DATA COMPRESSION FOR SIMULTANEOUS/SEQUENTIAL INFERENCE TASKS IN SENSOR NETWORKS

Mo Chen and Mark L. Fowler

Department of Electrical and Computer Engineering
State University of New York at Binghamton
{mchen0, mfowler}@Binghamton.edu

Andrew Noga

Air Force Research Lab
Rome, NY
Andrew.Noga@rl.af.mil

ABSTRACT

Sensor networks typically perform multiple inference tasks and compression is often used to aid in the sharing of data. Compression degrades the inference accuracy and should be optimized for the tasks at hand. Unfortunately, simultaneous optimization for multiple tasks is not generally possible - typically a fundamental trade-off exists that has not been previously explored. A particularly relevant and interesting scenario occurs with a task-driven sequence of inferences. This paper develops a framework for data-optimized data compression for the case of multiple inferences. In particular, the Fisher information matrix (FIM) is used to derive a suitable scalar distortion measure for multiple estimation tasks, while the Chernoff distance is used for decision tasks. Theoretical results are presented that support the use of this particular scalar FIM-based distortion. The method is demonstrated with application to the sequential problem of first detecting a common intercepted signal among sensors and then once detected progressing to the location of the source.

1. INTRODUCTION

It is important to design operational data compression methods that enable rapid sharing while causing only minimal degradation of the quality of the inferences to be made by the sensor network.

To design optimal decentralized compression algorithms that support inference tasks it is essential to have appropriate, useable metrics that measure the impact of rate reduction on the inference quality. In this paper, the Fisher information matrix (FIM) is used to assess the impact on estimation accuracy while the Chernoff distance is used to assess the impact on decision accuracy. The advantages of these distortion measures include: (i) the reciprocal of the FIM yields the Cramer-Rao Lower Bound on the variance of any unbiased parameters estimator and the Chernoff distance is a tight upper bound on probability of detection, (ii) both are independent of the specific choice of inference processing [1], (iii) both are additive for independent observations (which is very important because the network-wide optimal compression can be achieved in a decentralized way), (iv) they are invariant under application of invertible maps to the data and are decreased under application of quantization, which makes possible transform coding. Although these measures have been used for single inference tasks [2],[3], an important aspect not widely previously considered is that sensor systems generally have multiple inference tasks that have conflicting compression requirements [4],[5]. This paper explores how multiple measures (for the multiple inferences) drive the specification of new data compression methods needed to support multiple sequential and simultaneous inferences; we focus on the problem of emitter location using Time-Difference-Of-Arrival (TDOA) and the Frequency-Difference-Of-Arrival (FDOA).

In Section 2, we address how to map the FIM into various scalar distortion measures for simultaneous multi-parameter estimation. Then we address simultaneous estimation and detection. In Section 3, we propose and explore a novel task-embedded compression approach for sequential inference tasks.

2. SIMULTANEOUS INFERENCE TASKS

2.1 Transform Coding Framework

Suppose we have K sensor nodes that compress their received signal data to send to a fusion center that estimates an unknown parameter vector $\boldsymbol{\theta}$ and decides the presence of a common signal of interest at the nodes. At node S_k we model the data vectors as

$$\mathbf{x}_k = \gamma_k \mathbf{s}_k(\boldsymbol{\theta}) + \mathbf{w}_k , \tag{1}$$

where $\mathbf{s}_k(\boldsymbol{\theta})$ is an unknown deterministic signal vector dependent on the unknown deterministic parameter vector $\boldsymbol{\theta}$; γ_k represents a hypothesis function, that is, if $\mathbf{s}_k(\boldsymbol{\theta})$ is present, $\gamma_k \neq 0$; each $\mathbf{w}_k$ is an independent noise vector whose covariance matrix $\boldsymbol{\Sigma}_k$ is known or estimated. In this paper, for simplicity, we assume that $\mathbf{w}_k$ is zero-mean Gaussian noise. We orthonormal (ON) transform $\mathbf{x}_k$ to $\boldsymbol{\chi}_k$ (the coefficients $\{\chi_n\}_{n=1}^{N}$) with a unitary matrix $\boldsymbol{\Phi}$. The coefficients χ_n are then quantized using a set of bit allocations B = $\{b_n \mid n$=1,2,…, $N\}$. The compressed coefficient vector $\hat{\boldsymbol{\chi}}_k$ is

