Evaluating the Topological Quality of Watermarked Vector Maps

Ahmed Abubahia[a,*], Mihaela Cocea[a]

[a]*School of Computing, University of Portsmouth, Buckingham Building, Lion Terrace, Portsmouth PO1 3HE, UK*

Abstract

The pervasive use and exchange of digital content led to increased efforts in the research community for efficient approaches to protect intellectual property rights. While watermarking techniques have been used extensively for raster image format, watermarking approaches for the vector map format have been largely inspired from existing image watermarking techniques, without due consideration to the suitability of these techniques for this different data format. A key requirement of any watermarking approach of vector data is the preservation of the topological quality of the watermarked data. This is sometimes referred to as the invisibility of the watermark. For vector map data, the topological quality and invisibility are fundamentally different, but currently submerged into one and measured with error metrics borrowed from image watermarking, such as Root Mean Squared Error (RMSE) and Peak Signal to Noise Ratio (PSNR). Over the last 10 year, the research community on watermarking vector map data has repeatedly posed that error metrics alone are not appropriate for the evaluation of watermarked vector map topological quality. In this paper, a metric for measuring topological quality by measuring topological distortions is proposed based on topological properties of polygon-based vector maps. To evaluate the proposed metric, experiments with controlled watermarking capacity (i.e. how much is embedded) were run on maps of various sizes, using

*Corresponding author
Email addresses: ahmed.abubahia@port.ac.uk (Ahmed Abubahia), mihaela.cocea@port.ac.uk (Mihaela Cocea)

two popular embedding approaches, i.e. coordinate-based and distance-based embedding. The results indicate that the metrics allow comparisons between watermarked maps of different sizes and of different watermark sizes, and, thus, can be used to assess the quality of watermarked vector maps. The advantages and limitations of the proposed metric are discussed and further research directions are highlighted towards an agreed metric by the research community.
Keywords: digital watermarking, disclosure, fidelity metric, gap, geospatial map, information security, overlap, topology preservation.

1. Introduction

Geographical data have become widely available in digital format due to the advancement in computer devices, database systems, mapping applications and IT (Information Technology) [1, 2]. While the wide spread of geographical applications has brought many benefits for IT consumers [3, 4], it has also increased the necessity to protect digital geographical data from illegal distribution and modification [5–11].

Geographical data can be categorized into two types: vector and raster data [4, 12]. Vector data represents geographical information by using basic geometrical shapes such as points, lines and polygons [13], while raster data represents information in a matrix of cells or pixels of uniform size (e.g. satellite image data). Most geographical systems represent data in vector format [6, 14].

Watermarking of vector map data has been researched for the last 2 decades as a solution for the protection of this type of geographical data [12, 15–21]. It aims to conceal a watermark into the digital asset within a specific tolerance, which would not cause a considerable change so that the usability of the watermarked asset is not affected.

The vector map watermarking approaches can be categorised into two main categories: coordinate-based approaches [2], and distance-based approaches [22]. In coordinate-based approaches, the watermark is hidden in the the Cartesian coordinates' values within a specific tolerance, while in distance-based approaches, the watermark is hidden within the relations/links between the Cartesian coordinates, represented as distance measurements.

A key requirement of any watermarking approach is the quality preservation in the watermarked data [12, 23]. In the context of vector data, the quality preservation expresses that the original vector map is not affected by the concealed watermark, and is referred to as *fidelity*. Most often this is defined as the perceptual degree of similarity between the original vector map and the watermarked vector map. In the context of images (although used with vector map data as well) it is referred to as *invisibility*. In both cases, the emphasis is on the perceptual perspective [24] and is measured with error metrics, such as RMSE (Root Mean Squared Error) and PSNR (Peak Signal to Noise Ratio) (which is based on mean squared error). More details about the metrics used for invisibility of vector data can be found in [12, 25, 26].

While in the context of image watermarking the invisibility of the watermark can be taken to mean that the original image has preserved its quality [27], in the context of vector data, the quality of the map needs to be assessed in terms of the preservation of its topological properties, i.e. the geometrical shapes have not been distorted in the watermarking process. Although the need for a metric to assess topological quality preservation has been repeatedly highlighted [12, 28–30], few research works looked into this aspect [29, 31–34]. These works discussed the importance of topology preservation, and for particular applications looked at the effect of watermarking on some topological properties. To the best of our knowledge, a metric for quantifying topological distortion that can be used for assessing watermarked vector map topological quality has not yet been proposed.

In this paper, a metric based on topological properties of polygon-based maps is proposed. Here, the focus is on three topological rules, stating that the polygons need to be closed, that they should not have gaps between them and that they should not overlap. Consequently, a metric that quantifies to what degree these rules are broken is presented in this paper, i.e. how many polygon disclosures, gaps and overlaps are present, in proportion to watermark size. To evaluate the metric, experiments with the two different embedding approaches mentioned above and controlled watermarking capacity (i.e. how much is embedded) were run on maps of various sizes.

The rest of this paper is organized as follows: Section 2 reviews previous work on topology preservation in the context of digital vector map watermarking. Section 3 introduces the proposed metrics for measuring the polygon disclosure, overlap and gap aspects. Section 4 describes the experiments, including the data used and the experimental setup for the evaluation of the proposed metric. Section 5 discusses the experimental results, while Section 6 concludes the paper and outlines directions for future work.

2. Related Work

In this section, the topological aspects of vector data and the importance of their preservation are briefly outlined. Also an overview of previous work is introduced in relevance to addressing the issue of topological preservation when assessing watermarked vector map quality.

Unlike raster image data, vector map data has to follow topological rules that specify constraints for the shapes, e.g. lines and polygons, used in vector maps. The development of vector maps GIS tools (e.g. ArcGIS) [35] allows the identification of these errors, which allows them to be fixed. The value of the vector maps is related to the precision of the data, which allows spatial analysis [36]. While it is accepted that watermarking without any effect on the precision of vector map data is not possible [31], it is also clear that measuring the loss of precision only with error metrics, without checking the topology preservation, is not a good way to evaluate watermarked vector map data quality.

A recent review [12] outlines that the most used metrics for watermarked vector map fidelity are RMSE and PSNR, which are both error metrics based on the mean square error. The output of error metrics gives an indication of the precise loss caused by the watermarking process. Over the last 10 years, the research community on watermarking vector map data has repeatedly posed that error metrics are not appropriate for the evaluation of watermarked vector map topological quality [12, 28, 33].

A limited number of works have discussed topology preservation in the evaluation of watermarked vector maps [29, 31–33, 37]. These works are outlined below. In [31, 32], the authors used what they call an intersection test to verify if modifications occurred in the topology of line-based maps – more specifically, they assessed if lines that intersected previously to watermarking still intersect and if lines that should not intersect still do not intersect after watermarking. They report that they compared the values of the test before and after the watermark embedding, without details of how this was done, and that based on that comparison they concluded that topology was preserved.

In [29], the authors looked at polygon closure, data topology, error analysis and visual analysis. They also point out that in previous work data quality is mainly assessed through error metrics borrowed from image watermarking. They focused on tools for data inspection of watermarked vector data that allows visual identification of polygon disclosure, self-intersect, self-overlay and overlay for lines.

Like [29], in [33] the authors also focus on the visual inspection of topological issues without proposing a metric to quantify them; however, through this visual inspection, they stress the need for watermarking approaches that retain the topology of vector data and that the error analysis on its own is not an appropriate way of evaluating watermarking vector data approaches. In more recent work [37], data accuracy (i.e. the difference in coordinates values between the original and the watermarked map[1]) is discussed in relation to watermarked vector data quality of polyline-based maps. They talk about the assessment of distortion, but they only look at data accuracy and assess it with error metrics.

In summary, previous work highlighted the importance of topology preservation and proposed visual inspection for identifying distortions after water-

[1]some research uses the term fidelity to mean both data accuracy and invisibility; other research distinguishes between these terms, which is also the case for the work discussed here

marking. In this paper, to take this work further, a metric for quantifying topological distortions of polygon-based vector maps is proposed. The next section describes the proposed metric.

