

On the Evaluation and Comparison of Region Interpolation Methods

José Duarte¹, Paulo Dias¹ and José Moreira¹

¹ DETI/IEETA, University of Aveiro, Portugal
hfduarte@ua.pt

Abstract. This work presents a methodology for the comparison and evaluation of region interpolation methods proposed in the spatiotemporal databases literature. An evaluation performed using the methodology proposed is also presented. To the best of our knowledge this is the first time that such a methodology is discussed and presented in the spatiotemporal databases literature, and that region interpolation methods are evaluated and compared using the same dataset and using data from the evolution of real-world phenomena.

Keywords: Region Interpolation Problem, Spatiotemporal Databases, Evaluation, Comparison, Benchmark.

1 Introduction

Several methods have been proposed in the spatiotemporal databases literature to solve the region interpolation problem, i.e., to create moving regions from a set of observations. This allows the representation of the continuous evolution of a phenomenon in spatiotemporal databases, e.g., the evolution of a forest fire, the evolution of an oil spill or the evolution of biological cells. Region interpolation methods should be robust, correct, efficient and allow the development of operations for querying the evolution of deformable moving objects and the relationships that they establish with other objects. However, different methods can produce different results when used to represent the evolution of the same phenomenon. As a consequence, there is a need to compare and evaluate these methods. But, *How do we compare the output of two different methods? How do we measure the quality of the representation of the evolution of a phenomenon generated by a method? Which method is the best for a specific use case? Can an approximation error be defined between the representation and the actual evolution of a phenomenon?* To the best of our knowledge, these questions have not been discussed and considered in detail in the spatiotemporal databases literature. The contributions of this work are the following. We discuss and present a methodology for the comparison and evaluation of region interpolation methods, and we present an evaluation using the methodology presented here.

This paper is organized as follows. Section 2 presents a brief discussion on the state of the art of metrics and measures of similarity and a discussion about related work. Section 3 discusses and presents a methodology for the comparison and evalua-

tion of region interpolation methods. Section 4 presents an evaluation performed using the methodology proposed, and Section 5 presents the conclusions and future work.

2 Related Work

This section presents metrics and measures proposed in the state of the art to compute the similarity between two regions in 2D space, and related work presenting methodologies for the comparison and evaluation of different interpolations or interpolation methods.

2.1 Metrics to Compute the Similarity Between two Regions in 2D Space

Several metrics and measures¹ have been proposed to compute the similarity (dissimilarity) between two geometries or shapes [12], [2], [11]. These metrics and measures may be sensitive to certain characteristics of the data and may be appropriate for specific use cases or to measure specific characteristics [13]. To the best of our knowledge there is not a well-established set of metrics to evaluate and compare the quality of two different interpolations of a geometry.

In this work we are interested in metrics and measures to compute the similarity between two rigid geometries in 2D space with arbitrary complexity, e.g., a polygon and a multi-polygon. The geometries can have a different number of vertices and no correspondence between them is assumed.

The Procrustes distance is defined as the sum of the squared distances between the corresponding points of two discrete sets with the same number of elements [11]. The main disadvantages of using this metric include the following: (1) the distance between two geometries depends on the choice of the two corresponding sets of points (landmarks) representing them, (2) the landmarks have to be identified (this process may need the user intervention and can be subjective, e.g., *Which are the best landmarks to use and how many?*), (3) only part of the geometric information of the geometries is used to compute their distance.

The Hausdorff distance has been extensively studied in the literature [8]. It can be used to compute the dissimilarity between two sets of points with different sizes with no correspondences between them, and it has been used in different areas of study [13]. It is not rotation, reflection, and scale invariant [7], and it is very sensitive to noise and local distortion.

The Jaccard distance measures the dissimilarity between two finite sets, say A and B , and it is defined as $J_D = 1 - J(A, B)$, where $J(A, B) = |A \cap B| / |A \cup B|$ is the Jaccard Index. The Jaccard distance satisfies all the properties of a metric [6].

The PoLiS distance between two polygons a and b is defined as the average of the distances between each vertex in a and its closest point in the boundary of b [1]. Ac-

¹ A metric satisfies a set of well-defined properties. A measure is not a metric because it fails to satisfy one or more of these properties.

cording to the authors, it can be used with polygons with a different number of vertices (although it is not clear how it can be used to compute the similarity between, e.g., a multi polygon and a polygon with holes.), it is not topology independent, and seems to penalize missing estimated areas more than different generalizations of the boundary. The dissimilarity value can be underestimate if one of the polygons being compared has a much larger number of vertices. The authors also present a comparison between the Hausdorff, the Chamfer and the PoLiS distances when used to compute the similarity between building footprints. According to them, the Hausdorff distance can be thought of as a measure of highest dissimilarity and the normalized Chamfer distance as a measure of the overall average dissimilarity between two sets of points. Both distances are topology independent and sensitive to collinear points.

