Robust watermarking on copyright protection of digital originals

C Gu and X Y Hu
College of packaging & printing engineering, Tianjin University of Science & Technology, Tianjin, 300222, China
E-mail: guchong527@gmail.com, emma_huxy@yahoo.com.cn

Abstract. The issues about the difference between digital vector originals and raster originals were discussed. A new algorithm based on displacing vertices to realize the embedding and extracting of digital watermarking in vector data was proposed after that. The results showed that the watermark produced by the method is resistant against translation, scaling, rotation, additive random noise; it is also resistant, to some extent, against cropping. This paper also modified the DCT raster image watermarking algorithm, using a bitmap image as watermark embedded into target images, instead of some meaningless serial numbers or simple symbols. The embedding and extraction part of these two digital watermark systems achieved with software. Experiments proved that both algorithms are not only imperceptible, but also have strong resistance against the common attracts, which can prove the copyright more effectively.

1. Introduction
With the rapid development of computers and Internet, different kinds of digital originals have become quite popular. Meanwhile, the copy and transformation of them turned much easy that the traditional copyright protection is not enough any more. How to protect the digital originals’ copyright is a very pressing problem.

As a branch of Information Hiding technology, digital watermarking is a possible approach to counter abuses of digital multimedia data, such as texts, audio data, images, movies. Digital watermarking adds a structure called watermark to the target data object imperceptibly to the users and inseparably from the object, to identify the copyright and protect the legal right of the author.

Watermarking must have strong invisibility and robustness, because of the object here is both commodity and intellectual production. Digital watermark can be many forms, like images, words, numerals, etc. it shouldn’t affect the use of digital productions. However, when the dispute happens, the information encoded in the watermark can be used to identify the copyright owner or to detect tampering.

2. Digital watermarking for different digital originals
Two-dimensional prepress digital originals can be classified into either raster or vector digital data. Raster data represents an image composed by a 2D array of pixels. This kind of files is related to the resolution, so its quality can be easy depredated by rotation, scaling, and other geometrical transformations. Vector data files employ geometrical primitives such as points, lines, and polygons to represent objects using mathematics equations. Unlike the raster files, the vector digital files are able
to be scaled and rotated without loss of quality. Due to the difference between raster data and vector data, used algorithms of watermarking must be different.

However, the watermark used in the copyright protection of both vector and raster originals should meet the following basic requests:

- Imperceptibility: the embedded watermark must be invisible in the target originals;
- Robustness: the watermarking originals can resilient against some common changes: enlarge/shrink, rotated, object insertion/deletion, additive random noise, cropping, and so forth.
- Vindicability: the watermark can be extracted significantly and used for the copyright certification.

3. Watermarking algorithm of vector originals

3.1. Embedding

In vector graphics, almost all entities are defined by some attributes related to coordinates, so it is possible that the watermark can be embedded by the vertex coordinates. Meanwhile, the points, lines and other objects are usually located all over the file, so the embedded watermark information is hidden dispersedly in space; which can reduce the instability caused by the concentration of embedding.

In order to protect the original vertices, all the embedding experiments were realized by drifting coordinates and adding a new attribute. Find the original entities and drift the vertex coordinates first, and then add a unique attribute for each new vertex, so the watermark can be written in it later. The testing file in experiments is a DXF format digital map; figure 1 shows what the map looks like.

![Figure 1. DXF Digital map in experiments.](image)

The watermark embedding process is as follows, also shown in figure 2.

Step1, Search the file structure, find the locations of every stored coordinate. In DXF files, the coordinates are saved after a particular pair of group code and group values;

Step2, Read the $x$, $y$ coordinates (since the discussed objects are 2D vector files, so all the $z$ coordinates of the geometric entities in the data is always 0);

Step3, Reset the locations of the points within the precision allowed to modify, and rewrite the points into the file structure;

Step4, Add copyright watermark into the new coordinates, the watermark in experiment is “TUST Copyright”;

Step5, Set the drift-points to “invisibility” to avoid the visual impact of graphics files.
The amount of embedded watermark is increased with the entities number contained in the vector graphics; this can be seen from table 1. The more entities in the file, the more locations of embedded watermark will be found, and the better dispersion the watermarks have. In this way, the effects of malicious tampering or damaging attacks on the watermark would be much smaller.

