In: Gröger/Kolbe (Eds.), Proc of the 1st Intern. Workshop on Next Generation 3D City Models, Bonn 2005

AN INTEGRATED SEMANTICALLY CORRECT 2,5 DIMENSIONAL OBJECT ORIENTED TIN

A. Koch

Institute of Photogrammetry and GeoInformation (IPI), University of Hannover, Germany
Nienburger Straße 1, 30167 Hannover –
koch@ipi.uni-hannover.de

KEY WORDS: GIS, Adjustment, Integration, Modelling, DTM

ABSTRACT:

This paper presents an approach for the homogenisation and integration of two-dimensional GIS vector data and 2,5D Digital Terrain Models (DTM). Due to inconsistencies between the data which are caused by different surveying and production methods, different modellings and large time shifts between acquiring the data, the data sets must be homogenised. The homogenisation considers the accuracy of the planimetric coordinates of the 2D GIS data as well as the accuracies of the heights of the DTM points. The aim of the homogenisation is to produce 2,5D objects represented in a semantically correct way, i.e. the objects have to correspond to our view. For instance, a locally restricted lake geometrically can be represented by a horizontal plane, i.e. the heights of all points inside the lake must be identical. To give another example, a road should not exceed maximum slope and curvature values in driving direction and across. These characterisations are formulated by equalities and inequalities and are introduced into an inequality constrained least squares adjustment which is solved using the linear complementary problem (LCP). In a further step an integration of the data is applied, which is based on a constrained Delaunay triangulation. The aim of the approach is the creation of an integrated data set containing objects with respect to the third dimension.

1. INTRODUCTION

1.1 Motivation

Topographic vector data, the core data of a geographic information system (GIS), are currently two-dimensional. The topography is modelled by different objects which are represented by single points, lines and areas. In contrast, a digital terrain model (DTM) in most cases is a 2,5D representation of the earth's surface. By integrating these data sets the dimension of the topographic objects is augmented. However, inconsistencies between the data may cause a semantically incorrect result of the integration process.

Inconsistencies may be caused by different object modelling, different surveying and production methods and / or large time shifts between acquiring the data. For instance, vector data sets often contain roads modelled as lines or polylines. Attributes contain information on road width, road type etc. If the road is located on a slope, the corresponding part of the DTM often is not modelled correctly. When integrating these data sets, the slope perpendicular to the driving direction is identical to the slope of the DTM which does not correspond to the real slope of the road. Furthermore, data are often produced independently. The DTM may be generated by using lidar or aerial photogrammetry. Topographic vector data may be based on digitized topographic maps or orthophotos. These different methods may cause inconsistencies, too. An integration and a former homogenisation considering the semantics of the topographic objects lead to consistent data.

The homogenised and integrated data set is useful for many applications. For instance, good visualisations of 2,5D or 3D models of the topography need correct data and are important for flood simulations and risk management. The approach can also be used to produce correct orthophotos in areas with non-modelled bridges or other man-made objects within the DTM. Last, considering the semantics of objects it is also possible to verify the DTM. Mostly, topographic vector data are up-to-date because objects like roads and railways possess major priority in GIS. A DTM, however, may be more than ten years old. The approach will show discrepancies between the data and will allow to draw conclusions on the quality of both data sets.

1.2 Related work

This paper covers two different working fields: the homogenisation of geographic data and the integration of different data sources. The homogenisation includes the partition of differences between the data and the realisation of geometric constraints. Hettwer (2003) and Scholz (1992) have investigated the homogenisation of 2D cartographic data which possibly stem from different data sources and refer to different coordinate systems. Both use a least squares adjustment. Scholz solves the homogenisation using an equality constraint least squares adjustment. Geometric constraints are introduced by equalities. Hettwer introduces coordinates as direct observations, regularisation constraints are formulated using pseudo observations. The investigations are restricted to 2D data and no inequation constraints are introduced.

