Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

SEMIAUTOMATIC URBAN MAP GENERALIZATION USING A RASTER-VECTOR MODEL

UREÑA CÁMARA, M. A. y ARIZA LÓPEZ, F. J.
Grupo de Investigación en Ingeniería Cartográfica.
Dpto. de Ingeniería Cartográfica, Geodésica y Fotogrametría. Universidad de Jaén.
Campus "Las Lagunillas" s/n. 23071. Jaen (Spain). Tel.: +34 953 212 852 / +34 953 212 469
maurena@ujaen.es fjariza@ujaen.es
.

GENERALIZACIÓN SEMIAUTOMÁTICA DE MAPAS URBANOS USANDO UN MODELO RASTER-VECTOR

RESUMEN

En este trabajo se desarrolla un motor de generalización que incluye cuatro tipos de operadores: selección, generalización, algebra de mapas y morfología matemática. Debido al alto nivel de versatilidad de dicho motor, las soluciones que aquí se presentan son solo algunas de entre todas las posibles. El proceso de generalización se ha llevado a cabo para nueve cascos urbanos en el paso de escala de 1/25000 a 1/50000. Los resultados demuestran que el proceso de generalización mostrado reduce el tiempo de edición y la resolución de conflictos en comparación con los actuales procesos de generalización.

Palabras clave: Generalización cartográfica, medidas de similitud de imágenes, cartografía urbana, operadores de generalización.

ABSTRACT

This paper refers to a generalization engine that includes a toolbox containing four sets of operators: selection, generalization, map algebra and mathematical morphology (MM). Due to the high level of versatility of the generalization engine, the processes and solutions presented here are only a few among the many possible. The generalizations are taken from nine urban city block maps when reducing scale from S25k (1/25000 scale map) to S50k (1/50000 scale map). Our results demonstrate that this generalization process reduces the human workload of edition time and conflict solving in comparison with current vector processes.

Keywords: Cartographic generalization, automatic assessment indices, urban city-block maps, generalization operators.

Recibido: 4/3/2008
Aceptada versión definitiva: 15/7/2008

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

1. Introduction

Automatic generalization has been the subject of a great deal of research in cartography. It has been developed by public and private institutions: the National Oceanographic Administration Agency of the USA (Clayton, 1986 and Shea, 1988), the Institute Géographique National of France (Rousseau *et al*., 1994 and Barrault, 1995), the Instituto Geográfico Nacional of Spain (Iribas, 1997; Iribas, 2000), the National Centre of Geographic Analysis of the USA (Jasinski, 1990 and Fico, 1992), the European Organization for Experimental Photogrammetric Research (OEEPE, 1994), the AGENT project (1997-2000) and many others. Nevertheless, significant problems still exist when attempting to change scale by applying an automatic generalization procedure, particularly for urban city-block maps (Iribas, 2004).

Automatic generalization also attempts to solve the problem of subjectivity in manual generalization. This subjectivity is due not only to the lack of a set of "universal" rules of generalization, but also to the cartographic background and personal experience of each operator. Different results can also be obtained whilst working with the same element, even when it is the same person working on it. This is due to the absence of repetition, being each element different. On the other hand, due to the fact that the sequence of operations will always be the same, the digital procedure will always offer the same results when the data and the initial parameters are the same. Nevertheless, cartographic and geographic knowledge, on which manual generalization is based, needs to be incorporated into the digital context. However, there is a relative scarcity of formalized cartographic knowledge. This is one of the most complex difficulties that Cartography faces today.

Urban areas possess three important characteristics: high entity density, feature complexity and a high rate of change both in time and space. Due to these reasons, cartographic processes like automatic updating and generalization are critical when dealing with urban areas, yet because of their complexity, research on automatic generalization of urban zones is not as extensive as research on other topographic maps elements, such as linear features that comprise 80% of them (Ariza, 2002).

There are two main model options for dealing with generalization in digital mapping (Ureña, 2004): vector and raster. The vector model is based on object representation by coordinates. In this vector model, points, lines and areas (polygons) are the units which carry information and support the generalization process. Within the vector perspective, the generalization of urban maps is complex because of both the interdependence of objects and the semantic implications of this interdependence and of the generalization process. Ruas and Mackaness (1997) present the philosophical elements involved in such a generalization. Cuenin (1972) shows a very practical point of view within a manual framework, Powitz (1993) presents an example of how to generalize within an automated framework, and OEEPE (1994) illustrates the problems with commercial software applications.

The raster model is a space-primary data structure (Su *et al*., 1997) representing reality in terms of uniform, regular cells which are usually rectangular or square. In the raster model objects and spatial relationships between them are implicit (Ehlers, 1991). Raster generalization is

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

reasonable as this generalization is caused by a reduction in graphics space when a map scale becomes smaller. Nevertheless, raster generalization has received less attention than vector generalization, and most of the work on it is associated with remote sensing. Without any specific mention of MM, the work of Monmonier (1983) shows one of the first applications of the erosion operator in a raster mode area generalization applied to cartography. For Monmonier, raster generalization is more appropriate than vector for land use or land cover data generalization because the raster model approach promotes the partitioning of thin or insignificant polygons and the growing or linking of others. Ureña and Ariza (2000) present an example of applying MM to the generalization of area features with complex relations (city-blocks).

The generalization process employed by Spanish Mapping Agencies (SMA) is partially achieved using both manual and automatic operators. This process is more time consuming as more urban areas appear on the map. Two of these SMA (Iribas, 2004 and Baella, 2004) described the time needed as being up to 3 days (medium-size city generalization from S25k to S50k) for Instituto Geográfico Nacional and up to 2 days using the same map for Instituto Cartográfico de Cataluña. For this reason we have directed our efforts towards developing a generalization process which seeks a semi-automatic solution. This solution requires less revision on the part of the human operator, reducing the subjectivity and also increasing productivity.

Due to this reason, a complete generalization process based on BCNRV (Raster-Vector Numeric Database) has been developed. The technological bases of the process are shown in Ureña and Ariza (2005a, 2005b and 2005c). The generalization process focused on the change of scale from S25k to S50k, these being the common national scales in Spain. This work is also of great interest to our country because the new national map at S50k scale is currently obtained manually from the S25k national topographic map.

This paper comprises seven main sections: First of all, a general revision of the state of the art is included in order to introduce the scientific framework of this work. The second section presents a new data model called the BCNRV, which combines the strength of both basic models and constitutes the basis of the proposed methodology. Next, our generalization engine is presented. Modern generalization is mainly computational, and we have developed a complete set of generalization operators in a software system. For a complete understanding of the proposed methodology and the relevance of the results of this proposal a complete test on real data is needed. These data are presented in the sixth section. A general overview of the generalization process is included in the seventh section and the analysis process, both analytical and visual, in the next. Finally, main conclusions are stated.

2. State of the art

Recent years have been very fruitful, and many proposals, models, methods and examples dealing with the generalization of urban maps have appeared. In general, we can conclude that vector models have been increasing in complexity, and both vector and raster methods have achieved great advances in generalization operators, particularly in shape simplification and object displacement, but that no method, neither raster nor vector, has achieved a successful solution.

