Streaming Real-Time Audio and Video Data with Transformation-Based Error Concealment and Reconstruction*

Benjamin W. Wah, Dong Lin and Xiao Su
Department of Electrical and Computer Engineering
and the Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
E-mail: {wah, dlin, xiao-su}@manip.crhc.uiuc.edu
URL: http://manip.crhc.uiuc.edu

Abstract

One fundamental problem with streaming audio and video data over unreliable IP networks is that packets may be dropped or arrive too late for playback. Traditional error control schemes are not attractive because they either add redundant information that may worsen network traffic, or rely solely on the inadequate capability of the decoder to do error concealment. In this paper, we propose a simple yet efficient transformation-based algorithm in order to conceal network losses in streaming real-time audio and video data over the Internet. In the receiver side, we adopt a simple reconstruction algorithm based on interpolation, as sophisticated concealment techniques cannot be employed in software-based real-time playback. In the sender side, we design a linear transformation with the objective of minimizing the mean squared error, assuming that some of the descriptions are lost and that the missing information is reconstructed by simple averaging at the destination. We further integrate the transformations in case of video streaming in the discrete cosine transform (DCT) to produce an optimized reconstruction-based DCT. Experimental results show that our proposed algorithm performs well in real Internet tests.

1. Introduction

Increases in bandwidth and computational speed lead to growing interests in real-time audio and video transmissions over the Internet. Unlike circuit-switched telephone networks, the Internet is a packet-switched, best-effort delivery service, with no guarantee on the quality of service. As a result, packets carrying real-time data may be dropped or arrive too late to be useful.

Traditional audio and video compression algorithms are not robust to transmission errors. The sole objective of compression is to maximize coding gain, assuming error-free channels. Most video coding schemes rely on temporal-difference coding to achieve coding efficiency, thereby introducing a pervasive dependency structure into the bit stream. Hence, losses due to dropped packets or late arrivals result in the loss of subsequent dependent frames, leading to visual artifacts that can be long lasting and annoying.

To deliver audio and video data over the Internet in real time with high quality, an active research area is to develop simple, robust error-concealment and coding strategies.

Error concealment found in the literature are based on either *redundant* or *nonredundant* transmissions.

a) One strategy of adding redundancies inserts error-correction codes into packets before transmitting them and recovers data in case of loss [1, 5]. It is not robust because it was designed based on an unknown channel model. Another strategy exploits the time constraints of applications and arranges retransmissions in such a way that additional delay will not cause significant perception degradation [8]. However, it incurs additional bandwidth that is already a scarce resource in real-time transmissions.

b) Nonredundant transmissions, on the other hand, recover lost data from that received within a tolerable range using inherent redundancies of source data. Examples of such schemes include replaying the last packet received and waveform substitution that searches received packets for waveform segments resembling those of missing packets [14]. These schemes do not work well when the durations of bursty losses are long or when there are dependen-

* Research supported by a grant from Conexant Systems, Inc.

Proc. First International Conference on Web Information Systems Engineering, June 2000.

cies among packets. More complex algorithms [15, 16] exploit source-data properties, such as edge orientations and geometric structures, in order to perform recovery. Besides computationally expensive, designing the encoder at the sender independent of the decoder at the receiver may not result in high-quality reconstruction because the two are usually closely related.

Robust coding algorithms can roughly be grouped into layered coding and multiple-description coding.

a) In layered coding [4], data is partitioned into a base layer and a few enhancement layers. In networks with priority support, the base layer is normally assigned a higher priority so that it has a larger chance to be delivered error free when network conditions worsen. Layered coding has been popular with ATM networks but may not be suitable for Internet transmissions for two reasons. First, the Internet does not provide priority delivery service for different layers. Second, when the packet-loss rate is high and part of the base layer is lost, it is hard to reconstruct the lost bit stream since no redundancy is present.

b) Multiple-description coding (MDC) divides audio and video data into equally important streams such that the decoding quality with any subset is acceptable, and that better quality is obtained by more descriptions. It is assumed that losses happen to different descriptions are uncorrelated, and that the probability of losing all the descriptions is small. MDC implemented [10, 13] generally have high coding efficiency when all the descriptions are received but have mediocre reconstruction quality with one description. They are not applicable to high-loss conditions for which the ultimate perceived quality may be dominated by the reconstructed quality from one description.

Our proposed system consists of an efficient MDC-based strategy with nonredundant error concealment. It splits adjacent samples into multiple streams and code each separately. In contrast to previous approaches that design coding schemes at the sender, independent of reconstruction methods at the receiver, we design them jointly. At the receiver, we adopt a simple reconstruction algorithm based on average interpolation to facilitate real-time playback. At the sender, we design a *linear transformation* to minimize reconstruction error when some of the streams at the receiver are lost and are reconstructed using average interpolation from other streams. This approach leads to high reconstruction quality than those using one description, with only moderate increase in bit rate (about 20% to 30%).

Figure 1 illustrates the result of applying the transformation proposed on a segment of 16 voice samples. The SNR based on the transformed (resp. original) even samples and the reconstructed odd samples is 7.16 dB (resp. 5.23 dB).

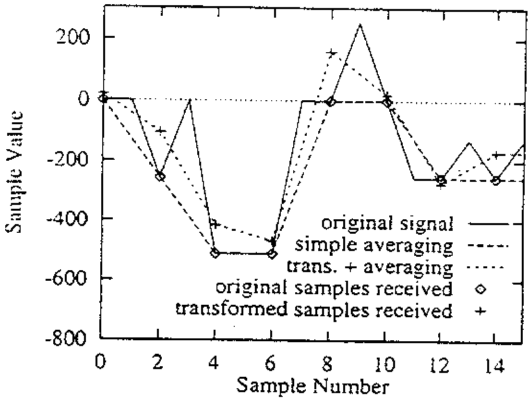


Figure 1. Reconstruction qualities between simple averaging (in dashed line) and averaging based on transformed samples (in dotted line), assuming only the even samples are received [3].

