Time Dilation of Videos

Optimizing Video Playback for Maximum Content Delivery

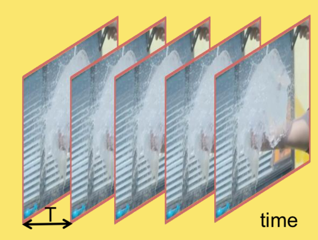
Jihui Jin and Erik Jorgensen

Introduction

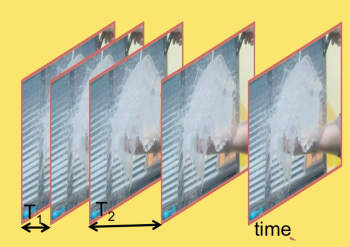
Context: Time manipulation of videos is not a new concept, but it is a tedious, manual process.

- Action scenes are slowed down for humans to fully process
- Mundane activities constitute too large a portion of our viewing time, but provide valuable context

Objective: Automatically compute a time-varying frame rate for optimal video playback



Original video

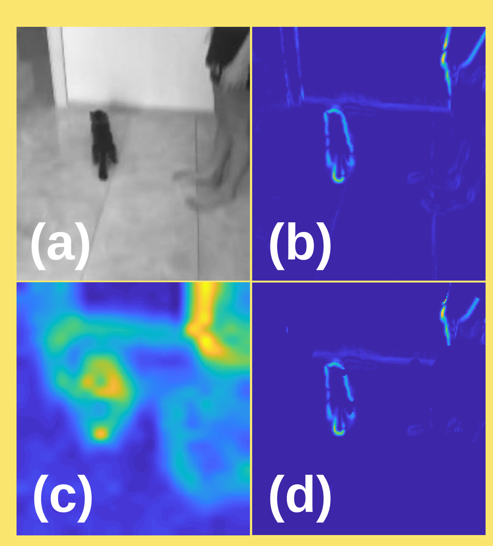


Time dilated video

Methods

1. Content Analysis

a. Original video frame as input
b. Calculate Optical Flow magnitudes
c. Saliency maps determine spatially interesting regions of frame.
d. Use Saliency map to mask extraneous flow vectors.



2. Energy Pooling

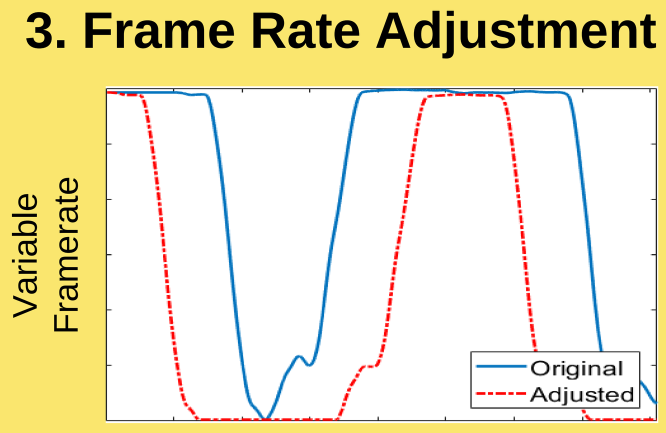
Energy Maps are converted into a 1D energy vector through pooling. We test the following three pooling functions:

$$\text{Minkowski} = \sum_{m=1}^{M}\sum_{n=1}^{N}\frac{Q[m,n]^p}{M \cdot N} \quad (1)$$

$$5_{num} = \frac{mean(Q) + Q_1 + med(Q) + Q_3 + \max(Q)}{5} \quad (2)$$

$$WP = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} w[m,n]\cdot Q[m,n]}{\sum_{m=1}^{M}\sum_{n=1}^{N} w[m,n]} \quad (3)$$

