Quality is in the Eye of the Beholder

Al Bovik

October 15, 2008

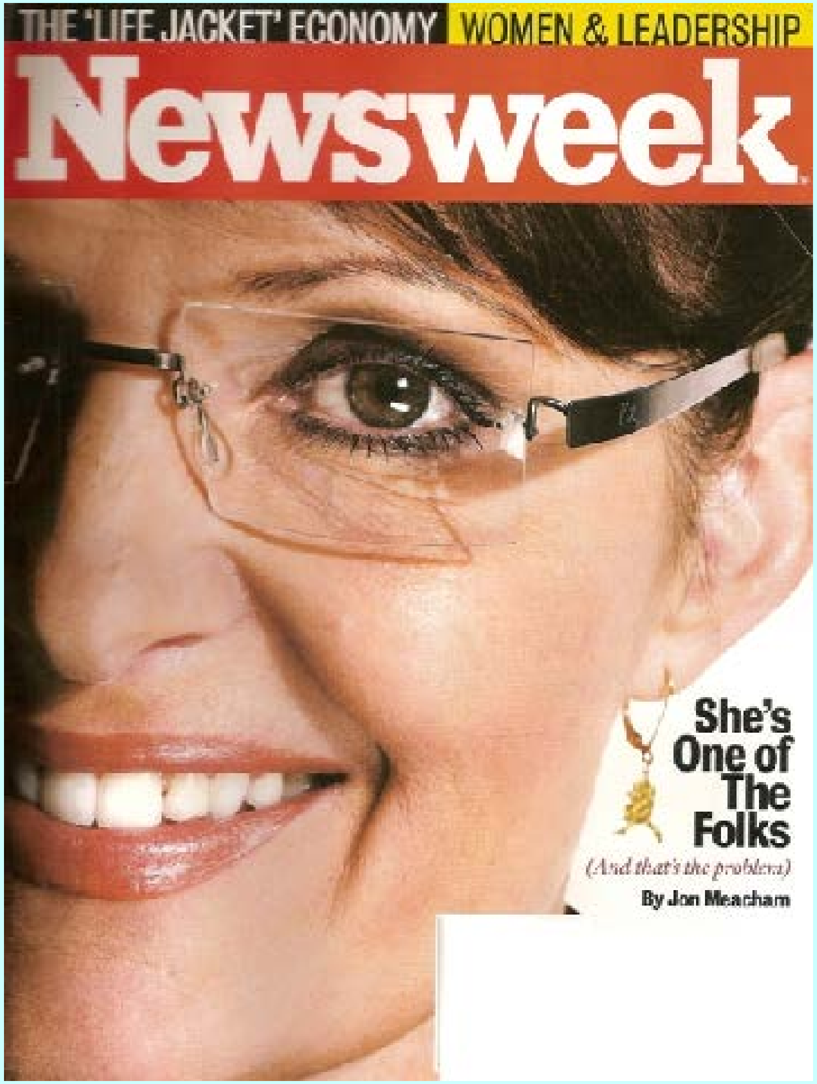


Image quality **too** good?

Theme #1

An analogy that I will develop:

- Assessing the quality of visual signals
- Measuring the fidelity of a visual communication system

are similar problems

A Classic Communication System

Tenet of Communication Theory

- The more known (that we can model) about

  transmitter

  channel

  receiver

  the better job of **communication**

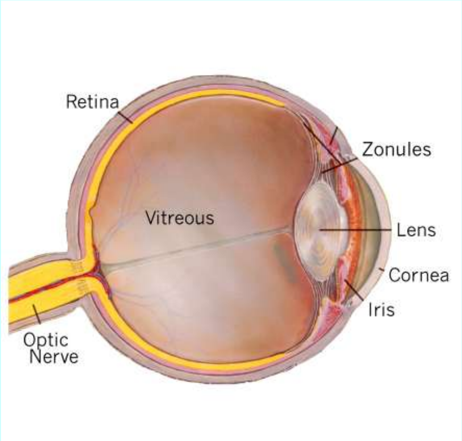
Image Quality Assessment

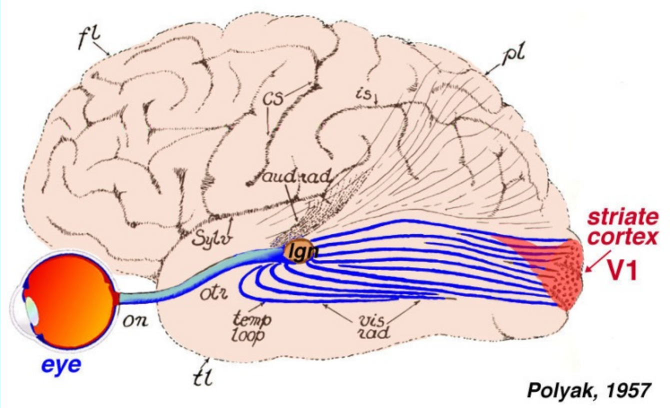
What are transmitter, channel, and receiver….?

