Coastline matching via a graph-based approach

Costas Panagiotakis1 · Smaragda Markaki1 · Eleni Kokinou2 · Harris Papadakis3

Received: 20 May 2022 / Accepted: 1 September 2022 / Published online: 30 September 2022
© The Author(s) 2022

Abstract
This paper studies the problem of unsupervised detection of geometrically similar fragments (segments) in curves, in the context of boundary matching. The goal is to determine all pairs of sub-curves that are geometrically similar, under local scale invariance. In particular, we aim to locate the existence of a similar section (independent of length and/or orientation) in the second curve, to a section of the first curve, as indicated by the user. The proposed approach is based on a suitable distance matrix of the two given curves. Additionally, a suitable objective function is proposed to capture the trade-off between the similarity of the common sub-sequences and their lengths. The goal of the algorithm is to minimize this objective function via an efficient graph-based approach that capitalizes on Dynamic Time Warping to compare the two sub-curves. We apply the proposed technique in the context of geometric matching of coastline pairs. This application is crucial for investigating the forcing factors related to the coastline evolution. The proposed method was successfully applied to global coastline data, yielding a bipartite graph with analytical point-to-point correspondences.

Keywords Coastline evolution · Coastline matching · Curve matching · Dynamic time warping

1 Introduction

A significant problem in computer vision and pattern recognition is the unsupervised detection of geometrically similar fragments (segments) in curves. In this work, we study the detection of common subcurves (commonalities) in two curves (e.g. 2D object boundaries) of arbitrary shape, without prior knowledge of their quantity, shape, and length. The user specifies the part on the first curve to be retrieved on the second given curve. The sub-curve matching is implemented under local scale invariance, i.e., in addition to rotation and translation, local scaling transformations can take place to allow matching of curves after various shape changes (e.g., due to coastline evolution). Figure 1 shows an application of the proposed method to represent the local scaling effect in the matching results between two 2D objects. Based on the detected commonalities (Fig. 1a) we present the matching solutions (Fig. 1b-d) from the best to the worst, according to the proposed distance score. In the heading of Fig. 1b-d, the distance rating is given.

This problem can be viewed as an extension of the jigsaw puzzle problem [1–4], where only rotational and translational transformations are supported. The jigsaw puzzle problem requires assembling a given number of puzzle pieces of arbitrary or specific shape (e.g., square) from a broken single object. The goal is to reassemble the given pieces in such a way that the original shape is recovered. Solving jigsaw puzzles is an amusing task and a challenging mathematical and engineering problem with a variety of applications. Scientific literature proposes methods that rely on computer vision algorithms, such as contour or feature recognition and deep-learning techniques that use color, texture, and shape-related features [5, 6].
Similar to the jigsaw puzzles problem, faced by knowledge recognition and data mining, it involves the detection of many mutual patterns in a mutual signal [7], time series [8], or string [9]. The detected commonalities are called motifs [8]. The goal of motif discovery algorithms is to find the repeating patterns for a range of lengths. The enumerated motifs can be used in many data summarization, retrieval, and compression tasks. The problem of detecting commonalities in videos was introduced by [10] as Temporal Commonalities Discovery (TCD). It has been applied to pairs of images called time series, which contain information about facial expressions or motion detection. It has also been treated in [11] as the detection of co-actions in multiple image sequences, in [12] as video co-segmentation for action extraction, and currently in [13, 14] as the temporal action of co-segmentation in video pairs. Other relevant problems include periodicity detection [15] and video co-summarization [16]. Commonality detection in image sequences is an interesting topic applied in various fields such as data mining, content retrieval, audio and natural language processing, image and video analysis, bioinformatics, economics, physics and others.

In the present work, we study commonalities between two 2D curves $A$ and $B$ under the assumption that the user specifies on the first curve the part to be retrieved on the second given curve. In this constellation, a pair of subcurves of $A$ and $B$, which is also perceived as a path in the distance matrix $D$ of all pairwise distances between the elements (frames) of $A$ and $B$, is a possible commonality. This type of matrix is shown in Fig. 1a, which shows a heat map on which three commonalities are shown as white curves. Large and small pairwise distances are represented by warm and cold colors, respectively. How dissimilar the partial curves of a commonality candidate are is reflected by the total cost of a path. Paths with low costs correspond to similar subcurves, while high costs correspond to dissimilar subcurves. A short path tends to indicate low cost. Nevertheless, it corresponds to a commonality of small subcurves, which is not so remarkable. According to [14], the trade-off between the length of the commonality path and its cost can be balanced to find all paths in $D$ that truly correspond to common sub-curves in $A$ and $B$. This is achieved by detecting multiple commonalities. Figure 1a (top left and bottom right) also presents the two given curves and the tangential angles (see Section 3.1). In addition, Fig. 1b, c and d show the point correspondences indicated by gray dotted lines for the three commonalities. The three solutions are projected by gray dotted lines on the two given curves $A$ and $B$ that are drawn with blue and yellow colors, respectively.

