An Approach for Reducing Morphological Operator Dataset and Recognize Optical Character based on Significant Features

Conference Paper · August 2015

DOI: 10.1109/ICACCI.2015.7275847

2 authors, including:

Ashis Pradhan
Sikkim Manipal Institute of Technology
10 PUBLICATIONS **77** CITATIONS

An Approach for Reducing Morphological Operator Dataset and Recognize Optical Character based on Significant Features

Ashis Pradhan
Computer Science and Engineering
Sikkim Manipal Institute of Technology
East-Sikkim, India
ashis.p@livemail.smu.edu.in

Mohan P. Pradhan
Computer Science and Engineering
Sikkim Manipal Institute of Technology
East-Sikkim, India

Amit Prasad
Computer Science and Engineering
Sikkim Manipal Institute of Technology
East-Sikkim, India

***Abstract*— Pattern Matching is useful for recognizing character in a digital image. Optical Character Recognition (OCR) is one such technique which reads character from a digital image and recognizes them. Line segmentation is initially used for identifying character in an image and later refined by morphological operations like binarization, erosion, thinning etc. The work discusses a recognition technique that defines a set of morphological operators based on its orientation in a character. These operators are further categorized into groups having similar shape but different orientation for efficient utilization of memory. Finally the characters are recognize d in accordance with the occurrence of frequency in hierarchy of significant pattern of those morphological operators and by comparing them with the existing database of each character.**

Keywords—Binary Image; Morphological patterns; Frequency count; Priority; Reduction data set; Recognition

I. INTRODUCTION

Attempts to develop an OCR system by engineers started form late 1920‘s. However the first commercial OCR system became available during 1950‘s, the technological revolution was moving forward rapidly and electronic data processing was becoming an important field.

In today‘s digital era, paper documentation is virtually an extinct process since digital memory is cheaper than real-time storage space and since such storage is less vulnerable to theft, damage or loss. The process of optically scanning text images and converting it to machine editable text form is done by Optical Character Recognition (OCR) system.

Various approaches are used for the design of OCR system. Matrix Matching converts each characters into a pattern within a, matrix, and then compares the pattern with an index of known characters. Its recognition is strongest on Monotype and uniform single column pages. Feature Extraction defines each character by the presence or absence of key features, including height, width, density, loops, lines, stems and other characters. Structural Analysis identifies characters by examining their sub-features of the image, sub-virtual and horizontal histograms. Its character repair capability is great for low quality text.

In Matrix Matching or Pattern Matching the effectiveness of the identification process greatly relies o n orientation and the shape of the morphological patterns used. The morphological patterns used should be capable of differentiating the patterns present in a binarized image of English alphabets.

The efficiency of the system thus depends o n the technique, type of patterns used and o n the reduction of data set using the rotation invariant process. This proposed works involves the following steps:

- Preprocessing of the image line rotation, binarization and thinning using morphological operation.
- Character segmentation for separation of each character.
- Feature extraction using the morphological patterns.
- Characters recognition using the frequency and priority of the morphological patterns for each character.

II. RELATED WORKS

The implementation of OCR has been done in different approaches with their own results. The common approaches involved include OCR using pattern system, neuronal network system and feature system. Mohammad Faisal et al. [1] in Optical Character Recognition implementation using pattern matching has given algorithm for implementation of OCR of optically scanned images by preserving font properties. This approach requires the matching of the binarized image of the acquired character with the existing template. For this purpose, the image is divided into 5 tracks and each track is divided into 8 sectors. A corresponding track-sector matrix is generated, identifying number of pixel in each region and is compared with existing template.

M. P. Pradhan et. al. [2] have proposed a method for automatic extraction and identification of text from an

978-1-4799-8792-4/15/$31.00

image using Friend Pattern Chain and Euclidean Distance. The proposed method takes into account certain pre-acquired knowledge to recognize the text. The method creates friend pattern chain and host repository for each and every character and associate weight with each frequency count in the friend pattern chain.

Soft-comp uting approaches include fuzzy, neural-net and genetic algorithms. Yang et al. [3] have proposed character recognition method of license plate number based on parallel BP neural networks. The character feature is put into the parallel neural networks for character recognition. Bhattacharya Debi Prasad et al. [4] have proposed an English Alphabet Recognition using chain co de and LCS. Here, chain code along with another significant characteristic, number of vertices and edges (chain code from one vertex to another) existing in each alphabet has been used to identify English capital letters. The technique used is independent of size and many styles of input characters. The experimental results show 95% accuracy on the eight different fonts that they have operated on.

