Path Integration: Issues concerning the integration of data suitable for the leisure user

Matthew Chapman, William Tompkinson, and Matthew Faulk

Abstract

With the release of a complete feature-based large-scale digital map in Great Britain, the ability to integrate supplementary data with existing topographic information can be realised. Through the application of data integration techniques within an ESRI environment, different scale data have been matched together. Problems experienced integrating a single linear path with corresponding multiple linear topographic features are discussed. The implementation of a non-standard GIS function for inferring visual prominence of features in the landscape is described. Issues relating to the incorporation of these results with integrated path data are presented with the inclusion of 3D objects as visual cues.

1. Introduction

With the increasing proliferation of spatially related information, there has been a growing need to integrate such data to facilitate more in-depth and efficient spatial querying than could be provided through individual data sources. This has led to the emergence of a variety of spatial techniques, which have been used to undertake data integration. These techniques have been classified by Cobb *et al.,* (1998) into a three-tier hierarchy whose categorisation is based on the types of data and the degree of semantic intelligence or knowledge required to undertake the processes, moving from basic data integration, to conflation, and finally fusion (see Figure 1).

Such techniques produce integrated data sets which are seen to represent 'the best of both worlds' through the combination of two data sets to produce a third that is more than the sum of its parts (Uitermark *et al.,* 2002). Typically data integration is undertaken to facilitate more in-depth and efficient spatial querying through the provision of enriched data models. Once data are integrated, the subsequent analysis is usually limited to the additional attribute features from the source data which have been added to the more accurate co-ordinate target data. There is an additional benefit, however, that can be gained by deriving additional information from the integrated data that could not be ascertained from either data set individually.

1.1 Leisure User

People undertaking outdoor pursuits such as walking or mountain biking use maps to guide them during their activity and aid them in planning the routes that they are going to take. A user has the ability to measure distance, assess difficulty of passage (in terms of slope gradient, changes in elevation and obstacles), and plan a route so that certain features in the landscape are visited. Currently they do not guide a routing decision based on likely visual prominence of landscapes and / or the 3D representations of objects.

In some situations a user may know of an area in the landscape that they wish to visit, for example to see a particular feature in the landscape. Consequently additional information that could help with this decision might give an indication of the degree to which certain features in the landscape can be seen, and therefore infer a measure of visual satisfaction.

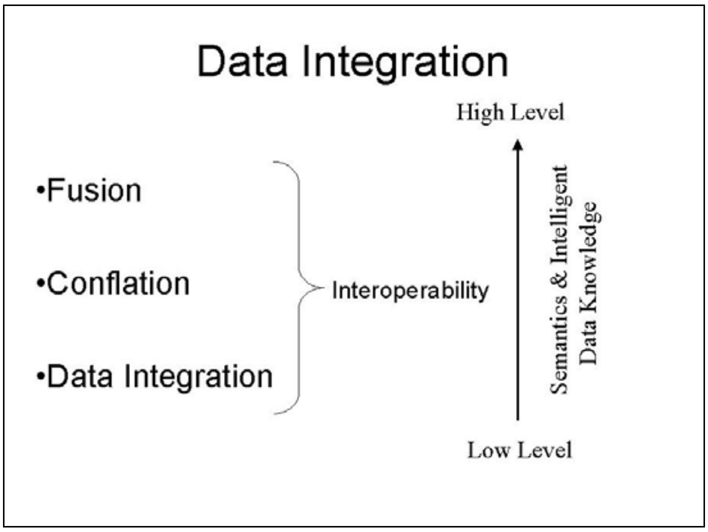

Figure 1. Data Integration Techniques (after Cobb *et al.,* 1998)

1.2 Approach

This paper outlines on-going research that applies the concept of data integration (after Cobb *et al.,* 1998) to combine 1:25,000 scale path data with large-scale topographic data set for a test area on the Isle of Wight (Great Britain). Following this, analysis is undertaken applying a perceptual area function (PAF) developed by Tompkinson (2001) for a variety of points along the integrated path network, together with a Digital Elevation Model (DEM). This quantified the degree of visual prominence of the landscape that could be viewed to provide ratings of views along particular routes to contribute to an enriched dataset for the leisure user. These results are then validated in the field. In addition, the further integration of 3D objects within the landscape to provide visual cues and improve on the available data for analysis are discussed.

