INT. J. GEOGRAPHICAL INFORMATION SCIENCE, 2002
VOL. 16, NO. 7, 699–709

Taylor & Francis
Taylor & Francis Group

Research Article

Quantitative measures for spatial information of maps

ZHILIN LI and PEIZHI HUANG*
Dept. of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University Kowloon, Hong Kong; e-mail: Lszlli@polyu.edu.hk

(Received 18 June 2001; accepted 10 December 2001)

Abstract. The map is a medium for recording geographical information. The information contents of a map are of interest to spatial information scientists. In this paper, existing quantitative measures for map information are evaluated. It is pointed out that these are only measures for statistical information and some sort of topological information. However, these measures have not taken into consideration the spaces occupied by map symbols and the spatial distribution of these symbols. As a result, a set of new quantitative measures is proposed, for metric information, topological information and thematic information. An experimental evaluation is also conducted. Results show that the metric information is more meaningful than statistical information, and the new index for topological information is more meaningful than the existing one. It is also found that the new measure for thematic information is useful in practice.

1. Introduction

For many centuries, the map has been used as a medium for recording and presenting geographical information, and has played an important role in human activities. On a map, geographical information is expressed with cartographic symbols. As the map is regarded as a communication tool, cartographers are interested in the effectiveness of map design and the information content of a map (Knopfli 1983, Bjørke 1996). The former can be studied either through theoretical analysis or through map evaluation experiments similar to a clinic survey, but it is outside of the scope of this study and there will be no further discussion in this paper. Indeed, this paper discusses the information content of a map.

Interest in map information dates back to the late 1960s following the publication of the work on quantitative measures of information by Shannon (1948) and Shannon and Weaver (1949), which is normally termed as 'information theory' and was applied in communication theory. 'Entropy' is a quantitative measure for the information content contained in a message. The pioneering work in quantitative measurement of map information was done by Sukhov (1967, 1970), who considered the statistics of different types of symbols represented on a map. The entropy of these symbols is computed using the proportion of each type of symbol to the total as the probability

*Present address: Department of Computer Science, Shenzhea University, China.

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810210149416

used in the formula (equation 1). This is a direct application of Shannon's information measure to cartography, and is indeed a kind of statistical information. Later, Neumann (1987, 1994) did some work on the topological information of maps. He (1994) demonstrated the measurement of topological information for a contour map using the information concept developed in communication theory (Newmann 1994). In his work, a dual graph is formed to record the topological relationship between neighbouring contour lines, and then the entropy of the dual graph is computed. Quantitative measures for map information have been used for comparing the information content between maps and images, maps at different scale, for evaluation of map design and so on (Knopfli 1983, Bjørke 1996).

However, it is clear that spatial information is more than simple statistical information and topological information. It may also contain geometric and thematic information. In other words, the spatial position and distribution of map symbols should also be considered when a quantitative measure is designed for spatial information. In this study, Voronoi regions of symbols have been employed to model the spatial distribution of map symbols and then a new set of quantitative measures of the spatial information on the map.

This introduction is followed by an evaluation of existing measures (§2). Based on the evaluation results, a set of new quantitative measures is then introduced (§3) and these new measures are experimentally evaluated (§4). Finally, some conclusions are drawn (§5).

2. Evaluation of existing quantitative measures for map information

As stated in the introduction, two important pieces of work on map information have been carried out previously, one for statistical information and the other for topological information. In order to introduce new measures, it seems pertinent to conduct an evaluation of existing work to reveal the advantages and shortcomings of such measures.

2.1. *The quantitative measure of information in communication: Entropy*

Shannon (1948) was the first person to introduce entropy in the quantification of information. He employed the probabilistic concept in modelling message communication. He believed that a particular message is one element from a set of all possible messages. If the number of messages in this set is finite, then this number or any monotonic function of this number can be regarded as a measure of the information when one message is chosen from the set, all choices being equally likely. Based upon this assumption, information can be modelled as a probabilistic process. He then introduced the concept of 'entropy' to measure the information content.

Let X be the random message variable, when the probabilities of different message choices are $P_1, P_2, \dots P_i, \dots P_n$. The entropy of X can be computed as follows:

$$H(X)=H(P_1, P_2, \dots P_n)=-\sum_{i=1}^{n} P_i \ln(P_i) \qquad (1)$$

Statistically speaking, $H(X)$ reveals how much uncertainty the variable X has on average. When the value of X is certain, $P_i=1$, then $H(X)=0$. $H(X)$ is at its maximum when all messages have equal probability.

