New Directions in Image and Video Quality Assessment

Al Bovik

Laboratory for Image & Video Engineering (LIVE)

The University of Texas at Austin

bovik@ece.utexas.edu

October 2, 2007

Prologue

I seek analogies between assessment of visual signal quality

and

measuring the fidelity of a communication system

in hopes

that the powerful tools of that mathematical discipline may be brought to bear

A Classic Communication System

Transmitter:
- Source
- Encoder
- Encrypter
- Modulator

Channel:
- Noise
- Interference
- Distortion
- Fading

Receiver:
- Decoder
- Decrypter
- Demodulator
- Interface

Basic Tenet of Communication Theory

- The more known about

  the **transmitter**

  the **channel**

  the **reciever**

  the better job of **communication** can be done

- Provided the models of transmitter, channel and receiver **are accurate**.

Image Quality Assessment

What are the transmitter, channel, and receiver….?

The Natural Image Transmitter

Photos of the natural image transmitter

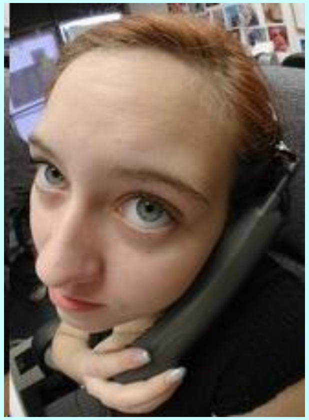
LABORATORY FOR IMAGE AND VIDEO ENGINEERING
LIVE
THE UNIVERSITY OF TEXAS AT AUSTIN

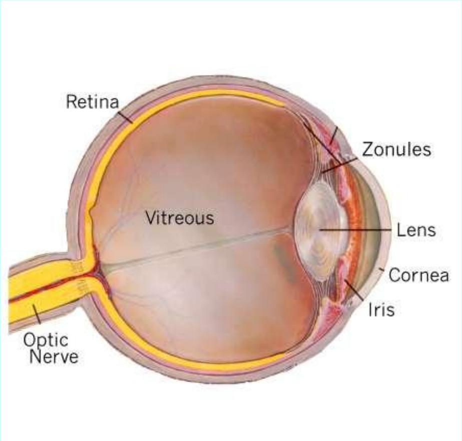
UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING

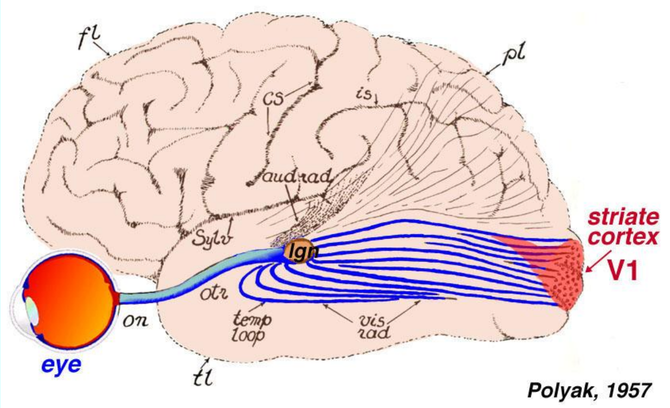
Natural Image Transmitter

- Also called "the real world."
- Produces ***Natural Image Signals*** – light fields emitted/reflected from objects.
- ***Natural Scene Statistics*** (NSS) are still being learned. These models are in a **nascent stage**.

The Natural Image Receiver





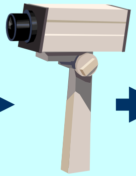


Depictions of the natural image receiver

The Natural Image Receiver

- Also called the ***Human Visual System*** (HVS).
- **Sophisticated models for**: optics, retinal neurons, post-retinal neurons, cortical receptive fields, gain control, masking, threshold visibility etc.
- Yet we have only begun to penetrate the exquisite sophistication of the HVS. **Our receiver model is very incomplete.**

