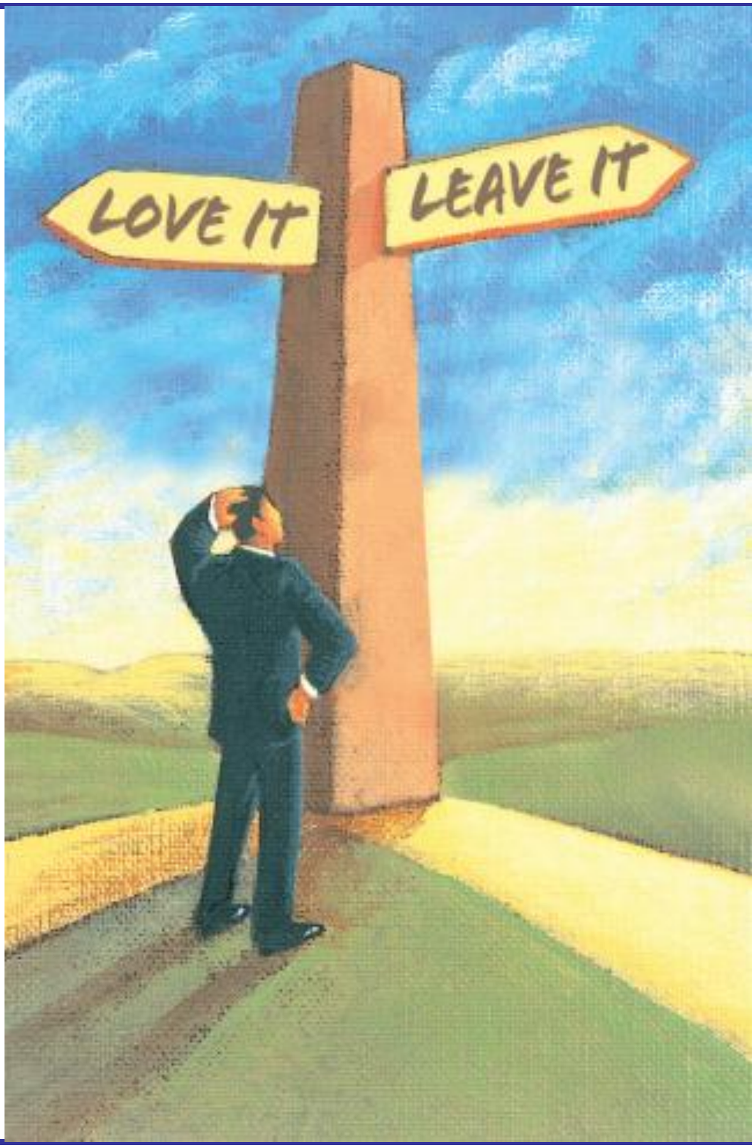
Mean Squared Error: Love it or Leave it?

An **Objective** Look at Image Quality Assessment

Santosh Chapaneri,
Assistant Professor,
EXTC, SFIT



Mean Squared Error:
Love it or Leave it?

Talk adapted primarily from:

Z. Wang, A. C. Bovik, “Mean Squared Error: Love it or Leave it?”, ***IEEE Signal Processing Magazine***, pp. 98-117, Jan **2009**

Image Quality

- Image Quality Assessment – one of the challenging field of digital image processing

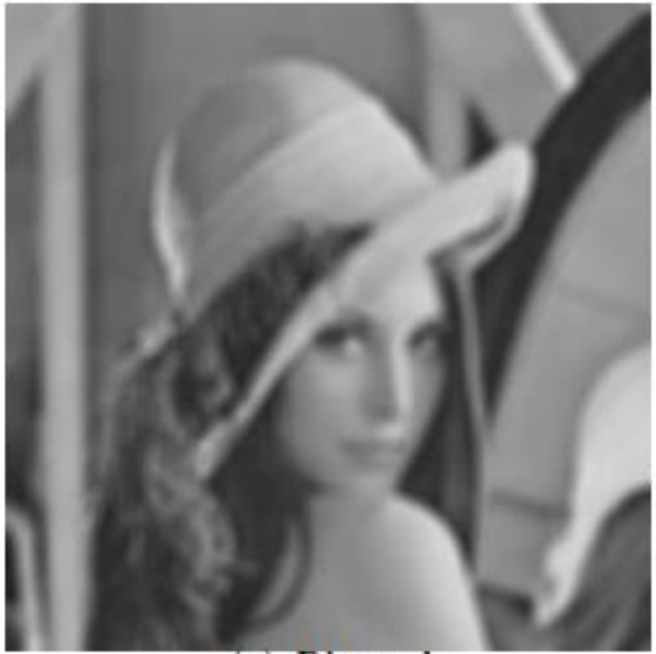
- Explosion of Video and Multimedia applications
  - **Biomedical Imaging**
  - **Video Telephony**
  - **Digital Cinema**
  - **HDTV**
  - Video over IP networks: **YouTube**, streaming video
  - Video over Wireless networks: in-flight entertainment, **Mobile TV**
  - Consumer Electronics: **Video Camcorders**