2. Study Area

The Isle of Wight is a small diamond shaped island (ca. 381 km^2), situated a few kilometres off the south coast of England near the cities of Southampton and Portsmouth. It spans 21 kilometres from Cowes in the North to St Catherine's Point in the south, and 37 kilometres from the Needles in the west to Bembridge Foreland in the east (Edwards, 1994). For the purposes of this study, two paths were selected to undertake the analysis, one running along the southern coast from Tennyson Down to the Needles and the second running across the coast by Alum Bay (see Figure 2).

3. Data Integration

The concept of data integration defined by Cobb *et al.,* (1998) conveys the process whereby incompatibilities among varying spatial data formats are resolved to allow their simultaneous analysis, display and processing within a GIS. "The problem of geographic data set integration is defined as establishing relationships between corresponding object instances, considering the differences between geographic data sets to be integrated" (Uitermark *et al.,* 2002). Data integration (and conflation) brings simultaneous representations of the data together (usually vector) where one data set is considered to have a better or desired attribution (source data) whilst the other has a more accurate or preferred geometrical co-ordinate information (target data). Typically this relies on both corresponding geographical features, such as roads being present in either data set e.g. the U.S. Census Bureau TIGER data and Digital Line Graph (DLG) data from the U.S. Geological Survey (Dallal, 2000) or the European standard Geographic Data File (GDF) and the German topographic cartographic spatial database Amtliches Topographisch-Kartographisches Informationssystem (ATKIS) (Walter & Frisch 1999). In the case of the Isle of Wight path data , the large scale topographic data does not have path data to directly correspond with the 1:25000 scale path data. Currently, this feature is represented by a mixture of features including 'established' paths, tracks and path / field boundaries. Consequently, the 1:25,000 path representation passes between these large scale corresponding features.

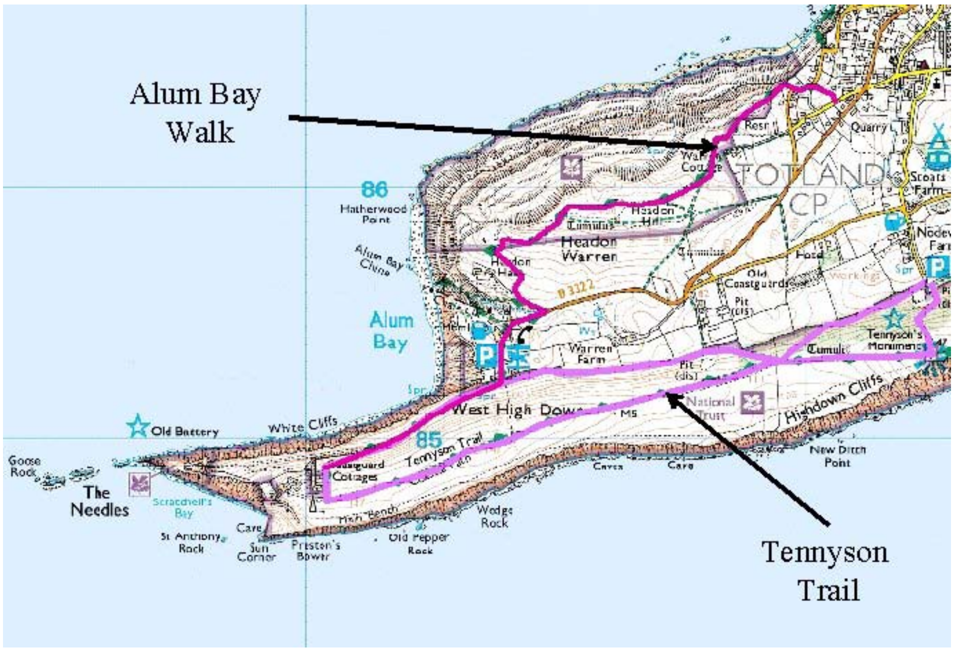

Figure 2. Location of the Isle of Wight paths used in this study (Source: 1:50 000 Scale Colour Raster. Reproduced from Ordnance Survey map data by permission of Ordnance Survey, © Crown copyright)

3.1 Path Integration

This approach of data integration is subdivided into three steps, which were undertaken in ArcGIS™ or through the execution of Arc Macro Language (AML)™ scripts to aid data integration in aligning the data.

