AUTOMATIC EXTRACTION OF LARGE BUILDINGS FROM HIGH RESOLUTION SATELLITE IMAGES FOR REGISTRATION WITH A MAP

V K Vohra and I J Dowman
University College London

Commision III, Working Group III/2

KEY WORDS: Feature Extraction, Boundary Matching, Match Points, Correlation, Registration.

ABSTRACT

Feature extraction plays an important role in extraction of meaningful structure in digital images. It aims at replacing the iconic representation of the image content by a symbolic image description, representing the essential parts of an image related to some task. In the work described in this paper the main interest is in extracting large buildings as polygonal features from high resolution satellite data which can be matched with maps for registration. The emphasis is on polygons which have a distinctive shape and can be extracted from vector as well as raster data. Algorithms used are described, for extraction of polygonal features from map and image, and also for matching map and image features to generate match points. The match points derived are used for determining transformation parameters. The registration of image to map is achieved by a applying bilinear resampling method using the parameters on the image. Results in terms of accuracy are given and future work is also discussed for fully automation of image to map registration.

1. INTRODUCTION

The quality of images, the basic models and evaluation algorithms, and also the way the data is represented are important parts for the efficiency of the image processing systems. Image processing is applied in several stages to extract important image information, to suppress redundant information and neglect information which is not used in the processes.

Feature extraction is a very active area of research for photogrammetry and remote sensing. Forstner (1993), Schenk (1993) and Sester (1993) have given comprehensive reviews of the current status and philosophy. Points, lines or polygons can be extracted as primitives, but work to identify and extract features is at present object oriented. Holm et al, (1995) has described a method of map to image registration for particular flat terrain with many lakes in Finland. The automation of the relative orientation process has been described by Haala et al, (1993) and by Hellwich et al, (1994). The basis of both these methods is the use of points determined by interest operators within an image pyramid, the matching of these points and the use of iconic and geometric constraints to eliminate false matches. Heipke (1993) has reviewed automation in orientation and work is in progress to use features such as roads or polygons instead of points to determine the orientation of images, either single or stereoscopic. Stevens et al, (1988) have attempted to match map features directly with images. Lee et al, (1993) have described an automatic method of registering an image to a map in two dimensions which uses non-rigorous methods but claims high speed and accuracy. Shahin et al, (1994) have been investigating the automatic detection of roads from SPOT data and matching these to road co-ordinates determined by GPS.

The present work leads to the automation of registration of an image to a map which is not yet solved in a flexible and general way. This is in contrast to the image to image registration problem which has been solved and can be applied to images from the same sensor or from different sensors. Automation of the registration processes is based on methods of feature extraction, as the main operation is concerned with finding a set of common points i.e. match points on the map and on the image. The features which are most suited to this are polygons because they can have a distinctive shape and can be found in urban and non urban areas. Suitable points are not easily found automatically from map data, but polygons can be extracted comparatively easily from both raster images and vector maps. Therefore, the extraction of polygonal features are considered in this paper to generate sets of match points from map and image for the registration of image with a map. The whole procedure is divided into three stages:

- Extraction of polygonal features from map and image.
- Matching polygonal features of map and image to generate a dense network of match points.
- Using match points for registration of image to map and assessing the results in terms of accuracy for the validation of the registration system.

2. FEATURE EXTRACTION

Large buildings can be seen very clearly and distinctly on the high resolution satellite images. Considering this the extraction of building features is planned for the test and the area of the Defence Research Agency site at Farnborough, UK is selected which shows large buildings and airfield facilities. The data used to extract buildings in polygonal form is taken from a high resolution Russian DD5 digital image with approximately 2.5m pixel size and from an Ordnance Survey 1:10 000 map in raster format.

It is important to bring the map and the image into same input data level i.e. edge representation for matching purpose. The map and the image, are seperately prepared here to get the boundaries of buildings in the map and the edges of buildings in the image.

2.1 Preparing the Map for Matching

The Ordnance Survey map of the test site is shown in Figure 1(a) which consists of 900 x 500 pixels in size. It is desired to have a map in vector format, but in the present work this is available from the Ordnance Survey only in raster format. It is essential to convert the map into vector format which can be used for matching with the image.