$$\begin{aligned}\hat{\boldsymbol{\chi}}_k &= \boldsymbol{\Phi}\mathbf{x}_k + \boldsymbol{\varepsilon}_k \\ &= \gamma_k \boldsymbol{\xi}_k(\boldsymbol{\theta}) + \boldsymbol{\omega}_k + \boldsymbol{\varepsilon}_k\end{aligned} , \tag{2}$$

where vector $\boldsymbol{\xi}_k(\boldsymbol{\theta}) = \boldsymbol{\Phi} s_k(\boldsymbol{\theta})$ holds the signal coefficients $\xi_{k,n}(\boldsymbol{\theta})$, vector $\boldsymbol{\omega}_k$ holds the noise coefficients $\omega_{k,n}$, and $\boldsymbol{\varepsilon}_k$ is the quantization noise vector with covariance $\mathbf{Q}_k = \mathrm{diag}\left\{ q_{k,1}^2, \ldots, q_{k,N}^2 \right\}$. We seek to allocate B to optimize an estimation/detection-centric distortion measure subject to rate constraint $\sum_n b_n \leq R$. The $\{q_{k,n}^2\}$ depend on b_n and can be calculated either by a closed form or an approximation [5].

2.2 Compression for simultaneous estimates

In this section, we assume that all sensors intercept the same signal ($\gamma_k = 1$) and address the compression to achieve trade-offs among the multiple unknown parameters $\boldsymbol{\theta}$. Without loss of generality, we focus on the two parameter case.

Whenever the sensor noises $\mathbf{w}_k$ are independent, the FIMs are additive so we only need to consider the effect of compression on the FIM of the data at a single sensor S_k. The FIM for $\boldsymbol{\theta}$ based on the data from sensor S_k is the 2×2 matrix

$$\mathbf{J}_k = 2\,\mathrm{Re}\{\mathbf{G}_k^H \Sigma_k^{-1} \mathbf{G}_k\}\,, \tag{3}$$

where $\mathbf{G}_k = \left[\partial \xi_k(\boldsymbol{\theta})/\partial\theta_1 \; \partial\xi_k(\boldsymbol{\theta})/\partial\theta_2\right]$ is an N×2 matrix of the signal's sensitivities to the parameters. The FIM specifies an information ellipse via $\boldsymbol{\theta}^T \mathbf{J}^{-1} \boldsymbol{\theta} = 1$ with semi-axes along the FIM's eigenvectors and whose lengths are proportional to the square roots of the eigenvalues – the larger this ellipse the better. Lossy compression of the transformed vector $\chi_k(\boldsymbol{\theta})$ changes the FIM, making the data inferior for estimation of $\boldsymbol{\theta}$ and thus shrinking and (perhaps) rotating the information ellipse. Under the model in (2) we have that the FIM after compression is

$$\hat{\mathbf{J}}_k = 2\,\mathrm{Re}\{\mathbf{G}_k^H (\boldsymbol{\Sigma}_k + \mathbf{Q}_k)^{-1} \mathbf{G}_k\}\,. \tag{4}$$

Note when we evaluate $\mathbf{G}_k$ in (4) and elsewhere, we have to replace the unknown $\xi_k(\boldsymbol{\theta})$ with the observed $\boldsymbol{\chi}_k$.

2.2.1. *Determinant of FIM*

The area of the FIM ellipsoid is proportional to $\sqrt{\lambda_1 \lambda_2}$, where λ_i is the i^{th} eigenvalue of the FIM. Thus, maximizing the area is equivalent to maximizing det($\mathbf{J}$), which has been used as an objective function in the solution of various engineering problems. It is well known that additive distortion yields simpler operational rate-distortion optimization schemes [6],[7], whereas the determinant provides a multiplicative distortion function. However, under fine quantization, we develop an additive distortion that allows simpler maximization of the determinant.

For convenience, consider that the noise is an i.i.d. process with variance σ_k^2 , i.e., $\boldsymbol{\Sigma}_k = \sigma_k^2 \mathbf{I}$ and that the signal data is real valued. The determinant of the post-compression FIM becomes

$$\det(\hat{\mathbf{J}}_k) = \frac{1}{\sigma_k^4} \det(\mathbf{G}_k^T \mathbf{G}_k) \det(\mathbf{I} - \mathbf{R}_k)\,, \tag{5}$$

where $\mathbf{R}_k = (\mathbf{G}_k^T \mathbf{G}_k)^{-1} \mathbf{G}_k^T \,\mathrm{diag}\{\ldots, q_{k,n}^2 / (\sigma_k^2 + q_{k,n}^2), \ldots\} \mathbf{G}_k)$.

