

ISSN 1816-0301

ИНФОРМАТИКА

ТОМ 15



3

ИЮЛЬ-СЕНТЯБРЬ
2018

ОТ РЕДАКЦИИ

В журнале «Информатика» публикуются оригинальные и обзорные статьи, описывающие результаты фундаментальных и прикладных исследований специалистов академического и вузовского профиля в области информатики и информационных технологий. Основной целью журнала является публикация наиболее значимых новых результатов в указанной области. Приветствуются статьи, описывающие заключительные результаты научных проектов и диссертационных исследований, открывающие новые направления исследований, которые находятся на стыке информатики и других наук.

К публикации приглашаются отечественные и зарубежные авторы, специалисты в области информационных технологий и молодые ученые.

В декабре 2017 г. журнал включен в базу данных Российского индекса научного цитирования (РИНЦ). С помощью инструментов и сервисов, доступных на платформе eLIBRARY (раздел «Личный кабинет»), можно самостоятельно корректировать список своих публикаций и цитирований в РИНЦ.

Журнал рассчитан на широкий круг специалистов в области информатики и информационных технологий.

Сайт журнала: <http://inf.grid.by>

Адрес редакции:

ул. Сурганова, 6, к. 305, г. Минск, Республика Беларусь

Тел. +375(017)284 26 22

E-mail: rio@newman.bas-net.by

ОБЪЕДИНЕННЫЙ ИНСТИТУТ ПРОБЛЕМ ИНФОРМАТИКИ
НАЦИОНАЛЬНОЙ АКАДЕМИИ НАУК БЕЛАРУСИ

ИНФОРМАТИКА

Том 15, № 3, июль-сентябрь 2018

Ежеквартальный научный журнал

Издается с января 2004 г.

Учредитель – Объединенный институт проблем информатики
Национальной академии наук Беларуси

Г л а в н ы й р е д а к т о р

Александр Васильевич Тузиков – генеральный директор Объединенного института проблем информатики Национальной академии наук Беларуси, Минск, Республика Беларусь

Р е д а к ц и о н н а я к о л л е г и я

- М. Я. Ковалев** – Объединенный институт проблем информатики Национальной академии наук Беларуси, Минск, Республика Беларусь (*заместитель главного редактора*)
- С. В. Абламейко** – Белорусский государственный университет, Минск, Республика Беларусь
- В. В. Анищенко** – Компания «СофтКлуб», Минск, Республика Беларусь
- П. Н. Бибило** – Объединенный институт проблем информатики Национальной академии наук Беларуси, Минск, Республика Беларусь
- М. Н. Бобов** – ОАО «АГАТ – системы управления» – управляющая компания холдинга «Геоинформационные системы управления», Минск, Республика Беларусь
- А. Б. Долгий** – Высшая инженерная школа Бретани, Нант, Франция
- А. Н. Дудин** – Белорусский государственный университет, Минск, Республика Беларусь
- А. А. Карпов** – Институт информатики и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург, Российская Федерация
- С. Я. Килин** – Президиум Национальной академии наук Беларуси, Минск, Республика Беларусь
- В. В. Краснопрошин** – Белорусский государственный университет, Минск, Республика Беларусь
- А. М. Крот** – Объединенный институт проблем информатики Национальной академии наук Беларуси, Минск, Республика Беларусь
- С. В. Кругликов** – Объединенный институт проблем информатики Национальной академии наук Беларуси, Минск, Республика Беларусь
- С. П. Кундас** – Белорусский национальный технический университет, Минск, Республика Беларусь

- Н. А. Лиходед** – Белорусский государственный университет, Минск, Республика Беларусь
- П. П. Матус** – Институт математики Национальной академии наук Беларуси, Минск, Республика Беларусь
- А. А. Петровский** – Белорусский государственный университет информатики и радиоэлектроники, Минск, Республика Беларусь
- В. А. Скляр** – Университет Авейру, Португалия
- Ю. С. Сотсков** – Объединенный институт проблем информатики Национальной академии наук Беларуси, Минск, Республика Беларусь
- А. Л. Стемпковский** – Институт проблем проектирования в микроэлектронике Российской академии наук, Москва, Российская Федерация
- Ю. С. Харин** – Научно-исследовательский институт прикладных проблем математики и информатики Белорусского государственного университета, Минск, Республика Беларусь
- А. Ф. Чернявский** – Институт прикладных физических проблем им. А. Н. Севченко Белорусского государственного университета, Минск, Республика Беларусь
- В. Н. Ярмолик** – Белорусский государственный университет информатики и радиоэлектроники, Минск, Республика Беларусь

ИНФОРМАТИКА

Том 15, № 3, июль-сентябрь 2018

Ответственный за выпуск *С. С. Мойсейчик*
Редактор *Г. Б. Гончаренко*
Корректор *А. А. Михайлова*
Компьютерная верстка *О. Б. Бутевич*

Сдано в набор 20.08.2018. Подписано в печать 21.09.2018.
Формат 60×84 1/8. Бумага офсетная. Гарнитура Таймс. Ризография.
Усл. печ. л. 13,5. Уч.-изд. л. 13,2. Тираж 50 экз. Заказ 8.

Государственное научное учреждение «Объединенный институт проблем информатики Национальной академии наук Беларуси».
Свидетельство о государственной регистрации издателя, изготовителя, распространителя печатных изданий № 1/274 от 04.04.2014.
ЛП № 02330/444 от 18.12.13.
Ул. Сурганова, 6, 220012, Минск.

THE UNITED INSTITUTE OF INFORMATICS PROBLEMS
OF THE NATIONAL ACADEMY OF SCIENCES OF BELARUS

INFORMATICS

Vol. 15, no. 3, Jule-September 2018

Published quarterly
Issued since January 2004

Founder – the United Institute of Informatics Problems
of the National Academy of Sciences of Belarus

Editor-in-Chief

Alexander V. Tuzikov – General Director of the United Institute of Informatics Problems
of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus

Editorial Board

- | | |
|--------------------------------|---|
| Mikhail Y. Kovalyov | – the United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus (<i>Deputy Editor-in-Chief</i>) |
| Sergey V. Ablameyko | – Belarusian State University, Minsk, Republic of Belarus |
| Uladimir V. Anishchanka | – SoftClub Ltd., Minsk, Republic of Belarus |
| Petr N. Bibilo | – the United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus |
| Mikhail N. Bobov | – Open Joint-Stock Company “AGAT – Control Systems – Managing Company of Geoinformation Control Systems Holding”, Minsk, Republic of Belarus |
| Alexandre B. Dolgui | – IMT Atlantique, Nantes, France |
| Alexander N. Dudin | – Belarusian State University, Minsk, Republic of Belarus |
| Alexey A. Karpov | – Institute of Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), St. Petersburg, Russian Federation |
| Sergey Ya. Kilin | – Presidium of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus |
| Viktor V. Krasnoproshin | – Belarusian State University, Minsk, Republic of Belarus |
| Alexander M. Krot | – the United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus |
| Sergey V. Kruglikov | – the United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus |
| Semen P. Kundas | – Belarusian National Technical University, Minsk, Republic of Belarus |

- Nikolai A. Likhoded** – Belarusian State University, Minsk, Republic of Belarus
Petr P. Matus – Institute of Mathematics of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus
Alexander A. Petrovsky – Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus
Valery A. Sklyarov – University of Aveiro, Portugal
Yuri N. Sotskov – the United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus
Alexander L. Stempkovsky – the Institute for Design Problems in Microelectronics of the Russian Federation Academy of Sciences, Moscow, Russian Federation
Yuriy S. Kharin – Research Institute for Applied Problems of Mathematics and Informatics of the Belarusian State University, Minsk, Republic of Belarus
Alexander F. Cherniavsky – A. N. Sevchenko Institute of Applied Physical Problems of the Belarusian State University, Minsk, Republic of Belarus
Vyacheslav N. Yarmolik – Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

INFORMATICS

Vol. 15, no. 3, Jule-September 2018

Issue Head *S. S. Maiseichyk*
Editor *H. B. Hancharenka*
Corrector *H. A. Mikhailava*
Computer Imposition *V. B. Butsevich*

Sent for press 20.08.2018. Output 21.09.2018.
Format 60×84 1/8. Offset paper. Headset Times. Riesography.
Printed sheets 13,5. Publisher's signatures 13,2. Circulation 50 copies. Order 8.

State Scientific Institution «The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus».
Certificate on the state registration of the publisher, manufacturer,
distributor of printing editions no. 1/274 dated 04.04.2014.
License for the press no. 02330/444 dated 18.12.13.
6, Surganov Str., 220012, Minsk, Republic of Belarus.

ISSN 1816-0301 (print)

СОДЕРЖАНИЕ

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ И РЕЧИ

- Ковалев В. А., Павленко Д. А., Снежко Э. В., Левчук В. А., Калиновский А. А.** Сравнительный анализ вычислительных платформ для бортового микромодуля предварительного распознавания изображений.....7
- Рыбенков Е. В., Петровский Н. А.** Синтез FPGA-архитектур банков фильтров на основе блочной лестничной факторизации в алгебре кватернионов (часть 2).....22
- Артемьев В. М., Наумов А. О., Кохан Л. Л.** Линейная адаптивная фильтрация случайных последовательностей на основе детерминированного подхода32
- Старовойтов В. В.** Уточнение индекса *SSIM* структурного сходства изображений41

ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ

- Бибило П. Н., Кириенко Н. А., Ланкевич Ю. Ю.** Логическая оптимизация многоуровневых представлений систем булевых функций на основе блочного разбиения и разложения Шеннона.....56

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

- Маркевич В. Э., Легкоступ В. В.** Модифицированный метод пропорционального наведения при ограниченном секторе сопровождения объекта управления.....71
- Ping P., Petrenko Y. N.** Flexible multiple access control algorithm in 5G mobile wireless system.....93

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

- Ерофеев В. Т., Урбанович А. И.** Математическая модель распространения электромагнитных волн в композитных средах со сфероидальными частицами102

УЧЕНЫЕ БЕЛАРУСИ

- Александр Васильевич Тузиков** (К 60-летию со дня рождения).....113

ISSN 1816-0301 (print)

CONTENTS

SIGNAL, IMAGE AND SPEECH PROCESSING

Kovalev V. A., Paulenka D. A., Snezhko E. V., Liauchuk V. A., Kalinovski A. A. Comparative analysis of computing platforms for onboard micromodule of provisional image recognition.....	7
Rybenkov E. V., Petrovsky N. A. Synthesis of FPGA architectures of block lifting-based filter banks in quaternion algebra (part 2).....	22
Artemiev V. M., Naumov A. O., Kokhan L. L. Linear adaptive filtering of random sequences on basis of deterministic approach.....	32
Starovoitov V. V. Enhancement of the structural similarity index <i>SSIM</i>	41

LOGICAL DESIGN

Bibilo P. N., Kirienco N. A., Lankevich Y. Y. Logical optimization the multilevel representations of systems of Boolean functions based on partitioning into blocks and Shannon decomposition	56
--	----

APPLIED INFORMATION TECHNOLOGIES

Markevich V. E., Legkostup V. V. Modified method of proportional guidance in case of the limited sector of tracking of the control object	71
Ping P., Petrenko Y. N. Flexible multiple access control algorithm in 5G mobile wireless system.....	93

MATHEMATICAL MODELING

Erofeenko V. T., Urbanovich A. I. Mathematical model of propagation of electromagnetic waves in composite media with spheroidal particles	102
--	-----

SCIENTISTS OF BELARUS

Alexander Vasilievich Tuzikov (On his 60 th birthday).....	113
--	-----

ISSN 1816-0301 (print)

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ И РЕЧИ*SIGNAL, IMAGE AND SPEECH PROCESSING*

УДК 004.9

Поступила в редакцию 06.04.2018

Received 06.04.2018

В. А. Ковалев, Д. А. Павленко, Э. В. Снежко, В. А. Левчук, А. А. Калиновский*Объединенный институт проблем информатики
Национальной академии наук Беларуси, Минск, Беларусь***СРАВНИТЕЛЬНЫЙ АНАЛИЗ ВЫЧИСЛИТЕЛЬНЫХ ПЛАТФОРМ
ДЛЯ БОРТОВОГО МИКРОМОДУЛЯ ПРЕДВАРИТЕЛЬНОГО
РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ**

Аннотация. Приводятся результаты анализа базового аппаратного и программного обеспечения существующих недорогих и коммерчески доступных вычислительных микроплатформ с целью выбора подходящего решения при разработке бортового микро модуля предварительного распознавания и отбора изображений подстилающих поверхностей заданных типов. Предполагается, что соответствующие варианты микро модуля могут быть установлены на борту малых космических аппаратов либо легких беспилотных летательных аппаратов (дронов). Рассматривается вариант микро модуля для дронов. Основным ограничением при выборе микроплатформы был ее небольшой вес (не более 300 г, включая камеру и интерфейсное оборудование) при относительно высокой производительности (время распознавания кадра цветного изображения размером 320×240 пикселей не более 300 мс). Другими важными ограничениями являлись невысокая цена и коммерческая доступность микроплатформы на рынке Беларуси. Приведенные в работе сведения могут быть полезны инженерам и научным работникам, разрабатывающим компактные бюджетные мобильные системы обработки, анализа и распознавания изображений.

Ключевые слова: микрокомпьютер, мобильная система, распознавание изображений, дрон, сверточная нейронная сеть

Для цитирования. Сравнительный анализ вычислительных платформ для бортового микро модуля предварительного распознавания изображений / В. А. Ковалев [и др.] // Информатика. – 2018. – Т. 15, № 3. – С. 7–21.

V. A. Kovalev, D. A. Paulenka, E. V. Snezhko, V. A. Liauchuk, A. A. Kalinovski*The United Institute of Informatics Problems of the National Academy
of Sciences of Belarus, Minsk, Belarus***COMPARATIVE ANALYSIS OF COMPUTING PLATFORMS
FOR ONBOARD MICROMODULE OF PROVISIONAL IMAGE RECOGNITION**

Abstract. This paper is devoted to the analysis of basic hardware and software of recent cheap and commercially available computing microplatforms for selecting an appropriate solution for development of an onboard micromodule for preliminary recognition and selection of images of underlying surface of given types. It is assumed that the corresponding versions of the micromodule can be installed on board of small spacecraft or light unmanned aerial vehicles (drones). In this paper we consider a variant of a micromodule for drones. When choosing a microplatform, the main limitations were its low weight (no more than 300 g, including camera and interface equipment) and its relatively high performance (time for frame processing of a color image 320×240 pixels is no more than 300 ms). Another important limitation was the low price and commercial availability of microplatform on the Belarusian market. The information provided in this paper could be useful for engineers and researchers who are developing compact budget mobile systems for processing, analyzing and recognition of images.

Keywords: microcomputer, mobile system, image recognition, drone, convolutional neural network

For citation. Kovalev V. A., Paulenka D. A., Snezhko E. V., Liauchuk V. A., Kalinovski A. A. Comparative analysis of computing platforms for onboard micromodule of provisional image recognition. *Informatics*, 2018, vol. 15, no. 3, pp. 7–21 (in Russian).

Введение. В последние годы наблюдается значительный прогресс в области разработки и использования малых космических и беспилотных летательных аппаратов (дронов). Это создает предпосылки для разработки различных систем автоматического сбора видеоданных, в том числе мобильных систем поиска, распознавания и отбора необходимых видеоданных непосредственно на борту летательного аппарата. Такая возможность избавляет оператора от длительного и утомительного просмотра отснятых фото- и видеоматериалов с целью выбора участков (объектов) интереса, поскольку система сбора данных фиксирует только изображения тех типов (классов), образцы которых были включены в полетное задание перед вылетом. Кроме того, в ряде случаев появляется дополнительная возможность существенной экономии энергии на борту, так как энергоемкие компоненты наблюдения, съемки и передачи данных могут включаться только в случае необходимости.

Целью настоящей работы являются анализ базового аппаратного и программного обеспечения недорогих, коммерчески доступных вычислительных микроплатформ и проведение комплекса вычислительных экспериментов для выбора подходящего решения при разработке микромодуля предварительного распознавания и отбора изображений подстилающих поверхностей на борту легких дронов. Перечислим основные функции микромодуля и ключевые технические решения:

1. Предполагается, что микромодуль является автономной системой автоматического распознавания изображений подстилающих поверхностей различных типов. Примерами указанных типов поверхностей могут служить леса различных видов, промышленные сооружения, водоемы, дороги, поля, кустарники, дачная застройка, сельхозугодия с растительностью, находящейся на различных стадиях роста, и т. п.

2. Конкретные целевые классы задаются в полетном задании либо в виде образцов соответствующих изображений в случае использования традиционной схемы распознавания типа «изображение – признаки – классификатор», либо в виде предобученной сверточной нейронной сети в случае использования технологии глубокого обучения (Deep Learning).

3. Встроенная видеокамера микромодуля используется только для решения задачи распознавания типа поверхности. При необходимости кадры видеопоследовательности (изображения низкого разрешения целевых классов полетного задания) могут сохраняться на внешнем носителе микромодуля для вспомогательных целей.

4. Взаимодействие микромодуля с основным бортовым микрокомпьютером дрона должно быть минимальным. По сути, оно должно сводиться к передаче ему в реальном времени номера текущего класса подстилающей поверхности. При этом класс «0» означает, что текущее изображение не относится ни к одному из целевых классов, заданных в полетном задании. Дополнительным параметром является вероятность (степень уверенности) отнесения к тому или иному классу. Принимается окончательное решение о съемке и записи (передаче) на землю изображения, а сама высококачественная съемка производится за пределами микромодуля.

5. Предполагается, что выбираемый микрокомпьютер является законченной вычислительной системой, снабженной всем необходимым интерфейсным оборудованием, операционной системой (ОС), средствами разработки программ и др. Вариант реализации микромодуля из отдельных специализированных компонентов (например, микроконтроллеров, цифровых сигнальных процессоров (Digital Signal Processor, DSP), чипсетов, систем на кристалле и др.) с помощью соответствующих схмотехнических решений не рассматривается.

Сравнительный анализ выполнен по методу стратегического планирования SWOT (Strengths, Weaknesses, Opportunities, Threats) [1], который заключается в выявлении факторов внутренней и внешней среды устройства и разделении их на четыре категории: сильные стороны (Strengths), слабые стороны (Weaknesses), возможности (Opportunities) и угрозы (риски) (Threats) использования анализируемого устройства. Сильные (*S*) и слабые (*W*) стороны являются факторами внутренней среды объекта анализа, т. е. тем, на что сам объект способен повлиять; возможности (*O*) и угрозы (*T*) являются факторами внешней среды, т. е. тем, что может повлиять на объект извне и при этом не контролируется самим объектом. Например, микрокомпьютер Raspberry Pi Zero Wireless обладает одноядерным процессором с частотой 1 ГГц и оперативной памятью 512 МБ, это факторы внутренней среды устройства. Факторами внешней среды являются возможные физические воздействия внешней среды, а также выбор алгоритма, языка про-

граммирования, разработка архитектуры программного обеспечения (ПО) микро модуля, которые сильно влияют на время распознавания одного изображения и энергопотребление.

SWOT-анализ может использоваться в любой ситуации принятия решений, когда определено целевое конечное состояние (цель). Способами применения SWOT являются сопоставление и преобразование. Сопоставление используется, чтобы найти конкурентное преимущество, сравнивая сильные стороны с возможностями. Преобразование заключается в превращении слабых сторон и угроз в сильные стороны и возможности. Если слабые стороны и угрозы не могут быть преобразованы, то их следует попытаться свести к минимуму или избежать. В данном случае целью является объективная оценка соответствия устройства техническому заданию. Также авторы стремились выявить конкурентные преимущества одного устройства над другим, найти слабые стороны и угрозы и либо преобразовать их в сильные стороны и возможности, либо избежать.

Эволюция электронно-вычислительных систем привела к тому, что компьютеры стали не только значительно мощнее, дешевле и доступнее для потребителя, но и многократно уменьшились в размерах. Появился целый рынок одноплатных микрокомпьютеров, способных решать разнообразные задачи (URL: https://en.wikipedia.org/wiki/Comparison_of_single-board_computers).

В настоящее время микрокомпьютером считается электронно-вычислительное устройство, сравнимое с кредитной картой по размерам и с персональными компьютерами предыдущих поколений по мощности. Например, общая реальная производительность для моделей Raspberry Pi Zero Wireless сравнима с Pentium 2 с частотой процессора 300 МГц, но при этом с гораздо лучшей графикой (URL: <https://www.raspberrypi.org/help/faqs/#performance-Speed>). Микрокомпьютер, или одноплатный компьютер, – это самодостаточный компьютер, собранный на одной печатной плате, на которой установлены микропроцессор, оперативная память, системы ввода-вывода и другие необходимые модули. Одноплатные компьютеры отличаются малыми размерами и дешевизной. Самые дешевые из них можно приобрести по цене 5–10 долл. США (URL: <https://www.slant.co/topics/1629/~single-board-computers>).

Особого упоминания заслуживают факты использования смартфонов в качестве универсальных микрокомпьютеров. Показательным в этом отношении является проект NASA PhoneSat [2], выполненный в рамках программы Small Spacecraft Technology Program по созданию наноспутников с использованием немодифицированных коммерческих смартфонов и запуску таких спутников на низкую околоземную орбиту. Проект был начат в 2009 г. в NASA Ames Research Center (Moffett Field, Калифорния). Стоимость компонентов, использованных для создания одного наноспутника, составляла от 3500 долл. для версии 1.0 и менее 7000 долл. для версии 2.0. Наноспутник построен по стандарту CubeSat, имеет размер 1U (10×10×10 см) и вес приблизительно 1 кг. Для сравнения в 2004 г. до начала проекта PhoneSat стоимость компонентов для создания классического CubeSat оценивалась в сумму от 20 000 до 40 000 долл. Стоимость коммерческого запуска наноспутника 1U на низкую орбиту составила около 85 000 долл. Еще один проект под названием STRaND-1 (Surrey Training, Research and Nanosatellite Demonstrator 1) является наноспутником 3U CubeSat (вес 4,3 кг). Он был разработан в Англии специалистами Суррейского космического центра (Surrey Space Centre, SSC) и компании Surrey Satellite Technology Ltd (SSTL) и выведен на орбиту на борту PSLV C-20 Rocket из Индии 25 февраля 2013 г. STRaND-1 стал первым спутником, в управлении которого использован смартфон.

1. Основные требования, предъявляемые к разрабатываемому микро модулю. Техническим заданием определяются следующие требования и ограничения на рассматриваемую в статье разработку:

- 1) масса микро модуля (включая камеру и интерфейсное оборудование) не более 300 г;
- 2) габариты (длина, ширина, высота) не более 12×6×4 см;
- 3) энергопотребление в активном режиме не более 5 Вт;
- 4) время распознавания одного кадра RGB-изображения размером 320×240 пикселей не более 300 мс;
- 5) количество типов изображений подстилающей поверхности (классов), распознаваемых программным обеспечением, не менее 10, в том числе распознаваемых одновременно, – не менее 3;

б) качество распознавания заданного типа изображений подстилающей поверхности (в зависимости от типа) не менее 85–95 %;

7) цена одного микромодуля при единичном производстве не более 500 долл.

На первом этапе были рассмотрены следующие типы вычислительных микроплатформ: несколько одноплатных микрокомпьютеров различных производителей, неттопы, смартфоны, изделия с формфактором материнских плат Mobile-ITX для x86-совместимых процессоров, платы формата PC/104 (или PC104), комплект разработчика NVIDIA Jetson TX2 Developer Kit, SMARC-модули стандарта SMARC версии 2.0, высокопроизводительный встроенный компьютер Manifold DJI, разновидности так называемых компьютеров на модуле (computer-on-module, COM).

В результате предварительного анализа выяснилось, что большинство упомянутых выше моделей микрокомпьютеров превышают предельно допустимые массогабаритные параметры. Согласно опубликованным техническим спецификациям энергопотребление многих моделей также превышает ограничение в 5 Вт при активном режиме работы. Кроме того, цена некоторых микрокомпьютеров оказалась слишком высока и (или) их приобретение проблематично на рынке Беларуси. Так, например, SMARC-модули [3] не подходят как по цене, так и по размерам. Размер платы-носителя для SMARC-модуля (20×21 см) значительно превышает предельно допустимые габариты (12×6×4 см). Цена самых дешевых SMARC-модулей начинается от 100 долл. без учета стоимости необходимой платы-носителя.

2. Тестируемые микроплатформы и их технические характеристики. Из всего многообразия микроплатформ для проведения сравнительного анализа были отобраны и приобретены четыре современные, недорогие и коммерчески доступные микроплатформы, которые являются типичными представителями своих классов (рис. 1).

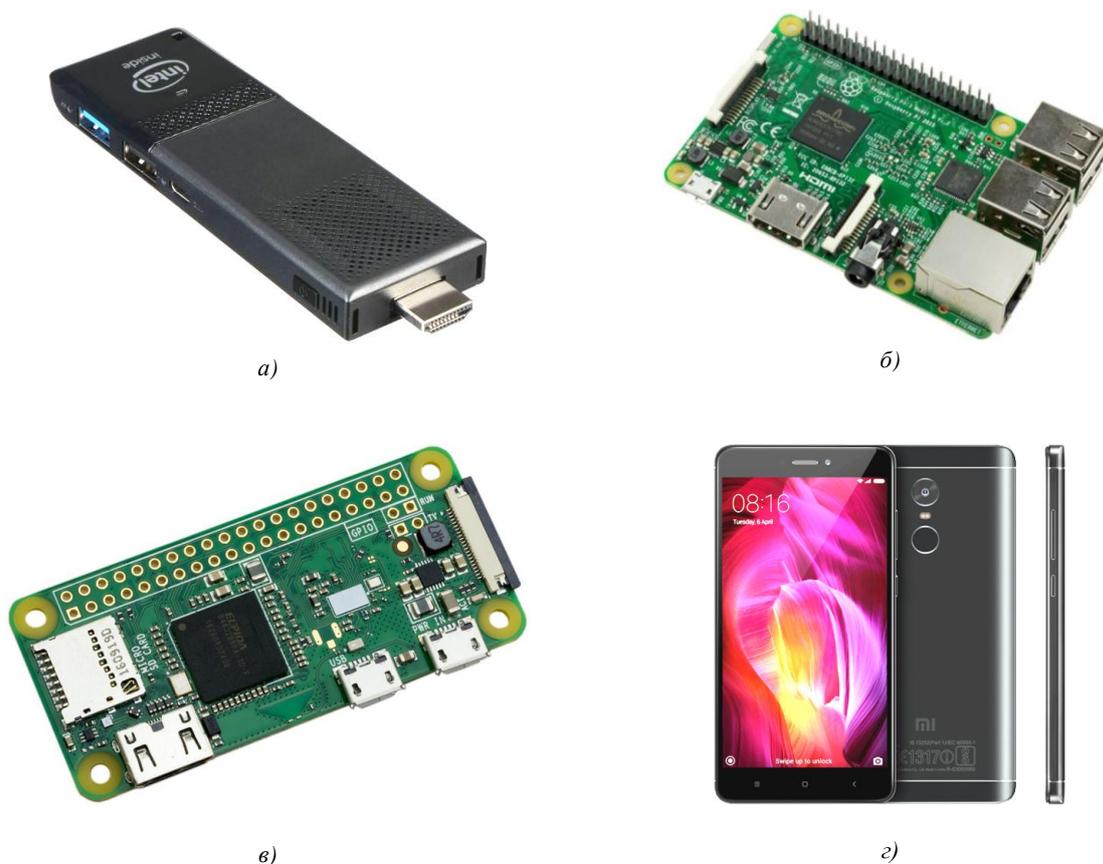


Рис. 1. Тестируемые образцы микроплатформ: а) Intel Compute Stick STK1AW32SC [4] с ОС Windows 10 Home Edition 32-bit; б) Raspberry Pi 3 Model B [5] с ОС Raspbian Stretch 4.9; в) Raspberry Pi Zero Wireless [6] с ОС Raspbian Stretch 4.9; г) Xiaomi Redmi Note 4 [7] с ОС MIUI 9.2 Global, основанной на Android 7.0 Nougat (современный бюджетный смартфон)

Цена устройств варьирует в пределах от 10 до 250 долл. и коррелирует с параметрами вычислительной мощности (количеством ядер процессора, тактовой частотой, объемом оперативной памяти, объемом энергонезависимой памяти), а также с наличием внешних интерфейсов и камеры (табл. 1). Все устройства удовлетворяют ограничению в 300 г, но Xiaomi Redmi Note 4 не проходит по габаритам (длина 15,1 см против предельно допустимых 12 см). Однако это превышение в определенном смысле компенсируется его небольшой толщиной в 0,9 см и, как следствие, небольшими габаритами. Диапазон рабочих температур для Raspberry Pi не указан, так как в официальных спецификациях таких данных нет. Модели Raspberry Pi собираются из коммерческих чипов разных производителей, каждый из которых имеет свой диапазон рабочих температур, микрочип LAN9514 для разъемов USB и Ethernet может работать в диапазоне температур от 0 до +70 °С, в то время как система на кристалле микрокомпьютера способна выдерживать температуры от –40 до +85 °С (URL: <https://www.raspberrypi.org/help/faqs/#performanceOperatingTemperature>).

Таблица 1

Основные технические характеристики тестируемых устройств

Характеристики	Intel Compute Stick	Raspberry Pi 3 Model B	Raspberry Pi Zero Wireless	Xiaomi Redmi Note 4
Дата выпуска, месяц/год	03/2016	02/2016	02/2017	01/2017
Рекомендованная розничная цена, долл.	129	35	10	250
Компания-производитель, страна	Intel Corporation, США	Raspberry Pi Foundation, Великобритания	Raspberry Pi Foundation, Великобритания	Xiaomi Inc., КНР
Длина, см	12,3	8,6	6,5	15,1
Ширина, см	3,8	5,7	3,0	7,6
Толщина, см	1,2	1,7	0,5	0,9
Масса, г	60	45	9	165 (с батареей)
Диапазон рабочих температур, °С	от 0 до +35	–	–	от 0 до +40
Модель центрального процессора (CPU)	Intel Atom x5-Z8300	ARM Cortex-A53	ARM1176JZF-S	Snapdragon 625 (MSM8953)
Частота процессора (CPU), ГГц	1,84	1,2	1	2,02
Количество ядер процессора, шт.	4	4	1	8
Модель графической карты (GPU)	Intel HD Graphics	Broadcom VideoCore IV 1080p60	Broadcom VideoCore IV 1080p30	Adreno 506 (650 МГц)
Размер оперативной памяти, ГБ	2	1	0,5	4
Размер встроенной памяти, ГБ	32	нет	нет	64
Разрешение специализированной (встроенной) камеры, Мп	нет	5 или 8	5 или 8	13, 5
Автофокусировка	нет	нет	нет	есть

Примечание: жирным шрифтом отмечены параметры, превышающие требования технического задания.

Известно, что для мобильных вычислительных устройств важной группой параметров, которые влияют на выбор платформы, являются параметры электропитания и потребляемой мощности. Соответствующие значения для рассматриваемых четырех вариантов представлены в табл. 2 и на рис. 2.

Таблица 2

Параметры электропитания и потребляемой мощности тестируемых устройств

Intel Compute Stick			Raspberry Pi 3 Model B			Raspberry Pi Zero Wireless			Xiaomi Redmi Note 4		
В	А	Вт	В	А	Вт	В	А	Вт	В	А	Вт
<i>Работа без нагрузки</i>											
5,0	0,80±0,10	4,00±0,50	5,0	0,35 ±0,01	1,75 ±0,05	5,0	0,12±0,01	0,60±0,05	4,2	0,12±0,02	0,51±0,08
<i>В нагруженном состоянии</i>											
5,0	1,70±0,30	8,50±1,50	5,0	1,34 ±0,10	6,70 ±0,50	5,0	0,35±0,05	1,75±0,25	4,2	0,7±0,05	2,94±0,21

Примечание: жирным шрифтом отмечены параметры, превышающие требования технического задания.

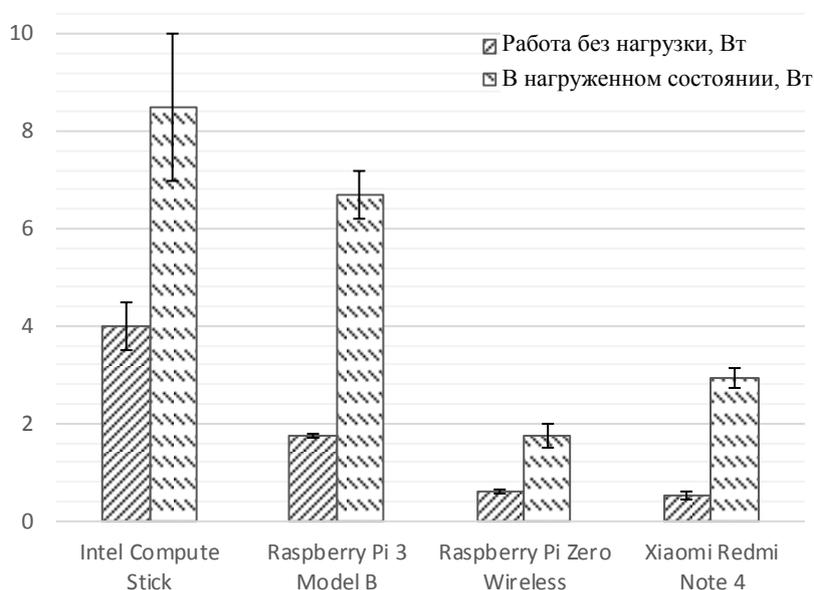


Рис. 2. Потребляемая мощность устройств

Несмотря на то что устройства используют напряжение питания 5 В (4,2 В для смартфона), различия в среднем значении потребляемого тока в нагруженном состоянии варьируют от 0,35 А у Raspberry Zero до 1,70 А у Intel Stick. Следовательно, средняя потребляемая мощность меняется от 1,75 до 8,50 Вт. Для всех устройств, кроме Xiaomi Redmi Note 4, следует учитывать, что при использовании внешней камеры типа Raspberry Pi дополнительно потребуется еще 0,25 А для обеспечения ее нормального функционирования. Поскольку в соответствии с входными требованиями (см. разд. 1) потребляемая мощность микромодуля в рассматриваемом случае не должна превышать 5 Вт, Intel Stick и Raspberry Pi 3 использоваться не должны. Кроме того, следует отметить, что из-за относительно большой потребляемой мощности и компактного формфактора микрокомпьютер Intel Stick в его оригинальном фирменном варианте снабжается охлаждающим вентилятором и имеет соответствующие отверстия в корпусе. При эксплуатации на открытом воздухе это может привести к попаданию влаги внутрь корпуса и поломке устройства.

3. Базовое программное обеспечение и его особенности. Многие современные микрокомпьютеры позволяют устанавливать на них те же самые ОС и ПО, что и на обычных персональных компьютерах. Это дает возможность вести разработку на рабочем компьютере, а затем с минимальными изменениями переносить результаты на микроплатформу, что упрощает и ускоряет работу над проектом.

Intel Compute Stick поставляется с предустановленной ОС Windows 10 Home Edition объемом 32 бита (платной). Специалисты утверждают [8], что на данное устройство можно также установить ОС Ubuntu 16.04, однако эта опция не тестировалась. Известно, что home-версия позиционируется как «домашняя» ОС, ориентированная на использование частными лицами вне офиса или предприятия. По этой причине у этой ОС отсутствуют многие сервисы. Например, невозможно отключить автоматические обновления системы, нет предустановленного Remote Desktop, групповых политик и т. д. ОС достаточно ресурсоемкая, потребляет сравнительно много оперативной памяти микрокомпьютера и занимает две трети объема встроенной энерго-независимой памяти.

Для устройств семейства Raspberry Pi (Raspberry 3 и Raspberry Zero) рекомендуемой ОС является Raspbian [9], основанная на Debian Linux. Последний релиз – это версия Raspbian Stretch 4.9, которая основана на ядре Linux и является бесплатной с лицензией GNU и открытым исходным кодом. Несмотря на то что на рассмотренных микрокомпьютерах установлены 32-битные версии Windows и Raspbian, установка 64-битных версий Windows и Linux также возможна, но в данном случае нецелесообразна, так как эти версии требуют больше оперативной и энергонезависимой памяти. В результате для установки и эффективного выполнения прикладного ПО распознавания изображений остается недостаточно ресурсов.

ОС MIUI 9.2 Global смартфона Xiaomi Redmi Note 4 основана на исходных кодах Android 7.0 Nougat. Пользовательский интерфейс отличается от интерфейса оригинального Android, но сама система MIUI совместима с приложениями, разработанными для ОС Android. Из приложения Android Market, которое доступно на ОС MIUI 9.2, можно загрузить и установить любое приложение для системы Android. Все три ОС поддерживают такие распространенные языки программирования и их среды, как Java, C/C++ и Python, но ОС Android достаточно слабо поддерживает Python. Сама ОС написана на языках программирования C/C++ и Java, которые являются оптимальными при разработке приложений для Android. Есть несколько проектов, которые разрабатывают интерпретаторы Python для Android (URL: <https://wiki.python.org/moin/Android>): BeeWare, Chaquopy, Kivy, pyqtdeploy, QPython, Termux и т. д. Однако при выборе таких интерпретаторов существует достаточно высокий риск, что в будущем некоторые из них перестанут поддерживаться разработчиками. Также есть все основания полагать, что поддержка языка программирования Python для Android со временем улучшится. Кроме того, существует так называемый Termux (URL: https://wiki.termux.com/wiki/Main_Page) – свободно доступный эмулятор Linux для ОС Android, который хорошо зарекомендовал себя при проведении вычислительных экспериментов, рассматриваемых ниже.

4. Результаты тестирования производительности микроплатформ. Вычислительная мощность микроплатформ играет весьма существенное значение при решении задачи предварительного распознавания изображений подстилающей поверхности в реальном времени. В рассматриваемом случае время распознавания одного кадра RGB-изображения размером 320×240 пикселей ограничивается величиной 300 мс, т. е. производительность микромодуля должна быть больше трех кадров в секунду. С учетом опыта решения различных задач распознавания цветных изображений в качестве базового алгоритма, используемого для тестирования, был принят алгоритм, основанный на представлении содержания изображений с помощью матриц совместной встречаемости цветов [10]. Цветовое пространство редуцировалось путем использования старших битов R-, G- и B-компонентов цвета по схеме 3-3-3 (512 цветов): из восьми битов каждого цвета отбрасывались менее значимые младшие пять битов и оставались более значимые три старших бита. Таким образом, на каждый цвет приходилось по три бита или девять битов на редуцированный RGB-цвет. Девять битов обеспечивают представление $2^9 = 512$ цветов. Значит, размерность результирующей матрицы совместной встречаемости 512×512. После вычисления матрицы входного изображения все ее ненулевые элементы выписывались в вектор, называемый дескриптором изображения.

При любом варианте построения алгоритма распознавания базовой операцией является оценка близости текущего изображения к образцам классов, заданных в полетном задании. Указанная операция выполнялась путем сравнения дескрипторов изображений и вычисления меры близости $L1$ (расстояния городских кварталов) в пространстве признаков. Метрика $L1$

есть сумма длин проекций отрезка между двумя точками n -мерного вещественного векторного пространства на оси координат, которая вычисляется по формуле

$$L1 = \sum_{i=1}^n |p_i - q_i|, \quad (1)$$

где $p = (p_1, p_2, \dots, p_n)$ и $q = (q_1, q_2, \dots, q_n)$ – векторы n -мерного пространства, а также дескрипторы изображения, полученные из матрицы совместной встречаемости цветов.

Время выполнения перечисленных базовых операций для всех четырех платформ при условии реализации на Python версии 2.7.13 приведено в табл. 3. Следует отметить, что алгоритм, основанный на использовании матриц совместной встречаемости цветов, применялся здесь в качестве теста на быстродействие анализируемых микроплатформ. Однако это не означает, что только данный алгоритм можно применять в разрабатываемом микромодуле для решения задачи распознавания.

В табл. 3 и на рис. 3 показано, что время обработки данных сильно изменяется в зависимости от применяемых библиотек. Расширение для Python на языке программирования C ожидаемо сокращает время обработки на всех микроплатформах: примерно в два раза по сравнению со случаем использования библиотеки обработки изображений Scikit-image и в три-пять раз – библиотеки обработки массивов NumPy.

Таблица 3

Время вычисления матрицы совместной встречаемости и дескриптора изображения 320×240 пикселей для 512 цветов для четырех тестируемых устройств, мс

Программное обеспечение	Intel Compute Stick	Raspberry Pi 3	Raspberry Pi Zero	Xiaomi Redmi Note 4
Расширение языка C для Python	19 ± 2	44 ± 1	55 ± 3	15 ± 1
Библиотека обработки изображений Scikit-image	32 ± 1	87 ± 1	105 ± 4	39 ± 1
Библиотека обработки массивов NumPy	63 ± 2	182 ± 1	275 ± 10	51 ± 1
Двойной цикл (цикл в цикле)	1 888 ± 4	3 918 ± 6	14 131 ± 209	1 560 ± 11

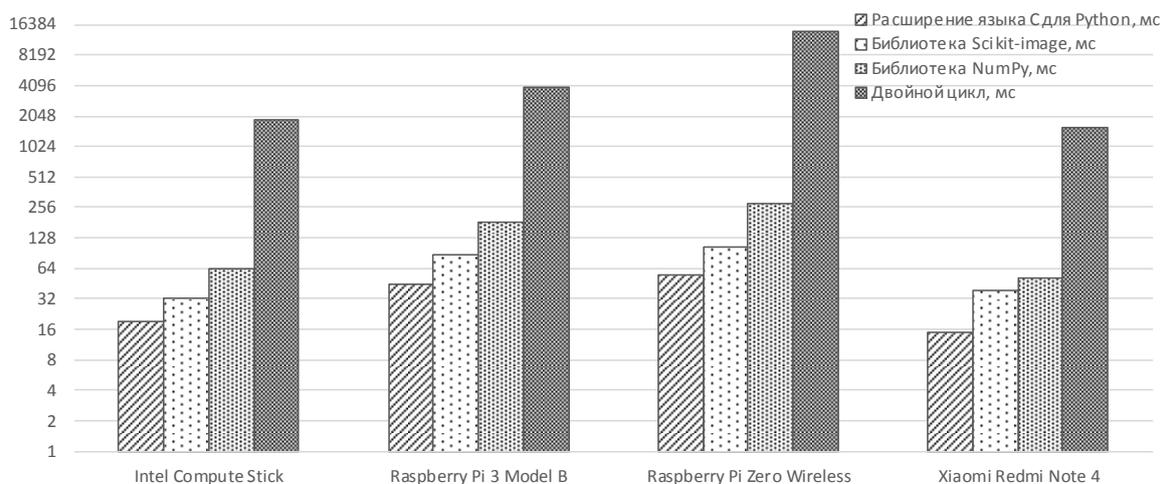
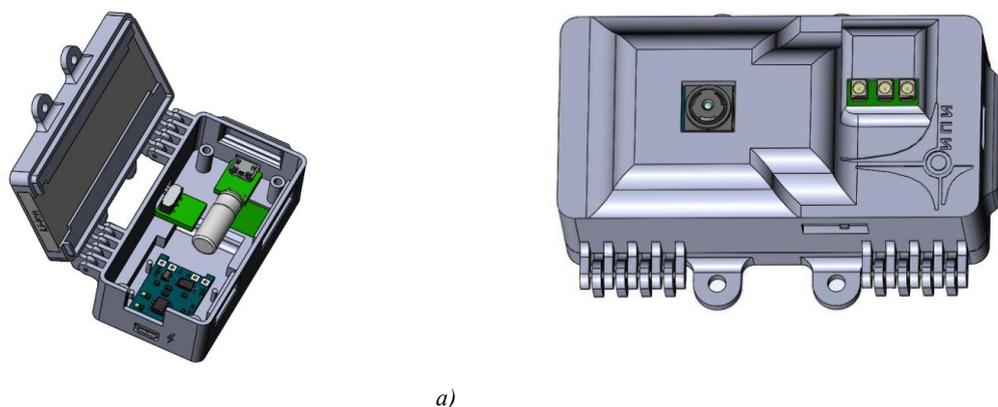


Рис. 3. Время вычисления матрицы совместной встречаемости и дескриптора изображения 320×240 пикселей для 512 цветов в логарифмическом масштабе

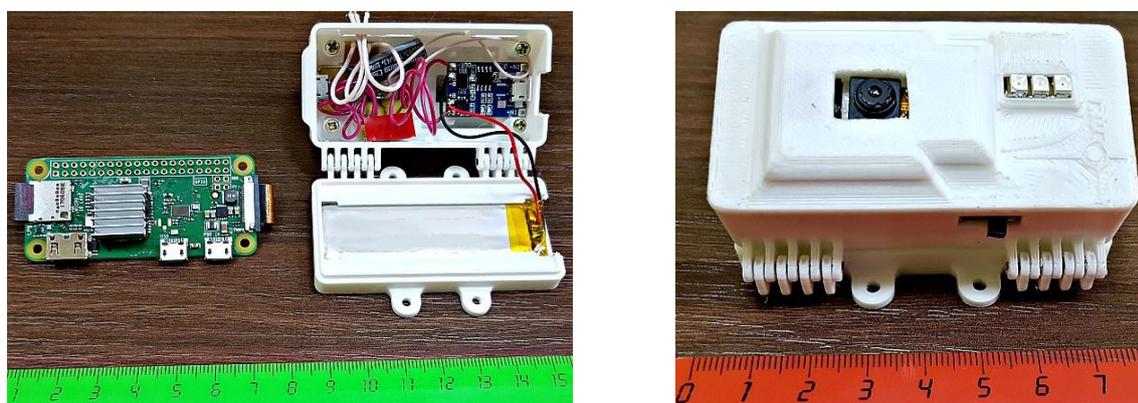
Следует отметить, что времена выполнения тестов на Intel Compute Stick и на смартфоне Xiaomi Redmi Note 4 отличаются лишь незначительно, хотя процессор смартфона по своим техническим спецификациям мощнее. Предположительно, это обусловлено тем, что на смартфоне интерпретатор Python работает через эмулятор и не поддерживается ОС Android напрямую. Измерения на смартфоне для ОС Android проводились с использованием Termux (эмулятора Linux для Android), часть из них была повторена на программе QPython для Android (URL: <http://www.qpython.com>). Как показали измерения, QPython работает медленнее, чем Termux.

5. Макет микро модуля. Полученные в результате сравнительного анализа четырех вычислительных микроплатформ данные позволили сделать вывод, что наилучшим микрокомпьютером для микро модуля оперативного распознавания изображений является модель Raspberry Pi Zero Wireless с ОС Raspbian. Она имеет минимальные энергопотребление, размеры, массу, стоимость и доступна для приобретения. Вычислительные эксперименты показали, что основной недостаток этой микроплатформы заключается в ее низкой вычислительной мощности и может быть преодолен за счет эффективной реализации алгоритмов распознавания на языке программирования C/C++.

В процессе реализации макета микро модуля был спроектирован и отпечатан на 3D-принтере корпус микро модуля, в котором размещаются микрокомпьютер Raspberry Pi Zero Wireless с камерой модели Raspberry Pi Zero V1.3 Mini Camera (суммарно 13 г), Li-Po-аккумулятор типа ROBITON емкостью 720 мАч (14 г), плата зарядки аккумуляторной батареи с конденсатором, сигнальные светодиоды и соединительные провода (рис. 4).



а)



б)

Рис. 4. 3D-модель (а) и макет (б) микро модуля

Макет микро модуля имеет размеры $7,3 \times 3,6 \times 2,5$ см и общую массу в собранном состоянии 57 г, что намного меньше ограничения в 300 г. Такие скромные массогабаритные характеристики позволят устанавливать микро модуль даже на легкие бюджетные дроны. Это создаст дополнительные конкурентные преимущества. Следует подчеркнуть, что в указанные 57 г входят суммарный вес автономного источника питания, обеспечивающего функционирование микро модуля на все время полета, и средства его зарядки, а также вес переключателя на внешний источник питания от сети в горячем режиме. При выполнении лабораторных испытаний с использованием дрона Parrot AR.Drone 2.0 текущий распознанный класс изображений в реальном времени выводится непосредственно на корпус макета микро модуля с помощью цветных светодиодных индикаторов, а копия текущего кадра изображения и результаты его распознавания передаются на ноутбук по Wi-Fi для упрощения процесса наблюдения и отладки ПО.

6. Состав и общие характеристики разрабатываемого программного обеспечения. Известно, что язык программирования Python обеспечивает быстрое прототипирование и реализацию ПО. Кроме того, он широко используется в качестве базового в большинстве программных фреймворков для сверточных нейронных сетей. Python не является идеальным инструментом разработки ПО встроенных систем, где обычно применяются C/C++ и Java. Поэтому прототипирование и разработка таких компонентов, как графические интерфейсы пользователя и другие вспомогательные функции, выполняются авторами на Python, а участки кода, критичные с точки зрения вычислительной эффективности, реализуются на C/C++ либо на расширениях C/C++ для Python.

ПО микро модуля состоит из наземной и полетной частей. Основной задачей наземной части является формирование полетного задания путем выбора образцов целевых классов изображений для традиционной схемы распознавания либо загрузка в микро модуль предобученной нейронной сети. Главная задача полетной части – оперативное распознавание изображений подстилающей поверхности, т. е. анализ текущего кадра видеопоследовательности и отнесение его к одному из классов, указанных в полетном задании. Данная задача решается двумя различными способами:

1. С помощью традиционной схемы распознавания. Традиционная схема включает выделение информативных признаков и применение одного из известных классификаторов типа kNN, SVM и др. В качестве признаков используются матрицы совместной встречаемости, подробно описанные в работе [10]. Высокая скорость распознавания обеспечивается за счет хранения только значимых (ненулевых) элементов матриц в виде списков типа «код элемента – значение», однопроходного алгоритма сравнения признаков и классификатора kNN с небольшими значениями k (от 1 до 10). В результате экспериментов было показано, что время распознавания одного кадра размером 320×240 пикселей составляет 14,5 мс в случае использования 256 цветов (схема 2-3-3 бит) и 55 мс в случае 512 цветов (разбитие цветового пространства по схеме 3-3-3).

