Exploiting Statistical Dependence for Compression

- Joint entropy
- Statistical dependence among color components
- Conditional entropy and bit-rate for lossless coding of sources with memory
- Image conditional entropy measurements
- Cross entropy
- Run-length coding
- Facsimile compression standards
Joint entropy

- Consider random vectors (with finite-alphabet components)
  \[ X = (X_0, X_1, \ldots, X_{m-1}) \]

- Entropy
  \[ H(X) = E[-\log_2 f_X(X)] = E[h_X(X)] \]

- Long-hand
  \[ H(X) \equiv H(X_0, X_1, \ldots, X_{m-1}) \]
  \[ = - \sum_{x_0} \sum_{x_1} \ldots \sum_{x_{m-1}} f_X(x_0, x_1, \ldots, x_{m-1}) \log_2 f_X(x_0, x_1, \ldots, x_{m-1}) \]

- Shannon’s Noiseless Source Coding Theorem: Consider a “vector source” generating i.i.d. random vectors \( X \). Joint entropy \( H(X) \) is achievable lower bound for bit-rate for encoding \( X \).
Color components

Standard Image 'Lena'

Red R

Green G

Blue B

Luminance Y

Chrominance Cb

Chrominance Cr
Joint entropy and statistical dependence

- Theorem

\[ H(X_0, X_1, X_2, \ldots, X_{m-1}) \leq H(X_0) + H(X_1) + \ldots + H(X_{m-1}) \]

Equality for statistical independence of \( X_0, X_1, X_2, \ldots, X_{m-1} \)

- Exploiting statistical dependence lowers bit-rate
- Statistically independent components can be coded separately without loss in efficiency
First order statistics of color components

- Histograms (relative number of occurrences) used in lieu of PMFs

<table>
<thead>
<tr>
<th>Component</th>
<th>Entropy (bpp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red (R)</td>
<td>7.25</td>
</tr>
<tr>
<td>Green (G)</td>
<td>7.59</td>
</tr>
<tr>
<td>Blue (B)</td>
<td>6.97</td>
</tr>
<tr>
<td>Luminance (Y)</td>
<td>7.23</td>
</tr>
<tr>
<td>Chrominance (Cb)</td>
<td>5.47</td>
</tr>
<tr>
<td>Chrominance (Cr)</td>
<td>5.42</td>
</tr>
</tbody>
</table>

- Image ‘Lena’, 512 x 512 pixels, 8 bits per color components
Statistical dependence among color components

- Image: ‘Lena’, 512 x 512 pixels, 8 bits per color component

<table>
<thead>
<tr>
<th>Fixed length</th>
<th>3 x 8 = 24 bpp</th>
<th>Fixed length</th>
<th>3 x 8 = 24 bpp</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H(Y,Cb,Cr)$</td>
<td>15.01 bpp</td>
<td>$H(R,G,B)$</td>
<td>16.84 bpp</td>
</tr>
<tr>
<td>$H(Y)+H(Cb)+H(Cr)$</td>
<td>18.12 bpp</td>
<td>$H(R)+H(G)+H(B)$</td>
<td>21.82 bpp</td>
</tr>
<tr>
<td>$\Delta H$</td>
<td>3.11 bpp</td>
<td>$\Delta H$</td>
<td>4.98 bpp</td>
</tr>
</tbody>
</table>

- Statistical dependence among R, G, B is stronger.
- Caveat: if joint sources Y, Cb, Cr or R,G,B are not treated as i.i.d., the possible gain by joint coding may be much smaller.
Exploiting Statistical Dependence Among Neighboring Pixels

- Break image into blocks
- Interpret each block as vector
- Block-by-block entropy coding
- Neglects dependencies across block boundaries

- Sliding window across image
- Interpret image as source with memory
- Perform conditional entropy coding
Markov random processes

- Consider discrete random process \( \{X_n\} \) with memory

- Stationarity

\[
f_{X_{i+i+m}} = f_{X_{0+m}} \quad \text{for all } i, m \in \mathbb{Z}, m \geq 0
\]

- Random process with finite memory ("Markov-p process")

\[
f_{X_m|X_{0:(m-1)}} = f_{X_m|X_{(m-p):(m-1)}}
\]

