An Enhanced Kashida-Based Watermarking Approach for Increased Protection in Arabic Text-Documents Based on Frequency Recurrence of Characters

Yasser M. Alginahi^{1,2*}, Muhammad N. Kabir³, Omar Tayan^{2,4}

¹ Dept. of Computer Science, Taibah University, Madinah, Saudi Arabia.

² IT Research Center for the Holy Quran and Its Sciences (NOOR), Taibah University, Madinah, Saudi Arabia.

³ Faculty of Computer Systems and Software Engineering, University Malaysia Pahang, 26300 Gambang, Pahang, Malaysia.

⁴ College of Computer Science and Engineering, Taibah University, Madinah, Saudi Arabia.

* Corresponding author. Tel.: +966540367388; email: alginahi@gmail.com, yginahi@taibahu.edu.sa

Manuscript submitted April 10, 2014; accepted September 29, 2014.

doi: 10.17706/ijcee.2014.v6.857

Abstract: With text being the predominant communication medium on the internet, more attention is required to secure and protect text information. In this work, an invisible watermarking technique based on Kashida-marks is proposed. The watermarking key is predefined whereby a Kashida is placed for a bit 1 and omitted for a bit 0. Kashidas are inserted in the text before a specific list of characters until the entire key is embedded. Two variations to the proposed method were developed, based on the frequency-recurrence properties of the characters. An advantage with the use of frequency recurrence statistics of Arabic characters was evident in this paper since it had enabled the dynamic variation of imperceptibility and robustness levels as required for a target application. The proposed methods proved to achieve the goal of document protection and authenticity with enhanced robustness and improved perceptual similarity with the original cover-text.

Key words: Arabic, text watermarking, Kashida, feature coding.

1. Introduction

The availability and distribution of digital text formats on the Internet in the form of websites, articles, on-line magazines, e-books, news, emails, chats, etc., made it easy to copy, tamper, plagiarize, sabotage, forge and reproduce text compared to other types of media. Online text documents have seen an exponential increase in use as compared to other types of multimedia since the invention of the Internet. However, not much attention has been given to the authentication and copyright protection of text. Therefore, the need to protect copyrights provided researchers with a new track of research, i.e., to produce watermarking techniques in order to protect such information. The research in this area started in 1991 and a number of text based watermarking methods have been proposed since then [1]. However, to-date, the research in specific areas of this field is far from satisfying the needs for such applications. For example, text watermarking techniques based on the syntactic approach of text and Natural Language Processing (NLP) algorithms are progressing slowly, as stated in [1], "NLP is an immature area of research so far and using in-efficient algorithms, efficient results in text watermarking cannot be obtained."

A digital watermark is a kind of marker embedded in media such as audio, text or image, in order to

identify ownership of the copyright of such media. Digital watermarking of any media is considered a branch of steganography and its main objective is to provide copyright protection for intellectual property and prevent illegal copying and diffusing, as illustrated in Fig. 1.

Fig. 1. Key parameters for watermarking and steganography.

Applications of digital watermarking techniques include: copy protection, copyright protection, source tracking, automatic monitoring and tracking of copyright material on the web, fingerprinting and content augmentation applications [2]. Both steganography and digital watermarking employ steganographic techniques to embed data; hence steganography aims for imperceptibility to human senses and digital watermarking tries to control the robustness as top priority. Since a digital copy of data is the same as the original, digital watermarking is considered a passive protection tool. It marks data, but does not degrade it nor controls access to the data.

Watermarking techniques go through several stages, starting from the generation and embedding of watermarks then the publishing stage (distribution and exposure to attacks) and finally, detection of watermarks as a mean of copyright protection. Fig. 2 shows the stages which digital watermarking goes through.

Fig. 2. Stages of digital watermarking.

There are two main forms of watermarking: visible and invisible [2]. Visible watermarking is usually in the form of logos, images or text which is used to identify the ownership of the media (text, videos, and images). In contrast, invisible watermarking is in the form of embedding data that is undetectable. With visible watermarking, the following parameters are very important: deterrence against theft, diminish commercial value without utility, discourage unauthorized duplication and identify the source. In invisible watermarking, the following parameters are very important: validation of intended recipient, non-reputable transmission, diminish commercial value without utility, and digital notarization and authentication.

