Efficient Tiling For Video Analytics

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

Maureen Daum, Brandon Haynes, Amrita Mazumdar, Magda Balazinska, Alvin Cheung[1]

Paul G. Allen School of Computer Science & Engineering, University of Washington

[1]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

Motivation

Video storage and indexing for efficient query processing.

Query: Run license plate detection on all cars.

Decode car pixels → Run license plate detector

Decode the entire frame

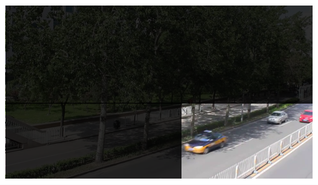
- Easy to store as encoded video
- Decode many irrelevant pixels

Decode only the car pixels

- Difficult to store as encoded video
- Decode only relevant pixels

Use tiling to decode only the region of the frame that contains car pixels

- Easy to store as encoded video
- Decode few irrelevant pixels



Strategy

- Split up video frames into independently decodable regions called “tiles”
- Set the tile layout using one of the following approaches:

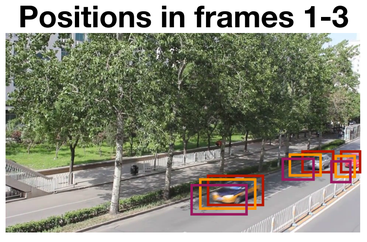
  Approach 1: Uniform tiles

  Approach 2: Non-uniform tiles around objects

  2.1: Large tiles around groups of objects

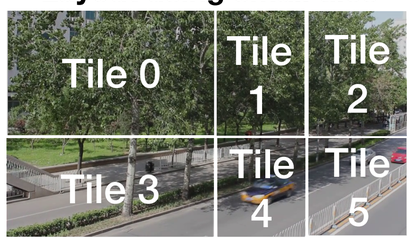
  2.2: Small tiles around individual objects
- Set the layout for a group of frames and update periodically
- Speed up queries by only decoding the tiles that contain pixels for a given query

Positions in frames 1-3

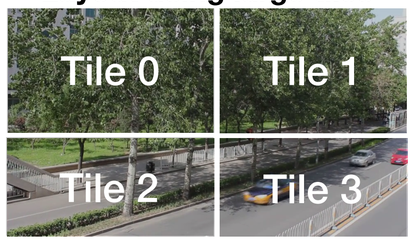


Layout using uniform tiles

Layout using small tiles



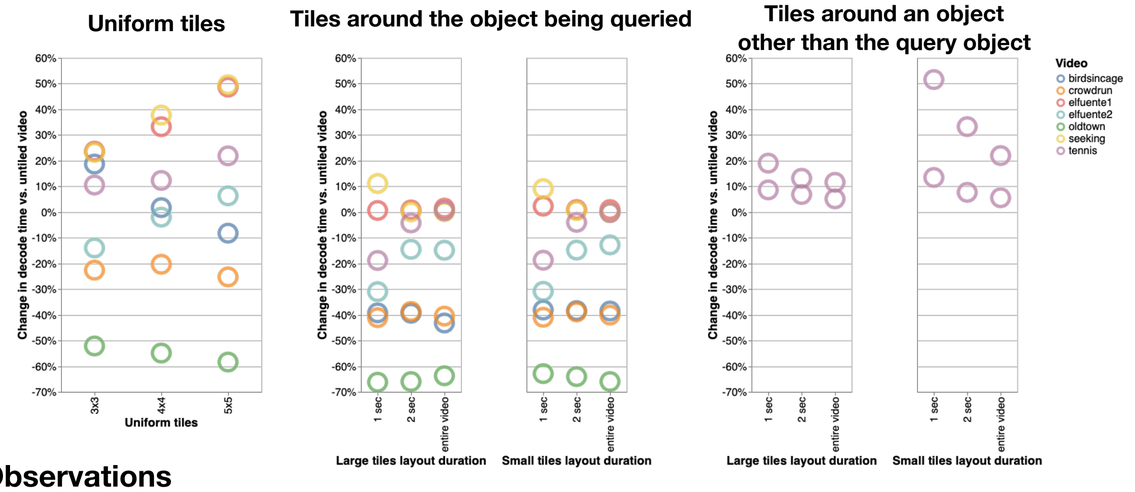
Layout using large tiles



Preliminary Results

- Run queries on videos from the Netflix public data set[2] to decode pixels for particular object types (e.g. “person”, “car”)
- Compare uniform tile layouts to layouts picked based on the locations of pixels being decoded
- Study the effect of updating the custom layouts after different durations

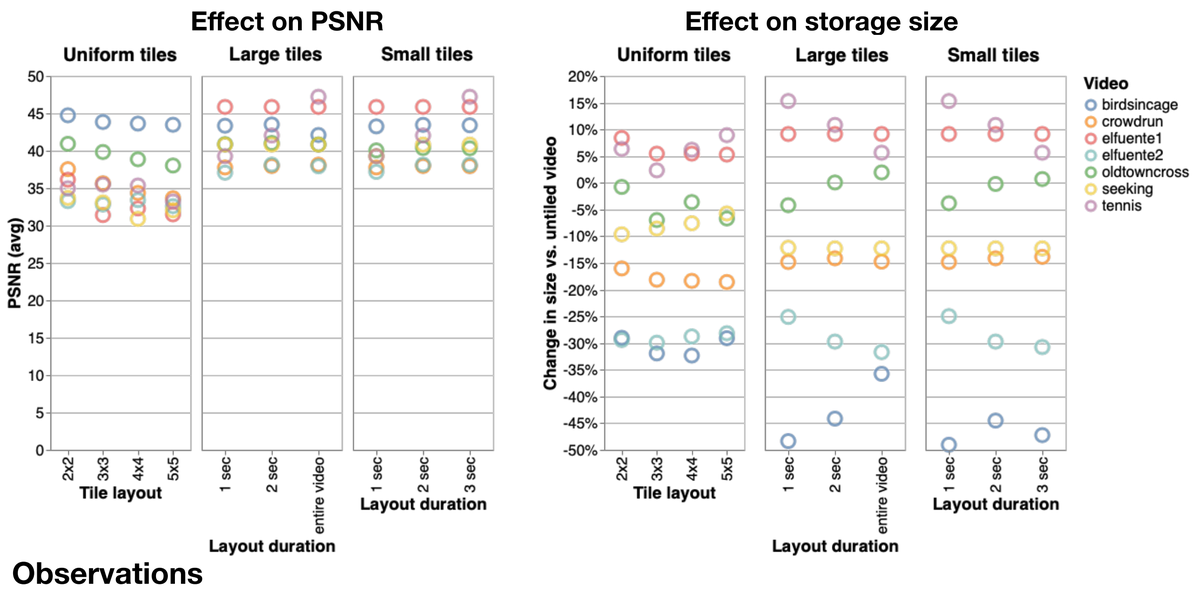
Effect of tiling on decode time



Observations

- Custom tile layouts reduce decoding time
- Tile layouts optimized for pixels different from the ones being queried can hurt performance

Effect of tiling on quality and storage size



Observations

- Custom tile layouts generally have better quality than uniform tiles (PSNR above 40 is considered lossless)
- Custom tile layouts sometimes lead to larger storage sizes. The size of the tiles depends on how they are encoded

Acknowledgements

This work is supported by the NSF through award CCF-1703051

[2]*https://github.com/Netflix/vmaf/blob/master/resource/doc/datasets.md*

Example video frame from UADetrac: http://detrac-db.rit.albany.edu