Quality

Comparison of different ***perceptual*** quality

(a) Reference image

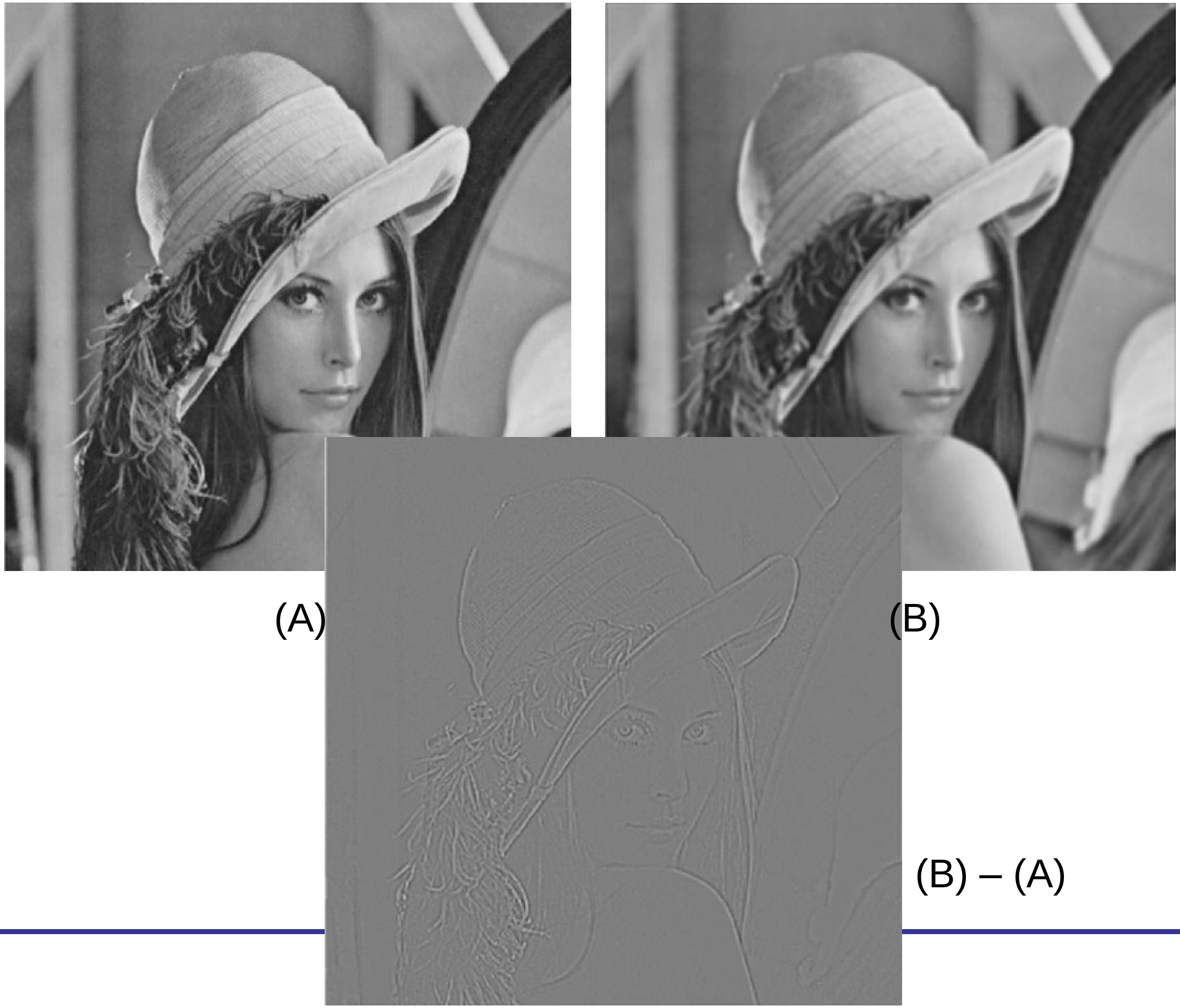
(b) Contrast enhancement



(c) Blurred

(d) JPEG compressed

Image Quality

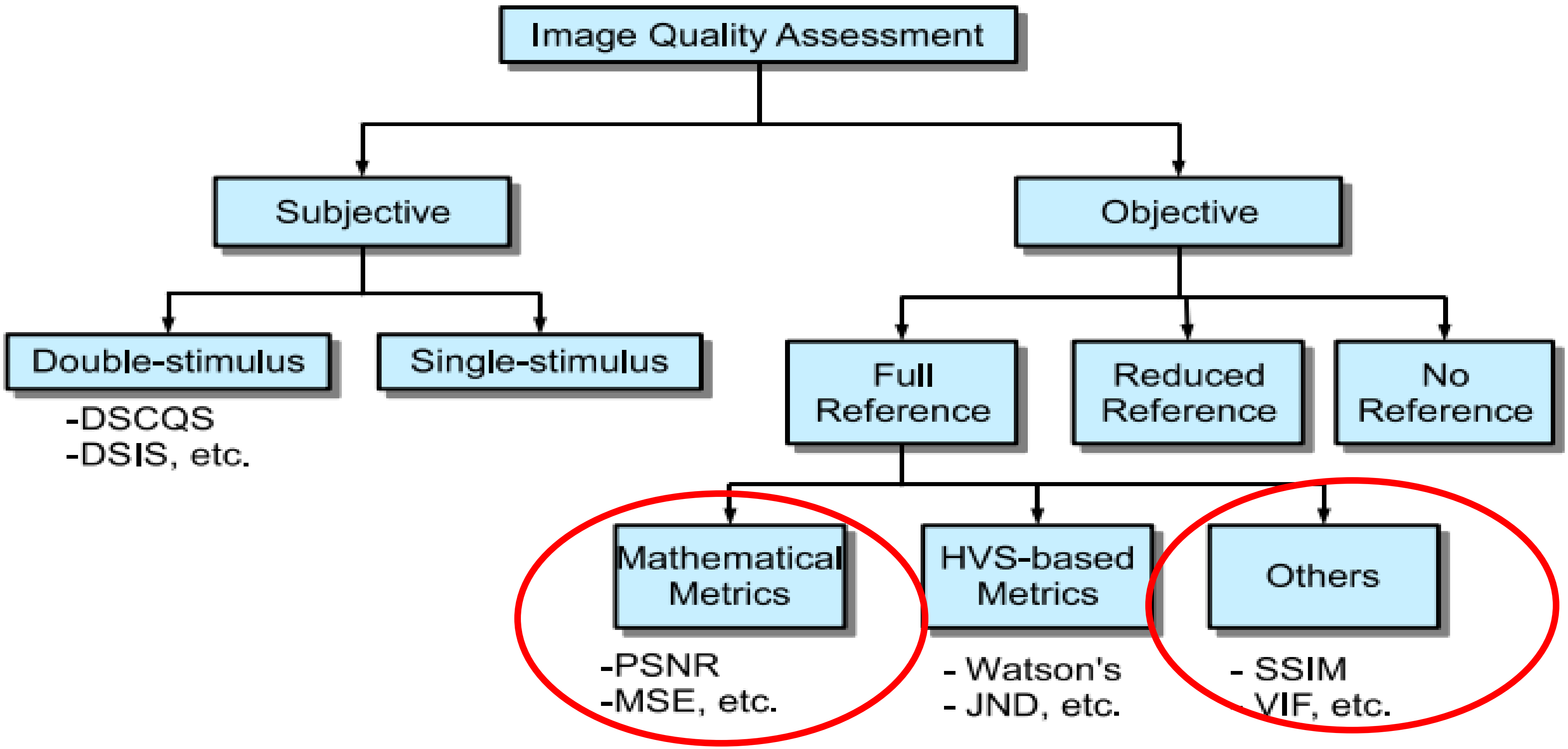


(A) (B)

(B) – (A)

Image Quality

- Several techniques exist for measuring image quality



Subjective Quality Assessment

- The best way to find quality of an image is to look at it because **human eyes are the ultimate viewer**.
- Subjective image quality is concerned with how image is perceived by a viewer and give his or her opinion on a particular image.

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

- Too Inconvenient, time consuming and expensive

Objective Quality Assessment

Philosophy

distorted signal = reference signal + error signal

Assume reference signal has perfect quality

Quantify **perceptual** error visibility

- **Representative work**
  - 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]

Objective Quality Metrics

- **Goal of Signal Fidelity measure**: provide a quantitative score to describe the degree of similarity between two signals
- Usually, one signal is pristine, and the other is distorted

- Suppose that $\boldsymbol{x} = \{x_i \mid i = 1, 2, \ldots, N\}$ and $\boldsymbol{y} = \{y_i \mid i = 1, 2, \ldots, N\}$ are two finite-length discrete signals (eg. visual images); N = number of signal samples (pixels)

- The MSE between these two signals is

$$\mathrm{MSE}(\mathrm{x}, \mathrm{y}) = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2$$

- The error signal is $e_i = x_i - y_i$

Objective Quality Metrics

Minkowski Error

$$d_p(\mathrm{x}, \mathrm{y}) = \left(\sum_{i=1}^{N} |e_i|^p \right)^{1/p}$$

$$\boxed{\mathrm{PSNR} = 10 \log_{10} \frac{L^2}{\mathrm{MSE}}}$$

For an 8bits/pixel monotonic signal, L is equal to 255.

- **MSE** – has remained the dominant quantitative performance metric in image and video processing
- **MSE** – ubiquitous preference for design engineers seeking to optimize their algorithms

Secret of MSE?

- Why is MSE so popular?
  - aka Why do we love MSE?
- When does MSE metric fail?
  - aka Should we leave MSE?
- If not MSE, what other metrics could be used?

Why do we Love MSE?

$$\mathrm{MSE}(\mathbf{x}, \mathbf{y}) = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2$$

- Simple & inexpensive to compute
- Memory-less
- Satisfies the conditions for distance metric
  - — nonnegativity: $\mathrm{MSE}(\mathbf{x}, \mathbf{y}) \geq 0$
  - — identity: $\mathrm{MSE}(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$
  - — symmetry: $\mathrm{MSE}(\mathbf{x}, \mathbf{y}) = \mathrm{MSE}(\mathbf{y}, \mathbf{x})$
  - — triangular inequality: $\mathrm{MSE}(\mathbf{x}, \mathbf{z}) \leq \mathrm{MSE}(\mathbf{x}, \mathbf{y}) + \mathrm{MSE}(\mathbf{y}, \mathbf{z})$
- Clear meaning – energy of error signal (satisfies Parseval's theorem)

So what’s wrong with MSE?

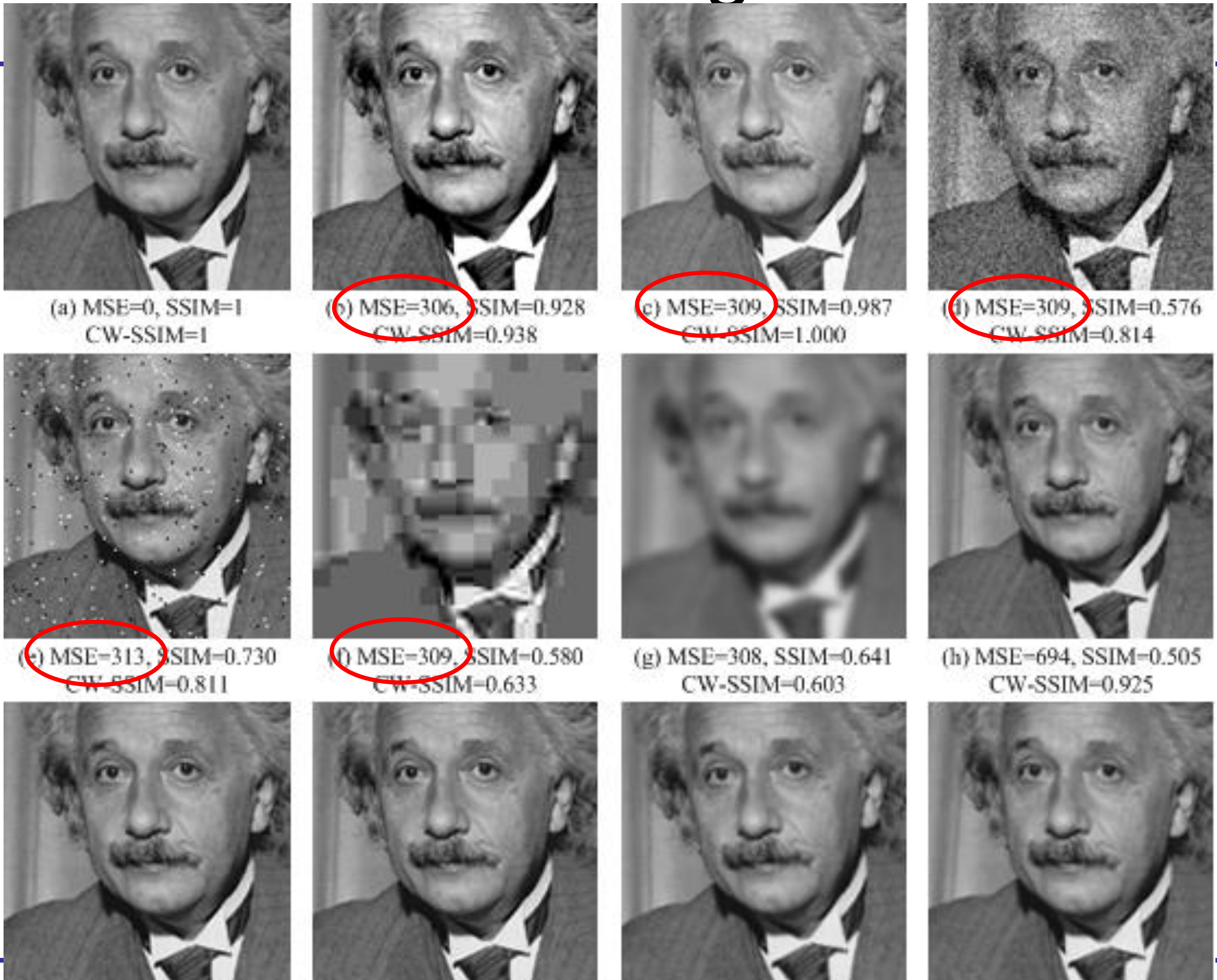
(a)

(b)

(c)

(d)

