The Theory and Applications of Multiple Description Coding

Guanjun Zhang
University of Notre Dame

February 13, 2003

1 A first look of multiple descriptions

Loss of descriptions of stochastic processes transmitted over unreliable channels is typical in communication network. An intuitive approach to address this problem is to send more than one descriptions of the same sources and hope that at least one description can survive. Each description is supposed to be good enough to meet some decoder requirement if it get through itself. So these descriptions carry a lot of similar information and are dependent. However, if all these descriptions are received, we also hope the combined results to be as good as possible. This requires that the descriptions are far apart and thus can not be individually good. The multiple description problem refers to "if an information source is described by two or more separate descriptions, what are the concurrent limitations on qualities of these descriptions taken separately and jointly?"[1]. In other words, the fundamental problem is how can we achieve tradeoff between making descriptions individually good and sufficiently different.

A multiple description coding(MDC) model with two channels and three receivers is shown in figure 1. An encoder is given a symbol sequence yielded by a stochastic process $\{X_k\}$, where X_k's are independent identically distributed (i.i.d.) according to some known distribution $p(x)$. Two descriptions of the same sequence, of rate R_1 and R_2, are sent to decoder 1 and 2 through two channels, respectively. The central decoder 0 receives information over both two channels while the other two only receive information over their respective channels. All decoders have their own specific quality requirement in terms of distortion, denoted by D_0, D_1 and D_2.

One key theoretical problem in MDC was answered by El Gamal and T. Cover in [2]. Their proof gave a general bound, in usual Shannon sense ,for the achievable set of values of the quintuple $(R_1, R_2, D_0, D_1, D_2)$, thus determine what are the achievable rates R_1 and R_2 given distortions D_0, D_1 and D_2. Ozarow further showed a tight bound for memoryless Gaussian source with with mean-square error(MSE). Unfortunately, these bound can only be achieved by coding arbitrarily long sequences, which are not applicable in real network.

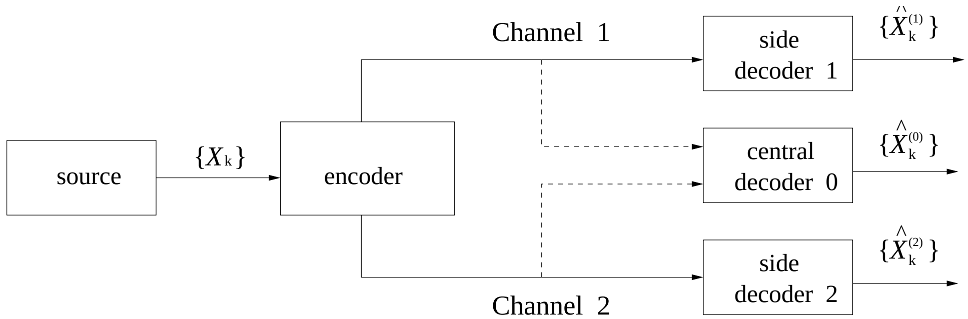

Figure 1: Multiple description coding model with two channels and three receivers.

Practical MDC algorithm are then proposed. They can be grouped into two catalogs: Multiple Description Scalar Quantization (MDSQ) and Multiple Description Correlating Transform (MDCT).

This report is organized in the following way: section 2 will cover background of shannon's rate distortion theorem and development of El Camel-Cover's proof of achievable rates in MD; section 3 summaries the basic ideas of MDSQ and MDCT and their applications in communications.

2 Conceptions and Theorem

2.1 Review of Shannon's rate distortion theorem

In real communication systems, distortion happens to information we are going to transmit due to limited energy, space and time people can spend. Quantization is widely used to reduce information rate to fit channel capacities while inevitably induces distortion. The basic problem in rate distortion theory is: given a source distribution and a distortion measure, what is the minimum rate description required to achieve a particular distortion? Or, equivalently, what is the minimum expected distortion achievable at a particular rate?[3]

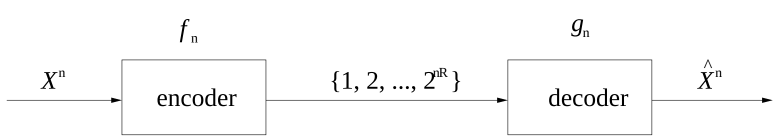

Figure 2: Normal point-to-point communication model.

