D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany.



# REVISION OF GEOGRAPHIC DATA: A FRAMEWORK

**J. Raul Ramirez**
The Ohio State University
Center for Mapping


## ABSTRACT

Demand for up-to-date geographic data is increasing due to the fast growth of Geographic Information System (GIS) technology. GIS requires up-to-date data in order to provide reliable answers to geographic queries. Outdated or incomplete geographic data could generate misleading or simply incorrect answers. Traditionally, updating geographic data (map revision) has been considered an extension of map production. Lately, Ramirez (1995, 1996) has developed pieces of an independent theory for revising cartographic products. These pieces are used here to present a general framework for geographic data revision. Time is the first element of this framework. The second element is the surface represented by the geographic data and its characteristics. The third element is the kind of modification that can occur in the geographic data. Defining the data sources to detect and record modifications in the geographic data is the fourth element. The fifth element is the set of techniques used to update the geographic data. The last element is integrating individual objects and relief into a consistent representation using conflation techniques.


## 1 INTRODUCTION

Geographic Information System (GIS) technology is a multibillion-dollar industry worldwide. Experts in many areas (from lawyers to engineers) are using GIS today at an increasing rate to analyze and answer questions related to geographic areas. GIS is without doubt the most efficient way to analyze and query geographic data!

A fundamental condition to use GIS is the existence of the appropriate geographic data. If a particular area does not have digital geographic data, then it is not possible to use GIS in that situation. The existence of geographic data is a necessary but not sufficient condition to use GIS. Geographic data must be representative of the situation to be studied. In other words, geographic data must contain the related information at the appropriate qualitative and quantitative levels. This constitutes a sufficient condition.

From the viewpoint of quality, there are several factors that affect geographic data. One of them is its currency. Outdated geographic data is not representative of the current geographic situation. GIS analysis and query using outdated geographic data will provide results that are valid for the situation the data represented. For example, in the United States the most precise nationwide coverage is provide by the 7.5-minute quadrangle series and its digital equivalent, the DLG files (available only for a small percentage of the whole country). The average age of the quadrangles is 25 years. The use of GIS based on this data provides answers that were valid 25 years ago!

The process of updating geographic data has been called map revision and is considered an extension of map production. In this paper, a framework for geographic data revision (map updating), independent of map production, is presented. This is divided into six sections:

(1) Study of time.
(2) Study of the surface represented.
(3) Modification that can occur on that surface.
(4) Data sources to detect and record changes.
(5) Techniques used to update the geographic data.
(6) Integration of individual objects and relief into a consistent representation.

## 2 TIME AND GEOGRAPHIC DATA REVISION

Revision of geographic data is time-dependent. Usually, the longer the amount of time that has passed since the last updating, the larger the number of changes in the geographic data and the greater the complexity of updating. Of course, changes in an area of interest are not systematic. There may be regions of the area where there are no changes or the magnitude of changes can be neglected, and other regions where major changes occurred and revision will require a great deal of effort.

There are also three different epochs to consider from the viewpoint of geographic data revision:

(1) The epoch of the data used in last revision ($\mathbf{T_P}$).
(2) The epoch of the data for the current revision ($\mathbf{T_N}$).
(3) The current epoch ($\mathbf{T_C}$).

$\mathbf{T_P}$ (epoch in the past) was the epoch of collection of the data sources used to generate or revise the cartographic product we want to revise today. Usually, that product was generated or revised at least five years ago. In some cases, as in the above example of the 7.5-minute quadrangle series, this may have happened twenty-five years ago! $\mathbf{T_N}$ (epoch in the near past) is the epoch of collection of the data sources to update a cartographic product (the task on hand). With current technologies, there is always a time interval of a few months to several years between the epoch when the data is acquired and the actual revision and updating process. Finally, $\mathbf{T_C}$ is the current epoch. This epoch is constantly changing and, therefore, it is impossible to have geographic data reflecting this epoch. What can be done is to have geographic data for the epoch $\mathbf{T_N}$ such that the difference $\mathbf{T_C - T_N}$ will be as close to zero as possible.

D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany.



From a practical viewpoint, it is important to keep this difference small in order to decrease the cost of revision. Figure 1 illustrates the relationship of the three epochs.

