РАСПОЗНАВАНИЕ ТЕКСТА В ДОПОЛНЕННОЙ РЕАЛЬНОСТИ

Зейналлы Теймур Эйюб оглы

студент Московский политехнический университет, РФ, г. Москва

Полубояринова Анастасия Сергеевна

студент Московский политехнический университет, РФ, г. Москва

Text recognition in augmented reality

Zeynalli Teymur Eyyub oglu

student Moscow Polytechnic University, Russia, Moscow

Anastasia Poluboyarinova

student Moscow Polytechnic University, Russia, Moscow

Аннотация. В данной статье рассматриваются способы распознавания текста в дополненной реальности средствами EmguCV и Vuforia для интеграции различного рода контента в художественную литературу. Преимущества и недостатки этих методов.

Abstract. This article explores the ways of recognizing text in augmented reality using EmguCV and Vuforia for integrating various kinds of content into fiction. Advantages and disadvantages of these methods.

Ключевые слова: OpenSV, EmguCV; дополненная реальность; программирование; распознавание образов.

Keywords: OpenSV, EmguCV; augmented reality; programming; pattern recognition.

В дополненной реальности в большинстве случаев приходится работать с таргетами (целями), за исключением тех случаев, когда реальность дополняется образами, позиция которых вычисляется, основываясь на GPS или относительно основной камеры. В данной статье будет описываться первый вариант. В таком случае важно распознавать образы, образы могут быть самыми различными, т.к. таргетом может быть что угодно: картинка, 3D объект (цилиндр, куб

– объект любой сложности), QR-код и даже текст. Проблема рассмотренная в данной статье – это распознавания текста в художественной литературе, для вставки в неё каких-либо моментов экранизации и прочего контента. Эта проблема решается в данный момент таким образом: на странице что-то берётся за цель - картинка, иллюстрация на странице, в том числе и сама станица, или QR код. Подобные картинки сканируются очень быстро, и дополненная реальность создаётся около них без проблем. Если за таргет брать не картинку или QR код, а сам текст, не обязательно весь, а какой-нибудь отдельный фрагмент, редкое словосочетание или предложение, что отличало бы эту страницу от всех остальных, и создавать около него дополненную реальность. Этот метод был бы универсальным для всех книг т.к. он работал бы с любым шрифтом, начертанием букв и кеглем.

Существует множество видов распознавания текстов, например распознавание каждой буквы по её картинке (т.е. сравнение образов), распознавание при помощи нейронных сетей и т.д. Но самый практикуемый и часто используемый способ – это оптическое распознавание (OCR – optical character recognition). Программ, которые используют именно этот метод распознавания (не в дополненной реальности) – очень много, например: ABBYY FineReader, OCR CuneiForm, Freemore OCR и т.д. В данной статье будут описаны алгоритмы распознавания Tesseract OCR.

Tesseract OCR – это open-source приложение, которое сравнивает каждый символ на картинке с эталонным символом. Эталонный символ является набором логических правил. Если символ на картинке похож на эталонный, то программа распознает его как эталонный.

Для начала нужно провести фильтрацию текста, убрать различные помехи медианным и монохромным фильтрами. Далее следует разбить текст на сегменты, вычислить положение каждой буквы в тексте, а потом уже приступать к распознаванию каждой буквы по отдельности. Если текст слишком мелкий, то его можно увеличить в несколько раз, но не настолько, чтобы он стал слишком размытым.

Как уже ранее было описано, символ сравнивается с эталонным. Эталонные символы хранятся в отдельном файле с расширением .traineddata. Просто так открыть его в текстовом редакторе не получится, этот файл на подобии файлов с расширениями .ttf или .fnt (файлы шрифтов). Для каждого языка есть отдельный такой файл типа eng.traineddata или rus.traineddata. Они считываются программой перед началом распознавания. Там могут храниться шаблоны букв для самых различных шрифтов, что позволит распознавать любые шрифты вплоть до рукописных.

