GERES: A PROTOTYPE EXPERT SYSTEM FOR THE GEOMETRIC RECTIFICATION OF REMOTELY-SENSED IMAGES

Mr. M.I. Heard & Dr. C.A. Higgins
Department of Computer Science
University of Nottingham,
NG7 2RD
United Kingdom.

Professor P.M. Mather,
Department of Geography
University of Nottingham,
NG7 2RD
United Kingdom.

Abstract

Remotely-sensed images contain a wealth of information that is inaccessible to users who require such data to be accurately located. The problem of geometric rectification encompasses many areas of specialist knowledge, including those dealing with spatial and statistical analysis. An example of a prototype expert system that has been developed to assist in the process of geometric correction is described.

Key Words: Remote-Sensing, Expert Systems, Geometric Rectification

1 INTRODUCTION

Remotely-sensed images have been widely used by the environmental and Earth science communities for over 15 years. Both photographic prints and digital imagery have been used for interpretation purposes, for example: to produce thematic maps showing land cover, to generate lineament maps for use in structural geology, or to create sea surface temperature maps for oceanographic purposes. The digital form of imagery has commonly been preferred, largely because it can be stored and manipulated using digital computers. Sophisticated software packages are now available to perform a wide range of image processing operations, ranging from contrast enhancement to filtering and classification. However, since the mid-1980s it has become apparent that remotely-sensed data must be used in conjunction with other spatial data if more and better information is to be extracted. For example, methods permitting the use of topographic maps to provide control points which were used to correct imagery in order to relate pixel values to their geographical location were developed in the late 1970s and early 1980s. More recently, the introduction of geographical information systems has allowed the user of remotely-sensed data to integrate these data with other sources of spatial information such as soil and geological maps as well as digital elevation models. Such integration will permit the more effective use of remotely-sensed data.

All but the most naive uses of remotely-sensed data require that those data are located in terms of a recognised coordinate system (either latitude/longitude or a grid defined by or related to a map projection) so that the remotely-sensed data, or products derived therefrom, can be accurately referenced on the surface of the Earth. Location is by definition the primary characteristic of spatial data. Geographical attributes that are spatially unreferenced are of little use. Other image-processing and information-extraction operations such as enhancement, filtering, transformation and classifications are, in this sense, subsidiary to the geometric operation which relates map and image coordinates in a defined and accurate fashion.

One of the difficulties facing the user of a spatial information system is the level and complexity of technical, as opposed to subject-matter, knowledge that must be acquired before such a system can be used successfully. The user may, for instance, be an ecologist interested in land-use change, and we will assume that he is an expert in this field who wishes to use remote sensing to provide input to his studies. The input he requires may be thematic (such as land cover maps at different dates) or parametric (such as estimates of one or more properties of the vegetation in particular areas; examples are biomass, vigour and percent spatial cover). Current technology, in the form of image processing systems, assumes that the ecologist is familiar with the terminology of image processing and with the technical details of each procedure that he wishes to employ. Additionally, the ecologist may be expected to be aware of the steps that are necessary for meaningful information to be extracted from raw or system-corrected remotely-sensed data. Such steps include sensor calibrations, to convert digital counts to top-of-the-atmosphere radiance, atmospheric corrections to remove or reduce the effects of scattering, geometric corrections needed to register image to image, or image to map, as well as corrections for illumination variations, view angle effects, and terrain effects. Clearly it is unreasonable to expect a scientist whose primary interests lie in the use of information deriving from the data to become an expert in the technicalities of data processing, photogrammetry, and environmental physics. The solution to this difficulty lies in (i) the establishment of active data centres which provide data which have already been corrected for the various effects listed above, and which are suitable for the intended use, and (ii) to incorporate expert knowledge into the image processing software that the scientist may use in the processing of his data.

2 OVERALL SYSTEM DESIGN

GERES was designed as a modular system with a straightforward and small control module. This allowed for the rapid development of a simple but functioning system that was then easily extended by adding new self-contained modules. Future enhancements require only new modules or the updating of existing ones. The modularity of GERES ensures the flexibility of the system and the ability to quickly track down and isolate bugs.

Basically, the system may be reduced to two components: the expert system itself and its GUI (Graphical User Interface). This split represents the separation between an "intelligent" knowledge-based system and the manner in which it communicates with the user: the GUI becomes the tool through which the user translates the satellite image and the map into a set of transformation functions via a list of co-ordinate pairs.

