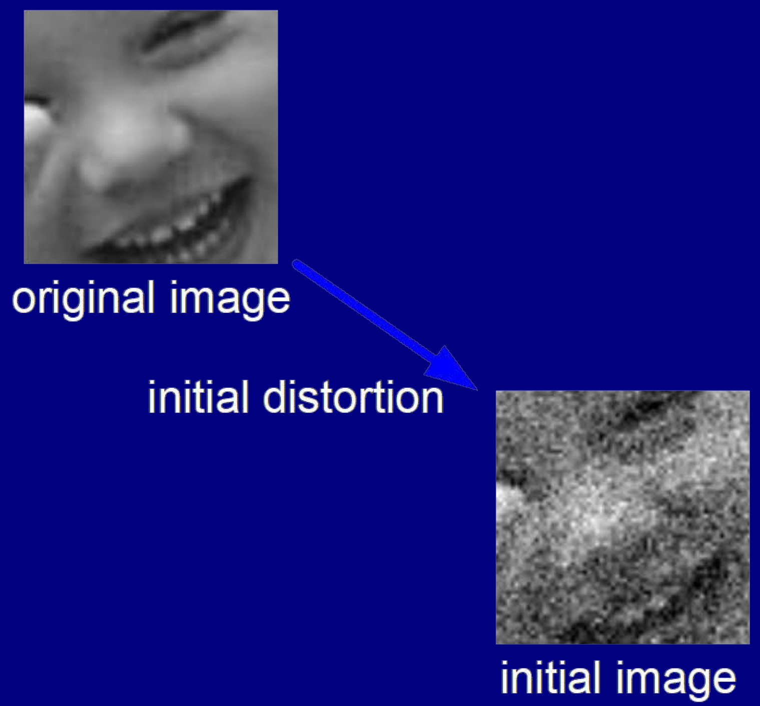
MSE=309, MSSIM=0.730

MAD Competition: MSE vs. SSIM



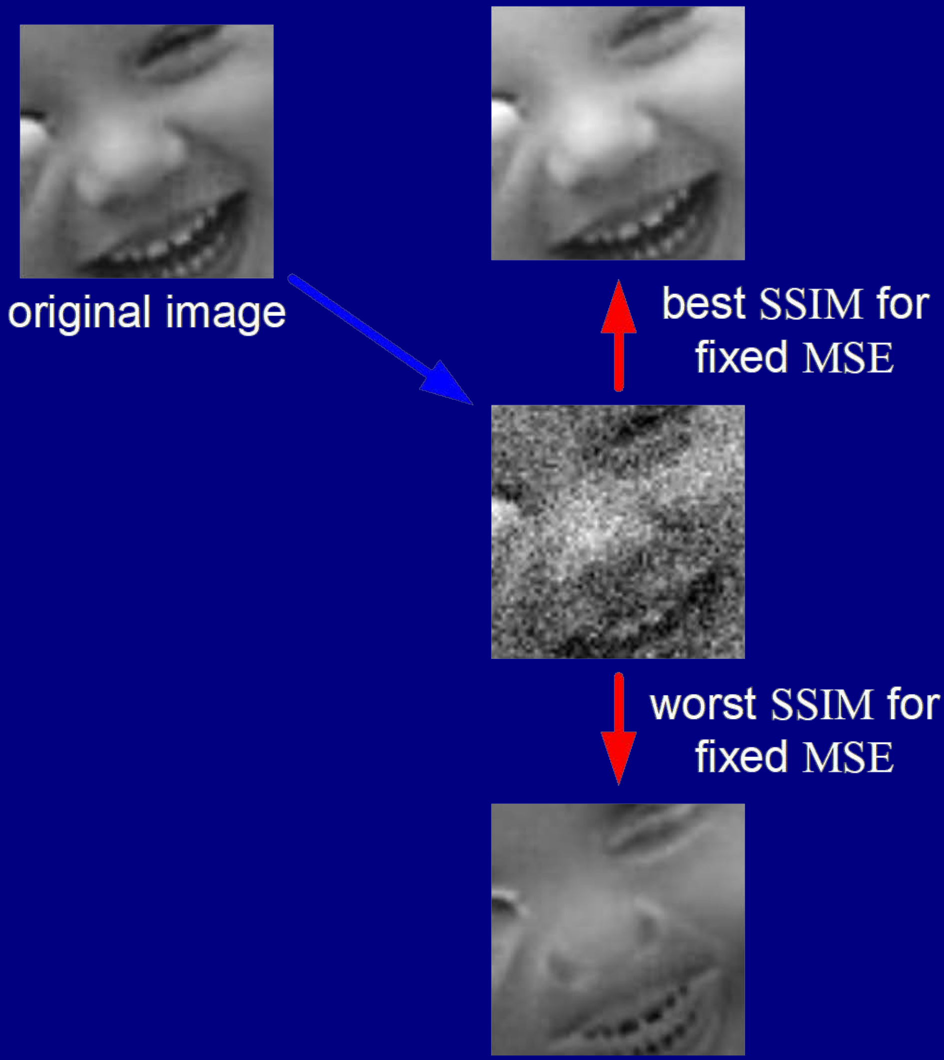
[Wang & Simoncelli ’04]

