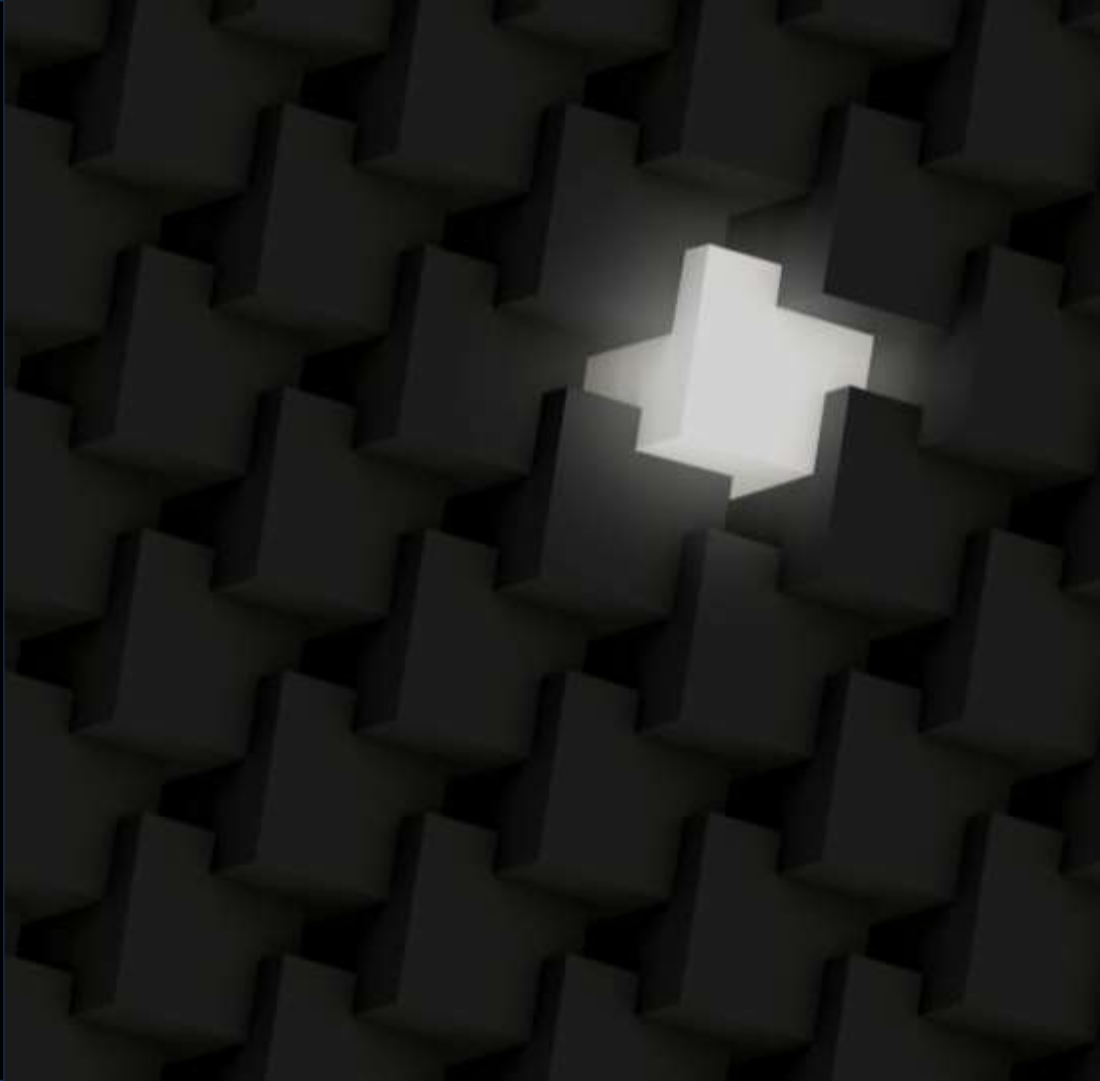
Carnegie Mellon University
Software Engineering Institute

RESEARCH REVIEW 2020

Video Summarization and Search (VidSum)

Ed Morris, Adam Harley, Kevin Pitstick, Rachel Brower-Sinning, Ben Cohen, Dan DeCapria, April Galyardt, Jeff Hansen, Ryan Meeuf, Jacob Ratzlaff



Document Markings

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.

DM20-0884

Agenda

Problem Overview

Transition Activities: Support to DoD

Research: 3D Tracking

RESEARCH REVIEW 2020

Video Summarization and Search (VidSum)

Problem Overview

VidSum Problem Statement

Problem: Aerial surveillance demands full attention to video by PED teams

- Manual, error-prone process
- Technical barriers including object detection, recognition and tracking
- Limitations result in poor pattern recognition in a surveilled region

Approach

- Improve DoD pattern recognition in aerial surveillance data by applying statistical analysis and machine learning technologies
- Work with CMU researchers to address core technology problems associated with object tracking

Achievements

- Influence on DoD pattern detection strategy
- “Reasoning” pathfinder for DoD
- 3D tracking state-of-the-art performance

Products

- Source code for data cleansing, statistical analysis, and ML-based pattern detection
- Source code to supplement training data
- Publications (2 accepted, 2 submitted)

Current Activities

- Transition: Support to DoD
- Research: 3D tracking

RESEARCH REVIEW 2020

Video Summarization and Search (VidSum)