The first step required the pre-processing and extraction of the representative target features from the large-scale topographic data. This was achieved through the use of standard query language on the large scale data, and the extraction of the integrated road network features. Initial examination of this data revealed in some instances that a few field boundaries surrounded the approximate 1:25,000 path network. However, a single feature to represent that path in the target data was obviously absent. This required the generation of a single centre-line path from the field boundaries. This was achieved through the use of generalisation techniques, although it is recognised that there are limitations to the current algorithms within ArcInfo™.

With the data in place the second step in the methodology required the addition of vertices to the 1:25,000 path data so that its total number of vertices matched those in the target data. Without this process the resulting data integration between the available vertices in the path data and the large-scale data would result in the path's arcs not completely aligning themselves, with the data matching only to the available vertices creating a ‘zigzag’ effect in the resulting integrated data in comparison to the target data.

The third step in the process integrated the data. This required the preferential order of the geometric elements to be set in the integration process to facilitate how the integrating path network should align itself to the nodes, segments and vertices in the target data. Following this, the data was integrated in a hierarchical order, firstly to any existing tracks or paths within the target data. If no matching feature was found in close proximity then the path network was aligned to alternative features, such as field boundaries. This was sequenced for each arc in the path network and resulted in an 1:25,000 path network integrated to large-scale topographic data (see Figure 3).

Figure 3. Integration of 1:25,000 path network to large-scale topographic data.

An examination of the results found that the use of the low-hierarchy concept of data integration had matched the data well, with the 1:25,000 path network integrating to large-scale topographic features. However, in a few exceptions it became clear that this approach was not able to integrate all the 1:25,000 path data. The main problem experienced was in the convergence of several field boundaries providing multiple options for the path data to match to (see figure 4). This is a limitation in the use of proximity to govern the integration of such data and the complications posed by the complex real world features on the ground. To overcome this problem would require the use of more semantic intelligence and a move to using conflation techniques to enable better decisions to be made about these instances. However, conflation techniques may still have problems in resolving such issues, which still rely on manual intervention to overcome un-resolved matches (Dallal, 2000).

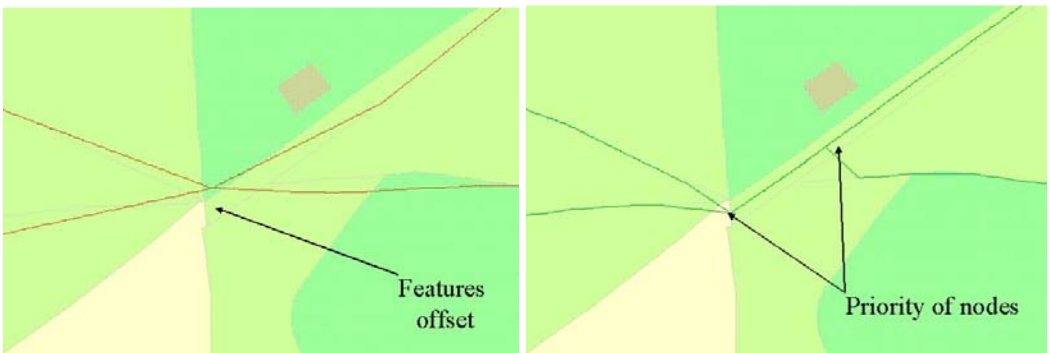

Figure 4. Data integration indecision with node priorities.

4. Perceptual Area Function

With the integration of the 1:25,000 path network with large-scale topographic data completed, the opportunity to derive additional spatial information was capitalised upon through the application of a Perceptual Area Function (PAF).

PAF is an enhancement of traditional visibility analysis, which is used to identify whether or not a particular feature in the landscape can be seen. Traditional visibility analysis does not, however, report the proportion of the area which can be seen within a person's field of view (FOV). The results of only taking into account whether a certain landscape feature can be seen, therefore, are thought to be the reason why for example, a photograph of a mountain range may appear to be disappointing. This results from the fact that the mountains in the distance may occupy a relatively low proportion of the FOV compared to other landscape features in the scene.

