Appendix 1 Graphical Rendering

1.1 Introduction

Rendering is the drawing of a real world object as it actually appears and, within computer graphics, specifically refers to the creation of a 3-D image that incorporates lighting effects to achieve this.

This section will briefly discuss the different models used to graphically render 3-D images. These models are based upon the position, orientation and characteristics of the surfaces and light sources illuminating them, however the actual physical interaction between visible light and surfaces is complex and consequently

graphics researchers have often approximated the underlying rules of optics and thermal radiation... Consequently many of the illumination and shading models traditionally used in computer graphics include a multiple of kludges, "hacks," and simplifications that have no firm grounding in theory, but that work well in practice. (Foley et al, 1990).

1.2 Rendering Models

The following sections discuss the two main rendering techniques, termed *local* and *global* illumination models. All models are, in general, physically based, being derived from physical principles of light and its interaction with different surfaces. A purely physically based model would be computationally intensive and so *localillumination* algorithms tend to incorporate approximations and simplifications to produce a graphically pleasing image. This method works on the principle of taking a surface and then incorporating visual enhancements where appropriate (e.g. directional light source, ambient light, material properties, reflections and transparency).

Globalillumination attempts to incorporate data on all reflected and transmitted light to produce a realistic image; this is opposed to using a simple ambient illumination term.

Both techniques require an input model (i.e. 3-D description) for the objects of interest, specification of the viewing and lighting geometry, calculation of the viewable surfaces (from the rendering viewpoint) and object colour assignment.

Local Illumination

This section describes the main techniques used in local illumination models, starting with simple illumination and slowly adding more complex effects.

1. Illumination

At its simplest, illumination is incorporated into a 3-D model by using non-directional light (i.e. no direct light source). This can be *self-luminosity* or, more realistically, *ambient light*, the latter of which can incorporate the reflective properties of an object. Most environments, however, have some kind of *point light source*, be it the sun or a lamp, from which light emanates in all directions. Object brightness depends upon the distance to, and direction of, the light source, whilst the object material properties control the type of reflection. Surface reflectance can be modelled using two main reflectance types:

1. *Diffuse (Lambertian) reflection* describes matte surfaces which appear equally bright from all angles. Diffuse light and colour can be added to make the shading more realistic. Further corrections can be added for atmospheric attenuation whereby more distant objects appear less intense than closer ones.

2. *Specular reflection* can be used to simulate shiny surfaces; simpler models assume equally radiating light (e.g. Phong), whilst newer models allow directionality to be introduced, in a manner similar to a spotlight.

2. Polygon Shading

Shading can be defined as *the calculation of light leaving a point*. To produce an image this can be performed by applying an illumination model at every visible point throughout a landscape, however this is computationally intensive

and so programmers have devised more efficient algorithms to approximate the visual rendering of this method for polygon models. These include:

1. *Constant shading* applies one illumination calculation to each polygon within a wire frame model; this is simply a method of sampling within the landscape.

2. *Interpolated shading* calculates illumination values at polygon vertices and then interpolates across the polygon between these values.

The above two methods suffer from the assumption that a polygon accurately represents the modelled surface, however for curved surfaces faceted banding becomes evident, an effect that is accentuated by the eye. However polygon wire frames are popular as visible surface algorithms are very efficient in this environment. *Gouraud shading* attempts to overcome these problems by interpolating polygon vertex illumination values. Polygon vertex normals are averaged from the surface normals of the polygon facets, vertex intensities (using a desired illumination model) are calculated and then intensities are linearly interpolated along and between edges.

Phong shading uses an alternative approach by interpolating the surface normal vectors rather than the intensity, a method which yields better results when using specular reflectance illumination models. As in the Gouraud shading, vertex normals can be averaged and then used to interpolate edge normals before going onto interpolate surface normals.

3. Surface Detail

Surface detail can be added to rendered polygons to make the final image more realistic. This can involve the overlay of *surface-detail polygons* which are ignored during the visible surface procedure, but then applies detail when the base polygon is shaded. This method is only suitable for gross detail and when more detailed textures are required a separate image is mapped onto the surface, a process known as *texture mapping*. *Bump mapping* modifies texture

mapping to take into account differences between the texture map and surface illumination conditions.