### Table 1. Numbers of embedded watermark.

| Entity numbers | Data-density  | Watermark capacity |
|----------------|--------------|--------------------|
|                | (coordinate numbers) | (bit numbers)     |
| 372            | 310          | 57                 |
| 682            | 596          | 57                 |
| 1463           | 1237         | 87                 |
| 2541           | 2074         | 101                |
| 4095           | 3296         | 240                |
| 6482           | 5309         | 360                |
| 8543           | 7018         | 480                |
| 10605          | 8980         | 540                |

### 3.2. Detection

Watermark detection is executed once opening the graphic file in the program. When the program opens a file, data structure will be read line by line, if we find the unique embedded watermark in the file, the password will be asked to enter, if the entered password is correct, watermark information will be displayed; otherwise the file can not be opened. The detection flow is shown in figure 3.
3.3. Experiments and Results

3.3.1. Perceptibility. Because all the drift-point coordinates are added “invisibility” attribute artificially, that means the locations that watermarks written in can not be seen, then embedded watermarks are totally invisible. Therefore, if the files with watermark be opened in graphics software such as AutoCAD, Illustrator, CorelDraw, etc, the watermark will not reduce the visual quality at all.

3.3.2. Resiliency against attacks. The watermark embedding is built on the basis of coordinates, so translation and scaling do not have any influence on the watermark. Other robustness evaluations of the embedded watermark are tested after various attacks, and then use the VC++ program to test. Table 2 shows the results of the experiment, in which the cropping test is detecting the cropping maps down to 1/16~1/2., the rates given in the table is the average rate of different methods.

Table 2. Error rates due to various attacks.

| Error rate (%) | Translation | Scaling | Rotation | Additive random noise | Cropping |
|----------------|-------------|---------|----------|-----------------------|----------|
|                | α=50 cm     | α=80 cm |          | 1/16                  | 1/8      | 1/4 | 1/2 |
| 0              | 0           | 0       | 0        | 2.1                   | 4.9      | 9.1 | 6.4 | 5.7 | 0 |

Experiments showed that the watermarks produced by the method are resilient against shift, scaling and rotation. When random noise having the amplitudes of either α = 50 cm or α = 80 cm is added to the vertex coordinate, the error rate of this algorithm is under 5%; the accuracy of watermark system under cropping is also within the acceptability. Therefore, if drift-point number reaches to a certain quantity, the robustness against cropping is going to be stronger.

4. Watermarking algorithm of raster originals

Raster images are made of many picture elements that we call pixels. When the pixels are small enough, and the distance between two of them short enough, people can not be able to tell. In this way, we can take advantage of human visual redundancy to embed watermarks, or embed watermarks into the position where human eye can not easily detect. The algorithm used in this paper is based on Discrete Cosine Transform (DCT) method of transform domain algorithm.

The host image in experiments is a 480×480 pixel grayscale image of Lena, which saved as tiff format, shown in figure 4, and the watermark image is school badge image of Tianjin university of science & technology, saved as Bmp format, shown in figure 5.

![Figure 4. Host image.](image1)

![Figure 5. Watermark image.](image2)

4.1. Embedding

The embedding starts with the area division step, which tries to produce 8×8 pixel arrays in the host image, since the height of 480 is divisible by 8, there will be 60 arrays along the height of the image and there will be also 60 arrays along the width of the image. Then each array will be transformed with DCT.

According to the study, Human Vision System is relatively sensitive to low frequency DCT coefficients, and high frequency part is fragile to attacks, which will affect the robustness of watermark. Therefore, the watermark should be embedded into the intermediate frequency DCT
coefficients of the host image. The colored part in figure 6 shows the arrays represent the intermediate frequency DCT coefficients.

![Diagram of 8x8 DCT coefficients array](image)

**Figure 6.** An $8 \times 8$ DCT coefficient array.

The level of variance reflects the smoothing degree of the divided arrays; the watermark image should be embedded into the position which has the largest variance here. So we calculate the average gray level $m$ and variance of the arrays $\delta^2$ by equation 1 and equation 2.

\[
m = \frac{1}{n^2} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} x(i, j) \]

\[
\delta^2 = \frac{1}{n^2} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} [x(i, j) - m]^2
\]

Where $n$ represents the number of row pixels, $x$ is the gray value of each pixel. After computing, the binary watermark image was transformed into one-dimensional row vector.

