An introduction to

OBJECTIVE ASSESSMENT OF IMAGE QUALITY

Harrison H. Barrett
University of Arizona
Tucson, AZ

Outline

- **Approaches to image quality**
- **Why not fidelity?**
- **Basic premises of the task-based approach**
- **Classification tasks**
  - **Figures of merit**
  - **Observers**
  - **Computational issues**
  - **Examples from current research**
- **Estimation tasks**
  - **Figures of merit**
  - **Estimation methods**
- **Summary, challenges for the future**

Approaches from the literature

Preference

Potter Stewart’s method
Structured preference (panels, check lists)

Fidelity measures

The best image is the one that looks the most like the object

(By what measure?)

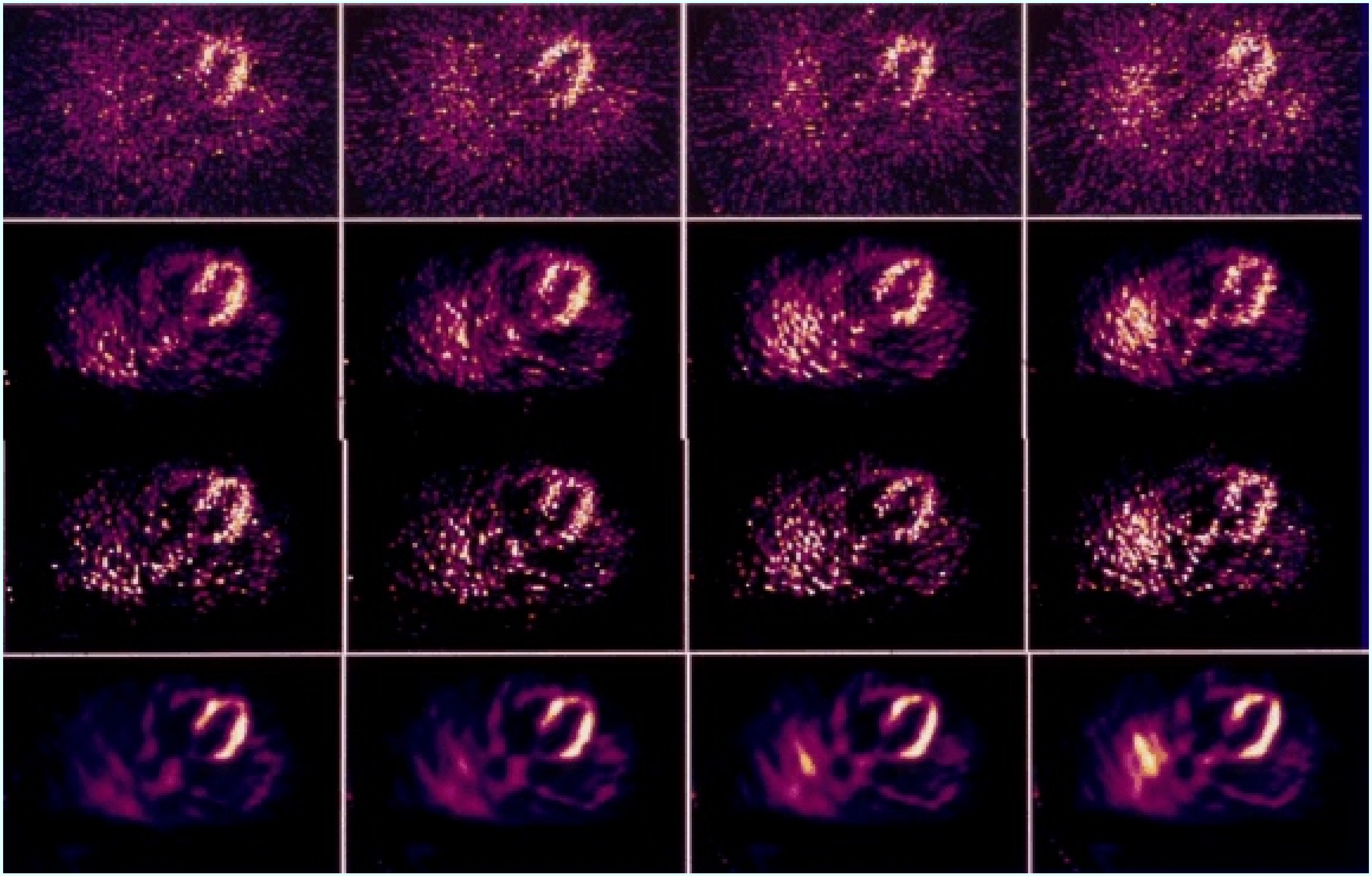
Perceptual difference measures

Information content

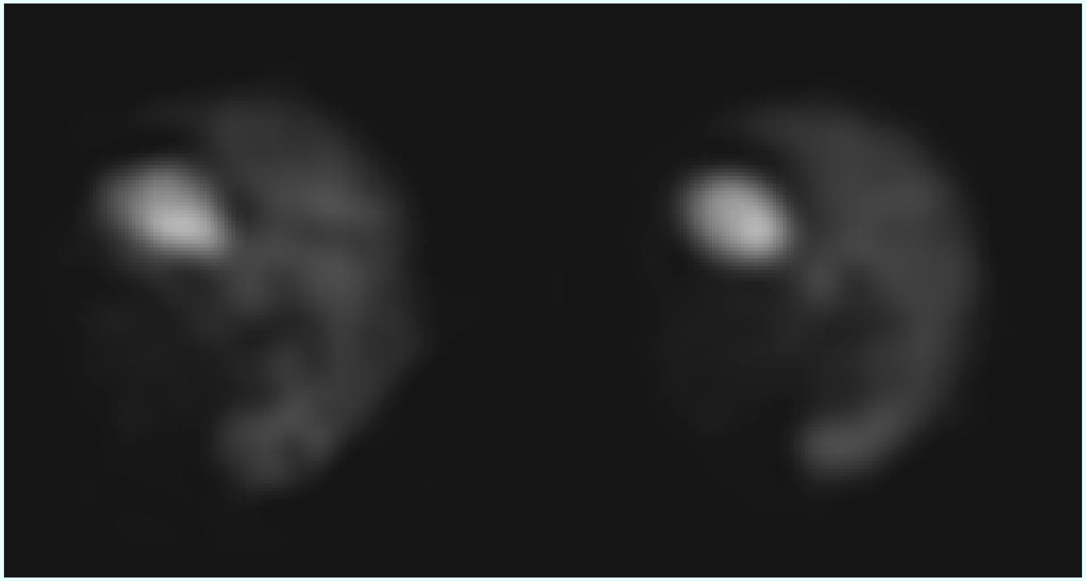
The best image is the one that gives the most information

(Information about what? To whom?)

It’s obvious ....



.... isn’t it?



(Hint: The good one is always on the right!)

Is fidelity a virtue?

If we think the best image is the one that looks most like the object, then ...

...mean-square error is a good figure of merit

Problem:

Objects are functions

Digital images are discrete vectors

How do you compare them?

Three ways to make object and image commensurable:

- Discretize the object

- Interpolate the image to a function