2. Путем распознавания с использованием сверточной нейронной сети. Жесткие ограничения базового микро компьютера по скорости вычислений и размеру оперативной памяти не позволили применить ни одну из широко известных нейронных сетей – AlexNet или GoogLeNet. Поэтому было принято решение разработать и использовать собственную компактную архитектуру сверточной сети. Основные решения и особенности реализации ПО распознавания изображений на базе сверточных нейронных сетей кратко описаны ниже.

Процесс решения задачи распознавания состоит из двух этапов: тренировки (обучения) нейронной сети на выборке изображений целевых классов подстилающей поверхности и, собственно, оперативного распознавания изображений с камеры на борту дрона. Обучение сети относится к наземной части и осуществляется с помощью мощного графического процессора. Распознавание кадров изображений с камеры не требует мощного процессора, поскольку при этом выполняются в основном операции свертки изображения с фильтрами, полученными на этапе обучения сети.

С учетом рассматриваемой задачи распознавания ограниченного количества классов, а также малой вычислительной мощности микро компьютеров была реализована упрощенная архитектура сверточной нейронной сети.

Сеть состоит из двух блоков, расположенных последовательно и содержащих сверточные слои. В первом блоке сверточный слой содержит 12 ядер свертки размером 3×3 пиксела, во втором блоке – 24 ядра. После каждого сверточного слоя расположен слой с активационной функцией ReLU: $f(x) = \max(0, x)$. За активационными слоями следуют субдискретизирующие слои, выполняющие нелинейное уплотнение карты признаков.

После упомянутых выше блоков расположена полносвязная сеть, выходом которой является еще один активационный слой с функцией Softmax.

Выходом сверточной сети для входного изображения является вектор размером пять элементов, что соответствует количеству распознаваемых классов подстилающей поверхности. Каждый элемент вектора представляет собой вероятность принадлежности входного изображения к соответствующему классу. Изображение относится к некоторому классу, если указанная вероятность превышает порог 0,99.

Реализация алгоритмов свертки выполнена на языке программирования C/C++ с последующей «ручной» оптимизацией кода, в результате чего время распознавания одного кадра составило 80 мс, т. е. 12 кадров в секунду. Однако эксперименты показали, что из-за ограниченности архитектуры нейросети и влияния всевозможных шумовых факторов (вариабельности освещения, изменения высоты полета дрона, вибрации камеры, большой вариабельности изображений «нецелевых» классов поверхностей) точность распознавания варьировала в широких пределах (от 75 до 95 %), что не соответствовало требованиям.

Решение было найдено в использовании каскада из двух нейросетей. Обе сети имели одну и ту же архитектуру, описанную выше. При этом первая из них обучалась разделять входной поток на изображения, относящиеся к группе целевых классов (т. е. к любому из них – обучение на объединении классов изображений) и «фону», а вторая сеть разбивала первый, объединенный, класс на целевые классы. В результате удалось достигнуть качества распознавания 90 % и более, хотя время распознавания возросло вдвое, т. е. до 160 мс. Более детальное рассмотрение вопроса качества распознавания изображений является предметом отдельной публикации.

7. Краткий SWOT-анализ тестируемых микроплатформ

Intel Compute Stick STK1AW32SC

Сильные стороны:

- почти в три раза быстрее, чем Raspberry Pi 3;
- предустановленная ОС Windows 10 Home для ряда пользователей может оказаться более простой и привычной, чем Linux или Android.

Слабые стороны:

- большое энергопотребление и, как следствие, потребность в более емком и тяжелом аккумуляторе;
- наличие вентилятора охлаждения и отверстий в корпусе, что в случае беспилотных летательных аппаратов может привести к попаданию влаги внутрь устройства и выходу его из строя;
- ограничения в настройках ОС Windows 10 Home Edition;
- превышение требования технического задания 2 (см. разд. 1) по длине (см. табл. 1);
- превышение требования технического задания 3 по энергопотреблению (см. табл. 2).

Возможности:

- мощный процессор позволяет использовать более сложные алгоритмы обработки данных;
- есть графическая карта Intel HD Graphics, которую потенциально можно использовать для реализации алгоритмов на основе сверточных нейронных сетей.

Риски:

- малый объем встроенной энергонезависимой памяти (32 Гб) создает риск переполнения хранилища;
- сложно подобрать небольшой, легкий, но энергоемкий аккумулятор.

Raspberry Pi 3 Model B

Сильные стороны:

- самый быстродействующий микрокомпьютер среди моделей ряда Raspberry Pi;
- обладает большим количеством портов для подключения внешних устройств.

Слабые стороны:

– достаточно высокое энергопотребление требует питания от более емкой и тяжелой батареи, что добавляет массу разрабатываемому мобильному устройству.

Возможности:

– использование более сложных алгоритмов обработки данных по сравнению с Raspberry Pi Zero;

– использование сверточных нейронных сетей для распознавания изображений.

Риски:

– линейный размер по ширине 5,7 см (см. табл. 1) вплотную подходит к максимальному размеру в 6 см (см. разд. 1, требование технического задания 2);

– как показывает практика, при высоких нагрузках на процессор и (или) температурах окружающего воздуха есть риск перегрева процессора и отключения системы;

– существует риск превышения требования технического задания 3 по энергопотреблению (см. табл. 2).

Raspberry Pi Zero Wireless

Сильные стороны:

– низкая цена (10–15 долл.);

– чрезвычайно малые массогабаритные параметры (9 г, 6,5 см в максимальном измерении);

– низкое энергопотребление.

Слабые стороны:

– малое количество портов для подключения внешних устройств;

– относительно низкая производительность.

Возможности:

– разработан специально для встраиваемых мобильных приложений.

Риски:

– есть риск того, что будет слишком медленно обрабатывать данные, особенно если использовать ресурсоемкие алгоритмы и плохо проработанную архитектуру ПО.

Xiaomi Redmi Note 4

Сильные стороны:

– наличие в комплекте всех необходимых элементов оборудования для разработки систем предварительного распознавания и отбора изображений;

– самый мощный центральный процессор из четырех протестированных микрокомпьютеров;

– самое низкое энергопотребление в режиме ожидания;

– относительно малое энергопотребление при больших вычислительных нагрузках;

– наличие металлического корпуса, защищающего от механических и электромагнитных воздействий;

– достаточно малые размеры, особенно по толщине.

Слабые стороны:

– сложность или полная невозможность добавить либо удалить компоненты оборудования устройства;

– слабая поддержка языка программирования Python для ОС Android;

– относительно большой вес (от 100 г и выше);

– превышение требования технического задания 2 по длине (см. табл. 1).

Возможности:

– разработка ПО на различных языках программирования, включая C/C++, Java, Kotlin, .NET C# и др.;

– устройство снабжено большим набором встроенных датчиков, которые можно задействовать при разработке мобильного решения.

Риски:

– разработка ПО под MIUI, основанного на Android, возможно, будет как-то отличаться от «классической» разработки под Android;

– у исполнителей отсутствует опыт разработки ПО на ОС Android.

Заключение. Результаты сравнительного анализа четырех вычислительных микроплатформ позволили сделать вывод, что наилучшим микрокомпьютером для микромодуля оперативного распознавания является модель Raspberry Pi Zero Wireless с ОС Raspbian. Эта модель имеет минимальные энергопотребление, размеры и массу, но при этом обладает достаточной вычислительной мощностью для решения задачи распознавания и классификации подстилающих поверхностей.

Сравнение вычислительных микроплатформ осуществлялось с помощью SWOT-анализа, согласно которому фактически только Raspberry Pi Zero Wireless удовлетворяет всем требованиям технического задания на разработку микромодуля. Также SWOT-анализ выявил низкую производительность Pi Zero. В результате было принято решение обойти данную проблему путем разработки участков кода, критичных с точки зрения вычислительной эффективности, на языке программирования C/C++.

Полученный макет микромодуля соответствует требованиям технического задания:

- 1) масса макета (включая камеру и интерфейсное оборудование) равна 57 г против максимальных 300 г;
- 2) габариты макета (длина, ширина, высота) 7,3×3,6×2,5 см против максимальных 12×6×4 см;
- 3) энергопотребление макета в нагруженном состоянии 1,75 Вт против максимальных 5 Вт;
- 4) время распознавания одного кадра RGB-изображения размером 320×240 пикселей с помощью матриц совместной встречаемости равно 55 мс и с помощью сверточной нейронной сети – 160 мс против максимальных 300 мс;
- 5) ПО макета распознает более десяти типов изображений подстилающей поверхности, в том числе четыре типа одновременно;
- 6) качество распознавания заданных типов изображений подстилающей поверхности (в зависимости от типа) составляет более 90 % для сверточной нейронной сети против минимальных 85–95 %;
- 7) предполагается, что при единичном производстве себестоимость одного микромодуля не превышает 50 долл. против максимальных 500 долл.

Существуют микрокомпьютеры с более быстрыми центральными и графическими процессорами, большим объемом памяти и количеством стандартных интерфейсов, чем у Raspberry Pi Zero Wireless, но они не удовлетворяют имеющимся техническим требованиям к разработке микромодуля. Тем не менее следует отметить преимущества смартфонов, которые, по мнению авторов, являются лучшими микрокомпьютерами, воплощают в себе все самые последние достижения в области встраиваемых систем и при этом коммерчески доступны. В соответствии с результатами SWOT-анализа смартфоны обеспечивают наилучшее соотношение между внутренними и внешними факторами: шесть сильных сторон у смартфона против двух-трех сильных сторон у остальных платформ. Совокупность сильных сторон и возможностей создает большое конкурентное преимущество для смартфонов перед другими микроплатформами. Основным недостатком смартфонов в данном случае является большой вес, что не позволяет использовать их для бюджетных дронов с низкой грузоподъемностью.

Работа была выполнена при финансовой поддержке проекта № 3.2.4.1 программы Союзного государства «Технология-СГ».

Список использованных источников

1. Sammut-Bonnici, T. SWOT analysis / T. Sammut-Bonnici, D. Galea // Wiley Encyclopedia of Management. – Vol. 12 : Strategic Management. – Wiley Online Library, 2014. – P. 1–8.
2. PhoneSat in-flight experience results / A. G. Salas [et al.] // Proc. of the Small Satellites and Services Symp. – Porto Petro, Majorca, 2014 [Electronic resource]. – 2014. – 19 p. – Mode of access: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20140008616.pdf>. – Date of access: 04.06.2018.
3. Дашевский, В. П. Обзор возможностей бортовых вычислителей на основе SMARC-модулей для робототехнических комплексов / В. П. Дашевский, М. М. Бизин // Докл. Томского гос. ун-та систем управления и радиоэлектроники. – Томск, 2015. – С. 91–97.

4. Intel® Compute Stick STK1AW32SC STK1A32SC. Technical Product Specification [Electronic resource] : Intel Corporation. – 2015. – Mode of access: https://www.intel.com/content/dam/support/us/en/documents/boardsandkits/computestick/STK1AW32SC_STK1A32SC_TechProdSpec.pdf. – Date of access: 04.06.2018.
5. Raspberry Pi Comparison Chart [Electronic resource] : ModMyPi LTD. – 2017. – Mode of access: <https://www.modmypi.com/download/raspberry-pi-comparison-chart.pdf>. – Date of access: 04.06.2018.
6. Raspberry Pi Documentation [Electronic resource] : Raspberry Pi Foundation with community contributions. – 2012–2018. – Mode of access: <https://www.raspberrypi.org/documentation>. – Date of access: 04.06.2018.
7. Mi Phone User Guide [Electronic resource] : Xiaomi Communications Co., Ltd. – 2016. – 37 p. – Mode of access: http://files.xiaomi-mi.com/files/Mi_Phone_User_Guide/Mi_Phone_User_Guide_eng.pdf. – Date of access: 04.06.2018.
8. Linuxium. Ubuntu flavoured 16.04 ISOs for Atom based Intel Compute Sticks [Electronic resource] : Customizing Ubuntu ISOs: Documentation and examples of how to use 'isorespin.sh'. – 2016. – Mode of access: <https://linuxiumcomau.blogspot.com/2016/05/ubuntu-flavoured-1604-isos-for-atom.html>. – Date of access: 04.06.2018.
9. Harrington, W. Learning Raspbian / W. Harrington. – Packt Publishing Ltd., 2015. – 154 p.
10. Kovalev, V. Color co-occurrence descriptors for querying-by-example / V. Kovalev, S. Volmer // Intern. Conf. on Multimedia Modeling. – Lausanne, Switzerland, 1998. – P. 32–38.

References

1. Sammut-Bonnici T., Galea D. SWOT analysis. *Wiley Encyclopedia of Management*, vol. 12 : Strategic Management, Wiley Online Library, 2014, pp. 1–8. doi: 10.1002/9781118785317.weom120103
2. Salas A. G., Attai W., Oyadomari K. Y., Priscal C., Schimmin R. S., Gazulla O. T., Wolfe J. L. PhoneSat in-flight experience results. *Proceedings of the Small Satellites and Services Symposium*. Porto Petro, Majorca, 2014, 19 p. Available at: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20140008616.pdf> (accessed 04.06.2018).
3. Dashevskij V. P., Bizin M. M. Obzor vozmozhnostej bortovyh vychislitelej na osnove SMARC-modulej dlja robototekhnicheskikh kompleksov [Overview of the capabilities of on-board calculators based on SMARC-modules for robotic complexes]. *Doklady Tomskogo gosudarstvennogo universiteta sistem upravlenija i radioelektroniki [Reports of Tomsk State University of Control Systems and Radioelectronics]*. Tomsk, 2015, pp. 91–97 (in Russian).
4. Intel® Compute Stick STK1AW32SC STK1A32SC. Technical Product Specification. Intel Corporation, 2015. Available at: https://www.intel.com/content/dam/support/us/en/documents/boardsandkits/computestick/STK1AW32SC_STK1A32SC_TechProdSpec.pdf (accessed 04.06.2018).
5. Raspberry Pi Comparison Chart. ModMyPi LTD, 2017. Available at: <https://www.modmypi.com/download/raspberry-pi-comparison-chart.pdf> (accessed 04.06.2018).
6. Raspberry Pi Documentation. Raspberry Pi Foundation with community contributions, 2012–2018. Available at: <https://www.raspberrypi.org/documentation> (accessed 04.06.2018).
7. Mi Phone User Guide. Xiaomi Communications Co., Ltd., 2016, 37 p. Available at: http://files.xiaomi-mi.com/files/Mi_Phone_User_Guide/Mi_Phone_User_Guide_eng.pdf (accessed 04.06.2018).
8. Linuxium. Ubuntu flavoured 16.04 ISOs for Atom based Intel Compute Sticks. Customizing Ubuntu ISOs: Documentation and examples of how to use 'isorespin.sh', 2016. Available at: <https://linuxiumcomau.blogspot.com/by/2016/05/ubuntu-flavoured-1604-isos-for-atom.html> (accessed 04.06.2018).
9. Harrington W. *Learning Raspbian*. Packt Publishing Ltd., 2015, 154 p.
10. Kovalev V., Volmer S. Color co-occurrence descriptors for querying-by-example. *International Conference on Multimedia Modeling*. Lausanne, Switzerland, 1998, pp. 32–38.

Информация об авторах

Ковалев Василий Алексеевич – кандидат технических наук, заведующий лабораторией анализа биомедицинских изображений, Объединенный институт проблем информатики Национальной академии наук Беларуси (ул. Сурганова, 6, Минск, Республика Беларусь). E-mail: vassili.kovalev@gmail.com

Павленко Дмитрий Анатольевич – инженер-программист, Объединенный институт проблем информатики Национальной академии наук Беларуси (ул. Сурганова, 6, Минск, Республика Беларусь). E-mail: dmitri.pavlenko@gmail.com

Снежко Эдуард Витальевич – кандидат технических наук, ведущий научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси (ул. Сурганова, 6, Минск, Республика Беларусь). E-mail: eduard.snezhko@gmail.com

Information about the authors

Vassili A. Kovalev – Ph. D. (Engineering), Head of the laboratory of Biomedical Images Analysis, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganova Str., 220012, Minsk, Republic of Belarus). E-mail: vassili.kovalev@gmail.com

Dzmitry A. Paulenka – Software Engineer, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganova Str., 220012, Minsk, Republic of Belarus). E-mail: dmitri.pavlenko@gmail.com

Eduard V. Snezhko – Ph. D. (Engineering), Leading Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganova Str., 220012, Minsk, Republic of Belarus). E-mail: eduard.snezhko@gmail.com

Левчук Виталий Андреевич – научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси (ул. Сурганова, 6, Минск, Республика Беларусь). E-mail: vitali.liauchuk@gmail.com

Калиновский Александр Александрович – научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси (ул. Сурганова, 6, Минск, Республика Беларусь). E-mail: gakararak@gmail.com

Vitali A. Liauchuk – Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Sarganova Str., 220012, Minsk, Republic of Belarus). E-mail: vitali.liauchuk@gmail.com

Alexander A. Kalinovski – Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Sarganova Str., 220012, Minsk, Republic of Belarus). E-mail: gakararak@gmail.com

ISSN 1816-0301 (print)
УДК 004.31

Поступила в редакцию 11.05.2018
Received 11.05.2018

Е. В. Рыбенков, Н. А. Петровский

*Белорусский государственный университет
информатики и радиоэлектроники, Минск, Беларусь*

СИНТЕЗ FPGA-АРХИТЕКТУР БАНКОВ ФИЛЬТРОВ НА ОСНОВЕ БЛОЧНОЙ ЛЕСТНИЧНОЙ ФАКТОРИЗАЦИИ В АЛГЕБРЕ КВАТЕРНИОНОВ (ЧАСТЬ 2)

Аннотация. В настоящее время методологии проектирования систем на кристалле основываются на высокопараметризованных IP-компонентах (IP – intellectual property), которые для конкретного целевого приложения обеспечивают широкий диапазон регулировки затрат ресурсов, форматов данных арифметики с фиксированной запятой и производительности системы. В статье предложена гибкая технология быстрого прототипирования архитектур процессоров целочисленных обратимых параунитарных банков фильтров в алгебре кватернионов (Int-Q-ПУБФ) на основе FPGA, в основу которой положен Q-MUL IP-компонент оператора умножения кватернионов на распределенной арифметике на сумматорах. Осуществлена реализация Int-Q-ПУБФ на FPGA Xilinx Zynq 7010, при этом восьмиканальный 8x24 Int-Q-ПУБФ имеет перфективную реконструкцию входных данных для заданного формата фиксированной запятой, малые аппаратные затраты и небольшую задержку конвейера по сравнению с известными решениями на CORDIC-процессорах и распределенной арифметике на памяти.

Ключевые слова: схема lossless-to-lossy, кватернионы, компрессия изображений, FPGA, банк фильтров, блочная лестничная факторизация

Для цитирования. Рыбенков, Е. В. Синтез FPGA-архитектур банков фильтров на основе блочной лестничной факторизации в алгебре кватернионов (часть 2) / Е. В. Рыбенков, Н. А. Петровский // Информатика. – 2018. – Т. 15, № 3. – С. 22–31.

E. V. Rybenkov, N. A. Petrovsky

Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus

SYNTHESIS OF FPGA ARCHITECTURES OF BLOCK LIFTING-BASED FILTER BANKS IN QUATERNION ALGEBRA (PART 2)

Abstract. Nowadays the methodology for designing systems on a chip is based on highly parameterized IP (intellectual property) components which provide a wide range of adjustment of resources, fixed point arithmetic data formats, and system performance for a specific application. The article describes a flexible technology for rapid prototyping of processor architectures for integer, invertible, paraunitary filter banks in quaternion algebra (Int-Q-PUFB) based on the FPGA Q-MUL IP-component as multiplication operator for quaternions on distributed arithmetic on adders. Implementation of Int-Q-PUFB on FPGA Xilinx Zynq 7010, with 8-channel 8x24 Int-Q-PUFB has a perfect reconstruction property of the input data for a fixed point format, small hardware resource utilization and a slight delay in the pipeline compared to the known solutions for CORDIC-processors and distributed arithmetic on the memory.

Keywords: lossless-to-lossy, quaternions, image compression, FPGA, filter bank, block-lifting factorization

For citation. Rybenkov E. V., Petrovsky N. A. Synthesis of FPGA architectures of block lifting-based filter banks in quaternion algebra (part 2). *Informatics*, 2018, vol. 15, no. 3, pp. 22–31 (in Russian).

Введение. Как было показано в [1], применение блочной лестничной параметризации оператора умножения кватернионов $M^+(Q)$ является более предпочтительным по сравнению с прямой лестничной параметризацией. Это дает возможность уменьшать количество операций округления $\approx 2,67$ раза, что положительно сказывается на этапе кодирования изображений, а реализация ступеней ULV-разложения $M^+(Q)$ на распределенной арифметике на сумматорах (DA-ADDER) [2] позволит уменьшить аппаратные ресурсы FPGA и повысить производительность операции Q-MUL по сравнению с распределенной арифметикой на памяти [3] и CORDIC-процессором [4].

В распределенной арифметике на сумматорах DA-ADDER [5] постоянные коэффициенты A_k вектора \mathbf{A} представляются в двоичной нормированной форме ($|A_k| < 1$), а непеременные компоненты x_k входного вектора \mathbf{X} – в форме DA-ROM:

$$y = \sum_{k=1}^L \left[\sum_{n=0}^{B-1} A_{kn} 2^{-n} \right] x_k = \sum_{n=0}^{B-1} \left[\sum_{k=1}^L A_{kn} \cdot x_k \right] 2^{-n},$$

где A_{kn} – биты (0 или 1), B – количество разрядов в каждом элементе входного слова данных. Поскольку A_{kn} может принимать только значение 0 или 1, то терм $\sum A_{kn} \cdot x_k = S_n$ представляет собой комбинацию x_k . В DA-ADDER термы $\sum A_{kn} \cdot x_k$ реализуются в виде дерева сумматоров. На каждом цикле работы устройства данные подаются на входы дерева сумматоров для вычисления частичных результатов ($\sum A_{kn} \cdot x_k$), которые складываются или вычитаются со значением регистра-накопителя. В схеме DA-ADDER связи между сумматорами фиксируются жестко, в результате чего постоянные коэффициенты не могут быть изменены без перестройки структуры дерева сумматоров.

Расчет обратимого Int-Q-ПУБФ на основе Q-MUL IP-компонента. FPGA-реализация Int-Q-ПУБФ предполагает минимизацию латентности Q-MUL IP-компонента, коэффициенты блочной лестничной факторизации $\mathbf{F}(\tilde{Q})$, $\mathbf{G}(\tilde{Q})$, $\mathbf{H}(\tilde{Q})$ которого должны выбираться с минимальным количеством единиц в двоичном коде их представления в распределенной арифметике на сумматорах DA-ADDER. Синтез Int-Q-ПУБФ строится на основе численных методов с ограничениями на аппаратную реализацию: длину слова (B) и число бит, равных единице (K) в двоичном коде лестничных коэффициентов $\mathbf{F}(\tilde{Q})$, $\mathbf{G}(\tilde{Q})$, $\mathbf{H}(\tilde{Q})$. Эффективность Int-Q-ПУБФ оценивается по следующим показателям:

1) частотной избирательности – максимальному ослаблению в полосе непропускания каналов банка фильтров ε_{SBE} [6]:

$$\varepsilon_{SBE} = \sum_{k=0}^{M-1} \int_{\omega \in \Omega_k} |H_k(e^{j\omega})|^2 d\omega,$$

где Ω_k – полоса задержки k -го фильтра банка фильтров;

2) коэффициенту эффективности кодирования (Coding Gain, CG) [6]:

$$CG = 10 \lg \frac{\frac{1}{M} \sum_{k=0}^{M-1} \sigma_{xk}^2}{\left(\prod_{k=0}^{M-1} \sigma_{xk}^2 \right)^{\frac{1}{M}}},$$

где σ_{xk}^2 – дисперсия сигнала в k -м канале банка фильтров. Показатель CG характеризует степень компактности энергии в субполосах Int-Q-ПУБФ. Значения дисперсии σ_{xk}^2 для M -канального банка фильтров определяются диагональными элементами автокорреляционной матрицы \mathbf{R}_{yy} выходного сигнала $y(n)$ [6]:

$$\sigma_{xk}^2 = [\mathbf{R}_{yy}]_{kk}, \mathbf{R}_{yy} = \mathbf{H}\mathbf{R}_{xx}\mathbf{H}^T,$$

$$[\mathbf{H}]_{kn} = h_k(L-1-n), k = (0, \dots, M-1), n = 0, \dots, L-1,$$

где \mathbf{R}_{yy} – автокорреляционная матрица выходного сигнала $y(n)$; \mathbf{R}_{xx} – автокорреляционная матрица входного сигнала $x(n)$; \mathbf{H} – ортогональные матрицы, которые определяются передаточными функциями банка фильтров $H_k(z) (k=1, \dots, M)$. Входной сигнал $x(n)$ с автокорреляционной матрицей \mathbf{R}_{xx} формируется на основе авторегрессионной модели первого порядка AR(1) с единичной дисперсией и коэффициентом корреляции 0,95. Данная модель сигнала соответствует обработке естественных изображений;

3) ошибке (ε_q) реконструкции сигнала банком синтеза, обусловленной квантованием лестничных коэффициентов $\mathbf{F}(\tilde{Q})$, $\mathbf{G}(\tilde{Q})$, $\mathbf{H}(\tilde{Q})$ в Int-Q-ПУБФ:

$$\varepsilon_q = \max(|\tilde{x}(n) - x(n)|),$$

где $\tilde{x}(n)$ – данные на выходе Int-Q-ПУБФ синтеза; $x(n)$ – данные на входе Int-Q-ПУБФ анализа.

Полярная форма представления кватерниона ($Q = |Q| \cdot e^{i\varphi} \cdot e^{j\psi} \cdot e^{k\chi}$ для $-\pi \leq \varphi \leq \pi$, $-\frac{\pi}{2} \leq \psi \leq \frac{\pi}{2}$, $-\frac{\pi}{2} \leq \chi \leq \frac{\pi}{2}$) позволяет сократить число степеней свободы до трех и легко параметризовать кватернионы с нормой $|Q|$, равной единице. Условие $|Q|=1$ является обязательным для использования блочной лестничной параметризации умножителя кватернионов. Минимальное значение K дает возможность эффективно применить параллельную распределенную арифметику на сумматорах.

Задача синтеза может быть сформулирована как задача поиска условного экстремума: найти такую точку $x^* = [\varphi_1^*, \psi_1^*, \chi_1^*, \dots, \varphi_i^*, \psi_i^*, \chi_i^*]^T \in X, i = (0, \dots, 2N-1)$, в которой целевая функция $f(x^*) = \min_{x^* \in X} f(x)$, $f(x) = -CG(x)$ имеет локальный минимум при заданных ограничениях $X = \{x | g_j(x) \leq 0, j = (1, \dots, p)\}$:

$$g_1(x) = \varepsilon_{SBE}(x) - \varepsilon_{\min SBE} \leq 0; \quad g_2(x) = \varepsilon_q(x) - \varepsilon_{\max q} \leq 0,$$

где $\varepsilon_{\min SBE}$ – минимально допустимое ослабление в полосе пропускания во всех каналах; $\varepsilon_{\max q}$ – максимально допустимый уровень ошибки реконструкции сигнала.

В качестве метода оптимизации выбран метод множителей Лагранжа [7]. Функция Лагранжа имеет вид

$$L(x, \mu^k, r^k) = f(x) + P(x, \mu^k, r^k),$$

$$P(x, \mu^k, r^k) = \frac{1}{2r^k} \sum_{j=1}^p \left\{ \left[\max(0, \mu_j^k + r^k g_j(x)) \right]^2 - (\mu_j^k)^2 \right\},$$

где $P(x, \mu^k, r^k)$ – штрафная функция, $\mu^k = (\mu_1^k, \dots, \mu_p^k)$ – вектор множителей Лагранжа, r^k – параметр штрафа, k – номер итерации.

Алгоритм синтеза лестничных коэффициентов восьмиканального Int-Q-ПУБФ:

1. Задать значения следующих параметров: начальной точки $x = x_0$, штрафа $r = [0, 1; 1]$, шага приращения параметра штрафа $C = [4; 10]$, погрешности решения $\varepsilon > 0$ (остановки алгоритма), вектора множителей Лагранжа μ , разрядности слова B лестничных коэффициентов, порядка факторизации N , $\varepsilon_{\min SBE}$, $\varepsilon_{\max q}$, K , $k = 0$.

2. Составить модифицированную функцию Лагранжа $L(x, \mu^k, r^k)$.

3. Найти точку $x^*(\mu^k, r^k)$ безусловного минимума функции $L(x, \mu^k, r^k)$ по x : $L(x^*, \mu^k, r^k) = \min_{x \in \mathbb{R}^n} L(x, \mu^k, r^k)$, при этом для определения параметров $CG(x)$, $\varepsilon_{SBE}(x)$, $\varepsilon_q(x)$ и функций ограничений $g_j(x)$ выполнить:

3.1. преобразование вектора $x = [\varphi_1, \psi_1, \chi_1, \dots, \varphi_i, \psi_i, \chi_i]^T$ в кватернионы P_i и Q_i согласно полярной форме представления;

3.2. вычисление кватерниона Q_{N-1} согласно формуле (19) из [1];

3.3. вычисление значения коэффициентов лестничных структур $\mathbf{F}_i(\tilde{Q})$, $\mathbf{G}_i(\tilde{Q})$, $\mathbf{H}_i(\tilde{Q})$ и матрицы перестановок \mathbf{P}_{pre} , \mathbf{P}_{post} для каждого кватерниона P_i и Q_i ;

3.4. вычисление для вектора параметров $x = [\varphi_1, \psi_1, \chi_1, \dots, \varphi_i, \psi_i, \chi_i]^T$ выхода $y(n)$ Int-Q-ПУБФ системы анализа-синтеза;

3.5. определение параметров $CG(x)$, $\varepsilon_{SBE}(x)$, $\varepsilon_q(x)$.

4. Вычислить значение функции $P(x^*, \mu^k, r^k)$:

если $|P(x^*(\mu^k, r^k), \mu^k, r^k)| \leq \varepsilon$, то

вернуть минимум функции Лагранжа $x^*(\mu^k, r^k)$ и перейти к п. 6;

иначе

пересчитать параметры штрафа r^{k+1} и множители μ_j^{k+1} для ограничений-неравенств: $r^{k+1} = C \cdot r^k$; $\mu_j^{k+1} = \max\{0, \mu_j^k + r^k g_j(x^*(\mu^k, r^k))\}$.

5. Положить $x^{k+1} = x^*(\mu^k, r^k)$, $k = k + 1$, и перейти к п. 2.

6. Выполнить преобразования целевых умножителей кватернионов к виду $\mathbf{M}^+(\cdot)$ в соответствии с табл. 1 из [7].

7. Для всех кватернионов определить компоненты векторов знаков \mathbf{SH} , \mathbf{SG} , \mathbf{SF} и изменить знаки лестничных коэффициентов $\mathbf{H}(\tilde{Q})$, $\mathbf{G}(\tilde{Q})$, $\mathbf{F}(\tilde{Q})$ в соответствии с табл. 2 из [1].

8. Конец алгоритма.

Начальные точки x_0 выбираются случайным образом, поскольку отсутствует информация о расположении глобального минимума функции $L(x, \mu^k, r^k)$. Синтез коэффициентов рекомендуется начинать при $B = 64$ и $K = 64$, после чего полученное решение x^* использовать как начальную точку x_0 , например, для $B = 12$ и $K = 3$. Иными словами, минимизация функции $L(x, \mu^k, r^k)$ в арифметике с «грубо» квантованными параметрами $B = 12$ и $K = 3$ приводит к остановке алгоритма синтеза Int-Q-ПУБФ через небольшое число итераций и в результате локальный минимум функции не достигается.

Значения коэффициентов лестничных структур вычисляются и квантуются для 24 возможных комбинаций кватерниона \tilde{Q} и матриц \mathbf{P}_{post} , \mathbf{P}_{pre} , из которых выбирается одна комбинация, удовлетворяющая следующим условиям: значения коэффициентов должны располагаться в диапазоне от -1 до 1 ; ошибка факторизации матрицы $\mathbf{M}^\pm(Q)$ в виде лестничной структуры должна быть минимальной; число бит K , равных единице, в лестничных коэффициентах, представленных в прямом двоичном коде, должно быть минимальным.

В табл. 1 приведены основные характеристики синтезированного восьмиканального Int-Q-ПУБФ анализа для различных ограничений на аппаратную реализацию. Анализ табл. 3 показал, что сбалансированное решение Int-Q-ПУБФ получается для схемотехнических ограничений $B=12$ и $K=3$. Квантование коэффициентов не повлияло на свойство полного восстановления (ошибка восстановления ε_q практически равна нулю). При этом Int-Q-ПУБФ характеризуется высоким значением показателя эффективности кодирования $CG = 9,49$ дБ и частотной избирательности $\varepsilon_{SBE} = -21,3$ дБ, а степень ослабления постоянной составляющей сигнала DC Att. равна -49 дБ.

Таблица 1

Характеристики восьмиканального Int-Q-ПУБФ для различных ограничений на аппаратную реализацию

Арифметика		ε_{SBE} , дБ	ε_q	CG, дБ	DC Att., дБ
B	K				
64	64	-20,6	$6,7 \cdot 10^{-16}$	9,37	-300
16	16	-20,6	$6,7 \cdot 10^{-16}$	9,37	-95,3
12	12	-20,6	$6,7 \cdot 10^{-16}$	9,37	-65,11
8	8	-20,2	$6,7 \cdot 10^{-16}$	9,39	-39,83
16	3	-21,3	$6,7 \cdot 10^{-16}$	9,49	-49,22
12	3	-21,3	$6,7 \cdot 10^{-16}$	9,49	-49,22

В табл. 2 приведены настроечные параметры (коэффициенты блочной лестничной параметризации \mathbf{F} , \mathbf{G} , \mathbf{H} , векторы знаков \mathbf{SF} , \mathbf{SG} , \mathbf{SH} и матрицы коммутации \mathbf{B}_{pre} , \mathbf{B}_{post}) Q-MUL IP-компонентов операторов $\mathbf{M}^\pm(\cdot)$ умножения кватернионов для данного восьмиканального 8×24 Int-Q-ПУБФ анализа с линейной фазово-частотной характеристикой (ФЧХ) и попарно зеркальной симметрией частотных характеристик канальных фильтров.

Таблица 2

Настроечные параметры восьмиканального 8×24 Int-Q-ПУБФ ($B = 12, K = 3$)

$\mathbf{M}^\pm(\cdot)$	$f_{11} = f_{22}$, $f_{12} = -f_{21}$	SF_1 , SF_2	$g_{11} = g_{22}$, $g_{12} = -g_{21}$	SG_1 , SG_2	$h_{11} = h_{22}$, $h_{12} = -h_{21}$	SH_1 , SH_2	\mathbf{B}_{pre} , \mathbf{B}_{post}
$\mathbf{M}^+(P_1)$	$+(2^{-5} + 2^{-4} + 2^{-3})$	+	$+(2^{-10} + 2^{-9} + 2^{-2})$	+	$-(2^{-4} + 2^{-3} + 2^{-1})$	-	[3142]
	$-(2^{-3} + 2^{-2} + 2^{-1})$	-	$+(2^{-8} + 2^{-5} + 2^{-3})$	-	$-(2^{-6} + 2^{-4} + 2^{-1})$	-	[1324]
$\mathbf{M}^-(P_2)$	$-(2^{-8} + 2^{-6})$	-	$+(2^{-5} + 2^{-4} + 2^{-3})$	+	$-(2^{-8} + 2^{-5} + 2^{-2})$	-	[4132]
	$-(2^{-7} + 2^{-5} + 2^{-2})$	+	$+(2^{-5} + 2^{-4} + 2^{-3})$	-	$+(2^{-8} + 2^{-6} + 2^{-5})$	-	[1342]
$\mathbf{M}^+(P_3)$	$+(2^{-4} + 2^{-3} + 2^{-2})$	-	$-(2^{-7} + 2^{-4} + 2^{-3})$	-	$-(2^{-5} + 2^{-2} + 2^{-1})$	+	[1234]
	$+(2^{-6} + 2^{-3} + 2^{-1})$	-	$+(2^{-6} + 2^{-3} + 2^{-2})$	-	$-(2^{-12} + 2^{-10})$	+	[1234]

Окончание табл. 2

$\mathbf{M}^{\pm}(\cdot)$	$f_{11} = f_{22},$ $f_{12} = -f_{21}$	$SF_1,$ SF_2	$g_{11} = g_{22},$ $g_{12} = -g_{21}$	$SG_1,$ SG_2	$h_{11} = h_{22},$ $h_{12} = -h_{21}$	$SH_1,$ SH_2	$\mathbf{B}_{pre},$ \mathbf{B}_{post}
$\mathbf{M}^+(Q_1)$	$+(2^{-6} + 2^{-4} + 2^{-2})$	-	$-(2^{-8} + 2^{-6} + 2^{-5})$	+	$+(2^{-5} + 2^{-3} + 2^{-2})$	-	[2431]
	$-(2^{-6} + 2^{-5} + 2^{-2})$	-	$+(2^{-8} + 2^{-6} + 2^{-4})$	-	$+(2^{-7} + 2^{-4} + 2^{-3})$	+	[1423]
$\mathbf{M}^+(Q_2)$	$-(2^{-6} + 2^{-5} + 2^{-1})$	+	$-(2^{-7} + 2^{-6} + 2^{-5})$	+	$+(2^{-7} + 2^{-3} + 2^{-1})$	+	[1432]
	$+(2^{-5} + 2^{-3} + 2^{-2})$	+	$-(2^{-8} + 2^{-5} + 2^{-4})$	+	$-(2^{-8} + 2^{-7} + 2^{-2})$	+	[1432]
$\mathbf{M}^-(Q_3)$	$+(2^{-4} + 2^{-2} + 2^{-1})$	-	$+(2^{-5} + 2^{-4} + 2^{-1})$	-	$+(2^{-7} + 2^{-4} + 2^{-3})$	-	[2314]
	$-(2^{-7} + 2^{-6} + 2^{-3})$	+	$-(2^{-8} + 2^{-4} + 2^{-1})$	-	$-(2^{-4} + 2^{-2} + 2^{-1})$	+	[1342]

Анализ амплитудно-частотных характеристик (АЧХ) $|H_k(e^{j\omega})|$ и ФЧХ $\arg(H_k(e^{j\omega}))$ банка фильтров анализа 8×24 Int-Q-ПУБФ (рис. 1, а) показывает, что Int-Q-ПУБФ характеризуется линейностью ФЧХ. Это обусловлено симметрией импульсных характеристик (рис. 1, б) канальных фильтров ($L=24$ отсчета). АЧХ системы анализа-синтеза $|T(e^{j\omega})|$ представляет собой прямую линию (рис. 1, а), что свидетельствует о наличии свойства полного восстановления в Int-Q-ПУБФ. На рис. 1, в изображены масштабирующая и вейвлет-функции. Вейвлет-функции являются гладкими и симметричными. Таким образом, Int-Q-ПУБФ может рассматриваться как многополосное вейвлет-преобразование, эффективность которого в сжатии изображений высокая.

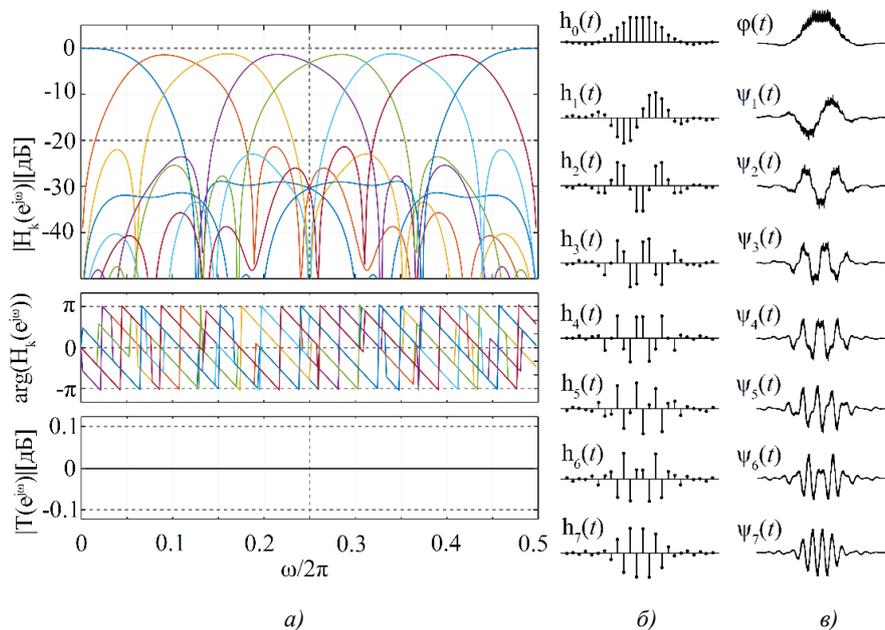


Рис. 1. Характеристики целочисленного восьмиканального 8×24 Int-Q-ПУБФ ($B = 12, K = 3$):

а) АЧХ, ФЧХ банка фильтров анализа и АЧХ системы анализа-синтеза; б) импульсные характеристики канальных фильтров; в) масштабирующая и вейвлет-функции

Сравнительный анализ Int-Q-ПУБФ для числа каналов $M = 8$, отсчетов импульсной характеристики ($L=16, 24$ и 32) и схемотехнических ограничений ($B = 12$ и $K = 3$) с другими известными преобразованиями (см. [8], табл. 2) по параметрам $CG, \varepsilon_{SBE}, DC Att.$ показывает,

что предложенный Int-Q-ПУБФ с блочной лестничной параметризацией характеризуется высокими показателями эффективности: 8×16 Int-Q-ПУБФ CG (9,64 дБ vs. 9,27 дБ), ε_{SBE} (-18,2 дБ vs. -18 дБ), DC Att. = -40 дБ; 8×24 Int-Q-ПУБФ CG (9,49 дБ vs. 9,38 дБ), ε_{SBE} (-21,3 дБ vs. -19,3 дБ), DC Att. = -49 дБ; 8×32 Int-Q-ПУБФ CG (9,48 дБ vs. 9,46 дБ), ε_{SBE} (-24,8 дБ vs. -18,9 дБ), DC Att. = -38 дБ, параметр DC Att. для схемотехнических ограничений ($B=64$ и $K=64$) равен -316 дБ и Q-ПУБФ могут рассматриваться как банки фильтров, не имеющие утечек по постоянному току, т. е. как оптимальные по показателю DC Att. Показатели эффективности для преобразований из работы [8, табл. 2] соответствуют реализации с плавающей запятой.

FPGA многоступенчатые конвейерные архитектуры Int-Q-ПУБФ

Реализация ступени Q-MUL IP-компонента. Рассмотрим процесс вычисления матриц перестановок \mathbf{B}_{pre} , \mathbf{B}_{post} и векторов-знаков \mathbf{SF} , \mathbf{SG} , \mathbf{SH} на примере оператора умножения кватернионов $\Gamma_4 \cdot \mathbf{M}^+(P_1 = p_1 + p_2i + p_3j + p_4k)$. В результате приведения кватерниона \mathbf{P}_1 к динамическому диапазону $[-1,1]$ модифицированный кватернион $\tilde{P}_1 = p_3 + p_1i + p_4j + p_2k$. Если определитель матрицы перестановок $\det(\mathbf{P}) = -1$, то в соответствии с правилом выбора оператора умножения [1] происходит смена оператора умножения на противоположный:

$$\mathbf{M}^+(P_1) = \mathbf{P}_{post} \mathbf{M}^-(\tilde{P}_1) \mathbf{P}_{pre},$$

$$\mathbf{P}_{pre} = \begin{bmatrix} 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{P}_{post} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Поскольку оператор умножения кватернионов всегда $\mathbf{M}^+(P)$, то в соответствии со строкой 3 табл. 1 [1] приведение целевого оператора умножения $\mathbf{M}^-(\tilde{P}_1)$ будет осуществлено, если выбрать

$$\mathbf{P}_{pre} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{P}_{post} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix}.$$

Матрицы перестановок \mathbf{B}_{pre} , \mathbf{B}_{post} и матрицы знаков \mathbf{R} , \mathbf{L} определяются следующим образом [1]:

$$\mathbf{B}_{pre} = |\mathbf{P}_{pre}| = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{B}_{post} = |\mathbf{P}_{post}| = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{R} = \mathbf{P}_{pre} \cdot \mathbf{B}_{pre}^{-1} = \text{diag}([1, 1, -1, 1]), \quad \mathbf{L} = \mathbf{B}_{post}^{-1} \cdot \mathbf{P}_{post} = \text{diag}([1, -1, -1, -1]).$$

Далее из строки 2 табл. 2 [1] определяются векторы знаков оператора $\Gamma_4 \cdot \mathbf{M}^+(P_1)$ умножения кватернионов:

$$\mathbf{SH} = [SH_1, SH_2]^T = [r_1, r_2]^T = [1, 1]^T, \quad \mathbf{SG} = [SG_1, SG_2]^T = [l_3 \cdot r_3, l_4 \cdot r_4] = [1, 1]^T,$$

$$\mathbf{SF} = [SF_1, SF_2]^T = [1, 1]^T.$$

Архитектура блока распределенной арифметики на сумматорах \mathbf{H} для оператора кватернионов $\mathbf{M}^+(P_i)$ изображена на рис. 2. Критический путь блока распределенной арифметики на сумматорах равен трем операциям сложения-вычитания, а умножителя кватернионов – четырем операциям сложения-вычитания согласно рис. 4 из работы [1]. Общая латентность умножителя кватернионов составит 12 операций сложения-вычитания. Матрицы \mathbf{B}_{post} , \mathbf{B}_{pre} и операции сдвига в блоке распределенной арифметики на сумматорах реализуются как коммутация соответствующих входных шин и поэтому не приводят к дополнительным затратам оборудования.

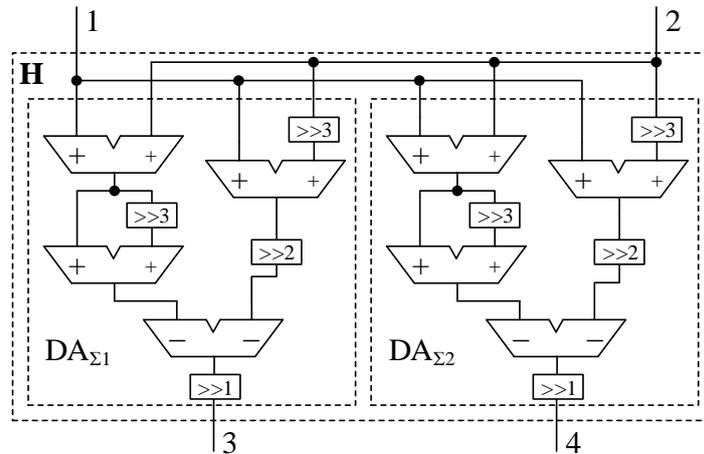


Рис. 2. Блок распределенной арифметики на сумматорах \mathbf{H}

Параллельно-поточный процессор восьмиканального 8×24 Int-Q-ПУБФ. Параллельно-конвейерная структура процессора восьмиканального 8×24 Int-Q-ПУБФ с попарно зеркальной симметрией частотных характеристик (см. [1], рис. 2, $N = 3$) на основе встроенных Q-MUL IP-компонентов с распределенной арифметикой на сумматорах изображена на рис. 3. Архитектура процессора Int-Q-ПУБФ включает два параллельно работающих многоступенчатых конвейера из Q-MUL IP-компонента. Следует отметить, что все умножители кватернионов имеют одинаковую структуру (это показано на рис. 3 черным треугольником вверху блока), но выполняют разные операции в соответствии с рис. 2 из работы [1].

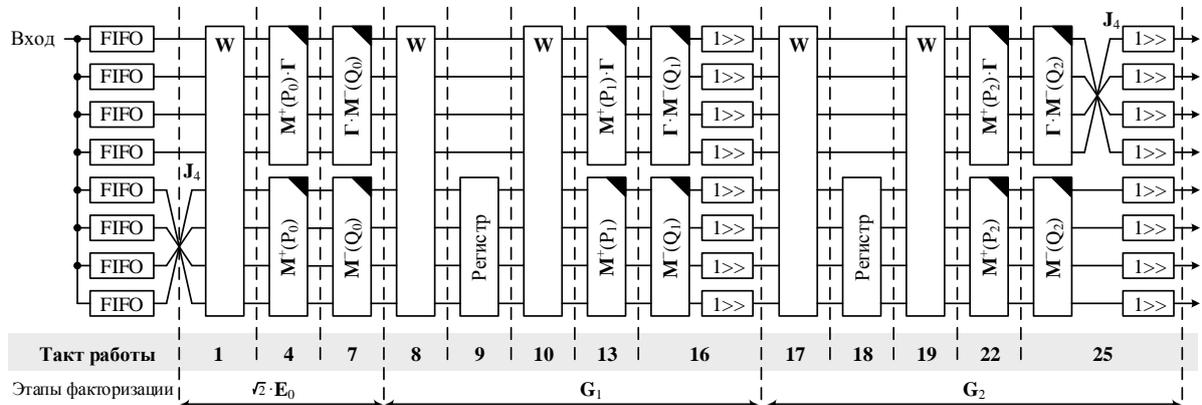


Рис. 3. Многоступенчатый конвейер восьмиканального 8×24 Int-Q-ПУБФ

Входной поток данных разбивается на части по восемь отсчетов, которые помещаются в соответствующие буферы памяти типа FIFO для согласования скорости работы процессора с частотой ввода данных в банк фильтров. Операция, определяющая период тактовой частоты конвейера процессора Int-Q-ПУБФ, – время выполнения ступени конвейера Q-MUL IP-компонента, состоящего из четырех операций сложения-вычитания. Операция задержки реализована на основе регистров и составляет один такт синхронизации процессора. Полное время задержки процессора равно 25 периодам синхронизации. Q-ПУБФ синтеза, как и Q-ПУБФ анализа, состоит из одинаковых компонентов, но переставленных в обратном порядке в соответствии с факторизацией полифазной передаточной матрицы банка синтеза [1].

