Theory of Spatial Similarity Relations and Its Applications in Automated Map Generalization

by

Haowen Yan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Geography

Waterloo, Ontario, Canada, 2014

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Automated map generalization is a necessary technique for the construction of multi-scale vector map databases that are crucial components in spatial data infrastructure of cities, provinces, and countries. Nevertheless, this is still a dream because many algorithms for map feature generalization are not parameter-free and therefore need human's interference. One of the major reasons is that map generalization is a process of spatial similarity transformation in multi-scale map spaces; however, no theory can be found to support such kind of transformation.

This thesis focuses on the theory of spatial similarity relations in multi-scale map spaces, aiming at proposing the approaches and models that can be used to automate some relevant algorithms in map generalization. After a systematic review of existing achievements including the definitions and features of similarity in various communities, a classification system of spatial similarity relations, and the calculation models of similarity relations in the communities of psychology, computer science, music, and geography, as well as a number of raster-based approaches for calculating similarity degrees between images, the thesis achieves the following innovative contributions.

First, the fundamental issues of spatial similarity relations are explored, i.e. (1) a classification system is proposed that classifies the objects processed by map generalization algorithms into ten categories; (2) the Set Theory-based definitions of similarity, spatial similarity, and spatial similarity relation in multi-scale map spaces are given; (3) mathematical language-based descriptions of the features of spatial similarity relations in multi-scale map spaces are addressed; (4) the factors that affect human's judgments of spatial similarity relations are proposed, and their weights are also obtained by psychological experiments; and (5) a classification system for spatial similarity relations in multi-scale map spaces is proposed.

Second, the models that can calculate spatial similarity degrees for the ten types of objects in multi-scale map spaces are proposed, and their validity is tested by psychological experiments. If a map (or an individual object, or an object group) and its generalized counterpart are given, the models can be used to calculate the spatial similarity degrees between them.

Third, the proposed models are used to solve problems in map generalization: (1) ten formulae are constructed that can calculate spatial similarity degrees by map scale changes in map generalization; (2) an approach based on spatial similarity degree is proposed that can determine when to terminate a map generalization system or an algorithm when it is executed to generalize

objects on maps, which may fully automate some relevant algorithms and therefore improve the efficiency of map generalization; and (3) an approach is proposed to calculate the distance tolerance of the Douglas-Peucker Algorithm so that the Douglas-Peucker Algorithm may become fully automatic.

Nevertheless, the theory and the approaches proposed in this study possess two limitations and needs further exploration.

- More experiments should be done to improve the accuracy and adaptability of the proposed models and formulae. The new experiments should select more typical maps and map objects as samples, and find more subjects with different cultural backgrounds.
- Whether it is feasible to integrate the ten models/formulae for calculating spatial similarity degrees into an identical model/formula needs further investigation.

In addition, it is important to find out the other algorithms, like the Douglas-Peucker Algorithm, that are not parameter-free and closely related to spatial similarity relation, and explore the approaches to calculating the parameters used in these algorithms with the help of the models and formulae proposed in this thesis.

Acknowledgements

It is a pleasure to thank the many people who made the completion of this thesis possible.

Above all, I would like to thank the members of my committee: Dr. Peter deadman, Dr. Jane Law and Dr. Su-yin Tan for their constructive comments and insightful advice.

A special thank goes to Dr. Douglas Dudycha for his kind help to me as the chairperson of the comprehensive exam.

A deep gratitude goes to my supervisor, Dr. Jonathan Li. I would thank him for his great advising work, for his insightful suggestions on every step I made in my study and research toward the completion of my thesis, for his conscientious input in reading and editing my writings, and most of all, for his unchanging encouragement and support throughout the years of my PhD study. This thesis would not have been possible without his help.

I would particularly thank the graduate program administrators, Mrs. Lynn Finch and Ms. Susie Castela, two amazing ladies with high professionalism and angel‘s hearts. Their kindness and their sincere help and support in my study are worth to be remembering forever.

I feel so indebted to my friends who helped me to collect various maps, do psychological experiments and analyze statistical data. They are Dr. Tao Liu, Dr. Zhonghui Wang, Mr. Yun Li, Miss Yaping Li, Miss Ruirui Zhang, Mr. yibo Niu, Mr. Libin Niu. Their friendships, their faith in my potential and their consistent encouragement and support are so invaluable to me.

My heart-felt thanks also go to those wonderful friends while my stay in Waterloo, Linlin Xu, Xiao Xu, Zhengzhong Si, Haiyan Guan, Wenxia Tan, and Yuanming Shu. They either provided good suggestions and sources for my research or gave me warm help and care in life. I sincerely cherish their friendships.

I appreciate very much the OGS (it has funded me from May 2011 to April 2012) and the NSERC (it has funded me from May 2012 to April 2014) for their financial support to the work describe in the thesis.

I am very grateful to the Chinese Academy of Survey & Mapping, the National Centre of Geomatics, China, the Map Academy of Gansu Province, China, and the Centre of Geomatics, Gansu Province, China. Many maps used in this thesis are provided by them.

I also thank my friends Dale Quesnel and Jean Quesnel very much, a very nice couple helping me in oral English as well as my life in Canada for three years.

Lastly, deepest gratitude goes to my life partner Weifang and my son Alex, who have been there every step of the way. I couldn’t have completed this thesis without their unwavering and unconditional love, patience, and encouragement. I really can’t thank them enough.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Background and Motivation

Multi-scale vector map database is one of the most fundamental components in the national spatial data infrastructure (NSDI), because vector map data provides geographically spatial positioning bases for various location-based services in the communities of politics, economy, military, environment, traffic, transportation, and telecommunication, etc., and plays an important role in the construction of digital cities (Yan, 2010).

Traditionally, a multi-scale map database of a region is built manually or semi-automatically by means of so-called "multiple-version method" (Wang, 1993), i.e. the maps of the region at multiple scales are digitized, processed, and saved in different databases that are characterized by their map scales to form a large database (Figure 1-1). For example, to build a digital map database containing maps at scales 1:10K, 1:50K, 1:250K and 1:1000K using the multiple-version method, the maps at the four scales are firstly collected and compile; and then they are digitized and edited; and last, the map data at each scale is stored in a corresponding map database, respectively. The combination of the databases at the four scales constitutes a multi-scale database of the region.

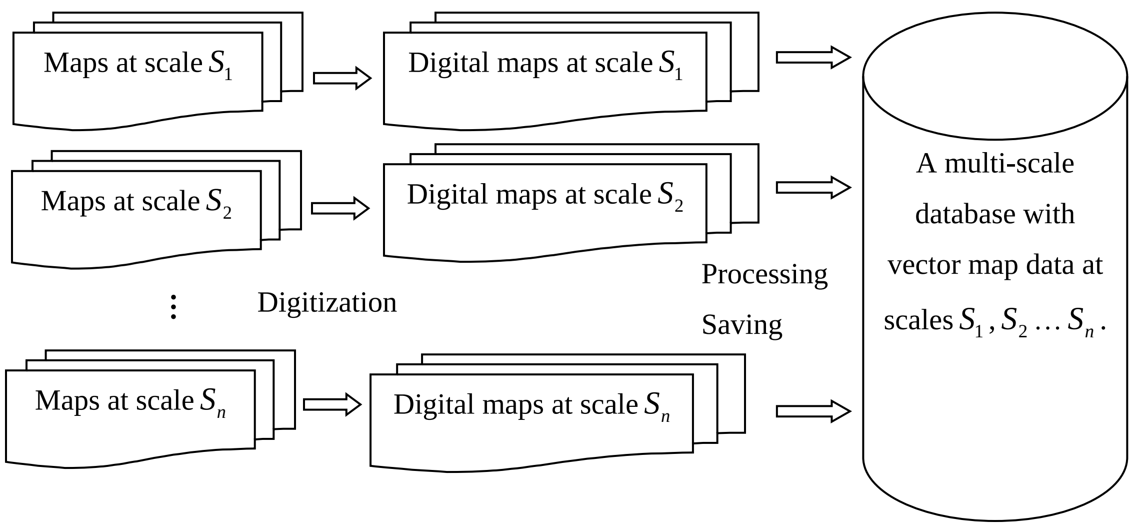


Figure 1-1 Construction of a multi-scale database using the multiple-version method.

The multiple-version method has dominated multi-scale map data generation for decades. As a result, almost all of the existing multi-scale vector map databases have been established using this method and these databases have been used in many countries for decades. Nevertheless, previous studies and practical applications have discovered that such multi-scale map databases have a number of shortcomings that need to be overcome (Wang, 1993; Ruas, 2001):

(1) repeated storage of map data at different scales generate a lot of redundant data in multi-scale map databases and leads to the waste of computer memory spaces;

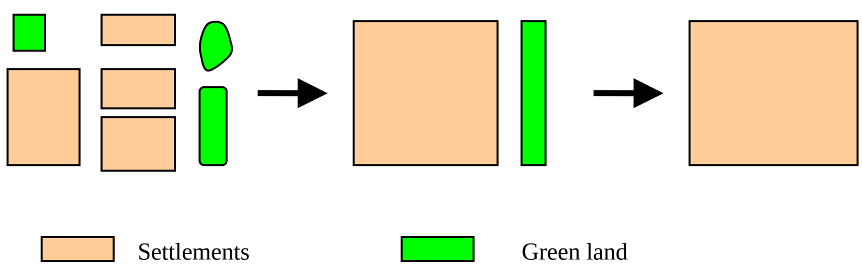
(2) storing multi-scale maps of a region greatly increases the quantity of the data and therefore decrease the efficiency of the data transmission via the Internet;

(3) consistency of the map data at different scales cannot be ensured due to repeated compilation and digitization of the maps at different scales of the same region; and

(4) renewal of multi-scale map databases is time-consuming and uneconomical.

1:100K **1:250K**

(a) Graphics transformation (http://wenku.baidu.com/view/50c230250722192e4536f6dd.html)



(b) Semantic transformation

Figure 1-2 Similarity transformation in map generalization

A most prospective method that can overcome the above disadvantages due to the multi-version method is automated map generalization (Ruas, 1998; Weibel and Jones, 1998). Automated map generalization is a technique for solving spatial conflicts and congestions that appear in the process of generating smaller scale maps from larger scale ones using various appropriate algorithms and operators (e.g. selection, displacement, simplification, etc.) under definite conditions (e.g. map scale, map purpose, etc.). If automated map generalization comes true in the construction of multi-scale map databases, cartographers do not need to do repeated compilation and digitization for building multiple versions of map databases, but build only one map database using the maps at the largest scale. When any of the map databases at the other scales is needed, they can produce it using the one at the largest scale by means of automated map generalization. This, undeniably, is an ideal method for building multi-scale map databases.

In essence, map generalization (it is also called cartographic generalization, sometimes) is a kind of similarity transformation in graphics and semantics. Take Figure 1-2(a) as an example: the islands on the map at scale 1:250K are generated from the map at scale 1:100K. Although the original map has been simplified in the process of scale change, the two maps of the same area keep their similarity in graphics. In Figure 1-2(b): combination of the polygons is a kind a similarity transformation in semantics.

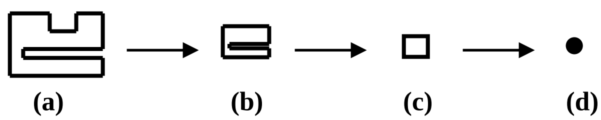


Figure 1-3 Generalization of a settlement.

It is evident that the similarity degree (or similarity value in some literature) between a generalized map and the original map and the scale of the generalized map are dependent on each other. The more the original map is generalized, the larger the scale changes from the original map to the generalized map (Figure 1-3). Nevertheless, no achievement has been made on quantitatively describing the relation which leads to the question “how to calculate the spatial similarity degree between a map and its generalized counterpart” (Yan, 2010) unsolved. This hampers the automation of map generalization, because:

(1) if similarity degrees are not known, a map generalization system/software does not know to what extent an original map should be generalized to produce a resulting map; and

(2) the system/software also does not know when to terminate a map generalization procedure if its parameters depend on spatial similarity degrees.

The above discussion reveals that calculation of similarity degrees of maps at different scales is of great importance in the automation of map generalization. Automated map generalization cannot be realized by cartographers and geographers until this problem is solved.

1.2 Significances of Spatial Similarity Relations

Similarity has aroused great interests of many researchers in the communities of cartography (Yan 2010), geographic information science (Rodríguez and Egenhofer, 2003; Rodríguez and Egenhofer, 2004), mathematics, psychology (Tversky, 1977) and computer science (Budanitsky and Hirst, 2001). As far as geography-related fields such as cartography and geographic information science are concerned, the significances of similarity relations at least can be seen in the theory of spatial relations, spatial description, spatial reasoning and spatial query/retrieval, spatial recognition, and automated map generalization.

1.2.1 Theory of Spatial Relations

Spatial relations, including distance, topological, direction, and similarity relations are essential tools for describing and expressing the geographic space, and they play important roles in the theories of Geo-Sciences. In the past decades, many achievements have been made on distance relations (Hong, 1994), topological relations (Egenhofer and Franzosa, 1991; Du et al., 2008), and direction relations (Peuquet, 1986; Goyal, 2000, Yan et al., 2006), but little work has been done on spatial similarity relations (Yan, 2010).

1.2.2 Spatial Description, Spatial Reasoning and Spatial Query/Retrieval

The function of similarity relations in spatial descriptions and reasoning is too evident to require strict academic proofs (Guo, 1997). Inductive reasoning and memory retrieval (Goldstone, 2004) depend on similarity to get cues from previous events. Similarity is also the basic element for analogical inference (Markman, 1997).

A well-known example of similarity relations used in spatial description and spatial reasoning is the setup of the theory of “plate tectonics” by German geologist Alfred Wegener (Figure 1-4). The theory is built on the old concepts of continental drift and describes the large scale motions of Earth's lithosphere. Obviously, the complementary similarity of the plate boundaries provides most strong and direct proof for this theory: the researchers found the phenomena by drawing the maps of continental boundaries (i.e. a kind of description of graphics similarity) and then matching the boundaries that have complementary similarity relations (reasoning using similarity).

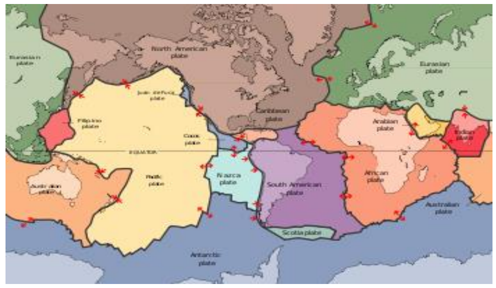


Figure 1-4 The tectonic plates of the world.

(from http://en.wikipedia.org/wiki/Plate_tectonics)

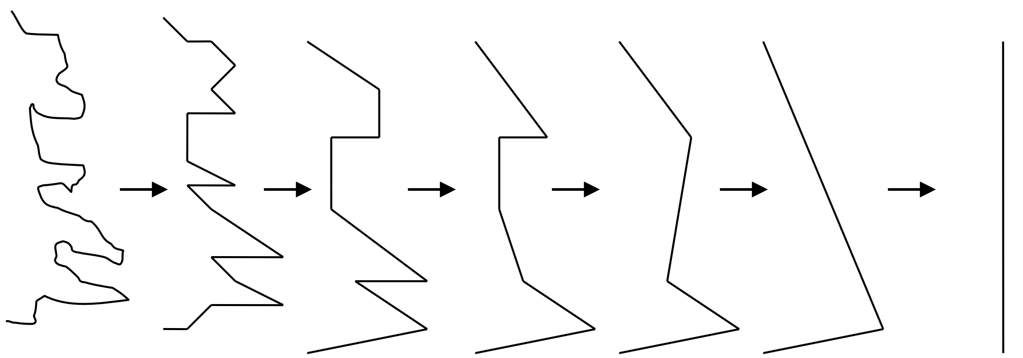
Spatial similarity plays the same role in the process of spatial information retrieval, spatial information integration, and spatial data mining (Rodríguez and Egenhofer, 2003; Rodríguez and Egenhofer, 2004). Using spatial similarities among spatial scenes to retrieve information, get interconnection among different databases, and find similar spatial objects or spatial phenomena have become or are becoming very common in many geographic information systems. For example, the similarity-based image query/retrieval has been used to substitute the match-based image query/retrieval (Petraglia et al., 2001) in recent decades. The main difference between the match-based and the similarity-based searches is: the result of a match-based search is a partition of the database in the set of images that match the query and the set of images that do not; while the result of a similarity-based search is a permutation of the whole database (Santini and Jain, 1996), to be exact in many cases, a sorting with respect to the similarity criterion.

1.2.3 Spatial Recognition

Similarity plays a fundamental role in human's spatial cognition (Li and Fonseca 2006). It serves as a principle for categorization (Tversky, 1977; Goldstone, 2004). Indeed, many theories assume that categorization depends on the similarity of the samples (Medin et al., 1993). It is popular that people tend to put those with more similarity into same groups. Such a typical example in geography is that geographers can easily classify relief into different categories (e.g. plateaus, hills, dunes, cliffs, etc.) according to the similarity degrees of the curvature, shapes and density of the contour lines on the maps. In addition, pattern recognition using images is a kind of similarity-based work, because one of its basic principles is to search the image to find the adjacent pixels having similar attributes (e.g. colour) with a prior known criterion (e.g. extracting a road from an image).

1.2.4 Automated Map Generalization

In automated map generalization, spatial similarity relation is of great significance to solve at least the following three problems.



(1) (2) (3) (4) (5) (6) (7)

Figure 1-5 Line simplification and similarity change

First, it can make some semi-automatic algorithms fully automatic.

An algorithm can generate maps at different levels of detail (LODs) using the same map if different generalization criterions are adopted. Such criterions are usually one or more thresholds in the algorithm. For example, in the Douglas-Peucker algorithm (Douglas & Peucker, 1973), distance tolerances are used as the thresholds in curve simplification. Different distance tolerances can generate different results if the Douglas-Peucker algorithm is used to simplify a curve (Figure 1-5). Nevertheless, as far as a map generalization software is concerned, such threshold values are not prior known but they usually need to be given by users or cartographers according to their experiences and experiments before the beginning of a map generalization project. The determination of the thresholds takes into account the original map scale and the resulting map scales as well as the purpose of the resulting maps. Input of the thresholds cannot avoid interrupting the map generalization procedure and therefore unfavorable to the full automation of map generalization.

Hence, it is necessary to find methods for automatically obtaining such threshold values. One of the evidences that cartographers can easily noticed is that the threshold values and map scales are dependent on each other. For example, in the Douglas-Peucker Algorithm, the greater the distance value, the simpler the curve will be simplified, and the smaller the resulting map scale should be. On the other hand, map scales are also closely related to similarity degrees between each generalized map and the original map.

If the approaches to calculating the similarity degree between two maps are known, it is possible to find out an approach for calculating the threshold value if the scale of the resulting map is given. Based on the threshold value, the algorithms can become parameter-free and therefore fully automatic. In this sense, calculation of similarity degrees between two maps is of great importance in automated map generalization.

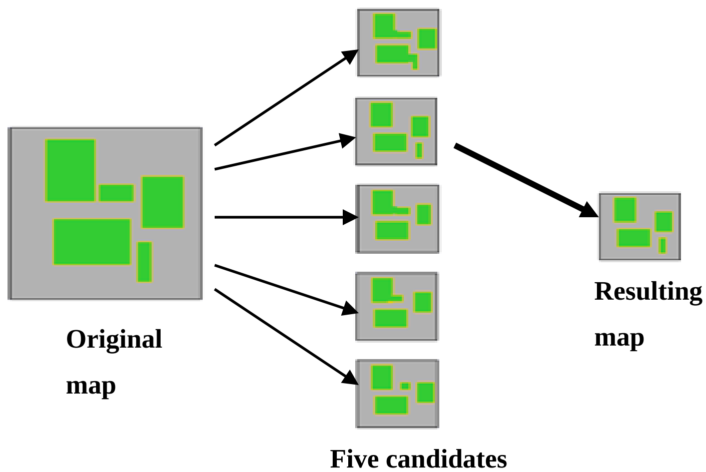


Figure 1-6 Multiple candidate maps in map generalization.

Second, it helps to determine when to terminate a map generalization algorithm/procedure.

Map generalization in the digital era depends on map generalization systems. A map generalization system is a combination of various algorithms. Generally, each algorithm is developed for generalizing specific map features. Although the Radical Law (Töpfer & Pillewizer, 1966) can determine "how many features can be retained on the resulting maps", "which features should be retained" and "to what extent the feature can be simplified" are unsolved yet. The two questions depend on calculation of spatial similarity relations between original map features and generalized map features which is in suspense by far; therefore, when to stop the relevant map generalization algorithms is unsolved yet.

It helps to select appropriate algorithms for map generalization systems (Yan et al., 2006).

Supposing that a map is given and it needs to be generalized to get another map at a specific scale. In manual way, it is usually true that different cartographers produce different maps (Figure 1-6). Here a problem arises: "which map is the best and should be the resulting map?" Cartographers solve this problem by comparing each of the generalized maps with an "imaginary" map (this map usually does not exist in the physical world but in the cartographers' brains "generated" by the cartographers' experiences and knowledge) and choose the one that has the greatest similarity degree with the "imaginary" map (Yan, 2010).

The same situation appears in map generalization aided by software systems: different algorithms usually generate many different maps using the same original map data, and the systems need to judge which map should be selected as the result map. Unfortunately, it is impossible to get those "imaginary" maps from experts' (i.e. cartographers') brains according to current research achievements in relevant communities, such as Mathematics, Cognitive Psychology, Computer Science and Geomatics.

An alternatively applicable way may be to calculate the similarity degree between the original map and each generalized map, and select the one with the greatest similarity degree as the resulting map. The reason for this is that: the more similar the two maps are, the more common information the two maps contain. This is coincident with the principle of information transmission in map generalization, i.e. map generalization should transmit information as more as possible.

In sum, approaches for calculating spatial similarity degrees take important roles in full automation of map generalization. So how to calculate similarity degrees between maps in multi-scale map spaces is worthy of a thorough investigation.

1.3 Classification of Objects in Multi-scale Map Spaces

It is necessary to give an introduction of the classification of the map objects prior to the presentation of the objectives of this study.

Above all, this work emphasizes on topographic maps.

A topographic map is a detailed and accurate graphic representation of cultural and natural features on the ground (Harvey, 1980). For many nations, topographic map series is an important resource in planning infrastructure and resource exploitation (Kraak and Ormeling, 1996). In the digital era, topographic maps are usually divided into different feature layers, and then the feature layers are separately digitized and stored to form databases (Harley and Woodward, 2005).

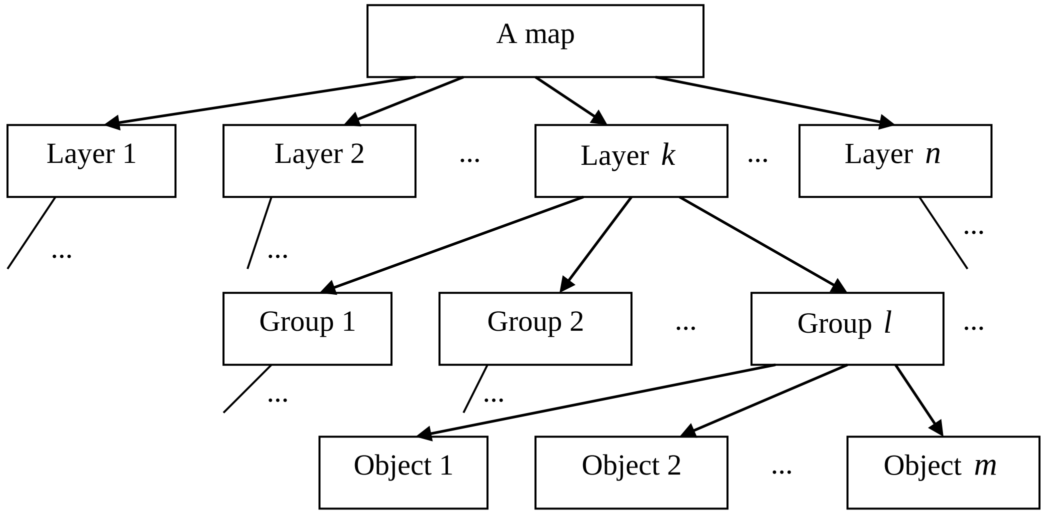


Figure 1-7 Hierarchy of topographic maps.

A topographic map can be generalized to get another map at a smaller scale. A map generalization process usually abides by so-called “divide-and-conquer” police to make it simple and efficient. To be exact, map generalization operators/algorithms generally operate on each of the feature layers, or on each group of objects, or even on individual objects, or on the whole of the map. After generalization, the individual objects and the groups of objects are organized to form feature layers, and the feature layers are organized and stored to form a new map. The theory of spatial similarity relation in this study aims at providing a tool to automate and control the process of map generalization; hence, the following four hierarchical levels of spatial similarity relations in topographic map generalization need to be calculated so that the four kinds of corresponding operators/algorithms can be developed in automated map generalization (Figure 1-7). They are

(1) spatial similarity relations between a map and its generalized counterparts at different map scales,

(2) spatial similarity relations between a map feature layer and its generalized counterparts at different map scales,

(3) spatial similarity relations between an object group and its generalized counterparts at different map scales, and

(4) spatial similarity relations between an individual object and its generalized counterparts at different map scales.

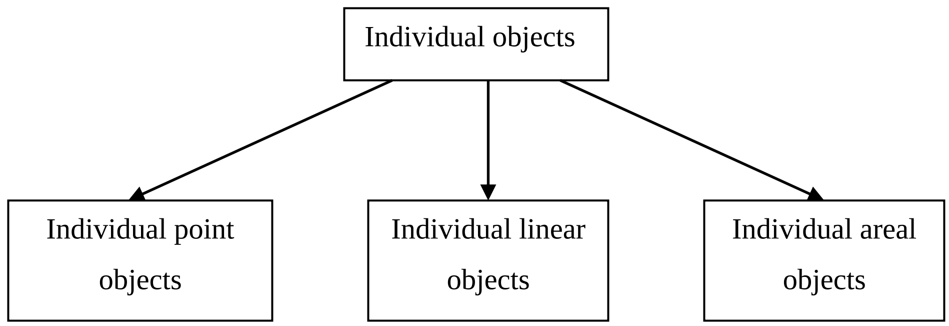


Figure 1-8 Classification of individual objects on maps

Individual objects on two-dimensional (2D) maps include individual point objects, individual linear objects and individual areal objects (Figure 1-8). They refer to discrete phenomena occurring at isolated locations and they are symbolized using separated points, lines, or polygons.

- Individual point object: such as a wells in a desert or a historic pavilion, usually represented using a point symbol on the map. It is zero-dimensional, small in size but important and need to be retained on the map.
- Individual linear object: such as a road, a river, or a ditch, symbolized using a line or a curve on the map. It is one-dimensional (1D).
- Individual areal object: such as a forest, a lake, or a parking lot. It is 2D and has length and width, and symbolized using a polygon.

Object groups can be classified into a number of categories according to the geometric characteristics of map features (Figure 1-9), i.e. point clouds, linear object groups, and areal object groups. Further, linear object groups are classified into parallel line clusters, intersected line networks and tree-like networks; areal object groups are classified into discrete polygon groups and connected polygon groups. The following gives a detailed explanation of these categories.

- Point cloud: such as control points in a region, trees alongside of a river or a road, etc.
- Parallel line cluster: a typical example of this is contour lines.

· Intersected line network: various roads in a city form an intersected line network.

· Tree-like network: a river basin consisting of a main stream and many branches forms a typical tree-like network.

· Discrete polygon group: such as settlements scattering in countryside.

· Connected polygon group: a typical example of this is the polygons on a land use map.

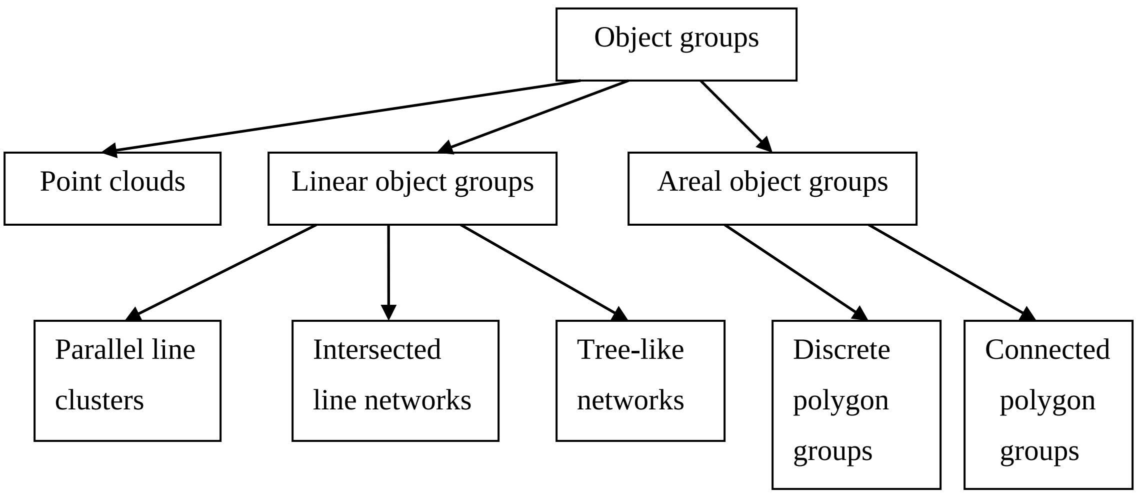


Figure 1-9 Classification of object groups on maps

1.4 Definitions of Map Scale Change

Map scale change is a most important concept that is used throughout the thesis and plays crucial role in many models and formulae, so it is separated from other concepts and defined here.

There are two maps M_0 *and* M_1. *Their scales are* S_0 *and* S_1, *respectively.* M_1 *is a generalized map of* M_0. *The ratio* $\frac{S_0}{S_1}$ *is called the map scale change from map* M_0 *to map* M_1.

1.5 Research Objectives

In order to construct the theory of spatial similarity relations and put it into use to improve the efficiency of many relevant algorithms in map generalization, the following objectives should be reached in this study.

- Fundamental theories of spatial similarity relations, including:

(1) a definition of spatial similarity relations in multi-scale map spaces;

(2) features of spatial similarity relations in multi-scale map spaces;

(3) factors that affect humans' judgments of spatial similarity relations in multi-scale map spaces; and

(4) a classification system for spatial similarity relations in multi-scale map spaces.

These problems are the basis of the calculation models/measures for spatial similarity relations. To prepare for constructing quantitative calculation models/measures, the definitions, features, and factors of spatial similarity relations should be given in mathematical languages in this research. Their correctness and validity should be systematically tested so that they can be acceptable by majority of people.

• Calculation approaches/models/measures of spatial similarity relations in multi-scale map spaces, including:

(1) approaches to calculating spatial similarity degrees between two individual point/linear/polygonal objects on maps at different scales;

(2) approaches to calculating spatial similarity degrees between two object groups (i.e. point clouds, parallel line clusters, intersected line networks, tree-like networks, discrete polygon groups, and connected polygon groups) on maps at different scales; and

(3) approaches to calculating spatial similarity degrees between a map and a generalized counterpart of the map at smaller scale.

• Application of the theories of similarity relations in automated map generalization, including:

(1) approach to calculating spatial similarity degrees between a map and its generalized counterpart at smaller scale taking map scale change as the independent variable;

(2) approach to calculating the distance tolerance of the Douglas-Peucker Algorithm; and

(3) approach to determining when a map generalization system or a map generalization algorithm should be terminated in the process of map generalization.

The three goals of the research are dependent on each other (Figure 1-10). The Fundamental theories of spatial similarity relations are the foundation of the research. The Calculation approaches to spatial similarity relations are based on the fundamental theories and are the main body and also the most important and most difficult part of the study. The applications of the theory verify the theories and models, and test their validity in the meanwhile. Only after successful applications of the theory in map generalization are the three objectives reached.

1.6 Scope of the Study

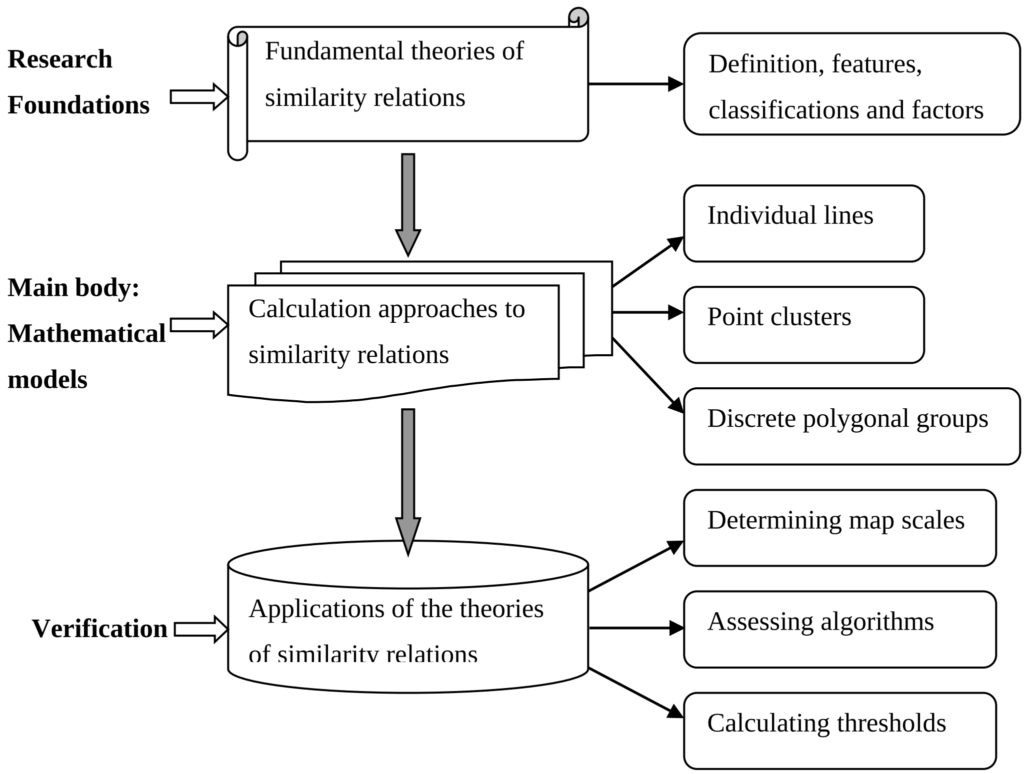


Figure 1-10 Relations among the research objectives

This study is limited in a scope that needs to be clarified before further discussion.

Above all, objects in this study generally refer to points, curves/lines, and polygons in two dimensional spaces (i.e. two dimensional map spaces). Elevations of the objects are not taken into account unless otherwise stated or specified.

Second, source data used in this study are vector map data unless otherwise stated or specified.

Third, correctness and validity of proposed models/approaches/measures should be tested by human being's spatial cognition, but not only by mathematical deductions; because judgments of spatial similarity root in and serve for human being's spatial cognition.

Fourth, although map scales are often used in this study, the research achievements should not be limited to a specified scope of map scale.

Last, although judgments of spatial similarity depend on people's age, gender, cultural background, etc., it is not preferable to test a large number and a great variety of subjects in our experiments at the beginning of constructing a theory. On the contrary, Survey on a small number of subjects may simplify the research work.

1.7 Thesis Outline

The thesis is divided into seven chapters to reach two goals: (1) establishment of the theory of spatial similarity relations in multi-scale map spaces, and (2) applications of the theory for the purpose of solving many related problems in automated map generalization.

Chapter 1: the background, significance and objectives of the study are addressed in the introduction, emphasizing on answering the questions "where does this study from?", "why is this topic worthy of serious studying?" and "what are researched in this thesis?"

Chapter 2: the achievements in the definitions, features and classification of similarity in various areas are reviewed, and existing models for calculating spatial similarity relations are discussed and their advantages and disadvantages are analyzed.

Chapter 3: the definitions, features and classification system of spatial similarity relations, and the factors that affect human's direction judgments are proposed and discussed in detail.

Chapter 4: the ten models for calculating spatial similarity degrees in multi-scale map spaces between various object pairs are constructed.

Chapter 5: the validity of the ten models is tested by psychological experiments.

Chapter 6: the theory of spatial similarity relations is used in automated map generalization and three goals are achieved: (1) the formulae for calculating the relations between map scale change and spatial similarity degree are constructed, (2) an approach to automatically determine when to terminate a map generalization algorithm/system is proposed, and (3) an approach for determining the distance tolerance used in the Douglas-Peucker Algorithm is presented.

Chapter 7: the overall summary, major innovations and contributions, limitations, and further research are presented in this concluding chapter.

Appendices: basic logic symbols are listed in appendix A which helps to understand the many formulae in the thesis; and my publications during the PhD study are listed in appendix B which may be assistance for examiners and committee members to evaluate my research work.

Chapter 2 Literature Review and Analysis

This chapter reviews the literature of the study.

This study emphasizes on three issues: the fundamental theory, the calculation models and the applications of spatial similarity relations; however, no applications of spatial similarity relations in automated map generalization can be found. Hence, this chapter only reviews the fundamental theory (including the definitions, features and classifications of spatial similarity relations) and the models for calculating similarity relations (including the models in various other disciplines and a number of raster-based models in geography).

2.1 Definitions of Similarity

Seemingly, similarity is a very simple concept. People encounter and use similarity almost every second in daily life. For example, people can recognize familiar persons by their faces if they meet. When judging the similarity of faces, someone may say that two human faces are similar if they have a common skin color, while someone else would require the identity of the geometric structure of facial features like the eyes, the nose, the mouth, etc.

Similarity also plays a crucial role in many fields in science (Gower, 1971; Bronstein et al., 2009). A typical example in geometry is "similar triangles": two triangles are similar if the three pairs of corresponding sides are proportional or two pairs of corresponding angles are congruent. In computer science, the definition of similarity, in many cases, is closely relative to character processing (e.g. comparing similarity of character strings). In pattern recognition, with a slight exaggeration, it may be true that all pattern recognition problems are based on finding methods for giving a quantitative interpretation of similarity (or equivalently, dissimilarity) between objects (Bronstein et al. 2008).

2.1.1 Definitions of Similarity in Various Fields

We cannot find unique definition of similarity from existing literatures. Every field has its criterion to define similarity for the purpose of solving a group of problems. Hence, the following discusses the definitions of similarity in several different fields, aiming at providing useful cues for our definition of spatial similarity relations in multi-scale map spaces.

Definition in Geometry

In geometry, two objects are called similar if both of them have the same shape. In other words, one of the two objects is congruent to the result of a uniform scaling (enlarging or shrinking), rotating, and repositioning of the other one. It is obvious that all circles are similar to each other (so are all squares and all equilateral triangles). On the other hand, two ellipses are not always similar to each other, nor are two hyperbolas.

People can easily judge whether two triangles are similar or not by comparing their corresponding angles or sides (Figure 2-1). However, if the concept of similarity extends to polygons with more than three sides, the criterion becomes different; because equality of all angles in sequence is not sufficient to guarantee similarity of two polygons. For example, all rectangles are not always similar (Figure 2-2).

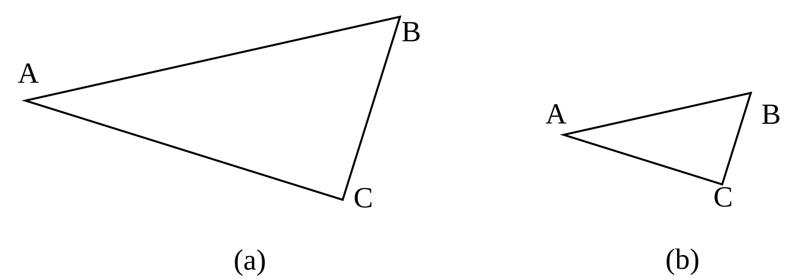
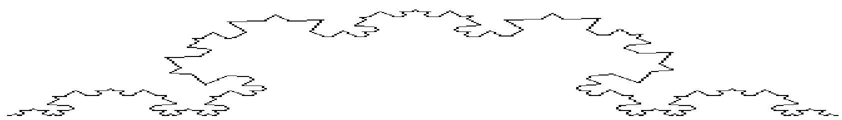


Figure 2-1 Similar triangles.

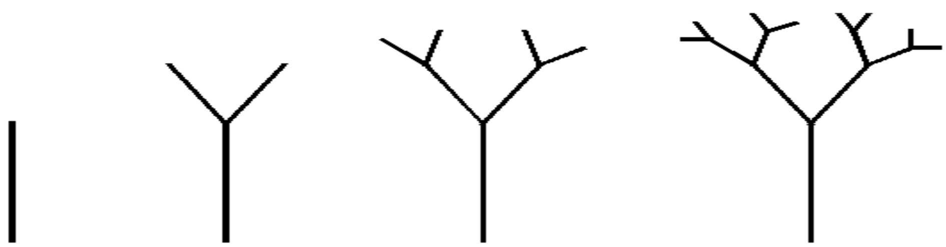
Self-similarity is another notable concept related to similarity in geometry, and it has also been a hot issue in geometry for decades. Self-similarity means an object is exactly or approximately similar to a part of itself. In other words, the whole has the same shape as one or more of the parts. Indeed, many geometric objects are statistically self-similar. Taking a coastline as an example (Figure 2-3(a)), parts of a coastline show the similar statistical properties at many scales (Mandelbrot 1967).

Figure 2-2 Dissimilar rectangles.

Fractal tree (Figure 2-3(b)) clearly shows the idea of self-similarity. Each of the branches is a smaller version of the main trunk of the tree. The main idea in creating fractal trees or plants is to have a base object and then to create smaller, similar objects protruding from that initial object. The angle, length and other features of these "children" can be randomized for a more realistic look. This method is a recursive method, meaning that it continues for each child down to a finite number of steps. At the last iteration of the tree or plant you can draw a leaf of some type depending on the nature of the plant or tree that you are trying to simulate.



(a) coastlines



(b) trees

Figure 2-3 Two examples of self-similarity.

Definition in Computer Science

In computer sciences, there are two important concepts that are closely related to similarity: similarity metrics and semantic similarity (Zadeh, 1971; Tennekes, 1984; Höhle, 1988; Ovchinnikov, 1991; El-Kwae & Kabuka, 1999; Belohlavek, 2000).

Similarity metrics (also called string metrics) are a class of metrics that are used for measuring similarity (closeness) and dissimilarity (distance) between two character strings for approximate matching or comparison in fuzzy string searching. The most commonly used string metric is the Levenshtein Distance, which is also named Edit Distance. The operation principle of the Levenshtein Distance is: compare the two input strings and return a score equivalent to the number of substitutions and deletions needed in order to transform one input string into another. Current research has expanded the metrics such as the Levenshtein distance to cover multiple media including phonetics, tokens, pictures, etc.

Semantic similarity (it is also known as semantic relatedness) is a concept used for assessing the likeness of the meaning/semantic content of a set of documents or terms within term lists by means of defining a metric (Budanitsky and Hirst, 2001). To be more concrete, such a metric can be a kind of topological similarity measured by a distance between words using ontology. Another term named “semantic relatedness” are usually used interchangeably with “semantic similarity”. However, semantic similarity is more specific than semantic relatedness, as the former includes some concepts (e.g. antonymy and meronymy) that have no relations with similarity, while semantic similarity does not. However, much of the literature uses these terms interchangeably, along with terms like “semantic distance”. In computer science, semantic similarity, semantic distance and semantic relatedness all mean "how much does term A have to do with term B?" To answer this question, two types of approaches that calculate topological similarity between ontological concepts have been developed, i.e. edge-based methods and node-based methods. They define a number between -1 and 1, or between 0 and 1, where 1 signifies extremely high similarity/relatedness, and 0 (or -1) signifies little-to-none similarity/relatedness.

Definition in Engineering

In engineering, similitude is a concept used for testing the similarity between two engineering models. An engineering model can be defined as “having similitude” with a real application on condition that they both share geometric similarity, kinematic similarity and dynamic similarity (Hubert 2009). Similarity and similitude are interchangeably used in this context. Similitude has been research in engineering community for decades. Some well-developed models have been used for solving a large number of engineering problems, and they have also been the basis of many textbook formulas. A typical application of similitude in engineering is to predict the performance of a new design by comparing it with an existing, similar design. In this case, the model is the existing design. Another use of similitude and models is in validation of computer simulations with the ultimate goal of eliminating the need for physical models altogether (Heller, 2011).

Main applications of similitude are in hydraulic and aerospace engineering. Here, similitude is used to test and evaluate fluid flow conditions with scaled models. Engineering models are used to study complex fluid dynamics problems where calculations and computer simulations are not reliable (Heller, 2011). Models are usually smaller than the final design, but not always. Scale models allow testing of a design prior to building, and in many cases they are a critical step in the development process.

Definition in Psychology

Similarity in psychology refers to the psychological nearness or proximity of two mental representations. A number of models/approaches for assessing the proximity of two mental representations have been developed in past research. They can be classified into four categories: mental distance approaches, featural approaches, structural approaches and social psychological approaches. Each of them is based on particular set of assumptions.

Mental distance approaches lay their foundation on an assumption that mental representations can be expressed as some concepts in a kind of mental space (Shepard, 1962). Usually, the concepts are represented using points in the space. Then the similarity between two concepts is a function that can be used to calculate the distance between two points (i.e. two concepts) in the space. If the distance between a pair of points is shorter than that of another pair of points, the concepts represented by the former two points are said to be nearer to each other than that of the latter two points.

To overcome the shortcomings in the mental distance approaches, the featural approaches (Tversky, 1977) were proposed. A typical shortcoming in the mental distance approaches is that they assume that spaces are symmetric (because the distance between any two points is the same regardless of which point we start from to calculate the distance). However, psychological similarity is not always symmetric. For example, in many cases, people can only state similarity in one direction. For example, it feels more natural to say “John Smith looks very like his father Alex Smith” than to say “Alex smith looks very like his son John Smith”. The featural approaches assess similarity between two objects by comparing a list of features that describe the properties of the object. The more commonalties they share, the more similar they are.

The basic idea of the transformational approaches (Hahn et al., 2003) developed to evaluate similarity independently of the type of mental representation is as follows: it assumes that any mental representation can be transformed into another one by a number of steps. So it is possible to define some necessary steps to complete this transformation. The more the number of steps in the transformation, the less similar the two representations are. However, Larkey and Markman (2005) found some evidences that are against this idea. Their work has shown that the number of steps to transform the colors and shapes of geometric objects does not predict people's similarity judgments for those objects.

In social psychology, researchers use similarity to describe the closeness or nearness of attitudes, personality, values, interests and culture match between people. It is interesting that research has revealed that interpersonal attraction is from similarity between people, and many forms of similarity have been shown to increase liking. For example, similarities in opinions, interpersonal styles, and amount of communication skill, demographics, and values have all been shown in experiments to increase liking. Several explanations have been offered to explain similarity increases interpersonal attraction. First, people with similar interests tend to put themselves into similar types of settings. For example, two people interested in literature are likely to run into each other in the library and form a relationship. Another explanation is that we notice similar people, expect them to like us, and initiate relationships. Also, having relationships with similar people helps to validate the values held in common. Finally, people tend to make negative assumptions about those who disagree with them on fundamental issues, and hence feel repulsion.

Definition in Music

Similarity does exist in music. For example, a man can easily recognize a familiar song that is being chanted by someone who is singing a little bit out of tune. It is musical similarity that has worked. In his judgment process, he compares the tune of the song which he is familiar with the one that is being sung. There are a number of types of musical similarity that has been research (Toussaint, 2006), such as metrical structure similarity, rhythmic pattern similarity, section structure similarity, modality structure similarity, etc.

Definition in Chemistry

Chemical similarity is an important concept in Chemo-informatics. It plays a significant role in predicting the structures and properties of chemical compounds, designing chemicals that have required structures and properties. Especially, it has been used in drug design studies by retrieving large databases that contain chemicals with anticipated structures or/and structures (Johnson and Maggiora, 1990; Nikolova and Jaworska, 2003). These studies are based on a "similar property principle": similar compounds have similar properties (Nikolova and Jaworska, 2003).

Chemical similarity is often described using a measure called "distance". The larger the distance is, the less similar the two chemicals should be. The distance can be expresses using two kinds of measures: Euclidean and non-Euclidean measures depending on whether the triangle inequality holds.

Definition in Geography

In geography, similarity is of great importance. It is known as spatial similarity relation, a subset of spatial relations which include topological, distance, direction and similarity relations. Similarity is one of the basic research issues in Geo-Sciences (Egenhofer and Mark, 1995a; Goodchild, 2006).

Yan (2010) proposed a definition for spatial similarity relation in light of the Set Theory, regarding it as a one-to-one correspondence of the properties of objects" (Zhou, 1993; Liang, 1999).

> *Suppose that A_1 and A_2 are two objects in the geographic space. Their property sets are C_1 and C_2, and $C_1 \neq \Phi$ and $C_2 \neq \Phi$. If $C_1 \cap C_2 = C_\cap \neq \Phi$, $C_\cap$ is called the spatial similarity relations of object A_1 and object A_2.*

The definition of spatial similarity degree was also discussed by Yan (2010).

> *Spatial similarity degree is a value between [0, 1]. It is use for evaluating the similarity relations of spatial objects.*

Based on the two definitions, Yan (2010) presented three deductions:

> *(1) The larger $C_\cap$, the larger the similarity degree of the two objects.*
>
> *(2) If $C_\cap = \Phi$, the two objects has no similarity property, therefore their spatial similarity degree is 0.*
>
> *(3) If $C_1 = C_2 = C_\cap$, the property sets of the two objects are wholly same, therefore their spatial similarity degree is 1.*

Further, a more general definition of spatial similarity relations for $k(k>2)$ objects in the geographic space was given, and the definition of spatial similarity relations in multi-scale map spaces is proposed.

> *Suppose that A is an object in the geographic space. It is symbolized as A_1、A_2……A_k separately on the maps at scales S_1、S_2……S_k. The property sets of A_i (i=1, 2, …, k) are C_1、C_2……C_k, and $C_i \neq \Phi$ (i=1, 2, …, k). If $C_1 \cap C_2 \ldots\ldots \cap C_k = C_\cap \neq \Phi$, $C_\cap$ is called the spatial similarity relations of the multiple representations of object A in multi-scale map spaces.*

The above definitions for similarity in geographic space lay the foundation on the Set Theory. It assumes that the similarity between objects can be assessed by a number of properties of the objects. The sum of the similarity degrees of the properties is the similarity degree between objects. The more common properties two objects possess, the more similar they are.

These definitions are still at conceptual level. The methods for quantitatively calculating similarity degrees are not mentioned yet.

2.1.2 Critical Analysis of the Definitions

An insight into the existing definitions of similarity in different fields may reveal many problems, and therefore present some interesting research topics.

- Each of the existing definitions of similarity is closely tied to a class of particular applications, or a form of knowledge representation, or assumes a particular domain model. Hence, they cannot be used interchangeably.
- It is obvious that all of the existing definitions of similarity have their underlying assumptions; however, they are not often given explicitly. Without knowing those assumptions, it is impossible to make theoretical arguments for or against any particular measures (Lin, 1998).
- All of the definitions are based on experiences. The comparisons and evaluations of the existing similarity measures are also based on empirical results.

To overcome the above shortcomings in the existing definitions, a number of rules listed in the following should be obeyed in our future research on the definitions of spatial similarity relations in multi-scale map spaces.

(1) The definition should have theoretical justifications. Definition of similarity should lay its foundation on mathematics and cognitive psychology. A mathematics-based definition can facilitate the quantitative representations and measurements of spatial similarity relations; while taking cognitive psychology into consideration can ensure that the results from quantitative measures of similarity are coincident with humans' intuition. In short, the definition of spatial similarity relations must be both mathematically correct and cognitively reasonable.

(2) The definition should be a universal and formal one in geographic space. Here, "universal" means the definition of spatial similarity relations should be applicable to different domains of geography where different similarity measures have previously been proposed, and also be applicable to the domains where no similarity measure has previously been proposed. To be concrete, the definition for spatial similarity relations should be applicable to geometric attributes and thematic attributes of spatial objects in 2-dimensional and 3-dimensional spaces. The definition should also be applicable to all four types of spatial data (i.e. nominal, ordinal, interval, and ratio data). "Formal" means the definition is not from personal experiences but based on the survey and tests of a number of people.

(3) The underlying assumptions of the definition of spatial similarity relations should be presented clearly. If possible, the assumptions should be mathematically expressed to facilitate the quantitative expressions of similarity measures.

Although a Set Theory-based definition of spatial similarity has been proposed (Yan 2010), it is conceptual. Its cognitive justifications, mathematical correctness, and universality in applications need to be verified.

2.2 Features of Similarity

2.2.1 Features of Similarity in Different Fields

Just like “different fields give different definitions of similarity”, different fields give different features of similarity.

In computer sciences, Cilibrasi and Vitanyi (2006) presented the features of similarity applicable for processing text strings. Let Ω be a nonempty set and R^+ be the set of non-negative real numbers. A distance function for describing the dissimilarity between two text strings is $D : \Omega \times \Omega \rightarrow R^+$. Based on this function, three features of similarity relations between text strings can be obtained.

(1) Equality: $D(x, y) = 0$, $iff \quad x = y$;

(2) Symmetry: $D(x, y) = D(y, x)$; and

(3) Triangle inequality: $D(x, y) \leq D(x, z) + D(z, y)$.

The value $D(x, y)$ is called the distance between $x, y \in \Omega$.

In psychology, the following four features of similarity have been discussed.

(1) Symmetry: It is based on two assumptions. The first one is that the similarity from A to B equals to the similarity from B to A; the second one is that judgments of similarity and difference are complementary (the more similarity, the less difference, and vice versa). Mathematically, it is $D(A, B) = D(B, A)$.

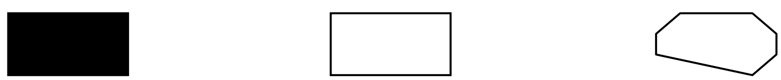
(2) Asymmetry and directionality: The contrast model proposed by Tversky (1977) has proved that “feature commonalities tend to increase perceived similarity more than feature differences can diminish it”. In addition, the structure alignment model has shown that similarity judgments focus on matching relations between items, while difference judgments focus on the mismatching attributes (Medin et al., 1990; Goldstone et al., 1991; Markman, 1996). Therefore, when A is more

similar to T than B is, it is still possible that A is also more different from T than B is.

(3) Minimality: $D(A,B) \geq D(A,A)$ (Tversky, 1977). This should be obvious, because it is impossible that the dissimilarity between identical objects is greater than that between different objects.

(4) Triangle inequality: $D(A,B) + D(B,C) \geq D(A,C)$ (Tversky, 1977).

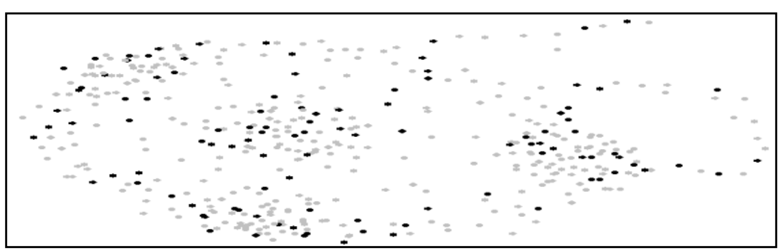
Where, $D(A,B)$ is the distance/dissimilarity function, similar to the one used in above discussion for the features in computer science.



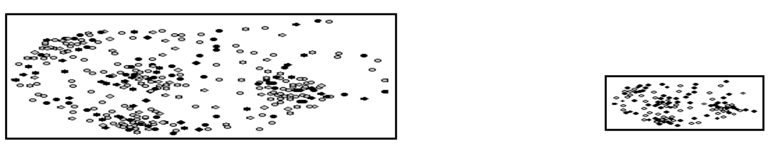
（a）settlement （b）vegetable land （c）vegetable land

Figure 2-4 An example of non-transitivity of spatial similarity relations.

In geography, Yan (2010) discussed the features of similarity relations applicable for objects in multi-scale map spaces.



(a) at scale 1:10K



(b) at scale 1:20K (c) at scale 1:50K

Figure 2-5 Similarity of point clusters at different scales.

(Source: Yan and Weibel, 2008)

(1) Reflexivity: any object has similarity relations with itself.

(2)Symmetry: if object A has similarity relations with object B, object B have the same similarity relations with object A.

(3)Non-transitivity: We cannot conclude that object A has similarity relations with object C, even if object A has similarity relations with object B, and object B has similarity relations with object A.

For example, in Figure 2-4, if take {shape, land type} as the properties for detecting similarity relations, the property set of (a), (b), and (c) are C_a={ rectangle, settlement}, C_b={rectangle, vegetable land}, and C_c={irregular polygon, vegetable land }. $C_a \cap C_b$ ={rectangle} denotes that the objects in (a) and (b) have similarity relations; $C_b \cap C_c$={vegetable land} denotes that the objects in (b) and (c) have similarity relations; but the conclusion that the objects in (a) and (c) have similarity relations cannot be made, for $C_a \cap C_c = \Phi$.

(4) Self-similarity on maps at multiple scales: Geographic objects can be symbolized using different patterns and symbols on maps at different scales. The objects on maps at different scales have spatial similarity relations.

(5) Scale-dependence of self-similarity degree at multi-scales: The spatial similarity degrees of objects on maps at different scales depend on scale change. The greater the scale span from the original map to a generalized map is, the less the similarity degree between two maps should be (Figure 2-5).

2.2.2 Critical Analysis of the Features

The following points can be gained by a comparison and analysis of the existing achievements in the features of similarity in the many fields.

First, the features of similarity in different fields are not always the same. Some applicable in one field may become inapplicable in the other field.

Second, mathematical expressions of the features of similarity in psychology and computer science have been developed, which is in favor of quantitative measurements of similarity.

Third, features in geography are qualitatively described, lacking both mathematical reasoning and psychological experiments to demonstrate their correctness and reasonability.

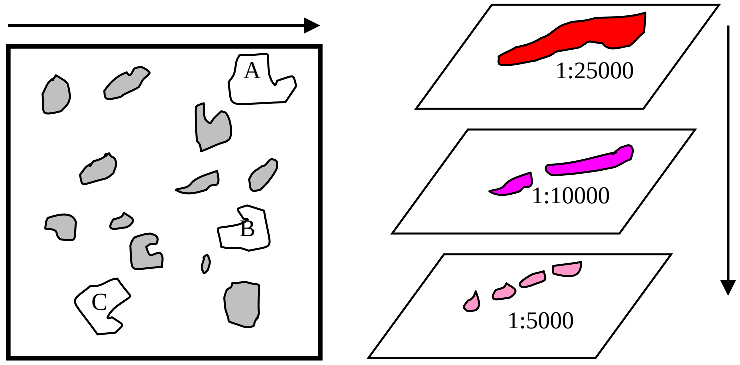
Last, some features (e.g. asymmetry) appearing in other fields have not been research in geography yet.

Hence, the following three issues are worthy of further investigation:

(1) to “borrow” features from other fields and test their applicability in geographic space;

(2) to give mathematic expressions of the features in geographic space; and

(3) to find psychological proofs to support the features in geographic space.



(a) horizontal similarity relations (b) perpendicular similarity relations

Figure 2-6 A scale-based classification system for spatial similarity relations.

(Source: Yan, 2010)

2.3 Classification for Spatial Similarity Relations

Generally, two rules must be obeyed in all classifications, i.e. completeness and exclusiveness. Completeness means the union of all subsets of the sub-categories equals to the whole set; while exclusiveness means the intersection of every two subsets is empty. To meet the demands of the two rules, appropriate criteria must be specified for the purpose of classification. Different criteria generate different categories from same things.

Based on the principles of "completeness" and "exclusiveness", Yan (2010) classified spatial similarity relations by the scales of objects (whether the objects are at same scale or different scales) on maps. If objects are at same scale, their similarity relations are called horizontal similarity relations; whereas if objects are at different scales, their similarity relations are called perpendicular similarity relations (Figure 2-6). Further, Yan (2010) researched on the perpendicular similarity relations, taking geometric attributes and thematic attributes of objects as the classification criterion, and proposed a detailed classification for it (Figure 2-7). However, the classification of horizontal similarity relations has not been touched yet.

