International Cartographic Association
Association Cartographique Internationale

New Authentication Method for Vector Geographic Data Based on Perceptual Hash

LiMing Zhang [a, *], XinGang Zhang [b]

[a] *Faculty of Geomatics, Lanzhou Jiaotong University, zhanglm8@gmail.com*
[b] *Faculty of Geomatics, Lanzhou Jiaotong University, 1441861764@qq.com*

* Corresponding author

Keywords: Perceptual Hash, Discrete Cosine Transform, Content Authentication, Robustness

Abstract:

In the Internet era, vector data can conveniently be stored, distributed and disseminated. This make it convenience for people to use them, and also make it easier for people to edit or modify them, which can reduce the credibility of Geo-information. Traditionally, authentication is done by using a hash function to generate a digital signature for data authentication. However, this method is very sensitive for change even one bit and it is appropriate for accurate authentication such as text. For geographic data, it may experience lossy compression, filtering distortion, geometric transformation, noise pollution, etc. during transmission and application, but it is consistent with the original data perception. Therefore, the traditional cryptography authentication method is not applicable to the robust authentication of geographic data.

Among various authentication techniques, perceptual hashing is a promising solution. Perceptual hashing is a type of unidirectional mapping of multimedia data sets to perceptual digest sets, that is, a multimedia digital representation with the same perceptual content is uniquely mapped into a piece of digital digest, and satisfies perceptual robustness and security. Since a perceptual hash value is a compact representation of the original content, it can be used for robust content authentication of vector geographic data. The advantage of perceptual hashing algorithms over traditional cryptographic hashing algorithms is that they can tolerate differences in quality and format. The same content is always mapped to the same hash value. This is very effective for robust authentication of geographic data.

In this work, we focus on vector geographic data content authentication by perceptual hash algorithms. In the research of vector geographic data authentication, there are usually some authentication methods based on statistics, rough representation of images and extraction of mutation points based on wavelet transform. However, these methods are more or less sensitive to geometric transformation, poor anti-aggression, high complexity, and poor robustness. In order to avoid the shortcomings of traditional authentication algorithms, this paper proposes a vector geographic data authentication algorithm combining DCT and perceptual hashing. The algorithm has high robustness and security against attack, translation, rotation and two types of collusion attacks. The flow of the algorithm is as follows:

a) Size normalization: The vector geographic data is uniformly resampled into 256×256 images to ensure the uniform length of the hash code generated.

b) Grayscale processing: The original data is converted into a 256-order grayscale image, with the aim of simplifying the calculation while ensuring the accuracy of the data.

c) Discrete cosine transform (DCT): The image is subjected to discrete cosine transform to obtain a 256 × 256 DCT coefficient matrix.

d) Zig-zag scanning: The DCT matrix is characterized by a gradual increase in frequency from the upper left corner to the lower right corner, and the energy related to the image content information is concentrated almost in the low frequency coefficient of the upper left corner of the matrix. Starting from the upper left corner, the matrix is scanned and extracted in a zig-zag line, thereby obtaining a one-dimensional sequence. The overall ranking trend of this one-dimensional series is that the frequency gradually increases and the effective information gradually decreases.

e) Reduce the DCT coefficient matrix: Take the first 2048 bits of the one-dimensional sequence in the previous step and form a 32*64 sub-matrix in order. This step greatly reduces the amount of data stored under the premise of retaining most of the image features.

f) Calculate the matrix mean: Calculate the mean of the coefficients in the 32*64 DCT coefficient submatrix, denoted as m.

g) Matrix binarization: The 32*64 submatrix is traversed in order from left to right and top to bottom. The assignment rules are as follows:

$$h(x)=\begin{cases}1 & if\ \ x\geq m\\ 0 & if\ \ x<m\end{cases}$$

Where x is the DCT coefficient of the element in the 32*64 submatrix, and $h(x)$ is the hash code corresponding to the element. According to the above steps, a 2048-bit hash code can be obtained.

h) Convert hash code to hexadecimal：The 2048-bit binary code is converted to a 256-bit hexadecimal code, which is the final hash code of the image. This code reflects the low frequency information of the data and preserves the image profile features.

i) Calculate Hamming distance：Hamming distance calculation method is used to measure the degree of difference between hash coding of original data and the minimum replacement number of hash coding of attacked data, and to calculate the similarity between them. If the similarity rate is greater than the threshold, it is determined to be unified data, otherwise it is different data.

This paper selects a county-level road data from Southeast China as the original vector geographic data. A common attack method for spatial information data such as point deletion attack, element deletion attack, translation attack, scaling attack, and rotation attack is applied to the data. The results show that the algorithm is robust to point deletion attacks, translation attacks, scaling attacks and rotation attacks, and it is also robust to the collusion attack method of rotation plus translation. Figure 1 shows a schematic diagram of various attack modes. Table 1 shows the Hamming distance and hash code similarity calculation results of the original data and the attacked data.

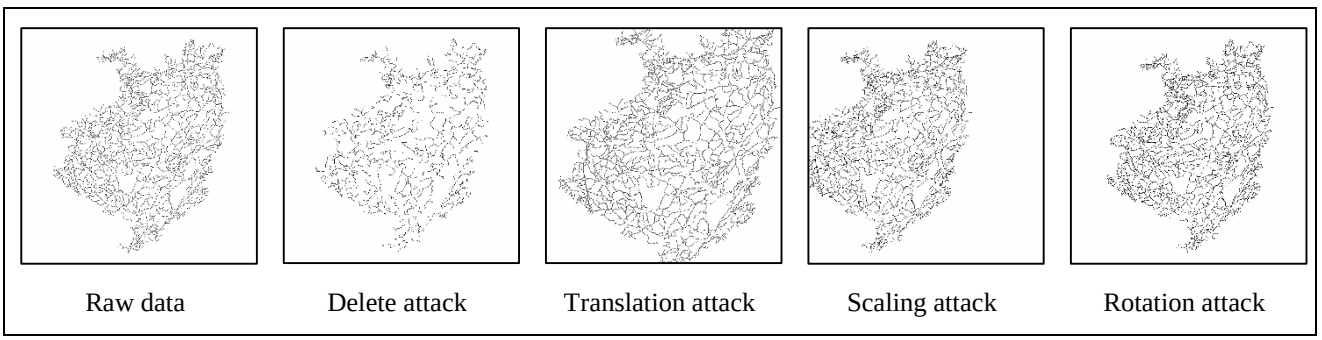

Figure. 1 Attack Modes

Attack Type	point deletion			element deletion			Translation		Scaling		Rotation		Translation &Rotation
Attack Ratio	5%	10%	50%	10%	20%	50%	5%	10%	150%	50%	1°	5°	5%、5°
Hamming Distance	0.021	0.023	0.050	0.250	0.456	0.415	0	0	0	0	0.085	0.122	0.122
Similarity Degree	0.979	0.977	0.950	0.750	0.544	0.585	1	1	1	1	0.915	0.878	0.878

Table. 1 Calculation results of Hamming distance and similarity

29th International Cartographic Conference (ICC 2019), 15–20 July 2019, Tokyo, Japan | https://doi.org/10.5194/ica-abs-1-431-2019