A good matching can proceed based purely on the knowledge of where edges of images and boundaries of map exist along with their gradient direction information. The following steps are taken to determine map region boundaries of buildings and their gradient direction.

2.1.1 Converting Buildings to Solid Objects

The Farnborough map area shows names of roads, roads and large buildings with stipples. In order to use the map for matching it is necessary to remove the stipples, roads and names from the map and produce boundaries of the buildings. This section discuss a method to remove the stipples to convert buildings to solid objects and the next two section deal with the removal of roads and names.

A powerful and flexible software, Human Information Processing Software (HIPS) system based in a UNIX environment is used for processing images. But, at this stage the conversion of buildings to solid objects is performed manually by using the drawing package in the Erdas Image system, because it is not available in the HIPS system.

All the features in the map have a value of digital number (DN) 255 on a background value of DN 0. The map is transferred to the Erdas Imagine system from HIPS system to give one unique DN value (127) to buildings to achieve homogeneity and converting buildings as solid objects. The solid buildings are separated from other objects like lines or letters of names attached to it by simply using the drawing package as shown in Figure 1(b). The building solid object map is then transferred to HIPS system for further processing.

2.1.2 Unique Identification to all Objects

It is necessary to remove small objects and clutter from the map which are not required for matching. This can be done by performing two steps. Firstly by giving unique DN value to each object which is described in this section and secondly by the size (number of pixels) of objects to be threshold.

