Adaptation for Task-Specific Compressive Imaging

Mark A. Neifeld
Electrical and Computer Engineering
College of Optical Sciences
University of Arizona

OUTLINE

1. Introduction
2. Statistical Adaptation
3. Information-Optimal Adaptation
1. Why are images typically measured as collections of pixels?
   • History = Original consumers were humans who wanted pretty pictures.
   • Physics = Image-formation using glass (or a pinhole) is “straightforward.”
   • Technology = Previous lack of electronic detection/post-processing.
   • Mind Set = If I can’t “see it” then it isn’t there.

2. What is the first thing we do after we measure 10Mpixel image?
   • Use compression to throw away the redundant parts.
   • Maybe we can push some compression into the measurement domain.
A Few Motivational Observations

1. Why are images typically measured as collections of pixels?
   - History = Original consumers were humans who wanted pretty pictures.
   - Physics = Image-formation using glass (or a pinhole) is “straightforward.”
   - Technology = Previous lack of electronic detection/post-processing.
   - Mind Set = If I can’t “see it” then it isn’t there.

2. What is the first thing we do after we measure 10Mpixel image?
   - Use compression to throw away the redundant parts.
   - Maybe we can push some compression into the measurement domain.
A Few Motivational Observations

1. Why are images typically measured as collections of pixels?
   • History = Original consumers were humans who wanted pretty pictures.
   • Physics = Image-formation using glass (or a pinhole) is straightforward.
   • Technology = Previous lack of electronic detection/post-processing.
   • Mind Set = If I can’t “see it” then it isn’t there.

2. What is the first thing we do after we measure 10Mpixel image?
   • Use compression to throw away the redundant parts.
   • Maybe we can push some compression into the measurement domain.

3. What might we gain by reducing the number of measurements?
   • Lower compressive post-processing requirements.
   • Lower size/weight/power requirements.
   • Lower bandwidth and/or storage requirements.
   • Increased measurement SNR.
   • Increased (task-specific) information transfer.

4. When I don’t have a choice I will accept a projective measurement (e.g., CT, MRI, ultrasound, …)
5. Well … you don’t have a choice!
Conventional Imaging is Already Compressive

Consider conventional b/w camera:

\[ g \sim f \rightarrow g = S \{ \left| \text{Fr} \left( E \exp[\text{i} QP] ; S_i \right) \right|^2 \} = H[f] \]

where:
- **Object**: \( f \in \mathbb{R}^+ \)
- **Field**: \( E \in \mathbb{C} \)
- **Measure**: \( g = H[f] \in \mathbb{R}^+ \)

**Conventional cameras** measure a visually pleasing representation of the object.

**Observations:**
- Compressive measurement \( \Rightarrow \) dimensionality reducing measurement
- Conventional camera integrates over pixel support (continuous object \( \Rightarrow \) discrete measurement)
- Conventional camera integrates over wavelength (continuous object \( \Rightarrow \) point measurement)
- Conventional camera flattens object from 3D onto 2D sensor

**Question**: So what’s all the hype about compressive imaging?
**Answer1**: We’re starting to think about *designing* alternate projections.
**Answer2**: Recent advances in *nonlinear* reconstruction.
**Computational Imaging and Alternate Projections**

**Alternate Projection**

**Overall System:** $\rightarrow \quad a = A[g] = A[H[f]] \sim f \quad \ldots$ we may still want a pretty picture

$\sim d \quad \ldots$ or we may want something else (e.g., decision)

**Implications:** Joint optimization of optical and post-processing degrees of freedom (MDO)

**Benefits:** Low implementation cost, flexible form factor, improved capabilities (e.g., reduced aberrations, extended depth of focus, ...) non-traditional metrics (e.g., information theoretic, detection theoretic, ...), novel/powerful optical components, ...

**Shoulders of Giants:** Goodman (1971), Dowski and Cathy (1995), George (2001), Brady (2002), Prasad (2003), ...

**Question:** What are the best alternate projections? (random is one example)

**Answer:** It depends upon your application (goals, resources, noise, ...)

**Surprise Answer:** Even pretty pictures can benefit from alternate projections!
Two-Class Detection Problem

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 $P_e$

$$T(r) = \frac{p(r|H_1)}{p(r|H_0)} \cdot \frac{\Pr(H_0)}{\Pr(H_1)};$$

$$P_e = \Pr(H_0 | H_1) \Pr(H_1) + \Pr(H_1 | H_0) \Pr(H_0)$$

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
**Required Number of Measurements**

1. Fix $T_{\text{int}}$ and $\sigma^2$ and consider $K$ measurements

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

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

\( M \) hypotheses:
\[
H_i : r = P^T(T_{\text{int}}G_i + n)
\]
\[
P \sim N \times L
\]

