

Geodatenintegration – Semantischer und geometrischer Abgleich von Vektordaten

Tobias Dahinden, Joachim Boljen, Nicole Ruhe, Sagi Dalyot, Malte Jan Schulze und Monika Sester

Zusammenfassung

Durch den Ausbau der Geodateninfrastruktur und der selektiven Verwendung von Geodaten kommt dem geometrischen Abgleich dieser Daten eine wachsende Bedeutung zu. Mit Zuordnungsverfahren und einem semantischen Abgleich können korrespondierende Objekte gefunden werden, die aneinander angepasst werden müssen. Für die geometrische Anpassung ist ein Ausgleichungsansatz nach bedingten Beobachtungen besonders geeignet. Hiermit können neben der Forderung nach Korrespondenz der Punkte auch noch weitere Bedingungen integriert werden, die für die Objektgeometrien gelten müssen. Zur automatischen Identifikation korrespondierender Geometrien werden verschiedene Kriterien vorgestellt. Die Möglichkeiten der Methode werden an realen Beispielen aufgezeigt.

Summary

Due to the development of geodata infrastructures and selective geodata application the geometrical adjustment of this data becomes more and more important. Corresponding objects may be determined with appropriate matching techniques and semantic integration. These objects have then to be adapted. The geometrical adaptation of geodata by adjustment procedures based on constraint observations provides several advantages. In order to find connections of corresponding geometries automatically, several selection criteria are presented. The proposed method is illustrated using real examples.

Schlüsselwörter: Integration, Matching, Homogenisierung, Ausgleichung

1 Allgemeines

Bei der gleichzeitigen Verwendung unterschiedlicher vektorieller Geobasis- und Geofachdaten sind die verschiedenen Datenmodelle, Granularitäten, Erfassungszeitpunkte, Zuständigkeiten und Zielsetzungen, unter denen die Daten erhoben werden, zu beachten. Die dadurch bedingte semantische und geometrische Inkonsistenz behindert sowohl die individuelle und selektive Nutzung der Daten als auch die im Rahmen einer Geodatenintegration erforderliche Aufgabenbündelung bezüglich ihrer Aktualisierung und inhaltlichen Erweiterung.

Die Interoperabilität zwischen Geodaten wird neben der Einhaltung der Standards zunehmend durch die räumliche Konsistenz der Daten bestimmt. Die Beseitigung der semantischen und geometrischen Konflikte

zwischen Daten dezentraler Quellen wird zur zentralen Aufgabe im Kontext von Geoinfrastrukturen.

Um die Geodaten abgleichen zu können, müssen zunächst Objekte einander zugeordnet, d.h. korrespondierende Objekte identifiziert werden (Matching). In einem zweiten Schritt werden die Daten geometrisch aneinander angeglichen. Dies kann stufenweise erfolgen, indem erst eine allgemeine Annäherung und anschließend eine lokale Anpassung durchgeführt werden. Ziel ist es, Objekte bzw. Objektteile, die geometrische Diskrepanzen aufweisen, aneinander anzupassen, sodass ein homogener Datenbestand entsteht. Dabei können zum einen Objekte vollständig angepasst werden (etwa die Anpassung eines Gebäudes in einen Datenbestand an das Gebäude im anderen Datensatz) oder nur partiell (z.B. die Anpassung einer Flurstücksgrenze an die benachbarte Straßenabgrenzung).

Im Artikel wird eine Methode dargestellt, mit der korrespondierende, aber unterschiedlich erfasste Objekte geometrisch mit einem Ausgleichungsansatz lokal aneinander angeglichen werden können. Dazu wird in Abschnitt 2 auf Lösungen des Problems der Objektzuordnung verwiesen. Für das weitere Vorgehen wird dann angenommen, dass nur passende Objekte aneinander zugeordnet werden. Es wird aufgezeigt, wie durch einen semantischen Abgleich geeignete Objektzuordnungen erfolgen. In Abschnitt 3 werden geometrische Diskrepanzen zwischen zwei Objekten oder Datensätzen aufgezeigt. Um die Diskrepanzen zu eliminieren, wird ein Ausgleichungsansatz nach bedingten Beobachtungen verwendet. In Abschnitt 4 wird auf einige Besonderheiten der Angleichung eingegangen. In Abschnitt 5 werden die Ergebnisse und Limitierungen der Methode anhand einiger Beispiele gezeigt und diskutiert. In der abschließenden Zusammenfassung werden die gegenwärtigen Einschränkungen erörtert, ein Ausblick auf zukünftige Arbeiten skizziert und Schwerpunkte der praktischen Anwendung angegeben.

2 Zuordnung korrespondierender Objekte

Bevor eine geometrische Angleichung durchgeführt werden kann, müssen korrespondierende Objekte in den Datensätzen bestimmt werden. Ist neben der Geometrie auch die Semantik der Objekte bekannt, so muss diese integriert berücksichtigt werden. Die Lösung für die Zuordnung ist meist speziell auf bestimmte Datensätze angepasst.

Die erste Umsetzung einer Objektzuordnung von Geodaten wird Rosen und Saalfeld (1985) zugeschrieben. Sie

kombinierten Straßendaten des United States Geological Survey mit solchen von United States Census. Dazu wurden korrespondierende Punkte anhand von Straßenkreuzungen bestimmt und die verbleibende Geometrie iterativ einander zugeordnet. Cobb et al. (1998) verwendeten zur Objektzuordnung neben der Geometrie auch Attributinformationen. Walter (1998) sowie Zhang (2009) untersuchten die Zuordnung von Daten mit unterschiedlichen thematischen Inhalten. Dabei werden in einem ersten Schritt potenzielle Zuordnungspaare gesucht und in einem zweiten Schritt die Paare bezüglich verschiedener Eigenschaften (Winkel, Richtung, Anzahl Kanten) verglichen.

Die Objektzuordnung kann über Distanzen (euklidisch, Fréchet, Hausdorff) zwischen Objekten erfolgen (Buthe-nuth et al. 2007; Siriba 2012). Die Annahme ist dabei, dass sich die Lage der Objekte nicht ändert. Die Objekte müssen aber in einem vergleichbaren Bezugssystem vorliegen. Diez et al. (2008) beschäftigten sich mit der Zuordnung von Daten mit unterschiedlichen Bezugssystemen, in die Straßennetze der Datensätze als Mengen aufgefasst werden. Basierend darauf wird die Bottleneck-Distanz der Mengen bestimmt und eine Anpassung durchgeführt.

Wenn sich die geometrische Repräsentation der Objekte in den Datensätzen entspricht, lassen sie sich durch ihre Form, Größe und Ausdehnung vergleichen (Arkin et al. 1991). Die Zuordnung lässt sich mit statistischen Tests (Mustière und Devogele 2008; Shnaidman et al. 2012) oder über Kostenfunktionen für die Anpassung (Belongie et al. 2002) bestimmen. Dabei wird oft von einem 1:1-Matching der Objekte ausgegangen. Eine Verallgemeinerung zu einem m:n-Matching liefern Li und Goodchild (2011).

Für die Homogenisierung von Geometriedaten im Kontext von Katasterdaten wurden verschiedene Ansätze vorgeschlagen (Kampshoff und Benning 2005, Gründig et al. 2007, Boljen 2010, Dalyot et al. 2012). Butenuth et al. (2007) fokussieren auf die Integration von Geofachdaten und topographischen Daten. Die unterschiedliche Genauigkeit, die den Ausgangsdaten zugrunde liegt, fließt dabei über eine gewichtete Mittelung oder eine Ausgleichung in die Anpassung ein.

Für die Zuordnung der Objekte ist es grundsätzlich wichtig, ihre semantische Ähnlichkeit zu betrachten. Dies erfordert eine Zuordnung der Datenschemata (Volz 2006). Die Bestimmung der semantischen Ähnlichkeit von Objekten kann durch Analyse der Objektartenkataloge erfolgen. In einigen Fällen dürfte die Zuordnung offensichtlich sein.

Es kann aber Fälle geben, in denen keine eindeutige Zuordnung vorliegt, etwa, wenn unterschiedliche Objektarten erfasst wurden. Unterschiede gibt es insbesondere dann, wenn die Datensätze aus unterschiedlichen Maßstäben stammen (vgl. Podrenek 2002, Kieler et al. 2009).

