Improved Prediction via Thresholding Transform Coefficients

ICIP 2018, Athens

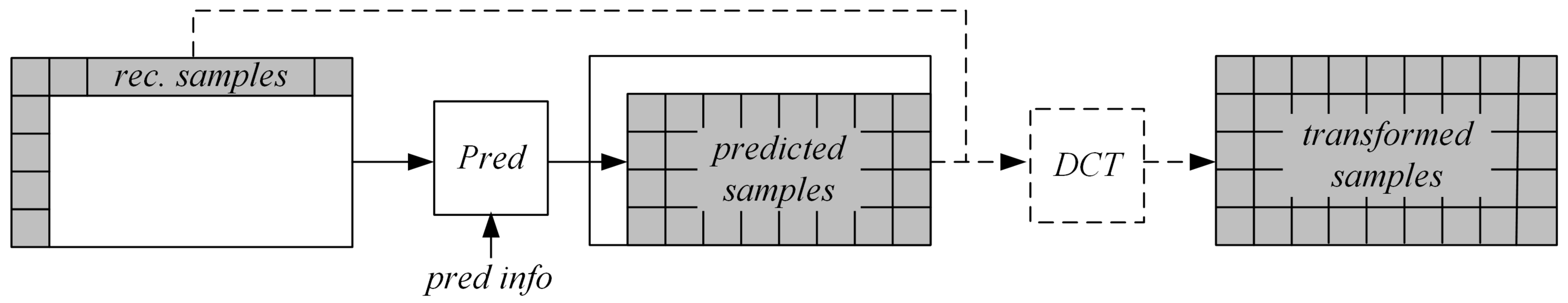
Michael Schäfer, Jonathan Pfaff, Jennifer Rasch, Tobias Hinz, Heiko Schwarz, Tung Nguyen, Gerhard Tech, Detlev Marpe, Thomas Wiegand

Fraunhofer HHI, Video Coding & Analytics Department

Fraunhofer Heinrich Hertz Institute

Motivation of research

Observation. A block-based video codec uses the reconstructed samples for generating a prediction. When transformed via DCT, the signal tends to be concentrated in a few low-frequency components.

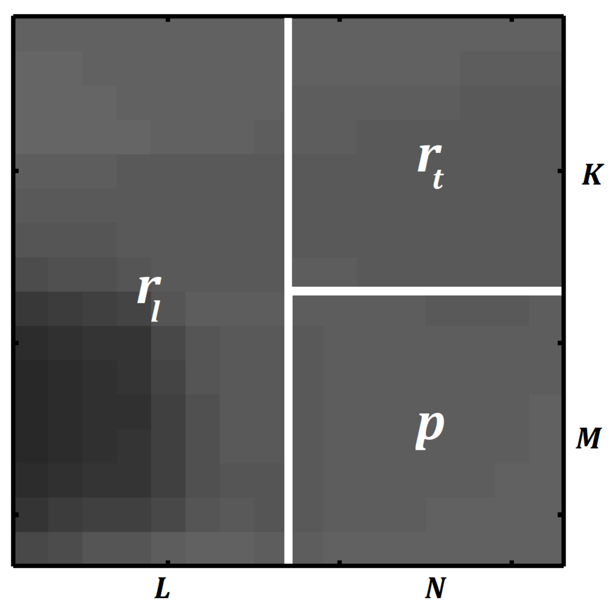


Compressed sensing theory. Assume a high-dimensional signal has a sparse representation in a suitable basis. Then it can be recovered from incomplete or distorted data by random linear measurements.

A question to answer. Is it possible to restore the original samples from the prediction by sparse regularization in the transform domain?

Description of the thresholding method

Initial value. Let $p \in \mathbb{R}^{M\times N}$ be the initial luma prediction. Let r_t and r_l be the reconstructed samples in the $K \leq M$ rows above and the $L \leq N$ columns to the left. Define the extended prediction $y \in \mathbb{R}^{(M+K)\times(N+L)}$ by arranging as below.



Forward transform. Map into the frequency domain via the orthogonal discrete cosine transform $W : \mathbb{R}^{(M+K)\times(N+L)} \to \mathbb{R}^{(M+K)\times(N+L)}$

$$Y = Wy. \tag{1}$$

Thresholding. Choose a threshold value $\tau > 0$. Set transform coefficients $Y_{11}, ..., Y_{(M+K),(N+L)}$ whose absolute value is less than the threshold τ to zero. Hence, the thresholded coefficients are given by

$$\hat{Y}_{ij} = \begin{cases} Y_{ij} & |Y_{ij}| \geq \tau \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

.

