Web Processing - Standardised GIS Analyses for Cable Route Planning

SEBASTIAN HEIDEN, Harz University of Applied Sciences, Germany

add as final part

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: datasets, neural networks, gaze detection, text tagging

ACM Reference Format:
Sebastian Heiden. 2022. Web Processing - Standardised GIS Analyses for Cable Route Planning. *J. ACM* 37, 4, Article 28 (February 2022), 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Sometimes, finding the shortest path is not sufficient. Additional parameters play also have to be taken into consideration. As the steepness of a road or the soil, play an important role for the building cost of a road or pipeline . For planing the additional routes for a power grid, additional aspects as legal regulations and acceptance by the local population have to be taken in consideration. Also technical aspects, as the effects on the grid stability might be further points to consider.[8]

source

what is the least cost path

The aggregation is implemented in a priority queue. Hence a point with a lower cost is evaluated earlier as a starting position to evalutate the cost of its neighbors, than a point with higher costs. The aggregation is implemented with early stopping. After the last end point has been has got an aggregated cost the aggregation stops and the back tracking starts.

2 METHODS

We retrieve a set of different spacial data-sets from public sources as a basis to create the cost raster. Field of study are the counties of Cuxhaven and Osterholz in the state of Lower Saxony, Germany. Areas protected by different European and National conservation laws are provided by the German Environment Agency as Web Feature Service (wfs) [1].

The nationwide land coverage (ATKIS) with a scale of 1:250000 are provided by the Federal Agency for Cartography and Geodesy [3]. The nationwide power grid (tags: 'power': line) has been retrieved via OpenStreetMap [7]. Local data as houses at Level of Detail 1 are offered by the State Office for Geoinformation and Land Surveying of Lower Saxony[6]. In addition, local planning geodata for the land usage are taken from 'Metropolplaner' (Planing data Lower Saxony & Bremen)[5]

Author's address: Sebastian Heiden, u38439@hs-harz.de, Harz University of Applied Sciences, Friedrichstrasse 57-59, Wernigerode, Saxony-Anhalt, Germany, 38855.

0004-5411/2022/2-ART28 $15.00
https://doi.org/XXXXXXX.XXXXXXX

PyWPS[2] is used to offer the least cost path algorithm as a Web Processing Service (wps). The initial implementation of the least cost path algorithm is based on the implementation for the QGIS-Plugin 'Least Cost Path'[4] in version 1.0.

The different layers from the different entities are optionally filtered, buffered and then rasterized. Filtering the layers of the files for special attributes enables to further differentiate. For examples makes it possible to differentiate between heath and uncultivated land. Adding a buffer can be used either used to convert a line objects as power grids and streets into a polygon with the correct physical width, or to add minimum distance from an existing of planed area to the new power grid. Each of theses rasterized layers are given a weight, or cost that expresses the cost of using land covered by this layer. The costs of all layers of is aggregated with the maximum function. Thus, an area covered by multiple layers is uniformly used with the highest costs. For every area in the study area, that is not covered by any layer, is given the default cost. The costs have been grouped into different levels (see table 1) starting from preferential areas with very low costs, via no restrictions, which is the default, used when no other layers are covering the local area, to restricted, strongly restricted and prohibited areas with high costs. These higher costs resemble the degree how much a local area should be avoided, while routing the path. The ratio of the higher costs to the lower costs directly translates into the additional diversion in pixels the algorithm is willing to go, for avoiding an area of high costs. Thus, as prohibited areas describe a legal obligation, not to use these areas or only to the utmost minimum, the weight that resembles the costs for these types of areas, has to be especially high.
The rasterization transforms a vector into a raster. The rasterization can be executed in to different ways. In both ways, the rasterization can be imaged, as the old vector is layered over the new raster grid with the new given resolution and the new affine transformation and the coordinate reference system of the vector. Both rasterization techniques differ in how the pixels are selected, that describe, that a geo-object overlays the pixel. The pixels can either be described as overlaid with a geo-object, when either the center of that pixel is overlaid by the geo-object, or any part of the pixel is overlaid. Hence, the version with any part overlaid is called all touched True. The version where a overlaid pixel center is require is thus called all touched False. All touched False is considered the default (see figure 1).

