4-1-2005

The Interval Categorizer Tesselation-Based Model for High Perfomance Computing

Marilton S. de Aguiar

Gracaliz P. Dimuro

Antonio C. da Rocha Costa

Rafael K.S. Silva

Comments:
UTEP-CS-04-10a.
Short version published in *Proceedings of the Workshop on State-of-the-Art in Scientific Computing PARA'04*, Lyngby, Denmark, June 20-23, 2004; extended version published in Jack Dongarra, Kaj Madsen, and Jerzy Wasniewski (eds.), PARA'04 Workshop on State-of-the-Art in Scientific Computing, Springer Lecture Notes in Computer Science, 2006, Vol. 3732, pp. 83-92.

HPC-ICTM: the Interval Categorizer Tessellation-Based Model for High Performance Computing

Marilton S. de Aguiar[1], Graçaliz P. Dimuro[1], Fábia A. Costa[1], Rafael K.S. Silva[2], César A.F. De Rose[2], Antônio C.R. Costa[13] and Vladik Kreinovich[4]

[1] Escola de Informática, Universidade Católica de Pelotas
Rua Felix da Cunha 412, 96010-000 Pelotas, Brazil
{marilton,liz,fabia,rocha}@ucpel.tche.br
[2] PPGCC, Pontifícia Universidade Católica do Rio Grande do Sul
Av. Ipiranga 6681, 90619-900 Porto Alegre, Brazil
{rksilva,derose}@inf.pucrs.br
[3] PPGC, Universidade Federal do Rio Grande do Sul
Av. Bento Gonçalves 9500, 91501-970 Porto Alegre, Brazil
[4] Department of Computer Science, University of Texas at El Paso
El Paso, TX 79968, USA
vladik@cs.utep.edu

Abstract. This paper presents the Interval Categorizer Tessellation-based Model (ICTM) for the simultaneous categorization of geographic regions considering several characteristics (e.g., relief, vegetation, land use etc.). Interval techniques are used for the modelling of uncertain data and the control of discretization errors. HPC-ICTM is an implementation of the model for clusters. We analyze the performance of the HPC-ICTM and present results concerning its application to the relief/land-use categorization of the region surrounding the lagoon *Lagoa Pequena* (RS, Brazil), which is extremely important from an ecological point of view.

1 Introduction

The ICTM (*Interval Categorizer Tessellation Model*) is a multi-layered and multi-dimensional tessellation model for the simultaneous categorization of geographic regions considering several different characteristics (relief, vegetation, climate, land use etc.) of such regions, which uses interval techniques [4, 9] for the modelling of uncertain data and the control of discretization errors.

To perform a *simultaneous categorization*, the ICTM proceeds (in parallel) to individual categorizations considering one characteristic per layer, thus generating different subdivisions of the analyzed region. An appropriate projection procedure of the categorizations performed in each layer into a basis layer provides the final categorization that allows the combined analysis of all characteristics that are taken into consideration by the specialists in the considered application, allowing interesting analyzes about their mutual dependency.

An implementation of the ICTM for the relief categorization of geographic regions, called TOPO-ICTM (*Interval Categorizer Tessellation Model for Reliable Topographic Segmentation*), performs a bi-dimensional analysis of the declivity of the relief function in just one layer of the model [1]. The data input are extracted from satellite images, where the heights are given in certain points referenced by their latitude and longitude coordinates. The geographic 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 or ecology analyst and it is directly associated to the refinement degree of the tessellation. Applications in Geophysics and Ecology were found, where an adequate subdivision of geographic areas into segments presenting similar topographic characteristics is often convenient (see, e.g: [2, 5]).

The aim of this paper is to present the ICTM model and describe a particular implementation of the model for clusters, called HPC-ICTM. We discuss the performance of the HPC-ICTM and present some results concerning the application of a 2-layered bi-dimensional model to the *relief/land use* categorization of the region surrounding the lagoon *Lagoa Pequena* (Rio Grande do Sul, Brazil), which is extremely important from an ecological point of view.

