12-1-2004

# The Multi-Layered Interval Categorizer Tesselation-Based Model

Marilton S. de Aguiar

Gracaliz P. Dimuro

Antonio C. da Rocha Costa

Rafael K.S. Silva

Fabia A. da Costa

***See next page for additional authors***



Comments:
UTEP-CS-04-38.
Published in Cirano Iochpe and Gilberto Camara (eds.), IFIP WG2.6 *Proceedings of the 6th Brazilian Symposium on Geoinformatics Geoinfo'2004*, Campos do Jordao, Brazil, November 22-24, 2004, pp. 437-454. ISBN 3901882200

**Authors**
Marilton S. de Aguiar, Gracaliz P. Dimuro, Antonio C. da Rocha Costa, Rafael K.S. Silva, Fabia A. da Costa, and Vladik Kreinovich

# THE MULTI-LAYERED INTERVAL CATEGORIZER TESSELLATION-BASED MODEL*

Marilton S. de Aguiar, Gracaliz P. Dimuro, Antônio C. da R. Costa, Rafael K. S. Silva and Fábia A. da Costa
*Escola de Informática, Universidade Católica de Pelotas*
*Pelotas/RS, CEP 96010-000, Brazil.*
{marilton,liz,rocha,rafaelk,fabia}@ucpel.tche.br

Vladik Kreinovich
*Department of Computer Science, UTEP*
*El Paso, TX 79968, USA.*
vladik@cs.utep.edu

**Abstract** The paper presents the results obtained by an implementation of the interval tessellation-based model for categorization of geographic regions according to the analysis of the relief function declivity, called *ICTM* (Interval Categorizer Tessellation-based Model). The analysis of the relief declivity, which is embedded in the rules of the model, categorizes each tessellation cell, with respect to the whole considered region, according to the (positive, negative, null) sign of the declivity of the cell. Such information is represented in the states assumed by the cells of the model. The overall configuration of such cells allows the division of the region into sub-regions of cells belonging to the same category, that is, presenting the same declivity sign. In order to control the errors coming from the discretization of the region into tessellation cells, or resulting from numerical computations, interval techniques are used.

**Keywords:** Cellular Automatons, Interval, Geophysics.

*This work was supported by the Brazilian funding agencies CNPq/CTPETRO/CTINFO and FAPERGS. V.K. was also partly supported by NASA under cooperative agreement NCC5-209, by Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-00-1-0365, by NSF grants EAR-0112968 and EAR-0225670, by the Army Research Laboratories grant DATM-05-02-C-0046, and by IEEE/ACM SC2003 Minority Serving Institutions Participation Grant.

## 1. Introduction

In (Aguiar et al., 2004a), it is presented a general tessellation-based model for categorizer tools that are able to subdivide a certain geographic region into sub-regions presenting similar characteristics, that is, belonging to the same range concerning a set of given observable properties.

In that paper, we also presented the so-called *2d-lc-ICTM*, which is a bi-dimesional multi-layered *ICTM* to analyze the variation of declivity of the function that maps a property of a given region, subdividing this region into sub-regions presenting the same behavior with respect to the declivity.

The number of the characteristics that should be studied determines the number of layers of the model. In each layer, a probably different analysis of the region is obtained. An appropriate projection of all layers to the basic layer of the model leads to a meaningful subdivision of the region and to a categorization of the sub-regions that consider the simultaneous occurrence of all characteristics, according to some weights. To control the errors coming from discretization and resulting from the numerical computations, interval techniques (Moore, 1979) are used to obtain a reliable categorization.

The tessellation-based model performs a bidimensional analysis of the declivity, using local rules for creation and categorization of sub-regions, giving the relative situation of each sub-region with respect the whole area, according to the states assumed by the cells. This work evolved directly from the analysis of the work (Coblentz et al., 2003).

The *ICTM* Model uses a structured mesh to constitute its tessellation. A structured bidimensional mesh is often simply a square grid deformed by some coordinate transformation. Each vertex of the mesh, except at the boundaries, has an isomorphic local neighborhood. In three dimensions, a structured mesh is usually a deformed cubical grid. Structured meshes are simpler than the non-structured ones, and require less computer memory, as their coordinates can be calculated, rather than explicitly stored. Structured meshes offer more direct control over the sizes and shapes of elements.

An immediate application is in Geophysics, where an adequate subdivision of geographic areas into segments presenting similar topographic characteristics is often convenient. See (Forman, 1995), for other applications related to the analysis of the relief.

The data input for the model are extracted from satellite images of the geographic region being analyzed, where the heights are given in certain points referenced by their latitude and longitude coordinates. This geo-

graphic region is represented by a regular tessellation that is determined by subdividing the total area into sufficiently small rectangular subareas, each one represented by one cell of the tessellation. This subdivision is done according to a cell size established by the geophysics analyst and it is directly associated to the refinement degree of the tessellation.

The categorization determined by each characteristic is performed in one layer of the model, generating different subdivisions of the analyzed region. For instance, a region can be analyzed according to its topography, vegetation, demography, economic data, etc.

A global categorization can be reached from the categorization of each layer through a projection procedure. This global categorization will determine a more reliable and significant subdivision combining the performed analysis in each characteristic.

There are many methods for image segmentation (Cooper, 1998; Fu and Mui, 1981; Lisani et al., 2003; Umbaugh, 1998) and the most commonly used techniques can be classified into two categories: (i) region extraction techniques, which look for maximal regions satisfying some homogeneity criterion, and (ii) edge extraction techniques, which look for edges occurring between regions with different characteristics. The main problem with most of these methods is that they are heuristic and frequently different methods give different results, and, therefore, it is desirable to produce reliable methods (see, e.g, (Aguiar et al., 2004b; Coblentz et al., 2003; Villaverde and Kreinovich, 1993)).

## 2. Clustering Techniques

Clustering is the unsupervised classification of patterns (observations, data items or feature vectors) into groups (clusters). The clustering problem has been addressed in many contexts and by researchers in many disciplines; this reflects its usefulness as one the steps in exploratory data analysis.