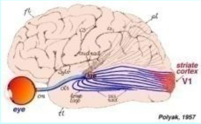
Overall Communication System



All front-end digital processing

classical channel

All back-end digital processing



Natural image signal

Sensing & digitizing

Mapping & display

Perceptual image signal

The Image Channel

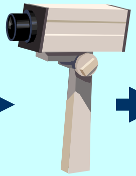
The Natural-Synthetic Image Transmitter

The Natural-Synthetic Image Receiver

LABORATORY FOR IMAGE AND VIDEO ENGINEERING
LIVE
THE UNIVERSITY OF TEXAS AT AUSTIN

UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING

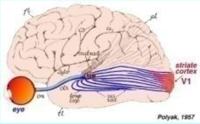
Sources of Image Distortion



All front-end digital processing

classical channel

All back-end digital processing



Natural image signal

Sensing & digitizing

Mapping & display

Perceptual image signal

The Image Channel

The Natural-Synthetic Image Transmitter

The Natural-Synthetic Image Receiver

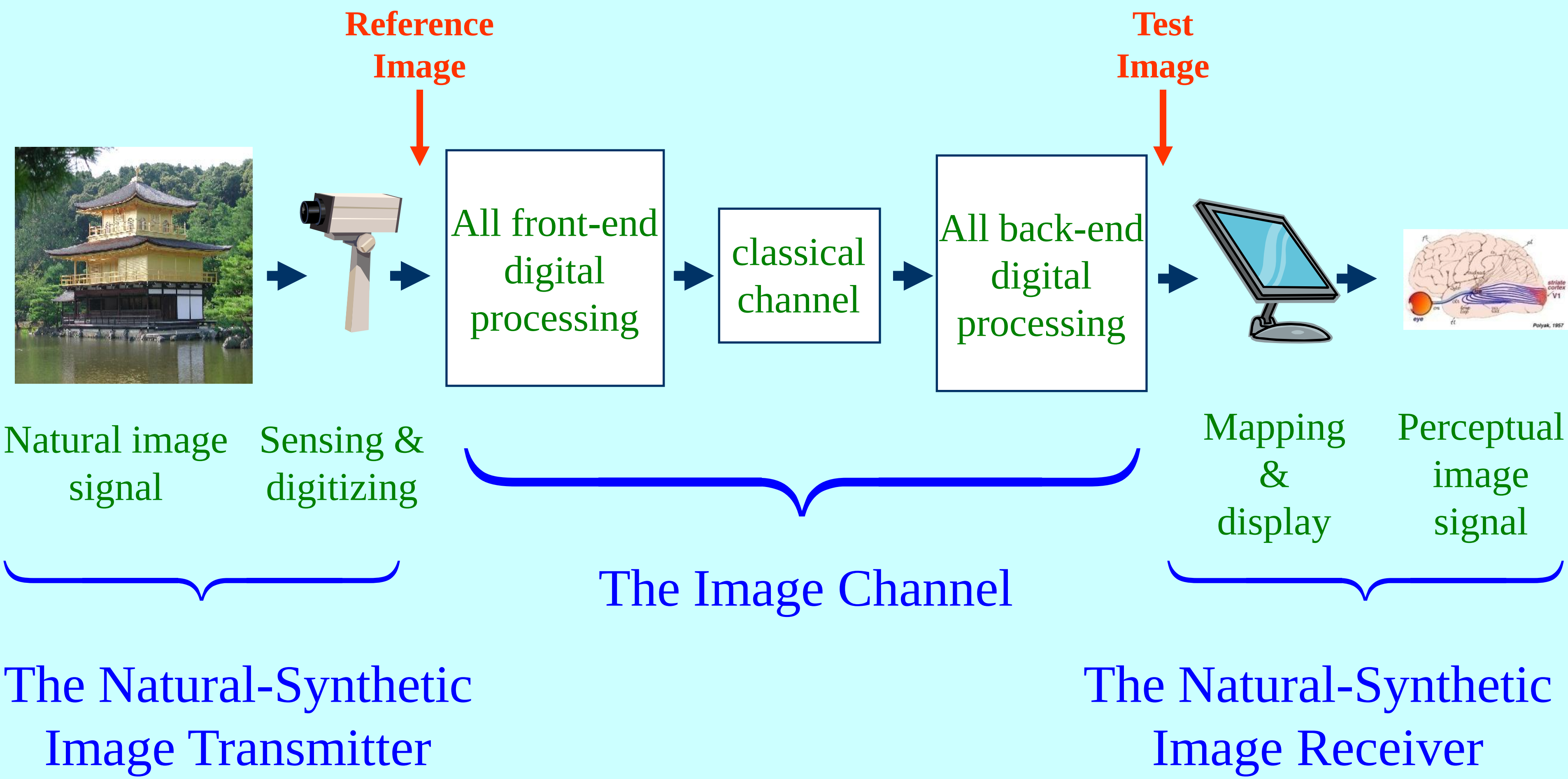
LABORATORY FOR IMAGE AND VIDEO ENGINEERING
LIVE
THE UNIVERSITY OF TEXAS AT AUSTIN

UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING

Four Classes of QA Algorithm

- "Full-Reference" QA
- "No-Reference" or Blind QA
- "Reduced-Reference" QA
- "Distortion-Specific" QA

“Full-Reference” QA



“Full-Reference” QA

- Depends on accurate statistical models of the transmitter: **How far does the test image depart from “normal behavior”?**

- Depends on accurate statistical models of the receiver: **How far does the test image depart from “normal appearance”?**

- Must be **baselined against human subjectivity** – large, statistically significant human studies.

LABORATORY FOR IMAGE AND VIDEO ENGINEERING
LIVE
THE UNIVERSITY OF TEXAS AT AUSTIN

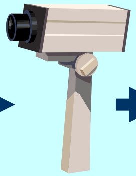
UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING

"No-Reference" or Blind QA

Reference Image??

Test Image

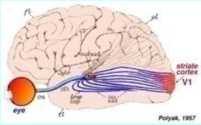
Reference Image??



All front-end digital processing

classical channel

All back-end digital processing



Natural image signal

Sensing & digitizing

Mapping & display

Perceptual image signal

The Image Channel

The Natural-Synthetic Image Transmitter

The Natural-Synthetic Image Receiver

LABORATORY FOR IMAGE AND VIDEO ENGINEERING

LIVE

THE UNIVERSITY OF TEXAS AT AUSTIN

UNIVERSITY OF TEXAS AT AUSTIN

UT ECE

ELECTRICAL & COMPUTER ENGINEERING


Perhaps it depends on whether the viewer is an ornithologist…
…. or a botanist….
Is this a "good quality" image?

Blind QA

- No reference image available. **Little progress** made on this problem. So, not covered here. ☹ …. a "Holy Grail" of image processing

- Will require profound insights into **natural image modeling** and **image appearance modeling.**

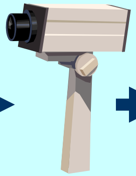
- I **do not** include "Blind QA with known distortion type"

"Reduced-Reference" QA

Reference Data

Test Image

Natural image signal



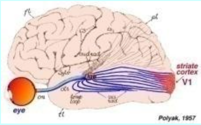
Sensing & digitizing

All front-end digital processing

classical channel

All back-end digital processing

Mapping & display



Perceptual image signal

The Image Channel

The Natural-Synthetic Image Transmitter

The Natural-Synthetic Image Receiver

"Reduced-Reference" QA

- No reference image, but **side information** sent with the transmitted image.

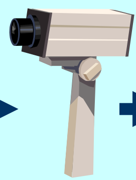
- **Side info:** Partial wavelet data; edge locations; local statistics, etc. Promising but requires application- & domain-dependent assumptions.

- Not discussed here.

"Distortion-Specific" QA

Distortion Model

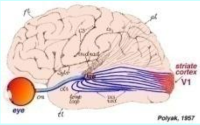
Test Image



All front-end digital processing

classical channel

All back-end digital processing



Natural image signal

Sensing & digitizing

Mapping & display

Perceptual image signal

The Image Channel

The Natural-Synthetic Image Transmitter

The Natural-Synthetic Image Receiver

LABORATORY FOR IMAGE AND VIDEO ENGINEERING
LIVE
THE UNIVERSITY OF TEXAS AT AUSTIN

UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING

“Distortion-Specific” QA

- Blind, reduced reference, **or** full reference.
- **Channel distortion(s) known,** e.g. JPEG blocking. Effective for specific applications.
- Not generic; not covered here.

Predictable distortion artifacts?