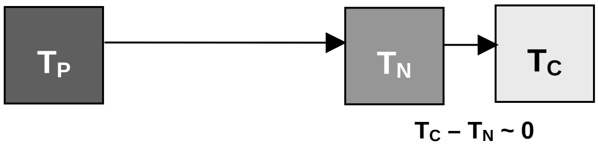


Figure 1. Time Line for Geographic Data Revision

## 3 THE SURFACE REPRESENTED

Geographic data represents the dynamic surface of the Earth. This surface is constantly changing. Changes are due to two major sources:

(1) Natural forces.
(2) Human actions.

Natural forces, on one hand, generate abrupt and usually radical changes. For example, the changes due to earthquakes, tornados, flooding, landslides, forest fires, etc. These changes usually modify the landscape in a radical fashion. Changes are obvious and easy to detect. On the other hand, natural forces also generate systematic and usually slow changes. These changes are caused by the constant action of such forces as gravity, tectonic masses, wind, running water, etc. These changes may not be obvious right away but with the passing of time they become obvious.

Human actions also generate abrupt and radical changes. The use of explosives causes the most common type of these changes. Sometimes, explosives are used with the specific purpose of changing the landscape as part of an engineering project, but many times they are used as the result of national or international conflicts. War, unfortunately, is one of the major reasons why geographic data need to be updated. Humans also generate predictable and unpredictable changes over time. For example, construction of many manmade features such as housing projects, shopping malls, new roads, etc., may be done over a larger time period but in these cases the outcome is known and predictable. Unpredictable or random changes are the result of unpredicted circumstances. For example, a road may be permanently closed because of changes in the demographics of a region; or because of economic hardship a city park may be permanent closed; or, because of a new baby, a homeowner may add a room to his/her house, etc. Unpredictable changes are also those changes, such as open-field mining and logging, whose outcome is unknown at the time and are only evident later on. Human actions can be summarized as resulting in two types of changes: predictable and unpredictable.

The features that change and the magnitude of the above changes are very important in geographic data revision. Changes in those features represented in the geographic data in consideration are the only types of changes we are concerned with. For example, if the geographic data in consideration does not include driveways, the construction of ten new driveways in the area will not require updating the data. For those features represented in the geographic data (including the relief), we are concerned only with those changes that can be represented at the scale of the cartographic product in consideration. Using the 7.5-minute quadrangle series as an example again, in agreement with the U.S. National Map Accuracy Standards at the scale of 1:24,000, not more than 10% of the well-defined planimetric points tested should have an error greater than 40 feet, and no more than 10% of elevations tested should have an error greater than one-half of the contour interval. Therefore, if a water stream changes its position by 5 feet in a period of five years, there is no need to update it in the geographic database if the difference between the ground position and the geographic representation of the stream in the digital data is less than 40 feet. Here it is assumed that the original geographic representation of the stream in the database is not perfect. Also, if the relief has changed by less than one-half of the contour interval (contour interval goes from 5 to 20 feet in the 7.5-minute series) there is no need to update it. Ramirez (1996) classifies geographic changes as shown in Table 1.

Table No. 1
Geographic Data Revision: Change Factors

| Origin | Frequency | Magnitude |
|---|---|---|
| Systematic | Constant | Small |
| Abrupt | Low | Large |
| Predictable | High | Large |
| Unpredictable | Medium | Medium |

## 4 MODIFICATIONS ON THE SURFACE

The third element of the framework is the kind of modification that can occur in the geographic data. Expressing these modifications and learning to handle them is the purpose of this section.

### 4.1 Changes on the Surface

Let us assume that geographic data is composed of objects representing the features and surface of the Earth. Examples of those objects are a highway, a building, a house, a political boundary, etc. Between the epoch difference, $\mathbf{T_N - T_P}$, four situations are possible for any object of interest at $\mathbf{T_N}$. An object:

(1) May no longer be on the ground.
(2) May have changed.
(3) May be unchanged.
(4) May not be represented.

