

Digital Address Identification From Handwritten Address In Postcards

V. Mekala, M. Manimegalai, K. Sasipriya, K. Selvakani, J. Sriram Gautham

Abstract: In this paper we present an idea of using character recognition for sorting the handwritten postcards in postal service. The Indian postal letter sorting system has been made to recognize the postcards based on PIN codes by using Barcode approaches. The objective of the project is to provide an alternative mean to the traditional sorting system which consumes more time for processing and sorting the postcards based on their respective areas. It also aims at eliminating the human errors which may occurs during manual sorting. This can be achieved by character recognition for segregating the postcards based on their respective district. The character recognition process is carried out by taking the postcard images as input and the addresses are converted into string of pixel values. The string of pixel values is processed with the datasets. As a result, the characters are recognized for each input image. When the entire recognition process is completed, the addresses are sorted based on their district and the addresses are uploaded to the corresponding excel sheet based on their district. This provides a solution to the existing problem in spending more time for processing and sorting the postcards and also helps to avoid human errors during manual sorting of postcards.

Index Terms: Character recognition, Python, OpenCV, Postal Address Sorting.

1. INTRODUCTION

The postal system is important in the development of mail transportation. But most of the postal systems in India are manually operated for sorting and processing. It has been observed that the manual sorting and processing leads to many disadvantages such as human errors, more processing time and manpower required. So, it is required that the manual sorting or processing of the mails must be transformed into fully automated processing method. The emergence of this method using OCR provides a solution of recognition of the handwritten text which leads to the automatic postal address recognition. Handwritten character recognition is an area of pattern recognition which defines the ability of the machine to analyse patterns and to identify the characters. The basic function of the recognition system is to detect and extract the data from the pattern of describing the object to the same pattern class and recognize the pattern. Handwritten character recognition is one such area where it is used to preserve the information to retrieve it at a later stage. A camera is used to capture image of the mail items and then the postal address is recognised with the help of OCR. Every individual human being has its style of writing. Sometimes even it is difficult for a human to identify and read the handwritten text of another human. However, text extraction in handwritten documents is becoming a challenging one while considering the scale and orientation of the characters because it spatially varying, moreover the distance between the lines are irregular and character may touch across words or text lines.

2 LITERATURE REVIEW

A webcam-based optical character recognition using MATLAB [13] takes the image as the input, gets text from that image and then convert it into characters. Each character gets segmented and the resulting character image is applied into the pre-processor for noise reduction and normalization. The feature vectors are then extracted from these characters for further classification. After classification, the identified characters are organized to recreate the original symbol strings and context may then be used to detect and correct the errors. OCR is a method to scan the document and create the document editable by converting the text into the machine-encoded form. The OCR system [14] consists of three major subphases namely pre-processing, segmentation and recognition. Image pre-processing is a sorting process that takes the input images resembling a photograph and the image process output could be a picture or a sequence of characteristics or parameters associated with the images. Segmentation contains three major steps like line segmentation, word segmentation, and character segmentation. After segmentation, the text is recognized using the machine-encoded formats of images. A handwritten text recognition based on Neural Network involves the processes of pre-processing, segmentation, feature extraction. This system [15] takes the handwritten characters as inputs, process the inputs, extract the feature vectors from the inputs. This system utilizes the 3-layer Artificial Neural Network using a supervised learning approach. The feature vectors are applied to the ANN along with the generated target vectors. The text is recognized based on the input samples of English Alphabets used in the ANN training process to make the applicability of system towards the new inputs. The trained system provides an average accuracy of more than 95% with the test images. The main purpose of this system is to increase the accuracy level of the test images. Segmentation Methods for Handwritten Character Recognition implemented using the pixel counting algorithm [16]. It is only focused on preprocessing and segmentation techniques. The preprocessing includes several steps such as digitization of input image. Then Segmentation can be implemented into its subcomponents. Segmentation of words, lines, and characters plays a major role in the recognition rate of the script. But it fails to provide satisfactory results while working with handwritten text images. The information of the printed or