MAD Competition: MSE vs. SSIM



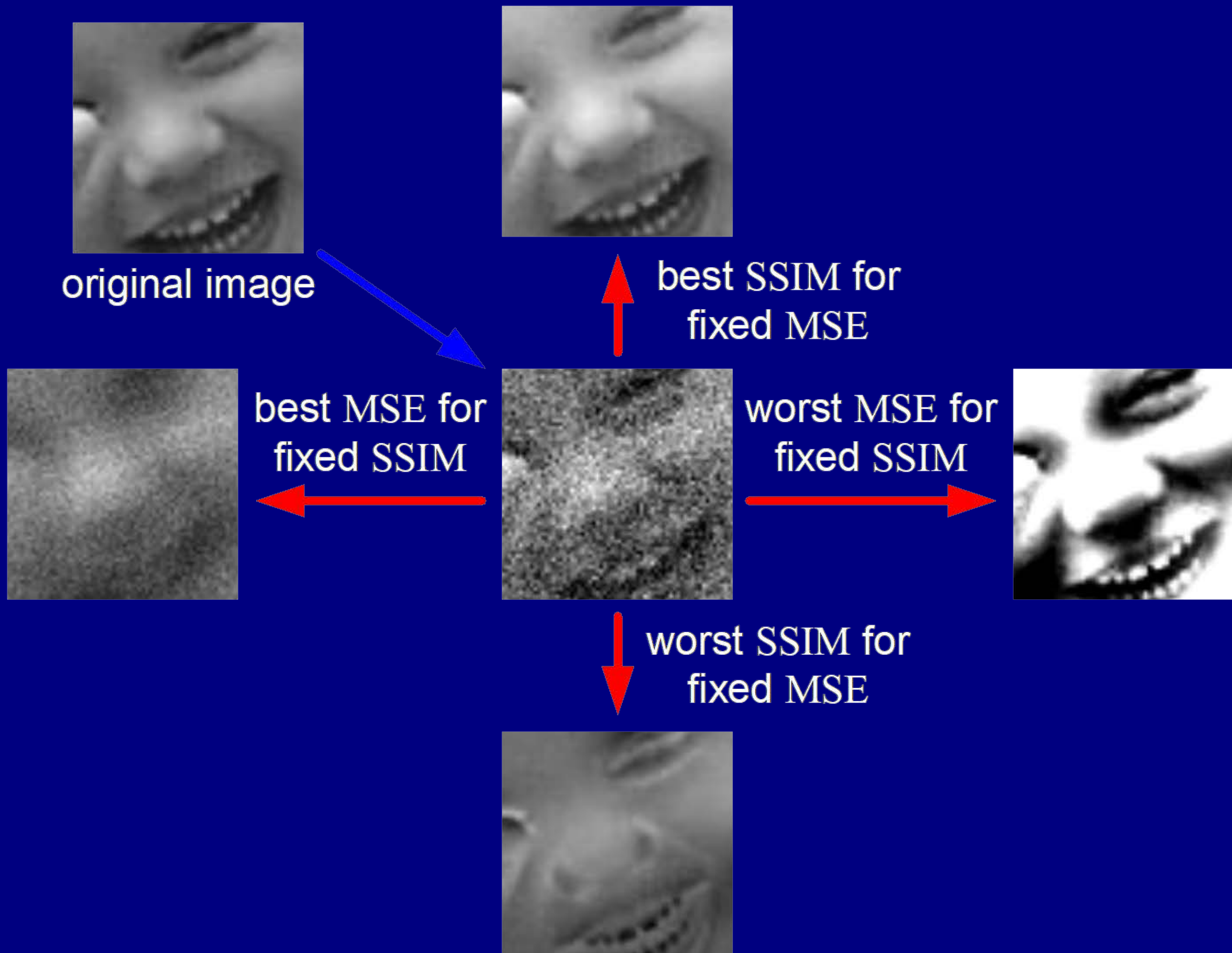
[Wang & Simoncelli ’04]