- Ignore the problem and do a simulation

Three ways to define mean-square error:

- Spatial average for one image

$$\|\text{object} - \text{image}\|^2$$

- Average over noise realizations

$$\langle \|\text{object} - \text{image}\|^2 \rangle_{\text{noise}}$$

- Average over noise realizations and object class

$$\langle \langle \|\text{object} - \text{image}\|^2 \rangle_{\text{noise}} \rangle_{\text{object}}$$

3 ways of making object and image commensurable

$\times$

3 ways of defining squared error

=

9 definitions of MSE

AND THEY GIVE VERY DIFFERENT ANSWERS

OTHER PROBLEMS WITH FIDELITY METRICS

☹ Sensitive to small changes in scale or orientation

☹ Sensitive to trivial grey-level mappings

☹ *In*sensitive to tradeoffs between noise and blur

☹ Very different images can have same MSE

Five Lenas, one MSE!

Z. Wang, A. C. Bovik and L. Lu,
IEEE International Conference on Acoustics,
Speech, & Signal Processing, May 2002.

The actual object

Lena Soderberg

But the biggest problem with MSE is that ......

.... it is independent of what you want to do with the image

BASIC PREMISE OF OBJECTIVE ASSESSMENT

A definition of image quality must specify:

☞ The task

What information do you want from the image?

☞ The observer

How will you extract the information?

☞ Object and image statistics

What set of objects are you looking at?
What is the measurement noise?

