International Journal of Computer Applications Technology and Research
Volume 12–Issue 02, 50-58, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1202.1007

Number Recognition Techniques to Automate Revenue Collection at Dodoma Mini Bus Terminals in Tanzania

Kelvin Rweshobora
Eastern Statistical
Training Centre,
Dar es salaam, Tanzania

Eliah Kazumali
Eastern Statistical
Training Centre,
Dar es salaam, Tanzania

Ali Khelef
Eastern Statistical
Training Centre,
Dar es salaam, Tanzania

Abstract: The manual revenue collection process currently operating in Tanzania, has deprived the Local Government Authority (LGA) of its potential income. Meanwhile, The Number Recognition (NR), is a modern technique that uses optical character recognition on images to obtain desired characters. The technique involves Image acquisition, pre-processing, character segmentation and Character recognition. This technique has been used in activities like license plate detection as Automatic Number Plate Recognition (ANPR). This research attempts to solve the problem by automating the current revenue collection process at the mini bus terminals in Tanzania taking a case of Dodoma Municipal using Number recognition techniques and Optical Character Recognition (OCR).

The research adopted a case study research design and simulation. In addition, the study used key informant interviews and observations to acquire a good understanding of the current operations at the mini bus terminals and the necessary requirements to achieve the main goal of the study. Simulation of the Number Recognition System (NRS) was achieved by using Matlab R2017a as a simulation tool on a Dell computer running windows 7 professional, Hard Disk Drive (HDD) 500 Gigabyte (GB), Random Access Memory (RAM) 4 Gigabyte.

The research, achieved a simulation for NRS with an accuracy of 0.988 for the character recognition of the captured Surface on mini buses. The researcher recommends a further study on the image acquisition process and messaging alert system, to completely automate the process.

Keywords: Image Processing and Recognition, Number Recognition, Revenue Collection

1. INTRODUCTION

Local Government Authorities (LGA) in Tanzania, are responsible for providing socio- economic services to its people. To manage this task Tanzania has embarked on decentralization with the aim of making the system of governance more accountable, more open and transparent, and more democratic [1].

Although the decentralization aimed at enabling local governments to support themselves, many local government authorities failed to support their activities and depended on the central government due to unstable and poor planned sources of revenue [2]. Local governments have many sources of revenue which include hotels, shops, restaurants, fish licensing fees and bus stand fees to mention a few. Owing to this inability to support themselves, the local government authorities adopted an outsourcing revenue collection as a means to solve the revenue collection problem[3]. [4] explains the potential of Private Public Partnership (PPP) as a way to solve the revenue collection problem. However, this did not entirely solve the problem as some councils reported loss of revenue due to outsourcing revenue collection.

Furthermore [5],reported the existence of corruption and fraud activities in the entire process of outsourcing and mismanagement of the whole

process. The report revealed that private collectors submitted less than what was collected and agreed in the contract. As reported by [6], collection of market due dropped by average of 47.6%. This reveals the presence of fraud activities in the outsourcing of revenue collection to private collectors.

To prevent fraudulent activities, revenue collection in bus terminals is done with an agent using an Electronic Fiscal Device (EFD). The fee is collected as the mini bus leaves the station and a receipt is handed to the driver or the conductor. This situation gives room for corrupt collectors to benefit by cheating and not actually using the device. These problems necessitate to automate the process at bus terminals to easy and maximize revenue collection.

2. LITERATURE REVIEW

2.1 Introduction

The presence of a diverse nature of number plate format and languages, has hindered the development of a single application that would be deployed worldwide for license plate recognition. This has led to the development of different system for different countries [7]–[10]. Several approaches have been used to overcome these problems across countries, Argentina used Intelligent Template Matching (ITM), Australia approached the problem by using both fixed and mobile systems and Egypt, approached the problem by the creation of an organized database to be used for the number plate recognition system [11]. Most of the research conducted is based on License plate recognition [12],[13].