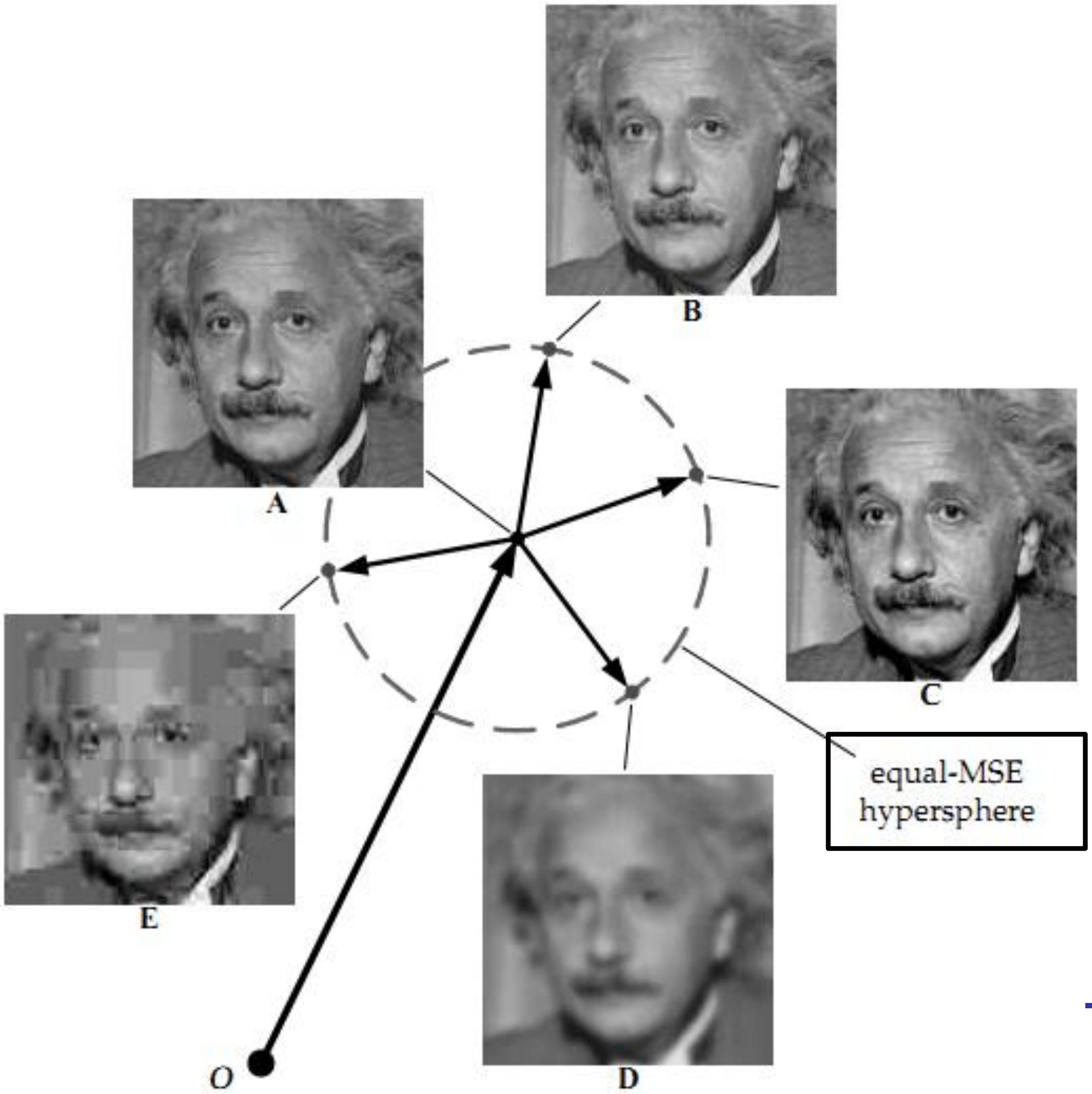
What's wrong with MSE?



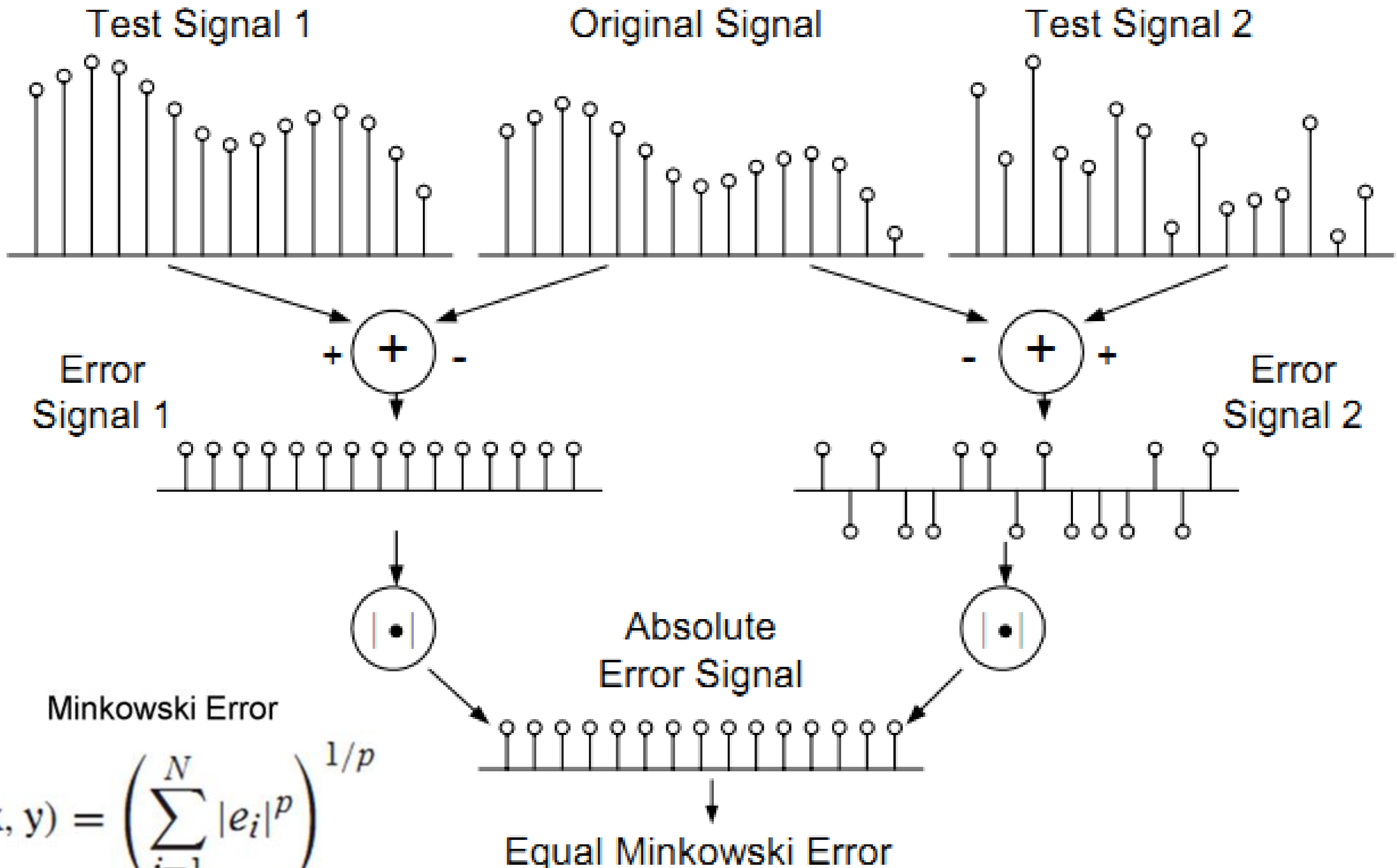
(a) MSE=0, SSIM=1 CW-SSIM=1
(b) MSE=306, SSIM=0.928 CW-SSIM=0.938
(c) MSE=309, SSIM=0.987 CW-SSIM=1.000
(d) MSE=309, SSIM=0.576 CW-SSIM=0.814
(e) MSE=313, SSIM=0.730 CW-SSIM=0.811
(f) MSE=309, SSIM=0.580 CW-SSIM=0.633
(g) MSE=308, SSIM=0.641 CW-SSIM=0.603
(h) MSE=694, SSIM=0.505 CW-SSIM=0.925
(i) MSE=871, SSIM=0.404 CW-SSIM=0.933
(j) MSE=873, SSIM=0.399 CW-SSIM=0.933
(k) MSE=590, SSIM=0.549 CW-SSIM=0.917
(l) MSE=577, SSIM=0.551 CW-SSIM=0.916

a) Reference
b) Contrast stretched
c) Luminance shift
d) Gaussian noise
e) Impulse noise
f) JPEG
g) Blurring
h) Zoom out
i) Shift to right
j) Shift to left
k) Rotation clockwise
l) Rotation anti-clockwise

What's wrong with MSE?



Trouble with Minkowski Error



Minkowski Error

$$d_p(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^{N} |e_i|^p \right)^{1/p}$$

Implicit Assumptions when using MSE

1) Signal fidelity is independent of temporal or spatial relationships between the samples of the original signal. In other words, if the original and distorted signals are randomly re-ordered in the same way, then the MSE between them will be unchanged.

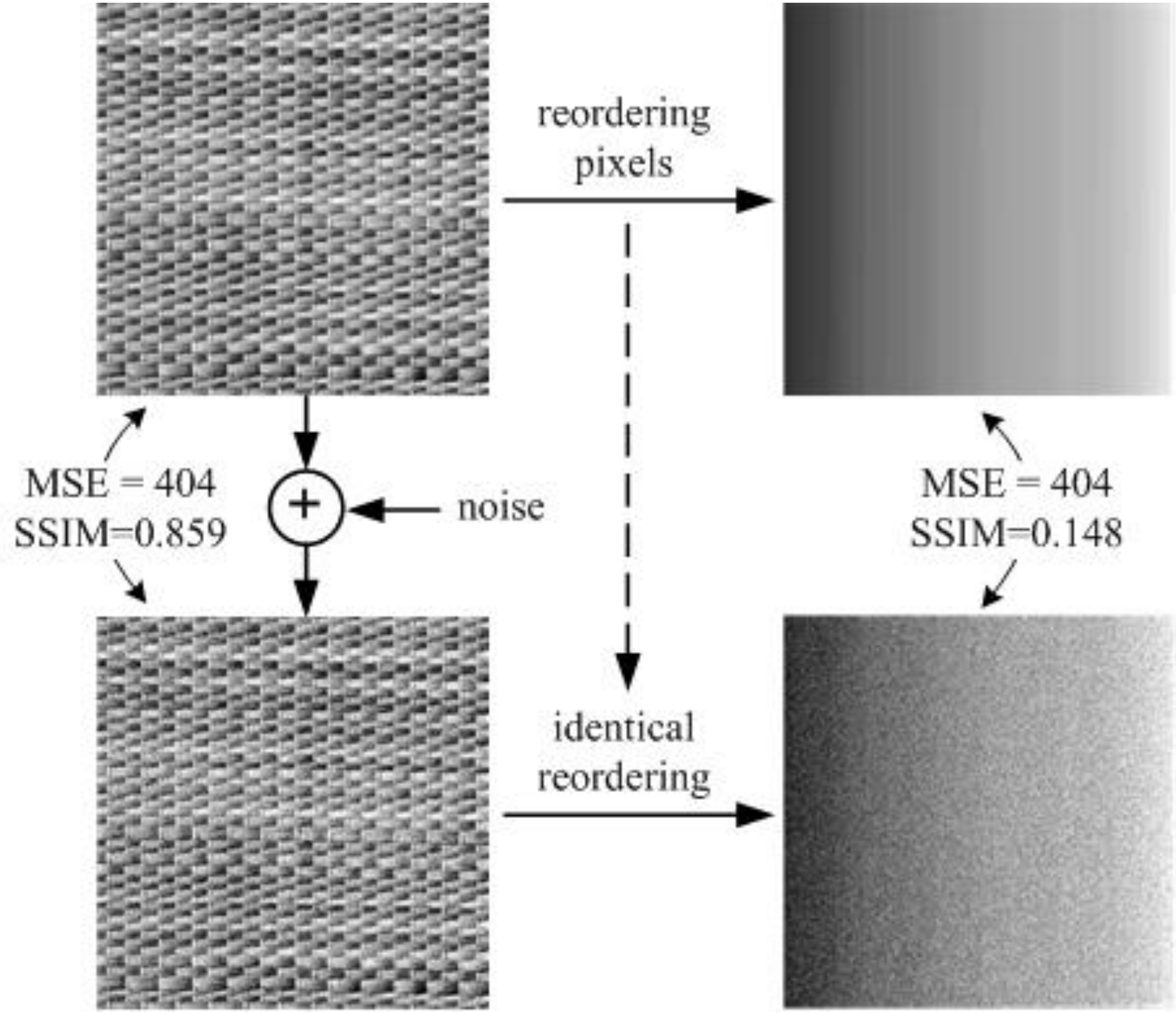
2) Signal fidelity is independent of any relationship between the original signal and the error signal. For a given error signal, the MSE remains unchanged, regardless of which original signal it is added to.

