SB-RAWVec – A Semi-Blind Watermarking Method for Vector Maps

Karina Magalhães
Instituto de Computação - UNICAMP
Campinas, SP, Brazil - 13084-971
Email: karina.magalhaes@students.ic.unicamp.br

Ricardo Dahab
Instituto de Computação - UNICAMP
Campinas, SP, Brazil - 13084-971
Email: rdahab@ic.unicamp.br

Abstract—RAWVec [1] is a private watermarking method for vector maps that uses a raster image as watermark. Visually recognizable watermarks can add extra information like integrity and authentication, while blind watermarks do not need the original data to be published for the detection. This work presents a semi-blind version of RAWVec, i.e. a method that uses the original vector map during the detection without revealing it, but maintains the watermark as a raster image.

I. INTRODUCTION

A watermark is some information added inconspicuously to digital data, usually to assure copyright detection on multimedia data, like images, videos and audio. Vector maps are images used in Geographic Information Systems (GIS), especially because they are easily scalable, have small file size and can be updated and maintained easily.

The original data must be embedded with the watermark, thus resulting in the new, copyrighted marked data. To test if some target data is an unauthorized copy of the marked data, a watermark detection procedure should be performed on the target data. This procedure usually requires some extra information to which we refer as the key.

Watermarks can be classified as private or blind according to the information needed during the detection algorithm. Private or non-blind watermarks use, along with the key and the data to be tested, the original data that was marked. Examples of non-blind watermarking techniques for vector are found in [2] and [3].

Blind watermarks use only the data to be tested and the key during the detection. Therefore, the detection is public, because it does not reveal any additional information and anyone can extract the watermark. Most watermarking methods for vector maps are blind, as in [4], [5], [6] and [7], because they usually have a wider range of application, while private watermarks are more efficient.

Another, less known, category [8] is that of Semi-Blind watermarks, in which the original map is watermarked and this marked version is used during the detection. Note that the original map is used during the detection, but it is not revealed.

In this paper we propose SB-RAWVec, an alternative version for RAWVec [1], a method for watermarking vector maps in spatial domain, i.e. the embedding algorithm modifies the coordinate of vertices. The main advantage of the present proposal is that it is semi-blind, while the original method uses a private technique.

This paper is organized as follows: in Section II some functions and concepts used on the algorithms are presented; in Section III we describe the embedding and detection algorithms for the original RAWVec method; in Section IV we describe the embedding and detection algorithms for SB-RAWVec; in Section V we analyze and compare the two methods and discuss their robustness against several attacks. Finally, in Section VI, we present conclusions and discuss future work.

II. BASIC CONCEPTS

The original RAWVec method and its semi-blind alternative use two functions $w()$ and $v()$ and a Point Pattern Matching Algorithm. These concepts are described in the following sections.

A. Functions $w()$ and $v()$

The function $w()$ modifies a square matrix $A_{n \times n}$ by shifting its elements. Let $A$ be a square matrix of order $n$; then $B = w(A)$, has $b_{ij} = a_{u,v}$, where $u = n - i + 1$ and $v = n - j + 1$. Figure 1 shows an example of an application of function $w$.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \Rightarrow w(A) = \begin{pmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{pmatrix}$$

Figure 1. An example of function $w(A)$

It is important to notice some properties of function $w()$:

(a) Property 1: $w(w(A)) = A$, for $A$ a square matrix.

(b) Property 2: $c * w(A) = w(c * A)$, for $c$ a real number, and $A$ a square matrix.

(c) Property 3: $w(A) + w(B) = w(A + B)$, for $A$ and $B$ squares matrices.

The watermark used in both methods is a bitmap image represented by a bi-dimensional matrix whose
elements store the color information or intensity for each pixel. The vector map must also be represented as a matrix, which can be obtained using a function \( v(M) \), that calculates the point representation \( P \) from a vector map \( M \). Vector maps are formed by geometric structures and, based on these structures’ points, it is possible to create the point representation \( P = v(M) \). Thus, each structure is decomposed into points, which are stored according to their type:

- **Punctual objects**: Structures described by points, like symbols and text. The coordinates of each point are stored.
- **Linear objects**: Structures described by point sequences, such as lines and polygons. The coordinates of each point in the sequence are stored.
- **Parameterized objects**: Structures described by parameters, like circles and ellipses. The object can be segmented in a few points, and these are stored; or the main points can be stored. If the object is a circle, for example, we can store a few points of the circle, or only its center.

