Abstract
GIS Generalization makes data less-detailed and less-complex for better analysis and representation at reduced scales. It allows interactive and batch processes for selecting, analyzing, generalizing, and post processing data based on user-specified specifications. Generalization involves many processes such as simplification; create centerlines, area aggregation, elimination, smoothing and symbolization.
This paper presents a new GIS generalization module which is developed using ArcGIS / Arc Info software and VBA programming language for semi-automatic generalization in the context of GIS. Different generalization functions are implemented. The developed module proves its efficiency and potentiality to be applied in different generalization processes from large-scale to a small scale maps. This module performs the generalization processes on both spatial and attributes database components. Many experiments are performed to evaluate the developed module using different data sets. The results prove that this module is efficient, accurate and successful.
The case study area is located in Cairo governorate. Various hardcopy maps with different scales ranging from 1:5,000 to 1: 1,000,000 are acquired. Hardcopy maps are transformed into digital maps under strict quality control procedures. Then, digital data are edited, edge matched, and transformed to a unique GIS data set having the same coordinate system. Then the developed GIS database is checked for accuracy, completeness, and consistency.
(Key Words: Generalization - Digital Conversion - Digital Maps - Geographic Information System)

1 MAIN ELEMENTS OF GENERALIZATION
Classification deals with creation of class corresponds to a meaningful grouping of locations. Grouping performs combination of two or more area features into a single area feature. Symbolization is polygon to point symbolization process. Simplification removes unwanted detail from a linear geometry by determining which points are critical to its portrayal and removing all other points. Smoothing smoothes the rough and jagged edges of a linear geometry so as to produce a more aesthetically pleasing appearance. Exaggeration increases the size of features than usual to make them readable in the generalized map, especially for important landmarks. Elimination deletes polygons with area less then minimum area or lines with length less than a specified minimum length.

2 ADVANTAGES AND PROBLEMS ASSOCIATED WITH MAP GENERALIZATION
1. Plotting time is often a bottleneck in many GIS applications. So, as the amount of data is reduced through the generalization process, the plotting speed is increased.
2. The generalization process reduces the storage size. Coordinate pairs are the bulk of data in many GIS. Hence, simplification may reduce a data set by 70% without changing the perceptual characteristics of some features, this results in significant savings in memory.

3. The generalization process increase the processing speed. This can be done through the following:
   - Faster vector-to-raster conversion
   - Faster vector processing. The time needed for many types of vector processing including translation, rotation, rescaling, cartometric analysis will be greatly reduced with a simplified data set
   - Many types of symbol-generation techniques will also be speeded up, e.g. many shading algorithms that calculate intersections between shade lines and polygonal boundaries. A simplified polygonal boundary will reduce both the number of boundary segments and also the number of intersection calculations required

On the other hand, When map generalization is applied, we will have the following problems:

1. Problems with plotter resolution when the scale is reduced.
2. As the scale of a digital map is reduced, the coordinate pairs are shifted closer together.
3. With significant scale reduction, the computed resolution could easily exceed the graphic resolution of the output device, e.g. a coordinate pair (0.1, 6.3) reduced by 50% to (0.05, 3.15) could not be accurately displayed on a device having an accuracy of 0.1.

3 Research Objectives
When considering the theoretical elements of why one should generalize, six such elements may be distinguished.
- The first theoretical element is that of **reducing complexity**. Here, complexity is a measure of the visual interaction of various graphic elements within a map. It is of paramount importance that the central result of the cartographer’s analysis is passed on to the end-user of the map. The number and the diversity of graphic elements within a given area impacts the map’s ability to communicate the mapped information to the reader. The complexity of an area increases as the scale is reduced and features within that area become cluttered in appearance.
- **Maintaining spatial accuracy** is the second theoretical element that motivates generalization. It is possible to measure to what extent this goal is reached, as spatial accuracy may be directly related to the displacement between the original and the generalized features of a map. This retention of spatial accuracy deals with geographical data — the cartographic features, such as points, lines and areas.
- The third theoretical element is **maintaining attribute accuracy**. In addition to considering the maintenance of spatial accuracy, one must also consider the accompanying attribute data that is associated with these spatial representations. This goal is, for the most part, purely numerical in nature. It makes use of both classification methods and statistical analysis. The overall objective is to minimize the unintentional alteration of feature attributes, which will affect the spatial representation of the features.
- The fourth objective is **maintaining aesthetic quality**. The aesthetic quality of a map is not an absolute measure, as it is a combination of a multitude of factors. These
factors include appropriate and consistent use of figure–ground relationships, overall balance, layout, typography styles and positioning and color or gray tones. A great deal of effort has been spent by the cartographers to establish guidelines for good cartographic design and, even though it is difficult to formulate specific rules, general guidelines are now being established.

