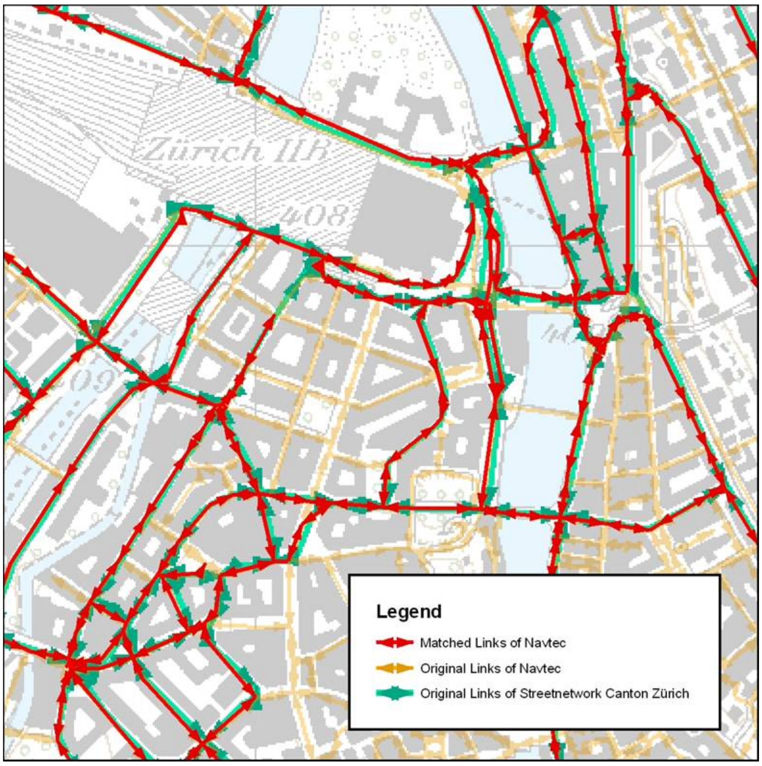


Automatic Matching of large detailed street networks – An algorithm with curve and attribute affinity and its evaluation

Urs Waldner, Institut for Spatial and Landscape Planning

Conference paper STRC 2005

STRC

5th Swiss Transport Research Conference
Monte Verità / Ascona, March 9-11, 2005

Automatic Matching of large detailed street networks – An algorithm with curve and attribute affinity and its evaluation

Urs Waldner
Institut for Spatial and
Landscape Planning
Zürich

Phone: 01 633 30 04
Fax:
email: waldner@nsl.ethz.ch

March 2005

Abstract

To produce and to maintain an area-wide streetnetwork with complete attributes it is useful to integrate information from different networks. The described GIS-methods calculate a matching table of identical routes, so that the attribute values can be compared and switched from one to another network. The algorithm calculates different matching criteria that describe the geometry and attribute affinity of routes from different networks and subsume these criteria to a matching table. To optimize and check the algorithm we digitized manually over 21'000 matching records.

Keywords

street network – GIS – network matching

1. Introduction

The Problem ist well known: In one network dataset the attribute information is good, but the small streets are only available in an other network dataset. A matching table would allow to use informations from two different networks by attribute value transfer.

In this paper a method to quantify the matching probability from link to link based on geometry and attribute affinity is described. The total link-link affinity index can also be used by a method using the link-node-link information.

To produce and maintain an area-wide streetnetwork with complete attributes it is useful to integrate information from different networks. The described GIS-methods calculate a matching table of identical routes, so that the attribute values can be compared and switched from one to another network. The algorithm calculates different matching criteria that describe the geometry and attribute affinity of routes from different networks and subsume these criteria to a matching table. To optimize and check the algorithm we digitized manually over 21'000 matching records. This practical GIS-approach is very robust and handles also complex city street network structures with different generalisation levels, but it can not compensate a big loss of position accuracy by generalizing the geometry.