Рабочая версия системы анализа-синтеза была получена с использованием библиотеки NUMERIC_STD. Моделирование разработанных VHDL-описаний проводилось в системе Xilinx ISE 14.7 и подтвердило корректность результата для кристалла Xilinx ZYNQ xc7z030-3fbg676. Анализ табл. 3 позволяет сделать вывод, что Int-Q-ПУБФ анализа и синтеза имеют одинаковые аппаратные затраты. Частота синхронизации работы Q-MUL IP-компонента составляет 200 МГц. Таким образом, производительность умножителя кватернионов равна 200 млн умножений кватернионов в секунду, а обработка входных данных восьмиканальным 8×24 Int-Q-ПУБФ ведется с максимальной скоростью 8×200 млн входных выборок в секунду.

Таблица 3

Результаты синтеза каждой ступени системы анализа-синтеза, шт.

Компонент	Slices	Slice Resisters	Slice LUTs	LUT-FF pairs
Int-Q-ПУБФ анализа	1588	3562	5604	5756
Int-Q-ПУБФ синтеза	1596	3623	5615	5677
Умножитель $M^+(P)$	159	217	431	469

Заключение. В работе предложена гибкая технология быстрого прототипирования архитектур процессоров Int-Q-ПУБФ на базе FPGA, в основу которой положен Q-MUL IP-компонент оператора умножения кватернионов на распределенной арифметике на сумматорах, обеспечивающий широкий диапазон регулировок параметров Int-Q-ПУБФ: амплитудных и фазочастотных характеристик, показателя эффективности кодирования, структурной регулярности, производительности, затрат ресурсов и форматов данных в арифметике с фиксированной запятой.

Список использованных источников

1. Рыбенков, Е. В. Синтез FPGA-архитектур банков фильтров на основе блочной лестничной факторизации в алгебре кватернионов (часть 1) / Е. В. Рыбенков, Н. А. Петровский // Информатика. – 2018. – Т. 15, № 2. – С. 29–44.
2. Petrovsky, N. A. Design and implementation of reversible integer quaternionic paraunitary filter banks on adder-based distributed arithmetic / N. A. Petrovsky, E. V. Rybenkov, A. A. Petrovsky // Signal Processing: Algorithms, Architectures, Arrangements, and Applications. – Poznan, 2017. – P. 17–22.
3. Petrovsky, N. Low read-only memory distributed arithmetic implementation of quaternion multiplier using split matrix approach / N. Petrovsky, A. Stankevich, A. Petrovsky // Electronics Letters. – 2014. – Vol. 50, no. 24. – P. 1809–1811. doi: 10.1049/el.2014.1775
4. Петровский, Н. А. CORDIC-техника для фиксированного угла вращения в операции умножения кватернионов / Н. А. Петровский, А. В. Станкевич, А. А. Петровский // Информатика. – 2015. – № 4(48). – С. 85–108.
5. Chang, T. S. New distributed arithmetic algorithm and its application to IDCT / T. S. Chang, C. Chen, C. W. Jen // IEE Proceedings – Circuits, Devices and Systems. – 1999. – Vol. 146, no. 4. – P. 159–163.
6. Vaidyanathan, P. P. Multirate Systems and Filter Banks / P. P. Vaidyanathan. – Englewood Cliffs, NJ : Prentice-Hall, 1993. – 911 p.
7. Корн, Г. А. Справочник по математике для научных работников и инженеров / Г. А. Корн, Т. М. Корн. – М. : Наука, 1974. – 832 с.
8. Li, B. A method for initializing free parameters in lattice structure of linear phase perfect reconstruction filter bank / B. Li, X. Gao // Signal Processing. – 2014. – Vol. 98. – P. 243–251.

References

1. Rybenkov E. V., Petrovsky N. A. Sintez FPGA-arhitektur bankov fil'trov na osnove blochnoj lestnichnoj faktorizacii v algebre kvaternionov (chast' 1) [Synthesis of FPGA architectures of block lifting-based filter banks in quaternion algebra (part 1)]. *Informatika [Informatics]*, 2018, vol. 15, no. 2, pp. 29–44 (in Russian).
2. Petrovsky N. A., Rybenkov E. V., Petrovsky A. A. Design and implementation of reversible integer quaternionic paraunitary filter banks on adder-based distributed arithmetic. *Signal Processing: Algorithms, Architectures, Arrangements, and Applications*. Poznan, 2017, pp. 17–22. doi: 10.23919/SPA.2017.8166830
3. Petrovsky N., Stankevich A., Petrovsky A. Low read-only memory distributed arithmetic implementation of quaternion multiplier using split matrix approach. *Electronics Letters*, 2014, vol. 50, no. 24, pp. 1809–1811. doi: 10.1049/el.2014.1775
4. Petrovsky N. A., Stankevich A. V., Petrovsky A. A. CORDIC-tehnika dlja fiksirovannogo ugla vrashhenija v operacii umnozhenija kvaternionov [CORDIC-techniques for fixed angle of rotation in multiplying operation of quaternions]. *Informatika [Informatics]*, 2015, no. 4(48), pp. 85–108 (in Russian).
5. Chang T. S., Chen C., Jen C. W. New distributed arithmetic algorithm and its application to IDCT. *IEE Proceedings – Circuits, Devices and Systems*, 1999, vol. 146, no. 4, pp. 159–163. doi: 10.1049/ip-cds:19990537
6. Vaidyanathan P. P. *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ, Prentice-Hall, 1993, 911 p.
7. Korn G. A., Korn T. M. *Spravochnik po matematike dlja nauchnyh rabotnikov i inzhenerov [A Handbook of Mathematics for Scientists and Engineers]*. Moscow, Nauka Publ., 1974, 832 p. (in Russian).
8. Li B., Gao X. A method for initializing free parameters in lattice structure of linear phase perfect reconstruction filter bank. *Signal Processing*, 2014, vol. 98, pp. 243–251. doi: 10.1016/j.sigpro.2013.11.016

Информация об авторах

Рыбенков Евгений Викторович – ассистент кафедры электронно-вычислительных систем, Белорусский государственный университет информатики и радиоэлектроники (ул. П. Бровки, 6, 220013, Минск, Республика Беларусь). E-mail: rybenkov@bsuir.by

Петровский Николай Александрович – кандидат технических наук, доцент кафедры электронно-вычислительных систем, Белорусский государственный университет информатики и радиоэлектроники (ул. П. Бровки, 6, 220013, Минск, Республика Беларусь). E-mail: nick.petrovsky@bsuir.by

Information about the authors

Eugene V. Rybenkov – Assistant Professor, Department of Computer Engineering, Belarusian State University of Informatics and Radioelectronics (6, P. Brovki Str., 220013, Minsk, Republic of Belarus). E-mail: rybenkov@bsuir.by

Nick A. Petrovsky – Ph. D. (Engineering), Associate Professor, Department of Computer Engineering, Belarusian State University of Informatics and Radioelectronics (6, P. Brovki Str., 220013, Minsk, Republic of Belarus). E-mail: nick.petrovsky@bsuir.by

ISSN 1816-0301 (print)

УДК 004.942

Поступила в редакцию 03.04.2018

Received 03.04.2018

В. М. Артемьев, А. О. Наумов, Л. Л. Кохан*Институт прикладной физики Национальной академии наук Беларуси, Минск, Беларусь***ЛИНЕЙНАЯ АДАПТИВНАЯ ФИЛЬТРАЦИЯ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НА ОСНОВЕ ДЕТЕРМИНИРОВАННОГО ПОДХОДА**

Аннотация. Рассматривается метод синтеза фильтров случайных последовательностей при отсутствии априорной статистической информации о характеристиках полезного сигнала и шумов. При синтезе используются лишь данные о текущих измерениях и ограниченный объем эмпирической информации, что приводит к необходимости применения детерминированного подхода на основе метода наименьших квадратов. В целях получения рекуррентного алгоритма фильтрации предлагается расширение структуры функции потерь метода за счет включения в ее состав дополнительного слагаемого, задающего экстраполяцию оценки на следующий период измерений. Оптимальная текущая оценка находится с учетом как результатов измерений, так и экстраполированных значений. Выбор функции экстраполяции осуществляется исходя из желаемого класса синтезируемого фильтра. В работе рассматривается вариант полиномиальной экстраполяции с учетом предшествующих оценок и измерений. Использование только предшествующих оценок приводит к структуре фильтра с обратной связью, а использование только предшествующих измерений формирует трансверсальный фильтр. Проводится математическое моделирование, и на конкретном примере оцениваются потери точности фильтрации за счет неучета априорной статистической информации.

Ключевые слова: случайные последовательности, детерминированный подход, адаптивная фильтрация, расширенный метод наименьших квадратов

Для цитирования. Артемьев, В. М. Линейная адаптивная фильтрация случайных последовательностей на основе детерминированного подхода / В. М. Артемьев, А. О. Наумов, Л. Л. Кохан // Информатика. – 2018. – Т. 15, № 3. – С. 32–40.

V. M. Artemiev, A. O. Naumov, L. L. Kokhan*Institute of Applied Physics of the National Academy of Sciences of Belarus, Minsk, Belarus***LINEAR ADAPTIVE FILTERING OF RANDOM SEQUENCES ON BASIS OF DETERMINISTIC APPROACH**

Abstract. The article studies the technique of synthesis of random sequence filters with unknown prior statistical information about the parameters of signal and noises. The synthesis uses only current measurements and a limited amount of empirical information, which leads to the necessity of using a deterministic approach based on the least squares method. In order to obtain a recursive filtering algorithm, it is proposed to extend the structure of the method loss function by including in loss function an additional term that defines the estimate extrapolation for the next measurement period. The optimal current estimate is based on both measurement results and extrapolated values. The extrapolation function is selected based on the desired class of synthesized filter. The paper considers a variant of polynomial extrapolation, taking into account previous estimates and measurements. The use of only previous estimates leads to the structure of the filter with feedback, while the use of only the previous measurements forms a transversal filter. Mathematical modeling was carried out and on particular example and the loss of filtering accuracy by not taking into account a priori statistical information was estimated.

Keywords: random sequences, deterministic approach, adaptive filtering, extended least-square method

For citation. Artemiev V. M., Naumov A. O., Kokhan L. L. Linear adaptive filtering of a random sequences on basis of deterministic approach. *Informatics*, 2018, vol. 15, no. 3, pp. 32–40 (in Russian).

Введение. Несмотря на большое число публикаций, задача фильтрации случайных процессов при ограниченных априорных данных о полезном сигнале и помехах продолжает оставаться актуальной. Существуют два адаптивных метода ее решения: непрямой и прямой. В первом случае измеренные данные используются для восстановления отсутствующей априорной информации и дальнейшей фильтрации известными методами. В прямых методах задача решается непосредственно по текущим измерениям без восстановления априорной статистики.

Использование для синтеза фильтров лишь текущих данных с учетом ограниченного объема эмпирической информации приводит к детерминированному подходу решения задачи [1]. При этом в качестве основного выступает метод наименьших квадратов [2], связанный с минимизацией текущих значений квадрата невязки решения. В настоящей работе исследуется прямой метод рекуррентной адаптивной фильтрации с использованием расширенного метода наименьших квадратов (РМНК).

Разработка адаптивных фильтров методом наименьших квадратов активно развивалась начиная с 1980-х гг. Одной из первых стала работа [3], где был разработан метод синтеза адаптивных решетчатых фильтров. Однако полученные алгоритмы не были рекуррентными, поскольку использовали весь текущий набор измерений, что приводило к непрерывному увеличению размерности задачи. Попытка нахождения рекуррентных алгоритмов сделана в работе [4] на основе геометрических представлений о решении задачи фильтрации, что привело к громоздким уравнениям, малодоступным для практики. Рекуррентный метод фильтрации был рассмотрен в работе [5], где показано, что рекуррентность позволяет существенно ускорить процесс адаптации. Сравнительная оценка сложности рекуррентных алгоритмов дана в работе [6]. Существенное продвижение в технологии метода синтеза рекуррентных адаптивных алгоритмов сделано в работе [7]. Авторы ввели в рассмотрение широко используемое в теории автоматического управления понятие пространства состояний и применили его для решения задач фильтрации методом наименьших квадратов. В частности, это упростило процедуру анализа ошибок в адаптивных фильтрах [8]. Итоги проведенных исследований были обобщены в монографиях [1, 9].

В опубликованных работах по синтезу адаптивных фильтров необходимым условием является обращение матриц, элементы которых определяются измеренными сигналами. Размерности этих матриц в процессе работы фильтра возрастают. Для перехода к рекуррентным алгоритмам в них используется процедура рекуррентного обращения матрицы. В общем случае она не дает точных результатов и требует дополнительных вычислений, что является недостатком метода.

Предлагаемый метод синтеза адаптивных фильтров при детерминированном подходе на основе расширенного метода наименьших квадратов [10, 11] позволяет находить структуру рекуррентного адаптивного фильтра и определять его параметры, не прибегая к обращению матриц.

Формулировка задачи. Предполагается, что исходный скалярный процесс является случайной последовательностью x_k , где $k = 0, 1, 2, \dots$ есть дискретное время. Будем считать, что модель этого процесса и ее статистические характеристики априорно неизвестны. Эмпирически полагаем, что значения процесса коррелированы во времени и время корреляции существенно выше периода измерений. Соответственно, возможна экстраполяция значений процесса \tilde{x}_k на момент времени k по его предшествующим значениям.

Посредством линейного безынерционного датчика проводятся следующие измерения:

$$z_k = h_k x_k + v_k, \quad (1)$$

где h_k – коэффициент чувствительности датчика, v_k – случайная ошибка измерений (шум измерений) с неизвестными статистическими характеристиками. Полагается, что длительность корреляции шума значительно меньше длительности корреляции сигнала.

Задача заключается в нахождении уравнений фильтра для текущей оценки \hat{x}_k значений исходного процесса x_k по результатам измерений z_k в условиях априорной статистической неопределенности, указанной выше.

При прямом методе адаптивной фильтрации в существующих работах производят непосредственное задание структуры фильтра, в рамках которого ищется решение [12]. В настоящей работе класс искомого фильтра определяется видом функции экстраполяции текущего значения оценки сигнала по данным, полученным в предыдущие моменты времени, и предшествующих измерений. Этот вид функции выбирается из числа известных [13] с набором пара-

метров, подлежащих оценке в процессе фильтрации. Представим их в виде вектора $\mathbf{a}_{n,k} = [a_{1,k}, a_{2,k}, \dots, a_{n,k}]^T$, где n – число параметров. Функцию экстраполяции обозначим как $\tilde{x}_k(\mathbf{a}_{n,k})$.

В основе решения лежит выбор функции потерь и задание критерия оптимальности. В РМНК используется квадратичная функция потерь, состоящая из нескольких слагаемых. Первое определяет потери за счет невязки решения и в скалярном случае задается функцией $(z_k - h_k \hat{x}_k)^2$. Для получения рекуррентного алгоритма вводится слагаемое $(\hat{x}_k - \tilde{x}_k(\mathbf{a}_{n,k}))^2$, где функция экстраполяции зависит от n неизвестных коэффициентов $\mathbf{a}_{n,k}$, подлежащих оценке наряду с \hat{x}_k . Чтобы оценки коэффициентов также были рекуррентными, вводится третье слагаемое – $\sum_{i=1}^n (a_{i,k} \hat{x}_{k-i} - a_{i,k-1} \hat{x}_{k-1-i})^2$. В итоге функция потерь принимает вид

$$J_k(\hat{x}_k, \mathbf{a}_{n,k}) = (z_k - h_k \hat{x}_k)^2 + (\hat{x}_k - \tilde{x}_k(\mathbf{a}_{n,k}))^2 + \sum_{i=1}^n (a_{i,k} \hat{x}_{k-i} - a_{i,k-1} \hat{x}_{k-1-i})^2. \quad (2)$$

Критерием оптимальной фильтрации является минимум этой функции относительно значений \hat{x}_k и параметров $\mathbf{a}_{n,k}$. Искомые оценки удовлетворяют необходимым условиям оптимальности

$$\frac{\partial J_k(\hat{x}_k, \mathbf{a}_{n,k})}{\partial \hat{x}_k} = 0, \quad \frac{\partial J_k(\hat{x}_k, \mathbf{a}_{n,k})}{\partial a_{j,k}} = 0, \quad j = \overline{1, n}, \quad (3)$$

откуда находятся уравнения оптимального фильтра и оценок коэффициентов функции экстраполяции. Структура адаптивного фильтра изображена на рис. 1.



Рис. 1. Структура адаптивного фильтра

Варьируя функцию экстраполяции, можно получить различные варианты алгоритма адаптивной фильтрации.

Уравнения фильтра с обратной связью. Фильтр с обратной связью (бесконечной памятью) находится путем задания функции экстраполяции регрессией

$$\tilde{x}_k(\mathbf{a}_{n,k}) = \sum_{i=1}^n a_{i,k} \hat{x}_{k-i}, \quad (4)$$

линейно зависящей от имеющихся оценок за n предыдущих моментов измерений. В этом случае уравнения оптимальности (3) приводятся к виду

$$\frac{\partial J_k(\hat{x}_k, \mathbf{a}_{n,k})}{\partial \hat{x}_k} = -h_k(z_k - h_k \hat{x}_k) + \left(\hat{x}_k - \sum_{i=1}^n a_{i,k} \hat{x}_{k-i} \right) = 0, \quad (5)$$

$$\frac{\partial J_k(\hat{x}_k, \mathbf{a}_{n,k})}{\partial a_{j,k}} = - \left(\hat{x}_k - \sum_{i=1}^n a_{i,k} \hat{x}_{k-i} \right) + (a_{j,k} \hat{x}_{k-j} - a_{j,k-1} \hat{x}_{k-1-j}) = 0, \quad j = \overline{1, n}. \quad (6)$$

Из соотношения (5) находится уравнение для оценки значений процесса \hat{x}_k :

$$\hat{x}_k = k_{1,k} \sum_{i=1}^n a_{i,k} \hat{x}_{k-i} + k_{0,k} z_k. \quad (7)$$

Здесь коэффициенты усиления $k_{1,k}$ и $k_{0,k}$ вычисляются по формулам

$$k_{1,k} = 1/(1+h_k^2), \quad k_{0,k} = h_k/(1+h_k^2). \quad (8)$$

Структура уравнения (7) изображена на рис. 2, где символом z^{-1} обозначена задержка сигнала на период измерений.

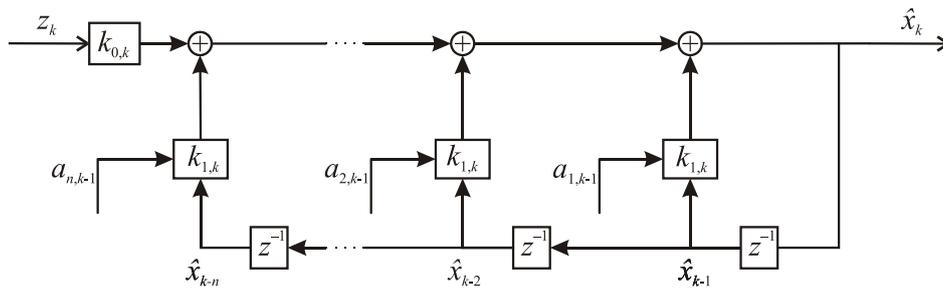


Рис. 2. Структура уравнения оценки процесса в адаптивном фильтре с обратной связью

Уравнения оценок коэффициентов функции экстраполяции находятся из выражений (4), (6) и в итоге приводятся к виду

$$a_{j,k} \hat{x}_{k-j} + h_k k_{0,k} \sum_{i=1}^n a_{i,k} \hat{x}_{k-i} = a_{j,k-1} \hat{x}_{k-1-j} + k_{0,k} z_k, \quad j = \overline{1, n}. \quad (9)$$

Эти уравнения позволяют находить произведения $a_{j,k} \hat{x}_{k-j}$, где множитель \hat{x}_{k-j} известен и определяется на n предыдущих моментах времени, и образуют систему из n линейных уравнений, решение которой не зависит от решения уравнения (7). Задача фильтрации начинается с решения уравнений (9), результаты которого затем используются для решения уравнений оценки (7).

Качество фильтрации задается величиной функции потерь (2), которая в оптимальном фильтре принимает минимальную величину $J_{k, \text{опт}}(\hat{x}_k, \mathbf{a}_{n,k})$. Для ее определения из уравнения (5) получаем выражение

$$h_k(z_k - h_k \hat{x}_k) = \left(\hat{x}_k - \sum_{i=1}^n a_{i,k} \hat{x}_{k-i} \right),$$

а из уравнения (6) – равенство

$$(a_{j,k}\hat{x}_{k-j} - a_{j,k-1}\hat{x}_{k-1-j}) = \left(\sum_{i=1}^n a_{i,k}\hat{x}_{k-i} - \hat{x}_k \right) = h_k(z_k - h_k\hat{x}_k).$$

Подставляя эти выражения в формулу (2), находим величину оптимальной функции потерь:

$$\begin{aligned} J_{k,\text{опт}}(\hat{x}_k, \mathbf{a}_{n,k}) &= (z_k - h_k\hat{x}_k)^2 + (\hat{x}_k - \tilde{x}_k(\mathbf{a}_{n,k}))^2 + \sum_{i=1}^n (a_{i,k}\hat{x}_{k-i} - a_{i,k-1}\hat{x}_{k-1-i})^2 = \\ &= (1 + nh_k^2)(z_k - h_k\hat{x}_k)^2. \end{aligned} \quad (10)$$

Таким образом, качество оптимальной фильтрации при детерминированном подходе определяется квадратом величины невязки и является случайной функцией времени. Используя формулу (1), а также учитывая, что разность $x_k - \hat{x}_k = e_k$ есть ошибка фильтрации, формулу (10) можно представить в виде $J_{k,\text{опт}}(\hat{x}_k, \mathbf{a}_{n,k}) = (1 + nh_k^2)(h_k e_k + v_k)^2$. Среднее значение этой величины по физическому смыслу близко к понятию дисперсии ошибки фильтрации $\sigma_e^2 = \langle e_k \rangle^2$. Треугольные скобки обозначают операцию нахождения математического ожидания.

Уравнения трансверсального фильтра. Функцию экстраполяции также можно сформировать на основе предшествующих измерений по формуле

$$\hat{x}_k = h_k^{-1} \sum_{i=1}^n a_{i,k} z_{k-i}. \quad (11)$$

В результате функция потерь (2) будет выглядеть следующим образом:

$$J_k(\hat{x}_k, \mathbf{a}_{n,k}) = (z_k - h_k\hat{x}_k)^2 + \left(\hat{x}_k - h_k^{-1} \sum_{i=1}^n a_{i,k} z_{k-i} \right)^2 + h_k^{-2} \sum_{i=1}^n (a_{i,k} z_{k-i} - a_{i,k-1} z_{k-1-i})^2.$$

Уравнения оптимальности (3) приводятся к виду

$$\frac{\partial J_k(\hat{x}_k, \mathbf{a}_{n,k})}{\partial \hat{x}_k} = -h_k(z_k - h_k\hat{x}_k) + \left(\hat{x}_k - h_k^{-1} \sum_{i=1}^n a_{i,k} z_{k-i} \right) = 0, \quad (12)$$

$$\frac{\partial J_k(\hat{x}_k, \mathbf{a}_{n,k})}{\partial a_{j,k}} = - \left(\hat{x}_k - h_k^{-1} \sum_{i=1}^n a_{i,k} z_{k-i} \right) + h_k^{-2} (a_{j,k} z_{k-j} - a_{j,k-1} z_{k-1-j}) = 0, \quad j = \overline{1, n}. \quad (13)$$

Из соотношения (13) следует уравнение оценки процесса

$$\hat{x}_k = k_{1,k} \sum_{i=1}^n a_{i,k} z_{k-i} + k_{0,k} z_k, \quad (14)$$

где коэффициенты усиления $k_{1,k} = h_k^{-1} / (1 + h_k^2)$, $k_{0,k} = h_k / (1 + h_k^2)$.

Структура уравнения оценки процесса (рис. 3) соответствует трансверсальному фильтру (с конечной памятью), оптимальные параметры которого определяются уравнениями

$$a_{j,k}z_{k-j} + h_k^2 k_{0,k} \sum_{i=1}^n a_{i,k} z_{k-i} = a_{j,k-1} z_{k-1-j} + h_k^2 k_{0,k} z_k, \quad j = \overline{1, n}. \quad (15)$$

Данные уравнения представляют собой систему линейных уравнений n -го порядка, решение которой позволяет находить значения $a_{i,k} z_{k-i}$, $i = \overline{1, n}$, и использовать их в уравнении оценки (14).

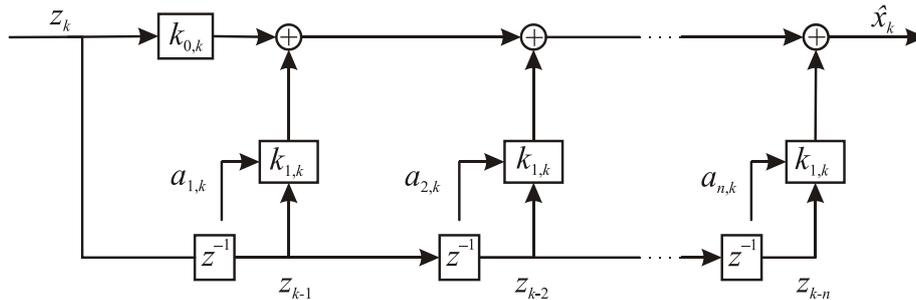


Рис. 3. Структура уравнения оценки процесса в адаптивном трансверсальном фильтре

Способом, изложенным в предыдущем разделе, можно получить следующее выражение для оптимальной величины функции потерь:

$$J_{k,\text{опт}}(\hat{x}_k, \mathbf{a}_{n,k}) = (1 + h_k^2 + nh_k^4)(z_k - h_k \hat{x}_k)^2 = (1 + h_k^2 + nh_k^4)(h_k e_k + v_k)^2, \quad (16)$$

где $e_k = x_k - \hat{x}_k$ – ошибка фильтрации. Как и в случае фильтра с обратной связью, величина функции потерь при детерминированном подходе пропорциональна квадрату невязки решения.

Результаты моделирования. Рассматривается задача, когда исходная случайная последовательность моделируется посредством стохастического конечно-разностного уравнения второго порядка:

$$\begin{aligned} x_{1,k} &= a_{1,1}x_{1,k-1} + a_{1,2}x_{2,k-1} + w_{1,k}, \\ x_{2,k} &= a_{2,2}x_{1,k-1} + a_{2,2}x_{2,k-1} + w_{2,k}. \end{aligned} \quad (17)$$

В системе уравнений (17) $w_{1,k}$ и $w_{2,k}$ полагаются центрированными дискретными белыми шумами. При моделировании приняты следующие значения параметров: $a_{1,1} = 0,98$, $a_{1,2} = 0,8$, $w_{1,k} = 0$, $a_{2,1} = 0$, $a_{2,2} = 0,9$, а дисперсия белого шума $\sigma_{w,2}^2 = 0,04$. Соответственно, дисперсия процесса $x_{1,k}$ $\sigma_{x_{1,k}}^2 = 100$, а длительность корреляции $\tau_{1,k} = 50$. Измеряется лишь процесс $x_{1,k}$ согласно линейному уравнению

$$z_k = h_{1,k}x_{1,k} + v_{1,k}, \quad (18)$$

где шум измерений $v_{1,k}$ является центрированным дискретным белым шумом с дисперсией $\sigma_{v,1}^2 = 4$, а коэффициент $h_{1,k} = 1$.

Цель моделирования состояла в оценке потерь, которые имеют место за счет неучета априорной статистической информации. Для сравнения использовались результаты оценок

ошибок фильтра Калмана (ФК) при известной априорной статистической информации. Известно, что в этом случае ФК дает оптимальные оценки по критерию минимума среднеквадратического отклонения (СКО) ошибки фильтрации $\sigma_{e,k}$. Для рассматриваемого примера на рис. 4 сплошной линией показана зависимость этой величины от времени. Стояла задача сравнения точности ФК и адаптивного фильтра при неизвестной модели процесса. Для этого проводилось моделирование в соответствии с уравнениями (7) и (8) при различных порядках экстраполятора ($n=1, 2, 3$). Усреднение результатов осуществлялось по 1000 реализациям процесса изменения ошибки фильтрации $e_{1,k}$. Значения СКО $\sigma_{e,k}$ адаптивного фильтра в установившемся режиме при различных значениях n показаны на рис. 4.

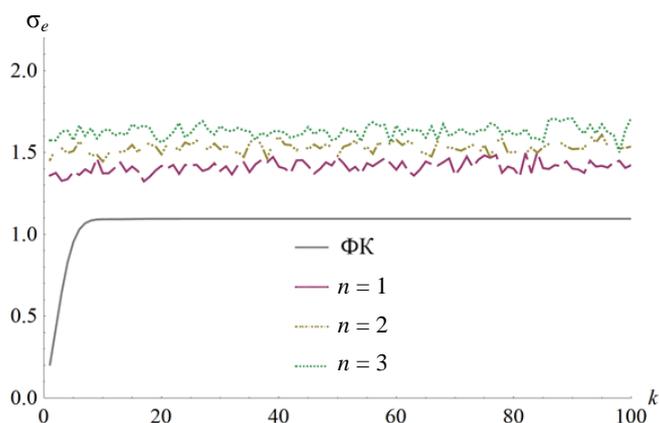


Рис. 4. Сравнительная оценка СКО $\sigma_{e,k}$ фильтра Калмана с адаптивным фильтром при различных значениях n порядка экстраполирующего полинома

Сравнение предлагаемого фильтра и ФК показывает, что ошибка адаптивного фильтра больше, чем у ФК. Это объясняется неучетом априорной информации. Наименьшее значение ошибок для данного конкретного примера имеет место при $n=1$. Как показывают результаты моделирования, при большей длительности корреляции $\tau_{1,k}$ минимальное СКО ошибки может оказаться при $n=2$. Минимум при $n=3$ встречается редко при очень больших длительностях корреляции.

Заключение. В работе предложен метод синтеза фильтров при отсутствии априорной статистической информации о характеристиках полезного сигнала и помех на основе детерминированного подхода, при котором используются лишь данные измерений, полученных ранее оценок и ограниченный объем эмпирической информации. Метод базируется на РМНК, позволяющем находить структуру рекуррентного оптимального фильтра и определять его параметры. Структура находится в классе фильтров, которые задаются видом функции экстраполяции, входящей в состав критерия оптимальности. Изучен вариант, когда эта функция является полиномом. В первом случае считается, что он зависит только от имеющихся предшествующих оценок процесса, и это приводит к структуре фильтра с обратной связью (бесконечной памятью). Во втором случае для формирования экстраполирующего полинома используются лишь предшествующие измерения и это приводит к структуре трансверсального фильтра (с конечной памятью). Параметры полиномов полагаются неизвестными, и их значения оцениваются в процессе фильтрации, что и делает фильтр адаптивным. На конкретном примере проведена сравнительная оценка точности адаптивного фильтра с результатами ФК. Оказалось, что СКО ошибок адаптивного фильтра больше, чем у ФК, примерно в полтора раза. Это является следствием неучета априорной статистической информации.

Список использованных источников

1. Haykin, S. S. *Adaptive Filter Theory* / S. S. Haykin. – N.J. : Prentice-Hall, 2002. – 936 p.
2. Фомин, В. М. Рекуррентное оценивание и адаптивная фильтрация / В. М. Фомин. – М. : Наука, 1984. – 388 с.
3. Mueller, M. S. Least-squares algorithms for adaptive equalizers / M. S. Mueller // *The Bell System Technical Journal*. – 1981. – Vol. 60. – P. 1905–1925.
4. Lev-Ari, H. Least-squares adaptive lattice and transversal filters: A unified geometric theory / H. Lev-Ari, T. Kailath, J. Cioffi // *IEEE Transactions on Information Theory*. – 1984. – Vol. 30. – P. 222–236.
5. Cioffi, J. M. Fast, recursive-least-squares transversal filters for adaptive filtering / J. M. Cioffi, T. Kailath // *IEEE Transactions on Acoustics Speech and Signal Processing*. – 1984. – Vol. 32. – P. 304–337.
6. Luk, F. T. Analysis of a recursive least-squares signal-processing algorithm / F. T. Luk, S. Qiao // *SIAM Journal on Scientific and Statistical Computing*. – 1989. – Vol. 10. – P. 407–418.
7. Sayed, A. H. A state-space approach to adaptive RLS filtering / A. H. Sayed, T. Kailath // *IEEE Signal Processing Magazine*. – 1994. – Vol. 11. – P. 18–60.
8. Yang, B. A note on the error propagation analysis of recursive least squares algorithms / B. Yang // *IEEE Transactions on Signal Processing*. – 1994. – Vol. 42. – P. 3523–3525.
9. Manolakis, D. G. *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering, and Array Processing* / D. G. Manolakis, V. K. Ingle, S. M. Kogon. – Boston : McGraw-Hill, 2000. – 796 p.
10. Артемьев, В. М. Линейная фильтрация многомерных случайных последовательностей расширенным методом наименьших квадратов / В. М. Артемьев, А. О. Наумов, Л. Л. Кохан // *Информатика*. – 2016. – № 4(52). – С. 20–25.
11. Артемьев, В. М. Нелинейная фильтрация случайных последовательностей расширенным методом наименьших квадратов / В. М. Артемьев, А. О. Наумов, Л. Л. Кохан // *Информатика*. – 2018. – Т. 15, № 1. – С. 60–69.
12. Цыпкин, Я. З. Оптимизация в условиях неопределенности / Я. З. Цыпкин // *Доклады АН СССР*. – 1976. – Т. 228, № 6. – С. 1306–1309.
13. Андерсон, Т. Статистический анализ временных рядов / Т. Андерсон. – М. : Мир, 1976. – 755 с.

References

1. Haykin S. S. *Adaptive Filter Theory*. N.J., Prentice-Hall, 2002, 936 p.
2. Fomin V. M. *Rekurrentnoe otsenivanie i adaptivnaya fil'tratsiya. Recurrent Estimation and Adaptive Filtering*. Moscow, Nauka Publ., 1984, 388 p. (in Russian).
3. Mueller M. S. Least-squares algorithms for adaptive equalizers. *The Bell System Technical Journal*, 1981, vol. 60, pp. 1905–1925. doi: 10.1002/j.1538-7305.1981.tb00302.x
4. Lev-Ari H., Kailath T., Cioffi J. Least-squares adaptive lattice and transversal filters: A unified geometric theory. *IEEE Transactions on Information Theory*, 1984, vol. 30, pp. 222–236. doi: 10.1109/tit.1984.1056882
5. Cioffi J. M., Kailath T. Fast, recursive-least-squares transversal filters for adaptive filtering. *IEEE Transactions on Acoustics Speech and Signal Processing*, 1984, vol. 32, pp. 304–337. doi: 10.1109/tassp.1984.1164334
6. Luk F. T., Qiao S. Analysis of a recursive least-squares signal-processing algorithm. *SIAM Journal on Scientific and Statistical Computing*, 1989, vol. 10, pp. 407–418. doi: 10.1137/0910027
7. Sayed A. H., Kailath T. A state-space approach to adaptive RLS filtering. *IEEE Signal Processing Magazine*, 1994, vol. 11, pp. 18–60. doi: 10.1109/79.295229
8. Yang B. A note on the error propagation analysis of recursive least squares algorithms. *IEEE Transactions on Signal Processing*. 1994, vol. 42, pp. 3523–3525. doi: 10.1109/78.340788
9. Manolakis D. G., Ingle V. K., Kogon S. M. *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering, and Array Processing*. Boston, McGraw-Hill, 2000, 796 p.
10. Artem'ev V. M., Naumov A. O., Kokhan L. L. Lineynaya fil'tratsiya mnogomernykh sluchaynykh posledovatel'nostey rasshirennym metodom naimen'shikh kvadratov [Linear filtering of random sequences using a least squares method with regularization]. *Informatika [Informatics]*, 2016, no. 4(52), pp. 20–25 (in Russian).
11. Artem'ev V. M., Naumov A. O., Kokhan L. L. Nelineynaya fil'tratsiya sluchaynykh posledovatel'nostey rasshirennym metodom naimen'shikh kvadratov [Nonlinear filtering of random sequences with extended least-squares method]. *Informatika [Informatics]*, 2018, vol. 15, no. 1, pp. 60–69 (in Russian).
12. Tsytkin Ia. Z. Optimizatsiya v usloviyakh neopredelennosti [Optimization in the conditions of uncertainty]. *Doklady AN SSSR [Proc. of the Academy of Sciences of the USSR]*, 1976, vol. 228, no. 6, pp. 1306–1309 (in Russian).
13. Anderson T. *Statisticheskiy analiz vremennykh ryadov. Statistical Analysis of Time Series*. Moscow, Mir Publ., 1976, 755 p. (in Russian).

Информация об авторах

Артемьев Валентин Михайлович – член-корреспондент Национальной академии наук Беларуси, доктор технических наук, профессор, главный научный сотрудник, Институт прикладной физики НАН Беларуси (ул. Академическая, 16, 220072, Минск, Республика Беларусь). E-mail: artemiev@iaph.bas-net.by

Наумов Александр Олегович – кандидат физико-математических наук, заведующий лабораторией, Институт прикладной физики НАН Беларуси (ул. Академическая, 16, 220072, Минск, Республика Беларусь). E-mail: naumov@iaph.bas-net.by

Кохан Леонид Леонидович – кандидат технических наук, старший научный сотрудник, Институт прикладной физики НАН Беларуси (ул. Академическая, 16, 220072, Минск, Республика Беларусь). E-mail: kokhanll@iaph.bas-net.by

Information about the authors

Valentin M. Artemiev – Corresponding Member of the National Academy of Sciences of Belarus, D. Sc. (Engineering), Professor, Chief Researcher, Institute of Applied Physics of the National Academy of Sciences of Belarus (16, Akademicheskaya Str., 220072, Minsk, Republic of Belarus). E-mail: artemiev@iaph.bas-net.by

Alexander O. Naumov – Ph. D. (Physics and Mathematics), Head of Laboratory, Institute of Applied Physics of the National Academy of Sciences of Belarus (16, Akademicheskaya Str., 220072, Minsk, Republic of Belarus). E-mail: naumov@iaph.bas-net.by

Leonid L. Kokhan – Ph. D. (Engineering), Senior Researcher, Institute of Applied Physics of the National Academy of Sciences of Belarus (16, Akademicheskaya Str., 220072, Minsk, Republic of Belarus). E-mail: kokhanll@iaph.bas-net.by

ISSN 1816-0301 (print)
УДК 004.932

Поступила в редакцию 12.06.2018
Received 12.06.2018

В. В. Старовойтов

*Объединенный институт проблем информатики
Национальной академии наук Беларуси, Минск, Беларусь*

УТОЧНЕНИЕ ИНДЕКСА *SSIM* СТРУКТУРНОГО СХОДСТВА ИЗОБРАЖЕНИЙ

Аннотация. Исследуются свойства популярной меры сравнения цифрового изображения с эталоном – индекса структурного сходства, называемого в литературе *SSIM*. Доказывается, что *SSIM* и производные от него функции не являются метриками. Описываются варианты модификации индекса *SSIM*. Показывается, что меры, подобные этому индексу, оценивают не качество изображений, а их пофрагментное сходство. Кроме того, отмечается, что усредненные оценки, выставляемые экспертами и называемые *MOS*, очень субъективны и не могут в точности коррелировать с вычисляемыми количественными оценками сходства сравниваемых изображений. Для подсчета индекса *SSIM* вычисляется матрица локальных оценок. Каждая оценка определяет сходство двух анализируемых пикселей с одинаковыми координатами с учетом соседних пикселей. Обычно в качестве индекса *SSIM* используется средняя арифметическая величина полученной матрицы. Вместо нее для усовершенствования индекса *SSIM* предлагается использовать параметр масштаба распределения Вейбулла, аппроксимирующего гистограмму локальных оценок индекса *SSIM*. На множестве изображений общедоступной базы TID2013 показывается, что предложенный параметр является более точной характеристикой сходства изображений, чем среднее множество локальных оценок.

Ключевые слова: цифровое изображение, мера сходства изображений, индекс *SSIM*, распределение Вейбулла
Для цитирования. Старовойтов, В. В. Уточнение индекса *SSIM* структурного сходства изображений / В. В. Старовойтов // Информатика. – 2018. – Т. 15, № 3. – С. 41–55.

V. V. Starovoitov

The United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Belarus

ENHANCEMENT OF THE STRUCTURAL SIMILARITY INDEX *SSIM*

Abstract. Properties of a popular measure of comparing a digital image with a reference – the index of structural similarity, called *SSIM* in the literature – are explored. It is proved that the *SSIM* and its derivative functions are not metrics. Variants of the index modification are described. It is shown that measures similar to this index evaluate not quality of images, but their similarity by fragments. Additionally, it is shown that the averaged expert assessments called *MOS* are very subjective and cannot exact correlate with numerical estimates of similarity of the compared images. To get the *SSIM* index, a matrix of local estimates is calculated. Each evaluation determines similarity of two analyzed pixels with the same coordinates taking into account neighboring pixels. Usually, the average arithmetic value of the obtained matrix is used as the *SSIM* index. Instead, to improve the *SSIM* index, it is proposed to use the scale parameter of the Weibull distribution, which approximates the histogram of the local index estimates. On a set of images from the public database TID2013, it is shown that the proposed parameter is a more accurate characteristic of image similarity than the mean of local estimates.

Keywords: digital image, image similarity measure, index *SSIM*, Weibull distribution

For citation. Starovoitov V. V. Enhancement of the structural similarity index *SSIM*. *Informatics*, 2018, vol. 15, no. 3, pp. 41–55 (in Russian).

Введение. В настоящее время подавляющее большинство глобальных оценок качества цифровых изображений вычисляется как среднее арифметическое множества локальных оценок. Вычисление индекса структурного сходства *SSIM* (от англ. Structural SIMilarity index) служит одним из популярных и широко цитируемых методов измерения сходства между двумя изображениями [1]. Индекс *SSIM* количественно описывает совпадение двух сравниваемых

изображений, хотя в англоязычной литературе его часто называют метрикой оценки качества изображений, имея в виду сравнение изображения с эталоном. Под эталоном подразумевается одно из сравниваемых изображений. Следует отметить, что данная функция метрикой не является, поскольку она не удовлетворяет метрическим аксиомам. Качество изображений этой функцией также не оценивается, поскольку сравнение низкокачественного изображения с самим собой дает максимальное значение индекса $SSIM = 1$, что должно свидетельствовать о его высоком качестве.

Отличительным признаком описанного подхода является то, что он учитывает структурные различия в паре изображений и наличие сильной корреляции в цвете соседних пикселей. Вычисление сходства выполняется в небольших окнах. Такие оценки сходства назовем локальными и обозначим $ssim$. Каждая из них оценивает яркость, контраст и структуру двух изображений в окне. На заключительном этапе в качестве глобального значения индекса $SSIM$ берется среднее арифметическое множества локальных оценок $ssim$.

Во всех исследованиях индекса $SSIM$ выполнялись сравнения изображений, представляющих одинаковые сцены с небольшими яркостными искажениями. Никто и никогда не применял этот индекс для сравнения изображений разных сцен или изображений, искаженных геометрически.

Для подобных сравнительных экспериментов используются специально созданные базы изображений. Одна из них, самая большая на данный момент, называется TID2013 [2]. Она содержит 25 эталонных изображений и 3000 их искаженных вариантов. Дополнительно в базе собраны 3000 усредненных субъективных оценок, выставленных экспертами для искаженных изображений, относительно эталонов. Такие оценки называют MOS (от англ. Mean Opinion Score). Все исследователи сравнивают субъективные оценки MOS с различными количественными оценками путем вычисления для них коэффициентов корреляции трех типов: Пирсона, Спирмена и Кендела, стремясь получить максимальную корреляцию.

В некоторых публикациях (см., например, [3, 4]) показано, что распределение Вейбулла хорошо описывает такие характеристики изображений естественных сцен, как контраст и текстура. Это может быть использовано для сегментации и классификации данных изображений. В настоящей статье исследуется возможность описания множества локальных оценок $ssim$ параметрами распределения Вейбулла и замены ими индекса $SSIM$, являющегося средним значением локальных оценок $ssim$.

Функция $SSIM$ и ее свойства

Определение функции $SSIM$. Вычисление локальных оценок $ssim$ в окнах x и y размером $N \times N$ пикселей выполняется по формуле [1]

$$ssim(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (1)$$

где μ_x – среднее значение в окне x ; μ_y – среднее значение в окне y ; σ_x^2 – дисперсия значений в окне x ; σ_y^2 – дисперсия значений в окне y ; σ_{xy} – ковариация в окнах x и y ; константы $c_1 = (k_1L)^2$ и $c_2 = (k_2L)^2$, $L = 255$, – значения динамического диапазона допустимых значений яркости; $k_1 = 0,01$, $k_2 = 0,03$ – экспериментально определенные константы.

Итоговый индекс $SSIM$ для изображений A и B вычисляется как среднее арифметическое локальных оценок по формуле

$$SSIM(A, B) = \frac{1}{M} \sum_{x,y} ssim(x, y),$$

где A и B – два изображения равных размеров, M – количество окон. Локальные оценки $ssim(x, y)$ можно вычислять в окрестности каждого пикселя или с некоторым шагом.

Фактически индекс $SSIM$ сконструирован из трех компонентов: оценки локальной яркости l , оценки локального контраста c и оценки локальной структуры s – посредством вычисления коэффициента корреляции Пирсона:

$$l(x, y) = \frac{2\mu_x \mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1},$$

$$c(x, y) = \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2},$$

$$s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x \sigma_y + c_3},$$

где $c_3 = c_2/2$.

В более общем виде локальная оценка сходства изображений в окрестности пиксела вычисляется по формуле

$$ssim(x, y) = l(x, y)^{p_1} \cdot c(x, y)^{p_2} \cdot s(x, y)^{p_3}, \quad (2)$$

где p_i – весовые коэффициенты. Если эти коэффициенты равны единице, формула (2) превращается в (1). Чаще всего на практике применяется формула (1).

При обобщении *SSIM* для изображений разных размеров можно применять окна x и y различных размеров пропорционально размерам изображений сцены, но с привязкой к объективно определенным центрам окон.

В некоторых приложениях требуется использовать неотрицательные локальные оценки. В таких случаях индекс *SSIM* может быть преобразован в нормализованное представление по формуле

$$NSSIM = (SSIM + 1) / 2. \quad (3)$$

Диапазон значений *NSSIM* будет в пределах $[0, +1]$.

Метрические свойства *SSIM*. В литературе индекс *SSIM* часто называют метрикой, однако он не является метрикой, поскольку $SSIM(A, A) = 1$, а не нулю. Кроме того, $SSIM(A, B)$ может иметь отрицательные значения. Следует отметить, что $SSIM(A, B) = SSIM(B, A)$.

Функция *NSSIM* также не является метрикой, так как для нее не выполняется аксиома неравенства треугольника. Это легко доказать через контрпример. Возьмем два изображения $A = 'Im01-01-2'$ и $B = 'Im01-01-1'$ из базы TID2013 и построим третье в виде негативной копии A , т. е. $C = 255 - 'Im01-01-2'$, тогда

$$NSSIM(A, B) = 0,9611 > NSSIM(A, C) + NSSIM(C, B) = 0,1045 + 0,1402.$$

В работе [3] утверждается, что компоненты функции *SSIM* можно преобразовать в две метрические функции:

$$S_1(x, y) = \sqrt{1 - l(x, y)}, \quad S_2(x, y) = \sqrt{1 - c(x, y)s(x, y)}.$$

Диапазон значений индекса *SSIM* лежит в пределах от -1 до 1 . Значение 1 достигается при полном совпадении сравниваемых изображений. Теоретически значение -1 должно достигаться при сравнении изображения A и его полной противоположности, однако это условие не выполняется. Анализ формул (1) и (2) показывает стремление к неопределенности индекса *SSIM* при сравнении, например, абсолютно белого и черного изображений. Чтобы избежать этого, введены положительные константы c_i . Функция *ssim*, реализованная в системе Матлаб, для таких изображений выдает ответ $SSIM(A, B) = 0$. Сравнение произвольных цветных изображений с их негативными копиями (т. е. пар изображений с меньшими отличиями) дает отрицательные значения *ssim*, но больше -1 . Найти пример пары изображений A и B , для которых $SSIM(A, B) = -1$, не удалось.

Связь *SSIM* с другими функциями. В работе [5] выполнен аналитический анализ индекса *SSIM*. Показано, что при равных значениях $\mu_x = \mu_y$ и $\sigma_x = \sigma_y$ индекс $SSIM = \rho$, т. е. индекс

упрощается до коэффициента корреляции Пирсона. Отметим, что на практике это маловероятная ситуация. Если один из четырех указанных параметров равен нулю, значение индекса *SSIM* также стремится к нулю, но его поведение нестабильно, поэтому в уравнениях (1), (2) добавлены небольшие константы.