Transition Activities: Support to DoD

Improving the Data: Data Cleansing and Smoothing

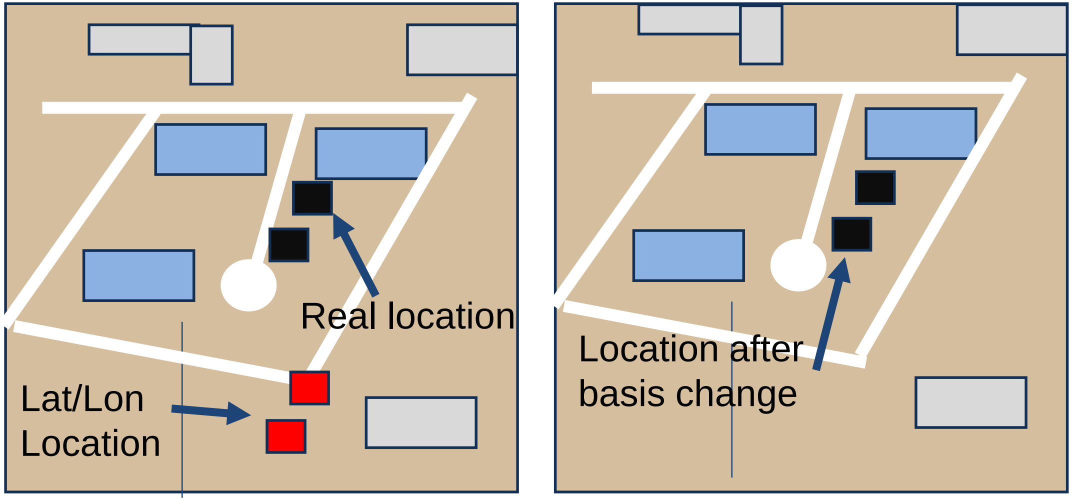
Problem: Data from aerial cameras is often "dirty"

- Imprecise lat/lon values due to onboard sensor inaccuracy and platform drift can lead to spurious/missing detections, bad tracking in downstream apps

Approach: Clean and smooth data prior to downstream processing

Implementation:

- Moving median smoothing
- Geo-registration corrections
  - Change of basis
  - Optical flow mismatch
- Kalman filtering



Example: Change of basis using 3 stationary objects

Pattern Analysis: Statistical Reasoning

Problem: Most activity is normal and harmless – some is not

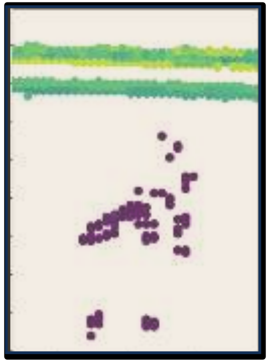
Approach: Use observations to build statistical PoL model

- map out “normal” (e.g., vehicles & people pathways, density)
- detect anomalous activities (specific to location and/or time)
- search for specific activities/interactions of interest

Implementation:

- Separate region into grid points based on camera attention
- Remove bad tracks
- Calculate grid point features (e.g., mean speed, heading, density)
- Detect anomalies by setting feature-based rules with thresholds

Surveillance tracks



Mean speed
(dark blue = slow, green = fast)

Pattern Analysis: Anomaly Detection

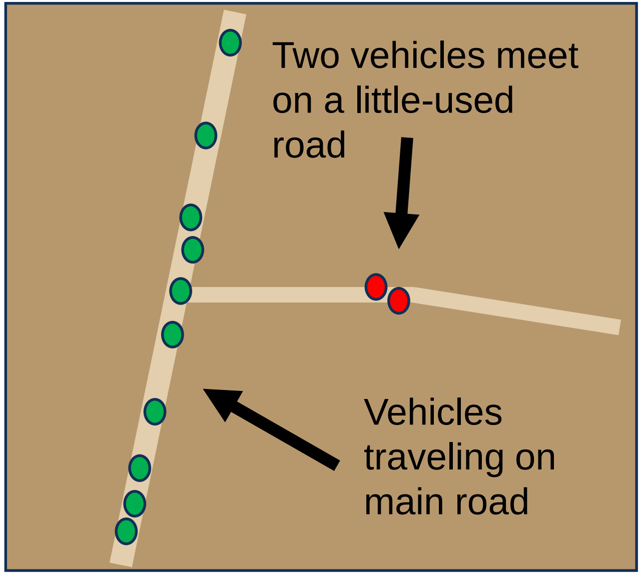
Problem: Most activity is normal and harmless – some is not

Approach: Use observations of a region to train an ML model to learn normal behavior in order to identify anomalous tracks and predict future tracks

Implementation:

- Train a long short-term memory (LSTM) autoencoder to reconstruct observed tracks
- Tracks with high reconstruction error are identified as anomalous tracks

Tracks not shown



Anomaly detector results:

- Perfect data (GPS)
- Reality not so pretty
- Importance depends on mission

Barrier to Progress: Poor Object Tracking

Problem 1: Best performing tracking algorithms correlate detections across 2D camera frames, but

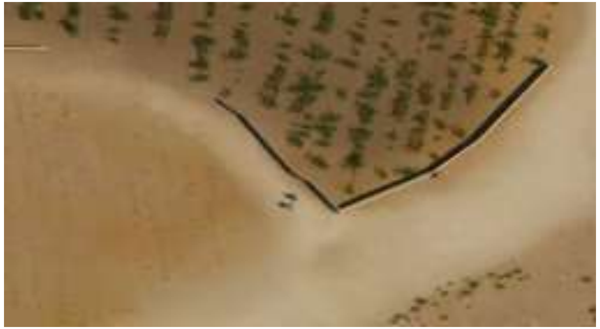
- Objects look different depending on viewpoint
- Occlusion throws trackers off
- Object coordinates within a frame are not a good predictor of where to look for the object in the future

