

ROBUSTNESS AND ACCURACY ASSESSMENT OF INVISIBLE WATERMARKING OVER GEOSPATIAL VECTOR DATA

Sangita Zope-Chaudhari¹, Parvatham Venkatachalam², Krishna Mohan Buddhiraju³

¹Research Scholar, CSRE, IIT Bombay, Mumbai, India, sangita.z@iitb.ac.in

²Professor and Head, CSRE, IIT Bombay, Mumbai, India, pvenk@csre.iitb.ac.in

³Associate Professor, CSRE, IIT Bombay, Mumbai, India, bkmohan@csre.iitb.ac.in

KEYWORDS— *Vector data, Fidelity, Attacks, Copyright Protection*

ABSTRACT— Fast development of Internet makes the process of data sharing very easy. However, this leads to unlimited copying and duplication of data. Digital watermarking has been used from many years to protect images from piracy. Geospatial vector data acquisition and generation is very complex and expensive task and it needs to be protected from unauthorized users. In this paper, we have considered distinctive characteristics of geospatial vector data and proposed wavelet based watermarking algorithm. Furthermore, we have examined the watermarked geospatial vector data against attacks specific to geospatial vector data. Accuracy assessment is done using conventional as well as specialized geospatial vector data quality assessment measures. Proposed method retains accuracy as well as robust against attacks like noise, compression, geometric transformations, addition/deletion of vertices and cropping.

I. INTRODUCTION

Geographical Information System (GIS) is a computer system for capturing, managing, integrating, manipulating, analyzing and displaying geospatial data. Geospatial data can be used in many ways. Disaster management can use parcel (land use, zonal, and cadastre) information for damage assessment after a disaster. Water resource management can be done using rainfall, land use, drainage, cadastral etc. Real estate applications are using land use data. Similarly geospatial data can be used in agriculture, environment monitoring, defense, transportation management, business etc.

For many years, the owners of geospatial data sets are interested in finding protection mechanisms against piracy. Earlier the threat was republishing of a geographic map without paying royalties. In such case, proving ownership was the problem. Some cartographers used to insert tiny errors in map to prove the ownership. Now analog map gets replaced by digital maps and the threat is that any of the legitimate owners can be source of an illegal copy which in turn is an exact equivalent to the original. In the last two decades, the growth of high speed computer networks and World Wide Web (WWW) have explored means of new business, scientific and social opportunities in the form of electronic publishing, real-time information delivery, data sharing, collaboration among computers, digital repositories and many more. It becomes easy to store and manipulate high quality digital maps efficiently using computers. Distribution of digital maps through Internet leads to unlimited copying and thus creates a real threat for map owners. Therefore, it becomes urgent to solve issue of copyright protection. Digital watermarking is one of the remarkable solutions available to deal with copyright protection

Digital watermarking algorithms are classified as spatial domain and transform domain algorithms depending on the approach used for embedding the watermark data. Spatial domain algorithms directly alter co-ordinates (or pixels) to hide watermark data. Transform domain algorithms alter frequency transform of data elements to embed watermark data. Transform domain includes Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT), and Discrete Wavelet Transform (DWT).

In vector map, there is high correlation among the neighboring vertices of same feature. Polyline feature is used for embedding watermark by Cao et al. (2010). Highly correlated vertices of polyline are grouped together and their median is calculated. Watermark bits are embedded in the median value of each group iteratively. This scheme produces less distortion due to high correlation between the vertices and also has higher payload capacity. In a blind watermarking scheme reported by Huo et al. (2010), polyline/polygon characteristics of map are used. Length/perimeter is calculated for all polyline/polygon in a map. Considering uniform step (key) of length/perimeter dynamic range, polylines/polygons are arranged in some groups. Watermark is inserted multiple times. This scheme is robust against attacks like geometric transformation, object order scrambling, swapping, vertex addition/deletion, and cropping. Cheng et al. (2010) have reported an algorithm based on data configuration of vector map along with theory of error correction. The features with more vertices are selected for watermark embedding. Even though objects are rotated and translated, angle between them remains unchanged. This feature is used by Kim(2010). An interior angle is calculated by applying cosine rule on three consecutive vertices of an

object. Watermark is inserted in the angle and vertices are transformed according to changed angle. In the algorithm proposed by Magalhaes and Dahab (2010), Vertices are shifted depending on the watermark bit values inserted into vector data.