В работе [6] показано, что пиковое отношение сигнала к шуму *PSNR* имеет аналитическую связь с индексом *SSIM*. Напомним, что для двух изображений *A* и *B*

$$PSNR(A, B) = 10 \log_{10} (255^2 / MSE(A, B)), \quad MSE(A, B) = \frac{1}{N} \sum_i^N (a_i - b_i)^2,$$

где *N* – число пикселей в каждом из изображений *A* и *B*; *a_i* и *b_i* – значения яркости пикселей.

Отношение *PSNR* можно переписать в виде

$$PSNR(A, B) \sim 10 \log_{10} \left[\frac{2\sigma_{xy}(l(A, B) - SSIM(A, B))}{255^2 SSIM(A, B)} + \left(\frac{\mu_A - \mu_B}{255} \right)^2 \right].$$

Функция *l(x, y)* описана выше. Выражение *PSNR* может быть упрощено для практических приложений, так как для многих типов искажений изображений значение *l(A, B) > 0,991*, т. е. примерно равно единице. Экспериментально в работе [4] показано, что для изображений, искаженных в результате применения фильтра Гаусса, а также сжатия алгоритмами JPEG и JPEG2000, оценки *PSNR* и *SSIM* при сравнении с оригиналами изображений имеют описанную выше нелинейную зависимость.

Существует класс функций, вычисляющих различия в цифровых изображениях. Методика их построения подробно описана в монографии [7]. Было показано, что оценка *MSE* является частным случаем этой методики. Основным способом анализа различий в 1990-е гг. было сравнение яркостей соответствующих пикселей. Особенностью предложенной методики был учет не только яркостей пикселей с одинаковыми координатами, но и их ближайшей окрестности. К сожалению, данная методика не получила широкого применения в построении мер сравнения изображений.

Индекс *SSIM* можно преобразовать в индекс структурного отличия *DSSIM* (от англ. structural DiSSIMilarity), который выражается через индекс *SSIM*:

$$DSSIM_1 = (1 - SSIM) / 2 \quad \text{либо} \quad DSSIM_2 = 1 - SSIM.$$

DSSIM_i(A, B) равен нулю при полном совпадении сравниваемых изображений (*A=B*) и не имеет отрицательных значений. Диапазон его значений от 0 до +1 в первом случае и до +2 во втором.

Функции *DSSIM_i* не являются метриками, так как для них не выполняется аксиома неравенства треугольника. Это легко доказать через контрпример. Возьмем изображение *A = 'Im02-01-2'* из базы TID2013 и построим два других: *B = A-1* (яркость в *A* уменьшаем на единицу) и негативную копию *A*, т. е. *C = 255 - A*. Тогда верно равенство

$$DSSIM_1(A, B) = 0,8605 > DSSIM_1(A, C) + DSSIM_1(C, B) = 0,000\,054\,561 + 0,8539.$$

Варианты модификации индекса *SSIM*

Использование разных весовых коэффициентов. В статье [8] представлен вариант подбора лучшего множества коэффициентов *p_i* для формулы (2). Автор выполнил достаточно большой перебор вариантов на изображениях одной из стандартных баз TID2008 и предложил использовать следующий набор коэффициентов: *p₁ = 0,61*, *p₂ = 0,077*, *p₃ = 0,241*. Однако применение этого набора коэффициентов в функции *ssim*, реализованной в Матлаб, дает в результате сравнения изображений разных сцен комплексные значения. Скорее всего, для другого набора изображений эти коэффициенты будут иными, т. е. указанные коэффициенты не являются оптимальными.

Мультимасштабный индекс MSSIM. Авторы индекса SSIM предложили его усовершенствованный мультимасштабный вариант [9]:

$$MSSIM(x, y) = [I_M(x, y)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(x, y)]^{\beta_j} [s_j(x, y)]^{\gamma_j}.$$

Для сравниваемых изображений строятся пирамиды Гаусса из M уровней (обычно $M = 5$) и для каждого уровня вычисляются компоненты c_j и s_j . Компонент I_M вычисляется только для яркости исходных изображений. Для упрощения вычислений авторы предлагают применять равные значения констант $\alpha_j = \beta_j = \gamma_j$ для всех уровней представления изображений и $\sum_{j=1}^M \gamma_j$.

Индекс SSIM вычисляется довольно долго, его модификация в виде индекса MSSIM вычисляется еще дольше при незначительном увеличении ранговых коэффициентов корреляции между величинами MSSIM и MOS, полученными на больших множествах изображений, поэтому индекс MSSIM практически не используется.

Интеграция разных оценок с помощью нейронных сетей. В статье [10] предлагается использовать искусственные нейронные сети для повышения количественной оценки с визуальной оценкой качества изображений MOS. Сформулированы требования к набору функций, используемому для вычисления входных данных сети:

- отбираются наиболее точные оценочные функции с максимальными коэффициентами ранговой корреляции относительно субъективных оценок MOS, выставленных экспертами;
- выбранные функции не должны использовать подобные признаки и иметь низкую взаимную корреляцию оценок одинаковых изображений.

Были выполнены эксперименты с шестью типами нейронных сетей и 38 вариантами выбора нейронов. В результате авторы построили интегрированную функцию оценки качества изображений путем отбора пяти лучших оценочных функций на базе четырехуровневой каскадной нейронной сети, что позволило повысить значения ранговых коэффициентов корреляции Спирмена и Кендалла на 0,07 по сравнению с индивидуальными результатами оценочных функций.

Анализ, выполненный в работе [10], показал, что разные функции по-разному оценивают искажения различных типов. Данное направление исследований перспективно, но для дальнейшего повышения точности оценок качества требуется объединять большее число оценочных функций, новые признаки и дополнительно учитывать особенности человеческого восприятия изображений.

Описание отдельных характеристик параметрами их статистических распределений. В статьях [3, 4] показано, что некоторые локальные характеристики изображений, например контраст, текстуру и др., можно описывать распределением Вейбулла и обобщенным нормальным распределением. В частности, для 75–90 % реальных изображений их локальный контраст, оцениваемый через статистику градиентов, очень хорошо описывается двухпараметрическим распределением Вейбулла. В работе Шольте [11] утверждается, что реакция человеческого мозга хорошо коррелирует со статистикой Вейбулла при анализе естественных изображений.

Распределение Вейбулла имеет два параметра: форму c и масштаб b . Функция плотности распределения Вейбулла для переменных $x > 0$ записывается в виде

$$f(x) = \frac{c}{b} \left(\frac{c}{b}\right)^{c-1} e^{-\left(\frac{x}{b}\right)^c},$$

где $0 < b$ – параметр масштаба, $0 < c$ – параметр формы.

Такие характеристики распределения Вейбулла, как среднее μ_{Weib} , медиана $Median_{Weib}$ и мода $Mode_{Weib}$, не совпадают с одноименными характеристиками, вычисляемыми для множества данных. Они определяются по формулам

$$\mu_{Weib} = \lambda \Gamma(1+1/c),$$

$$Median_{Weib} = \lambda (\ln 2)^{1/c},$$

$$Mode_{Weib} = \lambda ((c - 1)/c)^{1/c} \text{ при } c > 1, \quad Mode_{Weib} = 0 \text{ при } c \leq 1,$$

где $\Gamma(y)$ – гамма-функция от y .

Как правило, характеристики распределения Вейбулла не равны друг другу, что имеет место для аналогичных характеристик нормального распределения.

Отметим, что для некоторых совокупностей данных x кривая плотности их распределений близка к кривой нормального распределения. Например, распределение Вейбулла при параметре формы $b = 3,44$ выглядит как функция плотности нормального распределения за исключением хвостов распределения (рис. 1), а при $b = 1,0$ распределение Вейбулла идентично экспоненциальному распределению [12].

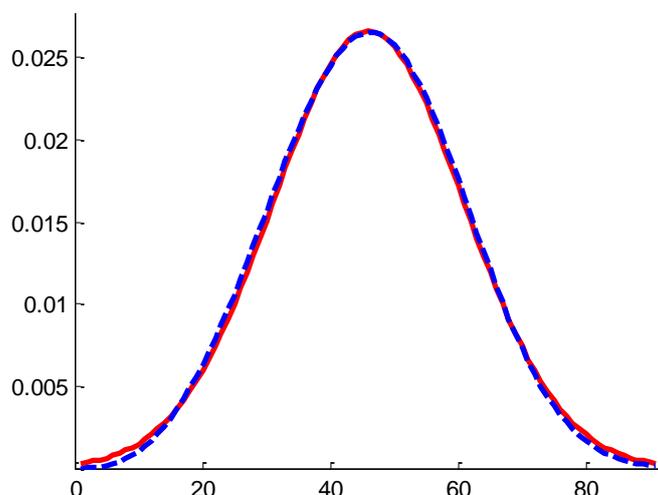


Рис. 1. График нормального распределения данных с параметрами $\mu = 45$, $\sigma = 15$ (красного цвета) и распределения Вейбулла с параметрами $c = 50$, $b = 3,44$ (синего цвета)

В рамках настоящего исследования была выдвинута следующая гипотеза: при формировании глобальной (единой) оценки сходства изображений вместо вычисления среднего значения множества локальных оценок лучше использовать параметр формы кривой распределения, описывающего гистограмму этого множества. Поскольку распределение Вейбулла и некоторые другие не работают с нулевыми и отрицательными данными, все локальные оценки *ssim* были нормализованы по формуле (3).

Гистограммы локальных оценок были аппроксимированы разными распределениями. Распределение Вейбулла оказалось самым гибким и лучше других описывало асимметричные данные. Далее анализировалось, что следует выбрать в качестве новой оценки совокупности локальных данных: параметр масштаба или формы кривой Вейбулла либо среднее, медиану, моду, коэффициенты асимметрии и эксцесса распределения Вейбулла. Согласно исследованиям У. А. Шухарта [13] и Д. Вилера [14] последние две характеристики бесполезны при анализе реальных данных, особенно имеющих экстремальные значения. В рассматриваемом случае некоторые локальные оценки всегда принимают крайние значения допустимого диапазона, например 1 или 0, поэтому коэффициенты асимметрии и эксцесса были отброшены. После первичных экспериментов параметр формы c также был отвергнут, поскольку его значение изменялось от 0 и до $+\infty$ для идентичных изображений, что усложняло получение финальных оценок.

Индекс *SSIM* как оценка качества изображений

База TID2013. База содержит 25 эталонных изображений, 24 из них взяты из базы изображений Kodak (<http://r0k.us/graphics/kodak/>), а 25-е синтезировано [2]. Ко всем изображениям применены 24 вида искажений (таблица), из них 12 типов искажений могут возникать при ре-

гистрации реальных изображений. Для каждого типа искажений создано пять вариантов изображений, которые имеют значения пикового отношения сигнал-шум $PSNR$, примерно равные 21, 24, 27, 30 и 33дБ. Всего в базе содержится 3000 искаженных цветных изображений размером 512×384 пиксела. Дополнительно для каждого искаженного изображения были получены субъективные оценки MOS его сравнения с эталоном в шкале [0, 9]. Для этого привлекали 971 эксперта из четырех стран.

Типы искажений изображений в базе TID2013

Тип искажения	Операции, при которых возникает
1. Аддитивный гауссовский шум	Регистрация изображений
2. Аддитивный шум в цветовых компонентах	Регистрация изображений
3. Пространственно-коррелированный шум	Цифровая фотография, сжатие
4. Маскированный шум	Сжатие
5. Высокочастотный шум	Сжатие
6. Импульсный шум	Регистрация изображений
7. Шум квантования	Регистрация изображений, гамма-коррекция
8. Гауссовское размытие	Регистрация изображений
9. Фильтрация шума	Шумоподавление
10. Сжатие алгоритмом JPEG	Сжатие
11. Сжатие алгоритмом JPEG2000	Сжатие
12. Сбои в канале передачи JPEG-изображений	Передача данных
13. Искажения при передаче JPEG2000-изображений	Передача данных
14. Искажения с низкой степенью оригинальности	Сжатие
15. Локальные блочные искажения	Регистрация изображений
16. Изменения яркости методом mean-shift	Регистрация изображений
17. Изменения контраста	Регистрация изображений, гамма-коррекция
18. Изменения насыщенности цвета	Регистрация изображений, сжатие
19. Мультипликативный гауссовский шум	Регистрация изображений, шумоподавление
20. Комфортный шум	Сжатие
21. Сжатие с потерями зашумленных изображений	Сжатие, шумоподавление
22. Квантование цветного изображения с дизерингом	Сжатие
23. Хроматические аберрации	Регистрация изображений
24. Восстановление изображений по разреженной сетке отсчетов	Сжатие, восстановление

Анализ корреляции величин $SSIM$ и MOS . Практически все исследователи оценивают корректность оценок сходства или качества изображений, полученных различными количественными функциями, степенью их корреляции с субъективными оценками MOS , которые обычно прикладываются к базам изображений. База TID2013 не является исключением. Для ее изображений коэффициенты ранговой корреляции Спирмена и Кендалла между величинами $SSIM$ и MOS согласно [2] равны 0,6370 и 0,4636, а между $MSSIM$ и MOS – 0,7872 и 0,6079 соответственно. Проверка дала немного отличающиеся значения корреляции $SSIM$ и MOS . Так, коэффициенты ранговой корреляции Спирмена и Кендалла при использовании стандартных функций системы Матлаб равнялись 0,6194 и 0,4391, а коэффициент корреляции Пирсона равнялся 0,5776.

На рис. 2 показана связь между величинами $SSIM$ и MOS для каждого из 3000 изображений базы TID2013. Красная прямая с уравнением $y = 5x + 0,2$ соответствует линейной аппроксимации этих оценок, зеленая кривая с уравнением $y = 10x^2 - 10x + 5,7$ – квадратичной аппроксимации. Приведенный график и коэффициенты корреляции указывают на слабую связь индекса $SSIM$ и субъективных оценок качества изображений MOS из базы TID2013.

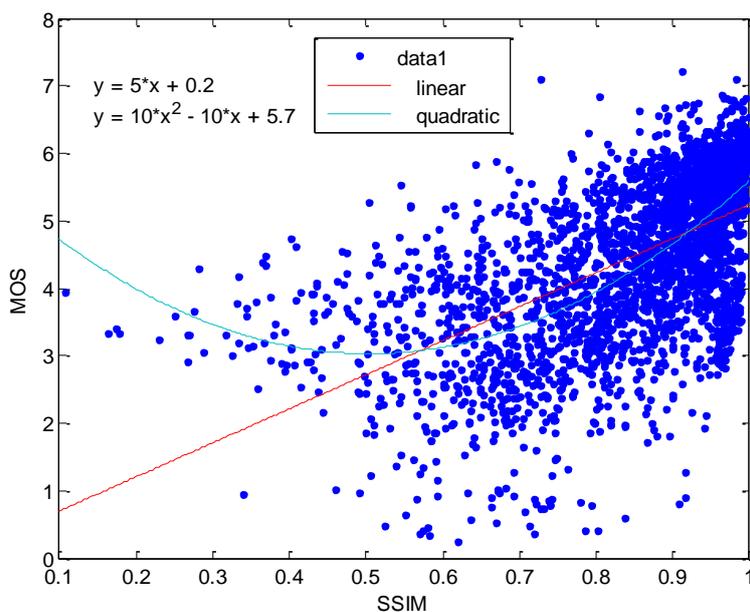


Рис. 2. График соответствия оценок *SSIM* и *MOS*, его аппроксимация прямой и параболой

В базе TID2013 есть изображения с локальными блочными искажениями (15-й тип искажений). Эти искажения формируются в результате наложения на оригинальные изображения небольших квадратных блоков одинакового размера и фиксированного цвета. В пяти вариантах искажений используются 10, 8, 6, 4 и 2 блока. Интересно, что многие эксперты поставили таким изображениям очень низкие оценки – менее четырех, а некоторым – менее трех из девяти возможных. При этом подавляющее большинство различных количественных оценочных функций, включая *SSIM*, подтвердило близость искаженных изображений к эталонным.

Можно констатировать, что оценки *MOS* изображений очень субъективны. При изменении группы экспертов такие оценки могут изменяться, и достаточно существенно. При этом процесс их получения достаточно долгий и затратный. Таким образом, они не могут служить в качестве эталонных оценок.

Экспериментальные исследования. В экспериментах использовались нормализованные по формуле (3) значения локальных оценок функции *ssim*, а для аппроксимации их распределения применялась функция *fitdist*. Обе функции реализованы в системе Матлаб. Их вычислительная сложность не оценивалась, но время выполнения функции *fitdist* было в 25 раз меньше, чем *ssim*.

Для пары изображений на рис. 3 среднеквадратичные отклонения теоретических кривых Гаусса и Вейбулла равны $MSE_{\text{Norm}} = 0,014\ 68$ и $MSE_{\text{Weib}} = 0,002\ 16$ соответственно, из чего следует, что распределение Вейбулла точнее аппроксимирует гистограмму локальных оценок. Это видно на рис. 4: кривой красного цвета показано приближение нормальным распределением, а синим – распределением Вейбулла. Гистограммы локальных оценок (рис. 4, 7, 9, 12) наглядно показывают, что при сравнении эталонных изображений с искаженными версиями многие из локальных оценок оказываются близки к единице. При этом гистограммы и их пики имеют явный сдвиг к единице, т. е. данные не распределены нормально. Такой же вывод подтверждает проверка распределения данных на нормальность, выполненная по критериям χ -квадрат, Колмогорова – Смирнова, Андерсона – Дарлинга, Лилиефорса.

Значение нормализованного индекса $SSIM_{\text{Norm}}$ для пары изображений на рис. 3 равно 0,7991, теоретический параметр масштаба b кривой Вейбулла равен 0,8386, а мода распределения Вейбулла равна 0,8308, т. е. они ближе к максимальному значению, равному единице, и близки между собой. Однако субъективная оценка сходства этих изображений достаточно низкая: $MOS = 2,742\ 86$. Данный факт подтверждает различие между рассматриваемыми субъективными и объективными оценками изображений.

Если сравнивать любое, даже максимально искаженное, изображение с самим собой, все локальные количественные оценки будут равны единице. Субъективная оценка будет зависеть от поставленной задачи – оценить сходство либо качество изображений.



Рис. 3. Изображение Im24 и его искаженный дизерингом вариант пятого уровня Im24-22-5

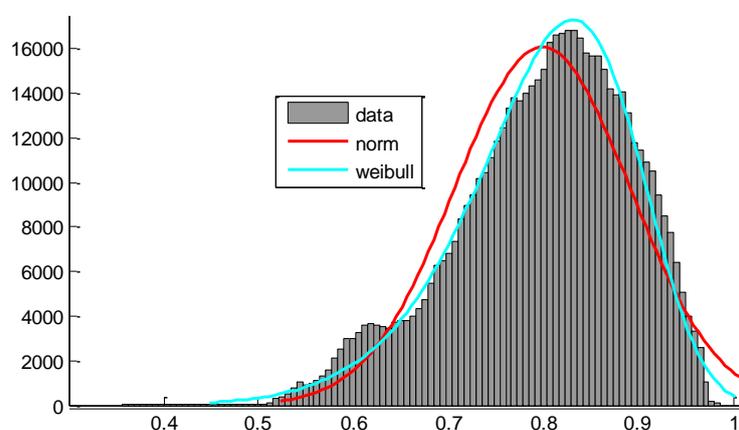


Рис. 4. Гистограмма локальных оценок *ssim* сходства изображений Im24, Im24-22-5 и ее приближение распределениями Гаусса и Вейбулла

В процессе исследования был выполнен анализ точности приближения гистограмм локальных оценок *ssim* распределениями Гаусса и Вейбулла. На рис. 5 показан график отсортированных разностей среднеквадратичного отклонения аппроксимации гистограмм локальных оценок *ssim* между эталонами и их искаженными копиями из базы TID2013 кривыми Гаусса и Вейбулла.

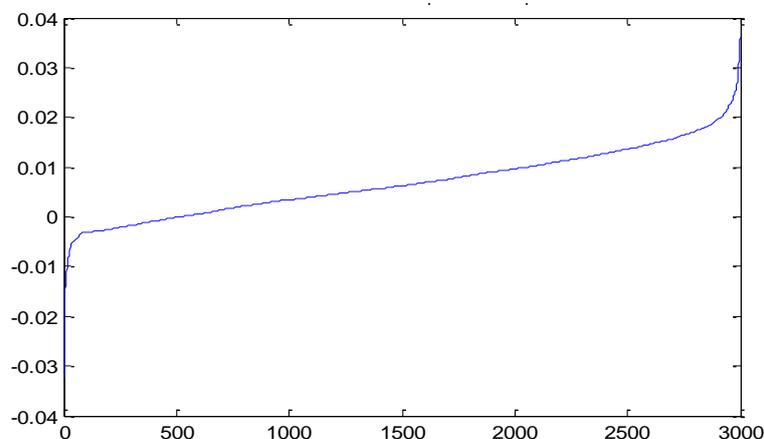


Рис. 5. График разностей среднеквадратичного отклонения аппроксимации гистограмм локальных оценок *ssim* функциями нормального распределения и Вейбулла

Минимальное значение разности равно $-0,0322$, максимальное – $+0,0362$. Для 514 искаженных изображений из 3000 (17,133 %) эта разность была отрицательной, т. е. для них нормальное распределение давало несколько лучшее приближение. Отсюда следует, что распределение Вейбулла в целом точнее нормального описывает совокупность локальных оценок *ssim* и его параметры можно использовать для формирования уточненного индекса *SSIM*.

Для полноты анализа было выполнено сравнение всех изображений, представленных в базе TID2013, друг с другом. Приведем несколько примеров (рис. 6–9).



Рис. 6. Эталонное изображение Im11 и его слабоконтрастная копия Im11-17-5

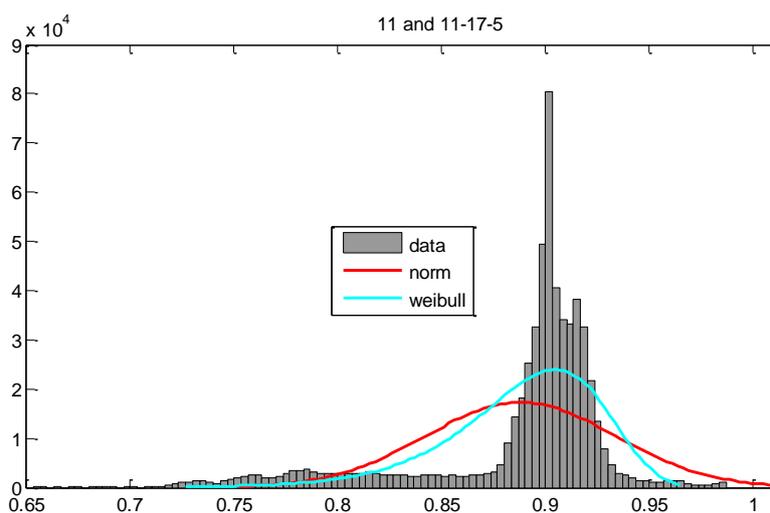


Рис. 7. Гистограмма нормализованных локальных оценок *ssim*(Im11, Im11-17-5) и ее аппроксимация функциями Гаусса и Вейбулла

Гистограммы на рис. 5 и 7 показывают, что при сравнении изображений одной сцены большинство локальных оценок близко к единице, т. е. локально изображения похожи. Однако гистограммы не симметричны относительно среднего значения, равного индексу *NSSIM*, а сдвинуты вправо. Их форма лучше описывается кривой Вейбулла, при этом среднеквадратичная погрешность такой аппроксимации меньше: $MSE_{Norm} = 0,052\ 09$, $MSE_{Weib} = 0,025\ 808$. Приведем оценки изображений, представленных на рис. 5: $ssim_{norm} = 0,8882$, $b = 0,9060$, $mode_{Weib} = 0,9049$, $MOS = 3,216\ 22$. Очевидно, что численные оценки больше отражают сходство данных изображений, а субъективная оценка *MOS* – качество искаженного изображения относительно эталонного.

Рассмотрим для примера оценки сравнения пары изображений Im01 и Im01-22-5 из базы TID2013 (рис. 8). Второе изображение – это копия первого после квантования цвета методом дизеринга. Усредненная субъективная оценка *MOS* искаженного изображения в качественной шкале [0–9] равна 2,1944.

На рис. 9 изображена гистограмма локальных оценок *ssim* пофрагментных отличий в данной паре изображений. Вершина красной кривой указывает на среднее значение множе-

ства локальных оценок, которое и является финальной оценкой индекса $NSSIM = 0,8372$. Синим цветом показано приближение данных кривой распределения Вейбулла с параметром масштаба $b = 0,8652$. Усредненная субъективная оценка MOS искаженного изображения равна 2,1944.



Рис. 8. Пара тестовых изображений Im01 и Im01-22-5 из базы TID2013

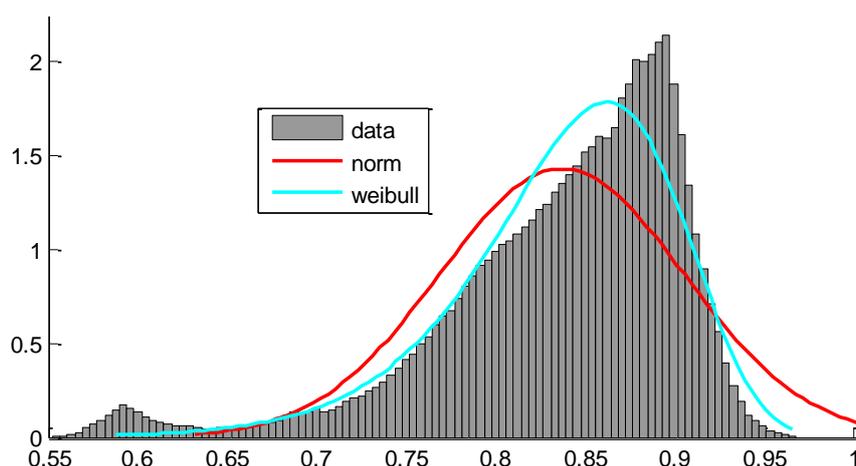


Рис. 9. Гистограмма локальных оценок $ssim$ пофрагментных отличий в паре изображений Im01 и Im01-22-5

На рис. 10 показаны отклонения теоретических кривых от реальных данных гистограммы, представленной на рис. 9, и среднеквадратичные отклонения для двух рассматриваемых распределений. Графики подтверждают, что кривая Вейбулла в данном примере имеет на порядок лучшее приближение, чем кривая Гаусса ($MSE_{Weib} = 0,001\ 7271$ против $MSE_{Gauss} = 0,017\ 174$).

Если убрать два десятка выбросов, то отклонения кривой Вейбулла приближенно описываются прямой, чего нельзя сказать о кривой Гаусса. Визуальный анализ пары изображений на рис. 8 показывает, что искаженное изображение очень похоже на оригинал, т. е. это изображение одной и той же сцены, но полного совпадения по цвету, текстуре и геометрическим деталям нет. Можно сделать вывод, что низкое значение MOS больше отражает субъективную оценку качества правого изображения по сравнению с оригиналом, а количественные оценки определяют сходство изображенной сцены. При этом параметр масштаба $b = 0,8652$ точнее, чем $NSSIM = 0,8372$, в целом характеризует локальные оценки, представленные на рис. 9.

Кроме сравнения искаженных изображений с их эталонами было выполнено выборочное сравнение изображений разных сцен и разного качества из базы TID2013 между собой, а также

этих изображений с их негативными копиями. Количество всех таких пар изображений очень велико ($6050 \times 3025 = 18\,301\,250$), поэтому получить для них оценки *MOS* практически невозможно.

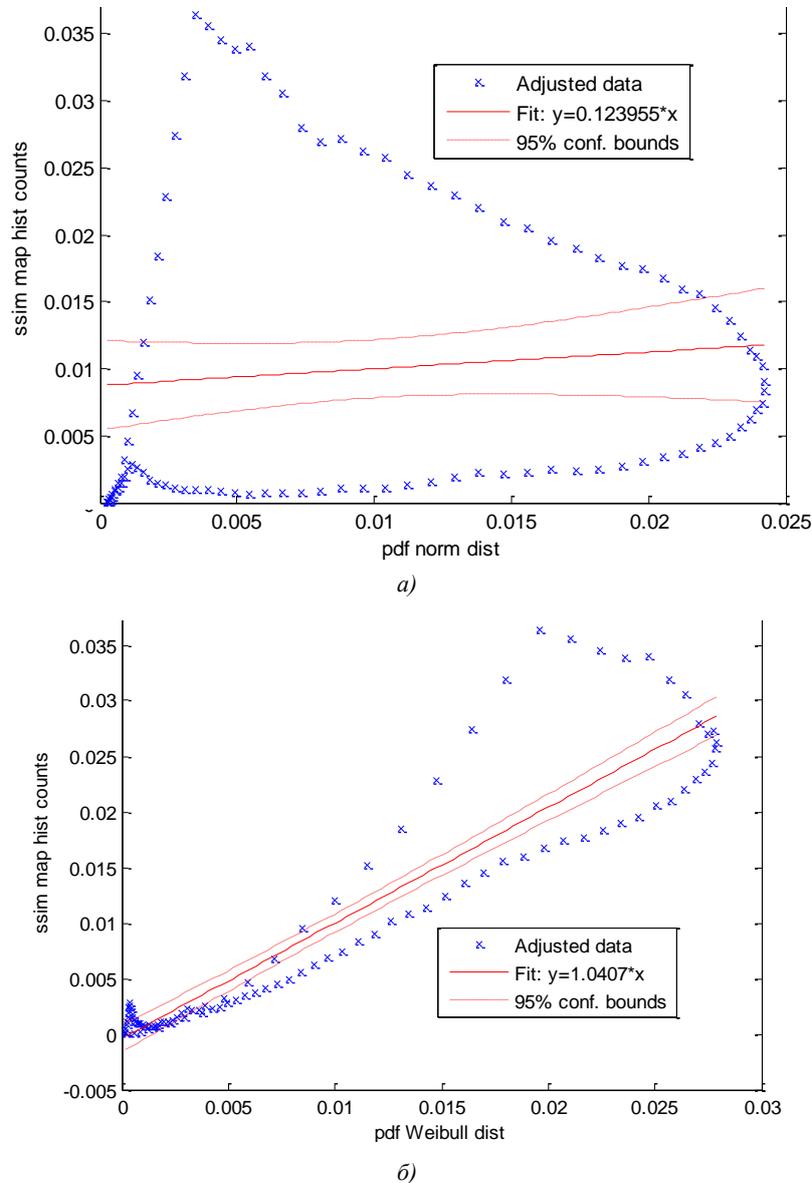


Рис. 10. Отклонения теоретических кривых Гаусса (а) и Вейбулла (б) от реального распределения локальных оценок *ssim*

Анализ результатов сравнения ряда пар разнообразных изображений подтвердил вывод о более точной аппроксимации гистограммы локальных оценок *ssim* распределением Вейбулла. На рис. 11 приведен пример пары изображений с разными типами гауссовского шума. Визуально они практически не отличимы, но все локальные оценки *ssim* (рис. 12) несколько меньше единицы, что означает несовпадение на локальном уровне, а их гистограмма не имеет нормального распределения. Глобальная оценка сходства, полученная из распределения Вейбулла, равна 0,9815, индекс $SSIM_{\text{Norm}} = 0,9732$. Среднеквадратичное отклонение кривой Вейбулла от графика гистограммы локальных оценок меньше, чем у кривой Гаусса: 0,0540 против 0,0676. Отметим, что коэффициент корреляции Пирсона между оценками $SSIM_{\text{Norm}}$ и параметром масштаба b распределения Вейбулла, полученный для 3000 изображений базы TID2013, равен 0,9923. Это говорит о их близости.

Значение параметра моды распределения Вейбулла во всех экспериментах было очень близко к значению параметра масштаба b этого распределения. Для уменьшения вычислений в качестве глобальной меры сходства двух изображений вместо среднего от множества локальных оценок $ssim$ предлагается использовать параметр масштаба b распределения Вейбулла, которое аппроксимирует гистограмму этих локальных оценок.



а)

б)

Рис. 11. Изображение Im01-01-1 с аддитивным гауссовским шумом (а) и оно же с мультипликативным гауссовским шумом (б)

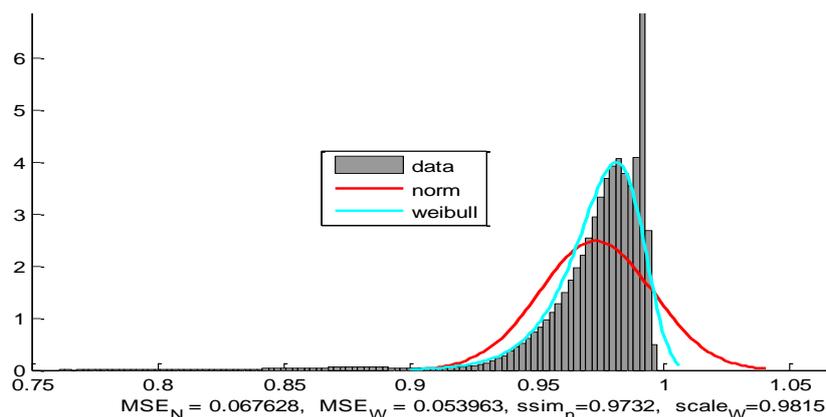


Рис. 12. Гистограмма локальных оценок $ssim$ пофрагментных отличий в паре изображений Im01-01-1 и Im01-19-1

Заключение. В работе исследованы свойства широко применяемого индекса структурного подобия изображений $SSIM$. Показано, что данная мера и варианты ее нормализации не являются метрическими функциями. По результатам исследований можно сделать вывод, что локальные количественные оценки $ssim$ характеризуют сходство изображений, а субъективная оценка MOS делает акцент на визуальные отличия искаженного изображения.

Были исследованы также варианты аппроксимации гистограммы локальных оценок сходства разными распределениями. Распределение Вейбулла оказалось самым гибким. Из трех характеристик этого распределения: параметра масштаба b , параметра формы c и медианы теоретической кривой Вейбулла – предложено в качестве глобальной меры сходства двух изображений вместо популярного индекса $SSIM$ использовать параметр масштаба b распределения Вейбулла. Показано, что значение этого параметра, вычисленное для множества локальных оценок $ssim$, точнее, чем их среднее арифметическое, отражает сходство или различие двух изображений. Для множества нормализованных локальных оценок $ssim$ данный параметр всегда будет находиться в диапазоне $[0, 1]$ и равняться единице, если изображения идентичны.

Работа частично выполнена в рамках проекта БРФФИ Ф18МС-028.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Image quality assessment: from error visibility to structural similarity / Z. Wang [et al.] // *IEEE Transactions on Image Processing*. – 2004. – Vol. 13, no. 4. – P. 600–612.
2. Image database TID2013: peculiarities, results and perspectives / N. Ponomarenko [et al.] // *Signal Processing: Image Communication*. – 2015. – Vol. 30. – P. 57–77.
3. Geusebroek, J. A six-stimulus theory for stochastic texture / J. Geusebroek, A. W. M. Smeulders // *International Journal of Computer Vision*. – 2005. – Vol. 62, no. 1–2. – P. 7–16.
4. Xue, W. Reduced reference image quality assessment based on Weibull statistics / W. Xue, X. Mou // *Proc. of the 2nd Intern. Workshop on Quality of Multimedia Experience*. – N.Y., 2010. – P. 11–16.
5. Brunet, D. On the mathematical properties of the structural similarity index / D. Brunet, E. R. Vrscay, Z. Wang // *IEEE Transactions on Image Processing*. – 2012. – Vol. 21, no. 4. – P. 1488–1499.
6. Hore, A. Image quality metrics: PSNR vs. SSIM / A. Hore, D. Ziou // *Proc. of the 20th Intern. Conf. on Pattern Recognition*. – Washington, 2010. – P. 2366–2369.
7. Старовойтов, В. В. Локальные геометрические методы цифровой обработки и анализа изображений / В. В. Старовойтов. – Минск : Ин-т техн. кибернетики НАН Беларуси, 1997. – 284 с.
8. Сидоров, Д. В. Модификация алгоритма SSIM / Д. В. Сидоров // *Прикладная информатика*. – 2010. – № 4. – С. 123–125.
9. Wang, Z. Multiscale structural similarity for image quality assessment / Z. Wang, E. P. Simoncelli, A. C. Bovik // *Proc. of the 37th Asilomar Conf. on Signals, Systems and Computers*. – USA, CA, 2004. – Vol. 2. – P. 1398–1402.
10. Еремеев, О. И. Интегрированная метрика визуального качества изображений при наличии эталона / О. И. Еремеев // *Системы обработки информации*. – 2014. – № 5. – С. 35–42.
11. Brain responses strongly correlate with Weibull image statistics when processing natural images / H. S. Scholte [et al.] // *Journal of Vision*. – 2009. – Vol. 9, no. 29. – С. 11–25.
12. Статистические методы. Распределение Вейбулла. Анализ данных : ГОСТ Р 50779.27–2017. – Введ. 10.08.2017. – М. : Госстандарт России : Изд-во стандартов, 2017. – 62 с.
13. Shewhart, W. A. *Statistical Method from the Viewpoint of Quality Control* / W. A. Shewhart. – Washington : Courier Corporation, 1939. – 155 p.
14. Wheeler, D. J. Problems with Skewness and Kurtosis. Part II. What do the shape parameters do? [Electronic resource] / D. J. Wheeler // *Quality Digest Daily*. – 2011. – Mode of access: <https://www.spcpress.com/pdf/DJW231.pdf>. – Date of access: 20.05.2018.

References

1. Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, vol. 13, no. 4, pp. 600–612.
2. Ponomarenko N., Jin L., Ieremeiev O., Lukin V., Egiazarian K., Kuo C. Image database TID2013: peculiarities, results and perspectives. *Signal Processing: Image Communication*, 2015, vol. 30, pp. 57–77.
3. Geusebroek J., Smeulders A. W. M. A six-stimulus theory for stochastic texture. *International Journal of Computer Vision*, 2005, vol. 62, no. 1–2, pp. 7–16.
4. Xue W., Mou X. Reduced reference image quality assessment based on Weibull statistics. *Proceedings of the 2nd International Workshop on Quality of Multimedia Experience*. N.Y., 2010, pp. 11–16.
5. Brunet D., Vrscay E. R., Wang Z. On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing*, 2012, vol. 21, no. 4, pp. 1488–1499.
6. Hore A., Ziou D. Image quality metrics: PSNR vs. SSIM. *Proceedings of the 20th International Conference on Pattern Recognition*. Washington, 2010, pp. 2366–2369.
7. Starovoitov V. V. Lokal'nye geometricheskie metody cifrovoj obrabotki i analiza izobrazhenij. *Local Geometric Methods of Digital Image Processing and Analysis*. Minsk, In-t teh. kibernetiki NAN Belarusi Publ., 1997, 284 p. (in Russian).
8. Sidorov D. V. Modifikacija algoritma SSIM [SSIM algorithm modification]. *Prikladnaja informatika [Applied Informatics]*, 2010, no. 4, pp. 123–125 (in Russian).
9. Wang Z., Simoncelli E. P., Bovik A. C. Multiscale structural similarity for image quality assessment. *Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers*. USA, CA, 2004, vol. 2, pp. 1398–1402.
10. Eremeev O. I. Integrirovannaja metrika vizual'nogo kachestva izobrazhenij pri nalichii jetalona [Integrated metric of visual quality of images in the presence of a standard]. *Sistemy obrabotki informacii [Information Processing Systems]*, 2014, no. 5, pp. 35–42 (in Russian).
11. Scholte H. S., Ghebreab S., Waldorp L., Smeulders A. W., Lamme V. A. Brain responses strongly correlate with Weibull image statistics when processing natural images. *Journal of Vision*, 2009, vol. 9, no. 29, pp. 11–25.
12. Statisticheskie metody. Raspreделение Vejbullla. Analiz dannyh. *State Standart R 50779.27–2017. Statistical methods. Weibull distribution. Data analysis*. Moscow, Gosstandart Rossii, Izd-vo standartov Publ., 2017, 62 p. (in Russian).
13. Shewhart W. A. *Statistical Method from the Viewpoint of Quality Control*. Washington, Courier Corporation, 1939, 155 p.
14. Wheeler D. J. Problems with Skewness and Kurtosis. Part II. What do the shape parameters do? *Quality Digest Daily*, 2011. Available at: <https://www.spcpress.com/pdf/DJW231.pdf> (accessed 20.05.2018).

Информация об авторе

Старовойтов Валерий Васильевич – доктор технических наук, профессор, главный научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси (ул. Сурганова, 6, 220012, Минск, Республика Беларусь).

E-mail: valerys@newman.bas-net.by

Information about the author

Valery V. Starovoitov – D. Sc. (Engineering), Chief Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Sorganova Str., 220012, Minsk, Republic of Belarus).

E-mail: valerys@newman.bas-net.by

ISSN 1816-0301 (print)

ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ*LOGICAL DESIGN*

УДК 519.7

Поступила 24.04.2018

Received 24.04.2018

П. Н. Бибилло, Н. А. Кириенко, Ю. Ю. Ланкевич*Объединенный институт проблем информатики
Национальной академии наук Беларуси, Минск, Беларусь***ЛОГИЧЕСКАЯ ОПТИМИЗАЦИЯ МНОГОУРОВНЕВЫХ ПРЕДСТАВЛЕНИЙ
СИСТЕМ БУЛЕВЫХ ФУНКЦИЙ НА ОСНОВЕ БЛОЧНОГО РАЗБИЕНИЯ
И РАЗЛОЖЕНИЯ ШЕННОНА**

Аннотация. Описываются результаты экспериментального исследования эффективности применения процедур оптимизации систем булевых функций, предварительно выполняемых при синтезе комбинационных схем. Процедуры используют алгоритмы разбиения систем функций на подсистемы и алгоритмы оптимизации многоуровневых представлений функций на основе разложения Шеннона. Показывается, что разбиение исходной системы булевых функций на подсистемы и оптимизация подсистем на основе разложения Шеннона с нахождением инверсных подфункций, входящих в разложения (BDDI-оптимизация), во многих случаях позволяют при синтезе уменьшать площадь комбинационных схем из библиотечных элементов. Совместная BDDI-оптимизация является более предпочтительным методом по сравнению с отдельной технологически независимой BDDI-оптимизацией, так как площадь схем, построенных по совместным BDDI, в подавляющем числе случаев меньше площади схем, построенных по отдельным BDDI.

Ключевые слова: булева функция, дизъюнктивная нормальная форма, диаграмма двоичного выбора, разбиение системы булевых функций, синтез логической схемы, КМОП-технология

Для цитирования. Бибилло, П. Н. Логическая оптимизация многоуровневых представлений систем булевых функций на основе блочного разбиения и разложения Шеннона / П. Н. Бибилло, Н. А. Кириенко, Ю. Ю. Ланкевич // Информатика. – 2018. – Т. 15, № 3. – С. 56–70.

P. N. Bibilo, N. A. Kirienko, Y. Y. Lankevich*The United Institute of Informatics Problems of the National Academy
of Sciences of Belarus, Minsk, Belarus***LOGICAL OPTIMIZATION THE MULTILEVEL REPRESENTATIONS OF SYSTEMS OF BOOLEAN
FUNCTIONS BASED ON PARTITIONING INTO BLOCKS AND SHANNON DECOMPOSITION**

Abstract. The results of experimental study of the effectiveness of optimization procedures for systems of Boolean functions which are used in the synthesis of combinational circuits are described. The procedures use algorithms for partitioning systems of functions into subsystems and algorithms for optimizing multilevel representations of functions based on Shannon decomposition. The Shannon decomposition uses the procedure for finding inverse subfunctions, contained in decomposition result (BDDI-optimization). It is shown that these procedures can reduce the area of combinational circuits from library gates in many cases in the process of synthesis. Joint BDDI optimization is more preferable method in comparison to separate technologically independent BDDI optimization, since the area of circuits built on joint BDDI is smaller than the area of circuits built on separate BDDI in most cases.

Keywords: Boolean function, disjunctive normal form, binary decision diagram (BDD), partitioning systems of functions, synthesis of logical circuit, CMOS-technology

For citation. Bibilo P. N., Kirienko N. A., Lankevich Y. Y. Logical optimization the multilevel representations of systems of Boolean functions based on partitioning into blocks and Shannon decomposition. *Informatics*, 2018, vol. 15, no. 3, pp. 56–70 (in Russian).

Введение. Логическая оптимизация различных представлений систем полностью определенных булевых функций обычно применяется непосредственно перед этапом технологического отображения (синтеза логических схем), на котором оптимизированные представления булевых функций покрываются описаниями базисных логических элементов [1]. В настоящее время для оптимизации описаниями булевых функций исходными являются многоуровневые представления, получаемые после этапа высокоуровневого синтеза – замены алгоритмических конструкций на языках VHDL, Verilog [2] логическими уравнениями и функциональными описаниями элементов памяти. Данные представления цифровых устройств применяются при автоматизированном проектировании, заменяя используемые ранее в качестве исходных данных таблицы истинности и системы дизъюнктивных нормальных форм (ДНФ) [1, 3].

В зависимости от целевой библиотеки синтеза и вида (формы) исходного описания систем булевых функций применяются различные методы предварительной оптимизации. Например, использование программируемых логических матриц (ПЛМ) потребовало разработки методов оптимизации сетей из ПЛМ по критериям площади и методов декомпозиции «больших» ПЛМ на сети ПЛМ с ограниченными параметрами [4]. Появление FPGA вызвало новый интерес к методам декомпозиции систем булевых функций [5, 6] и методам, основанным на разложении Шеннона и других подобных разложениях [7]. Использование предварительной логической оптимизации и последующее выполнение синтеза в промышленных синтезаторах являются перспективными направлениями для улучшения такого важного параметра синтезируемых логических схем, как площадь кристалла.

Основные методы технологически независимой оптимизации базируются на разложении Шеннона и позволяют получать оптимизированные BDD (Binary Decision Diagram – диаграммы двоичных решений) [8–10], называемые также бинарными диаграммами, двоичными решающими диаграммами и диаграммами двоичного выбора [7]. BDD и их модификации [11] широко используются при решении самых разнообразных задач, возникающих в различных областях, в том числе синтезе, построении тестов и верификации цифровых схем [12]. Результаты исследования [13] показали, что разложения Шеннона для систем булевых функций, в которых используются как коэффициенты разложений, так и инверсии коэффициентов, оказались эффективным способом предварительной логической оптимизации при синтезе КМОП-схем из библиотечных элементов (вентилей). Однако исследования эффективности применения классических BDD выявили тот факт, что наряду с совместными классическими BDD могут успешно применяться и BDD, построенные для подсистем функций, либо отдельные BDD, построенные отдельно для каждой из функций системы.

Настоящая работа посвящена экспериментальному исследованию эффективности применения при синтезе логических схем из КМОП-элементов трех оптимизационных процедур: разбиения исходного функционального описания систем полностью определенных булевых функций на блоки, ограниченные по числу входных и выходных переменных; построения матричных форм систем ДНФ для каждого из блоков; BDD-оптимизации представления каждого из блоков с минимизацией числа коэффициентов разложений Шеннона с точностью до инверсий. Эксперименты проводились для различных форм представления исходных данных (функциональных описаний) и одной и той же библиотеки логических элементов с различными вариантами разбиения функциональных описаний на блоки.

Формы представления систем булевых функций. Под векторной булевой функцией $f(x)$, $x = (x_1, x_2, \dots, x_n)$, будем понимать упорядоченную систему булевых функций $f(x) = (f^1(x), \dots, f^m(x))$. Широко известной в литературе [3, 4] формой представления систем булевых функций является пара матриц $\langle T^x, B^f \rangle$: троичная матрица T^x задания элементарных конъюнкций в виде троичных векторов и булева матрица B^f вхождений конъюнкций в ДНФ компонентных функций системы. В табл. 1 дан пример задания системы ДНФ векторной полностью определенной функции $f(x) = (f^1(x), f^2(x), f^3(x))$:

$$\begin{aligned}
 f^1 &= x_1 x_2 x_3 x_4 \vee \bar{x}_2 \bar{x}_3 \bar{x}_4 \vee x_1 x_2 x_4, \\
 f^2 &= \bar{x}_1 \bar{x}_2 \bar{x}_3 x_4 \vee x_1 \bar{x}_2 \bar{x}_3 \bar{x}_4 \vee x_1 x_2 \bar{x}_3 x_4 \vee \bar{x}_1 x_2 x_4 \vee x_3 x_4 \vee x_2 x_3, \\
 f^3 &= \bar{x}_1 \bar{x}_2 \bar{x}_3 x_4 \vee x_1 \bar{x}_2 \bar{x}_3 \bar{x}_4 \vee x_1 x_2 \bar{x}_3 x_4 \vee x_1 x_2 x_3 \bar{x}_4 \vee \bar{x}_1 x_2 \bar{x}_4.
 \end{aligned} \tag{1}$$

Таблица 1

Пример системы ДНФ трех булевых функций

T^x	B^f
$x_1 \ x_2 \ x_3 \ x_4$	$f^1 \ f^2 \ f^3$
0 0 0 1	0 1 1
1 0 0 0	0 1 1
1 1 0 1	0 1 1
1 1 1 0	1 0 1
- 0 0 0	1 0 0
0 1 - 0	0 0 1
0 1 - 1	1 1 0
- - 1 1	0 1 0
- 1 1 -	0 1 0

Систему булевых функций (табл. 1) можно представить также в виде взаимосвязанных логических уравнений, которые являются формулами, полученными командой `unpar` в синтезаторе `LeonardoSpectrum` [14] после выполнения синтеза схемы по VHDL-описанию этой системы. Особенностью таких формул, называемых еще *RTL-описаниями*, является то, что каждая из них содержит только одну логическую операцию над парой булевых переменных либо операцию инверсирования одной булевой переменной. В языке SF многоуровневое представление в виде взаимосвязанных логических уравнений обозначается как формат LOG [15].