Consequently, PAF was designed in AML™ to provide a qualitative rating as to whether a particular landscape feature is visually prominent within a FOV using trigonometric algorithms. "PAF estimates the apparent area of patch or pixel in steradians ([where] a steradian is the solid angle version of the radian and is defined as an angle at the centre of a sphere subtended by a circle with the same radius as the sphere)", which calculates the angular area of particular patches on the landscape as a function of their distance, angle and orientation relative to the observer (Tompkinson & Wadsworth, 2002). Consequently, grid cells closer and facing the viewer will have a greater PAF value than those further and tilted away (see Figure 5).

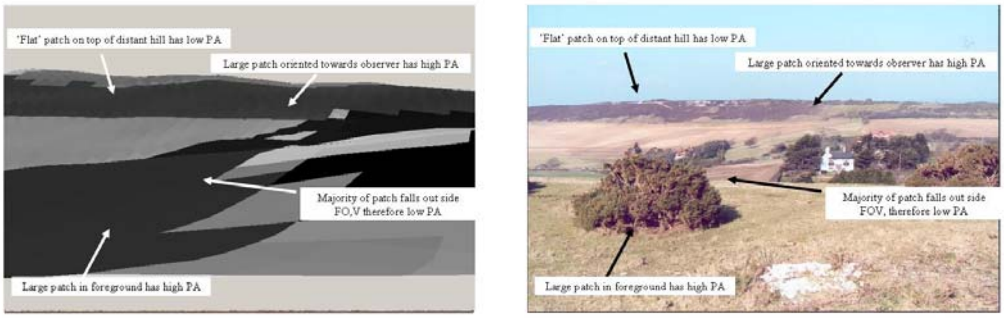

Figure 5. PAF analysis indicating areas of high (dark) and low (light) value.

Since PAF was originally written to process raster grid based data (from a DEM), these results were derived on a cell by cell basis. These results were then aggregated according to the boundaries of large-scale topographic features to highlight those that would be more visible within the landscape based on the mapping. PAF Analysis was undertaken for each of the two paths, Alum Bay and Tennyson Down, at 250m intervals to determine the visual prominence of features in the landscape in relation to a user walking those paths. These results were then integrated with the path data for each of the points assessed every 250m apart. The path geometry was then divided for each 250m tract and the PAF results were then associated from each point to the proceeding path tract in turn (see Figure 6). This would allow a better visual interpretation of the data by a leisure user, with the PAF score graded from light (lower PAF score) to dark grey (high PAF score). It is noted that there are some problems in representing the data in this manner, such as increased data volumes from splitting the path geometry. This would benefit from an improved network data model for storing such attribution. In addition the aggregation of the PAF values to a single node or link may benefit from a more fuzzy approach and requires further investigation.

Figure 6. Integrated PAF results for the Tennyson Trail

5. 3D Objects Introduction

During the course of this research it became apparent that whilst the integration of path data and the application of the Perceptual Area Function Score provided the user with a more informed spatial picture of the environment, a further refinement could be achieved with the aid of the integration of spatial data in the third dimension (3D). Of primary concern to this work was the ability to represent an object in 3D within a data model which will aid the decision-making of the user. This element of the research highlights the chosen method for the representation of a 3D object and points to further work required to enhance the overall model.

5.1 Enhancements to the visual appearance of the data

Users interested in the likely prominence of the view they can expect along the path network will also be interested in the overall view itself - to include objects sitting on the terrain surface. Hence, methods of modelling spatial objects which appear within this environment have been explored. Using the 2D objects contained within the Ordnance Survey large-scale topographic database, 3D objects have been generated. These objects have then been 'inserted' onto the terrain surface such that an enhanced visual scene is generated, with the aim of enhancing the integrated path data and PAF score results.

5.2 Methodology

In achieving the final results indicated in Figure 10, both the terrain and the 3D objects were generated separately and then using large-scale topographic data as the template, the 3D object was 'inserted' onto the terrain surface represented as a Digital Terrain Model (DTM). The procedure for achieving this is explained below.