Given the potential commonality for the two subcurves $s_A$ and $s_B$ of two curves $A$ and $B$, the equivalent commonality path and associated cost can be predicted by applying Dynamic Time Warping (DTW) [15]. DTW is a widely used method for optimal, nonlinear time alignment of two sequences and has been used for time series alignment [17]. According to [14], the exhaustive scheme of multiple commonality detection has a complexity of $O(|A|^3|B|^3)$ which is restrictive even for input boundaries with a small number of points. The method proposed in this paper, can be considered as a modification of MUCOS algorithm [14], which discovers multiple commonalities for input pairs of video sequences of many thousands of frames. This work also exploits the properties of the distance matrix $D$ and achieves state of the art performance with a computational complexity of $O(|A||B|\log(|A||B|))$ via an efficient graph-based approach.

The main contribution of this work concerns the application of unsupervised discovery to global coastal data to identify geometrically matched coastline pairs and produce a bipartite graph with the analytical point-to-point correspondences under local scale and rotation invariance in coastline pairs. Coastline matching is very important in various scientific disciplines that study coastal evolution at different time scales (in the present, in history and in geology) [18]. By implementing this work, we provide a highly accurate solution in coastline matching aiming to contribute to the study of
• erosional processes in the coastal environment [19]
• sedimentary budget of the coasts because of the coastal behaviour and mobility [20]
• seafloor spreading, coastal uplift and subsidence related to internal geodynamics

Another contribution is that the proposed methodology provides features for the study of the coastline evolution.

2 Related work

The partial matching of the boundaries of two two-dimensional objects can be seen as an application of the apictorial jigsaw puzzle reassembling problem, which is a very interesting scientific field. Apictorial jigsaw puzzles are called the puzzles in which the only information available is the geometric features of the pieces.

In addition to their theoretical significance, these puzzles also have excellent applications in various fields, such as archaeology for reconstruction of fragmented artefacts, art restoration and in biology [1, 3]. Such applications aim at the automatic reconstruction of a large number of irregularly shaped fragments coming from one or even several fragmented objects or shredded documents [2, 21], resulting in the elimination of tedious and time-consuming manual work. The detection of geometrical matching of coastlines can be considered as another interesting application of the apictorial jigsaw puzzle reassembling problem, which, to our knowledge, is studied for the first time in this work. The idea of geometrical coastline matching is based on the work of Alfred Wegener who suggested the hypothesis of continental drift (the basis for today’s model of plate tectonics) in 1912 [22].

Coastline matching with emphasis on computing the maximum and average discrepancy between sinuous lines was studied by [23]. Digital Elevation Models (DEM)s are used to define the boundaries of objects such as mountains and coastlines [24, 25]. Coastlines can also be automatically detected by sea and land segmentation in global or local range [26]. Detection accuracy and resolution can be improved by using a large number of repetitive, high-resolution, multispectral satellite images [27].

The remainder of this work is arranged as follows: Section 3 presents the various aspects of the problem, the properties of the problem, and ideas for dealing with the computational complexity of the problem. Section 4 presents algorithmic solutions applying the proposed formulation and the results. Section 5 presents the developmental features of the coastline. The experimental procedure and the results are given in Section 6. Finally, in Section 7 the summary and conclusions of this research are included.

3 Formulation

Let’s assume two given polygonal curves $A$ and $B$ of lengths $|A|$ and $|B|$, with $N_A$ and $N_B$ the number of their points, respectively. We also consider that the curves $A$ and $B$ have been resampled so that the length is the same for each line segment of the curves. This consideration allows us to calculate the commonalities and correspondences over pixels of equal size in the distance matrix.

The repetition sampling of equal length was also applied to the distance matrix of image sequences [14], which means that a common path on the distance matrix with slope equal to one, represents the correspondences between line segments of equal length one to one without scaling effect. To simplify notation, we assume that $A$ and $B$ have already been resampled. Therefore, in the following we use the notations $A$ and $B$ for the sampled curves. Figure 2(a) depicts the scheme of the proposed system architecture.

