Integrating 2D topographic vector data with a Digital Terrain Model – a consistent and semantically correct approach

Andreas Koch, Christian Heipke

Institut für Photogrammetrie und GeoInformation,Universität Hannover,
Nienburger Str. 1, 30167 Hannover, Germany

Abstract

The most commonly used topographic vector data are currently two-dimensional. The topography is modelled by different objects; in contrast, a digital terrain model (DTM) is a continuous representation of the Earth surface. The integration of the two data sets leads to an augmentation of the dimension of the topographic objects, which is useful in many applications. However, the integration process may lead to inconsistent and semantically incorrect results.

In this paper we describe recent work on consistent and semantically correct integration of 2D GIS vector data and a DTM. In contrast to our prior work in this area, the presented algorithm takes into account geometric inaccuracies of both, planimetric and height data, and thus achieves more realistic results. Height information, implicitly contained in our understanding of certain topographic objects, is explicitly formulated and introduced into an optimisation procedure together with the height data from the DTM. Results using real data demonstrate the applicability of the approach.

1. Introduction

Applications of geographic information systems (GIS) increasingly need consistent topographic data containing planimetric and height information. Examples include visualisation in terms of true orthophotos and photorealistic perspective views, e. g. for navigation purposes, environmental simulations and traffic safety applications, in which a road must be adequately modelled in three dimensions in order to predict the forces acting on a car during turns. Checking the consistency between planimetric and height data is also useful to assess data quality.

Historically, planimetric and height data do not share many similarities: they have been modelled differently, they have been acquired using different techniques, at different times and resolutions, and for different purposes, and they are stored using different data structures. Therefore, based on existing topographic data bases the required consistency can in general not be guaranteed. Since it is neither desirable nor economically feasible to acquire a completely new, consistent data set, data integration techniques must be developed that meet the described requirements using existing data, i. e. two-dimensional topographic vector data and digital terrain models (DTMs). In many countries such data are being or will be provided by the respective National Mapping Agencies as part of the reference geoinformation. As a side note, we recall, that data integration techniques were also applied in topographic paper maps, of course in a manual fashion: for example, height contour lines cross roads perpendicular to the driving direction, and river beds are usually visible in the contour lines.

Besides consistency, also correctness with respect to gravity and construction principles and manuals must be ensured, the latter is relevant for man-made objects only. This correctness, which depends on the object class label, is termed *semantic correctness* in this paper. To give a few examples, (a) inland water bodies can be considered to be horizontal, if we neglect wind, water currents and local gravitational differences; (b) rivers have a monotonous slope, since water flows downhill only; and (c) roads have constant width, and limited curvature and slope, since otherwise they could not fulfil their function, namely to ensure safe traffic movement.

The integration of two-dimensional topographic GIS data and DTMs has been dealt with to some extent in the literature over the last decade or so. First suggestions go back to Fritsch (1991) and Weibel (1993). Pilouk (1996), Lenk (2001), and Stoter (2004) derive a TIN (triangular irregular network) data structure, in which the triangulation is constrained by using the existing vector data as edges, in addition Lenk (2001) makes sure that the surface shape of the original DTM is preserved. This geometric integration, however, does not pay attention to semantic aspects of the objects to be integrated. These are mentioned by Rousseaux, Bonin (2003), who focus on the integration of 2D linear data such as roads, dikes and embankments into a DTM. The linear objects are transformed into 2.5D surfaces by using attributes (e. g. road width) of the GIS data base and the height information of the DTM. Slopes and regularization constraints are used to check semantic correctness of the objects. However, in case of incorrect results the correctness is not established or re-established.

In this paper we propose an approach for integrating 2D topographic GIS vector data and a DTM in a consistent and semantically correct way. The approach captures the semantics in mathematical equations and inequations, the data integration problem is solved through an optimisation approach based on least squares adjustment. We build upon earlier work (Koch, Heipke 2004), the extension presented in this paper consists in a formulation where not only DTM heights are subject to change to fulfil the formulated condition, but also the planimetric coordinates of the vector data are adjusted accordingly. In the next section we present the background and the mathematical description of the new algorithm, before presenting some results using real data sets from the State Surveying Authority of Lower Saxony.