- V. Mekala is an Assistant Professor in Kongu Engineering College, Erode, India. E-mail: mekalav.ece@kongu.edu
- M. Manimegalai is an Assistant Professor in Kongu Engineering College, Erode, India. E-mail: manimegalai@kongu.edu
- K. Sasipriya is a Bachelor Engineering student in Kongu Engineering College, Erode, India. E-mail: sasipriyakandasamy@gmail.com
- K. Selvakani is a student in Kongu Engineering College, Erode, India. E-mail: selvakanis37@gmail.com
- J. Sriram Gautham is a Bachelor Engineering student in Kongu Engineering College, Erode, India. E-mail: sriram.gautham1999@gmail.com

handwritten documents is stored by scanning the documents and storing them as image files [17]. The OCR technique is used to get back and store the information in particular text from these images automatically. The main objective of the OCR technique is to achieve the modification or conversion of any form of text in the handwritten or printed documents images into editable digital format. Marathi script consists of 13 vowels and 37 consonants making 50 alphabets [18]. Input handwritten document is scanned then these images are passed to the pre-processing stage. The colored or gray images are converted into Black and white images, then the noise gets removed. Then this cleared binary image is passed for segmentation. The textual document is converted into a machine-readable format which uses a digital camera or a scanner to capture different types of documents and convert all these documents into machine editable format like ASCII code [11] where characters are captured by a tablet digitizer. Handwritten character is recognized by pre-processing, segmentation, feature extraction, and classification techniques.

3 EXISTING SYSTEM

Optical Character Reader reads the Post mail piece destination address, Pin code and prints a Barcode for further sorting. Initially, in the automation of the sorting process the destination address block (DAB) is segmented from the envelope image using the Zip code and it has been located. The recognition process is performed after the extracted Zip code is segmented into individual digits. After recognition, an appropriate bar code of the destination address has been printed. Destination Address Block Segmentation algorithm is based on feature selection using a wavelet space. Brazilian postal envelopes were used in a greater number of experiments. The images were decomposed into wavelet space using decomposition with a basis to identify the salient points. The features based on the frequency of low and high were obtained. The images have a different layout and different background has been used. About 85% of accuracy was achieved. Convolutional locator networks (CLN) were used to locate the address block on machine-printed mail pieces of the US postal service (USPS). The DAB was considered as varied shapes, sizes, and justifications. The address block (AB) was obtained by using CLN by obtaining low-level features and high-level abstract features. CLN are used to detect the corners of Address Block. The features were obtained from each pair of pixels for both the images. The classification of the array is used for storing the extracted features. Then the Destination Address Block (DAB) was segmented after the images were tested based on the stored classification array by using the K-Nearest Neighbor (KNN) and Euclidean distance measure. It classifies them into text blocks and non-text blocks based on the obtained features by using a connected component (CC). The address blocks were merged to obtain with the segmented text blocks

4 PROPOSED SYSTEM

Character recognition plays a major role for sorting the postcards based on the district. This System captures the image of post mails in real-time with the help of web camera. Then the captured image undergoes pre-processing which includes techniques like colour code conversion, binarization, thresholding in which the image is converted from RGB image into grayscale image. The converted image is now modified

into white and black images by giving a threshold value to the image. Segmentation is used to segment each character of the handwritten image. After segmenting each character, the feature vectors for each character get extracted in the form of white and black pixels. The obtained pixels are then processed with the dataset of 28x28 pixels containing more than 60,000 images of handwritten characters and digits. As a result of maximum accuracy, the characters are recognized. After undergoing all the above techniques, the addresses of the postcards are recognized and displayed. The displayed addresses are sorted based on their respective district. The addresses are uploaded to the corresponding excel sheet to have the database of collected postcards in postal service that is created to serve the purpose of sorting the postcards based on their district. The addresses are continuously uploaded to the excel sheet each time when the images are captured, recognized and undergoes the process of sorting.

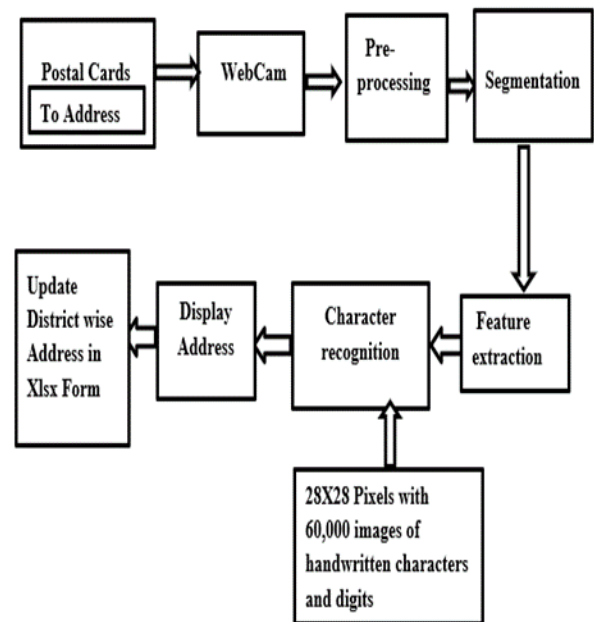


Figure 4.1 Block Diagram of Handwritten Postcards Sorting

5 IMPLEMENTATION

A handwritten postal card is converted into a binary image for digitalizing an input image by the computer technique using an image processing method. A postal card of an input image is processed with a particular pixel size of 28x28 of handwritten characters and digits. The image acquisition is done by capturing the image and saving it in a particular location. The saved image is then undergoing further process by providing the path of the image to the software program designed to convert the image into pixels. Figure 5.1 shows the input image of the handwritten postcards and it is saved in a particular path location.

