IMAGE PRIMITIVES: AUTOMATING IMAGE INTERPRETATION PROCEDURES IN TOPOGRAPHIC MAP PRODUCTION

W. Tompkinson*

*Research & Innovation, Ordnance Survey, Romsey Road, Southampton. SO16 4GU. United Kingdom.
Tel: +44 (0) 23 8079 2867, Fax: +44 (0) 23 8030 5072, Email: william.tompkinson@ordnancesurvey.co.uk

Keywords: Scene understanding, image interpretation, mapping.

Abstract

An outline is provided of a research programme at a national mapping agency, that aims to use computer vision tools modelled upon the theories of vision perception to improve the efficiency of the update of topographic maps. Based upon the underlying principles of top-down and bottom-up influences on image interpretation, case studies are presented that illustrate results produced after the application of robust algorithms for extracting low-level 'primitive' information from aerial photography. In addition to contributing to research that aims to produce automatic mapping systems, it is intended that algorithms for producing these primitives will form the basis of semi-automated tools that help photogrammetric operators update maps in an increasingly efficient manner. The work presented here is an outline of research programme, upon which the author gratefully acknowledges the contributions of Panja Sae-Ui [17], Paolo Gamba et al. [9, 10], Padraig Corcoran and Nick Donnelly.

1 Introduction

Ordnance Survey is the national mapping agency of Great Britain, where it is tasked with maintaining the national topographic database. This database is currently maintained through the extensive use of field and photogrammetric surveying data capture methodologies, where it is estimated that more than 5000 changes are made on a daily basis. Although the methods of data capture used are well established, are robust and have been refined over many years of experience, the organisation constantly strives to improve the efficiency of its procedures to improve its service to its customers.

An area where it is asserted that efficiency gains can be made is in photogrammetric data capture. At Ordnance Survey, photogrammetry is the process of interpreting aerial photography to help update maps. However, although the latest digital photogrammetric software is used on high-end computer workstations, photogrammetric data capture is still a predominantly manual procedure relying heavily upon the interpretation skills of the human operator.

Accordingly, with the aid of case studies, this paper will present on-going research aimed at improving the efficiency of map production procedures through the increased automation of image interpretation techniques.

2 Background

When extracting topographic objects that are suitable for mapping in an automated manner, it must be appreciated that any landscape is inherently complex, and that current machine intelligence does not match human intuition in the extraction of useful information [18]. Automated object extraction, has been a subject of research in the digital photogrammetric and computer vision communities for over 20 years, however, no universal edge detector exists that can “both identify and track edges with sufficient success” [1]. When concerned with urban mapping, the topographic objects that may need to be mapped include buildings, roads and car parks. The extraction of these objects can be hindered because there is rarely a common template that can be used to recognise them. For example, buildings are topographic objects that require copious semantic information to extract them successfully. These objects vary in size and configuration, and factors such as shadow and sun angle (especially on roof apexes) will cause confusion to an algorithm designed to extract the outlines of buildings [19]. Accordingly, it is very difficult to reliably extract real-world objects using solely low-level vision techniques, without the incorporation of complex high-level semantic reasoning.

The continuous nature of a remotely sensed dataset (for example, an aerial photograph) presents problems for the extraction of a multitude of different types of object that will exist in any one particular scene. In addition, in a comment that infers implications to computer vision in general, Bischof and Caelli [4] state that “although the literature abounds with techniques for the recognition of isolated 2D and 3D objects, the problem of efficiently detecting and recognising such structures in complex signals or scenes has not received much attention”. Bellman and Shortis [2] further indicate weaknesses in the research strategy of the photogrammetric community in particular. Firstly, [2] generalise the procedure for automated object extraction to the steps of object identification (through image interpretation), followed by the extraction of those objects to a database (through the precisely tracing the geometrical boundary of an object). It is then asserted that photogrammetric research has focussed upon the second of these steps and consequently has been very reliant upon the image interpretation skills of the human operator to
constrain the search space of such extraction algorithms (ibid.).

As a result, it is the aim of this work to develop effective strategies for understanding the scene presented in a remotely sensed image. From this comes a solid basis for automated object extraction from those scenes.

3 Image Primitives project

3.1 Rationale

Limitations in the current state-of-the-art of automated object extraction for mapping purposes and the complexities of remotely sensed data have been described in section (2). In appreciation of these limitations, research at Ordnance Survey is designed to utilise the tools that vision science can provide. The rationale for this research is that although it is the goal of any national mapping agency to make the process of capture and delivery of map data as automated a process as possible, there is great advantage to be gained from producing tools that complement rather than replace human expertise. As such, using the visual perception literature as inspiration, a series of projects have been designed that aim to better model human cognition and focus upon using the strengths of computer vision techniques, to improve the efficiency of image interpretation for topographic map production.