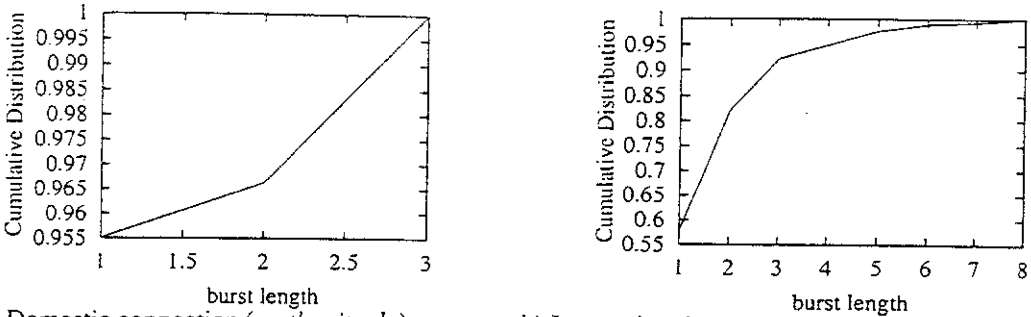
This paper is organized as follows. Section 2 studies packet-loss patterns of Internet transmissions and concludes that packet losses can be concealed effectively by interleaving and reconstruction. Sections 3 and 4 present in detail our transformation-based reconstruction algorithm and experimental results. Finally, Section 5 concludes the paper.

2. Internet Traffic Experiments

This section presents experimental results on loss characteristics for domestic and international connections.

During the experiments, the source computer periodically sent 2000 probe packets, at a rate of 100 packets per second and 500 bytes per packet, at the beginning of each hour over a 24-hour period to the echo port of each remote computer, and monitored the packets bounced back. (The sending rate and packet size were picked to reflect the upper bound on traffic in multimedia communications over the Internet. The packet size was picked to be smaller than the MTU of the Internet in order to avoid fragmentation. Results on other packet transmission rates can be found in the reference [3].) Statistics, such as sending and arrival times for each packet, was collected. To account for "delayed losses," each packet received had a scheduled "playback" time calculated from the arrival time of the first received packet and the difference of their sequence numbers. A packet was considered lost if it had been delayed by more than 200 msec of its scheduled playback time.

Our first set of experiments address the probability distribution of consecutive packet losses. Figure 2 clearly demonstrates that burst lengths of 1 and 2 are predominant. Even for the international connection with high losses, more than 80% of the losses were of burst length 1 or 2.



a) Domestic connection (*math.mit.edu*) b) International connection (*public.guangzhou.gd.cn*)

Figure 2. Probability distribution function of consecutive packet losses from *trace4.crhc.uiuc.edu.*

Table 1 lists for the international connection the conditional probability distribution of the next burst length, given the current burst length. For both sending rates, losses with burst length longer than 3 happened very infrequently, and one long burst did not imply that the next burst would also be long. For example, when packets were sent every 10 msec, the unconditional probability for the current burst length to be 4 and the next burst length to be greater than or equal to 4 is only $0.027 \times (1 - 0.938) \doteq 0.002$.

The fact that burst lengths are usually small (similar results have been shown in [2]) indicates that interleaving can be a good method to ease reconstruction. When the burst length is less than the interleaving factor, there are always parts of information received that can be used to recover the lost parts. For instance, with an interleaving factor of 2, a bursty loss of length 1 and a bursty loss of length 2 with samples belonging to different interleaving pairs can be recovered approximately. With an interleaving factor of 4, a bursty loss of length less than or equal to 3 and a burst length of 4, 5, and 6 with lost packets belonging to different interleaving sets can be recovered. In general, with an interleaving factor of i, it is possible to recover a bursty loss of length less than or equal to $i - 1$ and some of the bursty losses of length in the range $[i,\ (2i - 2)]$.

Let the total number of packets sent be n_p and the interleaving factor be i. Over all the interleaving sets, assuming that consecutive losses of length j, $j \le i$, happen $m_{i,j}$ times, and that n_s packets are lost (independent of i):

$$n_s = \sum_{j=1}^{i} j \times m_{i,j}. \tag{1}$$

We can derive $Pr(fail \mid loss, i)$, the conditional probability that a packet cannot be recovered for interleaving factor i. This happens when all the packets in an interleaving set are lost. From (1),

$$Pr(fail \mid loss, i) = \frac{i \times m_{i,i}}{n_s} \tag{2}$$

$Pr(fail \mid i)$, the unconditional probability that a packet cannot be recovered for interleaving factor i, is:

$$\begin{aligned} Pr(fail \mid i) &= Pr(fail \mid loss, i) \times Pr(loss) \\ &= Pr(fail \mid loss, i) \times \frac{n_s}{n_p} = \frac{i \times m_{i,i}}{n_p} \end{aligned} \tag{3}$$

Figure 3 plots $Pr(fail \mid i)$ for various interleaving factors and connections. $Pr(fail \mid i)$ drops quickly when the interleaving factor increases. For all times and both connections, $Pr(fail \mid i)$ is negligible when the interleaving factor is equal to or greater than 4. Moreover, an interleaving factor of 2 works well for domestic connections, achieving $Pr(fail \mid i)$ well below 3%. For the international connection (Figures 3a and 3b), an interleaving factor of 2 is not always enough because about 10-15% of the total losses will not be recoverable.

The above experimental results suggest that a small interleaving factor (between 2 to 4) is adequate. In most cases, an interleaving factor of 2 leads to good recovery.

3. Transformation-Based Reconstruction

In this section, we describe our proposed transformation-based reconstruction algorithm, with the objective of minimizing reconstruction error. Section 3.1 presents the derivations of a linear transformation of sound samples before interleaving in audio transmissions, while Section 3.2 describes the integration of the transformation with DCT in video transmissions in order to produce an optimized reconstruction-based DCT (ORB-DCT). We begin with an interleaving factor of two and extend to larger interleaving factors in Section 3.3.

3.1. Transformation Algorithm for Audio Signals

In our proposed method based on interleaving and average reconstruction, the sender performs interleaving and

Table 1. The correlation of burst lengths for $n_p = 20,000$ packets for the international connection to *public.guangzhou.gd.cn*. The data in each row represents the conditional distribution of the next burst length, given the current burst length listed in the first element in that row.