Summary to This Point

- I've attempted to begin casting image & video QA as a problem in **classical communications theory.**
- Much work to be done in **accurately modeling transmitter, receiver and channel**.
- As these **models improve**, I believe we will rely on principles of mathematical communication theory to **significantly improve** modern QA algorithms.

IQA Algorithms

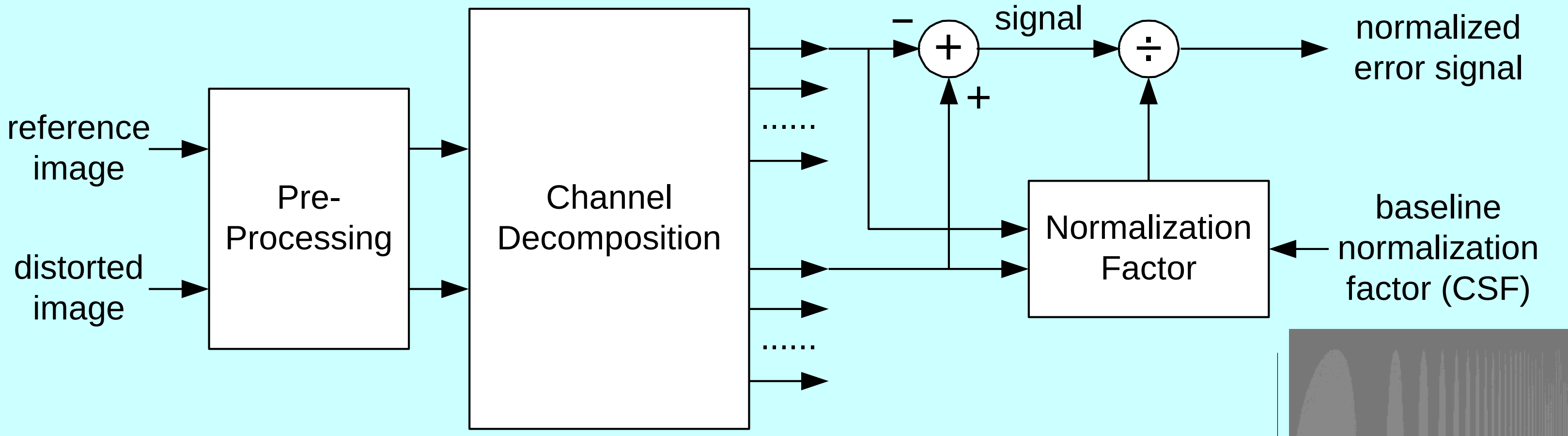
And now some discussion of existing IQA algorithms, old and new …

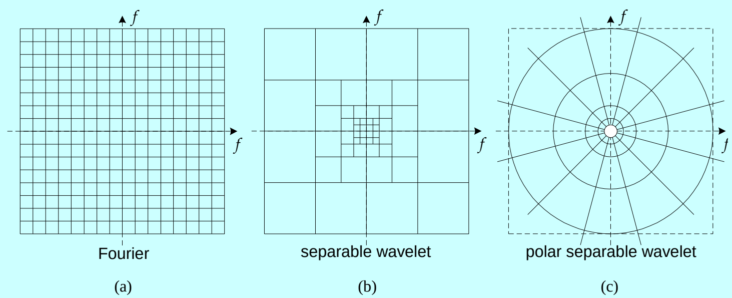
Receiver-Oriented Algorithms (FR)

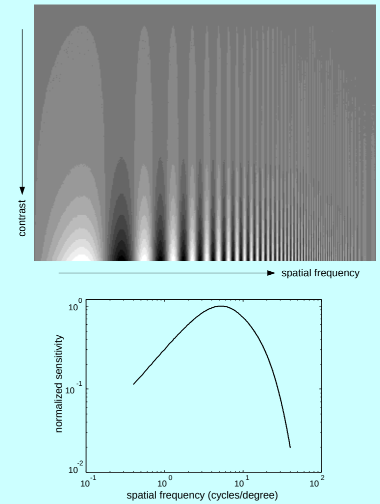
- Older approaches based on **models of human visual function** and on **measured perception**.
- Largely limited to front-end models of the visual pathway. **Idea: emulate the visual pathway.**
- Lubin, Daley, Watson, JNDMetrix, PQS, all generally complex/expensive.
- **No longer generally competitive for still images.**

Receiver-Oriented Algorithms

$$E = \left(\sum_m \sum_n \left| e(m,n) \right|^\beta \right)^{1/\beta}$$







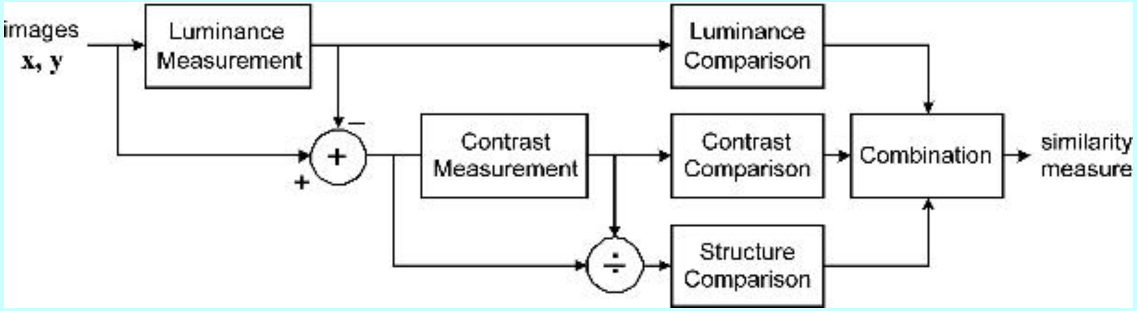
LABORATORY FOR IMAGE AND VIDEO ENGINEERING
LIVE
THE UNIVERSITY OF TEXAS AT AUSTIN

(Eventually) will again be the basis for the best QA algorithms (IMHO)

UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING

Structural Similarity (SSIM) Index

- Popular algorithm that uses weighted local (patch) image statistics:



$$\mathrm{SSIM}_{\mathbf{I},\mathbf{J}}(i, j) = L_{\mathbf{I},\mathbf{J}}(i, j) \cdot C_{\mathbf{I},\mathbf{J}}(i, j) \cdot S_{\mathbf{I},\mathbf{J}}(i, j)$$

local **luminance similarity**

local **structural similarity**

local **contrast similarity**

Wang & Bovik, *IEEE Signal Processing Letters*, March 02
Wang, Bovik, Sheikh & Simoncelli, *Trans on IP,* March 04

