Task Specific Information and Compressive Imaging

Mark A. Neifeld
University of Arizona
College of Optics and ECE Department
(and former DARPA PM 2009-2013)

OUTLINE

1. Compressive Measurement/Imaging
2. Task Specific Information Formulation
3. Information Optimal Static Measurements
4. Information Optimal Adaptive Measurements
Philosophical Underpinnings of Compressive Imaging

Images are Redundant

- This is true for all modalities and applications.
- It is true because the world is highly correlated in space, time, and wavelength.
- The world is made of objects ... not pixels.

This Should Make Us Uncomfortable

- Practically – we are spending resources on measuring redundant data.
- Theoretically – we know more than band-limited so we need to “fix” sampling theorem.
- Cybernetically – sometimes an image is never intended for human consumption.
The Weighing Problem

Sequential procedure
- P measurements needed
- SNR = SNR₀
- No inversion required

Multiplexed procedure
- P measurements needed
- SNR = (P/2) SNR₀
- Linear inversion required

Compressive procedure
- P boxes, S have non-zero weights
- O(S) measurements needed
- SNR = (S/2) SNR₀
- Nonlinear inversion required

Question 1: What does this have to do with optical imaging?
Question 2: What are the optimal combinations?
Feature-Specific Compressive Imaging

- Images are redundant – this why J2K works so well!
- Redundancy is a form of sparsity (e.g., energy compaction in DCT, wavelet, or PCA)
- Compressive imaging seeks to measure only the non-redundant parts

Conventional cameras measure a large number (N) of pixels
Compressive cameras measure a small number (M << N) features
Features are simply projections $y_i = (x \cdot f_i)$ for i=1, … M

Benefits of projective/compressive measurements include:
  - Increased measurement SNR $\Rightarrow$ improved image fidelity
  - Mode informative measurements $\Rightarrow$ reduced sensor power and bandwidth
  - Enable task-specific imager deployment $\Rightarrow$ information optimal

Sequential Compressive Imaging

- A single feature is measured in each time step (noise BW \( \sim M/T \))
- Photons collected on a single detector (measured signal \( \sim 1/M \))
- Unnecessary photons discarded in each time step (1/2)
- Reconstruction computed via post-processing

Characteristics

Parallel Compressive Imaging

- All M features are measured in a single time step (noise BW \( \sim 1/T \))
- Photons collected on M \( \ll N \) detectors (measured signal \( \sim 1/M \))
- Unnecessary photons discarded in each channel (1/2)
- Reconstruction computed via post-processing

Characteristics
Evaluate compressive imaging architectures for 16x16 block-wise feature extraction
All other conditions remain unchanged

Reconstruction RMSE versus Noise/BW

- Block-wise operation shifts crossover to smaller noise/bw
- RMSE Trend 1: Pipeline < Parallel < Sequential
- RMSE Trend 2: Compressive < Conventional for high noise/bw

σ₀ = 23.7
75% fewer detector measurements
Images are redundant – this is why image compression is so effective
Redundancy can be viewed as sparsity in some transform domain (L of N nonzero coefficients)

Nyquist was overly pessimistic. Stronger prior knowledge than band-limited is generally available.

How can measurements on f leverage this knowledge?
Traditional measurement requires O(N) samples.
Donoho proved that compressive measurements using random projections will require only O(L)

Reconstruct by solving an inference problem – what is the best estimate of x given (a) what I’ve measured, (b) the requirement for sparsity, and (c) anything else I know about x (e.g., positivity, prior distribution, …)?

$$x = \text{argmin} \{ ||x||_1 + \lambda ||g - Kw||^2 + \gamma P(x) \}$$

Sparsity Data Agreement Regularizer

**Nyquist was overly pessimistic. Stronger prior knowledge than band-limited is generally available**

Measurement cost now scales with information dimension not signal dimension
Examples with which you are already familiar:

- Sampling a band-limited waveform
- Sparse vector space
- Two-class decision problem: 1D sufficient statistic

- Natural images are redundant … this is why conventional compression works.
- More efficient measurement can be performed in correct basis.
- Random basis can be almost as good.
- Fundamental optical operation is projection … inner-product calculation.
Information content source requires probability density $\rho(r)$

Shannon Entropy: $J = -\int \rho(r) \log \rho(r) d^n r$

PROBLEM: In general $\rho(r)$ is very complex/unknown and high-dimensional
Task Specific Information Concept

- Information content is *task specific*.