2. An algorithm for consistent and semantically correct integration

Overview

Inconsistency and semantic incorrectness between topographic GIS vector data and a DTM can in principle have two reasons: either the planimetric coordinates of the vector data or the DTM heights are incorrect. Of course, a combination of the two effects is also possible. In contrast to earlier work where we only dealt with incorrect DTM heights, we now present an approach which can deal with both types of errors.

As in the earlier work we have chosen lakes, rivers, and roads as examples for topographic objects, because all of them contain implicit height information. The objects are modelled with the help of horizontal planes (lakes, road intersection areas) and tilted planes (roads, rivers). Details about object modelling are contained in Koch, Heipke (2004) and in Koch (2006).

The data structure we use for the integrated data set is a TIN. In a first step we convert linear objects to area objects through a buffering process, where the buffer width either comes from available attributes, or a default value is used. This conversion is necessary, since in the considered resolution the topographic objects we deal with all have a certain width in the landscape and thus are considered to be area objects.

The emphasize of our current work lies on the formulation of certain condition equations and inequations for the vector data and the DTM in order to enforce consistency and semantic correctness. These constraints are taken into account in an optimisation process based on least squares adjustment. The following assumptions have guided the selection of the constraints:

- The height information contained implicitly in the topographic objects must be captured explicitly in order to be introduced into the optimisation process.
- The data sets to be integrated can contain random and systematic errors, but they do not contain any gross errors (gross errors can and should be eliminated in a pre-processing step).
- The topographic vector data is separated into man-made and natural vector data:
  - The shape of the man-made vector data (e. g. roads) is considered to be generally correct, because it follows construction principles. Therefore, their position can only be changed as a whole. We use a 2D similarity transformation for this task.
  - The border of natural vector data (e. g. lakes, rivers) can vary also locally, we therefore consider the individual border coordinates as unknowns in the adjustment.
- The shape of the terrain should be preserved as much as possible.
- Neighbourhood consistency must be taken into account.

In most cases, an integration process involves a kind of compromise. We model the fact that some of the mentioned conditions can contradict each other by assigning weights to the individual equations. It is clear that a careful selection of the weights based on the quality of the input data is of major importance for obtaining meaningful results.

After the optimisation we perform the actual integration using a triangulation based on Lenk’s algorithm (Lenk 2001).

The optimization process

In the optimisation process, the heights of the topographic objects as derived from the DTM, and the DTM heights in the neighbourhood are considered as unknowns, together with the transformation parameters of the man-made topographic objects (4 per object) and all the planimetric coordinates of the natural topographic objects. These unknowns are estimated from a set of basic observation equations in a least squares adjustment, taking into account additional equation and inequation constraints.

The basic observation equations preserve the general position of the topographic objects, the shape of the terrain, and they ensure a smooth transition between changed and non-changed areas of the data set. The constraints capture consistency and semantic correctness. Equation constraints are formulated as observation equations with corresponding weights, thus the amount to which an equation constraint is actually fulfilled can be controlled by an adequate weight selection. The inequation constraints, on the other hand, are always fulfilled after the optimisation process.

Observation equations and constraints for planimetric coordinates of vector data

Man-made objects: For man-made objects such as roads the coordinates X_i, Y_i of the border polygon are improved through a two-dimensional similarity transformation resulting in a set of new coordinates X_i^t, Y_i^t. The unknowns are the translation $\hat{X}_0, \hat{Y}_0$ and the rotation and scale parameters $\hat{a}$ und $\hat{b}$, X_S, Y_S represent the centre of gravity of the object:

$$
\begin{aligned}
X_i^t &= \hat{X}_0 + \hat{a}\left(X_i - X_S\right) + \hat{b}\left(Y_i - Y_S\right) + X_S \\
Y_i^t &= \hat{Y}_0 - \hat{b}\left(X_i - X_S\right) + \hat{a}\left(Y_i - Y_S\right) + Y_S
\end{aligned}
\tag{2.1}
$$

