

USING A GUIDED VIRTUAL GAZE FOR THE ANALYSIS OF LOW RESOLUTION REMOTE SENSING IMAGES

Gilles COTTERET, Bernard MOULIN, Geoffrey EDWARDS
Centre of Research in Geomatics, Laval University
Quebec City, Quebec, Canada
Gilles.Cotteret@scg.ulaval.ca

G rard LIGOZAT
LIMSI-CNRS, Paris-Sud University
Orsay, France

ABSTRACT

In this work we propose a novel way to guide the extraction of linear features in remote sensing (RS) images for updating geographic information systems (GIS), using a virtual gaze technique. The proposed model takes advantage of two main ideas. On the one hand, the data we use are not only the RS images but also GIS data that is not necessarily current, and on the other hand, we take advantage of recent progress in the neuro-cognitive sciences by partially mimicking eye movements during visual search.

The model we propose avoids some traps in which classical methods currently fail. In particular, the proposed technique can be applied to low resolution or partially occluded images for which currently only human interpreters can process the image. The technique is also designed to be very fast and efficient when a quick GIS update is needed.

KEY WORDS

remote sensing applications, automated feature extraction, GIS updating, visual attention mechanism

1 Introduction

For a long time feature extraction from low resolution digital imagery has required important manual resources. A lot of efforts has been devoted to reducing the time required for this task via the use of semi-automated or automated extraction techniques. Linear features including hydrological networks [1], [2], [3] and [4] geological lineaments [5] and roads [6], [7], [8], [9], [10] and [11] have been particularly studied. In spite of this work, the extraction of curvilinear elements from remotely sensed images remains a significant challenge for data-processing algorithms. This is particularly the case when the target elements are close to the limit of resolution or when the images are disturbed.

On the one hand, classical methods use multispectral techniques such as supervised and unsupervised classification. On the other hand, the understanding of the human visual system and its modeling has made tremendous progresses since the early 1980's. Following David Marr's pioneering work [12], models of human visual attention have been significantly refined with the help of new computing techniques. Most of the latest models [13], [14], [15] are

strongly based on attentional theories and on research in neuropsychology.

It has been showed that eye movements during a visual search are not random but are guided by particular features of the visual scene [16] and [17]. For the present concern, the most important result is that the natural gaze of human beings appears to *follow* the salient traits within the image.

The visual search model has been considered as a strong bottom-up process for a long time. One of the most popular models has been proposed by Koch and Ullman [18]. Figure 1 shows an overview of this model, which is based on the idea of a saliency map that encodes the stimulus intensity for any position of the display. This saliency map receives inputs from primary visual processes (*pre-attentive computation of visual features*) and permits an efficient control strategy which allows the attention focus to scan the elements contained in the saliency map in a decreasing order of saliency. Later Koch and Ullman's model was modified by Itti and Kock [19] to integrate a top-down component, see bottom-right corner of Figure 1.

Some of the models of visual search are purely top-down [20]) and are mainly used for robot control. These task-dependent models have been inspired by the eye tracking experiments pioneered by Yarbus [16]. Yarbus showed that the recorded pattern of eye movements differs dramatically depending on the task being performed by a subject for a given display.

The extraction of curvilinear elements from remotely sensed images is a difficult problem from an algorithmic point of view; especially, when the targets are close to the limit of resolution or are partially occluded. Considering that:

1. the extraction of curvilinear elements from remotely sensed images remains a crucial issue;
2. the extraction of curvilinear elements is usually based on multispectral techniques;
3. visual search models are numerous and well known; and
4. eye movements are guided by particular features of a scene;