Eswara Babu P. et al. [5] have worked on rectification of document images captured by optical devices and proposed that even for mild skew in a document image which is invisible to human eye results in the failure of segmentation of complete characters and hence the recognition of characters. To eliminate the effect of skew, Hough Transform is used and the resultant segmented characters are recognized using a correlation technique.

Kavallieraton E. et al. [6] have proposed a handwritten character recognition algorithm based on structural characteristics, histograms and profiles. At first a 280 dimension vector is extracted for each character, consisting of histograms and profiles. One new histogram and two new profiles are introduced. The k-means algorithm is then used for classification of characters. The recognition system was first trained based on the NIST database for English characters for each of the categories like digits, uppercase and lower case characters separately. Then, it was tested on unseen cases of the corresponding categories. Experimental result showed varying accuracy from 72.8% to 98.8% dependent on the character category.

Nadira Muda et al. [7] have proposed an OCR (Optical character Recognition) system that recognizes the character by comparing two image of the alphabet. It involves creating database of character templates. For recognition, the input character is compared to each character template to find either an exact match or the template with the closest representation of the input characters the drawback is that the input character has to be rescaled to exact match of the template every time.

III. METHODILOGY

Initially, image is extracted, pre-processed and stored in a 2D array. This is performed to eliminate any sensor error involved in the image. A set of operators are defined which can be used for recognition. Following an algorithm to calculate the frequency count of each template for every scanned character and storing it. Priority is assigned to each operator based on its frequency count for every character.

Figure 1 shown below describes the general steps followed for designing the proposed system.

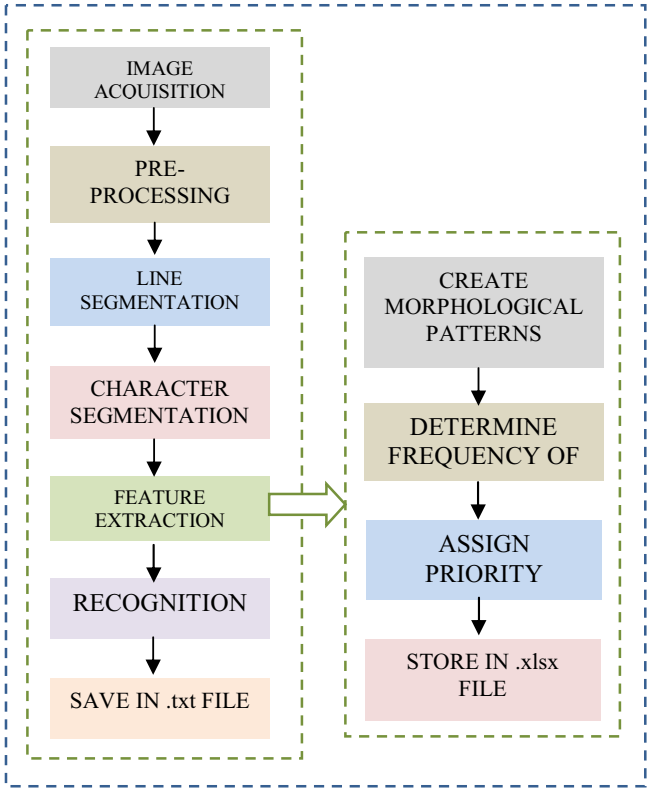

Fig. 1: Block diagram of a designed system

A. *IMAGE ACQUIS ITION*

OCR system read a gray scale or color image from the file specified by the string file name.

B. *CHARACTER SEGMENTATION*

When the image is ready to b e processed, the first step is to isolate each line of the text from the whole document horizontally. Once each line of the text is stored in a separate array, the next step is to isolate each character from the line vertically. The first and last black pixels that are detected vertically are the borders of the character.

C. *PRE-PROCESSING*

After reading a scanned image which is either in JPEG or BMP format, it should undergo the follo wing three conversion processes. They are:

1. Binarization of scanned image.
2. Thinning of image (skeletonization).
3. Pixel inversion.
4. Setting boundary pixels zero (black pixel).

