

**NewEyes**

**Проект №218 Съвременни системи за визуална информация**

1. **Тема**: NewEyes – Очила за незрящи Хора
2. **Автор:**

**Иван Мирославов Цветков**

ЕГН: 0346233967

Адрес: Климент Охридски №12, гр. Долни Дъбник

Тел: 0899348344

Имейл: [i.tsvetkovbg@gmail.com](mailto:i.tsvetkovbg@gmail.com)

МГ “Гео Милев” Плевен, 12 клас

1. **Ръководител:**

**Илко Симеонов**

Тел: 088 518 1556

Имейл: [ilko.simeonov2000@gmail.com](mailto:ilko.simeonov2000@gmail.com)

Позиция: Програмист

Проект №218, Съвременни системи за визуална информация 

**4. Резюме**

**4.1. Цели:**

Целта на проекта е да подобри живота на незрящите / хората с тежко увредено зрение, като бъде неизменен помощник в тяхното ежедневие. В зависимост от това от което се нуждае незрящия в дадена ситуация, са предвидени различни по вид и функция функционалности (модове). Всяка функционалност изпълнява строго дефинирана функция, като това помага за лесната използваемост на устройството.

Наличните функционалности, които ще се поддържат от устройството са:

Избираеми функции:

* Четене на книжен/печатен текст
* Четене на текст в необичайна среда
* Единично разпознаване на обекти
* Разпознаване на обекти в реално време
* Позвъняване на познат при нужда от помощ
* Разпознаване на QR & Barcode кодове
* Сканиране на касови бележки
* Разпознаване на валута / банкноти
* Разпознаване на главен цвят в снимка
* Разпознаване на лица и емоции

Автоматични функции работещи винаги след стартирането на устройството:

* Разпознаване на жестове с ръце
* Умен гласов асистент

**Програмният код е лицензиран под GNU General Public License v3.0**

За разработката на проекта се използва **Github**, което представлява децентрализирана система за контрол на версиите на файлове. Програмният код и примери от главните функционалности на проекта са достъпни на следния адрес:

<https://github.com/gonogo312/NewEyes>

Проект №218, Съвременни системи за визуална информация 

**4.2. Основни етапи в реализацията на проекта:**

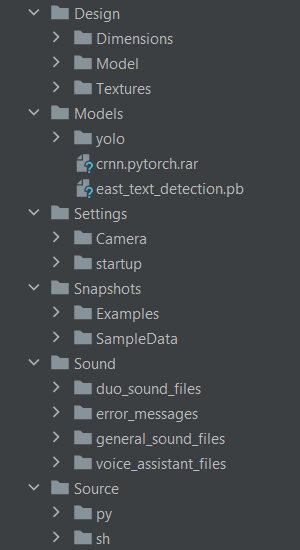
* Създаване на идеята
* Избиране на подходящи технологии
* Обмисляне на архитектурата и взаимодействието между различните модули
* Детайлно проучване нуждите на клиента
* План за реализация на проекта
* Разработка на устройството (хардуер/софтуер)
* Тестване на главните функционалности чрез помощта на незрящи доброволци
* Обработка на допълнителни клиентски изисквания / предложения
* Изкарване в продъкшън среда (готов за използване продукт)
* Бъдещи планове

**4.3. Ниво на сложност на проекта:**

В проекта са използвани множество модерни технологии и различни модули хардуер. Използваните технологии и библиотеки са проучени и тествани в реална среда, така че да изпълняват различните по вид функционалности на устройството по най-оптималният начин. Поради гореспоменатите причини определям „NewEyes“ като проект от високо ниво на сложност.

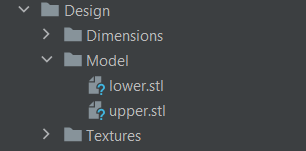
**4.4. Логическо и функционално описание на решението:**

**4.4.1. Генерална архитектура**



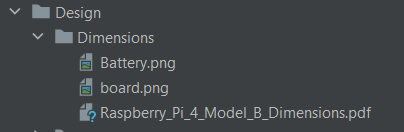
Проект №218, Съвременни системи за визуална информация 

* Design
  + Model



Upper.stl и съответно Lower.stl представляват двете части на модела от които е изградена кутията на устройството.

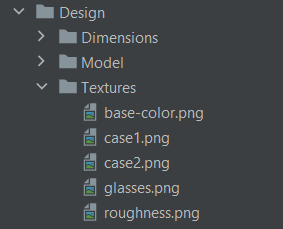
* + Dimensions



Съдържа схемите по които е изграден модела на кутията, съответно размерите на всеки елемент от хардуера в нея – едноплатков компютър Raspberry Pi 4, литиева батерия PiSugar2 Pro, SIM карта и други.