Data analysis underlies many computing applications, either in a design phase or as part of their on-line operations. Data analysis procedures can be classified as either exploratory or confirmatory, based on the availability of appropriate models for the data source. In (Berkhin, 2002), a key element in both types of procedures (whether for hypothesis formation or decision-making) is the grouping, or classification of measurements based on either (i) assumed model, or (ii) natural grouping (clustering) revealed through analysis.

Cluster analysis is the organization of a collection of patterns usually represented as a vector of measurements or a point in a multidimensional space into clusters based on similarity (Jain et al., 1999). Intuitively,

patterns within a valid cluster are more similar to each other than they are to a pattern belonging to a different cluster. An example of clustering is depicted in Figure 1. The input patterns are shown in Figure 1a, and the desired clusters are shown in Figure 1b.

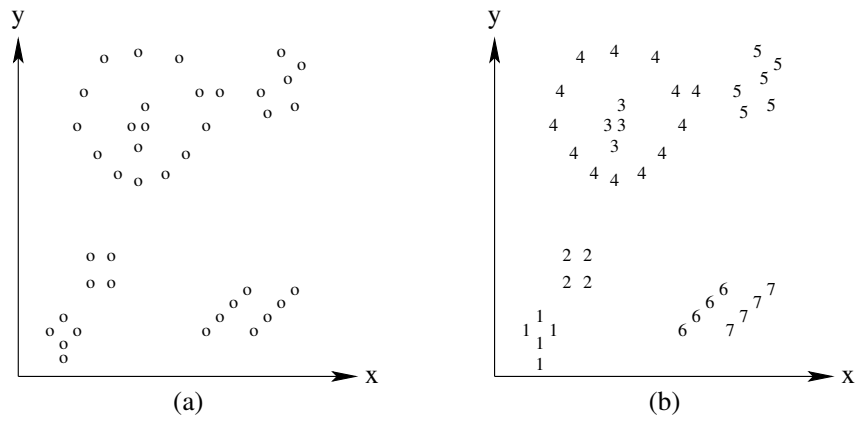


*Figure 1.* An Example of Data clustering

The points belonging to the same cluster are given the same label. There are a variety of techniques for representing data, measuring proximity (similarity) between data elements and grouping data elements. Clustering is useful in several exploratory pattern-analysis, grouping, decision-making, and machine-learning situations, including data minig, document retrieval, image segmentation, and pattern classification (Berkhin, 2002). However, in many such problems, there is little prior information (e.g., statistical models) available about the data, and the decision-maker must make as few assumptions about the data as possible. It is under these restrictions that clustering methodology is particularly appropriate for the exploration of interrelationships among the data points to make a preliminary assessment of their structure.

Typical pattern clustering activity involves the following steps (Jain and Dubes, 1988): (i) pattern representation; (ii) definition of a pattern proximity measure appropriate to the data domain; (iii) clustering or grouping; (iv) data abstraction, and (v) assessment of output.

Pattern representation refers to the number of classes, the number of available patterns, and the number, type and scale of the features available to the clustering algorithm. Some of this information may not be controllable by the specialist. Pattern proximity is usually measured by a distance function defined on pairs of patterns. For example,

the Euclidean distance can be used to reflect dissimilarity between two patterns (Diday and Simon, 1976).

The grouping step can be performed in a number of ways, the output clustering (or clusterings) can be (i) hard – the partition of the data into groups, or (ii) fuzzy – where each patterns has a variable degree of membership in each of the output clusters. Hierarchical clustering algorithms produce a nested series of partitions based on a similarity criterion for merging or splitting clusters.

Data abstraction is the process of extracting a simple and compact representation of a data set. Often, simplicity is either from the perspective of automatic analysis (so that a machine can perform further processing efficiently) or it is human-oriented (so that the representation obtained is easy to comprehend and intuitively appealing). In the clustering context, a typical data abstraction is a compact description of each cluster, usually in terms of cluster prototypes or representative patterns such as the centroid (Diday and Simon, 1976).

The study of clustering tendency, wherein the input data are examined to see if there is any merit to a cluster analysis prior to one being performed, is a relatively inactive research area (Jain et al., 1999). Cluster validity analysis, in contrast, is the assessment of a clustering procedure's output. Often this analysis uses a specific criterion of optimality. This type of evaluation is actually an assessment of the data domain rather than the clustering algorithm itself.

When statistical approaches to clustering are used, validation is accomplished by carefully applying statistical methods and testing hypotheses. There are three type of validation studies (Dubes, 1987; Dubes, 1993): (i) an external assessment of validity compares the recovered structure to an *a priori* structure; (ii) an internal examination of validity tries to determine if the structure is intrinsically appropriate for the data; and, (iii) a relative test compares two structures and measures their relative merit.

## 3. The *2d-lc-ICTM* Model

This section introduces the multi-layered interval categorizer tessellation-based model, formalized in terms of matrix operations. The single-layered *ICTM* was firstly presented in (Aguiar et al., 2004a). Here, we present the generalization of the number of the layers and the corresponding projection procedures.

This type of projection allows interesting analysis of the mutual dependency of the analysed characteristics. Each characteristic of the space is represented in a layer of the *ICTM* Model. Thus, by the in-

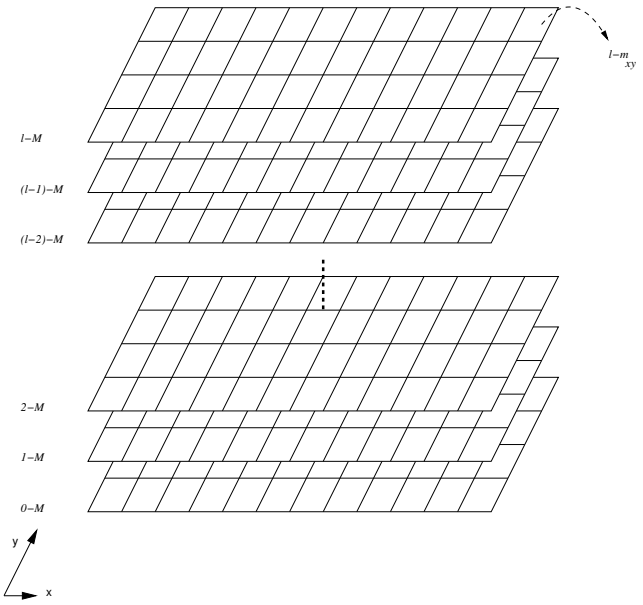


*Figure 2.* The *ICTM* multi-layered

dependency of the analysis, the subdivisions in each layer also occurs in a independently way. In the Figure 2, each bidimensional layer is represented by a label.