However, the invisible watermarking is considered more robust. Therefore, it is preferable that text watermarking should be easy to implement, imperceptible, robust, and adaptable to different text formats, have high information carrying capacity and should be effectively applied to print/digital proof. In other words, watermarking techniques provide security in a digital data by making imperceptible modifications in original document, which can be identified by a certain key through a certifying authority [2]. The benefits of digital watermarking are to confirm property ownership, to follow up of unauthorized copies, to verify validate and identify, to label documents, to control usage and protect content [3]. Finally, the continual exponential increase of digital information on the Internet presents a continual increase in the need to protect information.

In this work, we focus on confirming authenticity and integrity-robustness of sensitive text content whose primary motive may compromise on the need for secrecy in the communications channel during transmission. The motive here is that it may be required or even desirable that particular sensitive content should be freely propagated via multiple publishers/servers for wider outreach and dissemination. Hence, the well-understood relation between the client(s) and publishers/server now differs from the common one-to-one relation as in e-commerce transactions that had typically involved hashing or encryption algorithms being distributed between two or more known parties.

In this work, invisible watermarking Kashida-based approaches using character frequency recurrence properties are proposed for Arabic text documents. The proposed techniques explore all possible positions where Kashida could be placed before specific Arabic letters which always appear in Arabic scripts and are always connected to other characters from the right. This paper is organized as follows: Section 2; presents the characteristics of Arabic text; Section 3 introduces the related work; Section 4 explains the proposed enhanced Arabic Kashida-based watermarking; Section 5 provides the results and discussion; and finally, Section 6 concludes the paper with future perspectives.

2. Background: Characteristics of Arabic Scripts

The Arabic Alphabet consists of 28 characters and has many characteristics [4]-[5]. Arabic script possesses a cursive text even when printed and the letters are connected from the baseline of the word. It is written from right to left with the exception of numbers. Its letters change their shape depending on their position in the word. A single character can contain from one to four shapes for each character or ligature, depending on the implementation. The four possible shapes are: isolated; in which case the character is not linked to either the preceding or the following character, final; in which case the character is linked to the preceding character, but not to the following one, initial; in which case the character is linked to the following character but not to the preceding one, and finally, medial; in which case the character is linked to both the preceding and following characters [4]-[5].

From the Arabic alphabets in Table 1, six letters from the Arabic alphabet can only be connected from the right side (initial form) these are: ا، د، و، ز، ر، ذ. The appearance of any of these letters in the middle of the word form one or more sub-words, meaning they form more than one connected components in a single word, these sub-words may consist of one or more characters. Therefore, the shape of a character depends on the context. Examples of words containing more than one connected components include: (حراق), (حامد) and (رواف). Examples of single connected words include: عمر، محمد، مكة [4]-[5].

Arabic text contains diacritics which are marks written below or above text, in this work the text is assumed to contain no diacritics since most of the text found on the internet do not use diacritics. Several Arabic alphabet letters share the same shape, and are differentiated only in terms of the number and placement of dots on the letters. These dots may be referred to as desenders, if placed below the letters, or ascenders, if placed above the letter [5].

Table 1. Different Shapes of Arabic Characters [4]

Name	Isolated	Initial	Medial	Final
alif	ا			ا
baa	ب	ب	ب	ب
Taa	ت	ت	ت	ت
thaa	ث	ث	ث	ث
Jiim	ج	ج	ج	ج
haa	ح	ح	ح	ح
khaa	خ	خ	خ	خ
daal	د			د
dhaal	ذ			ذ
Raa	ر			ر
zaay	ز			ز
siin	س	س	س	س
shiin	ش	ش	ش	ش
saad	ص	ص	ص	ص
daad	ض	ض	ض	ض
Taa	ط	ط	ط	ط
dhaa	ظ	ظ	ظ	ظ
Ayn	ع	ع	ع	ع
ghayn	غ	غ	غ	غ
Faa	ف	ف	ف	ف
qaaf	ق	ق	ق	ق
kaaf	ك	ك	ك	ك
laam	ل	ل	ل	ل
miim	م	م	م	م
nuun	ن	ن	ن	ن
haa	ه	ه	ه	ه
waaw	و			و
Yaa	ي	ي	ي	ي

Another interesting property can be found in Arabic text documents, evident in the relative recurrence patterns of each character over many sample documents. Fig. 3 ranks the average recurrence rates over four sample documents to be used in further experimentation in this paper, arranged from highest to lowest frequency of recurrence. Clearly, it is seen that in the case of the four sample documents the dominant recurring characters can be exploited effectively for data-embedding/data-hiding purposes in this paper.