Die semantische Ähnlichkeit lässt sich durch eine Analyse existierender Datensätze bestimmen. Eine gängige

Tab. 1: Ähnlichkeitsmatrix der Objektarten A, B, C zu den Objektarten 15, 16, 17, 18 und 19.

	A	B	C
15	0,3	0,5	0,6
16	0,1	1	0,9
17	0,7	0,7	0,7
18	1	0,1	0,8
19	0	0,2	0,6

Methode ist, die Objekte zu überlagern und die korrespondierenden Objekte zu zählen oder ihren flächenmäßigen Anteil zu bestimmen. Auf diese Weise entstehen Ähnlichkeitsmatrizen (Tab. 1). Diese geben eine Rangfolge der Zuordnung von Objekten zwischen verschiedenen Datensätzen an (Kieler et al. 2007).

Gemäß Tab. 1 wäre eine Zuordnung der Objektart A zu Objektart 18, B zu 16 und C zu 16 optimal. Wenn der Fall eintritt, dass für ein Objekt der Objektart B sowohl ein Objekt der Objektart 17 als auch Objektart 19 als korrespondierendes Objekt infrage kommt, aber keines der Objektart 16, dann wird Objektart 17 der Vorrang gegeben.

Die Objektarten können sich bezüglich Minimaldimension und geometrischer Repräsentation in den Datensätzen unterscheiden. Daher ist es wichtig, nicht nur sich überlagernde Objektarten als ähnlich einzustufen, sondern auch solche, die möglicherweise aneinander grenzen. Dies lässt sich ebenfalls durch die eine Ähnlichkeitsmatrix ausdrücken (Stern und Sester 2011).

3 Geometrische Angleichung

3.1 Beobachtungsgleichungen

Die geometrische Angleichung von Vektordaten behandelt das Problem, zwei linienhafte oder polygonale Strukturen $\underline{x}_1$ und $\underline{x}_2$ einander anzupassen, deren Koordinaten einschließlich ihrer stochastischen Modelle $\underline{\Sigma}_{xx,1}$ und $\underline{\Sigma}_{xx,2}$ gegeben sind.

Um die Angleichung durchzuführen, wird der Ansatz von Boljen (2010) verwendet, wobei das dabei benutzte Gauss-Helmert-Modell durch eine Ausgleichung nach bedingten Beobachtungen ersetzt wird. Gemäß diesem lässt sich die Diskrepanz zwischen $\underline{x}_2$ und $\underline{x}_1$ durch die lotrechten Abstände d_{ij} der durch $\underline{x}_2$ gegebenen Punkte P_j , $j = 1, \dots, n_2$ bezüglich der durch $\underline{x}_1$ festgelegten Linien-segmente L_i ($P_a P_e$), $i = 1, \dots, m_1$ beschreiben. Das Ziel ist es, die Ausgangskordinaten mittels Ausgleichung so zu verändern, dass die Quadratsumme ihrer Verbesserungen beim Verschwinden der lotrechten Abstände d_{ij} minimiert wird. Abb. 1 zeigt den Zusammenhang von $\underline{x}_1$ (mit der Linie L_i bestehend aus dem Anfangspunkt P_a und dem Endpunkt P_e) und $\underline{x}_2$ (mit Punkt P_j) zur Querabweichung d_{ij} und Längsabweichung p_{ij} .

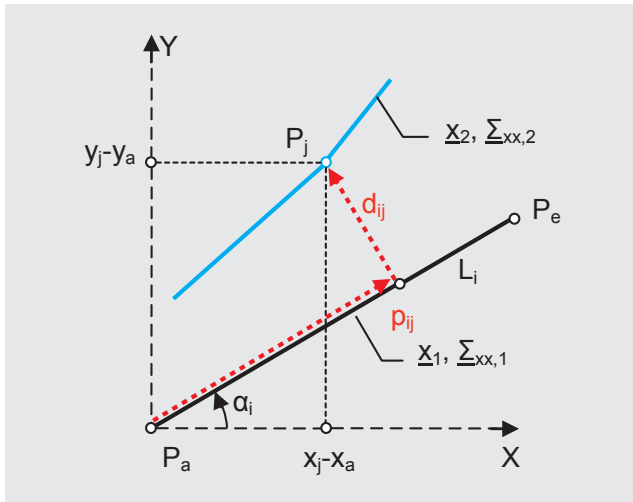


Abb. 1: Diskrepanz zwischen den Datensätzen ausgedrückt durch Längs- und Querabweichung p_{ij} bzw. d_{ij}

Für die Längs- und Querabweichung p_{ij} bzw. d_{ij} des Punktes P_j bezüglich der Linie L_i erhält man auf der Grundlage der Hesse'schen Normalform die Geradengleichung

$$\begin{aligned} p_{ij} &= (y_j - y_a) \sin \alpha_i + (x_j - x_a) \cos \alpha_i, \\ d_{ij} &= (y_j - y_a) \cos \alpha_i - (x_j - x_a) \sin \alpha_i, \\ i &= 1, \dots, m_1, \quad j = 1, \dots, n_2. \end{aligned} \quad (1)$$

Die Variable α_i steht dabei für die Richtung von L_i . Die Variable i läuft über die Liniensegmente von $\underline{x}_1$, die Variable j über die Punkte von $\underline{x}_2$.

Die auf die Querabweichungen d_{ij} bezogene funktionale Beziehung nach (1) kann für einen Ansatz nach bedingten Beobachtungen benutzt werden.

$$\begin{aligned} \underline{B}_x \underline{v}_x + \underline{w}_x &= \underline{0}, \\ \underline{\Sigma}_{ll,x} &= \sigma_o^2 \underline{Q}_{ll,x}. \end{aligned} \quad (2)$$

Leitet man die Beobachtungen d_{ij} nach den Koordinaten ab, dann erhält man als linearisierte funktionale Beziehung

$$\begin{aligned} &\left(1 - \frac{p_{ij}}{s_i}\right) \sin \alpha_i \cdot v_{x,a} - \left(1 - \frac{p_{ij}}{s_i}\right) \cos \alpha_i \cdot v_{y,a} \\ &+ \frac{p_{ij}}{s_i} \sin \alpha_i \cdot v_{x,e} - \frac{p_{ij}}{s_i} \cos \alpha_i \cdot v_{y,e} \\ &- \sin \alpha_i \cdot v_{x,j} + \cos \alpha_i \cdot v_{y,j} + d_{ij} = 0, \\ i &= 1, \dots, m_1, \quad j = 1, \dots, n_2. \end{aligned} \quad (3)$$

Dabei steht s_i für die Länge von L_i , v drückt die Verbesserungen an der jeweiligen Koordinatenkomponente aus (z.B. ist $v_{x,a}$ die Verbesserung der x-Komponente vom Punkt P_a).

Die in (3) aufgeführten Koeffizienten werden in die Matrix $\underline{B}_x$ aufgenommen. Dabei lässt sich die Matrix entsprechend der Zugehörigkeit der Verbesserungen in

zwei Untermatrizen $\underline{B}_{x,1}$ und $\underline{B}_{x,2}$ aufteilen. Die jeweilige Querabweichung d_{ij} tritt an die entsprechende Stelle des Widerspruchsvektors $\underline{w}_x$. Durch die Ausgleichung werden der jeweilige Punkt P_j und die zugehörige Linie L_i auf eine gemeinsame Gerade verschoben.

3.2 Interne Bedingungsgleichungen

Es gibt Eigenschaften, die innerhalb des jeweiligen Datensatzes $\underline{x}_1$ bzw. $\underline{x}_2$ bestehen und während des Abgleichs erhalten bleiben sollen. Von besonderer Bedeutung sind dabei Geradlinigkeit, Rechtwinkligkeit und Flächenerhalt. Die hier aufgezeigten Beziehungen gelten für die Datensätze $\underline{x}_1$ und $\underline{x}_2$ in gleichem Maße.

