Наука и Образование
МГТУ им. Н.Э. Баумана
Сетевое научное издание
ISSN 1994-0408

Наука и Образование. МГТУ им. Н.Э. Баумана. Электрон. журн. 2016. № 11. С. 99–112.

DOI: **10.7463/1116.0850126**

Представлена в редакцию: 03.10.2016
Исправлена: 17.10.2016

© МГТУ им. Н.Э. Баумана

УДК 004.932.75

Адаптация алгоритма локализации текстовых областей для видеопотока

Рудаков И. В.[1], **Ломовской И. В.**[1], **Сёмина В. А.**[1,*]

[*]valeriya.syomina@yandex.ru

[1]МГТУ им. Н.Э. Баумана, Москва, Россия

Алгоритм локализации текстовых областей, предназначенный для обработки изображений, адаптирован для видеопотока. Предложенное решение состоит из двух шагов: извлечение ключевых кадров и проведение локализации для отобранных изображений. Учтены влияние типа шрифта, языка и ориентации текста, а также сложности фона кадра на результат локализации текстовых областей. Для снижения количества ложных срабатываний алгоритма были добавлены эвристические правила. Экспериментально доказано улучшение качества обнаружения текстовых областей для видео как с простым, так и со сложным фоном, а также для различного разрешения.

Ключевые слова: видео, локализация текста, ключевые кадры, выделение границ изображения, ширина штриха, эвристики, Emgu CV

Введение

В современном мире видео играет важную роль в качестве средства массовой информации, распространяемого с помощью телевидения и сети Интернет. Ежедневно в Интернете, благодаря различным сервисам, блоггерам и обычным пользователям социальных сетей, появляются сотни и даже тысячи новых видео [1].

Для обработки данного медиа – контента часто возникает необходимость автоматически обнаруживать и извлекать текстовые области из кадров видеоролика для их дальнейшего распознавания. Следовательно, задача локализации текстовых областей является важным этапом распознавания выделенного текста [1].

Текст из видео может применяться для достижения разных целей:

1) индексации и ассоциативного поиска цифрового видео, что является важной функцией видео баз данных и мультимедийных библиотек [2, 3];
2) классификации и категоризации видеороликов, например, по жанру [3].

В настоящей работе была проведена адаптация алгоритма локализации текста для видеоряда с учетом следующих особенностей текста и изображений:

1. Различные типы шрифтов.

2. Различные языки.
3. Различные виды ориентации текстовых областей.
4. Сложность фона изображений.

1. Алгоритм локализации текстовых областей в видео

Видео представляет собой последовательность кадров, т.е. отдельных изображений. Поэтому для локализации текста на видео логично использовать те же самые алгоритмы, что используются для локализации текста на изображении. При этом стоит обратить внимание на тот факт, что большинство кадров, составляющих видео, являются схожими между собой. Поэтому необходимо осуществлять локализацию текста на изображениях, значительно отличающихся друг от друга. Для поиска подобных кадров, которые называются ключевыми, был выбран подход, основанный на анализе разницы граничной информации кадров [4, 5].

Для локализации текста был выбран гибридный алгоритм локализации [3, 6], так как он способен справляться с тексом различной ориентации, инвариантен к языку текста и предназначен для обработки изображений со сложным фоном.

Шаги модифицированного алгоритма локализации текстовых областей в видеопотоке представлены ниже.

Вход: видеоряд.

Выход: локализованные текстовые области для каждого ключевого кадра.

Начало

Шаг 1. Извлечение опорных кадров [3, 6].

Шаг 2. Для каждого опорного кадра.

Шаг 2.1. Преобразовать изображение в оттенки серого [7].

Шаг 2.2. Выделение границ изображения. На данном этапе использовался детектор границ Кэнни [8, 9].

Шаг 2.3. Построение карты ширины штриха [10].