MAD Competition: MSE vs. SSIM



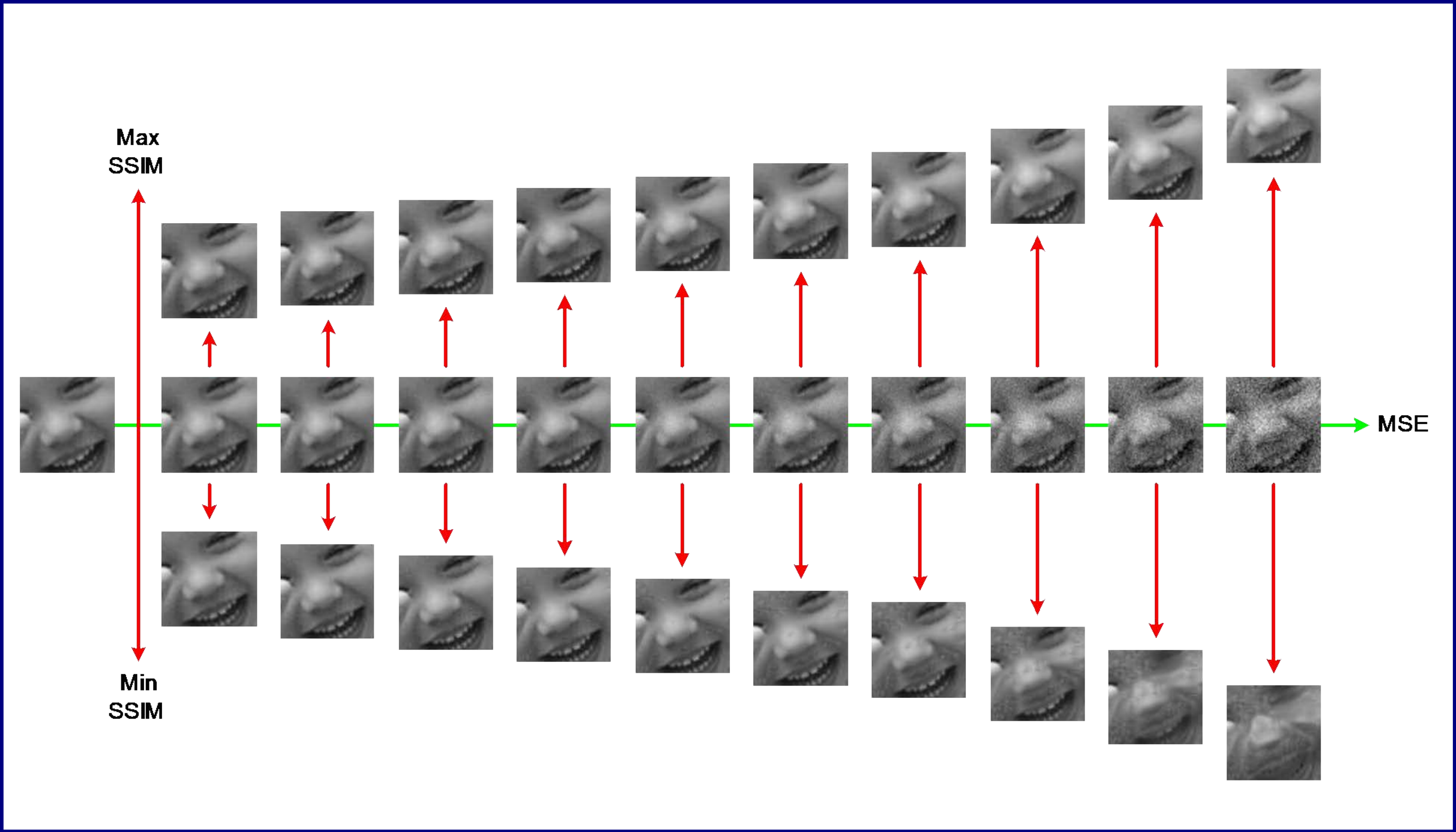
[Wang & Simoncelli ’04]