Therefore, maximizing $\det(\hat{\mathbf{J}}_k)$ is equivalent to maximizing $\det(\mathbf{I} - \mathbf{R}_k)$. If $\{e_1, e_2\}$ are the eigenvalues of $\mathbf{R}_k$, the latter product part of (5) becomes

$$\det(\mathbf{I} - \mathbf{R}_k) = 1 - \mathrm{tr}(\mathbf{R}_k) + e_1 e_2\,. \tag{6}$$

Proposition (see [5] for proof): *Under the fine quantization condition, maximizing* $\det(\hat{\mathbf{J}}_k)$ *is equivalent to minimizing* tr($\mathbf{R}_k$).

We can further simplify tr($\mathbf{R}_k$) as follows:

$$\mathrm{tr}(\mathbf{R}_k) = \sum\nolimits_{n=0}^{N-1} \frac{v_i q_{k,n}^2}{\sigma_k^2 + q_{k,n}^2} \tag{7}$$

where v_n is the n^{th} element of the vector $\mathbf{v}$ given by $\mathbf{v} = \mathrm{diag}\left(\mathbf{G}_k \left(\mathbf{G}_k^T \mathbf{G}_k\right)^{-1} \mathbf{G}_k^T\right)$. Thus, a compression algorithm for approximately maximizing the determinant of the FIM can be formulated as

$$\min_B \left\{ \sum\nolimits_{n=0}^{N-1} \frac{v_i q_{k,n}^2}{\sigma_k^2 + q_{k,n}^2} \right\} \text{ subject to } \sum_{n=1}^{N} b_i \le R\,. \tag{8}$$

We have seen that this works well even for coarse quantization.

Sometimes, one may wish to favor the accuracy of one parameter at the expense of the others. This allows user-imposed trade-offs between parameters, which is important in multiple parameter estimation problems where a user may favor accuracy on a subset of the parameters. The determinant can not fulfill this need [5], but the following perimeter approach can.

2.2.2. *Weighted trace of FIM*

The perimeter of the ellipse is quite complicated to compute exactly; however, it is approximately proportional to $\sqrt{\lambda_1 + \lambda_2}$. Thus, maximizing the perimeter is (approximately) equivalent to maximizing $\mathrm{tr}(\hat{\mathbf{J}}_k)$. Note that this allows importance-weighting on the accuracy of the two parameters: we can use a *weighted* trace $\mathrm{wtr}(\hat{\mathbf{J}}_k) = \alpha \hat{J}_{11} + (1-\alpha)\hat{J}_{22}$, where α is an importance-controlling parameter satisfying $0 \le \alpha \le 1$. However, a concern is the effect that compression can have on the tilt of the ellipse, which is not explicitly captured by these trace-based measures. This is of most concern when the ellipse is highly eccentric. The following theorem shows that this is not a serious issue because the post-compression ellipse will always reside inside the original ellipse; thus, for a highly eccentric original ellipse, compression is not able to greatly change the orientation of the ellipse.

Theorem (see [5] for proof): *The post-compression FIM ellipse* $\boldsymbol{\theta}^T \hat{\mathbf{J}}^{-1} \boldsymbol{\theta} = 1$ *lies inside the original FIM ellipse* $\boldsymbol{\theta}^T \mathbf{J}^{-1} \boldsymbol{\theta} = 1$.

In summary, the perimeter approach is to compress the transformed coefficients $\boldsymbol{\chi}_k$ to satisfy

$$\max_B \left\{ \alpha \hat{J}_{11}(\hat{\boldsymbol{\chi}}_k) + (1-\alpha)\hat{J}_{22}(\hat{\boldsymbol{\chi}}_k) \right\}, \text{ subject to } \sum_{n=1}^{N} b_n \le R\,. \tag{9}$$

where $\hat{J}_{ii}(\hat{\boldsymbol{\chi}}_k)$ is the ii^{th} element of the data-computed FIM and the maximization is done over all allocation sets B that satisfy the rate constraint. See [2] for the detailed computation of $\hat{J}_{ii}(\hat{\boldsymbol{\chi}}_k)$.