2.2 Methodologies to Compare Different Interpolations

In this work we are interested in methodologies to compare and evaluate the quality of two interpolations, in particular, interpolations of regions in 2D space.

In [3] the authors compare four different spatial interpolation methods for the generation of continuous surfaces from irregularly distributed data. The authors consider two different viewpoints, prediction and characterization, to evaluate the quality of the interpolation. Prediction considers that the best interpolation minimizes the prediction error at unknown points. Because the true value at an unsampled point is not known the authors use a procedure in which the interpolation is generated ignoring known data points. Then the values of those data points are compared with the values generated by the interpolation. Characterization considers that the surface generated by the interpolation must globally look like the actual surface that is not known. To overcome this situation, statistical characteristics obtained from the set of known data points are associated to the surface. Finally, indices are defined and used to measure the quality of the interpolation.

However, the methodology presented in [3] cannot be used as is in our context since, (1) region interpolation methods do not interpolate points, they interpolate 2D geometries (regions), and (2) we want to evaluate the quality of the interpolation not the quality of the surface generated as in the work cited.

Morphing techniques, used to interpolate a geometry between two known observations or states, are, in general, evaluated and compared qualitatively by observing their results, and region interpolation methods, proposed in the spatiotemporal databases literature, are in general evaluated w.r.t efficiency, robustness and genericity but not w.r.t the quality of the interpolation they generate. To the best of our knowledge there is no well-establish methodology to evaluate and compare the quality of two interpolations.

3 Methodology for the Evaluation and Comparison of Region Interpolation Methods for Spatiotemporal Databases

Real-world phenomena can go through different changes, e.g., (1) they can merge and split (e.g., cell division), (2) change to a different state (e.g., ice melting), leading to phenomena of appearing and disappearing, and (3) develop holes and concavities. These changes can occur in a more or less dynamic way following more or less well-defined patterns, e.g., the shape of fluids and of a forest fire can change dramatically in a short period of time, but the shape of an iceberg can change rather slowly. Each phenomenon can evolve differently depending on the circumstances with which it interacts. This makes the evaluation of the representation of their evolution complex. Furthermore, the level of complexity of a geometry representing the continuous evolution of a real-world phenomenon can change significantly during interpolation, e.g., it may start as a complex multi-polygon and change to a simple polygon during interpolation.

Evaluation can be qualitative and quantitative. Qualitative evaluation can be a powerful tool, e.g., the evaluation of an interpolation through visualization. However, certain characteristics are hard to evaluate visually. For example, when observing the evolution of an iceberg visually it may not be possible to detect area oscillations that, depending on its size and scale, can be more or less significant. A quantitative measurement of the evolution of the area during interpolation can measure these phenomena precisely. In general, a quantitative evaluation is less subjective, gives additional information about the geometries generated during interpolation, and quantifying things precisely can help make decisions, in particular, can help make decisions without the need for user intervention.

Assumption 1 A methodology for evaluation and comparison of region interpolation methods should consider these changes and complexity and include both a qualitative and a quantitative evaluation and comparison.

Assumption 2 The evolution of the characteristics of a phenomenon can be measured and quantified by studying (analyzing) the characteristics of the geometries generated during interpolation. These measurements can then be used for evaluating the quality of the interpolation w.r.t some known or expected values. For example, it seems to make sense to compare the similarity between the geometries generated during interpolation and a reference geometry or set of reference geometries.

Assumption 3 The study of the geometric similarity of the geometries generated during interpolation can give an idea about the characteristics (quality) of the interpolation method, e.g., *Do the in-between geometries change smoothly during interpolation or are there abrupt changes?*

3.1 Methodology

Following the previous assumptions, we construct a methodology by answering the following questions. *What to evaluate?*, *What are the characteristics of interest that should be evaluated?*, *Which metrics should be used?*, and *What is the meaning, the*

relevance, and the significance of the results obtained when using a specific metric or set of metrics? Our goal is to develop a methodology that is generic, not dataset, use case or application dependent. That (1) is efficient, to allow the processing of a large number of geometries with arbitrary complexity that can be very different from each other, both topologically and geometrically, (2) does not require the existence of a correspondence between the geometries being compared, and (3) uses metrics and measures, and methodologies used successfully in similar works.

