

Activity Recognition in Wide Aerial Video Surveillance Using Entity Relationship Models



Jongmoo Choi, Yann Dumortier, Jan Prokaj, and Gérard Medioni

Institute for Robotics and Intelligent Systems

{jongmooc, prokaj, medioni}@usc.edu, yann.dumortier@gmail.com

iCampus/MediaQ iWatch ✓

CT

Goal & Challenges

❖ Goal

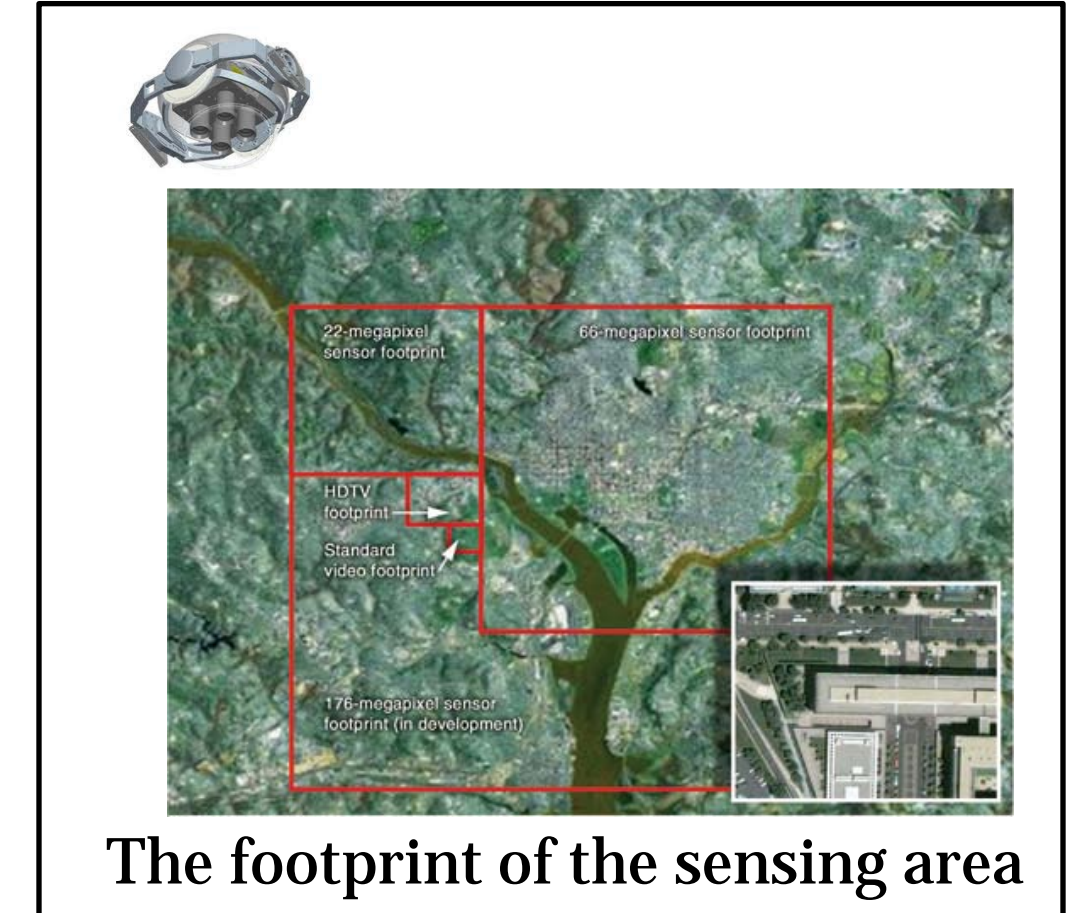
- Efficient activity recognition framework for wide area aerial surveillance (WAAS)
- Given partial, noisy vehicular tracks and possibly geo-registered imagery

