DATA CAPTURE FOR THE NINETIES : VTRAK

Robin Waters, Dr Dave Meader, Greg Reinecke
Laser-Scan
12343F Sunrise Valley Drive, Reston, VA, 22091

ABSTRACT

Geographic Information Systems (GIS) need cost effective data capture systems to feed their voracious appetites for data. Modern GIS require both raster(grid) data and vector data. Raster scanning of existing maps is a very attractive means of capturing data but the conversion of such images into intelligent, attribute-coded and topologically structured vector information is a non trivial exercise that has confounded many GIS project leaders. Laser-Scan pioneered interactive vector scanning with FASTRAK and LASERTRAK systems and have now implemented the same pragmatic and cost effective algorithms and techniques on a standard workstation accepting input from a variety of scanners. VTRAK, now in use in N. America, Europe, and Japan, is a unique system for applying the best algorithms to the right set of data at the right time. Fully automatic "vectorisation" or "raster to vector" programs must take a global approach to the raw data whereas VTRAK applies image processing, feature extraction, line following, symbol recognition and other techniques in a selective approach. All these can be applied at a standard workstation in different sequences depending on the source documents and the target database. This paper describes VTRAK, its justification, philosophy, operation, application and future development.

JUSTIFICATION

Map data capture is big business. In North America the map data conversion business will be worth more than a billion dollars over the next decade.

Utilities are investing in very large databases for geographic referencing of outside plant records.

Telephone companies are not only improving their records but automating the optimisation of their taxes.

Defense agencies are converting map products into digital form for faster updating, ease of use and direct input to some weapons systems.

Federal government, in the guise of many different agencies needs topographic and thematic information in GIS to provide a more cost effective service for the taxpayer.

Commercial *initiatives* range from the direct use by forest companies to the large number of data conversion services now available.

Despite increasing amounts of data becoming available in digital form direct from original surveys (satellite imagery, digital photogrammetry & total station surveys), the need to convert existing maps continues to grow.

UNIQUENESS OF MAPS

Maps have certain characteristics that distinguish them from, for example, engineering drawings (Woodsford 1988). These characteristics, detailed below, have a decisive effect on the methodology for converting them into digital form.

Maps are **accurate and to scale**. Measurement from the film or paper is the best that can be obtained; there are no dimensions written on the map. However, correction of data to a pre-existing grid or graticule will be vital.

Maps contain **very fine detail**. Lines of 0.1mm (4mil) or finer are commonplace. A high resolution scanner and the necessity for handling large images are required.

The **variety** of symbols and linestyles within a map series is very great. Worldwide the variations are almost infinite.

Maps are often **multicolored** with no access to the source separations.

Maps are generally part of a **continuum** with abutting or overlapping neighbours. Edge matching is non-trivial.

Maps are **multi-purpose** documents and conversion must be carried out with this in mind. Most individual users requirements are less expensive subsets of a general purpose conversion exercise and careful attention to specification and quality control is needed to ensure the support of a range of users.

Maps are of **variable quality**. Best results for data capture will always be obtained from scribed film originals on stable base material. However only blueline paper copies may be available, often overwritten with confusing detail.

Laser-Scan have been dealing with maps for over 15 years and recognise the unique nature of maps and the requirements of the users of geographic information whether it is to be used for the reproduction of maps by digital methods or for recording and analysis of data to produce answers to many geographic queries.

USER REQUIREMENTS

Vector data for digital mapping and GIS can vary in complexity from simple "spaghetti" linework to multi-attribute, object level, topologically structured data. The data capture process is not complete until the product has been tested and proven to conform to the specification and checked against the original input document.

The stages in a data capture process at which increasing complexity is added and/or at which quality is checked are critical to the efficiency of the process. Failure to check basic geometric accuracy until complete topological structuring has taken place will lead to massive reprocessing. However introduction of basic topology at

the time of vector feature extraction can help to achieve correct geometry at the first attempt.

VTRAK procedures can be designed to help this overall process and in particular to deal with the "hidden agenda" behind the overt specification (Woodsford 1988).

Accuracy. Typically specified as "half a line width" or perhaps "2 mil", accuracy normally includes implicit criteria for shape and "cartographic acceptability". VTRAK algorithms and immediate interactive overlay checking ensure quality.

