Indonesian Journal of Electrical Engineering and Computer Science
Vol. 24, No. 2, November 2021, pp. 1001~1008
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v24.i2.pp1001-1008

Framework of diacritic segmentation for Arabic handwritten document

Ahmed Abdalla Sheikh, Mohd Sanusi Azmi, Maslita Abd Aziz, Mohammed Nasser Al-Mhiqani, Salem Saleh Bafjaish
Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia

Article Info

Article history:

Received Feb 5, 2021
Revised Sep 9, 2021
Accepted Sep 16, 2021

Keywords:

Arabic handwritten segmentation
Diacritics segmentation
Image segmentation phase
Pre-processing phase
Region-based

ABSTRACT

In recent Arabic standard language and Arabic dialectal texts, diacritics and short vowels are absent. There are some exceptions have been made for the Arabic beginner learner scripts, religious texts and as well as a significant political text. In addition, the text without diacritics is considered ambiguous due to numerous words with different diacritic marks seem identical. However, this paper we present a framework for segmenting diacritics from Arabic handwritten document by using region-based segmentation technique. Since Arabic handwritten and Mushaf Al-Quran contain many diacritical marks. Hence, the diacritics must be properly extracted from Arabic handwritten document to avoid losing some good features. Furthermore, the proposed framework is devised specifically to segment diacritics from Arabic handwritten image, thus there will be no feature extraction, feature selection, and classification processes included. Besides, we will present the methodology that is used to fulfil the objectives of this paper. The pre-processing phases will be explained and more specifically segmentation phase for segmenting diacritics which is the phase we concentrate more in this article. Lastly, we will identify the proposed technique region-based segmentation to facilitate our development throughout the experimental process.

Corresponding Author:

Mohd Sanusi Azmi
Fakulti Teknologi Maklumat dan Komunikasi
Universiti Teknikal Malaysia Melaka
Jalan Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
Email: sanusi@utem.edu.my, ashiekh295@gmail.com

1. INTRODUCTION

Arabic is the most spoken language in the world with approaxmaely more than 420 people [1], [2], and its an extensively used alphabetic synthesis system on the planet, and it has 28 fundamental letters. The letters all together was initially used to make messages in Arabic, most undeniably the Quran the holy book of Allah swt. With the expansion of Islam, it came to be utilized form various vernaculars like Urdu, Pashto, Uyghur (in China), Ottoman Turkish and Spanish (in Western Europe). By then various changes and enhancements have been made to Arabic composed work substance, which understands additional characters and strokes. The modern strokes are so-called diacritics, and the explanation behind including these diacritics is to see letters of the practically identical or similar to Indicate and shape the sounds (vowels and tones) that are not passed on by the crucial letter set [3]. In Arabic language especially in (Quran Language) there's signs called by "diacritics" or "diacritical marks "which represent short vowels or other sounds if one of these diacritical marks overlooked, they might change the whole meaning of the word.

However, the current researches don't emphasize on secondary foreground image (diacritic) which considers as a significance part in Arabic handwritten in terms of identifying the originality of handwritten document, but rather they only considering about the primary foreground image such as huruf and subword. Moreover, the Arabic diacritics mark is an accompanying sign in a letter to amend the identical voice or to recognize the word from each other. The diacritics mark is immensely used in semantic languages such as Arabic. The objective behind these marks is to identify the structure of morphological, the linguistic function, the importance of semantic words and further lingual and sound features [4]. Diacritics marks in Arabic language are often omitted [5], compared to language like French that the diacritics should be present in the word [6]. To but this into the context, Figure 1, clarifies the differences among the original image, the current research which only focus on the primary image and as well as the secondary foreground image (diacritics).

Figure 1. Differences between original image, primary image, secondary foreground image

Besides, there are some existing techniques such as Voronoi and some others related that are not able to recognize the secondary foreground image in the image of Arabic handwritten in which they consider it as a noise object in the image. However, the propose of segmenting Arabic diacritics is to extract secondary foreground image (diacritics) that can be used as a feature instead of foreground image and as well as to identify the originality of the handwritten document. In this paper, a framework of extracting diacritics will be presented in order to produce a good image processing and extract the secondary features successfully.

