Environmental Analysis through integration of Geographical Information and Machine Vision systems

by

Paul D. Kelly, M.Eng.

A thesis presented on application for the degree of

DOCTOR OF PHILOSOPHY

Faculty of Engineering

The Queen's University of Belfast

School of Electrical & Electronic Engineering,

May 2004.

Chapter 3

Geographic Data Sources

Accurate spatial information (regarding the position of the camera, position of objects in the scene etc.) is essential for geo-referencing data so that it may be used as a baseline for change detection. Operating in an unstructured environment poses a problem because of the difficulty in determining the correspondence between pixels in an image and their 3-D world co-ordinates. Information on the heights of the objects in the scene is necessary, and a Geographical Information System is an obvious source for this information.

The majority of the GIS work described here could be classified as geographical information science (rather than the application of a geographical information system to a specific problem) and an attempt is made to make full use of the advanced analysis capabilities that GIS offers. Indeed GIS is also useful for actually performing the change detection (see Chapter 6), in addition to processing and geo-referencing input data (which will be discussed in this chapter).

3.1 Choice of GIS

The requirements for a GIS are as follows:

- Good data analysis capabilities; in order to make maximum use of all available information, data from various sources must be fused and verified in the GIS.
- Possible to extend the functionality of the GIS and add further modules to it; e.g. integration with an image processing system in this work.
- Able to handle three-dimensional data (although considering the elevation of an object as an attribute of its 2-D position, i.e. a 2.5-D representation, may be adequate)

To have the software easy to extend means it must be possible for the end-user to change the functionality, and for this it is necessary for the source code to be available. Otherwise change requests would have to be made to the vendor; this would be inconvenient and cause a delay and would not be possible unless paying for software support.

The FreeGIS Project (FreeGIS Project, 2003) provides information on many GIS-related software that are free and open-source. These range from simple data viewer applications to complex analysis tools for geographical data. GRASS (Geographic Resources Analysis Support System) (Neteler and Mitasova, 2002) was selected because of its overall competence in all the areas required. Its capabilities include well-developed 3-D visualisation (Mitasova et al., 1995) and advanced data analysis functions.

Using free (in terms of cost) software also adds value to the system, as it makes it easier for others to re-use and extend the work.

3.2 Co-ordinate Reference Systems

Throughout this work the existence of a three-dimensional Euclidean space will be assumed and referred to as the 3-D world space. However the Earth is an ellipsoidal shape; locations on its surface may be specified by the latitude and longitude angles of a line from the centre of the Earth to the point on the surface.

2-D location on Earth's surface To work with three-dimensional data, the angular ellipsoid position must be *projected* into a conventional Euclidean space. Projections are described in detail by Snyder (1987), for example.

A *datum* defines the model of the Earth that the latitude and longitude angles refer to, and a *projection* and *co-ordinate system* define the mathematical relationship between the latitude / longitude co-ordinates and the projected co-ordinates.

Conventionally in Northern Ireland the 'Ireland 1965' datum is used, which is based on the 'Modified Airy' ellipsoid model. The conventional co-ordinate system is the 'Irish Grid' (OSi and OSNI, 2000), which is based on the 'Transverse Mercator' projection.

Height The Irish Grid is a two-dimensional system and the position of a point on the Earth's surface is given in terms of its *easting* and *northing*. For the third dimension, conventionally orthometric height is used (i.e. height above sea level, in this case Belfast Lough Mean Sea Level).

This system can be used to uniquely identify any point in Ireland. However other datums exist in the world, based on different map projections and datums, e.g. the World Geodetic System 1984 (WGS84). Knowledge of the datum and co-ordinate system the data is referenced to is necessary when using data referenced to a system other than Irish Grid / Ireland 1965. This information will be used in Section 3.4.

3.3 GIS Data Sources

To generalise the system and enable it to be tested and used in a variety of different environments, a general source of topographic and elevation data that covers a wide area is required. The publicly-available data source maintained by the Ordnance Survey of Northern Ireland (OSNI) satisifies these requirements and was investigated as to its suitability.

Two-dimensional vector topographic data have been maintained for many years and are generally well-surveyed, of good quality and available for all areas (Perkins and Parry, 1996). The resolution of the data may be derived as follows; it is determined by the smallest element that can be discerned. The ratio of these two quantities is the scale of the image:

$$\text{resolution} = \text{smallest image element}/\text{scale}$$

For example, for a paper map at 1:2500 scale, the thinnest line that can be discerned on a map drawn on paper is 0.5 mm; the resolution is therefore the real-world feature size that this width represents and is given by

$$\begin{aligned} \text{resolution} &= \frac{0.5 \times 10^{-3}}{1/2500} \\ &= 1.25 \text{ metres} \end{aligned}$$

For an urban area map of 1:1250 scale, the resolution will be twice as fine, i.e. 62.5 cm.

Aerial photographs and satellite images may also be considered to be sources of GIS data in this context; see Section 3.3.2.

OSNI provides several elevation datasets, but they are not an integral part of the main vector product and it will be shown in Section 3.4.3 how they are even incompatible with each other. Producing an accurate elevation dataset is a substantial problem and will be discussed in section 3.4.

3.3.1 Import of OSNI vector data

Internally, OSNI stores the vector data in the proprietary Norwegian DST format (Evans, 2002). Sequentially-numbered tiles contain the data for all Northern Ireland; in rural areas the tiles cover an area of 1200×800 m at 1:2500 scale; in urban areas each tile covers a quarter of this area at 1:1250 scale. It is made available in several industry-standard formats, e.g. Ordnance Survey of Great Britain NTF (National Transfer Format), AutoCAD DXF and ESRI ArcView 'Shapefile'. GRASS has import routines for all these formats; a decision on which one to request the data in was made after the following considerations:

DXF Import (GRASS module v.in.dxf) Each layer in the DXF file, e.g. A Roads, B Roads, Houses, Commercial Buildings, High water line, etc. is represented as a separate vector layer in GRASS. Layers with additional attributes (spot heights and benchmarks, house numbers, road names etc.) also have an 'attributes' file. Where two or more layers co-incide, e.g. a shared boundary, the information (typically a line) is duplicated in the layers concerned.

