ISPRS Journal of Photogrammetry and Remote Sensing 65 (2010) 266–274

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

An evaluation of measures for quantifying map information

Lars Harrie [a,*], Hanna Stigmar [a,b]

[a] GIS Centre, Lund University, Sölvegatan 12, SE-223 62 Lund, Sweden

[b] National Land Survey of Sweden, SE-801 82 Gävle, Sweden

ARTICLE INFO

Article history:
Received 3 September 2007
Received in revised form
6 May 2009
Accepted 15 May 2009
Available online 12 June 2009

Keywords:
Cartography
Generalisation
Simplification
Visualisation
Analysis

ABSTRACT

A real-time map must not contain too much information. Therefore, we need measures of map information that could be guidelines for the selection of data layers and the real-time generalisation process. In this paper we evaluate measures of the *amount of information* and the *distribution of information*. The evaluation is performed by (1) defining measures, (2) implementing the measures, (3) computing the measures for some test maps, and finally (4) comparing the values of the measures with human judgement of the map information. For amount of information, we found that the measures *number of objects, number of points* and *object line length* had better correspondence with human judgement than *object area*. We also found that measures based on the size of Voronoi regions of objects (respectively points) can be used for identifying the distribution of information. The results are based on the testing of only building objects. Future work should extend the test, using all object types.

1. Introduction

A major issue in cartography is the usability of the map. For traditional paper maps this has been studied thoroughly. However, new technology has enabled new types of map usage such as interactive real-time applications using the web and mobile devices. These maps can be tailored for a specific purpose and even for a specific user (Reichenbacher, 2004; Gartner, 2004). This large freedom to tailor sets requirements on new analytical measures, or constraints, that describe the usability of the map.

This study is part of the project "The Swedish Planning Portal" (*Planeringsportalen*). By using this portal a user (employed at companies or governmental/local authorities) should be able to find planning information and especially geographic information related to physical planning. To enable the user to download and view the geographic information, web services – following the Open Geospatial Consortium (OGC) standards – will be set up. As for maps in general, it is important for a web-based map service like a planning portal that the information presented is as usable as possible. This means that is should be easy for the user to read and comprehend the maps.

To improve the usability of a map, cartographic generalisation is applied. In recent years, the generalisation research has tried to model the overall process of generalisation using constraints (Harrie and Weibel, 2007). A constraint can be seen as requirements that should be obtained in the generalisation process. The constraints can be classified into the following types (see Ruas and Plazanet, 1996; Weibel and Dutton, 1998; Harrie, 2003): *position*, *topology*, *shape*, *structural*, *functional* and *readability* (*legibility*). The five first types concern the representation, i.e. vital aspects of the map should not be lost in the generalisation process. The final type, readability constraints, concerns the ease with which the user can read the map.

There are two major types of readability constraints of a map. The first type concerns the visual perception. The map objects must be readable for a normal user. Robinson (Robinson, 1952, in MacEachren, 1995) suggested that cartographic objects should be designed considering human perception, using, for example, a definition of the smallest noticeable lettering size difference. For screen maps the paper map definitions can be rather coarse (Spiess, 1995), which is why specific definitions are needed.

The other type of readability constraints concerns the amount of map information. Even if the map objects, and features within the objects, are large enough the map reader cannot comprehend the map if it contains too much information (see Bjørke, 1996; Li and Huang, 2002). The amount of map information has an even greater importance in real-time maps than for traditional paper maps, as real-time maps should be read and understood relatively quickly. Therefore we should strive for establishing measures for the amount of information in a real-time map and let these measures act as constraints for selecting data layers and in the real-time generalisation process.

The aim of this study is to evaluate some measures of map readability that eventually should be used as constraints for the

* Corresponding author. Tel.: +46 46 222 01 55; fax: +46 46 222 83 91.
E-mail addresses: lars.harrie@nateko.lu.se (L. Harrie), hanna.stigmar@lantm.lth.se (H. Stigmar).

0924-2716/$ – see front matter
doi:10.1016/j.isprsjprs.2009.05.004

selection of data layers and in real-time generalisation. The paper is organised as follows. Section 2 includes a literature review of quantifying map information. In Section 3, we present our work. First, map readability is divided into the *amount of information* and the *distribution of information*; then, we propose some analytical measures for each category. These measures are evaluated in a case study. The paper ends with our conclusions.

2. Background

In order to present a suitable amount of information in a map we need some sort of measure or guidelines. This turns out to be a delicate problem. First we need to specify the word *information*; what is information, and how can it be measured? According to Kellog (1995), "information technically refers to a reduction in uncertainty about events". Information thus gives us a specification of the so-called events: what is important and what is not. How do we then measure this importance? Can we somehow quantify it?