Each coordinate must have a reference to its original object, making it possible to rebuild all the structures from the point representation.

**B. Point Pattern Matching**

A **Point Pattern Matching** algorithm compares two sets of points and determines a transformation that maps one set onto the other under a specified set of criteria. That is, given two sets of points \( P \) and \( Q \), it finds a transformation \( T \) which takes \( P \) to \( Q \); thus \( P = T(Q) \). This transformation \( T \) can be found even if some random noise is added or some points are removed from the sets of points.

There are several types of Point Pattern Matching algorithms, as described in [9]. These algorithms can be used to detect and remove any geometric attacks on marked vector maps, as well as attacks based on object reordering and insertion. The main disadvantage of these algorithms is their inefficiency. In contrast, the methods presented in this paper use the algorithm described in [10], in which the nearest neighbors of each point are calculated and iteratively several transformations are tested until a global transformation is found. This algorithm is the most efficient we have been able to find in the literature.

**III. The Original RAWVec Algorithm**

This section describes the original RAWVec method [1]. The embedding algorithm marks the vector map with the raster image by shifting the geometric structures’ coordinates. The detection algorithm uses the original vector map to extract a raster image that is then compared with the original one.

**A. Watermark Embedding**

Algorithm 1 shows the main steps for embedding a raster image \( R \) in the vector map \( M \) producing the marked vector map \( M' \). It uses a positive real constant \( C \) that controls the maximum shifting of a pixel, so it will stay within the maximum tolerable error of the vector map. We discuss this issue in more detail in Section V-A.

**Algorithm 1 Watermark Embedding**

**INPUT**: raster image \( R \), vector map \( M \) and constant \( C \). **OUTPUT**: marked vector map \( M' \).

1. Calculate the point representation \( P \leftarrow v(M) \).
2. Obtain matrices \( A_x \) and \( A_y \) from \( P \).
3. Resize raster image \( R \), creating the raster image \( E \).
4. Calculate \( B_x \leftarrow CE + A_x \); \( B_y \leftarrow Cw(E) + A_y \), for \( C \) a real constant.
5. Build the marked vector map \( M' \) from matrices \( B_x \) and \( B_y \).

Steps 1 and 2 build two square matrices – \( A_x \) for the \( x \) coordinate and \( A_y \) for the \( y \) coordinates – of point representation \( P = v(M) \), as described on Section II-A. Let \( t \) be the number of points stored in \( P \), \( n \) be the order of the matrices and \( p_i = (x_i, y_i) \) be the points stored in \( P \). Then

\[
\begin{align*}
\sqrt{t} = n, \quad & (1) \\
(A_x)_{ij} &= x_{n(i-1)+j} \quad \text{and} \quad (A_y)_{ij} = y_{n(i-1)+j} \quad (2)
\end{align*}
\]

Step 3 resizes the raster image \( R \), creating a new raster image \( E \) that has \( n \times n \) pixels, as illustrated in Figure 2.

**Figure 2. Resizing a raster image \( R \), creating a new raster image \( E \)**

Now that we have two matrices, \( A_x \) and \( A_y \), representing the vector map \( M \), and a matrix \( E \), representing the raster image, we are able to embed \( E \) in the vector map. Therefore, Step 4 creates two matrices, calculating \( B_x \) and \( B_y \):

\[
B_x = CE + A_x, \quad \text{and} \quad B_y = Cw(E) + A_y. \quad (3)
\]

Finally, in the last step, the marked vector map \( M' \) is built from matrices \( B_x \) and \( B_y \), as described in Section II-A.
B. Watermark Detection

The detection algorithm consists in the extraction of the watermark and its comparison with the original one. It extracts the watermark \( S \) from the target vector map \( N \), the one that is going to be tested, using the constant \( C \), the original watermark \( R \) and the original vector map \( M \). Therefore, this is not a blind watermark method. The basic steps are described below as Algorithm 2.