- Order-\( p \) conditional PMF describes stationary Markov process fully
Conditional entropy

- Consider two finite-alphabet r.v. $X$ and $Y$

$$H(X \mid Y) = E[-\log_2 f_{X \mid Y}(x, y)] = -\sum_y \sum_x f_{X,Y}(x, y) \log_2 f_{X \mid Y}(x, y)$$

$$= -\sum_y f_Y(y) \sum_x f_{X \mid Y}(x, y) \log_2 f_{X \mid Y}(x, y)$$

- Conditional entropy $H(X \mid Y)$ is average additional information, if $Y$ is already known

$$H(X,Y) = E[-\log_2 f_{X,Y}(X,Y)]$$

$$= E[-\log_2 (f_Y(Y) f_{X \mid Y}(X,Y))]$$

$$= E[-\log_2 f_Y(Y)] + E[-\log_2 f_{X \mid Y}(X,Y)]$$

$$= H(Y) + H(X \mid Y)$$
Conditional entropy (cont.)

- Independent random variables $X$ and $Y$

\[
H(X | Y) = E[-\log_2 f_{X|Y}(X,Y)]
= E[-\log_2 f_X(X)] = H(X)
\]

- Let $Y'$ be a random vector containing any subset of elements of another random vector $Y$

\[
H(X | Y) \leq H(X | Y')
\]

Equality, iff $X$ is conditionally independent from elements missing from $Y$.

Hence, extra prior information can only reduce uncertainty.
Bit-rate bound for Markov random process

- Consider discrete, stationary Markov-$p$ random process, characterized by conditional PMF $f_{X_p|X_0.(p-1)}$

- Achievable lower bound for bit-rate: Shannon’s Noiseless Source Coding Theorem:

\[
R \geq H\left( X_p \mid X_0.(p-1) \right)
\]

- Important to exploit statistical dependence

\[
H\left( X_p \mid X_0.(p-1) \right) \leq H\left( X_p \right)
\]
Switched variable length coding

- How to encode \( \{X_n\} \), such that the average code word length approaches \( H(X_p | X_{0:(p-1)}) \)?
- Idea: switch among a set of variable length codes for \( X \) (e.g., Huffman or Golomb), one for each “state” \( X^{(n-p):(n-1)} \)
- Number of states for 8-bit pixels

<table>
<thead>
<tr>
<th>( p )</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( 256 ) states</td>
</tr>
<tr>
<td>2</td>
<td>( 65,536 ) states</td>
</tr>
<tr>
<td>3</td>
<td>( 16,777,216 ) states</td>
</tr>
</tbody>
</table>
Statistical Dependence of Adjacent Pixels

Histogram of two horizontally adjacent pixels
(‘Lena’, 512 x 512 pixels, 8 bpp)
Calculating entropies for an individual image

- Assume that a particular image has its own entropy code – implies some overhead to send code book
- Obtain probabilities from the image histogram
- Joint entropy for two adjacent pixels (bound for joint coding of 2x1 blocks)

\[
H(S_0, S_1) = - \sum_{s_0} \sum_{s_1} f_{s_0,s_1}(s_0, s_1) \log_2 f_{s_0,s_1}(s_0, s_1)
\]

- First-order conditional entropy (bound for conditional coding using previous pixel)

\[
H(S_0 | S_1) = - \sum_{s_0} \sum_{s_1} f_{s_0,s_1}(s_0, s_1) \log_2 f_{s_0|s_1}(s_0 | s_1)
\]
Conditional entropy for an image

Image: ‘Lena’, 512 x 512 pixels, 8 bpp
All values in bpp

| component | $H(S_0)$ | $H(S_0 | S_1)$ | $H(S_0 | S_3)$ | $H(S_0 | S_4)$ |
|-----------|----------|----------------|----------------|----------------|
| Y         | 7.23     | 4.67           | 4.32           | 4.86           |
| Cb        | 5.47     | 3.80           | 3.58           | 3.85           |
| Cr        | 5.42     | 3.69           | 3.55           | 3.82           |
Cross Entropy