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

Table 1 shows some of the most important vector generalization studies. As it is shown in this table, at first certain elements were characterized and individual and group operators, based on manually-introduced parameters, were developed. Later all the efforts were applied to self-guided processes where parameters are automatically defined based on some of the constraints imposed on the full map. The main advantage of the vector model is the use of minimum description data to represent reality, so the operators can increase the computational complexity order without increasing the total calculation time. The second advantage is the precision achieved with regard to the position and description of the object.

Table 2 shows some of the generalization studies concerned with map generalization. Half of all the generalizations methods and operators described use MM, but the raster generalization methods are less developed than the vector ones, so they remain in operator definition state. The main disadvantage of these methods is the time required because of the number of operations performed on each cell of a raster image. However, these methods are quite robust and easily implemented.

As it is shown in table 1 and table 2, the vector process is more extensively studied than the raster one. This is due to two reasons, the first being that there is a huge number of vector Cartographic Numeric Databases compared to the raster ones. The second reason is that vector processes are less time consuming than raster.

On the other hand, this tendency is gradually being inverted. This may be due to an increasing complexity of generalization operators and processes, as it is shown in the Agent Project (1997-2000) and Brenner and Sester (2005). This increase in complexity is more important in vector methods than in raster ones.

In the way described in the previous paragraph, vector processes are enormously time-consuming (matrix inversions, combinatorial limit in relationships between objects), almost reaching the time needed with raster processes.

On the other hand, the greater precision of the vector process has to be taken into account. This is the reason why some authors propose a reverse transformation from raster to vector. We have succeeded in creating a new hybrid model for cartographic generalization. This is briefly described in the following section.

3. The raster-vector numeric database (BCNRV)

In order to achieve the goals of this generalization process, we have developed a model (Ureña, 2004) which uses the object paradigm, as explained by Buttenfield (1995), where each feature or relevant aspect of a phenomenon is a predefined object. We have named our proposal the BCNRV and it has the following characteristics:

- Each feature of interest of the real world is represented by one instance of a BCNRV object.
- Relationships between objects are stored as instances of BCNRV objects. In this way all topological relationships are explicitly shown in the model.

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

- Each BCNRV object is an isolated representation in its storage. This means that even if an object originally has a vector representation, it can be shown as a raster representation or as any other implemented option; and conversely, if the original data is raster, a vector representation can be used, so we have three representation models:
    - Raster representation model: Considered in order to take advantage of efficient raster computations such as morphological operators, Manhattan distances, etc. needed for the development of automatic generalization processes. Our raster model allows multiple data to be stored in one cell using key values with reference to a table register.
    - Vector representation model: Considered in order to take advantage of efficient vector modelling and computations such as network analysis, topological relations, etc. needed for the development of automatic generalization processes. The vector model allows us to use three-dimensional positioning with more precision.
    - Inherited representation model: This structure only refers to the representation models inherited from other objects by accessing the original data.
- Each explicit representation model has its own analysis operators independent of the attribute operators.

These characteristics confer great flexibility for visualization, query and change in the attributes of an instance. Furthermore, this structure provides a mechanism for storing multiple original data representations, with or without different resolutions, and the applicable operators.

4. The generalization engine

This section is concerned with our generalization engine. The engine implements selection and generalization operators which are then applied to a generalization process. This process involves gestalt selection among urban features (city-blocks/buildings), and this can only be achieved using the BCNRV briefly described in section 3.

Multiple operators have been developed and integrated into the system called the generalization engine. This is a workbench whose objective is to unify and simplify the testing process. We decided to use two different methods to implement this workbench:

1. Interactive method: User selects each operation and its parameters and runs operations one by one.
2. Virtual machine method: User programs all operators and parameters in a plain language and runs this program (this operation can be used in normal run mode or in debug mode).

The generalization engine has four operator classes that define a toolset, which allows the simulation of many cartographic operators:

1. Selection operators: Basic attribute selections for intervals and for a predefined number of objects have been developed. Thus, some spatial criteria have been added in order to extend selections to nearby objects.
2. Generalization operators: In the following section of this article we briefly describe the generalization operators which are considered to be the most important.

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

3. Map algebra operators: All the algebraic operators (addition, multiplication, and so on) have been included so that they can be applied to raster images.
4. MM operators: A set of morphological operators (dilation, erosion, complementary, opening and closing) can be applied automatically, and the decision for defining the origins of morphological kernels (for displacement operators) is left to the generalization module.

4.1. Generalization operators

Multiple generalization operators in the raster model have been developed to simulate those described by McMaster and Shea (1992):

- Amalgamation operator: Two amalgamation operators are implemented:
  1. Fusion of sides with distances less than a certain tolerance.
  2. Morphologic closing.

  There is also a special amalgamation operator which is applied to city zoning, the joining of objects and the extraction of objects of a given zone. The algorithm joins objects using distances, which is similar to vector amalgamation. The algorithm calculates all distances from perimeter pixels of one object to another, both of them in the same mesoblock. If this distance is less than the tolerance, then both pixels are joined. We call this solution an object extension (example shown in figure 1).
- Quasi-convex hull operator: This is an extreme case of the previous operator: the amalgamation operator is applied many times in order to achieve the quasi-convex hull (example shown in figure 2). If a set of objects is selected previously it is possible to create a segmentation or zoning of the city. This operator develops a shape simplification algorithm when applied to a single object.
- Erasing operator: Erases selected objects. The difference between this operator and other erasing operators is that they handle different objects. Therefore, if a street is erased here, two city-blocks/buildings are amalgamated and their topological relations erased. This implementation is based on the BCNRV structure proposed in Ureña and Ariza (2005b).
- Displacement operator: This operator has two approximations:
  1. MM approximation.
  2. Basic approximation: A composition of erasing from the old position and rebuilding into the new position is used.
- Street rebuilding operator: This algorithm is exclusively raster. It is very similar to the paint option of an image editor (like paint brush). The size of the brush can be controlled so that we can reconstruct different widths of streets.

5. Original data

Our research is mainly concerned with the development of a generalization process between S25k and S50k maps. For this reason the selected cases (city maps) are published at both scales: data at S25k scale as input data to the process and data at S50k for controlling the process. For a more in-depth investigation we selected maps corresponding to cities with different urban typologies, structures (planned and unplanned) and sizes (large, medium and small). This

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

heterogeneous selection allowed us to test our procedure with more accuracy, but it was not always possible to use the same source for both scales and also for all cases. Therefore, the original data for generalizing was selected from different cartographic products generated by Spanish agencies: Instituto Geográfico Nacional (IGN), Servicio Geográfico del Ejército (SGE), and Dirección General del Catastro (DGC). Table 3 shows an overview of their shape and other characteristics of interest for this work.

As it is shown in table 3, all Spanish urban maps represent city-blocks (or urban islands as described in Regnauld, 1998), contrary to European urban maps which represent buildings. So in this work we are dealing with the generalization of meso-objects and their relationships, instead of micro-objects (buildings).

6. The generalization process: applying the generalization engine

The generalization processes that can be applied using the previously described toolset are too numerous to be mentioned here. However, the generalization engine needs a guide in order to achieve its goals, and for this reason a complete process for the adjustment of operators and parameters, including expert revision of results, has been developed.