NTF Import (GRASS module m.in.ntf) The whole NTF file is represented in GRASS as just one vector layer containing every type of feature as lines. Every line is numbered sequentially; e.g. in the sample tile provided there are over 3000. These sequential numbers are all contained in the attributes file for the vector layer. The categories file distinguishes one type of feature from another, with a text description of the type of feature (e.g. ROAD_A, TOWNLAND) corresponding to each attribute number. The attributes (e.g. height values or road names) are not stored anywhere, and it may be a problem with the GRASS NTF import program; perhaps it only works properly with OSGB NTF data.

Shapefile Import (GRASS module v.in.shape) Shapefiles are imported as one vector file, similar to NTF, and individual layers (roads etc.) must be extracted individually before use. As well as the shapefile containing the same line data as the DXF and NTF files, for each tile OSNI also supplies a second shapefile with the lines converted to polygons (i.e. area boundaries) and the areas labelled according to the types of land use. This may be useful for some purposes but several errors (areas incorrectly labelled or unlabelled) were found close to the boundaries of the tiles, and manual editing would be required to correct the dataset and make it consistent across tile boundaries.

The shapefile format also has the facility to store multiple attributes for each vector object in a database-style DBF file, but this capability is not relevant in GRASS 5.0/5.3 (as used here), which can only store one attribute per vector object.

When this work was being performed (early 2002), the Shapefile import routine (v.in.shape) in GRASS was not very reliable (it has since been replaced with an improved version). The NTF import routine appears to be less than adequate, so the DXF data is a lot easier to work with for our purposes. DXF is also a more universal format and will be able to be viewed and edited with other software if necessary. Technical information and scripts for importing the data into GRASS are given in Appendix B.1.

Further investigation of the data imported by v.in.dxf revealed that the attribute labels had not been imported correctly into the GRASS vector data structure, and some of them would be deleted or ignored by GRASS. Because of the modular and diverse structure of GRASS, it is quite common to find individual modules that do not work correctly.

A workaround for the problem was devised, using an additional processing step involving a script written in the Perl programming language. Details of this are given in Appendix B.1. On occasion, the script has been contributed, via the GRASS mailing list, to other GRASS users who have been inconvenienced by the same problem.

Post-processing of vector data

The primary purpose of the vector data is to find the two-dimensional real world location (easting and northing) of landmarks in an image, which can be identified both through manual interpretation and automatically (by image processing). When combined with a source of elevation data (the third dimension)—Section 3.4—this location information will be used for camera pose estimation.

To identify the correspondence between images and the vector GIS data, the various topographical features contained in different vector layers may be merged. For example, in a countryside image field boundaries may be easily identified as such, however they may be stored in the vector data as 'Perimeter', 'Hedge', 'Fence', etc. and these layers should be combined to make identification easier. In addition, the area visible in an image may cover two or more map tiles and the same layers from all the tiles in the study area should be merged. This was done using standard GRASS functionality (v.patch).

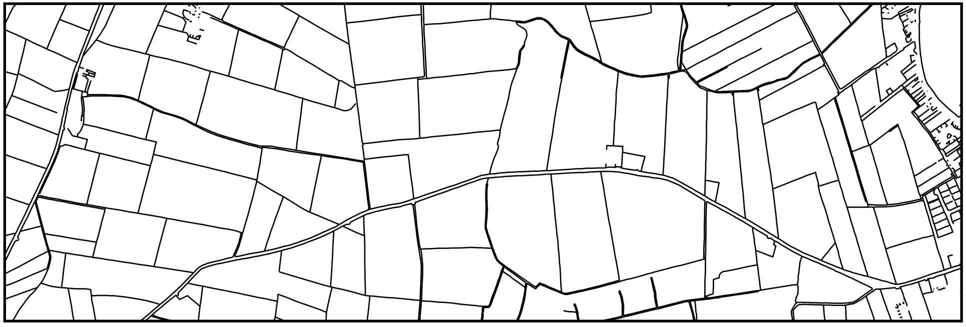
Figure 3.1 shows how this has reduced the amount of detail so as to contain only relevant features for landmark identification in the vector file, and how the data in two adjacent map tiles has been seamlessly merged ('fences' does not have a vertical tile boundary in the centre of the image).

3.3.2 Import of aerial photographs & satellite images

Aerial photographs (and remote sensing satellite images) in unprocessed form may be classified as either oblique or vertical. When obtained in raw format they may require ortho-rectification to form a plan-view (discussed in Section 1.5.5); this can be done using standard GIS functionality if a digital elevation model is available and the camera interior orientation parameters are known.

If the parameters are unknown and the image covers a small area of relatively flat ground it is possible to mark ground control points and perform an affine transformation. Again this is a standard GIS operation and the object of either of these two methods is to *geo-reference* the image so that it corresponds with the GIS co-ordinate system and can be referenced in the same way as a map or any GIS layer.

(a) All vector layers



(b) ‘fences’ layer

Figure 3.1: Selection of vector layers representing specific landmark types

Aerial photographs may also be supplied in already-geo-referenced format. The key point is that data in this format may be treated the same as any other geo-referenced raster layer in a GIS after some initial processing.

3.4 Creation of an Accurate High-Resolution Elevation Model

As mentioned in Section 3.3, two-dimensional GIS data is generally easily available and of good quality. Third-dimension elevation data is not available as universally and various sources of it have been investigated.

Relative Error Measurement Throughout this section errors and inconsistencies in the data sources presented will be described subjectively or relative to each other. Unfortunately, as it will be shown that there is no true correct data source available for the test areas, it is not possible to objectively quantify the error. Ultimately, using a GPS-based reference is probably the best solution to this problem, but it was not possible to take this any further in the time-frame available and is beyond the scope of this work.

3.4.1 OSNI elevation data sources

The first source to be investigated was the OSNI, which provides elevation data in the form of orthometric height (above mean sea level), referenced to the Irish Grid. This should be an advantage, in that no additional geo-referencing will be necessary for compatibility with the vector data. The approach followed is to attempt to combine data from the three primary sources of digital elevation data from OSNI:

1. Spot heights and benchmarks from Large Scale Vector Plans (1:1250 or 1:2500 scale)
2. 1:50000 contours (in vector format)
3. 50 metre resolution (Raster) Digital Terrain Data

Spot heights and bench mark elevation values are extracted from the large scale digital vector data. (In DXF format, they are in the CONTROL_TE layer of each map tile.) The elevation values are not actually attached as attributes to the points marking the measured heights, but to text boxes positioned near each point. This slight loss of positional accuracy will be discussed later.

