Applications of Task-Based Image Quality Assessment in Nuclear Medicine Imaging

Eric C. Frey, Ph.D.
Division of Medical Imaging Physics
Russell H. Morgan Department of Radiology and Radiological Science
Johns Hopkins University

Acknowledgements

- Organizers for inviting me
- Xin He, etc., at Johns Hopkins
- Mike King, Howard Gifford, etc., at U. Mass. Med.
- Jinyi Qi at U.C. Davis
- Paul Kinahan and coworkers at U. Wash.
- Gene Gindi and coworkers at Stony Brook

Scope of Talk

- Applications discussed limited to
  - SPECT and PET
  - Anthropomorphic Model Observers
    - No studies using purely human observers
    - No studies using ideal observers
  - Classification tasks
    - Detection
    - Detection + localization
    - Detection + characterization

Differences Between Slides and Handouts

- Fixed some typos
- Added some additional references
- The new slides will be available for download from:
  - http://mirlbcc1.jhoc1.jhmi.edu/public

Goals

- Describe choices that must be made to perform a task-based assessment
- Be as practical, detailed, and concrete as possible
- Give examples of applications of task-based assessment in range of nuclear medicine imaging applications
- Focus on methods rather than results

Outline

- Components of task-based assessment
  - Task models
  - Object model
  - Imaging system model
  - Observer model
  - Figure of merit (FOM)
- Examples of task-based assessment
  - Myocardial perfusion SPECT
  - Hepatic SPECT
  - Ga-67 SPECT
  - 2D vs. 3D PET tumor imaging
  - Optimization of MAP reconstruction for PET tumor imaging
- Summary

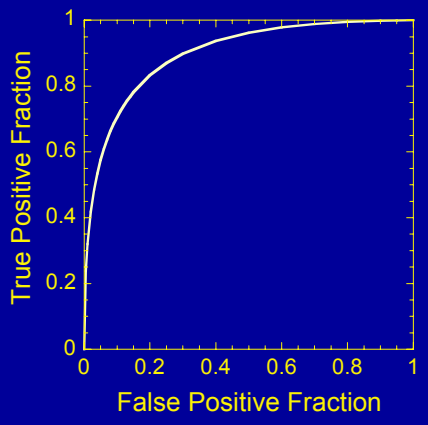
Task Models

- Detection
- Detection + Localization
- Detection + Characterization

Detection Task

- Observer provides
  - Rating of confidence that signal is present anywhere in image

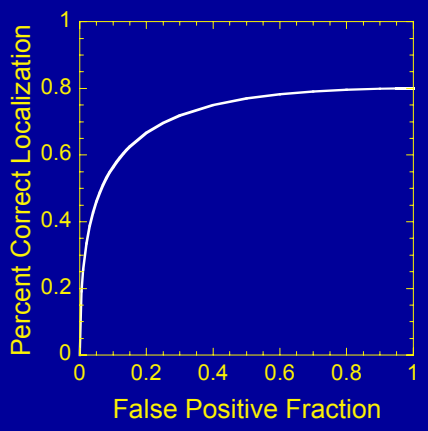
  or
  - Rating of confidence that signal is present at a particular point in image (SPKE)
- Does not realistically model all clinical tasks
- Analyze using ROC analysis
- Figure of merit
  - Area under ROC curve (AUC)
  - Range: [0.5,1.0]



Metz C. E. 1978 Seminars in Nuclear Medicine

Detection+Localization Task

- Observer provides
  - Location of suspected signal
  - Rating of confidence that the signal is at this point
- Task may be more clinically realistic for some applications
- Analyze using LROC analysis
- Figure of Merit:
  - Area under LROC Curve (A_{LROC})
  - Range: [0,1]

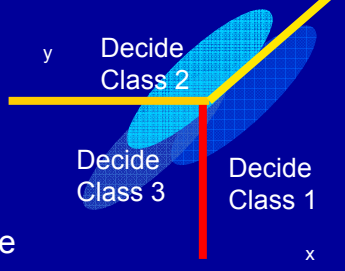


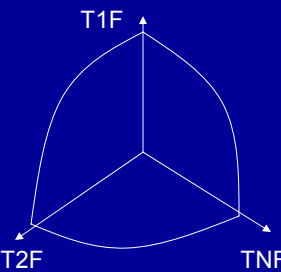
Swensson RG, Med Phys, 1996

Khurd and Gindi, IEEE Trans Med Imag 2005

Detection + Characterization

- Observer provides
  - Confidence rating that signal is present and class 1
  - Confidence rating that signal is present and class 2
- May be needed to model some clinical tasks
- Analyze using 3-class ROC analysis
- Figure of Merit
  - Volume under ROC surface (VUS)
  - Range [1/6, 1]





He, Metz, Links, Tsui, and Frey, IEEE Trans Med Imag, May, 2006

Outline

- Components of task-based assessment
  - Task models
  - **Object model**
  - Imaging system model
  - Observer model
  - Figure of merit (FOM)
- Examples of task-based assessment
  - Myocardial perfusion SPECT
  - Hepatic SPECT
  - Ga-67 SPECT
  - 2D vs. 3D PET tumor imaging
  - Optimization of MAP reconstruction for PET tumor imaging
- Summary

Object Model

- Background
  - Background known exactly (BKE)
  - Background known statistically (BKS)
- Signal
  - Signal known exactly (SKE)
  - Signal known statistically (SKS)

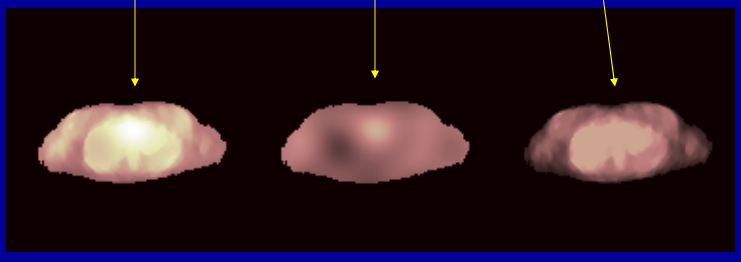
Background Models

- Background variability is essential
- Background variation should be as realistic as possible
- Ways to achieve this:
  - Statistically defined background
  - Single phantom, multiple signal locations
  - Collection (population of phantoms)