These days, the diacritics in Arabic are not considered as optional or additional to the language, they are becoming a significance part in the language itself, they are compulsory for the new leaners of Arabic such as children and foreigners. To put this into example, in order the writer wants to master the Arabic script writing style, he/she should initially master the writing and right placement of the diacritics as well, in order to compose a right and surely knew Arabic [7]. Additionally, Mushaf Al-Quran is the heavenly and most fast approaching book for many Muslims around the globe [8]. It tends to contain two forms printed or digital form, though it's in Arabic, but its written in the way that is dissimilar from any Arabic/Jawi document based as it encompasses "Diacritics" [9], [10].

At the point when the diacritics are accessible, the Arabic substance gives enough information about the correct explanation and the significance of the words [11]. In specific applications like Arabic substance to discourse, diacritics are essential to get the correct articulation [12]. Likewise, diacritics support to get the reference interpretation for discourse acknowledgment structures [13]. The changing of Arabic diacritics may change the word semantics. The issue is that most Arabic substance is customarily formed without diacritics, which makes it questionable. Likewise, diacritics in Arabic have two essential occupations: (i) they give a vocal guide, to empower perusers to display/articulate the substance adequately, and (ii) they disambiguate the normal significance of generally ambiguous words [14].

Apart of this, the absence of diacritical marks is a foundation of intricacy for processing Arabic automatic systems language that the sentence meaning cannot be determined easily [15]. Moreover, the shortage of Arabic diacritical marks in the sentences reflects the major reason of the muddle encounter throughout analysis stage [16]. Over and above, the study of [17] demonstrated that the diacritization automatic content increments manual quality labelling of the corpus.

Usually the segmentation algorithm challenges come from either type of writing or language processes ssuch as problems on natural language processing [18]. From type of writing point of view, what comes first is handwritten documents. The complexity of this arises from several sources. For instance, the writers mostly tend to the cursive nature writing. Another particular good example of this is the mode of writer which implies that one author may have more than one single writing style dependent on his/her mode. While on the other hand, from the texts handwriting perspective, Arabic is at the highest point of the most muddled dialects. Additionally, Arabic despite everything experiences lack of explores contrasted with different dialects, for example, Chinese and Latin. Moreover, the significant difficulties in Arabic are the plausibility of overlapping few characters with others which the make the segmentation ways of the words very ambiguous and difficult to detect [19]. Similarly, the language of Arabic text is written presently without diacritics and vowels which also recognized as (Tashkeel) which considerably makes the significance of the word equivocal and imprecise semantically, especially for non-local Arabic speakers [20].

For example, the Arabic word (كتب) if it composed without diacritic marks, it could have totally various implications. Truth be told, it could be interpreted as the words "write" "wrote" "books". Along these lines, diacritic marks are criticalness in Arabic language to disentangle the articulation and legitimate seeing particularly for non-local Arabic speakers while local speakers can anticipate the right importance of the word intellectually as per the specific situation [21]. The nowadays educated Arabic speakers can retrieve precisely the diacritics from document. The text with less diacritics turns into a wellspring of muddle for readers who are new in Arabic learning and as well as the learning disabilities people [22].

The diacritics in Arabic is considered a huge significance for non-native Arabic or new learners as there are numerous words that have the same writing style, but they are differ with regard to diacritics placement [23]. Furthermore, diacritic marks play a crucial role in meeting the criteria of ease of use of typographic text, for example, homogeneity, clarity and purity [24]. Changing the diacritic of a letter in a word could totally change its semantic or meaning. The circumstance is extremely entangled with polyglot content [13].

As shown in Figure 2, we have four similar words, but each has different diacritical marks. The word (لبس) has variety of meanings that differ from each other in term of diacritics placement in the word. For instance, the word (لَبِسَ) which means in English (Donned), word (لِبْسٌ) which means in English (Donned), word (لُبِسَ) which means in English (Was donned). As is clear, all these mentioned words have same writing style, but they have different diacritical marks placement. This consider as major challenges in Arabic language how to recognize the word when there is no diacritics placed in the word. Additionally, the variety of writing styles and the tendency within the same line make the process of the extraction/segmentation from Arabic handwritten even more challenging [21]. Thus, in this paper a framework for segmenting will be proposed to prepare image for extraction process using Arabic handwritten document

