International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 2, April 2024, pp. 1834~1841
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i2.pp1834-1841

Segmentation techniques for Arabic handwritten: a review

Ahmed Abdalla Sheikh[1], Mohd Sanusi Azmi[1], Waleed Abdel Karim Abuain[2], Maslita Abd Aziz[1]
[1]Department of Information and Communication Technology, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[2]Department of Computer Science, Applied College, Taibah University, Al-Madinah Al-Munawwara, Saudi Arabia

Article Info

Article history:

Received Sep 11, 2023
Revised Nov 7, 2023
Accepted Nov 29, 2023

Keywords:

Arabic handwritten
Handwritten segmentation
Image segmentation
Segmentation methods
Segmentation techniques

ABSTRACT

Image segmentation refers to the process of partitioning a page into distinct sections. This technique aims to improve and transform the image's representation into a more coherent and user-friendly format. Its common application involves identifying objects and boundaries (such as lines and curves) within images. However, this paper focuses on discussing segmentation methods specifically tailored for Arabic handwritten content. Dealing with the segmentation of Arabic handwritten material poses a significant challenge due to the diverse handwriting styles and the interconnection between Arabic letters. The paper will also touch on the classification of segmentation algorithms originally designed for modern documents, illustrating their adaptation in document processing. Furthermore, the paper will address the difficulties associated with segmenting Arabic handwritten content, including variations in writing style, the connected nature of Arabic characters, the complexity of Arabic cursive writing and as well as the diacritics challenges. Lastly, a concise overview of previously widely used segmentation techniques in various research endeavors will be provided.

Corresponding Author:

Ahmed Abdalla Sheikh
Department of Information and Communication Technology, Faculty of Information and Communication Technology, Unvirsiti Teknikal Malaysia Melaka
Jalan Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
Email: ashiekh295@gmail.com

1. INTRODUCTION

Image segmentation is a process where features with similar attributes are identified and grouped together. This procedure can involve statistical clustering, thresholding, edge detection, region identification, or a combination of these methods. Segmentation is typically a fundamental step in image analysis, marking the transition from treating individual pixels as the units of perception to dealing with coherent objects within the image that consist of multiple pixels. Furthermore, the segmentation phase holds significance because it impacts the accuracy of recognition and the origin of errors in recognition [1]. When segmentation is executed effectively, all subsequent stages of image analysis become more straightforward and uncomplicated. Image segmentation serves as a crucial and complex operation in image processing. The techniques employed in image segmentation aim to partition an image into coherent portions that exhibit similar characteristics, as shown in Figure 1.

The objective of segmentation is to streamline the representation of an image into a meaningful and analyzable format. In the field of image analysis, segmentation is the initial step [2]. It encompasses the division of an image into contiguous and meaningful segments or regions. Segmentation stands as a pivotal first phase for a variety of image analysis methodologies. The outcome of image segmentation in terms of its

accuracy and nature significantly influences the performance of these image analysis techniques [3]. This paper is dedicated to exploring segmentation techniques specific to Arabic handwritten documents. While Arabic characters serve as scripts for languages like Arabic, Uygur, Urdu, and Farsi, they have not gained as much attention in character recognition as Chinese or Latin characters have. Despite being more challenging due to their diverse forms, Arabic characters hold interest for researchers, particularly in the realm of automatically recognizing handwritten characters [4].

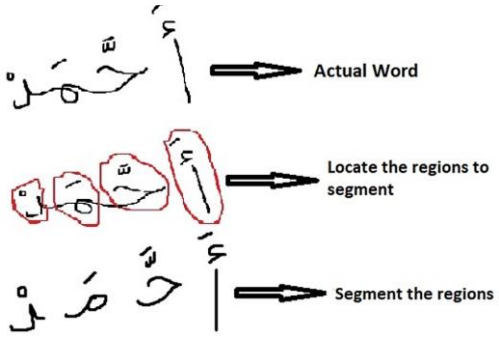

Figure 1. Example of image segmentation in Arabic handwritten word

2. BACKGROUND

There exist multiple approaches utilized for image segmentation, each with its own significance. Prominent segmentation techniques commonly employed encompass thresholding, edge detection, region-based, clustering, watershed, partial differential equation (PDE), and artificial neural network (ANN) methods, as illustrated in Figure 2. These techniques differ in their segmentation methods [5].

