Case-Based Reasoning and Software Agents for Intelligent Forest Information Management

Daniel Charlebois, David G. Goodenough, A.S. (Pal) Bhogal and Stan Matwin[1]

Pacific Forestry Centre, Victoria, BC, Canada
[1]University of Ottawa, Ottawa, ON, Canada

Abstract - **To perform forest information management, SEIDAM integrates forest cover descriptions, topographic maps and remote sensing imagery. SEIDAM relies on an on-line robotic data storage device, image and GIS metadata databases, software agents and a case-based reasoning system to deliver information to decision makers in a timely fashion. The image and GIS metadata databases contain information about the sources of data, where the data are stored, where they have been delivered and the processing they have undergone. The software agents perform the actual processing by running image analysis, GIS, database and other software to accomplish specific tasks. The case-based reasoning system relies on the software agents, past experience from domain experts and information from the metadata databases to determine what processing is required to deliver products satisfying user goals. This paper describes the intelligent inventory update function in SEIDAM and its AI methodology.**

I. INTRODUCTION

As part of NASA's Applied Information Systems Research Program, a project is being conducted to create an intelligent system (SEIDAM [1-3]) which manages and fuses remote sensing data from aircraft and satellites with multiple geographic information systems (GIS) in order to respond to queries about forests and the environment. SEIDAM (System of Experts for Intelligent Data Management) was begun in 1991 under the Applied Information Systems Research Program of NASA. SEIDAM partners include NASA, the Government of British Columbia (Forests and Environment, Lands and Parks), the Joint Research Centre at Ispra, the Royal Institute of Technology in Stockholm Sweden (KTH), MacDonald Dettwiler, and the universities of BC, Victoria, and Ottawa. SEIDAM integrates several key technologies: image analysis for remote sensing data, geographical information systems (both vector and raster based), artificial intelligence (AI), modeling (growth and yield) and visualization. User queries range from simple relational database queries to complex queries such as the update of digital forest cover maps. This second type of query may require SEIDAM to automatically perform image analysis tasks to extract surface features necessary to update the digital maps.

The focus of this paper is on the design of the AI components of the system, their integration into SEIDAM and how they successfully process remote sensing data and GIS data.

A brief overview of the SEIDAM components is presented in the next section. The following section gives a detailed description of an example of SEIDAM at work on a forest cover update problem. The flexibility of SEIDAM is discussed in the fourth section followed by some conclusions.

II. SEIDAM: THE SYSTEM

The components of SEIDAM are shown in figure 1. They are: the main expert system, the reasoning system (PALERMO - Planning and LEarning for Resource Management and Organization [4]), the Smart Access software agents (software agents [5-7] are systems that unobtrusively observe the manner in which they are used, adapt to the tasks for which they are used, as well as learn from the circumstances of their use), the remote sensing software agents, the GIS software agents, the image and GIS metadata databases, the image and GIS data recorded on a robotic mass data storage device, the SEIDAM knowledge base and the reasoning system's case-base.

The SEIDAM expert system provides a graphical user interface (GUI) through which users can enter simple queries that trigger a range of simple to complex processing tasks. It also controls the activation of the Smart Access software agents and the case-based reasoning system. The Smart Access software agents provide an intelligent interface between SEIDAM's knowledge base and an image metadata database and a GIS metadata database.

0-7803-3068-4/96$5.00©1996 IEEE

The case-based reasoning system [8-10] uses three sources of information to determine how user queries can be satisfied. The first source is its case-base where generalized solutions to previous problems are stored. The second source is the SEIDAM knowledge base where the Smart Access software agents place all metadata relevant to the current problem. The final source is a set of remote sensing and GIS software agents. Each agent creates a STRIPS-like [11] planning operator describing the task it performs.

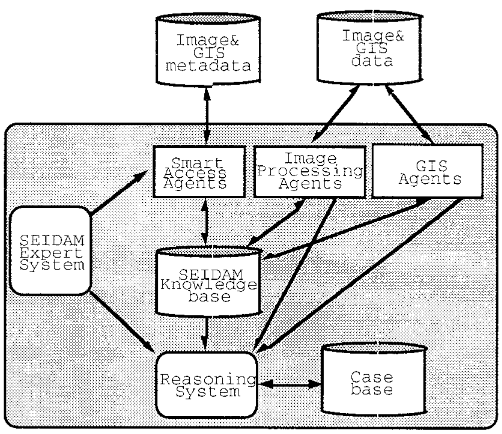

Figure 1 - SEIDAM

There are approximately 40 remote sensing and GIS software agents. They are implemented as expert systems that use third party software such as the PCI image analysis package and the Arc/Info GIS package, plus our own image analysis software, RIASSA. Each agent performs a specialized task such as copying files from their long term storage location to working directories; translating data between image file formats and GIS file formats; creating digital elevation models from point elevation data; applying topographic relief correction algorithms to remote sensing imagery, and performing analysis, segmentation and classification.