The system's two components may then be further broken down into smaller modules. Both, for example, having a module dedicated to establishing and maintaining a line of communication with the other. The GUI also has modules that deal with representing and enhancing the given map and satellite image, and with the display of the spatial distribution of GCPs. The expert system has modules that, for example, deal with the automatic detection of GCPs, with the calculation of the transformation functions and with the evaluation of the GCPs' spatial distribution.

These modules exist either as separate processes or as a distinct block of code within one process. For example, the expert system and the GUI are both separate processes, as are the image enhancement filters, whereas the code for matrix inversion and the calculation of spatial statistics form part of the same program. The user does not know, and does not need to know, of any such technical detail.

3 GRAPHICAL USER INTERFACE

The GUI is the only part of the system that the user interacts with. It is designed to be a user-friendly, easy-to-use front-end to the expert system. It allows the user to view both the image and the map at the same time, apply edge enhancement procedures to either or both, examine the spatial distribution of the GCPs and their residuals, and communicate with the expert system through a terminal-style window (see figure 1).

The image and the map may be either half-toned grey-scale images (suitable for monochrome display) or false-colour images; the windowing system automatically recognises the capabilities of the system it is running
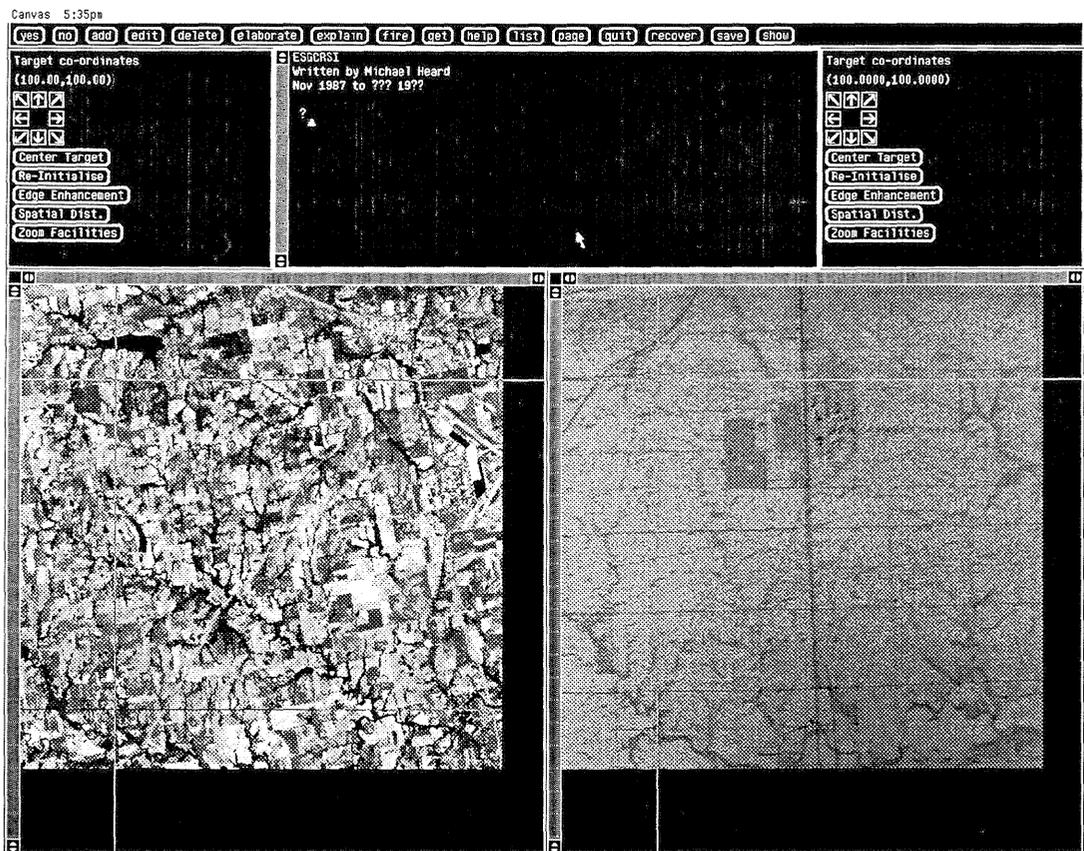
Figure 1: The Graphical User Interface.

on and responds accordingly. The size of the image or the map is handled by the use of a *scrollable* window; thus, the user can view any one part of the whole image at any time.

There are a number of tools that can be used on both the satellite image and the map which the user can apply to enhance his view of the current system status.

