



King Saud University
**Journal of King Saud University –
Computer and Information Sciences**

www.ksu.edu.sa
www.sciencedirect.com



Improved capacity Arabic text watermarking methods based on open word space

Reem A. Alotaibi^{a,*}, Lamiaa A. Elrefaei^{a,b,1}

^a *Computer Science Department, Faculty of Computing & Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia*

^b *Electrical Engineering Department, Faculty of Engineering, Shoubra Benha University, Cairo, Egypt*

Received 3 September 2016; revised 10 December 2016; accepted 28 December 2016

KEYWORDS

Arabic text watermarking;
Capacity;
Robustness;
Imperceptibility

Abstract Digital watermarking is used to protect text copyright and to detect unauthorized use. In this paper, two invisible blind watermarking methods for Arabic text are proposed. Since the pseudo-space is very small space used to force the connected characters to be isolated, it is added to the word space to hide binary bit “0” or “1”. In the first proposed method, the pseudo-space is inserted before and after normal word space based on dotting feature in Arabic text. The second proposed method inserts the pseudo-space and other three small or zero width spaces to increase the capacity, where the presence of them indicates bit “1” and the absence indicates bit “0”. The comparative results obtained by testing the proposed methods with some of existing watermarking methods using variable size text samples with different watermark lengths. The experiments show that the proposed methods have the highest capacity and higher imperceptibility than other watermarking techniques from the literature. The robustness of the proposed methods is tested under most of possible text attacks. They are robust against electronic text attacks such as: copying and pasting, text formatting and text tampering for tampering ratio up to 84%.

© 2017 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Digital watermarking techniques have been widely emerged in order to effectively protect multimedia copyright. Digital watermarking refers to the embedding process that inserts a watermark (i.e. Label, signature, or copyright) into several types of media. These types of media include text, audio, image, and video. The watermarking system involves two main processes: embedding of the watermark into the original data and extracting the watermark from watermarked data or attacked watermarked data. When designing a watermarking system, some of the basic requirements must be taken into account which vary depending on the use of the system (Stanković et al., 2012; Cox et al., 2007). The key watermarking

* Corresponding author.

E-mail addresses: reem.a.safran@gmail.com (R.A. Alotaibi), laelrefaei@kau.edu.sa, lamia.alrefaei@feng.bu.edu.eg (L.A. Elrefaei).

¹ Postal address: King Abdulaziz University, Female Campus, Room No. s109, Building No. 61, P.O Box 80221, Jeddah 21589, Saudi Arabia.

Peer review under responsibility of King Saud University.



Production and hosting by Elsevier

Table 1 The chosen spaces.

| Space name | Space unicode | Description | Example |
|------------------------|---------------|--------------------------------------------|------------|
| Normal space (NS) | U+0020 | Normal space used to separate words | سبحان الله |
| Pseudo-space (PS) | U+200C | Force the connected letters to be separate | سبحان الله |
| Thin space (TS) | U+2009 | 1/5 (sometimes 1/6) of an em wide | سبحان الله |
| Hair space (HS) | U+200A | Thinner than a thin space | سبحان الله |
| Zero width space (ZWS) | U+200B | No width | سبحانالله |

requirements are: capacity, imperceptibility, robustness and security. Capacity is the total number of hidden bits in an object. Imperceptibility is used to measure the difference between the original and watermarked object by noting any addition to the original object. Robustness is the ability to extract or detect the watermark after the watermarked object has been attacked. Security requirement is the difficulty of extracting the watermark without the destruction of the watermarked object.

Creating a hidden watermark in the text is the hardest kind due to the relative lack of unnecessary information within a text file compared to an image or audio files. The human sensitivity to text changes is higher than the sensitivity to other multimedia. Any text change must reserve the meaning, fluently, writing style and text value (Topkara et al., 2006; Jalil, 2010).

Text watermarking methods can be classified into: line-shift coding, word-shift coding, linguistic methods, open space method and other methods based on the language characteristic (Alotaibi and Elrefaei, 2015). In line-shift coding, the lines are shifted up or down to hide "0" or "1". In word-shift coding the words are moved horizontally to code secret bits. Linguistic methods aim to change the text structure (syntactic approach) or text content (semantic approach) (Bennett, 2004). Open space method depends on the exploitation of the white spaces by adding a space or more between words, sentences or at the end of lines to indicate on the existence of hidden bits (Bender et al., 1996).

The contributions of this paper are improving the embedding capacity by proposing two Arabic text watermarking methods, Method 1 and Method 2 and testing their robustness. The proposed methods utilize the open space between words but instead of using normal space as in Bender et al. (1996), small spaces or no width ones are used. Arabic letters do not take one form, but their shape varies depending on its location in the word (Alotaibi and Elrefaei, 2015). Pseudo-space is a non printing character when it comes before or after the letter, it forces the letter to take the final or initial form. Using it at the beginning or ending of the word does not change the word shape. The researchers in Alotaibi and Elrefaei (2016) inserted it before the normal space to indicate hidden data. In the proposed Method 1, Pseudo-space is added before and after the regular space to provide capacity as twice as the method presented in Alotaibi and Elrefaei (2016). In the proposed Method 2, four spaces: Pseudo-space, Thin space, Hair space and Zero width space, were chosen to be added to the normal space to give a very large capacity. The existence of these spaces is used to hide bit "1", and the absence of them is used to hide bit "0".

1.1. Overview of the used spaces

The proposed methods have adopted to embed the watermark in the Arabic text by inserting some spaces to the normal space. The chosen spaces are shown in Table 1 with their Unicode's (Whitespace character, 2016). Table 1 also shows examples of using these spaces between two words "سبحان" and "الله" to show how wide are they. The vertical pointer indicates where the Unicode space is inserted. Table 2 shows the used spaces in methods (Bender et al., 1996; Alotaibi and Elrefaei, 2016) and in the proposed watermarking methods.

The rest of this paper is organized as follows: Section 2 reviews the related works in Arabic text watermarking. The proposed methods are discussed in Section 3. Experiments are conducted to evaluate the capacity and imperceptibility of the proposed methods and to compare it with five of the Arabic text watermarking methods in sections 4.1 and 4.2. Robustness of the proposed methods is measured under most of the existing known text attacks in section 4.3. Section 5 concludes the paper.

2. Related work

Text watermarking methods in any language exploit the characteristics of the writing in that language or the general text characteristics. Arabic script has many characteristics such as: open space, kashida, diacritics and dotting.

Open space is a general characteristic used for data hiding in the host text as in the research work presented in Bender et al. (1996) by exploitation of the white spaces in the text document. The manipulation of white spaces done in three different ways: inter-sentence spacing, end-of-line spaces and inter-word spacing. Inter-sentence spacing is encoding secret information in form of binary string on the text based on spaces between sentences. It encodes a "0" by inserting one space or

Table 2 The used spaces.

| Watermarking method | The inserted spaces | Location of insertion |
|---------------------------------------------------|---------------------|-----------------------------------|
| Open space methods (Bender et al., 1996) | NS | Between words, lines or sentences |
| Pseudo-space method (Alotaibi and Elrefaei, 2016) | PS | Between words |
| Proposed Method 1 | PS | Between words |
| Proposed Method 2 | PS, HS, TS and ZWS | Between words |

Download English Version:

<https://daneshyari.com/en/article/6899061>

Download Persian Version:

<https://daneshyari.com/article/6899061>

[Daneshyari.com](https://daneshyari.com)