JOINT SOURCE DECODING IN LARGE SCALE SENSOR NETWORKS USING MARKOV RANDOM FIELD MODELS

Pradeepa Yahampath

Department of Electrical and Computer Engineering
University of Manitoba, Winnipeg, Canada
email:pradeepa@ee.umanitoba.ca

ABSTRACT

Scalable joint decoding of correlated observations transmitted using distributed quantization in a sensor-network is considered. In particular, quantized observations are modeled as a Markov-random filed (MRF), from which we construct a factor-graph for implementing the decoder using the well known sum-product algorithm. An attractive property of this approach is that the decoder complexity can be controlled by the choice of the clique structure used to define the Gibbs distribution of the MRF model. The experimental results obtained with a widely used correlated Gaussian observation model is presented, which demonstrate that substantial performance gains can be achieved by joint decoding based on simple clique structures and potential functions.

Index Terms— Distributed quantization, Markov-random fields, factor-graphs, sum-product algorithm

1. INTRODUCTION

Due to the recent advances in wireless sensor networks (WSNs), distributed source coding (DSC) and related issues are being widely investigated [1]. Since the observations produced by a dense sensor network tend to be highly correlated, DSC used at individual sensors can exploit this correlation to reduce the wireless transmission-bandwidth requirements. However, the computational complexity of the underlying optimal joint decoder grows exponentially with the network size. One can hope for near-optimal, but tractable decoding (complexity linear in network size), if the complex global decoder function can be successfully decomposed into a large number of simpler (local) functions. A particularly effective approach along these lines [2] is to use a *factor graph* to represent (perhaps approximately) the prior joint distribution of the quantized sensor outputs and then to apply the *sum-product (SP) algorithm* to compute the posterior joint distribution. In [2], the sensor observations are assumed to be Gaussian, and a factorized approximation to the prior joint distribution was obtained by minimizing the Kullback-Leibler distance (KLD). In [3] a more general approach (not requiring the Gaussian assumption) is presented wherein a Bayesian network learning is used to factorize the joint distribution.

This work was supported by a Discovery Grant from NSERC of Canada.

However, the aforementioned approaches do not explicitly use local dependencies very likely present in a sensor field, which can be used to greatly simplify the design of the joint decoder. The design procedures in [2, 3] involve search algorithms for optimization and the resulting factorization of the prior joint distribution lacks any structure. In contrast, this paper propose a simplified approach to designing a structured joint decoder based on Markov random filed (MRF) modelling of sensed observations. MRFs [4, 5, 6] are multi-dimensional generalizations of one-dimensional Markov processes widely used for modeling signals. We make use of the equivalence between an MRF and the Gibbs distribution [4], which allows the complete specification of the global probability distribution in terms of local conditional distributions. The Gibbs distribution can be used to directly represent the underlying joint decoder in desired factor graph form, and the complexity of the SP algorithm can be controlled by restricting the size of the *cliques* [4] taken into account.

The experimental results presented here reveal that joint decoders based on simple neighborhood structures and pair-cliques (pairwise interaction models) can yield substantial performance gains compared to independent decoders operating on individual sensors, as well as simpler joint decoders which operate on small clusters of sensors. It is worth recognizing that the proposed approach not only simplifies the design process, but also offers a flexible design compared to unstructured decoders considered in [2, 3]. For example, the addition or deletion of new nodes only require small localized changes to the decoder structure, and not a complete re-design as would be required by approaches in [2, 3] .