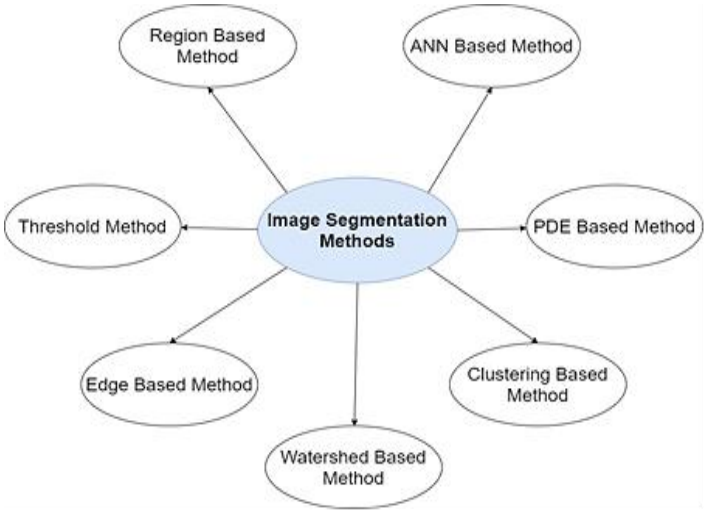

Figure 2. Classification of image segmentation techniques

2.1. Thresholding method

This method represents the most straightforward and uncomplicated approach to image segmentation. It distinguishes image pixels based on their intensity level. This method is suitable for images where foreground elements are lighter than the background. Moreover, there are three categories of thresholding: global thresholding, variable thresholding, and diverse thresholding. Furthermore, these three fundamental types of thresholding include [6], [7]:

a. Global thresholding: Employs a consistent threshold value (T) for the entire image.
b. Variable thresholding: Adapts threshold value (T) across the image.
c. Multiple thresholding: Utilizes multiple threshold values, like T0 and T1, determined via image histograms.

2.2. Edge based segmentation method

The edge-based segmentation techniques stand as well-established methods within image processing in their own right. These segmentation approaches, relying on rapid shifts in intensity values within an image, find their foundation in the fact that a sole intensity value does not adequately convey edge information. Two fundamental edge-based segmentation methods include gray histograms and gradient-based methods. To identify edges, basic edge detection techniques like Sobel, Canny, and Robert's operators can be employed. The result of these techniques yields a binary image essentially. These methods are grounded in structural techniques hinging on cut-off detection [8].

2.3. Region based segmentation method

The region-based segmentation method involves dividing the page into distinct regions sharing similar characteristics. It is a method for directly determining the region by iteratively encompassing neighboring pixels that share similarities and are connected to the seed pixel. Moreover, this approach encompasses a greater number of pixels to acquire additional information for characterizing the region. This method comprises two primary techniques [9]–[11]:

a. Region growing method: This method segments the image into various sections based on the expansion from seed pixels. These seeds can be manually assigned using prior knowledge or automatically designated according to specific applications.
b. Region splitting and merging method: This approach employs two fundamental techniques, splitting and merging, to divide the image into discrete areas. Region splitting involves iteratively separating image regions with similar attributes, while region merging combines neighboring comparable regions.

2.4. Clustering based segmentation method

The clustering-based segmentation technique involves dividing the image into clusters that contain pixels with similar attributes. Data clustering is a process that groups data elements into clusters, ensuring that elements within the same cluster are more similar to each other than to elements in other clusters. Two fundamental types of clustering exist [12], [13]:

a. Hard clustering: This method separates the image into a defined set of clusters, such that each pixel belongs to only one cluster. In other words, each pixel is exclusively assigned to a single cluster.
b. Soft clustering: This type of clustering is more organic, as strict separation is less likely due to the presence of noise. Soft clustering techniques are particularly useful for image segmentation scenarios where a clear division is not feasible. Thus, this technique offers greater flexibility compared to other methods [12].

