Applying the Effective Distance to Location-Dependent Data

JeHyok Ryu and Uğur Çetintemel

Department of Computer Science, Brown University
Providence, RI 02912, USA
{jhryu,ugur}@cs.brown.edu

Abstract. In wireless mobile environments, we increasingly use data that depends on the location of mobile clients. However, requested geographical objects (GOs) do not exist in all areas with uniform distribution. More urbanized areas have greater population and greater GO density. Thus the results of queries may vary based on the perception of distance. We use urbanization as a criterion to analyze the density of GOs. We propose the Effective Distance (ED) measurement, which is not a physical distance but the perceived distance varying based on the extent of urbanization. We present the efficiency of supporting location-dependent data on GOs with proposed ED. We investigate several membership functions to establish this proposed ED based on the degree of urbanization. In our evaluation, we show that the z-shaped membership function can flexibly adjust the ED. Thus, we obtain improved performance to provide the location-dependent data because we can differentiate the ED for very densely clustered GOs in urbanized areas.

Key words: Location-Dependent Data, Pervasive Computing

1 Introduction

The current trend in workspace is a change from wired and stationary to wireless and mobile. The white paper from UMTS Forum argues that the trend has already happened and will continue [1]. Boosters of this trend are renovative portable computing devices such as smart phones or laptops, and enhanced wireless technologies such as PCS networks, GPS and wireless sensor nodes. These boosters are gradually extending the space of the applications of various wireless mobile workspace and sensing systems [2–6]. For instance, they allow mobile clients to access their desired data regardless of their place and time. However, the accessed data may become obsolete whenever mobile clients change their locations. This kind of data is called location-dependent data (LDD), which corresponds to the results of a query according to the location of a mobile client [7–12].

For example, let's consider a request: "Find a nearby police station." The result may be different according to the mobile client's location at which the query was issued. Moreover, "nearby" has fundamental uncertainty, so it is difficult to represent its value accurately. The degree of distance is felt differently in an urbanized region and its suburbs, because the closeness between geographical objects varies according to the degree of urbanization. Similar problems may appear for a shopper or a business traveller.

Urbanization, which means the growth in the proportion of a population living in urban areas, accelerates expansion of the infrastructure, such as educational facilities, accommodations, medical centers, roads, and offices. Therefore, mobile clients may get many more results in response to their queries in an urbanized area compared to what they would get in a suburb. As a result, when we consider many results of a query at the urbanized area, like mega-cities or urban agglomerations [13], query processing needs to be differentiated from that for a suburb environment. Since geographical objects are deployed differently according to the degree of urbanization, we need an appropriate evaluation of distance measurement in these queries. The distance may also be uncertain, like the "nearby" example in this section. In this paper, we compare and evaluate several membership functions (mfs) to find an appropriate data management model for query processing for location-dependent data.

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 shows several mfs quantifying a sense of urbanized distance for location-dependent mobile query processing. We present the analysis on applied membership functions applying proposed urbanized distance in Section 4. In Section 5, we evaluate several mfs of the Effective Distance for LDD requests and analyze performance changes by the variation of simulated systems and mfs parameters. Section 6 concludes the paper.

2 Related Work

We first define the terminology we use in the rest of the paper. A mobile query (MQ) is mobile client's queries. If we classify these MQs, they may be two classes. The first, the query of the example in Sect.1 is a location-dependent mobile query (LDMQ). The LDMQ is a query for fixed GOs (geographical objects) such as the police station mentioned in Sect.1. If the queried data are moving objects such as a car, we'll use the term moving object mobile query (MOMQ). In both of the two MQs, mobile clients can move or stay. In this paper, we focus on the LDMQ.