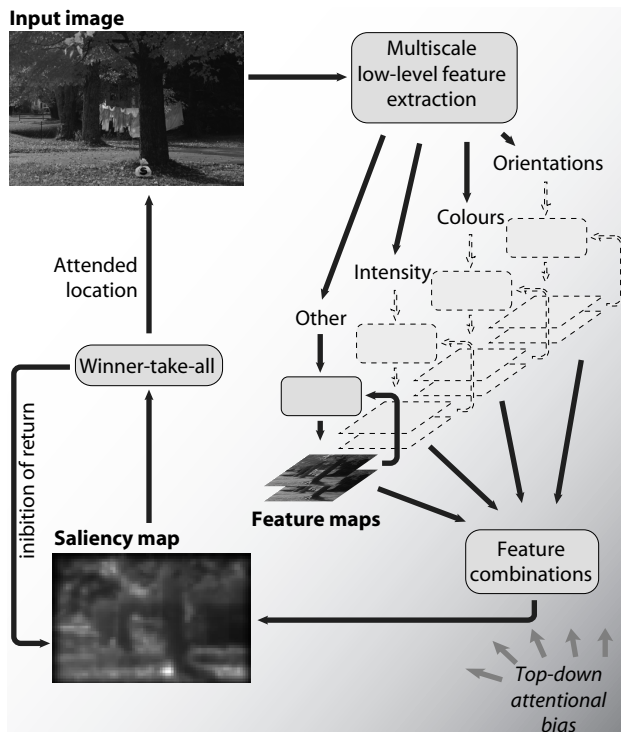


Figure 1. Flow diagram of a typical model for the control of bottom-up attention, simplified from [18] modified by [19] in order to introduce a top-down component.

we propose to use a technique of saccadic sampling inspired by human visual search to guide a virtual gaze for the identification of curvilinear elements in a remotely sensed image. This strategy could dramatically increase the efficiency of automated extraction techniques. It can also be useful to update geographic databases (such as used in GIS) in a fast and automated way.

The aim of our work is to develop a technique to quickly update geographic databases. The test database is used to enhance effective helicopter piloting in rescue operations using a helmet-mounted heads-up display for enhanced/synthetic vision. As a result, it is crucial that the database be as up-to-date as possible. The available data is obtained on the one hand from pre-existing geographic databases and on the other hand from recently acquired satellite images.

2 General Model

Updated and quickly assembled geographic databases are becoming increasingly crucial to location-based services such as effective piloting in rescue operations. As we saw previously, the extraction of curvilinear elements is still a significant challenge for data-processing algorithms even though it is a natural and immediate task for human interpreters. Extracting such elements for experts is easy and

direct.

We rely on the assumption that curvilinear elements (roads, rivers, lines of high voltage, *etc.*) already exist in the available geographic databases. However, they are often outdated representations of current land surfaces since they only reflect information that was exact at the time the map was made. In order to correct the geographic data, we use the outdated data to guide the search in the image. This strategy allows us to focus directly on locations in which the information has a good probability to be found; and we do this by mimicking human attention/focus patterns. It is then possible to accurately process and update the maps with the information newly collected from the image.

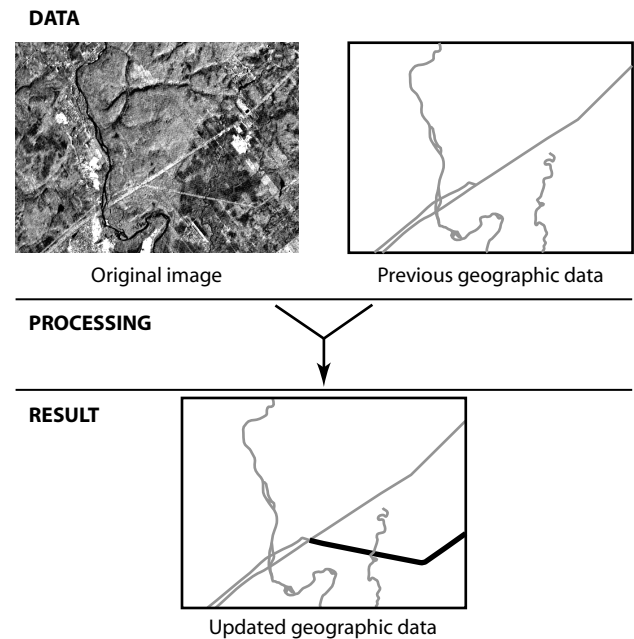


Figure 2. Project aim.

