1

Introduction

Sanjay Rana

1.1 EVERY *THING* HAS A SURFACE

A surface is the most fundamental shape of matter to us. Surfaces surround us in various forms, ranging from the undulating ground we stand on to this flat page in the book, so much so that the recognition of surfaces and their structure is crucial to our daily life. In addition, a number of non-physical spatial phenomena such as temperature and population density are also modelled as surfaces to aid visualisation and analyses. Despite the wide variety, conceptually the description of even the most complex surface forms is rather simple. Basic surface descriptors such as circle, box, flat, convex, and others can be combined together to derive any arbitrary shape, thus enabling the computer graphics animators in Hollywood to reconstruct dinosaurs.

It is fascinating to appreciate how different disciplines describe surfaces. It is also worthwhile to highlight the issues in surface representation as it reveals the level of abstraction when the increasingly massive surface datasets are stored in computers. Mathematicians have modelled surfaces primarily with an aim to decompose the surface into the *basic descriptors* or elements even if it meant oversimplification (e.g. by using the primary surface elements mentioned above), leading to potential loss of the structure. Such descriptions are generic (i.e. universal to all types of surfaces), formal, and robust, such as that required in computer-aided designing. The aim of the mathematical description is to produce a constrained *global model* of the surface. The other large group of surface researchers from the field of physical geography use more *compound descriptors* (e.g. valleys, mounds, scarps, drainage network) with more emphasis on the preservation of the structural information of the surface. Although the *compound descriptors* are more *natural*, their relevance

Topological Data Structures for Surfaces – An Introduction to Geographical Information Science. Edited by Sanjay Rana
 ISBN: 0-470-85151-1

is subjective to each individual; hence it is often difficult to derive an objective definition of surface features[1]. These researchers are more interested in the process that resulted in the surface; hence the descriptors are also symbolic of the factors in the process.

A relevant example of such fundamental dichotomy is the description of a terrain by these two disciplines. In order to achieve a simple and tractable model of terrain, a typical algebraic definition of terrain will be as follows:

> *A terrain is a smooth, doubly continuous function of the form $z = f(x, y)$, where z is the height associated with each point (x, y).*

Further,

> *The local maxima or peak of the terrain is a point with a zero slope and a convex curvature.*

Other terrain features are defined similarly using morphometric measures. Most physical geographers will, however, find these definitions very restrictive because (i) they assume away the absence of some common terrain features such as lakes and overhangs crucial to certain applications, for example, runoff modelling; and (ii) natural terrain features cannot be localised to a point because a peak with zero local slope will really be the exception rather than the rule in nature. In physical geography, the description of the terrain surface and terrain features is more indicative than precise. Therefore, as long as the shape of the terrain around *an area* could be classified into terrain feature type, it is the responsibility of the geographer to locate the position of the peak on terrain based on his/her expertise. Figure 1.1 shows the difference between an algebraic topology (as a triangulated irregular network (TIN)) and physical geography (drainage network) description of a terrain surface.

It therefore follows that a combination of these two ways of describing a surface should provide a complete and robust approach for describing surfaces. In other words, a data structure that could explicitly describe both the structure (e.g. hills and valleys) and the form (e.g. *xyz* coordinates) of a surface will be an ideal digital representation of the surface. A more general form of this requirement was stated by Wolf (1993). Wolf regarded an efficient surface data structure to be one that contains both the geometrical information (e.g. coordinates, line equations) and the topological information on the geometrical data (e.g. neighbourhood relationships, adjacency relationships) of the surface. But as you read the book, it will be clear that the construction of a topologically consistent surface data structure is a non-trivial task because real surfaces seldom obey the constraints required by topological rules. At this stage, I would like to propose a difference between the terms *surface data structure* and *surface data model*. I think the term surface data structure should merely imply a format for storing the geometric and topological information (e.g. point heights and adjacency relationships) in a single construction. On the other hand, surface data model should be an extended version of

[1] Wolf (1993, p24) highlights the importance for exact definitions quoting Werner (1988) and Frank et al. (1986).