DEFINITION 1 *A tessellation is a matrix $M$ with $n_r$ rows and $n_c$ columns. The entry at the $x$-th row and the $y$-th column is called the $xy$-cell of $M$.*

DEFINITION 2 *Considering a $n_c \times n_r$ tessellation $M$ and $l \in \mathbb{N}$, a multi-layered tessellation $\mathcal{L}$-$\mathcal{M}$ is the structure*

$$\mathcal{L}\text{-}\mathcal{M} = (1\text{-}M, \ldots, l\text{-}M)$$

*where the entry at the $l$-th layer, $x$-th row and $y$-th column is denoted by $l$-$m_{xy}$.*

## 3.1 The Interval Matrices

In topographic analysis, usually there are too many data, most of which is geophysically irrelevant. We then take, for each subdivision, the average value of the heights at the points supplied by the satellite photos, which are the entries of the tessellation $M$:

DEFINITION 3 *A layer $l$ of this tessellation $M$ is the $n_r \times n_c$ matrix $l\text{-}M^{abs} = [l\text{-}m_{xy}^{abs}]$, where the entry $l\text{-}m_{xy}^{abs}$ is the absolute value of the average height of the points represented by the $xy$-cell in the layer $l$ of $M$.*

We are interested in comparing the values corresponding to different cells, so we are not interested in absolute values, only in relative ones. To simplify the data of the matrix, we normalize them by dividing each $l\text{-}m_{xy}^{abs}$ by the largest $l\text{-}m_{max}$ of these values.

DEFINITION 4 *The relative matrix of layer $l$ $l\text{-}M^{rel}$ is defined as the $n_r \times n_c$ matrix given by*

$$l\text{-}M^{rel} = \frac{l\text{-}M^{abs}}{l\text{-}m_{max}}.$$

The heights are measured pretty accurately, so the only errors in the values $l\text{-}m_{xy}$ come from the discretization of the area in terms of the discrete set of tessellation cells. In other words, it is desirable to know the values of the relief function $h_{\xi\upsilon}$ for all $\xi$ and $\upsilon$, but only the values $h_{xy} \equiv l\text{-}m_{xy}^{rel} = \frac{l\text{-}m_{xy}^{abs}}{l\text{-}m_{max}}$ for $11, \ldots, 1n_r, \ldots, n_c1, \ldots, n_cn_r$, determined by division of the region in $n_rn_c$ cells, are used in the effective calculations.

In the following, we apply Interval Mathematics techniques to control the errors associated to the cell values. To see examples of the advantages of using intervals in solving similar problems see, e.g., (Coblentz et al., 2003; Kearfort and Kreinovich, 1996). For each $\xi\upsilon$, which is different from $xy$, it is reasonable to estimate $h_{\xi\upsilon}$ as the value $l\text{-}m_{xy}^{rel.}$ at the point $xy$ which is closest to $\xi\upsilon$, meaning that $\xi\upsilon$ belongs to the same segment of area as $xy$. For each cell $xy$, let $\Delta_x$ and $\Delta_y$ be the largest possible errors of the corresponding approximations considering the west-east direction and the north-south direction, respectively.

For fixed $y$, when $\xi > x$, the point $xy$ is still the closest until we reach the midpoint $x_{mid}y = \frac{(x\ +\ (x+1))}{2}y$ between $xy$ and $(x+1)y$. It is reasonable to assume that the largest possible approximation error $|l\text{-}m_{xy}^{rel} - h_{\xi y}|$ for such points is attained when the distance between $xy$ and $\xi y$ is the largest, i.e., when $\xi y = x_{mid}y$. In this case, the approximation error is equal to $|h_{x_{mid}y} - l\text{-}m_{xy}^{rel}|$.

LEMMA 1 *For fixed $y$, if $\xi > x$, then the approximation error $\epsilon$ is bounded by $0.50 \cdot |l\text{-}m_{(x+1)y}^{rel} - l\text{-}m_{xy}^{rel}|$.*

**Proof**. If the points $xy$ and $(x+1)y$ belong to the same segment of area, then the dependence of $n_{\xi y}$ on $\xi y$ should be reasonably smooth for

$\xi \in [x,(x+1)]$. On a narrow interval $[x,(x+1)]$, we can, with reasonable accuracy, ignore the quadratic and higher terms in the expansion of $h_{(\xi+\Delta\xi)y}$ and approximate $h_{\xi y}$ by a linear function. For a linear function $\xi \mapsto h_{\xi y}$, the difference $h_{x_{mid}y} - l\text{-}m^{rel}_{xy}$ is equal to the half of the difference $l\text{-}m^{rel}_{(x+1)y} - l\text{-}m^{rel}_{xy}$. On the other hand, if the points $xy$ and $(x+1)y$ belong to different segments, then the dependence $h_{\xi y}$ should exhibit some non-smoothness, and it is reasonable to expect that the difference $l\text{-}m^{rel}_{(x+1)y} - l\text{-}m^{rel}_{xy}$ is much higher than the approximation error. In both cases, the approximation error $\epsilon$ is bounded by $0.50 \cdot |l\text{-}m^{rel}_{(x+1)y} - l\text{-}m^{rel}_{xy}|$.□

LEMMA 2 *For fixed $y$, if $\xi < x$, then the approximation error $\epsilon$ is bounded by* $0.50 \cdot |l\text{-}m^{rel}_{xy} - l\text{-}m^{rel}_{(x-1)y}|$.