Ниже приведен пример простейшей программы распознавания текста, написанной на C#, используя библиотеку EmguCV (OpenCV для C#).

```
//Создание экземпляра класса Tesseract

//Здесь и указывается путь на файл .traineddata

Tesseract optical_character_recognizer = new Tesseract("", "rus", OcrEngineMode.Default);

//Загрузка картинки из файла

Image<Bgr, Byte> image = new Image<Bgr, byte>("D:/Desktop/2.png");

//Увеличение картинки в 2 раза

image = image.Resize(2, Emgu.CV.CvEnum.Inter.Cubic);

//Конвертирование картинки в градации серого

Image<Gray, byte> gray = image.Convert<Gray, Byte>();

//Отправка конвертированной картинки на распознавание
```

```
optical_character_recognizer.Recognize(gray);

//Получение массива букв

Tesseract.Character[] ch = optical_character_recognizer.GetCharacters();

//Обводка каждой буквы красным прямоугольником

foreach (Tesseract.Character c in characters)

 image.Draw(c.Region, new Bgr(Color.Red), 1);

//Вывод картинки с обведёнными буквами

pictureBox1.Image = image.ToBitmap();

//Получение всего текста

String recognized_text = optical_character_recognizer.GetText();

//Вывод текста

System.Diagnostics.Debug.Print("Результат " + recognized_text);
```

Проблем распознавания текстов много: шумы печати, "слипание" соседних символов, пятна и ложные точки вблизи символов, смещение символов вверх или вниз, изменение наклона символов и т.п. Этот метод используется и в дополненной реальности, а точнее распознавание текста в дополненной реальности базируется на этом методе, только с большими доработками.

Распознавание текстов в дополненной реальности так же практикуется. Существуют приложения для перевода (Google Translator), для решения математических задач (PhotoMath) и т.д.

Для работы с дополненной реальностью существует специальная платформа Vuforia. Это стандартная платформа для работы с дополненной реальностью, не имеющая аналогов. Эта платформа позволяет распознавать различного типа объекты (картинки, объекты кубической и цилиндрической формы, 3D объекты) в том числе и текст. После того, как объект распознан, на него накладываются различного типа другие объекты, дополняющие текущие.

Vuforia позволяет распознавать английский текст: отдельные слова или целые предложения. Программист задаёт слово, которое следует распознать и объект которым нужно дополнить реальность. Остальное сделает Vuforia. Так же можно и самому получить весь список распознанных слов и дальше с ними работать. Рассмотрим пример использования Vuforia.

Из папки Perfabs выбрать AR Camera, TextRecognition, Word и расположить их, как показано на рисунке. Рекомендуется так же скачать и интегрировать в проект VuforiaSamples, т.к. там находятся различного вида скрипты для корректной работы TextRecognition.

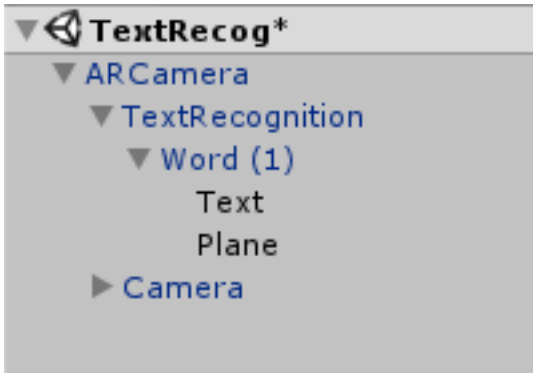

Рисунок 1. Размещение объектов в проекте

Искомое нами слово задаётся в TextRecognition и в Word. Plane – это объект которым дополняется данная книга, в данном примере объектом является флаг Англии, но он может быть чем угодно, например: другим словом, замещающим это, или 3D объектом, видео или аудиозаписью, каким-либо эффектом и т.д. Здесь координаты объектов важны только у объекта Plane. Координаты должны выбираться так, как мы хотим, чтобы реальность дополнялась относительно Word, т.к. Word – эта та самая плоскость, на которой будет находиться заданное слово или предложение, и относительно которой будет создаваться дополненная реальность. В данном примере флаг накладывается перпендикулярно и чуть выше Word. Ключевым словом в данном примере будет «How long». При наведении на ключевое слово, слово должно будет распознаться. Результат представлен на рисунке 3.

