Information-Theoretic Imaging
with Applications in
Hyperspectral Imaging and
Transmission Tomography

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
Collaborators

Faculty

Donald L. Snyder
William H. Smith
Daniel R. Fuhrmann
Jeffrey F. Williamson
Bruce R. Whiting
Richard E. Blahut
Michael I. Miller
Chrysanthi 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
Outline:

- Information Theory and Imaging
- Hyperspectral Imaging
  - Information Value Decomposition
- Transmission Tomography
  - Estimate in Exponential Family
  - Object-Constrained Computerized Tomography
- Alternating Minimization Algorithms
- Applications: HSI, CT
- Conclusions
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 $\rightarrow$ 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 Imaging

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.

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
Chrysanthe Preza

Digital Array Scanning Interferometer (DASI)
Hyperspectral Imaging

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

Scene Cube

2D scene at wavelength \( \lambda_0 \)

radiant spectrum at position (a,b)

"Scene Cube"

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

ideal data: \[ r(y) = \mu(y) \]
\[ \mu(y) = \iint h(y - x : \lambda)s(x : \lambda) dx d\lambda \]

\[ h(y - x : \lambda) = |h_a(y - x - \Delta : \lambda) + h_a(y - x + \Delta : \lambda)|^2 \]
\[ \Delta \] 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:
\[ r(y) = \text{Poisson}[\mu(y) + \mu_0(y)] + \text{Gaussian}(y) \]

data likelihood:
\[ E(r \mid \text{scene}) = -\log \left\{ \sum_{y=1}^{\infty} \sum_{n(y)=1}^{\infty} \frac{1}{n(y)} [\mu(y) + \mu_0(y)]^{n(y)} e^{-[\mu(y) + \mu_0(y)]} e^{-[r(y) - n(y)]^2 / 2 \sigma^2} \right\} \]
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 $\rightarrow$ Minimum I-divergence

$$S = \Phi A$$

- Poisson distributed data $\rightarrow$ loglikelihood function

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

- 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

\[
\begin{aligned}
\mathcal{L}(d : ¹) &= \prod_{y \in Y} [d(y) \ln g(y : ¹) + g(y : ¹)] \\
g(y : ¹) &= \int_{E} l_{0}(y; E) \exp \left[ \sum_{x} h(yjx)^{¹}(x; E) \right] + \bar{y}(y)
\end{aligned}
\]

- Mean unattenuated counts \( l_{0} \), mean background \( \beta \)
- Attenuation function \( \mu(x,E) \), \( E \) indexes energies

\[
¹(x; E) = \sum_{i=1}^{P} l_{i}(E)c_{i}(x)
\]

- Maximize over \( \mu \) or \( c_{i} \); equivalently minimize I-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

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

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** $E(\pi,B) = \{q: q_i = \pi_i \exp[\sum_j b_{ij} v_j]\}$

\[
p^{(l+1)} = \arg \min_{p \in L} I(p \parallel q^{(l)})
\]

\[
q^{(l+1)} = \arg \min_{q \in E} I(p^{(l+1)} \parallel q)
\]

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**

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

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

- **Second triangle equality: \(q\) 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

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

$$\sum_i r_i = 1, r_i \geq 0$$

- 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

\[ S = \Phi A \]

\( \Phi = \text{Endmembers, K Columns} \)

\( A = \text{Pixel Mixture Proportions} \)

SVD, Then Simplex Volume Minimization

**Graphs:**
- \( \phi_1 : \text{Aspen} \)
- \( \phi_2 : \text{Augite} \)
- \( \phi_3 : \text{Olivine} \)

**Diagrams:**
- Projected data and known constituents
- Initial Triangle
- Progress of Shrink-Wrap Algorithm
- Shrink-Wrap Algorithm: Final Result
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

Six Endmembers

Reflectance

$\Phi_1$ : Aspen

Reflectance

$\Phi_3$ : Olivine

Reflectance

$\Phi_4$ : Gypsum

Reflectance

$\Phi_2$ : Augite

Reflectance

$\Phi_5$ : Almandine

Reflectance

$\Phi_6$ : Colemanite

Reflectance

0 1 2 3

Wavelength

0 1 2 3

Wavelength
Information Theoretic Imaging: Application to Hyperspectral Imaging

- Problem Definition
- Optimality Criterion
- Algorithm Development
- Performance Quantification
- 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
CT Minimum I-Divergence Formulation

\[ l(d : ¹) = \sum_{y \in Y} [d(y) \ln g(y : ¹) + g(y : ¹)] \]

\[ g(y : ¹) = \sum_{E} I^0(y; E) \exp \left[ i \int_x h(y|x)¹(x; E) \right] + \bar{g}(y) \]

Maximize loglikelihood function over \( \mu \)  \( \Rightarrow \) minimize I-divergence