Points, which represent road intersections, are considered to be part of more than one road. Since for each road a separate set of equations of type (2.1) is used, this common point is lost without any further precautions. In order to preserve the topologic relationship between the roads, one constraint, formulated as an observation equation, is set up for each road ending in the intersection, where $\hat{X}_{int}, \hat{Y}_{int}$ denotes the unknown intersection point, and (as in all formulae throughout this paper) ***v*** stands for the residual of the observation equation:

$$
\begin{aligned}
0 + v &= \hat{X}_{int} - X_{int}^t \\
0 + v &= \hat{Y}_{int} - Y_{int}^t
\end{aligned}
\tag{2.2}
$$

For the remaining border polygon points X_i, Y_i of man-made objects basic observation equations to maintain the overall position are set up in the following way:

$$
\begin{aligned}
0 + v_i &= X_i^t - X_i \\
0 + v_i &= Y_i^t - Y_i
\end{aligned}
\tag{2.3}
$$

Natural objects: As mentioned above, for natural objects equations of type (2.1) are not used. Rather, individual border points can move separately, as shown in the basic observation equations (2.4). X_i, Y_i denote the original, $\hat{X}_i, \hat{Y}_i$ the unknown coordinates of the border polygon.

$$0 + v_i = \hat{X}_i - X_i$$
$$0 + v_i = \hat{Y}_i - Y_i \tag{2.4}$$

It must be ensured that despite movements of individual points of the border polygon remains an area without loops. This constraint is formulated by allowing the polygon angle α_j, which is a function of the sequence of polygon points P_{j-1}, P_j, and P_{j+1}, to change only by a small predefined amount $\Delta\alpha$ to form the resulting polygon angle α_j^*. α_j^* is a function of the unknown coordinates of P_{j-1}, P_j, and P_{j+1}, which are estimated in the optimisation procedure.

$$\alpha_j^* - \alpha_j \,|\, \le \Delta\alpha \tag{2.5}$$

This constraint can be formulated as a set of two inequations:

$$\alpha_j^* - \alpha_j \le \Delta\alpha$$
$$\alpha_j^* - \alpha_j \ge -\Delta\alpha \tag{2.6}$$

Topological aspects valid for all vector objects: Point movements can lead to different objects overlapping each other in an undesired way. We require the topology of objects to remain unchanged during the optimisation process. Fig. 1 shows an example, two objects A and B change their outline after the optimization. Fig. 1a depicts the original situation, Fig. 1b and 1c show two results, which change the topology of the objects and must therefore be avoided. Fig. 1d shows a possible point movement.

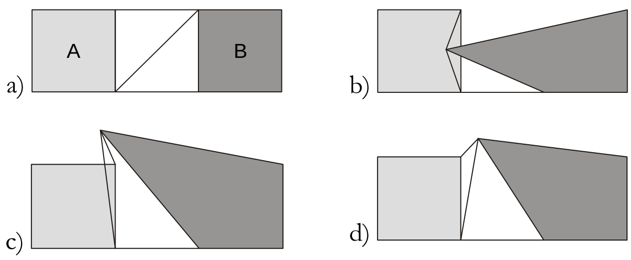

Fig. 1: Topologic relation between two objects A and B: (a) situation before optimisation, (b) and (c) invalid point movements, (d) valid point movement

If the GIS vector data are triangulated without considering the DTM points, possible and impossible situations can be separated based on an inspection of the individual triangles. In order to preserve topology the sense of orientation of the triangles connecting different objects must be maintained. This sense of orientation can be expressed by the triangle determinant D and its change dD (O'Rourke 1998). Assuming the determinant of a triangle with points P_i, P_j, P_k to be negative, the following inequation captures the constraint:

$$-dD \geq \begin{vmatrix} \left(\hat{X}_j - \hat{X}_i\right) & \left(\hat{X}_k - \hat{X}_i\right) \\ \left(\hat{Y}_j - \hat{Y}_i\right) & \left(\hat{Y}_k - \hat{Y}_i\right) \end{vmatrix} \tag{2.7}$$

For man-made objects the transformed coordinates X^t, Y^t take the place of the unknown coordinates $\hat{X}$, $\hat{Y}$.