Inverse transform. Use the inverse DCT and compute the new extended prediction signal $\hat{y} \in \mathbb{R}^{(M+K)\times(N+L)}$ as

$$\hat{y} = W^T\hat{Y}. \tag{3}$$

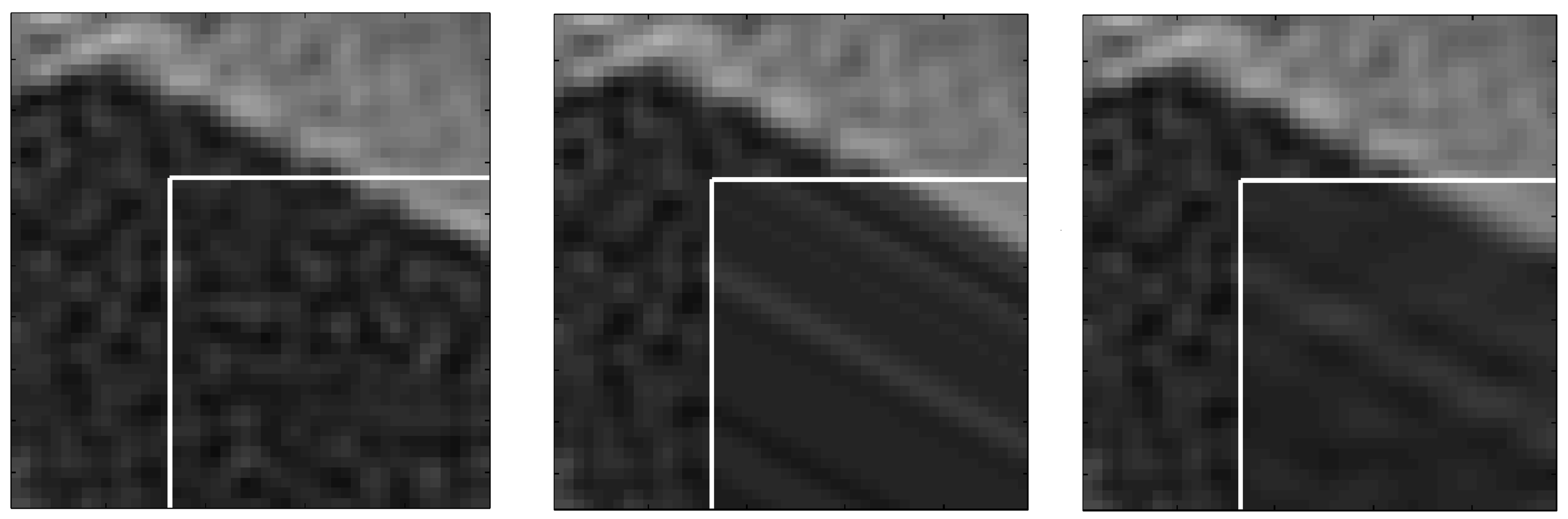
The existing prediction p is then replaced by the modified prediction $\hat{p} \in \mathbb{R}^{M\times N}$ whose entries for $i = 1, ..., M$ and $j = 1, ..., N$ are

$$\hat{p}_{ij} = \hat{y}_{(i+K),(j+L)}. \tag{4}$$

Sparse regularization.

- Goal: Force the prediction to have sparse transform coefficients in accordance with the reconstructed neighborhood.
- Impact of the neighborhood: The coding gain vastly decreases when the reconstructed samples are not used (experimentally validated).
- Choice of the threshold: Different values of τ need to be tested. Take into account the range of the transform coefficients Y and the QP.

Example



Left: Original luma samples. Middle: Angular prediction with extended boundary. Right: Result of the thresholding method.

Thresholding modes

Thresholding mode	Type	Nr. of thresholds τ	Nr. of extension sizes (K, L)
0	no thresholding	-	-
1	use thresholding	8	4

The thresholds are set as table values, depending on the QP.
The extension sizes are set as table values as well.

Experimental results in HEVC

- High-Tier, QP $\in \{22, 27, 32, 37\}$:

All Intra

Sequence name	Resolution	Y
BasketballDrive	HD	-3.92%
Tango	4K UHD	-4.90%
Rollercoaster	4K UHD	-5.17%
Crosswalk	4K UHD	-6.61%
FoodMarket	4K UHD	-8.04%
Overall		**-5.73%**

Random Access

Sequence name	Resolution	Y
Nebuta	4K UHD	-5.37%
Tango	4K UHD	-4.10%
Rollercoaster	4K UHD	-3.73%
Crosswalk	4K UHD	-4.43%
FoodMarket	4K UHD	-4.47%
Overall		**-4.42%**

- Main-Tier, QP $\in \{27, 32, 37, 42\}$:

All Intra

Sequence name	Resolution	Y
BasketballDrive	HD	-4.52%
Tango	4K UHD	-5.02%
Rollercoaster	4K UHD	-5.55%
Crosswalk	4K UHD	-6.79%
FoodMarket	4K UHD	-7.93%
Overall		**-5.96%**

Random Access

Sequence name	Resolution	Y
Nebuta	4K UHD	-5.81%
Tango	4K UHD	-3.02%
Rollercoaster	4K UHD	-2.98%
Crosswalk	4K UHD	-3.39%
FoodMarket	4K UHD	-3.25%
Overall		**-3.69%**

RD plot of *Food Market*, All Intra