Оптимизация многоуровневых представлений

Многоуровневое разложение Шеннона с нахождением инверсных подфункций. Еще одной используемой в данной работе формой задания систем булевых функций являются алгебраические многоуровневые представления на базе разложения Шеннона с применением инверсных подфункций (коэффициентов разложения).

Разложением Шеннона полностью определенной булевой функции $f = f(\mathbf{x})$ по переменной x_i называется представление

$$f = f(\mathbf{x}) = \bar{x}_i f(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n) \vee x_i f(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n). \tag{2}$$

Функции $f_0 = f(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n)$, $f_1 = f(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n)$ в правой части (2) называются коэффициентами разложения по переменной x_i , остаточными подфункциями либо просто *подфункциями*. Они получаются из функции $f = f(x_1, \dots, x_n)$ путем подстановки вместо переменной x_i констант 0 и 1 соответственно. Каждая из подфункций $f_0 = f(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n)$ и $f_1 = f(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n)$ может быть разложена по одной из переменных из множества $\{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n\}$. Процесс разложения подфункций заканчивается, когда все n переменных будут использованы для разложения либо когда все подфункции вырождаются до констант 0, 1. На каждом шаге разложения выполняется сравнение на равенство полученных подфункций и оставляется одна из нескольких попарно равных (с точностью до инверсии) подфункций. Если же подфункции разложения по переменной x_i равны

$$f(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n) = f(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n), \quad (3)$$

то переменная x_i называется *несущественной* (фиктивной) и $f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) = f(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$.

Под BDDI (Binary Decision Diagram with Inverse cofactors) в настоящей работе понимается ориентированный ациклический граф, задающий последовательные разложения Шеннона булевой функции $f(x_1, \dots, x_n)$ по всем ее переменным x_1, \dots, x_n при заданном порядке (перестановке) переменных, по которым проводятся разложения. Граф BDDI одной полностью определенной булевой функции содержит три вида вершин: функциональные вершины, соответствующие разлагаемым функциям и подфункциям (и их инверсиям); вершины-переменные; листовые вершины, соответствующие константам 0, 1. Ориентация дуг обычно не показывается, так как при изображении BDDI все дуги ориентируются сверху вниз. Функциональная вершина BDDI реализует одну функцию φ (подфункцию) либо две функции (подфункцию φ и ее инверсию $\bar{\varphi}$) [13]. Смысл выражения «с точностью до инверсии» заключается в том, что любую из пары взаимно инверсных подфункций разложения можно считать неинверсной подфункцией, тогда другая подфункция такой пары будет инверсной к выбранной.

Далее для $f(x) = (f^1(x), \dots, f^m(x))$ будут рассматриваться BDDI двух видов – совместные и отдельные. *Совместными* BDDI, представляющими $f(x)$, будем называть такие BDDI, которые построены по общей для всех компонентных функций $f^i(x)$ перестановке переменных. Если же BDDI для каждой из компонентных функций системы строятся независимо, т. е. так, что каждая из функций разлагается по своей перестановке переменных, то такие BDDI для системы функций будем называть *отдельными*.

Для системы функций совместная BDDI (рис. 1) реализует следующее многоуровневое представление:

$$\begin{aligned} f^1 &= \bar{x}_2 s_0 \vee x_2 s_1, f^2 = \bar{x}_2 s_2 \vee x_2 s_0, f^3 = \bar{x}_2 s_3 \vee x_2 s_2, s_0 = \bar{x}_3 s_6, s_1 = \bar{x}_3 s_4 \vee x_3 s_5, \\ s_2 &= x_3 s_5 \vee x_3 s_6, s_3 = \bar{x}_3 s_5, s_4 = x_1 s_6, s_5 = \bar{x}_1 s_6 \vee x_1 s_6, s_6 = x_4. \end{aligned} \quad (4)$$

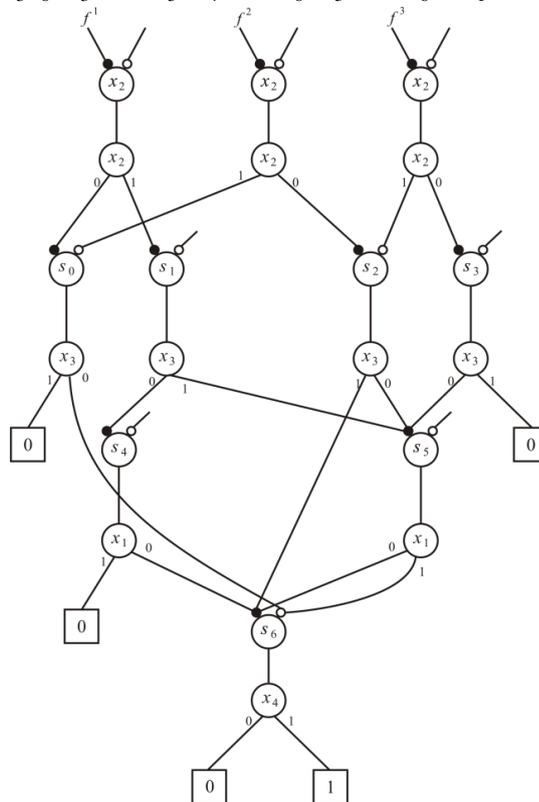


Рис. 1. Графическое представление уравнений (4)

Раздельные BDDI показаны на рис. 2. Функция f^1 (рис. 2, а) в форме BDDI имеет следующее многоуровневое представление:

$$f^1 = \bar{x}_1 a_0 \vee x_1 a_1, \quad a_0 = \bar{x}_2 a_2 \vee x_2 a_4, \quad a_1 = \bar{x}_2 a_2 \vee x_2 a_3, \quad a_2 = \bar{x}_4 \bar{a}_4, \quad a_3 = \bar{x}_4 a_4, \quad a_4 = x_3. \quad (5)$$

Функция f^2 (рис. 2, б) представлена в форме классической BDD [7] и реализует многоуровневое представление

$$f^2 = \bar{x}_2 b_0 \vee x_1 b_1, \quad b_0 = \bar{x}_3 b_2 \vee x_3 b_3, \quad b_1 = \bar{x}_3 b_1 \vee x_3, \quad b_2 = \bar{x}_4 b_4 \vee x_4 b_5, \quad b_3 = x_4, \quad b_4 = x_1, \quad b_5 = \bar{x}_1. \quad (6)$$

Функция f^3 (рис. 2, в) в форме BDDI имеет многоуровневое представление

$$f^3 = \bar{x}_1 c_0 \vee x_1 c_1, \quad c_0 = \bar{x}_3 c_2 \vee x_3 c_3, \quad c_1 = \bar{x}_3 \bar{c}_2 \vee x_3 c_3, \quad c_2 = \bar{x}_2 c_4 \vee x_2 \bar{c}_4, \quad c_3 = x_2 \bar{c}_4, \quad (7)$$

$$c_4 = x_4, \quad b_5 = \bar{x}_1.$$

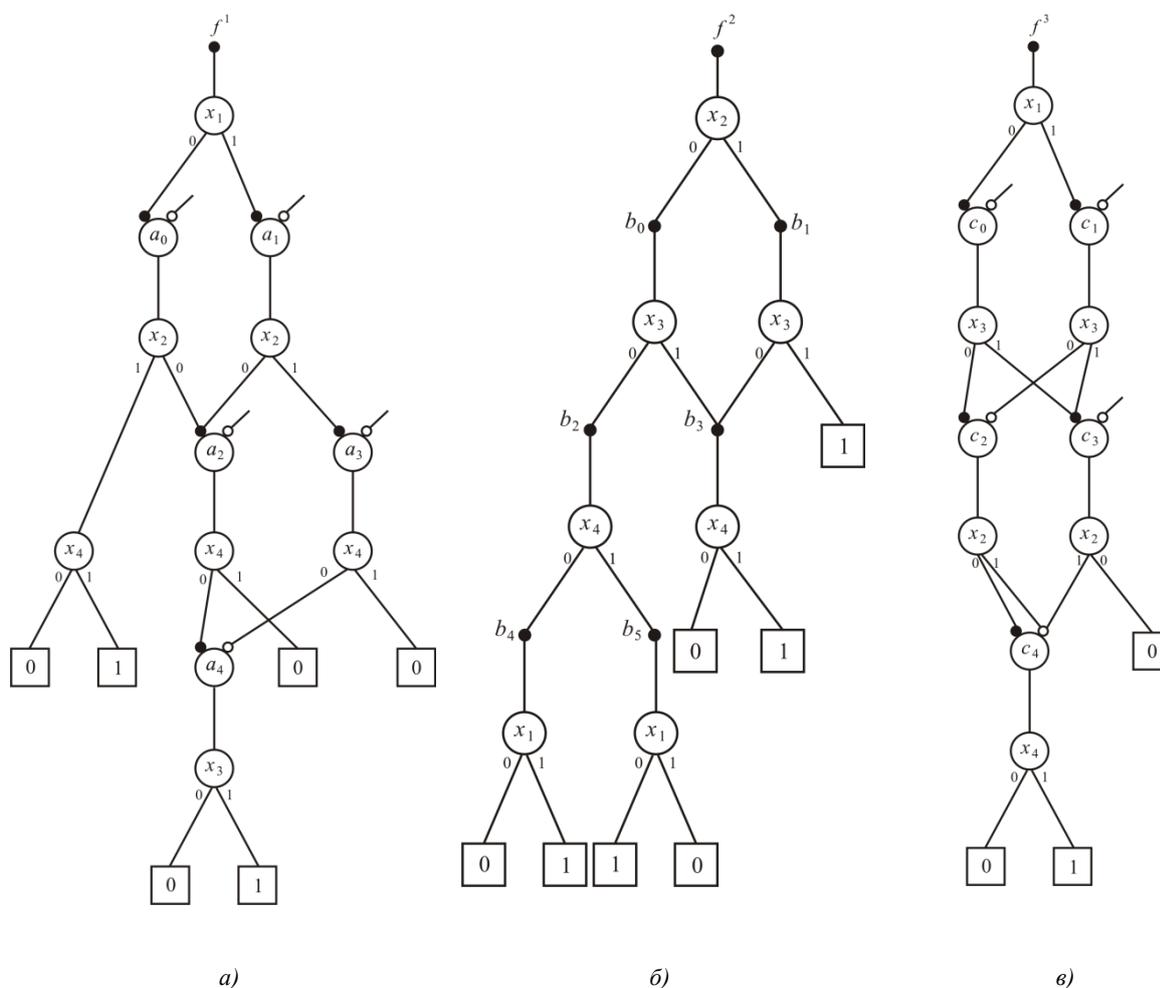


Рис. 2. Раздельные BDDI

Переход от многоуровневого представления к системе ДНФ булевых функций. Преобразование формата LOG в матричное представление системы функций основывается на устранении промежуточных переменных и получении зависимостей компонентных функций от входных переменных (аргументов) x_1, x_2, x_3, x_4 . Устранив промежуточные переменные в формулах (5)–(7), можно получить систему функций, эквивалентную заданной в табл. 1. Вместе с тем после осуществления перехода эквивалентная система ДНФ компонентных функций может быть представлена в другом виде (ДНФ могут иметь другую запись).

Разбиение многоуровневых представлений систем булевых функций на подсистемы. Одной из исследуемых оптимизационных процедур является процедура разбиения многоуровневых представлений систем булевых функций на подсистемы ограниченных размеров (блоки). Подобная задача встречается в процессе преобразования логических структур [16]. Известен подход, использующий выделение более крупных логических блоков, которые характеризуются сильной функциональной связью и строятся с использованием различных метрик кластеризации [17, 18].

Процедуру разбиения рассмотрим на графовой модели представления логической сети. Ограничениями на сложность каждого из блоков являются число r входных и число q выходных переменных блока. Текущим представлением логической сети является взвешенный ориентированный граф $G=(V, E)$. Вершине $v_i \in V$ соответствует логическое уравнение. Две вершины v_i, v_j графа G связывает дуга e_{ij} , принадлежащая множеству дуг E (дуга исходит из вершины v_i и заходит в вершину v_j), в том и только в том случае, когда выходная переменная уравнения, соответствующего v_i , является входной переменной уравнения, соответствующего v_j .

Задачу разбиения исходной системы логических уравнений на подсистемы можно представить как задачу разбиения множества V вершин графа G на множество непересекающихся подмножеств вершин $\{V_1, \dots, V_p\}$ графа (блоков), таких, что $V_i \subset V, V_i \neq \emptyset, V_i \cap V_j = \emptyset, \bigcup_{i=1}^p V_i = V$

для $i, j \in \{1, \dots, p\}, i \neq j$. Вершина графа может иметь числовые характеристики, которыми могут являться, например, числа входных и выходных переменных подсистемы логических уравнений, соответствующих этой вершине. В настоящей работе рассматривается задача разбиения графа, представляющего систему логических уравнений, на минимальное число p блоков V_i при выполнении заданных ограничений на числа входных (r) и выходных (q) переменных блоков V_i . Процедура разбиения реализует последовательный алгоритм формирования блоков.

Этапы построения каждого нового блока:

1. Выбор начальной вершины для блока и исключение ее из исходного множества вершин.
2. Построение множества смежных вершин для блока.
3. Просмотр множества смежных вершин и выбор очередной вершины для включения в блок. Выбранная вершина исключается из исходного множества.
4. Преобразование подсистемы логических уравнений, соответствующей данному текущему состоянию блока, с целью удаления внутренних переменных блока, т. е. устранение внутренних переменных блока.

Этапы 2–4 повторяются до тех пор, пока возможно включение уравнений в подсистему без нарушения заданных ограничений.

Перечисленные этапы можно представить как выбор вершин из исходного графа и объединение их в подграф (блок). Сформированный подграф будет характеризоваться параметрами, равными числу входных (r) и выходных (q) переменных в соответствующей подсистеме. Значения параметров каждого из блоков результирующего разбиения не должны превышать заданных ограничений.

В результате применения процедуры группирования с параметрами для блоков $r = 8$ (входов) и $q = 10$ (выходов) для RTL-описания, полученного командой untpar после синтеза схемы по табл. 1, получаем двухблочное разбиение (рис. 3).

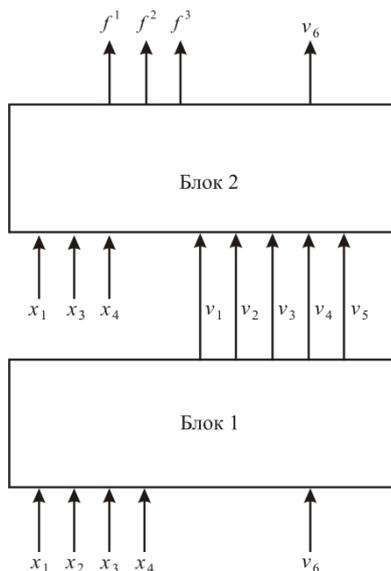


Рис. 3. Разбиение логических уравнений на два блока

Матричные функциональные описания полученных блоков в виде систем ДНФ представлены в табл. 2 и 3.

Таблица 2

Функции блока 1

x_1	x_2	x_3	x_4	v_6	v_1	v_2	v_3	v_4	v_5
0	0	-	-	0	0	1	0	0	0
-	0	-	1	0	0	1	0	0	0
0	-	-	0	0	0	1	0	0	0
-	1	-	0	0	0	1	0	0	0
0	1	-	-	0	0	1	0	0	0
0	1	-	-	-	0	0	0	1	0
-	1	-	0	-	1	0	0	0	0
-	0	-	0	-	0	0	1	0	0
-	-	0	-	-	0	0	1	0	0
-	-	1	-	-	0	0	0	0	1
-	1	-	-	-	0	0	0	0	1
-	-	-	1	-	0	0	0	0	1

Таблица 3

Функции блока 2

x_1	x_3	x_4	v_1	v_2	v_3	v_4	v_5	f^1	f^2	f^3	v_6
-	-	0	-	-	-	1	-	0	0	1	0
-	-	-	-	-	-	-	0	1	0	0	0
1	1	-	1	-	-	-	-	1	0	1	0
-	-	-	-	-	0	-	-	0	1	0	0
-	0	-	-	0	-	-	-	0	1	1	0
-	-	1	-	-	-	1	-	1	1	0	0
0	-	1	-	-	-	0	-	0	0	0	1

Обратим внимание на то, что выходная переменная (функция) $v_6 = \bar{x}_1 \bar{x}_4 \bar{v}_4$, где $v_4 = \bar{x}_1 x_2$, зависит от переменных x_1, x_2, x_4 , поэтому в схеме на рис. 3 нет обратных связей. Данный пример как раз иллюстрирует то, что объединение уравнений в блоки может покрывать груп-

пируемые уравнения различным образом. От параметров группирования зависят число и функциональность получаемых блоков.

Экспериментальное исследование. Общая идея организации экспериментов состояла в том, что различные оптимизационные (технологически независимые) процедуры для разных форм описания систем булевых функций выполнялись на языке SF в системе FLC [15], затем полученные минимизированные описания конвертировались в VHDL-описания и подавались на вход синтезатора LeonardoSpectrum [14], который выполнял построение (синтез) логических схем. Все схемы строились в одной и той же библиотеке КМОП-элементов (табл. 4). Сравнивались результаты исходного и повторного синтезов, осуществляемых с использованием оптимизационных процедур (и без их использования).

Таблица 4

Библиотека логических КМОП-элементов проектирования заказных СБИС

Элемент	Функция	Число транзисторов
GND	$y = 0$	1
VCC	$y = 1$	1
N	$y = \bar{A}$, инвертор	2
NX2	$y = \bar{\bar{A}}$, двукратный инвертор	4
NX4	$y = \bar{\bar{\bar{A}}}$, четырехкратный инвертор	8
NA	$y = \overline{AB}$	4
NO	$y = \overline{A \vee B}$	4
NAO	$y = \overline{(A \vee B)C}$	6
NOA	$y = \overline{(AB) \vee C}$	6
NA3O	$y = \overline{(A \vee B)CD}$	8
NO3A	$y = \overline{(AB) \vee C \vee D}$	8
NA3	$y = \overline{ABC}$	6
NO3	$y = \overline{A \vee B \vee C}$	6
LAT	D-триггер (по уровню)	20
DFF	D-триггер (по переднему фронту)	32
DFFRS	D-триггер (по переднему фронту) с установкой и сбросом	44
IBUF	$y = A$, входной буфер	4
OBUF	$y = A$, выходной буфер	10

Программы для экспериментов. Построение совместных BDDI производится программой BDD_Builder [13], применяющей при выборе очередной переменной разложения эвристику «минимальное число различных (с точностью до инверсии) подфункций». Эта программа в качестве исходных данных использует задание системы функций в виде матричного представления системы ДНФ, в языке SF такое представление называется SDF [15]. Построение отдельных BDDI осуществляется в системе FLC с помощью стратегий «декомпозиция по функциям», «BDDI-оптимизация листьев проекта», «устранение иерархии» [15].

Если исходная система функций задана многоуровневым представлением в виде логических уравнений, то для построения BDDI требуется осуществить переход от многоуровневого представления к двухуровневому (так часто в литературе называют представления булевых

функций в виде ДНФ). Такой переход программно реализован в системе FLC и называется ликвидацией внутренних переменных.

В экспериментах для блочного разбиения многоуровневых представлений используется алгоритм «полный с редукцией», представленный в работе [19]. Частным случаем группирования является разбиение системы функций на подсистемы, содержащие только одну компонентную функцию. В системе FLC для матричных форм функций такое разбиение осуществляется с помощью программы «Декомпозиция по функциям».

Примеры для экспериментов. Все примеры для экспериментов были взяты из практики проектирования цифровых схем. В эксперименте 1 сравнение проводилось на восьми примерах (Apx6, C8, Cht, Count, Dalu, Frg2, Too_large, X3) многоуровневых описаний комбинационной логики. В эксперименте 2 были использованы схемные реализации алгоритмических описаний промышленных VHDL-проектов. В экспериментах 3 и 4 сравнение проводилось на потоке из 28 примеров систем ДНФ булевых функций из библиотеки примеров [20] и на двух примерах Verg1, Verg2 матричных форм описания систем ДНФ булевых функций, взятых из практики проектирования.

Управление синтезом в синтезаторе LeonardoSpectrum. Для удобства потоки примеров обрабатывались единообразно с помощью соответствующих скриптов. Пример скрипта, который использовался при синтезе схем в LeonardoSpectrum:

```
clean_all;
set encoding Gray;
set modgen_select Smallest;
set asic_auto_dissolve_limit 500;
set auto_dissolve_limit 500;
read addm4.vhd;
load_library s3lib.syn;
set -hierarchy flatten
set effort standard
optimize -target s3lib -macro -area -effort standard -hierarchy flatten
report_area -cell_usage
```

Если требовалось получить RTL-описание (эксперименты 2 и 3) для повторного синтеза, то в конце скрипта добавлялись две команды:

```
unmap
auto_write addm4.vhd
```

В скрипте кроме значения **standard** жирным шрифтом выделены также важные для синтеза значения **flatten** параметра **hierarchy** и значение **500** параметра **asic_auto_dissolve_limit**. Установка значения **flatten** параметра **hierarchy** ориентирует синтезатор на выполнение синтеза с учетом устранения иерархии описания (схема синтезируется в виде одного блока), в этом случае имеется больше возможностей для оптимизации. Параметр **asic_auto_dissolve_limit** задает число «растворяемых» элементов при оптимизации. «Растворение» понимается как преобразование структурного описания подсхемы в функциональное. По умолчанию значение этого параметра равно 30. При эксперименте задается значение 500. При таком значении результирующие схемы имеют меньшую площадь, оптимизируются подсхемы большей размерности. Задание значения **Gray** параметра **encoding** важно при синтезе схем с памятью, схемы получаются более экономичными.

Эксперимент 1. Исходными данными являлись многоуровневые функциональные описания комбинационных схем в виде взаимосвязанных логических уравнений (формат LOG) языка SF. Для каждого из данных примеров строились три варианта логических схем и оценивалась их сложность (площадь S) в числе транзисторов.

Вариант 1.1. Исходное SF-описание переводилось в VHDL-описание и реализовывалось без выполнения предварительной оптимизации. Назовем этот вариант непосредственной схемной реализацией.

Вариант 1.2. Исходное SF-описание переводилось в систему ДНФ (внутренние переменные в описании устранялись). Затем выполнялась логическая оптимизация на основе разложения Шеннона, SF-описание переводилось в VHDL-описание, после чего синтезировалась логическая схема. Назовем этот вариант глобальной схемной реализацией на основе BDDI.

Вариант 1.3. Исходное SF-описание подвергалось процедуре разбиения на блоки, каждый блок SF-описания представлялся в виде системы ДНФ. Затем выполнялась логическая оптимизация на основе разложения Шеннона, SF-описание переводилось в VHDL-описание, после чего синтезировалась логическая схема. Назовем этот вариант схемной реализацией на основе блочного разбиения.

Эксперимент 2. Повторный синтез комбинационных частей схем, реализующих алгоритмические описания. Исходными данными являлись синтезируемые алгоритмические VHDL-описания цифровых устройств.

Вариант 2.1. После синтеза схемы по алгоритмическому VHDL-описанию оценивалась сложность комбинационной части схемы, структурное описание схемы (нетлист) переводилось в SF-описание. Затем элементы памяти выделялись в отдельный блок, логические комбинационные элементы – в комбинационную часть схемы и устранялась иерархия описания. Полученные логические уравнения представляли собой многоуровневое RTL-описание функций комбинационной части схемы с памятью.

Вариант 2.2. Повторный синтез комбинационного блока без оптимизации. Назовем такой вариант повторным синтезом от RTL-описания.

Вариант 2.3. Схемная реализация комбинационной части на основе блочного разбиения (см. вариант 1.3).

Эксперимент 3. Сравнение результатов синтеза от исходных матричных описаний комбинационных схем в виде систем ДНФ с результатами повторного синтеза от RTL-описаний и с результатами схемной реализации RTL-описаний на основе блочного разбиения. Исходными данными являлись матричные описания систем ДНФ (формат SDF). Каждый раз перед процедурой синтеза осуществлялась конвертация SF-описания схемы в VHDL-описание.

Выполнялись три варианта синтеза.

Вариант 3.1. Синтез по неоптимизированным исходным матричным представлениям систем ДНФ.

Вариант 3.2. Повторный синтез от RTL-описания схемы. RTL-описания были получены командой unpar после синтеза в библиотеке POWER [15].

Вариант 3.3. Разбиение RTL-описания на блоки, которые имеют восемь входов и два выхода, отдельная BDDI-оптимизация каждого из блоков, схемная реализация полученных оптимизированных представлений. В качестве исходных для варианта 3.3 были взяты те же RTL-описания, что и для варианта 3.2.

Эксперимент 4. Сравнение эффективности использования процедур оптимизации на основе совместных и отдельных BDDI. Исходными данными являлись матричные описания систем ДНФ (формат SDF).

Вариант 4.1. Конвертация исходного (неоптимизированного) SF-описания в VHDL-описание и его схемная реализация.

Вариант 4.2. Глобальная схемная реализация на основе оптимизированных совместных BDDI: оптимизация на основе совместной BDDI, конвертация BDDI в VHDL-описание и схемная реализация.

Вариант 4.3. Глобальная схемная реализация на основе оптимизированных отдельных BDDI: оптимизация на основе отдельных BDDI, конвертация отдельных BDDI в VHDL-описание и схемная реализация.

Рассмотрим результаты некоторых экспериментов (табл. 5) для примера системы функций из табл. 1. Синтезированная по варианту 4.2 логическая схема оказалась лучшей по площади, поскольку логические элементы, из которых она состоит, содержат наименьшее суммарное число транзисторов.

Таблица 5

Результаты экспериментов для системы ДНФ трех булевых функций

Варианты	Исходные данные	Площадь схемы по числу транзисторов, S	Число элементов схемы, L
4.1 и 3.1	Система ДНФ функций (см. табл. 1)	86	19
4.2	Совместная BDDI (см. (5))	74	17
4.3	Раздельные BDDI (см. (6)–(8))	86	18
3.3	BDDI, построенные по подсистемам функций (см. табл. 2 и 3)	78	15

Результаты экспериментов. Результаты проведенных четырех экспериментов представлены в табл. 6–9, где n – число аргументов булевых функций; m – число функций; k – число элементарных конъюнкций в системе ДНФ булевых функций; S – площадь логической схемы, задаваемая в числе транзисторов, из которых состоят элементы, лучшие решения выделены жирным шрифтом; L – число элементов логической схемы. Два последних столбца в табл. 6 и 7 относятся к варианту 1.3 (табл. 6) и варианту 2.2 (табл. 7) соответственно.

Таблица 6

Результаты эксперимента 1

Пример	n	m	Площадь схемы, S			Число блоков, p	Ограничения на параметры r, q блоков
			Вариант 1.1	Вариант 1.2	Вариант 1.3		
Арех6	135	94	1830	2110	1874	13	25, 100
C8	28	18	312	324	318	9	10, 100
Cht	47	36	680	670	652	7	10, 100
Count	35	16	256	256	256	8	8, 100
Dalu	75	16	1834	1396	1286	23	30, 100
Frg2	143	139	4224	–	3800	102	10, 100
Too_large	38	3	10902	9038	–	–	–
X3	135	99	3462	3788	3154	87	8, 100
Число лучших решений по площади			2	1	4		

Таблица 7

Результаты эксперимента 2

Пример	n	m	Площадь схемы, S			Число блоков, p	Ограничения на параметры r, q блоков
			Вариант 2.1	Вариант 2.2	Вариант 2.3		
Alu_avr	205	13	1472	1504	1700	40	15, 100
Alu_1	38	32	10934	11110	11340	323	10, 100
Timer	84	42	766	424	424	24	10, 100
VGA	150	106	2891	2604	3076	40	15, 100
simmetria	53	20	15390	15006	15988	361	10, 100
Число лучших решений по площади			2	3	0		

Таблица 8

Результаты эксперимента 3, параметры блоков: восемь входов, два выхода

Пример	n	m	k	Площадь схемы, S		
				Вариант 3.1	Вариант 3.2	Вариант 3.3
Add6	12	7	1092	5750	4414	4830
Addm4	9	8	512	4104	4108	4134
B12	15	9	431	258	262	236
B2	15	17	110	2506	2418	2430
B9	16	5	123	404	402	392
Dc2	8	7	58	452	372	354
Dk48	15	17	148	394	394	390
In0	15	11	138	2134	1926	1934
In2	19	10	137	2034	1822	1810
Intb	15	7	664	7340	7502	7946
M2	8	16	96	976	896	938
M3	8	16	128	1344	1194	1236
Newcpla1	9	16	38	428	382	376
Newtpla	15	5	23	212	210	214
P82	5	14	24	348	346	354
Root	8	5	256	958	900	900
Shift	19	16	100	806	1216	1186
Sqn	7	3	96	502	372	446
Sym10	10	1	837	1374	996	760
Tial	14	8	640	5704	5740	6024
Z5xp1	7	10	128	1644	1552	1568
Z9sym	9	1	420	1022	820	658
Число лучших решений по площади				4	10	9

Согласно результатам исследования эффективной процедурой технологически независимой оптимизации является глобальная оптимизация на основе совместных BDDI в тех случаях, когда от многоуровневого RTL-описания в виде логических уравнений удастся перейти к матричному описанию и воспользоваться программой оптимизации совместных BDDI. Если же не удастся перейти к такому описанию, то другой маршрут проектирования, а именно предварительное разбиение системы функций на подсистемы и последующее построение совместных BDDI (эксперимент 1), позволяет уменьшить площадь схемы. При этом уменьшение параметров блоков, на которые разбиваются логические уравнения, позволяет добиваться лучших результатов по площади.

В экспериментах 2 и 3 был использован метод повторного синтеза. Результаты эксперимента 3 показали, что повторный синтез комбинационных частей схем для проектов, содержащих элементы памяти, либо дает небольшой выигрыш, либо такой выигрыш (уменьшение площади) получить не удастся, при этом процедура разбиения на блоки только ухудшает решение. Смена библиотеки POWER на новую библиотеку (см. табл. 4) не изменяет общую ситуацию при синтезе: если оптимизация проведена хорошо, то для различных библиотек площадь схемы уменьшается по сравнению со схемой, получаемой по некачественному результату технологически независимой оптимизации.

Таблица 9

Результаты эксперимента 4

Пример	n	m	k	Вариант 4.1		Вариант 4.2		Вариант 4.3	
				S	L	S	L	S	L
addm4	9	8	512	4104	758	1332	241	1488	263
b12	15	9	431	258	48	262	50	306	59
b2	16	17	110	2506	516	5902	1046	3524	655
b9	16	5	123	404	79	440	90	402	82
in2	19	10	137	2034	392	1674	303	1260	238
intb	15	7	664	7340	1209	4588	781	4866	862
m181	15	9	430	266	49	276	51	300	57
mp2d	11	14	123	558	114	268	57	262	59
root	8	5	256	958	185	474	84	406	84
soar	83	94	529	2462	488	2204	432	1862	387
t3	12	8	152	292	57	252	53	250	52
tial	14	8	640	5704	952	8010	1384	4704	805
Verg1	17	61	2003	10284	1633	17282	2928	6996	1274
Verg2	18	63	2129	–	–	9600	1541	10648	1711
vtx1	27	6	110	282	60	574	117	408	87
x9dn	27	7	120	286	61	554	120	408	87
Число лучших решений по площади				5		3		8	

Согласно результатам эксперимента 4 применение совместных BDDI для исходных матричных форм задания функций является более предпочтительным приемом технологически независимой минимизации при синтезе логических схем из библиотечных элементов, так как площадь схем, построенных по совместным BDDI, в подавляющем числе случаев меньше площади схем, построенных по отдельным BDDI.

Заключение. Предварительная оптимизация исходных функциональных описаний систем булевых функций дает уменьшение площади при синтезе КМОП-схем, при этом конкурирующими методами являются совместные и отдельные BDDI, что аналогично ситуации с совместной и отдельной минимизацией систем булевых функций в классе ДНФ.

Повторные оптимизация (с использованием имеющихся программ) и синтез по RTL-описаниям, полученным по уже синтезированным схемам, практически не имеют смысла, так как могут лишь незначительно уменьшить площадь схемы либо даже увеличить ее. Такой вывод справедлив как для комбинационных схем (без элементов памяти), так и для схем с памятью и обусловлен тем, что в RTL-описаниях большой размерности, составленных из «мелких» уравнений, трудно выделить «связные» функциональные блоки и получить другое многоуровневое представление, по которому повторный синтез будет проведен более эффективно, чем синтез по исходному многоуровневому описанию, полученному в результате замены алгоритмических конструкций языка VHDL логическими уравнениями.

Список использованных источников

1. Брейтон, Р. К. Синтез многоуровневых комбинационных логических схем / Р. К. Брейтон, Г. Д. Хэттел, А. Л. Санджованни-Винченцелли // ТИИЭР. – 1990. – Т. 78, № 2. – С. 38–83.
2. Поляков, А. К. Языки VHDL и VERILOG в проектировании цифровой аппаратуры / А. К. Поляков. – М. : СОЛОН-Пресс, 2003. – 320 с.
3. Закревский, А. Д. Логические основы проектирования дискретных устройств / А. Д. Закревский, Ю. В. Поттосин, Л. Д. Черемисинова. – М. : Физматлит, 2007. – 592 с.
4. Закревский, А. Д. Логический синтез каскадных схем / А. Д. Закревский. – М. : Наука, 1981. – 416 с.

5. Sasao, T. FPGA design by generalized functional decomposition / T. Sasao // *Representations of Discrete Functions* ; ed. by T. Sasao, M. Fujita. – Kluwer Academic Publishers, 1996. – P. 233–258.
6. Scholl, C. *Functional Decomposition with Applications to FPGA Synthesis* / C. Scholl. – Kluwer Academic Publishers, 2001. – 288 p.
7. Бибило, П. Н. Применение диаграмм двоичного выбора при синтезе логических схем / П. Н. Бибило. – Минск : Беларус. навука, 2014. – 231 с.
8. Bryant, R. E. Graph-based algorithms for Boolean function manipulation / R. E. Bryant // *IEEE Transactions on Computers*. – 1986. – Vol. 35, no. 8. – P. 677–691.
9. Bryant, R. E. Ordered binary decision diagrams / R. E. Bryant, C. Meinel // *Logic Synthesis and Verification* ; ed. by S. Hassoun, T. Sasao, R. K. Brayton. – Kluwer Academic Publishers, 2002. – P. 285–307.
10. Meinel, C. *Algorithms and Data Structures in VLSI Design: OBDD – Foundations and Applications* / C. Meinel, T. Theobald. – Berlin, Heidelberg : Springer-Verlag, 1998. – 267 p.
11. Amaru, L. G. *New Data Structures and Algorithms for Logic Synthesis and Verification* / L. G. Amaru. – Springer, 2017. – 156 p.
12. Валидация на системном уровне. Высокоуровневое моделирование и управление тестированием / М. Чэнь [и др.]. – М. : Техносфера, 2014. – 296 с.
13. Бибило, П. Н. Использование полиномов Жегалкина при минимизации многоуровневых представлений систем булевых функций на основе разложения Шеннона / П. Н. Бибило, Ю. Ю. Ланкевич // *Программная инженерия*. – 2017. – № 3. – С. 369–384.
14. Бибило, П. Н. Системы проектирования интегральных схем на основе языка VHDL. StateCAD, ModelSim, LeonardoSpectrum / П. Н. Бибило. – М. : СОЛОН-Пресс, 2005. – 384 с.
15. Бибило, П. Н. Логическое проектирование дискретных устройств с использованием производственно-фреймовой модели представления знаний / П. Н. Бибило, В. И. Романов. – Минск : Беларус. навука, 2011. – 279 с.
16. Григорьян, С. Г. Конструирование электронных устройств систем автоматизации и вычислительной техники / С. Г. Григорьян. – Ростов н/Д : Феникс, 2007. – 303 с.
17. Кузовлев, В. И. Выявление высокоуровневых иерархических структур сверхбольших интегральных схем через сильно связанные логические группы / В. И. Кузовлев, Н. А. Иванова // *Вестник МГТУ им. Н. Э. Баумана. Сер. Приборостроение*. – 2016. – № 4. – С. 4–18.
18. Netlist and system partitioning / A. B. Kahng [et al.] // *VLSI Physical Design: From Graph Partitioning to Timing Closure*. – Springer, 2011. – Chap. 2. – P. 31–54.
19. Бибило, П. Н. Оптимизационные преобразования логической схемы на основе блочного разбиения / П. Н. Бибило, Н. А. Кириенко // *Информатика*. – 2009. – № 3(23). – С. 5–15.
20. Jeong, C. *Computer-Aided Design of Digital Systems* / C. Jeong // Department of Computer Science [Electronic resource]. – Mode of access: <http://www1.cs.columbia.edu/~cs6861/sis/espresso-examples/ex>. – Date of access: 20.03.2018.

References

1. Brayton R. K., Hachtel G. D., Sangiovanni-Vincentelli A. L. Sintez mnogourovnevnyh kombinacionnyh logicheskikh skhem [Multilevel Logic Synthesis]. *TIJER [TIJER]*, 1990, vol. 78, no. 2, pp. 38–83 (in Russian).
2. Polyakov A. K. *Yazyki VHDL i VERILOG v proektirovanii tsifrovoy apparatury. The VHDL and VERILOG Languages in Designing of Digital Hardware*. Moscow, SOLON-Press Publ., 2003, 320 p. (in Russian).
3. Zakrevskij A. D., Pottosin Ju. V., Cheremisinova L. D. Logicheskie osnovy proektirovanija diskretnykh ustrojstv. *Logical Bases of Design of Discrete Devices*. Moscow, Fizmatlit Publ., 2007, 592 p. (in Russian).
4. Zakrevskij A. D. Logicheskij sintez kaskadnykh skhem. *Logical Synthesis of Cascade Circuits*. Moscow, Nauka Publ., 1981, 416 p. (in Russian).
5. Sasao T. FPGA design by generalized functional decomposition. *Representations of Discrete Functions*. Kluwer Academic Publishers, 1996, pp. 233–258.
6. Scholl C. *Functional Decomposition with Applications to FPGA Synthesis*. Kluwer Academic Publishers, 2001, 288 p.
7. Bibilo P. N. Primenenie diagramm dvoichnogo vybora pri sinteze logicheskikh shem. *Application of Binary Decision Diagrams at Synthesis of Logical Circuits*. Minsk, Belarus. Navuka Publ., 2014, 231 p. (in Russian).
8. Bryant R. E. Graph-based algorithms for Boolean function manipulation. *IEEE Transactions on Computers*, 1986, vol. 35, no. 8, pp. 677–691.
9. Bryant R. E., Meinel C. Ordered binary decision diagrams. *Logic Synthesis and Verification*. Kluwer Academic Publishers, 2002, pp. 285–307.
10. Meinel C., Theobald T. *Algorithms and Data Structures in VLSI Design: OBDD – Foundations and Applications*. Berlin, Heidelberg, Springer-Verlag Publ., 1998, 267 p.
11. Amaru L. G. *New Data Structures and Algorithms for Logic Synthesis and Verification*. Springer Publ., 2017, 156 p.
12. Chen M., Qin K., Ku H.-M., Mishra P. Validaciya na sistemnom urovne. Vysokourovnevnoe modelirovanie i upravlenie testirovanie. *Validation at the System Level. High-Level Simulation and Testing Management*. Moscow, Tekhnosfera Publ., 2014, 296 p. (in Russian).
13. Bibilo P. N., Lankevich Yu. Yu. Ispol'zovanie polinomov Zhegalkina pri minimizacii mnogourovnevnyh predstavlenij sistem bulevykh funkcij na osnove razlozheniya Shennona [The use of Zhegalkin polynomials with minimization

of multilevel representations of systems of Boolean functions on the basis of the Shannon decomposition]. *Programmnyaya inzheneriya [Software Engineering]*, 2017, no. 3, pp. 369–384 (in Russian).

14. Bibilo P. N. Cistemy proektirovaniya integral'nyh skhem na osnove yazyka VHDL. StateCAD, ModelSim, LeonardoSpectrum. *Integrated Circuit Design Systems Based on the VHDL Language. StateCAD, ModelSim, LeonardoSpectrum*. Moscow, SOLON-Press Publ., 2005, 384 p. (in Russian).

15. Bibilo P. N., Romanov V. I. Logicheskoe proektirovanie diskretnykh ustrojstv s ispol'zovaniem produkcionno-frejmovoj modeli predstavlenija znaniy. *Logical Design of Discrete Devices with Use of Productional and Frame Model of Representation of Knowledge*. Minsk, Belarus. Navuka Publ., 2011, 279 p. (in Russian).

16. Grigor'yan S. G. Konstruirovaniye jelektronnykh ustrojstv sistem avtomatizacii i vychislitel'noj tekhniki. *Design of Electronic Devices of Automation Systems and Computers*. Rostov n/D, Feniks Publ., 2007, 303 p. (in Russian).

17. Kuzovlev V. I., Ivanova N. A. Vyyavlenie vysokourovnevnyh ierarhicheskikh struktur sverhbol'shih integral'nyh skhem cherez sil'no svyazannye logicheskie gruppy [Circuit detection through tangled logic structures]. *Vestnik MG TU im. N. E. Baumana, ser. Priborostroenie [Herald of the Bauman Moscow State Tech. Univ., Instrum. Eng.]*, 2016, no. 4, pp. 4–18 (in Russian).

18. Kahng A. B., Liening J., Markov I. L., Hu J. Netlist and system partitioning. *VLSI Physical Design: From Graph Partitioning to Timing Closure*. Springer, 2011, ch. 2, pp. 31–54.

19. Bibilo P. N., Kirienko N. A. Optimizatsionnye preobrazovaniya logicheskoy skhemy na osnove blochnogo razbieniya [Optimizing conversions of a logic circuit by partitioning into blocks]. *Informatika [Informatics]*, 2009, no. 3, pp. 5–15 (in Russian).

20. Jeong C. Computer-Aided Design of Digital Systems. *Department of Computer Science*. Available at: <http://www1.cs.columbia.edu/~cs6861/sis/espresso-examples/ex> (accessed 20.03.2018).

Информация об авторах

Бибилу Петр Николаевич – доктор технических наук, профессор, Объединенный институт проблем информатики Национальной академии наук Беларуси (ул. Сурганова, 6, 220012, Минск, Республика Беларусь). E-mail: bibilo@newman.bas-net.by

Кириенко Наталья Алексеевна – кандидат технических наук, доцент, Объединенный институт проблем информатики Национальной академии наук Беларуси (ул. Сурганова, 6, 220012, Минск, Республика Беларусь). E-mail: kir@newman.bas-net.by

Ланкевич Юрий Юрьевич – младший научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси (ул. Сурганова, 6, 220012, Минск, Республика Беларусь). E-mail: yurafreedom18@gmail.com

Information about the authors

Petr N. Bibilo – D. Sc. (Engineering), Professor, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganova Str., 220012, Minsk, Republic of Belarus).

E-mail: bibilo@newman.bas-net.by

Natalia A. Kirienko – Ph. D. (Engineering), The United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganova Str., 220012, Minsk, Republic of Belarus). E-mail: kir@newman.bas-net.by

Yury Y. Lankevich – Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganova Str., 220012, Minsk, Republic of Belarus). E-mail: yurafreedom18@gmail.com

ISSN 1816-0301 (print)

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
APPLIED INFORMATION TECHNOLOGIES

УДК 621.382.3

Поступила в редакцию 08.05.2018

Received 08.05.2018

В. Э. Маркевич, В. В. Легкоступ*ОАО «АЛЕВКУРП», Минск, Беларусь*

**МОДИФИЦИРОВАННЫЙ МЕТОД ПРОПОРЦИОНАЛЬНОГО НАВЕДЕНИЯ
ПРИ ОГРАНИЧЕННОМ СЕКТОРЕ СОПРОВОЖДЕНИЯ ОБЪЕКТА УПРАВЛЕНИЯ**

Аннотация. Рассматривается усовершенствованный вариант алгоритма, реализующего метод пропорционального наведения для управления сверхзвуковыми беспилотными летательными аппаратами (БПЛА) в ограниченном секторе углового сопровождения, связанном с линией визирования аэродинамической цели. Метод обеспечивает устойчивое наведение объекта управления с углами упреждения, не превышающими предельно допустимые. Приводятся аналитические выражения, описывающие алгоритм формирования команд управления и сигналов коррекции, адаптивных к параметрам движения. Сопровождение и наведение БПЛА при этом реализуются в широком диапазоне дальностей, высот и скоростей движения объекта перехвата. Представлены результаты математического моделирования задачи пространственного перехвата аэродинамической цели, а также сравнительного анализа и предварительной оценки точности наведения классических и модифицированных вариантов метода пропорционального наведения.

Ключевые слова: модифицированный метод пропорционального наведения, станция наведения ракет, сектор углового сопровождения, маневрирующая аэродинамическая цель, пространственная модель движения ОУ, ошибки наведения

Для цитирования. Маркевич, В. Э. Модифицированный метод пропорционального наведения при ограниченном секторе сопровождения объекта управления / В. Э. Маркевич, В. В. Легкоступ // Информатика. – 2018. – Т. 15, № 3. – С. 71–92.

V. E. Markevich, V. V. Legkostup

**MODIFIED METHOD OF PROPORTIONAL GUIDANCE
IN CASE OF THE LIMITED SECTOR OF TRACKING OF THE CONTROL OBJECT**

OJSC "ALEVKURP", Minsk, Belarus

Abstract. An improved variant of the algorithm realizing a method of proportional guidance for the control of hypersonic unmanned aerial vehicles (UAVs) in a limited sector of angular tracking associated with the aerodynamic target line is considered. The method provides stable guidance of the control object with angles of anticipation that do not exceed the maximum allowable limits. Analytic expressions describing the algorithm of calculation of guidance commands and correction signals adaptive to the motion parameters are given. The tracking and guidance of the UAVs are realized in a wide range of ranges, heights and speeds of the interception. The results of mathematical modeling of the problem of spatial interception of the aerodynamic target, as well as comparative analysis and preliminary estimation of accuracy of guidance of classical and modified variants of the method of proportional navigation are presented.

Keywords: modified method of proportional guidance, guidance station missiles, the angular sector tracking, aerodynamics maneuvering target, a spatial model of the missile, the pointing errors

For citation. Markevich V. E., Legkostup V. V. Modified method of proportional guidance in case of the limited sector of tracking of the control object. *Informatics*, 2018, vol. 15, no. 3, pp. 71–92 (in Russian).

Введение. Большинство современных алгоритмов управления сверхзвуковыми БПЛА различного назначения, предназначенных для поражения целей различного класса, базируются на модификациях метода пропорционального наведения (МПН) как частного случая метода наведения объекта управления (ОУ) в фактическую (мгновенную) точку встречи [1–8].

Согласно МПН кинематические нормальные ускорения ОУ должны быть пропорциональны угловой скорости вращения (УСВ) линии визирования (ЛВ) ОУ – цели. В векторной форме выражение для заданных ускорений ОУ имеет следующий вид [9–12]:

$$\mathbf{W}_p = N_0(t) \mathbf{M}(\Theta_2)(\boldsymbol{\omega}_0 \times \Delta \mathbf{v}), \quad \boldsymbol{\omega}_0 = \frac{(\Delta \mathbf{v} \times \Delta \mathbf{r})}{\Delta r^2}, \quad (1)$$

где $\Theta_2 = (\theta_2, \varphi_2, \gamma_2)^T$ – вектор углов ориентации ЛВ ОУ – цель в декартовой системе координат (СК), связанной с точкой расположения радиолокационной станции наведения (РЛСН); $\mathbf{W}_p = (W_{px} \ W_{py} \ W_{pz})^T$ – вектор ускорений ОУ в проекциях на оси СК, связанной с ЛВ ОУ – цель (для управления продольная компонента W_{pz} не используется); $\boldsymbol{\omega}_0 = (\omega_{0x} \ \omega_{0y} \ \omega_{0z})^T$ – вектор УСВ ЛВ ОУ – цель в проекциях на оси СК, связанной с ЛВ; $\Delta \mathbf{v} = \Delta \dot{\mathbf{r}} = \mathbf{v}_2 - \mathbf{v}_1$ – вектор разности скоростей сближения ОУ \mathbf{v}_1 и цели \mathbf{v}_2 ; $\Delta \mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ – вектор разности декартовых координат ОУ \mathbf{r}_1 и цели \mathbf{r}_2 ; $N_0(t)$ – навигационный параметр, зависящий от параметров движения ОУ $\mathbf{r}_1, \mathbf{v}_1$ и цели $\mathbf{r}_2, \mathbf{v}_2$ и изменяющийся, как правило, в диапазоне $N_0(t) \in (2; 10)$; $(\boldsymbol{\omega}_0 \times \Delta \mathbf{v}) = [\boldsymbol{\omega}_0, \Delta \mathbf{v}]$ – аналитическая запись векторного произведения.

Основными достоинствами МПН, используемого применительно к задаче перехвата скоростных и маневрирующих аэродинамических целей, являются:

- возможность реализации обстрела с произвольного ракурса как в переднюю, так и заднюю полусферу атаки;
- возможность формирования траектории наведения, обеспечивающей наведение на прямолинейно летящую (неманеврирующую) цель с минимальными перегрузками;
- слабая чувствительность метода к величине флуктуационной ошибки измерения угловых координат цели (в отличие от трехточечных методов телеуправления).