- **Maintaining a logical hierarchy** is the fifth theoretical element. It reflects the need for an ordering of the mapped features. A large city must be more readily noticeable than a small city and highways should be more prominent than country roads. This seems easy to achieve when considering a single class of features, but may become more difficult when the whole map is considered in a more holistic sense.

- The sixth element is that of **consistently applying generalization rules**, which seeks to ensure consistent and unbiased generalizations. The selection and application of generalization rules is a very subjective process and there is much variation in the way different cartographers choose to solve these problems.

So the main objectives of this research are:

- Study different map series and their characteristics.
- Formalize the Data Model of various map series.
- Study, evaluate, and analyze existing generalization software
- Develop an expert system which can carry out some generalization processes in a semi automatic process.

4 DEVELOPMENT OF THE GIS GENERALIZATION MODULE

There is a need for developing new powerful algorithms integrated with widely used GIS software packages, such as ArcGIS, to perform generalization processes in the integrated spatial and attribute database. This is very important for many reasons, such as reduction of database size for storage and processing purposes, performing certain types of analysis on small-scale digital maps and GIS databases, integration of multi data sources with different scales and formats, the need of semi automatic generalization module taking the input scale and the output scale to do the generalization model automatically without any human errors, and the need to take into consideration the standards of Egypt maps.

In the next section, a new technique is developed for performing some generalization tasks using the Visual Basic for Application (VBA) programming language, associated with the ArcGIS software package. These tasks are elimination, symbolization, grouping, and simplification. The proposed technique is capable of performing data abstraction and GIS database generalization. Many experiments are performed to evaluate the proposed technique using different GIS data sets that are developed separately in different projects from various sources with different scales and types.

4.1 **Generalization Procedure for Polygon Features (GPPF):**

The generalization procedure for polygon features is performed as follows:

1. Polygons are kept with the same feature definition.
2. Polygons with area less than \((\text{Accuracy in mm} \times \text{Reverse of Output Scale})^2\) will be Eliminated unless they are important, in this case they will be Symbolized (transformed to point feature) and then Eliminated.
3. polygons with in between distances less than (Accuracy in mm* Reverse of Output Scale) will be Grouped (Aggregated). If any of these features is important, it will be also Symbolized.

4. The polygons which have segments with length less than (Accuracy in mm* Reverse of Output Scale) will be Simplified.

The following figure Figure (1), shows the flow chart of the adopted procedure.

![Flow chart of Generalization Procedure for Polygons Features (GPPF)](image)

**Figure (1) Generalization Procedure for Polygons Features (GPPF)**

4.2 **Generalization Procedure for Linear Feature (GPLF):**

The generalization procedure for linear feature is performed as follows:

1. Lines are kept with the same feature definition.
2. Lines with length less than (Accuracy in mm* Reverse of Output Scale) will be Eliminated.
3. Lines with segments length more than (Accuracy in mm* Reverse of Output Scale) will be Simplified and Smoothed.

The following figure Figure (2), shows the flow chart of the adopted procedure.
4.3 **GIS Generalization Module Development:**
The comprehensive GIS module (Figure (3)) is developed using Visual Basic for Application (VBA) programming language on the top of the ArcGIS ArcInfo software.

