DPCM IN VIDEO CODING

Though not usually called DPCM, the most commonly used methods of video coding use a form of DPCM, e.g. MPEG, H.26X, HDTV, satellite "Direct TV", DVD.

Prediction is done on frame-by-frame basis, with prediction of a frame being the decoded reconstruction of previous frame, or a "motion-compensated" version thereof. In latter case, coder has a forward-adaptive component in addition to the backward adaption.

Example: (Netravali & Haskell, Digital Pictures, p. 331)

<table>
<thead>
<tr>
<th>Prediction Coefficients</th>
<th>MSPE</th>
<th>Pred. gain¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>left pixel, curr. frame</td>
<td>left pixel prev. frame</td>
<td>same pixel prev frame</td>
</tr>
<tr>
<td>1</td>
<td>-1/2</td>
<td>1/2</td>
</tr>
<tr>
<td>3/4</td>
<td>-1/2</td>
<td>3/4</td>
</tr>
<tr>
<td>7/8</td>
<td>-5/8</td>
<td>3/4</td>
</tr>
</tbody>
</table>

¹Based on educated guess that $\sigma^2 = 1000.$

DESIGN AND PERFORMANCE -- HIGH RESOLUTION ANALYSIS

Warning: Though this analysis is almost universally accepted, it has never been satisfactorily proven to be correct.

Assume source is stationary, zero-mean random process.

It can be shown that under ordinary conditions $\{(X_i,U_i,V_i,Y_i)\}$ is asymptotically stationary. So we assume it is stationary. Therefore,

$$D = E(X_i-Y_i)^2 = E(U_i-Q(U_i))^2$$ (same for all i)
Key assumption:
When $R$ is large and quantizer is well designed

$$D \equiv E(\tilde{U}_i - Q(\tilde{U}_i))^2$$

where $\tilde{U}_i = X_i - g(X_{i-1}, X_{i-2}, \ldots)$. $\tilde{U}_i \neq U_i$ but $\tilde{U}_i \equiv U_i$

Note: $\tilde{U}$ is a stationary random process.

Key assumption implies

least possible distortion of DPCM with rate $R$

$\equiv$ least possible distortion of SQ with rate $R$ encoding $\tilde{U}$, minimized over choice of predictor $g$

It follows that for FLC or VLC,

$$\delta_{dpcm}(R) \equiv \min_g \delta_{\tilde{U}, sq}(R) \equiv \frac{1}{12} \left( \min_g \sigma_{\tilde{U}}^2 \alpha_{\tilde{U}} \right) 2^{-2R}$$

where

$$\alpha_{\tilde{U}} = \begin{cases} \beta_{\tilde{U}}, & \text{for FLC} \\ \eta_{\tilde{U}}, & \text{for VLC} \end{cases}$$

Conclusions:

A good (but not necessarily optimal) way to design DPCM is to

(a) Choose $g$ to minimize $\sigma_{\tilde{U}}^2 \alpha_{\tilde{U}}$

(b) Choose $q$ to achieve $\delta_{\tilde{U}, sq}(R)$

Common situation:

The pdf of $\tilde{U}$ is similar to a scaled version of that $X$, which implies $\beta_{\tilde{U}} \equiv \beta_X$ and $\eta_{\tilde{U}} \equiv \eta_X$. Indeed, if $X$ is Gaussian source, then so is $U$, and these approximations become exact. In this case (a) becomes

(a') Choose $g$ to minimize $\sigma_{\tilde{U}}^2$

Therefore,

$$\delta_{dpcm}(R) \equiv \frac{1}{12} \mathcal{M} \alpha_X 2^{-2R}, \quad \text{where } \mathcal{M} = \min_g \sigma_{\tilde{U}}^2 = \text{MSPE}$$

$$S_{dpcm}(R) \equiv S_{sq}(R) + G_N \text{ dB}, \quad \text{where } G_N = 10 \log_{10} \frac{\sigma^2}{\mathcal{M}} \text{ dB} = \text{pred'n gain}$$

That is, gain of DPCM over SQ approximately equals the prediction gain.
EXAMPLE: FIRST-ORDER AUTOREGRESSIVE (AR) GAUSSIAN SOURCE.

Assume $X$ is stationary and

$$X_i = \rho X_{i-1} + Z_i$$  \hspace{1cm} (1)

where $-1 < \rho < 1$, $Z_i$'s are IID Gaussian with zero mean and variances $\sigma_Z^2$. $Z_i$ is independent of $X_j$, $j < i$. ($\rho = .9$ is a typical value.) Then

1) $EX_i = 0$:
   
   Derivation: Let $m = EX_i$. It does not depend on $i$ since $X$ is stationary.
   Taking expected values of both sides of (1) gives
   
   $$m = \rho m + 0,$$
   
   which implies $m = 0$.

