

Full citation: Holt, A., MacDonell, S.G., & Benwell, G.L. (1998) Assessing the degree of spatial isomorphism for exploratory spatial analysis, in Proceedings of the Third International Conference on Geocomputation (GeoComp'98). Bristol, UK, on CD-ROM.

Assessing the Degree of Spatial Isomorphism for Exploratory Spatial Analysis

Alec Holt, Stephen G. MacDonell and George L. Benwell

*Department of Information Science and Spatial Information Research Centre
University of Otago, P. O. Box 56
Dunedin, New Zealand
Email: aholt@commerce.otago.ac.nz*

Abstract

This research continues with current innovative geocomputational research trends that aim to provide enhanced spatial analysis tools. The coupling of case-based reasoning (CBR) with GIS provides the focus of this paper. This coupling allows the retrieval, reuse, revision and retention of previous similar spatial cases. CBR is therefore used to develop more complex spatial data modelling methods (by using the CBR modules for improved spatial data manipulation) and provide enhanced exploratory geographical analysis tools (to find and assess certain patterns and relationships that may exist in spatial databases). This paper details the manner in which spatial similarity is assessed, for the purpose of re-using previous spatial cases. The authors consider similarity assessment a useful concept for retrieving and analysing spatial information as it may help researchers describe and explore a certain phenomena, its immediate environment and its relationships to other phenomena. This paper will address the following questions: What makes phenomena similar? What is the definition of similarity? What principles govern similarity? and How can similarity be measured?

Generally, phenomena are similar when they share common attributes and circumstances. The degree of similarity depends on the type and number of commonalties they share. Within this research, similarity is examined from a spatial perspective. Spatial similarity is broadly defined by the authors as the spatial matching and ranking according to a specific context and scale. More specifically, similarity is governed by context (function, use, reason, goal, users frame-of mind), scale (coarse or fine level), repository (the application, local domain, site and data specifics), techniques (the available technology for searching, retrieving and recognising data) and measure and ranking systems.

The degree of match is the score between a source and a target. In spatial matching a source and a target could be a pixel, region or coverage. The principles that govern spatial similarity are not just the attributes but also the relationships between two phenomena. This is one reason why CBR coupled with a GIS is fortuitous. A GIS is used symbiotically to extract spatial variables that can be used by CBR to determine similar spatial relations between

phenomena. These spatial relations are used to assess the similarity between two phenomena (for example proximity and neighborhood analysis). Developing the concept of spatial similarity could assist with analysing spatial databases by developing techniques to match similar areas. This would help maximise the information that could be extracted from spatial databases. From an exploratory perspective, spatial similarity serves as an organising principle by which spatial phenomena are classified, relationships identified and generalisations made from previous bona fide experiences or knowledge. This paper will investigate the spatial similarity concept.

1. INTRODUCTION

Data exploring and re-use techniques will have an increasing impact on information technologies as more data is amassed. Case-based reasoning (Schank 1982), data mining and knowledge discovery (Fayyad 1997) are techniques used to search, recognize, extract, examine and predict decision knowledge from data. Earlier research by Holt (1996b) on advancing the exploratory data analysis (ESDA) techniques for GI focused on applying case-based reasoning (CBR) techniques. In particular he focused the *reuse* component of CBR and applied it to spatial phenomena. The next research direction focuses on determining methods to store (represent) spatial data in a case structure and how this affects the *retrieval* component of CBR. Researching the peculiarities of the *retrieval* component is important because of its role in selecting similar cases.

This paper details how cases are indexed for efficient retrieval and the similarity and weighting system between new and past cases. It is held that *spatial similarity* is an important concept for storing and retrieving cases. Spatial similarity will aid in determining clusters and feature detection for classification. This presupposes that it is possible to define spatial similarity. In this paper spatial similarity is defined as *the match between a source and a target for a particular scale and context*. The match is also determined by time, position and techniques. Time is the state of a phenomena at a particular instant, position is vital to utilise the spatial analysis functionality in a GIS, for example proximity, and the techniques are various retrieval, matching and ranking methods utilised to

retrieve and match similar phenomena. Similarity may be determined by any one of a number of methods including fuzzy membership (Zadeh 1965), rough sets (Pawlak *et al.* 1995) spatial auto-correlation and statistical techniques.

2. SIMILARITY

A dictionary definition of morphology is "*a science of form*". Isomorphism is defined as "*similarity of form*." The word isomorphism is used in this paper to indicate the broad focus in the similarity of spatial forms. Broad in the sense that similarity should not be limited to the formalisms of GIS systems. Similarity is more than that. Kant (1724-1804) says "*there is nothing more basic to thought and language than our sense of similarity; our sorting of things into kinds*."

This paper outlines previous studies on similarity assessment by various disciplines, especially psychology, philosophy and information science (computer science). This paper acknowledges that there are numerous disciplines including neuroscience, linguistics and statistics in which similarity has been researched but they are not detailed in this paper. This partial history of similarity studies is used as a motivation for proposing a novel theory of similarity called spatial-based similarity.

2.1 Cognitive Psychology

Similarity has been a topic researched in the psychology field for decades, for example, early researchers were Wallach 1958; Tversky & Krantz 1970; Tversky 1977. Recently there has been a huge resurgence in the topic. Similarity (or psychological distance) in psychology employs both descriptive and exploratory concepts (Knauff in Voß 1993). Similarity judgements are considered to be a valuable tool in the study of human perception and cognition and play a central role in theories of human knowledge representation, behaviour and problem solving. This paper aims to utilise similarity judgements as a tool to represent, retrieve, model and solve spatial dilemmas. Tversky (1977) describes the similarity concept as "*an organising principle by which individuals classify objects, form concepts, and make generalisations*". Classification, abstraction and generalisations are methods and techniques that underpin most GI systems. Therefore, similarity as defined by Tversky should be intuitive and useful to GI systems. Ellison (1997) suggests that human perceptions are often logically compatible with abstractions. Hampton (1997) also argues that many of our everyday concepts are built around similarity clusters. Ellison attempts to justify the claim that the future will be like the past by introducing the problem of induction, and proposes a solution based on similarity measures and topographic mapping. The premises of his solution are that; (i) Naturally occurring data and representations are embedded in spaces with non-trivial similarity structures and (ii) Natural cognitive mappings between spaces of representation are topographic mappings. MacLaury (1997), takes a different approach to similarity (from a cognitive science/anthropology perspective). He has researched a technique called Vantage Theory in an effort to procure a

testable model of categorization and the part played by judgements of similarity and difference. This approach is being used to propose the concept of *Spatial Vantages* (Holt & MacLaury In press) to investigate how spatial judgements can be made and to test its application for spatial categorisation.

2.2 Philosophy

Bain (1855, In Jurisica (1994)) realised the importance of studying similarity as a psychological problem. He defined a "Law of Principle of Similarity" as "*the tendency to be reminded of past occurrences and thoughts of every kind, through their resemblance to something present*." In Bain's work, resemblance is used as an undefined primitive term to define similarity. Similarity is used as one of two principles to explain learning (the other one is contiguity). He proposes that classifications be assembled by the notion of similarity. Again the usefulness of similarity is recognised by its ability to remedy from the past for the present. This concept is useful for spatial problem solving and classification.

2.3 Information Science

In information science the focus has been on implementing psychologically plausible theories of similarity. Information science terms dealing with similarity include, but are not limited to, indexing, sub-setting, retrieval, matching, ranking, solution space, clustering, trees, categorising, equal and equivalence. Information science research in the field of similarity could be grouped under the following headings; comparison functions, retrieval functions, evaluation functions and analysis functions. Various researchers from different information science disciplines are studying similarity. The results and ideas between some of these disciplines are interchangeable, because of the overlapping interests. The different disciplines include computer vision, graphic design, pattern recognition, image analysis, databases, artificial intelligence, remote sensing and GI systems.

