4 Statistical Dependency

4.1 Joint Entropy and Statistical Dependency

Extension of entropy definition to random vectors $\boldsymbol{X} = (X_0, X_1, \ldots, X_{N-1})$

Joint entropy of N random variables $X_i, i = 0, \ldots, N-1$ is defined as entropy of random vector $\boldsymbol{X}$ and a function of the joint PMF

$$H(\boldsymbol{X}) = \mathrm{E}\{-\log_2 p_{\boldsymbol{X}}(\boldsymbol{X})\} = \mathrm{E}\{h_{\boldsymbol{X}}(\boldsymbol{X})\}$$

Alphabet of random vector is discrete, elements can be enumerated in $A_{\boldsymbol{X}}$

$$\begin{aligned} H(\boldsymbol{X}) &= H(X_0, X_1, \ldots, X_{N-1}) \\ &= -\sum_{\boldsymbol{x} \in A_{\boldsymbol{X}}} p_{\boldsymbol{X}}(x_0, x_1, \ldots, x_{N-1}) \log_2 p_{\boldsymbol{X}}(x_0, x_1, \ldots, x_{N-1}) \end{aligned}$$

Shannon's noiseless source coding theorem

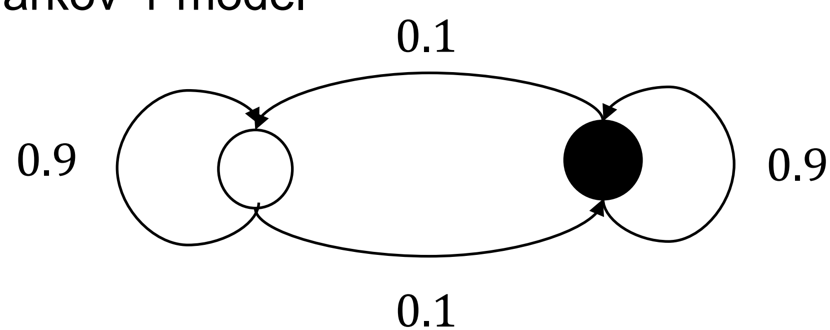
⇨ Joint entropy $H(\boldsymbol{X})$ is achievable lower bound for bit-rate R when encoding $\boldsymbol{X}$

Statistical dependence lowers entropy and hence bit-rate since generally

$$H(X_0, X_1, \ldots, X_{N-1}) \leq H(X_0) + H(X_1) + \ldots + H(X_{N-1})$$

Example for Joint Entropy: Markov Process

Consider two neighboring samples $\boldsymbol{X} = (X_0, X_1)$ of a binary image following a Markov-1 model



$A_X = \{\square, \blacksquare\}$

$P(\blacksquare) = 0{,}5$ $\xrightarrow{\text{Code}}$ "0"
$P(\square) = 0{,}5$ $\xrightarrow{\text{Code}}$ "1"

$R = 1$ bit / pixel

Entropy without considering correlation:

$$H(X) = -\sum_{x \in A_X} p_X(x) \log_2 p_X(x)$$

$= 1$ bit/pixel

Joint entropy for vector of two samples:

$$H(\boldsymbol{X}) = -\sum_{x_0 \in A_X} \sum_{x_1 \in A_X} p_{\boldsymbol{X}}(x_0, x_1) \log_2 p_{\boldsymbol{X}}(x_0, x_1)$$

$= 1.47$ bit / two pixel
$= 0.74$ bit / pixel

	Code	
$P(\square, \square) = 0{,}45$	→	"1"
$P(\blacksquare, \blacksquare) = 0{,}45$	→	"01"
$P(\blacksquare, \square) = 0{,}05$	→	"001"
$P(\square, \blacksquare) = 0{,}05$	→	"000"

$R = 1{,}65$ bit / two pixel
$= 0{,}83$ bit / pixel

Conditional Entropy

Conditional entropy of two finite-alphabet random variables X and Y

$$H(X|Y) = \mathrm{E}\{-\log_2 p_{X|Y}(X,Y)\} = -\sum_{y \in A_Y} \sum_{x \in A_X} p_{X,Y}(x,y) \log_2 p_{X|Y}(x,y)$$

$$= -\sum_{y \in A_Y} p_Y(y) \sum_{x \in A_X} p_{X|Y}(x,y) \log_2 p_{X|Y}(x,y)$$

Interpretation: conditional entropy is the average additional information we get from outcome of X if Y is already known