Structural Similarity (SSIM) Index

- Pointwise SSIM Index or SSIM Map:

$$\mathrm{SSIM}_{\mathbf{I,J}} = \left(\frac{2\mu_{\mathbf{I}}\mu_{\mathbf{J}} + C_1}{\mu_{\mathbf{I}}^2 + \mu_{\mathbf{J}}^2 + C_1}\right)\cdot\left(\frac{2\sigma_{\mathbf{I}}\sigma_{\mathbf{J}} + C_2}{\sigma_{\mathbf{I}}^2 + \sigma_{\mathbf{J}}^2 + C_2}\right)\cdot\left(\frac{2\sigma_{\mathbf{IJ}} + C_3}{\sigma_{\mathbf{I}}\sigma_{\mathbf{J}} + C_3}\right)$$

local **luminance similarity** local **contrast similarity** local **structural similarity**

- Mean SSIM Index

$$\mathrm{SSIM}(\mathbf{I,J}) = \left(\frac{1}{NM}\right)\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\mathrm{SSIM}_{\mathbf{I,J}}(i, j)$$

- Multiscale SSIM (MS-SSIM) operates over a dyadic pyramid (best SSIM performance)

Wang, Simoncelli & Bovik, *Asilomar, Nov 2003*

LABORATORY FOR IMAGE AND VIDEO ENGINEERING — LIVE — THE UNIVERSITY OF TEXAS AT AUSTIN

UNIVERSITY OF TEXAS AT AUSTIN — UT ECE — ELECTRICAL & COMPUTER ENGINEERING

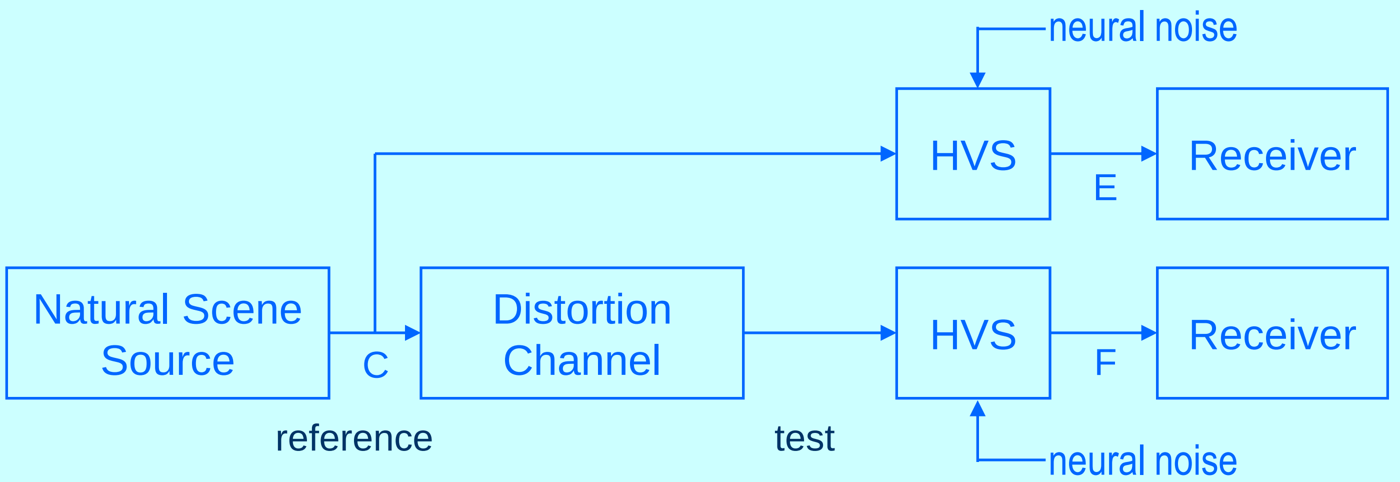
Comments on SSIM

- A **distortion-blind full-reference method**.
- Heuristic/Intuitive! **Not** derived from any **specific** image formation or perceptual models.
- ***Transmitter-Near-Optimal?*** Images are well described by local luminances (smooth patches), variances (textures), and structures (edges and details).
- ***Receiver-Near-Optimal?*** Perceptual quality depends on faithful rendering of local luminance, variation, and structure. And that quality perception of these is *separable*.

LABORATORY FOR IMAGE AND VIDEO ENGINEERING
LIVE
THE UNIVERSITY OF TEXAS AT AUSTIN

UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING

Visual Information Fidelity Index



$$\text{VIF} = \frac{I(C;F \mid z)}{I(C;E \mid z)}$$

- $I(C;F|z)$ is mutual information in the wavelet domain conditioned on scalar variance field z (estimated)

Sheikh, Bovik & DeVeciana, *Trans on IP*, Dec 05
Sheikh & Bovik, *Trans on IP*, Feb 06

VIF Index

- Distortion-Blind Full Reference IQA algorithm.
- Numerator measures the information that the HVS can extract ***from the distorted image***.
- Divisive normalization - by the information that the HVS can extract ***from the reference***.
- Simple models used:

  statistical transmitter model (GSM)

  channel model (blur + noise)

  simple wavelet + neural noise **receiver model**.

GSM Model – Statistical Transmitter Model

- **Image wavelet coefficients** modeled as Gaussian-scale mixture:

$$\boldsymbol{X} \sim z\boldsymbol{U}$$

  where z = space-varying variance field, and $\boldsymbol{U}$ are standard normal.

- Independent Gaussian when conditioned on variance.

- Simple, effective **transmitter model.**

Relative Performance

- The **LIVE Image Quality Assessment Database** – over 25,000 subjective judgements - Mean Opinion Scores (MOS). Recent Release 2 includes Differential MOS (DMOS) values as well.

- **Widely used and cited** - over 200 institutions have downloaded the (>1GB) LIVE database.

