An Improved Algorithm for Optical Character Recognition using Graphical User Interface Design

SHAHID MANZOOR[1], NIMRA WAHAB[1], M. K. A. AHAMED KHAN[2]
[1]Department of Electrical and Electronic Engineering,
Faculty of Engineering, Technology & Built Environment,
UCSI University, Kuala Lumpur, 56000,
MALAYSIA

[2]Departemnt of Mechanical Engineering,
Faculty of Engineering, Technology & Built Environment,
UCSI University, Kuala Lumpur, 56000,
MALAYSIA

Abstract: - Since the COVID-19 pandemic, numerous jobs have become necessary, including the storing and sharing of printed material across computers. One simple way to save data from printed papers to a computer system is to scan them first and then save them as images. However, it would be quite challenging to extract or query text or other information from these photo files to reuse this information. As a result, a method for automatically retrieving and storing information, particularly text, from picture files is required. Optical character recognition (OCR) is an ongoing research topic that aims to create a computer system capable of extracting and processing text from images. To accomplish successful automation, certain significant problems must be identified and addressed. The font properties of characters in paper documents, as well as image quality, are only a few of the latest problems. Characters may not be recognized correctly by the computer system because of many complexities. So, in this study, authors look into OCR in four different contexts and apply them to get our results. However, every OCR is further followed by these two steps. First, a comprehensive explanation of the challenges that may develop during the OCR phases is provided. The key phases of an OCR system are then executed, including pre-processing, segmentation, normalization, feature extraction, classification, and post-processing. It can be used with deep learning software to provide OCR data which is very useful for robotic and AI applications.

Key-Words: - Optical Character Recognition, Image Processing, Feature extraction, Segmentation, AI, RGB, Deep learning, Automatic number plate recognition, Graphical user interface.

Received: July 20, 2022. Revised: October 24, 2023. Accepted: November 27, 2023. Published: December 31, 2023.

1 Introduction

The basic or minimum requirements for data storage can be accomplished by paper files. You may store critical files in hand by reach and if well cataloged you'll find what you are looking for with ease. But there is a major catch to this, just like everything else around us, [1], [2], [3]. There are several reasons why data should not be stored on paper. Whether we're addressing privacy, the environment, or efficiency. This is where Optical Character Recognition comes into play, [4], [5].

OCR is an auspicious technique for converting handwritten letters or phrases into a digital version, [6], [7] as shown in Figure 1. It is a typical technique of digitizing printed texts so that they may be altered, searched, saved more compactly, and shown on the internet, [8].

Optical character recognition is one of the most intriguing and demanding topics in pattern recognition in the electronic era. This recognition system has evolved to include not just printed and handwritten characters, but even offline characters, [9], [10], [11]. Steps involved in a recognition of a system for OCR are crucial. Aside from that, to attain a good-prognosis, a recognition system relies heavily on a well-defined feature extraction method and a robust classifier, [12], [13], [14].

Presently, there is a great deal of interest in transferring the data contained in these paper records to a PC storage drive and then reusing this data via a searching procedure, [15], [16]. The image record can be composed by hand and analyzed by photograph, [17].

It translates pictures into an indisputable machine-encoded editable substance, [9], [10]. It sees simply those characters for which the system has been arranged using an unequivocal course of action calculation. There are numerous algorithms used and tried for English transcribed content, [11], [12], [14]].

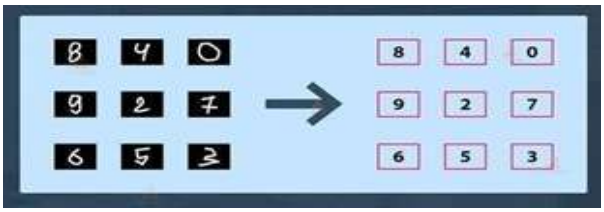

Fig. 1: Representation of OCR

In this paper, we investigate four OCR with different features of extraction, segmentation, and pre and post-image processing methods. First, we investigate the difficulties that may arise during the OCR phases. Furthermore, we create a graphical user interface (GUI) for the OCR system so that it can be applied to various processing and application tasks. These responsibilities include pre-processing, segmentation, normalization, feature extraction, classification, and post-processing. The findings demonstrate that the technology used in this paper was able to successfully retrieve data from both images and text. The literature review can be found in the coming section. Afterword, the System Design is discussed, followed by the "Methodology", and "Results Discussion" sections. Then finally the results are concluded which shows the layout of the system.