Bjørke (1996) discusses this matter and the use of Shannon information theory (or "The Mathematical Theory of Communication", Shannon and Weaver, 1964) in cartography. Previously, this approach has received some criticism as it does not cover all aspects of information in a map. The critics have argued that, as the theory decomposes the reality into simple elements, it misses the information in the map that is derived from the reader's previous knowledge. However, Bjørke points out that there are three aspects of information: syntactic, semantic, and pragmatic. While the syntactic aspect deals with the relationship among the symbols, the semantic aspect deals with the meaning of the symbols, and the pragmatic aspect with their application. The semantic and pragmatic aspects of information are very subjective. They depend to a great extent on the individual map reader; his/her preferences, opinions, and previous knowledge; but also on cultural and social factors, and the purpose of the map. Quantifying these aspects of the information is very complex, if not impossible. However, by separating the syntactic part of the information from the semantic and pragmatic parts, we can isolate the factual parts of the information, the objects themselves. Here we have a better opportunity to make a quantification, and we use information theory, as also argued by Bjørke.

One idea to quantify the map information is simply to count the number of objects in the map. However, looking at individual objects might not give a proper quantification as the map reader's subjective assessment has a major impact. How does the individual map reader determine what is one object? One segment of the road? One road line, from start to end points? One road network? Also, what impact does the visual distance have? Perhaps objects with different attribute types are regarded as different, while objects of the same attribute type are not. Another idea is to express the amount of information as the number of object points in the map. According to Biederman (1985), the human brain attaches great importance to the use of object points when reading and interpreting images; accordingly, these points would provide a suitable basis for the calculation of the amount of information. Yet some other ideas are to calculate the map area proportion covered by map objects, or the total line length of the objects (lines and polygons only).

Previous work on map readability is often based on the Shannon information theory (Shannon and Weaver, 1964). This theory is intended for message communication, and calculates the information content (entropy, H) in a sent message:

$$H = -[p_1 \log_2 p_1 + p_2 \log_2 p_2 + \cdots + p_n \log_2 p_n] \tag{1}$$

where the p_i are the probabilities for the messages or symbols i.

Sukhov (1967; 1970; in Li and Huang, 2002) applied this theory on cartographic communication in order to measure the information content in maps. First, the probabilities for the object types ($p_{IC,i}$) in the map are computed by

$$p_{IC,i} = \frac{K_i}{N} \tag{2}$$

where K_i is the number of symbols for object type i, and N is the total number of map symbols. Then, the entropy (H_{IC}) is defined as

$$H_{IC} = -\sum_{i=1}^{n} p_{IC,i} \log(p_{IC,i}). \tag{3}$$

However, as pointed out by Li and Huang (2002), this measure does not consider the spatial distribution of the objects. The entropy will be the same whether the objects are tightly assembled or more widespread, the same as applies for the amount of information measures described in the previous paragraph (number of objects, number of object points, object line length, and object area). Li and Huang argue that the spatial distribution of information is important, which is why the entropy calculation should also involve this aspect. Instead of using measures of the amount of information, spatially influenced measures are recommended. To identify the "region of influence", the empty spaces surrounding each map object, modelled as Voronoi regions, are used. Three measures are introduced: *geometric*, *topologic*, and *thematic*. The geometric measure calculates the entropy of the Voronoi regions. The probability (p_{SD}) for each object is calculated as the ratio between its Voronoi region size and the total map size:

$$p_{SD,i} = \frac{S_i}{S} \tag{4}$$

where S_i is the Voronoi area of the map objects $i = 1, 2, \ldots, n$, and S is the total map area.

The total map entropy for spatial distribution (H_{SD}) is then calculated as

$$H_{SD} = \sum_{i=1}^{n} p_{SD,i} \log p_{SD,i}. \tag{5}$$

Thus the map entropy is dependent on two properties: the number of regions (the entropy is larger with fewer regions) and the size differences of the regions (the more equally sized the regions are, the larger the entropy).

The topologic measure considers the Voronoi neighbours. Based on the ideas of Neumann (Neumann, 1994, in Li and Huang, 2002), the average number of neighbours for each Voronoi region (ANN) is computed as

$$ANN = \frac{N_S}{N_T} \tag{6}$$

where N_S is the sum of the neighbours for all map objects' regions, and N_T is the total number of map objects.

The thematic measure calculates the entropy based on the neighbour types. The assumption is here that the complexity increases when the objects are mixed, i.e., having neighbours of different types than themselves. To compute the thematic measure we start by calculating the probability (p_T) for each object region:

$$p_{T,j} = \frac{n_j}{N_i} \quad j = 1, 2, \ldots, M_i \tag{7}$$

where n_j is the number of neighbours of the same object type j, and N_i is the total number of neighbours.

The entropy of the object type i ($H_{Obj,i}$) is given as

$$H_{Obj,i} = \sum_{j=1}^{M_i} p_{T,j} \log(p_{T,j}) \tag{8}$$

and the total map entropy (H_T) is finally calculated as

$$H_T = \sum_{i=1}^{N} H_{Obj,i}. \tag{9}$$