PROPOSITION 1 *For the approximation error $\epsilon_x$,*

$$\epsilon_x \leq \Delta_x = 0.5 \cdot \min\left(|l\text{-}m^{rel}_{xy} - l\text{-}m^{rel}_{(x-1)y}|, |l\text{-}m^{rel}_{(x+1)y} - l\text{-}m^{rel}_{xy}|\right).$$

**Proof.** It follows from Lemmas 1 and 2.□

As a result, considering a given $y$, besides of the central values $l\text{-}m^{rel}_{xy}$, for each $x$, we get intervals $m^{x[\,]}_{xy}$ containing all the possible values of $h_{\xi y}$, for $x - \frac{1}{2} \leq \xi \leq x + \frac{1}{2}$.

COROLLARY 1 *Considering a fixed $y$, for each $x$, if $x - \frac{1}{2} \leq \xi \leq x + \frac{1}{2}$, then $h_{\xi y} \in l\text{-}m^{x[\,]}_{xy} = \left[l\text{-}m^{x^-}_{xy}, l\text{-}m^{x^+}_{xy}\right]$, where $l\text{-}m^{x^-}_{xy} = l\text{-}m^{rel}_{xy} - \Delta_x$ and $l\text{-}m^{x^+}_{xy} = l\text{-}m^{rel}_{xy} + \Delta_x$.*

Using an analogous argumentation, for a fixed $x$, it follows that:

PROPOSITION 2 *For the approximation error $\epsilon_y$,*

$$\epsilon_y \leq \Delta_y = 0.5 \cdot \min\left(|l\text{-}m^{rel}_{xy} - l\text{-}m^{rel}_{x(y-1)}|, |l\text{-}m^{rel}_{x(y+1)} - l\text{-}m^{rel}_{xy}|\right).$$

COROLLARY 2 *Considering a fixed $x$, for each $y$, if $y - \frac{1}{2} \leq \upsilon \leq y + \frac{1}{2}$, $h_{x\upsilon} \in l\text{-}m^{y[\,]}_{xy} = \left[l\text{-}m^{y^-}_{xy}, l\text{-}m^{y^+}_{xw}\right]$, where $l\text{-}m^{y^-}_{xy} = l\text{-}m^{rel}_{xy} - \Delta_y$, $l\text{-}m^{y^+}_{xy} = l\text{-}m^{rel}_{xy} + \Delta_y$.*

DEFINITION 5 *If $l\text{-}m^{x^\pm}_{xy} = l\text{-}m^{rel}_{xy} \pm \Delta_i$ and $l\text{-}m^{y^\pm}_{xy} = l\text{-}m^{rel}_{xy} \pm \Delta_j$, the interval matrices $l\text{-}M^{x[\,]}$ and $l\text{-}M^{y[\,]}$, associated with the relative matrix $l\text{-}M^{rel}$, are defined by the $n_r \times n_c$ interval matrices*

$$l\text{-}M^{x[\,]} = \left[l\text{-}m^{x[\,]}_{xy}\right] = \left[\left[l\text{-}m^{x^-}_{xy}, l\text{-}m^{x^+}_{xy}\right]\right] \text{ and}$$
$$l\text{-}M^{y[\,]} = \left[l\text{-}m^{y[\,]}_{xy}\right] = \left[\left[l\text{-}m^{y^-}_{xy}, l\text{-}m^{y^+}_{xy}\right]\right].$$

## 3.2 The Declivity Registers and the State Matrix

We proceed to a declivity categorization inspired by (Coblentz et al., 2003). We assume from the start that the relief approximation functions introduced by the tessellation-based model are piecewise linear functions. We cast the whole process as a kind of constraint satisfaction problem, where the tessellation-based model is in charge of finding a piecewise linear relief approximation function (and corresponding set of limit points between the resulting sub-regions) that fits the constraints imposed by the interval matrix. To narrow the solution space to a minimum, we take a qualitative approach to the relief approximation functions, clustering them in equivalence classes according to the sign of their declivity (positive, negative, null), thus making the tessellation-based model build a single qualitative solution to that constraint satisfaction problem, namely, the class of approximation functions compatible with the constraints of the interval matrix. We proceed as follows:

PROPOSITION 3 *Let $l\text{-}M^{x[\,]}$ and $l\text{-}M^{y[\,]}$ be interval matrices of layer $l$. For a given $xy$, if:*

*(i)* $l\text{-}m^{x^+}_{xy} \geq l\text{-}m^{x^-}_{(x+1)y}$, *then there exists a non-increasing relief approximation function between $xy$ and $(x+1)y$ (direction west-east).*

*(ii)* $l\text{-}m^{x^-}_{(x-1)y} \leq l\text{-}m^{x^+}_{xy}$, *then there exist a non-decreasing relief approximation function between $(x-1)y$ and $xy$ (direction west-east).*

*(iii)* $l\text{-}m^{y^+}_{xy} \geq l\text{-}m^{y^-}_{x(y+1)}$, *then there exists a non-increasing relief approximation function between $xy$ and $x(y+1)$ (direction north-south).*

*(iv)* $l\text{-}m^{y^-}_{x(y-1)} \leq m^{y^+}_{xy}$, *then there exists a non-decreasing relief approximation function between $x(y-1)$ and $xy$ (direction north-south).*

**Proof.** A sketch of the proof is given. In *(i)*, take, for example, $\mu_{xy} = l\text{-}m^{x^+}_{xy}$, $\mu_{(x+1)y} = l\text{-}m^{x^-}_{(x+1)y}$ and use a linear interpolation to define the values $\mu_{ky}$ for $x < k < x+1$. The proofs of *(ii)-(iv)* are similar.□

For each cell, four directed declivity registers[1] – $reg.e$ (east), $reg.w$ (west), $reg.s$ (south) and $reg.n$ (north) – are defined, indicating the admissible declivity sign of the function that approximates the relief function in any of these directions, taking into account the values of the neighbor cells. The analysis of declivity is done according of Prop. 3.