From an information science perspective, similarity can be described as a retrieval system that allows data to be compared for similarities. A user specifies the required data and the criteria for matching. The system retrieves all similar data. However, on occasions what is considered similar in one situation may not be similar in another. Thus, systems should take context into consideration by representing constraints on similarity matching (context) explicitly. Context allows the user to specify what parts of information representation to compare and what kind of matching criteria to use. This allows for excluding similar but irrelevant items. Context also allows us to constrain retrieved information in such a way that only relevant information is obtained. To assess similarity in different situations we need to be able to specify criteria for matching flexibly (Kolodner, 1993). This paper proposes to use the indexing technique in case-based reasoning to allow for this flexibility and to act as a context constraint.

2.3.1 Similarity in databases

Jagadish (1991) and Jagadish *et al.* (1995) researched similarity in a spatial database field and proposed an organization for a database of objects that permitted an efficient retrieval of objects with a shape similar to an input shape. For similarity judgments, an area-based similarity is used. Carbonell (1986) used similarity as one of the possible transmutations - a form of analogical inference. He defines similarity with respect to context (either implicitly or explicitly defined). However, he did not define features of similarity and dissimilarity. A way of using similarity and dissimilarity relations for inductive and deductive inferences is also provided. Kashyap & Sheth (1993) presented an approach to resolve schematic differences among semantically related objects in multi-database systems. They define semantic proximity as an attempt to characterize the degree of semantic similarity between two objects using the real world semantics. Key to their definition of semantic similarity is explicitly represented context. Another use of their approach is to represent uncertain information and to resolve data value incompatibility in multi-database system.

Juristica (1994) suggests that there are two possible approaches to implementing similarity-based retrieval systems;

1. Similarity relations among items are predefined. This approach is called a limited similarity in retrieval as the context is usually fixed.
2. Similar items are located by defining similarity relations at query time, allowing for flexibly changing context and criteria for matching. Such relations are defined as similarity in retrieval.

Juristica (1994) suggests that in general, similarity is a relation with three parameters: a set of relevant items, a context and an information base. In comparison Holt *et al.* (1997) use context, scale, repository, matching and ranking techniques and measure(s) to determine spatial similarity (Figure 1).

2.3.2 Image similarity

Image similarity is based on visual cues like size, shape, colour and texture. Research in image similarity focuses on the retrieval and recognition of the components of the image. World-wide projects such as Jacob, Virage in UCSD, Photobook in MIT, QBIC in IBM, KPX in Kodak and PressLink Online at PressLink are systems designed for the efficient storage and retrieval of relevant images and knowledge.

Jin *et al.* (1997) researched these *text and content* based retrieval systems and identified that retrieval requests are usually issued with partial information and it is difficult to describe visual cues. It was also noted that most retrieval methods are passive and do not possess the ability to understand query requests. Importantly they identified that humans are unsound in weighting image features quantitatively; however, are robust in accumulating knowledge, combining features and making complex judgements. Therefore, to improve from the inadequacies of current text-based and content-based retrieval systems,

Jin *et al.* (1997) proposed a two-stage image retrieval system, CBIR-VU. CBIR-VU goes beyond simple information retrieval to retrieving data on knowledge by accommodating knowledge acquisition in retrieval, and is able to handle complex queries with partial information.

In image analysis there have been many approaches to utilise spatial similarity for example, Richter, Gero & Sudweeks, Lee & Hsu, Coulon, Katey Borner, Angi Voß and Bartsch-Sporl & Tammer. Rather than describing these applications, a medical imaging example is provided.

In an image understanding architecture there are a number of tasks that employ a similarity measure/metric/notion. In segmenting an input image, a similarity measure is needed for separating feature clusters. In finding image cases a similarity measure is needed for calculating which cases are close to each other in the solution space. Similarity is defined by what the different image segments mean to an expert agent. One approach is to use explanations, such that, the system explains to itself, **why** the two representations of image segments are similar in this particular **context**. The answer to **why** depends on **context** Grimnes *pers com.* (1997).

Grimnes & Aamodt (1996) are concerned with the semantic similarity of cases, that is, what is considered similar by a radiologist is what defines the similarity "metric"? They view medical image interpretation as a design process. A clinically meaningful interpretation is a collection of subpart interpretations where all the subparts form a meaningful whole. As such the focus on similarity is both on how the whole is similar to the whole in another image, and equally on how each of the subparts are similar to subparts of other images. Therefore, it is underselling to define image similarity as $SM(A) \sim SM(B)$ where SM (Similarity Measure) is a function of an image (A/B) and \sim is some kind of (numerical) equality predicate. In a number of domains a more structurally/syntactically based similarity metric may be used, that is, maximum likelihood/c-means/grammar-parsing based artificial neural network. In some domains, however, there are semantic and contextual constraints that are difficult to capture with these methods.

Grimnes recognises that each metric have their advantages and disadvantages but suggests an advanced, learning and knowledgeable image understanding agent must probably be a hybrid that employs both knowledge poor and knowledge rich/demanding methods to achieve optimal retrieval, Grimnes *pers com.* (1997).

2.3.3 Similarity in remote sensing

Similarity has been researched previously by Jain and Hoffmann (1988) for pattern recognition. They designed a technique that used evidence-based reasoning to measure similarity between objects. More recently in the remote sensing field Agouris, *et al.* (1997) are concerned with the retrieval of images from image databases using query-by-sketch operations. Agouris, *et al.* (1997) propose to research beyond the typical and elementary metadata such

as color content. They base their approach on a *shape and geometry* oriented algorithm. They also use a least-squares methodology for shape and geometry similarity comparisons, as they suggest it offers excellent potential for ranking the matching images and is suitable for multi-scale applications. They aim to develop a general image query-by-sketch operation by analyzing geometry, shape, topology and semantics and provide an extension of query editing in space and scale for sequentially refining query operations.

2.3.4 Similarity in CBR

Research in CBR, an AI technique is what the authors focus on in this paper. It is realised there are other AI techniques which could be used for similarity assessment, for example, fuzzy logic and artificial neural networks.

Osborne & Bridge 1997 developed a similarity measurement framework used within CBR systems called similarity metrics. In their framework similarities are values from any data type on which a complete lattice is defined. Using the lattice allows a wide range of methods for measuring similarity. They suggest their approach is useful for data categorisation. Keane 1997 suggests that a reasonable computational level account of similarity is "some way off". One reason for this the low level of interest in the processes which shape the representation of items. Most emphasis on similarity judgement is focussed merely on the items. He illustrates his idea by using one computational instance from CBR. Keane 1997 proposes that various parts of the representation process can contribute to the perceived similarity of items. He then outlines a view which he favours called the Dynamic Similarity perspective. This view is supported by two sample psychological demonstrations in the judgement of similarity between (i) sentential descriptions of events and (ii) perceptual patterns that have been physically manipulated. Jeffery *et al.* (1997) have researched CBR using similarity and categorization from a multiple correspondence analysis. Their research relates to the use of visual cues for accessing and comparing the medical images of patients with a particular disease (pathology). They postulate that psychological similarity is captured in the spatial relations of items in a multiple correspondence analysis (MCA) scatter plot. Jeffery *et al.* (1997) suggest that similarity relations are conceptualised in the sense that two stimuli are similar psychologically if they appear close together in the similarity space. They also suggest that the psychological notion of the typicality of cases within a disease may be visualised as the distance of any case from the center of this map. They envision that it may also be possible to provide information using these scatter plots relating to the relative positions of cases in overlapping pathologies, for the identification of problem cases and to assist in the categorisation of new cases. Rodriguez (1997) has also researched CBR. He thinks flexibility is the most important factor in determining similarity. To achieve flexibility Rodriguez suggests the development of a context dependent similarity measure. His work presents a novel approach for determining the importance of the item characteristics by combining a memory of existing data with general domain knowledge into a number of fixed dimensions.