The integration of a DTM and 2D GIS data is an issue that has been tackled for more than ten years. Weibel (1993), Fritsch & Pfannenstein (1992) and Fritsch (1991) establish different forms of DTM integration: In case of *height attributing* each point of the 2D GIS data set contains an attribute "point height". By using *interfaces* it is possible to interact between the DTM program and the GIS system. Either the two systems are independent or DTM methods are introduced into the user interface of the GIS. The *total integration* or *full database integration* comprises a common data management within a data base. The terrain data often is stored in the data base in form of a triangular irregular network (TIN) whose vertices contain X,Y and Z coordinates. The DTM is not merged with the data of the GIS. The merging process, i.e. the introduction of the 2D geometry into the TIN, has been investigated later by several authors (Lenk, 2001; Klötzer, 1997). The approaches differ in the sequence of introducing the 2D geometry, the amount of change of the terrain morphology and the number of

points after the integration process. Nevertheless, all methods have in common that inconsistencies between the data are neglected and thus may lead to semantically incorrect results.
Rousseaux & Bonin (2003) focus on the integration of 2D linear data such as roads, dikes and embankments. The linear objects are transformed into 2,5D surfaces by using attributes 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. A new DTM is computed using the original DTM heights and the 2,5D objects of the GIS data.

This paper is organised as follows: In the following section semantics in terms of 2,5D topographic objects is explained. Then, the algorithm is presented in detail in section 3. Section 4 illustrates the results which have been derived using a simulated and a real data set. The paper concludes with a short outlook related to future work.

2. SEMANTICS OF 2,5D TOPOGRAPHIC OBJECTS

If we have a look at the topographic objects of a 2D GIS, there are several objects which contain information about the third dimension. The objects include no height values, rather they contain implicit height information. For example, a lake is a standing water. This means that a locally restricted lake can geometrically be represented by a horizontal plane. The plane is bordered by a closed polygon. Outside this polygon the heights of neighbouring points have to be higher than the lake. We do not know the height of the lake but we know the shape and the relation to the neighbouring terrain. To give another example, roads are usually non-horizontal objects. We certainly do not know the mathematical function representing the road. But we know from experience and from road construction manuals that roads do not exceed maximum slope and curvature values in road direction and across. If these characterisations are not considered when integrating the data, an semantically incorrect result of the integration is obtained. Figure 1 shows two examples of an integration. In the left part three lakes are presented. The height levels inside the lakes are not constant. Primarily, at the banks it seems to be that the water is flowing up. The right part of Figure 1 shows a road network. Within the DTM the roads are not modelled correctly. Thus, the slopes and curvatures of the roads within the integrated data set have extreme values. These representations do not correspond to our view, the results are not semantically correct.
Of course, beside these two objects all other objects are related to the third dimension, too. But it is often impossible to define general characteristics of their three-dimensional shape. For example, an agricultural field can be very hilly. But it is not possible in general to define maximum slope and cuvature values because these values vary from area to area.

3. THE ALGORITHM

The algorithm consists of two steps. One of these steps includes the establishment of the semantically correctness. This means that the topographic objects have to be represented in a semantically correct way. This is achieved by means of an optimisation where the semantics is introduced using mathematical equations and inequations (see 3.2). In a second step the integration of both data sets is applied which is based on a constrained Delaunay triangulation (see 3.3). The result is an integrated triangular irregular network (integrated TIN). After the integration the object polygon points are elements of the integrated TIN and their connections are edges of it. Therefore, all triangles of the TIN situated inside an object represent the object within the integrated data set.