To construct this methodology, we proceed as follows. We follow a strategy similar to [3], and we consider 3 perspectives for evaluation: the representation, the prediction and the characterization perspectives. The representation perspective evaluates the ability of a method to represent certain characteristics of interest that can occur during the evolution of real-world phenomena (this perspective is not presented in [3]). Table 1 presents the characteristics considered. The prediction perspective considers that a good interpolation minimizes the error at an unknown instant (not at a point as in [3]). The characterization perspective considers that the interpolation should globally represent an evolution similar to the actual evolution of the phenomena (here we consider the interpolation not a surface as in [3]).

Table 1. Representation Perspective Considered Characteristics.

Characteristic	Comments
Split, Merge	Ability to represent geometries that split or merge during interpolation. We define split as an evolution where n objects, faces or holes, form progressively from m objects, $n > m$. An object can only form from an object of the same type, and if k objects form from an object then $k > 1$. We define merge as an evolution where n objects form progressively from m objects, $m < n$. An object can only form from an object of the same type, and if k objects form from j objects then $k < j$, $k \geq 1$.
Concavities, Holes	1. Ability to represent the evolution of holes and concavities during interpolation, i.e., concavities and holes that exist in both the source and target observations. 2. Ability to represent the transformation of a concavity into n holes, $n \geq 1$, and a hole into m concavities, $m \geq 1$.
Translation, Rotation, Deformation	Ability to represent translation, rotation, and deformation, in particular non-uniform deformation, during interpolation.
Appearing, Disappearing	Ability to represent the appearance and disappearance of faces (regions), holes, and concavities during interpolation. We consider appearing and disappearing as an evolution that occurs progressively not instantaneously. Split and merge, and appearing and disappearing are considered different phenomena, i.e., objects do not appear (disappear) from (to) an existing object.

Representation. This perspective allows the methods to be characterized according to the types of phenomena that they can represent. Quantifying the ability of a method to represent a specific characteristic is, we believe, an open question for investigation.

Therefore, we propose a qualitative evaluation through visualization using a score defined in the set $\{0, 1, 2\}$, where 0 means that the method cannot represent the characteristic, 1 means that the method can represent the characteristic partially, and 2 means that the method can represent the characteristic. It is important to notice that in this perspective the quality of the representation is not evaluated.

Prediction. The geometry at an unknown (unsampled) instant is not known. We can, however, given a set of n observations of the evolution of a phenomenon, generate the interpolation between observations o_{i-1} and o_{i+1} and compare the result generated by the interpolation at instant t with the know observation o_i taken at that instant. Then, the prediction performance can be computed using, for example, the mean absolute error (MAE). Prediction performance can be evaluated w.r.t the area, the perimeter, the position, the geometric similarity, the number of faces, the number of concavities and the number of holes of the geometry, rotation and orientation (direction).

Characterization. We can assume that the actual evolution of a phenomenon is, in general, not known. However, we need some information to be used as a reference for comparison and evaluation. For example, Fig. 1 shows the evolution of a phenomenon generated by three region interpolation methods. *Which one generates the best representation?* This question can only be answered if we know or make assumptions about the actual evolution of the phenomenon. For example, (1) the evolution of the area or the perimeter during interpolation follows a specific pattern or function, (2) the geometry changes slowly or abruptly, (3) the geometry rotates a certain amount in a specific direction, (4) a certain characteristic or set of characteristics evolve in a certain way, e.g., a hole appears (disappears) during interpolation or a geometry splits into two geometries.

In this perspective we consider the evolution of the area, the perimeter, the position, rotation and direction, the number of faces and holes, and the evolution of the geometric similarity w.r.t a reference geometry or set of geometries during interpolation. This perspective includes a qualitative evaluation that is performed visually to help evaluate the quality of the representation.

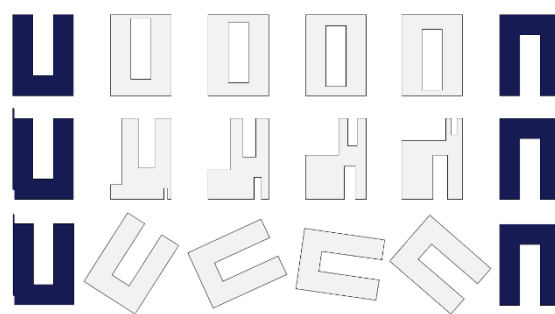


Fig. 1. Evolution of an object between two observations represented by three region interpolation methods. Librip (top). PySpatiotemporalGeom (middle) and Morphrip (bottom).