The completed list of layers and the processing applied to them, can be found in Supplement S1.

create the Supplement from the processing rules

The whole stack of the cost raster, via the aggregated costs to the resulting least cost path can is shown with an example for a starting cost raster of 50 m resolution and a rasterization with all touched False (see 2). Because of the early stopping strategy used in the aggregation sub-paths with high costs have not been fully computed.Instead the execution stops early after reaching the end point. Because the end point ends at a power transformer, which is a building type, the paths ends at in a Prohibited Area. Therefore

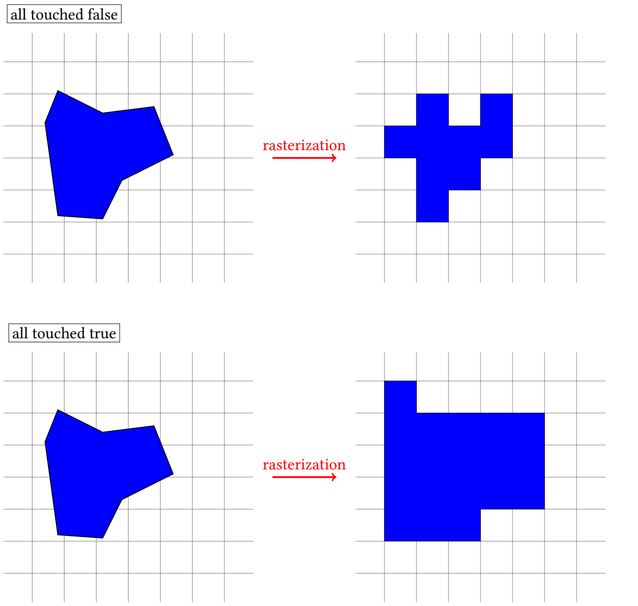
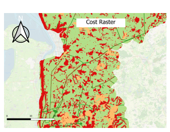


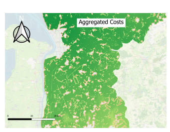
Fig. 1. Graphical example for rastering a vector (left blue), to a raster (right blue) with weither all-touched False (above), or all touched True (below).

Table 1. Used levels of costs, the applied numerical equivalent and example layer this cost have been used for.

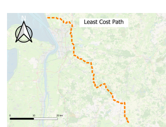
Cost Level	Cost	Example
Prohibited	500	Conversation areas as National Parks, Buildings
strongly Restricted	10	Conversation areas as Bird Reserve
Restricted	5	Protected Landscape Area, Industrial Areas, motorway, railway
No Restriction	0.5	Default
Preferential	0.1	Power Grid, Motorway and Railway Buffers



(a) Cost Raster.



(b) Aggregated Cost Raster.



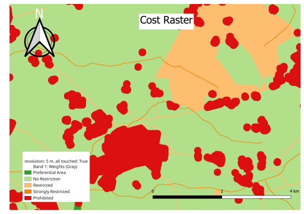
(c) Least Cost Path.

Fig. 2. Figures of Cost raster and the resulting aggregated Costs and the least cost path for a resolution of 50 m.

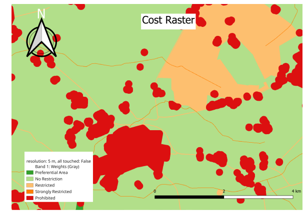
areas even farther away from the starting point have been explored first.

3 RESULTS

In this chapter we want to show the different cost raster, that were created from the very same set of layers, but computed for different resolutions. From this different raster the cost paths are calculated and compared. In the last step the raster with lower resolution were used to calculate in a way, that they shall result in similar paths as if the paths were computed from a high resolution raster.



(a) All touched: True.



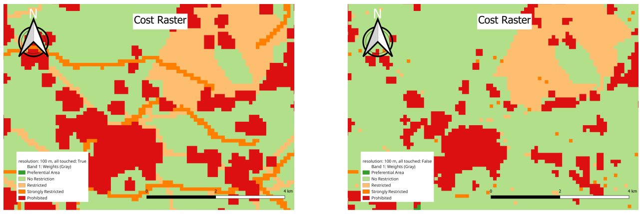
(b) All touched: False.

Fig. 3. Figures of the cost raster. Contrasting the for different values in all touched at a resolution of 5 m.

3.1 Cost Raster

The cost raster contains all the costs as weights for the geographic region of the study area. The regions outside the study area are given, a no-data value and will not be use for the calculation of the least cost path. It is decided to use the maximum value, for any place in the raster, that is covered by several rules at the same time. When the resolution that is use for the rasterization is smaller than the object size, than the effect of the rasterization with all touched True or False is limited. For all touched True any part of the pixel that is covered by the object, will attribute that whole pixel to the object. Thus, the object appears to be halve a pixel size larger in all directs. As can be seen in figure 3 that shows a detailed view of the costs for village of Beverstedt. Due to the maximum aggregation of the costs, the average cost of the raster of all touched true will be over estimated. All touched false will be a better description of the real size of the object.