Spearman Rank-Order Coefficient

	JP2K#1	JP2K#2	JPEG#1	JPEG#2	WN	GBlur	FF	All data
PSNR	0.9263	0.8549	0.8779	0.7708	0.9854 1	0.7823	0.8907	0.8755
JND	0.9646 2	0.9608	0.9599	0.9150	0.9487	0.9389	0.9045	0.9291
DCTune	0.8335	0.7209	0.8702	0.8200	0.9324	0.6721	0.7675	0.8032
PQS	0.9372	0.9147	0.9387	0.8987	0.9535	0.9291	0.9388	0.9304
NQM	0.9465	0.9393	0.9360	0.8988	0.9854 1	0.8467	0.8171	0.9049
Fuzzy S7	0.9316	0.9000	0.9077	0.8012	0.9199	0.6056	0.9074	0.8291
BSDM (S4)	0.9130	0.9378	0.9128	0.9231	0.9327	0.9600	0.9372	0.9271
SSIM(MS)	0.9645 2	0.9648 2	0.9702 1	0.9454 1	0.9805	0.9519	0.9395	0.9527 2
IFC	0.9386	0.9534	0.9107	0.9005	0.9625	0.9637 2	0.9556 2	0.9459
VIF	0.9721 1	0.9719 1	0.9699 2	0.9439 2	0.9828 2	0.9706 1	0.9649 1	0.9584 1

Sheikh, Sabir & Bovik, *Trans on IP*, Nov 06

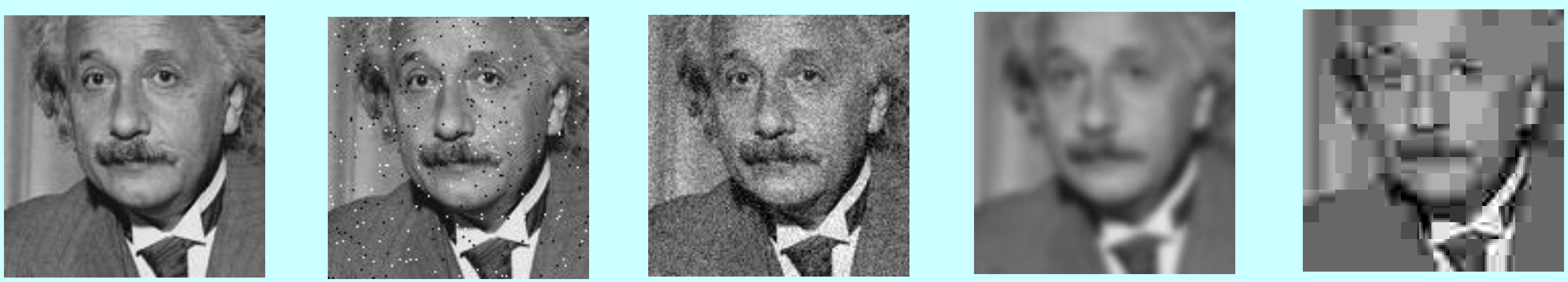
Note: IFC is VIF w/o divisive normalization

THE MSE

- For 40 years the Mean-Squared Error has dominated signal quality assessment
- As well as design and optimization

DUMP THE MSE!

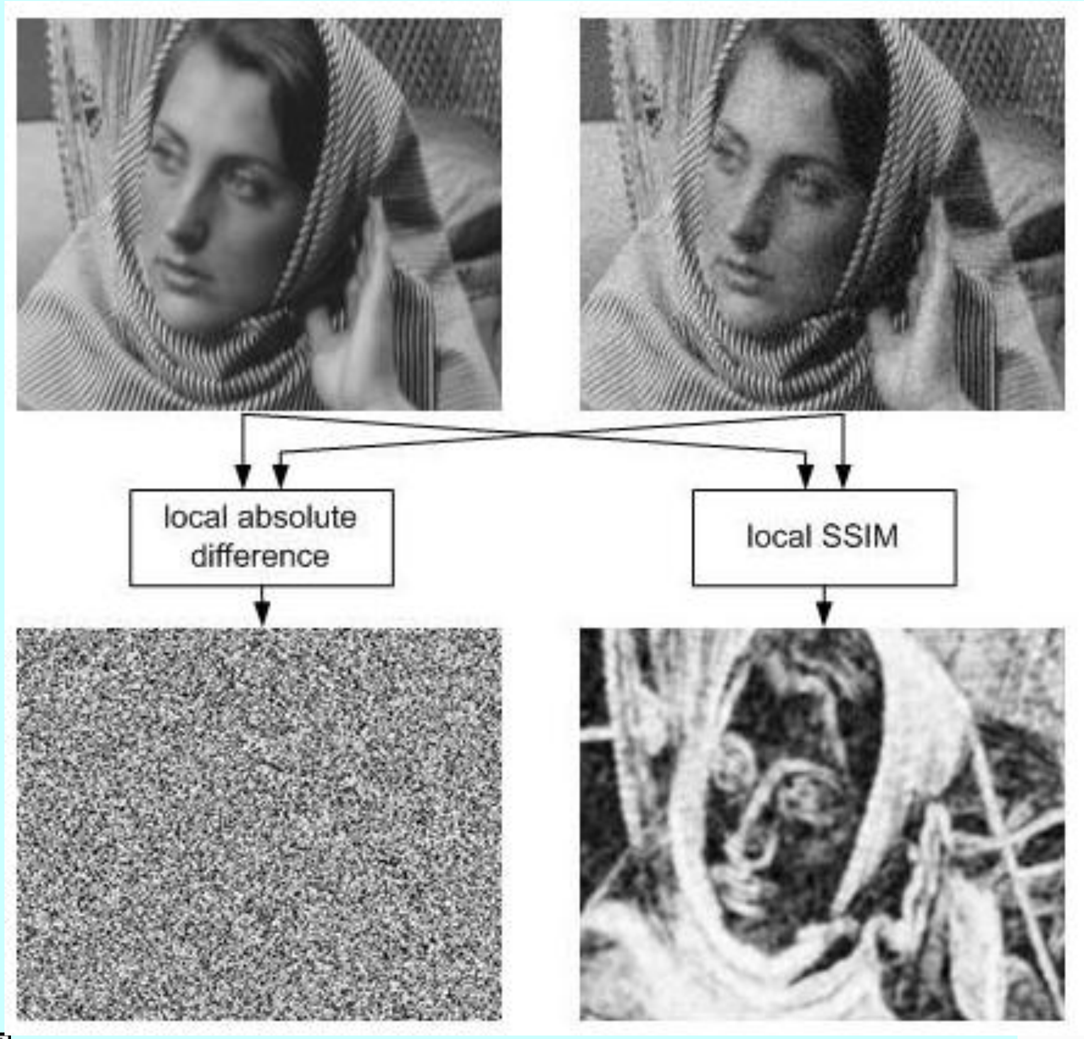
- The MSE and hence PSNR are (generally) **awful** measures of image quality.