1:50000 scale contours are obtained in DXF format and sampled at vertices (or 50 m intervals on straight sections), to give another series of sampled points. Sampling of the contours is done using the GRASS `v.to.sites` command.

The examples given here are for a test region comprising OSNI map sheet 242-13 (north of St. John's Point and west of Killough in Co. Down). This region contains a wide variation in elevation values ranging from the sea and a low-lying bog to a steep hill. This variety means there are a lot of contours. An A-class road, along which spot heights have been observed at regular intervals, passes through the region.

The map in Figure 3.2 shows the distribution of the data points in the test region.

3.4.2 Creation of gridded elevation model

The data sources have been sampled in such away that they are all 'sites data', i.e. consist of a list of sites defined by an easting and northing, with an associated elevation attribute. Each site may have multiple attributes, such as a parameter describing the accuracy of the elevation measurement, which is useful in advanced interpolation techniques (as in section 3.4.7).

The data sources were converted to sites format with a view to interpolating a surface, i.e. creating a raster layer of any arbitrary cell resolution based on the values of the scattered data points. The OSNI DTM data is already in this format, but the resolution of 50 m is too low when applied to interpretation of ground-level images—for areas near the camera the ground resolution visible in the image is very high.

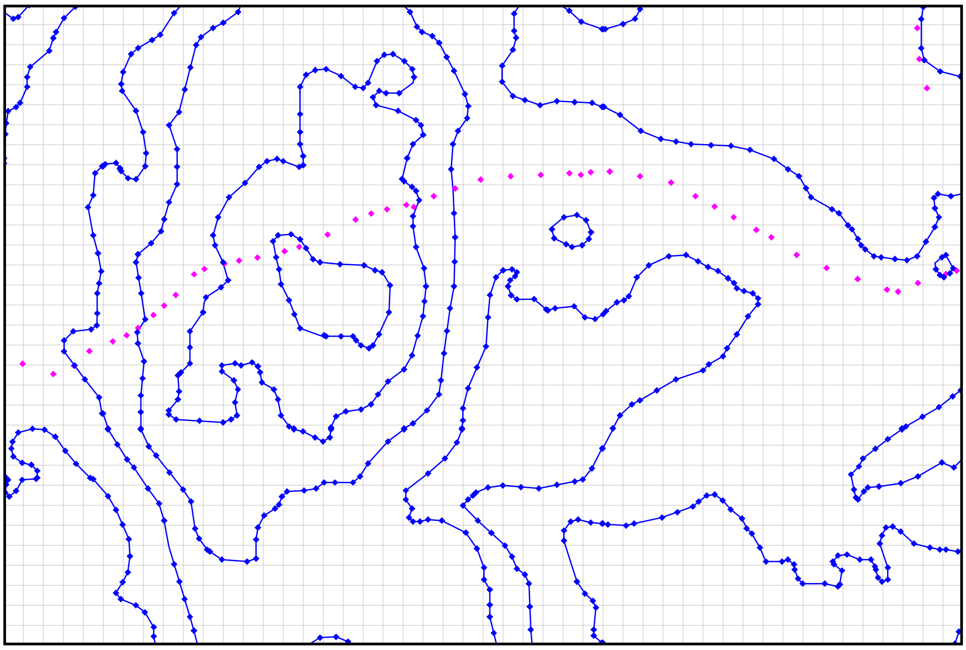


Figure 3.2: Contours and sampled points (blue) and spot heights (purple), for the St John's Point test region. Overlaid on grid showing resolution of OSNI raster DTM data

Interpolation of a raster DEM from the point data available, i.e. spot heights and sampled contour points, was attempted. The rest of this section describes the results of using different GIS interpolation algorithms for this.

Interpolation by Inverse Distance Weighting

This algorithm calculated the elevation value for each raster cell based on the values of and distances to the 12 nearest points in the sites layers. The interpolated output value $f(\mathbf{c}_o)$ at a raster cell $\mathbf{c}(x, y)$ is given by

$$f(\mathbf{c}_o) = \frac{\displaystyle\sum_{i=1}^{n} \frac{f(\mathbf{p}_i)}{|\mathbf{c}_o - \mathbf{p}_i|^2}}{\displaystyle\sum_{i=1}^{n} \frac{1}{|\mathbf{c}_o - \mathbf{p}_i|^2}} \tag{3.1}$$

where $\mathbf{p}_i$ $(i = 1, \ldots, n)$ are the n nearest input points $\mathbf{p}(x, y)$ and $\mathbf{c}_o - \mathbf{p}_i$ is the distance from the centre of cell $\mathbf{c}_o$ to point $\mathbf{p}_i$ (Neteler and Mitasova, 2002). Burrough (1986) recommends a value of n between 4 and 12, while stating that choosing the number of points to use in the interpolation involves a trade-off between computational efficency and precision. On a modern computer, computational efficiency was assumed to be of little concern and a value of $n = 12$ was used. Figure 3.3 shows the results of the IDW interpolation. In all the images in this section, a 2.5-D view of the landscape is shown

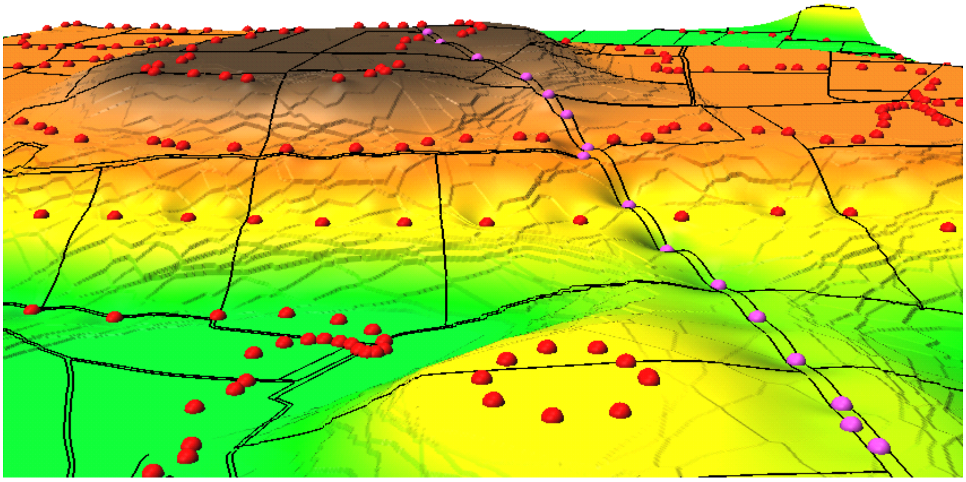


Figure 3.3: DEM interpolated using inverse-distance weighting (IDW): spot heights and contours alone used here.

from the same viewpoint. All the heights in the view (and subsequent images in this section) are exaggerated by a factor of approximately 3.5 for emphasis. The sampled contour sites are shown as red spheres and the spot heights and benchmarks as purple spheres. The layers representing perimeters from the large scale digital data have been overlaid in black to show the location of the road and fences etc.