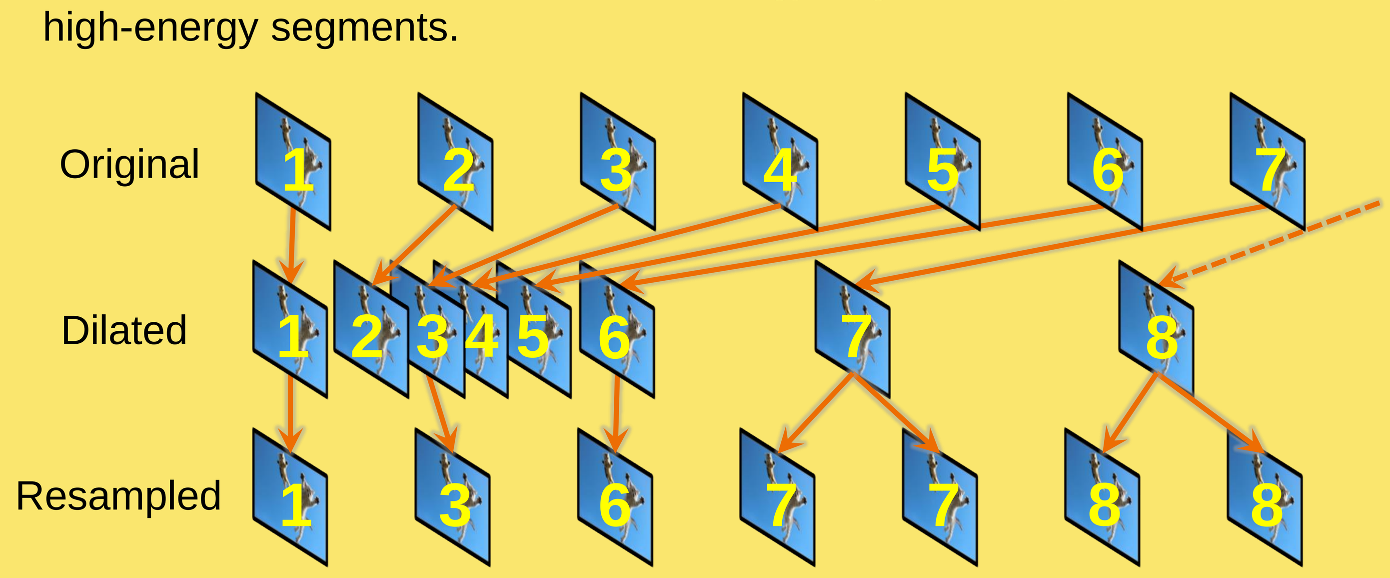
3. Frame Rate Adjustment



1. Median Filter to remove near-zero outliers
2. Smooth with moving average to eliminate jarring discontinuities
3. Time pad by preemptively slowing before high-energy frames, and delaying speedup afterward

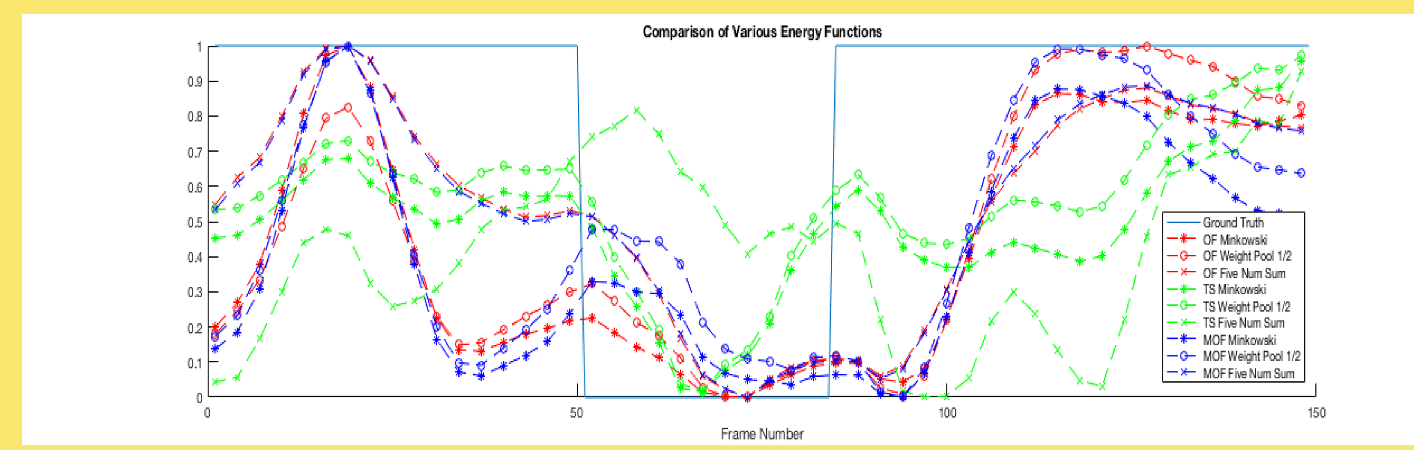
4. Frame Re-Sampling

After time dilation, choose frames closest to constant frame rate output instances. Throw out frames in low-energy segments, repeat frames in high-energy segments.