Observation equations and constraints for height data

Observation equations, equation and inequation constraints for the heights of DTM points and the coordinates of the topographic objects were presented in detail in (Koch, Heipke 2004). Therefore, only, a short summary of these equations will be given here. The difference to our new formulation is that heights, which need to be interpolated from neighbouring points, e. g. heights for the road centre axis, are now a function of the unknown planimetric position of the point under consideration.

DTM heights are introduced as:

$$0 + v_i = \hat{Z}_i - Z_i \tag{2.8}$$

Z_i refers to the original height of the DTM, $\hat{Z}_i$ denotes the unknown height, v_i is again the residual. If the considered point is part of the border polygon of a topographic object, Z_i has to be interpolated using neighbouring height information of the DTM.

In order to be able to preserve the slope of an edge connecting two neighbouring points P_j and P_k of the DTM TIN where one is part of the polygon describing the object, and the other one is a neighbouring point outside the object (and thus to control the general shape of the integrated DTM TIN) additional equations are formulated:

$$Z_j - Z_k + v_{jk} = \hat{Z}_j - \hat{Z}_k \tag{2.9}$$

The constraints used for horizontal and tilted planes are shortly described next. Heights Z_l of all points P_l lying in the area of a horizontal plane all have the same value $\hat{Z}_{HP}$. This fact is captured through the observation equation

$$0 + v_l = \hat{Z}_{HP} - Z_l \tag{2.10}$$

Heights Z_m for points P_m (X_m,Y_m) of the border polygon of the horizontal topographic objects are interpolated from the neighbouring DTM TIN points P_u, P_v, P_w, and the height difference between the unknown object height and the interpolated height is used to formulate the constraint:

$$0 + v_m = \hat{Z}_{HP} - Z_m\left(\hat{X}_m, \hat{Y}_m, Z_u, Z_v, Z_w\right) \tag{2.11}$$

A further constraint expresses the fact that for lakes, surrounding terrain points must have a larger height Z_i than the lake:

$$0 < \hat{Z}_{HP} - \hat{Z}_i \tag{2.12}$$

As mentioned, roads and rivers are modelled with tilted planes. Points P_r on such planes must fulfil the following constraint, where $\hat{a}_0, \hat{a}_1, \hat{a}_2$ are the unknown plane parameters:

$$0 + v_r = \hat{a}_0 + \hat{a}_1\hat{X}_r + \hat{a}_2\hat{Y}_r - \hat{Z}_r \tag{2.13}$$

Roads and rivers are further constrained by requiring the slope along the object to be smaller than a certain predefined threshold. Also, roads are assumed to have horizontal cross sections, for further details see Koch, Heipke (2004).

The optimisation problem including the inequation constraints is formulated as the linear complementary problem (LCP) and solved using the Lemke algorithm (Lemke, 1968; Schaffrin, 1981; Lawson & Hanson, 1995). Since the unknowns appear in a nonlinear form, the solution can only be found iteratively. It should be noted that the number of equations may change from iteration to iteration, because due to the changes of the planimetric coordinates of the topographic objects, it may be necessary to consider different points of the neighbourhood from iteration to iteration.

As mentioned above, adequate weights must be selected for all observation equations to obtain a meaningful result: the position and height coordinates have a certain geometric accuracy, and weights should be chosen accordingly. The weights of the equation constraints must be selected according to experience. Since the inequality constraints are automatically satisfied within the algorithm, the weights for the equality constraints together with the predefined thresholds (see above) determine the degree to which consistency and semantic correctness of the integrated data set is achieved.

3. Results

In this section we present results of a consistent and semantically correct integration of real topographic vector data and a DTM. We use the German *ATKIS Basis-DLM*[1] together with the DTM *ATKIS DGM5*. The geometric accuracy of the *Basis-DLM* is approximately ±3 m, the DTM heights have a standard deviation of about ±0,5 m.

