

OPTICAL CHARACTER RECOGNITION THROUGH MACHINE LEARNING
TECHNIQUES

(SECOND REVIEW DOCUMENT)

ABSTRACT

Design and developing a web application for personal identity verification. In this application identity of a person is scanned through a camera or device and information is extracted from this document . The information that extracted from the id is stored and cross matched with details in our hand. Through this proofing it can ensure that applicants are who they claim to be.

CHAPTER-1

INTRODUCTION

1. STATEMENT OF THE PROBLEM

Digital identity is the cornerstone of a complete digital experience that customers now expect. Common tasks such as enrolling for a new service, logging into an existing service, making changes to an existing account, or performing a payment all rely on customers being able to prove who they are. Without a verified digital identity, companies open themselves up to fraud, regulatory penalties and exposure to other security risks. It is with this in mind we are on this venture to develop a web based application for verifying the identity through machine learning techniques. Here the data extracted through OCR algorithms

CHAPTER-2

SYSTEM ANALYSIS

2.1 PRESENT SYSTEM

Traditional technique available for different service providers like government organisations , telecom and other sectors used are out dated methods and anyone can make fraud. The technique that used traditionally for verifying the identity is through document verification which is signed by a designated person by cross checking different other related documents and it is a time taking process. And also anyone can make forged documents and submit at verification process. Through our system the identity can verified easily with in a matter of time. The data is simply scanned and through OCR algorithms the data extracted and cross matched with an API for those identity which are provided

2.2 LIMITATIONS OF PRESENT SYSTEM

- a) Documents submitted for verification can be forged one
- b) The process of verification is time taking process.
- c) Reduces operational efficiency
- d) Authorities who are verifying can be bribe by personal

2.3 PROPOSED SYSTEM

A web application that developed for personal identity verification. In this application identity of a person is scanned through a camera or device and information is extracted from this document. The information that extracted from the id is stored and cross matched with details in our hand. Through this proofing it can ensure that applicants are who they claim to be.

2.4 ADVANTAGES AND FEATURES OF PROPOSED SYSTEM

Advantages of proposed system

Various Service providers like government, telecom and other sectors can leverage identity verification technology to

- Reduce identity fraud
- Adapt to local regulations and business needs
- Streamline customer on boarding across multiple channels
- comply with know your customer (KYC) regulations : Anti Money laundering, Combating the financing terrorism

Features of proposed system

SI.NO	FEATURES
1	REMOTE ID CAPTURE
2	EXTRACTING INFORMATION
3	TEMPLATE MAPPING
4	AUTOMATIC FORM FILLING
5	CROSS MATCHING DOCUMENT
6	VERIFICATION MESSAGE

1.Remote ID Capture

Document capture is done by providing real-time indications to user to help them to scan their ID with device camera .The device camera enables the capture of identity document data by taking a picture or simply uploading the documents.

2. Extracting Information

From captured image the data is extracted through optical character recognition algorithms. Optical Character Recognition (OCR) is a technique of reading or grabbing text from printed or scanned photos, handwritten images and convert them into a digital format that can be editable and searchable.

OCR has plenty of applications in today's business. A few of them are listed below:

- Passport recognition in Airports
- Automation of Data Entry
- License plates recognition
- Extracting business card information into a contact list
- Converting handwritten documents into electronic images
- Creating Searchable PDFs
- Create audible files (text to audio)

Some of the Open Source OCR tools are Tesseract, OCRopus

Here we use Tesseract Algorithm for extracting informations from an image.This algorithm is able to accurately decipher and extract text from a variety of sources! As per it's namesake it uses an updated version of the tesseract open source OCR tool. We also automatically binarize and preprocess images using the binarization so tesseract has an easier time deciphering images.

3. Template Mapping

Here we set a template for organizing the extracted information in an arranged manner. So the information that we extracted from an id is resulted in an ordered manner and stored in DB.

4. Automatic Form Filling

Automatic Form Filler is a simple, time saving WebApp that will automatically fill the website, Client will need to manually submit the forms. It will however save a lot of time filling out the forms.

5. Cross Matching Document

Cross Matching Document is comparing the two or more Response. The cross matching verifies the API and identifies whether it is valid or not. API verification uses a variety of increasingly sophisticated and frequently updated techniques.

