Overall Direction

-There exist two effective motion pattern mining algorithms developed in CRCV.

-Effective and efficient trajectory clustering is a contemporary problem in the data mining field.

-Modification of motion pattern algorithms can provide a solution to the data mining problem.
Motion Models Algorithm

- Trajectories
- Motion Patterns
- Optical Flow Representation Conversion
- (X,Y, ρ, θ) Data Representation
- Kmeans Clustering
- Connected Components
- Multivariate Gaussians
- Gathered 23 contemporary trajectory clustering papers.

- Identified commonly used and easily accessible trajectory datasets.

- Acquiring implementations of trajectory clustering algorithms for comparison
# List of Papers

<table>
<thead>
<tr>
<th>Title</th>
<th>Journal</th>
<th>Date</th>
<th>Cite</th>
<th>Data</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vector Field k-Means: Clustering Trajectories by Fitting Multiple Vector Fields</td>
<td>arxiv</td>
<td>2012</td>
<td>0</td>
<td>GeoLife GPS Trajectories, [1] HURDAT, Call Detail Records (CDR)</td>
<td>Coming in 2 weeks from 6/4</td>
</tr>
<tr>
<td>Profiling Moving Objects by Dividing and Clustering Trajectories Spatiotemporally</td>
<td>IEEE</td>
<td>2012</td>
<td>0</td>
<td>[28] [29] typhoon, taxi &amp; bus in Taipei</td>
<td>asked via email 6/6</td>
</tr>
<tr>
<td>Trajectory Analysis Using Spectral Clustering and Sequence Pattern Mining</td>
<td>JCIS</td>
<td>2012</td>
<td>1</td>
<td>urban intersection (not sourced)</td>
<td></td>
</tr>
<tr>
<td>Finding homogeneous groups in trajectory streams</td>
<td>ACM</td>
<td>2012</td>
<td>0</td>
<td>(4,1) bus &amp; truck in greece, Oregon wildlife</td>
<td>asked via email 6/6</td>
</tr>
<tr>
<td>Robust online trajectory clustering without computing trajectory distances</td>
<td>ICPP</td>
<td>2012</td>
<td>0</td>
<td>surveillance (not sourced)</td>
<td></td>
</tr>
<tr>
<td>Using Relative Distance and Hausdorff Distance to Mine Trajectory Clusters</td>
<td>IJES</td>
<td>2012</td>
<td>0</td>
<td>Hurricane Data</td>
<td></td>
</tr>
<tr>
<td>Learning trajectory patterns by clustering: Experimental studies and comparative evaluation</td>
<td>IEEE</td>
<td>2009</td>
<td>44</td>
<td>intersection surveillance (not sourced)</td>
<td></td>
</tr>
<tr>
<td>Trajectory pattern mining</td>
<td>ACM</td>
<td>2007</td>
<td>344</td>
<td>trucks in greece</td>
<td></td>
</tr>
<tr>
<td>Mining Frequent Trajectory Patterns from GPS Tracks</td>
<td>CISE</td>
<td>2010</td>
<td>1</td>
<td>taxis in Shenzhen (not sourced)</td>
<td></td>
</tr>
<tr>
<td>Learning Pedestrian Trajectories with Kernels</td>
<td>ICPR</td>
<td>2010</td>
<td>2</td>
<td>surveillance (not sourced)</td>
<td></td>
</tr>
<tr>
<td>A Complete Framework for Clustering Trajectories</td>
<td>IEEE</td>
<td>2009</td>
<td>2</td>
<td>bus &amp; truck in greece, Oregon wildlife</td>
<td>asked via email 6/6</td>
</tr>
<tr>
<td>A Trajectory Clustering Algorithm Based on Symmetric Neighborhood</td>
<td>World C congr.</td>
<td>2009</td>
<td>2</td>
<td>Oregon wildlife</td>
<td>asked via email 6/6</td>
</tr>
<tr>
<td>MoveMine: Mining Moving Object Databases</td>
<td>ACM</td>
<td>2010</td>
<td>29</td>
<td>movebank.org; weather.unisys.com/hurricane/at</td>
<td></td>
</tr>
<tr>
<td>Swarm: mining relaxed temporal moving object clusters</td>
<td>VLDB</td>
<td>2010</td>
<td>41</td>
<td>Movebank.org</td>
<td></td>
</tr>
<tr>
<td>Detection of risk factors using trajectory mining</td>
<td>JIIS</td>
<td>2011</td>
<td>15</td>
<td>ASL</td>
<td></td>
</tr>
<tr>
<td>Incremental Clustering for Trajectories</td>
<td>DASFAA</td>
<td>2010</td>
<td>31</td>
<td>SF Taxi GPS</td>
<td></td>
</tr>
<tr>
<td>Mining Spatio-Temporal Patterns in Trajectory Data</td>
<td>JIPS</td>
<td>2010</td>
<td>6</td>
<td>Synthetic data using spatio-temporal data generator G-TERD</td>
<td></td>
</tr>
<tr>
<td>TMC-pattern: holistic trajectory extraction, modeling and mining</td>
<td>SIGSPATIAL</td>
<td>2012</td>
<td>0</td>
<td>GeoLife</td>
<td></td>
</tr>
<tr>
<td>An Incremental DPMM-Based Method for Trajectory Clustering, Modeling, and Retrieval</td>
<td>IEEE</td>
<td>2013</td>
<td>0</td>
<td>ASL, Synthetic, Vehicle motion</td>
<td></td>
</tr>
<tr>
<td>Clustering and aggregating clues of trajectories for mining trajectory patterns and routes</td>
<td>VLDB</td>
<td>2011</td>
<td>0</td>
<td>CarlWeb</td>
<td></td>
</tr>
<tr>
<td>Mining Trajectories for Spatio-temporal Analytics</td>
<td>ICDMW (IBM)</td>
<td>2012</td>
<td>0</td>
<td>mentions testing with large scale real world data; no source provided</td>
<td></td>
</tr>
</tbody>
</table>
Trajectory Clustering: Partition-and-Group Framework