3. Metric for topological distortion

This section presents the proposed metric for judging the topological quality of watermarked GIS vector maps in line with the required standards for spatial data analysis tasks. Such standards are identified by several organisations working with and regulating the use of spatial data. Here, this paper follows the topological rules defined by the Environmental Systems Research Institute (ESRI), which supports the OCG[2] and ISO/TC211[3] geospatial standards.

ESRI defined a set of polygon-based shapefiles topology rules [4] to ensure the quality of polygon maps for spatial analysis tasks. In relation to the research of digital vector map watermarking, the significant rules are:

- Each polygon must be in the form of closed shape. A polygon is defined by a series of points, with the first point being the same as the last point; if the first and the last point are not the same, the polygon is not closed.
- Polygons must not overlap each other. This rule specifies that the interior of polygons must not overlap; polygons can only share edges or vertices.
- The map must not have gaps between polygons. This rule specifies that there should be no voids within a polygon or between neighboring polygons, so that all polygons form a continuous surface.

In this paper, three metrics are proposed in relation to these rules by quantifying the number of times the rules are broken proportionately to the size of the

[2] `http://www.opengeospatial.org/docs/is`

[3] `http://www.isotc211.org/`

[4] `http://help.arcgis.com/en/arcgisdesktop/10.0/help/001t/pdf/topology_rules_poster.pdf`

watermark. Also an overall metric as an average of the three metrics is defined, which can be used to compare topological problems across different watermarking approaches and map sizes. The metrics and the way they are calculated are described in the following subsections.

3.1. Polygon Disclosure

The polygon shape is formed by a sequence of vertices where the coordinates of the first point and the last point must be the same. Polygon disclosure occurs when this constraint is not met, i.e. the coordinates of the first and the last point are different.

In the watermarking process, there is a potential of having the polygon disclosure issue since the process of inserting the watermark is modifying the redundant bits of data, and the modification of different points may be done in different ways. For example, adding a watermark bit of 1 to the first point, while adding a watermark bit of -1 to the last point, would lead to disclosure.

Consequently, it is important to assess whether the polygon closure has been affected by the watermarking process. For this purpose, the condition used is that the coordinate value pair of the first point and the coordinate value pair of the last point must be the same, as shown in Equations (1) and (2).

$$F_x = L_x \tag{1}$$

and

$$F_y = L_y \tag{2}$$

where F_x is the x-coordinate of the first point, L_x is the x-coordinate of the last point, F_y is the y-coordinate of the first point and L_y is the y-coordinate of the last point.

The metric for polygon disclosure in the watermarked map is defined in Equation (3) as the proportion of disclosed polygons from all watermarked polygons:

$$M_1 = \frac{\sum_{i=1}^{n_w} d_i}{n_w} \tag{3}$$

where M_1 represents the disclosure metric, n_w represents the number of watermarked polygons and d_i is defined as in Equation (4):

$$d_i = \begin{cases} 1, & \text{if } F_x \neq L_x \\ 1, & \text{if } F_y \neq L_y \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

for each polygon i, where i takes values from 1 to n_w.

3.2. Overlap and Gap Identification

The overlap within the map polygons is a potential issue after inserting the watermark bits. This affects the map topology against the rule that the interior of polygons must not overlap, which means that an area cannot be shared by two or more polygons, i.e. polygons can only share edges or vertices. For example, the satisfaction of this topology rule is important for modeling administrative boundaries, such as voting districts, postal codes or land cover type.

The gaps between the map polygons could also be a consequence of the watermark insertion process, which has the effect of creating voids between adjacent polygons, while the topology rule requires that all polygons must form a continuous surface. This rule is significant in the context of spatial data analysis because it changes the perimeter of the surface. For example, when polygons define the type of soil in a particular area, there should be no gaps between polygons, i.e. the entire area needs to be defined in terms of the soil type; a gap would mean that the soil type (for the surface defined by this gap) is not known.

Algorithm 1 shows how the number of overlaps and gaps are identified. The *inpolygon* function in Matlab is used for this purpose, which establishes if a point is in or on the edge of a polygon. Thus, for all watermarked vertices, this function is applied with reference to the original polygon. If the watermarked vertex is within the original polygon, a gap is created, while if the watermarked vertex is outside the original polygon, an overlap is created.

Algorithm 1: Overlap_Gap_Calculation

Input : The original and watermarked maps: M_o, M_w

Output: *Gaps*, *Overlaps*

$sum_1 = 0$

$sum_2 = 0$

$sum_3 = 0$

for *each watermarked polygon* P_w *in the watermarked map* M_w **do**

  $[in, on] = inpolygon(x_{P_w}, y_{P_w}, x_{P_o}, y_{P_o})$

  // x_{P_w} and y_{P_w} are vectors holding the x and y coordinates values of the watermarked polygon P_w; x_{P_o} and y_{P_o} are vectors holding the x and y coordinates values of the corresponding original polygon P_o

  // in indicates if the points are inside or on the edge of the polygon; on indicates if the points are on the edge of the polygon

  $sum_1 = sum_1 + numel(x_{P_w}[in])$

  // the number of points inside or on the edge of the polygon

  $sum_2 = sum_2 + numel(x_{P_w}[on])$

  // the number of points on the edge of the polygon

  $sum_3 = sum_3 + numel(x_{P_w}[\sim in])$

  // the number of points outside the edge of the polygon

end

$Gaps = sum_1 - sum_2$

// the number of points inside the original polygons for the whole map

$Overlaps = sum_3$

// the number of points outside the original polygons for the whole map

return *Gaps*, *Overlaps*

The quantified measure for the overlap issue in the watermarked map is defined in Equation (5) as the proportion of overlapping polygons from all watermarked polygons:

$$M_2 = \frac{\sum_{i=1}^{V_w} V_{oi}}{V_w} \tag{5}$$

where M_2 represents the overlap metric, V_w represents the number of watermarked vertices and V_o represents the number of vertices placed outside their original polygon after watermarking, thus leading to overlaps.

The quantified measure for the gap issue in the watermarked map is defined in Equation (6) as the proportion of gaps between polygons from all watermarked polygons:

$$M_3 = \frac{\sum_{i=1}^{V_w} V_{g_i}}{V_w} \tag{6}$$

where M_3 represents the gap metric, V_w represents the number of watermarked vertices and V_g represents the number of vertices placed within their original polygon after watermarking, thus leading to gaps.

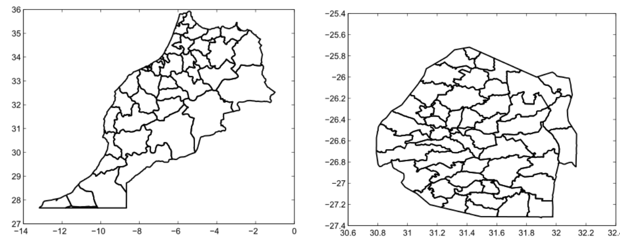
3.3. The Overall Metric

The overall metric is defined as the average of disclosure, overlap and gap measurements that were described in the previous subsections – see Equation (7).

$$M = \frac{\sum_{i=1}^{3} M_i}{3} \tag{7}$$

where M represents the overall fidelity metric, M_1 represents the disclosure metric, M_2 represents the overlap metric and M_3 represents the gap metric.

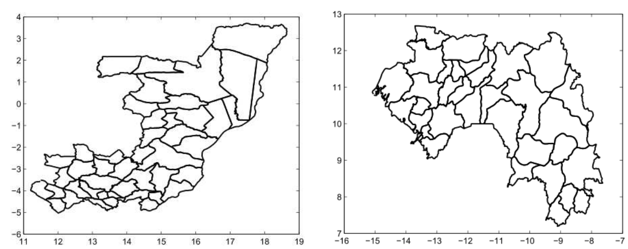
For all metrics, the values are between 0 and 1, where a value of 0 indicates no topology problems, and 1 indicates the maximum number of topology problems. For example, for the overall metric a value on 1 means that all watermarked polygons are disclosed and that overlaps and gaps take place for all watermarked vertices.



