

Geometric Analogies and Topographic Maps

Diarmuid O'Donoghue¹, Alan Loughlin¹, Amy Bohan², Mark T. Keane²

¹National University of Ireland, Maynooth,
Department of Computer Science / National Center for Geocomputation
Co. Kildare, Ireland
diarmuid.odonoghue@nuim.ie

²University College Dublin,
Department of Computer Science,
Dublin, Ireland.
{amy.bohan, mark.keane}@ucd.ie

Abstract

In this paper we describe the application of geometric analogies to the GeoComputation domain. We describe geometric analogies that include attributes (patterns and color). We identify two variants on an attribute matching algorithm that are required to solve these problems. Next we describe how we use one of these algorithms to identify targeted structures within a topographic database (map). We describe a how a problem in topographic datasets is solved using this technique, detailing results on four different datasets. Finally, we outline a variety of other problems that are being addressed, based on the same algorithm.

1 Introduction

At present, most maps of the world have now been transformed into digital media opening up new vistas for the flexible use of their contents by business, government and private individuals. However, the maintenance of such topographic data presents us with a new set of challenges that could benefit from automated solutions. In this paper, we present a novel solution to maintaining such maps using the analogical mapping techniques - see [Keane *et al*, 1994; Veale and Keane 1997; Falkenhainer *et al*, 1989; Ferguson and Forbus, 2000].

1.1 Problems Maintaining the OS MasterMap

The Topological Map of Great Britain (OS MasterMap) contains over 450 million non-overlapping polygons, describing ground cover across the country. Each polygon is composed of a number of lines defining the boundaries between polygons. Each recorded feature (line, polygon etc) is uniquely identified by a 16 digit Topographic-ID (ToID) number and the entire map contains over 2.5 billion ToIDs. Additionally, each polygon is categorized into one of ap-

proximately thirteen themes, including; *road, rail, building, inland waterway, made land, unmade land, roadside etc.*

There is a significant human work overhead associated with maintaining such a vast and ever-changing repository of information: more technically known as Data Quality Improvement (DQI). DQI is required because of several specific problems that arise in digital maps, including:

- Unclassified and Mis-classified polygons: Some polygons are either unclassified, or may be incorrectly classified.
- Sub-classification of polygons: Polygons can be further classified into sub-categories of the original categories, such as *building* having the sub-class *dwelling*, which is further divided into *terraced-house, semi-detached-house, and detached-house*. Work is currently under way on developing a formal ontology of topographic concepts.
- Obscured polygon segments: Topographic maps are 2D representations of 3D information. When one object obscures another, the obscured object becomes segmented. These obstructions must be detected and corrected.

Currently, these repeatedly occurring problems are mainly fixed by hand. Though there have been some attempts to automatically solve such problems (e.g., using computer vision and other techniques for classification [Keyes and Winstanley, 2001]), few automated solutions have been explored.

We believe that this GeoComputation domain is a very fertile ground for the application of AI techniques. We present a novel approach to some DQI problems by re-casting them as proportional analogies on geometric shapes (i.e., A:B::C:D IQ-test type analogies). The various problems that exist with map polygons can be cast as comparison problems where some partial description is completed by comparison to another complete description that is analogous. In the next section, we will outline proportional, geometric analogies before showing how they can be used in DQI.

1.2 Creative Interpretation in Maps

Before presenting our solution to the problems listed above, we make a number of observations on how people interpret topographic maps. For example, map users regularly encounter the obscured-polygon problem wherever a *road* passes beneath a *bridge*. This is because the bridge ensures that part of the road cannot be detected on the aerial photographs that are regularly used to make maps. Therefore the *road* appears to be obstructed by the *bridge*. But even novice map users have no difficulty in imagining that the *road* continues beneath the *bridge*. However, current geocomputation applications cannot support such simple inferences (route-planning applications typically use a separate “Transportation Network” layer that is manually derived from the topological map).