**Observations**

1. $K$ decreases monotonically with $T_{\text{int}}$
2. Performance saturates at $L = M - 1$
3. Note: no cost for larger $L$
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
  \[ \max \left\{ \sum_{k=1}^{L} \left| P_{kj} \right| ; j = 1 \ldots N \right\} = T_{int} \]

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

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 →

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

\[
FSI \Leftrightarrow H_i : r_k = P_k^T G_i + n
\]
\[
CONV \Leftrightarrow H_i : r_k = P_k^T (G_i + n)
\]

Decide using SHT

\[
G_i
\]

Update the priors and \( P_k \) to \( P_{k+1} \)

\( i^{th} \) target

New projections based on modified priors

After \( k - 1 \) measurements,
\[
P_k \leftrightarrow L \text{ dominant eigenvectors of } R_{\Delta_k}
\]
\[
R_{\Delta_k} = \sum_{j=1}^{M} \sum_{i=M+1}^{M+M} \Pr(H_i | r_k) \Pr(H_j | r_k) (G_i - G_j)(G_i - G_j)^T
\]

\[
Pr(H_i | r_k) = \frac{\Pr(r_k | H_i) \Pr(H_i | r_{k-1})}{\sum_{l=1}^{M} \Pr(r_k | H_l) \Pr(H_l | r_{k-1})}
\]

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

Recall that $\text{SNR} = 10 \log \frac{T_{\text{int}}}{\sigma^2}$

• Face 2 was originally chosen

• In this case correct decision at $K = 11$ iterations (note that $K$ is random)
Observations

1. AFSI outperforms SFSI; Improvement because of adaptation capability.

2. At Low $T_{int}$ AFSI with $L = 1$ performs best; noise-cost increases with $L$.

3. At $T_{int} = 0.1$, AFSI needs 5 times less measurements than SFSI.
Adaptive Conventional versus Compressive Imaging

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

**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 \( T_{int} \)
Average Time to Detection

\[ D = T_{int} \times E[K] \]

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

**Observations**

1. Higher \( T_{int} \) does not ensure faster detection.

2. There will exist an optimal value of \( T_{int} \) which ensures smallest average detection time.
What if Class Conditional Densities are Not Known?

- Assumption: Training object examples within each hypothesis are available

Different perspectives

- Object dimension: 32×32
- Number of training images in each class = 449
- Number of test images for each class = 50

- Test objects similar to training objects, however not contained in the training set
Class Conditional Density Estimation

- Parametric approach (P-AFSI): Assumes multi-variate Gaussian function for representing the class-conditional density
  - Requires the knowledge of mean $\mu_i$ and co-variance $\Sigma_i$, where $i = 1 \ldots M$
  - Mean and co-variance can be computed using the training data

$$p(r^{(k)} | H_i) = \frac{\exp(-[r^{(k)} - \mu_{i,k}]^T \Sigma_{i,k}^{-1} [r^{(k)} - \mu_{i,k}] / 2)}{(2\pi)^{kL/2} (\det(\Sigma_{i,k}))^{1/2}}$$

where $\mu_{i,k} = (P^{(k)})^T \mu_i$ and $\Sigma_{i,k} = (P^{(k)})^T \Sigma_i (P^{(k)}) + 2\sigma^2[I]_{kL}$

- Semi-nonparametric (NP-AFSI) approach uses:
  - Measurement history $r^{(k)}$ to find $n$ nearest neighbors, for $G$, within each hypothesis
  - Denote the set of $n$ neighbors for $G$ in the $i$th hypothesis as $R_i^{(k)}$
  - Use the elements from the set $R_i^{(k)}$ to form $p(r^{(k)} | H_i)$

$$p(r^{(k)} | H_i) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{(2\pi \cdot 2\sigma^2)^{kL/2}} \exp\left(-\frac{\| (P^{(k)})^T \tilde{G}_j - r^{(k)} \|^2_{L_2}}{4\sigma^2} \right)$$
NP-FSI versus P-AFSI

Observations:

1. At low $T_{int}$, NP-AFSI and P-AFSI with $L = 1$ perform best; noise-cost increases with $L$.

2. At $T_{int} = 1$ (i.e. SNR = -10 dB), P-AFSI requires $E[K] = 19, 27$ and $57$ for $L = 1, 2$ and $3$ respectively.

3. At $T_{int} = 1$ (i.e. SNR = -10 dB), NP-AFSI requires $E[K] = 12, 23$ and $38$ for $L = 1, 2$ and $3$ respectively.

Comparing with $E[K] \sim 2$ for AWGN case.

$M = 4, n = 201, P_e = 10^{-2}, \sigma^2 = 10$
Information-Optimal Adaptation
C(X) stochastically encodes X to produce scene Y

Example: for a detection task the virtual source variable X must be binary.