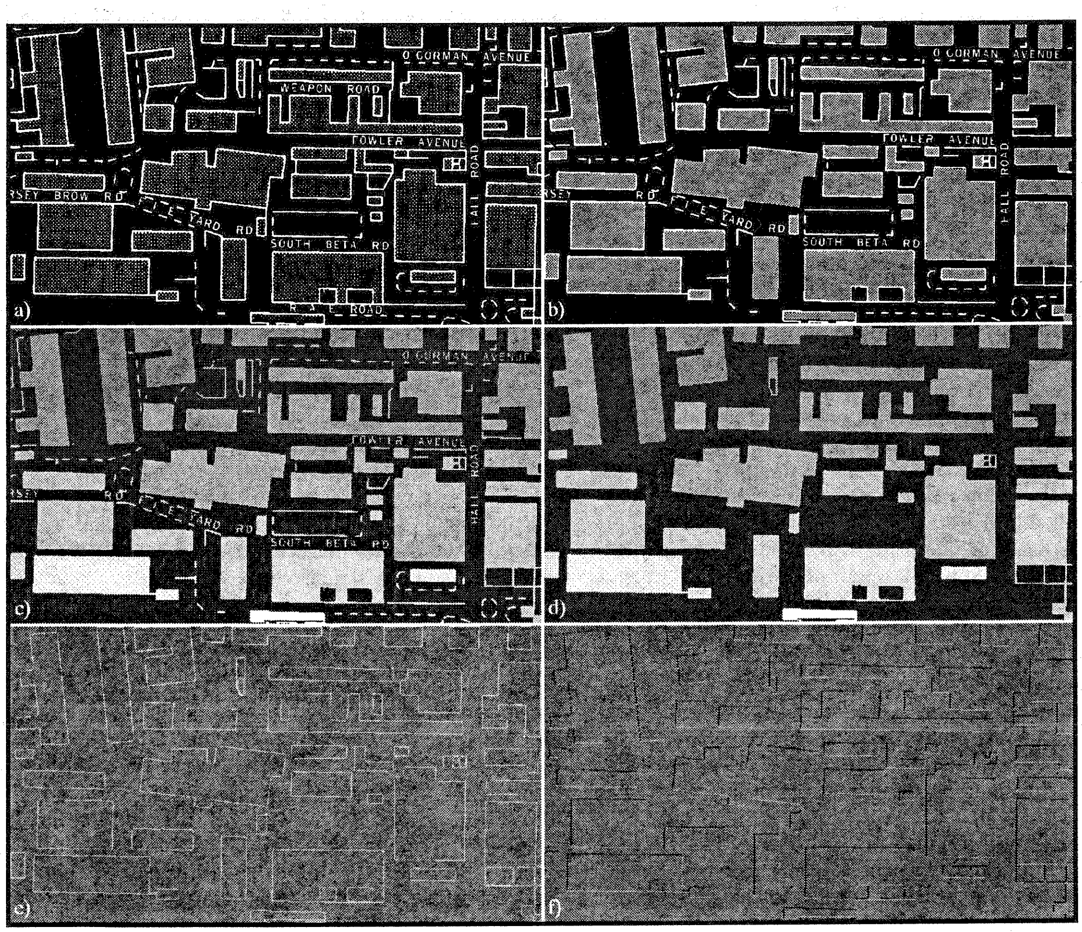

Figure 1. a) Ordnance Survey 1:10 000 map of Farnborough area, b) Map of buildings as solid objects of Farnborough area, c) Map of Farnborough area with unique identification to all objects, d) Map of Farnborough area with clutter information removed, e) Map of Farnborough area with building-boundary regions, f) Gradient direction of map boundaries of Farnborough area.

An algorithm is used which scans the map from top left to bottom right and while scanning it assigns value DN 1 to the first object it encounters and DN 2 to second object and so on to all objects in the map. Figure 1(c) shows each solid building object, separate roads and each letter of street or place names as a separate object in the map and all are assigned an individual unique DN numbers between 0-255.

2.1.3 Elimination of Clutter Objects

At this stage each object in the map has a unique DN value. An algorithm is used with a threshold size (number of pixels) value and below this value all objects are removed from the map. The algorithm first looks for an object having value DN 1 and counts number of pixels in that object, and if it is greater in size than the threshold size chosen then the object with DN 1 remains in the map, otherwise it is removed i.e. a zero DN value is given. In the same manner the algorithm looks for other DN value objects in the map, and removes clutter and small size objects. Figure 1(d) clearly shows a map of the solid buildings after removal of objects not required .

2.1.4 Map of Building Region Boundaries

To find the boundaries of the solid building regions, an algorithm is used which works in a very simple way. It scans the map of solid building regions from top left to bottom right. The first pixel of a solid object it encounters is considered as the first boundary pixel of that object. From that boundary pixel, the algorithm starts looking for such neighbouring pixels which have the same DN value of that object as well as lying adjacent to the background DN value i.e. DN 0, and it then traces the boundary of that object. After this solid building region boundary is traced, the algorithm looks for the next solid object, and in the same manner traces the boundaries of the object and the subsequent solid objects in the map [see Figure 1(e)].

2.1.5 Gradient Direction of the Map Boundaries

For determining the best match between the map and the image, edge pixel direction in the image and boundary pixel direction in the map are used. An algorithm is used to determine the directional component of each map boundary pixel. A two frame sequence input is used to apply this algorithm. The first frame contains the map boundaries and the second frame contains the solid regions from which the boundaries were defined. The output also results in two frames, first map boundaries, and second map direction as shown in Figure 1(f). The map with these two frames, boundaries and its directions, are ready at this stage to be used as input for matching .

2.2 Preparing the Image for Matching

The Farnborough subscene image is shown in Figure 2(a) which consists of 230 x 180 pixels. The aim here is to extract edges that define edges of the building regions and at the same time to suppress edges that do not represent building regions. The pre matching steps for preparing image for matching are described below:

2.2.1 Edge Preserve Smoothing

An edge preserving filter is applied prior to edge detection to strengthen the grey level discontinuities between different land cover types, and to reduce the detection of edges in areas of texture that are internal to regions. An algorithm used is an adaption of that outlined by Matsuyama et al, (1980) and Tomita et al, (1977). A window with nine masks (filters) is passed over the image and the variance is measured in 9 orientations (masks) around the central pixel. The orientation of minimum pixel variance is determined, and the mean of this is given to the central pixel. The selected orientation is never accross an edge. This is performed for each pixel in the image. The algorithm is iteratively applied to achieve maximum homogeneity of each region in the image as shown in Figure 2(b).