Шаг 2.4. Выделение связных компонент [11]. В качестве критерия объединения берется вычисленная на шаге 2.3 ширина штриха.

Шаг 2.5. Фильтрация полученных на шаге 2.4 регионов.

Шаг 2.6. Объединение регионов в пары.

Шаг 2.7. Объединение пар в цепочки. Слияние проводилось при условии совпадения крайних элементов и углов направления (ориентации).

Конец

Для фильтрации регионов (шаг 2.5) и объединения оставшихся областей в пары (шаг 2.6) использовались эвристические правила, описанные в [10].

Для обоих этих этапов алгоритма были добавлены эвристики:

1) Для этапа фильтрации (шаг 2.5):

$$\frac{N_r}{N_b} \in [1.5,\ 25], \qquad (1)$$

где N_r - число пикселей в регионе,

N_b - число пикселей в рамке, ограничивающей регион.

2) Для этапа группировки в пары (шаг 2.6):

$$\frac{|N_1 - N_2|}{Min(N_1, N_2)} \leq 2.0, \quad (2)$$

где N_1 и N_2 - число пикселей в первом и второй регионе соответственно.

2. Исследование результата работы предложенного решения

Для того чтобы иметь возможность проводить исследования, на базе предложенного решения было разработано программное обеспечения на языке программирования C# [12] под платформу .Net Framework [13] 4.5.1. Для разделения видеоряда на кадры использовалась обертка библиотеки OpenCV [14] под указанную платформу – Emgu CV [15].

Для оценки качества результата работы алгоритма использовались следующие метрики:

1. Вероятность ошибки 1 рода = $\frac{NDTB}{ATB}$ [1]
2. Вероятность ошибки 2 рода = $\frac{FDTB}{TDTB+FDTB}$ [1]
3. Вероятность потери данных = $\frac{MDTB}{TDTB}$ [1]

Существуют эталонные текстовые блоки и текстовые блоки, полученные в результате работы алгоритма. Совпадением двух текстовых блоков m_p будем называть отношение площади пересечения этих блоков к площади минимальной рамки, содержащей оба текстовых региона. Наилучшее совпадение двух текстовых областей вычисляется по следующей формуле:

$$m(r, R) = \max\{m_p(r, r_0) \mid r_0 \in R\}$$

1. Точность локализации = $\frac{\sum_{r\in(TDTB\cup FDTB)} \mathrm{m(r,ATB)}}{TDTB\cup FDTB}$ [1]

   Точность отражает вероятность того, что выделенные текстовые области являются корректными.

2. Полнота локализации = $\frac{\sum_{r\in(ATB)} \mathrm{m(r,}TDTB\cup FDTB)}{ATB}$ [1]

   Полнота – вероятность того, что корректные текстовые области были выделены.

3. F1 мера = $\frac{1}{\frac{\alpha}{\text{точность}} + \frac{1-\alpha}{\text{полнота}}}$ [1, 46]

где ATB – количество эталонных текстовых блоков,

TDTB – количество частично или полностью локализованных текстовых блоков,

NDTB – количество невыделенных блоков, содержащих текст,

FDTB – число выделенных блоков, не содержащих текст,

MDTB – количество частично локализованных текстовых блоков, то есть число блоков, обнаруженных с некоторой потерей данных,

α – коэффициент сбалансированности, обычно берется равным 0.5.

В качестве тестовых данных были взяты как специальная подборка видеороликов ICDAR 2015 [17], как и собственные видео и фото – данные.

Экспериментальная проверка работоспособности алгоритма состояла из двух этапов. На первом этапе обрабатывались отдельные изображения. Это позволило, с одной стороны, протестировать алгоритм, с другой стороны, подобрать значения эвристик. На втором этапе модифицированный алгоритм использовался для обработки видео. При этом использовались значения эвристических правил, найденные на первом этапе.