Problem 2: Best-performing tracking algorithms require many images in order to train object detectors, but

- Often relatively few images for many things that matter to DoD

Resulting in:

- Poor identification of objects
- Lost tracks
- Poor pattern detection due to poor tracking



Strategy: 3D Tracking

- Collaboration with Adam Harley and Dr. Katerina Fragkiadaki (advisor)
- Adam has turned it into a focus of his PhD thesis

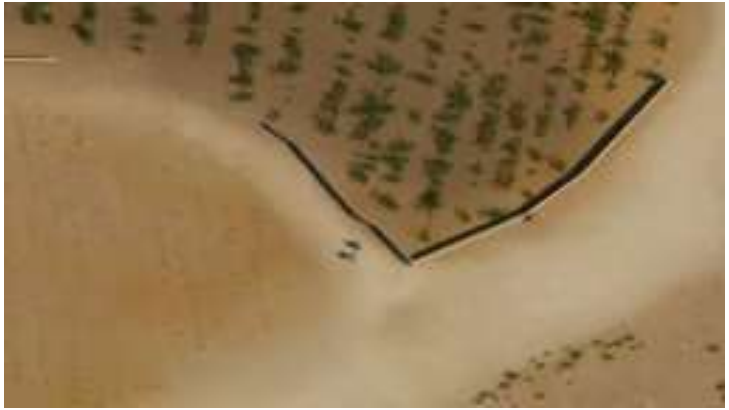
RESEARCH
REVIEW
2020

Video Summarization and Search (VidSum)

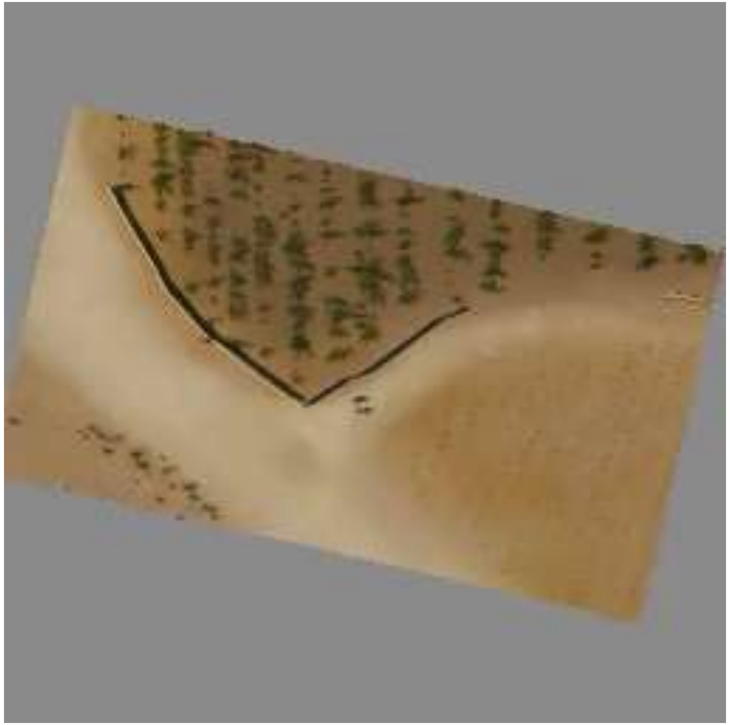
Research: 3D Tracking

Adam W. Harley, Yiming Zuo, Jing Wen,
Shrinidhi K. Lakshmikanth, Katerina Fragkiadaki