The paper is organized as follows. Section 2 presents the ICTM model and the categorization process. Some results on categorizations are shown in Sect. 3. The performance of the HPC-ICTM is discussed in Sect. 4. Section 5 is the Conclusion.

2 The ICTM Model

This section introduces the multi-layered interval categorizer tessellation-based model ICTM, formalized in terms of matrix operations, extending the results presented in a previous paper [1][5], for the single-layered model TOPO-ICTM.

A *tessellation*[6] is a matrix M, whose each entry[7] at the x-th row and the y-th column is denoted by m_{xy}. For $L \in \mathbb{N}$ and tessellations $M^1, \ldots, M^L$, an *L-layered tessellation* is a structure $\mathcal{M}^L = \left(M^1, \ldots, M^L\right)$, where each entry at the l-th layer, x-th row and y-th column is denoted by m^l_{xy}.

2.1 Using Interval Matrices

In many applications considered by the Geophysics and Ecologists, usually there are too much data to be analyzed, most of which is irrelevant. We take, for each subdivision of the geographic region, the average of the values attached to the points of each layer, which are the entries m^l_{xy} of the L matrices of the tessellation $\mathcal{M}^L$. To simplify the data, we normalize them by dividing each m^l_{xy}

[5] The proofs are omitted since they are similar to those presented in [1].

[6] To simplify the notation, we consider bi-dimensional tessellations only.

[7] For the application considered in this paper, the entries of the tesselation matrices are all non-negative. However, negative values may also be considered (e.g., when the data coming from the relief are determined with respect to the sea level).

by the largest m^l_{max} of these values, obtaining a relative value $rm^l_{xy} = \frac{m^l_{xy}}{|m^l_{max}|}$. The *relative matrix* of a layer l is given by $RM^l = \frac{M^l}{|m^l_{max}|}$.

Sometimes we have problems in representing the uncertain data provided by the sources of the considered application. Even if the values associated to the points are pretty accurate (like, for instance, the heights provided by satellite images), we have to deal with the errors that came from the discretization of the area in terms of the discrete set of the tessellation cells. We apply techniques from Interval Mathematics [4, 9] to control the errors associated to the cell values. See examples of using intervals in solving similar problems in [2, 4, 9].

Observe that, considering a layer l, for each $\xi\upsilon$ that is different from xy, it is reasonable to estimate a value $h^l_{\xi\upsilon}$, attached to the point $\xi\upsilon$ in the layer l, as the value rm^l_{xy} 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 in the layer l, let Δ^l_x be the largest possible error of the corresponding approximation considering the west-east direction. Then the approximation error ϵ^l_x is bounded by $\epsilon^l_x \leq \Delta^l_x = 0.5 \cdot \min(|rm^l_{xy} - rm^l_{(x-1)y}|, |rm^l_{(x+1)y} - rm^l_{xy}|)$. Now, for each cell xy in the layer l, let Δ^l_y be the largest possible error of the corresponding approximation considering the north-south direction. Therefore, the *approximation error* ϵ^l_y is bounded by $\epsilon^l_y \leq \Delta^l_y = 0.5 \cdot \min(|rm^l_{xy} - rm^l_{x(y-1)}|, |rm^l_{x(y+1)} - rm^l_{xy}|)$.

Thus, considering a given y in the layer l, the intervals $im^{x,l}_{xy} = [m^l_{x^-y}, m^l_{x^+y}]$, where $m^l_{x^-y} = rm^l_{xy} - \Delta^l_x$ and $m^l_{x^+y} = rm^l_{xy} + \Delta^l_x$, contain all the possible values of $h^l_{\xi y}$, for $x - \frac{1}{2} \leq \xi \leq x + \frac{1}{2}$. Similarly, for a fixed x in the layer l, for each y such that $y - \frac{1}{2} \leq \upsilon \leq y + \frac{1}{2}$, it follows that $h^l_{x\upsilon} \in im^{y,l}_{xy} = [m^l_{xy^-}, m^l_{xy^+}, m^l_{xy^-} = rm^l_{xy} - \Delta^l_y$, with $m^l_{xy^+} = rm^l_{xy} + \Delta^l_y$. For each layer l, the *interval matrices* associated to the relative matrix RM^l are denoted by $IM^{x,l}$ and $IM^{y,l}$.