Many researchers, have ventured in the development of effective and efficient character recognition algorithms. Number Recognition Systems date back as far as 1976 where they were first studied in the scientific development branch Police in UK [14],[15];[16]. The technique has been used in several areas such as the road, boarders and toll gates. License plate recognition, has assisted in solving several queries which include traffic monitoring, stolen vehicle monitoring and managing parking toll [17].

There have been many attempts to develop an efficient algorithm for character recognition part of the process. Some of these detection algorithms are Mathematical morphology, structuring element, media filtering and edge detection [18]. The License Plate Recognition process involves three steps which include Image license positioning, character segmentation and character recognition [19].

2.2 Number plate recognition systems world wide

Introduced in the year 1976, Number Plates Recognition (NPR) has since found a wide commercial application, making its research prospects demanding and scientifically interesting [20]. Countries like China, Europe, India and Malaysia have had a fair share of application from toll collection, intelligent traffic surveillance systems, law enforcement to car theft tracking [18]; [21].[22], [23]designed an algorithm that serve the private and commercial plates, for Tanzania mainland plates which are yellow and white. The aim was to exclude diplomatic plates which have different color and format. Some of the difficulties that the other encountered were the variation of Tanzania plates shape, size and color which pose challenges in the detection process Fig 1.

Figure 1. Tanzania Number Plates

This research has used minibus identification number assigned by Surface and Marine Transport Regulatory Authority (SUMATRA) for the detection process. A few challenges existed since some of numbers may be clear while others are not as shown in Fig 2.

Figure 2. SUMATRA Bus Identification Number (Source: Own Processing**)**

2.3 Payment techniques

Payment mechanisms differ from one payment process to another. The choice of the payment method depends on the ease of access and time required to complete the process. For a busy area like a bus terminal, time is a critical factor to consider. Putting this in mind, several toll collecting booths in the world deployed different methods with accordance to the nature of activities. Payment methods may include cash, pass cards or credit card [21]. When the user is paying the toll in cash, it requires much time which causes a delay at the toll booth. Korea Expressway Corporation (KEC) employed an electronic payment system that enables the driver to use the prepaid card to directly pay the toll fee. This reduced the delay at toll booth to about 3 seconds from 11 seconds. Figure 3 show an automated payment process that is adopted in many countries and areas like the airports and shopping malls.

Figure 3. Flow Chart on E-Payment (Source:[24]

3. METHOD

3.1 Research design

The study adopted a case study research design and design science. The design science approach was used in creating an artifact that mimics the operation of image acquisition and automatic revenue payment. The artifact was created with the focus on a modified four step frame work that was designed by Kopparapu as shown in Fig 4.

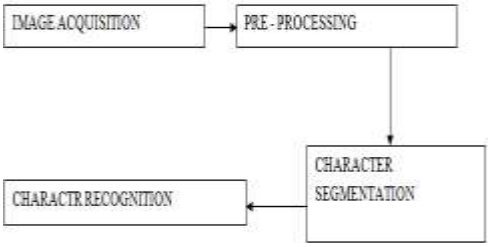

Figure 4. Flow Diagram (Source: Modified from [22]).

3.2 Image acquisition

The researcher used a mobile phone TECNO Y4 to capture image in the Red Green Blue (RGB) format for the study figure 3.1. Images were saved in a Joint Photographic Expert Group (JPEG also abbreviated as JPG) file format. The image acquisition process, involved the retrieval of images from a file in the computer system. These images were designed to meet the requirements for the image acquisition process.

3.3 Pre-processing

After capturing the image, it is then processed to convert it from the RGB to Gray scale. Since machine best understands grey images which are defined as a two-dimension function f(xy) where x and y are spatial coordinates and f is the light intensity at the region. A Gray scale image is a black and white representation of the converted true color image. Gray scale conversion was achieved through binarization. Detect MSER features function which was then applied to the gray scale image to obtain Region of Interest (RoI) through thresholding

3.4 Character segmentation

After obtaining the ROI, areas with the same color threshold were detected and considered as character segments. A mask is created to compare the region detected with the regions in the original image. The create Mask method returns a binary

image the same size as the input image, containing 1s inside the ROI and 0s everywhere else (MATLAB documentation, R2013)[25], [26]. This mask enables us to easily perform Optical Character Recognition (OCR) from the captured image.