The Natural Image Transmitter

Photos of
natural image transmitter

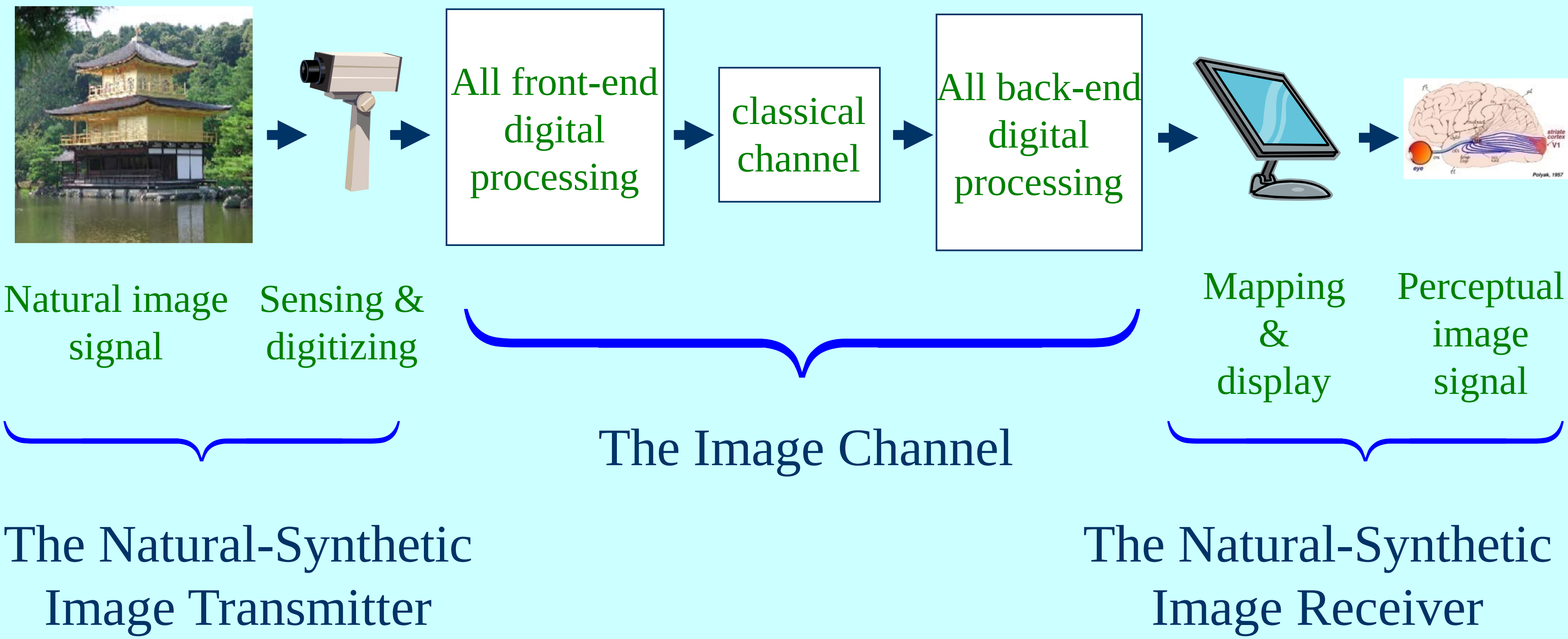
The Natural Image Receiver



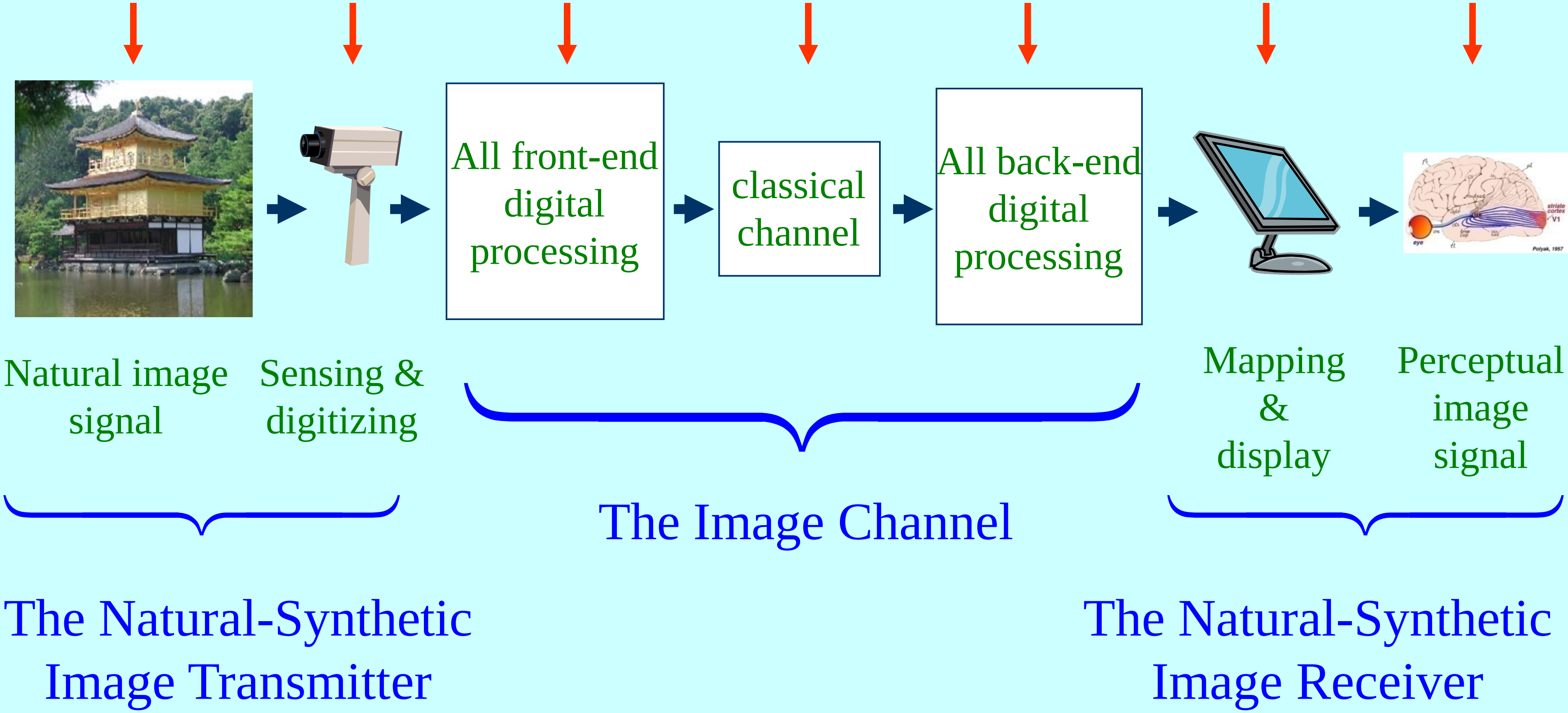


Depictions of natural image receiver

Overall Communication System



Sources of Image Distortion



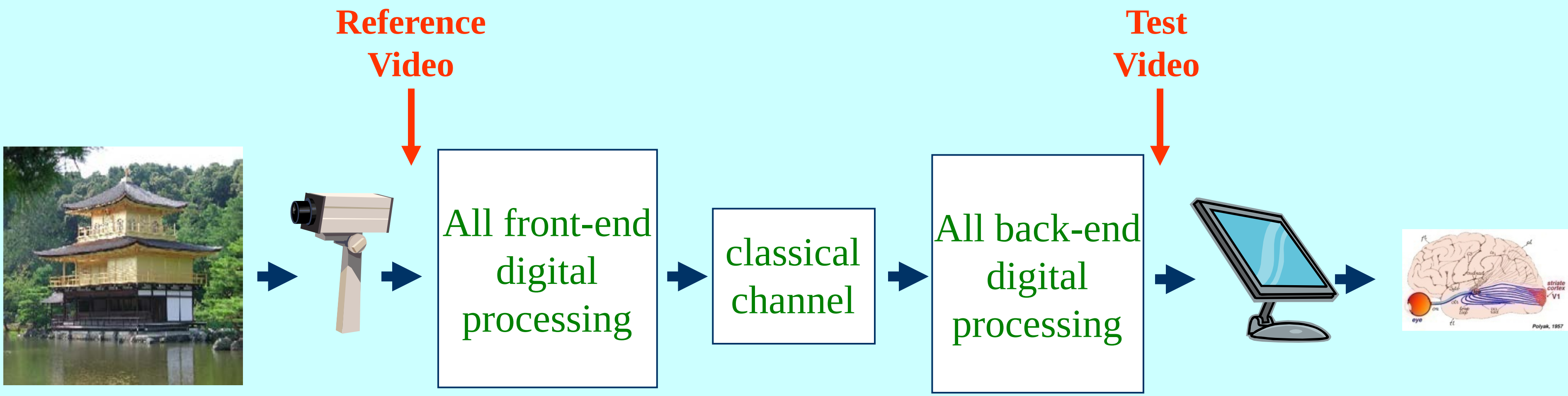
Theme #2

- **Quality Assessment Algorithms** are possible that **correlate highly** with **subjective judgment.**

“Nothing can be beautiful which is not true.”

