Adaptive STDM-Based PDF Documents Watermarking Algorithm Robust to Fixed Gain Attack

Jingcai Liu\textsuperscript{1}, Naixi Liu\textsuperscript{1}, Shundong Li\textsuperscript{2} and Daoshun Wang\textsuperscript{1,3, *}

\textsuperscript{1} Computer Science and Technology, Tsinghua University, Beijing 100084, China
\textsuperscript{2} School of Computer Science, Shaanxi Normal University, Xi’an 710062, China
\textsuperscript{3} Tsinghua University, Beijing, China
\textsuperscript{*} Corresponding Author: daoshun@mail.tsinghua.edu.cn

Abstract. Spread Transform Dither Modulation (STDM) has good performance in robustness against Additive White Gaussian Noise (AWGN) and re-quantization, so it has been widely used in image watermarking algorithms. However, STDM cannot resist Fixed Gain Attack (FGA). In the paper, we introduce an adaptive function into the traditional STDM algorithm, so that it can automatically adjust the amount of distortion according to the text line spacing in the document. Then we give an improved adaptive function to make our algorithm robust to FGA. Analysis and experimental results show that the algorithm proposed in this paper has good invisibility and high robustness against AWGN and FGA.

1. Introduction
With the rapid development of mobile Internet and communication technology, piracy and infringement are growing explosively, and the copyright protection of digital products has become an urgent problem to be solved. Digital watermarking technology is an effective method to protect the copyright of digital products. Watermarking can be used to hide digital data such as copyright, ownership or sequence code into text, image, audio and video \cite{1,2}. In case of copyright disputes, the hidden data can be extracted and used to validate its ownership \cite{3}.

The literature \cite{4} proposed a blind digital watermarking algorithm for PDF documents based on traditional Spread Transform Dither Modulation. The algorithm embeds the watermark into the character spacing. The author presents the acceptable distortion of the selected character’s horizontal coordinate, so that the algorithm achieves sufficient robustness under high density noises attacks while preserving sufficient transparency. The literature \cite{5} presented the Component Analysis Resistant-STDM watermarking scheme. Unlike the traditional STDM watermarking, the CAR-STDM uses multiple projection vectors at the encoder and the decoder. The results show that CAR-STDM preserves the robustness against AWGN and Salt & Pepper attacks, and achieves the security against the Blind Source Separation techniques, such as Principal Component Analysis and Independent Component Analysis.

In this paper, we proposed a new blind text watermarking algorithm that is highly robust against AWGN and FGA by introducing an adaptive function into traditional STDM. Section 2 describes the traditional STDM and presents the proposed algorithm in detail. The performance of the proposed approach is discussed in section 3. The conclusion and future work are given in section 4.
2. Proposed Algorithm

2.1. Spread Transform Dither Modulation

STDM is an implementation of the original QIM algorithm proposed by Chen and wornell [6], which introduces the idea of dither modulation and spread spectrum into the QIM system. The watermark bit \( m \) is embedded into the projection of the host signal through dither modulation, that is to say

\[
y = x + (\bar{Q}(x^T v, d_m, \Delta) - x^T v) \cdot v, \quad m \in \{0, 1\}
\]

(1)

Where \( v \) is the pseudo-random projection vector with unit length (or unit energy), \( \bar{Q} \) is the dither quantizer, \( \Delta \) is the quantization step, \( d_m \) is the dither signal, and

\[
\bar{Q}(x, d_m, \Delta) = \text{round}\left(\frac{x + d_m}{\Delta}\right) \cdot \Delta - d_m
\]

(2)

\[
d_m = -\frac{\Delta}{4} + \frac{\Delta}{2} \cdot m
\]

(3)

During transmission, the signal \( y \) will be distorted by various attacks, including common signal processing and malicious attacks. The STDM detector estimates the watermark bit by the quantization point closest to the projection of the distorted signal \( \hat{y} \) [7], thus the estimated watermark bit \( \hat{m} \) is as follow

\[
\hat{m} = \arg\min_{m=\{0,1\}} (\hat{y}^T v - \bar{Q}(\hat{y}^T v, d_m, \Delta))
\]

(4)