2.5. Watershed based methods

The commonly employed watershed-based method utilizes the concept of topological transformation. In this approach, intensity levels correspond to bowls formed at their minima points, from which water spills. As water reaches the rim of a bowl, adjacent bowls are amalgamated. The watershed method views the image's gradient as a topographic surface, designating pixels with steeper gradients as continuous boundaries [14].

2.6. PDE based method

The segmentation techniques based on PDE are rapid methods suitable for time-sensitive applications. Two key PDE techniques are noteworthy: the non-linear isotropic diffusion filter, enhancing edges, and the curved non-quadratic variation restoration filter, eliminating noise. The outcomes of the PDE technique result in blurred edges and boundaries that can be adjusted using nearby operators. The fourth-order PDE method is employed to diminish image noise, while the second-order PDE technique is used for more accurate edge and boundary detection [12].

2.7. Ann based method

Artificial neural network-based segmentation techniques replicate the learning processes of the human brain to achieve decision-making objectives. Currently, this method is widely employed for segmenting medical images, effectively isolating the desired image from its background. A neural network comprises numerous interconnected nodes, with each connection assigned a specific weight. This approach operates independently of PDE. The problem is converted into challenges that are addressed by the neural network. This technique comprises two fundamental stages: feature extraction and segmentation using the neural network [10].

3. RELATED WORK

In this section, it is explained the related work for Arabic handwritten segmentation starting from categorization of existing segmentation algorithms which we compare the capability of each technique, we also come across the challenges of Arabic handwritten segmentation. Furthermore, diacritics segmentation for Arabic handwritten document is discussed. Finally, a summary of the latest segmentation techniques that are used in this field is elaborated.

3.1. Categorization of present segmentation methods

Numerous methodologies for image segmentation have been proposed, each with distinct characteristics. However, these methods can be classified based on the anticipated outcomes of the corresponding image segmentation algorithms. Noteworthy segmentation algorithms include: X–Y cuts or projection profiles, the run length smoothing algorithm (RLSA), document spectrum analysis, whitespace assessment, constrained text lines [15], the Hough transform used for skew detection and correction in mushaf, line detection within the image space [16], Voronoi tessellation, and scale space analysis. An experiment by Shiekh *et al.* [17] employed a region-based technique for diacritics segmentation and produced a promising result. The Watershed transform, which represents ridges dividing areas drained by distinct river systems [18], is instrumental in defining boundaries for image region separation [19], also aiding in detecting non-linear spacing between characters [20]. Contour tracing initiates from a source point and traces the character boundary to the point of contact [21]. It is important to note that the majority of the aforementioned segmentation algorithms are primarily tailored for existing documents. Table 1 outlines and illustrates how these algorithms are utilized in image preprocessing. Additionally, a comparison of various segmentation techniques is presented in Table 1, building upon the groundwork of [15]. We have contributed to this table, particularly focusing on the last three sections: region-based, watershed transform, contour tracing.

Table 1. Classification of existing segmentation methods

	Existing research							
Segmentation algorithms	Printed document	Hand-written documents	Historical documents	Page segmentation	Text line segmentation	Word segmentation	Character segmentation	Diacritic segmentation
X-Y cuts	√	√	√	√	√	√	√	
RLSA	√	√	√	√	√	√		
Docstrum	√			√	√			
Whitespace analysis	√			√	√			
Constrained text line	√			√	√			
Hough transform	√			√	√	√		
Voronoi	√			√	√			
Scale space analysis	√	√	√	√	√	√		
Region-based		√						√
Watershed transform		√			√		√	
Contour traced		√		√	√		√	

3.2. Challenges of Arabic handwritten segmentation techniques

The Arabic language, with its rich history and extensive vocabulary, has remained largely unchanged for a significant duration. However, due to its widespread usage, it is considered linguistically abundant. The challenge of segmenting Arabic handwritten text is huge, due to diverse handwriting styles and the intricate connections between Arabic letters. Despite this, a reliable solution to this challenging issue has yet to emerge [22]. Additionally, even though handwritten text varies from person to person, the segmentation of such text into words, lines, and characters remains a complex endeavor [23]. Handling non-Latin scripts like Arabic presents even greater challenges; aside from writing variations, the presence of diacritical marks and the abundance of ascender and descender characters further complicate the segmentation process. To address this intricacy and leverage it as an advantage, especially considering Arabic's semi-cursive nature, it is imperative to develop robust segmentation techniques specific to Arabic handwritten content. These techniques should address the difficulties by reviewing existing segmentation methods suitable for Arabic handwritten documents [24]. Consequently, numerous researchers are actively seeking solutions to this issue, achieving commendable progress; however, further research is warranted to enhance the existing approaches to segmenting Arabic handwritten content.