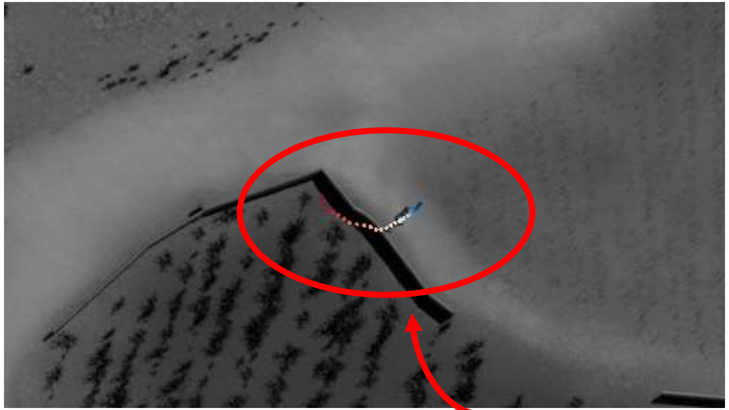
Detection and Tracking from Aerial Data

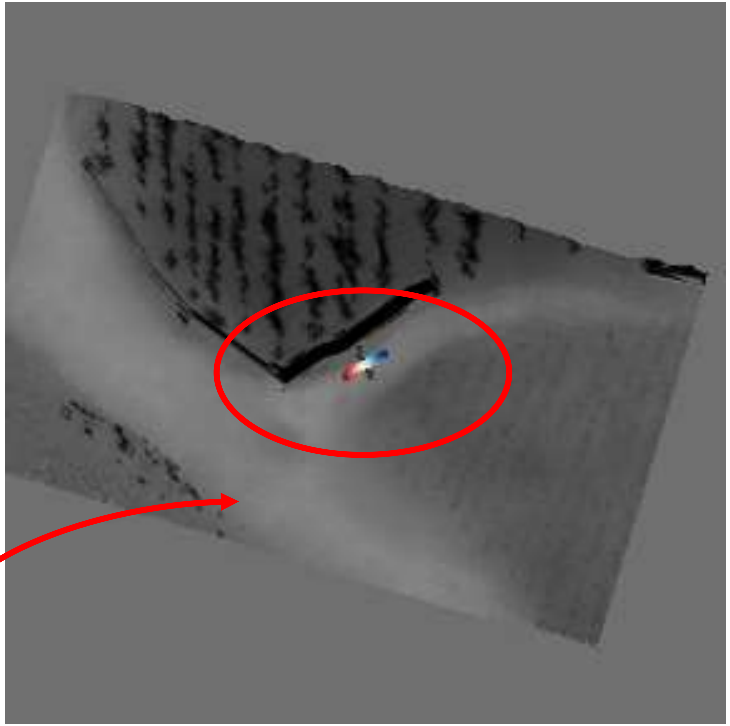


3D geometry can make things easier by stabilizing the observations



Detection and Tracking from Aerial Data





Trajectories that are complex in the raw video become simpler after stabilization

Academic Data

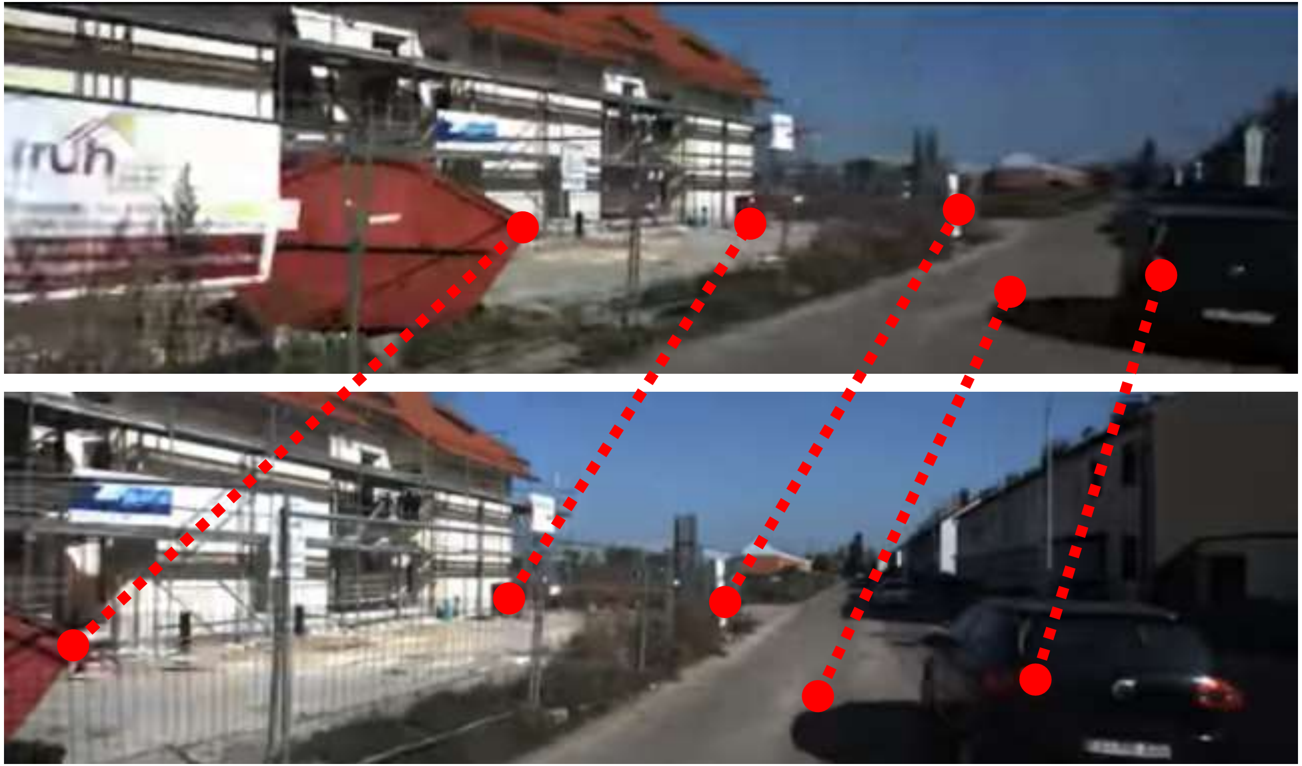
Existing academic data is not aerial, but we can explore the same techniques

RESEARCH REVIEW 2020

Video Summarization and Search (VidSum)