We propose a technique of saccadic sampling to cover the image in order to avoid total processing of the image. Figure 2 shows the inputs and the outputs of the system. The inputs (data) are the RS imagery and the outdated GIS data and the output is the updated GIS data. Hence, the goal of the system is to process this data in order to get an up-to-date database.

Three specific objectives must be realized in order to achieve this goal. These objectives correspond to the three main steps of the general model and they need to be realized one after another:

1. An efficient technique is needed to guide a virtual gaze over the visual scene in order to avoid useless processing of the entire image.
2. Within the local areas pointed by our virtual gaze, a technique is needed to extract the relevant local information, *i.e.* in our case curvilinear elements.

- Finally, a means of gathering all the local information into a global model is needed in order to update the database.

In this paper we focus on the first objective and we show how it is possible to guide the search in the visual zones in which relevant local information is located and avoid zones containing no relevant information.

3 Gaze fixations

The model we developed relies on a mixed top-down and bottom-up approach. The process is task-dependant due to the objects which are sought. However, the process is also bottom-up since our starting point corresponds to the data *i.e.* Geographic Information Systems (GIS) and satellite images.

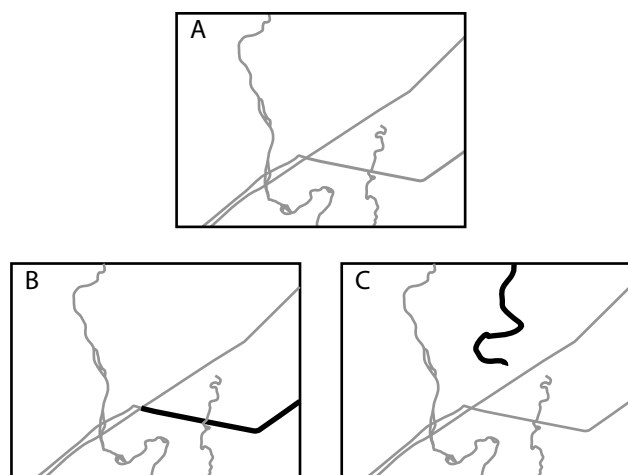


Figure 3. The two modes of the model. **A** Previous database. **B** Recognition mode : the target element exists in the database. **C** Discovery mode : the target element does not exist in the database.

There are two different modes in our approach in order to locate target objects in the image depending on whether the objects were already in the geographic database or not, see Figure 3. The first mode (**B** Figure 3) is called *recognition mode* whereas the second one (**C** Figure 3) is called *discovery mode*.

3.1 Recognition mode

The purpose of the recognition mode is to recognize the elements contained in the database and to validate them (to check if they still exist and are located at the right place, (**B** Figure 3) or to invalidate them (they may exist but not at the right place or they do not exist). In order to cover the image, a *priority list* is used. This priority list works like Koch and Ullman's saliency map but its creation is quite different.

The priority list is elaborated from the previous database in three steps. First, the curvilinear elements are selected in the GIS. Secondly, these curvilinear elements are sorted according to the complexity of each element. This sorting will allow an easier extraction later from the image. And thirdly, for each element a sort is realized in order to rank the different parts of the element according to their complexity. In other words, the priority list is doubly sorted; the first element in the list corresponds to the least complicated part of the least complicated element and so on. An easy way to get a complexity metric is to take advantage of the fact that curvilinear elements are represented in a vector mode in a GIS. In order to eliminate superfluous points along the linear elements considering the granularity of the database, the line is simplified using the Douglas-Peucker algorithm. We can then use a complexity metric corresponding to the linear density of the points representing the vector object.

The priority list, elaborated from the geographic database is a list of points which can be located on the image. The way the priority list is built offers two main advantages. First, it can be used as a saliency map and it is possible to scan it from the top to the bottom in order to focus on the image. But, secondly, it is also goal-oriented (top-down) since it follows elements one by one and only requisite elements.