2.2. Statistical information of a map: entropy of symbol types

Sukhov (1967, 1970) has adopted the entropy concept for cartographic communication but only the number of each type of symbol represented on a map is taken into account. Let N be the total number of symbols on a map, M the number of symbol types and K_i the number of symbols for the ith type. Then $N = K_1 + K_2 + \dots + K_M$. The probability for each type of symbol on the map is then as follows:

$$P_i = \frac{K_i}{N} \tag{2}$$

where, P_i is the probability for the ith symbol type, $i = 1, 2, \dots$ M.

The entropy of the map can be calculated as follows:

$$H(X) = H(P_1, P_2, \dots, P_M) = -\sum_{i=1}^{M} P_i \ln(P_i) \tag{3}$$

The shortcomings of this measure of map information are revealed by figure 1, which is modified from (Knopfli 1983). Both maps consist of three types of symbols, i.e. roads, buildings and trees, and have exactly the same number of symbols for each type. That is, there is a total of 40 symbols, i.e. 7 for roads, 17 for buildings and 16 for trees. Therefore, according to the definitions in equations (2) and (3), both maps shown in figure 1 have the same amount of information, $H = 1.5$. However, the reality is that the distributions of symbols on these two maps are very different. In figure 1(*a*), the map symbols are mostly located on the right side of the diagonal along the lower/left to the upper/right direction, and the tree symbols are scattered among buildings. There are two rows of buildings along the main road. However, in figure 1(*b*), there is an area of trees on the left side of the diagonal along the lower/left to the upper/right direction, and there is an area of buildings in the opposite direction. The roads almost follow the diagonal. Indeed, they represent different natures of spatial reality.

In other words, the entropy computed in this way only takes into account the number of symbols for each type, and the spatial arrangement of these symbols is

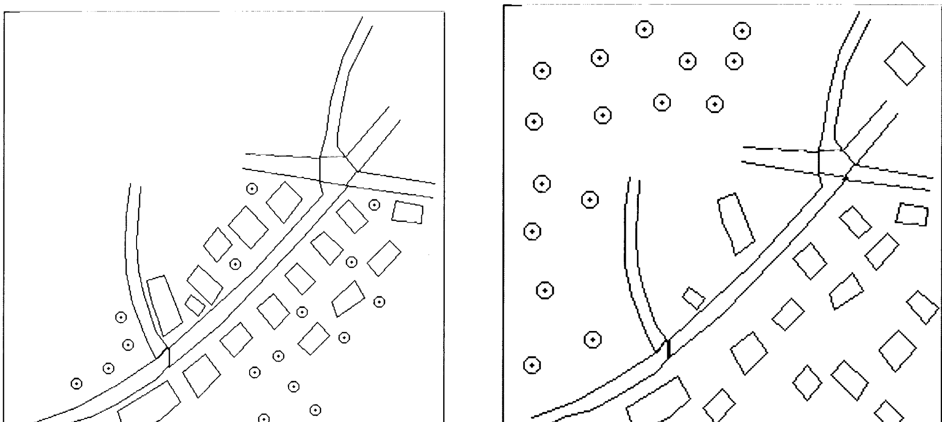


Figure 1. Two maps with the same amount of symbols but with different distribution (modified from Knopfli 1983).

completely neglected. Such a measure is purely statistical and thus is termed 'statistical information' in this paper. Indeed, it has little meaning in a spatial sense. Therefore, the usefulness of such a measure for maps is doubtful.

2.3. *Topological information of a map: entropy of neighbourhood*

Neumann (1994) proposed a method to estimate the topological information of a map. The method consists of two steps: (*a*) to classify the vertices according to rules such as their neighbouring relation and so on, to form a dual graph, and (*b*) to compute the entropy with equations (2) and (3). The method for the generation of this dual graph was put forward by Rashevsky (1955).