❖ Challenges

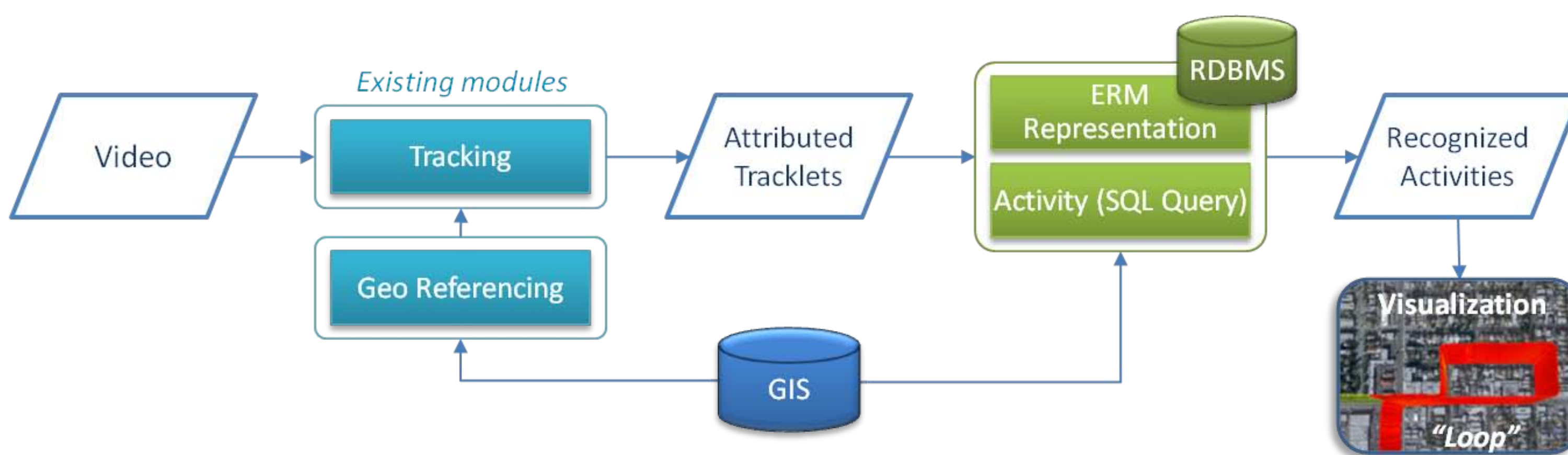
- Complex activities (multiple agents, hierarchical activities, sequence of events, composition of activities, ...)
- Scalability and efficiency

❖ Contributions

- Language is entity relationship model (ERM)
- Recognition ↔ RDBMS query
- Validation with real video and with GPS tracks



System Overview

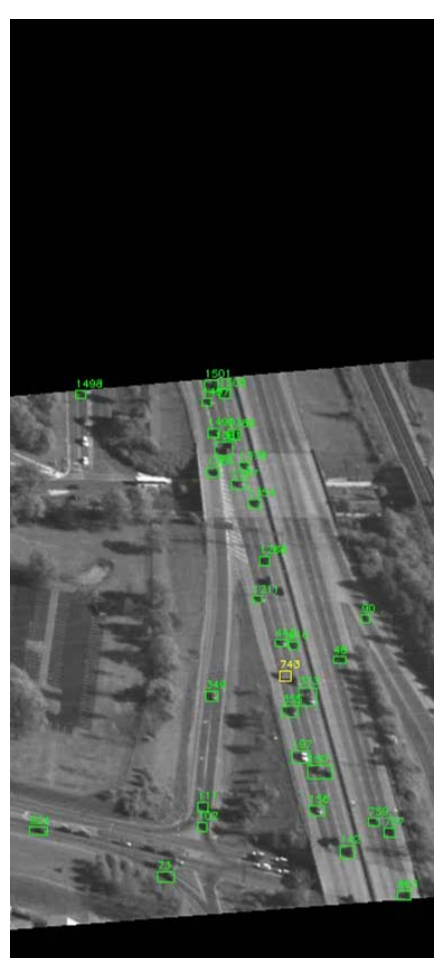


❖ ERM based activity recognition system

- Wide area aerial video sequence or GPS data
- Tracklets from Multi-Target Tracking
- Attributes for each tracklet
- ERM representation
- SQL query as activity recognition
- Visualization using standard tools
- Geo-spatial activity inference when GIS information supplied

