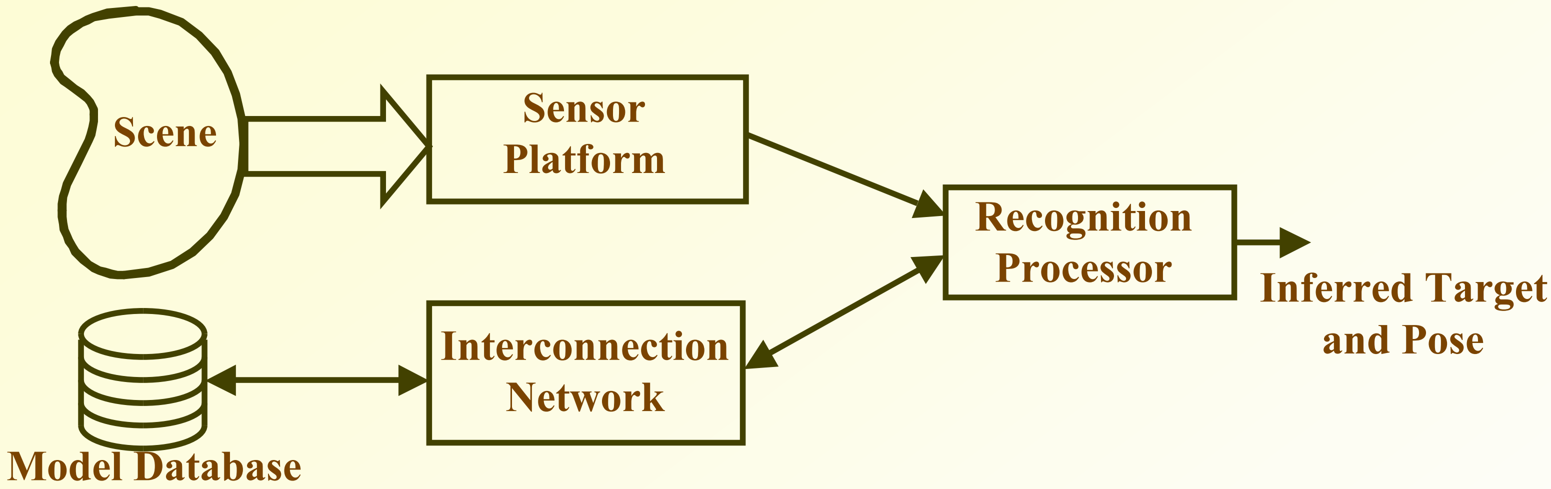
Object Recognition Systems with Dynamically Varying Resource Constraints

Is it true that "There's no theory for ATR?"

Michael D. DeVore
Systems and Information Engineering
University of Virginia
mdevore@virginia.edu