To add random sequence generated from the row vector into the intermediate frequency DCT coefficients of $n$ arrays that have the highest $\delta^2$, DCT coefficients must be calculated for each array, and the arrays need to undergo an Inverse Discrete Cosine Transform (IDCT) in order to be written back as an image, shown in figure 7.

![Watermarked image](image)

**Figure 7.** Watermarked image.

**Figure 8.** Extracted watermark.

4.2. Extraction

Prior to extraction, the host image and the watermarked image should be compared in DCT domain. By means of checking whether the DCT coefficients used to embed the watermark, the watermark can be reconstructed. According to the variance of the arrays, where the DCT is applied to the watermarked image can be determined.

After that, one-dimensional watermark sequence was generated by the order of sequence, as well as the complexity of the arrays, and reconstructs the watermark image by the watermark sequence, and makes it as an identification of the copyright.
4.3. Experiments and results

4.3.1. Perceptibility. We first tested perceptibility of the watermark using 15 volunteers, all non-expert in watermarking. We asked each one of them of she/he could tell a watermarked image if they are presented with a pair of same size images printed on a piece of paper, one watermarked and the other not. None could tell the watermarked image from the non-watermarked image.

In order to observe the image quality of watermarked image objectively, the PSNR (Peak Signal to Noise Ratio) value of the image is calculated using the equation 3 and if the PSNR value is greater than 35dB, the watermarked image is within acceptable degradation levels.

\[
PSNR = 10 \times \log\left(\frac{2^n - 1}{MSE}\right)
\]

Where \( n \) means the number of bits per sample value, the \( MSE \) represents mean square error between the host image and the watermarked image. The PSNR value obtained for this test case was 44.6032 dB. This is well above 35 dB thus implying that the watermarked image is within acceptable degradation levels, and an appropriate image quality of the image that is being watermarked is maintained.

4.3.2. Resiliency against attacks. Watermark’s robustness is based on whether the watermark can be extracted after experiencing a variety of attacks. In this paper, we chose several common attacks to test, such as modified, stained, cropping, format changing, printing and scanning. The similarity value calculated by equation 4 was used for determine the difference between the original watermark and extracted watermarks after attacking.

\[
sim = \frac{W \cdot \hat{W}}{\sqrt{W \cdot \hat{W}}}
\]

Where \( \hat{W} \) represents detected watermark; \( W \) represents original watermark; \( \sim \) means similarity value.

When the similarity value is over 0.7, we can identify that the watermark experiment is successful. Table 3 shows the results of the experiments under various attacks.

According to table 3, the similarity value between watermark without modification and the original watermark was 0.9275, and similarity values of other extracted watermarks from various attacked images are all over 0.8. The results showed that this algorithm can identify the copyright definitely.

| No modification | 7% modified | 12% cropping | 9% stained | reformatting | printing & scanning |
|-----------------|-------------|--------------|------------|--------------|-------------------|
| Similarity      | 0.9275      | 0.8487       | 0.8206     | 0.8096       | 0.9021            | 0.8017            |

5. Conclusion

In this paper, we analyzed the characteristics of graphics and images, studied their data properties, and found two kinds of visual redundancy location, and finally proposed two different watermarking algorithms for both two kinds of file. The experimental results determined the feasibility of the methods that could be the foundation for the further research in this field.

References

[1] Zhao Y 2006 Sequence watermarking robust to random geometrical attack(China: Chinese
[2] Ohbuchi R et al. 1999 *Proc. Int. Conf. on Computer Graphics* (Canada: Canmore) pp170-177

[3] Ohbuchi R, Ueda H and Endoh S 2002 *Proc. Int. Conf. on Multimedia and Expo* (Switzerland: Lausanne) pp1-4

[4] Ohbuchi R, Ueda H and Endoh S 2003 *Proc. Int. Conf. on the Shape Modeling* (Korea: Seoul) pp1-10

[5] Schyndel van R G, Tirkel A Z and Osborne C F 1994 *Proc. Int. Conf. on Image Processing* vol2 pp86-90

[6] Bender W, Gruhl D, Morimoto N 1995 *Techniques for data hiding: Proc. of the SPIE* (CA:San Jose) *Storage and Retrieval for Image and Vedio Database* vol3 pp2420-2425