3.1 Preprocessing

Before starting the optimisation the topographic objects of the GIS data have to be prepared. Mostly, objects like roads and paths are modelled by lines. In this paper data sets are used which cover a scale space of 1:5.000 to 1:25.000. Due to this, the objects should be represented by areas. Thus, they have to be buffered using an attribute road or path width. Another problem is that there may be large distances between neighbouring points of the original object lines. The calculation of the 2,5D shape of the objects using the height information of the DTM without considering the terrain between these neighbouring points can lead to erroneous results of the homogenisation. Therefore, additional points are introduced between the original points at which their number depends on the density of the original DTM points.
Both data sets, the DTM and the topographic objects of the GIS data, are triangulated using a constrained Delaunay triangulation (Figure 2). The DTM-TIN is needed because within the optimisation heights of the object points have to be interpolated. The GIS-TIN is needed because the topology between different topographic objects has to be considered

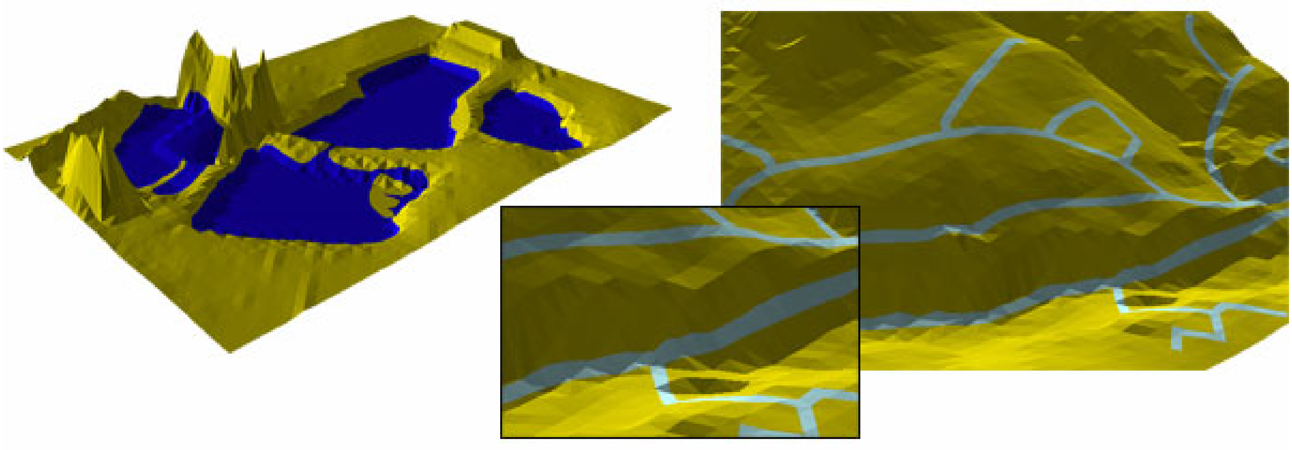

Figure 1. Results of an integration of a DTM and 2D vector data, no consideration of the semantics of the topographic objects

which is realised using inequality constraints (see chapter 3.2.3).

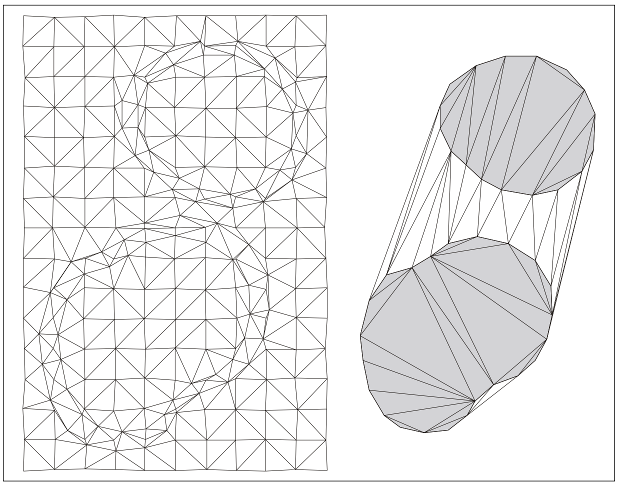

Figure 2. Left: DTM-TIN, right: TIN of the topographic objects (GIS-TIN)