Object Recognition Problem



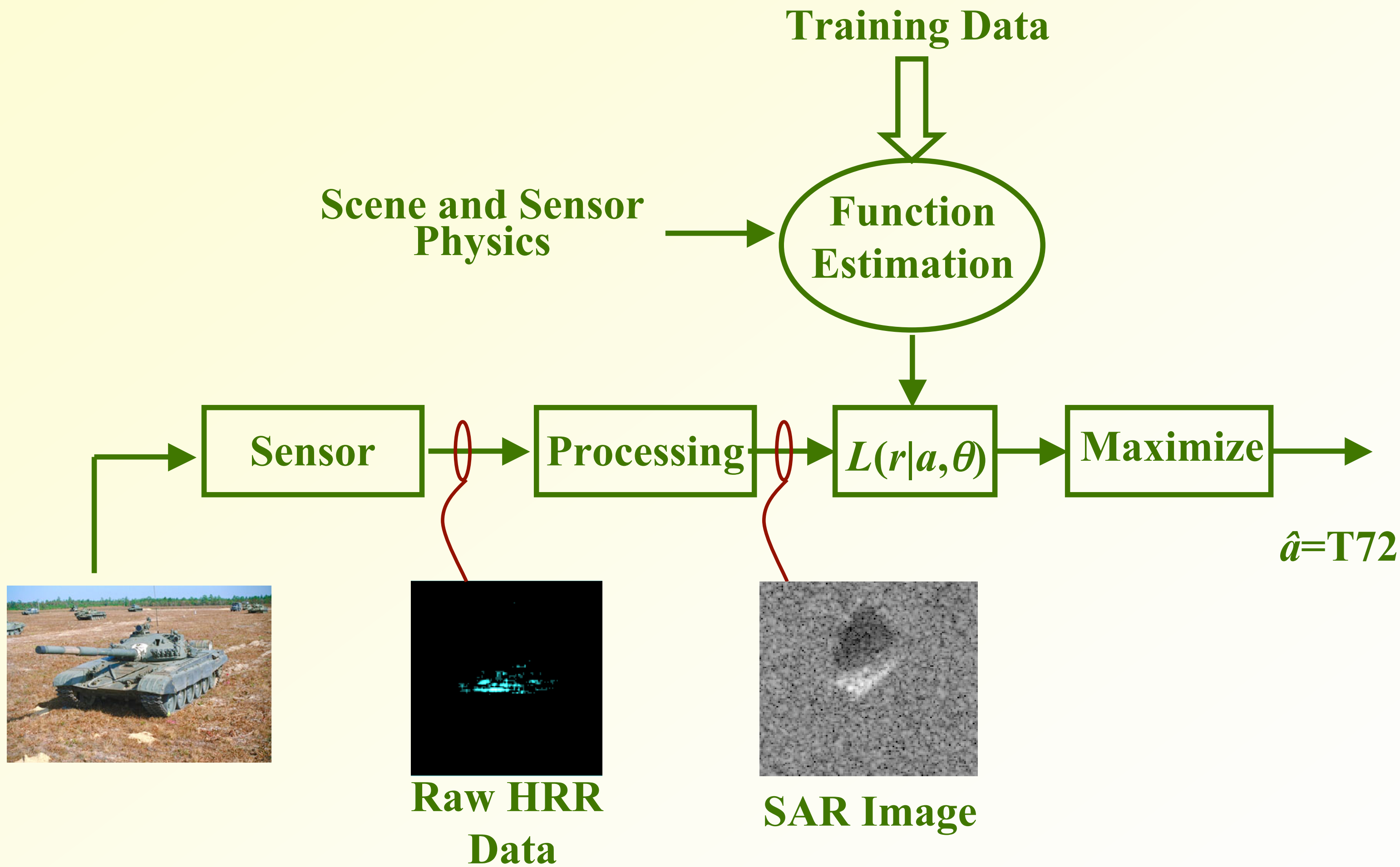
- Processor as a receiver from two sources:
  - Scene originates information carried through the sensor channel
  - Database of information on prospective targets carried through interconnection channel
- Apply information & communication theories to object recognition

What the Data Processing Inequality Doesn't Say

- Doesn't everyone think of recognition in this way?
  - Casually, yes
  - Formally, not really
- Data Processing Inequality:

  If $X \rightarrow Y \rightarrow Z$, then $\mathrm{I}(X;Y) \geq \mathrm{I}(X;Z)$
- Interpretation: If we want to infer X, we're better off using Y than some function $Z = G(Y)$
  - If G is invertible or a sufficient statistic, we lose nothing
- Problem 1: Sometimes our model for $p(X,Y)$ is not as good as for $p(X,Z)$
- Problem 2: Sometimes working with Y incurs too great a resource expenditure

Processing Example



Processing blocks can include image formation, feature extraction, dimensionality reduction, etc.

Unique Domain Features

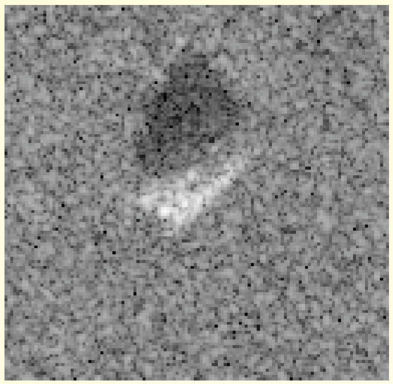
- Little control over source coding, channel coding, or modulation schemes
  - Many tools (i.e. Huffman coding) unavailable
  - Conditional output distributions must be estimated
- Unable to completely specify a null hypothesis
  - Segmentation and confuser rejection
- Source alphabet of extremely large cardinality
  - Consists of {objects} × {poses} × {articulations}
  - Codebook itself becomes subject of compression
- Often unable to exploit block-encoding
  - Rate-distortion theory and typical sets
- Often bit rate not the most relevant rate
  - System performance under time and resource constraints

Model Estimation

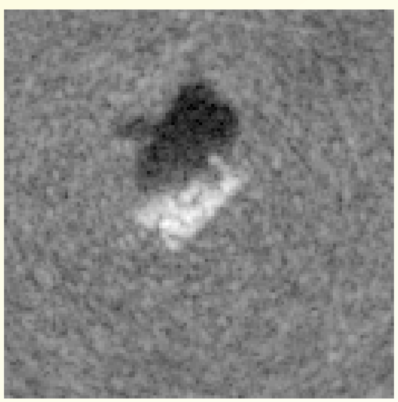
- Let a be a source symbol, θ a nuisance parameter, and r an observation
- In many communication problems, modulation $s(a,\theta)$ and channel transformation h are known
  - Models $p(r|a,\theta)$ are directly obtainable from $r = h[s(a,\theta)]$
- In recognition problems, we often must estimate $p(r|a,\theta)$ from observations $(r_1, a_1, \theta_1), \ldots, (r_n, a_n, \theta_n)$
  - Modulation and channel are neither known nor directly observable
  - Often referred to as "training"
- Generally restrict $p(r|a,\theta)$ to some parameterized family
- Special considerations for statistical assessment of model assumptions