Using the Set notation, the vector datasets for times $\mathbf{T_P}$ and $\mathbf{T_N}$ can be expressed as:

$M_P = \{m_p \mid m_P$ is the vector representation of a terrain or relief object, at a date $\mathbf{T_P}$, and scale $\mathbf{S}$, for purpose $\mathbf{P}\}$, (2)

and,

$M_N = \{m_N \mid m_N$ is the vector representation of a terrain or relief object, at a date $\mathbf{T_N}$, and scale $\mathbf{S}$, for purpose $\mathbf{P}\}$. (3)

Where $\mathbf{M_P}$ is the out-of-date geographic dataset and $\mathbf{M_N}$ is the up-to-date data set.

The set **M** representing the terrain for the date $\mathbf{T_P}$ in the past, can be expressed at the date $\mathbf{T_N}$ in the near past as:

$$M_P = \{D, C, U\}. \quad (4)$$

Where:

D = {*d*|*d* is the vector representation of an object that no longer exists},
C = {*c*|*c* is the vector representation of an object before change},
U = {*u*|*u* is the vector representation of an unchanged object}. (5)

Two additional sets need to be considered for the date $\mathbf{T_N}$: the set **N** of new terrain objects to be shown on the terrain representation (in agreement with the scale and purpose of the representation), and the set **H** of the objects that have changed. Sets **N** and **H** are defined as:

N = {*n*|*n* is the vector representation of a new object},
H = {*h*|*h* is the new vector representation of an object that has changed}. (6)

And, the revised geographic dataset may be expressed as:

$$M_N = (M_P - \{D, C\}) \cup N \cup H. \quad (7)$$

A closer representation of what may be realistic expected is:

$$M_N = U \cup N \cup \alpha C, \quad (8)$$

where:

$$H = \alpha C, \quad (9)$$

and,

$$\alpha C = \{\alpha_1 c_1, \alpha_2 c_2, \alpha_3 c_3, ...\},$$

where, $\alpha_i$ is the modification operator which converts the previous representation of object $\mathbf{c_i}$ into the new representation $\mathbf{h_i}$.

Identification and manipulation of the sets **D**, **U**, and **N**, and the transformation of set **C** into the set **H** are the goals of geographic data revision. This allows the creation of the set:

$$\mathbf{M_N = \{N, U, H\}} \quad (10)$$

for the time $\mathbf{T_N}$.

### 4.2 Introduction to Revision of Geographic Data

A highly automated solution for geographic data revision starts by assuming that there are at least three datasets:

- $\mathbf{M_P}$ (vector representation of the ground for time $\mathbf{T_P}$),
- $\mathbf{R_N}$ (ground image for time $\mathbf{T_N}$), and
- $\mathbf{DEM_N}$ (digital elevation data for all or part of the area of interest for time $\mathbf{T_N}$) of the same geographic area, and appropriate scale and resolution (or data sources to generate this set).

Additional datasets, such as $\mathbf{M_{IP}}$ (map image for time $\mathbf{T_P}$), $\mathbf{R_P}$ (raster image of the terrain at the time $\mathbf{T_P}$), $\mathbf{M_{MN}}$ (mobile mapping data at the time $\mathbf{T_N}$, in the near past), and $\mathbf{I_N}$ (any other types of images, such as SAR images), may also be available.

In the framework of "total revision" (Thompson, 1987), our goal is to generate up-to-date geographic data for the planimetric and relief representation of the area of interest for a time $\mathbf{T_N}$ in the near past. Therefore, it is assumed here that hypsographic information (more provable in the form of contour lines and spot elevations) is part of the set $\mathbf{M_P}$. A preliminary step in the revision process is the generation of the set $\mathbf{DEM_P}$ (if not available) from the corresponding hypsographic information.

The procedure starts by georeferencing all the datasets to a common coordinate system. Conceptually, we would like to have all existing data sets with a common origin, orientation, and scale. Warping of raster images (coordinate transformation and resampling), and coordinate transformation of vector data are used (whenever they are needed) to accomplish the georeferencing of all databases.

### 4.3 Revision of Planimetric Geographic Data

Revision of planimetric objects uses all available data sources. The procedure includes three steps.

Step One. The set $\mathbf{M_P}$ is overlaid with the set $\mathbf{R_N}$. The information of $\mathbf{M_P}$ is used to identify raster search areas. Raster search areas are marked and are used in step Two. As an example, let us assume that the set $\mathbf{M_P}$ is the road layer. The positional and attribute information of each road in this layer is used to identify areas on the raster image $\mathbf{R_N}$ where these features (if they still exist) should be located. A buffer zone is generated from each vector road, and they constitute the raster search areas.

