Automating Production of Digital Nautical Charts

David Opitz
The University of Montana
Computer Science Department
Missoula, MT 59812
opitz@cs.umt.edu

Stuart Blundell
Visual Learning Systems, Inc.
PO Box 8226
Missoula, MT 59807
sblundell@vls-inc.com

Chris Finlayson
Visual Learning Systems, Inc.
PO Box 8226
Missoula, MT 59807
cfinlayson@vls-inc.com

Andrew Mowry
Visual Learning Systems, Inc.
PO Box 8226
Missoula, MT 59807
amowry@vls-inc.com

Matthew Morris
Visual Learning Systems, Inc.
PO Box 8226
Missoula, MT 59807
mmorris@vls-inc.com

Abstract

This research discusses benefits of the synergistic use of Feature Analyst and the Production Line Tool Set (PLTS) over traditional methods in extracting features from scanned digital nautical charts. We start by describing algorithms to modify the line and point predictions from Feature Analyst to meet NIMA specifications. Though Feature Analyst is used traditionally with satellite imagery as opposed to scanned maps, test results demonstrate that Feature Analyst in its current form can automate several time-consuming tasks. Further we propose additional tools for improving the quality of results

including (a) manual clean-up tools designed for refining predictions and (b) an automated convert-to-line process.

Introduction

Because GIS has become the norm for handling of geospatial information, there is a need for converting vast numbers of paper maps into digital format. For maximum efficiency, partial automation of this task is essential. Some software products that extract features from scanned maps produce reasonably good results, but certain types of maps present special challenges. Digital Nautical Charts (DNCs) are one example. They require high standards of accuracy and contain a large number of feature types, many of which must contain specific attribute information. For example, depth soundings exist as decimal numbers on the charts and need to be extracted as precisely placed point features attributed with the depth.

Feature Analyst® is a software package that uses machine learning technology to automate extraction of object-specific features from image data. It uses user-defined training examples identify features such as buildings, roads, and land cover and returns them in vector format. Although it is commonly used with remotely sensed image data, Feature Analyst can be used on any images and performs well extracting features from scanned paper maps. Feature Analyst is an extension for ArcGIS, and optionally interfaces with ESRI's Production Line Tool Set (PLTS). The result is seamless end-to-end map production including raw feature extraction, automated cleanup, hand editing, symbology, attribution, quality control, and map layout.

In this paper we discuss the use of Feature Analyst for extraction of DNC features, including new tools designed specifically for scanned maps that refine initial polygon results and automatically converts them to lines and attributed points. The new tool will also allow for easier hand- editing procedures.

Automated Extraction of Digital Nautical Chart Elements

Below we outline DNC classification approaches. Also, because Feature Analyst returns all initial results in polygon vector format, we describe approaches to converting geometry to points and lines to conform to the National Geospatial Intelligence Agency (NGA, formerly NIMA) standards, as well as automated cleanup and quality control.

Point Features

To train Feature Analyst to recognize point features such as depth soundings and moorings, the user creates training examples that are centered on the features of interest. For depth soundings, we train Feature Analyst to recognize each of the digits 0-9, plus subscript versions of the digits 1-9. This is necessary because depth soundings are written without a decimal point. Decimal digits are written as subscripts. Thus, the soundings represent a nineteen-class problem and we give training examples for each digit (Fig. 1). Feature Analyst uses spatial context to recognize shape and size, which allows it to distinguish between the nineteen classes.
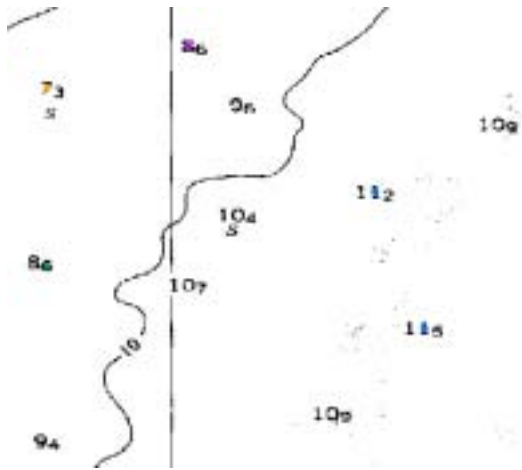
Figure 1. The user digitizes points as examples to train Feature Analyst to recognize the digits in nautical chart depth soundings. Feature Analyst must differentiate between nineteen types of digits.

We threshold the scanned chart to create a black and white image that requires less processing time and allows removal of extraneous colored symbols. When the user runs the classification, Feature Analyst decides, as a final step, which single-pixel result within a user-defined window most closely resembles a training example. This ensures that the center of each digit on the scanned map is represented by a point of the appropriate feature class. The initial results are polygon features the size of a single pixel, and the convert-to-point tool creates points representing centroids of polygons while maintaining distinct classes (Fig. 2).
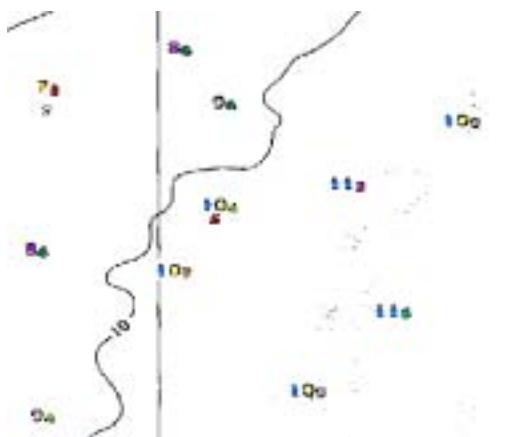

Figure 2. The initial soundings results include a separate point for each digit, with a class attribute representing the value of the digit.

3

Next, we use standard Feature Analyst hierarchical learning techniques to remove false positive results ("clutter"). This involves specifying correct and incorrect results and running the extraction again, allowing Feature Analyst to remove incorrect results while adding to its learned body of knowledge about each feature class.

A specialized "cluster points" tool then combines all of the digits in each sounding into a single point, which is automatically centered on the whole number digits, to meet the NGA specifications. Feature Analyst attributes each final point with the value of the soundings by interpreting the relative locations of the digits (Fig. 3).

The extraction procedure is similar for other point features, although most are extracted as single classes and do not require attribution until later in the workflow. Feature Analyst recognizes the shape of many point types, for example, the anchor symbols used to designate mooring locations.
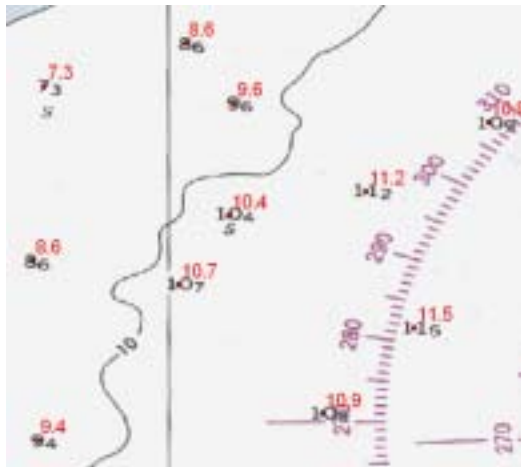

Figure 3. The final results are centered on the whole-number portion of the soundings, and attributed with the depth value (shown here as red labels).

Line Features

Several types of line features on the DNCs separate two different colors of fill. Because Feature Analyst works well extracting areas, we extract these types of lines by first classifying color separations and then converting their borders to lines. For example, with shoreline classifications, we classify all land as one class and all water as a second class. We choose training examples for each class to be representative of all possible colors and patterns within each area (Fig. 4). Feature Analyst aggregates the initial polygon results (Fig. 5), if necessary, with a tool that removes small features and fills holes smaller than a specified area. Next, we change the polygon edge to line geometry.

Because the initial lines often are often not located precisely (Fig.5), we have developed algorithms that relocate the lines to the black pixels in a black and white version of the scanned map. A line adjustment tool applies these algorithms in various sequences depending on user inputs. Figure 6 illustrates the results before and after the line adjustment. Below we describe the individual algorithms used by the tool.

Figure 4. Training examples used for classifying the color separations for land and water.
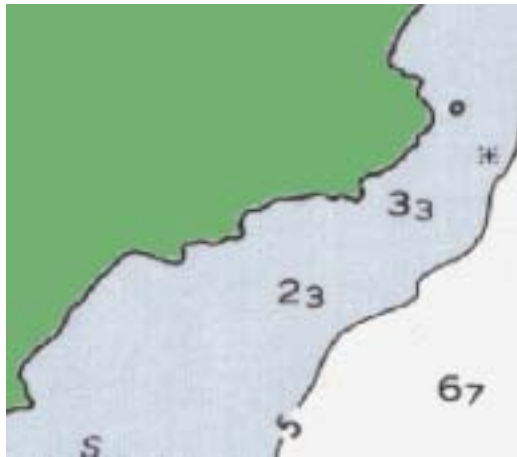

Figure 5. Initial results for the land color separation. The results also include a class for water, which has been turned off in this view.
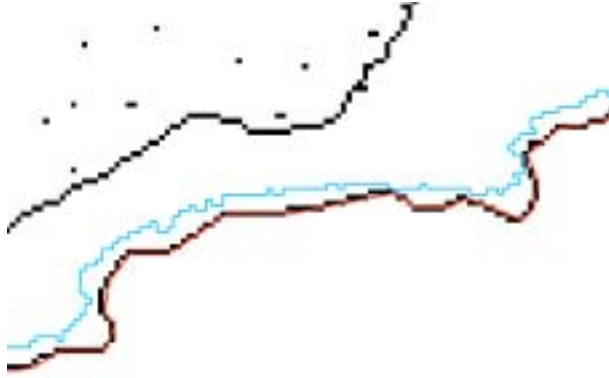
Figure 6. The line adjustment tool. The initial results are shown in blue, and the red line has been adjusted to the black and white raster image.

The "commit" algorithm checks that each vertex falls on a black pixel in the underlying raster, and if not, moves the vertex to the nearest black pixel within a specified distance window size. Thus, it moves the entire line onto the black lines in the raster. The "recenter" algorithm then moves the line to the middle of the black line in the raster. A culling algorithm removes excess vertices in order to smooth the line. The "recommit" algorithm locates instances where the line jumps from one underlying black line to another, and then back again, and relocates such segments. The "circumscribe gaps" algorithm locates places where bends in the lines have been cut off, and moves the line to follow the bends (Fig. 7). Finally, the "simplify" algorithm cleans line segments that are jumbled or doubled back on themselves.
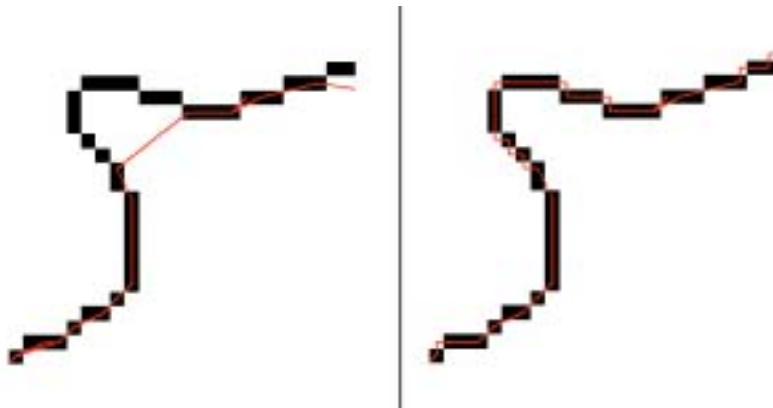

Figure 7. The "circumscribe gaps" algorithm, before (left) and after (right).

Feature Analyst includes additional tools for automatically jumping any gaps between line segments of a specified size and for removing dangles of less than a given length.

Some types of lines, such as certain bathymetry contours, separate areas of the same color, so Feature Analyst extracts the actual lines in these cases instead of color separations. The extraction procedure is the same, but the training examples are narrow polygons that represent the color of the line itself. Feature Analyst is able to use spatial association to learn to locate narrow areas of black, for example, that are surrounded by white on both sides. Previous color separations are used to mask unwanted areas, such as land, in order to simplify the procedure. We use a "convert to line" tool to collapse the initial polygon results to line features. User settings specify distances for jumping gaps and removing dangles, as well as a choice of line smoothing algorithms (Fig. 8). Finally, the line adjustment tool adjusts the line to the underlying raster as described previously (Fig. 9).
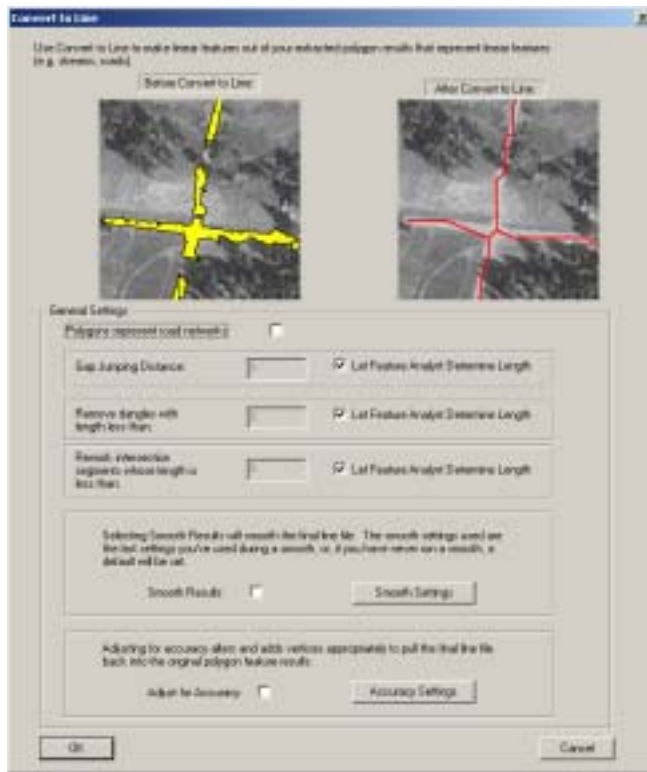

Figure 8. The interface for the Feature Analyst "convert to line" tool.

Figure 9. A bathymetry line extracted using Feature Analyst. The initial polygon results were converted to a line, jumping gaps where necessary, and the line was adjusted to the underlying raster.

Batch Processing

Once we obtain good extraction results for specific feature types on each type of DNC, it is possible to use the "learning model" that Feature Analyst creates to extract the same features on subsequent charts that are of the same source (i.e. NOAA, UK, Foreign, etc.). The learning model contains all information that the software learned from classification passes and automated cleanup steps such as the point clustering and attribution, conversion to line, and line adjustment. It is possible to select multiple image files and multiple learning models at once, allowing fully automated extraction of many feature types from multiple maps in a single step.

Hand Editing

Hand editing is often a necessary step before finalizing the extraction results. We have developed two tools for editing line and polygon features with maximum efficiency. Both tools allow several different editing functions, depending on the control key used and the location of clicks relative to existing features, to enable the user to perform many common editing tasks without switching tools. The Polygon Reshape tool allows extending and trimming polygon features, cutting holes and removing holes in polygons, and cutting of polygons into multiple parts. The Line Reshape tool allows connecting multiple features, adding and removing segments, repositioning segments, and connecting existing segments while automatically removing unwanted ones.

PLTS

ESRI's Production Line Tool Set (PLTS) is another extension for ArcGIS that facilitates production to meet specific cartographic standards, beginning with raw data. As such, it is ideally suited for finalizing features extracted with Feature Analyst to meet production guidelines such as those for DNCs.

After Feature Analyst extraction, the user can migrate the shape file results to a Geodatabase using Feature Analyst and PLTS tools. PLTS then allows attribution of features by selecting from a list of attributes in a "Favorites Geodatabase". Next, additional hand editing can be done using PLTS or Feature Analyst tools.

8

At this point, it is possible to import predefined symbology from an existing Geodatabase and aspects of geometry can be batch validated using predefined rules. PLTS records errors in an error table and allows the user to step through the errors in a map view. Next, visual validation is done. A PLTS "QC Grid" is created, which divides the data extent into a hierarchy of grids, allowing the user to step through the grids, checking for errors, while logging errors in the error table and recording the portions of the grid for which validation has been completed.

After additional quality control steps, the final results can be made into maps using the PLTS Map Production System. The user designs a standard map series, which is then used for production of individual map sheets by drawing in completed datasets.

Summary and Conclusions

Feature Analyst automates much of the work needed to convert scanned nautical charts into digital vector format. Feature Analyst tools designed specifically for this type of work convert initial extraction results to the proper geometry and allow automatic and manual cleanup of results. For nautical chart depth soundings, Feature Analyst differentiates between nineteen classes of numeric characters. Specialized tools return points at the center of each character then cluster these to a single point for each sounding, and attribute them with the proper depth. Information learned by Feature Analyst through user supervision can then be used to classify additional maps without further input.

When combined with the ESRI's Production Line Toolset (PLTS), Feature Analyst makes it possible for the entire extraction and map production process to take place within the ArcGIS environment. PLTS takes Feature Analyst results and allows editing, attribution, quality control, and final production to meet existing guidelines.

Future work will involve further integration of Feature Analyst with the PLTS environment and continued improvement of extraction techniques and existing tools for refining classification results.