Conceptual Representation of Digital Elevation Models

Rolando Quintero and Virginia Villegas

Laboratorio de Geoprocesamiento- Centro de Investigación en Computación - IPN,
Av. Juan de Dios Bátiz, Esquina con Miguel Othón de Mendizábal,
México, D.F., 07738. México
{quintero, villegas}@cic.ipn.mx

Abstract. In this work, we present a method for the conceptual representation of digital elevation models. This method has four stages, the first stage is focused on the conceptualization of objects and its intrinsic characteristics, the result of this stage is an ontology. The second stage is the mapping from conceptual characteristics to numeric values, for this purpose we will use fuzzy sets. The result of this stage is a set of bands that represent the behavior of the model. The third stage is based on carrying out the classification of the digital elevation model considering the bands obtained in previous stage; we use a semantic classification algorithm, which has been developed for this task. The last stage compiles the results of the classification and puts them in a data representation according to the ontology obtained in the first stage.

1 Introduction

Digital Elevation Models (DEMs) are performing an important role in several fields of Geographic Information Systems (GIS), including the environmental science, risk prevention, engineering and so on. The geometric characteristics of DEMs (resolution, coordinates, rows and columns numbers, etc.) describe the thematic aspects of the terrain, which are represented by qualities. Also, the use of the geometric shape terrain to analyze its distribution and the concentration of certain geospatial objects has been incorporated to DEMs. These representations have been traditionally applied to distinguish the dividing lines of water, drainage and other groups of the terrain objects.

The use of elevation models allow us to review six basic foundations: the relation between the terrain characteristics and the geomorphologic process, analysis of scales, analysis of changes in the surface, analysis of flow or surface movement, visualization approaches and topographic models [3].

DEM processing has been treated by using a numeric approaches [4][13]. In the present work we attempt to make a processing from the semantic point of view. We propose a procedure to represent DEMs in a semantic way. This process is depicted in Fig. 1.

© M. Moreno-Ibarra, S. Levachkine (Eds.)
Semantics in Geoinformatics Applications
Research in Computing Science 25, 2006, pp. 63-71

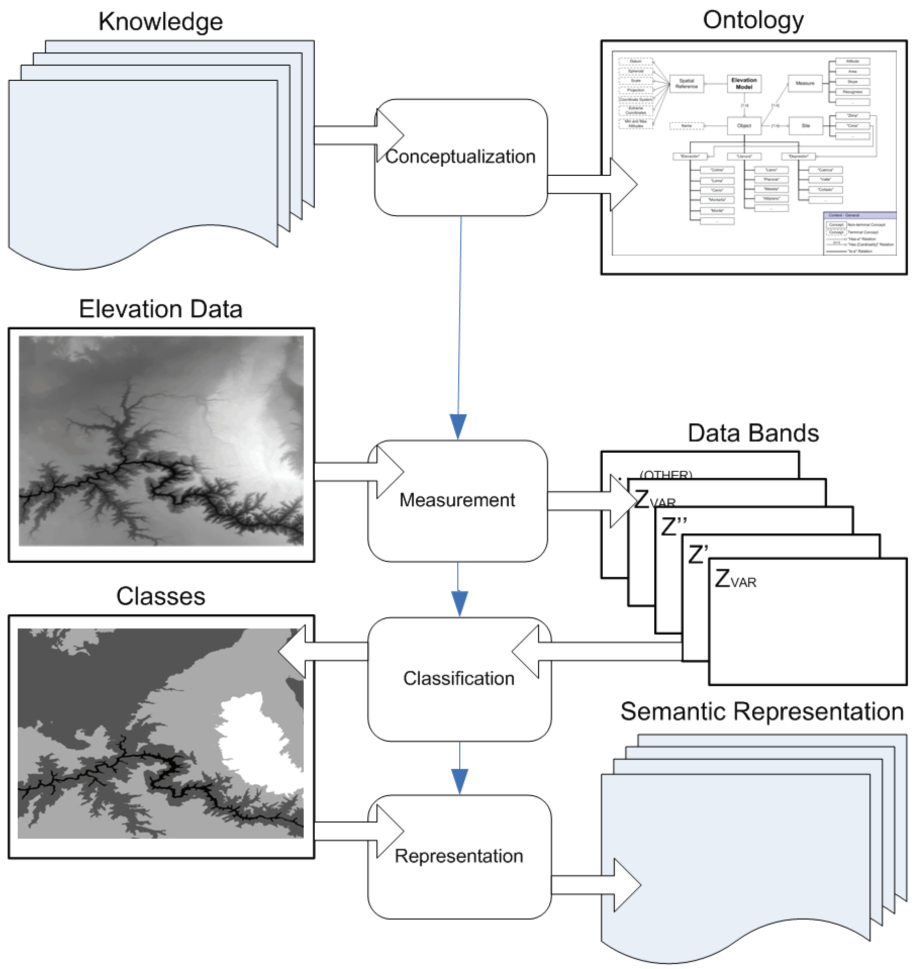

Fig. 1. Procedure proposed for conceptual representation

The spatial semantics is based on the description of the intrinsic properties of geospatial objects. These properties depend on the organization or the *status* of the object. For instance, the width and area of a polygon can provide a description. With this description, it is possible to generate specific standards based on the characteristics of the geospatial representation primitives (lines, points and areas), which define the behavior and relations between geographic objects.

Up-to-date, many analysis involved *quantitative characteristics*, therefore we propose to make a conceptualization focused on the *qualitative properties* of the geospatial objects and their relations in a general analysis. The characteristics that we consider to generate the conceptualization of the model are the follows: slope, surface, extension ruggedness and altitude.

The rest of the paper is organized as follows. Section 2 describes the conceptualization process. The measurement of geographic characteristics is presented in Section

3. In Section 4, we point out the semantic classification algorithm and the conceptual representation. Finally, Section 5 sketches out the conclusions of the research.

2 Conceptualization of spatial data in elevation models

The conceptualization of the characteristics on digital elevation models (elevations, depressions, plains, etc.) is the first stage of the method. It is done by taking the definitions, given in the Royal Spanish Academy [9], about the characteristics mentioned in section 1. Also, we use definitions given by National Institute of Statistics, Geography and Informatics of Mexico (INEGI) [5] related to the same characteristics. The analysis of those definitions provides as result ontology. This ontology represents the concepts that define the elevation model's characteristics, as well as the relations between those concepts. So, in order to represent the conceptualization of geographic characteristics[1], we will use ontology in the way described in [11]. In this work, we proposed to describe the ontology using two types of concepts (terminal and non-terminal ones) and only two types of relations (*"is-a"* for specialize concepts and *"has"* for aggregate concepts).

As it has been said, we define two types of concepts (C) in the ontology: *terminal* (C_T) and *non-terminal* (C_N) *concepts*. The first ones are concepts that do not use other concepts for defining their meaning (they are defined by "simple values"). The meaning of non-terminal concepts is defined by other concepts (terminal or non-terminal) (Eqn. 2).

$$C = C_N \bigcup C_T \tag{2}$$

Each concept has a set of aspects. They are properties and relations between geographic entities[2]. In the following, we shall use the term "relation" to denote unary relations/properties as [2]. From this point of view, all aspects of a terminal concept are simple, e.g. the type of all aspects that belongs to the set of primitive types (T_P), as shown in Eqn. 3.

$$\begin{aligned} &T_P = \{number, character, string, enumeration, struct\}, \\ &A = \{a_i \mid type(a_i) \in T_P\}, \end{aligned} \tag{3}$$

where T_P is the set of primitive types; A is the set of aspects.

Then, the set of *terminal concepts* is defined by Eqn. 4.

$$C_T = \{c(a_1, a_2, ..., a_n) \ni a_i \in A,\ i = 1,..,n\} \tag{4}$$

In the same way, the *non-terminal concepts* have at least one aspect that does not belong to T_P. It is denoted by Eqn. 5.

$$C_N = \{c(a_1, a_2, ..., a_n) \ni \exists a_i \notin A\} \tag{5}$$

---

[1] Geographic objects.

[2] Attributive data

where c is a concept.

Finally, the set of relations R is defined by the pairs that are associated to Γ and Φ, where Γ and Φ are non-reflexive, non-symmetric, and transitive relations (Eqn. 6).

$$R = R_\Gamma \cup R_\Phi = \{(a,b) \mid a\Gamma b,\ a \in C_N,\ b \in C\} \cup \{(a,b) \mid a\Phi b,\ a \in C_N,\ b \in C\} \qquad \mathbf{(6)}$$

Fig. 2 shows a fragment of the ontology used for the elevation model's conceptualization. In the figure, the concepts denoted by *"…"* (Three points) represents 'other' concepts that are presented in the same category of their "brothers" in the ontology. For instance, look at the concept *"depresión"* and its children concepts (*"cuenca"*, *"valle"* and *"collado"*), it is denoted by *"…"* that other *"depresión"* children can exist (*"cañada"*, *"cañón", "barranca"*, etc.). In the ontology we must define all required concepts for describing geospatial data in elevation models, according to INEGI [5].

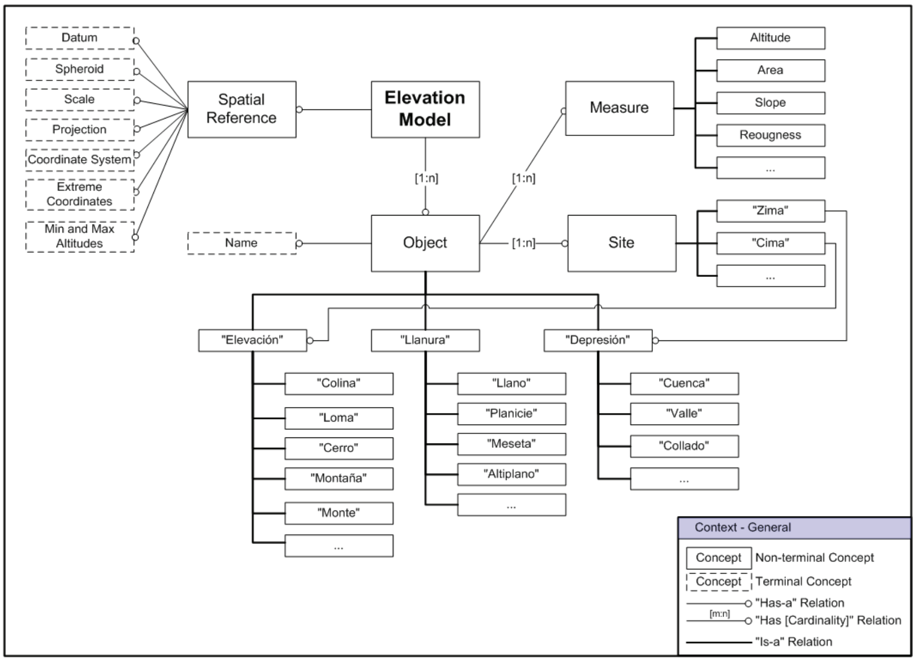

Fig. 2. Fragment of the ontology used for the conceptualization of geospatial data in elevation models. The ontology consists of a set of concepts and a set of relations. There are two types of concepts (terminal and non-terminal) as well as two types of relations (*has* and *is-a*).

3 Measurement of geographic characteristics

Once we have made the conceptualization, the second stage consists of assigning metrics to the concepts in the ontology. These metrics are ranges or procedures to obtain numeric values from the elevation data. For instance, the concept *"cima"* (highest point of mountains, mount and hills) can be found numerically from the elevation model. Other type of values determine the difference between concepts, as in the case of the concepts *"colina"* (natural land elevation, lower than a mountain) and *"monte"* (natural land elevation, higher than a hill), In this example, the boundary between the concepts is not clear; we only know that a *mount* is 'higher' than a *hill*. In cases like the last one, we will use the INEGI definitions in order to have a hint about the boundary. Anyway, we will consider that this boundary is always diffuse.

On the other hand, an elevation model is fundamentally a discrete function Z of world coordinates, represented by means a matrix containing the elevation data (Eqn. 7).

$$Z(m), \tag{7}$$

Also, some other metrics can be determined using the Z function as a base, for instance, the slope (Eqn. 8), curvature (Eqn. 9), the variance of altitude (Eqn. 10), among others.

$$Z' = \sqrt{\left(\frac{\Delta Z}{\Delta x}\right)^2 + \left(\frac{\Delta Z}{\Delta y}\right)^2}, \tag{8}$$

$$Z_x'' = \frac{\delta^2 Z}{\delta x^2}, Z_x'' = \frac{\delta^2 Z}{\delta y^2}, \tag{9}$$

$$Z_{VAR} = \frac{1}{MN}\sum_{i=1}^{M}\left(\sum_{j=1}^{N}\left(Z(i,j) - \overline{Z}\right)^2\right), \tag{10}$$

With these metrics we obtain 'pages' or 'data bands' that will be used in the classification process.

By using the characteristics modeled in the conceptualization stage, we can obtain the classification of vectors ω, that will be used later in the process. It is necessary to recall that the conceptualization stage provides a set of concepts (within ontology) with diffuse properties. For example, as we previously mentioned, in the case of the concepts *"colina"* and *"monte"*, the boundary between the concepts is not clear; we only know that a mount is 'higher' than a hill. In this case, we can distinguish with the statements 'A *"colina"* (hill) has low altitude and small area' and 'A *"monte"* (mount) has medium altitude and large area'. This carries us to define some fuzzy sets in order to map numeric values to conceptual ones (see Fig. 3).

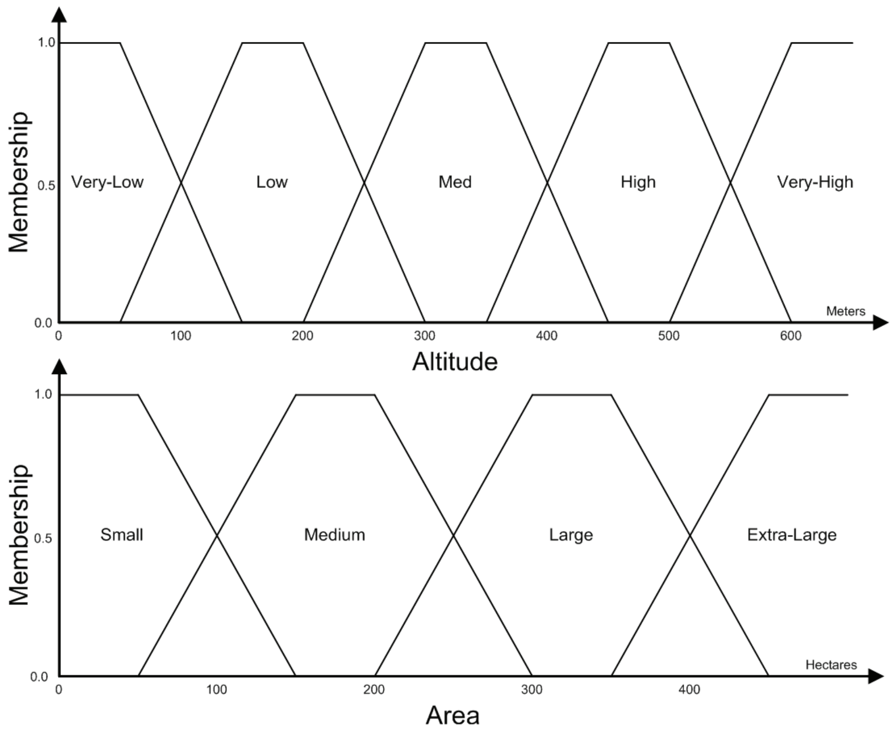

Fig. 3. Fuzzy sets obtained from the conceptualization of altitude and area of a geographic object.

4 Classification and representation of geometric characteristics

The third stage is the classification of the digital elevation model, according to the metrics that have been obtained of the conceptualization process[3]. The metrics that can be made on the pixels are the follows: altitude, slope, planar curvature and ruggedness [3]. Although in [7] is presented other measurements that can be used to this purpose. With these measurements to each pixel is obtained a set of "pages", which describes the behavior of the DEM according to each metric. For the classification, it is not necessary to have training stages nor computing the characteristic vectors, since the conceptualization provides the classes for the classification and the characteristic vectors. With this information, it is possible to apply some classification algorithm [1][6][8]. In [10] is sketched out an algorithm for semantic supervised segmentation to classify multispectral geo-images.

[3] It is important to mention, there are not so much metrics that can be made on DEMs. It can be a constraint for this procedure. However, future works are oriented to incorporate information of different sources associated to DEMs such as satellite images.

The result of the conceptualization stage is the set of semantic characteristic vectors Ω of the classification. From the Ω set, it is necessary to compute the mean and covariance matrix for each $\omega_i \in \Omega$. By attempting to determine the class or category for each pixel at a location x, it is necessary that each pixel contains a conditional probability, denoted by Eqn. 11:

$$p(\omega_i \mid x), i = 1, \ldots, M \qquad \textbf{(11)}$$

In [12], we describe the supervised clustering method for more details. Therefore, Bayes theorem provides potential means of converting knowledge of predictive correlations. The constraint (Eqn. 12) is used in the classification algorithm, since the $p(x)$ are known by training data, we assume that it is conceivable that the $p(\omega_i)$ is the same for each ω_i, due to this, the comparison is $p(x \mid \omega_i) > p(x \mid \omega_j)$.

$$x \in \omega_i \text{ if } p(x \mid \omega_i)p(\omega_i) > p(x \mid \omega_i)p(\omega_j) \text{ for all } j \neq i \qquad \textbf{(12)}$$

In this analysis, we assume that the classes have multidimensional normal distributions and each pixel can assign a probability of being a member of each class. After computing the probabilities of a pixel being in each of the available classes, we assign the class with the highest probability. The algorithm consists of the following steps:

[Step 1]. Given the number of classes ω_i by means of the conceptualization process, compute the maximum likelihood distribution and the covariance matrix for each class ω_i.

[Step 2]. For each image pixel, determine its semantic vector.

[Step 3]. Compute the probability of vector s to know if it belongs to each class ω_i.

[Step 4]. Obtain the coordinates (x, y) of the pixel, if the constraint (see Eqn. 13) is accomplished, then the pixel belongs to the class ω_i.

$$p(x \mid \omega_i)p(\omega_i) > p(x \mid \omega_j)p(\omega_j) \text{ for all } j \neq i \qquad \textbf{(13)}$$

[Step 5]. Repeat from the step 2 until all pixels in the elevation model can be classified.

Generally, the classification result emits *noise*, since this process is made at pixel level. As an additional step for the classification stage, it is necessary to determine "consolidated regions". This can be performed in several ways, the best option but more complex is the "semantic consolidation". However, it is necessary to conceptualize the problem to define rules by means of case study for its consolidation. For instance, the rule: *"if a decline is surrounded by a mountain, then the decline must be absorbed by the mountain"*. In this rule there are features already conceptualized, such as the *decline* and the *mountain*, but there are not yet other features conceptualized, i.e., *surrounded* and *absorbed*. An alternative, it is to use algorithms oriented to growth of regions, in which small regions are merged by big regions, until a boundary is defined among these regions.

The last stage consists of compiling the results about the classification to put in a conceptual way into a description[4]. Future works are focused on linking a toponym database to label regions according to their names. At this moment, the description is restricted to specify the existence of geographic characteristics and some important attributes such as surface, altitude mean, borders, etc.

In this research, we have not considered the conceptualization of topological relations that are presented in the geographic characteristics. However, we have envisaged incorporating these relations in a short time, because the topological relations are indispensable for consolidating semantics of regions, as well as for making a more descriptive representation.

By using this method, we attempt to approximate the semantics that contain DEMS, which will be useful to evaluate the semantic similitude between different models. According to this point, the evaluation uses concepts that are organized in a hierarchy[5]. With this focus, we attempt to use the qualitative measurements instead of quantitative measurements in order to evaluate the similitude and measure the distance between concepts stored in the conceptual representation.

5 Conclusions

In this work, we have presented a method to make a semantic representation of elevation models. The method has four stages: conceptualization, measurement, classification and representation. In the first stage, we propose the conceptualization of the objects that can be found within an elevation model, as well as the characteristics that define these objects. The result of this stage is the ontology that conceptualizes the domain. The second stage consists of mapping the conceptual characteristics to numeric values. We propose the use of fuzzy sets in order to make this mapping. The result of this stage is a set of data bands that represents the behavior of elevation model according to a specific characteristic. In the third stage, it is carried out the classification of the elevation model. In the classification process, the data bands obtained in the second stage are used as input for the semantic classification algorithm. The last stage takes the classification results, and puts them into a semantic representation according to the ontology obtained in first stage.

This work is in progress, so we have a lot of tasks to do. In a general way, it is necessary to grow the ontology with the conceptualization of topologic relations between geographic objects, as well as the metrics conceptualization. Also, we must set up the real values of fuzzy sets, introduced in Section 3, in order to make the values available to the classification process. Finally, we must implement the method[6], as well as the necessary tests.

---

[4] List of objects (and its attributes) found in model

[5] The term of hierarchy is defined as a data structure that stores concepts that are related by one relationship, in which the partitions are completed to represent the knowledge about a certain context.

[6] We have already implemented the semantic classification algorithm.

Acknowledgments

The authors of this paper wish to thank the CIC, SIP, IPN and CONACYT for their support.

References

1. Bandyopadhyay, S.: Satellite image classification using genetically guided fuzzy clustering with spatial information. *International Journal of Remote Sensing*, Taylor & Francis, Vol. 26, No. 3, (2005), pp. 579-593.
2. Berendt, B., Barkowsky, T., Freksa, C. and Kelter, S. I. E.: Spatial Cognition - An Interdisciplinary Approach to Representing and Processing Spatial Knowledge. *Spatial Representation with Aspect Maps*, Springer-Verlag, (1998) 157-175.
3. Etzelmüller, B and J.R. Sulebak: Developments in the use of digital elevation models in periglacial geomorphology and glaciology. *Physische Geographie*, Vol. 41, Jahrestagung SGmG Winterthur, (2000).
4. Evans, I.S.: General geomorphometry, in Goudie, A.S. (ed.), *Geomorphological Techniques*, (1981), pp.31-37.
5. Intituto Nacional de Estadística, Geografía e Informática de México (http://www.inegi.gob.mx).
6. Jianwen, Ma., Bagan H.: Land-use classification using RASTER data and self-organized neutral net-works. *International Journal of Applied Earth Observation and Geoinformation*, Elsevier, Vol. 7, No. 3, (2005), pp. 183-188.
7. Moreno M., Levachkine S., Torres M. and Quintero R.: Landform Classification in Raster Geo-Images, in A. Sanfeliu, J.F. Martínez-Trinidad, J.A. Carrasco-Ochoa (Eds.), *Lecture Notes in Computer Science*, Vol. 3287, Springer-Verlag, Berlin Heidelberg, (2004), pp. 558-565.
8. Nishii, R., Eguchi, S.: Supervised image classification by contextual AdaBoost based on posteriors in neighborhoods. *IEEE Transactions on GeoScience and Remote Sensing*, Vol. 43, No. 11, (2005), pp. 2547-2554.
9. Real Academia Española. 22ª edición. Madrid: Espasa Calpe, (2001).
10. Torres M., Guzman G., Quintero R., Moreno M. and Levachkine S.: Semantic Supervised Clustering to Land Classification in Geo-Images, in R. Khosla, R. J. Howlett, L. C. Jain (Eds), *Lecture Notes in Artificial Intelligence*, Vol. 3683, Springer-Verlag, Berlin Heidelberg, (2005), pp. 248-254
11. Torres M., Moreno M., Quintero R. and Fonseca F.: Ontology-driven description of spatial data for their semantic processing, in M.A. Rodríguez, I.F. Cruz, S. Levashkin and M. Egenhofer (Eds), *Lecture Notes in Computer Science*, Vol. 3799, Springer-Verlag, Berlin Heidelberg, (2005), pp. 242-249
12. Torres M., Moreno M., Quintero R., Guzmán G.: Applying Supervised Clustering to Land-sat MSS Images into GIS-Application, *Advances in: Artificial Intelligence, Computing Science and Computer Engineering, Research on Computing Science*, Vol. 10, (2004), pp. 167-176.
13. Wood, J.: The Geomorphological Characterisation of Digital Elevation Models, Ph. D. Thesis, Department of Geography, University of Leicester, UK, (1997).