This research programme is known operationally as the ‘Image Primitives project’. The work of Wolfe [22], in particular, is widely cited for its description of human visual search utilising both “stimulus driven” bottom-up and “user driven” top-down activation mechanisms. Similarly to other systems that aim to model visual search mechanisms (for example, the CITE system in [5] or GeoAida in [14]), it is intended in the Image Primitives project to produce a system that also incorporates both bottom-up and top-down procedures of visual search. However, the approach taken here differs from these systems, in that to increase the speed in which this research can be applied to a production environment, there is an increased emphasis upon applying theories of intermediate vision [21] to design tools that explicitly aid rather than replace existing human interpreters.

3.2 Strategy

Figure 1 illustrates a schematic of the work flow of this research. The overall message of this diagram is to show that given an image, simplistic ‘primitive’ information is first extracted. Then using the resulting collection of primitive information, these primitives are then used as components to both semi-automated tools that aid photogrammetric operators in their current tasks, and in on-going research that strives towards the long-term research goal of fully automated systems.

This research takes place under the assumption that the best method for mapping is one that is based upon human expertise. Consequently, in addition to information gleaned from the literature, choices made when deciding which primitives to extract are underpinned by on-going investigations such as, 1) cognitive studies aimed at identifying the types of information that human intuition looks for, and 2) usability studies to identify the exact tasks that an operator carries out.

3.2.1 Extraction of low-level primitives

The aim of this stage is to recognise and extract relatively simplistic objects and features in a robust manner. These primitives will either be discrete geometric objects such as polygons or lines, or region-based primitives that will provide an indication of areas in an image that reflect particular visual properties (such as texture).

Many of these primitive objects are chosen in accordance with visual perception literature that is concerned with bottom-up visual search procedures. For example, in the context of discrete geometric primitives, Biederman [3] explains the importance of T, Y and L–shaped corners, where each type of corner has a different status in the recognition task. Certainly within the context of building recognition these types of corners can convey pertinent information, where an L-shaped
corner could be inferred to be the corner of a building's outline, whilst T and Y-shaped corners might infer the location of a roof apex.

The underlying principle for focusing upon using primitives as the basis of recognition or extraction algorithms, is that a computer vision algorithm will be much more robust in recognising a simple object such as a geometric shape rather than the complete, semantically rich topographical object.

### 3.2.2 Combination routines

Once a collection of primitives has been assembled, mechanisms need to be constructed that aid map production procedures. As seen in figure 1, the products of this research will either be systems that strive towards fully automated data capture, or semi-automated tools that aid the photogrammetric operator.

The first option for using the collection of primitives is to utilise knowledge-based information modelling methods. Similarly to [14, 16], this research is addressing the feasibility of assembling real-world topographic objects from the collection of primitives (in a bottom-up manner) in addition to also being used to constrain the search space to find those objects (in a top-down manner).

The second option is aimed at developing tools for human operators that provide them with information for quicker interpretation of large remotely sensed scenes. Based upon ideas of multi-conjunction search in intermediate vision [21], such tools may combine combinations of primitives. In this case, the aim is not to compose complete topographic objects, but to refine the search space that an operator needs to interpret when understanding the content of a scene. One way that this multi-conjunction search might work could be through using a region-based primitive in combination with one or two geometric primitives to direct the operator to an area that is flagged as requiring update. The specific types of primitives used to aid such a search will depend upon the type of mapping task being undertaken at the time.

To illustrate how such a tool might help an operator, the hypothetical example of updating the location of new buildings in an image is given. In this example, a region-based primitive to identify uniform textures of a certain direction and orientation might first be implemented. This then refines the search space for the application of algorithms that search for combinations of specific geometric shapes such as L-shaped and Y-shaped corners. The result of applying such a tool would then highlight regions in a large, complex image for an operator to concentrate their efforts.

Obviously, the information provided to the operator by such a tool needs to be targeted so that it does help the operator. This further emphasises the importance of on-going and planned work to both analyse the cognitive processes involved in updating maps from images, and undertaking detailed analyses of the effect upon an operator's tasks that implementing such tools will incur.