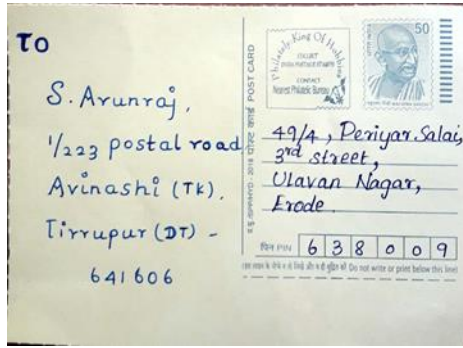


Figure 5.1 Image of postcard

The saved image is then cropped or resized to a particular value for finding the delivery address in the postcards. Figure 5.2 shows the cropped image after resizing the original image for finding the delivery address in the postcards.

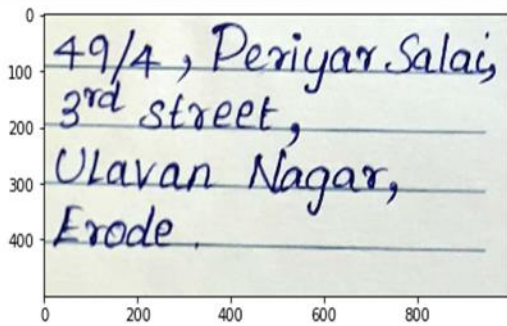


Figure 5.2 Cropped image of the delivery address

The resized image is then processed with the datasets of 28x28 pixels of more than 60,000 images of character and digits for the recognition of handwritten text in the postcards. Figure 5.3 shows the text written in the postcard after undergoing the recognition process.

49/4, Periyar Salas,
3% styect ,
Ulavan_Nagar,
Erode _*

Figure 5.3 Output of Recognized text

The recognition of the text in the input image is done when the input image is captured using the web camera. In this system, the real-time image is captured and processed with the datasets and gives the corresponding text as the output after recognition within a fraction of seconds.

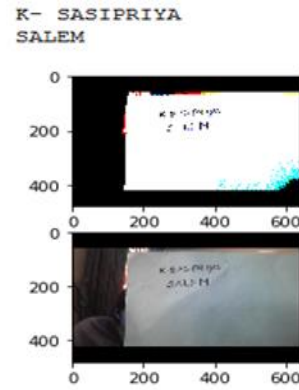


Figure 5.4 Test image 1

SRIRAM GAUTHAM
Bc Vaishnavi Gayden,
Ranga Nagay,
Avinashi

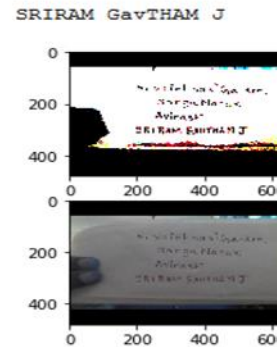


Figure 5.5 Test image 2

The Figure 5.4, Figure 5.5 shows the image captured in real-time using a web camera and the output for the respective images after recognition. Thus, the text in the captured image is processed and the feature vectors are obtained from each character in the form of strings of pixels.

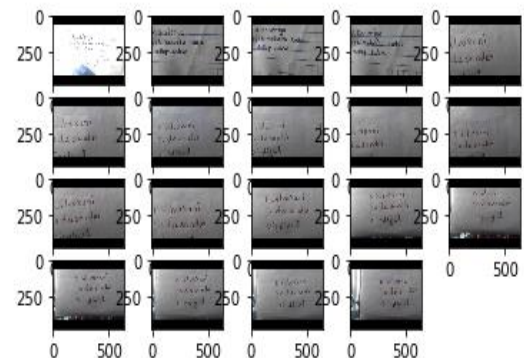


Figure 5.6 Set of Images of Handwritten Postcards

The Figure 5.6 shows the image of a set of 24 handwritten postcards. Initially, the count of the postcards need to be segregated are entered which will useful for sorting the daily postcards arriving the postcards. When the images of the postcards were collected, the character recognition process is initiated. This system is already trained with the datasets of characters and digits. First, the system detects the text in the input image and matches it against the trained dataset when the character is matched the handwritten text in the postcards

are displayed.

```

20-10-2019 12:00:23K+ Sasipri ya
101A, Vadakku
Malluye, Salem
K+ Sasiptiva
401A, Vadaxku kadu
Mallup Salem

20-10-2019 12:00:23K+ Sasiptiva
401A, Vadaxku kadu
Mallup Salem

20-10-2019 12:00:23
k:Selvakani
Vedo.sandu-x
a Dindigul

20-10-2019 12:00:23k:Selvakani
Vedo.sandu-x
a Dindigul
    
```