The above steps follow the production of a basic rendered image, however further details need to be considered in order to produce a realistic image. These include:

a. **Shadow** - shadow algorithms operate in relation to the point light source used in a scene; where multiple light sources are used, a relative shadow effect must be calculated for each surface.

b. **Transparency** - surface materials can be transparent or translucent. Simpler models ignore light refraction as the algorithms are significantly simpler and, through interpolation or filtering, calculate the colouring of the surface.

c. **Interobject Reflection** - where one surface reflects off another within an image, interobject reflection occurs and this can range from diffuse to specular. *Reflection mapping* attempts to address these issues however *it only provides an approximation to the correct reflection information* (Foley, *et al*, 1990).

4. Physically Based Shading

The Torrance-Sparrow model is a physically based model of a reflecting surface and has shown good correspondence to actual measurements. It assumes an isotropic collection of planar microscopic facets, each a perfect specular reflector. It has since been modified for computer graphics and incorporates the Fresnel equation for the specular reflection of unpolarized light of a non-conducting surface. Two further enhancements include the development of anisotropic illumination models which are better able to render preferentially oriented surface microfeatures and a correction for polarisation of light after it is reflected.

5. Further Enhancements

a. Extended Light Sources - models that are able to render “soft” shadows where there is only partial blocking of the point light source.

b. Camera Effects - post processing methods can add the effects of *depth of field* and *motion blur*.

Global Illumination

Global illumination models calculate the colour of a point using incoming light and light reflected and transmitted through different surfaces. Local illumination is concerned only with incoming light. The illumination models dealt with above have tended to use an single ambient light term in them and did not consider the position of the object or viewer or the effect of nearby objects in blocking ambient light. The *Ray Tracing* (view-dependent) and *Radiosity* (view-independent) methods are designed to incorporate these more complex interactions in a manner which hasn’t been achieved in the above models and are discussed below.

1. Ray Tracing

Ray tracing determines surface visibility by tracing imaginary rays of light from the centre of projection, through a viewing window to the objects of interest. For each pixel in the viewing window the colour is set to that of the object of *first intersection*. To calculate *shadow*, an additional ray is fired from the point of intersection to the light source and if a further intersection takes place then the area is in shadow and its colour contribution is ignored. A conditional statement within the algorithm can then go on to spawn *refraction* and *reflection* for specified objects. Each of these subsequent rays can then also spawn shadow, reflection and refraction rays which continues to a user specified maximum or until no object is intersected. Intensity is then computed from the last ray up, a node being a function of its children’s intensities.

Many modifications to conventional ray tracing have been made to increase efficiency, to more accurately model specular reflection and provide a better method than the simple point sampling. This final area has spawned methods

such as cone, beam and pencil tracing. Distributed ray tracing is a variation that uses a stochastic approach to ray tracing; it is able to offset the obvious effects of *temporal aliasing* (jagged edges) for the less obvious effects of noise.

One problem of these tracing methods is that they fail to account for indirect reflected and refracted light sources; to overcome this problem *backward ray tracing* attempts to trace the ray from the light sources to the viewer, however this method is usually used in tandem with conventional ray tracing, known as *bi-directional ray tracing*, in order to increase efficiency.

2. Radiosity

Ray tracing is able to model specular reflection and dispersionless refractive transparency effectively, however it still uses a single term to account for ambient light. *Radiosity* is the rate at which an object emits radiation and can be modelled within a closed environment such that all energy is either absorbed or reflected. This method removes the need to have an ambient light term as all interactions of radiation with a surface are accounted for. The method requires the computation of all light interactions, before a viewing position is selected and visible-surface and shading performed. This method has been refined since its introduction into computer graphics and can allow the progressive refinement of a rendered image with increasing iterations, rather than having to perform all the radiosity calculations first. Specular reflection has also been added into the radiosity calculations, however the computational overhead is quite high.