  **Detection task:**
  - Probability of presence/absence = $\frac{1}{2}$
  - Information content < 1 bit

  **Detection and Localization task:**
  - Probability of tank being absent = $\frac{1}{2}$
  - Probability of occurrence in each region = $\frac{1}{8}$
  - Information content < 2 bits

  **Classification task:**
  - Probability of each target the = $\frac{1}{2}$
  - Information content < 1 bit

- How do we *quantify* the *task specific information* (TSI)?
Task Specific Source Encoding

Virtual source $\rightarrow$ Encoding $\rightarrow$ C(X)

$C(X)$ stochastically encodes $X$ and produces scene $Y$

- Detection task: *presence/absence of target is of interest*
  - Virtual source variable $X$ must be binary.
  - $X = 1$ or 0 implies tank present or absent.

$X = 1$ (Tank present)  \hspace{1cm}  X = 0$ (Tank absent)
Task Specific Information Definition

Imaging chain block diagram

Virtual source \( \xrightarrow{X} \) Encoding \( \xrightarrow{Y = C(X)} \) Channel \( \xrightarrow{Z = H(Y)} \) Noise \( \xrightarrow{R} \)

- Imager is characterized by channel \( H \) and noise \( n \)
- Definition for Task Specific Information:

\[ TSI \equiv I(X; R) \leq J(X) \]

Mutual information between \( X \) and \( R \)

Always bounded by the entropy of \( X \)

Entropy \( J(X) \rightarrow \) maximum task specific information content
Computing Task Specific Information

- Measurement can be written as,

\[ R = H[C(X)] + n \]

zero-mean additive Gaussian noise

- Computing TSI is difficult for non-Gaussian source

- Use Verdu’s relation between mutual information and minimum mean square (mmse) estimation error

\[ TSI = I(X; R, s) = \int_0^s mmse(s')ds' \]

where \( mmse = \text{Trace}\left[H^T \Sigma_n^{-1} H(E_Y - E_{Y|X})\right] \), \( Y = C(X) \)

\[ E_Y = E\left[(Y - E(Y|R))(Y - E(Y|R))^T\right], \quad \text{(MMSE conditioned over } R) \]

\[ E_{Y|X} = E\left[(Y - E(Y|R,X))(Y - E(Y|R,X))^T\right]. \quad \text{(MMSE conditioned over } R \text{ and } X) \]
Some Encouraging Results

- Binary detection problem (tank present/absent) $\rightarrow$ 1 bit of TSI.
- Include positivity and energy conservation constraints.
- Compare various known measurement matrices with conventional imager.
- TSI optimization only over photon allocation (e.g., dwell time).

Even sub-optimal compressive measurement can provide 14.7x SNR improvement relative to conventional imaging.
Recent TSI Optimal Extension

- Binary detection problem (tank present/absent) $\rightarrow$ 1 bit of TSI.
- Include positivity and energy conservation constraints.
- Compare various known measurement matrices with conventional imager.
- TSI optimization yields fully optimal measurement vectors.

Information optimal static compressive measurement can provide 27.5x SNR improvement (i.e., to achieve $P_e=0.01$) relative to conventional imaging.
Conventional Imager (CONV) for recognition

- $T_{int} \sim$ Integration time per measurement
- $n \sim N(0, \sigma^2 I)$, $P \sim N \times 1$, $r \sim$ scalar
- $\Pr(H_0)$ and $\Pr(H_1)$ are respective priors for $G_0$ and $G_1$
- Define SNR as $10 \log \frac{T_{int}}{\sigma^2}$

**Bayesian paradigm:** For given priors minimize probability of error $P_e$

\[
T(r) = \frac{p(r \mid H_1)}{p(r \mid H_0)} \cdot \frac{\Pr(H_0)}{\Pr(H_1)}
\]

\[
P_e = \Pr(H_0 \mid H_1) \Pr(H_1) + \Pr(H_1 \mid H_0) \Pr(H_0)
\]

- Optimal projection vector for post-processing a conventional measurement

We would typically examine $P_e$ versus SNR … today we’ll look at something a bit different
1. Fix $T_{int}$ and $\sigma^2$ and consider K measurements