The previous discussion should answer the first two questions, i.e., ‘*What to evaluate?*’, and ‘*What are the characteristics of interest that should be evaluated?*’.

3.2 Which Metrics Should be Used?

We consider only rigid geometries in 2D space. We want to compute a full similarity between them, and we need to compute the similarity between two, potentially, very different shapes, e.g., between a polygon and a multi-polygon. In this work we do not consider the topology or a representation of the geometry to compute similarity. In the first case, a region interpolation method can impose the restriction that the topology of the geometry does not change during interpolation. Therefore, topological changes may not be appropriate to compare region interpolation methods. In the second case, we do not know how to construct an appropriate representation of a multi-polygon with an arbitrary complexity, and even if a representation could be constructed, e.g., by considering its polygons individually, we still do not know how to compare two such representations.

In the following, we define the properties of interest that a metric or measure should ideally have to be used in our scenario. Given three shapes, a , b and c , and the distance between two shapes $d(a, b)$, the metrics or measures to be used should have the following properties:

- *Nonnegativity*, $d(a, b) \geq 0$. *Identity*, $d(a, a) = 0$. *Uniqueness*, if $d(a, b) = 0$ implies $a = b$. *Symmetry*, $d(a, b) = d(b, a)$. *Triangular Inequality*, $d(a, b) + d(b, c) \geq d(a, c)$. Invariant to translation, rotation, and scale. Because the shape of the geometries being compared is expected to have non-uniform affine transformations, the property of being invariant under affine transformations is not considered critical in our scenario².
- Can handle shapes with an arbitrary geometry, with a different number of vertices, e.g., to compute the similarity between a polygon and a multi-polygon.
- Can compute the similarity between a large number of shapes in a reasonable amount of time.
- Use geometric characteristics (information), e.g., about the boundary, to compare two geometries.
- Are generic, i.e., are not specific to a dataset or application and can compute the similarity without the need of any prior knowledge of other shapes or the existence of a correspondence between the geometries being compared.

Therefore, we select the Jaccard distance, the Hausdorff distance³, and the Chamfer distance to compute the geometric similarity between two geometries. This is mostly motivated by the fact that these metrics can compute the similarity between geome-

² It is important to notice that if the geometries are aligned as a preprocessing step to measure their similarity, the alignment considered can influence the measurement. Computing the best alignment between two polygons can be hard and time-consuming.

³ The Hausdorff distance can be reformulated to satisfy all properties of a metric, i.e., $d_H(a, b) = \max(d(a, b), d(b, a))$, where $d(a, b)$ is the usual Hausdorff distance between a and b .

tries with an arbitrary complexity and a correspondence between them is not required. They compute the similarity from different perspectives. The Jaccard distance measures the dissimilarity w.r.t the area, the Hausdorff distance measures the highest dissimilarity, and the Chamfer distance measures the overall average dissimilarity.

3.3 Datasets Used for Evaluation

Ideally, diverse datasets of real data should be used for evaluation, in particular: datasets with specific cases where a particular characteristic occurs during interpolation should be used for representation evaluation, datasets with a ground truth should be used for prediction evaluation, and different datasets should be used for characterization evaluation.

4 Experimental Results

In this section we compare and evaluate three methods, two region interpolation methods proposed in the spatiotemporal databases literature and a method that is being developed to solve the region interpolation problem as part of a PhD thesis, i.e., Librip [5], PySpatiotemporalGeom [9] and Morphrip, respectively. Currently, Morphrip can only handle simple polygons. The evaluation is performed using a dataset extracted from a sequence of satellite images tracking the evolution of two icebergs in the Antarctic [10], following the methodology discussed and presented in this work.

It is important to note that (1) this is a preliminary evaluation, (2) the main objective is to apply the methodology presented in this work in practice, and (3) the methods being evaluated provide different options that can have a significant impact on the result of the interpolation. For example, PySpatiotemporalGeom allows the user to provide an explicit matching between the components of the source and target geometries (observations), i.e., that component s_i in the source geometry corresponds to component t_j in the target geometry (we use this option). Librip provides several matching strategies, including strategies defined by the user (we use the default option provided by the library).

In this study, we use the implementations provided by the authors. In some cases, the implementation provided (made available) is not a full implementation and/or has not been fully tested. As a consequence, we do not evaluate certain characteristics.

Finally, this section also serves the purpose of trying to answer the fourth question. *What is the meaning, the relevance, and the significance of the results obtained when using a specific metric or set of metrics?*

4.1 Representation Perspective Evaluation

We use synthetic data to perform the evaluation proposed in the representation perspective (the results are presented in Table 2). This is because we do not have real data where the various characteristics considered for evaluation occur. In Table 2, E.,

A., D., and # mean: evolving, appearing, disappearing, and not evaluated, respectively. We give score (see Section Methodology) 1 to Morphrip when evaluating rotation because although it can represent rotations, the method may consider that a rotation does not exist. We consider that the other two methods can represent rotations in the range $[0^\circ, [45^\circ, 90^\circ]^4$ (clockwise and counterclockwise) but we do not establish a precise range for each method. When evaluating E., A., and D. concavities we give a score of 1 to Librip because a concavity can unexpectedly evolve as a hole during interpolation, and a score of 1 to Morphrip because it can only handle simple polygons. When evaluating A. and D. faces we give a score of 1 to Librip because the existing face(s) disappear and then the new face(s) appear, i.e., it is not a continuous phenomenon (see Fig. 2 (bottom)). When evaluating Merge Faces we give a score of 2 to PySpatiotemporalGeom but notice that we are using the option that allows the user to define an explicit matching between the components of the observations (see Fig. 2 (top)). PySpatiotemporalGeom does not interpolate holes and faces appear instantaneously (see Fig. 2 (middle)).

Table 2. Representation Perspective Evaluation Results (Scores).

Characteristic	PySpatiotemporalGeom	Librip	Morphrip
Split Faces	0	0	0
Split Holes	0	0	0
Merge Faces	2	0	0
Merge Holes	0	0	0
Translation	2	2	2
Rotation			
45°	1	1	1
90°	0	0	1
180°	0	0	1
Deformation	2	2	2
E. Concavities	2	1	1
E. Holes	0	2	0
A. Concavities	2	1	1
A. Holes	0	2	0
A. Faces	0	1	0
D. Concavities	2	1	1
D. Holes	#	2	0
D. Faces	#	1	0
Concavity to Hole	0	0	0
Hole to Concavity	#	0	0

⁴ Here we do not use the terms small and large rotation because we would have to define them precisely.



Fig. 2. Merging faces using PySpatiotemporalGeom (top). Appearing faces using PySpatiotemporalGeom (middle) and Librip (bottom).

Table 2 shows that PySpatiotemporalGeom and Librip are able to represent a larger number of characteristics, and some characteristics are not represented by any of the methods. This could be an indication that there are still open problems for future investigation in the field. We also notice that a score in the range $\{0, 1, 2\}$ may be too coarse because two methods may represent a characteristic partially, but differently. It may also make sense to perform independent evaluations considering different types of geometries and characteristics that can occur during their evolution, e.g., perform an evaluation using only simple polygons or only polygons.

4.2 Prediction Perspective Evaluation

For the evaluation presented here we use observations taken from the evolution of an iceberg represented by a simple polygon with no holes, that changes relatively smoothly and slowly with a small rotation between consecutive observations (see Fig. 3). The prediction perspective evaluation was performed by generating moving regions between observations o_i and o_{i+2} and then comparing observation o_{i+1} with the geometry generated by the method at the middle of the interpolation between the two observations. Table 3 presents the results of the evaluation using the Mean Absolute Error (MAE). We did not remove outliers. In the following PySpatiotemporalGeom is abbreviated to PySptGeom to save space in the tables.



Fig. 3. Observations of the evolution of an iceberg.

When comparing PySpatiotemporalGeom and Librip, the area and the perimeter are the characteristics with the most significant differences. An important observation is the fact that these are global results that can be influenced by outliers, e.g., in the case of Librip the values observed for the perimeter and the Hausdorff distance are mostly affected by an outlier (the value of the first geometry ‘predicted’ by the method). Also, Librip obtained a small error w.r.t the expected number of holes. This is because Librip considers that the geometry develops a hole during interpolation.

Morphrip obtained the smallest error for the position, the area and the geometric similarity, i.e., for the Hausdorff and the Jaccard distances.

Table 3. Prediction Perspective Evaluation Results.