Here it is of great relevance to keep in mind the cartographer's thought processes when generalizing a city map. The laws of perception are one of the bases that control the generalization process. For this reason they is an important guide for the selection of objects (Thórisson, 1994; Rome, 2001; Anders, 2003).

The experience of the cartographer is very difficult to implement in an automatic generalization model. However, we have included part of this experience by means of a three-fold reviewing process. Each revision produces feedback which modifies operators/parameters of the generalization process.

Due to the above reason we consider that the result is a semiautomatic process, as shown in figure 3, which can reduce subjectivity when generalizing. Despite this reduction of subjectivity, an automatic solution can not be achieved, so, some expert edition is needed in order to obtain a cartographic solution. This process can be refined for each city and stored in the model as additional data or as a process script.

The generalization process shown in figure 3 can be subdivided into the following phases:

1. Selection of the city-blocks that need to be generalized and the groups of these city-blocks using gestalt laws and multiple parameters (area, orientation, etc).
2. Amalgamation of the selected objects in phase 1 using a distance parameter.
3. Refinement of the generalization of phase 2:
    a. Reconstruction of streets based on their level of importance and connectivity indices of the city.
    b. Simplification of city-block boundary using Serra's Number and shortest distance tolerance.
    c. Filling in of all inner courtyards of the city using minimum area criteria.

   d. Erasing of all city-blocks of an area smaller than a certain threshold.
   e. Erasing of all city-blocks presenting less than a predefined density.
4. Unifying the solutions from all previous phases and constraining the whole city to its quasi-convex hull envelope.

This generalization process has been implemented for both parallel and serial applications. We have determined that the critical phase is that of selection. We have therefore tested the selection phase using different object attributes, spatial distributions and individual object selections before proceeding to develop our first generalization process. The gestalt laws have been used to determine the similarities between objects, mainly city-blocks. Our results (see Ureña and Ariza, 2005a) have shown that these laws associate objects in the same way a cartographer would do during the generalization process.

This whole generalization process (figure 3) is an interpretation of the results of the previous work of the pool of experts (Ureña and Ariza, 2005a). This interpretation was applied to the operators/parameters in the BCNRV on three different occasions. So part of the experts' and authors' knowledge has been included in this process, yet even more knowledge is needed if we hope to attain the generalization level of the cartographer. For this reason, this knowledge must be embedded within the generalization engine. However, the current process is incomplete and thus needs the later modifications of the solutions by a cartographer (a semiautomatic process).

Despite the continuing need for involvement of the cartographer, our generalization increases the overall efficiency of the map production process. This is due to a reduction of time, for example less than a minute is required to obtain a S50k map of the city of Jaén from its original BCNRV (medium-size city generalization from S25k to S50k). The required edition time, after the generalized map is obtained, is about four hours in order to resolve 133 conflicts (see Table 6). Thus the complete generalization process takes about half a working day. However, one disadvantage of this generalization process is the time consumed in determining all the BCNRV objects and their later enrichment:- about three hours for the same city. But this pre-processing can be carried out in batch mode or as a background process.

7. Analysis

The results of the semi-automatic generalization process have been evaluated using two methods:

1. Visual analysis: An independent panel of experts evaluates the results. As it has been previously described, this process has been used to adjust the generalization engine (Ureña and Ariza, 2005a). Thus visual analysis has been used in two different ways: first for enhancing all the solutions in a feedback process, and second for evaluating results.
2. Analytical assessment analysis. By means of analytical indices some measures have been obtained and compared to experts' indications.

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

7.1. Visual analysis (evaluation by the experts)

The quality of this analysis depends on three factors: i) the size of the group of experts, ii) the heterogeneity in the cartographic background and mapmaking experience within the expert group, and iii) the inquiry system and its methodology.

For the first factor it is obvious that the larger the group is, the greater the representativeness and statistical soundness of the results are. Nevertheless, the possible number of experts depends upon what it is considered to be an expert, and that question is very difficult to answer. We use a more restricted idea of "expert", selecting people from mapping agencies and from university departments related to cartographic sciences. Even so, there is a great heterogeneity here which once again reveals the subjectivity of the generalization process. For the inquiry system we selected the poll system, using the experience of Ruas (2001). The objectives of the evaluation were to derive this information:

1. The rate of similarity among experts for the best generalized maps.
2. The rate of agreement of each expert for each map.
3. The conflicts found in each map and their proposed solution.
4. The similarity of the conflicts encountered by the experts.

With these four ideas in mind we designed a poll asking for opinions on:

1. The degree of global satisfaction with the result.
2. The number of conflicts. The experts were asked to indicate these visually by circling the conflictive areas, coded using the method developed in Ruas (2001). The table of codes is shown in table 4.

The number of experts polled was 26 (more than many other international studies). Each city was reviewed by 8 or 9 experts and the response rate was 65%.

7.2. Analytical assessment

The main idea is to analytically derive a set of indices or measures in order to automatically evaluate the results of a generalization process. These indices can be used to evaluate a characteristic of a single image, for instance its mean value, but also for comparing two or three images. These measures have been automatically applied to the generalized map, the same one evaluated by the experts, and to the original S25k map and to the S50k official generalized version. Of course these official solutions should not be considered as optimum and unique, but we believe they are sufficiently accurate. Following our previous studies (Ureña *et al*., 2001), we have used the set of measures that appear in Table 5. The first six have been adapted from various authors, such as Kullback (1959) and Kerstész *et al*. (1995). The others have been developed specifically for generalization assessment.

As it is shown in table 5, the measures are divided into three classes, based on their application:

1. Index value increased between two images: The two images are the manually generalized S50k and the automatically generalized S25k.

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

2. Two-image indices: The images used are the manually generalized S50k and the automatically generalized S25k.
3. Three-image indices: The images used are the S50k as the correctly classified image, the automatic generalized S50k as the classified image and the S25k original image resampled by a factor of 1/2 as another classified image.

Each individual measure is insufficient to define the best-generalized map, but when combining all of them there is a high rate of similarity with the best map selected by the experts (see the analysis shown in the next section). The method we have used to achieve this is detailed below:

1. Determining all the values for each index and each generalization.
2. Comparing all values to the best values in table 5 and marking the most similar one among all generalized maps for each index.
3. Selecting the modal result that has more marked indices than the others as the best generalized map.
4. In such a way the best solution can be determined using the following equation:

$$Best\ Solution = x \mid \forall x, y \in \{Automatically\ Generalized\ Map\ for\ a\ city\},\ Sum(x) \geq Sum(y)$$

where the function Sum is defined by:

$$Sum(x) = \sum_{i=1}^{number\ of\ indexes} j \mid \forall k_j \in \{abs(values\ for\ index\ j - best\ value)\} \wedge j = Int\left(1 - \frac{j - \min(k_j)}{\max(k_j) - \min(k_j)}\right)$$

7.3. Analysis of the results

In this section we describe the main results indicated by the expert group and a comparison with those indicated by the analytical measures. Table 6 shows some data of the evaluation of the experts. This evaluation shows the type (first modal and second modal) and number of conflicts detected by the experts, both for each type and globally, and some examples of individual conflicts are shown in table 7.