2.4 Calculation Models/Measures for Similarity Degree

Calculation models/measures for spatial similarity relations is a very new issue in the community of geographic information science (Nedas and Egenhofer, 2003), and few models/measures can be found in literature, except some borrowed from psychology and computer sciences. Indeed, quantitative description of spatial similarity is difficult to achieve. Guo (1997) ascribed this to two reasons. First, it is difficult to describe and express spatial similarity relations in mathematical languages. In other words, spatial similarity relation is less calculable than other spatial relations (e.g. distance, topological and direction relations). Second, spatial similarity relation is usually used to reveal complex and deeply-covered relations among spatial objects; therefore, it is not easy to find the principles and rules of spatial similarity relations. Li and Fonseca (2006) addressed that "spatial similarity is hard to address because of the numerous constraints of spatial properties and of the complexity of spatial relations". Since it is believed that spatial relations, mainly topology, direction, and distance, capture the essence of a scene's structure (Bruns and Egenhofer, 1996), most researchers focus on the similarity assessment of spatial relations.

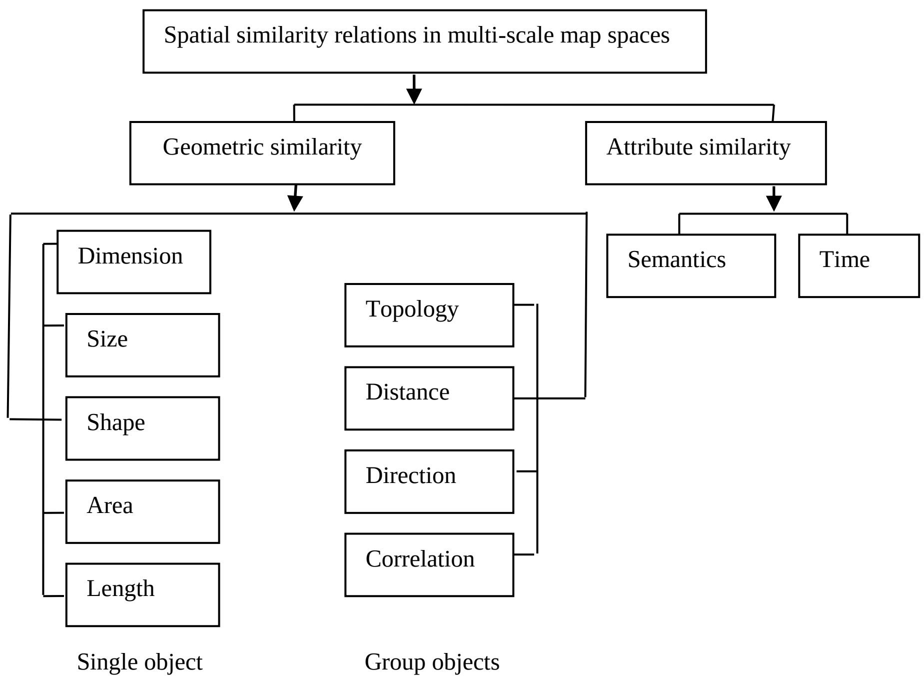


Figure 2-7 A classification system for perpendicular similarity relations.

(Source: Yan, 2010)

The models/measures for similarity in psychology and computer science are also presented and critically discussed in the following paragraphs along with that in geography, because they are the bases of both existing models and our future models for spatial similarity in geography. Then the potential work related to our research objectives will be proposed.

2.4.1 Models in Psychology

In the field of psychology, four similarity models are broadly accepted and used. They are the geometric model, the feature-based contrast model, the structure alignment model, and the transformation model. The four models are laid on the foundation which deems that similarity and difference are tightly related concepts. Definition of difference is usually coincident with distance between the representing points of two entities in a conceptual space. So the distance can be used as a measure of dissimilarity between the entities.

The geometric model is the dominant model in theoretical similarity analysis (Torgerson, 1965; Tversky, 1977; Goldstone, 2004; Li and Fonseca, 2006). The entities in this model are regarded as points in an arbitrarily dimensional space and the dissimilarity/difference of the entities is represented by the distance between the two corresponding points in that space (Tversky, 1977; Nedas and Egenhofer, 2003; Goldstone, 2004). Hence; it seems natural that the geometric model should obey the features of similarity in psychology including minimality, symmetry, and triangle inequality (Tversky, 1977; Thomas and Mareschal, 1997; Goldstone, 2004). However, Tversky's (1977) work has revealed that it is not the case with psychological notions of similarity; because humans' similarity judgments violate the above three features. Minimality is not obeyed since not all identical objects are equally similar. A simple example is that two complex objects that are identical (e.g. two trees) have more similarity than simpler identical objects (e.g. two leaves from the trees). Symmetry is violated because similarities in the metric space are the same no matter what the order of the comparison is, whereas similarities are believed to be asymmetric and directional. For example, a small model car is more similar to a car than a car is to a small model car since many features of the small model car come from cars. Triangle inequality is violated in some cases. For example, a lamp and a moon share an identical feature as both provide light; a moon and a ball share an identical feature as both are round; however a lamp and a ball share no feature in common (Tversky and Gati, 1982).

The feature-based contrast model lays its foundation on Set Theory. It assumes that objects are represented as collections of features, and similarities among objects are expressed as a feature-matching process among common and distinctive features (Tversky, 1977; Goldstone, 2004). Similarities of an object pair increase with its commonalities and decreases with its differences. The similarity of object A to object B is expressed as a linear function of the common and distinctive features. Here, the common features of an object pair are those elements in the intersection of the feature sets; the distinctive features of an object pair are those elements outside of the intersection of the feature sets. In this model, the similarity of an object pair increases with the size of the common features set and decreases with the size of the distinctive features set (Markman, 1993). Tversky (1977) claims that feature commonalities tend to increase perceived similarity more than feature differences can diminish it. In other words, commonalities get higher weights than differences do in the model.

The structure alignment model indicates that similarities come not only from the matching of common and different features, but also from the alignment of features (Markman, 1993). Medin et al (1993) proposed that structure and global consistency are more important in the process of similarity determination than simple local matches. It has been widely recognized that similarity comparisons involve structural alignment instead of simple feature matches (Markman, 1993; Medin et al., 1993). Usually, in the comparison of an object pair, the parts of one object must be aligned or placed in correspondence with the parts of the other object (Goldstone, 1994). In this model, outputs of a similarity comparison process include commonalities, aligned differences, and non-aligned differences (Medin et al., 1993).

The transformation model is one of the geometric models that measures similarity by means of transformational distance (Imai, 1977; Goldstone, 2004). The concept of transformational distance is defined as a function of the complexity that calculates the steps needed in the process of transforming the representation of one entity into the representation of another. The more steps are taken, the more dissimilar the two entities are. The transformation model is especially useful for visual configurations (Nedas and Egenhofer, 2003).

2.4.2 Models/measures in Computer Science

Similarity-based models/measures are mainly used in three areas in computer science, i.e. text processing, image recognition and graphics measurements. For text processing, various approaches and measures for similarity calculation among characters for the purpose of character recognition (Amin and Wilson, 1993; Natori and Nishimura, 1994) and words' semantic comparison (Guan, 2002) in the field of natural language processing have been researched for decades; for image recognition, content-based query in image databases is another hot issue closely related to similarity calculation. After a swift glance them, more attention here will be paid to the geometric similarity of graphics (e.g. shape, structure, distribution, configuration of graphics), because it is more closely related to geometric similarity of spatial objects which is useful in our research.

Vector graphics in a 2-dimensional space can be classified into three categories, i.e. points, lines/curves, and polygons. No method for similarity measurements between two vector point clusters has been found in literatures, except those Hausdorf distance-based ones for computing similarity between two point sets in two images (Huttenlocher, 1993). So the following paragraphs will discuss the measures/models for similarity measurements between curves/lines and between polygons, but ignore that between points.

● Measures/models/approaches/algorithms for similarity between two polygons

(1) Visibility-based approach (Avis and Elgindy, 1983)

A polygon is abstracted by means of its visibility graph, and two polygons are deemed similar whenever their graphs are cyclically isomorphic. This approach can deal with convex and concave polygons; however, it does not take complex polygons (e.g. a polygon with holes) into consideration.

(2) Polygon similarity estimation model (Cakmakov et al., 1992)

The model calculates the gravity centers of the two polygons; then it matches the vertices of the two polygons by sequential rotation and scaling. The similarity of the two polygons is computed using a deliberately defined function. This model is oriented to concave and convex simple polygons, and considers basic transformations such as translation, rotation and scaling of polygons. It also can be used for comparing two polygons with different vertices, though the results are usually unsatisfactory. Nevertheless, complex polygons are out the scope of this model.

(3) Turning function-based metric

A simple polygon is usually represented by describing its boundary using a circular list of vertices, expressing each vertex as a coordinate pair. For example, the visibility-based approach (Avis and Elgindy, 1983) and the polygon similarity estimation model (Cakmakov et al., 1992) use this kind of representation. An alternative representation of the boundary of a simple polygon is to give its turning function, i.e. expressing a polygon using its sides and turning angles. Arkin et al. (1991) proposed a turning function-based metrics. The basic idea of the metric is: the turning functions of the two polygons are constructed first; the distance (i.e. dissimilarity) between the two turning functions is calculated for substituting the dissimilarity between the shapes of the two polygons. This metric is only applicable to simple polygons.

In sum, it is clear that existing models/measures only consider the geometric aspects of simple polygons in similarity calculations. However, complex polygons, discrete polygonal groups and polygon coverages need to be considered; meanwhile, both geometric and attribute aspects of polygons should be taken into account in spatial similarity in multi-scale map spaces.

● Measures for similarity between curvers/lines

Similarity of curves plays important roles in a variety of different domains, such as analysis of stock market trends, protein shape matching, speech recognition, computer vision, etc. Here the curves are usually assumed to be represented as polygonal chains in the plane. The measures that have been used to assess their dissimilarity/similarity include the Hausdorff distance (Alt et al., 1995), the turning curve distance (Cohen et al., 1997), and the Frechet distance. Among them, the Frechet distance has received much attention as a measure of curve similarity (Alt et al., 2001). It belongs to a general class of distance measures that are sometimes called “dog-man” distances (Buchin et al., 2006), an imitation of a man and a dog walking along two curves from one endpoint to the other endpoint, on condition that the man holds an elastic leash at hand and neither of them can teleport (i.e. jump from one point to the next). The distance between the two curves is defined as a function of the leash length, typically minimized over all legal motions. The Frechet distance is the minimum (over all trajectories) of the maximum leash length needed for a fixed trajectory.

2.4.3 Models/measures in Music

On the one hand, qualitative similarity of melodies is popularly used. For example, when someone says “the two melodies are absolutely similar”, he is using an unconscious short-hand but neglects (or is unable) to identify the specific qualitative dimensions according to which the melodies are "close." Qualitatively speaking, two melodies may have similar pitch contours, similar structural tones, similar rhythms, similar harmony; they may evoke a similar mood, and/or express similar themes such as unrequited love, shame, or happiness, and/or be especially quiet, and/or simply have a similar duration. Both melodies may be strophic in form, or both may address a similar audience (e.g., children).

On the other hand, people sometimes attempt to use the qualitative properties by which two things may be deemed similar to characterize their quantitative similarity or degree of closeness. In some cases, a quantitative scale already exists making it possible to characterize directly the quantitative similarity for a given qualitative property. However, in some other cases, no quantitative scale exists as yardsticks. For quantitative data, a number of numerical and statistical methods have been devised as measures of similarity. For example, Pearson's coefficient of correlation provides a useful way of measuring the similarity of the rise and fall of two sets of numerical values. To determine whether the annual pattern of precipitation in Montréal is more similar to that of Melbourne, or of Miami, the monthly precipitation data are aligned and Pearson's coefficient of correlation can be calculated, and then we would find that Montréal correlates most strongly with Miami.

In measuring the similarity between two melodies, it is not easy to determine if there is some other (qualitative) dimension by which the two melodies exhibit a greater (quantitative) similarity. In some analytic tasks people may be most interested in determining which elements of a given set are most similar according to a pre-established qualitative dimension. In other tasks people may be interested in determining which qualitative dimension reveals the greatest similarity between two melodies. Not all data is quantitative in nature, so it is not always possible to apply parametric measures of similarity such as Pearson's correlation. Although many musical parameters may be represented quantitatively, it is not always possible to cast musical elements according to some quantitative yardstick. Often the information is in the form of discrete categories that cannot be ordered. In the case of non-quantitative data, an alternative way of calculating the degree of similarity between two melodies is to ask: how much "tinkering" is required in order to reach identity?

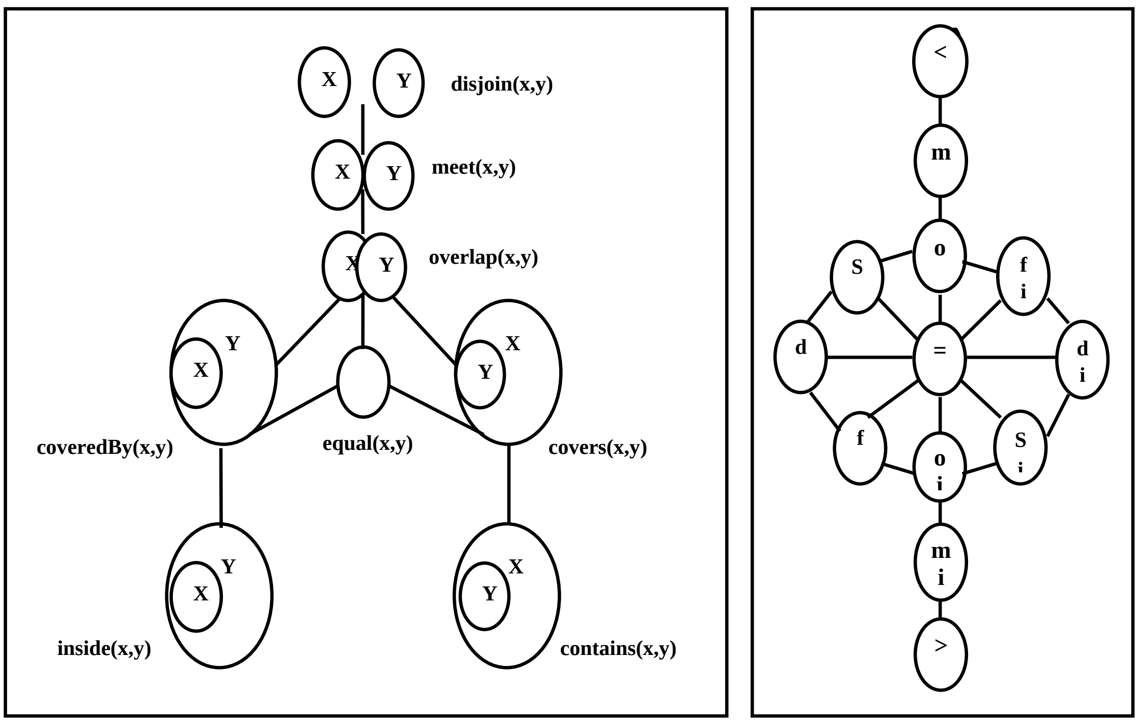
One of the most prevalent and intuitively appealing approaches to measuring quantitative similarity is to calculate the edit distance between two strings (Damerau, 1964; Levenshtein, 1966; Ullman, 1977). Briefly, the edit distance between two strings can be defined as the minimum number of basic modifications (insertions, deletions and substitutions) that must be performed on one string (source string) in order to make it identical to a second (target) string. Performing an insertion means augmenting the source string by adding a symbol, whereas a deletion means removing a symbol from the source string. A substitution is the replacement of a single symbol in the source string by another symbol, which could be the same, or different. If a replacement symbol differs from the symbol it replaces, the substitution is called a dissimilar substitution. For each type of edit operation we may define a numerical penalty representing the magnitude of the modification. For example, the operations of insertion and deletion might be defined as adding a nominal value of +1 to the edit distance. A substitution might be defined as adding a value of +1 if is dissimilar, and zero if it is not. A dissimilar substitution is logically equivalent to a deletion followed by an insertion, so if we assigned an edit-distance penalty of +2 rather than +1, then the substitution operation would be redundant.

2.4.4 Models/measures in Geography

Spatial similarity measurement is different from document/texts similarity assessment in which the focus is on matching keywords, because spatial similarity relations involve various elements, such as spatial relationships, spatial distribution, geometric attributes, thematic attributes, and semantic relationships. In addition, different applications may have different requirements and priorities on similarity elements, which make calculation/assessment of spatial similarity relations complicated and difficult. In sum, it is difficult for researchers to quantify spatial similarity relations due to at least the following two major reasons.

First, spatial similarity measurement is a cognitive process that is consistent with human's cognition; nevertheless, psychologists have not clearly known what has happened while people are judging spatial similarity relations.

Second, spatial relations, i.e. topological, direction and distance relations, capture the essence of a scene's structure (Bruns and Egenhofer, 1996) and play key roles in spatial similarity assessment; however, complexity of spatial relations and numerous constraints of spatial properties make spatial similarity relations hard to be addressed.



(a) (b)

Figure 2-8 Conceptual neighborhood of topological relations.

(a) Egenhofer's method, and (b) Freksa's method. (Revised from: Li and Fonseca, 2006)

Although it is not easy to calculated spatial similarity relations, many researchers have studied this issue and some achievements have been made. Many models/approaches/measures for similarity calculation/assessment are discussed in detail in the following sections, for the purpose of laying a good theoretical and methodological foundation for our new quantitative methods.

● Conceptual neighbourhood approach

The conceptual neighbourhood approach is the same as the transformation model in basic ideas, i.e. similarity in this model is measured according to the distance between two concepts in a network. It computes the shortest path between two nodes in the network. The distance is calculated as the number of edges between them (Rada et al., 1989; Budanitsky, 1999). The fewer edges between them on the network, the more similarities they share (Quillian, 1968).

A method based on the 9-intersection model (Egenhofer and Franzosa, 1991) is proposed by Egenhofer and Al-Taha (1992) to derived gradual changes of topological relationships. The principle of creating a conceptual neighbourhood of topological relationships by this method is illustrated in Figure 2-8(a). If changes in topological relations (e.g. scale, translation, and/or rotation) happen, the corresponding process can be described as a sequence of movements over the neighbourhood network. For example, if the distance from disjoin (x, y) to meet (x, y) is set as 1, the distance from disjoin (x, y) to covers (x, y) should be 3.

Figure 2-8(b) shows another method proposed by Freksa (1992), which creates the conceptual neighbourhood network based on Allen's 1-D interval relations (Allen, 1983). Papadias and Dellis (1997) extended this model into a higher dimensional space to address spatial relationship similarity on topology, direction and metric distance. Chang and Lee (1991) derived the conceptual neighbourhood network of 169 possible spatial relations between rectangles also from applying Allen's 1-D interval relations to orthogonal projections. Bruns and Egenhofer (1996) captured spatial relationship similarity over Chang and Lee's graph by combining the distance conceptual neighbourhood model. They describe the similarity measuring process as "one scene is transformed into another through a sequence of gradual changes of spatial relations. The number of changes required yields a measure that is compared against others, or against a pre-existing scale. Two scenes that require a large number of changes are less similar than scenes that require fewer changes."

● Projection-based approach

The projection-based model divides the two dimensional space with a horizontal line and a vertical line, taking a point as the reference (Frank, 1996; Ligozat, 1998). The four rays of the two lines represent the 4 cardinal directions: north, west, south, and east (Figure 2-9). The regions between these two lines represent the secondary directions, i.e. northwest, southwest, southeast, and northeast. It was argued that the projection model has advantages over the cone model (Frank, 1991) in implementation due to the rectangular nature of the directional partition (Goyal, 2000).

The projection-based approach projects spatial objects and their relations onto another space, which can be a vector space or a matrix space. By this way, the problem of similarity assessment is shifted from the comparison of objects in spatial scenes to that vector or matrix space. The famous 2D String symbolic representation is an example of projection-based approach (Chang et al., 1987), in which spatial objects and their relationships are represented by 2D strings along x and y axes. The similarity assessment between two scenes is then treated as it was a string matching. Chang defines three types of similarity criteria, type-0, type-1 and type-2. Type-0 is the most generous one. It is fulfilled when two objects have the same relationship on either the x- or the y-axis. Type-1 requires that two objects have the same relations on both the x- and y-axis. Type-2 requires not only two objects to have the same relations but also that they have the same rank of the relative positions.

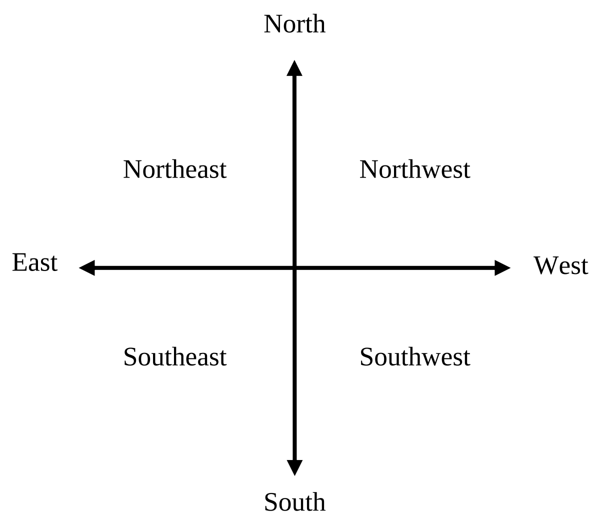


Figure 2-9 Directional space partition in the project-based approach.

●Combination of the conceptual neighbourhood approach and the projection-based approach

To measure distance similarity degrees, Goyal and Egenhofer (2001) proposed a method that combines the conceptual neighbourhood approach and the projection-based approach. In this hybrid method the directional space is projected into a 3×3 matrix, which represents the nine directions (north, northwest, west, southwest, south, southeast, east, northeast, and same). Each sector of the matrix specifies how much of a target object falls into the direction it represents. The similarity of a cardinal direction is determined by the least cost of transforming one direction-relation matrix into another one.

● Spatial relations-oriented model (the TDD model)

The TDD (Topology-Direction-Distance) model (Li and Fonseca 2006) provides a similarity measure that integrates four widely accepted conceptual similarity models (i.e. the geometric model, the feature contrast model, the transformation model, and the structure alignment model). The basic idea of the TDD model is: commonalities (C) and differences (D) between spatial scenes are measured; the final similarity measurement (S) is a combination of both, i.e. $S = C - D$. The structure alignment model considers that the parts of one object must be aligned or placed in correspondence with the parts of the other in the comparison of a stimulus pair. Therefore, the output of the similarity comparison process includes commonalities, alignable differences, and non-alignable differences. The TDD model treats alignable differences and non-alignable differences separately: D = (alignable difference + non alignable difference).

Table 2-1 Basic elements in the spatial measurement process.

(Revised from Li and Fonseca, 2006)

Level of comparison	**Types of similarity measured**		
Scene	Relationships	Spatial	Topological
			Direction
			Metric distance
			Distribution
		Non-spatial	Attributes
Object	Attributes	Geometric	Types of objects
		Thematic	Attribute comparison

The TDD model takes into account both relational similarity and attributes similarity (Table 2-1), and different weights are applied on relational similarity and attributes similarity, because they have different impacts on commonality judgment and difference judgment (Tversky, 1977) in similarity evaluation of a certain task context. In addition, the TDD model applies the order of priority (i.e. topology→direction→distance) into spatial similarity assessment and the relaxation of the transformation cost. Both features are implemented through the weight setting. The TDD model measures the similarity between spatial scenes (a spatial scene is comprised of spatial objects). A spatial scene in TDD model may include only one spatial object, or two spatial objects, or three or more spatial objects.

The TDD model is based on findings of psychological similarity research which stated that (1) the commonalities between a stimulus pair increase the similarity more than differences decrease it; (2) aligned differences affect the similarity more than non-aligned differences do; (3) the order of priority topology-direction-distance reflects the priorities of different types of spatial relationship in spatial similarity assessment; and (4) the difference between inter-group transformation cost and intra-group transformation cost which is consistent with the theory of categorization. Instead of measuring the distance between objects in traditional models, this model adopts Tversky's feature contrast model, which considers both commonality and difference in similarity assessment. It groups the topological relationships and introduces the concepts of inter- and intra-group transformation costs. The inter-group transformation cost has a higher value than the intra-group transformation cost.

● Spatial semantic-oriented models/measures

Although the World Wide Web (WWW) currently provides good access to data through a variety of search engines as long as the user knows the keywords that the data providers used, it falls short as a reliable access mechanism to information when purely syntactic comparisons cannot resolve ambiguities or fail to build connections to related or similar items that a data provider did not foresee. The Semantic Web (Berner-Lees et al., 2001) aims to overcome the limitations of WWW by incorporating explicitly modeled expressions of semantics into the search process. The provision of such explicit semantics may be seen as a much richer metadata model, with the goal to offer machine-readable and machine-executable metadata. The domain of geospatial information is particularly rich in this respect due to the varieties in human spatial languages for expressing and communicating spatial information. Naturally, a spatial similarity-based concept named "Semantic Geospatial Web" (SGW) appeared in recent years (Egenhofer, 2002; Fonseca and Sheth, 2003). SGW is envisioned as a new information retrieval environment that will facilitate meaningful access to geospatial information (Nedas and Egenhofer, 2003; Rodriguez and Egenhofer, 2004).

A set of methods developed by Nedas and Egenhofer (2003) for the retrieval of similar spatial information in spatial databases use Boolean operators, such as "not", "and", "or", to combine and integrate several similarity constraints. The methods take into account a 3-tuple {geometric attribute; thematic attribute, ID} in spatial similarity. Geometric attributes are associated with an object's topology and metric details, while thematic attributes capture spatial but non-geometric information. Because of this duality, their methods assess similarity among spatial objects at two procedures: geometric attribute assessment and thematic attribute assessment. The overall similarity value of two objects is a combination of their geometric and thematic similarity values. To combine the similarity values, the weighted mean values are used instead of two popular approaches: the geometric approach and the fuzzy-logic approach. This research in spatial similarity is from a conceptual rather than implementation point of view.

To determine semantic similarity among spatial entity classes, the Matching-Distance Similarity Measure (MDSM) was proposed (Rodriguez and Egenhofer, 2004), taking into account the distinguishing features of the classes (parts, functions, and attributes) and their semantic interrelations (is–a and part–whole relations). A matching process is combined with a semantic-distance calculation to obtain asymmetric values of similarity that depend on the degree of generalization of entity classes. MDSM's matching process is also driven by contextual considerations, where the context determines the relative importance of distinguishing features.

2.4.5 Critical Analyses of Existing Models/Measures

A number of insights can be gained from the analysis of existing models/measures for similarity assessments in psychology, computer science and geography.

(1) Similarity relation roots itself in humans' cognition; hence, the four Models for similarity calculations in psychology (the geometric model, the feature-based contrast model, the structure alignment model, and the transformation model) have been the bases of the existing models for similarity in geography and will still be a most important source of the models for spatial similarity in multi-scale map spaces in this study.

(2) Constructing a spatial similarity model needs to consider spatial aspects (including spatial relations, spatial distribution, spatial structure etc.) and attribute aspects (including geometric and thematic attributes, e.g. names, areas, length etc. of the objects) of spatial objects. Existing models put emphases on the attribute aspects and give little attention on spatial aspects (the TDD model considers topology, direction and distance, but it is not for multi-scale geographic spaces).

(3) Shape similarity between polygons and between curves/lines has been a hot issue in computer science for decades; however, few achievements have been made in comparing two polygons/curves with different vertices at different scales.

(4) Existing models consider similarity between only two single objects; while the spatial similarity relations between two groups of objects and between two maps have not been explored.

(5) “Scaling” has usually been taken as a parameter in existing models/measures for similarity calculation, where, “scaling” means simple enlargement and shrinkage of objects. This is wholly different from the concepts of “scaling” in map generalization that means simplification of objects due to map scale change.

2.5 Raster-based Approaches for Map Similarity Comparison

Besides vector-based models and measures for similarity calculation discussed in the previous sections of this chapter, many raster-based approaches have been proposed for map comparison (Berry, 1993; Hagen-Zanker, Straatman and Uljee, 2005; Hagen-Zanker and Lajoie, 2008; Hagen-Zanker, 2009). In addition, a raster-based software package has been developed to compute similarity degrees between raster maps or images (Visser and de Nijs, 2006). The following gives a brief summary of these approaches.

The raster-based approaches can be classified into two categories: one for comparing categorical maps (from Section 2.5.1 to Section 2.5.7) and the other for comparing numerical maps (in Section 2.5.8).

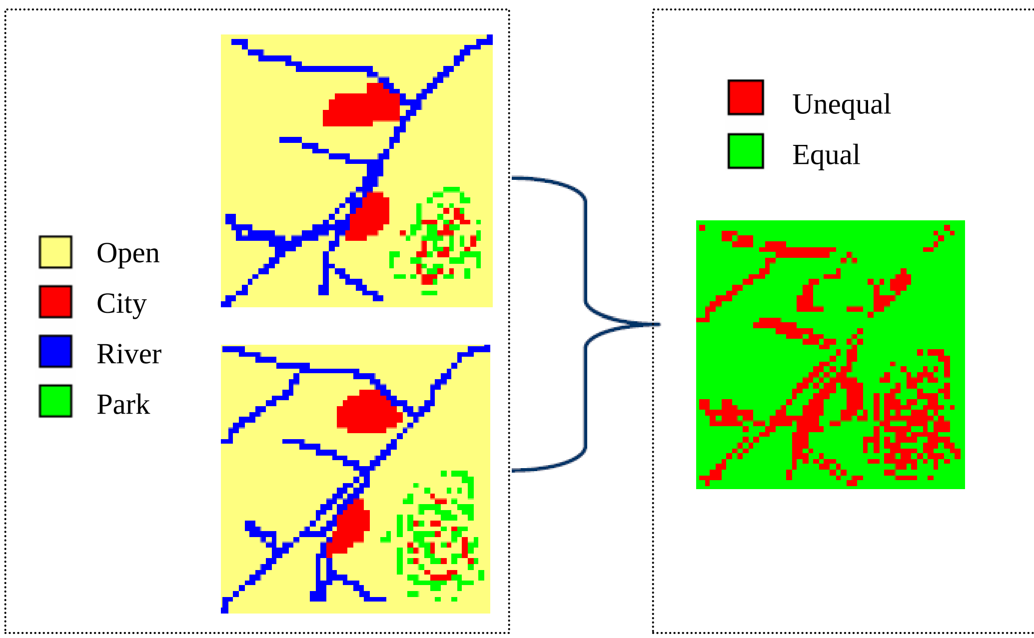


Figure 2-10 Raster-based similarity computation.

2.5.1 Per Category Comparison Method

The per category comparison method (Congalton, Oderwald and Mead, 1983; Maselli, Rudolph and Conese, 1996) performs a cell-by-cell comparison with respect to one category on the maps. It simultaneously gives the information about the occurrence of the selected category in both maps (Figure 2-10). This traditional technique is suspect because of possible map registration and error propagation problems. These boolean similarity operations often cannot adequately account for the uncertainty and complexity inherent in spatial information.

2.5.2 Kappa Comparison Method

The Kappa comparison method (Hagen, 2002) is based on a straightforward cell-by-cell map comparison, which considers for each pair of cells on the two maps whether they are equal or not. This results in a comparison map displaying the spatial distribution of agreement. This comparison method does not require any parameters.

Usually, over time only a small percentage of the land use area actually changes, while most locations keep the same. For those simulations with little change, the agreement will be high regardless of the quality of the model. In this case, Kappa simulation (Van Vliet, Bregt and Hagen-Zanker, 2011) corrects the agreement between two maps for the sizes of class transitions. By taking class transitions as the reference, rather than class sizes that Kappa comparison method uses, the absolute value of kappa can be interpreted.

2.5.3 Fuzzy Kappa Approach

Fuzzy Kappa approach to assessing similarity of categorical maps (Hagen, 2003; Hagen-Zanker et al., 2005) applies fuzzy set theory and involves both fuzziness of location and fuzziness of category to compare raster maps of categorical data. It obtains a spatial and gradual analysis of the similarity of two maps. The results from the comparison are basically in accordance with those of a visual inspection, because it distinguishes minor deviations and fluctuations within similar areas from major deviations. The main purpose of the Fuzzy Kappa map comparison is to take into account that there are grades of similarity between pairs of cells in two maps. Like its crisp counterpart, the fuzzy kappa is based on a cell-by-cell map comparison.

2.5.4 Fuzzy Inference System

The traditional a cell-by-cell map comparison may register a disagreement between cells even if this is due to a minor displacement between similar cells in the respective maps and the overall spatial patterns are essentially the same. To solve this problem, the fuzzy inference system comparison algorithm (Power et al., 2001) compares the characteristics of polygons rather than cells found in both maps. The calculation of the similarity is based upon a fuzzy inference system evaluation of these characteristics. The characteristics that are taken into account in this evaluation are area of intersection, area of disagreement and size of polygon. It has been shown that a fuzzy local polygon-by-polygon land use comparison is less affected by possible map registration problems because the fuzzy inference system indirectly fuzzifies the boundaries of the polygons. The local matching results from the fuzzy inference system for the project datasets demonstrate the advantage of the fuzzy approach over the Boolean comparison methods.

The fuzzy inference system approach is in essence asymmetrical, which means that the comparison of two maps is different depending on which map is considered to be the reference (or real) map and which is the comparison (or modeled) map.

2.5.5 Fuzzy Comparison with Unequal Resolutions

The Map Comparison Kit 3 (RISK, 2013) allows comparing maps of unequal resolution that cover the same area. The comparison takes place at the coarsest resolution of the two maps. Internally the comparison method transforms the crisp fine scaled map to a soft classified coarse one on the basis of percentages. The percentages are interpreted as degrees of similarity in a fuzzy set map comparison.

There are two options for evaluating similarity, either absolute or relative to the maximum attainable similarity.

2.5.6 Aggregated Cells

It is well established that the outcome of spatial analysis generally depends on the scale that it is conducted. The method of aggregated cells (Pontius jr, 2000; Pontius jr et al., 2004) aims to calculate scale-dependant similarities. Scale in this case is operationalized as aggregation level; the only parameter to this method is the aggregation factor, which must be a positive integer (natural) value. The method aggregates the original pixels taken in by categories to coarser maps where every cell is represented by a vector containing for each category the fraction of cover.

2.5.7 Moving Window-based Structure

The moving window based structure comparison method (Hagen-Zanker, 2006) compares maps on the basis of their local structure. Two types of structure are considered in the comparison; patch based structure and proportion based structure. These are sometimes also discerned as configuration and composition based structure. In this case, that denomination would be incorrect since the moving window in effect makes both approaches configuration based.

2.5.8 Numerical Comparison Methods

Six different cell-by-cell numerical comparison algorithms (McGarigal et al., 2002) are listed in Table 2-2. Accordingly, fuzzy numerical methods have been studied (McGarigal et al., 2002), considering fuzziness of location in the same manner that the fuzzy Kappa comparison does. The difference is that it applies to numerical maps, which means that the use of a categorical similarity matrix is not necessary (or possible).

Table 2-2 Six cell-by-cell numerical comparison algorithms

Operations	Explanations
b – a	difference
abs (b – a)	absolute difference
(b - a) / max (abs (b - a))	scaled difference
abs (b – a) /max (abs (b - a))	scaled absolute difference
b / a	relative difference
abs (b / a)	absolute relative difference

Note: the meaning of the logical operation can be found in Appendix A.

2.6 Chapter Summary

In order to lay a good foundation for constructing new models for calculating spatial similarity relations that can be used in automated map generalization, this chapter reviews, summarizes and analyzes the existing achievements in spatial similarity relations, including the definitions, features, classification systems, and calculation models/measures of similarity relations in various circles. Most importantly, this chapter summarizes the advantages and disadvantages of the existing achievements, and clearly shows the gap between the research objectives of this study and the existing achievements in this area.

Chapter 3 Concepts of Spatial Similarity Relations in Multi-scale Map Spaces[1]

This chapter explores the fundamental theories of spatial similarity relations in multi-scale map spaces, and aims at the four sub-objectives addressed in Chapter 1: (1) definitions of spatial similarity relations; (2) features of spatial similarity relations; (3) factors that affect humans' judgments of spatial similarity relations; and (4) a classification system for spatial similarity relations in multi-scale map spaces.

3.1 Definitions

Chapter 2 reviews the definitions of similarity in various fields, including geometry, computer science, engineering, psychology, music, chemistry, and geography. An insight into these definitions has gained that existing definitions are closely application-oriented, and based on corresponding assumptions, and lay their foundations on experiences. In other words, the existing definitions have their limitations, and cannot be used interchangeably. Hence, it is necessary to define spatial similarity relation in multi-scale map spaces by its own way in order to investigate this issue thoroughly.

Some rules need to be obeyed in defining spatial similarity relations in multi-scale map spaces in order to avoid the shortcomings existing in the definitions of similarity in other fields and to make the new definitions work well in automated map generalization. These rules require that the new definitions should be (1) expressed in mathematical language, (2) aligned with human's spatial cognition, and (3) formal, but not only based on personal experiences. In addition, the assumptions of the new definitions should be clearly presented in mathematical languages.

---

1 Partial of this Chapter has been published by: Yan H., 2010, Fundamental theories of spatial similarity relations in multi-scale map spaces, Chinese Geographical Science, 2010, 20(1): 18-22; partial has been submitted by: Yan H. & Li J., 2013, Features of spatial similarity relations, International Journal of Applied Earth Observation and Geoinformation, submitted on Dec. 11, 2013 (manuscript N0. JAG-D-13-00504); and partial has been submitted by Yan H. & Li J., 2013, Quantitative definition of spatial similarity relations in multi-scale map spaces, Earth Science Informatics, submitted on Dec. 9, 2013 (manuscript N0. ESIN-D-13-00126).

The following proposes the definition of spatial similarity relation in multi-scale map spaces. Before this, the definitions of similarity relation and spatial similarity relation need to be presented.

3.1.1 Definitions of Similarity Relation

Similarity relation can be defined descriptively and quantitatively.

Similarity relation has been descriptively defined over and again by many researchers in various research fields (Gower, 1971; Ramer, 1972; Lanczos, 1988; Hershberger & Snoeyink, 1992; Zhou, 1993), and its definitions also appear in huge dictionaries. To sum up, similarity relation can be simply described as:

a quality that makes one person or thing like another.

It covers two aspects:

1. quality or state of being similar: resemblance; and

2. comparable aspect: correspondence.

This definition presents a universal, qualitative description of similarity relations. Although it is useful for people to understand "similarity relation" intuitively, it cannot provide direct help to construct quantitative models for calculating similarity relations, because it lacks of a mathematical foundation.

Similarity relation is calculable; therefore it has been defined in mathematical language (Coxeter, 1961; Cederberg, 1989). In a general metric space (X, d) similarity relation can be expressed using a function f from the space X into itself that multiplies all distances by the same positive scalar r. To be exact, for any two points x and y, the following function can be true.

$$d(\mathrm{f(x)}, \mathrm{f(y)}) = \mathrm{r} \times d(\mathrm{x}, \mathrm{y}) \qquad 3\text{-}1$$

where, $d(\mathrm{x}, \mathrm{y})$ is the distance from x to y.

Weaker versions of similarity would for instance have f be a bi-Lipschitz function and the scalar r a limit:

$$\lim \frac{d(\mathrm{f(x)}, \mathrm{f(y)})}{d(\mathrm{x}, \mathrm{y})} = r \qquad 3\text{-}2$$

This weaker version applies when the metric is an effective resistance on a topologically self-similar set.

A self-similar subset of a metric space (X, d) is a set K for which there exists a finite set of similitudes $\{\mathrm{f}_s\}_{s\in S}$ with contraction factors $0 \le r_s < 1$ such that K is the unique compact subset of X (Martin, 1982) for which

$$\underset{s\in S}{U} f_s(\mathrm{K}) = \mathrm{K} \qquad 3\text{-}3$$

These self-similar sets have a self-similar measure μ^D with dimension D given by the formula

$$\sum_{s\in S} (\mathrm{r}_s)^D = 1 \qquad 3\text{-}4$$

which is often (but not always) equal to the set's Hausdorff dimension and packing dimension. If the overlaps between the $f_s(\mathrm{K})$ are "small", the following simple formula can be used for the measure of similarity relations:

$$\mu^D(f_{s1} \circ f_{s2} \circ \cdots \circ f_{sn}(K)) = (\mathrm{r}_{s1} \cdot \mathrm{r}_{s2} \cdots \mathrm{r}_{sn})^D \qquad 3\text{-}5$$

3.1.2 Definitions of Spatial Similarity Relation

Spatial similarity relation refers to the similarity relation in the geographic space (including map spaces). It comprises the similarity relations between individual objects and the similarity relations between object groups in the geographic space. For example, in Figure 3-1, people may be interested in either if Island A_1 is similar to Island A_2 or how similar Archipelago 1 and Archipelago 2 are.

Similarity refers to “comparable aspects”. To be exact, every object has a number of aspects. When people discuss the similarity relations between objects (or object groups), they usually compare the corresponding aspects of the two objects (or object groups) subconsciously in the process of similarity relation judgments.

In essence, similarity between two objects (or object groups) means one-to-one corresponding comparison of the properties of objects (Zhou, 1993; Liang, 1999). In light of the existing achievements (Li, 2000; Yan, 2010), the definition of spatial similarity relations may be developed based on Yan’s work (Yan, 2010) by means of the Set Theory. Because properties of the objects (object groups) generally weigh differently in human's similarity judgments, which should be taken into account in defining spatial similarity relations.

Definition

Suppose that A_1 *and* A_2 *are two objects in the geographic space. Their property sets are* P_1 *and* P_2, *respectively, and each of which has* n ($n > 0$) *elements* $P = \{p_1, p_2, ..., p_n\}$ *in it.* $P_1 = \{p_{11}, p_{12}, ..., p_{1n}\}$, *and* $P_2 = \{p_{21}, p_{22}, ..., p_{2n}\}$, *and their corresponding weights are* $W = \{w_1, w_2, ..., w_n\}$.

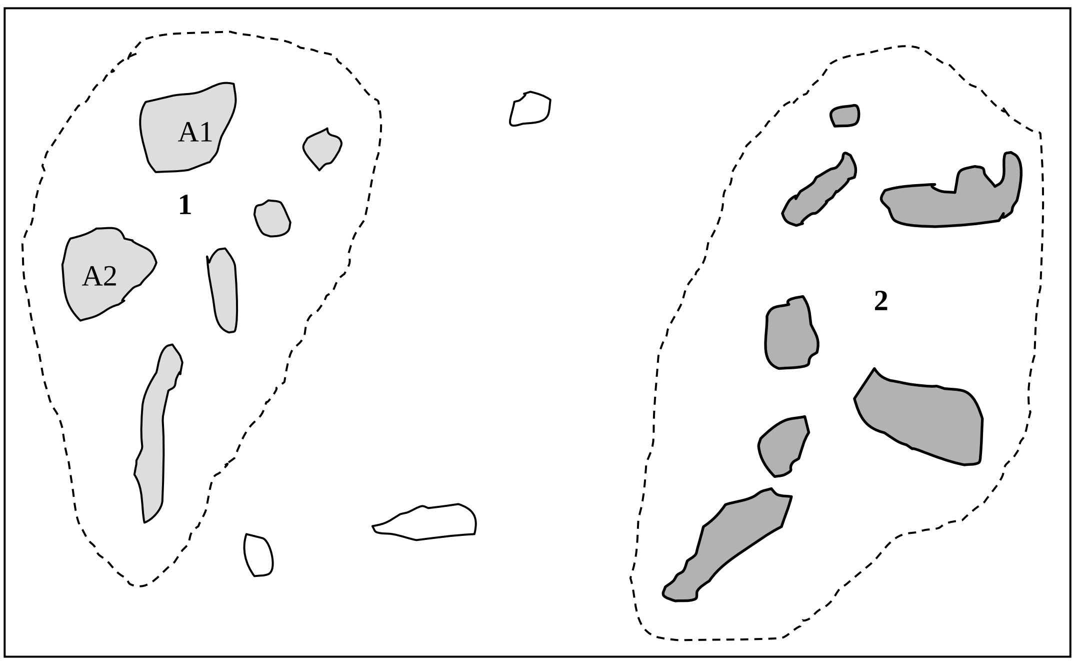


Figure 3-1 Spatial similarity relations on an island map.

Similarity relations between individual objects (Island A and Island B) or object groups (Archipelago 1 and Archipelago 2).

Let $Sim^{P_i}_{A_1,A_2} = f_i(p_{1i}, p_{2i})$. 3-6

$Sim^{P_i}_{A_1,A_2}$ *is called the spatial similarity relations of object* A_1 *and object* A_2 *at property* p_i. $i = 1, 2, ..., n$. *It is also named the spatial similarity degree between* A_1 *and* A_2 *at property* p_i, *and its value belongs to [0,1].*

Let $Sim(\mathrm{A}_1, \mathrm{A}_2) = \sum_{i=1}^{n} w_i Sim^{P_i}_{A_1,A_2}$. 3-7

$Sim(\mathrm{A}_1, \mathrm{A}_2)$ *is named the spatial similarity relations of object* A_1 *and object* A_2. $i = 1, 2, ..., n$. *It is also named the spatial similarity degree between* A_1 *and* A_2, *and its value is [0,1].*

Demonstration of the Definition

In order to explain the above definitions, the similarity relations between island A_1 and island A_2 in Figure 3-1 are taken as an example. The properties of island A_1 and island A_2 are $\mathrm{P} = \{\mathrm{Area, shape, arability}\}$, and the corresponding weights of the properties are $\mathrm{w} = \{0.3, 0.6, 0.1\}$ (these values are usually collected from experts and/or specific group of people by means of questionnaire surveys).

Here, the "area" of an island may be "large", "big", and "small", denoted by 3, 2, and 1, respectively; the "shape" of the island can be described using the number of edges of the polygon; and the "arability" may be "yes" or "no", denoted by 2 and 1. The property set of the two islands are $P_1 = \{2, 6, 1\}$, and $P_2 = \{2, 9, 1\}$, respectively.

The similarity relations of the two islands at the three properties are calculated and presented as follows. Here, f_1, f_2, and f_3 are three experience formulae by the author (they may be changed if necessary).

$$Sim^{P_1}_{A_1,A_2} = f_1(2, 2) = 1$$

$$Sim^{P_2}_{A_1,A_2} = f_2(6,9) = \frac{\vee(\mathrm{p}_{12}, \mathrm{p}_{22})}{(\mathrm{p}_{12}, \mathrm{p}_{22})/2} = \frac{\vee(6,9)}{(6+9)/2} = 0.8$$

$$Sim^{P_3}_{A_1,A_2} = f_3(1,1) = 1$$

Then the spatial similarity relations between A_1 and A_2 can be obtained.

$$Sim(\mathrm{A}_1, \mathrm{A}_2) = \sum_{i=1}^{3} w_i Sim^{P_i}_{A_1,A_2} = 1\times 0.3 + 0.8\times 0.6 + 1\times 0.1 = 0.88$$

Discussion

A couple of remarks can be made after a detailed analysis to the definition of spatial similarity relations.

First, this definition obviously lays its foundation on mathematics, and gives a quantitative expression of spatial similarity relations.

Second, objects in the geographic space have a number of different properties; but people are usually uncertain or ambiguous when they talk about similarity between two objects. In other words, people do not clearly know exactly what properties of the objects should be compared in their similarity assessments. Hence, work needs to be done to “extract” these properties from people’s brains.

Third, the weights of the properties in the definition are subjective values which depend on human’s experiences and knowledge. The more people are surveyed, the more accurate the weights are.

Last, the formulae for calculating spatial similarity relations should be formal so that the results are acceptable and reliable. Hence, experiments should be designed to test the reliability and the validity of the formulae.

3.1.3 Definitions of Spatial Similarity Relation in Multi-scale Map Spaces

Spatial similarity relations may exist either between objects on maps at same scale (e.g. A_1 and A_2 in Figure 3-1) or between objects at multiple different scales. As far as the latter is concerned, automated map generalization is an ideal source for obtaining such examples (e.g. Figure 3-2 and

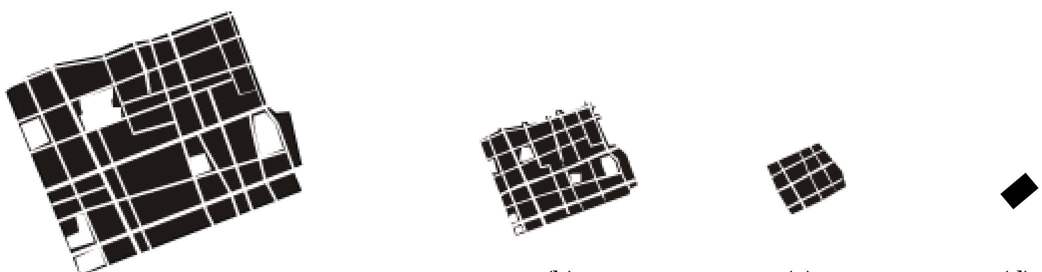
Figure 3-3). The spatial similarity relations between objects on maps at multiple different scales are named spatial similarity relations in multi-scale map spaces.

Although spatial similarity relation in multi-scale map spaces belongs to spatial similarity relations, it has a couple of characteristics that the other ones do not have.

First, the objects compared in spatial similarity relations in multi-scale map spaces are the same object in the geographic space. What are compared are actually the symbols of the objects on maps at different scales.

Second, although similarity relations in multi-scale map spaces refer to the similarity of the symbols of the same objects at different scales, it is different from the so-called self-similarity (Mandelbrot, 1967). Thus, the theory of self-similarity cannot be directly used to solve the problems in spatial similarity relations in multi-scale map spaces.

Third, properties of objects in multi-scale map spaces include attribute properties and spatial properties but no temporal properties, because all objects are the same one at different scales.



(a) (b) (c) (d)

Figure 3-2 Similarity relations of settlements at four different scales.

(a) scale S_1; (b) scale S_2; (c) scale S_3; and (d) scale S_4.

Definition

Suppose that A is an object in the geographic space. It is symbolized as A_1, A_2 ,……, A_n ($n>0$) separately on the maps at scales S_1, S_2 ,……, S_n. The property sets of A_1, A_2 ,……, A_n are P_1, P_2 ,……, P_n. If each property set has k ($k>0$) elements, and their corresponding weights are $W=\{w_1, w_2, ..., w_k\}$. The property sets are expressed as follows:

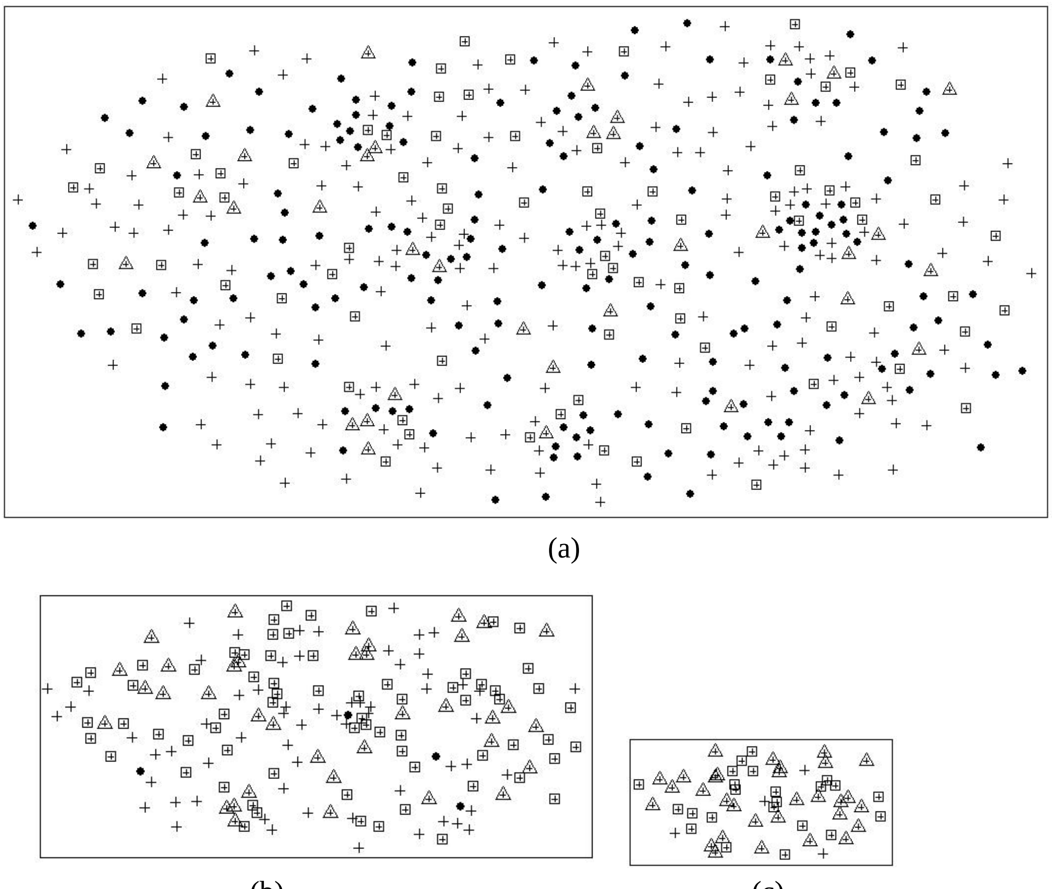
$$P_1 = \{p_{11}, p_{12}, ..., p_{1k}\};$$

$$P_2 = \{p_{21}, p_{22}, ..., p_{2k}\};$$

$$\ldots\ldots$$

$$P_n = \{p_{n1}, p_{n2}, ..., p_{nk}\} \qquad 3\text{-}8$$

$$Let\, Sim^{P_j}_{A_l, A_m} = f_i(p_{lj}, p_{mj}). \qquad 3\text{-}9$$



(a)

(b) (c)

Figure 3-3 Similarity relations of control points at three different scales.

(a) scale s_1; (b) scale s_2; and (c) scale s_3.

$Sim^{P_j}_{A_l, A_m}$ *is called the spatial similarity relations of object* A *at scale* l *and scale* m *regarding the* j^{th} *property. Here,* $i > 0$; $j > 0$; $l > 0$; $m > 0$. $Sim^{P_j}_{A_l, A_m}$ *is also named the* *spatial similarity degree of object A at scale l and scale m regarding the j^{th} property, and its value belongs to [0,1].*

$$Let\ Sim(A_l, A_m) = \sum_{i=1}^{k} w_i Sim_{A_l, A_m}^{P_i} . \qquad 3\text{-}10$$

$Sim(A_l, A_m)$ is named the spatial similarity relations of object A at scale l and scale m. Here, $l > 0$; $m > 0$. It is also named the spatial similarity degree of object A at scale l and scale m, and its value belongs to [0,1].

Discussion

The above presents two definitions regarding spatial similarity relations in multi-scale map spaces. Spatial similarity relations defined by them are one-to-one relations. To be exact, $Sim_{A_l, A_m}^{P_j}$ is the similarity relations of an object at two map scales regarding one property; and $Sim(A_l, A_m)$ is the similarity relations of an object at scale l and scale m.

In addition, the following points need to be noticed regarding the two definitions.

First, the two definitions give quantitative expressions of spatial similarity relations.

Second, selection of the properties used in spatial similarity relations is a subjective process. It closely related to people's nationalities, culture, age, gender, etc..

Third, the weight values of the properties should be obtained by psychological experiments, taking sufficient number of people as subjects and selecting sufficient number of appropriate objects as samples use in the experiments.

Last, validity of the definitions depends on users' judgments.

3.1.4 Definition of Difference

Difference is interchangeably used with similarity. Hence, it is defined here to facilitate our discussion. Suppose that A_1 and A_2 are two objects in the geographic space, difference can be expressed as:

$$Dif(A_1, A_2) = 1 - Sim(A_1, A_2) \qquad 3\text{-}11$$

3.2 Features

Previous work has revealed that similarity has a number of features in various fields (Table 3-1 lists the features that have been discussed in computer science, psychology and geography). The following will summarize and analyze these features, and prove whether they are applicable in the geographic space.

3.2.1 Equality

Equality of spatial similarity relations can be described as:

$$\forall(\mathrm{A}),\ Sim(\mathrm{A},\mathrm{A}) = 1 \qquad 3\text{-}12$$

This seems self-evident that every object in the geographic space is totally similar to itself.

Table 3-1 Features of similarity in various fields.

Fields / features	Computer Science	Psychology	Geography
Equality	√		
Symmetry	√	√	√
Asymmetry		√	
Triangle inequality	√	√	
Minimality		√	
Reflexivity			√
Non-transitivity			√
Scale-dependence			√
Self-similarity			√

Note: √ means the feature is applicable in the corresponding field.

3.2.2 Finiteness

Finiteness of spatial similarity relations can be described as:

$$\forall(\mathrm{A},\mathrm{B}),\ Sim(A,B) < \infty \qquad 3\text{-}13$$

The upper value is often set at 1 (creating a possibility for a probabilistic interpretation of the similitude).

3.2.3 Minimality

Minimality of spatial similarity relations can be described as:

$$\forall(\mathrm{A},\mathrm{B}),\ Sim(A,\mathrm{A}) \geq Sim(\mathrm{A},\mathrm{B}) \qquad 3\text{-}14$$

This feature should be obvious, because similarity between identical objects is greater than that between different objects.

3.2.4 Auto-similarity

Auto-similarity of spatial similarity relations can be described as:

$$\forall(A,B),\ Sim(A,B) = Sim(A,A) \Leftrightarrow A = B \qquad 3\text{-}15$$

This is obviously an inference from the previous feature “minimality”.

3.2.5 Symmetry (Reflectivity)

Symmetry (in other words, reflectivity) of spatial similarity relations can be described as:

$$\forall(A,B),\ Sim(A,B) = Sim(B,A) \qquad 3\text{-}16$$

This may be explained as: spatial similarity relations calculated from object A to B should be the same as that from B to A. For example, there are two cities A and B. It is obvious that spatial similarity compared from A to B is equal to that from B to A, no matter what properties of the two cities are compared.

Symmetry in the geographic space is conditional true. This will be discussed in the feature “weak symmetry”.

3.2.6 Non-transitivity

Non-transitivity of spatial similarity relations can be described as:

$$\forall(A,B,C),\ (Sim(A,B) > 0 \wedge Sim(B,C) > 0),\ \exists\ Sim(A,C) = 0. \qquad \text{3-17}$$

This feature means that that object A is similar to object B and object B is similar to object C does not guarantee that object A is similar to object C.

There are numerous examples regarding non-transitivity of spatial similarity relations in the geographic space. The following presents two of them.

◆ Example 1

In Figure 3-4, A is a city; B is a village with buildings and green land; and C is a small forest. Their properties "size" and "land cover" are selected for evaluate their similarity relations.

$W = \{0.5, 0.5\}$;

P_A ={large, built-up area};

P_B ={large, green land};

P_C ={small, green land}.

$\therefore\ Sim(A,B) = 0.5 > 0 \wedge Sim(B,C) = 0.5 > 0$

But $Sim(A,C) = 0$.

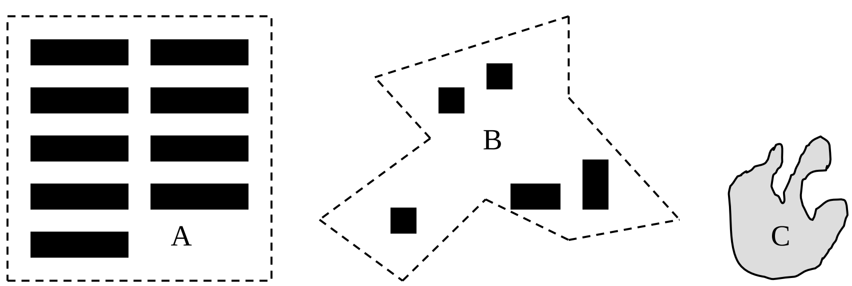


Figure 3-4 Example 1 for non-transitivity in the geographic space.