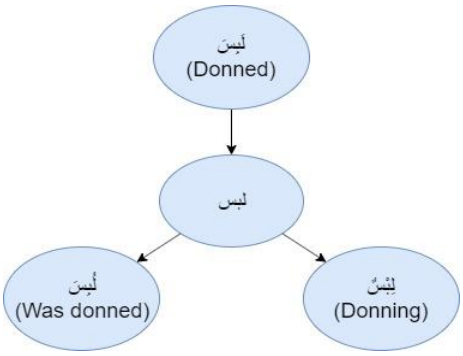

Figure 2. Arabic diacritics challenges

2. RESEARCH METHOD

In image processing there is a significance phase called Pre-processing. This phase major task is the preparation of the document to processed. However, the Pre-processing contains subsequence phases as following: Segmentation phase, which is segmenting image into multiple areas that simplifies the image representation into meaningful and easier to analyse. Figure 3 shows the pre-processing and segmentation phases. Feature extraction phase, which highlights on reducing the dimensionality process, where a set of initial raw variables is reduced into manageable groups, while still accurately and fully describing the original dataset. Feature selection phase, which otherwise called selection of attribute or variable, is a subset selecting process of relevant features such as variable and predictors. Classification phase, which concerns on labelling images into number of predefined categories [25]. Lastly, after conducting the pre-mentioned phases, the post-process image output will be produced which contain the desired features.

All these stages are very crucial as if an error occurred in one of them, it would hugely affect on the subsequent ones. However, in this paper our major concentrate would be on the segmentation phase as our main propose is to segment diacritics from Arabic handwritten based document by devising a suitable framework. In addition, our dataset will be an Arabic handwritten document image which is more challenging than Latin documents and that's majorly because of the semi-cursive nature of the its script which characterized by the attendance of the ascending and descending character and as well as the calligraphy [26].

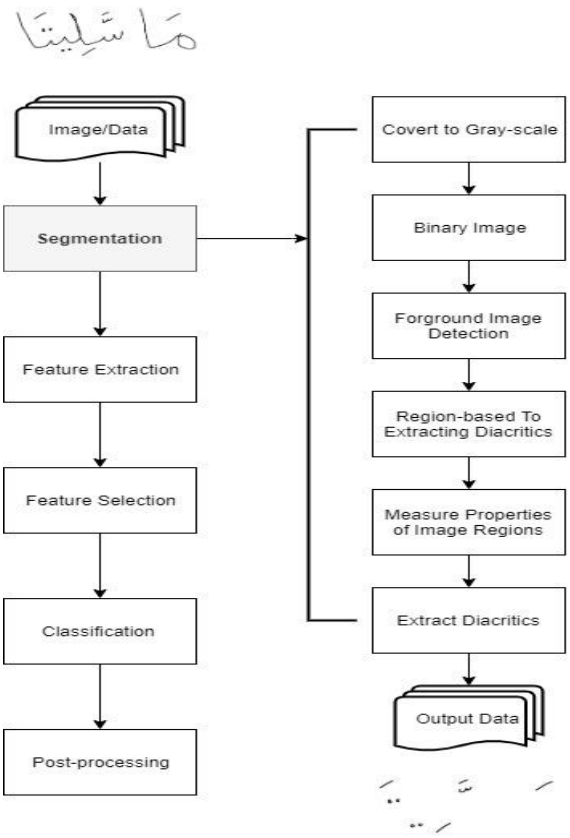

Figure 3. Pre-processing and segmentation phases

In the segmentation phase, there are a few steps displayed below, that enable us to segment diacritics starting from collecting the data, which is from Arabic handwritten document, converting image to grayscale, converting image to binary image, region-based to measure the area, which is proposed technique in this research, and lastly, extracting diacritics from image.