3) Signal fidelity is independent of the signs of the error signal samples.

4) All signal samples are equally important to signal fidelity.

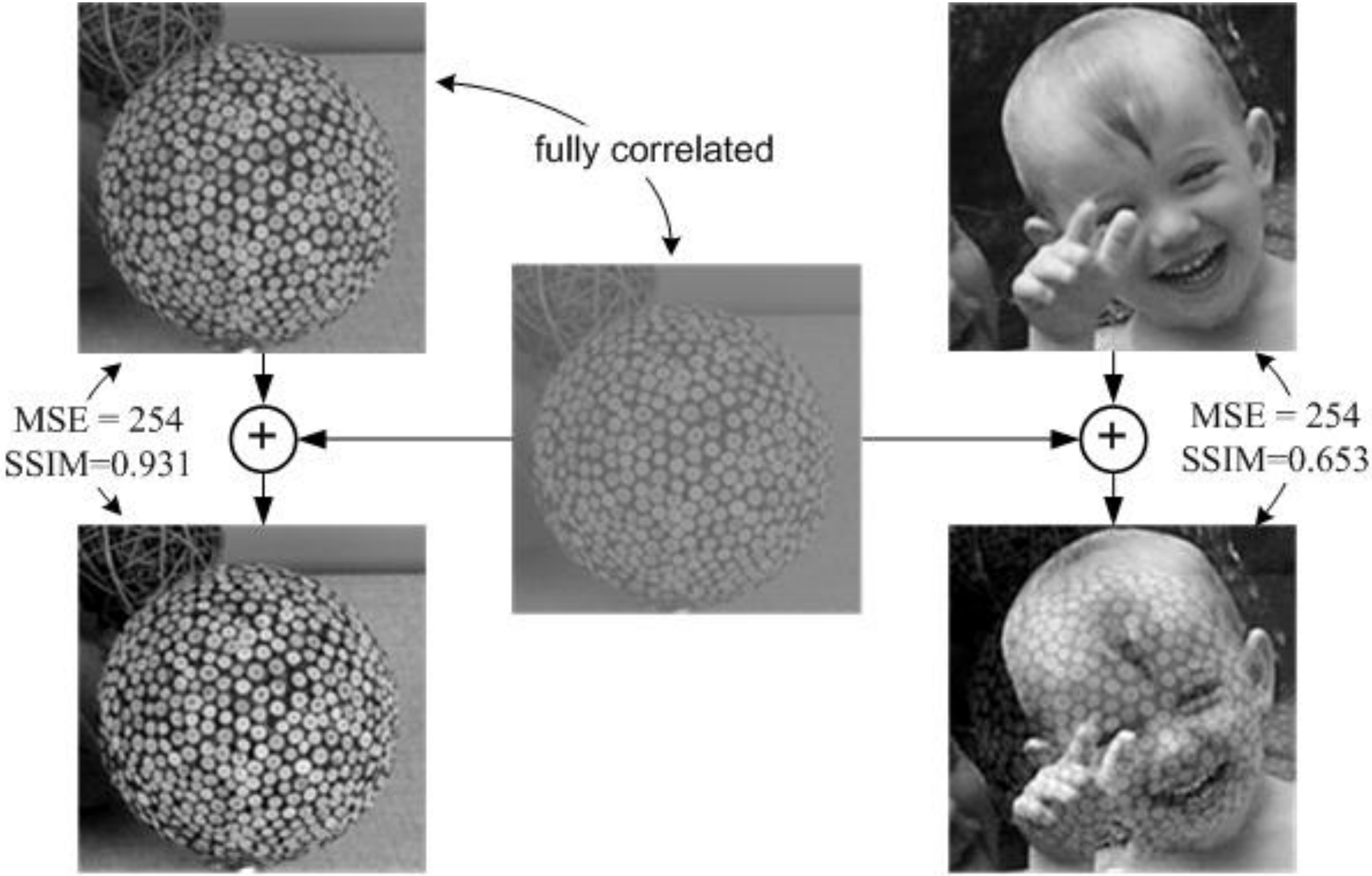
Disproving Assumption # 1

Signal fidelity independent of relationship between pixels



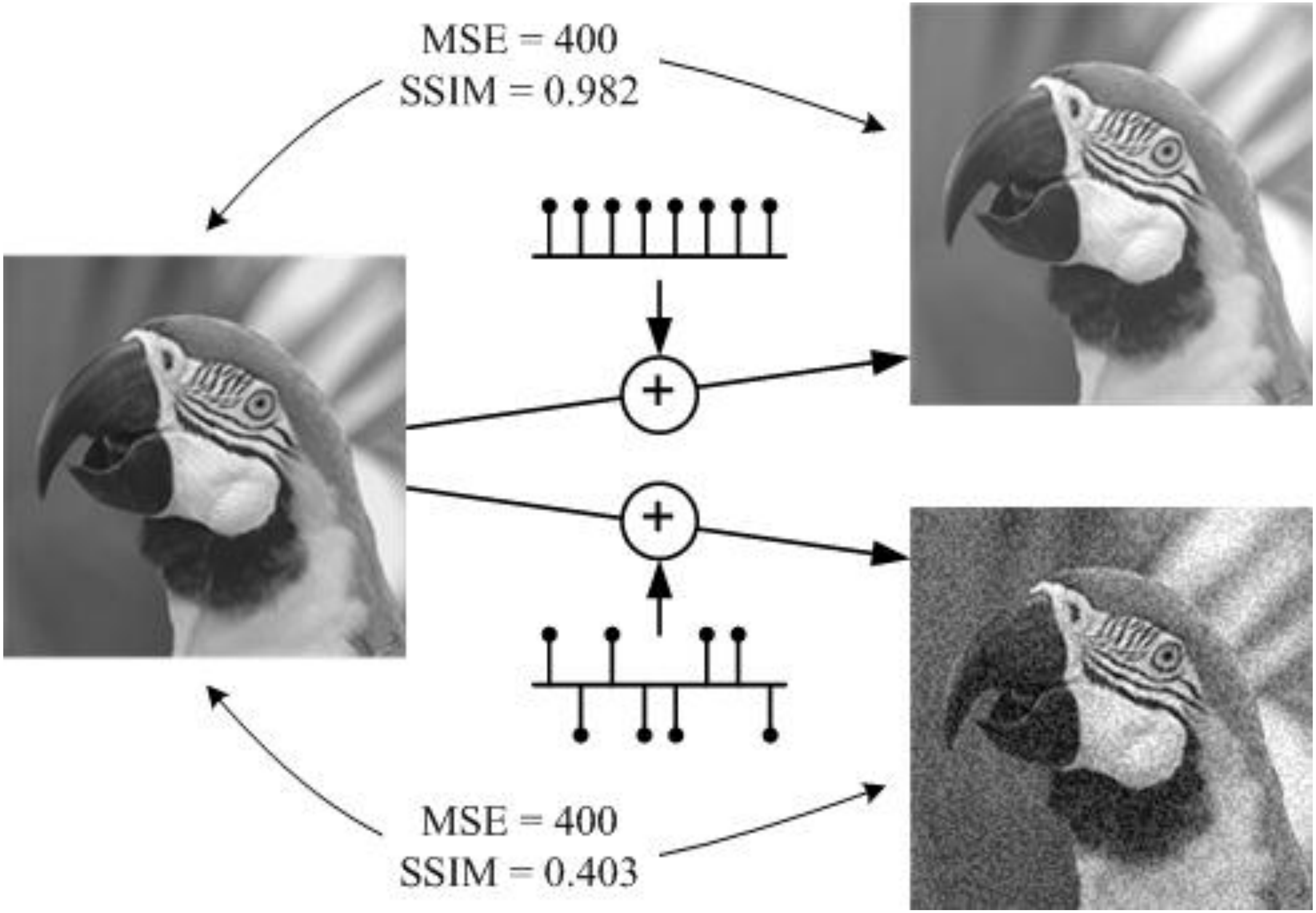
Disproving Assumption # 2

Signal fidelity independent of relationship between original and error signal



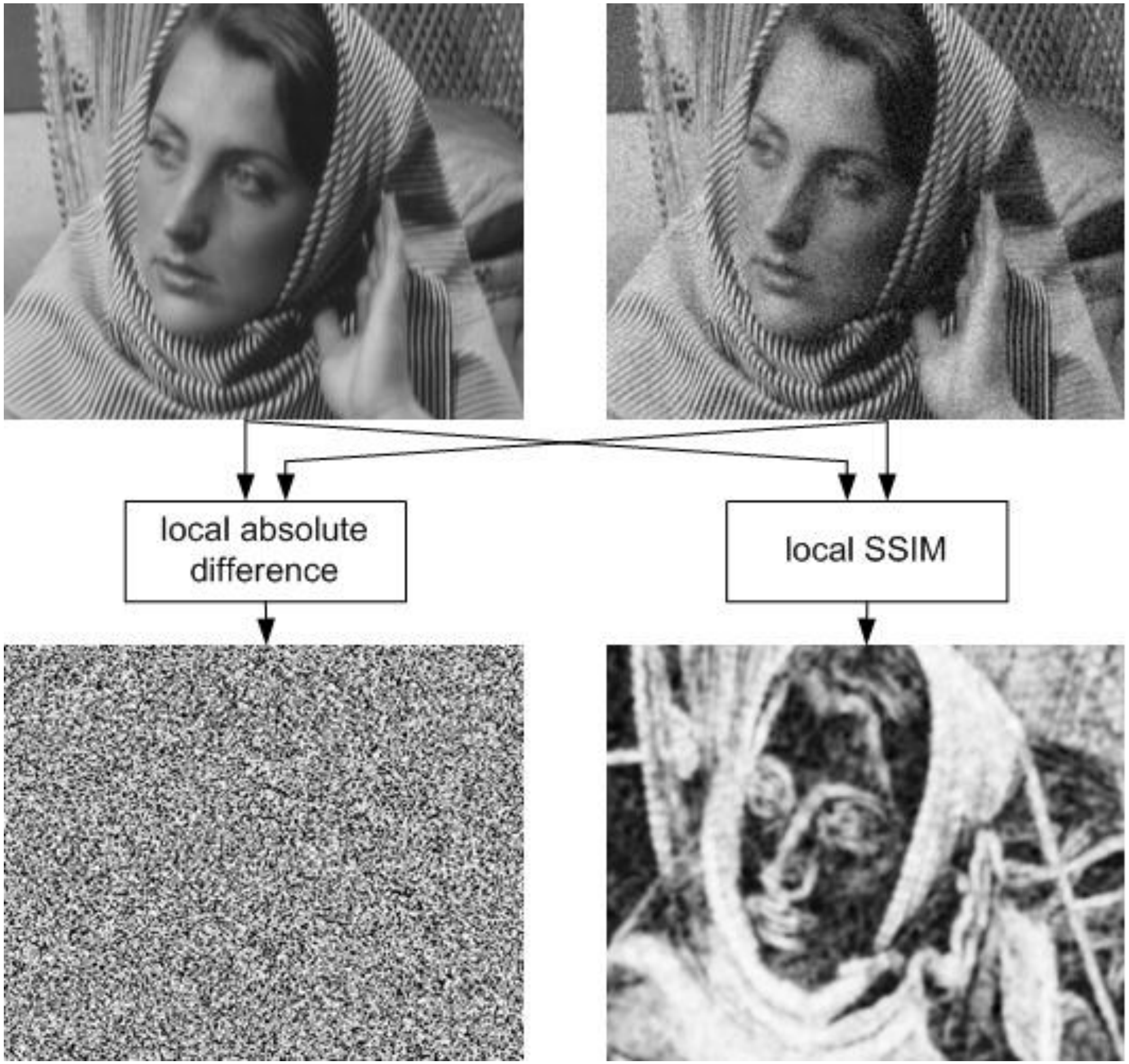
Disproving Assumption # 3

Signal fidelity independent of signs of error samples



Disproving Assumption # 4

All signal samples are equally important



New Paradigm: Structural Similarity

[1] Z. Wang, A. C. Bovik and L. Lu, “Why is image quality assessment so difficult?” ***Proc. IEEE Intl. Conf. Acoustics, Speech, and Signal Processing***, vol. 4, pp. 3313-3316, May 2002.

- In [1], a new philosophy has been proposed:

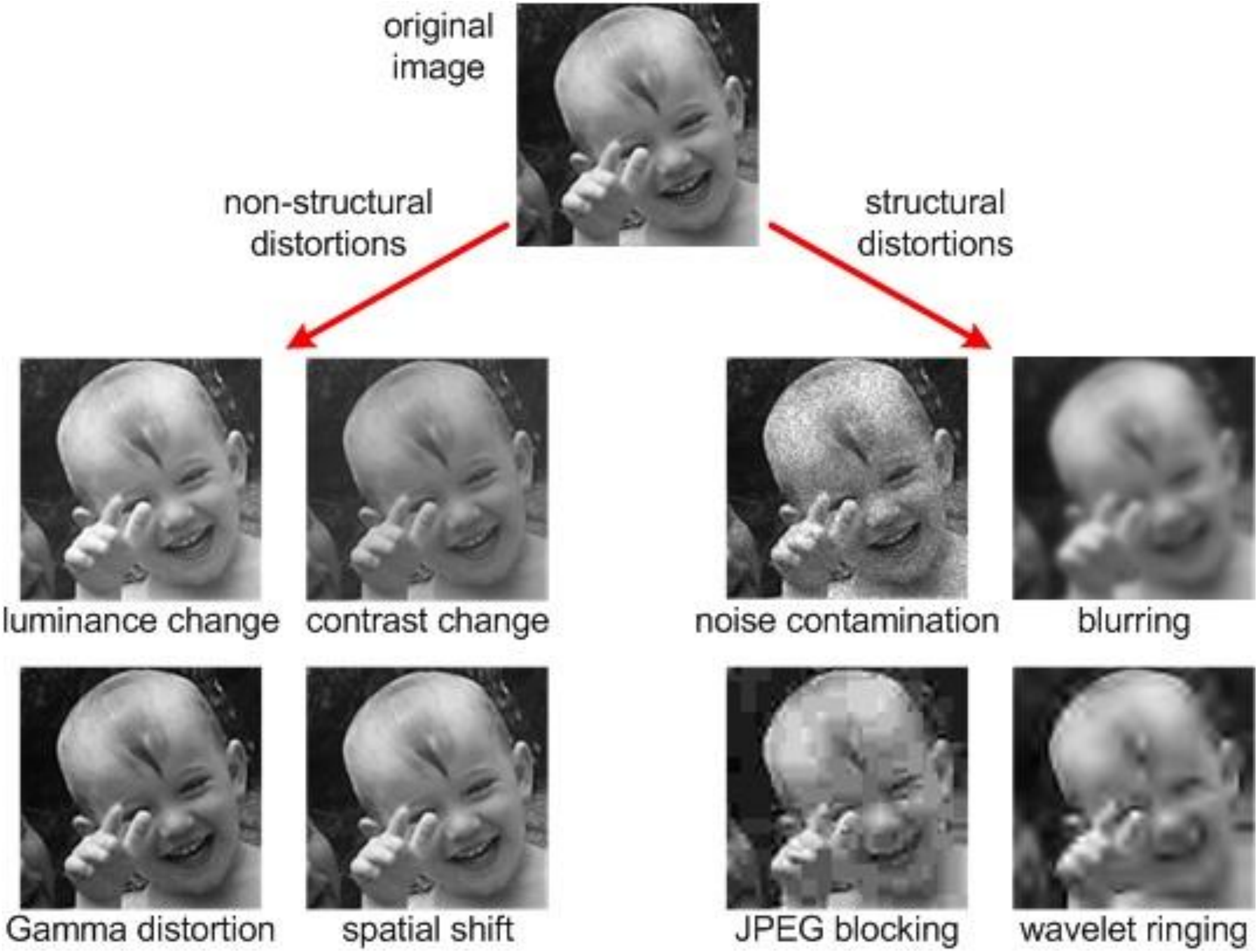
Philosophy

Purpose of human vision: extract **structural information**

HVS is highly adapted for this purpose