Burst length	Fraction of occurrence	Conditional probability distribution of the next burst length									
		1	2	3	4	5	6	7	8	9	≥10
Sending interval of 10 ms per packet and 500 bytes per packet (n_s = 7732)											
1	0.657	0.663	0.865	0.935	0.965	0.975	0.977	0.978	0.979	0.980	1.000
2	0.208	0.638	0.866	0.941	0.966	0.986	0.988	0.990			1.000
3	0.071	0.614	0.843	0.932	0.948	0.956	0.968	0.972		0.976	1.000
4	0.027	0.667	0.885	0.938	0.968	1.000					
5	0.012	0.733	0.889	0.933	0.978				1.000		
6	0.002	0.625	0.875		1.000						
7	0.002	0.800		1.000							
8	0.001	0.750		1.000							
9	0.001	0.667	1.000								
≥10	0.019	0.635	0.904	0.981						1.000	
Sending interval of 60 ms per packet and 500 bytes per packet (n_s = 6409)											
1	0.689	0.691	0.864	0.911	0.930	0.940	0.949	0.962	0.966	0.972	1.000
2	0.176	0.683	0.871	0.917	0.935	0.945	0.950	0.962	0.969	0.980	1.000
3	0.048	0.665	0.811	0.915	0.927	0.939	0.951	0.957	0.963	0.982	1.000
4	0.021	0.704	0.887	0.915	0.930	0.944	0.972				1.000
5	0.010	0.515	0.848		0.939		1.000				
6	0.008	0.690	0.793	0.931	0.966						1.000
7	0.011	0.784	0.973				1.000				
8	0.004	0.800	0.933		1.000						
9	0.007	0.750	0.917	1.000							
≥10	0.026	0.674	0.880	0.946	0.956						1.000

distributes related information in different packets, and the receiver reconstructs lost samples using the average of adjacent samples received [9]. Although the method improves the quality in the presence of isolated losses, the amount of aliasing distortions may be large when input signals have high frequency response and cannot be reconstructed accurately by simple averaging.

To improve the reconstruction quality, the sender transforms each original sample into a new sample before interleaving, packetization, and transmission (Figure 4). It transforms the samples in such a way that the reconstructed samples at the receiver will be the best approximation to the original ones on the average with respect to SNR, in case that half of the samples are lost.

There are two cases to be considered at the receiver: when only one of the two packets in an interleaving pair is received, and when both packets are received.

Case I: One packet in an interleaving pair is lost. Assume that the sender transforms the original data stream $\vec{x} = x_0, x_1, ..., x_{2N-1}$ (assuming $x_{2N} = 0$) into $\vec{y} = y_0, y_1, \cdots, y_{2N-1}$ by transformation T (unknown yet), and that the receiver only receives half of the stream. Without loss of generality, assume that all even samples $\vec{y}_{even} = y_0, y_2, \cdots, y_{2N-2}$ are received. After average reconstruction, the reconstructed stream, $\vec{\hat{y}} = \hat{y}_0, \hat{y}_1, ..., \hat{y}_{2N-1}$, is calculated as follows:

$$\hat{y}_i = \begin{cases} y_i, & i \text{ even} \\ \frac{y_{i-1}+y_{i+1}}{2} & i \text{ odd and } i \neq 2N-1 \\ \frac{y_{2N-2}}{2} & i = 2N-1 \end{cases} \tag{4}$$

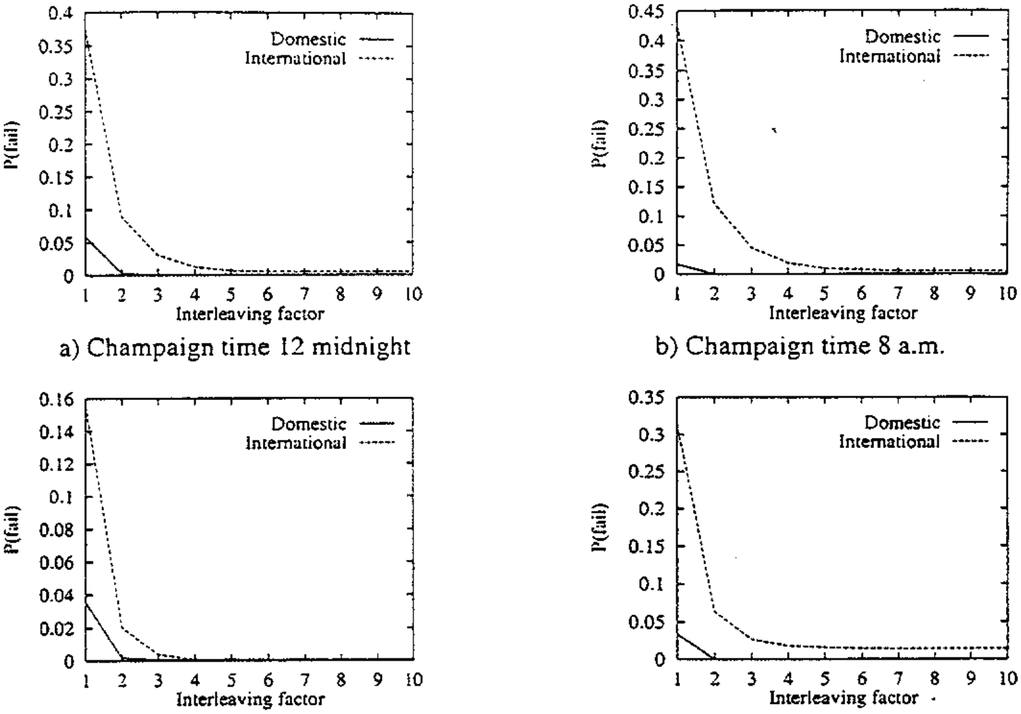
$\mathcal{E}_o$, the reconstruction error, is defined as:

$$\mathcal{E}_o = \sum_{i=0}^{2N-1} (x_i - \hat{y}_i)^2 = \sum_{n=0}^{N-1} (x_{2n} - y_{2n})^2 + \sum_{n=0}^{N-2} \left(x_{2n+1} - \frac{y_{2n}+y_{2n+2}}{2}\right)^2 + \left(x_{2N-1} - \frac{y_{2N-2}}{2}\right)^2. \tag{5}$$

To minimize $\mathcal{E}_o$, y_i, for any even i, must satisfy:

$$\frac{\partial \mathcal{E}_o}{\partial y_i} = 0, \quad i = 0, 2, \cdots, 2N-2. \tag{6}$$

After substituting $\mathcal{E}_o$ in (5) into (6), we get the following



a) Champaign time 12 midnight

b) Champaign time 8 a.m.

c) Champaign time 12 noon

d) Champaign time 8 p.m.