Fig. 3. Average frequency of Arabic characters for four test documents.

3. Related Work

The work in digital watermarking has seen an increase in the last two decades. Researchers in this area developed many watermarking techniques for different multimedia content. In this section, related work to digital text watermarking techniques will be presented. From surveying the literature, the work developed on digital text watermarking can be classified into two main categories: linguistic coding and formatting/appearance coding. The linguistic coding includes: syntactic and semantic, and the formatting/appearance coding includes: image-based, character/dot positioning, diacritics and Kashida. In addition to these main categories other approaches use combined text and image watermarking such as the work in [3].

Many watermarking approaches have been developed and used in watermarking text documents. In syntactic methods, the syntactic structure of the text is used to embed the watermarks [6]. The use of syntactic watermarking is limited since it can be easily attacked, resulting in the removal of the watermark. The semantic structure of text is used to embed the watermark; this is done by exploiting the text content to insert watermarks into the text techniques include: Synonym substitution [7]; techniques based on nouns and verbs [8], techniques based on text meaning representation [9] and techniques based on presuppositions which is a technique whereby the structure meaning and rearrangements are detected to embed watermark bits [10]. The major advantage of these methods is the protection of information in case of retyping or using OCR programs which is not the case for syntactic methods. However, the semantic approach may alter the meaning of the text and therefore may not be applicable to use on documents where the synonyms of words could provide a different meaning to the text, such as in poetry, English literature, religious books, ... etc.[11].

In formatting/appearance coding techniques, image-based techniques include methods based on text line shifting, word shifting [12] and character/feature coding [13]. In general, techniques based on shifting include adding bits or shifting bit positions resulting in shifting lines, words, paragraphs, adding space between characters or words ... etc.,[14]. However, techniques based on character/feature coding are applied on selected characters by providing slight modifications to the characters, in which such alterations include a change to an individual character's height or its position relative to other characters [13].

Other approaches are based on Kashida-markings, for instance [15] present a method useful for watermarking Arabic and Semitic scripts (used in languages such as Arabic, Urdu, Persian ...etc.). Such an approach is classified as feature coding which is implemented by exploiting the existence of the Kashida. The technique proposed uses pointed letters with a Kashida to hold bit 1 and un-pointed letters with Kashida to hold bit zero. This watermarking technique is implemented in two ways by adding the Kashida after letters or before letters. The work proposed by Gutub *et al.* in [16] is based on Kashidas where a secret message is embedded as a watermarking code. The initial stage of this method inserts Kashidas in the cover text for confusion purposes to ensure security. Then, the Kashidas are embedded based on the watermarking code which is obtained using positions of the remaining Kashidas not used in the initial stage. This method showed adequate capacity ratio and from the point of view of security it is concerned mainly about some watermarking removal attacks making it a good candidate to serve copyright applications.

In [17], the authors proposed an improved method of Arabic text steganography using the Kashida character. In this method, they modified their initial method which embeds a Kashida for a bit 0 and 2 Kashidas for a bit 1 which they assumed to be suspicious. The modification calls for avoiding the addition of a Kashida for a code of 2 consecutive 0s. For example a code of 000001 will not place a Kashida for the first four 0 bits (neither the first or second 0 bit-pairs) then a Kashida will be placed for the 5th bit (0) and 2 Kashidas for the bit 1. The authors claim this will remove the suspiciousness that may occur using the initial method they proposed.

Finally, not all techniques presented in this work could be applied to all Arabic text. Thus, too many techniques may not be used with sensitive Arabic text such as religious and formal documents, including the Holy Quran, since no modifications to the text position, meaning ... etc. is permitted within the watermarking-technique. Therefore, the proposed Kashida-based methods explained in the next section provide increased protection to Arabic documents and they can be easily applied in sensitive documents.