It can be seen that this method has resulted in a fairly smooth and realistic reproduction of the landscape, but there are many artificial 'steps', which make the DEM unsuitable for very detailed analysis. Also apparent are some points on the road where the purple spot heights are sitting in a 'crater' below the surrounding landscape. This suggests some kind of mismatch between the accuracy of the contours and spot height data.

Interpolation by Regularised Spline with Tension

Inverse distance weighting is a simple and intuitive technique and can produce perfectly adequate results for many purposes. However it can occasionally cause artifacts such as steps when viewed close-up and is not as suitable when the sites data are unevenly distributed, e.g. in the case of contours. Several interpolation methods have been developed to improve interpolation of site elevation data to DEM. Among these is the 'Regularised Spline with Tension' (RST) method (Mitášŏvá and Mitáš, 1993; Neteler and Mitasova, 2002), which is implemented in the GRASS command `s.surf.rst`.

The RST method works by simultaneously fitting a trend surface to the data while forcing the resulting surface to go as closely as possible through all the points using a radial basis function. The *smoothing factor* may be explicitly specified and if it is set equal to 0, the resulting surface will go through all the points. As can be seen in Figure 3.4, these twin aims of producing a smoothly varying surface while forcing it through all the data points has produced many anomalous peaks and craters in the areas where the two data sources are close together (i.e. where the road crosses contours, going up the hill.) Again this suggests there may be a discrepancy between the two data sources, but by manipulation of the RST parameters an attempt will be made to derive a satisfactory DEM.

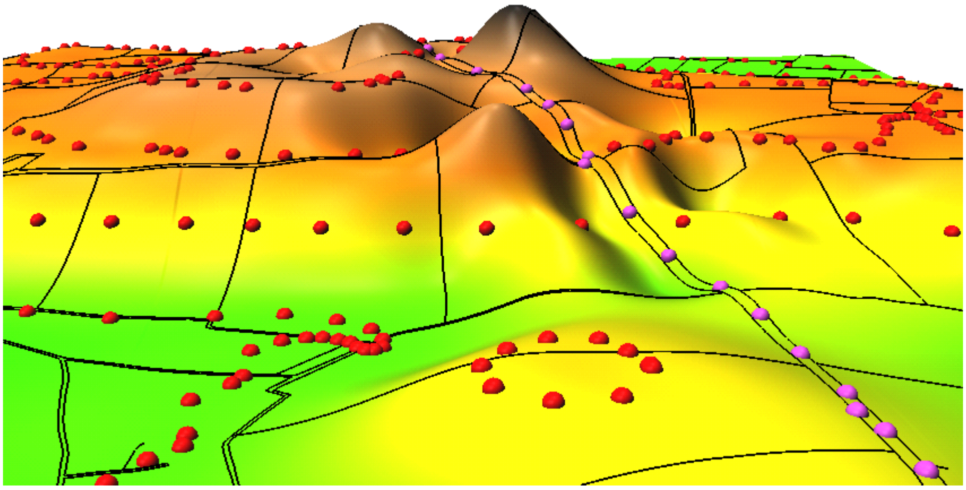


Figure 3.4: RST interpolation with smoothing = 0, i.e. DEM forced to pass through all data points. Default tension of 40 used. Only contours and spot heights, not OSNI DTM-derived data used here.

The equations describing the RST smooth-surface algorithm are presented by Neteler and Mitasova (2002).

The result in Figure 3.5 was obtained when the smoothing value was increased from 0 to 0.5, allowing the DEM to deviate from the points by a small amount if necessary, and the tension was reduced to 20 from the default of 40. The *tension* parameter controls whether the surface behaves like a ‘membrane’ (for high tension), i.e. it is stretched tightly over the data points and goes rapidly to trend everywhere else, or like a ‘thin steel plate’ (for low tension), where it is very stiff and changes only gradually. In this sense low tension also has a smoothing effect, i.e. localised anomalies will not have as big an influence on the overall smoothness of the DEM.

Although the DEM portrayed in Figure 3.5 is free from obvious distortion, it is known to have some deficiencies. For example, the dip between the small hill in the foreground of the image and the large hill in the background (marked ‘A’ in the figure) not as deep as it should be (when manually / visually compared with photographs of the same area). This is perhaps due to a lack of data points in this area, causing the surface to tend more towards the trend of the surrounding higher points.

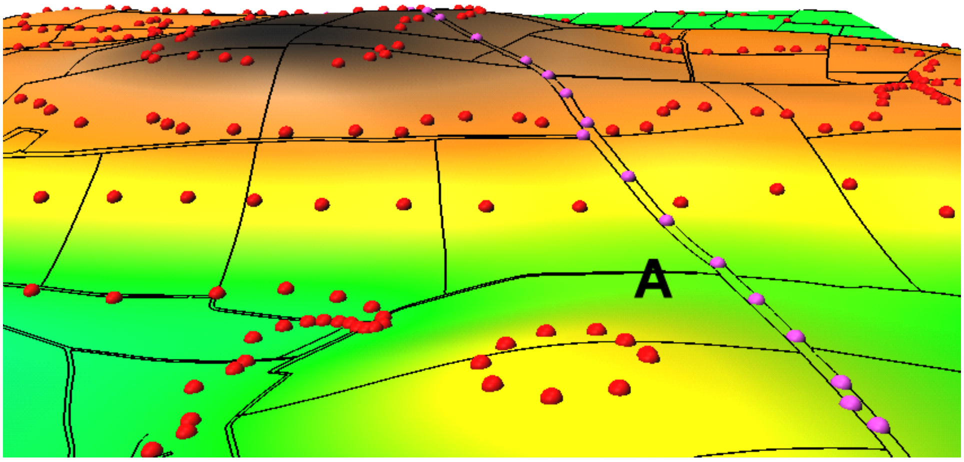


Figure 3.5: RST interpolation with smoothing = 0.5 and tension = 20.