2.2. Adaptability

In our algorithm, STDM is used to quantize the line spacing signal of PDF document to embed the watermark bits. The larger text elements have greater signal-to-noise ratio when subjected to the same distortion [8]. It means that larger text line spacing can bear more distortion. Chen [9] points out that when the projection of signal \( x \) on \( v \) is approximately uniformly distributed across the STDM quantization cell, the overall average expected distortion is

\[
D_s = \frac{\Delta^2}{12L}
\]

(5)

Where \( L \) is the length of the non-overlapping sub-signals divided by the host signal. When \( L \) remains unchanged, the overall average expected distortion is proportional to the quantization step. Therefore, we can map the text line spacing to the quantization step, so that the quantization step increases with the increase of the text line spacing. We define this mapping as an adaptive function. The adaptive function adopted by the algorithm in this paper is

\[
\Delta = g(p) = \alpha p^\beta
\]

(6)

Where \( p = x^Tv \), \( \alpha \) denotes the watermarking strength, \( \beta \) is the parameter of the adaptive function, and its solution and proof are shown in Appendix.

2.3. Fixed Gain Attack

As mentioned above, the output signal with watermark will be subjected to various attacks during the transmission, where Fixed Gain Attack (FGA) [10] refers to the signal received by the recipient is multiplied by a scaling factor \( \rho > 0 \), given as

\[
\hat{y} = \rho y
\]

(7)

The STDM-based watermarking algorithm is vulnerable to fixed gain attack. This is because the scaled signal easily deviates from the original quantization cell. In this paper, we optimize the adaptive function to achieve resistance to fixed gain attack, while maintaining our algorithm's adaptability, the adaptive function is as follow
\[ \Delta = \alpha p^\beta \parallel y^{i-1} \parallel_{\alpha \parallel y^{i-1}}^{1-\beta} \]  

(8)

Where \( p = x^T v \) is the projection of the \( i \)th sub-signal \( x \) in the host, \( y^{i-1} \) is the output of the \((i-1)\)th sub-signal in the host, \( \parallel y^{i-1} \parallel_\varepsilon \) denotes 1-\( \varepsilon \)-norm [11], although it cannot be called a norm in the strict sense, given as

\[ \parallel y \parallel_\varepsilon = \left( \frac{1}{L} \sum_{k=1}^{L} |y_k|^{\varepsilon} \right)^{1/\varepsilon}, \quad i = 1, 2, \ldots, H \]  

(9)

Where \( L \) is the length of the non-overlapping sub-signals divided by the host signal, \( H \) is the length of the watermark, \( y^i \) is the output of the \( i \)th sub-signal in the host. In order to simplify the writing form of the formula, the superscript of \( y^i \) is omitted in the later formula derivation.

In the case of only FGA, it can be known from equation (7) that \( \hat{y}^T v = \rho y^T v \), and

\[ \hat{\Delta} = \alpha (\hat{y}^T v)^\beta \parallel \hat{y}^{i-1} \parallel_{\alpha \parallel y^{i-1}}^{1-\beta} = \alpha (\rho y^T v)^\beta \parallel \rho y^{i-1} \parallel_{\alpha \parallel y^{i-1}}^{1-\beta} = \rho \hat{\Delta} \]  

(10)

then

\[ \tilde{m} = \arg\min_{m=0,1} [\hat{y}^T v - \tilde{Q}(\hat{y}^T v, \tilde{d}_m, \hat{\Delta})] \]

\[ = \arg\min_{m=0,1} \left\{ \rho y^T v - \left( r \left( \frac{y^T v}{\Delta} + \frac{2m - 1}{4} \right) - \frac{2m - 1}{4} \right) \cdot \rho \hat{\Delta} \right\} \]  

(11)

As can be seen from equation (11), the watermark bits extracted from the signal \( y \) that has never been attacked and the distorted signal \( \hat{y} \) that has been attacked by FGA are the same, which ensures the accuracy of our algorithm under the fixed gain attack. The following problem is to solve \( \beta \). According to equation (11), the conclusion of equation (24) in Appendix still holds, but

\[ t = \frac{\hat{p}}{\Delta} = \frac{\hat{p}}{\alpha \parallel y^{i-1} \parallel_{\alpha \parallel y^{i-1}}^{1-\beta}} = \frac{\hat{p}^{1-\beta}}{\alpha \parallel y^{i-1} \parallel_{\alpha \parallel y^{i-1}}^{1-\beta}} \]  