Figure 3. $Pr(fail \mid i)$**, probabilities of bursty losses that cannot be recovered, under various interleaving factors.**

matrix transformation:

$$\vec{y}_{even} = \begin{pmatrix} y_0 \\ y_2 \\ \vdots \\ y_{2N-4} \\ y_{2N-2} \end{pmatrix} = \mathbf{T} \times \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{2N-2} \\ x_{2N-1} \end{pmatrix} \qquad (7)$$

where

$$\mathbf{T} = \mathbf{A}^{-1} \times \mathbf{B} = \begin{pmatrix} 1 & \frac{1}{5} & 0 & \cdots & 0 & 0 & 0 \\ \frac{1}{6} & 1 & \frac{1}{6} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{6} & 1 & \frac{1}{6} \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{6} & 1 \end{pmatrix}^{-1} \qquad (8)$$

$$\times \begin{pmatrix} \frac{4}{5} & \frac{2}{5} & & & \cdots & & & & & \\ 0 & \frac{1}{3} & \frac{2}{3} & \frac{1}{3} & \cdots & & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & \cdots & \frac{1}{3} & \frac{2}{3} & \frac{1}{3} & & \\ & & & & \cdots & & & \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

In a way similar in deriving $\vec{y}_{even}$ in (7), the transformation for $\vec{y}_{odd}$ can be obtained.

Case II: Both packets in an interleaving pair are received. In this case, we can apply the following inverse transformation to restore $\vec{x}$ by combining the transformation for even and odd samples.

$$\vec{x} = \begin{pmatrix} \frac{4}{5} & \frac{2}{5} & & & & & & & & \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{3} & & & & & & & \\ 0 & \frac{1}{3} & \frac{2}{3} & \frac{1}{3} & & & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & \frac{1}{3} & \frac{2}{3} & \frac{1}{3} & \\ & & & & & & & \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ & & & & & & & & \frac{2}{5} & \frac{4}{5} \end{pmatrix}^{-1} \qquad (9)$$

$$\times \begin{pmatrix} 1 & 0 & \frac{1}{5} & 0 & & & & & & & \\ 0 & 1 & 0 & \frac{1}{6} & & & & & & & \\ \frac{1}{6} & 0 & 1 & 0 & \frac{1}{6} & & & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & \frac{1}{6} & 0 & 1 & 0 & \frac{1}{6} \\ & & & & & & & \frac{1}{6} & 0 & 1 & 0 \\ & & & & & & & & \frac{1}{5} & 0 & 1 \end{pmatrix} \vec{y}$$

After computing $\vec{x}$ using (9) at the receiver, the original stream $\vec{x}$ can be restored perfectly if $\vec{y}$ does not have any precision loss during processing and transmission.

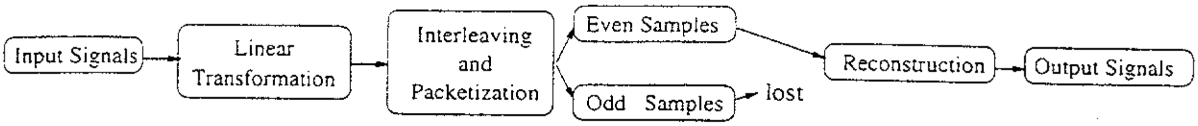


Figure 4. Process of transformation and reconstruction in two-way interleaving.

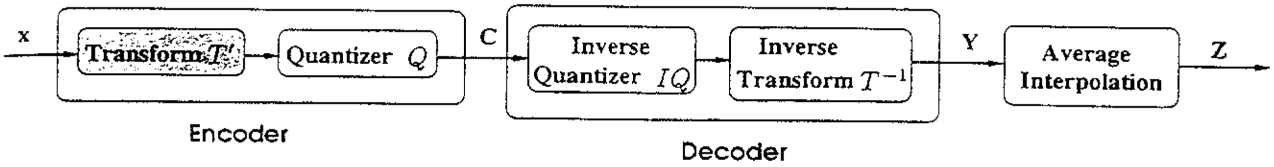


Figure 5. Basic building blocks of a modified codec. (The shaded block is our proposed ORB-DCT.)

3.2. Optimized Reconstruction-Based DCT for Video Signals

Assume that video data is partitioned into two streams, and that one of them is lost during transmission. In such a scenario, the original DCT and quantizer designs are not necessarily the best for reconstructing the lost data.

Figure 5 shows a simplified diagram of the basic building blocks in most state-of-the-art codecs. Our goal is to find a new transform T' in order to minimize the reconstruction error after interpolation, based on fixed quantization Q, inverse quantization IQ, and inverse DCT T^{-1}. That is,

$$\mathcal{E}_r \;=\; \| \underbrace{Interpolate(T^{-1}(IQ(\mathbf{c})))}_{decompression\,+\,reconstruction} - \mathbf{x} \|^2 . \qquad (10)$$

We assume inverse quantization IQ and inverse DCT T^{-1} to be fixed in order to keep our decoders standard-compliant. Consequently, our proposed transform coder can be used in real-time video-on-demand applications with standard-compliant decoders.

With quantization in place, the minimization of $\mathcal{E}_r$ becomes an integer optimization problem, where $\mathbf{c}$ in (10) takes integer values. Such optimizations are computationally prohibitive in real time. In the following, we derive an approximate solution that does not take into account quantization effects. Specifically, the objective to be optimized in the approximation is:

$$\mathcal{E}_r \;=\; \| \, Interpolate(T^{-1}(\mathbf{c})) - \mathbf{x} \, \|^2 . \qquad (11)$$

In the following, we derive ORB-DCT for intra-coded and inter-coded blocks separately.