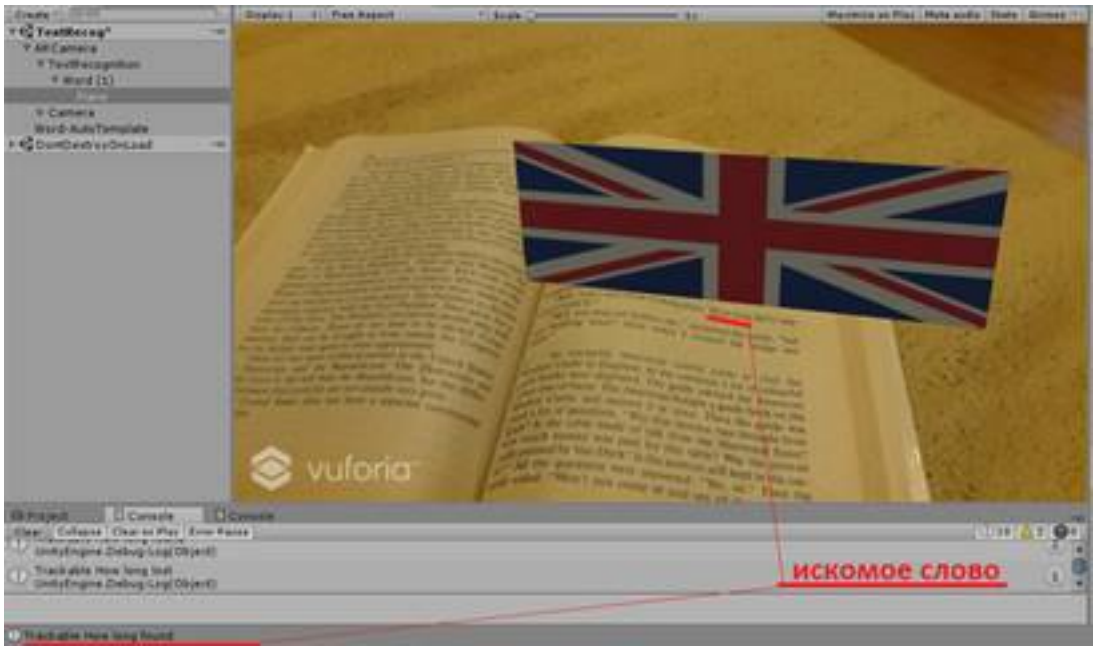

Рисунок 2. Результат распознавания

Данный пример был написан без единой строчки кода. Только нужно расположить правильно объекты.

Недостатков у данного метода много:

· скорость распознания зависит от количества слов, которых видит камера; - если выбирать некоторое множество ключевых слов (несколько объектов Word) то объект будет

накладываться на каждое это слово;

· в разных изданиях ключевые слова, которые должны быть на одной странице, могут быть на разных (в связи с разными шрифтами, например, или размерами страниц издания) тогда накладываться объект будет некорректно или вовсе ничего не распознается;

· если брать за таргеты отдельные предложения, то Vuforia не предусматривает переносы с одной строки на другую;

· Vuforia поддерживает только английский язык.

Список литературы:

1. Арсентьев Д.А. Выбор моделей для учебно-методического издания с использованием элементов дополненной реальности. – М.: Университетская книга: традиции и совре-менность: материалы научно-практической конференции. – 2015. – С. 14-17.

2. Ибрагимов В.В., Арсентьев Д.А. Алгоритмы и методы распознавания личности в усло-виях современных информационных технологий. М.: Вестник МГУМ имени Ивана Фе-дорова. – 2015. – № 1. – С. 67-69.

3. Ray Smith. An Overview of the Tesseract OCR Engine. URL: https://static.googleusercontent.com/media/research.google.com/ru//pubs/archive/33418.pdf (дата обращения 10.09.2018).

4. Методы распознавания текста. Разработка веб-сайтов, Программирование [Электронный ресурс]. – Режим доступа: https://habrahabr.ru/post/220077/ (дата обращения 12.09.2018).