Implementation Details

Tracklets from Imagery



Tracks from Multi-Target Tracking

- **Tracks:** a set of consecutive positions occupied by a vehicle, as inferred by a tracking module
- **Tracklets:** a linearly segmented atomic portion of a track.
- **Attributes for each tracklet:** time, location, turn, speed, heading, acceleration, and accumulated distance



Extracted tracklets from CLIF 2006

Simple Activity ("Loop")

Definition

$$Loop = \{x_i, x_j | (1 - \frac{\|x_i - x_j\|}{x_j.acc - x_i.acc}) > \theta, i < j, x_i.ID = x_j.ID\}$$

SQL

```
SELECT * FROM T1, T2
WHERE
T1.track_id = T2.track_id AND
(1 - (dist(T1, T2)/(T2.acc - T1.acc))) > \theta
```

❖ ERM representation

- Entity Relationship Models used in business
- Entity: tracklet, track, traffic rule, road segment, building, area, ...
- Relationship: tracklet -is on- road segment, road segment -has- traffic rule, ...
- Event: a relationship (e.g., "speeding": tracklet.speed > road.speed_limit)

❖ Activity recognition

- A collection of tracklets obeying certain properties
- RDBMS query

Experiments & Analysis

Data and Setup

Aerial video dataset

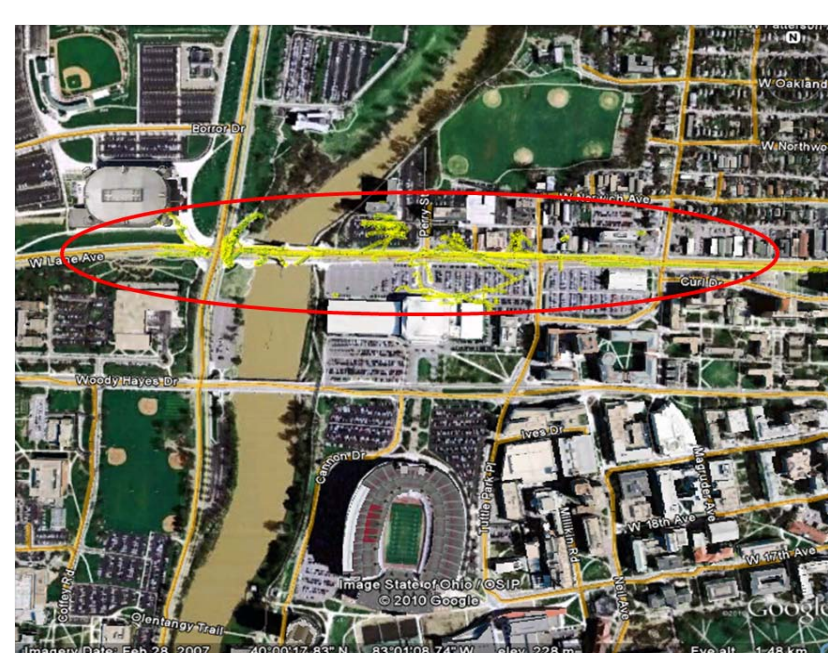
- Captured from an airborne sensor (CLIF 2006)
- A matrix of 6 cameras (image: 4008 x 2672)
- The footprint of the area is about 1 km²
- Activities: "Loop", "Three point turn", "Two point turn", "U-turn", "Entry", "Exit"

GPS tracks

- 10~40 minutes long
- Localization error is similar to video geo-registration
- Activities: "Loop", "Three point turn", "Stay", "U-turn"

Results and Conclusion

- Video dataset: precision (0.76), recall (0.86)
- GPS Tracks: precision (0.97), recall (0.90)
- Simple activity identified such as "U-turn"
- Multi-actor activity identified such as "source" and "sink"
- Geo-spatial activity identified such as "IsOnRoad(X)"