The image and GIS metadata databases provide information describing the owner, source, location, distribution privileges, processing history and other ancillary information relating to the remote sensing and GIS data available for SEIDAM. The GIS metadata database complies with the US' Federal Geographic Data Committee's Content Standards for Digital Geospatial Metadata. The image and GIS data are stored in a robotic mass storage device capable of containing over 1.1 terabytes of data. The robotic store is managed by a hierarchical management system that allows user's to treat the device as a large disk. The prime contractor for this robotic device was Digidyne Corporation in Vancouver, BC.

III. SEIDAM: THE SYSTEM AT WORK

One of the more complex problems SEIDAM must solve is the update of a digital forest cover map by using remote sensing images to detect forest depletion due to logging or other phenomena. This task can be carried out by overlaying a GIS digital forest cover map on an LANDSAT-TM image and allowing a user to digitize the lines around new clearcuts, thus creating updated polygons that can be added to the GIS. A geocoded TM image must also be corrected for topographic relief before it can be used for this purpose. Digital elevation models (DEMs) are used as the basis to correct the image for topographic relief. The example below describes how a TM image is prepared by using a DEM created from 1:20000 TRIM irregularly spaced elevation data.

The process is initiated by the user selecting from a list of functions the forest inventory update function in the SEIDAM forestry products window. Subsequently, the user specifies the spatial selection in the Spatial Specifications window. The administrative mapping structure for two provincial ministries guides the user to selecting digital GIS files either symbolically or by map sheet number. Once the map set selection is complete, SEIDAM activates the Smart Access software agents. Their task is to start the relational DBMS where the image and GIS metadata reside and submit SQL queries that will identify all of the GIS data available, as well as all of the TM imagery covering the selected maps. The metadata returned by the SQL queries will include the processing the data have undergone. The Smart Access agents then place this information in the SEIDAM knowledge base and shut down the DBMS.

SEIDAM then activates the case-based reasoning system to which it submits the user goal "update the forest GIS file using TM imagery over the selected map set". The reasoning system poles each of the remote sensing and GIS agents for their planning operator descriptions. It then attempts to find a case that solved a similar problem. Similarity is measured in terms of the user goal and sub-goals and the sub-goals of the case as described in SEIDAM's knowledge base. If a case is found, it is modified to account for the differences between the new user goal and the previous update problem. If a case is not found, the reasoning system will use goal regression to find a solution. The solution will consist of an ordered list of processing tasks that must be executed to update the forest cover. Each processing task corresponds to one remote sensing or GIS software agent. The reasoning system assesses the processing states of the remote sensing and GIS data and only includes in the solution agents for which processing is required.

The reasoning system activates each of the processing agents listed in the solution. As each agent executes, it modifies the contents of SEIDAM's knowledge base by adding and deleting information and performing the necessary processing. Upon successful execution, the information added and deleted from the knowledge base should agree with the add and delete lists described by the planning operator provided by each agent. This allows the reasoning system to insure that the processing carried out by each agent is successful before activating the next one.

The solution for the forest update problem, after Smart Access execution is:

```
copy_files_to_working_dir,
trim_to_dlg,
import_hydrology,
erdascp_lakes,
sieve_lakes,
lakes_to_bitmap,
trim_to_dem,
create_tin,
erdascp_dem,
copy_bitmap,
set_georeference_tm_large,
create_tm_small,
dem_to_pix_image_match,
set_georeference_tm_small,
tercom,
export_clear_cuts_to_pci,
digitize_clear_cuts,
export_clear_cuts_to_arc,
```

- `copy_files_to_working_dir`: this agent will copy the TM image and GIS files from the robotic mass storage device to a working directory on disk.
- `trim_to_dlg`: this agent translates TRIM GIS data into the digital line graph standard readable by Arc/Info.
- `import_hydrology`: this agent creates an Arc/Info hydrology coverage.
- `erdascp_lakes`: this agent translates the hydrology coverage into a PCI image file.
- `sieve_lakes`: since the hydrology data contains many small lakes, a smoothing filter is used to remove any lake smaller than nine pixels or 5500 m^2.
- `lakes_to_bitmap`: convert the smoothed hydrology image into a bit map.
- `trim_to_dem`: creates a point elevation file readable by Arc/Info from TRIM data.
- `create_tin`: create a DEM from a triangulated irregular network constructed with the point elevation file and place the result in an Erdas file format for input to PCI.
- `erdascp_dem`: create a PCI image file from the Erdas DEM file.
- `copy_bitmap`: copy the hydrology bit map into the DEM image file.
- `set_georeference_tm_large`: get geo-reference information from the TM image file and DEM image file and place it in SEIDAM's knowledge base,
- `create_tm_small`: create a small TM file that fits over the current map and copy a subset image from the original TM image file.
- `dem_to_pix_image_match`: copy DEM into small TM file.
- `set_georeference_tm_small`: add geo-reference information to small TM file.
- `tercom`: apply topographic relief to small TM file.
- `export_clear_cuts_to_pci`: export old clear cut vectors from Arc/Info to PCI.
- `digitize_clear_cuts`: allow user to digitize new clear cuts.
- `export_clear_cuts_to_arc`: export the new clear cut vectors from PCI to Arc/Info.
- `generate_products`: create any paper maps or tabular summaries of forest cover changes.

