Steps Involved in Text Recognition and Recent Research in OCR; A Study

K.Karthick, K.B.Ravindrakumar, R.Francis, S.Ilankannan

Abstract: ***The Optical Character Recognition (OCR) is one of the automatic identification techniques that fulfill the automation needs in various applications. A machine can read the information present in natural scenes or other materials in any form with OCR. The typed and printed character recognition is uncomplicated due to its well-defined size and shape. The handwriting of individuals differs in the above aspects. So, the handwritten OCR system faces complexity to learn this difference to recognize a character. In this paper, we discussed the various stages in text recognition, handwritten OCR systems classification according to the text type, study on Chinese and Arabic text recognition as well as application oriented recent research in OCR.***

Index Terms: Edge detection, Optical Character Recognition, OCR, Preprocessing stages, Text Recognition

I. INTRODUCTION

The technologyadventfoundan amazing and noble development curve for the last two centuries. For the last few decades, it is easy in using mouse and keyboard to assist as interfacing device between us and computer. However, the human-based communications ability to interact with a computer would make things easier to handler, but would be difficult to succeed for the investigators / researchers. Thepioneering achievements due to continuous research in man-machine communication may bring the scenario like interactions between human. Various techniques using magnetic stripe, speech recognition, identification using radio frequency, bar code, and Optical Mark Recognition (OMR), and OCR fulfill the automation needs in various applications. In this paper, we discussed about the handwritten OCR systems classification and the steps involved in OCR that is one of the automatic identification technique.

Revised Manuscript Received on 30 May 2019.
* Correspondence Author
K.Karthick*, Associate Professor, Department of EEE, GMR Institute of Technology, Rajam.
K.B.Ravindrakumar ,Professor, Department of EEE, Vel Tech Multi Tech Dr.RangarajanDr.Sakunthala Engineering College, Avadi,
R.Francis Assistant Professor, Department of Electrical Engineering, Annamalai University, Chidambaram,
S.Ilankannan,Research Scholar, Department of Electrical Engineering,Annamalai University.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

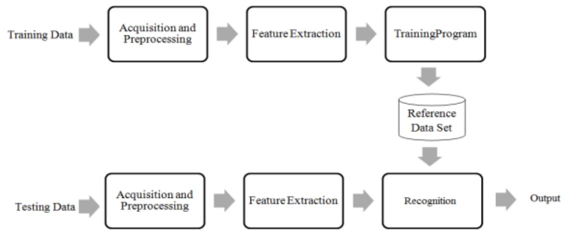


Fig.1 Steps Involved in Text Recognition

The low-level image processing involves primitive operation such as to reduce noise, contrast, enhancement and image sharpening. The major steps involved in text recognition system are shown in Figure 1.They are,

- Preprocessing
- Feature Extraction
- Recognition
- Post processing

In addition to the above steps segmentation and morphological processing also involved in the recognition process. These steps may be added before the feature extraction process.

II. PREPROCESSING

The preprocessing is a fundamental stage that is proceeding to the stage of feature extraction; it regulates the appropriateness of the outcomes for the consecutive stages. The OCR success rate is contingent on the success percentage of each stage.

A. Factors Affecting the Text Recognition Quality

Many factors influence the precision of character recognized using OCR. The factors are scan resolution, scanned image quality, printed documents category either photocopied or laser printer, quality of the paper, and linguistic complexities. The uneven illumination and watermarks are few factors faced in OCR system that influence the accuracy of OCR.

B. Significance of Preprocessing in Text Recognition

The preprocessing step is necessary to obtain better text recognition rate, using efficient algorithms of preprocessing creates the text recognition method robust using noise removal, image enhancing process, image threshold process, skewing correction, page and text segmentation, text normalization and morphological operations.