2 Literature Review

2.1 History of OCR

It appears that the first Optical Character Recognition device was developed in the late 1920s as presented in Figure 2 by the Austrian engineer Gustav Tauschek (1899-1945), who received a Reading Machine patent in Germany in 1929, followed by Paul Handel, who received a Statistical Machine patent in the United States. A text-containing image was passed before the window of the reading machine. The comparison device was a rotating disc or wheel with letter-shaped holes that had been housed within the objective lens. When the shape of the images and letter-shaped holes matched, the clockwork rotated the printing drum to the corresponding letter and printed it, [8]. The machine is shown in Figure 2.

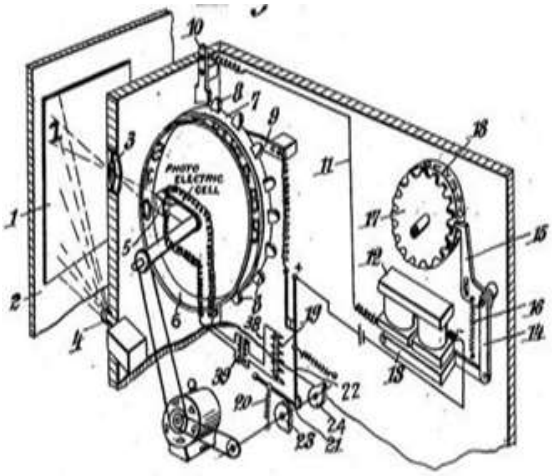

Fig. 2: The first machine invented for OCR

2.2 Isolated Versus Cursive

In languages with independent content, the characters do not combine. In cursive script, however, the situation is just the opposite. In cursive script, adjacent characters in words not only join together but also alter their shape based on their position, [9]. This adds a great deal of complexity and complication to the recognition process, necessitating an additional level of division to separate the characters in each word before proceeding. This may not be difficult to accomplish for various languages, therefore segmentation-free approaches are offered on multiple occasions. The segmentation-free approach aims to comprehend the word without separating it into individual letters.

2.3 Offline Vs Online

Off-line text recognition is attempting to recognize text that has previously been acquired in the form of pictures. A digital device such as a camera or scanner is often used to scan text data into the computer. Online recognition refers to real-time recognition while the user moves a pen to write anything. Text input through online recognition needed specialized gear, such as a tablet and a stylus. Online identification is regarded as less complicated because temporal information from pen traces, such as its speed and velocity of use, how pen lifts and what writing order, etc., is readily available, [8]. The first sort of system is one with a restricted vocabulary, such as postal address recognition and geographic names as represented in Figure 3.

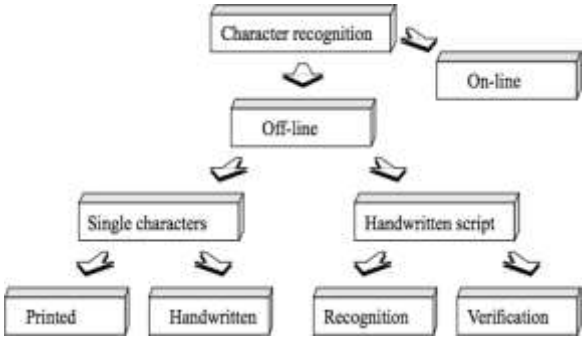

Fig. 3: Flow chart for online vs offline recognition

2.4 Street View House Numbers SHVN

Street View House Numbers is a certified image dataset that may be used to create and build AI and article recognition algorithms and computations with little data and information processing, formatting, and arranging requirements. This is an informative index of house numbers derived from Google Street View, as the name implies and presented in Figure 4. The job complexity is in the middle. The numbers come in a variety of forms, fonts, and composing styles, but each home number is positioned in the center of the view, so placement is not required. One disadvantage of SVHN is that the pictures are not high resolution and may be arranged in an unusual manner, [9], [10].


Fig. 4: Portrayal of OCR works on House Numbers

2.5 CAPTCHA

Since the internet is full of machines, vision tasks, primarily text decoding or CAPTCHA are a very popular way to tell them apart from humans and robots. Many of these documents are jumbled and twisted, making it difficult for a machine to decipher them as observed from Figure 5. It is doubtful if the creators or researchers of the CAPTCHA anticipated advancements in computer vision, but most texts are not difficult to answer nowadays, [11].