5.2.1 3D Object generation

A photogrammetry package was used to capture and then model prominent topographic objects in 3D from terrestrial photographs. With the camera calibrated (focal length, image size / number of pixels, etc.), a series of photographs were taken around the object. The criteria being that a number of common points must be visible in at least two of the photographs. Using the point, line and edge tools, points common to both (or more) photographs are then referenced in each of the images highlighted in Figure 7 below.

Figure 7: Mark and reference of common points for Tennyson's Monument, Isle of Wight.

This was repeated for each face of the monument so that a full 360° coverage of the object is achieved. Four photographs were taken to achieve this and the location for each relative to the monument is highlighted in Figure 8 below.

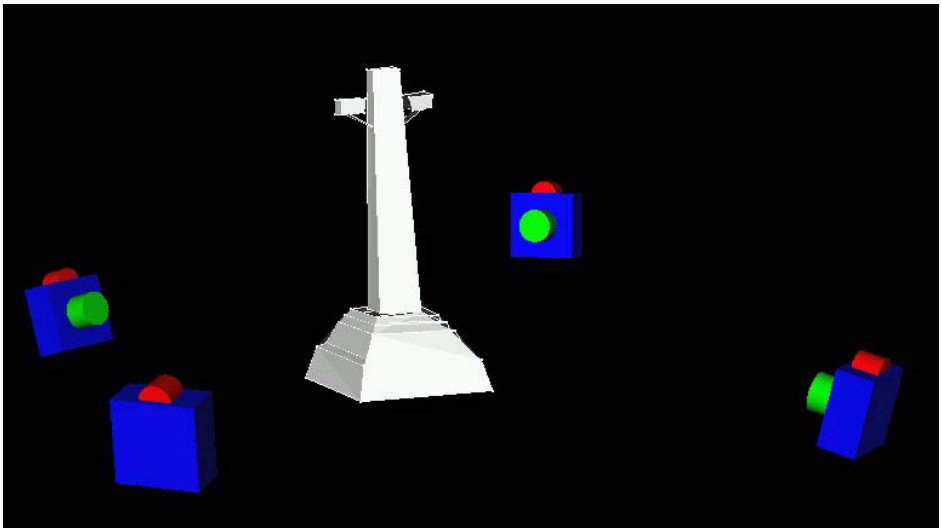

Figure 8: Location of the camera for each of the four photographs used in the modelling of Tennyson's Monument.

With the referencing of all points and lines for each of the four sides of the monument completed, a 3D wireframe model is generated. Specific faces can then be assigned to the model using three defined points. The wireframe and the transition to the full-faced 3D model is highlighted in Figure 9 below.

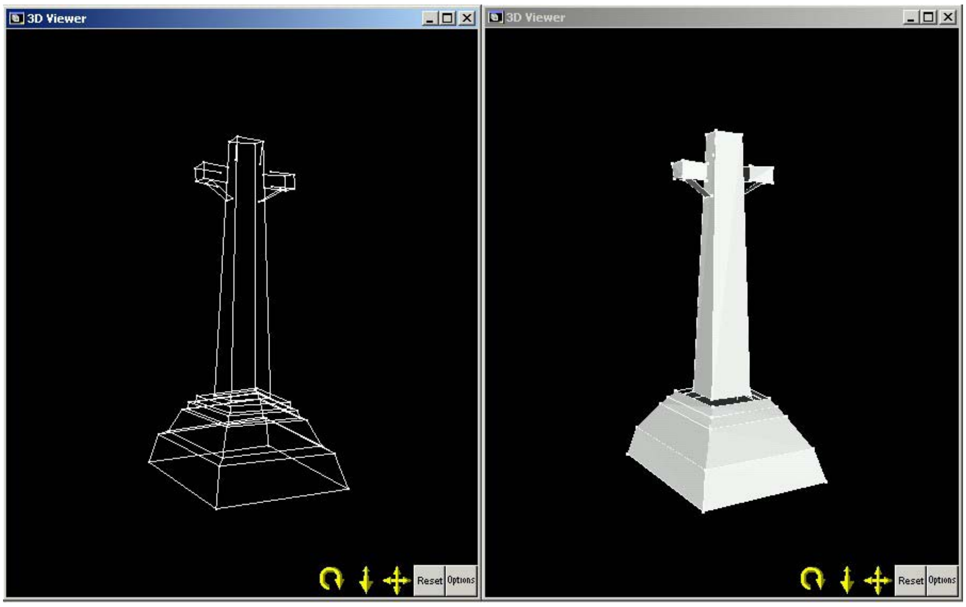

Figure 9: 3D Wireframe (left) and full surface model (right) for Tennyson's Monument, Isle of Wight.