3.2 Homogenisation by optimisation

In contrast to previous publications (Koch & Heipke, 2004) in this paper we also consider the accuracy of the horizontal coordinates of topographic objects. Therefore, the planimetric coordinates X,Y of the object polygons are introduced as unknown parameters within the optimisation process. Additionally, the heights of the objects and the heights of points of the neighbouring terrain have to be estimated. The aim of the optimisation is to produce coordinates which represent the topography in a semantically correct way. The semantic of the objects can be expressed by means of mathematical equations and inequations which can be formulated by observation equations and inequality constraints of an inequality constrained least squares adjustment:

$$\begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{v_c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{A_c} \end{bmatrix} \hat{\boldsymbol{x}} - \begin{bmatrix} \boldsymbol{l} \\ \boldsymbol{l_c} \end{bmatrix} \tag{1}$$

$$\boldsymbol{\Sigma}_{ll} = \sigma_0^2 \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_c \end{bmatrix} \tag{2}$$

subject to

$$\boldsymbol{b} \geq \boldsymbol{B}\,\hat{\boldsymbol{x}} \tag{3}$$

where the index c marks the observation constraints. The vectors $\boldsymbol{v}$ and $\boldsymbol{v_c}$ are the residuals, $\boldsymbol{l}$ and $\boldsymbol{l_c}$ are the reduced observations and the matrices $\boldsymbol{A}$, $\boldsymbol{A_c}$ and $\boldsymbol{B}$ contain the derivations of the observations, the observation constraints and the inequalities to the unknown parameters $\hat{\boldsymbol{x}}$. (2) is the stochastic model of the optimisation, it contains the accuracies and the accuracy relations of the observations respectively.
The optimisation is solved using the Linear Complementary Problem (LCP) (Fritsch, 1985; Schaffrin, 1981). For more details see also Koch, 2003.

In the following the observation equations, equality and inequality constraints are derived. The constraints are those observations and inequations which represent the semantics of the objects. The remarks are restricted to the object lake.

3.2.1 Basic observations: As mentioned before, the horizontal coordinates of the object polygon points are introduced as unknown parameters. The basic observations are:

$$\begin{aligned} 0 + v_{X_i} &= \hat{X}_i - X_i \\ 0 + v_{Y_i} &= \hat{Y}_i - Y_i \end{aligned} \tag{4}$$

where X_i, Y_i are the planimetric coordinates of point P_i. $\hat{X}_i, \hat{Y}_i$ are the coordinates which are estimated within the optimisation.

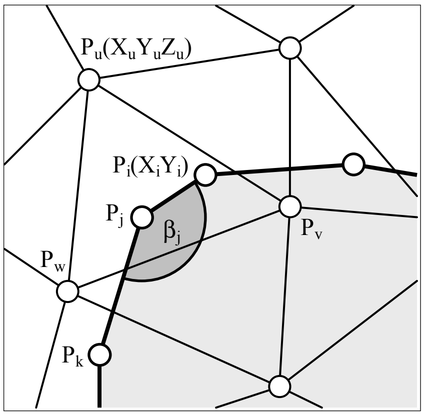

Figure 3. Part of an object lake (light grey area) bordered by its polygon points, DTM-TIN

To preserve the shape of the objects the angles between three successive points are introduced (see Figure 3):

$$\begin{aligned} &\beta_j(X_iY_iX_jY_jX_kY_k) + v_{\beta_j} = \\ &\beta_j^*(\hat{X}_i\hat{Y}_i\hat{X}_j\hat{Y}_j\hat{X}_k\hat{Y}_k) \end{aligned} \tag{5}$$

where β_j ist the angle using the original planimetric coordinates of the three object points and β_j^* ist the angle using the estimated planimetric coordinates of the three points.
The heights of neighbouring points as well as height differences between these points and points of the object polygons are further basic observations. In this case neighbouring points are original DTM points situated in a certain distance to the point of the object. They do not belong to any object considered in the algorithm. The heights are introduced as:

$$Z_u + v_{Z_u} = \hat{Z}_u \tag{6}$$

where Z_u is the height of the original DTM point and $\hat{Z}_u$ is the height which has to be estimated.