Estimate **structural information change**

Structural Distortions



Structural Distortions

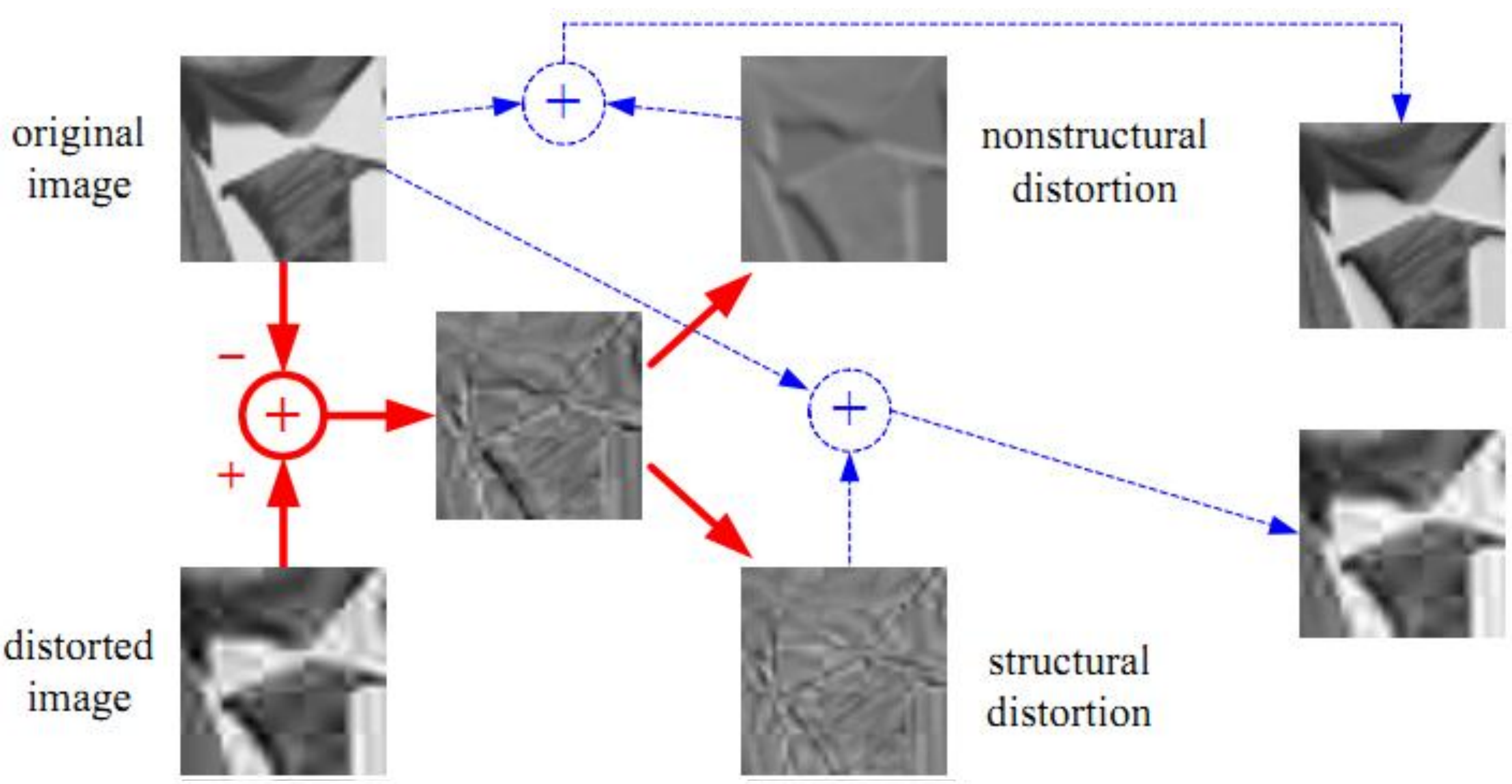


Image Quality Index (Q)

Z. Wang, A. Bovik, “A Universal Image Quality Index”, ***IEEE Signal Processing Letters*,** March 2002

Let $\boldsymbol{x}=\{x_i \mid i=1,2,\cdots N\}$ and $\boldsymbol{y}=\{x_i \mid i=1,2,\cdots N\}$ be the original and the test image signals, respectively. The quality index is defined as:

$$\boxed{Q=\frac{4\sigma_{xy}\,\bar{x}\,\bar{y}}{(\sigma_x^2+\sigma_y^2)[(\bar{x})^2+(\bar{y})^2]},}$$

loss of correlation, mean distortion, contrast distortion

where $\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i\,,\quad \bar{y}=\frac{1}{N}\sum_{i=1}^{N}y_i\,,$

$$\sigma_x^2=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2\,,\ \sigma_y^2=\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})^2\,,$$

$$\sigma_{xy}=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})\,.$$

The dynamic range of Q is [−1, 1].
The best value 1 is achieved if and only if $y_i=x_i$ for all $i=1,2,\cdots N$.