It is possible to generate photo-textures by specifying three points in the model and applying the relevant photograph to that face. However, due to data volume requirements and to the fact that we were only attempting to enhance and not replicate the view, it was felt that rendering photo-textures was not required.

5.2.2 3D Object 'Insertion'

The key element for the purpose of visualisation is ensuring that the 3D object is located in sympathy with the terrain surface. This is achieved by specifying real world coordinates for the 3D object generated (to include z or height values). As stated at the outset, the chosen objects are ones which have been mapped and are present in the large-scale topographic database. For Tennyson Monument, in 2D, this is the base polygon of the structure itself. By stating which of the four corners of the base polygon represent the four corners of the monument, it is possible to orientate the object correctly. By specifying their height, it is also possible to locate them on the terrain surface. Additional tools within the photogrammetry package allow for the correct translation and scaling of the object so that it is orientated correctly to the scale origin and has z values which are to scale.

5.4 3D Object Results

It has already been stated that in order to keep data volumes to a minimum, coupled with the irrelevance of photo textures to the PAF score, the 3D object is restricted to points, lines and faces (polygons). Specifically, the model is exported into ArcScene™ as a series of 3D Points, Lines, Edges, Surfaces, Curves and Cylinders in 3D DXF file format. Also, whilst obviously designed to assist the user on the scene they are likely to encounter, one does not wish to provide a full "VR" scene and thus spoil the enjoyment of undertaking the walk in the first place!

In applying the methodology stated above, it is possible to generate a 3D object which is not only in sympathy with an underlying terrain model but is also inserted onto the underlying 2D topographic polygon. The results can be seen in Figure 10 below. Although at present PAF has been developed to process raster data; if PAF is rewritten to accommodate 3D vector data structures as part of further work, it is hoped that users will have a more-informed view of the likely vistas they will experience as a result of combining the datasets of: path data, PAF data and 3D objects. Results from PAF using a potential completely 3D dataset would be more realistic since they could better incorporate the effect of obstructing surface structures that might be in the foreground of a given view.

