

Primary and Secondary Methods of Data Collection in GIS/LIS

Khagendra Thapa and Robert C. Burtch

ABSTRACT. Data collection is by far the most expensive part of establishing a geographic information system (GIS) or land information system (LIS). Most of the pundits of GIS/LIS tend to confine data collection to the digitization or scanning of existing maps without regard to the maps' accuracy, age, or quality. Consequently, it has been found that the spatial accuracy of the information obtained from existing maps is sometimes worse than that derived from the satellite imagery. This paper recommends classifying the methods of spatial data collection as primary or secondary. Primary methods refer to data collection either directly from the field or from recent aerial photographs and satellite imagery. Secondary methods refer to data collection from existing sources such as maps, charts, documents, etc. These two major divisions are further broken down into a variety of methods.

Introduction

Spatial data collection is one of the fundamental steps in the creation of a land information system (LIS). However, it is also one of the least understood and is often overlooked, especially by those who claim to be experts in geographic information systems (GIS). This paper will use "LIS" and "GIS" interchangeably. Digitizing or scanning existing analog maps have been the main methods used in data capture for GIS/LIS (Cowen 1988; Calkins 1984; Marble et al. 1984; Peuquet and Boyle 1984). The objective of GIS/LIS is to store, retrieve, transform, display, aggregate, and analyze spatial data to solve the complex environmental and social problems faced by decision makers, planners, and local, state, and federal governments. However, decision making based on spatial data may be flawed if we digitize information from a 50-year-old map. The idea behind GIS/LIS is to be able to answer the complex questions quickly and accurately. But if our information is out of date and inaccurate, the advantages of GIS/LIS are lost. This paper classifies the methods of data collection as primary or secondary, and explains different methods used within each class.

Classification of Methods of Data Collection in GIS/LIS

Spatial data collection pertaining to GIS/LIS may be classified into primary and secondary methods. The primary methods of spatial data collection refer to deriving data directly from the field or from photographs (terrestrial or aerial) and satellite imagery.

Khagendra Thapa is a professor and Robert C. Burtch is an associate professor in the Surveying Engineering Program, Construction Department, College of Technology, Ferris State University, Big Rapids, MI 49307.

Secondary methods refer to processes in which data are derived from existing documents, such as maps, charts, graphs, etc.

The primary methods of data collection can be further classified into the following categories:

1. Techniques of geodesy
2. Techniques of surveying
3. Techniques of photogrammetry
4. Techniques of remote sensing

Techniques of Geodesy¹

It must be emphasized that all primary methods of data acquisition require a frame of reference (geodetic networks) established by using geodetic surveying techniques, such as traditional terrestrial observations (theodolite, EDM); extraterrestrial processes such as satellite Doppler, Global Positioning System (GPS), Very Long Baseline Interferometry, and laser ranging to the moon and satellites; photogeodesy; and inertial positioning. All these observations must be based on a common frame of reference defined by a rotational ellipsoid of suitable shape, size, and orientation. The datum for height is the geoid, which approximately coincides with the mean sea level. The details of how to establish a datum may be found in Bomford (1980). The essential role of geodesy is to provide the point from which all the calculations and analysis of spatial data may be performed.

It is important to understand that if the reference frame or geometry is incorrect, it may be impossible to perform the spatial analysis (e.g., overlay, intersection, addition, deletion, etc.) of data. Therefore, it is important to establish a geodetic network in the county before starting an LIS. In a study of the geodetic reference system, Epstein and Duchesneau (1984) point out that (1) there is no inherent economic value in the geodetic reference system, (2) any economic benefits will be found outside of the system itself,

and (3) to evaluate the system, other products that are dependent on the reference system must be identified.

The economic importance of the geodetic reference system lies in its ability to provide a framework whereby measurements can be taken and to offer a tool for universal compatibility (Epstein and Duchesneau 1984). The latter ability gives secondary and tertiary users of spatial data the opportunity to relate diverse information from different sources with a high degree of accuracy across all states. Thus two or more products produced by different people for different purposes can be combined into a single map. The economic benefits are derived from the fact that subsequent users avoid the additional costs of obtaining control required for their purposes. If an existing network has an insufficient number of control points, then it needs to be densified by including more control points to supplement the current network. The Committee on Geodesy (1983) recommended that monumented points be available at a spacing of between 0.3 to 0.8 km in urban areas and between 1.6 to 3.2 km for rural areas. Because of the availability of kinematic GPS (Goat 1989), such stringent requirements of geodetic control may be relaxed. On the other hand, monumented control is the cornerstone of the judicial cadastre and our land tenure system.

Techniques of Surveying

Terrestrial Surveying. Terrestrial surveying is the traditional method of spatial data collection. Before the advent of photogrammetry, surveying was the sole method used. The plane table equipped with an alidade was the primary method of topographic surveying. Conventional surveying requires skilled and experienced field personnel who can not only observe angles, distances, and height differences accurately, but also can check the accuracy of the work right in the field before dismantling the instruments from a station.

Introduction of the electronic theodolite, or the so-called total station, has made conventional surveying easier and faster. Electronic theodolites operate much the same way as optical theodolites, and range in accuracy from 0.5 seconds to 10 seconds (Kavanagh and Bird 1989). When an electronic theodolite is integrated with or attached to an EDM, the resultant instrument is called a total station. It is sometimes referred to as an automatic total station because of the electronic system used in reading both horizontal and vertical angles as well as distances. Additionally, the total station can be interfaced with an electronic data collector, which stores all of the field data. Surveying is a reliable and accurate method of data col-

lection for GIS/LIS. However, it is not an economical method for areas larger than about 25 acres. Since the accuracy of surveying is a function of not only the equipment but also the observational procedure, atmospheric conditions, distance between points, and number of occupied stations, the errors over a block can accumulate quickly. Generally, photogrammetric methods are used to map large areas.

Control Surveying—GPS. GPS was initiated by the U.S. Department of Defense in 1973 as a space-based radio positioning and navigation system. The GPS constellation of NAVSTAR (NAVigation by Satellite Timing And Ranging) satellites was originally proposed to have 24 satellites in three orbital planes inclined at 55 degrees to the equator, with a period of 12 hours. However, due to the rigorous financial scrutiny at higher levels of government, the Air Force has been directed to develop the NAVSTAR GPS with an 18-satellite constellation plus three spare ones. In addition, seven satellites will always be ready on the ground.

When fully operational, sometime in 1993, the satellites will be placed in six equally spaced (in longitude) orbital planes with an inclination of 55 degrees to the equator. This will ensure visibility of at least four satellites at any time from almost anywhere in the world. The satellites will travel in nearly circular orbits at an altitude of approximately 20,200 km above the earth, with a time period of 11 hours and 58 minutes. The constellation of 18 satellites along with three active spares is known as the space segment of GPS.

GPS provides very accurate, three-dimensional position and velocity information. Moreover, it can also provide Coordinated Universal Time to anyone with a suitable receiver. The control segment of the NAVSTAR system consists of the five known control stations located in Hawaii; Colorado Springs, Colorado; Ascension Island; Diego Garcia; and Kwajalein. These control stations track all the GPS satellites and predict their orbits. One of the stations (Colorado Springs) is the Master Control Station (MCS). All control stations transmit data to MCS, which processes it, computes the satellite ephemeris and clock corrections, and transmits an updated ephemeris to the satellites.

There are two main points of observation used in GPS: pseudorange and carrier phase. Pseudorange is the distance between the satellite and the receiver, plus a small corrective term due to receiver clock errors. The carrier phase is the difference between the phase of the carrier signal of the satellite and the phase of the local oscillator within the receiver at the epoch of measurement (Leick 1987).

Every GPS satellite transmits in two L-band frequencies: $L_1 = 1575.4$ MHz and $L_2 = 1227.4$ MHz. These two frequencies may be modulated by various

signals. The advantage of using two L-band frequencies is that the first-order ionospheric effect may be eliminated.

Each GPS satellite has a high-precision oscillator with a fundamental frequency (f_0) of 10.23 MHz. All other frequencies ($L_1 = 154f_0$ MHz = 1575.4 MHz, $L_2 = 120f_0$ MHz = 1227.6 MHz and C/A code = $f_0/10 = 1.023$ MHz) are derived from it. C/A code is the course acquisition code (sometimes called standard code) and P-code is the precision code. Both are pseudorandom noise codes.

The L_1 carrier is modulated with both C/A and P-code, while L_2 is coded only with P-code. Both frequencies carry a D-code, which consists of the broadcast satellite message such as health and position of the satellite.

Some receivers require both P-code and C/A code while others work only with C/A code. It should be noted that C/A code is broadcast every millisecond and is available to anyone with a GPS receiver. However, the P-code is restricted to U.S. military and allies. The accuracy of surveying by GPS satellites depends on receiver satellite geometry, tropospheric and ionospheric effects, accuracy of the satellite ephemeris, and processing software used. Relative positioning accuracy using the currently available receivers ranges from about 5 mm to several centimeters.

In order to receive the signals from the GPS satellites, the horizon must be unobstructed above 15 degrees at the antenna location. This means that GPS receivers may not be used in cities with several high-rise buildings. Moreover, these receivers cannot be used in densely forested areas, unless the forest is cleared around the antenna location. The cost of a receiver is still too high for a practicing surveyor with a small business. In addition, the space segment is not yet complete. This, in turn, restricts the observing time to about four to six hours per day, depending on the location on the earth.

There are three different methods of positioning using GPS: (1) static, (2) kinematic, and (3) pseudokinematic. Static relative positioning involves two or more receivers in the survey. One can routinely perform positioning within millimeters, provided the distance between the receivers is not more than several kilometers. Static positioning requires that we observe at least one hour from each station. The time constraint is due to the satellite geometry, which will be poor if we only observe for a short period of time.

The basic observable used by GPS geodetic receivers is the carrier phase measurements. However, if the incoming signal is briefly interrupted, then this creates a cycle slip. When the contact is reestablished, the phase measurements will resume as if nothing has happened. However, the receiver loses track of the integer number of cycles. The slippage of cycles

can be detected and recovered during postprocessing of the observations, provided there are enough observations. Postprocessing of observations in GPS positioning involves double-difference phase readings (i.e., simultaneous phase measurements differenced once between two stations and again between two satellites). Double-difference observations eliminate the time bias both at the satellite and receiver. One advantage of the static mode of positioning is that there are enough observations to detect and correct for cycle slips.

The kinematic method of GPS positioning was first proposed by Remondi (1985), and is capable of providing geodetic accuracy with less than one minute of observations. This method is based on the fact that the theoretical value of the integer ambiguities can be calculated if the coordinates of the end points of the baseline are known. An integer ambiguity is the number of full wavelengths contained in the distance between the receiver and the satellite.

In the kinematic method of positioning, one can resolve the integer ambiguities in the following three ways (Goad 1989): (1) observing a baseline in the area of the kinematic survey, (2) observing an unknown baseline using static method, and (3) swapping antenna.

Integer ambiguity is first determined by adopting one of these methods, and then one receiver is kept static on the known point while the other receiver(s) moves to different, unknown stations, while making sure that at least four satellites are being tracked during the transportation of the roving receiver from point to point.

The kinematic method of positioning has two drawbacks (Ashkenazi and Summerfield 1989): It requires a geodetic receiver capable of taking carrier frequency phase measurements when it is moving, and it requires cycle-slip-free observations on at least four satellites during the move. If a cycle-slip occurs during the move, the integer cycle ambiguity cannot be recovered. The pseudokinematic method devised by Ashkenazi and Summerfield (1989) overcomes these disadvantages. In the pseudokinematic method, the known baseline with two stations is occupied in the first minute and then again at the end.

The unknown stations are occupied twice within a time interval of an hour or so. However, only a few observations (about a minute) need be taken at each visit. The advantages of this technique are that we can use geodetic GPS receivers that are not capable of tracking the satellites during the movement. In addition, there is no need to keep a lock on the satellites during the transportation of the roving receiver. Both pseudokinematic and kinematic methods of GPS positioning give centimeter-level accuracy.

Control Surveying—Inertial Positioning Systems (IPS). Surveying may also be performed using the refinements in inertial navigation and guidance systems. IPS is based on a computer-controlled gyrostabilized inertial platform and three orthogonally oriented accelerometers mounted on the platform. The accelerometers perform the function of spatial distance-measuring instruments, while the gyroscope operates as an angle-measuring instrument. An IPS consists of the following units: (1) inertial measurement, (2) data processing, (3) control and display, (4) data storage, and (5) power supply.

An IPS survey begins at a known point. The X, Y, Z coordinates or the latitude, longitude, and elevation of the point are input into the computer and the IPS measuring unit is allowed to orient itself at the initial point. As the vehicle moves, the three accelerometers measure the three-dimensional displacements, while the gyroscopes keep the platform oriented properly.

When the vehicle moves to the subsequent points in the traverse, errors in the system will accumulate. These system errors in turn will cause positional errors. If the vehicle is stopped, the acceleration and the velocity should be equal to zero. However, because of the changes in the local vertical and the system errors, the accelerometer outputs will give false values even after the system is stopped for a few minutes. The process of stopping the vehicle to measure the zero velocity is called zero velocity update (ZUPT). This process should be carried out every three to five minutes, depending on the accuracy specifications of the project. Higher accuracy in the results may be achieved by using more frequent ZUPTs. The IPS method is interpolative, and must be closed between two known points or should close back to the starting point.

Accuracies from an inertial survey are generally acceptable for most control survey and densification operations, since they generally meet second-order accuracy. Relative positioning on the order of 13 to 18 cm (0.4 to 0.6 ft) or better are possible. There are a number of advantages to the inertial surveying approach. First, it is a rapid surveying system. Second, it is not dependent upon line of sight. Third, there is the property of homogeneity, which means that the distribution functions are identical. Fourth, the results are given in coordinates. Fifth, the survey costs less.

Techniques of Photogrammetry

According to Slama (1980), photogrammetry may be defined as the art, science, and technique of obtaining reliable information about physical objects and the environment through recording, measuring, and

interpreting photographic images and patterns of electromagnetic radiant energy and other phenomenon. In a broad sense, photogrammetry involves the following: (1) photographing an object or obtaining the image of the objects, (2) measuring the positions of the object from the imagery and processed photographs, and (3) reducing the imagery to some useful form, such as a digital or topographic map.

Photogrammetry has been the major method of compilation of topographic maps based on measurements and information from aerial and space photographs over several decades. The process of photogrammetry for collecting the digital or analog topographic data consists of the following major steps: (1) planning and acquisition of aerial photography, (2) ground control, (3) aerotriangulation, (4) map compilation, and (5) editing.

Success of the photogrammetric method of data collection depends on the acquisition of proper photographs. Several factors are taken into account when planning for aerial photography (Moffit and Mikhail 1980; Slama 1980). Some of the important factors are

1. Purpose of photography
2. Scale of photography
3. Overlap between exposures
4. Allowable scale variation
5. Optical and mechanical characteristics of the camera
6. Film base and emulsion type used
7. Flying height used
8. Direction of orientation of topography
9. Relief displacement

Ground-control points that can be identified on aerial photographs are needed to provide control for photogrammetry. The purpose of the ground control in photogrammetry is to establish the orientation and position of each photograph in space relative to the ground. The position of the ground control enables the photographs to be used for model information, aerotriangulation, and subsequent compilation of planimetric and topographic maps. Ground control is usually provided by traversing, triangulation, leveling, GPS surveying, Transit satellites, and IPS. Currently, researchers (Merchant 1989) are trying to perform photogrammetric mapping without ground control. This will involve the location of the exposure station coordinates by kinematic GPS receivers on board the aircraft.

Orthophotography. A rectified photograph is an aerial photograph from which the image displacement due to tilt has been removed, transforming it into an equivalent vertical photograph. While rectification only compensates image displacement due to tilt, the

process of differential rectification eliminates both tilt and relief displacement of images resulting in a photographic image called an orthophoto. Orthophotos are thus photographic maps showing the location of images in their proper orthographic position and, because they are geometrically equivalent to line maps, they can be used to make measurements without correcting for image displacement.

Orthophotos are more expensive than a simple rectification of aerial photographs, but they are essential if the photographic map is of an area where terrain height varies considerably and if the map must meet National Map Accuracy Standards. Despite the costs, it has been reported that a 3-to-1 cost ratio can be found between orthophotos and line maps. Orthophotos are more expensive because instruments similar to those used in line map production, in which there is a heavy operator/instrument dependence, are required to control the height of the measuring mark used in the differential rectification process. But they have an added advantage over rectification, in that contour information can be added. Some image displacement still exists in orthophotography because the slit (a small opening in the viewing platen used in differential rectification) is larger than the differential changes that theoretically should occur. This may cause minor problems where sheer vertical changes exist, such as on buildings or cliffs. Finally, there is a difference in orthophotomaps and orthophotoquads, as defined by the U.S. Geological Survey. Orthophotoquads are black-and-white images in the quad format. They contain no contours and little cartographic notation. Orthophotomaps, on the other hand, are color-enhanced and include contour lines and names.

Photogrammetric Triangulation. Photogrammetry has been described by Ackerman (Chrzanowski and Dorner 1978) as the "tool par excellence" in an integrated survey system. Indeed, the advantages of photogrammetry lie in the multipurpose medium of the imaged point. This multipurpose nature of the photograph allows for the direct production of photographic products through rectification, more refined planimetric and topographic maps via restitution stereoplotters, computation of accurate ground coordinates for control densification projects using analytical techniques, delineation and identification of land-use and development patterns, and convenient storage of detail that existed at the time of photography that can be recalled later or compared with other photography at different dates. This means that the use of photogrammetry should be considered with a view to economy in the various applications of GIS/LIS. The cost advantages of photogrammetry are most pronounced when the number of points needed to

be measured is great, the level of detail to be plotted is dense, or the area is large. As with other systems, the costs increase with the demand for accuracy.

Photogrammetric triangulation can be performed a number of ways, with different accuracies. By far, the bundle adjustment method is the most precise. This is a simultaneous adjustment of the bundle of rays from all photographs to all measured ground points in a process that also recovers the orientation parameters simultaneously. A refinement to this process is the bundle adjustment with self-calibration. In this process, additional unknown residual systematic errors are carried along as unknown parameters in the adjustment process. This approach is very nearly a systems adjustment that provides not only the desired coordinates, but also an evaluation of the system performance within the whole process. There is a significant improvement in accuracy over the basic bundle adjustment, especially in blocks where there is only relatively sparse control available (Brown 1979). There are problems with self-calibration in view of writing specifications (Slama 1980). The first is defining or specifying the procedures to be used. This is due to the wide variety of methods employed in self-calibration. Second, there is not enough knowledge of the effects upon the other specified quantities. For this reason, the National Geodetic Survey has not included this method within its proposed specifications.

Lesser methods of treating photogrammetric observations are also available for control determination. They are derived from the analog triangulation procedures used earlier in photogrammetric history, and are based upon the best equipment and adjustment procedures available at that time. They are still in use in many operations. The first is a polynomial strip assembly in which photogrammetric models, relatively oriented, are adjusted into strips and then into blocks using a step-wise approach. A refinement to this is the independent model approach.

One of the major advantages of the photogrammetric bundle adjustment is that the accuracy of the computed ground coordinates is independent of the distance between the points. This means that the planimetric accuracies are almost uniform throughout the project, therefore requiring less primary control than conventional terrestrial surveying methods.

Analytical photogrammetry is concerned with the solution of photogrammetric problems mainly by pure mathematical modeling, using simple but highly precise coordinates of the image points. The cost advantage of analytical photogrammetry is significant. Conservative estimates show this to be 3 to 1 over first-order ground traversing. When one considers that height control can also be performed as a byproduct of the adjustment process, the economic advantage

increases even more. Further, the primary control requirements are such that a distribution of control points about the block periphery corresponding to every five photos is necessary. Not only will this ensure near uniformity of the planimetric accuracies, it also means that less control is required than in conventional techniques. In fact, the cost savings for this latter method may be sufficient to pay all of the photogrammetric densification costs, excluding those associated with monumentation (Brown 1979).

Digital Data from Photogrammetric Compilation. The main advantage of digitization from photogrammetric instruments during the stereocompilation is that accurate positions of the features can be obtained from the stereo model. In existing maps, features may have been generalized, omitted, or displaced. In addition, the photogrammetric method of obtaining digital data has the following advantages (Allam and Wong 1986; Elgarf 1986; Konecny 1980):

1. Data can be obtained in three dimensions (i.e., both planimetry and height information can be digitized)
2. It is more accurate than digitizing existing maps
3. The information portrayed in aerial photographs is current
4. Digital data collection using photogrammetric methods is flexible in terms of scale, data base, and choice of features to be digitized
5. The existing data base can be updated using photogrammetric methods
6. Digital data acquisition of heights by points, profiles, and contour lines can be performed according to distance, time, or coordinate increment

Gruen (1989) describes digital (photogrammetric) stations, which could be able to process the following kinds of data: (1) mono, stereo, and multi-image arrangements; (2) terrestrial, aerial, and satellite imagery; (3) different kinds of imaging sensors, sensor combinations, and nonimaging sensor data; and (4) digitized photographs and digital scenes.

In addition, digital stations can provide data in analog and digital form, includes data-base functions, and can manage large amounts of data. These stations promise to be useful in digital data collection for land information. It should be noted, however, that the photogrammetric method cannot be applied in areas where there is a dense vegetation cover, such as tropical rain forests. Further, cadastral boundary lines and place and road names are not visible in aerial photographs. Moreover, the type of road surface, and the location of underground and certain overground utilities, must be determined by field visits.

Techniques of Remote Sensing

Lillesand and Kiefer (1987) define remote sensing as the art and science of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area or phenomenon. Note that this definition of remote sensing includes photogrammetry, which was discussed in the previous section. In this section, we restrict our discussion of remote sensing as a method of data collection for GIS using electromagnetic energy sensors operated primarily from spaceborne platforms. Remote sensing basically involves the following processes (Goodenough 1988): (1) data acquisition, (2) preprocessing, (3) analysis, (4) accuracy assessment, and (5) information distribution.

The following are the elements of data acquisition (Lillesand and Kiefer 1987):

1. Propagation of energy through the atmosphere
2. Energy interaction with earth surface features
3. Retransmission of energy through the atmosphere
4. Airborne and spaceborne platforms
5. Generation of sensor data in digital form

Sensor-related preprocessing corrections, such as radiometric and geometric corrections, generally are performed by the data archiving and distributing agencies. However, in order to use the data for a GIS/LIS, additional radiometric, geometric, and atmospheric corrections need to be applied. These specialized corrections include (Goodenough 1988) corrections for radiometric distortion due to view angle, geometric correction for terrain relief, projection of imagery to a variety of map projections, and atmospheric corrections using meteorological data.

When integrating GIS/LIS and remote-sensing data, thematic data and attributes may be applied to guide the analysis and classification of remote-sensing data. Accuracy may be assessed based on class statistics and selected test sites derived from ground reference data. Finally, the information may be distributed in the form of maps, computer tapes, and images.

Integration of Remote-Sensing Data with GIS/LIS. Integration of GIS/LIS and remote-sensing data generally involves different data representations. In order to make the integration useful, there must be some kind of geometric registration between the different data sets. In order to use remote-sensing data, the existing GIS/LIS data must be in raster form. However, many agencies using GIS/LIS are unwilling to change the format of the existing large data sets. Algorithms are needed to convert raster to vector and vice versa.

GIS has many classes of information (i.e., layers or types of information). For example, a forest cover map may have as many as 1,800 different classes. However, most remote-sensing data processing systems can handle only up to 256 classes. Ironically, Goodenough (1988) found that most existing GIS have worse geometric accuracy than the data obtained from remote-sensing sources such as the Landsat Thematic Mapper and SPOT imagery. This is because the spatial data in GIS generally are obtained from digitizing existing maps that likely are inaccurate. Some of these digitized maps may be 50 years or more out of date. On the other hand, panchromatic SPOT imagery is accurate to 8.8 m (Gugan and Dowman 1988). Thematic mapper imagery is accurate to 7.2 m in planimetry.

Another problem (Goodenough 1988) is that the GIS class labels may not correspond to detectable remote-sensing classes. For instance, satellite data simply do not have the resolution to detect features such as small creeks overgrown with trees. Contextual information may be required for features such as parks, wooded areas, cemeteries, etc.

Generally, there are three steps involved in the integration of remote-sensing data in GIS/LIS: (1) extraction and processing of the GIS/LIS data, (2) analysis and combination of GIS/LIS and remote-sensing data, and (3) generation of the updated GIS/LIS.

Secondary Methods of Data Collection in GIS/LIS

Secondary methods of data collection may be classified as manual, semiautomatic (line following), and automatic (scanning). The manual method of digitization (i.e., conversion of analog maps, orthophotos, charts, etc., into digital form) is by far the simplest and the most common method of digitization. It consists of the following components: (1) digitizing tablet on which the map to be digitized is placed, (2) tracking device (e.g., cursor or pen), and (3) output device (e.g., floppy disk or hard disk).

In addition to these components, manual digitization also requires software that displays, stores, and performs the required transformations. The manual method of digitization is performed by simply moving the cursor along the line to be digitized and pressing the cursor button at all the points that need to be recorded to retain the shape of the line.

The graphic digitization systems may be classified into arm-type or restrained cursor and free cursor. The free-cursor type is free in the sense that there is a light, flexible wire attached to the cursor that does not obstruct the movement of the cursor on the tablet. Arm-type digitizers use either rotary or linear encoders. The rotary encoders are more susceptible to

errors and have an accuracy of 0.005" (Cameron 1982). Linear encoders are more stable and provide higher accuracy.

The free-cursor type of digitizers are categorized into electromechanical and electronic. However, the electromechanical type of digitizers are no longer manufactured because they are expensive to maintain.

There are two types of electronic free-cursor digitizers: incremental and absolute. The incremental electronic free-cursor digitizer lost its popularity because if the cursor is lifted from the table the reference to the origin is lost. According to Cameron (1982), there are four kinds of absolute electronic free-cursor digitizers: (1) sonic, (2) magnetostrictive, (3) electrostatic, and (4) electromagnetic.

The sonic digitizer, which is based on the generation of acoustic impulse, is not capable of high accuracy and, therefore, is not used in the mapping field. The accuracy of this digitizer is affected by air temperature and humidity.

The magnetostrictive digitizer is influenced by magnetic fields. In addition, it needs to be periodically remagnetized to keep a uniform magnetic field and minimize errors. Moreover, magnetostrictive digitizers also develop dead spots that cannot be remagnetized. Therefore, these digitizers do not have the reliability and stability required in GIS/LIS fields.

Electrostatic digitizers are affected by humidity, moisture on the work surface, pencil marks, and paper. As a result, electrostatic digitizers are being replaced by electromagnetic digitizers. Electromagnetic digitizers are very accurate and are not affected by temperature, humidity, or mechanical vibrations. These digitizers have a resolution of 0.001", with an accuracy of ± 0.003 " (Cameron 1982). The following digitizer parameters must be considered when selecting a digitizer for GIS/LIS applications (Cameron 1982): (1) stability, (2) repeatability, (3) linearity, (4) resolution, (5) skew, and (6) accuracy. The overall accuracy of a digitizer may be affected by the following factors (Cameron 1982): (1) cursor orientation, (2) temperature variation, (3) humidity, (4) cursor height, (5) drift, and (6) calibrated electronics.

Automated line-following techniques attempt to replace the hand and eye of a manual digitizing operator. This method of digitization is widely used for digitizing vector data (linear features) such as rivers, roads, coastlines, contour lines, etc. The operator of the line-following instrument places the pointer on a line to be digitized. Then the pointer automatically follows the line until one of the following situations is encountered (Faust 1987):

1. The pointer arrives at the edge of the map.
2. The line closes back at the starting point.
3. An impasse is encountered such that the line-

following software cannot decide which direction to go. The impasse may be caused by line weight problems in the manuscript, breaks in the line for labeling (e.g., contour values) or by lines that are so close to each other that the software is unable to decide.

To minimize the error in the line-following process, different map separates (e.g., contour lines, hydrographic features, roads, etc.) are used. Map or image scanning involves a computer-controlled instrument equipped with optics and detectors that can create the digital data from the document. These scanners need to be precise between 0.001" to 0.006" (Peuquet and Boyle 1984).

The scanners may be divided into the following four categories: (1) drum scanners, (2) photodiode strip scanners, (3) laser beam scanners, and (4) video camera scanners. The drum scanner is widely used in cartography to produce color separations for printing purposes. The document to be scanned is attached to the surface of the drum. An optical head is used to scan the document in one direction, and the drum motion provides the scanning in the other direction. The scan-line width may be controlled over a wide range by means of optical settings. In addition, the number of lines per inch is adjustable. However, it should be noted that the higher the resolution of the scanner, the longer it will take to scan the document. Digitization of a color map is achieved by the use of filters and by scanning the same document three times.

It is possible to construct photodiode strips with more than 1,000 sensors at a spacing of 0.001". The photodiode strip scanner operates by wiping these strips along the length of the document and repeating the process for adjacent strips. The scan time is about one hour for a document size of 24 by 30 inches using half-inch-wide strips at a resolution of 0.001". Laser scan systems may offer some significant advantages in speed and time. These scanners could be strips of full width with a resolution of 0.001".

Digitizing of documents (imaging) may be performed in real time at about a one-second rate using video camera scanners. Resolution of a video camera is normally 512 by 512 pixels. However, a few cameras with higher resolution are becoming available. It should be noted that the video camera digitized images will require geometric correction. This in turn requires ground-control point identification and selection. Other corrections such as lens distortion introduced in the video camera system may need to be applied. The camera is equipped with a charge coupled device. Finally, it should be pointed out that scanning methods of digitization may take only a few seconds to scan a document, but the vectorization

and addition of other information, such as attributes (e.g., contour heights), will take several minutes per document.

Advantages and Disadvantages of Secondary Methods of Data Collection

GIS is concerned with handling spatial information, such as overlay, intersection, and matching. These operations cannot be performed unless the geometry is correct. However, the geometry will not be correct if the data are collected from existing maps alone, for the following reasons:

1. Generally, maps are at least five years old; therefore, the information is not current.
2. Maps may be at different scales.
3. Maps may be on different datums, such as NAD 27 or NAD 83.
4. Maps may be at different projections, such as Lambert conformal or Transverse Mercator.
5. Not all maps satisfy National Map Accuracy Standards.
6. Maps may not contain all of the information desired.
7. Some accuracy is lost in the process of digitization or scanning of maps. Moreover, maps are generalized and some features may have been displaced and/or exaggerated, which introduces very significant positional error.
8. Paper maps do not really have the scale that is stated in the map because of the dimensional instability of the paper due to the change in temperature, humidity, age, and use or abuse of the map. That is why a bar scale is provided on a map.

The only advantages in using spatial information from maps are that it is much less expensive than field data collection and it is much faster than primary methods, although it takes a significant amount of time to scrub and clean the digitized data to get them in a usable form.

Conclusions

Data collection is by far the most expensive component of GIS/LIS. There are primary and secondary methods of data collection. The primary methods of data collection involve the techniques of geodesy, surveying, photogrammetry, and remote sensing. The secondary methods of data collection are classified as manual, semiautomatic, and automatic. Secondary methods of data collection may be faster and cheaper, but the data produced may not be current or accurate. The question of whether to use secondary methods of data collection depends on the purpose of the GIS.

For example, if one is interested in soil classification, it would be absurd to collect data using the primary methods.

ACKNOWLEDGMENT

We are very grateful to the Royal Institution of Chartered Surveyors (RICS) and the faculty research committee of the Ferris State University for providing us with research grants. Part of the funding was used to prepare this paper. We express our sincere thanks to Dr. Charles R. Schwarz, editor of this journal, for the meticulous, last-minute review of this paper.

REFERENCES

- Allam, M.M., and C.K. Wong. 1986. "An Integrated Digital Mapping System for Data Capture and Editing." *Proceedings of the Symposium on Photogrammetric and Remote Sensing Systems for Data Processing and Analysis*, vol. 26, part 2, pp. 511-520. Baltimore, Maryland: International Archives of Photogrammetry and Remote Sensing.
- Ashkenazi, V., and P.J. Summerfield. 1989. "Rapid Static and Kinematic GPS Surveying: With or Without Cycle Slips." *Land and Minerals Surveying*, vol. 7, no. 10, pp. 489-494.
- Bomford, G. 1980. *Geodesy*. London: Oxford University Press.
- Brown, D.C. 1979. "Photogrammetric Densification of Geodetic Nets, a Technology Whose Time Has Passed." Paper presented at the Aerial Triangulation Symposium, University of Queensland, Australia.
- Calkins, H.W. 1984. "Creating Large Data Files from Mapped Data." *Basic Reading in Geographic Information Systems*. Williamsville, New York: SPAD Systems Ltd.
- Cameron, E.A. 1982. "Manual Digitizing Systems." *Technical Papers of the 1982 ACSM/ASPRS Annual Convention*, Denver, Colorado.
- Chrzanowski, A., and E. Dorrer. 1978. *Proceedings, Standards, and Specifications for Integrated Surveying and Mapping Systems*. Munich, Germany: Hochschule der Bundeswehr München.
- Committee on Geodesy. 1983. *Procedures and Standards for a Multipurpose Cadastre*. Washington: National Research Council, National Academy Press.
- Cowen, D.J. 1988. "GIS Versus CAD Versus DBMS: What Are the Differences?" *Photogrammetric Engineering & Remote Sensing*, vol. 54, no. 11, pp. 1551-1555.
- Elgarf, T.M. 1986. "The Role of Mapping Company in (GIS) Field." *Proceedings of Geographic Information Systems Workshop, American Society for Photogrammetry and Remote Sensing*, Atlanta, pp. 189-197.
- Epstein, E.F., and T.D. Duchesneau. 1984. *The Use and Value of Geodetic Reference System*, pp. 36. Rockville, Maryland: Federal Geodetic Control Committee.
- Faust, N.L. 1987. "Automated Data Capture for Geographic Information Systems: A Commentary." *Photogrammetric Engineering & Remote Sensing*, vol. 53, no. 10, pp. 1389-1390.
- Goodenough, D.G. 1988. "Thematic Mapper and SPOT Integration with a Geographic Information System." *Photogrammetric Engineering & Remote Sensing*, vol. 54, no. 2, pp. 167-176.
- Goad, C.C. 1989. "On the Move With GPS." *P.O.B.*, vol. 14, no. 4, pp. 36-44.
- Gruen, A.W. 1989. "Digital Photogrammetric Processing Systems: Current Status and Prospects." *Photogrammetric Engineering & Remote Sensing*, vol. 55, no. 5, pp. 581-586.
- Gugan, D.J., and I.J. Dowman. 1988. "Topographic Mapping from Space Imagery." *Photogrammetric Engineering & Remote Sensing*, vol. 54, no. 10, pp. 1409-1414.
- Kavanagh, B.F., and S.J.G. Bird. 1989. *Surveying Principles and Applications*, second edition. Englewood Cliffs, New Jersey: Prentice Hall.
- Konecny, G. 1980. "How the Analytical Plotter Works and Differs from an Analog Plotter." *Proceedings of the Analytical Plotter Symposium and Workshop*, Reston, Virginia, pp. 31-75.
- Leick, A. 1987. "Positioning 2001." *Surveying and Mapping*, vol. 47, no. 3, pp. 181-189.
- Lillesand, T.M., and R.W. Kiefer. 1987. *Remote Sensing and Image Interpretation*, second edition. New York: John Wiley and Sons.
- Marble, D., et al. 1984. "Development of a Conceptual Model of the Manual Digitizing Process." *Basic Readings in Geographic Information Systems*. Williamsville, New York: SPAD Systems Ltd.
- Merchant, D.C. 1989. "Positioning the Photo Aircraft by the Global Positioning System Photogrammetric Challenges and Promises." *Technical Papers of the 1989 ASPRS/ACSM Annual Convention*, Baltimore, Maryland.
- National Research Council. 1983. *Procedures and Standards for Multipurpose Cadastre*, p. 197. Washington: National Academy Press.
- Peuquet, D.J., and A.R. Boyle. 1984. *Raster Scanning, Processing and Plotting of Cartographic Documents*. Williamsville, New York: SPAD Systems Ltd.
- Remondi, B.W. 1985. "Performing Centimeter Accuracy Relative Surveys in Seconds Using GPS Carrier Phase." *Proceedings of the First International Symposium on Precise Positioning*, Rockville, Maryland.
- Slama, C.C. 1980. *Manual of Photogrammetry*. Bethesda, Maryland: American Society for Photogrammetry and Remote Sensing.
- Schwarz, K.P. 1980. "Inertial Surveying System—Experiences and Prognosis." *The Canadian Surveyor*, vol. 34, no. 1, pp. 41-53.