INTELLIGENT DATA FUSION FOR ENVIRONMENTAL MONITORING

David G. Goodenough, Pal Bhogal, Daniel Charlebois, Stan Matwin[1] and Olaf Niemann[2]

Pacific Forestry Centre, Natural Resources Canada, Victoria, Canada
[1]Department of Computer Science, University of Ottawa, Ottawa, Canada
[2]Department of Geography, University of Victoria, Victoria, Canada

{dgoodenough, dcharlebois, pbhogal}@a1.pfc.forestry.ca
stan@csi.uottawa.ca

Abstract - In order to monitor the land environment, it is necessary to compare past and present states and to fuse remote sensing data from multiple sources. The past states of the vegetation for test sites in British Columbia are represented in GIS files and in past remote sensing imagery. The present state is shown in current airborne and satellite imagery. The data fusion involves radiometric calibration, geocoding, and topographic relief correction of the remote sensing data, integration of these data with the multiple GIS files, and analysis with the fused data. These operations and the subsequent analysis require complex software components. These software components require advanced expertise to correctly perform the data fusion operations. A methodology for creating a sequence of intelligent expert systems has been devised and applied to the problem of data fusion. An example will be given of the application of this intelligent system.

I. INTRODUCTION

In recent years, vast data repositories useful for understanding the current state of the biosphere and the trends in vegetation over large areas are becoming available. Use and analysis of such data can lead to better, sustainable utilization of resources. The largest vegetative component on the land surface of the earth is forestry. Natural coniferous forests provide 73% of the global industrial log supply. Global demand for wood fiber will require that an additional 77 million m^3 of timber be harvested annually. The increased demand can not be satisfied, and a shortage of industrial wood is expected in the future. Plantation forests supply less than 10% of the world's industrial wood. [1]. Consequently, in British Columbia old growth forests will continue to be the principal source of timber for three or more decades.

More than 40% of Canada's marketable timber can be found in British Columbia, often on rugged, mountainous terrain. In 1992, over 60% of the coastal lumber production was exported, accounting for 34% of the world's exports of coniferous lumber. Forest exploitation is and will remain an important economic sector in British Columbia. In 1991, B.C. forest industry had 8.6% of B.C. jobs and provided 11% of B.C.'s GDP.

The problem with the access, use, and analysis of the forest data is its diversity and complexity. This diversity is reflected in a variety of formats, media, and granularity of data. The problem is compounded by the complexity and heterogeneity of the computing environments in which the data resides, as well as the multitude of software tools which are needed to extract the required information from the data. In forestry, for example, the advanced information systems integrate forest cover descriptions, topographic maps, remote sensing, and application knowledge. Remote sensing data is essential for monitoring large areas. It is anticipated that the rate of data acquisition for the nation will reach one terrabyte per day by the year 2000. Data management systems for resources will need to make intelligent selections of the fused data in order to respond to users' goals, reduce complexity, and be more adaptable [2]. This paper describes the fusion of multidate LANDSAT imagery with GIS files for the estimation of rates of change in harvesting in Tofino Creek, a watershed in the Clayoquot Sound area on the west coast of Vancouver Island.

II. AN INTELLIGENT DATA FUSION SYSTEM

The SEIDAM (System of Experts for Intelligent Data Management) Project is conducted under the Applied Information Systems Research Program of NASA. SEIDAM is a complex system for forest and environmental monitoring that relies on extensive cooperation between expert systems and processing agents [3], [2].

The SEIDAM system will provide answers and products in response to user queries which trigger the fusing of those data necessary to answer the query. Once a solution to a problem has been formed, SEIDAM will use the processing agents at its disposal to execute each step in the solution. It is expected that, for each step, the agents will report on the state of the processing, thus allowing SEIDAM to learn about its successes and failures in order to refine its case bases and its knowledge of the processing agents. The processing agents will typically be third party software packages providing database management facilities (Ingres), geographical information systems (GIS; ESRI ARC/Ingres, GRASS, AVS Pamap GIS), image analysis programs (PCI, ENVI, LDIAS), visualization tools (AVS), etc.

