Distributed Source Coding

Hyun Sung Chang

May 17, 2006

1 Introduction

Having multiple sources, correlated with each other, to transmit, do we need more bits if they are separated than when they are together? This question was first raised in a seminal paper of Slepian and Wolf [1]. The answer is surprisingly "No," and what is even more interesting is that an important role of source compression is played by channel coding. This amazing result has been known as *Slepian-Wolf theorem* and followed by a fair amount of research works, e.g., extended by Wyner and Ziv [2] to lossy encoding of continuous-valued sources.

Unfortunately, however, the theorem has hardly found its place in practical applications for quite a long time. "Despite the existence of potential applications, the conceptual importance of distributed source coding has not been mirrored in practical data compression," said an article commemorating fifty years of information theory [3]. In fact, even a simple constructive coding example, beyond the theoretical and asymptotic result, had been unavailable until very recently a practical coding scheme was proposed by Pradhan and Ramchandran [4].

Since triggered by the work of Pradhan and Ramchandran, however, the Slepian-Wolf theorem has successfully redrawn attentions as a promising coding technique, particularly useful for modern applications such as distributed sensor network and low-power video coding, due to its capability of compressing distributed sources with low complexity at still high compression rates. This trend is well captured by the fact that the distributed source coding (DSC) solely comprises a regular session at many recent signal processing conferences.

2 Slepian-Wolf Theorem

The key result of Slepian-Wolf theorem can be summarized by the following: The minimal but still achievable amount of bits to represent two correlated sources X_1 and X_2 with little decoding error probability is equal to their joint entropy, $H(X_1, X_2)$, not increased by the separation between the sources. More precisely, a pair of rates (R_1, R_2), each required to represent X_1 and X_2, respectively, is achievable if and only if

$$R_1 \geq H(X_1|X_2) \ \text{ AND } \ R_2 \geq H(X_2|X_1) \ \text{ AND } \ R_1 + R_2 \geq H(X_1, X_2),$$

as illustrated on the R_1R_2 plane in Figure 1.

The achievability can be proven using the idea of "random encoding" and "jointly typical set decoding," while the proof for the converse part is quite straightforward [5]. Let us consider a corner point, $(R_1, R_2) = (H(X_1), H(X_2|X_1))$, in Figure 1. First, X_1 can be encoded, with R_1 bits, using a

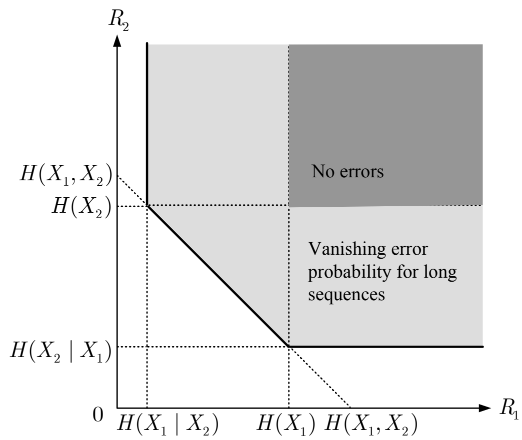

Figure 1: Slepian-Wolf theorem: admissible rate region for distributed compression of two statistically dependent i.i.d. sources X_1 and X_2. (Reproduced from [6])

single source compression scheme. Next, what the random encoding on X_2 actually does is to share 2^{nR_2} codewords among the typical X_2 sequences distinguishable if X_1 is additionally available at the decoder. In other words, the encoder decomposes X_2 into $(X_2', \tilde{X}_2)$ and encodes only $\tilde{X}_2$ with R_2 bits, discarding X_2'. Then, the decoder should guess X_2', with little error probability, using the correlation between X_1 and X_2'. This is possible from the channel coding theorem if we consider a virtual channel $(\mathscr{X}_2', p(x_1|x_2'), \mathscr{X}_1)$ because the input sequence whose rate is below the channel capacity, $I(X_1; X_2')$, can be recovered reliably. Note that $I(X_1; X_2') \leq I(X_1; X_2)$ and that R_2 goes minimum, i.e., to $H(X_2|X_1)$, when $I(X_1; X_2') = I(X_1; X_2)$. This argument suggests that the DSC is closely related to channel coding.

However, note that it tells only about the theoretical and asymptotic bound and not much about how to encode in practice. A subtle point here is about how to decompose X_2 into $(X_2', \tilde{X}_2)$ so that $I(X_1; X_2') = I(X_1; X_2)$. The next section will present some practical coding examples for specific types of correlation between X_1 and X_2 and also make the relationship to the channel coding more explicit.

3 Practical Coding

3.1 Modulo Encoding

Let X_1 and X_2 be scalar random variables correlated by $|X_1 - X_2| < \Delta_1/2$ for some integer Δ_1. The encoder compresses $\tilde{X}_2 := X_2 \mod \Delta_1$, rather than X_2, sharing the same codeword among all x_2's satisfying $\tilde{x}_2 = x_2 \mod \Delta_1$. Then, the required number of bits to encode $\tilde{X}_2$ is $\log \Delta_1$, equal to $H(X_2|X_1)$, if X_2 is uniformly distributed about X_1. The decoder can reconstruct X_2, without error, to $\tilde{X}_2 + \left\lfloor \frac{X_1 - \tilde{X}_2}{\Delta_1} + \frac{1}{2} \right\rfloor \cdot \Delta_1$ if X_1 is known.