(12)

and

\[ \hat{p} = \left( r \left( \frac{p}{\Delta} + \frac{2m - 1}{4} \right) - \frac{2m - 1}{4} \right) \cdot \Delta \]

\[ = \left( r \left( \frac{p^{1-\beta}}{\alpha \parallel y^{i-1} \parallel_{\alpha \parallel y^{i-1}}^{1-\beta}} + \frac{2m - 1}{4} \right) - \frac{2m - 1}{4} \right) \cdot \alpha \parallel y^{i-1} \parallel_{\alpha \parallel y^{i-1}}^{1-\beta} \]  

(13)

thus

\[ t = \left( r \left( \frac{p^{1-\beta}}{\alpha \parallel y^{i-1} \parallel_{\alpha \parallel y^{i-1}}^{1-\beta}} + \frac{2m - 1}{4} \right) - \frac{2m - 1}{4} \right)^{1-\beta} \cdot \alpha \parallel y^{i-1} \parallel_{\alpha \parallel y^{i-1}}^{1-\beta} \]  

(14)

Let \( s = \frac{p^{1-\beta}}{\alpha \parallel y^{i-1} \parallel_{\alpha \parallel y^{i-1}}^{1-\beta}} \), then equation (14) can be simplified as

\[ t = s^\beta \left( r \left( s + \frac{2m - 1}{4} \right) - \frac{2m - 1}{4} \right)^{1-\beta} \]  

(15)
The equation (15) and (27) are identical, thus the conclusion of $\beta$ in Appendix is also applicable in the case of FGA.

2.4. Watermark Embedding and Extraction

Assuming that the watermark bit sequence is $m = \{m_1, m_2, ..., m_H\}$, the watermark embedding algorithm is as follow:

1) $Key_p$ is used to scramble the line spacing signal $x = \{x_1, x_2, ..., x_N\}$ retrieved by the Parser, and the output signal $x_p = \{x_{p1}, x_{p2}, ..., x_{pN}\}$ is obtained.

2) The signal $x_p$ is divided into several sub-signals of length $L$. The watermark bit $m_i$ is embedded in sub-signal $x_p^i$, where $i = 1, 2, ..., H$, and $N \geq HL$.

3) The sub-signal $x_p^i$ is projected along the projection vector $Key_v$. Then the projection is sent to the adaptive function to evaluate the quantization step and the output signal $y_p^i$ is obtained by equation (1). Finally, the outputs of all the sub-signals are synthesized into the signal $y_p$. To further enhance the security of the watermarking system, the scrambling operation can be performed on the sub-signal $x_p^i$ again.

4) $Key_p$ is used again to perform inverse scrambling operation on the quantized signal $y_p$, and the encoded signal $y$ is output.

When the watermarked PDF document is transmitted through the channel, the recipient gets the distorted PDF document, the watermark extraction algorithm is as follow:

1) The same $Key_p$ in the embedding algorithm is used to scramble the distorted line spacing signal $\hat{y}$ retrieved by the Parser, and the output signal $\hat{y}_p$ is obtained.

2) The signal $\hat{y}_p$ is divided into several sub-signals of length $L$. The projection of the sub-signal $\hat{y}_p^i$ along the protection vector $Key_v$ is sent to the adaptive function to evaluate the quantization step. Then the watermark bit $\hat{m}_i$ is detected from the sub-signal $\hat{y}_p^i$ by equation (4).

3) The watermark bits detected in all sub-signals are synthesized to achieve the extracted watermark $\hat{m}$.

3. Experimental Results and Performance Analysis

Text watermarking algorithms generally use criteria such as imperceptibility, robustness, embedding capacity to evaluate the performance of the algorithm, and these criteria are mutually restrictive. Imperceptibility, that is, invisibility or transparency, requires that the modification of the original text document by the algorithm cannot significantly degrade the display quality of the document [12]. In practice, the imperceptibility is analyzed by comparing the difference between the original text document and the watermarked text document [13]. In this paper, MSE (Mean Square Error) and MXAE (MaXimum Absolute Error) are used to evaluate the imperceptibility of the text watermarking algorithms, which are defined as

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (X_i - \hat{X}_i)^2$$

$$MXAE = \max_{i} |X_i - \hat{X}_i|$$

Where $X$ is the line spacing signal of the original text document, $\hat{X}$ is the line spacing signal of the watermarked text document, $N$ is the length of the line spacing signal. MSE represents the average distortion introduced by the algorithm to the line spacing of the document, while MXAE represents the maximum distortion introduced by the algorithm.