2. PROBLEM STATEMENT AND PREVIOUS WORK

In Fig. 1, the random vector $\mathbf{X} = (X_1, \ldots, X_K)^T$ represents a set of correlated random variables observed by a field of spatially distributed sensors. Each sensor quantizes its observation, performs *binning* [1], and transmits the outcome over an independent channel to a central decoder which jointly forms the reconstruction values $\hat{X}_1, \ldots, \hat{X}_K$, based on the outputs received from all the sensors. We assume that the channels are discrete and memoryless. The binning, which is essentially an integer-to-integer mapping, is used to reduce the transmis-

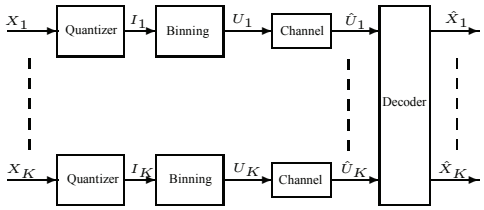

Fig. 1. The sensor network model considered in this paper.

sion bit rate of each sensor, by exploiting a priori known statistical dependencies among the sensors. Let the quantizer output, the transmitted index, and the channel output for the k-th sensor be $I_k \in \{1, \ldots, 2^{R_{Qk}}\}$, $U_k \in \{1, \ldots, 2^{R_{Tk}}\}$, and $\hat{U}_k \in \{1, \ldots, 2^{R_{Tk}}\}$ respectively, where R_{Qk} is the quantization rate and $R_{Tk} (\leq R_{Qk})$ is the transmission rate in bits/sample, $k = 1, \ldots, K$. For simplicity, we use the notation $\mathbf{I} = (I_1, \ldots, I_K)^T$ and $\hat{\mathbf{U}} = (\hat{U}_1, \ldots, \hat{U}_K)^T$. As usual, we use the mean square error (MSE) as the distortion measure, and hence the optimal joint decoder minimizes for each k, $E[X_k - \hat{X}_k(\hat{\mathbf{U}})]^2$. Then, it follows that the optimal source reconstructions are given by

$$\hat{X}_k^*(\hat{\mathbf{u}}) = \sum_{i=1}^{2^{R_{Qk}}} E\{X_k | I_k = i\} P\big(I_k = i | \hat{\mathbf{u}}\big), \qquad (1)$$

$k = 1, \ldots, K$, where $\hat{\mathbf{u}}$ is the observed channel output vector. By defining

$$\phi_k(i, \hat{\mathbf{u}}) = \sum_{\mathbf{I}: I_k = i} P(\hat{\mathbf{u}}|\mathbf{I}) P(\mathbf{I}), \qquad (2)$$

we have $P\big(I_k = i|\hat{\mathbf{u}}\big) = \phi_k(i, \hat{\mathbf{u}}) / \sum_j \phi_k(j, \hat{\mathbf{u}})$. The difficulty in evaluating this expression lies with the marginalization in (2) which has a computational complexity exponential in the number of sensors K. In order to address this issue, [2] considered factorization of the joint pmf $P(\mathbf{I})$ into factors which only involve only a small number of variables in each. Then, it is possible to compute (2) with a complexity that is linear in K, using a factor graph and the SP algorithm. More specifically, suppose that

$$P(\mathbf{I}) = \prod_{l=1}^{L} P(\mathbf{a}_l | \mathbf{b}_l), \qquad (3)$$

where $\mathbf{a}_l, \mathbf{b}_l \subset \mathbf{I}$ such that $\mathbf{a}_l \cap \mathbf{a}_m = \emptyset$ and $\cup_{l=1}^{L} \mathbf{a}_l = \mathbf{I}$. The chain rule expansion is one example, but without any assumptions about the conditional dependencies, the size of the set $\mathbf{b}_l$ in some of the factors can be quite large. The complexity of the SP algorithm is exponential in the number variables (*degree*) in each factor $P(\mathbf{a}_l|\mathbf{b}_l)$ [represented by a *factor node* in the factor graph]. In order to force the complexity to grow linearly with K, [2] proposes to find a close approximation $\tilde{P}(\mathbf{I})$ in which each factor involves only a fixed number of variables, so that the maximum degree $d_{\max} = \max_{1 \leq l \leq L} \{|\mathbf{a}_l| + |\mathbf{b}_l|\}$ of factor nodes in the factor graph is constrained. In [2], a search algorithm is used to find a factorization with a desired $d_{\max}$ value.