Transform domain methods depicted by (Tao et al.(2009), Kitamura et al.(2001), Liang et al.(2011), Lianquan and Qihong(2007), Zhu et al.(2008)) works on transformed coordinates of vector data. In the blind watermarking scheme proposed by Tao et al.(2009), phase of DFT is quantized using step size and then watermark is embedded in quantized phase. This method fairly deals with attacks like similarity transformation, Geometric transformation, and format conversion. DFT is used for embedding watermark in a set of polylines in vector map (Kitamura et al., 2001). The scheme described by Liang et al. (2011) has used feature domain along with transform domain. In this scheme, Minimum Bounding Rectangle (MBR) is computed and divided into uniform grids. Scrambled watermarking bits are embedded in medium frequency DCT coefficients of grid weighted array for each grid. This algorithm gives good imperceptibility and also robust against geometric transformation and simplification attack. In similar approach, feature vertices are extracted using Douglas-Peucker algorithm (Douglas and Peucker, 1973) to form feature image. DCT transform is performed on this feature image and watermark is embedded into the middle and low frequency coefficients. Finally inverse DCT transform is applied on the adjusted coefficients to get the watermarked data. This scheme results into good imperceptibility and robust against some common attacks(Lianquan and Qihong, 2007). In a non-blind algorithm depicted by Zhu et al.(2008), watermark bits are embedded in low frequency coefficients by applying integer wavelet transform on vector data. This algorithm can effectively resist the attacks like noise, data compression, point deletion and format exchange.

The motive behind watermarking of geospatial vector data is to stop illegal distribution, forgery, and tampering of valuable geospatial data. This technique also helps in data source tracing and authentication. Watermarking algorithm used for copyright protection of geospatial data can vary according to its characteristics. They are mainly classified as vector and raster data watermarking algorithms. The requirements for vector data watermarking are: (1) Good robustness against attacks; (2) Invisible watermarking scheme; (3) Precision and positional accuracy retention; (4) Topological relationship perseverance. All these requirements are taken into consideration while designing and evaluating the proposed algorithm.

This paper reports robustness and accuracy assessment of wavelet based watermarking of geospatial vector data. Section II describes the methodology of proposed system. In section III and IV, evaluation measures and experimental results are presented. Conclusions are drawn in section V.

II. PROPOSED METHODOLOGY

In watermark embedding process, wavelet transform based watermarking algorithm is proposed for copyright protection of vector data. As wavelet transform reflects the local features better and it is not sensitive to local modification, it is a good choice to use wavelet for watermarking.

A. Watermark Embedding

Shapefile is read and x-y coordinate sequence generated from all the vertices is considered as an input. Multilevel wavelet decomposition is performed on these x-y coordinate sequences, and then shuffled watermark is inserted in low frequency coefficients of decomposed data. The watermark is inserted multiple times to increase the robustness of the algorithm. The watermarked data is transformed by inverse wavelet transform.

Algorithm: Watermark Embedding

Input: Host Vector Data (V), Binary Watermark (W), Embedding strength (P)

Process:

- 1) Shuffle the input watermark (length= m) by adding pseudorandom sequence (PS) of same size.

$$W_{S_i} = W_i \oplus PS_i, \quad 0 \leq i \leq m \quad (1)$$

- 2) Apply 1-D Discrete Wavelet Transform (DWT) to x and y coordinates of vector data at third level.
- 3) Modify the DWT low frequency coefficients LL_{x_i} and LL_{y_i}

$$\begin{aligned} LL'_{x_i} &= LL_{x_i} + PW_{S_i} \\ LL'_{y_i} &= LL_{y_i} + PW_{S_i} \end{aligned} \quad \text{where } 0 < P \leq 1 \quad (2)$$

- 4) Retain high frequency coefficients as it is.
- 5) Apply inverse DWT to obtain the watermarked vector data.

Output: Watermarked Vector Data (V_w).

B. Watermark Extraction

Watermark extraction is the inverse procedure of embedding. Watermark is extracted by applying DWT on both original data and watermarked data. Low frequency coefficients of both are compared to detect watermark W .

Algorithm: Watermark Extraction

Input: Watermarked Vector Data (V'_w), Host Vector Data (V), Embedding strength (P)

Process:

- (1) Using 1-D DWT, obtain third level decomposition of the watermarked vector data .
- (2) Extract the shuffled watermark from low frequency band:

$$W'_{S_i} = \left(LL''x_i - LLx_i \right) / P \quad or$$

$$W'_{S_i} = \left(LL''y_i - LLy_i \right) / P \quad (3)$$

- (3) Obtain binary watermark w' by inverse shuffling the watermark obtained in step 2.