D. FEATURE EXTRACTION

Feature extraction is the process of getting information about a character in order to facilitate classification. This is an important part of the system to categorize a single character for recognition.

i. Defining Set of Morphological Operators:

The morphological operators are defined in matrix of size 3X3. The operators are scanned for its frequency in the identified character for recognition. Later the operators are grouped based on its similarities with variation in orientation.

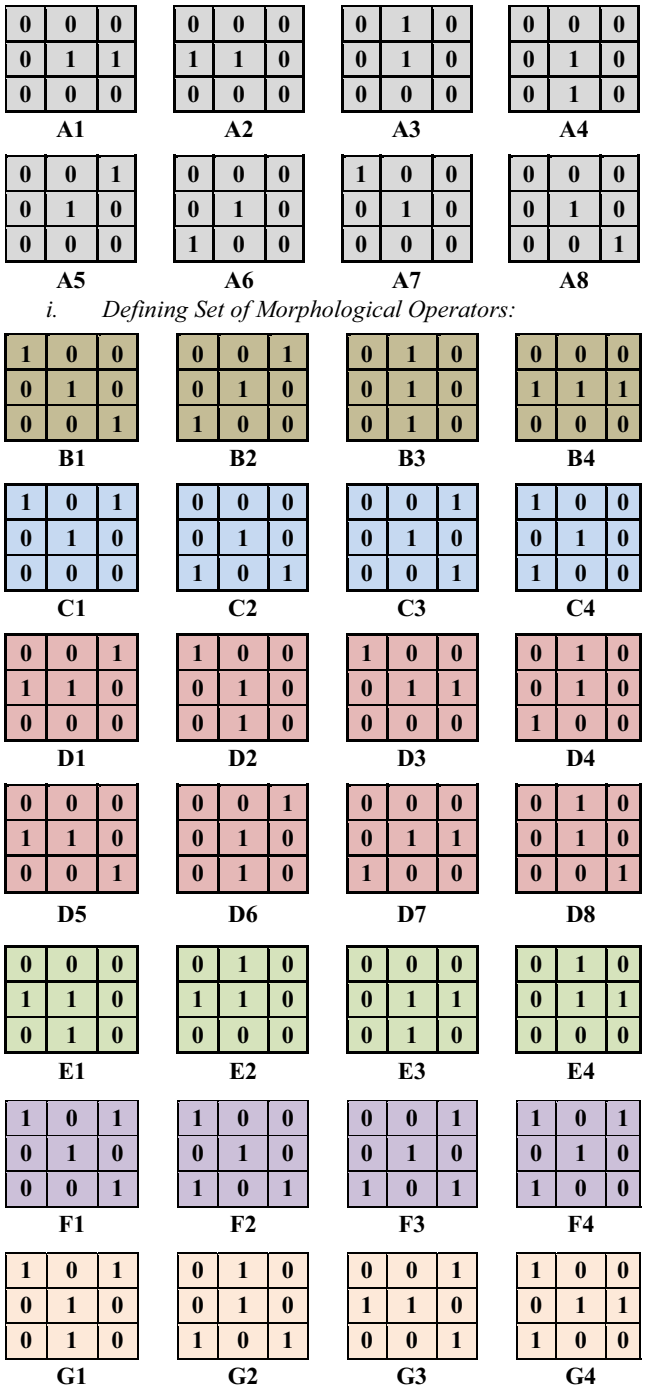

Fig. 2: Defined Morphological operator data set.

ii. Reduction of Data Set Using Rotation:

Each morphological operator represents certain characteristic patterns that are generally present in a binarized image of an character. There are 36 individual morphological patterns defined which are grouped in accordance with similarity in shape as shown in fig 2. While pre-processing, the binarized image is scanned from the top left corner to bottom right corner with 3x3 mask. If a centered pixel or white pixel (value 1) is encountered , then its neighboring 8 pixels are checked for any match from all the morphological patterns stored. If a match is found then account for that corresponding pattern is incremented.

n this process, we store all possible 3x3 mask or morphological patterns or structural element in memory which requires usage of some memory. In our case, there are 36 different 3x3 mask defined, which requires 324 bit reservation of memory, considering one bit for o ne cell.

However, it has been noticed that these set o f morphological pattern in groups resembles same feature based on the rotation of defined pattern. For example, if pattern A1 is rotated by 180º, it resembles pattern A2. Hence, by identifying the group of pattern with similarity based on rotation, we can reduce the storage space required for storing the set of defined structural element or pattern.