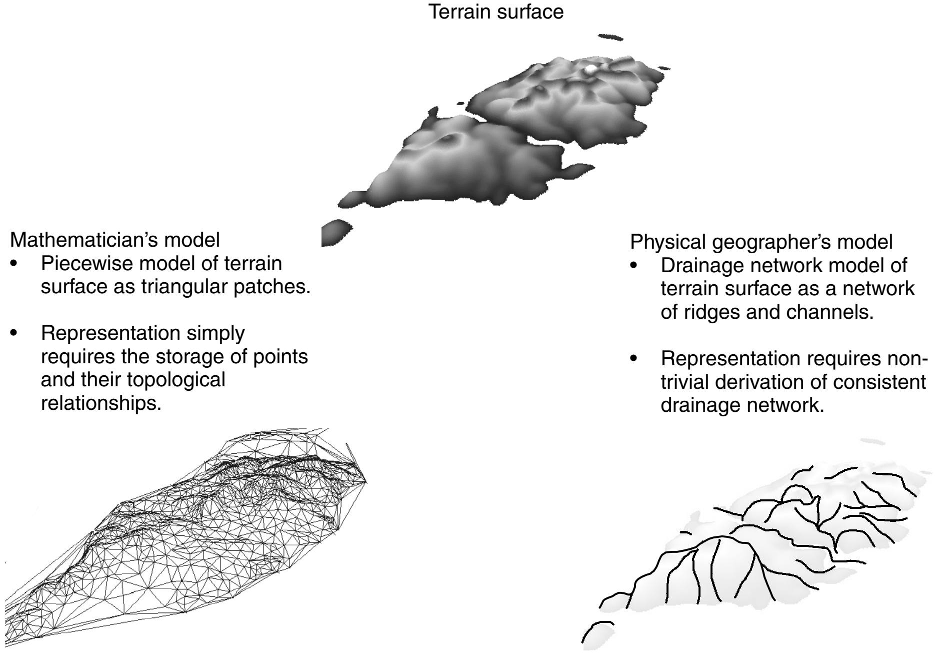

Figure 1.1 *Representation of the terrain surface into two different models depending upon the desired application*

the surface data structure in which additional metadata information characterising the surface (e.g. valleys, ridges, i.e. characteristic properties of surface) is also incorporated to produce a representation of the surface. In simple terms, a surface data model is a value-added product of the surface data structure and it explicitly represents the characteristics of the surface. Thus, all surface data models can be regarded as surface data structures but the opposite is not necessarily true.

1.2 TOPOLOGICAL DATA STRUCTURES FOR SURFACES

It is fairly straightforward to produce data structures that store the geometrical information about a surface. We simply need to collect certain points on the surface either on a regular lattice/grid or irregular locations. In fact, for many surface applications, only geometrical information is required for analyses. However, storing topological information has the following significant advantages:

- If we assume certain homogeneity in surface shape (e.g. smooth and continuous), using a topological data structure will reduce the number of points required to construct a surface. For example, by storing only certain morphologically important points (MIPs) (e.g. corners, inflexions) and their topological relationships, we could reconstruct the surface by means of interpolating between MIPs. Thus, the amount

of computer disk space required to store the surface will be reduced significantly. Helman and Hesselink (1991) reported 90% reduction in size on storing volumetric surface datasets in topological data structures.
- Topological relationships are a much more efficient way of accessing a spatial database, for example, sophisticated spatial queries such as clustering would be easily implemented on the basis of adjacency relationships.
- Topological data structures could provide a unified representation of the global structure of the surface. Thus, these data structures can be used for applications that require a uniform and controlled response from the entire surface such as morphing in computer graphics and erosion modelling in hydrology.
- Topological data structures will be useful for the visualisation of the structure of surfaces, particularly multi-dimensional surfaces. For example, Helman and Hesselink (1991) and Bajaj and Schikore (1996) reported that rendering of volumes surface datasets as skeleton-like topological data structures is more fast and comprehensible compared to traditional volume rendering.
- Bajaj and Schikore (1996) propose that topological data structures will be a simple mechanism for correlating and co-registering surfaces due to the embedded information on the structure of surfaces.

While the above benefits of topological data structures are applicable to all types of surfaces, it is uncertain which MIPs and topological relationships should constitute a universal surface topological data structure. Clearly, each surface should be characterised by MIPs suitable for a particular application. Many types of MIPs have been proposed by researchers in different disciplines and have been referred to by different names, for example, *landform elements* (Speight, 1976), *critical points and lines* (Pfaltz, 1976), *surface-specific features* (Fowler and Little, 1979), *symbolic surface features* (Palmer, 1984), *surface patches* (Feuchtwanger and Poiker, 1987) and *specific geomorphological elements* (Tang, 1992). The common aim of these classifications has been to provide a sufficient resemblance to the surface relevant to a particular application.