Liegen die Liniensegmente L_{k-1} (mit Punkt P_{i-1} und P_i) und L_k (mit P_i und P_{i+1}) auf einer Geraden, dann nimmt

$$\Delta g_i = (x_{i+1} - x_i)(y_i - y_{i-1}) - (x_i - x_{i-1})(y_{i+1} - y_i) \quad (4)$$

den Wert Null an. Um dies zu erreichen, müssen die betroffenen Ausgangskoordinaten mit Hilfe der zugehörigen Bedingungsgleichung

$$\begin{aligned} &(y_{i+1} - y_i) v_{x,i-1} - (x_{i+1} - x_i) v_{y,i-1} \\ &- (y_{i+1} - y_{i-1}) v_{x,i} + (x_{i+1} - x_{i-1}) v_{y,i} \\ &+ (y_i - y_{i-1}) v_{x,i+1} - (x_i - x_{i-1}) v_{y,i+1} + \Delta g_i = 0 \end{aligned} \quad (5)$$

verbessert werden.

Bilden die Liniensegmente L_{k-1} (mit P_{i-1} und P_i) und L_k (P_i und P_{i+1}) einen rechten Winkel, dann nimmt

$$\Delta r_i = (x_{i+1} - x_i)(x_i - x_{i-1}) + (y_{i+1} - y_i)(y_i - y_{i-1}) \quad (6)$$

den Wert Null an. Die Bedingungsgleichung lautet in diesem Fall

$$\begin{aligned} &-(x_{i+1} - x_i) v_{x,i-1} - (y_{i+1} - y_i) v_{y,i-1} \\ &-(x_{i+1} - 2x_i - x_{i-1}) v_{x,i} + (y_{i+1} - 2y_i - y_{i-1}) v_{y,i} \\ &+ (x_i - x_{i-1}) v_{x,i+1} + (y_i - y_{i-1}) v_{y,i+1} + \Delta r_i = 0. \end{aligned} \quad (7)$$

Da die Information der Geradlinigkeit bzw. Rechtwinkligkeit bei Vektordaten nicht explizit mitgeführt wird, muss der Datenbestand entsprechend überprüft werden. Dabei ist zu berücksichtigen, dass die zu tolerierenden Abweichungen nicht von der Standardabweichung der Punktkoordinaten, sondern aus formalen Gründen von deren Rundungsgenauigkeit Δk abhängig sind. Mit Hilfe der Standardabweichung σ_k der innerhalb der Grenzen Δk gleichverteilten Einzelwerte

$$\Delta \sigma_k = \frac{\Delta k}{\sqrt{12}} \quad (8)$$

kann ein multipler Test

$$|\Delta g_i|, |\Delta r_i| \leq \frac{y_{1-\alpha_0} \Delta k}{\sqrt{12}} \sqrt{\mathbf{e}_i^T \mathbf{B}_x \mathbf{B}_x^T \mathbf{e}_i} \quad (9)$$

als Entscheidungsmerkmal für die Geradlinigkeit bzw. Rechtwinkligkeit verwendet werden.

Der Flächenerhalt ist beispielsweise in digitalisierten Gebieten der Liegenschaftskarte relevant, wo sich zwischen der buchmäßig nachgewiesenen Fläche F und dem aus dem korrespondierenden Polygon der Vektordaten ermittelten Ergebnis ein Widerspruch

$$\Delta f = \sum_{i=1}^n x_i (y_{i+1} - y_{i-1}) - 2F \quad (10)$$

ergibt. Die Variable i läuft über die n Stützpunkte des Polygons (mit i modulo n). Um diesen Unterschied zu beseitigen, kann für jede einzelne Fläche die Bedingungs-

$$\sum_{i=1}^n v_{x,i} (y_{i+1} - y_{i-1}) - \sum_{i=1}^n v_{y,i} (x_{i+1} - x_{i-1}) + \Delta f = 0 \quad (11)$$

angesetzt werden.

3.3 Stochastisches Modell

In der Regel wird man sich allein schon aus Gründen des Speicherplatzbedarfs bei den in (2) aufgeführten Kovarianzmatrizen auf diagonale Strukturen und damit auf die Standardabweichungen der Koordinaten beschränken. Das bedeutet, dass man von stochastisch unabhängigen Größen ausgeht. Diese Unabhängigkeit ist bei einer gemeinsamen Entstehung allerdings fraglich. Zudem führt dann eine unterschiedliche Verteilung der Punkte zu einer Übergewichtung des dichteren Punktsatzes.

Dieser Effekt kann variabel gestaltet werden. Dazu wird die Kovarianz auf der Grundlage der Korrelation zwischen den jeweiligen Koordinatenkomponenten gemäß Boljen (2010, Gl. 7) modelliert.

4 Angleichungsumfang

4.1 Anpassungsbereich

Um die Querabweichungen d_{ij} zu bilden, ist jeder Punkt P_j gemäß (1) auf alle umliegenden Linien L_i abzubilden. Wenn davon ausgegangen wird, dass bereits eine semantische Zuordnung und eine allgemeine Annäherung stattgefunden haben, ist nur eine Angleichung von nahegelegenen Objekten sinnvoll. Deshalb werden die Abweichungen außerhalb des Intervalls

$$\begin{aligned} -p_c &\leq p_{ij} \leq s_i + p_c, \\ -d_c &\leq d_{ij} \leq d_c \end{aligned} \quad (12)$$

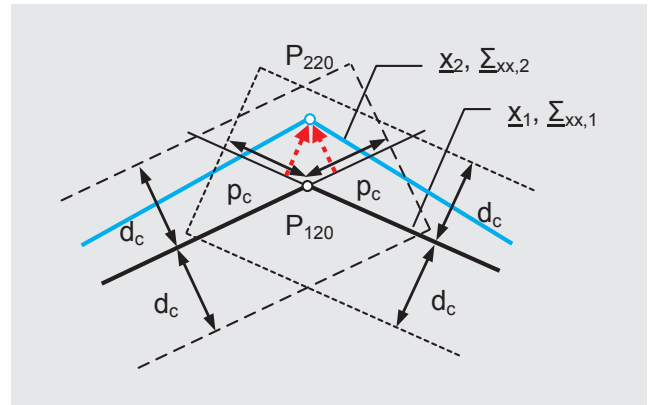


Abb. 2: Anpassungsbereich der Geometrieangleichung

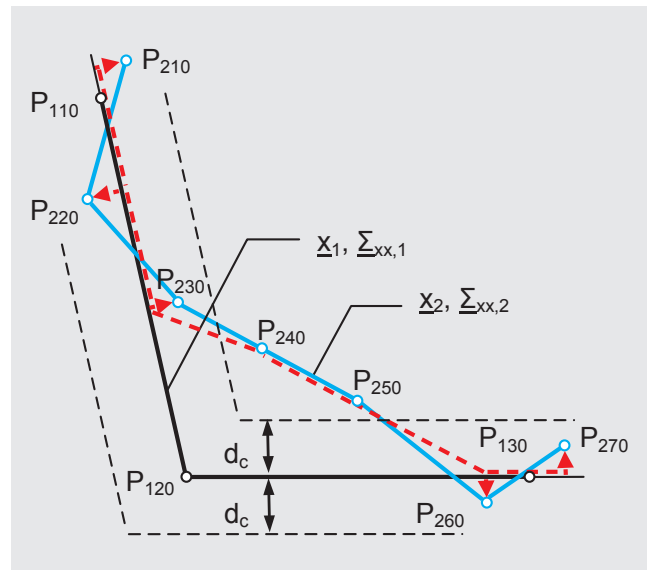


Abb. 3: Angleichung des blauen Datensatzes (x_2) an den schwarzen (x_1) innerhalb eines Anpassungsbereichs (gestrichelt). Ergebnis in rot.

von der weiteren Bearbeitung ausgeschlossen. Die Parameter p_c und d_c entsprechen der maximal zulässigen Anpassung. Abb. 2 veranschaulicht den durch (12) bestimmten Zuordnungsbereich.

Die Wirkungsweise stellt Abb. 3 dar. Dort ist x_1 in schwarz, x_2 in blau dargestellt. Die rot gestrichelte Linie zeigt das Anpassungsergebnis unter der Annahme, dass x_1 deutlich genauer ist als x_2 . Wenn die durch x_2 gegebenen Punkte Abweichungen zu x_1 aufweisen, die gemäß (12) innerhalb des Anpassungsbereichs liegen (P_{210} , P_{220} , P_{230} , P_{260} und P_{270}), werden diese bei der geometrischen Angleichung berücksichtigt. Die Punkte mit einer zu großen Abweichung (P_{240} und P_{250}) bleiben unberücksichtigt.

4.2 Strukturanpassung und Punktverdichtung

Der Approximationsgrad eines Objektes wird durch die Anzahl der konkret erfassten Punkte bestimmt. Die Beseitigung geometrischer Konflikte von einem Datensatz x_2 mit niedriger Punktdichte zu einem Datensatz x_1 mit höherer Punktdichte führt dazu, dass das Ergebnis

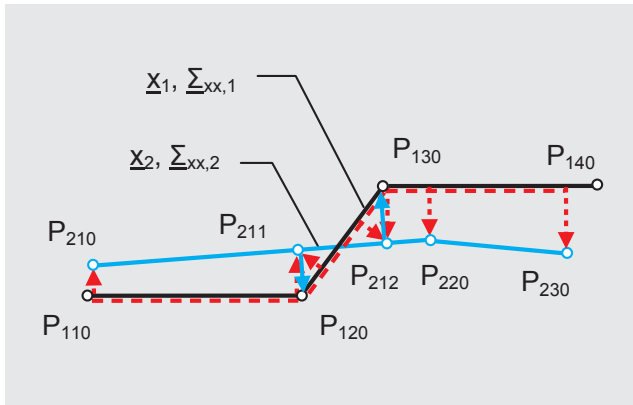


Abb. 4: Strukturanpassung durch hinzufügen zusätzlicher Punkte (P_{211} und P_{212}) an der Stelle der blauen Pfeile

durch die Ausgangsdaten mit der gröberen Struktur vorgegeben ist.

Ein derartiger Effekt kann im Sinne einer Glättung durchaus gewollt sein. Oftmals gibt es jedoch den Zwang, die Struktur des Datensatzes mit der höheren Auflösung aufrecht zu erhalten. In diesen Fällen ist die gröbere Gliederung so zu erweitern, dass sie die feinere Unterteilung des zweiten Datensatzes als Untermenge zusätzlich mit aufnimmt. Die Folge davon ist, dass die ursprünglich feinere zur gröberen und damit zur ergebnisprägenden Struktur wird.

Zur Realisierung werden alle Punkte P_i des Datensatzes $\underline{x}_1$ auf die Linienabschnitte L_j des Datensatzes $\underline{x}_2$ bezogen. Fällt die berechnete Längs- und Querabweichung in den Bereich

$$\begin{aligned} +p_c &\leq p_{ji} \leq s_i - p_c, \\ -d_c &\leq d_{ji} \leq d_c, \end{aligned} \quad (13)$$

dann ist ein zusätzlicher Hilfspunkt im Verhältnis p_{ji}/s_i einzuschalten. Durch die Festlegung von p_c bleiben die Punkte mit einer direkten Zuordnung bei der Struktur-erweiterung unberücksichtigt.

Die Punktverdichtung wird durch die Abb. 4 verdeutlicht. Um die feinere Gliederung der durch $\underline{x}_1$ (schwarz) gegebenen Daten zu erhalten, werden die orthogonalen Projektionen dieser Punkte (P_{120} und P_{130}) auf die zugehörigen Linien (blau) als weitere Zwischenpunkte (P_{211} und P_{212}) in den Datensatz $\underline{x}_2$ (blau) eingefügt (blaue Pfeile). Darauf aufbauend ergibt sich die Zuordnung für die geometrische Anpassung (rote Pfeile) der Datensätze.

4.3 Korrespondenzumfang

Es kann der Fall auftreten, dass einzelne Punkte P_j des Datensatzes $\underline{x}_2$ auf mehrere Liniensegmente aus $\underline{x}_1$ bezogen werden können. Dies ist sowohl bei Knotenpunkten, bei denen drei oder mehr Linien zusammenlaufen, als auch bei Punkten mit zwei korrespondierenden Linien innerhalb des Bereichs von p_c der Fall. In Abb. 5 betrifft dies den Knotenpunkt P_{250} und die Liniensegmente P_{220} und P_{230} .

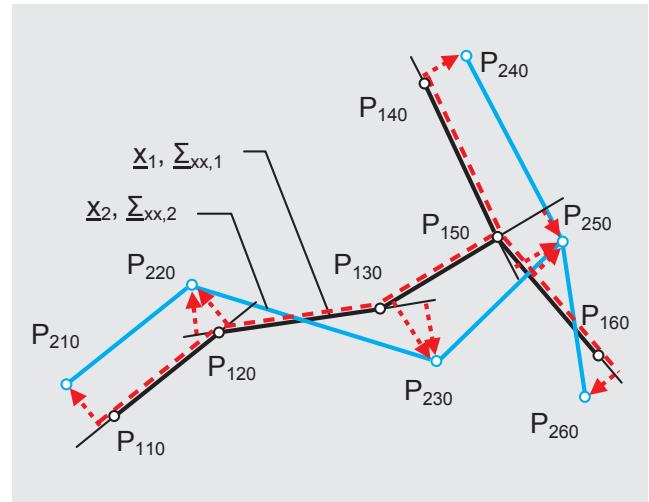


Abb. 5: Mehrfachkorrespondenzen

Die Mehrfachkorrespondenzen führen dazu, dass der betroffene Punkt P_j in Richtung des Schnittpunktes der korrespondierenden Linienzüge gezogen wird, bei einem iterativen Vorgehen kommen die Punkte zur Übereinstimmung. In Abb. 5 wird P_{220} auf P_{120} , P_{230} auf P_{130} und P_{250} auf P_{150} gezogen.

Da bereits zwei Korrespondenzen für das Verschmelzen der Punkte ausreichen, kann man sich in Knotenpunkten auf die beiden Bestimmungselemente beschränken, die sich möglichst rechtwinklig schneiden:

$$\text{abs}[\sin(\alpha_i - \alpha_k)] \rightarrow \max. \quad (14)$$

4.4 Bezugsprobleme

In detailreichen Datensätzen kann trotz vorgängiger Objektzuordnung und des in (12) gegebenen Selektionskriteriums ein einzelner Punkt P_j auf mehrere Linien bezogen werden. In diesem Fall zeichnen sich die zulässigen Kombinationen dadurch aus, dass ein Punkt P_i der Bezugsgeometrie in allen zu P_j korrespondierenden Liniensegmenten vertreten ist. Unpassende Zuordnungen können dadurch beseitigt werden, dass man den am häufigsten in $\underline{x}_1$ auftretenden Punkt P_i ermittelt und die zu P_j gehörenden Zuordnungen, in denen dieser Punkt nicht vorkommt, ausschließt.

Bezogen auf die in Abb. 6 dargestellte Situation bedeutet dies, dass die Korrespondenz $P_{221} \rightarrow P_{140}P_{150}$ eliminiert wird, da der in der Bezugsgeometrie ($P_{120}P_{130}$, $P_{130}P_{140}$, $P_{130}P_{180}$ und $P_{140}P_{150}$) des Ausgangspunktes (P_{221}) am häufigsten vertretene Punkt (P_{130}) in der Linie $P_{140}P_{150}$ nicht vorkommt. Die Korrespondenzen $P_{221} \rightarrow P_{120}P_{130}$, $P_{221} \rightarrow P_{130}P_{140}$ und $P_{221} \rightarrow P_{130}P_{180}$ bleiben erhalten.

Einige Korrespondenzfehler können nicht eindeutig lokalisiert werden, z. B. weil widersprüchliche Korrespondenzen mit der gleichen Häufigkeit auftreten. In diesen Fällen sollten die fraglichen Beziehungen insgesamt zurückgestellt werden, da fehlerhafte Zuordnungen zu einem lokalen Zusammenfallen der Geometrie führen.

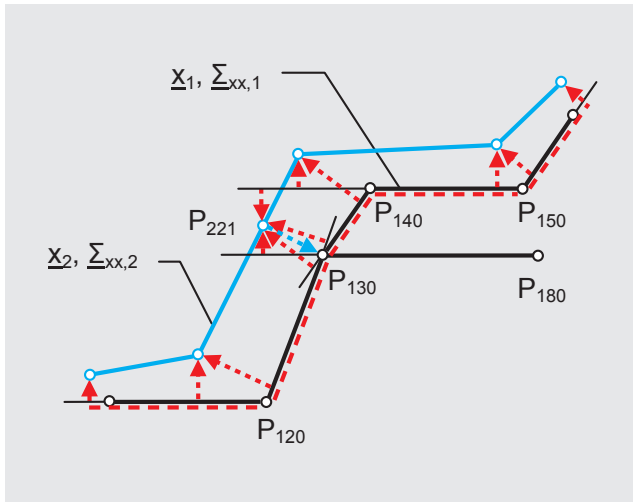


Abb. 6: Zuordnungsprobleme bei der Geometrieangepassung

Eine Vernachlässigung möglicher, aber nicht sicher erkannter Korrespondenzen (korrekte aber verworfene Zuordnungen) kann im Extremfall dazu führen, dass die Verbesserungen einzelner Koordinaten unbestimmt bleiben. Diese Lücken können durch eine Interpolation aus der Nachbarschaft geschlossen werden. Wendet man das Verfahren iterativ an, dann können Zuordnungsfehler durch schrittweise enger gefasste Zuordnungsparameter d_c und p_c letztendlich vollständig vermieden werden.

4.5 Angleichung ungleicher Geometrietypen (Linie – Fläche)

In einigen Datensätzen werden einzelne Objekte durch unterschiedliche Geometrietypen beschrieben. Dies gilt insbesondere für den Verkehrsbereich, der in ALKIS®-Daten als Fläche, in ATKIS®-Daten jedoch als Linie modelliert wird. Da keine direkte geometrische Zuordnung besteht, beschränkt sich die Anpassung in diesen Fällen auf die Einhaltung der Forderung, dass das linienhafte Element (Achse) innerhalb des flächenhaften Elements (Körper) verlaufen soll.

Für den Prozess wird eine Objektbreite b benötigt. Diese kann explizit in den Daten vorhanden sein (Trassenbreite, Gewässerbreite) oder sie wird aus der Größe des flächenförmigen Elements abgeschätzt. Zudem wird die Querabweichung zur linken d_{ij}^l und zur rechten d_{ij}^r Seite betrachtet.

Nachfolgend werden die Punkte P_j des Datensatzes $\underline{x}_2$ in analoger Anwendung von (1) auf die Linien L_i von $\underline{x}_1$ abgebildet und unter Berücksichtigung des Zuordnungsbereichs nach (12) auf plausible Lösungen beschränkt. Dabei wird jedoch der seitliche Toleranzbereich d_c um mehr als die halbe Objektbreite $b/2$ erweitert.

Da der Linienvorlauf $\underline{x}_2$ zumindest teilweise außerhalb der durch $\underline{x}_1$ gegebenen Fläche liegen kann, ist eine unmittelbare Festlegung der mit d_{ij}^l und d_{ij}^r bezeichneten linken und rechten Querabweichungen anhand der Vor-

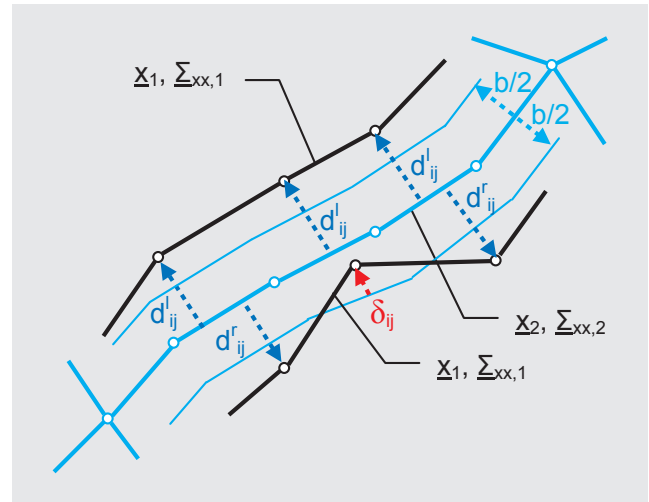


Abb. 7: Geometrieangepassung Linie – Fläche

zeichen nicht möglich. Unterstellt man jedoch, dass sich die Stützpunkte nahezu gleichmäßig auf die Ränder verteilen, dann können die Abweichungen d_{ij} , die oberhalb ihres Mittelwertes liegen, dem linken, die übrigen dem rechten Rand zugeordnet werden:

$$d_{ij}^l \geq \frac{1}{n} \sum_{i,j=1}^n d_{ij} \geq d_{ij}^r, \quad (15)$$

wobei hier n die Anzahl Verknüpfungen zwischen den beiden Datensätzen ist.

Für die Auswahl der nach (3) aufzustellenden Bedingungsgleichungen und die Berechnung der abzugleichenden Querabweichung δ_{ij} gilt gemäß Abb. 7

$$\begin{aligned} \delta_{ij} &= d_{ij}^l - \frac{b}{2} \leq 0, \\ \delta_{ij} &= d_{ij}^r + \frac{b}{2} \geq 0. \end{aligned} \quad (16)$$

Die durch (16) festgelegten Diskrepanzen δ_{ij} treten an die Stelle der in (3) aufgeführten Querabweichungen d_{ij} .

4.6 Modellangepassung

Führt die Geometrieangepassung auf Basis der ausgewählten Korrespondenzen zu einem nicht modellkonformen Ausgleichungsergebnis, dann sind die kritischen Korrespondenzen d_{ij} zu eliminieren, bis die durch

$$F_{r,\infty} = \frac{\mathbf{v}_x^T \mathbf{Q}_0^{-1} \mathbf{v}_x}{r \sigma_0^2} \leq F_{r,\infty; 1-\alpha} \quad (17)$$

formulierte globale Testentscheidung akzeptiert wird.

5 Beispiele

Im Folgenden werden einige Beispiele gezeigt, die den Ansatz anhand realer, inkonsistenter Datensätze verdeutlichen.

Der Angleichung der Daten ging eine Zuordnung voraus. Für die Beispiele wurden die Daten manuell anhand ihrer Semantik selektiert. Dadurch wurden mögliche Fehlzuordnungen eliminiert. Der vorgestellte Ansatz wurde mit verschiedenen Datensätzen getestet.

1. Angleichung von ATKIS®- und ALKIS®-Daten: Hier liegen die Diskrepanzen im Meterbereich (Abb. 8 bis 12).
2. Angleichung eines Straßennetzes eines provisorischen Katasters aus Kenia an photogrammetrisch erhobene Daten: Hier liegen die Diskrepanzen in einem Bereich bis 50 m (Abb. 13).
3. Angleichung von ALKIS®-Daten an der Landesgrenze verschiedener Bundesländer: Hier liegen die Diskrepanzen im Meterbereich (Abb. 14).

Die Prototyp-Implementierung des Algorithmus erfolgte mit Matlab.

Durch das stochastische Modell kann festgelegt werden, ob eine Geometrie präferiert und damit festgehalten wird, oder ob beide Datensätze mit ihren entsprechenden Genauigkeiten eingehen und somit eine mittlere Geometrie entsteht.

Die ersten Beispiele (Abb. 8 bis 10) zeigen die Anpassung linienhafter Elemente, die Anpassung flächenhafter Elemente erfolgt analog. Konkret handelt es sich um Vegetationsmerkmale aus ATKIS® (schwarz) und ALKIS® (blau). Der ALKIS®-Datensatz wurde als korrekt angenommen und wird nicht verändert.

Abb. 8 zeigt zwei Linien, die sich am Rand überlagern sollen, in der Mitte jedoch große Abweichungen aufweisen, die erhalten werden sollten. Als Anpassungsbereich d_c wurde 5 m gewählt.

Wie erwartet, findet die Angleichung dort statt, wo die Geometriedaten eine Diskrepanz < 5 m aufweisen. Der schwarz eingefärbte Datensatz passt sich in diesem Bereich gut an den blau eingefärbten an. Dort wo die Diskrepanz > 5 m ist, bleiben die ursprünglichen Formen erhalten.

In Abb. 9 wird der Einfluss des Anpassungsbereiches (12) dargestellt. Hier wurde der laterale Anpassungsbereich d_c auf 5 m (oben) respektive 3 m (unten) festgelegt. Die Diskrepanzen zwischen den anzupassenden Objekten liegen zwischen 2,5 m und 4,7 m.

Im oberen Teil von Abb. 9 ist der laterale Anpassungsbereich d_c für die Anpassung größer als die Diskrepanzen. Dadurch werden die Linien vollständig aneinander angeglichen. Im unteren Teil von Abb. 9 ist der Schwellwert kleiner gewählt worden als einige Diskrepanzen. Deshalb wird der schwarz eingefärbte Datensatz nicht vollständig an den blau eingefärbten angepasst.

Abb. 10 zeigt die Anpassung in einem Bereich, in dem sich die Anzahl Stützpunkte unterscheidet, es sollte also eine Strukturanpassung gemäß Abschnitt 4.2 erfolgen.

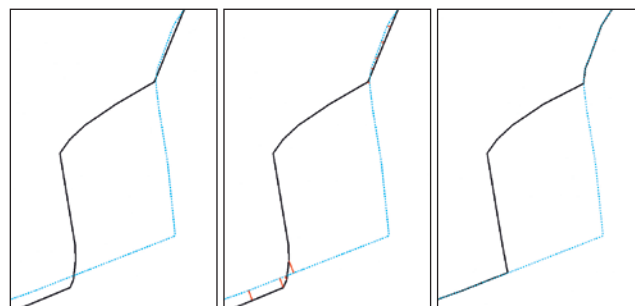


Abb. 8: Angleichung linienhafter Elemente. Links: Ausgangslage, mittig: Ausgangslage mit Zuordnungen (rot), rechts: Resultat

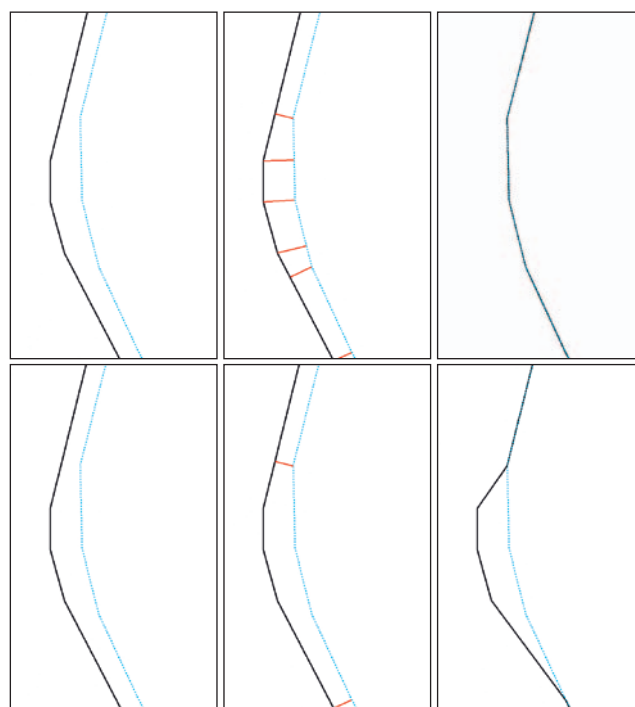


Abb. 9: Angleichung linienhafter Elemente mit Diskrepanzen < 5 m mit einem lateralen Anpassungsbereich von $d_c = 5$ m (oben) und $d_c = 3$ m (unten). Links: Ausgangslage, mittig: Ausgangslage mit Zuordnungen (rot), rechts: Resultat

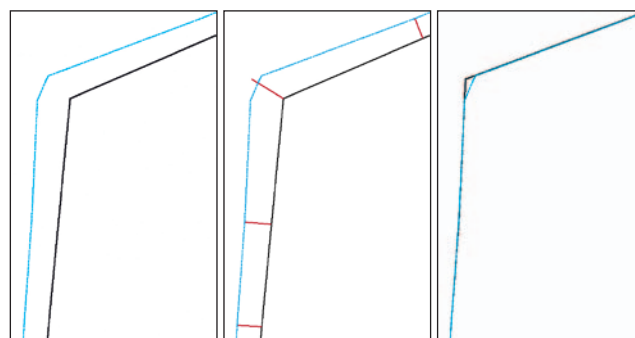


Abb. 10: Einseitige Anpassung bei ungleicher Anzahl Stützpunkte. Links: Ausgangslage, mittig: Ausgangslage mit Anpassungsvektoren (rot), rechts: Resultat

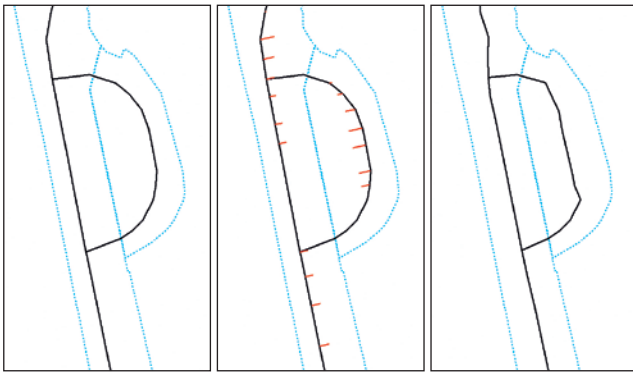


Abb. 11: Anpassung der Straßenachse (schwarz) an Straßenfläche (blau). Links: Ausgangslage, mittig: Ausgangslage mit Anpassungsvektoren (rot), rechts: Resultat

Allerdings würde ein neuer Stützpunkt innerhalb des Grenzbereiches p_c liegen. Deshalb findet gemäß (13) keine Strukturanpassung statt.

Die Geometrie wird so weit möglich zur Übereinstimmung gebracht. Im Bereich der Ecken findet zwar keine vollständige Angleichung statt, die Lösung liefert jedoch eine minimale Diskrepanz zwischen den Datensätzen. Um das Resultat aus Abb. 10 zu verbessern, könnte nun eine erneute Angleichung mit verkleinerten Grenzwerten (d_c und p_c) durchgeführt werden.

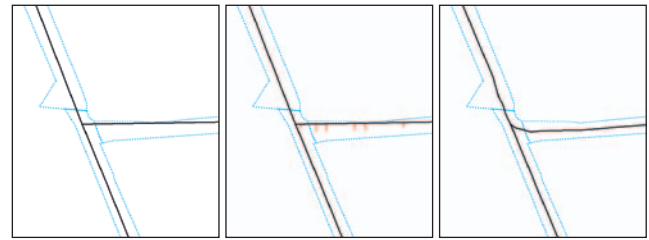


Abb. 12: Anpassung der Wegachse (schwarz) an Wegflächen (blau) bei einer T-Kreuzung. Links: Ausgangslage, mittig: Ausgangslage mit Anpassungsvektoren (rot), rechts: Resultat

Die Abb. 11 und 12 zeigen die Anpassung linienhafter Objekte an flächenhafte, wie sie in Abschnitt 4.5 beschrieben wurde. Konkret handelt es sich um Straßen- und Wegflächen aus ALKIS® und Straßen- und Wegachsen aus ATKIS®. Als lateraler Anpassungsbereich d_c wurde 5m gewählt, die Straßenbreite ist etwa 15m. In Abb. 11 wird das Verhalten bei benachbarten Flächenobjekten, in Abb. 12 die Anpassung im Bereich einer T-Kreuzung aufgezeigt.

