Imaging System Design and Analysis: 
*from optics through information extraction*
Joseph A. O’Sullivan

Electronic Systems and Signals Research Laboratory
Department of Electrical Engineering
Washington University

jao@essrl.wustl.edu
http://essrl.wustl.edu/~jao

Supported by: ONR, ARO, NIH, DARPA
Collaborators

**Faculty**
- Donald L. Snyder
- William H. Smith
- Daniel R. Fuhrmann
- Jeffrey F. Williamson
- Bruce R. Whiting
- Richard E. Blahut
- John C. Schotland
- Michael I. Miller
- Chrysanthe Preza
- David G. Politte
- James G. Blaine

**Students and Post-Docs**
- Michael D. DeVore
- Natalia A. Schmid
- Metin Oz
- Ryan Murphy
- Jasenka Benac
- Adam Cataldo
- Lee Montagnino
- Andrew Li
Outline:

- Systems View of Imaging
- Opportunities for Joint Design
- Roles of Information Theory
- Examples (as desired)
  - Alternating Minimization Algorithms
    Applications in CT, HSI
  - Performance-Complexity Tradeoffs
    Applications in SAR ATR
- Conclusions
Imaging System Design

• System Considerations
  – Focus on end-to-end performance
    *requires clear system goal(s)*
  – Component versus system optimization
  – Imaging System Components as Processors

• Irreversible Operations
  – Role of Sufficient Statistics

• Adaptability, Flexibility, Feedback

• Hardware, Software, Processing

• Roles of Information Theory
System Considerations

Information Flow

Electromagnetic waves → Sampling → DSP → Comm. → DSP → Display

Design and Analysis:  Forward and Backward
Motivates Adaptability and Feedback

Information flow forward and backward
Some Roles of Information Theory In Imaging Problems

Some Important Ideas:

- Roles depend on problem studied
- Key problems are in detection, estimation, and classification
- Information is quantifiable
- SNR as a measure has limitations
- Information theory often provides performance bounds
How to Measure Information

QUESTIONS:

Can we measure the information in an image?
Does one sensor provide more information than another?
Does resolution measure information?
Do more pixels in an image give more information?

ANSWER:

MAYBE

Information for what?
Information relative to what?

Ill-defined questions → clearly defined problem
Measuring Information

Information for what?
- Detection
- Estimation
- Classification
- Image formation

Information relative to what?
- Noise only
- Clutter plus noise
- Natural Scenery
Information Theory and Imaging

- Center for Imaging Science established 1995
- Brown MURI on Performance Metrics established 1997
- Invited Paper 1998
  J. A. O’Sullivan, R. E. Blahut, and D. L. Snyder,
- IEEE 1998 Information Theory Workshop on Detection, Estimation, Classification, and Imaging
- *IEEE Transactions on Information Theory* Special Issue on Information-Theoretic Imaging, Aug. 2000
- Complexity Regularization, Large Deviations, …
Information-Theoretic Image Formation

Joseph A. O'Sullivan, Senior Member, IEEE, Richard E. Blahut, Fellow, IEEE, and Donald L. Snyder, Fellow, IEEE

(Invited Paper)

Abstract—The emergent role of information theory in image formation is surveyed. Unlike the subject of information-theoretic communication theory, information-theoretic imaging is far from a mature subject. The possible role of information theory in problems of image formation is to provide a rigorous framework for defining the imaging problem, for defining measures of optimality, and for quantifying the statistical quality of the approximations.

To this end, the domain of information theory may be

IEEE Transactions on Information Theory, October 1998,
Special Issue to commemorate the 50th anniversary of
Claude E. Shannon's
A Mathematical Theory of Communication

Problem Definition →
Optimality Criterion →
Algorithm Development →
Performance Quantification
Hyperspectral Imaging at Washington University

WU Team

Donald L. Snyder
William H. Smith
Daniel R. Fuhrmann
Joseph A. O’Sullivan
Chrysanthi Preza

Digital Array Scanning Interferometer (DASI)
Hyperspectral Imaging

- Scene Cube $\rightarrow$ Data Cube
- “Drink from a fire hose”
- Filter wheel, interferometer, tunable FPAs
- Modeling and processing:
  - data models
  - optimal algorithms
  - efficient algorithms

2D scene at wavelength $\lambda_0$

"Scene Cube"