### 4 Case studies

This section outlines selected results from projects that have been completed to date. Detailed accounts of many of these projects are shortly to be submitted for publication in their own right. Work currently being undertaken in the Image Primitives project is aimed at producing the collection of primitives that can be used in either of the options discussed in section (3.2). The examples given in this section are methods for extracting discrete or region-based primitives for this collection.

It should be noted that the input data used in all of the investigations outlined in this section is scanned colour aerial photography. Since many computer vision algorithms are based upon the analysis of single intensity (digital number) values, these colour images are often converted to greyscale for the purposes of this research.

#### 4.1 Geometric primitives

Investigations outlined in this section are designed to identify the geometry and location of discrete geometric features of a particular type. Examples of geometric primitives to be extracted in this section of the programme, are different types of corners, polygons and lines.

##### 4.1.1 Polygons of constant interior and few corners

It was the aim of this investigation, undertaken for the Ordnance Survey by [17], to identify polygon shapes in the landscape that were most likely to be rectangular man-made topographic objects such as car-parks and buildings.

![Figure 2: The order of algorithms employed to recognise polygons of constant interior and not too many corners. After [17].](image)

Many man-made objects in the landscape can be assumed to be made up of rectangles and to often be uniformly composed of a single man-made material. For example a car park will usually be covered entirely in tarmac, with little or no texture variation across its expanse. Consequently, research has been undertaken in which polygons were identified in aerial images based upon a grouping of pixels that had constant intensity values across their interior, and just over four corners.
The procedure (as outlined in [17]) for identifying these polygons is illustrated in figure 2, while the results after applying each of the steps in the procedure are depicted in figure 3. These steps involve identifying edges in the image, then through the morphological processes of erosion and opening, regions that have been defined by the edges are shrunk and further delineated through the removal of spurious edges. Whilst the initial steps in the procedure were designed to separate regions using consistent defined edges, regions then needed to be labelled so that each group of pixels could be analysed independently. A size filter was then applied to identify regions that were not too small or too large. A dilation operation is then used to enlarge regions to their estimated original size before the erosion in the second step of the procedure.

Finally, [17] describes a rectangular fitting method that is used to select the final desired polygons from the image. A number of rectangular fitting techniques were experimented with, and the technique chosen utilised a particular corner finding algorithm. This corner finding algorithm is based upon an iterative linear segment splitting technique, where corners are either kept or rejected depending upon a given distance threshold between them. The distance threshold controls the sensitivity of the algorithm and as a result the number of corners that are accepted. Corners are further accepted or rejected through a recursive technique that tries to identify three corners that have a right angle between them. Once those three corners have been identified, since the algorithm is trying to fit a rectangle, a forth corner will then be searched for that fits with similar angle and distance constraints to the initial three corners.

Upon applying this procedure to one of the test images, the best set of results (in terms of false positives and false negatives) for that particular image can be seen in images (e) and (f) of figure 3. Image (e) illustrates the results in relation to existing topographic map data, where it can be seen that some of the polygons matched with objects in the map. However, polygons were also recognised that did not match with existing map data. This can be shown clearly in the image (f) where:

- Green rectangles are expected rectangles that have been manually generated from existing map data, but have not been recognised (false negatives).
- Yellow rectangles are polygons recognised using the described procedure, which match with those that are expected.
- Red rectangles are those polygons that have been recognised that do not match with those that are expected (false positives).

Many of the false positives can be accounted for when referring back to image (a) of figure 3, where they relate to shadows being cast by buildings in the image.

Although none of the tests produced matches for all those polygons that were expected, iterations of the parameters used in each of the stages of the procedure produced stable results in an empirical manner. In addition although shadows in particular result in false positives, these are not the fault of the low-level algorithm employed here, since they are still regions of uniform intensity relative to their surroundings.

![Figure 3: Exemplar results after applying procedure outlined in figure 2 to one of three test images. After [17]. Ordnance Survey data © Crown Copyright 2003](image)
been completed so far is a simple procedure that could form the basis of more sophisticated routines.

4.1.2 Corners of different geometrical characteristics

Given the importance of different types of corners and junctions between linear objects to visual perception, a project is currently under way that is addressing how such features can be extracted from aerial photography.

- L Junctions: Intersection of segment pairs must be outside at least one of the segments. However, this distance should not be more than an outer threshold. Whilst when inside a segment, distance should not be less than an inner threshold.
- X Crossings: Intersection of segment pairs must be inside both segments and more than an inner threshold.
- T Junctions: Intersection of segment pairs must be inside both segments but less than an inner threshold.