2.3.5 Similarity in GI systems

There are some distinctive groups currently researching similarity in the milieu of GI systems. These distinctive groups use a variety of techniques ranging from deviation from equivalence and feature matching to case-based reasoning. Possible uses of similarity range from inter-operability (Goodchild *et al.* 1998), conflation (Cobb *et al.* 1998), data retrieval (Holt & Benwell In Press); Flewelling 1997; Bruns & Egenhofer 1996), problem solving (Holt 1996b; Higham *et al.* 1996; Jones & Roydhouse 1994) and exploratory/interpretation (Holt & Benwell In Press).

Cobb *et al.* (1998) present a novel approach to combining maps and associated knowledge (conflation). For conflation they need to determine points which are identical between different maps. They describe feature matching and de-confliction and favour the use of using inexact reasoning concepts. They implement a system where each feature is considered as a set of attribute-value pairs. From this representation, a degree of matching similarity is determined. For numeric domains a membership matching function is used, while a similarity table is used for linguistic domains. By using a combination of the table and a fuzzy logic membership matching function a composite matching score is then computed from the combination of an expert system weight and the similarity table values.

Recent interest in similarity comes from a report by Goodchild, *et al.* (1998), which suggests similarity is relevant to inter-operability. It is relevant in that it allows a measure of the degree of which "two data sets, software systems, disciplines, or agencies use the same vocabulary, follow the same conventions, and thus find it easy to interoperate." Goodchild, *et al.* (1998) continue along the same vein and suggest that currently, it is only possible to inter-operate over a very narrow domain. Therefore, when considering similarity in the context inter-operability Goodchild, *et al.* (1998) say "the effort to achieve interoperability is thus an effort to extend domains, or to raise the threshold of similarity below which interoperability is possible." The authors assume the above could also be thought of for intra-operability.

Configuration similarity developed more recently as a form of content-based retrieval. Bruns and Egenhofer (1996) and Papadias & Egenhofer (1997) grapple with similarity initially by focussing their research on describing spatial structures and configurations to a high degree (in spatial databases). Once they realise the spatial shape or structure, and given a new instance, they can then equate similarity by counting the number of transforms it takes to morph from an unknown state to a known state (structure or configuration). Bruns and Egenhofer (1996) define similarity as "the assessment of deviation from equivalence". The question is how do we represent and measure "assessment of deviation" and how is "equivalence" defined? Bruns and Egenhofer (1996) use similarity for data retrieval and feature matching.

Egenhofer directs two current research projects with a focus on *similarity*. These include;

1. *Similarity* assessments based on spatial relations and attributes, funded by the National Imagery and Mapping Agency,
2. Heterogeneous geographic databases: spatial *similarity*, Advanced Research and Development Committee of the Community Management Staff.

The project includes research on numerous database issues including *spatial similarity* retrieval. Researchers include Egenhofer, Flewelling, Goyal, Paiva, Rodríguez & Beard (University of Maine), Bertolotto (Università di Genova, Italy), Freitas (INPE, Brazil), Sharma (Oracle) & Ubeda (INSA de Lyon, France).

In the *similarity assessments based on spatial relations and attributes* project spatial similarity measures are developed to overcome the shortcomings of traditional methods (precise spatial concepts, discrete data structures and boolean operators). Egenhofer's team propose similarity measures are based on spatial relations and attributes. Spatial relations are used to capture the distribution of spatial objects through a multi-scale model, allowing analysis of topological, directional and metrical relations. Attribute similarity is measured through a semantic network of feature classes.

The *spatial similarity* project investigates the changes detected whilst analysing multi-scale geographic databases among the different representations for the same geographic area, or different geographic locations. Spatial similarity can be derived using the concepts of the 4-intersection and its component invariants. We will extend this model to account for qualitative metric properties of spatial relations, and will develop formal models for assessing spatial changes. Egenhofer's team aim to also test their concept for 2-dimensional and 3-dimensional models.

Papadias and Delis (1997) define measures for modelling similarity of configurations. Papadias and Delis (1997) suggest configuration similarity has developed more recently as a complementary form of content based retrieval and that most approaches following methodology:

1. describe the set of spatial relations allowed in the expression of queries,
2. define measures of similarity between images based on the resemblance between spatial relations (and not on visual characteristics) and
3. (in some cases) propose algorithms for similarity retrieval.

Flewelling (1997) suggests recent similarity queries have been researched in the object-based spatial (Flewelling 1997; Bruns & Egenhofer 1996) and image database community (Flickner *et al.* 1995; Gudivada 1995; Gudivada & Raghavan 1995). There has been little research on the properties that similarity operators must fulfill and on the differences between field and object models. Flewelling (1997) proposes a solution to the differences between field and object models. He suggests that in order to measure the similarity of one field to another we must measure the similarity of the four field characteristics. He identifies these four fields as theme,

extent, time and value (samples) and says these can be used to derive a four dimensional distance representing the similarity of the two fields. A set of these field similarities could be generated against a user defined scenario (query) or a known state. Flewelling (1997) suggests that this will make it possible to retrieve fields from a database that are highly similar, (but not equivalent, to the users query) and to quantify that similarity.

The authors have identified the usefulness of similarity in GI systems (Holt & Benwell 1996). Holt (1996b) propose a spatial similarity system (SSS) which would allow GI systems the ability to recognise, retrieve, re-use, revise and retain from the past for the present and future. This concept is useful for spatial problem solving, data retrieval, classification and exploratory/interpretation (Higham *et al.* 1996; Holt & Benwell In Press).

There is an increased need for more GeoComputational techniques for data analysis, data mining and for exploratory analysis for certain applications (Holt 1997; Openshaw & Abraham 1996). This paper proposes that spatial similarity could be utilised both as a descriptive and exploratory concept in an attempt to satiate the GeoComputational need. The SSS is a spatial-artificial intelligence-hybrid and is under continuous research and development. The SSS has arisen from the belief that current GI systems are limited in their reasoning ability and case-based reasoning (CBR) can be integrated to support this deficiency. The primary use of such a system will be to develop reasoning techniques for discovering knowledge about areas that are considered to be *spatially similar*. CBR offers the ability to reason, explanation features, adaptation facilities, extended generalisation techniques, inference making abilities, constraining a search to the solution template, solution generation and the ability to validate and maintain knowledge bases. These features would aid planning, forecasting, diagnosis, design, decision making, problem solving and interpretation.

Holt and Benwell (1997) defined spatial similarity as "*those regions which, at a particular granularity (scale) and context (thematic properties) are considered similar.*" This definition has since been refined and illustrated in Figure 1. Similarity is influenced by the specific user (their goals), the application (the problem), the system developers and the available technology (software and hardware). It is important to realise that context in this definition is defined by the user and not automatically by the system. From a GI science perspective similarity can be defined as computing the degree of match, which is achieved by the retrieval, matching and ranking of geographical phenomena.

The degree of match to a set of criteria (parameters) and circumstances (application) also influence the degree of similarity. Another principle that governs similarity is determined by the user. The user selects a set of criteria, defines circumstances and biases the appropriate criteria to achieve the desired result. Therefore, based on a set of criteria selected by the user, similar instances can be found (Holt 1996b). It is not just the attributes that determines similarity: Dubitzky *et al.* (1993) adds to this by suggesting that "*The relation rather than the objects*

alone determines to a large degree the similarity between two situations". This paper attempts to build on this concept by including spatial relations to spatial data. It is the spatial relationships between situations that determine if they are spatially similar or not. Using proximity analysis available in GIS allows a relation to be formed between spatial data, which can be used as a similarity measure.

