Managing Dynamic Data

Change Detection and Feature Extraction Methods Support Census Bureau Map Updates

*By Charles O'Hara, GeoResources Institute, Mississippi State University (*www.gri.msstate.edu*), Mississippi State, Miss.*

From mapping's earliest days, mapmakers have used lines to abstractly depict the centerlines of linear features such as streams and roads. Today, centerline data modernization is an ongoing process in the United States.

The U.S Census Bureau is conducting a massive program to modernize its Topologically Integrated Geographic Encoding and Referencing (TIGER) file road centerline data (see "TIGER File Update Under Way," page 42). To complement the bureau's work, the GeoResources Institute (GRI) at Mississippi State University (MSU) is conducting ongoing research in partnership with several commercial companies, including DigitalGlobe, Definiens Imaging, ESRI and Smart Data Strategies. The group's goal is to create technologies and methodologies that detect changes and extract map-accurate features of interest for updating map products such as centerline files.

Multitemporal Image Data and Change Visualization

High-quality, co-registered image data acquired at different times can be used to automatically detect areas of change. Land use and land cover change analyses can be conducted in many ways, but this project focuses on three methods of change detection:

1. Image fusion for visual exploration of change.
2. Hybrid change analysis in which a variety of techniques are combined to optimize the detection of change areas and minimize errors of omission and commission.
3. Object- and feature-fusion change detection, which fuses a classification from a prior image to constrain the segmentation of a later image. This process enables information about the spectral, textural and other characteristic attributes of image-segmented objects within the two images to be mined for differences that would indicate specific types of land use and land cover change.

Figure 1. Images for the study were acquired during winter, leaf-off conditions for eight study areas across the country.

To conduct this project, bi-temporal, orthorectified DigitalGlobe QuickBird satellite data were collected from 2002-2004 for eight study areas of geographic, physiographic and demographic diversity across the United States. Images were acquired for nearly cloud-free conditions and near-nadir look angles to minimize shadows or oblique angle distortions.

To explore changes between the image scenes, images were evaluated for horizontal accuracy and close co-registration—a prerequisite to effective, automated change detection and multitemporal analysis. Once images met accuracy requirements, change exploration was conducted on entire image scenes (Figure 1).

Fusing image data for small areas provided a wealth of data about change for small features within residential areas

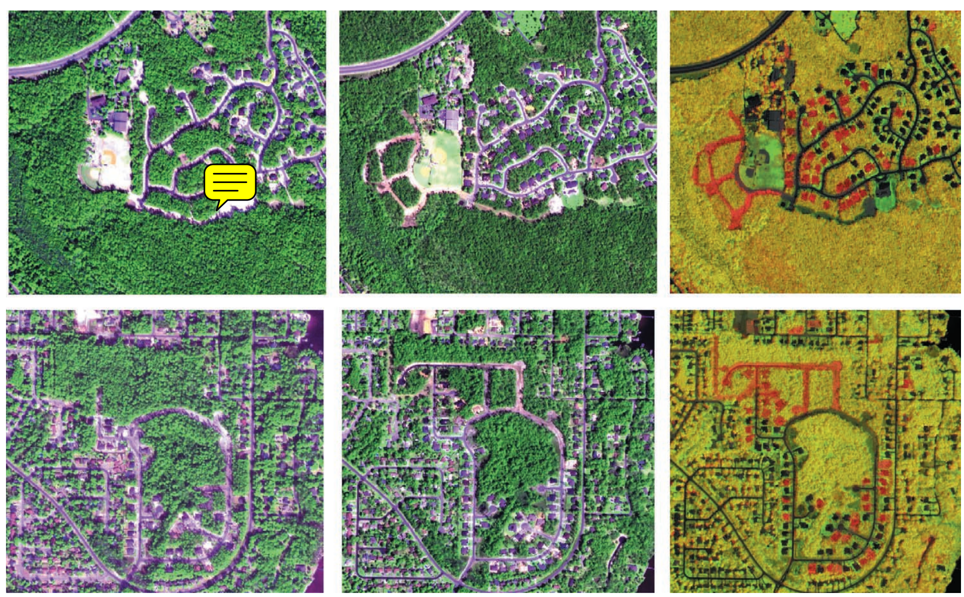

Figure 2. Color views of rural neighborhood areas for early images (left) are contrasted by later images of the same areas (middle). The right column shows unchanged areas in yellow (black in shade), and areas of change are shown in shades of red and green.

(Figure 2). These visual results support the use of co-registered, multitemporal, high-resolution satellite imagery for change detection and visualization, and illustrate how image data can be processed and manipulated to highlight locations of new features on a changing urban or rural landscape. Similar techniques may be employed to detect and visualize changes due to road construction (Figure 3) or to explore change details (Figure 4).

Classification and Feature Extraction

Extracting features of interest from high-resolution spectral images may be accomplished using many approaches. Traditional approaches to land cover and land use classification have focused on pixel-based classification wherein classes

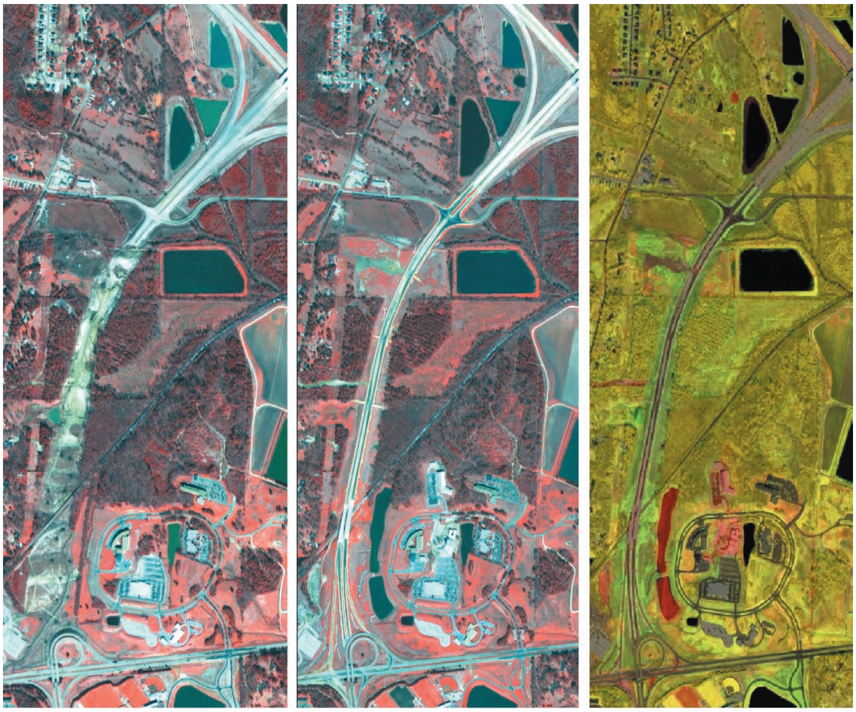

Figure 3. A 2002 QuickBird image of Starkville, Miss., (left) shows ongoing activity for a major roadway construction. The middle image shows the completed roadway, as well as storm water runoff impoundments and new facilities at the MSU Research and Technology Park. The change detection visualization (right) shows new roads, ponds and buildings in reddish shades. A panchromatic image is fused to the image to provide additional details and highlight areas of change.

were defined as a collection of pixels with similar multispectral characteristics. New object-based approaches to image processing divide the image into segmented areas with more or less homogeneous characteristics. The segmented areas are managed as objects with extensible attributes that an analyst can manipulate, allowing class membership functions to be defined.

In this project, object-based technologies are used for change detection and material classification of features. By partnering with DigitalGlobe and Definiens Imaging, the project has access to high-quality satellite data and personnel who wish to validate change detection and feature classification data products. The project team also has access to an enhanced version of Definiens Imaging's object-based classification software, eCognition Enterprise, which provides a platform wherein advanced, highly customized approaches to data, information, and object fusion are being developed and employed to provide enhanced change detection and feature extraction capabilities.

Using multitemporal images, areas of change may be explored, visualized and classified. As shown in Figure 5, new development features may be highlighted and used as a basis for managing updates to spatial data sets. Areas determined to have changed may be rapidly and efficiently located, and new features may be easily identified and mapped.

Tasks and Technologies

Using high-resolution imagery for detecting change and extracting features of interest for large geographic areas presents significant technical and managerial challenges. To develop a meaningful approach to address these challenges, the project considers the following technical and managerial tasks: data update status management, data processing task management, user needs assessment, technology outreach, raster data management, raster data quality assurance (QA), vector data management, data manipulation and transformation, data enhancement and fusion, change detection, feature extraction and classification, feature manipulation, and data update operations:

Data Update Status Management: An ESRI Inc. ARC-IMS geographic information system (GIS) application was designed and tested to provide status maps for processing based upon counties, as well as U.S. Geological Survey 1:24,000-scale quadrangle maps. A series of attributes were designed to control color-shade symbolization of status and allow summarized reports to be developed for status tracking. This allows areas to be tracked based upon processing status and steps within major status categories. Update requirements varied geographically and depended largely upon the policies of the state, county, tribe or local agency responsible for the update; the availability of funds to complete an update; the agency's technical readiness level; and the availability of data to use for an update.

Data Processing Task Management: A Web-centric, database-integrated application was designed that included an ARC-IMS application to provide task-assignment capabilities, report-generation capabilities and status maps for processing tasks. Defining access to data sets and tools was considered in a model based upon the Boeing-Autometric task management tools developed for the Shuttle Radar

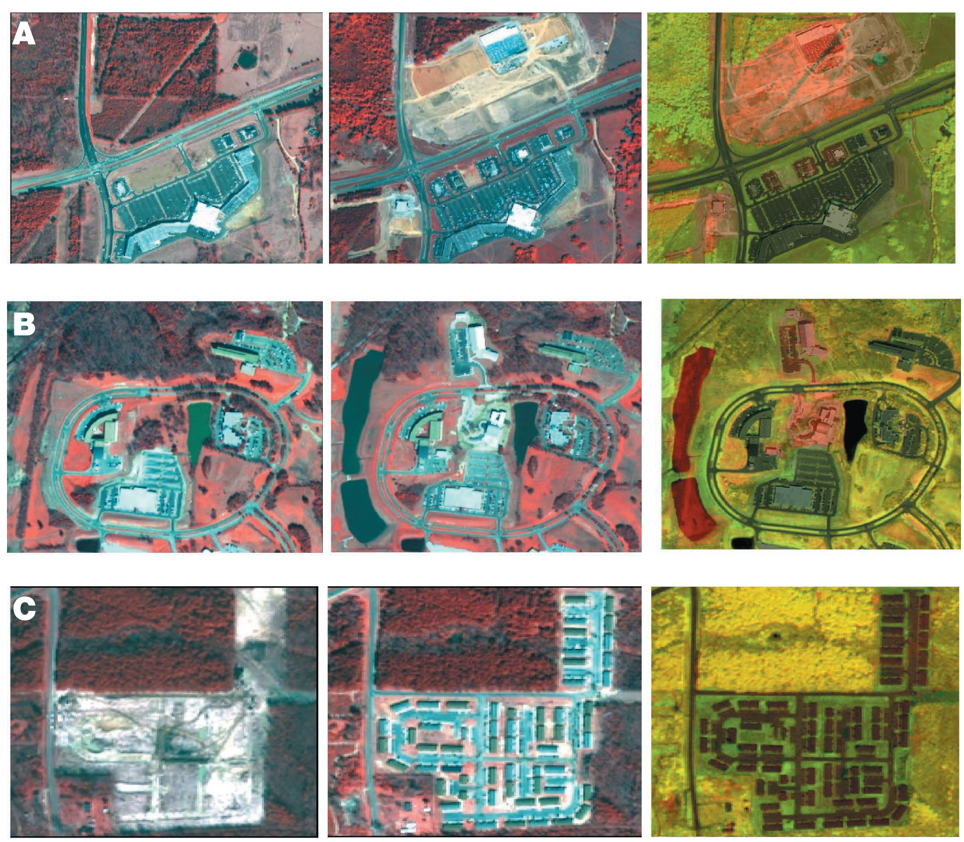

Figure 4. Early images (left) show preliminary conditions. The middle images show the same areas after land is cleared (A), buildings are constructed or major drainage features are put in place (B), or developments are built (C). The images on the right highlight areas and types of change.

Topography Mission (SRTM) processing project (*http://srtm.usgs.gov*). PixSell, an on-site subcontractor that led the SRTM technical development tasks, managed tasks and data updates.

User Needs Assessment: MSU's Social Science Research Center conducted user-needs assessment tasks through focus group meetings called Decision Support Labs (DSLs). The DSLs leverage software to build anonymous consensus, explore user issues, quantify results and generate automated reports. DSLs held in Mississippi included federal, state, county and local agency participants, and at the ESRI International 2004 User Conference to target Census Bureau agency staff and TIGER file users.

Technology Outreach: A multimedia product will include all project data, reports and materials relevant to government, educators, companies and the public.

Raster Data Management: Data acquisition was coordinated among project team members, DigitalGlobe contacts and agency data managers who exchanged data for the project areas. Advanced raster management strategies from ImageLinks and Intergraph Z/I Imaging were tested, and compression technologies were investigated to reduce file sizes while preserving information content.

Raster Data QA: The horizontal accuracy of the orthorectified image data used to extract new features or align previously existing features determines the degree to which the accuracy of the resultant vector data may be enhanced. A review of approaches to quantification of horizontal accuracy led to the development of in-house capabilities to quantify image accuracy.

Vector Data Management: A variety of vector data sets were acquired for each project area. The project team acquired data generated by local,

TIGER File Update Under Way

Early digital representations of street centerline data tied to street addressing information were compiled using address coding guides, with numbered intersections and address ranges assigned to digitized street block faces in the form of dual independent map encoding (DIME) files. By 1980, these files evolved and became known as the geographic base file (GBF)/DIME.

The 1990 Census saw the first implementation of a nationwide, seamless, digital map of centerlines with street address ranges known as the Topologically Integrated Geographic Encoding and Referencing (TIGER) file. The simple diagram in Figure 1 shows how the TIGER file associates address ranges with possible structures along a block between the starting and ending nodes of the street line segment, with even numbers on the right and odd numbers on the left (moving in the direction of increasing address number).

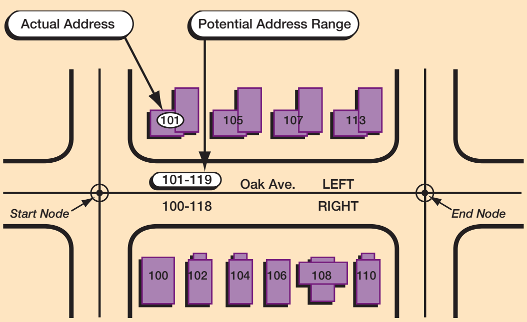

Figure 1. A 1997 map shows TIGER/Line address basics.

As part of an ongoing Master Address File (MAF)/ TIGER Accuracy Improvement Project (MTAIP), the U.S. Census Bureau is realigning TIGER to an accuracy of 7.6 meters or better (Figure 2). To accomplish this improved spatial accuracy, Census Regional Office staff personnel are contacting state, county, local and tribal governments to determine if they have data that meet Census Bureau requirements. When data are submitted, an independent accuracy check is made by the Census Bureau using points collected by a contractor. When centerline files aren't available, Harris Corp., MTAIP's main contractor, will obtain the data by the following methods:

- Using vector files as input to the realignment process.
- Checking with the private sector for available geographic information system (GIS) data.
- Subcontracting for vectors extracted from new or existing imagery.
- Subcontracting for street centerline collection using Global Positioning System (GPS) technology.

Figure 2. The MTAIP aims to improve the accuracy on the U.S. Census Bureau TIGER file by collecting accurate data from state, county, local and tribal agencies that have used highly accurate image data to realign centerline data sets.

Such work underscores the changing nature of spatial data development in modern society. A benchmark example of accomplishing spatial database updates through improved technology, MTAIP employs spatial information technologies to significantly improve the nation's important census data as part of a re-engineered 2010 census.

Census Bureau progress reports indicated that by the fall of 2004, data from more than 2,000 counties had been acquired from state, tribal, county and local sources (as well as from commercial GIS sources in some cases), and were expected to be acceptable from metadata provided. However, only two-thirds of the approximately 1,200 counties evaluated as of fall 2004 met requirements for spatial accuracy.

The current planned improvements schedule calls for data from more than 3,200 county areas to be submitted or acquired and processed by 2008. That means state, county, local and tribal agencies nationwide are improving their basemaps, acquiring highly accurate imagery for realigning street data (Figure 3), acquiring accurate data with GPS technology and improving associated address information to accompany their enhanced centerline data files. Given a nationwide source for enhanced centerline data, newly created, highly accurate information sources will properly line up with street data, GPS locations acquired in cars when used to overlay street data will properly show up on the road, and many new uses of data will be enabled through the availability of improved information.

Figure 3. High-resolution panchromatic (upper left), color (upper right) or color infrared (left) images from DigitalGlobe's QuickBird satellite may be used as a backdrop for screen digitizing. Such sources of accurate image data are used as a base layer for aligning road centerline data sets.

county, tribal or state agencies where such data were available. Vector data products play a significant role in the project, requiring the ability to track data updates and manage multiple-user access to edit, update and manipulate vector data sets.

Data Manipulation and Transformation: All data were processed and converted to WGS84 and maintained in a UTM projection for the appropriate zone for the location. This allowed uniform raster data management and the easy use of standard raster government and commercial products typically distributed in UTM.

Data Enhancement and Fusion: Satellite image data provided high-resolution panchromatic products that could be used to sharpen the co-registered multispectral data. Multiple approaches were considered and tested. Tools were developed to optimize data sharpening, and quality metrics were generated for sharpened and fused output.

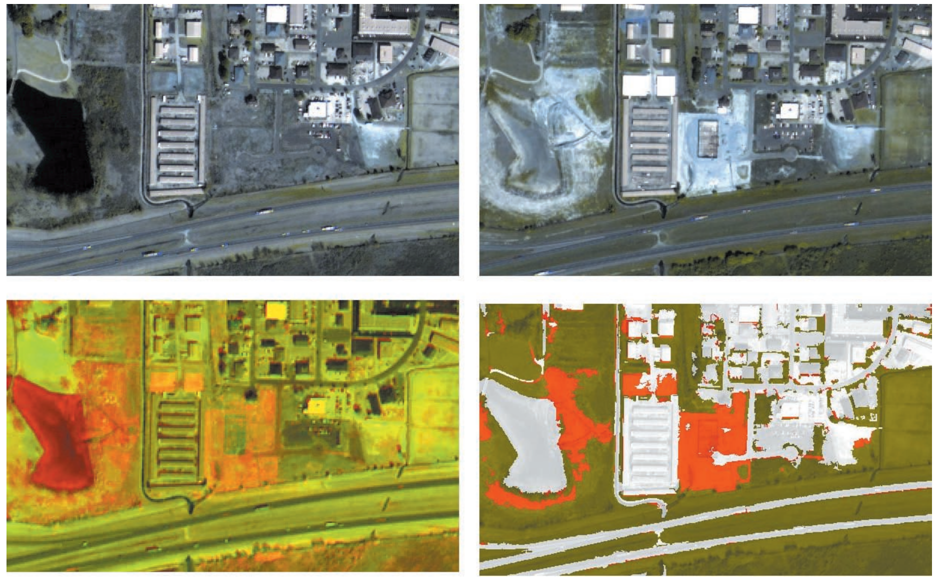

Figure 5. An early image of a commercial area beside a freeway (top left) contrasts with the later image (top right). The change visualization image (lower left) is accompanied by a classification map that shows changed areas, filled areas, cleared areas and new building features as red objects in the classification.

Change Detection: Considerable research has been conducted in this area, and several methods were considered. Ultimately, the team developed a set of standard methods to visually explore change, hybrid methods to determine change, and object methods to quantify and classify change.

Feature Extraction and Classification: The team developed approaches that optimize the identification of developed areas and the extraction of man-made linear features.

Feature Manipulation: A variety of methods were tested for manipulating extracted and classified road areas. Commercial software and internally developed tools were used to transform extracted features to centerline representation. Feature skeletonization in eCognition is one method that provided satisfactory results.

Data Update Operations: A solution was needed to provide an environment for merging new road centerline data with existing centerline data. Also, attribution of new data was required. The project worked with ESRI to define a solution using SDE and PLTS to form the basis of a production data update solution.

Ongoing Research

The MSU project demonstrates how high-resolution, multitemporal satellite image data can be a valuable tool for change detection and centerline feature updates. In addition, the project identifies many areas where improved technologies are needed for standardizing and streamlining data production and maintenance. Further technology deployment activities, partnerships and outreach programs are needed to provide important "middleware" between the Census Bureau and state and local groups to coordinate activities, provide enhanced data exchange, streamline data review and accuracy assessment tasks, and transfer needed technologies.

Because the Census Bureau can't offer partnership incentives that satisfy the needs of the data creators, other methods must be developed to provide needed research, technology transfer, outreach, collaboration and coordination. Support for ongoing research and related activities to improve the quality, usefulness and maintenance of spatial data must be continued. Improved geospatial data in a wide range of management and decision-support applications will benefit society in areas ranging from land and asset management to law enforcement and homeland security.

Author's note: This project was funded through a grant provided by the U.S. Small Business Administration (SBA), Grant No. SBAHQ-03-I-0023. SBA funding doesn't constitute an endorsement of any products, opinions or services. All SBA-funded programs are extended to the public on a nondiscriminatory basis