In contrast when the resolution the is much less, than the size of the object the described behavior changes. On one hand, while the area, of the pixel is increased for all touched True, also the area for which the cost is overestimated increases. On the other hand while the pixel size for all touched False increases, not only the over estimated area increases for that pixel as describes for all touched true, but in addition the object is less probable situated in the center of the pixel. The consequence of the pixel is, that will decreasing resolution that object is not included in the rasterziation. Hence, hence for all touched false lower resolution leads ot a loss of information. Because the default is relative small compared the over effect is a underestimation of the costs. The figure 4 shows for the resolution of 100 m, that while larger are still included in the map they appear to be much larger. On the contrary smaller objects might not be included. Objects that are small on in one dimension as streets will be included in a stochastic manner. Described by the likely of that a object of that sized overlaps with the pixel center. Thus, with having larger areas and more objects, with higher costs, all touched True rasterization might more likely lead to longer roots and more likely in blocking the direct spatial path.



(a) All touched: True. (b) All touched: False.

Fig. 4. Figures of the cost raster. Contrasting the for different values in all touched at a resolution of 100 m.

3.2 Least Cost Paths

For each resolution the costs paths were estimated for the rasterization with setting all_touched False and all_touched True.

The hypothetical path starts at a transformer about 6 km north the container terminal Bremerhaven and then end at a transformer in the south east of the Osterholz county.

The distance of both paths is calculated with the mean minimum distance. For every vertex P_i in the path L_1 the minimum between the vertex and the path L_2 is computed and afterwards the minimum distances are averaged (see equation 1).

$$d_{mean} = \frac{1}{|L_1|}\sum_{i=1}^{n} d_{min}(p_i, L_2)\Big| p_i \in L_1 \tag{1}$$

$$d_{max} = max(\sum_{i=1}^{n} d_{min}(p_i, L_2))\Big| p_i \in L_1 \tag{2}$$

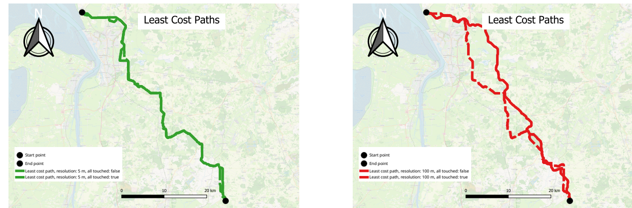
Hence, the mean minimum euclidean distance between the two paths can used to compute the similarity. As different table 2 shows the distance between the two paths decreases with increasing resolution. In addition, this tendency is depicted in figure 5 for the calculated cost paths of 5 m and 100 m resolutions.

In contrast the largest distance between the paths (see equation 2) will be used, to estimate the minimum distance to the areas that still should be examined to grantee a, that at least cost path found in this limited space is still likely to be optimal.

At the same time the differences in the aggregated costs stay constant. When normalize the aggregated costs by the resolution. Than On one hand it can be seen, that the all_touched False under estimate the costs and that this tendency scales linear with the resolution. On the other hand, the all_touched True least cost over estimates the aggregated costs on a linear scale of the resolution. Therefore, the difference for the normalized least cost paths is reduced on scale by the resolution.

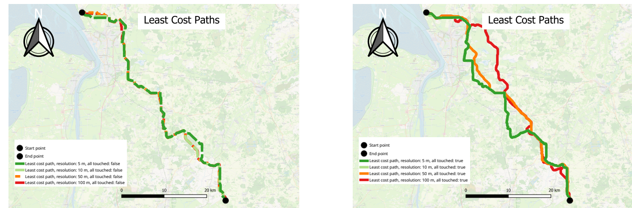
Comparing the all_touched True rasterization and all_touched False for the same resolution in contrast with the paths of all_touched False rasterization at the different resolutions the later paths are more similar. The mean average distance between the 100 m resolution run and the 5 m resolution run is 257.97 m.

The similarity of all all_touched False runs is higher than, the similar between the all_touched False and all_touched True runs with the same resolution, except for the highest resolution.



(a) resolution of 5 m. (b) resolution of 100 m.

Fig. 5. Figures of the least cost paths. Contrasting the paths for different resolutions. Paths computed from all touched false raster are indicated by dashed lines. Results from all touched True are indicated by continuous lines. Higher resolutions are indicated by the color green, lower resolutions by the color red. Using OpenStreetMaps as base map.



(a) all touched false. (b) all touched true.