Step Two. Low-vision edge detection is applied in each raster search area, together with geometric, logical, and/or cartographic constrains. A set $\mathbf{F_N}$ is generated as the result of this operation. Then, Boolean comparison (or similar approaches) between the sets $\mathbf{M_P}$ and $\mathbf{F_N}$ is performed. Two basic questions should be answered in this step:

(1) Is a ground object in the raster search area under consideration?
(2) Is that object equal to the one in the set $\mathbf{M_P}$? Four subsets are generated in this step: **U**, **D**, **C**, and **H**.

As the result of the comparison, three situations and their combinations are possible:

(1) Some or all the objects are unchanged in both datasets. The unchanged features form the subset **U**.
(2) Some objects are no longer in the image $\mathbf{R_N}$. They constitute the subset **D**.
(3) Some or all objects are in both datasets, but they have a different geometric outline. They constitute the subset **C**.

For some layers it may be possible that for the subsets **C** and **H**, each one needs to be replaced by two subsets ($\mathbf{C_1}$ and $\mathbf{C_2}$, $\mathbf{H_1}$ and $\mathbf{H_2}$, respectively). This is due to the magnitude of the changes experienced by ground objects. Hydrographic objects are an example of those whose location changes through time. Depending on the length of the time interval $\mathbf{(T_N - T_P)}$ the location of some objects may change so much that no part of them is found in the raster search areas. In this case, these features will be identified as belonging to the subset **D** (instead of **C**). Under these circumstances, the subsets **C** and **H** must be written as:

$$\mathbf{C} = \mathbf{C_1} \cup \mathbf{C_2}, \tag{11}$$

where,

$\mathbf{C_1}$ ={ $c_1/c_1$ is the vector representation of an object having at least an unchanged segment} (12)
$\mathbf{C_2}$ ={ $c_2/c_2$ is the vector representation of an object having no unchanged segment}

and,

$$\mathbf{H} = \mathbf{H_1} \cup \mathbf{H_2}, \tag{13}$$

where,

$\mathbf{H_1}$ ={ $h_1/h_1$ is the vector representation of the changed object $c_1$}, (14)
$\mathbf{H_2}$ ={ $h_2/h_2$ is the vector representation of the changed object $c_2$}.

The computation of $\mathbf{H_2}$ will be discussed as part of step Three.

Step Three. This step is the generation of the subset **N of** new objects and $\mathbf{H_2}$ (if needed). The goal of the solution discussed here for the revision of geographic data is to have a highly automated, not a fully automated solution. Therefore, some human operator interaction is expected. From this viewpoint, let us consider the set **N** expressed as:

$$\mathbf{N} = \mathbf{N_1} \cup \mathbf{N_2} \cup \mathbf{N_3}. \tag{15}$$

$\mathbf{N_1}$ is the subset of those new objects that can be found from the search of unchanged, no longer existing, or unchanged objects, based on geometric, logical, or cartographic relationships. For example, new roads are part of the road network; therefore, they are connected to previously existing roads. If the connections take place in the area of interest, then they are found by searching each raster search area. After that, edge detection techniques are used to generate the appropriate vector representation of the new roads.

$\mathbf{N_2}$ is the subset of those new objects generated with the help of additional datasets. As indicated earlier, there are two datasets carrying relief information: $\mathbf{DEM_P}$, and $\mathbf{DEM_N}$. These two datasets are compared (elevations for time $\mathbf{T_P}$ and $\mathbf{T_N}$ of the same geographic locations), and based on the differences found, additional raster search areas are generated, and planimetric objects of interest are looked for in those areas by using geometric, logical, or cartographic rules. Edge detection techniques are used to generate the vector representation of these objects.

Another type of data used for the generation of $\mathbf{N_2}$ is mobile mapping data. Mobile mapping data consist of three-dimensional coordinate values along specific objects (such as road centerlines), digital images with different orientation than conventional aerial images (for example, near-vertical images along the road), attributes (for example, road names), etc. This information is used as the geometric definition of additional terrain features, or to define buffer areas to search for additional objects.

Finally, $\mathbf{N_3}$ is the subset of those objects than cannot be found automatically and are collected by the operator during the last step of the process.