6. Verification Message

Sending a verified message is a two-step process: sending the message content to the user, through Web Application. Verification Message is a kind of a transactional message sent to a User after a certain condition is triggered. This message can be sent for the ID is valid or not.

2.5 Feasibility Study

The main aim of the feasibility study activity is to determine. Whether it would be financially and technically feasible to develop the product. The feasibility study activity involves analysis of the problem and collection of all relevant information relating to the product such as the different data items which would be input to the system the processing required to be carried out of these data, the output data required to be carried out of these data, the output data required to be produced by the system, as well as various constraints on the behaviour of the system.

DEPARTMENT OF MCA

- a) Technical Feasibility
- b) Economic Feasibility
- c) Operational Feasibility

Technical Feasibility

In technical feasibility the management determines whether the current level of technology can support the proposed system..Considering all the advantage of the proposed system time consumption for accessing a website and retrieval of information has been reduced

Operational Feasibility

Proposed system would be beneficial only if they can be turned in to information system that will meet the organization operating requirements. One of the main problems faced during the development of a new system is getting acceptance from user.

Economic Feasibility

Economically, this project doesn't raise any issue, as the project itself is planned as the website. The resource required for this project is minimum. This system doesn't demand any additional equipment.

- a) The resource required for our project is minimum and the system doesn't demand any additional equipment so our project is economically feasible
- b) The system is economic with the hospital's point of view
- c) It is cost effective as it eliminates the paper works

CHAPTER – 3

SYSTEM SPECIFICATION

3.1 MINIMUM SOFTWARE REQUIREMENTS

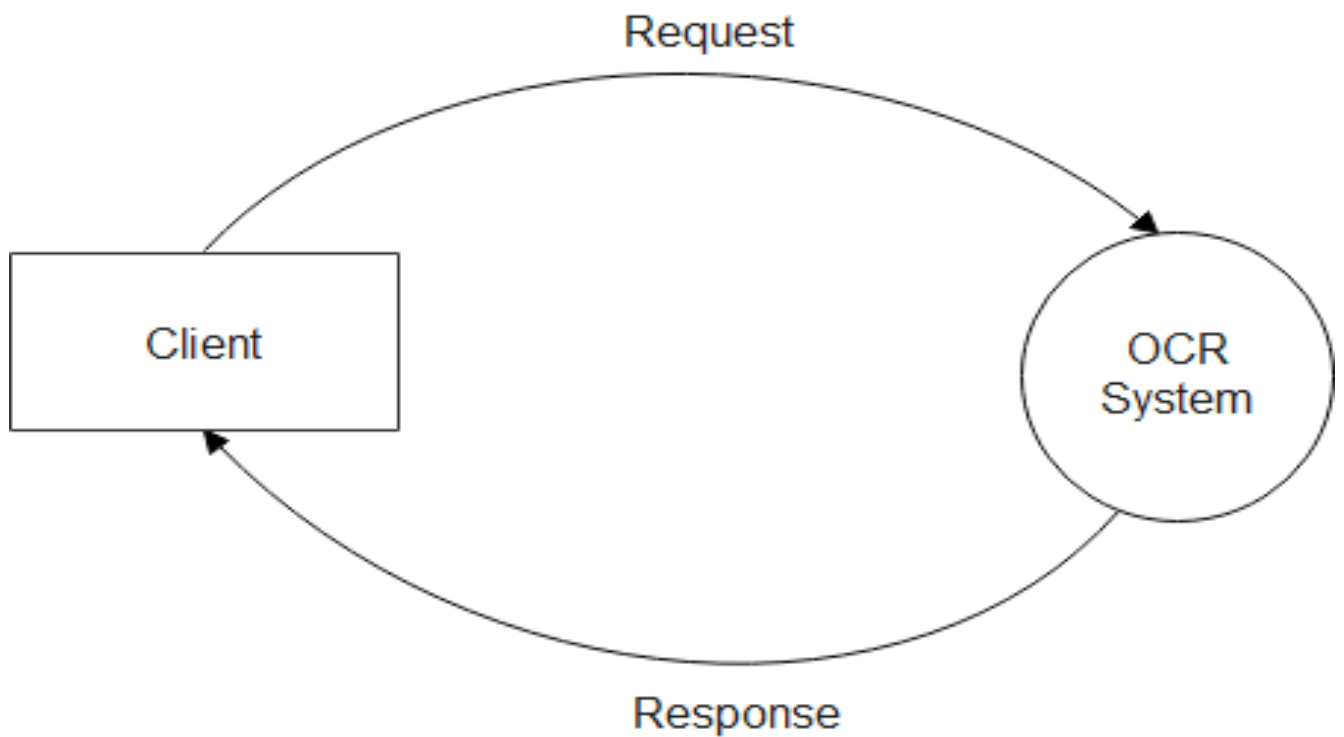
1.Operating System	:	Windows / Mac Os
2.Language	:	Python
3.Frontend	:	ReactJS
4.Framework	:	Flask
5.IDE	:	Pycharm
6.Libraries	:	Opencv, Pytesseract,

