

CircularTrip and ArcTrip: Effective Grid Access Methods for Continuous Spatial Queries

by

Muhammad Aamir Cheema

THE UNIVERSITY OF
NEW SOUTH WALES



SYDNEY • AUSTRALIA

A THESIS SUBMITTED IN FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTERS IN COMPUTER SCIENCE & ENGINEERING RESEARCH
(March 2007)

Supervisor: Dr. Xuemin Lin

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: **Cheema**

First name: **Muhammad**

Other name/s: **Aamir**

Abbreviation for degree as given in the University calendar: **Masters by research in Computer Science and Engineering (COMPER2665)**

School: **School of Computer Science and Engineering (CSE)**

Faculty: **Engineering**

Title: **CircularTrip and ArcTrip: Effective Grid Access Methods for Continuous Spatial Queries**

Abstract 350 words maximum: (PLEASE TYPE)

A k nearest neighbor query q retrieves k objects that lie closest to the query point q among a given set of objects P . With the availability of inexpensive location aware mobile devices, the continuous monitoring of such queries has gained lot of attention and many methods have been proposed for continuously monitoring the kNN s in highly dynamic environment. Multiple continuous queries require real-time results and both the objects and queries issue frequent location updates. Most popular spatial index, R-tree, is not suitable for continuous monitoring of these queries due to its inefficiency in handling frequent updates. Recently, the interest of database community has been shifting towards using grid-based index for continuous queries due to its simplicity and efficient update handling. For kNN queries, the order in which cells of the grid are accessed is very important. In this research, we present two efficient and effective grid access methods, CircularTrip and ArcTrip, that ensure that the number of cells visited for any continuous kNN query is minimum. Our extensive experimental study demonstrates that CircularTrip-based continuous kNN algorithm outperforms existing approaches in terms of both efficiency and space requirement. Moreover, we show that CircularTrip and ArcTrip can be used for many other variants of nearest neighbor queries like constrained nearest neighbor queries, farthest neighbor queries and $(k+m)$ -NN queries. All the algorithms presented for these queries preserve the properties that they visit minimum number of cells for each query and the space requirement is low. Our proposed techniques are flexible and efficient and can be used to answer any query that is hybrid of above mentioned queries. For example, our algorithms can easily be used to efficiently monitor a $(k+m)$ farthest neighbor query in a constrained region with the flexibility that the spatial conditions that constrain the region can be changed by the user at any time.

Declaration relating to disposition of project thesis/dissertation

I hereby grant to the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all property rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstracts International (this is applicable to doctoral theses only).

.....
Signature

.....
Witness

.....
Date

The University recognises that there may be exceptional circumstances requiring restrictions on copying or conditions on use. Requests for restriction for a period of up to 2 years must be made in writing. Requests for a longer period of restriction may be considered in exceptional circumstances and require the approval of the Dean of Graduate Research.

FOR OFFICE USE ONLY

Date of completion of requirements for Award:

THIS SHEET IS TO BE GLUED TO THE INSIDE FRONT COVER OF THE THESIS

Approval

This thesis entitled

**CircularTrip and ArcTrip: Effective Grid Access Methods
for Continuous Spatial Queries**

by

Muhammad Aamir Cheema

has been approved by the School of Computer Science and Engineering,
The University of New South Wales, Sydney Australia.

Supervisor: Dr. Xuemin Lin

The final copy of this thesis has been examined by the signatory and I find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Signed

Date

Copyright Statement

I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation. I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only). I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.

Muhammad Aamir Cheema

Signed

Date

Authenticity Statement

I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.

Muhammad Aamir Cheema

Signed

Date

Originality Statement

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgment is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Muhammad Aamir Cheema

Signed

Date

Dedication

I dedicate this thesis to my parents who have always been my nearest and reverse nearest neighbors¹ [KM00] and have been so close to me that I found them with me whenever I needed. It is their unconditional love that motivates me to set higher targets. I also dedicate this thesis to my sisters (Shazia Cheema and Nadia Cheema) and brothers (Muhammad Omer Cheema, Muhammad Umair Cheema and Muhammad Ammar Cheema) who are my nearest surrounders² [LLL06] and have provided me with a strong love shield that always surrounds me and never lets any sadness enter inside.

¹In a not-so-academic wording, your nearest neighbor is an object that is closest to you and your reverse nearest neighbor is any object for which you are the closest object.

²Your nearest surrounders are the objects that are nearest to you and surround you completely.

Acknowledgements

I am very thankful to my parents for their love, support and encouragement and for being with me on each and every step of my life. I am also very thankful to my eldest brother Muhammad Omer Cheema for being very supporting and motivating.

I wish to express my gratitude to my supervisor Dr. Xuemin Lin for his help and guidance. I feel thankful to him for his insightful advice and suggestions. I am also thankful to Dr. Wei Wang for giving me generous amount of time whenever I needed some help. I am grateful to Yidong Yuan with whom I worked during my research and learnt a lot. I feel myself very lucky that I got very friendly colleagues. In particular, I am thankful to Ying Zhang for being very helping and cooperative, Mahady Hasan for being my closest colleague (it is also true literally, he sits very close to me), Yi Luo especially for arranging group meetings, Wenjie Zhang especially for being the only other colleague who is working on moving objects and Bin Jiang especially for being the only masters student in our research group other than me.

I do not want to miss this opportunity to thank people who made my stay in Australia an enjoyable experience without which I would not have been able to concentrate on my research. First of all I am very thankful to Azeem Ahmad Cheema, my cousin who is more like brothers, for being so caring and fun loving that made my stay in Australia a memorable period of my life. I am also grateful to his family, my other brother Waseem Ahmad Cheema and his family for being very caring and affectionate. Their kids gave me so many reasons to smile. I also feel blessed that my brother and best friend Umair was also here with me and I never felt lonely in his presence. I am also thankful to Assad Mehmood for giving me company during coffee breaks.

I feel grateful to all my friends back in Pakistan for being in contact with me and encouraging me. I am thankful to my younger brother and my crime partner Ammar for keeping me informed about family events in his own special way. I am also thankful to

my best friends (they are best at finding something in me to complain about), Mohsin Hasan and Samar Cheema, for being very caring and for keeping me up-to-date of issues and events happening in Pakistan especially in our friends' circle. I am thankful to my intimate friend Irfan Ahmad for being too busy to be in contact. I am also thankful to my sweet cousin Farman Ahmad Bajwa for always remembering me and especially for always making me angry by telling me not to be angry. Finally, I am very grateful to Daisy for giving me so many valuable suggestions for this part of my thesis. The best advice she gave me was to include her name in acknowledgments.

I feel obliged to say “thanks” to so many people who have been doing so many good things for me but I am afraid the list may go forever. Same as mathematicians write all positive even numbers in very small space as $\{0, 2, 4, \dots, \infty\}$, below I try to write a shorter acknowledgements. Let k be a positive integer, T be my lifetime, and $dist(p, q)$ be the distance between the hearts of any two people p and q reflecting the love the person p has for q , I feel grateful to all people who have ever been among my k nearest neighbors and/or k reverse nearest neighbors over the time period T .

Abstract

A k nearest neighbor query q retrieves k objects that lie closest to the query point q among a given set of objects P . With the availability of inexpensive location aware mobile devices, the continuous monitoring of such queries has gained lot of attention and many methods have been proposed for continuously monitoring the k NNs in highly dynamic environment. Multiple continuous queries require real-time results and both the objects and queries issue frequent location updates. Most popular spatial index, R-tree, is not suitable for continuous monitoring of these queries due to its inefficiency in handling frequent updates. Recently, the interest of database community has been shifting towards using grid-based index for continuous queries due to its simplicity and efficient update handling. For k NN queries, the order in which cells of the grid are accessed is very important. In this research, we present two efficient and effective grid access methods, CircularTrip and ArcTrip, that ensure that the number of cells visited for any continuous k NN query is minimum. Our extensive experimental study demonstrates that CircularTrip-based continuous k NN algorithm outperforms existing approaches in terms of both efficiency and space requirement. Moreover, we show that CircularTrip and ArcTrip can be used for many other variants of nearest neighbor queries like constrained nearest neighbor queries, farthest neighbor queries and $(k + m)$ -NN queries. All the algorithms presented for these queries preserve the properties that they visit minimum number of cells for each query and the space requirement is low. Our proposed techniques are flexible and efficient and can be used to answer any query that is hybrid of above mentioned queries. For example, our algorithms can easily be used to efficiently monitor a $(k + m)$ farthest neighbor query in a constrained region with the flexibility that the spatial conditions that constrain the region can be changed by the user at any time.

Contents

1	Introduction	1
1.1	Spatial Access Methods	3
1.2	Spatial Queries	5
1.3	Contributions	8
1.4	Thesis Organization	9
2	Related Work	11
2.1	k Nearest Neighbors Queries	11
2.1.1	Snapshot k NN Queries	13
2.1.2	Continuous k NN Queries	15
2.1.3	k NN Queries in Spatial Networks	21
2.1.4	Approximate k NN Queries	23
2.1.5	Distributed Processing of k NN Queries	25
2.2	Variants of Nearest Neighbor Queries	28
2.2.1	Reverse Nearest Neighbor Queries	28
2.2.2	Constrained Nearest Neighbor Queries	33
2.2.3	Aggregate Nearest Neighbor Queries	36
2.2.4	All-Nearest Neighbor Queries	39
3	Grid Access Methods	42
3.1	Motivation	42
3.2	CircularTrip	44
3.2.1	CircularTrip on 2D Grid	45
3.2.2	Extension to Higher Dimensions	48
3.3	ArcTrip	48
3.4	Summary	51
4	CircularTrip Based Continuous kNN Algorithm	53
4.1	Motivations	54
4.2	System Overview and Data Structure	55
4.3	Initial k NN Computation	56
4.4	Continuous Monitoring	58
4.4.1	Handling Data Point Updates	59
4.4.2	Handling Query Updates	61
4.4.3	Complete Update Handling Module	63
4.4.4	Updating the Influence List	64

4.5	Performance Analysis	66
4.6	Experimental Study	70
4.6.1	Evaluating Efficiency	71
4.6.2	Evaluating Effect of Agility	72
4.7	Conclusion	74
5	Applications to Other Variants of NN Queries	75
5.1	Constrained Nearest Neighbor Queries	76
5.1.1	Continuous Pie-Region k CNN queries	77
5.1.2	Continuous Donut-Region k CNN Queries	80
5.1.3	Continuous Donut-Pie k CNN Queries	83
5.1.4	Continuous k CNN Queries Over Irregular Region	86
5.1.5	Discussion	87
5.2	Farthest Neighbor Queries	88
5.2.1	Initial Computation	88
5.2.2	Continuous Monitoring	89
5.2.3	Discussion	91
5.3	$(k + m)$ NN Queries	91
5.3.1	Technique	91
5.3.2	Discussion	92
5.4	Conclusion	93
6	Conclusion	94

List of Figures

1.1	Google Maps (http://maps.google.com)	2
1.2	Difficulty in Finding the Spatial Proximity of Objects	4
1.3	Effect of Different Distance Definitions on Nearest Neighbors	7
2.1	An Example of a NN Query on R-tree Index	14
2.2	YPK-CNN	18
2.3	SEA-CNN	19
2.4	CPM	20
2.5	Example of Safe Regions	26
2.6	Threshold-Based Algorithm for Monitoring a 3-NN Query	28
2.7	The Objects a and e are the Reverse Nearest Neighbors of q	30
2.8	Illustration of SAA	31
2.9	Half-Plane Pruning	32
2.10	The Monitoring Region of a Continuous RNN Query	33
2.11	Illustration of $mindist(q, M, R)$ and $minmaxdist(q, M, R)$	34
2.12	Computation of a Constrained Nearest Neighbor Query	35
2.13	Node N_1 can be pruned	37
2.14	Different Pruning Metrics for All-NN Queries	41
3.1	Grid Index	43
3.2	Minimal Set of Cells	43
3.3	CircularTrip	45
3.4	Next Intersected Cell is Either c_c or c_N	47
3.5	Minimum Angular Distance	48
3.6	ArcTrip	49
3.7	The Special Cell c_{spe} is Always the Adjacent Cell of c in D_{opp}	50
4.1	System Overview and Data Structure	56
4.2	A Nearest Neighbor Query	57
4.3	Handling Data Point Updates	61
4.4	Handling Query Updates	62
4.5	Handling Multiple Updates	64
4.6	Updating Influence Region	66
4.7	An Example of Updates Handling by CPM	68
4.8	An Example of Updates Handling by CircularTrip-based Algorithm	69
4.9	Effect of δ	72

4.10	Effect of k	72
4.11	Effect of N and n	72
4.12	Data Movement	73
4.13	Query Movement	73
4.14	Time Efficiency	73
5.1	Six Pie-Region Constrained NN Queries	78
5.2	A Continuous Pie-Region 1-CNN query	79
5.3	A Varying Pie-Region Continuous 1-CNN Query	80
5.4	A Donut-Region k CNN Query	81
5.5	Computation of a Donut-Region 1-CNN Query	82
5.6	A Varying Donut-Region Continuous 1-CNN query	83
5.7	Donut-Pie k CNN Query	84
5.8	Computation of a Donut-Pie 1-CNN Query	85
5.9	Computation of a Varying Donut-Pie Region 1-CNN Query	85
5.10	Continuous k CNN Query Over Irregular Region R	87
5.11	Computation of a Farthest Nearest Neighbor Query	89
5.12	Update Handling of a Farthest Neighbor Query	90
5.13	A $(k + m)$ NNs Query ($k = 3, m = 2$)	92

List of Tables

2.1	Depth-First Search	14
2.2	Best-First Search	15
3.1	Math Notations	44