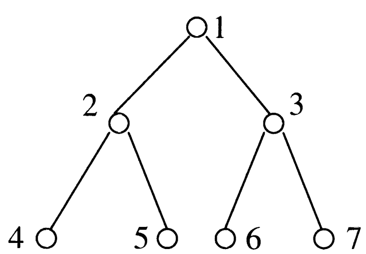
Figure 2(*a*) shows a dual graph which consists of seven vertices at three levels. There are three types of vertices, if they are classified by the number of neighbours. There are four vertices with only one neighbour, one vertex with two neighbours and two vertices with three neighbours. Then, $N=7$, $M=3$, thus, the probabilities of these three types of vertices are: $\frac{4}{7}$, $\frac{1}{7}$ *and* $\frac{2}{7}$. The entropy of this dual graph is then computed using equation (3) and the result is 1.38.

We now make a slight change to this graph by connecting the two vertices in the second level, as shown in figure 2(*b*). In this graph, there are four vertices with only one neighbour, one vertex with two neighbours and two vertices with four neighbours. The resultant entropy of this graph is exactly the same as that for figure 2(*a*) 1.38. However, it is clear that the graph shown in figure 2(*b*) looks more complex than that in figure 2(*a*). Thus, such topological information may not be able to reflect the true complexity of neighbour relations.

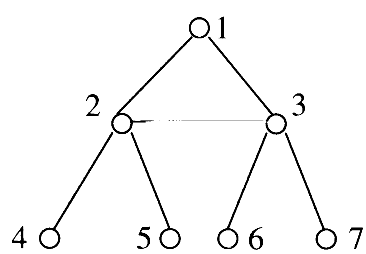
The question arising is 'how to form a dual graph for a given map?' It is, indeed, a difficult task to produce such a dual graph, e.g. for the map given in figure 1, because most of the map features are disjoint. A river network may be the type of feature convenient to form a dual graph. Indeed, in his study, Neumann (1994) produced a dual graph for a river network. He also tried to produce a dual graph for contour lines. This is possible because contour lines are nicely ordered according to their heights. The entropy computed by this method is only for the statistical distribution of vertex types instead of the spatial distributions.

Other types of information for a map

In fact, the usefulness of such topological information has also been questioned by Bjørke (1996). He provides another definition of topological information by



(a) A tree type of dual graph



(b) A dual graphs with network

Figure 2. Dual graphs for computation of topological information.

considering the topological arrangement of map symbols. He introduced some other concepts, such as positional entropy and metrical entropy. 'The metrical entropy of a map considers the variation of the distance between map entities. The distance is measured according to some metric' (Bjørke 1996). He also suggests to 'simply calculate the Euclidean distance between neighbouring map symbols and apply the distance differences rather than the distance values themselves'. The positional entropy of a map considers all the occurrences of the map entities as unique events. In the special case that all the map events are equally probable, the entropy is defined as $H(X)=\ln(N)$, where N is the number of entities.

3. New quantitative measures for spatial information of map

In the previous section, existing measures for information contents on a map have been reviewed and evaluated. Their limitation should be clear. It is now pertinent to introduce new measures in this section, which should be sound in theory. The usefulness in practice will be evaluated in §4.

3.1. *The line of thought*

Communication theory is based on order. It doesn't consider any spatial distribution. Therefore, it could be dangerous to follow the line of thought developed in communication theory. That is, a completely new line of thought must be followed.

It is a commonplace that a map contains the following information about features:

- (Geo)metric information related to position, size and shape.
- Thematic information related to the types and importance of features.
- Spatial relations between neighbouring features implied by distribution.

Therefore, a set of measures needs to be developed, one for each of these: metric, topologic and thematic information.

To consider metric information, the position of a feature is not a problem. On the other hand, the consideration of size and shape of a feature is not an easy job. One approach to describe the size of a feature is simply based on the size of the symbol. However, a serious deficiency with this absolute approach lies in its ignorance of the following facts:

- The size of a point symbol is always smaller than an areal symbol.
- The relative space of a feature, i.e. the empty space surrounding the feature, separates the feature from the rest. The larger the empty space surrounding the feature, the more easily it can be recognised.