The number of conflicts shown in table 6 indicates that more work is needed in order to achieve a semiautomatic state. However, the information acquired has been implemented in the generalization engine (Ureña and Ariza, 2005c) and these conflicts dramatically reduced. Thus, our engine has the capability to review and modify the result interactively. From table 6 we deduce that:

- Larger-size cities (Almería, Burgos or Bilbao) need a high generalization level.
- Smaller-size cities (Jaén, La Carolina, Logroño or Majadahonda) need a medium generalization level. Logroño is an exception to this rule.
- Planned cities with wide avenues (Pamplona) need the lowest generalization level.
- Generally, selection for the generalization process is best achieved using a perimeter selection. However, cities with many regular-shaped city-blocks achieve the selection via vertices.
- Conflicts 131 (see table 7) show us the need to improve the shape simplification algorithm.
- Conflicts 440 show the need for a higher exaggeration of streets (20% is sufficient).

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

The previous information is complemented with the experts' remarks. From all the remarks we highlight these (see table 7 for individual examples of more frequent conflicts as it was globally described in table 6):

- In general, the solutions are acceptable for visualization purposes. The current solutions can be used in web map servers.
- There is an extensive loss of spatial structure which could be reviewed by a cartographer.
- The solutions show a priority of city-blocks as opposed to other parts of the city.

In figure 4 we can see an example of the last two remarks. We have confirmed these ideas by locally calculating Urban Density with a maximum distance of 250 meters. Our tests show that the density of the generalization map ranges between -25% and 8% compared to the urban density of the original S25k map. This great percentage variation is due to the extensive loss of spatial structure indicated in the second experts' remark, while the high negative percentage (-25%) indicates a predominance of city-blocks over the other BCNRV objects. This leads us to a new assessment measure in order to improve detection of meso-blocks and guide the generalization process.

The main idea of these tests is that experts prefer the minimum urban density modification in order to obtain the best-generalized map, this being in general a -5% urban density.

On the other hand, we have compared results selected by the experts with results selected using assessment measures. This comparison is shown in table 8, allowing us to test the accuracy of our measures. However, we know that experts provided higher quantity and quality of information than the one that can be obtained using automatic measures. Our tests have shown that the assessment measures have these characteristics:

- Binary entropy, Tone and Ratio Urban/Non-urban are redundant, thus only one of them must be calculated. We advise the use of Binary Entropy.
- Percentage of Agreement and Kullback's Divergence are redundant as well. We advise using Kullback's Divergence because it is more reliable.
- Improvement and Deterioration indices have the same meaning as Correct Similarity and Incorrect Similarity.

As it is shown in table 8, six of the nine results are equal to those selected by the experts (7 if the second result for Jaen is considered). For this reason the set of our assessment measures are considered as a modal expert, with the advantage of efficiency (only a few seconds) and the disadvantage of the lack of information on the number of conflicts and their spatial distribution. This information is obtained by using local urban density in detecting conflictive zones of the city. However, some generalized maps like the ones of Bilbao and La Carolina do not coincide with the experts' decision. Nevertheless we conclude from the information shown in table 8 that if more than 60% of measures indicate only one result, then this result is the same as the experts'.

There is another interesting idea that can be obtained from table 8. This idea is that the best generalization map is not necessarily the map with least conflicts, contrary to what common sense would dictate. This shows us that the main issue involved in generalization is not the elimination of conflict in the generalized map as much as the general visual impact of the map or the

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

harmonization of the map as a whole entity. This part of the generalization process is the most difficult and at present it can only be achieved manually.

8. Conclusion

A new approach to cartographic generalization of urban maps has been presented. Our methodology is based on a mixed raster-vector data model, a generalization engine capability, a revised generalization process and the analysis and evaluation of results by means of visual and analytical measurements.

Recent studies on urban map generalization show two main tendencies: ever more complex vector options, and the increasing interest in raster possibilities. For this reason we have developed a mixed raster data model we call BCNRV, which uses the object paradigm to provide a mechanism for storing multiple original data representations, taking into account the advantages of both basic models.

The generalization engine has four sets of operators, the most complex to implement being amalgamation and shape simplification. The reviews indicate that morphological amalgamation is the best operator. However, amalgamation tolerance depends mainly on urban density, so it has been tested using local algorithms in order to obtain the differences between original data, and generalized data and to explain the experts' best choice.

The generalization process we have developed has its critical phase in the selection. Our results have shown that the gestalt laws associate objects in the same way a cartographer would do during the generalization process. However, the current process is incomplete and therefore it needs later reviews of the solutions by a cartographer (a semiautomatic process).

The reviews achieved by the expert group have defined the degree of generalization by means of the size and typology of city as it is deduced from table 6 in the analysis of the results of section 7.1.

On the other hand, as shown in table 8, results show that up to 78% of our map generalizations are equal to the ones selected by the experts. For this reason, the assessment measures have great importance in selecting the best generalized map. Furthermore, we conclude that if more than 60% of our assessment measures agree, then the solution selected by the measures is equal to that of the experts'.

The process shows three differences in comparison with other processes:

1. The use of raster and vector data under one data structure which generalises different parts of built-up areas.
2. Verification using visual analysis (polling people not related to the process) and automatic assessment measures.
3. The use of a high number of built-up areas which improves the reliability of the results.

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

The analyses in this paper, using various visual and analytic methods, show that our generalization process goes a long way toward achieving a semi-automatic generalization level that requires very few changes on the part of the cartographer. Our research is now centred on the use of assessment measures to guide the process for each zone in a city, always using the BCNRV conceptual model. Furthermore, with the use of local urban density indices we are now in possession of an automatic system for detecting the quality of our generalization process. This detection system is the key to developing a generalization process which is completely automatic (in all phases) and meets all S.M.A. requirements.

