Compression of Video Signals

Basic Concepts and Some Recent Advances in the Field

Yuriy A. Reznik

Codec Technologies Group
RealNetworks, Inc.
2601 Elliott Avenue, Suite 1000
Seattle, WA 98121

E-mail: yreznik@real.com

Outline

1. Introduction
   - A bit of history
   - Structure of a hybrid DCT+DPCM coder
   - Today's standards-based and proprietary algorithms
   - Pace of the progress (Girod's law)
2. Basic Concepts and Techniques
   - DPCM, Motion compensation
   - Transform - based coding
   - Quantization
   - Noiseless coding
3. Conclusions

A Bit of History

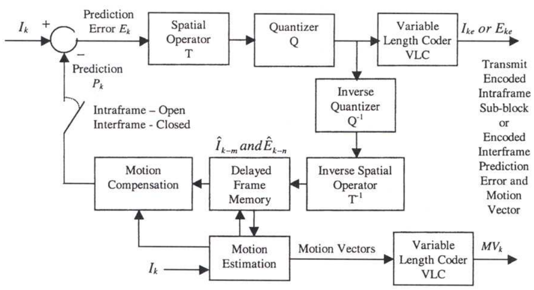
1. First techniques for coding of signals:
   - PCM (Reeves, 1938)
   - Delta-modulation (Deloraine, 1946)
   - DPCM (Cutlet, 1952)
2. Transforms
   - KLT (Karhunen & Loeve, 1948)
   - DFT $\sim$ KLT! (Pearl, 1973)
3. Intra-frame image/video coding
   - DCT (Ahmed, Natarajan, Rao, 1974)
4. Simple frame-difference coding (DPCM, scalar quantizer)
   - H.120 (1984)

A Bit of History (Cont'd)

5. First successful hybrid DPCM/DCT scheme
   - H.261 (1991)
     - motion compensation
6. Incremental improvements:
   - MPEG-1 (1993), MPEG-2/H.262 (1994)
     - 1/2-pixel motion compensation
   - H.263 (1996)
     - block-size adaptive motion compensation
   - H.26L (1999) $\rightarrow$ MPEG-4 AVC/H.264
     - 4x4 integer transforms
     - 1/4-pixel motion compensation
     - improved entropy coding

Structure of DPCM/DCT Codecs

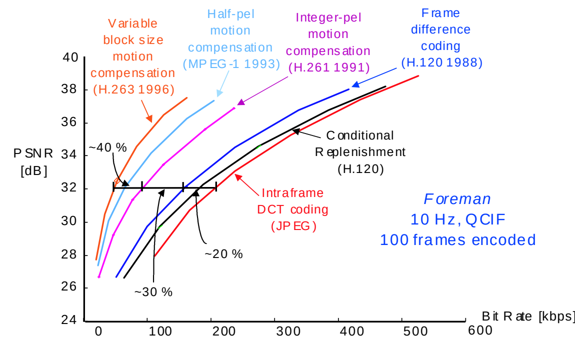
- Generalized structure (H.261,H.263,MPEG-1,2):



- Notation:
  - I_k – k-th input frame
  - P_k – predictor constructed using previous reconstructed frames $\hat{I}_{k-1} \ldots \hat{I}_{k-m}$
  - $E_k = I_k - P_k$ – prediction residual
  - $\hat{E}_k$ – reconstructed prediction residual

Pace of the Progress

- Performance of video codecs is improving by 0.5 dB every 12 to 18 months
  - B. Girod (1998)



- Performance metrics
  - I, $\hat{I}$ – original/reconstructed images, containing $m \times n$ pixels, of d bits each;
  - $MSE = \frac{1}{m\,n} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(I_{i,j} - \hat{I}_{i,j}\right)^2$;
  - $PSNR = 20 \log_{10} \left(\frac{2^d}{\sqrt{MSE}}\right)$

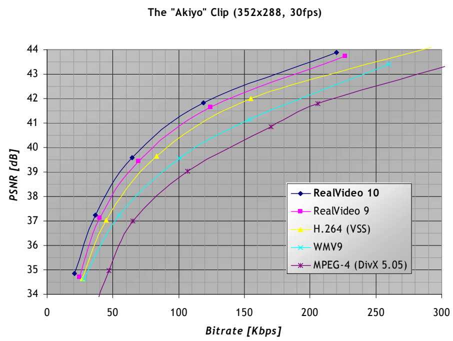
Key Improvements Since H.261:

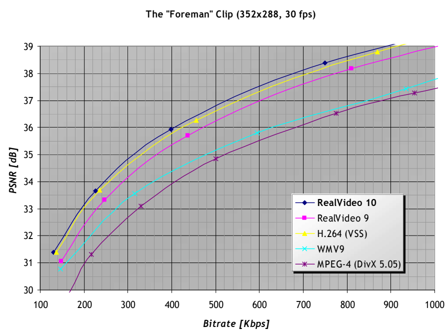
1. Half-pixel accurate motion compensation
   - +1.0..1.5dB gain (QCIF, 10Hz, 100Kbps)
   - introduced in MPEG-1 (1993)
2. Block-size adaptive motion compensation
   - +0.5..1.5dB gain (QCIF, 10Hz, 100Kbps)
   - introduced in H.263(Annex F) (1996)
3. 1/4-pixel-accurate motion compensation
   - +0.5..1.0dB gain (CIF, 24Hz, 1000Kbps)
   - proposed in MPEG-4 (ACE-profile) (1999)
   - different scheme in H.264/MPEG-4 AVC (2001)
4. 4x4 integer transforms + improved compander
   - +0.5..1.5dB gain (QCIF, 10Hz, 100Kbps)
   - introduced in H.26L (1999), improved in H.264/MPEG-4 AVC (2001).
5. Improved entropy coding schemes
   - 15-25% bitrate reduction (QCIF, 10Hz, 36dB)
   - first realized in RealVideo 8 (1999)
   - different schemes H.264/MPEG-4 AVC (2001)

Current Algorithms

- Standards-based:
  - MPEG-4 AVC / H.264 codecs (2003)
    - derivatives of the H.26L project (1999)
  - MPEG-4 (original version, 2001)
    - derivative of MPEG-2, H.263
    - not quite competitive these days
- Proprietary:
  - RealVideo 10 – RealNetworks
  - WMV 9 – Microsoft
  - VP6 – On2
- MPEG-4 wrapped in proprietary formats:
  - DivX, XdiV

Current Algorithms (RD-Plots)





Basic Concepts and Techniques

- Basic techniques:
  1. Predictive coding
     - Linear prediction
     - Motion compensation
  2. Transform - based coding
  3. Quantization
     - RD-theory
     - Direct vs. Successive quantization
  4. Noiseless coding
     - Coding of known sources
     - Universal coding

(items 1–3: skipped)

Noiseless Coding (Overview)

- Consider the following:

  $A = \{a, b, c\}$ - alphabet, $|A| = 3$

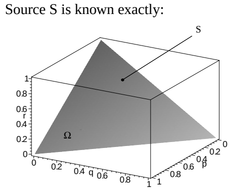
  Ω - a class of memoryless sources over A;

  $\Pr(a) = p$; $\Pr(b) = q$; $\Pr(c) = r$; $(p+q+r = 1)$

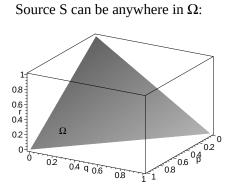
  $S \in \Omega$ - source to be encoded.

- Types of problems in noiseless coding:

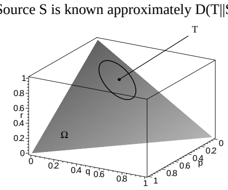
Source S is known exactly:



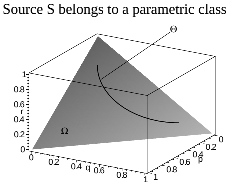
Source S can be anywhere in Ω:



Source S is known approximately D(T||S) < δ:



Source S belongs to a parametric class Θ:



Noiseless Coding (Definitions)

- We have the following:
  - $S \in \Omega$ - known memoryless source:

    $$P_S(\alpha) := \Pr(S \rightsquigarrow \alpha);$$

  - w – a word $|w| = n$ produced by S:

    $$P_S(w) = \prod_{\alpha \in A} P_S(\alpha)^{r_\alpha(w)},$$

    where $r_\alpha(w)$ denotes the number of letters α in w. Indeed: $\sum_{\alpha \in A} r_\alpha(w) = |w|$.
  - A **block code** ϕ:

    $$\phi : A^n \to \{0, 1\}^*$$

    where the output set is *decipherable*.

- Key parameters:
  - The **entropy** of source S:

    $$H(S) = -\sum_{\alpha \in A} P_S(\alpha) \log P_S(\alpha).$$

  - The **average cost** of code ϕ:

    $$C(\phi, n, S) = \frac{1}{n} \sum_{w \in A^n} P_S(w) |\phi(w)|.$$

  - The **average redundancy**:

    $$R(\phi, n, S) = C(\phi, n, S) - H(S).$$

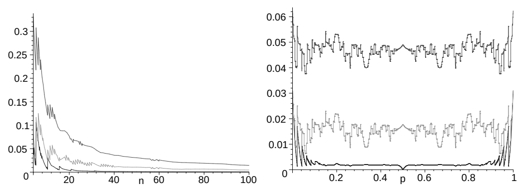
Noiseless Coding (Known Source)

- Redundancy rate decreases with n as:

$$R(\phi_S, n, S) = \frac{C}{n}$$

where C is a constant.

- Examples ($A = \{0, 1\}$, $\log_2 \frac{\Pr(1)}{\Pr(0)}$ is irrational):
  - Shannon code: $C_S = 1/2 + o(1)$
  - Huffman code: $C_H = 3/2 - 1/\ln 2 + o(1) \approx 0.0573..$
  - Gilbert-Moore code: $C_{GM} = 3/2 + o(1)$.