(polarization, propagation effects, time evolution)
Hyperspectral Imaging Likelihood Models

ideal data: \( r(y) = \mu(y) \)

\[ \mu(y) = \int \int h(y - x : \lambda) s(x : \lambda) \, dx \, d\lambda \]

\[ h(y - x : \lambda) = \left| h_a(y - x - \Delta : \lambda) + h_a(y - x + \Delta : \lambda) \right|^2 \]

\( \Delta \) shear vector for Wollaston prism

\( h(y - x : \lambda) \) wavelength dependent amplitude PSF of DASI

\( s(x : \lambda) \) scene intensity for incoherent radiation at \( x, \lambda \)

nonideal (more realistic) data:

\[ r(y) = Poisson(\mu(y) + \mu_0(y)) + Gaussian(y) \]

data likelihood:

\[ E(r \mid scene) = -\log \left\{ \sum_{y=1}^{Y} \sum_{n(y)=1}^{\infty} \frac{1}{n(y)} \left[ \mu(y) + \mu_0(y) \right]^{n(y)} e^{-\left[ r(y) - \mu(y) \right]} e^{-\left[ r(y) - n(y) \right]^2 / 2 \sigma^2} \right\} \]
Idealized Data Model

- Data spectrum for each pixel $s_j$
- Linear combination of constituent spectra

\[ s_j = \sum_{k=1}^{K} \phi_k a_{kj} \]

- Problem: Estimate constituents and proportions subject to nonnegativity; positivity of $S$ assumed
- Ambiguity if $\alpha > 0, \phi_1 - \alpha \phi_2 > 0, -\alpha \phi_1 + \phi_2 > 0$
- Comments: Radiometric Calibration; Constraints Fundamental
Idealized Problem Statement:
Maximum-Likelihood $\implies$ Minimum I-divergence

$$S = \Phi A$$

- **Poisson distributed data $\implies$ loglikelihood function**

$$l(S \mid \Phi A) = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ s_{ij} \ln \left( \sum_{k=1}^{K} \phi_{ik} a_{kj} \right) - \sum_{k=1}^{K} \phi_{ik} a_{kj} \right\}$$

- **Maximization over $\Phi$ and $A$ equivalent to minimization of I-divergence**

$$I(S \parallel \Phi A) = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ s_{ij} \ln \left( \frac{s_{ij}}{\sum_{k=1}^{K} \phi_{ik} a_{kj}} \right) - s_{ij} + \sum_{k=1}^{K} \phi_{ik} a_{kj} \right\}$$

- **Information Value Decomposition Problem**
Markov Approximations

- $X$ and $Y$ RV’s on finite sets, $p(x,y)$ unknown
- Data: $N$ i.i.d. pairs $\{X_i, Y_i\}$
- Unconstrained ML Estimate of $p(x,y)$

\[
S = \frac{n(x, y)}{N}
\]

- Lower rank Markov approximation
  
  $X \rightarrow M \rightarrow Y$

  $M$ in a set of cardinality $K$

- Factor analysis, contingency tables, economics

- Problem: Approximation of one matrix by another of lower rank


- SVD $\rightarrow$ IVD
CT Imaging in Presence of High Density Attenuators

Brachytherapy applicators
After-loading colpostats for radiation oncology

Cervical cancer: 50% survival rate
Dose prediction important

Object-Constrained Computed Tomography (OCCT)
Filtered Back Projection

FBP: inverse Radon transform
Transmission Tomography

- Source-detector pairs indexed by $y$; pixels indexed by $x$
- Data $d(y)$ Poisson, means $g(y; \mu)$, loglikelihood function

\[
I(d : \theta) = \sum_{y} [d(y) \ln g(y : \theta) \cdot g(y : \theta)]
\]

\[
g(y : \theta) = \prod_{E} I_0(y; E) \exp \left[ \sum_{x} h(y|x)^{\theta}(x; E) \right] + \bar{g}(y)
\]

- Mean unattenuated counts $I_0$, mean background $\beta$
- Attenuation function $\mu(x,E)$, $E$ indexes energies

\[
\mu^{(\theta)}(x; E) = \prod_{i=1}^{P} \frac{\mu^{(\theta)}(E)}{c_i(x)}
\]