Use Convex Decomposition Lemma

\( g \) is a marginal over \( q \)

\[ L(d) = \int p(y; E) \sum_{y \in Y} [d(y) \ln g(y : ¹) + g(y : ¹)] \]

\[ l(d; kg(y : c)) = \min_{p \in L(d)} l(p; kq) \]
**New Alternating Minimization Algorithm for Transmission Tomography**

\[
\min_{q \in E(I_0;H)} \min_{p \in L(d)} I(p, q)
\]

\[
\hat{q}^{(k)}(y; E) = I_0(y; E) \exp \sum_{i}^{\text{di}} i(E) h(y|x) \hat{c}_i^{(k)}(x)
\]

\[
\rho^{(k)}(y; E) = \hat{q}^{(k)}(y; E) \prod_{E^0} \frac{d(y)}{\hat{q}^{(k)}(y; E^0)}
\]

\[
\hat{\beta}_i^{(k)}(x) = \prod_{y \in Y} \prod_{E}^{\hat{c}_i^{(k)}(y; E)} i(E) h(y|x) \rho^{(k)}(y; E)
\]

\[
\hat{\beta}_i^{(k)}(x) = \prod_{y \in Y} \prod_{E}^{\hat{c}_i^{(k)}(y; E)} i(E) h(y|x) \hat{q}^{(k)}(y; E)
\]
New Alternating Minimization Algorithm for Transmission Tomography

\[ \hat{c}_i^{(k+1)}(x) = \hat{c}_i^{(k)}(x) \cdot \frac{1}{Z_i(x)} \ln \frac{b_i^{(k)}(x)}{\hat{b}_i^{(k)}(x)} \]

Interpretation: Compare predicted data to measured data via ratio of backprojections.
Update estimate using a normalization constant.

Comments: Choice for constants; monotonic convergence; Linear convergence; Constraints easily incorporated.
Iterative Algorithm with Known Applicator Pose
OCCT Iterations

Known Pose

OCCT
Magnified views around brachytherapy applicator

Truth

OCCT

FBP
Conclusions

- Information-Theoretic Imaging
- Alternating Minimization Algorithms
- Hyperspectral Imaging
  - Applying Formal Methodology
  - Data Collections
  - Sensor Modeling, Likelihoods
  - Broad, Ongoing Efforts
- CT Imaging
  - New Image Formation Algorithm
  - Simulations and Scanner Data
  - Modeling Issues: Physics Crucial
Physical Phenomena

Mathematical Abstractions

Reproduction Space

Scene Model

Sensor Models

$C' \subset C$

$\min d(\sigma, c)$

$\mathcal{L}(\sigma | c), \|\sigma - f(c)\|, I(\sigma \| f(c))$

Performance

$d(c, \hat{c}') = d(c, \hat{c}^*) + d(\hat{c}^*, \hat{c}')$

$C \in \mathbb{L}_2$

$\sigma \in \mathbb{O} \subset \mathbb{L}_2$
### Problem Definition

<table>
<thead>
<tr>
<th>Physical Phenomena:</th>
<th>Electromagnetic Waves, Physical Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematical Abstraction:</td>
<td>Function Spaces (L²), Dynamical Models</td>
</tr>
<tr>
<td>Sensor Models:</td>
<td>Projection Models, Statistics</td>
</tr>
<tr>
<td>Reproduction Space:</td>
<td>Finite Dimensional Approximation</td>
</tr>
</tbody>
</table>

### Optimality Criterion

<table>
<thead>
<tr>
<th>Deterministic:</th>
<th>Least Squares, Minimum Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stochastic:</td>
<td>Maximum Likelihood, MAP, MMSE</td>
</tr>
</tbody>
</table>
Algorithm Development

Iterative Algorithms: Alternating Minimizations
Expectation-Maximization

Random Search: Simulated Annealing,
Diffusion, Jump-Diffusion

Performance Quantification

Discrepancy Measures: Squared Error,
Discrimination

Performance: Bias, Variance,
Mean Discrimination

Performance Bounds: Cramer-Rao,
Hilbert-Schmidt
**Target Classification Problem**

Given a SAR image $r$, determine a corresponding target class $\hat{a} \in A$.

**Target Orientation Estimation**

Given a SAR image $r$ and a target class $\hat{a} \in A$, estimate target orientation.

