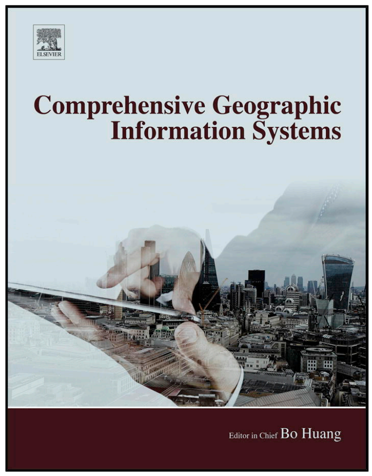

From Li, D., 2018. Geocoding and Reverse Geocoding. In: Huang, B. (Ed.), Comprehensive Geographic Information Systems. Vol. 1, pp. 95–109. Oxford: Elsevier.
http://dx.doi.org/10.1016/B978-0-12-409548-9.09593-2
ISBN: 9780128046609

Elsevier

1.08 Geocoding and Reverse Geocoding

Dapeng Li, Michigan State University, East Lansing, MI, United States

1.08.1 Introduction

Location plays a significant role in geography and geographic information systems (GIS). Georeferencing, defined as the general process of relating information to a geographic location, is an important concept in GIS (Hill, 2009). Geocoding, an important type of georeferencing technique, usually refers to relating street addresses to geographic coordinates (Goldberg, 2011). A broader definition of geocoding is not limited to address names but includes various kinds of geographic features. Geocoding can date back to the early days of digital mapping in the 1960s when the US Census Bureau used digital map databases to match addresses from transportation or health surveys to their corresponding census tracts or blocks (Cooke, 1998). With the rapid development of computer technologies (especially mobile computing) in the past few years, geocoding and reverse geocoding have become so ubiquitous that they have become a necessity in our daily life. For example, a geocoding operation is performed when a user gives a street address as the input to locate it in Google Map, while a reverse geocoding process occurs when the user searches nearby features with a specific geographic location. Moreover, reverse geocoding is also an indispensible functionality in various location-based services (LBS) applications.

Address geocoding has been the primary study focus in this field for decades since residential addresses have served as an important spatial attribute in various survey or historical record data in a variety of fields such as public health and transportation. These addresses need to be transformed to geographic locations before researchers, practitioners, or stakeholders can perform spatial analysis over the data for decision-making or policy making. The theoretical underpinnings, primary methods, various applications, and latest progress of address geocoding will be covered in this article such that the readers can develop a better understanding of this technique.

Geocoding/reverse geocoding can be generally categorized into two groups: conventional and online geocoding/reverse geocoding. Conventional geocoding/reverse geocoding practices are usually conducted by GIS professionals using existing GIS software, and users can have more control on the reference data and geocoding/reverse geocoding method. Online geocoding/reverse geocoding are mostly published by commercial companies as web services, and users can send requests according to predetermined format and receive results from these services. Compared with conventional offline geocoding/reverse geocoding, online services can be readily integrated into modern systems developed on different platforms or in different programming languages. This platform- and language-independent feature has made online geocoding/reverse geocoding services enjoy great popularity in modern information systems. This article also covers online geocoding/reverse geocoding.

The remainder of this article is organized as follows. Section "**Principles and Methods of Geocoding/Reverse Geocoding**" provides the principles, methods, and quality of geocoding/reverse geocoding and relevant metrics for quality evaluation. Section "**Geocoding/Reverse Geocoding Applications**" introduces the applications of geocoding/reverse geocoding. Privacy issues in geocoding/reverse geocoding are covered in section "**Location Privacy in Geocoding/Reverse Geocoding**." The challenges in geocoding/reverse geocoding are given in section "**Recent Trends and Challenges**." Finally, section "**Conclusion**" concludes with a summary of this work.

Comprehensive Geographic Information Systems, First Edition, 2018, 95–109

1.08.2 Principles and Methods of Geocoding/Reverse Geocoding

1.08.2.1 Principles

With substantial demand in a wide range of applications, geocoding/reverse geocoding has become a fundamental module of most GIS software packages. This section examines the principles and methods of geocoding/reverse geocoding. First and foremost, we need to understand the importance of geocoding/reverse geocoding in various applications before we dig into more details. Survey data have been widely used as input for analysis in many disciplines, for example, public health, geography, and sociology. In many cases, the address data are also collected from the subjects, which enable researchers to ask and answer questions from a spatial perspective. However, we need to use geocoding to transform addresses to geographic locations before we could perform spatial analysis to examine the spatial patterns of the phenomena. Equally important is the use of geocoding to associate collected data with various data (e.g., demographic data, socioeconomic status (SES) data, and environmental data) compiled by federal and local government agencies, organizations, and commercial companies and reveal the underlying patterns for policy making (Rushton et al., 2006). For example, the decennial census conducted by the US Census Bureau has been an important source of various demographic and SES data at different spatial levels (e.g., census blocks, block groups, and census tracts). The zonal system used by the US Census Bureau is characterized by a population-based subdivision of the United States. As a matter of fact, the US Census Bureau also sends out surveys to the households and performs geocoding to link the data with corresponding census zones. Geocoding provides an avenue for the researchers to link their data to a wide range of other data such that they could reveal new patterns or discover new knowledge. Fig. 1 demonstrates how geocoding can be used to aggregate the address data at different spatial units and link them with other existing data. Note that the users can use GIS software packages to spatially join the collected data to other data conveniently once they obtain the geographic locations of the address data using geocoding.

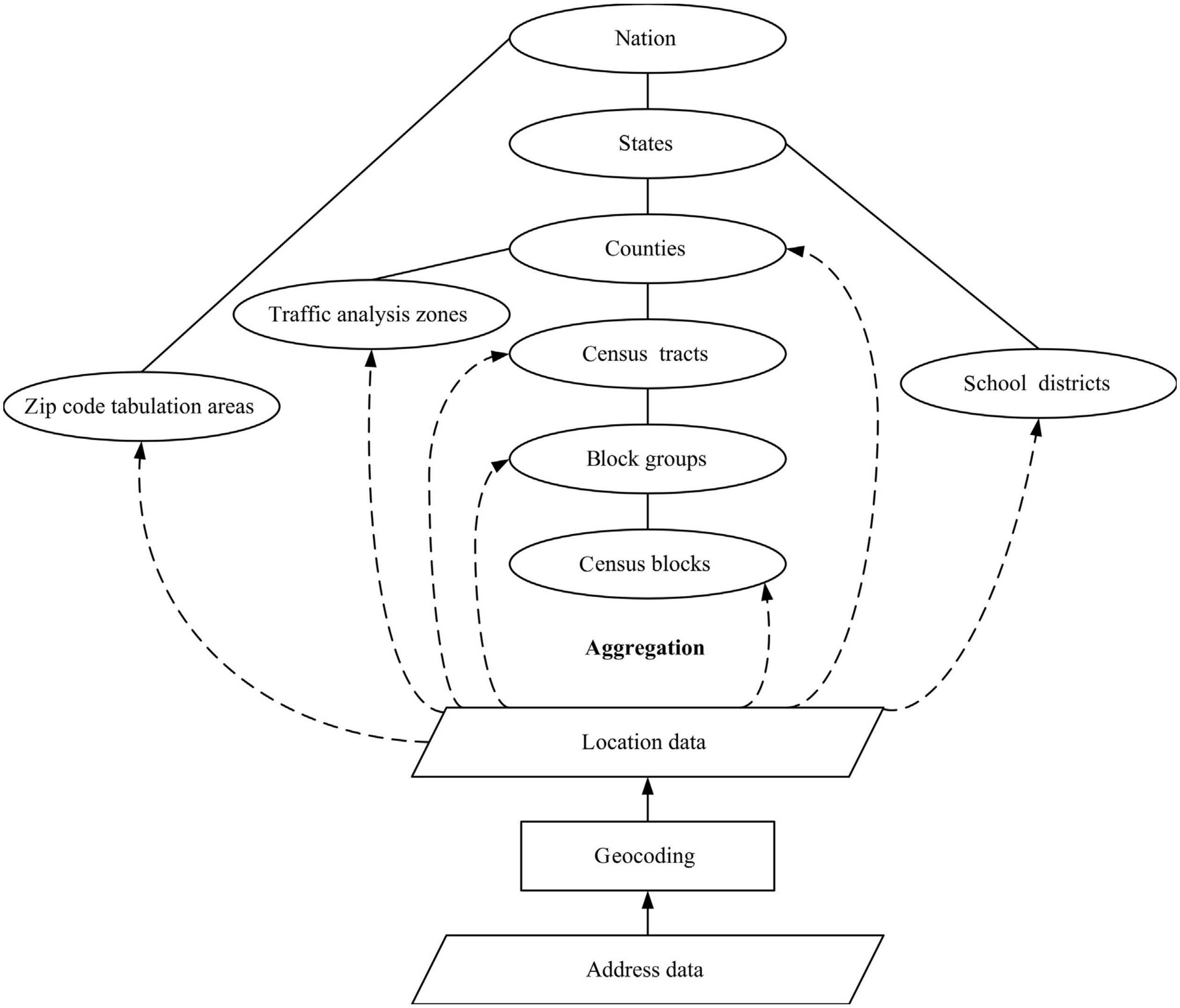

Fig. 1 Using geocoding to link input data to other data at different spatial scales.