2.2 The Categorization Process

To obtain a declivity categorization[8] in each layer, we assume that the approximation functions introduced by the model are piecewise linear functions[9]. The model determines a piecewise linear approximation function (and corresponding set of limit points between the resulting sub-regions) that fits the constraints imposed by the interval matrices. To narrow the solution space to a minimum, we take a qualitative approach to the approximation functions, clustering them in equivalence classes according to the signal of their declivity (positive, negative,

[8] This declivity categorization was inspired by [2]. By *declivity* we mean the tangent of the angle α between the approximation function and the positive direction of the horizontal axis. The declivity happens to be continuous, since the approximation functions of the model are total and have no steps. The declivity is positive, negative or null if $0 < \alpha < \frac{\pi}{2}$, $\frac{\pi}{2} < \alpha < \pi$ or $\alpha = 0$, respectively.

[9] Piecewise linear functions were considered for simplicity. A more general approach may consider other kinds of piecewise monotonic functions.

null). For each layer l, the model obtains the class of approximation functions compatible with the constraints of the interval matrices $IM^{x,l}$ and $IM^{y,l}$.

Proposition 1. *For a given xy in a layer l, it holds that:*

(i) Considering the direction west-east, if $m^l_{x^+y} \geq m^l_{(x+1)^-y}$, then there exists a non-increasing relief approximation function between xy and $(x+1)y$.
(ii) Considering the direction west-east, if $m^l_{(x-1)^-y} \leq m^l_{x^+y}$, then there exists a non-decreasing relief approximation function between $(x-1)y$ and xy.
(iii) Considering the direction north-south, if $m^l_{xy^+} \geq m^l_{x(y+1)^-}$, then there exists a non-increasing relief approximation function between xy and $x(y+1)$.
(iv) Considering the direction north-south, if $m^l_{x(y-1)^-} \leq m^l_{xy^+}$, then there exists a non-decreasing relief approximation function between $x(y-1)$ and xy.

Definition 1. *A* declivity register *of an xy-cell in a layer l is a tuple $dm^l_{xy} = (e^l_{xy}, w^l_{xy}, s^l_{xy}, n^l_{xy})$, where e^l_{xy}, w^l_{xy}, s^l_{xy} and n^l_{xy}, called* directed declivity registers *for east, west, south and north directions, respectively, are evaluated according to the conditions considered in Prop. 1:*

(a) For non border cells: $e^l_{xy} = 0$, if (i) holds; $w^l_{xy} = 0$, if (ii) holds; $s^l_{xy} = 0$, if (iii) holds; $n^l_{xy} = 0$, if (iv) holds; $e^l_{xy} = w^l_{xy} = s^l_{xy} = n^l_{xy} = 1$, otherwise.
(b) For east, west, south and north border cells: $e^l_{xy} = w^l_{xy} = s^l_{xy} = n^l_{xy} = 0$, respectively. The other directed declivity registers of border cells are also determined according to item (a).

The declivity register matrix of the layer l is the matrix $dM^l = \left[dm^l_{xy}\right]$.

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

Associating convenient weights to the directed declivity registers of a cell xy in a layer l, it is possible to obtain a binary encoding that represents the *state* sm^l_{xy} of such cell, given by $sm^l_{xy} = 1 \times e^l_{xy} + 2 \times s^l_{xy} + 4 \times w^l_{xy} + 8 \times n^l_{xy}$. The *state matrix* of layer l is given by $SM^l = \left[sm^l_{xy}\right]$. Thus, any cell can assume one and only one state represented by the value $sm^l_{xy} = 0..15$.