Therefore, this 36 uniquely defined data set were reduced to only 7 data set, which rotates itself to include all possible orientation of pattern within that group. Hence, the amount of memory required for storing these 7 patterns is reduced to only 63 bits.

iii. Determining Frequency Count:

The segmented character is traversed using a 3x3 mask element starting from top left corner to right bottom. For any significant pixel in its central position of mask, the neighborhood pixel location were identified and analyzed to identify the defined set of pattern.

For a given mask, the mask was traversed through entire image to count the number of occurrences (frequency) of that corresponding mask. The same process is carried for all defined set of morphological operator and stored the frequency count in tabular form of each operator for all character set which can be used as basis for recognition.

It has also been noticed from Table 1 that some pattern have least occurrences or frequency. So, it can be inferred that these least occurred pattern is the significant feature for particular character and can be used for recognizing correctly. Thus, it increases the efficiency of recognition phase.

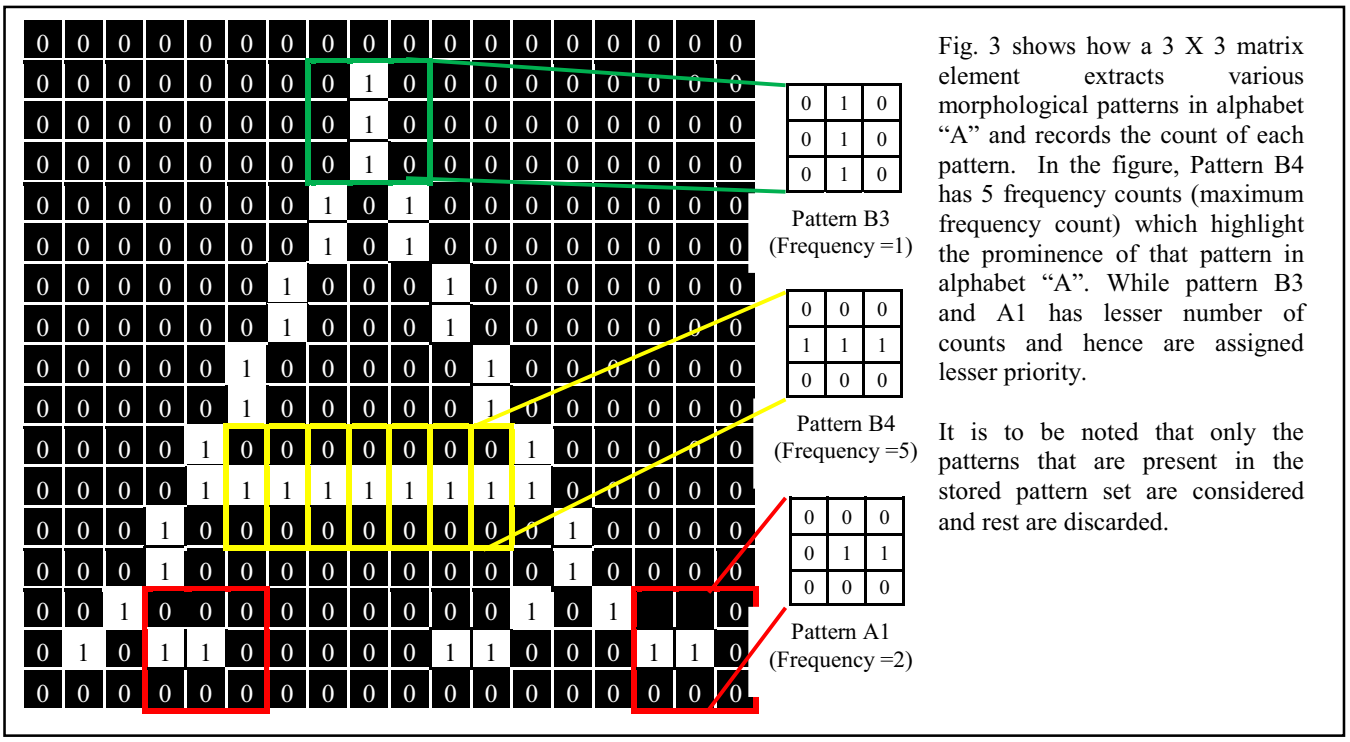

Fig. 3 shows how a 3 X 3 matrix element extracts various morphological patterns in alphabet "A" and records the count of each pattern. In the figure, Pattern B4 has 5 frequency counts (maximum frequency count) which highlight the prominence of that pattern in alphabet "A". While pattern B3 and A1 has lesser number of counts and hence are assigned lesser priority.