C. Preprocessing methods

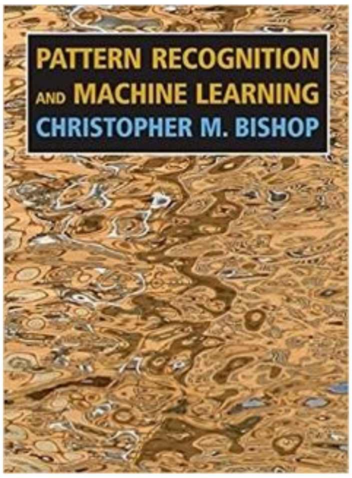
The majority of OCR application uses binary / gray images. The images mayhave watermarks and/or non-uniform background that make recognition process difficult without performing the preprocessing stage. There are several steps needed to achieve this. The initial step is to adjust the contrast or to eliminate the noise from the image called as the image enhancement technique. The next step is to do thresholding for removing the watermarks and/or noise followed by the page segmentation for isolating the graphics from the text. The next step is text segmentation to individual character separation followed by morphological processing. The morphological processing is required to add pixels if the preprocessed image has eroded parts in the characters.

D. Techniques involved in Image Enhancement

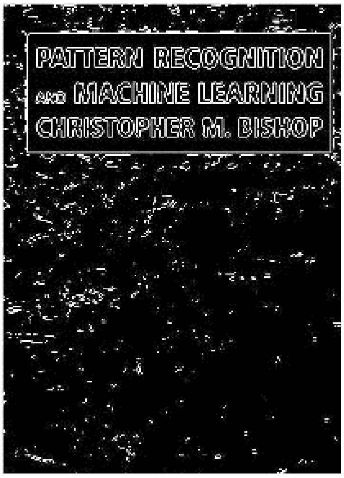
Image enhancement increasesimage quality for perception of humans by increasing contrast, minimizing blurring and removing noise (Nithyananda et al. 2016).

E. Spatial Image Filtering

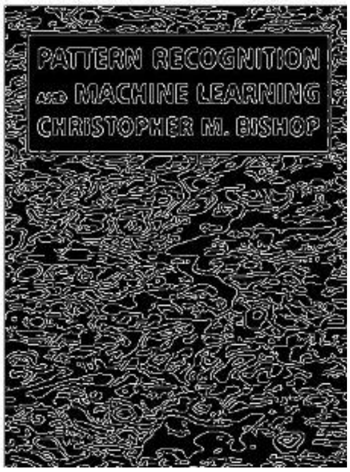
The filters are applied to defeat the high or low frequencypresent in the image. Eliminating the high frequencies in the image is smoothing, and the low frequencyelimination is enhancing or edge detection in the image. The following figure 2(a) shows the original image and 2 (b) & (c) shows the images applied withPrewitt and Canny edge detection methods. These filtering techniques may give effective text detection from images available in natural scene.



(a)



(b)



(c)

Fig.2 Edge Detection (a) Original Image(b) Prewittmethod (c) Cannymethod

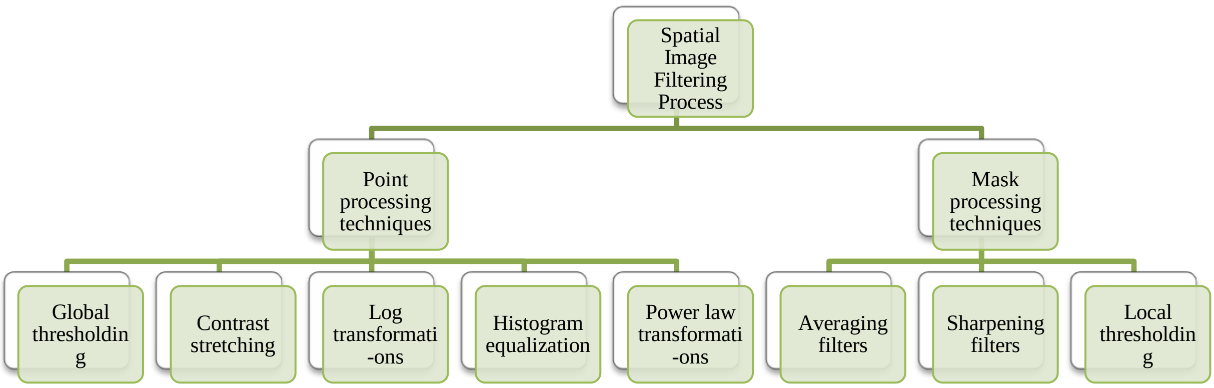


Fig. 3 Spatial filtering category

Spatial image filtering processing is categorized into mask and point processing. Figure 3 shows the spatial filtering categories. Point processing (Yi & Li 2013) adjusts the pixel value present in the image to generate the corresponding pixel value in the enhanced output image is expressed as,

