Performance-based assessment of reconstructed images

Ken Hanson

Contributions of Robert F. Wagner February 11, 2009 SPIE Medical Imaging Symposium

Toronto & Ottawa – 1976

- I met Bob at the1976 SPSE meeting in Toronto (Dainty and Shaw)
 - ► I went there to learn about image science
 - I had joined Los Alamos Scientific Lab in 1975
 - demonstrate feasibility and dose advantage of proton over x-ray CT
- As physicists, we had a similar view of the world we hit it off right away
- We drove together to Ottawa for Int. Congress on Medical Imaging
 - heard David Chesler talks about correlations in CT noise
- We conversed regularly during 80s and 90s; sorted out many ideas
 - ► by telephone
 - ► at SPIE Medical Imaging meetings
 - ► I visited Bob at BRH/CDRH once a year

Hawaii – 1986



OSA workshop on quantum-limited imaging; Hawaii – 1986

Image quality – task performance

- In 1980s, rms error of reconstructed images were used to judge reconstruction quality
- SPIE medical-imaging community, led by Bob Wagner, did not think this was not an appropriate metric
- We believed image quality should be related to ability to perform visual tasks
- With encouragement and advice from Bob, I set out to demonstrate this principle in context of tomographic reconstruction

ART reconstructions from limited data

- Projection data: 12 views, no noise; artifact limited
- ART (Algebraic Reconstruction Technique)
 - with/without non-negativity constraint
- ROC analysis for detection of low-contrast disks; "ideal" observer



 $d' = 0.90 \pm 0.15$

rms err = 0.109

d' = 2.09 ± 0.22 rms err = 0.074

5

ART reconstructions from complete data

- Projection data: 180 views, moderate noise; no artifacts, just noise
- ART (Algebraic Reconstruction Technique)
 - without/with non-negativity constraint
- rms error does **not** indicate usefulness of reconstruction



d' = 1.96 ± 0.21 rms err = 0.101

d' = 1.99 ± 0.21 rms err = 0.063

Rockville – 1988



CDRH - 1988

Collaboration with Bob –1988-1995

- Shortly after arriving at FDA in 1987, Kyle Myers and Bob began to collaborate with me to greatly extend this work
- We undertook a series of stimulating studies
 - human observer testing (with a lot of help from Art Burgess)
 - ► Rayleigh task higher order relying on mid-spatial frequency info.
 - ► variety of machine observers, incl. neutral net, Bayesian, ...
- In summary, we demonstrated that
 - relative performances for competing reconstruction algorithms could be assessed using machine observers
 - rms error does not always correlate with task performance

Simulating CT measurements & reconstruction

simulated scene



Rayleigh task

Using reconstruction, how well can one determine whether objects are bars or two dots?

Comparing reconstruction algorithms

- Simulated projection data: 8 views, moderate noise
- Reconstructions obtained with MEMSYS3
 - question: how strong should prior be, α ?
- Assessment based on binary decision between two alternatives
 - ► ROC analysis
- Human observers testing two alternative forced choice







Reconstructions for various values of α

Comparing reconstruction algorithms

- Graph shows typical results
- MEMSYS3 reconstruction algorithm entropic prior
- We compared
 - human observers
 - several "machine" observers
 - e.g. neural networks
- This work showed proper way to evaluate and compare image processing algorithms

Task performance for MEMSYS3 reconstructions



K J Myers, M P Anderson, D G Brown, R F Wagner, and K M Hanson, *Proc. SPIE* **2434**, pp. 828-837 (1995) [received best poster award]

Rockville – 1992



Bob explaining the Fischer Information Matrix on the board and working on his computer – 1992

Flagstaff - 1993



On an excursion during IPMI meeting in Flagstaff – 1993

Newport Beach – 1994



Beach session at SPIE MI meeting, Newport Beach – 1994

Los Alamos – 1995



On top of Los Alamos ski hill, after MaxEnt meeting in Santa Fe – 1995

San Diego – 2007



Lunch at SPIE Medical Imaging meeting – 2007

Memorial Website

http://sites.google.com/site/robertfwagnermemorial



Please send me photos and reminiscences to add to website – kmhpix@q.com

Bibliography

- Hanson, K. M., "Variations in task and the ideal observer," Proc. SPIE 419, 60–67 (1983)
- Wagner, R. F., K. J. Myers, G. D. Brown, M. J. T., and Burgess, A. E., "Higher order tasks: human vs. machine performance," *Proc. SPIE* **1090**, 183–194 (1989)
- Hanson, K. M., "Method to evaluate image-recovery algorithms based on task performance," *J. Opt. Soc. Amer. A* **7**, 45–57 (1990)
- Myers, K. J. and Hanson, K. M., "Comparison of the algebraic reconstruction technique with the maximum entropy reconstruction technique for a variety of detection tasks," *Proc. SPIE* **1231**, 176–187 (1990)
- Hanson, K. M. and Myers, K. J., "Rayleigh task performance as a method to evaluate image reconstruction algorithms," *Maximum Entropy and Bayesian Methods*, 303–312, Kluwer Academic, Dordrecht (1991)
- Myers, K. J. and Hanson, K. M., "Task performance based on the posterior probability of maximum-entropy reconstructions with MEMSYS 3," *Proc. SPIE* **1443**, 172–182 (1991)
- Wagner, R. F., Myers, K. J., and Hanson, K. M., "Task performance on constrained reconstructions: Human observers compared with suboptimal Bayesian performance," *Proc. SPIE* **1652**, 352–362 (1992)
- Myers, K. J., Wagner, R. F., and Hanson, K. M., "Binary task performance in images reconstructed with MEMSYS3: comparison of machine and human observers," *Maximum Entropy and Bayesian Methods*, 415–421, Kluwer Academic, Dordrecht (1993)

Bibliography

- Wagner, R. F., Myers, K. J., and Hanson, K. M., "Rayleigh task performance in tomographic reconstructions: Comparison of human and machine performance," *Proc. SPIE* **1898**, 628–637 (1993)
- Myers, K. J., Wagner, R. F., and Hanson, K. M., "Rayleigh task performance in tomographic reconstructions: comparison of human and machine performance," *Proc. SPIE* **1898**, 628–637, SPIE (1993)
- Myers, K. J., Wagner, R. F., Hanson, K. M., Barrett, H. H., and Rolland, J. P., "Human and quasi-Bayesian observers of quantum-, artifact-, and object-variability-limited images," *Proc. SPIE* **2166**, 180–190 (1994)
- Wagner, R. F., Myers, K. J., Brown, D. G., Anderson, M. P., and Hanson, K. M., "Toward optimal human and algorithmic observer performance of detection and discrimination tasks on reconstruction from sparse data," *Maximum Entropy and Bayesian Methods* 211–220 Kluwer Academic, Dordrecht (1995)
- Myers, K. J., Anderson, M. P., Brown, D. G., Wagner, R. F., and Hanson, K. M., "Neural network performance for binary discrimination tasks, part II: effect of task training, and feature pre-selection," *Proc. SPIE* **2434**, 828–837 (1995)