3. Hybrid Algorithms

Some authors have combined the radiosity and ray tracing methods to take advantage of each method's strengths in modelling ambient light and specular reflection respectively.

Ray tracing calculates the illumination signal for *each* shading point which means that even small or distant objects are captured as sharp images, however it is unable to trace indirect (ambient) illumination as it is coming from *all* objects and the method is only point sampling. It is therefore possible to miss significant sources of light. Conversely, radiosity algorithms are able to model

indirect light well, particularly multiple diffuse light sources. Such effects as soft shadows and colour bleeding are more easily produced, however specular reflection is reproduced less well due to a limiting resolution for the calculations of radiance.

Most hybrid algorithms begin with the view-independent radiosity calculations first in order to determine the ambient light. This is followed by the ray tracing to calculate the specular reflection.

1.3 DiscussionandConclusions

This appendix has given a brief introduction to computer based graphical rendering and so provided a framework with which to discuss its relevance to relief shading within GIS. Relief shading provides a special, simplified, situation for rendering. The object of interest (i.e. the DEM) is already a sampled surface. Simple shadow algorithms can determine whether any point on the surface is in shadow from a direct illumination source and then apply an interpolated shading algorithm. A constant term can also be used so that ambient light can be added. This *localillumination* model provides a fast and efficient solution for most scenarios. Unfortunately it is unable to accurately reproduce real world illumination conditions particularly those under low solar elevation angles; although the shadow algorithms are able to handle line-of-sight viewing adequately, the ambient light level is far too high as, in reality, this term is very sensitive in low light conditions. In addition, the algorithms do not add surface detail to account for the reflectance properties of the surface. A set specular reflector is assumed, generating high levels of reflected light. Ideally the lower levels of incident radiation associated with low solar elevation angles would be modelled, along with emitted and reflected light within the environment.

In conclusion, it is not surprising that computer relief shading is inappropriate for modelling earth surface radiation interactions as it is only broadly physically based with many simplifications designed to make the calculations more efficient. It is appropriate for the simple azimuth based experiments described in Chapter 5, however it is not suitable for modelling the effects of low solar

elevation of a topographically complex landscape with surfaces of varying reflectance characteristics.

Appendix 2 Solar Azimuth and Elevation Calculations

2.1 Introduction

Chapter 5 has presented conclusions for understanding the bias of lineament representation on satellite imagery. It went on to recommend the conditions through which optimum imagery can be obtained and presented calculations of solar azimuth and elevation (for Landsat ETM+) to help with this.

The following Appendix contains the algorithms necessary to calculate the solar azimuth and elevation angles for any given location, on any given day. This is specifically for the purpose of calculating these values for the days and times that a given satellite passes over a specific location. For example, Landsat ETM+, a near polar-orbiting satellite, has a swath width of ~180km. It has 233 paths which it takes 16 days to overfly, arriving back at its starting position. Each path is split into 248 rows (~180km long), covering the lit descending path (north pole to south pole) and unlit ascending path (south pole to north pole). The resulting grid is called the World Reference System (WRS) and is used to refer to all images. Each orbit takes approximately 90mins to complete and, due to the near-polar orbit and earth rotation, the paths aren't overflown in strict numerical order. Rather they follow a 7-day interval period, such that path 1 is overflown on day 1 and path 2 is overflown on day 8. As a result, satellite overpass times vary with both longitude and latitude, although adjacent paths have similar overpass times. In addition, the inclination of the satellite degrades through the year slightly altering the overpass times, although this is corrected once a year. Between years overpass times can vary by as much as 5 minutes.

The calculations have been summarised from Muneer (1997) where binaries and source code are available for individual use, and mainly utilise the work of Yallop (1992). The calculations below are based upon UTC time and require the latitude and longitude of the location and the time and day, month and year of the satellite overpass.

2.2 Solar Declination and the Equation of Time

All calculations are based around the *solar day*, which is measured as the time elapsed between the sun crossing a local meridian and the next time it crosses the same meridian. The solar day varies in length throughout the year, principally due to the tilt of the earth's axis and the angle swept by the earth-sun vector during a given period of time.