It is to be noted that only the patterns that are present in the stored pattern set are considered and rest are discarded.

Figure 3: Frequency count of pattern 'A1', 'B3' and 'B4' for alphabet "**A**"

After all the pixels in the character image are traversed by the 3X3 matrix element, the frequency count of each pattern is stored in database as shown in Table 1. Also, it is notified that the frequency of individual morphological operator varies with different character. Hence by plotting the histogram of individual character, the prominent or significant operator can be easily identified. The priority is then assigned in hierarchy for each operator based on the frequency count. Lastly, this priority is used as a basis for classification and recognition of testing character.

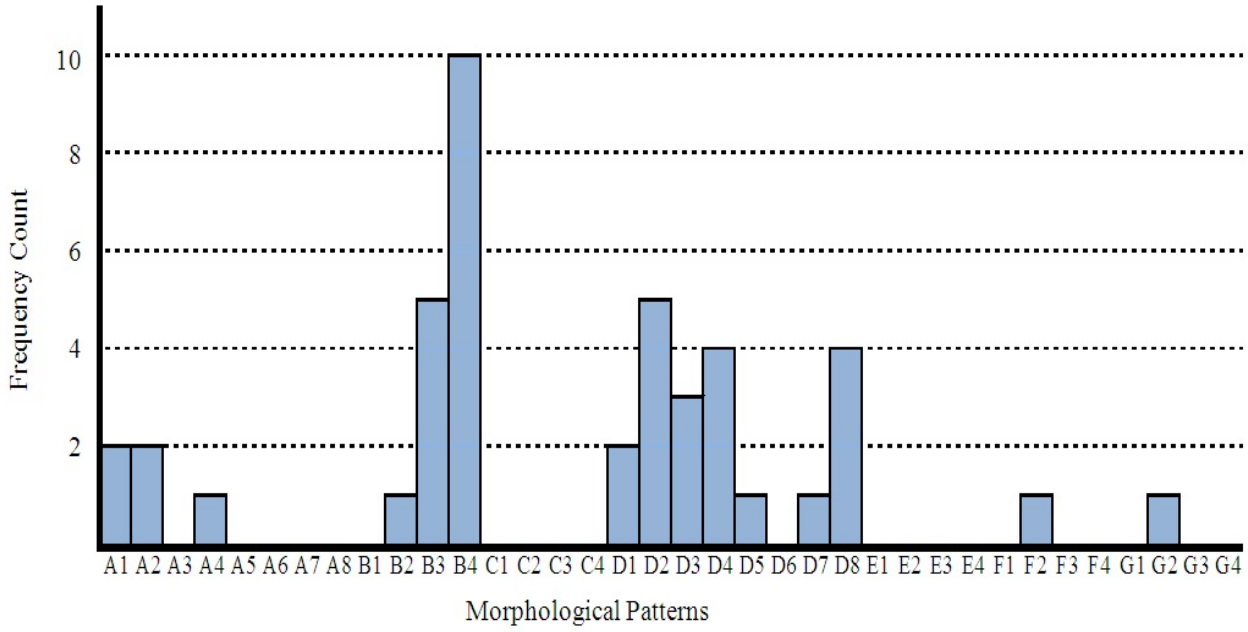

Fig. 4: Frequency count vs. morphological patterns bar graph for character "**A**"