	PySptGeom	Librip	Morphrip
Characteristic	MAE	MAE	MAE
Position	2,53	2,51	2,20
Area	119,89	106,91	26,33
Perimeter	1,48	4,73	1,54
Hausdorff Distance	6,46	6,92	1,45
Jaccard Distance	0,16	0,18	0,05
N° Faces	0	0	0
N° Holes	0	0,14	0

4.3 Characterization Perspective Evaluation

The characterization perspective evaluation was performed by creating moving regions between pairs of consecutive observations, i.e., between observations o_i and o_{i+1} . Then, we collected 100 observations (generated during interpolation) from each moving region and studied the evolution of several characteristics of the geometry during interpolation (see Fig. 4 to help clarify this). The geometric similarity is measured after the two geometries being compared have been aligned by using the iterative closest point (ICP) algorithm presented in [4] w.r.t the source geometry (any geometry can be used here). It is important to note that the alignment can have an impact in the measurement.

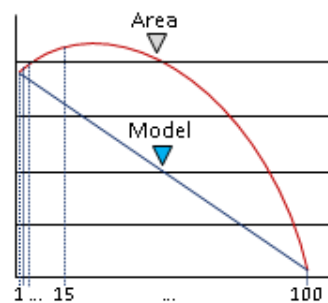


Fig. 4. Approximation of the evolution of the area during interpolation using the values taken from 100 observations, and the model assumed for the evolution of the area.

Table 4 presents the MAE of the area, the perimeter (P), the Hausdorff distance (H_D), and the Jaccard distance (J_D) computed using the results generated during interpolation and a model (function) assumed to represent the actual evolution of these characteristics. We assume a linear function as a model for all the characteristics considered, i.e., we assume that these characteristics evolve linearly during interpolation. PySpatiotemporalGeom seems to preserve the perimeter more than Librip, while Librip

seems to preserve more the area, and obtains better results for the Hausdorff and the Jaccard distances. In some cases, the two methods generate very similar results during interpolation. This explains why the maximum MAE values are equal for some characteristics, i.e., they occur during the evolution of a moving region where the two methods generate similar results. In average, Morphrip obtains the smallest values for the area and the geometric similarity, i.e., the Hausdorff and the Jaccard distances.

Table 4. MAE Measured during Interpolation.

	Librip	PySptGeom	Morphrip
MAE	Avg	Avg	Avg
Area	41,51	58,45	12,50
P	1,37	0,90	0,92
H _D	2,52	3,27	0,87
J _D	0,07	0,08	0,03

	Librip		PySptGeom		Morphrip	
MAE	Min	Max	Min	Max	Min	Max
Area	3,93	99,82	34,03	99,82	4,00	30,50
P	0,13	4,49	0,13	1,48	0,16	1,51
H _D	0,97	5,32	2,82	5,32	0,48	1,69
J _D	0,02	0,19	0,06	0,19	0,02	0,06

Table 5 presents the area minimum and maximum deviations, AmD and AMD, computed w.r.t the minimum and maximum area known values, respectively. The minimum and maximum perimeter deviations, PmD and PMD, computed in a similar way w.r.t the perimeter known values, the Hausdorff maximum deviation, HMD, computed w.r.t the Hausdorff distance between the two known observation (the source and target observation), and the maximum Jaccard distance deviation, JMD, computed in a similar way w.r.t the Jaccard distance between the two observations. The values in the table are presented as a percentage w.r.t the known minimum and maximum values, respectively. See also Fig. 5 to help clarify this.

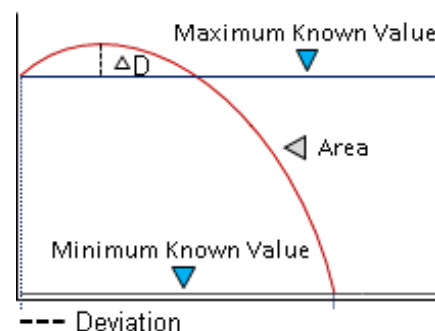


Fig. 5. Evolution of the area during interpolation. Maximum and minimum known values given by the source and target observations, and the maximum deviation of the area w.r.t the maximum known value.

Table 5. Deviation Measured during Interpolation.

	Librip		PySptGeom		Morphrip	
%	Avg		Avg		Avg	
AmD	0		0		0	
AMD	0,04		0,06		0,01	
PmD	0		0		0	
PMD	0,01		0		0	
HMD	2,86		4,36		0,02	
JMD	2,33		2,61		0,07	

	Librip		PySptGeom		Morphrip	
%	Min	Max	Min	Max	Min	Max
AmD	0	0	0	0	0	0
AMD	0,01	0,10	0,04	0,10	0	0,02
PmD	0	0	0	0	0	0
PMD	0,04	0	0	0	0	0
HMD	0	7,44	1,43	7,44	0	0,12
JMD	0	6,78	0,13	6,78	0	0,25

On average, the area and the perimeter deviations are relatively small and a deviation w.r.t the minimum known values was not observed. The maximum area deviation occurs between observations 41 and 67, and 47 and 56 for Librip and PySpatiotemporalGeom, respectively. The Hausdorff maximum deviation occurs between observations 45 and 100 (the last observation). Its average maximum deviation is 2.86% and 4.36% respectively. The Jaccard maximum deviation occurs between observations 37 and 100. The average maximum deviation is 2.33% and 2.61% for Librip and PySpatiotemporalGeom, respectively. The maximum HMD and JMD values are similar for both methods because they were observed during the evolution of a moving region where the two methods generate similar results.

