DCC 2021

Flow-grounded Dynamic Texture Synthesis for Video Compression

Suhong Wang, Xinfeng Zhang, Shanshe Wang, Siwei Ma and Wen Gao

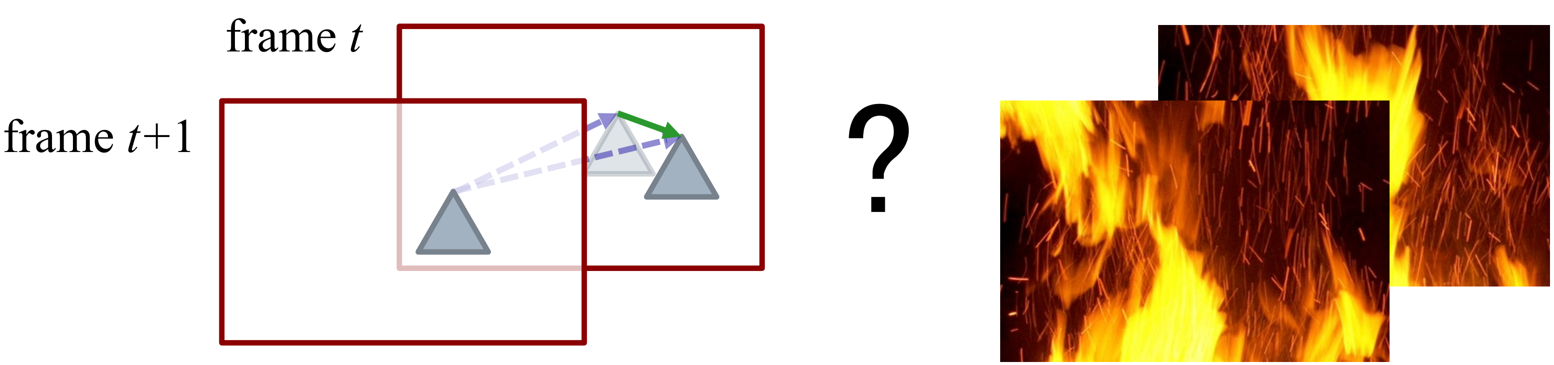
Institute of Digital Media, Peking University

Background

- **Video contents can be classified into three categories:**
  - Static contents (e.g. walls, fabric, surface of stones)
  - Activities (e.g. translation motion, rotation and scaling)
  - **Dynamic textures** (e.g. water surfaces, smoke, fire, clouds)
- **What are dynamic textures (DT) ?**
  - Time-varying motion patterns which exhibit certain temporal stationarity.
  - Usually existing non-linear motion.

Challenges

- **Block-based predictive coding scheme shows poor performance for DT contents.**
  - Rapid change over time and randomness.
  - Cannot obtain prediction blocks with high similarity in pixel level from reference frames.



Challenges

- **Block-based predictive coding scheme shows poor performance for DT contents.**
  - Prediction characteristics analysis:
    - Small block partition tends to occur for DT contents.
    - Residual of DT contents are much higher than other contents.

(a) non-DT sequence (b) DT sequence

Fig. 1 Comparisons of partition results.

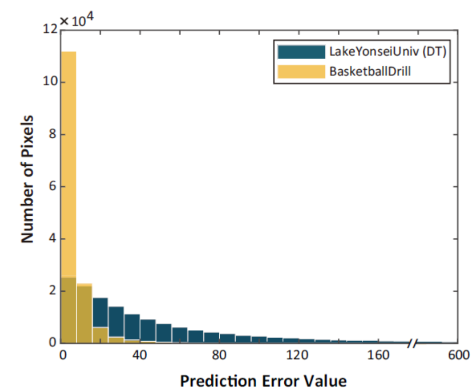


Fig. 2 Distribution of prediction error value of two sequences.

- **A more effective compression scheme should be investigated.**

Method: Overview

- An analysis-synthesis video compression scheme is proposed based on temporal characteristics of DT.
  - DT motion is first analyzed to generate flowlines and measure the period value of a given DT sequence.
  - After encoding key frames, flow-grounded DT synthesis is performed to replace traditional coding process of the un-coded frames.

