

SHAPE ASSESSMENT OF GENERALIZED BUILDING FOOTPRINTS

Yevgeniya Filippovska^a, Martin Kada^b

^aInstitute for Photogrammetry (ifp), University of Stuttgart
Geschwister-Scholl-Str. 24D, 70174 Stuttgart, Germany
yevgeniya.filippovska@ifp.uni-stuttgart.de

^bInstitute for Geoinformatics and Remote Sensing (IGF), University of Osnabrück
Barbarastr. 22b, 49076 Osnabrück, Germany
martin.kada@uni-osnabrueck.de

KEY WORDS: Comparison, Building, Metric, Generalization, Quality

ABSTRACT:

Along with the constantly improving acquisition, processing and visualization techniques, the amount of spatial data that is available today has dramatically increased. As a consequence, various data sets for one and the same area are often available, reflecting temporal changes or representations at different resolutions. Thus, there is a rising need to compare such data sets in order to assess their quality and pick out the proper one for a specific application. In this paper, the focus lies on the shape assessment of building ground plans for which various versions are available as a result of cartographic generalization. The commonly used characteristic for measuring the deviation between the outlines of polygonal objects is the Hausdorff distance. On generalized building data, however, the metric performs less intuitive from a human point of understanding. Examples resulting from the geometric simplification of footprints are given, the shortcomings of the Hausdorff distance discussed, and an alteration of the metric motivated therewith. This new metric does not allow for a direct connection of the two points of shortest distance, but restricts it to remain inside the polygonal difference between the two footprints. Besides the shortest distance, a variation of the metric is also shown to express the complexity of shape differences.

1. INTRODUCTION

In urban areas, buildings make up an essential portion of GIS-data. And even though 3D models are already captured for many cities, the predominant building type for comprehensive data sets is still 2D footprints. For various reasons, multiple versions of one and the same building might be available that differ both in geometry and shape. The reasons are various like data collections at different times, transformations by means of analyses (e.g. 3D building reconstruction) or for visualization purposes (e.g. generalization). While small geometric differences are naturally inevitable and insignificant for most applications, it is, however, important to be able to automatically identify with certainty large changes in location, extent and particularly in the shape of the objects; the last being the topic of the paper with a focus on data quality regarding generalized building footprints.

In order to compare the contours of two arbitrarily shaped objects, the prominent characteristic widely used both in the scientific area and in commercial products is the Hausdorff distance. This metric measures the distance between two non-empty, closed and bounded point sets in metric space (Herrlich et al., 2002). The definition, which states that each spatial object can also be considered an infinite compact point set, is standardized by ISO (Andrae, 2009; ISO 19107, 2003). The metric has been successfully used in such disciplines as image processing and computer vision. Since then, its application field is not only restricted to the measurement of distances between the locations of two objects, but it is used for the assessment of geometric similarity of two contours as well. Thus, in the approach from (Schlüter, 2001), the

Hausdorff distance is applied along with other characteristics for similarity-based classification of the contours of polygonal objects. In the work by Volz (2006), the Hausdorff distance is used among other parameters for the detection of correspondences between multiple representations of objects, a so called feature matching, in a road network. Due to its significance, the highest weight was assigned to this characteristic. In (Kreveld, 2006), the Hausdorff distance is referred to as a characteristic for the similarity measure of boundaries usable on any two objects for the automatic data correction after manual digitizing. It helps to detect, e.g., if the same data was captured several times like common boundary segments of neighboring objects. In this way, the topological consistency of a data set can be guaranteed or at least considerably improved.

In (Bouziani et al., 2010), the Hausdorff distance is particularly applied to compare the geometry of buildings extracted from high resolution satellite images to a reference map. This work is dedicated to the automatic change detection with regard to the shape of objects in urban areas such as streets, parking places, vegetation, water objects and building footprints. The Hausdorff distance is part of the quantitative evaluation of the results that are provided by this approach and used along with the analysis of identification accuracy (rate of correct change detection) and area ratio analysis (percentage of detected areal change).