Consider a normal point to point communication system, as shown in figure 2. Assume a discrete stochastic process produces sequence $X_1, X_2, \cdots, X_n$ with i.i.d. distribution

$p(x)$ and finite alphabet $\mathcal{X}$. The encoder describes the sequence X^n by an index from $\{1, 2, \cdots, 2^{nR}\}$ through an encoding function

$$f_n : \mathcal{X}^n \to \{1, 2, \cdots, 2^{nR}\} \tag{1}$$

The decoder reconstructs X^n by an estimate $\hat{X}^n$ through a decoding function

$$g_n : \{1, 2, \cdots, 2^{nR}\} \to \hat{\mathcal{X}}^n \tag{2}$$

We can define a symbol-to-symbol distortion function to measure the difference between X and $\hat{X}$.
Definition: A distortion function is a mapping

$$d : \mathcal{X}^n \times \hat{\mathcal{X}}^n \to R^+. \tag{3}$$

The above definition can be extended to sequence-based one in an average sense.
Definition: The distortion between sequence x^n and $\hat{x}^n$ is

$$d(x^n, \hat{x}^n) : \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i). \tag{4}$$

A rate distortion code $(2^{nR}, n)$ consists of f_n, g_n and a distortion measure D. Given the probability distribution of X, we can further define the distortion of the code as:
Definition: The distortion of code $(2^{nR}, n)$ is

$$\begin{aligned} D &= Ed(X^n, \hat{X}^n) \\ &= Ed(X^n, g_n(f_n(X^n))) \\ &= E\frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i) \\ &= \sum_{x_n} p(x^n) d(x^n, g_n(f_n(x^n))). \end{aligned} \tag{5}$$

Usually, a rate distortion pair (R, D) is said to be *achievable* if there exists a sequence of $(2^{nR}, n)$ codes with $Ed(X^n, g_n(f_n(X^n))) \leq D$ as $n \to \infty$. The closure of set of achievable (R, D) pairs consists a *rate distortion region*, which naturally leads to one important definition presented as following.
Definition: The rate distortion function $R(D)$ is the infimum of rates R such that (R, D) is in the rate distortion region of the source for a given distortion D.

The main theorem of rate distortion theory then points out that $R(D)$ can be related to the mutual information of original source and reconstructed representation.

Theorem: if $X_i,\ i = 1, 2, \cdots,$ are i.i.d. discrete finite alphabet random variables with probability mass function $p(x)$, then

$$R(D) = inf\, I(X; \hat{X}), \tag{6}$$

where $I(X; \hat{X})$ denotes Shannon mutual information of X and $\hat{X}$, and the infimum is taken over all joint probability mass function $p(x, \hat{x})$ such that

$$\sum_{x,\hat{x}} p(x, \hat{x}) d(x, \hat{x}) \leq D. \tag{7}$$

2.2 Achievable rate for multiple descriptions

The conceptions in the multiple description problem are defined as extension from the normal Shannon definitions. In case as shown in figure 1, suppose each decoder uses a specific distortion function d_0, d_1 and d_2, respectively. (R_1, R_2) is an achievable rate for distortion $\mathbf{D} = (D_0, D_1, D_2)$ if there exists a sequence of pairs of descriptions $f_n^{(1)} \in \{1, 2, \cdots, 2^{nR_1}\}$ and $f_n^{(2)} \in \{1, 2, \cdots, 2^{nR_2}\}$, and reconstructions $\hat{X}_1^n = g_n^{(1)}(f_n^{(1)})$, $\hat{X}_2^n = g_n^{(2)}(f_n^{(2)})$ and $\hat{X}_0^n = g_n^{(0)}(f_n^{(1)}, f_n^{(2)})$ such that $Ed_m(X_m^n, \hat{X}_m^n) \leq D_m$ with $m = 0, 1, 2$ as $n \to \infty$.