3. Локализация в зависимости от типа шрифта текстовых областей

Для учета типа шрифта было подобрано значение эвристики, осуществляющей фильтрацию регионов по вариации ширины штриха, для всех рассмотренных шрифтов, для рукописных [18] были также подобраны значения двух других эвристических правил. Результаты проведенного исследования представлены в табл.1 - 4. Классификация шрифтов проводится по [19, 20].

Таблица 1. Значение эвристики по ширине штриха для некоторых типов антиквы

Тип шрифта	**Значение эвристики по ширине штриха**	**Тип шрифта**	**Значение эвристики по ширине штриха**
Антиква старого стиля	≥ 1.2	Египетская брусковая антиква	≥ 1.4
Антиква нового стиля	≥ 1.5	Геометрическая брусковая антиква	≥1
Брусковая антиква	≥ 0.8	Итальянская брусковая антиква	≥ 4

Таблица 2. Значение эвристики по ширине штриха для некоторых гротесков

Тип шрифта	**Значение эвристики по ширине штриха**	**Тип шрифта**	**Значение эвристики по ширине штриха**
Старые гротески	≥ 1	Гуманистические гротески	≥ 1
Новые гротески	≥ 0.7	Прочие гротески	≥ 0.5
Геометрические гротески	≥ 0.5	-	-

Таблица 3. Значение эвристики по ширине штриха для некоторых готических шрифтов

Тип шрифта	**Значение эвристики по ширине штриха**
Текстура	≥ 2
Фрактура	≥ 1.5

Таблица 4. Значение некоторых эвристик для несвязных рукописных шрифтов

Название параметра	Значение параметра
Эвристика по ширине штриха	≥ 1
Соотношение средней ширины штриха и диаметра региона	≥ 15
Соотношение высот пар регионов	≥ 3

Верхняя граница значения эвристики, отвечающей за вариацию ширины штриха, для всех типов шрифтов зависит от сложности фона кадра.

Таким образом, можно сделать вывод, что предложенный алгоритм справляется с локализацией текста, выполненного рассмотренными типами шрифтов. Касательно рукописных шрифтов, алгоритм способен работать только с несвязными шрифтами.

4. Локализация в зависимости от языка текстовых областей

Исследование локализации текстовых областей были проведены для самых распространенных по числу говорящих языков мира [21], а именно, китайского, испанского, английского, арабского, русского, японского, немецкого и французского.

Оценка качества локализации проводилась на десяти различных кадрах для каждого языка, содержащих текст на соответствующем языке. Результаты проведенного исследования представлены в табл. 5.

Таблица 5. Значения метрик в зависимости от языка текстовых областей

Язык	Вероятность ошибки 1 рода	Вероятность ошибки 2 рода	Вероятность пропуска данных	Точность	Полнота	F1 - мера
Китайский	0.52	0.14	0.30	0.41	0.63	0.52
Испанский	0.09	0.18	0.08	0.86	0.91	0.88
Английский	0.07	0.17	0.06	0.87	0.93	0.89
Арабский	0.78	0.21	0.31	0.38	0.51	0.47
Русский	0.06	0.14	0.09	0.87	0.92	0.89
Японский	0.43	0.16	0.29	0.45	0.62	0.54
Немецкий	0.08	0.16	0.07	0.88	0.91	0.87
Французский	0.05	0.13	0.05	0.89	0.93	0.90

По результатам проведенного исследования можно сделать вывод, что разработанный алгоритм предназначен в основном для обнаружения текста на таких языках, как английский, русский, испанский, французский и немецкий. Значительно хуже алгоритм справляется с локализацией текста на японском и китайском языках. Для работы с такими языками, как арабский, алгоритм не предназначен вовсе.

Было установлено значение эвристики, отвечающей за вариацию ширины штриха для разных языков. Для английского, русского, испанского, французского и немецкого языков значение данного параметра будет зависеть от типа шрифта, сложности фона. Для китайского и японского языков значение вышеуказанной эвристики должно быть больше 4 – 5. Верхняя граница для любого языка зависит от сложности фона кадра.