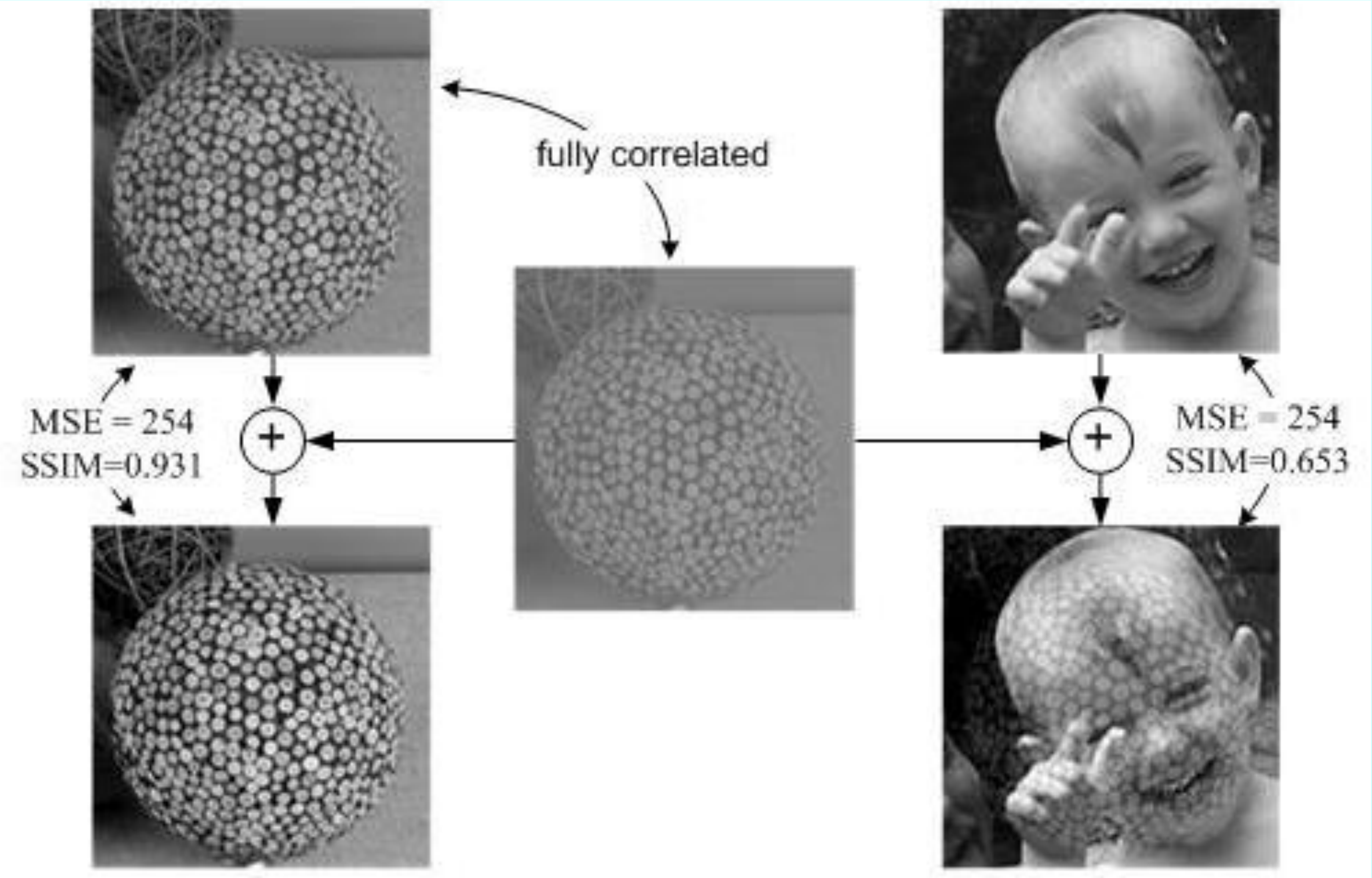
(a) | (b) MSE = 313 SSIM = 0.730 | (c) MSE = 309 SSIM = 0.576 | (d) MSE = 308 SSIM = 0.641 | (e) MSE = 309 SSIM = 0.580

Einstein altered by different distortions. (a) reference image; (b) impulse noise; (c) Gaussian noise; (d) blur; (e) JPEG compression.

DUMP THE MSE!!



DUMP THE MSE!!!



Towards Video Quality Assessment

- ***Frame-by-frame*** SSIM and VIF produces competitive results relative to sophisticated receiver-based algorithms.
- Video distortions are very different from pure spatial distortions and require **spatio-temporal measurements**.
- **Temporal masking** effects play an important role in the perception of spatial distortions.
- **Evaluation of VQA algorithms no mean task** (more later)

Video Distortions

- ***Spatio-temporal artifacts*** include *ghosting*, *motion blocking*, *motion compensation mismatches*, *mosquito effect*, *jerkiness*, *smearing*, and more.