3.1 Distance matrix and commonalities

We define the $N_A \times N_B$ matrix $D$ of pairwise distances for the points of the curves $A$ and $B$ (see Fig. 1). Depending on the type of curves and applications, different distance functions can be used to consider the relation of tangential angle (slope) ($D_a(i, j)$) and normalized translation ($D_t(i, j)$) differences between points $A(i), i \in \{1, ..., N_A\}$ and $B(j), j \in \{1, ..., N_B\}$ of the given curves $A$ and $B$, respectively. The parameter $\lambda \in [0, 1]$ controls the relation of $D_a(i, j)$ and $D_t(i, j)$ (e.g. $\lambda = 0.8$ gives greater weight to the tangential angle). Therefore, a general form of the distance matrix is given in Eq. 1:

$$D(i, j) = \lambda \cdot D_a(i, j) + (1 - \lambda) \cdot D_t(i, j)$$ (1)
$$D_a(i, j) = |\phi(A(i)) - \phi(B(j))|$$ (2)
$$D_t(i, j) = \pi \cdot \frac{\|A(i) - B(j)\|_2 - m}{M - m}$$ (3)
$$m = \min_{i,j}\|A(i) - B(j)\|_2$$ (4)
$$M = \max_{i,j}\|A(i) - B(j)\|_2$$ (5)

where the $\phi(A(i))$ denotes the tangential angle at the point $A(i)$, which is equal to the angle between the tangent line to the curve at the point $A(i)$ and the x-axis. $m$ denotes the minimum Euclidean distance between the two curves and is used for normalization to compare angles and translations.

A connected path of points $(x_i, y_i)$ represents a commonality in the image coordinate system. It holds for the image coordinate system that $\forall x_i, y_i$, $x_{i+1} \leq x_i, y_i \leq y_{i+1}$ as shown in Fig. 1. The coordinates of both curves $A$ and $B$ are arranged clockwise. Apart from this restriction, the paths are able to start and end anywhere in $D$. In this method, the given subcurve $q_A$ of the curve $A$ is defined
as \( q_A = [s_A, e_A], e_A \geq s_A \), where \( s_A \) and \( e_A \) are the starting and ending points of the subcurve, respectively. A commonality \( c = (q_A, q_B) \) of \( A \) and \( B \) is a pair of subcurves \( q_A \) (of \( A \)) and \( q_B \) (of \( B \)) that are similar subcurves (a problem solution). It should be noticed that \( q_A \) is given by the user and the goal is to compute the subcurve \( q_B \) that match with \( q_A \).

An example of a specific distance matrix \( D \) resulting from comparing the frames of two curves is presented in Fig. 1 which illustrates three commonalities. A commonality \( c = (q_A, q_B) \) describes the rectangle \( b_c \) in \( D \), where \( (s_A, e_A) \) is the lower left corner of the rectangle and \( (e_A, s_B) \) the upper right corner. The real correlation between the points of the subcurves \( q_A \) and \( q_B \) is determined by the path of minimum cost in \( D \) connecting \((s_A, e_A)\) to \((e_A, s_B)\). Basically, \( b_c \) is the bounding box of this path. The main notations used in this paper are summarized in Table 1.

### 3.2 Graph approach

Next, we create a directed graph \( G = (V, E) \) in order to be able to evaluate a commonality (see Section 3.3). The set of nodes \( V \) of this graph corresponds to the points of the distance matrix \( D \). Each node \( u = (i, j) \) is linked to a directed edge \( e \) to three nodes of its neighborhood \( v_1 = (i - 1, j), v_2 = (i - 1, j + 1) \) and \( v_3 = (i, j + 1) \). Let \( v \in \{v_1, v_2, v_3\} \), then each edge \( e = (u, v) \in E \) is associated with a weight

\[
w(u, v) = \begin{cases} D(u) + D(v), & v = (i - 1, j + 1) \\ \sigma \cdot (D(u) + D(v)), & \end{cases}
\]

(6)

The edge cost \( w(u, v) \) represents the sum of values of the distance matrix \( D \) at points \( u \) and \( v \) weighted by a factor \( \sigma \geq 1 \). In this work, we set \( \sigma = \sqrt{2} \). The factor \( \sigma \) is used to penalise the cases where there is no one-by-one correspondence.

Based on the aforementioned definition of graph \( G \), the commonalities can be naturally defined as paths on the directed graph \( G \). Similarly with [14], \( G \) is mainly used on the evaluation of the commonalities via the application of the Dijkstra algorithm on the Graph \( G \), as explained in the next Section.

### 3.3 Evaluating a commonality

The solution of the single commonality detection problem aims at finding the commonality \( c^* \) that minimizes a well-defined objective function \( \omega (c^* = \text{argmin}_c \ \omega (c)) \). Two sub-curves \( q_A = [s_A, e_A] \) and \( q_B = [s_B, e_B] \) define the candidate commonality \( c = (q_A, q_B) \). Similarly with [14], to evaluate this similarity, we propose an objective function \( \omega \) consisting of two terms:

- The cost \( P(c) \) of the commonality \( c \), defined as the cost of the shortest path from node \( u = (s_A, s_B) \) to node \( v = (e_A, e_B) \), can be estimated by applying Dijkstra’s Algorithm [28] on graph \( G \). The \( P(c) \) is equivalent to the sum of \( w(u, v) \) along the determined minimal path for \( c \). Therefore, the larger the cost, the less preferred the commonality \( c \).
- The product \( A(c) = (e_A - s_A) \cdot (e_B - s_B) \), which is analogous to the sizes of the two sub-curves.
Therefore, the objective function $\omega$ should favour the corresponding larger subcurves.