DEFINITION 6 *A declivity register of an $xy$-cell is a tuple*

$$reg = (reg.e, reg.w, reg.s, reg.n),$$

*where the values of the directed declivity registers are given by:*

*(a) For non border cells, considering the conditions given by Prop. 3: $reg.e = 0$, if (i) holds; $reg.w = 0$, if (ii) holds; $reg.s = 0$, if (iii) holds; $reg.n = 0$, if (iv) holds; $reg.e, reg.w, reg.s, reg.n = 1$, otherwise.*

*(b) For east, west, south and north border cells: $reg.e = 0$, $reg.w = 0$, $reg.s = 0$ and $reg.n = 0$, respectively[2]. The other directed declivity registers of border cells are also determined according to item (a).*

DEFINITION 7 *The declivity register matrix of the layer $l$ is defined as an $n_r \times n_c$ matrix $l\text{-}M^{reg} = [l\text{-}m_{xy}^{reg}]$, where the entry at the $x$-th row and the $y$-th column is the value of the declivity register of the corresponding cell.*

COROLLARY 3 *Considering the west-east direction, any relief approximation function $l\text{-}m_{xy}$ is either (i) strictly increasing between $xy$ and $(x+1)y$ if $l\text{-}m_{xy}^{reg.e} = 1$ (in this case, $l\text{-}m_{(x+1)y}^{reg.w} = 0$); or (ii) strictly decreasing between $xy$ and $(x+1)y$ if $l\text{-}m_{(x+1)y}^{reg.w} = 1$ (in this case, $l\text{-}m_{xy}^{reg.e} = 0$); or (iii) constant between $xy$ and $(x+1)y$ if $l\text{-}m_{xy}^{reg.e} = 0$ and $l\text{-}m_{(x+1)y}^{reg.w} = 0$. Similar results hold for the north-south direction.*

DEFINITION 8 *Let $w_{reg.e} = 1$, $w_{reg.s} = 2$, $w_{reg.w} = 4$ and $w_{reg.n} = 8$ be weights to be associated to the directed declivity registers. The state matrix is defined as an $n_r \times n_c$ matrix given by $l\text{-}M^{state} = \left[l\text{-}m_{xy}^{state}\right]$, where the entry at the $x$-th row and the $y$-th column is the value of the corresponding cell state, calculated as the value of the binary encoding of the corresponding directed declivity registers, given as*

$$\begin{aligned} l\text{-}m_{xy}^{state} &= w_{reg.e} \times l\text{-}m_{xy}^{reg.e} + w_{reg.s} \times l\text{-}m_{xy}^{reg.s} \\ &+ w_{reg.w} \times l\text{-}m_{xy}^{reg.w} + w_{reg.n} \times l\text{-}m_{xy}^{reg.n}. \end{aligned}$$

Thus, for given $xy$, the correspondent cell can assume one and only one state represented by the value $l\text{-}m_{xy}^{state} = 0..15$.

## 3.3 The Limit Matrix and the Constant-Declivity Sub-Regions

A limit cell is defined as the one where the relief function changes its declivity, presenting critical points (maximum, minimum or inflection points). To identify such limit cells, we use a limit register associated to each cell. The border cells are assumed to be limit cells.

*Table 1.* Conditions of non limiting cells $xy$

| Id | Conditions |
|---|---|
| 1 | $l\text{-}m^{reg.e}_{(x-1)y} = l\text{-}m^{reg.e}_{xy} = 1$ |
| 2 | $l\text{-}m^{reg.w}_{xy} = l\text{-}m^{reg.w}_{(x+1)y} = 1$ |
| 3 | $l\text{-}m^{reg.e}_{(x-1)y} = l\text{-}m^{reg.e}_{xy} = l\text{-}m^{reg.w}_{xy} = l\ mbox{-}m^{reg.w}_{(x+1)y} = 0$ |
| 4 | $l\text{-}m^{reg.s}_{x(y-1)} = l\text{-}m^{reg.s}_{xy} = 1$ |
| 5 | $l\text{-}m^{reg.n}_{xy} = l\text{-}m^{reg.n}_{x(y+1)} = 1$ |
| 6 | $l\text{-}m^{reg.s}_{x(y-1)} = l\text{-}m^{reg.s}_{xy} = l\text{-}m^{reg.n}_{xy} = l\ mbox{-}m^{reg.n}_{x(y+1)} = 0$ |

DEFINITION 9 *The limit matrix of the layer $l$ is defined as the $n_r \times n_c$ matrix given by $l\text{-}M^{limit} = \left[l\text{-}m^{limit}_{xy}\right]$, where the entry at the $x$-th row and the $y$-th column is determined as $l\text{-}m^{limit}_{xy} = 0$, if one of the conditions listed in Table 1 holds, and $l\text{-}m^{limit}_{xy} = 1$, otherwise.*

Analyzing the limit matrix it is easy to detect the existence of known relief configurations. The presence of limit cells allows the subdivision of the whole area into declivity categories.

DEFINITION 10 *The constant declivity sub-region associated to the non limiting cell $xy$, denoted $l\text{-}\mathcal{SR}_{xy}$, is inductively defined as follows:* (i) *$xy \in l\text{-}\mathcal{SR}_{xy}$;* (ii) *If $x'y' \in l\text{-}\mathcal{SR}_{xy}$, then all its neighbor cells that are not limiting cells also belong to $l\text{-}\mathcal{SR}_{xy}$.*

Observe that $l\text{-}\mathcal{SR}_{xy} = l\text{-}\mathcal{SR}_{x'y'}$ if and only if $x'y' \in l\text{-}\mathcal{SR}_{xy}$ (resp., $xy \in l\text{-}\mathcal{SR}_{x\prime y\prime}$ ). Definition 10 leads to a recursive algorithm similar to the ones commonly used to fill polygons in computer graphics.