Example: Complex Gaussian Family

T-72 Photo



SAR Image



Variance Image

Model pixel *i* as conditionally independent, complex Gaussian

$$p\left(r_i \middle| a,\theta\right) = \frac{1}{\pi\sigma_i^2(a,\theta)} e^{-\frac{\left|r_i - \mu_i(a,\theta)\right|^2}{\sigma_i^2(a,\theta)}}$$

Where: a **= target class** $\quad \sigma_i^2$ **= variance function**

θ **= target pose** $\quad \mu_i$ **= mean function**

If phase of r_i is uniform on $[0,2\pi)$, $\mu_i(a,\theta) = 0$

Further restrict function σ^2, for example:

- Class of piecewise constant functions
- Class of functions with a particular series representation

Null Hypothesis Issues

- Many classification problems share the following formulation:

  H_0: $r(t) \sim n(t)$ (pure noise)

  H_i: $r(t) \sim s_i(t) + n(t)$ (combination of signal and noise)

- Typical solution involves projecting $r(t)$ onto a set of functions whose span contains the $s_i(t)$
- Recognition algorithms must often reject "target-like" objects and work regardless of the scene background
- Model for r in the absence of a target (signal) is problematic

Example: Two-Sided Hypothesis Tests

- Suppose $r \sim \mathrm{CN}(r \mid \sigma^2)$ and choose H_0: $\sigma^2 = \sigma_0^2$ or H_A: $\sigma^2 \neq \sigma_0^2$
- Likelihood ratio test maximizes power at a single alternative
- Reject H_0 if

$$\ln \frac{p\left(r \middle| \xi^2\right)}{p\left(r \middle| \sigma_0^2\right)} = \ln \frac{\sigma_0^2}{\xi^2} + \frac{|r|^2}{\sigma_0^2} - \frac{|r|^2}{\xi^2} > \eta$$

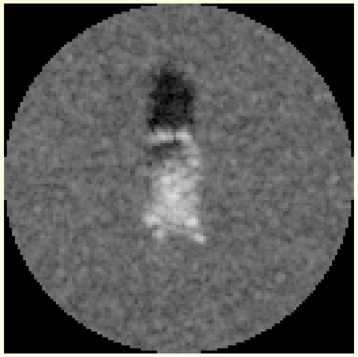
- Most likely alternative is $\xi^2 = |r|^2$
- Maximum power test at most likely alternative becomes an empirical relative entropy

$$D\left(p\left(r \middle| |r|^2\right) \middle\| p\left(r \middle| \sigma_0^2\right)\right) = \ln \frac{\sigma_0^2}{|r|^2} + \frac{|r|^2}{\sigma_0^2} - 1 > \eta$$

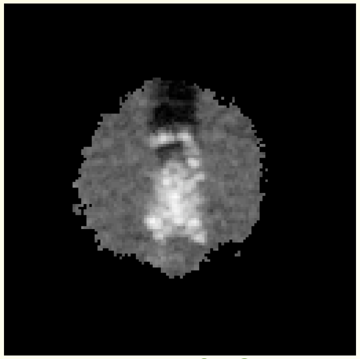
Application: Target Model Segmentation

- Object Recognition
- Unique Features
  - Model Estimation
  - Null Hypotheses
  - Codebook Representation
- Dynamic Reconfigurability
- Other Rates
- Directions

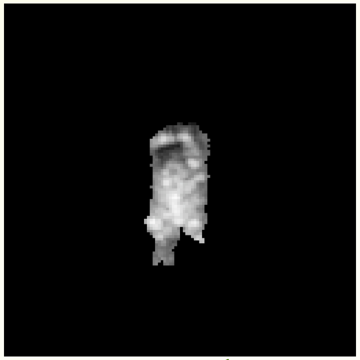
- Hypothesis test: pixel on clutter vs. not on clutter
- Pixelwise measure of information for discrimination
  $D(p_i \,\|\, p_0) > \eta$
- Yields an ordering of pixels by their empirical information relative to null-hypothesis