References

Anders, K. H. (2003): "A hierarchical graph-clustering approach to find groups of objects", Working Group on Map Generalization, International Cartographic Association. Paris, France.
Ariza, F. J. (2002): *Calidad en la producción cartográfica*. Madrid (Spain), Ra-Ma.
Bader, M. (2001): *Energy minimization methods for feature displacement in map generalization*. University of Zurich, Switzerland. Thesis.
Baella, B. (2004): *Personal communication*. Chief in charge of Automatic Cartography Area. Institut Cartográfic de Cataluña. Barcelona (Spain). February 2004.
Barrault, M. (1995): *Étude de la courbure en géométrie discrète*. IGN, Paris (France). Available as: /ign/cogit/Generalisation/Techn_Reports/courbure_polyligne.fr by ftp anonymous on sturm.ign.fr (192.33.147.11).
Brenner, C. and Sester, M. (2005): "Cartographic generalization using primitives and constraints", in *Proc. 22nd International Cartographic Conference*. A Coruña (Spain).
Buttenfield, B. (1995): "Object-oriented map generalization: modelling and cartographic considerations" in Müller, J.C., Lagrange, J.P. and Weibel, R. (Ed.): *GIS and generalization - Methodology and Practice. GISDATA 1*, London (U.K.), Taylor & Francis, pp. 91-105.
Clayton, V. H. (1986): *Cartographic generalization: a review of feature simplification and systematic point elimination algorithms*. Washington (U.S.), NOAA, U.S. Department of Commerce, NOAA Technical Report; 112 Charting and Geodetic Services series; 5.
Cuenin, R. (1972): *Cartographie générale. Tome 1. Notions générales et principes d'élaborations*. Paris (France), Eyrolles.
Ehlers, M. (1991): "Integration of remote sensing and GIS: data and data access", *Photogrammetric Engineering and Remote Sensing*, 55, 11, pp. 1619-1627.
Fico, F. J. (1992): *Opportunities for generalization in the Digital Chart of the World*. University California, Santa Barbara, CA. NCGIA Technical Report, 92-2.
Hangouët, J. F. (1998): *Approche et méthodes pour l'automatisation de la généralisation cartographic; application en bord de ville*. University of Marne La Vallée, France. Thesis.
Huilian, W.; Fang, W.; Hongyan, D.; Haizhong, Q.; Qi, H.; Linlin, Z. (2005): "The application of mathematical morphology and pattern recognition to building polygon", in Proc. 22nd International Cartographic Conference. A Coruña (Spain).
Iribas, F. (1997): "Obtención del MTN50 por generalización cartográfica del MTN25 Digital", *Mapping*, 38, pp. 22-29.
Iribas, F. (2000): *Diseño y elaboración de un procedimiento interactivo de obtención del mapa topográfico nacional a escala 1:50000 por generalización cartográfica del MTN 1:25000*. University of Cantabria (Spain). Thesis.

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

Iribas, F. (2004): *Personal communication*. Jefe del Servicio Regional del Instituto Geográfico Nacional en La Rioja. Instituto Geográfico Nacional, Ministerio de Fomento, La Rioja, Spain. February 2004.
Jasinski M. (1990): *The comparison of complexity measures for cartographic lines.* University of California, Santa Barbara. NCGIA Technical Report, 90-1.
Jones, Ch.; Bunday, G.; Ware, J. (1995): "Map generalization with a Triangulated Data Structure", *Cartography and Geographic Information Systems*, 22, 4, pp. 317-331.
Kertész, M.; Csillag, F.; Kummert, A. (1995): "Optimal tiling of heterogeneous images", *International Journal of Remote Sensing*, 16, 8, pp. 1397-1415.
Kullback, S. (1959): *Information theory and statistics*. New York (U.S.A.), Wiley.
Mackaness, W. (1995): "Analysis of Urban Road Networks to Support Cartographic Generalization", *Cartography and Geographic Information Systems*, 22, 4, pp. 306-316.
McMaster, R. and Shea, K. (1992): *Generalization in digital cartography*. Washington D.C. (U.S.A.), Association of American Geographers.
Monmonier, M. (1983): "Raster-mode area generalization for land use and land cover maps", *Cartographica*, 20, 4, pp. 65-91.
OEEPE (1994): *Interim Report May 1994*. Working Group on Generalization.
Powitz, B. (1993): *Zur Automatisierung der kartographishen generalisierung topographischen daten in geo-informations systemen*. University of Hannover (Germany). Nr. 185. Thesis.
Regnauld, N. (1998) : *Généralisation du bâti: Structure spatiale de type graphe et représentation cartographique*. University of Marne La Vallée (France). Thesis.
Rome, E. (2001): "Simultating perceptual clustering by gestalt principles", in *25th Workshop of the Austrian Association for Pattern Recognition*, ÖAGM/AAPR 2001. Berchtesgade, Germany, June 7-8.
Rousseau, D.; Rousseau, T.; Lecordix, F. (1994): *Expertise du map generalizer: logiciel de généralisation interactive d´Intergraph*. France, IGN-France.
Ruas, A. (1999): *Modèle de généralisation de données géographiques à base de contraintes et d'autonomie*. University of Marne La Vallée (France). Thesis.
Ruas, A. (2001): *Automatic Generalisation Project: Learning Process from Interactive Generalisation*. European Organization for Experimental Photogrammetric Research (OEEPE).Official Publication nº 39.
Ruas, A.; Mackaness, W. (1997): "Strategies for urban map generalization", in *ICA-Stockholm Workshop*, pp. 1387-1394.
Schylberg, L. (1993): *Computational methods for generalization of cartographic data in a raster environment*. Royal Institute of Technology of Stockholm (Sweden). Thesis.
Sester, M. (2000): "Generalization based on least squares adjustment", *International Archives of Photogrammetry and Remote Sensing*, 23, B4, pp. 931- 938.
Shea, S. K. (1988): *Cartographic Generalization*. Washington (U.S.A.), NOAA: U.S. Department of Commerce, NOAA Technical Report; 127 CGS; 5.
Su, B.; Li, Z.; Lodwick, G. (1996a): "Algebraic models for the elimination of area features in digital map generalization", in *Proc. Mapping Sciences '94*, pp. 297-307.
Su, B.; Li, Z.; Lodwick, G. (1996b): "Building mathematical models for the generalization of spatial data", in *Proc. Geoinformatics '96*, Wuhan Technical University of Technology. 1, pp. 417-424.
Su, B.; Li, Z.; Lodwick, G. (1997): "Morphological transformation for the elimination of area features in digital map generalization", *Cartography*, 26, 2, pp. 23-30.

Su, B.; Li, Z.; Lodwick, G. (1998a): "A morphological model for partial collapse of area features in generalization of spatial data", in *Proc. 8th International Symposium on Spatial Data Handling '98*. Vancouver, BC (Canada). IGU Geographic Information Science Group, pp. 690-700.

Su, B.; Li, Z.; Lodwick, G. (1998b): "Algebraic models for collapse operation in digital map generalization using morphological operators", *Geoinformatica*, 2, 4, pp. 359-382.

Thórisson, K. (1994): "Simulated Perceptual Grouping: An Application to Human-Computer Interaction", in *Proc. of the 16th Annual Conference of the Cognitive Science Society*. pp. 876 – 881.

Ureña, M. A.; Ariza, F. J. (2000): "Mathematical Morphology Applied to Raster Generalization of Urban City Block Maps", *Cartographica*, 37, 1, pp. 33 – 48.

Ureña, M. A.; García, J. L.; Reinoso, J. F.; Ariza, F. J. (2001): "Índices de Evaluación para Generalización Cartográfica de Cascos Urbanos en Raster", in *Proc. XIII Int. Congress of Graphic Engineering*. Badajoz (Spain).

Ureña, M. A. (2004): "Generalización de cascos urbanos en formato raster". University of Jaén (Spain). Thesis.

Ureña, M. A. and Ariza, F. J. (2005a): "Generalization of urban city-block (built-up areas) maps in raster-vector model", in Proc. 22nd International Cartographic Conference. A Coruña (Spain).

Ureña, M. A. and Ariza, F. J. (2005b): "Raster-Vector cartographic numeric database", in *Proc. 22nd International Cartographic Conference*. A Coruña (Spain).

Ureña, M. A. and Ariza, F. J. (2005c): "Generalization process for urban city-block maps", in *Proc. 22nd International Cartographic Conference*. A Coruña (Spain).