This book is primarily on the topological surface data structure called *surface network* (Pfaltz, 1976), which has been used in many disciplines because of its simple and universally applicable design. The book also discusses two other closely related data structures called the *Reeb graph* (Reeb, 1946) and the *contour tree* (Morse, 1968). I suppose some readers might be surprised to see the rather old lineage of the surface network. Surprisingly, however, these data structures have received little mention even in otherwise well-referenced texts (e.g. Koenderink and Van Doorn, 1998, Wilson and Gallant, 2000) despite a substantial, although I admit, irregular flow of research papers. This was indeed the main motivation behind this book. During the initial days when I was doing my Ph.D. on surface networks, I assumed that there was not much literature on surface networks, but, gradually, I started finding many works from all across the globe and from different disciplines, which was very encouraging. Hence, I decided to propose the book with the aim of putting together some key works on surface networks and the related data structures, so that future researchers could start from a single source.

Since each chapter in this book has a good introduction to the individual data structures, I will not define them here in detail. In this chapter, I will present instead the interesting history related to the development of these data structures followed by an overview of the chapters. In simple and general terms, the Reeb graph, contour tree, and surface network are graph-based surface data structures whose vertices are the local peaks, local pits, and local passes. The edges of the graph are the channels, connecting pits to passes, and ridges, connecting passes to peaks. Peaks, passes, and pits are together called the *critical points* of a surface, and ridges and channels are together called the *critical lines* of a surface. In my opinion, all the above surface data structures also qualify as surface data models because their construction is very much based on surface elements. Theoretically, any n-dimensional, smooth and doubly continuous surface can be represented as a surface network (Pfaltz, 1976); however, the most common implementations are limited to two- and three-dimensional surfaces.

The primary origin of these data structures lies in the realisation of the critical points and critical lines of the surface. Critical points can be defined as characteristic local surface features that are common to all surfaces and contain sufficient information to construct the whole surface, thus taking away the need to store each point on the surface. Critical points have a local zero slope, that is, $\mathrm{d}z/\mathrm{d}x = \mathrm{d}z/\mathrm{d}y = 0$, and three such critical points of the surface are local maxima ($\partial^2 z/\partial x^2 > 0$, $\partial^2 z/\partial y^2 > 0$ (also called *peaks, summits*), local minima ($\partial^2 z/\partial x^2 < 0$, $\partial^2 z/\partial y^2 < 0$) (also called *pits, immits*), and passes ($\partial^2 z/\partial x^2 > 0$, $\partial^2 z/\partial y^2 < 0$ or vice versa) (also called *knots, saddles, bars*). These critical points have been referred to in physical geography as the *fundamental topographic features*. In physical geography, the derivation of these features has traditionally been based on the overall shape of contours (i.e. using a regional context) on a topographic map rather than local morphometric properties. For example, a peak is identified as the centre of a closed highest contour bounded by lower contours. At any point on the surface, a line following the steepest gradient is called a *slope line* (also called *topographic curves* and *gradient paths*). Critical lines are a special pair of slope lines that originate and terminate at critical points. There are two types of critical lines, namely ridge line and channel line. A ridge line originates from a peak and terminates at a pass, and a channel line starts from a pass and terminates at a pit.

The concept of *critical points* of a surface and critical lines was proposed as early as the mid-nineteenth century by the mathematicians (De Saint-Venant, 1852 reported by Koenderink and van Doorn, 1998; Reech, 1858 reported by Mark, 1977). In physical geography, Cayley (1859), on the basis of contour patterns, first proposed the subdivision of the topographic surface into a framework of peaks, pits, saddles, ridge lines, and channels. Maxwell (1870)[2] extended Cayley's model and proposed the following relation between the peaks, pits, and passes:

$$\text{peaks} + \text{pits} - \text{passes} = 2 \tag{1.1}$$

This relation was later proved by Morse (1925) using differential topology and is also known as the *mountaineer's equation* or the *Euler–Poincaré formula* (Griffiths, 1981 reported by Takahashi et al., 1995). Maxwell also described the partition of the

[2] The anxiety with which Maxwell presented his paper is quite amusing. His note to the editor of the journal reads "An exact knowledge of the first elements of physical geography, however, is so important, and loose notions on the subject are so prevalent, that I have no hesitation in sending you what you, I hope, will have no scruple in rejecting if you think it superfluous after what has been done by Professor Cayley".

topographic surface into hills (areas of terrain where all slope lines end at the same summit) and dales (areas of terrain where all slope lines end at the same immit) on the basis of the fundamental topographic features.