Output: Binary watermark (W)

III. ERROR ANALYSIS MEASURES

Whenever watermark is embedded into host data, it leads to an error between original vector data and watermarked vector data. Here, we have used maximum error and mean square error (MSE) for watermark data accuracy assessment [12].

$$Max. Error = \max \left(\sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2} \right) \quad i = 1, 2, \dots, n \quad (4)$$

$$MSE = \frac{\sum_{i=1}^n \left[(x_i - x'_i)^2 + (y_i - y'_i)^2 \right]}{n} \quad i = 1, 2, \dots, n \quad (5)$$

where (x_i, y_i) is original vector data coordinates and (x'_i, y'_i) is coordinate of watermarked data. To evaluate similarity between original watermark and extracted watermark, Normalized Correlation (NC) is calculated. It is represented as:

$$NC = \frac{\sum_k (w_k \cdot w'_k)}{\sqrt{\sum_k (w'_k)^2}} \quad (6)$$

where w_k is original watermark, w'_k is extracted watermark, and k is the length of the watermark.

IV. EXPERIMENTS AND RESULTS

Proposed scheme is evaluated using 2-D polyline, polygon and road segment vector data of varying coordinate points. The watermark of size 20X40 pixels is used for watermarking. We have used Haar wavelet in watermarking algorithm implementation. Figure 1 shows input vector dataset and figure 2 shows original and shuffled watermark.

A. Error Analysis

Watermarking causes little distortion in input vector data. We have used maximum error and MSE to inspect data accuracy of vector data with watermark. Accuracy assessment is done using varying embedding strengths. Table 1 shows error analysis for various vector data at third decomposition level using Haar wavelet with varying

embedding strengths. From the various results obtained, it is noted that error increases as embedding strength increases.

