

Foundations for Summarizing and Learning Latent Structure in Video

Problem

The growing volume of streaming and archived surveillance video in the DoD is outpacing the ability of analysts to manually monitor and view it. There is a lack of automated tools available to assist analysts in monitoring real-time video or analyzing archived video.

Solution

Inspired by the past research of CMU Machine Learning professor Dr. Eric Xing, we investigated a new unsupervised video summarization pipeline that functions on extracted clips of objects in motion, rather than whole frames. The goal of the summary is to identify the key "object motion clips" occurring in the video.

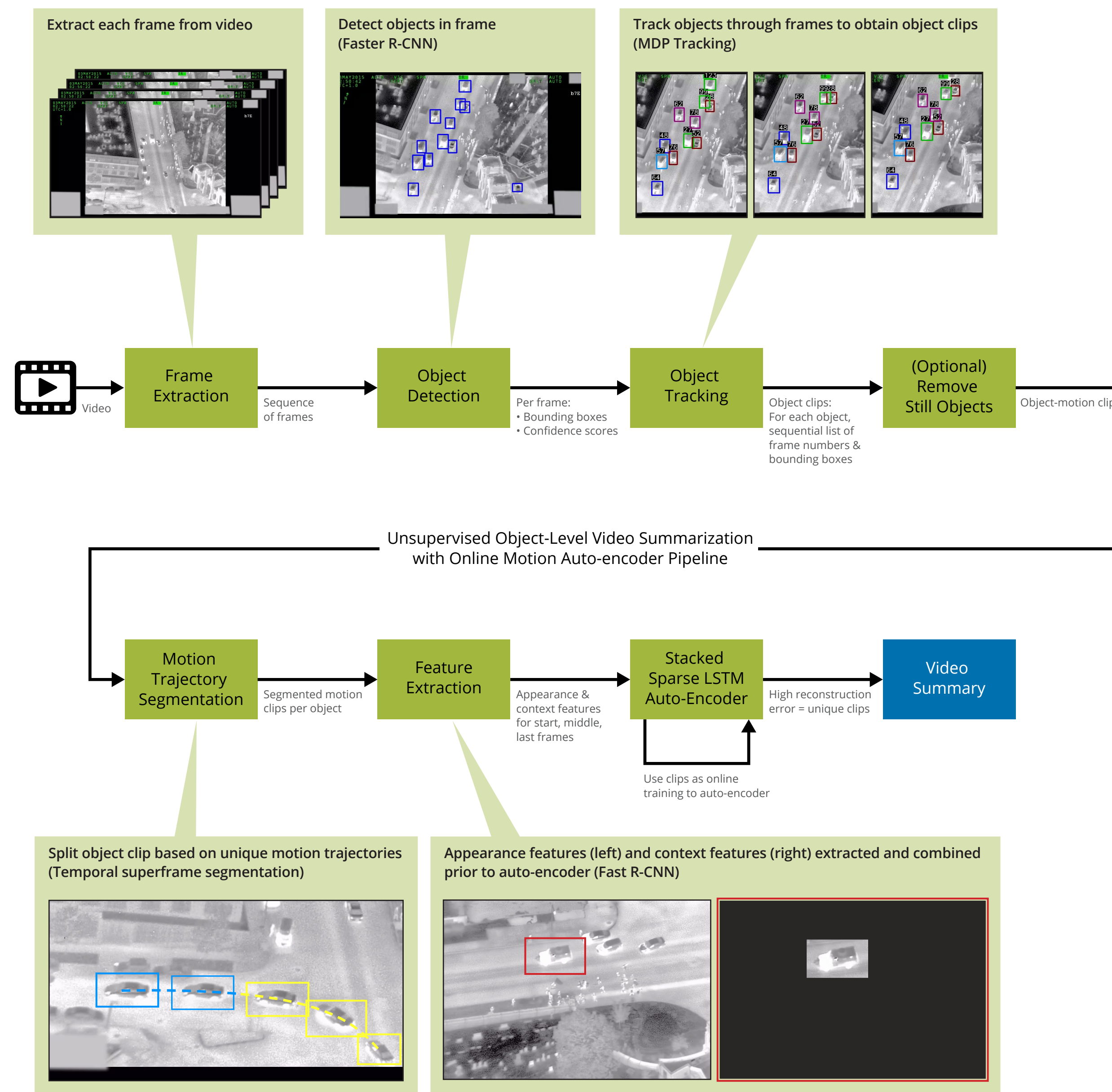
Approach

Unsupervised object-level video summarization with online motion auto-encoder

Offline Training: Initialize auto-encoder by training on random bounding boxes from representative video in order to learn a basic reconstruction capability