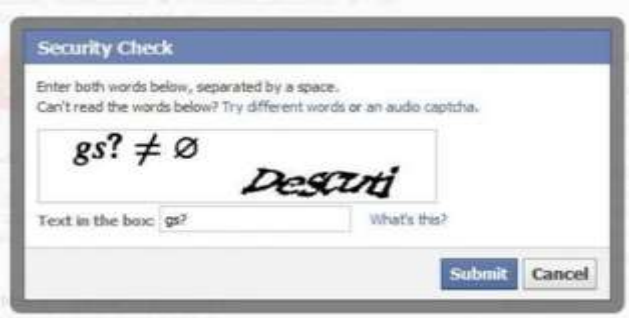

Fig. 5: Captcha to verify between Human and Robot

2.6 License Plates

Another challenge for OCR is License Plate Recognition, it isn't very hard but as to surveys it is a useful practice. As with most of the OCR jobs and tasks, this one involves detecting the license plate and identifying the characters on it as presented in Figure 6. Since the outline of the plate is relatively stable, some approaches rely on a basic reshaping process before identifying the process.

Fig. 6: License plate recognition

3 System Design

3.1 Picture Acquisition

Image acquisition is the act of converting a physical image into a digital format that can be controlled by computers. There are a variety of image sources for collecting pictures, such as scanning through a device or photographing them with an attached camera. The source image can be of different natures such as hand-written documents, printed documents, typewritten documents, etc.

3.2 Refining

After image acquisition processing or refining objects is to improve the nature or the quality of the

process. One of the refining process techniques is thresholding which aims to parallel the picture in light of some limited esteem. Various types of channels and filters can be applied such as min filter, max filter, averaging, etc. Other than that more unique activities like morphological operations, erosion, dilation, opening, and closing too can be performed.

3.3 Thresholding

Typically, the picture obtained from an examination or other acquisition procedure is in RGB or Indexed format. Such images contain color information for each pixel; for instance, RGB images contain RED, GREEN, and BLUE color values within the range 0-255 for each pixel, in addition to the pixel's intensity value. Once the images have been converted to grayscale, they contain data that has minimal influence on the identification process but adds unnecessary complication, [3]. This grayscale image is turned into a bi-level image to detect whether a pixel belongs to the foreground or background. Each pixel in a bi-level picture has just one of two potential values, either 1 or 0.

3.4 De-skewing

When text lines get tilted as a result of poor photocopying and scanning, the paper introduces skewness in the document. Preprocessing may also provide skewness correction which means taking care of adjusting the errors. This is referred to as de-skewing. Since a distorted text may not be adequately processed during the segmentation period, this de-skewing phase is critical. Another related concern is detecting orientation. It aids in the de-skewing of the data.

3.5 Feature Extraction

After isolating characters, the next step is to obtain as much detail as possible from each one to distinguish it from the others. Figure 7 represents feature extraction as another term for all this. All of the character's attributes are filtered out in this process. Four types of features are extracted mainly during the OCR process namely: Structural, Statistical, transformations, and matching of the basic templates and Correlation. Structural based Features consist of curves, endpoints, intersection points, zigzags, dots, loops, strokes. Other information such as direction, slope of the loops, and length can also be extracted. Statistical Features include pixel densities, zoning, crossing, moments, Fourier descriptors, and moments which are numerical indices of image regions. In Global transformations, the transformation transforms the pattern's pixel representation to a more abstract and compact shape.

We begin by normalizing all binary character images to a N×N matrix while keeping the original aspect ratio in mind. N is set to 60 in our example. We use the two sorts of characteristics that are given. Zones criteria are used for the first set of characteristics. The input image is split into two zones, vertical and horizontal zones and for each zone, the density of the character pixels is calculated.

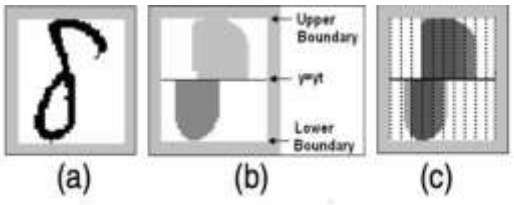

Fig. 7: Feature Extraction

3.6 Post Processing

OCR accuracy may be enhanced with the use of contextual analysis. The image's geometric and document context can be used to improve accuracy. To further enhance OCR accuracy, lexical processing using Markov models and dictionaries can be applied. There is also the option of using contextual interpretation to boost OCR effectiveness. The image's connotations of geometry and paper can aid in the elimination of errors.

4 Methodology

The OCR software includes the use of a graphical user interface. We'll start by designing the user interface. Write 'guide' in the command pane. A new box will appear. The graphical user interface (GUI) can be tailored to your requirements. This will assist you with creating a template that is simple and meets your needs. In this window, we will design our buttons and screens for input and output. Three push buttons of our GUI design were used: Input, Text Detection, and Output. And two screens show the display after the image has been chosen.

