A STUDY ON DATA-DRIVEN PROBYBILITY ESTIMATOR DESIGN FOR VIDEO CODING

Heiner Kirchhoffer, Christian Rudat,
Michael Schäfer, Jonathan Pfaff, Heiko Schwarz,
Detlev Marpe, and Thomas Wiegand

Overview

- The Enhanced Compression Model (ECM) version 3 is employed to study data-driven approaches for optimizing the probability estimator of its entropy coding stage

- ECM is based on the Versatile Video Coding (VVC) standard but includes additional coding tools that increase the compression efficiency

- In ECM, the probability estimator of each context model uses a weighted sum of two hypotheses with different associated adaptation rates

- This paper studies four alternative approaches in order to increase the compression efficiency

Review of the Probability Estimator of ECM

- For each context model, two probability estimates $p_0(t)$ and $p_1(t)$ are maintained
- The probability estimate for arithmetic coding is the weighted sum

$$p(t) = \big(p_0(t) + p_1(t)\big)/2$$

- After encoding (or decoding) a binary symbol (bin) $x(t)$, both estimates are updated according to the recursive equation

$$p_i(t+1) = \alpha_i \cdot p_i(t) + (1 - \alpha_i) \cdot x(t)$$

- A predefined value is used as initial probability estimate $p_i(1)$
- The adaptation parameter $\alpha_i = 1 - 2^{-r_i}$ controls the adaptation speed
- The inertia parameters r_i may be integers ranging from 2 to 9
- For each context model, predefined values are used for the inertia parameters

Data-driven hypotheses weighting (DHW)

- The number of hypotheses of the ECM estimator is increased from 2 to 14
- Fixed inertia parameters r_i in the range $1 \leq r_i \leq 14$ are used
- The probability estimate for arithmetic coding is the weighted sum

$$p_{DHW}(t) = \sum_{i=1}^{14} w_i \cdot p_i(t)$$

- The initial probability estimate $p(1)$ and the weights w_i are derived in a data-driven way

Data-driven weighting of latest bins (DWLB)

- The sequence of bins $x(t)$ is extended to an infinite sequence

$$\tilde{x}(t) = \begin{cases} x(t) & \text{if } t > 0, \\ p(1) & \text{otherwise.} \end{cases}$$

- The probability estimate of DWLB for arithmetic coding is given as:

$$p_{DWLB}(\mathrm{t}+1) = \sum_{i=0}^{D} \tilde{x}(t-i) \cdot \phi_i$$

- Parameter D corresponds to the number of previously observed bins
- The initial probability estimate $p(1)$ and the weights ϕ_i are derived in a data-driven way
- The probability estimator of ECM can be expressed as a DWLB estimator if $D \to \infty$
- Simulations are conducted using $D = 2048$ which turns out to be sufficiently large

Data-driven training of adaptation parameters (DTA)

- The probability estimate for arithmetic coding is the weighted sum of B hypotheses:

$$p_{DTA}(t) = \sum_{i=1}^{B} w_i \cdot p_i(t) = \sum_{i=1}^{B} w_i \cdot (\alpha_i \cdot p_i(t) + (1 - \alpha_i) \cdot x(t))$$

- In comparison to DHW, adaptation parameters α_i can be arbitrary real numbers (i.e., they are not restricted to stem from a fixed set of 14 values)
- The initial probability estimate $p(1)$, the adaptation parameters α_i, and associated weights w_i are trained in a data-driven way
- Simulations are conducted using two ($B = 2$) or three ($B = 3$) hypotheses

Heiner Kirchhoffer

Neural network-based weighting of latest bins (NNWLB)

- The DHW estimator $p_{DHW}(t)$ is combined with an estimator $p_{NN}(t)$ based on a 2-layer fully-connected neural network
- The probability estimate for arithmetic coding is given as weighted sum

$$p_{NNWLB}(t) = \beta \cdot p_{DHW}(t) + (1 - \beta) \cdot p_{NN}(t)$$

- The neural network uses the latest 200 bins as input
- The parameters of the DHW estimator, the parameter β, and the parameters of the neural network are derived in a data-driven way
- Two configurations are tested:
  - NNWLB_10: Uses 10 activations per hidden layer
  - NNWLB_100: Uses 100 activations per hidden layer

Heiner Kirchhoffer

Fraunhofer HHI

Experimental results

- The data set for data-driven optimization/neural network training is disjoint from the evaluation dataset
- Coding gains (Bjøntegaard delta rates based on $PSNR = 6 \cdot Y_{PSNR} + U_{PSNR} + V_{PSNR}$) relative to ECM:

Architecture (number of trainable parameters)	Transcoding			Estimation and coding		
	AI	RA	LB	AI	RA	LB
DHW (32)	-0.07%	-0.13%	-0.16%	-0.10%	-0.19%	-0.32%
DWLB (2053)	-0.08%	-0.14%	-0.17%	-0.11%	-0.22%	-0.35%
DTA_2 (7)	-0.06%	-0.11%	-0.14%	-0.06%	-0.16%	-0.22%
DTA_3 (9)	-0.07%	-0.12%	-0.15%	-0.08%	-0.18%	-0.28%
NNWLB_10 (2184)	-0.10%	-0.17%	-0.19%	-0.13%		
NNWLB_100 (30534)	-0.11%	-0.18%	-0.20%	-0.14%		

- Results for NNWLB 10/100 are not available for RA and LB due to high encoding runtimes.