Continuous Space-Time Video Resampling with Invertible Motion Steganography

Yuantong Zhang, Zhenzhong Chen
Wuhan University

Introduction

➢ **Task Introduction**:

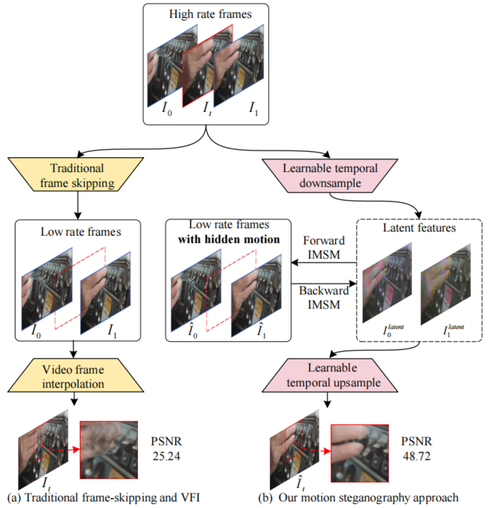
Space-time video resampling aims to optimize both spatialtemporal downsampling and upsampling processes to achieve high-quality video reconstruction.

➢ **Challenges**:

(1) Effective modeling of temporal resampling.

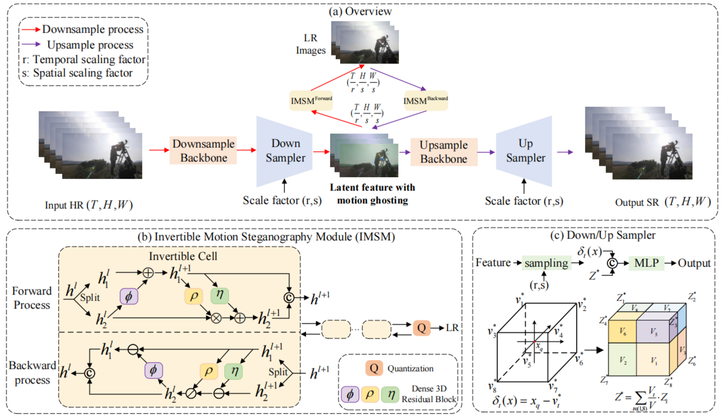
(2) Continuous space-time resampling.

➢ **Motivation:**



(a) Traditional ”frame-skipping and VFI”: Motion information is discarded, causing blur in reconstructed frames. (b) Our motion steganography approach: The IMSM reversibly embeds motion information into the low-frame-rate results in a imperceptible manner, aiding in frame reconstruction.

Methodology



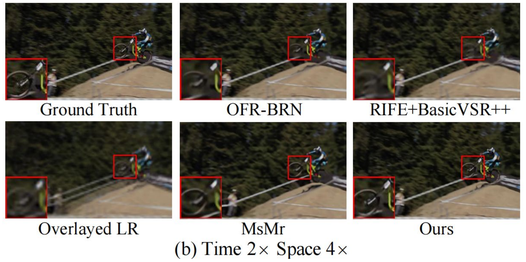
➢ The architecture and data flow of the proposed approach.
Downsampling process: The input HR frames are first downsampled to obtain latent features with motion ghosting, which are then passed through the IMSM to generate LR images (indicated by red arrows).
Upsampling process: The LR images are first converted back to latent features with inverse transformation. Then, multiple frames are mutually enhanced and upsampled to restore the SR output (indicated by purple arrows).

➢ The Invertible Motion Steganography Module (IMSM) is composed of invertible flow cells, which can embed motion information into LR images and reverse the process when needed.

➢ The Down/Up sampler modulates the input features to the desired shape using 3D local implicit modulation.

Experimental Results

Upscale rate time/space	Downsampler time/space	Reconstruction method	Params/M	Vimeo-90k RGB-Space	Vimeo-90k Y-Space	Vid4 RGB-Space	Vid4 Y-Space
2× / 1×	Nearest/-	XVFI [41]	5.7	34.16/0.9456	36.16/0.9603	28.50/0.9360	30.26/0.9477
		FLAVR [20]	42.1	36.27/0.9599	38.38/0.9716	29.39/0.9361	31.17/0.9467
		EDSC [8]	8.9	34.48/0.9489	36.35/0.9616	29.20/0.9372	31.05/0.9491
		IFRNet [22]	19.7	35.22/0.9544	37.12/0.9668	29.55/0.9406	31.39/0.9519
		RIFE [16]	20.9	35.25/0.9535	37.28/0.9662	28.41/0.9335	30.15/0.9454
		EMA-VFI [58]	65.7	36.08/0.9585	38.02/0.9697	29.80/0.9434	31.62/0.9538
		VFIFormer [31]	24.1	35.93/0.9581	37.88/0.9689	28.94/0.9400	30.70/0.9506
	STAA[50]	STAA [50]	15.9	45.01/0.9912	-	39.78/0.9926	-
	ours	Ours	11.7	48.03/0.9949	50.52/0.9976	43.75/0.9955	46.77/0.9974
2× / 4×	Nearest/Bicubic	EDSC [8]+BasicVSR++ [2]	8.9+7.3	32.79/0.9092	34.51/0.9252	24.48/0.7572	26.27/0.7900
		IFRNet [22]+BasicVSR++ [2]	19.7+7.3	32.88/0.9098	34.70/0.9268	24.52/0.7568	26.36/0.7917
		RIFE [16]+BasicVSR++ [2]	20.9+7.3	32.81/0.9091	34.56/0.9253	24.51/0.7582	26.37/0.7939
		ZSM [51]	11.1	33.48/0.9178	35.21/0.9323	24.82/0.7763	26.31/0.7976
		STDAN [46]	8.3	33.59/0.9192	35.50/0.9351	24.91/0.7832	26.28/0.8041
		MoTIF [4]	12.6	32.31/0.9098	34.03/0.9265	24.26/0.7527	25.79/0.7752
		RSTT [12]	7.7	33.66/0.9192	35.39/0.9336	24.94/0.7805	26.43/0.7994
		OFR-BRN [63]	11.77	33.71/0.9206	35.53/0.9358	25.21/0.7957	26.72/0.8141
		MsMr [64]	13.19	33.76/0.9214	35.57/0.9365	25.41/0.8044	26.87/0.8213
	STAA [50]	STAA [50]	16.0	34.53/0.9426	-	27.31/0.9173	-
	Ours	Ours	17.4	37.78/0.9624	40.57/0.9770	30.71/0.9336	32.76/0.9488



(b) Time 2× Space 4×

Our method significantly outperforms other comparison methods in both objective metrics and subjective quality.

Summary

➢ we tackle two critical challenges in video resampling. To preserve motion information during temporal, we develop the invertible motion steganography, which effectively embeds motion data from high-frame-rate videos into downsampled frames.

➢ To achieve flexible spatiotemporal resampling, we propose a 3D implicit modulation technique that allows for spatiotemporal resampling, including non-integer frame rate conversions.

➢ Our experiments demonstrate the effectiveness of these contributions across various video resampling tasks