TABLE I. FREQUENCY COUNT WITH THEIR PRIORITY

FREQUENCY OF MORPHOLOGICAL OPERATOR																										
Character	**A**	**P**	**B**	**P**	**C**	**P**	**D**	**P**	**E**	**P**	**F**	**P**	**G**	**P**	**H**	**P**	**I**	**P**	**J**	**P**	**K**	**P**	**L**	**P**	**M**	**P**
Operator		**A**		**B**		**C**		**D**		**E**		**F**		**G**		**H**		**I**		**J**		**K**		**L**		**M**
A1	2	2	1	1	0	X	1	1	1	1	1	1	0	X	2	2	1	1	1	1	2	2	1	1	2	2
A2	2	2	0	X	0	X	0	X	0	X	1	1	1	1	4	3	2	2	1	1	4	4	1	1	3	3
A3	0	X	0	X	1	1	0	X	2	2	1	1	0	X	0	X	0	X	0	X	0	X	0	X	1	1
A4	0	X	0	X	0	X	0	X	1	1	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
A5	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
A6	0	X	0	X	2	2	0	X	1	1	1	1	1	1	0	X	0	X	0	X	0	X	1	1	0	X
A7	0	X	0	X	0	X	0	X	0	X	1	1	1	1	0	X	0	X	0	X	0	X	0	X	0	X
A8	1	1	1	1	0	X	1	1	1	1	1	1	1	1	2	2	1	1	1	1	2	2	1	1	1	1
B1	1	1	3	3	2	2	1	1	2	2	1	1	3	3	0	X	0	X	1	1	6	6	1	1	3	3
B2	3	3	4	4	5	5	3	3	2	2	1	1	6	5	2	2	1	1	1	1	8	7	2	2	5	5
B3	5	4	16	5	8	6	22	6	14	3	13	4	10	6	24	5	15	3	15	3	10	8	15	5	27	7
B4	5	4	17	6	8	6	11	5	22	4	12	3	10	6	14	4	2	2	0	X	5	5	7	4	2	2
C1	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
C2	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
C3	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
C4	1	1	1	1	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
D1	2	2	2	2	3	3	3	3	2	2	1	1	2	2	2	2	1	1	2	2	2	2	3	3	2	2
D2	5	4	2	2	2	2	2	2	2	2	2	2	2	2	1	1	0	X	0	X	1	1	0	X	6	6
D3	3	3	3	3	4	4	4	4	1	1	2	2	4	4	4	3	2	2	2	2	3	3	1	1	3	3
D4	2	2	2	2	1	1	2	2	2	2	1	1	2	2	1	1	0	X	1	1	0	X	1	1	4	4
D5	1	1	3	3	3	3	4	4	1	1	2	2	4	4	2	2	1	1	2	2	2	2	0	X	1	1
D6	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
D7	0	X	1	1	3	3	2	2	1	1	1	1	3	3	2	2	1	1	2	2	2	2	2	2	1	1
D8	3	3	2	2	2	2	1	1	1	1	2	2	3	3	1	1	0	X	0	X	2	2	0	X	4	4
E1	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
E2	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
E3	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
E4	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
F1	0	X	0	X	0	X	0	X	0	X	0	X	1	1	0	X	0	X	0	X	0	X	0	X	0	X
F2	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
F3	1	1	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
F4	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
G1	0	X	1	1	0	X	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	0	X
G2	1	1	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	1	1
G3	0	X	1	1	1	1	0	X	1	1	1	1	0	X	1	1	0	X	0	X	0	X	0	X	0	X
G4	0	X	1	1	0	X	0	X	1	1	1	1	0	X	1	1	0	X	0	X	0	X	0	X	0	X