Robustness means the ability of the algorithm to resist the attacks such as retrieving, modifying, destroying and removing the original watermark [14,15]. Generally, Bit Error Rate (BER) is used to quantitatively analyze the robustness of the algorithm, which is defined as
BER = \frac{1}{H} \sum_{i=1}^{H} |W_i - \hat{W}_i| \quad (17)

Where \( W \) is the original watermark, \( \hat{W} \) is the extracted watermark, and \( H \) is the length of the watermark.

To ensure the fairness, we compared the proposed watermarking algorithms STDM-SA and STDM-SF with STDM [4] and RDM [11] under the same conditions. The parameter \( \beta \) of the adaptive function is 0.25. The watermark and projection vector are generated stochastically. For the length of the line spacing signal is limited, we assume that \( L=31, H=11, x=22 \), then repeat the calculation of the BER 500 times, and take the average value.

From figure 1, if the maximum distortion is the same, that is, \( m_{xae} = 0.4 \), the algorithms STDM and STDM-SA are highly robust to AWGN, and STDM-SA is slightly better, but quite vulnerable to FGA. The algorithm RDM has good robustness to FGA, and the watermark capacity is much higher than other algorithms, but the robustness to AWGN is very poor. The algorithm STDM-SF is robust to both FGA and AWGN. The reason that the robustness of STDM-SF to AWGN is weaker than that of STDM and STDM-SA is that the average distortion of STDM-SF is smaller than that of STDM and STDM-SA under the same maximum distortion as show in figure 2, and with the increase of the maximum distortion (watermarking strength), the robustness of the STDM-based algorithm to AWGN tends to be perfect.

Figure 2. Analysis of the robustness to AWGN.
4. Conclusion
In this paper, we have presented the adaptive STDM-based PDF documents watermarking algorithm, which is highly robust against FGA and AWGN by introducing an adaptive function into the traditional STDM. Experimental results show that the proposed scheme has considerable advantage in robustness against FGA and AWGN compared with the traditional STDM and RDM. However, the embedding capacity is relatively low, for the length of the line spacing signal in the document is limited. Further improvements such as higher watermark capacity and resistance to print-scan attack will be investigated in our future work.

5. Acknowledgements
This work was supported in part by the National Natural Science Foundation of China (No. 61972225, 61902164), China Mobile Research Fund Project (No. MCM20170407), the Key Laboratory of Digital Content Anti-Counterfeiting and Security, Forensics of State Administration of Press, Publication, Radio, Film and Television (SAPPRFT).

6. Appendix
According to the section 2.2, the adaptive function of the PDF document watermarking algorithm based on STDM is

$$\Delta = g(p) = ap^\beta$$

(18)

Where $p = x^Tv$, $\alpha$ denotes the watermarking strength, $\beta$ is the parameter to be determined. Both $\alpha$ and $\beta$ are greater than 0.

Without loss of generality, we assume that the host signal $x > 0$, that is, all elements of $x$ are greater than 0, and the signal $y > 0$, the projection vector $v > 0$, then $p > 0$. Let $\hat{p} = y^Tv > 0$, then

$$\hat{p} = (x + (\bar{Q}(x^Tv, d_m, \Delta) - x^Tv) \cdot v)^T \cdot v = \bar{Q}(p, d_m, \Delta)$$

(19)

Without any attack, we have $\hat{y} = y$, then equation (4) can be rewritten as

$$\hat{m} = \arg\min_{m = \{0, 1\}} |\hat{p} - (r(\frac{\hat{p}}{\Delta} + \frac{2m - 1}{4}) - \frac{2m - 1}{4}) \cdot \Delta|$$

(20)

Where $r$ denotes round function, and $\hat{\Delta} = \alpha \hat{p}^\beta$, then

$$dist_0 = |\hat{p} - (r(\frac{\hat{p}}{\Delta} - \frac{1}{4}) + \frac{1}{4}) \cdot \hat{\Delta}|$$