2. Fix $\sigma^2$ and $P_e$ and find required K for given $T_{int}$

3. Extend to M hypotheses (ad hoc but popular)

- **M hypotheses**: $H_i : r = P^T(T_{int}G_i + n)$
- $P \sim N \times L$

**Observations**

1. K decreases monotonically with $T_{int}$
2. Performance saturates at $L = M - 1$
3. Note: no cost for larger L

**Converges to K=1 for large $T_{int}$**

**Weighted between-class scatter matrix**

**Priors act as weights**

---

**M=4 Class Example**

**$P_e=10^{-2}$ $\sigma^2 = 10$**

---

**NEIFELD – UCLA 2015**
Compressive imaging measures linear projections of the scene irradiance optically.

- \( K \): Number of \( L \)-dimensional measurements
- Find \( K \) required to achieve \( P_e \) under photon count constraint -

\[
\max \left\{ \sum_{k=1}^{L} |P_{kj}|^2; j = 1 \ldots N \right\} = T_{int}
\]

\((M = 4, P_e = 10^{-2}, \sigma^2 = 10)\)

Observations
1. \( K \) decreases monotonically with \( T_{int} \)
2. Static CONV improves with \( L \) saturating at \( L = M - 1 \)
3. Static FSI is superior to static CONV for the same \( L < 4 \)
4. Static FSI rolls-over at \( L = 4 \) due to noise-cost \( \rightarrow \)

\( L=4 \) gains no new discriminating information but spends photons

Measurements consume finite resources
Sequential Hypothesis testing (SHT)

- At any stage of the experiment chooses one of these:
  1. **Accept** the hypothesis being tested
  2. **Reject** the hypothesis being tested
  3. **Continue** the experiment – make another measurement

Both FSI and CONV can exploit adaptation

Decide using SHT

After $k-1$ measurements, $P_k$ \leftrightarrow L dominant eigenvectors of $R_{\lambda_1}$

\[
R_{\lambda_1} = \sum_{j=1 \ldots M} \sum_{i=1 \ldots M} \Pr(H_j | r_i) \Pr(H_j | r_i) (G_i - G_j)(G_i - G_j)^T
\]

New projections based on modified priors

\[
Pr(H_j | r_k) = \frac{\Pr(r_k | H_j) \Pr(H_j | r_{k-1})}{\sum_{i=1 \ldots M} \Pr(r_k | H_i) \Pr(H_i | r_{k-1})}
\]
Adaptation Example – Evolution of Priors

- $L = 1, \ P_e = 10^{-2}, \ T_{\text{int}} = 0.1, \ \sigma^2 = 10$ (i.e., SNR = -20 dB)

- Face 2 was originally chosen

- In this case correct decision at $K = 11$ iterations (note that $K$ is random)
Adaptive Conventional versus Compressive Imaging

Observations

1. AFSI superior to ACONV
2. At Low $T_{int}$ AFSI with $L = 1$ performs best; noise-cost increases with $L$
3. At $T_{int} = 0.1$, AFSI requires 30 times less measurements than ACONV
4. Gain converges to unity at high $T_{int}$
Observations:
1. Adaptive compressive measurement offers opportunity for rapid decision-making.
2. Notice tradeoff with $T_{int}$ – indicates tension between measurement SNR and adaptation advantage.
3. TSI optimal adaptation at $T_{int} = 0.1$ offers a detection time advantage of $> 250x$. 

Average Detection Time $D = T_{int} \times E\{K\}$ 

- Conventional (ACONV)
- Eigenvectors (ACOMP)
- Adaptive Information Optimal (greedy) 

$(M=4, P_e=0.01, \sigma^2=10)$
Conclusions

1. Compressive measurement enables (nearly) arbitrary projections of the object space.
   - Weighing problem + image redundancy explains why this could make sense.
   - Straightforward implementation in optical domain.
   - Also potential implementations in spectroscopy, x-ray, RF, THz, ...

2. Projections may be optimized to remove redundancy (e.g., image reconstruction).

3. Projections may be optimized to maximize task performance (e.g., target detection).

4. Task specific information can be used to optimize projections.

5. Dramatic performance improvements are possible via information-optimal measurement.
   - Static projections with 30x SNR advantage relative to conventional imaging.
   - Adaptive solutions with 250x reduced time-to-detection relative to conventional imaging.