Вместе с тем существует ряд особенностей данного метода, снижающих в некоторых тактических ситуациях его эффективность:

- высокая чувствительность МПН к неконтролируемым случайным внешним возмущениям по нормальным ускорениям (противорадиолокационному или противоракетному маневру цели), зависящая от величины навигационного параметра метода;
- увеличение кривизны траектории при отсутствии сигналов компенсации систематических возмущающих воздействий (ускорения свободного падения, продольного ускорения ОУ);
- значительные углы упреждения ОУ относительно ЛВ цели, движущейся с большими угловыми скоростями относительно станции наведения ракет (СНР), которые достигают на начальном участке траектории величины (30...40)°.

В качестве примера на рис. 1 изображены геометрия взаимного перемещения ОУ и цели, а также системы координат, используемые при получении уравнений относительного и абсолютного движения ОУ и цели. В представленной схеме используются обозначения осей сопровождающих (динамических) и статических ортонормированных триэдров, связанных с полюсами соответствующих динамических объектов (СНР, ОУ и цели), взаимное положение которых характеризуется следующими ортами рассматриваемых СК: $\mathbf{I}_x, \mathbf{I}_y, \mathbf{I}_z$ – орты СК, связанной с фиксированной точкой визирования ОУ и цели, находящейся на земле (как правило, полюс данной СК определяется точкой стояния СНР); $\mathbf{I}_{x0}, \mathbf{I}_{y0}, \mathbf{I}_{z0}$ – орты СК ЛВ, характеризующие положение соответствующих осей данной СК; $\mathbf{I}_{x1}, \mathbf{I}_{y1}, \mathbf{I}_{z1}$ – орты связанной СК (СвСК), характеризующие положение продольной и поперечных осей (орт \mathbf{I}_{x1} и вектор скорости ОУ

можно считать коллинеарными в предположении о малости углов атаки и скольжения ОУ); $\mathbf{n}_{x1}, \mathbf{n}_{y1}, \mathbf{n}_{z1}$ – орты осей измерительной СК (ИСК), связанной с ЛВ ОУ (ИСК-1); $\mathbf{n}_{x2}, \mathbf{n}_{y2}, \mathbf{n}_{z2}$ – орты осей измерительной системы координат, связанной с ЛВ цели (ИСК-2); $\mathbf{r}_1 = \mathbf{x}_1, \mathbf{r}_2 = \mathbf{x}_2$ – векторы радиальной дальности ОУ и цели соответственно; \mathbf{v}_1 – вектор скорости ОУ; \mathbf{v}_2 – вектор скорости цели; $\Delta\mathbf{v} = \mathbf{v}_{12}$ – вектор относительной скорости сближения ОУ и цели (на рис. 1 не показан).

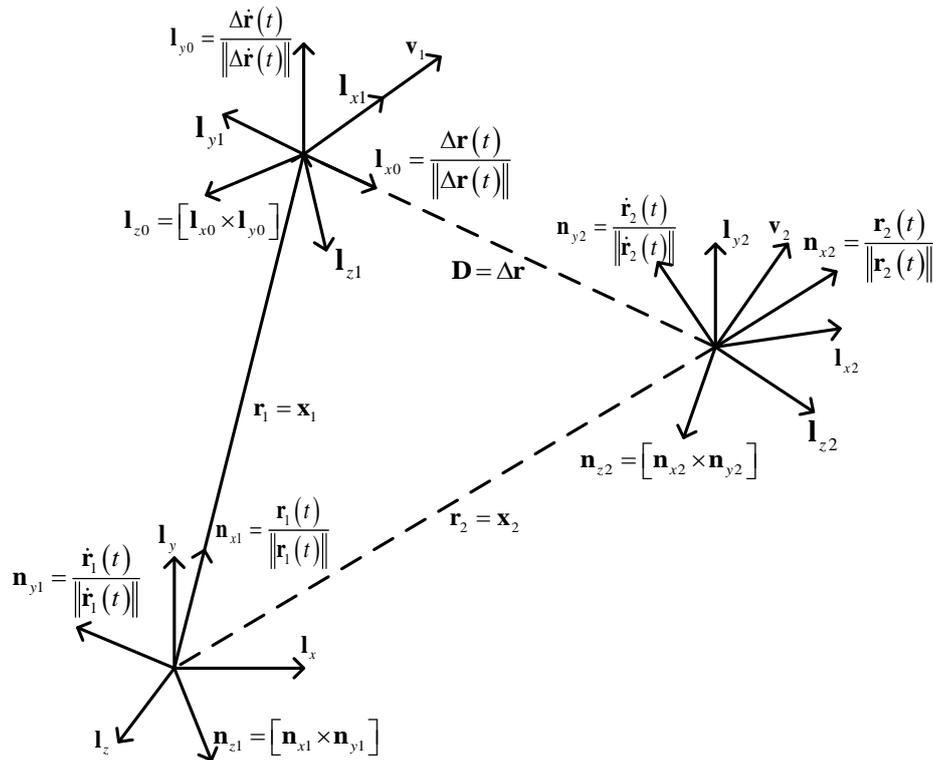


Рис. 1. Геометрия взаимного перемещения ОУ и цели, используемая для получения кинематических уравнений

Сопровождающий ортонормированный триэдр СК ЛВ характеризуется тремя ортами, которые связаны соотношениями вида

$$\mathbf{l}_{x0} = \frac{\Delta\mathbf{r}(t)}{\|\Delta\mathbf{r}(t)\|}, \mathbf{l}_{y0} = \frac{\Delta\dot{\mathbf{r}}(t)}{\|\Delta\dot{\mathbf{r}}(t)\|}, \mathbf{l}_{z0} = [\mathbf{l}_{x0} \times \mathbf{l}_{y0}]. \quad (2)$$

Аналогичные соотношения связывают орты сопровождающих триэдров остальных динамических СК (например, СвСК ОУ и цели). В качестве примера приведем один из вариантов исходной нелинейной модели перехвата, которая может быть использована для синтеза оптимального устройства управления методом аналитического конструирования агрегированных регуляторов (АКАР) с учетом технических ограничений на сектор сопровождения ОУ. Теоретические основы АКАР и синергетической теории управления нелинейными объектами изложены в [13, 14]. Уравнения относительного движения цели и ОУ имеют следующий вид [12]:

$$\begin{aligned} \dot{\theta}_2 &= \omega_{z2}, \\ \dot{\phi}_2 &= \frac{\omega_{y2}}{\cos(\theta_2)}; \end{aligned} \quad (3)$$

$$\begin{aligned}\dot{W}_{y2}^* &= k_y (W_{y2} - \operatorname{tg}(W_{y2}^*)), \\ \dot{W}_{z2}^* &= k_z (W_{z2} - \operatorname{tg}(W_{z2}^*)), \\ \dot{\omega}_{z2} &= \frac{W_{yp}^* - 2|\dot{D}|\omega_{z2}}{D} - \omega_{y2}^2 \operatorname{tg}(\theta_2), \\ \dot{\omega}_{y2} &= \frac{-W_{zp}^* - 2|\dot{D}|\omega_{y2}}{D} + \omega_{y2}\omega_{z2} \operatorname{tg}(\theta_2), \\ \dot{W}_{y2} &= \alpha_{yp} (U_y - W_{y2}), \\ \dot{W}_{z2} &= \alpha_{zp} (U_z - W_{z2}).\end{aligned}\tag{4}$$

Уравнения движения ОУ относительно СНР примут вид

$$\begin{aligned}\dot{\varepsilon}_j &= \omega_{zj}, \\ \dot{\beta}_j &= \frac{\omega_{yj}}{\cos(\varepsilon_j)}, \\ \dot{\varepsilon}_j^* &= k_\varepsilon (\varepsilon_j - \operatorname{tg}(\varepsilon_j^*)), \\ \dot{\beta}_j^* &= k_\beta (\beta_j - \operatorname{tg}(\beta_j^*)), \\ \dot{\omega}_{zj} &= \frac{W_{yj} - 2\dot{\rho}_j \omega_{zj}}{\rho_j} - \omega_{yj}^2 \operatorname{tg}(\varepsilon_j^*), \\ \dot{\omega}_{yj} &= \frac{-W_{zj} - 2\dot{\rho}_j \omega_{yj}}{\rho_j} + \omega_{yj}\omega_{zj} \operatorname{tg}(\varepsilon_j^*), \\ \dot{W}_{yj} &= \alpha_{yp} (U_{yj} - W_{yj}), \\ \dot{W}_{zj} &= \alpha_{zp} (U_{zj} - W_{zj}),\end{aligned}\tag{5}$$

$$\mathbf{U}_j = \mathbf{M}(\varepsilon_j, \beta_j) \mathbf{M}^T(\theta_2, \varphi_2) (U_{jx} \ U_{jy} \ U_{jz})^T,$$

где $\mathbf{W}_2 = (W_{x2} \ W_{y2} \ W_{z2})^T$ – компоненты вектора линейных (продольных и нормальных) ускорений ОУ в системе координат, связанной с ЛВ ОУ – цель (СК ЛВ); $\boldsymbol{\omega}_2 = (\omega_{x2} \ \omega_{y2} \ \omega_{z2})^T$ – компоненты вектора УСВ ЛВ ОУ – цель в СК ЛВ; D, \dot{D} – относительная дальность и скорость сближения ОУ и цели; $\alpha_p = \Delta f_p$ – коэффициент, определяющий полосу пропускания (инерционные свойства) неизменной части ОУ; U_y, U_z – сигналы (команды) управления на выходе регулятора; $\hat{\mathbf{W}}_{ц} = (\hat{W}_{xц}, \hat{W}_{yц}, \hat{W}_{zц})^T$ – вектор оценок нормальных ускорений цели, полученных в следящем фильтре оценивания параметров движения цели; $W_{yp}^* = (\hat{W}_{yц} - W_{y2}^*)$, $W_{zp}^* = (\hat{W}_{zц} - W_{z2}^*)$ – относительные управляющие нормальные ускорения ОУ; $\varepsilon_j, \beta_j, \rho_j$ – угол места, азимут и радиальная дальность ОУ (цели) относительно наземной РЛСН; ρ_j – радиальная скорость ОУ (цели); ω_{yj}, ω_{zj} – УСВ ЛВ СНР – ОУ (цели) в проекциях на оси измерительной СК, продольная ось которой связана с ЛВ; W_{yj}, W_{zj} – нормальные ускорения ОУ (цели) в проекциях на оси стационарной (измерительной) СК (СтСК); W_{yj}, W_{zj} – команды управления ОУ (задающие нормальные ускорения цели) на входе фильтра, моделирующего запаздывание в контуре управления ОУ (динамику маневра цели); $j \in (0, 1)$ – подстрочный индекс переменных состояний: 0 – ОУ, 1 – цель.

Решение рассматриваемой задачи в замкнутой форме возможно с привлечением положения теории аналитического конструирования агрегированных регуляторов [13,14]. Однако сложность аналитических соотношений и прогнозируемые трудозатраты на исследование полученных законов методами статистического моделирования могут нивелировать ожидаемый выигрыш от учета технических и технологических нелинейностей (ограничений). Исходя из вышеизложенного рассмотрим инженерный метод решения задачи синтеза метода наведения, который предполагает использование эвристического подхода¹ для обеспечения изложенных выше требований с учетом структурных ограничений.

Постановка задачи. На рис. 2 представлен сектор сопровождения ОУ в картинной плоскости с обозначениями угловых размеров секторов соответствующих антенных систем. Сечение сектора сопровождения образовано пересечением диаграмм направленности узконаправленных антенн (ДНА) станции сопровождения цели (ССЦ), антенн станции сопровождения ОУ и цели (ССРЦ) с линейным сканированием и станции передачи команд (СПК) узкого луча (УЛ) и широкого луча (ШЛ) с плоскостью, ортогональной оптической оси антенн (ЛВ цели).

Задача заключается в формировании управляющих команд таким образом, чтобы обеспечить удержание ОУ в пределах сектора сканирования ССРЦ, образованного следом от перемещающихся в пространстве главных лучей ДНА с заданными размерами. При этом должно быть обеспечено выполнение следующих требований к методу наведения:

- угловое упреждение (отставание) ОУ относительно ЛВ цели не должно превышать указанных выше значений;
- величины флуктуационных и кинематических нормальных ускорений (равно как и величины УСВ ЛВ ОУ – цель) в процессе наведения должны быть минимальными;

¹Эвристический алгоритм – это алгоритм решения задачи, правильность которого для всех возможных случаев не доказана, но про который известно, что он дает достаточно хорошее решение в большинстве случаев. Алгоритм решения задачи включает практический метод, не являющийся гарантированно точным или оптимальным, но достаточный для решения поставленной задачи, и позволяет ускорить решение задачи в тех случаях, когда точное решение не может быть найдено. В действительности может быть даже известно (т. е. доказано) то, что эвристический алгоритм формально неверен. Его все равно можно применять, если при этом он дает неверный результат только в отдельных, достаточно редких и хорошо выделяемых случаях или дает неточный, но все же приемлемый результат. Проще говоря, эвристика – это не полностью математически обоснованный (или даже «не совсем корректный»), но при этом практически полезный алгоритм.

– величины мгновенного линейного промаха и скорости его изменения в точке встречи не должны превышать допустимых значений, определяемых характеристиками боевого снаряжения ОУ.

Выполнение первого и последнего требований является безусловным, так как выход ОУ за пределы сектора сопровождения приведет к потере ответного сигнала и срыву процесса наведения. Нулевые значения промаха и его производной в точке встречи обеспечиваются соответствующим выбором коэффициентов штрафа при фазовых координатах.

Второе требование обусловлено тем, что величина необходимых нормальных ускорений ОУ зависит не только от неконтролируемых внешних ускорений (ускорения маневра, свободного падения, продольного ускорения ОУ), но и от значения навигационного коэффициента, которое можно выразить следующим образом [15]:

$$W_p(t) = \frac{N_0}{N_0 - 2} \left(1 - \left(\frac{D}{D_1} \right)^{N_0 - 2} \right) W_u, \quad (6)$$

где D, D_1 – текущая дальность и дальность начала маневра цели соответственно; W_u – ускорение маневра цели.

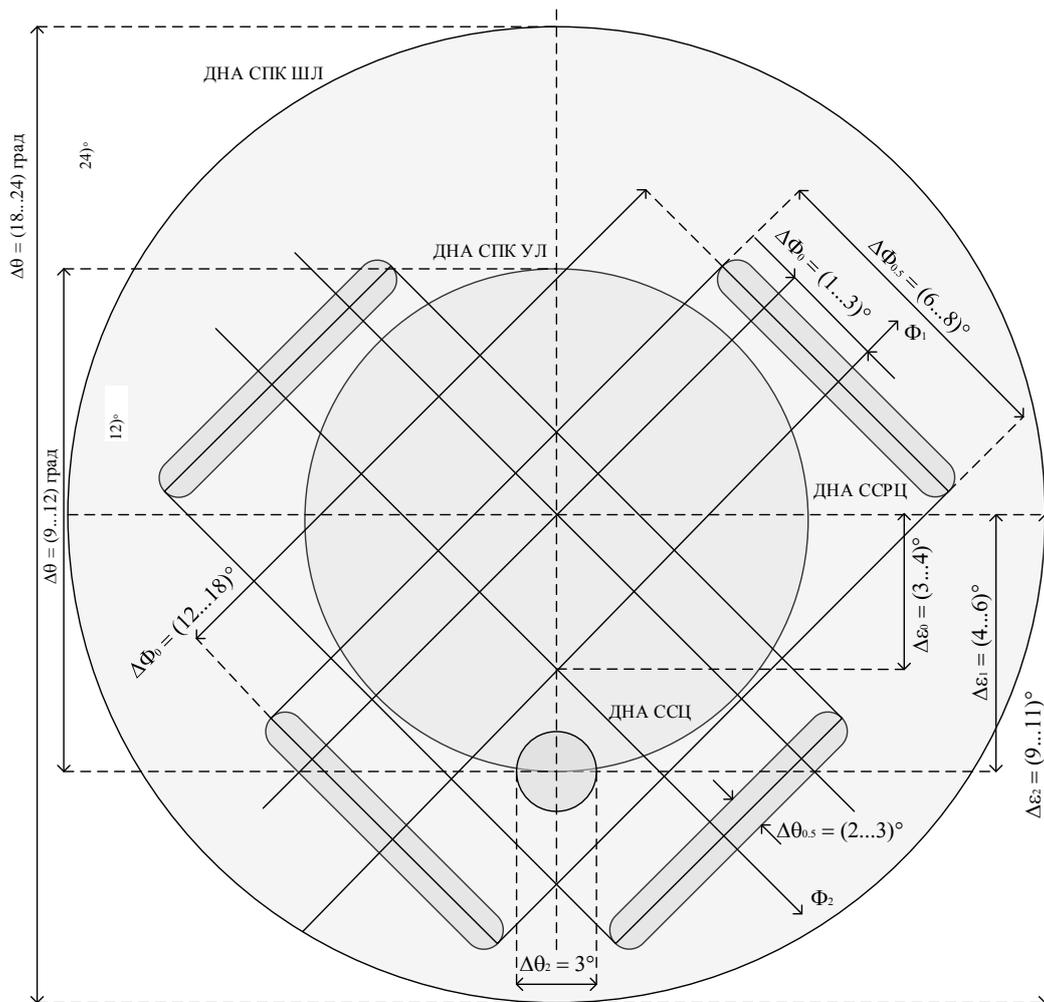


Рис. 2. Сектор углового сопровождения ракет в картинной плоскости, ортогональной ЛВ СНР – цель

При ограничении ускорений ОУ будет происходить неконтролируемый рост промаха h , величина которого с приемлемой для практики точностью определяется как [15]

$$h = \frac{W_u D_1^2}{2|\dot{D}|} \left(1 - \frac{W_p}{W_{mp}}\right) \left(1 - \frac{W_p}{W_{mp}}\right)^{\frac{2}{N_0-2}}, \quad (7)$$

где W_p – максимальные ускорения, развиваемые ракетой; W_{mp} – требуемые для наведения МПН ускорения ОУ; \dot{D} – относительная скорость сближения ОУ и цели.

Синтез оптимального закона перехвата вариационными методами приводит к необходимости решения многомерной нелинейной двухточечной краевой задачи в процессе наведения в реальном масштабе времени, что является весьма сложной и трудоемкой в алгоритмическом плане процедурой.

Использование методов синергетической теории управления позволяет получить оптимальное решение для задачи перехвата в замкнутой алгебраической форме. Однако громоздкость аналитических выражений и большой объем требуемых вычислений для оценки фазовых координат наряду со сложностью получения методики для практического анализа точности и динамических свойств существенно снижают прикладную ценность результатов синтеза.

Таким образом, одной из основных задач, стоящих перед разработчиками систем управления, является поиск научных и инженерных решений, позволяющих снизить влияние указанных факторов на характеристики точности, устойчивости и чувствительности системы наведения ракет при сохранении приемлемой для практической реализации технической сложности проектируемого устройства. Перечислим возможные варианты решения задачи наведения ОУ в условиях ограниченного сектора сопровождения:

- формирование дополнительных корректирующих сигналов (наряду с сигналами компенсации продольного ускорения, сил тяжести и маневра цели) в контуре оценки УСВ ЛВ ОУ – цель, зависящих от текущего углового упреждения относительно ЛВ СНР – цель, на основе простейших схем регулирования, например, нелинейных модификаций классического пропорционально-интегрального дифференцирующего (ПИД) регулятора;

- расчет кинематической (заданной) УСВ ЛВ ОУ – цель по расчетным кинематическим угловым координатам (УК) ОУ с последующим сравнением заданной и фактической УСВ ЛВ ОУ – цель и плавным переходом на классический МПН;

- комбинированное управление за счет плавного сопряжения траекторий трехточечных методов наведения (полного или частичного спрямление траектории, метода параллельного сближения) и двухточечных (например, МПН);

- комбинированное управление с использованием наведения с постоянным или переменным углом упреждения (метода погони) на начальном участке и МПН – на конечном участке;

- использование переменного навигационного коэффициента в выражениях для МПН, адаптивного к прогнозируемым параметрам встречи и позволяющего опосредованно управлять величиной углового упреждения;

- использование упрощенной эталонной модели движения ОУ для прогнозирования и предварительной оценки максимальных углов упреждения цели для последующей коррекции параметров МПН (навигационного коэффициента или сигнала компенсации);

- комбинация двух или более перечисленных способов на различных этапах наведения.

Из всего перечня способов учета технологических ограничений на допустимые углы упреждения ОУ ограничимся детальным рассмотрением подхода, основанного на использовании отдельного контура коррекции по сигналам углового рассогласования совместно с адаптацией навигационного коэффициента к параметрам наведения.

Модифицированный МПН для ограниченного сектора сопровождения БПЛА. Геометрия задачи пространственного наведения ОУ на цель изображена на рис. 3. Обозначения осей сопровождающих (динамических) и статических ортонормированных триэдров, связанных с полюсами (центрами масс) объектов, соответствуют приведенным ранее.

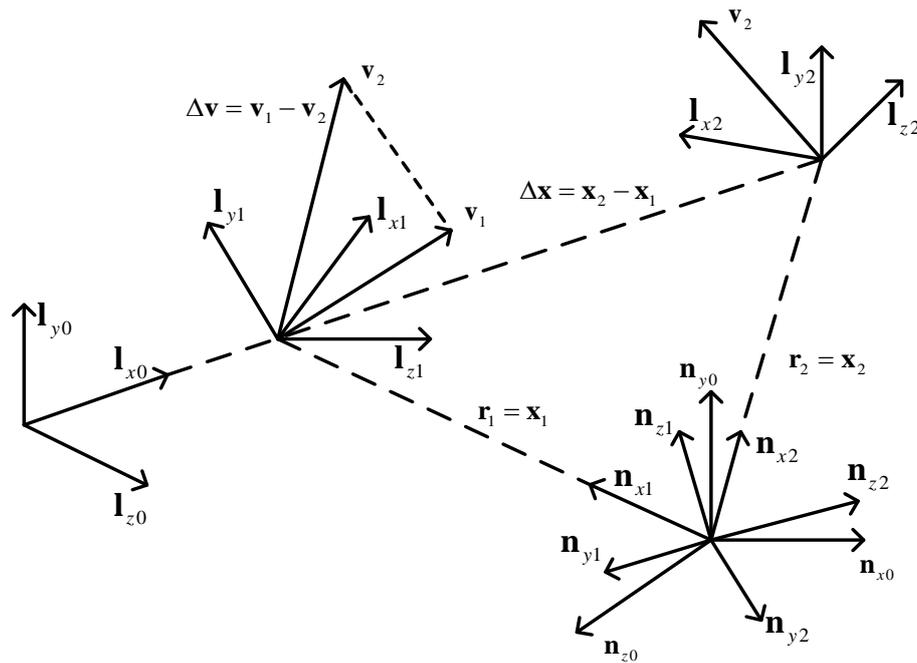


Рис. 3. Системы координат, используемые при рассмотрении взаимного перемещения ОУ и цели

Выражение для модифицированного МПН с использованием контура коррекции в общем случае имеет следующий вид (для вертикального канала наведения):

$$\begin{aligned}
 W_y &= N_y(t) \left(\omega_z + \Delta \bar{\omega}_z + \sum_{j=0}^2 \Delta \omega_{zj} \right), \\
 \Delta \dot{\omega}_z &= k_{od} \frac{\partial f(\Delta \varphi_z)}{\partial t} + k_{op} f(\Delta \varphi_z) + \Delta \zeta_z, \\
 \Delta \dot{\zeta}_z &= k_{oi} f(\Delta \varphi_z) + k_{oi} \delta \omega_z, \\
 \delta \dot{\omega}_z &= k_{\beta \omega} \left(\left(\Delta \omega_z - k_z \operatorname{th} \left(\frac{\Delta \omega_z}{k_z} \right) \right) - \delta \omega_z \right), \\
 \Delta \dot{\bar{\omega}}_z &= k_{\omega \alpha} \left(k_z \operatorname{th} \left(\frac{\Delta \omega_z}{k_z} \right) - \Delta \bar{\omega}_z \right), \\
 N_y(t) &= N_0(t) |\dot{D}(t)|,
 \end{aligned} \tag{8}$$

где $\dot{D}(t)$ – скорость сближения ОУ и цели; $\Delta \bar{\omega}_z$ – поправка по УСВ ω_z на выходе корректирующего устройства на основе модифицированного нелинейного ПИД-регулятора; k_z – нормировочный коэффициент функции $\operatorname{tanh}(x)$ для плавного ограничения; $k_{\omega \alpha}$ – коэффициент сходимости сглаживающего фильтра на выходе ограничителя ПИД-регулятора; k_{od}, k_{op}, k_{oi} – коэффициенты дифференцирующего, пропорционального и интегрирующего каналов коррекции по УСВ ЛВ; $\Delta \varphi_z, \Delta \zeta_z$ – переменные состояния ПИД-регулятора; $\Delta \omega_{zj}$ – корректирующая поправка по УСВ, обусловленная воздействием j -го внешнего возмущающего ускорения; $f(\Delta \varphi_z)$ – функция, обеспечивающая быстрый рост сигнала ошибки от входного рассогласования.

Функциональная зависимость навигационного параметра $N_0(t)$ подбирается по результатам математического моделирования. Диапазон изменения параметра составляет $N_0(t) \in (2...6)$ в зависимости от прогнозируемых условий встречи. Одна из типовых зависимостей имеет следующий вид:

$$N_0(t) = N_{y0} + \Delta N_0 \exp\left(-\frac{(t-t_0)}{\Delta t(\tau)}\right), \quad (9)$$

где $N_{y0} = 2$ – начальное значение навигационного коэффициента (реализуется метод наведения с постоянным пеленгом); $\Delta N_0 \in (2...4)$ – приращение параметра за время наведения; $\Delta t(\tau)$ – постоянная времени, зависящая от текущего значения подлетного времени τ и определяющая время достижения коэффициентом $N_0(t)$ установившегося значения. Величина $\Delta t(\tau)$ выбирается опытным путем или по результатам математического моделирования и может быть аппроксимирована как

$$\Delta t(\tau) = \Delta t_0 + \delta t_0 \operatorname{th}\left(\frac{\tau - \tau_0}{\delta \tau_0}\right), \quad (10)$$

где Δt_0 – минимальное начальное значение постоянной времени; $\delta \tau_0$ – величина, определяющая время нарастания параметра $\Delta t(\tau)$ от минимального до максимального значения; δt_0 – приращение величины $\Delta t(\tau)$.

Структурная схема нелинейного ПИД-регулятора изображена на рис. 4. Функциональный преобразователь $f(\Delta\varphi)$ на входе ПИД-регулятора выполняет следующие функции:

- обеспечивает нелинейную зависимость крутизны преобразования от величины входного сигнала ошибки $\Delta\varphi$;
- обеспечивает плавную коммутацию цепи коррекции при достижении граничного значения $\Delta\varphi_0(\Delta\dot{\varphi})$;
- исключает случайную коммутацию (дребезг) сигнала рассогласования вблизи граничного значения.

Назначением ограничителя сигнала ошибки на входе интегрирующей цепи является контролируемое увеличение сигнала регулирования по соответствующей компоненте выходного сигнала. Назначение ограничителя с обратной связью на выходе ПИД-регулятора – предотвращение насыщения выходного сигнала при больших величинах сигнала рассогласования на входе и уменьшение времени выхода из зоны ограничения. Типовые функции могут использовать зоны нечувствительности для плавного ввода сигнала ошибки в контур коррекции:

$$f(\Delta\varphi) = \left(\left| \Delta\varphi \right| < \Delta\varphi_0, 0, \varphi_0 \operatorname{tg}\left(\frac{\Delta\varphi - \Delta\varphi_0(\Delta\dot{\varphi}) \operatorname{sign}(\Delta\varphi)}{\varphi_0}\right) \right) \quad (11)$$

или

$$f(\Delta\varphi) = \left(\left| \Delta\varphi \right| < \Delta\varphi_0, 0, \sum_{j=0}^3 \kappa_j (\Delta\varphi - \Delta\varphi_0(\Delta\dot{\varphi}) \operatorname{sign}(\Delta\varphi))^j \right), \quad (12)$$

где $\Delta\varphi_0(\Delta\dot{\varphi})$ – смещение гистерезисного типа относительно нулевого значения сигнала рассогласования, при котором включается нелинейный контур коррекции; κ_j – постоянные полиномиальные коэффициенты.

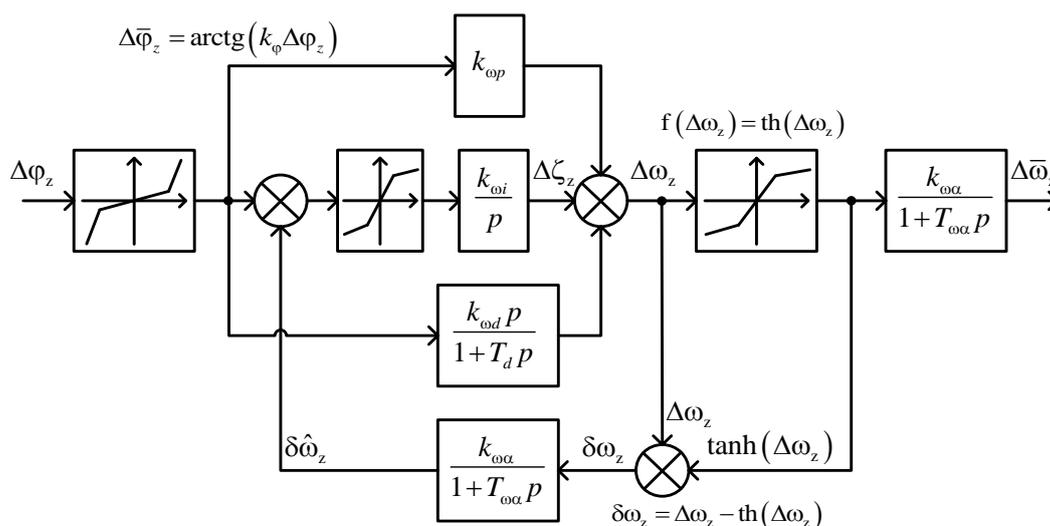


Рис. 4. Структурная схема нелинейного ПИД-регулятора канала углового упреждения

Рассмотрим возможные варианты поэтапного расчета управляющих ускорений ОУ. Первый вариант представляет собой векторную форму алгоритма расчета УСВ в проекциях на оси СК ЛВ и описывается следующими выражениями:

$$\begin{aligned}
 \boldsymbol{\omega}_2 &= \mathbf{M}_2 \cdot (\boldsymbol{\omega}_0 + \Delta\bar{\boldsymbol{\omega}}) + \sum_{j=0}^2 \Delta\boldsymbol{\omega}_j, \quad \boldsymbol{\omega}_0 = \frac{(\Delta\mathbf{x} \times \Delta\mathbf{v})}{(\Delta\mathbf{x}^T \Delta\mathbf{x})}, \\
 \Delta\dot{\boldsymbol{\omega}} &= k_{\omega d} \frac{\partial \mathbf{f}(\Delta\varphi_y, \Delta\varphi_z)}{\partial t} + k_{\omega p} \mathbf{f}(\Delta\varphi_y, \Delta\varphi_z) + \Delta\boldsymbol{\zeta}, \\
 \Delta\dot{\boldsymbol{\zeta}} &= k_{\omega i} \mathbf{f}(\Delta\varphi_y, \Delta\varphi_z) + k_{\omega l} (\Delta\boldsymbol{\omega} - \text{th}(\Delta\boldsymbol{\omega})), \\
 \Delta\dot{\bar{\boldsymbol{\omega}}} &= k_{\omega\alpha} (\text{th}(\Delta\boldsymbol{\omega}) - \Delta\bar{\boldsymbol{\omega}}), \\
 \Delta\boldsymbol{\omega}_j &= \sum_{j=0}^2 \frac{\Delta\mathbf{W}_j}{2|\dot{\mathbf{D}}|},
 \end{aligned} \tag{13}$$

где $\Delta\bar{\boldsymbol{\omega}}$ – корректирующая поправка для предотвращения выхода ОУ за пределы сектора сопровождения; $\Delta\boldsymbol{\omega}_j$ – компенсирующие поправки к фактической УСВ ЛВ, обусловленные возмущающим воздействием внешних ускорений; $\Delta\mathbf{x} = \mathbf{x}_2 - \mathbf{x}_1$ – разность координат цели и ОУ в проекциях на оси земной системы координат (ЗСК); $\Delta\mathbf{v} = \mathbf{v}_2 - \mathbf{v}_1$ – разность производных координат цели и ОУ в проекциях на оси ЗСК; $\mathbf{M}_2 = \mathbf{M}_2(\theta_2, \varphi_2)$ – матрица перехода от ЗСК к СК ЛВ, зависящая от углов ориентации ЛВ в ЗСК (θ_2, φ_2) ; $\Delta\mathbf{W}_j$ – поправки для компенсации возмущающего воздействия внешних неконтролируемых ускорений; $|\dot{\mathbf{D}}| = |\Delta\mathbf{v}|$ – модуль вектора скорости сближения; $k_{\omega d}, k_{\omega p}, k_{\omega i}$ – коэффициенты дифференцирующего, пропорционального и интегрирующего каналов коррекции по нормальным ускорениям; $\mathbf{f}(\Delta\varphi_y, \Delta\varphi_z)$ – векторная функция нелинейного преобразователя на входе ПИД-регулятора; $\mathbf{f}(\Delta\boldsymbol{\omega}) = \text{th}(\Delta\boldsymbol{\omega})$ – векторная функция векторного аргумента «гиперболический тангенс».

Управляющие нормальные ускорения определяются следующим образом:

$$\mathbf{W}_{\rho 2} = N_y(t) \mathbf{M}_3 \left(\boldsymbol{\omega}_2 \times \frac{\mathbf{v}_1}{|\mathbf{v}_1|} \right), N_y(t) = N_0(t) |\dot{D}_1|, \quad (14)$$

$$|\dot{D}_1| = \frac{|\dot{D}|}{\cos(\varepsilon_0) \cos(\beta_0)}, \dot{D} = \frac{\Delta \mathbf{x} \cdot \Delta \mathbf{v}}{|\Delta \mathbf{x}|},$$

где $\mathbf{M}_3 = \mathbf{M}_3(\boldsymbol{\Theta}_3) = \mathbf{M}_1 \mathbf{M}_2^T$ – матрица перехода от СК ЛВ к СвСК, зависящая от углов ориентации ОУ в СК ЛВ $\boldsymbol{\Theta}_3 = (\nu_3, \psi_3, \zeta_3)$, полученных с учетом пространственного крена ОУ при условии идеальной работы системы стабилизации; $N_0(t) \in (2 \dots 10)$ – навигационный коэффициент, зависящий от текущих условий наведения и прогнозируемых параметров встречи; $\mathbf{M}_1 = \mathbf{M}_1(\nu_1, \psi_1, \zeta_1)$ – матрица перехода от земной СК к СвСК, характеризующаяся углами тангажа ν_1 , рыскания ψ_1 и крена ζ_1 ОУ; ε_0, β_0 – угол места и азимут цели в СвСК ОУ (углы относительного пеленга цели).

Запишем альтернативный вариант оценки управляющих ускорений в СвСК:

$$\mathbf{W}_{\rho 1} = \mathbf{M}_3 \mathbf{W}_{\rho 2},$$

$$\mathbf{W}_{\rho 2} = N_y(t) \left(\mathbf{M}_2 \left(\boldsymbol{\omega}_0 \times \frac{\mathbf{v}_1}{|\mathbf{v}_1|} \right) + \Delta \mathbf{W} + \sum_{j=0}^2 \Delta \mathbf{W}_{0j} \right), \quad (15)$$

$$\Delta \dot{\mathbf{W}} = k_{wd} \frac{\partial \mathbf{f}(\Delta \varphi_y, \Delta \varphi_z)}{\partial t} + k_{wp} \mathbf{f}(\Delta \varphi_y, \Delta \varphi_z) + \Delta \mathbf{Q},$$

$$\Delta \dot{\mathbf{Q}} = k_{wi} \mathbf{f}(\Delta \varphi_y, \Delta \varphi_z),$$

где k_{wd}, k_{wp}, k_{wi} – коэффициенты дифференцирующего, пропорционального и интегрирующего каналов коррекции по нормальному ускорению.

Методика расчета компенсирующих (корректирующих) сигналов по нормальным ускорениям $\Delta \mathbf{W}_{0j}$ будет рассмотрена в следующем разделе.

Формирование корректирующих сигналов. В процессе наведения на ракету помимо управляющих нормальных ускорений в СК ЛВ действуют следующие задающие и возмущающие ускорения:

- вектор продольного ускорения ОУ \mathbf{W}_{x1} , обусловленного переменной величиной тяги двигателя и высотным градиентом параметров атмосферы;
- вектор ускорения свободного падения \mathbf{g} в заданной точке пространства;
- вектор ускорения цели \mathbf{W}_τ , вызванный случайным маневром.

С целью компенсации указанных возмущающих воздействий необходимо сформировать соответствующие сигналы компенсации. Для этого обобщим частную методику расчета вектора управляющих ускорений, изложенную в [10, 15]. Запишем в векторной форме уравнения для оценки ускорений, действующих на ОУ:

$$\mathbf{W}_\rho = \mathbf{W}_{x1} + \mathbf{W}_\rho + \mathbf{G}_0 + \mathbf{W}_2, \quad (16)$$

где $\mathbf{W}_{x1} = (W_{x1} \ 0 \ 0)^T$ – вектор продольного ускорения ОУ; \mathbf{W}_p – вектор фактических нормальных ускорений ОУ в СК ЛВ; \mathbf{G}_0 – вектор ускорения силы тяжести; \mathbf{W}_2 – вектор ускорения маневра цели.

Необходимо отметить, что расчет оценки вектора нормальных ускорений цели \mathbf{W}_2 является отдельной нетривиальной и специфической задачей, которая будет рассмотрена в дальнейших публикациях.

Вектор полных ускорений ОУ в проекциях на оси СК ЛВ запишется как

$$\mathbf{W}_p = k_1 \mathbf{1}_{x0} + k_2 \mathbf{1}_{y0} + k_3 \mathbf{1}_{z0}, \quad (17)$$

где k_1, k_2, k_3 – неопределенные коэффициенты, которые необходимо вычислить.

Введем векторные ограничения, которые формулируются следующим образом:

– проекция на продольную ось ОУ результирующего вектора ускорений равна величине собственного продольного ускорения ОУ;

– проекция на нормальные к ЛВ оси СК результирующего вектора ускорений ОУ определяется соответствующими величинами кинематических (заданных) ускорений.

Ограничения в аналитической форме примут следующий вид:

$$\begin{aligned} \mathbf{W}_p \cdot \mathbf{1}_{x1} &= W_{x1}, \\ \mathbf{W}_p \cdot \mathbf{1}_{y0} &= W_{y0} = +N_0(t) \left| \dot{D} \right| \omega_{z0}, \\ \mathbf{W}_p \cdot \mathbf{1}_{z0} &= W_{z0} = -N_0(t) \left| \dot{D} \right| \omega_{y0}. \end{aligned} \quad (18)$$

Подставив (17) в (18) и выполнив необходимые преобразования, получим

$$\begin{aligned} \mathbf{W}_{p1} &= \mathbf{M}_1 \mathbf{W}_{p0}, \\ \mathbf{W}_{p0} &= \mathbf{W}_0 + \left(\frac{W_{x1} - \mathbf{W}_0 \cdot \mathbf{1}_{x1}}{\mathbf{1}_{x0} \cdot \mathbf{1}_{x1}} \right) \mathbf{1}_{x0} - \mathbf{M}_0 (\mathbf{G}_0 - \mathbf{N}_2^T \mathbf{W}_2), \end{aligned} \quad (19)$$

где $\mathbf{W}_{p0} = (W_{px0} \ W_{py0} \ W_{pz0})^T$ – вектор ускорений ОУ в проекциях на оси СК ЛВ с учетом компенсации возмущающих (внешних) ускорений; $\mathbf{W}_0 = (0 \ W_{y0} \ W_{z0})^T$ – вектор кинематических ускорений в СК ЛВ; $\mathbf{G}_0 = (0 \ -9,81 \ 0)^T$ – вектор ускорения свободного падения в ЗСК, связанной с точкой стояния СНР; \mathbf{M}_0 – матрица перехода из ЗСК в СК ЛВ, характеризующаяся вертикальным θ_0 и горизонтальным ϕ_0 углами ориентации ЛВ РЦ; \mathbf{N}_2 – матрица перехода от ИСК цели к ЗСК, характеризующаяся углами места и азимута ε_2, β_2 ; $\mathbf{1}_{x0} \mathbf{1}_{x1}$ – произведение ортов соответствующих СК, которое определяется первым диагональным элементом матрицы перехода от СК ЛВ к СвСК.

Полученные аналитические соотношения являются обобщением алгоритма оценки команд управления ОУ, изложенного в [10] и не учитывающего воздействие гравитационного ускорения и ускорения маневра цели.

Метод расчета результирующих управляющих ускорений, основанный на прямой компенсации ускорений без учета ограничения $\mathbf{W}_p \cdot \mathbf{1}_{x1} = W_{x1}$ и предполагающий компенсацию возмущающих нормальных ускорений в СК ЛВ, приводит к следующему результату:

$$\begin{aligned} \mathbf{W}_{p1} &= \mathbf{M}_3(\Theta_3) \mathbf{O}_3 \mathbf{W}_{p0}, \mathbf{W}_{p0} = \mathbf{W}_0 + \Delta \mathbf{W}_0, \\ \mathbf{W}_{p0} &= \left(\mathbf{W}_0 - \mathbf{M}_0 \left(\mathbf{G}_0 - \mathbf{M}_1^T \mathbf{O}_1 \mathbf{W}_1 - \mathbf{N}_2^T \mathbf{W}_2 \right) \right), \end{aligned} \quad (20)$$

где $\mathbf{M}_3(\Theta_3)$ – матрица перехода от СК ЛВ к СвСК, зависящая от вектора углов Θ_3 ; \mathbf{O}_j – единичная матрица размером 3×3 с нулевым j -м диагональным элементом; $\Delta \mathbf{W}_0$ – вектор ускорений, компенсирующих возмущающее воздействие внешних ускорений.

Моделирование показывает, что второй вариант расчета ускорений ОУ дает некоторый выигрыш для типовых траекторий наведения (с точки зрения минимизации промаха). Заметные преимущества появляются лишь при обстреле интенсивно маневрирующих целей, а также целей, движущихся с большими угловыми скоростями.

В дальнейшем для математического моделирования и статистического анализа будем использовать соотношения (15) и (20).

Учет пространственного скручивания ИСК и СвСК. Вследствие пространственного движения ОУ и цели, а также из-за ошибок системы стабилизации ОУ по крену возникает расхождение между осями СК ЛВ и СвСК, что приводит к снижению запасов устойчивости контура наведения и увеличению случайной компоненты мгновенного промаха.

Ошибки системы стабилизации по крену вызваны возмущающим действием момента «косой обдувки» и не превышают в большинстве случаев (3...5)°. Пространственный крен ОУ появляется при повороте в двух плоскостях – по рысканию и тангажу.

Пространственный крен ОУ в гироскопной СК (ГСК) или СК свободного гироскопа устраняется системой стабилизации ОУ по крену. Остаточный крен ОУ в СК ЛВ при условии равенства нулю крена в ГСК называется углом пространственного скручивания [15].

Получим соотношения для расчета углов ориентации СвСК в СК ЛВ при условии работы системы стабилизации ОУ по крену. С этой целью запишем кинематические уравнения для расчета углов ориентации ОУ в ГСК:

$$\dot{\Theta}_4 = \mathbf{M}_\omega(\Theta_4) \Delta \omega_3, \quad (21)$$

$$\dot{\Theta}_4 = \begin{pmatrix} \dot{\vartheta}_4 \\ \dot{\psi}_4 \\ \dot{\gamma}_4 \end{pmatrix}, \mathbf{M}_\omega(\Theta_4) = \begin{pmatrix} 0 & \omega_{y1} & \omega_{z1} \\ 0 & \frac{\cos \gamma_4}{\cos \vartheta_4} & -\frac{\sin \gamma_4}{\cos \vartheta_4} \\ 1 & -\operatorname{tg} \vartheta_4 \cos \gamma_4 & \operatorname{tg} \vartheta_4 \sin \gamma_4 \end{pmatrix},$$

и преобразуем их при условии равенства угла крена в ГСК нулю ($\dot{\gamma}_4 = \gamma_4 = 0$):

$$\dot{\psi}_4 = \frac{\omega_{y1}}{\cos \vartheta_4}, \dot{\vartheta}_4 = \omega_{z1}, \quad (22)$$

где $\mathbf{M}_\omega(\Theta_4) = \mathbf{M}_\omega(\vartheta_4, \psi_4, \gamma_4)$ – матрица перехода от вектора УСВ, заданного в проекциях на оси соответствующей СК, к производным углов последовательного поворота относительно промежуточных осей (углов Эйлера – Крылова).

Спроецируем последовательно УСВ СвСК на оси неподвижной ГСК и на оси СК ЛВ:

$$\omega_3 = \mathbf{M}(\varepsilon_1, \beta_1) \mathbf{M}(\vartheta_0, \psi_0)^T \mathbf{M}(\vartheta_4, \psi_4)^T \omega_1, \quad (23)$$

где ω_3 – проекции УСВ СвСК на оси СтСК; ϑ_0, ψ_0 – углы ориентации ГСК в ЗСК, определяемые начальными углами ориентации ОУ (фиксируются в момент старта).

Запишем выражение для относительных УСВ ОУ в СК ЛВ:

$$\Delta\omega_3 = \omega_3 - \mathbf{M}_0(\theta_2, \varphi_2)\omega_0, \quad (24)$$

где $\mathbf{M}_0(\theta_2, \varphi_2)$ – матрица перехода от ЗСК к СК ЛВ.

Отсюда несложно получить кинематические уравнения для расчета вектора углов ориентации ОУ (СвСК) в СтСК Θ_3 при условии идеальной работы системы стабилизации по крену:

$$\dot{\Theta}_3 = \mathbf{M}_\omega(\Theta_3)\Delta\omega_3, \quad (25)$$

где Θ_3 – вектор углов ориентации СвСК в СК ЛВ при условии $\dot{\gamma}_4 = \gamma_4 = 0$.

Углы используются в матрице поворота для пересчета управляющих ускорений ОУ из СК ЛВ в СвСК, полученных в соответствии с выбранным методом наведения, согласно выражению (20):

$$\mathbf{W}_{p1} = \mathbf{M}_3(\Theta_3)\mathbf{O}_3\mathbf{W}_{p0}, \quad (26)$$

где $\mathbf{M}_1(\Theta_3)$ – матрица перехода от СК ЛВ к СвСК с учетом работы системы стабилизации по крену ОУ.

Следует напомнить, что МПН может быть реализован с использованием как наземных измерителей (СНР), так и бортовых (полуактивной или активной ГСН). В первом случае для расчета управляющих ускорений используются УСВ и углы тангажа, рыскания и крена скоростной СК (СкСК). При этом наведение осуществляется за счет управления ориентацией вектора скорости ОУ. Во втором случае (при отсутствии датчиков углов атаки на борту) для управления используются УСВ и углы ориентации СвСК (планера ОУ) в предположении малости углов атаки и скольжения $\alpha = (3...5)^\circ$, $\beta = (3...5)^\circ$ в большинстве практически важных приложений. При этом наведение выполняется путем управления ориентацией вектора скорости ОУ.

Результаты математического моделирования. Для оценки эффективности предлагаемой модификации МПН было выполнено математическое моделирование процесса наведения ОУ на аэробаллистическую цель. Варианты расчета траектории наведения ракет для заданных начальных условий приведены в табл. 1 и 2. Моделировалась динамика движения тактического истребителя, летящего прямолинейно и равномерно (варианты 1 и 3) и совершающего маневр кабрированием-пикированием с перегрузкой 80 м/с^2 (вариант 4), а также противорадиолокационной ОУ (вариант 2) для экстремальных по дальности, высоте, параметру и скорости цели точек зоны поражения. На рис. 5–16 изображены зависимости сферических координат цели ε_2, β_2 , ОУ ε_1, β_1 , относительных координат ОУ Φ_1, Φ_2 , а также компонент нормальных ускорений ОУ в скоростной СК W_x, W_y, W_z от полетного времени.

Таблица 1

Условия наведения, наблюдения и встречи ОУ с целью

Вариант расчета	Координаты точки встречи $x, y, z, \text{ м}$	Скорость цели, м/с	СКО ошибки измерения УК цели, угл. мин	СКО ошибки измерения дальности цели, м	СКО ошибки измерения скорости цели, м/с
1	–2000, 100, 15 000	500	1...2	2...4	3...5
2	2000, 14 000, 14 000	750	2...3	3...5	4...8
3	25 000, 20 000, 15 000	300	4...6	4...5	6...8
4	20 000, 10 000, 5000	300	2...4	1...3	4...6

Таблица 2

Ошибки наведения и мгновенный промах ОУ в точке встречи

Вариант расчета	m_w, σ_w для W_0 , ММПН / МПСТ, м/с ²	m_ζ, σ_ζ для h_0 , ММПН / МПСТ, м	m_γ, σ_γ для γ_0 , ММПН / МПСТ, м
1	(-65, 06) / (50, 04)	(19, 3) / (24, 5)	(15, 2) / (21, 4)
2	(55, 05) / (75, 06)	(27, 4) / (37, 3)	(24, 3) / (32, 3)
3	(09, 07) / (19, 11)	(16, 3) / (39, 5)	(14, 3) / (35, 4)
4	(45, 08) / (23, 03)	(31, 3) / (18, 4)	(29, 3) / (16, 3)

В процессе моделирования проведен сравнительный анализ точности наведения с использованием метода полного спрямления траектории (МПСТ) и предлагаемой модификации метода пропорциональной навигации (ММПН) для различных типов траектории полета цели и условий встречи. Кинематические углы в соответствии с МПСТ формировались согласно выражению

$$\varepsilon_3 = \varepsilon_1 + k_\varepsilon(t) \omega_{\varepsilon 1} \tau, \quad \beta_3 = \beta_1 + k_\beta(t) \omega_{\beta 1} \tau, \quad (27)$$

где ε_1, β_1 – текущий угол места и азимут цели; ε_3, β_3 – кинематический угол места и азимут ОУ; $\omega_{\varepsilon 1}, \omega_{\beta 1}$ – УСВ ЛВ цели; $k_\varepsilon(t), k_\beta(t)$ – коэффициенты упреждения.