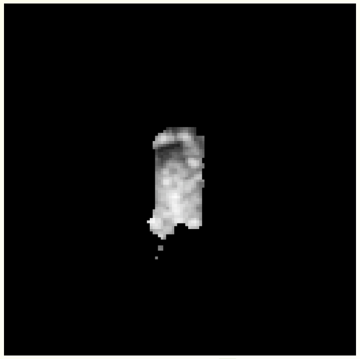
$\eta = 0$



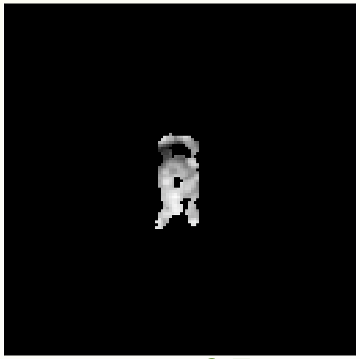
$\eta = 0.2$



$\eta = 1$



$\eta = 5$



$\eta = 25$

Codebook Representation

- Consider a sequence of approximations $p_1, p_2, \ldots$ with p_{n+1} a better approximation than p_n

$$\Pr[\text{error} \mid p_{n+1}] \leq \Pr[\text{error} \mid p_n]$$

- Let $C(p_n)$ be a measure of average resource consumption when approximation p_n is employed
- Since better approximations often involve higher complexity, we expect

$$C(p_{n+1}) \geq C(p_n)$$

- Static implementation by selecting n to satisfy constraints on $\Pr[\text{error} \mid p_n]$ and $C(p_n)$

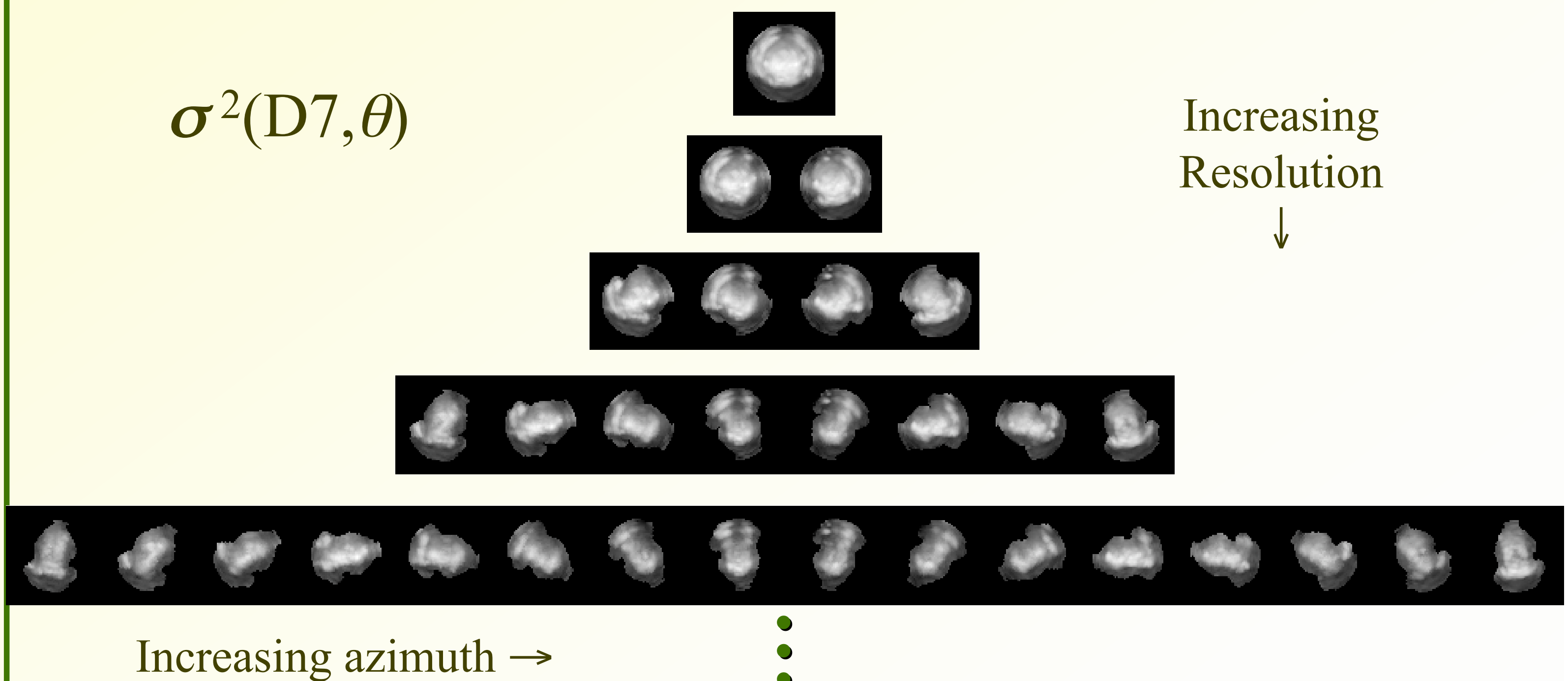
Example: Approximating Likelihoods

- Object Recognition
- Unique Features
  - Model Estimation
  - Null Hypotheses
  - Codebook Representation
- Dynamic Reconfigurability
- Other Rates
- Directions