3. MRF MODELING OF SENSOR OUTPUTS

For a thorough treatment of MRFs, see for example [4, 5]. We define an MRF with reference to an *undirected graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of vertexes (nodes) and $\mathcal{E}$ is the set of edges. Let each node in $\mathcal{V}$ represent a discrete random variable in $\mathbf{I}$. Two variables I_k and I_l are *neighbours* if there is an edge between the two corresponding nodes. Denote the set of neighbours of I_k by $\mathbf{N}_k$ and let $\mathbf{I}'_k = (I_1, \ldots, I_{k-1}, I_{k+1}, \ldots, I_K)$. Then $\mathbf{I}$ is an MRF if

$$P(I_k | \mathbf{I}'_k) = P(I_k | \mathbf{N}_k). \qquad (4)$$

This is clearly a good model for a dense sensor network wherein the output of a given sensor node is strongly dependent on a few other closely located sensors. As will be seen, MRF model allows us to exploit typical local dependencies in a sensor field to factorize the joint pmf $P(\mathbf{I})$ into simple factors, and thereby efficiently compute the posterior probabilities in (2) using the SP algorithm.

For simplicity, first consider a network of 9 sensors placed on a 3×3 grid, where the variables are labelled $I_1, \ldots, I_9$ based on a left-right, top-bottom scan. Suppose, the neighborhood of a sensor node is assumed to be the closest horizontal and vertical (4-connected or 4-c) neighbors . Then, using (4) together with the chain rule expansion, we immediately have

$$P(\mathbf{I}) = \prod_{k=1}^{K} P(I_k | I_l \in \mathbf{N}_k : l > k), \qquad (5)$$

as required in (3). The factor graph for this example is shown in Fig 2 (a). Note that, with 4-c neighbourhood, $d_{\max} = 3$. Clearly, improved performance of the decoder can be expected, if the size of the neighborhood is expanded. In our example, we may include the diagonal neighbors as well (8-connected or 8-c neighborhood). However, this will also increase $d_{\max}$ to 5, which substantially increases the computational complexity of the SP algorithm. In general, the chain rule expansion has the undesirable property that $d_{\max}$ depends on the size of the neighborhood. An alternative factorization in which $d_{\max}$ is independent of the neighborhood size is obtained by considering the equivalence between a MRF and the *Gibbs distribution* [4].

Given that $\mathbf{I}$ is a MRF with respect to some neighborhood structure, it has a Gibbs distribution given by [4] $P(\mathbf{I} = \mathbf{i}) = \frac{1}{Z} e^{-U(\mathbf{i})/T}$, where $U(\mathbf{i})$ is the *energy function*, T is the *temperature*, and Z is the normalizing constant. The energy function has the form $U(\mathbf{I}) = \sum_c V_c(\mathbf{I}_c)$, where $V_c(\mathbf{I}_c)$ is the *potential* of the *clique* c containing the set of variables $\mathbf{I}_c \subset \mathbf{I}$. A clique is a subset of nodes in the graph, in which all elements

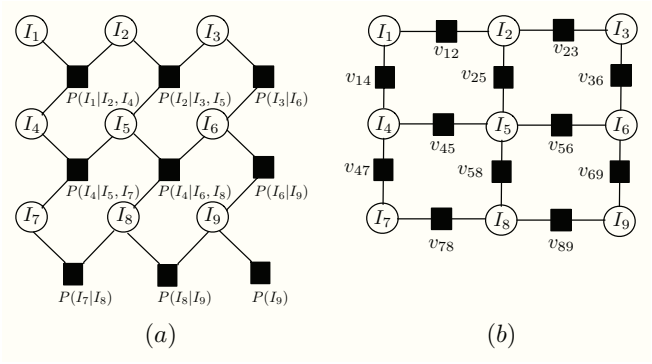

Fig. 2. Factor graph representations of an MRF with a 4-connected neighborhood structure: (a) chain-rule, (b) Gibbs distribution based on pair-cliques, where $v_{kl} = e^{-V_{kl}(I_k,I_l)/T}$.