Recent solutions to spatial problems have involved using previous similar spatial phenomena. Higham *et al.* (1996), for example, analysed tourist flow patterns, Jones & Roydhouse (1994) examined weather patterns and Holt (1996a) modelled the environment. Holt & Benwell (1997, In press) indicated that spatial similarity can be used to answer questions such as: *Are there spatial phenomena similar to the searched example? Which spatial phenomena have the certain criteria?*

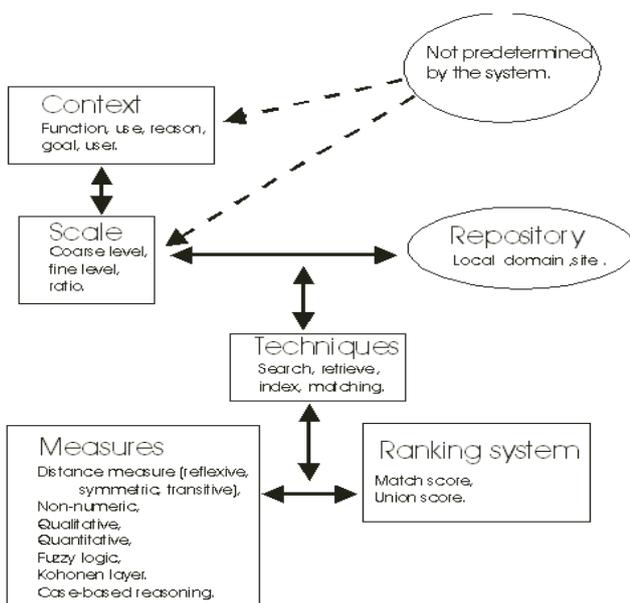


Figure 1. Components for determining spatial similarity.

3. SPATIAL SIMILARITY SYSTEM

A *spatial similarity system* should allow the user to detail their particular goal(s) and the application together into a set of parameters which can be executed upon and adjusted to calculate *spatial similarity*. The system would also allow results to be displayed indicating the degree of similarity through a matching and ranking measure. This would allow the user to select a set of textual and spatial (allow the user to click on a pixel/line/polygon and find the location of similar pixel/line/polygon(s)) parameters to be searched and to be adjusted (weights) accordingly for the application to get an indication of similarity between information stored and the new parameters entered into the system. The degree of similarity will be determined by a matching and ranking system. A characterisation of the similarity criteria that this paper uses, or is most pertinent to it, is the calculating of the degree of similarity. This is determined by using a statistical technique known as nearest neighbour weighting.

A spatial similarity system produces a map indicating the levels of similarity based on constraints defined by the user. The user had the choice to input the constraints as criteria they wanted fulfilled. As well as this the user could assign a weight suitable to the users expertise as to which criteria were the most important. *Idrisi for DOS* was used for analysis and *Visual Basic* for the user interface. The number of modules that can be executed from the command line in *Idrisi for DOS* for this exercise was limited to the following ten commands: COLOR, COLOR 85, DISTANCE, EXPAND, GROUP, MAINT, OVERLAY, RECLASS, SCALAR and WINDOW.

A typical query would be: "According to the control area (which has an altitude of 300m, slope of 25 degrees and an aspect of 160 degrees) find similar areas and indicate the degree of the similarity." Upon entering the criteria the user also has the option of assigning an appropriate weight (Figure 2). If the criteria have equal importance than the weights will be equal, otherwise the weights are assigned in a ratio as to their perceived or contextual importance of the criteria.

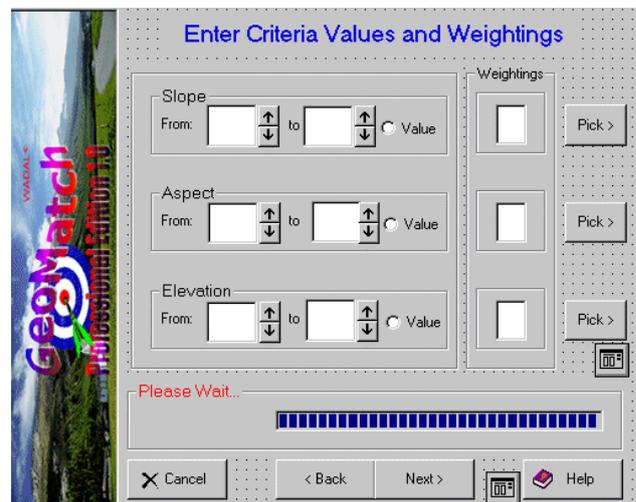


Figure 2. Enter criteria values and weightings.

The user query is then processed, which is a quantitative process using RECLASS and OVERLAY operators. The elevation image is RECLASS(ed) according to the criteria and then the dataset is used to generate two images for slope and aspect, using the SURFACE module. The three images will then be OVERLAY(ed) and RECLASS(ed) into a set of predetermined categories. A map is then produced indicating the various levels of similarity according to the users criteria and weights.

The level of similarity was determined by using the statistical technique known as nearest neighbour weighting. Using this method the category that the image pixel is part of is assigned a value of 1 in a RECLASS process. The categories adjacent to this category are assigned a value of 2, with the next adjacent categories given a value of 3. This is continued until every class in the dataset has been assigned a value. The higher the assigned value, the less similar the category. The resulting classification is then normalised. This process takes a range of categorisations for different mapped features and converts these into standardised units capable of

comparison with each other. This process will be carried out on the elevation, slope and aspect images (if they had weights assigned to them).

The normalised images will be OVERLAY(ed) to produce the solution image. This image is finally RECLASS(ed) into categories that are colour-coded for display. The resulting images (Figures 3 & 4) show the level of similarity of every pixel in the raster image.

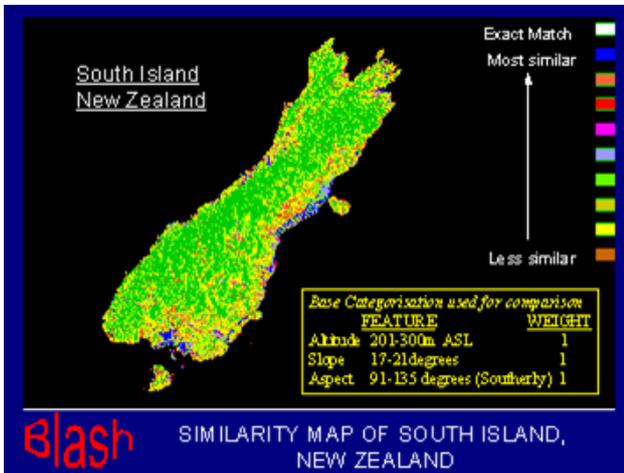


Figure 3. Similarity map with equal weightings.

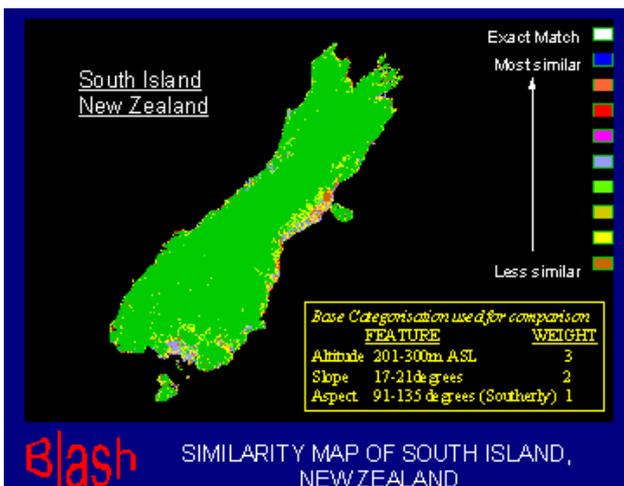


Figure 4. Similarity map with unequal weightings.

CBR offers the potential for improved functionality to current GIS. This is achieved in a complementary fashion as the functions they both have are executed in different methods (for example, retrieve and retain). The functions of GIS and CBR techniques which differ the most are their abilities and techniques for representing and storing data. The ability of CBR to learn is another component which separates it from a GIS. Data and knowledge in the form of cases are stored and represented so they can be retrieved quickly to suit particular requirements. This complicated storing method (bundles of knowledge) are indexed to allow new experiences to be saved. A sense of learning, therefore, is introduced. Other components offered by CBR include the reuse and revise (adapt) functions which current GIS software packages lack.

4. DETERMINING SIMILARITY

the degree of similarity between two matched features/values can be computed. (Kolodner 1995:346).

