ENTROPY ANALYSIS OF IMAGES USED IN DISASTER INFORMATION SYSTEMS

Serdar Bilgi, Cengizhan Ipbuker
Istanbul Technical University[1]

Keywords

Entropy, Statistical Information, Object Resolution in Disaster Information Systems, Matlab Image Processing Tools

Abstract

For designers, creating a conceptual data model is one of the challenging tasks of GIS. The main problem for such an automation task is how to quantify and formalize the properties of the process as an information system. Quantifying the efficiency and information content of the geometric data used by the information system is on paramount importance.

Analyzing the intensity of geometric objects in a disaster information system is a considerable factor in the installation progress and defining the optimal data. Optimal data is defined as the minimum necessary data for the aim and usage of the system which does not cause complexity on the screen.

Interest in the information amount in maps dates back to the late 1960s following the publication of work on quantitative measures of information. These works have come to be known as 'information theory,' and was also applied to communication theory.

Entropy theorem is used for measuring the statistical information. Entropy represents the measurement of the uncertainty of a system**.** This measurement does have an explicit link to information content and is associated with attempts to quantify information transfer within a communication system. Although entropy is the second law of thermodynamics, it was introduced in the quantification of information. The entropy concept for cartographic communication, takes into account the number of each type of objects represented on a map.

In this study, edge detection and thresholding methods and several algorithms of these methods are tested in Matlab software to determine the building objects in images. Prewitt, Canny, Roberts, Laplacian of Gaussian, Zero-Cross and Sobel algorithms are used for edge detection. Grey value algorithms are used for thresholding methods. The results are investigated in order to calculate the entropy.

1. Introduction

There are four components of GIS. These include data, hardware, software, and staff. Data is the most important and expensive component. Hardware and software constitute approximately 10%-20% of the overall budget, while staff costs comprise 10%; however, the budget for data consists of nearly 70% - 80% of the total system budget (Ekincioğlu, 1998).

Geographic Information Systems (GIS) are perfect tools for planning and disaster management.

Various disasters, including floods, fires, landslides, those resulting from the actions of people, and especially earthquakes have occurred. These disasters cause:

[1] *Dept. of Geodesy and Photogrammetry Eng., 34469, Istanbul-TURKEY, bilgis@itu.edu.tr, buker@itu.edu.tr*

- Social and cultural damages like death or the impairment of social living environments,
- Destruction of the environment like the spreading of hazardous materials and damaging wildlife,
- Materialistic damages like loss of agricultural products and collapse of city infrastructure

Efforts for reducing the loss associated with disasters consist of the studies and preparations before the disaster. Performing this task requires rapid and systematic access, and usage of the related information in the hazardous regions (Sahin et al., 2002).

The object is everything that can be monitored and has a historicity or a possibility existence in space. An object with a spatial reference should satisfy planimetric geometry, topological relations and attributes.

Object catalogue is the conceptual data model of a geographic information system and defines the frame of data which would be gathered for the system. Object catalogue classifies the object types with their geometric information and attributes under object groups. These geometric and non-geometric data describe the topological character of the region surface. In the planning stage, the object catalogue has to be formed due to the aim and usage of system. Some advantages of the object catalogue are:

- Modeling the GIS infrastructure with definite standards, thus preventing the enlargement of data without any control.
- Managing of the same data by different institutions.
- Efficiency increasing with interdisciplinary work.
- Elimination of data exchanging problems as a result of the standards prepared in the field.

Analyzing the intensity of topographic objects in a disaster information system is a considerable step in gathering data during the installation progress and defining the “optimal” data. Optimal data is defined as the minimum data needed for the aim and usage of the system which does not cause complexity on the screen. Thus, data excess and time-wasting are prevented by gathering optimal data in disaster information systems.

Maps have been created and used for centuries as abstractions of real World, and as such are capable of giving us a picture of our environment which distils its full complexity into an effective graphical rendering. By considering its general nature, the term ‘complexity’ can be defined as applicable to the description of cartographic representations (maps). Attempts have thus been made to measure complexity through practical testing (Fairbairn, 2006).