Similarly, when examining a map that depicts polygons of the category *building*, many of these structures are easily identified as *houses*. Regular collections of similarly sized suburban (semi-detached) *houses* can easily be seen, each with a *garden* and access to a *road*. So, while the exact type of *building* is not recorded on the map, the users understanding of the map is often much richer than the recorded data. Our model enriches the data that is stored in the electronic map, so that it more closely resembles what people are able to perceive.

One can view this as a relatively simple form of creativity where users apply their understanding of the real-world to the information presented on the map. This creates an understanding of the map that is *much* more useful than a strict reading of the presented data.

This paper describes a model that overcomes the problems listed above. This model displays some characteristics that are often associated with creativity, such as being directed, novel and useful (Ritchie, 2001). It is directed because its behavior is driven by some goal - such as repairing obstructions. It makes novel use of analogical comparisons to solve these problems - like identifying suburban houses. The usefulness of this model is that it enables information on a map to be elaborated, allowing other computational processes to make better use of the recorded data.

2 Geometric Proportional Analogies & Map

2.1 Geometric, Proportional Analogies

Geometric, proportional analogies are comparisons between two collections of geometric figures, called the *source* and *target* domains. A geometric analogy is of the form $A:B::C:D$, where A, B & C are given and D is derived from applying the transformation obtained from A to B, to C. We read such an analogy as A is-to B as C is-to D. Typically, the *source* domain (A:B) identifies some transformation(s), which must then be applied to C, yielding D (See Figure 1). For example, the analogy in Figure 1 centers on partitioning the polygon to the right of part A to produce part B. This partitioning transformation must then be applied to part C.

There are two key points to note in geometric, proportional analogies. First, the *transformation* found in the

source domain (i.e., the change between A and B). Second, the *mapping* between to the two domains: parts A and C are used to identify the inter-domain mapping that will yield part D. It is by mapping the transformation appropriately from the source domain that the missing target term (D) is found.

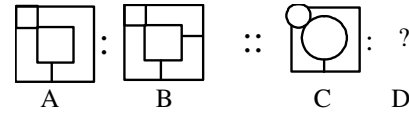


Figure 1: A Simple Geometric Analogy

2.2 Theoretical Background

The first computational model for solving geometric analogies was the ANALOGY model [Evans, 1967]. It solved geometric analogies involving plain figures - with objects possessing no color or pattern information. Evans' model also took graphic images as input and made use of the geometry of the presented figures. Additionally, Evan's Analogy model selected the best D from a given list of alternative solutions. Our analysis relies on more recent work in analogical mapping based on the Structure Mapping framework [Gentner, 1983; Falkenhainer *et al*, 1989; Keane *et al*, 1994].

In solving these proportional analogies we concentrate on topology rather than geometry; that is, we ignore the shape of the polygons and concentrate on how they are arranged [Bohan and O'Donoghue, 2001] (see sections 3 and 4 for justifications).. Our model uses a symbolic representation of the source and target domains. So, Figure 1 might be characterized by the following collection of predicates based on the labeling shown in Figure 2:

$line-adjacent(1,2)$ $line-adjacent(2,3)$
 $line-adjacent(3,4)$ $line-adjacent(4,1)$
 $point-adjacent(1,4)$

Having identified and labeled the distinct objects in the domain and characterizing their relations using $line-adjacent(x,y)$ and $point-adjacent(y,z)$ predicates, these representations are treated as geometric analogies. So, we identify the *mapping* between the descriptions of part A and part C, identifying the *structural isomorphism* between them. By concentrating on the topological relations in the mapping, different shaped polygons can be placed in correspondence between parts A and C. In this way, the top-left square of part A is mapped to the upper-left circle of part C (see Figure 1)

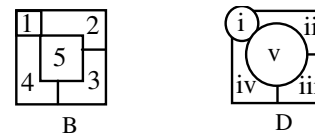


Figure 2: Labeling objects in the source and target

The description of the source domain includes the A-to-B *transformation* (unlike [Davies and Goel, 2001] we do not use a separate vocabulary to represent transformations, rely-

ing instead on different combinations of the same predicates). In Figure 1 the transformation involves the insertion of an extra polygon by splitting an existing polygon. Many geometric analogies involve transformations that insert and delete polygons from the source (and thus the target). Therefore the solution of the geometric analogy in Figure 1 is the application of this “insertion” transformation to the mapped equivalent in C – yielding D in Figure 3.