Evaluation of Effectiveness of Interpolation Techniques

In the preceding sections the effectiveness of some interpolation techniques has been evaluated subjectively. Local knowledge of the test region suggests that the evaluation is valid, however a more thorough treatment would require a quantitative analysis of the errors. This is difficult as there is no definitive correct data source, however two variations on a method that will give some indication of the viability of the techniques for this type of data are

1. Use only half the data points for the interpolation. Validate the data by sampling the resulting raster at the points not used in the first stage. The difference should be very small if the interpolation technique was valid.

2. Interpolate the raster surface using the contour data and validate it by sampling with the spot height data.

In this case it is not worthwhile to investigate the effectiveness of interpolation techniques using the data available, because of the general doubts about its quality. If sufficient high-quality data were available it may be worthwhile to investigate an optimal interpolation technique such as *kriging* (Burrough, 1986).

3.4.3 Error analysis for OSNI elevation data

RST interpolation should in theory work very well for the type of scattered data being used here. The fact that there are some large errors under certain parameters (shown clearly in Figure 3.4) indicates that there is a problem with the data. There follows an analysis of the discrepancies between the following data sources:

- Continuous raster Digital Elevation Models
    1. OSNI-supplied, 50 m resolution
    2. Interpolated from contours, also with 50 m resolution for comparison
- Discrete vector / sites ground height measurements
    1. Spot heights from large scale maps
    2. Contours from 1:50000 maps

The analysis technique used was to sample each DEM using each source of discrete height measurements and calculate the difference between the source and sampled data. If all the data were consistent the bias would be expected to be zero over a large number of sample points and the standard deviation small, mostly caused by interpolation errors / smoothing effects.

As well as the St John's Point test region, where data were available the results are also presented for a test region in Belfast. This comprises an area bounded by the 374400 and 371600 northings and 332600 and 334400 eastings, depicted in Figure 3.6. The region is very flat with some hills towards the south. Many main roads pass through so there are a lot of measured spot heights.

The results of the analysis for the two regions are tabulated in Table 3.1; all the units are metres. The top half of the table shows that the spot heights and contour-derived DEMs as a source of height information are clearly inconsistent with each other. The large biases and standard deviations explain the effects seen in Figure 3.4.

	St John's Point		Belfast	
Spot Heights	Bias	Std. Dev.	Bias	Std. Dev.
OSNI DTM-sampled	−0.54	3.10	+1.72	2.06
RST DEM-sampled	−0.88	3.07	+1.29	3.03
Contour Points				
OSNI DTM-sampled	−0.09	2.52	+0.24	2.52
RST DEM-sampled	−0.04	1.84	−0.05	2.21

Table 3.1: Analysis of Errors in Height Data (values in metres)

Figure 3.7 shows some of the inconsistencies even more clearly, in the form of an aerial (map) view of a small part of the St John's Point test region. There is a lot of information contained in this figure so it will be summarised in point form:

1. The spot heights are referenced to the Belfast Mean Sea Level datum (known as MSL Belfast) whereas the contours are referenced to the Malin Head Mean Sea Level datum (OSi and OSNI, 2000). The manually digitised contours have been corrected to MSL Belfast by subtracting 3.7 cm (in this case they are 29.963 m

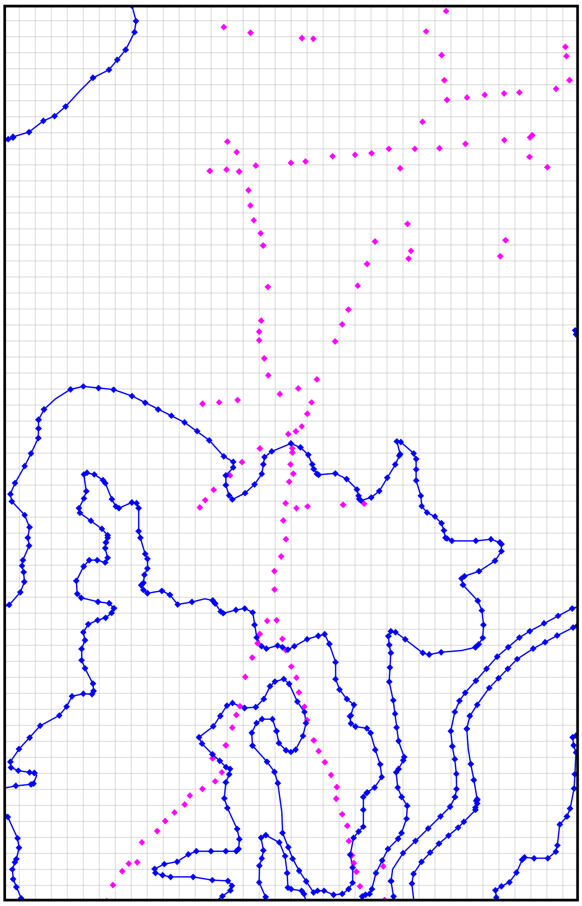


Figure 3.6: Contours and sampled points (blue) and spot heights (purple), for the Belfast test region. Overlaid on grid showing resolution of OSNI raster DTM data