$$O\ (a, b) = T\ [I\ (a, b)] \quad (1)$$

Where, I (a, b) is the input image, O (a, b) is the enhanced output image and T denotes the transformation among them.

F. Techniques of Point processing

Global thresholding: Image thresholding is the method of isolating the information from its background. Hence, this method is usednormallyto grey-level, or scanned color images and itis categorized as global and localthresholding. Globalmethod of thresholding chooses avalue of threshold for the completeimage from the intensity histogram (Mori 2010). Global thresholdingautomatically reduces a grey-level image to a binary image. The local adaptive thresholding method for each pixel it uses different values based on the information of local area. The figure 4 shows the global threshold applied using Otsu's method.

Contrast stretching method: The image contrast level may change because of incorrect setting or the poor illumination in the acquisition process.The equation of linear mapping expressed as,

$$A(x, y) = A_1 + \left(\frac{A_2 - A_1}{I_2 - I_1}\right)\left[I(x, y) - I_1\right] \quad (2)$$

Where A_1tends to 0 and A_2denotes to desired level. I_1 and $I_2$are the min and max values of the input gray range (Mori 2010).

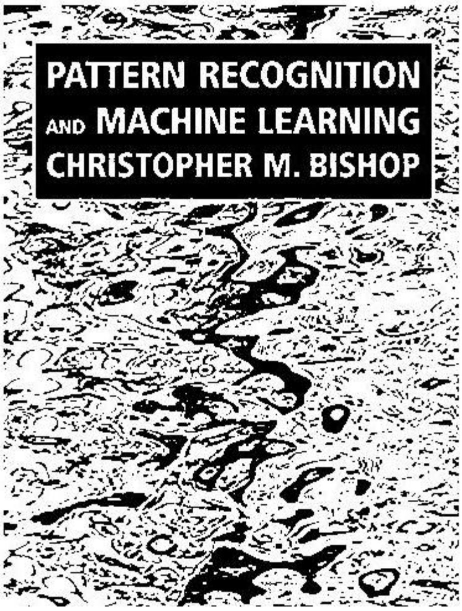
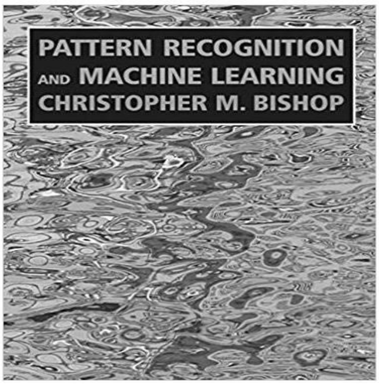


Fig. 4 Global thresholding using Otsu's method

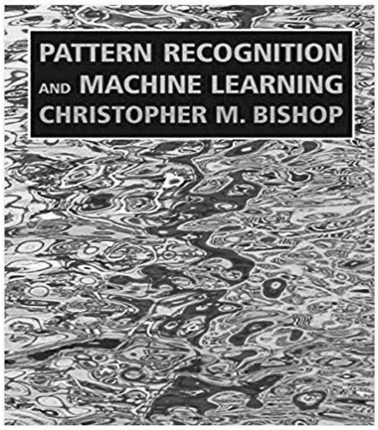
Log transformations: The common form of this method is expressed as,

$$S = K \log (1+r) \quad (3)$$

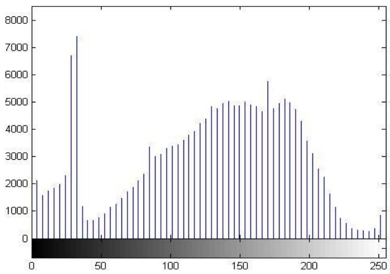
Here 'K' is a constant and the assumption is $r \geq 0$. (Gonzalez et. al. 2009).



