# Cell-based Model For GIS Generalization

**Bo Li, Graeme G. Wilkinson & Souheil Khaddaj**

*School of Computing & Information Systems*
*Kingston University*
*Penrhyn Road, Kingston upon Thames*
*Surrey, KT1 2EE*
*UK*
*b.li@kingston.ac.uk*, *G.Wilkinson@Kingston.ac.uk*, s.khaddaj@kingston.ac.uk

**Abstract.** Generalization is perhaps the most intellectually challenging task for cartographers. It has proved to be very difficult to automate. In this paper, a cell-based model is applied to GIS generalization, which represents a new methodology within the GIS context. Cellular Automata (CA) has found a place in many interesting real world applications, including the modeling and simulation of numerous systems, across many disciplines. CA has a number of unique advantages in geographical and environmental modeling. Thus, it has attracted growing attention in urban simulation because of its potential in spatial modeling. Geographical phenomena have extremely complex characteristics as a result of interactions among different components in a study area. CA provides a promising new approach to simulate and understand spatial phenomena. In this study, which is based on CA techniques, an extended neighborhood algorithm is used in the cell-based model to automatically generalize raster thematic maps derived from classified satellite images. An example of generalizing a land use map of Lisbon Bay in Portugal is given, which gives satisfactory results.

## 1. INTRODUCTION

In recent years, there has been a drive towards automating map generalization. As geographical information systems (GIS) have become more prevalent, the issue of generalization has increased in importance. Generalization is perhaps the most intellectually challenging task for cartographers, a proposition supported by the comparatively marginal success of computer algorithms in generalizing maps. Generalization is needed in order to represent information on an appropriate level of detail. As only a restricted amount of data can be represented on a certain level of detail, different pieces of information have to 'fight' for their representation on a specific aggregation level. This implies that generalization is an optimization problem, where different goals have to be satisfied simultaneously. It is a difficult procedure, as only a limited amount of data can be visualized, perceived and understood at a time. In recent years, the automatic generalization techniques have become popular. Due to its complex, diverse and non-deterministic nature, the generalization process has proved to be very difficult to automate, particularly because one is attempting to mimic a subjective and intuitive procedure (Joao, 1998).

The demand for spatial information continues to grow and satellite sensors represent a fast source of data compared to traditional map-making and aerial photo-interpretation. GIS have an important role in the successful creation of a map. There is a general agreement that GIS is more than just a tool for geographic and spatial database management, nor only a tool for automated-cartography and it is not only a set of procedures to manipulate and introduce remotely sensed data into the map-making process. GIS and satellite data, once integrated, can be successfully used for environmental monitoring, analysis, modeling and decision-making.

In this paper, a cellular automata model is applied to GIS generalization. This is a new application of cellular automata within the GIS context. We start by a brief overview of GIS generalization techniques. Then, cellular automata techniques are discussed. The cellular automata model is then presented. Finally some conclusions are drawn.

## 2. GIS GENERALIZATION

The map is an abstract model of reality and represents a medium for the comprehension, the record and the communication of spatial relationships and forms. As a communication medium the information represented in the map has been derived using cartographic generalization and design (Goffredo, 1998). The traditional process of map production, in fact, is based on manual data manipulation and visual interpretation of aerial photographs or satellite images. Thus, the experience, the intuition, the imagination and the inductive capability of the interpreter are instinctively combined together for the extraction of information at a high level of abstraction.

There are two main limits in spatial analysis (Wilkinson, 1993). Firstly, when organizing and manipulating data in order to emphasis the 'selected' information, other information is irreversibly destroyed; manual manipulation of data guided by the cartographer expert is in fact not repeatable (by another cartographer or by the same one at another time) being based on intuition and subjectivity. Secondly, there is lack of geographical precision in the majority of maps when real objects are generalized into nominal categories (for example, an area object is generalized into a point object). The use of GIS for spatial analysis requires accurate spatial location, therefore, in this context, cartography and GIS tools are not compatible. Satellite images and image processing techniques may be used to compensate the gap between map information and GIS data requirements, providing strategies can keep trace of the geographical relationship of the generalized object and its real position on the ground with raster analysis.

The rapid development of spatial information technologies (primarily the development of GIS and raster image processing and modeling tools) over the last few decades has allowed the space-time realization of many cellular models. The increasing availability of higher spatial resolution image data and the number of sources of remotely sensed imagery have provided a temporal information dimension from which time series analysis can provide stochastic representation of landscape transformations (Candau, 2000). Cellular models in particular offer a useful space-time modeling environment in which raster-based information on spatial and temporal landscape change (derived from remotely sensed imagery) and information on factors that influence change (e.g., topographic factors derived from a digital elevation model) can be brought together. These types of models provide effective ways of understanding the process of urban development as well as offering a means of evaluating the environmental and social consequences of alternative planning scenarios.

The study of cellular automata (CA) began as a theoretical field. Today, however, CA have found a place in many interesting real world applications, including the modeling and simulation of numerous systems, across many disciplines –including geography (Park, 1997). CA models are discrete-time system models with spatial extensions. The abilities of CA to model the complex order hidden in spatial detail have been demonstrated. CA has a number of unique advantages in geographical and environmental modeling. Firstly, CA is capable of generating very complex, a global spatial pattern by using simple, local transition rules. Secondly, CA can produce fractal structure, which is a natural representation of the hierarchy between local and global behavior. It generates global spatial behavior based on the local knowledge of individual cells. Thirdly, CA combine with the spatial information stored in GIS relatively easily. CA is simple in principles, wide in potential applications, and hierarchical in nature, and so they are powerful in theory. CA has attracted growing attention in urban simulation because of their potential in spatial modeling.

## 3. CELLULAR AUTOMATA

Cellular automata (CA) have found common place applications in statistical and theoretical physics and are linked to considerations of chaos theory and fractal geometry. More recently, cellular automata applications have found their way into 2-D applications in urban growth modeling. It is an approach with growing importance in discrete dynamic systems. CA theory was first introduced by John von Neumann in 1950s and gained considerable popularity two decades later through the work of John Conway in the game of life. CA is non-linear dynamic mathematical systems based on discrete time and space. The basic idea is very simple: a cellular automaton evolves in discrete time-steps by updating its states (i.e. cell value) according to a transition rule that is applied universally and synchronously to each cell at each time-step. The value of each cell is determined based on a geometric configuration of neighbor cells, which is specified as part of the transition rule. Updated values of individual cells then become the inputs for the next iteration. As iteration proceeds, an initial cellular configuration, which is a kind of cellular map containing an initial state of each cell, evolves based on the rules defined. The transition rules can be a deterministic or a probabilistic function of the neighborhood.

In terms of structure, this computational scheme is similar to the ones employed in the numerical manipulation of partial differential equations. The difference is that the state variable at each cell is only allowed to assume a small set of value and the transition functions do not assume an algebraic form, but may be deterministic or stochastic. Some scholars proposed that CA can be an alternative to differential equations in the modeling of physical phenomena. One important characteristic of CA is that complex global behavior across the whole cellular space may emerge from the application of simple local rules. The basic principle that drives the system through time is based on the notion that the states of cells change as a function of what is happening to other cells in their local neighborhood (Batty, 2000). Cellular automata models are purposed to be scale independent. The growth rules are integral to the data set being used because they are defined in terms of the physical nature of the location under study, thus producing a scale-independent model, though data sets are scale dependent themselves.

## 4. MODELING AND SIMULATION

In this section, a cell-based model is proposed. A classified satellite image of Lisbon Bay is used as an example to be generalized by the cell-based model. Here, the extended Moore neighborhood is applied to this model, which is shown in Figure 1. Basically, the state of the interested cell is compared to that of a group of neighbor cells that are adjacent to one another. Here, the combination of cell-based model and majority filtering is used to give a better result. The former one will keep the main feature and the structure of the image and the latter one will help to smooth it. It is Kept unchanged if it is the same state as the group being looked at, otherwise it is changed to the majority state of the local area.

The algorithm is given as follow:

```
For each iteration
{
    For every pixel in the image
    {
        If cell is the same state as its group made
        by several adjacent neighbor cells

            Keep the state of the cell
            unchanged

        Else choose the majority cells' value

    }

}
```

Using the output as the new input of the next iteration iterates the procedure. The state of cells stays stable after certain times of repetition.

Figures 2-3 give the original satellite image and generated one.

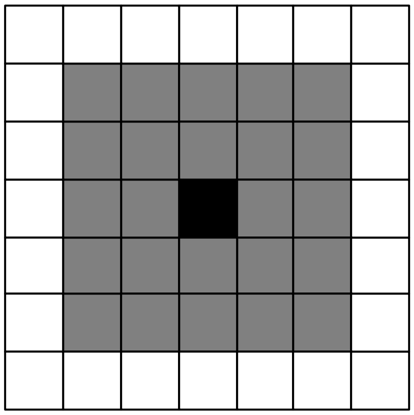

**Figure 1.** Extended Moore Neighborhood

**5. CONCLUSION**

From the used example it is obviously clear that the generalized image is closer to the manually made one at some point when use extended Moore neighborhood. Using Moore neighborhood doesn't give a very good result because it probably cannot capture the behavior of the system very accurately, especially when the size of the input image is big. The simulation shows that CA technique works for raster based map generalization. It is shown there is a promising future in this area. However, this model still needs more development and improvement. The main problem is that it cannot do any intelligent decision or judgement at this stage. In future work we will be considering applying some artificial intelligence technique to the CA model to help giving a better-generalized image.

## References


Batty, M. (2000): 'Geocomputational Modelling Using Cellular Automata'

Candau, J. (2000) : 'A Coupled Cellular Automaton Model for Land Use / Land Cover Dynamics', *4th International Conference on Intergrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs*. Banff, Alberta, Canada, September 2-8, 2000

Goffredo, S. (1998): '*Automatic Generalization of Satellite-derived Land Cover Information*', European Commission

Joao, E. M., (1998): *Causes and Consequences of Map Generalisation*, Taylor & Francis.

Park, S. and Wagner, D. F., (1997) : 'Incorporating Cellular Automata simulators as analytical engines in GIS', *Transaction in GIS*, pp213-231, vol.2, no. 3.

Wilkinson, G. G. (1993): 'The Generalisation of Satellite-derived Raster Thematic Maps for GIS Input', *GIS*, vol. 5

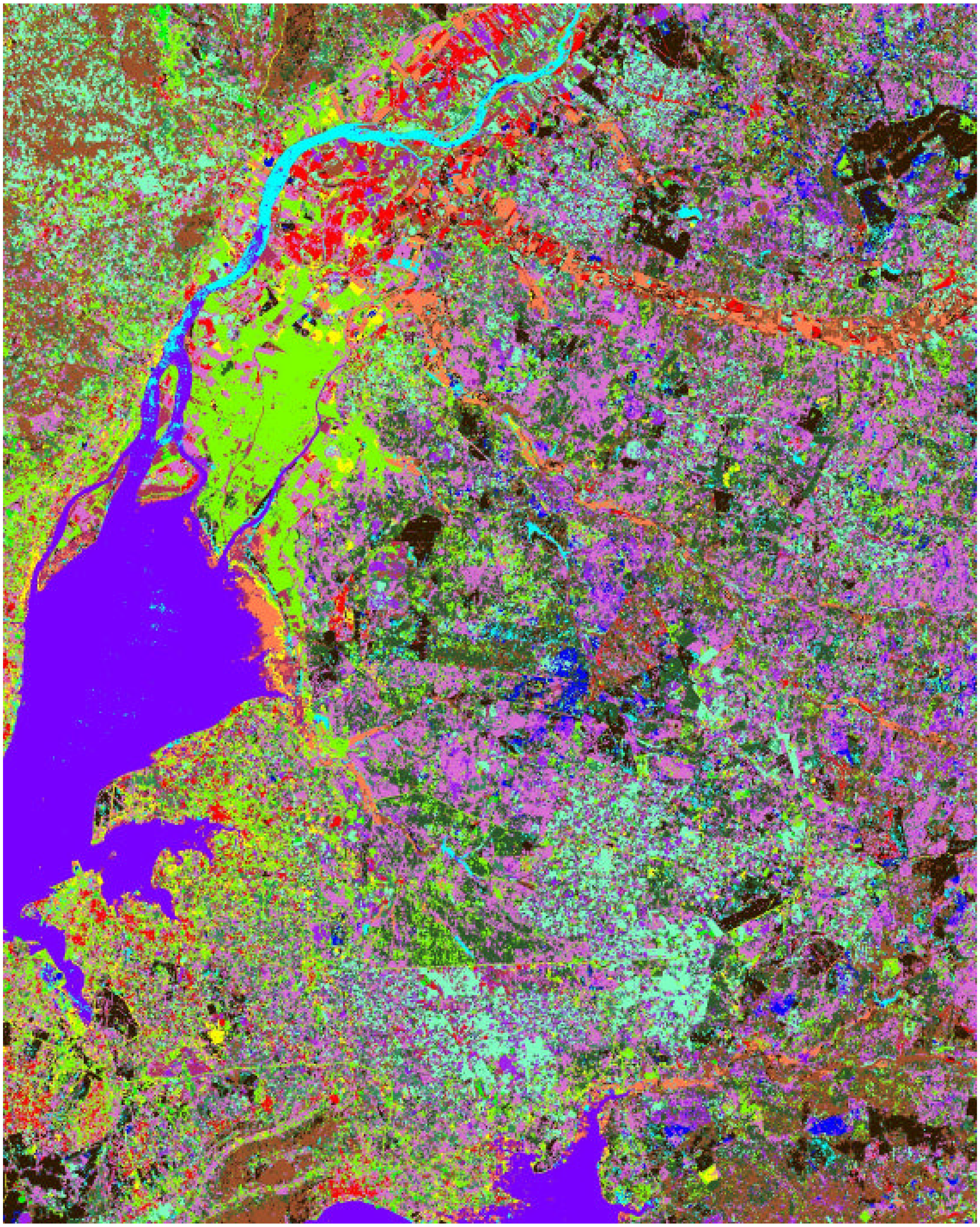

**Figure 2.** Classified satellite image of Lisbon Bay in Portugal

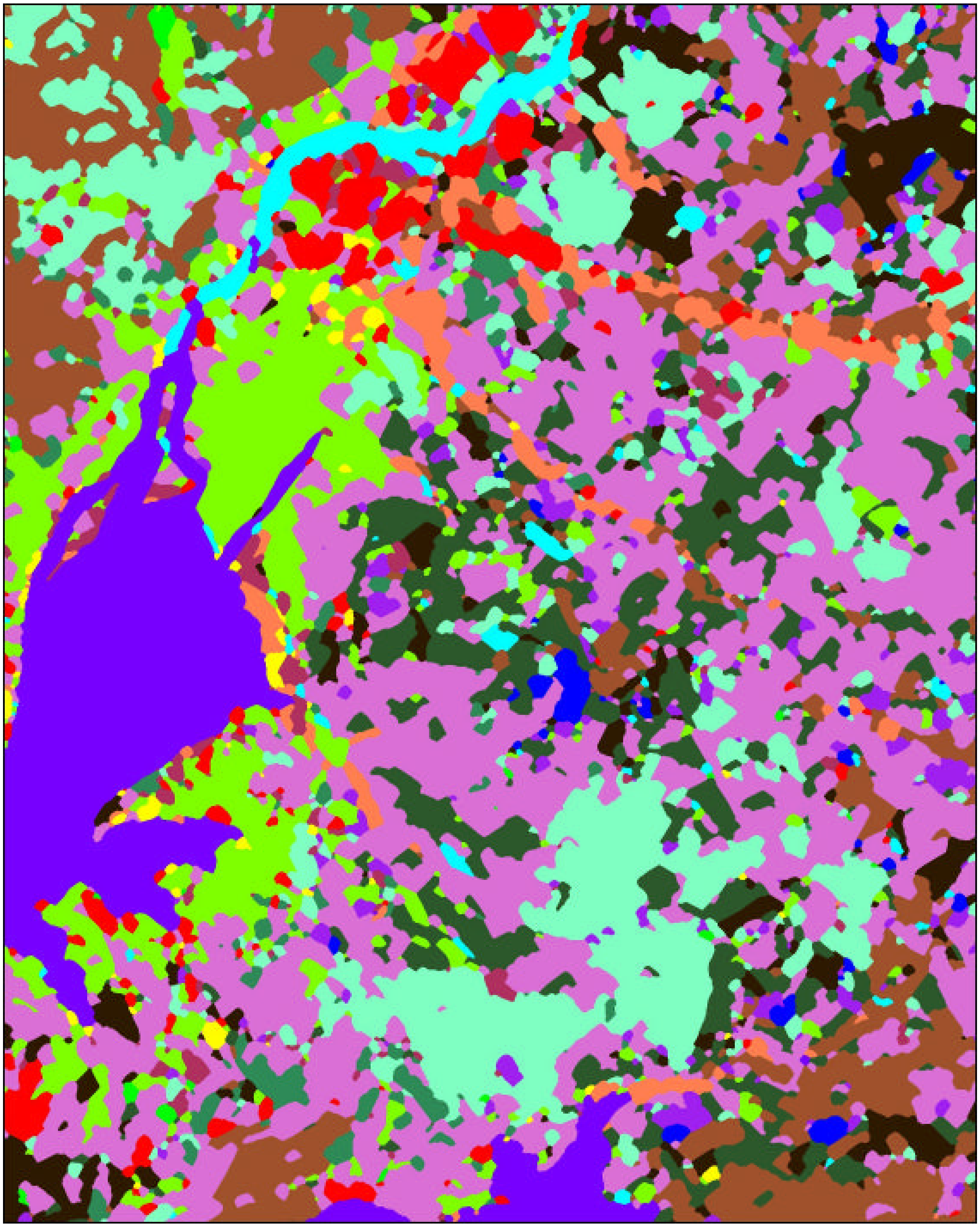

**Figure 3**. Generalized image using an extended Moore neighborhood CA model