For designers, creating a conceptual data model is one of the challenging tasks of GIS. The main problem for such an automation task is how to quantify and formalize the properties of the process as an information system (Bjørke, 1996). Thus, quantifying the efficiency and information content of the base map used by the information system is of paramount importance. However, there remains no metric measure for quantifying the cartographic efficiency of a map. As such, Shannon’s theory based on entropy may be a useful measure by which to optimize a map’s informational content.

In this study, edge detection and thresholding methods and several algorithms of these methods are tested in Matlab software to determine the building objects in images. Prewitt, Canny, Roberts, Laplacian of Gaussian, Zero-Cross and Sobel algorithms are used for edge detection. Grey value algorithms are used for thresholding methods. The results are investigated in order to calculate the entropy.

2. Map as Communication Tool

Maps have been used primarily as a medium for presenting geographical information. There are many criteria for a good map; one being information content. As a map is regarded as a communication tool, cartographers are interested in the effectiveness of both map design and the information content of a map (Knopfli 1983, Bjørke 1996).

Interest in map information dates back to the late 1960’s following the publication of quantitative measures of information by Shannon (1948) and Shannon and Weaver (1949).

This work has come to be known as 'information theory,' and was also applied to communication theory. 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 (Li and Huang, 2002).

3. Statistical Information of a Map: Entropy Analysis

Entropy is the second law of thermodynamics and, briefly defined, represents the measurement of the uncertainty of a system (URL 1). This measurement does have an explicit link to information content and is associated with attempts to quantify information transfer within a communication system (Fairbairn, 2006).

Sukhov (1967, 1970) has adopted the entropy concept for cartographic communication, but only took into account the number of each type of symbol represented on a map. Thus, Let **N** be the total number of symbols on a map, **M** the number of symbol types and $\mathbf{K_i}$ the number of symbols for the each type. The probability for each type of symbol on the map is then as follows:

$$N=K_1+K_2+.....+K_m \quad (3.1)$$

The probability for each type of symbol on the map is then as follows:

$$P_i=\frac{Ki}{N} \quad (3.2)$$

$\mathbf{P_i}$ is the probability for the ith symbol type, i=1, 2, . . . M.

The map entropy analysis requires the calculation of the following entropy quantities:

$$H(X)=H(P_1,P_2,....P_M)=-\sum_{i=1}^{M} Pi\ln(Pi) \quad (3.3)$$

4. Object Determination

In order to calculate the entropy value, number of buildings has to be known. Two methods and several algorithms are tested to define the objects in satellite images. Type of the objects is buildings. Used images are in "TIF" (**T**agged **I**mage **F**ile) format to use in Matlab software. Image Processing Tools of Matlab software was used for the study.

Image Processing: Image processing involves changing the nature of an image in order to either;

- Improve its pictorial information for human interpretation,
- Render it more suitable for autonomous machine perception (McAndrew, 2004).

4.1 Edge Detection Method

Edges contain some of the most useful information in an image. Edges might be used to measure the size of objects in an image; to isolate particular objects from their background; to recognize or classify objects (McAndrew, 2004). Several methods using different algorithms of Matlab software are used. Edge Detection algorithms are

- Prewitt algorithm
- Canny algorithm
- Roberts algorithm
- Laplacian of Gaussian algorithm
- Zero-Cross algorithm
- Sobel algorithm

4.2 Thresholding Method

Thresholding can be used to isolate objects from the background in the image. Grey level criteria of the pixels are used to determine the objects. If one pixel has a grey level value more than a threshold, it will be white, else it will be black. Two commands are used in Matlab software.

- imshow
- im2bw

5. Case Study

An application in introduction level is done for determining the objects in image. Firstly, the image which has a regular settlement is studied.

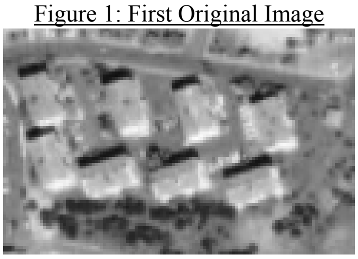
5.1 Edge Detection Method

Command line used in Matlab is given below. Only the algorithm name changes in 4th line. The first original monochrome image (Figure 1) is in .tif format including 8 buildings.

Command Lines

```
close all;
clear all;
I = imread('image_bw_v1.tif');
BW1 = edge(I,'prewitt'/'sobel'/…);
imshow(BW1);
```

Figure 1: First Original Image



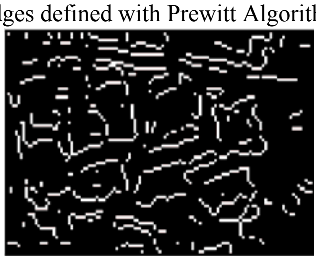
The results of the algorithms are shown in figure 1.

Figure 1. Results of Different Edge Detection Algorithms

Edges defined with Prewitt Algorithm

Edges defined with Canny Algorithm

Edges defined with Roberts Algorithm



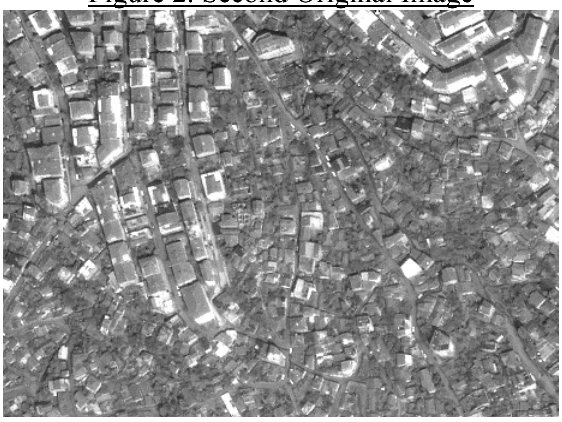
Edges defined with Laplacian of Gaussian Algorithm

Edges defined with Zero-Cross Algorithm

Edges defined with Sobel Algorithm

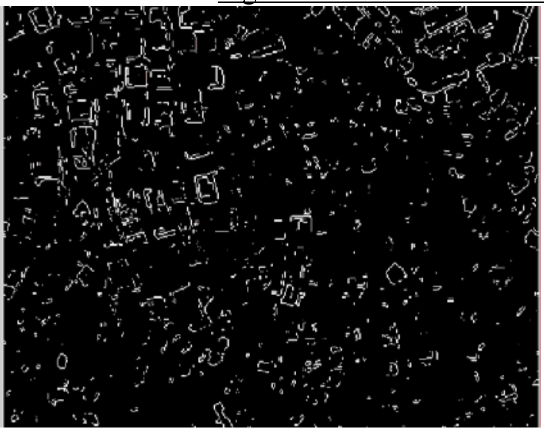
Second monochrome image is captured by Ikonos satellite with 1 m resolution and given in figure 2.

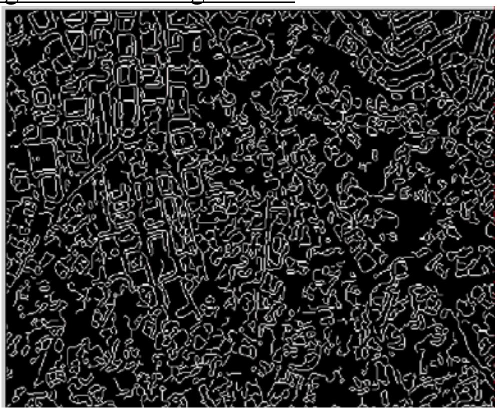
Figure 2. Second Original Image

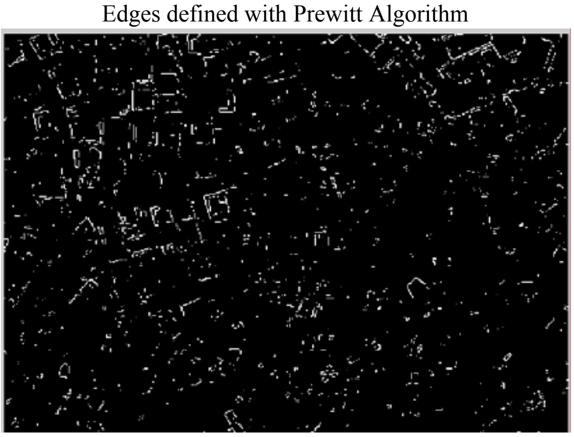


The results of the algorithms for second image are shown in figure 2.