3.2.1. Diacritic segmentation for Arabic handwritten

Undoubtedly, diacritics hold significant importance in the Arabic language, and they can be added to letters in various positions, such as above or below the letter. In traditional Arabic writing, native speakers often do not rely heavily on diacritics, as they can comprehend the meaning from the context of the sentence. However, for non-native Arabic readers, the absence of diacritics can pose challenges in grasping word meanings, potentially leading to misinterpretation of the text. Nonetheless, from a technical standpoint, these diacritics also complicate the segmentation process due to their challenging detection. As a result, many research papers consider diacritics as noise or unnecessary features in the context of image processing for Arabic handwritten content. Consequently, diacritics are not extensively emphasized in studies of this nature.

A study carried out by *et al.* [17] undertook an experiment to extract diacritics from handwritten images using a region-based approach. The researcher's methodology involved, first, capturing the connected components within a binary image. Subsequently, they assessed the characteristics of image regions essentially, measuring the properties of each connected component within the binary image. Given the relatively small size of diacritics, the researcher determined the thresholding properties of region area based on measurements of low area [25]. This implies that smaller regional areas correspond to smaller objects, specifically diacritics in this case. Conversely, larger region areas correspond to more significant objects, such as primary elements like words and characters. The result of this experiment is presented in Figures 3 and 4. Figure 3 represents the binary image that will be used to extract diacritics features from, while Figure 4 presents the outcome of the study which is extracted diacritics objects which are in other sense called secondary objects. The main reason for conducting this study is to clarify that diacritics objects in Arabic handwritten are of a paramount importance and removing them from the image could lead to losing some important features for recognizing the document originality in Arabic handwritten [17].

Figure 3. Primary objects

Figure 4. Secondary objects (diacritics)

3.2.2. Summary of previous segmentation techniques for Arabic handwritten document

Numerous segmentation methods for Arabic handwritten documents have gained significant traction in various research endeavors. Despite the absence of a dependable technique for segmenting Arabic handwritten document and the challenges posed by the diverse styles of handwriting, we have undertaken a concise literature review of past segmentation techniques specific to Arabic handwritten documents, as shown in Table 2 (see in appendix). These techniques serve multiple purposes in relation to segmenting Arabic handwritten documents, aiming to contribute to effective image processing solutions.

4. CONCLUSION

This article delves into a range of topics concerning techniques for image segmentation. The paper undertook a classification of image segmentation algorithms primarily designed for modern documents, aiming to elucidate how these algorithms are applied in document processing. Additionally, the challenges posed by segmentation techniques for Arabic handwritten content were addressed, stemming from variations in writing style, the interconnectedness of Arabic characters, and strategies to manage the intricacies of Arabic cursive writing. Furthermore, the methods employed in segmenting Arabic script were explained, providing insight into their accuracy and average performance rates. To conclude, a comprehensive overview

of previously widely employed segmentation techniques for Arabic handwritten documents was presented. In the future, there are plans to conduct further reviews of segmentation techniques specific to Arabic handwritten content, with the aim of identifying robust methods that align with the objectives of this research.

APPENDIX

Table 2. Summary of previous segmentation techniques for Arabic handwritten

Ref.	Year	Method/Technique	Datasets	Findings
[26]	2019	seam carving	Public database, Arabic handwritten documents	It demonstrated a line detection accuracy of 97.5% at a matching score of 90%. This technique can be utilized across diverse writing styles, including both printed and handwritten forms.
[27]	2021	Machine learning	Randomly selected images with different angles from Arabic handwritten documents	The segmentation approach put forth attained an impressive accuracy rate of 98.18%. Moreover, the suggested method holds potential for expansion into a novel application, capable of being employed for recognizing Arabic handwritten text without concern for its logical or sentence structure.
[28]	2019	Novel algorithm	IFN/ENIT: database of handwritten Tunisian town names, and AHDB: a database for Arabic word spotting and recognition	It exhibits high accuracy levels with exceptional efficiency by mitigating the issue of excessive segmentation, which arises from open character segmentation, and by adapting the segmentation in scenarios involving character adjacency or inaccuracies in word image segmentation.
[29]	2019	Hough transform	IFN/ENIT and Agriculture and Horticulture Development Board (AHDB)	The effectiveness of the suggested method's performance hinges on a specific evaluation approach that adequately gauges the performance for comparative analysis with other researchers' segmentation results. In general, the method performed well and yielded elevated accuracy and efficiency levels in the conducted experiments.
[30]	2021	In this paper, a new segmentation technique has been created and developed	2,500 words of Arabic handwritten samples	The recently employed segmentation algorithm divides the handwritten content into a series of strokes, which are then organized into clusters that showcase the nature of the composed word.
[31]	2022	hybrid approach	KHATT: KFUPM handwritten Arabic text database, and IFN/ENIT	The outcomes of the suggested method demonstrated strong and effective performance in segmenting words. This approach proves to be both efficient and precise for segmenting interconnected and overlapping Arabic characters.
[32]	2021	Projection Profile	Moroccan Quranic manuscript	The results obtained in terms of line segmentation hold promise. However, there were subpar outcomes in character segmentation, stemming from challenges related to complexity.
[33]	2023	Baseline segmentation (BS)	Arabic words from open source	In this paper, baseline segmentation is employed to extract characters from Arabic words. In addition, conventional neural network (CNN) is implemented for classification, and it showed a high accuracy rate compared to other approaches.
[34]	2020	CNN	SUST-ARG: is gathered by the Arabic Recognition Research Group within the School of Computer Science and Information Technology at Sudan University of Science and Technology	The experimental findings indicate that the suggested CNN outperforms other recognition methods when applied to the same dataset. The accuracy achieved for testing names is 99.14%.
[17]	2020	Region-based	Arabic handwritten documents	The proposed method is tested with nine Arabic handwritten images with different style of writing and managed successfully to extract diacritics form these images.

ACKNOWLEDGEMENTS

The authors of this article express their gratitude to Universiti Teknikal Malaysia Melaka and Faculty of Information Technology and Communication for providing excellent research facilities.

REFERENCES

[1] A. A. A. Ali, M. Suresha, and H. A. M. Ahmed, "A survey on Arabic handwritten character recognition," *SN Computer Science*, vol. 1, no. 3, May 2020, doi: 10.1007/s42979-020-00168-1.

[2] K. V. Sánchez, "Functional-communicative grammar (Spanish-German) for translators and/or interpreters," *Babel. Revue internationale de la traduction/International Journal of Translation*, vol. 47, no. 2, pp. 109–120, Dec. 2001, doi: 10.1075/babel.47.2.03san.

[3] S. Aksoy, J. C. Tilton, and Y. Tarabalka, "Image segmentation algorithms for land categorization," *Remote sensing Handbook V. 1 Remotely sensed data characterization, classification, and accuracies*, pp. 317–342, Oct. 2015.

[4] T. Sari, L. Souici, and M. Sellami, "Off-line handwritten Arabic character segmentation algorithm: ACSA," in *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition*, 2002, pp. 452–457, doi: 10.1109/IWFHR.2002.1030952.

[5] D. Kaur and Y. Kaur, "Various image segmentation techniques: A review," *International Journal of Computer Science and Mobile Computing*, vol. 3, no. 5, pp. 809 – 814, 2014.

[6] Y.-J. Zhang, "An overview of image and video segmentation in the last 40 years," *Advances in Image and Video Segmentation*, pp. 1–16, 2011, doi: 10.4018/978-1-59140-753-9.ch001.

[7] T. Lindeberg and L. I. Meng-Xiang, "Segmentation and classification of edges using minimum description length approximation and complementary junction cues," *Computer Vision and Image Understanding*, vol. 67, no. 1, pp. 88–98, 1997, doi: 10.1006/cviu.1996.0510.