5. Локализация в зависимости от ориентации текстовых областей

Исследование локализации проводилось для основных видов ориентации текстовых областей, а именно, горизонтальной, вертикальной, наклонной и ориентации по окружности.

По результатам проведенного исследования была построена табл. 6, содержащая информацию о значениях метрик в зависимости от типа ориентации текста.

Таблица 6. Значение метрик для текстовых областей различной ориентации

Тип ориентации	Вероятность ошибки 1 рода	Вероятность ошибки 2 рода	Вероятность пропуска данных	Точность	Полнота	F1-мера
Горизонтальная	0.04	0.10	0.03	0.88	0.90	0.89
Вертикальная	0.05	0.09	0.04	0.87	0.91	0.89
Наклонная	0.07	0.09	0.06	0.85	0.89	0.87
По окружности	0.06	0.11	0.04	0.86	0.90	0.88

Из таблицы, представленной выше, видно, что алгоритм одинаково справляется со всеми рассмотренными видами ориентации текстовых областей. Следовательно, можно сделать вывод, что значение эвристики, отвечающей за вариацию ширины штриха, будет зависеть от сложности фона кадра, типа шрифта и языка локализуемого текста, но не от его ориентации.

6. Локализация текстовых областей в зависимости от сложности фона кадра

Под изображением со сложным фоном будем понимать такое изображение, на котором, помимо текстовых областей, присутствуют объекты, размеры которых много больше размеров областей, содержащих текст.

Изображением с простым фоном будем считать такое изображение, на котором, кроме текстовых областей, не присутствуют другие объекты или размеры этих объектов меньше или соизмеримы с размерами текстовых областей

На рис. 1 и 2 приведены примеры локализации текстовых областей для кадров с простым и сложным фоном соответственно.