The earliest graph–theoretic representation of the topological relationships between the critical points of a terrain is the Reeb graph (Reeb, 1946; reported by Takahashi et al., 1995). Reeb graph basically represents the splitting and merging of equi-height contours (i.e. a cross section) of a surface as a graph. The vertices of the graph are the peaks, pits, and passes because the contours close at the pits and the peaks, and split at the passes. Consequently, the edges of the Reeb graph turn out to be the ridges and channels.

In a significant related development in mathematics, Morse (1925) derived the relationship between the numbers of critical points of sufficiently smooth functions (called *Morse functions* under certain constraints), which is known as the *Critical Point Theory* or *Morse Theory*. The generic nature and wide applicability of Morse Theory led to an expansion in the interest in the critical points of surfaces amongst various disciplines.

Warntz (1966) revived the interest of geographers and social science researchers in critical points and lines when he applied the Maxwell's "hills and dales" concept for socio-economic surfaces, referred to as the *Warntz network* (the term apparently coined by Mark, 1977).

A data structure identical to Reeb graph is the contour tree (Morse, 1968), also called the *surface tree,* by Wolf (1993). The contour tree represents the adjacency relations of contour loops. The treelike hierarchical structure develops because of the fact that each contour loop can enclose many other contour loops but it can itself be enclosed by only one contour loop. As is evident, the contour tree is the same as the Reeb graph except that it is separated by two decades. Kweon and Kanade (1994) proposed another similar idea called the *topographic change tree*. As in the case of the Reeb graph, the vertices of such a contour tree are the peaks, pits, and passes.

Pfaltz (1976) proposed a graph–theoretic representation of the Warntz network called *surface network* (Mark, 1977 used the term Pfaltz's graph). While the topology of the Pfaltz's graph was based on the Warntz network, Pfaltz added the constraint that the surface will have to be a Morse function. Since Pfaltz was in the computer science field, his work attracted the attention of researchers in three-dimensional surfaces such as in medical imaging, crystallography (e.g. Johnson et al., 1999, Shinagawa et al., 1991), and computer vision (e.g. Koenderink and Doorn, 1979). He also proposed a homomorphic contraction of the surface network graph to reduce the number of redundant and insignificant vertices. Along similar lines, Mark (1977) proposed a pruning of the contour tree to remove the nodes (representing contour loops) that do not form the critical points, i.e. the vertices of the contour tree, and called the resultant structure the *surface tree*. This essentially reduces the contour tree to the purely topological state of a Pfaltz's graph. It is easy to realise that the Reeb graph, Pfaltz's graph, and surface tree are fundamentally similar and are actually inter-convertible (Takahashi et al., 1995).

Wolf (1984) extended Pfaltz's graph by introducing more topological constraints in order that it be a consistent representation of the surface. He proposed assigning weights to the critical points and lines to indicate their importance in the surface and thus he proposed the name *weighted surface network* (WSN) for the Pfaltz's graph. He demonstrated new weights-based criteria and methods for the contraction of the surface networks. Later, Wolf suggested that in order to visualise the WSN for cartographic purposes and to make it useful for spatial analyses, the vertices could be assigned metric coordinates (Wolf, 1990). The resultant representation is termed *metric surface network* (MSN).

Recent works have mostly focused on the automated extraction of surface networks from raster (Wood and Rana, 2000, Schneider, 2003) and TIN (Takahashi et al., 1995), which will be discussed in detail in the following chapters.

1.3 OBJECTIVES OF THE BOOK

As mentioned earlier, while there is an extensive literature on other topological surface data structures (e.g. TIN and quadtrees by Samet, 1990a, b, van Kreveld et al., 1997), the topics of surface network, contour tree, and Reeb graph, proposed more than three decades ago, have only had irregular and scattered reports of the research on them. This gap is the main motivation of this book. Despite the unique inter-disciplinary scope of these data structures, there is generally a lack of awareness about their complete potential amongst modern researchers. The book is also timely because publications demonstrating all the promised potential of these data structures for practical applications such as visualisation of large datasets (e.g. Takahashi et al., 1995, Bajaj and Schikore, 1996), fast contour extraction (e.g. van Kreveld et al., 1997), generalisation and compression of surfaces (e.g. Rana, 2000a,b, Kraus and Ertl, 2001), and spatial optimisation (e.g. Rana, 2003a, Kim et al., 2003) have finally started coming out.