There have been a variety of proposals to assess similarity most of which are based on

1. geometric models,
2. Tversky's contrast model,
3. Structure mapping theory
4. models of representational change (Knauff in Voß (Ed) 1994).

In geometric models, similarity of two objects (a) and (b) is a monotonic function of the distance between their representations in a multidimensional space (Ortony, 1979). The fundamental disadvantage with the monotonic function approach is its inability to deal with asymmetry of similarity judgements (Knauff in Voß (Ed) 1994).

The Tversky contrast model assesses similarity between two instances by counting the number of matching and mismatching features. The disadvantage of this model is that it is not flexible enough to handle changes due to context. The advantages of this approach are efficiency and it is computationally inexpensive. Generally a measure of similarity is a distance measure, that is, a measure of the difference between a source dataset and a target dataset (Tversky 1977). Flewelling (1997) suggests that this concept is counter intuitive to the normal usage of similarity. He uses the following example, if two datasets have a high similarity, their difference is small. When the difference between two datasets is zero they are "the same". These datasets are "the same" if they have elements of the same type. Flewelling (1997) says "in order to assess similarity it is necessary to perform a difference operation over the set attribute measures for each pair of spatial datasets" Flewelling (1997:53).

Gentner and colleagues (Gentner 1983) (Gentner & Forbus 1991) in their structure mapping theory identify that a theory on similarity must "describe how the meaning of an analogy is derived from the meaning of its parts" Gentner (1983:155). The mapping principles are relations between objects, rather than attributes of objects and the definition of higher order relations. There are many approaches to similarity, which take this view. Some of the basic assumptions of such approaches were supported from a psychological point of view by (Knauff & Schlieder 1993).

In recent years these fixed description approaches were criticized, especially by Indurkha & O'Hara (Indurkha 1991 & 1992) (O'Hara 1992) (O'Hara & Indurkha 1993). They argue that the mechanism underlying such creative analogies is representational change (Indurkha 1992) or redescription (O'Hara 1992). The key idea of these approaches is a process by which new points of view can be created and these redescriptions can be useful for the matching process. Both authors focus on geometric proportional analogies (proportional analogies have the form A is to B, as C is to D).

4.1 Recognising similarity at different dimensions/scales

Scale affects spatial similarity. To understand and model spatial similarity the characteristics of scale and the affects of its changes (on information and analysis) need to be researched. Understanding scale variations is a complex topic as these variations in effect constrain the manner and in which information can be observed, represented and analysed. These constraints are the impetus for researchers, across all sciences that use geographic information, in an attempt to understand scaling.

Savitsky and Anselin (1997) say that;

"Issues of scale affect nearly every GIS application and involve questions of scale cognition, the scale or range of scales at which phenomena can be easily recognized, optimal digital representations, technology and methodology of data observation, generalization, and information communication".

Scale and resolution can have a significant effect on spatial patterns and processes according to Lilburne (1997). Scale dependence is where spatial pattern varies with scale. Different patterns emerge at different scales in most environmental systems. There is currently no objective methodology for determining the range and optimal scale at which a process operates, and contributes to a spatial pattern, despite this being critical for scaling or generalising models. There are no tools to help quantify the uncertainty that derives from modelling with data collected at different scales from the one of interest. The increasing availability of spatial data offers greater opportunities for spatial modelling and analysis at a variety of scales. This re-forces the need to outline to decision makers that scale related uncertainty and validity of data and models should be understood (Lilburne 1998).

Researchers in a variety of disciplines have been addressing the problems of scale and scaling. These include, for example, cartographers (Buttenfield & McMaster 1991), cognitive scientists (Voß 1993), computer scientists (Elmasri & Navathe 1994), ecologists (Ehleringer & Field 1993), (Cain *et al.* (1997), (Cullinan & Thomas 1992), geographers (Hudson 1992), geostatisticians (Wong & Amrhein 1996), hydrologists (Sivapalan & Kalma 1995) and remote sensing specialists (Cao & Lam 1997) (Quattrochi & Goodchild 1997). Consequently, there are a number of techniques in the literature that are of use in characterising scale of different spatial data types. These include measures of spatial autocorrelation, semivariograms, textual analysis, dimensional analysis, fractals, multi-fractals and statistical measures of variance and diversity.

Savitsky and Anselin (1997) say that much recent attention is focused on formalizing the study of scale .. (sic) ... and on exploring robust methods for the representation, analysis and communication of information across multiple scales.

Lilburne's (1998) research focuses on;

1. Establishing a set of techniques for measuring the operational scale of spatial processes and

determining an appropriate structure to model scale dependencies.

2. Implementing routines to calculate measures of scale in a GIS-based framework that is designed to facilitate an investigation of scale effects. This framework will be used to refine the set of scale measures (as above), based upon an analysis of scale effects of some environmental phenomena.

Hierarchy theory is seen by some researchers as a way forward to model the nesting of scale dependencies. Environmental gradients however, often overlap and the interactions between processes and scale are not necessarily hierarchical. By using biophysical datasets Lilburne intends to verify the appropriateness of hierarchical structures and investigate other representations including object orientation and logic.

Scale and spatial process are significant problems that are closely linked. It is possible to compute scale effects from static spatial data very easily and derive indicators of the effects from these. We cannot understand them unless we understand and/or can model the process involved. More emphasis should be placed on definitional aspects of space that can complicate expressions of spatial scale. Stevens (1946, in Flewelling 1997) identifies four scales of measurement, which are nominal, ordinal, interval and ratio. Each of which have specific characteristics which limit the types of valid operations executable. Recent work on the scaling behaviour of various phenomena and processes has shown that various processes are not linearly scaled (Savitsky & Anselin 1997). There needs to be more research on how various phenomena change through different scaling processors. There have been some attempts to describe the scaling behaviour by fractals, which have proven ineffective for many geographic phenomena because certain properties do not repeat across multiple scales. Hence, the research into multi-fractals which has shown some usefulness for characterizing the scaling behaviour of some phenomena. We are particularly interested in trying to understand the impacts that changes in scale have on the information content of databases.

Benefits of research into scale by Savitsky and Anselin (1997) that are applicable to similarity include;

1. the systematized bases for scale-related decision making,
2. the new methods for quantifying and compensating for the effects of scale in statistical and process models,
3. the improved understanding of cognitive issues of scale and
4. the design and development of multi-scale database.

New spatial analytical techniques and functions, which focus on determining scale and spatial similarity effects, underpin research in spatial data-mining. Ultimately this research may improve spatial modelling tools and the quality of information delivered to researchers and decision-makers.

4.2 Context

The context of data is not merely the attributes, it is also what the attributes are to be used for, their purpose. The purpose is the specific function (use, reason, goal) which the attributes are to be used for. To answer, what is similar between a source and a target depends on the context of the question. Different answers will be given for different contexts.

4.3 Techniques for measuring Similarity

It is recognised that numerous statistical analysis techniques exist, such as inverse distance weighting using linear, exponential or logarithmic functions as well as artificial intelligence (AI) tools such as Case-based Reasoning (CBR) to determine similarity. The following techniques are being researched by the authors for their possible use in GI systems to measure similarity.

4.3.1 Abstraction Hierarchy

Abstraction hierarchy is where the degree of similarity is computed in terms of the most specific common abstraction (MSCA) of the two values. Therefore, the more specific the MSCA the better the match (0 = least specific, 1 = most specific (Voß 1993)). Figure 5, indicates how through classification in the animal kingdom a Kea (a new Zealand bird) can be compared to a dog and a value for similarity can be calculated. In this case the value for the measure of similarity would be 0.2. In comparing a Kea to another Kea the value will be 1, meaning they are very similar and could, according to this hierarchy, be the same. Another example would be comparing a Kea to a Kiwi the answer would be 0.6. That means a Kea is more similar to a Kiwi than a dog.

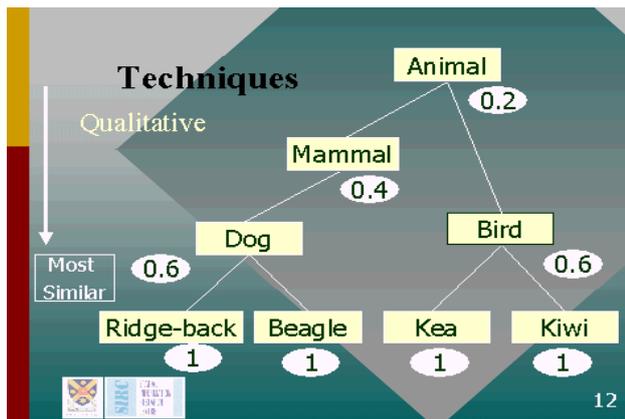


Figure 5. An example of an abstraction hierarchy.

4.3.2 Qualitative and quantitative distances

Qualitative and quantitative distances involves measuring the degree of match by calculating the distance between the two values on a qualitative scale. If two values are within the same qualitative region, then they are considered equal. Otherwise the distance between their qualitative regions provides a measure of their match score. The more regions separating two values the lower

the match score. This method is inaccurate for edge or border values. Two remedies are to define regions so that they overlap and then scrutinise the values that lie on the border of two regions (Voß 1993). For example in the age categories seen in Figure 6 below, for example, provide an instance where an attempt is made to measure the similarity between the age of people. According to the categories a person between 62 to 75 years is old and a person between 40 to 45 years is middle-aged. Therefore, the ages 40 and 62 are one qualitative region apart. The ages 35 and 65 are two qualitative regions apart. Figure 6. suggests that if a person falls into the young adult category then the numerical values of similar measures indicating the distance between the respective qualitative regions are illustrated.

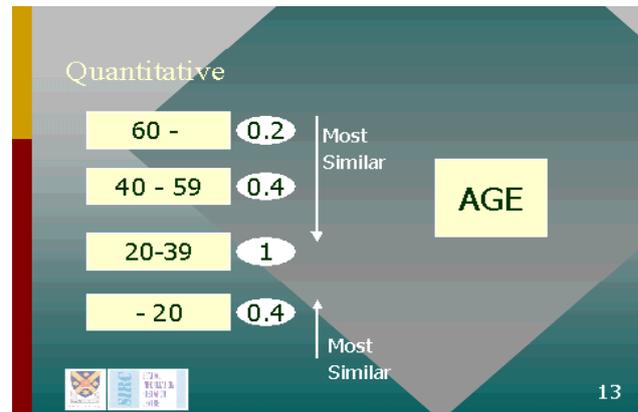


Figure 6. An example of using qualitative distances to measure similarity.

4.3.3 Other AI techniques

It is also possible to use the kohonen layer (Lees 1997), inverse distance matrix (Seixas & Aparico 1994) and fuzzy logic (Kasabov & Ralescu 1993) methods to calculate the similarity between phenomenon. Attempts to calculate spatial similarity were executed by spatial overlays and re-classification techniques (Holt 1996b; Black *et al.* 1997; Wallace *et al.* 1997). The authors favour the CBR approach applied to spatial data (Holt 1996b) because of its novel concept of re-using previous experiences. Subsequent research has highlighted that it is also a useful concept for determining similarity.

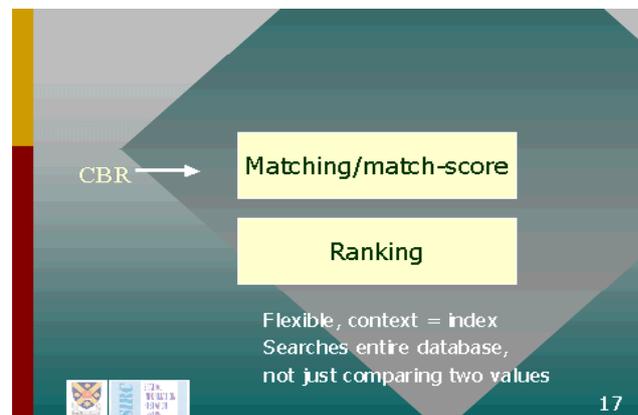


Figure 7. CBR method to calculate similarity.

CBR uses matching and ranking to derive similarity (Figure 7.). Matching is achieved through index and weights, while ranking is the total of the match score. CBR was useful as it offered flexibility in dealing with the concept of context (which we considered to be important in terms of similarity). CBR also searches and matches the entire database not just by comparing two values (Kolodner 1993). Most CBR systems the *nearest neighbour* matching technique for retrieval. Nearest neighbour algorithms are executed in a common fashion and this is represented in Figure 8.

$$\text{Similarity}(T,S) = \sum_{i=1}^n f(T_i, S_i) \times W_i$$

Figure 8. A typical nearest neighbour algorithm (Watson 1997:28).

Where;

T is the target case, S is the source case, n is the number of attributes in each case, i is an individual attribute from 1 to n, f is a similarity function for attribute i in cases T and S, W is the importance weighting of attribute i.

The nearest neighbour approach involves the assessment of similarity between stored cases and the new input case, based on matching and ranking each field and the respective weights. The user decides if certain features need weighting and if they do the various ratios between the weights of the features. One limitation of this approach is that retrieval times increase with the number of cases. This approach therefore, is more effective when the case base is relatively small (Watson 1994).

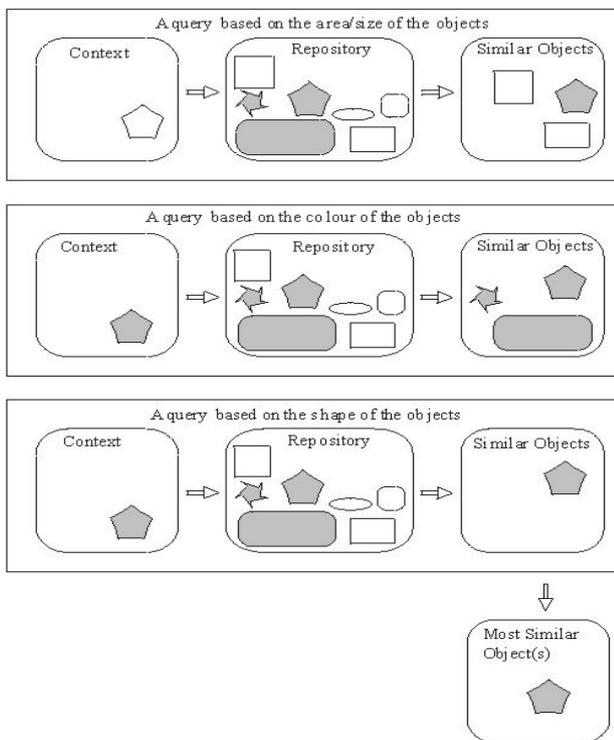


Figure 9. A non-numeric attempt to measure similarity.

4.3.4 A non-numeric technique

Most similarity measures use a numeric value to indicate the level of similarity. This numeric value is the result of matching and ranking techniques to provide a match score (the similarity value). On some occasions it may be incorrect to place a numeric value on a item, especially if we know little about the value and if the value is used in a secondary calculation. Figure 9. is an attempt to get a non-numeric measure of similarity, its graphical and the most similar item is a result of the union of a variety of queries and contexts.

5. CONCLUSION

The concepts outlined in this paper illustrate the data mining and data exploration benefits of determining spatial similarity. It also offers novel methods for searching and comparing complex geographical entities. This paper has proposed possible directions to advance current GIS techniques for analysing, searching, recognising and extracting information on spatial patterns. In particular this paper has outlined how an AI technique called case-based reasoning could help in achieving these proposed advances.

6. FUTURE WORK

Possible future research avenues include;

1. the incorporation of a *CBR-Neuro-Fuzzy* hybrid and investigate the extra robustness provided by using natural language programming techniques, in particular the experienced-based reasoning software which is a *Natural Language-Neuro-Fuzzy* hybrid,
2. implementing similarity as rules,
3. computing similarity in parallel,
4. researching user-computer dialogues effects on similarity,
5. researching typicality and asymmetry effects, diversity effects and contextual influences on similarity.