The spatial distribution of the GCPs collected by the expert system may be viewed separately from the satellite image and map in a compact, scaled-down window. The residuals from the computed transform at the ground control point locations may also be displayed in the same format.

The edge-enhancement procedures available at the present are the Laplacian and the Sobel filters. New filters can be added with ease as the filter procedures exist as separate modules that communicate with the GUI through a common interface. A new filter can be thus be any new program that conforms to the requirements of the interface.

4 THE EXPERT SYSTEM

The expert system encompasses several different areas of specialist knowledge, an expert rule base that knows how to access those areas, and an overall control structure that links the user with the expert rules. The rule base may be considered be the most important expert in this system as it links together all the different areas of knowledge.

The system is further complicated by the need to consider the range of users who may utilise its power: it has been designed both as a teaching aid, guiding the novice through the many problems of geometric registration, and as a tool to aid the expert. The two rôles are complementary and not, as may be supposed, contradictory. For all users, the system must provide the necessary mathematical tools to do the job in hand; any advice and decisions it makes are vital to guiding the inexperienced hand and, although not necessarily required by the expert, may prove useful to such a person.

The expert system is now examined in stages. First, there is a discussion the issues that arise from attempting to accommodate a variety of user types. Next the system's overall control structure is examined. This is followed by a description of the rule base and the way in which the system explains the decisions it reaches. Finally there is an analysis of the modules that deal with geometric registration, spatial distribution and the automatic detection of GCPs.

It is worth noting here that the expert system is written in a mixture of PROLOG and C. The programming logic language PROLOG is used to implement the rules of the expert system; PROLOG is a goal-driven (backward chaining) language that also has the necessary facilities for the utilisation of meta-rules. PROLOG, however, lacks an adequate library of mathematical functions that are needed by a number of the experts. Such mathematical functions are provided by the C language, that is used to implement 'number-crunching' routines.

4.1 Multi-Purpose Design Criteria

The system has been designed with the aim of interfacing well with both expert and novice users. It has been designed so that the novice may be guided, step by step, through the stages of geometric rectification and yet still allow the competent expert to use the system as a tool.

This functionality of the system is common to both types to user: they both require a window interface showing, at least, the map and the image at the same time, a set of mathematical and statistical tools to accomplish the task of geometric correction, a system to co-ordinate these tools and so on.

The two groups are separated only by the advice offered by the system: the novice (probably) requires copious amounts of detailed explanation whereas the expert will (usually) only need a terse pointer to the salient datum. Further, the decisions made by the system should not be rigid and should have a sufficient degree of flexibility so that the expert may override them, without it being obvious to the novice that they *are* decisions, let alone that they can be subverted.

Decisions Nonetheless, the decisions reached from a data set should be the same regardless of the user. For example, should the expert system

978

detect a clustering of GCPs it cannot know that it is the work of a novice for an expert may have made a mistake. The system should simply advise which of the points it thinks are suitable for disposal, and then wait for the user to enter his next command. The wording of the system's advice, as biased by the intensity of the clustering, should cause all users to take note; the expert, pursuing his own ends, can disregard such warnings, whereas the novice, unsure of his footing on this new and previously untrodden path, will need and ask for more information.

Thus is exposed a vital element in the design of the expert system: help and information is there for the *asking*. Although the system should produce advice and warnings after each action performed by the user, such communications should be limited to a line or two at most. Such terseness is necessary if the experienced user is to employ the power of the system without being hindered by unnecessary screenfuls of contemplation he can do without. However, the more lengthy advice *must* be quickly and easily available to the novice if he so requires it.

Help Further, there are two distinct breeds of information needed by the novice: there is a requirement for a generalised help facility and one for explaining any decisions reached by the expert system. The former is absolutely vital to the complete beginner who, confronting the system for the first time, knows little more than the fact that the satellite image needs correcting and needs reassurance over basic concepts such as 'GCP', 'least-squares fit', and so on. Such a user also requires detailed information on the sort of input that is acceptable to the expert system, not least of which is a list of 'commands' the system will recognise and act upon. The very first such command the system should recognise and be able to act upon is, of course, "help", and one of the first things that should be able to do is tell the user how to use the help facility. Such a system is enough (or at least *almost* enough) for the complete beginner, armed only with knowledge that typing "help" will aid him, to work his own way through the process of geometric correction.