There is a trade-off between the terms $A(c)$ and $P(c)$. Large $A(c)$ commonalities are to be preferred. At the same time, as $A(c)$ increases, so does $P(c)$. This trade-off is represented by the definition of the objective function $\frac{P(c)}{A(c)}$ [14]. In this work, the term $e_A - s_A$ is constant, since the $q_A = [s_A, e_A]$ is given by the user. Therefore, the objective function can be simplified to:

$$\omega(c) = \frac{P(c)}{e_B - s_B}. \tag{7}$$

### 3.4 Rotation invariant

The cost $P(c)$ used in the objective function of Eq. 7 does not support global rotations. This means that if an object is rotated, then the shortest path, as defined above, may not correspond to the requested solution. This is due to the fact that the weights are constant in the requested commonality path ($p$). More precisely, the values of the weights are equal to the global rotation angle, meaning that the sum of the weights square deviation of the weights ($SSD(p)$) corresponding to the constituent edges should be minimized. Equation 8 provides the sum of square deviation ($SSD$) of path $p$, where $e$ and $w(e)$ denote the edge of path $p$ and its weight, respectively. $\mu(p)$ denotes the average value of edge weights over the path $p$.

$$SSD(p) = \sum_{e \in p} (w(e) - \mu(p))^2 \tag{8}$$

$$\mu(p) = \frac{1}{p} \sum_{e \in p} w(e) \tag{9}$$

This problem is similar to Quadratic Shortest Path Problem which is NP-hard [29, 30]. The Quadratic Shortest Path Problem requires minimizing a quadratic objective function given the constraints of the shortest path. The objective of this problem is to detect the path with the lowest expected cost, while keeping the variance of the cost below a certain threshold.

Therefore, when applications require global rotation, $P(c)$ should be equal to the square root of the sum of square deviation of the edge weights over the commonality path ($P(c) = \sqrt{SSD(c)}$) in the objective function (Eq. 7). Algorithm 1 describes the proposed method that computes the shortest path minimizing the $SSD$ of the weights (see Eq. 8). The input of this algorithm is the given graph $G$, the starting $u$ and the ending $v$ nodes of the path. It is an iterative method. In each iteration until convergence, it computes the shortest path $p$ with minimum sum of square deviation (using Dijkstra algorithm) under the assumption that the $\mu(p)$ is provided. In each iteration, $\mu(p)$ is updated and the new path is computed taking into account the new estimation of $\mu(p)$.

Therefore, according to the proposed algorithm, the requested shortest path $p$ is obtained by applying the Dijkstra’s to a new graph $G'$ (procedure $DijkstraSP$ in Algorithm 1). $G'$ is dynamically updated so that the weights of this graph are equal to the square difference $(w(e) - \mu(p))^2$. The procedure, that updates the graph weights, is denoted by $graphUpdate$ in Algorithm 1.

**Algorithm 1** The proposed shortest path estimation method that minimizes the sum of weights square deviation over path $p$.

4 Commonality Detection via CD-RCP

The **CD-RCP algorithm** CD-RCP solves the Commonalities Discovery problem based on Reduced search space that is derived from the Critical Points of distance matrix $D$ [14], in order to reduce the candidate solutions of the 2D parameter space.

**CD-RCP** algorithm is a simpler version of the **MUCOS** method used in [14], where a new graph consists of weakly connected components, each associated with a single commonality. According to **MUCOS** method, the Johnson’s algorithm and the optimization of the objective function were performed independently in each connected component, achieving the decomposition of the whole problem into several, simpler problems. In this work, **CD-RCP** was applied directly to the graph $G$ and operates on the 2D space ($[s_B, e_B]$) of all possible commonalities (given $s_A, e_A$) under the constraints described above in the description of the **CD-RCP** method, with an additional extra constraint added that the candidate start and end points of a commonality should belong to the set of critical points of $D$. The set of critical points ($CP$) is defined by the union of the following two sets of points:

- $SET_D$: the points that belong to local minima of $D$
• \( \text{SET}_{GD} \): the points that belong to the local minima/maxima of the gradient magnitude of \( D \) (\(| \nabla D(x, y) | \)).

\[
\begin{align*}
CP &= \text{SET}_D \cup \text{SET}_{GD} \quad (10) \\
\text{SET}_D &= \{(x, y) : D(x, y) \text{ is local min}\} \quad (11) \\
\text{SET}_{GD} &= \{(x, y) : | \nabla D(x, y) | \text{ is local max/min}\} \quad (12)
\end{align*}
\]

Then, to discover all commonalities of two sequences \( A, B \) in \( CP \) set, \( \text{CD-RCP} \) operates as follows:

1. Compute the commonality that minimizes the proposed object function of Eq. 7 over all possible pairs of points belonging to the \( CP \) set, until we obtain an accepted solution\(^1\).
2. The process is terminated when the desired number of commonalities (e.g., specified by the user) is reached. Another option is to use an unsupervised termination criterion that optimizes a suitable objective function similar to the object functions proposed in [14] for the multiple commonality detection problem.