4. Proposed Enhanced Kashida-Based Arabic Text Watermarking

Arabic characters have different lengths and it is impossible to have a font style which could provide a uniform font size as is the case for Latin scripts [5]. Therefore, the purpose of this work is to utilize Kashida marks by inserting them between characters in words as part of the watermark encoding in order to protect the copyright and intellectual property of people or/and organizations.

The methodology followed in this work is to encode the original text document with Kashida according to a specific key which will be produced before the encoding process. The key consists of 48 bits and is 100111010110101010101000111101011000101101101001. In this work, the Kashida is placed after certain characters according to some condition. The flow chart for encoding the watermarking key is shown in Fig. 4. The encoding algorithm is presented in Algorithm 1.

Fig. 4. Flow chart of the proposed watermarking technique.

Algorithm 1: Algorithm for Proposed Watermarking Method:

Input: key k in binary format with length L ; set of characters S before which Kashida should be placed; set of characters $T := \{\text{و، ز، ر، ذ، د، ح، آ، ا، ء، إ، ؤ}\}$ after which no Kashida can be placed.

Output: Watermarked text

Start with the first bit of the key by setting its index $j := 0$.

```

2-   For  $i = 2$ : total characters ( $C$ ) of document.
3-       If (Condj satisfied) then
4-           Insert Kashida before character  $C_i$ 
5-           Increase index of key by 1
6-            $j := j + 1$ 
7-       end If
       If the end of the key is reached before the EOF then
       repeat the key sequence for the rest of the document by
       the following operation:
8-       If ( $j > L$ ) then
9-           Set  $j := 0$ 
10-      end If
11-  end For

```

Algorithm 1 is a variation of the technique originally proposed in [18], in this paper the proposed approaches explained in Algorithm 1 exploits the frequency recurrences of characters to yield improved encodings in the host document. By observing the frequency recurrence graph in Fig. 3, the characters are split into two sets (namely, A and B). Set A consists of the first 14 characters with large recurrence frequencies, while Set-B consists of the next 15 characters with low recurrence frequencies. In the first proposed variation, Method-A; set $S := A$ as the search-space for possible embedding, which is used in algorithm 1 with Cond_j as follows.

If $C_i \in S$ and $C_{i-1} \notin T$ and $(k_j = 1)$ then place a Kashida

This implies that if the current key bit is one, and the current character C_i exists in Set S , then a Kashida is placed provided that the previous character C_{i-1} does not belong to Set T .

In the second proposed variation, Method-B; two sets A and B are used in the search-space for possible embeddings, where $S = \{A, B\}$, and Cond_j with Method-B in Algorithm 1 for placing Kashida according to the key bit values can be given with the following rule:

If $(k_j = 1)$ and $C_i \in B$ and $C_{i-1} \notin T$ then place a Kashida, else if $(k_j = 0)$ and $C_i \in A$ and $C_{i-1} \notin T$ then place a Kashida, which means that if the current key bit is zero and the current character C_i exists in Set-A, then a Kashida is placed. Moreover, if the key bit is one and the current character C_i exists in Set B, then a Kashida is also placed.

5. Results and Discussion

Fig. 5(b) and Fig. 5(c) illustrate examples of how the Kashidas are inserted in a document using the two proposed approaches (Method-A and Method-B) based on recurrence frequencies of characters in the host document. In order to compare the results with our proposed methods, three Kashida-based methods from the literature [15]-[17] were chosen to facilitate attaining the comparative metrics. The comparative results were obtained by testing each method on four different documents with different lengths. From the results, it is observed that the average number of Kashidas per word is related to imperceptibility. The lower the average number of Kashida's per word, the higher the imperceptibility and therefore, the closer the perceptual similarity of the encoding scheme with the original cover-text. The proposed digital watermark techniques present high imperceptibility encoding since the original cover text and the watermark key are

perceptually indistinguishable; on the other hand, these are low capacity watermarking techniques which are more suitable for copyright protection and document-authenticity as opposed to data hiding (as in Steganography). This shows that the lower the capacity the higher the imperceptibility, in other words, imperceptibility is inversely proportional to capacity. The proposed methods circularly embed the watermark message N times in the host document.