Для модели использовались типовые структуры измерителей координат цели и ОУ, функционирующие совместно с системой управления силовыми электрическими следящими приводами антенны сопровождения цели и ОУ. СКО флуктуационной ошибки измерения УК цели задавалось в диапазоне $\sigma_\varphi = (2...6)$ угл. мин, СКО измерения дальности цели – в пределах $\sigma_r = (1...5)$ м, СКО измерения скорости цели – в пределах $\sigma_v = (3...8)$ м/с, математические ожидания (МО) ошибок измерения координат цели равны нулю. Начальные условия (координаты встречи и СКО ошибок измерения) представлены в табл. 1. Статистические оценки ошибок наведения и мгновенных промахов в момент встречи для различных методов наведения ОУ (табл. 2) характеризуются следующими параметрами:

- координатами точки встречи ОУ с целью $\mathbf{r}_m = (x_m \ y_m \ z_m)^T$ и скоростью цели V_u ;
- используемым методом наведения ОУ (ММПН, МПСТ), а также типом и интенсивностью маневра цели W_u (варианты расчетов 1, 3 и 4 соответствуют цели типа «истребитель», вариант 3 – цели типа «противорадиолокационная ракета»);
- СКО ошибок измерения УК σ_φ , дальности σ_r и скорости σ_v цели, определяющим величину флуктуационной нормальной перегрузки ОУ;
- МО и СКО величины мгновенного промаха² $\sigma_\gamma, m_\gamma, \gamma_0 = \|\gamma\|, \mathbf{h} = (0 \ \gamma_y \ \gamma_z)^T$ и ошибки наведения (линейного расстояния между ракетой и целью) m_ζ, σ_ζ в картинной плоскости $h_0 = \|\mathbf{h}\|, \mathbf{h} = (0 \ h_y \ h_z)^T$, а также МО и СКО нормальных ускорений ОУ в СКСК в момент встречи.

Величина мгновенного промаха рассчитывалась по формуле, учитывающей влияние ненулевого нормального ускорения ОУ в точке встречи:

$$\mathbf{h}_\gamma = \left[\boldsymbol{\omega} \times (\Delta \mathbf{v} + \Delta \mathbf{a} \cdot \tau_y) \cdot \tau_y \right] \cdot \tau_y, \quad \boldsymbol{\omega} = \frac{\Delta \mathbf{r} \times \Delta \mathbf{v}}{\Delta r^2}, \quad (28)$$

$$\tau_y = \frac{\Delta \mathbf{r} \cdot \Delta \mathbf{v}_0}{\|\Delta \mathbf{v}\|}, \quad \Delta \mathbf{v}_0 = \frac{\Delta \mathbf{v}}{\|\Delta \mathbf{v}\|}, \quad \Delta \mathbf{a} = \Delta \dot{\mathbf{v}},$$

где τ_y – экстраполированное время до встречи ОУ с целью (упрежденное время встречи); $\Delta \mathbf{a}$ – вектор полного ускорения ОУ в СКЛВ.

²Текущей мерой точности наведения, т. е. величиной, прогнозирующей наименьшее значение модуля относительной дальности, принято считать дальность, на которой ракета пролетит мимо цели, если с данного момента времени ракета и цель в относительном движении будут перемещаться прямолинейно (по касательной к относительной траектории). Эту дальность называют пролетом, или промахом. Возможны и другие меры текущей точности, например, расстояние, на котором ракета пролетит мимо цели, если с данного момента ускорение будет постоянным. В момент времени, когда вектор пролета совпадает с вектором дальности, расстояние между ракетой и целью минимально.

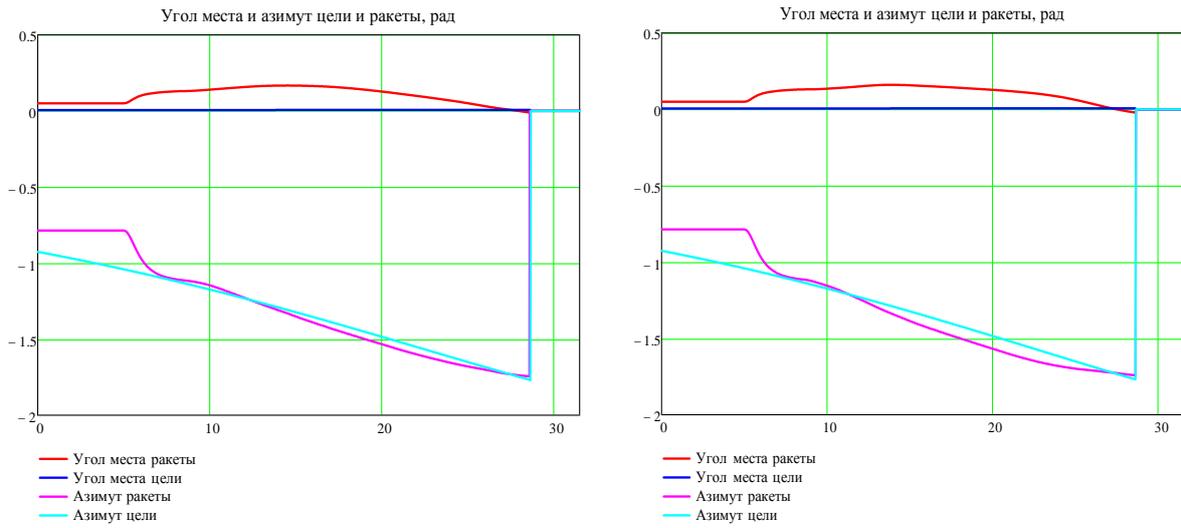


Рис. 5. Сферические координаты ОУ и цели при наведении методом ММПН (а) и МПСТ (б). Траектория 1

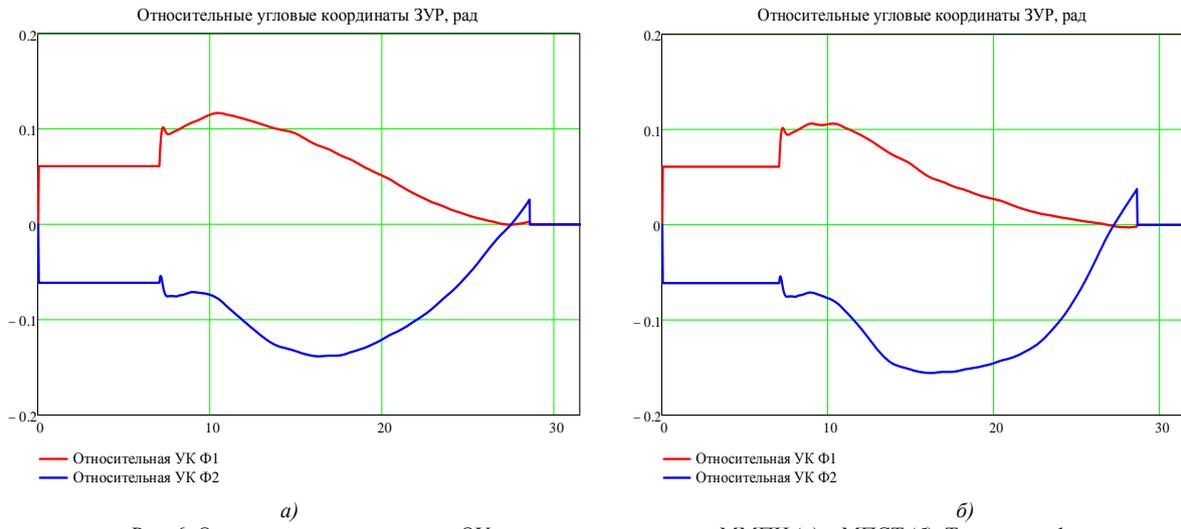


Рис. 6. Относительные координаты ОУ при наведении методом ММПН (а) и МПСТ (б). Траектория 1

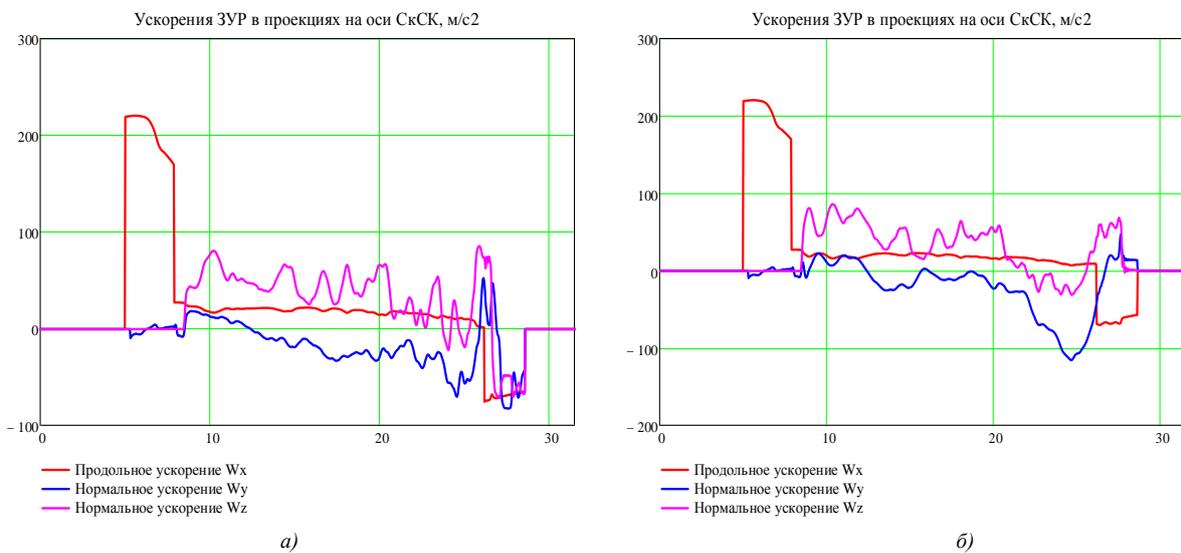
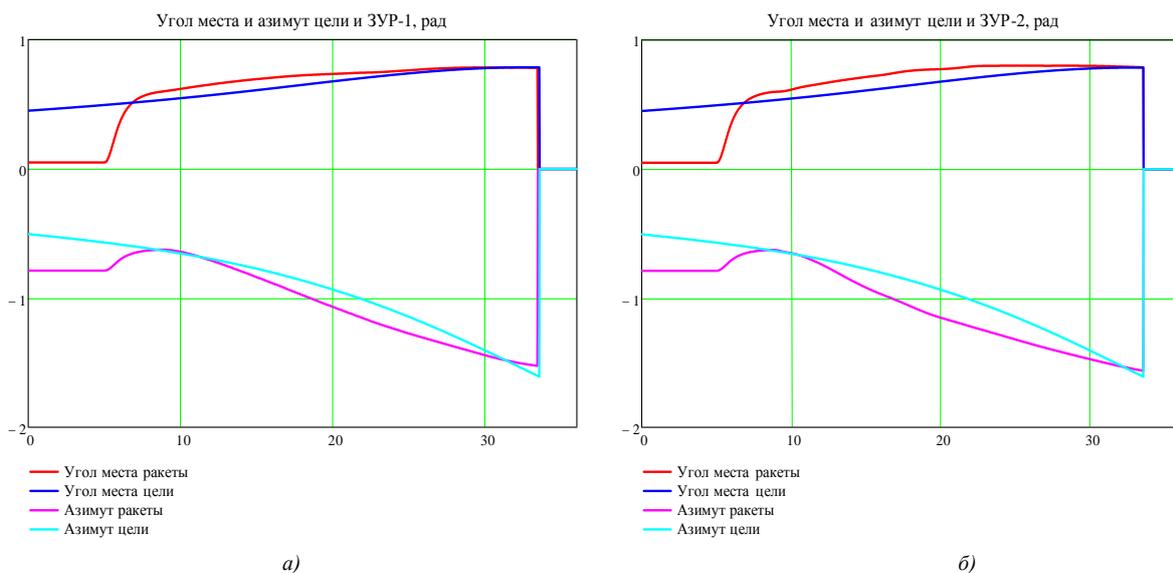
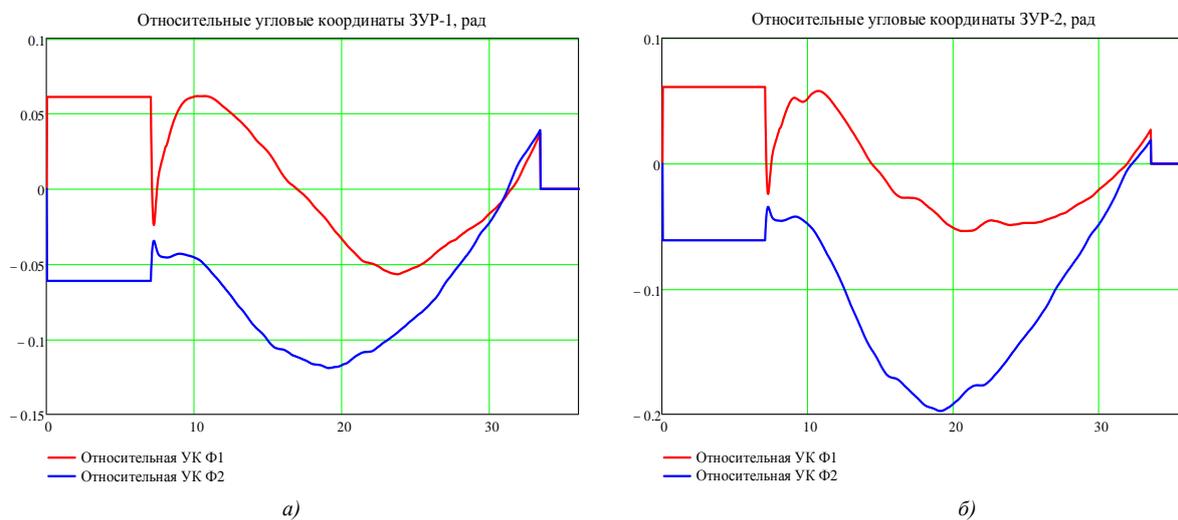
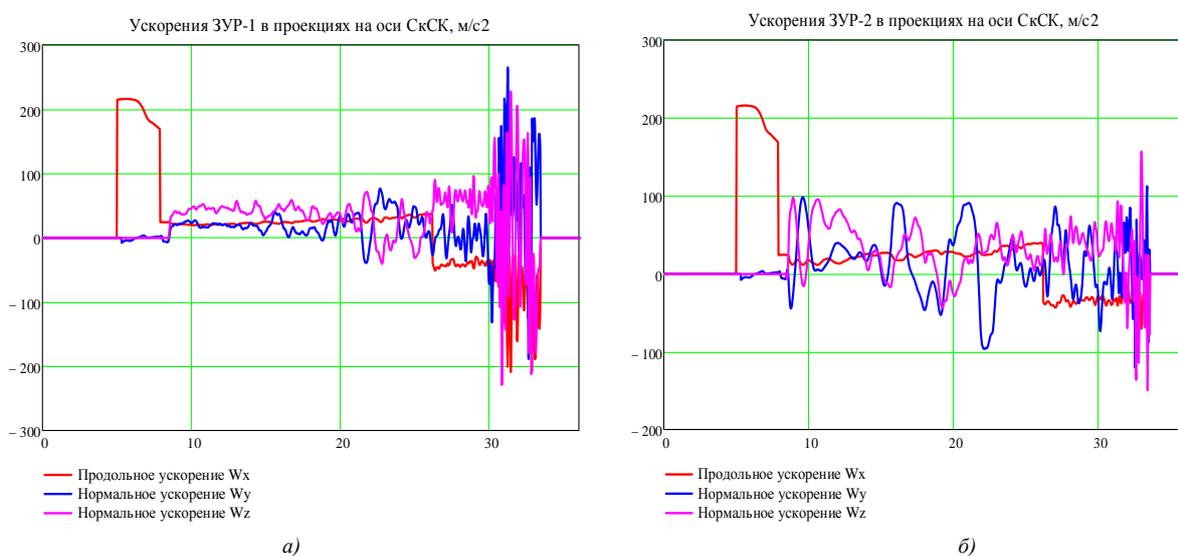


Рис. 7. Ускорения ОУ в СкСК при наведении методом ММПН (а) и МПСТ (б). Траектория 1

Рис. 8. Сферические координаты ОУ и цели при наведении методом ММПН (*a*) и МПСТ (*б*). Траектория 2Рис. 9. Относительные координаты ОУ при наведении методом ММПН (*a*) и МПСТ (*б*). Траектория 2Рис. 10. Ускорения ОУ в СкСК при наведении методом ММПН (*a*) и МПСТ (*б*). Траектория 2

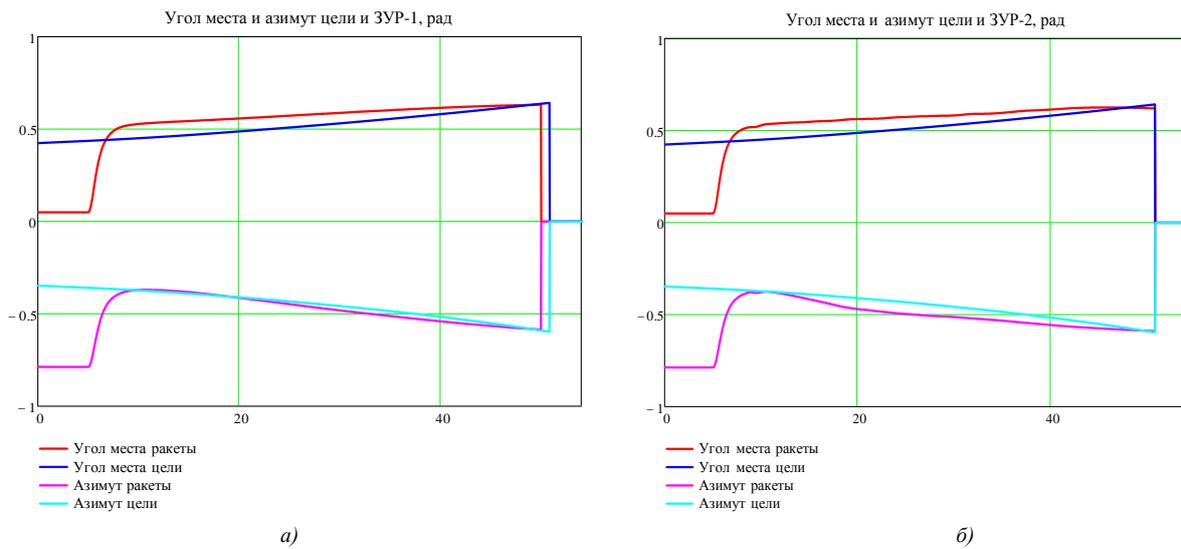


Рис. 11. Сферические координаты ОУ и цели при наведении методом ММПН (а) и МПСТ (б). Траектория 3

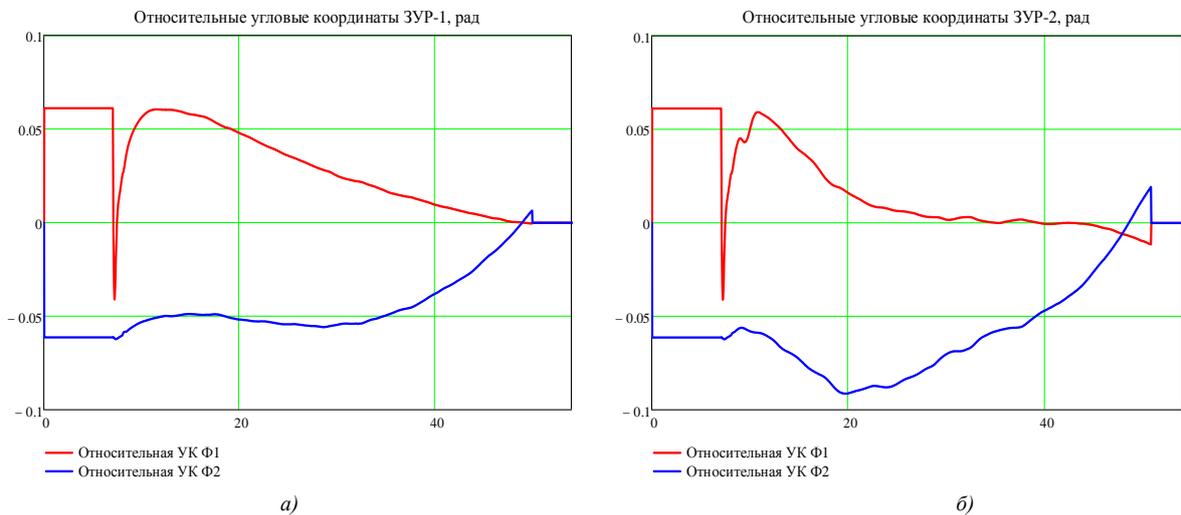


Рис. 12. Относительные координаты ОУ при наведении методом ММПН (а) и МПСТ (б). Траектория 3

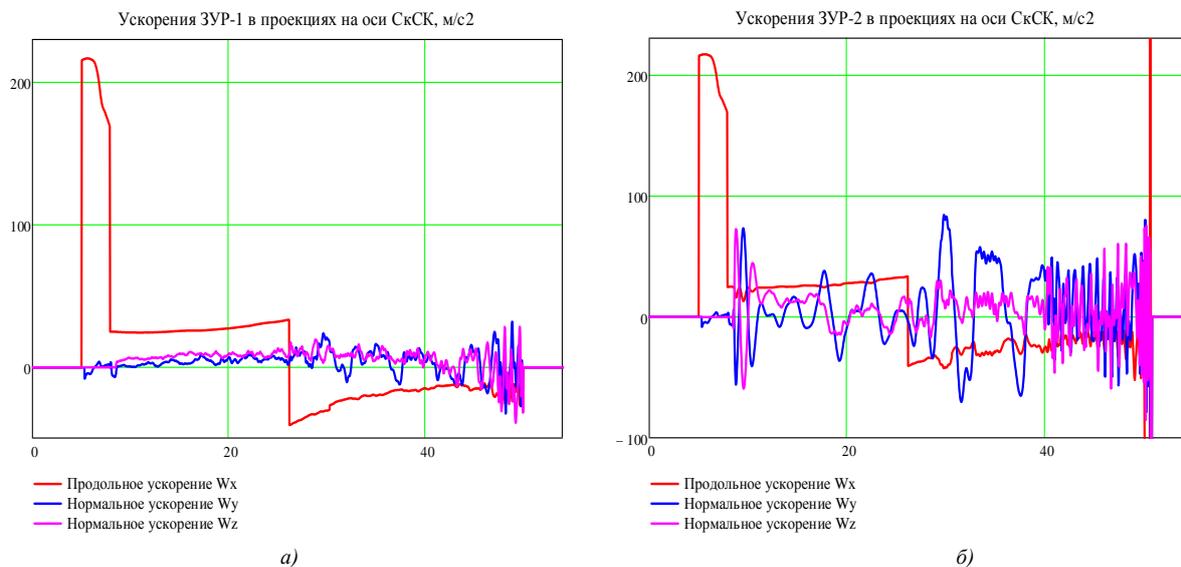


Рис. 13. Ускорения ОУ в СкСК при наведении методом ММПН (а) и МПСТ (б). Траектория 3

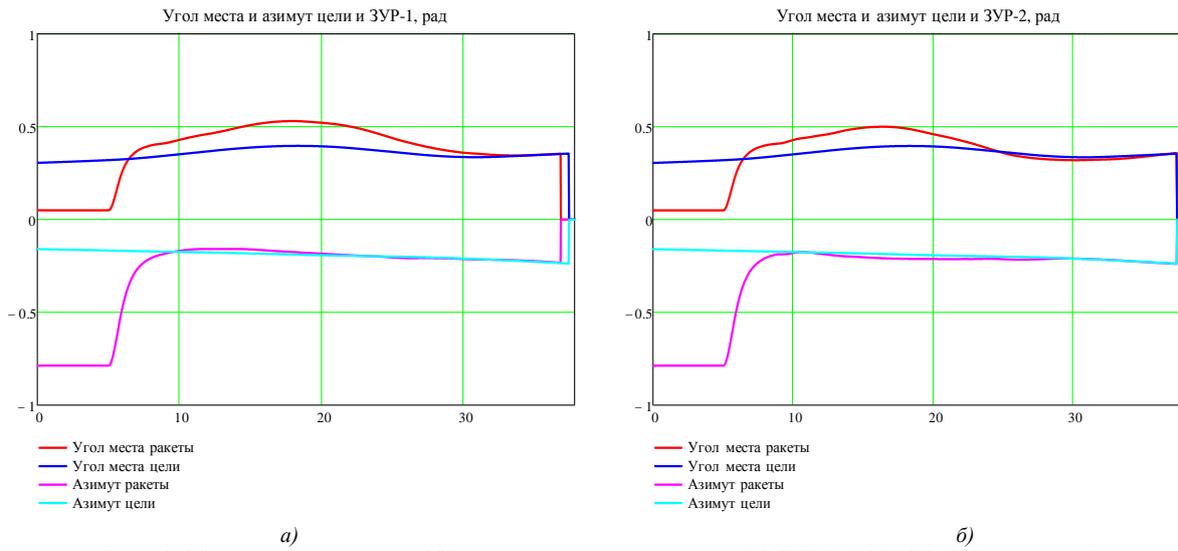


Рис. 14. Сферические координаты ОУ и цели при наведении методом ММПН (а) и МПСТ (б). Траектория 4

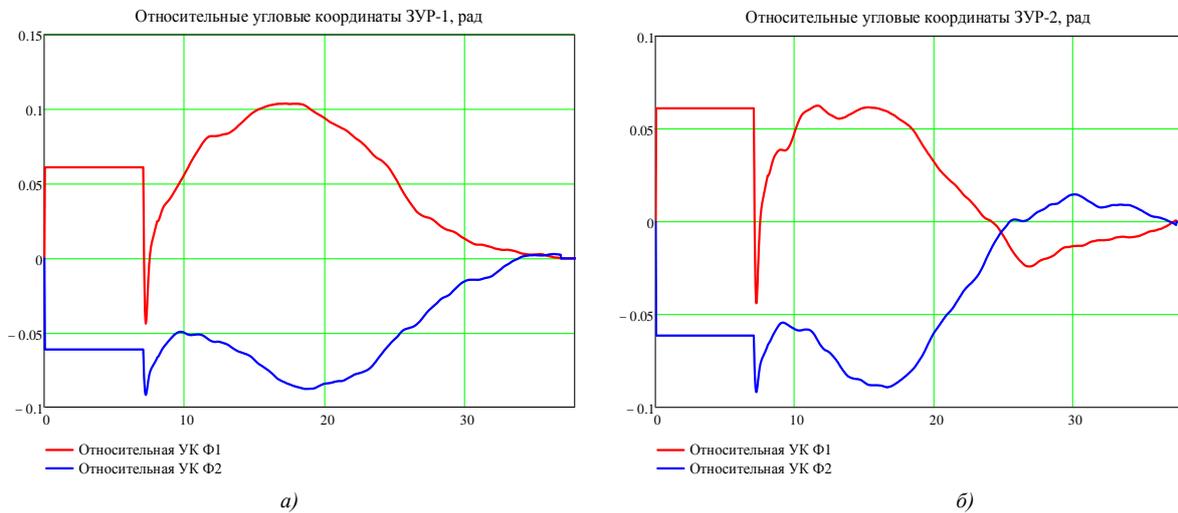


Рис. 15. Относительные координаты ОУ при наведении методом ММПН (а) и МПСТ (б). Траектория 4

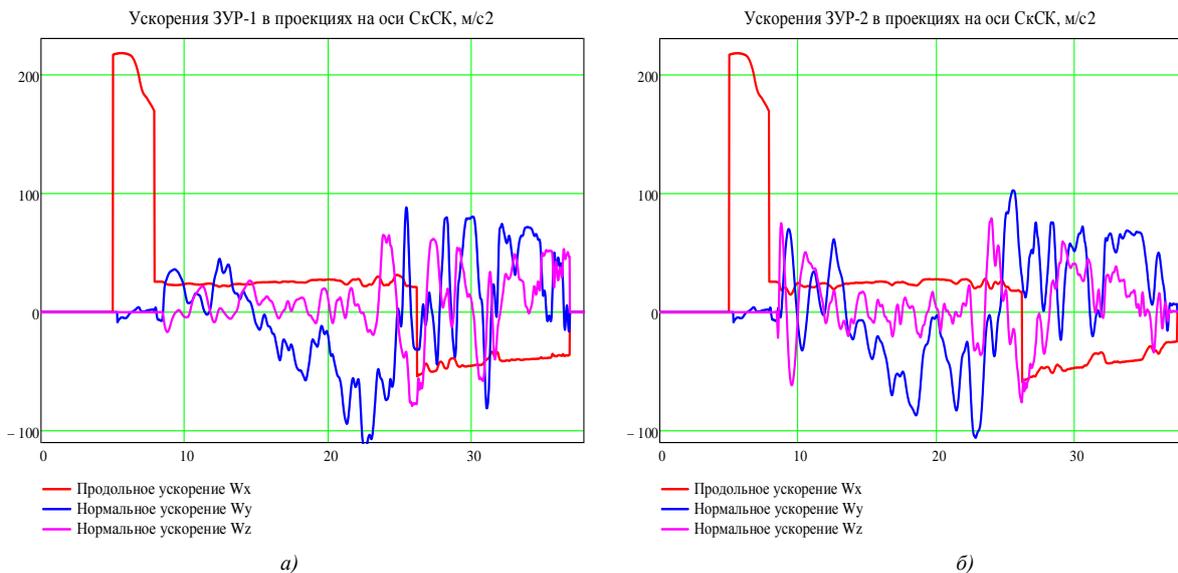


Рис. 16. Ускорения ОУ в СкСк при наведении методом ММПН (а) и МПСТ (б). Траектория 4

Альтернативная форма записи выражения для текущего промаха имеет вид [4, 5]

$$\begin{aligned} \mathbf{h}_\gamma &= \mathbf{h}_0 + \dot{\mathbf{h}}_0 \tau_y, \Delta \mathbf{a}_0 = \frac{\Delta \mathbf{a}}{\|\Delta \mathbf{a}\|}, \Delta \mathbf{v}_0 = \frac{\Delta \mathbf{v}}{\|\Delta \mathbf{v}\|}, \\ \mathbf{h}_0 &= (\Delta \mathbf{r} - \Delta \mathbf{v}_0 (\Delta \mathbf{r} \cdot \Delta \mathbf{v}_0)), \\ \dot{\mathbf{h}}_0 &= -(\Delta \mathbf{a}_0 (\Delta \mathbf{r} \cdot \Delta \mathbf{v}_0) + \Delta \mathbf{v}_0 (\Delta \mathbf{r} \cdot \Delta \mathbf{a}_0)). \end{aligned} \quad (29)$$

Линейная ошибка наведения $\mathbf{h}_\zeta = (0 \ h_y \ h_z)^T$ оценивалась в момент равенства дальностей цели и ОУ согласно следующим выражениям:

$$\begin{aligned} \Phi &= 0,707 \left((\Delta \varepsilon + \Delta \beta \cos(\varepsilon_2)) (\Delta \beta \cos(\varepsilon_2) - \Delta \varepsilon) \right)^T, \\ \mathbf{h}_\zeta (\Delta r = 0) &= r_1 (\Delta \varepsilon \ \Delta \beta)^T, \\ \mathbf{h}_\phi (\Delta r = 0) &= r_1 (\Phi_1 \ \Phi_2)^T, \Delta \varepsilon = \varepsilon_1 - \varepsilon_2, \Delta \beta = \beta_1 - \beta_2, \end{aligned} \quad (30)$$

где Δr – разность радиальных дальностей цели и ОУ; \mathbf{h}_ϕ – двухкомпонентный вектор линейной ошибки наведения в СК, повернутой на 45° относительно продольной оси.

Нормальные ускорения ОУ в момент встречи рассчитывались следующим образом:

$$\begin{aligned} \mathbf{W} &= [\boldsymbol{\omega} \times \mathbf{V}], \boldsymbol{\omega} = \mathbf{M}(\theta, \varphi) \frac{[\mathbf{v}_1 \times \mathbf{a}_1]}{\mathbf{v}_1^T \mathbf{v}_1}, \\ \mathbf{v}_1 = \dot{\mathbf{r}}_1 &= (v_{x1} \ v_{y1} \ v_{z1})^T, \mathbf{a}_1 = \ddot{\mathbf{r}}_1 = (\dot{v}_{x1} \ \dot{v}_{y1} \ \dot{v}_{z1})^T, \end{aligned} \quad (31)$$

где $\mathbf{V} = (V_x \ 0 \ 0)^T$ – вектор скорости ОУ в проекциях на оси СкСК; $\mathbf{M}(\theta, \varphi)$ – матрица связи ЗСК и СкСК, зависящая от скоростных углов тангажа и рыскания θ, φ ОУ.

Сравнительный анализ результатов моделирования, приведенных в табл. 2, позволяет сделать следующие выводы:

- оценки точности наведения ОУ на низколетящую цель для МПСТ и ММПН отличаются незначительно (это связано с отсутствием учета влияния зеркальных отражений от земли), при росте интенсивности флуктуаций по каналу цели предпочтительно использовать ММПН как менее чувствительный к ошибкам измерения на начальном и среднем участках траектории;

- для обстрела скоростной цели, летящей на больших высотах и предельном параметре, целесообразно использовать ММПН, который обеспечивает меньшее значение ошибки наведения при докритических углах упреждения (на рис. 9, б видно, что перехват цели на траектории 2 методом МПСТ в реальной ситуации невозможен, так как ОУ выходит за пределы сектора сопровождения);

- для обстрела цели, находящейся на предельных дальностях и высоте, целесообразно использовать ММПН ввиду меньшего уровня флуктуационных перегрузок ОУ, обусловленных ошибками измерения УК цели;

- при наведении ОУ на маневрирующую цель и при отсутствии канала оценивания нормальных ускорений цели предпочтение следует отдавать МПСТ как менее чувствительному к величине ускорения маневра и его продолжительности;

- для оценки величины нормальных ускорений целесообразно использовать не только траекторную информацию (величину первых и вторых производных угловых координат и дальности), но и сигнальную (ширину спектра флуктуаций отраженного сигнала), которая обрабатывается в устройстве адаптивной фильтрации параметров движения цели.

Заключение. В работе рассмотрен модифицированный МПН, учитывающий ограничения на сектор сопровождения ОУ, несовпадение измерительной и связанной СК, а также влияние неконтролируемых возмущающих воздействий, действующих на ОУ.

Сравнительный анализ существующих и вновь разработанных методов показывает перспективность использования ММПН для перехвата скоростных целей, а также целей, летящих на предельной дальности и высотах.

Снижение величины ошибок наведения при использовании ММПН обусловлено меньшим порядком астатизма контура наведения (структурой кинематического звена) и, как следствие, более медленным ростом интенсивности флуктуаций команд управления (УСВ вектора скорости ОУ) при увеличении СКО ошибок измерений.

Направлениями дальнейших исследований будут являться учет пространственных кинематических связей относительного движения цели и ОУ при формировании команд управления; разработка алгоритмов управления ракетой по крену для повышения точности наведения и вероятности перехвата (поражения) цели; разработка алгоритмов нелинейной оптимальной фильтрации для совместного оценивания координат и параметров движения; реализация эталонной модели ОУ, корректируемой по данным измерителей координат ОУ, для устойчивого сопровождения при больших углах упреждения.

Список использованных источников

1. Кельзон, А. С. Пропорциональная навигация как проблема кибернетики / А. С. Кельзон, О. В. Григорьева // Докл. АН СССР. – 1958. – Т. 121, № 3. – С. 432–435.
2. Кан, В. Л. О точных решениях уравнений пропорциональной навигации / В. Л. Кан, А. С. Кельзон // Изв. ВУЗов. Математика. – 1962. – № 1. – С. 50–56.
3. Пугачев, В. С. Системы управления и динамика полета ракет / В. С. Пугачев, И. Е., Казаков, Д. И. Гладков ; под ред. В. С. Пугачева. – М. : ВВИА им. проф. Н. Е. Жуковского, 1965. – 615 с.
4. Кринецкий, Е. И. Системы самонаведения / Е. И. Кринецкий. – М. : Машиностроение, 1970. – 236 с.
5. Батков, А. М. Проектирование систем наведения / А. М. Батков, А. А. Горский, В. Ф. Левитин ; под ред. Е. А. Федосова. – М. : Машиностроение, 1975. – 296 с.
6. Максимов, М. В. Радиоэлектронные системы самонаведения / М. В. Максимов, Г. И. Горгонов – М. : Радио и связь, 1982. – 304 с.
7. Казаков, И. Е. Авиационные управляемые ракеты : в 2 ч. Ч. 2 : Системы управления и динамика наведения управляемых ракет и бомб / И. Е. Казаков, А. Ф. Мишаков ; под ред. И. Е. Казакова. – М. : ВВИА им. проф. Н. Е. Жуковского, 1985. – 423 с.
8. Теория автоматического управления : учебник для ВУЗов : в 2 ч. Ч. 2 : Теория нелинейных и специальных систем автоматического управления / А. А. Воронов [и др.] ; под ред. А. А. Воронова. – М. : Высш. шк., 1986. – 504 с.
9. Dohi, N. Modified proportional navigation for a missile with varying velocity – comparison with the optimal guidance / N. Dohi, Y. Baba, H. Takano // The Proc. of 24th Intern. Congress of the Aeronautical Sciences (ICAS-2004). – Yokohama, 2004. – Vol. 7. – P. 207–215.
10. Palumbo, N. F. Modern homing missile guidance theory and techniques / N. F. Palumbo, R. A. Blauwkamp, J. M. Lloyd // Johns Hopkins APL Technical Digest. – 2010. – Vol. 29, no. 1. – P. 42–59.
11. Seunghyun, K. Robust proportional navigation guidance against highly maneuvering targets / K. Seunghyun, K. Jin // The Proc. of 13th Intern. Conf. on Control, Automation and Systems (ICCAS-2013). – Seoul, 2013. – P. 173–182.
12. Маркевич, В. Э. Аналитический синтез метода наведения сверхзвукового беспилотного летательного аппарата на основе многомерной нелинейной динамической модели / В. Э. Маркевич // Информатика. – 2017. – № 2(54). – С. 93–112.
13. Колесников, А. А. Проектирование многокритериальных систем управления промышленными объектами / А. А. Колесников, А. Г. Гельфгат. – М. : Энергоатомиздат, 1993. – 304 с.
14. Колесников А. А. Синергетическая теория управления. Инварианты, оптимизация, синтез / А. А. Колесников. – М. : Энергоатомиздат, 1994. – 344 с.
15. Кун, А. А. Основы построения зенитно-ракетных комплексов / А. А. Кун, В. Ф. Лукьянов, С. А. Шабан. – Минск : ВА РБ, 2001. – 327 с.

References

1. Kel'zon, A. S., Grigoryeva O. V. Proportsional'naya navigatsiya kak problema kibernetiki [Proportional navigation as a problem of cybernetics]. Dokl. AN SSSR [Reports of the Academy of Sciences of the USSR], 1958, vol. 121, no. 3, pp. 432–435 (in Russian).
2. Kan V. L., Kel'zon, A. S. O tochnyh resheniyah uravnenij proporcional'noj navigacii [About exact solutions of the equations of proportional navigation]. Izvestiya VUZov. Matematika [News of Universities. Mathematica], 1962, no. 1, pp. 50–56 (in Russian).
3. Pugachev V. S. (ed.), Kazakov I. E., Gladkov D. I. Sistemy upravleniya i dinamika poleta raket. Control Systems and Dynamics of Flight of Rockets. Moscow, VVIA im. prof. N. E. Zhukovskogo Publ., 1965, 615 p. (in Russian).

4. Krineckij E. I. Sistemy samonavedeniya. *Homing Systems*. Moscow, Mashinostroenie Publ., 1970, 236 p. (in Russian).
5. Batkov A. M., Gorsky A. A., Levitin V. F., Fedosov E. A. (ed.) Proektirovanie sistem navedeniya. *Design of Guidance Systems*. Moscow, Mashinostroenie Publ., 1975, 296 p. (in Russian).
6. Maximov M. V. Gorgonov G. I. Radioelektronnye sistemy samonavedeniya. *The Radio-Electronic Systems of Homing*. Moscow, Radio i svyaz Publ., 1982, 304 p. (in Russian).
7. Kazakov I. E., Mishakov A. F. Aviacionnye upravlyaemye rakety. T. 2: Sistemy upravleniya i dinamika navedeniya upravlyaemyh raket i bomb. *Aviation guided missiles. Vol. 2: Control systems and dynamics of guidance of guided missiles and bombs*. Moscow, VVIA im. prof. N. E. Zhukovskogo Publ., 1985, 423 p. (in Russian).
8. Voronov A. A. (ed.) Teoriya avtomaticheskogo upravleniya. T. 2: Teoriya nelinejnyh i special'nyh sistem avtomaticheskogo upravleniya. *Theory of Automatic Control. Vol. 2: Theory of Nonlinear and Special Systems of Automatic Control*. Moscow, Vysshaya shkola Publ., 1986, 504 p. (in Russian).
9. Dohi N., Baba Y., Takano H. Modified proportional navigation for a missile with varying velocity – comparison with the optimal guidance. *The Proceeding of 24th International Congress of the Aeronautical Sciences (ICAS-2004)*. Yokohama, 2004, vol. 7, pp. 207–215.
10. Palumbo N. F., Blauwkamp R. A., Lloyd J. M. Modern homing missile guidance theory and techniques. *Johns Hopkins APL Technical Digest*, 2010, vol. 29, no. 1, pp. 42–59.
11. Markevich V. E. Analiticheskij sintez metoda navedeniya sverhzhukovogo bespilotnogo letatel'nogo apparata na osnove mnogomernoj nelinejnoj dinamicheskoy modeli [Analytical synthesis of a method of guidance of the supersonic unmanned aircraft on the basis of multidimensional nonlinear dynamic model]. *Informatika [Informatics]*, 2017, vol. 2(54), pp. 93–112 (in Russian).
12. Kolesnikov A. A., Gelfgat A. G. Proektirovanie mnogokriterial'nyh sistem upravleniya promyshlennymi ob'ektami. *Design of Multicriterial Control Systems of Industrial Facilities*. Moscow, Energoatomizdat Publ., 1993, 304 p. (in Russian).
13. Kolesnikov A. A. Sinergeticheskaya teoriya upravleniya. Invarianty, optimizaciya, sintez. *Synergetic Theory of Control. Invariants, Optimization, Synthesis*. Moscow, Energoatomizdat Publ., 1994, 344 p. (in Russian).
14. Seunghyun K., Jin K. Robust proportional navigation guidance against highly maneuvering targets. *The Proceeding of 13th International Conference on Control, Automation and Systems (ICCAS-2013)*, Seoul, 2013, pp. 173–182.
15. Kun A. A., Luk'yanov V. F., Shaban S. A. Osnovy postroeniya zenitno-raketnyh kompleksov. *The Basic of Designed Anti-Aircraft Missile System*. Minsk, VA RB Publ., 2001, 327 p. (in Russian).

Информация об авторах

Маркевич Виталий Эдмундович – кандидат технических наук, ведущий научный сотрудник научно-исследовательского сектора ОАО «АЛЕВКУРП» (пр. Независимости, 95/7, 220043, Минск, Республика Беларусь). E-mail: mark.vit@tut.by

Легкоступ Виктор Валерьевич – аспирант БГУИР, младший научный сотрудник научно-исследовательского сектора ОАО «АЛЕВКУРП» (пр. Независимости, 95/7, 220043, Минск, Республика Беларусь). E-mail: legkostupvv@gmail.com

Information about the authors

Vitaliy E. Markevich – Ph. D. (Engineering), Leader Scientist, OJSC “ALEVKURP” (95/7, Nezavisimosti Ave., 220043, Minsk, Republic of Belarus). E-mail: mark.vit@tut.by

Victor V. Legkostup – Post-Graduate Student of BSUIR, Junior Researcher, OJSC “ALEVKURP” (95/7, Nezavisimosti Ave., 220043, Minsk, Republic of Belarus). E-mail: legkostupvv@gmail.com

ISSN 1816-0301 (print)
UDC 621.002

Received 07.05.2018
Поступила в редакцию 07.05.2018

P. Ping, Y. N. Petrenko

Belarusian National Technical University, Minsk, Belarus

FLEXIBLE MULTIPLE ACCESS CONTROL ALGORITHM IN 5G MOBILE WIRELESS SYSTEM

Abstract. With the development of mobile communication technology, wireless spectrum resources have become increasingly scarce. Researchers are eagerly looking for a new mobile communication technology which do not only meet user expectations but also improve wireless spectrum efficiency. The key technologies of the next generation wireless network research concentrates on complex multiple access control algorithms in the physical layer. This paper discusses the concept and practical aspects of non-orthogonal multiple access (NOMA) with a successive interference canceller at the receiver side. Using Matlab simulations, we justify, for multiple configurations, that the system-level performance achieved by NOMA is higher by more than 30 % compared to orthogonal multiple access (OMA) in maximum value.

Keywords: non-orthogonal multiple access, multiple access control algorithms, successive interference canceller

For citation. Ping P., Petrenko Y. N. Flexible multiple access control algorithm in 5G mobile wireless system. *Informatics*, 2018, vol. 15, no. 3, pp. 93–101.

П. Пинг, Ю. Н. Петренко

Белорусский национальный технический университет, Минск, Беларусь

АЛГОРИТМ ГИБКОГО УПРАВЛЕНИЯ МНОЖЕСТВЕННЫМ ДОСТУПОМ В МОБИЛЬНОЙ БЕСПРОВОДНОЙ СИСТЕМЕ 5G

Аннотация. С развитием технологий мобильной связи ресурсов беспроводного спектра становится недостаточно. Исследователи ищут новые технологии мобильной связи, которые будут не только отвечать ожиданиям пользователей, но и улучшат эффективность беспроводного спектра. Ключевые технологии исследования беспроводной сети следующего поколения основаны на сложных алгоритмах управления множественным доступом на физическом уровне. В статье предлагается новый алгоритм гибкого множественного доступа (неортогонального множественного доступа, НОМД), анализируется производительность системы для НОМД восходящей линии связи с усовершенствованным приемником последовательного подавления помех, применяемым на стороне базовой станции. С помощью Matlab показывается, что производительность системы, достигнутая в НОМД, выше на 30 % по сравнению с системой ортогонального множественного доступа.

Ключевые слова: неортогональный множественный доступ, алгоритм управления множественным доступом, фильтр сигналов

Для цитирования. Пинг, П. Алгоритм гибкого управления множественным доступом в мобильной беспроводной системе 5G / П. Пинг, Ю. Н. Петренко // Информатика. – 2018. – Т. 15, № 3. – С. 93–101.

1. Introduction of algorithm of orthogonal multiple access. In the field of wireless terminal application, the design of radio access technology is an important aspect to enhance system capacity in a cost-effective manner. Radio access technologies are typically characterized by multiple access schemes, e.g., orthogonal frequency division multiple access (OFDMA), which provide the means for multiple users to access and share system resources simultaneously. OMA is a reasonable choice for achieving good system level throughput performance in packet-domain services with a simplified receiver design. The receiver does not need to use a filter to distinguish subcarriers in OFDMA algorithm [1–3]. In fig. 1 the binary sequence signal processing is converted to frequency band signal by quadrature amplitude modulation (QAM) combine to fast Fourier transformation (FFT) and imaginary exponential function ($e^{j2\pi ft}$) to get OFDMA symbol.

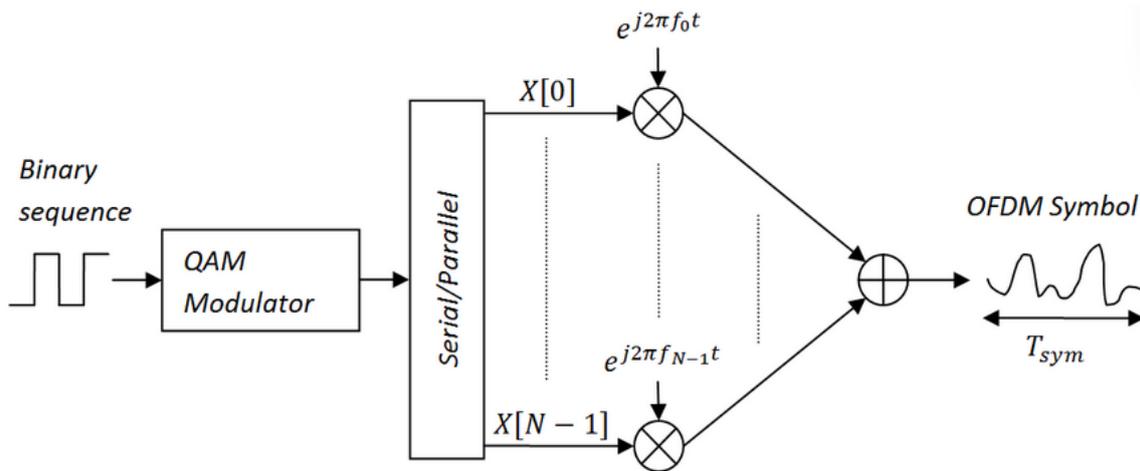


Fig. 1. The model of modulation in OFDMA

However, in order to boost further the spectrum efficiency in the future, more advanced multiple access algorithms are required in order to mitigate intra-cell and/or inter-cell interference. As a candidate multiple access scheme for frequency radio access, we propose an uplink NOMA scheme where multiple users are multiplexed in the power-domain on the transmitter side, and multi-user signal separation on the receiver side is conducted based on successive interference cancellation.