Statistically Defined Background

Lumpy Background

$$B = b\sum_{i=1}^{K} G(r_k, \sigma^2) + B_0$$



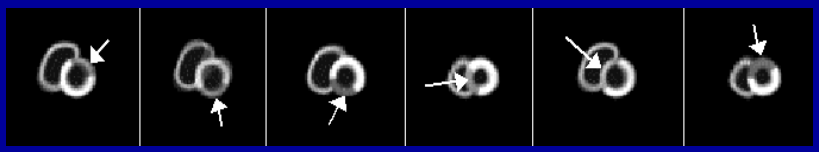
- Gaussian blobs with σ^2 =50 (pixel^2) at random locations r_k.
- Number of blobs K is a Poisson r.v. with mean = 100.
- b = 150.

J. P. Rolland and H. H. Barrett, *J Opt Soc of Amer A*, 1992.
Qi, IEEE Trans Nucl Sci, 2006

Courtesy of J.Y. Qi

Single Phantom
Multiple Signal Locations

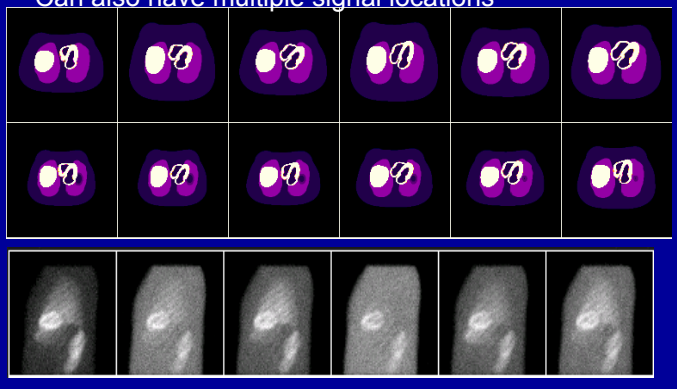
- Single heart and background
- Lesions at 6 locations



Frey et al. IEEE Trans Med Imaging, 2002

Multiple Phantoms

- NCAT Phantom
- 12 male, 12 female anatomies (females shown)
- Varied body, lung, liver and heart size, heart orientation
- Multiple randomly sampled organ uptakes for each anatomy
- Can also have multiple signal locations



He, Frey et al. IEEE Trans Nucl Sci 2004

Comparison of Background Definition Methods

- Statistically defined backgrounds
  - Easy to treat analytically
  - May not be sufficiently realistic for some applications
- Single phantom
  - Relatively simple to define object
  - Relatively easy to obtain projection data
  - Limited object variability
- Multiple phantoms
  - Substantial effort to define phantoms
  - Substantial effort to obtain projection images
  - Maximum realism

Types of Background

- Physical phantom
  - Accurate simulation of imaging system
  - Limited ability to model object variability
- Mathematical phantom (Frey, 2002; He, Frey, 2004)
  - May not model imaging system as well
  - Allows object variability limited only by computational resources
- Patient data (Gifford, TNS,2001)
  - Most realistic model of object and imaging system
  - Object variability limited only by availability of patient data
  - Limited control over object parameters

Signal Models

- Signal Known Exactly (SKE)
  - Most model observers are derived for and optimal for this case (Barrett, 1993)
  - My not be useful for optimizing images for human use (Yendiki 2006)
- Variable size and contrast (He, Frey, 2004)
  - More realistic
  - Have been used successfully with model observers
- Signal shape depends on application
  - Myocardial perfusion defect limited to myocardium
  - Tumors approximated by sphere or 3D Gaussian

Outline

- Components of task-based assessment
  - Task models
  - Object model
  - **Imaging system model**
  - Observer model
  - Figure of merit (FOM)
- Examples of task-based assessment
  - Myocardial perfusion SPECT
  - Hepatic SPECT
  - Ga-67 SPECT
  - 2D vs. 3D PET tumor imaging
  - Optimization of MAP reconstruction for PET tumor imaging
- Summary

Modeling the Imaging System

- Must realistically model imaging system
  - To test attenuation compensation, model attenuation accurately and include approximate scatter model
  - To test approximate scatter compensation, use more accurate (e.g., Monte Carlo) scatter simulation
- Should model continuous nature of activity distribution
  - E.g., simulate using pixelized phantoms with pixel size < ½ the size of projection bins

Components of Task Based Assessment Studies

- Task models
- Object model
- Imaging system model
- Observer model
- Figure of merit (FOM)

Imaging System Models

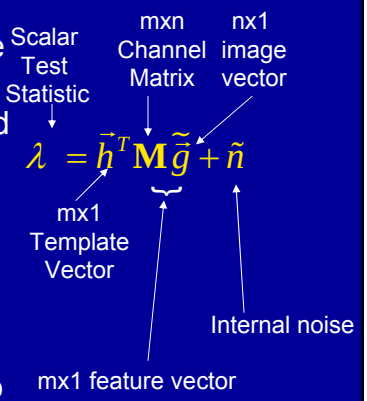
- Analytic simulation
  - Very computationally efficient
  - Hard to accurately model some physical processes (e.g., scatter in object and detector)
- Monte Carlo simulation
  - Can achieve high level of realism
  - Allows testing instrument designs without building them
  - Very computationally intensive
  - May need to combine analytic models for some parts of system with improve efficiency (e.g. X. Song, PMB, 2005)
- Physical Measurement (e.g. Chen, 2nd Biannual Workshop on Small-Animal SPECT Imaging, 2006)
  - Ultimate in accuracy
  - Hard to acquire large number of realizations
  - Requires access to real imaging system
  - Limited to physical phantoms and patients
- Hybrid (e.g. Gifford, TNS,2001)
  - Physical measurement for background
  - Monte Carlo or analytic simulation of lesion