Figure 2. Results of Different Edge Detection Algorithms

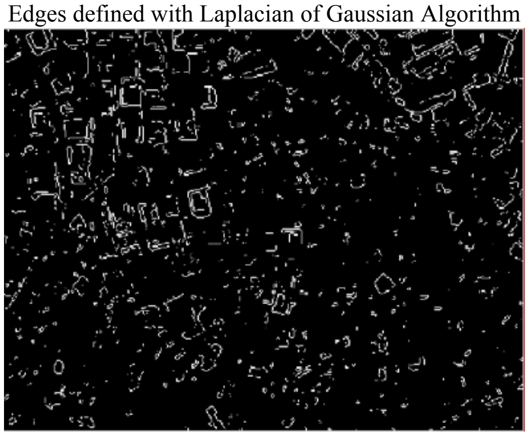

Edges defined with Prewitt Algorithm


Edges defined with Canny Algorithm


Edges defined with Roberts Algorithm


Edges defined with Laplacian of Gaussian Algorithm

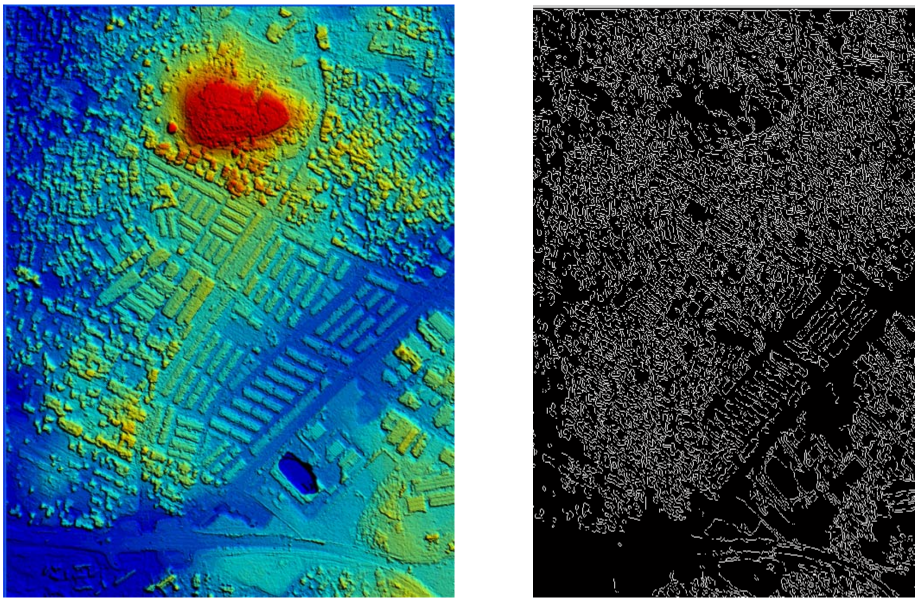

Edges defined with Zero-Cross Algorithm


Edges defined with Sobel Algorithm

Contrasts of the both images were increased but the results did not change. Edges of the buildings can be defined generally, but the lines were not closed. So instead of polygon geometry, polyline geometry could be produced. In other words, another solution must be added to the edge detection results for closing the polylines to polygons for the recognition of "one polygon means one separate object" by the computer.

Also a third image from Orbview satellite is tested with Canny algorithm of edge detection method (figure 3). Original image is colored due to the heights of the buildings. It is thought that determining the buildings by using their different colors from the ground, can give more significant results for defining the edges.

Figure 3. Orbview Satellite Image and Result of Canny Algorithm



Similarly, edges of the buildings can be defined generally, but the lines were not closed.

5.2 Thresholding Method

It is thought that determination of the buildings by using the grey level of their roofs can give more significant results for thresholding them from other objects in the image.
Two commands are used for thresholding method. The first one is "imshow" command. This command is for monochrome images only. Both usages of imshow command are tested for thresholding method in Matlab software. These are;

- Single threshold value for grey level
- Interval threshold values for grey level

Other command is "im2bw". This command is for both multispectral and monochrome images.

5.2.1 Single threshold value

Grey level of the pixel; more than s → it is white; less than s → it is black

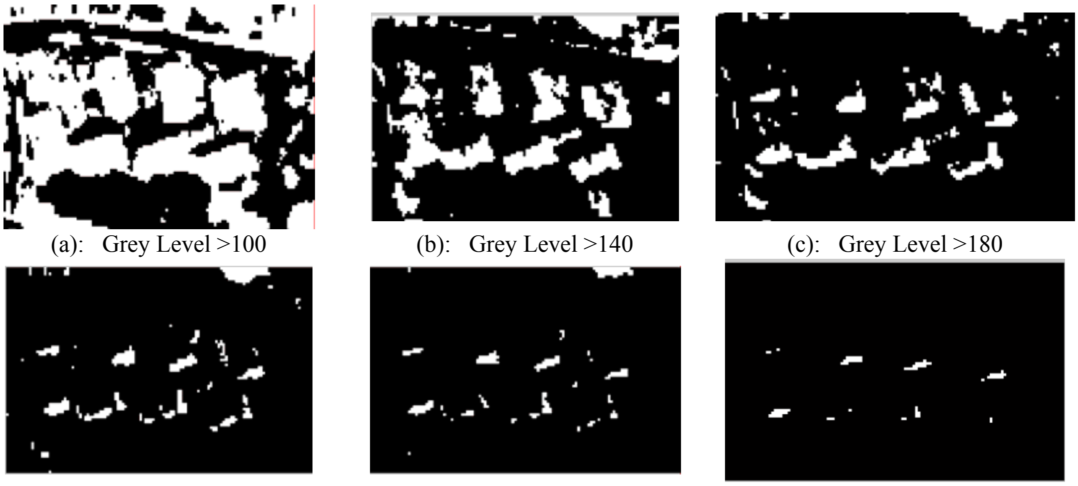
The first image is used for thresholding.

Command Lines

```
close all;
clear all;s=imread('image_bw_v1.tif');
imshow(s),figure,imshow(s>100);
```

Command lines of thresholding method in Matlab are shown above. "s" variable in the command means grey level value of the pixel. The results of thresholding algorithms are shown in figure 3.

Figure 3. Results of Single Threshold Value



(a): Grey Level >100 (b): Grey Level >140 (c): Grey Level >180
(d): Grey Level >200 (e): Grey Level >220 * (f): Grey Level >240

5.2.2 Interval Threshold Values

The first image is used for thresholding. If;

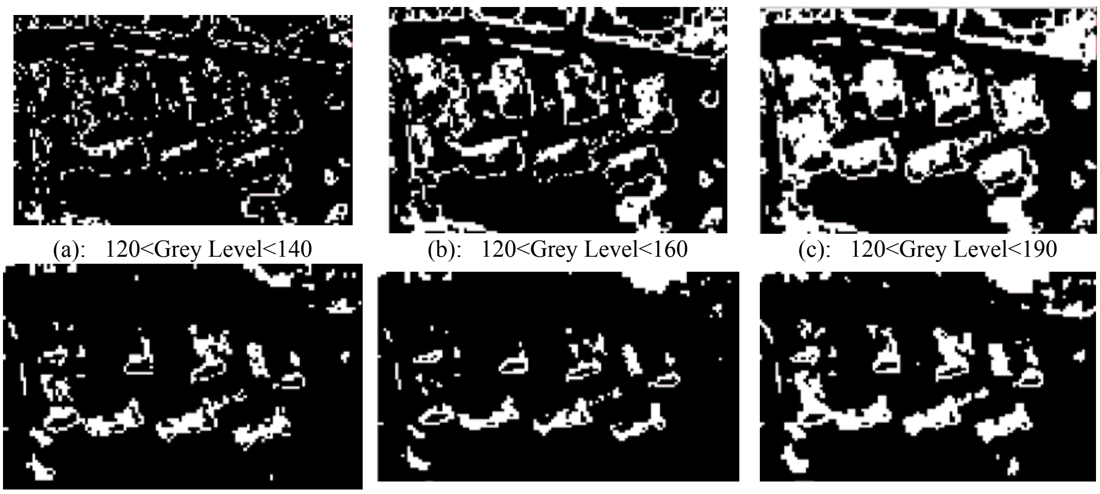
Grey level of the pixel;
- between s1 and s2 → it is white
- excluding → it is black

Command lines of thresholding method in Matlab are shown below. “s” variable in the command means grey level values of the pixel. The results of the algorithms are shown in figure 4.