All calculations of solar position are relative to the position of the earth in its orbit around the sun and require the use of solar time. The *Equation of Time* (EOT) is used to calculate the difference between clock time and solar time. Once computed, this is then corrected for the local time of the area of interest and the difference between this position and that of the longitude of the standard time meridian. Finally the *solar declination* (DEC) can be calculated. This is the angle between the earth-sun vector and the equatorial plane.

A variety of equations are available to calculate both the equation of time and the solar declination at varying degrees of accuracy. For ease of use and accuracy the methods of Yallop (1992) are followed, as reported by Muneer (1997). Although the calculations are simple to follow, many of the variables do not have a physical meaning, however they are described for completeness. The calculations are as follows:

```
t = { (UT/24) + d + [30.6m + 0.5] + [ 365.25(y – 1976)] – 8707.5 } /
36525
```

where

```
UT = Universal Time
y = year
m = month
d = day
min = minutes
s = seconds
t = day count from epoch J2000.0 (noon on 1 January 2000)

UT = h + (min/60) + (s/3600)
```

In addition if m > 2 then y = y and m = m – 3, otherwise y = y –1 and m = m + 9.

The following terms are then calculated. It may be necessary to add or subtract multiples of 360 (for G, L and GHA) to set them in the range of 0-360.

```
G     = 357.528 + 35999.05t
C     = 1.915 sin G + 0.020 sin 2G
L     = 280.460 + 36000.770t + C
a     = L – 2.466 sin 2L + 0.053 sin 4L
GHA   = 15UT – 180 – C + L – a
SHA   = [ 15(UT-12) ] + 180
e     = 23.4393 – 0.013t
DEC   = tan-1 (tan e sin a)
EOT   = (L – C – a)/15
```

where

```
G = mean anomaly
C = correction to centre
L = mean longitude
a = right ascension of sun (apparent) in degrees
GHA = Greenwich hour angle of sun in degrees
SHA = Solar Hour Angle
e = obliquity of the ecliptic
DEC = declination of sun (apparent) in degrees
EOT = Equation of Time
```

2.3 Solar azimuth and elevation

Once the above calculations are complete, the following algorithms are used to calculate solar azimuth (AZ) and elevation (ELE):

```
sin ELE     = sin Lat sin DEC – cos LAT cos DEC cos SHA

cos AZ      = [cos DEC (cosLAT tan DEC + sin LAT cos SHA)] / cos ELE
```

2.4 WorkedExample

The worked example uses a WRS scene from Ireland (Path:207. Row:23) for 1st December 2000. The overpass time was 11:20GMT and the scene centre co-ordinates are N53.104º and W8.203º. The following calculations are then made:

```
Longitude Correction = 8.203/15 = 0.5468

UT    = 11.03
UT Corrected = 11.03 + 0.5468 = 11.57
d     =  1
m     = 9
y     = 2000

t     = [ (11.57/24) + 1 + 275.9 + 8766 – 8707.5 ] / 36525
      = 0.009196

G     = 688.58
G (reset range) = 328.58

C     = -1.01610

L     = 610.5070
L (reset range) = 250.5070

a     = 249.0074

GHA   = -3.93
SHA   = 165.45

e     = 23.4392

DEC   = -22.0369

EOT   = 0.1677 x 60 =  10.062mins

ALT   = 13.81º
AZ    = 166.13º
```

2.5 Conclusions

Although the above calculations are straightforward, they can be lengthy. The purpose of calculating solar elevation and azimuth is not for images that have already been acquired (as this information is usually provided with the meta-data) but rather to perform these calculations over a series of dates to ascertain the best images to acquire. This can be performed manually, however the user will probably want to use the algorithms with the supplied Microsoft Excel spreadsheet to calculate these values. This has already been performed for every Landsat ETM+ WRS cell for the 1st of each month, based around an overpass cycle starting 13th August 2000. Finally, the above algorithms can be programmed manually to allow the user to perform their own calculations on any location for any satellite overpass time.