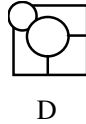


Figure 3: Solution to the Geometric Analogy in Figure 1

Transformations that delete polygons may be useful in “generalizing” topographic maps – that is, producing large scale maps from more detailed data sets [Ware *et al*, 2003]. However, work is still in its early stages.

2.3 Attributes in Geometric Analogies

In this paper we explore the application of geometric analogies that involve attribute matching and attribute transformations (see Figure 4). These attributes are vital in solving these analogy problems though they often receive less attention in more complex analogies [see Gentner, 1983; Falkenhainer *et al*, 1989]. While GeoRep [Ferguson and Forbus, 2000] and Galatea [Davies and Goel, 2001] also identify analogies in diagrammatic reasoning, neither involve similar attributes in this analogy process.

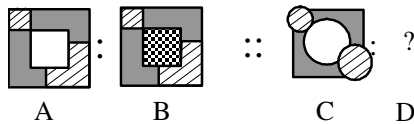


Figure 4: A Geometric Analogy with Attributes

We add the attribute information about each polygon to the predicate description (above). So part A of Figure 4 will be represented by similar predicate information as in Section 2.2, plus the following attributes: *striped(1)*, *gray(2)*, *striped(3)*, *gray(4)*, *plain(5)*.

2.3.1 Attribute Transformations & Matching

The source and target of Figure 4 will be represented by similar predicate information as in Section 2.2. However, the source domain of Figure 4 now defines a collection of attribute transformations, including:

plain(5) -> checkered(5)

The complete set of attribute transformations for this problem is:

striped -> striped

gray -> gray

plain -> checkered

Of course, we still identify the mapping between the source and target domains – and as these domains include attribute transformations then we must also identify the mapping between these attribute transformations. In the rest of this

section we show how these attribute transformations are used.

The addition of this attribute information complicates the solution process because there are multiple ways of identifying the mapping that occurs between the source and target attribute transformations. We define *Attribute Matching* as the process of determining the attribute changes in the transformation and mapping process. These matches can manifest themselves in two main ways: Global and Local attribute matches.

2.3.2 Global Attribute Matches

We define a *global attribute match* as a match with a 1-to-1 correspondence between the attribute transformations of the source and target domains. Figure 4 uses global attribute matching, to generate the solution depicted in Figure 5.

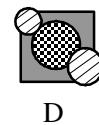


Figure 5: Solution to the Geometric Analogy in Figure 1

Table 1: Results of Different Attribute Matching algorithms

Attribute Transformation	
Global (Figure 4)	Local (Figure 6)
<i>striped -> striped</i>	<i>striped(1) -> striped(i)</i>
<i>gray -> gray</i>	<i>striped(3) -> dotted(iii)</i>
<i>plain -> checkered</i>	<i>gray -> gray</i>
	<i>plain -> checkered</i>

In Figure 4, *striped->striped* maps to *striped->striped*, *gray->gray* maps to *gray->gray* and *plain->checkered* maps to *plain->checkered* ([Bohan and O’Donoghue, 2001] describe examples of geometric analogies where non-identical attributes are matched). So in fact, not only is there a 1-to-1 correspondence between these attribute transformations, but the mapped attribute transformations are also identical. This is the type of attribute matching required by many geometric analogies that is useful here.

2.3.3 Local Attribute Matches

A *local attribute match* one where there is *no* 1-to-1 correspondence between the attribute transformation of the source and the target. Thus, in local attribute matches there is a 1-to-n correspondence between some of the attribute transformations.