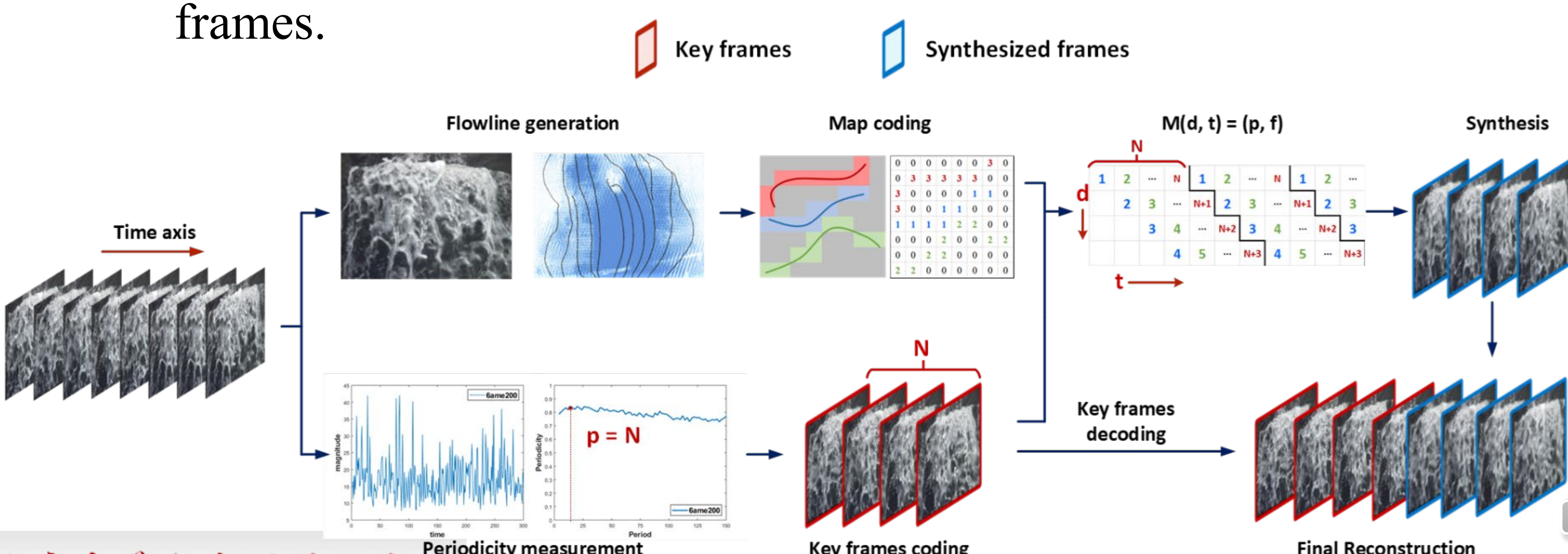


Fig. 3 The proposed scheme for DT compression.

Method

- **Particle movement for DT synthesis**
  - Textons in a DT sequence are denoted as a set of particles.
  - The synthesis procedure can be implemented by sequentially moving the particles along the flowlines over time at period *N*.