Appendix 3 Spatial Data Accuracy, Error and Error Assessment

3.1 Introduction

Geography is the study of phenomena with regard to their location. This is operational in both 3-dimensional space, as well as over time. As a result, a phenomenon's characteristics can change in space and time, which consequently affects relationships with other phenomena.

Geographical study tries to simplify the complex nature of relationships between phenomena in space and time by observing them and providing a rational framework through which they can be described and understood. It is the process of observation (phenomena location and attributes) that leads to the recording of spatial data. As geographers attempt to simplify (or abstract) the environment through their description of spatial data, it follows that spatial data are themselves necessary simplifications of our complex environment. Any simplification will incorporate error, whether the recording techniques are analogue or digital, precise or approximate.

This section is concerned with the issue of data quality; the manner in which errors are manifested, their propagation and how to minimise their effect. The phrase *data quality* assumes some standard by which a data set can be judged. Chrisman (1983) suggests that this should simply be a measure of the fitness for purpose. Data only becomes information when there is meaning attached to it and, for this to occur, it must happen within the context of the area of study. In addition to the discussion below, further general texts include Jones (1997), Chrisman (1994), and Aronoff (1989).

3.2 Accuracy

If spatial data are an abstraction of the real world, then a measure of data quality is the accuracy with which the real world are represented. *Error*, or the difference between observed and real values, is a measure of accuracy. Although simple in concept, accuracy assessment is actually more complicated

as it assumes that there is a truth or real world value which can ultimately be known. This very much depends upon the type and scale of the data which is collected. For example categorical attribute data can be considered ultimately knowable (i.e. a person is male or female), whereas numerical measurements (e.g. locational data) are only known to some degree of certainty. Consequently we cannot know what the real or true values are and so accuracy assessment can only be accomplished by comparison with the most accurate measurement available.

The terms *precision* and *accuracy*, are often used interchangeably and as a consequence, incorrectly. Precision refers to the detail used to report an observation (e.g. for numerical data this is the number of significant figures), whereas accuracy is the closeness to the real value. Observations should be reported at a precision equivalent to their accuracy.

3.3 Data Quality

Datasets are comprised of any number of phenomena (or data elements) which have characteristics that affect data quality. In addition, the dataset as a whole has characteristics that also affect data quality. This section will now discuss these elements.

3.3.1 Data Elements

Spatial data elements are made up of locational and attribute data. Locational data can be comprised of points and lines, both of which can suffer from positional error (i.e. a deviation in observed position from the actual position). Positional error can be of two main types:

1. Systematic
2. Random

A *systematic error*, also termed a *bias*, refers to systematic or regular deviations between the observed values and true values. Although ideally these should be zero, if the systematic error is recognised then it can be corrected. *Random errors* are more problematic as, due to their randomness, they are more difficult

to correct and consequently affect the accuracy of the data in unpredictable ways.

Attribute error is concerned with *logical consistency,* which is the maintenance of logical relationships between individual data elements, such that they are all observed in the same manner. For example a road can be mapped at its centre line or either boundary; for consistency the same method needs to be applied throughout.

3.3.2 Data Sets

At the dataset level there are four main elements that affect data quality. These are:

1. Completeness
2. Applicability
3. Compatibility
4. Consistency

Completeness refers to the overall inclusivity of all real elements that are required to be mapped. Datasets can be incomplete yet have high accuracy; again, quality is concerned with fitness for purpose and incomplete data may be suitable for certain applications. Although a separate measure of quality, completeness does involve elements of accuracy which are similar to logical consistency and attribute accuracy.

Resolution is an important element of completeness as it refers to the minimum recording distance. Although a separate concept from accuracy, it is normal practice to make this value smaller than the desired accuracy of the system. In comparison to vector systems, raster based systems often apply the term slightly differently, suggesting it is concerned with the smallest distance over which change is detectable. Regardless of which definition is used, phenomena of different sizes will be detectable with data of varying resolutions. In addition to resolution, systematic and random error may also affect completeness as they can cause the omission or commission of phenomena.