As map features share the empty space surrounding them, it is necessary to determine the share of each feature. In this case, the map space needs to be tessellated by feature-based tessellation (Lee *et al.* 2000). The Voronoi diagram seems to be the most appropriate solution. A Voronoi diagram is essentially a partition of the 2-D plane into N polygonal regions, each of which is associated with a given feature. The region associated with a feature is the locus of points closer to that feature than to any other given feature. Figure 3 shows the Voronoi diagram of the maps shown in figure 1. The polygonal region associated with a feature is normally called the 'Voronoi region' (or Thiessen polygon) of that feature, and it is formed by perpendicular bisectors of the edges of its surrounding triangles. Such a Voronoi region is a

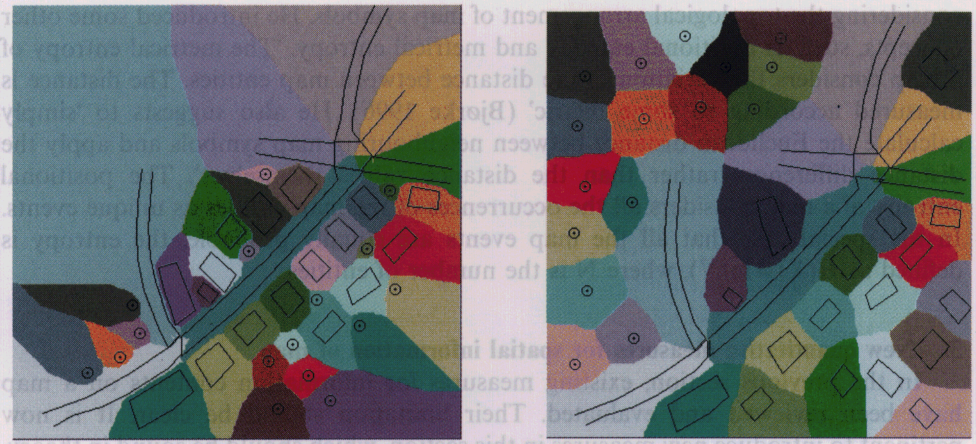


Figure 3. Voronoi diagrams of the maps shown in figure 1.

'region of influence' or 'spatial proximity' for a map feature. All these Voronoi regions together will form a pattern of packed convex polygons covering the whole plane (neither any gap nor any overlap). Thus a Voronoi diagram of a map feature is its share of its surrounding space.

Indeed, the Voronoi region is not only adequate for the determination of the share of surrounding empty space for a map feature, but also good for the neighbour relationship (Gold 1992). This is because the Voronoi region of a feature is determined by two factors, (*a*) the size of feature and (*b*) the neighbouring features. Indeed, Chen *et al.* (2001) have used Voronoi regions to describe spatial relations between map features.

For these reasons, the authors have attempted to relate the spatial information of map features to their Voronoi regions to develop a set of new quantitative measures. However, detailed discussion of the formation of Voronoi regions is outside the scope of this paper. Algorithms for the generation of a Voronoi region in vector mode have been presented by Okabe *et al.* (1992) and a raster-based algorithm has recently been proposed by Li *et al.* (1999). Therefore, no further discussion on this topic will be presented in this paper.

3.2. *(Geo)Metric information of a map: entropy of Voronoi regions*

(Geo)Metric information here considers the space occupied by map symbols only. In this case, an analogy to the entropy of a binary image is used. That is, if the space occupied by each symbol is similar, the map has a larger amount of information. If the variation is very large, the amount of information is smaller. This can be achieved by using the ratio between the Voronoi region of a map system over the enclosed area of the whole map as the probability used in the entropy definition. Let S be the whole area and be tessellated by S_i, $i=1, 2, \ldots N$. Such a probability can then be defined as follows:

$$P_i = \frac{S_i}{S} \tag{4}$$

The entropy of the metric information, denoted as H(M), can then be defined as follows:

$$H(M)=H(P_1, P_2, \ldots, P_n)=-\sum_{i=1}^{n}\frac{S_i}{S}(\ln S_i-\ln S) \quad (5)$$

$H(M)$ has its maximum when P_i has the same value for all $i=1, 2, \ldots N$. In other words, when the area S_i is equal. Mathematically,

$$H(M)_{max}=\mathrm{H}(\mathrm{P}_1, P_2, \ldots, P_n|_{P_1=P_2=\ \ldots\ =P_n})=\log_2 n \quad (6)$$

For example, the two maps shown in figure 4 have different amounts of metric information, although both are tessellated by nine polygons. The map in figure 4(*b*) has the maximum $H(M)$ for any tessellation into nine polygons.