In order to be able to preserve the slope of an edge connecting two neighbouring points the height difference between these points is introduced:

$$0 + v_{\Delta Z} = (\hat{Z}_L - \hat{Z}_u) - (Z_i(\hat{X}_i \hat{Y}_i Z_u Z_v Z_w) - Z_u) \tag{7}$$

where $\hat{Z}_L$ is the unknown height of the lake. Because the GIS data are two-dimensional, the height Z_i is calculated by linear interpolation at the unknown position $\hat{X}_i, \hat{Y}_i$, i.e. the height difference between the neighbouring point and the object point at the estimated position has to be preserved.

3.2.2 Observation constraints: A lake can be represented by a horizontal plane, i.e. the heights of all points situated inside the bordering polygon of the lake as well as the polygon points must have the same height level. The points inside the polygon are introduced as:

$$Z_v + v_{Z_v} = \hat{Z}_v \tag{8}$$

The points of the bordering polygon do not contain any height information. Therefore, the height Z_i has to be estimated at the unknow position $\hat{X}_i, \hat{Y}_i$.

$$0 + v_{Z_i} = \hat{Z}_L - Z_i(\hat{X}_i \hat{Y}_i Z_u Z_v Z_w) \tag{9}$$

3.2.3 Inequality constraints: The water of a lake is bordered by the object polygon points. This means that the heights of the neighbouring points have to be higher than the lake height. This constraint can be formulated using an inequality:

$$0 > \hat{Z}_L - \hat{Z}_u \tag{10}$$

where $\hat{Z}_u$ is the height of the neighbouring point outside the object which has to be estimated.
In order to prevent overlapping objects further inequalities are introduced which consider the topology of the GIS-TIN. An edge of an object polygon has a certain relation to a point which belongs to another object. The point is located on the left or on the right from the edge. This relation is formulated by an inequality constraint representing a determinant:

$$0 > \begin{vmatrix} \hat{X}_j - \hat{X}_i & \hat{X} - \hat{X}_i \\ \hat{Y}_j - \hat{Y}_i & \hat{Y} - \hat{Y}_i \end{vmatrix} \tag{11}$$

In case of (11) the constraint is that the point P with its coordinates X, Y is located on the right from the edge whose points are P_i and P_j because the determinant has to be smaller than zero.

3.3 Integration

The integration is based on a DTM TIN computed using a constrained Delaunay triangulation (Lee & Lin, 1986, Stoter, 2004). The object borders are splitted by introducing the intersection points between the DTM TIN and the object geometries (Steiner points) and the splitted object geometries are introduced as edges of the triangulation, the result is an irregular triangular network (TIN) - an integrated DTM TIN. For the sakes of completeness the algorithm is described in the following: The triangle which contains the first object point is used and the intersection points between the triangle and the object polygon are calculated. These points and the points of the object polygon in between are connected by new edges of the integrated TIN. Then, the left and right part of the object polygon inside the triangle are re-triangulated using a polygon triangulation. In this step the shape of the triangles is recursively optimised with respect to equal angles. After processing this triangle the neighbouring triangle in direction of the object polygon is used. The first intersection point is known from the calculations before, the second one is calculated and the integrated TIN is computed using a polygon triangulation again. This process is repeated for all 2D objects. When integrating the object geometry the heights of the Steiner points have to be interpolated using the heights of the estimated height values of the object points. Thus, the shape of the objects is preserved.

4. RESULTS

The results presented in this paper were determined using synthetic and real data sets. Single objects representing lakes were investigated.

4.1 Data sets

The synthetic data consist of a Digital Terrain Model containing mass points and structure elements. The mass points are approximately situated in a regular grid, the heights of the points, representing the surface of the DTM, are calculated using a sine function. The structure elements contain information about the topographic objects. In case of a lake they represent the water line and the break line between terrain and the bank. In Figure 4 the synthetic DTM and the corresponding GIS data set containing two lakes is shown. The DTM consists of 176 mass points and 71 points representing the structure elements (white lines). Random errors were added to the planimetric coordinates of the polygon points of the GIS data. Additionally, some points of the DTM inside the structure

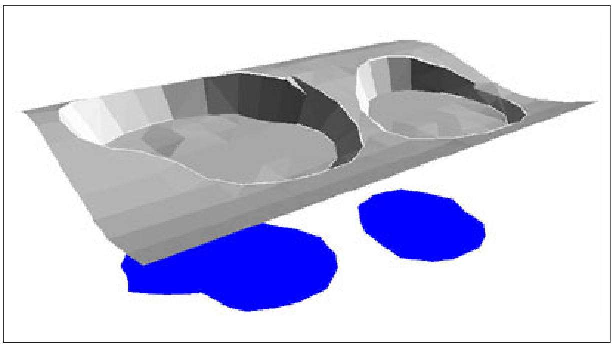

Figure 4. Synthetic data sets: DTM-TIN and two lakes of a GIS data set (blue)

element representing the water line contain random height errors.
The real data set representing lakes consists of three ATKIS Basis-DLM objects with 294 planimetric polygon points. The corresponding DTM contains 1.961 grid points and 1.047 points representing structure elements. The grid spacing is 12,5 meters.

4.2 Synthetic data

The optimisation using the synthetic data leads to 444 observations (basic observations and equation constraints). Additionally, 81 inequation constraints represent the relation between the neighbouring terrain and the lake heights and the two-dimensional neighbourhood relation between both objects, respectively. Against this, 177 parameters have to be estimated. During the preprocessing step just three points were added to the bordering polygons of the lakes because of oversized distances between neighbouring points of the objects (see chapter 3.1).
First investigations were carried out using the same accuracy for all observations. This means that the covariance matrix of the observations and consequently also the weight matrix are unit matrices. The result is semantically correct. The heights of all DTM points situated inside the bordering polygon of a lake are identical because just one unknown height per lake was introduced. Additionally, the terrain outside increases, i.e. the heights of all neighboured points outside the objects are higher than the estimated lake heights.

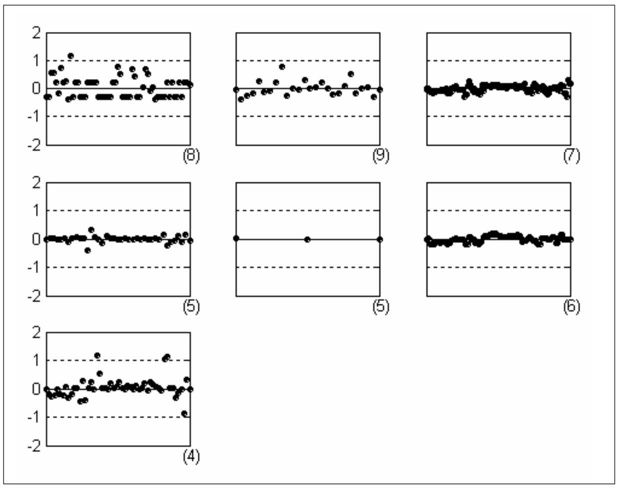

Figure 5. Different groups of residuals (see text for details)