5 **GIS DATABASE DEVELOPMENT FOR THE CASE STUDY AREA**
The GIS databases for the case study area are developed from various hardcopy map series with scales ranging from 1:5,000 to 1:1,000,000 using the following procedure:
1- Acquire maps for the case study area in Cairo with scale (1:5,000 & 1:50,000 & 1:250,000 & 1:1,000,000).
2- Convert the maps to digital format with WGS84 datum and the Egyptian Coordinate System (ETM).
3- Apply the edge matching concept to form one layer for each feature class.
4- Build the topology for all feature classes.
5- Build the GIS database for all feature classes.
6- Add field with name "scale" in all layers which defines the important feature in various scales. This value is equal to 1,000,000 or 250,000 or 50,000.
7- The following rules are taken into consideration during work:
   - All hospitals and schools in layer urban in scale 5000 must be converted into landmarks in small scale (50,000), this means that all hospital and schools in layer urban in scale 5,000 store value (50,000).
   - All roads are existing in scale 1:5,000; the main roads only are kept in scale 1:50,000 (store 50000 in the scale field).
   - The railways connecting main villages are kept in scale 1:250,000 and 1:1,000,000 (store 1:250,000 or 1:1,000,000 in the scale field).
   - All canals are exist in scale 1:5000, the main canals only are kept in scale 1:50,000 (store 50,000 in the scale field), the large canals connecting main villages are kept in scale 1:250,000 and 1:1,000,000 (store 1:250,000 or 1:1,000,000 in the scale field).

More details are given in the following sections.

5.1 Data storage and processing
GIS data are organized in layers and levels. Individual data/levels are stored in individual data files. Each layer is stored in a group of files. ArcGIS software store the layers in Geodatabase file, it is physically organized for proper and easy data management and retrieval.

Data required for this project are available in two different forms: spatial and attributes. Spatial data are related to specific locations and represented in a vector format. Vector data includes maps with various scales in hardcopy format. Hard copy maps of a scale 1:50,000 produced by the Egyptian Survey Authority (ESA) and the Military Survey Department (MSD) are used for the case study area in Cairo. Hard Copy maps of a scale 1:250,000 and scale 1:1,000,000 produced by the Military Survey Department (MSD) will be used for the same area. Attribute data are those related to geographic features. They linked to the features in the layer's table.

The developed layers are:
1. 1:5,000 GIS database
2. 1:50,000 GIS database
3. 1:250,000 GIS database
4. 1:1,000,000 GIS database

5.2 GIS Attribute Data Handling
GIS attribute data are numerous. They include numerical and text data. Numerical and text data are often recorded on paper and scattered amongst a multitude of different conventional registers and files. Attribute data handling includes three successive procedures, namely data preparation, entering, and editing. Quality Control (QC) procedures must be applied through these operations.

Data acquisition is controlled with data availability at different organizations, roles and regulations, data characteristics, and good preparation of data collection structure. The data fields of the collection forms should correspond to the data fields of the database.

Quality control procedures are done manually and using automatic verification functions. Automatic verification functions retrieve only formal errors. Incorrect information can only be detected and edit by manual copy reading. The links between spatial and attribute data is checked at the end. The following tables summarize the
characteristics of the used GIS database layers in the practical experiments for evaluating the developed comprehensive GIS generalization module.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Example</th>
<th>Source</th>
<th>Projection</th>
<th>Layer_Name</th>
<th>Definition</th>
<th>Attributes</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>5,000</td>
<td>Cairo</td>
<td>ESA</td>
<td>ETM</td>
<td>Builds Urban areas</td>
<td>Polys polys</td>
<td>Some Names Some Names</td>
<td>1988</td>
</tr>
<tr>
<td>50,000</td>
<td>Cairo</td>
<td>Military Survey</td>
<td>UTM</td>
<td>Lakes Urban areas</td>
<td>Gravels pols</td>
<td>Some Names Some Names</td>
<td>2000</td>
</tr>
<tr>
<td>250,000</td>
<td>Cairo</td>
<td>Military Survey</td>
<td>UTM</td>
<td>Graves Urban areas</td>
<td>Agriculture pols</td>
<td>Some Names Some Names</td>
<td>1997</td>
</tr>
<tr>
<td>1000,000</td>
<td>Cairo</td>
<td>ESA</td>
<td>Geographic</td>
<td>Capitals City</td>
<td>Villages Markaz Points Points Points</td>
<td>Type , class</td>
<td>1968</td>
</tr>
</tbody>
</table>

