Exploiting Vector Map Properties for GIS Data Copyright Protection

Ahmed M. Abubahia
School of Computing
University of Portsmouth
Portsmouth, United Kingdom
Email: ahmed.abubahia@port.ac.uk

Mihaela Cocea
School of Computing
University of Portsmouth
Portsmouth, United Kingdom
Email: mihaela.cocea@port.ac.uk

***Abstract*—Geographic Information System (GIS) vector maps have become more widely available, prompting a need to prevent their unauthorized use. This is commonly done through the use of a digital watermark, with many approaches applying techniques from image map watermarking, without exploiting the particular properties of vector map data. In previous work we showed that using k-medoids clustering and the bounding box property of vector maps in the embedding process leads to increased robustness against simplification (removing vertices from vector data) and interpolation (adding new vertices to the data) attacks, which may distort the watermark and prevent the identification of the map owner. In this paper we show that the advantages of using the bounding box property are maintained even with a different clustering approach (k-means), and argue that they would hold regardless of the method used for identifying the watermark embedding locations in the map.**

***Keywords*-Geographic Information System; Vector Map Data; K-means Clustering; Data Protection; Security; Digital Copyright;**

I. INTRODUCTION

Advancements in Geographical Information Systems (GIS) technology, including capabilities of mapping, monitoring, modeling, management and measurement [1], [2], led to an increased employment of GIS maps in many applications such as government and public services [3], business and service planning [4], logistics and transportation [5], and environmental studies [6].

The production of a GIS map involves a time-consuming process of analysis and the use of well-trained specialists, accurate hardware and licensed software tools. Therefore, given the high cost of producing GIS maps, the producers of these maps are interested in preserving their copyright.

Moreover, these maps are liable to be illegally copied, modified or distributed due to their digital nature. Consequently, there is a compelling need for copyright protection to combat illegal use of GIS maps [7], [8].

GIS data can be represented in the form of two main models [9]: raster and vector data. The raster model (image) stores the geographic information into a form of grid cells, where each cell represents the natural corresponding value on the ground (e.g. color scale). The vector data model stores the geographic information into geometrical entities which have properties such as length, a starting point and an ending point [10]. GIS vector data is defined by a sequence of coordinates, and includes shapes such as points, polylines and polygons [11]. In this paper the focus is on the vector format of GIS data.

In response to the copyright protection issue, many digital watermarking methods have been proposed in literature, e.g. [12]–[14]. Nevertheless, GIS data received less attention than images, audio, texts and videos in the field of watermarking research, as pointed out in several recent review papers [11], [15], [16].

In addition, although the use of partition clustering methods for GIS map applications has seen an increase in recent years (e.g. [17]–[19]), little research in GIS map watermarking takes advantage of data mining methods in general, and clustering in particular. Two partitioning clustering methods have been used in this area: k-means [20] and k-medoids [12], [21]. In the context of GIS map watermarking, partition clustering methods use the distance between map vertices to divide the map vertices into a set of clusters with the purpose of identifying locations for embedding the watermark, and provides the advantage of ensuring the distribution of the watermark across the entire map.

In previous work [12] we showed that the use of a particular property of vector maps, called a bounding box, leads to an increased resistance of the watermarked map to malicious modification called attacks, and in particular, to interpolation (i.e. adding vertices to the map) [20] and simplification (i.e. deleting vertices from the map) [22] attacks. Moreover, we argued that the proposed approach maintains a good trade-off between capacity (the number of inserted watermark bits) and fidelity (the quality of the map after watermark insertion). This trade-off is very important for vector map data, as its value stems from its accurate locations properties; therefore, it is important that the watermark does not affect the precision of the locations (i.e. it has good fidelity), while at the same time it provides enough watermark bits to ensure the map's copyright protection (i.e. it has good capacity).

In this paper we argue that using the bounding box property of vector maps maintains the resistance to interpolation and simplification attacks even when a different clustering approach is used for identifying locations for embedding the

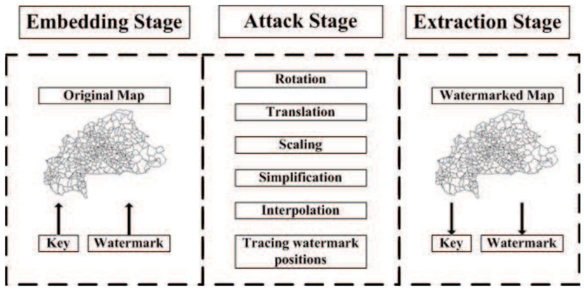


Figure 1. Digital GIS Map Watermarking System

Position	Field	Value	Type
Byte 0	File Code	9994	Integer
Byte 4	Unused	0	Integer
Byte 8	Unused	0	Integer
Byte 12	Unused	0	Integer
Byte 16	Unused	0	Integer
Byte 20	Unused	0	Integer
Byte 24	File Length	File Length	Integer
Byte 28	Version	1000	Integer
Byte 32	Shape Type	Shape Type	Integer
Byte 36	Bounding Box	Xmin	Double
Byte 44	Bounding Box	Ymin	Double
Byte 52	Bounding Box	Xmax	Double
Byte 60	Bounding Box	Ymax	Double
Byte 68*	Bounding Box	Zmin	Double
Byte 76*	Bounding Box	Zmax	Double
Byte 84*	Bounding Box	Mmin	Double
Byte 92*	Bounding Box	Mmax	Double