– John Ruskin

Full-Reference Quality Assessment



- Need accurate **models of transmitter.**
- Need accurate **models of the receiver**

Two Relevant Algorithms

Two still image quality assessment (IQA) algorithms **relevant** to later discussion …

Structural Similarity (SSIM) Index

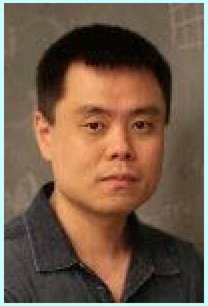
- Weighted local (patch) **image statistics** create a **SSIM map**:

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

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

- Mean SSIM Index

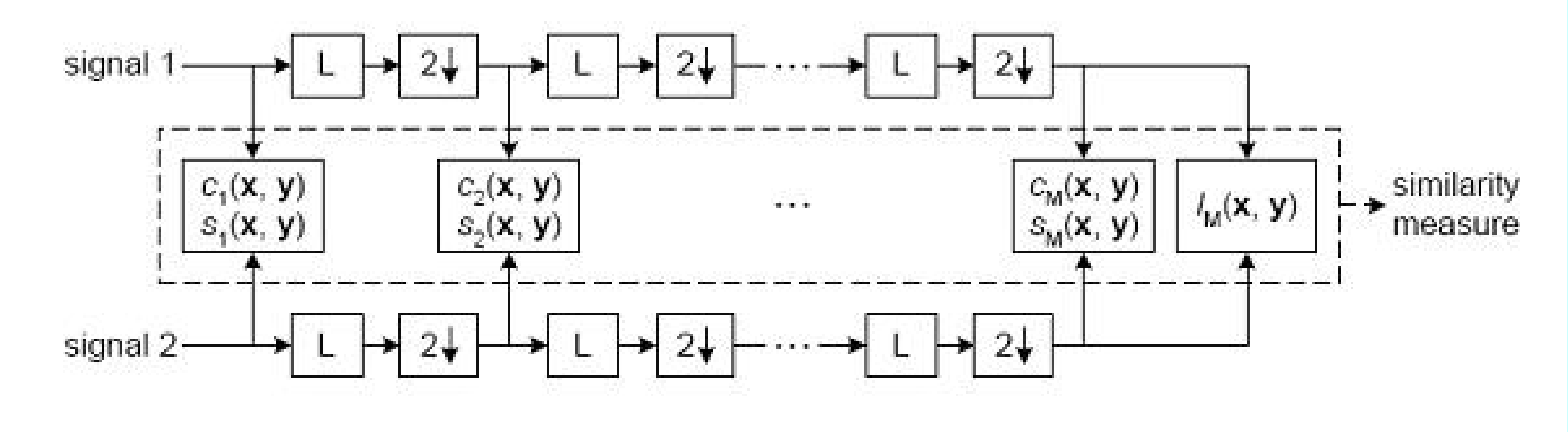
$$\mathrm{SSIM}(\mathbf{I,J}) = \left(\frac{1}{NM}\right)\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\mathrm{SSIM}_{\mathbf{I,J}}(i, j)$$



Zhou Wang

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

Multi-Scale SSIM

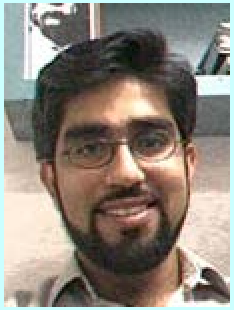


SSIM calculated over *scale space*

Multi-scale SSIM (MS-SSIM) calculated on dyadic pyramid yields better performance

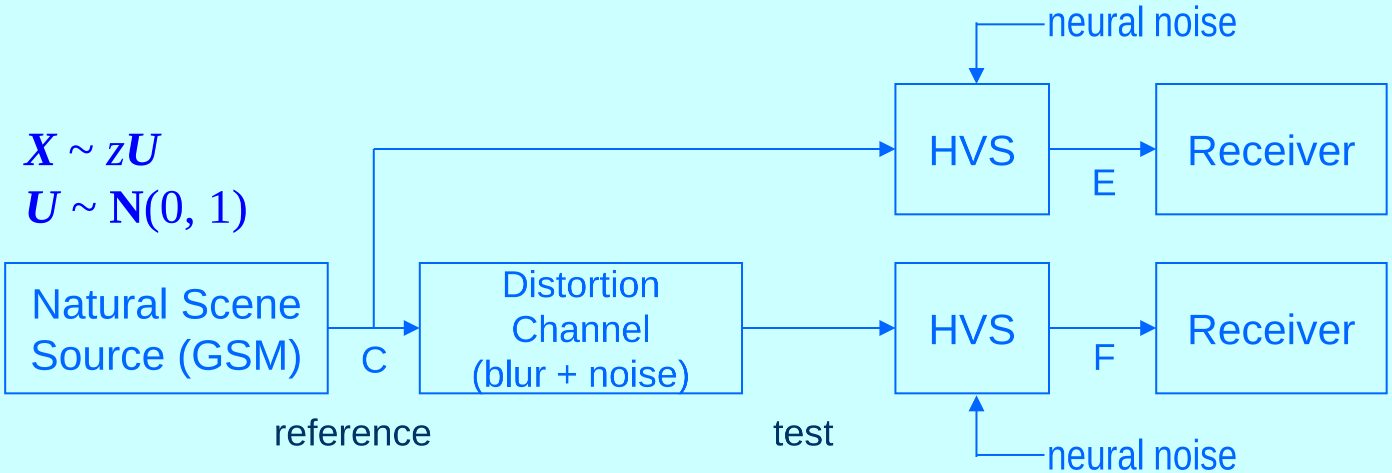
Wang, Simoncelli & Bovik, *Asilomar, Nov 2003*

Visual Information Fidelity Index



Hamid Sheikh

$\boldsymbol{X} \sim z\boldsymbol{U}$
$\boldsymbol{U} \sim \mathbf{N}(0, 1)$



$$\mathrm{VIF} = \frac{I(C;F\,|\,z)}{I(C;E\,|\,z)} = \frac{\text{info HVS can extract from distorted image}}{\text{info HVS can extract from original image}}$$

- $I(C; F|z)$ = **mutual information** in wavelet domain conditioned on variance field z

Sheikh & Bovik, *Trans on IP*, Feb 06

SSIM and VIF are Related

- **Under GSM model** we have been able to show

  VIF and Multi-scale SSIM

  are essentially identical.

- Consequently, the **efficacy of MS-SSIM is explained** in information-theoretic sense under GSM model.

Seshadrinathan & Bovik, *ICIP 2008*

Relative Performance

Yuck!

LIVE Image Quality Assessment Database: >25,000 subjective (DMOS) judgments.

Spearman Rank-Order Correlation Coefficient (SROCC)

	JPEG2K #1	JPEG2K #2	JPEG #1	JPEG #2	WN	Gaussian Blur	Fast Fading Noise	All Data
PSNR	0.93	0.86	0.88	0.77	**0.99** 1	0.78	0.89	**0.82**
JND	**0.96** 2	0.96	0.96	0.92	0.95	0.94	0.91	0.93
DCTune	0.83	0.72	0.87	0.82	0.93	0.67	0.77	0.80
PQS	0.94	0.92	0.94	0.90	0.95	0.93	0.94	0.93
NQM	0.95	0.94	0.94	0.90	**0.99** 1	0.85	0.82	0.91
Fuzzy (S7)	0.93	0.90	0.91	0.80	0.92	0.61	0.91	0.83
BSDM (S4)	0.91	0.94	0.91	0.92	0.93	**0.96** 2	**0.94** 2	0.93
VSNR	0.95*	0.95*	0.91*	0.91*	**0.98** 2	0.94	0.91	0.89
MS-SSIM	**0.96** 2	**0.97** 1	**0.97** 1	**0.95** 1	**0.98** 2	0.95	**0.94** 2	**0.95** 2
VIF	**0.97** 1	**0.97** 1	**0.97** 1	**0.94** 2	**0.98** 2	**0.97** 1	**0.97** 1	**0.96** 1

*Data available only for combined JPEG & JPEG2K results

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

Theme #3

- QA algorithms are **not just interesting research problems.**
- They are practical ways of **benchmarking image processing algorithms** of every flavor.
- They can **remove the human element** when deciding algorithm performance…..
- …. while still accounting for **human judgment** of performance.