3.2.1 ORB-DCT for Intra-Coded Blocks

Assume that the original frame is divided into blocks of size 8×16 pixels. After ORB-DCT, block $\mathbf{X}$ is transformed into two blocks $\mathbf{C}_1$ and $\mathbf{C}_2$, each of size 8×8, corresponding to blocks of odd-numbered and even-numbered pixels, respectively. Since the derivations are similar, we only show the derivations for $\mathbf{C}_1$.

Our objective is to find $\mathbf{C}_1$ in order to minimize (11). After inverse DCT, output $\mathbf{Y}_1$ can be calculated as follows:

$$\mathbf{Y}_1 = \sum_{i=1}^{8}\sum_{j=1}^{8} C_{i,j}\mathbf{b}_i\mathbf{b}_j^T , \qquad (12)$$

$$\text{where} \quad \mathbf{b}_i = \left\{ \frac{1}{2}\alpha_i \cos\frac{(2k-1)(i-1)\pi}{16} \right\}_{k=1,2,\ldots,8} ,$$

$C_{i,j}$ is the $(i,j)^{th}$ element in $\mathbf{C}_1$, $\mathbf{b}_i$ is the i^{th} basis vector of DCT, $\alpha_1 = \frac{1}{\sqrt{2}}$, and $\alpha_i = 1$ for $i = 2, 3, \ldots, 8$.

Putting (12) in matrix form gives:

$$\mathbf{Y}_1 = (\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_8)_{8\times 8} , \qquad (13)$$

$$\text{where} \quad \mathbf{p}_k = \sum_{i=1}^{8}\sum_{j=1}^{8} C_{i,j}\mathbf{b}_i b_{j,k} \quad k = 1, \ldots, 8, (14)$$

$b_{j,k}$ is the k^{th} component of basis vector $\mathbf{b}_j$. The interpolated pixels $\mathbf{Z}$ is then obtained by inserting even-numbered columns as the average of columns from $\mathbf{Y}_1$, with the boundary column duplicated:

$$\mathbf{Z} = \left(\mathbf{p}_1 \quad \frac{\mathbf{p}_1 + \mathbf{p}_2}{2} \quad \mathbf{p}_2 \quad \frac{\mathbf{p}_2 + \mathbf{p}_3}{2} \quad \cdots \quad \mathbf{p}_8 \quad \mathbf{p}_8 \right)_{8\times 16} . \qquad (15)$$

$\mathbf{Z}$ can also be expressed as:

$$\mathbf{Z} = \sum_{i=1}^{8}\sum_{j=1}^{8} C_{i,j}\mathbf{b}_i\mathbf{e}_j^T , \qquad (16)$$

$$\text{where} \quad \mathbf{e}_j = \left(b_{j1} \quad \frac{b_{j1} + b_{j2}}{2} \quad b_{j2} \quad \ldots \quad b_{j8} \quad b_{j8} \right)^T (17)$$

We define $\mathbf{e}_j$ as an extended basis vector for reconstruction purpose. The distortion between the original and the received and reconstructed pixels is:

$$\mathcal{E}_r = \left\| \sum_{i=1}^{8} \sum_{j=1}^{8} C_{i,j} \mathbf{b}_i \mathbf{e}_j^T - \mathbf{X} \right\|^2 . \qquad (18)$$

To minimize $\mathcal{E}_r$ with respect to $\mathbf{C}$, we first linearize each matrix into a vector by raster-scan order, *i.e.*, following the first row by the second row in a matrix, and so on. The following notations are defined after linearization:

$$\begin{aligned} \vec{\mathbf{u}} &= (C_{i,j})_{(8\times 8)} \\ \vec{\mathbf{v}}_{8(i-1)+j} &= \mathbf{b}_i \mathbf{e}_j^T{}_{(8\times 16)} \qquad i,j = 1,2,\ldots,8. \\ \vec{\mathbf{w}} &= (X_{i,j})_{(8\times 16)} \end{aligned}$$

We further define matrix $\mathbf{V}$ as:

$$\mathbf{V} = (\vec{\mathbf{v}}_1 \quad \vec{\mathbf{v}}_2 \quad \vec{\mathbf{v}}_3 \quad \ldots \quad \vec{\mathbf{v}}_{64}) . \qquad (19)$$

Then (18) can be rewritten as follows:

$$\mathcal{E}_r = \| \mathbf{V}\vec{\mathbf{u}} - \vec{\mathbf{w}} \|^2 , \qquad (20)$$

where $\mathbf{V}$ is a 128×64 matrix, $\vec{\mathbf{u}}$, a 64×1 vector, and $\vec{\mathbf{w}}$, a 128×1 vector. Since the linear system of equations $\mathbf{V}\vec{\mathbf{u}} = \vec{\mathbf{w}}$ is an over-determined one, there exists at least one least-square solution $\vec{\mathbf{u}}$ that minimizes (20) according to the theory of linear algebra [6]. Specifically, the solution $\vec{\mathbf{u}}$ with the smallest length $|\vec{\mathbf{u}}|^2$ can be found by first performing SVD decomposition of matrix $\mathbf{V}$:

$$\mathbf{V} = \mathbf{S}\ [diag(w_j)]\ \mathbf{D}^t, \qquad j = 1,2,\ldots,64, \qquad (21)$$

where $\mathbf{S}$ is a 128×64 column-orthogonal matrix, $[diag(w_j)]$, a 64×64 diagonal matrix with positive or zero elements (singular values), and $\mathbf{D}$, a 64×64 orthogonal matrix. Then the least-square solution can be expressed as:

$$\vec{\mathbf{u}} = \mathbf{D}\ [diag(1/w_j)]\ \mathbf{S}^T\ \vec{\mathbf{w}}. \qquad (22)$$

In the above diagonal matrix $[diag(1/w_j)]$, the element $1/w_j$ is replaced by zero if w_j is zero. Therefore, ORB-DCT is a product of three matrices: $\mathbf{D}\ [diag(1/w_j)]\ \mathbf{S}^T$.

To derive the ORB-DCT transform for $\mathbf{C}_2$, simply replace $\mathbf{e}_j, j = 1,2,\ldots,8$, in (17) by the following:

$$\mathbf{e}_j = \left(b_{j1} \ \ b_{j1} \ \ \frac{b_{j1}+b_{j2}}{2} \ \ b_{j2} \ \ \ldots \ \ \frac{b_{j7}+b_{j8}}{2} \ \ b_{j8} \right)^T .$$

The rest of the steps are similar.