Then, the rate distortion region $R(\mathbf{D})$ in MD for given distortion $\mathbf{D}$ is the closure of the set of achievable rate pairs (R_1, R_2) which induce distortions no larger than $\mathbf{D}$. Recall the rate distortion function defined in the point-to-point communications, it is natural to think of what is the boundary of the multiple description rate distortion region.

The contribution of Cover and El Gamal to MD problem was to prove that there exists a bound of the rate distortion region and it can be related to the mutual information of sources and representations.

Theorem[2]: if $X_i,\ i = 1, 2, \cdots,$ are i.i.d. discrete finite alphabet random variables with probability mass function $p(x)$. Let $d_m,\ m = 0, 1, 2$ be bounded distortion measures. An achievable rate region for distortion $\mathbf{D} = (D_0, D_1, D_2)$ is given by the convex hull of all (R_1, R_2) such that

$$\begin{aligned} &R_1 > I(X; \hat{X}_1) \\ &R_2 > I(X; \hat{X}_2) \\ &R_1 + R_2 > I(X; \hat{X}_0, \hat{X}_1, \hat{X}_2) + I(\hat{X}_1; \hat{X}_2) \end{aligned} \tag{8}$$

for some joint probability mass function $p(x, \hat{x}_0, \hat{x}_1, \hat{x}_2)$ such that

$$\begin{aligned} &Ed_1(X_1^n, \hat{X}_1^n) \leq D_1 \\ &Ed_1(X_2^n, \hat{X}_2^n) \leq D_2 \\ &Ed_1(X_0^n, \hat{X}_0^n) \leq D_0. \end{aligned} \tag{9}$$

To prove the theorem, we might need to use the conception of jointly typical sequence (or strongly ϵ-typical sequence) and conditional jointly typical sequence. The Steps of proof are summarized as following:

- Construct jointly typical sequences, denoted by index, for individual channels 1 and 2 with rate R'_1 and R'_2, respectively;
- Construct conditional jointly typical sequence k for decoder 0 given each jointly typical sequences i and j in the last step;
- Encoding: find the all jointly typical sequences (i, j, k) of X, $\hat{X}_1$, $\hat{X}_2$ and $\hat{X}_0$. If such sequences can be found, then the distortions are satisfied. Separate the typical sequence in step 2 into two sequences k_1 and k_2. Then send (i, k_1) over channel one and (j, k_2) over channel two;
- Reconstruct $\hat{X}_1$ given (i, k_1), $\hat{X}_2$ given (j, k_2) and $\hat{X}_0$ given (i, j, k);
- Show that as length of sequence $n \to \infty$, probabilities of all types of errors in three decoders goes to zero, and the rates satisfy the boundary proposed in the theorem.

The proof of the theorem in length can be found in the same reference.

3 Applicable MDC

Today's packet network has the property of packet loss, which is very similar to the channel models used in MDC coding. Two main types of approaches have been proposed to exploit MDC to real network communications. Both assume that long block codes can not be used. Their objective is not to completely conquer the impairment of channels, but to obtain some satisfactory performance, whose requirement is less restricted than former.

3.1 MD Scalar Quantization(MDSQ)

MDSQ refers to the use of two separate scalar quantizers in source coding to yield two descriptions of a scalar source sample. A third decoder makes use of both descriptions. The main idea is to design the two scalar quantizers with different quantization steps. If the combined quantization steps are finer than those of the two individual ones, the third decoder can definitely recover more precise information with lower distortion.

Two examples of MDSQ are given in figure 3. The simplest form of MDSQ is presented in figure 3(a). Each quantizer, Q_1 and Q_2, output quantization index of the same sample, for example, $Q_1 = K_1$ and $Q_2 = K_2$. The reconstruction process of each side decoder will use the central values of cell $K_i,\ i = 1, 2$ to estimate the original value. While decoder 0 can center on the intersection of cell $K_i,\ i = 1, 2$, which means the central distortion is less than the others. If we denote the number of cells in decoder $i = 0, 1, 2$ as N_i, the design of quantizers in this example allow $N_0 = N_1 + N_2 - 1$. Asymptotically, if $R_1 = R_2 = R$, then D_0, D_1 and D_2 are all $O(2^{-2R})$.