**Likelihood Model Approach:**

- $p_{R|\Theta,a}(r|\theta,a)$ - Conditional Data Model
- $p_{\Theta,a}(\theta,a)$ - Prior on orientation (known or simply uniform)
- $P(a)$ - Prior on target class (known or simply uniform)
Training and Testing Problem: Function Estimation and Classification

- Labeled training data: target type and pose
- Log-likelihood parameterized by a function: mean image, variance image, etc.
Function Estimation and Classification

- Functions are estimated from
  - sample data sets only
  - physical model datasets only (PRISM, XPATCH, etc.)
  - combination of these

- Training sets are finite

- Computational and likelihood models have a finite number of parameters

- Estimation error

- Approximation error

- Some regularization is needed
Function Estimation

Let $S$ and $T$ be two target types, $f_S$ and $f_T$ the corresponding functions

$$f_S : \Theta \times D \rightarrow \mathbb{R} \text{ or } \mathbb{R}_+$$

Spatial domain $D$:
- Surface of object
- Rectangular region in $\mathbb{R}^2$
- Continuous

Pose space $\Theta$:
- Typically SE(3)
- Simpler to consider SE(2), SO(2)
- $\Theta$ is continuous

Assume a null function $f_0$
Function Estimation: Sieve

Function space $F$

Family of subsets $F_\mu$, $\mu > 0$

- Maximum likelihood estimate of $f$ exists for each $\mu > 0$
- $\bigcup_{\mu > 0} F_\mu = F$

Ulf Grenander

Spline sieves $\mu = 1/Q$, where $Q =$ number of coefficients.
Nonnegative functions, nonnegative expansions.
Pierre Moulin, et al.

$$f(\lambda) = \sum_{k \in \Lambda_\mu} a(k) \psi_k(\lambda), \quad a(k) \geq 0$$
Sieve Geometry

\[ \Phi_2 \cdot \phi_1 \cdot f_2 \cdot f_1 \cdot f_S \]

\[ F \]

\[ F_1 \quad F_2 \]
Discrepancy Measures

Estimation:
Relative entropy between parameterized distributions
\[ d(f_S, \phi_\mu) \quad \text{approximation error} \]
Expected relative entropy for estimated functions
\[ E\{d(\phi_\mu, f_\mu)\} \quad \text{estimation error} \]

Classification:
Probability of error, rate functions
Decompose into approximation and estimation error

Examine performance as a function of \( \mu \), optimize
Motivation

- Many reported approaches to ATR from SAR
- Performance and database complexity are interrelated
- We seek to provide a framework for comparison that:
  - Allows direct comparison under identical conditions
  - Removes dependency on implementation details

Publicly available SAR data from the MSTAR program
Approaches: Conditionally Gaussian

Model each pixel as complex Gaussian plus uncorrelated noise:

\[ p_{R|\Theta,A}(r|\theta,a) = \prod_i \frac{1}{\pi(K_i(\theta,a)+N_0)} e^{-\frac{|r_i|^2}{K_i(\theta,a)+N_0}} \]

GLRT Classification and MAP Estimation:

\[ \hat{a}_{\text{Bayes}}(r) = \arg\max_a \max_k \hat{p}(r|\theta_k,a) \]
\[ \hat{\theta}_{\text{HS}}(r,a) = \arg\max_{\theta_k} \hat{p}(r|\theta_k,a) \]
Approach

- Select 240 combinations of implementation parameters
- Execute algorithms at each parameterization
- Scatter plot the performance-complexity pairs
- Determine the best achievable performance at any complexity

BMP2 Variance Image at 6 Sizes

ZIL131 Variance Image at 6 Sizes

Six different image sizes from 128x128 to 48x48
System Parameters and Complexity

Approximate $\alpha(\theta,a)$ and $\sigma^2(\theta,a)$ as piecewise constant in $\theta$

Implementations parameterized by:
- $\nu$ - number of constant intervals in $\theta$
- $d$ - width of training intervals in $\theta$
- $N^2$ - number of pixels in an image

Database complexity $\equiv \log_{10}(\# \text{ floating point values} / \text{ target type})$
Performance and Complexity

Forty combinations of angular resolution and training interval width.

Variance image of a T62 tank
1 Window trained over 360°

Variance images of a T72 tank
72 Windows trained over 10°
Problem Statement

- Directly compare conditionally Gaussian, log-magnitude MSE, and quarter power MSE ATR Algorithms
  - identical training and testing data
  - identical spatial and orientation windows
- Plot performance vs. complexity
  - probability of classification error
  - orientation estimation error
  - log-database size as complexity
- Use 10 class MSTAR SAR images

Approach

- Select 240 combinations of implementation parameters
- Execute algorithms at each parameterization
- Scatter plot the performance-complexity pairs
- Determine the best achievable performance at any complexity
Approaches: Log-Magnitude

Minimize distance between $r_{dB} = 20 \log |r|$ and dB templates

$$d^2(r_{dB}, \mu_{LM}) = \| r_{dB} - \mu_{LM} \|^2$$

Make decisions according to:

$$\hat{a}_{LM}(r) = \arg \min_a \min_k d^2(r_{dB}, \mu_{LM}(\theta_k, a))$$

$$\hat{\theta}_{LM}(r|a) = \arg \min_{\theta_k} d^2(r_{dB}, \mu_{LM}(\theta_k, a))$$

