Dynamic Modality Switching Aided Object Tracking using an Adaptive Sensor

Matthew J. Hoffman, Ph.D^., mjhsma@rit.edu
Burak Uzkent, Ph.D. Student*, bxu2522@rit.edu
Anthony Vodacek, Ph.D.*, vodacek@cis.rit.edu

^School of Mathematical Sciences, Rochester Institute of Technology
*Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology

Funded by
the Dynamic Data Driven Applications Systems Program,
Air Force Office of Scientific Research
Problem statement

• In a close view surveillance, one can rely on features such as vehicle texture, shape, corners, edges, or color histograms. The drawback is lack of area coverage.

• Wide-area ground vehicle tracking often uses Doppler radar. *(Position and Radial Velocity)*

• Vehicle tracking systems with single modalities are very likely to fail in cluttered scenes.

• Wide-area surveillance problem is a challenging task due to several factors:
  1. Image resolution versus coverage
  2. Vehicle motions are severely constrained by the environment
  3. Frequent occlusions
  4. Many potential target vehicles
Approach

• Sensors that can observe different modalities (hyperspectral, polarization, thermal) can add useful information to the problem

• Unfortunately, they also result in a significant increase in data transfer which can make real-time tracking infeasible

• Adaptive sensors can change this by eliminating collection of unnecessary data

• Here we consider real-time tracking of a single, user selected vehicle

• Sensor control is enabled through Dynamic Data Driven Applications Systems principles
Recent Work

Journal Papers


- Integrating Hyperspectral Likelihoods in a Multi-dimensional Assignment Algorithm for Aerial Vehicle Tracking, In review.

Conference Papers

- Spectral Validation of Measurements in a Vehicle Tracking DDDAS, ICCS 2015

- Background image understanding and adaptive imaging for vehicle tracking, SPIE Defense Commercial Systems Conference, 2015

- Efficient integration of spectral features for vehicle tracking utilizing an adaptive sensor, SPIE Electronic Imaging, 2015.
The Rochester Institute of Technology Multi-object spectrometer (RITMOS) is considered for adaptive spectral data and panchromatic image acquisition.

Two micro-mirror arrays deflect the light towards the spectrograph. The others are kept still.

The limitations of the performance-driven RITMOS multi-modal sensor

- Panchromatic image of a scene is collected around 0.1 s.
- Spectral data is collected over first 3 ms. Longer integration times can be traded with number of pixels.
- Other sensors may have different constraints.
- Here we have the spectral sensor collect a 200x200 ROI within a larger panchromatic image.
- The control of the spectral sampling is part of the DDDAS control.

A panchromatic image consists of a single band and is usually displayed as a grey scale image.

* A panchromatic image consists of a single band and is usually displayed as a grey scale image.
Synthetic Image Generation

- DIRSIG allows us a knowable ground truth for use in performance evaluation.
- Imagery in a variety of modalities can be produced with DIRSIG
- Simulation of Urban Mobility (SUMO) platform is used to simulate a large number of vehicles.
- Noise is added to the DIRSIG image to simulate realistic video
- A moving platform is also simulated (new addition)
Vehicle Simulation

- 86 vehicles from a SUMO simulation are placed in a DIRSIG scene that contains three intersections.

- Traffic is added in intersections to increase difficulty/realism.

- Different vehicles types (car, truck, van, etc.) are assigned.

- One of 24 paint models is randomly assigned to the vehicles.

- The simulation is 130 seconds long.
Motion Segmentation

- Previous work using a stationary platform have used a Median filter method to model the background over some number of frames.

- This method is less effective with a moving platform due to noise from registration.

- Instead, spectral information is used to reduce the search space and then find target matches.
Motion Segmentation

- Vegetation is identified using NDVI
- It is then excluded from the target search space
• A nonlinear SVM with radial basis function kernel is used as a classifier
• 2700 samples are used for training
• 87.4% accuracy on 2000 test samples
Motion Segmentation

- When the classification is complete, the image is segmented into road, vegetation, and other
- Other includes vehicles and also buildings, paths, etc.
Feature Matching

- Local spectral histograms are used in a sliding window technique to compute spectral similarity.
Feature Matching

- Multi level Otsu method provides threshold
- Morphological closing yields candidate blobs
Feature Matching

- Buildings can still match vehicle pixels spectrally
- To mask out buildings we introduce another classifier
- This uses a linear SVM cascaded with Histogram of Oriented Gradients (HoG) features
- The SVM is trained using 2500 vehicle and non-vehicle samples at five different times to get different illuminations
- Test samples are classified with 92.75% accuracy
Imaging and Detection Summary

- After training, these all run fairly efficiently

<table>
<thead>
<tr>
<th>Module</th>
<th>Vegetation Detection</th>
<th>Road Detection</th>
<th>Spectral Histograms</th>
<th>HoG-SVM Detector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run time</td>
<td>0.002 s.</td>
<td>008-0.1 s.</td>
<td>0.13-0.15 s.</td>
<td>0.05 s.</td>
</tr>
</tbody>
</table>

Runtimes with an 2.9GHz i7 processor

- Further speedup could be gained with parallel processing (even GPUs)
Data Association and Filtering

- Once the target candidate blobs are identified, assignment is performed using the multi-dimensional assignment (MDA) algorithm.