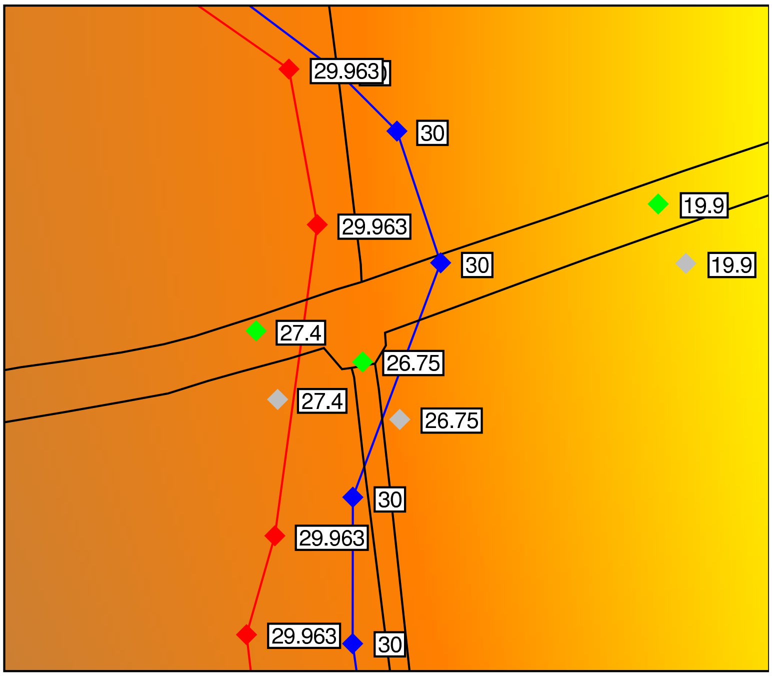


Figure 3.7: Detailed view of discrepancies between different Data Sources: A) Contours [Blue: from OSNI digital data, Red: manually digitised from paper map] B) Spot heights [Grey: from OSNI digital data, Green: manually digitised from paper map] C) Black: Perimeter layers from OSNI digital data

instead of 30 m). The digital data contours have not yet been similarly corrected as this factor is insignificant compared to other differences.

2. The grey and green markers for the spot heights refer to the same set of points. The green points have been manually input from a scanned paper map where the point they referred to is obvious, e.g. a cross in the centre of the road. The grey points have been automatically extracted from the OSNI digital data where they were attached to a text box containing the elevation value. The site marked in the figure is the lower left corner of each text box.

3. The position of the contour line is clearly wrong, as it can be seen that the 30 m contour line passes *below* the 26.75 m benchmark. Perhaps the positional precision of the contours in the DXF file is stated over-accurately for the quality of the data, and a very high smoothing value should be applied to the contours.

The errors for the sampled spot heights in Table 3.1 are similar for both the OSNI-supplied elevation model and that derived from the contours by RST interpolation. This may be explained by the fact that the OSNI data was also originally derived from the contours (Meneely and Graham, 2002) The interpolation algorithm used is not known.

Therefore the second half of Table 3.1 is essentially a comparison of interpolation algorithms: the unknown algorithm used by OSNI and the RST algorithm as implemented in GRASS. The very small bias values (several centimetres) indicate that the contours were indeed the source of the DEMs, and the larger standard deviations can be explained by the coarse resolution (50 m). In fact the RST-interpolated DEM has a smaller bias, particularly for the Belfast region and a smaller standard deviation, suggesting it is the better algorithm. The relatively large bias of +0.24 m above the contour heights for the OSNI DTM in the Belfast region suggests that perhaps the (unknown) interpolation algorithm was not suited to the relatively flat terrain there. However more factors would need to be evaluated for a full comparison of the interpolation results, and it is not relevant for this work.

As already mentioned, unfortunately there is no definitive correct data source that results can be objectively compared to, but the inconsistencies found in the OSNI data suggest that there may be a general problem with the quality of the data from that source. As a result of this two further sources of elevation data were investigated. Acquisition and initial processing of these sources is described in the next sections.

3.4.4 ASTER satellite digital elevation model

One such source of additional elevation data is the 30 m resolution ASTER DEM; this has a precision of 7 m both horizontally and vertically (Welch et al., 1998). This is available free over the internet from the NASA Earth Observing System Data Gateway in the USA. The data is derived from stereopairs of images from the ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) satellite using computer stereocorrelation techniques (Hirano et al., 2003).

The coverage of the Earth is not complete but the satellite has flown over parts of Northern Ireland several times. On two of these the sky was substantially cloud-free; the image from 21 October 2000 at 11:47am was selected and a request submitted for creation of a custom DEM. This process takes several months but once the DEM has been created it is available for immediate download by any user.

The data is a raster coverage and is available in the format of a geo-referenced TIFF image (GeoTIFF) file. It is in the Universal Transverse Mercator (UTM) projection, referenced to the WGS84 datum. To use it in the GIS with the OSNI vector data, etc., it must be transformed to the Irish Grid / Ireland 1965 system.

This can be accomplished using the re-projection commands in GRASS, `r.proj`, `v.proj` and `s.proj` for raster, vector and sites data respectively. When this work was commenced GRASS only supported re-projection and not datum transformation, however datum transformation capabilities were added as part of this work. While substantial work was involved to get the system working properly, it mainly involved changes to GRASS libraries and Makefiles, and familiarisation with GRASS and C development. The mathematical implementation of datum transformation was already part of the PROJ.4 projections software library (RemoteSensing.org, 2004) and will not be discussed here.

The changes made to GRASS were extensive and the work included:

- familiarisation with GRASS' current interface to PROJ.4
- familiarisation with recent datum-related improvements to PROJ.4
- modification of GRASS to use the new PROJ.4 API
- detection and correction of many bugs in GRASS' projection handling
- updating r.proj, v.proj and s.proj to remove and correct earlier only partly-functional attempts at adding datum transformation functionality

All these improvements were accepted back into the official GRASS source code and can be browsed at the source code repository http://freegis.org/cgi-bin/viewcvs.cgi/grass/ .

The re-projection is accomplished in GRASS by the following procedure

1. Set up a separate WGS84 / UTM location
2. Import the GeoTIFF ASTER data using the `r.in.gdal` command, which supports many common raster formats.
3. Re-start GRASS in the Irish Grid location, set the resolution to 30 m (ASTER resolution) and use `r.proj` to re-project and transform the raster data.

However, the bilinear or cubic interpolation implementations used by r.proj (Schroeder, 2003) cause distortion and errors at the edge of data areas, e.g. near the coast. This is because no-data cells are filled with the values of neighbouring cells before the interpolation calculation is done using the normal algorithm, rather than using a special-case parital-data algorithm. This is obviously important in the coastal St John's Point test region and so an alternative approach was used.

