Geo-Mobile Query-by-Sketch

David Caduff*
Department of Geography, University of Zurich – Irchel
Winterthurerstr. 190, CH-8057 Zurich, Switzerland
E-mail: caduff@geo.unizh.ch
*Corresponding author

Max J. Egenhofer
Department of Spatial Information Science and Engineering,
5711 Boardman Hall, University of Maine, Orono, ME, USA
E-mail: max@spatial.maine.edu

Abstract: The advent of wireless technology, such as cellphones, PDAs, tablet PCs, and sub-notebooks, allows transferring portions of traditional, desktop-based GIS technology to mobile environments. This paper introduces Geo-Mobile Query-by-Sketch, a sketch-based spatial querying system for mobile GIS environments that combines techniques for spatial querying with mobile technologies. The system implements an adaptive client-server architecture, which copes with restrictions of mobile environments, such as fluctuating bandwidth and frequent disconnections. The core concept analyzed is the *mobile sketch*, a multi-representation data structure of a sketched scene, which enables an adaptation strategy that is tailored to the available transmission rates. We analyze the transmission cost of Geo-Mobile Query-by-Sketch and develop a protocol that optimizes the adaptation level in order to guarantee quality of service.

Keywords: mobile and wireless GIS, spatial queries, sketch-based query languages, location-based services, spatial information retrieval in mobile environments

Reference to this paper should be made as follows: Caduff, D. and Egenhofer, M. J. (2006) 'Geo-Mobile Query-by-Sketch', *Int. J. of Web Engineering and Technology*.

Biographical notes: David Caduff received his B.S. degree in Geodesy and Geoinformatics from the Institute of Technology and Management, Muttenz, Switzerland, and his M.S. degree in Spatial Information Science and Engineering from the University of Maine, USA in 1998 and 2003, respectively. He is currently pursuing a Ph.D. in Geography at the University of Zurich, Switzerland. His research interests are in spatial cognition, location-aware services, and navigation systems.

Dr. Max J. Egenhofer (Ph.D. 1989, University of Maine; Dipl.-Ing. 1985, University of Stuttgart, Germany) is the director of the National Center for Geographic Information and Analysis (NCGIA) at the University of Maine, professor in Spatial Information Science and Engineering, and cooperating professor in Computer Science. He has published over 100 articles in refereed journals and conference proceedings relating to geographic information science, has co-organized over 20 international scientific meetings and has served on the program committees of over 100 conferences and workshops. He currently serves on four international academic advisory boards and sits on the editorial boards of five journals. Dr. Egenhofer is the recipient of the 2003 Researcher of the Year Award of the University Consortium for Geographic Information and Analysis.

1 INTRODUCTION

Handheld devices, such as cellphones, PDAs, tablet PCs and sub-notebooks, are becoming increasingly popular as tools for querying spatial data in mobile environments (Artem, 2000). They enable the combination of mobile appliances with wireless technology, which in turn allows designers of geographic information systems (GISs) to transfer parts of the core GIS technology from the desktop into the users' hands, leading to innovative spatial information appliances (Egenhofer and Kuhn, 1998). This transition is iterative, where some initial migrations for graphical presentation (Reichenbacher, 2004) and vector data transmission (Bertolotto and Egenhofer, 1999, Buttenfield, 2002) have been investigated already. This paper offers another perspective with an analysis of the implications of mobility on spatial information retrieval, particularly the design of tailored spatial query languages.

Retrieval of spatial information is one of the key operations of a GIS. While conventional query languages, such as SQL, use text-based statements, which work well within data domains where data can easily be stored in tables, these languages lack expressiveness and flexibility within more complex domains, such as images, maps, or other spatially related, multi-dimensional data (Egenhofer, 1992). Images and especially maps, however, are an integral part of most GISs and, therefore, query methods for GISs need to be sufficiently expressive and efficient.

Visual information retrieval systems have promoted novel techniques to query spatial data more intuitively (Egenhofer, 1996, Blaser, 2000a) as they stress the use of visual tools to formulate queries, thereby focusing more directly on the end result than the text-based SQL approach, because an example of a user's query is used as part of the formulation of a query statement. This paper studies visual information retrieval techniques for sketched queries (Egenhofer, 1996, Saund and Moran, 1994, Haarslev and Wessel, 1997) in the context of *mobile* GISs. We use the term *geo-mobile queries* for these types of queries. A sketch-based user interface for information retrieval allows a user to formulate a query as a hand-drawn sketch that represents the spatial scene he or she wants to find in a spatial database. The goal of this paper is to investigate the interplay between mobile clients, wireless networks, and static servers, and demonstrate the practicability of extending Spatial-Query-by-Sketch (Egenhofer, 1996) from a static to a mobile environment. This migration is called *Geo-Mobile Query-by-Sketch*.

For spatial queries in mobile environments, particular attention needs to be paid to query languages that are responsive to bandwidth fluctuations, frequent disconnections, and various constraints of the mobile devices, such as their limited input bandwidths. Since GIS applications on PDAs are unlikely to have all relevant data sources readily available in the device, *response time* is another critical aspect of information retrieval in mobile GIS environments.

Client-server architectures in mobile environments differ from the classical, wired architectures in many aspects (Satyanarayanan, 1996); therefore, this paper is concerned with identifying constraints of mobility that affect system *and* user behavior, so that properties of geo-mobile query systems can be defined. These properties are then applied to designing an adaptive client-server architecture that allows efficient geo-mobile querying of spatial data under varying conditions of the mobile environment.

Adaptation in a mobile client-server environment consists of three main steps: (1) resource monitoring, (2) an adaptation strategy, and (3) the adaptation process (Jing et al., 1999, Katz, 1994). Resource monitoring deals with identifying vital resource parameters for the application, while the adaptation strategy defines how these parameters influence adaptation for a specific system. Finally, the adaptation process controls the functionality of the application. Adaptation includes both the client and the server; therefore, we use for Geo-Mobile Query-by-Sketch a *mobile sketch*, which is derived from the sketched spatial scene and contains additional information about the client for guiding the adaptation process. The mobile sketch propagates the level of adaptation from the client to the server and contains a symbolic representation of the sketched scene, which is used for completion of the query process. We focus on the query formulation, because the resulting concepts and findings are so generic that they apply to a wide range of applications. The presentation and analysis of the results, on the other hand, are application-specific and, therefore, should be investigated separately.

The result of this investigation is a client-server architecture for geo-mobile querying that implements application adaptation in order to overcome the limited use of current technology in mobile environments. The investigation of the workflow and dataflow for sketch-based information retrieval systems in client-server architectures demonstrates the suitability of handheld devices as appropriate tools for performing geo-mobile queries. It is shown that adaptation to the mobile environment is necessary in order to ensure efficiency of geo-mobile queries.

The remainder of this paper is structured as follows: Section 2 discusses the client-server architecture of Geo-Mobile Query-by-Sketch. Its core internal representation—the *mobile sketch*—is introduced in Section 3. Based on the three different levels of representation of a mobile sketch an adaptation strategy is developed in Section 4. Section 5 analyzes the transmission cost of the mobile sketch's representation levels and develops a protocol to optimize the adaptation process in order to guarantee quality of service. Section 6 draws conclusions and indicates items for future research.

2 ARCHITECTURE FOR GEO-MOBILE QUERY-BY-SKETCH

The architecture of Geo-Mobile Query-by-Sketch is based on application-aware adaptation (Katz, 1994, Satyanarayanan et al., 1995) and the extended client-server model (Satyanarayanan, 1996, Jing et al., 1999, Rao, 2000). Since client and server share the responsibility of executing spatial queries, the adaptation logic—a function that uses quality-of-service reports to configure parameters of multimedia components (Ruiz et al., 2004)—resides on both client and server, such that the application on the mobile client is able to react to changes of the resources in the mobile environment. The request for adaptation is then propagated to the server in order to adjust the server's functionality.

2.1 Geo-Mobile Client

The client for Geo-Mobile Query-by-Sketch is the central part of the query system as it hosts—in addition to the user interface—the adaptation logic. The architecture of the mobile client consists of the operating system, which runs on top of the hardware, a middleware layer that acts as a mediator between system resources and applications, and the geo-mobile query processor.

The role of the operating system and the middleware layer is to monitor scarce resources and to respond to external events. The resource monitor keeps track of the resources, allocates the available resources among competing applications, and notifies the applications of changes to these resources. Complementarily, the role of the application is to adapt to changing conditions by using the information and resources provided by the resource monitor. The application reacts to the changes by switching to a different level of functionality so that it can guarantee best performance.

A change in resources affects both the user interface and the generation of the query statement. The user interface adapts to the change by enabling or disabling a specific set of functions for the current level of adaptation, while the query statement is issued in the form of a mobile sketch. The *mobile sketch* (Section 3) is a digital representation of the user input (i.e., the sketched scene) and reflects the effects of adaptation to available resources in terms of informative content and metadata of the sketched scene. Data transfers between client and server and requests for adaptation on the server are based on a transfer mechanism that utilizes the mobile sketch as a control protocol.

2.2 Geo-Mobile Server

The server of Geo-Mobile Query-by-Sketch, as the main component of the retrieval mechanism, is responsible for processing sketched queries against a spatial database. The role of the server in the adaptation process is passive, because the client monitors the mobile environment and decides about the extent of adaptation. For every query statement (i.e., mobile sketch) that gets transmitted, the server identifies the level of representation (Section 3.4) and executes the appropriate tasks.

The query process is partitioned in such a way that the steps on the client and the steps on the server are complementary, resulting in a digital sketch that can be processed against a database. For instance, a mobile client with poor resources would parse the user input, create the mobile sketch, and transmit it instantly to the server. The server resumes the querying process and generates the objects, creates the digital sketch, processes the query against the database, and finally prepares the result for presentation to the user.

Sketched queries typically generate a set of candidates ordered by best approximation (Egenhofer, 1996), which the server prepares for presentation to the user. The presentation of the result is based on a set of parameters of the mobile environment in order to guarantee efficient result browsing. Such parameters—including screen size of the mobile client, color and depth—are captured in the form of a user profile on the client and transmitted to the server.

3 MOBILE SKETCH

Spatial-Query-by-Sketch (Egenhofer, 1996) involves sketch parsing (x/y-coordinates of points and timestamp), object processing, digital sketch generation, and query processing (Figure 1). The result of the first three steps is a digital representation of a spatial scene consisting of spatial objects (points, lines, and regions) and the binary spatial relations among these objects. This digital representation of the sketched scene is the *digital sketch* (Blaser and Egenhofer, 2000), which is used to query a spatial database.

The focus of Geo-Mobile Query-by-Sketch is the tailoring of the digital sketch to a mobile environment so that one coordinates the workflow between client and server. This tailored representation is called the *mobile sketch.* The mobile sketch reflects important characteristics that are required to derive the digital sketch on the server, which is the final, meaningful representation of a sketch used for querying spatial databases.

3.1 Properties of the Mobile Sketch

The sketched scene is basically a collection of user-drawn spatial objects, their spatial relations, and some annotations. In order to investigate how objects and relations influence the configuration of the mobile sketch we need to know what properties characterize a sketched scene. The considerations in this section are based on a survey of people's sketching habits (Blaser, 2000b), which was conducted in order to establish an empirical basis for modelling and processing sketched spatial queries.

Sketch Complexity

Drawn objects are the primary elements in a sketch and are typically abstract and generalized representations of their real-world counterparts (Blaser, 2000b). Examples include such entities as buildings, road systems, and railroads. A sketched scene described by a collection of abstract objects may reflect any arbitrary level of complexity. The number of strokes per object increases if the sketch complexity increases, or if unusual objects are drawn; therefore, the level of complexity directly influences the size of the mobile sketch.

The summary of the sketched objects that appear in the sketching survey shows a multitude of object classes, ranging from common objects like buildings and road systems, to more complex objects, as for instance topographic features. The buildings and road systems contribute 53% of all objects contained in the sketch. The major four classes, which additionally include directions and symbols, contribute 72%. The remaining object classes are present in sketches only occasionally (Blaser, 2000b).

The object classes may be further categorized as either *Line Objects* or *Region Objects*, depending on their dimensionality. This categorization allows us to analyze the complexity of the sketch at the level of single line strokes. The survey shows that most line objects in a sketch are straight lines, and that curved lines or complex lines are drawn only occasionally. Similarly, most region objects are box-like or complex objects.

Quantitative Aspects

A sketched scene may reflect any arbitrary spatial configuration present in the users' mental map. As a result, it is difficult to establish an accurate estimation of the quantitative aspects of the sketched scene. Nevertheless, the quantitative aspects of the sketch may be characterized based on two essential aspects: the *number of objects per sketch*, and the *number of binary relations* present in the sketch.

The number of objects per sketch is an important quantitative aspect of a sketched scene. Objects are instances or logical entities in a sketch and substantially define the semantic of the sketch. The typical number of objects per sketch is 14 and the standard deviation is 3.3, which is 23% of all the objects in the sketch (Blaser, 2000b).

The similarity assessment between two sketched scenes is based on an association graph consisting of a set of binary relations (Rodríguez et al., 2003), which establish the relations among the single objects in the sketch, and hence, captures the properties of the spatial configuration. The number of binary relations contained in the set influences the size of the mobile sketch substantially. The worst-case scenario includes all possible binary relations (Equation 1), which causes the association graph to grow by $O(n^2)$, with n being the number of objects contained in the sketch.

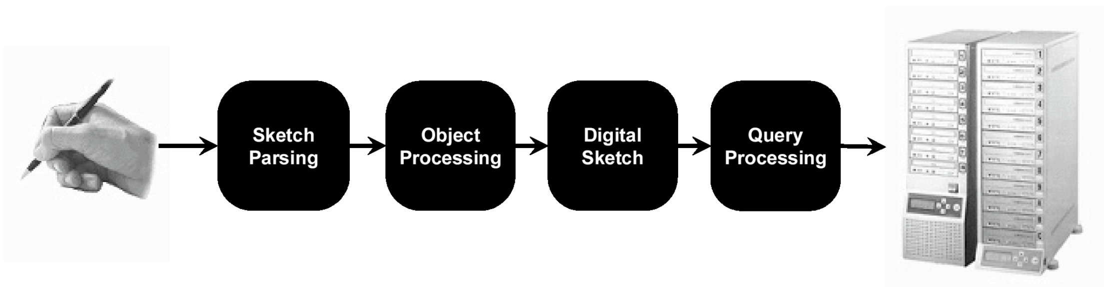
$$m = \frac{n!}{2*(n-2)!} = \frac{n*(n-1)}{2} \qquad (1)$$

In a best-case scenario, the association graph consists of a subset of binary relations. In this case, the number of binary relations in the association graph is linearly dependent on the number of objects contained in the sketch. Consequently, the association graph grows by $O(n)$.

3.2 Components of the Mobile Sketch

The mobile sketch is an abstraction of the digital sketch that efficiently captures the vital properties of a sketch at different levels of object representation. The main objective of this approach is to reduce the amount of data transmitted between client and server. The structure of the mobile sketch consists basically of four modules.

Figure 1 The functional objects involved in the Spatial-Query-by-Sketch process



Module 1—Mobile Sketch Signature*:* A mobile sketch begins with a signature containing metadata of the sketch, such as its creation time, the document size, and its history. The most important information conveyed by the mobile sketch signature, however, is the level of representation of the mobile sketch, which enables the server to perform the remaining steps to create the digital sketch.

Module 2—Hardware Profile*:* This profile contains system-specific information describing the properties of the mobile client. For instance, mobile clients may differ in terms of screen size, color depth, and resolution. This information is crucial for effective presentation of the query results on the mobile client.

Module 3—User Profile*:* The user profile contains preferences that are set by the user on the mobile client. It includes the selected association graph model for the digital sketch, weights for binary spatial relations, and thresholds for result presentation.

Module 4—Data Module*:* The data module reflects the sketch that the user draws on the touch-sensitive screen. Unlike the other three modules, which remain essentially the same for all adaptation levels, the content of the data module may change depending on the chosen level of representation of the mobile sketch.

The purpose of the four modules of the mobile sketch is to enable the server to complete the generation of the digital sketch, support the query process, and supply parameters for a valuable presentation of the results.

3.3 Generating a Mobile Sketch

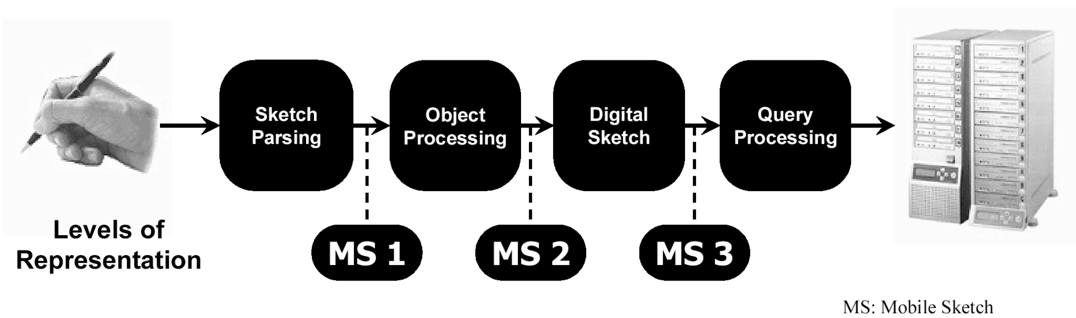
The scheme used to generate the mobile sketch may be described as a lossy compression technique applied on the digital sketch. Lossy compression techniques involve a compression such that if expanded less information may be available then what was in the original (Sayood, 2000). In return, such techniques generally obtain much higher compression ratios than what is possible with lossless compression. Unlike most compression techniques, however, no decomposition and no compression algorithms are required to create multiple representations and reduce the amount of data of the mobile sketch. The degree of complexity (i.e., the actual object representation) of the mobile sketch is directly dependent on the number of steps involved in the creation of the sketch. This approach reduces not only the amount of data, but also the required infrastructure to create the mobile sketch. Therefore, it is the ideal approach for an adaptive application on a mobile client.

The sketch generation scheme is based on the functional objects of the query process: sketch parsing, object processing, and digital sketch generation. Each of the three functional objects generates a mobile sketch at a different computational level (Figure 2). The mobile client selects an appropriate level of adaptation based on the available resources in the mobile environment, that is, the application running on the handheld device determines what steps are ideally performed on the mobile client given the set of parameters for the actual mobile environment. Accordingly, the selected level of adaptation defines the level of representation of the mobile sketch, since only the selected steps of the query process are executed. Such a mobile sketch is then sent to the server, where the generation of the mobile sketch is completed. Subsequently, the mobile sketch is converted into a digital sketch that can be used for the query against the database.

3.4 Levels of Representation

The mobile sketch produced by the generation scheme may reflect any of the three different levels of representation. The term multiple representations in GISs refers to changes in the geometric and topological structure of a digital object that may occur with under varying resolutions at which that object is encoded for computer storage, analysis and depiction (Buttenfield, 1993). Accordingly, the term *level of representation* in the context of Geo-Mobile Query-by-Sketch does not refer to spatial details, since the sketched scene is the same for the mobile sketch and the digital sketch. Instead, level of representation refers to the following three qualitative and quantitative aspects of the mobile sketch, termed *MS 1–MS 3*.

Figure 2 The generation scheme for multiple representations of the mobile sketch



***MS 1—Simplified Line Strokes**:* Simplified line strokes are the lowest representation possible in a mobile sketch. They consist of connected, time-stamped points (i.e., x- and y-coordinates and the corresponding creation time) as drawn by the user on the touch-sensitive user interface. Consequently, this level of representation requires the least powerful infrastructure and produces the smallest amount of data.

***MS 2—Geographic Objects**:* This representation reflects the user input in terms of geographic objects (i.e., lines and regions), ASCII text, or symbols. The geographic objects consist of interconnected line segments and a set of properties; however, this level of representation contains no information about spatial relations among the sketched objects. The significant difference compared to the lowest representation level is the higher degree of computing resources required to process the user input in order to generate the objects and their properties.

***MS 3—Digital Sketch**:* This level of representation corresponds to the digital sketch as used for the query processing. It consists of a set of distinguishable sketched objects and the corresponding binary spatial relations between them. The attributes and properties of a single object are the same at *MS 3* as at *MS 2*. The generation of the association graph of binary spatial relations and the assessment of topological, metric, and direction attributes requires appropriate CPU and memory resources; therefore, it is executed only if these resources are available. At this level, the amount of data increases drastically with the addition of the binary relations.

The mobile sketch facilitates an adaptive geo-mobile system architecture that guarantees an appropriate level of workload on the client as well as an amount of data that corresponds to the available network bandwidth. In order to achieve such adaptation to the mobile environment, a strategy is needed that guides the adaptation.

4 MOBILITY-AWARE ADAPTATION STRATEGY

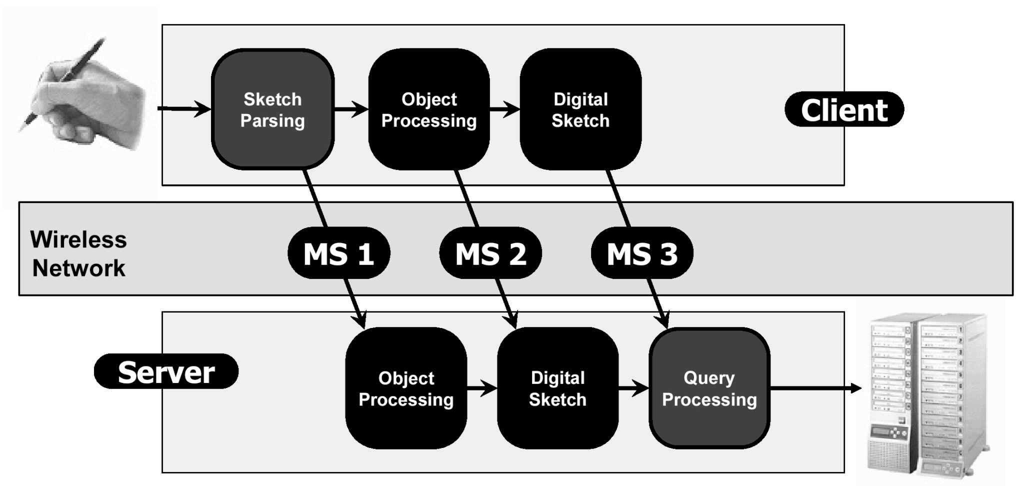
Mobility-aware adaptation involves dynamic partitioning of the functionality between mobile host and server. By varying the partition of duties, however, we also vary the functionality of the user interface and above all, the quality of data produced on the mobile host. Consequently, adaptation involves the trading of data quality and user experience for resource consumption. The architecture for Geo-Mobile Query-by-Sketch (Section 2) captures this notion of data degradation through the three different levels of representation of the data produced on the mobile client (i.e., the mobile sketch).

4.1 Complementary Distribution of Query Processing

The representation levels define the degree to which data delivered to the server require further processing. The lowest representation level results from scarce resources on the mobile client, requiring in turn full server support. The second level of representation applies when processing power, memory, and energy supply are abundant, but wireless communication with the server is not reliable. The application on the mobile client produces the highest level of representation if the handheld device *and* the wireless network provide sufficient resources. Figure 3 illustrates the distributed query process and the three possible scenarios.

The choice of an appropriate adaptation level is based on predefined policies. Adaptation policies capture different system behaviour in a flexible and customizable manner. The policies govern a discrete adaptation algorithm.

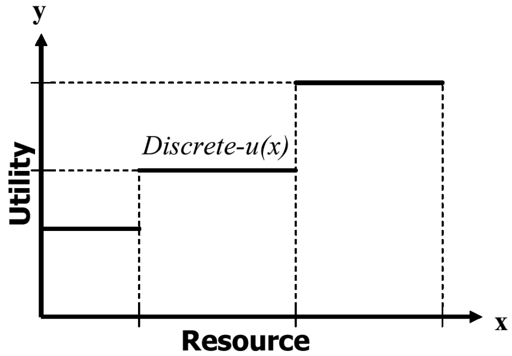
Figure 3 The mobility-aware adaptation strategy: the functional objects are located on client and server and generate mobile sketches at different levels of resolution



4.2 Adaptation Policy

In order to adapt to a changing environment, a system must evaluate its present situation and try to change the situation to another configuration that guarantees acceptable performance; therefore, the role of the adaptation policy is central to capturing the application-specific responses to the available resources. Since it is difficult to obtain an analytic expression that takes into account all possible parameters (e.g., CPU cycles, memory I/O operations, jitter), a profile-based modeling is used to approximate the mapping of the available resources onto the application functionality.

Figure 4 Example of a discrete utility curve



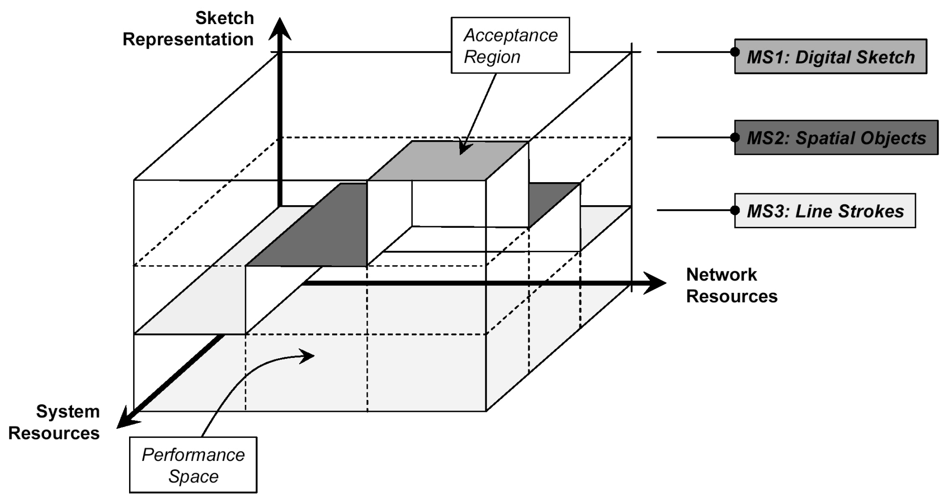
A profile-based, discrete adaptation policy allows applications to move up along step or staircase-shaped utility curves, rounding off the assigned parameters to the lower discrete parameter value (Figure 4). This discrete adaptation considers the portion of a step to determine a discrete representation for the mobile sketch, requiring complete increments of single steps to support multiple representations.

The user profile defines the relation between the single steps of the utility curve, that is, it defines the thresholds that guide the transition from one step of the utility curve to the next. We utilize a compression factor (i.e., low, middle, high) to divide the resource into three parts, each representing a representation level for the mobile sketch. The compression factor divides the resource in a 1:2:3 (i.e., low to middle to high) ratio. The application adapts the thresholds for the steps every time the user changes the compression factor.

We define two spaces that reflect the available resources with respect to the multiple representations: (1) a *resource space* R and (2) a *performance space* P. The resource space R is dimensioned by resource characteristics in the mobile environment, which define the operational spectrum of the application. We utilize a two-dimensional resource space that includes system characteristics (i.e., CPU power and memory) and network properties (i.e., bandwidth and latency). Each axis reflects the resources assigned to the application by the middleware. Performance space P is dimensioned along user-oriented parameters and includes three *acceptance regions* (i.e., representations of the mobile sketch). The acceptance region for a specific representation AR of P is defined as the region in which the application is considered to be working properly with the current parameters. The adaptation model is illustrated in Figure 5.

The application adapts its functionality every time it detects changes in the parameters received from the resource monitor. In addition, three protocols guide the adaptation in case of abrupt changes of the wireless connection (i.e., a disconnected protocol, a weak connection protocol, and a connected protocol). The disconnected protocol redirects the user input into a file that is stored locally. As soon as the connection is reestablished, the connection protocol or weak connection protocol checks for such files and prepares them for transmission to the server.

Figure 5 The discrete adaptation model: available system and network resources define the type of mobile sketch that is generated



5 ANALYSIS OF TRANSMISSION COSTS

The amount of information contained in the different representations of the mobile sketch increases from representation level to representation level. In this section, we examine how the file size of the mobile sketch changes as a result of the representation level. We also investigate how this size of the sketch is related to the transmission speed, which is crucial for an efficient communication in network-based systems. The framework for studying the performance of the mobile sketch consists of two components: (1) one that deals with the properties of the sketched scene and establishes the foundation for the estimation of the size of the mobile sketch and (2) one that deals with the transmission cost model, which explains assumptions related to the network and the transmission mechanism.

5.1 Transmission Cost Model

For Geo-Mobile Query-by-Sketch we define transmission cost as the time that is required to send a sketch from a mobile client to the server over a wireless link. The wireless link is the intrinsic bottleneck of mobile client-server computing. In contrast, the time required to pre-process sketches on the mobile client and the time required for completion of the query on the server will converge due to advancements in mobile computing technology (Pitoura and Bhargava, 1994, Satyanaranayanan, 1996).

Transmission cost may also be described as *message latency*, that is, the sum of transmission time and propagation latency. Transmission time is the time required to get all bits accepted by the receiver. The transmission time for a mobile sketch is dependent on the available bandwidth, network protocols, and the quality of the link. For instance, the message latency in packet data networks is the sum of statistical multiplexing, serialization delay, and forwarding delay.

The second component, propagation latency, is the time that one bit needs to reach its destination. It depends on the distance between sender and receiver, as well as the communication medium. We focus on the amount of data and bandwidth, and assume that transmission protocols take care of results of queuing effects within the network, which affect message latency (e.g., loss, jitter). Hence, we assume that the transmission cost is equal to the transmission time and that the transmission time is a linear function of the data size and the bandwidth.

The purpose of the transmission cost analysis is to better understand how mobile sketches at different levels of representation influence the performance of geo-mobile querying. Mobile sketches basically consist of two separate parts: a *static part* and a *dynamic part*.

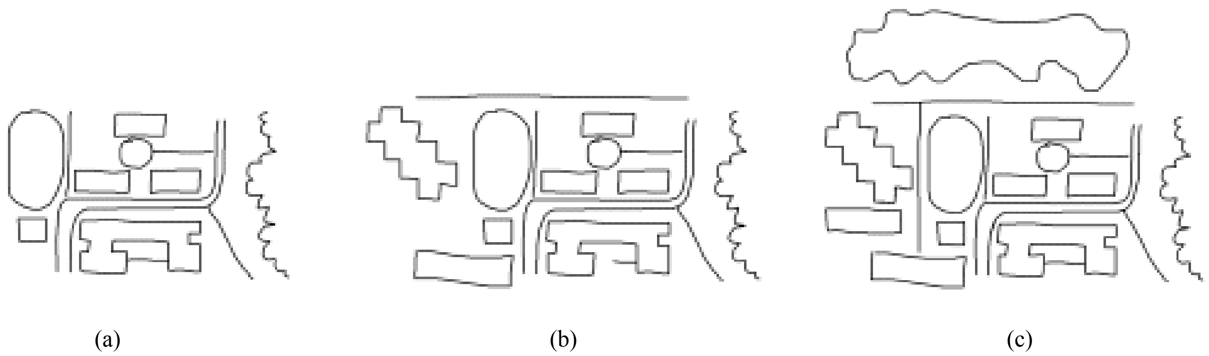
The static part is constant in size as it describes the overhead of the mobile sketch. It contains the sketch signature and information pertaining to the user profile and hardware characteristics. The dynamic part contains the data that describe the sketched scene. Sketched scenes vary in content depending on the task at hand, with the result that the number of objects, and hence the resulting file size, varies from sketch to sketch.

The base of the assessment is a set of three sketches that represent various levels of complexity and quantitative characteristics. The three sketches incorporate typical sketch characteristics. The first sketch represents a simple spatial configuration enclosing only a few elements (Figure 6a). The second sketch contains the number of elements of an average sketch (Figure 6b), whereas the third sketch contains a large number of elements (Figure 6c).

At the lowest level of representation (*MS 1*), a mobile sketch consists essentially of a set of individual line strokes. Therefore, the focus for the configuration of the mobile sketch at this level is on the total number of strokes per sketch.

The second level of representation (*MS 2*) contains additional information about objects created from the original strokes. Hence, the criteria for drawing the three sketches at this level are the total number of objects and the type of the objects (i.e., lines and regions showing various levels of complexity).

Figure 6 The three sketches that build the base of the assessment: (a) with only a few elements, (b) with an average number of elements, and (c) with a large number of spatial elements.



The highest level of representation (*MS 3*) contains the association graph of binary relations among the objects, which conveys implicit information about the sketch configuration. For the purpose of this analysis, we investigate two cases of the association graph setup. First, we assess the resulting file size in a best-case scenario, which means that the association graph is reduced to its minimum and, therefore, yields the smallest file size. The second case represents the worst-case scenario, which generates an association graph that includes all possible binary relations among the objects. Table 1 summarizes the quantitative properties of the three sketches at the three representation levels.

Table 1 Quantitative characteristics of the three sketches' representation levels.

			Sketch Nr. 1	2	3
MS1:	**Strokes**	per Object	3	6	9
		per Sketch	45	90	135
MS2:	**Objects**	Total	12	15	18
	Line	Straight	3	4	5
		Curved	1	1	1
		Complex	1	1	1
	Region	Box	3	4	5
		Complex	1	2	3
		Square	1	1	1
		Circle	1	1	1
		Oval	1	1	1
MS3:	**Binary Relations**	Best Case	12	15	18
		Worst Cae	66	105	153

The results of the analysis are summarized in Figure 7. Figure 7a shows the resulting file sizes of the three sketch configurations with respect to the level of representation of the mobile sketch. In addition to the three original sketches, it shows the average file size for each level of representation. Figure 7b shows the transmission time for all four cases as a function of file size and transmission rate.

5.2 Quantitative Characteristics

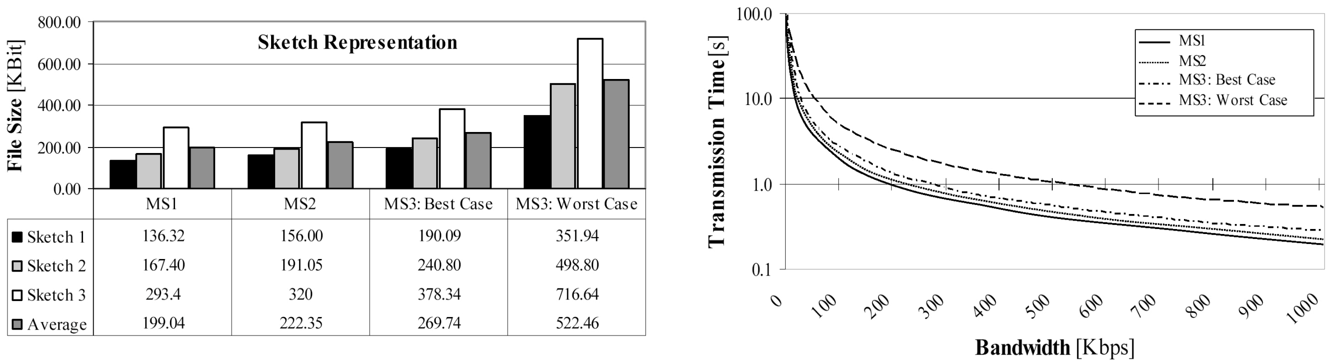
The file size of a mobile sketch is a function of several factors (e.g., representation level, sketch composition, user profile). The association graph, however, influences the amount of data contained in a mobile sketch substantially, and may be considered the main factor for the resulting file size.

The file size increases gradually between the mobile sketch containing the simplified line strokes, the sketch containing the geographic objects (i.e., about 9 to 12%), and the sketch containing a subset of the association graph (i.e., around 19% to 22% compared to the sketch containing the geographic objects). Conversely, the file size increases significantly (i.e., approximately 100%) between the mobile sketch containing a reduced set of binary relations and the sketch containing the full association graph. Figure 7a shows the file size for all three sketches as a function of the representation level, which illustrates this behaviour. Consequently, the selection of the appropriate association graph is crucial for the resulting file size, and needs to be considered accordingly.

5.3 Transmission Time in Context

In order to understand the influence of the file size and the resulting transmission time on the query process, and hence on the end-user, we need to put these numbers into the context of the currently available wireless infrastructure. Today's wireless infrastructure can generally be categorized as Wireless Private Area Networks (WPANs), Wireless Local Area Networks (WLANs), and Wireless Wide Area Networks (WWANs).

Figure 7 Results of the transmission cost analysis: (a) file sizes and the average for the mobile sketches used in the analysis, and (b) transmission times as a function of file size and data rate



(a) (b)

Geo-mobile query systems may theoretically be implemented in networks in any of these categories. For our analysis, however, we focus on WWANs only, because, the available bandwidth in WPANs (e.g., Bluetooth with a data rate of 1 Mbps) and WLANs (e.g., IEEE 802.11 offers 2 Mbps, IEEE 802.11b up to 11 Mbps) is far larger than the file size of mobile sketches. The transmission time in these types of wireless networks shrinks to a fraction of a second, and hence, becomes irrelevant to the end-user.

The available bandwidth in WWANs, however, is an issue for geo-mobile queries since it is rather limited. For instance, in the United States the current predominant wireless protocol, High Speed Circuit Switched Data (HSCSD), operates with bearers that offer a bandwidth of 9.6 Kbps. In the ideal case, four of these bearers are bundled together to form a 39.6Kbps channel for communication. The Global System for Mobile Communication (GSM) offers approximately the same bandwidth as HSCSD, while the bandwidth for the General Packet Radio Service (GPRS) is slightly better, depending on the number of available bearers (i.e., 8 bearers a 21.4Kbps). As a result, the transmission of mobile sketches using HSCSD, GSM, or GPRS may result in long waiting periods for the end-user.

For instance, the average file size of the three sketches for the different levels of representation ranges from 200KBit up to 520KBit. Given the transmission rates for WWANs range from 9.6Kbps (GSM with one bearer) to 171.2Kbps (GPRS with 8 bearers a 21.4Kbps), we obtain transmission times ranging from 10 seconds to 55 seconds for GSM, and from 1.2 second to 3.1 seconds for GPRS. While the first example (i.e., GSM) shows clearly that file size is an issue, the transmission times for the second example (i.e., GPRS) seem to satisfy our needs. However, these numbers are purely theoretical since they do not include any delays deriving from queuing effects within the network. Hence, even with GPRS, file size is an important issue when transmitting mobile sketches in wireless networks.

These two examples illustrate the variation of transmission rates present in WWANs. Wireless technology for WWANs will become more sophisticated and reliable in the future, offering similar bandwidth as today's WPANs and WLANs. A first example that testifies the impressive growth in bandwidth for WWANs is the Universal Mobile Telecommunications System (UMTS), which is not largely implemented yet, but promises bandwidths of 1Mbps and more. Wireless WANs, however, are influenced by many factors, which affect their stability. As a result, applications that rely on WWAN infrastructure need to ensure quality of service despite fluctuating bandwidth.

5.4 Optimizing the Geo-Mobile Adaptation Process

The geo-mobile adaptation process is static, since it requires the user to define the representation level that triggers the adaptation. Based on the insight about file size characteristics and transmission time of a mobile sketch, we investigate how to optimize this process and make the adaptation dynamic.

The cost of communication for geo-mobile queries is proportional to the size of data that have to be transmitted; therefore, the amount of elements contained in the mobile sketch directly influences the communication cost. The findings in the previous sections provide the base for an accurate estimate of the file size for all levels of representation. Based on this estimate, we can predict the communication cost (i.e., the transmission time for each level of representation). This estimate, together with a quality of service parameter in terms of the desired response time, provides the system with the parameters necessary to adapt the level of representation to the given infrastructure.

The increase in file size behaves approximately the same for all three sketches; therefore, we use the difference in file size to calculate a file size factor for each level of representation, relatively to the lowest level of representation, which is accurate enough to guide the adaptation process.

If we multiply this factor with the file size for the line strokes (i.e., lowest representation level), which we know immediately after the line strokes have been drawn, we obtain an efficient estimate of the file size. The dynamic adaptation process is based on the estimation of the file size and the desired response time, as defined by the user.

The estimation of the file size allows us to create a table containing response times for the different representations. If we scan this table and eliminate all values that do not satisfy the desired response time, we find the best representation level for the given parameters.

Table 2 Lookup table of expected transfer times for dynamic adaptation. The gray area represents response times that exceed the QoS parameter defined by the user, which is set to 10 seconds in the example.

	Transfer Time for Wireless WANs [sec]				
Bandwidth [Kbps]	MS1	MS2	MS3 Best Case	MS3 Worst Case	Max. Time Difference
2	99.5	111.2	134.9	261.2	161.7
5	39.8	44.5	53.9	104.5	64.7
10	19.9	22.2	27.0	52.2	32.3
20	10.0	11.1	13.5	26.1	16.2
30	6.6	7.4	9.0	17.4	10.8
40	5.0	5.6	6.7	13.1	8.1
50	4.0	4.4	5.4	10.4	6.5
60	3.3	3.7	4.5	8.7	5.4
70	2.8	3.2	3.9	7.5	4.6
80	2.5	2.8	3.4	6.5	4.0
90	2.2	2.5	3.0	5.8	3.6
100	2.0	2.2	2.7	5.2	3.2
120	1.7	1.9	2.2	4.4	2.7
150	1.3	1.5	1.8	3.5	2.2
200	1.0	1.1	1.3	2.6	1.6
300	0.7	0.7	0.9	1.7	1.1
500	0.4	0.4	0.5	1.0	0.6
1000	0.2	0.2	0.3	0.5	0.3
1500	0.1	0.1	0.2	0.3	0.2
2000	0.1	0.1	0.1	0.3	0.2

Consider the average file size of the three sketches as an example. The file size for the simplified line strokes allows us to estimate the file size for all representations. In the next step, the system calculates the response time of each representation level for a range of transmission rates, as listed in Table 2. The desired response time acts as a filter to eliminate occurrences that do not satisfy the criteria. Based on the resulting table, the system can dynamically choose the best level of representation for the available bandwidth.

6 CONCLUSIONS AND FUTURE WORK

This paper introduced Geo-Mobile Query-by-Sketch and investigated the implications of mobility on sketch-based information retrieval systems. Important issues in this context include challenges related to mobile technology, wireless communication, and the architectural structure of information retrieval systems in such environments. The main objective of this paper was to identify the major factors that influence the extension of sketch-based query techniques from static to dynamic mobile environments. The paper first identified characteristics of mobile computing and the challenges involved therein and then created an architectural framework that permits efficient sketch-based querying in mobile GIS environments.

The transmission cost analysis shows that for the transmission of a typical sketch over today's typical WWANs, the lowest level of representation of a mobile sketch yields best transmission times. The lowest level of representation yields file sizes that are approximately 60% less than the file size of the digital sketch containing all binary relations. The analysis also shows that the transmission times are good enough to guarantee response times that keep a user's attention (i.e., shorter than 10 seconds).

Geo-Mobile Query-by-Sketch introduces a new adaptive approach for geo-mobile querying about which many variations and extensions remain to be explored. One important aspect of geo-mobile querying is context. Three types of context that need to be addressed by future research are spatial context dependency, context dependency of application adaptation, and task-based context dependency.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation under grant number IIS-9613646 and the National Geospatial-Intelligence Agency under grant number NMA201-01-1-2003. Max Egenhofer's work has been further supported by NGA grant NMA401-02-1-2009 and by the National Science Foundation under grant numbers EPS-9983432 and IIS-9970123. We are grateful for the feedback from reviewers of and participants at the *Fifth International Workshop on Web and Wireless Geographical Information Systems*, where an earlier version of this paper (Caduff and Egenhofer 2005) was presented.

REFERENCES

Artem, G. (2000) 'Management of Geographic Information in Mobile Environments', Department of Computer Science and Information Systems, University of Jyväskylä, Jyväskylä, Finland.

Balasundaram, V. and Kennedy, K. (1989) 'A techniques for summarising data access and its use in parallelism enhancing transformations', *Proceedings of ACM SIGPLAN'89 Conference on Programming Language Design and Implementation*, Portland, OR, pp. 41-53.

Balley, S., Parent, C., Spaccapietra, S. (2004) 'Modeling Geographic Data with Multiple Representations', *International Journal of Geographical Information Science*. **18**, 327-352.

Bertolotto, M. and Egenhofer, M. (1999) 'Progressive Vector Transmission', In C. Bauzer Medeiros (Ed.): *ACM-GIS '99, Proceedings of the 7th International Symposium on Advances in Geographic Information Systems*, Kansas City, MO, pp. 152-157.

Bircsak, J., Craig, P., Crowell, R., Harris, J., Nelson, C.A. and Offner, C. (2000) 'Extending openMP for NUMA machines: the language', *Workshop on OpenMP Applications and Tools (WOMPAT)*, San Diego, July.

Blaser, A. (2000a) 'Sketching Spatial Queries', Ph.D. Thesis, University of Maine, Orono.

Blaser, A. (2000b) 'A Study of People's Sketching Habits in GIS', Spatial Cognition and *Computation*. **2**, 393-419. .

Blaser, A. and Egenhofer, M. (2000) 'A Visual Tool for Querying Geographic Databases', In Di Gesù, V., Levialdi, S., and Tarantino, L. (Eds.): *Advanced Visual Interfaces (AVI'2000)*, Palermo, Italy, pp. 211-216.

Buttenfield, B. (1993) 'NCGIA Research Initiative 3—Multiple Representations', Technical Report, National Center for Geographic Information and Analysis, Buffalo.

Buttenfield, B. (2002) 'Transmitting Vector Geospatial Data across the Internet', In M. Egenhofer and D. Mark (Eds.): *Geographic Information Science, Second International Conference,* Boulder, CO, *GIScience 2002*, Lecture Notes in Computer Science **2479**, pp. 51-64.

Caduff, D. and Egenhofer, M. (2005) 'Geo-Mobile Queries: Sketch-Based Queries in Mobile GIS-Environments', In K.-J. Li and C. Vangenot (Eds.): *Web and Wireless Geographical Information Systems, 5th International Workshop*, Lausanne, Switzerland, Lecture Notes in Computer Science **3833**, pp. 143-154.

Egenhofer, M. (1992) 'Why not SQL!', *International Journal of Geographical Information Systems,* **6,** 71-85.

Egenhofer, M. (1996) 'Spatial-Query-by-Sketch', In *VL '96: IEEE Symposium on Visual Languages*, Boulder, CO, pp. 60-67.

Egenhofer, M. and Kuhn, W. (1998) 'Beyond Desktop GIS', In *GISPlaNet*, Lisbon, Portugal (http://www.spatial.maine.edu/~max/BeyondDesktopGIS.pdf).

Haarslev, V. and Wessel, M. (1997) 'Querying GIS with Animated Spatial Sketches', In *IEEE Symposium on Visual Languages*, Capri, Italy, pp. 197-204.

Jing, J., Helal, A. and Elmagarmid, A. (1999) 'Client-Server Computing in Mobile Environments', *ACM Computing Surveys,* **31,** 117-157.

Katz, R. (1994) 'Adaptation and Mobility in Wireless Information Systems', *IEEE Personal Communications,* **1,** 6-17.

Pitoura, E. and Bhargava, B. (1994) 'Building Information Systems for Mobile Environments', In *Proceedings of the Third International Conference on Information and Knowledge Management*, Gaithersburg, MD, pp. 371 - 378.

Rao, S. (2000) 'An Architecture for Adaptable Wireless Networks', Master's Thesis, Massachusetts Institute of Technology, Boston, MA.

Reichenbacher, T. (2004), 'Mobile Cartography—Adaptive Visualization of Geographic Information on Mobile Devices', Ph.D. thesis, Technical University Munich, Germany.

Rodríguez, A., Egenhofer, M., and Blaser, A. (2003) ‘Query Pre-Processing of Topological Constraints: Comparing a Composition-Based with a Neighborhood-Based Approach’, In: T. Hadzilacos, Y. Manolopoulos, J. Roddick, and Y. Theodoridis (Eds.): *Advances in Spatial and Temporal Databases: 8th International Symposium, SSTD 2003, Santorini Island, Greece,* Lecture Notes in Computer Science **2750**, pp. 362-379*,*

Ruiz, P., Botía, J., and Gómez-Skarmeta, A. (2004) ‘Seamless Multimedia Communications in Heterogeneous Mobile Access Networks’, In Chadwick, D., Binczewski, A., Campanella, M., Clarke, P., le Guigner, J.-P., Hanss, T., López, D., Reijs, V., Sartzetakis, S., Wierenga, K., and Williams, S. (Eds.): *Proceedings of TERENA 2004: One Step Ahead, The 20th Trans European Research and Education Networking Conference*, Rhodes, Greece (http://www.terena.nl/ publications//tnc2004-proceedings/ papers/ruiz.pdf).

Satyanarayanan, M. (1996) ‘Fundamental Challenges in Mobile Computing’*,* In *Symposium on Principles of Distributed Computing*, Philadelphia, PA, pp. 1-7.

Satyanarayanan, M., Noble, B., Kumar, P. and Price, M. (1995) ‘Application-Aware Adaptation for Mobile Computing’, *Operating Systems Review*, **29**, 52-55.

Saund, E. and Moran, T. (1994) ‘A Perceptually-Supported Sketch Editor’, In *Proceedings of the ACM Symposium on User Interface Software and Technology,* Marina del Rey, CA, pp. 175 - 184.

Sayood, K. (2000) *Introduction to Data Compression,* Academic Press, San Diego, CA.