Проект №218, Съвременни системи за визуална информация 

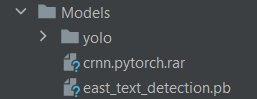
* + Textures



Съдържа отделните части на модела като изображение, цялостният модел и цветова гама на модела.

* Models

Yolo, CRNN.PythonTorch и east\_text\_detection моделите дават способността на устройството да разпознава текст и обекти по най-оптималният начин в съотношение с изчислителната мощност на хардуерните компоненти.



* + Yolo

Съвкупност от различни модели, използвани за разпознаване на обекти.

* + CRNN.PythonTorch

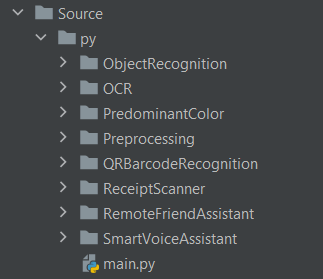
Модел и код за трениране на модела. Прилага се при разпознаването на текст в необичайна среда (функционалност №2) – четене на текст от табели, тролейбуси, коли и други идентични източници.

* + East\_text\_detection.pb

Подобен модел, използван при четенето на текст.

Проект №218, Съвременни системи за визуална информация 

* Source
* **Python**



* + - OCR

NewEyes.Source.python.OCR.book\_ocr дава функционалността за четене на печатен текст.

NewEyes.Source.python.OCR.crnn\_ocr предоставя функционалността за четене на текст в необичайна среда.

* + - ObjectRecognition

NewEyes.Source.python.ObjectRecognition.yolo съдържа класовете и методите за разпознаването на обекти, тяхната позиция в пространството и разстоянието до всеки обект.

NewEyes.Source.python.ObjectRecognition.start\_obj\_recog

предоставя функционалността за разпознаване на обекти от снимка.

NewEyes.Source.python.ObjectRecognition.realtime\_yolo\_obj\_recog предоставя

функционалността за разпознаване на обекти в реално време.

* + - PredominantColor

NewEyes.Source.python.PredominantColor.predominant\_clr дава функционалността за разпознаване на главен цвят от снимка.

Проект №218, Съвременни системи за визуална информация 

* + - Preprocessing

NewEyes.Source.python.Preprocessing.image\_processing пази методите използвани за обработката на снимките.

* + - QRBarcodeRecognition

NewEyes.Source.python.QRBarcodeRecognition.qr\_barcode функционалност за разпознаване на QR & Barcode кодове в реално време.

* + - ReceiptScanner

NewEyes.Source.python.ReceiptScanner.receipt\_scanner съдържа функционалността за сканирането и четенето на касови бележки.

* + - RemoteFriendAssistant

NewEyes.Source.python.RemoteFriendAssistant.duo\_google пази функционалността за позвъняване на друг човек чрез мобилният му номер и фийд от камерата чрез платформата Duo.

* + - SmartVoiceAssistant

NewEyes.Source.python.SmartVoiceAssistant.voice\_recog дава функционалността за Speech-to-Text, което се използва при функционалността гласов асистент.

* + - HandGestureRecognition

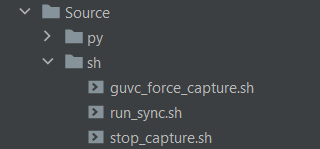
NewEyes.Source.python.HandGestureRecognition дава функционалността за разпознаване на жестове с ръце.

* + - Main.py

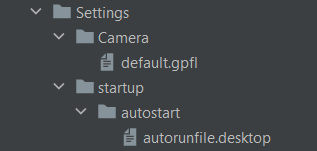
Съдържа основният код, чрез който потребителят навигира през списъка с функционалностите, изпълнява избрана функционалност. При стартиране се създават нишки (Threads) за функционалностите SmartVoiceAssistant и HandGestureRecognition.

Проект №218, Съвременни системи за визуална информация 

* + **Bashsc**



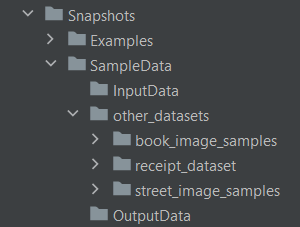
* + - Guvc\_force\_capture.sh прави снимка чрез камерата на очилата чрез програмата guvcview (софтуер използван за обработката на UVC потоци).
    - Stop\_capture.sh изключва guvcview след като се направи и запази снимката.
    - Run\_sync спрямо въведена часова зона, синхронизира часовника на устройството чрез NTP (Network Time Protocol).
* Settings



* Camera
  + NewEyes.Settings.Camera.default.gpfl съдържа настройки за конфигуриране на параметри като Focus, Brightness, Contrast, Saturation и други, използвани при правенето на снимка чрез софтуера guvcview.