Figure 2. The Header of Polygon-based Shapefile, ESRI [29]

watermark. To assess this claim, we compare the original k-means approach of Huo et al. [20] with a modified k-means approach using the bounding box property.

The two approaches are also compared in terms of the trade-off between fidelity and capacity, to assess the influence of using the bounding box property on this trade-off. Moreover, we argue that the advantages of using the bounding box property would hold regardless of the methods used for identifying the watermark embedding locations in the map.

The rest of this paper is organized as described in the following. In Section II, GIS watermarking research terminology and requirements are briefly explained, while Section III gives a detailed overview of relevant previous work. Section IV presents the k-means approach with the use of the bounding box property and Section V presents the comparison results of the two clustering approaches, i.e. our modified k-means approach using the bounding box property and the k-means approach of Huo et al. [20]. Section VI concludes the paper and outlines directions for future work.

II. Research Background

A GIS map watermarking system includes two main stages: embedding and extraction (Fig. 1). The embedding stage aims to insert a watermark (e.g. digital binary sequence) into the GIS vector map points, by using a specific computing approach; the embedding space is often the Cartesian coordinates [22], [23]. The extraction stage refers to obtaining the watermark from the host GIS data in order to retrieve the original map. There are three key requirements for a reliable GIS watermarking system: fidelity, capacity and robustness [11], [15].

The fidelity requirement refers to the similarity degree between the watermarked map and the original map in the sense that the watermark insertion process should not noticeably affect the shape and quality of the host map [24].

The capacity requirement refers to the amount of inserted watermark bits into the host GIS map. In addition, the watermark bits should be well-distributed over the whole digital map for securing the watermark. The more watermark bits are inserted, the more the host map is changed, which may lead to a decrease in fidelity [21]. Therefore, fidelity and capacity need to be balanced to achieve good security with minimal loss of quality.

The robustness requirement refers to the resilience of the watermark against a potential set of modifications, referred to as attacks, to the host GIS map. Resistance to these attacks is important because they could seriously change the map shape in terms of vertices' coordinates values, and, as a consequence, making the process of watermark extraction more difficult; this, in turn, would jeopardise the identification of the rightful owner of the data. There are several attacks that are relevant for vector map data:

1) rotation attacks, which mean using a specific angle to turn the GIS map around its center [25];
2) translation attacks, which involve moving the whole map by a specific distance towards a specific direction [26];
3) scaling attacks, which refer to the use of a specific value, in both axes, to alter the size of the GIS map [25];
4) simplification attacks, which involve the removal of some vertices from the GIS vector map [22];
5) interpolation attacks, which consist of adding new vertices into the GIS vector map [27].

Despite the use of ESRI shapefiles in GIS vector data watermarking research, e.g. [21], [20], [28], the advantage of the shape bounding box feature in the shapefile header has not been exploited in the watermarking context apart from our previous approach [12]. As shown in Fig. 2, the bounding box properties we are interested in are the minimum and maximum coordinates' values in both horizontal and vertical axes.

III. Related Work

To the best of our knowledge, there are only three published approaches that used partition clustering methods for protecting the copyright of GIS vector maps. In this section, these approaches are reviewed in relation to the trade-off between fidelity and capacity, and in relation to their vulnerability to interpolation and simplification attacks.

Huo et al. [20] presented an approach that used k-means partition clustering for inserting a watermark into a GIS vector map composed of a small number of polygons, based on ESRI shapefile format, according to the polygons' mean centers (i.e. the mean of vertices' coordinates values in the polygon). Although their fidelity achievement is considerably high, the capacity of the watermark was relatively low for the size of the GIS map they used. Therefore, this approach, does not achieve a good trade-off between fidelity and capacity; in addition, this approach is vulnerable to simplification and interpolation attacks.

To improve the approach of Huo et al. [20], we presented an approach [21] that used k-medoids-based partition clustering for inserting watermark bits into a set of GIS vector maps composed of a small number of polygons, and we used mean polygons' centers for identifying the optimum position to insert watermark bits into the GIS maps. Although this approach improved the trade-off between fidelity and capacity, it did not address the vulnerability to simplification and interpolation attacks. Moreover, both approaches did not consider the case of larger maps.