MAD Competition: MSE vs. SSIM



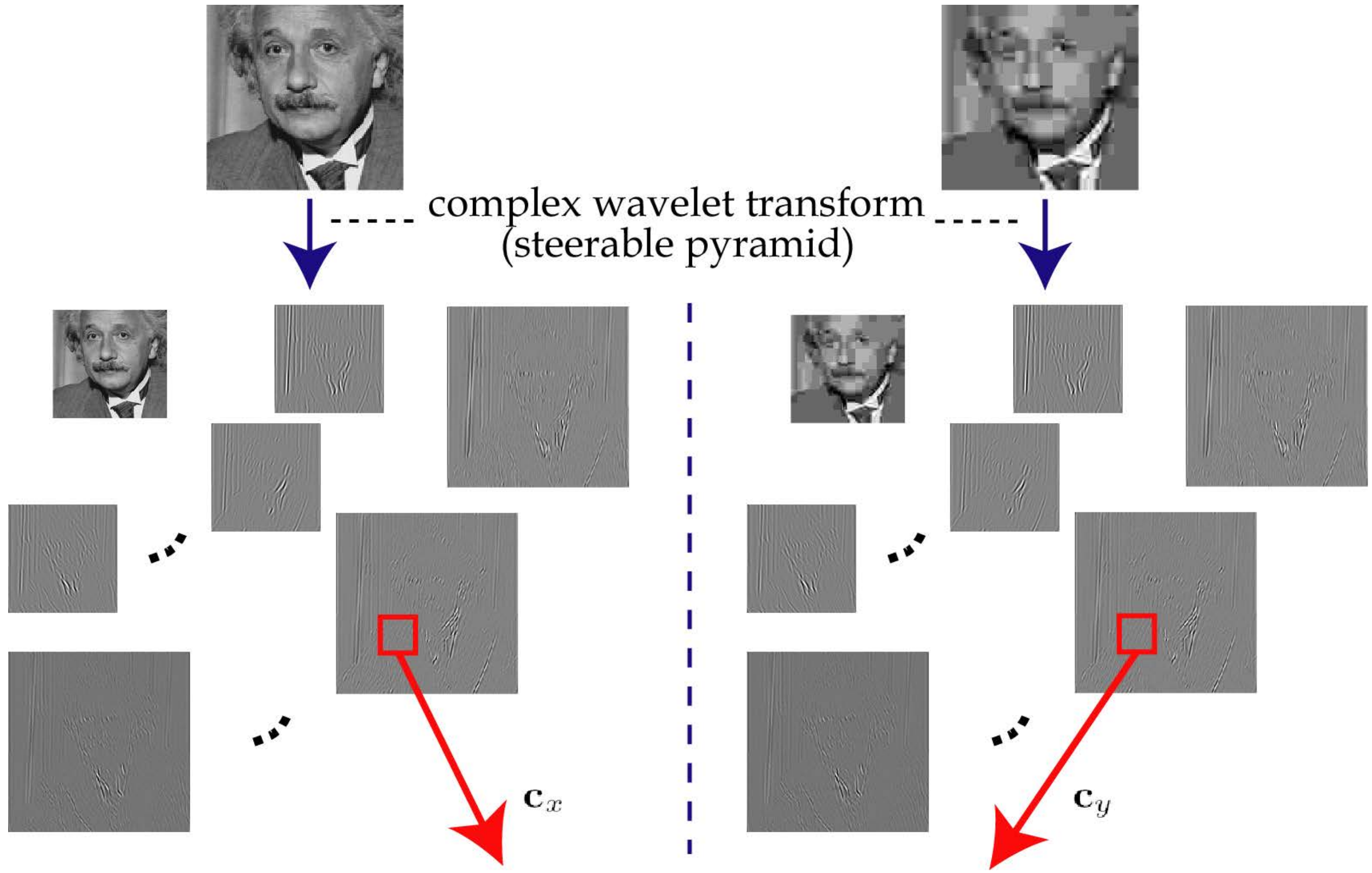
[Wang & Simoncelli '04]