FREQUENCY OF MORPHOLOGICAL OPERATOR																										
Character	**N**	**P**	**O**	**P**	**P**	**P**	**Q**	**P**	**R**	**P**	**S**	**P**	**T**	**P**	**U**	**P**	**V**	**P**	**W**	**P**	**X**	**P**	**Y**	**P**	**Z**	**P**
Operator		**N**		**O**		**P**		**Q**		**R**		**S**		**T**		**U**		**V**		**W**		**X**		**Y**		**Z**
A1	1	1	0	X	1	1	0	X	1	1	0	X	1	1	0	X	0	X	0	X	2	2	1	1	1	1
A2	2	2	0	X	1	1	1	1	2	2	0	X	1	1	2	2	2	2	3	3	4	4	3	3	1	1
A3	1	1	0	X	0	X	0	X	0	X	0	X	2	2	0	X	1	1	2	2	0	X	0	X	1	1
A4	0	X	0	X	0	X	0	X	0	X	1	1	0	X	0	X	0	X	0	X	0	X	0	X	0	X
A5	0	X	0	X	0	X	0	X	0	X	1	1	0	X	0	X	0	X	0	X	0	X	0	X	0	X
A6	0	X	0	X	0	X	0	X	0	X	1	1	0	X	0	X	0	X	0	X	0	X	0	X	1	1
A7	0	X	0	X	0	X	0	X	0	X	1	1	0	X	0	X	0	X	0	X	0	X	0	X	0	X
A8	2	2	0	X	1	1	0	X	1	1	0	X	0	X	2	2	2	2	3	3	2	2	2	2	0	X
B1	10	4	4	3	1	1	7	5	5	4	3	2	1	1	2	2	1	1	3	3	6	5	3	3	2	2
B2	2	2	6	4	3	3	8	6	3	3	5	4	0	X	1	1	3	3	5	5	11	6	5	4	9	4
B3	23	5	13	6	16	5	13	7	15	6	5	4	17	4	26	5	5	4	16	8	0	X	6	5	1	1
B4	1	1	8	5	9	4	5	3	5	4	6	5	8	3	4	4	0	X	0	X	1	1	1	1	16	5
C1	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
C2	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
C3	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
C4	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
D1	1	1	2	1	2	2	0	X	2	2	1	1	1	1	2	2	0	X	1	1	3	3	1	1	2	2
D2	3	3	3	2	1	1	3	2	2	2	3	2	1	1	0	X	5	4	6	6	3	3	2	2	0	X
D3	2	2	3	2	3	3	6	4	6	5	5	4	1	1	2	2	2	2	1	1	4	4	2	2	1	1
D4	1	1	3	2	1	1	3	2	1	1	1	1	0	X	2	2	5	4	5	5	2	2	2	2	5	3
D5	1	1	3	2	3	3	5	3	5	4	6	5	2	2	1	1	1	1	1	1	2	2	1	1	1	1
D6	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
D7	1	1	2	1	1	1	1	1	2	2	1	1	2	2	3	3	2	2	4	4	2	2	2	2	2	2
D8	2	2	3	2	1	1	3	2	2	2	4	3	0	X	1	1	6	5	7	7	3	3	2	2	0	X
E1	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
E2	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
E3	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
E4	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
F1	1	1	0	X	0	X	0	X	0	X	1	1	0	X	0	X	0	X	1	1	0	X	0	X	0	X
F2	1	1	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	1	1	0	X	0	X
F3	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
F4	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	1	1
G1	0	X	0	X	1	1	0	X	1	1	0	X	0	X	0	X	2	2	2	2	1	1	1	1	0	X
G2	1	1	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	1	1	0	X	1	1
G3	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X
G4	0	X	0	X	1	1	0	X	1	1	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X

IV. RECOGNITION

The basic principal used in recognition phase is the hierarchical based elimination based on existence of significant feature. Recognition of the character is based on the pre-acquired knowledge of the frequency count and priority of each character stored as a reference. While the frequency count plays an important role in determining the accuracy, the priority assignment plays a key role in reducing the search space for the recognition process. Reduction of Search Space: It becomes a complex process to match the frequency count of each pattern in the test character to the frequency count of each pattern of every single reference character stored. The search space can be reduced using priority assignment. In order to recognize a character, firstly, only the pattern with the highest priority along with frequency count, of the test character is matched with the pattern (along with its frequency count) with highest priority of all character stored in repository.

Case I: If only a single character in the repository is found whose highest priority pattern is same and has frequency count that matches with the test character, the a match is declared.

Case II: If there are more than one character in repository whose highest priority pattern matches, but not all of their frequency count, then only the pattern whose frequency count matches is selected and a match is declared.

Case III: If there are more than one character in repository whose highest priority pattern matches along with their frequency count, then the search space (containing the matched characters) is reduced by considering only those characters in search space, whose second highest priority pattern along with their frequency count matches and so on until a match is found.

Note: A threshold value can be set up to check the matching of frequency count. Doing so will help increase the flexibility of the recognition process.

Finally the designed system was tested by considering same data (same font type and size) from other operating system like Linux and Mac. Hence, through this result analysis, it was found that although the operator count doesn't match exactly, but the pattern of occurrence was same as compared to Windows OS. This leads to challenges when the frequency distribution pattern doesn't rely on character size.

It is noted that the character keyed in through keyboard, whose font type and size is same, is found to have different frequency count for same character in different operating system. The figure shown below describes the pattern show by same character with similar properties in different operating system.