TASKS

- CLASSIFICATION (outcome is a label for each image)

    Signal detection
    Discrimination
    Pattern recognition

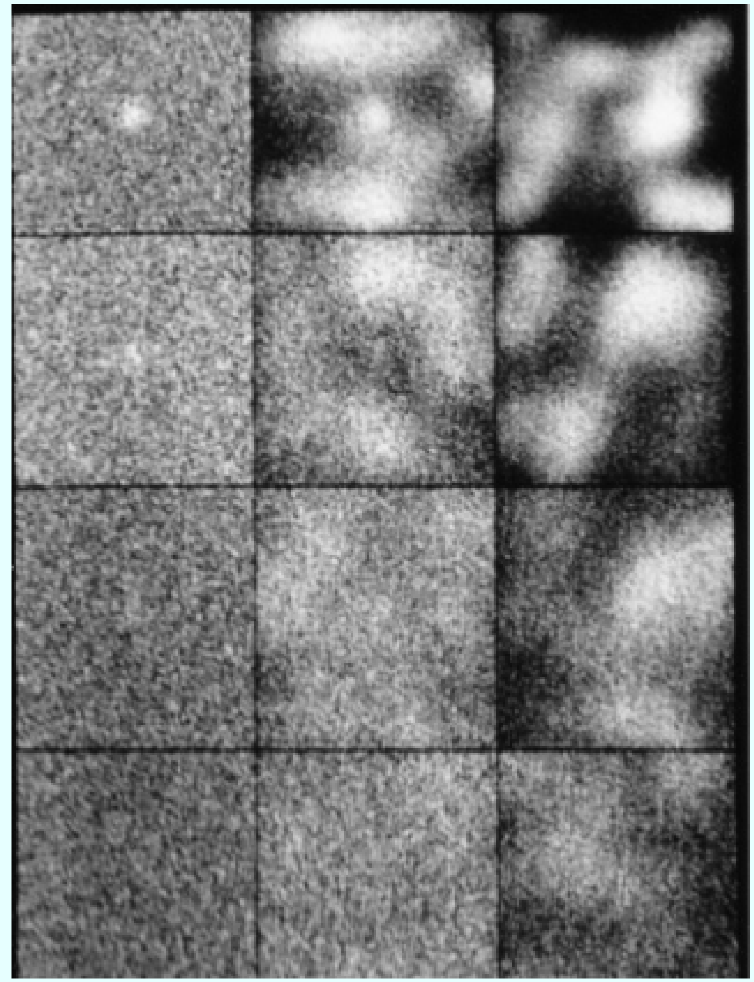
- ESTIMATION (outcome is one or more numbers)

    Absolute brightness, changes in brightness
    Integral over a region of interest
    Mensuration
    Localization
    Image reconstruction?

IMPORTANCE OF OBJECT AND IMAGE STATISTICS IN SIGNAL DETECTION

Signal detection performance degrades with more Poisson noise (top to bottom) or more structure in object background (left to right).

Work of Jannick Rolland



OBSERVERS FOR CLASSIFICATION TASKS

- HUMANS
- IDEAL (BAYESIAN) OBSERVER
- IDEAL LINEAR OBSERVER
  (HOTELLING OBSERVER)
- MODEL OBSERVERS THAT PREDICT HUMAN
  PERFORMANCE (ANTHROPOMORPHIC)

Binary (two-alternative) classification tasks
Example: signal detection

	Observer responds: "Yes"	Observer responds: "No"
Signal present	True positive	False negative
Signal absent	False positive	True negative

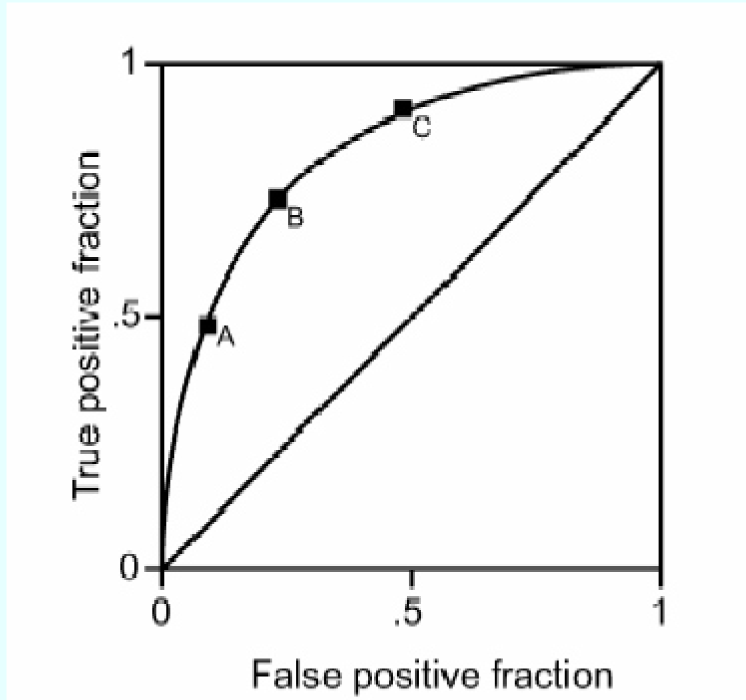
True-positive fraction (TPF) = probability of detection
False-positive fraction (FPF) = false-alarm rate