In three-dimensional space, the Hausdorff distance enables to measure the maximum distance between two surfaces. A frequent example given is the comparison of a triangulated surface to its simplified version (Luebke et al., 2003). Since the accurate com-

putation of the Hausdorff distance in 3D is very complex, the approximate deviation will be evaluated instead, though the methods of its computation can vary. In (Cignoni et al., 1998) one of the two surfaces, the reference surface, will be approximated by a regular distributed grid with given resolution, the value of which depends on the required accuracy of the result. After that, the shortest distances will be computed from each vertex of this grid to the second surface.

In another approach, the pattern parallel to any of the two triangle sides, which the surface of the object is compound of, is used in (Aspert et al., 2002) to approximate each of them separately. At last, only the parts of the triangulation surface with the expected maximum distance to the other surface will be approximated in (Guthe et al., 2005).

As in the case of similarity measure, the objects are compared to their references for the purpose of quality evaluation as well. The primal difference between these two tasks consists in their goal. Thus, similarity measure basically aims to identify if an object fits a certain search pattern or if the compared objects are alike. From the decision that is made depends, e.g., whether the object will be included into the search result. Or it could be determined that the objects are different representations of the same phenomenon and will be considered appropriately with all the consequences that follow. For this reason, it is common in the field of similarity measure to operate with thresholds. Here, the attention is principally paid only to the fact whether the distance value exceeds a given limit or remains within it. Thereby, the specification of the Java Conflation Suit explains that the explicit calculation of the Hausdorff distance for similarity measure of objects and the subsequent detection of their correspondence is not necessarily within the scope of this approach, although this concept exactly underlies it (Blasby et al., 2003). Instead, it is just sufficient to know that the maximum distance between the objects and their contours does not exceed the so called distance tolerance, which can be modeled by the generation of a buffer. In this case, polygonal objects will be mutually tested whether they completely stay within the buffers of each other; buffers around the objects themselves and around their boundaries are considered separately.

In contrast to the similarity measure, the value of the quality measurement itself is of major interest for the process of quality evaluation, because the sensible classification of results can only be application-specific. The possibility of direct usage of the Hausdorff distance for data quality assessment is proclaimed in (Hangouët, 2006). It is shown that it can be applied to characterize the geometric aspect of quality of such object types as points, lines and particularly closed lines. Moreover, this work suggests to compare not only individual objects, but also to indicate the correspondence between whole data sets.

However, the Hausdorff distance turns out to perform less effective for building footprints, which is discussed in section 2. It is shown that the distance computed in such a way is not always comprehensible by a human or can help one to detect the relevant shape differences of two building footprints. In section 3, an alteration of the Hausdorff metric is proposed, which restricts the computation of the maximum distance between the contours to be

within the polygons of the symmetric difference. As the distance measured by a straight line does not always run completely inside these polygons, the second distance metric measures the length of a polyline that lies within the aforementioned area (section 4). The results are discussed in the last section.

2. DIFFICULTIES IN MAXIMUM DISTANCE IDENTIFICATION BETWEEN TWO GROUND PLANS

Since the exact mathematical definition of the Hausdorff distance is difficult to implement for continuous objects, it is common to calculate an approximation of this characteristic, called the vertex Hausdorff distance. It considers only the distances from the vertices of one object to the edges of the other one and contrariwise, where the notion of this characteristic originates from (Schlüter, 2001) and (Volz, 2006). In fact, the maximum distance between two objects often originates at a vertex lying on either of the outlines (if their line segments run in approximately similar direction). In this case, the vertex Hausdorff distance completely coincides with the original characteristic, which makes the accurate computation redundant.

Nevertheless, this approach fails if the compared objects are shaped quite differently (Hangouët, 1995). Figure 1 illustrates a typical case for such a situation. The shape of the initial ground plan presents an open ring with its right part eliminated by generalization. As a consequence, a U-shaped object is generated. It is significant for such a shape change that the distances from the deleted element, considering the entire points of its outline to the second contour, are the shortest at the vertices of this element. The longest distance between the two contours in this position is shown in Figure 1a as a thin dashed line. However, the distance increases moving from the end point to the middle part of the vertical segment of the deleted element. The actual maximum distance from the generalized contour point is to be identified under consideration of all points delimiting the deleted element and lies on its longest side. According to this, there are two distances from this point to the generalized contour available with exactly the same length (cp. Figure 1b).