2.2.3. *Compression for joint TDOA/FDOA*

Two received signals at sensors S_1 and S_2 with unknown TDOA n_d and FDOA v_d can be modeled by:

$$\begin{aligned} x_1[n] &= s[n - (n_0 + n_d/2)]e^{j(v_0 + v_d/2)} + w_1[n] \\ x_2[n] &= s[n - (n_0 - n_d/2)]e^{j(v_0 - v_d/2)} + w_2[n] \\ n &= -N/2, -N/2+1, \ldots, N/2 \end{aligned} \tag{10}$$

where $s[n]$ is a complex baseband signal, v_0 and n_0 are unknown nuisance parameters that need not be estimated, and $w_i[n]$ are uncorrelated complex Gaussian white noise with variances σ_i^2 . The signal-to-noise ratios (SNR) at the i^{th} sensor is SNR_i. We assume here that signal $x_1[n]$ at sensor S_1 is compressed and sent to sensor S_2 where it is decompressed and then used with $x_2[n]$ to perform joint TDOA/FDOA estimation.

According to (10), and motivated by [2], the weighted-trace-based TDOA/FDOA distortion measure is

$$\mathrm{wtr}(\hat{\mathbf{J}}_1) = \alpha \sum_{j=1}^{M} \left(\frac{f_j^2 \sum_{n \in \{j \text{ block}\}} |c_n|^2}{\sigma_k^2 + q_{k,j}^2} \right) + (1-\alpha) \sum_{j=1}^{M} \left(\frac{t_j^2 \sum_{n \in \{j \text{ block}\}} |c_n|^2}{\sigma_k^2 + q_{k,j}^2} \right). \quad (11)$$

where $\{c_n\}$ are wavelet packet coefficients of $\mathbf{x}_1$ and are grouped into M blocks; $\{q_{k,j}^2\}$ are the quantization noise variance in the j^{th} block, and f_j and t_j are the frequency and central time, respectively, of the j^{th} block. Bits are allocated to the coefficients using the method of [7] to maximize (11) for a given rate R. Determinant-based TDOA/FDOA distortion is similarly defined in (7).

Simulation results for the weighted trace method and the determinant method are shown in Figure 1, where the MSE method is also shown for comparison. We see the inherent trade-off between TDOA and FDOA that is controlled by the choice of α; it is possible to adaptively choose α [4],[5]. In contrast, only one operating point is obtained from the determinant method.

2.3 Compression for simultaneous estimation/detection

The fusion sensor might need to detect a signal first and then estimate parameters. The Chernoff distance of all the sensors' data $\{\boldsymbol{\chi}_k\}_{k=1}^{K}$ is the sum of the Chernoff distance of $\boldsymbol{\chi}_k$ when the noises are independent. Although Chernoff distance has been used [3] to quantify the degradation of compression, here we explore the tradeoffs between conflicting requirements of detection and estimation. In order to keep the distortion measure scalar and additive as well as simple, we use the following. A FIM-Chernoff-based distortion measure for joint detection and multiple estimations is given by

$$\max_{B} \left[\beta \psi(\hat{\mathbf{J}}(\hat{\boldsymbol{\chi}}_k)) + (1-\beta) \mu_s(\hat{\boldsymbol{\chi}}_k) \right], \text{ subject to } \sum_{n=1}^{N} b_i \le R, \quad (12)$$

where ψ represents any form of the FIM measures in Section 2.2, $\mu_s(\hat{\boldsymbol{\chi}}_k)$ is the Chernoff bound of $\hat{\boldsymbol{\chi}}_k$ and β is a parameter used to control the relative importance of estimation accuracy and detection error. See [3] for details on how $\mu_s(\hat{\boldsymbol{\chi}}_k)$ is proportional to SNR_k when $\mathbf{w}_k$ is Gaussian. The measure in (12) will be further considered in compression for sequential inference tasks.

3. COMPRESSION FOR SEQUENTIAL TASKS

In a sensor network there are cases where inference tasks are naturally done sequentially and therefore the sharing of data can also occur sequentially. For example: first the data is shared among the sensors to decide if K sensors have intercepted signals from the same source; then, an estimation task would be performed to locate the source. As we know, the compression requirements for the different tasks are different and usually conflicting. Therefore, optimized compression to handle sequential tasks requires what we call "task-embedded compression" here: *the transmitting sensor constructs the optimal task-embedded data stream to send only the data needed to supplement the already-delivered data for the current task.*

On the other hand, to meet aggressive time-line requirements between the m^{th} and (m+1)th task, it may be desirable to force the previous m stages to send data that would ultimately be helpful in the $(m+1)^{\text{th}}$ stage. For example, we wish to reduce the $(m+1)$-th stage bit rate without reducing inference quality at the $(m+1)^{\text{th}}$ stage. In the previous m stages, we could allocate some bits which are eventually beneficial to the inference accuracy of the $(m+1)^{\text{th}}$ stage. The inference qualtiy will be worse at the first m stages but the total rate up to the $(m+1)$-th stage will be about the same. This can lead to a situation where in each stage there might be multiple simultaneous inference tasks that generally have conflicting data compression requirements; thus, we must use (12) and choose a suitable β to achieve the proper tradeoff.