Since address data are the primary input for geocoding, we need to develop a good understanding of address data before we could perform geocoding/reverse geocoding procedures more effectively. An address is a piece of text that describes and identifies the location of a specific residence. Different countries may have different postal systems. However, a residential address is usually composed of a series of components, for example, house number, street name/type, city name, state/province name, and zip code. Note that duplicates may exist for one component of an address within a postal system but the whole combination should always be unique. In most cases, a postal system will have a set of standards for addresses. For example, the United States Postal Service (USPS) has its postal addressing standards that describe both standardized address formats and contents. It is important that the address data adhere to these standards during data collection such that they could be readily processed during geocoding. Note that there exist many web services that can verify and validate addresses and can give the users a standardized address. These address standardization tools have been widely used in various software systems to improve the accuracy of address data. In addition to residential addresses, post-office box (POB) addresses are also widely used. Since the location of a POB address usually differs from the location of a physical residence, the use of POB addresses poses challenges to geocoding and subsequent analysis (Hurley et al., 2003). Thus, geocoding practitioners need to take into account address type during the geocoding procedure.

A geocoding procedure is usually composed of several components: the input data, reference datasets, a processing method, and output (Goldberg et al., 2007; Karimi et al., 2004; Levine and Kim, 1998), as shown in **Fig. 2**. As mentioned earlier, the input data usually come from surveys or other sources and are characterized by records with address information. The reference datasets could be in a variety of forms, for example, address point data, road data, and parcel data. A processing algorithm uses the reference datasets to match an input address record to a specific geographic location (latitude/longitude). Note that the reference data should maintain a certain level of accuracy; otherwise, the locations produced from geocoding will be inaccurate. After the geocoding procedure, each address record that can be geocoded is associated with a geographic location.

Fig. 2 could also be used to demonstrate the building blocks of a reverse geocoding procedure. The input data for reverse geocoding will be a set of geographic coordinates, which could be acquired through stand-alone Global Positioning Systems (GPS) units or GPS receivers installed on various mobile devices. A processing algorithm looks up the geographic features in the reference datasets to obtain a feature or features that are close to the input point and satisfy a set of constraints from the user. Thus, reverse geocoding could be considered as a spatial query process that retrieves the features around a given input point from the reference datasets and returns the results to the user. Note that the features in reverse geocoding are not limited to addresses and can also include a variety of feature types. These features are usually termed points of interest (POIs). The output of a reverse geocoding procedure is a feature or a set of features that satisfy certain constraints (e.g., distance and feature type) from the user.

1.08.2.2 Methods

The methods for geocoding/reverse geocoding rely on the reference datasets, and users need to develop a good understanding of various reference datasets first. Reference data could use different data models, for example, point, polyline, and polygon. **Fig. 3** gives some examples of the popularly used reference data: address point, parcel, road, and other spatial units (e.g., counties) data. Note that the reference data should be standardized and appropriately indexed before they are used for geocoding. As for the reference data for reverse geocoding, they should be spatially indexed to speed up spatial query.

Fig. 4 illustrates the general workflow of a geocoding procedure, which includes tokenization, standardization, and address matching (Goldberg et al., 2007). Specifically, in the context of geocoding, tokenization refers to the practice of splitting a whole address into different components. Then, the standardization step transforms each component into standardized form such that it can be used for address matching. In the address matching step, the components of an input address record are used to match the address to a record in the reference datasets. Note that the users can set relevant constraints for the matching step. A scoring algorithm is usually used to assign a score to a matched record. For example, if all the components of an input address have been matched with a record in the reference dataset and an accurate location can be determined, this geocoded record will have a high score; otherwise, if only parts of the components can be matched and the returned geographic point is derived from the reference datasets, the matching can have a low score. Finally, if the matching is successful, a geocoded location will be returned as the result of the geocoding procedure.

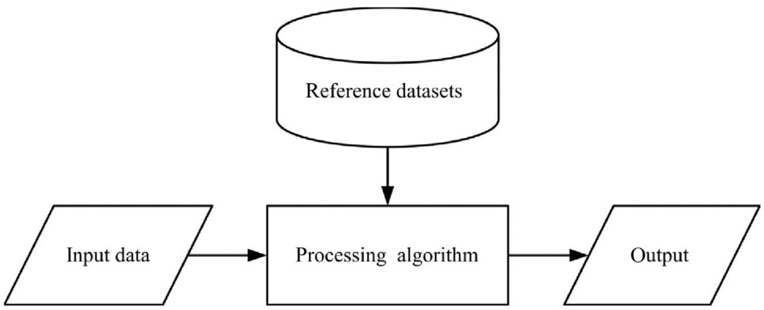

Fig. 2 Components of the geocoding procedure.

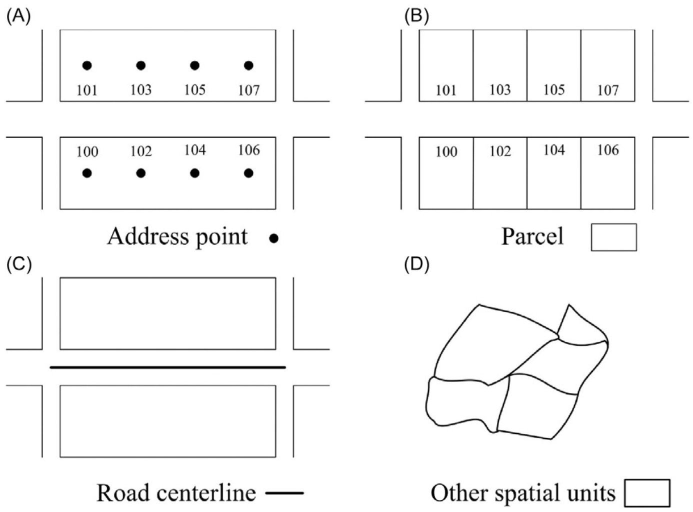

Fig. 3 Examples of various reference data.

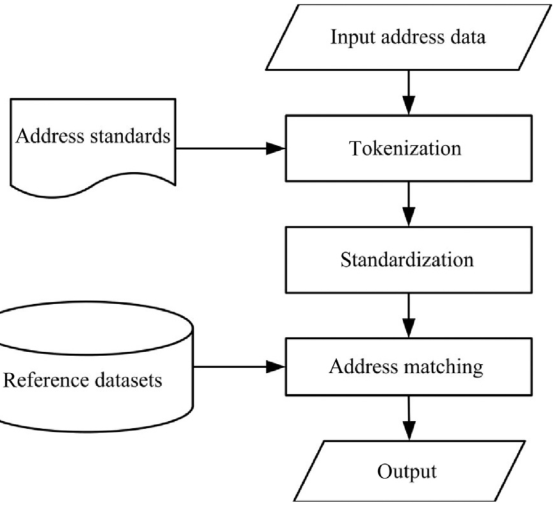

Fig. 4 The general workflow of a processing algorithm.

Address point data use points to represent address locations, and each record includes the address information, which can be used to match the address data during geocoding. The point feature associated with each record could be derived from parcel data. Furthermore, high-resolution remote sensing image data have also been used to improve the spatial accuracy of the point location. For example, a point could be positioned based on the footprint of the residence. Each record in the reference address point dataset contains the different components (e.g., house number, prefix direction, street name, street type, city, state, and zip code) of an address. During the geocoding process, the processing algorithm first splits an input address record into corresponding components and then compares it with the reference data to match it to a point location. Note that there could be many variations in the components of an address. For example, according to the USPS, the standard suffix abbreviation for "AVENUE" is "AVE." However, commonly used suffixes or abbreviations also include "AV" and "AVEN," to name a few. Moreover, there could also be many misspellings in the input address records during data collection. Thus, these issues should also be taken into account in the processing algorithm. Techniques such as soundex (a phonetic index system that indexes information based on word sounds) can be used to deal with these issues (Zandbergen, 2008). It is also worth mentioning that it is expensive to compile and maintain large address point datasets. Countries such as Australia, the United Kingdom, and Canada have national address point datasets. In the United States, although many counties and states have built their address point datasets, a national address point database does not exist at this moment.

Parcel data have also been widely used as the reference datasets in geocoding, and Fig. 3B gives an example of the layout of parcel data. Similarly, each record in the parcel data has address and geometry information. Note that while fine-grain address point data could be used to link an address with a building structure, parcel data could at most link an address with a parcel or the centroid of a parcel. The matching process could also be done in the same manner. The compilation and maintenance of large parcel data is also costly. In the United States, parcel data are usually managed by local government agencies, which limits EL id="del19" orig="s"; the use of parcel data in nationwide geocoding practices. It should also be noted that since the centroid of a parcel is usually used as the result for an input address in a geocoding procedure, parcel sizes will have impacts on positional accuracy.

The use of road data in geocoding has enjoyed great popularity in the past few decades. One important reason is the availability of national road datasets from the US Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) products. The road datasets from the TIGER/Line dataset have been widely used in geocoding practices. The users can join the road dataset to an address range dataset such that each road contains address range information. Due to the growing customer needs for geocoding, the US Census added Address Range Feature shapefiles (ADDREFEAT) to the TIGER/Line products beginning in 2011. Note that while the TIGER road datasets could also be used for other purposes (e.g., transportation studies), the address range feature data are specially designed for address geocoding. Fig. 5 shows the address feature data from the 2015 TIGE/Line products for Salt Lake City, Utah. Note that some road segments are not included in this dataset because they have no addresses (residences) associated with them. Each record includes a series of attributes that are important for geocoding, and Table 1 lists some key attributes.