- Model the SAR image of a bulldozer as a function of azimuth

$$r_i \sim \mathrm{CN}(0,\ \sigma_i^2(a,\theta)\ )$$

- Likelihood function depends on parameter function σ_i^2
- Sequence of piecewise constant approximations



Dynamic Reconfigurability

- Seek algorithms that dynamically adjust to fit requirements
  - can't necessarily determine n ahead of time
- Let $\Delta C(p_{n+1})$ be the additional resources consumed using p_{n+1} assuming problem with p_n already solved
- Good designs characterized by

$$C(p_{n+1}) \approx C(p_n) + \Delta C(p_{n+1})$$

- Produce a sequence of answers (a_1,θ_1), (a_2,θ_2), … with increasing accuracy and resource consumption

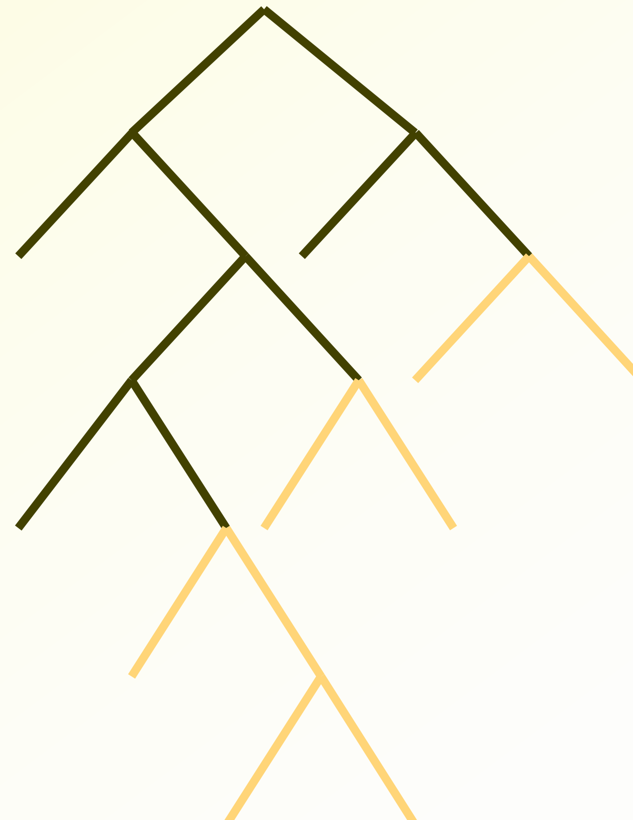
$$C'(p_{\mathrm{n}+1}) = C(p_1) + \Delta C(p_2) + \ldots + \Delta C(p_{n+1})$$

  - stop when resource allocation exhausted

Example: Delta Cost Functions

- Let cost be average number of bits read from database
- Divide azimuth into N_d non-overlapping intervals of width d

$$\tilde{\sigma}^2_{d,i}(\theta_k, a) = \frac{1}{d} \int_{\frac{2\pi k}{N_d} - \frac{d}{2}}^{\frac{2\pi k}{N_d} + \frac{d}{2}} \sigma_i^2(\theta, a)\, d\theta$$



Approximations d and $d/2$ are hierarchically related:

$$\tilde{\sigma}^2_{d,i}(\theta_k, a) = \frac{1}{2}\left[\tilde{\sigma}^2_{\frac{d}{2},i}(\theta_{2k}, a) + \tilde{\sigma}^2_{\frac{d}{2},i}(\theta_{2k+1}, a)\right]$$

Sequence Selection

Object Recognition
Unique Features
Dynamic Reconfigurability
Incremental Costs
Sequence Selection
Other Rates
Directions