Проект №218, Съвременни системи за визуална информация 

* Startup/autostart
* NewEyes.Settings.startup.autostart.autorunfile.desktop съдържа инструкции чрез които при включване на устройството се изпълнява основният код, main.py.
* Snapshots



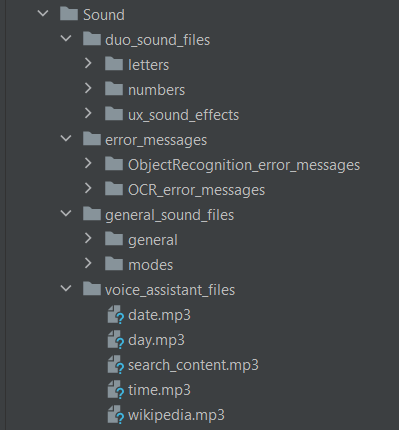
* Examples

Съдържа примери от главните функционалности и техните резултати.

* SampleData/other\_datasets съдържа примери използвани при тестването на функционалностите и dataset използван при тестването на функционалността ReceiptScanner.

Проект №218, Съвременни системи за визуална информация 

* Sound



Съдържа аудио файловете във формат \*.mp3 и \*.wav използвани при различните функционалности, разделят се в 4 основни категории:

* General\_sound\_files
* Duo\_sound\_files
* Error\_messages
* Voice\_assistant\_files

**4.4.2. Използвани библиотеки:**

* **GoogleSpeech –** Speech класа позволява създаването на \*.mp3 файлове с даден текст.
* **GpioZero –** Button класа позволява достъпа до бутоните.

Проект №218, Съвременни системи за визуална информация 

* **OS & MPG312 –** OS библиотеката позволява извикването на MPG312 music player, който прави възможно извикването и изпълняването на аудио файлове.
* **Imutils –** Предоставя базови функции за обработка на снимки, като някои от тях са non\_max\_suppression, label\_contour, four\_point\_transform, put\_text и други.
* **Pytesseract –** Енджин, използван при разпознаването на текст от снимка, предоставящ различните модове/опции още наречени Page Segmentation Modes и OCR Engine Mode или PSM и OEM. Някои от използваните модове в проекта са:

**PSM**

1. **Automatic page segmentation with OSD.**
2. **Fully automatic page segmentation (No OSD)**

**OEM**

1. **Neural nets LSTM engine only.**
2. **Legacy + LSTM engines.**
3. **Default**

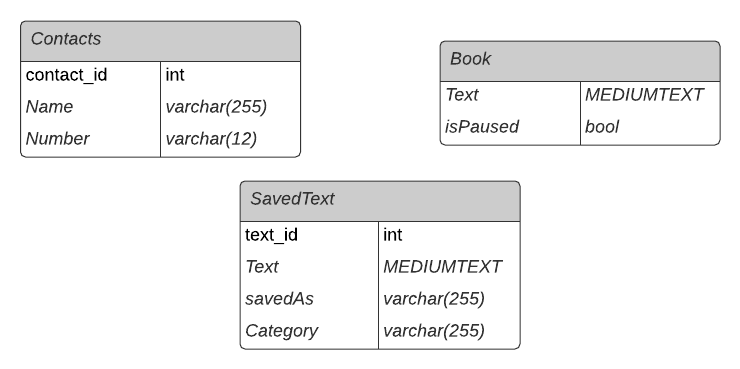
* **Subprocess –** Библиотека предоставяща функционалността за създаване на нови процеси използвани за изпълняването на функциите на устройството изпълняващи се в реално време.
* **Threading –** Класът Thread дава възможност за създаването на нови нишки използвани за изпълнението на функционалности като SmartVoiceAssistant, HandGestureRecognition.
* **Pyzbar –** Декодира засеченият QR или Баркод.
* **TextBlob –** TextBlob API позволява поправка на грешно прочетен текст чрез NLP (Natural Language Processing), който дава възможността детайлен анализ на върнатият текст. Някои от възможностите на библиотеката са: Sentiment analysis, Classification, Parsing, Spelling correction и други.

Проект №218, Съвременни системи за визуална информация 

* **OpenCV –** Предоставя множество фундаментални функции използвани при обработката на снимки и заснемане в реално време.

