

ОБНАРУЖЕНИЕ И ЛОКАЛИЗАЦИЯ ТЕКСТОВЫХ ОБЛАСТЕЙ НА ПОЛУТОНОВЫХ ЦИФРОВЫХ ИЗОБРАЖЕНИЯХ

А. В. Коренной, Д. С. Юдаков, С. В. Дедов, В. П. Стражник

ВУНЦ ВВС «Военно-воздушная академия им. проф. Н. Е. Жуковского и Ю. А. Гагарина»

Поступила в редакцию 20.09.2015 г.

Аннотация. В работе разработан алгоритм обнаружения и локализации текстовых областей на цифровых изображениях. Алгоритм основан на развитии метода обнаружения горизонтальных перепадов яркости. В качестве достаточной статистики используется результат построчной корреляционной обработки изображения. Приведены результаты локализации текстовых областей на изображении разработанным способом.

Ключевые слова: изображения, текстовая область, локализации текстовых областей, цифровые видеопоследовательности и изображения.

Annotation. We consider an algorithm of detection and localization of text fields in digital images. The algorithm is based on the development of the detection method of horizontal gradients of brightness. As a result of a sufficient statistic used the progressive correlation image processing. Results of localization of text fields in an image using the developed method are presented.

Keywords: The images, text fields, localizations of text fields, digital videosequences and images.

1. ВВЕДЕНИЕ

Задача распознавания текста на изображениях сложных графических сцен подразумевает локализацию и распознавание текста в ситуации, когда изображения кроме текста могут содержать большое количество другой графической информации, которая практически не поддается фильтрации [1].

Сложные графические сцены – это изображения, характеризующиеся неоднородностью фона, отсутствием чётких критериев отличия фона от текста, большой вероятностью разнообразных искажений. Примером сложных графических сцен могут являться кадры из кинофильмов с субтитрами, фотографии уличных вывесок, сделанные цифровым фотоаппаратом и т. д.

Существующие на данный момент системы распознавания текста ориентированы на бумажные документы, то есть на типовые условия, в которых однотонный текст расположен на однотонном контрастном фоне. Такие изображения могут содержать нетекстовую

информацию, например, рисунки, графики. Однако эта информация легко локализуется, отделяется от текстовых областей [1, 2].

В отличие от распознавания изображений бумажных документов, распознавание текста на сложных графических сценах осложняется тем, что текст на таких изображениях часто не отделен от прочей информации явно, а является частью этой информации. Невозможно заранее предугадать, в какой области изображения расположен текст, какое он имеет искажение. Соответственно, усложняется задача поиска, локализации текста.

Частичные решения задачи поиска текста на изображениях сложных графических сцен уже описаны в научно-технической литературе. Например, широко известны системы для распознавания государственных номерных знаков автомобилей на изображениях, полученных с видеоканалов камер слежения [1, 3]. Решение этой задачи облегчается тем, что количество форматов номерных знаков ограничено, размер и положение номерного знака относительно автомобиля фиксированы, а текст знака всегда находится на контрастном фоне.

© Коренной А. В., Юдаков Д. С., Дедов С. В., Стражник В. П., 2015

Другой пример частного решения поставленной задачи – распознавание метки даты и времени на сканированном изображении фотографии [1, 4]. Эта задача подразумевает лишь поиск цифр в определенной области изображения.

Решение задачи качественной локализации текста в сложных графических сценах в общем виде позволит расширить область практического применения систем распознавания текста.

2. ОБЗОР СОВРЕМЕННЫХ ПОДХОДОВ К ОБНАРУЖЕНИЮ ТЕКСТОВЫХ БЛОКОВ НА ЦИФРОВЫХ ИЗОБРАЖЕНИЯХ

Классически задача распознавания текста делится на три фазы:

- анализ изображения и выделение текстовых областей;
- распознавание символов;
- сборка текста и экспорт данных во внешний формат.

