

# **ERECTION OF INTERESTING GEOSPATIAL REGIONS IN THE CONTEXT OF ON-THE- SPOT LEARNING**

SHIVENDRA PRASAD TIWARI



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
**INDIAN INSTITUTE OF TECHNOLOGY DELHI**

JULY 2016

**©Indian Institute of Technology Delhi (IITD), New Delhi, 2016**

# **ERECTION OF INTERESTING GEOSPATIAL REGIONS IN THE CONTEXT OF ON-THE- SPOT LEARNING**

by

**SHIVENDRA PRASAD TIWARI**

Department of Computer Science and Engineering

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the



**INDIAN INSTITUTE OF TECHNOLOGY DELHI**

JULY 2016

# Certificate

The thesis entitled **Erection of Interesting Geospatial Regions in the Context of On-the-Spot Learning** being submitted by **Mr. Shivendra Prasad Tiwari** to the **Indian Institute of Technology Delhi**, for award of the degree of **Doctor of Philosophy** is a record of original bona-fide research work carried out by him. He has worked under my guidance and supervision, and has fulfilled the requirements for the submission of this thesis, which has attained the standard required for a Ph.D. degree of this institute.

The results presented in this thesis have not been submitted elsewhere for the award of any other degree or diploma.

**Dr. Saroj Kaushik**

Professor,

Department of Computer Science and Engineering,

Indian Institute of Technology Delhi

New Delhi

## **Acknowledgements**

I would thank god for offering me a courage and determination to pursue the higher education in the field of Computer Science and Engineering. I would like to tender my heartfelt thankfulness to my supervisor Prof. Saroj Kaushik for her priceless supervision, continuous inspiration and enthusiastic support in every stage of the work. Her deep insight into problems, imaginative ideas, technical guidance and sufficient encouragement was a constant source of inspiration and motivation. Her critical comments on my technical writing, eye for detail and persistent encouragement strongly motivated me to strive hard to achieve the high targets set by her in completion of this thesis. Her religious, humane and positive outlook towards the life has been a great source of inspiration even in the non-academic side of life.

I am extremely grateful to the members of my Student Research Committee (SRC) – Prof. Huzur Saran, Prof. Aditeshwar Seth, Prof. Brejesh Lall, who have been very helpful by offering suggestions and advice. I thank my Ph.D. colleagues, specially, Priti Jagwani, Sunita Tiwari, and Kuntal Dey for their help on reviewing my work and offering useful suggestions.

My parents and my wife Sunita Sharma showed immense patience and provided me great support during the course of my work. Their patience, sacrifice, inspiration and moral support are of special mention. This thesis is dedicated to them.

**Shivendra Prasad Tiwari**

# Abstract

The increasing requirements to handle more information on mobile devices are putting more demands on context sensitive computing and visualization than ever. Thus, it would require having a Location-based Information System (LBIS) that changes the level of granularity based on the user interest. Current Location Based Service (LBS) systems use lowest level of the Point of Interests (POI) details for information queries, which does not serve the next generation LBS requirements such as the storytellers, and on-demand information granularity. There is lack of digitization in remote areas, context aware selection of interesting objects, and the solutions must comply with the existing LBS infrastructure. The proposed research deals with the following problems:

*Lack of digitization* – several developing countries still do not have the digitized road network, and POIs. Mobile technology helps generating user movement footprints i.e. trajectory data. We have used trajectory patterns for automating spatial data collection, interesting region analysis, etc.

*Unrated POIs* – millions of POIs are still unrated due to the lack of awareness or interest on manual participation. The trajectory analysis is another way of estimating the popularity. We have proposed automatic popularity estimation based on the weighted user check-in and visit patterns.

*Erection and storage model of the ROI* – we have proposed a model that includes the erection of the ROIs along with storage and retrieval with weighting heuristic. The proposed storage methods include administrative boundaries, dynamic spatial chunking and a hybrid method that achieves the advantages of both the ROI creation and storage models.