Key points:

There is always some decision threshold for detection
∴ there is always a tradeoff between TPF and FPF

The receiver operating characteristic (ROC)

Different points on the ROC curve are generated by applying different decision thresholds to some test statistic

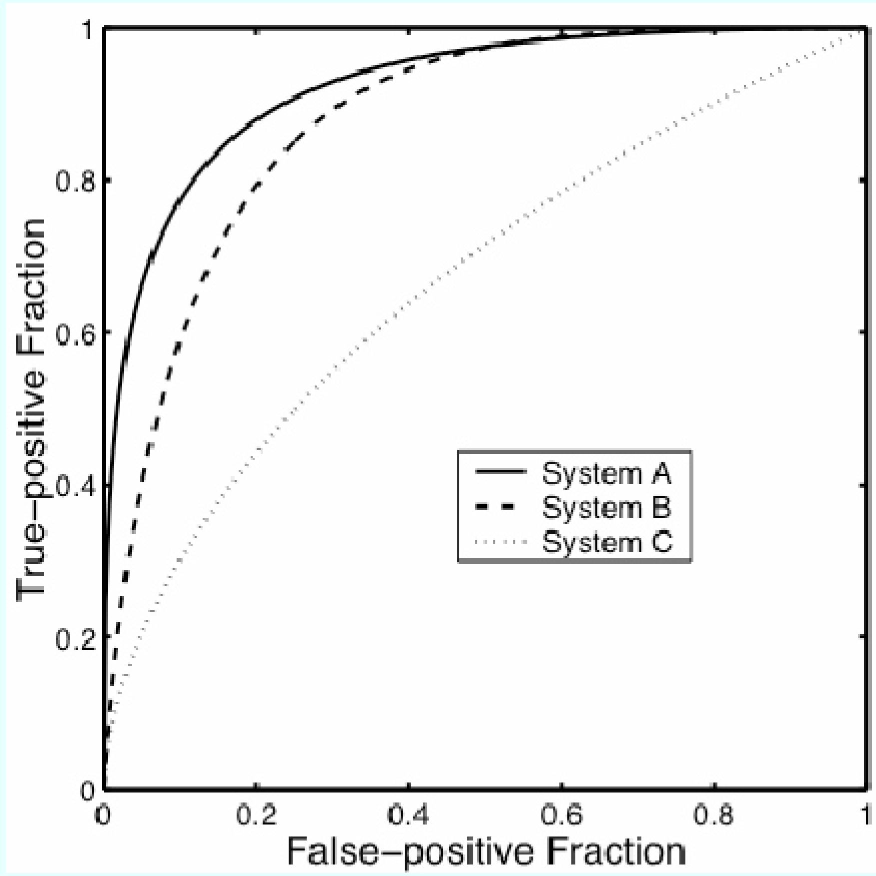


Ideal system (for a binary detection task):
True-positive fraction = 1
False-positive fraction = 0
Area under the curve (AUC) = 1

Worthless system: AUC = 0.5

∴ **AUC is a figure of merit for detection**

USING AUC TO COMPARE SYSTEMS



System A is better than B, and B is better than C
(for this particular detection task and observer)

HUMAN OBSERVERS

- Performance on binary classification tasks can be measured by psychophysical experiments
- Results can be plotted as ROC curves
- Area under the ROC curve can be used as FOM

PROBLEMS WITH HUMAN OBSERVERS

☹ Psychophysical studies are time-consuming and difficult to control

☹ Many observers and images are required for good statistical accuracy

☹ Tests of statistical significance are controversial

ALTERNATIVE: Mathematical model observers

Ideal
Hotelling
Anthropomorphic

WHAT IS THE IDEAL OBSERVER?