Most of the previous work in mobile environments has assumed that mobile clients access data in one region or one cell [14, 15]. But the possibility of different MQs has become increasingly considered. Madria et al., [8] introduced the use of concept hierarchies based on location data. These hierarchies define mappings among different granularities of locations. In this work, the concept hierarchies were used for the level of cities and states only. But a useful symbolic model for the location may include smaller hierarchies such as roads, streets, boulevards, and avenues. There is no obvious difference among the terms road, street, boulevard or avenue. Since all of these terms are commonly used in most cities, concept hierarchies do not distinguish between them in terms of precision but simply classify them as a group at a similar precision level in the hierarchy.

Dunham and Kumar [7] introduced an approach to manage LDD that has their intrinsic characteristics. Previous approaches of managing data of GOs in wireless mobile environments did not usually consider the relationship between data objects and their geographical distance. They defined a spatial replication, which may have different correct data at any point in time. They also introduced the concept of data region to explain their view of the data. A data region is different from a cell of wireless mobile environments and identifies data of GOs. They showed an implementation architecture for LDD in which a requested query is routed to the appropriate data region to service the LDD correctly. Ryu et al., [16] also evaluated the LDMQ in two kinds of data regions: geographical data regions (GDRs), which have an initial set of data objects for GOs at its own geographical area, and proposed LDD regions (LDRs).

Xu and Lee [17] investigated issues for querying wireless location-dependent data (WLDD). This work identified five query types based on access to location information. A cell ID is used for location binding. Local queries such as "List the local hotels," search the database starting from the current BS to the root until the results are found. In Non-Local cases, a query, e.g. "Find the weather in Cell #8," is redirected to the corresponding cell, and then follows the same method as local queries. Geographically Clustered and Geographically Dispersed queries have to be processed in a cluster of cells and in every cell, respectively. Likewise, a nearest query is processed in the current cell and the nearest cell to the farthest cell until the query is satisfied. In our work, LDMQs are assumed to follow the processing in the same manner as [7, 17, 16] in terms of how the queries are routed.

In [18], the authors proposed the method of dynamic attributes whose values change continually as a function of time. From the point of view of the queried objects, they investigated moving objects for MOMQs. But our managed data objects correspond to the geographical objects that are not moving, such as restaurants, hotels, and airports.

The relationship between the location granularity and the precision of results causes tradeoffs in LDMQ processing. Location models [11] were summarized in two models. They depend on the system's underlying location identification technique. Two models for representing locations are the geometric model and the symbolic model. The geometric model has compatibility across heterogeneous systems. However, it can be costly and complex. For example, the geometric model includes the latitude-longitude pair returned by GPS (Geographical Positioning Systems) or a set of coordinates defining the area's bounding geometric shape, such as a polygon. In the symbolic model, logical and real-world entities describe the location space. Entities can be buildings, streets, cities, or system-defined elements such as wireless cells, and are uniquely identifiable by a hierarchical naming system. This model typically has coarser location granularity than the geometric model because it stresses the representation of relationships between logical entities rather than their precise coordinates. And converting locations among heterogeneous systems is difficult. But discrete and well-structured symbolic location information is easier to manage. In terms of the GOs, the process of urbanization was already very advanced in more developed regions, where 75 percent of the population lived in urban areas in 2000 [13, 19]. This means that a number of GOs were already built in the urbanized areas. In these areas, the expansion rate of GOs is slower than the rates in other areas. In this context, to achieve efficient support for LDD in urbanized areas, it is necessary to incur the cost of LDD organization in advance. The LDD organization have been discussed in [16]. Since new GOs are not constructed frequently, frequent updates to LDD organization are not required. Thus, the degree of urbanization to which an area is urbanized should be considered together.