MAD Competition: MSE vs. SSIM



[Wang & Simoncelli ’04]

Extension: Complex Wavelet SSIM



$$\tilde{S}(\mathbf{c}_x, \mathbf{c}_y) = \frac{2\sum_{i=1}^{N} |c_{x,i}||c_{y,i}| + K}{\sum_{i=1}^{N} |c_{x,i}|^2 + \sum_{i=1}^{N} |c_{y,i}|^2 + K} \cdot \frac{2\left|\sum_{i=1}^{N} c_{x,i}\, c_{y,i}^*\right| + K}{2\sum_{i=1}^{N} |c_{x,i}\, c_{y,i}^*| + K}$$

[Wang & Simoncelli, *ICASSP* '05]

Image Matching without Registration

Standard patterns: 10 images

Database: 2430 images

Correct Recognition Rate:

MSE: 59.6%   Complex wavelet SSIM: **97.7%**

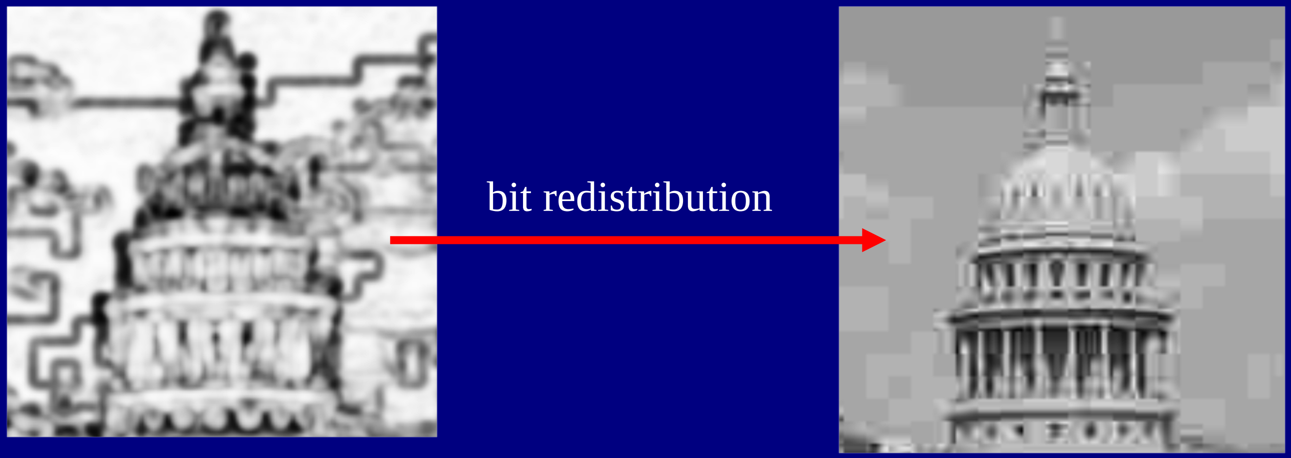
[Wang & Simoncelli, *ICASSP* '05]