2.2.2 Region Segmentation

Segmentation is the splitting up of an image into regions which hold properties distinct from their neighbours, and it is generally approached from two points of view: by detection of edges that separate regions, or by the extraction of regions directly. Using a histogram derived from the edge preserving smooth image and thresholding it at value DN 38 resulted in direct extraction of building regions including some clutter as shown in Figure 2(c).

2.2.3 Edge Enhancement

Edge enhancement determines, for each pixel in the image, its edge strength and the direction of the gradient of the edge at that point. This is obtained by image differentation, which is itself achieved by the convolution of various kernels with the image. The Sobel Operator is used which consists of two kernels (X and Y) and are passed accross the region segmentated Farnborough image. The strength of the edge at the central pixel of the kernel and its gradient is determined for each pixel in the image as:

$$\text{Strength} = \sqrt{[(\text{Result of convolving kernel X})^2 + (\text{Result of convolving kernel Y})^2]} \quad (1)$$

$$\text{Direction} = \tan^{-1}[(\text{Result of convolving kernel X}) / (\text{Result of convolving kernel Y})] \quad (2)$$

The result of edge enhancement is shown in Figure 2(d).

2.2.4 Non Maximal Suppression

The result of edge enhancement shows edges of two or more pixels thick. Non Maximal suppression seeks to remove those edgels (pixels) that are not local maxima, thus sharping the representation of edges effectively to thin the edge to a single pixel width. An algorithm is used which considers the edge strength and gradient direction information. The edge pixel is passed if the two neighbouring pixels along the gradient direction are less than or equal to it in strength, if not the pixel is set to zero. The result of Non Maximal Suppression algorithm is shown in Figure 2(e).

2.2.5 Alter Directions of Edge Pixel Gradient

The edge gradient direction is a useful element in the matching procedure, where edge pixel gradient directions are compared to map boundary pixel gradient directions to obtain good matches. However, there is always some rotational difference between the image and the map spaces. To compensate this, an adjustment is required to the gradient directions calculated for the edge pixels. Four control point pairs are selected in well distributed manner manually from the map and the image, and are used to define a similarity geometric transformation between the map and the image using an algorithm. The parameters of this transformation are used to find the rotational differences, which in this case is -39.3925 in degrees to alter each edge pixel gradient direction.