are neighbors of each other. The above sum is taken over all possible cliques associated with the given graph. The form of $V_c(\cdot)$ is not unique, an issue to be addressed below. In any case, it follows that

$$P(\mathbf{I}=\mathbf{i}) = \frac{1}{Z}\prod_c e^{-V_c(\mathbf{i}_c)/T}, \qquad (6)$$

which is a factorization similar to (3).While the factors in this expression are not conditional probabilities, the overall expression can still be expressed in factor-graph form, as shown in Fig.2 (b). Note that Z will not affect the final result due to normalization of ϕ_k in (1). It is easy to see that, with this factorization, $d_{\max}$ is the maximum size of a clique. Thus, the complexity of the SP algorithm can be controlled by limiting the size of the cliques taken into consideration. In this paper, we only use single variable cliques and cliques containing pairs of neighbors in the 8-c neighborhood.

The choice of the clique potential V_c to be used depends on the spatial process being quantized. In this paper, we assume that the spatial process $\mathbf{X}$ under observation is smoothly varying, and hence use a quadratic potential function to model the interaction between a pair of quantized observations I_k and I_l, i.e., $V_{kl}(i_k, i_l) = \beta_{kl}(\tilde{x}_k - \tilde{x}_l)^2$, where $\tilde{x}_k$ is the reconstruction value for the quantizer index i_k. Then, considering only single- and pair-variable cliques, the Gibbs energy of the MRF $\mathbf{I}$ based on quadratic clique potentials is

$$U(\mathbf{i}) = \sum_k \alpha_k \tilde{x}_k^2 + \sum_k \sum_l \beta_{k,l}(\tilde{x}_k - \tilde{x}_l)^2, \qquad (7)$$

where $\beta_{k,l} = 0$ if I_k and I_l are not neighbours. The local interaction coefficients α_k and $\beta_{k,l}$ are to be estimated from quantized sensor observations. It now follows that the conditional distribution of I_k, given its neighbors is

$$P(I_k = i_k|\mathbf{N}_k) = \frac{e^{\alpha'_k \tilde{x}_k^2 + \sum_l \beta'_{k,l}\tilde{x}_k\tilde{x}_l}}{\sum_{\tilde{x}_k} e^{\alpha'_k \tilde{x}_k^2 + \sum_l \beta'_{k,l}\tilde{x}_k\tilde{x}_l}}, \qquad (8)$$

where $\tilde{x}_l$ denote neighbours of $\tilde{x}_k$ and α'_k and $\beta'_{k,l}$ are constants that depend on α_k and $\beta_{k,l}$. Note that, the (discrete) local distribution given by (8) has the same form as the (continuous) local distribution of a Gauss-MRF (GMRF) [5]. Thus, this model is good when the observed spatial process is a GMRF and if the quantizers have sufficiently high resolution, in which case the pmf of a quantized variable is approximately proportional to the pdf of the quantizer input [7, Sec. 5.6].