Research Part 1/3: Learning to track objects in 3D without labels

Corresponding Static Points

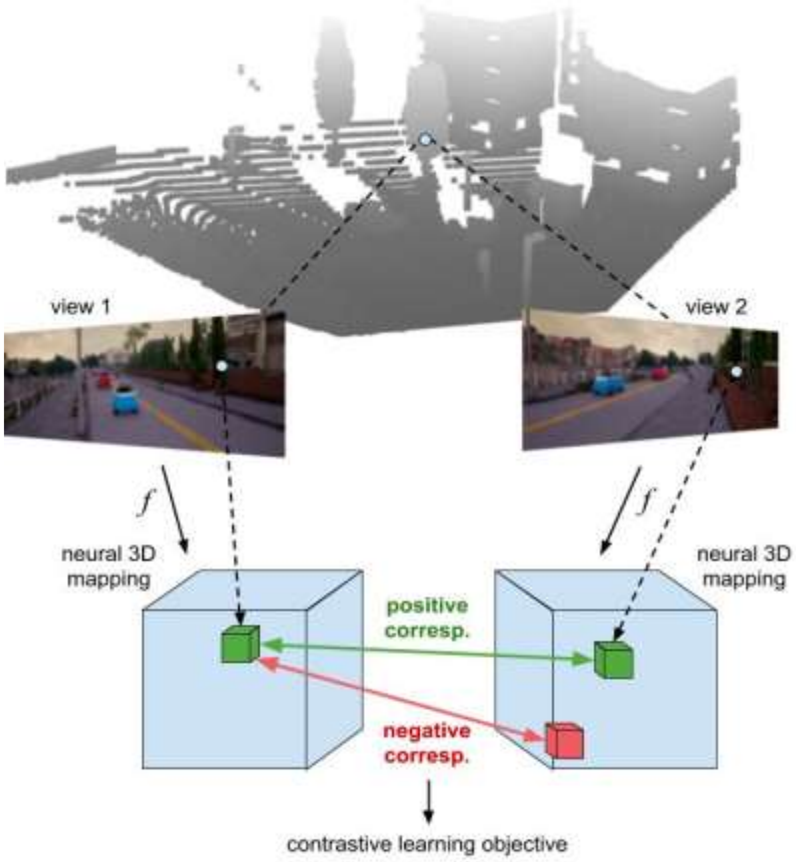


Using geometry we can correspond static points. If we train features to correspond these points visually, maybe we can use the same features to track moving points.

Training from Static Points

Given 2 viewpoints of the same object:

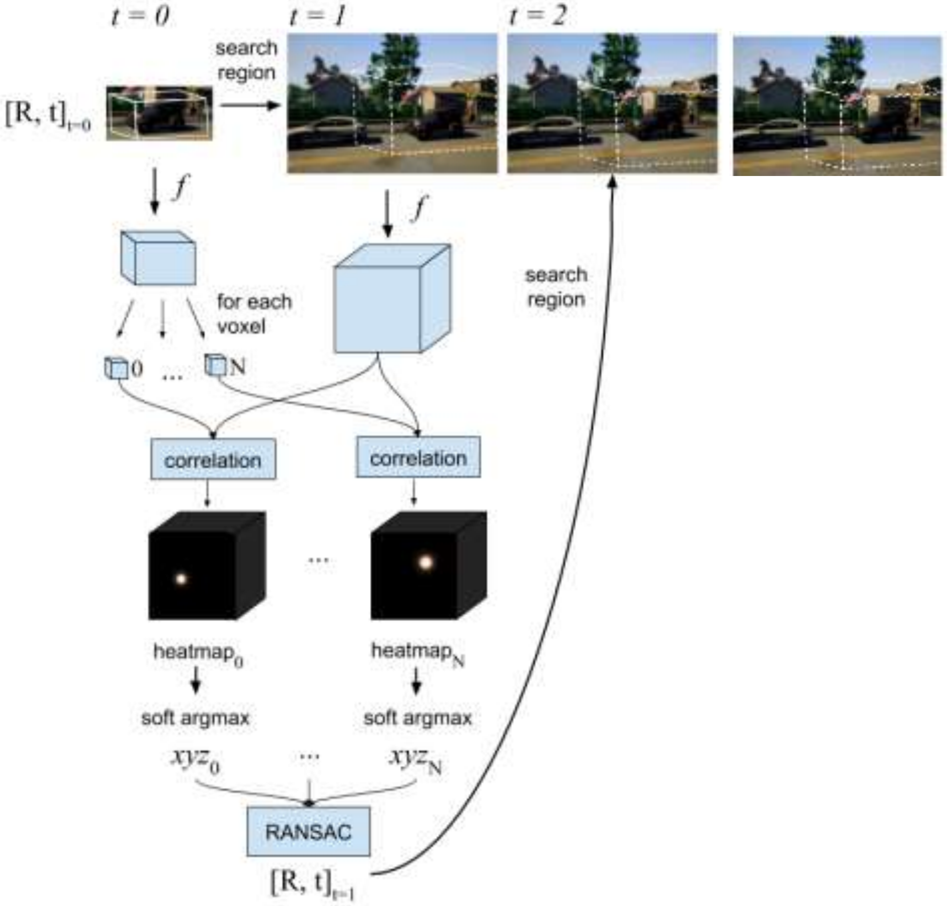
- Generate a neural 3D mapping for each
- Identify the corresponding voxel pair in the two mappings
- Treat all other mappings as negative correspondences
- Train the features to indicate the correspondences automatically



Tracking Moving Objects

Given the bounding box of the target object:

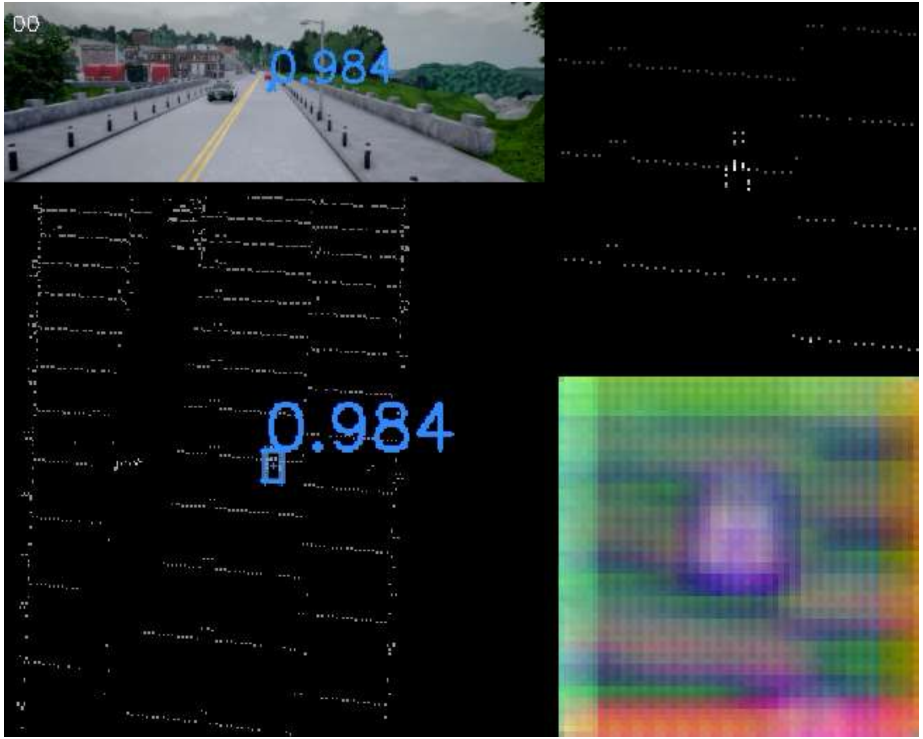
- Generate features for the object
- Generate features for the search region
- For each voxel of the object, compute its correlation with the search region
- Estimate the total motion with RANSAC
- Update the box



Tracking Moving Objects: Qualitative Results

Perspective view

Birds eye view

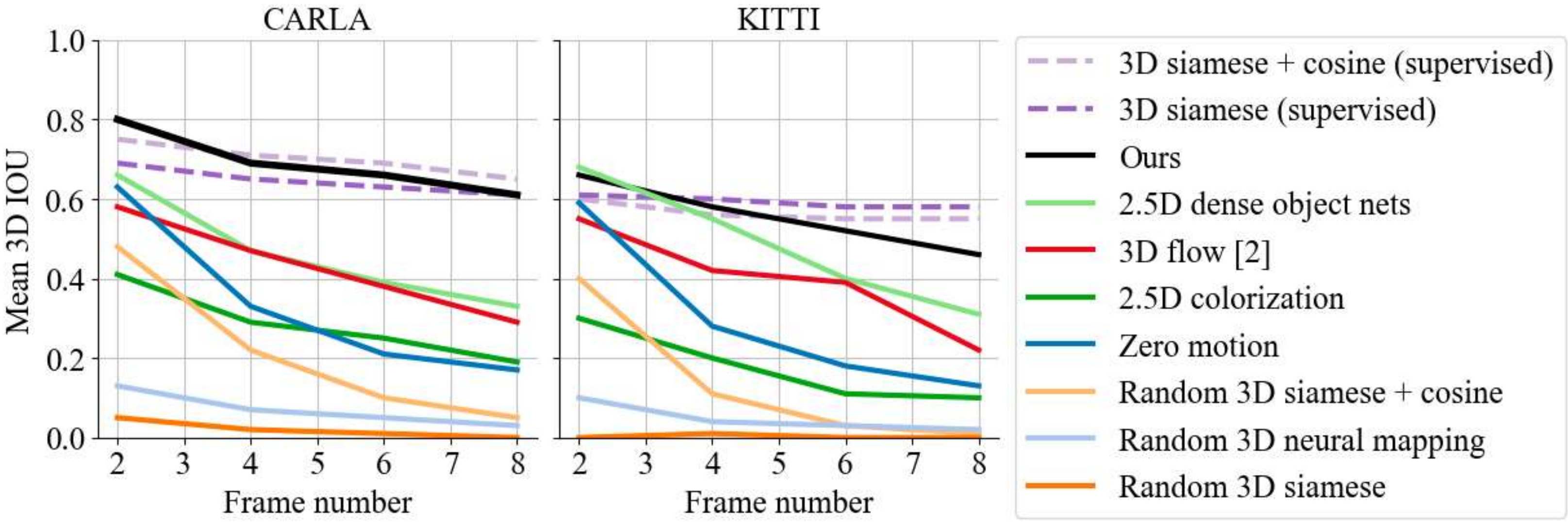


Search region

Search region features

- Tracking is mostly successful.
- Boxes "jump around" since this is frame-by-frame tracking (no motion prior).
- Works in simulation and in the real world.

Tracking Moving Objects: Quantitative Results



- Improves on unsupervised tracking algorithms
- Approaches supervised tracking algorithms

Tracking Moving Objects: Contributions

1. **We show that learning correspondence from static 3D points causes 3D object tracking to emerge.**
2. We introduce a neural 3D mapping module that simplifies prior works on 3D inverse graphics.
3. We introduce a method to train for correspondence in dynamic scenes — simply drop moving parts!