At first, all of the letters and numbers are stored separately in a folder on our desktop for character recognition. For practice, all of the alphabets in capital and small letters, as well as numbers 0-9, for training purposes are being used. For example- M, m, N, n, etc.

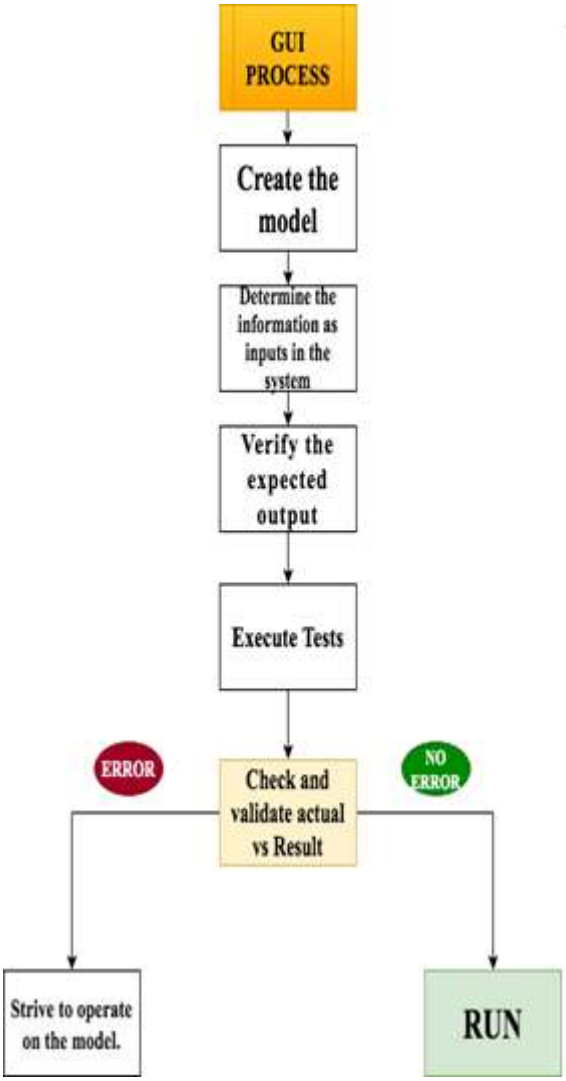

Fig. 8: Flow chart for GUI Processing Design

4.1 GUI Process

Graphical User Interface (GUI) verification is the practice of testing an application's user interface. Menus, checkboxes, buttons, colors, fonts, sizes, icons, text, and pictures are all part of a graphical user interface. GUI testing is performed as indicated in Figure 8 to ensure the functioning and usability of design components as a user of an application under test. Most online exchanges are not intended to be definitive, but rather to enable and actively support growth. GUI testing is critical to the software's successful release since it verifies the user experience. GUI testing contributes to the delivery of high-quality, user-friendly software. In the end, users have increased customer acquisition and satisfaction.

By the design requirements and objectives Initially, a test script was created to follow and read all of the stages. Two windows are included in the design below. Then proceeding GUI will read the picture and detect each letter or number.

5 Results and Discussion

To begin, we will select an image for the folder as an example. The picture is next subjected to the text detection command, which recognizes each letter separately. The first example's image reads out the basic sentence "HOW ARE YOU?" as shown in Figure 9. The output, as shown in the Figure 10, reads the supplied sentence without mistake. the code is successful, and the GUI display the correct detected text.

Fig. 9: Text test input

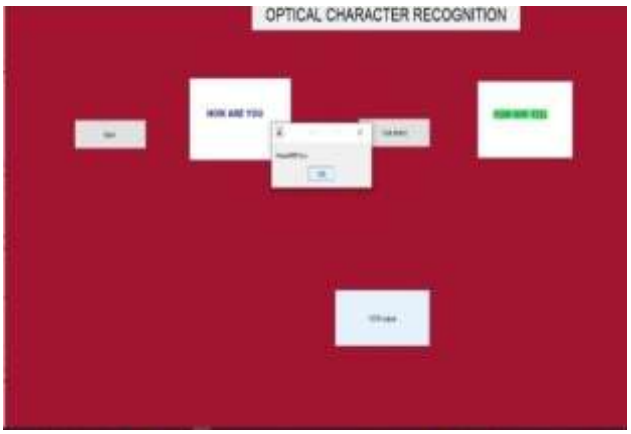

Fig. 10: Coded GUI for processing image data

In the subsequent case, we'll examine license plates. This will allow us to display both numbers and letters at the same time. Automatic number plate recognition (ANPR) is another term for recognizing license plates. The system is installed at the entrance for security management of a highly restricted region such as military zones or areas around major government buildings such as Parliament, the Supreme Court, and so on. The created technology identifies the vehicle and then captures a picture of it.