Further work that assesses the implications of applying alternative tolerances to select the corners, is required. Even so, it is clear that the techniques developed here allow relatively easy selection of discrete objects of defined geometry that can be used in further intermediate-level recognition tasks.

4.2 Region-based primitives

Unlike those techniques described in section (4.1) that were aimed at identifying the locations and geometries of discrete geometric objects, the aim of the region-based primitive work is to identify areas of generic perceptual characteristics across an image. For example, areas might be delineated according to an overall colour, or texture, across a particular area within an image. As asserted earlier in this paper, the advantage of utilising region based methods within image interpretation, is that they help constrain the search space for either the human operator or an automated recognition technique.

4.2.1 Regions of homogeneous texture

Figure 4 illustrates the results from an on-going research project that aims to identify intermediate geometric objects constructed from line segments. Line segments are first constructed from the aerial image using techniques based upon connectivity weighted Hough and Rotation transforms [10]. Following this, junctions are identified through intersections between segment pairs using the following rules (after [9]) based upon angles and crossing-points between those segments:

- L Junctions: Intersection of segment pairs must be outside at least one of the segments. However, this distance should not be more than an outer threshold. Whilst when inside a segment, distance should not be less than an inner threshold.
- X Crossings: Intersection of segment pairs must be inside both segments and more than an inner threshold.
- T Junctions: Intersection of segment pairs must be inside both segments but less than an inner threshold.

Figure 5: Schematic diagram of the feature recognition and feature selection produced to produce texture clusters for analysis in this investigation.

The important role that texture perception plays in human vision is explained in [15]. With this in mind, the purpose of this case-study is to assess how regions of homogeneous texture could be best differentiated within an image in an
unsupervised manner. In particular, it was wished that procedures were robust to the size of the image dataset provided (i.e. scale of the dataset) and the spatial location of textures to each other.

Figure 5 shows the three different procedures for producing texture clusters that were tested in this work. Each procedure applied a method for extracting textures the available textures present in the image, followed by a method for selecting and grouping those textures into clusters. So while Gabor filters [6] and Gaussian Markov Random Fields (GMRF) [7], are methods for extracting texture features; K-means and Mean-shift [8] are feature selection routines.

<table>
<thead>
<tr>
<th>Algorithm Combination</th>
<th>Orientation Variance</th>
<th>Scale Variance</th>
<th>Average $S_{Dbw}$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMRF and k-means</td>
<td>0.00832</td>
<td>0.00772</td>
<td>1.610</td>
</tr>
<tr>
<td>GMRF and mean shift</td>
<td>0.0084</td>
<td>0.127</td>
<td>1.260</td>
</tr>
<tr>
<td>Gabor filters and k-means</td>
<td>0.0098</td>
<td>0.0097</td>
<td>1.109</td>
</tr>
</tbody>
</table>

Table 1: $S_{Dbw}$ variance for scale and orientation, and Average $S_{Dbw}$ value for each algorithm.

To test the robustness of these procedures relative to the scale of the dataset, four sizes of image subset were used that ranged from 175x175 pixels to 257x257 pixels. To assess robustness to the spatial location of textures in an image, this has been tested in these initial experiments by rotating the image, in this case to rotations of $0^\circ$, $13^\circ$, $72^\circ$ and $132^\circ$. Without significant modification, the algorithms employed here are too computational expensive to process images larger than 256x256 pixels. Future investigations will need to focus upon optimising these techniques for ever larger remotely sensed images, in addition to employing increasingly sophisticated methodologies for assessing robustness relative to spatial location of textures.

A cluster validity index known as $S_{Dbw}$ [11] is used here to assess the differentiation between the texture clusters in terms of their compactness and the inter-cluster density between them. Table 1 shows the variance in the average $S_{Dbw}$ index values across the scales and orientations of images used in this study. Since the variance of the validity index in all cases is very small, it suggests that all the texture models are robust relative to scale of the dataset and spatial location of those textures.

It is further shown in table 1, in the average $S_{Dbw}$ validity values, that the combination of Gabor filters followed by k-means clustering produces the most defined texture regions from these images. It is also helpful to note after studying figure 6, that in addition to producing the mathematically most separated clusters, the texture model of Gabor filters followed by k-means produces the visually 'cleanest' regions of texture too.

4.2.2 Regions of high visual saliency

The purpose of this case-study is to illustrate how regions in an image can be identified using a computational modelling technique, that can be inferred to be 'attractive' to a human first viewing that image. Further research may then apply a similar computational model as this within a software tool that presents information to a user that is pertinent to their tasks. Accordingly this is an example of a project that has used a practical cognitive study to compare the results of applying a computational model to isolate primitive information in an image, with observed characteristics of human interpretation.