- Image data
  The image is collected from Arabic handwritten documents and as well as Al Quran since its originally handwritten documents. Arabic contains many diacritical marks [27].
- Grayscale image
  Obviously, some of images come with different colors, thus, there is a requirement for transformation grayscale picture so as to make the process of segmenting the diacritics easier. Converting image to grayscale is somehow necessary for cost reduction, make easy implementing binary algorithm, and as well as reduce color.
- Binary image
  It's the major phase in preprocessing, binary image has two values for each pixel which are black and white [28]. Be that as it may, this step is made from unique picture in order to identify the significance parts of the picture.
- Foreground image detection
  Next when a binary image is made, the foreground image will take place to spot areas in Arabic handwritten text that contain diacritics in the Image.
- Proposed technique
  In this paper, the technique is used to segment diacritics is region-based segmentation which is a technique that determines the region directly as the principal objective of segmentation is parceling the picture into series of areas and regions. Each image has its own area measurement to extract the diacritics from due to differences of pixels from one another [29].
- Extract diacritics/Output Data
  After conducting the pre-mentioned steps, the final step will take place to obtain image output that contains only diacritics. Thus, the aim of the research is fulfilled.

This paper presents region-based segmentation as proposed technique. As we already know, the basic objective of segmentation is to fragment picture into areas. This segmentation method allows us to measure the area of the region to extract diacritics from image. The area measurement for each image are differ from one another [30]. Despite the challenges that some images that contain an overlapped and connected component in the region, is difficult to recognize between the primary object and secondary object (diacritics) due to differences of writing style, overlapped and several connected components for each image and as well as the different pixels from one image another. Region-based remains the simplest technique among existence techniques that is suitable for segmenting diacritics. The pseudocode for this technique is illustrated as shown in Figure 4.

```
Start
1.0  Read image input
2.0  Pre-processing image
        2.1  convert to gray-scale
        2.2  convert to binary
3.0  Find connected components
4.0  Measure the properties of image region
5.0  Find the area in the region
6.0  Create LabelMatrix function
7.0  Output image result
End
```

Figure 4. Pseudocode region-based

The pseudocode defined in Figure 4 is elaborating more the segmentation phase of diacritics mentioned in Figure 3, but in a very detailed way. To be more accurate, this pseudocode is designed to implement the operational framework technically in order to get a segmented diacritic image as output. However, the pseudocode for this study contains several steps that enable us to extract diacritics from image such as: reading image input which implies image from Arabic handwritten document. Preprocessing that contains gray-scale and binary image to facilitate the segmentation process. Finding connected component which is in binary picture. Measure the properties of image meaning that the arrangement of properties of each associated segment in binary picture is measured. Finding the area in the region according to the higher and lower area for instance the area for image 1 is >=240 && area <= 279. Create Label Matrix function for creating a binary image containing connected component returned by image, and it retrieves the minimum numeric class for the number of objects. Lastly, the result output will be shown as extracted diacritics features from image of Arabic handwritten.

3. RESULTS AND DISCUSSION

The main function of the segmentation phase is to separate the region of the interest from background. Image segmentation is one of the most fundamental implementations in image processing [31]. The images input was at that point thresholded so we can separate the diacritics straightforwardly. The purpose for segmenting diacritics is to kill the primary content of the text while keeping the diacritics immaculate. However, in this paper the proposed method has been experimented in MATLAB Software, also was tested using Arabic handwritten images that contains diacritics. Region-based functions were implemented for measuring image area as well as for segmenting diacritics. The following Figure demonstrates the proposed method mechanism of how the area of the region is measured and what best area measurement is suitable for Arabic handwritten document in order to extract diacritics and this may vary from one image to another as shown in Figure 5.