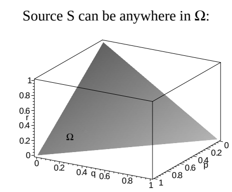
Average redundancy rates of Gilbert-Moore, Shannon, and Huffman block codes (left: fixed source $\Pr(0) = 1/8$; right: fixed block $n = 32$, $p = \Pr(0)$).

Noiseless Coding (Unknown Source)

- We know only that $S \in \Omega$

- The solution is to find a code ϕ_Ω:

$$R(\phi_\Omega, n, \Omega) = \inf_{\phi \in \Phi} \sup_{S \in \Omega} R(\phi, n, S) \ .$$



- The code ϕ_Ω is called a **universal block code** for a class of sources Ω.

- Achievable redundancy rate (Krichevsky,1975):

$$R(\phi_\Omega, n, \Omega) = \frac{1}{n}\left[\frac{|A|-1}{2}\log n + C\right] = O\left(\frac{\log n}{n}\right)$$

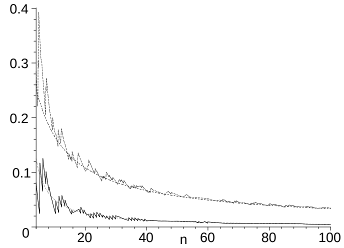
which is slower than $O\left(\frac{1}{n}\right)$ convergence rate of codes for known sources.

Noiseless Coding (Comparison)

- Again, if the source S is known exactly, then the redundancy of its optimal encoding is decreasing as: $R \sim \frac{C}{n}$,

- but if we only know that the source is memoryless, we attain: $R \sim C\ \frac{\log n}{n}$.

- Example:



Block Shannon code vs. universal code under a binary source with $\Pr(0) = 1/8$.

Noiseless Coding (Construction)

- Construction of universal codes:
  - probability estimation + encoding

- Block encoding can be done by either:
  - Arithmetic encoders (Rissanen, Pasco, 1976)
  - Enumerative codes (Lynch, Davisson, 1966, Babkin, Shtarkov, 1968-74, . . .)

- Probability estimators:
  - Laplace estimator:

$$P_L(w) = (m-1)!\frac{r_{\alpha_1}(w)! \dots r_{\alpha_m}(w)!}{(|w| + m - 1)!},$$

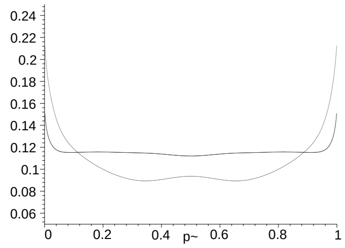
  - Krichevsky-Trofimov (KT) estimator:

$$P_{KT}(w) = \frac{\Gamma(m/2)}{\Gamma(1/2)^m}\frac{\prod_{\alpha \in A}\Gamma(r_\alpha(w) + 1/2)}{\Gamma(|w| + m/2)},$$

where m is cardinality of the alphabet A, $r_\alpha(w)$ - frequencies of symbols in a word w, and $\Gamma(x)$ is the Γ-function.

Noiseless Coding (Estimation)

- Performance of KT vs. Laplace estimators:



Redundancy rates of block universal codes under a binary source with Pr(0) $= p$. Block size $n = 32$.

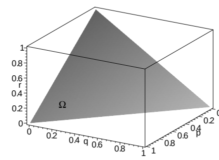
- KT-estimator achieves uniform convergence on Ω.

- Both estimators can be implemented in an incremental fashion; e.g.:

$$P_L\,(w\,\alpha) = P_L\,(w)\,\frac{r_\alpha(w)+1}{|w|+m}\,.$$

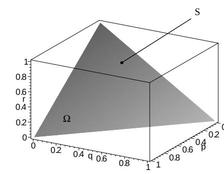
Noiseless Coding (Other cases)

Source S can be anywhere in Ω:



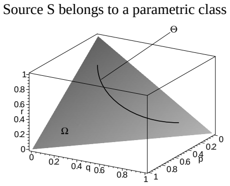
$$R \sim \frac{|A|-1}{2\,n}\log n\,;$$

Known sample of length l produced by S:



$$R \sim \frac{|A|-1}{2\,n}\log\frac{n+\ell}{\ell}\,;$$

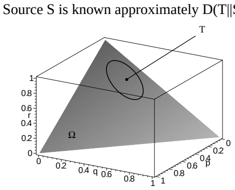
Source S belongs to a parametric class Θ:



$$R \sim \frac{d}{2\,n}\log n\,;$$

when $\Theta \subset \mathbb{R}^d$

Source S is known approximately D(T||S) < δ:



$$R \sim \frac{|A|-1}{2\,n}\log\,(n\,2\,\delta)\,;$$

References

1. T.M. Cover and J.A. Thomas, Elements of Information Theory, Wiley, 1991.

2. R. E. Krichevsky, Universal Data Compression and Retrieval, Kluwer, 1993.

3. S. Graf and H. Luschgy, Foundations of Quantization for Probability Distributions, Springer, Lecture Notes in Mathematics, 1730, Berlin, 2000.

4. R.M. Gray, Source Coding Theory, Kluwer, 1990.

Questions?