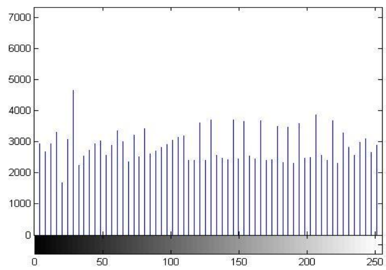
(a)



(b)



(c)



(d)

Fig. 5 (a) Original Image (b) Contrast Enhancement using Histogram Equalization (c) Original image**e**Histogram(d) Histogram of the processed image

Histogram equalization: Histogram equalization is a global type method thatgives the histogram of the whole pixels (0-255) range and this technique can be applied in image understanding problems to normalize variations in illumination.For each level of brightness 'j' in the actual image, the level value (k) for the new pixel can be calculated as,

$$k = \sum_{i=0}^{j} N_i / T \quad (4)$$

WhereT is thetotal pixels in an image (Russ 2007). The contrast enhancement of an image using histogram equalization and its histogram has been shown in figure 5.

G. Mask Processing Techniques

Averaging filters: Average filter also known as mean filter is auncomplicated method for smoothing images.It minimizes the intensity variation amount between adjacent pixels also used to decrease or remove the noise. The averaging filter will be acting as a low-pass frequency filterthatreduces the intensity of spatial derivatives in the image. Mean filterdepends on a kernel that represents thesize and shapeof the neighborhood pixelsto be sampled duringmean value computation. The 3×3 square kernel or mask is shown in Figure 6 (Mori 2010);other large masks like 5×5, 7x7, and 9x9 could be used for smoothing in severe case.

$$\frac{1}{9}\times\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix} \qquad \frac{1}{16}\times\begin{bmatrix}1&2&1\\2&4&2\\1&2&1\end{bmatrix}$$

Average filter — Average Weighted filter

Fig.6 Averaging Kernels3×3 Used in Average Filter

Local thresholding: Local thresholding techniques can be applied to the document images that are having complex background like water marked images or images having non-uniform illumination in its background.The global thresholding method fails to separate the foreground from the backgroundbecause the histogram of images may have more than two peaks.

H. Segmentation

The two main segmentationtypesare external and internal segmentation. External type segmentation splits the page layout into the logical units. It is thedecisivecomponent of the text analysis, as it is an essential step before the off-line text recognition. It extractstheparagraph, sentence or words. The internal segmentation decomposes an image of a series of texts into sub-images. It gives the extraction of letters, especially, in cursive written words.

I. Morphological Processing

The morphological filtering techniquecomprises opening and closing, erosion and dilation, thinning and skeletonization. This practicesuitableonly on binary images (Phillips 2000).
Dilation and erosion are the two morphological masking and threshold techniquesthat increase or decrease thesize of the object. Erosion process erodes or removes the pixels in edges and makes an object smaller; however, dilation process adds pixels around the object edges and creates an object larger.

J. Feature Extraction

Feature extraction is the stage to eliminate redundancy from the data. Theclassification accuracy can be enhanced by selecting or searching most relevant features (Mohamad et al. 2015).

Feature set should have a sufficient discriminating power to enable the accurate classification even among very similar symbols. Thus, features which are beneficial for classification ensure that objects from different classes have different values; objects from the identical class have similar feature values. A good feature set should be efficient about computation time.

K. Text Recognition

Considering the type of text, machine-printed and handwritten OCR methods are two majorpart of interest in the text recognition. The difficulties for fixed and multi-font OCR are solved with little constraints.The individual character separation of full text image are very important step in recognition.

The documents generated with good quality paper yieldstest recognition accuracy as 99%. However, the commercially available products recognition rates much dependent on the quality of paper and ink, the age of the documents.
The problem of the script could be split into two types as the cursive and normal script. In practice, it is difficult to draw a clear difference among them. A mixture of these two types could frequently be observed. On the writing style basis and the complexity of the segmentation stage, there are five phases of the problem in handwritten textrecognition that is shown in Figure 7 (Arica &Vural 2001).