Alternatively, use a form of normalization:

$$d^2(r_{dB} - \overline{r_{dB}}, \mu_{LM}(\theta_k, a) - \overline{\mu_{LM}(\theta_k, a)})$$

Approaches: Quarter Power

Minimize distance between $r_{QP} = |r|^{1/2}$ and quarter power templates

$$d^2 (r_{QP}, \mu_{QP}) = \|r_{QP} - \mu_{QP}\|^2$$

Make decisions according to:

$$\hat{a}_{QP} (r) = \arg\min_a \min_k d^2 (r_{QP}, \mu_{QP} (\theta_k, a))$$
$$\hat{\theta}_{QP} (r, a) = \arg\min_k d^2 (r_{QP}, \mu_{QP} (\theta_k, a))$$

Or, normalized by vector magnitude:

$$d^2 \left( \frac{r_{QP}}{\|r_{QP}\|}, \frac{\mu_{QP}(\theta_k, a)}{\|\mu_{QP}(\theta_k, a)\|} \right)$$

S. W. Worrell, et al., SPIE 3070, 1997
Discussions with M. Bryant of Wright Laboratory
Conditionally Gaussian Results
## Performance-Complexity Legend

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Complexity</th>
<th>Number of Piecewise Constant Intervals</th>
<th>Training Window Width</th>
</tr>
</thead>
<tbody>
<tr>
<td>○○</td>
<td>w = 1, d = 360°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>××</td>
<td>w = 2, d = 180°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>⊕⊕</td>
<td>w = 3, d = 120°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>★★</td>
<td>w = 3, d = 240°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>□□</td>
<td>w = 4, d = 90°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>◊◊</td>
<td>w = 4, d = 180°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▽▽</td>
<td>w = 5, d = 72°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>△△</td>
<td>w = 5, d = 144°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>← ←</td>
<td>w = 6, d = 60°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▼ ▼</td>
<td>w = 6, d = 120°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>★★</td>
<td>w = 8, d = 45°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>⊕ ⊕</td>
<td>w = 8, d = 90°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>× ⊙</td>
<td>w = 8, d = 60°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▼ ▽</td>
<td>w = 15, d = 24°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▽ ▽</td>
<td>w = 15, d = 48°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>△ △</td>
<td>w = 18, d = 20°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>△ ▽</td>
<td>w = 18, d = 40°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>★ ⊗</td>
<td>w = 18, d = 40°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>⊕ ⊗</td>
<td>w = 30, d = 12°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▽ ⊗</td>
<td>w = 30, d = 24°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>◊ ⊗</td>
<td>w = 36, d = 10°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▽ ⊗</td>
<td>w = 36, d = 20°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>⊕ ⊗</td>
<td>w = 72, d = 5°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▽ ⊗</td>
<td>w = 72, d = 10°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>⊕ ⊗</td>
<td>w = 72, d = 15°</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>⊕ ⊗</td>
<td>w = 72, d = 20°</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Forty combinations of number of piecewise constant intervals and training window width**
Log-Magnitude Results

Recognition without normalization

Arithmetic mean normalized
Quarter Power Results

Recognition without normalization

Recognition with normalization
Normalized Conditionally Gaussian Results

Results
Side-by-Side Results

Comparison in terms of:
- Performance achievable at a given complexity
- Complexity required to achieve a given performance
## Target Classification Results

- **Probability of correct classification: 97.2%**

### Target Classification Results Table

<table>
<thead>
<tr>
<th></th>
<th>2S1</th>
<th>BMP 2</th>
<th>BRDM 2</th>
<th>BTR 60</th>
<th>BTR 70</th>
<th>D7</th>
<th>T62</th>
<th>T 72</th>
<th>ZIL131</th>
<th>ZSU 23 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>2S1</td>
<td>262</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BMP 2</td>
<td>0</td>
<td>581</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BRDM 2</td>
<td>5</td>
<td>3</td>
<td>227</td>
<td>1</td>
<td>0</td>
<td>14</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>BTR 60</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>193</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BTR 70</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>184</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>D7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>271</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T 62</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>259</td>
<td>11</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>T 72</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>582</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ZIL131</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>272</td>
<td>0</td>
</tr>
<tr>
<td>ZSU 23 4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>271</td>
</tr>
</tbody>
</table>
Conclusions

- General problem in training/testing posed as estimation/classification
- Method of sieves (polynomial splines chosen)
- Comprehensive performance-complexity study for
  - Ten class MSTAR problem
  - Conditionally Gaussian model
  - Log-magnitude MSE
  - Quarter power MSE
- Provided a framework for direct comparison of alternatives and selection of implementation parameters
- Analysis ongoing