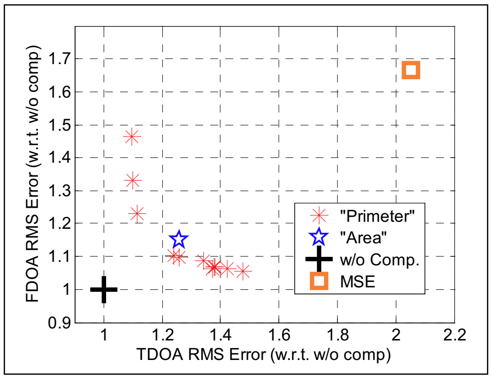

Figure 1: Comparison between determinant and weighted trace methods for compression ratio 3:1 and SNR_{i} = 15 dB.

3.1 Example: sequential detect-then-TDOA

Consider the scenario where multiple sensors are deployed to detect and then locate RF emitters. At first the sensors would share collected data for the purpose of detecting if they have jointly intercepted a common signal. After detection, further data is shared among the sensors to estimate the emitter's position using the TDOA method. In these two sequential stages, our task-embedded compression can be applied as follows: (i) the data stream that is shared during the detection phase is optimally compressed for detection, then (ii) send the additional data "layer" needed to optimally estimate TDOA. As stated in Section 2.3, the Chernoff-distance-based distortion measures for the detection task depends on the SNR of post-compressed data; however, according to [2], the FI-based distortion for TDOA depends on quadratically-frequency-weighted DFT coefficients.

Stage 1: Maximizing SNR for the Detection Task:

$$\underset{\{b_{n1}\}}{\mathit{Maximize}} \left[\beta \times \sum_{n=-N/2}^{N/2} \frac{n^2 |X_k[n]|^2}{\sigma^2 + q_k^2(b_{n1})} + (1-\beta) \times \sum_{n=-N/2}^{N/2} \frac{|X_k[n]|^2}{\sigma^2 + q_k^2(b_{n1})} \right] \quad (13)$$

$$\text{subject to } \sum_{n=0}^{N-1} b_{n1} \le R_D$$

where $X_k[n]$ is the DFT of the data $\mathbf{x}_k$, $q_k^2(b_{n1})$ is the quantization noise power as a function of b_{n1} bits allocated to that coefficient, R_D is the rate for Stage 1, and the parameter β controls the tradeoff between TDOA estimation and detection. Setting β = 0 causes this allocation to be done with no consideration of the Stage 2 task of TDOA estimation; however, increasing β forces more consideration of the subsequent Stage 2 task.

Stage 2: Maximizing FI for the TDOA Estimation Task:

In this stage only TDOA is of interest and its accuracy needs to be refined. Thus we will maximize

$$\underset{\{b_{n2}\}}{Maximize}\left[\sum_{n=-N/2}^{N/2}\left(\frac{n^2\left|X_k[n]\right|^2}{\sigma^2+q_k^2(b_{n1}+b_{n2})}\right)\right] \quad \text{subject to} \quad \sum_{n=0}^{N-1} b_{n2} \le R_E \text{ , (14)}$$

where R_E is bit budget for the Stage 2. and the $\{b_{n1}\}$ are the bits already allocated in the first stage.

The case considered above is illustrated in Figure 2 where the rates for different tasks are held fixed and performance is changed during the trade-off: the variation of β enables the tradeoffs for the detection during the first detection stage and the location accuracy during the second estimation stage.

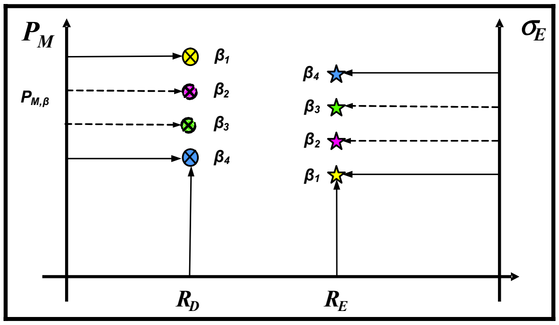

Figure 2: Conceptual illustration of the trade-off accomplished via choice of the β parameter.