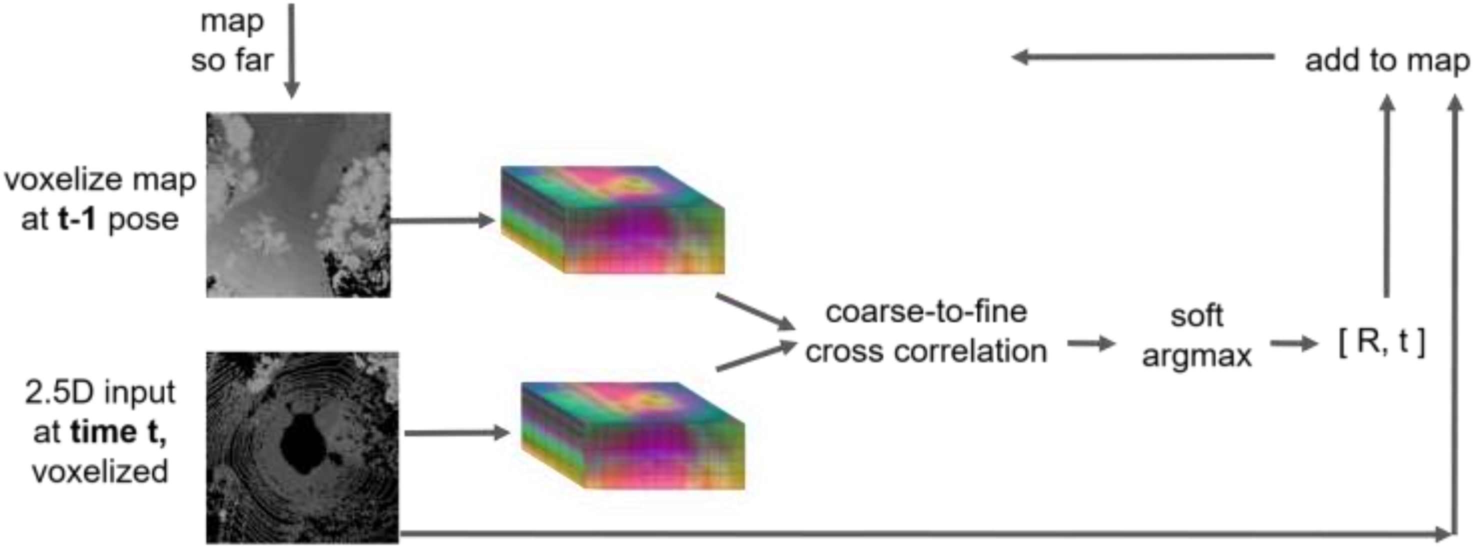
RESEARCH REVIEW 2020

Video Summarization and Search (VidSum)

Research Part 2/3: Estimating Camera Motion (Egomotion)

Estimating Camera Motion (Egomotion)

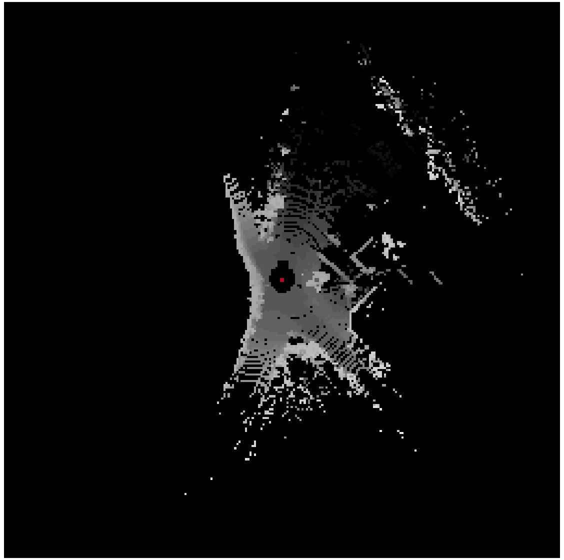
- Input: 2.5D (RGB+Depth) video
- Output: camera's rotation, translation at each timestep

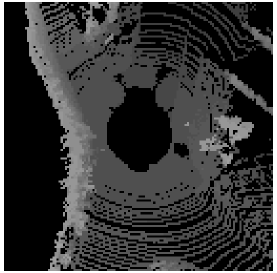


Egomotion: Qualitative Results

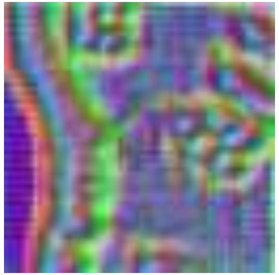
Input RGB

Model-aggregated map





3D occupancy input



3D feature output

- The model builds a "feature map" of the world while travelling through it.
- If the map gets corrupted, everything fails, so it is important to only make "good" updates to the map.

Egomotion: Quantitative Results

	Mean endpoint error (in meters) after 100 frames
Ours - no map, no coarse-to-fine	8.525
Ours - no map	4.914
Ours - full	1.627
Orbslam2-stereo	0.2993

KITTI Odometry Validation Set Results

Egomotion: Contributions

1. **We introduce a neural egomotion module that is capable of map-building.**

2. We are closing the gap between the “deep” and “traditional” methods, both in terms of method and accuracy. This paves the way for more general systems, that succeed in domains where the handcrafted features fail.