2. Material

2.1 Steet Network Datasets

The following networks have been used:

2.1.1 Microdrive 99

Area:	Switzerland and Europe
Street classes:	Freeways and some major streets
Node pos. accuracy:	50 - 200 m
Line position accuracy:	10- 500 m
Attribut quality:	High
Source:	Microdrive
Remarks:	Original dataset for IVT-Street-Network with same node coordinates

Fig. 1: Green links (Microdrive 99), Red links (IVT/ARE)

2.1.2 Gis-streetnetwork for canton Zurich

Area: Kanton Zürich
Street classes: Freeways, major streets and important medium streets
Node position accuracy: 1 - 3 m
Line position accuracy: 1 - 5 m
Attribut quality: Very High
Source: Kanton Zürich
Remarks: Some error by network consistancy

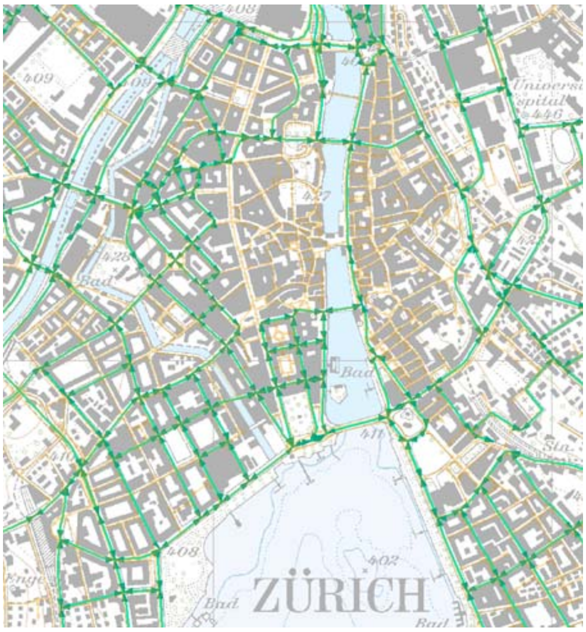


Fig. 2: Green-Blue links

2.1.3 Navtec

Area: Test dataset available for greater zurich area
Street classes: All streets, also without driving rights
Node position accuracy: 5 - 20 m
Line position accuracy: Medium
Attribut quality: Very High
Source: Kanton Zürich
Remarks: Some errors by network consitancy

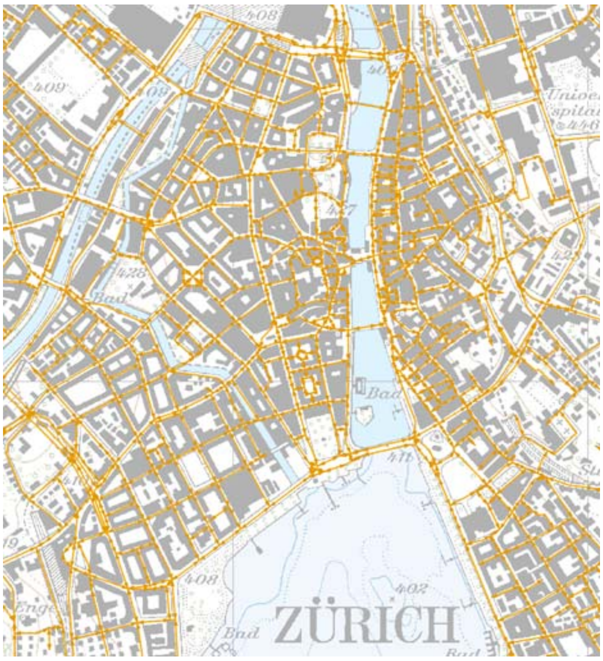


Fig. 3: Navtec network.

3. Method

The algorithum to generate the matching table uses information form geometry and some basic attribute information like street classes if available.

3.1 Software Framework

The program was realised as a prototype to develop the method in two programs:

1. In ArcGIS a Visual Basic for Application (VBA) Sript using ArcObjects writes from all possible matching candidates a new Feature Class. All informations about the geometry proximity are attributes of this Feature Classes in a Personal Geodatabase (Database Format: Access). A configuration database is used to make the program easy adaptable to different networks.