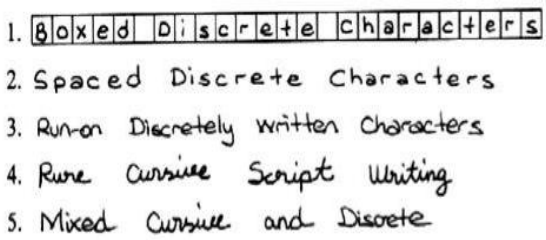


Fig. 7 Five Stages of Handwritten Word Recognition Problem(Arica &Vural 2001)

Chinese Character Recognition

Chinese character recognition is areallycomplicated due to thecomplicated structures, huge number of categories, and resemblanceamong characters, font unevenness or writing styles.Chinese characters have exclusive structures compared to other western language character that causes technical challenges to its detection process. The traditional characters of 5,401 are added in a regular set in Taiwan of China. The three Chinese character sets that contain6,763, 20,902 and 27,533 characters are used as the National standards in the Mainland of China.In academic investigation experiments, usually 3,755 characters are considered as per GB2312-80 standard.

Most of the Chinese characters havecomparatively independent substructurescalled radicals which are composed of poly-line strokes or straight-lines.
Wang et al. (2014) described about homologous offline handwritten Chinese text image. The following Table 1 shows the text recognition results comparison of Chinese offline handwritten text proposed by Wang et al. (2014). In Table GLM, TLM, rTLM indicates general language model, topic language model and restricted topic language model respectively.

The performance parameters can be calculated as,

$$\text{Recognition Accuracy} = \frac{X_t - X_{se} - X_{de} - X_{ie}}{X_t} \quad (6)$$

$$\text{Segmentation Error Rate} = \frac{X_{sce}}{X_t} \quad (7)$$

$$\text{Interpretable Recognition Error Rate} = \frac{X_{resc}}{X_t} \quad (8)$$

Where X_t is the characters count in the recognized text image, X_{de} is the total number of deletion error, X_{se} is the total number of substitution error, X_{ie} is the number of insertion error, X_{sce} is the total number of segmentation contributed errors and X_{resc} is the total number of recognition errors whereas the segmentation is correct.

Table 1 Recognition Results

Data Set	RAR %	ISER %	RER %
GLM	86.94	4.93	4.10
TLM	88.5	4.55	3.40
rTLM (0.05)	88.15	4.61	3.62
rTLM (0.1)	88.08	4.58	3.68
rTLM (0.15)	87.90	4.53	3.96

Arabic Character Recognition

Arabic language is one among the top five spoken languages in the world and 25 countries in Islamic region are using as official languages. It is spoken as mother tongue by more than 250 millionsof people.Arabic characters are used in different languages such as Kardi, Farsi, Urdu, and Chawi. The wide spectrum of characteristics of Arabic language inflicts the requirement for more researches. Arabic language sentence formation is done from right to left that is dissimilar with English language. The recognition should be carried out according to its direction.
The characters of Arabic language may stack on each other also its words may overlap each other horizontally. The characters have single, double or triple dots that are inserted below or above the body of the letter. Table 2 shows the text recognition rare of Arabic text and Table 3 shows the segmentation accuracy of Urdu text that has influence of Arabic text.

Table 2 Arabic Text Recognition Rate

Author	Text type	Findings
SamirElons et al. (2013)	Arabic sign language	Achieved 90% recognition rate for the Arabic sign static postures.
Rashid et al. (2013)	Printed Arabic	99% as recognition rate for the dataset APTI
Rez et al. (2015)	Handwritten Arabic	Achieved 87.4% as recognition rate for IFN/ENIT dataset (32,492 city names)

Table 3 Segmentation Accuracy of Urdu Text

Author	Segmentation method	Segmentation Accuracy
Husain et al. (2002)	CC labeling &method of centroid to centroid distance	100%
H. Malik & M.A. Fahiem (2009)	Horizontal pixel count, freeman chain codes, label matrix, vertical scanning	99.4%