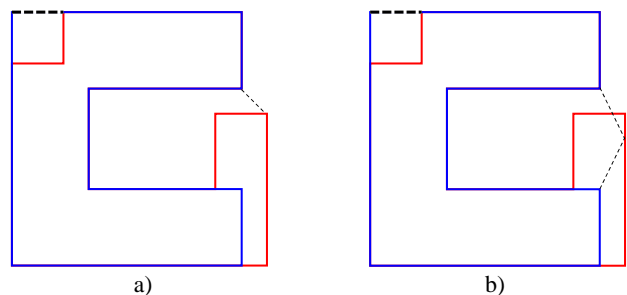


Figure 1. Result of the approximated vertex Hausdorff distance in comparison to the real Hausdorff distance.

The calculation of the maximum distance is basically meant to identify the area of highest discrepancy between two contours, which must not exceed a tolerance threshold. As it is shown in

Figure 1 by the bold dashed line, the maximum distance is not just quantitatively underestimated, but even falsely located.

A further question that is just as important for the development of a new characteristic is to which extent the solutions are actually intuitive to understand. Basically, the correspondence between the parts of the two contours is based on the shortest distances between the points of their outline, which can result in an incomprehensible result for the human viewer.

As Figure 2a presents, the maximum distance between the two objects runs from the upper right corner of the eliminated element to the closest edge of the generalized ground plan and is shown with a bold dashed line. However, a human viewer would perceive the maximum distance in all likelihood as the thin dashed line. It can be explained by the fact that the eliminated element rather corresponds to the low part of the generalized ground plan to which it was connected to. A similar example can be seen in Figure 2b. In this case, the maximum distance runs between the upper vertex of the eliminated element and the left edge of the generalized ground plan, although this element belongs rather to the bottom side of the original contour.

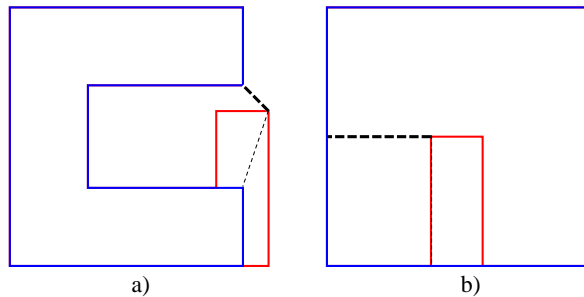


Figure 2. Computed maximum distances between two contours compared to what a human might perceive.

Figure 2 clearly demonstrates that the Hausdorff distance as a characteristic for computing the maximum distance between two building ground plans does not always deliver adequate results. Thus, neither the accurate mathematical solution nor its approximation may correspond to human perception. In order to overcome this disadvantage, it is suggested to evaluate the maximum distance between the two contours not by means of the shortest distances between their points, but considering each eliminated part as a whole.

Summarizing all above, it is to point out three properties, which arise with the identification of the maximum distance between the two contours and cannot be solved by using of the Hausdorff distance: the new characteristic must be able to detect erroneous situations produced by the generalization, the results should be intuitive to a human viewer, and the computation must be simple.

3. MODIFIED HAUSDORFF DISTANCE

3.1 Maximum distance within the symmetric difference

In order to avoid a false identification of the correspondence between the elements of the original and generalized polygon, the computation of the Hausdorff distance can be based not on their entire set of vertices, but evaluated within each individual eliminated element i instead. For this purpose, it must be first determined, what segments belong either to the original or to the generalized ground plan, which are accordingly identified as sets of points (O) and (G). After that, the Hausdorff distance (HD_i) is determined between these two original and generalized polylines forming the eliminated element. The overall maximum distance between the ground plans (HD_{Mod}) is the maximum value of all Hausdorff distances within the eliminated elements computed in the same way. Altogether they compose a set of distances (HD).