- Quality Assessment of videos distorted by such processes must rely on **effective handling of motion.**

Motion and Optical Flow

- Simple method – Differential-VIF (Sheikh-Bovik '05)
- VIF operating in the wavelet ***derivative*** domain. Improved performance relative to frame-by-frame.
- Current (developing) approach: model optical flow and ***measure video quality along the motion trajectories***

Performance

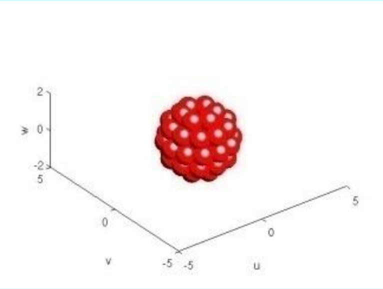
Spearman Rank Order Correlation Coefficient (SROCC) between subjective and objective scores for different quality metrics. *Proponent P8 is the best performing metric tested by the VQEG in terms of SROCC

Quality Model	SROCC
PSNR	0.786
Proponent P8 (Swisscom)*	0.803
Frame-by-Frame SSIM (Wang '04)	0.812
D-VIF (Sheikh '05)	**0.849**

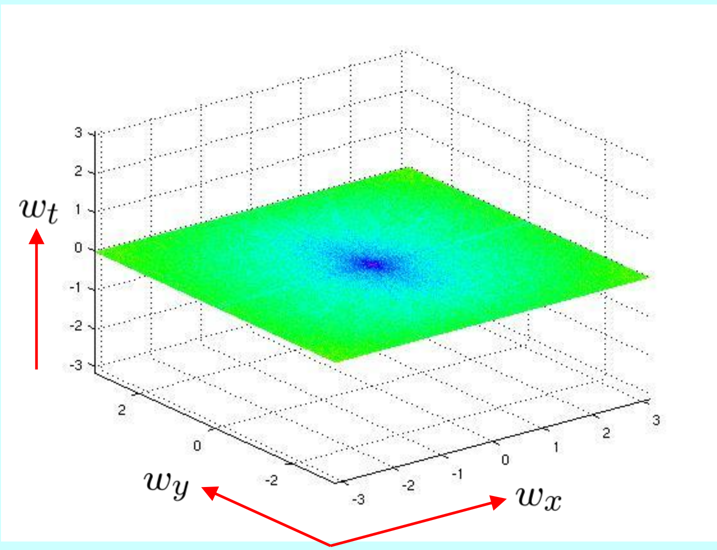
Current: Optical Flow Modeling

- **Assume:**
  - Video segments (without **scene changes**) consist of *local (instantaneously) translating image patches.*

- **Model:**
  - Combines GSM model for natural images with local patch translation model: local motion induces **spatio-spectral planes of higher energy**
  - **3-D Gabor filterbank** – based optical flow algorithm deployed to detect motion energy

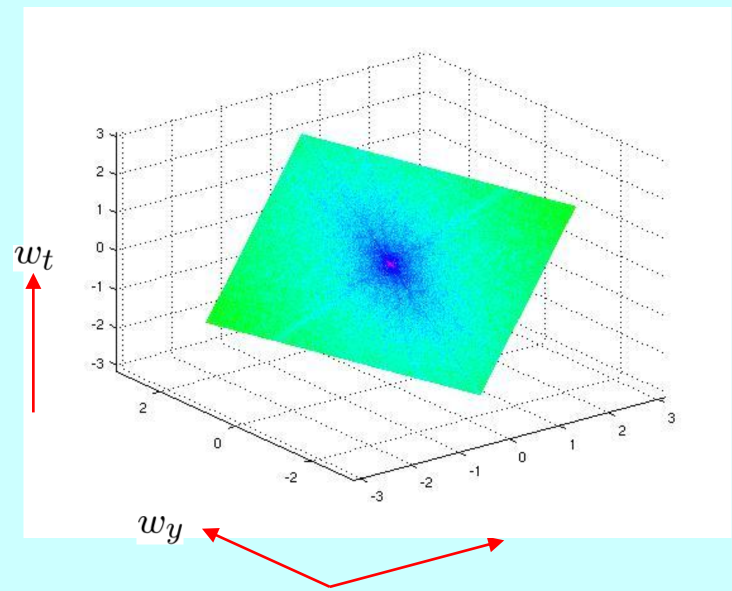


Illustration



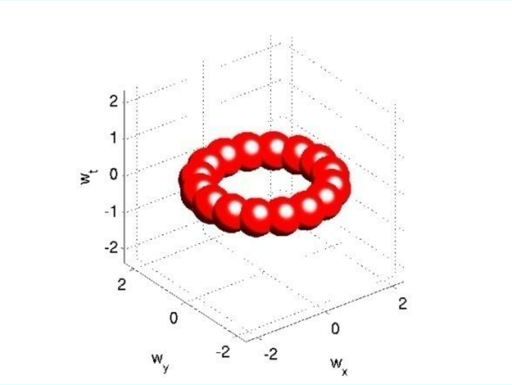
Fourier Transform
of a static sequence

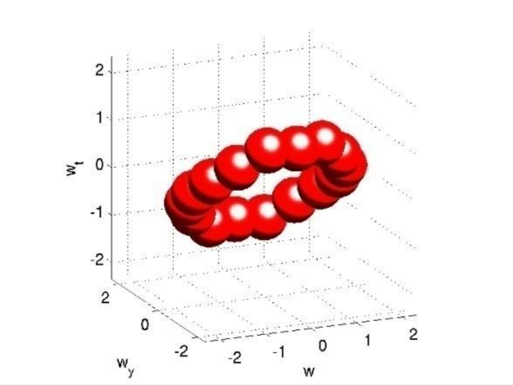
Fourier Transform of
a sequence in motion



Filter Subset Selection

- Subsets of Gabor filterbank used for VQA: those that intersect the local motion plane





Filterbank automatically chosen for (left) static sequence (right) translating sequence.

Proposed Video SSIM (V-SSIM)

- We compute V-SSIM in the complex wavelet domain.

$$\text{V-SSIM}(\vec{f},\vec{g}) = \frac{\sum_{i=1}^{N} \left| f_k \, g_k \right| + K}{\sum_{i=1}^{N} \left| f_k \right|^2 + \left| g_k \right|^2 + K}$$

- Subband coefficients *f*, *g* computed from the active Gabor filters.

- V-SSIM – **still under development.**

Proposed Video VIF (V-VIF)

- The V-VIF Index from i^{th} **active** sub-band :

$$\text{V-VIF}_i(x) = \frac{I\left[C_i(x); F_i(x) \middle| \hat{z}_i(x)\right]}{I\left[C_i(x); E_i(x) \middle| \hat{z}_i(x)\right]}$$

- Variance field $\hat{z}$ estimated from **sub-band energies along the motion trajectories.**
- Measures info the HVS can extract from the ***distorted video,*** normalized by the info the HVS can extract ***from the reference video.***

- V-VIF **still under development.**

A LIVE Video Quality Database

- We have begun to create a **LIVE VQA Database** of generic power freely available to the research community.

