ISSN 2278-9723
DOI:10.31838/ijccts/10.01.03

Research Article

Untangling Pancard By Designing Optical Character Reader Tool Box By Correlating Alpha Numeric Character

MADUGURI SUDHIR[1], KASANI MANEESHA[2], GUJJARLAPUDI ANUDEEPTHI[3], TOOMATI ANUSHA[4], ACHANA CHANDINI[5]

[1]Department of Electronics and Communication Engineering, Assistant Professor, KKR& KSR Institute of Technology and Sciences, Guntur

[2,3,4,5]Department of Electronics and Communication Engineering, UG student KKR& KSR Institute of Technology and Sciences, Guntur

Email: Sudhir3801@gmail.com [1], maneeshakasani@gmail.com[2], anudeepthi.5302@gmail.com[3], anushatoomati@gmail.com[4],chandini.achana@gmail.com[5]

Received: 07.01.22, Revised: 13.02.22, Accepted: 14.03.22

ABSTRACT

Nowadays our society is mostly depending on the digitalization and automation. The main aim of this project is developed OCR code for decoding the pan card number. OCR is an optical character recognition and is the mechanical or electronic translation of images. OCR is a field of researching pattern recognition, artificial intelligence and machine vision.

In this project we can decode the pan card number and it gives the details such as income tax transactions, number of loans taken, cibil score, returns and loan status of the card holder so that we can know the every information of the holder. We can get the information easily about how many are paying the income tax to the government regularly. For this we are using MATLAB software. In this we can track the details of owner by taking the photograph of pan card and that will be stored, processed in the OCR and details will be displayed

Keywords: OCR, templates, correlation, labeling.

I. Introduction

A permanent account number (PAN) is a ten-character alphanumeric identifier, issued in the form of a laminated "PAN card", by the Indian Income Tax Department, to any person who applies for it or to whom the department allots the number without an application. It can also be obtained in the form of PDF file.

A PAN is a unique identifier issued to all judicial entities identifiable under the Indian Income Tax Act 1961. The income tax PAN and its linked card are issued under section 139A of the Income Tax Act. It is issued by the Indian Income Tax Department under the supervision of the Central Board for Direct Taxes (CBDT) and it also serves as an important proof identification.

Fig.1: Pan card Image

II. Proposed Method

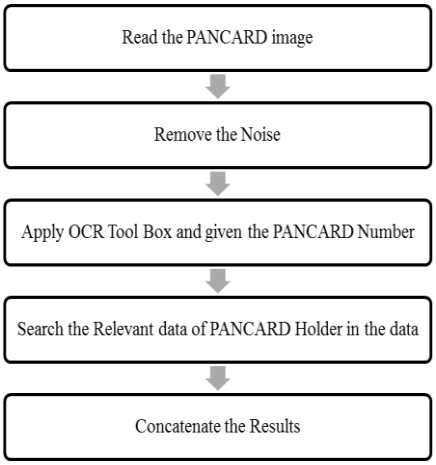

Fig.2:Flow Chart

We observed many disadvantages in the existing system to overcome the disadvantages discussed earlier we have proposed a new method in which we developed OCR function to scan the PAN card.

So, as to develop this OCR function we had written an image processing algorithm for extracting the holder details, such as transaction history, loan status, etc.

OCR technology is used for automating the data extraction from an image file then converting the text into machine readable form to be used for data processing like editing and searching. It provides a complete form processing and documents capture solution. Usually, OCR uses a modular architecture that is open scalable and work flow controlled. It includes forms definition, scanning, image pre-processing, and recognition capabilities. OCR is a technology that converts different types of documents such as pdf files or image captured by digital camera into editable and searchable data.

The proposed method mainly designed for decoding the PAN to track the details by scanning pan card or taking photograph of pan card, that will be stored, processed in the OCR and the details will be displayed.

It is a simple procedure for acquiring holder details such as transaction history, loan status, etc., just by scanning the pan card.

2.1 Advantages of Proposed Method

- It offers more security.
- Time saving process.
- Flexibility for customers.
- Quick Response.
- Low cost.
- Proofreading
- Layouts

2.2 Algorithm explanation

Our project start with reading of PANCARD image. The next step is used to remove the noise and select the desired area in the pancard. Now the the selected area is seperated and OCR is applied on the selected area. Now the text in image is converted into text. Now the decoded text information is serched in the database. Now the information of the person is displayed on the screen.