Note that a grid-structure of sensors is not necessary to apply the MRF-model [6]. The only requirement is to define the neighborhood of each sensor, so that a graph satisfying the Markov property can be constructed.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, some experimental results demonstrating the performance of MRF-based joint decoders are presented. We model the spatial process observed by the sensor network according to $X_k = Y_k + W_k$, where $k = 1, \ldots, K$ denotes the sensor, $\{Y_k, k = 1, \ldots, K\}$ is a correlated Gaussian random field (GRF), and W_k is an iid Gaussian variable representing measurement noise of the sensor k, i.e., $E\{W_k W_l\} = 0$ for $k \neq l$ and $E\{W_k^2\} = \sigma_w^2$. All variables have zero mean. Note that, our source model is not an MRF, but a correlated random field- we simply model this by an MRF. The covariance matrix of the GRF $\mathbf{Y}$ is assumed to follow the power exponential model [8], $E\{Y_k Y_l\} = \sigma_y^2 e^{(-d_{kl}/\Delta)^2}$, where d_{kl} is the distance between the sensors k and l, and $\Delta > 0$ a constant. We use $\Delta = 0.28$ and choose σ_y^2 so that $\sigma_X^2 = E\{X_k^2\} = 1$. In this set-up, the signal-to-noise ratio (SNR) of each sensor is given by $\gamma = \sigma_y^2/\sigma_w^2$. We consider a network of $K = 240$ sensors, laid out on a grid of 16×15.

We use a 3-bit ($R_{Qk} = 3$) Lloyd-Max scalar quantizer for a unit-variance Gaussian variable in each sensor. The transmission rate of each sensor is reduced to 1 bit/sample ($R_{Tk} = 1$) using a binning function designed by considering non-overlapping clusters of 3 neighboring sensors (designing binning functions for larger clusters is too complex). The binning functions are optimized using the simulated annealing algorithm described in [9]. We assume that each sensor transmits its output to the decoder over a binary symmetric channel (BSC) with known bit error rate (BER).

In the following, we compare the performance of four types of decoders: (a) *MRF-Gibbs* joint decoder based on (6) and (7), (b) *MRF-CR* ("chain rule") joint decoder based on (5), (c) an array of *cluster decoders*- each of which operating on a cluster of 3 sensors used to design the binning functions [each decoder uses a look-up table of codewords optimized for 3 bits it receives from the 3 respective sensors], and (d) an array of *independent decoders* [one per sensor]. In the case of the MRF-Gibbs decoder, we use the 8-c neighborhood, while for the MRF-CR decoder the 4-c neighbour is used (due to complexity constraints).

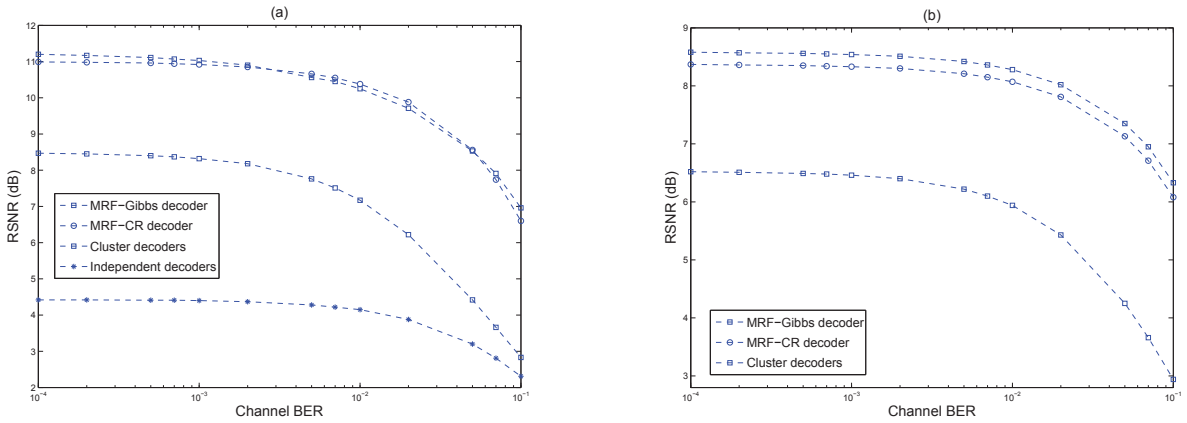

Fig. 3. A comparison of decoder performance: (a) sensor SNR $\gamma = 40$ dB and (b) $\gamma = 10$ dB. These estimates have been obtained by using a training set of 100,000 observation vectors.