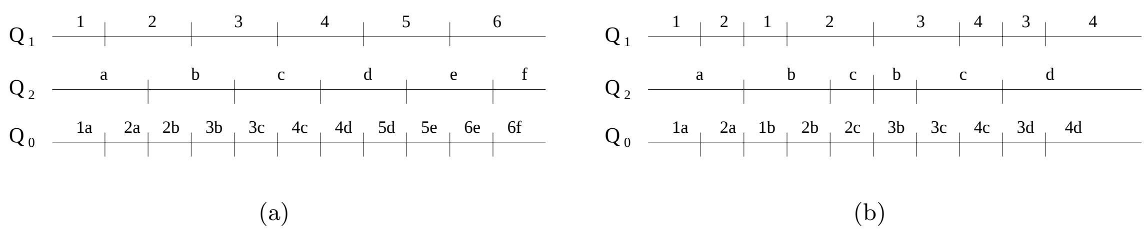

Figure 3: Examples of MD scalar quantizers. (a) The simplest form of MDSQ. (b) An MDSQ based on Vaishampayan's quantizer index assignment.

Obviously, if we can have N_0 increase faster, we can have lower central distortion. figure 3(b) shows an MDSQ example based on Vaishampayan's work [4]. Each quantizer uses less steps for them selves compared to 3(a), while the central decoder can still get similar distortion as 3(a). It systematically balances the central and side distortions while maintaining optimal joint asymptotic decay of distortion with rate. If the condition $R_1 = R_2 = R$ is kept, then D_0 $O(2^{-4R})$. A better central distortion is achieved with the price of lower side distortions.

3.2 MD Correlating Transform(MDCT)

Pairwise correlating transforms provide an alternative to achieve tradeoff in MDC. It was first introduced by Y. Wang, M. Orchard and A. Reibman in [5]. Use of MDCT is based on observations of the effect of correlation . Assume we have two independent zero-mean Gaussian random variables X_1 and X_2 with variances σ_1^2 and σ_2^2. Two channels are equally likely to fail. Uniform quantization, orthogonal transforms and MSE distortion measures are used. Instead of concerning individual distortion D_1 and D_2, we use the average distortion $\bar{D}_1 = \frac{1}{2}(D_1 + D_2)$. The central distortion is just the quantization distortion. If we introduce correlation between two random variables by using transform

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \tag{10}$$

it can be proven that the central distortion is increase of a factor of

$$\Gamma = \frac{(\sigma_1^2 + \sigma_2^2)/2}{\sigma_1 \sigma_2} \tag{11}$$

while the average side distortion is reduced by a related amount. By using specific orthogonal transforms, intermediate tradeoffs can be obtained.

V. Goyal et. al. gave optimal transforms to encode two-tuple sources with jointly Gaussian distributions in MDC and extent the transforms to generalized situations: sending N variables through M channels.[1]

4 Open Problems

The theoretical achievable rate distortion region of MDC is completely known only for a memoryless Gaussian source with distortion measure MSE[6]. The exact region for sources with other distributions has not been fully determined.

References

[1] Vivek K. Goyal and Jelena Kovacevic. Generalized multiple description coding with correlating transforms. *IEEE Trans. Info. Theory*, 47(6):2199 – 2224, 09 2001.

[2] A. El Gamal and Thomas M. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Info. Theory*, IT-28(6):851 – 857, Nov. 1982.

[3] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., U.S., 1991.

[4] V. A. Vaishampayan. Design of multiple description scalar quantizers. *IEEE Trans. Info. Theory*, 39:821 – 834, May 1993.

[5] Y. Wang M. Orchard and A. R. Reibman. Multiple description image coding for noisy channels by pairing transform coefficients. *Proc. IEEE Workshop Multimedia Signal Processing*, pages 419 – 424, June 1997.

[6] L. Ozarow. On a source-coding problem with two channels and three receivers. *Bell Syst. Tech. J.*, 59(10):1909 – 1921, Oct. 1980.