To be exact, the notion of the eliminated element implies the deleted as well as the newly generated parts of the original building ground plan after generalization. Conceptually, this approach is based on the assumption that only the intersection points of the two ground plans express the direct correspondence between these objects. From this, it follows that the discrepancy of the two contours will be measured between the sections running between the same intersection points of the ground plans. The course of action of this approach consists of the following steps:

1. By means of intersecting the original and generalized ground plan, the eliminated parts, so called intrusions and extrusions, are detected, as Figure 3a and b show. After the notion of set theory, they can be identified altogether as symmetric difference (SD).

$$SD = (O \setminus G) \cup (G \setminus O) \quad (3.1)$$

2. In order to identify the original and generalized components, the contour of each individual eliminated element will be successively intersected with the contour of the original and generalized ground plan. These polylines are accordingly shown with red and blue in Figure 3c and d.

$$\partial O_i = \partial SD_i \cap O \quad (3.2)$$

$$\partial G_i = \partial SD_i \cap G \quad (3.3)$$

3. The computation of the one-sided Hausdorff distance between the original and generalized components $hd_i(O)$ and conversely $hd_i(G)$ is based on the vertex Hausdorff distance, which means the shortest distances will be identified from the vertices of the initial polyline to the edges of the second polyline.

$$hd_i(O) = d(\partial O_i, \partial G_i) = d((\partial SD_i \cap O), (\partial SD_i \cap G)) \quad (3.4)$$

$$hd_i(G) = d(\partial G_i, \partial O_i) = d((\partial SD_i \cap G), (\partial SD_i \cap O)) \quad (3.5)$$

- The maximum distance within an eliminated element of a ground plan is the maximum of the both reciprocal Hausdorff distances as Figure 3e presents.

$$HD_i = \max(hd_i(O), hd_i(G)) \quad (3.6)$$

- The maximum of all Hausdorff distances computed for each eliminated element makes the modified Hausdorff distance between the initial and generalized building ground plans.

$$HD_{Mod} = \max(HD) \quad (3.7)$$

Figure 3 demonstrates clearly that the described approach reaches the expected outcome in this case. The important advantage of this method consists in its simplicity. The desired correspondence between the segments of the original and generalized ground plan can be determined while avoiding an additional analysis of the situation. Nevertheless, further examination is required to determine in which other cases and to what extent there will be differences in the use of the modified compared to the original Hausdorff distance, which the next section is dedicated to.

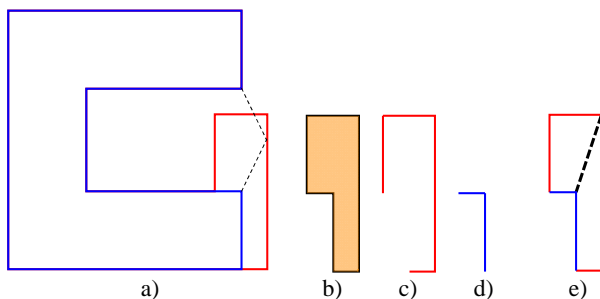


Figure 3. Calculation of the modified Hausdorff distance

In anticipation of the following metrics, all computing steps listed above remain the same over all metrics described in this paper. Only step 3 is altered in order to get more precise results. First, a shortest path is calculated from two given points (subsection 4.1) and then a maximum shortest path between the line segments of the symmetric difference is presented afterwards (subsection 4.2).

3.2 Original vs. modified Hausdorff distance

The examples given in Figure 2 illustrate that the discrepancy between the outcomes using the original and the modified Hausdorff distances, is basically caused by the diversely determined correspondence between the elements of the initial and generalized ground plan. In this case, the distance of the original characteristic is shorter than the modified one. At the same time, both characteristics provide identical results, if the maximum points between the two contours lie on the outline of the eliminated element. However, even the equality of the resulting distances computed by means of these two characteristics cannot guarantee that this maximum distance completely runs within the eliminated element. As it is shown in Figure 4, it is typically the case if the symmetric difference of the two ground plans includes concave

elements. Then the parts of the direct connection of the two points from the modified distance can lie outside the symmetric difference.

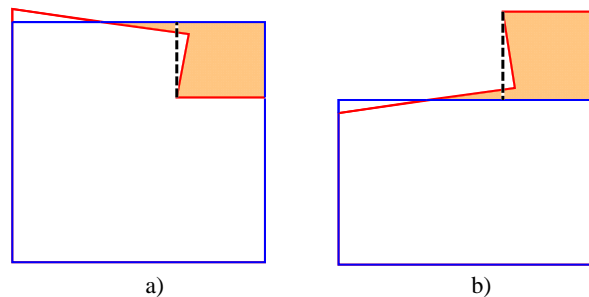


Figure 4. Hausdorff distance and modified Hausdorff distance for concave elements.

These examples open an interesting question, how in general the maximum distance between two contours containing concave elements is to be evaluated. The result of the original and the modified Hausdorff distance is compared in Figure 5a and b. The Hausdorff distance identifies in this case a perpendicular line from one of the topmost vertices of the original contour to the upper edge of the generalized one. As the perpendicular point does not belong to the symmetric difference, the modified characteristic provides a different result. Assuming that the distance between the two contours has to be measured only within the eliminated element, it should follow the fine dashed line in Figure 5 c.

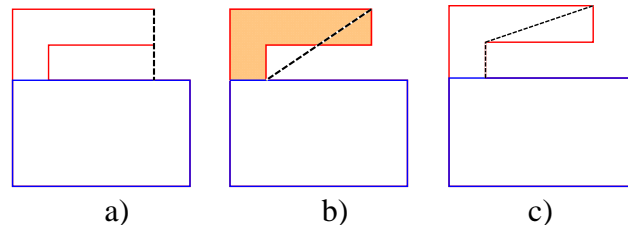


Figure 5. a) Hausdorff distance, b) modified Hausdorff distance, c) shortest path within the eliminated ground plan element

Let the third characteristic be identified as the shortest path within the eliminated element. Whereas the original and the modified Hausdorff distance characterize the maximum deviation between the outlines of the two objects, the shortest path rather defines the shape complexity of the eliminated element. Figure 5 demonstrates that this characteristic can also be considered for the quality evaluation of generalization along with the maximum distance between the original and generalized object. Section 4 discusses in detail the geometric meaning and methods for computing this distance.

4. MAXIMUM SHORTEST PATH

4.1 Maximum distance with regard to concave elements

The modified Hausdorff distance as well as the original characteristic is measured as a straight line between two points identified by its algorithmic computation. However, these points of maximum distance cannot always be connected with a single line segment on the condition that it must completely run within an eliminated element. Rather a polyline with several segments has to be used. The difference between the direct distance and the connected line segments that lie inside an element of the symmetric distance can be interpreted as its shape complexity. The computation of such a polyline can be associated with the shortest path problem from graph theory which consists of finding a path with minimal weight of all its connected edges between the two given nodes in a weighted graph (Sedgewick, 1988). The shortest path itself is an alternating sequence of nodes and edges never repeated (Gould, 1988).

The graph is generated on the basis of the corresponding polygon of the symmetric difference SD_i with n points and sides which contains the modified Hausdorff distance. Let $G = (p, q)$ be a weighted graph of order (number of nodes) $p = n$ or $p = n + 1$ if the modified distance consists of one perpendicular base point to a polygon side. Firstly, all possible connections between the graph nodes will be generated as illustrated in Figure 6a. In the next step, those edges that only partially run within an eliminated element will be rejected (cp. Figure 6b). The remaining connections, including the edges composing the contour of the eliminated element, will be considered as the edges of the new graph. Their weights w result from the Euclidean distance between the corresponding start nodes. Finally, the shortest path between the two given nodes of the generated graph is computed as it is, for instance, shown in Figure 6c. This single-pair shortest path can be solved, for example, by means of the Dijkstra algorithm (Cormen et al., 2001).

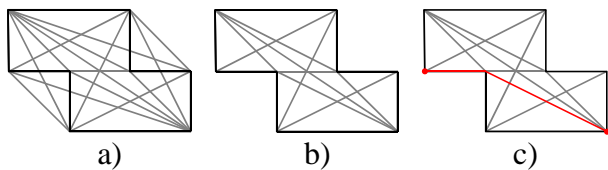


Figure 6. a) All connections between points of the eliminated element, b) the resulting graph, c) the shortest path within the eliminated element.