3 Quantifying Urbanized Sense of Distance

In processing the LDMQ, the location binding may be performed by two types, according to the expression of distance. The first type expresses explicit measured distance in a LDMQ. In the second type the LDMQs have uncertain distances. In the example of Sect.1,

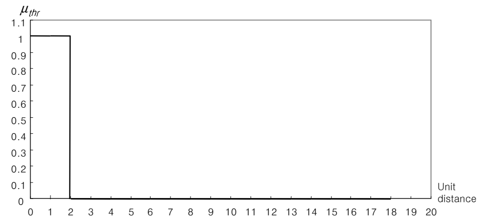

Fig. 1. Threshold mf for ED

we looked into the second type, using the "nearby" term as a representative example. The "nearby" distance in an urbanized area makes us feel or expect shorter distances to GOs than expected in a less urbanized suburb. Usually, this arises from the fact that a number of geographical objects are densely deployed in an urbanized area, such as many stores and facilities on a busy street. In this context, we define a new urbanized distance as follows:

Definition 1. *Effective Distance (ED) is a perceived distance that reflects a human's feeling according to the degree of urbanization.*

For example, we expect that the distance from a mobile client's current location to a hotel in urban agglomerations is probably very near. But we also expect that the distance in a rural area may be relatively far away. We'll call this distance ED. In an urbanized area, the binding location based on the ED is relatively narrower than the ED for the suburb. But, there is no exact boundary between an urban area and the suburb. This requires a method to represent the degree of urbanization for the ED. Thus, we consider mfs for sophisticated binding of the degree of ED because geographical objects are usually distributed more densely in urbanized areas. In the first case, the ED may be estimated by a threshold mf:

$$\mu_{thr}(x) = \begin{cases} 1, & 0 \leq x \leq a \\ 0, & a < x \end{cases} \tag{1}$$

If a given "near" corresponds to "a" for the ED in the mf μ_{thr} and the value of a is 2, the ED by the threshold mf can be shown as in Fig.1. The output value is either 1 or 0, depending on the criterion of ED for input value x. This threshold style may be applied in cell-based wireless networks. However, there is a shortcoming that definitely divides the boundary among the cells in a wireless region, though the boundary is indivisible. This is also the intrinsic limit of cell-based location management.

$$\mu_{tra}(x) = \begin{cases} 1, & 0 \leq x \leq a \\ \frac{s-x}{s-a}, & a < x < s \\ 0, & s \leq x \end{cases} \tag{2}$$

In the second consideration, a trapezoid mf may also be applied for the ED. When the value of 'a' and 's' is 2 and 18 in the mf μ_{tra}, respectively, the ED by the trapezoid mf can be presented as in Fig.2. The value of s may be set for the scale of a serviced area which is possible to request LDD. This trapezoid style makes up for the weak point in the threshold mf. But ED in trapezoid mf decreases regularly to the distance within the range of 'a' to

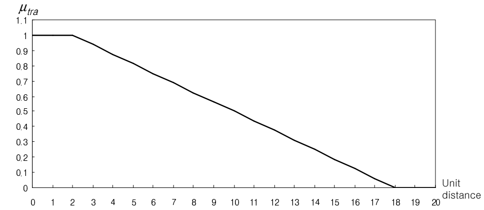

Fig. 2. Trapezoidal mf for ED

's' in μ_{tra}. Thus, it is difficult to differentiate the degree of ED for urbanization in which many geographical objects are densely deployed. Thus, the trapezoid style is not flexibly congruent with the concept of ED. Finally, we investigate the z-shaped mf:

$$\mu_z(x) = \begin{cases} 1, & 0 \leq x \leq a \\ \frac{1}{1+e^{b(x-c)}}, & a < x \end{cases} \tag{3}$$

If the given distance of 'near' is also 2 for the value of 'a' in mf μ_z, Fig.3 shows the z-shaped mf. Nearby locations after the distance 2 in Fig.3 have almost the same memberships but are slightly lower than the membership for the location of given distance 2 for 'a'. This may minutely represent that geographical objects in an urbanized area are deployed very close together. We can adjust the value of 'b' to the slope of this curve for the degree of ED, and 'c' extends the range of urbanized distance 'a' in the range from 0 to 1 the degree of membership. So, 'c' is the given 'a' plus a distance 'ext.' The distance ext can be adjusted if the scale of given urban area is extended. If we adjust the value of 'c' and 'b' in mf μ_z, it represents more subtly the membership degree of ED for clustered geographical objects located after given nearby distance 'a' in a more urbanized area. Thus, we expect that mf μ_z may resolve the problems in mf μ_{thr} and μ_{tra} and flexibly represent for the ED, especially the indefinite boundary between urbanized areas and suburbs.

For instance, the value 'a' may be considered as the radius of a data region. In μ_z, adjusting the values of 'c' and 'b' allows us to represent the degree of urbanization. We will consider the sensitivity between two variables to give us guidelines for our evaluation in the performance study.

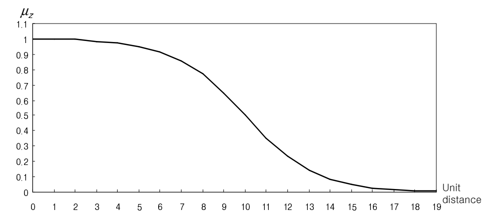

Fig. 3. z-shaped mf for ED

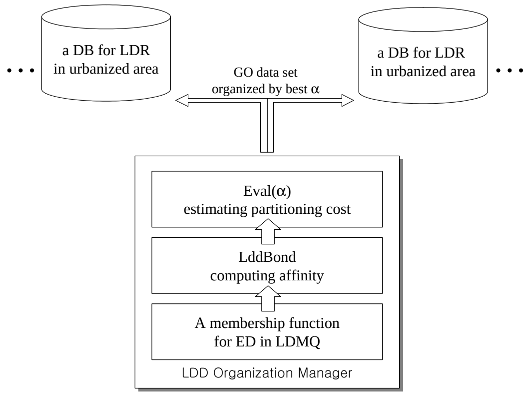

Fig. 4. Steps organizing LDRs

4 Effective Distance in Data Region

4.1 Data Region with ED

A model constructing Data Region for the LDD, proposed in [16], was called the Location-dependent Data Region (LDR). We use the data region to apply proposed ED (refer to [16] for details). To make the paper self-contained, however, we will briefly discuss the LDR with the ED. A data region manages its own data for GOs. A data region may be differently constructed according to the applied mobile computing environments. For instance, if a cell-based location management model is applied, a data region may consist of a data set for GOs covered by one base station (BS) or more BSs. This data region is not clear because the data regions depend on the evaluation of the distance for the mobile client's request on the LDD. We consider that the membership degree of the ED stated in Sect.3 is reflected in constructing the data regions. When we construct the data regions, we apply the membership degree of the ED on the physical distance as well as the access locality of the location of the requested LDD, and the location of the data. We summarize organizing LDRs adapted ED as follows. We assume the knowledge of the physical location of GOs. Let the entire service field be F. The F consists of a set of regions $R = \{r_1, r_2, \ldots, r_u\}$. Let us also suppose that $D = \{d_1, d_2, \ldots, d_n\}$ is the serviced data of GOs in F, and $Q = \{q_1, q_2, \ldots, q_t\}$ is the set of LDD requests. First, one of the mfs for ED from (1), (2), and (3) is used for the degree of distance between LDD. That is data in set D. Second, the frequency and locality between two LDD accessed together is considered. In other words, organizing LDRs reflects the frequency at each location from where queries are issued and serviced data are accessed by the same request. We use these relationships between two LDDs by an affinity, called *LDDBond*. Figure 4 illustrates the model organizing LDRs. With this organization, we exploit mfs for proposed ED. The LDD Organization Manager (LOM) warms up organizational steps with a given number of LDD requests. This manager may work in the Location Service Center (LSC) and be linked to Mobile Support Stations (MSSs). An LOM covers LDRs in the serviced regions. We assume that each LDR is linked to a database for the GOs in its region. The *LddBond* represents the possibility that two data objects may be grouped in the same data region. It reflects the frequency of accessed data objects d_i and d_j at the regions from which queries are issued, and the degree of the ED for the distance not only between d_i and d_j, but also from the queries issued locations to each accessed data objects.

4.2 Adjusting the ED with mf μ_z

In the mfs, the threshold mf is a baseline to be compared with trapezoidal and Z-shaped mfs. We can expect μ_{tra} to give a more appropriate degree of ED than that of μ_{thr}, but it is difficult to differentiate the degree of the very densely located GOs in urban areas. In terms of that, μ_z gives two parameters to adjust the ED according to the level of urbanization. When we apply the membership function μ_z, the sensitive data set of geographical objects affected by ED corresponds to the objects located in the urban area after the given near distance. This expectation is due to the reason that the ED for urbanization differentiates minutely the membership degrees of distance for the GOs, which are closely clustered together. Thus, we need to analyze how sensitive the changes of two variables in μ_z affect the degree of ED so that it also changes the performance.

Let μ_z be $f(x; b, c)$, where x is the given input distance to be evaluated for ED. We can find how sensitive the output of f is to a change in the input at different values 'b' and 'c' by evaluating the partial derivatives, $\partial f/\partial c$ and $\partial f/\partial b$:

$$\frac{\partial f}{\partial c} = \frac{\partial \frac{1}{1+e^{b(x-c)}}}{\partial c} = b(1+e^{b(x-c)})^{-2}e^{bx}e^{-bc} \tag{4}$$

$$\frac{\partial f}{\partial b} = \frac{\partial \frac{1}{1+e^{b(x-c)}}}{\partial b} = -(x-c)(1+e^{b(x-c)})^{-2}e^{b(x-c)} \tag{5}$$

To find the effect on outputs according to the changes of the variables, we take $|\partial f/\partial c|/|\partial f/\partial b|$, and have the result, $|b|/|c-x|$. We can recognize the sensitivity from $b/|c-x| < 1$ and $b/|c-x| > 1$ by c and b, respectively. Since b is greater than 0, we have the two ranges as follows:

$$c - b < x < c + b \quad if\ b > |c - x| \tag{6}$$

$$x < c - b,\ x > c + b \quad if\ b < |c - x| \tag{7}$$

From the above expressions, we can see that the degree of ED is more sensitive to a change in 'b' when 'b' is smaller than 'c', where the given distance 'x' for ED in a LDD request is less than $c-b$ or greater than $c+b$ from (7). As we can see from (6), the degree is more sensitive to a change in 'c' if the distance x to evaluate the degree of ED is $c - b < x < c + b$. In this paper, the region where $x < c - b$ (x, c, and $b \geq 0$) is an important target area to finely differentiate the degree of ED where so many GOs are densely located, due to urbanization. We discuss this issue further with simulation results in Sect.5

5 Evaluation

We simulate the access number of data region by LDD requests in its LDRs by the ED and GDRs by the physical distance. We measure the improvement from the difference

Table 1. Setting the parameters of the simulated environment with urbanization

Parameter	Description	Setting
KndOfGO	No. of kinds of serviced geographical objects	30
NoMC	No. of mobile clients	600
NoLDR	No. of LDRs	12×12
NoSvcGO	No. of serviced geographical objects	400
SKW	Skewness for the Zipf distribution	0.55 0.77, 0.99
MLQ	Distribution of mobile clients' locations issuing LDD requests	Zipf with *SKW*
NrScp	The scope of near sense (value 'a' in *mf*s)	2
Zb	The variable *b* of Z-shaped *mf* (4 values in *NrScp*<b<c)	2, 2.5, 3, 3.5
Zc	The variable *c* of Z-shaped *mf*' (2+*NrScp*)	4
DgrUrbn	Degree of urbanization	Zipf with *SKW*

of the total access number at LDRs and GDRs. From the improvement, we compare the performance of the ED using the three mfs in Sect.3, especially the z-shaped mf based on the analysis in Sect.4.2. We set the parameters of our simulated environment as shown in Table 1.

We apply random distributions used in [20] for MLQ and $DgrUrbn$ in Table 1. The distribution of mobile clients' locations issuing queries (MLQ) is modeled using a Zipf distribution, because we want to simulate LDD requests crowding on the urban area according to the anticipated urbanization. This means that the urbanized area corresponds to a small part of an entire service field. We simulate the degree of urbanization by a Zipf distribution of geographical objects in the serviced field ($DgrUrbn$). Thus, the higher skewness is (0.99 in Table 1), the smaller part of the serviced field is set to be concentrated most of all GOs, thereby simulating a very urbanized area. The $KndOfGO$ is the number of types of geographical objects (such as hotels, restaurants, shopping malls, bus stops, and police stations). We distribute the given geographical objects in a grid of 12 by 12 GDRs. We set the $NrScp$ for the given near sense by 2 distance units. The unit distance may be measured in meters, kilometers or miles according to the applied environments. We assume that the $NrScp$ as the half size of a data region. Each mobile client issues one request for LDD. An request accesses one or two LDD by mobile clients, and the number of issued requests per $KndOfGO$ is an average of 30 requests by the uniform distribution.

5.1 Simulation Results and Analysis

We apply three degrees of urbanization ($DgrUrbn$ in Table 1). As we described in Sect.4.2, the results of performance are mainly affected by quantifying the ED for urbanization. Especially, the $mf\mu_z$ makes effect on the results by variations of 'c' and 'b' with the degree of urbanization. Based on (6) and (7), the urban area where GOs are densely located is evaluated and affected by 'b' more than that by 'c.' So, we show the results in Fig.5, which gives more changes by the variation of 'b' in the ranges of $x < c - b$ in (7). For 'c', we fix the value to assume a urban area. If we would like to enlarge the size of the urban area, then we can adjust 'c' by 'ext' of Sect.3 correspondingly. We performed 600 LDD requests for each variation of 'b' for μ_z, and the other mfs based on three $DgrUrbn$. Figure 5 follows our analysis well in Sect.4.2. That is, the point of highest improvement in each 'b' of μ_z is moved to a lower value of 'b' close to the given $NrScp$ according to

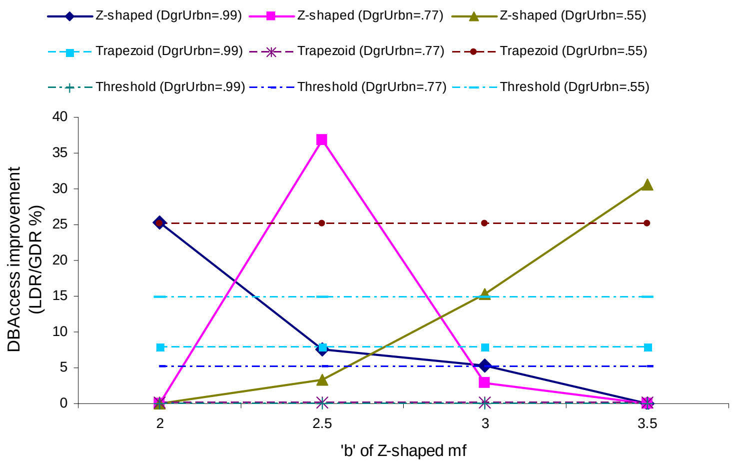

Fig. 5. Performance comparison by the ED in urban areas by three degree of urbanization

higher $DgrUrbn$. This means higher $DgrUrbn$ made densely located GOs and issued the LDD request, and the outputs are affected by 'b' when 'x' to be evaluated for ED is in $x < c - b$ of (7). So, in the case of the lower urbanized area and the range x from (6), the best improvement of results start to show in the range over $x < c - b$. In the results, μ_{thr} is a baseline to find minimum results. Furthermore, μ_{tra} relatively performed well but it does not flexibly represent the ED. So, from these results, we can recognize that μ_z differentiates the degree of ED well using flexible adjustments for different urbanized areas. It also effects the performance much more than the other mf.

6 Conclusions

Location-dependent queries may have results that are uncertain depending on the evaluation of distance. Due to the degree of urbanization, we may have a different sense of distance, called Effective Distance. We proposed a method to support LDD by the membership degree of Effective Distance based on urbanization. In our evaluation, we performed LDD requests based on LDRs by applying several membership functions for the Effective Distance from the degree of urbanization. Our experiments presented improved support for LDD and the analysis of the affects of the sensitivity of membership functions of ED on performance. The results reveal that a refined membership reflecting the characteristics of the environment, such as urbanization, introduces more possibilities of sophisticated pervasive computing services in next generation applied environments (such as advanced mobile environments and wireless sensor networks). In future work, we will investigate the sensitive relationship between urbanization and membership functions in more detail, and consider the pervasive applications with sensor nodes in the growth of urban areas.

References

1. Forum, U.: Mobile evolution shaping the future. White paper, UMTS Forum, http://www.umts-forum.org/ (2003)
2. Rao, B., Minakakis, L.: Evolution of mobile location-based services. Communications of the ACM **46**(12) (2003) 61–65
3. Imielinski, T., Badrinath, B.R.: Querying in highly mobile and distributed environments. In: Proceedings of International Conference on Very Large DataBase. (1992) 41–52
4. Woo, A., Madden, S., Govindan, R.: Networking support for query processing in sensor networks. Communications of the ACM **47** (2004) 47–52
5. Beaver, J., Sharaf, M.A., Labrinidis, A., Chrysanthis, P.K.: Location-aware routing for data aggregation in sensor networks. In: 1st Geo Sensor Networks Workshop. (2003) 1–18
6. Srivastava, M., Hansen, M., Burke, J., Parker, A., Reddy, S., Saurabh, G., Allman, M., Paxson, V., Estrin, D.: Wireless urban sensing systems. In No.65, C.T.R., ed.: University of California at Los Angeles. (2006) 199–210
7. Dunham, M.H., Kumar, V.: Location dependent data and its management in mobile databases. In: 9th International Workshop on Database and Expert Systems Applications. (1998) 414–419
8. Madria, S.K., Bhargava, B., Pitoura, E., Kumar, V.: Data organization issues for location-dependent queries in mobile computing. ADBIS-DASFAA, LNCS **1884** (2000) 142–156
9. Zheng, B., Lee, D.L.: Location-dependent queries in a multi-cell wireless environment. In: Proceedings of the 2nd ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE01). (2001) 54–65
10. Seydim, A.Y., Dunham, M.H., Kumar, V.: Location dependent query processing. In: MobiDE. (2001) 47–53
11. Lee, D.L., Xu, J., Zheng, B., Lee, W.C.: Data management in location-dependent information services. IEEE Pervasive computing (2002) 65–72
12. Gao, X., Hurson, A.R.: Location dependent query proxy. In: ACM Symposium on Applied Computing. (2005) 1120–1124
13. Division, U.N.P.: World urbanization prospects: The 2001 revision (2001)
14. Pitoura, E., Samaras, G.: Data Management for Mobile Computing. Kluwer Academic Publishers (1997)
15. Tan, K.L., Ooi, B.C.: Data Dissemination in Wireless Computing Environments. Kluwer Academic Publishers (2000)
16. Ryu, J., Song, M., Hwang, C.S.: Organizing the ldd in mobile environments. IEICE TRANS. INF. & SYST. **E86-D**(9) (2003) 1504–1512
17. Xu, J., Lee, D.L.: Querying location-dependent data in wireless cellular environment. In: WAP Forum/W3C Workshop on Position Dependent Information Services. (2000)
18. Wolfson, O., Xu, B., Chamberlain, S., Jiang, L.: Moving objects databases: Issues and solutions. In: Proceedings of International Conference on Scientific and Statistical Database Management (SSDBM '98). (1998) 111–122
19. Division, U.N.P.: Future world population growth to be concentrated in urban areas of world. Department of Economic and Social Affairs Press Release (http://www.un.org/esa/population/unpop.htm (2002)
20. Nascimento, M., Theodoridis, Y.: Benchmarking spatio-temporal databases: The gstd software. http://www.cti.gr/RD3/GSTD (1998)