A *limiting cell* is the one where the relief function changes its declivity, presenting critical points (maximum, minimum or inflection points). According to this criteria, any non-limiting cell should satisfy one of the conditions listed in Table 1. The border cells are assumed to be limiting. To identify the limiting cells, we use a *limiting register* λm^l_{xy} associated to each xy-cell of a layer l, defined as:

$$\lambda m^l_{xy} = \begin{cases} 0 \text{ if one of the conditions listed in Table 1 holds;} \\ 1 \text{ otherwise.} \end{cases} \quad (1)$$

The *limiting matrix* of the layer l is $\lambda M^l = \left[\lambda m^l_{xy}\right]$. Analyzing this matrix, we proceed to the subdivision of the whole area into constant declivity categories.

Table 1. Conditions of non limiting cells

Conditions	Conditions
$e^l_{(x-1)y} = e^l_{xy} = 1$	$n^l_{xy} = n^l_{x(y+1)} = 1$
$w^l_{xy} = w^l_{(x+1)y} = 1$	$s^l_{x(y-1)} = s^l_{xy} = 1$
$e^l_{(x-1)y} = e^l_{xy} = w^l_{xy} = w^l_{(x+1)y} = 0$	$s^l_{x(y-1)} = s^l_{xy} = n^l_{xy} = n^l_{x(y+1)} = 0$

Definition 2. *The constant declivity sub-region $\mathcal{SR}^l_{xy}$, associated to a non limiting cell xy in a layer l, is inductively defined: (i) $xy \in \mathcal{SR}^l_{xy}$; (ii) If $x'y' \in \mathcal{SR}^l_{xy}$, then all its neighbor cells that are not limiting cells also belong to $\mathcal{SR}^l_{xy}$.*

Definition 2 leads to a recursive algorithm similar to the ones commonly used to fulfill polygons.

2.3 The Projection Procedure

The basis layer λM^π is used to receive the *projection* of the limiting cells of all layers. This projection is useful for the identification of interesting information, such as: (i) the cells which are limiting in all layers; (ii) the projection of all sub-areas; (iii) the certainty degree of a cell to be limiting etc.

Two projection algorithms were proposed. In the first algorithm (*Type I*, Fig. 1), if a certain cell is a limiting cell just in one layer then it will be projected on the basis layer also as a limiting cell. A weight $0 \leq w_l \leq 1$, for $l = 1, \ldots, L$ is associated to each layer, so that $w_l = 1$ ($w_l = 0$) indicates that the layer l is (is not) selected for the projection. Then, the projection of limiting cells on the basis layer is given by $\lambda M^\pi = \bigvee_{l=1}^{L} \lambda M^l \times w_l$. In the second algorithm (*Type II*, Fig. 2), each layer may present different degrees of participation in the determination of the projection[10]. In this case, the projection of limiting cells on the basis layer is given by $\lambda M^\pi = \bigvee_{l=1}^{L} \lambda M^l \times \overline{w}_l$, where $\overline{w}_l = \dfrac{w_l}{\sum_{i=1}^{L} w_i}$ are the normalized weights.

3 Some Results on Categorizations

This section presents the *relief* and *land use* categorizations obtained for the region surrounded the lagoon *Lagoa Pequena* (Rio Grande do Sul, Brazil). These

[10] Moreover, the sum of the weights may not be equal to 1 in the case that the analyst does not have a clear perception of the importance of each layer in the whole process.