### 4.4 Revision of Relief Data

Revision of relief data is based on the datasets $\mathbf{DEM_P}$, $\mathbf{DEM_N}$, $\mathbf{M_{MN}}$ (if available) and planimetric changes.

Step One. The sets $\mathbf{DEM_P}$ and $\mathbf{DEM_N}$ are compared and buffer search areas are generated. These search areas are compared with the planimetric changes ($\mathbf{C_P}$) and set $\mathbf{M_{MN}}$ and a consistent list of buffer areas is generated.

Step Two. Dense digital elevation models of buffer areas are generated from $\mathbf{R_N}$ images. Relief changes are evaluated in agreement with national map accuracy standards and buffer areas with significant changes are marked.

Step Three. Contour lines are generated for buffer areas with significant changes by using digital terrain-model technology. The set $\mathbf{M_{MN}}$ is used as breakline information (if possible). These new contour segments are conflated with those from unchanged areas generating a consistent relief representation.

Step Four. Merging the dataset $\mathbf{DEM_N}$ with the data collected in Step Two generates a consistent **DEM**.

## 5 DATA SOURCES

Planimetric revision is based on a minimum of three datasets: $\mathbf{M_P}$ or a $\mathbf{M_{IP}}$, $\mathbf{R_N}$, and $\mathbf{DEM_N}$. From a practical viewpoint, if there is not a digital vector dataset $\mathbf{M_P}$ of the area of interest for the epoch $\mathbf{T_P}$, then one can be generated from the digital image $\mathbf{M_{IP}}$ of the corresponding map. The fundamental assumption in this framework is that the updated geographic database $\mathbf{M_N}$ will have only the same type of objects as the $\mathbf{M_P}$ dataset. Therefore, it is assumed in this framework that such a database exists or can be generated from its digital image. A typical example of **M** datasets is the Digital Line Graph (DLG) data from the U.S. Geological Survey (USGS).

The set $\mathbf{R_N}$ is the digital image of the area of interest at the appropriate resolution and accuracy. In general it will be assumed in this framework that that image is a digital orthophoto. A desirable but not necessary condition is that the images covering the area of interest will perfectly fit the area of interest without major overlap. The accuracy of these images must be higher than the accuracy of the vector digital data for the $\mathbf{T_P}$ epoch. This of course is necessary to preserve the level of accuracy of the vector dataset **M**. For example, if we want to update 1:24,000 DLG files, the digital orthophotos must be of accuracy higher than 1:24,000. Otherwise, the updated dataset **M**, due to error propagation during the updating process, will have an accuracy smaller than the one needed for 1:24,000 maps.

The resolution of the digital image is important in order to be able to extract all the pertinent objects. Of course, ground objects that are not present in the digital images cannot be extracted from them. If, for example, you need to extract ground objects of 1 foot-length and the image's pixel resolution is 1 meter, you cannot extract those objects. Typical examples of the set $\mathbf{R_N}$ are the digital orthophoto quarter-quadrangles at the scale 1:12,000 generated by the USGS.

The last dataset required for planimetric revision is $\mathbf{DEM_N}$. The assumption here is that this dataset is a sub-product of the generation of digital orthophotos. Two conditions are important for this dataset.

(1) Ideally, this dataset will include a dense coverage of elevation values describing brake lines, break points, and homogeneous areas.
(2) The accuracy of the elevation values will be appropriate for generation of the corresponding digital orthophotos.

As indicated earlier, we assume here the existence of digital relief data for the epoch $\mathbf{T_P}$. This dataset may be in the form of contour lines or DEM. If it is in the form of contour lines, then there is a need to generate a DEM from the contour lines. Thus, two different DEM files (for epochs $\mathbf{T_P}$ and $\mathbf{T_N}$) will be available at the beginning of the revision process.

Relief revision is based on a minimum of three datasets: the $\mathbf{DEM_P}$, $\mathbf{DEM_N}$, and $\mathbf{C_P}$. The meaning of the two first datasets has been discussed earlier. The set $\mathbf{C_P}$ is generated during the revision of planimetric objects and contains a set of polygons describing the detected locations of those changes.