- Both kinematic and spectral likelihoods are used in the MDA algorithm.

- 5 time levels are used.

- Filtering is then performed using a Gaussian Sum Filter.

- 17 components are used here.

- Each of these Gaussian components can be assigned a different prediction model as part of a multiple model framework (stop, turn, constant velocity).
Experiments

Tracking Parameters
We present results from 13 vehicles in the scene using the following parameters:

- Number of Gaussians: 17
- Dimensionality in S-D: 5
- Number of Monte Carlo Runs: 100
- Number of Frames to Terminate a Track: 7

Performance Metrics

Recall = \frac{\#TP}{\#TP + \#FN}

Precision = \frac{\#TP}{\#TP + \#FP}

Track Purity(T_rP) = \frac{\text{# of times true measurement is assigned}}{\text{duration of track}}

Target Purity(T_gP) = \frac{\text{# of times true measurement is assigned}}{\text{duration of ground truth}}

Cluster Density Reduction(CDR) = 1 - \frac{\#TP \text{ in ROI}}{\#\text{Vehicles in ROI}}

TP= True Positives
FP= False Positives
FN= False Negatives
## Results

<table>
<thead>
<tr>
<th>ID</th>
<th>Recall</th>
<th>Precision</th>
<th>TrP</th>
<th>TgP</th>
<th>CDR (%)</th>
<th>Frames</th>
<th>Paint Model</th>
<th>Tree Occlusions per Frame (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car 34</td>
<td>90.11</td>
<td>96.07</td>
<td>90.25</td>
<td>89.79</td>
<td>66.46</td>
<td>22-127</td>
<td>Dirty White</td>
<td>15.53</td>
</tr>
<tr>
<td>Car 35</td>
<td>86.47</td>
<td>95.40</td>
<td>31.39</td>
<td>17.00</td>
<td>53.41</td>
<td>23-156</td>
<td>Clean White</td>
<td>23.08</td>
</tr>
<tr>
<td>Car 36</td>
<td>69.86</td>
<td>84.74</td>
<td>25.76</td>
<td>12.44</td>
<td>55.82</td>
<td>33-156</td>
<td>Dark Blue</td>
<td>25.00</td>
</tr>
<tr>
<td>Car 37</td>
<td>66.76</td>
<td>95.84</td>
<td>48.59</td>
<td>31.15</td>
<td>60.95</td>
<td>38-126</td>
<td>Light Blue</td>
<td>25.88</td>
</tr>
<tr>
<td>Car 38</td>
<td>98.31</td>
<td>94.28</td>
<td>98.31</td>
<td>98.31</td>
<td>73.68</td>
<td>61-138</td>
<td>Light Red</td>
<td>14.86</td>
</tr>
<tr>
<td>Car 39</td>
<td>98.14</td>
<td>97.03</td>
<td>98.24</td>
<td>98.24</td>
<td>66.13</td>
<td>40-156</td>
<td>Light Red</td>
<td>6.19</td>
</tr>
<tr>
<td>Car 40</td>
<td>83.62</td>
<td>84.17</td>
<td>85.83</td>
<td>85.83</td>
<td>74.27</td>
<td>71-156</td>
<td>Dark Blue</td>
<td>30.49</td>
</tr>
<tr>
<td>Car 41</td>
<td>99.98</td>
<td>98.82</td>
<td>82.38</td>
<td>80.72</td>
<td>60.61</td>
<td>26-130</td>
<td>Green</td>
<td>10.89</td>
</tr>
<tr>
<td>Car 42</td>
<td>96.21</td>
<td>94.58</td>
<td>89.00</td>
<td>89.00</td>
<td>29.93</td>
<td>27-141</td>
<td>Dirty White</td>
<td>6.31</td>
</tr>
<tr>
<td>Car 43</td>
<td>98.82</td>
<td>98.16</td>
<td>29.40</td>
<td>29.40</td>
<td>46.63</td>
<td>38-156</td>
<td>Clean White</td>
<td>2.61</td>
</tr>
<tr>
<td>Car 44</td>
<td>72.20</td>
<td>87.15</td>
<td>52.82</td>
<td>41.21</td>
<td>75.28</td>
<td>42-154</td>
<td>Dark Red</td>
<td>17.76</td>
</tr>
<tr>
<td>Car 45</td>
<td>83.35</td>
<td>95.79</td>
<td>81.43</td>
<td>81.43</td>
<td>78.27</td>
<td>40-156</td>
<td>Dark Blue</td>
<td>23.01</td>
</tr>
<tr>
<td>Car 46</td>
<td>95.26</td>
<td>97.96</td>
<td>93.33</td>
<td>93.33</td>
<td>72.08</td>
<td>51-123</td>
<td>Dirty White</td>
<td>11.59</td>
</tr>
<tr>
<td>Average</td>
<td>87.62</td>
<td>93.85</td>
<td>69.75</td>
<td>65.22</td>
<td>62.11</td>
<td>39-144</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