The first data sets, called *3 lakes*, consists of three lake objects with 294 planimetric polygon points, covering a relatively flat area of 450 x 650 m². The corresponding DTM contains 1.961 grid points, and in addition additional 118 points representing geomorphologic information (break lines etc.). In a pre-processing step both groups were merged using a constrained Delaunay triangulation to form a TIN. Prior to the integration, at the border to the lakes inconsistencies were clearly visible.
The results for *3 lakes* data set are indeed consistent and semantically correct. They are shown in Tab. 1 and in Fig. 2. For the main types of equations the table contains the standard deviation of the observations as well as the number and size of the resulting residuals. It can be seen that major position changes occur in planimetry. Although the shape of the objects remains more or less the same, the minimum and the maximum values of the residuals amount to three times the introduced standard deviation. From Fig. 2 it is visible that these changes occur mainly at the border polygon points. Apparently the original border points of the lake polygons lie outside the actual lake and are now moved into the water, since the water height is mainly dictated by large number of points inside the lake, which were considered to be rather accurate. While this result is consistent and semantically correct, a somewhat smaller weight for the heights inside the lakes would have probably resulted in smaller and more realistic planimetric point movements.

---

[1] ATKIS stands for Authoritative Topographic Cartographic Information system and represents the German national reference geoinformation database. The *Basis-DLM* (basic digital landscape model) contains the highest resolution and is approximately equivalent to a topographic map 1:25.000; the *DGM5* is a hybrid data set containing regularly distributed points with a grid size of 12,5 m and additional geomorphologic information.

Type of equation		Standard deviation [m]	Residuals			
			No.	Mean [m]	Min. [m]	Max. [m]
Planimetric position (2.4)	X	3.0	690	-0.47	-9.69	7.24
	Y	3.0	690	-0.28	-8.37	8.79
Heights of border polygon (2.11)		0.5	690	-0.24	-1.90	0.75
Heights outside the border (2.8)		0.5	531	-0.05	-0.35	0.95
Height differences (2.9)		2.0	3279	-0.19	-1.72	0.59

Tab. 1: Results for real data set *3 lakes*. For the main types of equations the table contains the standard deviations of the observations as well as the number and size of the residuals.

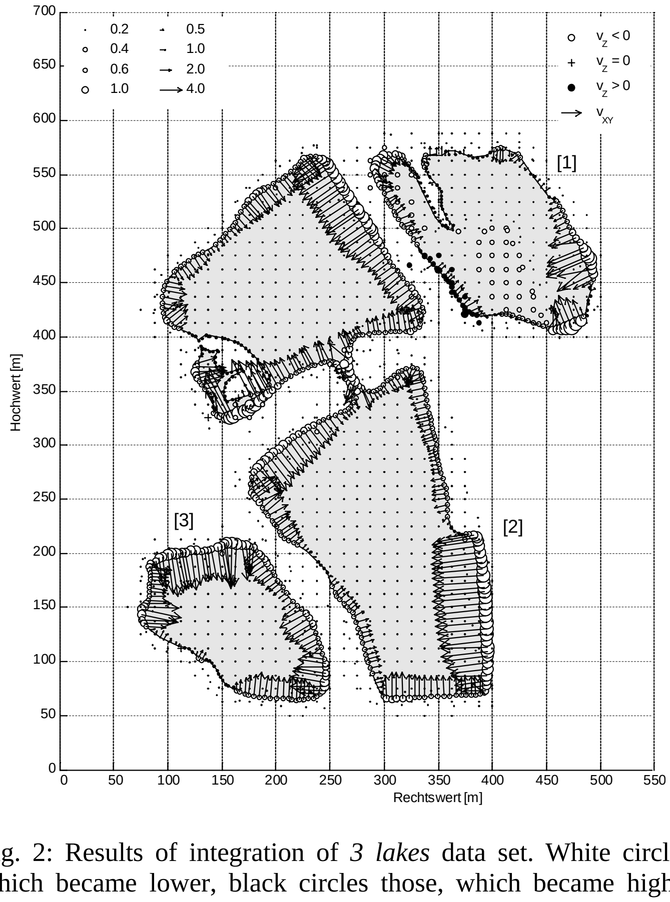

Fig. 2: Results of integration of *3 lakes* data set. White circles denote heights which became lower, black circles those, which became higher, arrows depict planimetric point movement

The lake no. 1 to the upper right of Fig. 2 shows a somewhat different behaviour than the other two. Some heights in the middle of the lake become significantly lower. The reason was found to be a break line running through the lake, which constitutes a gross error in the data set.