Online Processing & Training: For each video being processed:

1. Detect object bounds in each frame
2. Track objects through frames
3. (Optional) Remove clips of still objects
 - Useful for stationary camera
4. Segment object clips based on unique motion trajectories
5. Extract appearance and context features from the start, middle, and last frame of each clip
6. Feed the clip features through a three-layer, sparse, long short-term memory (LSTM) auto-encoder
 - Clips with high reconstruction error are collected as the "summary"
 - As clips are processed, use them as online training for the auto-encoder



	Sparse Coding	Stacked Sparse Auto-encoder	Stacked Sparse Auto-encoder	Stacked Sparse LSTM Auto-encoder (OURS)
AUC score	0.4252	0.4354	0.5680	0.5908
AP score	0.1542	0.1705	0.2638	0.2850
F-measure	0.1284	0.1662	0.2795	0.2901

Table 1: Object-level summarization results between competing approaches on Orangeville dataset

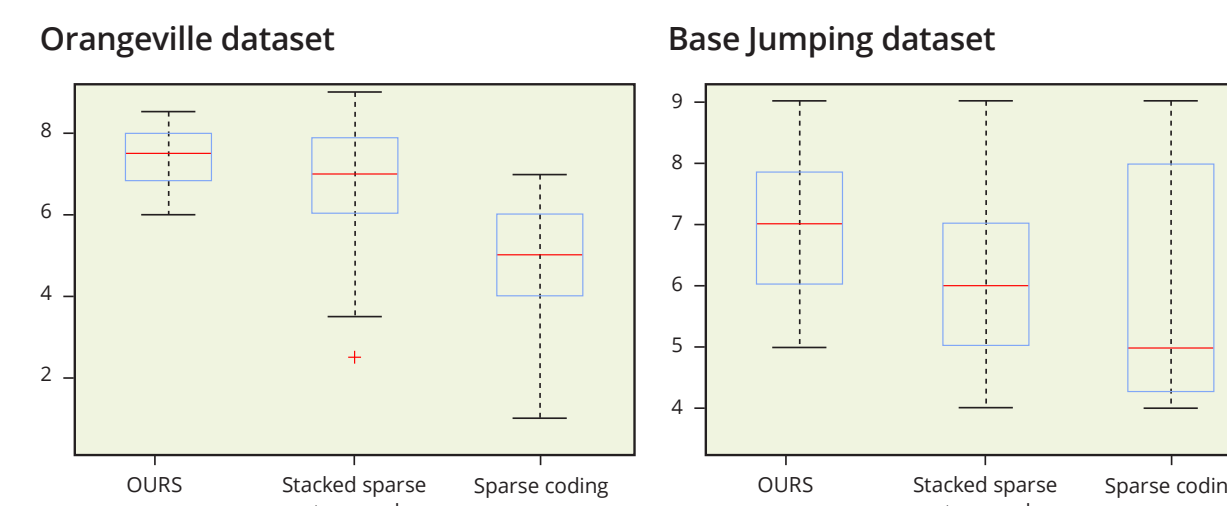


Figure 1: User study evaluation scores between competing approaches on Orangeville and Base Jumping datasets

Results

Quantitative: Our results matched or exceeded competing algorithms on benchmark datasets (see Table 1)

Qualitative: In a user study, our approach received higher and more consistent evaluation scores (see Figure 1)

Artifacts

Software

- Prototype utilizing the unsupervised, online, object-level video summarization pipeline
- Video Markup Tool for annotating spatial-temporal object clips within video

Paper

- Submission to IEEE Transactions on Cybernetics

Datasets

- "Orangeville" benchmark for object-level summarization - dataset & annotations
- Annotations & model for vehicle detection in FBI IR surveillance data

Future Work

FY18 Project: Summarizing and Searching Video

- Finish investigation of current pipeline to summarization of full-motion video (FMV) datasets
- Unsupervised activity clustering, utilizing object-motion clips as basis
- AFRL collaboration to explore applying analysis techniques to existing DoD problems
 - e.g., Nothing Significant to Report

In summary, we investigated object-level video summarization as an approach to identify the key segments in video with the final goal of applying these techniques to analyze real DoD surveillance data.

Copyright 2017 Carnegie Mellon University. All Rights Reserved.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

Internal use:* Permission to reproduce this material and to prepare derivative works from this material for internal use is granted, provided the copyright and "No Warranty" statements are included with all reproductions and derivative works.

External use:* This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other external and/or commercial use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

* These restrictions do not apply to U.S. government entities.

Carnegie Mellon® is registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM17-0738

Foundations for Summarizing and Learning Latent Structure in Video