$$N = \frac{\text{No. of extendable characters in the document}}{\text{Watermark length (bits)}}$$

This consequently improves the robustness, since when a particular segment of the host document is modified due to common signal processing operations, the entire watermark can be extracted from other portions of the document. The proposed enhanced Kashida-based methods are robust compared to other methods found in the literature [15]-[17]. The comparison of all techniques on four different documents with different lengths using the same watermarking key is shown in Table 2-Table 5. From the tables, it can be checked that the capacity ratio and average number of Kashidas per word using the two proposed approaches Method-A and Method-B are lower than those using other Kashida methods found in literature. It can be furthered observed that no major difference occurs between the results of the two proposed methods based on the character frequency recurrence properties, but the level of capacity and imperceptibility varies depending on the number of characters used in each approach.

الهرولة في سحب الإيداعات من البنوك لأن رأس المال جبان وهذا ما تناولته وكالات الإعلام المختلفة. قيام العديد من المؤسسات المالية بتجميد منح القروض للشركات والأفراد خوفاً على صعوبة استردادها.

(a). Original text.

الهرولة في سحب الإيداعات من البنوك لأن رأس المال جبان وهذا ما تناولته وكالات الإعلام المختلفة.

↑↑↑ ↑ ↑ ↑ ↑↑↑↑↑↑↑ ↑↑ ↑↑ ↑↑↑↑ ↑ ↑ ↑ ↑ ↑ ↑↑↑

0 0 0 1 0 1 0 1 0 1 01 01 101 0 1 1 1 0 01

قيام العديد من المؤسسات المالية بتجميد منح القروض للشركات والأفراد خوفاً على صعوبة استردادها.

↑ ↑↑ ↑↑ ↑↑ ↑↑ ↑ ↑↑↑↑ ↑ ↑ ↑↑↑↑ ↑ ↑↑↑ ↑↑↑ ↑↑↑ ↑↑↑↑ ↑↑↑↑ ↑↑↑↑ ↑↑↑↑

1 01 1 1 00 1 1 0 0 1 0 1 1 011 0 1 00 0 1 1 0 1 0 1 1 1 1

(b). Watermarked text using Proposed Method A

الهرولة في سحب الإيداعات من البنوك لأن رأس المال جبان وهذا ما تناولته وكالات الإعلام المختلفة.

↑ ↑

1011 11 0 0 01 01 0 1 01 01 011 0 1 0 1 1 1 0 0 1

قيام العديد من المؤسسات المالية بتجميد منح القروض للشركات والأفراد خوفاً على صعوبة استردادها.

↑ ↑

0 1 0 1 0 1 0 1 0 1 1 0 101 11 00 110 01 0110 11 01 0 0 01 10

(c). Watermarked text using Proposed Method B.

Fig. 5. An example of a sampled text after applying the watermark key.

Table 2. Comparison Results Using Document 1

	Document 1				
	Total Characters	No. of words	Capacity	Capacity ratio	Average No. of Kashidas per word
Method 1[15]	3423	357	788	23.0	2.2
Method 2[16]	3423	357	630	18.4	1.8
Method 3[17]	3423	357	2264	66.1	6.3
Method 4[18]	3423	357	114	3.3	0.3
Proposed modified Method-A	3423	357	478	13.98	1.34
Proposed modified Method-B	3423	357	528	15.44	1.48

Table 3. Comparison Results Using Document 2

	Document 2				
	Total Characters	No. of words	Capacity	Capacity ratio	Average No. of Kashidas per word
Method 1[15]	2133	214	460	21.6	2.1
Method 2[16]	2133	214	392	18.4	1.8
Method 3[17]	2133	214	1408	66.0	6.6
Method 4[18]	2133	214	80	3.8	0.4
Proposed modified Method-A	2133	214	280	13.13	1.30
Proposed modified Method-B	2133	214	321	15.05	1.50

Table 4. Comparison Results Using Document 3

	Document 3				
	Total Characters	No of words	Capacity	Capacity ratio	Average No. of Kashidas per word
Method 1[15]	1582	156	386	24.4	2.5
Method 2[16]	1582	156	280	17.7	1.8
Method 3[17]	1582	156	1018	64.3	6.5
Method 4[18]	1582	156	44	2.8	0.3
Proposed modified Method-A	1582	156	218	13.81	1.40
Proposed modified Method-B	1582	156	241	15.28	1.55