Fig. 11: Image for identifying number plate

This concept was developed in 1976 at the Police Scientific Development Branch in the United Kingdom. However, withthe advancement of digital cameras and the rise in computing capability over the previous decade, it has garnered a lot of attention. It is the capacity to automatically extract and recognize the characters of a car license plate from a picture. In essence, it comprises a camera or frame grabber capable of capturing a picture, identifying the location of the number in the image, and then extracting the characters for a character recognition tool to convert the pixels into numerically readable characters. ANPR may be utilized for a variety of purposes, ranging from speed enforcement and tool collection to parking lot management, etc. It may also be used to identify and preventa wide range of criminal actions, as well as to regulate security in highly restricted locations such as military zones or regions around important government offices.

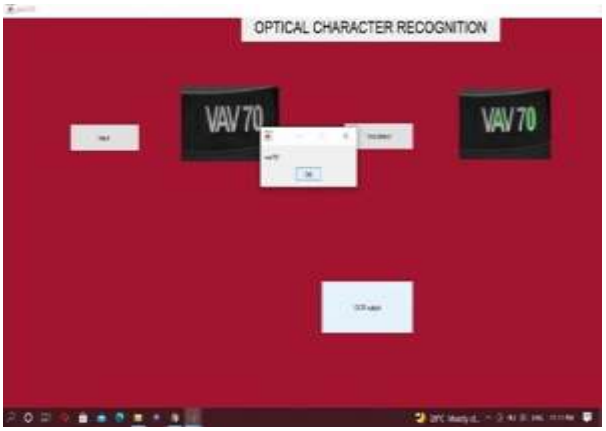

Fig. 12: GUI processes the input image of Figure 11

To begin, the camera is connected to the PC through Matlab®. The camera is connected through the USB port.

Various photos of automobiles with various colors and construction kinds are captured and saved on a computer. During the processing, the various impacts of the daylighting are also evaluated. As seen in Figure 3, the pictures are in RGB format and have a resolution of 800 600 pixels.

To detect the car number, an optical character recognition (OCR) method is used. The cropped picture obtained in the second phase is inverted, which means that all white pixels areturned to black and all black pixels to white. The lettering is now white, while the backdrop of the plate is black. Individual lines in the text are separated using the line separation technique before using OCR. Number plate recognition is done for two car number plates as shown in Fig. 11, stationary position, and fig.13, moving car. The results of GUI processing are shown in Fig. 12 and Fig. 14 for successful extraction of the number plate.


Fig. 13: Moving car input number plate picture

Fig. 14: GUI processes the image input of Fig. 13

The line separation sums the pixel values of each row. If the resultant sum of the row is zero, the row contains no text pixels; if the consequent sum of the row is more than zero, the row contains text. The line begins with the first resultant sum greater than zero, and ends with the first resultant sum equal to zero. The start and end values of the line are used to crop the first line of text. A similar method is utilized to divide the second line of text. After separating the lines in an extracted car license plate, the column-by-column line separation process is done to separate the individual characters. The separated characters are then stored in separate variables. Currently, OCR is applied to compare each character to the whole alphabetic database.

6 Conclusions and Further Work

Given the fact that several algorithms, methods, and strategies for optical character recognition in scene images have been developed, there are not enough literature reviews on this subject. We presented an

arrangement of various approaches, algorithms, and strategies in this work. It is envisioned that this comprehensive assessment would give insight into the ideas involved and, maybe, stimulate future progress in the field. To begin, we discussed the major challenges of OCR, followed by a detailed discussion of the main important phases, architecture, proposed algorithms, and techniques of OCR. We emphasize that when designing any application related to OCR, one must pay close attention to each phase to achieve a highly accurate character recognition rate. However, we are unable to provide detailed methods for each phase since they are dependent on datasets, application particulars, and parameter specifications. Finally, important OCR applications and a brief OCR history are presented.