(21)

$$dist_1 = |\hat{p} - (r(\frac{\hat{p}}{\Delta} + \frac{1}{4}) - \frac{1}{4}) \cdot \hat{\Delta}|$$

(22)

Thus

$$dist^2_0 - dist^2_1 = \Delta^2 \left( r(\frac{\hat{p}}{\Delta} - \frac{1}{4}) - r\left(\frac{\hat{p}}{\Delta} + \frac{1}{4}\right) + \frac{1}{2}\right) \left( r\left(\frac{\hat{p}}{\Delta} - \frac{1}{4}\right) + r\left(\frac{\hat{p}}{\Delta} + \frac{1}{4}\right) - 2\frac{\hat{p}}{\Delta}\right)$$

(23)

Let $t = \hat{p}/\hat{\Delta} > 0$, then equation (23) can be simplified as

$$dist^2_0 - dist^2_1 = \Delta^2 \left( r(t - \frac{1}{4}) - r\left(t + \frac{1}{4}\right) + \frac{1}{2}\right) \left( r\left(t - \frac{1}{4}\right) + r\left(t + \frac{1}{4}\right) - 2t\right)$$

(24)

It can be proved that the function $h(t) = (r(t - 1/4) - r(t + 1/4) + 1/2)(r(t - 1/4) + r(t + 1/4) - 2t)$ is a periodic function with a period of one in $[0, +\infty]$. If $n < t < n + 0.5$, then $dist^2_0 - dist^2_1 < 0$, i.e. $dist_0 < dist_1$, and if $n + 0.5 < t < n + 1$, then $dist^2_0 - dist^2_1 > 0$, i.e. $dist_0 > dist_1$, where $n \in \{0, 1, 2, ...\}$.

From equation (19), we have
\[
\hat{p} = \left( r \left( \frac{p}{\Delta} + \frac{2m - 1}{4} \right) - \frac{2m - 1}{4} \right) \cdot \Delta = \left( r \left( \frac{p^{1-\beta}}{\alpha} + \frac{2m - 1}{4} \right) - \frac{2m - 1}{4} \right) \cdot \alpha p^\beta
\] (25)

Thus
\[
t = \frac{\hat{p}^{1-\beta}}{\alpha} = \left( \frac{p^{1-\beta}}{\alpha} \right)^{\beta} \cdot \left( r \left( \frac{p^{1-\beta}}{\alpha} + \frac{2m - 1}{4} \right) - \frac{2m - 1}{4} \right)^{1-\beta}
\] (26)

Let \( s = p^{1-\beta}/\alpha \), then equation (26) can be simplified as
\[
t = s^\beta \left( r \left( s + \frac{2m - 1}{4} \right) - \frac{2m - 1}{4} \right)^{1-\beta}
\] (27)

If the watermark bit embedded in the host signal \( x \) is 0, then
\[
t = s^\beta \left( r \left( s - \frac{1}{4} \right) + \frac{1}{4} \right)^{1-\beta}
\] (28)

It can be proved that if \( n < s < n + 3/4 \), then
\[
n^\beta \left( n + \frac{1}{4} \right)^{1-\beta} < t < \left( n + \frac{3}{4} \right)^\beta \left( n + \frac{1}{4} \right)^{1-\beta}
\] (29)

And if \( n + 3/4 < s < n + 1 \), then
\[
\left( n + \frac{3}{4} \right)^\beta \left( n + \frac{5}{4} \right)^{1-\beta} < t < \left( n + 1 \right)^\beta \left( n + \frac{5}{4} \right)^{1-\beta}
\] (30)

As known, the STDM detector is a minimum distance detector, in order to extract the watermark bit 0 from the host signal \( x \), it must be ensured that \( \text{dist}_0 < \text{dist}_1 \), that is, \( n < t < n + 0.5 \). If \( \beta = 1 \), then \( t = 1/\alpha \), \( \alpha \) is an adjustable parameter, hence \( t \) may fall outside \([n, n + 0.5]\), and the detected watermark bit is 1, resulting in an error in the detected watermark. If \( \beta > 1 \), then the left expression of inequality (29) is less than \( n \), and the right expression is greater than \( n + 0.5 \), \( t \) may also fall outside \([n, n + 0.5]\). If \( 0 < \beta < 1 \), then the left expression of inequality (29) is always in \((n, n + 0.5)\), and the right expression of inequality (30) is always in \((n + 1, n + 1.5)\), so there must be
\[
\left( n + \frac{3}{4} \right)^\beta \left( n + \frac{1}{4} \right)^{1-\beta} < n + \frac{1}{2}
\] (31)
\[
\left( n + \frac{3}{4} \right)^\beta \left( n + \frac{5}{4} \right)^{1-\beta} > n + 1
\] (32)