◆ Example 2

Figure 3-5 shows a map with three linear objects. A is a road; B is a ditch; and C is an administrative boundary. Their properties "origination" and "line type" are selected for evaluate their similarity relations.

$W = \{0.5, 0.5\}$;

P_A ={man-made, straight};

P_B ={man-made, curve};

P_C ={natural, curve}.

$\therefore\ Sim(A,B) = 0.5 > 0 \wedge Sim(B,C) = 0.5 > 0$

But $Sim(A,C) = 0$.

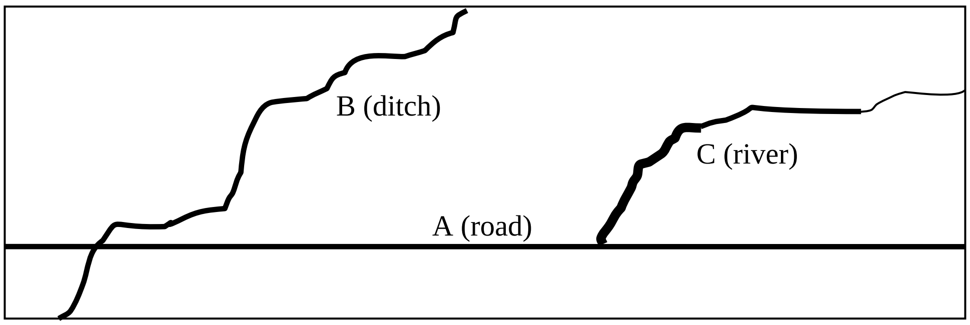


Figure 3-5 Example 2 for non-transitivity in the geographic space.

3.2.7 Weak Symmetry

Weak symmetry of spatial similarity relation refers to such kind of cases: that A is similar to B does not always mean B is similar to A . This may be expressed using a formula:

$$\exists(A,B),\ Sim(A,B) \neq Sim(B,A) \qquad 3\text{-}18$$

For example, in our daily life people are accustomed to say “John is like his father” but seldom say “John’s father is like his son.”

Such examples also exist in the geographic space. For example, in China people usually say “North Korea is similar to China” but do not say “China is similar to North Korea.” This comparison is related to historical and geographic reasons.

3.2.8 Asymmetry

If A is more similar to T than B is, it is still possible that A is also more different from T than B is. This is called Asymmetry of spatial similarity relations and may be expressed as:

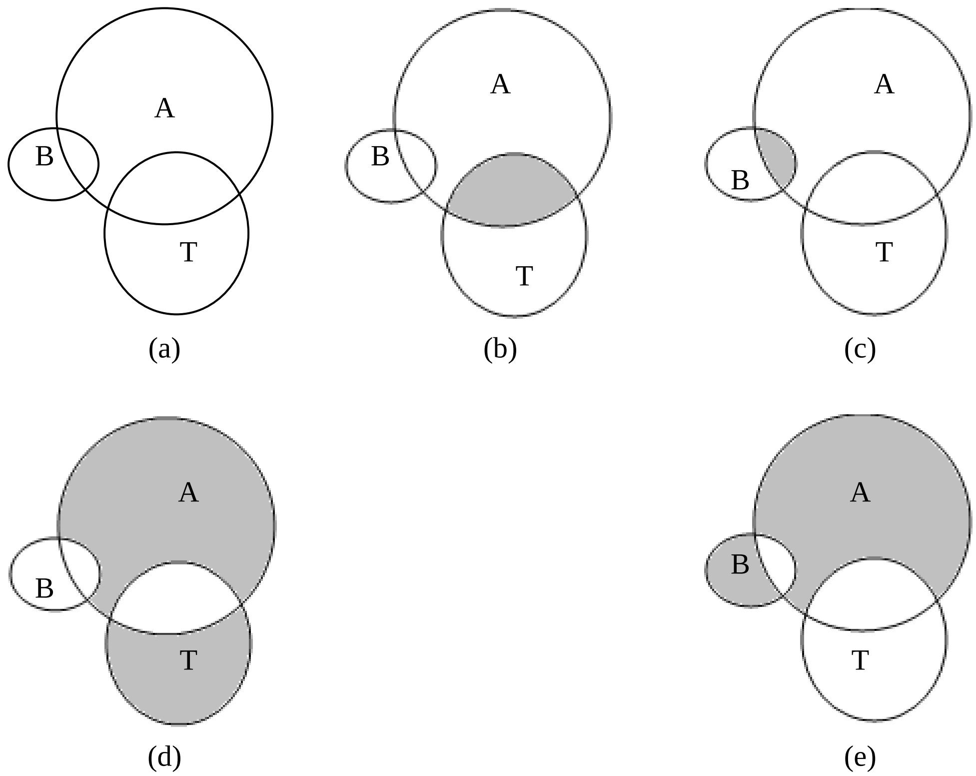


Figure 3-6 Explanation of asymmetry.

(a) Three objects A, B and T; (b) $Sim(A,T)$; (c) $Sim(A,\mathrm{B})$; (d) $Dif(A,\mathrm{T})$; and (e) $Dif(A,\mathrm{B})$. So $Sim(A,\mathrm{T}) \geq Sim(\mathrm{A}, B)$, and $Dif(A,\mathrm{T}) \geq Dif(\mathrm{A}, B)$.

$$\forall(A,B,T),\ \text{if}\ Sim(A,\mathrm{T}) \geq Sim(\mathrm{A},B),\ \exists\ Dif(A,\mathrm{T}) \geq Dif(\mathrm{A},B) \qquad 3\text{-}19$$

An explanation of this feature is shown in Figure 3-6. A is a house; B is a tree; and T is a pavilion. There are totally five elements in the property set: history, origination, owner, size, and environment. Possible values of the properties are:

History: ancient, modern, unknown;

Origination: natural, man-made, unknown;

Owner: public, private, unknown;

Size: large, big, small; and

Environment: excellent, good, bad.

The property set of object A, including all of the five properties, is:

$$P_A = \{\text{ancient}, \text{man}-\text{made}, \text{public}, \text{big}, \text{bad}\}.$$

The property set of object B, if including history, owner, and size, is

$$P_B = \{\text{ancient}, \text{private}, \text{small}\}.$$

The property set of object B, if including origination, size, and environment, is

$$P_B^{'} = \{\text{man}-\text{made}, \text{small}, \text{bad}\}.$$

The property set of object T, if including history, origination, owner, and environment, is

$$P_T = \{\text{ancient}, \text{man}-\text{made}, \text{public}, \text{good}\}$$

The property set of object T, if including owner, size, and environment, is

$$P_T^{'} = \{\text{public}, \text{small}, \text{good}\}$$

When the similarity between A and T is considered, P_T is selected and the weights are

$$W_{P_T} = \{0.25, 0.25, 0.25, 0.25\}.$$

When the difference between A and T is considered, $P_T^{'}$ is selected and the weights are

$$W_{P_T^{'}} = \{0.3, 0.3, 0.4\}.$$

When the similarity between A and B is considered, P_B is used and the weights are

$$W_{P_B} = \{0.3, 0.3, 0.4\}.$$

When the difference between A and B is considered, $P_B^{'}$ is used and the weights are

$$W_{P_B^{'}} = \{0.3, 0.3, 0.4\}.$$

By the above data, the similarity relations can be obtained:

$$Sim(\mathrm{A},\mathrm{T}) = .25\times1 + .25\times1 + .25\times1 = .75$$

$$Sim(A,B) = .3\times1 = .3$$

$$Dif\,(A,T) = 1 - Sim^{P_{T'}}(\mathrm{A},\mathrm{T}) = 1 - .3\times1 = .7$$

$$Dif\,(A,B) = 1 - Sim^{P_{B'}}(\mathrm{A},\mathrm{B}) = 1 - .3\times1 + .4\times1 = .3$$

In conclusion, $Sim(A,\mathrm{T}) \ge Sim(\mathrm{A},B)$, $\exists$ $Dif\,(A,\mathrm{T}) \ge Dif\,(\mathrm{A},B)$

3.2.9 Triangle Inequality

Triangle inequality of spatial similarity relations can be described as:

$$\forall(A,B,C),\ \mathrm{Sim(A,B)} + \mathrm{Sim(B,C)} \ge \mathrm{Sim(A,C)} \qquad \text{3-20}$$

Triangle inequality of similarity in the geographic space refers to such case: the similarity degree between object A and object B plus that of B and C is greater than that of A and C. The following gives an example to explain this feature.

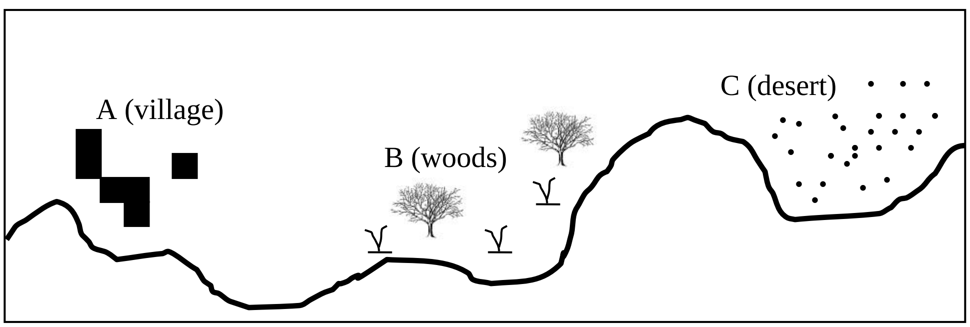


Figure 3-7 An example for triangle inequality in the geographic space.

Suppose that there are three objects alongside of a river bank, they are a village, a patch of woods, and a desert (Figure 3-7). Their property set contains three elements: history, size, and owner. Possible values of these elements are as follows:

History: ancient, modern, current, unknown;

Size: large, small; and

Owner: public, private, unknown.

The property sets of the three objects are:

$P_A = \{\text{modern}, \text{small}, \text{public}\}$;

$P_B = \{\text{current}, \text{small}, \text{private}\}$; and

$P_C = \{\text{ancient}, \text{small}, \text{public}\}$.

Corresponding weights of the properties are:

$W = \{0.3, 0.4, 0.3\}$.

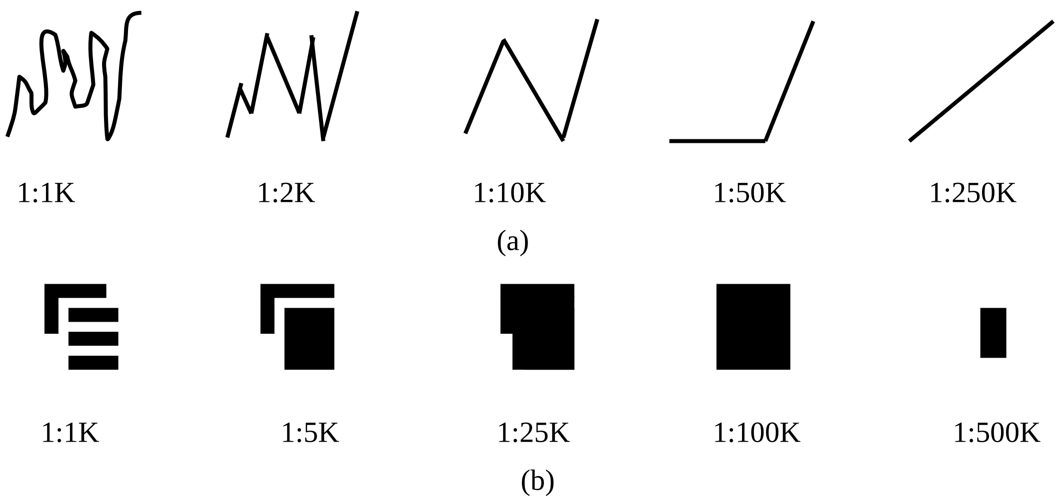


Figure 3-8 Generalization and scale change.

Hence, we have

$Sim(A, B) = 0.4 \times 1 = 0.4$;

$Sim(B, C) = 0.4 \times 1 = 0.4$;

$Sim(A, \mathrm{C}) = 0.4 \times 1 + 0.3 \times 1 = 0.7$;

$\therefore\ (\mathrm{Sim(A, B)} + \mathrm{Sim(B, C)}) = 0.8 \geq \mathrm{Sim(A, C)} = 0.7$

3.2.10 Scale-dependence

Scale dependence in multi-scale map spaces may be explained in this way: if object A at scale S is gradually generalized to objects A_1, A_2,……, A_n ($n > 0$) on maps at scales S_1, S_2,……, S_n, and $S_1 > S_2 > ... > S_n$. If objects A_1, A_2,……, and A_n are compared with A, respectively, taking their spatial properties (shape, the number of edges, etc.) and attributes as the properties, the following function should be correct.

$$Sim(\mathrm{A}, \mathrm{A}_1) > \mathrm{Sim}(\mathrm{A}, \mathrm{A}_2) > ... > \mathrm{Sim}(\mathrm{A}, \mathrm{A}_n) \qquad 3\text{-}21$$

To express this feature in a simple way: the more an object is simplified (generalized), the less similar it is if compared with the original object.

By Formula 3-21, it is easy to deduce Formula 3-22:

$\forall(A), s = f(A, \mathrm{A}_s)$ is a monotonic decreasing function. 3-22

where, A is an object on the map; A_s is the simplified A at scale s.

Figure 3-8 shows two examples to demonstrate this formula.

3.3 Factors in Similarity Judgments

Factors that affect human's similarity judgments play important roles in constructing models for calculating similarity relations as well as designing methods for evaluating the validity of the models. Although progress regarding the factors in similarity judgments has been made in past work (Rodr ǵuez and Egenhofer, 2004; Li and Frederico, 2006), the achievements are not systematic and incapable of supporting our further research. Hence, this section will thoroughly explore the factors in spatial similarity judgments, aiming at answering the following two questions that take core roles in human's similarity cognition.

Question 1: what factors take effect in similarity judgments?

Question 2: do these factors have different effects in the process of human's spatial recognition? And if so, how can the weights of the factors be obtained?

To answer question 1, the factors used in spatial similarity judgments are firstly classified into two categories, i.e. factors for individual objects and factors for object groups, because spatial similarity assessment is usually performed between individual objects or object groups. Here, the meaning of object group is similar but not equal to “scene” (Bruns and Egenhofer; 1996).

To answer question 2, many psychological experiments need to be done using a number pairs of individual objects and object groups; and then the statistical data from the experiments should be analyzed to determine the weights of the factors.

3.3.1 Factors for Individual Objects

Factors for individual objects in spatial similarity judgments refer to attributes of the objects. These attributes are classified as geometric attributes and thematic attributes (Li and Frederico, 2006). Geometric attributes are those attributes that relate to geometric features of the spatial objects, e.g., location, length, area, slope, and shape. Thematic attributes identify or describe the thematic features of spatial objects, such as population, road types, or the time of an event.
Three types of individual objects are considered here. They are individual point objects, individual linear objects and individual areal objects.

Individual Point Objects

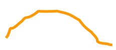
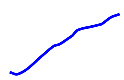
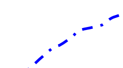
Individual point objects on maps refer to those small but important objects in the geographic space needing to be represented on maps, such as pavilions, isolated houses, pagodas, monuments, signposts alongside roads, oil wells, etc. Their attributes that should be considered in spatial similarity judgments include:

- Location
- Shape
- History
- Owner
- Area, etc.

Individual Linear Objects

Linear symbols are used to represent the geographic objects and the events that are localized on lines (e.g. lines of watershed) and the demarcating lines (e.g. borders of regions, states) and in order to mark objects that have linear character, that are not manifested by its width in a scale (e.g. rivers or roads). Linear symbols may be contours, roads, rivers boundaries, etc. on topographic maps (Table 3-2) as well as power transmission lines, pipelines, land type demarcating lines, etc. on thematic maps. It is impossible and unnecessary to enumerate all kinds of individual linear objects/phenomena. Here, three kinds of important individual linear features on topographic maps are selected as representatives, i.e. rivers, roads, and contour lines. The factors for each of them in spatial similarity judgments are addressed, respectively.

Table 3-2 Examples of individual linear objects on maps

Symbols	Features	Symbols (cont.)	Features (cont.)
	Index contour line		High way
	Intermediate contour line		Secondary high way
	Supplementary contour line		Light duty road
	Depression		Unimproved road
	Levee		Trail
	National boundary		Stream
	Provincial boundary		Intermittent river

◆ River

- Width
- Depth
- Length
- Curvature
- Elevation

- The number of branches
- Navigability
- The number of harbors
- History
- Owner
- Sediment concentration, etc.

◆ Road

- Width
- Length
- Curvature
- The number of crosses
- Construction status: If the road is started, planned, closed for maintenance, or completed.
- Road access: If the road is open to the public or is part of a restricted, private area.
- Priority: The road's priority indicates the type of traffic that the road handles, its physical geometry, and its connectivity. Some roads are bigger, support more traffic, and are more universally recognized than others.
- Type of route: It can range from highways to trails.
- The number of lanes
- Max speed
- Divider: It separates the flow of the traffic and prevents a turn.
- Direction: It means one-way or two-way on the road.
- Elevation
- Surface type
- Road condition
- Popularity: It tells how well-known the road is, e.g. city-wide, country-wide, or world-wide.
- Grade levels: If the road segment is underpass, overpass or on the ground.
- Bicycle and pedestrian access.

◆ Contour line

- Length

- Elevation
- Curvature
- Closed: Whether the contour is a closed curve or not?
- Type: What type is the contour, an index, an intermittent, or a supplementary contour?
- Contour interval
- Location: What does the contour represent, a plateau, a depression, a saddle, a hilltop, a ridge, or a valley?
- Accuracy: This refers to the elevation accuracy of the contour line.
- Scale of the map

Individual Areal Objects

Individual areal objects refer to those topologically separated objects that are represented on maps using polygonal symbols, such as settlements/buildings, water bodies, forests, etc. Table 3-3 shows a number of areal symbols usually used to represent individual areal objects on topographic maps.
The three kinds of individual areal objects, i.e. buildings, lakes, and forest, on topographic maps are selected as representatives, and their factors that affect human's spatial similarity judgments are addressed, respectively.

◆ Building

- Area
- Height
- The number of stories
- Population
- Roof type: Whether the building is waterproof, or sunscreen?
- Construction material: If the building is made from wood, or concrete, etc.?
- Owner
- Price
- Status: Whether the building is in construction, in maintenance, or in use?
- Construction time

◆ Lake

- Location

• Area

• Depth

• perimeter

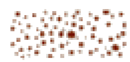
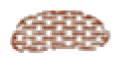
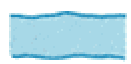
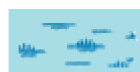
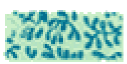
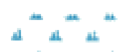
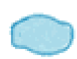
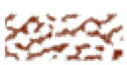
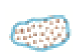
• Status: Whether it is a seasonal or a perennial lake?

• Navigability

• Origination: Whether the lake is formed by remnants of glaciers, blocked rivers, or rivers that fill natural basins?

• Bottom status: Whether the lake is covered by mud?

Table 3-3 Examples of individual areal objects on maps

Symbols	Features	Symbols	Features
	woodland		Gravel beach
	Low brush		Tailings ponds
	Planted vegetation		Perennial river
	Cultivated vines		Swamp
	Dense, tropical trees		Rice field
	Sand		Perennial lake
	Intricate surface		Dry lake

◆ Forest

• Area

• Perimeter

• Species

• History

• Owner

• Price

- Mean height of the trees
- Precipitation
- Temperature

3.3.2 Factors for Object Groups

To judge similarity relations at the level of object groups (or scenes, though slightly different), people usually pay more attention to the relations between the objects in the groups but ignore the geometric attributes and thematic attributes of individual objects (Li and Fonseca, 2006). Generally, three types of spatial relations (i.e. topological relations, direction relations, metric distance relations and distribution relations) and one type of non-spatial relations (i.e. attributes) are taken into account and regarded as the crucial factors that affect human's spatial similarity judgments if two object groups are compared.

Topological Relations

Topological relations often capture the configuration of an object group—topology matters, metric refines (Egenhofer and Mark, 1995b). "Topological relations are attractive in similarity cognition as they are largely immaterial to subtle geometric variations and when they get changed usually significant alterations occur. If several of such changes occur, a chain reaction gets triggered."(Bruns and Egenhofer, 1996). Initially two relations are slightly changed, or still just one. The new scene is still similar. After more and more changes occur, the new scene becomes less and less similar. In this sense, the change is gradual, from equivalent to high similar, then to less and less similar.

The concept of "gradual change" has been used to quantify similarity of topological relations by many researchers in recent years (Egenhofer and Al-Taha, 1992; Egenhofer and Mark, 1995a; Bruns and Egenhofer, 1996; Li and Fonseca, 2006).

Figure 3-9 shows the gradual changes of topological relations, discriminating among pairs of objects.

Nevertheless, Figure 3-9 is not systematic enough to quantitatively express topological relations, and there are some errors in the costs. For example, in Figure 3-9(a), there are three different answers in five cases for the cost direct and indirect from "overlap" to "equal".

(1) Overlap→Equal, the cost is 3;

(2) Overlap→contain→equal, the cost is 5;

(3) Overlap→contain & Meet→equal, the cost is 5;

(4) Overlap→contain & Meet→Contain→equal, the cost is 6; and

(5) Overlap→ Contain→contain & Meet→ equal, the cost is 6.

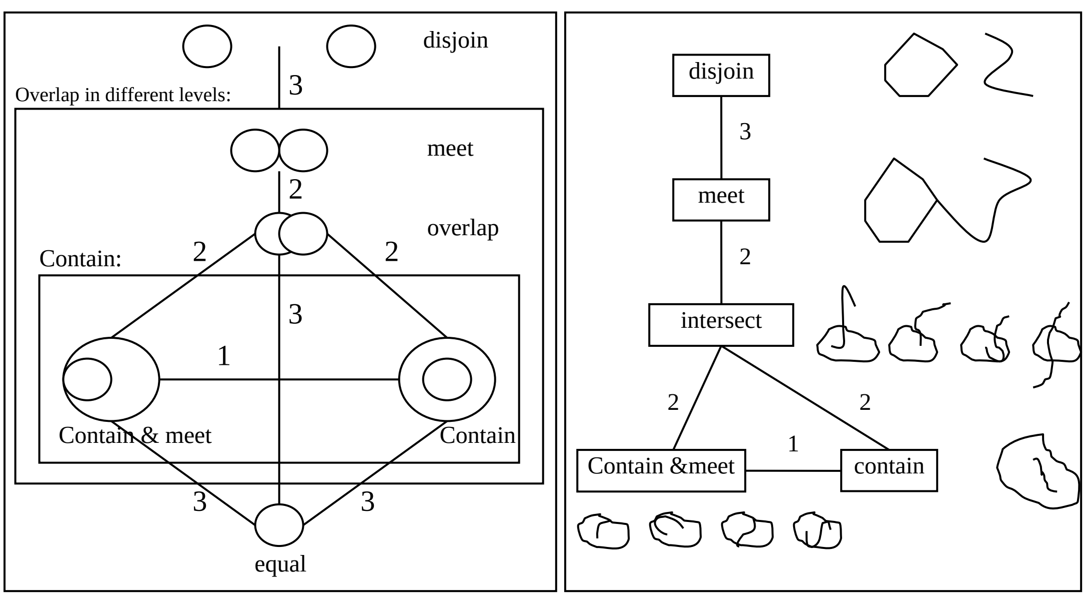
The three answers are ambiguous. On contrary, they should be intuitively equal in human's cognition. To correct this error, some improvements have been made, and a refined and systematic version of the transformation costs is proposed here (Figure 3-10). The main idea of the improvements is as follows:

(1) Transformations between "disjoint" and "meet" and between "intersect or overlap" and "equal" are viewed as major changes; thus, the cost on each of their edges is 4. While the other changes are minor changes and each of their costs should be less than 4.

(2) The cost of a direct transformation between any two topological relations should be equal to that of an indirect transformation. In other words, the sum of the costs between specified two topological relations should be identical no matter which route is selected.

To ensure the nationality of gradual changes of topological relations, the improvements inherit the basic principles of gradual changes of topological relations proposed and tested by Bruns and Egenhofer (1996); on the other hand, the improvements make the costs between any two relations are equal. This is obviously coincident with human's spatial cognition. For example, in Figure 3-10, the transformation cost from "overlap" direct or indirect to "equal" is always equal 4.

This improved theory of "gradual changes of topological relations" will be used in defining topological similarity relations between object groups in Chapter 4.



(a) (b)

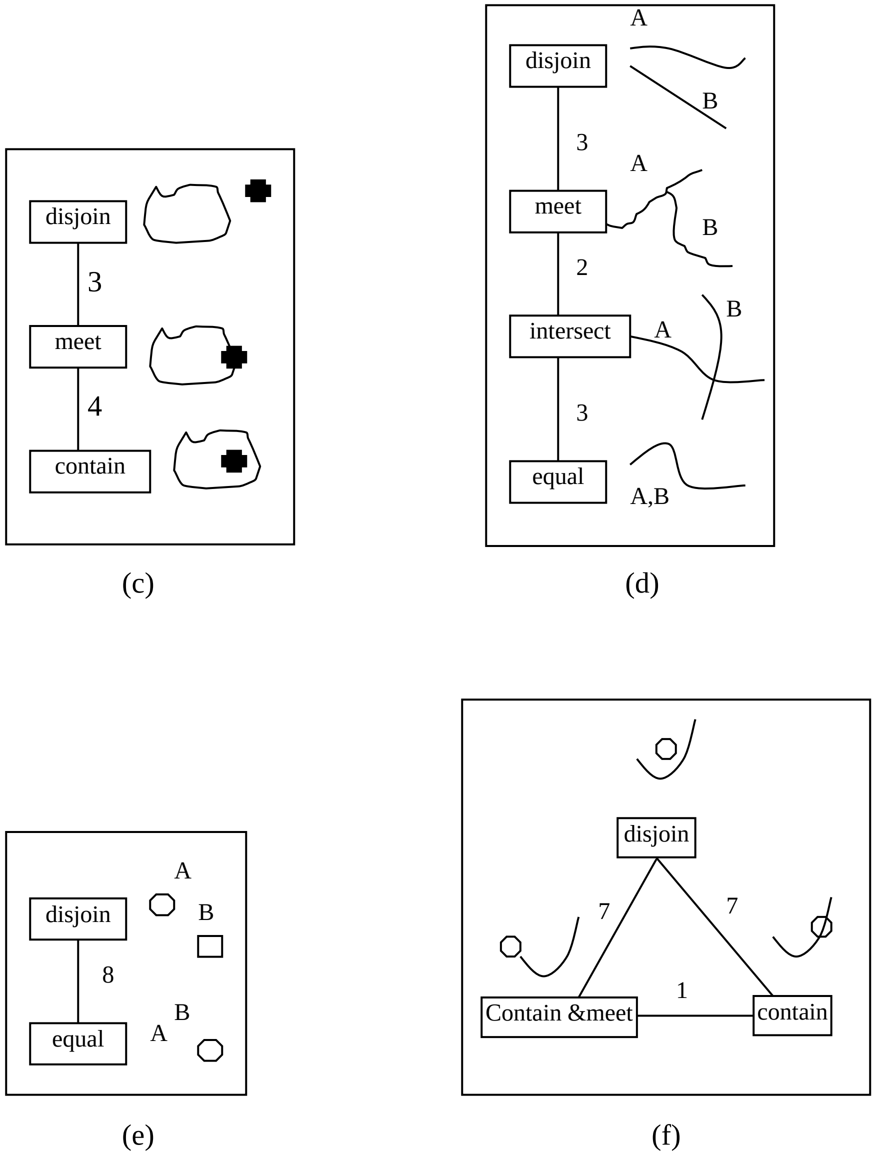


Figure 3-9 Gradual changes of topological relations.

The digit on the edge denotes the transformation cost or the weight between the two adjacent topological relations. (a) two polygons; (b) a polygon and a line; (c) a polygon and a point; (d)two lines; (e) two points; and (f) a line and a point.

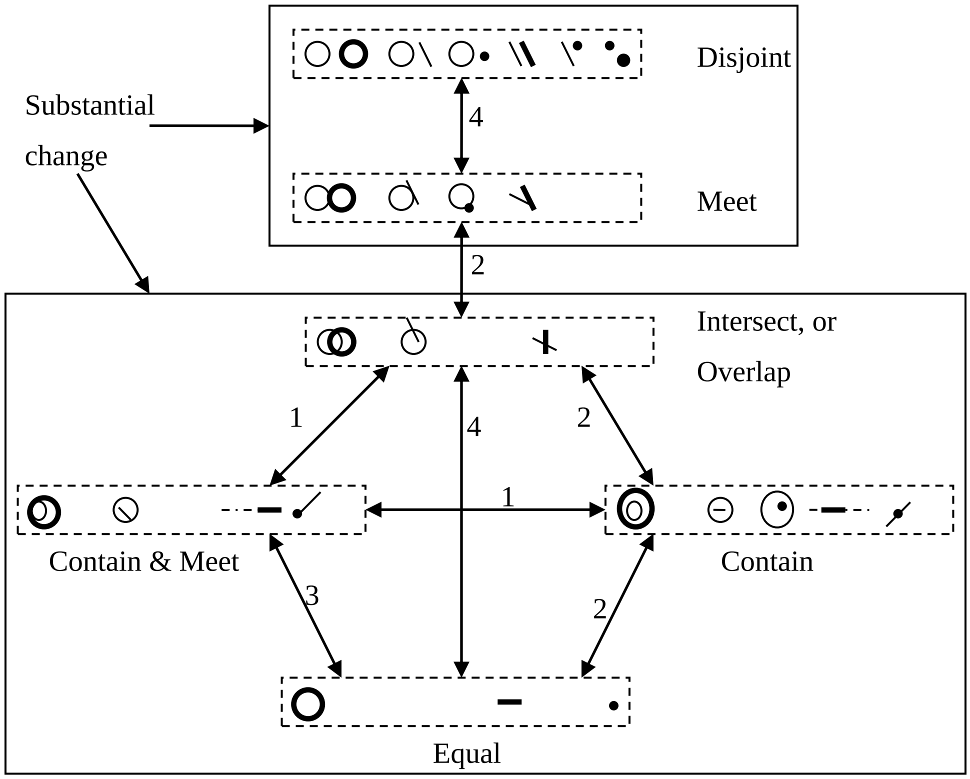


Figure 3-10 Transformation costs (or weights) in topological relations

Table 3-4 Costs in topological relation transformations

	Disjoint	**Meet**	**Overlap/intersect**	**Contain & meet**	**Contain**	**Equal**
Disjoint	**0**	4	***6***	***7***	***8***	***10***
Meet	4	**0**	2	***3***	***4***	***6***
Overlap/intersect	***6***	2	**0**	1	2	4
Contain & meet	***7***	***3***	1	**0**	1	3
Contain	***8***	***4***	2	1	**0**	2
Equal	***10***	***6***	4	3	2	**0**

Notes: bold italic underlined digits, such as "***8***", are calculated using the other digits.

The values are listed in table 3-4. Using this table, the costs between any two topological relations can be obtained.

Direction Relations

Two methods have been address in previous work, i.e. the 16-direction system proposed by Bruns and Egenhofer (1996) and the 9-direction system proposed by Li and Fonseca. Actually, it is not appropriate to specify a direction system before the resolution/scale of the discussed spatial similarity relations are decided, because spatial similarity relations may also be described at different levels of detail, which usually cannot be well expressed using a specified, unchangeable resolution/scale. Indeed, at least three direction systems are usually used in our daily life, i.e. 4-direction system, 8-direction system, and 16-direction system (Figure 3-11). Because there is an additional “same” direction (Yan et al., 2006) in each of the direction systems, they are sometimes called 5/9/17-direction system, instead.

A number of rules are used to quantify the gradual change of direction relations in each of the direction systems. The 8-direction system is taken as an example to facilitate the following discussion (in Figure 3-11(b)).

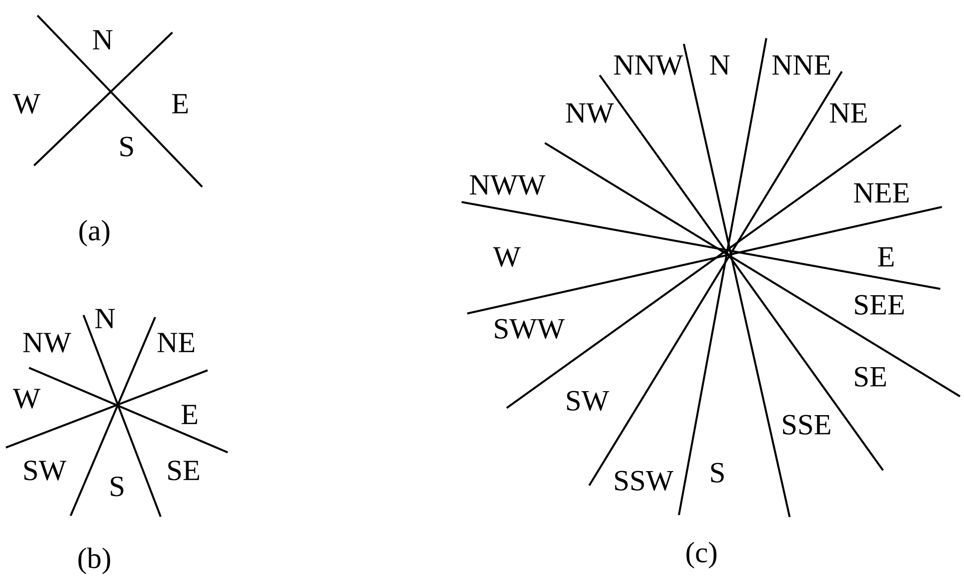


Figure 3-11 Three different direction systems.

- The cost between any two neighbouring directions is 1. For example, the cost between N and NE is 1, because they are neighbouring.

• The cost between any two directions is the sum of the cost in the gradual transformation from the one direction to the other direction. But this value should not be more than half of the total direction number of the direction system. For example, the cost between W and E is 4, because it covers the route W→NW→N→NE→E, which takes four steps; while the cost between W and SE is 3 but not 5, because route W→SW→S→SE is shorter than route W→NW→N→NE→E, and the later takes 5 steps which is greater than half of the total direction numbers of the direction system (i.e. 4).

Metric Distance Relations

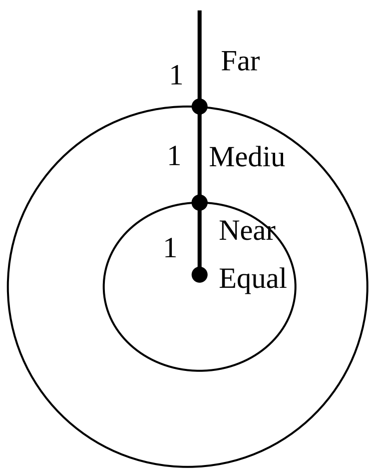


Figure 3-12 Qualitative descriptions of distance relations.

Table 3-5 Costs in direction relation transformations in the 8-direction system

	N	NE	E	SE	S	SW	W	NW
N	0	1	2	3	4	3	2	1
NE	1	0	1	2	3	4	3	2
E	2	1	0	1	2	3	4	3
SE	3	2	1	0	1	2	3	4
S	4	3	2	1	0	1	2	3
SW	3	4	3	2	1	0	1	2
W	2	3	4	3	2	1	0	1
NW	1	2	3	4	3	2	1	0

Qualitative distance relations are difficult to define for general spatial objects, because the terms and concepts used for describing qualitative distance are quite subjective and sensitive to the scale of the spatial data being considered. Bruns and Egenhofer (1996) use four terms “zero”, “very close”,

“close”, and “far” to express the order of such relations, while Li and Fonseca (2006) use “equal”, “near”, “medium”, and “far”, instead. This paper adopts the later in qualitative distances (Figure 3-12), and defines that the transformation cost between any two neighbouring distances is 1 (i.e. between equal and near, between near and medium, and between medium and far).

The core problem of this metric distance relation is to define a criterion that can transform quantitative distance relations into qualitative ones. They can be defined after a couple of prerequisites are defined.

Above all, “directly adjacent” between two objects need to be defined.

Given that there are two objects A and B in a scene, C represents an arbitrary object in the scene. The conclusion “ A and B are directly adjacent” can be made, if and only if no object intersects with an arbitrary line segment L that direct connects the boundaries of A and B. Of course, L has no other intersection with A and B except for its starting point and the end point at the two boundaries.

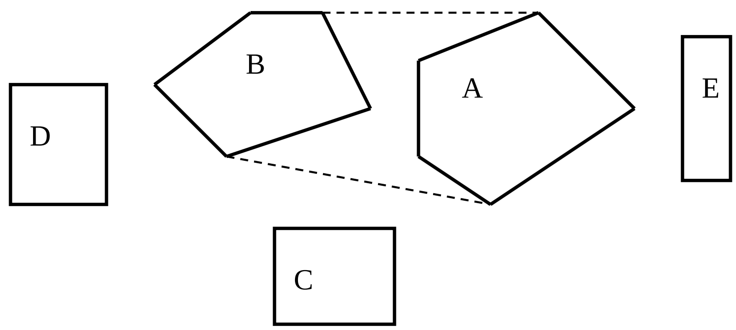


Figure 3-13 Concept of “directly adjacent”.

An example is shown in Figure 3-13 to demonstrate the concept “direct adjacent”.

Then the mean distance between directly adjacent objects can be calculated, supposing that there are totally N pairs of objects that are directly adjacent.

$$\overline{D} = \sum_{i=1}^{n} d_i / N \qquad 3\text{-}23$$

where, $\overline{D}$ is the mean distance; and d_i is the distance between the i^{th} pair of objects.

The four terms in qualitative description of distance relations may be given, based on the above two definitions, supposing that the distance between A and B is d_{AB} .

Equal: if A and B are topologically equal, or intersected/overlap, or $d_{AB} = 0$, they are “equal”.

Near: if $d_{AB} \le \overline{D}$, A and B are “near”.

Medium: if $\overline{D} < d_{AB} \le 2\overline{D}$, A and B are “medium”.

Far: if $d_{AB} > 2\overline{D}$, A and B are “far”.

To quantitatively express the qualitative distance refers to express each of the four terms using corresponding digital values. The values are usually obtained by psychological experiments.

Attributes

Attribute is a similarity factor that measures the internal attribute of an object group that consists of two or more spatial objects. The attributes are composed of two parts, i.e. geometric attributes and thematic attributes, and each part includes many attributes. The attributes are either quantitative (usually expressed using digital values) or qualitative (usually expressed using descriptive words or terms).

Suppose that there are two object groups A and B, each of them have n attributes. Their overall attribute similarity may be expressed as:

$$Sim_{attribite}(A,B) = \sum_{i=1}^{n} w_i Sim(\text{Attritute}^{A_i}, \text{Attribute}^{B_i}) \quad \text{3-24}$$

where, $Attribute^{A_i}$ is the i^{th} attribute of A; $Attribute^{B_i}$ is the i^{th} attribute of B; w_i is the weight of the i^{th} attribute.

3.3.3 Psychological Tests for Determining the Weights of the Factors

Although the factors that affect human’s spatial similarity judgments have been presented, and the idea for quantifying the factors has also been addressed in the previous sections, a crucial problem regarding the factors has not been solved yet, i.e. the weights of the factors are unknown. Because the weights depend on human’s cognition, psychological experiments are employed to determine the weights here. The experiments are divided into two parts: Experiment 1 is for object groups, and Experiment 2 is for individual objects.

The following gives a detailed description of the experiments.

◆ Basic information of the test

Time: October 12, 2013.

Place: Lanzhou Jiaotong University, P.R. China.

Subjects: 52 students at undergraduate or graduate level, 24 female and 28 male. Their age ranges from 17 to 27. All subjects are majoring in or have majored in geography and related communities, including 27 in geographic information science, 17 in cartography, 2 in surveying, 3 in human geography, and 3 in physical geography.

It is not easy to recruit enough subjects. To carry out this task, an advertisement was posted in the webpage of Lanzhou Jiaotong University, China, about 20 days before the psychological tests. Every subject is required to register his/her basic information (e.g. name, age, gender, major/career and contact information) in a table.

◆ Goal of the test

(1) to get the weights of topological relations, direction relations, distance relations and attributes of object groups in human's spatial similarity judgments; and

(2) to get the weights of the attributes (geometric attributes and thematic attributes) of individual spatial objects in human's spatial similarity judgments.

◆ Steps in the two experiments

Step 1: Select the factors that need to be tested and design the structure of the answer sheet. In Experiment 1, because it is for object groups, topological relations, direction relations, distance relations and attributes need to be considered. In Experiment 2, because it is for individual objects, only attributes need to be considered.

Step 2: Systematically design the samples that are used in the experiments. There are totally six types of object group in the 2-dimensional space, i.e. point-point, point-line, point-polygon, line-line, line-polygon and polygon-polygon; therefore, six samples corresponding to the six types of object group are constructed. In each sample four changes are designed to show the corresponding four factors (i.e. topological relation, direction relation, distance relation, and attribute). The samples are shown in Figure 3-14 and Figures 3-16 to Figure 3-20.

Step 3: Distribute the samples to the subjects, and explain the regulations to them.

Each of the subjects is invited respectively to participate in the test. Firstly, the subject is given one of the samples and an answer sheet; then the subjects are told that the four transformations in the sample; last, they are required to compare the original graph with each of the four transformations, describe their similarity degree using a decimal, and ensure that the sum of the four decimal is equal to 1.

Step 4: Collect the test sheets, analyze the data.

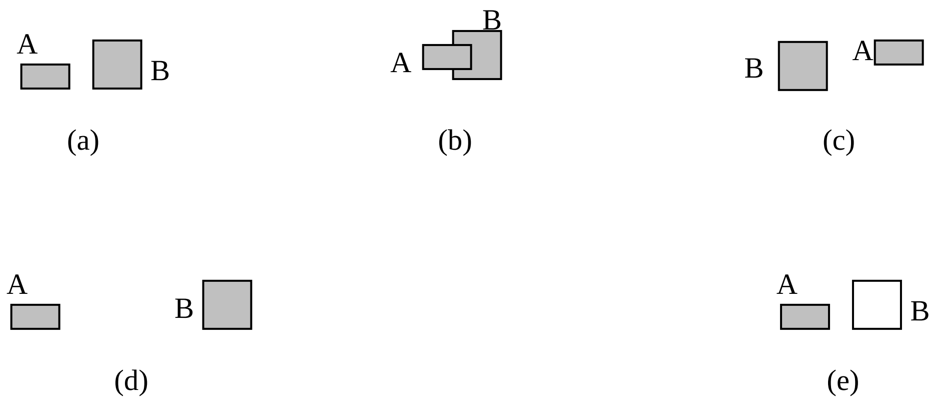


Figure 3-14 Factors for polygon-polygon groups in similarity judgments.

(a) original object group with two objects A and B; (b) topological transformation; (c) direction transformation; (d) distance transformation; and (e) attribute transformation.

◆Experiment 1: for object groups

Figure 3-14 illustrates the three transformations in topological relations, direction relations, distance relations, and attributes, respectively. The subjects are required to answer the following questions on the answer sheet according to the instructions (see Figure 3-15).

Please use a decimal to denote the weights of topological relations, direction relations, distance relations, and attributes after evaluating corresponding similarity changes in this example. The sum of the four weights should be 1.

(1) Weight of the topological relations _____________

(2) Weight of the direction relations_______________

(3) Weight of the distance relations _______________

(4) Weight of the attributes _______________

Figure 3-15 Answer sheet used in Experiment 1.

The same answer sheets are used in the other samples of Experiment 1.

Because the three relations may exist between six kinds of object pairs, i.e. polygon-polygon, polygon-line, polygon-point, line-line, line-point, and point-point, the other five kinds of examples are also used in the experiments (from Figure 3-16 to Figure 20).

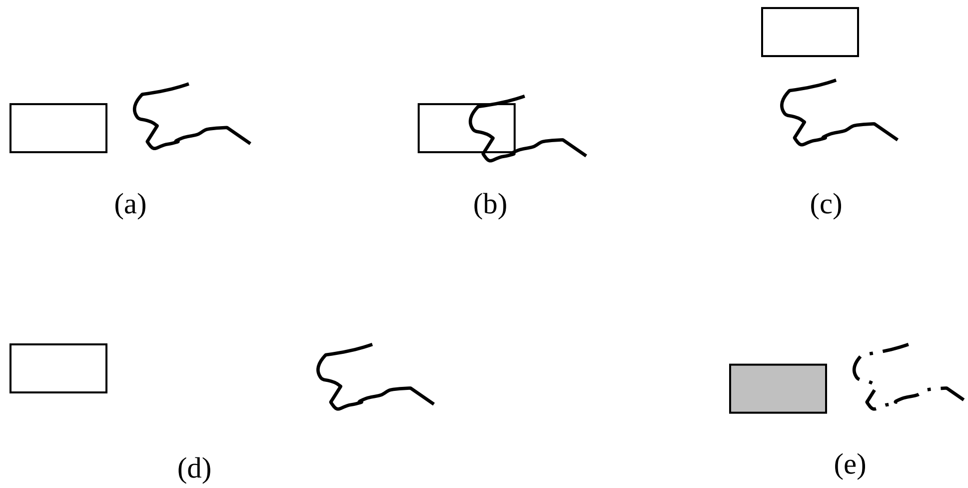


Figure 3-16 Factors for polygon-line groups in similarity judgments.

(a) original object group; (b) topological transformation; (c) direction transformation; (d) distance transformation; and (e) attribute transformation.

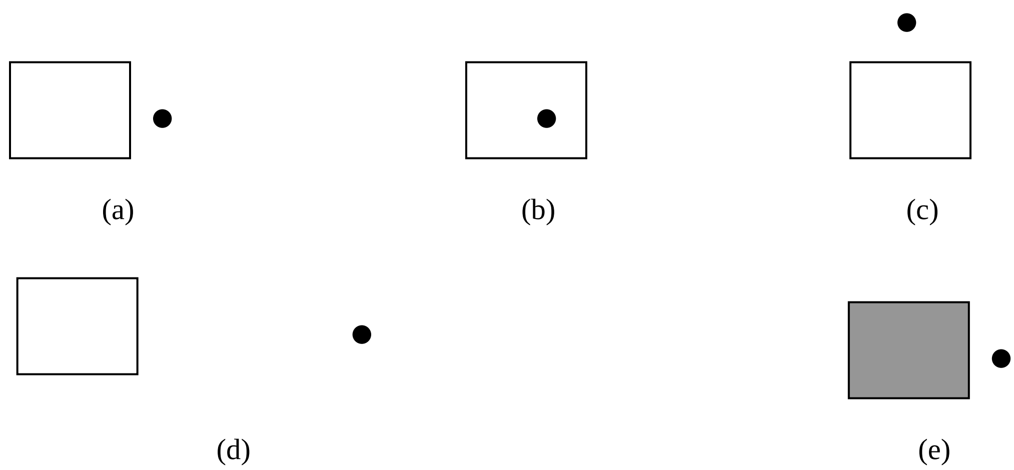


Figure 3-17 Factors for polygon-point groups in similarity judgments.

(a) original object group; (b) topological transformation; (c) direction transformation; (d) distance transformation; and (e) attribute transformation.

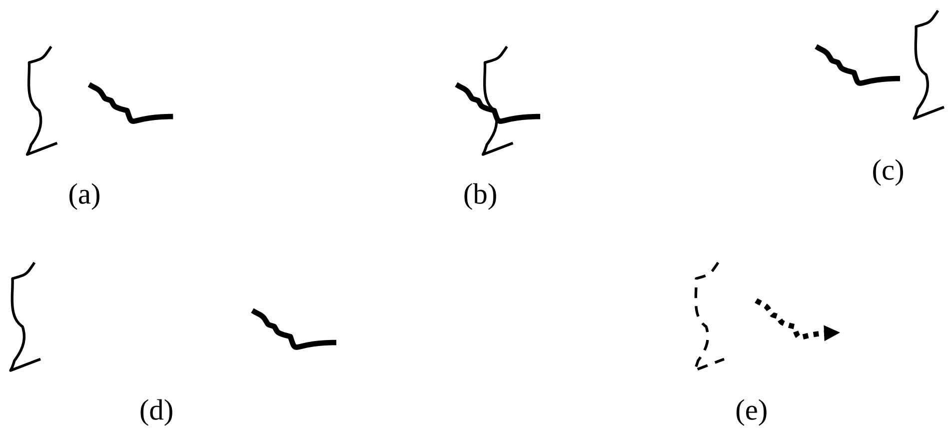


Figure 3-18 Factors for line-line groups in similarity judgments.

(a) original object group; (b) topological transformation; (c) direction transformation; (d) distance transformation; and (e) attribute transformation.

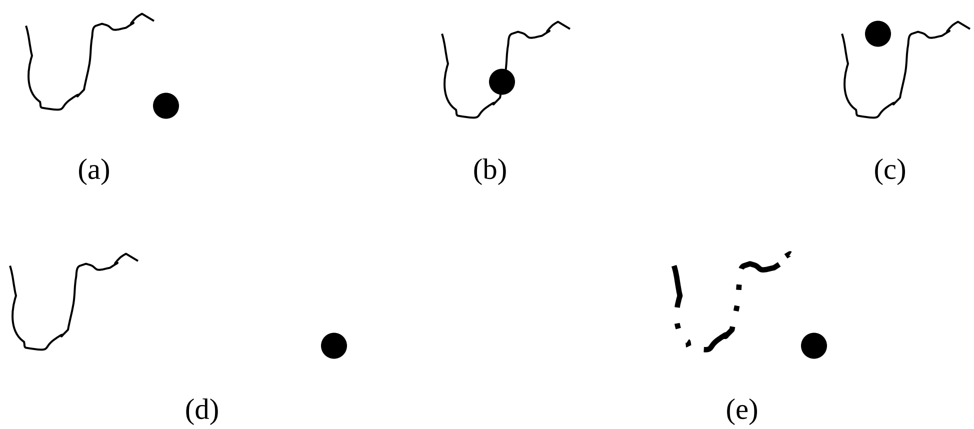


Figure 3-19 Factors for line-point groups in similarity judgments.

(a) original object group; (b) topological transformation; (c) direction transformation; (d) distance transformation; and (e) attribute transformation.

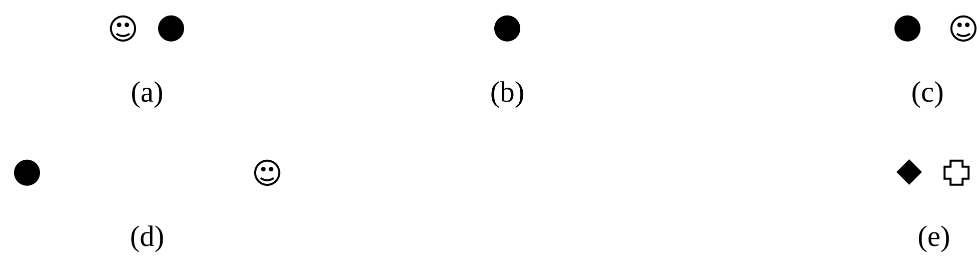


Figure 3-20 Factors for point-point groups in similarity judgments.

(a) original object group; (b) topological transformation; (c) direction transformation; (d) distance transformation; and (e) attribute transformation.

The answers to Experiment 1 are listed in Table 3-6, by which the mean value of each weight can be obtained.

Table 3-6 Weights of the four factors of the object groups

	Total weights obtained from the 52 subjects			
	Topological	Direction	Distance	Attribute
Figure 3-15	13.00	10.92	16.12	11.96
Figure 3-16	13.00	11.44	16.64	10.92
Figure 3-17	10.92	12.48	15.60	13.00
Figure 3-18	11.44	13.52	16.64	10.40
Figure 3-19	10.92	13.00	16.12	11.96
Figure 3-20	10.92	15.60	15.60	9.88
Standard deviation	1.07	1.08	0.44	1.01

$$w_{topologival} = \sum_{1}^{6} w_i^{topological} / (52 \times 6) = 0.22 \qquad 3\text{-}25$$

$$w_{direction} = \sum_{1}^{6} w_i^{direction} / (52 \times 6) = 0.25 \qquad 3\text{-}26$$

$$w_{distance} = \sum_{1}^{6} w_i^{distance} / (52 \times 6) = 0.31 \qquad 3\text{-}27$$

$$w_{attribute} = \sum_{1}^{6} w_i^{attribute} / (52 \times 6) = 0.22 \qquad \text{3-28}$$

Where, $w_i^{topological}$, $w_i^{direction}$, $w_i^{distance}$ and $w_i^{attribute}$ correspond to the data listed in Table 3-6.

The standard deviations of the four weights obtained from the 52 subjects are listed in Table 3-6. Accordingly, the standard deviations of the four weights for per subject are 1.07/52=0.021, 1.08/52=0.021, 0.44/52=0.008, 1.01/52=0.019. The percentages of the four standard deviations in the corresponding weights are 0.021/0.22=9.5%, 0.021/0.25=8.4%, 0.008/0.31=2.6%, 0.019/0.22=8.6%. This shows that the subjects' recognition to the four weights is stable.

◆ Experiment 2: for individual objects

(a) (b) (c)

Figure 3-21 Factors for an individual areal object in similarity judgments.

(a) Original object; (b) change of geometric attributes; and (c) change of thematic attributes.

Attributes of spatial objects consists of geometric attributes and thematic attributes. There are many geometric attributes and numerous of that of the thematic attributes in map spaces, and the attributes that are used for similarity judgments change on different occasions; thus it is impossible to get the weights of all attributes that can be popularly accepted. To simplify this problem, only geometric attributes and thematic attributes are differentiated, and therefore two weights corresponding to each of them are considered.

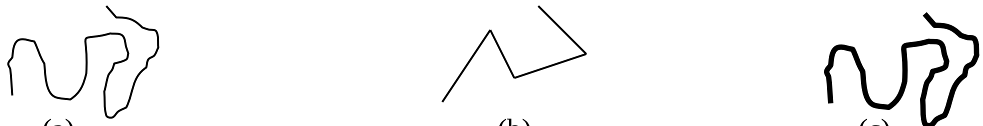
(a) (b) (c)

Figure 3-22 Factors for an individual point object in similarity judgments.

(a) Original object; (b) change of geometric attributes; and (c) change of thematic attributes.

Three categories of individual spatial objects (i.e. points, lines and polygons) are enumerated in the following three samples (Figure 3-21, Figure 3-22 and Figure 3-23) and the answer sheet (Figure 3-24) together with the samples are presented to the subjects. Two transformations are shown in each of the three samples.

The answers of the experiment 2 are listed in Table 3-7, by which the mean value of each weight can be obtained.



(a) (b) (c)

Figure 3-23 Factors for an individual linear object in similarity judgments.

(a) Original object; (b) change of geometric attributes; and (c) change of thematic

$$w_{Geometric} = \sum_{1}^{3} w_i^{Geometric} / (52 \times 3) = 0.53 \qquad 3\text{-}29$$

Please use a decimal to denote the weights of the geometric attributes and the thematic attributes after evaluating corresponding similarity changes in this example. The sum of the two weights should be 1.

(1) Weight of the geometric attributes_______________

(2) Weight of the thematic attributes_______________

Figure 3-24 Answer sheet used in Experiment 1.

$$w_{Thematic} = \sum_{1}^{3} w_i^{Thematic} / (52 \times 3) = 0.47 \qquad 3\text{-}30$$

Where, $w_i^{Geometric}$ and $w_i^{Thematic}$ correspond the data listed in Table 3-7.

Table 3-7 Weights of geometric and thematic attributes from the 52 subjects

	Geometric attributes	**Thematic attributes**
Figure 3-21	27.56	24.44
Figure 3-22	32.24	19.76
Figure 3-23	22.36	29.64
Standard deviation	3.309	3.309

The standard deviations of the two weights obtained from the 52 subjects are listed in Table 3-7. Accordingly, the standard deviations of the two weights for per subject are 3.309/52=0.064, 3.309/52=0.064. The percentages of the two standard deviations in the corresponding weights are 0.064/0.53=12.1%, 0.064/0.47=13.6%. This shows that the subjects' recognition to the two weights is stable.

3.4 Classification

Classification of spatial similarity relations not only presents the relations of every aspects of spatial similarity relations, but also helps to organize relevant research work clearly. This section addresses the classification of spatial similarity relations in the geographic space and on line maps.

3.4.1 A Classification System of Spatial Similarity Relations in Geographic Spaces

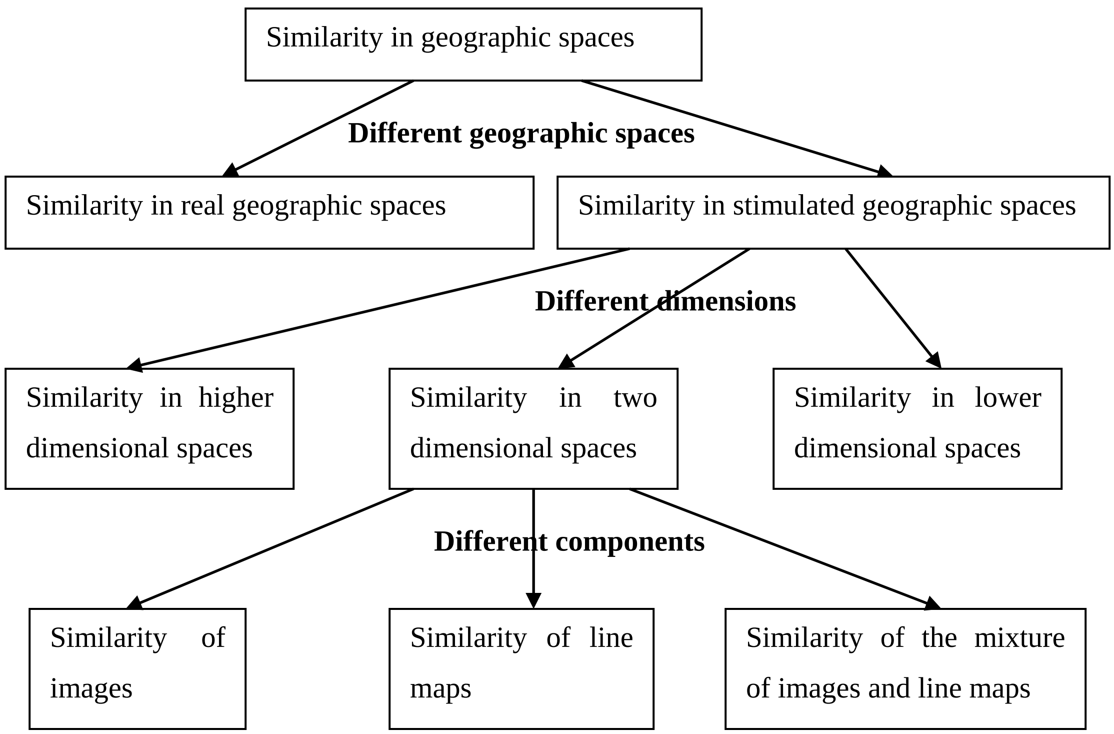


Figure 3-25 A classification system of similarity in geographic spaces.

As far as similarity in geography is concerned, it may be classified into two categories: similarity in real geographic spaces and similarity in analoge geographic spaces (Figure 3-25), taking geographic spaces as the criterion in the classification. This study is only interested in similarity in map spaces; hence stimulated geographic spaces are further classified into "similarity in higher dimensional spaces", "similarity in two-dimensional spaces", and "similarity in lower dimensional spaces" according to their dimensions. Similarity in two dimensional spaces comprises “similarity of images”, “similarity of line maps” and “similarity of the mixture of images and line maps”.

Sililariry in multi-scale map spaces is belong to the simialrity of line maps. It needs to be classified further.