Challenge to the Community

- For **decades** we've been **eyeballing** image processing results or using the MSE/PSNR.
- My challenge to image processing algorithm designers: **assess and report your results** using a perceptually significant IQA/VQA metric
- Restoration; denoising; deblocking; reconstruction; representation; compression; inspection; network and wireless channel benchmarking, **etc etc**

Theme #4

- **Perceptual optimization is a next big thing.**
- **Or should be!**

What Excites Me

- **Perceptual optimization** using Quality Indices as **objective functions!**

- What we've "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**

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

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

Sumohana Channappayya

original

blur+noise

SSIM-Optimal 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 solved this **quasi-convex** problem in a **near closed form** computationally efficient manner.

Channappayya and Bovik, *Trans IP*, Aug 2008

Local MMSE-optimal

SSIM-optimal

Theme #5

Video Quality Assessment is more important, harder, and requires better modeling than still image QA.

Digital Video is Taking Over the World

“Without impermanence, nothing is possible”

- *Thich Nhat Hanh*

“Motion is the very essence of what has hitherto been called matter”

- *Lord Kelvin*

Video Distortions

- There **many** distortions that occur commonly in video.

- **Spatial = "Mostly Spatial"**

  - **Blocking artifacts** (compression)
  - **Ringing** (compression)
  - **Mosaicking** (block mismatches)
  - **False contouring** (quantization)
  - **Blur** (acquisition or compression)
  - **Additive Noise** (acquisition or channel)

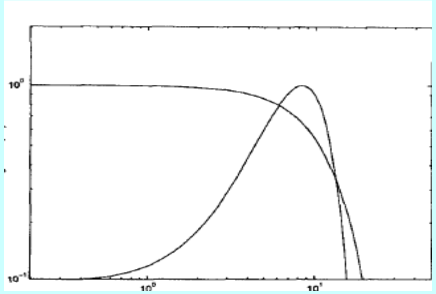
Temporal Distortions

- **Temporal = “Mostly Temporal”**
  - **Ghosting** (poor motion)
  - **Motion blocking** (propagation of block artifacts)
  - **Motion compensation mismatches** (ambiguity)
  - **Mosquito edge effects** (poor correction of ringing)
  - **Packet loss/error concealment** (ARQ, FEC)
  - **Stationary area fluctuations** (texture flutter)
  - **Jerkiness** (temporal aliasing)
  - **Smearing** (slow acquisition)

Whew!

Competitive VQA Algorithms

- **Frame MS-SSIM/VIF**[1] – MS-SSIM/VIF applied to frames
- **“Swisscom P8” - Leading** VQEG FRTV Phase 1 Test proponent.
- **Video Quality Metric (VQM)**[2] from NTIA (an ANSI and ISO **standard)**. Leading VQEG Phase 2 Test proponent (non-public study) .
- No prior VQA algorithm has used **motion estimates** or **motion tuning –** to compute VQA along **motion trajectories.**
- Some have used very simple **temporal filtering** w/o motion handling.



[1]Wang, Lu & Bovik, *Image Commun.* ’04
[2]Pinson & Wolf, *IEEE Trans Broadcasting*, ‘04

Motion handling offers the greatest potential for improving VQA algorithms.

Perception of Motion

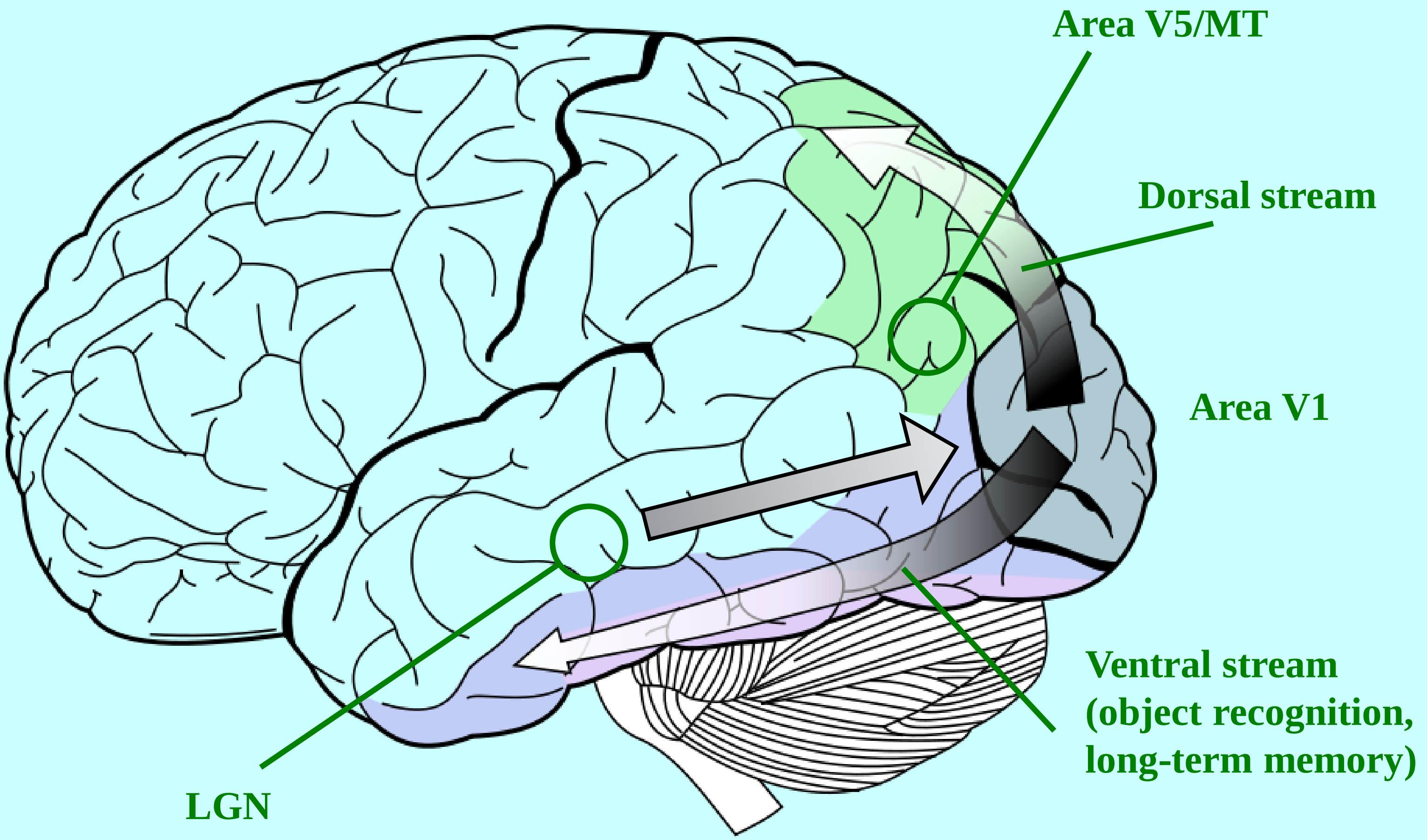
- The ***dorsal stream*** of visual data passes through **Area V1** of primary visual cortex to **Area V5 (Area MT – middle temporal)**

- **Area V1:** Multichannel space-time decomposition of visual data occurs in V1: **patterns, direction, speed localized**

- **Space-time data passed to Area MT**, where space-time data is integrated into **motion estimates**[1]

[1]Amongst other tasks such as visual location, eye and arm movements, etc.