2. In Access a Visual Basic for Application (VBA) script creates and executes the SQL-Queries to rank matching candidates and write out the resulting matching table. For performance reason it was necessary to use SQL instead of pointers like recordsets to make the ranking.

This set up allows an easy development and control of the algorithmus with GIS. The integration of both programs in a installable ArcGIS-Extension is easy to make when a ODBC Connection to the Personal Geodatabase (Access) is used to execute the ranking and selecting SQL-Statements.

3.2 Algorithmus

3.2.1 First Selection with groups of street classes

If street class informations are available in both networks, this informations should be used to avoid coarse errors. Usually the different classification praxis in both networks does not allow to make a simple conversion table, but allows to exclude some streets from the choice set when the discrepancy in street class is to high. The street classification can also be made from indirect informations like the flag "velocity and capacity greater than zero", which allows to exclude streets with ban on driving.

In the program there non-overlapping groups of streets can be defined in the network 1 with a choice set of possible street classes in network 2 (overlapping by network 2 allowed). The group definitions for Network 1 and 2 constitute SQL-WHERE-Statments, which can be developed and checked with ArcGIS and are part of the configuration database (table "tblLineType" (Example table see apendix).

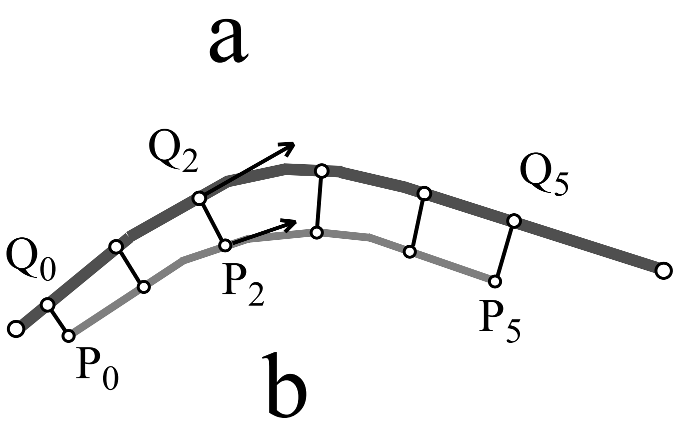
3.2.2 Second Reselection with Buffer

The minimal distance between two lines of a maching candidate pairs must be abouve a fix size.

3.2.3 Geometry gap indicators

For each link *a* in network 1 the described preselection process with groups of street classes and buffers gives a set of possible matching candidates in network 1, called link *b* (See Fig 1).

Fig. 4: Two matching candidates link *a* (network 1), Link *b* (network 2). Link *a* is split up in Points P_0 ... P_5. Q_n is the nearest Point on link *a* from P_n.



a: Line of network 1, can be used on zero, one or both direction

b: Line of network 2, can be used on zero, one or both direction

P_i Points on b at regular distance between Start and End including EndsPoint

Q_i Nearest Point on b to P_i on a

α_i Direction of tangent to a by Point P_i

β_i Direction of tangent to b by Point Q_i

Using these Definitions the following geometry affinity indexes have been calculated:

Table 1: List of geometry affinity index

Indicator	Definition	Significant point on Network	Aggregation formula
StartGap	Distance from start point of *b* to nearest point of *a*	P_0: Start Point of Polyline	StartGap = $\overline{P_0 Q_0}$
EndGap	Distance from End point of *b* to nearest point of *a*	P_n: End Point of Polyline	EndGap = $\overline{P_n Q_n}$
BetweenGap (BETWGA)	Distance from between point on *b* to nearest point on *a*	P_i: Points at regular distance between Start and End including Ends	BetweenGap = $\sum_{i=0}^{n} \overline{P_i Q_i}$
BetweenAngle (BANGLE)	Angle Difference of the tangent on *b* in the betweenpoint and the tangent by Q_i, the nearest point to P_i on *a*	P_i: Points at regular distance between Start and End *including* Ends	BetweenAngle = $\sum_{i=0}^{n} \left(1 - \lvert\cos(\alpha - \beta)\rvert\right)^2$
FractionInside (LENI)	Part of *b* inside a buffer arround *a*	Whole Line *b*	$\frac{len(binside)}{len(btot)}$
Courviness (COUR)	Summarized angle of courves at each vertex of the Polyline of link *b*	All Vertex of the Line *b*	$\sum_{v} \beta_v - \beta_{v-1}$