Table 5. Comparison Results Using Document 4

	Document 4				
	Total Characters	No of words	Capacity	Capacity ratio	Average No. of Kashidas per word
Method 1[15]	7026	725	1568	22.3	2.2
Method 2[16]	7026	725	1240	17.6	1.7
Method 3[17]	7026	725	4140	58.9	5.7
Method 4[18]	7026	725	234	3.3	0.3
Proposed modified Method-A	7026	725	1036	14.76	1.43
Proposed modified Method-B	7026	725	1109	15.79	1.53

6. Conclusions

In conclusion, existing Kashida approaches in the literature were found to be vulnerable to security-attacks due to their high-rate of Kashida embeddings in their encoding scheme which were used mainly for steganographic applications. In this paper, an enhanced-Kashida encoding method was proposed

in order to achieve our goal of document copyright protection and authentication-verification whilst reducing the vulnerability to security-attacks in the watermarked document. In addition, the two approaches (Method-A and Method-B) based on character frequency recurrence properties were proposed. The proposed encoding algorithms proved to yield enhanced robustness and improved imperceptibility as compared to the other Kashida based methods in literature whilst achieving our goal with a relatively low watermark capacity ratio. Advantageously, the use of frequency recurrence statistics of Arabic characters had enabled the dynamic variation of imperceptibility and robustness levels as required for a given target application as demonstrated in the Results section. This shows the applicability of the presented methods for applications that include copyright protection, document-authenticity verification and document tamper-proofing.

Acknowledgment

The authors would like to thank and acknowledge the IT Research Center for the Holy Quran (NOOR) at Taibah University for their financial support during the academic year 2012/2013 under research grant reference number NRC1-126.

References

- [1] Jalil, Z., Mirza, A., & Sabir, M. (2010, February). Content based zero-watermarking algorithm for authentication of text documents. *International Journal of Computer Science and Information Security*, 7(2).
- [2] Thapa, M., Sood, S. K., & Sharma, M. (2011). Digital watermarking: current status and key issues. *International Journal of Advances in Computer Networks and its Security*, 1(1), 327-332.
- [3] Ranganathan, S., Ali, A., Kathirvel, K., & Kumar, M. M. (2010). Combined text watermarking. *Int'l Journal of Computer Science and Information Technologies*, 1(5), 414-416.
- [4] Alginahi, Y. M. (2013). A survey on Arabic character segmentation. *International Journal on Document Analysis and Recognition*, 16(2), 105-126
- [5] Zeki, A. M. (2005). The segmentation problem in Arabic character recognition: the state of the art. *Proceedings of First International Conference on Information and Communication Technologies* (pp. 11-26).
- [6] Atallah, M. J., *et al.* (April, 2001). Natural language watermarking: Design, analysis, and a proof-of-concept implementation. *Proceedings of the Fourth Information Hiding Workshop: Vol. LNCS 2137* (pp. 25-27). Pittsburgh, PA.
- [7] Ensen, C. D. (2001). Fingerprinting text in logical markup languages. In G. I. Davida, & Y. Frankel (Eds.), *Lecture Notes in Computer Science* (pp. 433-445). Springer Verlag, Berlin, Heidelberg.
- [8] Sun, X., & Asimwe, A. J. (2005, August). Noun-verb based technique of text watermarking using recursive decent semantic net parsers. *Lecture Notes in Computer Science*, 3612, 958-961.
- [9] Peng, L., *et al.* (2009). An optimized natural language watermarking algorithm based on TMR. *Proceedings of 9th International Conference for Young Computer Scientists*.
- [10] Macq, B., & Vybornova, O. (January, 2007). A method of text watermarking using presuppositions. *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*.
- [11] Niimi, M., Minewaki, S., Hoda, H., & Kawaguchi, E., (2003). A framework of text-based steganography using SD-Form semantics model. *Proceedings of Pacific Rim Workshop on Digital Steganography 2003*, Kyushu Institute of Technology, Kitakyushu, Japan, July 3-4.
- [12] Low, S. H., Maxemchuk, N. F., Brassil, J. T., & O'Gorman, L. (April, 1995). Document marking and