III. RECENT RESEARCHES IN OCR

A. Multilingual Character Segmentation and Recognition

P. Sahare and S. B. Dhok (2018) have developed the Multilingual Character Segmentation and Recognition Schemes for Indian Document Images. They presented the robust algorithms for character segmentation and recognition that supports the multilingual Indian document images with intermixed texts of Devanagari and Latin scripts.The First two features are formed with respect to the character center pixel, whereas the third feature calculation has been calculated using the text pixels neighborhood information. The k-Nearest Neighbor classifier is used for recognizing the input character. They carried out comprehensive experiments on different databases that contain both printed and handwritten texts. The highest segmentation and recognition rates obtained as 98.86% and 99.84% respectively.The Segmentation Rate (SR), Over-Segmentation Rate (OSR) and Bad Segmentation Rate (BSR) obtained by them in the segmentation process are shown in following Table 4. Table 5 shows the comparative analysis of Recognition rate obtained by them on different databases.

Table 4 Segmentation results

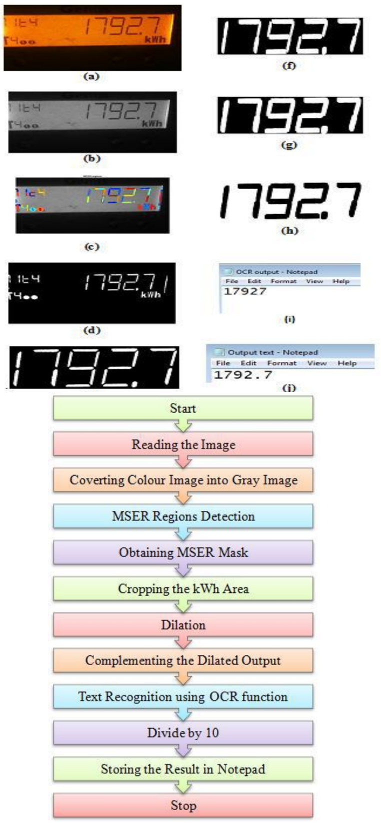
Database	SR %	OSR %	BSR %
Tobacco-800	97.71	0.22	2.07
Proprietary database – Latin Script	98.86	0.18	0.96
Proprietary database - Devanagari script	97.26	0.20	2.54

Table 5 Comparative analysis of recognition algorithm on different databases

Database	Recognition Time	Recognition Rate %
CVSLD	2.71	97.33
CPAR	3.252	98.26
Chars74k Latin script database	5.881	97.10

B. Electricity Billing using OCR

Karthick et al. 2017 has proposed the novel method of electricity billing using OCR. The flowchart and results of text recognition process has been shown in figure 8.



(k)

Fig. 8 OCR Results (a) Original image (b) Grey image (c) Detected MSER (d) Binary mask (e) Cropped image (f) Horizontally dilation (g) Vertically dilation (h) Complemented image (i) OCR output (j) Final output (k) Flow Chart of Text Recognition Process in Electricity Billing (Karthick et al. 2017)

C. Fast Scanning of a Robotic Eye

Kim et al (2018) presents the dynamics-based deblurring method that improves the performance of OCR during fast scanning of robotic eye. The human perception has the comparable frame rate of 24 frames per second (fps) like the camera system. It has the capability of generating rapid motion.

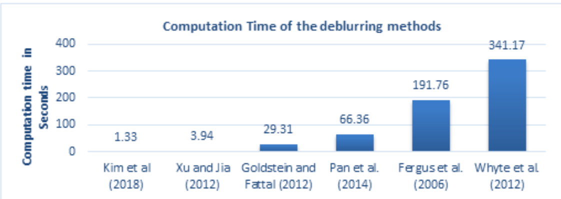


Fig. 9 Comparison of Computation Time of the deblurring methods

Figure 9 shows the computation time taken by the various approaches. It is observed that Kim et al (2018) found the optimal results in reduction of computation time.