To improve the robustness of our approach [21], we extended it [12] by using k-medoids-based clustering and polygon bounding box information in ESRI shapefiles, for inserting watermark bits into a set of GIS vector maps composed of larger numbers of polygons. In this approach, we used the polygons bounding boxes centers to identify the optimum locations in the GIS map for inserting the watermark bits. This approach achieved: (1) robustness to both simplification and interpolation attacks, (2) a considerable increase in the trade-off between fidelity and capacity and (3) reliability of the approach for GIS vector maps composed of larger number of polygons.

Regardless of the partition method used, we argue in this paper that the use of bounding box property of GIS vector map for locating the watermark bits into polygons' vertices has a significant implication on protecting the GIS vector map copyright, especially in terms of addressing the vulnerability to simplification and interpolation attacks, while preserving a good trade-off between fidelity and capacity. To asses this, we compare a k-means clustering approach based on the bounding box centers of polygons with the k-means approach of Huo et al. [20].

IV. K-means clustering with bounding boxes approach

This section presents our approach based on k-means partition clustering using the bounding box information in the ESRI shapefile. We compare the results of this approach with the work of Huo et al. [20], which used the mean centers of polygons. The purpose is to establish the role of the bounding box property in addressing the vulnerability to simplification and interpolation attacks, and to investigate if the trade-off between fidelity and capacity is preserved.

A. Embedding Locations Identification

The embedding stage involves the identification of locations for embedding the watermark and the insertion of the watermark bits in the identified locations.

The location identification involves three consecutive steps: computing the bounding box centers for each polygon, applying k-means clustering to the polygons' computed centers, and calculating mean distance values (the locations for inserting the watermark bits). Each of these steps is described in the following.

Step 1 : Computing Bounding Box Centers

Each polygon in the GIS vector map has a defined bounding box, which identifies the boundaries of each polygon in the map; the coordinates for the bounding box are available in the shapefile [29], as illustrated in Fig. 2. Polygons' bounding box centers are calculated in both axes, as shown in Equation (1) and Equation (2), respectively.

$$x_c = \frac{x_{min} + x_{max}}{2} \tag{1}$$

$$y_c = \frac{y_{min} + y_{max}}{2} \tag{2}$$

where: x_c and y_c are the coordinates of polygon's center in both x and y axes respectively; x_{min} is the minimum vertex coordinate in x-axis; x_{max} is the maximum vertex coordinate in x-axis; y_{min} is the minimum vertex coordinate in y-axis; y_{max} is the maximum vertex coordinate in y-axis. x_{min}, x_{max}, y_{min} and y_{max} are each of 8-byte length [29].

Unlike the approach to calculating the polygons' centers, based on bounding boxes as explained above, Huo et al. [20] calculate polygons' centers by summing up all vertices coordinates for each polygon and dividing the sum by the number of vertices minus one; the minus one is due to the last vertex coordinates being the same as for the first vertex, according to the polygon shapefile format [29]. These polygons' average centers are quite sensitive to the total number of vertices in a polygon; consequently, adding (interpolation) or removing (simplification) some vertices, will change the average value of the polygons' centers. In contrast, the bounding box centers are independent from the total number of vertices in a polygon; consequently, the use of this property plays a significant role in achieving the required robustness to both simplification and interpolation attacks.

Step 2 : Clustering Polygons' Bounding Boxes Centers

The k-means method is used to cluster the bounding box centers in order to determine the positions for embedding the watermark. The k-means clustering method is relatively simple, easy to implement, and needs a predefined number of clusters (k). We experiment with different numbers of k to explore different values for capacity and the effect they have on fidelity. More specifically, we experiment with values of k that represent approximately 25%, 33% and 50% of the total number of polygons. The resulting centroids are kept as a secret key ($key1$).

Step 3 : Distance Calculation

For each cluster centroid identified at the previous step, unlike the previous approaches [12], [21], [20], the distance length is calculated by measuring the distance from the polygon bounding box top right corner to its center, where the center is calculated as described in Step 1. Equation (3) illustrates the way of computing the distance length of selected polygons. This approach is adding increased robustness to the simplification and interpolation attacks due to the independence of this distance of the number of vertices in a polygon.

$$L_c = \sqrt{(x_c - x_{max})^2 + (y_c - y_{max})^2} \tag{3}$$

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 up right bounding box corner coordinates in x and y axes, respectively.

The values of bounding box distance lengths for all selected polygons are stored as a secret key ($key2$) and they represent the selective positions for embedding the watermark.