- Written by Lee, Han, Whang in 2007
- Serves as foundation for MoveMine set of works
- 357 citations

Figure 3: An example of trajectory clustering in the partition-and-group framework.
• Preciseness vs Conciseness

• Characteristic points – points where the behavior of trajectory changes rapidly
• MDL (Minimum Description Length) principle
• L(H) conciseness (hypothesis)
• L(D|H) preciseness
The optimal partitioning of a trajectory should possess two desirable properties: preciseness and conciseness. Preciseness means that the difference between a trajectory and a set of its trajectory partitions should be as small as possible.
MDL Principle

MDL cost = L(H) + L(D|H)
– L(H) represents the sum of the length of all trajectory partitions (conciseness)
– L(D|H) represents the number of segments that deviate from actual trajectory (preciseness)
– We need to find the optimal partitioning that minimizes L(H) + L(D|H), achieving the optimal tradeoff between preciseness and conciseness.
Clustering

- Based on DBSCAN
- Parameters common to TRACLUS and DBSCAN
  - $\varepsilon$ – the maximum distance
  - MinLns – minimum number of line segments in a cluster
- Parameter unique to TRACLUS:
  - Trajectory cardinality of a cluster:
    \[ \text{PTR}(C_i) = \{ \text{TR}(L_j) \mid \forall L_j \in C_i \} \]
TRACLUS

- Parameter selection
  - $\varepsilon$ - simulated annealing
  - MinLns – average number of lines at an optimal $\varepsilon$

- Complexity –
  - $O(n^2)$
  - Depending on organization and indexing of data (line segments), complexity can be reduced to $O(n \log n)$
1476 Atlantic hurricane trajectories recorded from 1851 to 2011.


Algorithm Comparison

UCF

Vector Field
Kmeans
Truck Data

GPS tracks from concrete trucks around Athens from 2002. 1100 trajectories.


Algorithm Comparison

UCF

TraClus
Deer 1995 dataset

- 32 trajectories
- 20,065 data points

Used in the following papers:


- Jae-Gil Lee, Jiawei Han, Xiaolei Li, and Hector Gonzalez. 2008. TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. Proc. VLDB Endow. [cited by 83]

- Zhenhui Li, Jae-Gil Lee, Xiaolei Li, and Jiawei Han. 2010. Incremental clustering for trajectories. In Proceedings of the 15th international conference on Database Systems for Advanced Applications - Volume Part II (DASFAA’10), [cited by 31]


- Yu Zhang and Dechang Pi. 2009. A Trajectory Clustering Algorithm Based on Symmetric Neighborhood. In Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering - Volume 03 (CSIE ’09), [cited by 2]

- Jae-Gil Lee, Jiawei Han, and Xiaolei Li. 2008. Trajectory Outlier Detection: A Partition-and-Detect Framework. In Proceedings of the 2008 IEEE 24th International Conference on Data Engineering (ICDE ’08).[cited by 112]
Performance Comparison

UCF

TRACLUS
Elk 1993 dataset
● 43 trajectories
● 47,204 points

Used in the following papers:

• Zhenhui Li, Jae-Gil Lee, Xiaolei Li, and Jiawei Han. 2010. Incremental clustering for trajectories. In Proceedings of the 15th international conference on Database Systems for Advanced Applications - Volume Part II (DASFAA’10), [cited by 31]


• Yu Zhang and Dechang Pi. 2009. A Trajectory Clustering Algorithm Based on Symmetric Neighborhood. In Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering - Volume 03 (CSIE ’09), [cited by 2]

• Jae-Gil Lee, Jiawei Han, Xiaolei Li, and Hector Gonzalez. 2008. TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. Proc. VLDB Endow. [cited by 83]

• Jae-Gil Lee, Jiawei Han, and Xiaolei Li. 2008. Trajectory Outlier Detection: A Partition-and-Detect Framework. In Proceedings of the 2008 IEEE 24th International Conference on Data Engineering (ICDE ’08). [cited by 112]
“Swainson’s Hawks converged in eastern Mexico on the Gulf of Mexico coast. Southward, these hawks followed a narrow, well-defined path through Central America, across the Andes Mountains in Columbia, and east of the Andes to central Argentina where they all spent the austral summer. Swainson’s Hawks northward migration largely retraced their southward route.”

Performance Comparison

UCF

TRACLUS
African Buffalo dataset

- 165 trajectories
- 26,609 points

Used in the following works:

- Zhenhui Li, Ming Ji, Jae-Gil Lee, Lu-An Tang, Yintao Yu, Jiawei Han, and Roland Kays. 2010. MoveMine: mining moving object databases. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data (SIGMOD '10). [cited by 29]

- Zhenhui Li, Bolin Ding, Jiawei Han, and Roland Kays. 2010. Swarm: mining relaxed temporal moving object clusters. In Proceedings of VLDB Endowment. [Cited by 41]
Performance Comparison

UCF

TRACLUS