The objective of this book is to bring together some key earlier and modern researches on these data structures to rejuvenate these topics and fuel ideas for future research. Some of the important features of this book are as follows:

- Popular morphometric feature extraction algorithms, useful in drainage analysis, computer vision, and information organisation, are described with practical examples with links to the directions for future research.
- A comprehensive and condensed treatment of these data structures, unpublished elsewhere, has been made available to the reader.
- This is a multidisciplinary area of research and this volume provides accessible content to practitioners in a range of fields.

1.4 OVERVIEW OF THE BOOK

The book is divided into two main parts. The first part deals with concepts, automated extraction, and issues related to the Reeb graph, contour tree, and surface network. The second part of the book presents a number of applications of these topological surface data structures.

The primary audience for this book are postgraduates and professionals. As can be clearly seen from the diverse background of the authors, this book will appeal to members of a number of disciplines such as Geographic Information Science, Computer Science, and other sciences involved in the morphometric analysis of surfaces.

Owing to the very practical nature of this book and its deliverables, I am tempted to believe that the commercial organisations, particularly those involved in research and development of solutions, would be keen to explore the ideas presented in this book. I feel that many of the ideas are still in their early phase of development with promising outputs, which could also provide topics for postgraduate and higher-level research.

All chapters contain a basic to intermediate discussion on the data structures; hence each chapter is quite self-contained and could be read independently. The following brief descriptions should give a general idea about each chapter so that the reader can decide to follow the book at his/her own pace and order.

1.4.1 Part I – Concepts and implementation

Part I starts with a chapter by Gert Wolf (Chapter 2), who extended the work of Pfaltz (1976). Wolf describes the construction of the surface network graph and how weights could be assigned with edges and nodes to indicate their importance for both the macro- and the micro-structure of the underlying surface in great and lucid detail. The graphs thus obtained – termed *weighted surface networks* (WSNs) – represent a powerful tool for characterising and generalising topographic surfaces. The technique of generalising surface networks using two graph–theoretic contractions is explained. The chapter then proposes an improvement of the purely topological WSNs, by attaching (xy)-coordinates to the nodes resulting in an MSN. Finally, the generalisation process of a real landscape taken from the Latschur region in Austria is shown.

Shigeo Takahashi (Chapter 3) describes algorithms for extracting surface network and Reeb graphs. The fully automated algorithms extract these data structures from an input TIN by simply generating a linear interpolation over the elevation samples instead of computationally expensive higher-order interpolations. Furthermore, the extracted features are correct in the sense that they maintain topological integrity (e.g. the Euler–Poincaré formula for critical points) inherited from the properties of smooth surfaces. The present algorithms are robust enough to handle troublesome datasets such as noisy or stepwise discrete samples, and such robustness is demonstrated with several experiments on real terrain datasets. The resultant configuration of critical points allows us to tackle other geographical information systems (GIS) related issues, which is also discussed in this chapter.

Bernhard Schneider and Jo Wood (Chapter 4) describe two methods for the identification of MSNs from raster terrain data. This chapter includes a discussion of some of the computational issues that have previously prevented automated construction of MSNs, and presents two new automated methods that may be applied to raster terrain data. The chapter first describes a simple and robust bilinear polynomial approximation method followed by an advanced bi-quadratic polynomial method.

Marc Kreveld and others (Chapter 5) describe an efficient method to construct the contour tree and to obtain seed sets that are provably small in size. The contour algorithm can be used for regular and irregular meshes.

Silvia Biasotti and others (Chapter 6) introduce and describe the concept of extended Reeb graph (ERG). First of all, a useful overview of the definition of critical points and Morse complexes for smooth manifolds is given. It is followed by a description of some existing methods that extend these concepts to piecewise linear 2-manifolds, focusing, in particular, on topological structures as surface networks and quasi-Morse complexes, available for analysis and simplification of triangular meshes. To avoid the dependency of such structures on the locality of the critical point definition, they propose to consider critical areas and influence zones instead of usual ones, and give their formal definition. They base the ERG representation on this characterization and compare it with the surface network structure. Finally, they describe the ERG structure and its construction process, also introducing several examples from a real terrain from New Zealand.

1.4.2 Part II – Applications

This is perhaps the more interesting part of the book and has taken the most effort to put together.