B. Watermark Bits Insertion

The concept of zero watermarking [30] is utilized in our proposed watermark embedding process. Zero watermarking aims to exploit some of the host GIS data characteristics in order to generate a more robust watermark. In this case, the topological characteristic of the host GIS data that is used, is the distance length of polygons.

The watermark is constructed by adding or subtracting a bit value of 1 from the distance length of polygons. The watermark is embedded by applying odd-even indexing [31], [20], as outlined in Equation (4).

$$W_i = \begin{cases} T-1, & \text{if } OES(I) = odd \\ T+1, & \text{if } OES(I) = even \end{cases} \tag{4}$$

where: W_i is the ith bit value of the watermark; OES stands for Odd-Even Status; I is the order index of the distance length value in the matrix; T is the value of the 4th digit of the distance length value, after the decimal point [20].

The index of each distance value is used in this approach, instead of using an additional random sequence proposed by [20], to get more consistent positions for embedding the watermark. This consistency sum up both: (a) the indexing as a vital role in the clustering process, and (b) maintaining the security of the watermark position by storing the index values as a key instead of utilizing a random sequence that is not relevant to the used data. This also offers the ability to control the watermark capacity in order to preserve the map fidelity, whereas the use of a random sequence [20] will limit that choice of control.

As shown in Equation (4), the watermark is embedded by comparing the OES (Odd-Even Status) of the I and T variables. The conditions are set based on two scenarios as in the following:

- If the OES of I is odd, 1 will be subtracted from the value of T.
- In contrast, 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^*. This new distance length values are stored as another secret key ($key3$), to secure the positions in which the watermark is embedded. The change rate α_c is calculated as depicted in Equation (5):

$$\alpha_c = \frac{L_c^*}{L_c} \tag{5}$$

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 Equations (6) and (7).

$$v_x^* = \alpha_c v_x + x_c(1 - \alpha_c) \tag{6}$$

$$v_y^* = \alpha_c v_y + y_c(1 - \alpha_c) \tag{7}$$

where: v_x^* and v_y^* are the new vertices' coordinates after embedding the watermark according to the aforementioned condition, in Equation (4); v_x and v_y are the original vertices' coordinates before inserting the watermark bits.

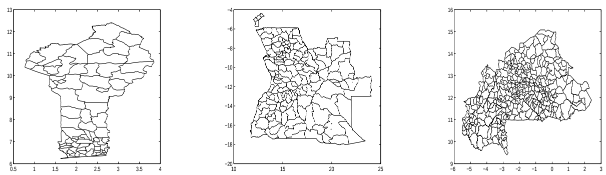
Embedding the watermark bits into the distance length values has the advantage of providing robustness to rotation, translation and scaling attacks.

In rotation and translation attacks the entire map is shifted either by turning the map around to a specific angle or by moving the entire map in a specific direction. These modification apply the same shift to all coordinate values of vertices; this consistent change signifies that the distance values will remain the same. Consequently, the distance lengths are not affected by rotation and translation attacks.

Scaling attacks involve a change in the size of the map by a particular scaling factor. This scaling factor can be determined by dividing the distance values of the attacked map by the distance values of the original map (i.e. $key2$). Consequently, the original map can be restored from the attacked map by applying the complementary scaling factor to the attacked map.

C. Watermark Bits Extraction

Our proposed approach is characterized by blindness and flexibility. Blindness means that the original vector map is not needed in the watermark extraction process, while flexibility means that the watermark extraction process can be implemented in similar way as presented in the watermark embedding process.



(a) Benin (222 polygons) (b) Angola (501 polygons) (c) Burkina Faso (1046 polygons)

Figure 3. The GIS vector maps used in the experiments.

The bounding box centers are computed for each polygon and the k-means method is used to divide the polygons' computed centers into k-clusters in the same way as illustrated in Step 2 (Section IV-A). The results are compared with the stored $key1$ to identify if there have been some modification applied to the vector map; the comparison with the other stored keys (see below) has the same purpose.

The distance length for the watermarked map are calculated in the same way as illustrated in Step 3 (Section IV-A). The recalculated distances are compared to the stored $key2$ and $key3$, to ensure that the vector map has the embedded watermark bits (1 or -1) in order to go further for the extraction stage. Both $key2$ and $key3$ help in retrieving the watermarked map to its original form, which maintain the robustness to rotation, translation and scaling attacks, as discussed in Section IV-B.

V. Experiments and Discussion

Three maps were used to compare the k-means approach using the bounding box property with the k-means approach using the mean centers of polygons. As shown in Fig. (3a), (3b) and (3c), the used GIS maps are polygon-based maps that represent administrative boundaries of 3 countries in Africa: Benin (222 polygons), Angola (501 polygons) and Burkina Faso (1046 polygons). These GIS vector maps are freely available, in ESRI shapefile format, from the Natural Earth website.[1]

ESRI Shapefiles (.shp) are produced by ESRI [2], and considered as a popular format for geographic information system applications [2]. 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 [29].

For the watermark embedding and extraction processes, we implemented the two approaches in MATLAB[3] version R2014b (8.4.0.150421) on 64-bits windows-PC.

The effect of simplification and interpolation attacks on the two approaches has been investigated through the following experiment on the map of Burkina Faso, i.e. the map with the largest number of polygons (1046). As shown in Table I, the map of Burkina Faso contains a total of 113996 vertices. The watermark was inserted in a third of the whole map, i.e. 349 polygons were watermarked containing 39375 vertices, referred to as watermarked vertices in Table I. The total number of removed or added vertices is 7875, which represents 6.9% of the map vertices and 20% of the watermarked vertices.

For each watermarked polygon, 20% of the vertices were removed (for the simplification attack) or added (for the interpolation attack), and the changes in the computed distance values were calculated. The differences are illustrated in Fig. (4) and Fig. (5) and they point out that the interpolation and simplification attacks result in changes when the mean centers approach is used, but have no effect on the bounding box approach. Consequently, the bounding box approach is robust to simplification and interpolation attacks.

Although the changes in the mean distance values when using the mean centers may seem small, they are significant because they distort the watermark, which may lead to the loss of its copyright. Moreover, these small changes may also mean that the quality of the map is still quite high, thus allowing the attackers to use it without the liability of copyright infringement.

The robustness to the simplification and interpolation attacks is ensured by using the bounding box centers due to their independence of the number of vertices in a polygon. In other words, removing or adding new vertices would not affect the main four corners of the bounding box, thus leaving the value of the polygon center unchanged. In contrast, the mean centers approach [20] is vulnerable to these attacks because the center of a polygon is calculated as the average values of vertices' coordinates, thus depending on the number of vertices. Consequently, the removal or addition of vertices will affect the values of the polygons' centers calculated with this approach.

Here, we assume that the attacker will not remove each of bounding box corners' coordinates because removing each of these coordinates will lead to a considerable change in polygon shape and make the map unusable due to the loss of its quality. An attacker would normally be interested in preserving the quality of the map when removing its watermark, so that they can still use it.

To further test the bounding box approach, we investigated the effect of this approach on the capacity and fidelity metrics. Table II compares the results of our approach using the bounding box information with the results for the approach using mean centers [20], to investigate the advantage of using polygons' bounding box based centers over the traditional polygons' mean centers in achieving a good trade-off between fidelity and capacity. The difference between the compared approaches is in the definition of the centers of polygons, i.e. using the bounding box information as explained in Section 4.1 vs using the mean of polygon

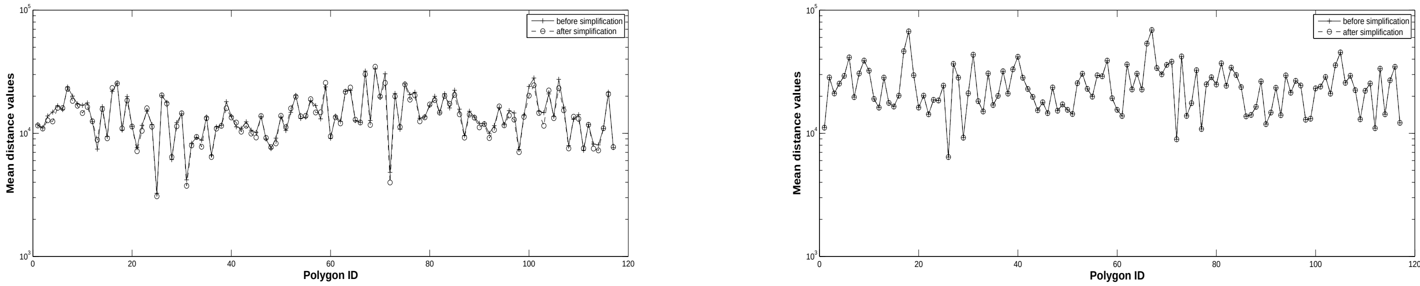
[1] http://www.mapmakerdata.co.uk.s3-website-eu-west-1.amazonaws.com/library/stacks/Africa/index.htm

[2] http://www.esri.com/

[3] http://www.mathworks.co.uk/

Table I
THE PERCENTAGE OF ADDED/REMOVED VERTICES IN RELATION TO THE TOTAL NUMBER OF VERTICES IN THE BURKINA FASO MAP

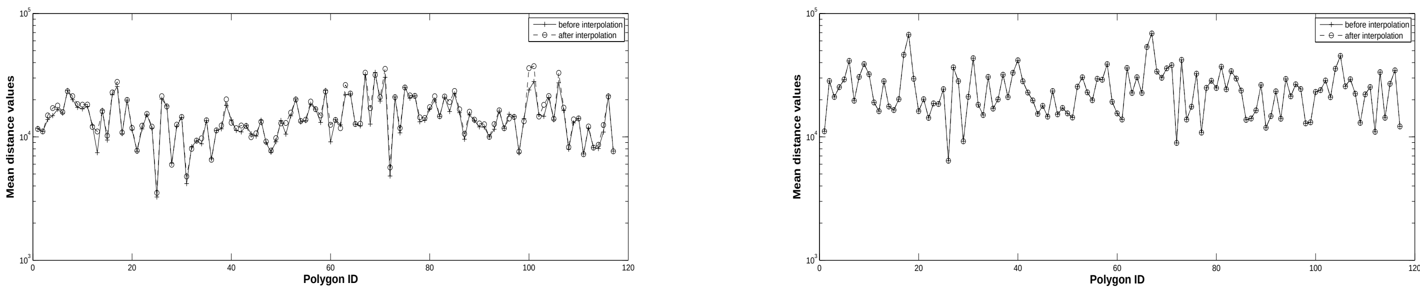
The total number of vertices in the map	113996
The number of watermarked vertices in the map	39375
The number of removed/added vertices	7875



(a) Using mean centers [20]

(b) Using Bounding Box centers

Figure 4. Changes after the simplification attack.



(a) Using mean centers [20]

(b) Using Bounding Box centers

Figure 5. Changes after the interpolation attack.

vertices coordinates in the approach of [20]. Consequently, the difference in results can be attributed to a certain degree to the use of the bounding box properties.

The fidelity metric aims to measure the imperceptibility of the watermark and reflects its degree of invisibility. This invisibility is measured by using PSNR (Peak Signal to Noise Ratio), in decibels [21], [20]. There is no specific range for PSNR values but a higher PSNR would normally indicate that the data is of higher quality [32]. The typical values are considered to be between 30 and 50 dB, in the context of digital images [33]. In order to use this metric, we stored the watermarked GIS maps in JPEG image format.

On the other hand, capacity refers to the number of vertices in the host GIS map, which carry the watermark bits. The importance of the watermark capacity is specified by its vital implication on increasing the watermark robustness to cropping attacks. Cropping is the process of cutting some parts of the host GIS map [34]. Consequently, it is important not only to have high capacity, but also to have the watermark distributed across the entire map.

As shown in Table II, the trade-off between capacity and fidelity is achieved by increasing the number of vertices that carry the watermark bits (capacity) while keeping higher watermark invisibility measured by PSNR (fidelity).

In addition, three different proportions of map size, i.e. 25%, 33% and 50%, were used to observe the effect of increased capacity and its effect on fidelity. These proportions represent approximately a quarter, a third and half of the number of polygons in the used maps. The relation between the map size proportions and the number of clusters is illustrated in the following for each of the three maps used in the experiments. Thus, for the map of Benin, 25%, 33% and 50% corresponds to 56, 74 and 111 clusters, respectively; for the map of Angola, 25%, 33% and 50% corresponds to 126, 167 and 251 clusters, respectively; and for the map of Burkina Faso, 25%, 33% and 50% corresponds to 262, 349 and 523 clusters, respectively.

When looking at the results for the 25% sizes of the three maps in Table II, we notice that the capacity values for the approach proposed in this paper (bounding box-based

Table II
THE RESULTS OF BOUNDING BOX APPROACH VERSUS MEAN POLYGON CENTERS USING K-MEANS

	Our approach		Huo et al. approach [20]	
No. of Clusters (= No. of Polygons)	Capacity (No. of vertices)	Fidelity (PSNR)	Capacity (No. of vertices)	Fidelity (PSNR)
Benin Map (25%)	1278	40.5223	1113	39.7769
Benin Map (33%)	1452	40.1054	1239	40.0029
Benin Map (50%)	2492	37.4146	2363	36.9682
Angola Map (25%)	4009	43.6947	3902	43.6013
Angola Map (33%)	5365	41.9369	5219	41.9134
Angola Map (50%)	9631	39.7193	9507	39.7081
Burkina Faso Map (25%)	15242	40.3900	15171	40.1572
Burkina Faso Map (33%)	19017	39.5909	18930	38.9722
Burkina Faso Map (50%)	31147	36.2769	31012	36.2584

k-means), i.e. 1278, 4009 and 15242, are higher than those of Huo et al. [20], i.e. 1113, 3902 and 15171. At the same time, we notice that the fidelity values are also higher in the approach using the bounding box compared with the approach using the mean centers, despite the increase in capacity. The same can be observed for the 33% and 50% sizes on all three maps.