ACKNOWLEDGEMENTS

The authors acknowledge;

The assistance provided by the New Zealand Foundation for Research, Science and Technology, (Grant UOO605) awarded to Professors Geoff Kearsley (Centre for Tourism) and Rob Lawson (Department of Marketing) University of Otago.

The support from the Information Science Department and the research of the WADAL and BLASH teams from the 1997 postgraduate paper on spatial information systems.

The correspondence with Linda Lilburne and various Landcare research grants including, LRIS (C09626) objective 3 & 8, and soil quality (C09629).

REFERENCES

- Aamodt, A. & E. Plaza, 1994 Case-based Reasoning: Foundational Issues, Methodological Variations and System Approaches. *Artificial Intelligence Communications*, Vol.7, No.1.
- Agouris, P., Stefanidis, A., & M. J. Egenhofer, 1997 I. Q. Image Query by Sketch <http://www.spatial.maine.edu/~peggy/IQ.html>.
- Black, W. Hutchinson, G. & T. K. Siang 1997 System Design and Implementation of a Spatial Similarity System: BLASH. Information Science Dept, INFO408 Report, University of Otago, Dunedin, New Zealand, 54 pages.
- Bruns, T. & M. Egenhofer 1996 Similarity of Spatial Scenes, in: M.-J. Kraak & M. Molenaar (eds.), *Seventh International Symposium on Spatial Data Handling*, Delft, The Netherlands Taylor & Francis, pp. 173-184.
- Buttenfield, B. P., & R. B. McMaster (editors), 1991. *Map Generalization: Making Rules for Knowledge Representation*. New York: Longmont Scientific and Technical.
- Cain, D. H., K. Ritters, & K. Orvis. 1997. A Multi-Scale Analysis Of Landscape Statistics. *Landscape Ecology*, 12(4), p. 199-212.
- Cao, C., & N. S.-N. Lam. 1997. Understanding the scale and resolution effects in remote sensing and GIS. In *Scale in remote sensing and GIS*, D. A. Quattrochi & M. F. Goodchild, eds., Lewis Publishers, p. 57-72.
- Carbonell, J. G. 1986 Derivational analogy. a theory of reconstructive problem solving and expertise acquisition. In R.S. Michalski, J.G. Carbonell, and T.M. Mitchell, eds. *Machine Learning : An Artificial Intelligence Approach*, Vol. 2, pages 371--392. Morgan Kaufman Publishers, Los Altos, California.
- Cobb, M. A., Chung, M. J., Foley III, H., Petry, F. E., Shaw, K. B. & H. V. Miller, 1998, A Rule-based Approach for the Conflation of Attributed vector Data. *GeoInformatica: An International Journal on Advances of Computer Science for Geographical Information Systems*, Vol. 2, Number 1, 7-37.
- Cullinan, V. I., & J. M. Thomas. 1992. A comparison of quantitative methods for examining landscape pattern and scale. *Landscape Ecology*, 7(3), p. 211-227.
- Dubitzky W. Carville F. & J. Hughes 1993 Case-level Knowledge Modelling in CBR, *Irish Journal of Psychology*, 14:3 :478-479.
- Ehleringer, J. R., & C. B. Field (editors), 1993. *Scaling Physiological Processes, Leaf to Globe*. New York: Academic Press, Inc.
- Ellison T. M. 1997, Induction and Inherent Similarity. *SimCat 97 An Interdisciplinary Workshop on Similarity And Categorisation*, November, Department of Artificial Intelligence, University of Edinburgh.
- Elmasri, R. & Navathe, S. B. 1994, *Fundamentals of Database Systems*. The Benjamin/Cummings Publishing Company, Redwood City, C.A.
- Fayyad, U. M. 1997 Editorial. *Data Mining and Knowledge Discovery*, 1(1): 5-10.
- Flewelling, D. M. 1997, *Comparing Subsets from Digital Spatial Archives: Point Set Similarity*. Ph.D., University of Maine, Orono, Maine.
- Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D. & P. Yanker, 1995 Query by Image and Video Content: The QBIC System. *IEEE Computer* 28(9): 23-32.
- Gentner, D. & Forbus, K. D. 1991 MAC/FAC: A model of similarity-based retrieval. In *Proceedings of the 13th Annual Conference of the Cognitive Science Society*, p 504-509, Chicago.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2):155-170.
- Goodchild, M. F, Egenhofer, M. J. & R. Fegeas., 1998, Interoperating GISs. Report of a Specialist Meeting held under the auspices of the Varenus Project. http://www.ncgia.ucsb.edu/conf/interop97/interop_toc.html
- Grimnes M. and Aamodt A., 1996, A two layer case-based reasoning architecture for medical image understanding. In *Advances in Case-Based Reasoning, Third European Workshop, EWCBR-96* Smith and Faltings (eds.), pp. 164-178.
- Grimnes M. 1996 Personal communication.
- Gudivada, V. 1995 On Spatial Similarity Measures for Multimedia Applications, in: *SPIE*, pp. 363-372.
- Gudivada, V. and V. Raghavan 1995 Design and Evaluation of Algorithms for Image Retrieval by Spatial Similarity, *ACM Transactions on Information Systems*, 13 (2): 115-144.
- Hampton J. A. 1997, Similarity and Categorization. *SimCat 97 An Interdisciplinary Workshop on Similarity And Categorisation*, November, Department of Artificial Intelligence, University of Edinburgh.
- Higham, E. C. Holt, A. & G.W. Kearsley, 1996 Tourist Flow Reasoning: The Spatial Similarities of Tourist Movements. In the *Proceedings of the 8th Annual Colloquium of the Spatial Information Research Centre*, Otago University, Dunedin, New Zealand, pp 69-78.
- Holt, A. 1996a Allowing the Environment to Model Itself. *Environmental Perspectives, A Quarterly Newsletter published by the Environmental Policy & Management Research Centre*, Issue 10: 6-7.
- Holt, A. 1996b Incorporating A New Computational Reasoning Approach to Spatial Modelling. In the *proceedings of The 1st International Conference on GeoComputation*, University of Leeds, Leeds, England. 1: 427-442.