[8] M. Salem, S. Al-Amri, N. V Kalyankar, and S. D. Khamitkar, "Image segmentation by using edge," *International Journal on Computer Science and Engineering*, vol. 02, no. 03, pp. 804–807, 2010.

[9] A. M. Siddiqui, A. Ghafoor, and M. R. Khokher, "Image segmentation using multilevel graph cuts and graph development using fuzzy rule-based system," *IET Image Processing*, vol. 7, no. 3, pp. 201–211, 2013, doi: 10.1049/iet-ipr.2012.0082.

[10] N. Senthilkumaran and R. Rajesh, "Image segmentation - A survey of soft computing approaches," in *ARTCom 2009 - International Conference on Advances in Recent Technologies in Communication and Computing*, 2009, vol. 1, no. 2, pp. 844–846, doi: 10.1109/ARTCom.2009.219.

[11] S. Angelina, L. P. Suresh, and S. H. K. Veni, "Image segmentation based on genetic algorithm for region growth and region merging," in *2012 International Conference on Computing, Electronics and Electrical Technologies, ICCEET 2012*, 2012, pp. 970–974, doi: 10.1109/ICCEET.2012.6203833.

[12] M. Yambal and H. Gupta, "Image segmentation using fuzzy C means clustering: a survey," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, no. 7, pp. 2927–2929, 2013.

[13] V. K. Dehariya, S. K. Shrivastava, and R. C. Jain, "Clustering of image data set using k-means and fuzzy k-means algorithms," in *Proceedings - 2010 International Conference on Computational Intelligence and Communication Networks, CICN 2010*, 2010, pp. 386–391, doi: 10.1109/CICN.2010.80.

[14] W. X. Kang, Q. Q. Yang, and R. P. Liang, "The comparative research on image segmentation algorithms," in *Proceedings of the 1st International Workshop on Education Technology and Computer Science, ETCS 2009*, 2009, vol. 2, pp. 703–707, doi: 10.1109/ETCS.2009.417.

[15] N. Nikolaou, M. Makridis, B. Gatos, N. Stamatopoulos, and N. Papamarkos, "Segmentation of historical machine-printed documents using adaptive run length smoothing and skeleton segmentation paths," *Image and Vision Computing*, vol. 28, no. 4, pp. 590–604, Apr. 2010, doi: 10.1016/j.imavis.2009.09.013.

[16] S. S. Bafjaish, M. Sanusi, M. Nasser, A. Ramzani, and H. Mahdin, "Skew detection and correction of mushaf Al-Quran script using hough tTransform," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 8, 2018, doi: 10.14569/IJACSA.2018.090852.

[17] A. A. Shiekh, 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, Apr. 2020, doi: 10.11591/ijeecs.v18.i1.pp478-484.

[18] A. Souhar, Y. Boulid, E. Ameur, and M. Ouagague, "Segmentation of Arabic handwritten documents into text lines using watershed transform," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 6, 2017, doi: 10.9781/ijimai.2017.08.002.

[19] K. Oh, S. Kim, I. Na, and G. Kim, "Text line segmentation using AHTC and watershed algorithm for handwritten document images," *International Journal of Contents*, vol. 10, no. 3, pp. 35–40, Sep. 2014, doi: 10.5392/IJoC.2014.10.3.035.

[20] A. S. Kavitha, P. Shivakumara, G. H. Kumar, and T. Lu, "A new watershed model based system for character segmentation in degraded text lines," *AEU - International Journal of Electronics and Communications*, vol. 71, pp. 45–52, Jan. 2017, doi: 10.1016/j.aeue.2016.11.007.

[21] I. Ullah, M. Sanusi, M. Ishak, and Y. M., "Segmentation of touching Arabic characters in handwritten documents by overlapping set theory and contour tracing," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, 2019, doi: 10.14569/IJACSA.2019.0100519.

[22] S. Wshah, Z. Shi, and V. Govindaraju, "Segmentation of Arabic handwriting based on both contour and skeleton segmentation," in *2009 10th International Conference on Document Analysis and Recognition*, 2009, pp. 793–797, doi: 10.1109/ICDAR.2009.152.