Figure 5.7 Image of Address Recognized

The Figure 5.7 shows the text recognized after the images are captured. The postcard count is entered in the loop and the images are captured up to the loop and the text is recognized with the time and date of when the image was captured. In this system, the real-time image also contains cursive text which is captured and processed with the datasets and gives the corresponding text as the output after recognition.

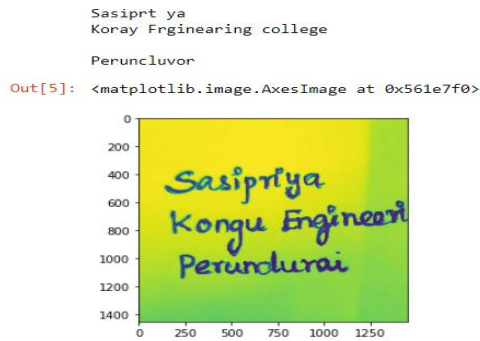


Figure 5.8 Cursive Test Image 1

The Figure 5.8, Figure 5.9 shows the images of the cursive text and its corresponding output after recognition.

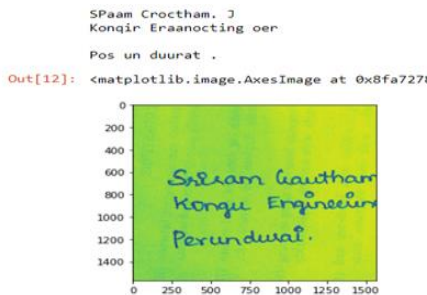


Figure 5.9 Cursive Test Image 2

6 RESULT AND DISCUSSION

When the input image is provided, the system detects the address in the image and extract the feature vectors from the image in the form of strings of pixels. When the features are matched with the training dataset the address gets recognized and displayed. The recognized address in the postcard is then sorted in the excel sheet based on their district.

Clipboard	Font	Alignment							
A1									
	A	B	C	D	E	F	G	H	I
5		20-10-2019 17.09.25K	Sasipriya101A, Vadakku	KadyMalluw, salem					
6		20-10-2019 17.09.25K	Sasipriya401A, Vadakku	kadyMalluw, salem					
7		20-10-2019 17.09.25K	Sasipriya101A, Vadakku	kadyMalluw, salem					
8		20-10-2019 17.09.25K	Sasipriya401A, Vadakku	kadyMalluw, salem					
9		20-10-2019 17.09.25K	Sasipriya101A, Vadakku	kadyMalluw, salem					
10		20-10-2019 17.09.25K	Sasipriya101A, Vadakku	kadyMalluw, salem					
11		20-10-2019 17.09.25K	Sasipriya101A, Vadakku	kadyMalluw, salem					
12		20-10-2019 17.09.25K	Sasipriya101A, Vadakku	kadyMalluw, salem					
13		20-10-2019 17.09.25K	Sasipriya101A, Vadakku	kadyMalluw, salem					
14		20-10-2019 17.09.25K	Sasipriya101A, Vadakku	kadyMalluw, salem					
15		20-10-2019 17.09.25K	Sasipriya101A, Vadakku	kadyMalluw, salem					
16		20-10-2019 17.09.25K	Sasipriya101A, Vadakku	kadyMalluw, salem					
17									
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									
28									
29									
30									
31									

Figure 6.1 Spreadsheet Result-1 of Sorted Address

Clipboard	Font						
A1							
	A	B	C	D	E	F	G
1		20-10-2019 17.32.50S+	SelvaPeriyatGoErode				
2		20-10-2019 17.32.50S+	SelvaPeriyatogErode				
3		20-10-2019 17.32.50Se	Se waPeriyat atErode				
4							
5							
6							

Figure 6.2 Spreadsheet Result-2 of Sorted Address

Clipboard	Font						
A1							
	A	B	C	D	E	F	G
1							
2							
3		04-10-2019 10.43.33'e	ly akante da sandurDindigul				
4		04-10-2019 10.43.33elvakante	da sandurDindigul				
5		04-10-2019 10.43.33elvaka	nt2 d a sandurDindigul				

Figure 6.3 Spreadsheet Result-2 of Sorted Address

The Figure 6.1, Figure 6.2, Figure 6.3 represents the resultant spreadsheet where the address in the postcards is marked with the corresponding date and time of when the input image was captured. The Addresses are saved in their respective spreadsheet columns based on their district.