3.4.2 A Classification System of Spatial Similarity Relations on Line Maps

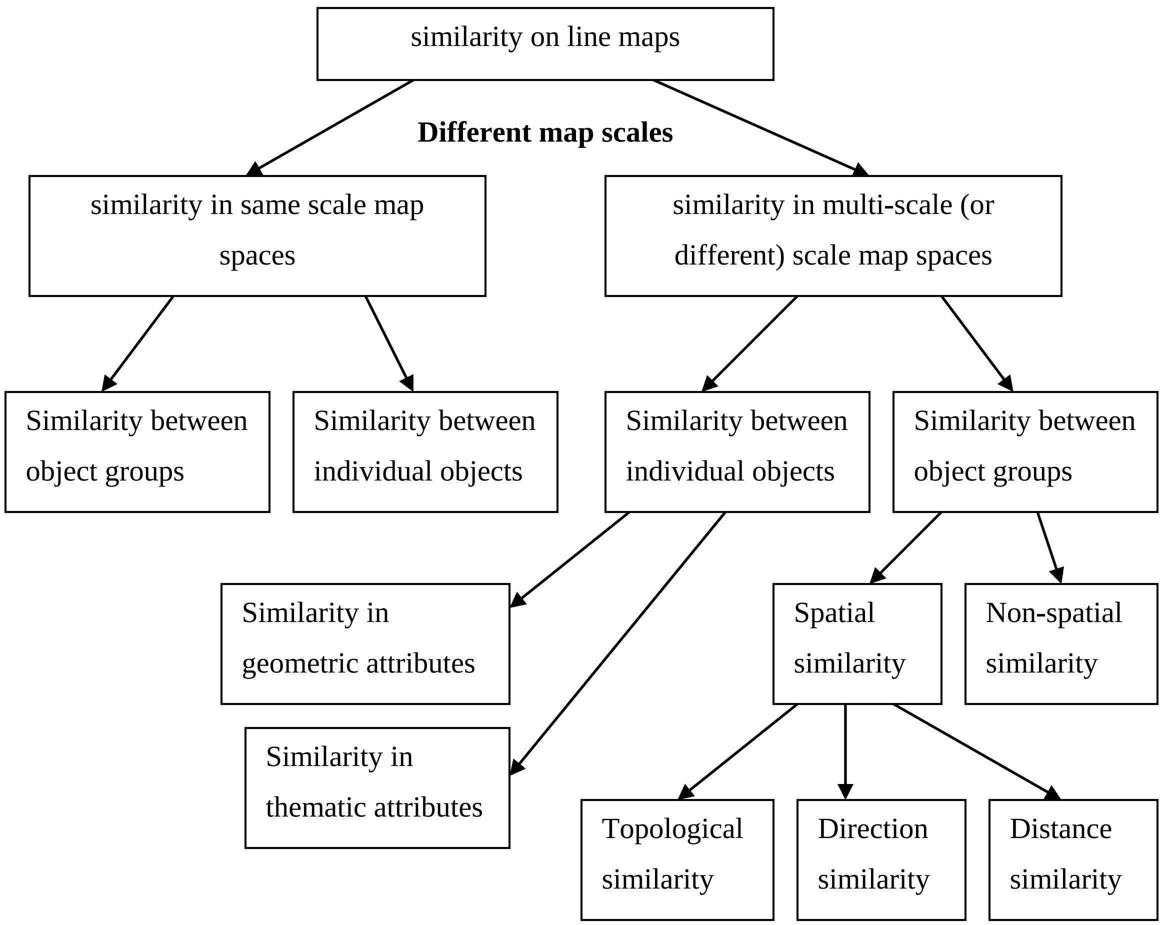


Figure 3-26 A classification system of similarity on line maps

Figure 3-26 present a classification system for spatial similarity relations on line maps.

First, spatial similarity relations on line maps can be classified into similarity in same scale map spaces and similarity in multi-scale (different) scale map spaces. The former is called horizontal similarity relations, considering the similarity between objects at same map scale; the latter is called perpendicular similarity relations (Yan, 2010), focusing on the similarity of objects at different map scales, which is the emphasis of this study.

Second, similarity in multi-scale scale map spaces may be evaluated either between individual objects or between object groups.

Figure 3-27 presents an example to demonstrate this concept: a cluster of land parcel and an individual land parcel on the map at scale 1:10K is generalized to generate the graphs at scales 1:25K and 1:50K. That what spatial similarity relations are changed between the individual parcels and

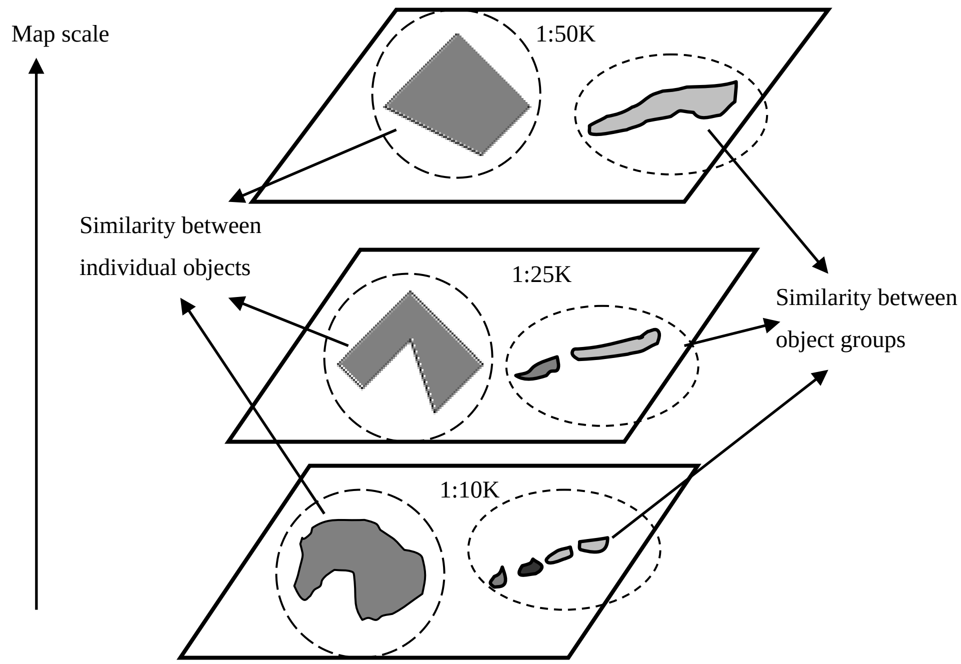


Figure 3-27 An example of similarity in multi-scale scale map spaces.

between the original parcel group and the generalized counterparts is of great interests to many cartographers.

Third, suppose that an individual object A is at scale S_A and another individual object B is at scale S_B and $S_A \neq S_B$, similarity relations between A and B is a kind of similarity between individual objects in multi-scale map spaces. Such similarity is evaluated by both the geometric attributes and thematic attributes.

If $S_B < S_A$ and A and B are different symbols of the same object (see Figure 3-27 for an example), the spatial similarity relation between them is of significance to automated map generalization.

Last, spatial similarity relations between object groups in multi-scale map spaces include either different object groups or different symbols of the same object groups at different scales. To evaluate such kind of spatial similarity relations, both non-spatial similarity (including geometric and thematic attributes) and spatial similarity (including topological, directional and distance relations) should be taken into account.

3.5 Chapter Summary

This chapter addresses the fundamental issues of spatial similarity relations.

It first proposes the definitions of similarity relation, spatial similarity relation and spatial similarity relation in multi-scale map spaces.

Second, it addresses the features of spatial similarity relations, including equality, finiteness, miniality, auto-similarity, symmetry, non-transitivity, weak symmetry, asymmetry, triangle inequality, and scale-dependence.

Third, it proposes the factors that affect human's direction judgments. These factors include the ones for individual objects and the ones for object groups. The psychological experiments are designed to get the weights of the factors in spatial similarity judgments.

Last, a classification system for spatial similarity is presented.

Chapter 4 Models for Calculating Spatial Similarity Degrees in Multi-scale Map Spaces[2]

It is a challenge work to propose new models for calculating spatial similarity degrees between objects in multi-scale map spaces. In this chapter, ten new models are proposed. Three models are for individual objects and the other seven models are for object groups. To be exact, the former comprises the models for individual point objects, individual linear objects and individual areal objects, and the latter comprises the models for point clouds, parallel line clusters, intersected line networks, tree-like networks, discrete polygon groups, connected polygon groups and maps.

4.1 Models for Individual Objects

As proposed in Chapter 3, two factors that affect human's spatial similarity judgments should be taken into consideration in constructing the models for individual objects, i.e. geometric attributes and thematic attributes.

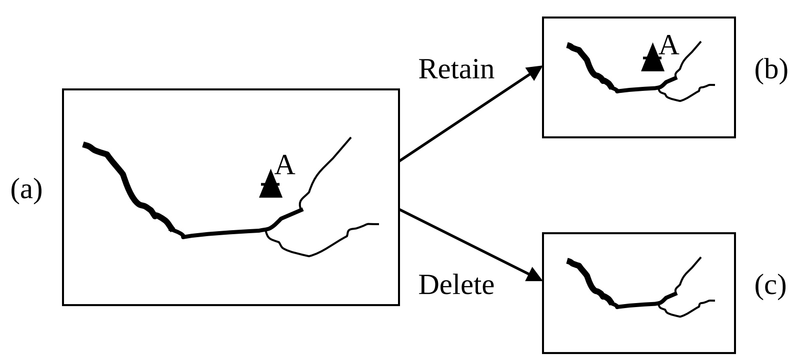


Figure 4-1The individual pavilion A can be retained or deleted.

[2] Partial of the chapter has been published by: Yan H., Li J. and Wang Z., 2013, An approach to computing direction relations between separated object groups, GeoScientific Model Development, 6(2): 1591-1599; and partial has been published by: Yan H. and Li J., 2013, An approach to simplifying point features on maps using the Multiplicative Weighted Voronoi Diagram, Journal of Spatial Science, 58(2): 291-304.

4.1.1 Model for Individual Point Objects

In map generalization, an individual point object cannot be simplified, which means its geometric attributes and thematic attributes cannot be changed on the map. The operations that can be executed to it are “deletion” or “retaining” (Figure 4-1). Thus, the similarity degree of a point object A at scales l and m can be calculated using the following formula, given that $l > m$.

$Sim(\mathrm{A}_l, \mathrm{A}_m) = 0$ *, if* A *is deleted from the map at scale* m *; else,*

$$Sim(\mathrm{A}_l, \mathrm{A}_m) = 1. \qquad 4\text{-}1$$

4.1.2 Model for Individual Linear Objects

Measuring curve similarity is a fundamental problem in many application fields, including graphics, computer vision, cartography and geographic information science (Alt and Godau, 1995; Alt et. Al., 1998; Yan, 2010). An individual linear object on the map may be a line segment (e.g. a short trail), a curve (e.g. a zigzag country road), or a closed curve (e.g. a boundary of a province or a country, or a closed contour on a map). When it is generalized, its geometric attributes may be changed (e.g. removal of curvatures from a zigzag contour line) and its thematic attributes can also be modified (e.g. change of river grade). Thus, a generic model that takes into account both geometric attributes and thematic attributes of an individual linear object may be constructed, based on Formula 3-9 and Formula 3-10, given that the original map scale is k and the resulting map scale is m .

$$Sim(\mathrm{A}_k, \mathrm{A}_m) = w_{thematic} Sim^{thematic}_{A_k, A_m} + w_{geometric} Sim^{geometric}_{A_k, A_m}. \qquad 4\text{-}2$$

where, $w_{thematic}$ is the weight of thematic attributes of the individual linear object; $w_{geometric}$ is the weight of geometric attributes of the individual linear object; $Sim^{thematic}_{A_k, A_m}$ is the spatial similarity degree of object A at scale k and scale m ; and $Sim^{geometric}_{A_k, A_m}$ is the spatial similarity degree of object A at scale k and scale m .

Formula 4-2 can be simplified to get Formula 4-3, because cartographers pay most of their attention to the geometric attributes of individual linear objects and ignore their thematic information. This conclusion is also supported by the previous psychological experiments in Chapter 3.

$$Sim(\mathrm{A}_k, \mathrm{A}_m) = Sim^{geometric}_{A_k, A_m}. \qquad 4\text{-}3$$

A Formula for Calculating Shape Similarity between Lines

Shape is viewed as the most crucial, sometimes the only, geometric factor for describing planar curves (Douglas and Peucker, 1973; Mokhtarian and Mackworth, 1992). This is also the case in multi-scale representation of individual lines in map spaces. Therefore, similarity of shape of an individual line at two scales is an appropriate substitute for the similarity of the line at two the scales. It can be expressed as

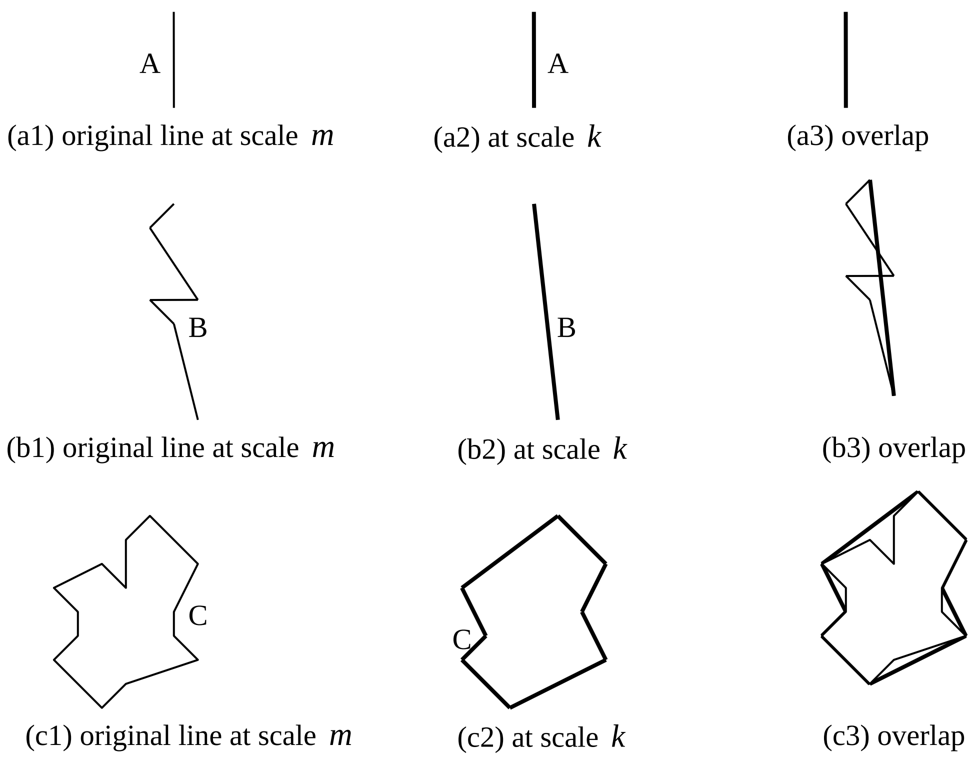


Figure 4-2 Overlap of an individual line and its generalized counterpart.

$$Sim(\mathrm{A}_k, \mathrm{A}_m) = Sim^{shape}_{A_k, A_m}. \qquad 4\text{-}4$$

Hence, the following will propose a method for calculating similarity degrees of the shapes of individual lines in multi-scale map spaces based on the concept "coincidence summary" used to assess the similarity between maps (Berry, 1993). "Coincidence summary" used the percentage of the map area in agreement (or disagreement) between the two maps to indicate the overall similarity. In vector analysis maps are intersected to generate the areas of the son-and-daughter polygons to summarize the type of disagreement. In grid-based analysis the process simply involves noting the number of grid cells falling into each category combination.

Based on coincidence summary and human's intuition in similarity judgments, similarity between two lines on the map can be evaluated by comparing the common length of the two lines. After overlapping the two lines at two different scales and matching their corresponding endpoints, their common length may be easily calculated (Figure 4-2), and their similarity degree can also be obtained.

$$Sim_{A_k,A_m}^{shape} = \frac{l}{L} \quad 4\text{-}5$$

where, L is the length of the original line at scale m ; and l is the common length of the line at scale m and the simplified line at scale k .

Formula 4-5 can work well in many cases. For example, in Figure 4-2, the three similarity degrees are

$$Sim_{A_k,A_m}^{shape} = \frac{l}{L} = 1.00$$

$$Sim_{B_k,B_m}^{shape} = \frac{l}{L} = 0.00$$

$$Sim_{C_k,C_m}^{shape} = \frac{l}{L} = 0.32$$

However, it sometimes gives inappropriate results. For example, in Figure 4-2(b1), (b2) and (b3), the similarity degree of line B at scale m and scale k is $Sim_{B_k,B_m}^{shape} = \frac{l}{L} = 0$, because the length of the intersection of the two lines is 0. This conclusion is obviously discrepant with human's spatial cognition in daily life.

Improvement of the Formula

To compensate for the shortcoming, an improve formula is proposed here, taking into account the distance between the two lines.

$$Sim_{A_k,A_m}^{shape} = \sum_{i=1}^{n} w_i l_i / L \quad 4\text{-}6$$

where, L is the length of the original line; n is the number of the line segments contained in the resulting line; l_i is the length of the i^{th} line segment of the resulting line; and w_i is the weight of l_i, which can be calculated by

$$w_i = 1 - \frac{\overline{d_i} l_i}{\sum_{j=1}^{n} \overline{d_j} l_j} \quad \text{4-7}$$

where, n, l_i and l_j are the same as that in Formula 4-6; and $\overline{d_i}$ is the mean distance between l_i and the original line, and it is the distance from the midpoint of l_i to the original line.

Using Formula 4-6, the similarity degrees in Figure 4-2 can also be calculate.

$$Sim^{shape}_{A_k, A_m} = \sum_{i=1}^{n} w_i l_i / L = 1.00$$

$$Sim^{shape}_{B_k, B_m} = \sum_{i=1}^{n} w_i l_i / L = 0.78$$

$$Sim^{shape}_{C_k, C_m} = \sum_{i=1}^{n} w_i l_i / L = 0.55$$

These results are obviously more reasonable.

4.1.3 Model for Individual Areal Objects

Individual areal objects refer to individual polygons. Many objects on maps are represented using polyugons, such as settlements, waterbodies, forest, etc. If the scale of these maps becomes smaller, the boundaries of the polygons need to be simplified so that they can be adaptive to the new map scale. As far as the generalization of an individual polygon is concerned, cartographers usually need to consider the consistency of the shape of the polygon at different scales, and ignore the other attributes including the thematic attributes and the other geometric attributes (Douglas & Peucker, 1973); thus, similarity of shape of individual polygons at different scales can be viewed as the similairy of the polygon at different scales.

The similarity degree of shape of an arbitrary individual polygon P at scale k and scale m can be simply calculated by

$$Sim^{shape}_{P_k, P_m} = 1 - \frac{Abs\left|A_{P_k} - A_{P_m}\right|}{A_{P_k}} \quad \text{4-8}$$

where, A_{P_k} is the area of polygon P at scale k, and A_{P_m} is the area of polygon P at scale m.

It should be noted that polygons discussed in this study are simple polygons. A polygon is called a simple polygon if it contains no holes, and its non-adjacent edges do not intersect with each other.

4.2 Models for Object Groups

It has been proposed in Chapter 3 that four factors affecting human's spatial similarity judgments regarding object groups need to be taken into account, i.e. topological relations, direction relations, distance relations and attribute relations. The following sections addresses the models for calculating similarity degrees of various object groups in multi-scale map spaces, mainly considering the above four factors.

The problem that is going to be addressed can be described as follows:

Suppose that A_l *is an object group consisting of* N_l *objects on the map at scale* l, A_m *is a generalized object group consisting of* N_m *objects at scale* m *.The property set of* A_l *and* A_m *is* $P = \{P_{Topological}, \mathrm{P}_{Direction}, P_{Distance}, P_{Attribute}\}$ *, and the corresponding weight set is* $W = \{W_{Topological}, \mathrm{W}_{Direction}, W_{Distance}, W_{Attribute}\}$ *.it is required to calculate* Sim_{A_l, A_m} .

According to Formula 3-10, $Sim(\mathrm{A}_l, \mathrm{A}_m) = \sum_{i=1}^{4} w_i Sim_{A_l, A_m}^{P_i}$.

where, $w_i \in W$ and $P_i \in P$; $i = 1, 2, 3, 4$.

The value of $w_i \in W$ are obtained by the psychological experiments in Chapter 3. Therefore, if $Sim_{A_l, A_m}^{P_{Topological}}$, $Sim_{A_l, A_m}^{P_{Direction}}$, $Sim_{A_l, A_m}^{P_{Distance}}$ and $Sim_{A_l, A_m}^{P_{Attribute}}$ in the new models can be calculated, $Sim(\mathrm{A}_l, \mathrm{A}_m)$ can be obtained. Thus, the following sections focus on the calculation of $Sim_{A_l, A_m}^{P_{Topological}}$, $Sim_{A_l, A_m}^{P_{Direction}}$, $Sim_{A_l, A_m}^{P_{Distance}}$ and $Sim_{A_l, A_m}^{P_{Attribute}}$ of object groups at scales l *and* m .

4.2.1 Model for Point Clouds

Many natural and man-made features appear on maps like point clouds. For example, the control points in Figure 4-3 can be viewed as point clouds when they are displayed on a separated map layer. If the map are reduced to a smaller scale one, the point clouds need to be simplified so that they are

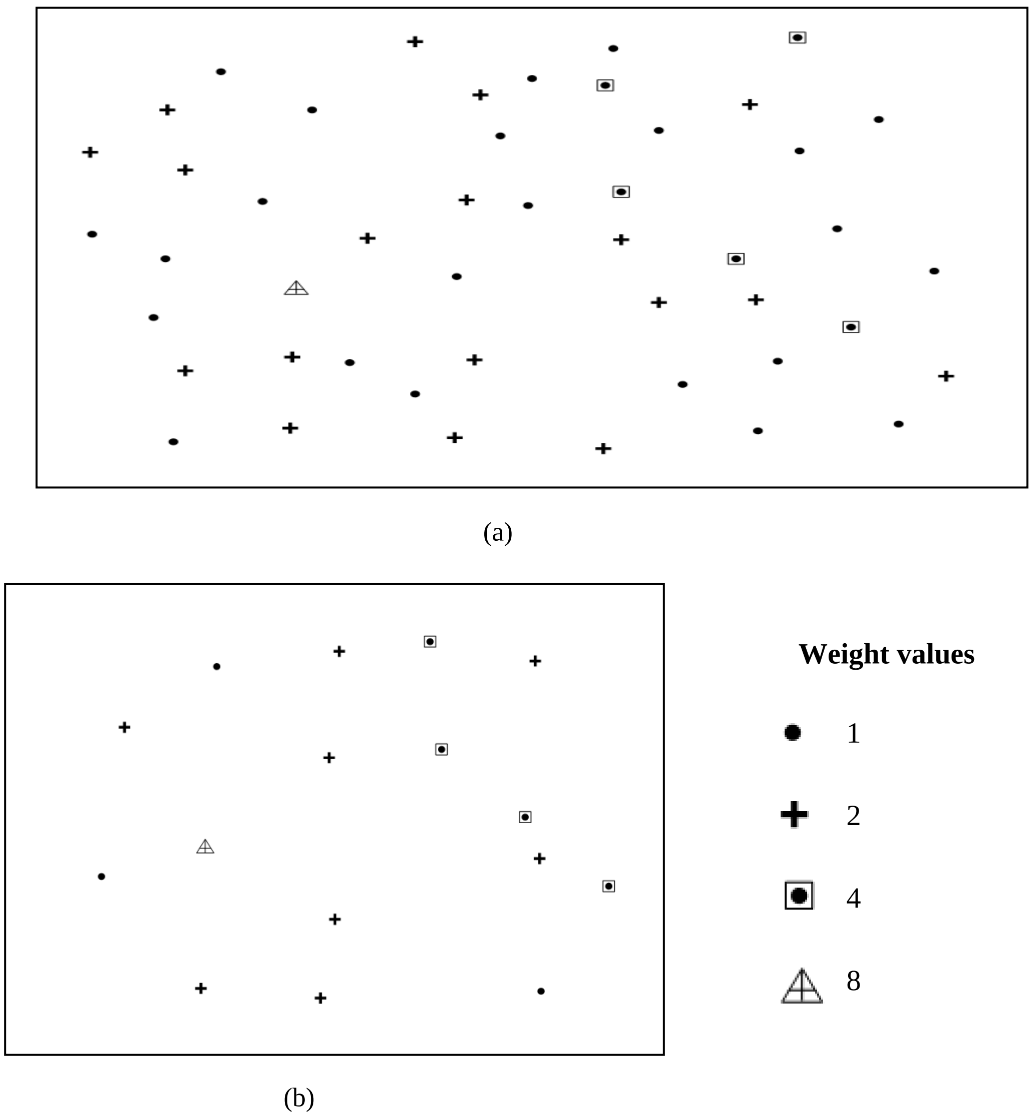


Figure 4-3 An example of point clouds and generalized point clouds.

(a) Control points of a region on the map can be viewed as point clouds when they are displayed on a separated map layer; and (b) generalized control points.

legible, which means some less important points should be deleted from the original map. The control points in Figure 4-3(b) are generalized by the map in Figure 4-3(a), which shows that the point with greater weight values have more probabilities to be retained on the generalized map.

Similarity in Topological Relations

Above all, the definition of topological relation among points is given here, using the concept of the *k*-order Voronoi neighbour.

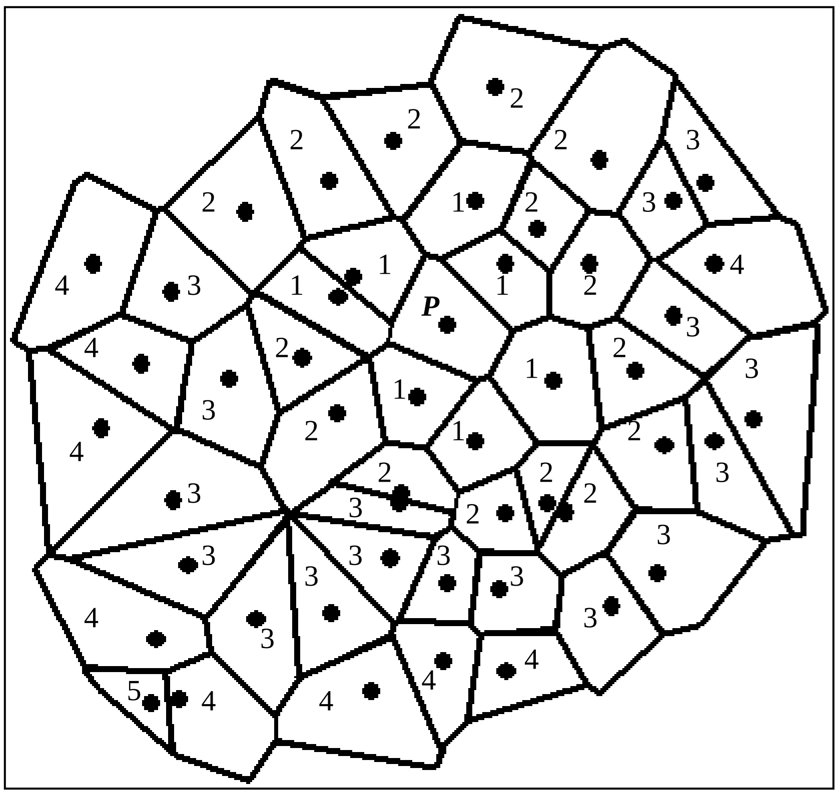


Figure 4-4 The definition of *K*-order Voronoi neighbors.

The number n = 1, 2, 3, 4, 5 in each Voronoi polygon denotes that the corresponding point is an *n*-order neighbor of point *P*.

(1) Point P is the 0-order Voronoi neighbour of itself;

(2) if the Voronoi polygon of point Q shares a common edge with that of a (k-1)-order Voronoi neighbour of P, Q is defined as a k-order Voronoi neighbour of P. Where, k=1,2,…; and

(3)1-order Voronoi neighbours of Point P are called the topological neighbours of P.

Figure 4-4 shows 1-order to 5-order Voronoi neighbours of point P. It is easy to know that point P totally has 7 1-order Voronoi neighbours.

The following two rules are usually obeyed by cartographers to guarantee that topological relations among points can be preserved well in the process of map generalization.

(1) The deletion of adjacent points is generally unacceptable by cartographers in practice if the change of map scale does not have a large span (e.g. from 1:10K to 1:25K). For example, in Figure 4-4, it is not unsatisfactory to delete points P along with any of its neighbours when the map is generalized from 1:10K to 1:25K.

(2) In point cloud generalization, simultaneous deletion of a point and some of its 1-order neighbours possibly makes those distant points become neighbours, which leads to distant things abruptly becoming related. In theory, this operation is contrary to the First Law of Geography: "everything is related to everything else, but near things are more related than distant things" (Tobler 1970, pp.234). For example, in Figure 4-5, if point P and its 1st-order neighbours P_1 and P_2 are deleted, points P_3 and P_4 (they are 3rd-order neighbours of each other) will be 1st-order neighbours and abruptly become closely related; whereas, if only point P is deleted, points P_1 and P_4, 2nd-order neighbours of each other, will become 1st-order neighbours, which is obviously more natural and acceptable by map readers.

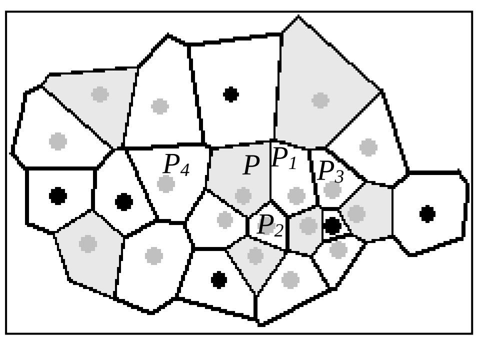


Figure 4-5 The principles of point deletion.

The similarity degree in topological relations of a point cloud at two map scales can be defined as:

$$Sim_{A_l,A_m}^{P_{Topological}} = \frac{\sum_{i=1}^{N_m} n_m^i}{\sum_{i=1}^{N_m} n_l^i} \qquad 4\text{-}9$$

where, N_m is the number of points retained on the map at scale m; for the i^{th} point on the map at scale m, n_l^i is the number of its 1st-order Voronoi neighbours on the map at scale l; and for the i^{th} point on the map at scale m, n_m^i is the number of common 1st-order Voronoi neighbours of the i^{th} point on the map at scale m and on the map at scale l.

Similarity in Direction Relations

Point objects on maps are seldom moved before and after map generalization, so the change of their direction relations can be ignored. In other words, $W_{Direction}$ can be viewed as equal to zero, thus its similarity degree does not need to be further discussed.

Similarity in Metric Distance Relations

Relative local density is a metric distance measure to evaluate the density variations between points before and after generalization. The relative local density of the i^{th} point is defined as:

$$r_i = \frac{R_i}{\sum_{k=1}^{n} R_k} \qquad \text{4-10}$$

where, r_i is the relative local density of the i^{th} point; n is the total number of the points; R_i is the absolute local density of the i^{th} point which is defines as:

$$R_i = \frac{1}{A_i} \qquad \text{4-11}$$

where, A_i is the area of the Voronoi polygon containing the i^{th} point.

This definition for absolute local density is a variation of the one given by Sadahiro (1997: pp52) 'a ratio of the local density at the certain location to the summation of local density over the region' while the definition here is the inverse of the area of the Voronoi polygon of the point. The improvement of the latter definition compared with the former is that the latter can give absolute (and relative) local density of every point while the former cannot. This makes the comparison of density changes point to point before and after generalization possible.

Based on the definition of relative local density, similarity degrees of a point cloud at different scales in metric distance relations can be given as follows:

Suppose that R^l is an array for recording all of the values of the relative density on the map at scale

l; the i^{th} element of R^l is r_i^l. R^m is an array for recording all of the values of the relative density on the map at scale m; the i^{th} element of R^m is r_i^m. To compare the change of relative local density point by point on the two maps, the following strategy is employed:

(1) Check R^l, and delete r_i^l if the i^{th} point on the map at scale l has been deleted;

(2) Sort R^l in increasing order and the elements in R^m are arranged according to the sequences of the values of the corresponding points in R^l; and

(3) To quantify to what extent the two arrays of relative local density are similar, the monotonicity ratio of R^l and R^m is defined:

$$Sim_{A_l,A_m}^{P_{Distance}} = 1 - \frac{n_a}{N_m} \qquad 4\text{-}12$$

where, N_m is the number of points on the map at scale m; n_a is the number of the monotonically abnormal elements in R^m (if the i^{th} element is larger than the $(i+1)^{th}$ in R^m, the i^{th} element is termed monotonically abnormal).

It is obvious that the larger $Sim_{A_l,A_m}^{P_{Distance}}$, the better the relative local density is preserved.

Similarity in Attributes

Importance value is usually used as a comprehensive index to evaluate the change of importance values of a point cloud over the whole region. Mean importance value is defined as

$$\overline{I} = \frac{\sum_{i=1}^{n} I_i}{n} \qquad 4\text{-}13$$

where, $\overline{I}$ is the mean importance value; I_i is the importance value of the i^{th} point; and n is the number of points in the point cloud.

The similarity degree of a point cloud in attributes at two different scales is

$$Sim_{A_l,A_m}^{P_{Attribute}} = \frac{abs\left|\overline{I_l} - \overline{I_m}\right|}{\overline{I_l}} \qquad 4\text{-}14$$

where, $\overline{I_l}$ is the mean importance value of the point clouds at scale l; $\overline{I_m}$ is the mean importance value of the point clouds at scale m; and $abs\left|\overline{I_l} - \overline{I_m}\right|$ is a mathematic absolute value.

Resulting Formula

$$Sim(A_l, A_m) = \frac{W_{Topological}}{w} Sim_{A_l, A_m}^{P_{Topological}} + \frac{W_{Distance}}{w} Sim_{A_l, A_m}^{P_{Distance}} + \frac{W_{Attribute}}{w} Sim_{A_l, A_m}^{P_{Attribute}} \qquad 4\text{-}15$$

where, $w = W_{Topological} + W_{Distance} + W_{Attribute}$.

4.2.2 Model for Parallel Line Clusters

Here, a parallel line cluster specifically refers to contour lines. Apparently, contour lines are approximately parallel curves on maps (Figure 4-6).

Similarity in Topological Relations

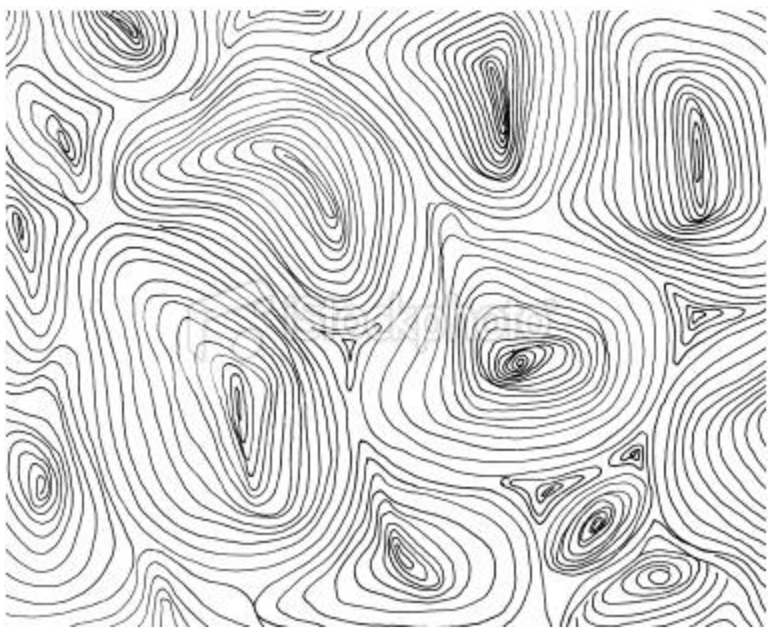
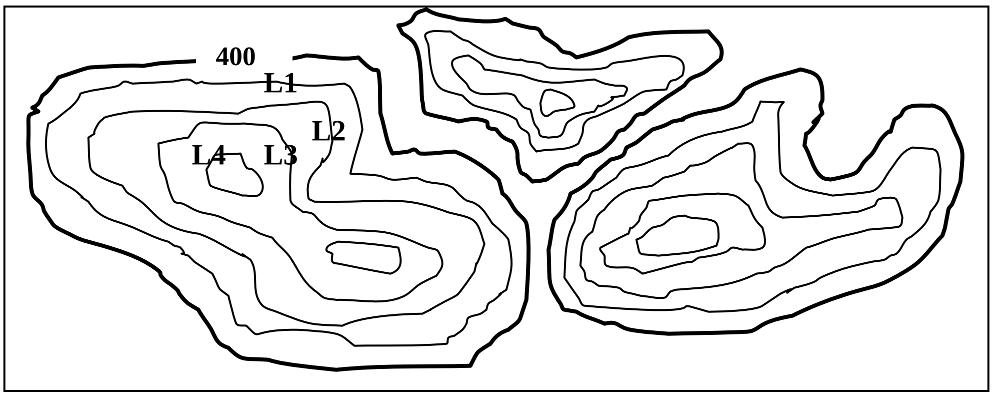


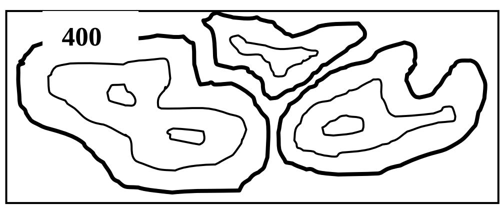
Figure 4-6 A contour map: contours are approximately parallel.

There are totally two types of topological relations between two contour lines, i.e. topologically neighbouring and topologically contained. If the elevations of two topologically adjacent contour lines are equal, they are called topologically neighbouring, and they are "brothers" of each other; otherwise, they are called topologically contained. The contained one calls the other one "father" and otherwise "son". For example, in Figure 4-7(a), the three index contour lines are topologically neighbouring; and the contour line $L1$ and the index contour line marked "400" are topologically contained.

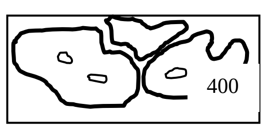
In process of map generalization, if the contour intervals of the original map and the resulting map are different, some contour lines need to be deleted. This inevitably lead to the change of topological relations among contour lines.



(a) Original contours at scale l. The contour interval is 10m.



(b) Generalized contours at scale m. The contour interval is 20m.



(c) Generalized contours at scale k. The contour interval is 40m.

Figure 4-7 Change of topological relations of contour lines in map generalization.

Supposing that L is a contour line at scale l and N_L^l is a value for quantitatively expressing the topological relations of L with other contour lines, it can be calculated by

$$n_L^l = F_L^l + S_L^l + B_L^l \tag{4-16}$$

where, F_L^l, S_L^l and B_L^l are the number of fathers, sons and brothers of L at scale l.

The similarity degree of a cluster of contour lines (say A) at scale l and at scale m can be defined as:

$$Sim_{A_l,A_m}^{P_{Topological}} = \frac{\sum_{j=1}^{N_m} n_j^m}{\sum_{i=1}^{N_l} n_i^l} \tag{4-17}$$

where, $\sum_{i=1}^{N_l} n_i^l$ is the value of the total quantitative topological relations of the contour lines at scale l;

and $\sum_{j=1}^{N_m} n_j^m$ is that at scale m.

Similarity in Direction Relations

It is generally not allowed to move contour lines on maps in the process of map generalization; therefore, direction relations between contour lines are not changed after map generalization. In other words, its weight $W_{Direction}$ can be viewed as equal to zero; hence, its spatial similarity degree does not need to be further discussed.

Similarity in Metric Distance Relations

Metric distance relation of contour lines can be evaluated using the density of contour lines, which is defined as:

$$D_{Contour} = \frac{\sum_{i=1}^{n} L_i}{A} \tag{4-18}$$

where, A is the area occupied by the contour lines; n is the number of the contour lines; and L_i is the length of the i^{th} contour line.

The similarity degree of the contour lines in metric distance relations can be calculated by

$$Sim_{A_l,A_m}^{P_{Distance}} = \frac{D_{Contour}^m}{D_{Contour}^l} \tag{4-19}$$

where, $D_{Contour}^l$ is the density of contour lines on the map at scale l; and $D_{Contour}^m$ is the density of contour lines on the map at scale m.

Similarity in Attributes

Attribute change of contour lines on topographic maps in map generalization generally refers to the change of contours' names and functions (e.g. a 70m index contour line on the map whose contour interval is 10m may become an intermediate contour on the map when its contour interval changes to 20m) caused by the change of contour interval. Hence, this similarity degree can be calculated by

$$Sim_{A_l,A_m}^{P_{Attribute}} = \frac{C_l}{C_m} \qquad 4\text{-}20$$

where, C_l is the contour interval of the original map, and C_k is the contour interval of the generalized map.

Resulting Formula

$$Sim(\mathrm{A}_l,\mathrm{A}_m) = \frac{W_{Topological}}{w} Sim_{A_l,A_m}^{P_{Topological}} + \frac{W_{Distance}}{w} Sim_{A_l,A_m}^{P_{Distance}} + \frac{W_{Attribute}}{w} Sim_{A_l,A_m}^{P_{Attribute}} \qquad 4\text{-}21$$

where, $w = W_{Topological} + W_{Distance} + W_{Attribute}$.

4.2.3 Model for Intersected Line Networks

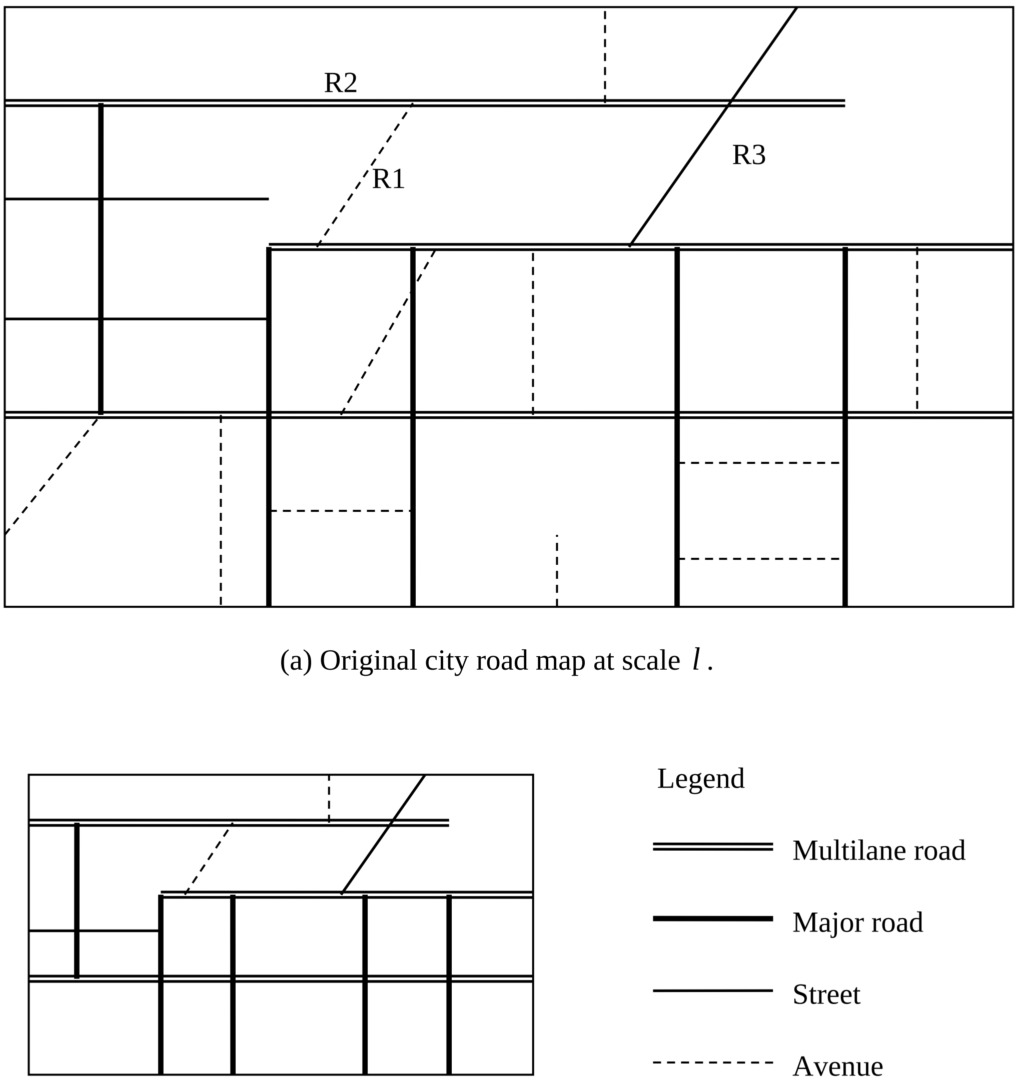
Intersected line networks majorly refer to road networks on maps. The roads in a region usually intersected with each other and form a network (Figure 4-8).

Similarity in Topological Relations

To calculate similarity of two road networks in topological relations, it is necessary to know the difference of topological relations between two road networks at different scales. To achieve this goal, an approach to quantitatively calculate the topological relations of a road network and to calculate the difference of topological relations between two road networks is proposed here.

There are totally two topological relations between two roads on the map, i.e. topologically disjoint (e.g. R1 and R3 in Figure 4-8(a)) and topologically intersected (e.g. R2 and R3 in Figure 4-8(a)). An $n \times n$ matrix A may be used to record the topological relations of a road network containing n roads, and it is assumed that

$A_{ij} = 1$ and $A_{ji} = 1$ if the i^{th} road in the road network is intersected with the j^{th} road; or else $A_{ij} = 0$ and $A_{ji} = 0$.



(a) Original city road map at scale l.

(b) Generalized city map at scale m.

Figure 4-8 A road network at two scales.

Suppose that the matrix for recording the topological relations of the original road network at scale l is B with $N_l \times N_l$ elements, and that for the generalized road network at scale m is C with $N_m \times N_m$ elements, the spatial similarity degree in topology can be calculated by

$$Sim_{A_l, A_m}^{P_{Topological}} = 1 - \frac{D_{Topological}}{N_l \times N_l} \quad \text{4-22}$$

where, $D_{Topological}$ is the topological differences between the two road network. It can be calculated using the following method described in computer language C.

Step 1: let $D_{Topological}$=0;

Step 2: take an element C_{ij} from C starting from $i = 0$ and $j = 0$. C_{ij} denotes the topological relations between the i^{th} road and the j^{th} road on the map at scale m.

Step 3: search B for the element B_{pq} that can also record the topological relations of the i^{th} road and the j^{th} road on the map at scale m.

Step 4: If no $B_{pq} = C_{ij}$ can be found, $D_{Topological}$++.

Step 5: i++; j++.

Step 6: if $i > N_m$ or $j > N_m$, end the procedure; else go to step 3.

Similarity in Direction Relations

The positions of the roads on the original map and on the generalized map are the same, so their direction relations are not changed. Therefore, this similarity is ignored in spatial similarity calculation and does not need to be discussed further.

Similarity in Metric Distance Relations

Similarity of road networks in metric distance relations can be evaluated base on road density, a concept popularly appearing in other communities, such as animal conservation (Butler et al., 2013) and remote sensing (Zhang et al., 2002). Road density (D) is defined as the ratio of the length (L) of the region's total road network to the region's land area (A).

$$D = \frac{L}{A} \quad \text{4-23}$$

Map generalization may lead to the decrease of the number of roads on the map and enlarge the distance among roads, and therefore reduce the road density. Hence, we have

$$Sim_{A_l,A_m}^{P_{Distance}} = \frac{D_m}{D_l} \qquad 4\text{-}24$$

It is obvious that the more roads are deleted, the less D_m is, and the less $Sim_{A_l,A_m}^{P_{Distance}}$ is, which means the similarity degree between the original road network and generalized one decreases with the number of roads in map generalization.

Similarity in Attributes

Similarity in Attributes of road networks can be calculated by a factor named "significance value" which depends on several attributes such as road type, road class, road condition, road grade, etc. To simplify the problem, road class is used to represent the differences of road attributes. Each of the road classes is denoted by a number named class value, and the higher the road class, the greater the class value.

$Sim_{A_l,A_m}^{P_{Attribute}}$ may be calculated by

$$Sim_{A_l,A_m}^{P_{Attribute}} = \frac{\sum_{j=1}^{n_m} L_j^m \times C_j^m}{\sum_{i=1}^{n_l} L_i^l \times C_i^l} \qquad 4\text{-}25$$

where, L_i^l is the length of the i^{th} road in the road network at scale l; C_i^l is the class value of the i^{th} road in the road network at scale l; L_j^m is the length of the j^{th} road in the road network at scale m; and C_j^m is the class value of the j^{th} road in the road network at scale m.

Here, $\sum_{i=1}^{n_l} L_i^l \times C_i^l$ can be viewed as the total class value of the road network at scale l, while $\sum_{j=1}^{n_m} L_j^m \times C_j^m$ is the total class value of the road network at scale m. Thus, $Sim_{A_l,A_m}^{P_{Attribute}}$ represents the percentage of the total class values of the two road networks.

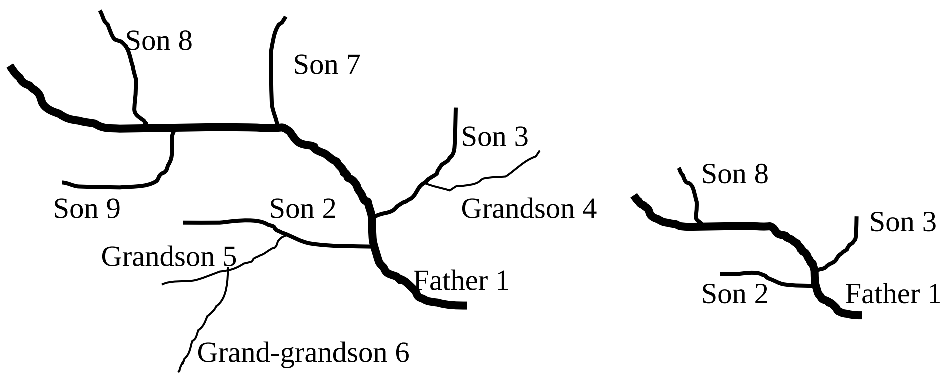
Resulting Formula

$$Sim(A_l, A_m) = \frac{W_{Topological}}{w} Sim_{A_l,A_m}^{P_{Topological}} + \frac{W_{Distance}}{w} Sim_{A_l,A_m}^{P_{Distance}} + \frac{W_{Attribute}}{w} Sim_{A_l,A_m}^{P_{Attribute}} \qquad 4\text{-}26$$

where, $w = W_{Topological} + W_{Distance} + W_{Attribute}$.

4.2.4 Model for Tree-like Networks

The graph of a river basin comprising a main stream and several tributaries is like a tree on the map. Hence, river basins are often studied using the concept of "tree structure", taking their main streams as trunks and tributaries as brunches (La Barbera & Rosso, 1989; Ross, 1999). The tree structures are called "tree-like networks" in this section.



(a) (b)

Figure 4-9 A river basin.

(a) original tree-like network; and (b) generalized tree-like network.

Similarity in Topological Relations

The main stream and its brunches of a tree-like network may be called "root" and "leaves" in computer science. They can also be called "father", "sons", "grandsons" and "grand-grandsons", etc. to facilitate our discussion, the latter is adopted in this section. Figure 4-9(a) presents such a tree-like network. Their relations can be recorded in a tree data structure (Knuth, 1997) in Figure 4-10, which shows their descendent relations clearly. The topological relations of a tree-like network are mainly descendent relations.

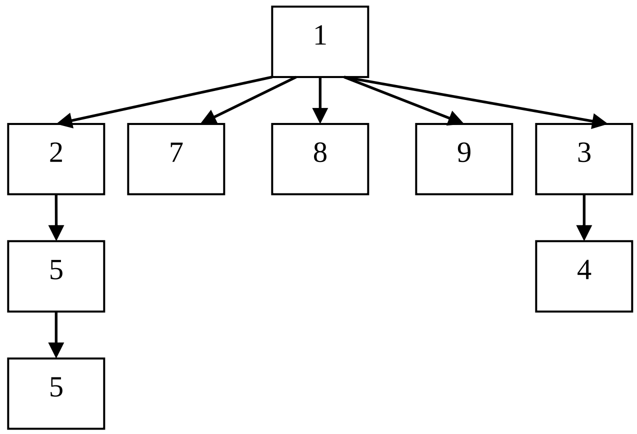


Figure 4-10 Tree data structure of the network for Figure 4-9(a).

If the tree-like network is generalized, some branches is probably deleted, which changes the topological relations among the father and his children. Apparently, $Sim_{A_l,A_m}^{P_{Topological}}$ depends on the number of the topological relations changes taken place in the process of map generalization.

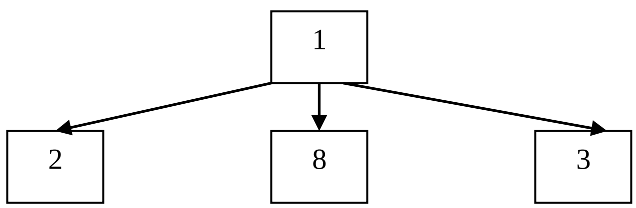


Figure 4-11 Tree data structure of the network for Figure 4-13(b).

$$Sim_{A_l,A_m}^{P_{Topological}} = \frac{N_{Topological}^{m}}{N_{Topological}^{l}} \quad 4\text{-}27$$

where, $N_{Topological}^{l}$ is the total number of topological relations of the tree-like network at scale l (if the main stream and a tributary or two tributaries are father-son relations, there exists a topological relation in the tree-like network); and $N_{Topological}^{m}$ is the total number of topological relations of the tree-like network at scale m.

For example, Figure 4-10 and Figure 4-11 show the father-son relations of the two tree-like networks in Figure 4-9.

$$Sim_{A_l,A_m}^{P_{Topological}} = \frac{3}{8} = 37.5\%$$

Similarity in Direction Relations

Rivers on topographic maps are spatial objects with most high accuracy, and they are essential in spatial positioning, and their positions are not allowed to be modified. Hence, no direction relations are changed among the components of a river basin, and their spatial similarity degrees in direction relations do not need to be discussed further.

Similarity in Metric Distance Relations

Density of river network is often used to represent metric distance relations of a river basin.

$$D_{river} = \frac{L_{river}}{A_{river}} \quad \text{4-28}$$

where, D_{river} is the Density of the river network; A_{river} is the area of the river basin; and L_{river} is the total length of the main stream and tributaries of the river network.

By the definition of density of river network, the spatial similarity degree of the river network in metric distance relations can be obtained.

$$Sim_{A_l,A_m}^{P_{Distance}} = \frac{D_{river}^{m}}{D_{river}^{l}} \quad \text{4-29}$$

where, D_{river}^{l} is the density of the original river network at scale l ; D_{river}^{m} is the density of the generalized river network at scale m .

Similarity in Attributes

Although many geometric and thematic attributes are used in river network generalization, stream order is the most popularly accepted one. Stream order is a comprehensive index calculated by a combination of several attributes such as length of the river, width of the river, and level of the river etc. A number of encoding rules have been proposed for calculating stream order (Figure 4-12), e.g. Horton, Strahler, Shreve, and Branch (Horton, 1945; de Serres and Roy, 1990; Thomson and Brooks, 2002; Zhang, 2006). Each of the rules has its advantages and disadvantages, which is not necessary to be further discussed in this section. Here, the Branch rule proposed by Zhang (2006) is utilized and it calculates stream order by

$$S_{order} = n + 1 \qquad \text{4-30}$$

where, n is the totally number of children the stream owns.

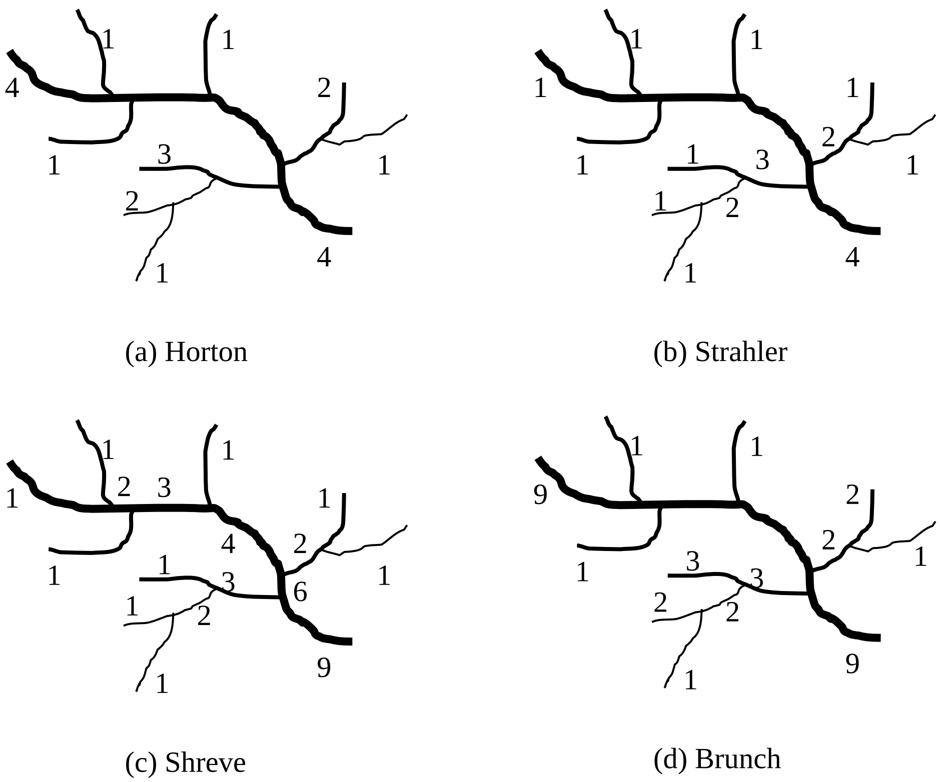


Figure 4-12 Four encoding rules for ordering streams.

Both Figures 4-12(d) and 4-13 illustrate the principle of the Branch rule.

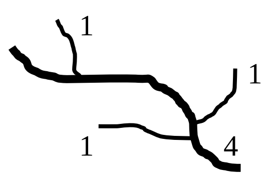


Figure 4-13 Branch encoding for the generalized river network in Figure 4-12(d).

The similarity in attributes of a river network at two scales can be obtained by a comparison of the attribute changes between the original river network at scale l and the generalized river network at scale m.

$$Sim_{A_l,A_m}^{P_{Attribute}} = \frac{\sum_{i=1}^{n} S_i^m}{\sum_{i=1}^{n} S_i^l} \qquad 4\text{-}31$$

where, $\sum_{i=1}^{n} S_i^m$ is the sum of the stream order in the generalized river network; $\sum_{i=1}^{n} S_i^l$ is the sum of the stream order of the streams in the original river network but also existing in the generalized river network.

For example, in Figures 4-12(d) and 4-13,

$$Sim_{A_l,A_m}^{P_{Attribute}} = \frac{7}{15} \approx 46.7\%$$

Resulting Formula

$$Sim(\mathrm{A}_l, \mathrm{A}_m) = \frac{W_{Topological}}{w} Sim_{A_l,A_m}^{P_{Topological}} + \frac{W_{Distance}}{w} Sim_{A_l,A_m}^{P_{Distance}} + \frac{W_{Attribute}}{w} Sim_{A_l,A_m}^{P_{Attribute}} \qquad 4\text{-}32$$

where, $w = W_{Topological} + W_{Distance} + W_{Attribute}$.

4.2.5 Model for Discrete Polygon Groups

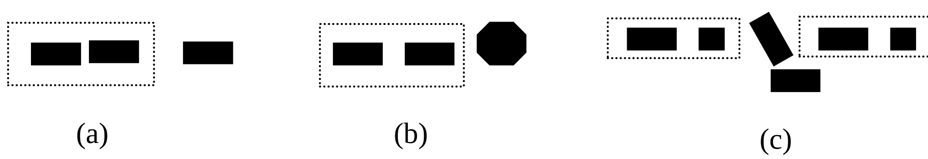


Figure 4-14 Settlements grouping.

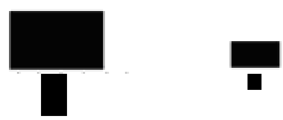
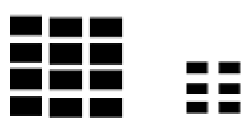
(a) proximity: two close settlements form a group; (b) similarity: only the two buildings of same size and shape form a group; and (c) common direction: only those objects that are arranged in the same directions form a group. Settlements in each of the dotted rectangles form a group.