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

TABLES

Table 1. Some vector generalization studies ordered by year.

Year	Author(s)	Issue
1995	Mackaness	Define cities as a set of relations in the transportation network (streets), urban morphology and spatial distribution of urban use. The structure of the city is constrained, but also constraining, because of the interaction of these elements.
1995	Jones, Bunday and Ware	Reality is divided using a Constrained Delaunay Triangulation (CDT). The CDT is used to implement simplification, collapse, exaggeration and amalgamation operators.
1998	Hangouët	Categorical generalization: Conversion to equal-area rectangles and a spatial redistribution of these objects (self-generated lineal redistribution taking into account orientations). The process also includes elimination operators if conflict arises.
1998	Regnauld	Structuring operator: Reduces conflicts inside an urban island (city-block) using a displacement algorithm. Amalgamation operator: Using displacement of a small object towards an object of greater area or filling of gaps among buildings; here convex hulls of objects are used.
1999	Ruas	Definition of a three-level hierarchy of objects: micro, meso, and macro. Generalization method (subsequently inherited by Agent Project) with four phases: Enrichment of the objects, testing of some their attributes against map specifications, automatic proposal of at least one algorithm to solve conflicts, repetition from the beginning. The previous sequence is applied to all micro objects using the constraints of the macro level (self-guided generalization).
2000	Sester	Simplification of buildings: Reduces the perimeter detail based on maximum and minimum side distance. Displacement operator: Based on shape parameters, the minimum and critical distances and position. These parameters are minimized using a least square adjustment.
2000	Agent Project	Definition of an agent: a member or something that can solve a problem using teamwork. All the agents are micro, meso or macro (Ruas, 1999). Macro agents trigger the activity of micro agents. When the triggered activity returns to its caller the map specifications are reached.
2001	Bader	Energy minimization methods are used for feature displacement (lines and buildings).
2005	Brenner and Sester	Definition of primitives (geometry, constraints and discrete behaviour) in order to achieve cartographic generalization. After each edition (generalization process or revision) the primitives find an overall solution which minimizes certain criteria while enforcing strict constraints simultaneously.

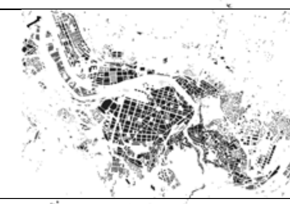
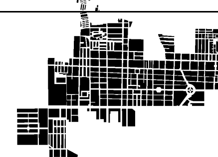
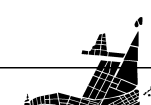
Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

Table 2. Some raster generalization studies ordered by year.

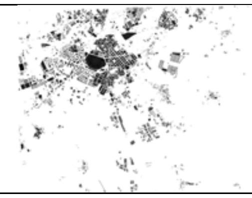
Year	Author(s)	Issue
1983	Monmonier	One of the first studies that uses a raster model and is justified by its greater efficiency for semi-automatic areal generalization. Monmonier proposes an algorithm which maintains multiple classes by reducing the number of pixels. This reduction is based on a weighted average of the minicells (sets of pixels).
1993	Schylberg	Amalgamation: A simplified version of aggregation only applied to area features. Simplification: An operator that reduces details of area features. Elimination: Algorithm to delete small or unsuitable features for the purpose of mapping.
1996a, 1996b, 1997, 1998a, 1998b	Su, Li and Lodwick	Generalization based on MM. Elimination and aggregation (amalgamation) of areas by using opening and closing operations with circular kernels whose size depends on scale reduction.
2005	Huilian *et al.*	Preserve shape characteristic boundary and the same area after generalization. The generalization is carried out in four steps: MM generalization using two templates and three directions, analysing Freeman chain code to determine patterns, simplifying these patterns and finally a reversed coordinate rotating algorithm to obtain a vector representation.

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

Table 3. Original data used in the generalization process.

City Name	Scale	Source Productor (1)	Digital	Area (Hm^2)	Structure (2)	Representation (resampled from original data)
Almería	S25k (S2k)	DGC	✓	928	Mixed	
Bilbao	S25k	IGN	✓	1156	Mixed	
	S50k					
Burgos	S25k	IGN	✓	1007	Mixed	
	S50k					
Jaén	S25k	IGN	✓	374	Mixed	
	S50k	SGE	✗			
La Carolina	S25k (S2k)	DGC	✗	124	Chessboard	
	S50k	IGN				
Logroño	S25k	IGN	✓	977	Designed	
	S50k					
Majadahonda	S25k	IGN	✓	316	Designed	
	S50k					
Mengibar	S25k (S2k)	DGC	✗	172	Nutted	

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

City Name	Scale	Source Productor (1)	Digital	Area (Hm^2)	Structure (2)	Representation (resampled from original data)
	S50k	IGN				
Pamplona	S25k	IGN	✓	1320	Designed	
	S50k					

Notes: 1) IGN = Instituto Geográfico Nacional de España; SGE = Servicio Geográfico del Ejército; DGC = Dirección General del Catastro.
2) Chessboard = very regular and squared morphology; Designed = a designed morphology; Nutted = chaotic morphology, mainly characteristic of Middle Age cities; Mixed = complex morphology with districts belonging to different structure.

Table 4. Table codes for conflict reviewing. The codes are selected from Ruas (2001). Some examples of these conflicts are shown in Table 7.

Feature Type	Code	Description
Isolated Polygon	120	Too thin
	121	Too small
	131	Loss of shape characteristic
Set of Polygons	421	Loss of relative building size
	440	Polygons too close
	462	Polygon and hole erroneous
	464	Loss of polygon spatial structure (gestalt)
	480	Over density of polygons
	481	Homogeneity of density erroneous

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

Table 5. Assessment measures used in analysis. Next to the name of the measure the acronym we use to refer to it has been added.

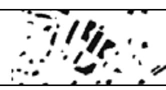
Measure		Range	Best Value	Definition
Description	**Name**			
Index value increase between two images	Moran Index I (**M.I.**)	[-2,2]	0	Spatial autocorrelation of the image
	Binary Entropy (**B.E.**)	[-1,1]	0	Minimum number of storage bits without loss of data
	Tone (**T.**)	[-1,1]	0	Mean value of the image
	Urban/Non-Urban (**R.U.N.**)	[-∞,+∞]	0	Ratio urban non-urban pixel
Two-image indices	Percentage of Agreement (**P.A.**)	[0,1]	1	Percentage of correctly classified pixels
	Kullback's Divergence (**K.D.**)	[0,1]	0	Degree of similarity between two images
	Self-correlation Divergence (**S.D.**)	[0,1]	0	Extension of previous measure with the autocorrelation characteristic
Three-image indices	Improvement (**3.I.**)	[0,1]	1	Correctly classified pixels minus Previous correctly classified pixels
	Deterioration (**3.D.**)	[0,1]	0	Incorrectly classified pixels minus Previous incorrectly classified pixels
	Correct Similarity (**3.C.S.**)	[0,1]	1	Correctly classified pixels in the three images
	Incorrect Similarity (**3.I.S.**)	[0,1]	0	Incorrectly classified pixels in the three images

Table 6. Best generalized urban city-block maps from the expert selection. For each map, the conflict that appears in most zones and the next (second model) is shown. In Table 4 the description for each code is shown. An example of these conflicts is shown in Table 7.