Image Quality Index (Q)

$$Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\bar{x}\bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$$

correlation coefficient between $\boldsymbol{x}$ and $\boldsymbol{y}$
range is $[-1, 1]$

how similar the mean values of $\boldsymbol{x}$ and $\boldsymbol{y}$ are
range of $[0, 1]$

how similar the contrasts of the images are
range of $[0, 1]$

loss of correlation,
mean distortion,
contrast distortion

Image Quality Index (Q)

MSE=200

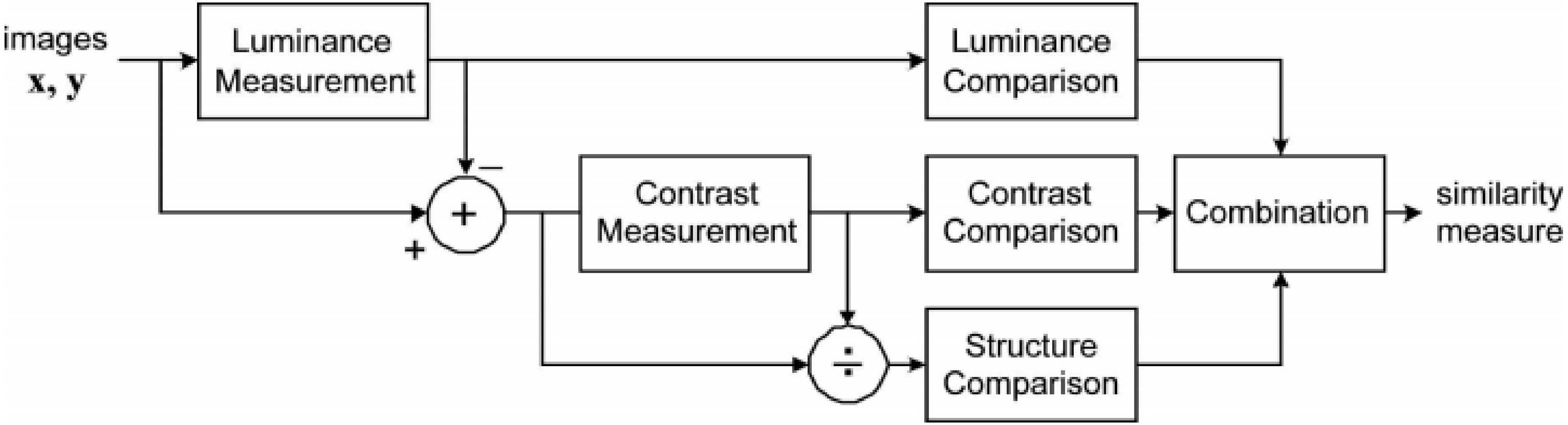
Q=0.3483

MSE=200

Q=0.6594

Structural Similarity Index (SSIM)

Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity," ***IEEE Transactions on Image Processing*** , vol. 13, no. 4, pp. 1-14, April **2004**



Luminance = *Reflectance* x *Illumination* of object
(independent of structure)

Effects of *Luminance* and *Contrast* are canceled out

Structure comparison on ***Luminance*** and ***Contrast*** **normalized** signals

Structural Similarity Index (SSIM)

$$\mathrm{SSIM}(\vec{x},\vec{y}) = f\left(l(\vec{x},\vec{y}), c(\vec{x},\vec{y}), s(\vec{x},\vec{y})\right)$$

$$l(\vec{x},\vec{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \qquad c(\vec{x},\vec{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \qquad s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

The SSIM index is applied locally rather than globally => because image features are highly non-stationary.
Using local windows provides a **quality map** of the image => valuable information about visual quality.

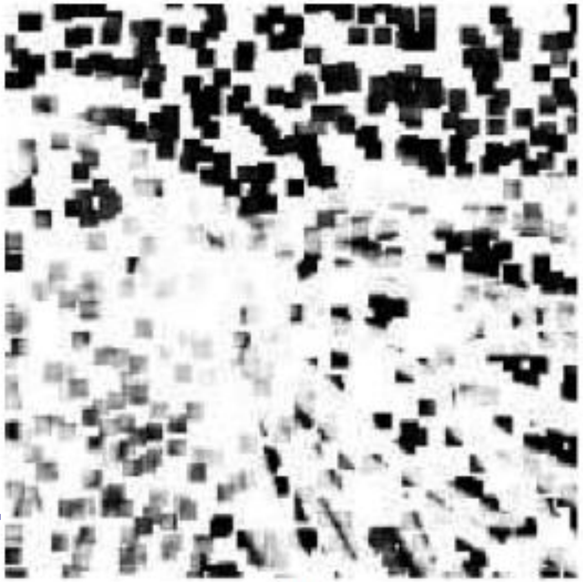
The above quantities are computed in a **local sliding window**, i.e. moved pixel-by-pixel over the entire image.
To avoid blocking artifacts, the resulting values are weighted using a circularly symmetric 11 x 11 Gaussian function.

$$\boxed{\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = \frac{(2\,\mu_x\,\mu_y + C_1)\,(2\,\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)\,(\sigma_x^2 + \sigma_y^2 + C_2)}} \qquad \boxed{\mathrm{MSSIM}(\vec{X},\vec{Y}) = \frac{1}{M}\sum_{i=1}^{N}\mathrm{SSIM}(\vec{x_i},\vec{y_i})}$$

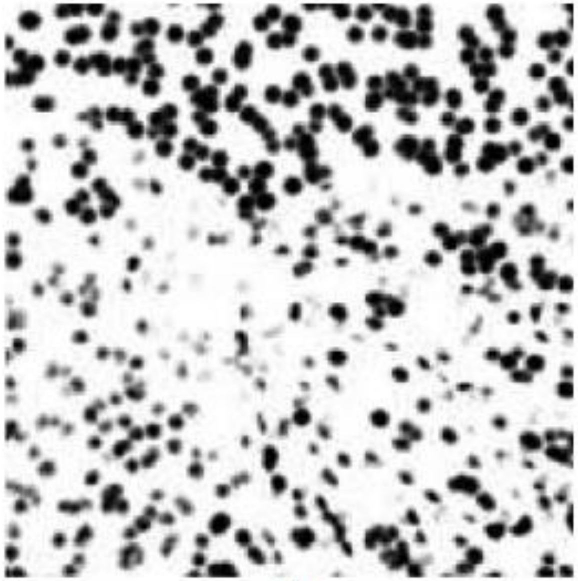
Structural Similarity Index (SSIM)

(a)

(b)



(c)



(d)

a) Original Image

b) Impulse Noise Contaminated Image

c) SSIM Index Map using Square Window

d) SSIM Index Map using Gaussian Window

Structural Similarity Index (SSIM)

(a)

(b)

(c)

(d)

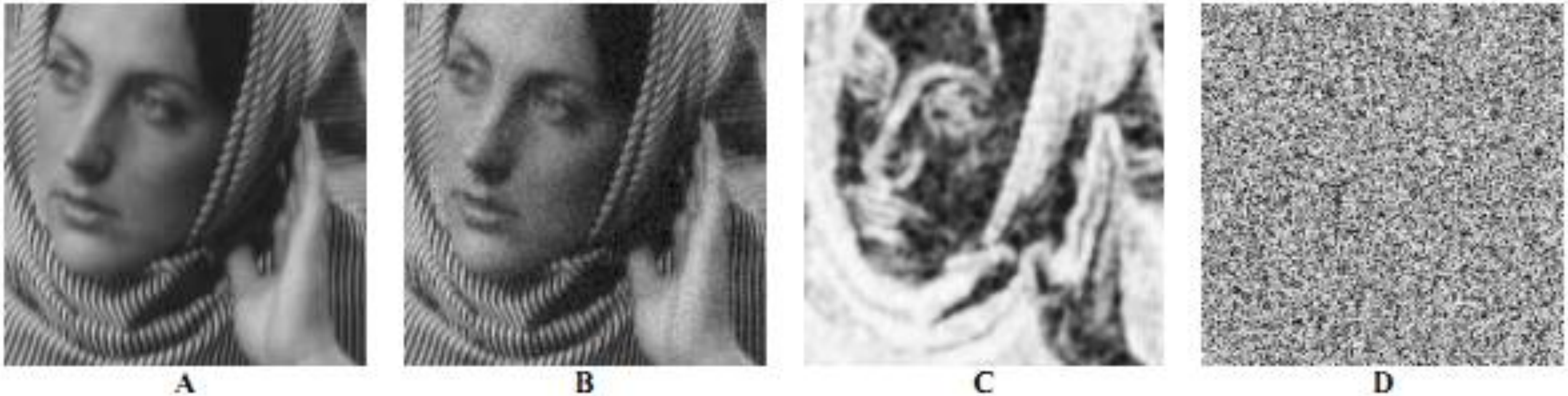
(a) Reference image

(b) JPEG compressed image

(c) Absolute error map of the distorted image

(d) SSIM index map of the distorted image

Structural Similarity Index (SSIM)