*Erection of ROIs using POI clusters* – we have proposed an Adjacent Grid Block Selection (AGBS) method for efficiently finding boundary points of the POI clusters on the road network.

*Lack of organized information sources* – freely available web data, and e-Learning systems are extended to support the location based learning features. Fusion of navigation infrastructure and e-learning system is proposed in order to create information rich ROIs.

In order to validate the proposed solutions for the individual problems, we have implemented the prototypes using trajectory database provided by Microsoft Research, POI data collected from HERE Maps and freely available Wikipedia content as source of learning content associated with specific geospatial regions.

# Table of Contents

<i>Certificate</i> .....	<i>i</i>
<i>Acknowledgements</i> .....	<i>iii</i>
<i>Abstract</i> .....	<i>v</i>
<i>Table of Contents</i> .....	<i>vii</i>
<i>List of Figures</i> .....	<i>xi</i>
<i>List of Tables</i> .....	<i>xv</i>
<i>Acronyms</i> .....	<i>xvii</i>
<b>Chapter 1 Introduction</b> .....	<b>1</b>
<b>1.1 Location Based Services: Basics</b> .....	<b>1</b>
1.1.1 Point of Interests (POI) .....	4
1.1.2 Region of Interests (ROI) .....	5
1.1.3 Geospatial Data Structures.....	7
1.1.4 Key Players in LBS System.....	8
<b>1.2 Architectures of an LBS Ecosystem</b> .....	<b>9</b>
1.2.1 Generic LBS Ecosystem .....	10
1.2.2 Middleware Oriented Architecture .....	11
1.2.3 Location Platform by Kivera .....	12
1.2.4 Hybrid Architecture .....	14
1.2.5 Evaluation and Comparisons .....	16
<b>1.3 On-the-Spot Learning (OTSL)</b> .....	<b>17</b>
<b>1.4 Motivation</b> .....	<b>20</b>
<b>1.5 Challenges</b> .....	<b>22</b>
<b>1.6 Proposed Research</b> .....	<b>23</b>

1.6.1 Outline of the Thesis .....	23
<b>Chapter 2 Erection of Interesting Regions using Shared Nearest Neighbor (SNN) based Trajectory Clustering .....</b>	<b>27</b>
<b>2.1 Introduction .....</b>	<b>27</b>
<b>2.2 Related Work .....</b>	<b>31</b>
<b>2.3 Problem Definition and Challenges.....</b>	<b>33</b>
<b>2.4 Proposed Solution .....</b>	<b>36</b>
2.4.1 Data Modeling and Pre-processing .....	37
2.4.2 Trajectory Labelling and Storage Structure .....	42
2.4.3 SNN Based Trajectory Clustering .....	46
<b>2.5 Implementation and Evaluation .....</b>	<b>50</b>
2.5.1 Evaluation of Improved-SNN Using Synthetic Geospatial Points.....	51
2.5.2 Evaluation with GeoLife Trajectory Dataset .....	54
2.5.3 Evaluation with Animal Trajectory Dataset .....	57
<b>2.6 Conclusion.....</b>	<b>59</b>
<b>Chapter 3 Estimating Popularity of Interesting Geospatial Regions using GPS Trajectory Mining .....</b>	<b>61</b>
<b>3.1 Introduction .....</b>	<b>61</b>
<b>3.2 Related Work .....</b>	<b>63</b>
<b>3.3 Problem Definition and Challenges.....</b>	<b>65</b>
<b>3.4 Proposed Solution .....</b>	<b>68</b>
3.4.1 Data Modeling and Pre-processing .....	68
3.4.2 User Check-in and Visitor’s Registry .....	69
3.4.3 Home Location and User Categorization .....	72
3.4.4 Popularity Estimation.....	73
<b>3.5 Implementation and Evaluation .....</b>	<b>76</b>

3.6 Conclusion .....	82
<b><i>Chapter 4 Erection of ROI using kNN Graph Based POI Clustering on the Road Network</i></b> .....	<b>85</b>
4.1 Introduction.....	85
4.2 Related Work.....	89
4.3 Problem Definition and Challenges .....	91
4.4 Distance Modeling on Road Networks .....	93
4.5 AGBS kNN-Join Computation .....	96
4.5.1 Grid Block Generation .....	96
4.5.2 AGBS kNN-Join.....	99
4.6 kNN graph Based Cluster Generation .....	102
4.6.1 kNN Graph Generation using AGBS .....	102
4.6.2 kNNG Partitioning and Boundary Nodes Detection Through Threshold Filter .....	103
4.7 Implementation and Evaluation.....	103
4.7.1 kNN-Join: Gorder vs AGBS .....	104
4.7.2 Boundary Points Detection: BORDER vs AGBS-BORDER.....	105
4.8 Conclusion .....	107
<b><i>Chapter 5 Erection of ROI using Dynamic Landmark Chunking</i></b> .....	<b>109</b>
5.1 Introduction.....	109
5.2 Related Work.....	111
5.3 Problem Definition and Challenges .....	113
5.4 Proposed Solution.....	114
5.4.1 Weighted Heuristics Based Saliency Computation.....	117
5.4.2 Fuzzy Reasoning Based Saliency Computation.....	118
5.4.3 Extensions of the Basic Solution.....	120

<b>5.5 Implementation and Evaluation .....</b>	<b>121</b>
5.5.1 Performance Evaluation.....	122
5.5.2 Accuracy of Landmarks Ranking .....	125
<b>5.6 Conclusion.....</b>	<b>128</b>
 <b>Chapter 6 Extracting ROI Content using Existing E-Learning Systems and Web Data</b>	
<b>Sources.....</b>	<b>131</b>
 <b>6.1 Introduction .....</b>	<b>131</b>
<b>6.2 Related Work .....</b>	<b>132</b>
<b>6.3 Problem Definition and Challenges.....</b>	<b>134</b>
<b>6.4 Proposed Solution .....</b>	<b>135</b>
6.4.1 Architecture of ROI Based Information System .....	136
6.4.2 Storage Structure of Region of Interests (ROI) .....	138
6.4.3 Location Based Learning Operations .....	141
<b>6.5 Implementation and Evaluation .....</b>	<b>146</b>
6.5.1 Spatial Search: Getting Linguistic Labels for Query Point .....	148
6.5.2 Geocoding: Linguistic Search Using OpenStreetMap.....	149
6.5.3 Fetching Region/Landmark Details .....	150
6.5.4 Accuracy of Underlying Services and Landmarks Learning Content .....	155
<b>6.6 Conclusion.....</b>	<b>156</b>
 <b>Chapter 7 Conclusion and Future Work .....</b>	
<b>Bibliography.....</b>	<b>161</b>
<b>Appendices.....</b>	<b>171</b>
<b>Publication out of the Thesis .....</b>	<b>195</b>
<b>Biography of the Author .....</b>	<b>197</b>

# List of Figures

Fig. 1.1 Image Showing Flight Movement.....	6
Fig. 1.2 Region Details as ROI .....	7
Fig. 1.3 Generic Location Based System Architecture [88].....	10
Fig. 1.4 Middleware Based Architecture of LBS Systems [116] .....	11
Fig. 1.5 Architecture of Kivera Location Platform .....	13
Fig. 1.6 Hybrid Architecture diagram of publish/subscribe (p/s) based OTSL system. ....	15
Fig. 1.7 Example application of the on-the-spot learning .....	19
Fig. 1.8 High-level block diagram of the proposed solution for OTSL.....	24
Fig. 2.1 Example of stationary trajectory - the GPS points are spotted in a random pattern.....	29
Fig. 2.2 Visual comparison of the clustering methods – (a) DBSCAN (MinPts=4, Eps=9.75); (b) DBSCAN (MinPts=4, Eps=9.92); (c) K-means Clustering and (d) SNN Clustering. ....	30
Fig. 2.3 System diagram of the proposed method for erection of Region of Interest.....	37
Fig. 2.4 Map Matching example at the runtime in the dense parallel road scenarios.....	39
Fig. 2.5 Runtime Map Matching example in the flyover and bridge scenarios.....	39
Fig. 2.6 Typical trajectory patterns for the interesting regions.....	44
Fig. 2.7 Logical representation of the trajectories when stored in spatial grid structure. ....	45
Fig. 2.8 An example of clipping the trajectories to fit into the spatial grids - the common points are inserted into all the sharing cells to improve the clustering efficiency. ....	45
Fig. 2.9 Storage structure of trajectory label indices in a cell of the spatial grid .....	46
Fig. 2.10 The geographic domain is divided into a set of spatial grids: searching for neighbors within the current cell in the 0 <sup>th</sup> iteration; and within the surrounding cells in the 1 <sup>st</sup> iteration onwards. ....	49
Fig. 2.11 Performance distribution of the SNN clustering steps for k=10, Eps = 3, and MinPts = 7.....	51
Fig. 2.12 Performance distribution of the SNN clustering steps for k=20, Eps = 7, and MinPts = 14.....	52
Fig. 2.13 Performance comparison of the SNN clustering for k=10, Eps = 3, and MinPts = 7 .....	53

Fig. 2.14 Performance comparison of the SNN clustering for $k=20$ , $Eps = 7$ , and $MinPts = 14$ .....	53
Fig. 2.15 Example GPS trajectory: first few rows of a trajectory file from a user recorded in 2008.....	54
Fig. 2.16 Fixed window check-in: The Ruins of Vuanmingyuan Beijing .....	56
Fig. 2.17 SNN clustering based ROI: The Ruins of Vuanmingyuan Beijing .....	56
Fig. 2.18 Fixed Window check-in: Chemical Engg Inst. Tsinghua University China.....	56
Fig. 2.19 SNN clustering based ROI: : Chemical Engg Inst. Tsinghua Univ. China .....	56
Fig. 2.20 Visual comparison of TRACUS and SNN Clusters with Deer(1995) dataset. ....	58
Fig. 2.21 Visual comparison of TRACUS and SNN Clusters with Elks (1993) dataset.....	58
Fig. 2.22 SNN Clusters with Deer(1993) dataset for different $k$ , $Eps$ and $MinPts$ parameters.....	59
Fig. 2.23 SNN Clusters with Cattle(1993) dataset for different $k$ , $Eps$ and $MinPts$ parameters .....	59
Fig. 3.1 The place's popularity depending on different scenarios and factors. ....	62
Fig. 3.2 Overall component diagram for proposed popularity estimation. The popularity value is saved in the POI database for further user in various applications. ....	69
Fig. 3.3 Demonstration of the stay region/POI matching with the GPS trajectories .....	71
Fig. 3.4 Block diagram of popularity estimation explaining rating computation process. ....	73
Fig. 3.5 Fuzzification of the stay time and travel distance of the visitors and computing the weighted frequency (visitor importance). ....	75
Fig. 3.6 Fuzzification of the total weighted frequency of a place and computing the final rating based on the weighted frequency.....	76
Fig. 3.7 Category distribution of the places in the underlying POI database.....	77
Fig. 3.8 Rating trend of the POI – most of the POIs are unrated.....	79
Fig. 3.9 Place visit frequency distribution – capped at 100. ....	79
Fig. 3.10 Original user rating distribution provided in the POI database. ....	80
Fig. 3.11 Comparison of ratings distribution – frequency based rating .....	81
Fig. 3.12 Comparison of ratings distribution –weighted frequency based rating. ....	81
Fig. 4.1 kNN Graph for $k=3$ . The outgoing edges show the kNN relationships; however, the incoming edges denote the Reverse k-Nearest Neighbor (RkNN) relationships among the nodes .....	86

Fig. 4.2 kNN vs RkNN: (a) The points P1, P3 and P4 mark the point P2 as their nearest neighbor. (b) The point P2 marks P1, P3 and P4 as its reverse-nearest neighbor. ....	88
Fig. 4.3 A road network with the point of interest $P_i$ on the edges. The edges have weights, and nodes are numbered by $N_i$ . ....	94
Fig. 4.4 The storage structure of the nodes and the point of interests. ....	96
Fig. 4.5 The multilevel hierarchical structure for the I/O and CPU cost optimization ....	97
Fig. 4.6 Illustration of spatial grid-based join operation. The corresponding grids are searched for the KNN-join objects first.....	100
Fig. 4.7 AGBS Block Traversal in their adjacency neighboring order.....	101
Fig. 4.8 kNN-join: Comparison between Gorder and AGBS with variable k values .....	104
Fig. 4.9 kNN-join: Comparison between Gorder and AGBS with variable dataset size.....	105
Fig. 4.10 Boundary point detection: Comparison between BORDER and AGBS-BORDER with variable k .....	106
Fig. 4.11 Performance comparison: BORDER vs AGBS-BORDER with variable dataset size .....	106
Fig. 5.1 Illustration of the Region of Interest around the route .....	115
Fig. 5.2 Block diagram of the proposed POI selection approach. ....	119
Fig. 5.3 Screen shot of the Fuzzy Inference Engine for Saliency Computation of the Landmarks .....	120
Fig. 5.4 Performance comparisons with increasing speed; fixed content size $t_c = 10$ mins .....	122
Fig. 5.5 Performance comparisons with increasing content play time; fixed speed = 50km/h .....	123
Fig. 5.6 Performance of different methods of POI chunking, variable speed, content time $t_c=10$ mins .....	124
Fig. 5.7 Performance of different methods of POI chunking - Fixed speed = 50km/h, variable content size. ....	125
Fig. 5.8 Search results – landmark-ranking comparisons among the distance based sorted list, popularity based sorted list and the saliency values based sorted list. ....	126
Fig. 5.9 Error on search results – ranking error among the distance based sorted list, popularity based sorted list and the saliency values based sorted list.....	127
Fig. 6.1 Architecture of the ROI based Information System.....	137

Fig. 6.2 The Hierarchical Storage of ROI data .....	138
Fig. 6.3 Landmark Chunking Based Hierarchy .....	139
Fig. 6.4 Hybrid ROI: chunked ROIs at lower level, and administrative records at higher levels .....	140
Fig. 6.5 Learning content geocoding process: conventional geocoding is used to associate the coordinates with the learning materials.....	142
Fig. 6.6 Content reverse-geocoding process .....	144
Fig. 6.7 The learning sight attributes and learning resource properties (data modeling) .....	145
Fig. 6.8 User attributes and preferences (data modeling) .....	146
Fig. 6.9 Fetching the location based content from the web .....	147
Fig. 6.10 Sample geocoding response from OpenStreetMap.....	148
Fig. 6.11 Reverse-geocoding performance comparisons between different available services .....	150
Fig. 6.12 Geocoding performance comparison between different service providers .....	151
Fig. 6.13 An example of the façade content response from Wikipedia for Golden Gate Bridge.....	152
Fig. 6.14 The performance comparison between façade content method vs detailed content method .....	153
Fig. 6.15 The data size comparison between the façade content vs the detailed content.....	153
Fig. 6.16 The performance comparison when the number of the POIs in the cluster increase. Switching to the administrative levels of the location improves the performance. ....	154
Fig. 6.17 The data size transfer comparison with increasing number of POIs. ....	155

# List of Tables

Table 1.1 Comparison of the LBS Architectures for Different Attributes .....	17
Table 2.1 Trajectory data used in the experiment - distribution of trajectories by distance .....	55
Table 2.2 Trajectory data used in the experiment - data collection duration.....	55
Table 2.3 Trajectory data used in the experiment - distribution of trajectories by effective duration.....	55
Table 3.1 Ratings comparison of top 20 places in Beijing visited by the trajectory users .....	78
Table 5.1 The salience factors contributing in landmark selection for dynamic chunking. ....	116
Table A.1 Performance distribution of the SNN clustering steps for k=10, Eps = 3, and MinPts = 7 for randomly generated geo-coordinates.....	173
Table A.2 Performance distribution of the SNN clustering steps for k=20, Eps = 7, and MinPts = 14 for randomly generated geo-coordinates.....	173
Table A.3 Performance comparison of the SNN clustering methods for k=10, Eps = 3, and MinPts = 7 for randomly generated geo-coordinates.....	173
Table A.4 Performance comparison of the SNN clustering methods for k=20, Eps = 7, and MinPts = 14 for randomly generated geo-coordinates.....	174
Table A.5 The top visited place in Beijing out of 26700 Point of Interests (POI) collected from HereMaps	177
Table A.6 The rating trend of 26807 POIs in Beijing collected from HereMaps [49] .....	179
Table A.7 Places category distribution in the place in Beijing out of 26700 Point of Interests (POI) collected from HereMaps [49].....	179
Table A.8 Gorder vs AGBS kNN-Join: Comparison with Fixed Dataset (2000 Points), Fixed Block Size and Variable k Values .....	180
Table A.9 Gorder vs AGBS kNN-Join: Comparison with Variable Size Dataset, Fixed Block Size and Fixed k Values (k=50).....	180
Table A.10 BORDER vs AGBS-BORDER: Comparison with Fixed Dataset (3000 Points), Fixed Block Size and Variable k values .....	181
Table A.11 BORDER vs AGBS-BORDER: Comparison with Variable Size Dataset, Fixed Block Size and Fixed k Values (k=50).....	181

Table A.12 Performance comparisons of different landmark search methods with increasing speed fixed content size $t_c = 10$ mins .....	184
Table A.13 Performance comparisons of different landmark search methods with increasing content play time fixed speed = 50km/h .....	185
Table A.14 Performance comparisons of different methods of POI chunking, variable speed, content time $t_c=10$ mins .....	185
Table A.15 Performance comparisons of different methods of POI chunking - Fixed speed = 50km/h, and variable Content Size) .....	186
Table A.16 Places ranking comparison and error values. Error = (place rank -1)/total candiate places.....	190
Table A.17 Geocoding performance using different mapping services for the individual point of interests in San Francisco, CA, USA. ....	191
Table A.18 Reverse-Geocoding performance using different mapping services for the individual point of interests in San Francisco, CA, USA. ....	191
Table A.19 Time comparison with individual POIs: initial download time requirement using Façade vs Detailed content request methods .....	192
Table A.20 Data size comparison with individual POIs: initial network data transfer size using Façade vs Detailed content request methods .....	192
Table A.21 Initial time requirement with increasing number of POIs using Façade vs Detailed content request methods.....	193
Table A.22 Initial Data transfer with increasing number of POIs using Façade vs Detailed content request methods.....	193

# Acronyms

AGBS	Adjacent Grid Block Selection
CLDC	Connected Limited Device Configuration
FC	Façade Content
GML	Geography Markup Language
GPS	Global Positioning System
GSM	Global System for Mobile Communication
HG	Hierarchical Grid
KML	Kilobyte Markup Language
kNN	k-Nearest Neighbor
LBL	Location Based Learning
LBS	Location Based Service
LDT	Location Determination Technology
LMS	Learning Management System
LSP	Location Based Service Provider
MIF	MapInfo Interchange Format
MMD	Multiple Multimedia Data
MMD	Min-Max Distance
NNQ	Nearest Neighbor Query
OGC	Open Geospatial Consortium
OMG	Optimal Meeting Grid
OSA	Open Services Architecture
OSM	Open Street Map
OTA	Over-The-Air
OTSL	On-The-Spot-Learning
PCA	Principal Component Analysis
POI	Point Of Interest
RkNN	Reverse k Nearest Neighbor
ROI	Region Of Interest
SAM	Spatial Access Method
SNN	Shared Nearest Neighbor
TBT	Turn-By-Turn