In the case of a map, it is clear that for the same number of features, the entropy will be larger if the symbols are more evenly distributed. However, it is clear that such entropy is related to the number of map symbols, and thus it would not be convenient to compare two maps with a different number of symbols. In order to overcome this shortcoming, entropy can be normalised as follows:

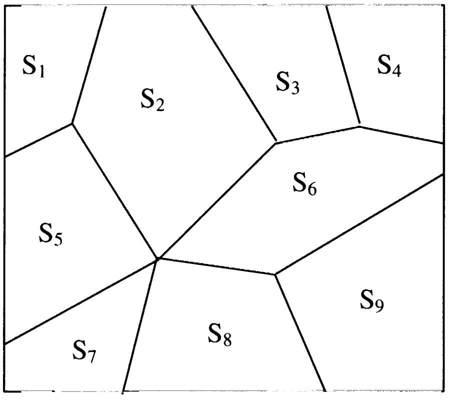
$$H(M)_N=\frac{H(M)}{H(M)_{max}} \quad (7)$$

Another possible measure is the ratio, R_M, between the mean of the areas (m_A) and the standard deviation σ_A.·

$$H(TM)=\sum_{i=1}^{N}H_i(TM) \quad (8)$$

$$\sigma_A=\sqrt{\frac{\sum_{i=1}^{n}(A_i-m_A)^2}{n-1}} \quad (9)$$

$$R_M=\frac{m_A}{\sigma_A} \quad (10)$$



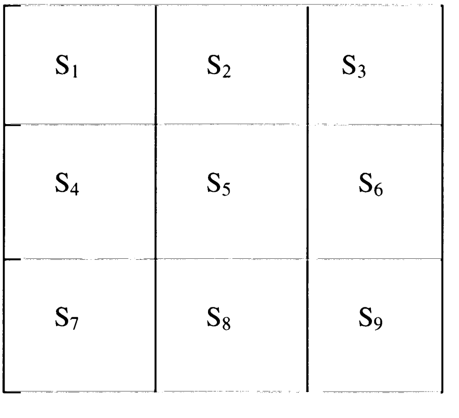


Figure 4. Two different tessellations of an area, resulting in two different amounts of metric information.

3.3. *Topological information of a map: Voronoi neighbours*

As has been discussed in the previous section, the construction of dual graphs for map features is a difficult task because the vast majority of map features are disjoint. However, with the Vonoroi region, all features have been connected together to form a tessellation. The generation of a dual graph for map features could be replaced by the dual graph of the Voronoi region of these features. This is illustrated in figure 5. Figure 5(*a*) is the Voronoi region and figure 5(*b*) is the corresponding dual graph.

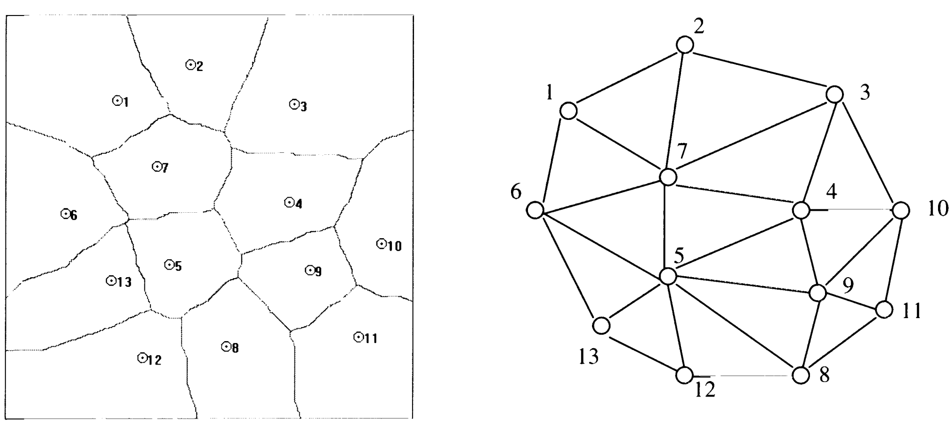
The entropy of this map can then also be computed, as for the graphs in figure 2. The entropy computed using the number of nodes in the graph is that of the distribution of different kinds of vertices (§2.3). It does not really reflect the complexity of the dual graph directly. Indeed, it is sometimes misleading, as shown in the case of figure 2. Therefore, a new index needs to be designed. As the complexity of a dual graph can be indicated by the number of neighbours for each vertex, this number is already a good measure. In order to compare the complexity of the dual graph with different vertices, the average number of neighbours for each vertex may be used as a value to indicate the complexity of a dual graph.