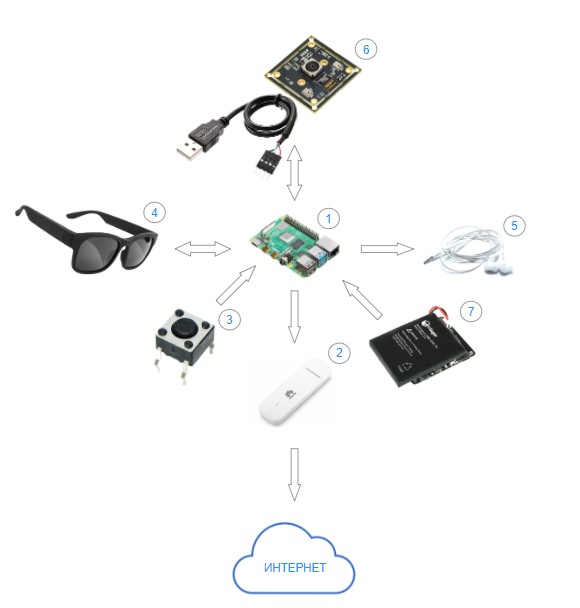
* **MySql.Connector –** Предоставя достъп до локалната MySQL база данни, което позволява изпълняването на заявки за четене и промяна на данни.
* **SpeechRecognition –** Дава възможността за Speech-to-Text.
* **Други библиотеки –** Numpy, argparse, random, time, datetime, sys, io, signal, pyttsx3, pygame, skigame, pynput, webbrowser, pyautogui, wikipedia, re.

Структура на MySQL базата данни:



Проект №218, Съвременни системи за визуална информация 

**4.4.3. Хардуерна архитектура и компоненти**



1. Компютър Raspberry Pi 4 Model B

2. Huawei 4G Dongle

3. 3x Бутона

4. Bluetooth очила с вградени говорители

5. Слушалки (по избор)

6. Arducam 4K 16MP камера

7. ​Pisugar2 | Литиева Батерия | 5000mAh

Проект №218, Съвременни системи за визуална информация 

**4.4.4. Дизайн на системата**





Проект №218, Съвременни системи за визуална информация 

**4.4.5. Външен вид на готовият продукт**



**4.5. Реализация:**

**4.5.1.** Избор на технологии

За разработката на проекта са проучени в детайли и избрани набор от модерни технологии и инструменти.

Технологиите включват програмен език Python, MySQL база данни, Pytesseract енджин за извличане на текст от снимка, OpenCV обработка на снимки и други.

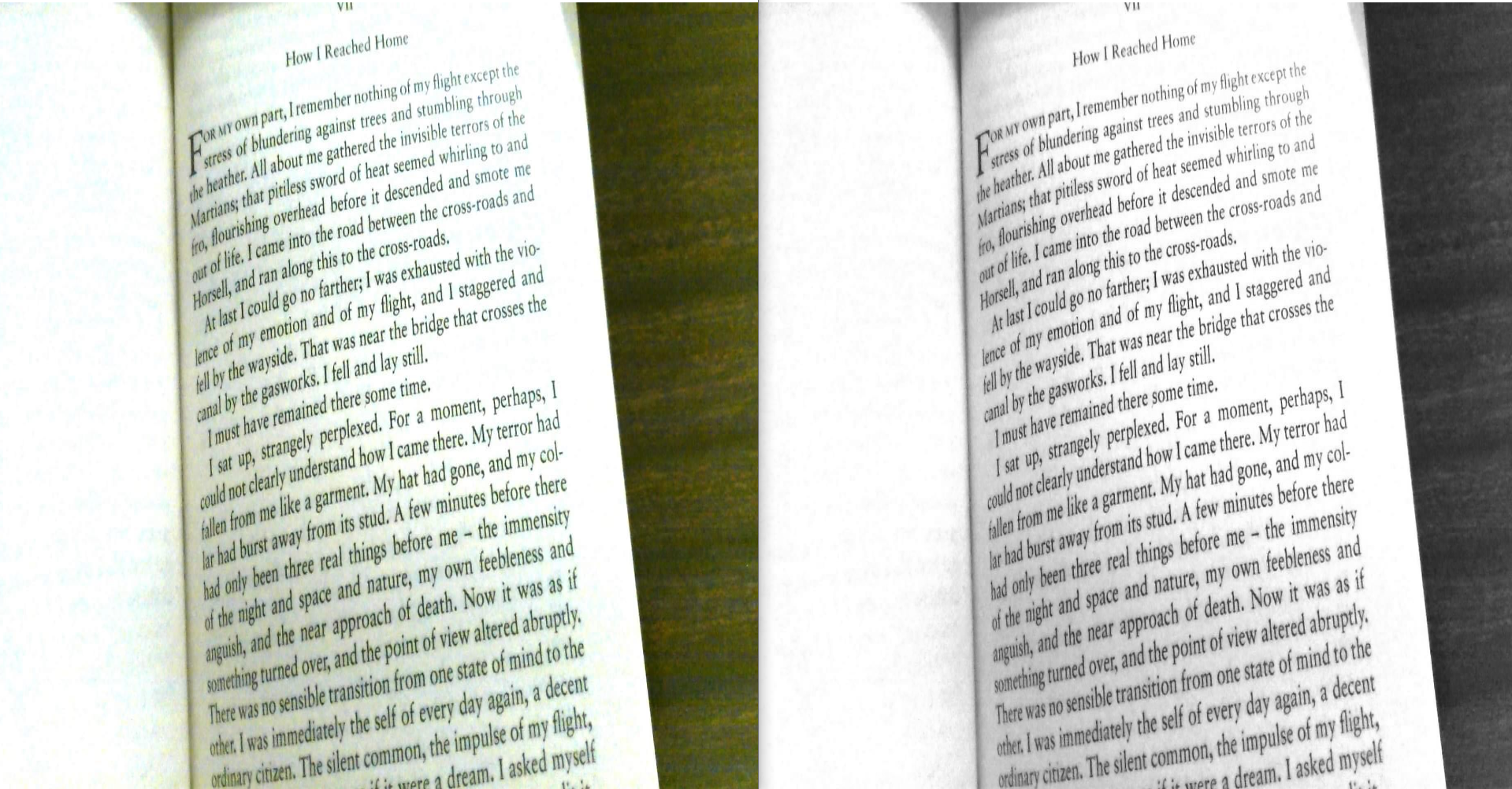
Проект №218, Съвременни системи за визуална информация 