Outline

- Components of task-based assessment
  - Task models
  - Object model
  - Imaging system model
  - **Observer model**
  - Figure of merit (FOM)
- Examples of task-based assessment
  - Myocardial perfusion SPECT
  - Hepatic SPECT
  - Ga-67 SPECT
  - 2D vs. 3D PET tumor imaging
  - Optimization of MAP reconstruction for PET tumor imaging
- Summary

Model (Mathematical) Observers

- Most anthropomorphic observers are linear
- Dot product of observer template and image plus internal noise gives test statistic (analogous to rating value)
- Internal noise models variability of human observers
- Use channel model to better model human observers
- Addition of non-linear components to Hotelling observer (Zhang, Pham and Eckstein, TMI, 2006)



Barrett, 1993

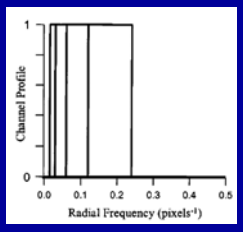
Abbey and Barrett, JOSA, 2001

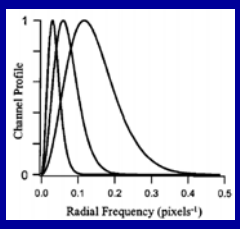
Commonly Used Model Observers

- Channelized Non-Prewhitening Observer (CNPWO)
  - Simpler to implement than CHO
  - Template is simply the signal
  - Good agreement in several studies
- Channelized Hotelling Observer
  - Models ability of human observers to adapt to noise
  - Used in a large number of Nuclear Medicine related studies
  - Generally good agreement with human observer study results

Channel Shape

- Constant Q (SQR) Channels
  (e.g., Frey, 2002)
  - Non-overlapping
  - Used in many studies
  - Generally good agreement
- Difference of Gaussians (DOG)
  (Gifford and King, 2000)
  - Channels overlap
  - Used in fewer studies
  - To date, no significant advantage in predicting human observer performance compared to SQR channels





Channel Properties

- Should have zero response at 0 frequency
- Start channel can be important parameter that may need to be optimized for good agreement with human observers
- Bandwidth for n-th channel typically proportional to 2^n
- Number of channels
  - Typically 3-5
  - More channels increases difficulty of estimating observer
  - Depends on image size
  - Chosen based on relative size of noise and and signal
- Shifted to suspected signal location

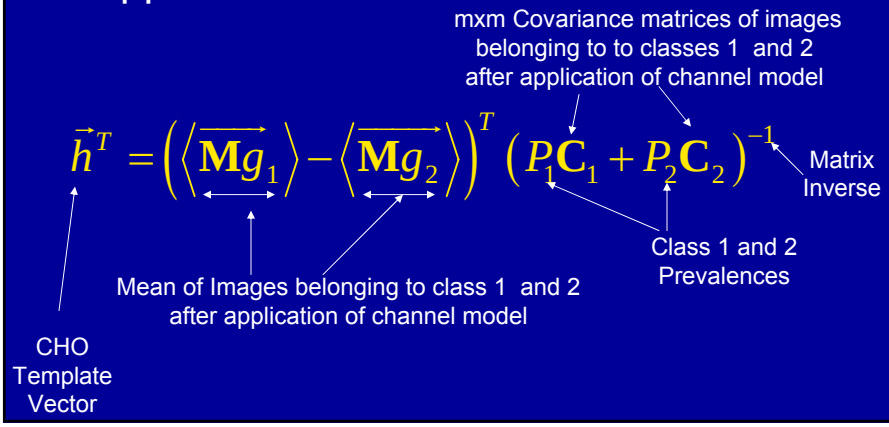
Observer Internal Noise

- Can improve absolute agreement between model observers and human observers
- Generally not needed when only rankings of methods are important

Abbey and Barrett, JOSA, 2001

Estimating the CHO Template

- CHO template is a function of the mean and covariances of the images after application of the channel model



Analytic Estimation of CHO Template

- See J.Y. Qi, 2004 and Khurd, 2005c
- Extremely fast (1000x faster than ensemble methods)
- Based on estimates of mean and covariance at convergence
- Are approximate (but good approximation)
- Does not work for all algorithms (e.g., OS-EM)
- Does not allow modeling of non-linear processes

Monte Carlo Estimation of the CHO Template

- Simple to implement
  - Use sampling techniques to generate many sample images
  - Compute requisite mean and covariance matrices
- Completely general
  - Applicable to any background model
  - Applicable to any image processing (even nonlinear)
- Requires large number of images
  - Minimum of ~4x number of elements in covariance matrix
  - Typically use several hundred samples
  - Large computation (weeks of CPU time) and storage demands (100s of GB of disk space)

Outline

- Components of task-based assessment
  - Task models
  - Object model
  - Imaging system model
  - Observer model
  - **Figure of merit (FOM)**
- Examples of task-based assessment
  - Myocardial perfusion SPECT
  - Hepatic SPECT
  - Ga-67 SPECT
  - 2D vs. 3D PET tumor imaging
  - Optimization of MAP reconstruction for PET tumor imaging
- Summary

Figures of Merit

- Detection Task
  - CHO SNR
  - Detectability (d_A)
  - Area under ROC curve (AUC)
- Detection+Localization
  - Area under LROC Curve (A_{LROC})
- Detection+Characterization
  - Volume under ROC Surface (VUS)

Relationships between Detection Task FOMs

- SNR is given by:
- If test statistics for two classes $\{\lambda_1\}$ and $\{\lambda_2\}$ are independent and Gaussian distributed, then d_A=SNR
- AUC and d_A are monotonically related