One that:

- Maximizes area under the ROC curve
- Maximizes true-positive fraction at any specified false-positive fraction
- Minimizes Bayes risk

Neat result: By any of these criteria, the test statistic is always the likelihood ratio

THE LIKELIHOOD RATIO

$$\Lambda(\mathrm{g}) = \frac{\mathrm{pr}(\mathrm{g}|\mathrm{H}_1)}{\mathrm{pr}(\mathrm{g}|\mathrm{H}_0)}$$

H_0 = null hypothesis (e.g., signal absent)
H_1 = alternative hypothesis (e.g., signal present)

Points on ROC curve for ideal observer are generated by comparing $\Lambda(\mathbf{g})$ to different thresholds

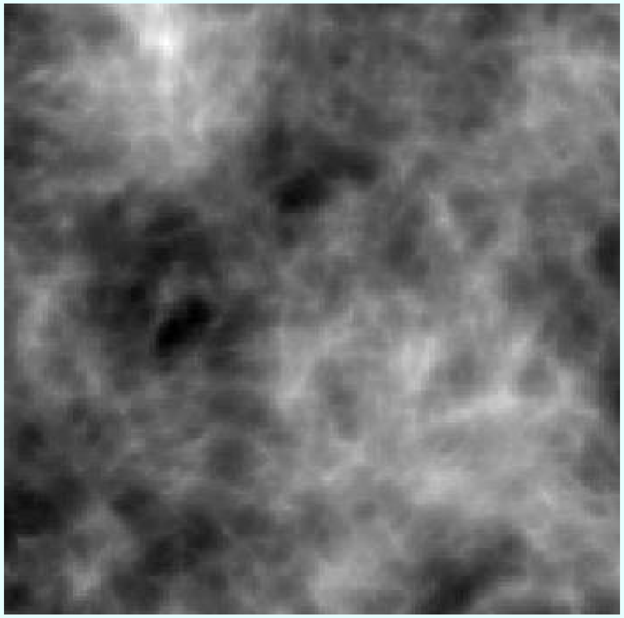
DIFFICULTIES WITH THE IDEAL OBSERVER

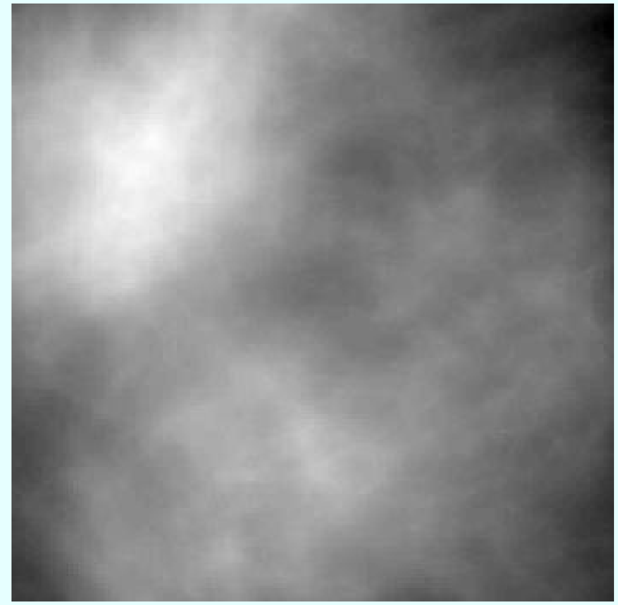
- ✏ Requires immense amount of prior knowledge (huge-dimensional PDFs)
- ✏ Usually a nonlinear function of the image data
- ✏ Difficult to compute except for a few textbook cases
- ✏ Not obviously related to humans or other practical observers

NEW COMPUTATIONAL TOOLS FOR IDEAL OBSERVER

- Likelihood-generating function
  - All statistical properties of LR in one scalar function!
- Realistic but tractable object models
  - Analytic forms for object statistics
- Characteristic functionals
  - Propagate object statistics through noisy, low-resolution imaging systems
- Markov-chain Monte Carlo
- Dimensionality reduction with minimal information loss
  - Channelized ideal observer

STATISTICAL MODELS FOR RANDOM OBJECTS





These simulations look very much like real medical images, *but their complete statistics are known analytically to all orders*!

Question: Can we do the same for astronomy?

THE HOTELLING OBSERVER

Harold
Hotelling

- Has full knowledge of the mean vectors and covariance matrices
- Uses that knowledge to compute an optimum linear discriminant

$$t(\mathbf{g}) = \mathbf{w}^t\mathbf{g}$$

- Maximizes a detectability measure called the Hotelling trace
  or Hotelling SNR
- Also maximizes AUC if $t(\mathbf{g})$ is normally distributed

COMPUTATIONAL ISSUES

☺ Good news – Hotelling doesn’t need image PDFs

☹ Bad news – It does need image covariance matrix

☹☹☹ Worse news – You gotta invert it!