Table (1) GIS database layers utilized for evaluating the developed GIS generalization module (polygons - Points)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Example</th>
<th>Source</th>
<th>Projection</th>
<th>Layer_Name</th>
<th>Definition</th>
<th>Attributes</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>5,000</td>
<td>Cairo</td>
<td>ESA</td>
<td>ETM</td>
<td>Road Edges Canals</td>
<td>Drains Lines</td>
<td>Some Names, Class</td>
<td>1988</td>
</tr>
<tr>
<td>50,000</td>
<td>Cairo</td>
<td>Military Survey</td>
<td>UTM</td>
<td>Roads Railway</td>
<td>Lines Some Names, Class</td>
<td>2000</td>
<td></td>
</tr>
<tr>
<td>250,000</td>
<td>Cairo</td>
<td>Military Survey</td>
<td>UTM</td>
<td>Roads Railway</td>
<td>Lines Some Names, Class</td>
<td>1997</td>
<td></td>
</tr>
<tr>
<td>1000,000</td>
<td>Cairo</td>
<td>ESA</td>
<td>Geographic</td>
<td>Roads Railway</td>
<td>Lines Some Names, Class</td>
<td>1968</td>
<td></td>
</tr>
</tbody>
</table>

Table (2) GIS database layers utilized for evaluating the developed GIS generalization module (Lines)

5.3 Evaluation of the developed GIS Generalization Module

Many experiments are performed to evaluate the developed modules using different data sets. In the following sub-sections, experiments details, the study area, and input data are given. According to the procedures are described later we will run the module and test the outputs, then the assessment criteria are outlined.

The following figure, summarizes these experiments.

Figure (4) experimental work for testing the developed GIS generalization module
5.3.1 Polygon Features

After the GIS data preparation we make the following experiment using the developed application. The input is urban areas for scale 1: 5000, Figure (5) and the output should be urban areas for scale 1: 50,000. We have some attribute data should be kept in scale 1:50,000, Figure (6) also we have some land marks should be kept in scale 1:50,000, Figure (7). We can see that the tolerance is automatically calculated using the output scale but it is editable, so the user can change the tolerance.

Selecting the input scale, Figure (5-11), feature layer is urban area, input layer, writing the output name, selecting the output scale, and click on the generalize button. The output of urban map after running the generalization Procedure for Polygons Features (GPPF) can be shown in Figure (8). The important land mark like schools, hospitals, and police stations are symbolized and converted into points with all the attributes, Figure (9).

We can compare between the input attributes and output attributes using the Table (3).
Figure (9) The final Generalized Urban map scale (1 : 50000)

Figure (10) Urban with scale (1 : 5,000) and the Generalized Urban with Scale (1 : 50,000)

Table (3) Attributes of Urban Areas Before and After Generalization

<table>
<thead>
<tr>
<th>Input Data Description</th>
<th>Output Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of object: Polygon Feature</td>
<td>Type of object: Polygon Feature</td>
</tr>
<tr>
<td>Class</td>
<td>Class</td>
</tr>
<tr>
<td>Number of records: 2456</td>
<td>Number of records: 200</td>
</tr>
<tr>
<td>Attributes: FID, Shape, AREA, PERIMETER, BUILDSGK, BUILDSG:ID, BUILDINGSSK:ID, BUILD:K, AREA, BUILDK_PERIMETE, ARABIC_NAME</td>
<td>Attributes: FID, Shape, AREA, URBAN50K, URBAN50:ID, URBAN50K AREA, URBAN50K_PERIMETE</td>
</tr>
</tbody>
</table>

The comparison between generalized urban layer and the actual Urban layer with scale 1:50,000 is made to evaluate the performance of the GIS generalization module including spatial and attribute components
- The spatial difference
- The attributes differences is discussed

**Analysis of results**
After analyzing the results, the following remarks could be made:
- The output of the generalization is more accurate than the maps of scale 50,000 because the source of data is more accurate and no shift in coordinates is performed.
- The output of generalized map is shifted than the maps with the same scale
- The attributes of the urban layer in scale 5000 is kept in the generalized layer
- The processing time is short compared with manual generalization
- The effort made is little compared with manual generalization
- No human decisions during the generalization are required.
- We can concentrate on the maps with large scale, and then we can produce the small scale maps from it using the developed module.
- The overlaying between small scale and large scale will be very easy.