Gaussian Noise Contamination

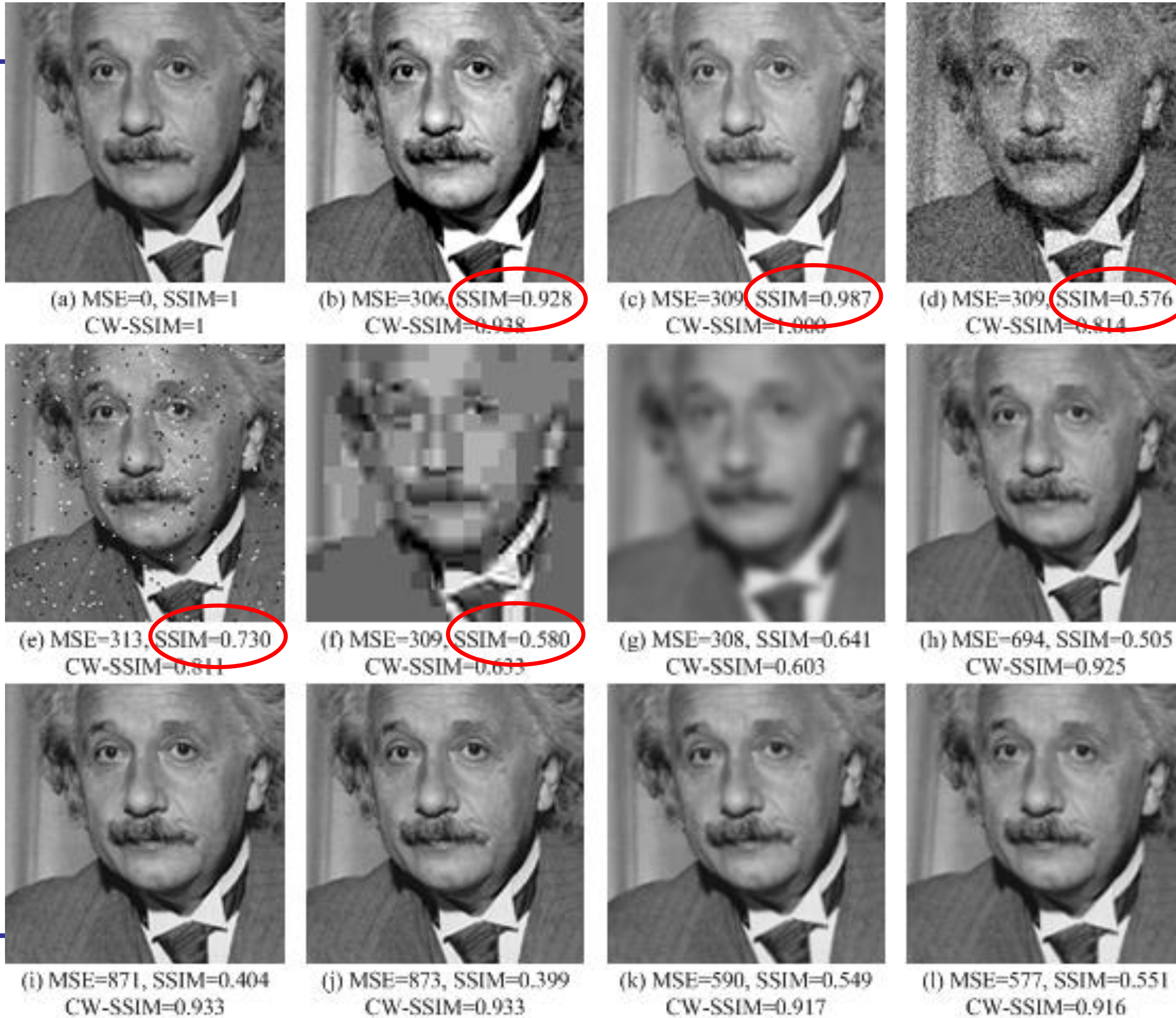
Original Image

SSIM Index Map

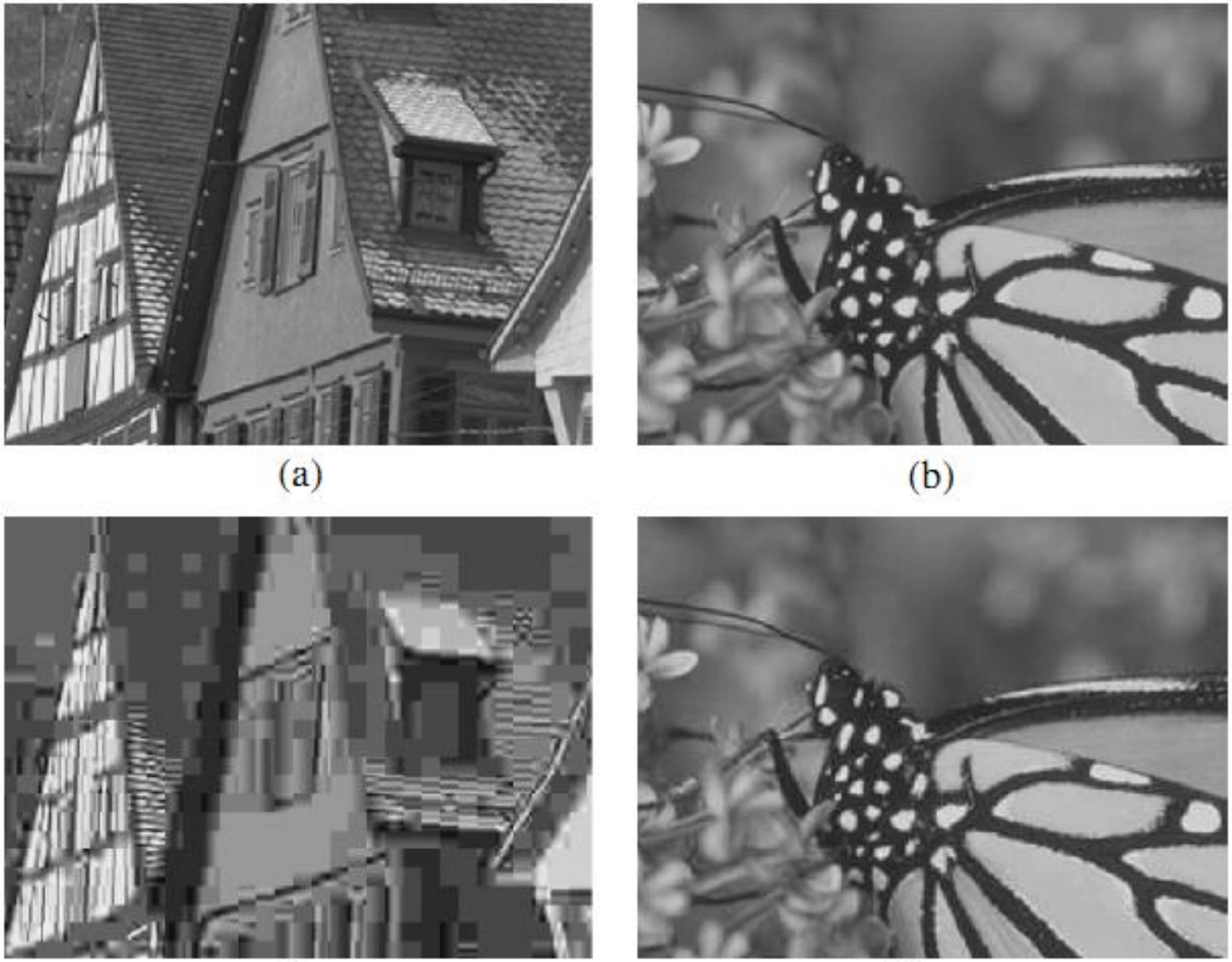
Absolute Error Map

JPEG2000 Compressed

SSIM is better than MSE



SSIM is better than MSE



(a) (b)

(c) (d)

MSSIM = 0.7118

MSSIM = 0.9898

SSIM – End of road?

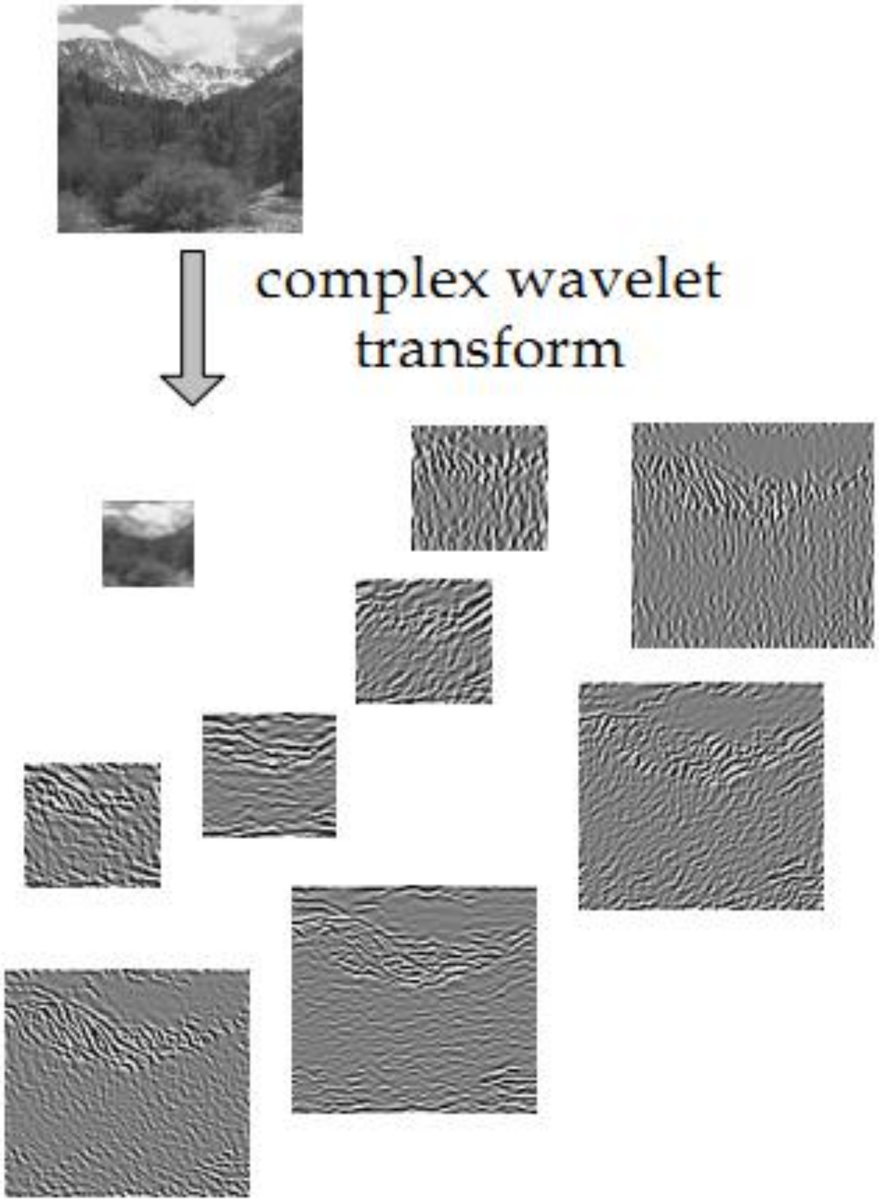
Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," ***IEEE Trans. Image Processing*** , vol. 13, no. 4, pp. 1-14, April 2004

We consider the proposed SSIM indexing approach as a **particular implementation** of the **philosophy of structural similarity**, from an image formation point of view.