Representation. Different features require different digital representation that can only be ensured by recognition of the class of feature prior to extraction from the image. VTRAK provides this facility. In fact, as a result, completely different algorithms and parameters can be applied to different features.

Abstraction. Many GIS objects are inferred from the actual map depiction; they are not explicit. The centre line of cased roads, the centre point of a symbol, and the indefinite bounding polygon of a symbolised swamp are all particular examples handled by VTRAK.

Selection and Completeness. Different data capture specifications applied to the same map will require that a selection process is carried out. Failure to do this at an early stage in conversion will lead to wasted and often counter productive processing. VTRAK blends selection of parts of the image for background processing with overt interactive selection during feature extraction and special vector edit functions to optimise selection.

Quality Control. The cost of conversion varies in direct relation to the quality control procedures used. The quality assurance programme carried out by the end user or his agent is a factor that must be taken into account in any conversion programme. A VTRAK system can be tuned to handle the most stringent tests including random remeasurement, checking of feature quality with respect to attributes and topology checking.

VTRAK OPERATION

Configuration. VTRAK operates on standard Digital Equipment Corporation (DEC) VAXstations but with background processing, scanner interfacing and network handling capable of being carried out on other VAX processors in a cluster. Laser-Scan is a multi-national OEM for DEC and can supply complete turnkey systems anywhere in the world. The one additional item added to the workstation for VTRAK operation is the Mapstation Console. This replaces the mouse as a pointing device and provides extra function buttons which are the most ergonomic means of interacting with the display. All modes of VTRAK provide pop-up menus; keyboard input is only needed for the entry of file names.

Scanning. VTRAK will accept binary or greyscale raster data from a variety of scanners which may be chosen for speed, accuracy, resolution and other parameters to handle the required maps. Transfer to the immediate VTRAK environment can be by tape, network or direct on-line connection. Laser-Scan will configure particular scanners for turnkey systems. Parameters that need to be known for each image are type of scanner, image polarity, scanner resolution and size (in pixels) of the image. Small parts of the image can then be viewed to set parameters for thresholding, orientation, color separation and zoom factors.

Image Zone Definition. Some whole images and some areas of other images are suitable for fully automatic 'Autopass' feature-extraction in background mode. The areas to be selected for this are zoned by coloring them green at this stage with a variable sized rectangular window interactively moved across the image.

Autopass. VTRAK background mode feature-extraction is called 'Autopass'. While most of the feature extraction parameters are identical to those used in an interactive mode, there are additional functions such as differentiation of lines by their width, selection by minimum size and creation of an extra file of edit requests, where the automatic process realises that human intelligence will be required.

Interactive Feature Extraction. The heart of VTRAK is the interactive mode used for setting all feature extraction parameters and for interactive (but semi automatic) extraction of features selected by the operator. This mode is used for all areas not zoned green for autopass. In fact autopass can be delayed until after an interactive session so that difficult features can be extracted before the background process 'screws them up' ! Interactive VTRAK typically requires the following simple operations :

select feature with cursor controlled by tracker ball

press precoded start button that selects suitable parameters for that type of feature, codes the feature with the appropriate attribute(s), and starts the feature extraction process.

watch process of feature extraction which can be halted or reversed at any time and which can be displayed at any zoom factor.

guide process (if desired) in complex areas. At this stage VTRAK can also accept features digitised 'manually' off the screen image using the cursor at any zoom factor. This is how 'imaginary' features may be input or very low qality image data handled.

accept feature when completed. A single button press signifies the end of that feature, the writing of that feature to the output vector file and the 'paint-out' of that feature from the raster image so that it will

not be repeated or be able to confuse further processes. (Paint-out also happens in the Autopass mode).

edit features with image background. Presentation of the complete set of features overlayed on the original image gives a final edit and check capability with several special purpose functions. Exact registration is achieved via corner points or other registration marks on the image and a systematic pan and zoom routine gives a completeness and accuracy check. Procedures for checking attributes by selective display, for showing topological queries and for sensible handling of text and symbol features can also be introduced. Automated attribute coding for multiple-coded features or the building of 'objects' from basic features can also be supported.

output processing. VTRAK is a data capture system that can provide inputs to a variety of GIS or digital mapping systems including Laser-Scan's own comprehensive suite of programs. Other systems supported are Arc-Info, Intergraph, Synercom, Autocad and general purpose output to USGS DLG files or a variety of military formats.

VTRAK feature extraction algorithms are proprietary and do not use the 'skeletonising' approach used by many raster to vector systems. They use all available pixel information to extract the centreline of a multi pixel feature or the centre of a symbol. Similarly, nodes in line networks are extracted from an array of pixels and not just from the junction of skeletonised lines. VTRAK does not suffer from 'hairs' and 'junction kinks' often seen in standard vectorised output. VTRAK algorithms are also designed to 'fail safe' and call for help rather than struggle with bad data.

The emphasis of the VTRAK philosophy is toward flexibility stemming from the recognition that no two map specifications are alike and that every map is unique in itself. VTRAK provides managers with the ability to give the actual operators of the system as much or as little flexibility as their training and the type of work demand. Tuning of parameters can be reserved for management or left to the experienced operator.

NORTH AMERICAN APPLICATIONS

VTRAK has been benchmarked in N. America against competing products for utility and topographical mapping applications. These benchmarks have shown that VTRAK is superior to all other products as a cost effective scanned data capture system. VTRAK is now installed at the Canada Centre for Geomatics (Energy Mines and Resources) where it is used primarily for the 1:50000 series.

USGS 'quad' sheets are ideal sources and some special features of VTRAK are applicable to these maps :

single operation removal of contour labels and joining of contour gap across label.

simultaneous heighting of multiple contours.

recognition, orientation and coding of multiple house symbols.

measurement of road centre lines from cased road representation.

extraction of all types of dashed & dotted lines (eg boundaries, tracks, intermittent rivers).

topological structuring of river and road networks to DLG standards.

Some of the facilities are used on the LASERTRAK scanners at USGS where contours (hypsography) and rivers (hydrography) are being captured for the National Cartographic Database (Moreland 1986).

Utility maps require a different emphasis because they typically have less linework, more text and symbols and are often of lower quality. There are no simple answers but VTRAK gives a superb environment for the development of very efficient routines with menu, keyboard or voice entry of attributes, on-line intervention in critical areas and creation of network topologies. Both land base and plant data can be captured and kept separate so that, for example, the land base can be sold to other utilities, local government and other users (Cross 1987).

Thematic maps of soils, geology, water resources, geodemographic data, vegetation, city zoning boundaries etc. are vital inputs to GIS. VTRAK not only enables the user to input the linework with basic coding, but also provides polygon building and coding facilities.

EUROPEAN APPLICATIONS

Most of the facilities described above apply equally to European maps except that the range of map specifications is much greater and perhaps more centrally regulated. Laser-Scan have been very involved in setting the UK standards for digital mapping and have developed specialised routines for digitising Ordnance Survey large scale maps (1:1250, 1:12500) and for quality controlling the process to the National Joint Utilities Group (NJUG) standards.

In continental Europe the base mapping for local government and utility applications is the cadastral map used primarily for recording land ownership and for taxation. These maps are highly accurate and typically feature parcels surrounded by boundary lines which connect beacons denoted by hollow circles. The accurate

measurement of these circles, their topological connection to others via the boundary network and the addition of attributes to both beacons and parcels are fundamental to the data capture process. VTRAK has 'beacon recognition' facilities and enables the operator to capture a parcel at a time together with associated attributes.

CONCLUSION

During the **1970**s extravagant claims were made about the efficacy of raster scanning and vectorisation : Laser-Scan developed the FASTRAK line following scanners, which are still in use today.

During the **1980**s realisation that current raster scan capabilities were way ahead of the software for raster to vector conversion led to the increased use of raster mode drawing management systems and to the proliferation of conversion houses using well tried but time consuming manual digitising systems. Laser-Scan introduced the LASERTRAK vector scanner which is in use in the USA, UK, Japan, Italy, Norway and the Middle East.

During the **1990**s both raster drawing management and vector database systems will coexist and will need to exchange data for efficient use of GIS resources. Laser-Scan have introduced VTRAK to meet the need for feature extraction from raster map images as and when it is required.

VTRAK is available now !

REFERENCES

Cross D. A. 1987, Intelligent Scanning of Maps and Plant : European AM/FM Conference, Montreux

Moreland D. K. 1986, Development of an Automated Laser-Based Data Capture System at the US Geological Survey : ACSM-ASPRS Spring Convention, Washington

Woodsford, P. A. 1988, Vector Scanning of Maps: Proceedings of IGC Scanning Conference, Amsterdam