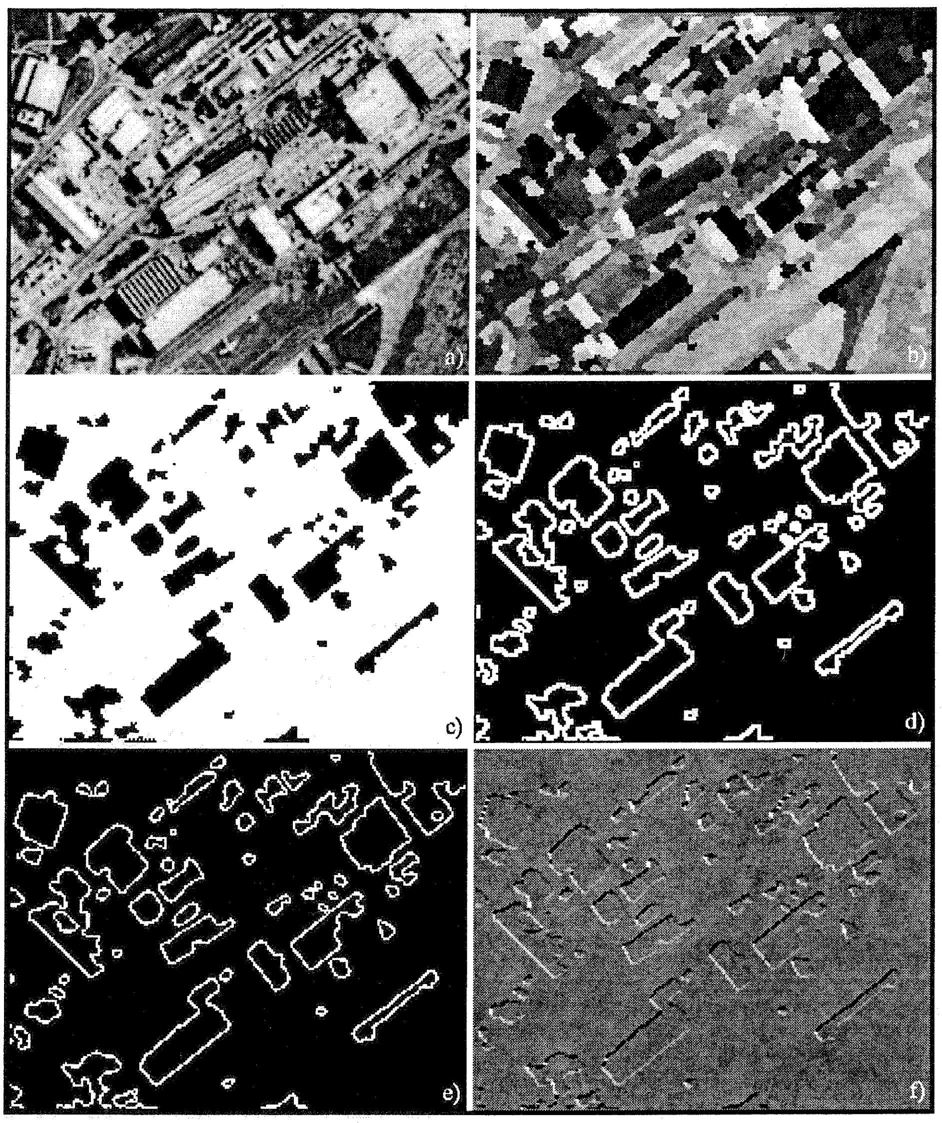

Figure 2. a) Histogram equalisation contrast stretched subscene of Farnborough image, b) Edge preserve smoothness of Farnborough subscene, c) Building region segmentation of Farnborough subscene, d) Edge enhancement of Farnborough subscene, e) Farnborough subscene - edges to the local maxima, f) Edge gradient direction of Farnborough subscene.

The algorithms used above for the preparation of the image, resulted in a two frame output, the first is edge strength and the second edge gradient direction which is shown in Figure 2(f). The image with these two frames, edge strength and its directions, is ready at this stage to be used as input for the matching with the map.

3. MAP AND IMAGE MATCHING

The matching routine owes much of its inspiration from Maitre et al, (1989), and the use of dynamic programming method for matching purpose is described by Newton et al, (1994). The core of the algorithm is a routine for matching map boundaries to image edges. Four control points are selected manually in well distributed pattern for an initial 2D transformation between the map and the image. The matching routine with these control points project boundary pixels from the map into the image space to predict edge pixel position. Matching is defined by two components of a cost function:

• The distance between predicted edge pixel and the edge pixel under consideration is the first cost component.

• The difference in their gradient directions is the second component of cost.

Only one edge pixel can be matched to one map boundary pixel, and the edge pixel under consideration with minimum cost is considered as the best match.

The use of these dynamic programming method generated 923 match points which are shown in Figure 3, and also produced a match point file containing co-ordinates of these match points in the map space and its corresponding points in the image space. This match point file is the source for the further processing to register image with the map.

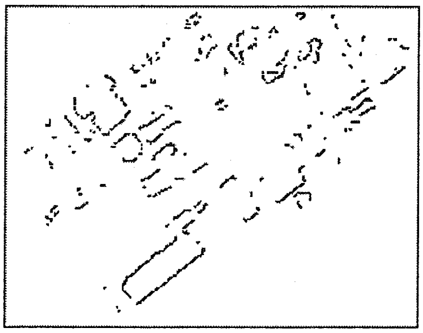

Figure 3. Matched points of map-image pixels.

4. REGISTRATION OF IMAGE TO MAP

The matching of polygons produced 923 match points and these points were used to find transformation parameters for the registration of the image to the map. An affine transformation gave a rmse of 2.0 pixel using match points. This is not an acceptable result. It is thought that the clutter as shown in Figure 2(c) may be the cause of large errors.

A post region segmentation processing is performed to remove the clutter as is shown in Figure 4.