5.3.2 Linear Features
After the GIS data preparation we make the following experiment using the developed application. The input is roads for scale 1: 50,000, Figure (12) and the output should be roads for scale 1: 1,000,000, Figure (14). We can see that the tolerance is automatically calculated using the output scale but it is editable, so the user can change the tolerance.
Selecting the input scale, Figure (13), feature layer is urban area, input layer, writing the output name, selecting the output scale, Figure (13), and click on the generalize button. The output of roads after running the generalization Procedure for Line Features (GPLF) can be shown in Figure (14). We can compare between the input attributes and output attributes using the Table (5-5).
Analysis of results
After analyzing the results, the following remarks could be made:

- The generalized roads date is 2000 and the actual roads date is 1968
- The output of generalization is more accurate than the maps of scale 1,000,000 because the source of data is more accurate
- The output of generalized map is shifted than the maps with the same scale
- The attributes of the roads layer in scale 1:50000 is kept in the generalized layer in scale 1:1000000
- The processing time is short compared with manual generalization
- The effort made is little compared with manual generalization
- No human decisions during the generalization are required.
- We can concentrate on the maps with large scale, and then we can produce the small scale maps from it using the developed module.
- The overlaying between small scale and large scale will be very easy.
6 CONCLUSION

1. Generalization is the process of changing the form of map features, usually performed when the map is transformed from one map scale to a smaller scale. It is a scale-dependent procedure. The smaller, the scale the greater generalization of map objects is necessary.

2. Generalization is very important for many reasons, such as reduction of database size for storage and processing purposes, performing certain types of analysis on small-scale digital maps and GIS databases as well as integration of multi data sources with different scales and formats.

3. Various algorithms have been developed addressing individual generalization tasks. However, there was very limited research for developing a comprehensive system that can handle integrated tasks and dealing with GIS database generalization where two types of generalization tasks are required: database generalization and cartographic generalization.

4. Generalization allows interactive and batch processes for selecting, analyzing, generalizing, and post processing data based on user-supplied specifications.

5. The attributes of GIS data layers are kept in the output maps.

6. A GIS database is developed from existing hardcopy Egyptian maps covering part of Greater Cairo governorate with scales ranging from 1: 5,000 to 1: 1,000,000. This database is utilized for carrying out all the experiments.

7. A robust generalization module should start with specifying input and output scales, processed layers, and tolerances, and then proceed with semi-automatic generalization process that execute necessary generalization functions in the correct manner.

8. A new comprehensive module is developed for performing some generalization tasks using the VBA programming language, associated with the ArcGIS software package. The proposed technique is capable of performing data abstraction and GIS database generalization. The efficiency of the developed generalized modules are assessed by comparing the outputs with the existing spatial and attribute databases within the pre-developed GIS. Experimental work proves that these modules are efficient and successful.

9. The developed GIS generalization module is ready to use in Egypt, and it might need some modifications to be use in other countries where other maps specifications are used.

10. The developed GIS generalization module can work in different projection and different datum.

11. Using the developed GIS generalization module, the attributes of the generalized layers are kept.
12. The developed GIS generalization module does not need much interactive work.
13. The output of the developed generalization module is more accurate than the maps with small scale because the source of the data is more accurate.

7 RECOMMENDATIONS
The some comprehensive recommendations for future research and development are as follows:

1. There is a need to develop full automatic generalization module taking the input scale and the output scale to do the generalization model automatically without any human errors.

2. The generalization is very important as it eliminate work duplication and save money and time so we should use this technique in the authorities, the governmental agencies, and the private sector companies.

3. There is a need for developing generalization software using open source GIS softwares to reduce the cost.

4. There is a need to develop new summarized GIS attribute database related to the produced generalized digital map.

5. There is a need for developing generalization software to limit the interactive processes for the incomplete and unresolved features.
**LIST OF REFERENCES**


12. Youssef Ibrahim Abdelaziz, Prof. of Surveying and Photogrammetry, Faculty of Engineering Cairo University, has wrote a paper for Scale map. (Scale Map for Surveying Map), Under processing.