Decision Justification The expert system must also be able to explain any decisions it reaches. Indeed this is one of the fundamental criteria for a system to be judged to be an expert: there are many definitions of an expert system (e.g. (Jackson, 1986), or (Hayes-Roth et al, 1983)), ranging from all-encompassing generalisations to the mathematically precise, but is almost impossible to find one that fails to include a justification element. The explanation of decisions is vital to both novice and expert: the novice will need to be told what he has done 'wrong' and why the suggested improvement is indeed so; the expert can never have confidence in a complex tool that cannot clearly exhibit its line of reasoning and, conceivably, might lack knowledge in one particular area of the system.

4.2   Expert System: Overall Control Structure

In order to register successfully a remotely-sensed image with a map it is necessary to build a list of GCPs that conform to certain criteria: the rules in the expert rule base attempt to ensure that the GCPs meet those conditions. The overall control structure is designed to encourage and help the user build and maintain a list of GCPs, but does not make any of the above decisions; this is the function of the expert rule base.

A number of commands are provided for the explicit manipulation of the GCP list. Ground control points, each in the form of a pair of co-ordinates, may be entered via the keyboard or 'grabbed' from the display or even offered by the expert system itself. They can then be examined and edited or discarded as necessary. In this sense the expert system is driven by the user: the control structure waits for user input, acts upon it, consults the expert rule base (which then consults expert modules as needed, makes decisions, acts upon them and advises the user appropriately), and then loops back to wait for more user input. This controlling loop only terminates when the expert rule base decides that the list of GCPs conform to all the necessary criteria and the satellite image can be said to be registered with the map.

As well as those designed to help the user construct the list of GCPs, the user can also issue commands to control various aspects of the window display, and ask the expert to explain its decisions. This is the limit of the user's control over the expert system. He cannot directly access the rule base nor any of the expert modules since access to the expert modules is controlled by the expert rule base, access to which is in turn controlled by the control loop.

4.3   The Expert Rule Base

This expert system has been built to achieve the goal of registering a satellite image with a map using the method of ground control points. The solution to this problem is a list of GCPs that conform to certain criteria: that there are a sufficient number of GCPs, they are evenly distributed, that their associated residuals are small, that the polynomial transformation functions are of the correct order, and so on. The GCP list is built, as explained above, by the user entering and altering the locations of GCPs, one at a time. The expert system must advise the user each time the GCP list is updated on how it fails in being a possible solution to the problem in hand.

Production Systems The problem is neatly solved using a system based on production rules (see (Jackson, 1986a), (Davis and King, 1977), Waterman (1985) p.63-69 for an introduction and overview of production systems). A production system consists of a rule base, an interpreter that decides which rules to fire, and 'working memory', that can hold information be it raw data, processed data or a list of rules to be fired. Each rule in the set has the form:-

```
If condition 1 is true,
    and condition 2 is true,
    ...
    and condition n is true,
            then perform actions A, B, ...and N.
```

PROLOG provides an ideal environment for implementing a production system: its provides an interpreter and working memory, leaving the expert rules to be encoded as PROLOG statements.

The Expert Rules As mentioned above, the rules in the rule base are designed to ensure that the user-built list of GCPs constitute a valid 'solution' to the problem of registering a satellite image with a map. The problem can be said to be solved when that list meets all the criteria encoded in the rule base. Thus, the rule base lies at the heart of the expert system for it alone determines when an image has been successfully matched with a map. All the necessary intelligence and knowledge, from the many specialist domains, is set out in that list.

Implicit in the above paragraph is the idea that the rules are more 'negative' than is usually the case: each rule represents a potential flaw in the list of GCPs. Thus the rules are more strictly of the form:-

```
If condition 1 is true,
    and condition 2 is true,
    ...
    and condition n is true.
            then the list is lacking in this manner.
```

The many actions have become just one, and it is of the same form each time: a warning to the user that he is failing in some manner, although the textual output tries to be a little less critical and a little more helpful in tone. The action becomes a suggestion to the user that will improve the suitability of the GCP list. A few concrete examples will now be given for the sake of clarity.

One of the simplest rules in the rule base is:-

```
If there are less than twenty points in the list,
    then more GCPs need to be found.
```

The condition to be met is the number of GCPs in the list: if it is less than twenty the user is advised to find more. The explanation associated with this advice is that twenty is the minimum for a statistically valid sample.

A more complex example is:-

```
If X is one of SE, NE, NW or SW,
    and there are no points in the X quadrant
                        of the image,
    and there at least four points in the GCP list,
            then more GCPs need to be found in the
                    X quadrant of the image.
```

This rule is a crude attempt to achieve an acceptable spatial distribution of the points by forcing the user to locate at least one GCP in each quadrant of the image. This rule is not sufficient in itself to force an acceptable spatial distribution of the GCPs but it is useful in indicating to the user that he should pay more attention to the position of the GCPs he chooses.