7 CONCLUSION AND FUTURE SCOPE

Thus, the system has proved to provide an alternative means for recognizing and sorting the postcards in the post office and also helps in avoiding human errors. It has put an end to the tedious process of sorting the postcards and at the same time, it has saved the time that is previously spent on sorting the postcards. The efficiency of the system depends on the resolution (density of pixels) of the input image and type of the image. The result showed that the proposed system requires recognition for cursive characters which can be implemented in the future to improve the system performance. This system can be implemented in the post office for the segregation of postcards instead of manual sorting.

REFERENCES

- [1] Hussian Salim Qasim (2015), "A Letter Recognition Data using Neural Network" International Journal of Scientific and Engineering Research, Vol 4, Issue 5.
- [2] Jederberg.M, Simonyan.K, Vedaldi.A, and Zisserman.A (2015), "Deep Structured Output Learning for Unconstrained Text Recognition," in International Conference on Learning Representations, pp 1-10.
- [3] Madhurai Maheshwari, Deepesh Namdev, Saurabh Maheshwari (2018), "A Systematic Review of Automation in Handwritten Character Recognition" International Journal of Applied Engineering Research, Vol-13.
- [4] Malakar, Samir et al., (2012) "Text line Extraction from Handwritten Documents using spiral run-length Smearing Algorithm. "IEEE International Conference on Communications, Devices and Intelligent Systems (COCDIS).
- [5] Merial A, Prof. Jiberal Jos (2017), "Feature Extraction of Optical Character Recognition: Survey", International Journal of Applied Engineering Research ISSN 0973-4562 Vol-12.
- [6] Neumann.L and Matas.J(2012), "A Method for text localization and recognition in real world Images," in Asian Conference on Computer Vision.
- [7] Nitrogiannis, Konstantinos, Basilis Gatos, and Ioannis Pratikakis (2013), "A Performance Evolution Methodology for Historical Document Image Binarization" IEEE International Conference on Document Analysis and Recognition.
- [8] Romero, Veronica et al., (2013) "Handwritten Text Recognition for Marriage Register Book", International Conference on Document Analysis Recognition, IEEE.
- [9] Russakovsky.O, Deng, Hu.Su.J, Krause.J, Satheesh.S, Ma.S, Huang.Z, Karpathy.A, Khosla.A, Bernstein.M, Berg.A.C, and Fei-Fei.L (2015) "ImageNet Large Scale Visual Recognition Challenge, "Journal on Computer Vision, vol. 115, no. 3, pp.211-252.
- [10] Sebastian Sudholt, Gernot A.Fink (2017), "A Deep Convolutional Neural Network for word Spotting in Handwritten Documents" Department of Computer Science TU Dortmund University.
- [11] Surya Nath.R.S, Afseena.S (2015), "Handwritten Character Recognition A Review" International Journal of Scientific and Research Surya Publications, Volume5, Issue.
- [12] Savitha Attigeri (2018), "Neural Network Based Handwritten Character Recognition System" International Journal of Engineering and Computer Science ISSN:2319-7242, Vol 7, Issue 3 March 2018, pp 23761-23768.
- [13] Dhiraj Kumar Jasrotia, Aarti Malik (2018), "Webcam-based optical character recognition using MATLAB", International Journal of Engineering Sciences and Research Technology, Vol 7, pp 216-222.
- [14] Maninder Kaur, Manjeet Kaur (2017), "A Brief Review on Optical Character Recognition Techniques", International Journal of Computer Science and Mobile Computing, Vol 6, Issue 2, pp 95-100.
- [15] Ahmed Mahi Obaid, Hazem M.El Bakry, M.A.Eldosuky, A.I.Shehab (2016), "Handwritten Text Recognition Based on Neural Network", International Journal of Advanced Research in Computer Science and Technology, Issue1, Vol 4.
- [16] Narmata Dave (2015), "Segmentation Methods for Handwritten Character Recognition", International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol 8, pp 155-164.
- [17] Karez Abdulwahhab Hamad, Mehmet Kaya (2016), "A Detailed Analysis of Optical Character Recognition Technology", International Journal of Applied Mathematics, Electronics and Computers, Issue 4, pp 244-249.
- [18] Nikita V Borse, Prof I.R.Shaikh (2015), "Handwritten Text Extraction from documents", International Journal of Emerging Technology and Advanced Engineering, Vol 5, Issue 8