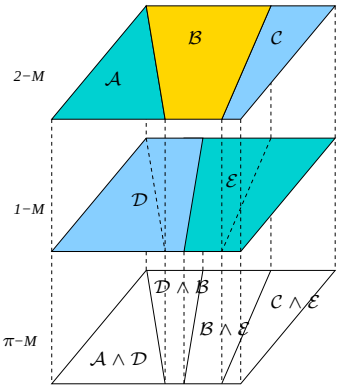

Fig. 1. *Type I* projection procedure

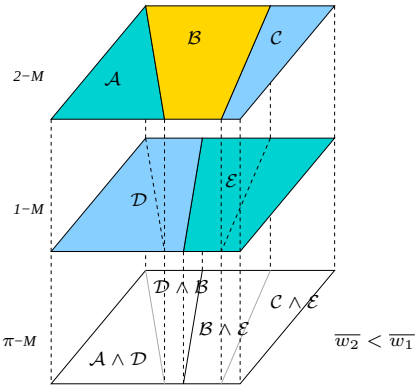

Fig. 2. *Type II* projection procedure

analyzes are to be used for the environment characterization of that region, aiming to give subsidies for its integrated preservation/management.

Figure 3 shows the location of the lagoon and a *land use* categorization of the region surrounded it, which shall be combined with relief categorizations. For the portion of the LANDSAT image[11] shown in Fig. 4(a), the `ICTM` produced the relief categorization presented in Fig. 4(b), for the Digital Elevation Model (DEM), and in Fig. 4(c), for a 3D visualization of this categorization. Figure 4(c) shows the ICTM relief characterization given in terms of the state and limiting matrices, where a pleistocene marine barrier can be distinguished.

4 Performance Analysis

The parallel implementation of the `ICTM` model for clusters was done using the library MPI (*Message Passing Interface*) [7]. Each process of the parallel program is responsible for a categorization performed in one of the layers of the model, in a modified *master-slave* schema, where the slave processes receive the information sent by the master process and, after executing their jobs, they generate their own outputs. The master process is responsible for loading the input files and parameters (the data and the radius), sending the radius value for the L slave processes to start the categorization process in each layer. The file system is shared, that is the directory with the input files is accessible by all the cluster's nodes.

For the analysis of the performance, we consider three tessellation matrices: M_1 (241×241), M_2 (577×817) and M_3 (1309×1765). The results were processed by the CPAD (Research Center of High Performance Computing of PUCRS/HP, Brasil), with the following environment: (i) heterogenous cluster with 24 machines: ($e800$) 8 machines with two processors Pentium III 1Ghz and

[11] The coordinates are Datum SAD69 (South American Datum 1969) and UTM (Universal Transverse Mercator), Zone 22 South Hemisphere.

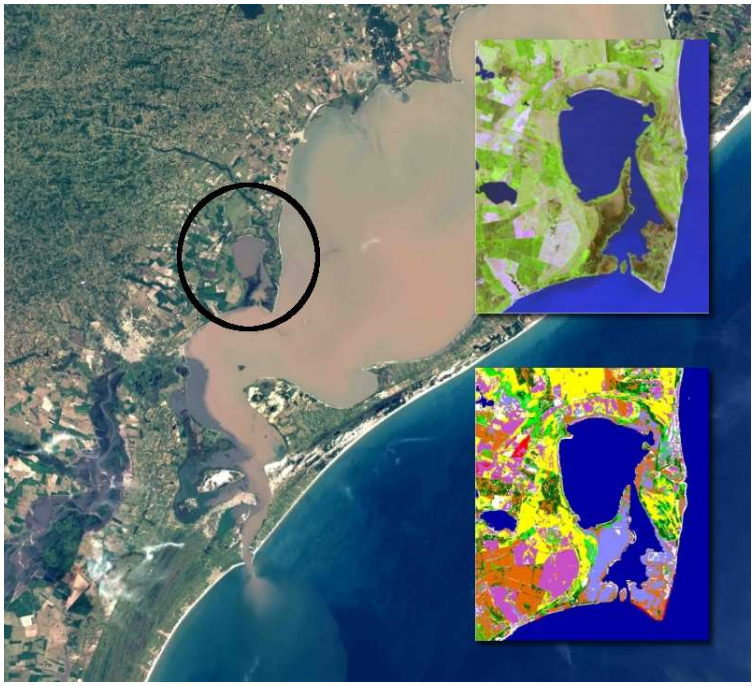

Fig. 3. Land use Map of the region surrounded the *Lagoa Pequena* (light blue: wetland, dark blue: water, yellow: crops and past, purple: transitional, light green: riparia forest, dark green: restinga forest, red: lagoon beaches, white: without classification)