Under the same philosophy, **other approaches** may emerge that could be significantly different from the proposed SSIM algorithm.

Creative investigation of the concepts of the structural information and structural distortions are likely to drive the success of these innovations.

Extensions of SSIM



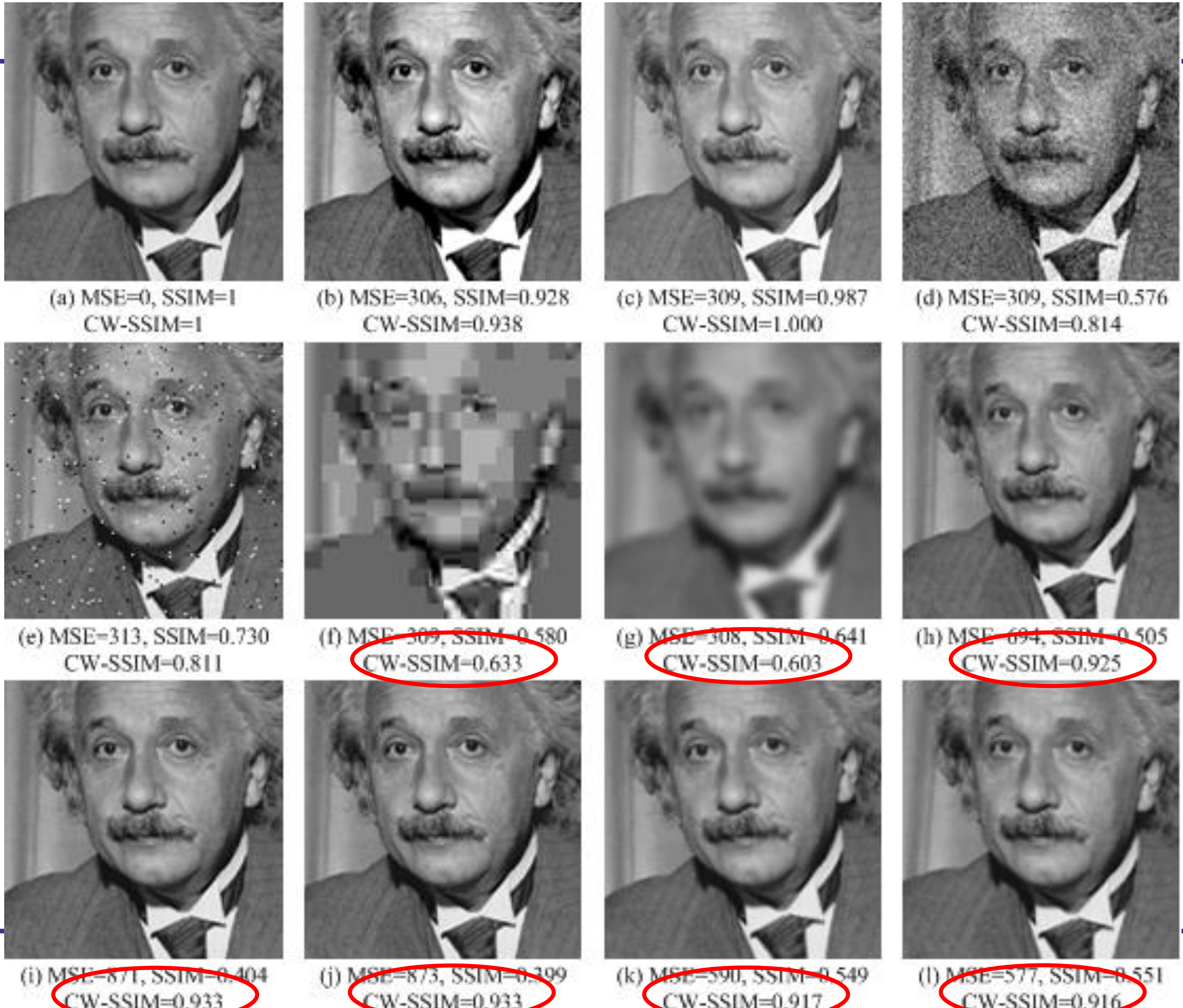
- **Complex wavelet SSIM**
  - Motivation: robust to translation, rotation and scaling

$$\mathrm{SSIM}(\boldsymbol{x},\boldsymbol{y}) = \frac{2\left|\sum c_x \cdot c_y^*\right| + C}{\sum |c_x|^2 + \sum |c_y|^2 + C}$$

c_x, c_y : complex wavelet coefficients in images x and y

[Wang & Simoncelli, *ICASSP* '05]

CW-SSIM is even better than SSIM



(a) MSE=0, SSIM=1
CW-SSIM=1

(b) MSE=306, SSIM=0.928
CW-SSIM=0.938

(c) MSE=309, SSIM=0.987
CW-SSIM=1.000

(d) MSE=309, SSIM=0.576
CW-SSIM=0.814

(e) MSE=313, SSIM=0.730
CW-SSIM=0.811

(f) MSE=309, SSIM=0.580
CW-SSIM=0.633

(g) MSE=308, SSIM=0.641
CW-SSIM=0.603

(h) MSE=694, SSIM=0.505
CW-SSIM=0.925

(i) MSE=871, SSIM=0.404
CW-SSIM=0.933

(j) MSE=873, SSIM=0.399
CW-SSIM=0.933

(k) MSE=590, SSIM=0.549
CW-SSIM=0.917

(l) MSE=577, SSIM=0.551
CW-SSIM=0.916

Application: Image Matching

Standard patterns: 10 images

Database: 2430 images

Correct Recognition Rate:

MSE: 59.6%; SSIM: 46.9%; Complex wavelet SSIM: **97.7%**

[Wang & Simoncelli, *ICASSP* '05]

Research Community Loves SSIM

• **Watermarking/data hiding**	[Alattar '03, Noore *et al.* '04, Macq *et al.* '04 Zhang & Wang '05, Kumsawat *et al.* '04]
• **Image denoising**	[Park & Lee '04, Yang & Fox '04 , Huang *et al.* '05 Roth & Black '05, Hirakawa & Parks '05]
• **Image enhancement**	[Battiato *et al.* '03]
• **Image/video hashing**	[Coskun & Sankur '04, Hsu & Lu '04]
• **Image rendering**	[Bornik *et al.* '03]
• **Image fusion**	[Zheng *et al.* '04, Tsai '04, Gonzalez-Audicana *et al.* '05]
• **Texture reconstruction**	[Toth '04]
• **Image halftoning**	[Evans & Monga '03, Neelamani '03]
• **Radar imaging**	[Bentabet '03]
• **Infrared imaging**	[Torres '03, Pezoa *et al.* '04]
• **Ultrasound imaging**	[Loizou *et al.* '04]
• **Vision processor design**	[Cembrano *et al.*, '04]
• **Wearable display design**	[von Waldkirch *et al.* '04]
• **Contrast equalization for LCD**	[Iranli *et al.* '05]
• **Airborne hyperspectral imaging**	[Christophe *et al.* '05]
• **Superresolution for remote sensing**	[Rubert *et al.* '05]

Modern State-of-the-Art Approaches

- Multi-Scale SSIM (MS-SSIM), 2003
- Complex-Wavelet SSIM (CW-SSIM), 2005
- Visual Information Fidelity criterion (VIF), 2006
- Visual Signal to Noise Ratio (VSNR), 2007
- Blind Image Quality Index (BIQI), 2010
- Information Content Weighted SSIM (IW-SSIM), 2010
- Motion-based Video Integrity Evaluation (MOVIE) index for video quality assessment, 2010
- Feature Similarity Metric (FSIM), 2011

… lots of opportunity for **<u>research</u>** and **<u>innovation</u>**

Challenges Ahead…

Quantifying Visual Aesthetics

(a)

Presence of Border

(b)

Computer-generated effects

Challenges Ahead…

Quantifying Visual Aesthetics

(a)

(b)

Images with 'bad' aesthetic rating : (a)it is unclear if the poor aesthetic rating is influenced by the content and (b) it is unclear if the poor rating corresponds to poor quality or poor aesthetics

Challenges Ahead…

Quantifying Visual Aesthetics

(a)

(b)

(a) An image with poor aesthetics and poor quality, but possibly good content (for the dog owner for example). (b) An image with high aesthetic, quality and (possibly) content rating.

Challenges Ahead…

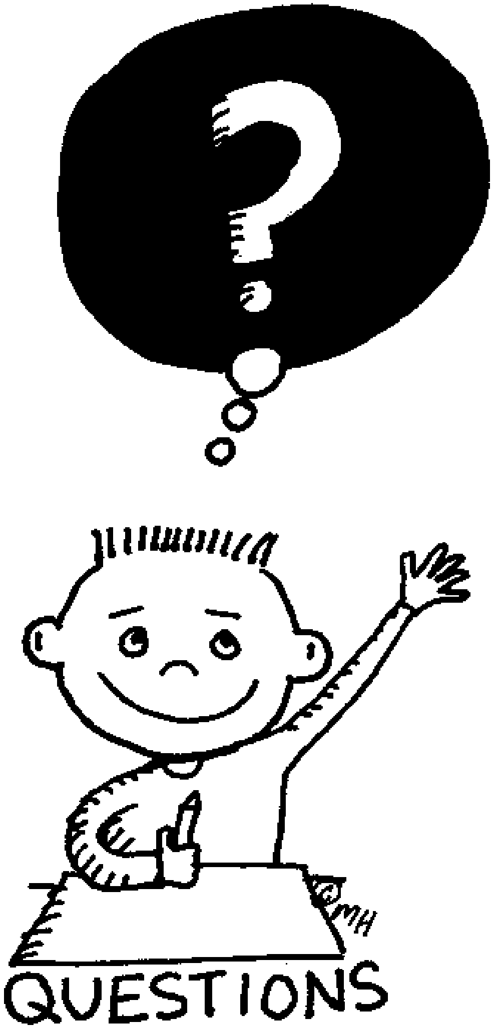
Quantifying Visual Aesthetics

(a)

(b)

(a) An image with high aesthetic appeal and good quality. (b) The same image with high appeal but with poor quality.

Quantifying Visual Aesthetics is a ***challenging*** problem!


QUESTIONS