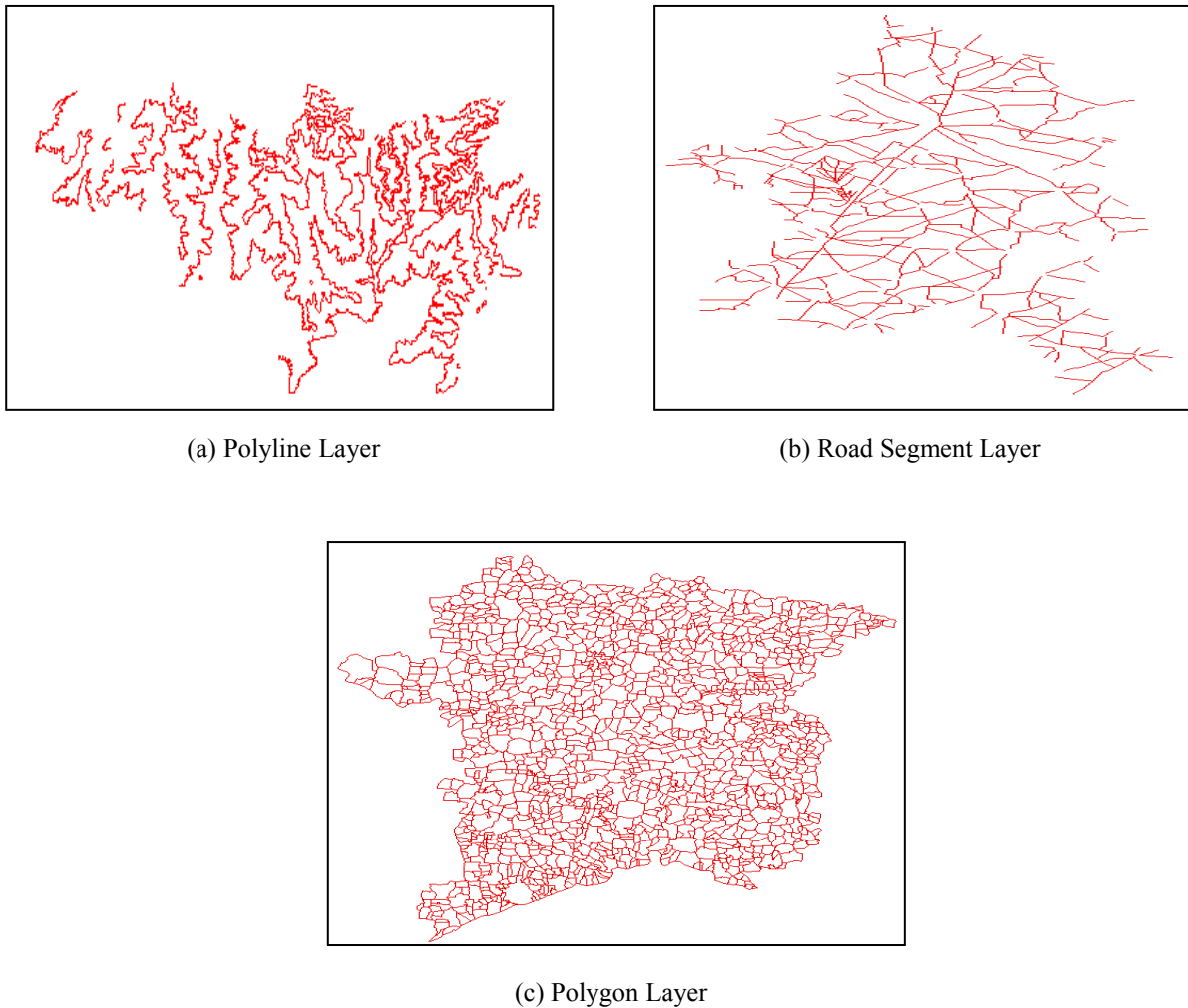


Figure 1: Input Vector Dataset

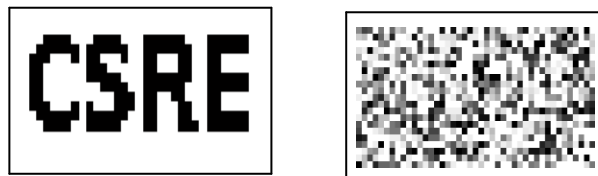


Figure 2: Original and Shuffled Watermarks

B. Robustness Evaluation

Robustness of watermarking algorithm denotes its efficiency to oppose some common data processing operations. Normalized correlation coefficient is calculated between original watermark and extracted watermark from attacked vector data. Robustness of the proposed watermarking scheme is evaluated against following attacks:

1. Imposition of random noise: Random noise with amplitude 0.01 units is added into vertex coordinates of watermarked data.
2. Translation: Vector map coordinates(x and y) are translated by 10 units.
3. Scaling: vector map is uniformly shrunk by 0.5 times (down scaling) and uniformly enlarged by scaling factor 3 (up scaling).










4. Cropping: 10% and 40% cropping of watermarked vector data is done and then the watermark is extracted from it.
5. Co-ordinate addition/deletion: approximately 10% new vertices are added into watermarked vector data and 5% of vertices (non feature) are deleted from the watermarked data.

Table 1 Error Analysis for Different Vector Data with Varying Embedding Strength at Third Level Wavelet Decomposition using 'Haar' Wavelet

| Vector Data | Error | Embedding Strength(P) | | | |
|--------------------|------------|-----------------------|-----------|-----------|-----------|
| | | 0.01 | 0.1 | 0.3 | 0.7 |
| Polyline Layer | Max. Error | 5.16E-5 | 5.16E-4 | 1.5E-3 | 3.5E-3 |
| | MSE | 4.2021E-10 | 4.2021E-8 | 3.7819E-7 | 2.059E-6 |
| Polygon Layer | Max. Error | 5.45E-5 | 5.45E-4 | 1.6E-3 | 3.7E-3 |
| | MSE | 4.9688E-10 | 4.9686E-8 | 4.4718E-7 | 2.4346E-6 |
| Road Segment Layer | Max. Error | 5.6905E-5 | 5.6905E-4 | 1.8E-3 | 3.9E-3 |
| | MSE | 8.6240E-10 | 8.412E-8 | 7.5713E-7 | 4.1222E-6 |

Table 2 shows detected watermarks and correlative coefficients for the above listed attacks on polygon watermarked data at third level wavelet decomposition. The proposed method has good robustness against noise and vertex deletion attack. It is also observed that the algorithm offers good robustness against cropping attack as watermark is embedded multiple times. Finally, for the attacks like format change of watermarked data, addition of vertex coordinates, translation, and scaling the watermark is extracted perfectly with correlative coefficient equal to one.