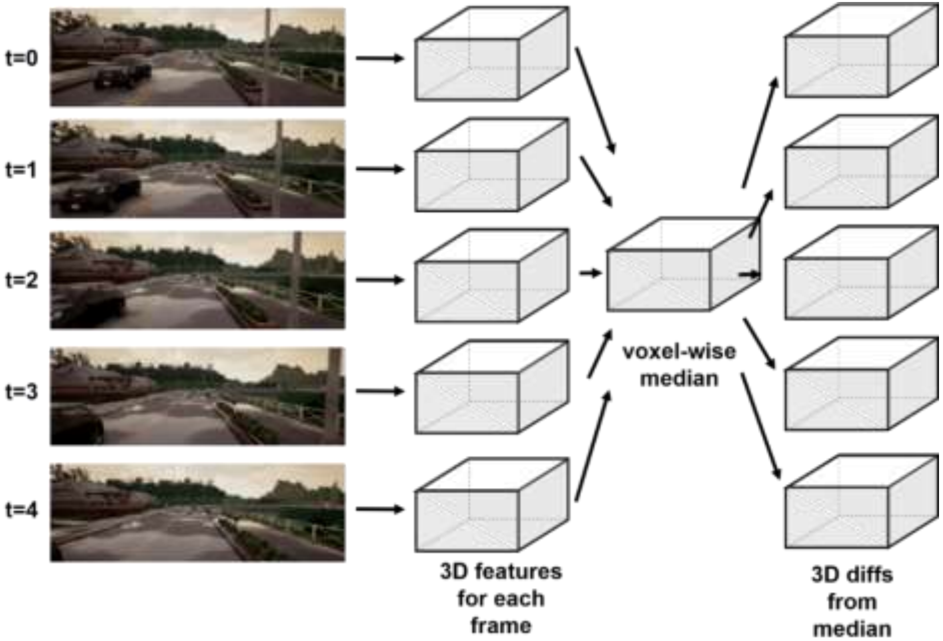
RESEARCH
REVIEW
2020

Video Summarization and Search (VidSum)

Research Part 2/3: Object Discovery

Object Discovery: Process

What happens when you do not have enough data to train good detectors, or require a process that does not need human intervention to track objects?



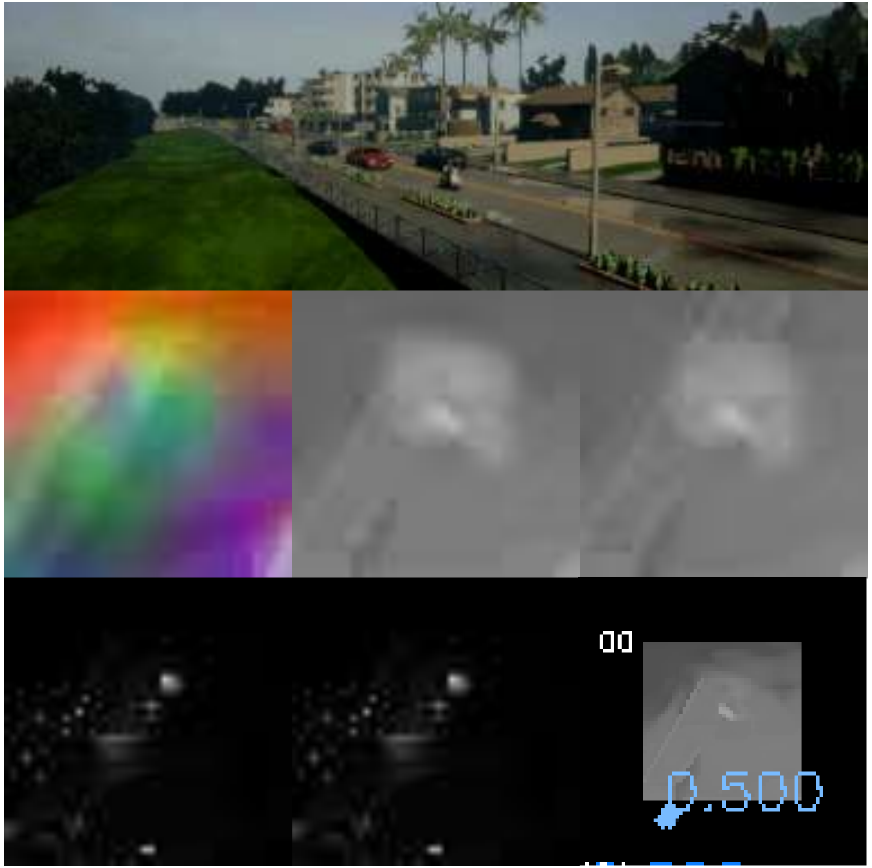
- Extract 3D features for each frame
- Determine voxel-wise median
- Determine the difference from the median for each frame

Object Discovery: Qualitative Results

RGB

Feat,
occ,
bkg

diff1,
diff2
tracklets



- The "median of the scene" is visibly empty - no cars or bikes. This is what makes the subtraction work.
- The largest differences from the median (big blobs) highlight moving objects.
- When an object is detected, we track it with our previous (unsupervised) method.

Object Discover: Contributions

We have shown that object discovery is relatively easy if

- we appropriately exploit the geometry of the scene
- we leverage long time horizons, where the “median” is a stable estimate of the background

Summary

Current Activities

- Transition : Support to the DoD
- Research: 3D Tracking

Next Steps:

- Continued work with DoD to improve pattern recognition from aerial surveillance data
- Continued research on 3D tracking by Adam Harley and the CMU team

More Information

Ed Morris
ejm@sei.cmu.edu

Grace Lewis
glewis@sei.cmu.edu

Adam Harley
aharley@cmu.edu

Dr. Katerina Fragkiadaki
katef@cs.cmu.edu