After the agents have all executed, SEIDAM, once again, activates the Smart Access software agents. Their task now is to start the relational DBMS and submit Data Manipulation Language (DML) commands to update the image and GIS metadata databases to reflect the processing that has been carried out on the image and GIS files. Once this has been completed, the Smart Access agents will move the image and GIS data into the appropriate locations on the robotic data storage device.

IV. DISCUSSION

The procedure described above can be expanded to cover other sensors, automated segmentation and labeling of clear-cuts, or the use of other GIS software. For example, suppose the Smart

Access agents do not find TM data corresponding to a particular map sheet, but do find airborne optical AVIRIS data. The AVIRIS data would be copied to the working directory, and additional agents invoked to select the appropriate bands, perform atmospheric correction with MODTRAN 3, and to geocode the imagery. The topographic correction of AVIRIS data to the same accuracy as TM is difficult and it would likely be necessary to relax the geocoding standards of the GIS products.

The imagery would be segmented by creating an edge image based on edges in any band, inverting this edge image, finding the valleys in the inverted edge image, and building homogeneous segments. The domain expert would have specified in training the agents which filters to use for each sensor, the noise smoothing to apply, and the minimum mapping unit for segment size. The user would still need to validate the automated clear-cut labeling, but the automated process would dramatically reduce user efforts. Intelligent agents for such segmentation have been created and implemented in SEIDAM. They replace the "`digitize_clear_cuts`" agent in the automated solution.

SEIDAM connects to two other GIS packages, GRASS and PAMAP. Both are raster-oriented GIS packages. GRASS is free, but lacks an effective connection to a relational DBMS. It would not be able to support the many attributes of forest polygons. PAMAP can connect to Oracle and achieve functionality for the forest update similar to Arc/Info.

Finally, the update procedure has been modified to support the creation of forest fire fuel maps for planning responses to fire outbreaks.

V. CONCLUSION

The SEIDAM sequencing of intelligent agents for forest inventory update has been described. SEIDAM allows the rapid creation of processing chains which enhance software reuse, which are flexible, and which can operate over an ATM optical fibre network. Metadata databases are used to ensure easy access to a terabyte of remote sensing and GIS data held in a robotic data store. Future work includes the expansion of the update function to incorporate automated segmentation, use of airborne sensors, and the integration of a second DBMS, Oracle.

VI. REFERENCES

[1] S. Matwin, D. Charlebois, D. G. Goodenough, and P. Bhogal, "Machine Learning and Planning for Data Management in Forestry," *IEEE Expert Systems*, vol. 10, pp. 35-41, 1995.

[2] D. G. Goodenough, D. Charlebois, and S. Matwin, "Automating Reuse of Software for Expert Systems Analysis of Remote Sensing Data," *IEEE Trans on GRS*, vol. 32, pp. 525-533, 1994.

[3] D. G. Goodenough, P. Bhogal, D. Charlebois, S. Matwin, and O. Niemann, "Intelligent Data Fusion for Environmental Monitoring," presented at IGARSS'95, Florence Italy, 1995, pp. 2157-2160.

[4] D. Charlebois, D. G. Goodenough, S. Matwin, A. S. P. Bhogal, and H. Barclay, "Planning and Learning in a Natural Resource Information System," presented at Canadian AI, Toronto ON Canada, 1996, pp. (in press).

[5] T. Mitchell, S. Mahadevan, and L. Steinberg, "LEAP: A learning apprentice for VLSI design," presented at Ninth IJCAI, LA, 1985, pp. 573-580.

[6] T. Mitchell, R. Caruana, D. Freitag, J. McDermott, and D. Zabowski, "Experience with a learning personal assistant," *Communications of the ACM*, vol. 37, pp. 80-91, 1994.

[7] O. Etzioni and D. Weld, "A Softbot-Based Interface to the Internet," *Communications of the ACM*, vol. 37, pp. 72-79, 1994.

[8] J. G. Carbonell, "Derivational analogy: A theory of reconstructive problem solving and expertise acquisition," in *Machine Learning: An Artificial Intelligence Approach*, vol. II, R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, Eds.: Morgan Kaufman, 1986.

[9] K. J. Hammond, *Case-Based Planning: Viewing Planning as a Memory Task*. Boston MA: Academic Press, 1989.

[10] M. Veloso, "Learning by Analogical Reasoning in General Problem Solving," : Carnegie Mellon University, 1992.

[11] R. E. Fikes, P. E. Hart, and N. J. Nilsson, "STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving," *Artificial Intelligence*, vol. 2, pp. 189-208, 1971.