Modulo-encoding fails to reach the Slepian-Wolf bound, i.e., $H(X_2|X_1)$, for other kinds of distribution. For instance, if X_2 is Gaussian with its mean and variance given by X_1 and $\Delta_1^2/36$, modulo-encoding falls below the bound by about half a bit [7].

3.2 Syndrome Encoding

Let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ be binary random vectors correlated by $d_H(\boldsymbol{X}_1, \boldsymbol{X}_2) \le \Delta_2$, where d_H denotes the Hamming distance. Any group of $\boldsymbol{x}_2$'s which differ from each other by $(2\Delta_2 + 1)$ or more bits can share the same codeword, say $\tilde{\boldsymbol{X}}_2$, because they become distinguishable if combined with $\boldsymbol{X}_1$. In other words, if $\boldsymbol{x}_2^{(1)}$, the actual value of $\boldsymbol{X}_2$, shares the codeword with $\boldsymbol{x}_2^{(2)}$, $d_H(\boldsymbol{x}_1, \boldsymbol{x}_2^{(2)}) \ge d_H(\boldsymbol{x}_2^{(1)}, \boldsymbol{x}_2^{(2)}) - d_H(\boldsymbol{x}_1, \boldsymbol{x}_2^{(1)}) \ge (2\Delta_2 + 1) - (\Delta_2) = \Delta_2 + 1$, by the triangle inequality, and thus $\boldsymbol{x}_2^{(2)}$ cannot come along with $\boldsymbol{x}_1$, leaving no room for confusion at the decoder if only $\boldsymbol{x}_1$ is available, because it is sufficiently far apart from $\boldsymbol{x}_2^{(1)}$. This notion of "separation" is exactly the same as appearing in the context of error correcting codes, and $\tilde{\boldsymbol{X}}_2$ corresponds to a coset index, called syndrome, in the linear block code [4].

Note that the above argument was made on binary sources. For continuous-valued sources, e.g., Gaussian sources, lossy coding based on rate-distortion theory [2] is used instead. In practice, it is considered equivalent to the quantization followed by the lossless Slepian-Wolf encoding.

3.3 Convolutional Encoding

For the same correlation condition as in the syndrome coding, i.e., $d_H(\boldsymbol{X}_1, \boldsymbol{X}_2) \le \Delta_2$, a convolution code can also be utilized instead of a linear block code. In this case, the error pattern of $\boldsymbol{X}_2$ behaves like a coset index in the linear block code. Therefore, it should be identified, e.g., using Viterbi decoding algorithm, and then encoded for transmission.

For lossy coding, more powerful channel coding techniques such as turbo code and low-density parity-check (LDPC) code may be exploited for greater bit saving, as found in [6],[7].

4 Remark: In Contrast to Conventional Source Coding

Finally, this section will briefly describe conventional joint source coding to highlight how the DSC is different from the conventional coding and why the DSC is preferred in a specific application, low-power video coding.

Let X_1 and X_2 be located together at a single source. In order to asymptotically achieve the rate of $H(X_2|X_1)$, X_2 may be encoded according to the conditional probability distribution $p(x_2|x_1)$. However, in many real situations, an encoder often needs to be non-asymptotic and agnostic to the conditional distribution. In this setting, a decorrelation process is usually taken for X_2 to extract components orthogonal to X_1, say $\tilde{X}_2$, from X_2. Because $\tilde{X}_2$ is not correlated with X_1 any longer, we expect $H(\tilde{X}_2)$ to approach $H(X_2|X_1)$.

For video coding example, the motion compensation is a kind of such decorrelation processes conducted to remove temporal redundancy between two frames. Indeed, it is the most complicated part in the video coding. For low-power video encoding, e.g., in wireless surveillance applications, a low-complexity coding scheme is quite desirable, while decoding can be done efficiently on a much more powerful machine. Because the DSC does not require such a decorrelation process, shifting the complexity to the decoder, it has been recognized suitable for the new applications. However, even with the latest practical coding schemes, there still exists a performance gap between the DSC and the conventional coding scheme [6], which should be further narrowed.

References

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory,* vol. 19, no. 4, pp. 471–480, July 1973.

[2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory,* vol. 22, no. 1, pp. 1–10, Jan. 1976.

[3] S. Verdú, "Fifty years of Shannon theory," *IEEE Trans. Inform. Theory,* vol. 44, no. 6, pp. 2057–2078, Oct. 1998.

[4] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," *IEEE Trans. Inform. Theory,* vol. 49, no. 3, pp. 626–643, Mar. 2003.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York, NY: John Wiley & Sons, 1991.

[6] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.

[7] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "A distributed source coding technique for correlated images using turbo-codes," *IEEE Commun. Lett.,* vol. 6, no. 9, pp. 379–381, Sep. 2002.