Expert Rule Interpreter It is evident at this juncture that it is necessary to apply the rule base to the GCP list every time it is updated. Further, it may be possible to fire more than one of the rules in the rule base after any one particular update of the GCP list; indeed it is highly likely that this will be the case until the final solution is achieved. For example, the user may have less than twenty points in the list and they may exhibit clustering as well as having other undesirable features. Which rule or rules should be fired?

Clearly the standard PROLOG interpreter is not powerful enough in this situation, for it would fire only the first rule that satisfied all the conditions and would "backtrack" to other solutions only when forced to do so. This approach is too predictable and inflexible. Instead, *all* rules that meet the prevailing conditions should be considered, and a appropriate subset of these can then be activated. The following points should be considered when addressing the question of which rule or rules should be fired:-

- The user should not be overwhelmed with information, as this will only serve to confuse the issue.

- The user should not have the same advice repeatedly offered to him after each update. Clearly it is neither desirable nor necessary that the user be continually informed of the large residual associated with one particular GCP.

- The user should be given pertinent advice. For example, the information that a newly entered GCP has caused a cluster of GCPs to appear on the image is more important than the datum that that there are not enough points in the list for it to be a statistically valid sample.

Firing only one of the rules from the expert rule base ensures that the user is not inundated with advice. The leaves the question of which rule to fire. To this end each rule is given a base 'importance weighting'. After an update the current 'importance weight' of each rule in the rule is calculated: this is presently the base weight or zero if either the rule cannot be fired (i.e. one or more of its conditions cannot be met) or it was fired after the last update. The rule with the greatest importance weight is fired, and its advice is offered to the user. If all the rules have a current weight of zero no rule is fired and no advice is offered to the user.

The system can be made to behave along the lines indicated in the above considerations by attaching the correct weights to each of the different rules. This is, as it stands, an admittedly crude system but it does have the advantage of working. It is hoped that a more subtle but equally effective system of weighting will be found to work in the future.

4.4 Explaining Decisions

Near the centre of all expert systems lies the decision justification process. No user can have confidence in a system which generates advice without accompanying with reasons. GERES makes only one decision whilst it is running: does the list of GCPs represent a transformation between satellite image and map? Thus, there are two basic answers to justify and two different methods of justification.

When, in the expert system's opinion, the user has successfully rectified the image, it must justify that opinion. It does by enumerating the general requirements for geometric rectification: that there are a statistically valid number of GCPs; that they have an acceptable spatial distribution; that the transformation functions are accurate; and so on. It will then explain any or all of these points further to the user as required,

More detailed explanation may be required when the system decides that the GCP list fails to meet the criteria in some manner. Thus, although it explains that, for example, the points are poorly distributed it should also indicate where they are lacking and where they are clustered. Sometimes such decisions are self-explanatory (e.g. "There are no GCPs located in the western half of the image.") and sometimes further detail needs to be given (e.g. when asked to explain the general message "The GCPs are badly distributed." the system should inform the user of the tests applied and the results produced).

5 THE GEOMETRIC REGISTRATION PROBLEM

5.1 Transformation Functions

These typically take the form of the low-order polynomial functions. For example,

$$E = a + bX + cY$$

where

- E is the map easting;

- X, Y are the image co-ordinates;

- a, b, c are constants to be determined.

The method of least-squares fit, traditionally used to find the best curve describing a set of data points, has been applied to the geometric correction problem to give the most accurate transform possible in each case (Hardy, 1978). In the case of the linear polynomial it gives us, in the first instance, three simultaneous equations to solve for each transformation function.

The 'correctness' of the transformational functions is measured on both a global and point basis using a combination of residuals, the difference between the observed GCP location and that calculated using the function, and statistics proposed by Ford and Zanelli (Ford and Zanelli, 1985). A further test requires that the rms (root-mean-square) error should be less than 1 pixel edge.

5.2 Spatial Distribution

A set of GCPs with a poor spatial distribution will lead to poor transformation functions (see (Mather, 1976), for examples). Any clustering or linear arrangement of GCPs give deceptive results for the functions will give small error residuals around the clustered areas and along the lines but will get steadily greater away from such features. The Clark-Evans and the Ripley-Silverman tests (Ripley, 1981) have been implemented along with quadrant-counting procedures to give an indication of the spatial distribution of the points to the user.