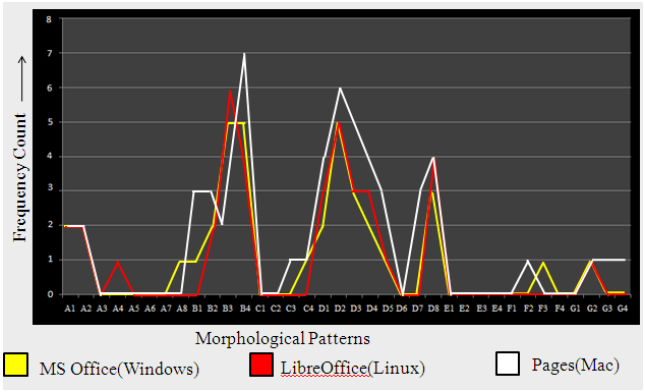

Fig. 5: Pattern representing frequency in different OS for character "A"

It is seen that the behavior of occurrence of pattern is same, however it is not accurate. This shows that the accuracy of designed system has accuracy of 97.19% in MS Office (Windows), 77.75% in Libre Office (Linux) and 63.705% in Pages (Mac).

TABLE II. ACCURACY ASSESMENT ON DIFFERENT ENVIROMENT

	Recognition Rate (%)		Error Rate (%)	
	Upper Case	Lower Case	Upper Case	Lower Case
WINDOWS	98.02	96.36	1.98	3.64
LINUX	77.23	78.27	22.77	21.73
MAC	61.88	65.53	38.12	34.47

V. CONCLUSION

The efficiency of the OCR technique described in this paper depends on the set of morphological patterns we define. Hence, it is required that we include only those patterns that play an important role in the recognition of character. Finally, we discard those patterns that do not have any significance in the recognition process, to reduce the data set.

The feature extraction process can be made faster by using rotation invariant technique. While frequency count and the priority assignment decide the performance of the recognition process, it is only priority assignment that helps in reducing the search space during the recognition process.

REFERENCES

[1] Faisal Mohammad, Jyoti Anarase, Milan Shingote, Pratik Ghanwat, —Optical Character Recognition Implementation using Pattern Matching□, International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (2), 2014, 2088-2090.

[2] Mohan P. Pradhan, M. K. Ghose, —Efficient Technique for Automatic Extraction and Identification of Text from an Image using Friend Pattern Chain and Euclidean

Distance□, Internation al Journal of Computer Applications (0975 – 8887), Vol. 71 – No.10, May 2013.

[3] Feng Yang and Fan Yang et al., Character Recognition Using Parallel BP Neural Network, Proc. of International Conference on Audio, Language and Image Processing(ICALIP),1595-1599 pp., 2008.

[4] Debi Prasad Bhattacharya, Susmita Koner, ―English Alphabet Recognition using Chain code and LCS□, Indian Journal of Computer Science and Engineering (IJCSE), Vol. 3 No. 2 Apr-May 2012.

[5] P. EswaraBabu, K. Ravi Kumar, Dr. G. Lavanya Devi, ―Recognition of characters in document images using Morphological operation□, International Journal of Science Engineering and Advance Technology, IJSEAT, Vol 1, Issue 6, November 2013.

[6] Ergina Kavallieratou, K Sgarbas, Nikos Fakotakis, G Kokkinakis, "Handwritten word recognition based on structural characteristics and lexical support", Proceedings of Seventh International Conference on Document Analysis and Recognition, 2003, pp. 562-566.

[7] Muda, Nadira, et al. "Optical Character Recognition By Using Template Matching (Alphabet)." Online available at http://umpir.ump.edu.my/969/1/NaCSES-2007-086_Optical_ Character_Recognition_By_Using_Templ.pdf", [accessed on 16 march 2015]

[8] Wang, L. and Pavlidis, T., Direct gray-scale extraction of features for character recognition, IEEE Trans. PAMI, Vol. 15, No.10, 1053–1066 pp., 1993.

[9] M. Salmani Jelodar, M.J. Fadaeieslam, N. Mozayani, M. Fazeli, ―A Persian OCR System using Morphological Operators□, Interna tional Journal of Computer, Information, Systems and Control Engineering Vol:1 No:4, 2007.

[10] Ravi Kumar, Anurag Anand, Nikunj Sharma, ―Recognition Of English Characters By Codes Generated Using Neighbour Identification□, International Journal of Applicatio n or Innovation in Engineering and Management Vol: 2, Issue 4, April 2013

[11] Seethalakshmi R., Sreeranjani T.R., Balachandar T., ―Optical Character Recognition for printed Tamil text using Unicode ―Journal of Zhejiang University SCIENCE ISSN 1009-3095.

[12] Jonath an J. Hull, Alan Commike, et.al., ―Multiple Algorithms for Handwritten Character Recognition□, International Workshop on Frontiers in Handwritten Recognition, Montreal, Canada, April 2-3, 1990, pp-117-129.