Flow of Visual Data

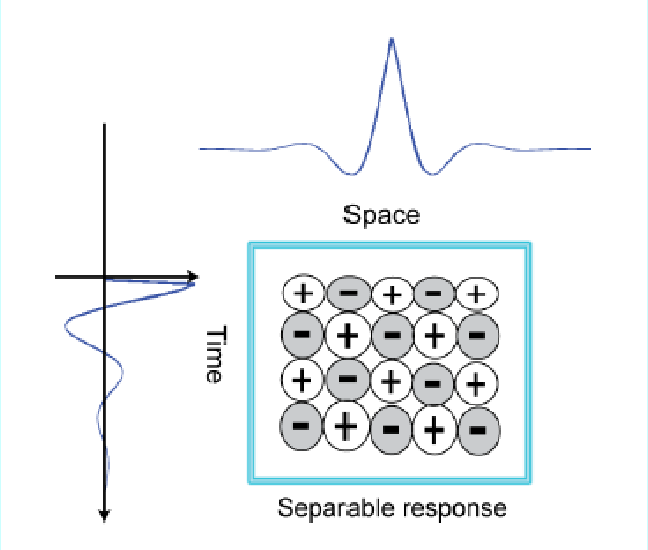


Area V1 Models

- **V1 Spatial** receptive field model: **Gabor functions** in quadrature pairs.

- **V1 Temporal** receptive field model: **Causal** gamma-modulated sinusoids

- Space-time responses **separable**



3-D Gabor Model

- **Causality** not required for full-reference QA (not typically real-time)

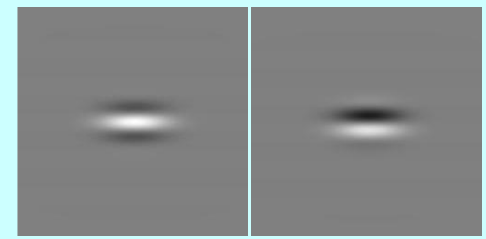
- **Separable time** and **space Gabor filters**:

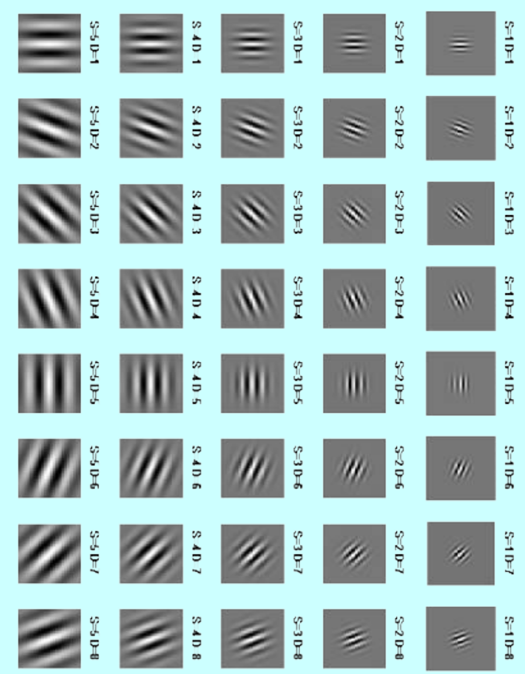
$$g(x,y,t) = K\left[e^{-t^2/2\gamma^2} e^{2i\pi(w_0 t)}\right] e^{-\left[\left(x/\lambda\right)^2 + y^2\right]/2\sigma^2} e^{2i\pi(u_0 x + v_0 y)}$$

- **Optimally localized** in space-time-frequency.

Brief History of 2-D Gabor Functions

- **1980**: **1-D** Gabor model **of V1** cortical fields (Marcelja)
- **1985**: **2-D** uncertainty-optimal **Gabor model of V1** cortical fields (Daugman); now dominant V1 spatial model.
- **1986**: First proposed for **textured image analysis**; now dominant texture filter primitives (Bovik, Clark, Geisler, Turner)
- **1987**: First proposed for **motion computation**; now dominant optical flow basis functions (Heeger; Fleet & Jepson 1990)
- **1989**: First proposed for **stereo**; now dominant stereo phase matching basis functions (Fleet & Jepson 1989)
- **1993**: Dominant primitives for **Iris Recognition** (Daugman 1993)
- **1999**: Dominant primitives for **Face Recognition** (Wiskott 1999)





A Spatio-Temporal VQA Algorithm

- We've recently created a Video Quality index that performs quite well:

MOtion-based **V**ideo **I**ntegrity **E**valuation index,[1] or **MOVIE index**

- **Spatial & temporal** distortion assessment
- Operates in subband (Gabor) space-time-frequency
- **Assesses temporal quality along computed motion trajectories**
- Models **Area MT motion tuning and motion weighting**
- Embodies **visual masking**
- Combines principles from **SSIM** and **VIF**
- **Information-theoretic optimal** under natural scene statistic model

Kalpana Seshadrinathan

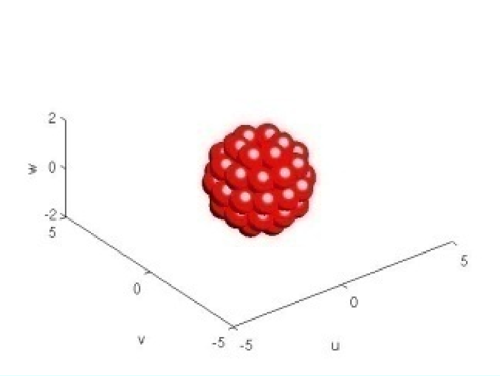
[1]Seshadrinathan & Bovik, "Spatio-temporal Quality Assessment of Natural Videos," *IEEE Trans Image Processing*, submitted, 2008.

Spatio-Temporal Decomposition

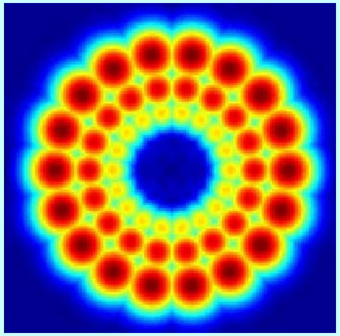
- The MOVIE index is defined as a product:

  (Spatial MOVIE) x (Temporal MOVIE)

- In both: videoss (reference *f* and test *t*) decomposed by a **3-D multi-scale Gabor filterbank**



3-D Gabor filterbank in
frequency space (one scale only)



Slice through 3-D spatial
Gabor filterbank
in frequency space

Outline of Spatial MOVIE Index

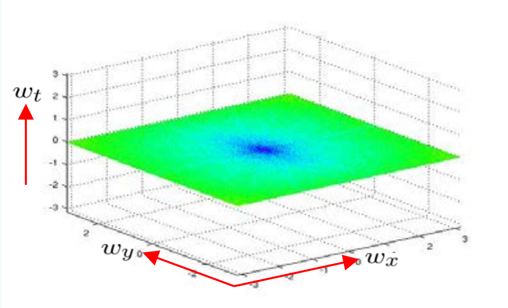
- Compare **amplitude responses** of 3-D Gabor filters to **test** and **reference videos**.
- Local **Gabor-domain SSIM/VIF-like computation** is made.
- **Mutual masking principle** is used (masking on both reference and test video)
- Overall **Spatial MOVIE Index** pools quality scores over scale/band, space, and time.