**4.5.2.** Описание насофтуерните решения

При разработката на различните функционалности са използвани различен набор от стъпки и инструменти при обработката за постигане на оптимални резултати. Използвани алгоритми и техники при разработването на модовете са следните:

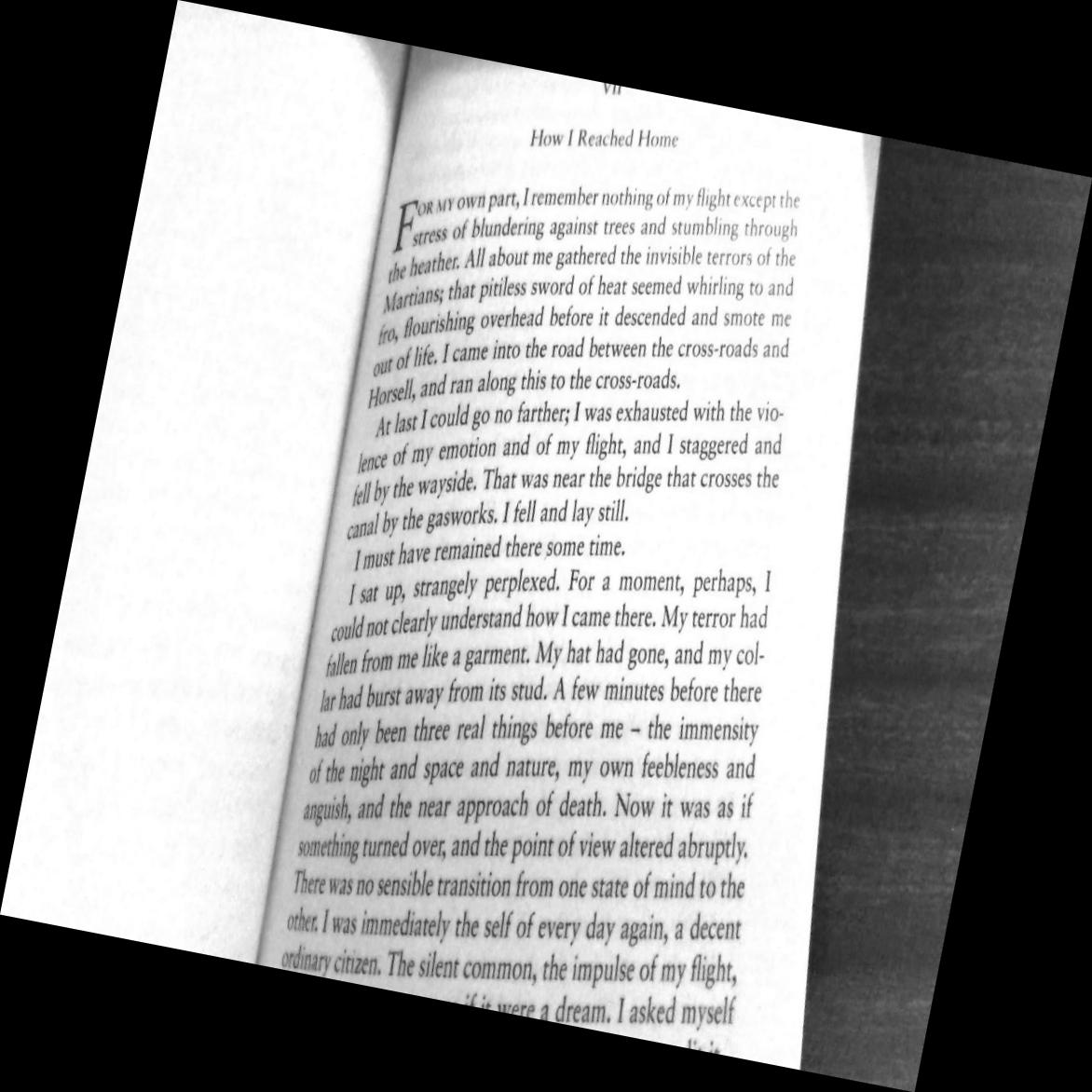
1. Алгоритъм при обработката на снимки съдържащи печатен текст