The research in this paper and in our previous work [12] used clustering for identifying the embedding locations. Clustering has the advantage of ensuring a good distribution of the watermark across the entire map, thus adding resilience to cropping attacks. Other approaches for the identification of locations, however, when used in conjunction with the bounding box property, should still preserve the robustness to simplification and interpolation attacks.

In terms of the trade-off between capacity and fidelity, the experimental results reported in this paper, as well as previous results using k-medoids clustering [12] indicate that the use of the bounding box centers led to a good trade-off between capacity and fidelity; more specifically, the results indicate an increase in fidelity even when there is an increase in capacity. Further experimentation would be needed to assess the role of the bounding box in achieving this results compared with the role of the two other main factors: the approach for identifying the embedding locations and the watermark insertion approach. Although the results indicate that the bounding box plays a role in the trade-off between capacity and fidelity, we cannot separate this effect from the two factors mentioned above.

VI. CONCLUSIONS

In this paper, we investigated the influence of using the bounding box property for protecting the copyright of digital GIS vector maps, by comparing a k-means clustering method that used the bounding box property with the earlier work of Huo et al. [20] using the mean centers of polygons.

Using bounding box centers increases the robustness to the simplification and interpolation attacks due to the independence property of bounding box centers from the number of vertices in a polygon, in contrast to mean centers of polygons, which are dependent on the number of vertices.

The effectiveness of our approach is assessed by looking at both fidelity and capacity aspects. The experiments demonstrate that the computation of bounding box centers has a considerable implication on the trade-off between the fidelity and the capacity metrics, and resulted in higher fidelity as capacity increased.

The PSNR fidelity metric was used for consistency with the work of [20]. This measurement is often used in image watermarking research, and is not necessarily the best metric for GIS vector map [23]. Our future work will investigate different fidelity measurements for identifying more suitable metrics for GIS vector map type of data.

Building on our previous approach based on k-medoids clustering [12] which demonstrate the advantages of using bounding box property for promoting the research of GIS map copyright protection, our approach in this paper stresses that these advantages are maintained even with a k-means clustering based approach, and can reasonably conclude that they would hold regardless of the clustering method used for identifying the watermark embedding locations in the map.

Further research and experiments will be carried out on GIS vector map properties to strengthen the research of GIS vector map watermarking in response to the needs and requirements in copyright protection applications.

REFERENCES

[1] P. Burrough, R. McDonnell, and C. Lloyd, *Principles of Geographical Information Systems*. OUP Oxford, 2013.

[2] P. Longley, M. Goodchild, D. Maguire, and D. Rhind, *Geographic Information Systems and Science*. John Wiley and Sons, 2011.

[3] R. Kingston, "Public participation in local policy decision-making: the role of web-based mapping," *The Cartographic Journal*, vol. 44, no. 2, pp. 138–144, 2007.

[4] M. Angelaccio, A. Basili, B. Buttarazzi, and W. Liguori, “Using geo-business intelligence to improve quality of life,” in *The 1st European Conference on Satellite Telecommunications*. IEEE, 2012, pp. 1–6.

[5] F. Camelli, J.-M. Lien, D. Shen, D. W. Wong, M. Rice, R. Löhner, and C. Yang, “Generating seamless surfaces for transport and dispersion modeling in GIS,” *The Journal of Geoinformatica*, vol. 16, no. 2, pp. 307–327, Aug. 2012.

[6] G. Busch, “GIS-based tools for regional assessments and planning processes regarding potential environmental effects of poplar SRC,” *The Journal of BioEnergy Research*, vol. 5, no. 3, pp. 584–605, 2012.

[7] R. Ciptasari and K. Sakurai, *Multimedia Information Hiding Technologies and Methodologies for Controlling Data*. IGI Global, 2012, ch. Multimedia Copyright Protection Scheme Based on the Direct Feature-Based Method, pp. 412–437.

[8] X. Wang, C. Shao, X. Xu, and X. Niu, “Reversible data-hiding scheme for 2-d vector maps based on difference expansion,” *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 311–320, Sept 2007.

[9] G. F. Bonham-Carter, *Geographic information systems for geoscientists: modelling with GIS*. Elsevier, 2014.

[10] M. Kennedy, *Introducing Geographic Information Systems with ArcGIS.* John Wiley and Sons, 2013.

[11] T. A. Abbas and M. J. Jawad, “Digital vector map watermarking: Applications, techniques and attacks,” *The Oriental Journal of Computer Science and Technology*, vol. 6, no. 3, pp. 333–339, 2013.

[12] A. Abubahia and M. Cocea, “A clustering approach for protecting GIS vector data,” in *The 27th International Conference on Advanced Information Systems Engineering*, 2015, pp. 133–147.

[13] F. Peng, Y. Liu, and M. Long, “Reversible watermarking for 2D CAD engineering graphics based on improved histogram shifting,” *The Journal of Computer-Aided Design*, vol. 49, pp. 42 – 50, 2014.