A number of features on maps are represented using discrete polygonal symbols such as settlements, green lands, ponds, islands, etc. In map generalization, such kinds of polygonal symbols are often clustered and processed taking group as unit.

As one of the most popular features on topographic maps, settlement group is selected as a representative to discuss the model for calculating similarity degrees. Indeed, settlements are regarded as groups in automated map generalization in past research work (Bader and Weibel, 1997; Ruas, 1998; Boffet and Rocca Serra, 2001; Regnauld, 2001; Christophe and Ruas, 2002; Rainsford and Mackaness, 2002; Li et al., 2004; Bader et al., 2005), which is theoretically supported by a number of Gestalt principles (Palmer, 1992; Rock, 1996) such as proximity, similarity, and common directions/orientation (Figure 4-14).

Similarity in Topological Relations

Table 4-1 Operations and topological changes in settlement group generalization.

Operations	Examples	Topological change
Aggregation		Yes
Collapse		No
Displacement		No
Exaggeration		No
Elimination		Yes
Simplification		No
Typification		Yes

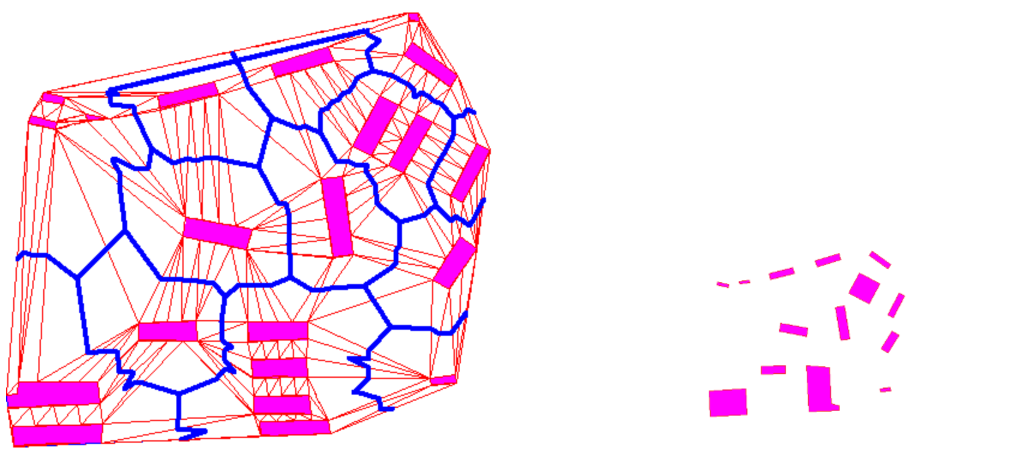
A number of operations can be exerted to generalize settlement groups, including aggregation, collapse, displacement, exaggeration, elimination, simplification and typification. Some of them cause changes of topological relations in the process of settlement group generalization (Table 4-1), while some other do not.

It is necessary to compare the topological relations of a settlement group before and after map generalization to obtain the topological change so that spatial similarity in topological relations can be calculated.

Apparently, every two settlements in the original settlement group are topologically separated; hence there are $N_l \times (N_l - 1)$ topologically disjoint relations in the original settlement group. The number of topologically disjoint relations (i.e. $N_m \times (N_m - 1)$ in the generalized settlement group depends on the number of the settlements in the generalized group. The difference of disjoint relations between the two settlement groups reveals the changes of similarity degree.

$$Sim_{A_l, A_m}^{P_{Topological}} = \frac{N_m \times (N_m - 1)}{N_l \times (N_l - 1)} \qquad 4\text{-}33$$

Taking Figure 4-15 as an example, their similarity degree is



(a) Original group with 21 settlements (b) Generalized group with 14 settlements

Figure 4-15 Topological similarity of a settlement group in map generalization.

$$Sim_{A_l, A_m}^{P_{Topological}} = \frac{14 \times 13}{21 \times 20} = 65\%$$

Similarity in Direction Relations

Direction relations among settlements are possibly be changed in the process of map generalization due to operations such as aggregation, displacement and elimination. A natural thought to calculate $Sim_{A_l, A_m}^{P_{Direction}}$ is to record and compare the direction relations of the settlement group before and after map generalization.

Direction relations between two settlements can be described using direction group (Yan, Chu and Li et al., 2006). Direction group is based on a fact that people often describe directions between two objects using multiple directions but not a single direction (Peuquet & Zhan, 1987; Hong, 1994; Goyal, 2000); therefore description of direction relations should use multiple directions, i.e. direction group. A direction group consists of two components: the azimuths of the normals of direction Voronoi Diagram (DVD) edges between two objects and the corresponding weights of the azimuths.

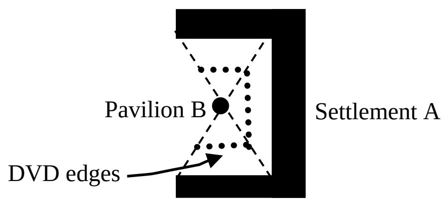


Figure 4-16 An example of direction group.

40% of B is to the west of A; and 30% of B is to the north of A; and 30% of B is to the south of A.

For example in Figure 4-16, the direction relations between the pavilion B and the settlement A can be expressed as $Dir(A,B)=\{<N,30\%>,<S,30\%>,<W,40\%>\}$ by means of direction group.

To record direction relations among settlements, two matrixes are defined: $N_l \times N_l$ matrix B_l is for the original settlement group at scale l, and $N_m \times N_m$ matrix C_m is for the generalized settlement group at scale m. Each element in B_l and C_m is a direction group for recording direction relations between two settlements.

It is necessary to calculate the intersection of B_l and C_m in order to obtain $Sim_{A_l,A_m}^{P_{Direction}}$. The basic idea is: take an element b_{ij} from B_l. Here, b_{ij} represents the direction relations between the i^{th} settlement and the j^{th} settlement in the original settlement group. Then search for C_m to find an element, say c_{kp}, that totally or partially represents the direction relations between the i^{th} and the j^{th} generalized settlements. Compare c_{kp} and b_{ij} to get their intersection. In the eight-direction system, if c_{kp} and b_{ij} are totally same, their intersection value is 8. Otherwise, their intersection value is the number of the common directions. After comparing each element in B_l with the elements in C_m, the total intersection value n_m can be obtained. This value denotes the common direction relations between the original settlement group and the generalized settlement group. Hence, we have

$$Sim_{A_l,A_m}^{P_{Direction}} \frac{n_m}{8N_l(N_l-1)} \quad \text{4-34}$$

where, $8N_l(N_l-1)$ is the total direction relations among the settlements in the original settlement group.

Similarity in Metric Distance Relations

$$Sim_{A_l,A_m}^{P_{Distance}} = 1 - \frac{abs(\overline{D_l} - \overline{D_m})}{\overline{D_l}} \quad \text{4-35}$$

where, $\overline{D_l}$ is the mean settlement density of the original settlement group; and $\overline{D_m}$ is the mean settlement density of the generalized settlement group.

The mean settlement density of a settlement group ($\overline{D}$) may be calculated by

$$\overline{D} = \frac{\sum_{i=1}^{n} A_i}{S} \quad \text{4-36}$$

where, S is the total area of the region occupied by the settlement group, comprising the area of the settlements and the area of common space; n is the number of the settlement in the group; and $\sum_{i=1}^{n} A_i$ is the total area of the settlements in the group.

Similarity in Attributes

Settlement attributes (e.g. height, building material, etc.) are seldom taken into consideration in map generalization; thus, they have little effect to similarity relations. In other words, the attribute similarity degree of a settlement group before and after map generalization does not change, and does not need to be further investigated.

Resulting Formula

$$Sim(A_l, A_m) = \frac{W_{Topological}}{w} Sim_{A_l,A_m}^{P_{Topological}} + \frac{W_{Direction}}{w} Sim_{A_l,A_m}^{Direction} + \frac{W_{Distance}}{w} Sim_{A_l,A_m}^{P_{Distance}} \quad \text{4-37}$$

where, $w = W_{Topological} + W_{Direction} + W_{Distance}$.

4.2.6 Model for Connected Polygon Groups

Categorical maps consist of connected polygons. These categorical spatial patterns are typically the result of mapping, classification, or modeling exercises that produce maps of land cover or some other categorical representation of a landscape (Boots and Csillag, 2006; Remmel and Fortin, 2013). Here it is selected as a representative for addressing similarity relations between connected polygon groups.

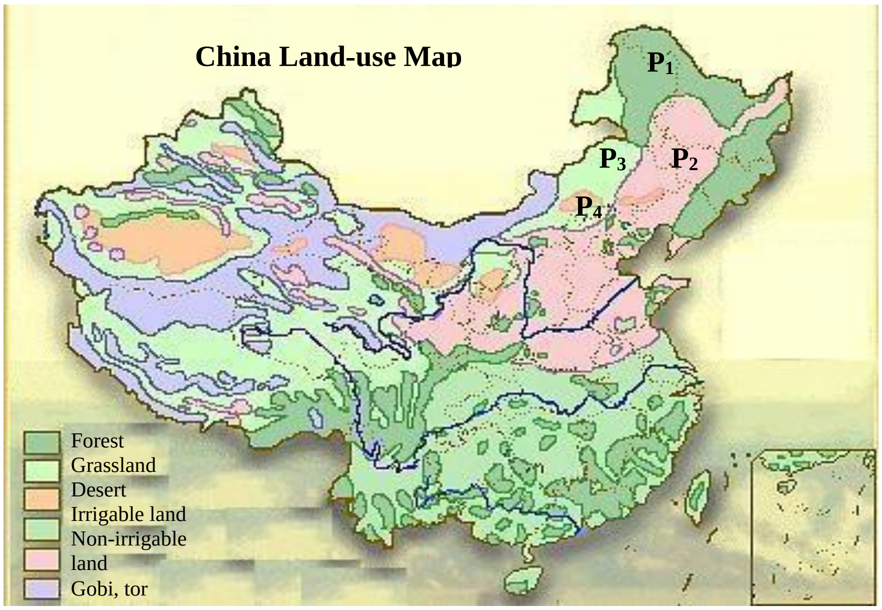


Figure 4-17 A land-use map consists of connected polygons.

Similarity in Topological Relations

There exist three types of topological relations between polygons on categorical maps, i.e. topologically disjoint, topologically adjacent and topologically contained. “Inside” is an inverse of “topologically contained”, therefore they may be viewed as the same type. For example, in Figure 4-17 (Revised from http://ishare.iask.sina.com.cn/f/13293700.html), P_1 and P_4 are disjoint; P_1 and P_2 are adjacent; and P_4 is contained by P_3.

To get $Sim_{A_1,A_m}^{P_{Topological}}$, it is necessary to record the topological relations of the two connected polygon groups and then compare the two maps before and after generalization and get the intersection of their topological relations .

Suppose that a $N_l \times N_l$ matrix B is used to record topological relations of the original connected polygon group, element b_{ij} in B records the topological relations between the i^{th} polygon and the j^{th} polygon; and a $N_m \times N_m$ matrix C in the same way is used to record topological relations of the generalized connected polygon group.

The following algorithm can be used to calculate the intersection of B and C:

Step 1: let $N_{same} = 0$.

Step 2: take the first element, say b_{ij}, from B.

Step 3: traverse C from the first element to the last element and compare each element of C with b_{ij}. If there exists an element in C representing the topological relations of the same objects in the original group and the topological relations are the same, $N_{same}++$.

Step 4: if b_{ij} is not the last element of B, take the next element from B and still name it b_{ij}, and go to Step 3.

Step 5: end the procedure.

Based on this calculation, we have,

$$Sim_{A_l, A_m}^{P_{Topological}} = \frac{N_{same}}{N_l \times N_l} \quad 4\text{-}38$$

Similarity in Direction Relations

No direction between polygons is changed, because the polygons are not moved before and after map generalization. Thus, similarity in direction relations is usually ignored.

Similarity in Metric Distance Relations

No operations in the process of map generalization can change the distance relations between connected polygons. Thus, similarity in metric distance relations is usually ignored.

Similarity in Attributes

To get the similarity in attributes, it is a feasible way to overlap the original connected polygon group with the corresponding one after map generalization. Indeed, the literature regarding spatial analysis is crowded with the ideas that measure, describe, or compare categorical spatial patterns (Uuemaa et al. 2009) using vector-based (Milenkovic, 1998; Liu, 2002; Sadahiro, 2012) and raster-based approaches (Gustafson, 1998; Hagen, 2003; Csillag and Boots, 2004). Here, a raster-based approach is employ to calculate the intersection of the original polygons and generalized polygons.

Suppose that an $N \times N$ regular grid network is used to rasterize the two polygon groups, respectively, their intersection (i.e. the number of the grids with same attribute in the two polygon groups) is $N_{intersection}$. We have

$$Sim^{P_{Attribute}}_{A_l,A_m} \frac{N_{intersection}}{N \times N} \quad 4\text{-}39$$

A problem that should be noticed is the value of N, because the great N is, the more accurate and the lower efficient the rasterization is and vice versa. Here, $N = \sqrt{A}$. A is the least polygon in the original polygon group.

Resulting Formula

$$Sim(\mathrm{A}_l, \mathrm{A}_m) = \frac{W_{Topological}}{w} Sim^{P_{Topological}}_{A_l,A_m} + \frac{W_{Attribute}}{w} Sim^{P_{Attribute}}_{A_l,A_m} \quad 4\text{-}40$$

where, $w = W_{Topological} + W_{Attribute}$.

4.3 Model for Calculating Spatial Similarity Degrees between Maps

Previous sections of this chapter view a map as a combination of a number of separated feature layers (i.e. individual objects and object groups), and propose a series of models for calculating spatial similarity degrees of each of the feature layers before and after map generalization. No doubt, these models can be used to assess the change of similarity degrees of individual map feature layers. Nevertheless, it is usually necessary to overlap the generalized features layers to form a resulting map before they are put into practical use. Hence, problems arise here: how can the similarity degree between the original map and the resulting map obtained? Can it calculated by the models for calculating similarity degrees of individual feature layers at different scales?

This section will try to solve the problems by integrating the previous models for calculating similarity degrees of individual feature layers to form a comprehensive model. This new model will be a vector-based model, because those ones for individual feature layers are vector-based, too.

Because it seems considerably difficult to construct a generic model for all types of maps, only topographic map is taken as a representative for addressing the idea of the new integrated model here. A detailed description of a topographic map may be given first.

Suppose that there is a topographic map T_l at scale l, it consists of N feature layers. The numbers of objects in the N feature layer are $n_1, n_2, ..., n_N$, respectively. A generalized counterpart of T_l is the topographic map T_m at scale m. It consists of M feature layers. The numbers of objects of the M feature layer are $m_1, \mathrm{m}_2, ..., \mathrm{m}_M$, respectively. $N_l = n_1 + n_2 + ... + n_N$; $N_m = m_1 + m_2 + ... + m_M$.

The four types of similarity relations between two topographic maps need to be considered, i.e. topological similarity, direction similarity, metric distance similarity, and attribute similarity. The degrees of the four types of similarity relations are denoted by $Sim_{T_l,\mathrm{T}_m}^{Topological}$, $Sim_{T_l,\mathrm{T}_m}^{Direction}$, $Sim_{T_l,\mathrm{T}_m}^{Distance}$, and $Sim_{T_l,\mathrm{T}_m}^{Attribute}$, respectively.

The similarity between the two maps ($Sim_{T_l,\mathrm{T}_m}^{Map}$) is:

$$Sim_{T_l,\mathrm{T}_m}^{Map} = w_1 Sim_{T_l,\mathrm{T}_m}^{Topological} + w_2 Sim_{T_l,\mathrm{T}_m}^{Direction} + w_3 Sim_{T_l,\mathrm{T}_m}^{Distance} + w_4 Sim_{T_l,\mathrm{T}_m}^{Attribute} \quad \text{4-41}$$

where, w_1 , w_2 , w_3 and w_4 are the weights of $Sim_{T_l,\mathrm{T}_m}^{Topological}$, $Sim_{T_l,\mathrm{T}_m}^{Direction}$, $Sim_{T_l,\mathrm{T}_m}^{Distance}$, and $Sim_{T_l,\mathrm{T}_m}^{Attribute}$, respectively. They are obtained by psychological tests which have been addressed in Chapter 3. Thus, the following sections will focuses on the calculation of the four types of similarity degrees.

4.3.1 Similarity in Topological Relations

It is necessary to compute, record, and compare the topological relations between the two maps to get their intersection for the purpose of calculating their topological similarity degree.

The methods for computing topological relations among map objects (including points, lines and polygons) have crowded literature for decades (Egenhofer et al., 1994; Clementini et al., 1994; Bjørke, 2004; Du et al., 2008; Formica et al., 2013), which provides sufficient theoretical and technical supports for obtaining topological relations.

Two matrixes are defined to record topological relations among objects on the two maps: $N_l \times N_l$ matrix B_l is for the map at scale l, and $N_m \times N_m$ matrix C_m is for the generalized map at scale m. Each element in B_l and C_m is a positive integer for indicating a topological relation between two objects. Their corresponding relations are listed in Table 4-2.

An algorithm is proposed to compute the intersection of B_l and C_m .

Step 1: let $N_{same} = 0$.

Step 2: take the first element, say b_{ij}, from B_l.

Step 3: traverse C_m from the first element to the last element and compare each element of C_m with b_{ij}. If there exists an element in C_m representing the topological relations of the same objects in the original map and the topological relations are the same, $N_{same}++$.

Step 4: if b_{ij} is not the last element of B_l, take the next element from B_l and still name it b_{ij}, and go to Step 3.

Step 5: end the procedure.

Table 4-2 Integers for denoting topological relations

Topological relations	**Recorded values**
Disjoint	1
Meet	2
Overlap/intersect	3
Contain & meet	4
Contain	5
Equal	6

Base on this calculation, we have

$$Sim_{T_l,T_m}^{Topological} = \frac{N_{same}}{N_l \times N_l} \quad \text{4-42}$$

4.3.2 Similarity in Direction Relations

Directional similarity degree between two maps depends on the change of the direction relations between the two maps. Therefore, the direction relations among the objects of the map at scale l and that at scale m should be calculated, recorded and compared.

Direction group (Yan et al., 2006) is employed to describe direction relations between arbitrary two objects. To record direction relations among objects on the two maps, two matrixes are defined: $N_l \times N_l$ matrix B_l is for the map at scale l, and $N_m \times N_m$ matrix C_m is for the generalized map at scale m. Each element in B_l and C_m is a direction group for recording direction relations between two settlements. The eight-direction system is used here.

The basic idea for comparing B_l and C_m is: take an element b_{ij} from B_l. Here, b_{ij} represents the direction relations between the i^{th} object and the j^{th} object on the original map at scale l. Then search for C_m to find an element, say c_{kp}, that totally or partially represents the direction relations between generalized the i^{th} object and the j^{th} object on the map at scale m. Compare c_{kp} and b_{ij} to get their intersection, i.e. the number of same directions. In the eight-direction system, if c_{kp} and b_{ij} are totally same, their intersection value is 8. After comparing each element in B_l with the elements in C_m, the total intersection value $N_{direction}^{intersection}$ can be achieved. This value denotes the common direction relations between the original map and the generalized map, by which the direction similarity between the two maps can be obtained.

$$Sim_{T_l,T_m}^{Direction} = \frac{N_{direction}^{intersection}}{8N_l(N_l - 1)} \qquad 4\text{-}43$$

4.3.3 Similarity in Metric Distance Relations

Metric distance relations of a topographic map can be described using the Voronoi Diagram, because Voronoi Diagram has been regarded as an ideal tool in tessellation of 2-dimensional map spaces (Aurenhammer, 1991) and description of spatial relations (Chen et al., 2001). Concept of the Voronoi Diagram has already been extended from the tessellation of point clusters to that of spatial objects, including points, lines, and polygons. Figure 4-18 illustrates the principle of the Voronoi Diagram for spatial objects (Li et al., 1999).

For point objects, they may be put into a group and regarded as a point cloud. Formulae 4-11, 4-12 and 4-13 can be used to calculate and compare the values of relative local density of all points at scales of l and m, and then get the similarity degree of the point objects ($Sim_{T_l,T_m}^{Distance,\mathrm{Point}}$) before and after generalization.

All linear objects can be put together and regarded as a line cluster. The three pairs of formulae, i.e. formulae 4-19 and 4-20, Formulae 4-24 and 4-25, and Formulae 4-28 and 4-29, that express the same idea “distance relations of linear objects can be described using density of linear objects”, can be employed to calculate the similarity degree of the linear objects ($Sim_{T_l,T_m}^{Distance,\mathrm{Linear}}$) before and after generalization. It should be noted that the area occupied by each of the linear objects is the area of its Voronoi polygon (Li et al., 1999).

For areal objects, they can be regarded as discrete polygon group though they may be connected polygons or discrete polygons. Formulae 4-35 and 4-36 may be used to calculate the similarity degree of the area objects ($Sim_{T_l,\mathrm{T}_m}^{Distance,\mathrm{Areal}}$) before and after map generalization. The area occupied by each of the polygons is the area of its Voronoi polygon. For connected polygons, their density, according to Formula 4-35, should be equal before and after map generalization.

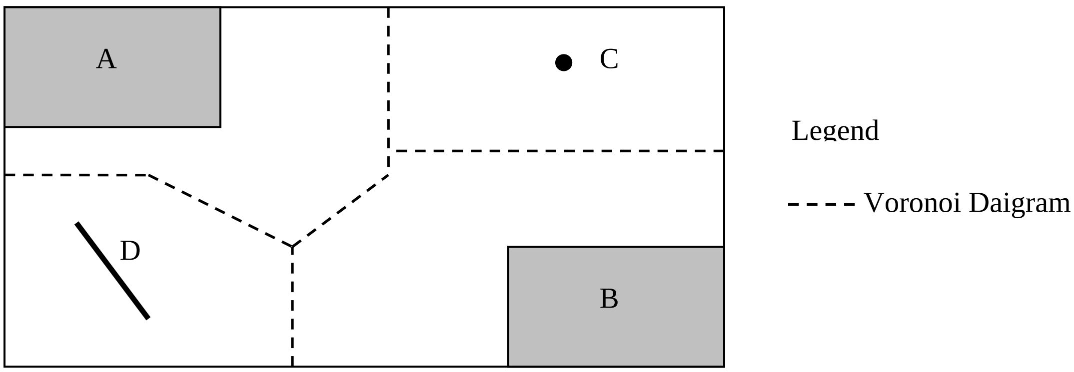


Figure 4-18 Voronoi Diagram of spatial objects.

The distance similarity degree of the map before and after generalization can be expressed as:

$$Sim_{T_l,\mathrm{T}_m}^{Distance} = \frac{N_l^{Point}}{N_l} Sim_{T_l,\mathrm{T}_m}^{Distance,\mathrm{Point}} + \frac{N_l^{Linear}}{N_l} Sim_{T_l,\mathrm{T}_m}^{Distance,\mathrm{Linear}} + \frac{N_l^{Areal}}{N_l} Sim_{T_l,\mathrm{T}_m}^{Distance,\mathrm{Areal}} \qquad 4\text{-}44$$

where, N_l^{Point} is the number of the points on the map at scale l; N_l^{Linear} is the number of the linear objects on the map at scale l; and N_l^{Areal} is the number of the polygons on the map at scale l.

$\frac{N_l^{Point}}{N_l}$, $\frac{N_l^{Linear}}{N_l}$ and $\frac{N_l^{Areal}}{N_l}$ are the weights. The greater the number of a type of objects on the map, the greater the weight value.

4.3.4 Similarity in Attributes

Topographic maps show physical and human-made features of the Earth and regard all of the feature layers as the same importance by default (Harvey, 1980; Barber, 2005). Hence, the attribute weights for all of the feature layers are equal.

$$Sim_{T_l,T_m}^{Attribute} = \frac{\sum_{i=1}^{N} Sim_{T_l,T_m}^{Layer_i^{Attribute}}}{N} \qquad 4\text{-}45$$

where, $Sim_{T_l,T_m}^{Attribute}$ is the attributes similarity degree between the two topographic maps; $Sim_{T_l,T_m}^{Layer_i^{Attribute}}$ is the attributes similarity degree between two the i^{th} map feature layers of the two topographic maps.

$Sim_{T_l,T_m}^{Layer_i^{Attribute}}$ may be calculated by a formula in the previous sections of this chapter. The formula can be decided according to the type of the features.

4.4 Chapter Summary

This chapter aims at proposing the models for calculating spatial similarity degrees of various types of objects at different map scales. Totally ten types of objects are concerned, i.e. (1) individual points, (2) individual lines, (3) individual polygons, (4) point clouds, (5) parallel lines clusters, (6) intersected line networks, (7) tree-like networks, (8) discrete polygon groups, (9) connected polygon groups, and (10) maps, and the model for calculating spatial similarity degrees between each of the ten types of objects before and after map generalization are proposed. All of the proposed models are based on vector data.

Chapter 5 Model Validations

People are accustomed to taking spatial similarity relation as a qualitative factor to describe the geographic space (Guo, 1997); therefore whether the quantitative values of spatial similarity relations calculated by the proposed models coincide with human's spatial cognition is worth validating so that the questions like "Are the similarity degrees calculated by the models the same as that of my recognition?" and "Are the calculated similarity degrees acceptable by most people?" can be answered. For this purpose, this chapter focuses on validating the ten models proposed in Chapter 4, aiming at proving that the models are acceptable to majority of people.

5.1 General Approaches to Model Validation

Correctness of models is often addressed through model verification and validation (Schlesinger, 1979; Carson, 2002; Banks et al., 2010). Model verification is defined as "ensuring that the computer program of the computerized model and its implementation are correct" (Sargent, 2011). Model validation is usually defined to mean "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model" (Naylor and Finger, 1967; Schlesinger et al., 1979) and is the definition used here. This study assumes that the correctness of the computer program can be ensured and therefore setting model verification aside. It only emphasizes on the validation of the proposed models for calculating spatial similarity degrees in map multi-scale spaces.

Generally, a model is developed for a specific purpose or application, and its validity is determined with respect to that purpose and application. If the purpose is to solve a variety of problems, the validity of the model should be determined with respect to solving all of those problems. Hence, numerous experimental conditions are required to define the domain of the applications of the model. A model is viewed as valid if for each of the experimental conditions its accuracy is always within its acceptable range. Usually, the acceptable range of accuracy for each model should be specified prior to starting the development of the model or very early in the model development process.

It is often too costly and time consuming to determine that a model is absolutely valid over the complete domain of its intended applicability. Instead, tests and evaluations are conducted until sufficient confidence is obtained that a model can be considered valid for its intended application (Sargent, 1982; 1984). If a test determines that a model does not have sufficient accuracy for any one of the sets of experimental conditions, then the model is invalid. However, determining that a model has sufficient accuracy for numerous experimental conditions does not guarantee that a model is valid everywhere in its applicable domain.

There are four basic approaches for determining whether a model is valid (Sargent, 2010; 2011). Each of the approaches requires conducting model validation as a part of the model development process.

Above all, a frequently used approach is for the model development team itself to make the decision as to whether a simulation model is valid. A subjective decision is made based on the results of the various tests and evaluations conducted as part of the model development process.

Second, if the size of the simulation team developing the model is small, a better approach is to have the model users involved with the model development team in deciding the validity of the simulation model, i.e. the focus of determining the validity of the simulation model moves from the model developers to the model users.

Third, a third (independent) party can be used to decide whether the simulation model is valid. The third party is independent of both the simulation development team and the model sponsors/users. The approach should be used when developing large-scale simulation models, whose developments usually involve several teams. The third party needs to have a thorough understanding of the intended purpose(s) of the simulation model.

Last, a scoring model can be employed to decide whether a model is valid (Balci, 1989; Gass, 1983; 1993). Scores are determined subjectively. A simulation model is considered valid if its overall and category scores are greater than some passing score(s). This approach is seldom used in practice, because the passing scores are usually decided in subjective way, and the scores may cause over confidence in a model, or the scores can even be used to argue that one model is better than another.

In sum, model validation is critical in the development of a simulation model. Nevertheless, no specific approach can easily be applied to determine the “correctness” of all models, and no algorithm exists to determine what techniques or procedures to use. Every simulation presents a new and unique challenge to the model development team.

5.2 Strategies for Validating The New Models

Each of the proposed models in Chapter 4 is a simulation of cartographers’ similarity judgment process regarding corresponding map features or maps in map generalization. As is well known, in human being's cognition, the spatial relations are typically qualitative, approximate, categorical, or topological rather than metric or analog. They may even be incoherent, that is, people may hold beliefs that cannot be reconciled in canonical three-dimensional space (Tversky et al., 2006). On the other hand, these models, if proved correct, can substitute for cartographers to judge spatial similarity in map generalization so that full automation of map generalization can be implemented; thus, whether the models have sufficient accuracy is of great importance. Thus, three strategies are employed to form a comprehensive approach to ensure the validity of the newly proposed models due to the above reasons. They include theoretical justifiability, third part involvement, and experts' participation.

Strategy 1: theoretical justifiability

The models for calculating similarity degrees in this study are for map generalization and aim at automating the algorithms used in generalizing various map layers and maps. Hence, this study firstly classifies the research object into 10 categories that can be directly operated by the algorithms (i.e. individual points, individual lines, individual polygons, point clouds, parallel lines clusters, intersected line networks, tree-like networks, discrete polygon groups, connected polygon groups, and maps). Then the 10 models are constructed in accordance with the 10 categories of objects. This ensures that all potential algorithms that use spatial similarity degrees in map generalization have been taken into consideration.

To ensure the difference between the similarity degrees calculated by the new models and the ones judged by human beings can be as small as possible, all of the major factors that affect human's spatial similarity judgments in map generalization have been taken into consideration to construct the new models. Cartographers consider spatial relations and non-spatial relations of spatial objects in map generalization. The former includes topological relations, direction relations, and distance relations; while the latter refers to attributes of spatial objects. To simulate cartographers' thinking process accurately, the four factors (i.e. the three spatial relations and one non-spatial relation) are all used in the models. This, though cannot ensure the simulation models match cartographers' judgments well, provides a theoretically plausible way for calculating spatial similarity degrees.

Strategy 2: third party involvement

To obtain the weights of topological relations, direction relations, distance relations, and attributes of spatial objects in human's spatial similarity judgments, a number of subjects are invited and sample data are distributed to them to know the weights of the four factors. The average values of these weights are directly used in the new models.

Strategy 3: experts' participation

Now that the proposed new models are used as substitutions of cartographers (i.e. the experts in map generalization), it is justifiable to survey a number of experienced cartographers by psychological experiments to know to what extent they agree with the results calculated by the new models.

Strategies 1 and 2 have been used in the construction of the new models and presented in previous sections.

The following sections introduce Strategy 3, i.e. using psychological experiments to test the validity of the new models. The design of the psychological experiments is presented first; then a number of samples are shown and the psychological surveys are implemented. Finally, the data collected from the experiments are analyzed and discussed, and some conclusions are drawn.

5.3 Psychological Experiment Design

◆ Basic information of the experiments

Time: October 20, 2013.

Place: Lanzhou Jiaotong University, P.R. China.

Subjects: 50 students at undergraduate or graduate level, 24 female and 26 male. Their ages range from 17 to 27. Each of the subjects has least six months experience in making maps. All subjects are majoring or have majored in geography and related communities, including 16 in geographic information science, 22 in cartography, 9 in surveying, 3 in geography.

An advertisement is posted in the webpages of Lanzhou Jiaotong University and Gansu Map Institute for the purpose of recruiting enough subjects who are experienced in mapping and/or geographic information systems.

◆ Goal of the experiments

(1) to know the confidence level of the new models; and

(2) to know if the models can be used in automated map generalization.

◆ Procedure of the experiments

Step 1: Preparation of samples

Totally 10 types of objects are prepared, i.e. (1) individual points, (2) individual lines, (3) individual polygons, (4) point clouds, (5) parallel lines clusters, (6) intersected line networks, (7) tree-like networks, (8) discrete polygon groups, (9) connected polygon groups, and (10) maps.

For each type of the objects, at least three samples, either real or analogous, should be prepared. Each sample consists of the original objects at a larger scale and five counterparts of generalized objects at smaller scales; and the similarity degree between the original objects and each counterpart of the generalized objects calculated by the corresponding new models is given.

To ensure that each sample is a good representative of the corresponding type of the ten object groups and to ensure that the original map/object group can be correctly generalized, four experienced cartographers are invited to provide samples and generalize the maps.

Step 2: Psychological experiments

Each of the subjects are invited to participate in the experiments, respectively. The samples are printed and distributed to each of the subjects one by one. After getting a sample (e.g. Figure 5-35) and five decimals (e.g. Figure 5-36) for describing the similarity degrees, the subject is required to evaluate the similarity degree between the original map and each of the generalized ones, and are required to tell if the similarity degrees are acceptable.

Step 3: Statistical analysis

The similarity degrees calculated by the new models and that obtained from the experiments are listed in Table 5-1. After statistical analysis on these data, the spatial similarity degrees calculated by the new models and map scale changes as well as the number of the subjects that agree/disagree with the calculated credibility of the spatial similarity degrees are listed in Table 5-2.

5.4 Samples in Psychological Experiments

5.4.1 Rules Obeyed in Sample Selection

Totally ten types of samples are considered. Every type has three or four samples. All of the samples used in the experiments are shown below. All maps in the experiments are not shown to exact scales.

It is evident that the more samples are used in the psychological experiments, the better the results will be. However, it is not possible to use all sample of objects (object groups) in the geographic space in the experiments. A feasible way is to pursue a balance between the number of the samples and the accuracy of the experiments. Hence, some rules are employed in selecting the samples for each of the experiments so that the balance can be reached.

• At least three samples should be selected for each category of the objects. In each sample, five generalized results of the original objects (or object groups) are shown. Therefore, after psychological experiments, at least 15 coordinate pairs can be obtained with spatial similarity degrees and map scale change as coordinates. This ensures that enough points can be supplied for constructing the relations between spatial similarity degree and map scale change as coordinate by curve fitting.

• The ten categories of objects discussed in this thesis are all taken into consideration so that the samples can include all types of objects on topographic maps.

• The samples in each category of objects (or object groups) should be obviously different from each other so that they can be good representations of other objects of corresponding category. To guarantee good representation of the samples, many experienced cartographers have been invited to design examples for each category of objects and choose typical samples from three map databases owned by the Chinese Academy of Survey and Mapping, the National Centre of Geomatics, China, and the Map Academy of Gansu Province. The differences of the samples in each category can be seen by the figure captions in Figure 5-1 to 5-34.

5.4.2 Samples Used

Individual Points

Three individual point objects are used in the experiments (Figures 4-19, 4-20 and 4-21). It is not possible to simplify a point symbol; hence their symbols are all the same at different map scales.

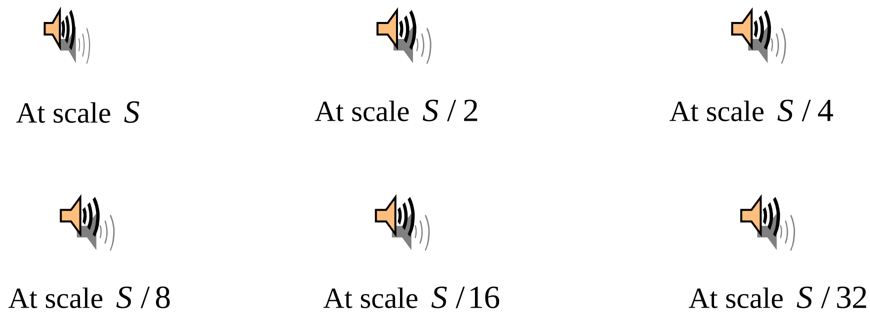


Figure 5-1 Experiment 1: a broadcasting station at different map scales.

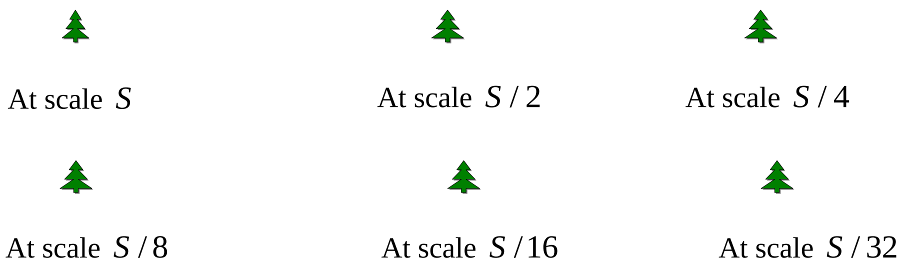


Figure 5-2 Experiment 2: an individual tree at different map scales.

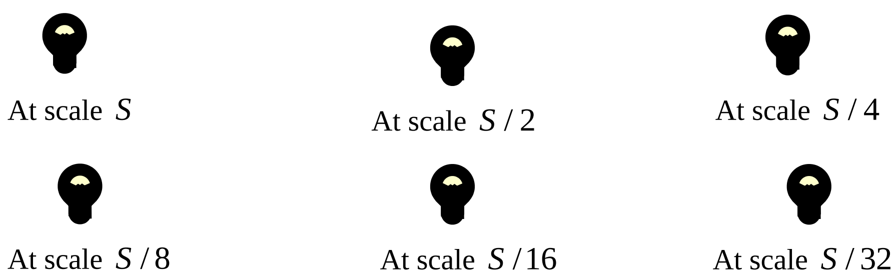


Figure 5-3 Experiment 3: a traffic light at different map scales.

Individual Lines

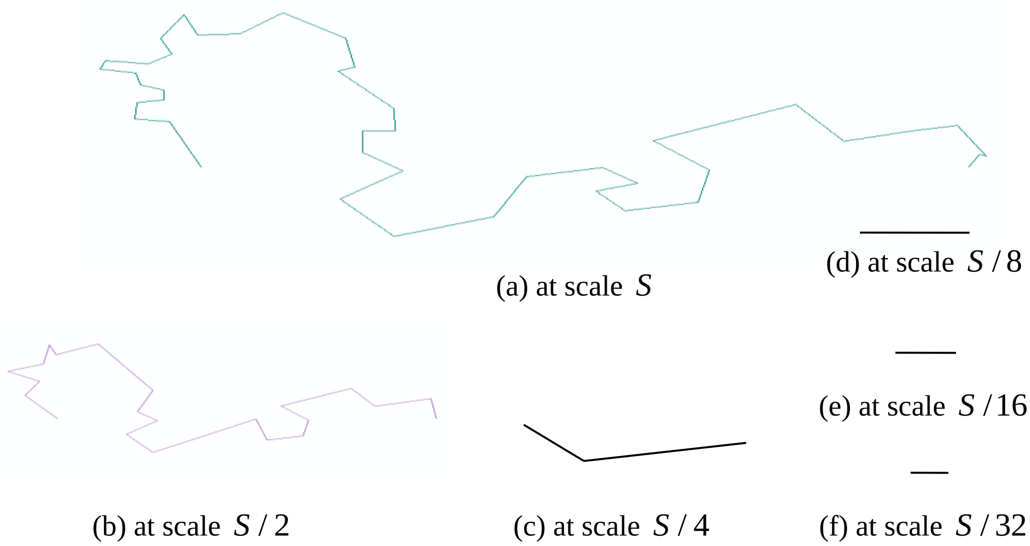


Figure 5-4 Experiment 4: a road at different map scales.

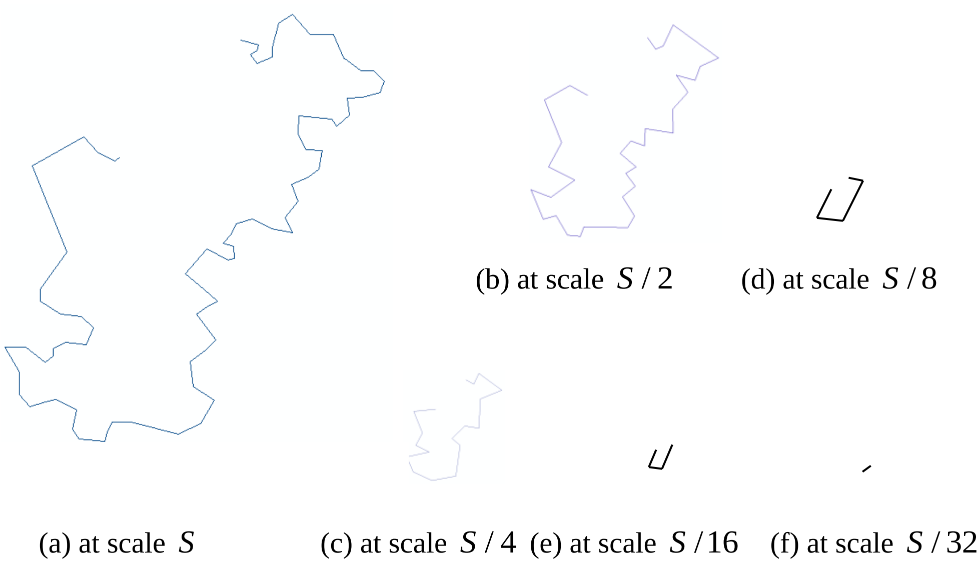


Figure 5-5 Experiment 5: a segment of a boundary line at different map scales.

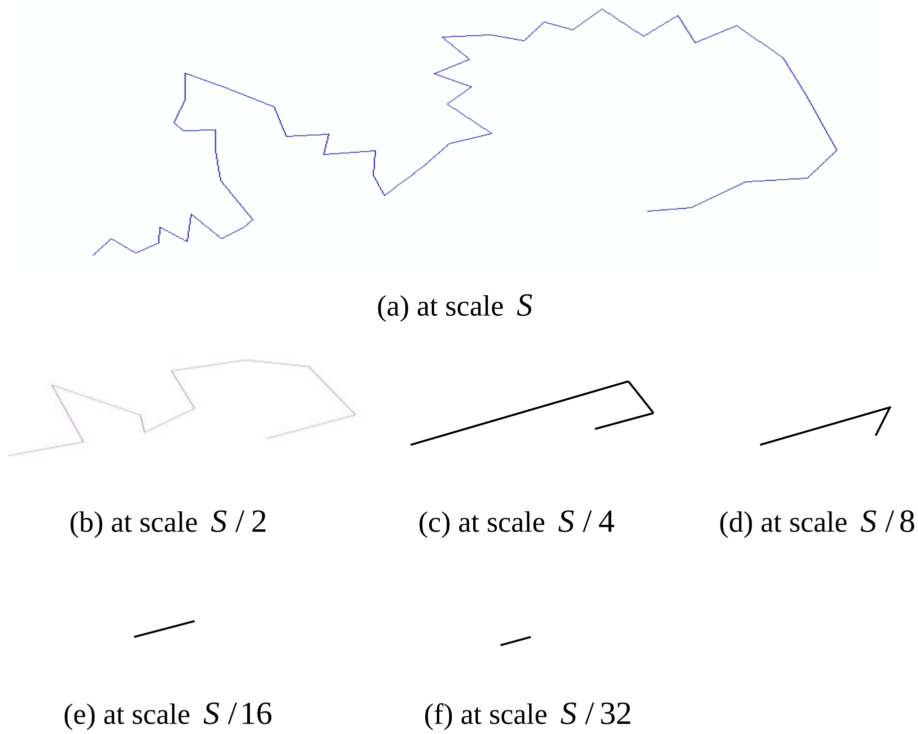


Figure 5-6 Experiment 6: a coastline at different map scales.

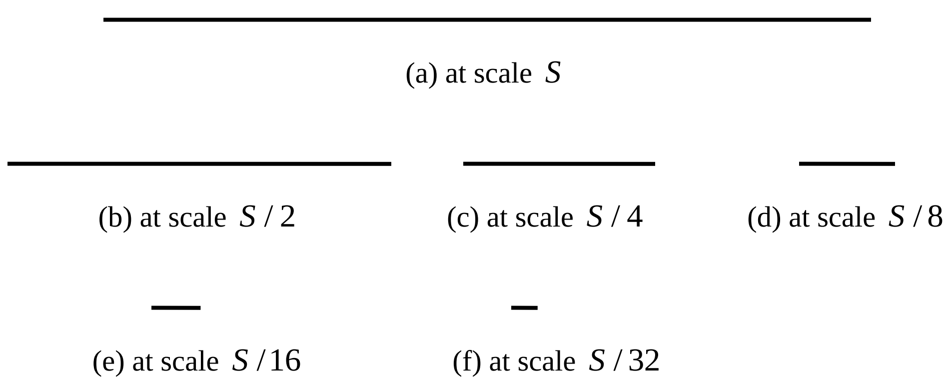
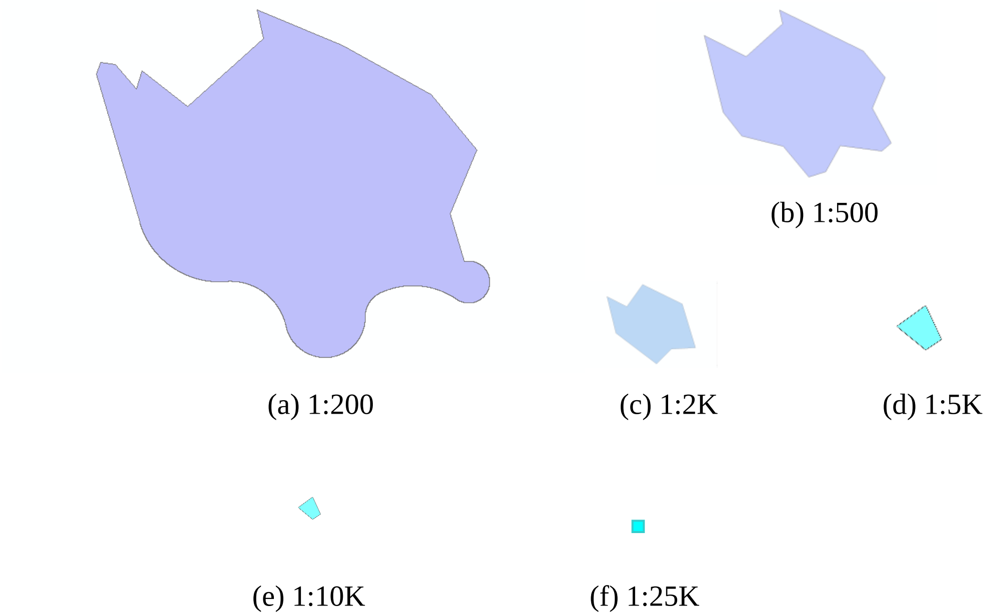


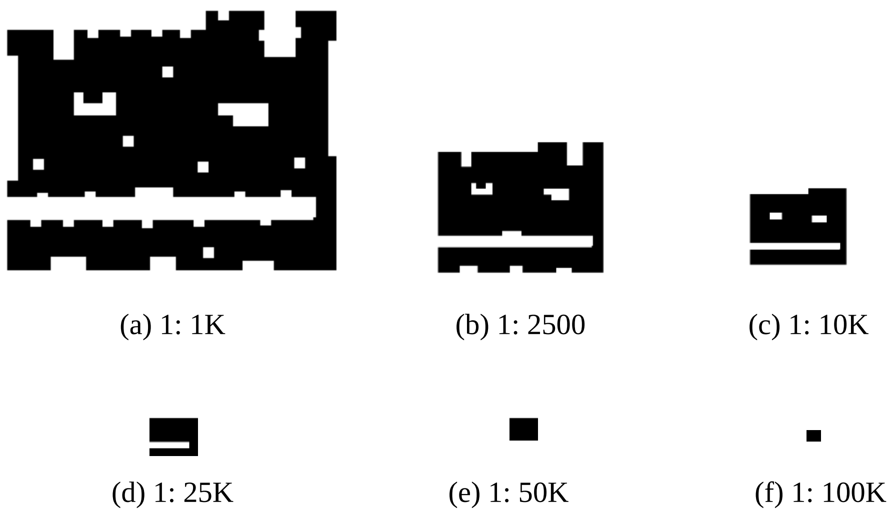
Figure 5-7 Experiment 7: a ditch at different map scales.

Individual Polygons



(a) 1:200 (b) 1:500 (c) 1:2K (d) 1:5K (e) 1:10K (f) 1:25K

Figure 5-8 Experiment 8: a pool at different map scales.



(a) 1: 1K (b) 1: 2500 (c) 1: 10K (d) 1: 25K (e) 1: 50K (f) 1: 100K

Figure 5-9 Experiment 9: a settlement at different map scales.

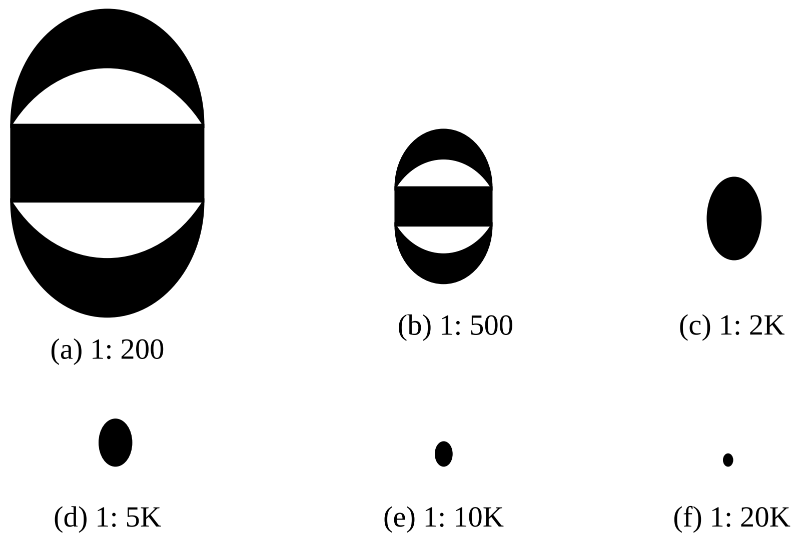


Figure 5-10 Experiment 10: an opera house at different map scales.

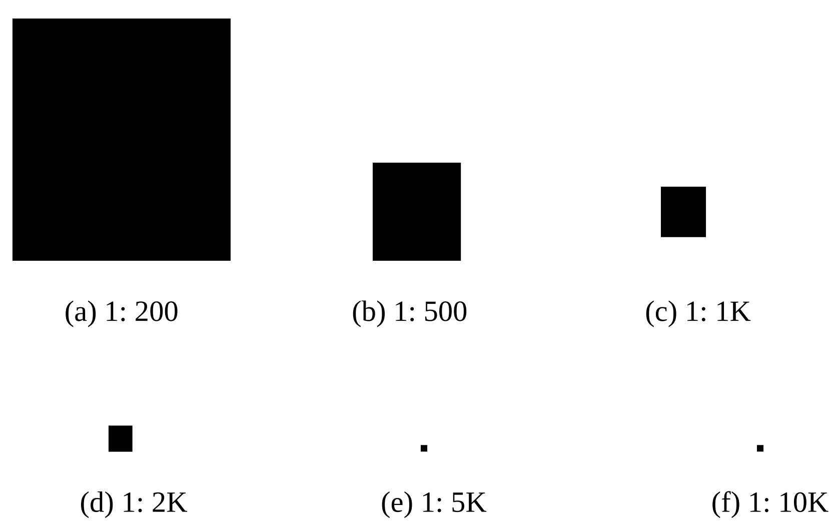
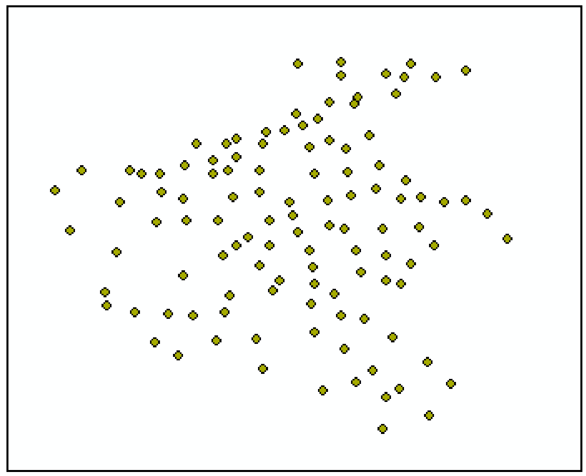
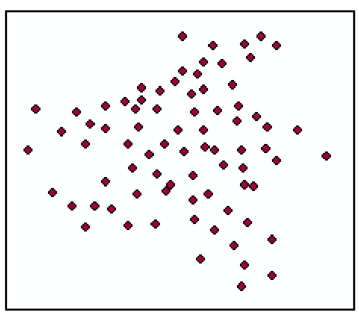


Figure 5-11 Experiment 11: a townhouse at different map scales.

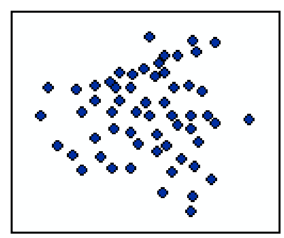
Point Clouds



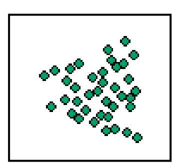
(a) 1: 10K, 113 points



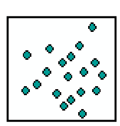
(b) 1: 20K, 78 points



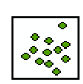
(c) 1: 1, 50K, 58 points



(d) 1: 100K, 38 points

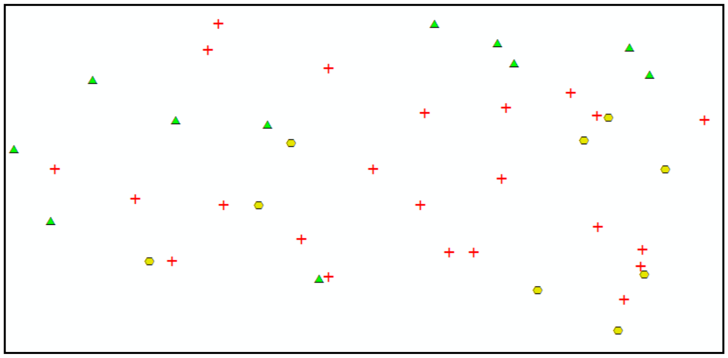


(e) 1: 250K, 19 points

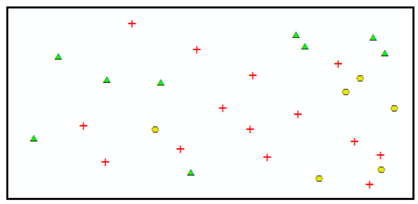


(f) 1: 500K, 12 points

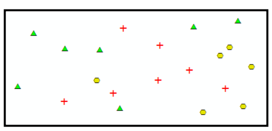
Figure 5-12 Experiment 12: point clouds at different map scales.
The weights of all points are equal.



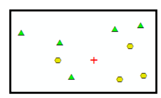
(a) 1: 10K, 43 points.



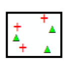
(b) 1: 20K, 29 points retained.



(c) 1: 50K, 20 points retained



(d) 1: 100K, 10 points retained.



(e) 1: 250K, 6 points retained.

(f) 1: 500K, 3 points retained.

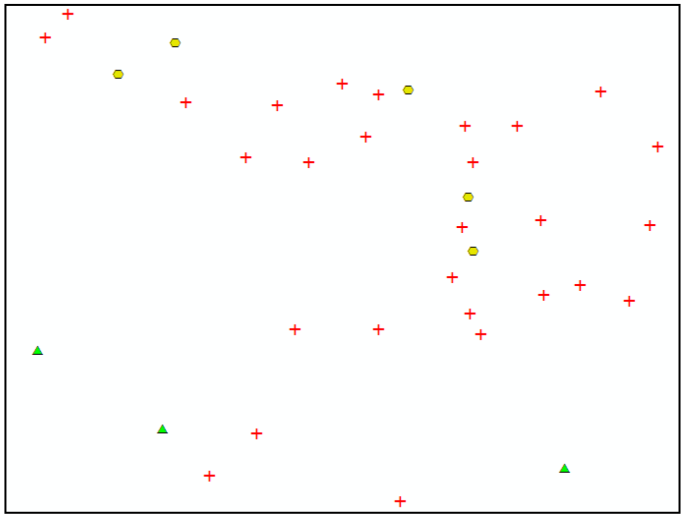
Legend:

First class control point. The weight is 4.

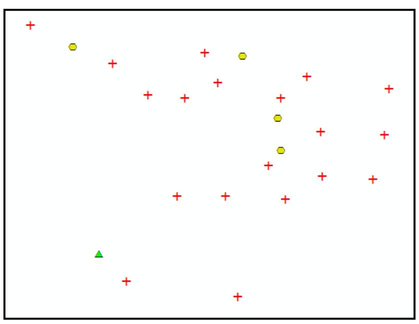
Second class control point. The weight is 2.

Third class control point. The weight is 1.

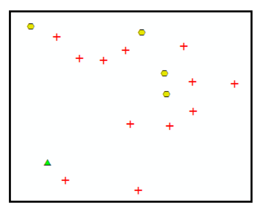
Figure 5-13 Experiment 13: control points in a regular area at different scales.



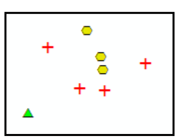
(a) 1: 10K, 36 points.



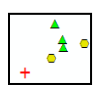
(b) 1: 20K, 24 points retained.



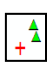
(c) 1: 50K, 17 points retained.



(d) 1: 100K, 8 points retained.



(e) 1: 250K, 6 points retained.



(f) 1: 500K,3 points retained.

Legend:

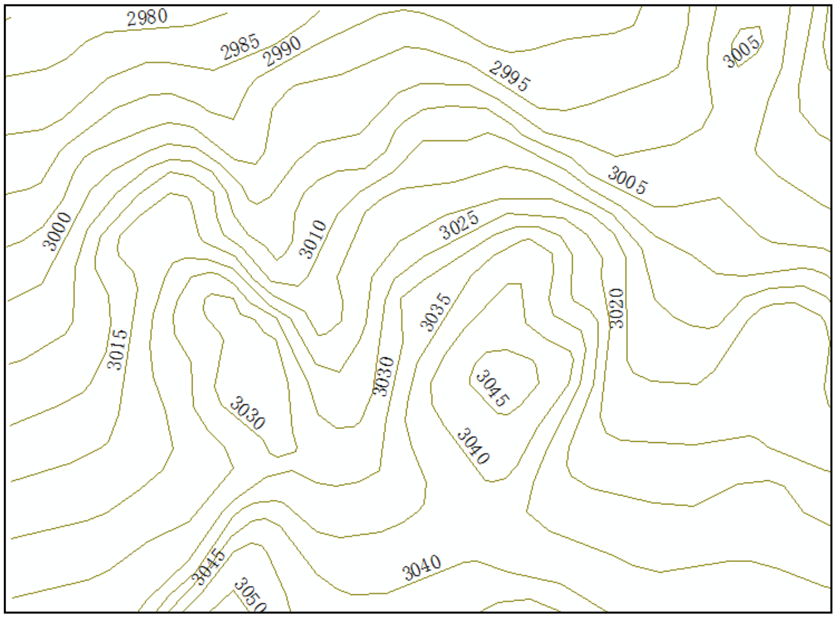
First class, weight is 4.

Second class, weight is 2.

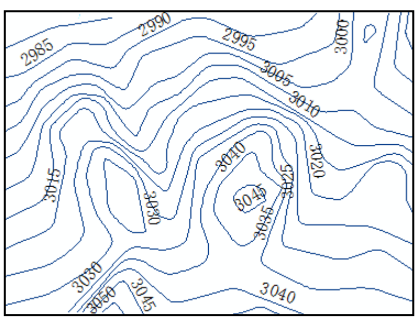
Third class, weight is 1.

Figure 5-14 Experiment 14: control points in an irregular area at different scales.

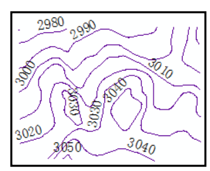
Parallel Lines Clusters



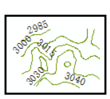
(a) 1: 10K.



(b) 1: 20K.



(c) 1: 50K.