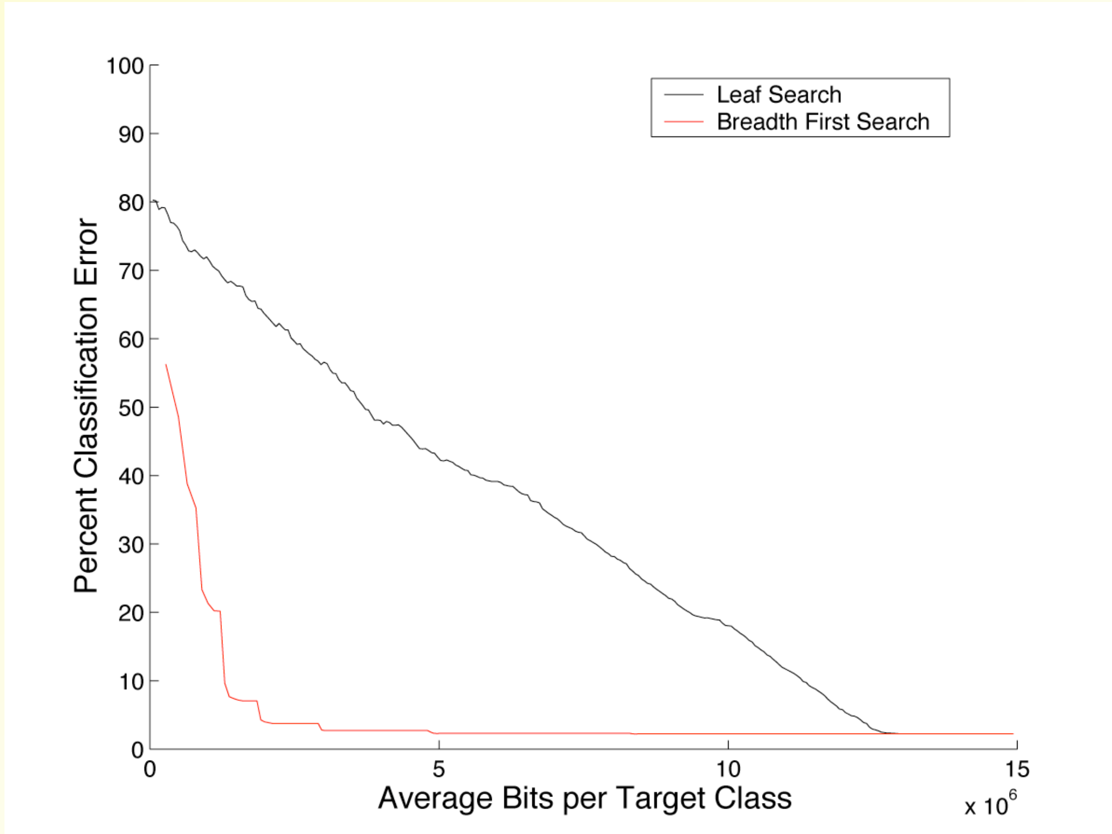
- Selection of sequence p_n drastically affects the parametric curve $\Pr[\text{error} \mid p_n]$ vs. $C'(p_n)$
- Good designs decrease error rapidly at start of sequence
  - useful results even if search is terminated early
  - can make use of additional resources if available
- Example: Error probability vs. database communication
  - Design #1: “Leaf Search”

    Refine sequential 1.4º intervals
  - Design #2: “Breadth First”

    Divide the most likely interval

Example: Search Algorithm

Error rate vs. bits transmitted from database to processor



- Classification depends on extent of search
- Eventually, search covers all possibilities
- Breadth-first search quickly finds good solutions (a, θ)
- Small overhead present with ordered searches

Other Consumption Measures

- Network bandwidth is one of many types of resources
- Other average rates of resource consumption:
  - Elapsed time per classification
  - CPU cycles per classification
  - Database (magnetic) storage per model class
  - Power dissipation
- Changing resource consumption rates due to:
  - Variation in application requirements
  - Reallocation of resources to higher priority tasks
  - Damaged or offline computation elements
  - Disrupted communication paths
  - Power considerations

Example: Throughput

Time to process through approximation p_m includes time to:

- distribute SAR image to each CPU
- process each approximation until local memory is full
- process each remaining approximation

$$T_{\text{chip}} = \frac{S_c}{\text{BW}}\left\lceil \log_2 (P+1) \right\rceil + \sum_{l=1}^{l_{\text{mem}}} 2^{l-1} N_{\text{T}} \tau_{d_l} + \sum_{l=l_{\text{mem}}+1}^{m} 2^{l-2} N_{\text{T}} \left(\tau_{d_l} + \tau'_{d_l} \right)$$

Where:

S_{c} = bits per SAR image

P = number of processors

BW = network bandwidth

N_{T} = number of target classes

τ_d = average time per template at approximation p_d exploiting hierarchy.
Receive variance, compute variance, and compute likelihoods.

τ'_d = average time per template without exploiting hierarchy.
Receive variance and compute likelihood.

