Region-based models for motion compensation in 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.
- **Translational models** are not sufficient to handle real world motion types such as **camera zoom**, **shake**, **pan**, ... where large objects are split into **multiple coding blocks**, whereas an homography could be used instead.
- However, **a single homography** is **not always sufficient**, especially for **non-planar scenes** and scenes with **strong parallax** [1].

(a) (b)

Fig. 1: Example inputs: (a) current frame, (b) reference frame.

Proposed Approach

- We present here a **region-based inter-prediction scheme** able to compensate **complex scene motion**, by estimating multiple geometric models [2].
- We also introduce an **affine photometric correction** for each geometric model to compensate **local illumination distortions**.

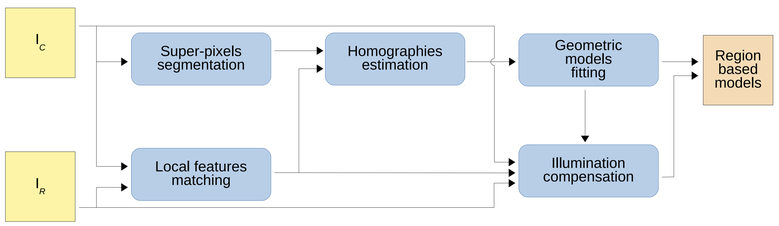


Fig. 2: Flowchart of the proposed solution.

Local Models Estimation

- The current frame is clustered into **super-pixels** via **SLIC**:

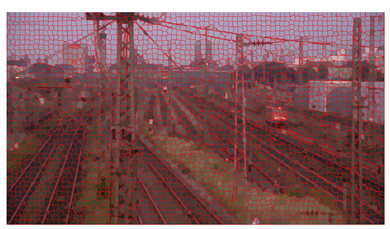


Fig. 3: Super-pixel segmentation of the current frame.

- **SIFT** local feature descriptors are detected and extracted, then **matched** via an approximate **nearest neighbour** algorithm.

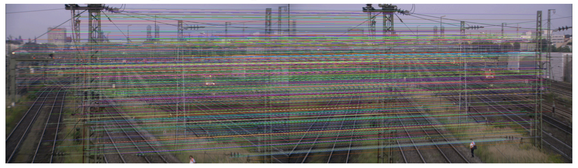


Fig. 4: Keypoint matches between the current and reference frame.

- From the matched keypoints, **homographies** H_k, are estimated for each **super-pixel** via **RANSAC**:

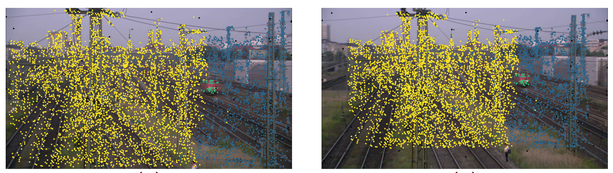
$$H = \begin{bmatrix} s_x.\cos(\theta) & -s_y.\sin(\theta+\sigma) & t_x \\ s_x.\sin(\theta) & s_y.\cos(\theta+\sigma) & t_y \\ k_x & k_y & 1 \end{bmatrix}$$

Region-based Models Estimation

- The **geometric models** are recursively **re-estimated** and fitted [3] to keep only the most significant transformation models, by minimizing an objective function:

$$E(f) = \overbrace{\sum_{p\in P} D_p(f_p)}^{\text{data cost}} + \overbrace{\sum_{(p,q)\in N} V_{pq}(f_p, f_q)}^{\text{smooth cost}} + \overbrace{\sum_{L\subseteq\mathcal{L}} h_L.\delta_L(f)}^{\text{label cost}}$$

- *data cost*: Symmetric Transfer Error term (fidelity term)
- *smooth cost*: Delaunay triangulation of SIFT keypoints (spatial coherence)
- *label cost*: restricts the number of models



(a) (b)

Fig. 5: matched region descriptors.