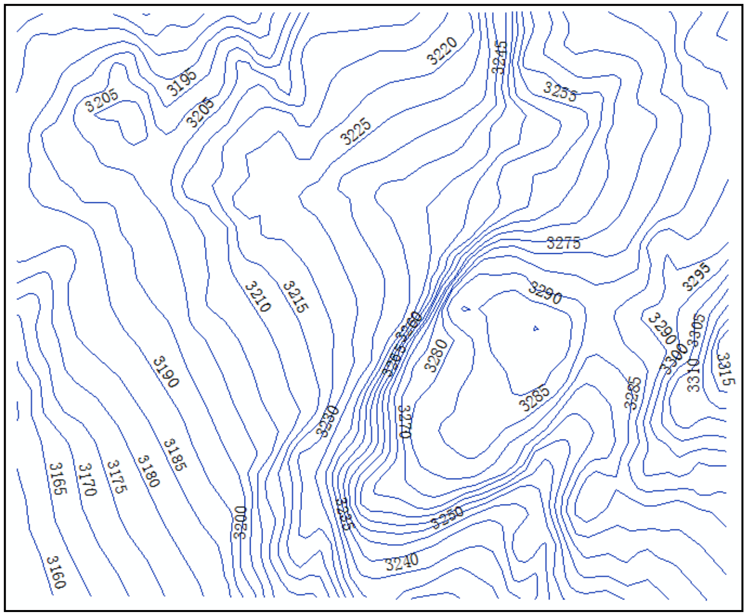


(d) 1: 100K.

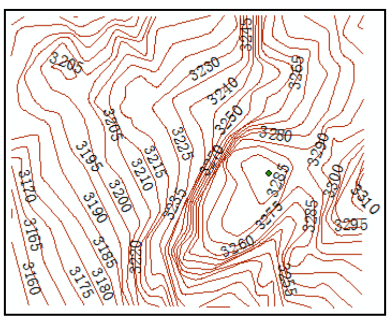
(e) 1: 250K.

(f) 1: 500K.

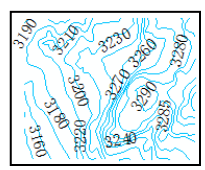
Figure 5-15 Experiment 15: contours representing a gentle hill at different scales.



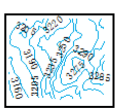
(a) 1: 10K.



(b) 1: 20K.



(c) 1: 50K.

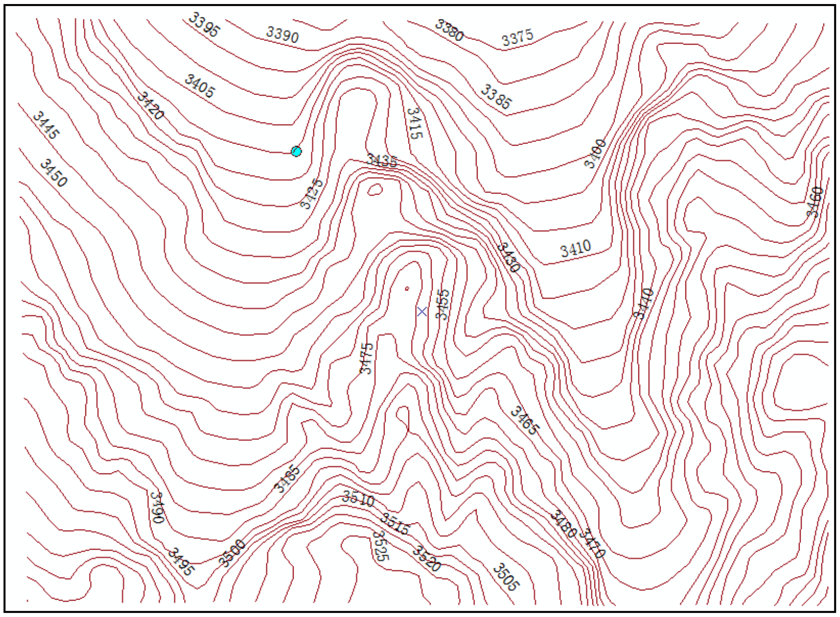


(d) 1: 100K.

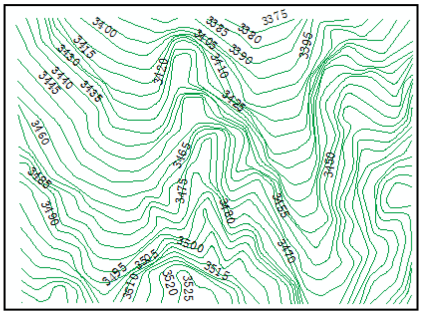
(e) 1: 250K.

(f) 1: 500K.

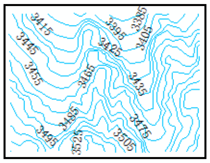
Figure 5-16 Experiment 16: contours representing a steep slope at different scales.



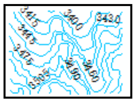
(a) 1: 10K.



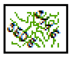
(b) 1: 20K.



(c) 1: 50K.



(d) 1: 100K.



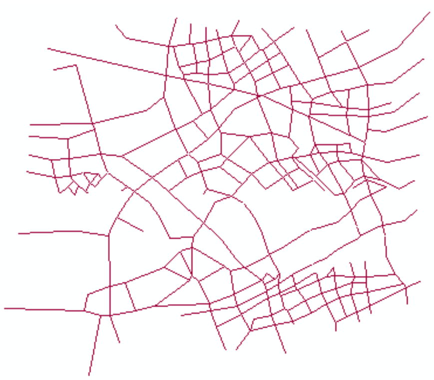
(e) 1: 250K.

(f) 1: 500K.

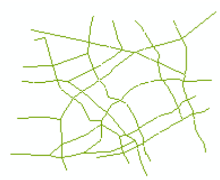
Figure 5-17 Experiment 17: contours representing a gulley at different scales.

Intersected Line Networks

(a) 1: 10K.

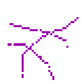


(b) 1: 20K.



(c) 1: 50K.

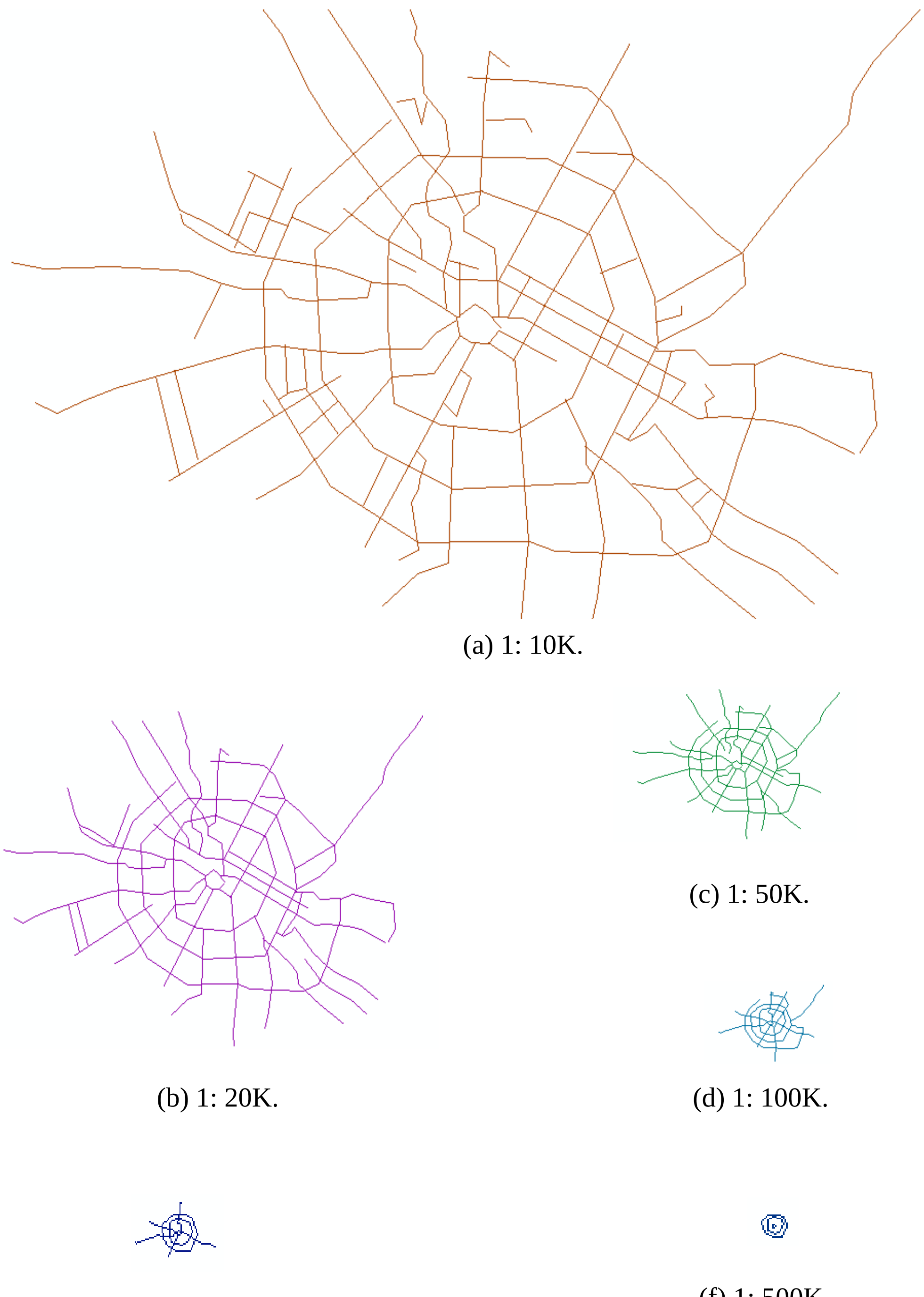
(d) 1: 100K.



(e) 1: 250K.

(f) 1: 500K.

Figure 5-18 Experiment 18: an ordinary road network at different map scales.



(a) 1: 10K.

(b) 1: 20K.

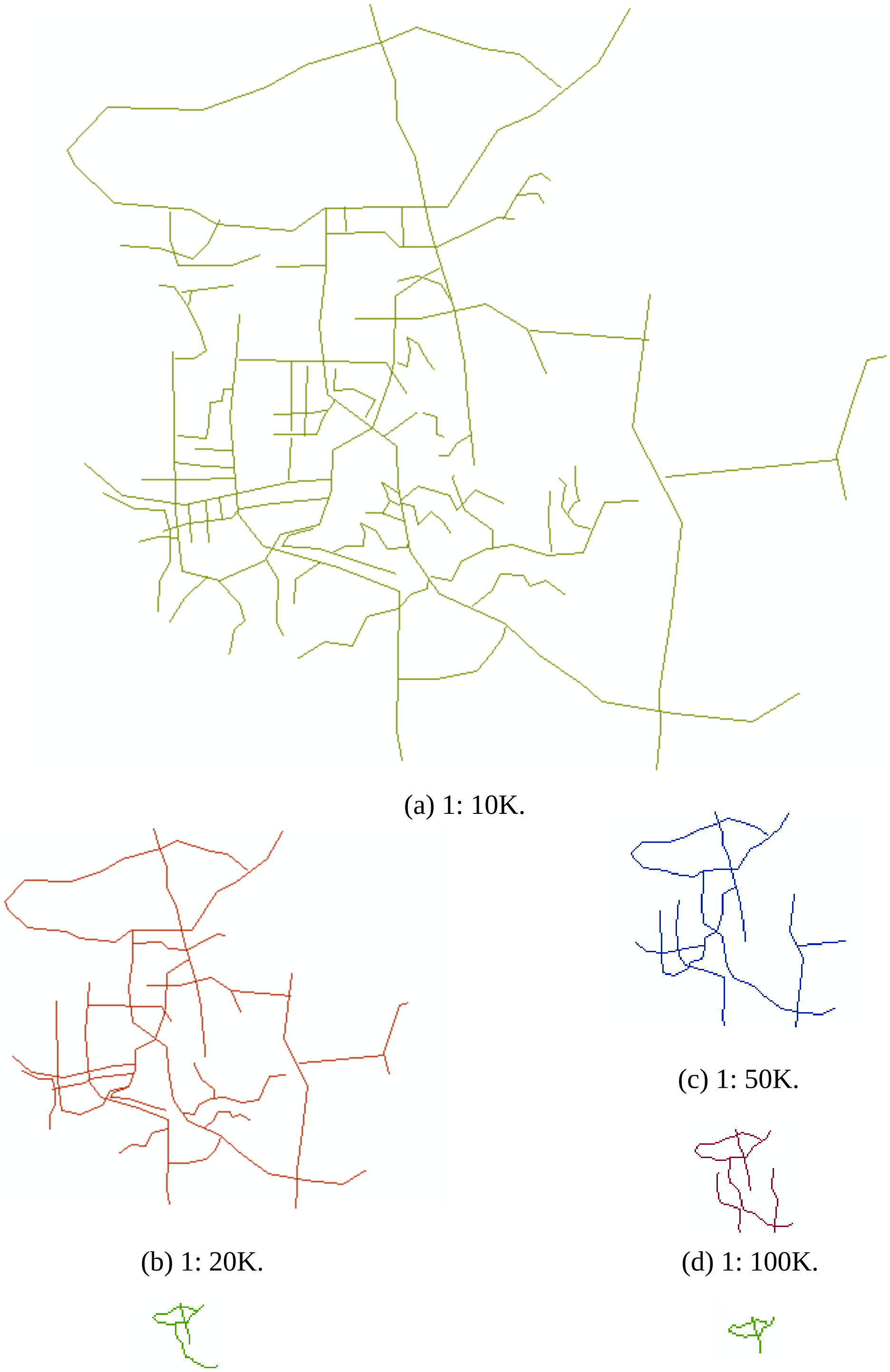
(c) 1: 50K.

(d) 1: 100K.

(e) 1: 250K.

(f) 1: 500K.

Figure 5-19 Experiment 19: a road network with ring roads at different map scales.



(a) 1: 10K.

(b) 1: 20K.

(c) 1: 50K.

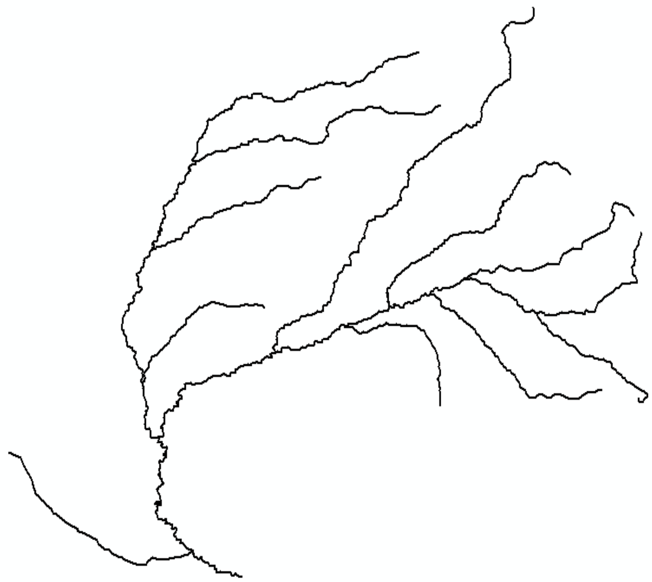
(d) 1: 100K.

(e) 1: 250K.

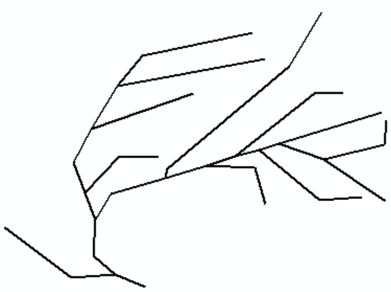
(f) 1:500K.

Figure 5-20 Experiment 20: a road network with zigzag roads at different map scales.

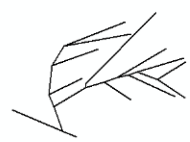
Tree-like Networks



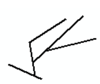
(a) 1: 10K.



(b) 1: 25K.



(c) 1: 50K.

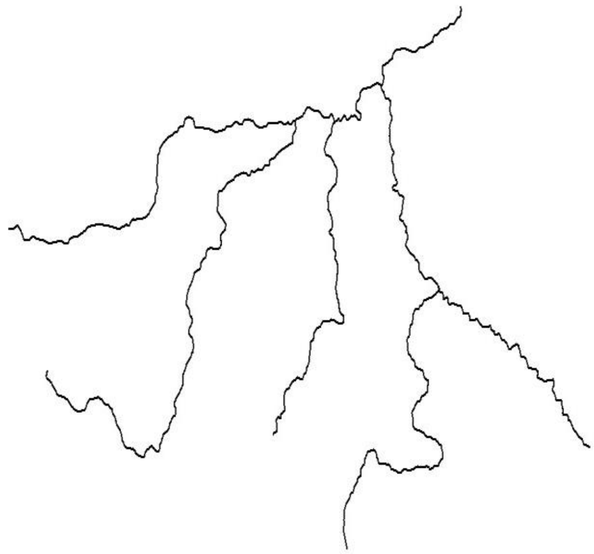


(d) 1: 250K.

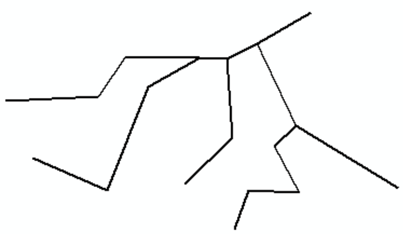
(e) 1: 500K.

(f) 1: 1M.

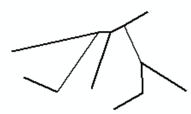
Figure 5-21 Experiment 21: a river network at different map scales.



(a) 1: 10K.



(b) 1: 25K.



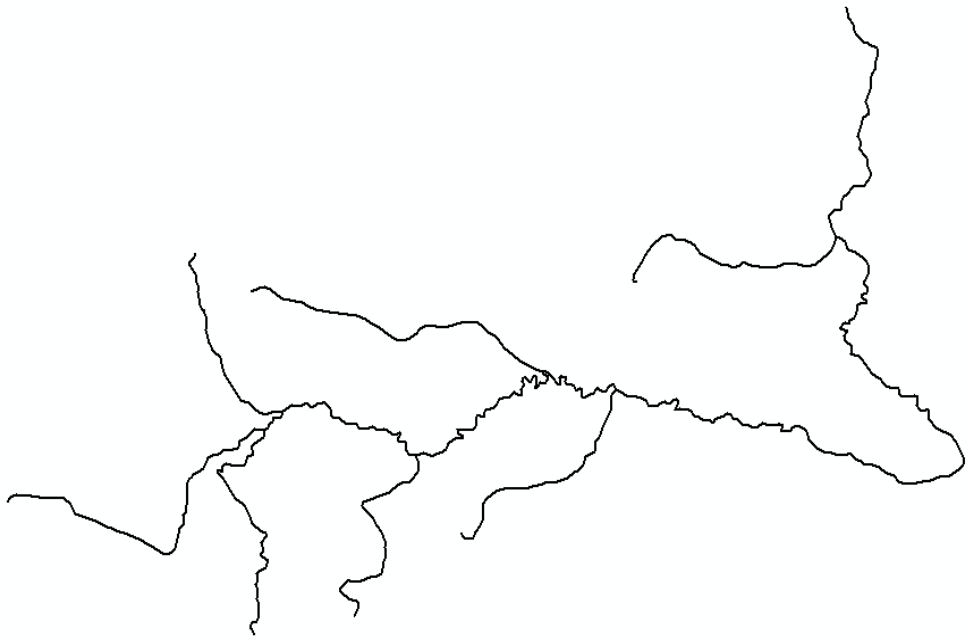
(c) 1: 50K.

(d) 1: 250K.

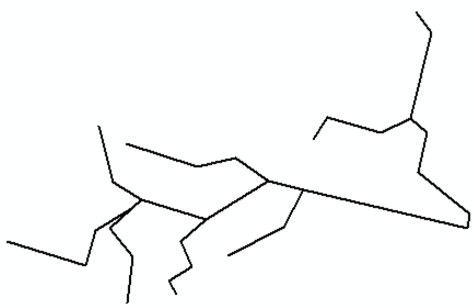
(e) 1: 500K.

(f) 1: 1M.

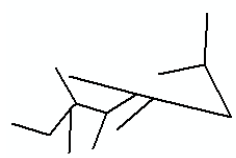
Figure 5-22 Experiment 22: a river network at different map scales.



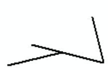
(a) 1: 10K.



(b) 1: 25K.



(c) 1: 50K.



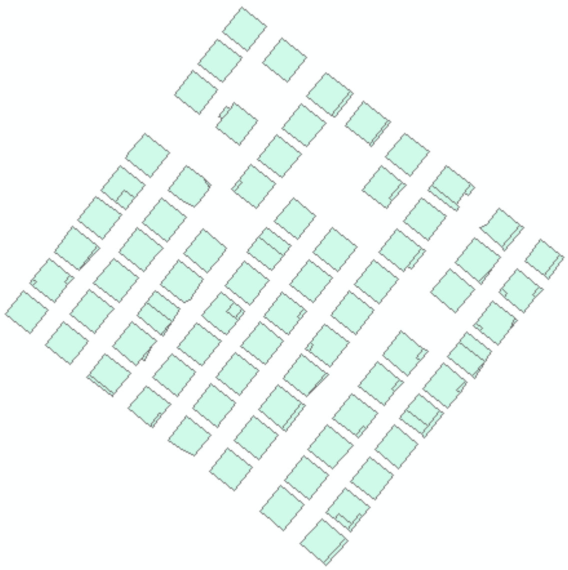
(d) 1: 250K.

(e) 1: 500K.

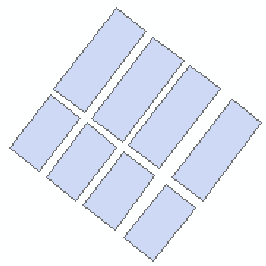
(f) 1: 1M.

Figure 5-23 Experiment 23: a river network at different map scales.

Discrete Polygon Groups



(a) 1: 10K.



(b) 1: 25K.

(c) 1: 50K.

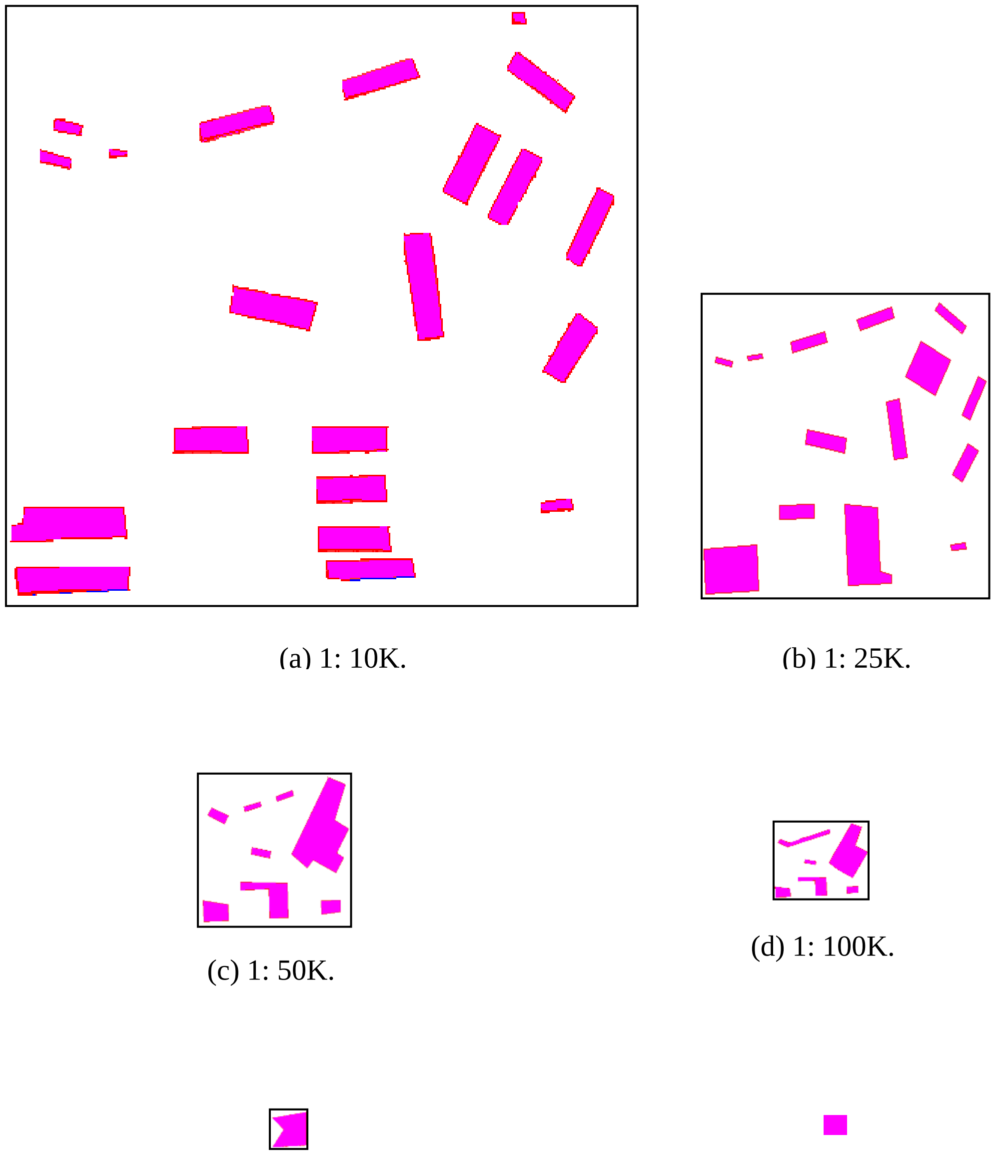
(d) 1: 100K.

(e) 1: 250K.

(f) 1: 500K.

Figure 5-24 Experiment 24: regularly-shaped and distributed settlements.

The settlements are rectangular shaped and regular distributed in a block at different map scales.



(a) 1: 10K. (b) 1: 25K.

(c) 1: 50K. (d) 1: 100K.

(e) 1: 250K. (f) 1: 500K.

Figure 5-25 Experiment 25: simple settlements at different map scales.

The settlements have simple and rectangular shapes, and have different orientations and much parallelism.

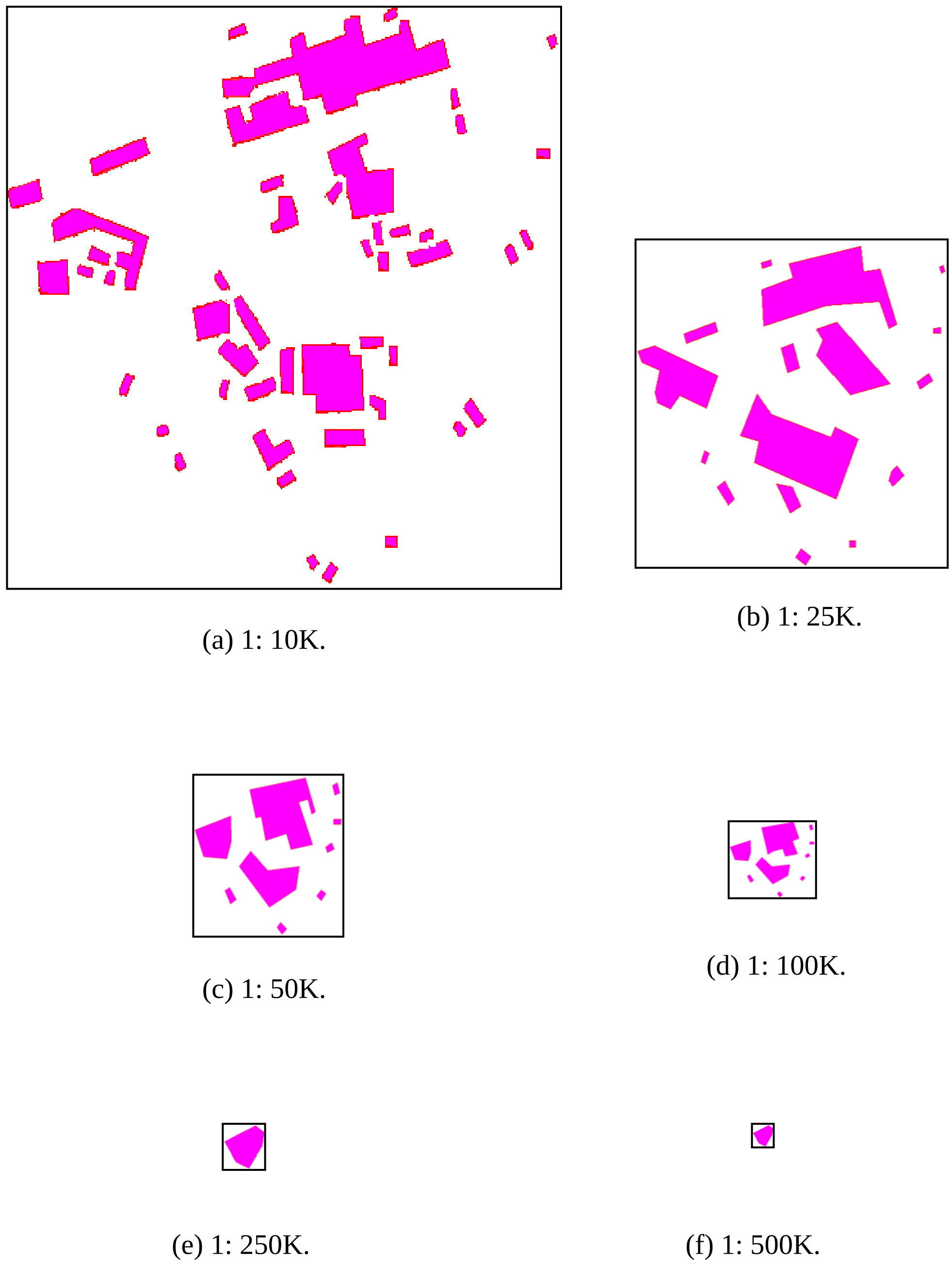
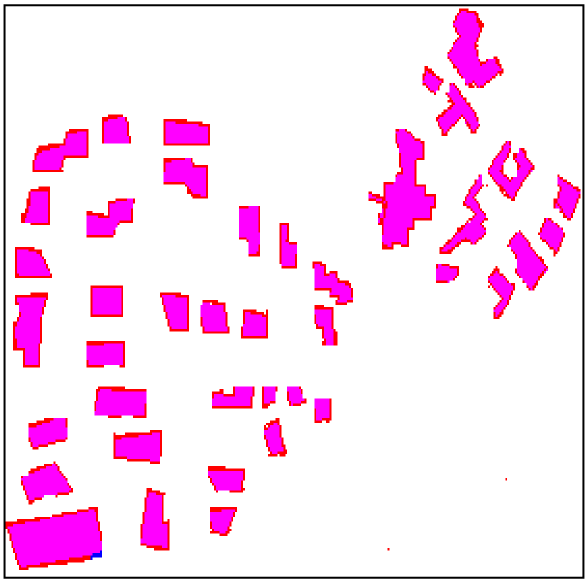
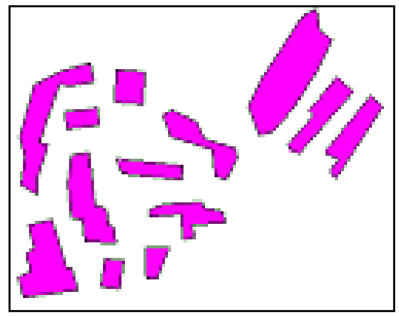


Figure 5-26 Experiment 26: complex settlements at different map scales.

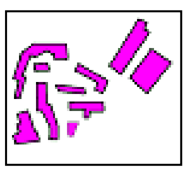
The settlements are complex-shaped but basically orthogonal in the corners, and show different orientations and little parallelism.



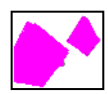
(a) 1: 10K.



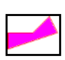
(b) 1: 25K.



(c) 1: 50K.



(d) 1: 100K.

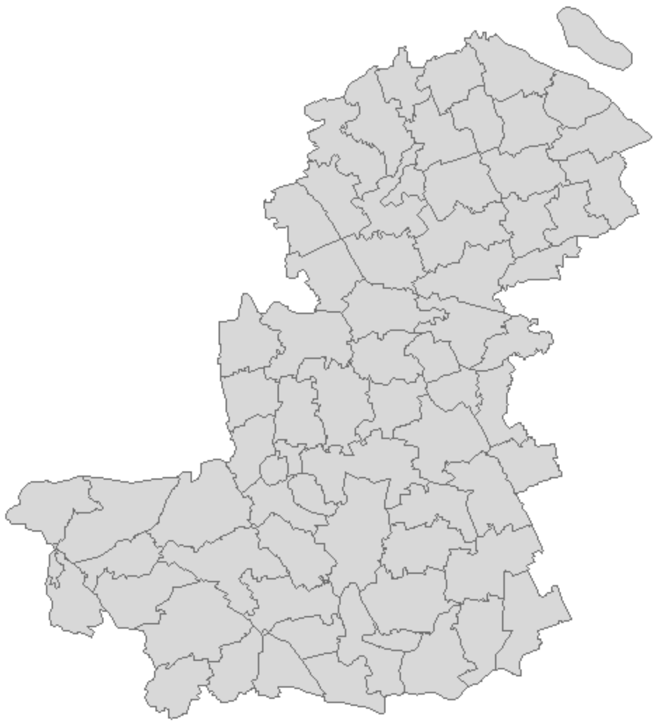


(e) 1: 250K.

(f) 1: 500K.

Figure 5-27 Experiment 27: irregular-shaped settlements at different map scales. The settlements have complex and non-convex shapes with arbitrary angles in the corners, and have arbitrary orientations and little parallelism.

Connected Polygon Groups

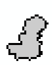


(a) 1: 500.

(b) 1: 1K.

(c) 1: 2.5K.

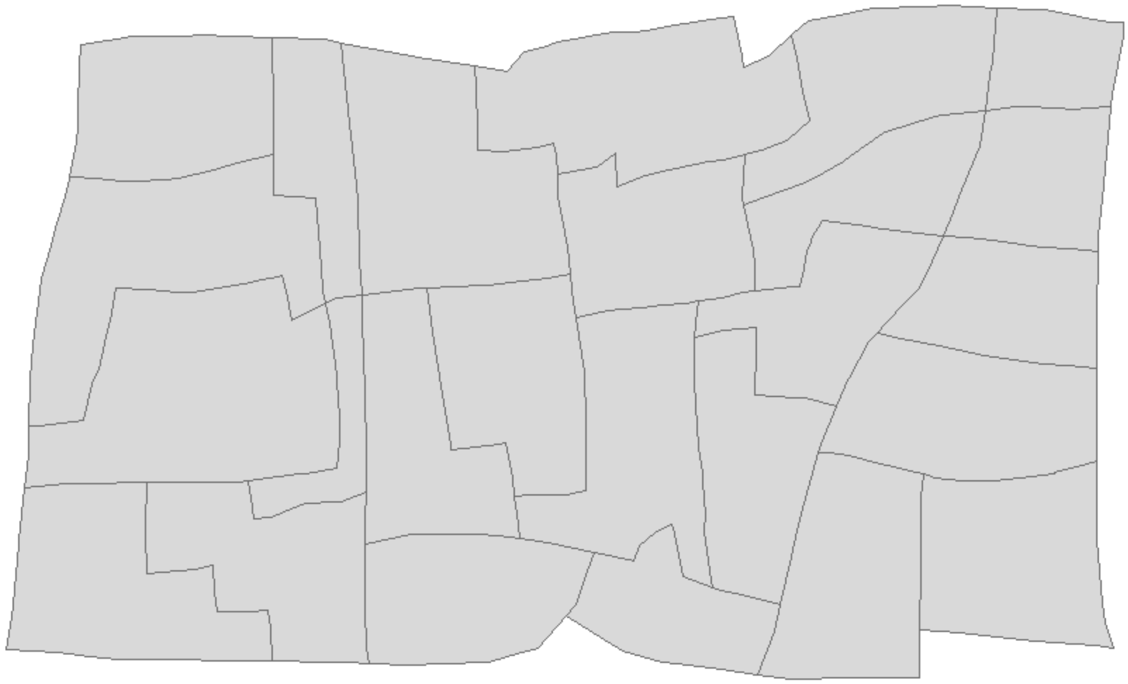
(d) 1: 5K.



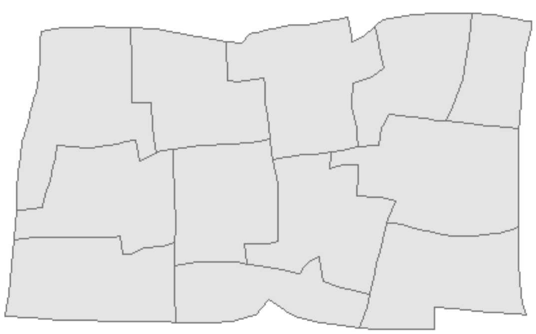
(e) 1: 10K.

(f) 1: 25K.

Figure 5-28 Experiment 28: a township consisting of patches at different map scales.



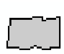
(a) 1: 2K.



(b) 1: 5K.

(c) 1: 10K.

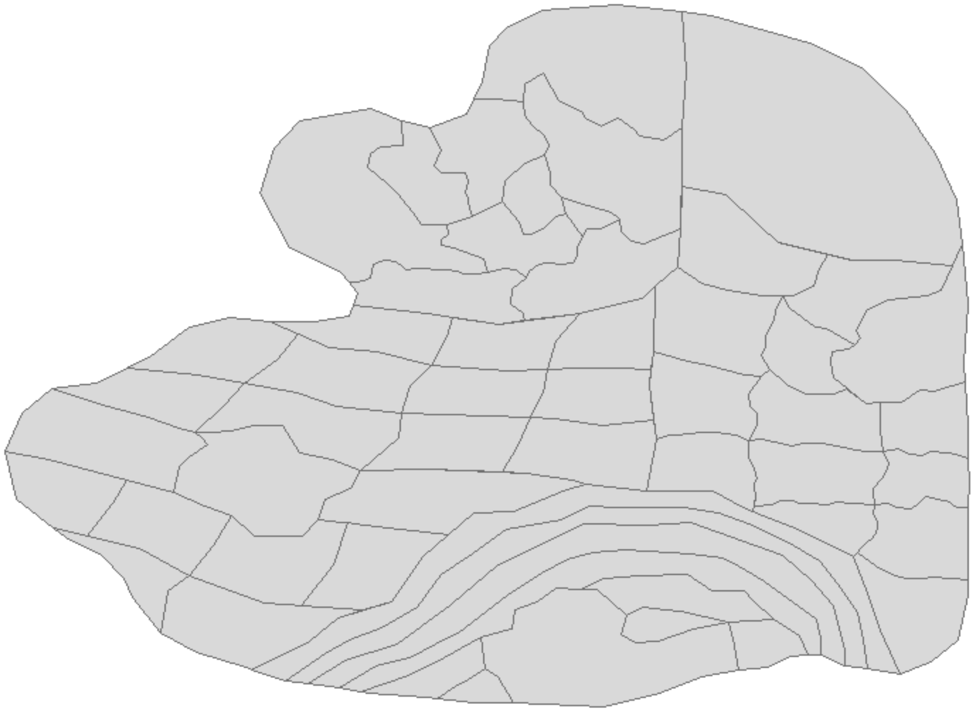
(d) 1: 20K.



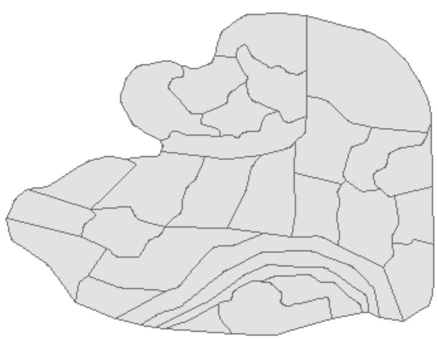
(e) 1: 50K.

(f) 1: 100K.

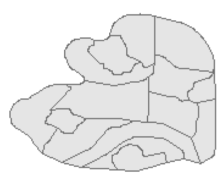
Figure 5-29 Experiment 29: polygonal boundary map at different scales.



(a) 1: 2K.



(b) 1: 5K.



(c) 1: 10K.

(d) 1: 20K.

(e) 1: 50K.

(f) 1: 100K.

Figure 5-30 Experiment 30: Connected polygonal farmlands at different map scales.

Maps

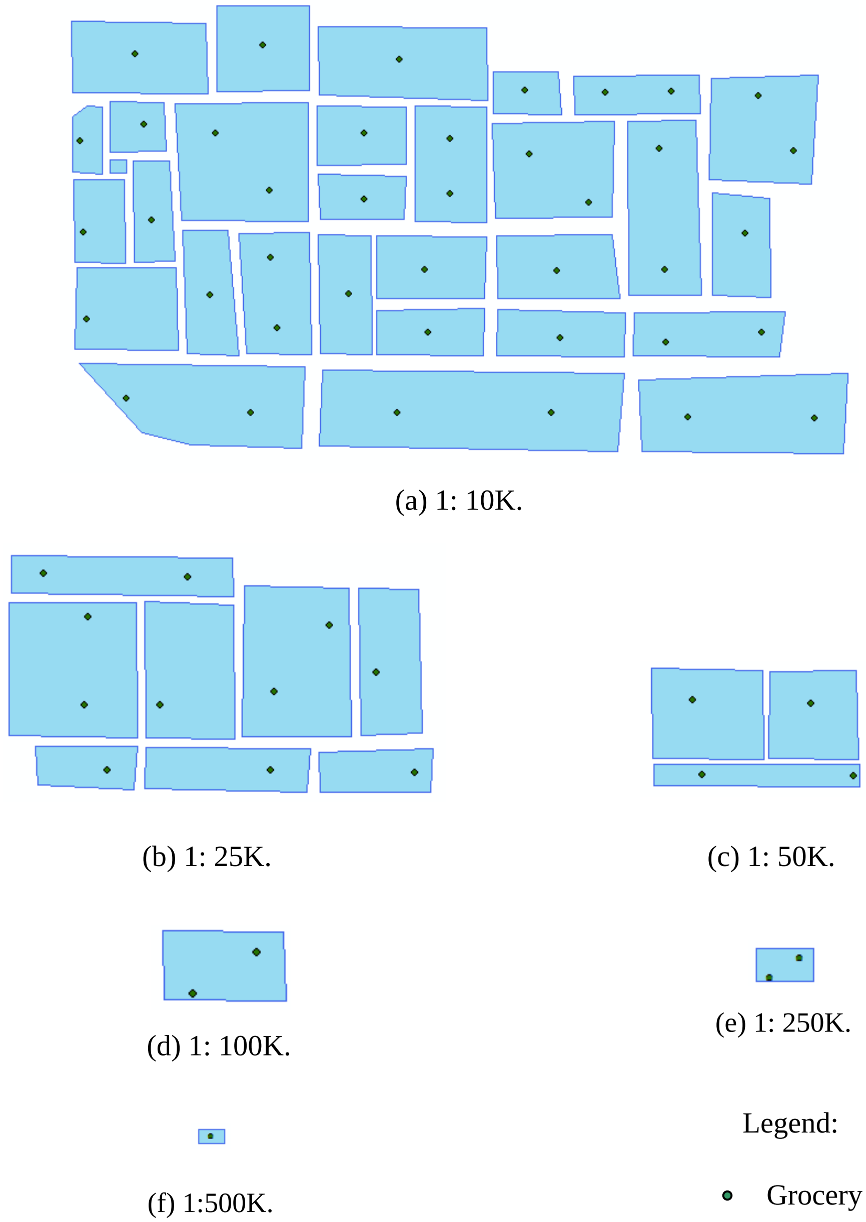


Figure 5-31 **Experiment 31: A street map at different map scales.**

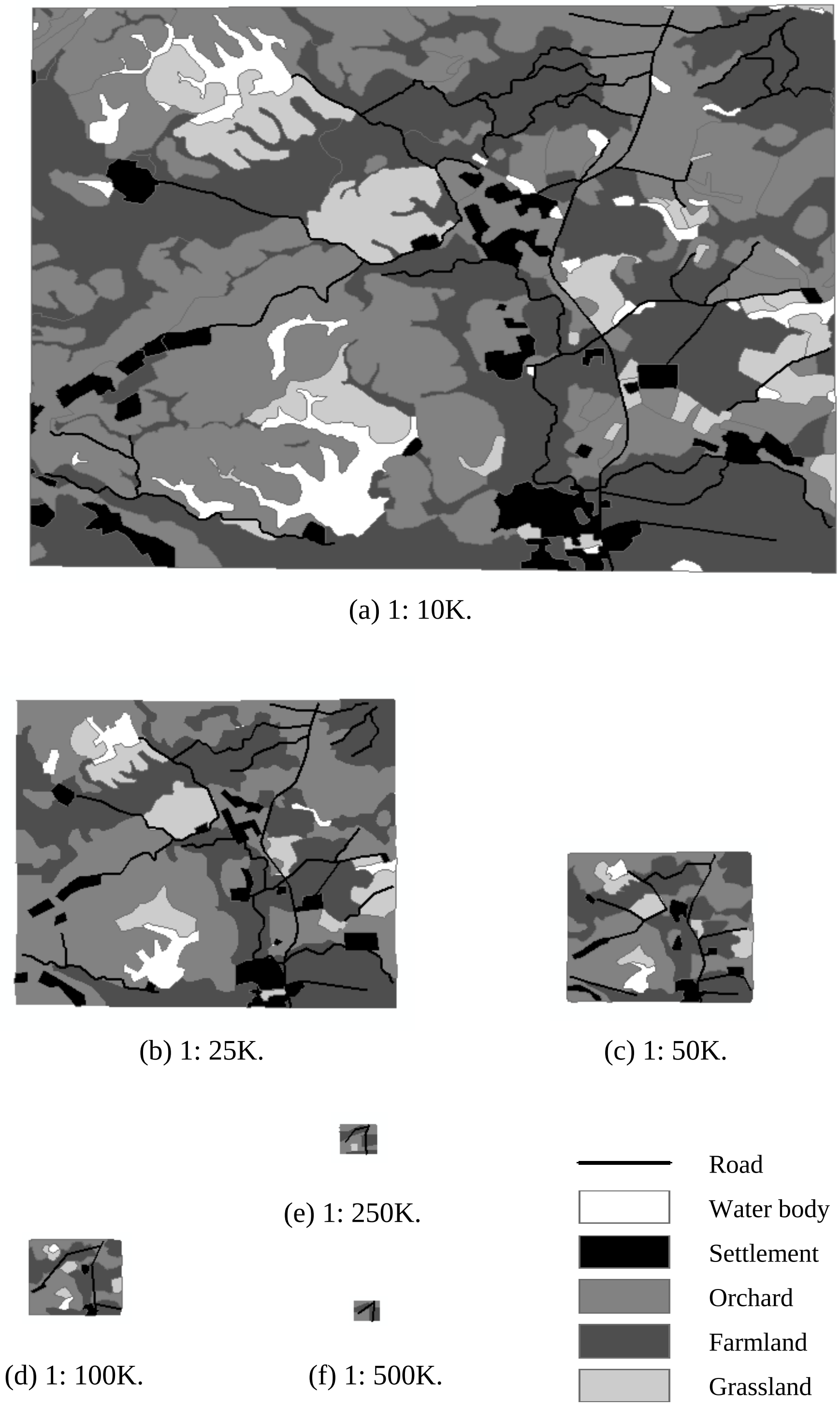
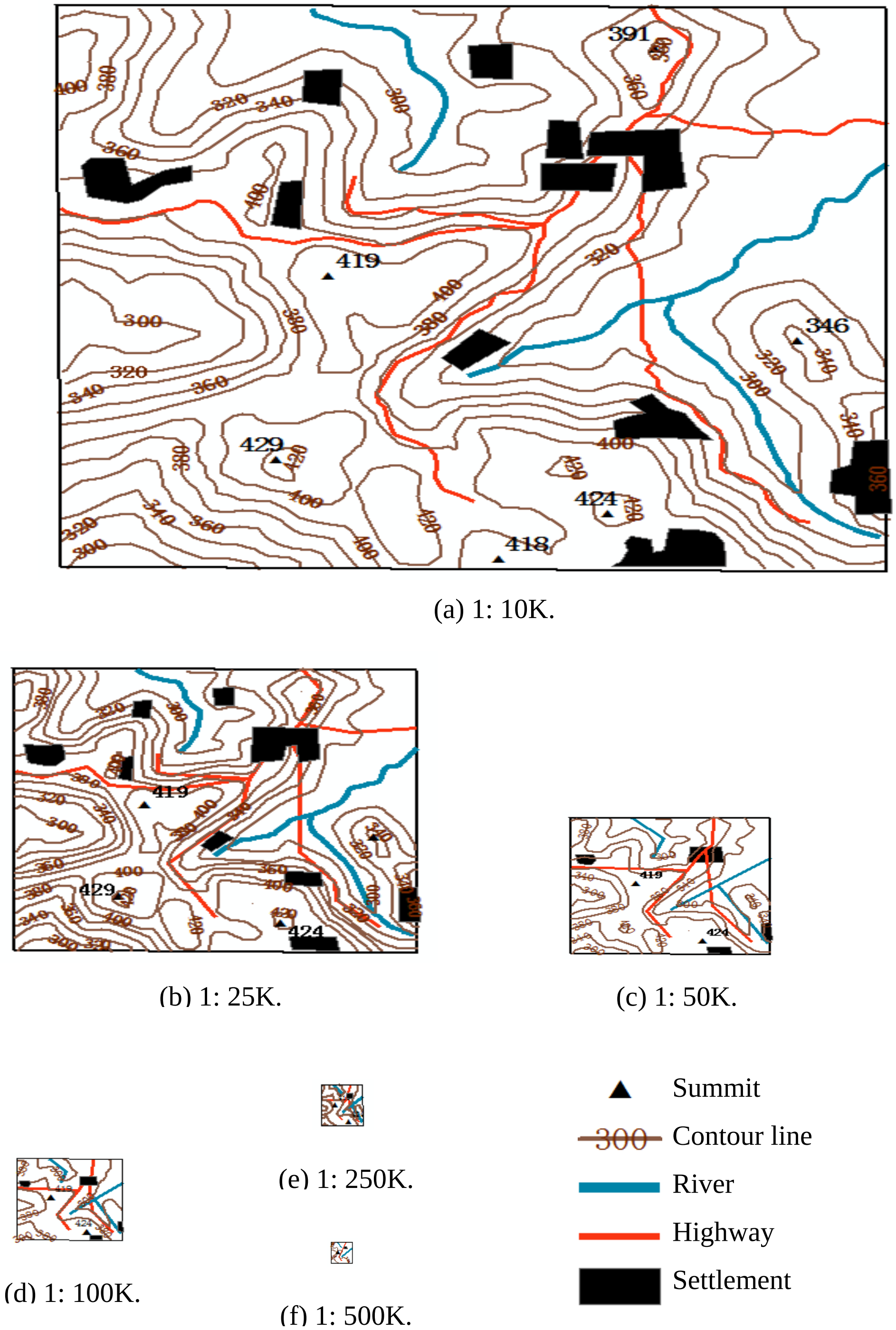


Figure 5-32 Experiment 32: a categorical map with irregular patches at different map scales.



(a) 1: 10K.

(b) 1: 25K.

(c) 1: 50K.

(d) 1: 100K.

(e) 1: 250K.

(f) 1: 500K.

Figure 5-33 Experiment 33: a topographic map at different map scales.

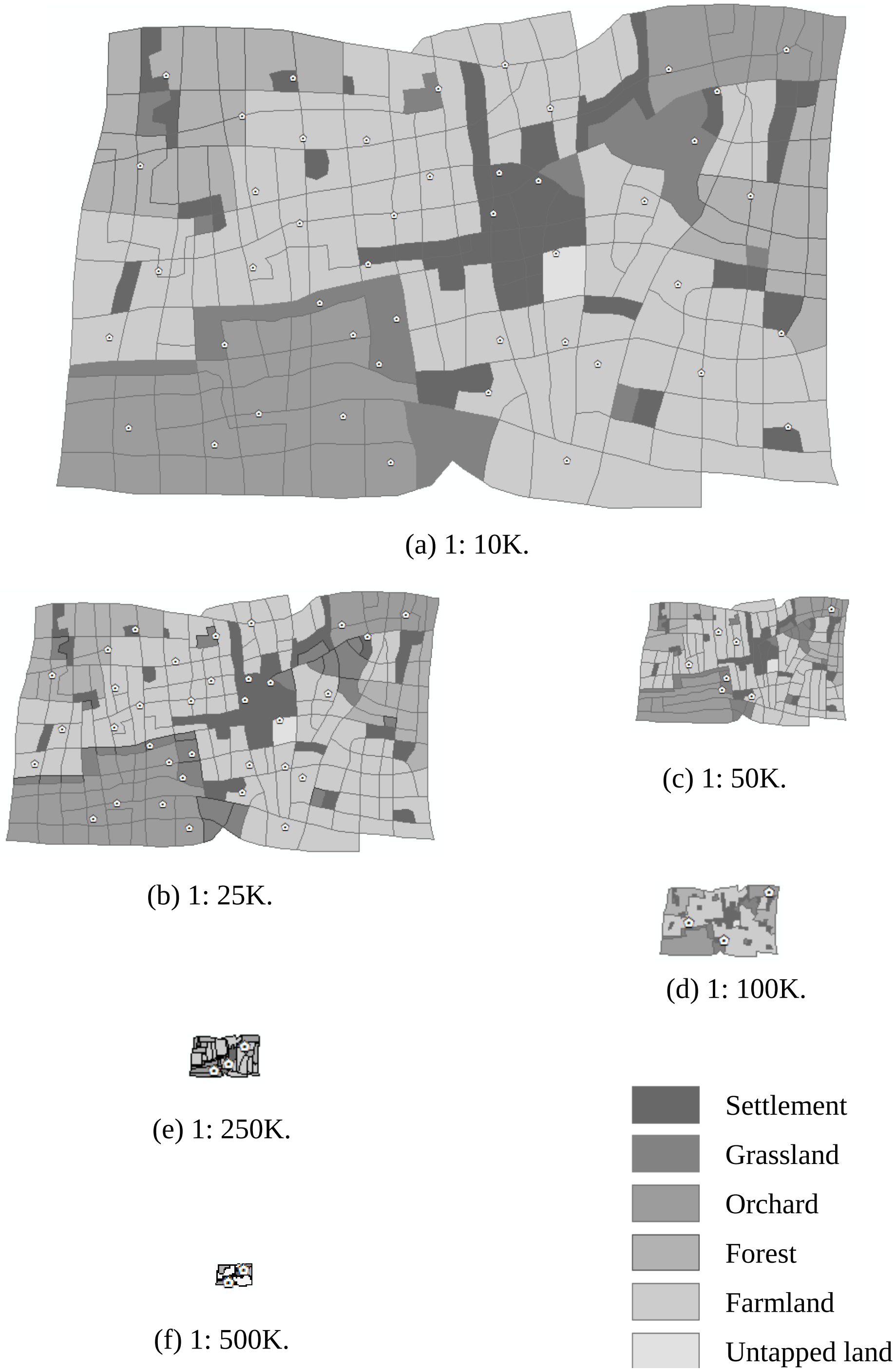


Figure 5-34 Experiment 34: a categorical map with regular patches at different map scales.

5.5 Statistical Analysis and Discussion

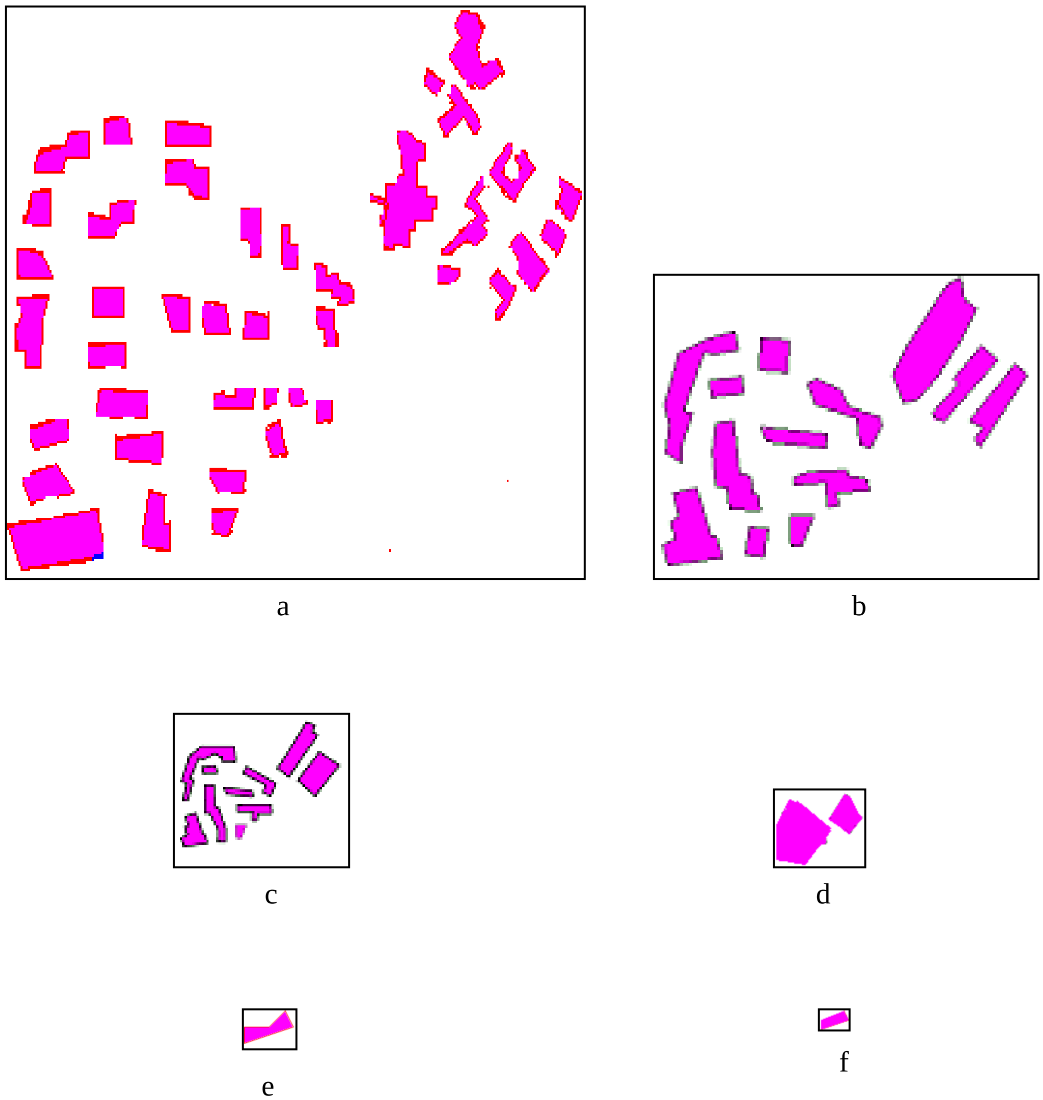


Figure 5-35 A sample used in the psychological experiments.

The above shows a map at four different scales. Below gives two groups of fractions in A and B. Each group comprises five values, representing the five similarity degrees between (a) and each of the other five objects/maps.

Similarity degrees

$Sim_{a,b}^{Map}=0.77$, $Sim_{a,c}^{Map}=0.45$, $Sim_{a,d}^{Map}=0.32$, $Sim_{a,e}^{Map}=0.00$, $Sim_{a,f}^{Map}=0.00$.

You are required to complete the following work.

◆ Tick at appropriate positions to tell if you can accept the similarity degrees in A.

A is acceptable (____)   A is not acceptable (____)   I have no idea (____)

◆ Use three values in [0,1] to represent the describe similarity degrees between (a) and the other five maps, respectively.

Value 1: (______)   Value 2: (_______)   Value 3: (________)

Value 4: (______)   Value5: (_______)

Figure 5-36 The answer sheet used in the psychological experiments.

The similarity degrees calculated by the new models and obtained from the subjects in the experiments (from Figure 5-1 to Figure 5-34) are listed in Table 5-1.

