An Affine-Linear Intra Prediction With Complexity Constraints

ICIP 2019,
Taipei

Michael Schäfer, Björn Stallenberger, Jonathan Pfaff,
Philipp Helle, Heiko Schwarz, Detlev Marpe, Thomas Wiegand

Fraunhofer HHI, Video Coding & Analytics Department

Fraunhofer
Heinrich Hertz Institute

Motivation of research

Observation. Given modern computational capabilities, it is possible to obtain new intra prediction modes as outcome of a training experiment; *Pfaff et al. 2018*.

A question to ask. Given the computational burden of the conventional intra prediction modes as upper bound, are these predictors optimal among all?

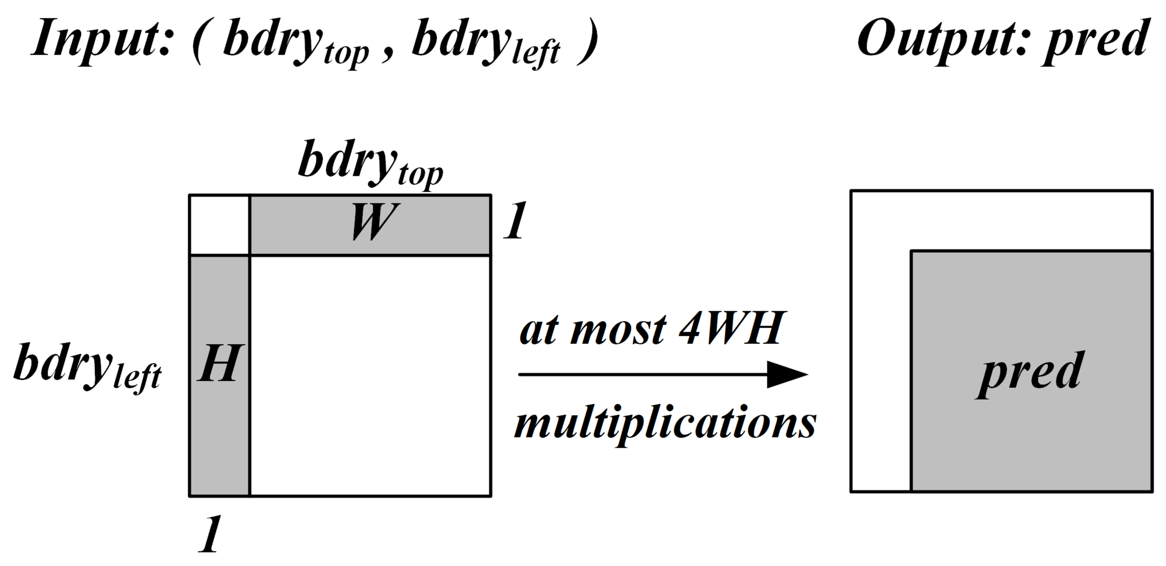

Fig. 1 Generation of an intra prediction signal.

Our goal. Train affine-linear predictors which use one line of reconstructed boundary samples as input and require at most four multiplications per sample to predict.

Description of the trained predictors

Overview. We propose to train $N = 35$ intra prediction modes on a large set of high resolution images as training data. The prediction consists of the following three steps:

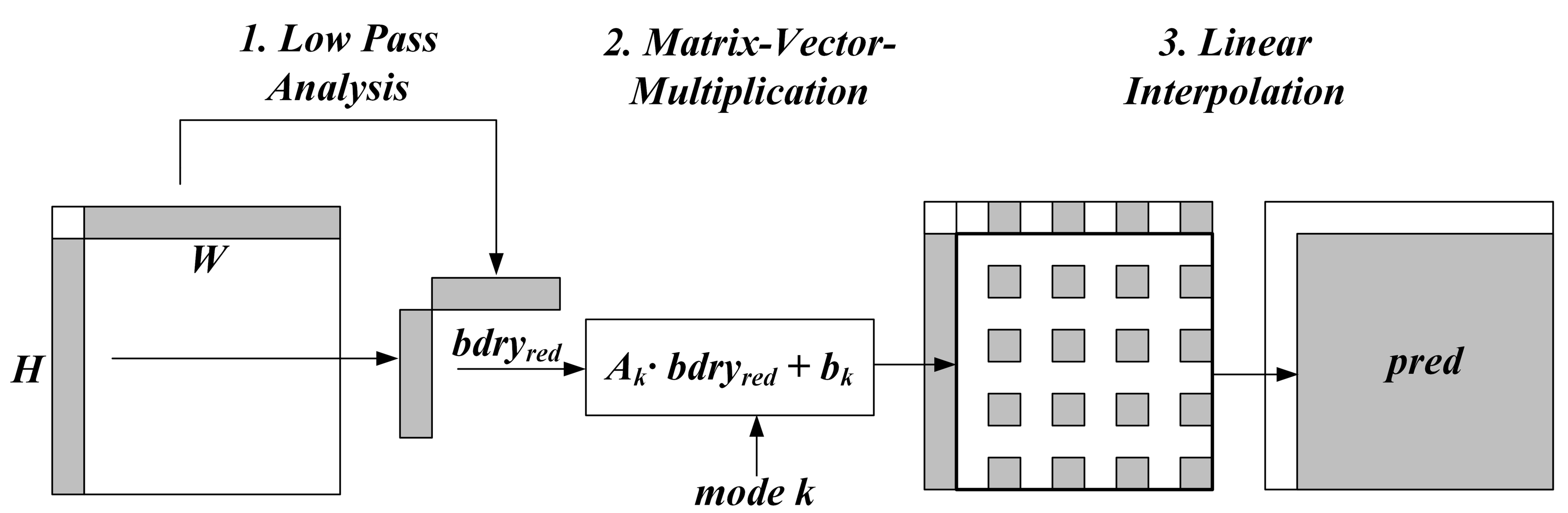

Fig. 2 Flow chart of the trained intra prediction.

Low pass analysis. The low-pass-filtered boundary bdry_{red} consists of two samples along each axis in the case of (4,4)-blocks and four samples else. Given the width $W = 4 \cdot 2^n, n \geq 0$, one computes

$$\text{bdry}_{red}^{top}[i] = \begin{cases} \frac{1}{2^n}\sum_{j=0}^{2^n-1} \text{bdry}^{top}[2^n i + j], & \text{if } n > 0, \\ \frac{1}{2}\left(\text{bdry}^{top}[2i] + \text{bdry}^{top}[2i+1]\right), & \text{else.} \end{cases} \tag{1}$$

The low subband of the left boundary bdry_{red}^{left} is obtained analogously.

Matrix-Vector-Multiplication. Given the prediction mode k, the low subband of the prediction signal pred_{red} is computed as

$$\text{pred}_{red} = A_k \cdot \text{bdry}_{red} + b_k. \tag{2}$$

The dimension (W_{red}, H_{red}) of the low-pass signal pred_{red} equals $(4,4)$ if $\max(W,H) \leq 8$. In any other case holds

$$(W_{red}, H_{red}) = (\min(W,8), \min(H,8)). \tag{3}$$

The expression (2) requires not more than $4WH$ multiplications.

Linear interpolation. Assume $W \geq H$. The vertically interpolated signal pred_{red}^{up} is given for $y = 0, ..., H_{red} - 1, x = 0, ..., W_{red} - 1$ as

$$\begin{aligned} \text{pred}_{red}^{up}[x][2y+1] &= \text{pred}_{red}[x][y], \\ \text{pred}_{red}^{up}[x][2y] &= \frac{1}{2}(\text{pred}_{red}[x][y-1] + \text{pred}_{red}[x][y]). \end{aligned} \tag{4}$$

The step (4) is carried out n times until $2^n \cdot H_{red} = H$.

Training design.

- We train a single joint layer and N different linear output layers.
- Simplify after training by multiplying the weights from both layers.
- The predictors for shapes $(4,4)$, $(8,8)$ and $(16,16)$ are trained jointly in one run using a recursive quad-tree.

Memory assessment.

Shape (W,H)	Input dim	Output dim	Nr. of (A_k, b_k)	Bits per entry	Memory in kB
$W = H = 4$	4	16	18	10	1.8
$\min(W,H) \leq 8$	8	16	18	10	3.24
else	8	64	18	10	12.96

Loss function

Given the DCT-transformed residuals of prediction mode k by $c_k = T(\text{org} - \text{pred}_k)$, we approximate the bit-rate of the residuals as

$$L(\text{org}, k) = \sum_{i=1}^{WH}(|(c_k)_i| + \alpha g(\beta|(c_k)_i| - \gamma)) = \sum_{i=1}^{WH} l((c_k)_i), \tag{5}$$

where α, β, γ are hand-tuned parameters. The total loss of a block is modelled as the sum of L and the signalling cost for the mode index k.

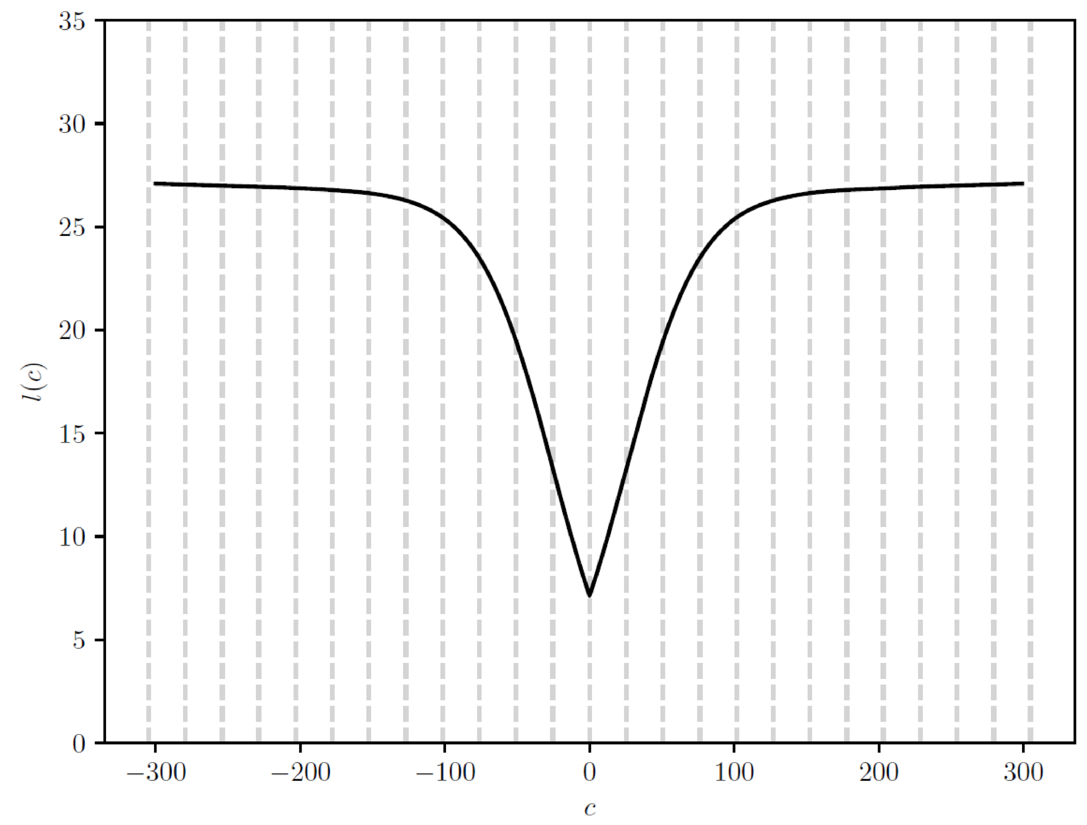

Fig. 3 Profile of the function l in (5).

It penalizes nonzero coefficients. For large coefficients, the curve flattens.

Experimental results and conclusion

All Intra	Y	Enc Time	Dec Time
Class A1	-1.38%	152%	104%
Class A2	-0.75%	151%	103%
Class B	-0.79%	155%	101%
Class C	-0.86%	154%	100%
Class E	-1.11%	151%	98%
Overall	-0.95%	153%	101%

Table 1 BD-rate savings over VTM-3.0 (CTC; *JVET-L1010*).

- Novel data-driven training of affine-linear intra prediction modes
- The predictors employ subband decomposition
- Good trade-off between memory, complexity and bit-rate savings