Let, N_s be the sum of the numbers of neighbours for all vertices and N_T the total number of vertices in a dual graph. Then, the average number of neighbour for each vertex is:

$$H_T = \frac{N_s}{N_T} \tag{11}$$

3.4. *Thematic information of a map: Entropy of neighbour types*

Thematic information is related to the thematic types of features. It is clear that, if a symbol has neighbours all of the same thematic type, then the importance of this symbol is very low, in terms of thematic meaning. On the other hand, if a symbol has neighbours of different thematic types, it should be regarded as having higher



(a) Voronoi diagram (b) Dual: Triangulation

Figure 5. A Voronoi diagram and its dual graph. (*a*) Voronoi diagram. (*b*) Dual: triangulation.

thematic information. Here, the neighbours are also defined by the immediately-neighbouring Voronoi regions. For example, the symbol 5 in figure 5 (*a*) has symbols 7, 6, 13, 12, 8, 9, and 4 as neighbours.

Based on this assumption, the thematic information of a map symbol can then be defined. Suppose, for the *i*th map symbol, there are in total N_i neighbours and M_i types of thematic neighbours. There are in total n_j neighbours for *j*th thematic type. Then the probability of the neighbours with *j*th thematic type is as follows:

$$P_j = \frac{n_j}{N_i} \qquad j = 1, 2, \dots M_i \tag{12}$$

The thematic information of the *i*th map symbol is then as follows:

$$H_i(TM) = H(P_1, P_2, \dots, P_{M1}) = -\sum_{j=1}^{Mi} \frac{n_j}{N_j} \ln\left(\frac{n_j}{N_j}\right) \tag{13}$$

Suppose there are in total N symbols on a map; the total amount of thematic information for this map is then

$$H(TM) = \sum_{i=1}^{N} H_i(TM) \tag{14}$$

4. An evaluation

In the previous section, a set of new measures has been proposed for the spatial information of a map. It is appropriate to conduct some experimental tests on the usefulness of these new measures and also to see whether these new measures are more meaningful than existing ones.

4.1. *Metric information vs statistical information*

The first test is on metric information. The two maps in figure 1 were used. The corresponding Voronoi regions are shown in figure 3. The results for the entropy of Voronoi regions (equation 5) and the ratio (equation 10) between mean (equation 8) and standard deviation are listed in table 1.

From table 1, it is clear that the map shown in figure 1(*b*) contains more metric information than that in figure 1(*a*). Considering the fact that they should have the same amount of statistical information, as pointed out in §2, it seems logical to claim that these measures are more appropriate than the statistical information.

4.2. *Topological information: new* versus *old*

The second test is on the topological information. Using the new index, the results for the two graphs in figure 2 would be different. In figure 2(*a*), there are seven vertices and the total number of neighbours for all vertices is twelve. The average number of neighbours for each vertex is 1.7. In figure 2(*b*), there are also seven vertices, but the total number of neighbours for all vertices is fourteen. The

Table 1. Metric information of the two maps in figure 1. (The area of the map is a unit).

	$H(M)$	$R_M(\%)$	S_E	$\sigma(\%)$
Map in figure 1 (*a*)	4.2848	80.51	0.025	2.84
Map in figure 1 (*b*)	5.1260	96.32	0.025	1.51

average number of neighbours for each vertex is 2.0. It is then clear that figure 2 (*b*) is more complex than figure 2 (*a*).

To further elaborate the adequacy of this new measure, the index values for the Voronoi regions shown in figures 3 and 5 are also computed and listed in table 2. It shows that the map shown in figure 1(*a*) is more complex than that shown in figure 1(*b*). This is because the three symbols are mixed into the building symbols.

4.3. *Thematic information*

It is intuitive that that the more types of (meaningful) symbols in a map, the greater thematic information contents it has. With the same type of symbols, the distribution of the symbol is the only thing that matters in the computation of thematic information. The symbol distributions in figure 1(*a*) and figure 1 (*b*) are different although the number of types is identical, thus the thematic information of these two maps will be different.

