Deep frames interpolation for video compression

Jean Bégaint[1,2], Franck Galpin[1], Philippe Guillotel[1], Christine Guillemot[2]

[1]Technicolor R&I, [2]INRIA, Rennes, France

Context

- Video codecs are primarily designed assuming that **rigid, block-based, two-dimensional displacements** are suitable models to describe the motion taking place in a scene, or otherwise rely on splitting complex motions into several PUs
- **Translational models** are not sufficient to handle real world motion types such as **camera zoom**, **shake**, **pan**, ...
- **Video frame interpolation** methods have recently benefited from significant improvements by leveraging **deep convolutional networks**, *e.g.:* for slow-motion applications.

Fig. 1: Ref 0

Fig. 3: Ref 1

Fig. 2: Interpolation result

Fig. 4: Ground truth

Deep frames interpolation

- Given two consecutive input frames, a network is used to interpolate an intermediate frame via a **dense pixel-wise motion estimation and compensation**
- Each frame coded from past and future references can benefit from an additional interpolated reference
- Extension of the classic *random access* configuration mode to support extra reference frames

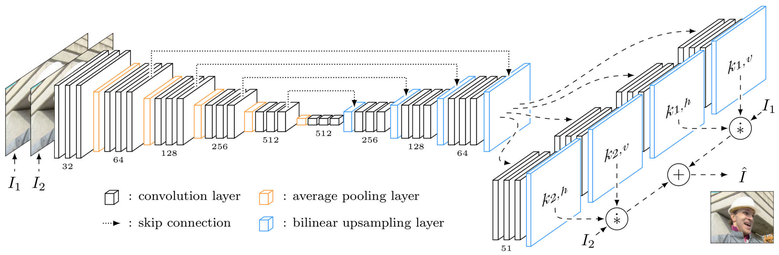


Fig. 5: *SepConv* architecture from Niklaus *et al.* [1]

Coding Scheme

- Interpolated frame used as an extra reference frame (SKIP, MERGE, INTER still applicable)
- Mode signaling performed at the frame and PU level, with a CABAC context from top and left PU blocks.

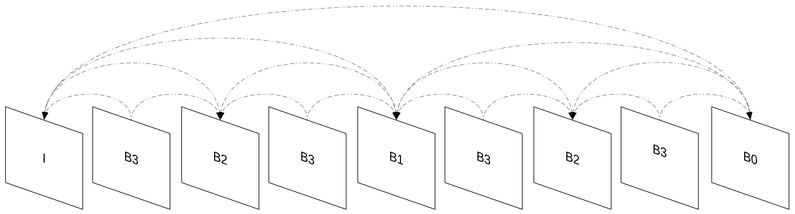


Fig. 6: Illustration of the additional references for a GOP of 8 frames

Experimental Results

- Rate-distortion performances are computed with the Bjontegaard metric using QP 22, 27, 32, 37 (luma only), in *random access* configuration (GOP 16).
- Coding experiments demonstrate that the proposed method outperforms HEVC

Compression performances

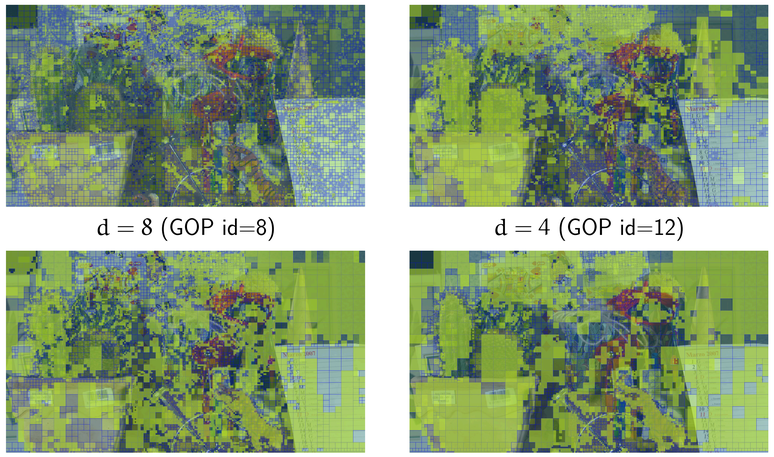
Tab. 1: BD-rate gains *w.r.t* HEVC.

Class	Sequence	SepConv-l1	SepConv-lf	DVF [2]	C. Flow [3]
B	BQTerrace	**-6.38%**	-5.94%	–	-2.77%
	BasketballDrive	**-1.43%**	-1.05%	–	-0.35%
	Cactus	**-7.36%**	-7.07%	–	-4.56%
	Kimono1	**-2.05%**	-1.71%	–	-0.23%
	ParkScene	**-2.36%**	-2.14%	–	-0.41%
C	BQMall	**-3.44%**	-2.89%	-0.10%	-0.34%
	BasketballDrill	**-3.87%**	-3.77%	-0.71%	-1.17%
	PartyScene	-2.53%	**-3.08%**	-0.14%	-0.37%
	RaceHorses	**-1.30%**	-1.01%	-0.03%	-0.47%
D	BQSquare	-3.31%	**-5.48%**	0.10%	0.06%
	BasketballPass	**-2.93%**	-2.51%	-0.13%	-0.34%
	BlowingBubbles	-2.08%	**-2.29%**	-0.13%	-0.24%
	RaceHorses	**-2.21%**	-1.84%	-0.06%	-0.29%
Affine	CStoreGoods	**-0.45%**	-0.29%	–	-0.25%
	DrivingRecorder1	**-1.37%**	-0.93%	0.04%	-0.07%
	DrivingRecorder2	**-1.24%**	-1.03%	0.01%	-0.24%
	LakeWalking	**-0.08%**	0.00%	0.03%	-0.05%
	ParkSunny	**-1.34%**	-1.34%	–	-0.82%
	ParkWalking	**-0.54%**	-0.32%	–	-0.03%
Other	foreman	**-2.92%**	-2.82%	0.08%	-0.32%
Mean		**-2.46%**	-2.38%	-0.09%	-0.89%

Proposed Mode Selection

- The same network is used for all interpolation distances (d)

Fig. 7: *Cactus* sequence (QP=22), d: reference frames distance



d = 8 (GOP id=8) d = 4 (GOP id=12)

d = 2 (GOP id=14) d = 1 (GOP id=15)

Encoded from the interpolated references

Signaling cost

Tab. 2: With and without signaling

Signaling	Class B	Class C	Class D	Affine
Yes	-3.93%	-2.60%	-2.62%	-0.86%
No	-4.15%	-2.90%	-3.00%	-1.04%

Run times

Tab. 3: Mean interpolation times

Method	Platform	Interpolation time (seconds)		
		Class B	Class C	Class D
Celiu Flow	CPU	52.53	9.90	2.43
DVF	GPU	–	–	0.13
SepConv	GPU	0.99	0.30	0.10

Full-frame interpolation (*worst* case)

Future work

- Beyond linear and uniform motions

(a) Ref 0 (b) Ref 1 (c) Interpolation (d) Groundtruth

References

[1] S. Niklaus *et al.*, "Video frame interpolation via adaptive separable convolution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 261–270.

[2] Z. Liu *et al.*, "Video frame synthesis using deep voxel flow," in *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, 2017, pp. 4473–4481.

[3] C. Liu *et al.*, "Beyond pixels: exploring new representations and applications for motion analysis," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

jean.begaint@technicolor.com

DCC 2019, Snowbird, UT, USA