Perceptual Image Compression

- **General Idea**

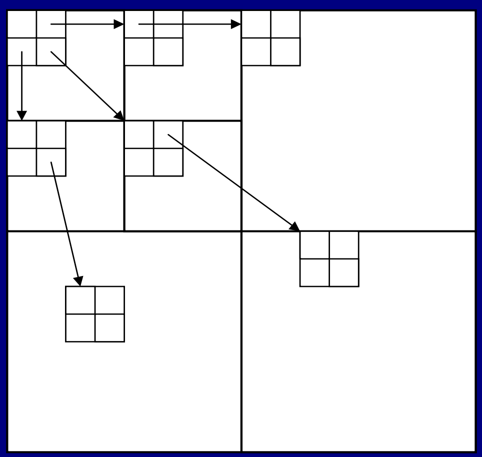
  - Transform image signal into "perceptually uniform" space

- **Implementations**
  - Perceptual weighting + uniform quantization
  - Equivalently, *perceptually adaptive* quantization
  - Net effect: bit redistribution, perceptually

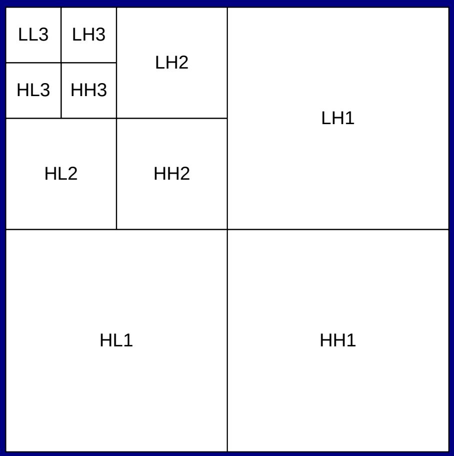


Perceptual Image Compression

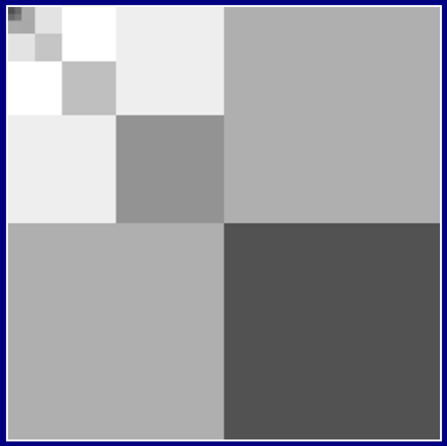
- **Frequency Weighting**
  – JPEG quantization table; JPEG2000 subband weighting



wavelet tree



wavelet subbands



subband weighting

- **Masking**
  – JPEG2000 neighborhood/self masking

Perceptual Image Denoising

Goal: estimate random variable (vector) $\boldsymbol{x}$

Given observation $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$, find estimator $\hat{\boldsymbol{x}}(\boldsymbol{y})$, such that

Bayes Least Square (BLS)

$$\hat{\boldsymbol{x}}_{BLS} = \arg\min_{\hat{\boldsymbol{x}}} E[(\boldsymbol{x} - \hat{\boldsymbol{x}})^2 \mid \boldsymbol{y}]$$

Solution: $\hat{\boldsymbol{x}}_{BLS} = E[\boldsymbol{x} \mid \boldsymbol{y}]$

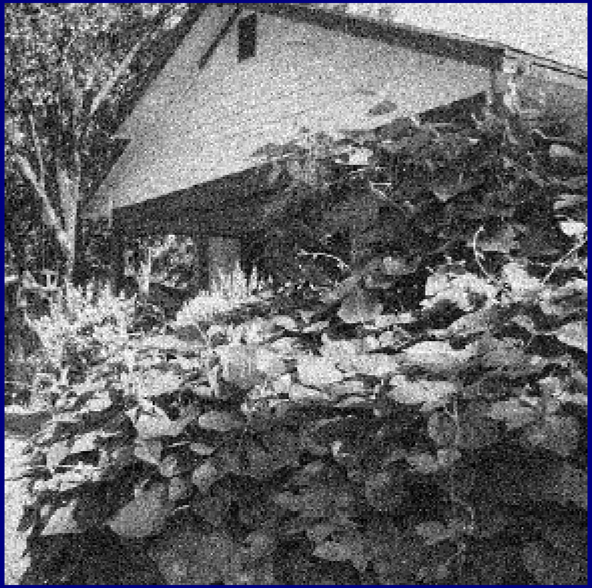
Bayes Max SSIM (BMS)

$$\hat{\boldsymbol{x}}_{BMS} = \arg\max_{\hat{\boldsymbol{x}}} E[SSIM(\boldsymbol{x}, \hat{\boldsymbol{x}}) \mid \boldsymbol{y}]$$

What's the solution?