The total computational cost of the \( CP-RCP \) algorithm is \( O(N_A \cdot N_B) \) for computing the distance matrix. Then, taking into account that the number of elements of \( CP \) set (\(| CP | \)), the total computational cost is given by \( O(| CP |^2 \cdot N_A \cdot N_B \cdot \log(N_A \cdot N_B)) \) for the \( \text{CD-RCP} \) execution.

5 Coastline evolution features

After performing commonality detection, the evolution of the coastline can be measured via the geometrical matching for pairs of coastlines. In this work, we have used the area based distortion (\( PA \)) feature to measure the distortion of the corresponding coastlines. \( PA \) yields zero distortion when the commonality is a line segment, meaning that there is a point-to-point matching \( l_{sc} = \{[s_A, e_A], [s_B, e_B]\} \), which is also consistent with human intuition.

The area based distortion (\( PA \)) is defined by the normalized area of the polygonal shape \( S_a(c) \). \( S_a(c) \) is given by the region lying between the given commonality \( c \) and the corresponding line segment \( l_{sc} \). The normalization is implemented by dividing the area of \( S_a(c) \) by half the area of the bounding box \( b_c \). \( PA \) can be seen as a distance metric between the given commonality curve and the corresponding line segment \( l_{sc} \), that measures the requested distortion. Figure 2(b) shows how the distortion metric \( PA \) is computed via a synthetic example. It depicts a synthetic commonality \( c \) with blue points. The yellow cells correspond to region \( S_a(c) \). According to this example, \( PA = 25\% \).

6 Experimental evaluation

Commonality detection methods that are similar to the \( \text{CD-RCP} \), were compared with state-of-the-art commonality detection methods, i.e., \( TCD \) [10], which was used by Guo et al. [12] and the Segmental DTW (SDTW) method [31], showing that the proposed schemata outperform other state-of-the-art methods for commonality detection [13, 14]. In this work, we experimentally evaluate the performance of \( \text{CD-RCP} \) on real-world datasets with coastlines. The code implementing the proposed method together with experimental results are publicly available at 2. In order to assess the computational efficiency of the proposed method and its applicability to coastlines, we measured the computational time as a function of distance matrix size for the case of Northeastern Africa-Arabian Peninsula. We performed the running time experiments on a laptop with an Intel Core i7 2.2GHz processor and 32GB of RAM memory. To study how the image size affects the computational time, the distance matrix has been resized according to three predefined percentages (25%, 50%, 100%) of the original size of 1 MP distance matrix. The corresponding computational time of \( \text{CD-RCP} \) is 1 sec, 0.7 sec and 0.5 sec for 1 MP, 0.5 MP and 0.25 MP distance matrices.

6.1 Datasets

The experimental evaluation of \( \text{CD-RCP} \) was performed for both synthetic (Fig. 1) and real data. The real datasets of coastlines correspond to various areas around the world and come from the European Environment Agency\(^3\), and the National Oceanic and Atmospheric Administration (NOAA), National Centers for Environmental Information\(^4\). Specifically, we present the following pairs of coastlines, most of them share common geological origin:

1. Hudson’s Bay (Canada), Coats-Southampton Islands (Fig. 3), these areas present the same structural and paleogeographic evolution in previous geological times [32].
2. Aegean Sea (Greece), Paros-Naxos (Fig. 4(b)) subjected to geologically recent rapid subsidence [33].

\(^1\)A solution is accepted if and only if it has no intersections with the already accepted commonalities.

\(^2\)https://sites.google.com/site/costaspanagiotakis/research/CDRCP

\(^3\)https://www.eea.europa.eu/data-and-maps/data/eea-coastline-for-analysis-1/gis-data/europe-coastline-shapefile

\(^4\)https://www.ngdc.noaa.gov/mgg/shorelines/data/gshhg/latest/
3. Aegean Sea (Greece), Milos-Kimolos (Fig. 4(d)) that belong to the Milos volcanic field [34].
4. Ionian Sea (Greece), Kefalonia-Zakynthos (Fig. 4(f)) where internal geodynamics strongly affects the modern geomorphology of this area [35].
5. Australia-New Guinea (Fig. 4(h)), sea level rise 8000 years ago, caused flooding of the land bridge between Australia and New Guinea [36].
6. Indonesian Archipelago, Borneo-Sulawesi Islands (Fig. 4(j)) where plate tectonics and the spatiotemporal
variations of environments are considered important factors related to biogeography and accumulation of natural res [37].