3.2.2 ORB-DCT for Inter-Coded Blocks

For inter-coded blocks, output $\mathbf{Y}_1$ after inverse DCT, as shown in (12), is the residual block after motion prediction. Denote its corresponding reference block as:

$$\mathbf{R} = (\mathbf{r}_1 \quad \mathbf{r}_2 \quad \ldots \quad \mathbf{r}_8)_{8\times 8} . \qquad (24)$$

Then the interpolated data $\mathbf{Z}$ is the sum of two terms after motion compensation:

$$\begin{aligned} \mathbf{Z} &= \left(\mathbf{p}_1 \quad \frac{\mathbf{p}_1+\mathbf{p}_2}{2} \quad \mathbf{p}_2 \quad \frac{\mathbf{p}_2+\mathbf{p}_3}{2} \quad \ldots \quad \mathbf{p}_8 \quad \mathbf{p}_8 \right) \\ &+ \left(\mathbf{r}_1 \quad \frac{\mathbf{r}_1+\mathbf{r}_2}{2} \quad \mathbf{r}_2 \quad \frac{\mathbf{r}_2+\mathbf{r}_3}{2} \quad \ldots \quad \mathbf{r}_8 \quad \mathbf{r}_8 \right) \\ &= \sum_{i=1}^{8} \sum_{j=1}^{8} C_{i,j} \mathbf{b}_i \mathbf{e}_j^T + \mathbf{R}'. \end{aligned} \qquad (25)$$

Substituting the above equation into (18) results in the reconstruction error for inter-coded blocks.

$$\mathcal{E}_r = \left\| \sum_{i=1}^{8} \sum_{j=1}^{8} C_{i,j} \mathbf{b}_i \mathbf{e}_j^T - (\mathbf{X} - \mathbf{R}') \right\|^2 . \qquad (26)$$

To derive ORB-DCT in this case, we note that only vector $\vec{\mathbf{w}}$ is different as compared to the case of intra-coded blocks. From (22), it is obvious that the transform itself does not depend on $\vec{\mathbf{w}}$; therefore, ORB-DCT retains the same form.

In short, a uniform transform of ORB-DCT exists for both intra- and inter-coded blocks. For intra-coded blocks, it is applied to an original block $\mathbf{X}$ to produce transform coefficients $\mathbf{C}_i, i = 1,2$; whereas for inter-coded blocks, it is applied to interpolated motion-predicted blocks $(\mathbf{X} - \mathbf{R}')$.

Like DCT, ORB-DCT is also a row-column-separable transform. To compute a transform coefficient of ORB-DCT by a row-column approach, it takes 40 floating-point multiplications and 37 floating-point additions. In the future, we plan to study fast implementations of ORB-DCT, similar to what was done in deriving fast DCT.

3.3. Transformations to cope with longer bursts

In the above derivations, audio and video signals are assumed to be partitioned into two streams. Statistics in Section 2 suggests that domestic sites generally have bursty losses of one packet in duration, whereas international sites may have bursty losses of three or more packets. Hence, an interleaving factor of 2 is not always enough, but 4 is generally adequate.

To handle burst lengths of four, our proposed transformation can be extended in two ways. First, we can derive an optimal transformation assuming that only one stream is received. The transformations obtained this way are overly pessimistic because they assume that three streams are always lost. In practice, it is possible for no, one, two, or three streams to be lost when sent in the Internet.

Second, we can construct 4-way interleaving using a combination of 2-way interleaving [12]. For instance, in streaming video signals, we first interleave the original frame $\tilde{\mathbf{x}}$ in the horizontal direction into two streams, $\tilde{\mathbf{x}}_{h1}$

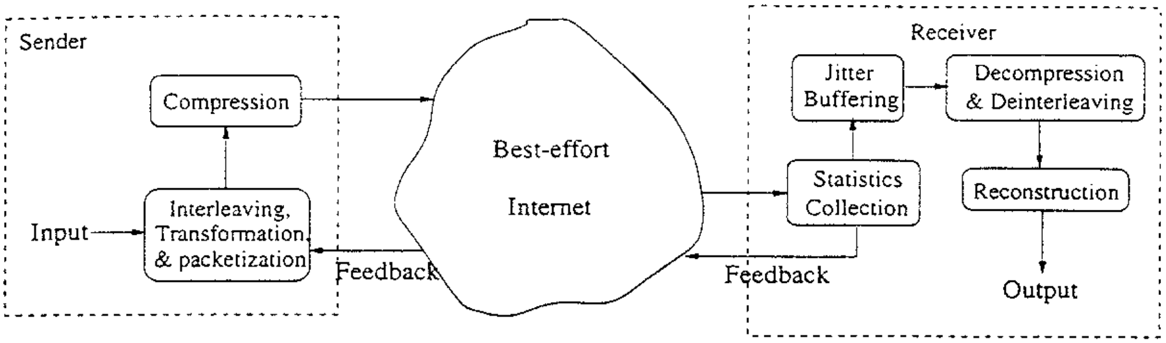


Figure 6. Transmission system prototype with feedback on loss statistics.

and $\tilde{x}_{h2}$, and transform them. Similarly, we perform interleaving and transformations in the vertical direction and get two additional streams. The four streams, $\tilde{z}_{h1,v1}$, $\tilde{z}_{h1,v2}$, $\tilde{z}_{h2,v1}$ and $\tilde{z}_{h2,v2}$, are then sent in distinct packets to the receiver. If one of the streams is missing in a direction, then reconstructions are carried out by interpolation of the stream received. If both streams are missing in a direction, then reconstructions are carried out by interpolation of the stream reconstructed/received in the other direction.

In a similar way, the decomposition of audio signals into 4-ways interleaving is done by a combination of two 2-way interleaving [11].

4. Experimental Results

In this section, we evaluate our proposed reconstruction method by prototyping both video and audio systems and by performing tests on the Internet under realistic loss behavior. Figure 6 shows the components of our real-time transmission system that has compression (ADPCM and H.263), interleaving, reconstruction, statistics collection, and buffering to remove jitters [7].

For a fair evaluation under the same traffic conditions, we did trace-driven simulations by applying the same trace of packets lost in real Internet transmissions on transformed and original sequences.

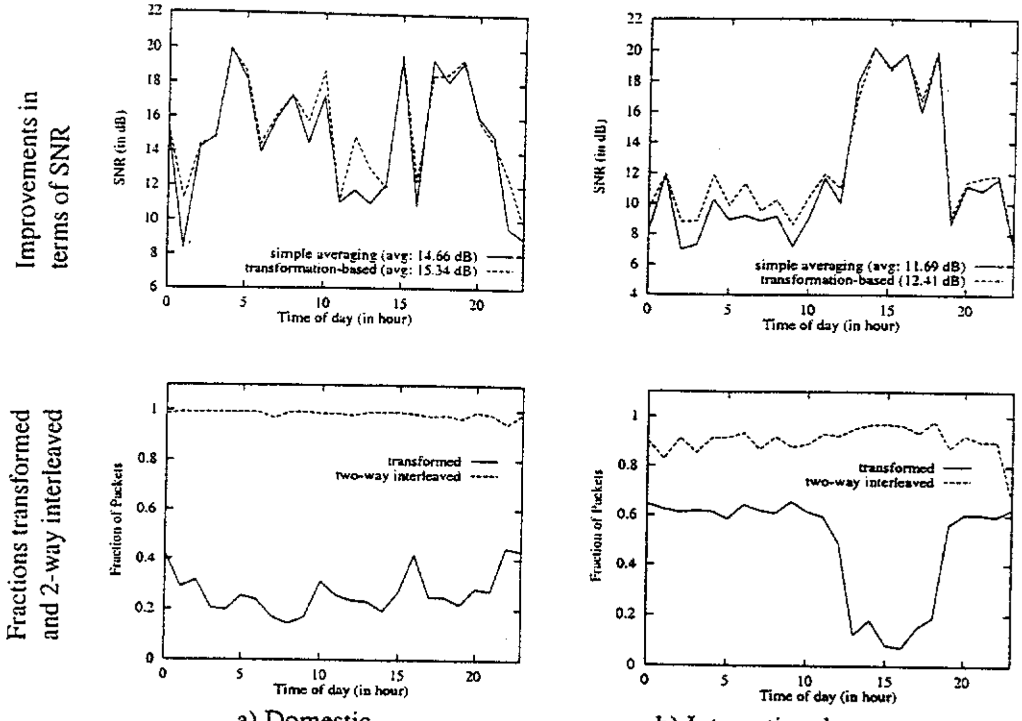
In collecting traffic traces, we sent 512-byte packets for video and 250-byte packets for audio periodically from our home site in Champaign to remote echo servers at a rate of 20 packets/sec for video and 32 packets/sec for audio. From the packets echoed back, we recorded the sequence numbers and sending and arrival times and determined packet losses based on the sequence numbers recorded. The packet-loss rate estimated was likely to be pessimistic since each packet traversed a round trip. We set the jitter-buffer size to be comparable to the standard deviation of packet inter-arrival times so that any packet that arrived later than its scheduled arrival time plus the jitter time was considered lost.

4.1. Experiments on Audio Streaming

Obviously, reconstruction quality depends on the loss rate and whether transformation was performed by the sender. When losses are low, the sender should determine a priori whether to transform input samples before they are sent, whereas when losses are high, transformations almost always help reduce reconstruction errors, provided that a suitable interleaving factor is chosen. Hence, the receiver in our prototype sends run-time statistics on losses and burst lengths to the sender periodically. Based on this information, the sender chooses the best interleaving factor and whether to perform transformation before sending data to the receiver.

The results are shown in Figure 7. The top graphs compare the SNR between simple averaging and reconstruction based on transformed input data. Using transformed input data, the reconstruction quality was almost always better or equal. For the domestic connection with low losses, reconstruction based on transformed data can achieve about the same quality as reconstruction based on simple averaging. In this case, most of the quality loss is due to compression. For the international connection with high losses, reconstruction based on transformation can improve over reconstruction based on simple averaging by about 0.7 dB on the average, with a peak improvement of over 2 dB.

The solid lines in the bottom graphs of Figure 7 measure the fraction of packets that were transformed. For the domestic connection, around 20% of the packets were transformed. In contrast, for the international connection, up to 65% of the packets were transformed. The dashed lines plot the fraction of packets that were two-way interleaved (the remaining packets were four-way interleaved). For the domestic connection, almost all packets were two-way interleaved, whereas 10%-30% of the packets might be four-way interleaved for the international connection when loss rates were high or burst lengths were larger than 2.



a) Domestic

b) International

Figure 7. Reconstruction qualities on audio streaming over a 24-hour period for the domestic and international connections.

4.2. Experiments on Video Streaming

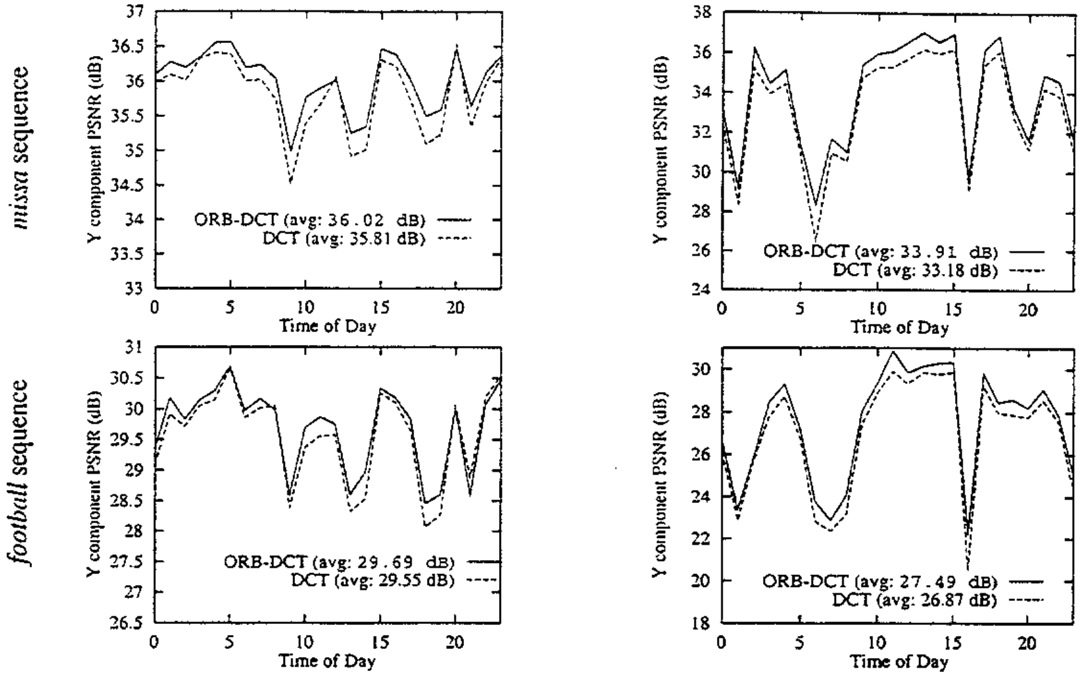
Our experiments were done using two video sequences in CIF (352 × 288) YUV format: *missa* (Miss America) consisting of 150 frames in a typical video conferencing sequence with slow head-and-shoulder movements, whereas *football* has 90 frames in a fast-motion movie.

Our experiments to apply traffic traces consist of a sender process and a receiver process. The sender process was in charge of compressing and packetizing video frames, and mapping packet losses to GOB losses of each frame. The number of streams (2 or 4) was set periodically every 0.5 sec at the sender according to feedback information on GOB losses of frames from the receiver. In our simulations, we assume that the receiver collected GOB-loss information every 0.5 sec before sending the information to the sender, and that the network delay was constant at 0.5 sec. The receiver process was in charge of decompressing coded streams, deinterleaving them and performing reconstruction. For every GOB of each frame, any lost information was reconstructed by average interpolation using adjacent pixels. The reconstructed frame was sent back to the decoder as a reference for future inter-coded frames. If the entire GOB was lost, it was reconstructed by copying the corresponding GOB from the last received frame.

Figure 8 compares the reconstruction quality over a 24-hour period for both domestic and international connections. For the domestic (resp. international) connection, the new ORB-DCT transform yields comparable playback quality, with 0.15 – 0.2 dB (resp. 0.4 – 0.7 dB) improvement in PSNR on the average. The higher improvement for the international connection is expected since our transformation is designed to optimize reconstruction performance under high and bursty loss scenarios.

5. Conclusions and Future Work

In this paper, we have proposed and applied a new transformation-based reconstruction algorithm for real-time voice and video streaming over the Internet and have found consistent improvement in quality of the data received. The transformations are derived based on the way signals are reconstructed at the receiver and the loss behavior in the network. They are general and can be extended easily to other interpolation-based reconstruction algorithms. Our future work is focused on extending the method to low-bit rate compression methods, such as CELP, that involves new objectives not based on SNR and on developing transformations under bandwidth constraints.



a) Domestic

b) International

Figure 8. Reconstruction qualities on video streaming over a 24-hour period for the domestic and international connections.

References

[1] A. Albanese, J. Bloemer, and J. Edmonds. Priority encoding transmission. In *Proc. Foundations of Computer Sciences*, pages 604-612, Santa Fe, NM, 1994.

[2] J. C. Bolot. Characterizing end-to-end packet delay and loss in the Internet. *High-Speed Networks*, 2(3):305–323, Dec. 1993.

[3] D. Lin. *Real-Time Voice Transmissions over the Internet*. M.Sc. Thesis, Dept. of Electrical and Computer Engineering, Univ. of Illinois, Urbana, IL, Dec. 1998.

[4] M.Normura, T.Fujii, and N.Ohta. Layered packet-loss protection for variable rate coding using DCT. In *Proc. of Int'l Workshop on Packet Video*, Sept. 1988.

[5] H. Ohta and T. Kitami. A cell loss recovery method using fec in ATM networks. *IEEE Journal on Selected Areas in Communications*, 9(9):1471-1483, December 1991.

[6] L. Rade and B. Westergren. *Mathematics Handbook for Science and Engineering*. Studentlitteratur Birkhauser, 1995.

[7] R. Ramjee, J. Kurose, D. Towsley, and H. Schulzrinne. Adaptive playout mechanisms for packetized audio applications in wide-area networks. In *Proc. 13th Annual Joint Conf. IEEE Computer and Communications Societies on Networking for Global Commmunication*, volume 2, pages 680–688, 1994.

[8] I. Rhee. Error control techniques for interactive low-bit rate video transmission over the Internet. In *Proc. SIGCOMM'98*, 1998.

[9] J. Suzuki and M. Taka. Missing packet recovery techniques for low-bit-rate coded speech. *IEEE Journal on Selected Areas in Communications*, 7(5):707–717, June 1989.

[10] V. A. Vaishampayan. Design of multiple description scalar quantizer. *IEEE Trans. on Information Theory*, 39(3):821-834, May 1993.

[11] B. W. Wah and D. Lin. Transformation-based reconstruction for real-time voice transmissions over the internet. *IEEE Trans. on Multimedia*, 1(4):342–351, Dec. 1999.

[12] B. W. Wah and X. Su. Streaming video with transformation-based error concealment and reconstruction. In *Proc. Int'l Conf. on Multimedia Computing and Systems*, volume 1, pages 238–243. IEEE, June 1999.

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

[14] O. J. Wasem, D. J. Goodman, C. A. Dvordak, and H. G. Page. The effect of waveform substitution on the quality of PCM packet communications. *IEEE Trans. on Acoutics, Speech, and Signal Processing*, 36(3):342–348, Mar. 1988.

[15] W.Kwok and H.Sun. Multi-directional interpolation for spatial error concealment. *IEEE Trans. on Consumer Electronics*, 39(3):455-460, Aug. 1993.

[16] W. Zeng and B. Liu. Geometric structure based directional filtering for error concealingment in image/video transmission. In *Proc. SPIE Wireless Data Transmission at Information Systsmes/Phontonics East'95*, pages 145-156, October 1995.