Perceptual Image Denoising

original image



noisy image

least-square denoised image

max-SSIM denoised image

[Channappayya & Bovik '06]

Other Perceptually-Inspired IP System Design

- **Video Compression**
  - Quantizer design
  - Rate control: macro-block, frame, and group-of-pictures levels
- **Image Restoration**
- **Image Enhancement**
- **High Dynamic Range Image Compression**
- **Image Interpolation and Inpainting**
- **Image Halftoning**
- **Image (texture) Segmentation and Classification**
- **Image (texture) Synthesis**
- **……**

Real-World Challenges

- **Example: Video Distribution Industry**
  - Poor video quality: No. 1 reason why viewers stop watching
    66% viewers abandon a video when playback quality is poor, more than too many ads (62%) or multiple pauses (51%)
    [Limelight Networks, Dec. 2015]

- **Reality**
  - Real VQA restricted to lab environment or small-scale testing
  - QoS factors (bitrate, packet loss, …) used in network
  - Pseudo QoE measure (rebuffering, …) used at end user devices
  - True QoE assessment is limited (at hosting server) or missing (at end user devices)

Real-World Challenges

- **Accuracy**
  - High correlation with human opinions
- **Complexity**
  - Faster than real-time and low implementation/storage cost
- **Ease**
  - Easy-to-understand and easy-to-use

Real-World Challenges

- **Cross-Device Assessment**

reference

test images (shown on different devices)

the same reference image

IQA?

Real-World Challenges

- **Cross-Resolution Assessment**

reference

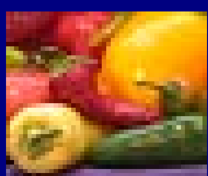
test images
(scaled & transcoded)

IQA?

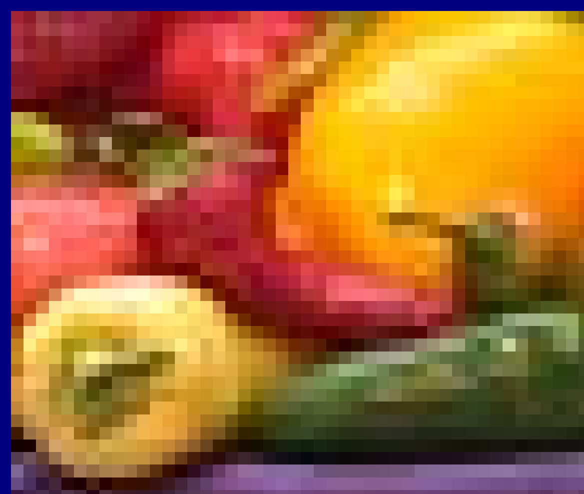
Real-World Challenges

- **Cross-Resolution Assessment**

reference



test images (interpolated)



IQA?

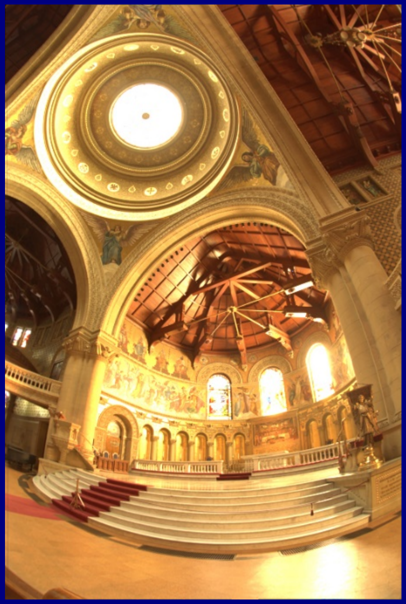
Real-World Challenges

- **Cross-Dynamic Range Assessment**

reference

test images (tone mapped)

HDR
Image



IQA?

Real-World Challenges

- **Multi-Exposure Fusion**

reference

test image (fused)

IQA?

Real-World Challenges

- **Color-to-Gray Conversion**

reference

test images

IQA?