In implementing the MRF-Gibbs decoder, we need to estimate the interaction coefficients α_k and β_{kl} from observed data (quantizer outputs). On the other hand for the MRF-CR decoder, we need to estimate the conditional probabilities in (5). In this paper, we assume that the MRF $\mathbf{I}$ is homogeneous (ignoring edge effects) and that the interaction coefficient for two variables I_k and I_l only depends on the distance between the sensors k and l. Thus, we have $\alpha_k = \alpha$ for all k and $\beta_{kl} = \beta^{HV}$ for horizontal and vertical cliques and $\beta_{kl} = \beta^{D}$ for diagonal cliques (we set $T = 1$). The parameter estimation was carried out using the *pseudo-likelihood* method [6].

In our experiments, we consider two scenarios: (i) $\gamma = 40$ dB, which corresponds to a high correlation between neighboring sensors and (ii) $\gamma = 10$ dB, which corresponds to a relatively weaker correlation among neighboring sensors. The performance of the decoders, in terms of reconstruction signal-to-noise ratio (RSNR) vs channel BER is shown in Fig. 3. In the case of $\gamma = 40$ dB [Fig. 3(a)], we note that the Gibbs factorization (which uses 8 neighbors) does not provide a noticeable improvement over the chain-rule expansion (which only uses 4 neighbours). The reason is that, as a result of high correlation, given the observations of vertical/horizontal neighbors, the quantized sensor outputs are nearly conditionally independent of the diagonal neighbors. This can be seen from the estimated values of interaction coefficients for $\alpha = -0.5$, $\beta^{HV} = 2.4$, and $\beta^{D} = 0.1$. On the other hand, for $\gamma = 10$ dB, we found that $\alpha = -0.35$, $\beta^{HV} = 0.6$, and $\beta^{D} = 0.2$. In this case [Fig. 3(b)], the lower SNR at the sensor inputs makes the quantized values less correlated (hence the lower RSNR of all decoders), and in turn makes the contribution of diagonal neighbours to the conditional pmf in (8) more significant. While the performance gain of Gibbs factorization over the chain-rule expansion in our experimental-case is modest, these results show that the former approach provides an effective means of constructing low-complexity decoders with extended neighbourhood structures which can improve the performance.

5. CONCLUDING REMARKS

MRF modeling is a simple but effective approach to constructing scalable joint decoders for sensor networks and related situations. While we used quadratic energy functions, there may be other choices [5] that fit different source models better. A useful extension to this work is the use of 3-dimensional MRF models in the decoder to exploit the spatio-temporal correlation in a sensor network.

6. REFERENCES

[1] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Mag.*, pp. 80–94, Sept. 2004.

[2] J. Barrõs and M. Tüchler, "Scalable decoding on factor trees: A practical solution for wireless sensor networks," *IEEE Trans. Commun.*, vol. 54, no. 2, pp. 284–294, Feb. 2006.

[3] R. Yasaratna and P. Yahampath, "Design of scalable decoders for sensor networks via Bayesian network learning," *IEEE Trans. Commun.*, under review.

[4] R. Kindermann and J. L. Snell, *Markov random fields and their applications*, Contemporary Mathematics. American Math. Soc., Providence, RI, 1980.

[5] J. Besag, "Spatial interactions and the statistical analysis of lattice systems," *J. Royal. Stat. Soc. B*, vol. 36, 1974.

[6] J. Besag, "Statistical analysis of non-lattice data," *The Statistician*, vol. 24, no. 3, 1975.

[7] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992.

[8] J. O. Berger, V. De Oliveria, and B. Sansó , "Objective Bayesian analysis of spatially correlated data," *J. American Stat. Assoc.*, vol. 96, pp. 1361–1374, Dec. 2001.

[9] P. Yahampath, "Predictive vector quantizer design for distributed source coding," in *Proc. IEEE ICASSP*, Apr. 2007, pp. III–629–632.