- Affine **photometric compensations** are estimated for each detected **region**:

$$\underset{\alpha,\beta}{\operatorname{argmin}} \sum_{P} |Y'(x'_p) - (\alpha.Y(x_p) + \beta)|^2$$

Coding Scheme

- Region-based models are estimated between the ***I*** frame and each ***B*** frame of the GOP.
- The parameters of the *selected* geometric and photometric models are stored in the bitstream for each frame.

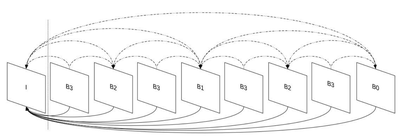


Fig. 6: Coding scheme example for a GOP of 8 frames in *random access* mode.

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** (**RB**) **outperforms** both **HEVC** [4] and a **global motion compensation** (**GLB**).
- The additional luminance compensation (**GLB-L**, **RB-L**) contributes up to -1.2% of the BD-rate improvements (-0.33% on average).

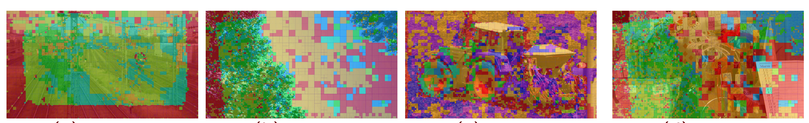
Tab. 1: BD-rate gains *w.r.t* HEVC.

	BD-rate	GLB	GLB-L	RB	RB-L
	CStoreGoods_720x1280	-2.29%	-2.29%	-2.58%	**-2.71%**
	DrivingRecorder1_720x960	0.16%	0.16%	-0.90%	**-1.05%**
	DrivingRecorder2_720x960	0.11%	0.07%	-1.63%	**-1.81%**
	LakeWalking_720x960	-7.06%	-7.06%	-9.26%	**-10.45%**
(1)	ParkSunny_720x1280	-0.13%	-0.37%	-0.28%	**-0.74%**
	ParkWalking_720x1280	-2.04%	-2.04%	-2.73%	**-3.11%**
	bluesky_1920x1080	-3.90%	-4.21%	-3.97%	**-4.32%**
	station2_1920x1080	-11.28%	-11.75%	-12.13%	**-13.27%**
	tractor_1920x1080	0.53%	0.16%	**-0.27%**	**-0.27%**
	average	-2.88%	-3.04%	-3.75%	**-4.19%**
	B_Cactus_1920x1080	0.02%	-0.03%	-3.16%	**-3.17%**
	city_1280x720	0.13%	0.13%	0.00%	**-0.26%**
	in_to_tree_1280x720	-0.28%	-0.28%	**-0.54%**	**-0.54%**
(2)	shields_1280x720	-2.06%	-2.06%	**-2.13%**	**-2.13%**
	average	-0.55%	-0.56%	-1.46%	**-1.52%**
	Total average	-2.16%	-2.27%	-3.04%	**-3.37%**

Tab. 2: Complexity increases *w.r.t.* HEVC.

Complexity	GLB	GLB-L	RB	RB-L
Encoder	190%	190%	300%	302%
Decoder	174%	173%	196%	199%

Proposed Mode Selection



(a) "station2" (b) "bluesky" (c) "tractor" (d) "Cactus"

Fig. 7: Red and orange blocks are coded with classical intra and inter prediction, respectively. Predictions made by our mode are depicted with the other colors.

References

[1] S. Parker *et al.*, "Global and locally adaptive warped motion compensation in video compression," in *ICIP 2017*.

[2] J. Bégaint *et al.*, "Region-based prediction for image compression in the cloud," *IEEE TIP 2018*.

[3] A. Delong *et al.*, "Fast approximate energy minimization with label costs," *IJCV 2012*.

[4] G. J. Sullivan *et al.*, "Overview of the high efficiency video coding (HEVC) standard," *IEEE TCSVT 2012*.

jean.begaint@technicolor.com

PCS 2018, San Francisco, CA, USA