## 3.4 The Base Layer

The $\pi\text{-}M^{limit}$ layer of the *ICTM* model is used to receive the projection of the limiting cells of a set of layers. This projection is useful for the identification of interesting information, such as: (i) the cells which are limits in all layers; (ii) the projection of all sub-areas; (iii) the certainty degree of a cell to be limiting, etc.

Firstly, for this type of the analysis we propose two projection algorithms. In the first algorithm (type I), if the cell is a limit cell just in one layer then it will be projected in the base layer as limit cell. Thus, this projection method (Fig. 3) obtains all sub-regions found in all layers.

DEFINITION 11 *Each layer $l$ of the $n$-dimensional tessellation $M$ has associated a weigth $0 \leq w_i \leq 1$, for $i = 1, \ldots, l$.*

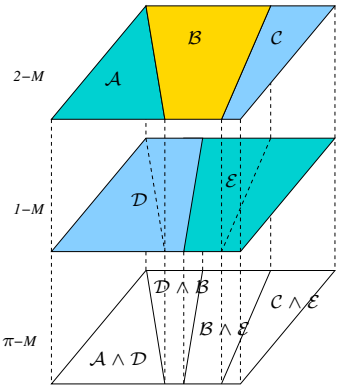


*Figure 3.* Type I Projection

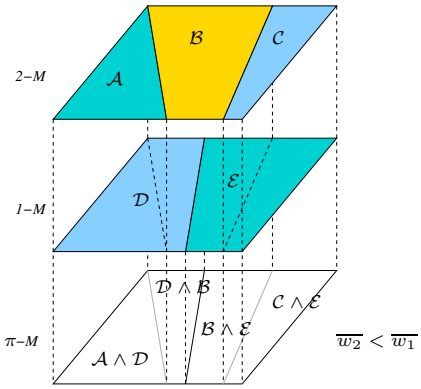


*Figure 4.* Type II Projection

In this algorithm, $w_i = 1$ ($w_i = 0$) indicates that the $i$-$M$ layer is (is not) selected for the projection.

DEFINITION 12 *Considering a bidimensional tessellation $M$ with $l$-layers, the $\pi$-$M_{xy}^{limit}$ projection (type I) denoting all limit cells of selected layers over the base layer is*

$$\pi\text{-}M_{xy}^{limit} = \bigvee_{i=1}^{l} i\text{-}M_{xy}^{limit} \times w_i, \forall xy \in M.$$

In the second projection method (type II), each layer may present different degrees of participation in the determination of the projection (Fig. 4). This degree of participation should be set by an especialist. Moreover, the layers can have non-normalized weights[3] since their sum may not be 1 (or 100%).

DEFINITION 13 *Considering the weights $0 \leq w_i \leq 1$ associated to $l$ layers of the tessellation $M$, the normalization of these weights denoted by $\overline{w_i}$ is given by:*

$$\overline{w_i} = \frac{w_i}{\sum_{j=1}^{l} w_j}, \text{ for } i = 1, \ldots, l.$$

DEFINITION 14 *Considering a bidimensional tessellation $M$ with $l$ layers and the normalized weights $\overline{w_i}$ associated to each layer, the $\pi$-$M_{xy}^{limit}$ projection (type II) of the $l$ layers over the base layer is given by:*

$$\pi\text{-}M_{xy}^{limit} = \sum_{i=1}^{l} i\text{-}M_{xy}^{limit} \times \overline{w_i}, \forall xy \in M.$$

In this projection method (Fig. 4), the limit cells in $\pi$-$M$ layer also indicates the certainty degree of these cells being limits according to the weights stipulated. The limit register $\pi$-$M_{xy}^{limit}$ will take values between 0 and 1. If the layer has weight $\overline{w_i} = 0$ then this layer isn't selected for the projection.

## 4. Some practical results

This section presents some results that this work has already reached and also some notes for future works. Up to the moment, the model of parallel processing that is being implemented creates independent processes for each analyzed property, each one of these processes performing sequentially all the *ICTM* steps (presented in the sections 3.1– 3.4).

However, it is also possible to expect for an increase of performance when the steps of the *ICTM* model are processed in parallel. Besides, the tessellation can be divided to be processed separately using an algorithm (i.e. the Schwarz algorithm (Chan and Mathew, 1994)) for overlapping domain decomposition. Domain decomposition methods are techniques for solving partial differential equations based on a decomposition of the spatial domain of the problem into several subdomains.

The implementation of the model is naturally parallel since the analysis is performed on the basis of local rules. Considering that the evaluation of each cell (a work unit) is independent of the others, this problem maps very well into the Single Program Multiple Data (SPMD) class of parallel applications. Our implementation, taking advantage of such characteristic, uses the MPI standard on top of a distributed processing cluster.

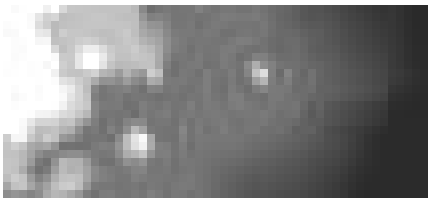

*Figure 5.* Graphic representation of DEM of 1000m

The results below use Digital Elevation Models of resolutions 1000 and 500 meters whose coordinates[4] are (i) Upper-left corner at ($X = 427559m$, $Y = 6637852m$) and (ii) Lower-right corner at ($X = 480339m$, $Y = 6614507m$).

The Fig. 5 represents the DEM of 1000m, including 24 rows and 53 columns, and was obtained from the digitalization of topographic maps

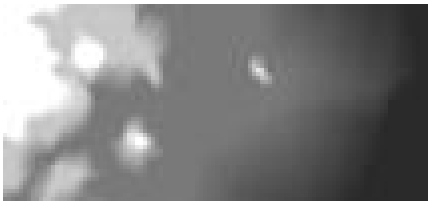

*Figure 6.* Graphic representation of DEM of 500m

at 1:1.000.000 scale. The Fig. 6 represents the DEM of 500m, with 49 rows and 106 columns, and was obtained from the interpolation of digitalized topographic maps at 1:1.000.000 scale.

*Table 2.* Number of Categories

| Radius | DEM1000m | DEM500m |
|---|---|---|
| 1 | 76 | 230 |
| 2 | 62 | 197 |
| 5 | 36 | 143 |
| 10 | 22 | 125 |
| 20 | 18 | 108 |
| 40 | 18 | 83 |

Some practical results are presented in Table 2. It can be observed that the number of categories obtained by the *ICTM* analysis is inversely proportional to the neighborhood radius. Moreover, it can be noted that for the resolution of 500m where each point has an smaller area (approx. four times) than the corresponding point in the DEM of resolution 1000m, the number of categories also follow (approx.) this factor. In this case, the DEM of 500m of resolution (of this region) does not present new sub-areas to be categorized, it just shows the sub-areas in more detail. However, if the region has bigger declivity variations then probably this factor cannot be verified anymore.

In plain areas, bigger neighborhood radius generate good approximations for the categories. For example, for the region A in the Fig. 7, its representations with bigger radius (Figs. 9 and 8) indicate reasonable approximations for this declivity degree. However, regions with bigger declivity variations obtained better aproximations with smaller radius.

Regions with bigger number of categories (i.e. region B, Fig. 7) are indicative that an analysis more detailed must be done (observe the same region in Fig. 10). The *ICTM* Model is regulated by two aspects: (i)

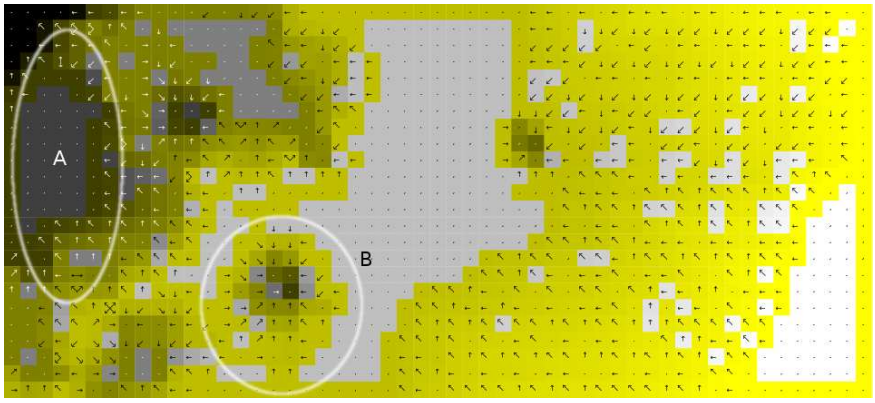


*Figure 7.* DEM of 1000m and $radius = 1$

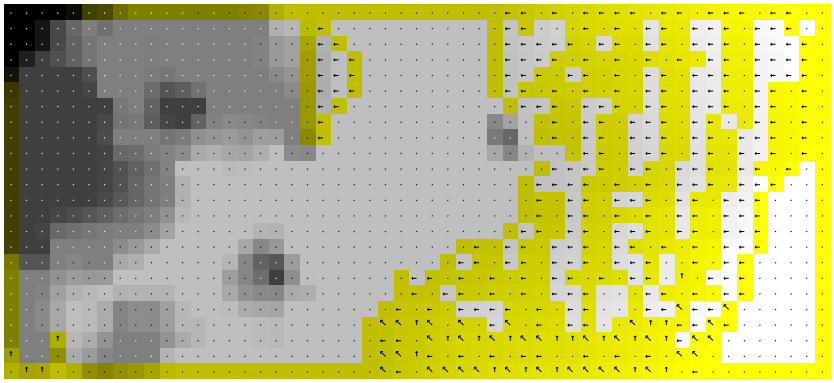

*Figure 8.* DEM of 1000m and $radius = 20$

by the spacial resolution of the digital elevation model, and (ii) by the neighborhood radius of the cell.

Thus, regions with an agglomeration of limit cells can be studied with more details by just augmenting the resolution of the altimetry data, or by reducing the neighborhood radius. In the *ICTM* model, the state of a cell in relation to its neighbors in declivity terms can be verified instantaneously, contrasting with the usual analyses (See Fig. 11).

In such type of information (countour lines) the cell has a status globally defined and in some cases just allowing a matching with distant cells. In the *ICTM* model, the information is punctual, associating all properties at the same point. Moreover, the size of the area, the DEM's resolution and the degree of variation of the declivity are the most impor-