City	Best Generalization		First modal conflict		Second modal conflict		Total conflicts in the selected map
	Selection Attributes (Plus Orientation)	Degree of Generalization (H=High, M=Medium, L=Low)	Code	Number of conflicts	Code	Number of conflicts	
Almería	Area	H	464	19	440	19	96
Bilbao	Perimeter	H	121	56	464	28	165
Burgos	Vertices	H	464	52	121	44	220
Jaén	Perimeter	M	464	34	121	36	133
La Carolina	Vertices	M	464	30	440	9	64
Logroño	Perimeter	M	131	20	12	19	93
Majadahonda	Vertices	L	121	19	464	6	68
Mengíbar	Perimeter	H	464	23	440	7	51
Pamplona	Perimeter	L	121	21	464	12	71

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

Table 7. Experts' most indicated conflict for the best result of each city.

City	Original data	Conflict code			
		121	131	440	464
Almería					
Bilbao					
Burgos					
Jaén					
La Carolina					
Logroño					
Majadahonda					
Mengíbar					
Pamplona					

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

Table 8. Results of each assessment measures for each selected map and comparison between the best results polled by the expert group and the best results calculated using assessment measures.

City	Map		Δ(1 Image)				2 images			3 images				Measures	Experts'	
	Selection Attributes (Plus Orientation)	Degree of Generalization	M.I.I	B.E.	T.	R.U.N.	P.A.	D.K.	S.D.	3.I.	3.D.	3.C.S.	3.I.S.	Best map	Best map	Map with less conflict
Almería	Area	M														
	Perimeter	M														
	Perimeter	H														
	Area	H														
Bilbao	Perimeter	M														
	Perimeter	H														
	Vertices	M-H														
	Vertices	H														
Burgos	Perimeter	H														
	Perimeter	H														
	Vertices	H														
	Vertices	H														
Jaén	Perimeter	M														
	Perimeter	H														
	Perimeter	H														
	Vertices	H														
La Carolina	Perimeter	M														
	Perimeter	H														
	Vertices	M														
	Vertices	H														
Logroño	Perimeter	M														
	Perimeter	M														
	Perimeter	H														
	Vertices	H														
Majadahonda	Perimeter	M														
	Area	M														
	Vertices	M-L														
	Vertices	H														
Mengíbar	Perimeter	H														
	Perimeter	H														

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

	Vertices	H														
	Vertices	H														
Pamplona	Perimeter	L														
	Perimeter	H														
	Area	L														
	Vertices	M														
NOTE: Almería assessment measures were calculated with only S25k image (Grey cells cannot be calculated). The acronyms are shown in Table 4.																

FIGURES

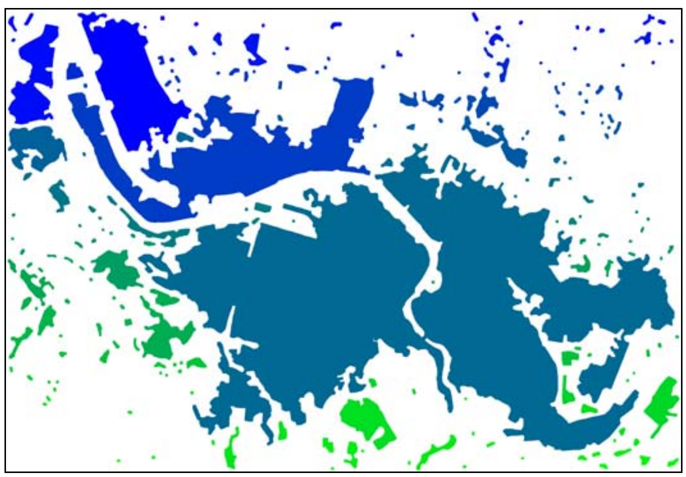
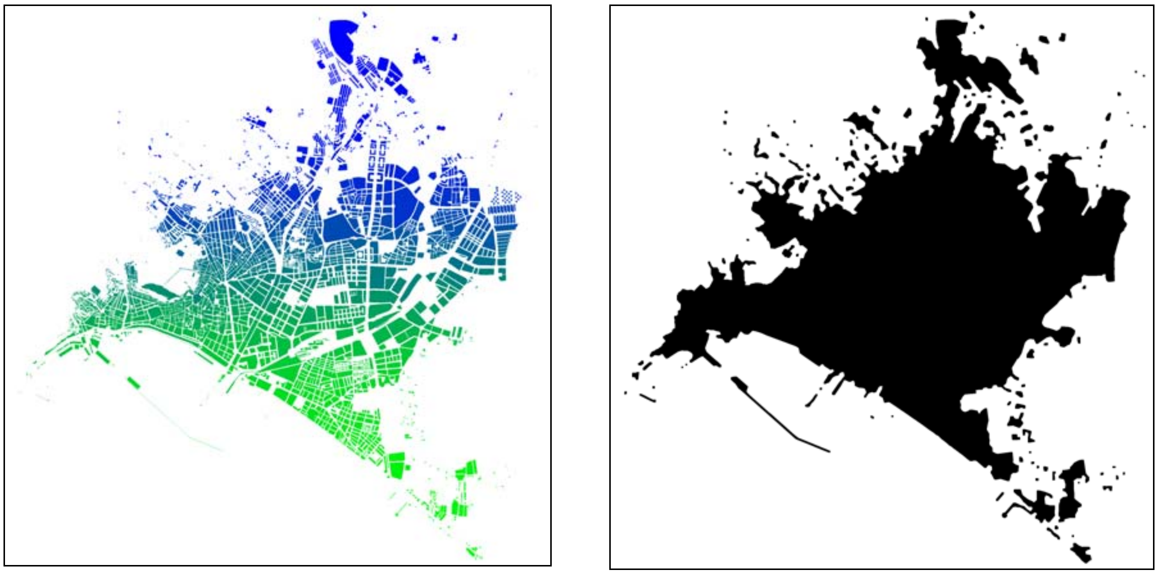


Figure 1. Example shows a complete use of amalgamation of zonified objects in the city of Bilbao. Each colour represents a different zone.

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157



(a) (b)

Figure 2. Example of Quasi-convex hull operator in developing city perimeter from isolated city-blocks.

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

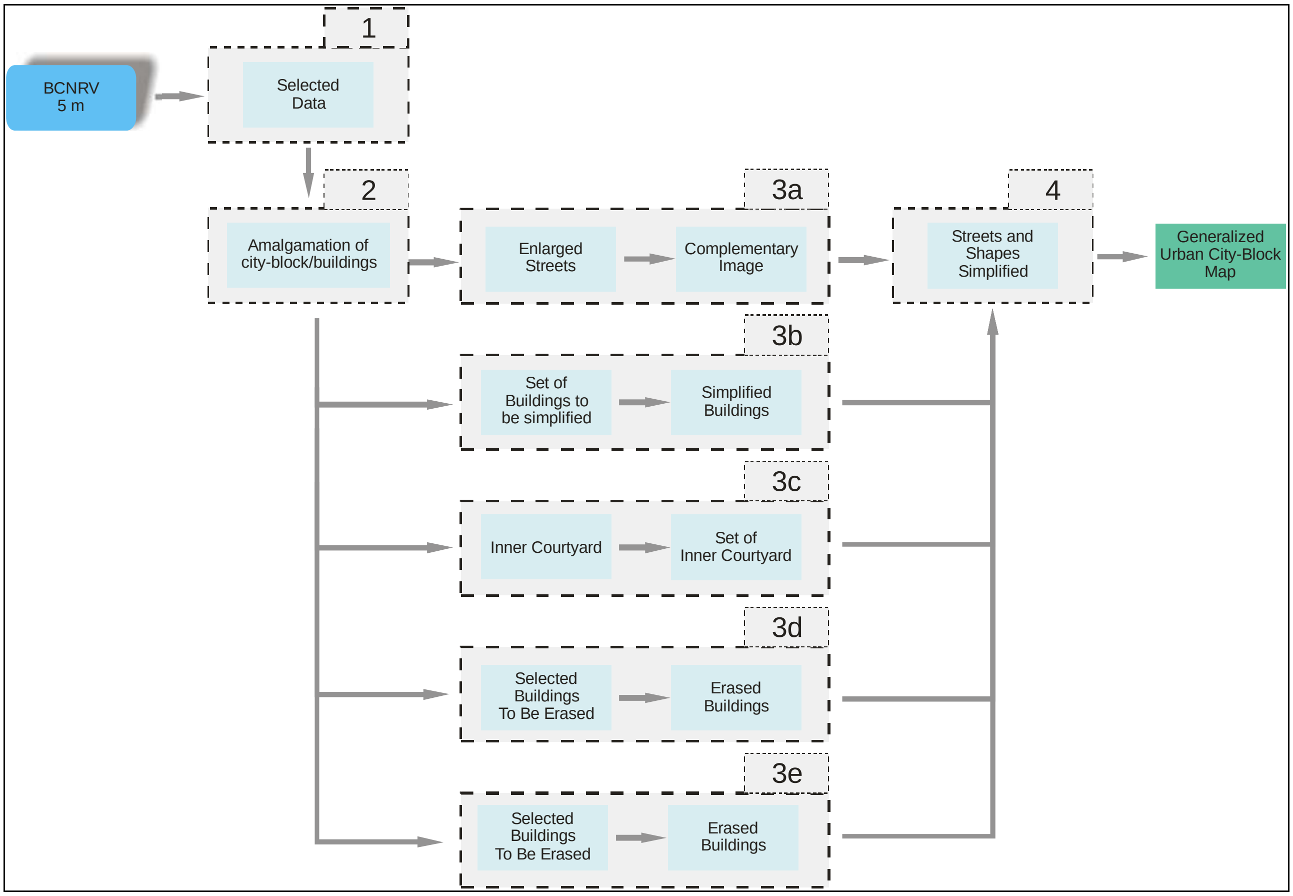


Figure 3. Semiautomatic generalization process. Numbers in dashed boxes represent the phases and sub-phases of the generalization process.

Ureña Cámara, M. A. y Ariza López, F. J. (2008): "Semiautomatic urban map generalization using a raster-vector model", GeoFocus (Artículos), nº 8, p. 225-250. ISSN: 1578-5157

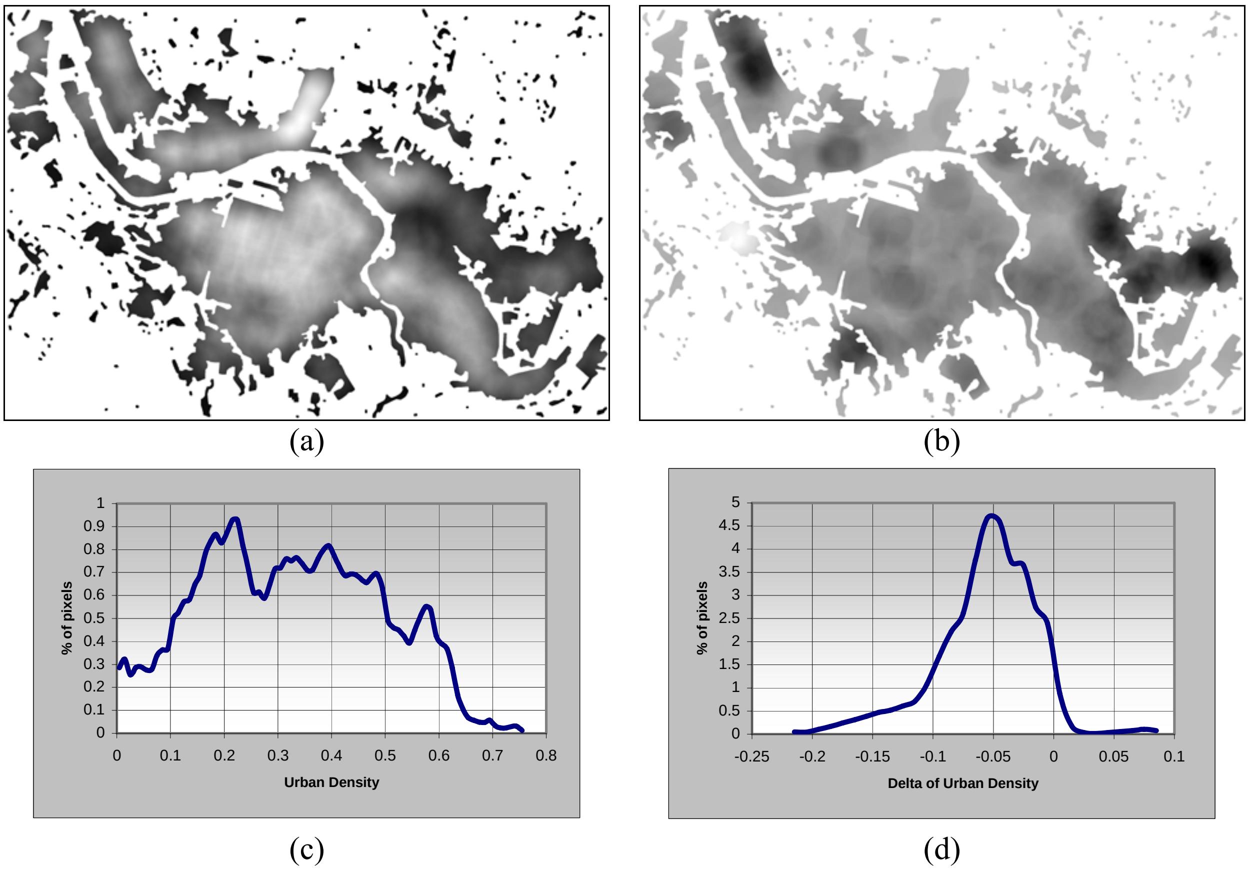


Figure 4. Local urban density in original map (a) and generalized map (b) of the city of Bilbao. The (c) and (d) parts represent the histograms of both (a) and (b) images.