Imager is characterized by channel $H$ and noise $n$

Definition for Task Specific Information:

$$TSI \equiv I(X; R) \leq J(X)$$
Computing Task-Specific Information

- Measurement can be written as,

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

Channel operates on coding of virtual source

- Computing TSI is difficult for non-Gaussian source

- Use Verdu’s relation between mutual information and \textit{mmse} estimation error

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

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

\[
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) \)
TSI can be used to analyze and compare imagers and projections.

- TSI metric measures *functional content* of measured signals.
- Working to apply TSI metric within optimization loop.
Information Theoretic Adaptation

Measurement model at $k^{th}$ step (when class-conditional objects are known):

$$\mathbf{r}_k = \mathbf{P}_k^T \mathbf{G} \mathbf{x} + \mathbf{n}$$

where $\mathbf{r}_k \sim L \times 1$, $\mathbf{P}_k \sim N \times L$, $\mathbf{n} \sim N(0, 2\sigma^2 \mathbf{I}_L)$

- $\mathbf{P}_k$ is the projection matrix at the $k^{th}$ step
- $\mathbf{G}$ is the object matrix
- $\mathbf{x}$ is a random indicator vector

Recognition TSI is given by:

$$J_k = I(\mathbf{r}^{(k)}; \mathbf{x})$$

where $I(\mathbf{r}^{(k)}; \mathbf{x})$ is the mutual-information between $\mathbf{r}^{(k)}$ and $\mathbf{x}$, and

$$\mathbf{r}^{(k)} = [\mathbf{r}_1^T, ..., \mathbf{r}_{k-1}^T, \mathbf{r}_k^T]^T = [(\mathbf{r}^{(k-1)})^T, \mathbf{r}_k^T]^T$$

$$\mathbf{P}^{(k)} = [\mathbf{P}_1, ..., \mathbf{P}_k]$$

Goal is to maximize $J_k$ at each step with respect to $\mathbf{P}_k$

$$J_k = I(\mathbf{r}^{(k)}; \mathbf{x}) = I(\mathbf{r}^{(k-1)}; \mathbf{x}) + I(\mathbf{r}_k; \mathbf{x} | \mathbf{r}^{(k-1)})$$

$$= J_{k-1} + I(\mathbf{r}_k; \mathbf{x} | \mathbf{r}^{(k-1)})$$

Conditional information $J_{k|k-1} = I(\mathbf{r}_k; \mathbf{x} | \mathbf{r}^{(k-1)})$

- This implies that the projection basis $\mathbf{P}_k$ must be designed such that $J_{k|k-1}$ is maximized
- Analytic computation of $J_{k|k-1}$ is mathematically intractable
  - Numerical optimization of $J_{k|k-1}$ is required
Optimization of Conditional Information

Problem formulation: At $k^{th}$ step

$$\max_{P_k} [J_{k|k-1}], \text{ such that } \max_l \left\{ \sum_{j=1}^{L} |P_{k,l,j}| ; l = 1...N \right\} = T_{\text{int}}.$$  

**Algorithm for maximizing $J_{k|k-1}$:**

1. At the iteration index $i = 0$, initialize the projection basis $P_k = P(0)$ satisfying the photon constraint

2. Update the projection basis after updating the gradient $Z(i) = \frac{\partial J_{k|k-1}}{\partial P_k}$

   $$P_k(i + 1) = P_k(i) + \mu Z(i).$$

3. Normalize $P_k(i + 1)$ to satisfy the photon constraint

4. Set $i \leftarrow i + 1$ and go to step 2 until $\sum \sum |P_k(i + 1) - P_k(i)| < \varepsilon$.

Note: We will refer to this approach as CI-AFSI.
CI-AFSI versus Statistical AFSI

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

Observations

1. CI-AFSI outperforms AFSI
2. CI-AFSI suffers with no noise-cost with increasing \(L\)
3. CI-AFSI works better for all \(T_{\text{int}}\); Faster convergence to \(E[K] = 1\)
4. Scatter-matrix approximation (Popt) works good at low-values of \(T_{\text{int}}\)
Observations

1. Higher $T_{int}$ does not ensure faster detection

2. CI–AFSI achieves minimum $D$ at $T_{int} = 0.1$ compared to $T_{int} = 0.2$ for AFSI
Conclusions

1. All imagers measure linear projections of the object space.
2. Compressive imaging (FSI) enables the *design* of a custom projection basis.
3. Task-specific design can offer substantial performance benefits.
4. Adaptation exploits results of previous measurements to define current projection.
5. Adaptive FSI strives to make optimal use of every photon.
   - Statistical AFSI
   - Information optimal AFSI