Fig. 6. Figures of the least cost paths. Contrasting the changes of the least cost paths for the different results, depending on the parameter all_touched. Paths computed from all touched false raster are indicated by dashed lines. Results from all touched True are indicated by continuous lines. Higher resolutions are indicated by the color green, lower resolutions by the color red. Using OpenStreetMaps as base map.

Hence, in an overall perspective paths of the all_touched False runs stay in a similar region, while paths of the all_touched True coverage strongly to the paths all_touched False runs. This behaviour is depicted in figure 6. On a more detailed level, it can be seen, that also the Paths of all_touched converges the all_touched True path. But the extent is smaller. The length of the the paths only differs to a maximum of about 10 %. The least path distance for higher resolutions can be lower, because new paths, between regions that are forbidden in higher resolution can open. On the other side the length of the paths can increase, because with higher resolution more vertices will be used.

The zonal stat (see table 3) for a buffer of 100 m (5 m) around the path has been used, to estimate the percentage of which costs levels are around the path. When using all_touched True rasterization with higher resolution the tendency is to use a higher percentage of the Preferential Level and less the NoRestriction Level. The ratio of the 100 m buffered least cost path, strongly shifted to Levels lower costs. There is no strong tendency for the all_touched False least cost paths.

Table 2. Least cost paths as length for the different resolution (res) of the raster, including the mean minimum distance and the maximum minimum distance and the aggregated (agg.) costs. From the agg. costs the differences of the agg. costs and the corrected (corr) agg. by resolution are given.

res /m	$length_{al=f}/m$	$length_{al=t}/m$	d_{mean} /m	d_{max}/m	agg. $cost_{al=f}$	agg. $cost_{al=t}$	Δ costs	corr agg. $costs_{al=f}$	corr agg. $costs_{al=t}$
5	76136.27	78002.00	126.04	1065.00	18665.923	19616.756	-850.00	93329.60	97584.77
10	75430.10	77936.57	277.92	1590.00	8931.245	9731.175	-799.95	89312.45	97311.75
25	75422.85	78422.85	313.75	1621.15	3354.869	3872.656	-517.78	83871.73	96816.40
50	76135.02	70619,95	1140.01	4950.00	1409.023	2300.073	-891.05	70451.15	115003.65
100	76283.80	74120.73	1946.41	6016.64	640.516	1572.268	-931.70	64051.60	167226.80

3.3 Execution time

In theory, the execution time should increase with the square of the resolution, because higher resolutions result in a higher number of pixels and thus data points the aggregated costs needs to be calculated for. A full logarithmic fit for several repetitions of the execution shows, that the execution time scales with power of 2.1997 ± 0.007 of the inverse resolution. With the caveat of a low number of samples this can equally be success fully be fitted to a second degree polynomial of the inverse resolution with a r^2 for the test set of 0.99 and a the squared inverse of the resolution with a r^2 for the test set of 0.99. Hence, the order of magnitude

The total execution time consists of two parts. The aggregation of the costs and the back tracking of the least cost to find the path.

back tacking scales with number of points.

used regions

3.4 Faster Processing of the Cost Path Algorithm

The final least costs paths should between the least cost paths of the lower resolution, with a tendency to be nearer to the paths resulted by the all touched false rasterization. The first step is to optimize the calculation speed is, only to calculated the least cost paths for this smaller area. Another method, is to improve the prediction of the medium resolution itself and thus reduce the need for a computation in higher resolution.

3.4.1 Compare least cost paths, for overlay of both rasterizations. As all touched true rasterization overestimates and all touched true underestimates the real costs. A weighted mean will describe the real costs more precise. As present work indicates, that the weight should favor the all touched false raster. The best weight should be the percentage of the pixel, which is covered by the object, but this can not be computed in this work. Thus, the best weight has to be searched. An alternative might be to compute the cost raster in higher resolution, but than to reproject the to a smaller resolution with a using a (linear) interpolation of the of the weights.

While the aggregated thus can be speed up. The time needed for the back tracking stays unchanged.

3.4.2 Compare least cost paths, for down sampled cost paths. As an alternative to the super position of the all touched true and false raster for the same resolution, the all touched false raster was down sampled to 10 m, 25 m, 50 m and 100 m with bi-linear interpolation. With this method smaller structures still can be seen in the cost raster, although the resolution is reduced.

3.4.3 Restrict search to a buffered around the least cost paths. Construct a polygon form the two lines. Buffer the polygon with twice the max minimum path distance.