Table 2 Robustness against attacks

| Attacks | | Extracted Watermark (Proposed scheme) | Normalized Correlation (NC) |
|--------------------------|--------------|---|-----------------------------|
| Noise | |  | 0.903269 |
| Compression | |  | 0.876492 |
| Translation | |  | 1 |
| Scaling | Up Scaling |  | 1 |
| | Down Scaling |  | 1 |
| Coordinate Addition | |  | 1 |
| Coordinate Deletion (5%) | |  | 0.954062 |
| Cropping (25%) | |  | 0.999430 |
| Cropping (40%) | |  | 0.978564 |

C. Topology Relation Inspection

Topology relations for single polyline and polygon layer are considered for evaluation of the proposed scheme. Here, cluster tolerance and rank is considered as 0.001 and 1 respectively for both polyline and polygon data. For different embedding strengths, topological relations are checked for watermarked vector data. The topological rules "Must not overlap", "Must not intersect", "Must not self-overlap", and "Must not self-intersect" are used for testing topological relationship. It has been observed that no lines are overlapping themselves and other lines but some of the lines are crossing themselves and each other. Also, the embedding strength considered in watermark embedding has enormous effect on topological relationship between the features. Table 3 shows topological errors for different embedding strength.

Table 3 Topological Errors for Different Vector Data for varying embedding Strength

| Vector Data | Topology error | Embedding strength(P) | | | |
|--------------------|--------------------------|---------------------------|-----|-----|-----|
| | | 0.01 | 0.1 | 0.3 | 0.7 |
| Polyline Layer | Must not intersect | 2 | 2 | 4 | 6 |
| | Must not self- intersect | 0 | 0 | 0 | 1 |
| Polygon Layer | Must not overlap | 1 | 4 | 4 | 6 |
| | Must not have gaps | 1 | 1 | 2 | 9 |
| Road segment Layer | Must not self- intersect | 0 | 0 | 2 | 4 |
| | Must not have dangles | 0 | 0 | 4 | 5 |

V. CONCLUSION

As geospatial data have stricter data quality and robustness requirements than any other data, it is vital to do the evaluation in terms of positional accuracy as well as topological accuracy. In this paper, we have described and evaluated wavelet based watermarking scheme. We have used Haar wavelet and third level of decomposition as we are getting best results over there. It has been observed that this scheme is giving good results for embedding strength 0.1. As embedding strength increases, positional accuracy as well as topological accuracy gets ruined. Also, this scheme is robust against noise, compression, scaling, translation, addition/deletion of vertices, and cropping attack.

REFERENCES

- Cao, L., Men, C., Li, X. 2010. Iterative embedding-based reversible watermarking for 2D-vector maps. In Proceeding of 17th IEEE International Conference on Image Processing, Hong Kong, pp. 3685-3688.
- Cheng, F., Yin, H., Zhang, X., Zhang, D. 2010. A digital watermarking algorithm for vector map. In Proceedings of International Conference on Challenges in Environmental Science and Computer Engineering, Wuhan, China, pp. 101-103.
- Douglas, D., Peucker, T. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Canadian Cartographer*, Vol. 10(2), pp. 112–122.
- Huang, L., Zhou, W. Jiang, R. Li, A. 2010. Data quality inspection of watermarked GIS vector map. In proceedings of 18th International conference on Geoinformatics, Beijing, China, pp. 1-5.
- Huo, X., Seung, T., Jang, B., Lee, S., Kwon, K. 2010. A watermarking scheme using polyline and polygon characteristic of shapefile. In Proceedings of 3rd International Conference on Intelligent Networks and Intelligent Systems, Shenyang, China, pp. 649-652.
- Kim, J. 2010. Robust vector digital watermarking using angels and a random table. *Advances in Information Sciences and Service Sciences*, Vol. 2(1), pp.79-90.
- Kitamura, I., Kanai, S., Kishinami, T. 2001. Copyright protection of vector map using digital watermarking method based on discrete Fourier transform. In Proceedings of IEEE International Geosciences and Remote Sensing Symposium, Sydney, Vol. 3, pp. 191-193.
- Liang, B., Rong, J., Wang, C. 2011. A Vector maps watermarking algorithm based on DCT domain. *Journal of Photogrammetry and Remote Sensing*, Vol. 38(1), pp. 118-121.
- Lianquan M., Qihong, Y. 2007. A digital map watermarking algorithm based on discrete cosine transform. *Journal of Computer Applications and Software*, Issue 1, pp. 146-148.
- Magalhaes, K., Dahab, R. 2009. SB-RAWVec - A Semi-Blind Watermarking Method for Vector Maps. In Proceedings of IEEE International Conference on Communications, Dresden, Germany, pp. 1-6.
- Tao, S., Dehe, X., Chengming, L., Jianguo, S. 2009. Watermarking GIS data for digital map copyright protection. In Proceeding of 24th International Cartographic Conference, Santiago, Chile, pp.1-9.
- Zhu, C., Yang, C, Wang, Q. 2008. A watermarking algorithm for vector geo-spatial data based on integer wavelet transform. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing, Vol. 37(B4), pp. 15-18.