Command Lines

```
close all;
clear all;
s=imread('image_bw_v1.tif');
imshow(s),figure,imshow(s>50 & s<100);
```

Figure 4. Results of Interval Threshold Values



(a): 120<Grey Level<140 (b): 120<Grey Level<160 (c): 120<Grey Level<190
(d): 170<Grey Level<220 * (e): 180<Grey Level<240 * (f): 160<Grey Level<240

As the intervals of grey level decreases, it becomes difficult to select the buildings. Best results in test are obtained the intervals 170< s <220 and 180< s <240.

5.2.3 Single threshold value for Multispectral Images

Colored image is used for this command. The image is in .tif format and includes 8 buildings.

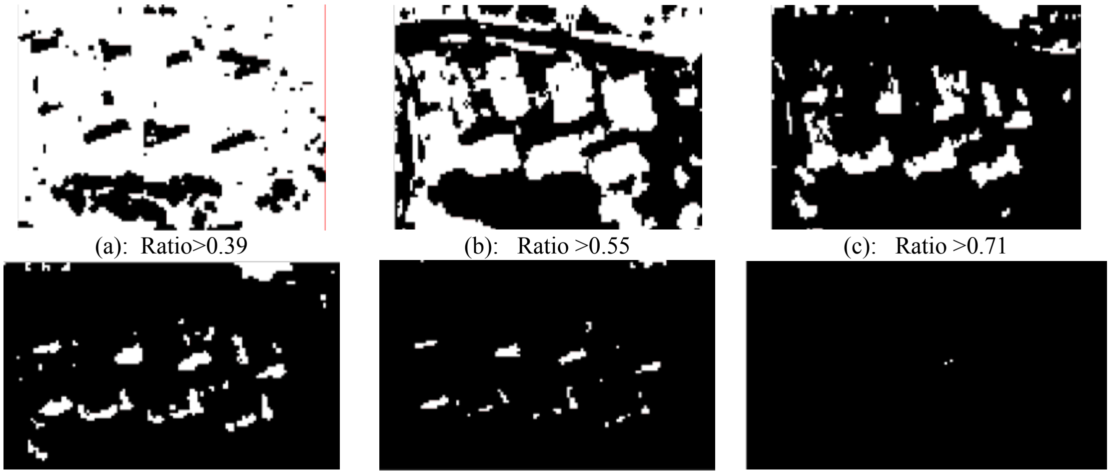
Command Lines

```
close all;
clear all;
s=imread('image_bw_v1.tif');
imshow(s),figure,im2bw(s,0.43);
```

Figure 4. Original Image

Command lines of thresholding method in Matlab are shown above. The number used after im2bw command is between 0 and 1 (0|s|1). This number means grey level/white value (Grey Level/255) of the pixel. The results of the algorithms are shown in figure 5.

Figure 5. Results of Single Threshold Value



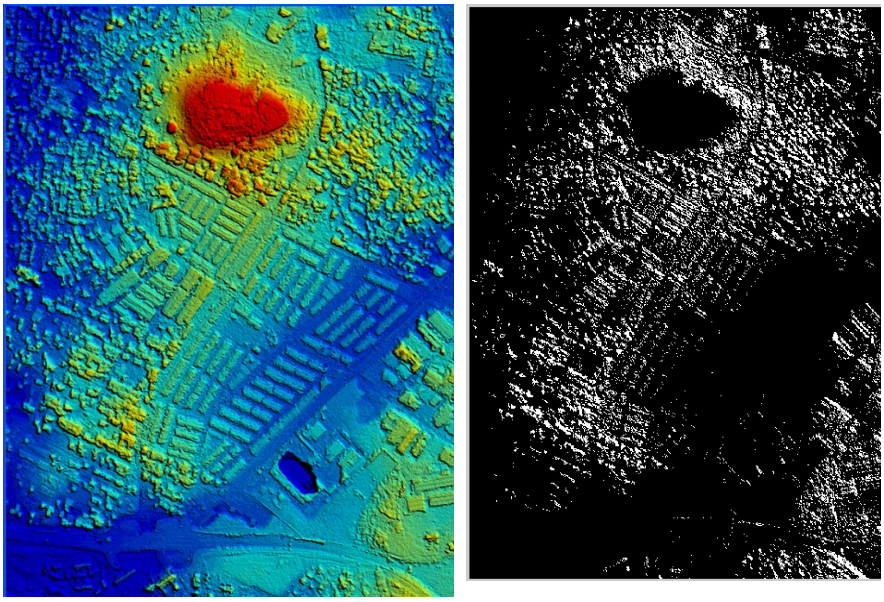
(a): Ratio>0.39 (b): Ratio >0.55 (c): Ratio >0.71

(d): Ratio >0.78 (e): Ratio >0.86 * (f): Ratio >0.98

As seen on the original image, roofs are red color. Red color objects cannot be seen frequently, as much as roofs, in nature. For this reason, test has more successful results than monochrome image.

Also an image from Orbview satellite is tested with thresholding method. Original image is colored due to the heights of the buildings. It is thought that selecting the buildings by using their different colors from the ground can give more significant results for defining the buildings.

Figure 5. Original Image and Result of Thresholding Method



Point clusters are obtained at the result. Especially, similar colored buildings with the ground color (buildings have less height) could not be selected successfully.

6. Conclusion

Entropy is computed with the number of objects in the image. The result is relevant with existence probability for each type of object through the total number of objects. When the value of X (P_1, P_2, …) is certain, P_i=1, then H(X)=0. H(X) is at its maximum when all probabilities for each type of object have equal values (Li and Huang, 2002).

Map entropy analyses can produce different quantitative measures for interpreting information efficiency during the map creation process in information systems. These measures may give some statistical evidence for determining the object resolution of digital cartographic products.

As seen in the results, determination of building objects using different methods is not successful yet. Some new methods are being tested still. New tests are not only for buildings, but also for other type of objects on the images. So it is thought that all type of objects could be determined on the images and the results could be taken into account for calculation of entropy.

7. References

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

Ekincioğlu, İ. (1998). Konumsal Bilgi Sistemlerinin Kurumlar Arası Kullanılması. (PhD Thesis, YTU Institute of Science and Technology).

Fairbairn D. (2006). Measuring Map Complexity. The Cartographic Journal, 43(3), 224-238.

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.

Li, Z. and Huang, P. (2002). Quantitative measures for spatial information of maps. International Journal of Geographical Information Science, 16(7), 699 – 709.

McAndrew Alasdair. (2004). An Introduction to Digital Image Processing with Matlab Notes, School of Computer Science and Mathematics, Victoria University of Technology.

Sahin, M., Uçar, D., Baykal, O., Türkoğlu, H., Tarı, E., İpbüker, C., Musaoğlu, N., Göksel, Ç., Coşkun, M.Z., Kaya, Ş., Yiğiter, R., Erden, T., Karaman, H., Yavaşoğlu, H., Bilgi, S.,

Üstün, B. (2002), TABIS: Development of a National Database Using Geographical Information Systems and Remote Sensing System and Standards for a Disaster Management Decision Support System Project, I.T.U., Istanbul.
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 Mathematical 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.
URL 1: http://www.biltek.tubitak.gov.tr, Last Accessed September 2007.

Author Biography

Serdar BILGI

He is a Geodesy and Photogrammetry Engineer and has his Master of Science degree on 2003 from Istanbul Technical University Science and Technology Institute on "Spatial Object Modeling of Disaster Information Systems". He is Research Assistant in Istanbul Technical University, Faculty of Civil Engineering, Department of Geodesy and Photogrammetry, Division of Cartography, since 2001.

Cengizhan IPBUKER

He is a Geodesy and Photogrammetry Engineer and has his Master of Science degree on 1994 from Istanbul Technical University Science and Technology Institute on "Geometric Accuracy of Satellite Images for Producing Map". He is Assoc. Prof. Dr in Istanbul Technical University, Faculty of Civil Engineering, Department of Geodesy and Photogrammetry, Division of Cartography, since 2000.