7. Caribbean Sea, Cuba-Haiti Islands (Fig. 4(l)), these areas present the same structural and paleogeographic evolution in previous geological times [38].

8. Northeastern Africa-Arabian Peninsula separated by the Red sea (Fig. 4(m)), that is a young ocean basin experiencing rapid geomorphological and environmental changes [39].

9. Africa-South America (Fig. 4(p)), remnants of the Gondwana supercontinent, the geological linking between Africa and South America is crucial for the spatial reconstruction of the Gondwana supercontinent [40].

6.2 Matching results

Hereafter, we present the matching of the five pairs of coastlines showing the results of the CD-RCP method (Figs. 3 and 4). Figure 3 depicts the results of the different stages of the proposed CD-RCP algorithm for the Coats and Southampton coastal pair. Figure 3(a) and (b) show the tangential angles after re-sampling and the corresponding boundaries of the Coats and Southampton Islands (Canada). Figure 3(c) shows the error about the detected commonality path. Figure 3(d) shows the detected commonality (white curve) projected on the corresponding distance matrix. Finally, the point correspondences (grey dotted lines) of the detected commonalities projected on the two given coastlines are provided in Fig. 3(e). Similarly with the previous example, we present the experimental results of the proposed CD-RCP algorithm for the pairs of Paros-Naxos (Fig. 4(a) and (b)), Milos-Kimolos (Fig. 4(c) and (d)), Kefalonia-Zakynthos (Fig. 4(e) and (f)) in the Aegean and Ionian Sea. Additionally, the CD-RCP algorithm has been successfully applied to the Australia-New Guinea (Fig. 4(g) and (h)), Borneo-Sulawesi Islands (Fig. 4(i) and (j)), Cuba-Haiti Islands (Fig. 4(k) and (l)), Northeastern Africa-Arabian Peninsula separated by the Red sea (Fig. 4(m) and (n)) and Africa-South America (Fig. 4(o) and (p)). Ground truth data [33, 34, 36, 39, 41] confirm the common geological origin of the coastline’s pairs used in the present work, thus strengthening the results of the CD-RCP method.

The proposed method is robust under scale and rotation transforms (see 3.4). To study the scale and rotation invariance, we have applied several random scale and rotation transforms on real data getting similar results. Figure 5 depicts two such examples under different scales and rotations for Southampton Island scale = 25%, rotation = 40 degrees (Fig. 5(a)) and scale = 50%, rotation = −45 degrees (Fig. 5(b)). In both cases, the results are quite similar with the original one (Fig. 5(e)).

Hereafter, we present the experimental results concerning the distortion feature analyzed in Section 5. Table 2 depicts the area based distortion (AD), the corresponding coastline length (Length), and the distance (Distance) between the corresponding coastlines for the nine pairs of coastlines (see Section 5). Finally, the last column shows the ratio between the Distance and the Length. The pairs are sorted in ascending order according to the AD distortion metric. The distortion AD metric is correlated to the ratio Distance/Length, with about 0.63 Pearson’s correlation coefficient, meaning that as the ratio Distance/Length increases, the distortion also probably increases. According to Table 2, the pair Northeastern Africa-Arabian Peninsula yields the lowest coastline distortion. This is because the rifting that separated the Arabian Peninsula from Northeastern Africa and formed

![Fig. 5](https://via.placeholder.com/150)

This is the detected commonality projected on the two given coastlines of Coats and Southampton Islands under different scales and rotations for Southampton Island (a) scale = 25%, rotation = 40 degrees (b) scale = 50%, rotation = −45 degrees.

 Springer
Table 2  Results of several features for the nine pairs of coastlines. The pairs are sorted in ascending order according to AD

| Pairs of coastlines                  | AD   | Coastline length (km) | Coastline distance (km) | Ratio |
|--------------------------------------|------|-----------------------|-------------------------|-------|
| Africa-Arabian Peninsula             | 6.8% | 3000                  | 300                     | 0.10  |
| Coats and Southampton Islands        | 8.5% | 90                    | 85                      | 0.94  |
| Cuba-Haiti                           | 9.5% | 350                   | 200                     | 0.57  |
| Africa-South America                 | 11.3%| 7000                  | 4500                    | 0.64  |
| Australia-New Guinea                 | 11.5%| 1300                  | 900                     | 0.69  |
| Borneo-Sulawesi                      | 11.6%| 600                   | 300                     | 0.50  |
| Kefalonia-Zakynthos                  | 13.8%| 43                    | 35                      | 0.81  |
| Paros and Naxos                      | 18.0%| 10                    | 9                       | 0.90  |
| Milos and Kimolos                    | 27.3%| 2                     | 2                       | 1.00  |