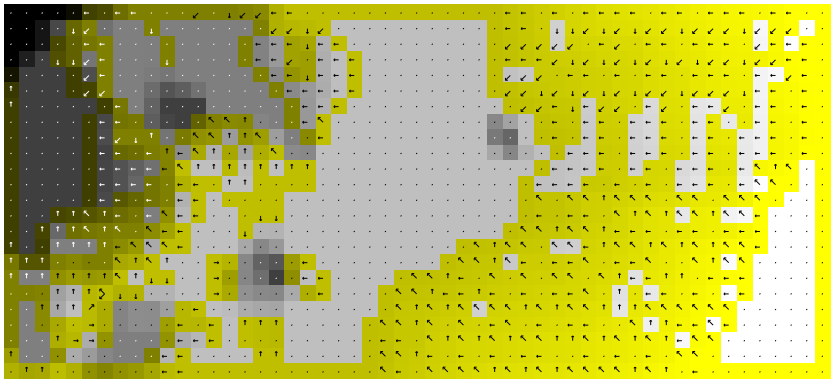

*Figure 9.* DEM of 1000m and $radius = 5$

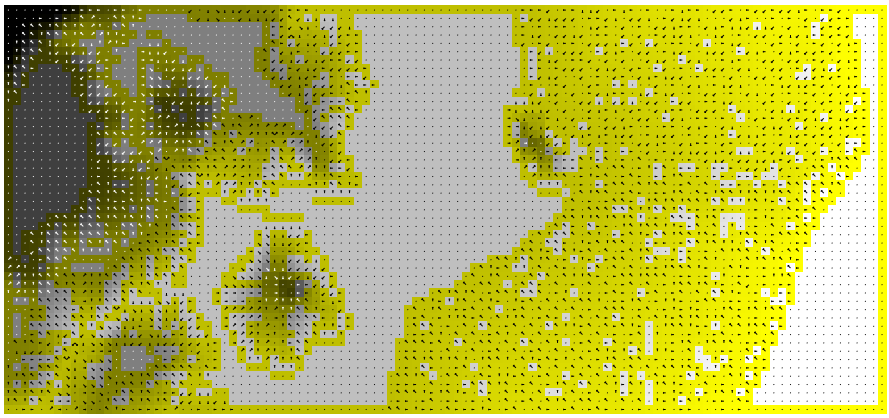

*Figure 10.* DEM of 500m and $radius = 1$

tant properties in the determination of a more appropriate neighborhood radius for a significant categorization.

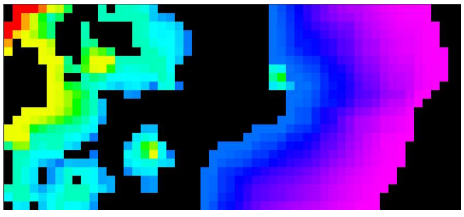

*Figure 11.* Contour lines – DEM of 1000m

The results obtained up to here indicate that regions with lesser variation of declivity are receptive to the *ICTM* model with bigger radius. In contrast, regions with great variations of declivity suggest radius smaller. The Fig. 12 indicates an classification according to the degree of declivity of the DEM with resolution of 1000 meters.

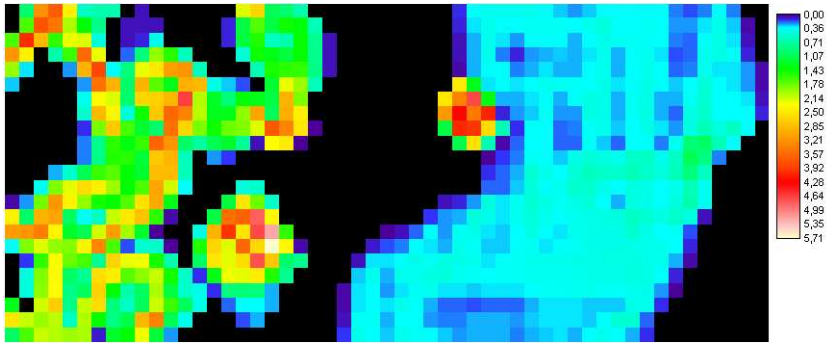


*Figure 12.* Slope Degree – DEM of 1000m

Evidently, small regions hold only lesser radius, therefore bigger radius of neighborhood tend to smooth a very great area of the cell in question.

## Notes

1. This paper uses the dot notation of the object-oriented programming languages to represent the components of a data structure ( e.g., *reg.e* denotes the component *e* of the data structure *reg*).

2. This is consistent with the relief function being a constant in the border cells.

3. The weights are naturally complementary when the analist have a clean perception of the importance of each layer for the process. If this correlation isn't possible *a priori*, then this feature may be an useful tool.

4. These coordinates are UTM 22S (South Hemisphere) and Datum SAD69 (South America Datum)

## References

Aguiar, M. S., Costa, A. C. R., and Dimuro, G. P. (2004a). ICTM: an interval tessellation-based model for reliable topographic segmentation. *Numerical Algorithms*. (to be published).

Aguiar, M. S., Dimuro, G. P., Costa, A. C. R., Finkelstein, A., and Kreinovich, V. (2004b). Separating components in interval-valued images. *Reliable Computing*. (to be published).

Berkhin, P. (2002). Survey of clustering data mining techniques. Technical report, Accrue Software, San Jose, CA.

Chan, T. F. and Mathew, T. P. (1994). Domain decomposition algorithms. *Acta Numerica*, pages 61–143.

Coblentz, D., Kreinovich, V., Penn, B., and Starks, S. (2003). Towards reliable subdivision of geological areas: Interval approach. In Reznik, L. and Kreinovich, V., editors, *Soft Computing in Measurements and Information Acquisition*, pages 223–233. Springer-Verlag, Berlin-Heidelberg.

Cooper, M. C. (1998). The tractability of segmentation and scene analysis. *International Journal on Computer Vision*, 30(1):27–42.

Diday, E. and Simon, J. C. (1976). Clustering analysis. In Fu, K. S., editor, *Digital Pattern Recognition*, pages 47–94. Springer Verlag.

Dubes, R. C. (1987). How many cluster are best?–an experiment. *Pattern Recognition*, 20(6):645–663.

Dubes, R. C. (1993). Cluster analysis and related issues. In Chen, C. H., Pau, L. F., and Wang, P. S. P., editors, *Handbook of Pattern Recognition & Computer Vision*, pages 3–32. World Cientific Publishing Co., River Edge, NJ.

Forman, R. T. T. (1995). *Land Mosaics: the ecology of landscapes and regions*. Cambridge University Press.

Fu, K. S. and Mui, J. K. (1981). A survey on image segmentation. *Pattern Recognition*, 13(1):3–16.

Jain, A. K. and Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall Inc., Englewood Cliffs, NJ.

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323.

Kearfort, R. B. and Kreinovich, V., editors (1996). *Applications of Interval Computations*. Kluwer, Dordrecht.

Lisani, J. L., Moisan, L., Monasse, P., and Morel, J. M. (2003). On the theory of planar shape. *Multiscale Modeling and Simulation*, 1(1):1–24.

Moore, R. E. (1979). *Methods and Applications of Interval Analysis*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

Umbaugh, S. E. (1998). *Computer Vision and Image Processing*. Prentice Hall, Upper Saddle River, New Jersey.

Villaverde, K. and Kreinovich, V. (1993). A linear-time algorithm that locates local extrenma of a function of one variable from interval measurements results. *Interval Computations*, 4:176–194.