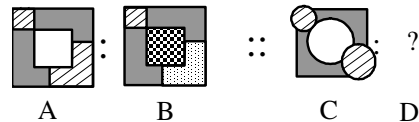


Figure 6: Local attribute matches

The analogy depicted in Figure 6 uses the same topology as that in Figure 4, but identifying the attribute matches is more complex. The source domain (parts A and B) identi-

fies two possible transformations for the *striped* attribute – *striped* and *dotted* (see column 2 of Table 1). These attribute transformations have to be dealt with separately.

$striped(1) \rightarrow striped(i)$
 $striped(3) \rightarrow dotted(iii)$

So for the object *i* in the target the *striped*->*striped* transformation is applied, while for the object *iii* the *striped*->*dotted* transformation is applied (see Figure 7).



D

Figure 7: The solution to Figure 6

Of course, all global attribute matches can be solved with the global attribute matching algorithm – although this would involve a large amount of duplicate information. However, solving the simpler problems with the more complex algorithm would fail to take account of the inherent simplicity in these problems.

In practice, we have found that most DQI problems can be solved using global attribute matches, though the possibility of more complex local attribute matches occurring should not be ruled out

3 Maps and Analogies

Topographic maps are stored as a database of interconnected features, representing information at the levels of *point*, *line* and *polygon*. We use geometric analogies as a basis for identifying specific problem situations (as discussed in Section 1) within a topographic database (map).

The map segment illustrated in Figure 8 is centered on a *road* with two *roadside* polygons, all of which cross an underlying *river*. The data collection process often fails to represent the obscured segment of the river, thereby resulting in a segmentation of the river into multiple, seemingly independent polygons. These segmentations mean that most topographic datasets do not represent complete structures like *rivers* (*roads* and *railways*) because they become segmented into many separated polygons. Such obscurities are a particular problem for topographic data and manually correcting them is laborious and error-prone. Obscurities are one of the problems that analogies can solve.

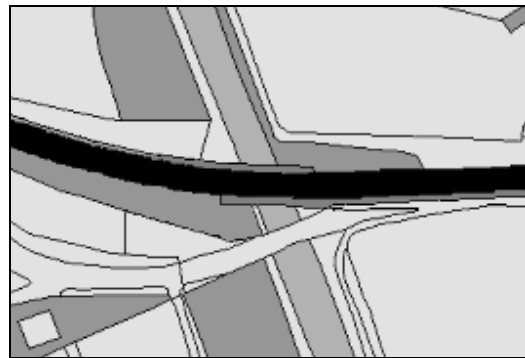


Figure 8: Topographic Map with a Road (running horizontally across the Figure) obscuring a River (vertical)

3.1 Problem Representation

The largest structure stored in a topographic map is the polygon. However, to fit in with our geometric analogies, we require larger structures that have some internal domain topology. Thus we introduce a new level of resolution for dealing with topographic data. Each domain is a collection of multiple polygons that we call a *locality*. Each *locality* consists of one central polygon plus all polygons that are immediately adjacent to that central polygon. Thus, the entire map is covered by numerous over-lapping localities.

Each locality is described by three types of information. Firstly, a locality has a unique identifier (ToID) for each polygon in that locality. Secondly, it records the categories of all polygons contained in it. Finally, each locality records the topological structure between all polygons in that locality, using the *line-adjacent* and *point-adjacent* predicates (that is useful in the “Problem Correction” described in section 3.4). The representation of a locality creates a collection of predicates and attributes similar to that listed earlier.

3.2 System Architecture

The topographic map that contains occluded polygons is loaded into the “ArcInfo” GIS (Geographic Information System) application program. An ArcInfo script was written to generate the locality descriptions. Each locality description is then passed to the structure matching program CSM (Cartographic Structure Matching) [Winstanley *et al*, 2000; O’Donoghue and Winstanley, 2001; O’Donoghue *et al*, 2003] that compares each locality description against known problem situations. When an isomorphic match is detected between the two descriptions, and when the categories of all mapped, polygons are identical (forming a local attribute match) – then a problematic locality has been identified, and is therefore dispatched to the appropriate repair process. Because the two relations used in describing localities are commutative, the CSM mapping model ensures that all valid combinations of argument order are investigated, in order to find an optimal mapping.