Hotelling observer, signal-known-exactly (SKE) task

Basic equations boil down to:

$$t(\mathbf{g}) = \mathbf{w}^t\,\mathbf{g}$$

$$\mathbf{w} = \mathbf{K}^{-1}\mathbf{s}$$

$$\mathrm{SNR}^2 = \mathbf{s}^t\,\mathbf{K}^{-1}\mathbf{s}$$

$\mathbf{s}$ = signal vector

$\mathbf{K}$ = covariance matrix

Basic problem at a glance:

$$\mathrm{SNR}^2 = \mathbf{s}^t \, \mathbf{K}^{-1} \, \mathbf{s}$$

Example from tomography:

128 × 128 detector, 64 projection angles

s has 10^6 elements

K is a $10^6 \times 10^6$ matrix

$\therefore$ **K** has 10^{12} elements!

MEGALOPINAKOPHOBIA

Megalopinakophobia (mĕg´ ə-lō pēn´ ə-kō fō´ bē-ə) *n*.

1. Fear of large matrices.

2. Abnormality characterized by excessive fondness for Fourier methods.

3. Fear of large paintings

[< Gk. *megas,* great + *pinakas,* matrix or painting + *phobos*, fear]

(Courtesy of Georgios Kastis)

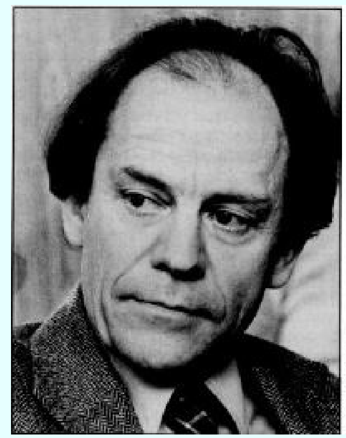
WHAT TO DO INSTEAD OF HUGE MATRIX INVERSIONS:

- Use decomposition of the covariance: $\mathbf{K} = \mathbf{K}_{noise} + \mathbf{K}_{obj}$
- Use simulated noise-free images to estimate $\mathbf{K}_{obj}$
- Solve for Hotelling template iteratively
  Fiete et al., J. Opt. Soc. Am., 1987
- Use matrix-inversion lemma
  - Reduces $N \times N$ problem to $J \times J$, where
    N = number of pixels, J = number of image samples
- Use Neumann series for $\mathbf{K}^{-1}$
- Reduce dimensionality by judicious feature extraction
  Barrett et al., Proc. SPIE 3340, 1998

ANTHROPOMORPHIC MODEL OBSERVERS
(MODELS THAT PREDICT HUMAN PERFORMANCE)

- USUALLY LINEAR OBSERVERS
- INCORPORATE SPATIAL-FREQUENCY CHANNELS

David Hubel



Torsten Wiesel

- REQUIRE “INTERNAL NOISE” TO ACCOUNT FOR HUMAN UNCERTAINTIES
- OTHERWISE IDEAL!

Why do we use anthropomorphic models?

- Basic perceptual studies

  Compare psychophysical studies with model, gain insights into human observers

- Objective assessment of imaging hardware

  Computationally simpler than ideal or Hotelling

  Faster and more accurate than psychophysical

- Assessment of processing algorithms and displays

  Ideal or Hotelling observer *gives no information* about algorithms

  Purpose of algorithm or display is to match the raw data to the characteristics of the human observer

VALIDATION OF HUMAN OBSERVER MODELS

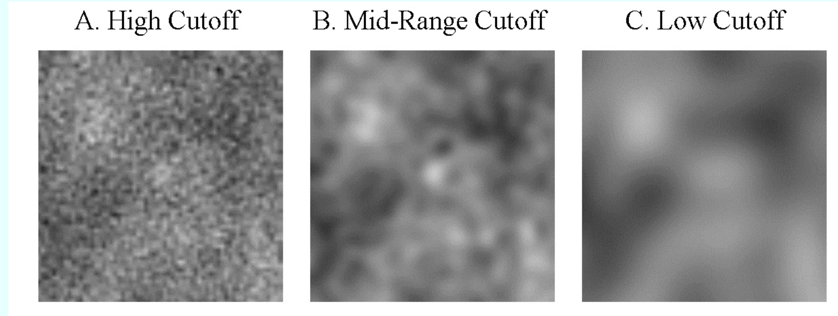
- Artificial backgrounds
- Phantoms
- Real clinical images

Procedure:

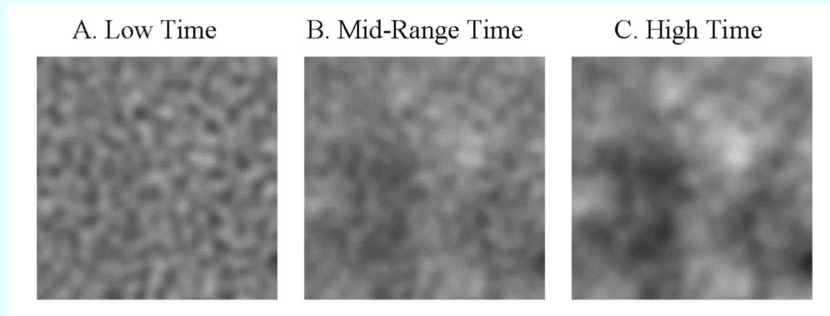
Generate many images
Estimate low-dimensional covariance on channel outputs
Compute Hotelling SNR
Compare to psychophysical studies

Abbey et al., J. Opt. Soc. Am. A, 2001

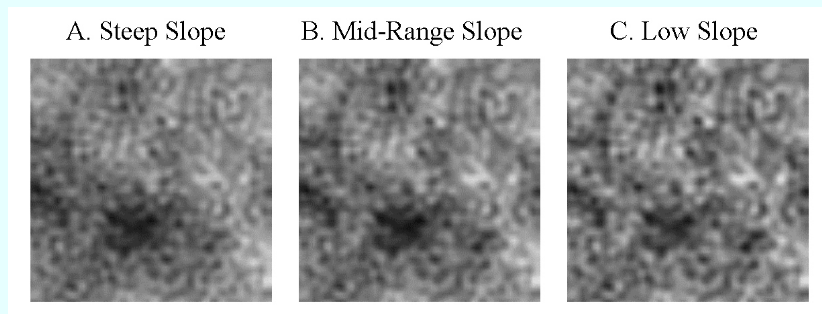
Regularization study
(Variation of filter cutoff)



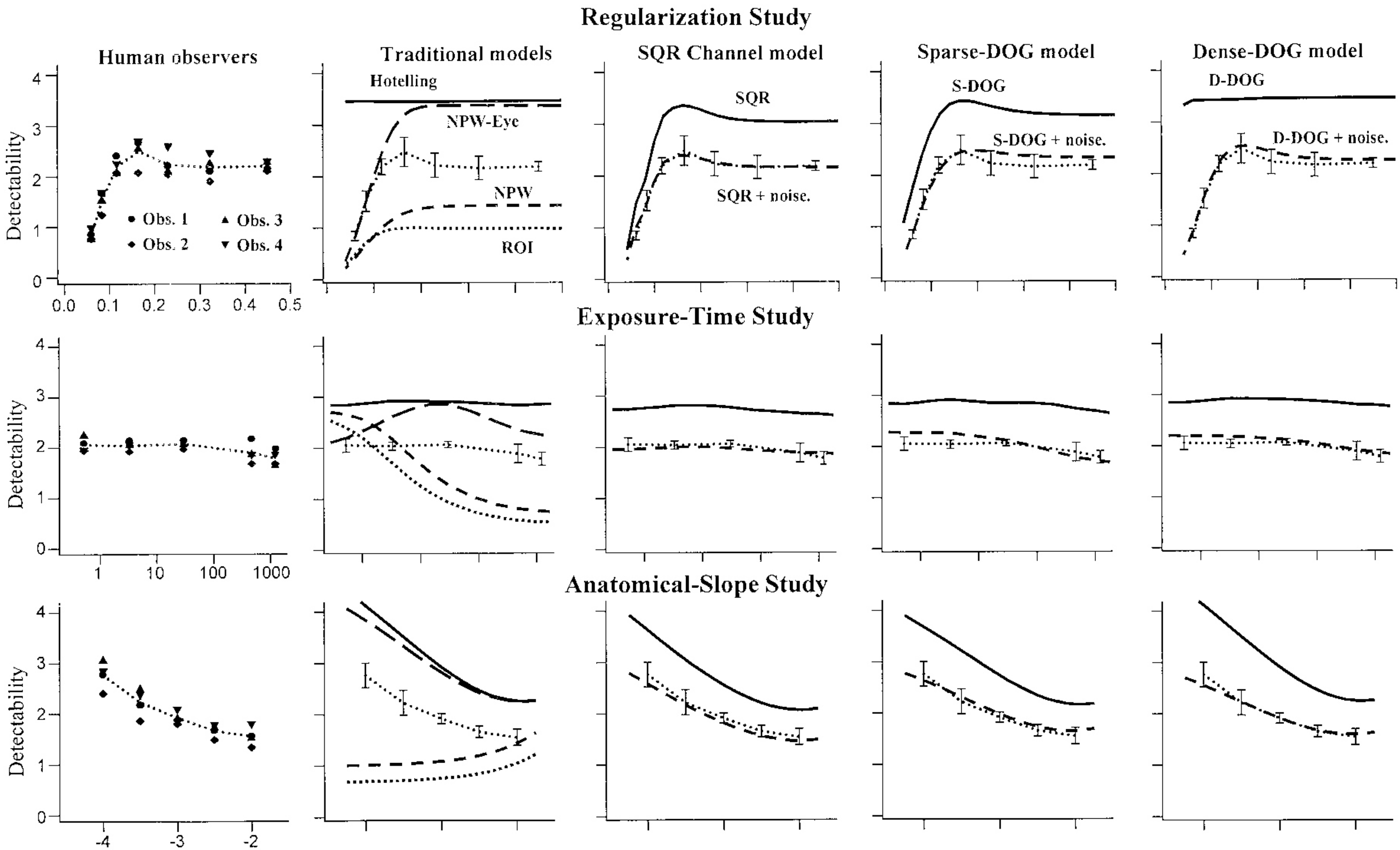
Exposure-time study
(Trades off Poisson
and anatomical noise)