As we emphasized before, the database, is for the most part, out-dated and inaccurate with respect to the reality shown by the remote sensing image. Depending on the different datum used and the image corrections it is not unusual to have shifts or distortions between the data from the geographic database and the image. Therefore it is important, when the targeted element is not at the expected location, to search the neighborhood of the position given by the database. Consequently we use some conceptual buffer zones, called *search corridors*, around the curvilinear objects in which the targeted objects are expected to be found.

3.2 Discovery mode

The discovery mode takes place after the recognition mode, *i.e.* the image is already segmented by the objects recognized in recognition mode. Hence, it is easy to partition the image into different areas, (Figure 4). In this mode the aim is to discover objects which were not previously in the database; that is elements different from the curvilinear elements recognized during the previous mode. The exploration is therefore located out of the search corridors, *i.e.* in the striped areas of Figure 4.

The way of covering the image in the discovery mode also uses a priority list, but its construction differs from the previous one and depends on the user strategy. The objective is to fill in the list in order to make the exploration as efficient as possible. That is, we want a minimum of points in the list for a maximum set of target objects found. The strategy adopted follows. The exploration strategy is de-



Figure 4. Example of image subdivision after the recognition mode.

pendent on the order of the different categories of points in this list. Hence, we introduce the following types of points in order of ranking:

1. Points corresponding to seeds (belonging either to road, river or lines of high voltage seeds) labeled during the recognition mode. It allows the process to focus directly on objects which are surely present in the image.
2. Points located in the periphery of the image. Starting from the observation that a large majority of the target curvilinear elements are part of a network [2], even for occluded objects, the probability of finding the elements on the image periphery is high.
3. Points chosen randomly in the striped areas. In occluded images or when the object sizes are close to the limit of resolution, some element seeds could be missed from the two previous point sets. Elements which are not part of a network (abandoned forest roads, tunnel exits, etc) must be considered also. Therefore, within the areas previously delimited some random points are selected in order to reach a realization density previously determined by the user. A closely related technique has been used already with Lévy Flights (see [14]).

The virtual gaze scans the list of points which can itself be continuously updated (cf. §4). Once the list is empty the process stops but another round (discovery mode) could be restarted if some new seeds have been found.

4 Image cover

As indicated earlier, the objective of our technique is to be able to focus on the relevant information in a remotely sensed image even if this information is partially occluded, or close to the limit of resolution. Human subjects do not

have any particular problem to follow a road partially occluded in a remotely sensed image. This is possible thanks to the interpretation of intensity and the direction of the object coupled with a saccadic method of covering the image which allows a good integration of the jigsaw pieces.

4.1 Saccadic behavior

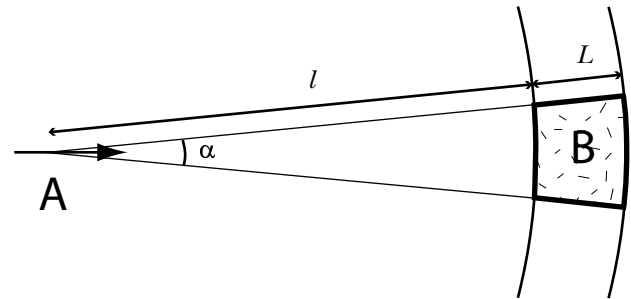


Figure 5. Saccade area. A Starting point, given direction. B Target saccade area.

Given the priority lists in each mode, the image is covered following specific constraints. On the one hand, the priority list serves as a saliency map. Therefore, it is accessed in a decreasing order of saliency. On the other hand, a constraint is introduced so that the virtual gaze exhibits saccadic behavior. This behavior is determined by the orientation (and the intensity) of the object at the starting point (A Figure 5). From this point, and following the orientation of the object, we determine an area (B Figure 5) in which the next saccade is to be located. As shown in Figure 5, the area is determined by three parameters, α , l and L . These parameters are calculated for each saccade, taking into account the object's complexity.