Example: Throughput

Object Recognition

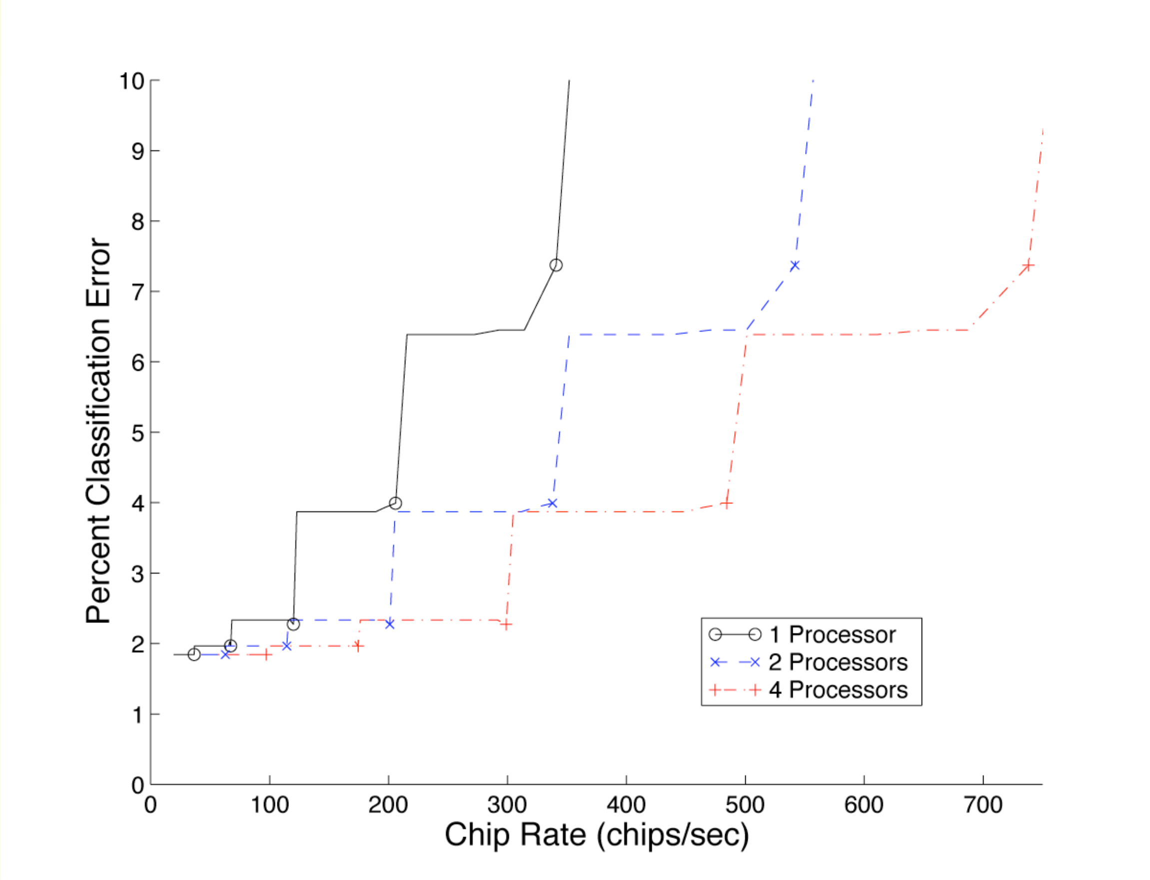
Unique Features

Dynamic Reconfigurability

Other Rates

Directions

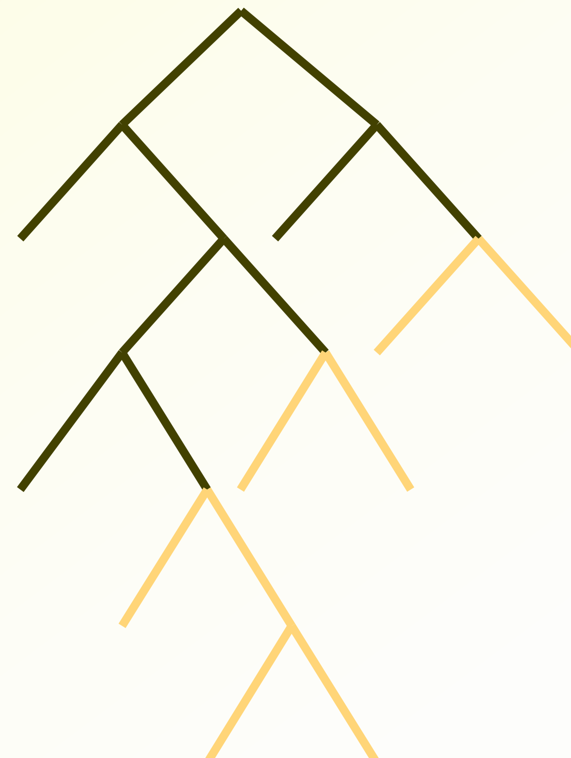
- High throughput corresponds to coarse approximations
- Markers denote doubling in # of representation intervals