Das linienhafte Objekt liegt innerhalb der zugeordneten Fläche. Das entspricht dem gewünschten Verhalten, da es sich bei den Linien, zumindest näherungsweise, um die Repräsentation der Mittelachse der flächenhaften

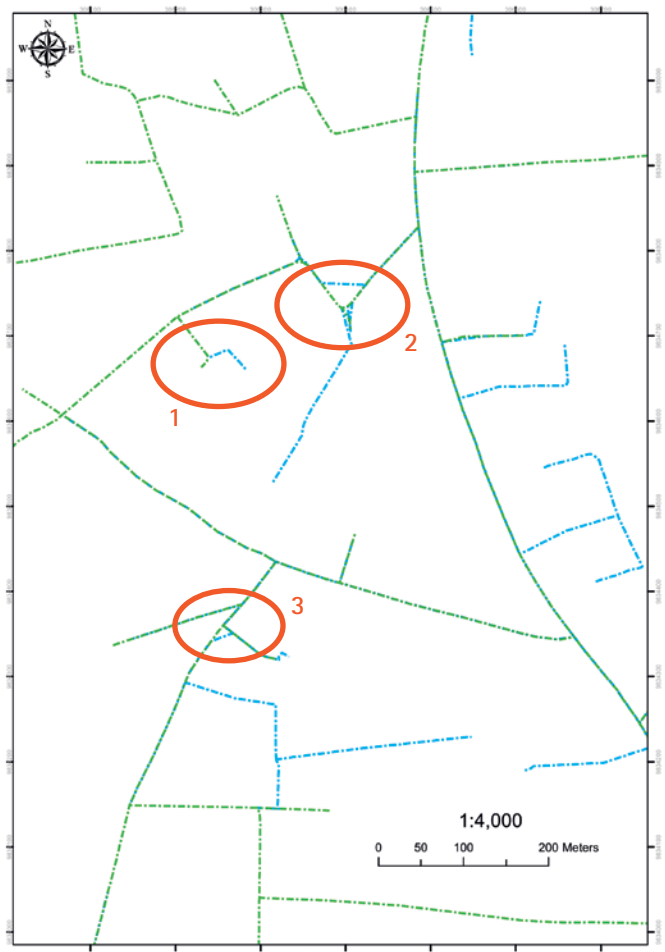
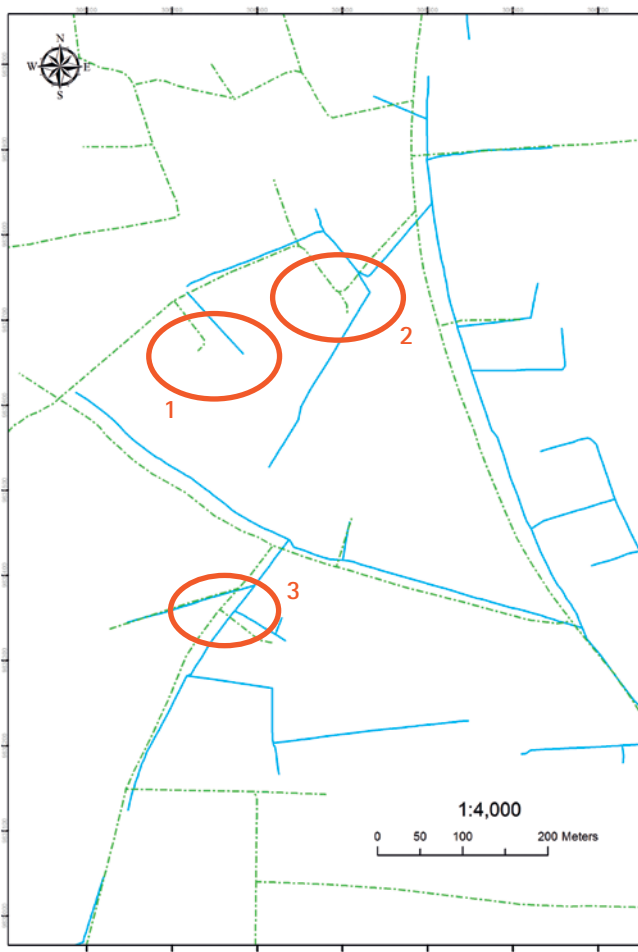


Abb. 13: Anpassung eines Straßennetzes aus einem provisorischen Kataster an photogrammetrisch erhobene Daten. Links: Ausgangslage, rechts: Resultat

Objekte handelt. Das gewünschte Resultat ergibt sich auch bei Flächenobjekten, bei denen die Hauptrichtungen schwach ausgeprägt sind (Abb. 11). Im Kreuzungsbereich (Abb. 12) bleibt die Kreuzungsstruktur erhalten. Es verändert sich allerdings die Form der waagrechten Linie. Sie erscheint leicht geknickt.

Abb. 13 zeigt die Anpassung von zwei Straßennetzen. Das eine Netz stammt aus einem provisorischen Kataster (blau), das andere wurde aus Luftbildern erfasst (grün). Die Diskrepanzen zwischen den beiden Datensätzen liegen in einem Bereich bis zu 50 m. Entsprechend wurde der laterale Anpassungsbereich d_c mit 50 m gewählt. Die Anpassung erfolgte ohne die Geradlinigkeit zu erhalten.

In den Bereichen, in denen die Zuordnung offensichtlich ist, gelingt diese gut. Hervorzuheben sind die drei eingekreisten Stellen:

Bei 1 ist zu erkennen, dass die Geradlinigkeit nicht erhalten wurde. Die blaue Linie wurde so weit möglich angepasst. Wo keine Anpassung möglich war, blieb sie an ihrer bisherigen Lage. Zwischen dem angepassten Linien- und dem bisherigen bleibt eine Verbindung bestehen. Dies führt zu einem Rechtsknick in der blauen Linie und zu einer Straßenform, die so in den Ausgangsdaten nicht vorhanden war. Eine Verbesserung ist möglich, wenn man die in der Angleichung unbestimmt gebliebenen Punkte im Rahmen einer Interpolation nachführt (vgl. Abschnitt 4.4).

Bei 2 entstehen mangelhafte Angleichungen, da diese nur lokal erfolgen: Es werden nicht sich entsprechende Kreuzungspunkte aneinander angeglichen, sondern die Punkte nur minimal auf eine Linie verschoben. Dies liegt in der relativ geringen Anzahl und Verteilung der Stützpunkte im Verhältnis zum Anpassungsbereich begründet. Abhilfe würde hier eine vorgängige, grobe Anpassung, z. B. nach Siriba (2012), schaffen.

Bei 3 wird ein Stützpunkt der vertikalen blauen Linien an die horizontale grüne Linie angepasst, da die vertikale Linie teilweise außerhalb des Anpassungsbereichs liegt.

Abb. 14 zeigt die Anpassung von zwei ALKIS®-Datensätzen im Bereich der Grenze zweier Länder. Die Diskrepanzen liegen hier im Submeterbereich. Als Anpassungsbereich d_c wurde 5 m gewählt. Die Anpassung erfolgt gegenseitig, da in beiden Datensätzen dieselben, vermarkten Stützpunkte erfasst sein dürften und für beide dieselbe Genauigkeit angenommen werden kann.

Durch die Anpassung werden die Lücken zusammengezogen. Knotenpunkte mit Mehrfachkorrespondenzen fallen wie vorgesehen auf denselben Punkt.

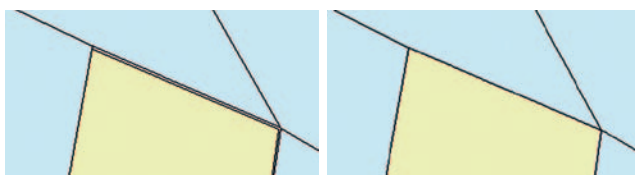


Abb. 14: Geometrische Anpassung von Katasterdaten entlang einer Landesgrenze. Links: Ausgangslage, rechts: Resultat

6 Zusammenfassung

Die vorgeschlagene Angleichung setzt eine geeignete, vorausgehende Objektzuordnung voraus. Zudem kann es sinnvoll sein, eine Grobanpassung der Geometrie vorzunehmen. Sind diese Kriterien erfüllt, bietet die aufgezeigte Methode geeignete Möglichkeiten, Daten mit bestimmten Genauigkeitsansprüchen aneinander anzugleichen. Dabei kann die Genauigkeit der Grunddaten beim Eliminieren der Diskrepanzen berücksichtigt werden. Falls die Diskrepanzen außerhalb eines Anpassungsbereichs liegen, werden diese zudem nicht eliminiert.

Eine Schwierigkeit besteht darin, diesen Anpassungsbereich zu bestimmen. Wird er zu klein gewählt, wird die Anpassung teilweise unterbunden (vgl. Abb. 9). Wird er zu groß gewählt, kollabieren einzelne Objekte wegen einer fehlerhaften Zuordnung. Um den Anpassungsbereich adaptiv bestimmen zu können, ist es notwendig, eindeutig übereinstimmende Objekte zu kennen. Von diesen Objekten kann die Diskrepanz bestimmt und als Richtwert für den Anpassungsbereich verwendet werden. Dabei wird erreicht, dass der optimale Anpassungsbereich an die konkrete Aufgabenstellung angepasst wird und lokal variieren kann.

Die Methode lässt es zu, Zusatzbedingungen wie Geradlinigkeit, Rechtwinkligkeit oder Flächenerhalt zu berücksichtigen. Damit können wichtige Eigenschaften eines Datensatzes erhalten bleiben. Eine Erweiterung wäre, den Objekten zusätzliche Formstabilität zu geben. Damit würde sich der Abgleich nicht auf einzelne Punkte und Linien sondern auf Objekte beziehen. Zudem könnten Bedingungen zwischen Objekten (z. B. Minimalabstand bei Doppellinien) eingeführt werden.

Weiter umfasst die Methode eine Möglichkeit, Objekte mit unterschiedlichem Geometrietyp aneinander anzupassen. In den Beispielen (Abb. 11 und 12) wurden dazu Straßen und Wege verwendet.

Insgesamt hilft die Methode, Geodaten aneinander anzugleichen und Informationen von einem Datensatz auf den anderen zu übertragen. Damit ist es möglich, konsistente, verlässlichere und aktuellere Daten anzubieten.

Bezüglich der praktischen Anwendung geht es mittelfristig vor allem um die schrittweise Anpassung des ALKIS®- und ATKIS®-Datenbestandes. Kurzfristig wird man sich darauf beschränken müssen, die Aktualisierungsprozesse zu bündeln. Für die tatsächliche Nutzung bedeutet dies, die Geometrie und Semantik der Fortführungsdaten übergangsweise sowohl an den ALKIS®- als auch ATKIS®-Datenbestand anzupassen. Sonderaufgaben, wie der Abgleich der Geobasisdaten an den Landesgrenzen, sind wegen der einfachen Zuordnungsproblematik bereits jetzt möglich (vgl. Abb. 14).

Mittelfristig wird man sich der Aufgabe zuwenden müssen, Geofachdaten an die Geobasisdaten anzupassen. Dabei geht es nicht nur um eine erleichterte modulare Verwendung aller Daten in einer Geodateninfrastruktur,

sondern auch um die Nutzung von Fortführungsinformationen für die Aktualisierung der Geobasisdaten.

Um die sich überlagernden Arbeitsprozesse im Interesse eines minimalen Ressourceneinsatzes zusammenzuführen, muss die benutzungsbezogene Trennung des ALKIS®- als auch ATKIS®-Datenbestands letztendlich überwunden werden. Die speziellen anwendungsbezogenen Forderungen müssen langfristig durch Selektions- und Filterprozesse aus einem integrierten Datenbestand abgeleitet werden.

Literatur

- Arkin E.M., Chew L.P., Huttenlocher D.P., Kedem K., Mitchell J.S.B.: An Efficiently Computable Metric for Comparing Polygonal Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(3), S. 209–216, 1991.
- Belongie S., Malik J., Puzicha J.: Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(24), S. 509–522, 2002.
- Boljen J.: Geometrischer Abgleich von Vektordaten auf der Basis eines Gauss-Helmert-Modells. *zfv* 135 (5), S. 314–323, 2010.
- Butenuth M., Gösseln G.v., Heipke C., Lipeck U., Sester M., Tiedge M.: Integration of heterogeneous geospatial data in a federated database. *ISPRS Journal of Photogrammetry and Remote Sensing* 62(5), S. 328–358, 2007.
- Cobb M.A., Chung M.J., Foley H., Petry F.E., Shaw K.B., Miller H.V.: A rule-based approach for the conflation of attributed vector data. *Geoinformatica* 2, S. 7–36, 1998.
- Dalyot S., Dahinden T., Schulze M.J., Boljen J., Sester M.: Geometrical adjustment toward the alignment of vector databases. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* I-4, S. 13–18, 2012.
- Diez Y., Lopez M.A., Sellarès J.A.: Noisy Road Network Matching. In: Cova J.T., Miller H.J., Beard K., Frank A.U., Goodchild M.F. (eds.), *GIScience* 2008, LNCS 5266, S. 38–54, 2008.
- Gründig L., Gielsdorf F., Aschoff B.: Merging Different Data Sets Based on Matching and Adjustment Techniques. In: *FIG Working Week*, Hong Kong, China, 2007.
- Li L., Goodchild M.F.: An optimisation model for linear feature matching in geographical data conflation. *International Journal of Image and Data Fusion* 2(4), S. 309–328, 2011.
- Kampshoff S., Benning W.: Homogenisierung von Messdaten im Kontext von Geodaten-Infrastrukturen. *zfv* 130 (3), S. 133–147, 2005.
- Kieler B., Sester M., Wang H., Jiang J.: Semantic Data Integration: Data of Similar and Different Scales. *Photogrammetrie Fernerkundung Geoinformation* 6, S. 447–457, 2007.
- Kieler B., Huang W., Haunert J., Jiang J.: Matching River Datasets of Different Scales, *Advances*. In: Sester M., Bernard L., Paelke V. (eds.), *GIScience: Proceedings of 12th AGILE Conference on GIScience, Lecture Notes in Geoinformation and Cartography*, S. 135–154, 2009.
- Mustière S., Devogele T.: Matching Networks with Different Levels of Detail. *GeoInformatica* 12(4), S. 435–453, 2008.
- Podrenek M.: Aufbau des DLM50 aus dem Basis-DLM und Ableitung der DTK50 – Lösungsansatz in Niedersachsen. In: *KS, Band 6, Kartographie als Baustein moderner Kommunikation*, S. 126–130, 2002.
- Rosen B., Saalfeld A.: Match Criteria for Automatic Alignment. *Proceedings of Auto-Carto VII*, 1985.
- Shnaidman A., Shoshani U., Doytsher Y.: Genetic algorithms: a stochastic approach for improving current cadastre accuracies. *Survey Review*, 44(325), S. 102–110, 2012.
- Siriba D.: Conflation of Provisional Cadastral and Topographical Datasets: Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover, ISSN 0174-1454, Nr. 298, Deutsche Geodätische Kommission (DGK), Reihe C 673, 2012.
- Stern C., Sester M.: Towards extraction of constraints for integrating environmental spatial data in digital landscape models of lower resolution – work in progress. *Proceedings of the Joint Workshop Geographic Information on Demand*, Paris, <http://generalisation.icaci.org>, 2011.
- Volz S.: Modellierung und Nutzung von Relationen zwischen Mehrfachrepräsentationen in Geo-Informationssystemen. *Dissertation Universität Stuttgart*, 2006.
- Walter V.: Zuordnung von raumbezogenen Daten – am Beispiel der Datenmodelle ATKIS und GDF. *Universität Stuttgart, Deutsche Geodätische Kommission (DGK), Reihe C 480*, 1997.
- Xiong D.: A three-stage computational approach to network matching. *Transportation research part C*, S. 71–88, 2000.
- Zhang M.: Methods and Implementations of Road-Network Matching. *Dissertation, Technische Universität München*, 2009.

Anschrift der Autorinnen/Autoren

Monika Sester | Tobias Dahinden | Sagi Dalyot | Malte Jan Schulze
 Institut für Kartographie und Geoinformatik
 Leibniz Universität Hannover
 Appelstraße 9a, 30167 Hannover
 tobias.dahinden@ikg.uni-hannover.de
 monika.sester@ikg.uni-hannover.de

Joachim Boljen | Nicole Ruhe
 Landesamt für Vermessung und Geoinformation Schleswig-Holstein
 Mercatorstraße 1, 24106 Kiel
 joachim.boljen@LVermGeo.landsh.de
 nicole.ruhe@LVermGeo.landsh.de