- Maximize over $\mu$ or $c_{i\mu}$; equivalently minimize l-divergence
- Comment: pose search $c(x) = c_a(x; \theta) + c_b(x)$
Alternating Minimization Algorithms

- Define problem as $\min_q \phi(q)$
- Derive Variational Representation: $\phi(q) = \min_p J(p,q)$
- $J$ is an auxiliary function $p$ is in auxiliary set $P$
- Result: double minimization $\min_q \min_p J(p,q)$
- Alternating minimization algorithm

$$
\begin{align*}
p^{(l+1)} &= \arg \min_{p \in P} J(p, q^{(l)}) \\
q^{(l+1)} &= \arg \min_{q \in Q} J(p^{(l+1)}, q)
\end{align*}
$$

Comments: Guaranteed Monotonicity; $J$ selected carefully
Alternating Minimization Algorithms: I-Divergence, Linear, Exponential Families

- Special Case of Interest: $J$ is I-divergence

- Families of Interest:
  Linear Family $L(A,b) = \{p: Ap = b\}$
  Exponential Family $E(\pi,B) = \{q: q_i = \pi_i \exp[\sum_j b_{ij} v_j]\}$

\[
p^{(l+1)} = \arg\min_{p \in L} \mathcal{I}(p \| q^{(l)})
\]
\[
m_{(l+1)} = \arg\min_{q \in E} \mathcal{I}(p^{(l+1)} \| q)
\]

Csiszár and Tusnády; Dempster, Laird, Rubin; Blahut; Richardson; Lucy; Vardi, Shepp, and Kaufman; Cover; Miller and Snyder; O'Sullivan
Alternating Minimization Example

- **Linear family**: \( p_1 + 2p_2 = 2 \)
- **Exponential family**: \( q_1 = \exp(v), q_2 = \exp(-v) \)

\[
\min_{q \in E} \min_{p \in L} I(p \| q)
\]
Information Geometry

- \( \text{I-divergence is nonnegative, convex in pair } (p,q) \)
- \( \text{Generalization of relative entropy, example of f-divergence} \)
- \( \text{First triangle equality: } p \text{ in } L \)

\[
p^* = \arg \min_{p \in L} I(p \| q) \Rightarrow I(p \| q) = I(p \| p^*) + I(p^* \| q)
\]

- \( \text{Second triangle equality: } q \text{ in } E \)

\[
q^* = \arg \min_{q \in E} I(p \| q) \Rightarrow I(p \| q) = I(p \| q^*) + I(q^* \| q)
\]
Variational Representations

• **Convex decomposition lemma.** Let $f$ be convex. Then

$$f\left(\sum_{i} x_i\right) \leq \sum_{i} r_i f\left(\frac{1}{r_i} x_i\right)$$

$$\sum_{i} r_i = 1, r_i \geq 0$$

• **Special Case:** $f$ is ln

$$\ln\left(\sum_{i} q_i\right) = -\min_{p \in P} \sum_{i} p_i \ln \frac{p_i}{q_i}$$

$$P = \left\{ p : \sum_{i} p_i = 1 \right\}$$

• **Basis for EM; see also De Pierro, Lange, Fessler**
Shrink-Wrap Algorithm for Endmembers

\[ S = \Phi A \]

\( \Phi = \) Endmembers, K Columns
\( A = \) Pixel Mixture Proportions

SVD, Then Simplex Volume Minimization
Alternating Minimization Algorithms for Hyperspectral Imaging

\[ S = \Phi A \]

Given \( \Phi \) and \( S \), estimate \( A \).

Uses I-Divergence Discrepancy Measure.
Information Value Decomposition Applied to Hyperspectral Data

- Downloaded spectra from USGS website
- 470 Spectral components
- Randomly generated A with 2000 columns
- Ran IVD on result
Information Value Decomposition
Applied to Hyperspectral Data
Information Theoretic Imaging: Application to Hyperspectral Imaging

- Likelihood Models for Sensors
- Likelihood Models for Scenes
- Spectrum Estimation and Decomposition
- Performance Quantification
- Applications to Available Sensors
Hyperspectral Imaging

Ongoing Efforts

• Scene Models Including Spatial and Wavelength Characteristics
• Sensor Models Including Apodization
• Orientation Dependence of Hyperspectral Signatures
• Expanded Complementary Efforts
• Atmospheric Propagation Models
• Performance Bounds
• Measures of Added Information