* Grayscale



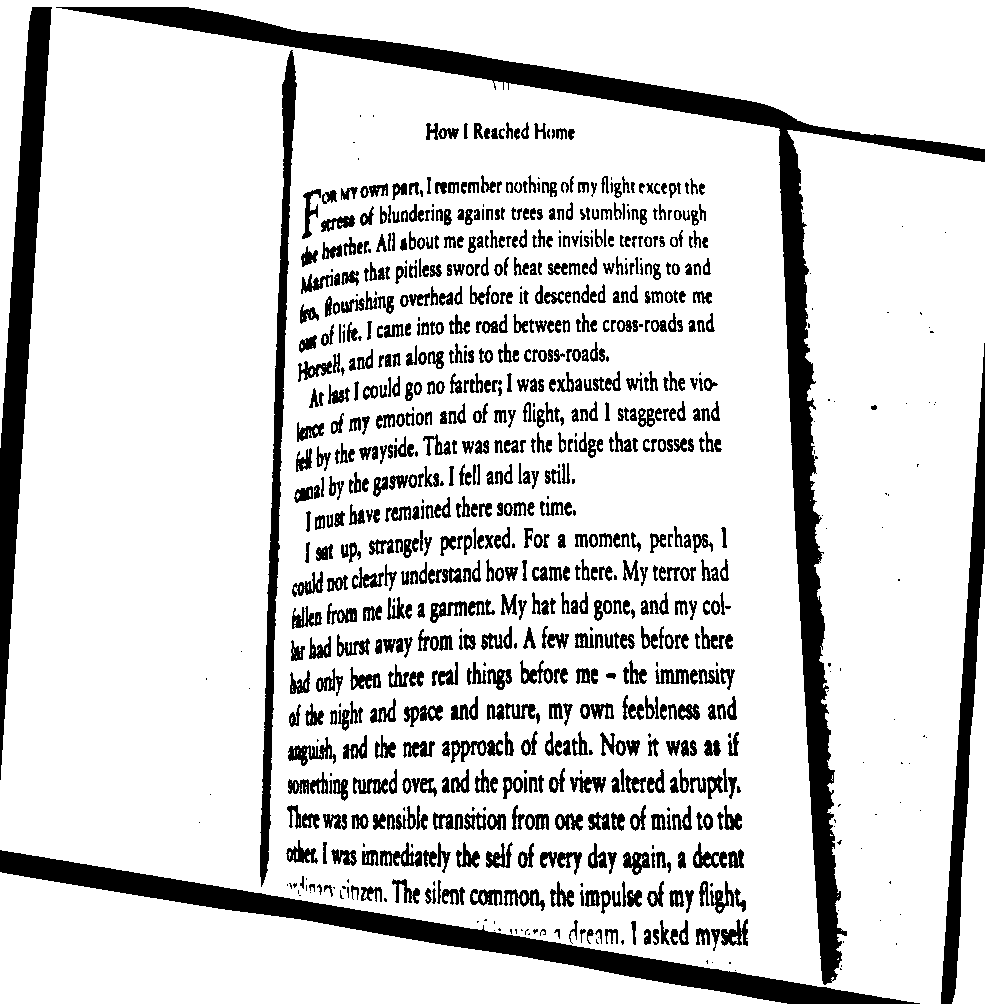
Проект №218, Съвременни системи за визуална информация 

* Deskew - Изправяне на снимката спрямо ориентацията на текста.

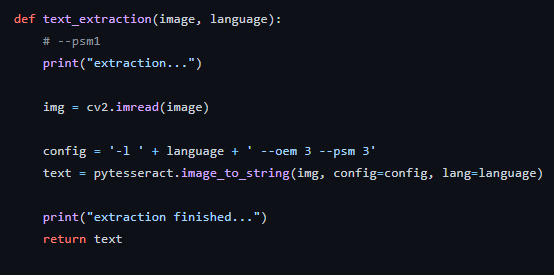


* OSD - Проверка на ориентацията, идентифициране на езика, при допусната грешка в ориентацията на текста, снимката се преработва отново.
* Dilation - Морфологична операция, увеличаваща белият регион от снимката.
* Noise Reduction - Намаляване на шума от снимката, чрез използването на median blur.
* Adaptive Thresholding - Поради способността да задава автоматично thresholding стойности за всеки пиксел се намалява загубата на информация от снимката до минимум.