Основная задача представленной работы заключается в первой фазе – необходимо найти текст в сложной графической сцене.

В [1] приведен обзор методов, предлагаемых в научно-технической литературе для решения этой задачи.

Работа [5] описывает идею автоматической генерации признаков, используемых для распознавания. Основная идея здесь заключается в том, что признаки для классификатора можно не составлять вручную, а создавать их с помощью машинного обучения. В качестве базы для обучения предлагается использовать так называемые патчи, имеющие размер 8×8 пикселей. Дальнейшая обработка изображения, по мнению авторов, может быть сведена к вычислению обученных признаков на интересующих областях изображения.

Основная идея ряда работ заключается в том, что буквы и слова на изображении, как правило, имеют постоянную толщину штриха. Поэтому для выявления таких объектов, по мнению авторов, перспективно использовать алгоритм SWT. Действительно, разумно

ожидать, что ширина штриха в букве должна быть приблизительно постоянна, равно как и буквы в одном слове должны иметь похожую ширину. Признак тем более полезен, так как использовать его возможно как в качестве одного из признаков при классификации регионов, так и в качестве признака при «собирации» регионов в слова.

Границы символов в рамках описанного подхода могут вычисляться, например, с помощью Canny Edge Detector. Затем в направлении градиента находится парная граничная точка. Если градиент в этой точке примерно коллинеарен градиенту в первой точке, то все точки между ними заполняются значением найденной ширины. Затем, при втором проходе по найденным на первом проходе отрезкам вычисленное значение ширины ограничивается медианным значением вдоль отрезка. При этом отметим, что предложенный метод потребует определённых дополнительных вычислительных затрат для борьбы с ошибками на углах и некоторыми другими специфическими для него эффектами.

Работа Y. Kunishige и соавторов [6] опирается на идею использования «контекста окружения» (environmental context). Основная мысль заключается в использовании информации о том, что окружает область-«кандидата». Иными словами, предлагается анализировать тот фон, на котором находится регион изображения, возможно, являющийся текстовым. Идея базируется на эмпирическом предположении, что вероятность наличия текста, например, на травяном покрове или на небе – низка.

Авторский коллектив [7] предлагает сначала детектировать на изображении так называемые SURF-точки. Предполагается, что если на исследуемом изображении присутствуют буквы, то они будут плотно такими точками покрыты. Дополнительно к этому вычисляется визуальная заметность (visual saliency). Вместе SURF и saliency будут представлять собой $(128 + 1)$ – мерный вектор признаков. На этом векторе предполагается провести обучение классификатора AdaBoost.

В работе [8] к аналогичной задаче предложен несколько иной подход. В общих чертах,

в указанной работе решается задача поиска точечного текста, но для обнаружения точек, составляющих буквы, применяется хорошо известный алгоритм FAST. Затем производится эвристическая фильтрация ложных кандидатов, объединение точек в буквы, букв – в слова, после чего применяется классификатор SVM [9] для детектирования текстовых областей.

С одной стороны, все перечисленные работы прогнозируют хорошие результаты при поиске текстов, изначально состоящих из обособленных точек (например, такими текстами набираются цифры, указывающие срок годности на упаковке продуктов). С другой стороны, на сложных графических сценах, где вполне возможно присутствие «неточечного» текста (рекламные щиты, вывески магазинов, автомобильные номера и т. д.), предложенные авторами алгоритмы дадут не слишком хорошие результаты – именно в силу изначальной своей ориентированности на работу с точками.

3. АЛГОРИТМ ОБНАРУЖЕНИЯ И ЛОКАЛИЗАЦИИ ТЕКСТОВЫХ ОБЛАСТЕЙ НА ЦИФРОВЫХ ИЗОБРАЖЕНИЯХ

Наибольшие проблемы с выделением областей, содержащих текст, на цифровых изображениях возникают для искаженных изображений: изображений в которых текст расположен под углом; малоконтрастных изображений (цвет символов близок к цвету фона); изображений имеющих сложный фон, аппликации и т. д. [10]. Современная информационно-поисковая система должна решать задачи локализации текстовых областей при наличии любой из указанных выше проблем.

