

# ОБНАРУЖЕНИЕ ТЕКСТОВЫХ РЕГИОНОВ НА ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ МОДИФИЦИРОВАННОГО ДЕТЕКТОРА FASTEXT

Хрулев П.А.<sup>1</sup>, Логвин В.И.<sup>2</sup> Email: Khrulev17101@scientifictext.ru

<sup>1</sup>Хрулев Павел Алексеевич – магистрант,  
кафедра вычислительной техники;

<sup>2</sup>Логвин Василий Игоревич – магистрант,  
кафедра систем автоматического управления и контроля,  
Национальный исследовательский университет Московский институт электронной техники,  
г. Москва

**Аннотация:** предложен метод обнаружения текстовых регионов на изображении с использованием модифицированного детектора FASText. В основе алгоритма лежит сравнение интенсивности рядом лежащих пикселей между собой для определения ширины штриха и поиска специфичных ключевых точек. Для улучшения характеристик работы детектора был использован алгоритм неточной локализации текстовых регионов и повышения их контрастности. Работа алгоритма проверена на стандартном наборе изображений, представленном на конференции ICDAR 2013.

**Ключевые слова:** компьютерное зрение, обработка изображений, обнаружение текста, текстовый регион.

## TEXT REGION LOCALIZATION IN IMAGES USING MODIFIED FASTEXT DETECTOR

Khrulev P.A.<sup>1</sup>, Logvin V.I.<sup>2</sup>

<sup>1</sup>Khrulev Pavel Alekseevich – undergraduate,  
COMPUTER ENGINEERING DEPARTMENT;

<sup>2</sup>Logvin Vasily Igorevich – undergraduate,  
AUTOMATIC CONTROL SYSTEM DEPARTMENT,  
NATIONAL RESEARCH UNIVERSITY OF ELECTRONIC TECHNOLOGY,  
MOSCOW

**Abstract:** proposed text region localization detector – modified FASText. Detector based on pixel intensity comparison for searching stroke-specific keypoints. The stroke-specific keypoints produce 2 times less region segmentations and still detects 25% more characters than the commonly exploited MSER detector and the process is 4 times faster. Algorithm of increasing the contrast of text was used to improve the performance of the original FASText. Algorithm performance was tested on ICDAR 2013 images dataset.

**Keywords:** computer vision, image processing, text localization, text region.

УДК 004.932.75'1

### Введение

С распространением Интернета и различных мобильных устройств произошел огромный рост генерации разнообразного фото- и видеоконтента, содержащего различные сцены из жизни людей. Этот контент содержит большое количество информации, которую можно обнаружить и распознать для дальнейшего использования. В качестве такой информации могут выступать различные объекты, действия, текст. Распознавание этой информации позволяет описать сцены из жизни на языке, понятном не только человеку, но и компьютеру. Для обнаружения и распознавания объектов используются методы компьютерного зрения.

При распознавании объектов из фото- и видеоконтента одной из самых интересных и актуальных задач является задача обнаружения и распознавания текста. Текст является очень важным источником информации. Методы его обнаружения и распознавания применяются в большом числе прикладных приложений, таких как поиск определенной информации по фото- и видеофрагментам, навигация автоматических устройств, перевод текста в реальном времени, классификация фото- и видеоконтента.

Определение местонахождения и распознавание текста в сценах окружающего нас мира является сложной задачей. Основные трудности можно разделить на три категории:

1. Разнообразие в представлении текста. В сравнении с обработкой документов, когда все слова имеют один цвет, шрифт и расположение, а также лежат на одном белом фоне, обработка текста из различных сцен жизни несет следующие трудности: слова могут быть расположены в разных местах, могут иметь разный цвет, направление и масштаб.

2. Сложность фонового изображения. Текст может находиться где угодно. Такие элементы изображения, как знаки, различные декорации, узоры, могут быть неотличимы от текста, вследствие чего возникает много ошибок при определении реального местоположения текста.

3. Влияние помех. На распознавание текста также могут отрицательно влиять многочисленные внешние факторы, такие как шум, размытие, плохое разрешение и так далее.

Для решения задачи определения местоположения текста используется большое число различных алгоритмов, основанных на использовании локальных особенностей изображений. Одним из таких алгоритмов является алгоритм FASText [2].

#### **Алгоритм FASText**

FASText представляет собой алгоритм, основанный на использовании локальных особенностей изображения. В ходе его работы происходит поиск ключевых точек, которые затем объединяются в текстовые регионы. Он работает значительно быстрее, чем схожие с ним и широко известные алгоритмы SWT [3] и MSER [6]. Также в результате его работы происходит гораздо меньше ложных срабатываний. Основы работы алгоритма описаны далее.

Основная идея работы алгоритмов SWT, MSER, FASText заключается в том, что буквы и слова на изображении, как правило, имеют постоянную толщину штриха [1]. Для работы со штрихом в алгоритме FASText вводится понятие ключевой точки. Каждая ключевая точка представляет собой один пиксель изображения. Ключевая точка определяет, является ли пиксель частью текста. В алгоритме FASText определены два класса ключевых точек:

1. Точка окончания штриха.
2. Точка изгиба штриха.

Для поиска ключевых точек происходит сравнение интенсивности рассматриваемого пикселя с интенсивностями пикселей, расположенных внутри круга, обрамляющего рассматриваемый пиксель. Диаметр круга является параметром данного алгоритма. Обычно используется круг диаметром 12 пикселей. Определение принадлежности пикселя к одному из классов ключевых точек происходит с использованием эвристик, описанных в работе [2]. Каждый пиксель вокруг рассматриваемого помечается определенным символом в зависимости от того, насколько различна его интенсивность в сравнении с рассматриваемым. Существует три вида символов:

1. Темнее.
2. Ярче.
3. Подобен.