Проект №218, Съвременни системи за визуална информация 

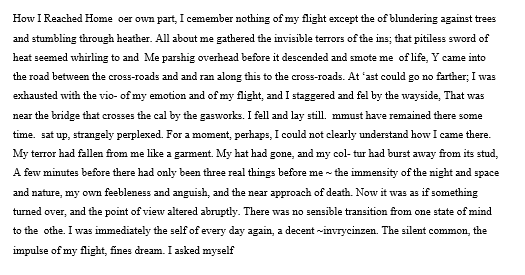


* Извличане на текста с Pytesseract



Проект №218, Съвременни системи за визуална информация 

**Резултат:**



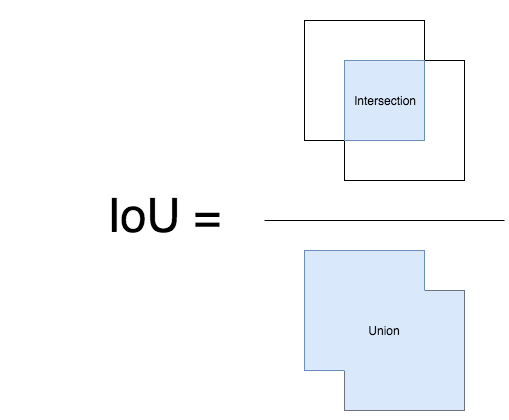
1. Алгоритъм при обработката на снимки съдържащи текст в необичайна среда

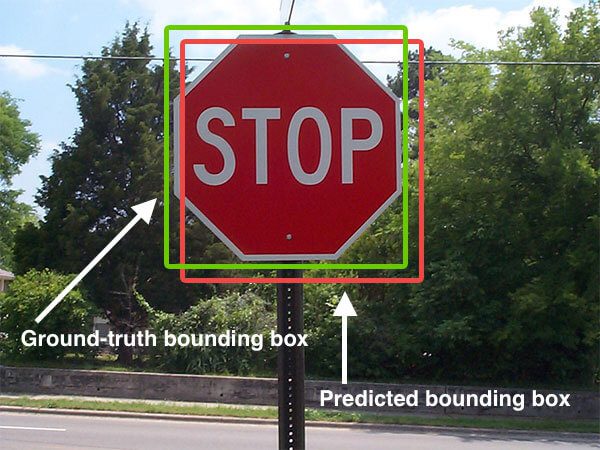
За обработката на снимки в необичайна среда се използва невронна мрежа тип FCN – Fully Convolutional Network.

* При обработката, на снимката се засичат Bounding Boxes на отдените парчета текст чрез FCN моделът на невронната мрежа.
* След засичането на всички налични Bounding Boxes, Non-Maximum Suppression техниката премахва онези Bounding Boxes, които са с коефициент на точност – confidence, по-нисък от зададеният. В нашият случай, коефициентът на точност е 0.6, при което се елиминират повечето засечени Bounding Boxes на обектите.

Проект №218, Съвременни системи за визуална информация 

* След филтрацията на кутиите чрез Non-Maximum Suppression, чрез IoU – Intersection Over Union техниката, намираме сечението на останалите Bounding Boxes.
* Извличаме текста от всяка резултантна кутия чрез Pytesseract.





Проект №218, Съвременни системи за визуална информация 

1. Алгоритъм при обработката на снимки съдържащи обекти

При обработката се използва система за засичане на обекти наречена YOLO – You Look Only Once. Използваната невронна мрежа от YOLO е от тип CNN – Convolutional Neural Network. Предимството на YOLO е в бързината и точността, като за разлика от други невронни мрежи като R-CNN – Region-based Convolutional Neural Network и FAST R-CNN, които се нуждаят от множество повторения на стъпките, YOLO извършва всички стъпки само веднъж. Стъпките през които YOLO минава за разпознаване на обектите от дадена снимка са следните:

Термини:

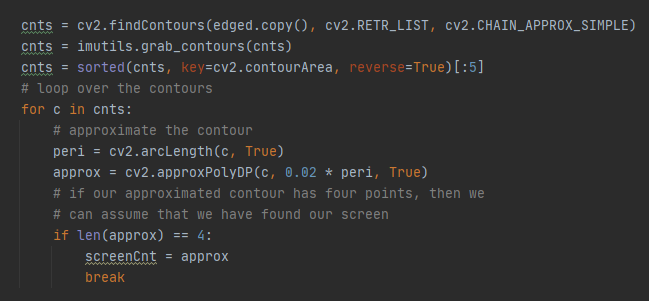
Classification – Спрямо вида на обекта се задава клас, който отговаря на характеристиките на дадения клас.