With given road or address range feature data from TIGER/Line products, the input addresses can be geocoded using an address interpolation algorithm. Fig. 6 gives a demonstration of address interpolation in street geocoding. Each road segment in the reference datasets includes the address range on the left and right side of the road, and the house number from the input address record is used to interpolate the address location. It is worth mentioning that the address range for each side of the road segment can contain more addresses than the true address range. For example, the true address range on the left side of the road in Fig. 6 is 101–107, while the address range in the reference data is 101–109. Note that the offsets along the road segment and perpendicular to the road

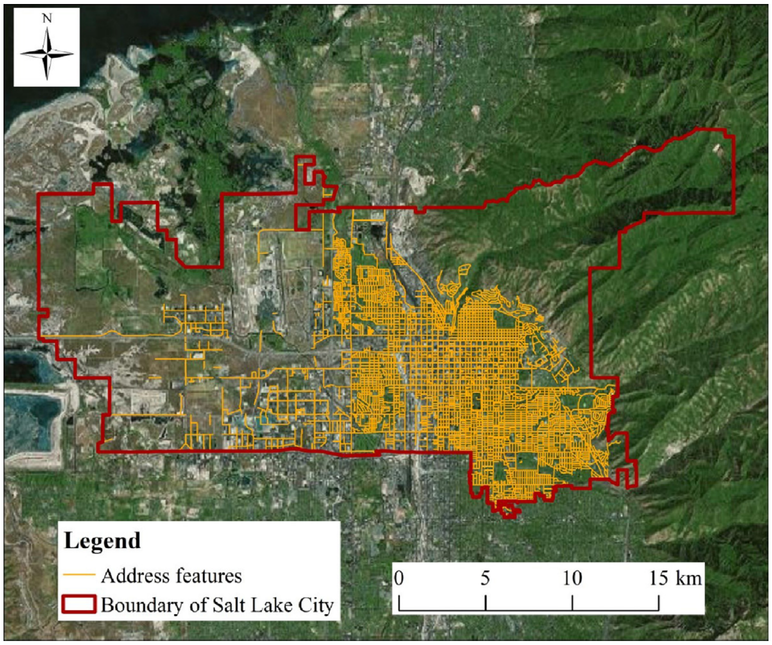

Fig. 5 Address range feature data in Salt Lake City.

Table 1 Key record attributes in the address range feature dataset

Field name	*Description*
FULLNAME	The full name of the road segment
LFROMHN	The beginning address number on the left side of a road segment
LTOHN	The ending address number on the left side of a road segment
RFROMHN	The beginning address number on the right side of a road segment
RTOHN	The ending address number on the right side of a road segment
ZIPL	The zip code on the left side of a road segment
ZIPR	The zip code on the right side of a road segment

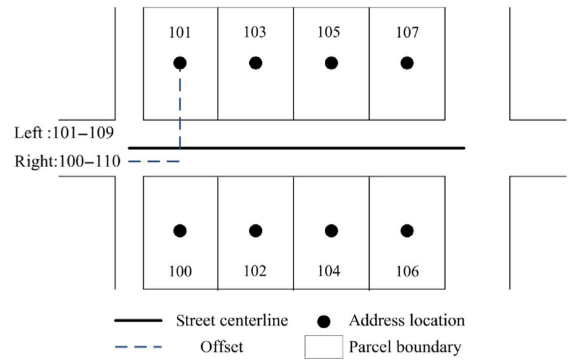

Fig. 6 Address interpolation in street geocoding.

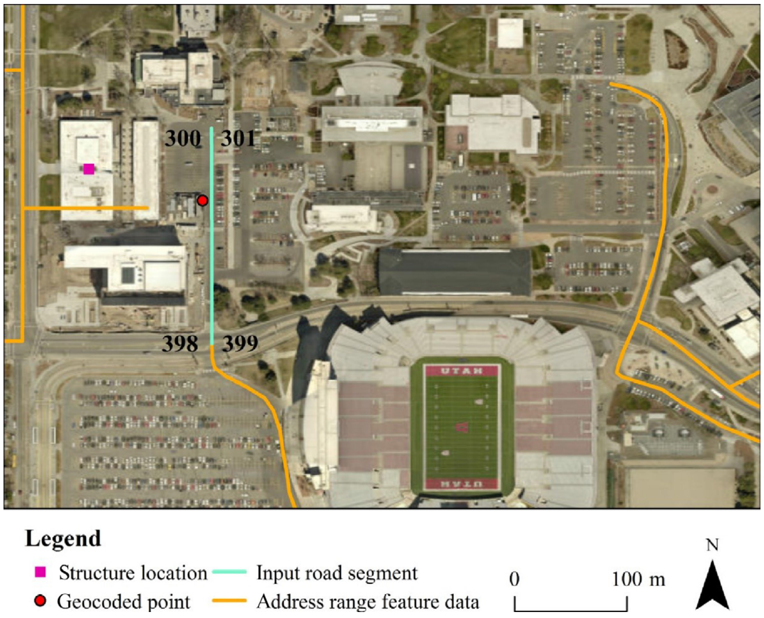

Fig. 7 A geocoded address using the address range feature data in ArcGIS.

segment are usually given in the reference datasets. The offsets are used to compute the location of the geocoded point. It should also be noted that a road segment is directional in geocoding, and the starting and end node information can be used to determine the left and right side of the road. In the United States, the national TIGER/Line datasets are available to the public at no cost, which makes street geocoding a widely used method in geocoding practices.

Fig. 7 gives an example of using address range feature data for geocoding. Specifically, the input address is "332 S 1400 E, SALT LAKE CITY, UT, 84112"; the reference data used are the address range feature data shown in Fig. 5, and geocoding was performed in ArcGIS 10.1. Note that the start point of the input road segment is located at the northmost node while the endpoint at the southmost node. The address ranges on the left and right side of the road are 301–399 and 300–398, respectively. The house number 332 from the input address, the address range data, and the road geometry were used to compute the location of the address. The structure location of the input address is also shown in Fig. 7, and there is a displacement between the interpolated point from geocoding and the true structure location. This displacement can represent the positional accuracy, which will be discussed in more detail in section "**Geocoding/Reverse Geocoding Quality**."

In addition to the aforementioned methods, other areal unit features (e.g., zip code, city, county, and state boundary lines) in different zoning systems can also be used for geocoding (Goldberg and Cockburn, 2012). Similar to parcel geocoding, when these areal unit features are used as the reference data for geocoding, the returned results are the centroids of these features. Note that the centroid computed for an irregular area unit can fall within another areal unit, which can cause significant errors and impact the results (Goldberg and Cockburn, 2012). It is also worth mentioning that areal unit features have also been widely used in reverse geocoding. This process is also termed spatial join in GIS. For example, Nguyen et al. (2016) made use of the census tract boundary data to match geotagged tweets to the census tracts they fall within such that demographic and SES data can be linked to study the social environment at the neighborhood level.

Note that the geocoding procedure has evolved over time with the improvement of reference datasets and processing algorithms. For example, the quality of reference data has been improved significantly from the early Dual Independent Map Encoding data to the widely used TIGER road data and the recent TIGER address range feature data. A study done by Zandbergen et al. (2011) examined the positional accuracy of TIGER 2000 and 2009 data, and the results indicate that TIGER 2009 data are consistently more accurate than TIGER 2000 data. It should also be noted that the choice of the resolution of reference data for geocoding depends on the context and the methods to be used in subsequent analysis procedure. At the same time, a significant amount of research has been done to develop new processing algorithms to further improve the geocoding procedure. For example, many composite methods have been proposed to ensure that more address records can be successfully matched during the geocoding procedure.

1.08.2.3 Online Geocoding/Reverse Geocoding

Online geocoding/reverse geocoding refers to web services that can provide geocoding/reverse geocoding services to the users, and the past few years have witnessed the great popularity of online geocoding/reverse geocoding (Kounadi et al., 2013; Roongpiboonsopit and Karimi, 2010a). Compared with conventional geocoding/reverse geocoding, online geocoding/reverse geocoding services have several advantages. First, the users do not need to compile, process, or manage reference datasets in an online setting. Second, online geocoding services provide easy-to-use, language and platform-independent representational state transfer (REST) application programming interfaces (APIs) such that the users can conveniently perform geocoding operations. For example, most online geocoding/reverse geocoding services use Extensible Markup Language (XML) or JavaScript Object Notation (JSON) as the output formats, which enables the users to integrate these services into different computer systems. However, online geocoding services also come with a number of disadvantages. First, the users cannot manage or control the reference data, the constraints, or the parameters for geocoding/reverse geocoding, which makes it difficult to evaluate the quality. Second, most service providers have set count limits for the number of queries the users can make without any cost, and the users need to buy special licenses before they could use these online services to make large batches of requests. Another constraint is that the users need internet access to use these online services.

Online reverse geocoding services play a significant role in LBS. In a typical LBS application, a user retrieves POIs close to his or her current location for a certain purpose. For example, users can use a mobile application named Yelp in the United States to search for different types of nearby restaurants through its reverse geocoding services. Table 2 lists several popular free reverse geocoding services and their count limits. Note that it is very expensive to compile, update, and maintain the data behind these online reverse geocoding services. Most companies provide their free online reverse geocoding services to basic users with a daily count limit and commercial services to advanced users at a certain price. Since the count limit usually applies to the device, these free online reverse geocoding services will suffice for most client-side applications, and only those server-side applications that make a large number of reverse geocoding requests will need special licenses.

With the growing demand for online geocoding/reverse geocoding services and the rapid development of computing technologies, more and more federal or state government agencies have started to build and publish their own online services. For example, the US Census Bureau has published an online geocoding/reverse geocoding service for public use. In addition to the standardized REST APIs, the US Census also provides a user-friendly online graphical user interface that enables users to upload address data and perform a batch processing operation. Another advancement made by the new US Census online services is that they enable the users to choose the reference data for geocoding/reverse geocoding. The users can use either the most current TIGER/Line data or the 2010 TIGER/Line data for geocoding/reverse geocoding. This gives the users more control over the reference data used in geocoding/reverse geocoding. Another example for online geocoding/reverse geocoding services developed by government agencies comes from the Automated Geographic Reference Center (AGRC) – the GIS Department of Utah in the United States. AGRC has published its online geocoding/reverse geocoding services to satisfy various data needs in Utah. One novel feature provided by AGRC's services is that they enable the users to choose a spatial reference for geocoding/reverse geocoding. For example, the user can choose to use either a geographic coordinate system (e.g., World Geodetic System (WGS) 1984) or a projected coordinate system (e.g., Web Mercator) for geocoding/reverse geocoding. Furthermore, a number of other features (e.g., allowing the use of different types of geometry for spatial queries) are still under development. Moreover, AGRC also provides the examples of using these services in different programming languages and publishes the source code on GitHub. In summary, all these endeavors made by federal, state, or local government agencies will accelerate the development of online geocoding/reverse geocoding services and play a significant role in data sharing in the era of open science.

Table 2 Examples of free online reverse geocoding services

Service name	*Company*	*Limitations*	*Output format*
GeoNames	Marc Wick	2000 requests/h	XML/JSON
Google Maps	Google	2500 requests/day	XML/JSON
MapQuest	AOL	5000 requests/day	XML/JSON
Nominatim	Nominatim	1 request/s	XML/JSON

1.08.2.4 Geocoding/Reverse Geocoding Quality

Since geocoding usually serves as one step and provides location data for subsequent analysis and computation in many research endeavors, its quality has significant impacts on subsequent procedures and the final results (Zandbergen et al., 2012). Thus, we need to develop a better understanding of the quality of the geocoding procedure before we could analyze the data and interpret the results correctly. Popular metrics to evaluate geocoding quality include positional accuracy, match rate, and repeatability (Zandbergen, 2009b). Online reverse geocoding has been widely used in various LBS applications, and positional accuracy is also an important metric to evaluate reverse geocoding quality.

Positional accuracy is an issue that can never be overemphasized in geocoding and reverse geocoding. Positional errors in geocoding/reverse geocoding can propagate to subsequent spatial analysis and influence the results directly, and a significant body of research has been conducted to examine error propagation in various applications (Jacquez, 2012; Karimi et al., 2004; Zandbergen et al., 2012). With the popularity of mobile computing in the past few years, positional accuracy in geocoding/reverse geocoding is becoming more significant. Fig. 8 illustrates the locations and displacements relevant to the positional accuracy of geocoding/reverse geocoding. The detailed descriptions are listed in Table 3. In a geocoding scenario, the user can obtain a location point p_2 for an input address with the true location p_3. The distance d_2 between p_2 and p_3 can represent the positional accuracy of the geocoding procedure. If address point data are used as the reference data in geocoding, d_2 is the displacement between the address point and its true location; if street centerline data are used, the accuracy of the road geometry and the offsets along and perpendicular to the road segment both contribute to the d_2. Note that p_3 can be derived from GPS units, large-scale parcel maps, or georeferenced high-resolution remote sensing image to compute d_2 and evaluate positional accuracy in geocoding practices (Bonner et al., 2003; Cayo and Talbot, 2003; Curtis et al., 2006; Roongpiboonsopit and Karimi, 2010a; Strickland et al., 2007; Zandbergen and Green, 2007; Zandbergen et al., 2011).

With the advent of GPS-enabled mobile devices, mobile computing has enjoyed great popularity in the past few years, and online reverse geocoding services have been widely used in various LBS applications. In high-density urban areas, the GPS signal may be impacted by obstructions such as buildings, which results in inaccuracy in the location derived from GPS-enabled mobile devices; moreover, locations derived from Wi-Fi and cellular positioning have large errors and cannot satisfy the accuracy needed by many LBS applications (Zandbergen, 2009a). A typical LBS application is used as an example to demonstrate the positional accuracy

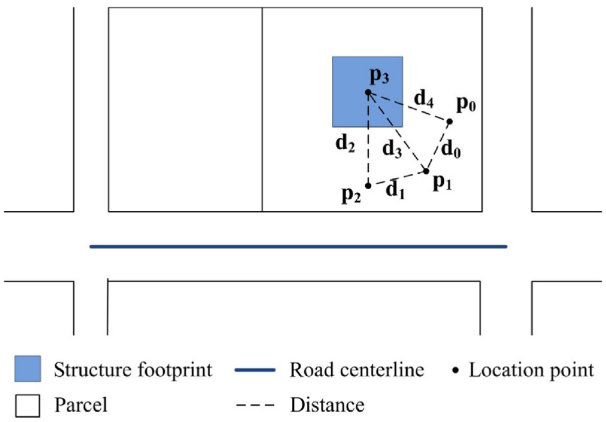

Fig. 8 Illustration of positional accuracy in geocoding and reverse geocoding.

Table 3 Positional accuracy of geocoding/reverse geocoding

Parameter	*Description*
p_0	The true location of a user's location
p_1	The input query point for the reverse geocoding
p_2	The derived point from geocoding/reverse geocoding
p_3	The true location of a target structure (centroid of the structure footprint)
d_0	The displacement between the input point and the true location of it
d_1	The distance between the input point and the derived point from reverse geocoding
d_2	The displacement between the derived point from geocoding/reverse geocoding and the true location of the target address/POI
d_3	The distance between the input point for reverse geocoding services and the true location of the target address/POI
d_4	The distance between the true location of the input point and the true location of the target address/POI

of online reverse geocoding services and mobile devices. As shown in Fig. 8, in the context of a LBS application, p_0 denotes the true location of a user, p_1 is the input query point derived from a mobile device, p_2 is the derived point from a reverse geocoding service, and p_3 is the true location of the POI. Theoretically, the true distance d_4 between the user and the POI should be used in LBS applications; however, since the displacement d_0 caused by mobile devices and d_2 from online reverse geocoding services are usually unavoidable, the distance d_1 is what the user derives in a LBS application. One example of d_0 is the displacement caused by GPS receivers: the user gets his or her location p_1 from GPS receivers, while the true location is p_0. It is worth mentioning that most modern operating systems for mobile devices provide relevant APIs to derive geographic locations. Note that locations derived from different sources (e.g., GPS receivers, Wi-Fi access points, and cell powers) vary in positional accuracy (Zandbergen, 2009a). A study conducted by Zandbergen and Barbeau (2011) examined the accuracy of assisted GPS (A-GPS) on mobile phones under diffident conditions (static and dynamic outdoor tests and static indoor test), and the results indicate that A-GPS is accurate enough to satisfy the needs of most LBS applications. However, it may still be challenging to use locations derived from mobile devices in some special LBS applications (e.g., locating a specific POI in urban areas with high POI density). Thus, developers should take into account positional accuracy when they design and implement reverse geocoding services and LBS applications.

Match rate, also coined completeness, is defined as the percentage of data records that are successfully matched (Zandbergen, 2008). It has become a routine to include the match rate of the geocoding procedure in most studies. It should be noted that match rate itself cannot represent the quality of a geocoding procedure. For example, a geocoding procedure based on address point data could have a lower match rate than one based on road centerline data. However, the geocoded points derived from the former are usually more accurate than the results from the latter in terms of positional accuracy. Although there is no consensus on the minimum match rate for each specific application, many studies have been conducted on this topic. For example, Ratcliffe (2004) used Monte Carlo simulation to replicate a declining match rate, and the result indicates that the estimated minimum reliable match rate for crime mapping is 85%.

Repeatability refers to how sensitive the results are to variation in different components of geocoding (e.g., reference datasets, and matching algorithm) (Zandbergen, 2008). Whitsel et al. (2006) used addresses in 49 US states to evaluate geocoding services provided by four vendors, and the results revealed that significant differences exist in match rate, positional accuracy, and concordance between established and derived census tracts. Note that these metrics can also be used to evaluate online geocoding services (Roongpiboonsopit and Karimi, 2010a,b). It is worth mentioning that when evaluating geocoding quality, we should take into account all these metrics; otherwise, the conclusion could be biased.

1.08.3 Geocoding/Reverse Geocoding Applications

Geocoding and reverse geocoding have been widely used in a variety of fields. This subsection gives a brief review of various kinds of applications of geocoding/reverse geocoding. Specifically, the applications of geocoding in health, crime, and traffic accident studies are covered, and the applications of reverse geocoding concentrate on various LBS applications.

As noted, geocoding has been extensively used in health studies, for example, measures of accessibility, disease cluster analysis, and exposure analysis. Accessibility studies in public health deal with measuring people's accessibility to ability to access a wide range of resources that have impacts on their health and well-being, for example, food outlets (Barnes et al., 2015; Vandevijvere et al., 2016), health care facilities (Luo and Wang, 2003; Wan et al., 2012), and tobacco stores (Cantrell et al., 2016). Geocoding is used to locate both facilities and residences in accessibility-related studies. In disease cluster analysis, geocoding is employed to transform the patients' addresses into geographic coordinates such that spatial analysis can be performed to detect clusters (Glatman-Freedman et al., 2016). Exposure analysis focuses on examining the impacts of pollution on people's health and well-being, and geocoding is used to locate the residences of the subjects so that they can be linked to the environmental pollution data (Bellander et al., 2001; Ganguly et al., 2015). Another important application of geocoding is to transform addresses into geographic points in crime studies, for example, crime mapping and analysis (Andresen et al., 2016; Ratcliffe, 2002, 2004) and residency restrictions for sex offenders (Rydberg et al., 2016; Zandbergen and Hart, 2009). Moreover, geocoding has also been widely used to georeference police crash records to analyze traffic accidents and identity the road links that are more likely to cause crashes (Erdogan et al., 2008).

Reverse geocoding has been widely used in a myriad of LBS applications. These applications are usually in an online setting. Online reverse geocoding services provide addresses or various POIs to the users to satisfy their needs. Note that these online services usually offer user-friendly interfaces such that users can make queries using different spatial or aspatial constraints. Moreover, these online services can be readily integrated into various computer systems on different platforms. Besides satisfying the users' information needs in LBS applications, reverse geocoding has also been employed to build semantics for people's GPS trajectories (Cao et al., 2010; Lv et al., 2016), which can help us develop a better understanding of people's exposure to social, cultural, and physical environments.

1.08.4 Location Privacy in Geocoding/Reverse Geocoding

Since geocoding/reverse geocoding deals with sensitive location data, a great deal of attention is being paid to the privacy issues in various geocoding/reverse geocoding applications (Armstrong et al., 1999; Cassa et al., 2006; Curtis et al., 2006; Kounadi et al., 2013; Kounadi and Leitner, 2014; Tompson et al., 2015; VanWey et al., 2005). For example, in the United States, many public health researchers and practitioners deal with sensitive individual data in their work. Although the Health Insurance Portability and Accountability Act issued by the US Department of Health and Human Services in 1996 has relevant regulations to protect individually identifiable health information, geolocation privacy issues still exist because of the difficulty to quantify the risk of disclosure. As mentioned earlier, it is a common practice to aggregate the collected data at a specific spatial level such that other datasets can be linked and further analysis is performed. And privacy issues will arise when large-scale maps are published to present the results to the public because detailed map data can be easily reverse geocoded to derive individual location and information. The practice of using published maps and/or other related data to recover the identity of an individual and infer individual information is termed reengineering (Curtis et al., 2006). Armstrong and Ruggles (2005) gave two examples of possible privacy violation in geocoding/reverse geocoding applications: one is reengineering individual-level information from dot maps and the other one rediscovering user location in LBS applications. Meanwhile, they also point out that the concepts of privacy issues change with the development of technologies. For example, the widely available high-resolution remote sensing image data could not only be used to construct more accurate structure-level reference data to improve geocoding quality but also be used to recover individual-level information. Thus, the development of technologies provides us an avenue to improve geocoding/reverse geocoding quality and also poses new challenges in privacy protection.

In order to reduce disclosure risk, a substantial body of research has been conducted on preserving individual confidentiality in geocoding applications. Specifically, one simple method is to aggregate individual information to various geographic areas (e.g., census tracts, zip code zones, and counties) such that individual information will be protected. Note that if aggregated data are used in subsequent analysis, the details of spatial patterns could be lost, which makes it difficult to detect clusters (Armstrong et al., 1999). Furthermore, data aggregation can also lead to the modifiable unit area problem (Openshaw, 1984). Armstrong et al. (1999) listed some alternative methods (e.g., individual affine transformations, random perturbation, aggregation, neighbor information, and contextual information) that can mask individual information and preserve valuable information, and this line of research has attracted significant research attention in the past few years (Wieland et al., 2008; Zhang et al., 2015). A recent summary of the popularly used geographic masking methods was given by Kounadi and Leitner (2015). According to Armstrong et al. (1999), the methodology used to evaluate the effectiveness of geographic masking methods covers three aspects: (1) preserving information, (2) preserving links to other spatial data, and (3) preserving confidentiality. Note that the information to be preserved can be reflected in many forms (e.g., pairwise relations, event–geography relations, clusters, trends, and anisotropies) and different characteristics are used to store the information (e.g., distance, orientation, and directionality) (Armstrong et al., 1999). Thus, when evaluating the capabilities of geographic masks to preserve information, we need to take into account the application context (i.e., what information we should prioritize to preserve). For example, if we are to study the spatial pattern of disease incidences, the priority will be placed on preserving the spatial pattern information; however, if we study the effects of environmental exposure on people's health conditions, we may need to take into consideration preserving the distance between residences and the pollution sources.

With the popularity of various LBS applications, location privacy has also attracted a significant amount of attention. In LBS applications, users provide their locations as the input for LBS services to search and retrieve relevant POIs. This process poses significant disclosure risk to the users because the attackers can retrieve the location and identify the user (Wernke et al., 2014). Many countermeasures have been put forward to protect location privacy: regulatory strategies, privacy policies, anonymity, and obfuscation (Duckham and Kulik, 2006). Among these countermeasures, k-anonymity is a fundamental method that has been widely used in many studies (Gedik and Liu, 2008; Krumm, 2009). The concept of k-anonymity was originally proposed to protect privacy in the data released from a data holder, and a release is k-anonymous if each person's information is indistinguishable from at least $k-1$ other persons' information in the same release (Sweeney, 2002). Location k-anonymity can be achieved when the location information from a mobile user cannot be distinguished from the location information of at least $k-1$ other mobile users (Gruteser and Grunwald, 2003). Location k-anonymity has been extensively used as a metric for confidentiality preservation in many studies (Tribby et al., 2015; Zhang et al., 2015). It should be noted that we should take into account privacy issues if we use any individual-level location data in our research.

1.08.5 Recent Trends and Challenges

1.08.5.1 Accessibility

The Euclidean distance has been widely used in various geocoding/reverse geocoding applications. For example, the Euclidean distance between the geocoded point and the true address location is computed to evaluate the positional accuracy of a geocoding procedure. Note that we need to take into account the application context when evaluating the impacts of positional accuracy on the final results. For example, for certain studies that employ the spatial analysis methods that rely on Euclidean distance, the positional error of the geocoded locations can influence the results directly, while for the studies that use analysis methods based on other types of distances, positional accuracy may not directly reflect variations in the results. For example, Euclidean distance can be used as an effective measure in environment exposure studies but may cause large errors in accessibility studies. **Fig. 9** lists different measures of accessibility that differ in accuracy. Note that Euclidean distance is a very coarse measure of accessibility because people need to travel within the road network to access different resources. Thus, most health studies that involve measuring people's access to health-related resources use network-based distance or travel time as the measure of accessibility (Delamater, 2013; Wan et al., 2012). However, it is worth mentioning that most health studies that use network-based distance or travel time to measure

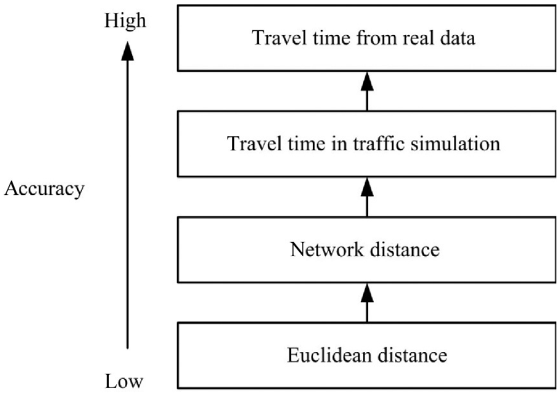

Fig. 9 Accuracy of different measures of accessibility.

accessibility do not consider traffic congestion and the travel delays caused by traffic lights. A more accurate travel time can be estimated using traffic simulation, which can take into account these aspects by using relevant travel demand and traffic lights data. Finally, with the rapid development of big data technologies and intelligent transport systems (ITS), travel time can be derived from the real traffic data, which will provide a more accurate, detailed measure of accessibility for various applications.

In addition to various geocoding applications, the development of measures of accessibility also has implications on many reverse geocoding applications. Specifically, one basic need in many LBS applications is to locate and access a nearby POI. Although a few navigational LBS applications (e.g., Google Map) provide travel times based on real traffic data to the users, a majority of reverse geocoding services and LBS applications still use Euclidean or network distance. It is expected that big data and ITS can be leveraged to enrich reverse geocoding services and resolve this issue.

1.08.5.2 The Temporal Dimension

Traditional research on geocoding/reverse geocoding has primarily focused on the spatial dimension (e.g., spatial accuracy), and the temporal dimension still remains underresearched. The temporal dimension in this context can have several aspects. First, note that the reference data used in geocoding/reverse geocoding are not static and commercial data companies or government agencies will add new records or update existing records during their maintenance. When users perform geocoding/reverse geocoding operations, they are using a snapshot of the reference data. While users can choose to use a specific reference dataset in convention geocoding, they have little control over the reference data when using commercial online geocoding services. Most commercial companies that provide online geocoding/reverse geocoding services take it for granted that users will need the most current reference data for geocoding and have neglected the situations in which the users need to use historical reference data to link to other historical datasets. As mentioned earlier, some online services like the US Census geocoding/reverse geocoding services have started to provide different referent datasets for the users, and this issue concerning the temporal aspect of the reference data in these online services could be solved with the development of technology and the increasing awareness of the importance of this problem. The implications of using reference datasets compiled in different time periods on academic research lie in that they differ in data quality, and this can influence subsequent analysis and the final results (Zandbergen et al., 2011). More research should be conducted to more thoroughly examine the impacts of data quality on spatial analysis and result interpretation.

Another aspect of the temporal dimension concerns the temporal attributes of various POIs used in reverse geocoding for LBS applications. For example, the hours of operation for the POIs can be valuable for a myriad of LBS applications. A great many commercial reverse geocoding services have included hours of operation of the POIs in the databases. For example, Google maintains its own POI database and provides reverse geocoding services to the users through the Google Places APIs. The users can get the detailed information such as hours of operation about a specific POI, which provides great convenience to the users when they look for certain POIs. However, note that the hours of operation information for a specific could be subject to change due to holidays or other reasons, which indicates that it will be more valuable if these reverse geocoding service providers can connect with the business owners and maintain this information in a dynamic manner. With the rapid development of the internet of things, this could be achieved to further increase the value of various reverse geocoding services.

The temporal dimension of the POIs has also attracted significant attention in academia. From a supply and demand perspective, the hours of operation of POIs can reflect the supply side and play a significant role in measuring people's access to relevant resources. Studies on food access have incorporated time into accessibility measurements (Chen and Clark, 2013, 2016), which provides an avenue to discover new scientific findings. Moreover, recent studies have also proposed new methods to enrich reverse geocoding services. For example, user check-in data have been used to construct temporal signatures and further enrich the temporal information of the POIs (McKenzie and Janowicz, 2015; McKenzie et al., 2015). These endeavors based on data-driven methods will further improve the quality of reverse geocoding services and have great potential in designing better LBS applications that can better satisfy the users' needs (Majid et al., 2013).

1.08.5.3 Indoor Positioning

With the rapid development of mobile computing, the past few years have witnessed the great popularity of indoor positioning, and studies have shown that technologies such as Bluetooth and Wi-Fi can be readily used for indoor positioning (He and Chan, 2016; Li et al., 2015; Liu et al., 2007). The advent of indoor positioning poses many new challenges to geocoding/reverse geocoding practitioners. Mapping companies have incorporated indoor maps into their mapping practice such that users can locate various POIs more conveniently. Fig. 10 gives an example of indoor mapping in Google Map for the Hollywood & Highland building in Los Angeles, United States. Note that the physical address for the whole building is "6801 Hollywood Blvd, Los Angeles, CA 90028," and many POIs are located inside the building. In this case, indoor positioning can be integrated with mobile mapping to better facilitate pedestrian navigation in indoor environments, and relevant geocoding/reverse geocoding services need to be developed such that users can readily locate relevant POIs in indoor environments. Indoor positioning has brought about many new indoor LBS applications, and indoor location-based retailing is one of the novel applications (Thamm et al., 2016). Small-sized beacons can be conveniently attached to different products, and when a consumer approaches an item, the smartphone can sense the beacon, and relevant descriptions of the item can be retrieved and displayed to the user. In this regard, relevant geocoding/reverse geocoding services can also be developed to satisfy these new needs in indoor environments.

1.08.5.4 Privacy in the Mobile Age

Mobile computing has become so pervasive that more and more mobile devices are being used in our daily life, and various mobile applications on smartphones or smart watches can easily record people's locations and trajectories, which pose significant risks on people's privacy. For example, a great many studies have shown that private information such as identity and home and workplace locations can be inferred from GPS trajectory data by using other ancillary data and advanced data mining methods (Krumm, 2007; Zheng, 2015). Recent studies reveal that even the trajectory data from geotagged social media posts can be used to infer various private data (Kim et al., 2016; Luo et al., 2016). These new challenges have attracted significant attention, and a myriad of studies

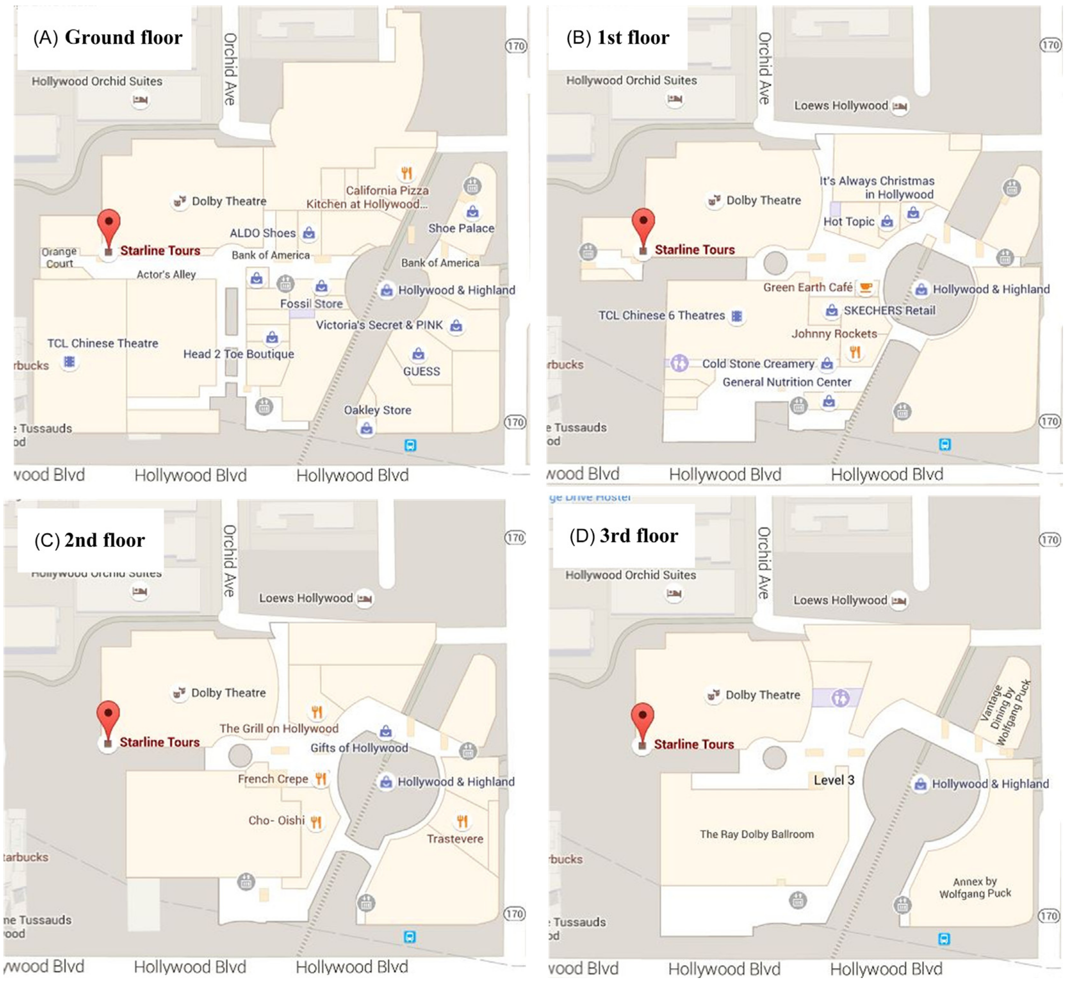

Fig. 10 An example of indoor mapping in Google Map.

have been conducted to develop relevant countermeasures to protect privacy from being exposed in the trajectory data (Fanaeepour et al., 2015; Seidl et al., 2016). Furthermore, the rapid development of big data and new positioning technologies (e.g., indoor positioning) will cause many new privacy issues, and more research should be conducted to thoroughly examine these issues and protect user privacy in the mobile age.

1.08.6 Conclusion

This work summarizes the principles, methods, and applications of geocoding/reverse geocoding and introduces the metrics for evaluating the quality of geocoding/reverse geocoding procedures. Furthermore, the pros and cons of online geocoding/reverse geocoding services are also covered. The privacy issues in geocoding/reverse geocoding are introduced, and relevant countermeasures for privacy protection are also discussed. Finally, the recent trends and challenges in geocoding/reverse geocoding are included to shed light on potential future research directions in this field. In summary, it is necessary for geocoding/reverse geocoding users and researchers to develop a good understanding of the theoretical underpinnings before they can use it correctly and more effectively and contribute to its development.

References

Andresen, M.A., Linning, S.J., Malleson, N., 2016. Crime at places and spatial concentrations: Exploring the spatial stability of property crime in Vancouver BC, 2003–2013. Journal of Quantitative Criminology 1–21. http://dx.doi.org/10.1007/s10940-016-9295-8.

Armstrong, M.P., Ruggles, A.J., 2005. Geographic information technologies and personal privacy. Cartographica: The International Journal for Geographic Information and Geovisualization 40 (4), 63–73. http://dx.doi.org/10.3138/RU65-81R3-0W75-8V21.

Armstrong, M.P., Rushton, G., Zimmerman, D.L., 1999. Geographically masking health data to preserve confidentiality. Statistics in Medicine 18 (5), 497–525.

Barnes, T.L., Bell, B.A., Freedman, D.A., Colabianchi, N., Liese, A.D., 2015. Do people really know what food retailers exist in their neighborhood? Examining GIS-based and perceived presence of retail food outlets in an eight-county region of South Carolina. Spatial and Spatio-Temporal Epidemiology 13, 31–40. http://dx.doi.org/10.1016/j.sste.2015.04.004.

Bellander, T., Berglind, N., Gustavsson, P., Jonson, T., Nyberg, F., Pershagen, G., Järup, L., 2001. Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. Environmental Health Perspectives 109 (6), 633–639.

Bonner, M.R., Han, D., Nie, J., Rogerson, P., Vena, J.E., Freudenheim, J.L., 2003. Positional accuracy of geocoded addresses in epidemiologic research. Epidemiology 14 (4), 408–412. http://dx.doi.org/10.1097/01.EDE.0000073121.63254.c5.

Cantrell, J., Pearson, J.L., Anesetti-Rothermel, A., Xiao, H., Kirchner, T.R., Vallone, D., 2016. Tobacco retail outlet density and young adult tobacco initiation. Nicotine & Tobacco Research 18 (2), 130–137. http://dx.doi.org/10.1093/ntr/ntv036.

Cao, X., Cong, G., Jensen, C.S., 2010. Mining significant semantic locations from GPS data. Proceedings of the VLDB Endowment 3 (1–2), 1009–1020. http://dx.doi.org/10.14778/1920841.1920968.

Cassa, C.A., Grannis, S.J., Overhage, J.M., Mandl, K.D., 2006. A context-sensitive approach to anonymizing spatial surveillance data. Impact on Outbreak Detection 13 (2), 160–165. http://dx.doi.org/10.1197/jamia.M1920.

Cayo, M.R., Talbot, T.O., 2003. Positional error in automated geocoding of residential addresses. International Journal of Health Geographics 2 (1), 1–12. http://dx.doi.org/10.1186/1476-072x-2-10.

Chen, X., Clark, J., 2013. Interactive three-dimensional geovisualization of space-time access to food. Applied Geography 43, 81–86. http://dx.doi.org/10.1016/j.apgeog.2013.05.012.

Chen, X., Clark, J., 2016. Measuring space-time access to food retailers: A case of temporal access disparity in Franklin County, Ohio. The Professional Geographer 68 (2), 175–188. http://dx.doi.org/10.1080/00330124.2015.1032876.

Cooke, D.F., 1998. Topology and TIGER: The Census Bureau's contribution. In: Foresman, T.W. (Ed.), The history of geographic information systems: Perspectives from the pioneers. Upper Saddle River, NJ, Prentice Hall.

Curtis, A.J., Mills, J.W., Leitner, M., 2006. Spatial confidentiality and GIS: Re-engineering mortality locations from published maps about Hurricane Katrina. International Journal of Health Geographics 5 (1), 1–12. http://dx.doi.org/10.1186/1476-072x-5-44.

Delamater, P.L., 2013. Spatial accessibility in suboptimally configured health care systems: A modified two-step floating catchment area (M2SFCA) metric. Health & Place 24, 30–43. http://dx.doi.org/10.1016/j.healthplace.2013.07.012.

Duckham, M., Kulik, L., 2006. Location privacy and location-aware computing. In: Drummond, J. (Ed.), Dynamic and mobile GIS: Investigating change in space and time. CRC Press, Boca Raton, FL, pp. 34–51.

Erdogan, S., Yilmaz, I., Baybura, T., Gullu, M., 2008. Geographical information systems aided traffic accident analysis system case study: City of Afyonkarahisar. Accident Analysis & Prevention 40 (1), 174–181. http://dx.doi.org/10.1016/j.aap.2007.05.004.

Fanaeepour, M., Kulik, L., Tanin, E., Rubinstein, B.I., 2015. The CASE histogram: Privacy-aware processing of trajectory data using aggregates. GeoInformatica 19 (4), 747–798. http://dx.doi.org/10.1007/s10707-015-0228-8.

Ganguly, R., Batterman, S., Isakov, V., Snyder, M., Breen, M., Brakefield-Caldwell, W., 2015. Effect of geocoding errors on traffic-related air pollutant exposure and concentration estimates. Journal of Exposure Science and Environmental Epidemiology 25 (5), 490–498. http://dx.doi.org/10.1038/jes.2015.1.

Gedik, B., Liu, L., 2008. Protecting location privacy with personalized *k*-anonymity: Architecture and algorithms. IEEE Transactions on Mobile Computing 7 (1), 1–18. http://dx.doi.org/10.1109/TMC.2007.1062.

Glatman-Freedman, A., Kaufman, Z., Kopel, E., Bassal, R., Taran, D., Valinsky, L., Shohat, T., 2016. Near real-time space-time cluster analysis for detection of enteric disease outbreaks in a community setting. Journal of Infection 73 (2), 99–106. http://dx.doi.org/10.1016/j.jinf.2016.04.038.

Goldberg, D.W., 2011. Advances in geocoding research and practice. Transactions in GIS 15 (6), 727–733. http://dx.doi.org/10.1111/j.1467-9671.2011.01298.x.

Goldberg, D.W., Cockburn, M.G., 2012. The effect of administrative boundaries and geocoding error on cancer rates in California. Spatial and Spatio-temporal Epidemiology 3 (1), 39–54. http://dx.doi.org/10.1016/j.sste.2012.02.005.

Goldberg, D.W., Wilson, J.P., Knoblock, C.A., 2007. From text to geographic coordinates: The current state of geocoding. URISA Journal 19 (1), 33–46.

Gruteser, M., Grunwald, D., 2003. Anonymous usage of location-based services through spatial and temporal cloaking. In: Paper Presented at the Proceedings of the 1st International Conference on Mobile Systems, Applications and Services, San Francisco, California.

He, S., Chan, S.H.G., 2016. Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons. IEEE Communications Surveys & Tutorials 18 (1), 466–490. http://dx.doi.org/10.1109/COMST.2015.2464084.

Hill, L.L., 2009. Georeferencing: The geographic associations of information. The MIT Press, Cambridge, MA.

Hurley, S.E., Saunders, T.M., Nivas, R., Hertz, A., Reynolds, P., 2003. Post office box addresses: A challenge for geographic information system-based studies. Epidemiology 14 (4), 386–391. http://dx.doi.org/10.1097/01.ede.0000073161.66729.89.

Jacquez, G.M., 2012. A research agenda: Does geocoding positional error matter in health GIS studies? Spatial and Spatio-temporal Epidemiology 3 (1), 7–16. http://dx.doi.org/10.1016/j.sste.2012.02.002.

Karimi, H.A., Durcik, M., Rasdorf, W., 2004. Evaluation of uncertainties associated with geocoding techniques. Computer-Aided Civil and Infrastructure Engineering 19 (3), 170–185. http://dx.doi.org/10.1111/j.1467-8667.2004.00346.x.

Kim, M.G., Kang, Y.O., Lee, J.Y., Koh, J.H., 2016. Inferring tweet location inference for twitter mining. Spatial Information Research 1–15. http://dx.doi.org/10.1007/s41324-016-0041-y.

Kounadi, O., Leitner, M., 2014. Why does geoprivacy matter? The scientific publication of confidential data presented on maps. Journal of Empirical Research on Human Research Ethics 9 (4), 34–45. http://dx.doi.org/10.1177/1556264614544103.

Kounadi, O., Leitner, M., 2015. Spatial information divergence: Using global and local indices to compare geographical masks applied to crime data. Transactions in GIS 19 (5), 737–757. http://dx.doi.org/10.1111/tgis.12125.

Kounadi, O., Lampoltshammer, T.J., Leitner, M., Heistracher, T., 2013. Accuracy and privacy aspects in free online reverse geocoding services. Cartography and Geographic Information Science 40 (2), 140–153. http://dx.doi.org/10.1080/15230406.2013.777138.

Krumm, J., 2007. Inference attacks on location tracks. In: LaMarca, A., Langheinrich, M., Truong, K.N. (Eds.), Pervasive Computing: 5th International Conference, PERVASIVE 2007, Toronto, Canada, May 13–16, 2007, Proceedings. Springer, Berlin, Heidelberg, pp. 127–143.

Krumm, J., 2009. A survey of computational location privacy. Personal and Ubiquitous Computing 13 (6), 391–399. http://dx.doi.org/10.1007/s00779-008-0212-5.

Levine, N., Kim, K.E., 1998. The location of motor vehicle crashes in Honolulu: A methodology for geocoding intersections. Computers, Environment and Urban Systems 22 (6), 557–576. http://dx.doi.org/10.1016/S0198-9715(98)00034-9.

Li, X., Wang, J., Liu, C., 2015. A bluetooth/PDR integration algorithm for an indoor positioning system. Sensors 15 (10), 24862.

Liu, H., Darabi, H., Banerjee, P., Liu, J., 2007. Survey of wireless indoor positioning techniques and systems. IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews 37 (6), 1067–1080. http://dx.doi.org/10.1109/TSMCC.2007.905750.

Luo, W., Wang, F., 2003. Measures of spatial accessibility to health care in a GIS environment: Synthesis and a case study in the Chicago region. Environment and Planning B: Planning and Design 30 (6), 865–884.

Luo, F., Cao, G., Mulligan, K., Li, X., 2016. Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. Applied Geography 70, 11–25. http://dx.doi.org/10.1016/j.apgeog.2016.03.001.

Lv, M., Chen, L., Xu, Z., Li, Y., Chen, G., 2016. The discovery of personally semantic places based on trajectory data mining. Neurocomputing 173 (Part 3), 1142–1153. http://dx.doi.org/10.1016/j.neucom.2015.08.071.

Majid, A., Chen, L., Chen, G., Mirza, H.T., Hussain, I., Woodward, J., 2013. A context-aware personalized travel recommendation system based on geotagged social media data mining. International Journal of Geographical Information Science 27 (4), 662–684. http://dx.doi.org/10.1080/13658816.2012.696649.

McKenzie, G., Janowicz, K., 2015. Where is also about time: A location-distortion model to improve reverse geocoding using behavior-driven temporal semantic signatures. Computers, Environment and Urban Systems 54, 1–13. http://dx.doi.org/10.1016/j.compenvurbsys.2015.05.003.

McKenzie, G., Janowicz, K., Gao, S., Gong, L., 2015. How where is when? On the regional variability and resolution of geosocial temporal signatures for points of interest. Computers, Environment and Urban Systems 54, 336–346. http://dx.doi.org/10.1016/j.compenvurbsys.2015.10.002.

Nguyen, Q.C., Kath, S., Meng, H.-W., Li, D., Smith, K.R., VanDerslice, J.A., Li, F., 2016. Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity. Applied Geography 73, 77–88. http://dx.doi.org/10.1016/j.apgeog.2016.06.003.

Openshaw, S., 1984. The modifiable areal unit problem. Geo Books, Norwich, UK.

Ratcliffe, J.H., 2002. Damned if you don't, damned if you do: Crime mapping and its implications in the real world. Policing and Society 12 (3), 211–225. http://dx.doi.org/10.1080/10439460290018463.

Ratcliffe, J.H., 2004. Geocoding crime and a first estimate of a minimum acceptable hit rate. International Journal of Geographical Information Science 18 (1), 61–72. http://dx.doi.org/10.1080/13658810310001596076.

Roongpiboonsopit, D., Karimi, H.A., 2010a. Comparative evaluation and analysis of online geocoding services. International Journal of Geographical Information Science 24 (7), 1081–1100. http://dx.doi.org/10.1080/13658810903289478.

Roongpiboonsopit, D., Karimi, H.A., 2010b. Quality assessment of online street and rooftop geocoding services. Cartography and Geographic Information Science 37 (4), 301–318. http://dx.doi.org/10.1559/152304010793454318.

Rushton, G., Armstrong, M.P., Gittler, J., Greene, B.R., Pavlik, C.E., West, M.M., Zimmerman, D.L., 2006. Geocoding in cancer research: A review. American Journal of Preventive Medicine 30 (2), S16–S24. http://dx.doi.org/10.1016/j.amepre.2005.09.011.

Rydberg, J., Grommon, E., Huebner, B.M., Pleggenkuhle, B., 2016. Examining the correlates of sex offender residence restriction violation rates. Journal of Quantitative Criminology 1–23. http://dx.doi.org/10.1007/s10940-016-9303-z.

Seidl, D.E., Jankowski, P., Tsou, M.-H., 2016. Privacy and spatial pattern preservation in masked GPS trajectory data. International Journal of Geographical Information Science 30 (4), 785–800. http://dx.doi.org/10.1080/13658816.2015.1101767.

Strickland, M.J., Siffel, C., Gardner, B.R., Berzen, A.K., Correa, A., 2007. Quantifying geocode location error using GIS methods. Environmental Health 6 (1), 1–8. http://dx.doi.org/10.1186/1476-069x-6-10.

Sweeney, L., 2002. Achieving *k*-anonymity privacy protection using generalization and suppression. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (05), 571–588. http://dx.doi.org/10.1142/S021848850200165X.

Thamm, A., Anke, J., Haugk, S., Radic, D., 2016. Towards the omni-channel: Beacon-based services in retail. In: Abramowicz, W., Alt, R., Franczyk, B. (Eds.), Business Information Systems: 19th International Conference, BIS 2016, Leipzig, Germany, July, 6–8, 2016, Proceedings. Springer International Publishing, Cham, pp. 181–192.

Tompson, L., Johnson, S., Ashby, M., Perkins, C., Edwards, P., 2015. UK open source crime data: Accuracy and possibilities for research. Cartography and Geographic Information Science 42 (2), 97–111. http://dx.doi.org/10.1080/15230406.2014.972456.

Tribby, C.P., Miller, H.J., Brown, B.B., Werner, C.M., Smith, K.R., 2015. Assessing built environment walkability using activity-space summary measures. Journal of Transport and Land Use 9 (1), 187–207. http://dx.doi.org/10.5198/jtlu.2015.625.

Vandevijvere, S., Sushil, Z., Exeter, D.J., Swinburn, B., 2016. Obesogenic retail food environments around New Zealand schools: A national study. American Journal of Preventive Medicine 51 (3), e57–e66. http://dx.doi.org/10.1016/j.amepre.2016.03.013.

VanWey, L.K., Rindfuss, R.R., Gutmann, M.P., Entwisle, B., Balk, D.L., 2005. Confidentiality and spatially explicit data: Concerns and challenges. Proceedings of the National Academy of Sciences 102 (43), 15337–15342. http://dx.doi.org/10.1073/pnas.0507804102.

Wan, N., Zou, B., Sternberg, T., 2012. A three-step floating catchment area method for analyzing spatial access to health services. International Journal of Geographical Information Science 26 (6), 1073–1089. http://dx.doi.org/10.1080/13658816.2011.624987.

Wernke, M., Skvortsov, P., Dürr, F., Rothermel, K., 2014. A classification of location privacy attacks and approaches. Personal and Ubiquitous Computing 18 (1), 163–175. http://dx.doi.org/10.1007/s00779-012-0633-z.

Whitsel, E.A., Quibrera, P.M., Smith, R.L., Catellier, D.J., Liao, D., Henley, A.C., Heiss, G., 2006. Accuracy of commercial geocoding: Assessment and implications. Epidemiologic Perspectives & Innovations 3 (1), 1–12. http://dx.doi.org/10.1186/1742-5573-3-8.

Wieland, S.C., Cassa, C.A., Mandl, K.D., Berger, B., 2008. Revealing the spatial distribution of a disease while preserving privacy. Proceedings of the National Academy of Sciences 105 (46), 17608–17613. http://dx.doi.org/10.1073/pnas.0801021105.

Zandbergen, P.A., 2008. A comparison of address point, parcel and street geocoding techniques. Computers, Environment and Urban Systems 32 (3), 214–232. http://dx.doi.org/10.1016/j.compenvurbsys.2007.11.006.

Zandbergen, P.A., 2009a. Accuracy of iPhone locations: A comparison of assisted GPS, WiFi and cellular positioning. Transactions in GIS 13, 5–25. http://dx.doi.org/10.1111/j.1467-9671.2009.01152.x.

Zandbergen, P.A., 2009b. Geocoding quality and implications for spatial analysis. Geography Compass 3 (2), 647–680. http://dx.doi.org/10.1111/j.1749-8198.2008.00205.x.

Zandbergen, P.A., Barbeau, S.J., 2011. Positional accuracy of assisted GPS data from high-sensitivity GPS-enabled mobile phones. Journal of Navigation 64 (03), 381–399.

Zandbergen, P.A., Green, J.W., 2007. Error and bias in determining exposure potential of children at school locations using proximity-based GIS techniques. Environmental Health Perspectives 115 (9), 1363–1370.

Zandbergen, P.A., Hart, T.C., 2009. Geocoding accuracy considerations in determining residency restrictions for sex offenders. Criminal Justice Policy Review 20 (1), 62–90.

Zandbergen, P.A., Ignizio, D.A., Lenzer, K.E., 2011. Positional accuracy of TIGER 2000 and 2009 road networks. Transactions in GIS 15 (4), 495–519. http://dx.doi.org/10.1111/j.1467-9671.2011.01277.x.

Zandbergen, P.A., Hart, T.C., Lenzer, K.E., Camponovo, M.E., 2012. Error propagation models to examine the effects of geocoding quality on spatial analysis of individual-level datasets. Spatial and Spatio-temporal Epidemiology 3 (1), 69–82. http://dx.doi.org/10.1016/j.sste.2012.02.007.

Zhang, S., Freundschuh, S.M., Lenzer, K., Zandbergen, P.A., 2015. The location swapping method for geomasking. Cartography and Geographic Information Science 1–13. http://dx.doi.org/10.1080/15230406.2015.1095655.

Zheng, Y., 2015. Trajectory data mining: An overview. ACM Transactions on Intelligent Systems and Technology 6 (3), 1–41. http://dx.doi.org/10.1145/2743025.