[23] A. M. M. Algaradi, M. S. Azmi, I. E. A. Jalil, A. F. A. Hashim, and A. A. M. Al-Malki, "The trend of segmentation for Arabic handwritten touching characters," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022, doi: 10.14569/IJACSA.2022.0130159.

[24] P. Xiu, L. Peng, X. Ding, and H. Wang, "Offline handwritten Arabic character segmentation with probabilistic model," *International Workshop on Document Analysis Systems*, pp. 402–412, 2006, doi: 10.1007/11669487_36.

[25] A. A. Shiekh, M. S. Azmi, M. A. Aziz, M. N. Al-Mhiqani, and S. S. Bafjaish, "Framework of diacritic segmentation for Arabic handwritten document," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 24, no. 2, pp. 1001–1008, Nov. 2021, doi: 10.11591/ijeecs.v24.i2.pp1001-1008.

[26] M. Daldali and A. Souhar, "Handwritten Arabic documents segmentation into text lines using seam carving," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, 2019, doi: 10.9781/ijimai.2018.06.002.

[27] S. I. Saleem, A. M. Abdulazeez, and Z. Orman, "A new segmentation framework for Arabic handwritten text using machine learning techniques," *Computers, Materials and Continua*, vol. 68, no. 2, pp. 2727–2754, 2021, doi: 10.32604/cmc.2021.016447.

[28] A. A. A. Ali and M. Suresha, "An efficient character segmentation algorithm for recognition of Arabic handwritten script," in *2019 International Conference on Data Science and Communication, IconDSC 2019*, 2019, pp. 1–6, doi: 10.1109/IconDSC.2019.8817037.

[29] A. A. A. Ali and M. Suresha, "Efficient algorithms for text lines and words segmentation for recognition of Arabic handwritten script," *Advances in Intelligent Systems and Computing*, vol. 882, pp. 387–401, 2019, doi: 10.1007/978-981-13-5953-8_32.
[30] 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.
[31] O. A. Boraik, M. Ravikumar, and M. A. N. Saif, "Characters segmentation from Arabic handwritten document images: hybrid approach," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 4, pp. 395–403, 2022, doi: 10.14569/IJACSA.2022.0130447.
[32] A. Alghamdi, D. Alluhaybi, D. Almehmadi, and S. Bin Siddeq, "Text segmentation of historical Arabic handwritten manuscripts using projection profile," in *2021 National Computing Colleges Conference (NCCC)*, 2021, pp. 1–6.
[33] A. Belabiod and A. Belaïd, "Line and word segmentation of Arabic handwritten documents using neural networks," Diss. LORIA-Université de Lorraine; READ, 2018.
[34] M. E. Mustafa and M. Khalafallah, "A deep learning approach for handwritten Arabic names recognition," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 1, 2020, doi: 10.14569/IJACSA.2020.0110183.

BIOGRAPHIES OF AUTHORS

Ahmed Abdalla Sheikh received his BSc in computer science (database management system) in 2017 and MSc in computer science (software engineering and intelligent) from the Universiti Teknikal Malaysia Melaka (UTeM) in 2019. His research interests include image processing, image segmentation, machine learning, and computer vision systems. He can be contacted by this email: ashiekh295@gmail.com.

Mohd Sanusi Azmi is an associate professor Department of Software Engineering, Universiti Teknikal Malaysia Melaka (UTeM). He received BSc. M.Sc. 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 to the feature extraction domain. He has proposed a novel technique based on geometry features used in Digit and Arabic based handwritten documents. He can be contacted by this email: sanusi@utem.edu.my.

Waleed Abdel Karim Abuain received his PhD in artificial intelligence in 2016 from the Artificial Intelligence Technology Center of the National University of Malaysia, Department of Computer Science (UKM). Since 2016, he is currently working as a full-time assistant professor in the Department of Computer Science at Taibah University, Applied College, Saudi Arabia. His research interests include artificial intelligence, machine learning and optimization, data science, computer vision, deep learning, and the internet of things. He can be contacted by this email: wabuain@taibahu.edu.sa.

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. She can be contacted by this email: maslita@utem.edu.my.