Means of test statistics from classes 1 and 2

$$SNR^2 = \frac{\left(\bar{\lambda}_1 - \bar{\lambda}_2\right)^2}{\left(\sigma_{\lambda_1}^2 + \sigma_{\lambda_2}^2\right)}$$

Variances of test statistics from classes 1 and 2

Analytic Estimation of FOMs

- Analytic expressions are available for SNR (J.Y. Qi 2004, 2006)
- Rapid estimation of A_{LROC} is available (Khurd 2005c)
- Expressions are approximate (but very good)
- Do not provide ROC or LROC curves
- Are related to AUC only if test statistics are independent and Gaussian distributed

Monte Carlo Estimation of FOMs

- Apply CHO template to ensemble of images to obtain ensemble of test statistics

Then

- Compute means and variances needed to compute SNR

or

- Analyze test statistics using codes (e.g., LABROC, CLABROC) for analyzing estimating the ROC or LROC Curve and AUC or A_{LROC}
- Take advantage of extensive statistical machinery developed for analyzing human observer data

Outline

- Components of task-based assessment
  - Task models
  - Object model
  - Imaging system model
  - Observer model
  - Figure of merit (FOM)
- Examples of task-based assessment
  - Myocardial perfusion SPECT
  - Hepatic SPECT
  - Ga-67 SPECT
  - 2D vs. 3D PET tumor imaging
  - Optimization of MAP reconstruction for PET tumor imaging
- Summary

Example 1
Myocardial Perfusion SPECT

- Frey 2002, Sankarin 2002, He 2004, Narayanan 2002
- Task
  - Myocardial defect detection
  - Signal position known exactly (ROC task)
- Questions
  - What is optimal number of iterations?
  - What is optimal combination of compensations?
  - What is optimal number of iterations?
  - What is optimal orbit (180 or 360°)?
  - Can acquisition time be reduced?

Object Models

- MCAT Phantom population
  - 24 anatomies
    - 8 male flat diaphragm
    - 8 male w/raised diaphragm
    - 8 female
  - Constant heart size, shape and orientation
  - Organ size and shape varied with phantom
  - Constant organ activity concentration ratios
  - 6 defect locations, each with constant size
  - 12% defect contrast
  - 1/8 clinical count level

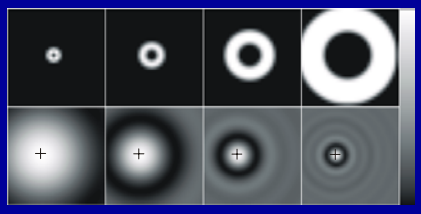
- NCAT Phantom population
  - 24 anatomies
    - 12 male
    - 12 female
  - Randomly varied heart size, shape, and orientation
  - Randomly varied organ sizes and shapes
  - Randomly varied organ activities concentration ratios
  - 6 defect locations, size varied by +/- 25%
  - Defect contrast varied from 10-35%
  - Clinical count level

Imaging System Model

- MCAT Phantom Study
  - Analytical simulation
  - Simulation voxel size was ½ of reconstructed voxel size
- NCAT Phantom Study
  - Monte Carlo simulation
  - Simulation voxel size was ½ of reconstructed voxel size

Observer Model

- CHO
- Applied to 32x32 short axis slices
- 4 SQR channels
  - Start freq and freq width of first channel=1/64 pixel^{-1}
- Ensemble estimation of template (training) using 864 images
- Internal noise not included



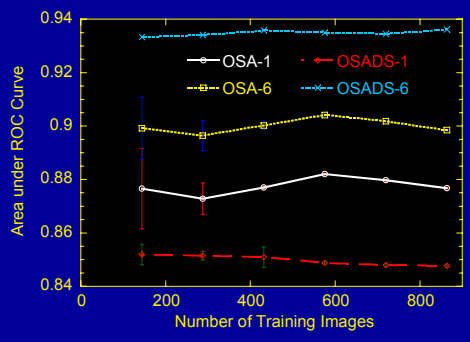
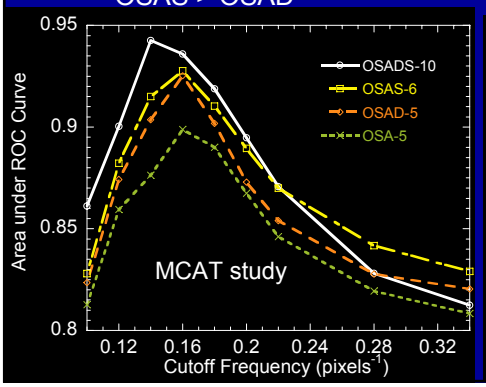


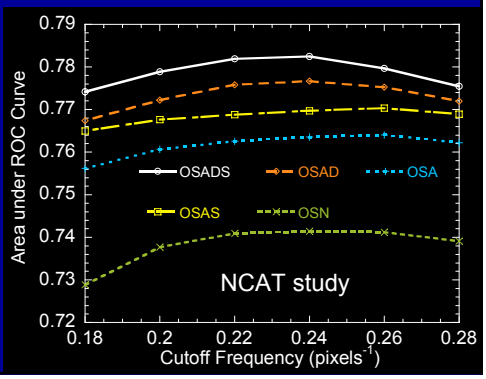
Figure of Merit

- Area under ROC Curve
- Estimated using CLABROC program from Metz et al
- Ensemble of testing images equal to # of training images (864)
- Tested for statistical significance of differences using CLABROC (takes into account correlations due to use of same data ensemble for all data treatments)