Applicability refers to the overall suitability of data for an intended purpose, analysis or manipulation. For example certain interpolation routines may be inappropriate when converting digitised contours to digital elevation models, whilst population data is inappropriate to map glacial landforms. Data therefore has to be fit for purpose and suitable for whatever techniques are to be applied to the data.

Compatibility refers to the appropriate integration and use of different data sets within a study. For example, the same co-ordinate system should be used between data sets. Data sets should therefore be consistent in the way that they are captured, stored and analysed.

Finally, *consistency* requires the maintenance of a logical relationship between phenomena. Within a GIS this is typically enforced through the topological model (e.g. all polygons must be complete).

3.4 Sources of Errors

So accuracy and error have been discussed with respect to the observations and their real values. However spatial data, and its use, is concerned with more than simply recording values, but storing, analysing and displaying them. At every stage in the process of using spatial data, error can be introduced and, as such, data quality is concerned not so much with removing error, but understanding and managing it. The following section briefly describes the different stages at which error can be incorporated.

Data collection is concerned the initial acquisition of a primary data source from which spatial data will be mapped. Much primary data will be remotely sensed and can contain inaccuracies or biases. The operator may have little control over how the data are collected, but an appreciation of data limitations and management is essential. Secondary data may also be used and these will have their own data collection errors, as well as errors incorporated through their production.

The following stage is *data input* and involves an operator extracting appropriate elements from the input data. Data input is usually of high precision, but accuracy will depend upon the quality of the input data and the experience and ability of the operator.

After data has been captured in an appropriate manner, it will be in a format which can be *manipulated and analysed*. It is at these stages that further error can be introduced; examples include co-ordinate transformations, raster to vector conversion, interpolation and overlay operations.

Finally, *data storage* requires an appropriate number of significant digits otherwise numerical degradation occurs, whilst data transfer can cause the loss of data between different data formats.

3.5 Accuracy Assessment

Accuracy is defined as the deviation of our observations from the real or actual values and, for numerical data (e.g. locational data), these real values will not be exactly known. Accuracy assessment still requires some measure of accuracy and this can be dealt with statistically by providing a numerical assessment of the expected accuracy. In this scenario, observations can be viewed as experiments involving repeated observations. If these measurements are Normally (Gaussian) distributed then we can use the dataset characteristics to provide estimates of expected accuracy. An alternative solution is to compare the dataset values with ones of higher accuracy and so use them as a surrogate for the real values. In this manner we can gain a better understanding of the types of error present.

Attribute data can be numerical or categorical; numerical data can be assessed in a manner similar to that used for locational data. Categorical data are treated according to accuracy classification methods developed within the remote sensing community through the use of a misclassification matrix. Using higher accuracy data, a matrix of correctly and incorrectly classified phenomena or objects can be drawn up and give an indication of the overall accuracy.

Although attribute and locational error are often treated separately they can be intimately linked. Where attributes are applied to phenomena such as soil units, changing an attribute because it has been incorrectly classified can lead to a locational change of that unit. Other boundaries (e.g. political) are fixed and so attribute changes do not affect location.

3.6 Conclusions

Although spatial researchers are particularly concerned with accuracy, it is error that is the central topic. All spatial data are inherently wrong and contain errors. It is the job of the researcher to understand error and minimise it to a level where it is fit for the purpose it is to be used for. Spatial data cover a wide variety of disciplines and as such methods for dealing with error are various. The earliest approaches to error management came from surveying, photogrammetry and geodesy, where reliable measurements are obtained through repeated measures and the use of least-squares estimation. In contrast, social scientists have developed statistical tools from other disciplines to meet their needs and have consequently inherited their approaches. These are mostly designed to deal with errors from sampling.

GIS has to take these approaches and provide a way of incorporating and using them within a technical environment. It will then be possible to make users aware of the error that is inherent in their data and so provide methods of managing it. Within earth science this can provide minimum error guidelines, whilst for social science an awareness of the error can be reported and incorporated into the result.