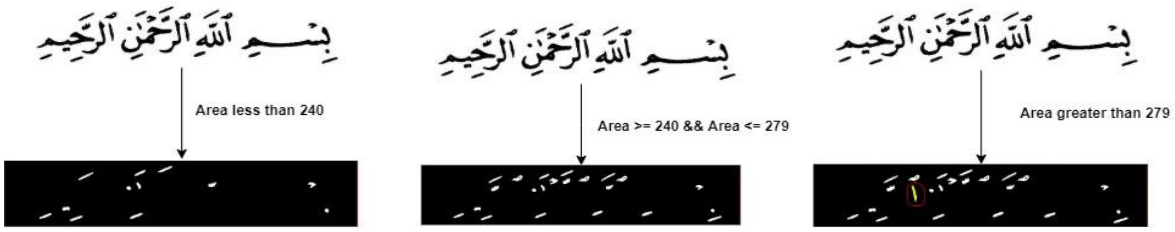

Figure 5. Experimental framework result

Figure 5 shows three experiment result using Arabic handwritten document that was obtained from the source [32]. However, the region-based segmentation mechanism works as the more objects Arabic handwritten contains the less accurate result we can get. Conversely, the less objects Arabic handwritten contains the better result we can obtain which is diacritic. As shown in Figure 5 the area less than 240 shows us image that missing some diacritic features which probably means this area of region not precise. While on the other hand, the area of region that range between 240 and 279 has obtained us the fully features of diacritics in the image, thus, this area measurement is appropriate for this specific handwritten document. Lastly, any area measurement who is greater than 279 shows the primary objects such as text, huruf, which means the area went beyond the boundaries of region. In summary, as the region-based segmentation technique is the easiest technique for implementing this experiment it has its shortage when it comes to image of handwritten document that contains many diacritics features the image will be shown contain less diacritics than the actual image is. Hence, future work we must emphasize on enhancing the proposed technique to make it very accurate and robust in extracting diacritics from images of Arabic handwritten document.

4. CONCLUSION

In this paper, a framework for segmenting diacritics in Arabic handwritten document was presented. The study emphasizes on extracting diacritics from documents and use it as features by using region-based segmentation technique which is partitioning the image into series of regions and determines the region directly and as well as measuring the properties of the region area in order to obtain the diacritics from image. Two phases were explained such as pre-processing phases and particularly segmentation phase. Moreover, the result in this study, is an experiment conducted in MATLAB software which carried out the proposed technique to obtain image containing diacritics features. Future work for this study is to conduct as many researches as possible in order to find several techniques that able to recognize the diacritics as features and as well as segmented accurately.

ACKNOWLEDGEMENTS

The authors thank the Ministry of Education for funding this study through the following grants: FRGS/1/2017/ICT02/FTMK-CACT/F00345. Gratitude is also due to Universiti Teknikal Malaysia Melaka and Faculty of Information Technology and Communication for providing excellent research facilities.

REFERENCES

[1] A. M. F. Al Sbou, "A survey of Arabic text classification models," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 6, pp. 4352-4355, 2018, doi: 10.11591/ijece.v8i6.pp.4352-4355.

[2] H. K. Tayyeh, M. S. Mahdi, and A. S. A. Al-Jumaili, "Novel steganography scheme using Arabic text features in Holy Quran," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 3, pp. 1910-1918, 2019, doi: 10.11591/ijece.v9i3.pp1910-1918.

[3] M. Lutf, X. You, Y. M. Cheung, and C. L. P. Chen, "Arabic font recognition based on diacritics features," *Pattern Recognit.*, vol. 47, no. 2, pp. 672-684, 2014, doi: 10.1016/j.patcog.2013.07.015.

[4] F. Debili and A. Hadhémi, "Voyellation automatique de l'arabe," *Semitic '98: Proceedings of the Workshop on Computational Approaches to Semitic Languages*, 1998, pp. 42-49, doi: 10.3115/1621753.1621761.

[5] B. Sbartai and H. Mouloud, "Évaluation de la vulnérabilité sismique en milieu urbain : Application à la ville de Constantine," *Conference: 31èmes Rencontres de l'AUGC*, 2013, pp. 1-10.

[6] A. Chennoufi and A. Mazroui, "Morphological, syntactic and diacritics rules for automatic diacritization of Arabic sentences," *Journal of King Saud University-Computer and Information Sciences*, vol. 29, no. 2, pp. 156-163, April 2017, doi: 10.1016/j.jksuci.2016.06.004.

[7] M. Lutf, X. You, and H. Li, "Offline Arabic handwriting identification using language diacritics," *2010 20th International Conference on Pattern Recognition*, 2010, pp. 1912-1915, doi: 10.1109/ICPR.2010.471.

[8] L. N. M. Esa, M. A. Morshidi, and S. M. M. Zailani, "Development of pose estimation algorithm for Quranic Arabic word," *TELKOMNIKA (Telecommunication Comput. Electron. Control)*, vol. 16, no. 4, pp. 1633-1641, 2018, doi: 10.12928/TELKOMNIKA.v16i4.9048.