The start and end points of the shortest path are defined as the same points as the modified Hausdorff distance. This way, the two characteristics can be easily compared. The complexity of a shape is computed as described in Formula 4.1. Here, the higher the value, the more complex the shape of the eliminated element appears. This complexity metric can help a generalization process to (re-) consider elements for elimination during simplification.

$$Complexity = ShortestPath_{SinglePair} - HD_{Mod} \quad (4.1)$$

Thus, the main points of the presented approach are: the maximum distance can be calculated on basis of the shortest path within the eliminated part, so that it will consist not of a straight line, but of several line segments. And this shortest path is computed between the two specified points, identified as the most distant by the modified Hausdorff distance

4.2 The shortest path from point to contour

As explained in the previous subsection, the new characteristic for the shape complexity results from the comparison of the modified Hausdorff distance and the shortest path between the same vertices running within an eliminated element. Nevertheless, the shortest path calculated in such a fashion cannot be considered an adequate independent characteristic and this restricts its further interpretation. The main reason of this limitation is determined by the fact that the shortest path between the two given points must not necessarily coincide with the shortest path from a point on one contour to the other contour. Figure 7 demonstrates an appropriate example for such a case. Here, it can be clearly seen that the shortest path (thin dashed line) between the start and end points of the modified Hausdorff distance (bold dashed line) in Figure 7a is longer than the one between the original (red) and generalized (blue) line segments of the eliminated element (Figure 7b).

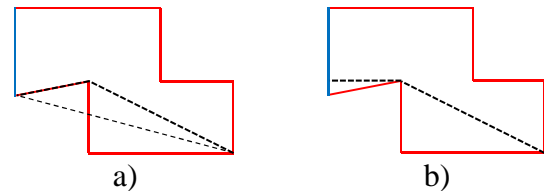


Figure 7. a) Shortest path between two specified points and b) from point to contour.

Therefore, the shortest path between the vertices of the modified Hausdorff distance cannot be accepted as a final solution for the maximum distance between the contours. In order to overcome this problem, an approach analog to the notion of the Hausdorff distance can be applied. The only difference consists in the use of the shortest paths instead of the Euclidean distances. Thus, the two most distant points belonging to the outline of an eliminated element and lying respectively on the intersection of this outline with the original and generalized contour must be determined. The final outcome results as the maximum from the maximum shortest paths between the original and generalized line segments of the symmetric difference computed reciprocally.

In the first step, the shortest paths from each point of the first contour to the other contour will be calculated. Subsequently, a weighted graph will be generated on the basis of this polygon as explained in subsection 4.1. The perpendiculars from all the vertices on one contour to each line segment of the other one will

also be included. As a result, the connection to the perpendicular foot points that are completely running within an eliminated element (black) will be added to the already generated edges (grey) of the graph. After weighting all of the edges of the new graph, the shortest paths from all the vertices of the initial contour to the other one can be computed successively.

First, the shortest paths from a start node to all other nodes of the graph will be computed. Then, the path with the minimum length will be picked from all the shortest paths from the start node to all the nodes belonging to the other contour. This path presents the sought-for shortest path from a given start point lying on the initial outline to the other outline. Figure 8b shows an arbitrary start point on the original contour marked with red and the alternatives for the end points of the shortest path marked with blue. The shortest path from this start point to the generalized contour is shown with a bold dashed line in Figure 8c.

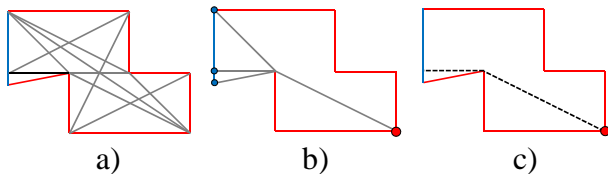


Figure 8. a) Additional connection (black) in the graph, b) shortest path alternatives from start point on original (red) to all end points on generalized contour (blue), c) result.

4.3 The maximum shortest path between the contours

In order to gain the final metric, the maximum from all shortest paths from each point of one contour to the other contour will be chosen, which can be defined as a one-sided maximum distance between the two contours. After identification of the one-sided maximum distance from the second contour to the initial one, the maximum of these two distances presents the maximum distance between the original and generalized contour to be determined within the given eliminated element.

5. CONCLUSION

In this work, it is demonstrated that the Hausdorff distance as well as its approximation, the vertex Hausdorff distance, do not always correspond to what a human viewer would perceive as a maximum difference between two building contours. As a result, two new characteristics were suggested, the calculation of which is analog to the Hausdorff distance and which are both based on the symmetric difference of two objects. This enables to compare the outlines of these objects not as a whole using the shortest distance between their points, but to evaluate their deviation within each eliminated element.