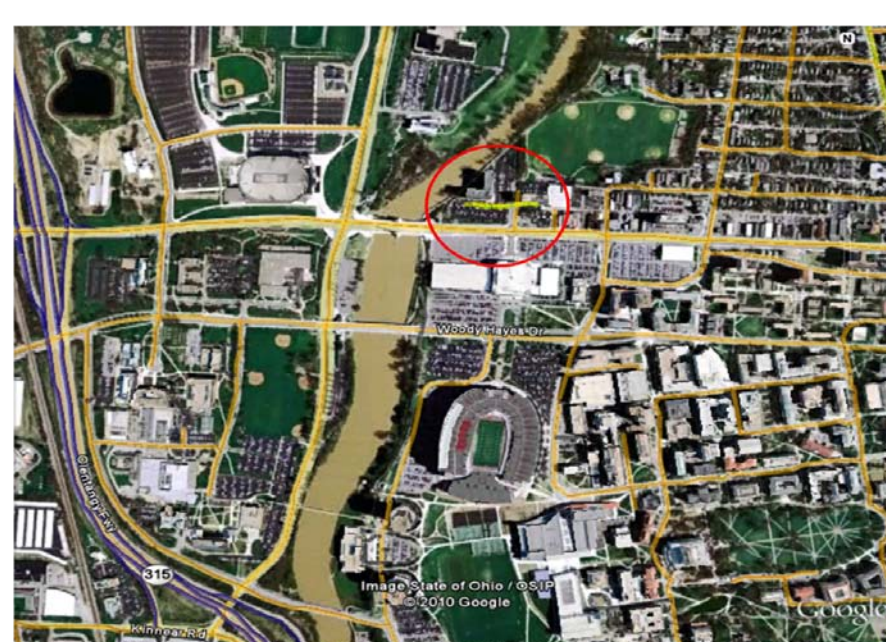
"IsOnRoad(X)"



"U-turn"



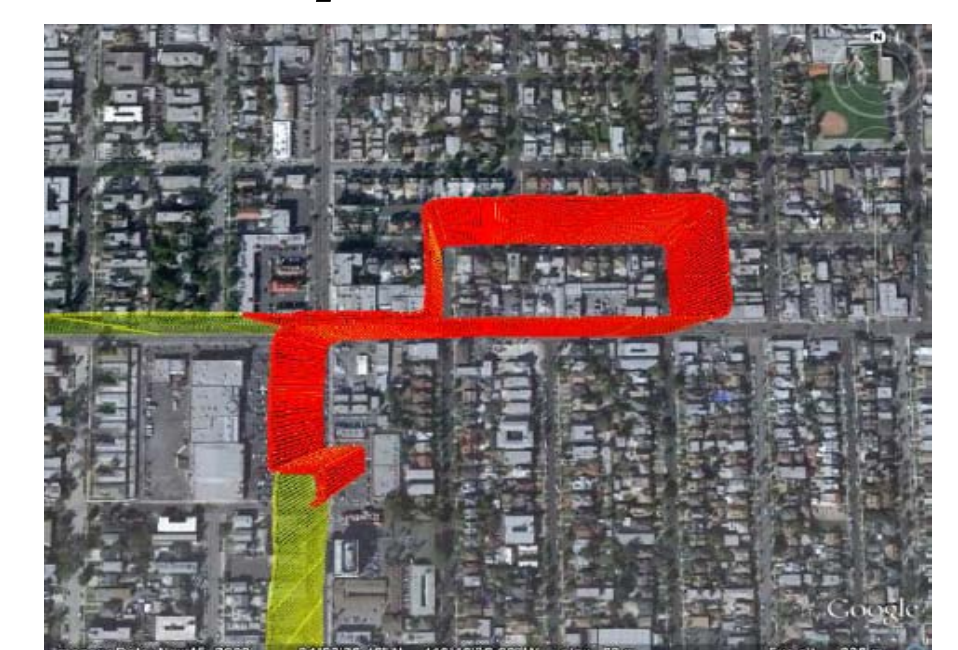
"loops" from a track



"source"



Rendering result ("source")



"loop" (zoom-in)

Examples of identified activities from GPS tracks

Examples of identified activities from aerial video