2.3 Modules in algorithm

Our algorithm contains 03 modules those are

01) Module-01 Removing Noise
02) Module-02 Designing OCR Tool Box
03) Module-03 Searching Algorithm

III Ocr Tool

OCR stands for optical character recognition. Recognizing text in images is useful in many computer vision applications such as image search, document analysis, and robot navigation. The OCR function provides an easy way to add text recognition functionality to a wide range of applications.

The OCR functions returns the recognized text, the recognition confidence, and the location of the text in the original image. We can use this information to identify the location of misclassified text within the image.

OCR performs best when the text is located on a uniform background and is formatted like a document. When the text appears on a non-uniform background, additional pre-processing steps are required to get the best OCR results.

Optical Character Recognition is a technology that enables you to convert different types of documents, such as scanned paper documents, PDF files or images captured by a digital camera into editable and searchable data.

Fig.3: OCR

Optical character recognition or optical character reader (OCR) is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo) or from subtitle text superimposed on an image (for example: from a television broadcast)

V. Results

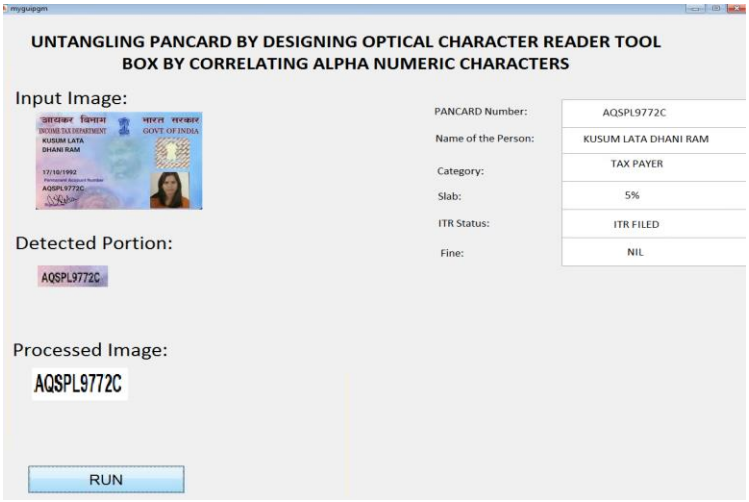

Fig.4:Result

IV Advantages

1. Data Security
2. Cost Reduction
3. Human Efforts will be Reduced
4. High Accuracy
5. 100% Text-searchable Document
6. Makes Documents Editable
7. Disaster Recovery
8. Purchase & Sale of immovable Assets
9. Identity Proof
10. TaxDeductions

Conclusion

In this project by using MATLAB software we tracked the details of owner by taking the photograph of pan card. We developed the OCR code for decoding the pan card number. Later we displayed every information of the card holder such as income tax transactions, no. of loans taken, cibil score, returns and loan status of the card holder.

References

1. R. Smith. An overview of the Tesseract OCR Engine, Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) Vol 2 (2007), pp. 629-633.
2. Smith, R., D. Antonova, and D. Lee. Adapting the Tesseract Open Source OCR Engine for Multilingual OCR. Proceedings of the International Workshop on Multilingual OCR, (2009).
3. Maduguri Sudhir, Cenikala Baswanth Vignesh, Chintalacheruvu Jairam, Dasari Anitha, Tunuguntla Sreeja, “Image Based Vehicle Number Plate Decoding,” International Journal

of Research in Advent Technology, Vol. 5, No. 1, January 2017.
4. R.Adolphs Cognitive neuroscience of human social behavior Nature Reviews Neuroscience 4 (3)2003 165–178.
5. J.F. Cohn, Foundations of human computing: facial expression and emotion, in: Proceedings of the ACM International Conference on Multimodal Interfaces, 2006, pp. 233–238.
6. J.F. Cohn, P. Ekman, Measuring facial action by manual coding, facial EMG, and automatic facial image analysis, in: J.A. Harrigan, R. Rosenthal, K.R. Scherer (Eds.), Handbook of Nonverbal Behavior Research Methods in the Affective Sciences, 2005, pp. 9–64. 11