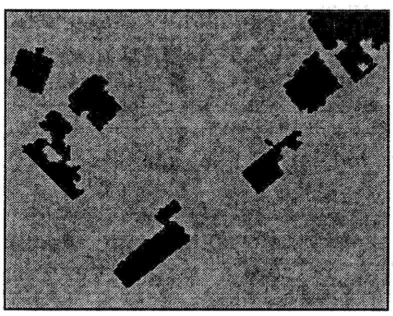

Figure 4. Building region segmentation without clutter.

The edges of buildings are extracted from the image. Then the matching of polygons in the map and the image are performed and which resulted in 393 match points. An affine transformation on 393 match points resulted in a rmse of 1.8 pixel. Figure 5 shows a graph of the magnitude of the absolute residual vector against frequency. A graph showing stepwise a curve raised an interesting question. Is the difference of height of buildings creating stepwise curve? It is clear that the perspective geometry of the images will cause displacement of the roof line with respect to the building line. Buildings of differenent heights will therefore cause different errors. This appears to be the effect shown in Figure 5 in which building heights are clustered at certain levels causing clustering in the errors. An analysis of the errors and known building heights showed this hypothesis to be valid. Further more the errors are similar for each building edge. This can also be illustrated by using the projection of a 3D object on to a 2D plane if the position of the sensor is known as shown in Figure 6. In this figure BcdB' is the area which will produce large residuals and AA' is the line with very low residuals.

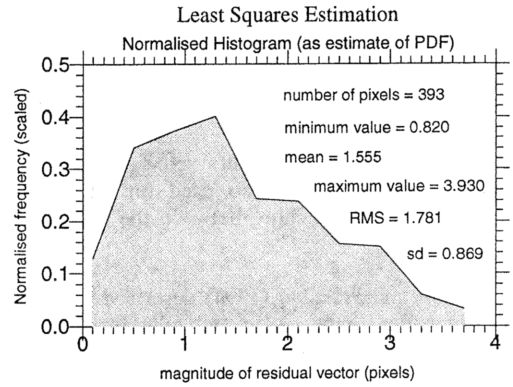

Figure 5. Residual vector vs. frequency of 393 match points.

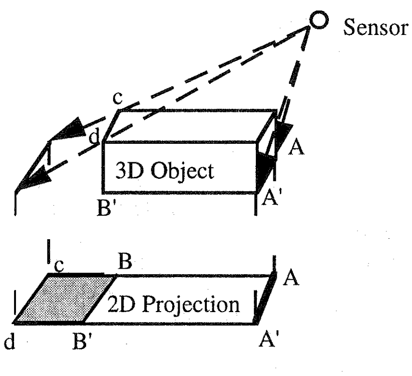

Figure 6. Model of perspective distortion.

A selection of 64 good points on undistorted edges gave a rmse of 0.32 pixels after an affine transformation. This demostrates that the method can generate sufficient planimetric points for absolute orientation.

The application of bilinear resampling using the parameters on the image resulted in the registration of the image to the map as shown in Figure 7.

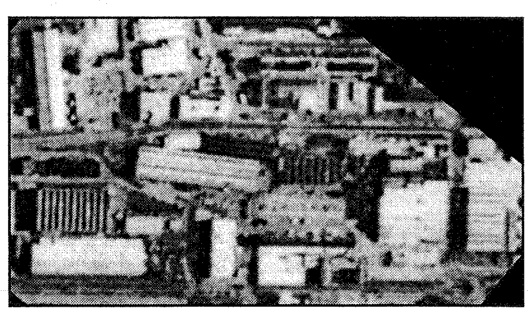

Figure 7. Farnborough subscene registered to the map.

5. DISCUSSION AND CONCLUSIONS

The system developed in this work is semi-automatic for the registration of the image to the map. A few components of the system are performed manually as mentioned below:

- converting buildings to solid objects in preparing the map for matching with the image.

• manually selection of four control point pairs in well distributed pattern to define a similarity transformation parameters between the map and image which are used to find the rotational differences to alter each edge pixel gradient direction.

• the same four control point pairs selected to alter edge pixel gradient direction are also used for an initial 2D transformation for the matching between the map and the image.