В данной статье предложен комбинированный алгоритм локализации областей, содержащих текст, на изображении. Алгоритм основан на том, что в большинстве случаев части изображений, содержащие текст, имеют ярко выраженный горизонтальный перепад яркости между буквами и фоновым рисунком, который можно подчеркнуть при помощи вторых горизонтальных простран-

ственных разностей между пикселями изображения. Обнаружение этих областей осуществляется с помощью субоптимальной корреляционной обработки.

В качестве исходного будем рассматривать цифровое дискретное изображение размером M на N точек, а элементы такого изображения будем называть пикселями.

Необходимо отметить, что алгоритм работает с полутоновыми изображениями, поэтому необходимо предварительно преобразовать цветное изображение в полутоновое, или выбрать один из цветовых каналов, в зависимости от сложности фонового рисунка, например, красный.

На первом этапе производится удаление постоянной составляющей, что позволяет сузить динамический диапазон яркости изображения. С целью более качественного выделения резких перепадов при помощи вторых пространственных разностей, а также избавиться от негативного влияния постоянной составляющей в случае ограниченного по пространственным координатам поля [11].

На следующем этапе вычисляются вторые дискретные разности по горизонтали, что аналогично непрерывному пространственному дифференцированию

$$F_p(x, y) = \frac{\partial^2 F_I(x, y)}{\partial x^2} = F_I(x-1, y) - 2F_I(x, y) + F_I(x+1, y), \quad (1)$$
$$x \in [1, N], \quad x \in [1, M],$$

где $F_I(x, y)$ – значение яркости пикселя с координатами x и y исходного изображения; $F_p(x, y)$ – значение яркости пикселя с координатами x и y преобразованного изображения.

Выполнение данной операции позволяет определить области изображения, в которых наблюдаются резкие изменения яркости. Текстовая область будет содержать большое количество скачков на ограниченном пространственном промежутке, что позволит определить их границы.

Необходимо отметить, что все дальнейшие вычисления производятся со значениями пикселей, переведенными в веществен-

ный формат чисел для исключения потери динамического диапазона сигнала.

В связи с тем, что физически значения яркости могут принимать только положительные значения, а после проведения операции (1) граница текстовой области будет описываться, как положительными, так и отрицательными скачками, перейдем к абсолютным значениям функции $F_p(x, y)$.

Переход к абсолютным значениям позволяет при дальнейшей обработке избежать потери части сигнала за счет суммирования составляющих с противоположным знаком, описывающих границы текстовой области при помощи положительных и отрицательных скачков.

На следующем этапе реализуется процедура обнаружения областей содержащих текст. Данная процедура сводится к формированию достаточной статистики изображения и сравнения ее с порогом.

В качестве достаточной статистики $U(x, y)$ используется результат построчной корреляционной обработки изображения $F_p(x, y)$ с некоторой опорной функцией $T(x)$, т. е.

$$U(x, y) = \sum_{x'=1}^n T(x-x')F_p(x', y), \quad n < N, \quad (2)$$

где N – длина строки изображения в пикселях, n – размер эталона в пикселях.

Если опорная функция $T(x)$ будет точно соответствовать изображению строки текста, достаточная статистика $U(x, y)$ будет обеспечивать максимум отношения текст/фон. Однако для реальных изображений выбрать опорную функцию, точно соответствующую искомому тексту, достаточно сложно, так как отсутствуют априорные сведения об искомом текстовом блоке (относительный размер букв, форма текста, тип шрифта, наклон букв). Поэтому обнаружение с использованием достаточной статистики (2) будет субоптимальным.

В качестве эталона лучше всего подойдет равномерный строб размером n пикселей. Применение данного эталона позволит получить максимум взаимной корреляции на участках, где после обработки по (1) возникает решетка из перепадов яркости. Таким об-

разом, выражение для эталона будет иметь вид

$$T(x) = \frac{1}{n}(1(x) - 1(x-n)), \quad (3)$$

где $1(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$ – функция Хевисайда.

Размер эталона n выбирается из следующих соображений: учитывается, что средняя длина русского слова составляет 5.28 (английского 5) символов [12]; средняя ширина символа определяется на основе априорных сведений о разрешении в кадре, о расстоянии от камеры до экспозиции, об относительных размерах шрифтов.

Для представленных в статье изображений средний размер символа составляет 10 пикселей. На основании вышеизложенного при проведении экспериментов использовался параметр $n = (5.28 + 1)10 = 63$ (здесь учтены пробелы между словами).

Для обнаружения всех текстовых областей в кадре, используется решающее правило, по которому на основании сравнения с порогом h будет принято решение о принадлежности пикселя к текстовому блоку (решение «1»), или нет (решение «0»)

$$U_s(x, y) = \begin{cases} 1, & \text{при } U(x, y) > h \\ 0, & \text{при } U(x, y) < h. \end{cases} \quad (4)$$

Таким образом, бинарное изображение $U_s(x, y)$ будет являться результатом обнаружения текстовой формы на исходном изображении.

Для определения порога h целесообразно использовать вероятностный критерий. В данном случае будем использовать критерий Неймана-Пирсона.

В качестве фона будем рассматривать весь кадр изображения, принимая во внимание, что текст занимает относительно малую площадь и на статистические характеристики распределения яркости во всем кадре значения не оказывает.

Для определения закона распределения значений достаточной статистики $U(x, y)$ построим ее гистограмму $P(U)$ и проанализируем ее.

На рис. 1 а) и б) представлены исходные тестовые изображения, распределения значе-

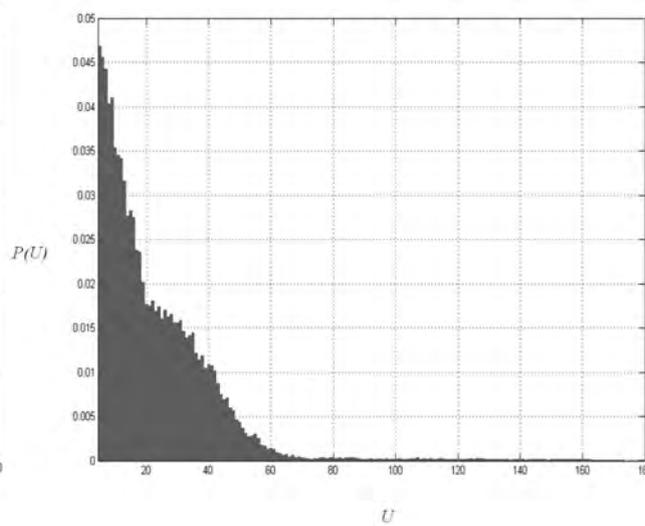
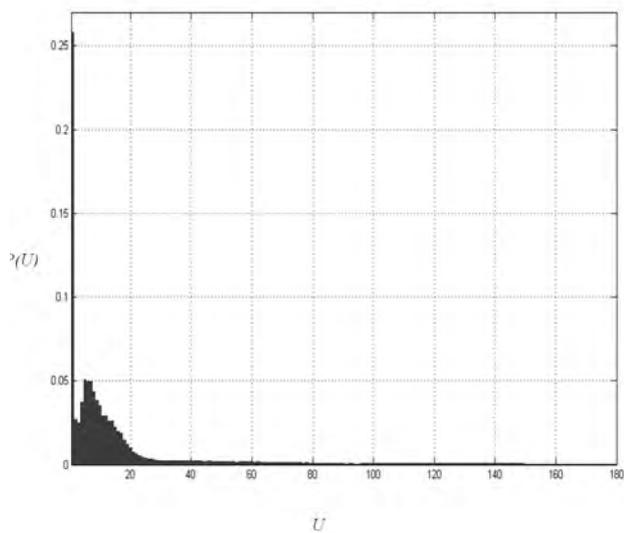
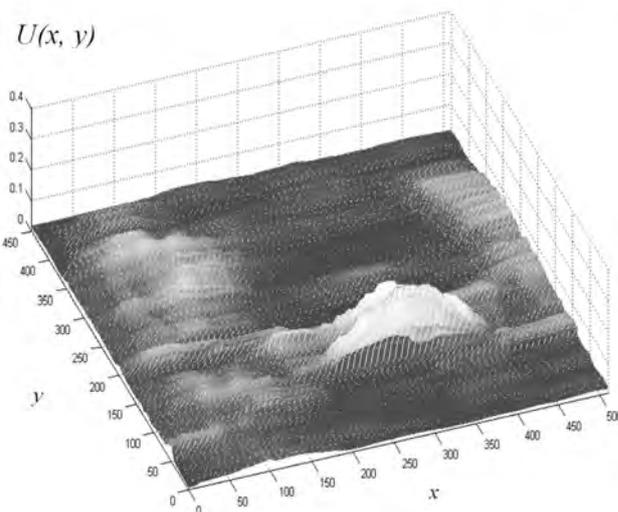
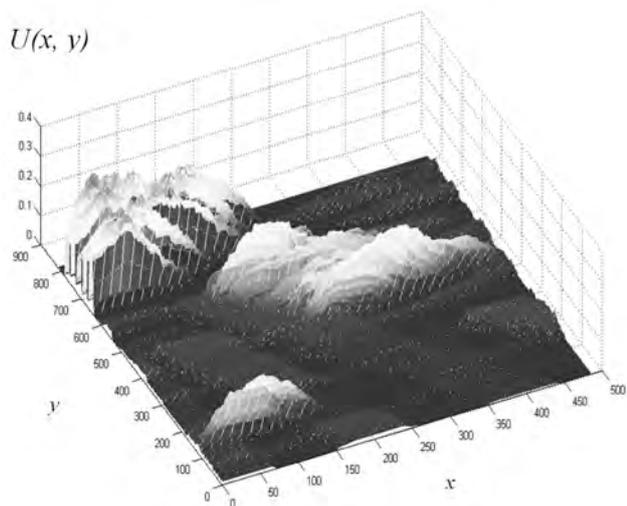


Рис.1. Тестовые изображения, распределение значений функции $U(x, y)$ для этих изображений и соответствующие им гистограммы

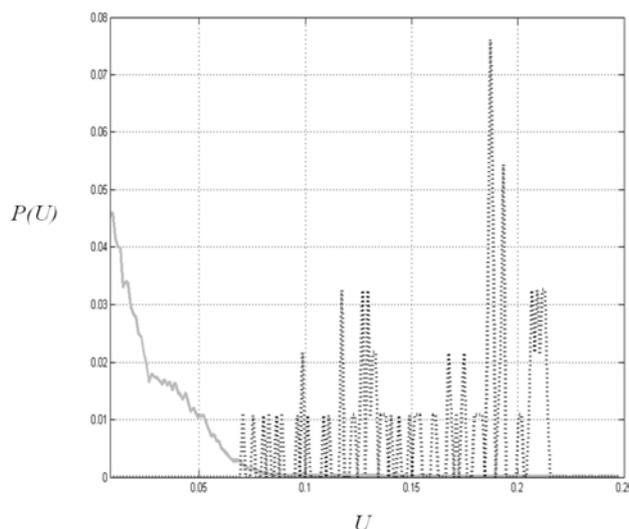


Рис.2. Сравнение гистограмм для областей с текстом и без

ний функции $U(x, y)$ для этих изображений и их гистограммы $P(U)$ соответственно.

Для расчета вероятностных характеристик обнаружения необходимо знать закон распределения значений достаточной статистики (2) для области изображения содержащей текст. Результаты исследований показали, что описать текстовые области на изображениях каким-либо из известных законов распределения достаточно сложно. Так, например, на рис. 2 приведена гистограмма достаточной статистики $U(x, y)$ участка изображения, соответствующего автомобильному номеру (пунктирная линия). Для сравнения сплошной линией показана гистограмма достаточной статистики $U(x, y)$ всего изображения. Очевидно, что в данном случае для определения вероятностных характеристик обнаружения целесообразно использовать численные методы.

Порог h для обнаружения области, содержащей текстовую форму, будем вычислять численно на основе анализа гистограммы при заданной вероятности ложной тревоги по выражению

$$1 - P_{\text{лм}} = \frac{1}{m} \sum_{i=1}^k n_i, \quad k = \frac{h}{\Delta x}, \quad (5)$$

где n_i – число экспериментальных точек в k интервале гистограммы, Δx – размер интервала гистограммы, m – общее количество точек функции $U(x, y)$.

На последнем этапе к полученному по (4) бинарному изображению, применяется морфологическая операция наращивания [13], для удаления разрывов обнаруженных областей. После проведения данной операции производится определение параметров обнаруженных текстовых областей (координаты, длина ширина и т. д.) на основе алгоритма поиска связанных контуров [13]. Каждый обнаруженный блок текста нумеруется и подсчитывается количество пикселей входящих в данный блок (площадь), а также вычисляются его координаты (координаты верхней левой точки) и геометрические размеры (длина и ширина).

Для наглядной демонстрации работы представленного способа приведем примеры обнаружения текстовых областей на подборке изображений содержащих рекламу с различным шрифтом и на фоне различной сложности.

На рис. 3 представлены результаты работы предложенного алгоритма.

Анализ рис. 3 а) и б) показывает: в случае с рис. 3 а) обнаружено 8 блоков текста из 10, сложности с обнаружением возникли для текста расположенного под большим углом; в случае с рис. 3 б) блоки текста локализованы точно, пропусков текста и ложных срабатываний не наблюдается.

При моделировании разработанного способа на ЭВМ на изображениях, содержащих



Рис.3. Результаты обработки тестовых изображений

текст различного стиля, размера и ориентации, было обнаружено и локализовано 76 % текстовых областей. В качестве тестовых изображений использовались изображения кадров рекламы. На простом однородном фоне процент обнаруженных текстовых форм приближается к 100 %, на сложных фонах наблюдаются ложные срабатывания и частичные пропуски текста. Повысить качество обнаружения и локализации текстовых форм можно путем предварительной настройки алгоритма (соответствующий выбор опорной функции $T(x)$ и ее размера). Например, при решении задачи выделения автомобильных номеров размер номерной рамки в кадре заранее известен.

Следует отметить, что при разработке алгоритма, задача распознавания обнаруженного текста не ставилась, необходимо было определить лишь наличие текста и площадь кадра занимаемую текстом.

4. ЗАКЛЮЧЕНИЕ

В данной работе представлен алгоритм обнаружения и локализации текстовых областей на цифровом изображении, основанный на развитии метода обнаружения горизонтальных перепадов яркости. В основе предложенного алгоритма лежит формирование достаточной статистики изображения и сравнения

ее с порогом. В частности, предложено выражение для формирования порога на основе анализа гистограмм двумерного поля, полученного в результате построчной корреляционной обработки исходного изображения. Представлены результаты применения предложенного алгоритма для локализации текста на изображениях содержащих рекламу. Представленный алгоритм обладает простотой реализации и позволяет получить характеристики достаточные для решения прикладных задач, таких как борьба с рекламой, спамом или поиска информации о рассматриваемой сцене (название города, улицы, даты, количественные и временные показатели).

СПИСОК ЛИТЕРАТУРЫ

1. Андрианов А. И. Локализация текста на изображениях сложных графических сцен / А. И. Андрианов // «Современные проблемы науки и образования». Электронный научный журнал. – № 3. – 2013. – URL: <http://www.science-education.ru/109>.
2. Запрягаев С. А., Сорокин А. И. Сегментация рукописных и машинных текстов методом диаграмм Вороного / С. А. Запрягаев, А. И. Сорокин // Вестник Воронеж. гос. ун-та. Сер. Системный анализ и информационные технологии. – 2010. – № 1.

3. Веснин Е. Н., Царев В. А. Оптимизация процесса обработки данных в системах распознавания буквенно-цифровых меток движущихся объектов // Интеллектуальные системы и компьютерные науки: Материалы IX международной конференции. – М. : Изд-во механико-математического факультета МГУ, 2006. – Т. 2, Ч. 1. – С. 73–79.

4. Shahab A., Shafait F., Dengel A. Bayesian Approach to Photo Time-Stamp Recognition // The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011, P. 1039–1043.

5. Epshtein B., Ofek E., Wexler Y. Detecting Text in Natural Scenes with Stroke Width Transform // 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – V. 5. – San Francisco, 2010.

6. Kunishige Y., Yaokai F., Uchida S. Scenery Character Detection with Environmental Context // The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011. – P. 1049–1053.

7. Uchida S., Shigeyoshi Y., Kunishige Y., Yaokai F. A Keypoint-Based Approach Toward

Scenery Character Detection // The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011. – P. 819–823.

8. Du Y., Ai H., Lao S. Dot Text Detection Based on FAST Points // The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011. – P. 435–439.

9. Jung C., Liu Q. F., Kim J. Accurate text localization in images based on SVM output scores / C. Jung, Q. F. Liu, J. Kim // Image and Vision Computing. – V. 27. – 2009. – P. 1295–1301.

10. Коренной А. В. Математические модели полутоновых изображений / А. В. Коренной. – Радиотехника, 2007. – N 8. – С. 79–81.

11. Богословский А. В. Обработка многомерных сигналов / А. В. Богословский. – М. : Радиотехника, 2013. – 168 с.

12. Шаров С. А. Статистика слов в русском языке / С. А. Шаров. – URL: http://www.lingvisto.org/artikoloj/ru_stat.html.

13. Гонсалес Р., Вудс Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М. : Техносфера, 2005. – 1072 с.

Коренной Александр Владимирович – д.т.н., профессор кафедры средств связи (и авиационных комплексов связи) ВУНЦ ВВС «Военно-воздушная академия имени профессора Н. Е. Жуковского и Ю. А. Гагарина»

Юдаков Дмитрий Сергеевич – к.т.н., преподаватель кафедры средств связи (и авиационных комплексов связи) ВУНЦ ВВС «Военно-воздушная академия имени профессора Н. Е. Жуковского и Ю. А. Гагарина»
E-mail: yds12345@rambler.ru

Дедов Сергей Владимирович – к.ф.-м.н., начальник методического отдела учебно-методического центра ВУНЦ ВВС «Военно-воздушная академия имени профессора Н. Е. Жуковского и Ю. А. Гагарина»

Стражник Владимир Петрович – к.т.н., начальник кафедры эксплуатации радиотехнических средств (обеспечения полетов) ВУНЦ ВВС «Военно-воздушная академия имени профессора Н. Е. Жуковского и Ю. А. Гагарина»

Korennoi Aleksandr V. – Communication Department, Zhukovsky-Gagarin Air Force Academy, Voronezh, Prof.

Yudakov Dmitriy S. – Communication Department, Zhukovsky-Gagarin Air Force Academy, Voronezh; PhD, assistant
E-mail: yds12345@rambler.ru

Dedov Sergey V. – Communication Department, Zhukovsky-Gagarin Air Force Academy, Voronezh; PhD, the chief of a methodical department.

Strazhnik Vladimir P. – Communication Department, Zhukovsky-Gagarin Air Force Academy, Voronezh; PhD, the chief of department.