TOWARDS LARGE SCALE DISTRIBUTED CODING

Sharadh Ramaswamy, Kumar Viswanatha, Ankur Saxena, Kenneth Rose

Department of Electrical and Computer Engineering
University of California at Santa Barbara, CA-93106.
E-mail: {rsharadh, kumar, ankur, rose}@ece.ucsb.edu

ABSTRACT

This paper considers the problem of distributed source coding for a large sensor network. A typical shortcoming of current approaches to true distributed coding is the exponential growth of the decoder codebook size with the number of sources in the network. This growth in complexity renders many traditional approaches impractical for even moderately sized sensor networks. Inspired by our recent results on fusion coding for selective retrieval, we propose a new distributed coding approach that scales to a large number of sources. Central to our approach is a "bit-subset selector" module whose role is to judiciously extract an appropriate subset of the received bits for decoding per individual source. This, together with joint design of all system components, enables direct optimization of the decoder complexity-distortion tradeoff, and thereby the desired scalability. Experiments on both real and synthetic data-sets show considerable gains over heuristic schemes.

Index Terms— Distributed Coding, Sensor Network, Compression, Quantization, Codebook Complexity

1. INTRODUCTION

Distributed source coding (DSC) has been studied extensively in both the information theory and signal processing research communities, ranging from the pioneering work of Slepian -Wolf [1] and Wyner-Ziv [2] to more practical approaches inspired on channel coding principles [3, 4]. An alternative approach to designing distributed coders springs directly from the source coding perspective [5, 6]. Unlike channel coding-inspired methods, they make no restrictive prior assumptions on source distributions and incur low to zero delay during decoding. The block diagram of the canonical DSC system is shown in Fig.1.

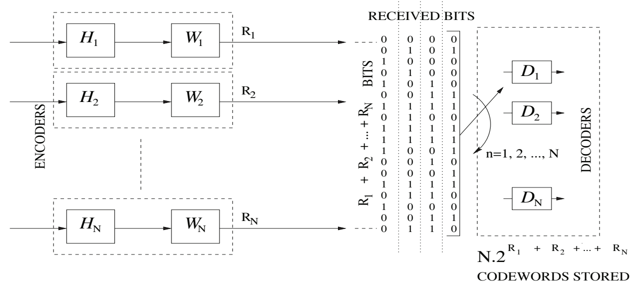

Fig. 1. Typical DSC scenario

Fig. 2. Example of a WZ-map

Distributed coding for a large number of sources is, in theory, a trivial extension of the two source case, but the exponential codebook size growth with the number of sources, makes it nonviable in most practical applications. Just to illustrate, consider 20 sensors, transmitting information at 2 bits per sensor. The base station receives 40 bits from which it reconstructs estimates of the signals perceived by the 20 sensors. This implies that the decoder has to maintain a codebook of size $20 \cdot 2^{40}$ which requires about 175 Terabytes of memory. In general, for N sources transmitting at R_s bits, the total decoder codebook size would be $N \cdot 2^{NR_s}$. Clearly, the design of optimal low-storage distributed coders is a crucial problem whose solution has existential ramifications to application of DSC in real world sensor networks.

While there is a considerable volume of high quality DSC work in the literature, most of coders have been developed and demonstrated for 2-3 sources. To the best of our knowledge, the main attempt so far to scale DSC in a practical scenario [4], is by first clustering the sources based on their statistics (assumed to be jointly Gaussian) and then using principles from channel coding to achieve gains. In this paper, inspired by the principles of fusion coding [7, 8], we introduce a bit-subset selector at the decoder, which allows the decoder to use any subset of received bits to decode any source. The subset sizes control the source reconstruction rates and codebook sizes. Thus, a direct tradeoff between decoder storage complexity and reconstruction distortion is possible. Unlike [4], our framework for large scale distributed quantization does not assume any restrictive prior knowledge of source statistics beyond a training set, and is based on source coding principles. Experiments with both real and synthetic data sets show that our approach reduces codebook complexity, by factors reaching $16X$, for roughly the same distortion over heuristic source grouping methods.

The rest of the paper is organized as follows. In section 2 we describe the conventional DSC setting and the challenge of scaling to large networks. In section 3, we propose a new framework and formulate a Lagrangian to quantify the codebook complexity-distortion tradeoff. In section 4 we derive the necessary conditions for optimality. We summarize the system design algorithm in section 5, and the simulation results in section 6.