Results
Reconstruction Parameters

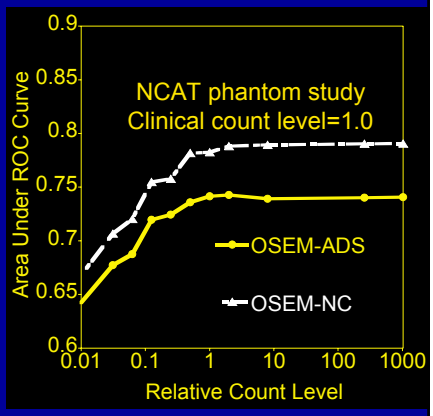
- For MCAT study:
  - AUC much larger (despite 1/8 count level and lower contrast)
  - Optimal cutoff frequency much lower and sharper
  - Optimal number of iterations was larger
  - OSAS > OSAD





Results
Effect of Count Level

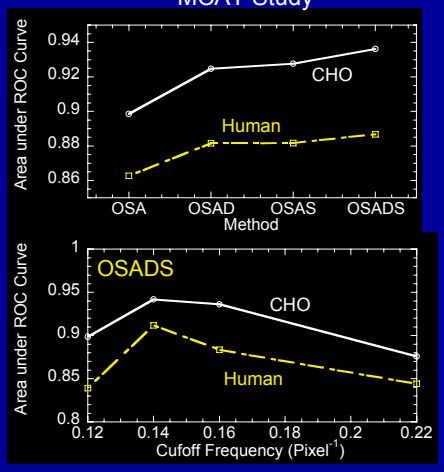
- Demonstrates tradeoff between anatomical noise and statistical noise
  - Low count level: performance decreases with increasing noise
  - High count level: performance relatively constant
  - Very high count level: performance not perfect due to anatomical noise



Including realistic levels of object variation is essential!

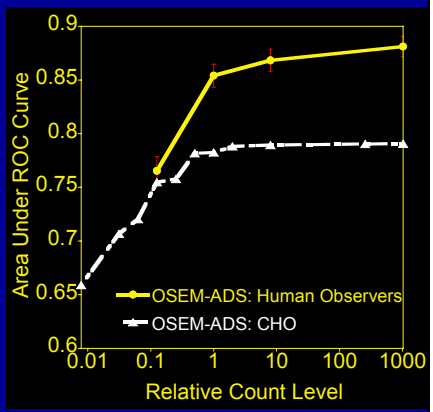
Results
Comparison w/ Human Observers

MCAT Study



Correlation coef for AUCs: 0.904
Correlation coef for rankings: 0.964
CHO performed better than human

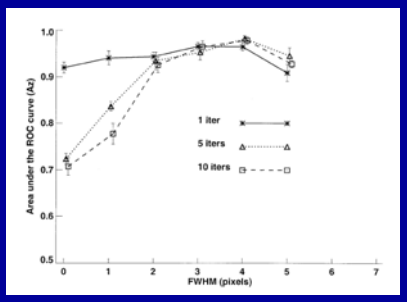
NCAT Study



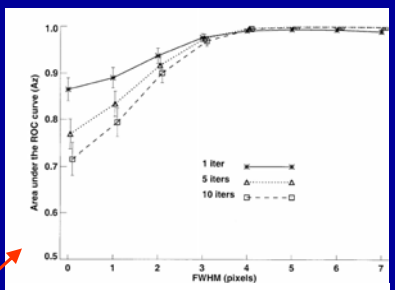
Correlation coef for AUCs: 0.986
CHO performed worse than human

Effect of Quantization

Human Observers

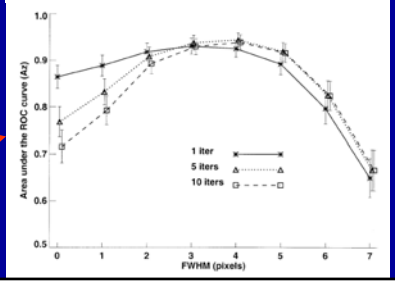


CHO



CHO w/o Quantization Noise Model
Correlation Coef=0.604

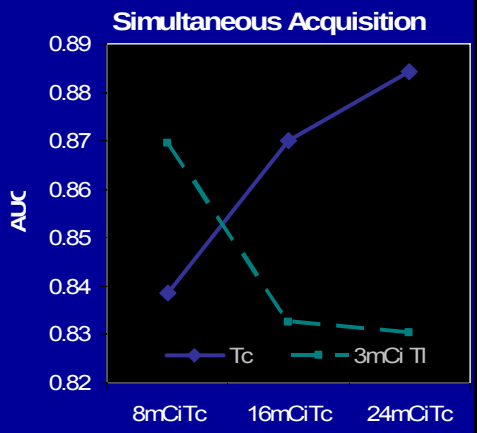
CHO w/ Quantization Noise Model
Correlation Coef=0.870



From Narayanan 2002

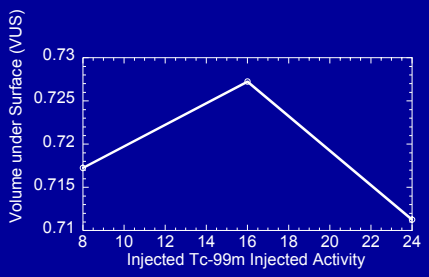
Simultaneous Dual Isotope Myocardial Perfusion SPECT

- Crosstalk:
  - Tc photons corrupt Tl image
  - Tl images corrupt Tc imag
- Optimizing relative injected activity (IA) cannot be done using simple detection task



Simultaneous Dual Isotope Myocardial Perfusion SPECT

- Must explicitly consider 3-class nature of task:
  - Use two images
  - To classify patient as
    - Normal
    - Reversible defect
    - Non-reversible defect



	Normal (Class 1)	Reversible (Class 2)	Non- reversible (Class 3)
Stress (^{99m}Tc)	N	D	D
Rest (^{201}Tl)	N	N	D