[14] M. Urvoy, D. Goudia, and F. Autrusseau, “Perceptual DFT watermarking with improved detection and robustness to geometrical distortions,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 7, pp. 1108–1119, 2014.

[15] P. Bhanuchandar, M. Prasad, and K. Srinivas, “A survey on various watermarking methods for GIS vector data,” *The International Journal of Computer and Electronics Research*, vol. 2, no. 3, pp. 5–7, 2013.

[16] J. Wu, F. Yang, and C. Wu, “Review of digital watermarking for 2D vector map,” in *IEEE International Conference on Green Computing and Communications, Internet of Things and Cyber, Physical and Social Computing*, 2013, pp. 2098–2101.

[17] J. Choi, D. Lee, and H. Jung, “Knowledge discovery and integration: A case study of housing planning support system,” in *Mobile, Ubiquitous, and Intelligent Computing*, ser. Lecture Notes in Electrical Engineering, J. Park, H. Adeli, N. Park, and I. Woungang, Eds. Springer Berlin Heidelberg, 2014, vol. 274, pp. 287–291.

[18] A. Croitoru, A. Crooks, J. Radzikowski, and A. Stefanidis, “Geosocial gauge: A system prototype for knowledge discovery from social media,” *International Journal of Geographic Information Science*, vol. 27, no. 12, pp. 2483–2508, Dec. 2013.

[19] J. Han, J. G. Lee, and M. Kamber, “An overview of clustering methods in geographic data analysis,” in *Geographic Data Mining and Knowledge Discovery*. Taylor and Francis Group, 2009.

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

[21] A. Abubahia and M. Cocea, “Partition clustering for GIS map data protection,” in *The 26th International Conference on Tools with Artificial Intelligence*, 2014, pp. 830–837.

[22] S. Jianguo, K. Liang, and X. Songzhu, “Research of lossless digital watermarking technology,” *The Journal of Applied Mechanics and Materials*, vol. 333, pp. 1219–1223, 2013.

[23] X. Niu, C. Shao, and X. Wang, “A survey of digital vector map watermarking,” *International Journal of Innovative Computing, Information and Control*, vol. 2, no. 6, pp. 1301 – 1316, Dec. 2006.

[24] J. Nin and S. Ricciardi, “Digital watermarking techniques and security issues in the information and communication society,” in *Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on*, March 2013, pp. 1553–1558.

[25] S.-H. Lee and K.-R. Kwon, “Vector watermarking scheme for GIS vector map management,” *The Journal of Multimedia Tools and Applications*, vol. 63, no. 3, pp. 757–790, Oct. 2013.

[26] W. Xun, H. Ding-jun, and Z. Zhi-yong, “A robust zero-watermarking algorithm for 2D vector digital maps,” in *Computer, Informatics, Cybernetics and Applications*, ser. LNEE, X. He, E. Hua, Y. Lin, and X. Liu, Eds. Springer Netherlands, 2012, vol. 107, pp. 533–541.

[27] N. Wang and C. Men, “Reversible fragile watermarking for locating tampered blocks in 2D vector maps,” *The Journal of Multimedia Tools and Applications*, vol. 67, no. 3, pp. 709–739, Jan. 2013.

[28] N. Wang, H. Zhang, and C. Men, “A high capacity reversible data hiding method for 2D vector maps based on virtual coordinates,” *The Journal of Computer-Aided Design*, vol. 47, pp. 108–117, Feb. 2014.

[29] ESRI, “ESRI shapefile technical description,” Environmental Systems Research Institute, USA, Tech. Rep., July 1998.

[30] X. Wang, D. Huang, and Z. Zhang, “A robust zero-watermarking algorithm for vector digital maps based on statistical characteristics.” *The Journal of Software*, vol. 7, no. 10, pp. 2349–2356, Oct. 2012.

[31] B. Wu, W. Wang, and D. Miao, “2D vector map watermarking based on spatial relations,” in *The International Conference on Earth Observation Data Processing and Analysis*, 2008, pp. 32–37.

[32] Q. Huynh-Thu and M. Ghanbari, “Scope of validity of PSNR in image/video quality assessment,” *Electronics Letters*, vol. 44, no. 13, pp. 800–801, 2008.

[33] R. Hamzaoui and D. Saupe, “Fractal image compression,” in *Document and Image Compression*, M. Barni, Ed., 2006.

[34] Q. Zhao, L. Sui, C. Wang, and X. Yin, “Publicly verify the integrity of the geographical data using public watermarking scheme,” in *Geo-Informatics in Resource Management and Sustainable Ecosystem*, ser. Communications in Computer and Information Science, 2013, vol. 398, pp. 646–652.