Algorithm 2 Watermark detection

**INPUT:** raster image \( R \), vector map \( M \), constant \( C \) and the vector map to be tested \( N \).

**OUTPUT:** watermark \( S \).

1. Embed the watermark \( R \) on the vector map \( M \), using the algorithm described in Section III-A and returning the marked vector map \( M' \).
2. Calculate the point representations \( P \leftarrow v(M') \) and \( Q \leftarrow v(N) \).
3. Compare the point representations \( P \) and \( Q \) using Point Pattern Matching, returning a transformation \( T \), where \( Q = T(P) \).
4. Build matrices \( A \) and \( B \) from \( v(M) \), and \( B' \) and \( B'' \) from \( T^{-1}(Q) \).
5. Calculate \( D_x \leftarrow \frac{B_x - A_x}{C} \), \( D_y \leftarrow \frac{w(B_y) - w(A_y)}{C} \).
6. Calculate \( D \leftarrow \frac{D_x + D_y}{2} \).
7. Resize the watermark \( D \) to its original size, returning the watermark \( S \).

The first step of the detection algorithm is to embed the watermark \( R \) on the vector map \( M \), resulting on the marked map \( M' \). This marked vector map must be compared with the vector map \( N \), using the Point Pattern Matching Algorithm [10], so that any transformation \( T \) used in an attack is found and removed from the vector map \( N \).

Now we are able to build four matrices \( A \) and \( B \) from \( v(M) \), and \( B' \) and \( B'' \) from \( T^{-1}(N) \). Matrices \( A \) and \( B \) represent the original vector map and matrices \( B' \) and \( B'' \) represent the vector map \( N \) that may be \( M' \). So, we must calculate the watermark in vector map \( N \) and compare it with the original one. The watermark \( D \) is calculated using the four matrices \( A \), \( B \), \( B' \) and \( B'' \), the constant \( C \) and the function \( w \):

\[
D = \frac{D_x + D_y}{2},
\]

where

\[
D_x = \frac{B_x - A_x}{C} \quad \text{and} \quad D_y = \frac{w(B_y) - w(A_y)}{C}.
\]

Finally, the watermark \( D \), with size \( n \times n \), must be resized to its original size – the size of watermark \( R \) – producing watermark \( S \). Watermarks \( S \) and \( R \) must be compared using a probabilistic algorithm and the human eye. According to this comparison we can conclude whether or not the vector map \( N \) was marked.

IV. SB-RAWVec Algorithm

This section presents the modifications made on the original RAWVec method, resulting in a semi-blind alternative. Two marked vector maps are created, one must be published as the marked vector map, referred as public map, and the other is only used during the detection algorithm, referred as the detection map. The original vector map is not needed during the detection and a new constant \( I \) is added to help control the maximum shifting of a pixel along with the \( C \) constant.

A. Watermark Embedding

Algorithm 3 shows the steps for the watermark embedding of the new method. The main difference appears on the Step 4, in which the raster image is divided in two matrices, resulting in two marked vector maps.

Algorithm 3 Watermark Embedding

**INPUT:** raster image \( R \), vector map \( M \) and two constants \( C \) and \( I \).

**OUTPUT:** public marked vector map \( M_1 \) and detection marked vector map \( M_2 \).

1. Calculate the point representation \( P \leftarrow v(M) \).
2. Obtain matrices \( A \) and \( B \) from \( P \).
3. Obtain matrices \( A \) and \( B \) from \( P \).
4. Decompose matrix \( E \) into random matrices \( E_1 \) and \( E_2 \) such that \( E = E_1 + E_2 \), and \( 0 \leq z \leq I \) for each \( z \) an element of \( E_1 \).
5. Calculate \( B_1x \leftarrow A_x + CE_1 \); \( B_1y \leftarrow A_y + CW(E_1) \), for \( C \) a real constant.
6. Calculate \( B_2x \leftarrow A_x - CE_2 \); \( B_2y \leftarrow A_y - CW(E_2) \), for \( C \) a real constant.
7. Build the public marked vector map \( M_1 \) using matrices \( B_1x \) and \( B_1y \).
8. Build the detection marked vector map \( M_2 \) using matrices \( B_2x \) and \( B_2y \).