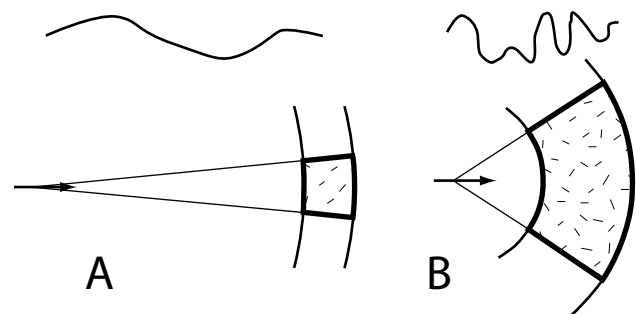


Figure 6. Different types of saccade areas. A Simple curvilinear object, B Complex curvilinear object.

We saw in Section 3.1 that curvilinear objects are represented in a vector mode in the data base (GIS) and they are simplified using the Douglas-Peucker algorithm. A complexity metric consisting of the linear density of the

points representing the vector object is used. The more complex the object is, the larger α is and the smaller l and L are. Figure 6 illustrates this. In A the target area is small because the curvilinear object is simple, *i.e.* l and L are large and of similar values, α is small. In this case the amplitude of the saccade is important. When the object is more complex (B Figure 6) α is large and l and L are small but significantly different from each other ($L \gg l$). This allows the target area to be wide and close to the starting point. Therefore, the saccadic amplitude will be short.

Hence, the next point should be in this determined area. If the first point of the priority list is located in the area, it will be the one chosen. If not, depending on the situation (a point of the priority list is located in the area or not), it will be either the first point of the priority list located in the area¹ if there is one or else the center of the area.

4.2 Information extraction

Figure 7 shows an example of a saccadic path on a visual scene for two curvilinear elements. For legibility purposes only the locations for which information has been found (squares) are shown.

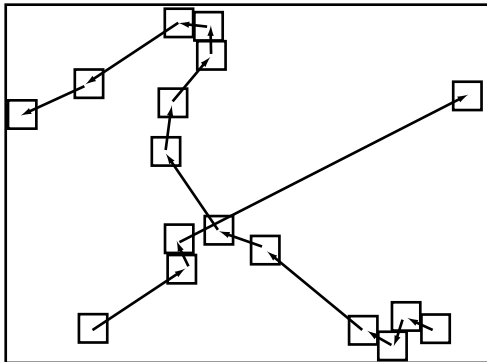


Figure 7. Example of a saccadic path for 2 objects. Each arrow is a saccade and each square (window) a fixation point.

We saw previously how the location of the next saccade is determined by a compromise between the priority list and some constraints due to the saccadic behaviour. In the case of recognition mode, another constraint is applied in order to preserve tracking during the corridor search. In order to be able to realize a saccade sequence such as the one in Figure 7, the orientation of the element on the image is needed for each fixation point.

First of all, the element is searched in the determined area (B Figure 5). Second, a window is centered on this area. And third, a filter is applied to this window in order to extract the linear features. The size of the window depends

¹Except if the point corresponds with another element in the recognition mode.

on several parameters. The problem is to get enough local information to be able to compute the next saccade.

Hence, the difficulty is to find a good ratio for the window size. A window that is too large would increase the processing time, while a window that is too small would not allow enough information extraction.

We use a classical remote sensing extraction technique. The Sobel filter seems to be the most appropriate for two main reasons. The extraction using a Sobel filter is not computed according to any particular direction *i.e.* there is no privileged direction and the algorithm is very well known and relatively fast. The direction of the extracted elements is then computed in order to determine the next saccade. In the case where there is no information in the window (no element extracted by the filter), the focus is moved to a zone next to the current window that is *still* in the target area.

5 Conclusions

The system implementation in C++ is still in progress. Using the UML formalism, we have designed a system architecture which takes into account all the ideas presented in this paper, as well as the processing of the GIS updating which has been summarized in Section 2.