3.5 Character recognition

Involved identification of the recognized numerals and matching them to the pre-existing database and checking if the number exist. The numbers were recognised and extracted from the image using the OCR function of MATLAB. Only numbers were checked because the SUMATRA mini bus identification number contains numbers and some short code representing the region.

3.6 Sample

The researcher has used a purposive sampling technique to achieve the targeted sample. Forty (40) plate numbers were used. Sampled mini buses were chosen based on the clarity/ visibility of their SUMATRA number Fig 5. This sampling technique was appropriate for the research as it saved time and enabled the researcher to get the right minibus with correct clearly visible SUMATRA numbers.

Figure 5. Captured SUMATRA Number (Source: Own Processing)

3.7 Data collection

To collect images of plate numbers to be used in model development, the researcher used a digital camera for capturing images of the side bus numbers (SUMATRA number) based on the number of samples required for the study.

3.8 Simulation tool

MATLAB as a simulation tool was used for this study. The software choice was influenced by the presence of a rich literature of image recognition systems developed and the presence of required libraries, for the image processing to demonstrate a complete simulation of the revenue collection process, database to represent the electronic payment process was designed by using Microsoft access database and linked to MATLAB by using Open Database Connectivity (ODBC) drivers.

4. RESULTS

4.1 Image acquisition

Image was acquired from a set of pre-designed images. Images are custom designed to meet the system specifications which are font style (Arial Unicode MS), font size (36) and character spacing (Normal). To acquire the image from a file in the computer, the imread() function of MATLAB was used. The image should reside in the same location or a path has to be defined.

>> image=imread('num6.jpg'); # Read image for processing

>> inshow(image); # Displays true color image Fig 6

International Journal of Computer Applications Technology and Research
Volume 12–Issue 02, 50-58, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1202.1007

Figure 6.Acquired Image

Figure7.Gray Scale Image Character Segmentation

4.2 Pre-processing

The acquired true color image Red Green Blue (RGB) is converted to the grayscale intensity image by the rgb2gray () function that converts RGB images to grayscale by eliminating the hue and saturation information while retaining the luminance. This is achieved by forming a weighted sum of the R, G, and B components Fig 7:

0.2989 * R + 0.5870 * G + 0.1140 * B

>> grayimage=rgb2gray(image); # Converts the true color image to a gray scale image

>> imshow(grayimage); # Displays the grayscale image

The detectMSERFeature extraction method with the gray image and a threshold value of 5 is used to segment the character part from the rest of the image thus obtaining the Region of Interest Fig 8.

>>imgRegions=detectMSERFeatures(grayimage,'ThresholdDelta',0.5); #threshold conversion

smtRegionsPixels=vertcat(cell2mat(imgRegions.PixelList))

imshow(image); hold on; # image display

plot(imgRegions,'showPixelList',true,'showEllipses',false); plotting the threshold results on the image figure 8.

International Journal of Computer Applications Technology and Research
Volume 12–Issue 02, 50-58, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1202.1007

Figure 8.Threshold Image

4.3 Character recognition

After the characters are segmented through the threshold and Maximally Stable Extremal Regions (MSER) feature detection, OCR function is applied to recognize and capture the characters which are then stored in the ocrtext object. By using the ocrtext.Word ( ) function, the required text can be displayed or used by other function in the developed system model. The "OCRTXT" object captures all text on the image. To obtain the required text, the ". Word" property is used. >> text=ocrtxt.Word ( ). In order to obtain the desired character, the object OCRTXT.Word(2:3,1) was used to specify the character from the row two and three of the first column in the OCRTXT object Fig 9.