- We shall provide **subjective scores** (MOS, DMOS) for the distorted videos.

Towards a Video Quality Database

- VQEG Phase-I FR-TV database has significant limitations. Most reference and distorted videos are ***interlaced*** - hence visual artifacts in the *reference* as well as distorted video sequences.

- **De-interlacing is inappropriate** in a VQA framework.

- The VQEG database consists *only* of compression-related artifacts produced by e.g., H.263 and MPEG codecs.

Towards a Video Quality Database

- **Acquiring high-quality, progressive scanned, copyright free source videos is difficult.** We've obtained ~ 12 HD videos.
- We've created a GUI to perform **Single Stimulus Continuous Quality Evaluation (SSCQE)** experiments - subjects provide a *time-dependent* index of quality - well suited to applications such as video monitoring and quality control.
- Our psychometric study will be done in consultation with noted visual psychologists and frequent collaborators L. Cormack and W. Geisler.
- Envision that the resulting database, **with a wide diversity of distortions,** will prove more challenging than current VQEG database, and will enable more rigorous performance evaluation of QA systems.

Beyond QA: Using QA Indices for Other Things

- SSIM Index is well-suited to other applications, since not specific to any receiver or transmitter models.
- We are exploring its utility for **other types of signals and other applications.**
- For example we have developed **automated inspection systems** based on SSIM:
  - The US Postal Service is using SSIM **to evaluate letter-reading cameras**.
  - The US Mint is using SSIM to **detect minted coin defects.**

SSIM Applications by Others

- SSIM has been used for signal fidelity/quality assessment in many applications, including

  text recognition, palmprint verification, face recognition, image fusion, content retrieval/indexing, image/video compression, watermarking, denoising, color image quality, retinal and see-through wearable displays, video hashing, and visual surveillance, etc.

- In very diverse areas: digital camera design, IR imaging, MRI imaging, remote sensing, ATR, chromosome imaging, industrial control, etc.
- Deployed in popular public-domain software such as the MSU *Video Quality Measurement Tool* and the award-winning freeware H.264 codec *x.264*

What Really Excites Me

- **Perceptual optimization** using image / video quality indices!
- Much of what we have "optimally" designed over the past 30+ years should be re-examined
- Signal restoration, denoising, enhancement, reconstruction, compression, display, quantization, scaling, recognition, detection, tracking …. etc etc etc
- Seek optimization using accurate perceptual measures– rather than "perceptual data."

LABORATORY FOR IMAGE AND VIDEO ENGINEERING
LIVE
THE UNIVERSITY OF TEXAS AT AUSTIN

UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING

Example: Optimal Linear Image Restoration

- Classic blur + noise

$$\mathbf{y} = \mathbf{g} * \mathbf{x} + \mathbf{n}$$

- MMSE approach: find best linear filter that minimizes

$$E\left[\left(\hat{\mathbf{x}} - \mathbf{x}\right)^2\right]$$

over all

$$\hat{\mathbf{x}} = \mathbf{h} * \mathbf{y}$$

SSIM-Optimal Linear Image Restoration

- **Maximum SSIM approach**: find best linear filter that maximizes statistical SSIM Index:

$$Stat-SSIM(\tilde{\mathbf{x}},\tilde{\mathbf{y}})=\left(\frac{2\mu_{\mathbf{x}}\mu_{\mathbf{y}}+C_1}{\mu_{\mathbf{x}}^2+\mu_{\mathbf{y}}^2+C_1}\right)\left(\frac{2E\left[\left(\tilde{\mathbf{x}}-\mu_{\mathbf{x}}\right)\left(\tilde{\mathbf{y}}-\mu_{\mathbf{y}}\right)\right]+C_2}{E\left[\left(\tilde{\mathbf{x}}-\mu_{\mathbf{x}}\right)^2\right]+E\left[\left(\tilde{\mathbf{y}}-\mu_{\mathbf{y}}\right)^2\right]+C_2}\right)$$

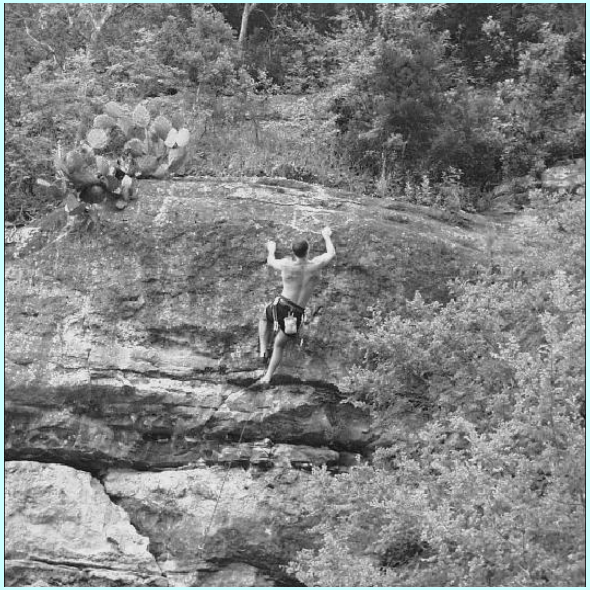
  over all

$$\hat{\mathbf{x}}=\mathbf{h}*\mathbf{y}$$

- We have solved this problem in a near closed form, computationally efficient manner.

LABORATORY FOR IMAGE AND VIDEO ENGINEERING LIVE THE UNIVERSITY OF TEXAS AT AUSTIN

UNIVERSITY OF TEXAS AT AUSTIN UT ECE ELECTRICAL & COMPUTER ENGINEERING



(a)

(b)

blur+noise

MMSE

(c)

(d)

SSIM-optimal

Questions?

LIVE's IQA/VQA Sponsors

AT&T

NSF

KLA Tencor
Accelerating Yield

AFRL
THE AIR FORCE RESEARCH LABORATORY
LEAD | DISCOVER | DEVELOP | DELIVER
AFOSR

21st CENTURY
TECHNOLOGIES

THE DEPARTMENT OF THE TREASURY
1789

UNITED STATES
POSTAL SERVICE®

NATIONAL
INSTRUMENTS

LABORATORY FOR IMAGE AND VIDEO ENGINEERING
LIVE
THE UNIVERSITY OF TEXAS AT AUSTIN

BOEING

UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING