AUTOMATED FEATURE EXTRACTION: A HISTORY

Philip Kern, Program Manager – The Americas
Earth Imaging Solutions Centre
Intergraph Corporation
230 Business Park Blvd.
Madison, AL 35757
phil.kern@intergraph.com

ABSTRACT

Automation of the feature extraction process has been the “Holy Grail” of the photogrammetric data collection industry for many years. This paper discusses the various attempts at automating feature extraction from the late 1980’s to the present including rasterization of line maps, automated text recognition, automated DTM collection and the results of DARPA grant research. The pitfalls of film-based imagery are discussed as well as the advantages of digital imagery. New techniques of image processing show promise such as hierarchical feature classification. One of these packages is discussed in detail, but at the scales needed for photogrammetric planimetric feature collection, l only about 80% of features are successfully extracted. Those missed or misclassified still take longer to fix than to collect the data manually to begin with.

INTRODUCTION

Automation of the feature extraction process has been the “Holy Grail” of the photogrammetric data collection industry for many years. In the late ‘80s and into the early 1990’s automated paper map conversion (rasterization) was popular. Both automated and semi-automated methods were used. The automated methods were very good at line rasterization but not very good at rasterizing text. Only about an 80% success rate was achievable and it took far more time to correct the improperly rasterized text.. Semi-automated software was written that would prompt the user for text entry. This was much more robust and much faster than trying to correct the error prone automated text recognition.

In 1990, the possibility of automating robust DTM collection from stereo pairs was realized in digital photogrammetry. This was a real breakthrough for extracting DTM points to generate contours and breaklines. Of course, image processing were available to analyze satellite imagery and produce GIS data directly through the use of supervised and unsupervised image classification for small scale image data.

At about the same time that DTM collection had become automated, DARPA (Defense Advanced Research Projects Agency) issues a tender for speeding up and automating feature extraction. Intergraph competed for this project along with many other vendors and research organizations. The Intergraph team was the successful in winning the tender and proceeded with the work.

After much hard work with some of the best minds in business and academia, work with many different types of algorithms for automating feature extraction including SNAKES, it was finally concluded that at the time, a well designed digitizing menu would make the operator far more productive than the automated procedures. As in the automated text recognition case previously mentioned, achieving an 80% success rate was not cost effective, less expensive to start from scratch rather than fixing what the computer got wrong

THE QUEST

What we are trying to achieve is the quest for the Holy Grail in photogrammetric feature collection. The reason this effort is referred to as the Holy Grail is that it is so difficult and has major benefits if it can be automated completely. It is the last major portion of the photogrammetric workflow that is still very labor intensive. If this can be cracked, methods for automating the entire workflow are not far behind.

Unfortunately, this portion of the workflow is one of the most difficult at planimetric feature collection scales. Image features vary from region to region and possibly from house to house, and even photo to photo, especially in the film-only days. Film was not a consistent medium, therefore the automation would not work well. Textures such as those on roofs vary greatly, e.g. asphalt roof shingles, vs. metal roofing, vs. tile roofing. Another important issue is the software's ease of use, an easy method of training the software is necessary. And of course, a much higher success rate than 80% must be achieved.

PROMISING TECHNIQUES

Better than the classic image processing techniques, image segmentation and hierarchical classification has been able to show very promising results. An example of this is the E-cognition software from Definens that has been used successfully has been used successfully in extracting image data. As can be seen in Figure 1, the resulting

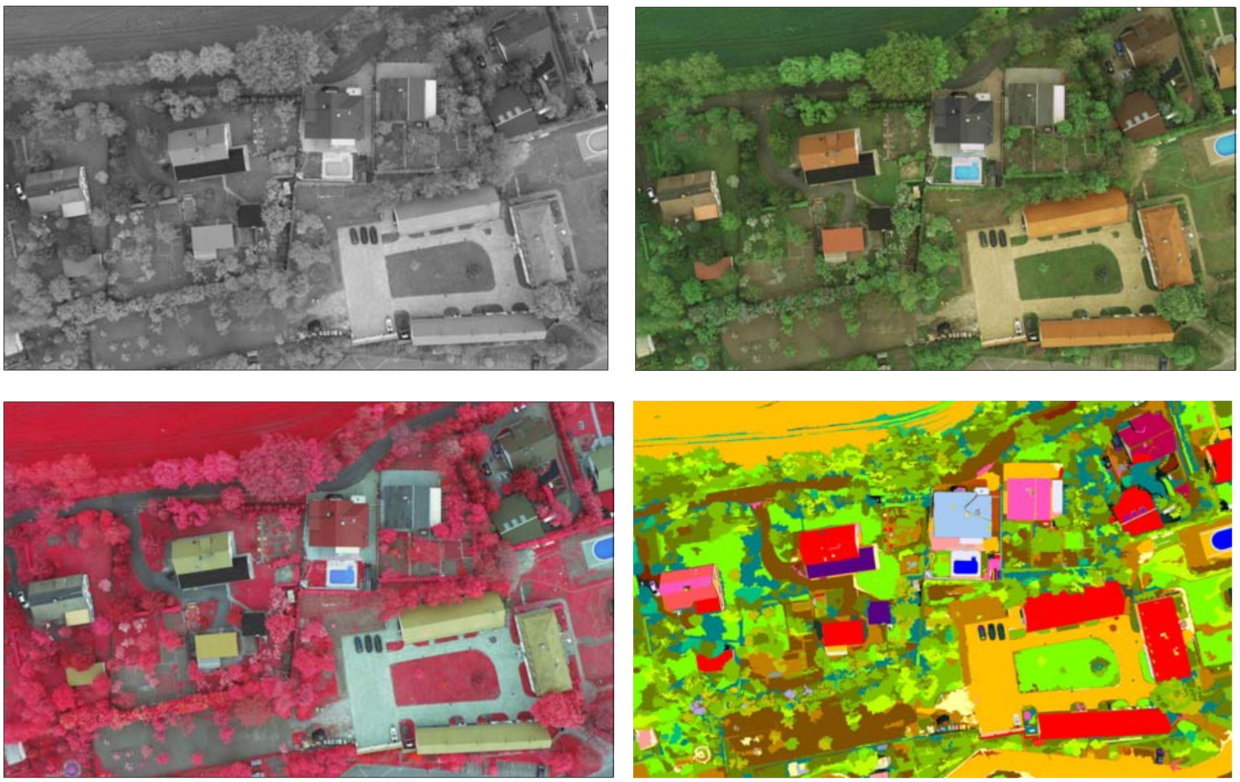

Figure 1. Portions of an image from the Digital Mapping Camera (DMC) over a portion of Germany. Couter-clockwise from upper right – color, panchromatic, color IR, and classified image from E-cognition. Data courtesy of ILV Wagner.

classified image can contain many classes. In this example a 4 band stacked data set was used as shown. Even though all of the roofs are tile roofs, there are at least six classes of roof in the image due to slight differences in color, texture, and shadow. This illustrates that something as seemingly simple as finding roofs in order to collect roof outlines can become very complex.

This type of classification system can be very detailed but can also be quite complex to use and to extract useful results.

As a result of the NGA Softcopy Search program, Intergraph was looking for software that would do automatic feature extraction to be included in the suite of products being assembled for NGA. The software needed to be easy to use, customize, and train to find feature classes. Intergraph chose Feature Analyst from Visual Learning Systems.

Feature Analyst had already enjoyed great success in many U.S. government agencies for many reasons. Among other applications, Feature Analyst claims to acquire defense "Features of Interest"; road, building and drainage extraction as well as vegetation and land cover mapping and tying this data into an underlying GIS database. As it turns out, the software can do many things very well.

For point features, it does a fairly good job of extracting features. During extraction it uses an oibject's color, size, shape, texture, pattern, shadow, and spatial association. An example of this can be seen in Figure 2.

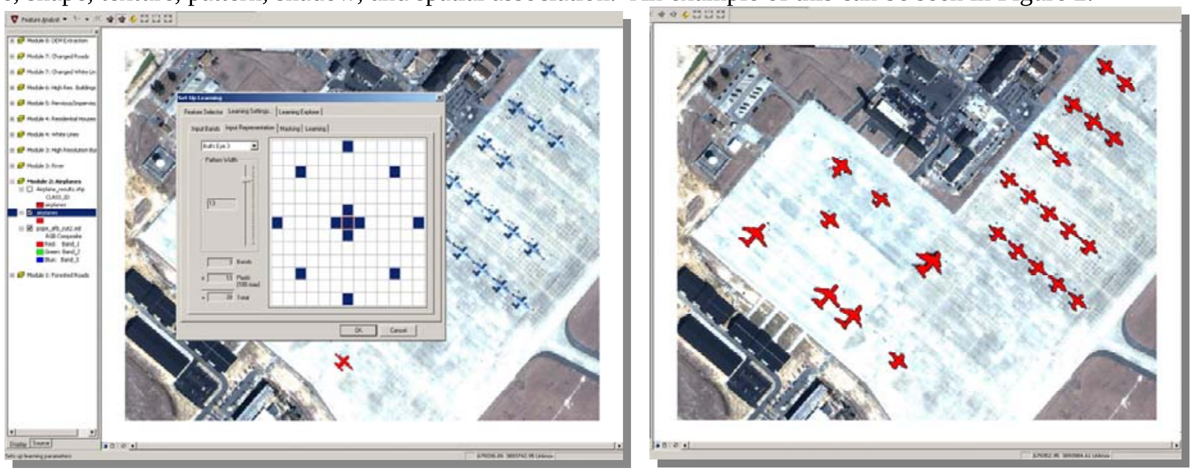

Figure 2. Feature Analyst tool for point feature extraction and training on the left, on the right the results of the feature identification.

The software has also proven its accuracy in digitizing test performed by several organizations. As can be seen in Figure 3, both land cover and drainage areas were manually digitized. The same operators who digitized the areas trained the software to extract the same features.

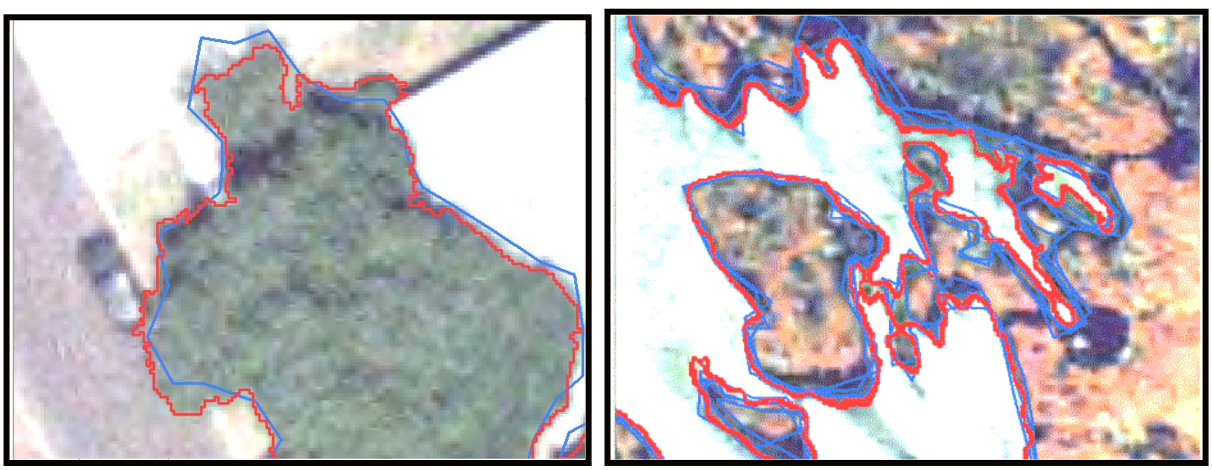

FIGURE 3. Land cover digitizing test on the left and drainage digitizing test on the right. Manual digitizing from four different analysts is shown in blue. Automated extraction results from training sets created by the same four analysts are shown in red.

The software also does a very good job of extracting features using change detection. So how well does this all work? Have we found our Holy Grail? Unfortunately, for the scales typically used in planimetric mapping, Feature Analyst is better at identifying point data. In order to collect features more consistently and get them into a GIS database, the software needs help for planimetric feature extraction.

This is where tools such as Intergraph's Geo-Intelligence Production System (GIPS) can supplement and enhance the feature extraction process. Using tools such as validation routines and templates, the software suite organizes the workflow into a smooth process. Data validation of attribute data, geometry and connectivity are integrated into the process. In addition, data management tools for revision, conflict detection and resolution and data history are performed and maintained.

These tools allow automated feature extraction in concert with geometric validation. They also provide automated dimensional attribution as part of the workflow. An example of this is shown in Figure 4.

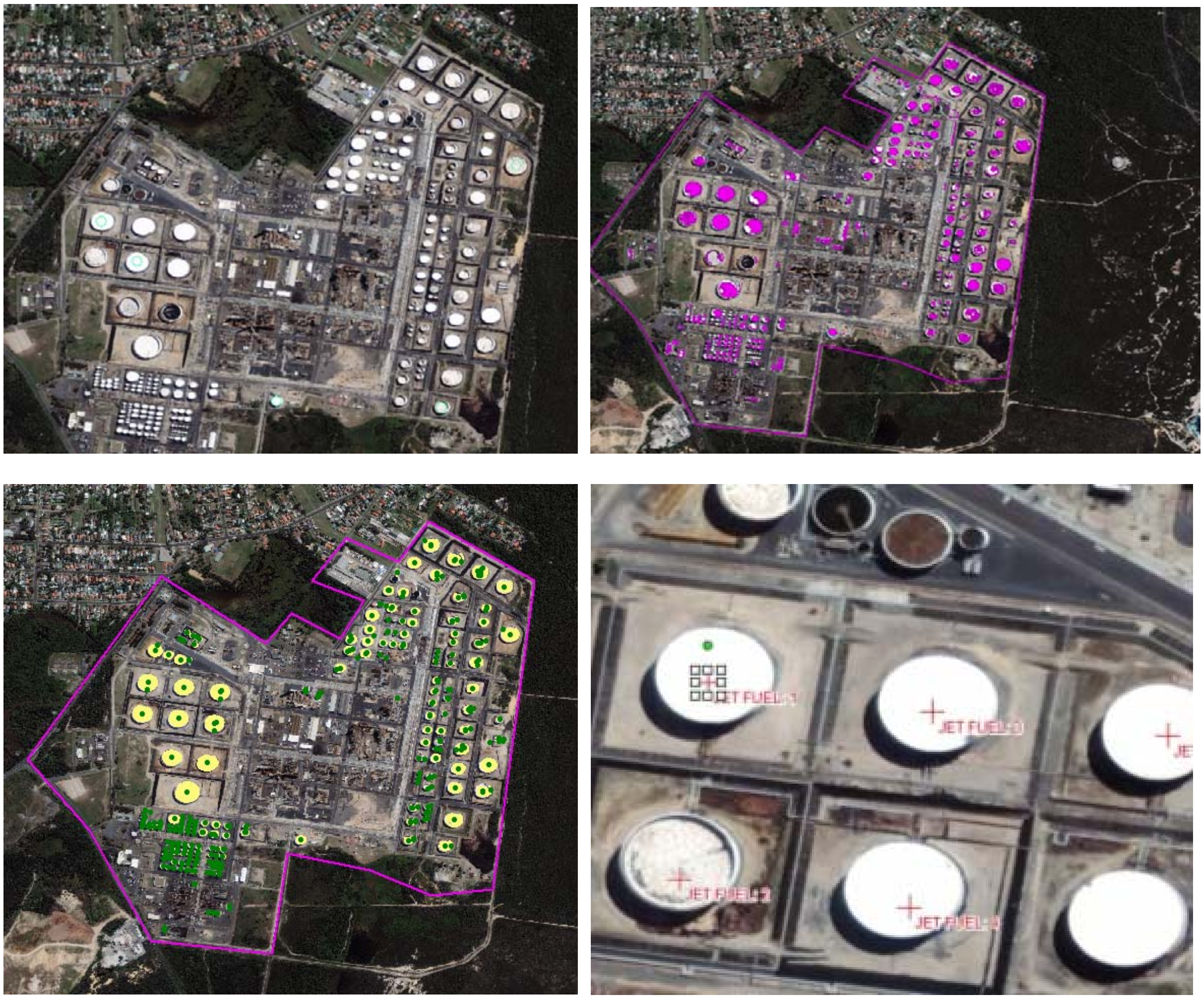

FIGURE 4. Example of feature extraction and attribution workflow. Fuel tanks in upper left image are used as a training data set to locate all fuel tanks in the upper right image. Lower left image converts the area features into point features. A tank size may then be calculated to further refine extraction and attribution. Finally the attributed tanks are shown in the lower right image.

CONCLUSIONS

At this point in time, automated feature extraction is a tool that is useful for partially automating the workflow, it has gotten easier to train the software tools and to use them. It can be trained on small areas and used on large areas. Still only about an 80% success rate for the features we are trying to extract. This rate will need to come close to 100% for roads and houses to be really productive. We are close but we haven't found the Grail yet.