Low Rate Compression of Video With Dynamic Backgrounds

Solomon Garber, Ryan Marcus*, Antonella DiLillo and James Storer

Brandeis University *MIT CSAIL

{solomongarber,dilant,storer}@brandeis.edu ryanmarcus@csail.mit.edu

Introduction

The content in the background of a video is often of low semantic importance, but may be nonetheless difficult to compress due to complex dynamic motion. In video regions where semantic importance is low, visual quality becomes more important than signal fidelity metrics such as PSNR or SSIM in concealing compression artifacts.

- We present a system for compressing videos with dynamically moving backgrounds and semantically salient foregrounds
- Input streams are semantically segmented by a MaskRCNN[3] segmentation network.
- Foreground, background, and mask streams are stored separately

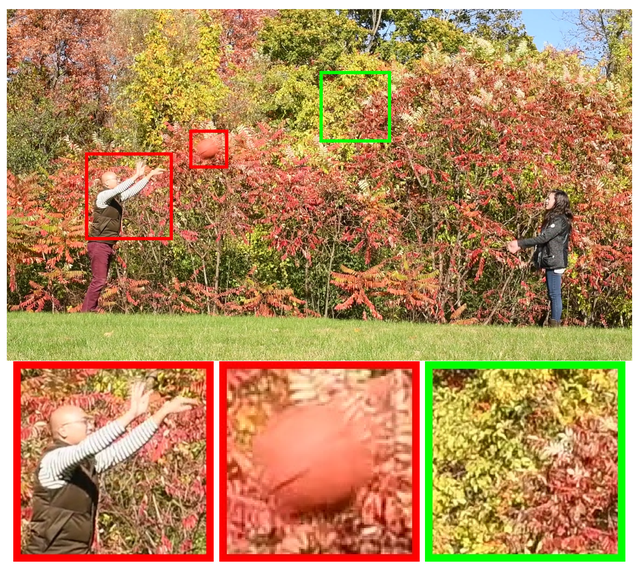


Fig. 1: Example frame containing semantic foreground (red) and dynamic background (green). Our system targets different objectives in different regions of the frame, prioritizing signal fidelity in the semantic regions and perceptual quality in the background.

Semantic Foreground Segmentation

Using semantic masks generated by a pre-trained MaskRCNN, the foreground regions of a video can be separated and compressed using a traditional codec such as AVC or HEVC.

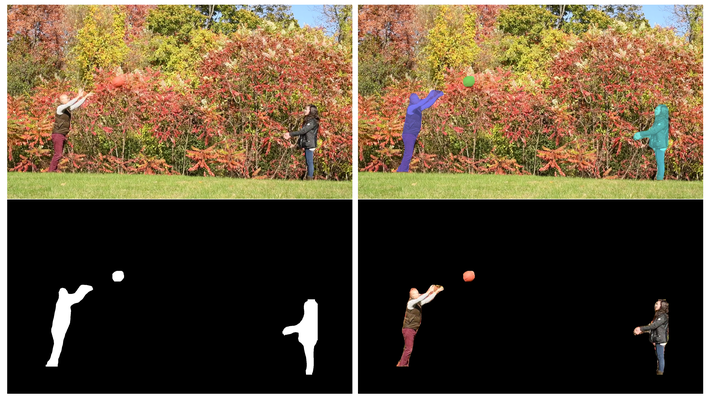


Fig. 2: Input frame (top left), instance level semantic segmentation masks generated by MaskRCNN (top right), with corresponding foreground mask (bottom left) and foreground content (bottom right). The bottom foreground and mask streams are compressed with AVC or HEVC and stored separately.

Dynamic Background Sprites

Dynamic background motion can be spectrally analyzed[1] stored as a dynamic motion sprite [2]

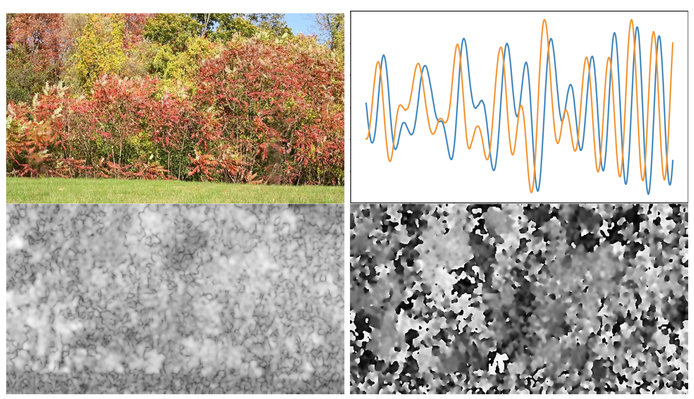
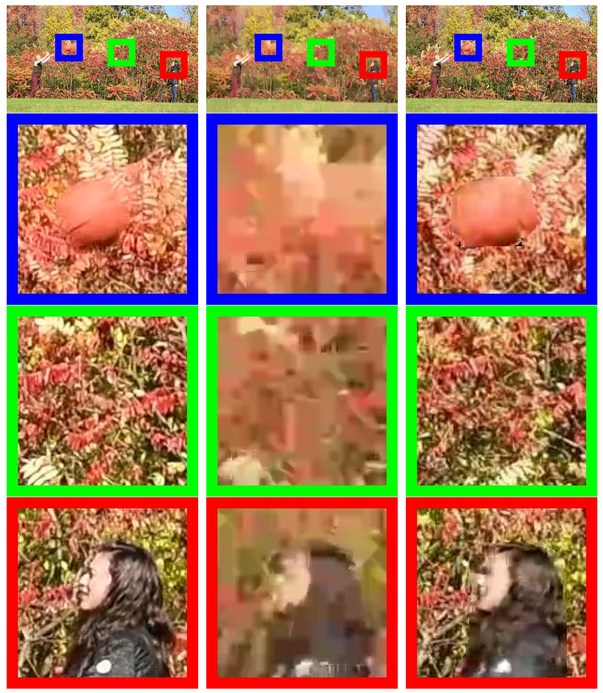


Fig. 3: Example dynamic motion sprite containing static background (upper left), horizontal motion signal (upper right), and the log amplitude and phase of the complex mode of horizontal motion (lower left and right).

Results

Below is comparison between our system and off-the-shelf ffmpeg implementation of AVC. On left is a still frame taken from a sample video, encoded in camera at original high quality AVC 34 Mbit/s. Middle is the same frame taken from a low quality ffmpeg AVC of the video, at 258 Kbit/sec. Right is the same frame compressed by our method at 258Kbit/s. Click the frames to view the video or go to: **rm.cab/semvid1**



References

[1] Abe Davis, Justin G Chen, and Frédo Durand. “Image-space modal bases for plausible manipulation of objects in video”. In: *TOG* 34.6 (2015), p. 239.

[2] Solomon Garber et al. “Compact Representations of Dynamic Video Background Using Motion Sprites”. In: *2019 Data Compression Conference (DCC)*. IEEE. 2019.

[3] Kaiming He et al. “Mask R-CNN”. In: *ICCV* (2017), pp. 2980–2988.