This enabled the possibility to run a 2.5 m resolution cost raster and clip it to the extent of the polygon. This clipped 2.5 m raster for all touched true did only slightly change the path. While the all touched False raster, leads to a totally new segment at the end of the path. Due to the small resolution, that a small path with the extent of about 2.5 m width became passable. This small path is a power line next to road between which build a cone in a protected landscape area. The street and the protected landscape area are both restricted areas, while the power line is preferential. This showing, that the way the cost raster is created in the first place can play a vitual role, in the end result. So that a nuance, can cause a detour. When this behavior occurs, the polygon can not grantee to include the least cost path. Therefore this polygon should be overlayed with a polygon around the shortest path.

3.4.4 Restrict search to reachable points. For this purpose the aggregated cost as to converted back into a raster.

3.4.5 Compare the different Solutions.

3.4.6 Check solution for least cost path between different points.

4 DISCUSSION

As shown need of computation power increases with the power of two with the resolution. While the error only scales linear with the resolution.

As the original least cost path QGIS plug-in the used algorithm stops the least aggregation of the costs after the final end point has been found. This early stopping might result in sub optimal paths around the end points.

As a any buffer around buildings are describes as protected areas, power poles are also included. In further works, these should be excluded.

Difference aggregation of the rule of a cost raster with max vs sum

Scaling with multiple endpoints. Not algorithmic speed-ups as precompilation, JIT, vector units, multi-core

5 RELATED WORKS

6 CONCLUSION

ACKNOWLEDGMENTS

. . .

REFERENCES

[1] 2015. Schutzgebiete in Deutschland. https://geodienste.bfn.de/schutzgebiete?lang=de

Table 3. ratio (r) of Category percentages of each least cost path for a buff of 100 m (5 m) around the least cost path.

resolution /m	all touched	$r_{Preferential}\%$		$r_{NoRestriction}\%$		$r_{Restricted}\%$		$r_{stronglyRestricted}\%$		$r_{Prohibited}\%$	
5	False	4.7	(5.4)	58.7	(58.9)	8.8	(8.4)	0.7	(0.7)	27.1	(26.7)
10	False	19.6	(33.5)	68.5	(64.5)	1.0	(0.8)	0.8	(0.3)	10.1	(0.9)
25	False	19.2	(34.2)	68.9	(64.9)	1.0	(0.2)	0.7	(0.1)	9.7	(0.6)
50	False	20.4	(33.2)	68.0	(66.2)	0.9	(0.1)	0.7	(0.0)	10.1	(0.5)
100	False	21.1	(30.7)	69.1	(68.8)	1.1	(0.0)	0.7	(0.0)	7.9	(0.4)
5	True	18.9	(28.5)	67.3	(66.4)	1.3	(1.6)	1.0	(0.5)	11.5	(3.0)
10	True	18.9	(33.7)	66.6	(63.4)	1.6	(1.4)	1.4	(0.6)	11.5	(1.0)
25	True	18.7	(31.9)	65.5	(65.5)	2.0	(1.3)	2.5	(0.7)	11.4	(0.6)
50	True	9.1	(13.0)	75.7	(83.0)	3.9	(2.0)	4.2	(1.6)	7.1	(0.4)
100	True	7.0	(10.1)	73.8	(81.9)	5.5	(3.9)	8.5	(3.6)	5.2	(0.4)

[2] 2016. Welcome to the PyWPS 4.3.dev0 documentation! — PyWPS 4.3.dev0 documentation. https://pywps.readthedocs.io/en/latest/index.html
[3] 2021. Digitales Landschaftsmodell 1:250 000 (Ebenen). https://gdz.bkg.bund.de/index.php/default/open-data/digitales-landschaftsmodell-1-250-000-ebenen-dlm250-ebenen.html
[4] 2022. LeastCostPath/dijkstra_algorithm.py at master · Gooong/LeastCostPath. https://github.com/Gooong/LeastCostPath
[5] 2022. Metropolplaner. https://metropolplaner.de/metropolplaner/
[6] 2022. OpenGeoData.NI. https://opengeodata.lgln.niedersachsen.de/#lod1
[7] Geoff Boeing. 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65 (Sept. 2017), 126–139. https://doi.org/10.1016/j.compenvurbsys.2017.05.004
[8] Benjamin Schäfer, Thiemo Pesch, Debsankha Manik, Julian Gollenstede, Guosong Lin, Hans-Peter Beck, Dirk Witthaut, and Marc Timme. 2022. Understanding Braess' Paradox in power grids. *Nature Communications* 13, 1 (Sept. 2022), 5396. https://doi.org/10.1038/s41467-022-32917-6 Number: 1 Publisher: Nature Publishing Group.

A RESEARCH METHODS

A.1 Part One

A.2 Part Two

B ONLINE RESOURCES

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009