The second data set, called *roads*, consists of a network of 13 small roads. Most of them are connected at both ends, some are dead end roads. The data set consists of rolling terrain with height differences of about 50 m and covers an area of 575 x 400 m², it consists of 1.551 DTM grid points and 27 break lines. Before the integration, inconsistencies are clearly visible. The weights were again chosen according to the geometric accuracy of the input data, the constraints were introduced with high weights.

The results for the *roads* data set are similar to those for *3 lakes*. Again, a consistent and semantically correct result was achieved. The main changes could be observed in the planimetric position of the topographic objects, and in particular in the dead end roads in the rougher terrain. One of the points in steep terrain was moved by more than 10 m. In contrast to those roads ending in intersections, the position of the dead end roads is obviously not stabilised through equation (2.2).

4. Conclusions and outlook

This paper presents an approach for the consistent and semantically correct integration of a DTM and 2D topographic GIS data. The algorithm is based on a Delaunay triangulation and a least squares adjustment including inequality constraints derived from the implicitly available height information of topographic objects, and is solved by converting the approach into a linear complementary problem (LCP). In contrast to our earlier work, we not only adjust DTM heights, but also the planimetric position of topographic objects. Thus, vector and height data can be introduced with their respective geometric accuracy.

The approach was tested using a number of real data sets, taken from the German ATKIS. The results of two of these data sets have been presented in this paper. In all cases, a consistent and semantically correct result was achieved, which is not self understood as such, because the equation constraints are introduced as observations equations and are controlled via weight selection.

While the results are very promising, the proper selection of weights remains a difficult problem which requires some experience. Another open question is whether our approach can be transferred from the aggregation level we currently work at (ATKIS Basis-DLM) to other scales, e. g. a more detailed scale, in which e. g. consistency plays a very important role for visualisation. In addition, the geomorphologic information available in the DTM should be considered explicitly in an extended version of the algorithm. Finally, if a complete GIS data set is to be integrated with a DTM, aspects such as the propagation of planimetric changes from objects with implicit height information to neighbouring objects also need to be dealt with. These are the issues we currently work on.

Acknowledgement

We express our deep gratitude to the State Surveying Authority of Lower Saxony (Landesvermessung und Geobasisinformation Niedersachsen; LGN) for significantly supporting this research.

References

Fritsch, D., 1991: Raumbezogene Informationssysteme und digitale Geländemodelle. DGK, Reihe C, Nr. 369.

Koch, A., 2006: Semantische Integration von zweidimensionalen GIS-Daten und Digitalen Geländemodellen, Dissertation, Universität Hannover.

Koch, A.; Heipke, C., 2004: Semantically Correct 2.5D GIS Data – the Integration of a DTM and Topographic Vector Data. In: Fisher, P. (Ed.): Developments in Spatial Data Handling. Berlin : Springer, pp. 509-526.

Lawson, C. L.; Hanson, R. J., 1995: Solving Least Squares Problems. Philadelphia : Soc. for Industr. and Appl. Mathematics,, 337 p.

Lenk, U., 2001: 2.5D-GIS und Geobasisdaten - Integration von Höheninformation und Digitalen Situationsmodellen. DGK, Reihe C, Nr. 546.

O'Rourke, J., 1998: Computational Geometry in C. 2nd Ed. Cambridge : Cambridge University Press, 376 p.

Pilouk, M., 1996: Integrated Modelling for 3D GIS. PhD Thesis. ITC Publication Series No. 40, Enschede.

Rousseaux, F.; Bonin, O., 2003: Towards a coherent integration of 2D linear data into a DTM. Proceedings of the 21st ICA Conference, pp. 1936-1942.

Schaffrin, B., 1981: Ausgleichung mit Bedingungs-Ungleichungen. AVN, No. 6, pp. 227-238.

Stoter, J., 2004: 3D Cadastre. Netherlands Geodetic Commission, Publications on Geodesy 57, 327 p.

Weibel, R., 1993: On the Integration of Digital Terrain and Surface Modeling into Geographic Information Systems. Proceedings AUTOCARTO 11. Minneapolis, Minnesota, pp. 257-266.