[9] S. S. Bafjaish, M. S. Azmi, M. N. Al-Mhiqani, and A. A. Sheikh, "Skew correction for mushaf Al-Quran: A review," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 17, no. 1, pp. 516-523, 2020, doi: 10.11591/ijeecs.v17.i1.pp516-523.

[10] S. S. Bafjaish, M. S. Azmi, M. N. Al-Mhiqani, A. R. Radzid, and H. Mahdin, "Skew Detection and Correction of Mushaf Al-Quran Script using Hough Transform," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 402-409, 2018, doi: 10.14569/ijacsa.2018.090852.

[11] Y. Hifny, "Smoothing Techniques for Arabic Diacritics Restoration," *The Twelfth Conference on Language Engineering (ESOLEC'12)*, 2012, pp. 6-12.

[12] Y. Hifny *et al.*, "ARABTALK: An implementation for arabic text to speech system," *Conference: nemlar. conference*, 2006, vol. 14.

[13] D. Vergyri and K. Kirchhoff, "Automatic diacritization of Arabic for acoustic modeling in speech recognition," *Semitic '04: Proceedings of the Workshop on Computational Approaches to Arabic Script-based Languages*, 2004, pp. 66-73, doi: 10.5555/1621804.1621822.
[14] M. Jarrar, F. Zaraket, R. Asia, and H. Amayreh, "Diacritic-Based Matching of Arabic Words," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 18, no. 2, pp. 1-21, 2019, doi: 10.1145/3242177.
[15] A. Said, M. El-sharqwi, A. Chalabi, and E. Kamal, "A Hybrid Approach for Arabic Diacritization Arabic Diacritization : Linguistic Description," *Conference: 18th International Conference on Applications of Natural Language to Information Systems, NLDB 2013*, June 2013, pp. 53-64, doi: 10.1007/978-3-642-38824-8_5.
[16] M. Boudchiche and A. Mazroui, "Evaluation of the ambiguity caused by the absence of diacritical marks in Arabic texts : statistical study," *2015 5th International Conference on Information & Communication Technology and Accessibility (ICTA)*, 2015, pp. 1-6, doi: 10.1109/ICTA.2015.7426904.
[17] H. Bouamor *et al.*, "A Pilot Study on Arabic Multi-Genre Corpus Diacritization Annotation," *Proceedings of the Second Workshop on Arabic Natural Language Processing*, 2015, pp. 80-88, doi: 10.18653/v1/W15-3209.
[18] M. N. Al-Mhiqani, R. Ahmad, Z. Z. Abidin, W. Yassin, A. Hassan, and A. N. Mohammad, "New insider threat detection method based on recurrent neural networks," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 17, no. 3, pp. 1474-1479, 2019, doi: 10.11591/ijeecs.v17.i3.pp1474-1479.
[19] Y. Osman, "Segmentation Algorithm for Arabic Handwritten Text based on Contour Analysis," *2013 International Conference On Computing, Electrical And Electronic Engineering (ICCEEE)*, 2013, pp. 447-452, doi: 10.1109/ICCEEE.2013.6633980.
[20] A. A. Sheikh, M. S. Azmi, M. A. Aziz, M. N. Al-Mhiqani, and S. S. Bafjaish, "Diacritic segmentation technique for Arabic handwritten using region-based," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 18, no. 1, pp. 478-484, 2020, doi: 10.11591/ijeecs.v18.i1.pp478-484.
[21] O. E. Shaaban, "Automatic Diacritics Restoration for Arabic Text," Masters thesis, King Fahd University of Petroleum and Minerals, 2014. [Online]. Available: https://eprints.kfupm.edu.sa/id/eprint/139145
[22] I. Zitouni and R. Sarikaya, "Arabic diacritic restoration approach based on maximum entropy models," *Comput. Speech Lang.*, vol. 23, no. 3, pp. 257-276, 2009, doi: 10.1016/j.csl.2008.06.001.
[23] M. A. Abuzaraida, M. Elmehrek, and E. Elsomadi, "Online handwriting Arabic recognition system using k-nearest neighbors classifier and DCT features," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 3584-3592, 2021, doi: 10.11591/ijece.v11i4.pp3584-3592.
[24] M. Hssini and A. Lazrek, "Design of Arabic Diacritical Marks," *Int. J. Comput. Sci. Issues*, vol. 8, no. 3, pp. 262-271, 2011.
[25] A. A. H. Hassan, W. M. Shah, M. F. Iskandar, M. N. Al-Mhiqani, and Z. K. Naseer, "Unequal clustering routing algorithms in wireless sensor networks: A comparative study," *J. Adv. Res. Dyn. Control Syst.*, vol. 10, no. 2-Special Issue, pp. 2142-2156, 2018.
[26] A. Souhar, Y. Boulid, E. Ameur, and M. Ouagague, "Segmentation of Arabic Handwritten Documents into Text Lines using Watershed Transform," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 4, no. 6, pp. 96-102, 2017, doi: 10.9781/ijimai.2017.08.002.
[27] D. A. Mohammed, A. A. H. Mezher, and H. S. Hadi, "Off-line handwritten character recognition using an integrated DBSCAN-ANN scheme," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 14, no. 3, pp. 1443-1451, 2019, doi: 10.11591/ijeecs.v14.i3.pp1443-1451.
[28] L. B. Melhem, M. S. Azmi, A. K. Muda, N. J. Bani-Melhim, and M. Alweshah, "Text line segmentation of al-quran pages using binary representation," *Advanced Science Letters*, vol. 23, no. 11, pp. 11498-11502, 2017, doi: 10.1166/asl.2017.10315.
[29] M. Kaur and P. Goyal, "A Review on Region Based Segmentation," *International Journal of Science and Research (IJSR)*, vol. 4, no. 4, pp. 2319-7064, 2015.
[30] MathWorks, "Measure properties of image regions," 2021. [Online]. Available: https://www.mathworks.com/help/images/ref/regionprops.html
[31] K. V. Sánchez, "Functional-Communicative Grammar (Spanish-German) for Translators and/or Interpreters: A Project," *Babel. Rev. Int. la Trad. / Int. J. Transl.*, vol. 47, no. 2, pp. 109-120, 2001, doi: 10.1075/babel.47.2.03san.
[32] King Fahd Complex for the Printing of the Holy Qur'an, "Digital Mushaf of Madinah," 2021. [Online]. Available: https://dm.qurancomplex.gov.sa/hafsdownload