the Gulf of Aden and Red Sea, is geologically young (about 25 million years ago). The low coastline distortion means that the coastlines of Northeastern Africa-Arabian Peninsula still can retain common geomorphological characteristics even though these two areas experience rapid geomorphological and environmental changes [39]. To the opposite, the coastline pairs of Paros and Naxos, and Milos and Kimolos Islands reveal the highest coastline distortion. For both cases the relative high distortion value could be due to the internal geodynamic activity and the strongly prevailing oceanographic conditions in this area that rapidly change the coastline morphology.

7 Conclusions

In this work, the problem of partial matching between two given curves is solved by a graph-based approach (CD-RCP algorithm). The proposed method has been successfully applied to global coastal data to identify geometrically matched coastline pairs and produce a bipartite graph with the analytical point-to-point correspondences as well as the local matching accuracy under local scale and rotation invariance. In addition, previous works on geological and environmental topics confirm the results of the distortion for different pairs of coastlines around the world. In conclusion, the CD-RCP algorithm is capable of geometrically matching pairs of coastlines and can be applied to other pattern recognition applications.

Acknowledgements This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH - CREATE - INNOVATE B cycle (project code: T2EDK-03135).

Funding Open access funding provided by HEAL-Link Greece.

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding authors on reasonable request.

Declarations

Conflict of Interests Authors declare that they have no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Paumard, M.-M., Picard, D., Tabia, H.: Deepuzzle: Solving visual jigsaw puzzles with deep learning and shortest path optimization. IEEE Trans. Image Process. 29, 3569–3581 (2020)
2. Naiman, A.E., Farber, E., Stein, Y.: Physical match Informatica 43(2) (2019)
3. Huroyan, V., Lerman, G., Wu, H.-T.: Solving jigsaw puzzles by the graph connection laplacian. SIAM J. Imaging Sci. 13(4), 1717–1753 (2020)
4. Markaki, S., Panagiotakis, C.: Jigsaw puzzle solving techniques and applications: a survey. Vis. Comput. 1–17 (2022)
5. Noroozi, M., Favaro, P.: Unsupervised learning of visual representations by solving jigsaw puzzles. In: European Conference on Computer Vision, pp. 69–84. Springer (2016)
6. Ostertag, C., Beurton-Aimar, M.: Matching ostraca fragments using a siamese neural network. Pattern Recogn. Lett. 131, 336–340 (2020)
7. Mueen, A., Chavoshi, N.: Enumeration of time series motifs of all lengths. Knowl. Inf. Syst. 45(1), 105–132 (2015)
8. Fuchs, E., Gruber, T., Nitschke, J., Sick, B.: On-line motif detection in time series with swiftmotif. Pattern Recogn. 42(11), 3015–3031 (2009)

9. Azmi, A.M., Al-Ssulami, A.M.: Discovering common recurrent patterns in multiple strings over large alphabets. Pattern Recogn. Lett. 54, 75–81 (2015)

10. Chu, W.-S., Zhou, F., Torre, F.D.L.: Unsupervised temporal commonality discovery. In: European Conference on Computer Vision, pp. 373–387. Springer (2012)

11. Xiong, C., Corso, J.J.: Coaction discovery: segmentation of common actions across multiple videos. In: International Workshop on Multimedia Data Mining, pp. 17–24. ACM (2012)

12. Guo, J., Li, Z., Cheong, L.-F., Zhou, S.Z.: Video co-segmentation for meaningful action extraction. In: IEEE International Conference on Computer Vision, pp. 2232–2239. IEEE (2013)

13. Papoutsakis, K., Panagiotakis, C., Argyros, A.: Temporal action co-segmentation in 3D motion capture data and videos. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)

14. Panagiotakis, C., Papoutsakis, K., Argyros, A.: A graph-based approach for detecting common actions in motion capture data and videos. Pattern Recogn. 79, 1–11 (2018)

15. Panagiotakis, C., Karvounas, G., Argyros, A.: Unsupervised detection of periodic segments in videos. In: IEEE International Conference on Image Processing, pp. 923–927. IEEE (2018)

16. Chu, W.-S., Song, Y., Jaimes, A.: Video Co-Summarization: Video Summarization by Visual Co-Occurrence. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)

17. Morel, M., Achar, C., Kulpa, R., Dubuisson, S.: Time-series averaging using constrained dynamic time warping with tolerance. Pattern Recogn. 74, 77–89 (2018)