Table 5-1 Similarity degrees obtained by three different methods.

Experiment No.	$Sim_{a,b}^{V}$, $Sim_{a,c}^{V}$, $Sim_{a,d}^{V}$, $Sim_{a,e}^{V}$, $Sim_{a,f}^{V}$	$Sim_{a,b}^{E}$, $Sim_{a,c}^{E}$, $Sim_{a,d}^{E}$, $Sim_{a,e}^{E}$, $Sim_{a,f}^{E}$
1	1.00, 1.00, 1.00, 1.00, 1.00	1.00, 1.00, 1.00, 1.00, 1.00
2	1.00, 1.00, 1.00, 1.00, 1.00	1.00, 1.00, 1.00, 1.00, 1.00
3	1.00, 1.00, 1.00, 1.00, 1.00	1.00, 1.00, 1.00, 1.00, 1.00
4	0.87, 0.64, 0.38, 0.38, 0.38	0.86,0.49,0.34,0.25,0.21

5	0.91, 0.78, 0.52, 0.44, 0.36	0.91,0.67,0.51,0.35,0.18
6	0.75, 0.55, 0.44, 0.35, 0.26	0.78,0.57,0.40,0.24,0.19
7	1.00, 1.00, 1.00, 1.00, 1.00	1.00, 1.00, 1.00, 1.00, 1.00
8	0.95, 0.88, 0.73, 0.65, 0.55	0.88,0.76,0.52,0.37,0.28
9	0.91, 0.82, 0.66, 0.52, 0.52	0.88,0.75,0.56,0.36,0.27
10	1.00, 0.55, .055, 0.55, 0.55	0.95,0.64,0.49,0.41,0.33
11	1.00, 1.00, 1.00, 1.00, 1.00	1.00, 1.00, 1.00, 1.00, 1.00
12	0.76, 0.57, 0.36, 0.21, 0.15	0.89,0.79,0.63,0.54,0.41
13	0.82, 0.62, 0.36, 0.19, 0.12	0.86,0.74,0.60,0.45,0.36
14	0.71, 0.58, 0.40, 0.18, 0.11	0.85,0.71,0.55,0.44,0.37
15	0.95, 0.88, 0.67, 0.45, 0.36	0.91,0.78,0.60,0.49,0.39
16	0.93, 0.83, 0.76, 0.51, 0.42	0.91,0.79,0.64,0.50,0.38
17	0.96, 0.86, 0.75, 0.55, 0.40	0.91,0.81,0.65,0.51,0.38
18	0.77, 0.52, 0.31, 0.22, 0.18	0.89,0.76,0.57,0.42,0.36
19	0.75, 0.55, 0.37, 0.28, 0.19	0.90,0.76,0.60,0.48,0.31
20	0.68, 0.49, 0.34, 0.28, 0.16	0.88,0.75,0.61,0.48,0.37
21	0.82, 0.55, 0.27, 0.21, 0.17	0.80,0.71,0.54,0.40,0.34
22	0.63, 0.49, 0.32, 0.22, 0.15	0.83,0.69,0.52,0.40,0.25
23	0.74, 0.56, 0.29, 0.23, 0.15	0.83,0.71,0.52,0.41,0.26
24	0.68, 0.38, 0.31, 0.16, 0.16	0.73,0.60,0.44,0.34,0.27
25	0.82, 0.58, 0.33, 0.21, 0.15	0.84,0.67,0.51,0.38,0.24
26	0.85, 0.51, 0.31, 0.22, 0.14	0.84,0.67,0.51,0.34,0.27
27	0.74, 0.47, 0.29, 0.25, 0.14	0.82,0.67,0.52,0.42,0.27
28	0.88, 0.76, 0.61, 0.44, 0.28	0.89,0.73,0.60,0.48,0.35
29	0.74, 0.57, 0.55, 0.38, 0.21	0.87,0.66,0.56,0.44,0.37
30	0.85, 0.72, 0.65, 0.46, 0.22	0.88,0.77,0.63,0.48,0.36
31	0.53, 0.39, 0.23, 0.22, 0.15	0.75,0.60,0.47,0.39,0.32
32	0.82, 0.67, 0.46, 0.33, 0.18	0.84,0.72,0.58,0.45,0.34
33	0.80, 0.69, 0.47, 0.27, 0.17	0.85,0.75,0.60,0.49,0.36
34	0.88, 0.68, 0.46, 0.39, 0.21	0.88,0.73,0.58,0.46,0.36

Notes : the following variables are applicable to Figures 5-1 to 5-34.

$Sim^V_{a,b}$: similarity degree between (a) and (b) calculated by the new model.

$Sim^V_{a,c}$: similarity degree between (a) and (c) calculated by the new model.

$Sim^V_{a,d}$: similarity degree between (a) and (d) calculated by the new model.

$Sim^V_{a,e}$: similarity degree between (a) and (e) calculated by the new model.

$Sim^V_{a,f}$: similarity degree between (a) and (f) calculated by the new model.

$Sim^E_{a,b}$ **:** similarity degree between (a) and (b) given by the subjects.

$Sim^E_{a,c}$ **:** similarity degree between (a) and (c) given by the subjects.

$Sim^E_{a,d}$ **:** similarity degree between (a) and (d) given by the subjects.

$Sim^E_{a,e}$ **:** similarity degree between (a) and (e) given by the subjects.

$Sim^E_{a,f}$ **:** similarity degree between (a) and (f) given by the subjects.

The spatial similarity degrees calculated by the new models and map scale changes as well as the number of the subjects that agree/disagree with the calculated credibility spatial similarity degrees are listed in Table 5-2.

Table 5-2 Similarity degree and map scale change

Experiment No.	$Sim^V_{a,b}$, $Sim^V_{a,c}$, $Sim^V_{a,d}$, $Sim^V_{a,e}$, $Sim^V_{a,f}$	$DScale_{a,b}$, $DScale_{a,c}$, $DScale_{a,d}$, $DScale_{a,d}$, $DScale_{a,d}$	N_{Agree}, $N_{Disagree}$, N_{Noidea}
1	1.00, 1.00, 1.00, 1.00, 1.00	2, 4, 8, 16, 32	50, 0, 0
2	1.00, 1.00, 1.00, 1.00, 1.00	2, 4, 8, 16, 32	50, 0, 0
3	1.00, 1.00, 1.00, 1.00, 1.00	2, 4, 8, 16, 32	50, 0, 0
4	0.87, 0.64, 0.38, 0.38, 0.38	2, 4, 8, 16, 32	50, 0, 0
5	0.91, 0.78, 0.52, 0.44, 0.36	2, 4, 8, 16, 32	50, 0, 0
6	0.75, 0.55, 0.44, 0.35, 0.26	2, 4, 8, 16, 32	48, 0, 2

7	1.00, 1.00, 1.00, 1.00, 1.00	2, 4, 8, 16, 32	50, 0, 0
8	0.95, 0.88, 0.73, 0.65, 0.55	2.5, 10, 25, 50, 125	50, 0, 0
9	0.91, 0.82, 0.66, 0.52, 0.52	2.5, 10, 25, 50, 100	50, 0, 0
10	1.00, 0.55, .055, 0.55, 0.55	2.5, 10, 25, 50, 100	50, 0, 0
11	1.00, 1.00, 1.00, 1.00, 1.00	2.5, 5, 10, 25, 50	50, 0, 0
12	0.76, 0.57, 0.36, 0.21, 0.15	2, 5, 10, 25, 50	50, 0, 0
13	0.82, 0.62, 0.36, 0.19, 0.12	2, 5, 10, 25, 50	50, 0, 0
14	0.71, 0.58, 0.40, 0.18, 0.11	2, 5, 10, 25, 50	50, 0, 0
15	0.95, 0.88, 0.67, 0.45, 0.36	2, 5, 10, 25, 50	50, 0, 0
16	0.93, 0.83, 0.76, 0.51, 0.42	2, 5, 10, 25, 50	50, 0, 0
17	0.96, 0.86, 0.75, 0.55, 0.40	2, 5, 10, 25, 50	50, 0, 0
18	0.77, 0.52, 0.31, 0.22, 0.18	2, 5, 10, 25, 50	50, 0, 0
19	0.75, 0.55, 0.37, 0.28, 0.19	2, 5, 10, 25, 50	49, 0, 1
20	0.68, 0.49, 0.34, 0.28, 0.16	2, 5, 10, 25, 50	48, 0, 2
21	0.82, 0.55, 0.27, 0.21, 0.17	2.5, 5, 10, 50, 100	47, 0, 3
22	0.63, 0.49, 0.32, 0.22, 0.15	2.5, 5, 10, 50, 100	49, 0, 1
23	0.74, 0.56, 0.29, 0.23, 0.15	2.5, 5, 10, 50, 100	48, 0, 2
24	0.68, 0.38, 0.31, 0.16, 0.16	2.5, 5, 10, 25, 50	49, 0, 1
25	0.82, 0.58, 0.33, 0.21, 0.15	2.5, 5, 10, 25, 50	50, 0, 0
26	0.85, 0.51, 0.31, 0.22, 0.14	2.5, 5, 10, 25, 50	48, 0, 2
27	0.74, 0.47, 0.29, 0.25, 0.14	2.5, 5, 10, 25, 50	50, 0, 0
28	0.88, 0.76, 0.61, 0.44, 0.28	2, 5, 10, 20, 50	50, 0, 0
29	0.74, 0.57, 0.55, 0.38, 0.21	2.5, 5, 10, 25, 50	50, 0, 0
30	0.85, 0.72, 0.65, 0.46, 0.22	2.5, 5, 10, 25, 50	50, 0, 0
31	0.53, 0.39, 0.23, 0.22, 0.15	2.5, 5, 10, 25, 50	50, 0, 0
32	0.82, 0.67, 0.46, 0.33, 0.18	2, 5, 10, 25, 50	50, 0, 0
33	0.80, 0.69, 0.47, 0.27, 0.17	2, 5, 10, 25, 50	50, 0, 0
34	0.88, 0.68, 0.46, 0.39, 0.21	2, 5, 10, 25, 50	50, 0, 0

Notes: “the figure” in following variables refers to Figures 5-1 to 5-34.

$DScale_{a,b}$: map scale change from (a) to (b) in the figure.

$DScale_{a,c}$: map scale change from (a) to (c) in the figure.

$DScale_{a,d}$: map scale change from (a) to (d) in the figure.

$DScale_{a,e}$: map scale change from (a) to (e) in the figure.

$DScale_{a,f}$: map scale change from (a) to (f) in the figure.

N_{Agree} : the number of the subjects that can accept the three similarity degrees calculated by the new model.

$N_{Disagree}$: the number of the subjects that disagree with the three similarity degrees calculated by the new model.

N_{Noidea} : the number of the subjects that have no idea about the three similarity degrees calculated by the new model.

A number of insights can be gained from the statistical data listed in Table 5-1, Table 5-2 and the experiments.

First, similarity degrees are closely related to map scale change. It is obvious from Table 5-2 that the similarity degrees increase with the corresponding map scale changes in any of the experiments. The smaller the similarity degree between two objects/maps, the bigger the map scale change. This conclusion can also be easily drawn from the similarity degrees obtained from the subjects in the experiments.

Second, people are accustomed to describing spatial similarity relations qualitatively and fuzzily; however, quantitative spatial similarity relations do exist and are used in many communities such as cartography, environment and geography. People sometimes describe spatial similarity degrees or compare the degree of similarity between spatial objects using accurate values (for example, somebody may say: “this small building is 20% similar to that tall one but 90% similar to that short one”).

Third, each of the percentages of the subjects that agree with the similarity degrees calculated by the new models is between 94% and 100%. Therefore, the ten new models are acceptable to the majority of people in the experiments.

Fourth, average deviation between the similarity degrees calculated by the new models and that given by the subjects is 0.045, which shows that the similarity degrees calculated by the new models are high accuracy.

The average deviation is calculated by Formula 5-1.

$$\overline{D} = \sum_{i=b}^{f} abs(Sim_{a,i}^{V} - Sim_{a,i}^{E}) / (34 \times 5) \qquad 5\text{-}1$$

where, $i = b, c, d, e, f$.

Last, the new models are tested selecting 50 experienced cartographers as subjects, which makes the experiments go easily. On the other hand, it limits the varieties of the subjects and therefore decreases the credibility of the experimental results.

5.6 Chapter Summary

This chapter aims at validating the new models.

Firstly, it introduces the four basic approaches generally used for validation of simulation models, including the approach depending on the model development team, the approach depending on the model users and the model development team, the approach depending on a third party, and the score model.

Secondly, it proposes the four strategies for invalidating new models, which include Strategy 1: theoretical justifiability, Strategy 2: third party involvement, Strategy 3: comparison with existing approaches, and Strategy 4: experts' participation. Because Strategies 1 and 2 have been addressed in previous sections, it emphasizes on the other two strategies.

Therefore, it then gives a design of a series of psychological experiments, and presents a number of samples to do the experiments taking many experienced cartographers as subjects. The subjects are required to tell if the similarity degrees calculated by the new models are acceptable. In addition, they need to tell the similarity degrees between the spatial objects.

Finally, the data from the experiments are analyzed, and some conclusions are drawn.

Chapter 6 Applications of Spatial Similarity Relations in Map Generalization

It has been mentioned in Chapter 1 that this study mainly aims at solving three problems: (1) fundamental theory of spatial similarity relations in multi-scale map spaces, (2) calculation approaches/models/measures of spatial similarity relations in multi-scale map spaces, and (3) application of the theories of spatial similarity relations in automated map generalization. Now that the first two problems have been touched, it is pertinent to address the third one which includes the following three important issues.

(1) to find an approach to determine the relations between spatial similarity degree and map scale change in map generalization;

(2) to find an approach to determine when to terminate a map generalization algorithm/procedure; and

(3) to find an approaches to calculate the threshold values of a specific algorithm. Here, the threshold values refer to those depend on spatial similarity degrees of the corresponding objects at different map scales but are input by human-interruption while the algorithm is executed in a map generalization system.

6.1 Relations between Map Scale Change and Spatial Similarity Degree

Previous psychological experiments have discovered that similarity degree increases with map scale change, but a quantitative description of their relations is unknown yet. Therefore, the following sections focus on this problem and try to solve it using mathematical methods. Because ten models have been proposed for the corresponding ten types of objects in map generalization, they need to be researched, respectively.

6.1.1 Description of the Problem

Suppose that there is a map M_0 at scale S_0, it is generalized to produce N maps M_1, M_2,…, M_N at scales S_1, S_2,…, S_N, respectively, and $S_1 > S_2 > ... > S_N$. $C_i^{scale} = \frac{S_0}{S_i}$ is the map scale change from map M_0 to map M_i, and S_i is the scale of map M_i; Sim_{M_0,M_i} is the similarity degree between map M_0 and map M_i, where, $i=1,2,...,N$.

The question is: how to get a quantitative relation between C_i^{scale} and Sim_{M_0,M_i}? This question can be divided into two parts: "if C_i^{scale} is known, how to obtain Sim_{M_0,M_i}?" and "if Sim_{M_0,M_i} is known, how to obtain C_i^{scale}?" Each of them corresponds with an expression that regards Sim_{M_0,M_i} and C_i^{scale} as the independent variables, respectively.

$$Sim_{M_0,M_i} = f(C_i^{scale}) \quad 6\text{-}1$$

$$C_i^{scale} = g(Sim_{M_0,M_i}) \quad 6\text{-}2$$

Some applications of the two expressions can be found in the communities of cartography and geographical information science. For example, decision-makers (e.g. urban planners) are often seen using a number of maps of an area at different scales in order to get different levels of detail of the region. They may say: "how similar the maps are!" Nevertheless, many decision-makers even do not know what quantitative similarity is, let alone to tell the similarity degrees between the maps at multiple scales.

In academic research work, such as map generalization, as well as in our daily life, Expression 6-1 is much more popularly used than Expression 6-2, because people usually know the map scales (i.e. C_i^{scale}) but seldom know the similarity degrees (i.e. Sim_{M_0,M_i}). This situation is popular in automated map generalization.

Hence, the following sections will focus on Expression 6-1.

6.1.2 Conceptual Framework for Solving the Problem

To simplify the expression, let $x = C_i^{scale}$ and $y = Sim_{M_0,M_i}$. Expression 6-1 is transformed to

$$y = f(x) \quad 6\text{-}3$$

In Table 5-2, it is easy to get that each experiment provides five pairs of x, y that can be viewed as five pairs of coordinates in the Cartesian coordinate system, i.e. ($Sim_{a,b}^V$, $DScale_{a,b}$),

($Sim^{V}_{a,c}$, $DScale_{a,c}$), ($Sim^{V}_{a,d}$, $DScale_{a,d}$), ($Sim^{V}_{a,e}$, $DScale_{a,e}$), and ($Sim^{V}_{a,f}$, $DScale_{a,f}$). For example, the five pairs of coordinates in the experiment 5 are (0.91, 0.500), (0.78, 0.250), (0.52, 0.125), (0.44, 0.0625), and (0.36, 0.03125). Three or four experiments are employed to test each category of objects in the experiments; hence each category of objects has 15 or 20 pairs of coordinates.

To find the relation between $x = C_i^{scale}$ and $y = Sim_{M_0,M_i}$ means to get formulas that can calculate y by x. Because the relation between them is apparently can be expressed using empirical formulae, the curve fitting approach is employed to construct formulae using the experimental data for the ten categories of objects.

Curve fitting is a process of constructing a curve or a mathematical function that has the best fit to a series of data points (Kolb, 1984; Arlinghaus, 1994). Fitted curves should capture the trend in the data across the entire range, and can be used as an aid for data visualization to infer values of the function where no data are available and to summarize the relationships among two or more variables. Thus, it may be employed to substantiate Function 6-3.

The curve fitting employed here comprises the following three steps, which is addressed in detail before it is put into use.

(1) Determine the data points that are used in the curve fitting.

All of the data points obtained from the experiments may be adopted. In addition, a special point (1.000, 1.000) can be added in the point set obtained from the experiments for each category of objects. This point refers the situation that a map, an object, or an object group is totally similar to itself; thus, its similarity is 1.00 and its map scale change is 1.00, too.

(2) Select some functions as candidates.

An infinite number of generic forms of functions can be chosen as candidates for almost any shape curves. It is not easy to select an appropriate function from numerous candidates to fit a series of points, because an inappropriate candidate may be either under-fit or over-fit.

Potential candidate functions usually used in curve fitting comprise polynomials, power functions, logarithmic functions, and exponential functions. Previous experiments have revealed that the candidate functions should be monotonic decrease functions, so only 1st and 2nd order polynomials can be considered, because the other polynomials (e.g. 3rd and 4th order polynomials) have $n-2$ (n is the order of the polynomial) inflection point(s) which indicates that the curve is not monotonic. Hence, the following functions will be taken into account.

$$y = a_1 x + a_0 \tag{6-4}$$

$$y = a_2 x^2 + a_1 x + a_0 \tag{6-5}$$

$$y = a_2 e^{a_1 x} + a_0 \tag{6-6}$$

$$y = a_1 \ln(\mathrm{x}) + \mathrm{a}_0 \tag{6-7}$$

$$y = x^a \tag{6-8}$$

(3) Calculate the coefficient(s) of each function.

The least square method (Lanczos, 1988), a widely use method, is used to pick the coefficient(s) of each function that best fits the curve to the data points.

(4) Compare the functions to determine the best fit one.

R^2, i.e. R-squared, is usually used to compare the candidate functions. The greater an R^2, the better its corresponding curve. Thus, the curve with the greatest R^2 among all of the candidates is the best curve fitting the point set.

R^2 can be calculated by the following method.

There is a function $y = f(x)$. Its dependent variable y has n modeled/predicted values $\hat{y}_i$ and n observed values y_i. Here, $\mathrm{i}=1, 2,...,\mathrm{n}$.

$\bar{y}$ is the mean of the observed data:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$, where, n is the number of observations.

The "variability" of the data set is measured through different sums of squares:

$SS_{Total} = \sum_{i=1}^{n}(y_i - \bar{y})^2$: the total sum of squares (proportional to the sample variance);

$SS_{\text{Regression}} = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2$: the regression sum of squares, also called the explained sum of squares; and

$SS_{\text{Residual}} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$: the sum of squares of residuals, also called the residual sum of squares.

The most general definition of the coefficient of determination is:

$$R^2 \equiv 1 - \frac{SS_{\text{Regression}}}{SS_{Total}} \quad 6\text{-}9$$

R^2 is a statistic that gives some information about the "goodness" of fit of a model. In regression, the R^2 coefficient of determination is a statistical measure of how well the regression line approximates the real data points. An R^2 of 1 indicates that the regression line perfectly fits the data.

6.2 Formulae for Map Scale Change and Spatial Similarity Degree

The formula for each of the ten types of objects is constructed using the method discussed in Section 6.1.2, and is implemented by means of Microsoft EXCEL (Version 2010).

The following three steps are carried in determining each of the ten formulae.

(1) Point used: present the points used in the curve fitting. The coordinates are from the table 5-2.

(2) Candidate curves: illustrate the candidate curves, the corresponding formulae and R^2 .

(3) formula: present the selected formula directly, and give a short explanation if needed.

6.2.1 Individual Point Objects

- Points used (6 points, obtained from the data of Experiments 1, 2 and 3 listed in Table 5-2)

  (1, 1.00)

  (2, 1.00), (4, 1.00), (8, 1.00), (16, 1.00), (32, 1.00).

- Candidate curves

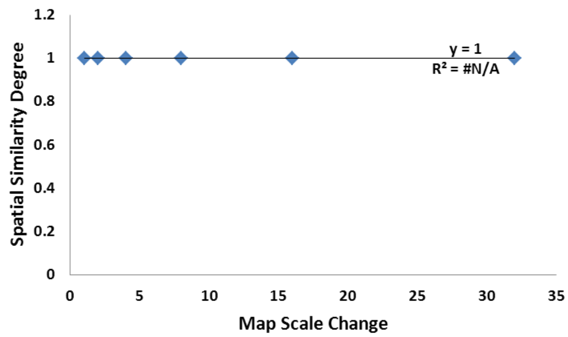


Figure 6-1 Curve fitting for individual points.

The feature of the point set is too obvious to describe using other curves but a horizontally straight line, because all of the coordinates of y are equal to 1.

• Formula

$$y=1 \tag{6-10}$$

6.2.2 Individual Linear Objects

• Points used (21 points, obtained from the data of Experiments 4, 5, 6 and 7 listed in Table 5-2)

(1, 1.00)

(2, 0.87), (4, 0.64), (8, 0.38), (16, 0.38), (32, 0.38),

(2, 0.91), (4, 0.78), (8, 0.52), (16, 0.44), (32, 0.36),

(2, 0.75), (4, 0.55), (8, 0.44), (16, 0.35), (32, 0.26),

(2, 1.00), (4, 1.00), (8, 1.00), (16, 1.00), (32, 1.00).

• Candidate curves

Totally 21 points are taken into account. 16 points are used in the curve fitting and the other five points (the last line in the point set) are considered separately, because they are from the experiment that tested on a horizontal line segment, and the five resulting points are collinear.

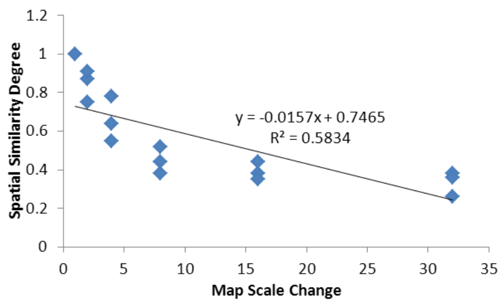
The five candidate curves are shown in Figure 6-2.

• Formula

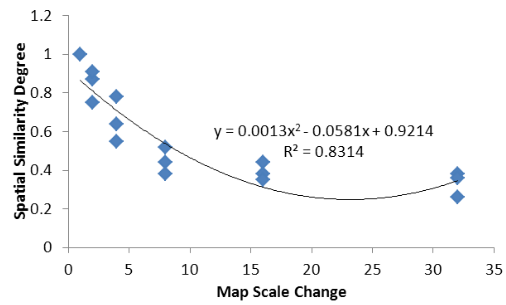
Function $y=1.0164x^{-0.343}$ should be selected, because its corresponding R^2 is the greatest.

Considering the special case, the function should be:

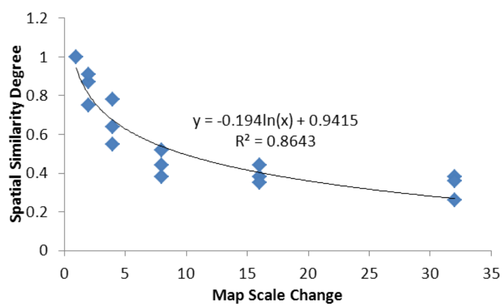
$$y=\begin{cases}1, \text{ if the original line is a straight line, else} \\ 1.0164x^{-0.343}\end{cases} \quad \text{6-11}$$



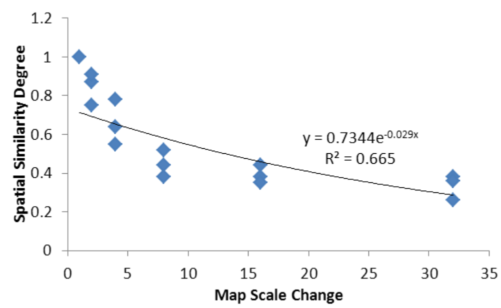
(a)



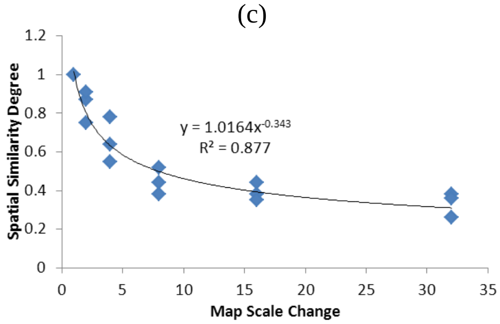
(b)



(c)



(d)



(e)

Figure 6-2 Curve fitting for individual linear objects.

6.2.3 Individual Areal Objects

• Points used (21 points, obtained from the data of Experiments 8, 9, 10 and 11 listed in Table 5-2)

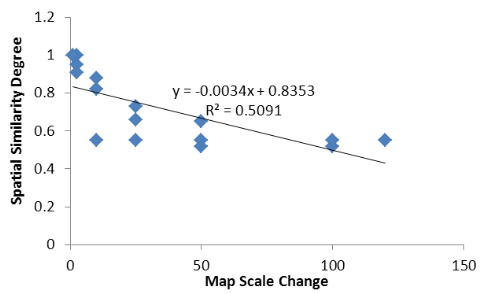
(1, 1.00)

(2.5, 0.95), (10, 0.88), (25, 0.73), 50, 0.65), (125, 0.55),

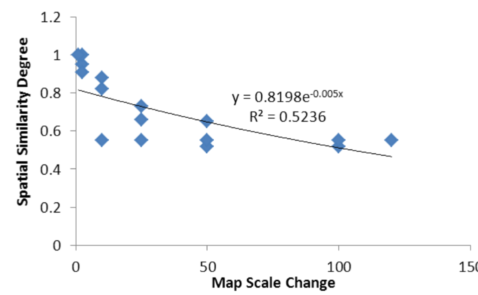
(2.5, 0.91), (10, 0.82), (25, 0.66), 50, 0.52), (100, 0.52),

(2.5, 1.00), (10, 0.55), (25, 0.55), 50, 0.55), (100, 0.55),

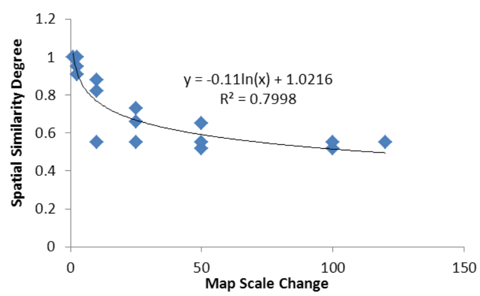
(2.5, 1.00), (5, 1.00), (10, 1.00), (25, 1.00), 50, 1.00).



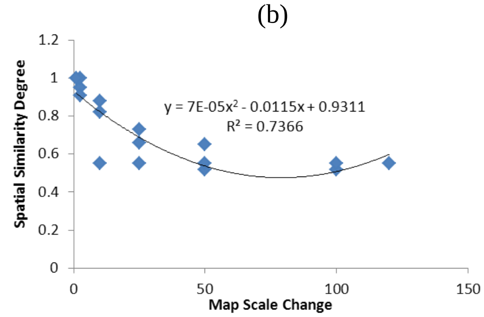
(a)



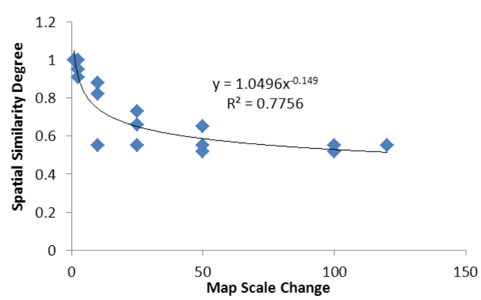
(b)



(c)



(d)

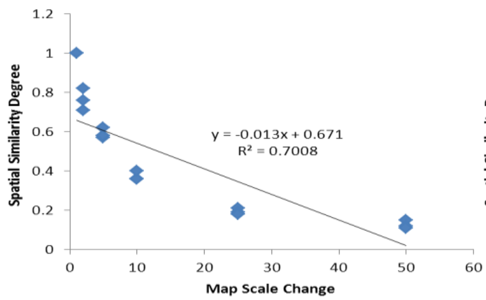


(e)

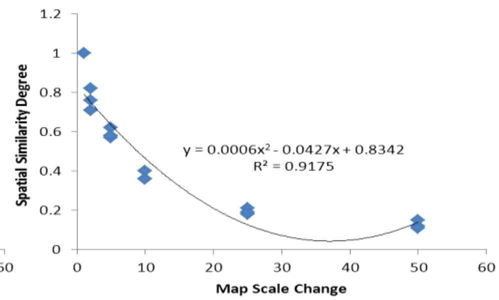
Figure 6-3 Curve fitting for individual areal objects.

• Candidate curves

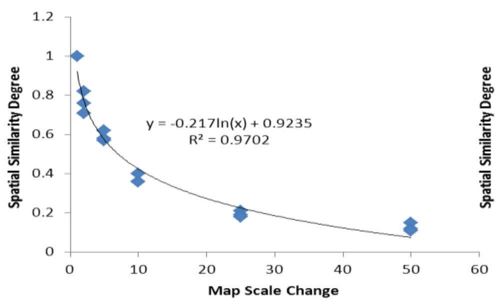
Totally 21 points are taken into account. 16 points are used in the curve fitting and the other five points (the last line in the point set) are considered separately, because they are from the experiment that tested on a polygon (a square-shaped building) that does not need to be simplified at any scale.



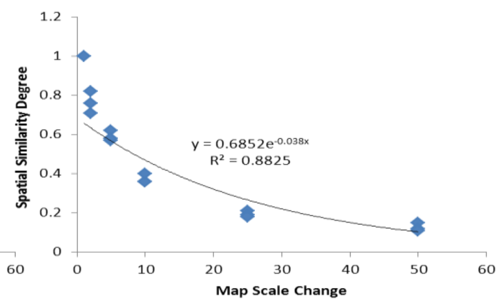
(a)



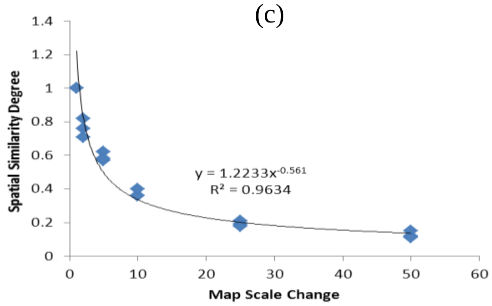
(b)



(c)



(d)



(e)

Figure 6-4 Curve fitting for point clouds.

The five candidate curves are shown in Figure 6-3.

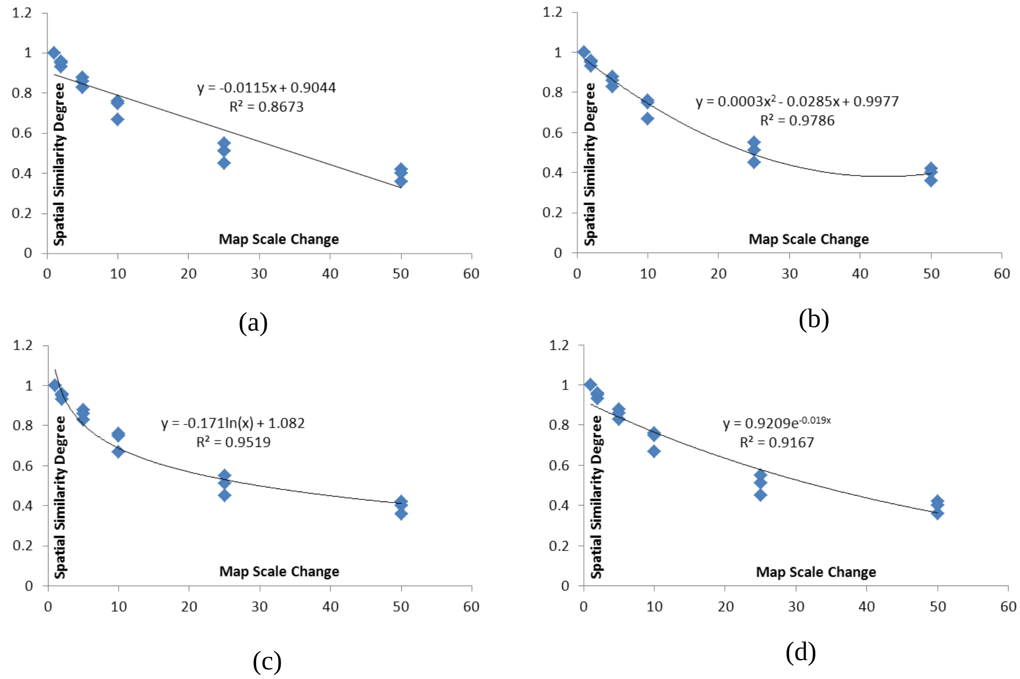
- Formula

Function y=-0.11ln(x)+1.0216 should be selected, because its corresponding $R^2 = 0.7998$ is the greatest.

Considering the special case, the function should be:

$$y=\begin{cases}1, \text{ if the original polygon is a square, else} \\ -.011\ln(x)+1.0216\end{cases} \quad \text{6-12}$$

6.2.4 Point Clouds



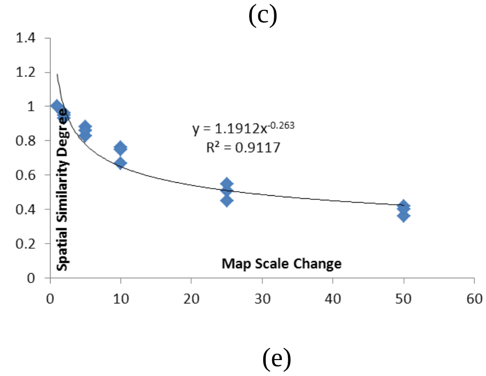


Figure 6-5 Curve fitting for parallel line clusters.

• Points used (16 points, obtained from the data of Experiments 12, 13 and 14 listed in Table 5-2)

(1, 1.00)

(2, 0.76), (5, 0.57), (10, 0.36), (25, 0.21), (50, 0.15),

(2, 0.82), (5, 0.62), (10, 0.36), (25, 0.19), (50, 0.12),

(2, 0.71), (5, 0.58), (10, 0.40), (25, 0.18), (50, 0.11).

• Candidate curves

The five candidate curves are shown in Figure 6-3.

• Formula

The resulting function should be

$$y\text{=-}0.217\ln(x)+0.9235 \qquad \text{6-13}$$

because its corresponding $R^2 = 0.9702$ is the greatest in the five R^2 of the candidate curves.

6.2.5 Parallel Line Clusters

• Points used (16 points, obtained from the data of Experiments 15, 16 and 17 listed in Table 5-2)

(1, 1.00)

(2, 0.95), (5, 0.88), (10, 0.67), (25, 0.45), (50, 0.36),

(2, 0.93), (5, 0.83), (10, 0.76), (25, 0.51), (50, 0.42),

(2, 0.96), (5, 0.86), (10, 0.75), (25, 0.55), (50, 0.40).

• Candidate curves

The five candidate curves are shown in Figure 6-5.

• Formula

The resulting function should be

$$y\text{=}0.0003x^2 - 0.0285x + 0.9977 \qquad \text{6-14}$$

because its corresponding $R^2 = 0.9786$ is the greatest in the five R^2 of the candidate curves.

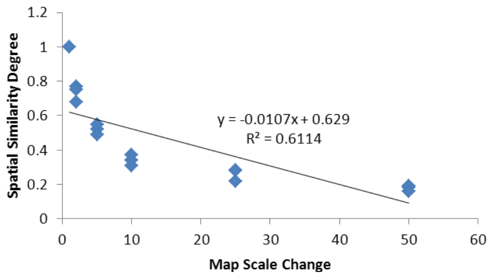
6.2.6 Intersected Line Networks

• Points used (16 points, obtained from the data of Experiments 18, 19 and 20 listed in Table 5-2)

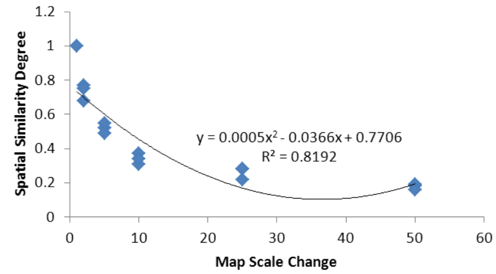
(1, 1.00)

(2, 0.77), (5, 0.52), (10, 0.31), (25, 0.22), (50, 0.18),

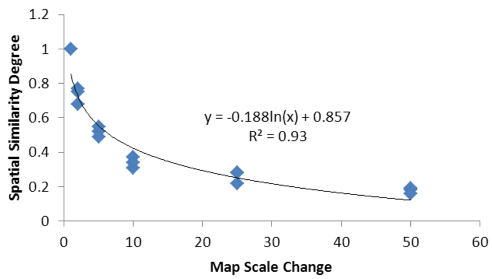
(2, 0.75), (5, 0.55), (10, 0.37), (25, 0.28), (50, 0.19), (2, 0.68), (5, 0.49), (10, 0.34), (25, 0.28), (50, 0.16).



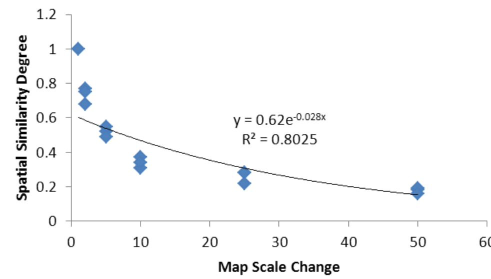
(a)



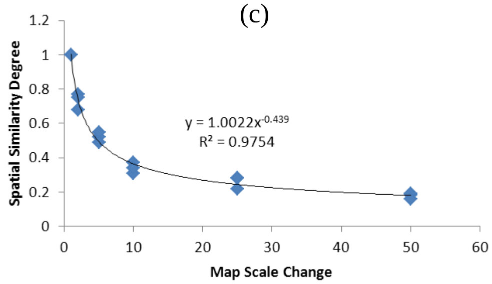
(b)



(c)



(d)



(e)

Figure 6-6 Curve fitting for intersected line networks.

- Candidate curves

The five candidate curves are shown in Figure 6-6.

- Formula

The resulting function should be

$$y=1.0022x^{-0.439} \qquad \text{6-15}$$

because its corresponding $R^2 = 0.9754$ is the greatest in the five R^2 of the candidate curves.

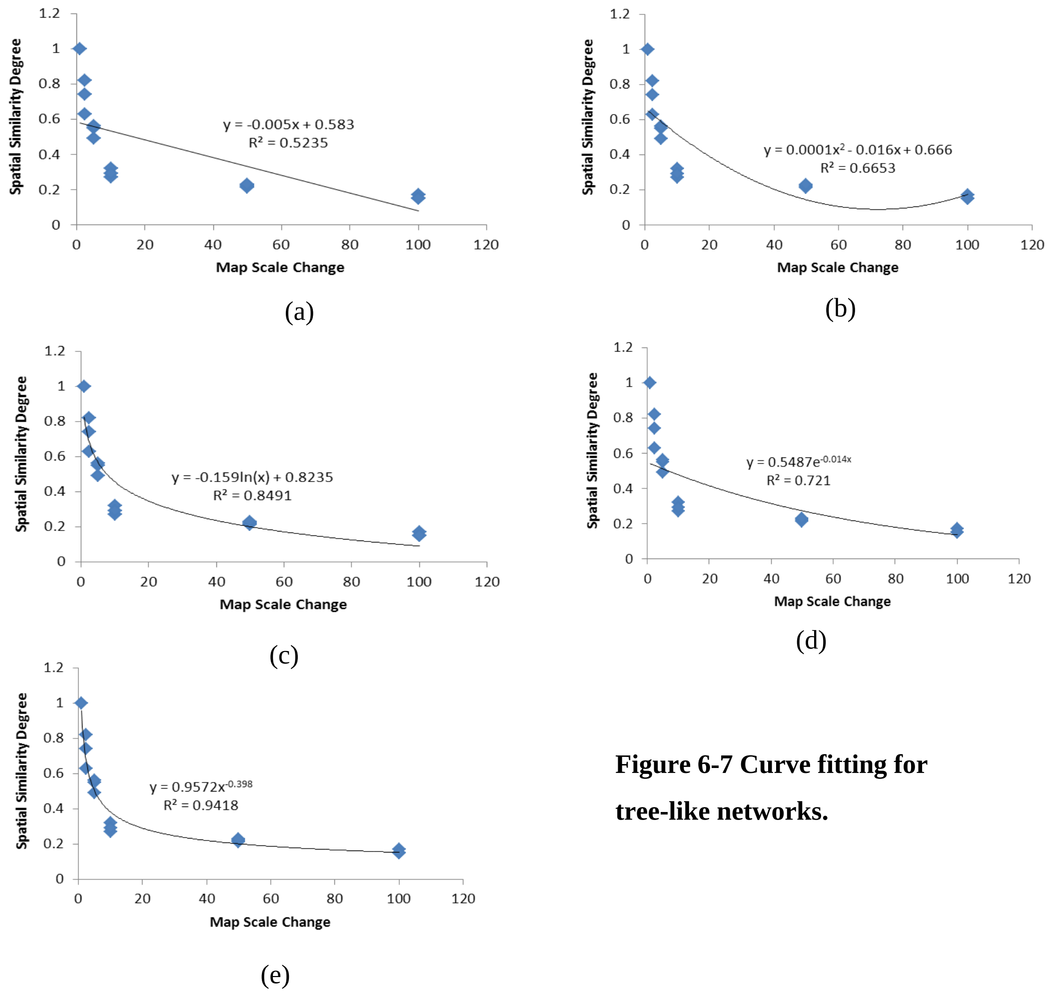


Figure 6-7 Curve fitting for tree-like networks.

6.2.7 Tree-like Networks

- Points used (16 points, obtained from the data of Experiments 21, 22 and 23 listed in Table 5-2)

  (1, 1.00)

  (2.5, 0.82), (5, 0.55), (10, 0.27), (50, 0.21), (100, 0.17),

  (2.5, 0.63), (5, 0.49), (10, 0.32), (50, 0.22), (100, 0.15),

  (2.5, 0.74), (5, 0.56), (10, 0.29), (50, 0.23), (100, 0.15).

- Candidate curves

  The five candidate curves are shown in Figure 6-7.

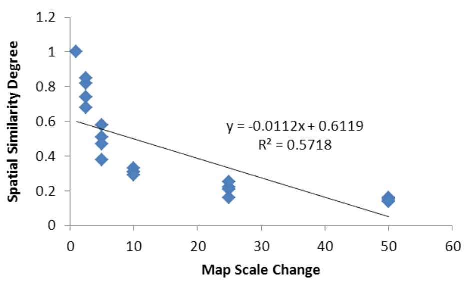
• Formula

The resulting function should be

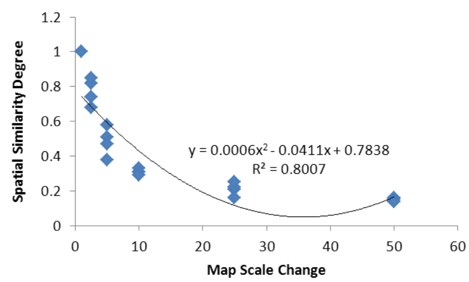
$$y = 0.9572x^{-0.398} \qquad \text{6-16}$$

because its corresponding R^2 is the greatest.

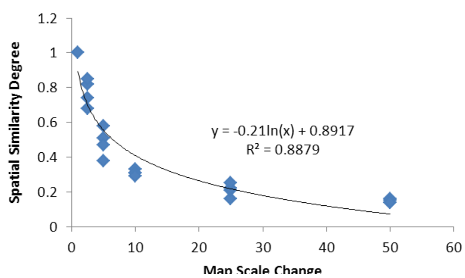
6.2.8 Discrete Polygon Groups



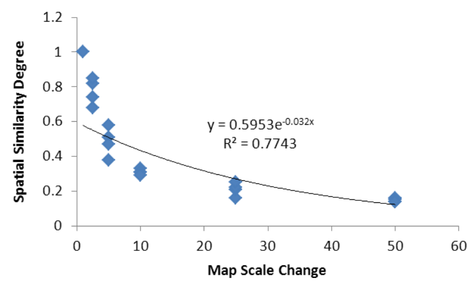
(a)



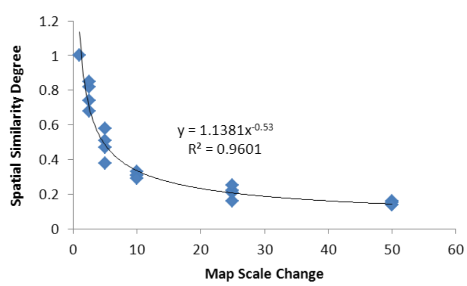
(b)



(c)



(d)



(e)

Figure 6-8 Curve fitting for discrete polygon groups.

• Points used (21 points, obtained from the data of Experiments 24, 25, 26 and 27 listed in Table 5-2)

(1, 1.00)

(2.5, 0.68), (5, 0.38), (10, 0.31), (25, 0.16), (50, 0.16), (2.5, 0.82), (5, 0.58), (10, 0.33), (25, 0.21), (50, 0.15),

(2.5, 0.85), (5, 0.51), (10, 0.31), (25, 0.22), (50, 0.14),

(2.5, 0.74), (5, 0.47), (10, 0.29), (25, 0.25), (50, 0.14).

- Candidate curves

The five candidate curves are shown in Figure 6-8.

- Formula

The resulting function should be

$$y = 1.1381x^{-0.53} \qquad 6\text{-}17$$

because its corresponding $R^2 = 0.9601$ is the greatest in the five R^2 of the candidate curves.

6.2.9 Connected Polygon Groups

- Points used (16 points, obtained from the data of Experiments 28, 29 and 30 listed in Table 5-2)

(1, 1.00)

(2, 0.88), (5, 0.76), (10, 0.61), (20, 0.44), (50, 0.28),

(2.5, 0.74), (5, 0.57), (10, 0.55), (25, 0.38), (50, 0.21),

(2.5, 0.85), (5, 0.72), (10, 0.65), (25, 0.46), (50, 0.22).

- Candidate curves

The five candidate curves are shown in Figure 6-9.

- Formula

The resulting function should be

$$y = -0.187\ln(x) + 0.9973 \qquad 6\text{-}18$$

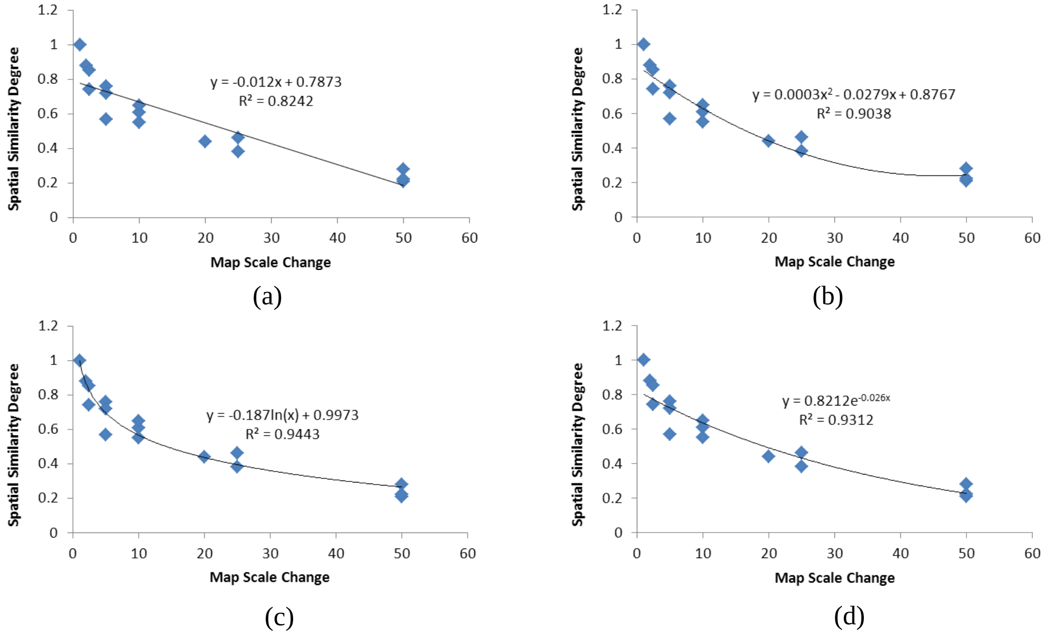
because its corresponding $R^2 = 0.9443$ is the greatest in the five R^2 of the candidate curves.

6.2.10 Maps

- Points used (21 points, obtained from the data of Experiments 31, 32, 33 and 34 listed in Table 5-2)

(1, 1.00)

(2.5, 0.53), (5, 0.39), (10, 0.23), (25, 0.22), (50, 0.15),



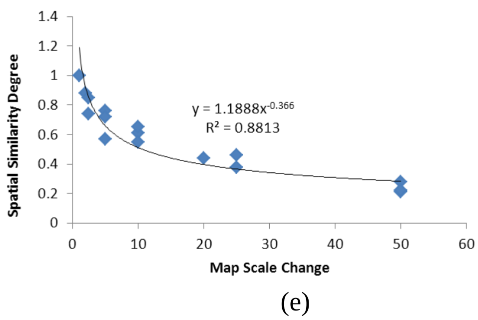


Figure 6-9 Curve fitting for connected polygon groups.

(2, 0.82), (5, 0.67), (10, 0.46), (25, 0.33), (50, 0.18),

(2, 0.80), (5, 0.69), (10, 0.47), (25, 0.27), (50, 0.17),

(2, 0.88), (5, 0.68), (10, 0.46), (25, 0.39), (50, 0.21).

- Candidate curves

The five candidate curves are shown in Figure 6-10.

- Formula

The resulting function should be

y=-0.194ln(x)+0.9118 6-19

because its corresponding $R^2 = 0.8502$ is the greatest in the five R^2 of the candidate curves.

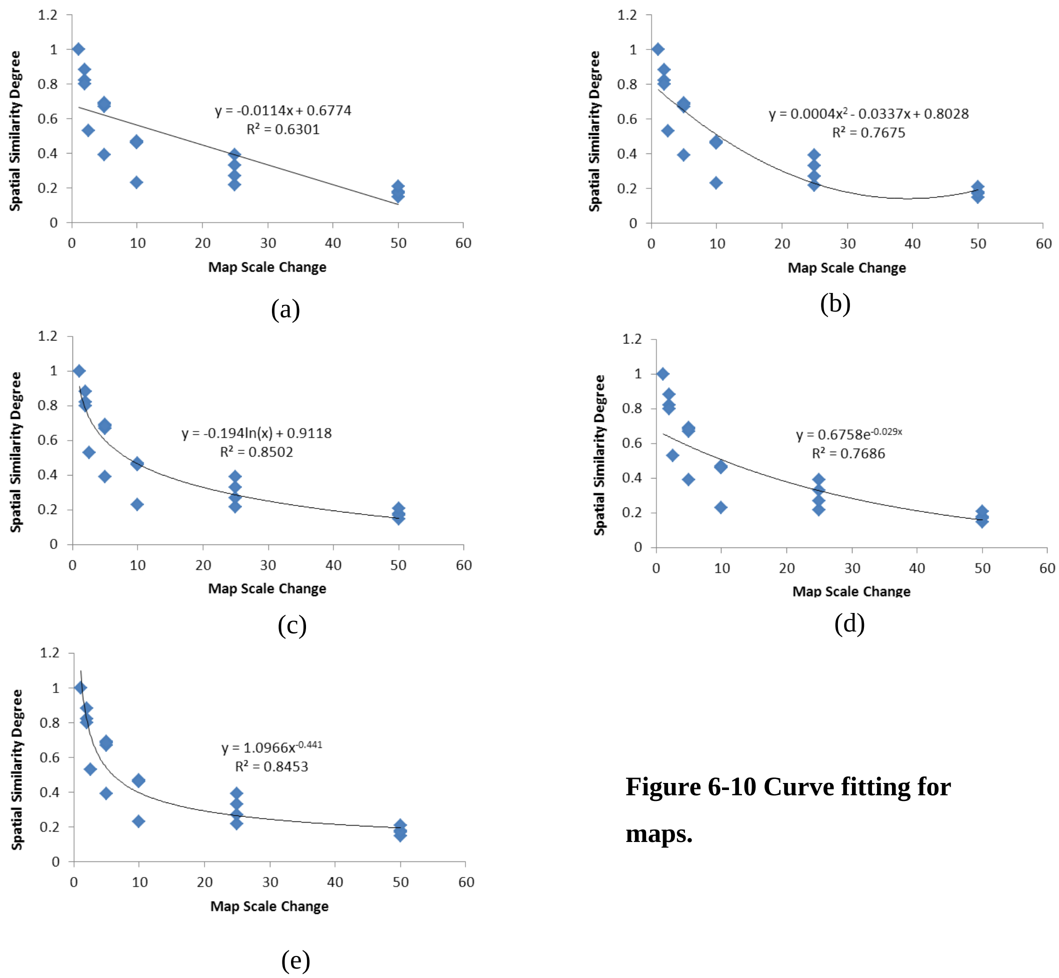


Figure 6-10 Curve fitting for maps.

6.3 Discussion about the Formulae

Table 6-1 Formulae for calculating spatial similarity degrees using map scale changes.

Type of objects	Type of the formula	Formula
Individual point objects	Polynomial (Linear)	$y=1$
Individual linear objects	Power	$y=\begin{cases} 1, \text{ if the original line is a straight line, else} \\ 1.0164x^{-0.343} \end{cases}$

Individual areal objects	Logarithmic	$y=\begin{cases} 1, \text{ if the original polygon is a square, else} \\ -.011\ln(x)+1.0216 \end{cases}$
Point clouds	Logarithmic	y=-0.217ln(x)+0.9235
Parallel line clusters	Polynomial	$y=0.0003x^2-0.0285x+0.9977$
Intersected line networks	Power	$y=1.0022x^{-0.439}$
Tree-like networks	Power	$y=0.9572x^{-0.398}$
Discrete polygon groups	Power	$y=1.1381x^{-0.53}$
Connected polygon groups	Logarithmic	y=-0.187ln(x)+0.9973
Maps	Logarithmic	y=-0.194ln(x)+0.9118

Some insights and conclusions can be gained from the formulae for calculating the relations between map scale change and spatial similarity degree.

(1) There are four logarithmic functions, four power function and two polynomials (the linear function can be viewed as a special case of polynomials) in the selected functions; but no exponential function is used (Table 6-1). Thus, it is necessary but difficult to provide an identical formula for different types of objects.

(2) The ten formulae in Table 6-1 can be used to calculate the spatial similarity degree (y) if the map scale change (x) between an original map and a generalized map are given.

On the other hand, the corresponding inverse functions of the ten formulae can be obtained, which can be used to calculate the map scale change between a map and its generalized version if their spatial similarity degree is known.

(3) The domain of the ten formulae is identical, i.e. $x \in (1, \infty)$; and their corresponding range is also identical, i.e, $y \in [0,1]$.

(4) The formulae can be used to interpolate any values belonging to the domain (and belonging to the range if the invers functions are used), though the formulae only have been experimented by a few commonly used map scales.

For example, there is a road network at scale 1:1000, if it is generalized to get three maps at scale 1:1950, 1:5650, and 1:270000, respectively. The spatial similarity degrees can be calculated by formula y=1.0022x$^{-0.439}$.

For the map at scale 1:1950,

$$y=1.0022x^{-0.439}=1.0022\times1.950^{-0.439}\approx0.748.$$

For the map at scale 1:5650,

$$y=1.0022x^{-0.439}=1.0022\times5.650^{-0.439}\approx0.469.$$

For the map at scale 1:270000,

$$y=1.0022x^{-0.439}=1.0022\times270^{-0.439}\approx0.086.$$

(5) The formulae are based on limited number of psychological experiments. Hence, they can be "adjusted" by using more samples in the experiments.

6.4 Approach to Automatically Terminate a Procedure in Map Generalization

Map generalization is a process that simplifies an original map for the purpose of producing a smaller scale map. In semi-automated map generalization, this process is implemented by a series of algorithms. The map is usually divided into many feature layers, and each feature layer is generalized by one or more algorithms.

As is well known, each algorithm is a simulation of cartographers' work in map simplification, which means it generalizes the corresponding map feature layer gradually and tentatively, and presents intermediate maps one by one to cartographers to determine which one is satisfactory and if the generalization can be terminated. The disadvantage of this process is obvious: human's interference decreases the efficiency of map generalization, and increases the uncertainty of the resulting map (because it is possible for different cartographers to select different maps as the resulting map). A crucial reason for cartographers' determining when an algorithm can be terminated is that no appropriate methods have been developed for calculating spatial similarity degrees between maps and between map feature layers.

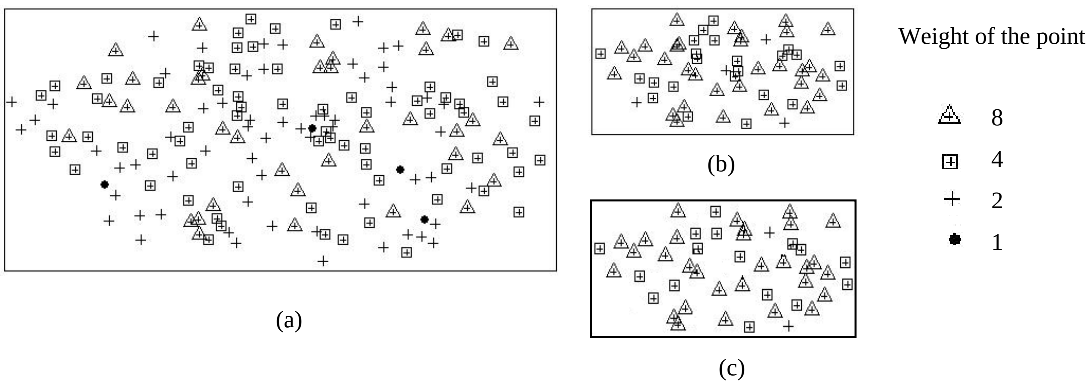


Figure 6-11 Demonstration of the point cloud generalization algorithm.

(a) a point cloud with 173 points at scale 1:10K. The number of points weighted 1 is 4 points, and the number of points weighted 2 is 63, and the number of points weighted 4 is 69, and the number of points weighted 8 is 37; (b) generalized point cloud at scale 1:100K with 58 points retained, among which the number of points weighted 2 is 4, and the number of points weighted 4 is 23, and the number of points weighted 8 is 31; and (c) generalized point cloud at scale 1:100K with 49 points retained, among which the number of points weighted 2 is 2, and the number of points weighted 4 is 18, and the number of points weighted 8 is 29.

Now that a series of models have been proposed for calculating spatial similarity degrees, an approach to automatically determining when to terminate an algorithm or a system composed of many algorithms in map generalization is proposed here.

