Divergence-Based Adaptive Extreme Video Completion

Majed El Helou, Ruofan Zhou, Frank Schmutz, Fabrice Guibert and Sabine Süsstrunk

IEEE International Conference on Acoustics, Speech, and Signal Processing

ICASSP 2020

Introduction

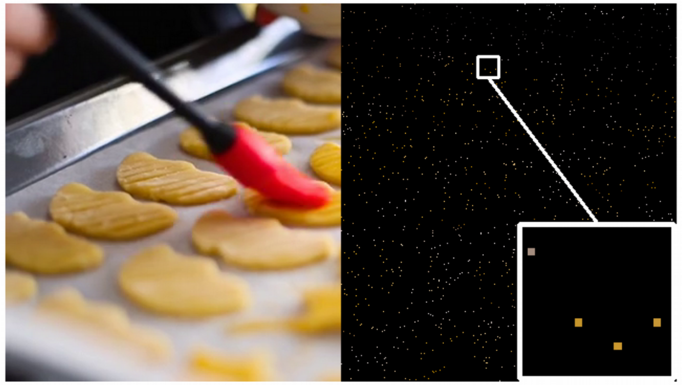
- A very computationally-cheap approach to sample/compress images or videos is to randomly sample pixels at extremely low rates.

- This is advantageous for instance in low-power IoT devices, or for emergency-deployed network infrastructures.

- The saved pre-processing computations however come at the cost of a challenging reconstruction.

Background

- Most inpainting/hole filling techniques rely on the presence of at least a sufficient continuous portion of images.
- Methods applicable to extreme image completion are very computationally expensive. EFAN[1] proposes an efficient such method.



1% sampling

- We propose to extend EFAN from extreme image to extreme video completion.

[1] Radhakrishna Achanta, Nikolaos Arvanitopoulos, and Sabine Süsstrunk, "Extreme image completion," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 1333–1337.

EFAN

- We refer to EFAN as EFAN2D. Its interpolation follows a normalized filtering of the sparse image *I* by the special filter G. (*i*,*k*) are pixel locations, and *N* a neighborhood around *i*.

$$J[i] = \frac{\sum_{k \in \mathcal{N}} G[i,k] I[k]}{\sum_{k \in \mathcal{N}} G[i,k]}$$

$$G[i,k] = e^{-0.5 \frac{(i_x - k_x)^2 + (i_y - k_y)^2}{\sigma^2}}$$

EFAN3D: Video Extension

- A straight-forward extension is EFAN3D, where a temporal aspect is added to the filter.

$$J[i] = \frac{\sum_{k \in \mathcal{N}} G[i,k] G_t[i,k] I[k]}{\sum_{k \in \mathcal{N}} G[i,k] G_t[i,k]}$$

$$G_t[i,k] = e^{-0.5 \frac{(i_z - k_z)^2}{\sigma_t^2}}$$

Results of EFAN2D and EFAN3D

EFAN2D

EFAN3D

Adaptive-Depth EFAN (ADEFAN)

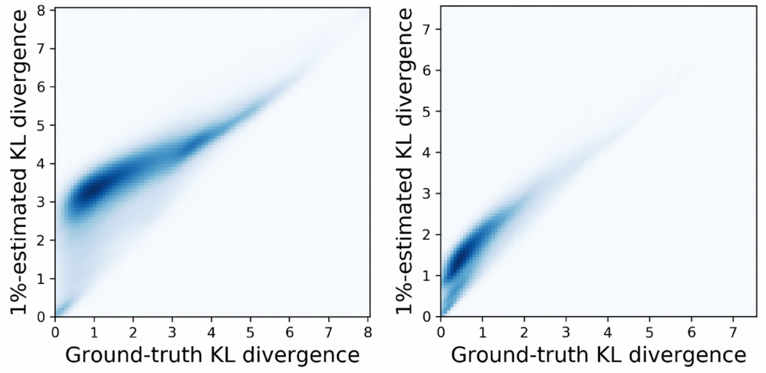
- Filtering frame by frame causes excessive flickering, and 3D filtering can cause over-smoothness of motion.
- We propose to adapt the depth of the filter based on color motion.
- **Challenge**: pixels are extremely sparse.

ADEFAN Cont'd

- The key idea of ADEFAN is to leverage KL-divergence between temporally-adjacent windows to adapt the filter's depth.

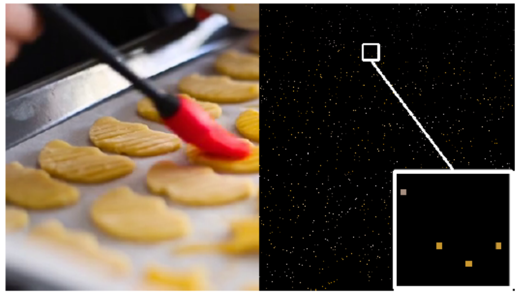
$$f(div_{next}, fr_{max}) = \left\lceil \frac{fr_{max}}{1 + \beta div_{next}} \right\rceil$$

- Window sizes are pre-calibrated and adjust to the sampling rate.



(a) 80 × 80 windows (b) 160 × 160 windows

Visual Results



(a) Reference | 1%-sampled frame

(b) EFAN2D (20.75 dB)

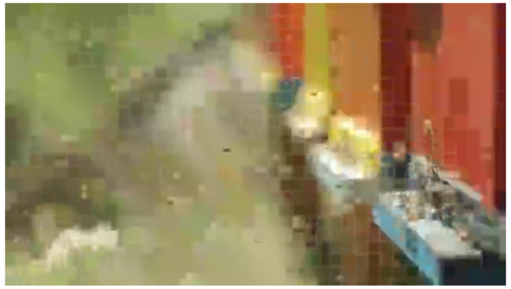
(c) EFAN3D (20.39 dB)

(d) ADEFAN (**22.02 dB**)

(a) Reference video frame

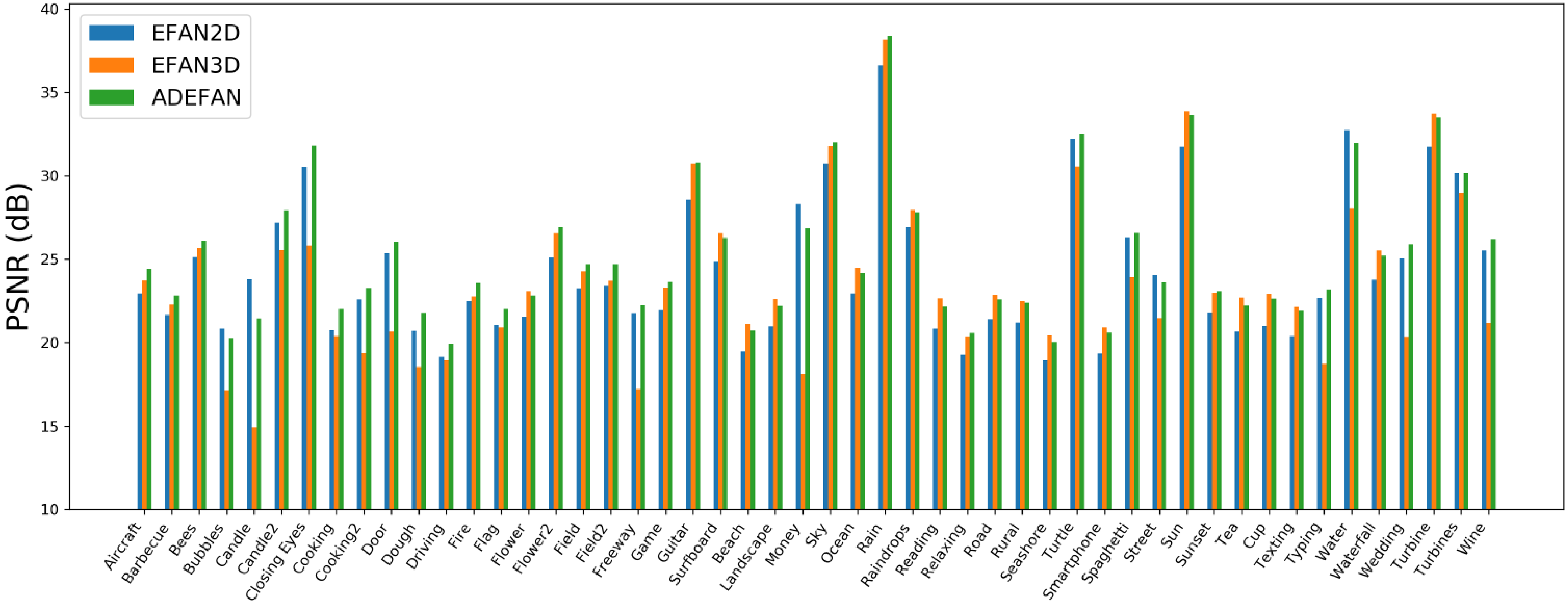
(b) ADEFAN (1%) (26.53 dB)