BIOGRAPHIES OF AUTHORS

Ahmed Abdalla Sheikh received his BSc in Computer Science (Database Mangement System) in 2017 and MSc in Computer Science (Software Engineering and Intelligent) from the Universiti Teknikal Malaysia Melaka (UTeM) in 2019. His research intrests include image processing, image segmentation, computer vision systems.

Mohd Sanusi Azmi is an Associate Professor Department of Software Enginering, Universiti Teknikal Malaysia Melaka (UTeM). He received BSc. Msc and Ph.D from Universiti Kebangsaan Malaysia (UKM) in 2000, 2003 and 2013. He is the Malaysian pioneer researcher in identification and verification of digital images of Al-Quran Mushaf. He is also involved in Digital Jawi Paleography. He actively contributes in the feature extraction domain. He has proposed a novel technique based on geometry feature used in Digit and Arabic based handwritten documents.

Maslita Abd Aziz is a senior lecturer at Faculty of Information and Communication Technology, UTeM. She finished her first degree at Universiti Utara Malaysia (UUM) in 1995 with BSc in Information Technology (with Hons.) and later her MSc from Rochester Institute of Technology, New York, USA in 1998 with MSc in Information Technology (with Hons) specializing in Software Development and Management. Her research interests are in information retrieval, specifically on code retrieval of how to assist programmers during system development or learning the language

Mohammed Nasser Al-Mhiqani received his BSc in Computer Science (Computer Networking) in 2014, and MSc in Computer Science (Internetworking Technology) from the Universiti Teknikal Malaysia Melaka (UTeM) in 2015. Currently, he is a PhD student at the Universiti Teknikal Malaysia Melaka (UTeM). His research interests include cyber security, cyber-physical system security, insider threats, machine learning, and image Processing.

Salem Saleh Bafjaish received his Bachelor Degree from Staffordshire University, Malaysia in 2014 in computing (Software Engineering) and his Master Degree from UTeM university department of Information and Communications Technology (Software Engineering and Intelligent) in 2019. His current research interests are document analysis, image processing, computer vision, machine learning.