The raster DEM was converted to a regularly-spaced grid of sites representing the average height of each 30 m × 30 m cell, the sites located at the centre of each cell. These sites were then re-projected using the s.proj command; projection of point sites can be done with arbitrary precision. Finally the projected sites were interpolated to a raster using the RST algorithm (`s.surf.rst` command), which has previously been shown to give good performance (Sections 3.4.2 and 3.4.3). The coastal outline (High Water Mark Mean Tides) from the OSNI 1:50000 contours was used as a mask to limit interpolation to the landward side.

While the ASTER data is now geo-referenced to the Irish Grid, the actual height values are only relative to the approximate sea level height as determined by the satellite's navigation system. It is possible to generate an absolute DEM if ground control points are provided; however the GCPs are required to be specified in terms of their pixel locations in the original ASTER stereo images from which the DEM is generated. These must be purchased, although before August 2002 they had been available free, as their derived products (e.g. DEMs) still are.

Thus a further transformation will be necessary to make the ASTER heights compatible with other data. This will be described in Section 3.4.6 after a fourth data source has been introduced.

3.4.5 Kinematic GPS data

Kinematic GPS measurements may be used to determine the exact location of a moving platform containing a camera (see Chapter 5). But if the offset of the moving platform above the terrain surface is known, then these measurements may also be used to augment the other sources of ground elevation data.

A dual-frequency Leica GPS receiver is operated in kinematic mode at its maximum sampling frequency of 10 Hz. The mobile data is post-processed together with data from a static base station and this results in centimetre-level accuracy.

Section 5.2.2 discusses processing of the GPS data in more detail. The positional output is in terms of latitude / longitude co-ordinates referenced to the WGS84 datum (easily transformed to Irish Grid / Ireland 1965 in GRASS) and WGS84 ellipsoidal height. This can be transformed to MSL Belfast orthometric height by subtracting the OSGM02 geoid model (Forsberg et al., 2002).

OSGM02 models the geoid, i.e. the Earth's gravitational field (which determines orthometric height). The model consists of a table of values denoting the separation between the surface of the WGS84 ellipsoid and MSL Belfast at the vertices of a 1 km grid. Over Ireland the WGS84 ellipsoid is always below sea level; the values in the table are always positive, and of the order of 60 m.

These points were imported into GRASS and interpolated to give a raster layer. Then at each of the imported GPS points, the value of this raster layer was interpolated and subtracted from the GPS height to give the MSL Belfast orthometric height. Finally the offset of the GPS antenna above the terrain surface was subtracted to give the actual ground elevation at each point. Transformation between the OSGM02-corrected GPS height data and the other sources will be discussed in the next section.

3.4.6 Conversion to OSGM02 (reference elevation datum)

As mentioned in Section 3.4.5, kinematic GPS has been used to measure the location of the camera. To ensure consistency, the digital elevation model should be referenced to the same vertical datum, i.e. MSL Belfast as calculated using the OSGM02 model.

OSGM02 is now considered the definitive vertical datum in Great Britain (Ordnance Survey, 2002). The principle behind this is that GPS-determined heights are an order of magnitude more precise than those determined using traditional levelling techniques. It is unclear if this also applies to Northern Ireland however an accuracy (RMS error) of < 4 cm is cited for this region and the model will be used in this work.

In this section the relationship between OSGM02 and the other data sources is analysed. GPS heights were measured along a 7.2 km loop of road in the St John's Point area. The elevation sources to be analysed were interpolated into raster layers at high resolution (1 m) and sampled using bi-linear interpolation at the GPS points.

As OSGM02 (GPS heights) defines MSL Belfast (OSNI heights), there would be expected to be a high degree of correlation between the two sets of elevation values. This correlation was tested by using linear regression to calculate a scaling factor and offset that would transform interpolated contour heights to OSGM02. This was done using a computer spreadsheet, which is useful because a statistical summary of the regression output is given. This allows to determine if the calculated transformation is a good fit to the data. Table 3.2 summarises the results of the linear regression analyses in this section; these results will be interpreted in the rest of the section.

Description of variables	R^2	Intercept		1st Ind. Var.		2nd Ind. Var.	
		Value	Std. Err	Value	Std. Err	Value	Std. Err
Contours & Geoid	0.93997	-937.42	116.35	0.8966	0.0028	16.74	2.08
Contours	0.93940	-0.4592	0.0434	0.8915	0.0028		
Spot Hts & Geoid	0.99801	-265.92	43.63	1.0056	0.0009	4.70	0.78
Spot Hts	0.99798	-2.7157	0.0198	1.0057	0.0009		
ASTER & Geoid	0.87318	6415.67	207.76	0.8147	0.0045	-114.47	3.71
ASTER	0.84807	8.059	0.056	0.7846	0.0048		

Table 3.2: Results of linear regression analyses

The co-efficient of determination (also known as R^2) is the proportion of the variation in the dependent variable (i.e. the calculated OSGM02 height) that can be explained by variation in the independent variables(s) (Mezei, 1990). For the interpolated contour data the relationship was found to be

$$h_{\mathrm{OSGM02}} = 0.8915 \times h_{\mathrm{contour}} - 0.4592 \tag{3.2}$$

and R^2 was 0.939, which implies a good correlation. Additional terms were included experimentally in the regression analysis, namely the square of the contour height, and the OSGM02 geoid separation, but neither of these gave any significant improvement in R^2. If the standard error of any co-efficient approaches the magnitude of that co-efficient it indicates that the corresponding variable does not have a strong influence on the output and its effect can be neglected (Mezei, 1990). This is the case for the addition of geoid separation to the contour calculation (lines 1 and 2 in the table), for example. As the contour height is an orthometric height it would already be expected to be 'corrected' for the geoid, however it was included experimentally. The 6.1% unexplained variation can be assumed to be the result of information lost due to unmodelled variations in height within the contour intervals.

To test this assumption a 1 metre resolution DEM along the main road was interpolated using only the spot heights on the road. As these are located at frequent intervals throughout the area of interest (the GPS points are also located on the road) it is a reasonable assumption that there will be no unmodelled variations in height along the road here.

This assumption appears to be correct as with the spot heights R^2 was 0.998, indicating that almost all the variation in the calcualted OSGM02 height can be explained by variation in the input spot heights. However the transformation equation deserves closer attention:

$$h_{\text{OSGM02}} = 1.0057 \times h_{\text{MSL}} - 2.7157 \tag{3.3}$$

i.e. there is negligible scale error but the OSGM02 heights are all offset by approximately 2.7 m below the surveyed spot heights. It is highly unlikely that OSNI surveying techniques (Perkins and Parry, 1996) could have been so bad as to introduce an error of 2.7 m in the measured heights. This result would appear to cast doubt on the validity of OSGM02 as a transformation to MSL Belfast, certainly the claimed accuracy of < 4 cm in Northern Ireland.

However the requirements of consistency between the surveyed camera position and GIS data used for perspective transformation outweigh the requirement for compatibility with existing elevation data. N.B. the accuracy of the two-dimensional OSNI vector data is not in any way affected by errors or otherwise in the height data.

The third source of elevation data to be matched to the OSGM02 vertical datum is the ASTER DEM. Attempting linear regression using a scaling of the height and addition of an offset gave an R^2 of only 0.84 in this case. With the addition of the OSGM02 geoid separation as a second independent variable, the value of R^2 rose to 0.87. This is understandable as the ASTER DEM did not incorporate any geoid correction. No further increase in correlation could be obtained by adding squared or cubed terms for either variable. The transformation equation is

$$h_{\text{OSGM02}} = 0.8147 \times h_{\text{ASTER}} - 114.47 \times h_{\text{geoid}} + 6415.67 \tag{3.4}$$

As mentioned earlier, a principal aim of attempting to use the ASTER data was to model the variations between contour heights and away from measured spot heights. However the fact that the ASTER DEM has more unexplained variation than the interpolated contour DEM (13% rather than 6%) is a small cause for concern and mitigates against this advantage.

Visual inspection of the ASTER data shows no obvious errors in the region of interest and indeed the opposite—some land features such as small bogs that are not really modelled at all in the contour-derived DEM are clearly visible in the ASTER DEM. There may again be small variations that are not modelled by the 30 m resolution of the data, however it is assumed that most of the 13% unexplained variation derives from uncertainity in the calculation of the relative ground heights using the satellite ephemeris data (since no ground control points were provided for the DEM generation).

It is argued that despite this 7% more unexplained variation, considering the method of acquisition and properties of the ASTER data still leads to the conclusion that it is a better overall model of the ground elevation than the interpolated contour data. In the latter case the original data (actual ground elevation) has already been sampled to create contours, with data being lost (i.e. variations within contour intervals) in the process. This lossy data was then interpolated to create the DEM. However the ASTER data is used in its original continous coverage format, where an elevation value has been calculated for every 30 m-side square on the ground.

Although this argument is only justified using theory and subjective evidence (see discussion of minor land features above), ASTER should in general model all features although the accuracy of placement of those features will be less than the contour-derived DEM. In the next sub-section a method of improving the accuracy where it is most important, i.e. near the camera, will be discussed. The relative merits of the ASTER and contour-derived data would form a worthwhile area of further research.

3.4.7 Final combined DEM

The RST interpolation modules in GRASS allow for the combination of data sources of different or even variable accuracy, by using the variable smoothing feature. This allows for the association of a 'smoothing factor' with each data point. Points known to be accurate can be given a low smoothing factor, while those with a high margin of error or subject to noise (for example) can be given a higher smoothing factor (Cebecauer et al., 2002).

This allows valid data sources even of low precision to make a worthwhile contribution to the output. For example, the GPS heights are known to be very accurate with an RMS error of only a few centimetres (available in the GAMIT output for each point) and so will be given a low smoothing factor. In regions close to their locations, the combined DEM will be largely determined by their values. The less accurate

ASTER data will have a high smoothing factor, causing it not to have much influence here. However, in regions away from the main road where there are no GPS points the ASTER data will strongly influence the output, despite its high smoothing factor. This is because there is no more accurate data (i.e. with a low smoothing factor) available for that region.

This arrangement allows the ASTER data to contribute to an accurate DEM despite some concerns over its accuracy. Where accuracy (and consistency with the measured camera position) is most important, i.e. near the camera locations on the road, the accurate GPS heights will override the ASTER data. The ASTER DEM can better model land features further from the road and camera where absolute height accuracy is not quite as important.

One of the most important and relevant advantages of this technique derives from the fact that for most of the system tests described in the rest of this thesis, GPS positioning equipment was used to determine the camera position. If the same GPS data is also used to determine the elevation model, the two positions will be consistent and this is likely to lead to more useful results and a more meaningful evaluation of the methods.

It was thus decided to use a combined ASTER / GPS digital elevation model for the rest of the work.

3.5 Conclusions

The need for accurate geographical data was identified and the requirements for a geographical information system were specified. The GRASS GIS was selected for use because of its competence in the areas required.

Details were presented of the co-ordinate reference systems in common use in Northern Ireland. Sources of geographical data were then identified and evaluated as far as possible in terms of data coverage and accuracy. OSNI was identified as the only source of large-scale vector feature data. The resolution of the data was derived and recognised as sufficiently accurate for this work. Existing literature was also found to endorse the quality of OSNI vector data.

Sources of elevation (third dimension) data were found to be available as three separate products from the OSNI range. Attempts were made to merge these into one higher-quality dataset using GIS data processing techniques, in particular raster interpolation. A statistical error analysis was carried out and revealed that the sources were inconsistent with respect to each other.

Further sources of elevation data (ASTER DEM and measured GPS heights) were obtained and pre-processed to conform to the the Irish Grid. Substantial modifications and improvements were made to the re-projection functionality in the GRASS GIS to enable this pre-processing.

Linear regression was used to correct the elevation values in these sources to a common reference vertical datum consistent with the OSNI height data. This incorporated correction for the geoid (i.e. a model of the Earth's gravitational force). The OSNI height data was also corrected in this way to verify the theoretical assertion that no geoid correction is needed for it. Good correlation was obtained in most cases with all data sources and sources for any unexplained variation were postulated.

Interpolation with variable smoothing was used to interpolate the very accurate GPS heights and less accurate ASTER data into a combined DEM. Good results from visual inspection, together with the inherent consistency with a GPS-measured camera position, led to this DEM being selected as the one to use for further work. The next chapter describes how these data sources (vector feature data and interpolated raster DEM) are used during 3-D visualisation and perspective transformation.

The interesting and quite unexpected (in terms of the apparent problems found with publicly available data sources) results found in this chapter certainly merit further research and investigation.