3.3 Problem Detection

We will begin by describing how we solve the problem of occluded *roads*, *riders* and *railways* using the geometric analogy approach. We identify the following seven different types of occlusion, namely:

- 1) Road over River
- 2) Structure over River
- 3) Rail over River
- 4) Road over Road
- 5) Rail over Road
- 6) Path+road+path over River
- 7) Path+road+path over Rail

Each of these 7 problem situations are also described by a special locality, recording the polygons, their categories and the topology formed by these polygons. These 7 descriptions form a simple case-base that serves to identify these specific problem structures within the topographic map.

Early testing revealed that occasionally, false obscurities were being detected at opposite ends of the obscuring polygon. To overcome this problem we supplement the CSM match with a simple distance metric to ensure the width of the obscured polygon is below some maximum threshold. This is the only geometric feature used to detect obscurities.

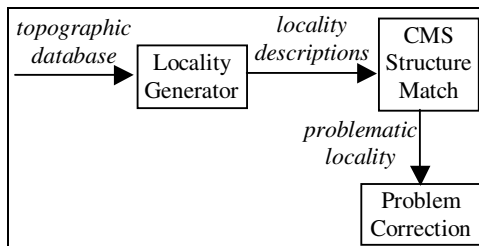


Figure 9: System Architecture

3.4 Problem Correction

Having identified the problem structure we can now repair them. This correction process corresponds to the A-to-B transformation of the geometric analogies. Correcting the problems detected with the structure mapping approach is actually a very straight-forward process. This is partly because the topologically sensitive Problem Detection process also identifies which are the obscuring polygons and which are the obscured polygons. That is, the seven problem situations that identify both the obscuring and obscured polygons – and because we have an isomorphic mapping between this information and the locality description, we therefore know the corresponding information for the problem locality.

Armed with this information, an augmented locality description is returned to the ArcInfo application program. This identifies the intersection between the *road* and each of the adjacent *river* polygons. Generating these line-segments is greatly facilitated by the ToID numbers that uniquely identify each polygon in the map – and thus in our problematic locality. ArcInfo generates two line-segments that correspond to two sides of the obscured river polygon. It is then an easy process to generate two straight-lines to represent

the missing sides, and thereby generate a new polygon that corresponds to the obscured segment of the river.

While two sides of the obscured polygon are inferred from the topology of the *locality* description, there is no guarantee that obscured edges will be straight lines. However, in practice we have found that the obscured polygons are sufficiently accurate for most requirements. The crucial factor is that the previously disconnected *river* polygons are reconnected and the river can now be treated as a composite object.

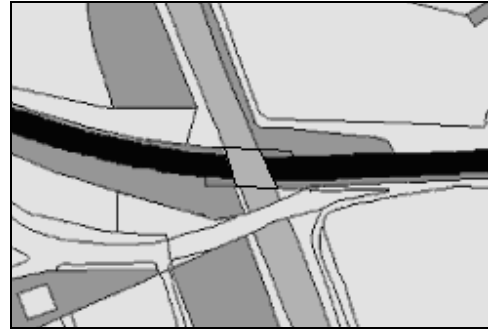


Figure 10: Reinserting the Obscured Polygon

4 Testing

We tested the polygon generation process on three different subsets of OS Master Map, depicting a portion of the regions of *Moffat*, *Port Talbot* and *Birmingham*. Each of these subsets of the MasterMap Topological Layer, represents a region of a few square kilometers upon which to test our solution. These map segments gave us a total of 43,000 polygons covering *urban*, *suburban*, *industrial*, *parkland* *rural*, and *mountainous* regions – each region represents a different challenge to the application, having a different distribution of polygons from the 13 basic categories.