Anatomical-slope study
(Varies fractal dimension
of anatomical background)



Model observer: Channelized Hotelling Observer (CHO) with and without internal noise (Myers et al., JOSA A, 1987)



Regularization Study
Human observers
Traditional models
SQR Channel model
Sparse-DOG model
Dense-DOG model
Hotelling
NPW-Eye
NPW
ROI
SQR
SQR + noise.
S-DOG
S-DOG + noise.
D-DOG
D-DOG + noise.
Obs. 1
Obs. 2
Obs. 3
Obs. 4
Detectability
0.0
0.1
0.2
0.3
0.4
0.5
Exposure-Time Study
1
10
100
1000
Anatomical-Slope Study
-4
-3
-2

ESTIMATION TASKS

$\mathrm{f}(\mathbf{r})$ = object distribution

$\Theta\{\mathrm{f}(\mathbf{r})\}$ = parameter of interest

$\mathbf{g}$ = data vector

$$\mathbf{g} = \mathcal{H}\{\mathrm{f}(\mathbf{r})\} + \mathbf{n}$$

$\mathrm{f}(\mathbf{r}) = \mathrm{f}_{\mathrm{meas}}(\mathbf{r}) + \mathrm{f}_{\mathrm{null}}(\mathbf{r})$ where $\mathcal{H}\,\mathrm{f}_{\mathrm{null}}(\mathbf{r}) = 0$

$\Theta\{\mathrm{f}(\mathbf{r})\}$ is *estimable* if and only if $\Theta\{\mathrm{f}(\mathbf{r})\} = \Theta\{\mathrm{f}_{\mathrm{meas}}(\mathbf{r})\}$

For estimable parameters, can define bias, variance, MSE

Pixel values are almost never estimable

MAXIMUM-LIKELIHOOD ESTIMATES

For an estimable parameter, can define bias, variance, MSE

Cramer-Rao bound defines minimum achievable variance

Estimator that achieves this bound is *efficient*

ML estimate is:

- Asymptotically unbiased
- Asymptotically efficient
- Efficient if an efficient estimator exists

But ML estimate uses no prior information

FIGURES OF MERIT FOR ESTIMATION

MSE of ML estimator

Cramer-Rao bound

Fisher information

Bayes risk

CONCLUSIONS

- ✯ Logical definitions of image quality must specify the task and observer as well as object and image statistics
- ✯ Anthropomorphic or real human observers should be used to evaluate algorithms and displays
- ✯ Ideal or ideal-linear observers should be used to evaluate imaging hardware on detection tasks
- ✯ Estimation task performance meaningful only for estimable params.
- ✯ Computation of observer performance is challenging, but methodology is developing rapidly
- ✯ Realism in system and object modeling is paramount
- ✯ The ultimate goal is full system optimization