Step 1: calculate the spatial similarity degree between the original objects/map and the resulting objects/map using the corresponding appropriate formula (i.e. Formula 6-10 to Formula 6-19).

Step 2: simplify the objects/map using the algorithm/system, which generates a series of intermediate objects/maps after each round of generalization. Calculate the spatial similarity degree between the original objects/map and the intermediate objects/map using the corresponding model proposed in Chapter 4. The spatial similarity degree between the original objects/map and the intermediate objects/map generated after the i^{th} round of generalization is called y_i.

Step 3: if $y_{\mathrm{i}} > y$, go to step 2;

else if In this case, the model that is adopted is formula 4-15, because the type of the generalized objects belongs to point clouds.

and $\mathrm{i}=1$, go to step 4;

else if $abs(y_{\mathrm{i}} - y) \geq abs(y_{\mathrm{i-1}} - y)$, the intermediate objects/map generated after the $(i\text{-}1)^{th}$ round of generalization is the resulting objects/map;

else, go to step 4.

Step 4: take the intermediate objects/map generated after the i^{th} round of generalization as the resulting objects/map, and end the procedure.

This approach can be demonstrated by means of simplifying a point cloud using the Voronoi-based algorithm (Yan & Weibel, 2008). Suppose that a point cloud map at scale 1:10K (Figure 6-11) is simplified using the Voronoi-based algorithm to generate a map at scale 1:100K. In this case, the model that should be adopted to calculate similarity degree taking map scale change as dependent variable is Formula 4-15, because the type of the generalized objects belongs to point clouds. Hence, the similarity degree can be obtained:

$$y\text{=-}0.217\ln(\mathrm{x})+0.9235\text{=-}0.217\ln(10)+0.9235 \approx 0.42$$

The point cloud is deleted by iteratively constructing Voronoi diagrams. It generates an intermediate point cloud after each round of deletion. The spatial similarity degree between the original point cloud and each intermediate point cloud is calculated using the corresponding model proposed in Chapter 4 and it is compared with y.

After the 5th round of deletion (Figure 6-11(c)), $y_5 = 0.38$, so $y_5 < y$.

According to the previous calculation using the same model for calculating y_5, y_4=0.44. Thus, it is clear $abs(y_5 - y) > abs(y_4 - y)$, and the resulting point cloud should be the one obtained after the 4th round of deletion (Figure 6-11(b)).

6.5 Calculation of the Distance Tolerance in the Douglas-Peucker Algorithm

To simplify geometry to suit the displayed resolution, various line simplification algorithms exist, while the Douglas-Peucker Algorithm is the most well-known (Ramer, 1972; Douglas & Peuker, 1973; Hershberger & Snoeyink, 1992). This algorithm is for reducing the number of points in a curve that is approximated by a series of points. The purpose of the algorithm is, given a curve composed of line segments, to find a similar curve with fewer points. The simplified curve consists of a subset of the points that defined the original curve.

6.5.1 The Douglas-Peucker Algorithm and Its Disadvantages

Given that the original curve is an ordered set of points or lines and the distance tolerance is $\varepsilon > 0$, the algorithm keeps/deletes points by recursively dividing the curve (Figure 6-12).

The algorithm firstly automatically marks the first and last point to be kept. Secondly, it finds the point (this point is called the worst point) that is furthest from the line segment with the first and last points as end points. If the distance from the point to the line segment is less than ε, then any points currently not marked to keep can be discarded. Otherwise, if the point furthest from the line segment is greater than ε from the approximation then that point must be kept. The algorithm recursively calls itself with the first point and the worst point and then with the worst point and the last point. When the recursion is completed a new output curve can be generated consisting of all (and only) those points that have been marked as kept.

Although the Douglas-Peucker Algorithm has been the most popular algorithm used in line simplification in map generalization, it is not a fully automatic algorithm, because cartographers often need to input the distance tolerance ε in the execution of the algorithm which decreases the efficiency of line simplification. Hence, it is of importance to find an approach to calculate ε.

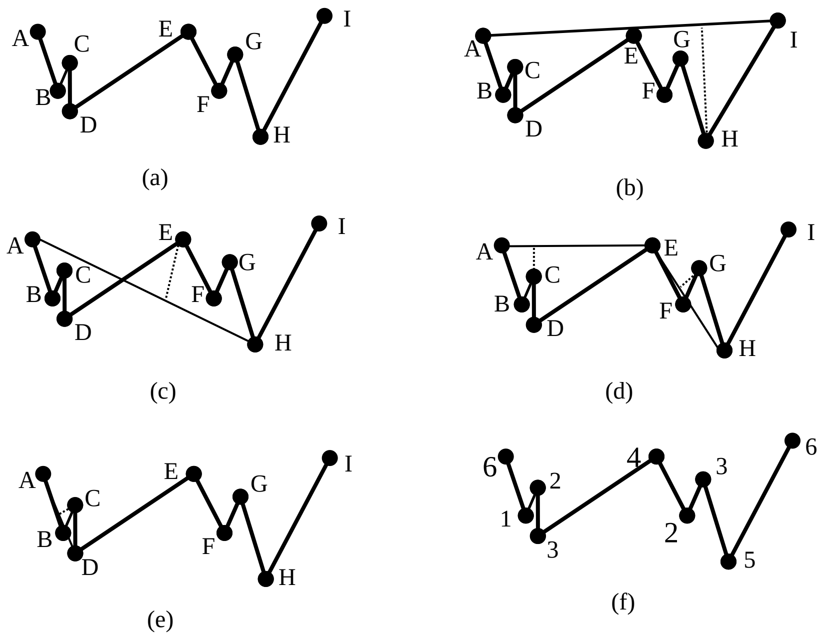


Figure 6-12 Principle of the Douglas-Peucker Algorithm.

(a) original curve; (b) link AI and keep points A, I and H, because A and I are the first point and the last point, and H is the farthest point to AI and the distance is greater than ε; (c) link AH and keep point E, because it is the farthest point to AE and the distance is greater than ε; (d) link AE and EH, and keep D and G, because they are the farthest points to AE and EH, respectively, and the two distances are greater than ε; (e) link AD and keep C, and link EG and keep F, because they are the farthest points to AD and EG, respectively, and the two distances are greater than ε; and (f) B is the last point to be kept, and the number, say m, beside each point denoting that this point can be deleted in the m^{th} round of deletion. Here, ε is supposed to tend to be 0 in order to demonstrate the algorithm clearly.

6.5.2 Approach to Calculating the Distance Tolerance for the Douglas-Peucker Algorithm

The problem can be described as follows: a series of digital line maps at a specific scale (say, S_0) are planned to be generalized to produce the maps at a given smaller scale (say, S_1) using the Douglas-Peucker Algorithm. How can the distance tolerance (ε) be obtained so that the execution of the Douglas-Peucker Algorithm becomes fully automatic?

The problem can be analyzed and solved in the following way.

Above all, a theoretical spatial similarity degree (say, y_T) can be calculated by Formula 6-11 because the map scale change (i.e. S_0 / S_1) can be obtained. Considering ε has no effects to straight lines in map generalization, y_T can be calculated by $y_T = 1.0164 \times (\frac{S_0}{S_1})^{-0.343}$.

Secondly, ε is an imperial and therefore fuzzy value, and it is the criterion for all curves simplification. On the other hand, ε is closely related to spatial similarity degrees and map scale changes in curve simplification. A greater ε means a greater map scale change and a smaller spatial similarity degree between the original and the simplified curve. In addition, ε does not serves for one curve. Therefore, it should be plausible to obtain ε by calculating its relations with spatial similarity degrees taking the curves on the original maps as samples.

Third, y_T is the criterion for evaluating if a curve simplified from another curve at scale S_0 is appropriate to appear on the map at scale S_1, which means the similarity degree (say, y_P) between the simplified curve and the original curve should be approximately equal to y_T. On the other hand, a simplified curve corresponds to a distance tolerance. Hence, if y_P can be obtained, its corresponding distance tolerance is possible to be calculated. According to previous work in Chapter 4, Formula 4-6 may be used to calculate y_P.

Last, because each simplified curve corresponds to a y_P, and each y_P corresponds a distance tolerance, it is reliable to select a number of curves so that an average value of a number of distance tolerance can be obtained as the resulting distance tolerance.

Suppose that $\varepsilon_o = 0$, N ($N > 1$) curves on the original maps are selected as samples. ε_o can be used as the substitute of ε to determine the order of point selection of each curve (e.g. Figure 6-12). Obviously, the order of point deletion can be got by the reversion of the point selection.

As far as a sample curve is concerned, after the order of point deletion is determined, a series of intermediate curves (e.g. Figure 6-13) may be formed when the original curve is gradually simplified according to the point deletion order calculated by the Douglas-Peucker Algorithm if the distance tolerance is 0.

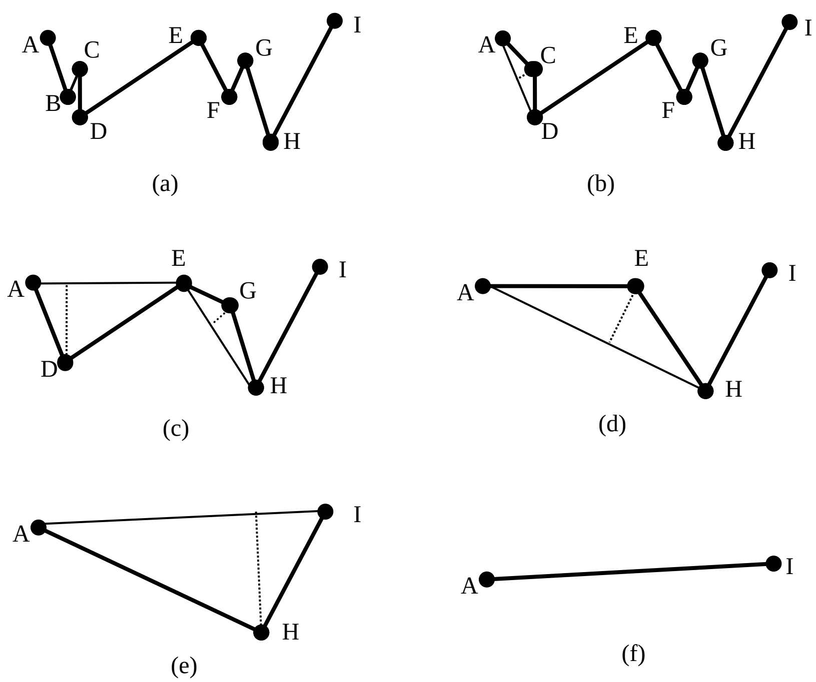


Figure 6-13 Gradual deletion of the points, taking Figure 6-12 as an example.

(a) original points; (b) B is deleted in the 1st round of deletion; (c) C and F are deleted in the 2nd round of deletion; (d) D and G are deleted in the 3rd round of deletion; (e) E is deleted in the 4th round of deletion; and (f) H is deleted in the last round of deletion, and only the first and the last points are retained.

After the k^{th} round of point deletion, the similarity degree (y_{P}^{k}) between the new curve and the original curve is calculated using Formula 4-16. If $y_{\mathrm{P}}^{k} > y_{T}$, continue with the next round of point deletion; else end point deletion procedure and determine which curve is more appropriate to be viewed as the resulting curve.

If $abs(y_{\mathrm{P}}^{k} - y_{T}) < abs(y_{\mathrm{P}}^{k-1} - y_{T})$ the curve formed after the k^{th} round of point deletion is the resulting curve; otherwise, the one formed after the $(k-1)^{th}$ round of point deletion is the resulting curve. The greatest distance (i.e. the dotted line in Figure 6-13) in the previous round of point deletion used to evaluate if a point can be retained is the distance tolerance. For example, if the curve in Figure 6-13(d) is the resulting curve, the length values of dotted lines in Figure 6-13(b) and Figure 6-13(c) are compared. Obviously, the length of the dotted line from point D to line AE is the distance tolerance.

If the distance tolerance of the i^{th} sample curve is ε_i , ε is the average of the N tolerance values obtained from the N sample curves.

$$\varepsilon = \frac{\sum_{i=1}^{N} \varepsilon_i}{N} \tag{6-20}$$

6.5.3 An Example for Testing the Approach

To test the approach for calculating the distance tolerance of the Douglas-Peucker Algorithm, a topographic map at scale 1: 100K is selected which comprises a number of contour lines (Figure 6-14). The study region is in Hubei Province, China. The left-bottom of the map is (10402877.801, 3046627.198), and the right-top is (10404262.01, 3050953.432). The contour interval is 20 metres. The purpose here is to get the distance tolerance of the Douglas-Peucker Algorithm so that the map can be generalized to get a map at scale 1:200K.

Firstly, the theoretical spatial similarity degree may be calculated by Formula 6-11:

$$y_T = 1.0164 \times (\frac{S_0}{S_1})^{-0.343} = 1.0164 \times (\frac{\frac{1}{100000}}{\frac{1}{200000}})^{-0.343} = 1.0164 \times 2^{-0.343} = 0.80133 .$$

Secondly, three contour lines are used as the representatives to calculate the distance tolerance, i.e. L1, L2 and L3 in Figure 6-14.

Third, each of the three contours is "simplified" using the Douglas-Peucker Algorithm. Taking L1 as an example, in the process of line simplification, record each intermediate distance tolerance and calculate the similarity degree (say, y_P) between each intermediate simplification result and the original contour. Select the intermediate contour when its corresponding y_P is most close to y_T . Then the corresponding distance tolerance (say, ε_{L1}) is viewed as the most appropriate distance tolerance for simplifying L1.

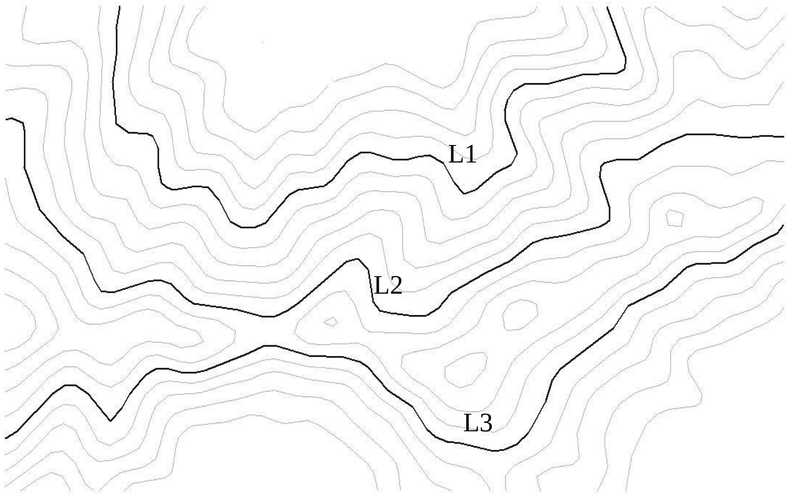


Figure 6-14 Original topographic map at scale 1:100K.

By the same method, the corresponding distance tolerance for L2 (say, ε_{L2}) and L3 (say, ε_{L3}) can also be obtained.

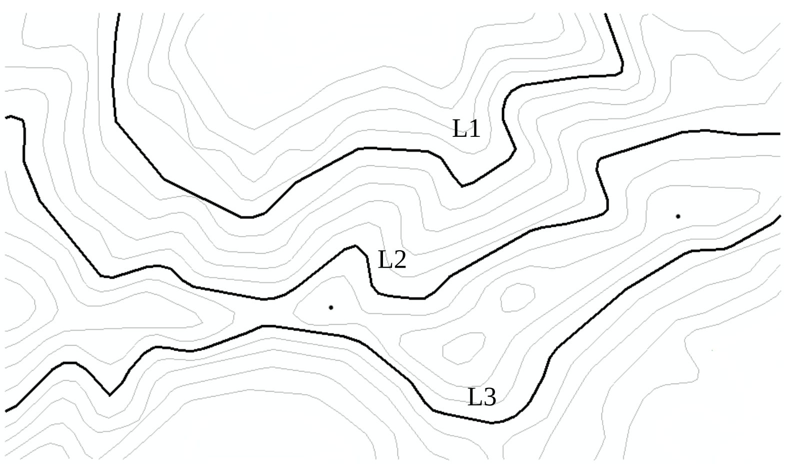


Figure 6-15 Simplified topographic map at scale 1:200K.

ε_{L1} =111.23m

ε_{L2} =131.77m

ε_{L3} =124.14m

Therefore, the resulting distance tolerance can be obtained by Formula 6-20,

$$\varepsilon = \frac{\sum_{i=1}^{N} \varepsilon_i}{N} = \frac{111.23 + 131.77 + 124.14}{3} = 122.57\text{m}$$

Using $\varepsilon = 122.57$m as the distance tolerance in the Douglas-Peucker Algorithm to simplify the contour lines, the resulting map can be generated (Figure 6-15).

After simplification of the contour lines, a question arises: “is the resulting map acceptable?” In other words, “is the tolerance distance appropriate?” The Radical Law (i.e. the Principle of Select) proposed by Töpfer and Pillewizer (1966) may be used to evaluate the resulting map, because it gives a formula for calculating the number of points retained on the resulting maps:

$$N_r = N_o \sqrt{\frac{S_o}{S_r}} \qquad \text{6-21}$$

where, N_r is the number of points that retained on the resulting contours; N_o is the number of points on the original contours; S_o is the denominator of the original map scale; and S_r is the denominator of the resulting map scale.

Taking Figure 6-14 (original map) and Figure 6-15 (resulting map) as an example, $S_o = 100K$, $S_r = 200K$. According to the original database, $N_o = 988$; so

$$N_r = N_o \sqrt{\frac{S_o}{S_r}} = 988\sqrt{\frac{100K}{200K}} = 698$$

In the light of the resulting database, the number of the points retaining on the contours is 655. Hence, the deviation is $D = \frac{698 - 655}{698} = 6.2\%$. This reveals that the resulting contours are acceptable.

6.6 Chapter Summary

This chapter addresses the three typical applications of spatial similarity relations in automated map generalization.

First, it discusses the relations between map scale change and spatial similarity degree in map generalization and proposes a general approach to quantitatively describing their relation. Further, ten formulae corresponding to the ten types of objects are given that can calculate spatial similarity degrees regarding map scale change as independent variables.

Second, it presents an approach for terminating a map generalization algorithm/system if the original map scale and resulting map scale are given. The approach is demonstrated taking a Voronoi-based algorithm for point clouds simplification as example.

Third, it proposes an approach to calculating the distance tolerance used in the Douglas-Peucker Algorithm in curve simplification. The approach serves for map generalization and can obtain the distance tolerance if the original map scale and the resulting map scale are known. The traditional Douglas-Peucker Algorithm may become fully automatic with the help of this approach.

Chapter 7 Conclusions and Recommendations

7.1 Overall Summary

This thesis focuses on spatial similarity relation. It aims at proposing the fundamental theory of spatial similarity relation and the models for calculating spatial similarity relations in multi-scale map spaces. The theory and the models can serves for automated map generalization.

For the purpose of obtaining quantitative relations between spatial similarity degree and map scale change, this study classifies the research objects into ten categories, and presents three major objectives in Chapter 1: (1) fundamental theories of spatial similarity relations, including the definitions, features, classification systems of spatial similarity relations, and the factors that affect humans' judgment of similarity in 2-dimensional map spaces; (2) approaches to calculating spatial similarity relations between two individual objects, or between two object groups, or between two maps in multi-scale map spaces; and (3) applications of the theories of similarity relations in automated map generalization, including calculating similarity degrees between a map and its generalized counterparts, calculating the threshold values of the Douglas-Peucker Algorithm, and determining when a map generalization system/algorithm can be terminated.

A systematic review of literature is presented in Chapter 2, including the definitions and features of similarity in various communities, a classification system of spatial similarity relations, and the calculation models of similarity relations in the communities of psychology, computer science, music, and geography, as well as a number of raster-based approaches for calculating similarity degrees between maps/images. The review not only summarizes previous achievements in spatial similarity relations and lays a theoretical foundation for this study, but also clearly shows the gap between previous achievements and the research objectives of this study.

Chapter 3 investigates the fundamental theory of spatial similarity relations systematically. It gives a definition of spatial similarity relations/degrees based on the Set Theory, addresses the ten features of spatial similarity relations and the factors that affect human's spatial similarity judgments, and proposes a classification system of spatial similarity relations. The weights of the factors that affect human's spatial similarity judgments have been achieved by psychological experiments. The first objective of the study has been reached in this chapter.

Chapter 4 proposes the ten models for calculating spatial similarity degrees of the ten types of objects at different map scales, i.e. individual points, individual lines, individual polygons, point clouds, parallel lines clusters, intersected line networks, tree-like networks, discrete polygon groups, connected polygon groups, and maps. Each of thses ten models takes into account the corresponding factors that affects human's similarity judgments and uses the weights of the factors obtained by psychological experiments presented in Chapter 3.

In Chapter 5, four strategies are employed, i.e. theoretical justifiability, third part involvement, comparison with existing approaches, and experts' participation, to validate the ten models. The first three strategies are briefly addressed; on the contrary, the last one, various psychological experiments accompanied by the third strategies, is discussed in detail. This has proved that the ten models are acceptable and therefore can be put into use in map generalization.

In Chapter 6, the proposed ten models are used in map generalization at three aspects . First, they are used to construct the ten formulae that can determine quantitative relations between spatial similarity degree and map scale change of the corresponding ten types of objects in map generalization. Second, an approach is proposed based on the ten models that can determine when to terminate a map generalization system/algorithm in the process of map generalization. Third, the model are used to calculate the distance tolerance of the Douglas-Peucker Algorithm so that the algorithm can become fully automatic in map generalization.

7.2 Contributions

Although various achievements have been made on similarity relations in many fields including image processing, few books and articles can be found that research on spatial similarity relations in vector map spaces. This study emphasizes on approaches to calculating spatial similarity degrees in multi-scale map spaces, and has made innovative contributions in the following aspects.

First, the fundamental issues of spatial similarity relations are explored, i.e. (1) a classification system is proposed that classifies the objects processed by map generalization algorithms into ten categories; (2) the Set Theory-based definitions of similarity, spatial similarity, and spatial similarity relation in multi-scale map spaces are given; (3) mathematical language-based descriptions of the features of spatial similarity relations in multi-scale map spaces are addressed; (4) the factors that affect human's judgments of spatial similarity relations are proposed, and their weights are also obtained by psychological experiments; and (5) a classification system for spatial similarity relations in multi-scale map spaces is proposed.

Second, the models for calculating spatial similarity degrees for the ten types of objects in multi-scale map spaces are proposed, and their validity is tested by psychological experiments. If a map (or an individual object, or an object group) and its generalized counterpart are given, the models can be used to calculate the spatial similarity degrees between them.

Third, the proposed models are used to solve problems in map generalization: (1) ten formulae are constructed that can calculate spatial similarity degrees by map scale changes in map generalization; (2) an approach based on spatial similarity degree is proposed that can determine when to terminate a map generalization system or an algorithm when it is executed to generalize objects on maps, which may fully automate some relevant algorithms and therefore improve the efficiency of map generalization; and (3) an approach is proposed to calculate the distance tolerance of the Douglas-Peucker Algorithm so that the Douglas-Peucker Algorithm may become fully automatic.

7.3 Limitations

Despite having made many achievements in spatial similarity relation, the theory and the approaches proposed in this study possess several limitations.

First, spatial similarity relations are usually describe using qualitative terminologies, and people, including cartographers and geographers, are not accustomed to quantitative descriptions of spatial similarity relation; hence, it is difficult for cartographers and geographers to accept and use the mathematical formulae and models proposed in this study in short period of time.

Second, the proposed formulae and models are based on psychological experiments. As is well known, the more subjects and samples (i.e. maps and objects) the experiments possess, the more accurate the experiments are, and the better the models and the formulae are. Nevertheless, the number of the surveyed subjects and the number of used samples in the psychological experiments are limited, which is a negative aspect for the accuracy of the formulae and the models.

As a final note, spatial similarity relation roots itself in human's spatial cognition. It may be slightly different from people to people due to their difference in age, gender, educational background, culture, etc. Thus, the adaptability of the models and formulae should be taken into consideration before they are widely used.

7.4 Recommendations for Further Research

Further research of this issue may target on the following areas.

First, more experiments should be done to improve the accuracy and adaptability of the proposed models and formulae. The new experiments should select more typical maps and map objects as samples, and find more subjects from different cultural background.

Second, is it possible to design an identical and simple model for the ten models proposed in Chapter 4 that can calculate spatial similarity degrees between two maps/objects at different scales? In the meanwhile, is it possible to construct an identical and simple formula for the ten formulae proposed in Chapter 6 that can calculate spatial similarity degree taking map scale change as independent variable?

The significance of solving the two problems is too evident to discuss further.

Third, it is important to find the algorithms and operators that are not parameter-free and closely related to spatial similarity relation and map scale change. More importantly, it is worth exploring the approaches for automatically obtaining the parameters used in these algorithms and operators with the help of the models and formulae proposed in this study. Progress in this area may lay good foundation for fully automation of map generalization.

Additionally, it is of great useful to tell the similarity degree of two arbitrary vector maps. The ability to objectively compare maps is fundamental to map analysis yet is often neglected by far, and visual comparison is far too limited. The theory of spatial similarity relation in multi-scale map spaces provides a way for comparing maps, whether the theory can be extended to compare maps in general map spaces is worth of further investigation.

Appendix A List of Basic Logic Symbols

Symbol	Name	Explanation	Example
⇒ → ⊃	implies; if... then	A ⇒ B is true just in the case that either A is false or B is true, or both. → and ⊃ may mean the same as ⇒	x = 2 ⇒ x2 = 4 is true, but x2 = 4 ⇒ x = 2 is in general false (since x could be −2).
⇔ ≡ ↔	if and only if; iff; means the same as	A ⇔ B is true just in case either both A and Bare false, or both A and B are true.	x + 5 = y + 2 ⇔ x + 3 = y
¬ ˜ !	not	The statement ¬A is true if and only if A is false. A slash placed through another operator is the same as "¬" placed in front.	¬(¬A) ⇔ A x ≠ y ⇔ ¬(x = y)
∧ • &	and	The statement A ∧ B is true if A and B are both true; else it is false.	n < 4 ∧ n >2 ⇔ n = 3 when n is a natural number.
∨ + ǁ	or	The statement A ∨ B is true if A or B (or both) are true; if both are false, the statement is false.	n ≥ 4 ∨ n ≤ 2 ⇔ n ≠ 3 when n is a natural number.
⊕	xor	The statement A ⊕ B is true when either A or B, but not both, are	(¬A) ⊕ A is always true. A ⊕ A is always

⊻		true. A ⊻ B means the same.	false.
⊤ T 1	top, verum	The statement ⊤ is unconditionally true.	A ⇒ ⊤ is always true.
⊥ F 0	bottom, falsum	The statement ⊥ is unconditionally false.	⊥ ⇒ A is always true.
∀ ()	for all; for any; for each	∀ x: P(x) or (x) P(x) means P(x) is true for all x.	∀ n ∈ N: n2 ≥ n.
∃	there exists	∃ x: P(x) means there is at least one x such that P(x) is true.	∃ n ∈ N: n is even.
∃!	there exists exactly one	∃! x: P(x) means there is exactly one x such that P(x) is true.	∃! n ∈ N: n + 5 = 2n.
:= ≡ :⇔	is defined as	x := y or x ≡ y means x is defined to be another name for y (but note that ≡ can also mean other things, such as congruence). P :⇔ Q means P is defined to be logically equivalent to Q.	cosh x := (1/2)(exp x + exp (−x)) A XOR B :⇔ (A ∨ B) ∧ ¬(A ∧ B)
⊢	provable	x ⊢ y means y is provable from x (in some specified formal system).	A → B ⊢ ¬B → ¬A
⊨	entails	x ⊨ y means x semantically entails y	A → B ⊨ ¬B → ¬A

Appendix B List of Publications during the PhD Study

• Peer Reviewed Journal Papers

H. Yan, Z. Wang, J. Li, 2013, An approach to computing direction relations between separated object groups, GeoScientific Model Development, 6(5): 1591-1599.

H. Yan, J. Li, 2013, An approach to simplifying point features on maps using the Multiplicative Weighted Voronoi Diagram，Journal of Spatial Science, 58(2): 291-304.

H. Yan, L. Zhang, Q. Li, S. Li, 2013, A vector–based efficient algorithm for computing the differences between two complex polygons, Application Research of Computers, 30(10): 3192-3194. (in Chinese)

H. Yan, Y. Liu, J. Cao, 2013, Analysis and prediction of the underground water depth and its influencing factors in Minqin Oasis, Science of Soil and Water Conservation, 11(2): 45-50. (in Chinese)

H. Yan, B. Wang, 2013，A MWVD-based algorithm for point cluster generalization, Geomatics and Information Science of Wuhan University, 38(9): 1088-1091. (in Chinese)

H. Yan, J. Li, Blind watermarking technique for topographic map databases, 2012, Applied Geomatics, 4(4): 85-93.

H. Yan and J. Li, 2011, A blind watermarking approach to protecting geospatial data from piracy, International Journal of Information and Education Technology, 1(2): 94-98.

H. Yan, J. Li, H. Wen, A key points-based blind watermarking approach for vector geospatial data, Computers, Environment and Urban Systems, 2011, 35(6): 485-492.

• Conference Papers

H. Yan, T. Liu, W. Yang, J. Li, 2013, Spatial similarity relations in multi-scale map spaces: basic concepts and potential research issues, In: the Proceedings of the ICA Workshop on Generalization and Map Production, Dresden, Germany, August 23-24, 2013.

H. Yan, W. Yang, J. Li, 2013, Point cluster simplification using weighted voronoi diagram, In: the Proceedings of the Fourth International Conference on Earth Observation for Global Changes and Annual Conference of Canadian Institute of Geomatics(CIG), Ryerson University, Toronto, Ontario, Canada, June 5-7, 2013.

H. Yan, 2011, Watermarking algorithm for vector point clusters, In: the 7th International Conference on Wireless Communications (WiCOM), Networking and Mobile Computing, Wuhan, China. September 23-25, 2011.

H. Yan, Y. Tan, W. Yang, 2011, Blind watermarking algorithm for topographic map data, In: International Conference on Engineering and Information Management (ICEIM2011), Chengdu, China, April 2011. ISSN: 1877-7058.

• ***Books and Book Chapters***

H. Yan, M. Wang and Z. Wang, Computational Geometry: Algorithms in Spatial Data Handling, Beijing: Science Press, 2012. (in Chinese)

H. Yan, Chapter 7: digital map generalization based on spatial relations, in: Z. Li, J. Wang and F. Wu (eds.), Progress in Digital Map Generalization, Beijing: Science Press, 2011. (in Chinese)

H. Yan and T. Ai, Chapter 5: library of the rules in digital map generalization, in: Z. Li, J. Wang and F. Wu (eds.), Progress in Digital Map Generalization, Beijing: Science Press, 2011. (in Chinese)

Bibliography

Allen J., 1983, Maintaining knowledge about temporal intervals. Communications of the ACM 26(11): 832-843.

Alt H., and Godau M., 1995, Computing the Fr ́echet distance between two polygonal curves. International Journal of Computational Geometry and Applications, 5:75–91.

Alt H., Behrends B., and Blomer J., 1995, Approximate matching of polygonal shapes, Annals of Mathematics and Artificial Intelligence, 13: 251–266.

Alt H., Fuchs U., Rote G., and Weber G., 1998, Matching convex shapes with respect to the symmetric difference. Algorithmica, 21: 89–103

Alt H., Knauer C., and Wenk C., 2001, Matching polygonal curves with respect to the Frechet distance, In LNCS: Symposium on Theoretical Aspects of Computer Science (STACS), Berlin: Springer, pp. 63–74.

Amin A. and Wilson W.H, 1993, Hand-printed character recognition system using artificial neural networks, Proceedings of the Second International Conference on Document Analysis and Recognition, pp.943-946.

Arkin E.M., Chew L.P., Huttenlocker D.P., Kedem K., Mitchell J.S.B., 1991, An efficient computable metric for comparing polygonal shapes, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(3): 209-216.

Arlinghaus S.L., 1994, PHB Practical Handbook of Curve Fitting, CRC Press.

Aurenhammer F., 1991, Voronoi Diagram- a survey of a fundamental geometric data structure, ACM Computing Surveys, 23(3): 345-405.

Avis D. and Elgindy H, 1983, A combinatorial approach to polygon similarity, IEEE Transactions on Information Theory, 29:148-150.

Bader M., and Weibel R., 1997, Detecting and resolving size and proximity conflicts in the generalisation of polygon maps, in Proceedings of the 18th International Cartographic Conference, Stockholm, Sweden, pp.1525-1532.

Bader M., Barrault M., and Weibel R., 2005, Building displacement over a ductile truss, International Journal of Geographical Information Science, 19(8-9): 915-936.

Balci O., 1989, How to assess the acceptability and credibility of simulation results, In the Proceedings of 1989 Winter Simulation Conference, E.A. MacNair, K.J. Musselman and P. Heidelberger (eds.), Piscataway, New Jersey: IEEE, pp.62-71.

Banks J., Carson J.S., Nelson B.L., Nicol D.M., 2010, Discrete-event System Simulation (Fifth edition), Upper Saddle River: Pearson Education.

Barber P., 2005, The Map Book, Weidenfeld and Nicolson, pp. 232, 250.

Belohlavek R., 2000, Similarity relations in concept lattices. Journal of Logic and Computation, 10(6): 823-845.

Berner-Lee J., Handler J., and Lassila O., 2001, The Semantic Web. Scientific American, 184(5), 34-43.

Berry J.K., 1993, Beyond Mapping: Concepts, Algorithms and Issues in GIS, Wiley and Sons, 246 pages.

Bjørke J.T., 2004, Topological relations between fuzzy regions: derivation of verbal terms, Fuzzy Sets and Systems, 141(3): 449-467.

Boffet A., and Rocca Serra S., 2001, Identification of spatial structures within urban blocks for town characterization, in Proceedings of the 20th International Cartographic Conference, Beijing, China. (CD-ROM)

Boots B. and Csillag F., 2006, Categorical maps, comparisons, and confidence, Journal of Geographical Systems, 8:109–118.

Bronstein A.M., Bronstein M.M., and Kimmel R., 2008, Numerical Geometry of Non-rigid Shapes, Berlin: Springer.

Bronstein A.M., Bronstein M.M., Bruckstein A.M., and Kimmel R., 2009, Partial similarity of objects, or how to compare a centaur to a horse, International Journal of Computer Vision, 84: 163-183.

Bruns H.T., and Egenhofer M.J., 1996, Similarity of spatial scenes, in: Proceedings of the 7th International Symposium on Spatial Data Handling, Delft, The Netherlands, pp. 4A.31-42.

Buchin K., Buchin M. and Wenk C., 2006, Computing the Frechet distance between Simple polygons in polynomial time, in: the Proceedings 22th ACM Symposium on Computational Geometry (SoCG), pp. 80–88.

Budanitsky A. and Hirst G., 2001, Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures, Workshop on WordNet and Other Lexical Resources, the Second Meeting of the North American Chapter of the Association for Computational Linguistics. Pittsburgh

Budanitsky A., 1999, Lexical Semantic Relatedness and Its Application in Natural Language Processing. Computer Systems Research Group, University of Toronto, Technical Report.

Butler L.K., Ries L., Bisson I., Hayden T.J., Wikelski M.M., Romero L.M., 2013, Opposite but analogous effects of road density on songbirds with contrasting habitat preferences, Animal Conservation, 16(1): 77-85.

Cakmakov D., Arnautovski V., and Devcav D., 1992, A model for polygon similarity estimation, Proceedings of CompEuro'92: Computer Systems and Software Engineering, Hague , the Netherlands, pp.701-705.

Carson J., 2002, Model verification and validation, in: E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds., Proceedings of the 2002 Winter Simulation Conference. http://informs-sim.org/wsc02papers/008.pdf.

Cederberg J.N., 1989, A Course in Modern Geometries, Chapter 3.12 Similarity Transformations, Springer, pp. 183–189.

Chang C. and Lee S., 1991, Retrieval of similarity pictures on pictorial databases. Pattern Recognition 24(7): 675-680.

Chang S.K., Shi Q.S. and Yan C.W., 1987, Iconic indexing by 2-D strings. IEEE Transactions on Pattern Analysis and Machine Intelligence, 9(6): 413-428.

Chen J., Li C., Li Z. and Gold C., 2001, A Voronoi-based 9-Intersection model for spatial relations, International Journal of Geographical Information Science, 15(3): 201-220.

Christophe S., and Ruas A., 2002, Detecting building alignments for generalisation purposes, In: Richardson D.E. and van Oosterom P. (eds.), Proceedings of the 10th International Symposium on Spatial Data Handling, Berlin: Springer-Verlag, pp. 419-432.

Cilibrasi R., Vitanyi P., 2006, Similarity of objects and the meaning of words, In: Proceedings of the 3rd Annual Conference on Theory and Applications of Models of Computation (TAMC'06), LNCS, Vol.3959, pp:21-45.

Clementini E., Sharma J., Egenhofer M.J., 1994, Modeling topological spatial relations: strategies for query processing, Computers and Graphics, 18(6): 815–822.

Cohen S.D. and Guibas L., 1997, Partial matching of planar polylines under similarity transformations, in: Proceedings of the 8th ACM-SIAM Symposium on Discrete Algorithms, pp. 777–786.

Congalton R.G., Oderwald R. and Mead R.A., 1983, Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques, Photogrammetric Engineering and Remote Sensing, 49: 1671-1678.

Coxeter H.S.M., 1961, Introduction to Geometry: §5 Similarity in the Euclidean Plane, John Wiley and Sons, pp. 67–76.

Csillag F. and Boots B., 2004, Toward comparing maps as spatial processes. In: Fisher P. (eds) Developments in Spatial Data Handling. Springer, Heidelberg, pp. 641

Damerau F.J., 1964, A technique for computer detection and correction of spelling errors. Communications of the ACM. 7(3): 171-176.

de Serres B., Roy A., 1990, Flow direction and branching geometry at junctions in Dendritic river networks, The Professional Geographer, 42 (2): 149-201.

Douglas D., and Peucker T., 1973, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, The Canadian Cartographer, 10(2): 112-122.

Du S., Qin Q., Wang Q., Ma H., 2008, Evaluating structural and topological consistency of complex regions with broad boundaries in multi-resolution spatial databases, Information Sciences, 178(1): 52–68.

Du S.H., Qin Q.M., Wang Q. and Ma H.J., 2008, Reasoning about topological relations between regions with broad boundaries, International Journal of Approximate Reasoning, 47(2):219-232.

Efrat A, Guibas L.J, Har-Peled S., Mitchell J.S.B, and Murali T.M., 2002, New similarity measures between polylines with applications to morphing and polygon sweeping, Discrete and Computational Geometry, 28(4): 535-569.

Egenhofer M. and Al-Taha K., 1992, Reasoning about gradual changes of topological relationships. in: Frank, A. U., Campari, I., and Formentini, U., (Eds.), Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. Pisa, Italy.

Egenhofer M. and Franzosa R., 1991, Point-set topological spatial relations, International Journal of Geographical Information Systems. 5(2):161-174.

Egenhofer M. and Mark D., 1995a, Naive geography, spatial information theory—a theoretical basis for GIS. in: Frank A and Kuhn W (eds.). Proceedings of International Conference COSIT '95. Semmering, Austria. 988: 1-15.

Egenhofer M. and Mark D., 1995b, Modeling conceptual neighbourhoods of topological line-region relations, International Journal of Geographical Information Systems, 9(5): 555-565.

Egenhofer M., 2002, Toward the Semantic Geospatial Web. In ACM-GIS, McLean, VA, edited by A. Voisard and S.-C. Chen (ACM Press), pp. 1–4.

Egenhofer M.J., Clementini E., Di Felice P., 1994, Topological relations between regions with holes, International Journal of Geographical Information Systems, 8(2): 129-144.

El-Kwae E.A. & Kabuka M.R., 1999, A robust framework for content-based retrieval by spatial similarity in image databases. ACM Transactions on Information Systems, 17(2): 174-198.

Fonseca F., and Sheth A., 2003, The geospatial semantic Web, Research Priorities Series, University Consortium for Geographic Information Science. http://www.personal.psu.edu/faculty/f/u/fuf1/Fonseca-Sheth.pdf

Formica A., Pourabbas E. and Rafanelli M., 2013, Constraint relaxation of the polygon-polyline topological relation for geographic pictorial query languages, Computer Sciences and Information Systems, 10(3): 1053-1075.

Frank A.U., 1991, Qualitative spatial reasoning about cardinal directions. in: Proceedings of the 7th Austrian Conference on Artificial Intelligence, pp.157-167.

Frank A.U., 1996, Qualitative spatial reasoning: cardinal directions as an example. International Journal of Geographical Information Systems, 10(3): 269-290.

Freksa C., 1992, Temporal reasoning based on semi-intervals. Artificial Intelligence 54: 199-227.

Gass S.I., 1983, Decision-aiding models: validation, assessment, and related issues for policy analysis, Operations Research, 31(4): 601-663.

Gass S.I., 1993, Model accreditation: a rationale and process for determining a numerical rating, European Journal of Operational Research, 66(2): 250-258.

Goldstone R.L., 1994, Similarity, interactive activation and mapping. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(1): 3-28.

Goldstone R.L., 2004, Similarity. in: R.A.Wilson and F.C.Keil, (eds.), MIT Encyclopedia of the Cognitive Sciences. MIT Press.

Goodchild M.F., 2006, The law in geography, Technical Report in GIS, the GIS Institute, University of Zurich.

Gower J.C., 1971, A general coefficient of similarity and some of its properties, Biometrics, 27(4): 857-871.

Goyal R. and Egenhofer M., 2001, Similarity of cardinal directions. in: Jensen C., Schneider M., Seeger B., and Tsotras V., (eds.), 7th International Symposium on Spatial and Temporal Databases.

Goyal R.K., 2000, Similarity assessment for cardinal directions between extended spatial objects, PhD Thesis, The University of Maine, USA.

Guan Y., Wang X.L., Kong X.Y., and Zhao J., 2002, Quantifying semantic similarity of Chinese words from HowNet, in: the Proceedings of the 1st International Conference on Machine Learning and Cybernetics, Beijing, November 4-5, 2002.

Guo R Z, 1997. Spatial Analysis. Wuhan: Press of Wuhan Technical University of Surveying and Mapping. (in Chinese)

Gustafson E.J., 1998, Quantifying landscape spatial pattern: what is the state of the art? Ecosystems, 1: 143-156.

Hagen A., 2002, Multi-method assessment of map similarity. Paper presented at the 5th AGILE Conference on Geographic Information Science, Palma, Spain.

Hagen A., 2003, Fuzzy set approach to assessing similarity of categorical maps, International Journal of Geographical Information Science, 17: 235-249

Hagen A., 2003, Fuzzy set approach to assessing similarity of categorical maps, International Journal of Geographical Information Science, 17(3): 235-249.

Hagen-Zanker A. and Lajoie G., 2008, Neutral models of landscape change as benchmarks in the assessment of model performance, Landscape and Urban Planning, 86: 284-296.

Hagen-Zanker A., 2006, Map comparison methods that simultaneously address overlap and structure, Journal of Geographical Systems, 8: 165-185.

Hagen-Zanker A., 2009, An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation, International Journal of Geographical Information Science, 23 (1): 61-73.

Hagen-Zanker A., Straatman B. and Uljee I., 2005, Further developments of a fuzzy set map comparison approach, International Journal of Geographical Information Science, 19(7): 769-785

Hagen-Zanker A., van Loon J., Maas A., Straatman B., de Nijs T. and Engelen G., 2005, Measuring performance of land use models: an evaluation framework for the calibration and validation of integrated land use models featuring cellular automata. Paper presented at 14th European Colloquium on Theoretical and Quantitative Geography, Tomar, Portugal.

Hahn U., Chater N., and Richardson L.B., 2003, Similarity as transformation, Cognition, 87(1): 1–32.

Harley J.B. and Woodward D., 2005, The History of Cartography Vol.3: Cartography in the European Renaissance. Chicago and London: University of Chicago Press.

Harvey P.D.A, 1980, The history of Topographical Maps: Symbols, Pictures and Surveys, Thames and Hudson. Pp.9

Heller V., 2011, Scale effects in physical hydraulic engineering models, Journal of Hydraulic Research, 49(3): 293–306.

Hershberger J. & Snoeyink J., 1992, Speeding up the Douglas–Peucker line-simplification algorithm, in: the Proceedings of the 5th Symposium on Data Handling, pp.134–143. UBC Technical Report TR-92-07, available at http://www.cs.ubc.ca/cgi-bin/tr/1992/TR-92-07 on October 2, 2013.

Höhle U., 1988, Quotients with respect to similarity relations. Fuzzy Sets and Systems, 27(1), 31-44.

Hong J., 1994, Qualitative distance and direction reasoning in geographic space, PhD Thesis, University of Maine, USA.

Horton H., 1945, Erosional development of streams and their drainage basins- hydrophysical approach to quantitative morphology, Bulletin of the Geological Society of America, 56: 275-370.

Hubert C., 2009, Turbulent air-water flows in hydraulic structures: dynamic similarity and scale effects, Environmental Fluid Mechanics, 9(2):125–142.

Huttenlocher D.P., Klanderman G.A., and Rucklidge W.G., 1993, Comparing images using the Hausdorff distance, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15:850–863.

Imai S., 1977, Pattern similarity and cognitive transformations. Acta Psychologica, 41: 443-447.

Johnson A.M. and Maggiora G.M., 1990, Concepts and Applications of Molecular Similarity, New York: John Willey & Sons.

Knuth D., 1997, The Art of Computer Programming: Fundamental Algorithms, Third Edition. Addison-Wesley, Section 2.3: Trees, pp. 308–423.

Kolb W.M., 1984, Curve fitting for programmable calculators, Syntec Inc.

Kraak M.J., and Ormeling F., 1996, Cartography: Visualization of Spatial Data, Longman.

La Barbera P. & Rosso R., 1989, On the fractal dimensions of stream network. Water Resource Research, 25 (4): 735-741

Lanczos C., 1988, Applied Analysis (Reprint of 1956 Prentice–Hall edition), Dover Publications, pp. 212–213.

Larkey L.B. and Markman A.B., 2005, Processes of similarity judgment, Cognitive Science, 29: 1061–1076.

Levenshtein V.I., 1966, Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady, 10: 707-710.

Li B. and Fonseca F.T., 2006, TDD - A comprehensive model for qualitative spatial similarity assessment, Spatial Cognition and Computation, 6(1): 31-62.

Li C., Chen J. and Li Z., 1999, A raster-based method for computing Voronoi Diagrams of spatial objects using dynamic distance transformation, International Journal of Geographical Information Science, 13(3): 209-225.

Li JC., 2000, From graphic similitude to object similitude. Journal of Sichuan University of Science and Technology. 19(2):126-128. (in Chinese)

Li Z., Yan H., Ai T, Chen J., 2004, Automated building generalization based on urban morphology and gestalt theory, International Journal of Geographical Information Science, 18(5): 513-534.

Liang J.X, 1999. Fundamental concepts of similarity relation and its weak equivalence properties. Systems Engineering-theory and Practice. 28(7):106-111. (in Chinese)

Ligozat G., 1998, Reasoning about cardinal directions. Journal of Visual Languages and Computing, 9: 23-44.

Lin D., 1998, An information-theoretic definition of similarity, in: the Proceedings of the 15th International Conference on Machine Learning, San Francisco, CA: Morgan Kaufmann, pp. 296–304.

Liu Y.L., 2002, Categorical database generalization in GIS, PhD Thesis, Utrecht University, The Netherlands.

llman J.R., 1977, A binary n-gram technique for automatic correction of substitution, deletion, insertion and reversal errors in words. Computer Journal, 20(2): 141-147.

Mandelbrot B., 1967, How long is the coast of Britain? Statistical Self-Similarity and Fractional Dimension. Science, New Series, 156(3775): 636-638.

Markman A.B., 1993, Structural alignment during similarity comparisons, Cognitive Psychology, 25: 431-467.

Markman A.B., 1997, Constraints on analogical inference. Cognitive Science 21(4): 373-418.

Martin G.E, 1982, Transformation Geometry : an Introduction to Symmetry, New York: Springer-Verlag.

Maselli F., Rudolph A. and Conese C., 1996, Fuzzy classification of spatially degraded Thematic Mapper data for the estimation of sub-pixel components. International Journal of Remote Sensing, 17: 537–551.

McGarigal K. Cushman S.A. Neel M.C. and Ene E., 2002, FRAGSTATS: spatial pattern analysis program for categorical maps. www.umass.edu/landeco/research/fragstats/fragstats.html.

Medin D.L., Goldstone R.L. and Gentner D., 1993, Respects for similarity, Psychological Review,100(2): 254-278.

Milenkovic V.J., 1998, Rotational polygon overlap minimization and compaction, Computational Geometry-theory and Applications, 10(4): 305-318.

Mokhtarian F. and Mackworth A.K., 1992, a theory of multi-scale, curvature-based representation for planar curves, IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(8): 789-805

Monserud R.A. & Leemans R., 1992, Comparing global vegetation maps with the Kappa statistic, Ecological Modeling, 62: 275-293.

Natori N. and Nishimura L., 1994, Dynamics-based active learning for handwritten character recognition, Proceedings of ICN94, pp.2875-2880.

Naylor T.H. and Finger J.M., 1967, Verification of computer simulation models, Management Science, 2: B92-B101.

Nedas K.A. and Egenhofer M.J., 2003, Spatial similarity queries with logical operators, in: the Proceedings of the 8th International Symposium on Spatial and Temporal Databases, LNCS 2750, pp. 430-448.

Nikolova N. and Jaworska J., 2003, Approaches to measure chemical similarity-- a review, QSAR and Combinatorial Science, 22 (9-10): 1006–1026.

Orpen K.S., & Huron David, 1992, Measurement of similarity in music: a quantitative approach for non-parametric representations, Computers in Music Research, 4: 1-44.

Ovchinnikov S., 1991, Similarity relations, fuzzy partitions, and fuzzy orderings. Fuzzy Sets and Systems, 40(1), 107-126.

Palmer S.E., 1992, Common region: a new principle of perceptual grouping, Cognitive Psychology, 24(2): 436-447.

Papadias D. and Dellis V., 1997, Relation-based similarity. in: the Proceedings of the 5th ACM Workshop on Advances in Geographic Information Systems, Las Vegas, NV, pp.1-4.

Petraglia G., Sebillo M., Tucci M., and Tortora G., 2001, Virtual images for similarity retrieval in image databases, IEEE Transactions on Knowledge and Data Engineering, 13 (6): 951-967.

Peuquet D. & Zhan C.X., 1987, An algorithm to determine the directional relation between arbitrarily-shaped polygons in the plane, Pattern Recognition, 20(1): 65-74.

Peuquet D., 1986, The use of spatial relationships to aid spatial database retrieval, the Proceedings of the 2nd International Symposium on Spatial Data Handling (SDH). Zurich, Switzerland.

Pontius J.R.G., 2000, Quantification error versus location error in comparison of categorical maps. Photogrammetric, Engineering and Remote Sensing, 66(8): 1011-1016.

Pontius jr R.G., Huffaker D. and Denman K., 2004, Useful techniques of validation for spatially explicit land-change models, Ecological Modeling, 179: 445-461

Power C., Simms A., and White R., 2001, Hierarchical fuzzy pattern matching for the regional comparison of land use maps, International Journal of Geographical Information Science, 15(1): 77-100.

Quillian M., 1968, Semantic Memory. in: Minsky, M., (eds.), Semantic Information Processing. MIT Press, Cambridge, MA, pp.

Rada R., Mili H., Bicknell E., and Blettner M., 1989, Development and application of a metric on semantic nets. IEEE Transactions on Systems, Man, and Cybernetics, 19(1):17-30.

Rada R., Mili H., Bicknell E., and Blettner M., 1989, Development and application of a metric on semantic nets. IEEE Transactions on Systems 19(1): 17-30.

Rainsford D. and Mackaness W., 2002, Template matching in support of generalization of rural buildings, In Richardson D.E. and van Oosterom P. (eds.), in: the Proceedings of the 10th International Symposium on Spatial Data Handling, Berlin: Springer-Verlag, pp. 137-151.

Ramer U., 1972, An iterative procedure for the polygonal approximation of plane curves, Computer Graphics and Image Processing, 1(3): 244–256.

Regnauld N., 2001, Contextual building typification in automated map generalization, Algorithmica, 30(2): 312-333.

Remmel T.K. and Fortin M.J., 2013, Categorical, class-focused map patterns: characterization and comparison, Landscape Ecology, 28:1587–1599.

RISK BV, 2013, Map Comparison Kit 3: User Manual, Research Institute for Knowledge Systems BV, http://www.riks.nl/mck.

Rock I., 1996, Indirect Perception. London: MIT Press.

Rodríguez A. and Egenhofer M., 2003, Determining semantic similarity among entity classes from different ontologies. IEEE Transactions on Knowledge and Data Engineering, 15(2): 442-456.

Rodríguez A., and Egenhofer M., 2004, Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure. International Journal of Geographical Information Science, 18(3):229-256.

Rodriguez M.A., and Egenhofer M.J., 2004, Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure, International Journal of Geographic Information Science, 18(3): 229–256.

Ross R. 1999, Fractal relation of mainstream length to catchment area in river networks. Water Resource Research, 27 (3): 381-387.

Ruas A. 2001, Automating the generalization of geographical data. In Proceedings of the 20th International Cartographic Conference. Beijing, China, pp. 1943–1953.

Ruas A., 1998, A method for building displacement in automated map generalization, International Journal of Geographical Information Science, 12(8): 789-803.

Sadahiro Y., 1997, Cluster perception in the distribution of point objects. Cartographica, 34: 49–61.

Sadahiro Y., 2012, Exploratory analysis of polygons distributed with overlap, Geographical Analysis, 44(4): 350-367.

Santini S., and Jain R., 1996, Similarity queries in image databases, in the Proceedings og the IEEE International Conference on Computer Vision and Pattern Recognition, San Francisco, CA, pp. 646-651.

Sargent R.G., 1982, Chapter IX Verification and validation of simulation models, F.E. Cellier (ed.), Progress in Modeling and Simulation, London: Academic Press, pp.159-169.

Sargent R.G., 1984, Chapter 19 Simulation model validation, in: T.I. Oren, B.P. Zeigler and M.S. Elzas (eds.), Simulation and Model-Based Methodologies: An Integrative View, Heidelberg, Berlin: Springer-Verlag, pp.537-555.

Sargent R.G., 2010, A new statistical procedure for validation of simulation and stochastic models, Technical Report SYR-EECS-2010-06, Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, New York.

Sargent R.G., 2011, Verification and validation of simulation models, in: S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu (eds.), Proceedings of the 2011 Winter Simulation Conference, Baltimore, MD, pp. 166-183.

Schlesinger S., 1979, Terminology for model credibility, Simulation, 32 (3): 103-104.

Shepard R.N., 1962, The analysis of proximities: Multidimensional scaling with an unknown distance function, International Psychometrika, 27 (2): 125–140.

Tennekes H., 1984, Similarity relations, scaling laws and spectral dynamics. F.T.M Nieuwstadt et al. (eds.), Atmospheric Turbulence and Air Pollution Modeling, Reidel Publishing Company, Dordrecht, The Netherlands, pp. 37-68.

Thomson R., Brooks R., 2002, Exploiting perceptual grouping for map analysis, understanding and generalization: the case of road and river networks, in: D. Blostein , Y.B. Kwon, (eds.), Graphics Recognition: Algorithms and Applications, LNCS, Vol.2390, Berlin: Springer, pp: 141-150.

Tobler W.R., 1970, A computer movie simulating urban growth in the Detroit region. Economic Geography, 46: 234–240.

Töpfer F. & Pillewizer W., 1966, The principles of selection, The Cartographic Journal, 3(1): 10-16.

Torgerson W.S., 1965, Multidimensional scaling of similarity. Psychometrika 30: 379-393.

Toussaint G.T., 2006, A comparison of rhythmic dissimilarity measures, FORMA, 21(2): 129–149.

Tversky A., 1977, Features of similarity, Psychological Review, 84 (4): 327–352.

Tversky B., Agrawala M., Heiser J., Lee P., Hanrahan P., Phan D. & Daniel M.P., 2006, Cognitive design principles for automated generation of visualizations, Applied Spatial Cognition: From Research to Cognitive Technology, 53-75.

Uuemaa E., Antrop M., Roosaare J. and Mander U., 2009, Landscape metrics and indices: an overview of the use in landscape research. Living Reviews in Landscape Research, 3: 1–28.

Van Vliet J., Bregt A. and Hagen-Zanker A., 2011, Revisiting Kappa to account for change in the accuracy assessment of land-use change models, Ecological modeling, 222: 1367-1375.

Visser H. and de Nijs T., 2006, The map comparison kit, Environmental Modeling & Software, 21: 346-358.

Visser H. and de Nijs T., 2006, The map comparison kit, Environmental Modelling & Software, 21: 346-358.

Wang J Y, 1993. Principles of General Map Generalization. Beijing: Surveying and Mapping Press. (in Chinese)

Weibel R. and Jones C.B., 1998, Computational perspectives on map generalization, GeoInformatica, 2(4): 307-314.

Yan H.W., and Weibel R., 2008, An algorithm for point cluster generalization based on the Voronoi diagram. Computers and Geosciences, 34(8): 939-954.

Yan H.W., 2010, Fundamental Theories of spatial similarity relations in multi-scale map spaces, Chinese Geographical Science, 20(1):18-22.

Yan H.W., Chu Y.D., Li Z.L. and Guo R.Z., 2006, A quantitative direction description model based on direction groups, Geoinformatica, 10(2):177-196.

Yan H.W., Li Z.L. and Ai T.H., 2006, System for automated cartographic map generalization. Chinese Geographic Science, 16 (2): 165-170.

Yan H.W., Weibel R. & Yang B.S., 2008, A multi-parameter approach to automated building grouping and generalization, Geoinformatica, 12 (1):73-89.

Zadeh L.A., 1971, Similarity relations and fuzzy orderings. Information Sciences, 3(2), 177-200

Zhang Q., Wang J., Peng X., Gong P., Shi P., 2002, Urban built-up land change detection with road density and spectral information from multi-temporal Landsat TM data, International Journal of Remote Sensing, 23(15): 3057-3078.

Zhang Q.N., 2006, Generalization of drainage network with density differences, Acta Geodaetica et Cartographica Sinica, 35(2): 191-196.

Zhou L.M., 1993. Similarity theories. Beijing: Press of Science and Technology of China. (in Chinese)

http://data.worldbank.org/indicator/IS.ROD.DNST.K2

http://en.wikipedia.org/wiki/List_of_logic_symbols

http://en.wikipedia.org/wiki/Similarity_(geometry)

http://ishare.iask.sina.com.cn/f/13293700.html

http://uwaterloo.ca/graduate-studies/thesis/preparing-your-thesis

http://uwaterloo.ca/graduate-studies/thesis/thesis-regulations http://www.compassdude.com/map-symbols.shtml

http://www.gismanual.com/mapping/

http://www.gsd.harvard.edu/gis/manual/lookup/

http://www.humdrum.org/Huron/Publications/orpen.similarity.text.html

http://www.innovativegis.com/basis/MapAnalysis/Topic10/Topic10.htm

http://www.innovativegis.com/basis/present/GIS02_similarity/GIS02_similarity.htm

http://www.merriam-webster.com/dictionary/similarity

http://www.riks.nl/mck/index.php

http://zh.wikipedia.org/wiki/%E9%80%BB%E8%BE%91%E7%AC%A6%E5%8F%B7%E8%A1%A
8

https://support.google.com/mapmaker/answer/1098048?hl=en

https://support.google.com/mapmaker/answer/1098056

https://uwspace.uwaterloo.ca/ldap-login;jsessionid=D66DC41DA42ECA492142BA2E3755A2AA

http://highered.mcgraw-hill.com/sites/0073022845/student_view0/chapter11/section3/

http://kobus.ca/seminars/ugrad/NM5_curve_s02.pdf