5.3 Automatic Detection of GCPs

The current method employed for the automatic detection of GCPs involves passing the image through a series of filters, such the Sobel edge-enhancement filter, until a thinned binary image is obtained. This image is then scanned for the crossing points of lines. The image is scanned left to right, top to bottom and each black (set) pixel is considered in turn. The implemented algorithm considers it to be a junction (i.e. a GCP) if at least three distinct 'roads' can be traced from it.

6 FURTHER DEVELOPMENTS AND INTEGRATIONS

Although a functioning system, GERES, having been developed as a prototype, is in no sense complete. There exists room for enhancing the internal working of the expert system, developing the functionality of existing modules and integrating new areas of specialist knowledge into the system in the form of modules.

One new knowledge domain that could be added is that of image segmentation and classification. Applying the techniques afforded by this speciality would allow 'unusable' regions of the images (or possibly even the map) to be detected and compensated for. Regions of the image covered by, for example, water or cloud will not yield any GCPs and should be taken into account when calculating the spatial distribution statistics and advising the user on which areas are lacking in GCPs. Further, any GCPs detected from such regions should be rejected out of hand.

The automatic recognition of GCPs could be much improved from is current state where only possible locations on the satellite image are suggested to the user. GCPs are pairs of co-ordinates; the position on the map corresponding to a crossroad on the satellite should determined and then both locations should be presented to the user. Two approachs to this problem may be tried:

1. Use of the same pattern recognition techniques to the (bitmap image of the) map so as to produce a list of co-ordinates. Using the transformation functions, this list can then be merged with compiled from the satellite image to form a list of possible GCPs. This method can only be used when at least three GCPs have been registered with the system.

2. Examination of each of the satellite locations and attempt to find the corresponding position on the map. This would probably have to be guided by transformation functions but it would also utilise contextual information (such as the number of vectors at the junction, the angles between them and so on).

Further, there is the question of which of the automatically-detected GCP should be suggested to the user first; currently the system presents them in the order in which it receives them from the module. This question is closely tied in with the problem of determining whether a location extracted by pattern recognition algorithms would be a junction as far as the human eye and mind are concerned for the routines will outline "roads" (e.g. crop patterns) ignored by the user.

At present, each image is considered on its own merits. This could be extended by examining the history of the images passing through the system.

For example, a sequence of images from the same overall area would provide information useful in showing the change of land use over time. The expert system could be expanded to interact with Geographical Information Systems (GIS's), which could make use of such information.

Satellite images may be classified as being of high, medium or low resolution. The expert system has been designed to deal the medium resolution image, which can be dealt with using the relatively simple technique of ground control points, but could be easily extended to other types of imagery that could be corrected using a similar system. This has been achieved with AVHRR images using GCP "chips" (typically fragments of coastline), as opposed to single locations, without encountering many additional problems. High-resolution images, usually produced by aircraft, present more of a challenge, principally because of the yawing, pitching and rolling of the sensor platform.

7 CONCLUSION

Geometric correction can logically be considered to be one of the fundamental preprocessing operations to be carried out on remotely-sensed images in that the information contained within the image cannot be properly utilized in conjunction with other spatial data unless the image is locationally-referenced. The prototype system described in this paper offers an answer to the problems resulting from the increasing complexity of the operations that the user of remote sensing data faces in attempting to extract information from the data and enter it into a GIS. Enhancements and extensions of the prototype system are described; these will further improve the system's capabilities and provide an efficient means of checking the geometric quality of remotely-sensed data.

References

Davis, R. and King, J., 1977. An overview of production systems. In: Elcock, E. and Michie, D. (eds) *Machine Intelligence 8*. Ellis Horwood, pp. 300-332.

Ford, G.E. and Zanelli, C.I., 1985. Analysis and quantification of errors in the geometric correction of satellite images. *Photogrammetric Engineering and Remote Sensing*, **51**:1725-1734.

Hardy, J.R., (1978). Methods and accuracy of location of Landsat MSS points on maps. *Journal of the British Interplanetary Society*, **31**:305-311.

Hayes-Roth, F., Waterman, D. and Lenat, D., (eds) (1983). *Building expert systems*. Addison-Wesley, pp. 16.

Jackson, P. (1986). *Introduction to expert systems*. Addison-Wesley, pp. 1.

Jackson, P. (1986). *Introduction to expert systems*. Addison-Wesley, pp. 29-51.

Mather, P.M. (1976): *Computational methods of multivariate analysis in physical geography*. John Wiley and Sons, pp. 120-130.

Ripley, B.D, (1981): *Spatial statistics*. Wiley.