- Holt, A. 1997 GeoComputation-neologism, or gambit towards progressive research? *Environmental Perspectives*, A Quarterly Newsletter published by the Environmental Policy & Management Research Centre, Issue 15:5-6.
- Holt, A. & G. L. Benwell, (1996), Case-based Reasoning and Spatial Analysis. *Journal of the Urban and Regional Information Systems Association*, 8(1) :27-36.
- Holt, A. & G. L. Benwell 1997 Using Spatial Similarity for Exploratory Spatial Data Analysis: Some Directions. The 2nd International Conference on GeoComputation, University of Otago, Dunedin, New Zealand, pp. 279-288.
- Holt, A. & G. L. Benwell (In Press) Applying Case-based Reasoning to Spatial Phenomena, *The International Journal of Geographical Information Science*. 30 pages. (accepted for publication).
- Holt, A., Higham, E. C. & G.W. Kearsley, 1996 Elucidating International Tourist Movements: An Intelligent Approach. In the proceedings of Tourism Down Under II: A Tourism Research Conference, Centre for Tourism, University of Otago, Dunedin, pp. 167-180.
- Holt, A., Higham, E. C. & G.W. Kearsley, 1997 Predicting International Tourist Flows: Using a Spatial Reasoning System. *The Pacific Tourism Review: An Interdisciplinary Journal*. 1:4 30 pages.
- Holt, A. & R. E. MacLaury, (In Press) Spatial Vantages: Understanding Spatial Similarities. *Language Sciences: A Special Issue on Vantage Theory*, ed. N. Love.
- Hudson, J., 1992. Scale in space and time. In R. F. Abler, M. G. Markus, & J. M. Olson (Editors), *Geography's Inner Worlds: Pervasive Themes in Contemporary American Geography*. New Brunswick, NJ: Rutgers University Press, pp. 280-300.
- Indurkha, B., 1992 Metaphor and cognition: An interactionist approach. Kluwer Academic Publishers, Dordrecht.
- Indurkha, B., 1991 On the role of interpretive analogy in learning. *NGC*, 8(4):385-402.
- Jagdish, H. V. 1991 A retrieval technique for similar shapes. In the proceedings of the 10th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, pages 208-217, Denver, Colorado.
- Jagdish, H. V., Mendelzon, A. O. & T. Milo. 1995 Similarity based queries. *PODS*.
- Jain, A. K. & R. Hoffmann, 1988 Evidence-based recognition of 3D Objects. *IEEE Transactions On Pattern Analysis And Machine Learning*, Vol. 10, No. 6. pp. 783-801.
- Jeffery, N. Teather, D. & Teather, B. A., 1997, Case-Based Training Using Similarity and Categorization from a Multiple Correspondence Analysis. *SimCat 97 An Interdisciplinary Workshop on Similarity And Categorisation*, November, Department of Artificial Intelligence, University of Edinburgh.
- Jin, J. S., Greenfield H., & Kurniawati R., 1997 CBIR-VU: a new scheme for processing visual data in multimedia systems. *Lecture Notes in Computer Science: Visual Information Systems*, Leung C. H. C., Springer Verlag, pp40-65.
- Jones, E. K. & A. Roydhouse, 1994 Spatial Representations of Meteorological Data for Intelligent Retrieval. *The Sixth Annual Colloquium of the Spatial Research Centre, Proceedings*. Eds. G.L. Benwell and N.C. Sutherland. Dunedin, New Zealand. pp.45-58.
- Juristica, I. 1994 How to Retrieve Relevant Information?. In Russell Greiner (Ed.): *Proceedings of the AAAI Fall Symposium Series on Relevance*, New Orleans, Louisiana.
- Kasabov, N. & A. Ralescu 1993 *The Basics of Fuzzy Systems: Fuzzy System Applications*. A tutorial at The First New Zealand International Two-stream Conference on Artificial Neural Networks and Expert Systems 49 Pages.
- Keane M. 1997, Dynamic Similarity: The Zany World of Processing Similarity. *SimCat 97 An Interdisciplinary Workshop on Similarity And Categorisation*, November, Department of Artificial Intelligence, University of Edinburgh.
- Knauff, M. 1993 Introduction. Voß, A. (Ed.) *Similarity Concepts and Retrieval Methods*. FABEL-Report No. 13 Druck: Gesellschaft für Mathematik und Datenverarbeitung mbH (GMD), Sankt Augustin.
- Knauff M. & C. Schlieder, 1993 Similarity assessment and case representation in case-based design. In M. M. Richter, S. Wess, K. D. Althoff, and F. Maurer, editors, *First European Workshop on Case-Based Reasoning (EWCBR'93)* Vol.1, p 37-42.
- Kolodner J. 1993 *Case-Based Reasoning*. San Mateo, Morgan Kaufmann, Publishers.
- Lees, B. G. 1997 Data Questions in GeoComputation. The 2nd International Conference on GeoComputation, University of Otago, Dunedin, New Zealand, pp. 289-296.
- Lilburne, L. (Ed), 1997. *Proceedings of the workshop: Modelling the environment - the scaling problem*. Landcare Research NZ Ltd., Lincoln. New Zealand, 74 pages.
- Lilburne, L. 1998. Scale issues in environmental data modelling. Internal Report, Landcare Research NZ Ltd., Lincoln, New Zealand.
- MacLaury R. E. 1997, Vantage Theory in Cognitive Science: An Anthropological Model of Categorization and Similarity Judgement. *SimCat 97 An Interdisciplinary Workshop on Similarity And Categorisation*, November, Department of Artificial Intelligence, University of Edinburgh.
- O'Hara, S, 1992 A model of the 'redescription' process in the context of geometric proportional analogy problems. In K. P. Jantke, editor, *Proceedings of the International Workshop on Analogical and Inductive Inference*, pages 268-293. SpringerVerlag.

- O'Hara, S. & B. Indurkha, 1993 Incorporating (re) interpretation in case-based reasoning. In M. M. Richter, S. Wess, K. D. Althoff, and F. Maurer, editors, EWCBR p 154-159, Kaiserslautern.
- Openshaw, S. & R. J. Abraham 1996 GeoComputation. In the proceedings of The 1st International Conference on GeoComputation, University of Leeds, Leeds, England. 1: 665-666.
- Ortony, A. 1979 Beyond literal similarity. *Psychological Review*, 86:161-180.
- Osborne, H. & D. Bridge 1997, Models of Similarity for Case-Based Reasoning. SimCat 97 An Interdisciplinary Workshop on Similarity And Categorisation, November, Department of Artificial Intelligence, University of Edinburgh.
- Papadias, D. & M. J. Egenhofer, 1997, Algorithms for Hierarchical Spatial Reasoning. *GeoInformatica: An International Journal on Advances of Computer Science for Geographical Information Systems*, Vol. 1, Number 3, 251-274.
- Papadias, D. & Delis, B. 1997 Relation-based Similarity. Proceedings of the 5th ACM Workshop on GIS, Las Vegas, ACM Press.
- Pawlak, Z., Grzymala-Busse, J. W., Slowinski, R. & W. Ziarko, 1995, Rough Sets. *CACM* 38(11): 88-95.
- Quattrochi, D. A., & M. F. Goodchild (editors), 1997. *Scaling in Remote Sensing and GIS*. Boca Raton, FL: CRC/Lewis Publishers, Inc.
- Rodriguez A. R. 1997, Combining Different Domain Models into a Contextual Similarity Function. SimCat 97 An Interdisciplinary Workshop on Similarity And Categorisation, November, Department of Artificial Intelligence, University of Edinburgh.
- Savitsky, B. & Anselin, L. 1997. Scale. http://www.ncgia.ucsb.edu/other/ucgis/research_priorities/paper6.html.
- Schank, R. 1982, *Dynamic memory: A theory of leaning in computers and people*. Cambridge University press. New York.
- Seixas J. & J. Aparico 1994 A Framework for Spatial Reasoning the Task of Image Interpretation. EGIS, <http://www.odyssey.ursus.maine.edu/gisweb/spatdb/egis/eg94015.html>
- Sivapalan, M., & J. D. Kalma, 1995 Scale problems in hydrology: Contributions of the Robertson Workshop. *Hydrological Processes* 9(3/4):243-250.
- Tversky, A. 1977 Features of Similarity. *Psychological Review* 84(4): 327-352.
- Tversky, A. & D. H. Krantz, 1970 The dimensional representation and the metric structure of similarity data. *Journal of Mathematical Psychology*, Vol. 7. p572-597.
- Voß, A. 1993 *Similarity Concepts and Retrieval Methods*. FABEL-Report No. 13 Druck: Gesellschaft für Mathematik und Datenverarbeitung mbH (GMD), Sankt Augustin.
- Wallace, D. Fraser, W. & Nicol L. 1997 *System Design and Implementation of a Spatial Similarity System: WADAL incorporating GeoMatch*. Information Science Dept, INFO408 Report, University of Otago, Dunedin, New Zealand, 62 pages.
- Wallach, M. A. 1958. On Psychological similarity. *Psychological Review*, 65(2):103-116.
- Watson, I.D. 1994 The Case for Case-Based Reasoning. In, Proc. Information Technology Awareness in Engineering Conference, 21-22 November 1994, London.
- Watson, I. D. 1997 *Applying Case-based Reasoning: Techniques for Enterprise Systems*. Morgan Kaufmann Publishers, Inc. California 289 pages.
- Wong, D., & C. Amrhein (eds), 1996 *The Modifiable Areal Unit Problem*. Special issue of *Geographical Systems* 3:2-3.
- Zadeh, L. 1965 Fuzzy sets. *Information and Control*, Vol.8. pp 338-353.