- For most of these vehicles, the track closely mirrors the true vehicle trajectory, even through obscurations and traffic lights.
Challenges

• If a vehicle becomes obscured and then slows down the predictions will overshoot, potentially causing target loss

• If other spectrally similar vehicles are nearby, then we can assign the wrong track

• The background modeling could be used to identify obscurations in advance and change strategy
Current/Future Work

- Up to this point, we have been using simulated data.

- We are working test the system on real data and we have three systems we are investigating
  - WASP-Lite sensor at RIT (multi-spectral)
  - Visual band light field spectrometer (~400-700nm)
  - High-speed line scanning VNIR HIS on a gimbal (DURIP project; ~400-1000nm)

- Once data sets are developed they can be shared with the community
WASP Lite

- WASP Lite has 7 spectral bands and will be mounted on a tall building to capture vehicle movement
- WASP Lite has thermal and can also capture polarization
- Capable of 1-2 frames per second
WASP Lite

- WASP Lite has 7 spectral bands and will be mounted on a tall building to capture vehicle movement.
- WASP Lite has thermal and can also capture polarization.
- Capable of 1-2 frames per second.
Light Field Spectrometer

- System from MITRE uses micro-lens and filter arrays to capture spectral video
HSI Low-Rate Video Rate Sensor

- Received DURIP funding to purchase a system capable of hyperspectral video
- Sensor will be used to create data sets for tracking
- Sensor will also be used for a Navy project on realtime classification of littoral zones
- We are in the process of purchasing
HSI Low-Rate Video Rate Sensor (DURIP)

- Generate low-rate video HSI (~3 Hz) via nodding in the along track direction:
  - e.g. nod up and down ~100 lines (record during both upward and downward motions and unpack and reorient image blocks later)
  - Produce cube (100x1846 spatial pixels)x(spectral bands) every ~0.3 sec
  - In some applications: steer in azimuth/zenith to a position, record via nodding, then steer to new position and repeat
Tony Requirements:
Range = 100m to 2,000m
Pixels on Target = 10 min.

Vehicle Length = 3m

Camera: Headwall MicroHE
Focal Plane: 1846 (spatial) x 369 (spectral bands)
Detector Readout Architecture: rolling shutter or global shutter
Pixel Pitch: 6.5 microns
Spectral Bandwidth: 400nm – 1000nm
Frame Rate = 3Hz
Focal Length = 12mm
SNR (by band regions): TBD bandwidth regions
Edge of Field Fall-off (cos⁴ effect) = 0.56 edge of field

<table>
<thead>
<tr>
<th>Range</th>
<th>Focal Length</th>
<th>GSD</th>
<th>Pixels on Target</th>
<th>Imaging Window Pixels (h x w)</th>
<th>Vehicle Motion¹ (pixel displacement per frame)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100m</td>
<td>12mm</td>
<td>0.054m</td>
<td>55</td>
<td>5.4m x 100m</td>
<td>75 pixels</td>
</tr>
<tr>
<td>200m</td>
<td>12mm</td>
<td>0.108m</td>
<td>27</td>
<td>11m x 200m</td>
<td>38 pixels</td>
</tr>
<tr>
<td>350m</td>
<td>12mm</td>
<td>0.189m</td>
<td>15</td>
<td>19m x 350m</td>
<td>21 pixels</td>
</tr>
<tr>
<td>500m</td>
<td>12mm</td>
<td>0.271m</td>
<td>11</td>
<td>27m x 500m</td>
<td>4 pixels</td>
</tr>
</tbody>
</table>

¹ Vehicle speed = 30mph = 44 ft/sec = 12.6m/sec, 3Hz = 4.1m vehicle displacement