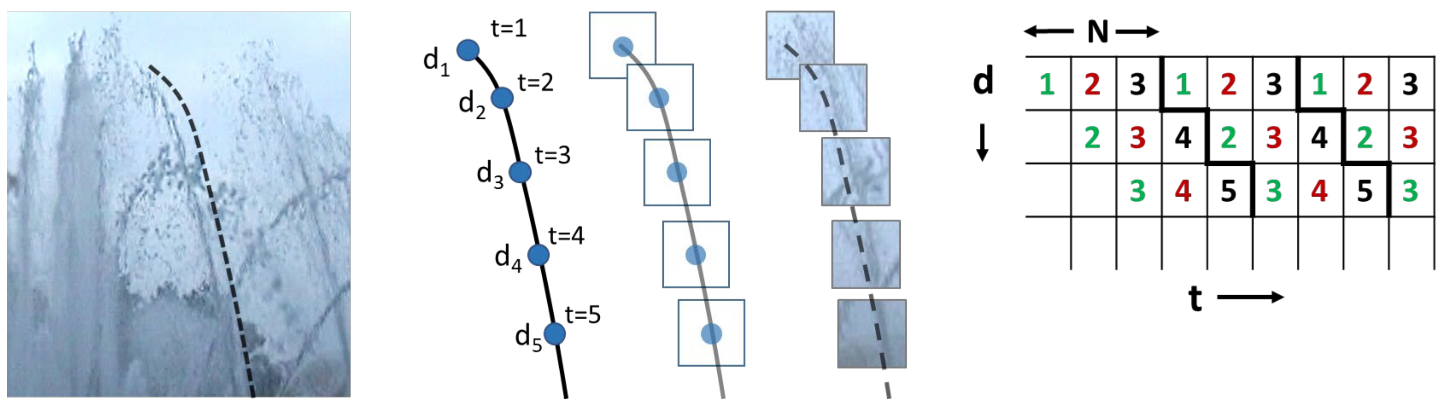


Fig. 4 DT synthesis process.

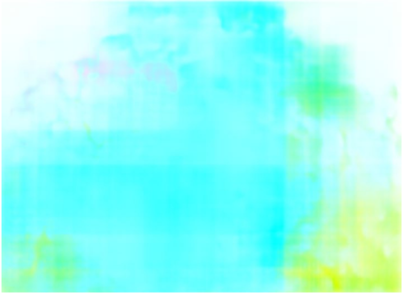
Method

- **Flowline generation**
  - For a spatial point x, the differential equation is defined as follows:

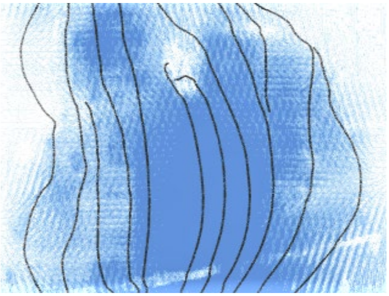
$$\frac{dx}{dt} = v(x)$$

  - With s_0 represents the starting point and u refers to the integration variable. The position at time t is given by:

$$s(t) = s_0 + \int_{0 \le u \le t} v(s(u))du, \qquad s_{i+1} = s_i + v(s_i) \cdot d_t$$



(a) Optical flow fields



(b) Generated flowlines

Fig. 5 An example of flowline generation process.

Method

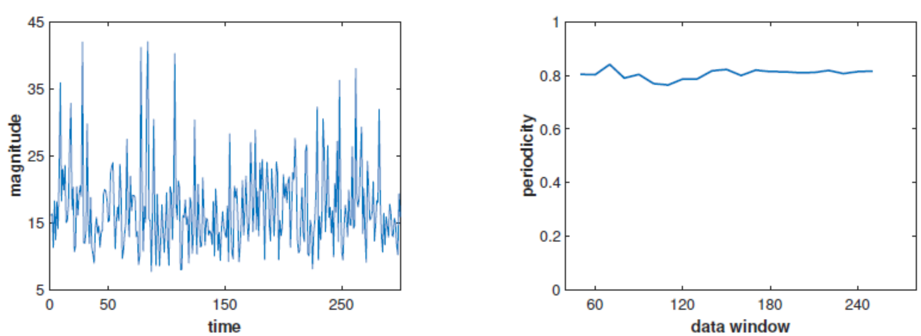
- **Periodicity measurement**
  - Calculate the motion feature A_n of each frame by wavelet transform.
  - Applying SVD to detect periodicity of the motion signal.

$$A_\mathrm{n} = USV^T,\ S = diag(s_1, s_2, \cdots, s_r: 0)$$

$s_1 \gg s_2$ indicates a nearly periodic component of the length n.

  - To find the most appropriate period value:

$$N = argmax_n \left(\frac{1}{W}\sum_{i=1}^{W} 1 - \frac{s_2(n,i)}{s_1(n,i)}\right)$$



(a) original signal (b) periodicity

Fig. 6 Periodicity analysis for a DT sequence.