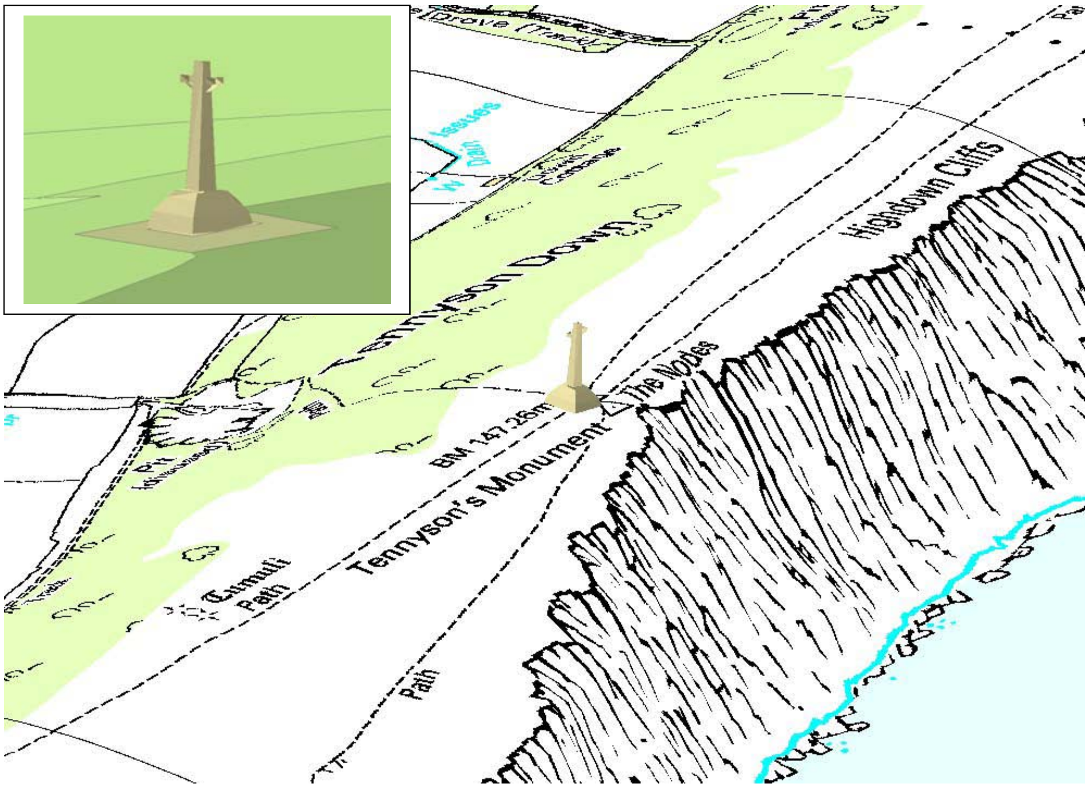

Figure 10: Final output indicating how the 3D object has been inserted onto the underlying 2D topographic polygon (inset) and subsequently onto the terrain surface (main image). (Source: 1:10 000 Scale Raster & LandForm PROFILE®. Reproduced from Ordnance Survey map data by permission of Ordnance Survey, © Crown copyright)

6. Conclusion

This approach has demonstrated how two different scale data sets can be matched using data integration techniques, however, it has been noted that further work is required to resolve mismatches between the source and target data. This could be improved through the implementation of more semantic information and knowledge with conflation techniques and the use of software such as the ArcGIS™ extension MapMerge™ produced by ESEA. These issues may still prove difficult to resolve, however, owing to differences between the data they could nonetheless require manual intervention. Following the data integration of two differently scaled data sets the ability to benefit from their alignment to derive additional information that neither data set could provide was capitalised upon through the implementation of PAF. This analysis has the potential to enrich the data model and further, with 3D objects, to assist users in making more informed decisions regarding the routes they take to ensure they visit and can view landscape features. However, further investigation is required to determine the optimum way the PAF data is conveyed to the leisure user.

7 Acknowledgements:

The authors are grateful to Research & Innovation at Ordnance Survey for giving us the opportunity to undertake this "shed" innovation project and for the assistance of Peter Bunting of Ordnance Survey in this work.

8 References:

Cobb, M. A., Petry, F. E. & Shaw, K. B. (1998) Uncertainty issues of conflation in a distributed environment. GIS/LIS'98, 9-12 November 1998.

Dallal, S. L. (2000) Conflation of TIGER with DLG. Available at http://www01.giscafe.com/technical_papers/Papers/paper035/ (Accessed 10/01/03)

Edwards, A-M (1994) Isle of Wight Walks. Ordnance Survey Pathfinder Guide. Ordnance Survey and Jarrold Publishing, UK.

Tompkinson, M.W.J (2001) From userability to noticeability. Unpublished MSc thesis. University of Nottingham.

Tompkinson, W., & Wadsworth, R. (2002) From visibility to Noticeability: An investigation into the feasibility of calculating perceptual areas and identifying visually prominent parts of the landscape. GIS Research UK (GISRUK), 10th Annual Conference, Sheffield, United Kingdom, 3-5 April 2002, Session 6A.

Uitermark, H., van Oosterom, P, Mars, N. & Molenaar, M. (2002) Ontology-based geographic data set integration. 5th AGILE conference on geographic information science, Balearic Islands, Spain, 25-27 April 2002.

Walter, V. & Fritsch, D. (1999) Matching spatial data sets: a statistical approach. International Journal of Geographical Information Science, 13, 5, 445-473.

Author Information:

Matthew Chapman, William Tompkinson & Matthew Faulk, Research Scientists, Research & Innovation, Ordnance Survey, Romsey Road, Southampton, United Kingdom. SO16 4GU.

Phone: (+44) (0)23 8030 5014
Fax: (+44) (0)23 8030 5072
Email: mchapman@ordsvy.gov.uk