As explained in [20], this research has applied the widely used Itti and Koch saliency-based attention model [12,13], to aerial photography. Such models of visual attention tend to focus on the initial time course of gaze control immediately after an image is presented - typically the first few hundred milliseconds. Consequently the Itti and Koch model is being tested against the results of an investigation that has used human volunteers to study the effect of visual saliency across a time course, in relation to expertise and the effect of top-down influences.

An experiment was carried out that compared the results of human interpreters to the results of the computational model. A technique from cognitive science known as the 'letter probe'
was used to investigate visual saliency in the participants. In this technique, a grid of numbers is flashed very briefly (up to 100ms) onto the screen immediately after the image has disappeared, and participants needed to name the numbers they saw. To assess the contribution of prior (top-down) knowledge to this task, the images used were aerial photographs categorised according to the landscapes of agriculture, wilderness, residential (suburban), city centre (urban) and industry.

Figure 7: A saliency map produced by the Itti and Koch [12] computational saliency model. Ordnance Survey data © Crown Copyright 2004.

Figure 7 illustrates a saliency map produced after applying the Itti and Koch model to an aerial photograph, where regions of high visual saliency are identified with white pixels. Generally speaking, the saliency model looks for pixels of high brightness compared to their surroundings, followed by an assessment of where junctions and changes in orientation appear in groups of pixels in the image. Accordingly it can be seen that bright white buildings especially, attract high saliency values.

Figure 8: Mean saliency of computational model in areas that humans found to be most salient in each landscape type.

To illustrate a comparison of Itti and Koch's model with the characteristics of the human participants, Figure 8 depicts the mean saliency of values given by the model, in the regions of the images (categorised by landscape type) that also attracted the participants' visual attention. It should be noted that there were statistically significant differences (p<0.0001) between the results of images that represented different landscape types. It can further be seen that the model was more reliable in areas where there were very well defined shapes relative to their surroundings, such as in industrial areas. Further analysis of the results indicated that as a general rule, the effect of saliency decreased over time, presumably due to the effect of top-down influences.

Although further studies are required using increasingly sophisticated methodologies (such as those that use eye-tracking equipment), it is apparent that this model in particular is biased towards very low-level factors such as variations in radiance. Similar computational models now need to be developed that better take into account intermediate-level features such as geometrical objects.

5 Conclusion

This paper has outlined on-going research into improving the efficiency of image interpretation procedures for the update of topographic maps. Although a strategy is employed here that is similar to other approaches taken to research into automatic object recognition, the approach taken here differs to many in its appreciation of the need to use computer vision to design tools that help rather than replace highly skilled human operators.

An account has been provided of projects that utilise computer vision techniques to recognise lower-level primitive information in aerial photography. Especially in the case of discrete geometric primitives, the parameters required can be very sensitive to variations in the different landscapes that will be depicted in each image that a particular algorithm is run upon.

Ultimately, when designing tools for operators, it may be desirable to use computer vision to isolate information that a human is not very well suited to extracting themselves. However, it is asserted that the state-of-the art of computer vision, and visual attention in particular, is still at the level of mimicking human perception. As a result this further iterates the requirement for algorithms to be designed on the basis of improved understanding of the cognitive aspects and specialist user tasks of photogrammetric operators.

Nevertheless it is intended that this line of research will enable a mapping agency to fully embrace that advantages that computer vision can bring in a flexible and effective manner.

Acknowledgements

The author is grateful for the contributions of researchers not based at Ordnance Survey, these include: Panja Sae-Ui for his work on consistent interior polygons whilst being sponsored by Ordnance Survey. Paolo Gamba, Fabio Dell’Acqua and Gianni Lisini for their work on extracting corners as part of an Ordnance Survey research contract. Also to Padraig Corcoran for undertaking the work on texture whilst occupying a visiting scientist position at Ordnance...
Survey. Finally to Nick Donnelly, Kyle Cave, Helen Sockett, Luke Phillips and Natalie Lock for their assistance in the investigation on visual saliency.

Also to Clare Davies, Layla Gordon, David Holland, Isabel Sargent and Doreen Boyd at Ordnance Survey.

References


Disclaimer

This article has been prepared for information purposes only. It is not designed to constitute definitive advice on the topics covered and any reliance placed on the contents of this article is at the sole risk of the reader.