2) $\sigma_X^2 = \frac{\sigma_Z^2}{1-\rho^2}$  \hspace{1cm} ($= 5.26 \sigma_Z^2$ if $\rho = .9$)

   Derivation: Let $s = EX_i^2$. Taking expected square value of both sides of (1) gives,
   
   $$s = \rho^2 s + 2\rho EX_i Z_i + \sigma_Z^2.$$

   Since $X_i-1$ and $Z_i$ are independent,
   
   $$EX_i-1 Z_i = EX_i-1 EZ_i = 0.$$

   Therefore, $s = \rho^2 s + \sigma_Z^2 \Rightarrow s = \frac{\sigma_Z^2}{(1-\rho^2)}.$$

3) $R_X(k) =$ autocorrelation function $= E X_iX_{i+k} = \sigma_X^2 \rho^{|k|}$

   $\Rightarrow \rho =$ correlation coefficient of $X = \frac{\text{cov}(X_iX_{i+1})}{\sigma_X^2}$

   Derivation: $R_X(0) = E X_i^2 = \sigma_X^2$. We now show $R_X(k) = \rho R_X(k-1)$.

   $$R_X(k) = E X_iX_{i-k} = E (\rho X_{i-1} + Z_i)X_{i-k} = \rho E X_{i-1}X_{i-k} + EZ_iX_{i-k} = \rho R_X(k) + 0.$$

4) Best predictor for $X_i$ based on $X_{i-1}, X_{i-2}, ...$ is $\tilde{X}_i = \rho X_{i-1}$ with MMSPE $= \sigma_Z^2$

   Derivation: It's not easy to use $a = K^{-1}r$. However, we may use the
   orthogonality principle, which indicates that $\tilde{X}_i$ is optimal if and only if
   
   $$E (X_i - \tilde{X}_i)X_{i-j}, \text{ for } j = 1, ..., N.$$

   We find $E (X_i - \rho X_{i-1})X_{i-j} = E (\rho X_{i-1} + Z_i - \rho X_{i-1})X_{i-j} = E Z_iX_{i-j}$

   $$= EZ_i EZ_{i-j} = 0, \text{ since } Z_i \text{ indep. of past } X's.$$

   It follows from orthog. principal that $\tilde{X}_i = \rho X_{i-1}$ is optimal.

   MMSPE $= E(X_i - \rho X_{i-1})^2 = E(\rho X_{i-1} + Z_i - \rho X_{i-1})^2 = E Z_i^2 = \sigma_Z^2$

   Therefore, for $N \geq 1$, the prediction gain (which is the gain of DPCM over SQ) is

   $$G_N = 10 \log_{10} \frac{\sigma_X^2}{\text{MMSPE}} = 10 \log_{10} \frac{1}{1-\rho^2} \Rightarrow 7.2 \text{ dB if } \rho = .9.$$
Consider a stationary, Gaussian source and large $R$.

For $k$-dimensional transform coding:
$$\delta_{\text{tr}}(k,R) \equiv \frac{1}{12} |K^{(k)}|^{1/k} \alpha_2 2^{-2R}$$

For DPCM with $k$th-order linear prediction
$$\delta_{\text{dpcm}}(k,R) \equiv \frac{1}{12} \mathcal{M}_k \alpha_2 2^{-2R}$$

where $\mathcal{M}_k$ is MSPE of optimal $k$th-order linear prediction for $X_I$ from $X_{i-k}, \ldots, X_{i-1}$.

Fact A: $|K^{(k)}|^{1/k} \geq \mathcal{M}_k$.

Proof: $|K^{(k)}|^{1/k} = (\alpha_X^2 \prod_{i=1}^{k-1} \mathcal{M}_i)^{1/k}$ this was proved in the transform coding notes
$$\geq \mathcal{M}_k \quad \text{because all terms being averaged are } \geq \mathcal{M}_k$$

It follows from this fact that DPCM with $k$th-order prediction is at least as good as $k$-dimensional transform coding.
Fact B: \[
\lim_{k \to \infty} M_k = \lim_{k \to \infty} |K^{(k)}|^{1/k} = \exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln S(\omega) \, d\omega\right\}
\]

Proof: First we note that since the \( M_k \)'s are nonnegative and nonincreasing, they converge to a limit. Secondly,
\[
|K^{(k)}|^{1/k} = \left(\sigma_X^2 \prod_{i=1}^{k-1} M_i\right)^{1/k},
\]
and the latter expression converges to \( \lim_{k \to \infty} M_k \) because when one takes the geometric average of the first \( k \) terms of a convergent sequence (the \( M_k \)'s), that geometric average converges to the same value as the sequence. Hence,
\[
\lim_{k \to \infty} |K^{(k)}|^{1/k} = \lim_{k \to \infty} M_k
\]

It follows from this fact that for large \( k \), DPCM and Transform Coding have the same performance.

However, consider the example of a first-order AR Gaussian source.

Then, since \( M_1 = \lim_{k \to \infty} M_k = Q \), "one-step prediction error"
\[
\delta_{dpcm}(1,R) \equiv \delta_{tr}(\infty,R)
\]

The SNR gain in dB of DPCM with kth-order linear prediction over k-dimensional transform coding is
\[
10 \log_{10} \frac{\delta_{tr}(1,R)}{\delta_{dpcm}(k,R)} = 10 \log_{10} \frac{(1-\rho^2)^{(k-1)/k}}{1-\rho^2} = \frac{10}{k} \log_{10} (1-\rho^2)
\]

Similarly, for an Nth-order AR Gaussian source,
\[
\delta_{dpcm}(N,R) \equiv \delta_{tr}(\infty,R)
\]
WHAT COULD BE WRONG WITH THE HIGH-RESOLUTION ANALYSIS?

For large $R$, we assumed $E(U-Q(U))^2 \equiv E(\tilde{U}_i-Q(\tilde{U}_i))^2$.

This is obviously true because both sides are approximately 0.

But it hasn't been shown that $E(U-Q(NU))^2 / E(\tilde{U}_i-Q(NU)^2) \to 1$ as $N \to \infty$.

So it is not clear that as $R$ increases, the pdf of $\tilde{U}$ becomes so similar to that of $U$ that $\delta_{U,sq}(R)/\delta_{U,sq}(R) \to 1$.

Since both terms in the ratio go to zero, it wouldn't take much of a difference between the terms to make the ratio not converge to one.

WHAT IF THE TRADITIONAL HIGH-RESOLUTION ANALYSIS IS WRONG?

I don't think it could be wrong by much.

Nitadori made an analysis of DPCM that takes into account the fact that $\tilde{U}$ contains quantization as well as prediction errors. See Jayant and Noll, p. 278, 279. I don't believe this analysis has been checked thoroughly for correctness.

Naraghi-Pour (UM Thesis, 1987) has shown that the best predictor is not always the predictor that minimizes the $E\tilde{U}^2$.

DESIGN WITHOUT THE HIGH-RESOLUTION ASSUMPTION

A "hypothetical" design algorithm.

Make an initial choice of the predictor and iterate the following two steps.

1. Optimize quantizer for a given predictor
   a) Make an initial choice of the quantizer $Q^{(0)}$.
      Let $k = 1$.
   b) Find the steady-state pdf $f^{(k)}(u)$ of $U$ assuming the quantizer is $Q^{(k-1)}$.
      (There is an iterative algorithm for doing this.)
   c) Let $Q^{(k)}$ be the optimal quantizer for pdf $f^{(k)}(u)$.
   d) Let $k = k+1$ and go to b).

2. Optimize predictor for a given quantizer
   a) Make an initial choice of predictor $a^{(0)}$.
      Let $k = 1$.
   b) Find the steady-state moments of $X_i, Y_{i-1}, ..., Y_{i-N}$: $EX_i Y_{i-j}$'s, $EY_{i-m} Y_{i-n}$'s.
   c) Let $a^{(k)}$ be the optimal predictor for these statistics.
   d) Let $k = k+1$ and go to b).
FURTHER NOTES:

• It is not clear that for an optimal DPCM system, $Q$ is the optimal quantizer for $U$.
  One might think that if $Q$ weren't optimal for $U$, then one could improve DPCM by making it optimal for $U$. However, changing $Q$, changes the pdf of $U$, so it's not clear that changing $Q$ would help.

• Similarly, it is not clear that for an optimal DPCM system, $g$ is the optimal predictor for $X$ based on $Y_{i-N},...,Y_{i-1}$.
  One might think that if $g$ weren't optimal for $X$, then one could improve DPCM by making it optimal. However, changing $g$, changes the moments of the $Y$'s, so it's not clear that changing $g$ would help.

• DPCM can be made backward adaptive in various ways, e.g. by adapting the quantizer or the predictor.

• Jayant and Noll is an excellent reference on fancier versions of DPCM.

ANALYSIS WITHOUT THE HIGH-RESOLUTION ASSUMPTION

• There are iterative algorithms for computing actual distortion.