- identification using both line and word shifting. *Proceedings of Infocom 1995* (pp. 853–860). Boston, MA.
- [13] Davarzani, R., & Yaghmaie, K. (2009). Farsi text watermarking based on character coding. *Proceeding of International Conference on Signal Processing Systems* (pp. 152–155).
- [14] Yang, H., *et al.* (June, 2004). Text document authentication by integrating inters character and word spaces watermarking. *Proceedings of IEEE International Conference on Multimedia and Expo* (pp. 955-958). Taibei.
- [15] Gutub, A., Ghouti, L., Amin, A., Alkharobi, T., & Ibrahim, M. (2007). Utilizing extension character 'Kashida' with pointed letters for Arabic text digital watermarking. *Proceedings of International Conference on Security and Cryptography* (pp. 329-332), Barcelona, Spain.
- [16] Gutub, A., Al-Haidari, F., Al-Kahsah, K., & Hamodi, J. (Feb., 2010) e-Text watermarking: utilizing 'Kashida' extensions in Arabic language electronic writing. *Journal of Emerging Technologies in Web Intelligence*, 2(1).
- [17] Gutub, A., Al-Alwani, W., & Mahfoodh, A. B. (Dec. 2010). Improved method of Arabic text steganography using the extension 'Kashida' character. *Bahria University Journal of Information and Communication Technology*, 3(1).
- [18] Alginahi, Y. M., Kabir, M. N., & Tayan, O. (Nov., 2013). An enhanced Kashida-based watermarking approach for Arabic text-documents. *Proceedings of IEEE International Conference on Electronics, Computer and Computation* (pp. 301-304).

Yasser M. Alginahi received a Ph.D. degree in electrical engineering from the University of Windsor, Windsor, Ontario, Canada, and a master of science degree in electrical engineering and a bachelor degree in biomedical engineering from Wright State University, Dayton, Ohio, U.S.A.

He is an associate professor at the Department of Computer Science, Deanship of Academic Services at Taibah University, Madinah, Saudi Arabia. He is also the consultation unit coordinator at the IT Research Center for the Holy Quran and its Sciences, Taibah University, Madinah, Saudi Arabia. He published a book entitled "Document Image Analysis" and he has published over 70 journal and conference papers. He worked as a principal investigator and co-principal investigator on many funded research projects by the Deanship of Scientific Research at Taibah University and other organizations such as King Abdul-Aziz City of Science and Technology. His current research interests are Quran security, information security, document image analysis, pattern recognition, OCR, modeling and simulation and numerical computations. Dr. Alginahi is a licensed professional engineer, Ontario, Canada, a member of Professional Engineers Ontario, a senior member of IACSIT and IEEE since 2010. And he had received Taibah University 2013 Research Excellence Award.

Muhammad Nomani Kabir received his PhD degree in computer science at the University of Braunschweig, Germany. He is currently a senior lecturer at the Faculty of Computer Systems and Software Engineering of University of Pahang in Malaysia. Previously, he was an assistant professor in computer science at Taibah University, Saudi Arabia and American International University-Bangladesh. His research interests include issues related to numerical methods, embedded systems, network security and cryptography.

Omar Tayan completed his undergraduate degree in computer and electronic systems from the University of Strathclyde, Glasgow, UK, and his PhD degree in computer networks, Department of Electronic & Electrical Engineering from the same university. He has been working as an assistant professor at the College of Computer Science and Engineering (CCSE) and IT Research Center for the Holy Quran and Its Sciences (NOOR) at Taibah University in Saudi Arabia since 2007. He was a consultant to the Strategic and Advanced Research and Technology Innovation Unit at the university for four years and is one of the founding members of the "IT Research Center for the Holy Quran and Its Sciences (NOOR)" at Taibah University, Al-Madinah Al-Munawwarah, Saudi Arabia.

His research interests include: information security, e-learning technologies, performance modeling and evaluation, high-speed computer networks and architectures, software simulation techniques and queuing theory, wireless sensor networks for intelligent transportation systems, networks-on-chip (NoC) and optical networks (ONs).