4.1 Results

We ran the “Obscured Polygon Insertion” application on the three topographic datasets.

Table 2: Results summary for the different map regions

Region	Description	Number of polygons	Identified Obscurities
Moffat	Mountain, rural, town	11293	47
Port Talbot	Rural, suburban, industrial	5198	10
Birmingham	Urban, suburban, parkland	26632	14

These same maps were presented to two human reviewers who were asked to manually identify all obscuring bridges on the same maps. The reviewers were given a printout of the map on which to locate the bridges. As well as being given the printed map, the reviewers had access to the map through the “ArcInfo” application. (This allowed the reviewer, for example, to view only the road themes and this

made identifying bridges a very straight forward task). Additionally, the reviewers were given as much time as required to identify all obscuring bridges. This process identified all obscuring bridges in the three maps, allowing us to examine the accuracy of our repair process.

As can be seen from Table 3 the application proved very accurate, generating 100% of the obscured polygons in many situations. Additionally, in all cases where the obscuring bridge was identified, the newly inserted polygon had the correct location, dimensions and was assigned to the correct category (theme).

5 Emerging Applications

Our geometric analogy solution is also being applied to a number of other problems. Three of these problems are briefly discussed in this section.

Table 3: Accuracy of Obscured-polygon Insertion

Problem	Topographic Database		
	Moffat	Port Talbot	Birmingham
	Accuracy%	Accuracy%	Accuracy%
Road over River	85	66	0
Structure over River	100	100	100
Rail over River	100	100	100
Road over River	66	100	100
Rail over Road	50	100	100
Path+road+ path over River	58	22	25
Path+road+ path over Rail	100	70	100

5.1 Correcting Mis-Classification

Correctly classifying all polygons in a topographic database is a major concern as it directly impacts on the usefulness of maps to the end user. Most automated classifiers are based on the description of individual objects – examining their size, border length and geometry. But, some polygons shapes are particularly ambiguous and object-based classifiers become unreliable [Keyes and Winstanley, 2001]. Detecting and correcting mis-classifications is, therefore, an important task to automate [O’Donoghue and Winstanley, 2001].

The central polygon in Figure 11 (below) is an example of a polygon that can be incorrectly classified by object-based classifiers - as its shape and size are almost identical to those of a typical dwelling. Again, taking the context into account through the use of *locality* matching, we can see that this polygon is easily identified as a *road-junction* polygon (not a building).



Figure 11: An Ambiguously-Shaped Central Polygon

5.2 Sub-Categorization

Currently the topographic maps produced by the OS UK, categorize polygons into one of 13 different classes (or themes). However work is ongoing to develop a formal ontology containing many more detailed classes. We are developing an application to identify sub-categories of many of the 13 basic themes.

Our application categorizes several classes of polygon into a number of sub-categories. Many of the following are related to dwellings, which are of particular interest to market research agencies, which make use of topographic databases.

- 1) Building > dwelling > {detached-residence, semi-detached-residence, terraced-residence}
- 2) Building > out-house (associated with a dwelling)
- 3) Road > {road-segment, cul-de-sac, T-junction, X-junction}

However, this work is still in its early stages.

5.3 Composite Object

Composite objects consist of collections of individual polygons. For example a *homestead* may consist of a *dwelling* with surrounding *garden* and *out-houses*. Identifying such composite objects can improve a maps’ usefulness to many user groups. Large features like rivers and roads consist of individual polygon segments. Propagating (say) the name of a river to all adjacent segments makes use of the incremental mapping process [Keane, 1990]. However, this is still in progress.

6 Conclusions

We show how geometric, proportional analogies that use attributes (e.g., color, pattern information) can be used to solve many maintenance problems in digital maps. Two different attribute matching algorithms were discussed, that are required to generate solutions to these problems. This attribute information operates “on top of” the topology of the presented polygon information.