2. THE TYRANNY OF DIMENSIONS

We begin with a description of the DSC system. Consider N correlated sources, $\{X_i,\ i = 1 \ldots N\}$ transmitting information at rate R_i, respectively, to the fusion center. The encoding consists of 2

The work was supported by the NSF under grant CCF-0728986

stages. The first stage is the discretization of the source-space by a high-rate quantizer (a practical engineering necessity, see e.g., [5, 6], which partitions the input source space into a finite number of regions N_i i.e.,

$$\mathcal{H}_i : \mathcal{R} \rightarrow \mathcal{Q}_i = \{1, ..., N_i\}$$

The second stage, which we call the 'Wyner-Ziv map' (WZ-map), is a module that relabels the N_i quantizer regions with a smaller number, 2^{R_i}, of labels, which if properly designed exploits the correlation between the sources and aids in distributed compression [5, 6]. Mathematically, for each source i, the WZ map is the function,

$$\mathcal{W}_i : \mathcal{Q}_i \rightarrow \mathcal{I}_i = \{1, ..., 2^{R_i}\}$$

and the encoding operation can be expressed as:

$$\mathcal{E}_i(x_i) = \mathcal{W}_i(\mathcal{H}_i(x_i)), \forall i \tag{1}$$

We use $\mathcal{I} = \mathcal{I}_1 \times \mathcal{I}_2 \cdots \mathcal{I}_N$ to denote the set of all possible received symbols at the decoder. The total number of bits received by the decoder is $R_r = \sum_{i=1}^{N} R_i$. Decoder $\mathcal{D}_i$ for source X_i, is given by

$$\mathcal{D}_i : \mathcal{I} \rightarrow \hat{\mathcal{X}}_i \subseteq \mathcal{R}$$

The decoder needs to store a unique reconstruction for each possible received bit sequence. Therefore for N sources each transmitting at R bits, the total number of entries in the look-up table scales as $|\mathcal{I}| = 2^{\sum_{i=1}^{N} R_i} = 2^{NR}$ [1]. Note that the above DSC approach is different from classical 'Slepian-Wolf' coding applied on quantized indices. But even there, at minimum sum rate, the set of jointly ϵ-typical sequences $A_\epsilon^{(n)}$ (for block length n) has cardinality $|A_\epsilon^{(n)}| \geq (1-\epsilon)O(2^{n(H(X_1,...,X_N)-\epsilon)}) = (1-\epsilon)O(2^{n(\sum_1^N R_i - \epsilon)})$! In most prior work, distributed coding was performed for a few (typically $2-3$) sources, with the implicit assumption of design scalability with network size. What was often overlooked was the exponential growth of the size of the look-up table. Unless a severe restriction on N is imposed, such systems cannot conceivably be designed or implemented.

3. LARGE SCALE DISTRIBUTED QUANTIZATION

We now describe our approach to large scale distributed quantization. Inspired by the recent results on the fusion coding for selective retrieval [7, 8], we introduce an additional module at the decoder which we term the bit-subset selector. For each source, the bit-subset selector determines the subset of the received bits that shall be used in reconstructing a particular source. For each subset of received bits, a separate codebook is used to reconstruct the corresponding source. Fig. 3 represents our system in a block diagram form.

Formally, the bit subset selector is the mapping :

$$S : \{1, ..., N\} \rightarrow \mathcal{B} = 2^{\{1,...,R_r\}} \tag{2}$$

where $R_r = \sum_{i=1}^{N} R_i$ and $\mathcal{B}$ is the power set (set of all possible subsets) of the set $\{1, \ldots, R_r\}$. For source i, the bit-subset selector uses the the bits indicated by $\mathcal{S}(i)$ for decoding. This implies a decoding rate of $R_{d_i} = |\mathcal{S}(i)|$ bits, which would necessitate a codebook of size $2^{|\mathcal{S}(i)|}$, where $|\cdot|$ is used to denote set cardinality. For different choices of $\mathcal{S}(i)$, reconstructions of source i at different distortion levels are possible and thus, reconstruction quality can be traded against the required codebook size.

[1]Note that even for a cascaded coding system, the lookup table at each of the N successive stages of the tree becomes exponentially large.