(c) ADEFAN (2%) (**28.09 dB**)



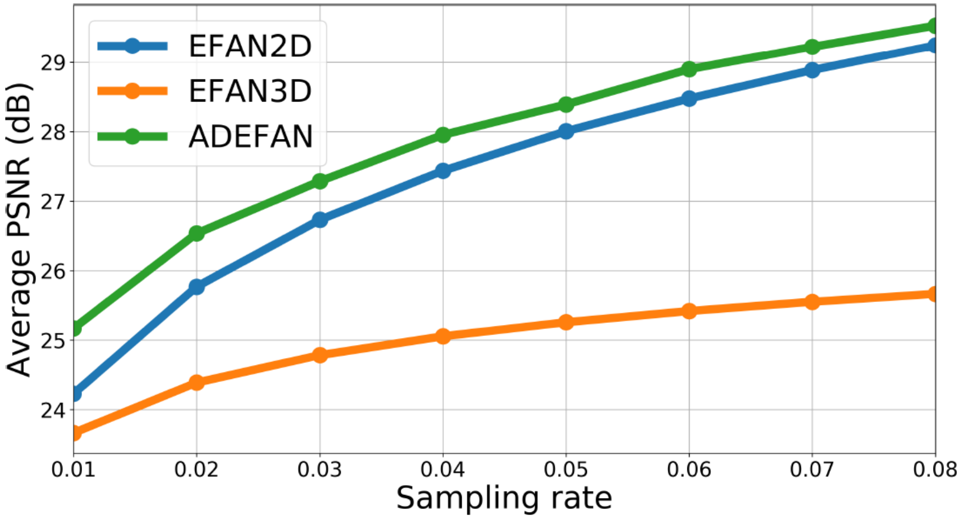
(d) MPEG-4 (≈2%) (26.03 dB)

Results on 50 videos (with 1% sampling)



Experimental Results Cont'd

Evaluation across sampling rates



Evaluation through a survey

Opinion Score	EFAN2D	EFAN3D	ADEFAN
Reconstruction	<u>5.53</u>	3.72	**6.52**
Visual Quality	<u>5.05</u>	4.12	**6.34**

original ADEFAN

EFAN3D EFAN2D

original ADEFAN

EFAN3D EFAN2D

Conclusion

- We present an algorithm, ADEFAN, for extreme video completion.
- ADEFAN can reconstruct videos from 1% of pixels, sampled randomly.
- Our reconstruction results are both more accurate and more visually pleasing than the available baselines.
- ADEFAN provides an efficient video encoding algorithm.

Thank you

https://github.com/majedelhelou/ADEFAN