SEIDAM incorporates a fiber optic communication link with ATM protocol allowing distributed data management and application processing. Two terrabytes of land data have been collected and are stored in an EXB-120 robotic data store. The hierarchical storage management system (Open Vision) is connected to a SEIDAM meta data database for images and GIS files. The workstations are linked by Ethernet and ATM. The expert systems for

0-7803-2567-2/95 $4.00 © 1995 IEEE

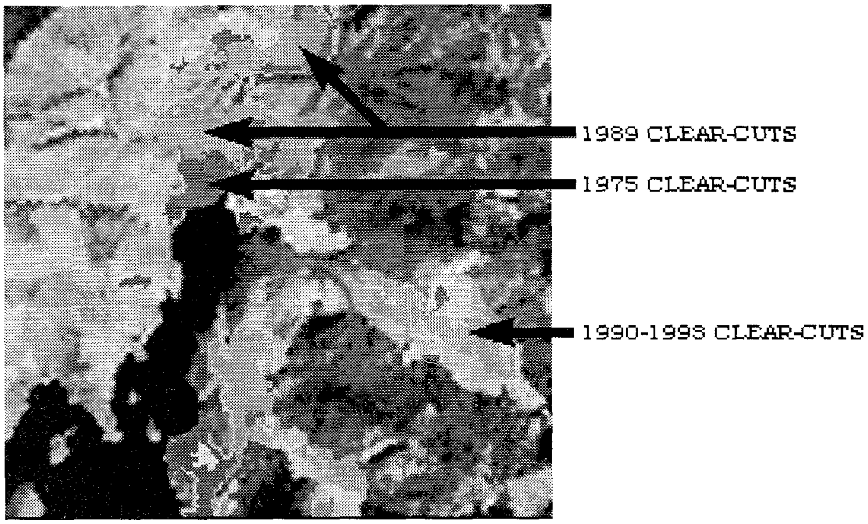

Figure 1: Change detection of forest cover in Tofino Creek showing themes for clear-cuts from 1975 to 1993.

SEIDAM execute on SUN workstations. The user interacts through TCL/TK and Motif windows. User prompts from application software are substantially reduced by the availability of an extensive knowledge base, stored expertise, and cases (examples) provided by domain experts.

III - FUSING REMOTELY SENSED AND GIS DATA

The process of fusing remote sensing data into GIS includes the following procedures: data acquisition, data correction, geocoding or georeferencing to a common map base, detection and identification of desired objects, analysis, visualization, error analysis and the creation of the final products.

In a GIS it is possible to overlap many descriptive layers, such as forest cover, soils, hydrology, etc [4]. In an image analysis system, one can also bring together in raster form many sources of data, such as LANDSAT, SAR, AVIRIS, etc. These combinations permit the application of a large number of data analysis techniques. The fusing of GIS and remotely sensed data in SEIDAM enable the following data sets to be used together:

- multi-temporal remote sensing imagery;
- multi-resolution imagery from several remote sensing instruments and platforms;
- digital ungeneralized topographic data;
- digitized maps of various land attributes on NAD'27 or NAD'83 datum such as forest cover, soil types, forest practices decision boundaries, hydrography, etc.,
- socio-economic data,
- point measurements of atmospheric and surface attributes, such as radio sonde data, soil moisture, etc.

As an example, consider a resource manager who wishes to know how the forest cover has changed in Tofino Creek over the last 40 years.

Topographic GIS data for Tofino Creek exists as ungeneralized, digital elevation points. In addition, we also have GIS data for forest cover, hydrology, soils, geographic names, etc. The topographic data have been transformed from elevation points to digital terrain models for integration with remotely sensed data. The site is approximately 15 kilometers by 18 kilometers in dimension.

LANDSAT MSS imagery from 1975, and LANDSAT TM imagery from 1989, 1990 and 1993 are available for measuring the change in forest cover over time. For each data set, the implications of topographic relief displacement were studied, and all data sets were registered to a common geographical coordinate system and spatial scales. In the TM data, channel 6 (thermal band) was omitted. A DTM was fused with the MSS and TM data to minimize misclassification resulting from slope and aspect variations. Clustering (ISODATA, [5]) was carried out on each data set, and objects corresponding to clear-cuts from different periods, as well as regeneration areas, were created. Finally, the clustering reports were used to quantitatively determine the areal extent of clear-cuts from different times. This approach avoided the atmospheric variations from scene to scene.

Figure 1 shows the change in forest cover between 1975 and 1993 in Tofino Creek. Also available is the forest cover area for 1940. In Figure 1, the clear-cuts from 1975 are visible as darker gray areas. More recent clear-cuts appear as lighter shades of gray.

IV. A PLANNING-BASED APPROACH TO DATA FUSION

Figure 2 represents the top level architecture of the AI portion of our system. First, a user query is translated into a conjunction of planner goals. The Planner/Controller system is the core of our approach. It relies on four different knowledge sources to perform its task.

The first source, the meta-knowledge, describes the information the system has: what data sets, in what formats, with what accuracies, and at what cost, are available to answer queries. If the meta-knowledge indicates that the answer to the query is available, the Planner creates a plan that tells the agent where to retrieve the data, and where to deliver it.

The second knowledge source is a case-base of plans whose execution by the agent provides a result that may partially satisfy the query. This case-based reasoning component uses a transformational analogy approach much like that used in CHEF [6]. The retrieval mechanism selects classes of plans that are similar up to the objects which are the arguments of the operators. Modifications of the objects manipulated by the processing agents produce an adjusted plan that more closely satisfies the query.

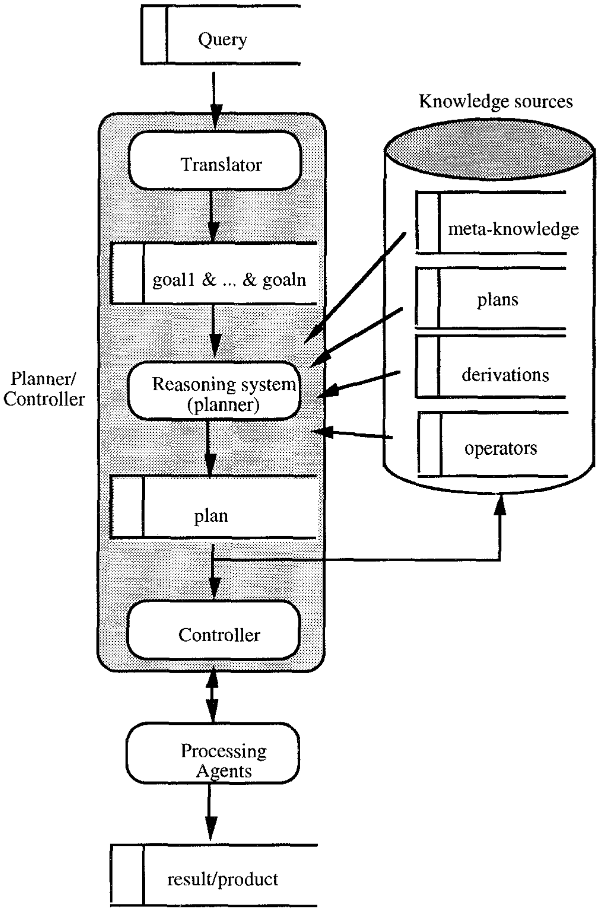

Figure 2. Architecture of PALERMO.

The third knowledge source is a case-base of derivations. When transformational analogy is unable to create a plan that satisfies all the goals, we must resort to some search-based technique to fill this gap. This activity involves creating a plan by choosing adequate operators. When no additional knowledge is available to help in this choice, operator selection decisions may be subject to expensive backtracking. Derivations suggest what operator ought to be chosen at a given step in the planning process, if a similar decision has in the past resulted in the derivation of a successful plan [7]. This approach has been studied by Veloso during the development of the NoLimit [8] system.

Definitions of planning operators provide the fourth knowledge source for the planner. After the plan has been constructed by a combination of transformational analogy, derivational analogy, and search, the resulting plan is executed by Processing Agents. Processing agents are *Software agents* (also known as *apprentices* [9, 10] and *softbots* [11]) 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. Each agent is composed of four different components: a rule-based expert system, a frame-based knowledge-base, a suite of third party software, and a set of command operators that are used to send instructions to the third party software.

The planning system is responsible for assembling plans to answer queries and to create products which can be the result of assigning a sequence of tasks to the processing agents over a given data set and knowledge base.

V. DATA FUSION DESCRIPTION

In this section, we will describe answering a query. The query presented to the system is: "How has the forest cover changed in Tofino Creek over the last 40 years."

The system will first try to find a similar case that satisfies the query. There are several approaches to establishing similarity. The approach taken in here is to declare two goals similar if they differ only in subterms of their expressions (e.g.: *map_sheet(092b042, location, 'Greater Victoria Watershed')* and *map_sheet(092F023, location, 'Tofino Creek')* are similar whereas *map_sheet(092b042, location, 'Greater Victoria Watershed')* and *tm_image(PG01512-1.bil_image, location, 'Greater Victoria Watershed')* are not). If it cannot find a similar case, it will try goal-regression.

Let us assume that a case exists in the knowledge base that can satisfy goals similar to those expressed in the query. The case would be retrieved by the reasoning system and adapted to provide for the differences, thus creating a new plan. More details about how this is achieved can be found in [12] and [2]. A set of processing agents has been trained [13] to perform the processing described in the case. For the query above, these steps are:

(1) find the appropriate TM data, remove channel 6, the thermal channel, and load the data onto disk,
(2) find the appropriate GIS files and load the data onto disk,
(3) geocode all remotely sensed data to a common geographic reference system,
(4) perform topographic correction on all multi-temporal TM and MSS data to be used in the analysis,
(5) fuse GIS data and the remotely sensed imagery,
(6) cluster each data set,
(7) create themes corresponding to clear-cuts from

various time frames,

(8) generate reports from the clustering algorithms, to determine the areas of clear-cuts, and plot them as a function of time.

After the successful execution of the agents, the image shown in Figure 1 is created as well as Figure 3 showing the forest cover change in Tofino Creek.

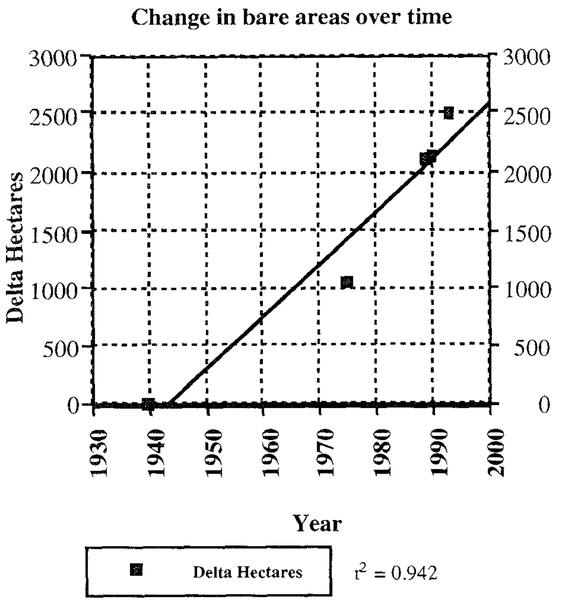

Figure 3. Forest cover change from 1940 to 1993 as a result of data fusion.

The result in Figure 3 shows that the rate of harvesting in Tofino Creek has accelerated dramatically over time with over 2500 hectares cut by 1993.

VI. CONCLUSION

In this paper we have argued that the problem of environmental monitoring involves many diverse sources of GIS and remote sensing data. An intelligent system for fusing these data sources has been described. An example of responding to a query about the changes in forest cover in an environmentally sensitive area has been presented. This system (SEIDAM) has been implemented using a robotic data store, a meta data database, and an ATM network. Future work includes the data fusion of AIRSAR and AVIRIS data for more sophisticated queries related to distributed timber volume (biomass) and canopy chemistry.

VII. ACKNOWLEDGMENTS

The SEIDAM Project is supported by NASA's Applied Information Systems Research Program, Natural Resources Canada, Industry Canada, the B.C. Ministry of Forests, and the B.C. Ministry of Environment, Lands and Parks. We also acknowledge research support from the Natural Sciences and Engineering Research Council of Canada.

VIII- REFERENCES

[1] D. Charlebois, D. G. Goodenough, and S. Matwin, "Machine Learning from Remote Sensing Analysis," presented at IGARSS'93, Tokyo, Japan, 1993.

[2] S. Matwin, D. Charlebois, D. G. Goodenough, and P. Bhogal, "Machine Learning and Planning for Data Management in Forestry," *IEEE Expert Systems*, vol. (in press), 1995.

[3] D. G. Goodenough, D. Charlebois, S. Matwin, D. MacDonald, and A. J. Thomson, "Queries and Their Application to Reasoning with Remote Sensing and GIS," presented at IGARSS'94, Pasadena CA, 1994.

[4] M. Ehlers, D. Greenlee, T. Smith, and J. Star, "Integration of Remote Sensing and GIS: Data and Data Access," *PE&RS*, vol. 57, pp. 669-675, 1991.

[5] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*: Addison-Wesley Publishing Company, 1974.

[6] E. D. Sacerdoti, "The Non-Linear Nature of Plans," presented at IJCAI'75, 1975.

[7] R. C. Schank, *Dynamic Memory: A Theory of Learning in Computers and People.* Cambridge MA: Cambridge University Press, 1982.

[8] G. J. Sussman, "A Computational Model of Skill Acquisition," : MIT, 1973.

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

[10] 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.

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

[12] D. Charlebois, D. G. Goodenough, and S. Matwin, "Case-Based Reasoning in an Intelligent Information System for Forestry," presented at SPIE'94, Orlando, FL, 1994.

[13] S. Matwin, D. Charlebois, and D. G. Goodenough, "Training Agents in a Complex Environment," presented at IEEE: Conference on Artificial Intelligence Applications, Los Angeles, California, 1995.