Localization – Намиране на местоположението на обект и създаване на Bounding Box около него.

* Локализация на обектите и задаване на Bounding Boxes около обектите, коефициент на точност и класификация за всеки обект.
* Използване на техниката Non-Maximum Suppression.
* Използване на техниката Intersection over Union.
* Извличане на данните за обектите от всяка резултантна кутия.

1. Алгоритъм при сканирането на касови бележки

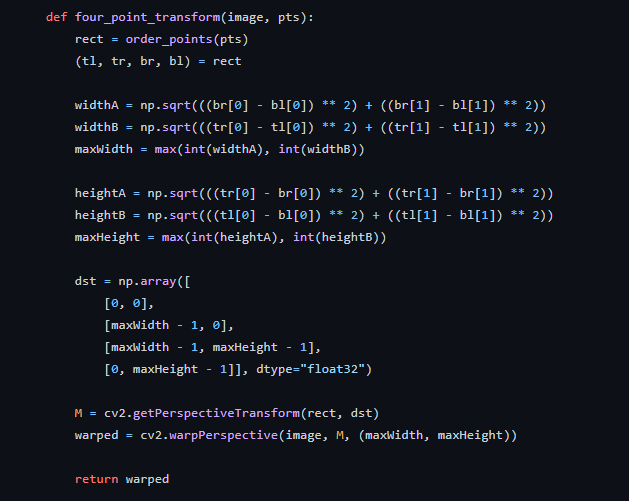
* Edge Detection - Намиране на върховете на касовият бон.



Проект №218, Съвременни системи за визуална информация 



* Преобразуване на засечения регион в 2D картина, чрез което енджинът за извличане на текст се справя оптимално при извличането на текст от касови бележки.



Проект №218, Съвременни системи за визуална информация 

**Резултат от обработката:**



За обработката на снимки при разпознаването на валута / банкноти и тяхната стойност се обработват данните по идентичен начин както при сканирането на касови бележки.

Проект №218, Съвременни системи за визуална информация 

**4.6. Описание на приложението:**

Менюто на устройството е проектирано по лесен за навигиране и използване начин от страна на потребителя. Преди използване се поставят очилата, като устройството седи в джоба/ръката на незрящия човек. 3D структурата и тежестта на устройството го прави удобно при използване за дълъг период от време.

* Устройството се стартира чрез Switch от дясната страна на кутията.
* Избираеми опции и активация: При натискането на бутона за активиране на избрана функционалност, компютъра подава сигнал към камерата и прави снимка или съответно пуска Realtime Video Stream от който да вземе всеки фрейм по отделно. Като след обработката, от говорителите на очилата се изговаря съответното съобщение.
* Функции работещи в реално време: По всяко време след стартирането, независимо от избраната чрез бутоните опция, гласовият асистент изчаква команда за активация. При изричането на дефинираната команда, гласовият асистент позволява притока на входни данни. Пример за това са известните гласови асистенти като Siri, Alexa и други. Някои от функциите на гласовият асистент са активиране на функционалност, търсене в интернет, пускане на музика, спиране и пускане на звуков файл, запазване в базата данни прочетен текст от книга. Разпознаването на жестове с ръце позволява функции като спиране, пускане на звуков файл, намаляване на звука от говорителите и други.

Практически примери на някои от функционалностите (стар модел на очилата с работещи главни функционалности за четене на текст и разпознаване на обекти) и интервю с Инж. Павлин Ангелов, член на Управителният Съвет на Съюзът на Слепите Хора в България са достъпни на следния адрес:

<https://sparkvision.tech>

Проект №218, Съвременни системи за визуална информация 

**4.7. Заключение:**

NewEyes е устройство, което дава на незрящите хора онази самостоятелност и независимост, която не са имали преди. Проектът има за цел да подобри живота на незрящите хора, като бъде неизменен помощник в тяхното ежедневие. Уникалното на NewEyes е това, че е един от първите по рода си проект, които улесняват живота на незрящите хора и хората с тежко увредено зрение, като им помагат с ежедневни дейности. Основните проблеми на проекта бяха успешно разрешени. За в бъдеще се очаква разрастване, подобряване и добавяне на нови функционалности, като голяма част от тях идват от обратната връзка и контакта с незрящите хора от Съюза на Слепите Хора в България използващи устройството ежедневно.

Основните предимства на NewEyes са бързодействието и точността, затова ще се обърне внимание на следните характеристики на устройството:

* Хардуерна оптимизация
* Бързодействие при изпълнение
* Точност при изпълнение
* Подобряване на наличните функционалности
* Добавяне на нови функционалности
* Дистанционни софтуерни ъпдейти
* Други