Steps 1 to 3 are exactly the same as in the original method. A point representation for the vector map \( M \) is created, as well as two matrices \( A \) and \( B \), representing the \( x \) and \( y \) coordinates respectively. These two matrices must be square and have the same size. Next, the raster image is resized, in order to obtain a square matrix \( E \) with the same size as \( A \) and \( B \).

Matrices \( A \) and \( B \) represent the positions for the objects in the vector map, and matrix \( E \) represents the color for each pixel in the raster image.

Step 4 represents the creation of two matrices \( E_1 \) and \( E_2 \), using the positive integer constant \( I \) and the matrix \( E \). This step is based on Visual Cryptography [11], in
which an image is divided into two, and can be recreated by superimposing the two shares. In our case, the image can be reconstructed by adding the two shares. Note that the image $E_2$ is not visual, as it can have negative elements.

Using $A_x$, $A_y$ and $E_1$, it is possible to obtain two matrices $B_{1x}$ and $B_{1y}$, and therefore build the public marked vector map $M_1$:

$$B_{1x} = A_x + CE_1, \quad \text{and} \quad B_{1y} = A_y + Cw(E_1). \quad (6)$$

And, using $A_x$, $A_y$ and $E_2$, it is possible to obtain two other matrices $B_{2x}$ and $B_{2y}$, and build the detection marked vector map $M_2$:

$$B_{2x} = A_x - CE_2, \quad \text{and} \quad B_{2y} = A_y - Cw(E_2). \quad (7)$$

Note that the sum used in the calculation of the public map $M_1$ is replaced by a subtraction in the calculation of the detection map $M_2$.

The error added to $M_1$ can be controlled by the constants $C$ and $I$, thus forcing the error added to $M_2$; therefore, $M_2$ can only be used during the detection algorithm. This is detailed in Section V-A.

### B. Watermark Detection

Although it does not use the original map $M$, the detection algorithm uses a Point Pattern Matching algorithm to detect and remove attacks. This is possible due to the information hidden on the detection map $M_2$. Algorithm 4 shows the basic steps for this algorithm.

The first two steps are simple, they only create the point representations and the respective matrices for the two maps: $B_{2x}$ and $B_{2y}$ for the detection map $M_2$ and $B_x$ and $B_y$ for the target map $N$. The first main difference between this algorithm and Algorithm 2 is Step 4, in which the hidden information from the detection map is obtained.

From the embedding algorithm, described in IV-A, we have that $B_{2x} = A_x - CE_2$, $B_{2y} = A_y - Cw(E_2)$ and $E = E_1 + E_2$, so we can extract the information needed for the Point Pattern Matching algorithm from $B_{2x}$ and $B_{2y}$, as follows:

$$B'_{2x} = B_{2x} + CE = A_x - CE_2 + CE = A_x + C(E - E_2) = A_x + CE_1 = B_{1x}$$

$$B'_{2y} = B_{2y} + Cw(E) = A_y - Cw(E_2) + Cw(E) = A_y + Cw(E - E_2) = A_y + Cw(E_1) = B_{1y}$$

If the target map $N$ is a tampered version of public map $M_1$, we can compare it with $M_1$ using a Point Pattern Matching algorithm. This algorithm will find any transformation or modification that maps $N$ onto $M_1$, comparing matrices $B'_{2x}$ and $B'_{2y}$ with matrices $B_x$ and $B_y$.

Step 6 removes the transformation and modification found by the Point Pattern Matching algorithm from $B_x$ and $B_y$, resulting in two new matrices $B'_x$ and $B'_y$. Now we are able to calculate the extract watermark and verify if the target map $N$ has been stolen:

$$D = \frac{D_x + D_y}{2}, \quad (8)$$

where

$$D_x = \frac{B'_x - B_{2x}}{C} \quad \text{and} \quad D_y = \frac{w(B'_x) - w(B_{2y})}{C}. \quad (9)$$

Since the target map $N$ is $M_1$ tampered, $B'_x$ and $B'_y$ will be equivalent to $B_{1x}$ and $B_{1y}$, so:

$$D_x = \frac{B'_x - B_{2x}}{C} = \frac{B_{1x} - B_{2x}}{C} \quad \frac{(A_x + CE_1) - (A_x - CE_2)}{C} = \frac{CE_1 + CE_2}{C} = \frac{E_1 + E_2}{C} = E$$

The calculation of $D_y$ is more complex because the $w()$ function is used; therefore the three properties described in II-A are essential:
The maximum shifting of a pixel is given by $\pm C \cdot \text{max}(\epsilon)$, where $\epsilon \in R$. This control refers only to the intensity of the pixels of the raster image and no change on the intensities of the pixels of the raster image is needed. The maximum shifting of a pixel in this case is given by $\pm CI \sqrt{2}$. This control refers only to the error added to the vector map. Furthermore, an attack and the other for the maximum shifting of a pixel i.e., the answer from 5 people to the question “Do you think that the image $R$ is the image $D$ after some modifications?” Each positive answer represents 1 and each negative answer represents 0, and $h$ is the sum of the answers. The Pearson correlation coefficient $r$ is based on the pixels intensities and it is calculated as follows:

$$r = \frac{\sum_{i,j} (r_{ij} - I_{mR})(s_{ij} - I_{mS})}{\sqrt{(\sum_{i,j} (r_{ij} - I_{mR})^2)(\sum_{i,j} (s_{ij} - I_{mS})^2)}}. \quad (10)$$

### B. Attacks

Most attacks on watermarks for vector maps are based on the objects and vertex information of the vector maps [13]. Geometric attacks consist on transformations, such as rotations and translations, on the marked vector map with the goal of removing or degrading the watermark without damaging the vector map. Other attacks consist of: reordering or removal of some objects of the vector map; reordering and removal of vertices of some objects of the vector map; and adding random noise.

All these attacks can be detected and removed, to a certain extent, by a Point Pattern Matching algorithm that compares the vector map to be tested with a marked vector map. The transformations are found and returned by the algorithm, as well as added random information and points or objects removed from one of the maps. Therefore, the Point Pattern Matching is essential to the robustness of these methods.
Figures 5 and 6 show the extracted watermark of a simple test, without any attacks, for the original RAWVec method and the SB-RAWVec method, respectively. Figures 7 and 8 show the extracted watermark of a test with a combined attacked vector map for the same two methods. In this case, the marked vector map was rotated $10^\circ$, translated, scaled to double of its original size, $11\%$ of its points were reordered, and some random noise was added. The coefficients $r$ and $h$ are calculated for all the tests, as described in V-A. Figure 4 shows the original watermark used during the tests. Several other tests can be found in [14] (in portuguese).

The original images used as watermarks showed in Figures 4, 5, 6, 7 and 8 are small ($60 \times 68$ pixels), so that several copies can be added to a single map as pictured in Figure 2. This redundancy allows for an easier retrieval of the right information about a pixel, because during the extraction, more than one copy of each pixel is analyzed.

The original vector map used in this paper was produced by Robert Ford, which can be found in http://www.go.dlr.de/pdinfo_dv/xfig_Examples/transit.fig.

VI. CONCLUSIONS AND FUTURE WORK

In this work, the RAWVec method for watermarking vector maps is revised and an alternative SB-RAWVec method is presented. The new method is as efficient and robust as the original one, since they both make use of a Point Pattern Matching algorithm and redundancy to detect and remove attacks.

This new method has the same advantages as the original one - such as the use of an image as the watermark - with the addition of the semi-blinding feature. Although it may seem a marginal improvement, blind and semi-blind schemes are more practical, since the original map may not be available for a number of reasons, limiting the application of the non-blind method.

Another advantage of our new method is a better control of the maximum shifting of a pixel, i.e. the quality of the marked vector map, now controlled by two constants.

Other types of attacks on SB-RAWVec remain to be studied; such as those more commonly seen in watermarking for other medals like video, audio and bitmap images.

REFERENCES