Figure 5 shows different groups of residuals. Each graph represents one group of observations which obtain one accuracy value and one weight respectively. Beneath the graphs the corresponding observation equation (basic observation, equation constraint) is specified. For instance, graph (8) are the residuals which arise from equation (8). Here, approximately half of the residuals of this group are negative and their absolute values are nearly identical. This is because some of the points have identical height values which are systematically higher than the mean value of the other points inside the structure elements representing the lake. Graph (5) represent the angle between three successive points of the bordering polygon of the object. In Figure 5 there are two graphs representing this observation equation. This is because it is possible to specify different accuracies for the angles in original points of the object and in points which arise from the preprocessing step of the approach (see chapter 3.1). The biggest residuals of this group are nearly -0,4 and +0,3 which correspond to -23 and +17 degree respectively. This means that the object geometry is not preserved. Just a small number of points have new planimetric coordinates which differ more than half a meter from their original one.
Introducing higher accuracies for the observation equation (5) leads to objects whose geometry is as before, i.e. the residuals of this group of observations are all nearly zero. Additionally, the changes of the planimetric coordinates of the object polygon points are larger than before. This means that the position of the objects have changed over a wieder area. Also, the observation equation (4) and (7), the height differences as well as the heights of neighbouring points, lead to larger residuals. But the estimated lake heights are nearly the same as before.
A better accuracy of 0,01 and therefore a higher weight for the observation constraint (8) does not change the results of the optimisation. This is because the estimated lake heights of both investigations before are nearly identical to the mean value of the heights whose points are situated inside the bordering polygons of the objects.

4.3 Real data

It is important to know how accurate are the input data to obtain a realistic result of the optimisation process. Mostly, this information is missing. Just global values for all DTM points of one region whose data were acquired by a specific method exist. Additionally, a global value for the accuracy of all planimetric coordinates is available. This reduced information makes it impossible to get a result which can be used to update the data bases of the topographic data.

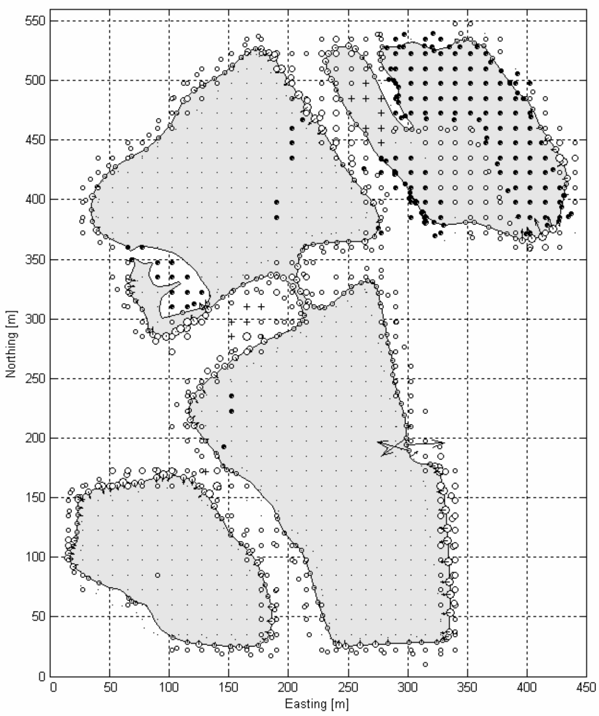

Figure 6. Position and absolute values of residuals

In the past structure elements representing the water line were often acquired by terrestrial measurements. The lake height, i.e. the heights of all DTM points inside the lake got the same height value. Thus, the DTM in these regions is of high accuracy and the observation equation (8) representing the heights inside the lake got an accuracy of 0,01 meter. The heights of the bordering polygon are derived by linear interpolation. These values are less accurate, which got a value of 1 m. The planimetric coordinates of the lake points have a global accuracy of 3 m. To prevent large geometric changes of the objects observation equation (5) also got a high accuracy of 0,01.

Figure 6 shows the residuals obtained by the optimisation using the mentioned accuracies. The white circles represent negative residuals of the heights, the black one are positive residuals. The arrows represent changes of the planimetric positions of the lake points. The crosses present positions where the residuals are zero. The larger the circles and the arrows are the larger are the absolute values of the corresponding residuals.

Obviously, the north-eastern lake contains large residuals with alternating signs. The results point out gross errors in the data sets. A large part of the lake of the GIS data are systematically higher than most of the other points inside the lake. A high accuracy of these values lead to an estimated lake height which is nearly the mean value of the heights of all points inside the lake. In the bank the heights are reduced, i.e. the residuals are negative. Just a small number are positive. These values are caused by the inequation constraint (10). A 2,5D view of the result is given by Figure 7.