Outline of Temporal MOVIE Index

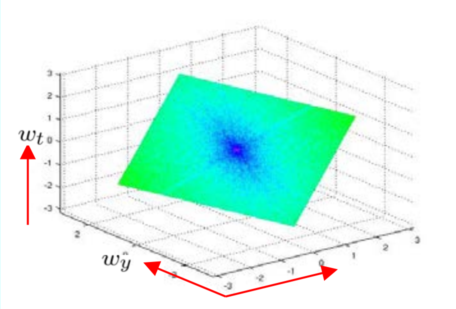
- Evaluates **temporal quality** along **motion trajectories** computed using **3-D Gabor phase-based optical flow (Fleet *et al*, 1990).**

- **Local motion** of patches gives rise to **orientations** in space-time frequency:

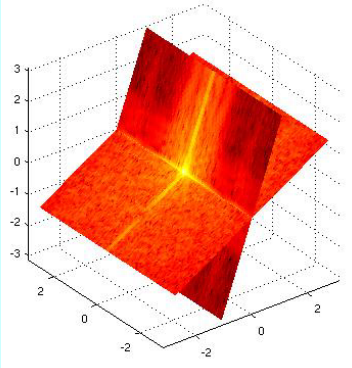
Fourier transform of static patch



Fourier transform of patch in motion



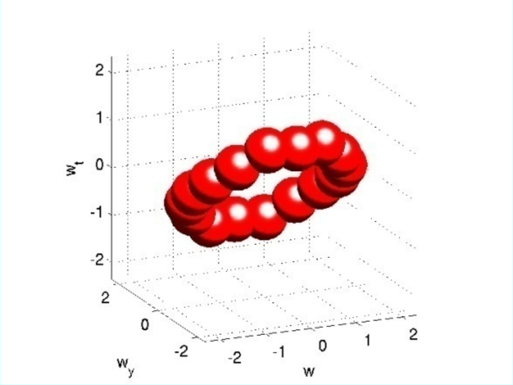
- Temporal MOVIE may be viewed as finding misalignments between local orientations of flow



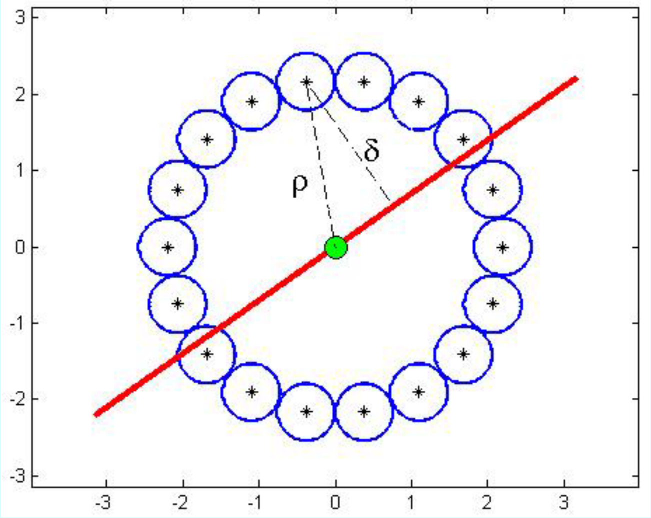
Misaligned spectra of corresponding patches from reference and distorted videos.

Area MT Motion Tuning Model

- Outputs of Gabor filters combined **to allow motion tuning.**[1]
- Gabor responses **weighted by function of distance** from translational plane.
- Filters close to plane given **excitatory weights** α_k, others **inhibitory weights**.



Excitatory responses at one scale



Motion plane and filters at one scale - weights are functions of ρ and δ

$$\alpha = \frac{\rho - \delta}{\rho}$$

shifted/scaled across responses to be zero-mean.

[1]Simoncelli & Heeger, *Vision Research*, 1998.

Motion-Tuned Responses

- The **motion tuned amplitude responses** of 3-D space-time Gabor filters to **test** and **reference videos** are compared.
- A **Gabor-domain SSIM/VIF-like computation** is made.
- Overall **Temporal MOVIE Index** pools quality scores values over scale/band, space, and time.

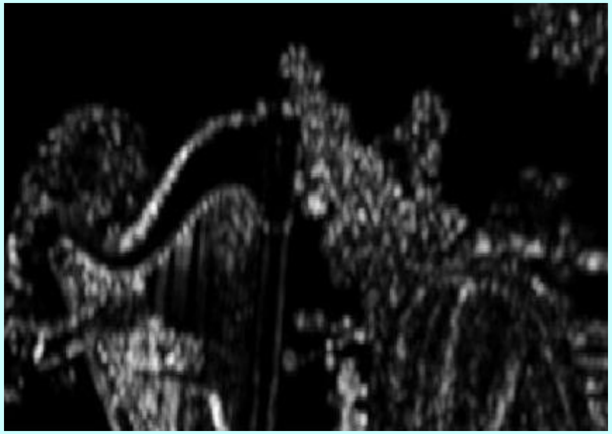
MOVIE Index Maps

- Spatial & temporal MOVIE indices displayed as **Quality Map Videos** (bright = larger errors).

Reference

Test

Temporal MOVIE map



Spatial MOVIE map

View video

Final MOVIE Index

- Overall **MOVIE Index** is separable combination of Spatial and Temporal MOVIE Indices:

$$MOVIE(f,g) = Spatial\ MOVIE(f,g) \times Temporal\ MOVIE(f,g)$$

- MOVIE contains **no tuned parameters**. No dataset training.

Performance of MOVIE

- Assessed on **VQEG FRTV Phase 1 Dataset**.
- 20 reference sequences, 16 distortions of each
- 4 videos are **artificial animations** (floating letters on constant background, etc) - **unnatural**
- Scores tabulated in following Tables.

Performance Comparison

Spearman Rank Order Correlation Coefficient (SROCC) Comparison

Quality Model	SROCC
PSNR	0.79
Proponent P8 (Swisscom)*	0.80
Frame SSIM (Wang ’04)	0.81
MOVIE	**0.83**
MOVIE (no animations)**	**0.86**

*Proponent P8 = best performing metric tested by VQEG

MOVIE is designed using natural scene statistic model. Animations (constant regions with step edges) don’t satisfy NSS models. Other indices’ behavior **varies little when animations are removed.

THEME #6

- A **publicly available** Video Quality Database is badly needed.
- **Nobody** is happy with the VQEG Phase 1 database
- **No other publicly-available VQA database**

A LIVE Video Quality Database

- We are making available a **LIVE VQA Database** of generic power freely available to the research community.

- We provide **subjective scores** (DMOS) for the distorted videos from a large human study.