The thematic information for the two maps shown in figure 1 is also computed and shown in table 3. It is very clear that the map shown in figure 1(*a*) has more thematic information because the tree symbols are scattered around building symbols. On the other hand, the thematic information contained by the map shown in figure 1(*b*) is lower because the three types of symbols are quite clustered. Therefore, the thematic information defined in this way also seems very meaningful.

5. Conclusion

In this paper, the existing quantitative measures for map information are evaluated. It is pointed out that these are only measures for statistical information and some sort of topological information, but do not consider the spaces occupied by symbols and their spatial distribution. As a result, a set of new quantitative measures is proposed, for metric information, topological information and thematic information. In these measures, the Voronoi regions of map features play a key role, which not only offer metric information but also some sort of thematic and topological information. An experimental evaluation is also conducted. Results show that metric information is more meaningful than statistical information and the new index for topological information is also more meaningful than the existing one. It is also found that the new measure for thematic information is useful in practice.

Quantitative measurement of the information content of maps is an important

Table 2. The average number of neighbours for figures 3 and 5.

	N_T	N_S	H_T
Figure 3(*a*)	40	206	5.15
Figure 3(*b*)	40	188	4.70
Figure 5	13	54	4.15

Table 3. Thematic information of the two maps in figure 1.

	Thematic Information $H(TM)$
Map in figure 1 (*a*)	28.2
Map in figure 1 (*b*)	16.4

issue in spatial information science. It has been used for comparing the information content between maps and images, maps at different scale, for evaluation of map design, and so on (Knopfli 1983, Bjørke 1996). Effective quantitative measures are of great importance not only for understanding the characteristics of spatial information but also for the effective use of spatial information.

Acknowledgment

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. PolyU 5094/97E).

References

Bjørke, J. T., 1996, Framework for entropy-based map evaluation. *Cartography and Geographical Information Systems*, **23**, 78–95.

Chen, J., Li, C. M., Li, Z. L., and Gold, C., 2001, A Voronoi-based 9-Intersection model for spatial relations. *International Journal of Geographical Information Science*, **15**, 201–220.

Gold, C. M., 1992, The meaning of 'Neighbour'. In *Theories and Methods of Spatial-Temporal Reasoning In Geographical Space*, (Lecture Notes in Computer Science, No. 39), edited by A. Frank, I. Campari, and U. Formentini, (Berlin: Springer-Verlag), pp. 220–235.

Knopfli, R., 1983, Communication theory and generalization. In *Graphic Communication and Design in Contemporary Cartography*, edited by D. R. F. Taylor (New York, Chichester: John Wiley & Sons Ltd), pp. 177–218.

Lee, Y. C., Li, Z. L., and Li, Y. L., 2000, Taxonomy of space tessellation. *ISPRS Journal of Photogrammetry and Remote Sensing*, **55**, 139–149.

Li, C., Chen, J., and Li, Z. L., 1999, A raster-based method for computing Voronoi diagrams of spatial objects using dynamic distance transformation. *International Journal of Geographical Information Science*, **13**, 209–225.

Neumann, J., 1987, Gnoseological aspects of improving the geometrical component of the space-time model in cartography. In *Proceedings, 13th International Cartographic Conference ICA*, Morelia, Mexico, IV: pp. 569–584.

Neumann, J., 1994, The topological information content of a map: an attempt at a rehabilitation of information theory in cartography. *Cartographica*, **31**, 26–34.

Okabe, A., Boots, B., and Sugihara, K., 1992, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams* (Chichester: John Wiley).

Rashevsky, N., 1955, Life, information theory and topology. *Bulletin of Mathematical Biophysics*, **17**, 229–235.

Shannon, C. E., 1948, A mathematical theory of communication. *The Bell System Technical Journal*, **27**, 379–423 & 623–656.

Shannon, C. E., and Weaver, W., 1949, *The Mathematcal Theory of Communication* (Urbana, Illinois: University of Illinois Press).

Sukhov, V. I., 1967, Information capacity of a map entropy. *Geodesy and Aerophotography*, **X**, 212–215.

Sukhov, V. I., 1970, Application of information theory in generalization of map contents. *International Yearbook of Cartography*, **X**, 41–47.