Abbasi-Dezfouli and Freeman (1994) described a method of chain coding using shape, size and relative geometry of patches to distinguish correctly matching patches from a pair of SPOT images without using any ground control points or ephemeris data. The same method can be applied for registration of image to map to select a set of well distributed polygons which will then be the input to the fine matching algorithms discussed above. This method will allow the removal of manual the component of selecting control points from the system. Some initial results for this method using aerial photographs are reported by Dowman et al, (1996) at this congress.

The buildings have shown large residuals on the matched points due to perspective distortion of the buildings. These are systematic displacements which are related to the camera look angle and the orientation of the buildings. It is necessary to remove the systematic errors first from the matched points, which will then allow a better accuracy of registration of image to map.

For full absolute arientation and the removal of perspective distortion height information is required. This aspect has also been discussed in Dowman et al, (1996).

6. FUTURE WORK

Future work is concerned with developing algorithms which can replace the manual components of the semi-automatic system and convert it into fully automatic system for the registration of image to map. The system will incorporate different types of sensor data, multi-resolution data and different landcover types will also be considered to make the system robust. The system will also include a model of the distortion which will identify and remove the perspective effects.

ACKNOWLEDGMENTS

Acknowledgment is made to the Space department, DRA, Farnborough, UK who supplied the DD5 High Resolution satellite data and the height of buildings of the Farnborough test area as a part of a DRA contract. Authors also feel grateful to the The Natural Environment Research Council for providing a research studentship for the development of a fully automatic image to map registration system.

REFERENCES

Abbasi-Dazfouli M and Freeman T, 1994. Patch matching in stereo-images based on shape. International Archives of Photogrammetry and Remote Sensing, 30(3):1-8.

Dowman I J, Morgado A, Vohra V, 1996. Automatic registration of images with maps using polygonal features. International Archives of Photogrammetry and Remote Sensing, 31(3).

Forstner W, 1993. A future of photogrammetric research. Geodesia 93-8:372-383. Haala B, Hahn M, Schmidt D, 1993. Quality and performance analysis of automatic relative orientation. SPIE volume 1993:140-150.

Heipke C, 1993. Performance and state-of-the-art of digital stereo processing. Photogrammetric Week 1993, Wichmann, Karlsruhe:173-183.

Hellwich O, Heipke C, Tang L, Ebner H, Mayr W, 1994. Experiences with automatic relative orientation. International Archives of Photogrammetry and Remote Sensing, 30(3):370-378.

Holm M, Parmers E, Andersson K, Vuorela A, 1995. Nationwide automatic satellite image registration system. Integrating Photogrammetric Techniques with scene analysis and machine vision II, SPIE Volume 2486:156-167.

Lee A J, Carender N H, Knowlton D J, Bell D M, Bryan J K, 1993. Fast autonomous registration of Landsat, SPOT, and digital imagery. SPIE volume 1993:68-79.

Maitre H and Wu Y, 1989. A Dymanic Programming Algorithm for Elastic Registration of Distorted Pictures Based on Autoregressive Model. IEEE Transactions on Acoustics, Speech, and Signal Processing, 37(2).

Matsuyama T and Nagao M, 1980. A structural analysis of complex aerial photographs. Plenum Press, New York. 199 pages.

Newton W, Gurney C, Sloggett D, Dowman I, 1994. An approach to the automated identification of forests and forest change in remotely sensed images. International Archives of Photogrammetry and Remote Sensing, 30(3):607-614.

Schenk T, 1993. Image understanding and digital photogrammetry. Photogrammetric Week 1993, Wichmann, Karlsruhe:197-208.

Sester M, 1993. Modelling strategies in Photogrammetry. Photogrammetric Week 1993, Wichmann, Karlsruhe:21-32.

Shahin F and Novak K, 1994. The exterior orientation of digital images by road matching. International Archives of Photogrammetry and Remote Sensing, 30(2):174-181.

Stevens A, Morris A C, Ibbs T J, Muller J-P, 1988. Automatic Generation of Image Ground Control Features From a Digital Map Database. Proceedings ISPRS Congress, Kyoto, 1988, Comm. II/3:402-413.

Tomita F and Tsuji S, 1977. Extraction of Multiple Regions by Smoothing in Selected Neighbourhoods. IEEE Trans. on Systems, Man and Cybernetics, Volume SMC-7:107-109.