Similarly, if the watermark bit embedded in the host signal \( x \) is 1, there must be
\[
\left( n + \frac{1}{4} \right)^\beta \left( n - \frac{1}{4} \right)^{1-\beta} < n
\] (33)
\[
\left( n + \frac{1}{4} \right)^\beta \left( n + \frac{3}{4} \right)^{1-\beta} > n + \frac{1}{2}
\] (34)

From the above analysis, the parameter \( \beta \) of the adaptive function is the value that simultaneously satisfy the inequalities (31), (32), (33) and (34). Take solving inequality (34) as an example, let
\[
f(\beta) = \left( n + \frac{1}{4} \right)^\beta \left( n + \frac{3}{4} \right)^{1-\beta} - \left( n + \frac{1}{2} \right)
\] (35)

Then \( f(0) = 1/4 > 0 \), \( f(1) = -1/4 < 0 \). Since \( f'(\beta) = (\ln(n + 1/4) - \ln(n + 3/4))(f(\beta) + n + 1/2) < 0 \), so \( f(\beta) \) decreases monotonically in \([0,1]\), and has a unique zero point
that is, \( f(b) = 0 \), then

\[
b = \frac{\ln \left( n + \frac{1}{2} \right) - \ln \left( n + \frac{3}{4} \right)}{\ln \left( n + \frac{1}{4} \right) - \ln \left( n + \frac{3}{4} \right)}
\] (36)

To make the inequality (34) true, then \( f(\beta) > 0 \), that is, \( 0 < \beta < b_{\min} = 0.3690 \). Similarly, the solutions of inequalities (31), (32) and (33) are \( 0 < \beta < 0.5 \), \( 0 < \beta < 0.4368 \), \( 0 < \beta < 0.5 \) respectively. In conclusion, \( 0 < \beta < 0.3690 \).

7. References

[1] Jane O, Elbaşi E and Ilk H 2014 Journal of Applied Research and Technology 12 750-61
[2] Kamaruddin N, Kamsin A, Por L and Rahman H 2018 IEEE Access 6 8011-28
[3] Chaubey P and Singhadia A 2015 International Journal of Scientific Research & Engineering Trends 1 112-5
[4] Bitar A, Darazi R, Couchot J and Couturier R 2017 Multimedia Tools and Applications 76 143-61
[5] Hatoum M, Darazi R and Couchot J 2018 Proceedings of the 15th International Joint Conference on e-Business and Telecommunications (Porto) pp 420-7
[6] Chen B and Wornell G 2001 IEEE Trans. Inf. Theory 47 1423-43
[7] Li X 2011 Digital Watermarking Based on Spread Transform Dither Modulation (Shandong: Shandong University) pp 17-9
[8] Brassil J, Low S and Maxemchuk N 1999 Proceedings of the IEEE 87 1181-96
[9] Chen B 2000 Design and Analysis of Digital Watermarking, Information Embedding, and Data Hiding Systems (Cambridge: Massachusetts Institute of Technology) pp 61-7
[10] Hatoum M, Darazi R and Couchot J 2020 Multimed Tools Appl 79 1887-919
[11] Pérez-González F, Barni M, Abrardo A and Mosquera C 2004 IEEE 6th Workshop on Multimedia Signal Processing number I (Siena: IEEE) pp 139-42
[12] Alkawaz M, Sulong G, Saba T, Almazyad A and Rehman A 2016 Security and Communication Networks 9 6365-78
[13] Taleby Ahvanooye M, Dana Mazraeh H and Tabasi S 2016 Information Security Journal 25 191-6
[14] Matheson L, Mitchell S, Shamoon T, Tarjan R and Zane F 1998 Financial Cryptography (Berlin: Springer) pp 227-40
[15] Potdar V, Song H and Elizabeth C 2005 2005 3rd IEEE International Conference on Industrial Informatics (Perth: IEEE) pp 709-16