The architecture and the various modules have been customized and optimized. The vector objects have been designed and the programming language has been chosen in order to get the best efficiency possible. Indeed, the ultimate objective is to update a rescue aircraft's database. Therefore, this updating process has to be performed very quickly.

This work is a good alternative to classical techniques used for GIS updating. It presents some interesting advantages and perspectives, especially where classical methods fail. The principal strengths are:

1. The focalization on areas of the image in which the relevant information is located and consequently, the ability to avoid non-relevant zones.
2. The possibility of processing bad images for automated GIS updating.
3. The high speed of the processing because of the technique.

With this work we propose a new way of updating geographic databases. The integration of cognitive science results in different domains of application opens interesting and new perspectives.

6 Acknowledgments

This research is supported by GEOIDE (GEOmatics for Informed DEcisions, <http://www.geoide.ulaval.ca>) under a Ph.D. scholarship of the MNG#ELD project.

References

- [1] C. Dillabaugh, K. Niemann, and D. Ridcharson. Semi-automated extraction of rivers from digital imagery. *GeoInformatica*, 6(3), 2002, 263–284.
- [2] M. Dobie, P. Lewis, and M. Nixon. Curvilinear network extraction from remotely sensed image. *IAPR Workshop on Machine Vision Applications (IAPR MVA'94), December 13–15*, 123–126, Kawasaki, Japan, 1994.
- [3] B. Yee. An expert system for planimetric feature extraction. *Proceedings of IGARSS'87 Symposium, May 18–21*, 321–325, Ann Arbor, Michigan, USA, 1987.
- [4] Y. Zhang. A method for continuous extraction of multispectrally classified urban rivers. *Photogrammetric Engineering and Remote Sensing*, 66(8), 2000, 991–999.
- [5] J. Wang and P.J. Howarth. Edge following as graph searching and hough transform algorithms for lineament detection. *IGARSS'89, May 10–14*, volume 1, 93–96, Vancouver, B.C., Canada, 1989.
- [6] M. Barzohar, M. Cohen, I. Ziskind, and D. Cooper. *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, chapter Fast Robust Tracking of Curvy Partially occluded Roads in Clutter in Aerial Images, 277–286. Birkhauser Verlag, Switzerland, 1997.
- [7] M. Fischler, J. Tenenbaum, and H. Wolf. Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique. *Computer Graphics and Image Processing*, 15, 1981, 201–223.
- [8] D. Geman and B. Jedynak. An active testing model for tracking roads in satellite images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1), 1996, 1–14.
- [9] D. McKeown and J. Denlinger. Cooperative methods for road tracking in aerial imagery. *IEEE Computer Vision and Pattern Recognition, June 5–9*, 662–672, Ann Arbor, Michigan, USA, 1988.
- [10] N. Merlet and J. Zerubia. New prospects in line detection by dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4), 1996, 426–431.
- [11] A. Zlotnick and P. Carnine. Note: Findind road seeds in aerial images. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 57(2), 1993, 243–260.
- [12] D. Marr. *Vision : A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman and Co, San Francisco, 1982.
- [13] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12), May 2000, 1489–1506.
- [14] D. Brockmann and T. Geisel. The ecology of gaze shifts. *Neurocomputing*, 32-33, 2000, 643–650.
- [15] J. Machrouh. *Perception attentive et vision en intelligence artificielle*. PhD thesis, University Paris XI, Orsay, France, December 2002.
- [16] A. Yarbus. *Eye movements and vision*. Plenum, New York, 1967.
- [17] R. L. Gregory. *L'œil et le cerveau : la psychologie de la vision*. Neurosciences & cognition. De Boeck Universit, Bruxelles, 2000.
- [18] C. Koch and S. Ullman. Shifts in selective visual attention: Towards the underlying neuronal circuitry. *Human Neurobiology*, 4, 1985, 219–227.
- [19] L. Itti and C. Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3), March 2001, 194–203.
- [20] L. Itti. Visual attention. M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks, 2nd Ed.*, 1196–1201. MIT Press, 2003.