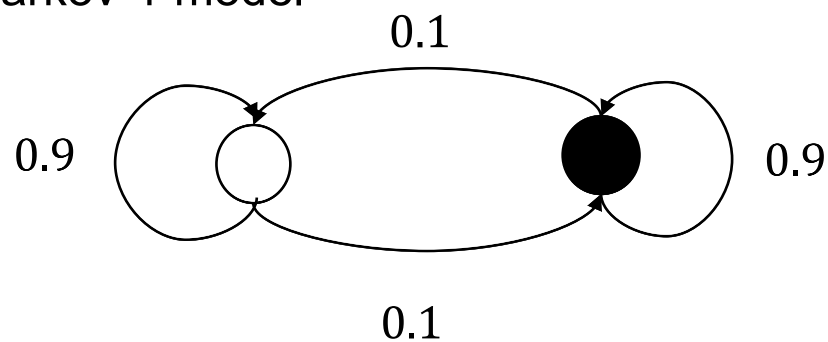
$$\begin{aligned} H(X,Y) &= \mathrm{E}\{-\log_2 p_{X,Y}(X,Y)\} \\ &= \mathrm{E}\{-\log_2 (p_Y(Y) p_{X|Y}(X,Y))\} \\ &= \mathrm{E}\{-\log_2 p_Y(Y)\} + \mathrm{E}\{-\log_2 p_{X|Y}(X,Y)\} \\ &= H(Y) + H(X|Y) \end{aligned}$$

Independent random variables X and Y

$$\begin{aligned} H(X|Y) &= \mathrm{E}\{-\log_2 p_{X|Y}(X,Y)\} \\ &= \mathrm{E}\{-\log_2 p_X(X)\} = H(X) \end{aligned}$$

Example for Conditional Entropy: Markov Process

Consider two neighboring samples $\boldsymbol{X} = (X_0, X_1)$ of a binary image following a Markov-1 model



$A_X = \{\square, \blacksquare\}$

$P(\blacksquare) = 0{,}5$	Code $\rightarrow$	"0"
$P(\square) = 0{,}5$		"1"

R = 1 bit / pixel

Conditional entropy of samples X_0 and X_1:

$$H(X_0|X_1) = -\sum_{\forall x_0}\sum_{\forall x_1} p_{X_0,X_1}(x_0,x_1)\log_2 p_{X_0|X_1}(x_0,x_1)$$
$= 0.47$ bit / pixel

$P(\square, \square) = 0{,}45$		"1"
$P(\blacksquare, \blacksquare) = 0{,}45$	Code $\rightarrow$	"01"
$P(\blacksquare, \square) = 0{,}05$		"001"
$P(\square, \blacksquare) = 0{,}05$		"000"

Joint entropy for vector of two samples:

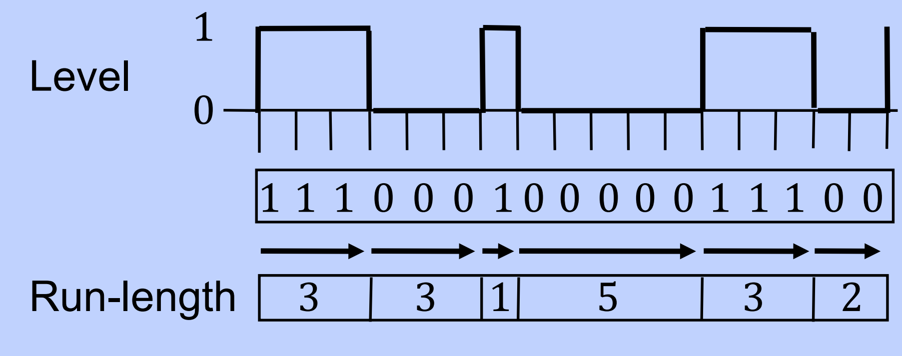
$H(X_0, X_1) = H(X_1) + H(X_0|X_1)$
$= 1$ bit / pixel $+ \, 0.47$ bit / pixel
$= 1.47$ bit / two pixel