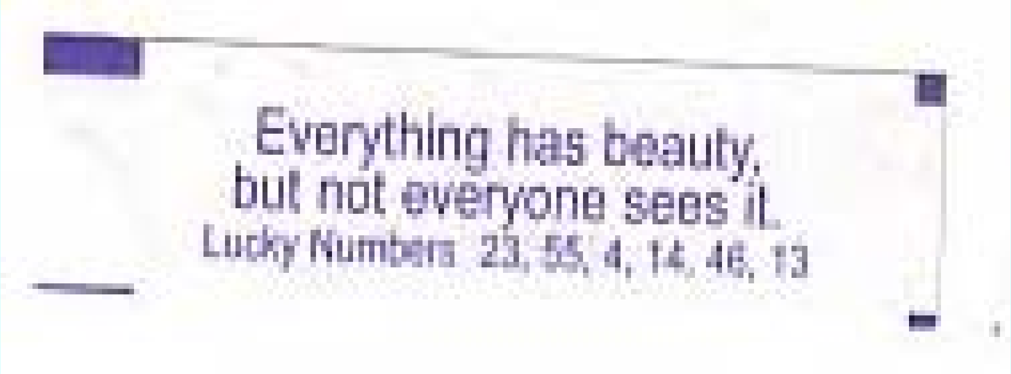
Kalpana Seshadrinathan

Rajiv Soundararajan

“Everything has beauty, but not everyone sees it.”

- Confucius

Corollary: All videos have distortions, but not everyone sees them the same way.


Everything has beauty,
but not everyone sees it.
Lucky Numbers 23, 55, 4, 14, 46, 13

Towards a Video Quality Database

- VQEG Phase-1 FRTV database **limitations**:
  - Reference & distorted videos ***interlaced***
  - *Only* compression-related artifacts; e.g., H.263 and MPEG.
  - Distorted videos have **poor perceptual separation**.

LIVE Video Quality Database

- 10 reference videos supplied by Technical University of Munich free of charge.
- All **progressively scanned** YUV420, 768x432, 10s duration, 25 fps and 50 fps.
- **Diverse assortment of distortions** - more challenging VQEG, enabling more rigorous performance evaluation of VQA systems:
  - Compression artifacts from modern codecs **(MPEG-2, H.264)**
  - **Packet loss errors** from **wireline (IP)** environment
  - **Packet loss errors** from **wireless** environment
- Each reference video subjected to **15 distortions**

Distorted Videos

- **(ISO[1]) MPEG-2** distortions: bitrates **700 Kbps - 4 Mbps**.
- **(JVT[2]) H.264** distortions: bitrates **200 Kbps - 5 Mbps**.
- **Simulated (VCEG[3]) IP errors** on H.264 stream. Loss rates: 3%, 5%, 10%, and 20%. Packetization: 1-4 slices/frame. Both I- and P-frame losses.
- **Simulated (VCEG[3]) wireless errors** on H.264 stream. Multiple slices/frame: short packets (~200 bytes). Both I-and P-frame losses.
- For each distortion, **perceptual separation of degradations** emphasized.
- **Example (low-res): H.264, 7Mbps, 3% packet loss, 4 packets/frame** (IP channel)

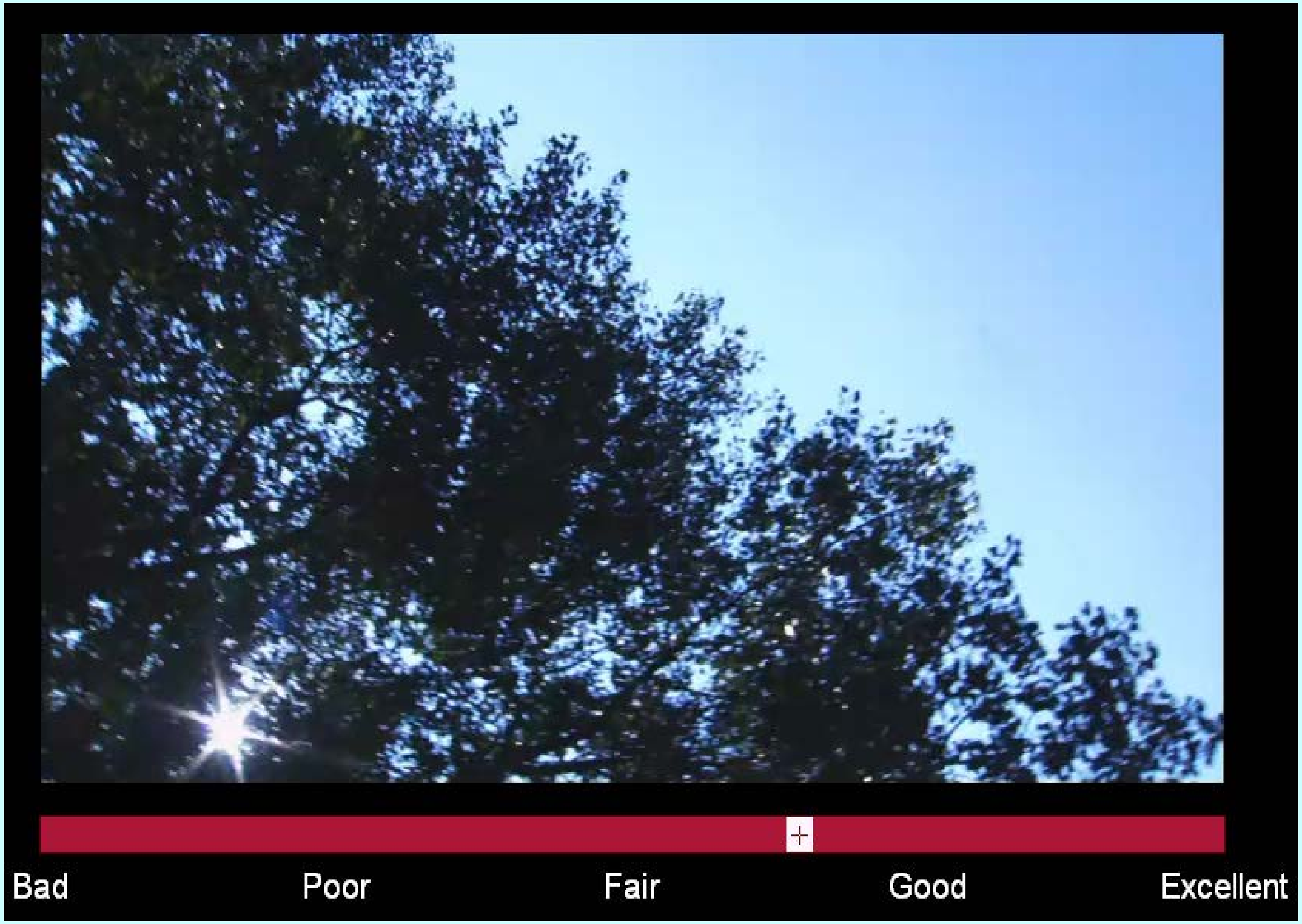
[1]ISO = International Organization for Standardization
[2]JVT = Joint Video Team
[3]VCEG = Video Coding Experts Group

Subjective Study

- 38 subjects viewed **150 test videos** (+10 hidden references) in two ½-hour sessions.
- **Single Stimulus Continuous Quality Evaluation (SSCQE)** (hidden reference) – continuous evaluation allows for fine gradations in subjective quality assessment.
- Subjects also **discretely** scored each video **at the end**.