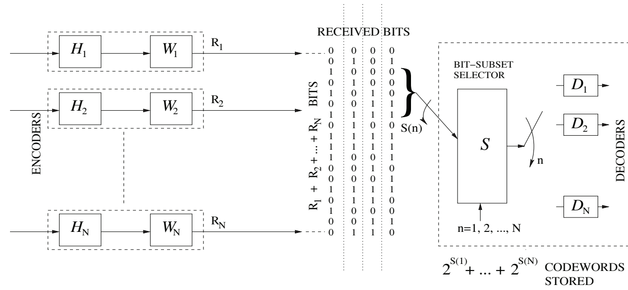

Fig. 3. Proposed Setup

The decoder for each source is now modified to be the mapping

$$\mathcal{D}_i : \mathcal{I} \times \mathcal{B} \rightarrow \hat{X}_i \in \mathcal{R} \tag{3}$$

We define the decoder complexity of the system, C, as the average codebook size,

$$C = \frac{1}{N} \sum_{i=1}^{N} 2^{R_{d_i}} = \frac{1}{N} \sum_{i=1}^{N} 2^{|\mathcal{S}(i)|} \tag{4}$$

The average reconstruction distortion is measured as:

$$D = E[\sum_{i=1}^{N} \gamma_i d_i(X_i, \hat{X}_i)] \tag{5}$$

where $d_i : R \times R \rightarrow [0, \infty)$ is a well-defined distortion measure and $0 \leq \gamma_i \leq 1 : \sum_{i=1}^{N} \gamma_i = 1$ are used to weigh the relative importance of each of the sources to the total distortion. Hereafter, we will specialize to the squared error distortion. In practice, we only have access to training sets and not the actual source distributions. Assuming ergodicity, expectation is approximated by simple averaging over the training set. If the training set is denoted as $\mathcal{T}$, we measure distortion as :

$$D = E[\sum_{i=1}^{N} \gamma_i (X_i - \hat{X}_i)^2] = \frac{1}{|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} \sum_{i=1}^{N} \gamma_i (x_i - \hat{x}_i)^2 \tag{6}$$

where $\mathbf{x} = \{x_1 \ldots x_N\}$. The trade-off between distortion and decoder complexity is controlled by a Lagrange multiplier $\lambda \geq 0$ and by optimizing the weighted sum of the two quantities. From (4) and (6), the Lagrangian cost to be minimized is

$$\begin{aligned} L &= D + \lambda C \\ &= \frac{1}{|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} \sum_{i=1}^{N} \gamma_i (x_i - \mathcal{D}_i(I, S(i)))^2 \\ &\quad + \frac{\lambda}{N} \sum_{i=1}^{N} 2^{|S(i)|} \end{aligned} \tag{7}$$

where $I = [\mathcal{E}_1(x_1), \ldots, \mathcal{E}_N(x_N)]^T$. The design objective is to find $\{\mathcal{E}_i\}$, $\{\mathcal{D}_i\}$ and $\mathcal{S}$ that minimize L.

4. NECESSARY CONDITIONS FOR OPTIMALITY

In this section, we derive the necessary conditions for optimality of all modules in the proposed approach to distributed coding.

4.1. Optimal Encoders

Let $\mathcal{T}_{i,j} = \{\mathbf{x} \in \mathcal{T} : \mathcal{H}_i(x_i) = j\}$. Then, from (7), the optimum WZ-map is given by:

$$\mathcal{W}_i(j) = k^* = \arg\min_{k \in \mathcal{I}_i} \sum_{\mathbf{x} \in \mathcal{T}_{i,j}} \sum_{l=1}^{N} \gamma_l (x_l - \mathcal{D}_l(I_{i,k}, S(l)))^2 \quad (8)$$

where

$$I_{i,k} = [\mathcal{E}_1(x_1), \ldots, \mathcal{E}_{i-1}(x_{i-1}), k, \mathcal{E}_{i+1}(x_{i+1}), \ldots, \mathcal{E}_N(x_N)]^T$$

4.2. Optimal Bit-Subset Selector

For fixed encoders and decoder codebooks, the optimum subset of bits to be used to estimate each source is given by:

$$S(i) = e^* = \arg\min_{e \in \mathcal{B}} \sum_{\mathbf{x} \in \mathcal{T}} (x_i - \mathcal{D}_i(I, e))^2 + \frac{\lambda}{N} \times 2^{|e|} \quad (9)$$

where $I = [\mathcal{E}_1(x_1), \ldots, \mathcal{E}_N(x_N)]^T$.

4.3. Optimal Decoders

If $I = [\mathcal{E}_1(x_1), \mathcal{E}_2(x_2) \ldots \mathcal{E}_N(x_N)]^T$ represents the bits received from all the sources, and if e represents the positions of bits selected by a decoder, then we use I_e to denote the index obtained by extracting the bits in I at the positions indicated by e. Differentiating the expression for L, (7), with respect to the reconstruction values and equating it to zero gives the optimal expression for the decoder to be:

$$\hat{x}_i(I, e) = \mathcal{D}_i(I, e) = \frac{1}{|\mathcal{F}|} \sum_{\mathbf{x} \in \mathcal{F}} x_i \quad (10)$$

where $\mathcal{F} = \{\mathbf{x} \in \mathcal{T} : ([\mathcal{E}_1(x_1), \mathcal{E}_2(x_2) \ldots \mathcal{E}_N(x_N)]^T)_e = I_e\}$.