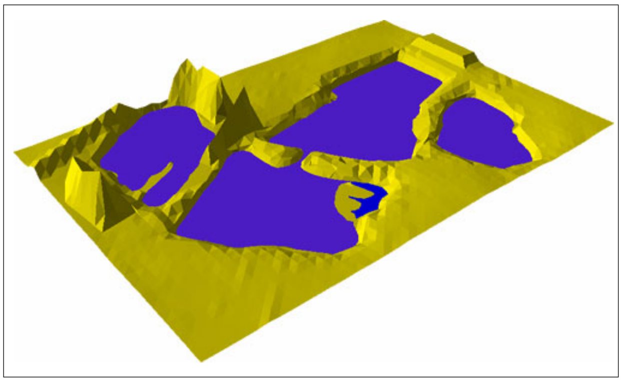

Figure 7. Result of a homogenisation and an integration

5. OUTLOOK

This paper presents an approach for the homogenisation and integration of a 2,5D DTM and 2D GIS vector data. In contrast to previous publications we also consider the accuracy of the horizontal coordinates of topographic objects. Therefore, the planimetric coordinates of the objects are introduced as additional unknown parameters within the optimisation process. First investigations were carried out using the object lake. The results are satisfying, the integrated data are semantically correct. But they depend on the information about the quality of the data. If this information is not available we have to draw conclusions about the accuracy using information about the acquisition method and other important facts.

In the future other objects have to be investigated. Primarily, the weights of the different observations are much more important for other objects because the correctness of the data set depends on the weighting of different groups of observations.

ACKNOWLEDGEMENT

This research was supported by the surveying authority of Lower Saxony Landesvermessung und Geobasisinformation Niedersachsen (LGN). We also express our gratitude to LGN for providing the data.

REFERENCES

Fritsch, D., 1985: Some Additional Information on the Capacity of the Linear Complementary Algorithm. In: Grafarend, E. ; Sanso, F. (Eds.): Optimisation and Design of Geodetic Networks. Berlin : Springer, pp. 169-184.

Fritsch, D. ; Pfannenstein, A., 1992: Comceptual Models for Efficient DTM Integration into GIS. In: Proceedings EGIS'92, Third European Conference and Exhibition on Geographical Information Systems. Munich, pp. 701-710.

Klötzer, F., 1997: Integration von triangulierten digitalen Geländemodellen und Landkarten. Diploma thesis at the Institute of Informatics, Rheinische Friedrich-Wilhelms-Universität Bonn, unpublished.

Koch, A., 2003: Semantically correct integration of a digital terrain model and a 2D topographic vector data set. Proceedings of ISPRS workshop on challenges in geospatial analysis, integration and visualization II, Stuttgart, CD.

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.

Lee, D.-T. ; Lin, A.-K., 1986: Generalized Delaunay triangulations for planer graphs. Discrete Comput. Geom., No. 1, pp. 201-217.

Lemke, C. E., 1968: On complementary pivot theory. In: Dantzig, G. B., Veinott, A. F. (Eds.): Mathematics in the Decision Sciences, Part 1, pp. 95-114.

Lenk, U., 2001: - 2.5D-GIS und Geobasisdaten - Integration von Höheninformation und Digitalen Situationsmodellen. Wiss. Arb. Fachr. Verm. Universität Hannover Nr. 244 and DGK bei der Bayer. Akad. D. Wiss., Reihe C, Nr. 546. Diss., Univ. Hannover.

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

Schaffrin, B., 1981: Ausgleichung mit Bedingungs-Ungleichungen. AVN, 6, pp. 227-238.

Stoter, J., 2004. 3D Cadastre. NCG, 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 11th International Symposium on Computer Assisted Cartography (AUTOCARTO 11). Minneapolis, Minnesota, pp. 257-266.