## 6 TECHNIQUES TO UPDATE GEOGRAPHIC DATA

As indicated in the previous section, revision of geographic data requires identifying search areas on the ground image for the epoch $\mathbf{T_N}$, and then to search those areas and to identify ground objects. Those ground objects need to be classified. If a particular object is of the same type as the corresponding vector object used to define the search area, then they need to be compared. Comparison will show if the object has changed. If the object has not changed, no additional processing of it is needed. If the object has changed, the magnitude of the changes needs to be evaluated. If they are greater than the corresponding positional standard or if their nature has changed (for example, a dirt road is now a paved road), then the newest version of the object goes into the geographic database replacing the previous one.

A similar process is repeated for each object stored in the geographic database for the epoch $\mathbf{T_P}$. As a result of this process, those objects no longer existing on the ground are identified and eliminated from the geographic database. Then, every current object is evaluated for irregularities and geometric, logical, and cartographic rules are applied to them. This results in the identification of some (or all) new ground objects. Finally, an operator checks the validity of the results and adds, removes, and modifies the geographic database. All of this solution is based on the “Human in the Loop” paradigm. And, as indicated elsewhere, our goal is to develop a highly automated solution for the revision of geographic data.

Figure 2 shows the major implementation steps of the solution tested at the Center for Mapping.

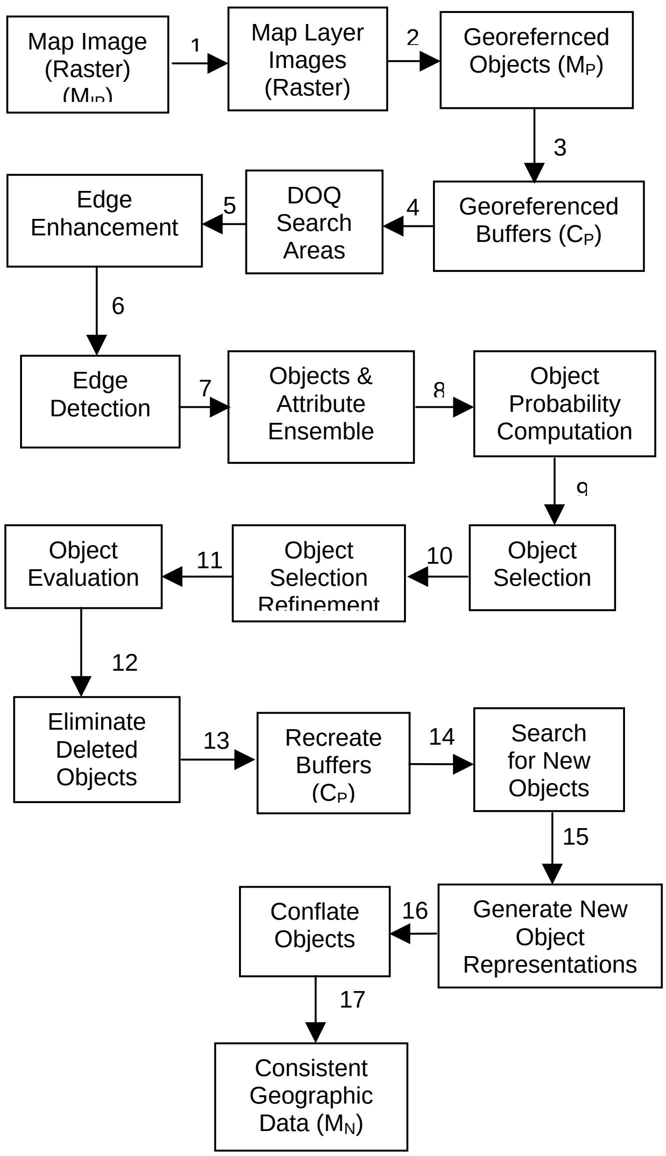


Figure 2. The Revision Model

Steps 1 and 2 are optional. If there is no digital vector data $\mathbf{M_P}$, then these two steps will be used to generate it. Step 1 separates the digital image into different classes (coverages) such as roads, water, buildings, etc.

Step 2 generates attributed vector representation for the objects of each class. These will constitute the $\mathbf{M_P}$ dataset.

Step 3 generates the boundaries of the search areas. Boundaries are generated using the georeferenced information, the attribute data, and an operator. A major component of the system is Machine Learning, which is used to train the system. The operator starts by selecting some buffer areas, and the system computes a set of parameters based on the gray-level intensity, geometric and logical characteristics, attributes, and coordinate values, and uses them later to select automatically more buffer areas. The operator looks at the results and accepts or rejects part or all of them. The system improves its decision-making capabilities from the operator's decisions. The output is a set of boundary values describing the different search areas.

Step 4 uses the values from Step 3 to isolate the areas of the images where each particular object will be searched. The goal of this step is to simplify the automatic search by knowing in advance the characteristic of the object. The output is a list of pixels per object where the search will be conducted.

The next step, Step 5, uses nonlinear filters to enhance each raster region (Liu, Wang, and Ramirez, 1998). Generally, a smoothing filter is used to make a local gradient more reliable. The most common filter used is the Gaussian. This results in blurred edges while reducing the noise. At the Center for Mapping we have developed a nonlinear smoothing filter based on orientation-sensitive probability measure. This filter is very robust and image enhancement results are very satisfactory.

Step 6 is edge detection. We use Multi-Scale Adaptive Segmentation techniques (McCane and Caelli, 1997). As indicated by Stassopoulou, Caelli, and Ramirez (1998), "the basis of the adaptive, multi-scale segmenter is to use a recursive, hierarchical multi-scaled procedure where a region at one scale is segmented at a finer scale if the variance between regions at the finer scale accounts for a significant amount of the variance of the region at the original scale." This method uses Canny edge detector for edge extraction. This is followed by the generation of closed regions at different scales. A region is replaced by those regions at a finer scale based on the evaluation of its variance. If the variance of the original region can be accounted for, or modeled, by the between variance of the regions at the finer scale, then the region is replaced.

The goal of Step 7 is the ensemble of objects and their attributes. The results from Step 6 (the closed objects) are evaluated by the operator as part of the machine learning process and attributes are added to each object. This is done initially by the operator, and later by the system.

Step 8, Object Probability Computation, uses Bayesian Network techniques to compute the conditional probability of each object. We have implemented a prototype Bayesian Network for buildings. The factors used to evaluate buildings in our prototype system are:

(1) Rectangularity of the object.
(2) Fitting of polygon area and corresponding segmented region.
(3) Area of the region.
(4) Slope of the region.
(5) Material of the region.
(6) Solar Azimuth.
(7) Sun Elevation.
(8) Average intensity of the region.
(9) Road adjacency.
(10) Road Orientation.
(11) Shadow presence.

The outcome is the conditional probabilities for each building. Bayesian networks for other types of objects will be developed in the future.

In the Step 9, objects are selected. Fundamentally, those with a high probability are selected as belonging to the class of objects in consideration, and those with low conditional probability are rejected. We are still investigating the limits for high and low probabilities. In general, we believe that objects with a conditional probability of .80 or greater can be considered as having a high probability. Objects with a conditional probability of .40 or less may be considered as having a low probability, but this is an ongoing debate. Objects with an intermediate probability (for example, with probabilities between .40 and .80) are evaluated in the next step.

Step 10 re-evaluates those objects with intermediate probabilities. The new evaluation is done by considering the influence and effect of those objects accepted or rejected in the previous step. Once the new conditional probabilities are computed, objects are selected or rejected. Those still with intermediate probability are marked and presented to the operator for his/her decision.

In the next step, Step 11, all the objects selected in Steps 9 and 10 are evaluated with respect to the existing data in the set $\mathbf{M_P}$. Evaluation is done from two different perspectives:

(1) Positional accuracy: Has an object changed its location on the ground above the permissible amount?
(2) Object Class: Does a particular ground object belong to the same class that it belonged to before?

The results of the answers to these questions are decisions about keeping or replacing an object from the geographic database.

In Step 12 objects no longer on the ground are removed from the geographic database. These objects are recognized as a result of comparing the set $\mathbf{M_P}$ with the objects selected in Steps 10 and 11.

Step 13 recreates the boundaries of the search areas based on the current information about ground objects that have not changed, the new locations of objects that have changed, and the removal of those objects no longer on the ground.

In Step 14 the data generated in Step 13 is used to search for new ground objects. As indicated before, we look for irregularities among the recognized ground objects. An example of these irregularities is the widening of a road at a point of its path. The junction with a new road may cause this type of irregularity.

Step 15 uses the techniques described in Steps 6, 7, 8, 9, and 10 to generate new ground object representations. These data are added to the geographic database and the techniques described in Steps 13 and 14 are used with these new features.

Step 16 processes features of a particular class and features of different classes to produce a consistent representation of ground objects. The result is a consistent geographic dataset $\mathbf{M_N}$ as shown in Step 17.

## 7 INTEGRATION OF OBJECTS

A very important part of the revision process described in this paper is the integration of the geographic data into a consistent image of the area represented. Step 16 deals with this issue and the purpose of this section is to give an extended explanation of our approach toward data integration.

Fundamentally, our approach for geographic data revision is based on individual ground-object evaluation. A given ground object is compared with its previous representation in the set $\mathbf{M_P}$ and actions are taken based on that comparison. A predictable result of this approach is that when all objects represented are analyzed as part of an holistic representation, many irregularities and mistakes are found.

Objects need to be integrated at two different levels:

(1) At the class level. For example, all roads need to be made consistent; all boundaries need to be made consistent.
(2) At the global level. All classes need to be made consistent. For example, roads, relief, water, and boundaries must be made consistent.

Data integration is known as conflation. Saalfeld (1993) defines conflation as, “the compilation or reconciliation of two variants of a map of the same region.” In our case, for the Level 1 above, we consider the dataset of the existing invariant objects of a class as one of the two sets; the other set is the one formed by the modified and new object representations belonging to that specific class. For Level 2, we consider the sets of two different classes to start the process, and after that the conflated class set and one additional class set are considered as the datasets to be conflated.

Conflation can be accomplished by “rubber-sheeting.” As indicated by Saalfeld (1993), “rubber-sheeting helps bring matched features of two versions of the same map into exact alignment. This alignment serves two purposes: (1) it unclutters the representation by showing a single representation for features that have been matched; and (2) it brings candidates for matching into proximity for easy decision making.” In our particular case, an additional complexity to the problem is the fact that classes of objects do not have common features and may not even have explicit common points. This requires the use of raster images (digital orthophotos) as an additional/complementary data source. Another major problem in our case is the fact that we are interested in the “total revision” concept. This requires generating a consistent representation of the ground objects and the relief. Because of this, geometric — and/or topologica l— based conflation is not enough to generate a consistent integration of geographic data. Our goal is to develop a conflation theory based on geometric, topologic, contextual, cartographic and natural constraints to integrate the revised geographic data.

As part of the prototype system currently in development at the Center for Mapping, only geometric and topologic constrains are implemented, but raster images are used as an additional data source. We are planing to extend this research to incorporate contextual, cartographic, and natural constrains.

## 8 CONCLUSIONS

We have presented a framework for revision of geographic data as an independent process of map production. This framework integrates ideas from cartography, photogrammetry, image understanding, and belief networks. The framework, in general, is independent of the scales of the maps, although it has been illustrated with examples related to the 1:24,000 USGS quadrangle maps and their digital equivalent, the DLG files.

A prototype system has been built as part of a National Imaging and Mapping Agency (NIMA) sponsored research project. The prototype system has been used to revise building, water, roads, and vegetation objects.

## REFERENCES

Liu, X, Wang D., Ramirez, J.R., A Nonlinear Smoothing Framework Using Orientation-Sensitive Probability Measure and its Application in Robust Edge Detection.

McCane, B., Caelli, T., 1997 Multi-Scale Adaptive Segmentation Using Edge and Region-Based Attributes, Proceedings of the First International Conference of Knowledge-Based Intelligent Electronic System, pages 72-81.

Ramirez, J.R., 1996, Spatial Data Revision: Current Research and its Influence in GIS, Proceedings of Position, Location, and Navigation Symposium, Atlanta, Georgia, pages 137-144.

Saalfeld, A.J., 1993, Conflation: Automated Map Compilation, Report IRI-9017393, Computer Vision Laboratory, University of Maryland, page 7.

Stassoupoulou, A., Caelli, T., Ramirez, J.R., 1998, Building Detection Using Bayesian Networks

Thompson, M.M., 1987, Maps for America, U.S. Department of the Interior Geological Survey, pages 96-107.