5. ALGORITHM FOR SYSTEM DESIGN

Given the necessary conditions for optimality of the proposed distributed coding system, a natural design rule is to iteratively optimize the different modules. When each module is optimized, it leads to a decrease in the Lagrangian cost. Since there are only a finite number of source partitions possible, convergence to a locally optimal solution is guaranteed. By varying λ, the trade-off between decoder complexity and distortion is controlled and an operational complexity-distortion curve is obtained. To mitigate the issue of local minima, the system is optimized over multiple random initializations. Global optimization techniques such as deterministic annealing [6], can be used to avoid poor local minima but are beyond the scope of this paper.

5.1. Design Complexity

We first consider the optimization of the WZ maps. The complexity of this step grows as $\mathcal{O}(|\mathcal{T}|N(\sum_{i=1}^{N} N_i 2^{R_i}))$, which is polynomial in the number of sources[2]. In contrast, the design of the bit-subset selector is a high complexity step. In order to find the best bit-subset selector for any decoder, every possible combination of bits should be considered for source reproduction at the corresponding decoder. This will necessitate $\mathcal{O}(N2^{R_r})$ calculations and also imply a storage of $N\sum_{k=1}^{R_r}\binom{R_r}{k}2^k = N(3^{R_r} - 1)$ codewords during design. Note that R_r grows linearly with N. Thus, the number of computations and codewords to be maintained grows exponentially with N which complicates design.

5.1.1. *Low-Complexity Design Scheme*

Instead of finding the best among all possible bit-subset selector settings, at each step of optimization an incrementally better bit selection is chosen. Moreover, the search is restricted to only those bit-subsets that differ from the current bit-selector setting in exactly one position, i.e., restricted to the R_r bit subsets at a Hamming distance of 1 from the current bit-subset. The design complexity of this technique is $\mathcal{O}(NMR_r)$, where M is the number of iterations across all bits allowed by the designer. Note that now the design complexity is polynomial in N. In our simulations, we observed that the bit subset selector usually converged to a local optimum within 3 iterations ($M = 3$). Additionally, during each iteration, only R_r codebooks need to be maintained.

6. RESULTS

We tested our proposed algorithm extensively on both synthetic and real sensor network data and obtained the operational complexity (C) vs. the distortion (D) (C-D curve) for each of the data sets. In all our simulations we considered a transmission rate $R_i = 2$ bits per source and $\gamma_i = 1 \; \forall i$. The high rate quantizer partitioned the input space into 32 regions. To be fair, we used the same high rate quantizers for our technique and the competitors. For all methods, we report the best performance over several random initializations (limited to 40). The maximum average complexity allowed for the decoder codebook was 1024 (10 bits) [3].

As a competitor, we grouped the sources heuristically based on their correlations, making sure that the sources with higher correlations are grouped together and applied conventional distributed coding for each group. We varied the number of sources in each group to obtain the distortion at different complexities. As another competitor, we also considered the performance of the canonical distributed source coder at a lower transmission rate of $R_i = 1$ bit per source.

Lastly, from (4), note that only a finite number of complexities are possible. The popular 'time sharing' argument is not applicable here because using a small memory for a long time and a very large memory for a short time would still necessitate the decoder to have a very large memory[4].

6.1. Synthetic Datasets

We first considered synthetic Gaussian sources, $\mathcal{N}(0, 1)$, with correlation exponentially decaying with the distance. Specifically, the correlation coefficient between sources X_i and X_j, is $\rho_{ij} = \rho^{|i-j|}$. In our simulations, $\rho = 0.95$ and training set length $|\mathcal{T}| = 200,000$. The results reported are on a test set of length $200,000$.

6.1.1. *5 Sources*

Fig. 4 shows the performance of the proposed schemes for 5 sources. Our approach outperforms the heuristic grouping scheme by about $1\,dB$ at a complexity of 2^7. Our method gains $2.5dB$ compared

[2] Note that the design of the high rate quantizers for each source is performed independently using the Lloyd-Max scheme, which is a relatively low complexity step.

[3] Note that the smallest complexity we consider is 4, as we force the bit-subset selector to select at least the 2 bits sent by the corresponding encoders.

[4] We still connect all the points on the C-D curve for clarity.