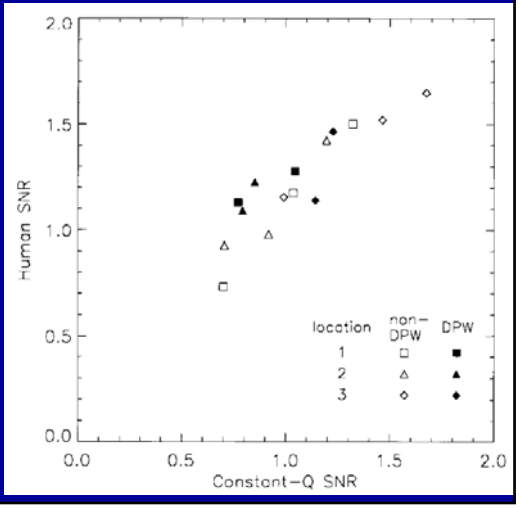
Example 2 Hepatic SPECT

- Gifford 2000
- Task: SKE, BKE (ROC)
- Object Model: Zubal phantom, 3 spherical tumors
- Imaging System Model: MC Simulation
- Observer Model
  - CHO
    - Constant Q w/4 channels
    - D.O.G. w/ 3 channels
  - Ensemble training
  - Explored several combinations of channel parameters
- FOM: CHO SNR

Results
Correlation with Human Observers

- Rank correlation
  - Constant Q: 0.92
  - D.O.G.: 0.84

Comparison of SNR for Human and CHO w/ Constant Q channel model



Example 3
Ga-67 SPECT Imaging

- Gifford 2005
- Task: Detection + Localization (LROC)
- Object
  - MCAT phantom
  - Randomly sampled lesions placed around lymph nodes
- Image Formation
  - Monte Carlo Simulation
  - 2D and 3D reconstruction/processing strategies

Observer Model

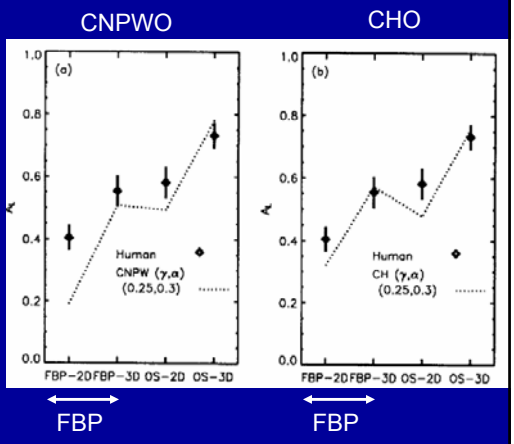
- Used observers with search model including region of uncertainty
- Single and multi-slice observers
- Internal noise appropriate for LROC task
- NPWO
- CNPWO
- CHO
- Constant Q channels with 3 different start frequencies
- Ensemble training (320 images) and testing (100 images)

Figure of Merit

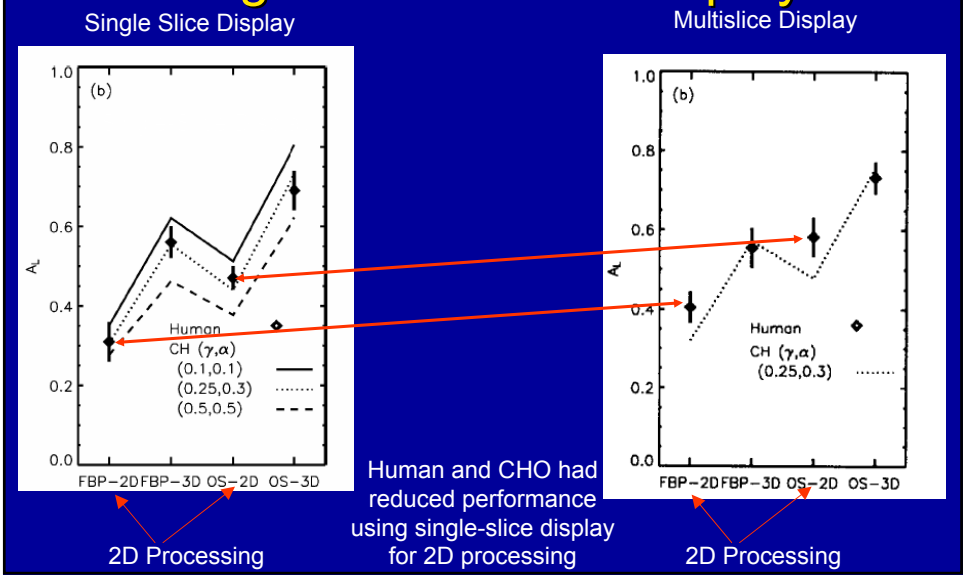
- Test statistics estimated from 100 test images
- A_{LROC} computed using Swenson's ML algorithm

Results
Multislice Display

- Both CNPWO and CHO observers predicted rankings of methods
- Good quantitative agreement with appropriate internal noise
- CHO provided better predictions for FBP reconstruction strategies

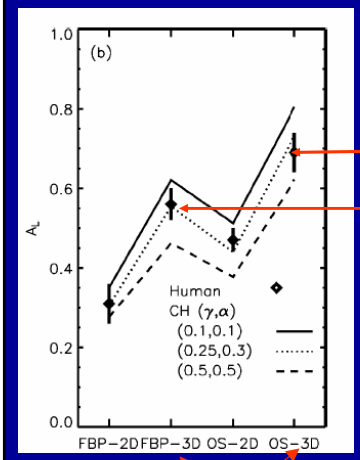


Results
Single vs Multislice Display

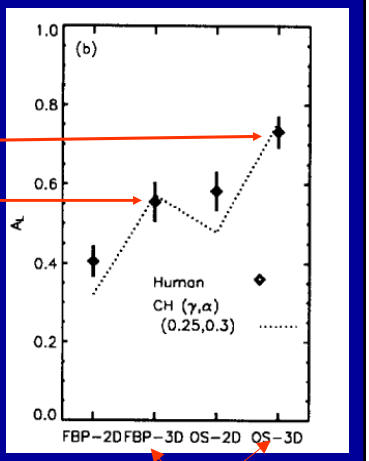


Results
Single vs Multislice Display

Single Slice Display



Multislice Display



3D Processing

Human and CHO had similar performance using single or multi-slice display for 3D processing

3D Processing

Conclusions

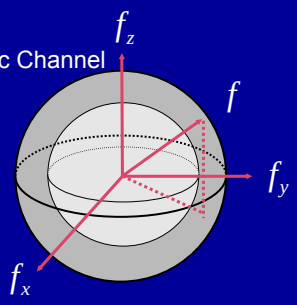
- Can model LROC task for both single and multislice observer
- For 3D processing, single-slice observer may be sufficient
- For some algorithms, CNPWO may predict human performance as well as CHO, and is easier to train, for some tasks

Example 4
2D vs 3D PET Tumor Imaging

- Kim 2004, Lartizen 2004
- Task: Detection (ROC)
- Object:
  - MCAT phantom
  - Multiple lesions per phantom
    - Several sizes and contrasts
- Image Formation
  - Analytic model
  - 2D and 3D acquisition for same activity
  - 2D acquisition with higher activity

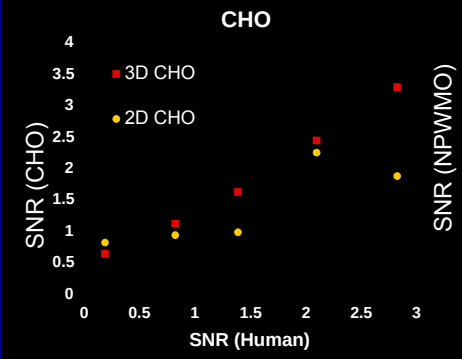
Observer Model

- CHO
  - 2D and volumetric channels
  - Constant Q channel model
  - No internal noise
- NPWO
- Image Contrast
- Ensemble training
- FOMs
  - Detectability index d_A computed from SNR for
    - CHO
    - NPWO
  - Image Contrast

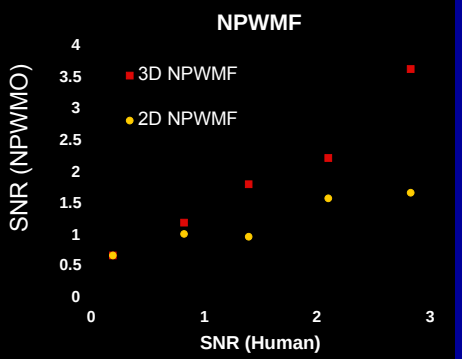


Results
Planar vs Volumetric Observer

- No axial smoothing



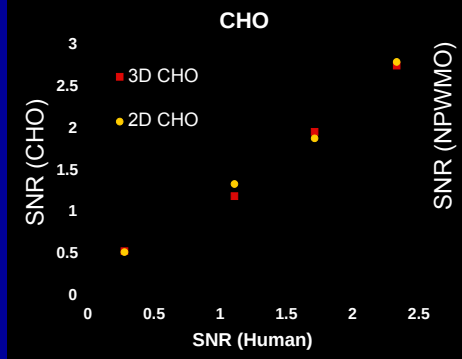
	2D CHO	3D CHO
Correlation Coef.	0.71	0.99



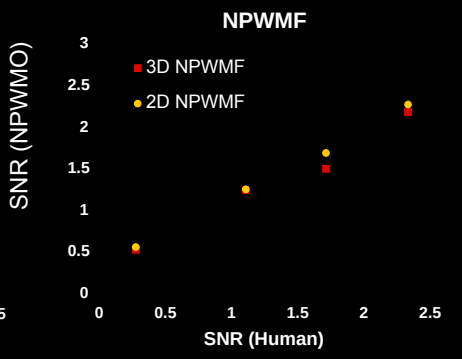
	2D NPWMF	3D NPWMF
Correlation Coef.	0.91	0.96

Results
Planar vs Volumetric Observer

- With axial smoothing



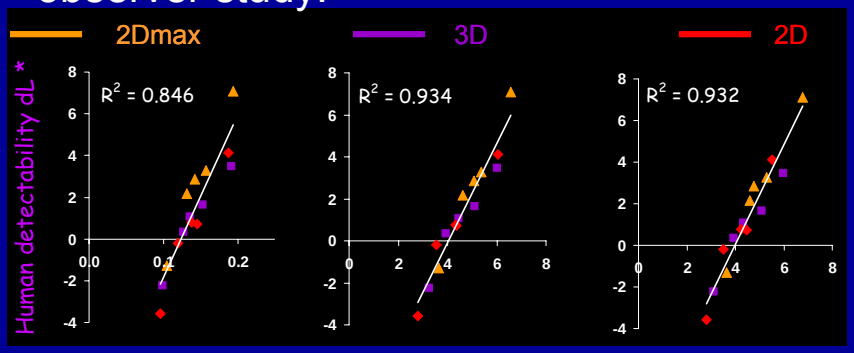
	2D CHO	3D CHO
Correlation Coef.	0.99	0.99



	2D NPWMF	3D NPWMF
Correlation Coef.	1.0	0.98

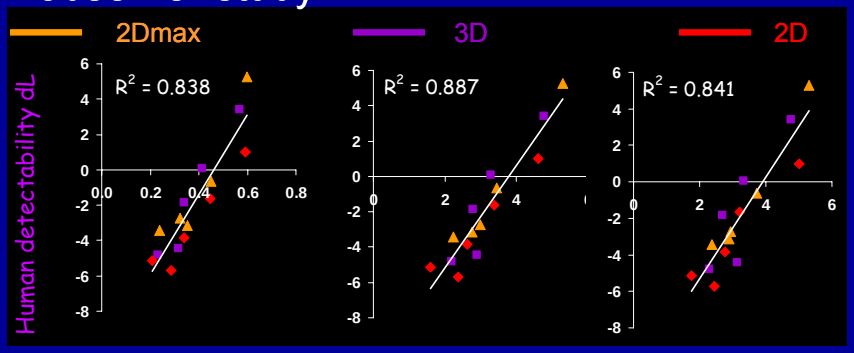
Results
Lesions in Lungs

Comparison with results of AFROC human observer study.