- Assume that an entropy code is used which is optimal for probabilities $g_{s_0,s_1}(s_0,s_1)$ rather than the "true" probabilities.
- Obtain "true" image probabilities $f_{s_0,s_1}(s_0,s_1) \neq g_{s_0,s_1}(s_0,s_1)$ from histogram.
- Joint cross entropy for two adjacent pixels (bound for coding of 2x1 blocks):

$$H_{\text{cross}}(f_{s_0,s_1} \parallel g_{s_0,s_1}) = E_{f_{s_0,s_1}} \{-\log_2 g_{s_0,s_1}(s_0,s_1)\} - \sum_{s_0} \sum_{s_1} f_{s_0,s_1}(s_0,s_1) \log_2 g_{s_0,s_1}(s_0,s_1)$$

- First-order conditional cross entropy (bound for coding using previous pixel):

$$H_{\text{cross}}(f_{s_0|s_1} \parallel g_{s_0|s_1}) = -\sum_{s_0} \sum_{s_1} f_{s_0,s_1}(s_0,s_1) \log_2 g_{s_0|s_1}(s_0,s_1)$$
Kullback Leibler divergence

- Kullback-Leibler divergence (increase due to not using “true” probabilities)

\[
D_{KL}(f \parallel g) = H_{cross}(f \parallel g) - H_{cross}(f \parallel f) \geq 0
\]

- Not symmetric

\[
D_{KL}(f \parallel g) \neq D_{KL}(g \parallel f)
\]

- Also known as “relative entropy”
Run-length coding

- For sources that emit “runs” of identical symbols
- Replace a sequence \( \{x_n\} \) by a shorter sequence of symbol pairs \( \{a_k, r_k\} \) such that
  \[
  x_n = a_k \quad \text{for all } n : \sum_{j=1}^{k-1} r_j < n \leq \sum_{j=1}^{k} r_j
  \]
- Entropy coding (e.g., Huffman coding) of new symbol pairs \( \{a_k, r_k\} \)
- For binary images, \( a_k \) can be omitted.
Statistical model for binary images of line-art

- **Markov-1 model** [Capon, 1959]

![Diagram](attachment:image.png)

- **State probabilities**

\[
\Pr \{W\} = 1 - \Pr \{B\} = \frac{p_{BW}}{p_{BW} + p_{WB}}
\]

- **Run-length distributions**

\[
\Pr \{k \text{ successive white pixels}\} = (1 - p_{WB})^{k-1} p_{WB} \quad \text{for } k = 1, 2, 3, \ldots
\]

\[
\Pr \{k \text{ successive black pixels}\} = (1 - p_{BW})^{k-1} p_{BW} \quad \text{for } k = 1, 2, 3, \ldots
\]

- **Golomb code for geometric distribution**
Measured parameters for Capon model

<table>
<thead>
<tr>
<th>Document</th>
<th>Weather Map</th>
<th>Printed Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr {W}</td>
<td>0.887</td>
<td>0.935</td>
</tr>
<tr>
<td>Pr {B}</td>
<td>0.113</td>
<td>0.065</td>
</tr>
<tr>
<td>(P_{WB})</td>
<td>0.027</td>
<td>0.024</td>
</tr>
<tr>
<td>(P_{BW})</td>
<td>0.214</td>
<td>0.347</td>
</tr>
<tr>
<td>(H\left( X_i \mid X_{i-1} \right))</td>
<td>0.241 bpp</td>
<td>0.215 bpp</td>
</tr>
</tbody>
</table>

[Kunt, 1974]
Facsimile compression standards

- Standards by the ITU-T (formerly CCITT)
  - T.4 (Group 3)
    - used by all fax machines over PSTN (public switched telephone network)
    - 1-d modified Huffman code (MH) or 2-d MMR code optional
  - T.6 (Group 4): fax over digital networks (e.g., ISDN)
    - Always 2-d MMR code
    - Less error-resilient than Group 3

- Picture formats
  - Horizontal resolution: 1728 pixels/line (1664 active) $\Rightarrow$ 8.05 pixel/mm
  - Vertical resolution
    - Standard mode: 3.85 lines/mm (978 lines/page)
    - Fine mode: 7.7 lines/mm (1956 lines/page)
    - Very-fine mode: 15.4 lines/mm (3912 lines/pages)
Group 3 fax: modified Huffman code

- Lengths of runs of white pixels and black pixels encoded within a scan line
- Each run represented as
  
  white runs: \( r_w = 64 \times r_{w/\text{make-up}} + r_{w/\text{term}} \)
  black runs: \( r_b = 64 \times r_{b/\text{make-up}} + r_{b/\text{term}} \)

- “Make-up” and “termination” run-lengths encoded independently
- Two separate Huffman code tables for white and black runs (“make-up” and “termination” codes) based on the statistics of 8 representative documents
- Shortest code words (2 bits) for black runs of length 2 and 3
- Shortest code words (4 bits) for white runs of lengths 2 . . . 7
- Same code tables also used as part of Group 4 fax standard (MMR)
- Special EOL codeword for each line, 6x EOL as end of page
2-d fax coding

- MMR – Modified modified READ (relative address designate)
- Optional enhancement for Group 3, mandatory for Group 4
- For error resilience, Group 3 standard requires that at most \( K-1 \) lines are encoded 2-d, with \( K=2 \) for standard resolution and \( K=4 \) for fine resolution.
**Modified modified READ algorithm**

- Encode black and white run lengths relative to reference line above
- Mode #1: Pass Mode

If $b_2 < a_1$: issue pass code word 0001

Update pointers: $a_0 \leftarrow b_2$, find new $b_1, b_2$
Modifying modified READ algorithm (cont.)

- Mode #2: Vertical Mode

If $|a_1 - b_1| \leq 3$: encode difference $a_1 - b_1$ using table

<table>
<thead>
<tr>
<th>$a_1 - b_1$</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0000011</td>
<td>000011</td>
<td>011</td>
<td>1</td>
<td>010</td>
<td>000010</td>
<td>0000010</td>
</tr>
</tbody>
</table>

Update pointers: $a_0 \leftarrow a_1$, $b_1 \leftarrow b_2$ etc.
Mode #3: Horizontal Mode

If neither Pass Mode nor Vertical Mode:
encode next two runs 1-d, using Group 3 modified Huffman code

Update pointers: $a_0 \leftarrow a_2$, find new $a_1, a_2, b_1, b_2$
Compression efficiency of fax standards

<table>
<thead>
<tr>
<th>CCITT Document Number</th>
<th>Standard resolution (2,052,864 pixels)</th>
<th></th>
<th>Fine resolution (4,105,728 pixels)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-d MHC</td>
<td>2-d READ</td>
<td>1-d MHC</td>
<td>2-d READ</td>
</tr>
<tr>
<td>1</td>
<td>133,095</td>
<td>93,196</td>
<td>266,283</td>
<td>141,826</td>
</tr>
<tr>
<td>2</td>
<td>123,930</td>
<td>53,366</td>
<td>247,443</td>
<td>80,550</td>
</tr>
<tr>
<td>3</td>
<td>244,028</td>
<td>138,411</td>
<td>487,485</td>
<td>229,375</td>
</tr>
<tr>
<td>4</td>
<td>436,450</td>
<td>366,055</td>
<td>871,983</td>
<td>553,942</td>
</tr>
<tr>
<td>5</td>
<td>253,509</td>
<td>162,186</td>
<td>506,283</td>
<td>257,548</td>
</tr>
<tr>
<td>6</td>
<td>191,347</td>
<td>78,577</td>
<td>381,905</td>
<td>132,509</td>
</tr>
<tr>
<td>7</td>
<td>428,028</td>
<td>357,130</td>
<td>855,841</td>
<td>539,152</td>
</tr>
<tr>
<td>8</td>
<td>238,221</td>
<td>89,654</td>
<td>476,624</td>
<td>137,560</td>
</tr>
<tr>
<td>Average</td>
<td>256,076</td>
<td>167,322</td>
<td>511,731</td>
<td>259,058</td>
</tr>
</tbody>
</table>

[Yasuda, 1980]
Reading

- Wiegand + Schwarz, Chapter 3
- Taubman + Marcellin, Chapter 2.3 – 2.5