3.2 MINIMUM HARDWARE REQUIREMENTS

1.Processor	:	Intel i5 or AMD Ryzen 3
2.RAM	:	3 GB
3.Hard Disk Drive	:	300GB
4.Peripherals	:	Keyboard, Mouse,Camera, Monitor

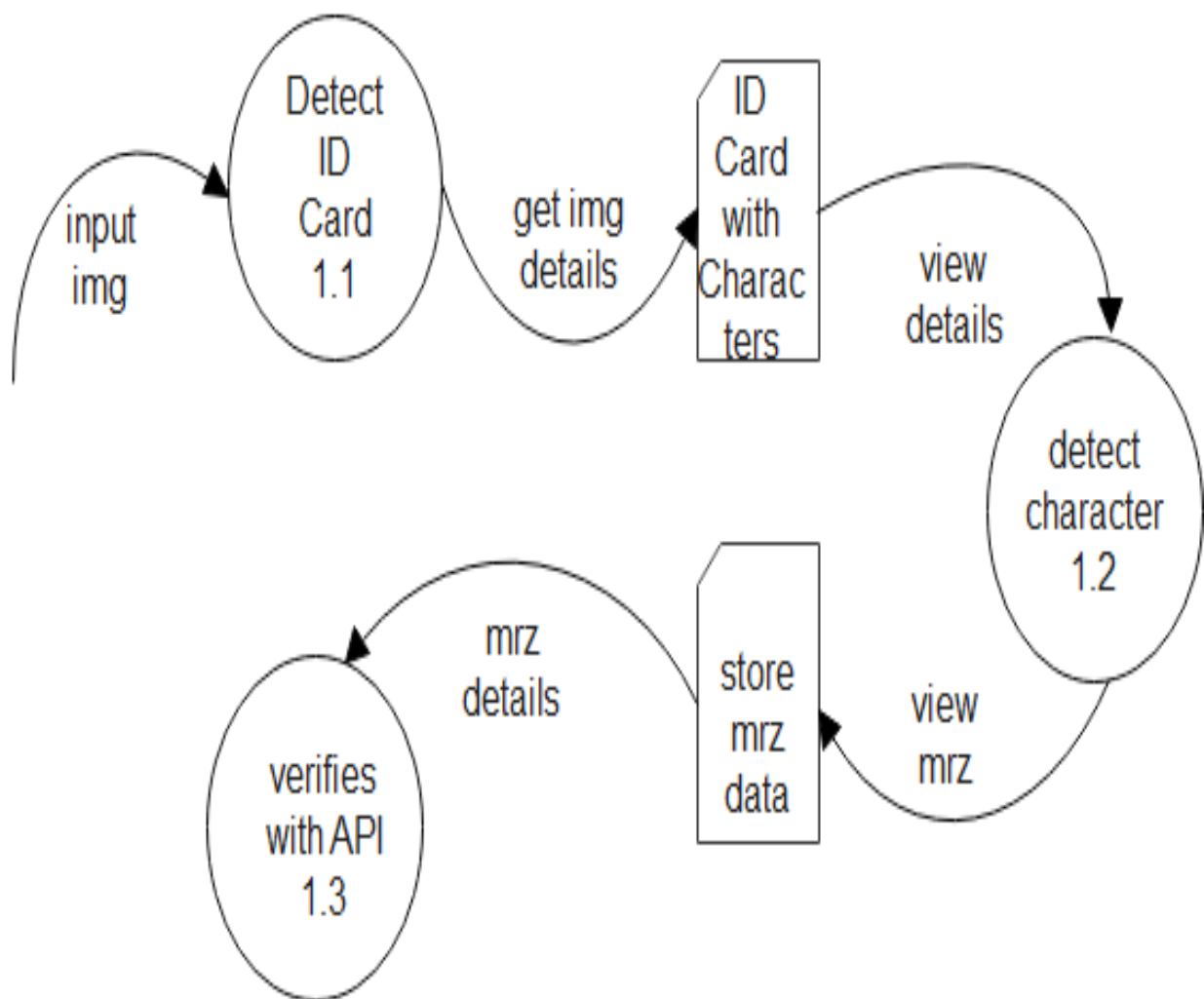
CHAPTER – 4
SYSTEM DESIGN

4.1 CONTEXT LEVEL DIAGRAM

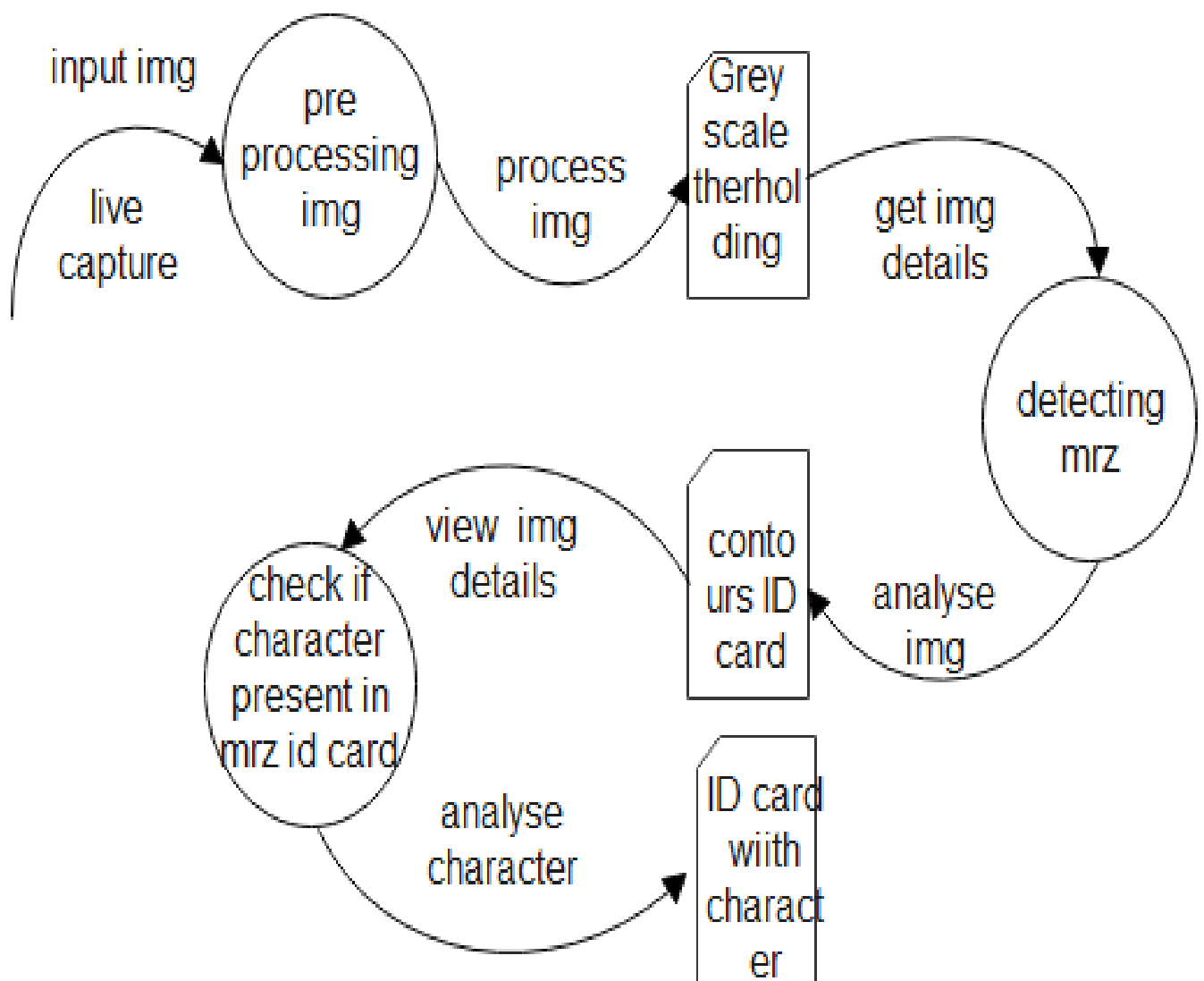


4.2 DATA FLOW DIAGRAM

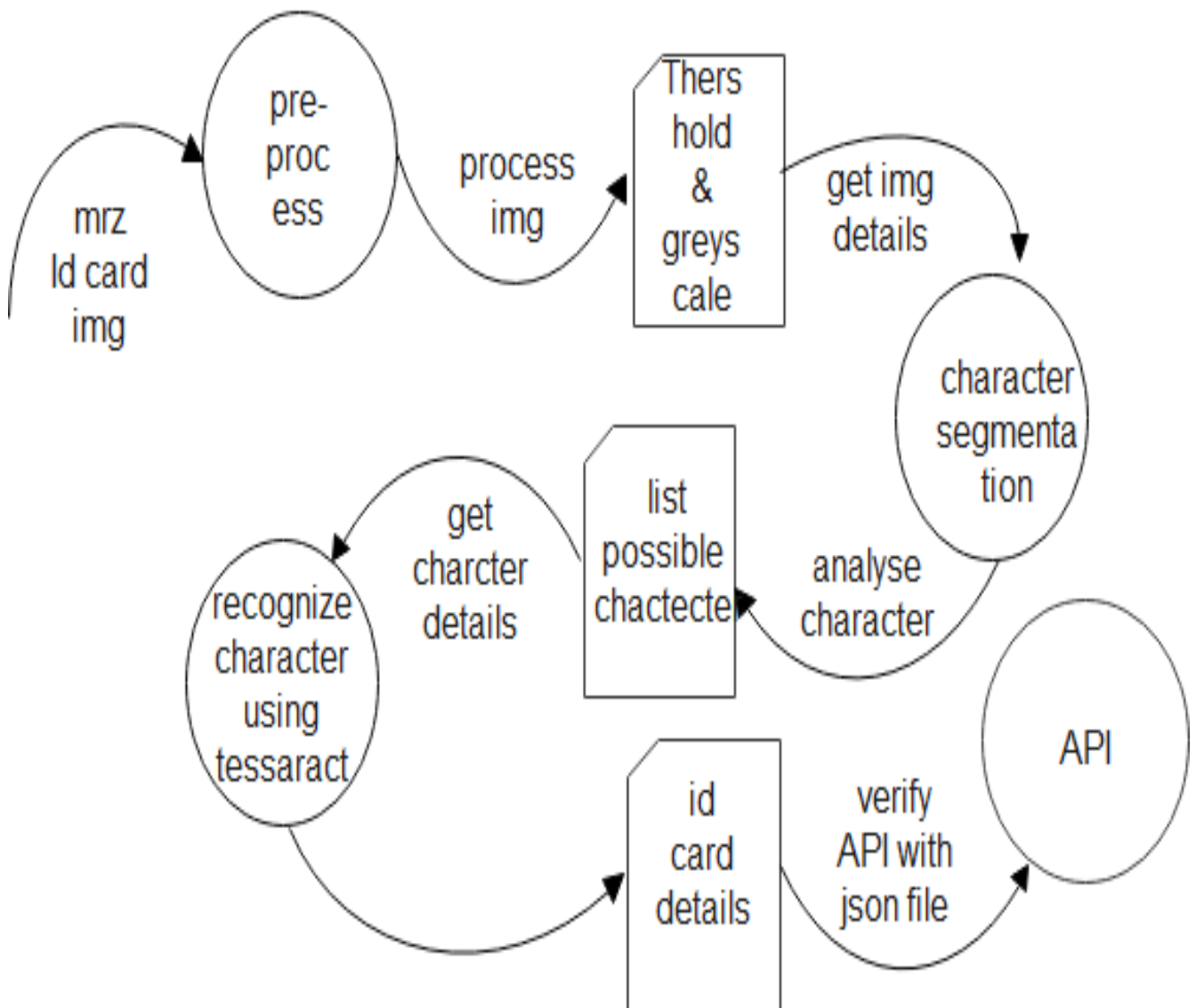
LEVEL 1: DFD OF OCR SYSTEM



LEVEL 2 : DFD OF CAPTURED IMAGE PROCESSING



LEVEL 3: DFD OF MACHINE READABLE ZONE IN ID CARDS



4.3 DESIGN OF EACH SUB SYSTEM