Рис. 1. Результат локализации для кадра с простым фоном

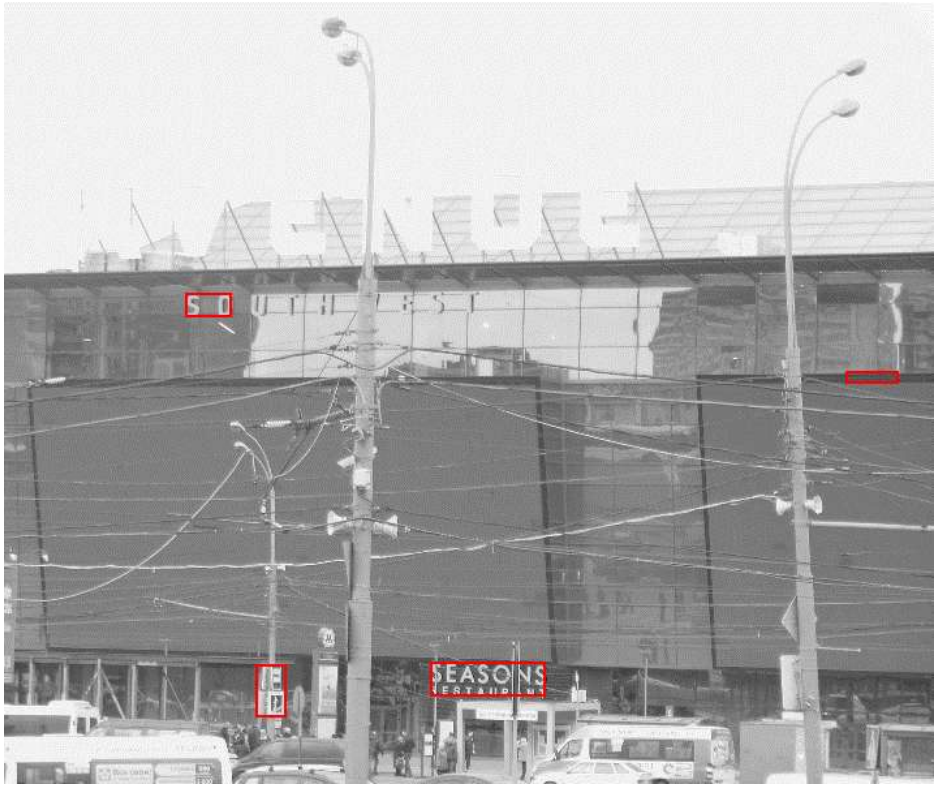


Рис. 2. Результат локализации для кадра со сложным фоном

По результатам по работы была построена табл. 7.

Таблица 7. Значения метрик для кадров со сложным и простым фоном

Тип фона кадра	Вероятность ошибки 1 рода	Вероятность ошибки 2 рода	Вероятность пропуска данных	Точность	Полнота	F1 - мера
Сложный	0.09	0.24	0.07	0.79	0.88	0.84
Простой	0.02	0.04	0.01	0.90	0.93	0.91

В результате проведенного исследования были установлены рекомендованные значения эвристики, производящей фильтрацию регионов по вариации ширины штриха.

Для изображений с простым фоном нет жестких требований для данного параметра, зависящих именно от сложности фона. Его можно взять ≥ 0.5. В данном случае оно будет в большей степени определяться шрифтом текста и его языком.

Для изображений со сложным фоном рекомендуется, чтобы данная эвристика лежала в диапазоне от 0.5 до 1. Данный разброс значений должен гарантировать оптимальную локализацию, то есть приемлемое соотношение полноты и точности. Однако стоит учитывать, что такое ограничение на значение эвристики может лишить возможности проводить локализацию для некоторых типов шрифтов и языков.

7. Локализация текста в видеопотоке

Исследование результата локализации для видео проводилось в зависимости от разрешения и сложности фона кадров с учетом добавленных эвристик (см. (1) и (2)) и без них.

В результате проведенного эксперимента были получены табл. 8 – 11.

Таблица 8. Значения метрик для видео с простым фоном без использования добавленных эвристик

Разрешение Видео	Вероятность ошибки 1 рода	Вероятность ошибки 2 рода	Вероятность пропуска	Точность	Полнота	F1 - мера
640x480	0.13	0.24	0.12	0.76	0.78	0.77
800x600	0.10	0.20	0.11	0.80	0.83	0.81
1024x600	0.09	0.16	0.09	0.82	0.86	0.84

Таблица 9. Значения метрик для видео со сложным фоном без использования добавленных эвристик

Разрешение Видео	Вероятность ошибки 1 рода	Вероятность ошибки 2 рода	Вероятность пропуска	Точность	Полнота	F1 - мера
640x480	0.25	0.74	0.23	0.25	0.58	0.35
800x600	0.22	0.38	0.19	0.61	0.69	0.64
1024x600	0.19	0.29	0.17	0.70	0.80	0.75

Таблица 10. Значения метрик для видео с простым фоном с использованием добавленных эвристик

Разрешение Видео	Вероятность ошибки 1 рода	Вероятность ошибки 2 рода	Вероятность пропуска	Точность	Полнота	F1 - мера
640x480	0.11	0.19	0.10	0.79	0.81	0.80
800x600	0.09	0.16	0.09	0.83	0.85	0.84
1024x600	0.08	0.12	0.07	0.86	0.89	0.87

Таблица 11.Значения метрик для видео со сложным фоном с использованием добавленных эвристик

Разрешение Видео	Вероятность ошибки 1 рода	Вероятность ошибки 2 рода	Вероятность пропуска	Точность	Полнота	F1 - мера
640x480	0.18	0.62	0.17	0.39	0.69	0.49
800x600	0.15	0.31	0.11	0.68	0.78	0.72
1024x600	0.13	0.20	0.10	0.79	0.86	0.82

Из полученных в ходе данного эксперимента результатов можно сделать вывод, что использование добавленных эвристик приводит к улучшению результата локализации, как для видео с простым, так и со сложным фоном. Более того, для видеоряда со сложным фоном рост качества обнаружения происходит быстрее, так как добавление новых эвристических правил позволяет отсеять большее количество областей, не являющихся текстом.

Заключение

В результате проделанной работы адаптирован для видео алгоритм, изначально предназначенный для локализации текстовых областей на изображениях. Для этапов фильтрации и группировки оставшихся регионов в пары добавлены эвристические правила, благодаря которым удалось улучшить результат обнаружения текстовых областей в видеопотоке.

Установлено, что предложенное решение справляется со всеми рассмотренными типами шрифтов и языками. Также алгоритм предназначен для локализации текста разной ориентации и способен обрабатывать как видео с простым, так и со сложным фоном.

Список литературы

1. Dutta A., Pal U., Shivakumara P., Ganduli A., Bandyopadhya A., Tan C.L. Gradient based approach for text detection in video frames. Available at: https://www.comp.nus.edu.sg/~tancl/publications/c2009/ICSIP2009-3.pdf, accessed 17.09.2016.
2. Зотин А.Г. Методы и алгоритмы обнаружения наложенных текстовых символов в системах распознавания изображений со сложной фоновой структурой. Режим доступа: http://www.dissercat.com/content/metody-i-algoritmy-obnaruzheniya-nalozhennykh-tekstovykh-simvolov-v-sistemakh-raspoznavaniya (дата обращения 17.09.2016).

3. Никитин И.К. Обзор методов комплексного ассоциативного поиска видео. Режим доступа: https://www.researchgate.net/publication/274732904_obzor_metodov_kompleksnogo_associativnogo_poiska_video (дата обращения: 17.09.2016).
4. Wolf W. Key frame selection by motion analysis // IEEE Intern. Conf. on Acoustics, Speech and Signal Processing: ICASSP' 96: Proceedings. Wash.: IEEE, 1996. Vol. 2. DOI: 10.1109/ICASSP.1996.543588
5. Khushboo Khurana, Chandak M.B. Key frame extraction methodology for video annotation. Available at: https://www.academia.edu/3255959/KEY_FRAME_EXTRACTION_METHODOLOGY_FOR_VIDEO_ANNOTATION, accessed 17.09.2016.
6. Фраленко В.П. Локализация текстовых фрагментов на смешанном фоне: краткий научный обзор. Режим доступа: http://psta.psiras.ru/read/psta2014_2_33-45.pdf (дата обращения: 17.09.2016).
7. Горьков А. О цветовых пространствах. Режим доступа: https://habrahabr.ru/post/181580/ (дата обращения: 17.09.2016).
8. Canny J. A computational approach to edge detection. Available at: https://perso.limsi.fr/vezien/PAPIERS_ACS/canny1986.pdf, accessed 17.09.2016.
9. Буй Т.Т.Ч., Спицын В.Г. Анализ методов выделения краев на цифровых изображениях. Режим доступа: http://www.tusur.ru/filearchive/reports-magazine/2010-2-2/221.pdf (дата обращения: 17.09.2016).
10. Epstein B., Ofek E., Wexler Y. Detecting text in natural scenes with stroke width transform. Available at: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/1509.pdf, accessed 17.09.2016.
11. Connected component labelling. Available at: http://aishack.in/tutorials/connected-component-labelling/, accessed 17.09.2016.
12. Троелсен Э. Язык программирования C# 5.0 и платформа .NET 4.5. 6-е изд. М.: Вильямс, 2013. 1312 с. [Troelsen A. Pro C#5.0 and the .NET 4.5 framework. 6th ed. Berk.: Acad. Press, 2012. 1312 p.]
13. .NET: Develop high performance application in less time, on any platform. Available at: https://www.microsoft.com/net/, accessed 17.09.2016.
14. Open CV: Open source computer vision. Available at: http://opencv.org/, accessed 17.09.2016.
15. Main page: Emgu CV. Available at: http://www.emgu.com/wiki/index.php/Main_Page, accessed 17.09.2016.
16. F1 score. Available at: https://en.wikipedia.org/wiki/F1_score, accessed 17.09.2016.
17. ICDAR 2015. Robust reading competition. Available at: http://rrc.cvc.uab.es/?ch=3&com=introduction, accessed 17.02.2016.