For the simulation results below, similar to Section 2.1.3, we assume that signal $\mathbf{x}_1[n]$ at sensor S_1 is compressed sequentially and sent to sensor S_2, where the sequential tasks of detect-then-TDOA are performed. Moreover, we impose the conditions: during the first detection stage, in order to satisfy a rate constraint R_D, the data is compressed with a compression ratio of 8:1 (without entropy coding); on the other hand, during the second stage estimation, in order to satisfy a rate constraint R_E, Stage 2 bits are added such that the total compression ratio is 5:1 (without entropy coding). In terms of practical coding, unlike the simultaneous case, quantizers must be changed to embedded quantizers in order to support the task-embedded data stream between the sensors. Simulation results are shown in Figure 3, where the actual probabilities in the detection stage were not evaluated because that would take tremendous time.

The results in Figure 3 show the various SNR-σ_E points that can be achieved for fixed R_D and R_E; the points in the upper right corner favor detection while those in the lower left corner favor estimation. Note that the values of the post-compression detection-stage SNR were quite low; this is due to the large compression ratio that was imposed. The coherent processing gain of cross-correlation-based detection will raise these values significantly. Furthermore, detection accuracy is not only dependent on post-compression SNR but also on the size of data; any increase in SNR indicates a significant decrease in error probability for large sample sizes. The specific operating values shown in Figure 3 are of less importance than the fact that the simulation results verify that the algorithm achieves a curve of trade-off points as expected.

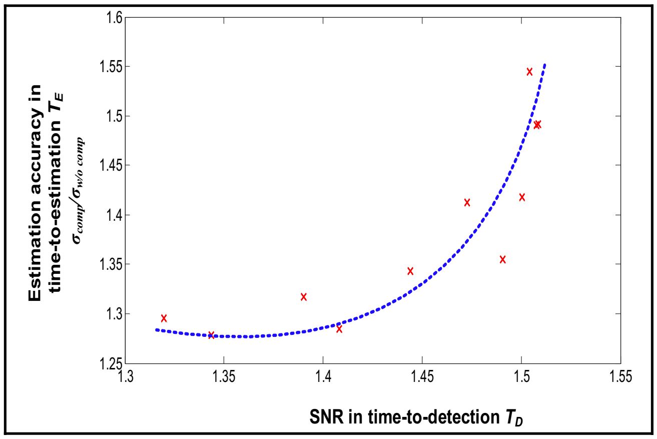

Figure 3: Simulation results illustrating the achieved trade-offs.

4. CONCLUSION

By recognizing that multiple inferences (simultaneous and/or sequential) generally have conflicting compression requirements, we developed the theory and several algorithms for both the general simultaneous estimation problem and the simultaneous estimation/detection problem. We also proposed and explored a task-embedded compression approach to support sequential inferences tasks. More importantly, all algorithms are based on additive rate-distortion measures and are therefore very effective to be directly applied in current distributed sensor systems.

5. REFERENCES

[1] H.V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlayge.

[2] M. L. Fowler and M. Chen, “Fisher-information-based data compression for estimation using two sensors,” to appear in *IEEE Trans. on Aero. Elect. Systems.* Available at http://www.ws.binghamton.edu/fowler/

[3] A. Jain, P. Moulin, M. Miller, K Ramchandran, “Information-theoretic bounds on target recognition performance based on degraded image data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, No.9, Sep. 2002, pp.1153-1166

[4] M. Chen and M. L. Fowler, “Geometry-Adaptive Data Compression For TDOA/FDOA Location,” *IEEE ICASSP* 2005, Philadelphia, PA, pp. IV1069 – IV1072.March 18 – 23, 2005.

[5] M. Chen, PhD Dissertation, *Data Compression for Inference Tasks in Wireless Sensor Network*, State University of New York at Binghamton, 2005, Available at http://www.ws.binghamton.edu/fowler/

[6] A. Ortega and K. Ramchandran, “Rate distortion methods for image and video compression,” *IEEE Signal Processing Magazine*, vol. 15, Nov. 1998, pp. 23 – 50.

[7] Y. Shoham and A. Gersho, “Efficient bit allocation for an arbitrary set of quantizers,” *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, No. 9, September 1988, pp. 1445-1453.