The analysis of urban population distributions has been one of the central subjects of human geography since Clark's pioneering study in the 1950s (Clark 1951, 1958). Following Clark, most initial studies dealt with population distributions in terms of a population density with respect to the distance from the centre of a city (i.e. one-dimensional distributions). These studies were extended to two-dimensional distributions by use of the trend surface analysis. This analysis, however, has difficulty in interpreting the estimated coefficients. To overcome this difficulty, Okabe and Masuda (1984) proposed a method for analysing population surfaces in terms of surface networks. In the first chapter of this part, Atsuyuki Okabe and Atsushi Masuyama (Chapter 7) propose a new method for measuring topological similarity between activity surfaces in terms of modified counter trees. This method has an advantage in measuring not only local similarity but also global similarity. They develop an algorithm for implementing the proposed method and apply the method to urban population surfaces in Japan and show topological similarity among Japanese urban population surfaces.

In Chapter 8, Valerio Pascucci discusses a technique that reduces the user responsibility to infer implicit information present in the data, by computing topological features like maxima, minima, or saddle points and determining their relationships. He first introduces the formal framework, based on Morse theory and homology groups, necessary to analyse the critical points of a scalar field and to classify the shape of its level sets. He discusses a set of algorithms that map this mathematical formalism into an efficient pre-processing of the data. The chapter concludes with a discussion on user interfaces that present intuitively the computed information and a demonstration of examples from real scientific datasets such as subatomic particle collision.

Martin Kraus and Thomas Ertl (Chapter 9) show how to apply the concepts of scalar topology to the volume visualisation of structured meshes. This chapter discusses the role of topology-guided downsampling in direct and indirect volume visualisation. While most algorithms related to scalar topology work on unstructured meshes, topology-guided downsampling is a recently published downsampling algorithm for structured meshes, for example, Cartesian grids. The main goal of this technique is to preserve as many critical points as possible, that is, to preserve as much as possible of the topological structure of the original scalar field in the downsampled scalar field.

After presenting this downsampling algorithm, they discuss its application in indirect volume visualisation with isosurfaces and in direct volume rendering. Particular emphasis is made on the interplay with recent developments in direct volume rendering, namely the use of programmable per-pixel shading for pre-integrated volume rendering. They also show how to employ topology-guided downsampling in the generation of hierarchical volume data that can be rendered with the help of programmable per-pixel shading.

Jason Dykes and I (Chapter 10) present a surface networks–based features modelling approach for the visualisation of dynamic maps. This chapter extends the proposals of Valerio Pascucci (Chapter 8). Despite their aesthetic appeal and condensed nature, dynamic maps are often criticised for the lack of an effective information delivery and interactivity. We argue that the reasons for these observations could be due to their information-laden quality, lack of spatial and temporal continuity in the original map data, and a limited scope for a real-time interactivity. We demonstrate, with the examples of a temporal and an attribute series of a terrain and a socio-economic surface, respectively, how the re-expression of the maps as the surface network, spatial generalisation, morphing, graphic lag, and the brushing technique can augment the visualisation of dynamic maps.

Traditional surface texture parameters take a statistical approach to characterisation using only the height and location information of the individual measured points. Many applications, that is, lubrication, paintability of a surface, anodized extruded aluminium, and so on, require the characterization of features such as peaks, pits, saddle points, ridge lines, course lines, and so on, and the relationships between these features. That is a pattern recognition approach. In Chapter 11, Paul Scott proposes a topological characterisation of surface texture. This is based on the topological relationships between critical points and critical lines on the metrological surface and incorporated into a WSN. He then removes the predominant insignificant critical points, caused by measurement noise, and Gert Wolf's graph contractions (see Chapter 2), and then assesses the relationships between the significant surface features.

In Chapter 12, Jeremy Morley and I propose the advantages of using the fundamental topographic features forming the surface network of a terrain, namely the peaks, pits, passes, ridges, and channels, as the observers or the targets in visibility computation. We demonstrate that considerable time can be saved without any significant information loss by using the fundamental topographic features as observers and targets in the terrain. The optimisation is achieved because of a reduced number of observer–target pair comparisons, which we call the *Reduced Observers Strategy* and *Reduced Targets Strategy*. The method has been demonstrated for a gridded digital elevation model. Owing to this selected sampling of observers in the terrain, there is an underestimation of the viewshed of each point. Two simple methods for assessing this uncertainty have been proposed.

In the conclusion of the book, I raise a number of unresolved issues related to the data structure model, their automated generation, and generalisation.