We described how detecting some problem situations in topographic maps, can be treated as an instance of solving a geometric analogy that includes attribute information. We examined the results obtained by running our application on three separate topographic maps – from different regions around Great Britain. Finally, we illustrated how the presented technique is being adapted to address a number of other problems in the Geocomputation domain.

ACKNOWLEDGEMENTS

The authors would like to thank: Glen Hart, Ordnance Survey Research Center, UK for his assistance; The British Council for their support; and Leo Mulhare and Adam Winstanley, Department of Computer Science, National University Ireland, Maynooth, for their assistance.

References

- [Bohan and O'Donoghue, 2000] A. Bohan, D. O'Donoghue "A Model for Geometric Analogies using Attribute Matching", 11th AICS Conf., Aug. 23-25, Galway, Ireland, 2000.
- [Davies and Goel, 2001] J. Davies, A. Goel, "Visual Analogy in Problem Solving", Proceedings of IJCAI, 2001.
- [Evans, 1967] T.G Evans, "A Program for the Solution of a Class of Geometric Analogy Intelligence-Test Questions", in Semantic Information Processing (Ed.), M. Minsky, MIT Press, 1967.
- [Falkenhainer *et al.*, 1989] B. Falkenhainer, K. Forbus, D. Gentner, "The Structure Mapping Engine: Algorithm and Examples", Artificial Intelligence, 41, 1-63, 1989.
- [Ferguson and Forbus, 2000] R. Ferguson, K. Forbus, "GeoRep: A Flexible Tool for Spatial Representation of Line Drawings", Proc. 18th Conf. on Artificial Intelligence, Austin, Texas, 2000.
- [Gentner, 1983] D. Gentner, "Structure-Mapping: A Theoretical Framework for Analogy", Cognitive Science, 7, 155-170, 1983.
- [Keane *et al.*, 1994] M.T. Keane, T. Ledgeway, S. Duff, "Constraints on Analogical Mapping: A Comparison of Three Models", Cognitive Science, 18, 387-438, 1994.
- [Keane, 1990] M.T. Keane, "Incremental Analogising: Theory and Model" in Gilhooly, K.T. Keane, M.T. Logie R. & Erdos G. (Eds.) Lines of Thinking (Vol. 1) New York: Wiley, 1990.
- [Keyes and Winstanley, 2001] L. Keyes, A.C. Winstanley, "Using Moment Invariants for Classifying Shapes on Large Scale Maps", Journal of Computers, Environment and Urban Systems 25(1), 119-130, January 2001.
- [O'Donoghue *et al.*, 2003] D. O'Donoghue, A.C. Winstanley, L. Mulhare, and L. Keyes "Applications of Cartographic Structure Matching", IEEE - IGARSS Conf., July 21-25, Toulouse France, 2003.
- [O'Donoghue and Winstanley, 2001] D. O'Donoghue, A.C. Winstanley "Finding Analogous Structures in Cartographic Data", 4th Conference on G.I.S. in Europe (AGILE), Brno, Czech Republic, April, 2001.
- [Ritchie, 2001] Ritchie, G. "Assessing Creativity", Proc. of AISB Symposium on AI and Creativity, York, England, March, 2001.
- [Veale and Keane, 1997] T. Veale, M.T. Keane, "The Competence of Sub-Optimal Theories of Mapping on Hard Problems", Proc. 15th IJCAI, Nagoya, Japan, 1997.
- [Ware *et al.*, 2003] J. Ware, I Wilson, J.A. Ware, "Reducing Graphic Conflict in Reduced Scale Maps", Proceedings 11th GISRUK, p 1-10, London, UK, April, 2003.
- [Winstanley *et al.*, 2000] A.C. Winstanley, D. O'Donoghue, and L. Keyes "Topographical Object Recognition through Structural Mapping", 1st International Conference on Geographic Information Science - GIScience 2000, Savannah, Georgia, USA, October 28-31, 2000.