18. Рукописные шрифты. Режим доступа: http://www.fonts-online.ru/fonts/handwritten (дата обращения: 17.09.2016).
19. Наглядная классификация шрифтов. Режим доступа: http://infogra.ru/typography/naglyadnaya-klassifikatsiya-shriftov (дата обращения: 17.09.2016).
20. Классификация [шрифтов]. Режим доступа: http://www.paratype.ru/help/class/ (дата обращения: 17.09.2016).
21. Список наиболее распространенных языков. Режим доступа: https://ru.wikipedia.org/wiki/Список_наиболее_распространённых_языков (дата обращения: 17.09.2016).

Science & Education of the Bauman MSTU
Electronic journal
ISSN 1994-0408

Science and Education of the Bauman MSTU, 2016, no. 10, pp. 99–112.

DOI: **10.7463/1116.0850126**

Received: 03.10.2016
Revised: 17.10.2016

© Bauman Moscow State Technical Unversity

The Adapting of Text Detection Algorithm for Video

I.V. Rudakov[1], **I.V. Lomovskoi**[1], **V.A. Semina**[1,*]

*valeriya.syomina@yandex.ru

[1]Bauman Moscow State Technical University, Moscow, Russia

Keywords: video, text detection, key frames, image edge detection, stroke width, heuristic, Emgu CV

The paper considers adaptation of image text detection algorithm for video. The suggested solution consists of two steps: key frames extraction and chosen images text detection. Key frames are extracted by the method based on analysis of difference in edge information of frames. The hybrid algorithm was chosen for frames text detection. Such choice is due to the fact that this class of algorithms is capable of coping with the text of different orientation, is invariant to the text language and aimed at image processing with complex background.

The paper pays a special attention to two stages of the text detection algorithm, namely, filtering of regions and merging the remaining ones in pairs. Some heuristics were used for said purposes. The algorithm adaptation is to add two heuristics for these stages of algorithm in order to improve text detection quality.

The known metrics (type I and type II type error probability, data loss probability, recall, precision, F1 - score) were used to assess detection quality. A ICDAR 2015 video set was taken as test data. The experiment results found that the suggested solution enables us to cope with text detection of various types of the font, language, and orientation and also to process video frames with complex background. One more conducted experiment has proved improvement of text detection quality for video both with simple and complex background and for its different resolutions as well.

The result of this work is an adapted algorithm for text detection in video.

The result of experiments is a set of recommendations for values of some heuristics depending on the type of text font, language, and orientation and video frame background complexity.

The advanced direction of further study is to add new heuristics or modify already used ones in order to reduce false positives.