Figure 9.Detected Character

Table 1.OCR results for recognition Accuracy

Sumatra Number Image	Recognised Characters	Ocr Word Confidences
	400 DOM	0.8908 (Sumatra number) 0.8978 (region)
	500 DOM	0.8908 (Sumatra number) 0.8978 (region)
	490 DOM	0.9111 (Sumatra number) 0.8776 (region)
	7000 DOM	0.9278 (Sumatra number) 0.8706 (Region)
	900 DOM	0.8912 (Sumatra number) 0.8741 (region)

The table 1 displays the probability of returning the accurate reading from the captured SUMATRA number. The word confidences ranging from 0 to 1, shows the probability of the captured number being correct. The findings show that, the designed model has a recognition accuracy of 0.90234 which is an average of the five attempts, to recognize different SUMATRA numbers. Finding the average of recognition show the extent to which the obtained results are accurate. The sum of the overall Word Confidences is obtained and divided

by the number of images captured and processed to obtain the degree of accuracy of the recognition process equation (1).

$$ARWC = \frac{Sum\ of\ Word\ Confidence}{Number\ of\ processed\ images} \quad (1)$$

Where ARWC is Average Recognition Word Confidences

4.4 Results recommendation

This research main objective was to simulate the automation of the revenue collection process at the Dodoma Municipal mini bus terminals. Due to limitations encountered, the research used own processed images and proposed a Microsoft publisher template, for the creation of the SUMATRA numbers that will be similar for all the mini busses, operating at the Municipal. The template will eliminate variation among SUMATRA numbers on mini busses. The research also recommends a standard or guide from SUMATRA for both number and position of the number, on the side of the mini bus. The guide will create conducive environment for the development and implementation of the simulated revenue collection process.

5. CONCLUSION

Number recognition and character recognition are growing technologies with a number of applications from the health sector, engineering, education to business. The potential to solve these problems has seen these technologies being applied in our daily lives, including a face recognition for security, object identification and number plate recognition for traffic purposes.

This study has employed number recognition technique and OCR to obtain SUMATRA numbers on the side of mini buses, at the Dodoma Municipal bus terminals for the automation of revenue collection. The developed simulation has the potential to solve revenue collection problems facing LGA's. Deployment of such processes at the desired terminals would increase the revenue collection and eliminate if not reduce, fraud activities which the current manual system is prone to.

6. ACKNOWLEDGMENTS

Our great thanks to all who have contributed in one way or another in the completion of this paper.

7. REFERENCES

[1] B. E. S. Mgonja and C. Tundui, "Institutional Impacts of the Local Government Reform Program on Good Local Governance in Tanzania," *International Journal of Academic Research in Business and Social Sciences*, vol. 2, no. 5, pp. 206–222, 2012.

[2] L. Government, "The United Republic of Tanzania Priminister's Office , Final Report A Study on LGAs Own Source Revenue Collection," 2013.

[3] M. Lukio, L. Mrutu, and P. Mganga, "Outsourcing or Giving all Out? Experience from Tanzania Local Government Revenue Collection," *Journal of Public Administration and Governance*, vol. 6, no. 3, pp. 1–11, 2016, doi: 10.5296/jpag.v6i3.8695.

[4] H. P. Ngowi, "Public-Private Partnership (PPPs) in the Management of Municipalities in Tanzania – Issues and Lessons of Experience Mzumbe University," *African Journal of Public Administration and Management*, vol. 17, no. 2, pp. 1–18, 2006.

[5] O.-H. Fjeldstad, L. Katera, and E. Ngalewa, "Outsourcing Revenue Collection : Experiences from Local Government Authorities in Tanzania," *Public Adm*, no. until, 2005.

[6] P. P. Mabhuye, "The Performance of Outsourced Revenue Collection System in Local Government Authorities Case Study of Kasulu District," 2013.

[7] C. Patel, D. Shah, and A. Patel, "Automatic Number Plate Recognition System (ANPR): A Survey," *Int J Comput Appl*, vol. 69, no. 9, pp. 21–33, 2013, doi: 10.5120/11871-7665.

[8] A. Sarkale, K. Shah, A. Chaudhary, and T. P. N, "A Literature Survey: Neural Networks for object detection," 2018. [Online]. Available: www.viva-technology.org/New/IJRI

[9] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," May 2019, [Online]. Available: http://arxiv.org/abs/1905.05055

[10] Lubna, N. Mufti, and S. A. A. Shah, "Automatic number plate recognition:A detailed survey of relevant algorithms," *Sensors*, vol. 21, no. 9, pp. 1–35, 2021, doi: 10.3390/s21093028.

[11] A. Puranic, "Vehicle Number Plate Recognition System : A Literature Review and Implementation using Template Matching," vol. 134, no. 1, pp. 12–16, 2016.

[12] A. goyal and R. Bhatia, "Automated Car Number Plate Detection System to detect far number plates," *IOSR J Comput Eng*, vol. 18, no. 04, pp. 34–40, 2016, doi: 10.9790/0661-1804033440.

[13] A. Verma and D. Kumar Singh, "Text Deblurring Using OCR Word Confidence," *I.J. Image, Graphics and Signal Processing*, vol. 1, no. 1, pp. 33–40, 2017, doi: 10.5815/ijigsp.2017.01.05.

[14] J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," *IEEE Access*, vol. 8. 2020. doi: 10.1109/ACCESS.2020.3012542.

[15] F. A. Memon, "Vehicle Number Recognition System for Automatic," 2012.

[16] N. Bohra, M. Lather, N. Chawla, and A. Kumari, "License Plate Recognition," no. Icicc, pp. 1–5, 2020.

[17] A. Badr, M. M. Abdel, A. M. Thabet, and A. M. Abdelsadek, "Automatic number plate recognition system," *Annals of the University of Craiova, Mathematics and Computer Science Series*, vol. 38, no. 1, pp. 62–71, 2011.

[18] P. S. Krishna, "Automatic Number Plate Recognition by Using Matlab," vol. 2, no. 4, pp. 1–7, 2015.

[19] S. Gupta, R. Singh, and H. L. Mandoria, "A Review Paper on License Plate Recognition System," *iMedPub Journals*, pp. 6–8, 2020, doi: 10.36648/ijircce.05.01.04.

[20] R. Panahi and I. Gholampour, "Accurate Detection and Recognition of Dirty Vehicle Plate Numbers for High-Speed Applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 767–779, Apr. 2017, doi: 10.1109/TITS.2016.2586520.

[21] D. Mitra and S. Banerjee, "Automatic Number Plate Recognition System : A Histogram Based Approach," vol. 11, no. 1, pp. 26–32, 2016, doi: 10.9790/1676-11142632.

[22] I. Bulugu and P. Zhijun, "Volume 2 Issue 12, December 2013 www.ijsr.net An Improved Method for Tanzania Number

Plate Location and Segmentation Based on Mathematical Morphology and Regional Features of an Image," *International Journal of Science and Research*, 2010, [Online]. Available: www.ijsr.net

[23] J.-W. Hsieh, "Morphology-based license plate detection in images of differently illuminated and oriented cars," *J Electron Imaging*, vol. 11, no. 4, p. 507, 2002, doi: 10.1117/1.1501575.

[24] S. K. Kopparapu, "Mobile Phone Based Vehicle License Plate Recognition for Road Policing," pp. 1–7, 2015, [Online]. Available: http://arxiv.org/abs/1504.01476

[25] P. S. Krishna, "Automatic Number Plate Recognition by Using Matlab," vol. 2, no. 4, pp. 1–7, 2015.

[26] J. Hansen, "A Matlab Project in Optical Character Recognition (OCR)," *DSP Lab, University of Rhode Island*, vol. 6, 2002.