$R = 1.65$ bit / two pixel
$= 0.83$ bit / pixel

4.2 Run-Length Coding

Sequences of identical values

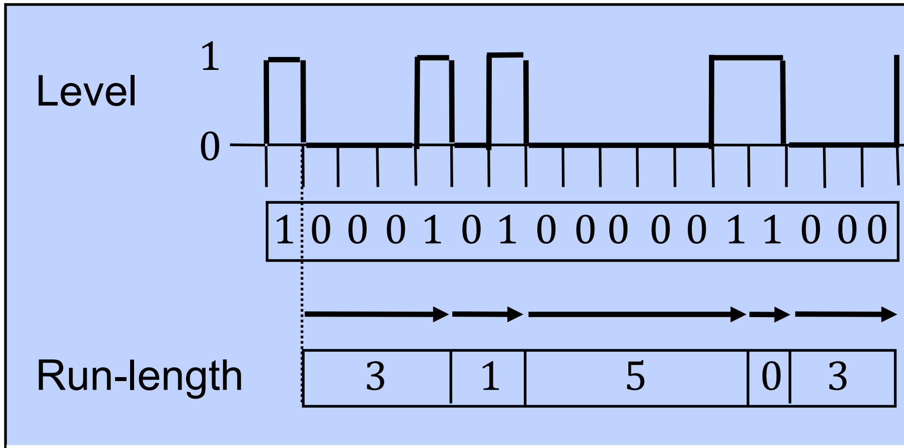
- Represented by their value a and number of repetitions r
- Symbols a and r are entropy coded, e.g. using Huffman coding
- For binary values a can be omitted



- Levels have almost equal probability

Applications are especially in

- bi-level images like text, graphics, maps,…
- differential coding (DPCM)



- Coding only the more frequent level
- The run length of level “0” includes the jump and one run of “1”

Disadvantage

- Sensitive to transmission errors

Run-Level Coding

Sparse signal

- Contains mostly zero and only few non-zero (non-binary) amplitudes
- Example: grey-level text documents

Signal value	8	3	0	4	0	0	1	0	0	0	2	1	0	0	0	0

Zero run-length	0	0	1	2	3	0	4
level	8	3	4	1	2	1	0

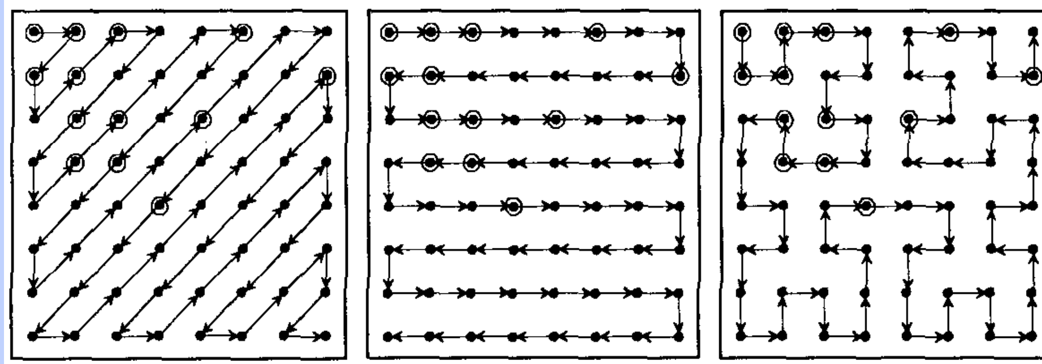
Run-level coding

- Send run-length r of preceding zeros and level a of non-zero symbol

Extension to images

- Coded using predefined two-dimensional scan orders

Run-level coding of spectral coefficients:



0,0,0,1,0,2,2,0,2,2,6,6,9,21 | 0,0,0,2,2,5,0,1,0,1,8,0,4,28 | 0,0,0,0,0,3,2,0,0,17,22,4,2,1

Zigzag scan | Line by line scan | Peano-Hilbert scan

4.3 Fax Compression Standards

Standards for transmission of facsimile documents released by ITU-T:

- T.4 (Group 3): fax machines connected to PSTN
  - 1D modified Huffman code
  - 2D MMR code optional
- T.6 (Group 4): fax over digital networks, e.g. ISDN
  - Always 2D MMR code
  - Less error resilient

Scanning resolution

- Horizontal sampling at 1728 pixels/line corresponding to approx. 8 pixel/mm
- Vertical sampling at three different modes
  - Standard mode: 3.85 lines/mm
  - Fine mode: 7.7 lines/mm
  - Very fine mode: 15.4 lines/mm

Comparison Group 3 and Group 4 Fax

Group 3 fax: Modified Huffman code (MHC)

- Run-lengths of white and black pixels coded within scan line
- Two separate Huffman code tables for white and black runs
- Code tables obtained from statistics of representative documents
- Special EOL codeword for each line, 6x EOL signals end of page

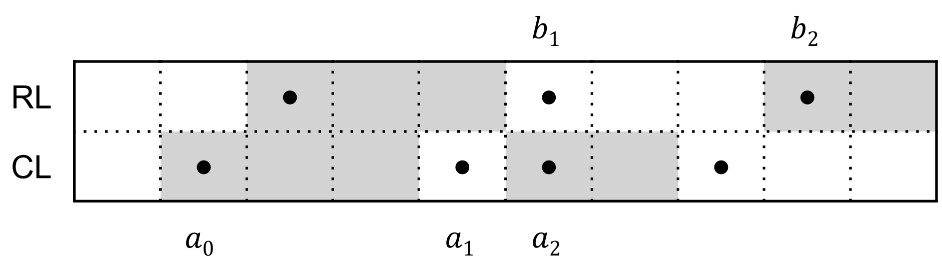
Group 4 fax: Modified modified READ (MMR)

- READ = relative address designate
- Optional for Group 3, mandatory for Group 4
- Black and white run-lengths are coded relative to reference line above
- Number K of predictively coded scan lines is signaled
- For error resilience K can be limited ($K = 4$ for Group 3)

MMR Algorithm

Consider two neighboring scan lines:

First line → reference line (RL), second line → coding line (CL)



"•" = CP

Definitions a_0 current pixel location

a_1, a_2 the CPs on CL

b_1, b_2 the CPs on RL

such that a_1 is the first CP to the right of a_0

a_2 is the second CP to the right of a_0

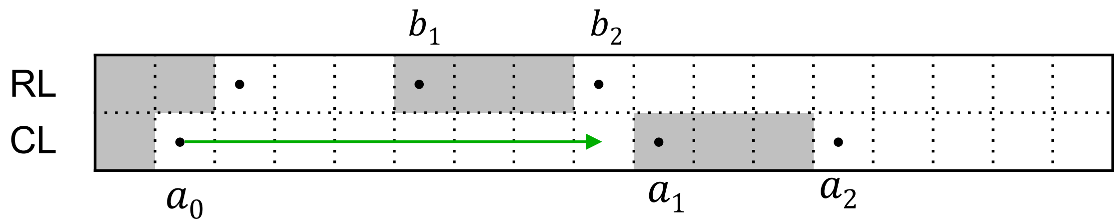
b_1 is the first CP to the right of a_0 that has the same color as a_1

b_2 is the first CP to the right of b_1 that has the same color as a_2

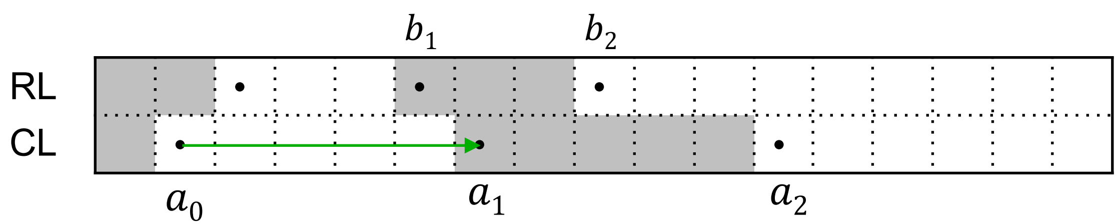
MMR Algorithm

(a) Choose one of three modes and (b) move to end of green arrow

Pass mode (P): if b_2 lies strictly to the left of a_1 ⇨ send [pass code]

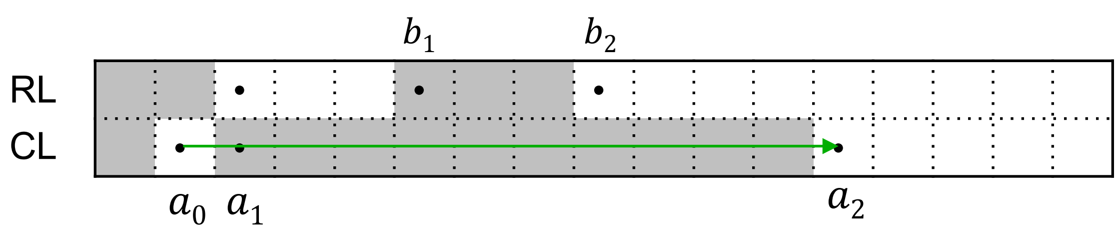


Vertical mode (V): if not pass mode and a_1, b_1 are no more than T pels apart from each other ⇨ send $[V_x]$ $(x = 0, \pm 1, \ldots, \pm T)$



here: $V_x = V_{+1}$
typically: $T = 2$

Horizontal mode (H): if none of the above, send [horizontal code] and $[a_1 - a_0]$ as well as $[a_2 - a_1]$



Pseudo Code for MMR Algorithm

Modified modified READ algorithm

Initialization: $a_0 = 0$

Step 1: Detect a_1, b_1, and b_2
If (**P** mode, i.e. $b_2 < a_1$)
- send [pass code] symbol
- move a_0 to the pixel below b_2
- go to Step 1

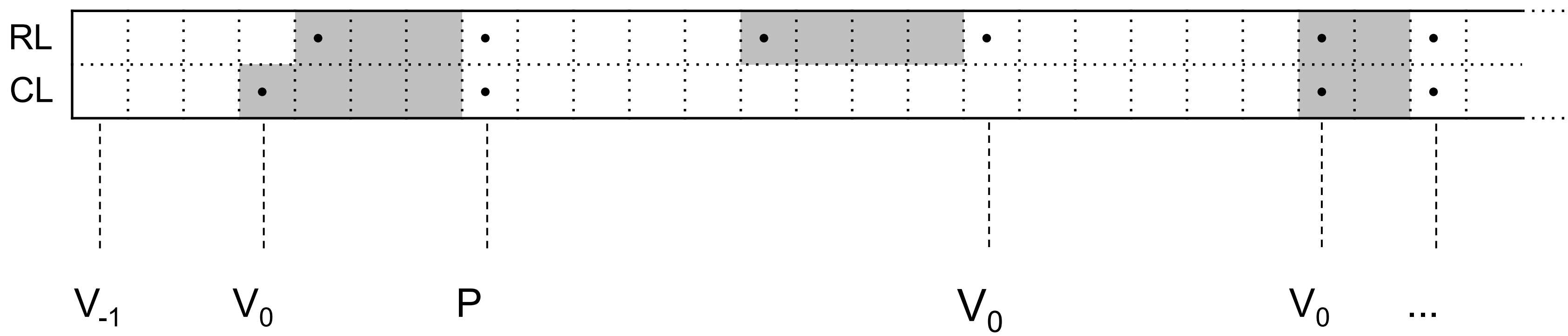
Step 2: If (**V** mode, i.e. $|a_1 - b_1| \leq T$)
- send $[V_x]$ $(x = 0,+/-1,\ldots,+/-T)$
- move a_0 to a_1
- go to Step 1

Step 3: (**H** mode)
- detect a_2
- send [horizontal code] + $[a_1 - a_0]$ + $[a_2 - a_1]$
- move a_0 to a_2
- go to Step 1

end

Example for MMR Coding

Coding mode for changing line



Compression efficiency of T.4 and T.6

Standard resolution		Fine resolution	
1D MHC	2D READ	1D MHC	2D MMR
356,076	167,322	511,731	259,058

- Average number of bits for test set of 8 documents

Statistical Dependency - Summary

- Joint entropy is lower bound for vector coding
- Run-length and run-level coding exploit dependency
- Fax compression standards based on relative address coding