При этом используется параметр алгоритма «смещение», определяющий компромисс между неточностью в работе алгоритма и числом пропущенных текстовых символов.

После определения ключевых точек происходит их сегментация для формирования символов и отделения их от фона. При этом происходит сравнение интенсивности ключевых точек с заданным пределом. Пределом является значение наибольшей или наименьшей интенсивности внутри множества пикселей вокруг рассматриваемого. После этого для формирования символа используется алгоритм заливки. Результатом работы детектора ключевых точек являются найденные текстовые регионы, представляющие собой символы.

Для того, чтобы уменьшить число ложных срабатываний детектора ключевых точек производится классификация регионов. Для этого используется классификатор Gentle AdaBoost. Для его работы определены 4 свойства регионов точек, которые рассчитываются в ходе процесса сегментации.

В результате проведения классификации получены верифицированные текстовые регионы, содержащие символы. Из этих символов на данный момент работы алгоритма не составлены слова. Поэтому необходимо провести соединение текстовых регионов для формирования последовательностей из символов. Для этого используется алгоритм поиска соседей на основе сравнения центроид текстовых регионов.

На завершающем этапе работы алгоритма происходит выделение сформированных строк из полученных текстовых регионов. То есть непосредственное выделение прямоугольников, содержащих текст.

Математическое описание работы алгоритма представлено в [2].

#### **Модификация алгоритма FASText**

В ходе исследования работы алгоритма было обнаружено, что чаще всего не локализованные текстовые регионы находятся в местах слабой контрастности текста. Поэтому было предложено на стадии поиска ключевых точек использовать какой-либо алгоритм для повышения контрастности текста на изображении.

Чтобы не увеличивать контрастность всего изображения необходимо примерно определить текстовые регионы простым, быстрым и, относительно, неточным способом. Для этого был использован метод поиска текстовых регионов с помощью детектора границ Кенни. Данный детектор производит поиск

горизонтальных и вертикальных границ текстовых регионов. Далее используется морфологическая операция – открытие, определяющая границы региона.

После нахождения регионов происходит увеличение контрастности пикселей, находящихся в них. Для этого используются фильтр высоких частот и размытие по Гауссу.

Данные изменения в работе алгоритма немного замедлили его работу, однако улучшили результаты его работы на используемом наборе изображений.

#### Результаты

Работа модифицированного алгоритма FASText проверялась на стандартном наборе изображений, представленном на конференции ICDAR 2013. Для анализа правильности работы алгоритма этот стандартный набор содержит эталонные текстовые регионы для каждого из изображений.

Для сравнения работы алгоритмов для локализации текста используются следующие метрики [1]:

1. Точность локализации. Эта метрика отражает вероятность того, что выделенные текстовые области являются корректными.

2. Полнота локализации. Эта метрика отражает вероятность того, что корректные текстовые области были выделены.

3. F-мера. Эта совокупная метрика, обобщающая две предыдущих.

Результаты сравнения работы алгоритмов представлены в таблице 1.

Таблица 1. Результаты сравнения работы алгоритмов

Метод	Точность локализации, %	Полнота локализации, %	F-мера, %	Время работы с одним изображением, с
FASText [2]	84	69	77	<b>0.15</b>
<b>Модифицированный FASText</b>	<b>86</b>	<b>76</b>	<b>80</b>	<b>0.20</b>
TextFlow [7]	85	76	80	0.94
<b>CTPN [8]</b>	<b>93</b>	<b>76</b>	<b>88</b>	<b>0.14 GPU</b>
CCTN [4]	90	83	86	1.3 GPU
CNN [5]	93	73	82	0.5 GPU

Из таблицы сравнения можно сделать вывод, что на текущий момент самые качественные результаты демонстрируют алгоритмы, ядром которых являются сверточные нейронные сети. Для использования таких нейронных сетей требуются большие вычислительные ресурсы. Обычно для их работы используют GPU. Если алгоритм локализации текста должен использоваться на мобильных устройствах, то пока что использование нейронных сетей не представляется возможным из-за ограниченной вычислительной мощности. Поэтому, для этого предлагается использовать модифицированный алгоритм FASText. Он демонстрирует достаточно хорошие результаты для решения задач, которые могут возникнуть при необходимости локализации текста с использованием камеры мобильного устройства. При использовании данного алгоритма среднее время обработки одного изображения увеличилось на 0.05 с по сравнению с оригинальным алгоритмом FASText, однако при этом улучшились и показатели работы алгоритма.

#### Список литературы / References

1. Андрианов А. И. Локализация текста на изображениях сложных графических сцен // Современные проблемы науки и образования, 2013. № 3.
2. Busta M., Neumann L., Matas J. FASText: Efficient unconstrained scene text detector // IEEE International Conference on Computer Vision, 2015.
3. Epshtein B., Ofek E., Wexler Y. Detecting text in natural scenes with stroke width transform // Proceedings of International Conference on Computer Vision and Pattern Recognition, 2010. Pp. 2963-2970.
4. He T., Huang W., Qiao Y., Yao J. Accurate text localization in natural image with cascaded convolutional text network, 2016.
5. He T., Huang W. Text-attentional convolutional neural network for scene text detection, 2016.
6. Neumann L., Matas J. A method for text localization and recognition in real-world images // 10<sup>th</sup> Asian conference on Computer vision, 2010. Pp. 770-783.
7. Tian S., Pan Y., Huang C. Text Flow: A unified text detection system in natural scene images, 2015.
8. Tian Z., Huang W., He T., He P., Qiao Y. Detecting text in natural image with connectionist text proposal network, 2016.