(a) Map of Morocco (47 polygons, 7523 vertices)

(b) Map of Swaziland (53 polygons, 7678 vertices)

Figure 1: Dataset 1.



(a) Map of Congo-Brazzaville (46 polygons, 12511 vertices)

(b) Map of Guinea (56 polygons, 21304 vertices)

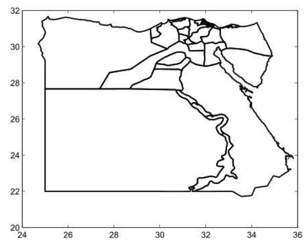
Figure 2: Dataset 2.

4. Experiments

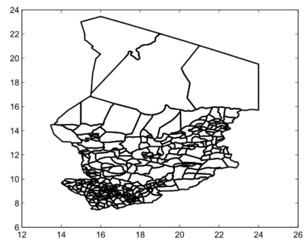
This section describes the experiments that are conducted for the evaluation of the proposed metrics, including the data used and the way of controlling the embedding of the watermark to assess the comparability of the results across maps and watermarks of different sizes.

4.1. Data Description and Experimental Setup

To evaluate if the metrics allow comparisons for maps of different sizes in terms of number of polygons and number of vertices, four datasets (of two maps each) combining high and low numbers of polygons and vertices were used, respectively:

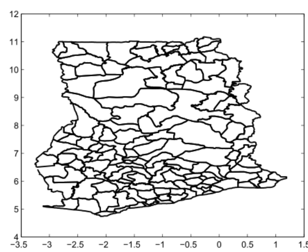


(a) Map of Egypt (129 polygons, 5992 vertices)

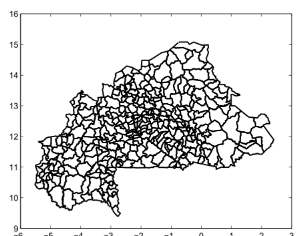


(b) Map of Chad (347 polygons, 19542 vertices)

Figure 3: Dataset 3.



(a) Map of the Ghana (138 polygons, 243329 vertices)



(b) Map of Burkina Faso (351 polygons, 113996 vertices)

Figure 4: Dataset 4.

- Dataset 1 includes maps with small number of polygons and small number of vertices.
- Dataset 2 includes maps with small number of polygons and large number of vertices.
- Dataset 3 includes maps with large number of polygons and small number of vertices.
- Dataset 4 includes maps with large number of polygons and large number of vertices.

Within each dataset, the two maps are chosen to represent opposite ratios of number of polygons to number of vertices, i.e. one map has on average a smaller number of vertices per polygon compared with the other map in the same dataset.

Also, the size of the watermark is controlled, i.e. 25%, 33% and 50% of the original map, to show that the metrics can be used to compare watermarked maps not only of variable map size, but also variable watermark size.

Table 1 lists the maps of the four datasets, their number of polygons and vertices, the average number of vertices per polygon, as well as the number of polygons that correspond to the proportions of 25%, 33% and 50%, which are used when embedding the watermark. Figures 1 to 4 illustrates the eight maps of the four datasets.

Table 1: The datasets (D) with corresponding number of polygons (#P), vertices (#V) and number of polygons for proportions of map size.

D	Map	#P	#V	Avg	Proportions		
					25%	33%	50%
1	Morocco (MOR)	47	7523	160	12	16	24
	Swaziland (SWA)	53	7678	144	14	18	27
2	Congo-Brazzaville (CNG)	46	12511	271	12	16	23
	Guinea (GIN)	56	21304	380	14	19	28
3	Egypt (EGY)	129	5992	46	33	43	65
	Chad (CHA)	347	19542	56	87	116	174
4	Ghana (GHA)	138	243329	1763	35	46	69
	Burkina Faso (BUF)	351	113996	324	88	117	176

The proposed metrics are defined in relation to the watermark size to allow comparison across maps and watermarks of different sizes. This relativity to the watermark size should results in our experiments in similar metrics values for all the maps within the same dataset, as well as across all datasets. In other words, the experiments were set up to show that regardless of map size, comparisons on the distortions introduced by watermarking still can be made.

The maps used in our experiments are freely available, in ESRI shapefile format, from the map maker website[5]. Maps that are freely available Were used to facilitate the development of benchmarks in the context of vector data, as one of the important aspects of bringing research in this area forward, by making it possible to compare different developments.

ESRI Shapefiles (.shp) are produced by ESRI[6], and considered as a popular format for geographic information system applications [1]. They have several key features: small storage space, easy reading and writing, fast shape editing, storing both spatial and attribute information, and supporting point, polyline and polygon geometry types [38].

The two most-known watermark embedding approaches were implemented in MATLAB version R2014b (8.4.0.150421) on a 64-bits Windows-PC. The way watermarks of different sizes were embedded, is explained in the following section.

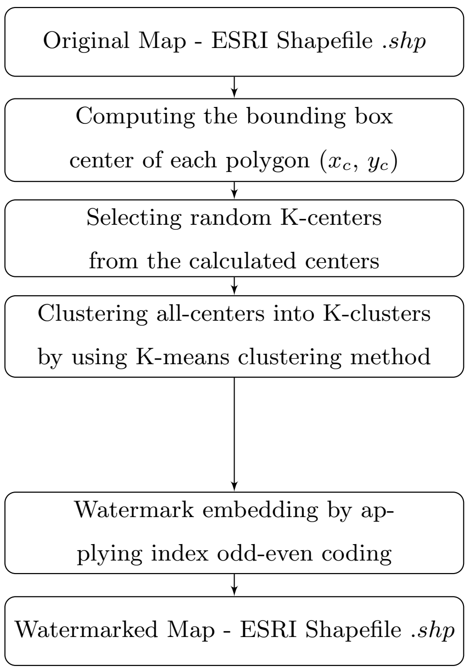
4.2. Watermark Insertion Process

For the watermark embedding process, two main prevalent approaches were used and compared: (1) a coordinate-based approach (shown in Fig.5a) and (2) a distance based approach (shown in Fig.5b). These approaches have shown, practically, a better resilience to map changes/attacks such as: rotation, translation, scaling, simplification and interpolation [39, 40]. In both approaches, clustering is used to control the size of the watermark in relation to map size, as well as distribute the watermark throughout the map. Clustering is used to identify locations in the map for embedding the watermark [30].

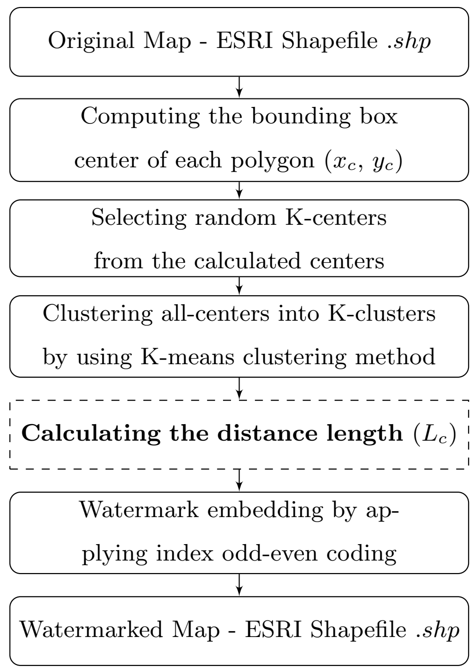
Both approaches mentioned above uses the bounding box property in ESRI shapefiles, which identifies the boundaries of each polygon in the map [38]. Polygons' bounding box centers are calculated in both axes, as shown in Equation 8:

[5] `http://www.mapmakerdata.co.uk`

[6] `http://www.esri.com/`



(a) The coordinates based approach



(b) The distance based approach

Figure 5: Two different watermark insertion approaches

$$x_c = \frac{x_{min} + x_{max}}{2} \quad \& \quad y_c = \frac{y_{min} + y_{max}}{2} \tag{8}$$

where x_c and y_c are the coordinates of a polygon's center in x and y axes respectively; x_{min} is the minimum vertex coordinate in the x-axis; x_{max} is the maximum vertex coordinate in the x-axis; y_{min} is the minimum vertex coordinate in the y-axis; y_{max} is the maximum vertex coordinate in the y-axis; x_{min}, x_{max}, y_{min} and y_{max} are each of 8-byte length [38].

The k-means clustering method is used to cluster the bounding box centers, as the polygons' representatives, in order to determine the positions for embedding the watermark. More precisely, through this process, a number of polygons are identified as locations for embedding the watermark. The k-means method is relatively simple, easy to implement, and needs a predefined number of clusters (k) – see reference [39] for more detail. The experiments were set up with values of k that represent approximately 25%, 33% and 50% of the total number of polygons. In this way, the size of the watermark is controlled, which allows evaluating the proposed metrics for different watermark sizes.

The watermark is constructed by adding or subtracting a bit value of 1 from either x and y vertex coordinate values (coordinate-based approach) or distance length values (distance-based approach) within the selected polygons (identified by k-means clustering).

The watermark is embedded by applying odd-even indexing, which is one of the most popular embedding approaches [41], [22], [40], [39], [30]. This approach is formally represented as in Equation (9).

$$W_i = \begin{cases} T - 1, & \text{if OES(I)=odd} \\ T + 1, & \text{if OES(I)=even} \end{cases} \tag{9}$$

where W_i is the ith bit value of the watermark; OES stands for Odd-Even Status; I is the order index of the watermark embedding position value; T is the value of the 4th digit of the embedding position value, after the decimal point. The following two subsections detail the embedding procedure for the coordinate-based and distance-based approaches.

4.2.1. Coordinates-based Embedding

In this approach, the embedding space is the x and y vertex coordinate values. The watermark is embedded by comparing the OES (Odd-Even Status) of I which represents the sequential order of the vertex within the set of polygon's vertices. As shown in Equation (10), the conditions are set based on two scenarios: (a) if the OES of I is odd, 1 will be subtracted from the value of T, which represents the 4th bit after the decimal point of the x and y vertex coordinate values; (b) if the OES of I is even, 1 will be added to the value of T.

$$v_x^* = v_x \pm 0.0001 \quad \& \quad v_y^* = v_y \pm 0.0001 \tag{10}$$

where v_x^* and v_y^* are the new vertices' coordinates after embedding the watermark according to the aforementioned condition, in Equation (9); v_x and v_y are the original vertices' coordinates before inserting the watermark bits.

4.2.2. Distance-based Embedding

In this approach, the embedding space is the mean distance length values. The distance length is calculated by measuring the distance from the polygon bounding box top right corner to its center, as illustrated in Equation (11).

$$L_c = \sqrt{(x_c - x_{max})^2 + (y_c - y_{max})^2} \tag{11}$$

where L_c is the distance length; x_c and y_c are the center coordinates in x and y axes, respectively; x_{max} and y_{max} are the top right bounding box corner coordinates in the x and y axes, respectively.

As shown in Equation (9), the watermark is embedded by comparing the OES (Odd-Even Status) of the I variable, which represents the order index of the mean-distance length values. Similarly to the coordinate-based approach, the conditions are set based on two scenarios: (a) if the OES of I is odd, 1 will be subtracted from the value of T; (b) if the OES of I is even, 1 will be added to the value of T.

After applying the OES to change the values of L_c, the new values of distance length will be represented by L_c^*. The change rate α_c is calculated as depicted in Equation (12):

$$\alpha_c = \frac{L_c^*}{L_c} \tag{12}$$

The change rate α_c is used to change all vertices of polygons that belong to each cluster's center on the basis of the embedding condition, as given in Equation (13).

$$v_x^* = \alpha_c v_x + x_c(1 - \alpha_c) \quad \& \quad v_y^* = \alpha_c v_y + y_c(1 - \alpha_c) \tag{13}$$

Both embedding approaches should lead to contrasted readings in overlaps and gaps as the size of the watermark increases; the same should occur for disclosures for the coordinate-based approach (the distance-based approach does not lead to disclosures). In other words, the more watermark bits are included, the more issues with topology will occur. As a metric should allow comparison across different map sizes, as well as watermark size (and not simply penalise bigger watermarks), the metrics are defined as the number of topological issues (disclosures/gaps/overlaps) relative to the watermark size. Consequently, similar metrics were expected across the maps of different size and across the different sizes of watermarks, with some expected variety due to the randomness involved in the selected polygons for embedding (with varying numbers of vertices) and the odd-even status of the embedding locations; these random variations are further discussed in the next section.

Consequently, to show the reliability of the overall metric, the experimental results should show the following:

1. The disclosure metric for the coordinate-based approach will depend on the number of vertices in the watermarked polygons, thus leading to variations unrelated to the map size or watermark size; if all watermarked polygons have an even number of vertices, there will be no disclosures, while if all watermarked polygons have an odd number of vertices, all will have disclosures. The probability for a watermarked polygon to have either an odd or an even number of polygons is 0.5; thus, for higher numbers of watermarked polygons, the $M1$ metric would be expected to have values around 0.5, while

for fewer watermarked polygons, a higher variety would be expected in the metrics' values.

2. The gaps and overlaps metrics for both embedding approaches should have very similar values; since all watermarked vertices will lead to either a gap or an overlap, two phenomena are expected: (a) approximately half of the vertices will lead to gaps and half to overlaps, which would results in values of approximately 0.5 for metrics $M1$ and $M2$; (b) when the previous does not happen due to randomness, there will be a complementarity between the number of gaps and overlap, i.e. the more gaps, the fewer overlaps;

3. The overall metric for the coordinate-based approach will follow the variation in the disclosure metric, as it is an average of the disclosure, overlaps and gaps metrics, and the overlaps and gaps metrics should display little variation;

4. The overall metric for the distance-based approach should be very similar for all maps and all watermark sizes, as there are no disclosures for this embedding approach, and the overlaps and gaps metrics should be complementary (i.e. the more gaps, the fewer overlaps).

The next section presents the results and discusses them in terms of our expectations outlined above.

5. Results and Discussion

This section presents the results of our experiment in relation to the three metrics corresponding to the three topology rules for polygons, as well as the overall metric. The results are discussed in relation to the experimental setup and the expectations outlined in the previous section.

The disclosure metrics for all datasets are given in Table 2 and Fig. 6; this is just for the coordinate-based approach, as for the distance-based approach there are no disclosures due to the embedding process.

As expected, the results show an increase in disclosures proportionate to the watermark size, i.e. the larger the watermarks, the higher the number of

Table 2: The disclosure metric for the coordinate-based embedding method; Notes: n_w = number of watermarked polygons; D = number of disclosures; M_1 = disclosure metric.

Dataset	Map	n_w	Coordinate	
			D	M_1
1	MOR (25%)	12	3	0.25000
	MOR (33%)	16	3	0.18750
	MOR (50%)	24	9	0.37500
	SWA (25%)	14	8	0.57143
	SWA (33%)	18	9	0.50000
	SWA (50%)	27	15	0.55556
2	CNG (25%)	12	4	0.33333
	CNG (33%)	16	6	0.37500
	CNG (50%)	23	11	0.47826
	GIN (25%)	14	9	0.64286
	GIN (33%)	19	11	0.57895
	GIN (50%)	28	17	0.60714
3	EGY (25%)	33	14	0.42424
	EGY (33%)	3	22	0.51163
	EGY (50%)	65	29	0.44615
	CHA (25%)	87	50	0.57471
	CHA (33%)	116	69	0.59483
	CHA (50%)	174	92	0.52874
4	GHA (25%)	35	18	0.51429
	GHA (33%)	46	26	0.56522
	GHA (50%)	69	38	0.55072
	BUF (25%)	88	44	0.50000
	BUF (33%)	117	61	0.52137
	BUF (50%)	176	86	0.48864

Table 3: The overlap metrics for coordinate-based and distance-based embedding methods; Notes: V_w = number of watermarked vertices; O = number of overlaps; M_2 = overlap metric

Dataset	Map	V_w	Coordinate		Distance	
			O	M_2	O	M_2
1	MOR (25%)	2105	1067	0.50689	1094	0.51971
	MOR (33%)	2729	1382	0.50641	1386	0.50788
	MOR (50%)	4275	2165	0.50643	2225	0.52047
	SWA (25%)	1808	922	0.50996	1093	0.60454
	SWA (33%)	2793	1419	0.50806	1559	0.55818
	SWA (50%)	4174	2119	0.50767	2424	0.58074
2	CNG (25%)	3510	1770	0.50427	1860	0.52991
	CNG (33%)	4194	2115	0.50429	1682	0.40105
	CNG (50%)	6036	3043	0.50414	2720	0.45063
	GIN (25%)	6277	3138	0.49992	3115	0.49626
	GIN (33%)	9046	4526	0.50033	4397	0.48607
	GIN (50%)	13887	6947	0.50025	6930	0.49903
3	EGY (25%)	4055	2065	0.50925	2126	0.49824
	EGY (33%)	2855	1478	0.51769	1612	0.56462
	EGY (50%)	4504	2328	0.51687	2467	0.54774
	CHA (25%)	4887	2538	0.51934	2486	0.50870
	CHA (33%)	6933	3595	0.51853	3782	0.54551
	CHA (50%)	10004	5187	0.51849	5082	0.50800
4	GHA (25%)	59299	29417	0.49608	30301	0.51099
	GHA (33%)	94058	46648	0.49595	49442	0.52565
	GHA (50%)	133860	66401	0.49606	70292	0.52513
	BUF (25%)	26270	13206	0.50270	13886	0.52859
	BUF (33%)	36404	18304	0.50280	18677	0.51305
	BUF (50%)	54854	27593	0.50303	29217	0.53263

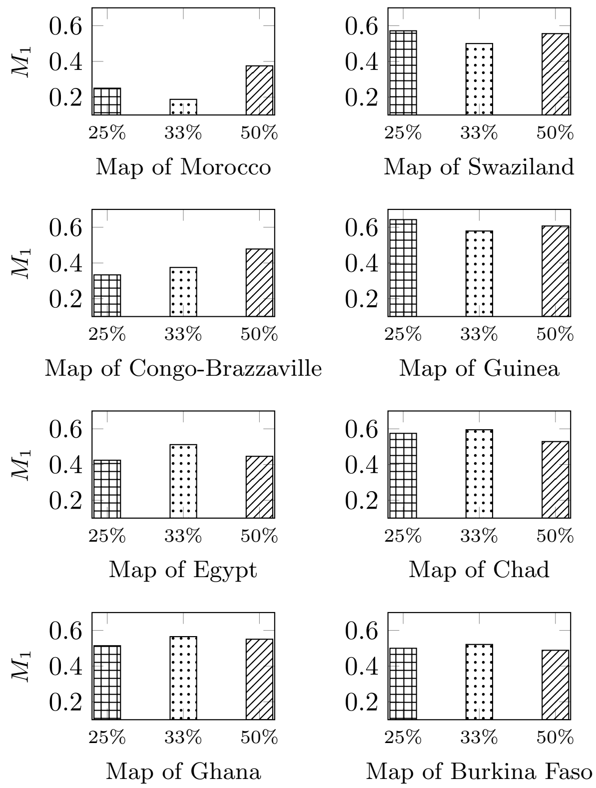


Figure 6: Coordinate-based method disclosure metrics (M_1).

disclosures – see the 4th column (D) in Table 2. The M_1 metric does not entirely preserve this proportions (see Fig. 6) due to the randomness involved in the odd-even status of the number of vertices in a polygon, i.e. if the watermark is added to a polygon with an odd number of vertices, there will be no disclosure, while if the watermark is added to a polygon with an even number of vertices, there will be a disclosure.

When looking at the variations of the M_1 metric for the same map with different watermark sizes, it is noticeable that these are relatively small with most differences smaller than 0.09. The biggest variations take place for the MOR (0.19) and CNG (0.15) maps, which is not surprising since these are the maps with the smallest number of polygons (at it is known that the randomness effect stabilizes for larger numbers). Unsurprisingly, the smallest variation occurs for

BUF (0.03), which is the map with the highest number of polygons.

The experimental results for the overlap metric (M_2) are displayed in Table 3, Fig. 7 and Fig. 8, for both watermarking approaches.

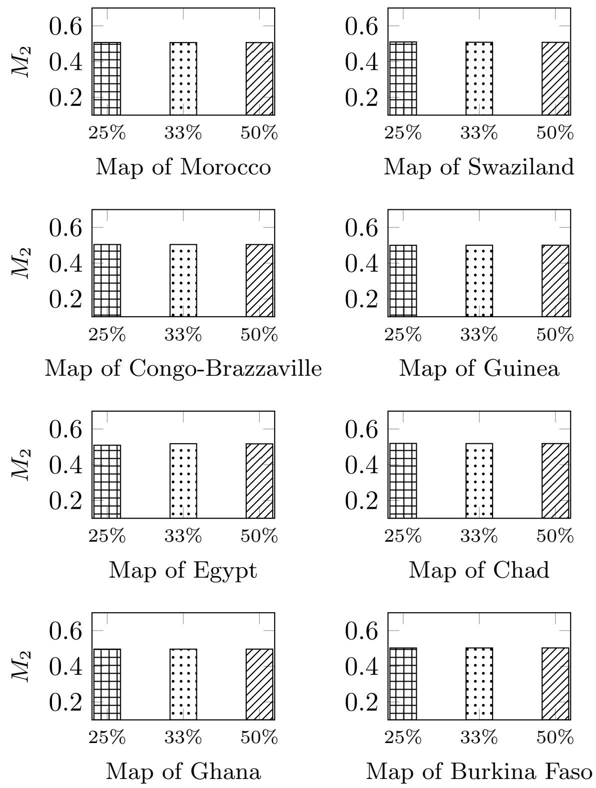


Figure 7: Coordinate-based approach overlap metric (M_2).

As expected, the higher the number of watermarked vertices, the higher the number of overlaps (columns 4 and 6 in Table 3). The only exception to this is for the Map of Egypt, where the 33% watermark results in fewer watermarked vertices than the 25% watermark. This is due to our embedding procedure in which a number of polygons is selected in which the watermark is inserted, thus, the number of watermarked vertices overall depends on the number of vertices in each polygon selected for embedding. In the case of the Map of Egypt–33%, the polygons selected for the embedding of the watermark had fewer vertices overall than the polygons selected for the Map of Egypt–25%.

As expected, for both embedding approaches, overlaps metrics are very similar regardless of map size and watermark size. For the same maps with different watermark sizes, for the coordinate-based approach, the average difference is 0.00109 with a standard deviation of 0.00221. For the distance-based approach, the average is 0.03041 and the standard deviation is 0.03166.

Overall, the overlap metric for all maps ranges between 0.49595 and 0.51934 for the coordinate-based approach and between 0.40105 and 0.60454 for the distance-based approach. Thus, it is noticeable that the coordinate-based approach leads to more similar values than the distance-based approach.

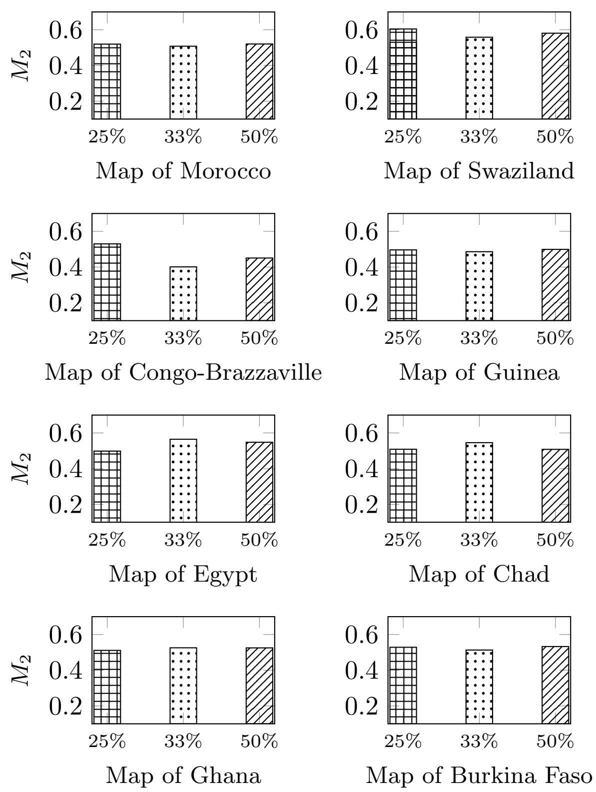


Figure 8: Distance-based approach overlap metric (M_2).

Table 4, Fig. 9 and Fig. 10 displays the gap metrics for both coordinate-based and distance-based approaches. As expected, the more vertices are watermarked, the more gaps occur, with the exception for the Map of Egypt men-

Table 4: The gap metrics for coordinate-based and distance-based embedding methods: Notes: V_w = number of watermarked vertices; G = number of gaps; M_3 = gaps metric.

Dataset	Map	V_w	Coordinate [2]		Distance [22]	
			G	M_3	G	M_3
1	MOR (25%)	2105	1038	0.49311	1011	0.48029
	MOR (33%)	2729	1347	0.49359	1343	0.49212
	MOR (50%)	4275	2110	0.49357	2050	0.47953
	SWA (25%)	1808	886	0.49004	715	0.39546
	SWA (33%)	2793	1374	0.49194	1234	0.44182
	SWA (50%)	4174	2055	0.49233	1750	0.41926
2	CNG (25%)	3510	1740	0.49573	1650	0.47009
	CNG (33%)	4194	2079	0.49571	2512	0.59895
	CNG (50%)	6036	2993	0.49586	3316	0.54937
	GIN (25%)	6277	3139	0.50008	3162	0.50374
	GIN (33%)	9046	4520	0.49967	4649	0.51393
	GIN (50%)	13887	6940	0.49975	6957	0.50097
3	EGY (25%)	4055	1990	0.49075	2141	0.50176
	EGY (33%)	2855	1377	0.48231	1243	0.43538
	EGY (50%)	4504	2176	0.48313	2037	0.45226
	CHA (25%)	4887	2349	0.48066	2401	0.49130
	CHA (33%)	6933	3338	0.48147	3151	0.45449
	CHA (50%)	10004	4817	0.48151	4922	0.49200
4	GHA (25%)	59299	29882	0.50392	28998	0.48901
	GHA (33%)	94058	47410	0.50405	44616	0.47435
	GHA (50%)	133860	67456	0.50394	63565	0.47487
	BUF (25%)	26270	13064	0.49730	12384	0.47141
	BUF (33%)	36404	18100	0.49720	17727	0.48695
	BUF (50%)	54854	27261	0.49697	25637	0.46737

tioned previously for overlaps - since the gap metric, like the overlap one, is influenced by the total number of vertices in the watermarked polygons, the same effect occurs.

For the same maps with different watermark sizes, for the coordinate-based approach the average difference is 0.00120 and the standard deviation is 0.00235. For the distance-based approach, the average is 0.03108 and the standard deviation is 0.03125.

Overall, the gap metrics range between 0.48147 and 0.50405 for the coordinate-base approach and between 0.39546 and 0.59895 for the distance-based approach. Similar the overlaps metric, it is noticeable that a smaller range occurs for the coordinate-based approach compared with the distance-based approach.

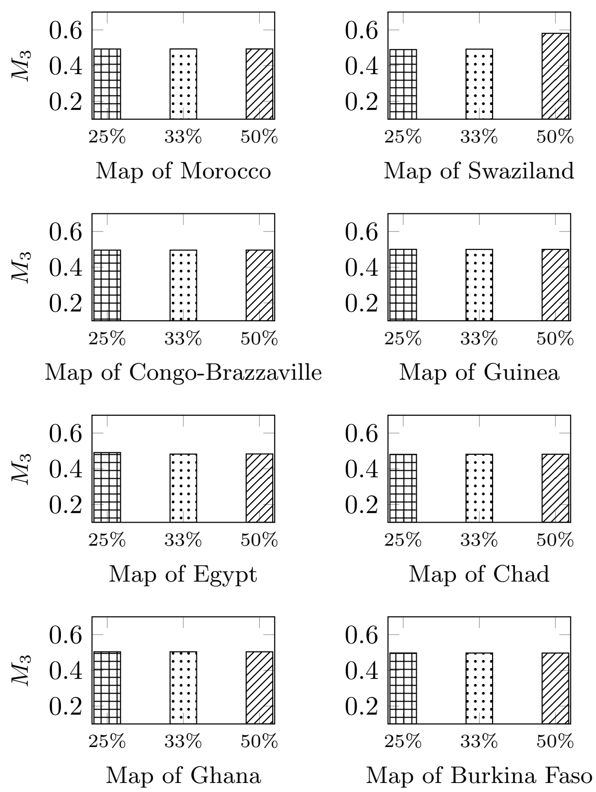


Figure 9: Coordinate-based approach gap metric (M_3).

For the overall metrics, the results are displayed in Table 5, Fig. 11 and

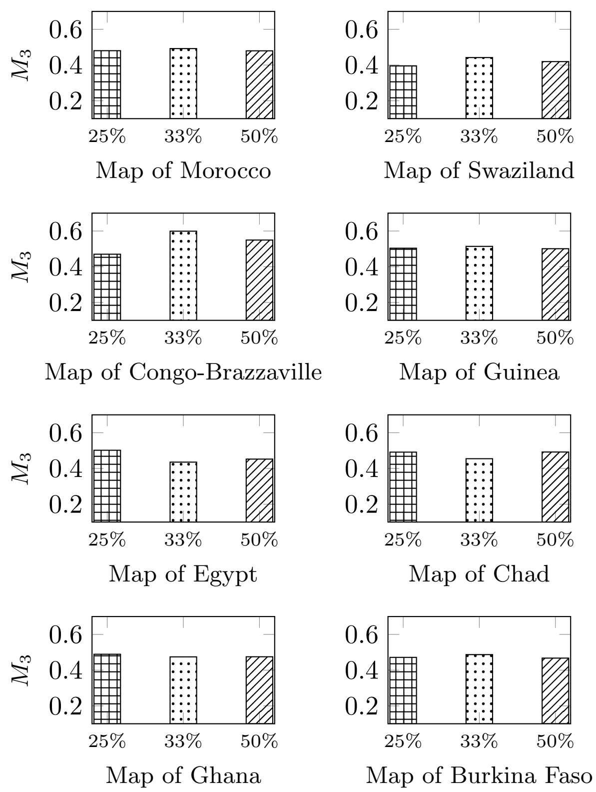


Figure 10: Distance-based approach gap metric (M_2).

Fig. 12. For the coordinate-based approach, the overall metric values are between 0.39583 and 0.54762, while for the distance-based approach the metrics are 0.33333 for all maps and all watermark sizes. For the distance-based approach, the same values are occurring due to the lack of disclosures (thus, the lower value) and the complementarity between gaps and overlaps (i.e. a watermarked vertex will lead to either a gap or an overlap), i.e. when more gaps occur, there are fewer overlaps (as reflected in the $M2$ and $M3$ metrics).

For example, the SWA (25%) map has a large number of overlaps reflected in a high $M2$ metric, i.e. 0.60454, and a lower number of gaps reflected in a low $M3$ metric, i.e. 0.39546 (the two metrics add up to 1); the $M2$ and $M3$ metrics add up to 1 for all maps. As there are no disclosures, and each metric has the same weight, the overall metric becomes 1/3, i.e 0.33333.

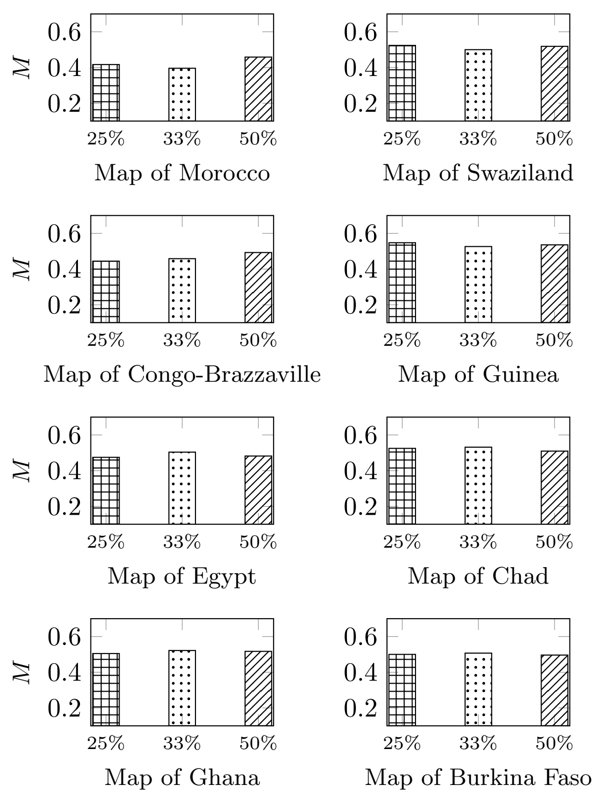


Figure 11: Coordinate-based overall metric (M).

Table 5: The overall metric (M) for coordinate-based and distance-based embedding methods.

Dataset	Map	Coordinate M	Distance M
1	MOR (25%)	0.41667	0.33333
	MOR (33%)	0.39583	0.33333
	MOR (50%)	0.45833	0.33333
	SWA (25%)	0.52381	0.33333
	SWA (33%)	0.50000	0.33333
	SWA (50%)	0.51852	0.33333
2	CNG (25%)	0.44444	0.33333
	CNG (33%)	0.45833	0.33333
	CNG (50%)	0.49275	0.33333
	GIN (25%)	0.54762	0.33333
	GIN (33%)	0.52632	0.33333
	GIN (50%)	0.53571	0.33333
3	EGY (25%)	0.47475	0.33333
	EGY (33%)	0.50388	0.33333
	EGY (50%)	0.48205	0.33333
	CHA (25%)	0.52490	0.33333
	CHA (33%)	0.53161	0.33333
	CHA (50%)	0.50958	0.33333
4	GHA (25%)	0.50476	0.33333
	GHA (33%)	0.52174	0.33333
	GHA (50%)	0.51691	0.33333
	BUF (25%)	0.50000	0.33333
	BUF (33%)	0.50712	0.33333
	BUF (50%)	0.49621	0.33333

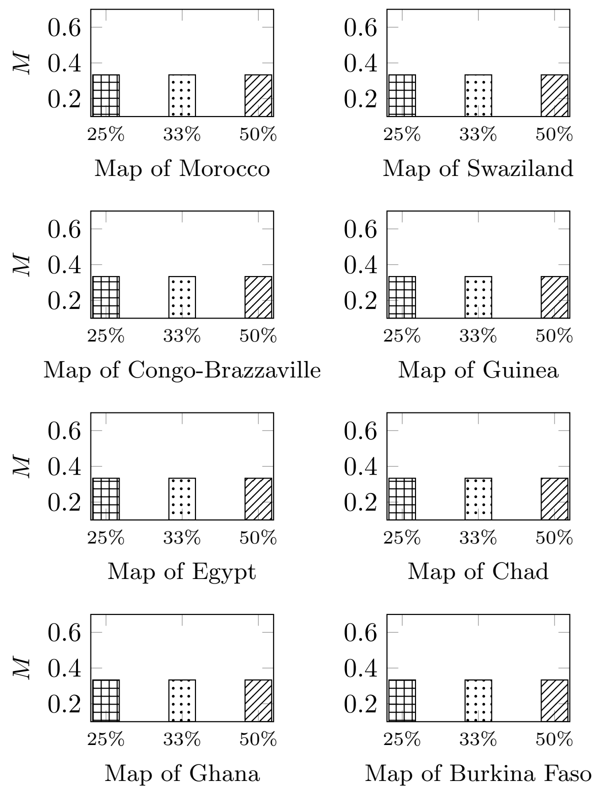


Figure 12: Distance-based overall metric (M).

The experiments were set up with the purpose of showing that the metrics allow comparisons between maps of different sizes, as well as different watermark sizes. More specifically, this work looked at a variety of maps grouped into four datasets covering the different combination of number of polygons and number of vertices. Moreover, within the same dataset, maps that had opposite ratios of numbers of vertices per polygon were chosen. The results show that the metrics are comparable across this variation in map size properties, with a few exceptions explained by the randomness involved in the embedding process.

By looking at different watermark sizes, the metrics were tested in terms of their accurate reflection of the number of distortions. As the number of distortions are proportionate to the size of the watermark, an increase in the number of distortions were expected as the size of the watermark increased, which has been shown in the results. Because the metrics are defined as the number of distortions relative to the size of the watermark, it is expected that the metrics for the same map with the different watermark sizes would be very similar, with only small differences in values.

The results showed this consistency in the values of the metrics between the same map with watermarks of different size. The results were more consistent for the overlap and gap metrics than for the disclosure metric for the coordinate-based approach. The higher variability in the disclosure metric could be explained as a consequence of the odd-even indexing used in the embedding process. Another aspect related to the higher variability in the disclosure metric is the fact that the disclosure metric is defined in relation to the number of watermarked polygons, while the overlap and gap metrics are defined in relation to the number of vertices. As the number of polygons has a smaller range than the number of vertices, the metrics show more variation for the disclosure metric.

6. Conclusions and Future Work

In this paper, the importance of a metric to assess topological distortions in watermarked vector maps is discussed, and a metric for polygon-based vector maps is proposed. This paper looked at three distortions that can occur when polygon topology rules are broken in the watermarking process: polygon disclosures, overlaps and gaps.

Maps and watermarks of different sizes were used, as well as two different watermarking approaches to test the metrics; thus, four datasets were used, where each dataset had varying degrees of size in terms of number of polygons and number of vertices. Each dataset contained two maps, which had opposite ratios of number of vertices per polygon. By using k-means clustering to embed the watermark, the size of the watermark is controlled and experimented with three sizes corresponding approximately to 25% (16–117 polygons), 33% (12–88 polygons) and 50% (24–176 polygons) of the number of polygons in the original maps. The results indicate that the metrics allow comparisons between watermarked maps of different sizes and of different watermark sizes, and, thus, can be used to asses the quality of watermarked vector maps.

The proposed metric described and tested in this paper is a first step towards a standard metric for watermarked vector map quality that assesses topological distortion. Further research and experiments will be carried out on addressing the problem of the randomness in the map polygon indexes associated with odd-even coding to further understand the behavior of the metric in extreme cases. Also, the possibility of introducing different weights for the different topological aspects will be investigated.

References

References

[1] P. Burrough, R. McDonnell, C. Lloyd, Principles of Geographical Information Systems, Oxford University Press, 2013.

[2] X. Wang, C. Shao, X. Xu, X. Niu, Reversible data-hiding scheme for 2-d vector maps based on difference expansion, IEEE Transactions on Information Forensics and Security 2 (3) (2007) 311–320.

[3] N. Ren, Q. Wang, C. Zhu, GIS Spatial Data Updating Algorithm Based on Digital Watermarking Technology, Vol. 482 of Communications in Computer and Information Science, Springer Berlin Heidelberg, 2015, Ch. Geo-Informatics in Resource Management and Sustainable Ecosystem, pp. 143 –150.

[4] A. Okabe, GIS-based Studies in the Humanities and Social Sciences, CRC Press, 2016.

[5] C. Wang, W. Wang, B. Wu, Q. Qin, A watermarking algorithm for vector data based on spatial domain, in: 1st International Conference on Information Science and Engineering, 2009, pp. 1959–1962.

[6] S.-H. Lee, K.-R. Kwon, Vector watermarking scheme for gis vector map management, Journal of Multimedia Tools and Applications 63 (3) (2013) 757–790.

[7] N. Ren, Q. sheng Wang, C. qing Zhu, Selective authentication algorithm based on semi-fragile watermarking for vector geographical data, in: Geoinformatics, IEEE 22nd International Conference on, 2014, pp. 1–6.

[8] Y. Peng, M. Yue, A zero-watermarking scheme for vector map based on feature vertex distance ratio, Journal of Electrical and Computer Engineering 2015 (2015) 35:1–35:6.

[9] T. Zhang, H. Zhang, H. Hou, A. Abd El-Latif, A robust watermarking for 2D CAD graphic based on coordinates modification, in: The International Conference of Electronic Engineering and Information Science, 2015, pp. 435 – 438.

[10] N. Wang, X. Zhao, C. Xie, Rst invariant reversible watermarking for 2d vector map, International Journal of Multimedia and Ubiquitous Engineering 11 (2016) 265–276.

[11] B.-J. Jang, S.-H. Lee, E.-J. Lee, S. Lim, K.-R. Kwon, A crypto-marking method for secure vector map, Multimedia Tools and Applications (2016) 1–34.

[12] A. Abubahia, M. Cocea, Advancements in gis map copyright protection schemes - a critical review, Multimedia Tools and Applications 76 (10) (2017) 12205–12231.

[13] X.-J. Huo, S.-H. Lee, S.-G. Kwon, K.-S. Moon, K.-R. Kwon, A watermarking scheme for shapefile-based gis digital map using polyline perimeter distribution, Journal of Korea Multimeadia Society 14 (5) (2011) 595 – 606.

[14] J. Bong-Joo, L. Suk-Hwan, L. Sanghun, K. Ki-Ryong, Progressive vector compression for high-accuracy vector map data, International journal of Geographic Information Science 28 (4) (2014) 763–779.

[15] V. Doncel, N. Nikolaidis, I. Pitas, An optimal detector structure for the fourier descriptors domain watermarking of 2d vector graphics, IEEE Transactions on Visualization and Computer Graphics 13 (5) (2007) 851–863.

[16] P. Bas, T. Furon, A new measure of watermarking security: The effective key length, IEEE Transactions on Information Forensics and Security 8 (8) (2013) 1306–1317.

[17] N. Wang, H. Zhang, C. Men, A high capacity reversible data hiding method for 2d vector maps based on virtual coordinates, Computer-Aided Design 47 (2014) 108 – 117.

[18] S. Jian-Guo, Z. Guo-Yin, Y. Ai-Hong, W. Jun-Peng, A reversible digital watermarking algorithm for vector maps, International Journal of Network Security 16 (1) (2014) 40 – 45.

[19] M. Asikuzzaman, M. Alam, A. Lambert, M. Pickering, Imperceptible and robust blind video watermarking using chrominance embedding: A set of approaches in the dt cwt domain, IEEE Transactions on Information Forensics and Security 9 (9) (2014) 502–517.

[20] N. Wang, Reversible fragile watermarking for locating tampered polylines/polygons in 2d vector maps, International Journal of Digital Crime and Forensics 8 (1) (2016) 1–25.

[21] D. Eppstein, M. T. Goodrich, J. Lam, N. Mamano, M. Mitzenmacher, M. Torres, Models and algorithms for graph watermarking, in: M. Bishop, A. C. A. Nascimento (Eds.), Information Security, Springer International Publishing, Cham, 2016, pp. 283–301.

[22] X.-J. Huo, K.-S. Moon, S.-H. Lee, T.-Y. Seung, S.-G. Kwon, Protecting gis vector map using the k-means clustering algorithm and odd-even coding, in: The 17th Korea-Japan Joint Workshop on Frontiers of Computer Vision, 2011, pp. 1–5.

[23] N. Wang, X. Zhao, H. Zhang, Block-based reversible fragile watermarking for 2d vector map authentication, International Journal of Digital Crime and Forensics 7 (3) (2015) 60–80.

[24] M. Urvoy, D. Goudia, F. Autrusseau, Perceptual dft watermarking with improved detection and robustness to geometrical distortions, IEEE Transactions on Information Forensics and Security 9 (7) (2014) 1108–1119.

[25] Z. Peng, M. Yue, X. Wu, Y. Peng, Blind watermarking scheme for polylines in vector geo-spatial data, Multimedia Tools and Applications 74 (24) (2015) 11721–11739.

[26] Q. Huynh-Thu, M. Ghanbari, Scope of validity of psnr in image/video quality assessment, Electronics Letters 44 (13) (2008) 800–801.

[27] X. Wang, K. Pang, X. Zhou, Y. Zhou, L. Li, J. Xue, A visual model-based perceptual image hash for content authentication, IEEE Transactions on Information Forensics and Security 10 (7) (2015) 1336–1349.

[28] X. Niu, C. Shao, X. Wang, A survey of digital vector map watermarking, International Journal of Innovative Computing, Information and Control 2 (6) (2006) 1301 – 1316.

[29] L. Huang, W. Zhou, R. Jiang, A. Li, Data quality inspection of watermarked GIS vector map, in: The 18th International Conference on Geoinformatics, 2010, pp. 1–5.

[30] A. Abubahia, M. Cocea, Partition clustering for gis map data protection, in: The 26th International Conference on Tools with Artificial Intelligence, 2014, pp. 830 – 837.

[31] J. Kim, Robust vector digital watermarking using angles and a random table, Advances in Information Sciences and Service Sciences 2 (4) (2010) 79–90.

[32] J. Kim, Vector map digital watermarking using angles, in: 6th International Conference on Networked Computing and Advanced Information Management, 2010, pp. 417–423.

[33] S. Zope-Chaudhari, P. Venkatachalam, Evaluation of spatial relations in watermarked geospatial data, in: The 3rd ACM SIGSPATIAL International Workshop on GeoStreaming, ACM, New York, NY, USA, 2012, pp. 78–83.

[34] S. Neyman, Y. Wijaya, B. Sitohang, A new scheme to hide the data integrity marker on vector maps using a feature-based fragile watermarking algorithm, in: Data and Software Engineering (ICODSE), 2014 International Conference on, 2014, pp. 1–6.

[35] D. J. Maguire, Arcgis: General-purpose gis software, in: Encyclopedia of GIS, Springer International Publishing, Redlands, USA, 2015, pp. 1–8.

[36] S. S. Maraş, H. H. Maraş, B. Aktuğ, E. E. Maraş, F. Yildiz, Topological error correction of gis vector data, International Journal of Physical Sciences 5 (5) (2010) 476–483.

[37] S. Zope-Chaudhari, P. Venkatachalam, K. M. Buddhiraju, Assessment of distortion in watermarked geospatial vector data using different wavelets, Geo-spatial Information Science 18 (2-3) (2015) 124–133. arXiv:http://dx.doi.org/10.1080/10095020.2015.1071064.

[38] ESRI, Esri shapefile technical description, Tech. rep., Environmental Systems Research Institute, 380 New York Street, Redlands, CA 92373-8100, USA (1998).

[39] A. Abubahia, M. Cocea, Exploiting vector map properties for gis data copyright protection, in: The 27th International Conference on Tools with Artificial Intelligence, 2015, pp. 575 – 581.

[40] A. Abubahia, M. Cocea, A clustering approach for protecting gis vector data, in: Advanced Information Systems Engineering, 27th International

Conference, CAiSE 2015, June 8-12, 2015, Proceedings, Springer International Publishing, Stockholm, Sweden, 2015, pp. 133–147.

[41] B. Wu, W. Wang, D. Miao, 2d vector map watermarking based on spatial relations, in: The International Conference on Earth Observation Data Processing and Analysis, 2008, pp. 532–537.