The first characteristic, the modified Hausdorff distance, identifies the two most distant points lying on the contour of the eliminated element by means of a straight line. Since in the case of concave elements, this distance can run out of the eliminated

element, a further characteristic, the maximum shortest path, was proposed. In contrast to the modified Hausdorff distance, it can consist of several line segments so that it completely remains within an eliminated element. It has to be noted, that this distance is really a length of the path. The difference between the modified Hausdorff distance and the shortest path between the same end nodes can be considered as the complexity of an eliminated element and used for evaluation of the importance of this element for the whole shape of an object.

Although the first characteristic is sufficient for most of cases, the shortest path enables to cover all of them without any exception though its computation is somewhat more complex. In the future, the suggested characteristics will be applied to large data sets in order to demonstrate their advantages.

6. REFERENCES

- Andrae, C., 2009. *Spatial Schema: ISO 19107 und ISO 19137 vorgestellt und erklärt*. OpenGIS essentials: Die Geo-Standards von OGC und ISO im Überblick. Wichmann, Heidelberg, 232 pp.
- Aspert, N., Santa-Cruz, D. and Ebrahimi, T., 2002. Mesh: Measuring Errors between Surfaces Using the Hausdorff Distance, In: *Proc. of the IEEE International Conference in Multimedia and Expo*, Lausanne, Switzerland, pp. 705-708.
- Blasby, D., Davis, M., Kim, D. and Ramsey, P., 2003. *GIS Conflation using Open Source Tools*, JCS Conflation Suite, pp. 32.
- Bouziani, M., Goïta, K. and He, D.-C., 2010. Automatic Change Detection of Buildings in Urban Environment from Very High Spatial Resolution Images Using Geodatabase Using Prior Knowledge. In: *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1): 143-153.
- Cignoni, P., C. Rocchini and Scopigno, R., 1998. Metro: Measuring Error on Simplified Surfaces. In: *Computer Graphics Forum*, 17(2): 167-174.
- Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C., 2001. *Introduction to Algorithms*. McGraw-Hill Higher Education.
- Gould, R., 1988. *Graph theory*. The Benjamin/Cummings Publishing Company.
- Guthe, M., Borodin, P. and Klein, R., 2005. Fast and Accurate Hausdorff Distance Calculation between Meshes. In: *Journal of WSCG*, 13(2): 41-48.
- Hangouët, J.-F., 1995. Computation of the Hausdorff Distance between Plane Vector Polylines. In: *Proceedings of the International Symposium on Computer-Assisted Cartography Charlotte*, North Carolina, pp. 1-10.
- Hangouët, J.-F., 2006. Spatial Data Quality Assessment and Documentation. In: R. Devillers and R. Jeansoulin (Editors), *Fundamentals of Spatial Data Quality. Geographical Information Systems Series*, pp. 211-235.

Herrlich, H., Hušek, M. and Preuß, G., 2002. Hausdorff-Metriken und Hyperräume, Felix Hausdorff, Gesammelte Werke: *Gründzüge der Mengenlehre*. Springer-Verlag, Berlin, Heidelberg, pp. 866.

ISO 19107, 2003. *Geographic Information - Spatial schema*.

Kreveld, M.v., 2006. Computational Geometry: Its objectives and relation to GIS. In: P.J.M.v. Oosterom and M.J.v. Kreveld (Editors), *Geo-information and Computational Geometry*. NCG, Delft, pp. 1-8.

Luebke, D. et al., 2003. *Level of Detail for 3D Graphics. Computer Graphics and Geometric Modeling*. Morgan Kaufmann, San Francisco, 490 pp.

Schlüter, D., 2001. *Hierarchisches Perzeptives Gruppieren mit Integration dualer Bildbeschreibungen*, Bielefeld, pp. 222.

Sedgewick, R., 1988. *Algorithms*. Computer Science. Addison-Wesley, 659 pp.

Volz, S., 2006. *Modellierung und Nutzung von Relationen zwischen Mehrfachrepräsentationen in Geo-Informationssystemen*. Doctoral Thesis, Universität Stuttgart.