DYNAMIC TEXTURE PREDICTION FOR H.264/AVC INTER CODING

Stojanovic A., Ohm J.-R. - RWTH Aachen University
{stojanovic, ohm}@ient.rwth-aachen.de

ICVSS 2010
Registration and Video Analysis

Abstract

We propose an extension to a H.264/AVC encoder that improves the coding performance for sequences containing dynamic textures. Based on a model introduced in [1], we compute a prediction frame from already encoded frames. The encoder decides whether to use the synthesized content using an RD decision. A new macroblock mode along with an additional reference frame with flexible position are introduced for the purpose. The latter are used by means of signaling to the decoder to synthesize certain regions in an efficient way. A bitrate reduction of up to 30% over H.264/AVC at equal PSNR is achieved.

Dynamic Texture Model

For a given sequence $\{y(t)\}_{t=1\ldots\tau}$, y_{m} being the temporal mean of the sequence, we have from [1]:

$$x(t) = Ax(t-1) + Bv(t) \quad (1)$$

$$y(t) = Cx(t) + y_{\mathrm{m}} + w(t) \quad (2)$$

$$y_d(t) = Cx(t), \quad (3)$$

with initial condition for the state vector $x(0) = x_0$, an unknown stationary distribution $q(\cdot)$ with $v(t) \overset{i.i.d.}{\sim} q(\cdot)$, and given noise $w(t) \overset{i.i.d.}{\sim} p_w(\cdot)$.

References

[1] Soatto, S.; Doretto, G.; Ying Nian Wu; "Dynamic textures," in *ICCV 2001*

[2] Stojanovic, A.; Kosse, P.; "Extended dynamic texture prediction for H.264/AVC inter coding," in *ICIP 2010*

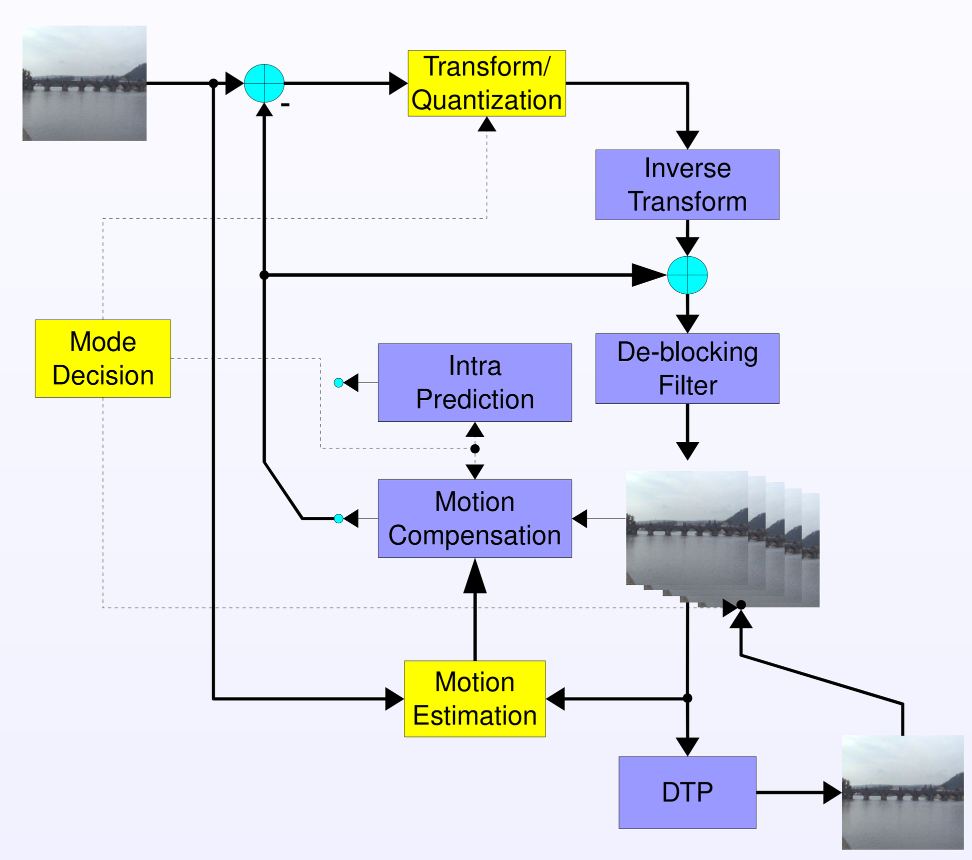
[3] `http://www.ient.rwth-aachen.de/~stojanovic/Demo/`

Acknowledgements

The present project is supported by the National Research Fund, Luxembourg.

Principle

A Dynamic Texture Prediction (DTP) module is integrated into the coding loop.



Adaptation to prediction

Only few reference pictures are typically used, and only those are use to generate the predicted reference frame. Therefore we propose an adapted version of the model, see [2]:

$$x(t) = Ax(t-1) \quad (4)$$

$$y(t) = Cx(t). \quad (5)$$

After deriving C and A from pictures in the picture buffer, we have:

$$x(t+1) = Ax(t) \quad (6)$$

$$y(t+1) = Cx(t+1). \quad (7)$$

Coding Options

Additional encoding options:

- DT reference frames (red)
- DT skip mode (NS)

Encoder Settings

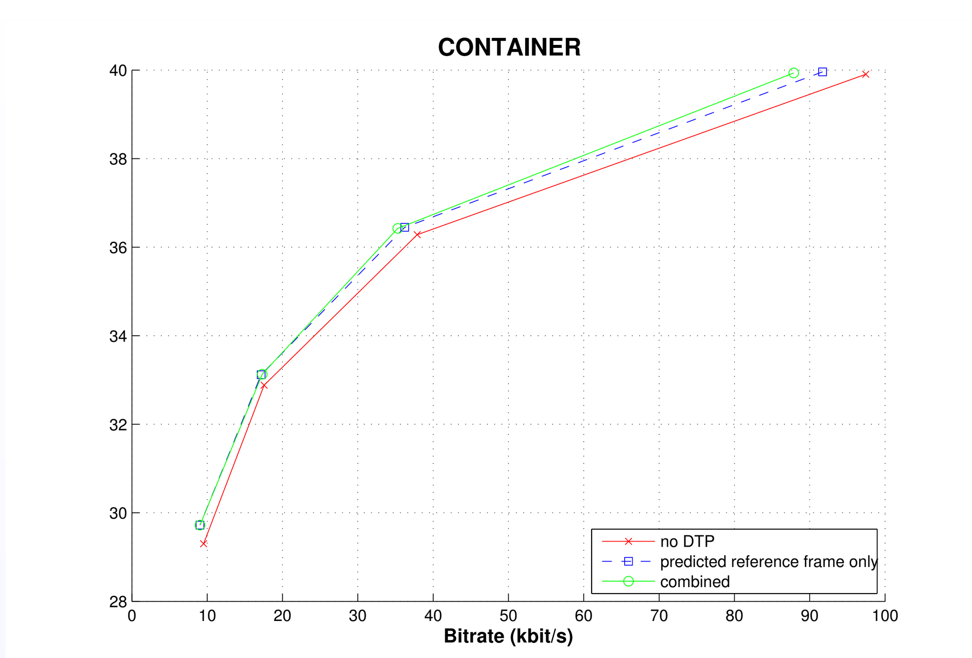
Parameter	Value
GOP Structure	IPPP
QP I	22, 27, 32, 37
QP P	23, 28, 33, 38
Search Range	32 Pixels
FREXT Profile	High
RDO	On
Entropy Coding Mode	CABAC
Frame Rate	30 frames/s
Reference Frames	5
F-Pel Distortion Metric	SAD
H-Pel Distortion Metric	Hadamard SAD
Q-Pel Distortion Metric	Hadamard SAD

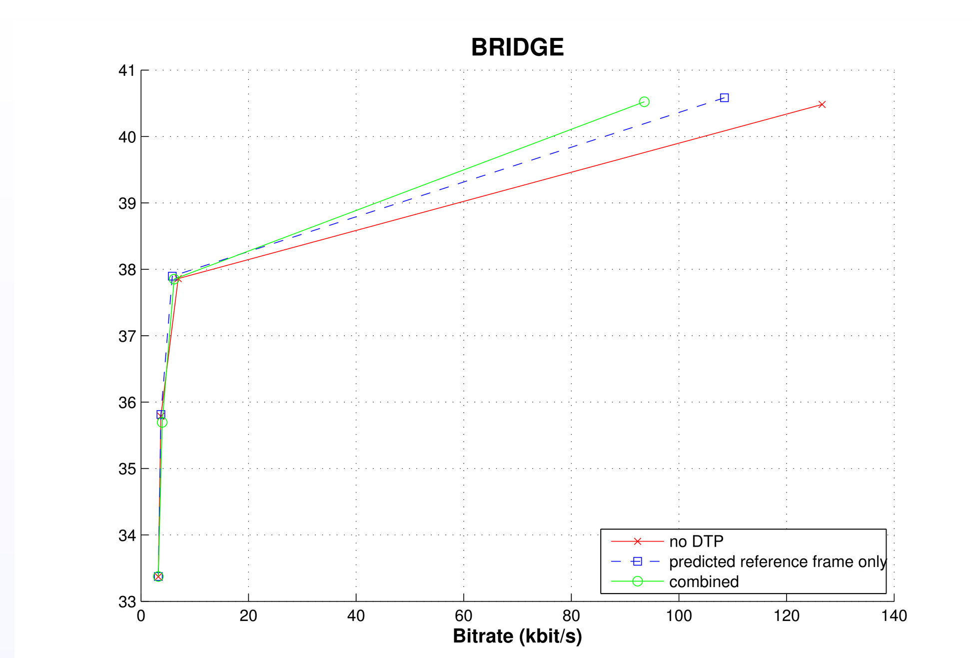
Outlook

Problems yet to be solved:

1. Handling camera pan and zoom. For preliminary results see [3].
2. Criterion for synthesis beyond common RD.

Results





Sequence	Source	Predicted Reference Frame		Combined	
		Δ PSNR [dB]	Δ Rate [%]	ΔPSNR [dB]	ΔRate[%]
Bridge-close	Orig.	0.058	-2.192	0.097	-3.647
Container	Orig.	0.385	-7.922	0.438	-8.866
Preakness	Orig.	0.094	-2.232	0.131	-3.044
Rushhour	Crop.	0.084	-2.590	0.109	-3.003
Sean	Orig.	0.095	-1.729	0.102	-1.849
Average		0.1432	-3.333	0.1754	-4.082