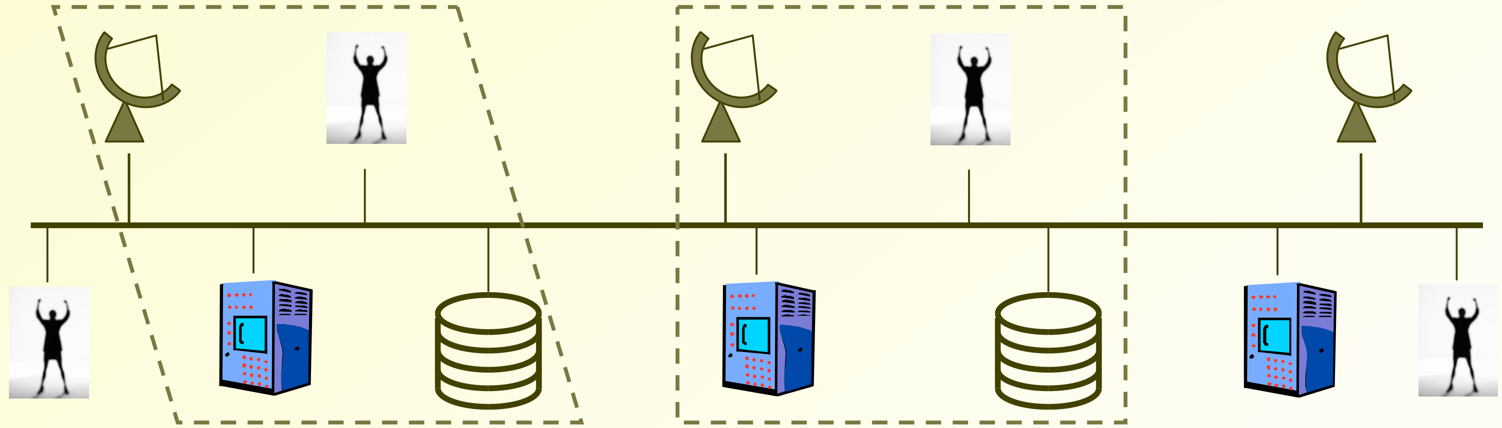
With: Chip Rate = $1/T_{chip}$ | 1 GHz clock | 64 bit read per clock cycle
4 target classes | 0.5 CPI | 25 target locations
10 Gbps interconnection

Directions: Rate-Recognition Theory

- **Theoretical development relating accuracy to cost**
- **Successively-refinable data representations**
  - **Information theory: M. Effros, Cover & Equitz, B. Rimoldi**
  - **Image processing: J. Shapiro, Said & Pearlman**
  - **Wavelet analysis: D. Donoho, I. Daubechies, et al.**
- **Successively-refinable recognition**
  - **Recognition error ↔ distortion**
  - **Log-time, log-space, etc. ↔ rate**
  - **Rate-distortion → rate-recognition**



Directions: Networked ATR Systems



Goal: Near optimal use of resources in a **dynamic** environment

- System operates within constraints
  - Accuracy, bandwidth, computational capability, throughput
- Demands and resources may change during an engagement
  - Damage, jamming, new capabilities, preemption, budgeting, etc.
- Successively refinable search algorithms to adjust operating point on the fly

Joint work with Washington University and NAWC-WD, China Lake

Issues in Networked ATR

- Dynamic bandwidth availability: Co-design of
  - Search algorithms exploiting nested model families
  - Object representations to support search algorithms
  - Data compression optimized for recognition
- Dynamic Environments
  - Models for network resource consumption
  - Achievable accuracy surfaces
  - Characterize robustness relative to varying resources
  - Find all feasible resource allocations given accuracy and resource constraints
- System Design
  - System architecture
  - Partitioning effort across distributed elements
  - Modules which can operate in concert or isolation

Directions: Shape Representation for 3+D Objects

Object class is often determined by shape

Hypothesize and test approaches require shape models

Much graphics research focused on generating photo-realistic imagery

Spiral Staircase by LightWork Design Ltd.

Issues in Shape Representation

- Representations optimized for recognition
  - Probabilistic shape models capturing random deformations
  - Probabilistic texture models capturing surface reflectance
  - "Random texture over a random shape"
  - Successively refinable representations
- Practical Uses
  - Model estimation from discrete samples
  - Recognition and inference algorithms
  - Manipulation
- Relationship to spectral estimation problems
  - Reflectance $r(\theta)$ is a random process in viewing angle
  - Estimate autocovariance function $\xi(\theta_1, \theta_2) = \Sigma\, b_{i,j}\phi_i(\theta_1)\phi_j(\theta_2)$ from samples

Conclusions

- Toward a theory for object recognition
- Build upon communication and information theories
- Incorporate resource constraints as a primary consideration
- Extensions to distributed networked systems
  - Vision Information Systems
- Object models defined in the domain of the objects themselves

"Make theory-based recognition systems a real-time reality"