18. Garcin, M., Le Cozannet, G.: The driving factors of coastal evolution: toward a systemic approach. In: Climate Change and Sea Level Rise: Coastal Vulnerability and Societal Impacts (2013)

19. Chandrasekar, N., Mujabar, P.S.: Computer application on evaluating beach sediment erosion and accretion from profile survey data. Comput. Geosci. 44(4), 503–508 (2010)

20. Baux, N., Murat, A., Poizot, E., Méar, Y., Gregoire, G., Lesourd, S., Dauvin, J.-C.: An innovative geo-statistical sediment trend analysis using geochemical data to highlight sediment sources and transport. Comput. Geosci. 26(2), 263–278 (2022)

21. Lalitha, K., Das, S., Menon, A., Varghese, K.: Graph-based clustering for apictorial jigsaw puzzles of hand shredded content-less pages. In: International Conference on Intelligent Human Computer Interaction, pp. 135–147. Springer (2016)

22. Wegener, A.: Die entstehung der kontinente. Geol. Rundsch. 3(4), 276–292 (1912)

23. Masclet, A., Devogele, T., Le Berre, I., Hénaff, A.: Coastline matching process based on the discrete fréchet distance. In: Progress in Spatial Data Handling, pp. 383–400. Springer (2006)

24. Panagiotakis, C., Kokinou, E.: Unsupervised detection of topographic highs with arbitrary basal shapes based on volume evolution of isocontours. Computers & Geosciences 102, 22–33 (2017)

25. Leonowicz, A.M., Jenny, B., Hurni, L.: Automatic generation of hypsometric layers for small-scale maps. Computers & geosciences 35(10), 2074–2083 (2009)

26. Liu, C., Xiao, Y., Yang, J.: A coastline detection method in polarimetric sar images mixing the region-based and edge-based active contour models. IEEE Trans. Geosci. Remote Sens. 55(7), 3735–3747 (2017)

27. Dai, C., Howat, I.M., Larour, E., Husbys, E.: Coastline extraction from repeat high resolution satellite imagery. Remote Sens. Environ. 229, 260–270 (2019)

28. Dijkstra, E.W. et al.: A note on two problems in connexion with graphs. Numerische mathematik 1(1), 269–271 (1959)

29. Rostami, B., Malucelli, F., Frey, D., Buchheim, C.: On the quadratic shortest path problem. In: International Symposium on Experimental Algorithms, pp. 379–390. Springer (2015)

30. Hu, H., Sotirop, R.: On solving the quadratic shortest path problem. Inform. J. Comput. 32(2), 219–233 (2020)

31. Park, A.S., Glass, J.R.: Unsupervised pattern discovery in speech. IEEE Transactions on Audio Speech, and Language Processing 16(1), 186–197 (2008)

32. Heywood, W.W., WW, H., BV, S.: Geology of southampton, coats, and mansel islands, district of keewatin northwest territories (1976)

33. Evelyridou, N., Melini, D., Pirazzoli, P., Vassilopoulos, A.: Evidence of repeated late holocene rapid subsidence in the syl cyclades (greece) deduced from submerged notches. Int. J. Earth Sci. 103(1), 381–395 (2014)

34. Vougioukalakis, G.E., Satow, C.G., Druitt, T.H.: Volcanism of the south aegaean volcanic arc. Elements: An International Magazine of Mineralogy, Geochemistry, and Petrology 15(3), 159–164 (2019)

35. Kokinou, E., Papadimitriou, E., Karakostas, V., Kamberis, E., Vallianatos, F.: The kefalonia transform zone (offshore western greece) with special emphasis to its prolongation towards the ionian abyssal plain. Mar. Geophys. Res. 27(4), 241–252 (2006)

36. Kennett, B.L., Chopping, R., Blewett, R.S.: Preview of the australian continent: a geophysical synthesis. Preview (199),39–48 (2019)

37. Wilson, M.E., Moss, S.J.: Cenozoic palaeogeographic evolution of sulawesi and borneo. Palaeoecogr. Palaeoclimatol. Palaeoecol. 145(4), 303–337 (1999)

38. Calais, E., De Lépinay, B.M.: Strike-slip tectonic processes in the northern caribbean between Cuba and hispaniola (windward passage). Mar. Geophys. Res. 17(1), 63–95 (1995)

39. Rasul, N.M., Stewart, I.C.: The Red Sea. Springer, Berlin (2015)

40. Konopásék, J., Sláma, J., Košler, J.: Linking the basement geology along the africa-south america coasts in the south atlantic. Precambrian Res. 280, 221–230 (2016)

41. Lykousis, V.: Sea-level changes and shelf break prograding sequences during the last 400 ka in the aegaean margins: subsidence rates and palaeogeographic implications. Cont. Shelf Res. 29(16), 2037–2044 (2009)

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.