IV. CONCLUSION

The computer vision and digital image processing are fast-growing fields that are essential in many aspects of other areas like multimedia, artificial intelligence, robotics and much more. Image analysis involves the study of segmentation, feature extraction, and classification techniques. Humans interact quite naturally with each other over writing and speech, similarly human – computer interaction would make things exciting and easier to the user.From the study it is found that optimal results can be obtained with less computation time as well as multilingual character segmentation and recognition also possible with better rate. It is to be noticed that the segmentation free approach using DNN is also possible in OCR. Our work may bridge the knowledge on automatic interaction between the human-system and system – system interaction.

REFERENCE

1. Arica, N &Vural, FTY 2001, 'An overview of character recognition focused on off-line handwriting', IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 31, no. 2, pp. 216-233.
2. Cherneta, DS, Druki, AA &Spitsyn, VG 2016, 'Development of multistage algorithm for text objects recognition in images', *International Siberian Conference on Control and Communications (SIBCON)*, Moscow, pp. 1-5.
3. Gonzalez, RC, Woods, RE &Eddins, SL 2009, 'Digital Image Processing using Matlab', Gatesmark Publishing, USA.
4. Husain,S. A. "A multi-tier holistic approach for Urdu Nastaliq recognition," *International Multi Topic Conference, 2002. Abstracts. INMIC 2002.*, Karachi, 2002, pp. 84-84.
5. Karthick, K &Chitra, S 2017, 'Novel Method for Energy Consumption Billing Using Optical Character Recognition', Energy Engineering: Journal of Association of Energy Engineers, vol. 114, no. 3, pp. 64-76, ISSN:1998595.
6. Kim, M. D. and Ueda,J. "Dynamics-Based Motion Deblurring Improves the Performance of Optical Character Recognition During Fast Scanning of a Robotic Eye," in IEEE/ASME Transactions on Mechatronics, vol. 23, no. 1, pp. 491-495, Feb. 2018
7. MalikH. and Fahiem,M. A. "Segmentation of Printed Urdu Scripts Using Structural Features," *2009 Second International Conference in Visualisation*, Barcelona, 2009, pp. 191-195. doi: 10.1109/VIZ.2009.12
8. Mohamad, MA, Hassan, H, Nasien, D &Haron, H 2015, 'A Review on feature extraction and feature selection for handwritten character recognition', International Journal of Advanced Computer Science and Applications, vol. 6, no. 2, pp. 204-212.
9. Mori, M 2010, 'Character recognition', Sciyo Publisher, Croatia.
10. Nithyananda, CR, Ramachandra, AC &Preethi 2016, 'Survey on Histogram Equalization method based Image Enhancement techniques', International Conference on Data Mining and Advanced Computing (SAPIENCE), Ernakulam, pp. 150-158.
11. Phillips, D 2000, 'Image Processing in C', R & D Publications, Lawrence, Kansas, USA.
12. Rashid,S.F. Schambach,M.P. Rottland,J. Nüll,S.V.D. 'Low resolution Arabic recognition with multidimensional recurrent neural networks', In *Proceedings of the 4th International Workshop on Multilingual OCR* (MOCR '13). ACM, New York, NY, USA, Article 6, 5 pages, 2013.
13. SahareP. and Dhok,S. B. "Multilingual Character Segmentation and Recognition Schemes for Indian Document Images," in *IEEE Access*, vol. 6, pp. 10603-10617, 2018.
14. SamirElons,A.,Abull-ela,M.,Tolba,M.F., 'Pulse-coupled neural network feature generation model for Arabic sign language recognition', IET Image Processing, 2013, 7, (9), pp. 829-836, DOI: 10.1049/iet-ipr.2012.0222.
15. Wang,Y. DingX. and Liu,C. "Topic Language Model Adaption for Recognition of Homologous Offline Handwritten Chinese Text Image," in *IEEE Signal Processing Letters*, vol. 21, no. 5, pp. 550-553, May 2014.
16. Yi, X & Li, T 2013, 'Design of a multi-format fixed point processing element for graphics processing', 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA), Toronto, pp. 567-570.