Where *StartGap* and *EndGap* are simple distances. The two index *BetweenGap* and *BetweenAngle* summarize the distance and direction affinity at many points of the lines. The geometry index *FractionIndex* is the result of cliping the link *b* with a Buffer-Polygon arround link *a*. *FractionIndex* it is easy programmable with standart GIS-Commands.

3.2.4 Third Reselection with value ranges for affinity indexes

For each geometry affinity index, a range of allowed values has been defined. These limits are defined by an interactive manual process considering the percentil and controlling it on the screen. It will not be necessary anymore in the next version.

3.2.5 Transformation of affinity indexes

The different affinity indexes are transformed and aggregated to one single affinity index (*TotalW*). First is to aggregate the some similar indexes

EndsLenW : Combination of StartGap and EndGap

EndsLenW = (Max(StartGap,EndGap))^2

DistanceW = *BetweenGap* * $(Max_{DistanceW} - Min_{DistanceW})$

DirectionW = DirectionM * $(Max_{DirectionW} - Min_{DirectionW})$

DistWindW: BetweenGap weighted with couviness

DistWindW = Log(DistanceW * (1 - Winding1 * ($Max_{Winding1}$- $Min_{Winding1}$)) *($Max_{DistWindW}$-$Min_{DistWindW}$)

TotalW: Summarize over all index

TotalW = LengthW + EndsLenW + DistanceW + DirectionW + DistWindW + DirWindW

3.2.6 Last Reselection

For each link *a* all matching candidates *b* are ranked with the Index *TotalW*. The line length of all selected links *b* should not be longer than the line length of all links *a* plus a fix and length-dependent tolerance.

3.3 Manually generated reference matching table

To check the algorithm a independent manually edited reference matching table was created with a special ArcGIS VBA macro Toolbox. The Toolbox had 3 Buttons with the following functionality:

Write: Rewrite the matching tables records for selected Link *a* and all selected Links *b*. Zoom to all selected links in both networks to make a visual check.

Clear: Clears the Current Selection in both networks for all records with the selected link *a*.

