Digital Capture of Geographic Feature Data for Surveys

Sarah M. Nusser

Department of Statistics and Center for Survey Statistics and Methodology, Iowa State University
219 Snedecor Hall, Ames, IA 50011-1210, nusser@iastate.edu

Abstract

In some surveys, it is of interest to record geospatial data such as coordinates for housing unit locations, road networks, agricultural field boundaries, and water body boundaries as part of the survey data collection process. These data objects are represented as shape data (i.e., points, lines, polygons), and are inherently different from coded and fixed format questionnaire responses. Geographic information system (GIS) software can be used to record shape data, but the differences between a GIS software environment and that of a standard computer-assisted survey instrument are substantial. Questions arise in balancing the linear structure of protocol enforcement with the iterative interactions that occur when working with geographic information; how to represent and configure feature editing environments for unskilled users; and whether constraints should be placed on the digitizing process to balance the accuracy of geographical feature delineation with the time spent recording the data. GIS software environments for survey data collection were explored via investigations with two agencies in the US Department of Agriculture. One study was conducted with National Agricultural Statistics Service field interviewers, who had not previously been exposed to computer-assisted data collection. Enumerators used pen-based tablet computers and a simplified ESRI ArcView application to delineate field boundaries on digital photographs during interviews with farm operators. Data were collected from observers and enumerators on the interview experience and software use. In a second study, Natural Resource Conservation Service staff were trained in the use of GIS, but varied in the depth of their prior experience with GIS software. Participants delineated water bodies on high resolution photographs using a simple ESRI ArcView application on desktop computers. The goals were to investigate the properties of area measurements at different scales in relation to work effort and to gain information on how users approached the delineation task in this environment through observation and digital capture of software sessions. We present results from these studies and on-going software development efforts.

1. Introduction

The use of geographic information in survey data collection is being explored by many survey organizations. Research is being conducted on methods to use global positioning system (GPS) track data during survey operations and to capture coordinates that depict the location of housing units (Murphy and Nusser 2003). For surveys that collect data on land characteristics, protocols and software are being explored for collecting geographic features such as agricultural field boundaries, water body boundaries, stream courses, and transportation networks (Nusser 2004).

Geographic data for locations and feature extents are represented in a *shape* format, i.e., as a point, line or polygon. Recording and manipulating shape data is strikingly different from working with alphanumeric codes or even free text captured via a questionnaire. For example, most questionnaire response data obtained are derived from a limited response set (e.g., pre-coded items) or specified format (e.g., number of months for time periods). While the type of geographic features to be collected will generally be known, the actual location, shape and extent of the feature to be delineated are not generally limited to a finite set of options. Further, although boundaries or linear features are the focus of data collection, the delineated feature is often used to generate a secondary numeric survey measurement, such as the area or length of the feature or its relationship to other geographic features.

The types of software tools required to collect these data are also different from traditional computer-assisted survey instruments (CASIs). In a standard CASI, a pre-defined and relatively static view of the question item and response options is presented. The most common interaction is to enter a response and navigate to the next question. When working with geographic data, the shape delineation is typically depicted against a contextual backdrop (e.g., map or photographic image). In contrast to interacting with a question and response screen in a fixed layout, the data collector “grabs and moves” the geographic data. Manipulating geographic data involves complex forms of scrolling known as *panning* (arbitrary lateral movement by grabbing the image or using an eight-way scroll arrow) and *zooming* (identifying a focused region for viewing or using a tool to set the data scale at a different resolution). Further, instead of selecting a code from a fixed list, shape data for a linear or area feature are created, e.g., via an *edit* tool, by recording a linear sequence of connected points to depict the feature.

Editing of recorded geographic feature data occurs in two distinct steps. First, the initial draft of the feature shape commonly requires modification once the draft delineation has been compared to the image backdrop or checked against the respondent's perceptions. This kind of data modification is relatively rare in well-structured survey instruments, and the edits required to refine geographic features are more intricate than corrections to alphanumeric data. Second, as with questionnaire data, edit checks may also be applied to feature data recorded by the data collector. Edit rules may include format checks (e.g., closed polygons) or comparing adjacent shapes for overlaps; conflicts may be resolved by applying dominance rules, for example, to clip features into non-overlapping polygons.

Geographic information system (GIS) software is the primary tool for manipulating geographic images and shape data and delineating and editing shape data. Clearly, GIS software functions will be an essential component of a CASI for geographic feature collection. However, without substantial custom programming, commercial GIS software may have significant limitations for use in survey data collection. GIS interface conventions and associated terminology differ greatly from standard office and CASI software conventions, which increases the difficulty of training data collectors who have not previously used GIS applications. To accommodate a broad range of applications, commercial GIS software also offers an enormous and often daunting suite of tools for data display, manipulation and storage, at the cost of considerable computational overhead. For survey data collection, only a small subset of these tools is needed. While it is possible to restrict the tool set within a GIS application, the flexibility in what a user can do with tools is often too broad for a structured survey protocol. Further, some of the most important tools for a survey data collector (e.g., editing) are not optimized to minimize the burden associated with using the tool or to maximize the clarity of the consequences of actions being taken by the data collector.

An approach to developing CASIs for recording geographic feature data is needed that recognizes these essential differences in how alphanumeric questionnaire data and geographic shape data are collected and manipulated. With this challenge in mind, we worked with two U.S. Department of Agriculture (USDA) agencies to explore user responses and possible software approaches to creating a CASI to support collection of geographic features. We focused in particular on geographic feature data that require hand-delineation by the enumerator or data collector; we did not consider geo-referenced data collection that relies on an external device, such as a GPS receiver that records points coordinates or navigation paths.

The goal of this paper is to discuss preliminary research into design issues for survey instruments that support geographic feature data collection. We describe methods and results from two users studies conducted in collaboration with the National Agricultural Statistics Service (NASS) and the Natural Resources Conservation Service (NRCS). We then briefly discuss our findings in the context of prototype software for collecting natural resource feature data.

2. Collecting Agricultural Tract Data From Farm Operator Interviews

2.1. Objectives

A user study was conducted in collaboration with the NASS that involved a simplified GIS application to record agricultural field boundaries during interviews with farm operators. The overall objective of the study was to evaluate the feasibility of using a computer-based approach to capturing digital boundaries for areas operated by producers within sample area segments. Here, we focus on the specific goal to obtain information on how GIS software can be designed to effectively support users with little computer experience. Study outcomes related to enumerator experiences and implications for NASS data collection are presented in Nusser (2004).

2.2 Study Design

The user study was conducted in November 2003 in Iowa. The field setting was modeled after the June Area Survey, conducted annually by NASS to obtain data on current and planned agricultural production for the upcoming growing season. Field interviewers, who had not previously been exposed to computer-assisted data collection, used pen-based tablet computers to delineate field boundaries on digital photographs during interviews with farm operators.

During the June Area Survey, NASS enumerators collect data about current and planned agricultural activities for all land tracts within a sample of area segments. Most area segments in Iowa are 640 acres (the size of a 1 mi x 1 mi square). Survey data collection involves identifying the boundaries of all land managed by a producer within the sampled area segment and then collecting data on each field operated by the producer. Shapes are delineated for each *tract* boundary within the segment. A tract is defined to be all land in the segment that is operated by a producer. The tract is further divided into one or more *fields*, each of which is also delineated. A field corresponds to an individual parcel of land that has a common land use, such as farmstead or crop field. For cropped fields, the crop to be planted in the field, total field acre and cropped field acres are also collected. Currently, tract and field boundaries identified by the respondent are recorded using a wax pencil on a large black and white aerial photograph with a scale of 8 inches to 1 mile (i.e., a scale ratio of 1:7920), and acreage data are entered on a paper questionnaire.

The study was conducted with six enumerators, selected to represent a range of interviewing experience (1-2 years to over 5 years) and computer experience (none, desk/laptop users, computer-aided design software user). Twelve Iowa area segments with historically high response rates and a range of field configurations were identified to be included in the study; fields were geographically distributed from western to eastern Iowa in the central tier of counties. Each enumerator was assigned two segments, generally in their own assignment area.

The interview process involved verifying the respondent information, identifying and delineating tract and field boundaries on the computer, entering an identification number and acreage of each tract and the identification number, the land use and the acreage of each field. Enumerators were trained on the use of the tablet, the GIS software interface, and using the software tools to collect the data.

A simple GIS application to support the interview process was developed by NASS using ESRI ArcView 3.3. The application was customized to offer the few tools needed to digitize boundaries and collect the reported crop and cropped acres. Image navigation functions included zoom tools to examine at smaller or larger regions of the image (zoom in and out to view the image with more or less detail, respectively) and a pan tool for scrolling laterally across the photographic image. Data collection support tools were developed for delineating a tract boundary that included all fields operated by the producer within the area segment (as more than one polygon, if necessary) and delineating each individual field operated by the producer. No polygon editing was allowed due to time constraints associated with developing an edit tool in ArcView that would be simple for a naive computer user to apply. Thus, when a mistake was made in digitizing, the enumerator had to re-enter the polygon. Automatically-spawned dialog boxes allowed enumerators to enter the tract id and respondent-supplied tract acres, and field id and respondent-supplied field acres, land use, and if the field was cropland, the planned crop acres. Auxiliary tools were provided to assist the enumerator in checking progress and supporting the operator. These included linear and aerial measurements tools and a simple summary table of the field and tract data entered thus far for the area segment. Edit checks were implemented for tabular data, including farm operator self-reported acres; no additional edit checks were included for collected shape data.

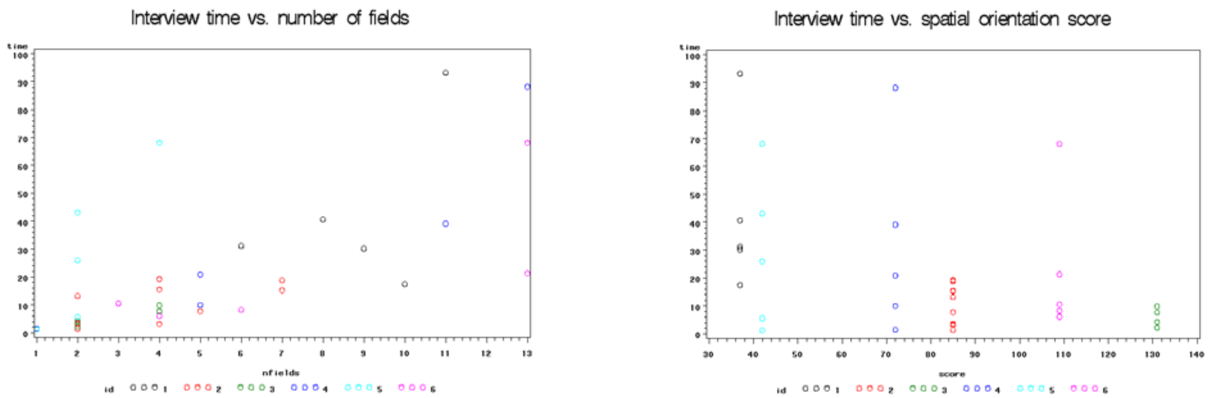
During the study, NASS and Iowa State University (ISU) staff accompanied the enumerators to gather observations on the interview process. After each interview, enumerators and observers completed questionnaires that requested information on interactions with the respondent and with the software. A debrief session was conducted to elicit information from the enumerators about their experiences and suggestions about how the computer-assisted methods could be improved. A cognitive test measuring spatial orientation was administered to enumerators prior to this session (Ekstrom et al. 1976). The software use time during each interview was extracted from the GIS system files.

Questionnaire data from each interview were summarized as frequencies. A mixed model regression of software use time on the number of fields per respondent and the enumerator’s spatial orientation score was conducted, using area segment within enumerator as a random effect.

2.3 Results

Thirty five interviews were completed by the enumerators in this study. Five enumerators conducted four or five interviews, and one enumerator conducted 11 interviews. The average number of fields per respondent was 5.3, with a range of one to 13 fields per respondent. The shape of the fields varied from rectangular to highly irregular. As reported in Nusser (2004), the interview process was well accepted by enumerators and respondents, with most of the tracts being reasonably easy to digitize.

Figure 1. Plot of interview software use time in relation to number of fields delineated during the interview (left) and to the interviewer's spatial orientation test score (right). Each interviewer's data is represented by a different color.



Effort was measured by the amount of time spent using the software during the interview process, which ranged from 1 minute to 1.5 hours and averaged 22.4 minutes per respondent. The regression of software use time on the number of fields per respondent and the enumerator's spatial orientation score indicated a strong relationship between software use time and number of fields ($p < .01$) and a modest correlation between the spatial orientation score and use time ($p = .06$). Graphs depicting these relationships are presented in Figure 1.

Observations and comments made during the debrief session indicated that enumerators would have benefited from a more formal, step-by-step outline of how to use the software to capture data during the interview. During these sessions, enumerators with lower spatial orientation scores expressed more confusion on how to effectively use the software than those with higher scores.

With respect to software issues, enumerators and observers indicated that it would be highly desirable to avoid the use of double taps, perhaps by using a single-tap software button or hardware button. Double taps are used to "turn off" or toggle out of the editing mode, where shapes are digitized. Because the double-tap toggle out of the digitizing mode was hard for enumerators to reliably execute, many errors were made that led to confusion about whether or not the software was still in editing mode. When the enumerator did not successfully toggle out of editing mode but thought that s/he had, the next set of taps were recorded as feature boundaries, but the enumerator was expecting a different action to occur and became frustrated and confused. Another interface problem was that the software used a gray line to depict the digitizing line during delineation, which was difficult to see against a black and white image; ArcView did not permit the color of this line to be altered.

3. Delineating Water Bodies on High-Resolution Low-Altitude Photographs

3.1 Study Objectives

A user study was conducted in collaboration with the Natural Resources Conservation Service (NRCS) in preparation for developing software to collect digital geographic feature data for the 2004 National Resources Inventory. A simple GIS application was used by mock data collectors to record water body boundaries on a high-resolution low-altitude photographic image. The specific objectives of the study were to investigate (1) the impact of digitizing scale on measurement error and effort required to record the data, (2) how data collectors hindered or helped by GIS interfaces, and (3) how data gatherers approach digitizing tasks.

3.2 Study Design

The study was conducted in May 2004 on desktop computers in an Iowa State University GIS computer laboratory. Six NRCS staff with varying levels of exposure to GIS software participated in the study. A few days before the study, each participant was trained for four hours on the basic use of ArcView software and selected tools to record and edit feature boundaries. The specific task given to data collectors was to digitize water bodies on aerial photographs. The exercise was preceded by several hours of training and practice to enable the participants to become comfortable with the software task prior to commencing with the formal exercise.

Three experimental factors were of interest in this study: the size of the water body, the distinctness of the water body boundary, and the scale at which the boundaries were digitized. There were three water body size classes: small (0.1-10 acres), medium (10-40 acres), and large (more than 40 acres). Two edge distinctness categories were identified for small and medium water bodies: distinct (clear edge) and vague (edge was shadowed or occluded by tree cover); large water bodies all had a combination of distinct and vague edges. Between these two factors, five feature type (size and edge clarity) conditions were studied. The scale at which a water body was delineated was the third factor, and five digitizing scales were studied: 1:7920, 1:6000, 1:4000, 1:2000, and 1:1000. For a ratio of 1:7920, 1" represents 1/8 mile or 660 feet, which is the *smallest* scale in the study, or the view with least detail and the most visible land area. A ratio of 1:1000 is equivalent to 1" = 83 ft or ~0.02 mile, which the *largest* scale, or the view with most detail and the least visible land area.

For each water body size and edge clarity category, two training and five experimental digital photographs were obtained, resulting in a total of 10 training photographs and 25 experimental photographs. A single water body of the appropriate category was identified for each image. For a given water body size and edge clarity category, image replicates and scale levels were randomly assigned to participants using a graeco-latin square design.

An ESRI ArcView 3.3 software application was developed to support the exercise that offered a constrained set of tools. The software displayed the selected image at the prescribed scale with a symbol indicating which water body should be digitized. No zooming was allowed. Polygons were created by recording a series of vertices (points) that were connected in sequence. Available software tools allowed the user to create a polygon; edit a polygon by adding, deleting, moving a vertex; delete a polygon; pan; and set a snapping tolerance. If a snapping tolerance is set, the endpoint of the delineated line will be "snapped" to the nearest recorded vertex when the endpoint is within the snapping tolerance of the vertex.

Each participant was asked to digitize water bodies in the order prescribed by the experimental design. Attributes of the final delineation were stored, including the area of the polygon, its circumference, and the number of vertices recorded. The difference between the area of a water body digitized by a given participant at a specified scale and the area of a gold standard delineation for the water body was calculated. Screen capture software was also used to record the working session for each water body completed by each participant. The resulting video was used to determine the amount of time spent with the initial delineation and subsequent editing, and to provide information on user behaviors in using the software to accomplish the task.

Analysis of variance was used to estimate the effect of water body size class, edge clarity, and digitizing scale on effort measures (number of vertices and time spent delineating) and area measures (the delineated area, the difference between the delineated area and the gold standard area). For area measures, we also considered both the absolute area and the fraction of the gold standard area that was erroneously included (commission errors or overshot boundaries) and the same two variables for erroneously omitted areas (omission errors or undershot boundaries). When results are cited for pairs of related variables, such as for both effort variables, the largest p-value is reported unless otherwise noted. Videos and observations by those conducting the experiment were reviewed for further information on how users approached the task and used the software.

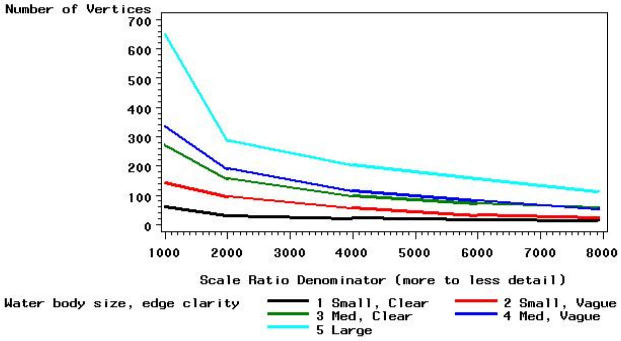
3.3 Results

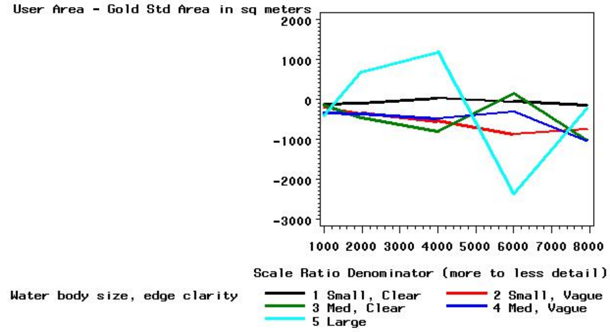
Analysis of variance results for effort as expressed by the number of vertices and time spent digitizing indicted that the size of the feature ($p < 0.001$) and the scale of the delineation ($p < 0.0001$) were significant, but the clarity of the water body edge was not. An interaction between feature type and delineation scale was also significant ($p < 0.001$). Larger water bodies and larger scales (more detailed view, least visible area) required more effort (Figure 2, left plot).

On the other hand, there were no statistically significant differences for the absolute or relative differences between the participant's delineated water body area and the gold standard water body area, indicating that feature size, edge clarity and digitizing scale did not have an impact the accuracy of delineated areas in these experiment (Figure 2, right plot). The largest difference observed was roughly 3 acres for a large water body (~11,000 sq meters), with most differences falling well below this value.

In general, the size of erroneously included and omitted areas increased with larger water bodies ($p = .001$), while the fraction of erroneously included and omitted areas were smaller with larger water bodies ($p <= 0.01$). When scales were smaller

Figure 2. Graphs of mean number of vertices (left) and difference in area between user-digitized and gold standard water body area (on right) in relation to the digitizing scale, where 1000 indicates that the scale with the most detail, but the smallest viewable area and 8000 indicates the scale with the least detail, but the largest viewable area.





(more zoomed out, less detail), the erroneously excluded area and the fraction of erroneously included and omitted areas were larger ($p = 0.06$ for fraction of erroneously included area, $p < 0.01$ for other two variables). The interaction between feature type and scale was significant only for the area of erroneously excluded areas ($p < .001$), where larger water bodies were associated with larger excluded areas at smaller scales.

Observation of user responses to the constrained tool set and fixed digitizing scale indicated that it was easier for new ArcView users to learn procedures than it was for those who had prior GIS experience. Experienced users were accustomed to a broader suite of tools and more flexibility in how the tools could be used. Some experienced GIS users were quite frustrated by the inability to zoom in, and felt that the lack of detail at the smaller scales negatively affected their digitizing quality.

Users found the latency involved with refreshing the image to be quite frustrating. Images were served from a server on a local network, but when a lot of panning was required, as was the case for large water bodies being digitized at larger scales, the refresh rate seemed to significantly impede progress and create user frustration. Observations also indicated that, given the importance of the panning tool in a fixed scale environment, the interaction mode for the pan tool required too much effort. Finally, participants noted that although the boundaries of digitized areas were visible, the size of the boundary sometimes obscured small objects that were important to evaluating the quality of the recorded boundary.

Users generally approached the digitizing task by recording an initial boundary and then editing the boundary if needed. One participant had small motor movement problems, and that individual preferred to draw a boundary surrounding the water body, and then pull the vertices towards the edge of the water boundary to make the final determination.

4. Discussion

4.1. Introduction

Computer-assisted survey instrument (CASI) software is used to record data collected as part of a statistical survey. The CASI plays an important role in reducing the potential for nonsampling error and minimizing interviewer burden during the data collection process. This is largely accomplished by enforcing the proper implementation of protocols and providing usable and efficient tools that support accurate capture of information. While software principles for questionnaire-based data collection are reasonably well understood (Couper et al. 1998, Nusser et al. 1997, Nusser and Thompson 1998), research to guide the development of CASI software for digital capture of geographic features is limited. As with questionnaires, there is the question of how best to obtain a "response" that leads to an accurate measurement. In addition, because of the more complex interaction required by recording and editing shape data, more consideration needs to be given to the user interface and software functionality as a vehicle to improve the efficiency and accuracy of the data collection.

Although these studies focus on both of these issues, their settings differ. The goal of NASS study was to learn about enumerator interactions with the respondent and GIS-based survey software during an interview. Participants were new to computer-assisted interviewing and GIS. NASS interviewers were provided a limited set of GIS tools, but considerable flexibility in how those tools were used to manipulate the geospatial data. The NRCS study focused on user interactions with GIS-based software for a data collection task that did not involve eliciting information from other humans. Participants had more experience with office computers and in some cases GIS software, and had formal GIS training before the study was conducted. The software application offered more sophisticated editing tools, although the absence of a zoom tool created a more restrictive manipulation environment for viewing geospatial data. Despite these differences, several common threads emerge, and the range of settings covered by these studies sheds light on issues associated with data collectors as software users and with software design for survey instruments that support geospatial data collection.

4.2 Software design

Interface and tool design options offered by GIS software are somewhat problematic in a survey setting, in part because the initial driver for GIS software development was unstructured interaction tasks to *visualize* geospatial data, rather than structured actions to *create* geospatial data objects. Toggling, which invokes and turns off tool functions, is often used in GIS software. However, users found the way toggling worked to be frustrating because of the types of mouse actions used and because the cues that signify the active mode were too subtle. Many icons used by GIS manufacturers are also foreign to novice GIS users and make it difficult for interviewers to learn the software. For example, users reported issues with remembering which zoom direction was represented by the "+" (zoom in) and "–" (zoom out) icons. Rusch et al. (2004) verified this confusion in a usability study that compared a range of standard GIS icons with alternative icons as input to the development of Census Bureau address canvassing software. Ideally, a survey instrument for collecting geospatial data would provide more familiar and efficient interaction methods, particularly to reduce the hand movements required to perform common tasks.

Computer-assisted survey software typically enforces restricted paths or offers limited tool sets to the data collector to support proper protocol implementation. For example, the directional flow of a questionnaire is often prescribed by data-driven rules in a computer-assisted survey instrument. The ability to implement this kind of control in a commercial GIS system is somewhat limited. It is possible to implement software protocols that dictate the order in which the data collection proceeds for macro-level tasks (e.g., record the tract before the field, record identifiers before other attributes). It is more difficult and is often undesirable to enforce the sequence of micro-level steps of the process. In working with early drafts of the software, we found that integrating a high degree of flow control in the details of recording geospatial data severely limits the kind of flexibility required for the ad hoc nature of user interactions when recording shape data.

It is also possible, and in some cases beneficial, to control the environment for digitizing. Observations during the NRCS study suggest that some users will spend too much time relative to the accuracy gains achieved. When collecting tens of thousands of polygons, this can have a negative impact on study costs. Based on the results of the NRCS study, the scale should probably be fixed to avoid spending too much time on detailed delineation that contributes little to reducing the measurement error. In our setting, given the types and sizes of geometric objects that we were collecting, study results suggest a scale of 1:4000 may strike a balance between reducing digitizing errors and spending too much time to achieve a very accurate polygon. Of course, the appropriate scale for other surveys will depend on the types, sizes and number of features being recorded and the desired precision of the data. The NRCS study also indicate that more flexibility is needed in the viewing scale when evaluating the image or shape prior to and after the digitizing step.

One of the more challenging aspects of the NRI study was the latency when traveling across a reasonably large local network. The technology used for image display refreshes the screen every time the image view is altered (e.g., via panning or zooming). While this was less of a problem in the June Area Survey study because of the local storage of the images, there was still some latency in the local refreshing of the screen. The resulting delays could be costly and often frustrate data collectors working under pressure.

4.3 Survey interviewers and data collectors

With respect to software users, prior experience has a profound effect on how the data collector or interviewer approaches the task. Fortunately, even for novice users with fears about their ability to use the software, a constrained interface is reasonably simple to learn. At the same time, novice users do require quite a bit of practice and a very structured training regimen to learn how to reliably perform the inherently more complex task. This is partly due to their lack of exposure to some of the interaction modes of digitizing, including navigation via panning and zooming as well as the act of digitizing. Progress is needed on simplifying tools and making interaction modes available that are more consistent with users' interaction habits from other software.

For experienced GIS users, the switch to a constrained protocol is difficult. There is a tendency to want to apply quality standards that may not relevant and to want to fall back on the much broader suite of tools used in other GIS work. It will be necessary to set the stage for such users so that they understand that this is not a GIS application, but rather a survey data collection application with a different set of objectives that drive what should be done and how.

Finally, at least in the NASS study, we see evidence that users will respond differently in relation to their ability to mentally manipulate spatial images. A positive correlation between measures of spatial ability and field staff productivity and software use has been observed in other studies with Census Bureau field staff (Murphy and Nusser 2003). Further, human-computer interaction research indicates that persons with lower scores on spatial cognition tests have more difficulty using computer software (Norman, 1994; Vicente, Hayes, & Williges, 1986). Research is underway at the Census Bureau to examine how measures of spatial ability, wayfinding strategies, and map reading ability are related to software use and productivity measures. One hypothesis is that persons who have more trouble processing visual information tend to prefer text-based lists of steps or icons, similar to findings from navigation preference studies (Lawton 1996, Streeter and Vitello 1986). These studies suggest that both training and software design must be carefully constructed to support a range of styles and abilities to work with computers and spatial information.

4.4 Prototype survey instrument software

The Iowa State University Center for Survey Statistics and Methodology is responsible for developing survey systems for the National Resources Inventory (NRI) survey program under a cooperative agreement with NRCS. Part of the NRI data collection process involves recording shape data that represent features such as water bodies and streams, built-up land, and transportation within each of 70,000 area segments per year. In the past, data collectors recorded shape data on mylar sheets and used planimeters to obtain measurements. The Center is developing a computer-assisted survey instrument (CASI) to collect shape data digitally. Considerations from these studies, plus the desire to vastly streamline the software and improve our control over the interface design, led us to develop custom survey instruments for NRI survey data collection, rather than build these instruments using commercial GIS software.

In the current CASI design, each major step of the process has a separate view. A major step, for example, is to identify and digitize shape data for a specific feature category such as transportation. The tools used to record data for any feature type function in the same way. However, within each feature view, only the appropriate tools are offered for the feature type, and tools are customized to automatically label the polygons with the correct feature type and enforce size restrictions for eligible features. Survey instrument work flow is handled by allowing users to switch between different views (e.g., via tabs on the interface). The interface can also be set up to easily navigate back to a prior step, while preventing the data collector from skipping to a future step until the appropriate condition is triggered. Context for progress in the data collection process is provided by graying out references to steps that have not been started.

Interaction methods have been constructed to provide multiple options for how the user toggles in and out of different functions (i.e., modes). For example, once the user has selected which feature type s/he will record, the data collector can use a keyboard shortcut, toolbar button, or mouse tap to leave the editing mode. This allows users can use the interaction style that is most natural. An additional innovation is that the software takes advantage of a graphics accelerator card to provide real-time manipulation of images with the speed seen in video game software. This has eliminated the latency issue, and greatly improved the usability of the software.

We have developed three classes of computer-assisted checking rules for geospatial data that mirror those of coded questionnaire data: rules that check whether a valid geometric format is recorded for the feature (e.g., polygon is closed), internal consistency checks among shape objects within the area segment (e.g., no overlapping polygons), and consistency rules comparing attributes of shape objects with other survey data collected for the sample unit (e.g., whether a sample point located within a polygon has a land cover/that use is consistent with the polygon label). Several protocol rules and checks that were manually implemented by the data collector in prior surveys are now being shifted to the software. For example, NRI has specific dominance rules for different types of land that may be overlapping. Rather than asking the data collector to apply these rules, dominance and clipping rules are applied by the software after data are collected. Rules are also being designed to check the consistency of polygon attributes with non-shape data collected at sample points inside the area segment in real-time, which was not possible in the past because the topological relationships were not represented digitally.

Acknowledgements

This research was supported in part by the National Science Foundation under Grant No. EIA-9983289, a portion of which was funded by the National Agricultural Statistics Service, and Cooperative Agreement No. 68-3A75-3-43 with the USDA Natural Resources Conservation Service. NASS research staff Eric Waldhaus, Ray Roberts and Dan Beckler; Dean Thompson of NRCS; and Howard Butler and Marc Rogers at Iowa State University all contributed substantially to this research. Andrew Vardeman, Scott Hanrath and Todd Krueger at Iowa State University are designing and developing the prototypes discussed in this research. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation, the National Agricultural Statistics Service, or the Natural Resources Conservation Service.

References

Couper, M. P., Baker, R. P., J. Bethlehem, J., Clark, C. Z. F., Martin, J., W. L. Nicholls, II, W. L., O'Reilly, J. M. (1998). *Computer Assisted Survey Information Collection*. Wiley, New York.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.

Lawton, C.A. (1996). Strategies for indoor wayfinding: The role of orientation. *Journal of Environmental Psychology*, 16:137-145.

Murphy, E. D. and Nusser, S. M. (2003). Evaluating user interfaces for accommodation of individual differences in spatial abilities and way-finding strategies. *Proceedings of HCI International 2003*, 4:1005-1009.

Norman, K. L. (1994). Spatial visualization -- a gateway to computer-based technology. *Journal of Special Education Technology*, *XII*(3), 195-206.

Nusser, S. M. (2004). Computer-assisted data collection for geographic features. *2004 Proceedings of the Section on Survey Research Methods,* American Statistical Association. [CD-ROM]

Nusser, S. M. and Thompson, D.M. (1997). Networked hand-held CASIC. Proceedings of *Symposium 97: New Directions in Surveys and Censuses*, Statistics Canada, p. 361-364.

Nusser, S. M., Thompson, D. M. and DeLozier, G. S. (1996). Using personal digital assistants to collect survey data. *1996 Proceedings of the Section on Survey Research Methods,* American Statistical Association, p. 780-785.

Rusch, M. L., Murphy, E., and Hourcade, J.P. (2004). Icon design study questionnaire and results, LAMI Version 1.0, Human-Computer Interaction Memorandum Series #71, Statistical Research Division, U.S. Census Bureau.

Streeter, L. A. and Vitello, D. (1986). A profile of drivers' map reading abilities. *Human Factors*, 28:223-239.

Vicente, K., Hayes, B. C., & Williges, R. C. (1987). Assaying and isolating individual differences in searching a hierarchical file system. *Human Factors*, *29*, 349-359.