Although state-of-the-art OCR allows for high-accuracy text recognition, we believe that OCR has many more useful uses. We want to employ OCR in the future for such practical applications for everyday personal usage. We intend to combine mobile devices with OCR in a single OCR solution. Some of our upcoming OCR-based applications include an automatic book reader and a receipt tracker, [15], [17]. OCR is no longer just matching or seeing, with new technology of deep learning it is now entered into a new phase and it can recognize the text after scan and then convert it to different meaning in full applications[15]. With deep learning software, it provides more robust extraction of information and high-quality insight and it also can be used in the robotics field and integrated for artificial intelligence applications like Chat-GPT, [16], [17].

References:

[1] Satti, D. A., Offline Urdu Nastaliq OCR for Printed Text using Analytical Approach. *MS thesis report Quaid- i- Azam University*: Islamabad, Pakistan. 2013, [Online]. https://www.cle.org.pk/research/theses.htm (Accessed Date: January 24, 2023).

[2] Mahmoud, S.A., & Al-Badr, B., Survey and bibliography of Arabic optical text recognition. *Signal processing*, 41(1), 2017, pp 49-77.

[3] Bhansali, M., & Kumar, P, An Alternative Method for Facilitating Cheque Clearance Using Smart Phones Application. *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, 2(1), 2018, pp 211-217.

[4] Qadri, M.T., & Asif, M, Automatic Number Plate Recognition System for Vehicle Identification Using Optical Character Recognition presented at *International IEEE Conference on Education Technology and Computer, Singapore*, 2019. Singapore:

[5] A.S. Abutaleb. Automatic thresholding of gray-level pictures using two-dimensional entropy. *Computer Vision, Graphics, and Image Processing*, 47 (1): pp 22–32, 2014.

[6] A.M. AL-Shatnawi, K. Omar, and A.M. Zeki. Challenges in thinning of Arabic text. *In ICGST International Conference on Artificial Intelligence and Machine Learning (AIML-11)*, Dubai. United Arab Emirates, pages 127–133, 2016.

[7] Lazaro, J., Martín, J.L, Arias, J., Astarloa, A., & Cuadrado, C, Neuro semantic thresholding using OCR software for high precision OCR applications. *Image and Vision Computing*, 28(4), 2016, pp.571-578.

[8] Lund, W.B., Kennard, D.J., & Ringger, E.K., Combining Multiple Thresholding Binarization Values to Improve OCR Output presented in *Document Recognition and Retrieval XX Conference*, California, USA, 2015. USA: SPIE.

[9] Shaikh, N.A., & Shaikh, Z.A, A generalized thinning algorithm for cursive and non-cursive language scripts presented in *9th International Multitopic Conference IEEE INMIC*, Pakistan, 2015. Pakistan.

[10] Shaikh, N.A., Shaikh, Z.A., & Ali, G, Segmentation of Arabic text into characters for recognition presented in *International Multi Topic Conference, IMTIC*, Jamshoro, Pakistan, 2018.

[11] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*. Series B (Methodological), pages 1–38, 2019.

[12] Mochurad, L. "A New Approach for Text Recognition on a Video CARD." *Computer systems and information technologies*, 2022, https://doi.org/10.31891/csit-2022-3-3.

[13] Karthick, K., Ravindrakumar, K.B., Francis, R., & Ilankannan, S., Steps Involved in Test Recognition and Recent Research in OCR ; A Study, 2019, Corpus ID: 212448703.

[14] Faizullah, S., Ayub, M.S., Hussain, S., & Khan, M.A., A Survey of OCR in Arabic Language: Applications, Techniques, and Challenges. *Applied Sciences*, 2023, 13(7), 4584; https://doi.org/10.3390/app13074584.

[15] Adyanthaya, S.K., Text Recognition from

Images: A Study, *International of engineering research and technology (IJERT), NCCDS* 2020, vol.8, Issue 13, pp. 118-120.

[16] Mittal, R., & Garg, A.. Text extraction using OCR: A Systematic Review. 2020 Second International Conference on *Inventive Research in Computing Applications (ICIRCA)*, 357-362.

[17] Pal, M., & Santra, D,. Handwritten Character Recognition. *SSRN Electronic Journal*, 2023, https://dx.doi.org/10.2139/ssrn.4539983.

Contribution of Individual Author to the Creation of a Scientific Article

- Shahid Manzoor provided the main idea and designed the methodology and formulation of the project. Finalized the simulation results using MATLAB®.
- Nimra Wahab executes the design and constructs the main MATLAB® codes and completes the GUI design.
- M. K. A. Ahmed Khan was involved in constructive discussions and formulation of ideas and analysis throughout the project.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare.