SAHARA
Semi-Automatic Help for Aerial Region Analysis

P. Druyts[1], W. Mees[1], D. Borghys[1], C. Perneel[1], M. Acheroy[1], J. L. Valero[2]

[1]Signal and Image Centre, Royal Military Academy, Belgium
http://www.sic.rma.ac.be
Av. de la Renaissance, 30
1000 Bruxelles, Belgium
Pascal.Druyts@elec.rma.ac.be, Wim.Mees@asge.rma.ac.be, Dirk.Borghys@elec.rma.ac.be

[2]Western European Union Satellite Centre
Apartado de Correos, 511
Torrejon de Ardoz, 28850 Madrid, Spain

Abstract - **Sahara is a WEU funded project. Its aim is to show the potential usefulness of image interpretation techniques to help photo-interpreters. Typically, this could include drawing a sketch of the airport, looking for changes with respect to a reference geographic database, etc. The need for tools that fasten satellite imagery interpretation will become even more stringent with the next generation of satellites that will provide a very high amount of high resolution data. We therefore used aerial multi-sensor imagery with resolution expected in the near future for commercial satellite imagery (VIS 2m, SAR 5m, etc.).**

As a test case, the project was dedicated to the interpretation of airports. In this scope the main objects of interest like runways, taxiways, shelters, airplanes and buildings are searched for.

In this paper, we show that object-related detectors are not able to solve the problem by their own. We then propose a scheme in which high level knowledge may be introduced in an efficient way. This knowledge is used to steer the interpretation and to enforce a coherent result. A GUI allows the user to follow the evolution of the interpretation and to have a clear view of the high level knowledge that is used. The user may also intervene at any moment to help the interpretation, by correcting errors, stopping the research in a non promising areas, etc.

Key words - **3D scene analysis, artificial intelligence, expert system, blackboard system, fuzzy production rules, road detection, building detection, shelter detection, airport interpretation.**

I. Introduction

For decades photo-interpreters have been looking at photographs, using a magnifying glass to observe smaller details and using some computation aid to perform elementary conversions from distances measured on the photograph to approximate real world distances, knowing an average scale for the photograph. Nowadays they use a computer for visualising scanned images, performing resolution changes and some simple computing but apart from that not a lot has changed. Instead of drawing on the paper copies of the photographs and preparing a report on an ancient typewriter, a photo-interpreter now vectorizes on-screen and uses some fancy word processor but he is essentially still performing the same basic operations.

Most commercial software packages which combine aerial image manipulation with the management of vector-style geographical information still require the user to perform the passage from raster to vector information manually. Some have limited automatic line-following capabilities but these are most often not useful in an operational environment. What these software packages excel in however is their huge libraries of image processing techniques, filtering the image in order to obtain a smoother, noise-filtered image, to make the image sharper and eliminate some unwanted blurring effect, to detect edges, etc.

Unfortunately, experience shows that in practice most photo-interpreters don't use these features. They claim that the human visual system is the best possible adaptive filter and that whichever type of filtering can never result in an image with a higher information content than the original image. Therefore they prefer to perform their interpretation on the original, unfiltered image.

When one looks at the evolution in the field of earth observation data acquisition systems, it becomes obvious that the flux of aerial and satellite imagery will grow exponentially, whereas the number of trained photo-interpreters will not. Therefore the real challenge for computer system developers in the field of scene analysis is to build a semi-automatic system which relieves the human photo-interpreter from the routine part of his work, so that he can focus on those parts of the job for which his human intelligence and intuition are indispensable. In this paper we will present a possible layout for such a system.

To interpret a scene, a number of objects of interest should be searched. Some detectors devoted to objects of interest in the scope of runway interpretation are briefly presented in section III. We will show in section II that these detectors are highly inefficient if used alone because it would then be necessary to search for everything everywhere using only local information. Therefore, in section IV, we will discuss additional knowledge that should be incorporated into the system. A system design in which this additional knowledge may be incorporated will be presented in section V. A prototype of this system has been developed and will be described in section VI. Finally, in section VII we will present our conclusions.

II. Searching for objects of interest

For each scene interpretation problem, a number of objects of interest may be defined. These objects may be those that are of direct interest for the user but also some other objects that once found may help the system to detect the first category. For the airport interpretation problem, the objects of interest could be runways, taxiways, shelters, airplanes, buildings, etc. Even if the user is only interested in the detection of shelters, runways and taxiways will remain objects of interest for the system because shelters are difficult to detect and it is necessary to use the taxiway position to reach an efficient detection.

For each type of object of interest, a 'local detector' is developed. These detectors are based on a model of the searched object but do not take into account constraints that may exist between different objects (a building may not stand in the middle of the runway, etc.).

These detectors are clearly unable to solve the interpretation problem by their own. A brute force approach in which all detectors are launched together and in which the objects detected by all those detectors are added in the world model and used as final interpretation would be a highly inefficient approach for several reasons amongst which :

- A scene is highly structured and it is a waste of time and of computational power to search for everything everywhere. A strategy should be used.
- Some detectors use a rough model of the object. If many parts of the image that do not include the object are compatible with that rough model, a lot of false alarms will be found; the detector is quick and dirty. However, within reasonable limits of false alarm rate, those detectors may be quite efficient because it is possible to reject false alarms by using a more precise model of the object to check all the candidates. There is an important gain in time because the complete model has only to be applied to the object candidates. Furthermore, the detailed knowledge about an object is often quantitative and difficult to embody in a detector. We will use a fuzzy logic approach to model that knowledge.
- Some prior knowledge may help the local detectors. This knowledge may be known from the beginning on or become available at a certain stage of the interpretation (e.g. the type of airport may be found and more information about the expected type of shelters may become available). This knowledge may also be in a form that may not be directly used by the detector. As an example the expected size of the object may be useful for the detector and may be known a priori in meters whereas the detector can use it if expressed in pixels. In our approach, parameters are used for each information (in a form directly usable by the detector) that could help the detector. It is the task of the high level system to pass this information to the detector whenever it becomes available eventually after some conversion (meters to pixels, ...).
- Using only local information, it may be quite difficult to decide whether the detected structure is really the object of interest. A lot of false alarms will appear and could be rejected by global constraints encoding the confidence in a given arrangement of objects.
- When choosing parameters for a detector, one is often faced with trade-offs. Some parameter may lead to a high level of detection, but could imply a huge consumption of computation time and memory. In the extreme case, the detector could not even run within the limited resources available on the machine. It is then more efficient to search first for the most salient objects by using restrictive parameters. This implies a limited consumption of computer resources and leads to few false alarms. High level knowledge may then be used to steer the interpretation and a more detailed search may be performed in regions of interest.
- It is difficult to manage a pool of detectors; a lot of parameters are hidden within those detectors. They may either be hard-coded or passed as parameters. It is quite difficult to have a global view of all these parameters and to select good values for them, a trial and error procedure is often used. In our approach, all those parameters will be centralised in the local detector manager. This allows for a global view of all the parameters and their value may be computed by some rules introduced by an expert. These rules may use any information about the scene (contrast, etc.). It is also planed to add an evaluation module that could provide a feedback for an automatic tuning of these parameters by learning. This evaluation could also be used to rate the detectors and to call only those that are appropriate for the problem at hand. Learning has to be considered with caution to avoid the black box effect. The expert must keep the control on the system but nevertheless, some limited learning within predefined bounds or subject to confirmation by the expert could help and ease the management of a big system.

III. Local detectors used for airport interpretation

In this section, we will briefly describe some detectors that have been implemented for runway, taxiway and building detection. Detectors dedicated to other types of objects have also been build but will not be presented due to space limitation.

A. Runway detection

Two runway detectors have been developed; one for medium (5-10) resolution and one for low resolution (10-20m). At medium resolution, the two borders of the runway are well detected by means of an edge extraction algorithm. Those borders are searched and approximated by lines. A histogram of directions is then computed. The runways borders should be amongst the longest lines in the image and thus lead to local maxima in the histogram. Having found the direction of the runway, the lines in that direction are extracted and grouped using a proximity criterion. The bounding boxes of the groups are then computed and those with dimensions compatible with a runway are kept as runway candidates.

On low resolutions images, it becomes difficult to extract the borders of the runway by means of a contour extraction algorithm. A bar detector is more appropriate but it does not give a precise measure of the width of the road. A tool has been developed to extract the borders corresponding to a bar; this provides a precise measure of the width of the bar. The bars with compatible width and in good prolongation are then grouped. The width and length of the groups are used to keep only those with appropriate length and width.

B. Taxiway detection

Two methods (contour and region based) have been developed for taxiway detection. The first is based on contour extraction and searches for structures composed of groups of parallel lines in good prolongation and sufficiently straight. This approach gave good results; the model is quite generic but nevertheless quite discriminative because only few false alarms are detected.

In some cases however, the model is not valid anymore; this is the case when the borders of the taxiway are highly deformed by parking aprons in front of shelters for example. It appeared that in some of those pathological cases, a region based approach may solve the problem. In such an approach, homogeneous regions are searched and processed to find the middle and the width of the candidate taxiway. If the width and shape of the candidate taxiway are appropriate, that candidate is kept. It has to be noted that to detect the middle of the taxiway, a skeleton based method is not appropriate because it does not work when the borders are deformed. A specific tool has thus been developed for this purpose.

Using the two methods just described and the available spectral bands, a fusion of the detected taxiways is performed. This leads to a very efficient detection. Indeed, a taxiway will only be missed if it is neither homogeneous in any spectral band nor well described by the parallel border model.

C. Building detection

Most buildings are closed shapes composed of straight lines and near 90 degree corners. The detector uses the extracted contours and tries to connect them by filling gaps to get closed contours composed of straight lines and near 90 degree corners. Many such closed paths may exist in an image. To avoid a combinatorial explosion in the number of explored paths, the search is limited by starting the search only near detected shadows (in the visible image) or near bright regions on the SAR image.

IV. Adding scene analysis knowledge

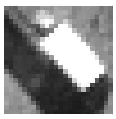
Scene analysis knowledge is *qualitative knowledge*. Reichgelt [1] describes qualitative knowledge as any kind of knowledge that doesn't always allow a correct and consistent match between the represented objects and the real world but can nevertheless be used to get an approximate characterisation of the behaviour of the modelled domain.

A complex, real-world problem such as scene analysis cannot be formalised in a nice and neat way. Most often not all the information needed is available. The available information will furthermore not be 100% correct and consistent and may thus prove and disprove a fact with the same theory. Therefore inference techniques have to be used which deal with knowledge that may be in part incorrect, incomplete and inconsistent.

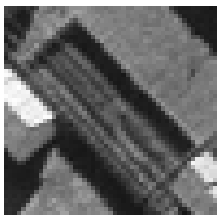
We propose to classify the scene analysis knowledge according to following scheme [2][3]:

- scene description
  - ◊ *geographical database*: when a geographical database of a sufficiently large scale (e.g. 1/10.000), so as not to undergo displacements due to generalisation, is available, it will be consulted by the scene analysis system in order to collect a priori knowledge.
  - ◊ input by a human operator: a human operator could prepare a scene by marking already certain key elements in the scene serving as a priori knowledge for the automatic system. He may also follow the evolution of the semi-automatic interpretation process using the graphical user interface and intervene by adding or removing objects in order to steer the system or correct the intermediate partial solution.
- scene independent knowledge
  - ◊ *interpretation strategy*: it is well known that our eyes don't move in a smooth and continuous manner when viewing an image. They go briefly over numerous fixation points, separated by jumps, and concentrate on those features conveying salient information [4]. The human mind doesn't scan from left to right and top to bottom like most image processing algorithms do. A trained image interpreter will steer his focus of attention based on hypotheses generated by previously interpreted objects as well as on a set of standing operating procedures. For instance in a suburban area first look for roads in a down-sampled copy of the image and after that look for buildings alongside the roads in the full-resolution image. For the image in Figure 1 the attention will first be focused on the houses and the road and only thereafter on smaller details such as the driveways, cars, swimming pools, and so on.
  - ◊ *generic constraints*: this knowledge implements general physics laws (e.g. relationships between a building and its shadow), administrative regulations regarding land-use as well as the experience a human operator has acquired after years of practice, expressed as series of rules of thumb. It can best be represented by a set of global rules acting on clusters of objects, of the same or of different types. The rules will judge the geometric relationships between the objects within a cluster [5]. In a multi-sensor system this knowledge will be sensor-independent.
  - ◊ *object-type specific knowledge*: this is the knowledge which allows one to distinguish a certain type of object from all other types. It is most often expressed as a list of conditions on image features such as contours, texture and the gray-value histogram (e.g. a building is rectangular or L-shaped with certain limits on its dimensions). This knowledge will in general be sensor-dependent.

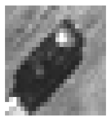
Figure 1: aerial photo



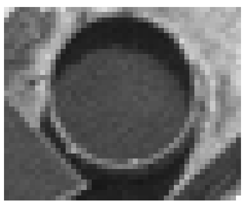
(a)



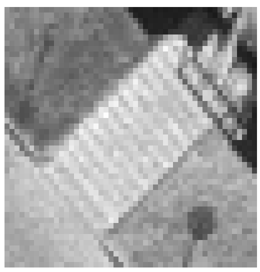
(b)



(c)



(d)



(e)

Figure 2: local versus global knowledge

The interaction between the local and global part of the scene analysis knowledge is very important since neither of both can do the job alone. This is illustrated in Figure 2 where five parts of an image are shown which were cut out of an aerial image (Figure 1). Given only the local information of the objects themselves and a very limited part of their immediate surroundings, it is clearly difficult to identify them. This might be our first guess :

(a) truck with tractor and trailer
(b) factory building
(c) car
(d) cooling tower of power plant or a silo
(e) flight of stairs

However, when we look at these objects in the original image, from their position relative to the houses and roads we learn that we had it completely wrong. These are the correct answers:

(a) small building in backyard
(b) low greenhouse
(c) car
(d) swimming pool
(e) driveway

This goes to show that the relative position with respect to other objects is indispensable when trying to recognize an object in an image.

V. General system lay-out

As explained in section II, it would be highly inefficient to search for everything everywhere. We will therefore base our problem solving strategy on the technique of generating and fusing uncertain and partial solutions to construct solutions, using an *island-driven* approach. A relatively reliable partial hypothesis is designated as an island of certainty and the hypothesis building process pushes out from this island in a number of directions into the ocean of uncertainty surrounding it.

Local detectors are then called in an ordered and efficient manner according to the strategy and the global rules. Most salient and easy to extract objects are searched first, giving clues to where other objects should be searched. The confidence in the detected objects permanently evolves according to the local and global rules allowing false alarms to be rejected when the confidence becomes too small.

This leads to an efficient combination of bottom-up and top-down approaches. Starting from a bottom-up approach to search for the most salient object, all gathered knowledge is used to determine what to search and where to search it. When this knowledge becomes more precise, the search method tends to a top-down approach.

Fitting all the above-mentioned types of knowledge in a single knowledge representation scheme would involve compromising and would thus result in a sub-optimal solution. The general strategy can be well represented using a *goal-reduction* scheme [6]. The generic constraints will be expressed by human photo-interpretation experts. They will use natural language rules, based on vague terms. This set of rules will be imposed using a *fuzzy production rule system*. The knowledge related to each object type is necessary when evaluating or extracting a single object without taking into account other objects in the scene. This knowledge varies strongly from one type of object to another. It will be integrated in the implementation of local detectors, each one dedicated to a specific type of object. Different sensors will require different local detectors for specific object-types. Note that the knowledge encoded in the detector is usually based on a rough model and more precise constraints on the object may be expressed by human photo-interpretation experts. Those constraint will be encoded in a similar manner as the generic constraint and allow for the rejection of a number of false alarms.

The *blackboard* problem-solving model is particularly well suited for this type of complex problems because it supports the incremental development of solutions, can apply different types of knowledge and can adapt its strategy to a particular problem situation [7] [8]. It allows the use of independent *knowledge sources* in order to represent the different types of knowledge as shown in Figure 3.

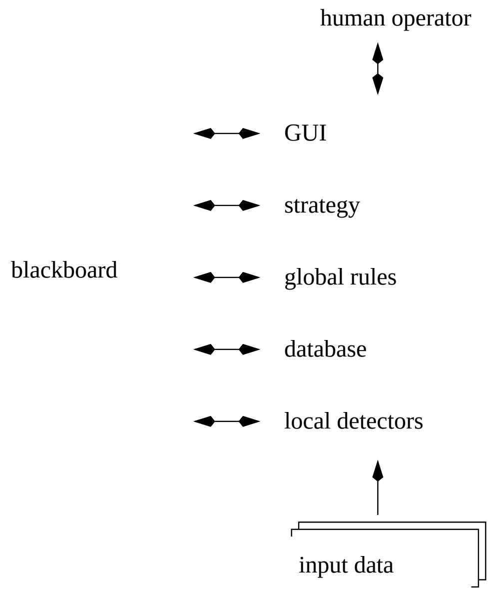


Figure 3: system design

VI. Prototype definition

At present a prototype has been developed which implements the system described in section V. The central blackboard, the different knowledge sources and the local detectors are all built as independent executables. The blackboard and the different knowledge sources communicate using TCP/IP sockets as inter-process communication technique. The local detectors are actually scripts which call a series of executables with specific parameters and in a well-defined order. They are launched by the local detector manager knowledge source.

A. Results & Future work

Our preliminary tests have shown that the system functions well for the typical configurations of input data which were considered when designing the local detectors and when writing the global rules and the strategy. Until now, no complete evaluation has been performed but a visual evaluation lets us say that the runway and taxiway structure as well as the shelters are quite well extracted.

It is quite clear however that in order to test the validity of the global rules a much larger amount of test images is needed. This will also show the shortcomings of the already developed local detectors in certain specific circumstances and thus lead to the development of some extra, complementary local detectors.

The system has a distinct advantage over other systems which only incorporate the typical local information handling routines because it can itself operate these local routines in an ordered and efficient manner, with varying parameters, allowing for a certain number of false candidates since it will thereafter weed out most of the false candidates using the knowledge contained in the global rules knowledge source. The high-level, expert system part here plays the role of a filter, eliminating the ‘‘noise peaks” corresponding to false candidates. Simulations with noisy detectors have shown that this will only function up to a certain level of false candidates since then the system will start discovering fake structures in the set of false candidates, which will reinforce its confidence in these false candidates and may even suppress correct candidates in the same region. This shows that the usage of high level knowledge is quite efficient to obtain a reliable interpretation but does not relieve the need for good local detectors.

In all, we can say that the way in which the knowledge is structured in the system is the major advantage of the approach we've presented. The user has a good grip on the knowledge the system applies thanks to the clear logic as to which knowledge goes where. This is enforced by the graphical interface that is currently being developed. This GUI will allow the user to analyse all the knowledge sources in an efficient and clear way. Amongst other, the user will be able to follow the evolution of the state of the system in the strategy, to intervene on it, to analyse which rule rejected a true candidate and eventually to modify the rule.

The way in which the objects, representing the solution, are combined with the rules, raising or lowering their confidence, in a single network in the world-model has shown to be very useful since it allows the user to examine immediately what the influence of every single one of the global rules is on each of the objects to which it is applied.

What we have ascertained however is the fact that some ideas we had were hard to implement with the structure as it is at present. We will therefore add in the next version of the system some extra knowledge sources, one of which will realise perceptual grouping at the level of the objects in order to detect different types of structures of objects, such as buildings lying on a parabola, cars equally spaced in a parking lot, etc. The usefulness of an evaluation module that would allow for some learning as explained in II was also felt.

VII. Conclusion

Due to the ever increasing amount of available earth observation information, the need for semi-automatic systems, which aid the human expert in his analysis, will continue to grow. Aside from the well-known point and raster data manipulation techniques with which actual systems are already equipped, these semi-automatic aids will furthermore allow a user to integrate a part of his knowledge into the system in the form of a general strategy, global rules and local descriptions.

These systems will then be able to relieve the human expert from the routine part of his work and allow him to focus on ‘‘special cases” or on the interpretation of *why* certain features occur at certain positions without first having to extract them manually from the input data.

It is obvious that an experienced human data analyst will always outperform any artificial system when it comes down to the completeness of the analysis or the handling of exceptional cases. A human operator on the other hand has the disadvantages of a higher operating cost and a dislike for routine duties. Therefore if we combine both systems, we will obtain an increase in productivity with the same quality as when compared to the human expert alone. The SAHARA project aims at demonstrating this assertion by means of a prototype.

REFERENCES

[1] Han Reichgelt, “Knowledge Representation: An AI Perspective”, Ablex Publishing Corporation, 1990.

[2] Wim Mees, “A general framework for automatic aerial photo interpretation”, Third Euroconference on Geographic Information Systems, Karlsruhe (Germany), September 1995.

[3] Wim Mees, “Scene analysis: fusing image processing and artificial intelligence”, IEEE-CS looking .forward, Spring 1997.

[4] K.N. Leibovic, “Science of Vision”, Springer-Verlag, New York, 1990.

[5] Wim Mees, “Image interpretation using fuzzy expert systems”, International Conference on Digital Signal Processing (DSP95)}, pages 511-518, Limassol, Cyprus, June 1995.

[6] Patrick-Henry Winston, “Artificial Intelligence”, Addison-Wesley, 2 edition, 1984.

[7] H.-Penny Nii, Blackboard systems part i: The blackboard model of problem solving and the evolution of blackboard architectures”, The AI Magazine, pages 38-53, summer 1986.

[8] H.-Penny Nii, “Blackboard systems part ii: Blackboard application systems, blackboard systems from a knowledge engineering perspective”, The AI Magazine, pages 82-106, summer 1986.