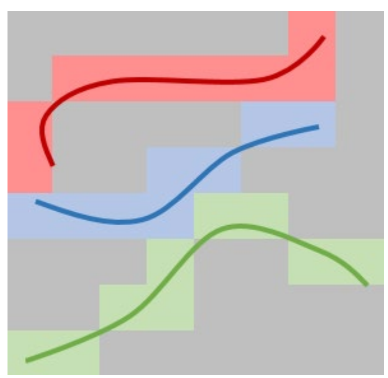
Method

- **Integration into a VTM-10.0**
  - Number of key frames is set according to the analyzed period N.

$$N_{final} = k \cdot GOPsize, k = \lfloor N/GOPsize \rfloor$$

$$N_{key} = 2 \times (N_{final} - 1)$$

  - Flowline distribution needs to be sent to decoder side. Run length coding is extended in this work to encode the map.



0	0	0	0	0	0	3	0
0	3	3	3	3	3	0	0
3	0	0	0	0	1	1	0
3	0	0	1	1	0	0	0
1	1	1	1	2	2	0	0
0	0	0	2	0	0	2	2
0	0	2	2	0	0	0	0
2	2	0	0	0	0	0	0

Fig. 7 Map coding for flowlines.

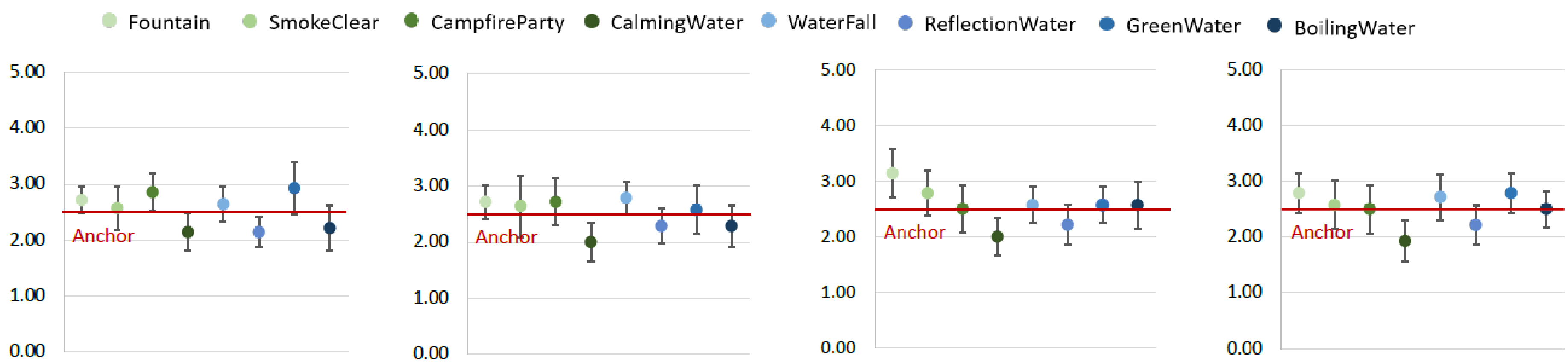
Experiments (1/2)

- **Bitrate Savings**
  - $\Delta Rate = (R - R_p)/R \times 100\%$

Resolution	Sequence	Database	Period	Bitrate Saving(%)	
				RA	LDB
1024×1024	Fountains	SJTU 4K	16	67.85	51.99
	SmokeClear	BVI textures	32	33.82	22.53
512×512	CampfireParty	SJTU 4K	16	52.61	56.84
	CamlingWater	BVI textures	32	42.01	45.06
256×256	WaterFall	HomTex	16	18.96	21.79
	ReflectionWater		16	57.38	63.92
	GreenWater		16	86.62	87.57
	BoilingWater		32	16.87	18.70
Average				**47.02**	**46.05**
Encoding Time(%)				**37.31**	**31.14**

Experiments (2/2)

- **Subjective evaluation**
  - A resulting MOS value of 2.5 indicates that anchor and proposed method were rated to have the same visual quality.
  - These DT contents which are difficult to encode can use synthesized results with comparable visual quality.



(a) QP = 22 (b) QP = 27 (c) QP = 32 (d) QP = 37

Thanks!