Experimental Results

Pooling Functions as Energy Vectors



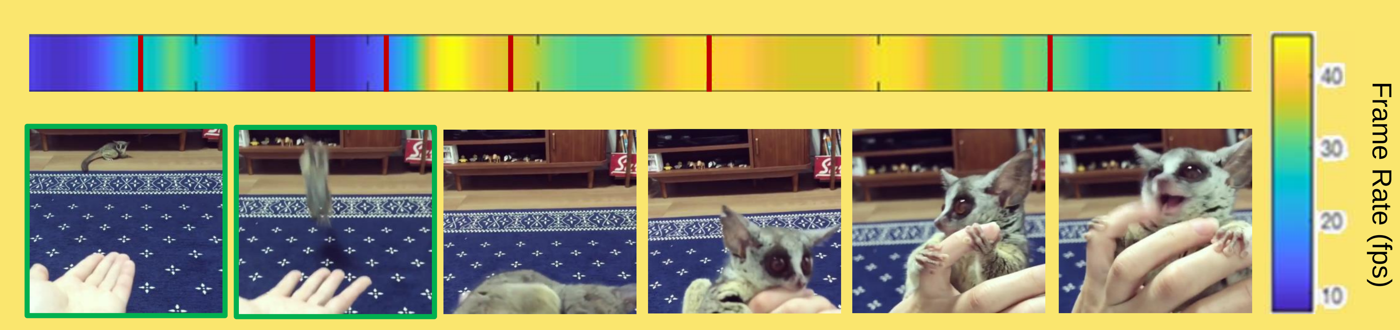
Pooling functions applied to Optical Flow Magnitudes (red), Time-Weighted Saliency Maps (green), and Saliency-Masked Optical Flow (blue). Graph generated on 'cat_wall_climb' video. Subjective ground truth labels frames with moving cat as 1 and empty frames as 0.

- Optical Flow and Masked Optical Flow energy functions correlate better with subjective ground truth.
- Weighted Pooling and Minkowski pooling functions perform better on average, producing the least jarring effects on final output.

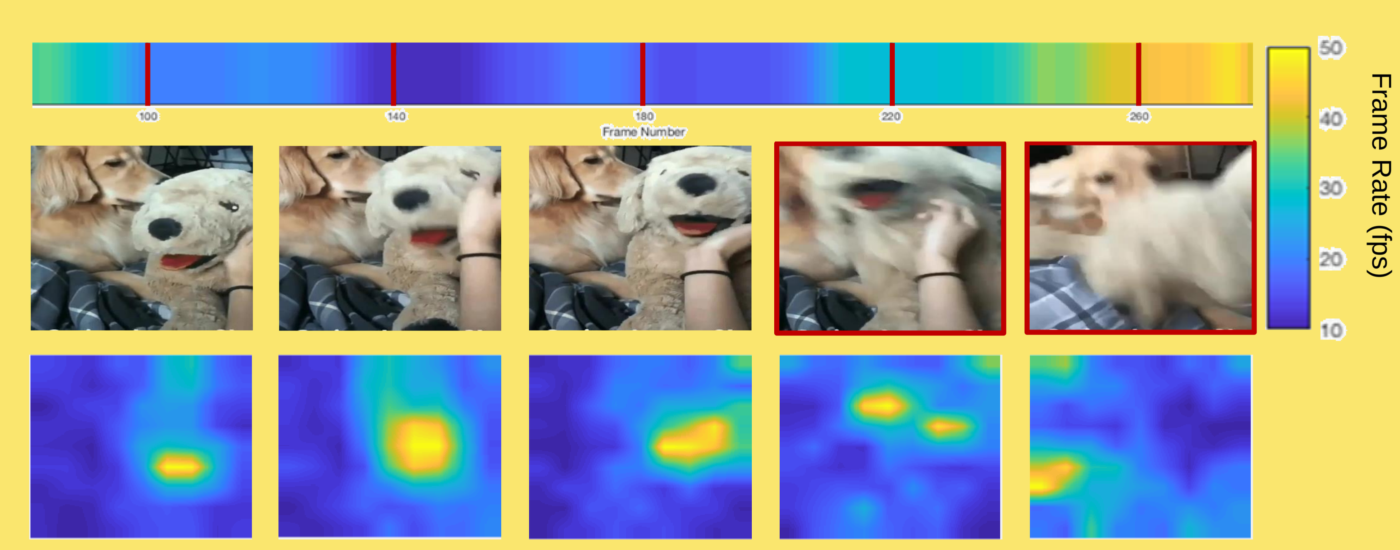
Results

Results are calculated using masked optical flow and weighted pooling with $p = {}^1/_2$. We represent each video with a heat map of frame rates. Frame number increases left to right with selected frames for interpretation.

'cat_wall_climb' - correctly detects and slows down the rapid movements of the cat.



'zaboomafoo' - correctly detects and slows down the jump in the beginning of video while speeding up the later frames that have little movement.



'dog_and_stuffedDog' - fails to detect and slow down the rapid movement of the dog biting the toy (frames in red). The algorithm is sensitive to the inconsistency of the saliency map and low contrast when calculating optical flow.

Contribution and Conclusion

We present a framework to solve the unexplored problem of automatic and adaptive time manipulation.

- Time dilation yields impressive results on videos with simple object movement and minimal frame jitter.
- Results vary with the energy function, but generally correlate with our subjective ground truth
- Results sensitive to poor optical flow/video saliency calculations due to motion jitter, low contrast, and fast or occluded movement, etc.
- This new framework sets a base for future work to provide a temporally optimal viewing experience.