Visual Interface

Visual Interface



Processing Subjective Scores

- **Difference scores** per session (hidden reference):

$$d_j(i, k) = s_{j\text{-}ref}(i, k) - s_j(i, k)$$

$s_j(i, k)$ = score assigned to **video *j*** by **subject *i*** in **session *k***

- Z-scores per session:

$$Z_j(i, k) = [d_j(i, k) - \mu_j(i, k)] / \sigma_j(i, k)$$

- Subject rejection using ITU-R BT 600.11

- DMOS of video = **average of Z-scores**

Performance of VQA Indices on LIVE Video Quality Assessment Database

Yuck!

Linear Correlation Coefficient (LCC) after Nonlinear Regression

Algorithm	Wireless	IP	H.264	MPEG-2	All
PSNR	0.46	0.41	0.48	0.38	0.40
SSIM	0.55	0.54	0.66	0.58	0.54
Multi-Scale SSIM	0.71	0.72	0.69	0.69	**0.74**
Speed-weighted SSIM[1]	0.58	0.58	**0.72**	0.64	0.60
VNSR[2]	0.70	**0.74**	0.65	0.68	0.69
VQM[3]	**0.74**	0.65	0.63	**0.80**	0.72
MOVIE	**0.81**	**0.73**	**0.77**	**0.75**	**0.77**

[1]Wang & Li, *J. Opt Soc. Amer.*, 2007.
[2]VSNR = "Visual Signal-to-Noise Ratio" – Chandler & Hemami, *IEEE Trans Image Process.*, 2007
[3]VQM = "Video Quality Metric" – Pinson & Wolf, *IEEE Trans Broadcasting*, 2004. Currently the ANSI/ISO standard.

Performance of VQA Indices on LIVE Video Quality Assessment Database

Yuck!

Spearman Rank-Order Correlation Coefficient (SROCC)

Algorithm	Wireless	IP	H.264	MPEG-2	All
PSNR	0.43	0.32	0.43	0.36	0.37
SSIM	0.52	0.47	0.66	0.56	0.53
Multi-Scale SSIM	**0.73**	0.65	**0.71**	0.66	**0.74**
Speed-weighted SSIM[1]	0.56	0.47	**0.71**	0.62	0.59
VNSR[2]	0.70	**0.69**	0.65	0.59	0.68
VQM[3]	0.72	0.64	0.65	**0.78**	0.70
MOVIE	**0.79**	**0.67**	**0.72**	**0.75**	**0.75**

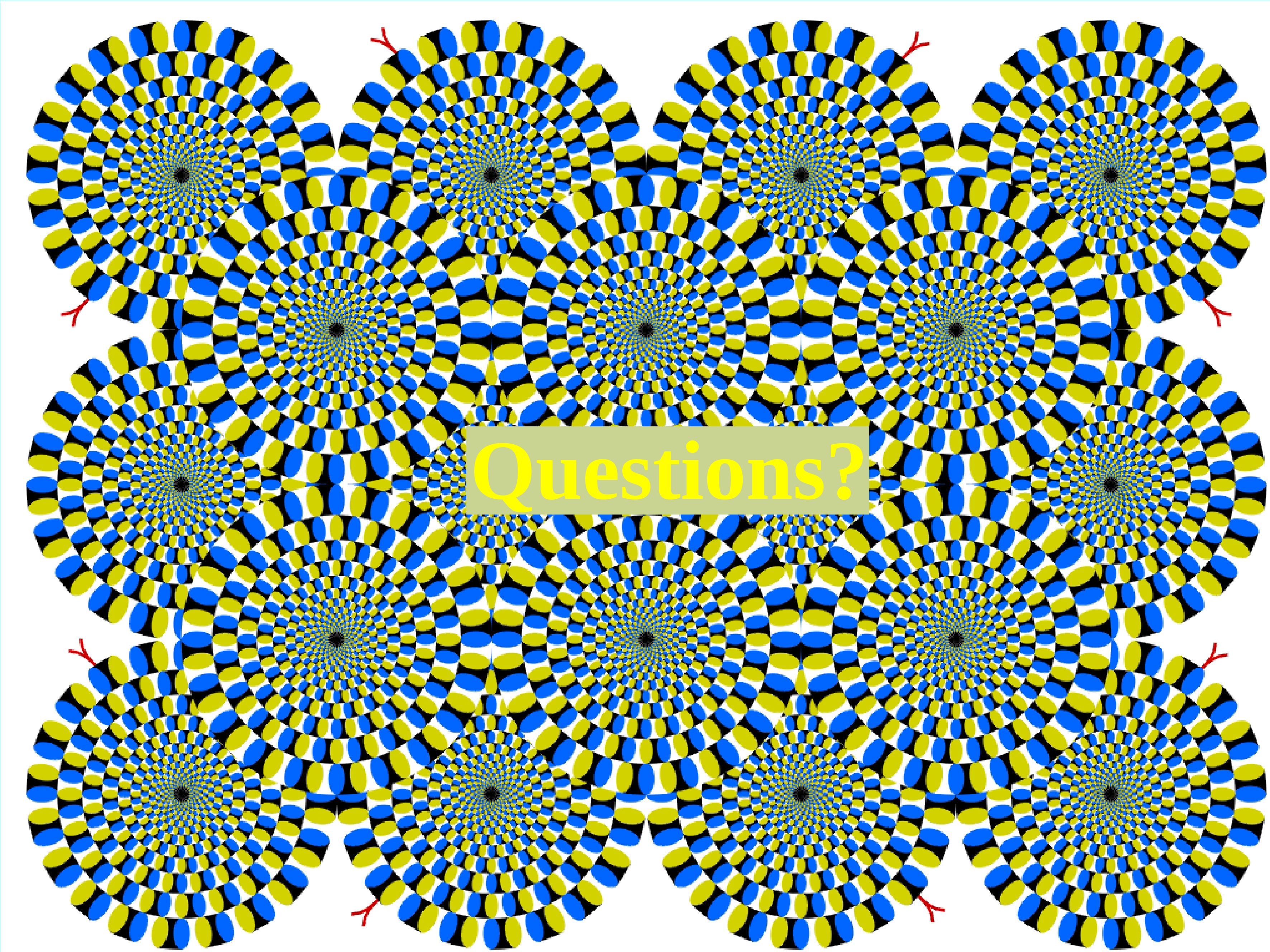
[1]Wang & Li, *J. Opt Soc. Amer.*, 2007.
[2]VSNR = "Visual Signal-to-Noise Ratio" – Chandler & Hemami, *IEEE Trans Image Process.*, 2007
[3]VQM = "Video Quality Metric" – Pinson & Wolf, *IEEE Trans Broadcasting*, 2004. Currently the ANSI/ISO standard.

The Future

- **Algorithms: Blind quality assessment** for applications as well as the general blind problem.

- **Datasets:** LIVE VQA Database expanding – more distortion types

- **Human Data:** Use the **continuous-scale human data** we collected for improve VQA algorithm development & algorithm analysis.


Questions?

LIVE's VQA Sponsors

AT&T

Texas Instruments

NSF

National Instruments

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

Video Clarity
Tools for Video Analysis

BOEING