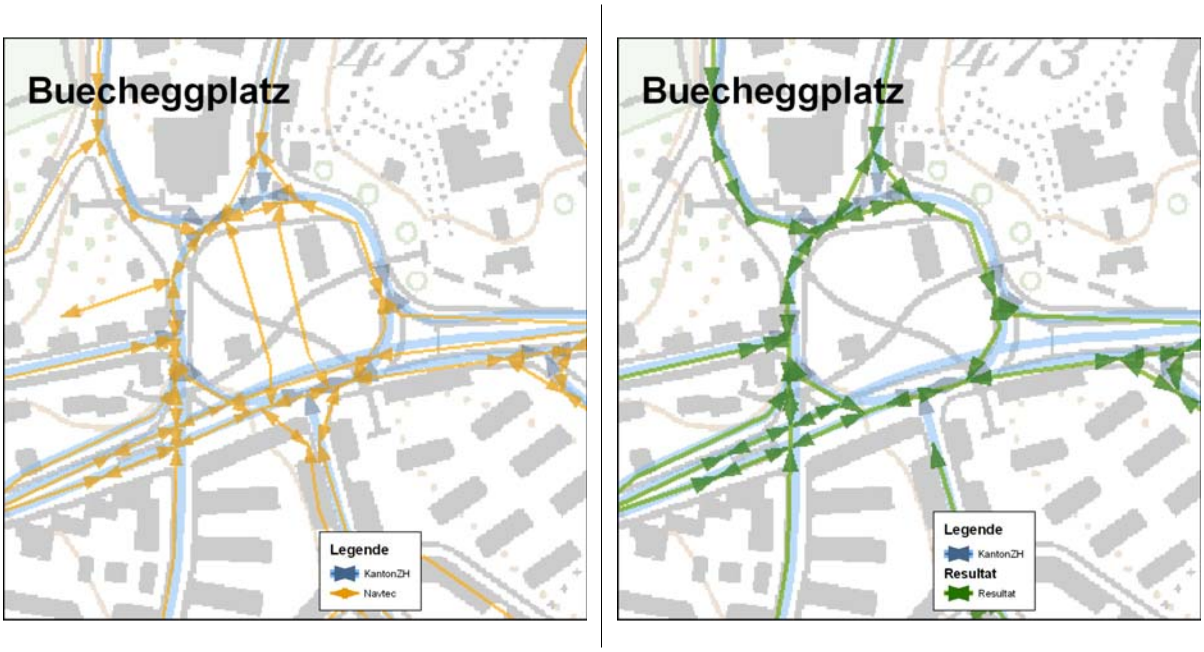
Link: Automatic selection of all corresponding links *b* in network 2 based to the current selection in network 1.

4. Results

4.1 Matching of “Gis-streetnetwork for canton Zurich” with “Navtec”

The matching of the "GIS-Strassennetz des Kantons Zürich" with the detailed Navtec Networks applyed to a complex crossover Bucheggplatz in the city of Zürich with a aterial in a tunnel see Fig5:

Figure 5 a) Original Networks from Buecheggplatz, City of Zurich.
b) Matched Links on Navtec-Network 2.



By difference in the generalization level it creates some missings. The crosstab from all matching candidates of the algorithm against the manually digitized datasets shows that 84 % of the matches are correct (Table 1).

Table 2 Crosstab between with current used algorithm (see Chapter 3) and manually matched reference dataset.

		Algorith Match			
		No Match	Match	Total	% Correct
Manual Match	No Match	22449	3413	25862	0.868
	Match	2933	12015	14948	0.803
	Total	25382	15428	40810	0.844

The Person Correlation (Table 3) between these simple factors shows that the simple FractionInside index has the best correlation to the manual edited reference datas. Startgap and Endgap are quite the same, because the Navtec-Network has two oposite unidirectional links-objects.

Table 3 Pearson Correlation between manual matched and geometry affinity index.

	Manual	StartGap	EndGap	Between Gap	Between Angle	Fraction Inside	Courvi ness
Manual	1	-0.231	-0.231	-0.334	-0.417	-0.726	0.086
StartGap	-0.231	1	-0.074	0.675	0.091	0.249	-0.019
EndGap	-0.231	-0.074	1	0.675	0.092	0.249	-0.019
BetweenGap	-0.334	0.675	0.675	1	0.140	0.369	-0.034
Bet.-Angle	-0.417	0.091	0.092	0.140	1	0.491	-0.066
Fract.-Inside	-0.726	0.249	0.249	0.369	0.491	1	-0.183
Courviness	0.086	-0.019	-0.019	-0.034	-0.066	-0.183	1

A binary choice model shows a little bit better results than the used algorithm (87 % instead of 84 %) see table 3):

Table 4 Crosstab between with current used algorithm (see Chapter 3) and manually matched reference dataset.

		Algorith Match			
		No Match	Match	Total	% Correct
Manual Match	No Match	22426	3436	25862	0.867
	Match	1595	13353	14948	0.893
	Total	24021	16789	40810	0.876

Table 5: Binary choice model based on affinity index.