By observing the interpolation generated by Librip and PySpatiotemporalGeom and the graph of the functions that approximate the values collected during interpolation (see Fig. 6 for an example) we observe the following. The evolution of the perimeter is similar in both methods and approximately linear. In this case, the difference in the results obtained is mostly explained by the fact that in some cases Librip generates holes during interpolation (see Fig. 7 (top)). When a small rotation exists Librip tends to represent the evolution of the area in a slightly quadratic way. As the rotation increases this representation tends to be clearly quadratic. PySpatiotemporalGeom represents the evolution of the area in a quadratic way in both situations. In general, the representation is non-monotonic. Except for the cases where the two methods generate similar results and where Librip generates holes during interpolation, the geometric similarity (measured using the Hausdorff and the Jaccard distance) evolves more smoothly in the case of Librip.

In the case of PySpatiotemporalGeom the Hausdorff distance tends to increase until a maximum is reached, usually approx. at the middle of the interpolation, and then starts to decrease (see Fig. 6). Morphrip obtained the smallest deviation and the geo-

metric similarity evolves more smoothly during interpolation. Fig. 6 presents an approximation of the evolution of the Hausdorff distance during the interpolation between two observations when using the three methods.

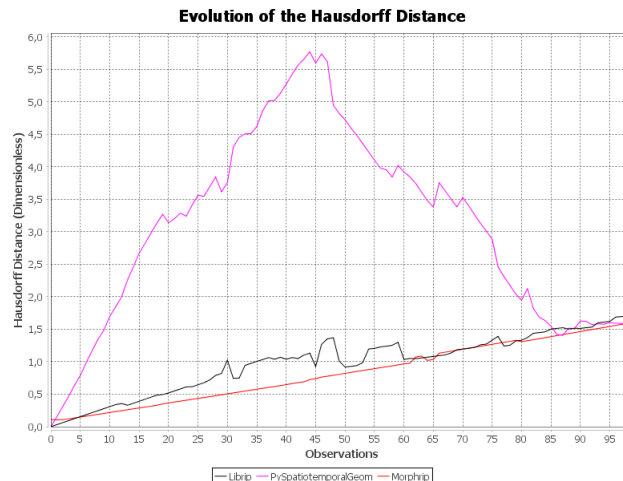


Fig. 6. Evolution of the Hausdorff distance between a pair of consecutive observations when using Librip (black), PySpatiotemporalGeom (magenta), and Morphrip (red).

Discussion.

Morphrip seems to generate the most natural interpolation. Librip may generate holes that appear and disappear during interpolation even if the known observations used to create the moving region are simple polygons with no holes. PySpatiotemporalGeom seems to generate an interpolation where the area increases at the middle. The maximum area deviation observed was 10%. The results presented here should not be generalized and used to fully characterize the methods evaluated. It is important to note that the main goal of this work is to propose a methodology for evaluation and comparison of region interpolation methods. A more detailed comparison and evaluation of these methods requires appropriate datasets. The area is a good metric for finding phenomena of shrinkage and enlargement during interpolation. The perimeter is difficult to interpret, e.g., the perimeter can increase in two different scenarios, one where the area of a geometry increases and the other where it decreases. However, it can give relevant information if the evolution of the perimeter is an important factor to consider. The Jaccard Index may give results easier to interpret than the Jaccard distance, and both metrics are not a good measure of similarity. We need a scale to interpret the results of the Hausdorff distance and establish the degree of similarity between two geometries. It would be interesting to use more metrics to establish similarity. Overall, we need a variety of datasets representing the evolution of phenomena with different characteristics, e.g., rigid, fluid, dynamic and biological, that have a large number of observations, provide a ground truth, and represent all the changes that can occur during the evolution of a phenomenon.

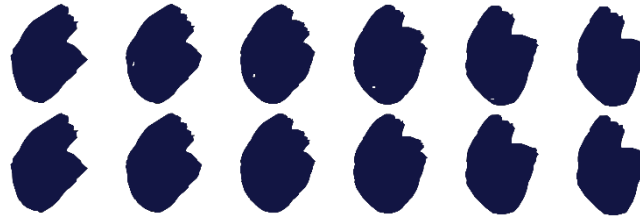


Fig. 7. Example where Librip generates a hole during interpolation. Librip (top). PySpatiotemporalGeom (bottom).

5 Conclusions and Future Work

We presented and discussed a methodology to compare and evaluate region interpolation methods developed in the context of spatiotemporal databases. We performed an evaluation using real data to show the methodology being used in practice. To the best of our knowledge, this is the first time that such a methodology is discussed in detail in the spatiotemporal databases literature and that region interpolation methods are compared and evaluated using real data. Future work includes (i) finding robust methods proposed in the literature to align two geometries before measuring their similarity, and to find the rotation and the orientation of a geometry, and (ii) performing a more detailed evaluation using more datasets. This is a first step that can make contributions to the establishment of a benchmark for the evaluation of region interpolation methods.

6 Data and Software Availability

Software used:

- Librip <http://dna.fernuni-hagen.de/secondo/> (VM Appliance, Version. 4.1.2).
- PySpatiotemporalGeom <https://pypi.org/project/pyspatiotemporalgeom/>.
- Morphrip <https://hfduarte@bitbucket.org/hfduarte/morphrip.git>.

The results were obtained and collected using a java application running in a Windows 10 OS. This application uses the region interpolation methods through a middle layer that can communicate with C++, Python and Secondo. The application and the data used in this paper can be found at:

<https://hfduarte@bitbucket.org/hfduarte/evaluation-of-region-interpolation-methods.git>.

A video showing the application is available at:

https://drive.google.com/open?id=1FzhOUTcEXhXHoEOfDcBfea_w0LFWs6E6.

Acknowledgments We thank the anonymous reviewers and Mark McKenney for their comments and suggestions to improve this paper. José Duarte has a research grant awarded by the FCT (Foundation for Science and Technology) with reference PD/BD/142879/2018. This work is partially

funded by FCT in the context of the projects UID/CEC/00127/2013 and POCI-01-0145-FEDER-032636.

References

1. Avbelj, J., Müller, R., Bamler, R.: A Metric for Polygon Comparison and Building Extraction Evaluation. *IEEE Geoscience and Remote Sensing Letters* 12(1), 170–174 (2015).
2. Basri, R. et al.: Determining the similarity of deformable shapes. *Vision Research* 38(15), 2365–2385 (1998).
3. Caruso, C., Quarta, F.: Interpolation methods comparison. *Computers & Mathematics with Applications* 35(12), 109–126 (1998).
4. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354–3361. IEEE, Providence, RI, USA (2012).
5. Heinz, F., Güting, R. H.: Robust high-quality interpolation of regions to moving regions. *GeoInformatica* 20(3), 385–413 (2016).
6. Kosub, S.: A note on the triangle inequality for the Jaccard distance. *Journal of Pattern Recognition Letters* 120(1), 36–38 (2018).
7. Latecki, L. J., Lakamper, R.: Shape similarity measure based on correspondence of visual parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(10), 1185–1190 (2000).
8. M. Tang, M. Lee, Kim, Young, J.: Interactive Hausdorff distance computation for general polygonal models. *ACM Transactions on Graphics* 28(3), 9 (2009).
9. McKenney, M., Frye, R.: Generating Moving Regions from Snapshots of Complex Regions. *ACM Trans. Spatial Algorithms Systems* 1(1), 1–30 (2015).
10. Moreira, J., Dias, P., Amaral, P.: Representation of continuously changing data over time and space. In: *IEEE International Conference on e-Science*, pp. 111–119. IEEE, Baltimore, MD, USA (2016).
11. Perner, P.: Determining the Similarity between Two Arbitrary 2-D Shapes and its Application to Biological Objects. *International Journal of Computer & Software Engineering* 3(139), 12 (2018).
12. Veltkamp, R. C.: Shape matching: Similarity measures and algorithms. In: *International Conference on Shape Modeling and Applications*, pp. 188–197. IEEE, Genova, Italy (2001).
13. Veltkamp, R. C., Latecki, L. J.: Properties and Performance of Shape Similarity Measures. In: Batagelj, V., Bock, H., Ferligoj, A., Žiberna, A. (eds) *Data Science and Classification*, vol. 1, pp. 47–56. Springer, Heidelberg (2006).