References

1. Dutta A., Pal U., Shivakumara P., Ganduli A., Bandyopadhya A., Tan C.L. Gradient based approach for text detection in video frames. Available at: https://www.comp.nus.edu.sg/~tancl/publications/c2009/ICSIP2009-3.pdf, accessed 17.09.2016.
2. Zotin A.G. *Metody i algoritmy obnaruzheniia naloshennykh tekstovykh simvolov v sistemakh raspoznavaniia izobrazhenij so sloshnoj fonovoj structuroj* [Methods and algorithms for detection of the superimposed text characters in systems of recognition of images with complicated background structure] Available at: http://www.dissercat.com/content/metody-i-algoritmy-obnaruzheniya-nalozhennykh-tekstovykh-simvolov-v-sistemakh-raspoznavaniya, accessed 17.09.2016 (in Russ.).
3. Nikitin I.K. *Obzor metodov kompleksnogo assotsiativnogo poiska video* [An overview of complex content-based video retrieval methods]. Available at: https://www.researchgate.net/publication/274732904_obzor_metodov_kompleksnogo_associativnogo_poiska_video, accessed 17.09.2016 (in Russ.).
4. Wolf W. Key frame selection by motion analysis. *IEEE Intern. Conf. on Acoustics, Speech and Signal Processing: ICASSP' 96*: Proceedings. Wash.: IEEE, 1996. Vol. 2. DOI: 10.1109/ICASSP.1996.543588
5. Khushboo Khurana, Chandak M.B. Key frame extraction methodology for video annotation. Available at: https://www.academia.edu/3255959/KEY_FRAME_EXTRACTION_METHODOLOGY_FOR_VIDEO_ANNOTATION, accessed 17.09.2016.
6. Fralenko V.P. *Lokalizatsiia tekstovikh fragmentov na smeshannom fone: kratkij nauchnyj obzor* [Localization of text fragments on mixed background]. Available at: http://psta.psiras.ru/read/psta2014_2_33-45.pdf, accessed 17.09.2016 (in Russ.).
7. Gor'kov A. *O tsvetovykh prostranstvakh* [About color spaces]. Available at: https://habrahabr.ru/post/181580/, accessed 17.09.2016 (in Russ.).
8. Canny J. A computational approach to edge detection. Available at: https://perso.limsi.fr/vezien/PAPIERS_ACS/canny1986.pdf, accessed 17.09.2016.
9. Buj T.T.Ch., Spytsin V.G. *Analiz metodov vydeleniia kraev na tzifrovykh izobrazheniiakh* [Analysis of methods for edge detection on digital images]. Available at: http://www.tusur.ru/filearchive/reports-magazine/2010-2-2/221.pdf, accessed 17.09.2016 (in Russ.).
10. Epstein B., Ofek E., Wexler Y. Detecting text in natural scenes with stroke width transform. Available at: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/1509.pdf, accessed 17.09.2016.
11. Connected component labelling. Available at: http://aishack.in/tutorials/connected-component-labelling/, accessed 17.09.2016.
12. Troelsen A. *Yazyk programmirovaniia C# 5.0 i platforma .NET 4.5* [Pro C#5.0 and the .NET 4.5 framework]. 6th ed. Moscow, Williams Publ., 2013. 1312 p. (in Russ.).
13. .NET: Develop high performance application in less time, on any platform. Available at: https://www.microsoft.com/net/, accessed 17.09.2016.
14. Open CV: Open source computer vision. Available at: http://opencv.org/ , accessed 17.09.2016.
15. Main page: Emgu CV. Available at: http://www.emgu.com/wiki/index.php/Main_Page, accessed 17.09.2016.

16. F1 score. Available at: https://en.wikipedia.org/wiki/F1_score, accessed 17.09.2016.
17. ICDAR 2015. Robust reading competition. Available at: http://rrc.cvc.uab.es/?ch=3&com=introduction, accessed 17.02.2016.
18. *Rukopisnye shrifty* [Handwritten fonts]. Available at: http://www.fonts-online.ru/fonts/handwritten, accessed 17.09.2016 (in Russ.).
19. *Nagliadnaia klassifikatsiia shriftov* [Visual classification of fonts]. Available at: http://infogra.ru/typography/naglyadnaya-klassifikatsiya-shriftov, accessed 17.09.2016 (in Russ.).
20. *Klassifikatsiia [shriftov]* [Classification of fonts]. Available at: http://www.paratype.ru/help/class/, accessed 17.09.2016 (in Russ.).
21. *Spisok naibolee rasprostranennych yazykov* [The list of most common languages]. Available at: https://ru.wikipedia.org/wiki/Список_наиболее_распространённых_языков, accessed 17.09.2016 (in Russ.).