	B	S.E.	Wald	Sig.
StartGap	-.003	.001	4.786	.029
EndGap	-.003	.001	4.824	.028
BetweenGap	-.003	.003	1.044	.307
BetweenAngle	-.060	.004	183.867	.000
FrectionInside	-4.813	.069	4839.999	.000
Courviness	-.112	.014	61.363	.000
Constant	1.884	.023	6530.778	.000

4.2 Matching of “Microdrive” with “Gis-streetnetwork for Canton Zurich”

Table 2 Crosstab between with current used algorithm (see Chapter 3) and manually matched reference dataset.

		Algorith Match			
		No Match	Match	Total	% Correct
Manual Match	No Match	22426	3436	25862	0.867
	Match	1595	13353	14948	0.893
	Total	24021	16789	40810	0.876

Table 3: Binary choice model based on affinity index.

	B	S.E.	Wald	Sig.
StartGap	-.003	.001	4.786	.029
EndGap	-.003	.001	4.824	.028
BetweenGap	-.003	.003	1.044	.307
BetweenAngle	-.060	.004	183.867	.000
FractionInside	-4.813	.069	4839.999	.000
Courviness	-.112	.014	61.363	.000
Constant	1.884	.023	6530.778	.000

5. Conclusion

This practical GIS-approach is very robust and handles also complex city street network structures with different generalisation levels. The match rates are enough for use cases like to generate localized emmission datas. To have match rates over 90 % the planned extension with a short path algorithm based on a affinity cost index must been implemented (see Chapter 6).

6. Outlook: Further Extension

The described reselection process can be extended to a more robust algorithm which uses also the node-link information.

For each link *a* a subset of the network 2 can be build using the first and second reselection steps (Chapter 3.2.1 and 3.2.2). For both ends (E_s, E_e) of the link *a* (network 1) three nearest points N_i and N_j on different links *b* of the network 2 were searched. The shortest path on a subset of the network 2 will be searched between all Points N_i and N_j using the affinity index *TotalW* between link *a* and *b* as cost-function. These nine solutions from N_i to N_j can be ranked based on the distance E_sN_i and E_eN_j and the summarized affinity index over all used pairs of link *a* and *b*. It is also possible to add a from/to field to the matching table which describes the part of the link *b* that matches to link *a*.

Apendix: Groups of street classes

Table 2: Definition Querys for Matching Example KanNav: "Gis-streetnetwork for canton Zurich" and "Navtec". Field "*RICHTPLANTYP*" and "*NT_KAT*" are street classification fields. Field "*KAP*" is the capacity of this link.

SQL Where Clause for network 1 (Gis-streetnetwork for canton Zurich)	SQL Where Clause for network 2 (Navtec)
RICHTPLANTYP = 10	NT_KAT < 4 AND KAP <> 0
RICHTPLANTYP = 11	NT_KAT < 5 AND KAP <> 0
RICHTPLANTYP = 12	NT_KAT < 5 AND KAP <> 0
RICHTPLANTYP = 20	(NT_KAT < 5 OR NT_TYP = 9) AND KAP <> 0
RICHTPLANTYP = 21	(NT_KAT < 5 OR NT_TYP = 9) AND KAP <> 0
RICHTPLANTYP = 22	(NT_KAT < 5 OR (NT_TYP = 6 AND NT_KAT = 5) OR (NT_TYP = 6 AND NT_KAT = 7)) AND KAP <> 0
RICHTPLANTYP = 30	((NT_KAT = 4) OR (NT_TYP = 9 AND NT_KAT = 5) OR (NT_TYP = 11 AND NT_KAT = 5)) AND KAP <> 0
RICHTPLANTYP = 31	((NT_TYP = 6 AND NT_KAT = 5) OR (NT_TYP = 9 AND NT_KAT = 5)) AND KAP <> 0
RICHTPLANTYP = 40	((NT_TYP = 9 AND NT_KAT = 5) OR (NT_TYP = 10 AND NT_KAT = 4) OR (NT_TYP = 11 AND NT_KAT = 5) OR (NT_TYP = 12 AND NT_KAT = 7) OR (NT_TYP = 15 AND NT_KAT = 7)) AND KAP <> 0