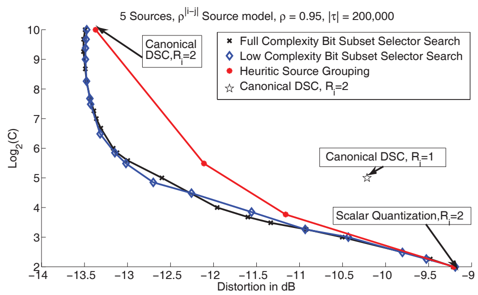

Fig. 4. 5 Sources

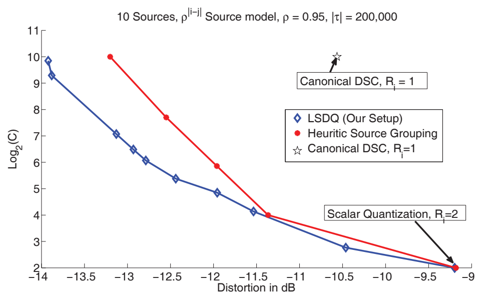

Fig. 5. 10 Sources

to the distortion achieved by the canonical DSC transmitting at 1 bit/source. At some points on the curve, due to issues with local minima, the performance of the full complexity search is marginally worse compared to the low complexity search. Note that, one extreme of the C-D curve closely approximates scalar quantization and the other extreme corresponds to canonical distributed coding, both at 2 bits/source.

6.1.2. 10 Sources

Next, we considered 10 synthetic Gaussian sources. Fig. 5 shows the comparison. At a complexity of 2^{10}, our method gains 1 dB over heuristic grouping and $3.5\,dB$ over canonical DSC at 1 bit/source. In terms of codebook size, our scheme requires $16X$ smaller complexity at a distortion of $-13\,dB$.

6.2. Real Dataset

The next data set considered is the Intel Berkeley Research center dataset [5]. The data had 50,000 samples of temperature, light, humidity and voltage readings from 15 sensors, which is equivalent to 60 sources. Sources were scaled to have zero mean and unit variance. For this dataset, our approach does $3dB$ better than the heuristic scheme at a complexity of 2^9. It gains $16X$ in codebook size at a distortion of $-21dB$. The gains in case of the real dataset are more substantial than for the synthetic datasets due to the fact that our method is capable of capturing higher order dependencies, while the heuristic grouping is based only on second order statistics.

[5] Available at http://db.csail.mit.edu/labdata/labdata.html

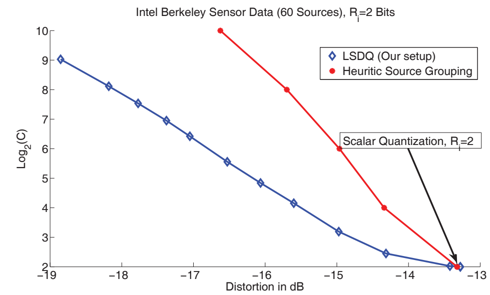

Fig. 6. Real Data Set (60 Sources)

7. CONCLUSIONS

We proposed a new setup to design a large-scale distributed coding system and operate at practical codebook sizes. We formulated a Lagrangian cost and proposed an iterative design algorithm to optimize the performance tradeoff between complexity and distortion. Simulation results show that the proposed algorithm has significant gains over heuristic schemes.

An important issue that needs to be considered in design of such distributed coding systems is the required training set, the length of which also grows exponentially with the number of sources to ensure good generalization. Yet another problem is the presence of local minima that result in suboptimal designs. Future work will focus on these and related problems

8. REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE. Trans. on Information Theory*, vol. 19, pp. 471–480, Jul 1973.

[2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. on Information Theory*, vol. 22, pp. 1–10, Jan 1976.

[3] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (discus): Design and construction," *IEEE Trans. on Inform. Theory*, vol. 49, pp. 626–643, 1999.

[4] G. Maierbacher and J. Barros, "Source-optimized clustering for distributed source coding," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Dec 2006, pp. 1–5.

[5] D. Rebollo-monedero, R. Zhang, and B. Girod, "Design of optimal quantizers for distributed source coding," in *IEEE Data Compression Conferecne*, Mar 2003, pp. 13–22.

[6] A. Saxena, J. Nayak, and K. Rose, "On efficient quantizer design for robust distributed source coding," in *IEEE Data Compression Conference*, Mar 2006, pp. 63–72.

[7] J. Nayak, S. Ramaswamy, and K. Rose, "Correlated source coding for fusion storage and selective retrieval," in *IEEE International Symposium on Information Theory*, Sep 2005, pp. 92–96.

[8] S. Ramaswamy, J. Nayak, and K. Rose, "Code design for fast selective retrieval of fusion stored sensor-network/time-series data," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Apr 2007, vol. 2, pp. II–1005–II–1008.