256MB of memory, ($e60$) 16 machines with two processors Pentium III 550Mhz and 256MB of memory; (ii) Ethernet and Myrinet networks; (iii) MPICH version of MPI Library; (iv) 1 front-end machine with two processors Pentium III 1Ghz and 512MB of memory. Table 2 presents the results of sequential implementation processed in the cluster front-end, and also the results of the parallel implementation of `ICTM`. As expected, the machine $e800$ is faster then the machine $e60$. However, in some cases, the *ethernet network* was faster than *myrinet network*, due to the low volume of inter-processor communication. In general, the difference between the performance of the two networks is observed when the processors require a lot of message exchanges. Notice that, even when the number of layers increased, a significant time variation was not observed.

One interesting feature of the parallel implementation is the partition of input data, which reduces the amount of memory to be stored in a single processor. We observe that, in the sequential implementation, the data of all layers has to be stored in a unique processor. Table 3 presents the *speedups* for the tests realized. Observe that a parallel implementation may be considered when the `ICTM` presents more than one layer. However, it becomes really necessary in the case that it has a great amount of layers, since, in this case, a sequential implementation is practically not feasible.

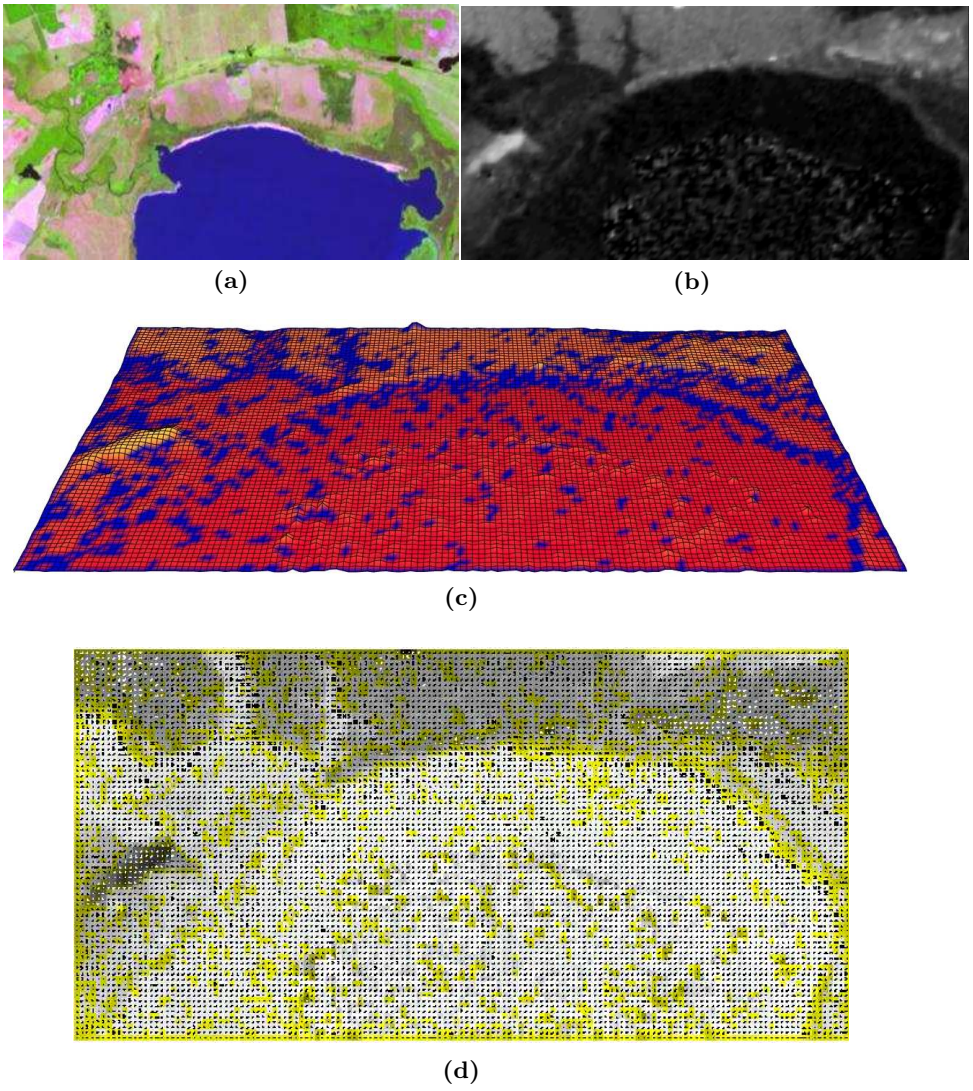

Fig. 4. Relief categorizations of a portion of the region surrounding the *Lagoa Pequena*: (a) LANDSAT image, coordinates: Upper-Left Corner (X: 390735, Y:6512015), Lower-Right Corner (X: 402075, Y:6505685), Pixel Size (X: 30m, Y: 30m); (b) ICTM DEM categorization; (c) ICTM 3D categorization; (d) ICTM status-limits categorization

5 Discussion and Conclusion

In the categorizations produced by the ICTM, the state of a cell in relation to its neighbors, concerning the declivity, is shown directly by arrows (see Fig. 5), which has been considered a very intuitive representation, by the ecologists,

Table 2. Analysis of `ICTM` results (T_s is the time of sequential implementation and $T_{e60-Eth}$, $T_{e800-Eth}$, $T_{e60-Myr}$, $T_{e800-Myr}$ are the times of parallel implementations)

Matrix	# of Layers	T_s	# of Proc.	$T_{e60-Eth}$	$T_{e800-Eth}$	$T_{e60-Myr}$	$T_{e800-Myr}$
M_1	3	1.947s	4	2.015s	1.340s	1.998s	1.209s
M_1	5	3.298s	6	2.228s	1.510s	2.263s	1.254s
M_2	3	16.041s	4	11.180s	6.038s	12.003s	6.248s
M_2	5	27.015s	6	11.383s	6.194s	12.011s	6.331s
M_3	3	15m50.871s	4	50.277s	27.635s	54.742s	28.982s
M_3	5	35m14.322s	6	50.551s	27.665s	55.657s	29.068s

Table 3. *Speedups* for the tests realized

Matrix	# of layers	e60-Eth	e800-Eth	e60-Myr	e800-Myr
M_1	3	0.97	1.45	0.97	1.61
M_1	5	1.48	2.18	1.46	2.63
M_2	3	1.43	2.66	1.34	2.57
M_2	5	2.37	4.36	2.25	4.27
M_3	3	18.91	34.41	17.37	32.81
M_3	5	41.83	76.43	37.99	72.74

since most geographic information systems present this kind of result by the usual color encoding of declivity, with no indication of direction.

The ICTM is regulated by two aspects, namely, the spacial resolution of the DEM and the neighborhood radius of the cell. Thus, regions with an agglomeration of limiting cells can be studied with more details by just increasing the resolution of altimetry data, or reducing the neighborhood radius. In plain areas (see Fig. 5 (region A)), a large neighborhood radius indicated reasonable approximations for the declivity degree. However, regions with too much declivity variation (see Fig. 5 (region B)) obtained good approximations only with small radius. The number of categories obtained is always inversely proportional to the neighborhood radius and to the area of a tessellation cell.

The analysis of some related works concerning image segmentation [3, 6, 8, 10] turns out that those methods are, in general, heuristic, and, therefore, the `ICTM` model presented here is more reliable (for other works, see, e.g.: [2, 11]).

Future work is concerned with the aggregation of a dynamic structure, based on cellular automata [12], for the modelling of the dynamic of populations in a ecological context.

Acknowledgments. We are very grateful to Alex Bager for his suggestions concerning ecological aspects. We also thank the referees for the valuable comments. This work is supported by the CNPq and FAPERGS. V.K. was also supported by NASA under cooperative agreement NCC5-209, by Future Aerospace Science and Technology Program Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under the 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.

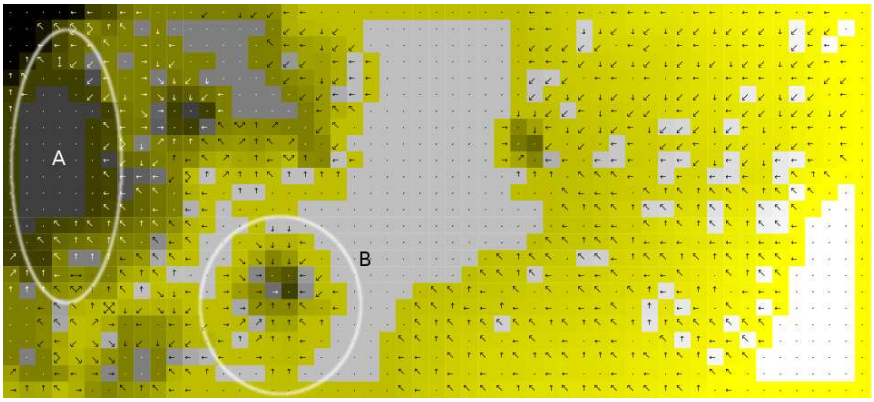

Fig. 5. Relief categorization of a portion of LANDSAT image with coordinates: Upper-left corner (X:427559m, Y:6637852m), Lower-right corner (X;480339m, Y:6614507m).

References

1. M.S. Aguiar, G.P. Dimuro, and A.C.R. Costa. TOPO-ICTM: an interval tessellation-based model for reliable topographic segmentation. *Numerical Algorithms* 37(1): 3–11, 2004.
2. D. Coblentz, V. Kreinovich, B. Penn, and S. Starks. Towards reliable sub-division of geological areas: interval approach. In L. Reznik and V. Kreinovich, editors, *Soft Computing in Measurements and Information Acquisition*, pages 223–233, 2003. Springer-Verlag.
3. M. C. Cooper. The Tractability of Segmentation and Scene Analysis. *International Journal on Computer Vision*, 30(1): 27–42, 1998.
4. R.B. Kearfort and V. Kreinovich (eds.). *Aplications of Interval Computations*. Kluwer, Dordrecht, 1996.
5. R.T.T. Forman. *Land Mosaics: the ecology of landscapes and regions*. Cambridge University Press, Cambridge, 1995.
6. K.S. Fu and J.K. Mui. A Survey on Image Segmentation. *Pattern Recognition*, 13(1): 3–16, 1981.
7. W. Gropp, E. Lusk e A. Skjellum. *Using MPI: portable Parallel Programming with the Message-Passing Interface*, MIT Press, 2nd ed., 1999.
8. J.L. Lisani, L. Moisan, P. Monasse, and J.M. Morel. On The Theory of Planar Shape. *Multiscale Modeling and Simulation*, 1(1): 1–24, 2003.
9. R.E. Moore. *Methods and Applications of Interval Analysis*. SIAM, Philadelphia, 1979.
10. S.E. Umbaugh. *Computer Vision and Image Processing*. Prentice Hall, New Jersey, 1998.
11. K. Villaverde and V. Kreinovich. A Linear-Time Algorithm that Locates Local Extrenma of a Function of One Variable from Interval Measurements Results. *Interval Computations*, 4: 176–194, 1993.
12. S. Wolfram. *Cellular Automata and Complexity: selected papers*. Addison-Wesley, Readings, 1994.