Results
Lesions in Liver

Comparison with results of AFROC human observer study.



Conclusions

- Volumetric (multi-slice) observer needed when there is only 2-D processing
- CHO performed poorly in soft tissue
  - Perhaps due to presence of multiple lesions
- Contrast did not work well in lungs or liver
- NPWO worked well in both tissues

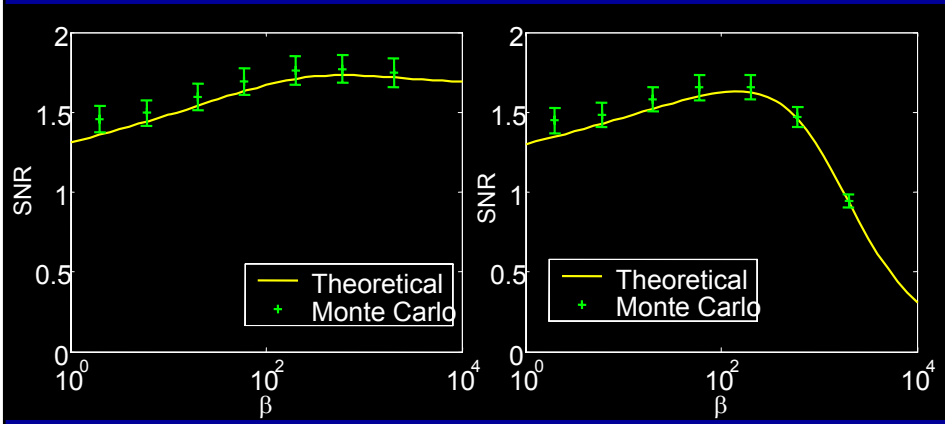
Example 5
Optimization of Penalty Parameters

- J.Y. Qi 2004, 2006a, 2006b
- Task: Detection (ROC)
- Object Model
  - Torso w/ Lumpy Background
  - Spherical lesions with varying size and position
- Image Formation
  - Analytic 2D simulation
  - Penalized likelihood reconstruction

Observer Model

- CHO
- Channel Model
  - SQR and D.O.G. Channels
- Computed using fast (non ensemble) estimation method
- Estimated from noisy and noise-free images
- F.O.M. CHO SNR

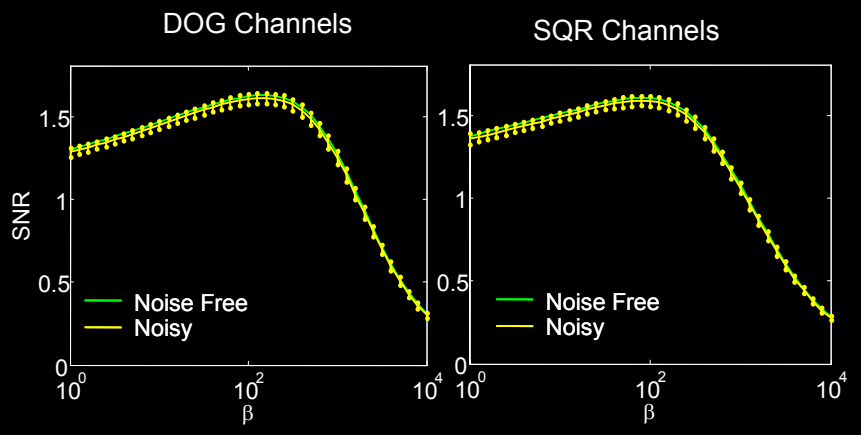
Comparison of Fast and Ensemble Estimation of CHO SNR



Internal Noise σ^2= 0 Internal Noise σ^2= 0.005

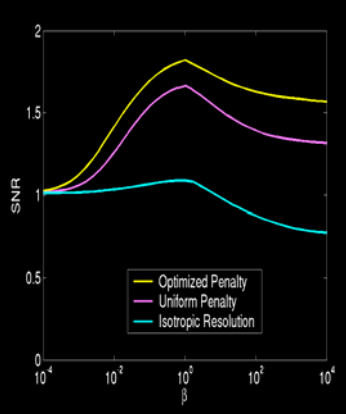
Estimation of FOM from Noisy Images

- Can estimate SNR from noisy images

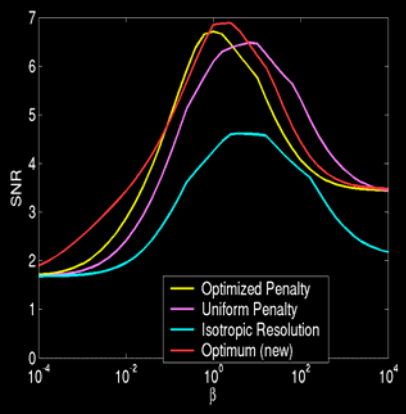


- From 500 noisy data sets

Optimization of Regularization Parameters



5-mm lesion



15-mm lesion
(with 5-mm lesion penalty functions)

Conclusions

- CHO SNR can be estimated rapidly for penalized likelihood reconstructions
- Can be estimated from noisy data
- Potential to allow for fast optimization of MAP parameters from patient data

Summary

- Model observers have been applied to optimization and evaluation of a wide variety of nuclear medicine imaging problems
- Generally good agreement between human and model observers
- There is a large body of literature describing methods that work
- Must take care to use appropriate task, object, image formation, and observer model to obtain meaningful results
- Validation of observer models still remains important area of research