2. Concept of NOMA algorithm. In this section, we focus on the principle of NOMA and OFDMA technology as air interface technology for current and future mobile wireless broadband access systems. The next section provides the concept considerations of NOMA with a successive interference canceller at the receiver side. In the remaining part of the article, we present performance analysis of NOMA algorithm model system, successive interference canceller capacity analysis for uplink and we propose multiuser scheduling scheme to defuse these challenges.

NOMA is the result of an evolution of digital transmission techniques and advances in the digital signal processing methods and technologies. The basic idea of NOMA is to actively introduce interference technology in transmission at the transmitter. A receiver with successive interference canceller can achieve the correct demodulation. The complexity of the system has been improved by using the successive interference canceller algorithm, however, it can also improve the spectral efficiency. The essence of NOMA algorithm is in increasing the spectral efficiency by increasing the complexity of the receiver [4].

The comparison of multiple access algorithms

Terminal multiple access algorithm	Non-orthogonal code division multiple access	Orthogonal OFDMA	Non-orthogonal successive interference canceller
The waveform of signals	Single-carrier	OFDM (DFT-s-OFDM)	OFDM (DFT-s-OFDM)
Link adaptability	High-speed transfer protocol control	Adaptive modulation coding	Adaptive modulation coding + power allocation
Graphical	<p>CDMA, WCDMA</p>	<p>OFDMA LTE, WIMAX</p>	<p>NOMA</p>

OFDMA algorithm is still used in the sub-channel transmission of NOMA algorithm (table). The sub-channels are orthogonalities which do not interfere with each other [5]. The resources which

according to discrete Fourier transform orthogonal multiplexing on a sub-channel can be shared by multiple users. The transmission is non-orthogonally on the same sub-channel between different devices. At this point, the signal of devices will have interference between each other [6]. The receiver has to use successive interference canceller technology for multi-user detection purposes due to this reason. At the transmitter, different devices on the same sub-channel use power multiplexing algorithm in transmission. The power resource of a signal is allocated according to NOMA algorithm for different devices [7]. Therefore, the signal power of each device is different at the receiver. The successive interference canceller receiver performs interference elimination based on signal power algorithm (power size in a certain order) to achieve the purpose of different devices [8].

2.1. NOMA algorithm based serial interference cancellation system. The serial interference cancellation (SIC) receiver has a great improvement in performance compared with the conventional detectors. The modified structure of SIC system is easy to achieve. The basic principle of serial interference cancellation is to gradually subtract the maximum signal power from the user interference in the receiving terminal. The received signal will be uniquely recovered from SIC detector. The multiple access interference caused by the user signal is subtracted when a user signal is determined. This keeps circulating until all the multiple access interference is eliminated as shown in fig. 2. The out of band noise is eliminated maximally when a mixed signal $y(t)$ passing through the band pass filter. Then this mixed signal enter into the verdict. The order of single judgment operation according to its power value (high power signal sequence first, because of the highest power is easiest to capture). The output value of each time cycle is the signal decision of the maximum power user and subtract the multiple access interference caused by this user. In this way, multiple access interference can be reduced to a minimum (the weaker the signal is, the greater the benefit will be) and the reliability of detection will be greatly increased. The multi time cycle structure takes the output signal of the previous cycle as the input signal of the next cycle to repeat the operation in detection, amplitude estimation, detection. T_b is a timer which strictly controls the system clock synchronization of each cycle detection signal.

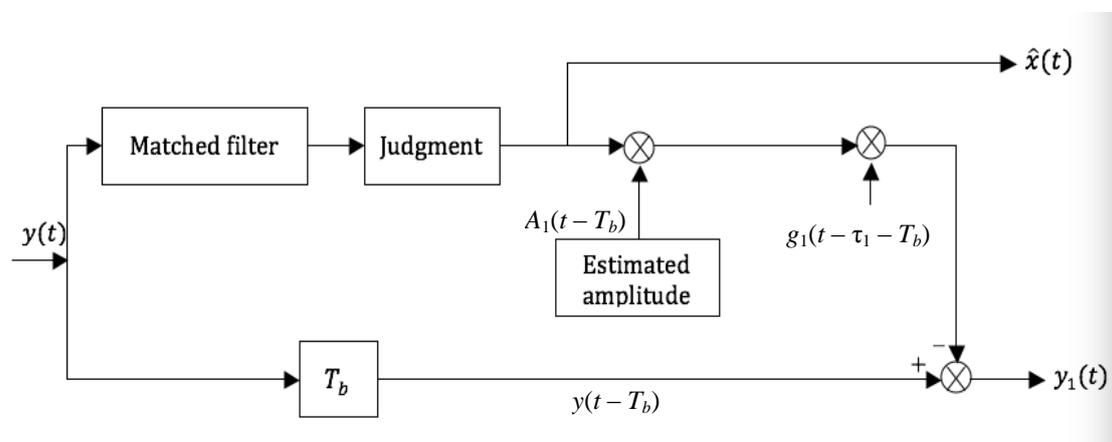


Fig. 2. The structure of SIC detector

NOMA is a multiplexing scheme that utilizes an additional new domain, i.e., the power domain, which is not sufficiently utilized in previous systems. The NOMA in power domain is different from power control in 4G wireless system [8]. Different transmit power is allocated to different transmitter device according to the base station NOMA algorithm on one side; on the other side the SIC receiver could distinguish signals from devices in different power [8].

2.2. Performance analysis of NOMA algorithm model system. In order to analyze the performance of the NOMA system, we take a downlink as an example. There has a simplified model for the NOMA downlink as fig. 3. The transmitter and the receiver adopt a mode which in single illuminator – multiple receiver [9, 10].

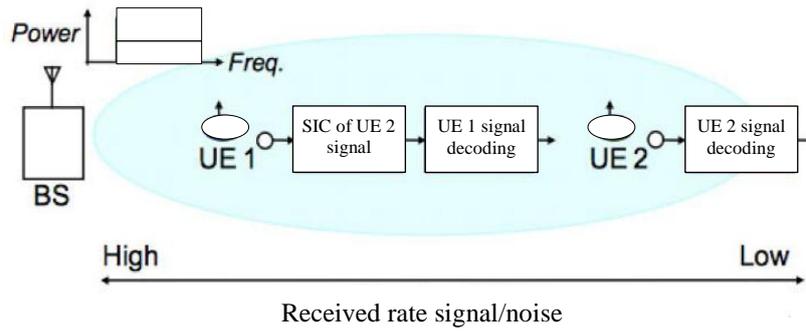


Fig. 3. Simplified model for the NOMA downlink

The signal is UE I ($I = 1, 2$), the channel bandwidth is 1 Hz, and the device signal is $E[|X_i|^2]=1$ in this model. The signal after encoding in superposition is: $X = \sqrt{P_1}X_1 + \sqrt{P_2}X_2$. Among them, P_i is the power of the UE I ($I = 1, 2$), the total power of signal from devices as in (1)

$$\sum_{i=1}^2 P_i = P. \tag{1}$$

In this fig. 4 the UE with high channel gain (UE 1) is allocated less power and the UE with low channel gain (UE 2) is allocated more power ($P_1 < P_2$). Such large power difference facilitates the successful decoding (with high probability) and thus the successful cancellation of the signal designated to UE 2 (being allocated high power) at UE 1 receiver.

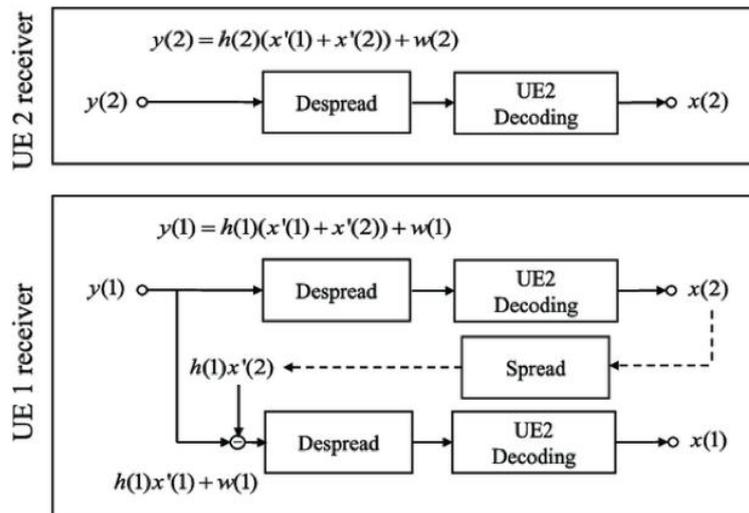


Fig. 4. The process of signal detection by user UE I

The example on a sub-channel shared by 2 users which as mentioned above, the superimposed signal is could be expressed as

$$x = x(1) + x(2). \tag{2}$$

Among the formula (2), $x(i)$, $i = 1, 2$ represents the signals which from 2 users respectively. In addition, at UE 2 receiver, the signal designated to UE 2 is decoded directly by treating the interference from the signal designated to UE 1 (being allocated low power) as noise. As formula (3)

$$y_i = h_i X + w_i \tag{3}$$

$y(i)$ is the input mixed signal, where $h(i)$ is the complex channel coefficient between user terminal (UE- i) and the base station (BS). Term $w(i)$ denotes additive white Gaussian noise (AWGN) including inter-cell interference. The power spectral density of $w(i)$ is $N_{0,i}$. In downlink NOMA, the SIC process is implemented at the UE receiver. The optimal order for decoding is in the order of decreasing channel gain normalized by noise and inter-cell interference power, $|h_i/2/N_{0,i}$ (called simply channel gain in the remaining). Based on this order, we assume that any user can correctly decode the signals of other users whose decoding order comes before the corresponding user. Thus, UE- i can remove the inter-user interference from the j -th user whose $|h_j/2/N_{0,j}$ is lower than $|h_i/2/N_{0,i}$. In this 2-UE case, assuming that $|h_1/2/N_{0,1}| > |h_2/2/N_{0,2}|$, UE 2 does not perform interference cancellation since it comes first in the decoding order. UE 1 first decodes x_2 and subtracts its component from received signal y_1 , and then next, it decodes x_1 without interference from x_1 . Assuming successful decoding and no error propagation, the throughput of UE- i , R_i , is represented as formula (4)

$$R_1 = \log_2 \left(1 + \frac{P_1|h_1|^2}{N_{0,1}} \right), R_2 = \log_2 \left(1 + \frac{P_2|h_2|^2}{N_{0,2} + P_1|h_1|^2} \right). \quad (4)$$

It can be seen that by adjusting the power allocation ratio, P_1/P_2 , the base station can flexibly control the throughput of each UE. Clearly, the overall cell throughput, cell-edge throughput, and user fairness are closely related to the power allocation scheme adopted.

2.3. Serial interference cancellation channel capacity analysis for uplink. An additive white Gaussian noise channel model as an example to analyze in this section. The uplink access model of two users is represented as formula (5):

$$y = x_1 + x_2 + n, \quad (5)$$

where n is Gaussian white noise, user power is p_i . According to the AWGN channel capacity theory which propose by Shannon could infer that the information transmission rate R_1 , R_2 (two different terminal devices) satisfying the following conditions as:

$$\begin{cases} R_1 < \log \left(1 + \frac{P_1}{N_0} \right); \\ R_2 < \log \left(1 + \frac{P_2}{N_0} \right); \\ R_1 + R_2 < \log \left(1 + \frac{P_1 + P_2}{N_0} \right), \end{cases} \quad (6)$$

$R_1 < \log \left(1 + \frac{P_1}{N_0} \right)$ and $R_2 < \log \left(1 + \frac{P_2}{N_0} \right)$ in formula (6) indicate that the signal transmission rate of 2 users can not exceed point to point channel capacity by power p_i respectively. $R_1 + R_2 < \log \left(1 + \frac{P_1 + P_2}{N_0} \right)$ in formula indicate that the total channel capacity of the system can not exceed the capacity of the point to point transmitted by power $P_1 + P_2$. When user 1 reaches the maximum channel transmission rate, the user 2 can be expressed as:

$$R_2 = (R_1 + R_2) - R_1 = \log \left(1 + \frac{P_1 + P_2}{N_0} \right) - \log \left(1 + \frac{P_1}{N_0} \right) = \log \left(1 + \frac{P_2}{P_1 + N_0} \right). \quad (7)$$

The formula (7) can be expressed in a mathematical coordinate axis as follow.

The signal from user 2 is first detected at the receiving terminal in practical applications (one of them includes the multiple access interference and noise from the user 1). The channel capacity of user 2 is $\log \left(1 + \frac{P_2}{P_1 + N_0} \right)$ as shown in fig. 5. In this moment, the value of user 2 is need to subtract from the received signal after the successful decoding.

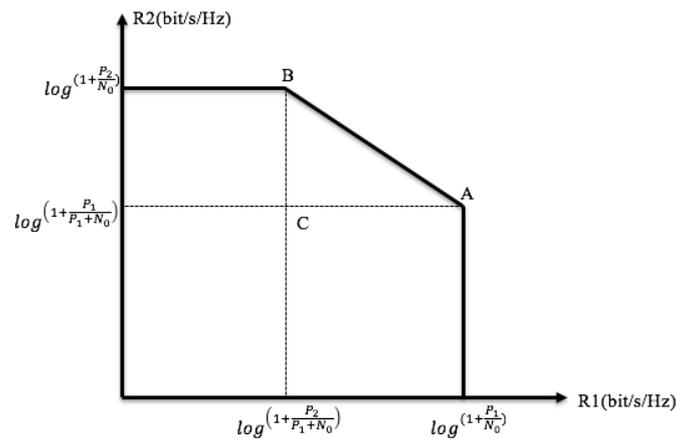


Fig. 5. Analysis of the capacity in an AWGN uplink channel from two users

Then the user 1 is detected at the receiving terminal. At this time, in virtue of the detected only contain signal (user 1) and noise, thus the maximum reachable rate of user 1 is $\log^{(1+\frac{P_1}{N_0})}$ (the channel capacity characteristics at A points). It is equally proved that at any point on line segment AB satisfies the maximum wireless channel capacity. At this time, the fairness of the user channel must be considered when the value of power is much difference between user 1 and user 2 at present (we suppose that the signal power of the user 2 is much greater than the power of user 1). The point A is more likely to be fair in fig. 6. Therefore, the A point is the best serial interference cancellation channel capacity for uplink [11].

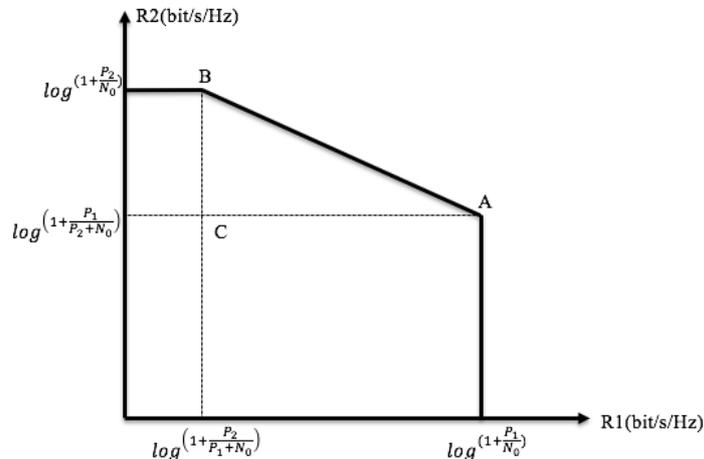


Fig. 6. Analysis of channel capacity in difference power from two users

2.4. Proposed multiuser scheduling scheme. In this section, the proportional fairness (PF) based multiuser scheduling scheme proposed for NOMA uplink is described. First, we present the PF scheduling metric of a set of NOMA multiplexed users [12]. Then, we introduce the sub-band assignment method for NOMA users with the constraint of contiguous resource allocation. In NOMA with a SIC [13], more than one user can be simultaneously scheduled using the same sub-band. The scheduling candidates' S should contain all the sets of users in a cell. In order to reduce the computation of scheduling, all the possible combinations of users in a cell are firstly exhaustively searched at the beginning of scheduling [13]. In this procedure, users within a cell are multiplexed as user set S , $1-|S| - N_{max}$. Each set S includes one or more users but no more than N_{max} users. Therefore, in a cell with K users in total, the number of scheduling user candidate sets, N_{us} as in

formula (8). All the users are uniformly distributed in the cells and the number of user per cell is assumed to K

$$N_{us} = \binom{K}{1} + \binom{K}{2} + \dots + \binom{K}{N_{max}}. \quad (8)$$

The system performance, such as average user throughput and cell edge user throughput, is greatly associated with the scheduling policy [12]. The PF scheduler has been proved as an effective method to achieve good tradeoff between capacity and fairness by maximizing the PF metric as scheduling objective function [14]. The average user throughput at time instance $t + 1$ of user i is expressed as in formula (9)

$$T_i(t + 1) = \left(1 - \frac{1}{t_c}\right) \times T_i(t) + \frac{1}{t_c} \sum_b R_b(i, t), \quad (9)$$

where t_c is the time index representing a sub-frame index, t_c is the length of time window for throughput averaging, $R_b(i_b(l)|s_b)$ is the throughput of user $i_b(l)$, s_b is a neighbor sub-band, t is at time when data comes in a sub-band user, $i_b(l)$ is the channel gain user, i_b could be defined as $g_b(i_b(l))$, including the large scale channel gain, distance-dependent loss and shadowing loss [14]. According to the Shannon first law known W is the bandwidth, $R_b(i_b(l)|s_b)$ is the wireless channel capacity in NOMA, $h_b(i_b(l))$ is the N_r -dimensional channel coefficient vector of the link between user $i_b(l)$ and the serving base station at sub-band b , which consists of path-loss, shadow fading, and small scale fading coefficients, $p_b(i_b(j))$ denotes the transmission power of user $i_b(l)$ at sub-band b , $n_b(i_b(l))$ is the noisy at sub-band b . In a scheduled user set S_b at sub-band b is expressed as in formula (10)

$$R_b(i_b(l)|s_b) = W \log_2 \left(1 + \frac{\|h_b(i_b(l))\|^2 p_b(i_b(l))}{\sum_{j \in S_b, g_b(j_b(l)) > g_b(i_b(l))} |h_b^H(i_b(l)) * h_b(i_b(l))|^2 p_b(i_b(j)) + n_b(i_b(l))} \right). \quad (10)$$

At a sub-band b , the scheduler calculates the scheduling metric of each user set to be scheduled as in formula (11)

$$f_b(S_b) = \prod_{i \in S_b} \left(1 + \frac{R_b(i|S_b)}{(t_c - 1) T_i(t)} \right), \quad (11)$$

where $f_b(S_b)$ is the product of the average user throughput among users in a user set S_b . The scheduler selects the user set maximizing the scheduling metric of each sub band [15].

According to these formulas could illustrate the cumulative distribution function probability curves [16] of the user throughput with maximum number of multiplexed user ranging from 1 to 3 as shown in fig. 7, SIC as $N_{max} > 1$ in the NOMA.

Using the proposed PF-based scheduling scheme can achieve better throughput compared to OFDMA for the most region of the cumulative distribution function curve. Because of the user throughput in OFDMA is limited by the orthogonal bandwidth allocation, which results in reduction of the bandwidth for the respective users. On the contrary, NOMA with a SIC can allow all users to utilize the overall bandwidth, irrespective of the channel conditions. However, in the cell-edge throughput region, NOMA achieves worst performance, due to the increased inter-cell interference. Furthermore, the degradation of the cell-edge performance becomes severe as the maximum number of multiplexed users $N_{max} = 3$. The result is the total transmission power per-cell in the NOMA with a SIC is larger than that of OFDMA.

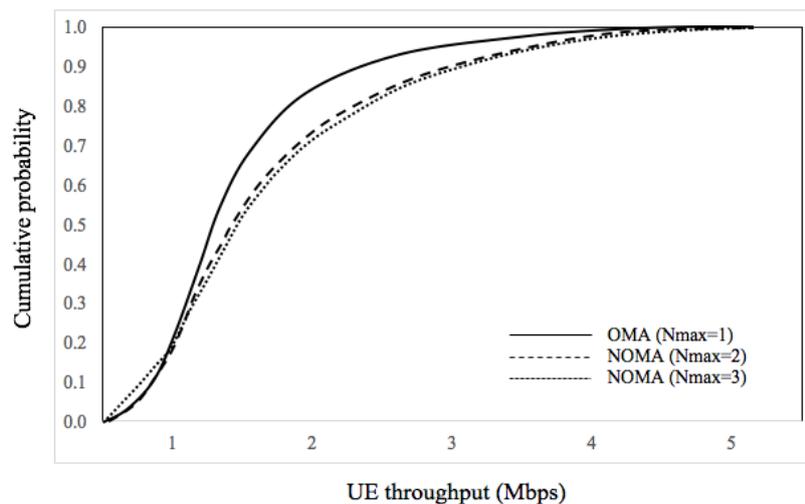


Fig. 7. Cumulative distribution function of user throughput of non-orthogonal access with different maximum number of non-orthogonal multiplexed users

Conclusion. The specific technical standard of 5G has not been formulated yet at present, spectrum efficiency is a key developing direction for wireless communication system from the research results which released by major international organizations. Further advantages of NOMA are the very efficient spectrum usage and, with digital signal processing being cost-effective and flexible, also low-complexity application of the Massive MIMO principle [17]. From this point of view, the NOMA algorithm can satisfy both the demand rate and the spectral efficiency of mobile services.

References

1. 3GPP TS36.300. *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN)*. Available at: https://www.etsi.org/deliver/etsi_ts/136300_136399/136300/09.04.00_60/ts_136300v090400p.pdf (accessed 13.05.2018).
2. 3GPP TR36.913. *3GPP; TSG RAN; Requirements for further advancements for E-UTRA (LTE-Advanced)*. Available at: https://www.etsi.org/deliver/etsi_tr/136900_136999/136913/09.00.00_60/tr_136913v090000p.pdf (accessed 14.05.2018).
3. 3GPP TR36.814. *Further advancements for E-UTRA physical layer aspects*. Available at: <http://www.qtc.jp/3GPP/Specs/36814-900.pdf> (accessed 03.06.2018).
4. Eisuke F., Akishige N., Mamoru H. OFDM Wireless Transmission Technique for System Beyond IMT2000. Available at: <https://www.fujitsu.com/downloads/MAG/vol38-2/paper15.pdf> (accessed 05.06.2018).
5. Yuan Q. LTE easily to progress. *Publishing House of Electronics Industry*. Beijing, China, 2014, vol. 1, pp. 21–71.
6. Tong X. J., Luo T. Principles and applications of OFDM mobile communication. *Post and Telecommunication Press*. Beijing, 2009, vol. 1, pp. 56–81.
7. Beena R. B., Ankit C., Neha S. Orthogonal frequency division multiplexing and its applications. *International Journal of Science and Research*, 2013, vol. 2, pp. 325–328.
8. Takeda T., Higuchi K. Enhanced user fairness using nonorthogonal access with SIC in cellular uplink. *Vehicular Technology Conference (VTC fall)*. San Francisco, 2011, vol. 20, pp. 1–5.
9. Benjebbour A., Anxin Li, Saito Y. Y., Kishiyama Y., Harada A., Nakamura T. System-level performance of downlink NOMA for future LTE enhancements. *Globecom 2013 Workshop-Emerging Technologies for LTE-Advanced and Beyond-4G*. Atlanta, 2014, vol. 1, pp. 136–195.
10. Saito Y., Kishiyama Y., Benjebbour A., Nakamura T., Li A., Higuchi K. Non-orthogonal multiple access (NOMA) for future radio access. *IEEE 77th Vehicular Technology Conference*. Dresden, 2013, pp. 205–221.
11. Otao N., Kishiyama Y., Higuchi K. Performance of non-orthogonal access with SIC in cellular downlink using proportional fair-based resource allocation. *International Symposium on Wireless Communication Systems (ISWCS 2012)*. Paris, 2012, pp. 476–480.
12. Higuchi K., Kishiyama Y. Non-orthogonal access with successive interference cancellation for future radio access. *The IEEE VTS Asia Pacific Wireless Communications Symposium*. Kyoto, 2012, pp. 211–246.
13. Umehara J., Kishiyama Y., Higuchi K. Enhancing user fairness in non-orthogonal access with successive interference cancellation for cellular downlink. *International Conference on Computational Science*. New Orleans, 2012, pp. 343–367.

14. Takeda T., Higuchi K. Enhanced user fairness using non-orthogonal access with SIC in cellular uplink. *IEEE 74th Vehicular Technology Conference*. San Francisco, 2011, vol. 5, pp. 68–82.
15. Endo Y., Kishiyama Y., Higuchi K. Uplink non-orthogonal access with MMSE-SIC in the presence of inter-cell interference. *11th International Semantic Web Conference*. Boston, 2012, pp. 261–265.
16. Jason L., Szymon R., Ravi R., Princeton U. Adaptive numerical cumulative distribution functions for efficient importance sampling. *Eurographics Symposium on Rendering*. Konstanz, Germany, 2005, pp. 11–20.
17. Ping P., Petrenko Y. N. The physical layer technologies in 5g mobile communication system. *International Conference on Physics, Computing and Mathematical Modeling (PCMM)*. Shanghai, China, 2018, pp. 111–117.

Information about the authors

Pei Ping – Post Graduate student at the Department of Information Systems and Technologies, Belarusian National Technical University (25/3, building 20, Nezavisimosty Ave., 220114, Minsk, Republic of Belarus). E-mail: peiping123456.love@gmail.com

Yury N. Petrenko – Ph. D. (Engineering), Associate Professor at the Department of Information Systems and Technologies, Belarusian National Technical University (25/3, building 20, Nezavisimosty Ave., 220114, Minsk, Republic of Belarus.). E-mail: ypetrenko@bntu.by

Информация об авторах

Пей Пинг – аспирант кафедры «Информационные системы и технологии», Белорусский национальный технический университет (пр. Независимости, 25/3, к. 20, 220114, Минск, Республика Беларусь). E-mail: peiping123456.love@gmail.com

Петренко Юрий Николаевич – кандидат технических наук, доцент кафедры «Информационные системы и технологии», Белорусский национальный технический университет (пр. Независимости, 25/3, к. 20, 220114, Минск, Республика Беларусь). E-mail: ypetrenko@bntu.by

ISSN 1816-0301 (print)

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ**MATHEMATICAL MODELING**

УДК 517.958:537.8

Поступила в редакцию 04.05.2018

Received 04.05.2018

В. Т. Ерофеев¹, А. И. Урбанович²¹*Учреждение БГУ «НИИ прикладных проблем математики и информатики», Минск, Беларусь*²*Белорусский государственный университет, Минск, Беларусь***МАТЕМАТИЧЕСКАЯ МОДЕЛЬ РАСПРОСТРАНЕНИЯ
ЭЛЕКТРОМАГНИТНЫХ ВОЛН В КОМПОЗИТНЫХ СРЕДАХ
СО СФЕРОИДАЛЬНЫМИ ЧАСТИЦАМИ**

Аннотация. Разработана математическая модель, которая описывает распространение монохроматических электромагнитных волн в среде с пространственной дисперсией, содержащей вытянутые вдоль заданного направления сфероидальные частицы. Исходная классическая интегро-дифференциальная модель для электромагнитных полей в среде с пространственной дисперсией преобразована с точностью до величин третьего порядка малости в дифференциальную модель, в которой интегро-дифференциальные уравнения Максвелла представлены в виде системы дифференциальных уравнений второго порядка. При этом электрическая и магнитная поляризации среды представлены через операторы Лапласа. Система уравнений решена аналитически, и построена полная система четырех прямых и четырех обратных, распространяющихся в противоположных направлениях электромагнитных волн. Аналитическое представление полей содержит вектор, определяющий направление распространения плоских волн. Волновые числа полей также зависят от направления их распространения, что указывает на анизотропный характер разработанной математической модели.

Ключевые слова: математические модели, интегро-дифференциальная модель, электромагнитные монохроматические волны, плоские поля, сфероидальные частицы, анизотропные среды, пространственная дисперсия, аналитическое моделирование

Для цитирования. Ерофеев, В. Т. Математическая модель распространения электромагнитных волн в композитных средах со сфероидальными частицами / В. Т. Ерофеев, А. И. Урбанович // Информатика. – 2018. – Т. 15, № 3. – С. 102–112.

V. T. Erofeenko¹, A. I. Urbanovich²¹*Institution of the Belarusian State University**“Research Institute for Applied Problems of Mathematics and Informatics”, Minsk, Belarus*²*Belarusian State University, Minsk, Belarus***MATHEMATICAL MODEL OF PROPAGATION OF ELECTROMAGNETIC WAVES
IN COMPOSITE MEDIA WITH SPHEROIDAL PARTICLES**

Abstract. A mathematical model describing the propagation of monochromatic electromagnetic waves in a medium with spatial dispersion containing spheroidal particles of the along prescribed direction has been developed. The initial classical integro-differential model for electromagnetic fields in a medium with spatial dispersion is transformed, within the third-order infinitesimal, to the differential model, where the integro-differential Maxwell equations are represented as a system of second-order differential equations. In this case electrical and magnetic polarizations of the medium are given in the Laplace operators. This system of equations is analytically solved; a complete system of four forward and four backward counter-propagating electromagnetic waves is formed. The analytical representation of the fields includes a vector determining the propagation direction of plane waves. Wave numbers of the fields also depend on their propagation directions pointing to anisotropic character of the developed mathematical model.

Keywords: mathematical models, integro-differential model, electromagnetic monochromatic waves, plane fields, spheroidal particles, anisotropic medium, space dispersion, analytical modeling

For citation. Erofeenko V. T., Urbanovich. A. I. Mathematical model of propagation of electromagnetic waves in composite media with spheroidal particles. *Informatics*, 2018, vol. 15, no. 3, pp. 102–112 (in Russian).

Введение. Разработка математических методов моделирования распространения излучений электромагнитных волн в композитных материалах является актуальным направлением исследований в математической физике [1]. Как правило, композиты представляют собой однородные матрицы, содержащие материальные неоднородности (частицы) различных типов, разнообразие которых велико [2–6]. Анализ таких материалов требует применения специальных математических моделей, адекватно описывающих электрические и магнитные свойства композитных материалов. При этом существенную роль в моделях играют отношения между размерами частиц и длинами волн в вакууме, в материале матрицы и материале неоднородностей. В статьях [2, 3] разработаны модели киральной и биизотропной сред, содержащих сферические частицы со спиральной проводимостью поверхности. В работе [4] исследована среда, содержащая случайно распределенные идеально проводящие спирали, в [6] изучаются среды с биизотропными сферическими частицами. Для исследования электродинамических свойств композитных материалов рассчитываются коэффициенты эффективности экранов из композитов [4, 7, 8].

Одним из классов композитных материалов являются среды с пространственной дисперсией [9–12]. Для них электромагнитные свойства в точке зависят от значений напряженностей электрического и магнитного полей в окрестности рассматриваемой пространственно-временной точки. В данном случае электрическая и магнитная поляризации среды выражаются через интегральные операторы [1, 11], входящие в уравнения электродинамики. Аналитическое решение уравнений Максвелла с интегральными членами требует разработки специальных методов. В связи с этим строятся дифференциальные модели сред, которые позволяют исследовать уравнения и получать их аналитические решения. В статье [12] разработана модель, описывающая распространение электромагнитных волн в материале с пространственной дисперсией, который содержит сферические частицы.

Учитывая влияние размеров формы частиц композитов [5] на свойства анизотропии, в настоящей работе строится математическая модель, описывающая распространение электромагнитных волн в среде с пространственной дисперсией со сфероидальными частицами. В случае вытянутых сфероидальных ориентированных вдоль оси Oz частиц, размеры которых значительно меньше по сравнению с длиной электромагнитной волны, распространяющейся в среде, поля представлены в виде разложений Тейлора. В разработанной модели использовано конечное число слагаемых рядов и опущены относительно малые члены бесконечных рядов. В результате интегро-дифференциальные уравнения Максвелла преобразованы к системе уравнений Максвелла, содержащих дифференциальные операторы второго порядка. Для полученных уравнений построена система аналитических решений в виде плоских электромагнитных полей, распространяющихся в среде.

Интегро-дифференциальная модель биизотропной среды с пространственной дисперсией. Рассмотрим пространство R^3 с декартовой системой координат $Oxyz$, заполненное однородной матрицей с материальными параметрами: $\varepsilon = \varepsilon_r \varepsilon_0$, $\mu = \mu_r \mu_0$ – диэлектрическая и магнитная проницаемости; ε_0 , μ_0 – электрическая и магнитная постоянные. В матрице случайным образом распределены частицы двух сортов в виде вытянутых сфероидов $(x^2 + y^2)/a_j^2 + z^2/b_j^2 = 1$, направленных вдоль координаты Oz ; a_j , b_j – малая и большая оси сфероидов ($j = 1, 2$). Среду из матрицы со сфероидальными частицами будем называть анизотропной средой с пространственной дисперсией. Для математического моделирования такой среды используется эквивалентная однородная среда с пространственной дисперсией. Комплексные амплитуды \vec{E}, \vec{H} электромагнитного поля с временной зависимостью $\exp(-i\omega t)$ в такой среде подчиняются уравнениям

$$\operatorname{rot} \vec{H} = -i\omega(\varepsilon \vec{E} + \vec{P}), \quad \operatorname{rot} \vec{E} = i\omega(\mu \vec{H} + \vec{m}), \quad (1)$$

где ω – круговая частота. Электрическая и магнитная поляризации среды с пространственной дисперсией определяются объемными интегралами по пространственным переменным $\vec{r}_0(x_0, y_0, z_0)$:

$$\vec{P}(\vec{r}) = \frac{\varepsilon_0}{V_1} \int_{D_{1M}} P(\vec{r} - \vec{r}_0) \vec{E}(\vec{r}_0) d\vec{r}_0; \quad (2)$$

$$\vec{m}(\vec{r}) = \frac{\mu_0}{V_2} \int_{D_{2M}} m(\vec{r} - \vec{r}_0) \vec{H}(\vec{r}_0) d\vec{r}_0, \quad (3)$$

где D_{jM} – вытянутый сфероид с осями a_j, b_j и центром в точке $M(x, y, z)$, точка $M_0(x_0, y_0, z_0) \in D_{jM}$, $\vec{r} = (x, y, z)$; $V_j = \frac{4}{3}\pi b_j a_j^2$ – объем сфероида D_{jM} .

Воспользуемся системой вытянутых сфероидальных координат $O\xi\eta\varphi$ [13, с. 39]:

$$\begin{aligned} x &= f \left[(\xi^2 - 1)(1 - \eta^2) \right]^{\frac{1}{2}} \cos \varphi, & y &= f \left[(\xi^2 - 1)(1 - \eta^2) \right]^{\frac{1}{2}} \sin \varphi, \\ z &= f\xi\eta, & 1 < \xi < \infty, & \quad -1 < \eta < 1, \quad 0 \leq \varphi < 2\pi. \end{aligned} \quad (4)$$

Для аналитического задания сфероида D_{jM} введем вытянутые сфероидальные координаты $M \xi_j \eta_j \varphi_j$ с началом координат M , которые связаны с декартовыми координатами формулами

$$\begin{aligned} x_0 &= x + f_j \left[(\xi_j^2 - 1)(1 - \eta_j^2) \right]^{\frac{1}{2}} \cos \varphi_j, & y_0 &= y + f_j \left[(\xi_j^2 - 1)(1 - \eta_j^2) \right]^{\frac{1}{2}} \sin \varphi_j, \\ z_0 &= z + f_j \xi_j \eta_j, & f_j &= \sqrt{b_j^2 - a_j^2}, \quad 1 < \xi_j < \infty, \quad -1 < \eta_j < 1, \quad 0 \leq \varphi_j < 2\pi. \end{aligned} \quad (5)$$

Поверхность сфероида S_j в сфероидальных координатах определяется уравнением

$$\xi_j = c_j = \frac{b_j}{\sqrt{b_j^2 - a_j^2}}.$$

Дифференциальная модель биизотропной среды с пространственной дисперсией специального вида. Система интегро-дифференциальных уравнений (1)–(3) описывает распространение электромагнитных волн в произвольной анизотропной среде с пространственной дисперсией. Для частного случая среды преобразуем интегро-дифференциальную модель к более простой дифференциальной модели уравнений с частными производными второго порядка.

Выразим координату ξ сфероидальной системы координат $O\xi\eta\varphi$ через декартовы координаты x, y, z системы $Oxyz$, используя (4). Получим

$$\xi = F(x, y, z, f) = \frac{1}{f\sqrt{2}} \left[x^2 + y^2 + z^2 + f^2 + \left[(x^2 + y^2 + z^2 + f^2)^2 + 4(x^2 + y^2)z^2 \right]^{\frac{1}{2}} \right]^{\frac{1}{2}}. \quad (6)$$

В случае специальной среды подынтегральные функции для поляризаций (2), (3) положим равными $P(\vec{r} - \vec{r}_0) = \bar{K}_1(F(\vec{r} - \vec{r}_0; f_1))$, $m(\vec{r} - \vec{r}_0) = \bar{K}_2(F(\vec{r} - \vec{r}_0; f_2))$, где $\bar{K}_j(\xi) = K_j(\xi/c_j)$, $K_j(p)$ – заданные функции, $0 \leq p \leq 1$. В итоге получим формулы

$$\begin{aligned} \vec{P}(\vec{r}) &= \frac{\varepsilon_0}{V_1} \int_{D_{1M}} \bar{K}_1(F(x-x_0, y-y_0, z-z_0; f_1)) \vec{E}(\vec{r}_0) d\vec{r}_0, \\ \vec{m}(\vec{r}) &= \frac{\mu_0}{V_2} \int_{D_{2M}} \bar{K}_2(F(x-x_0, y-y_0, z-z_0; f_2)) \vec{H}(\vec{r}_0) d\vec{r}_0. \end{aligned} \quad (7)$$

Методику построения дифференциальной модели опишем с помощью следующей теоремы.

Теорема 1. С точностью до величины третьего порядка малости система интегро-дифференциальных уравнений (1)–(3) с поляризациями (7) эквивалентна системе дифференциальных уравнений

$$\begin{aligned} \text{rot } \vec{H} &= -i\omega \left(\varepsilon_{\Pi} \vec{E} + P_1 \Delta \vec{E} + P_2 \frac{\partial^2}{\partial z^2} \vec{E} \right), \\ \text{rot } \vec{E} &= i\omega \left(\mu_{\Pi} \vec{H} + m_1 \Delta \vec{H} + m_2 \frac{\partial^2}{\partial z^2} \vec{H} \right), \end{aligned} \quad (8)$$

где Δ – оператор Лапласа,

$$\varepsilon_{\Pi} = \varepsilon_0 \left(\varepsilon_r + F_1 \int_1^{c_1} \bar{K}_1(\xi) (3\xi^2 - 1) d\xi \right), \quad \mu_{\Pi} = \mu_0 \left(\mu_r + F_2 \int_0^{c_2} \bar{K}_2(\xi) (3\xi^2 - 1) d\xi \right),$$

$$P_2 = \frac{1}{10} \varepsilon_0 f_1^2 F_1 \int_1^{c_1} \bar{K}_1(\xi) (3\xi^2 - 1) d\xi, \quad F_j = \frac{f_j^3}{b_j a_j^2},$$

$$P_1 = \frac{1}{2} \varepsilon_0 f_1^2 F_1 \int_1^{c_1} \bar{K}_1(\xi) (\xi^2 - 1) \left(\xi^2 - \frac{1}{5} \right) d\xi,$$

$$m_2 = \frac{1}{10} \mu_0 f_2^2 F_2 \int_1^{c_2} \bar{K}_2(\xi) (3\xi^2 - 1) d\xi,$$

$$m_1 = \frac{1}{10} \mu_0 f_2^2 F_2 \int_1^{c_2} \bar{K}_2(\xi) (\xi^2 - 1) \left(\xi^2 - \frac{1}{5} \right) d\xi.$$

Доказательство. Компоненты векторов

$$\vec{E}(\vec{r}_0) = E_x(\vec{r}_0) \vec{e}_x + E_y(\vec{r}_0) \vec{e}_y + E_z(\vec{r}_0) \vec{e}_z,$$

$$\vec{H}(\vec{r}_0) = H_x(\vec{r}_0) \vec{e}_x + H_y(\vec{r}_0) \vec{e}_y + H_z(\vec{r}_0) \vec{e}_z$$

разложим в ряды Тейлора в окрестности точки M . Для построения простой модели ограничимся слагаемыми ряда до второго порядка включительно, пренебрегая величинами третьего порядка малости [12]:

$$\begin{aligned}
 E_\alpha(\vec{r}_0) &= E_\alpha(\vec{r}) + \frac{\partial E_\alpha(\vec{r})}{\partial x}(x_0 - x) + \frac{\partial E_\alpha(\vec{r})}{\partial y}(y_0 - y) + \frac{\partial E_\alpha(\vec{r})}{\partial z}(z_0 - z) + \\
 &+ \frac{1}{2} \left[\frac{\partial^2 E_\alpha(\vec{r})}{\partial x^2}(x_0 - x)^2 + \frac{\partial^2 E_\alpha(\vec{r})}{\partial y^2}(y_0 - y)^2 + \frac{\partial^2 E_\alpha(\vec{r})}{\partial z^2}(z_0 - z)^2 \right] + \\
 &+ \frac{\partial^2 E_\alpha(\vec{r})}{\partial x \partial y}(x_0 - x)(y_0 - y) + \frac{\partial^2 E_\alpha(\vec{r})}{\partial x \partial z}(x_0 - x)(z_0 - z) + \frac{\partial^2 E_\alpha(\vec{r})}{\partial y \partial z}(y_0 - y)(z_0 - z),
 \end{aligned} \tag{9}$$

$$\begin{aligned}
 H_\alpha(\vec{r}_0) &= H_\alpha(\vec{r}) + \frac{\partial H_\alpha(\vec{r})}{\partial x}(x_0 - x) + \frac{\partial H_\alpha(\vec{r})}{\partial y}(y_0 - y) + \frac{\partial H_\alpha(\vec{r})}{\partial z}(z_0 - z) + \\
 &+ \frac{1}{2} \left[\frac{\partial^2 H_\alpha(\vec{r})}{\partial x^2}(x_0 - x)^2 + \frac{\partial^2 H_\alpha(\vec{r})}{\partial y^2}(y_0 - y)^2 + \frac{\partial^2 H_\alpha(\vec{r})}{\partial z^2}(z_0 - z)^2 \right] + \\
 &+ \frac{\partial^2 H_\alpha(\vec{r})}{\partial x \partial y}(x_0 - x)(y_0 - y) + \frac{\partial^2 H_\alpha(\vec{r})}{\partial x \partial z}(x_0 - x)(z_0 - z) + \frac{\partial^2 H_\alpha(\vec{r})}{\partial y \partial z}(y_0 - y)(z_0 - z),
 \end{aligned}$$

где $\alpha = x, y, z$.

Вычислим компоненты P_α электрической поляризации (7). Подставляя (9) в (7), получим

$$\begin{aligned}
 P_\alpha(\vec{r}) &= \epsilon_0 \left(I_0^{(1)} E_\alpha(\vec{r}) + I_1^{(1)} \frac{\partial E_\alpha(\vec{r})}{\partial x} + I_2^{(1)} \frac{\partial E_\alpha(\vec{r})}{\partial y} + I_3^{(1)} \frac{\partial E_\alpha(\vec{r})}{\partial z} + \right. \\
 &+ I_{11}^{(1)} \frac{\partial^2 E_\alpha(\vec{r})}{\partial x^2} + I_{22}^{(1)} \frac{\partial^2 E_\alpha(\vec{r})}{\partial y^2} + I_{33}^{(1)} \frac{\partial^2 E_\alpha(\vec{r})}{\partial z^2} + \\
 &\left. + I_{12}^{(1)} \frac{\partial^2 E_\alpha(\vec{r})}{\partial x \partial y} + I_{13}^{(1)} \frac{\partial^2 E_\alpha(\vec{r})}{\partial x \partial z} + I_{23}^{(1)} \frac{\partial^2 E_\alpha(\vec{r})}{\partial y \partial z} \right);
 \end{aligned} \tag{10}$$

$$I_0^{(j)} = \frac{1}{V_j} \int_{D_M} \bar{K}_j(F(\vec{r} - \vec{r}_0; f_j)) d\vec{r}_0, \quad I_1^{(j)} = \frac{1}{V_j} \int_{D_M} \bar{K}_j(F(\vec{r} - \vec{r}_0; f_j))(x_0 - x) d\vec{r}_0,$$

$$I_2^{(j)} = \frac{1}{V_j} \int_{D_M} \bar{K}_j(F(\vec{r} - \vec{r}_0; f_j))(y_0 - y) d\vec{r}_0, \tag{11}$$

$$I_3^{(j)} = \frac{1}{V_j} \int_{D_M} \bar{K}_j(F(\vec{r} - \vec{r}_0; f_j))(z_0 - z) d\vec{r}_0,$$

$$\begin{aligned}
I_{11}^{(j)} &= \frac{1}{2V_j} \int_{D_{jM}} \bar{K}_j(F(\vec{r} - \vec{r}_0; f_j))(x_0 - x)^2 d\vec{r}_0, \\
I_{22}^{(j)} &= \frac{1}{2V_j} \int_{D_{jM}} \bar{K}_j(F(\vec{r} - \vec{r}_0; f_j))(y_0 - y)^2 d\vec{r}_0, \\
I_{33}^{(j)} &= \frac{1}{2V_j} \int_{D_{jM}} \bar{K}_j(F(\vec{r} - \vec{r}_0; f_j))(z_0 - z)^2 d\vec{r}_0, \\
I_{12}^{(j)} &= \frac{1}{V_j} \int_{D_{jM}} \bar{K}_j(F(\vec{r} - \vec{r}_0; f_j))(x_0 - x)(y_0 - y) d\vec{r}_0, \\
I_{13}^{(j)} &= \frac{1}{V_j} \int_{D_{jM}} \bar{K}_j(F(\vec{r} - \vec{r}_0; f_j))(x_0 - x)(z_0 - z) d\vec{r}_0, \\
I_{23}^{(j)} &= \frac{1}{V_j} \int_{D_{jM}} \bar{K}_j(F(\vec{r} - \vec{r}_0; f_j))(y_0 - y)(z_0 - z) d\vec{r}_0.
\end{aligned} \tag{11}$$

Для вычисления интегралов (11) введем две сфероидальные системы координат $M \xi_j \eta_j \varphi_j$ с началом в точке M (5). Интегралы (11) вычислим в сфероидальных координатах, подставляя (5) в (11), учитывая (6) и полагая $d\vec{r}_0 = f_j^3 (\xi_j^2 - \eta_j^2) d\xi_j d\eta_j d\varphi_j$.

Получим следующие значения интегралов:

$$\begin{aligned}
I_0^{(j)} &= \frac{f_j^3}{V_j} \int_1^{c_j} \bar{K}_j(\xi_j) \int_0^1 \int_{-1}^1 (\xi_j^2 - \eta_j^2) d\eta_j d\varphi_j d\xi_j = \\
&= \kappa_j = \frac{f_j^3}{b_j a_j^2} \int_1^{c_j} \bar{K}_j(\xi) (3\xi^2 - 1) d\xi, \\
I_1^{(j)} &= I_2^{(j)} = I_3^{(j)} = 0, \quad I_{12}^{(j)} = I_{13}^{(j)} = I_{23}^{(j)} = 0, \\
I_{11}^{(j)} &= \frac{f_j^5}{2V_j} \int_1^{c_j} \bar{K}_j(\xi_j) \int_0^1 \int_{-1}^1 (\xi_j^2 - \eta_j^2) (\xi_j^2 - 1) (1 - \eta_j^2) \cos^2 \phi_j d\eta_j d\phi_j d\xi_j = \\
&= \frac{\pi f_j^5}{2V_j} \int_1^{c_j} \bar{K}_j(\xi_j) (\xi_j^2 - 1) \int_{-1}^1 (\xi_j^2 - \eta_j^2) (1 - \eta_j^2) d\eta_j d\xi_j = g_j = \\
&= \frac{f_j^5}{2b_j a_j^2} \int_1^{c_j} \bar{K}_j(\xi) (\xi^2 - 1) \left(\xi^2 - \frac{1}{5} \right) d\xi, \quad I_{22}^{(j)} = g_j, \\
I_{33}^{(j)} &= \frac{f_j^5}{2V_j} \int_1^{c_j} \bar{K}_j(\xi_j) \xi_j^2 \int_0^1 \int_{-1}^1 (\xi_j^2 - \eta_j^2) \eta_j^2 d\eta_j d\varphi_j d\xi_j = g_j + \tau_j, \quad \tau_j = \frac{1}{10} \kappa_j f_j^2.
\end{aligned} \tag{12}$$

Подставив значения интегралов (12) в (10), получим формулы для декартовых компонент вектора электрической поляризации

$$P_\alpha(\vec{r}) = \varepsilon_0 \left(\kappa_1 E_\alpha(\vec{r}) + \tau_1 \frac{\partial^2}{\partial z^2} E_\alpha(\vec{r}) + g_1 \Delta E_\alpha(\vec{r}) \right).$$

Тогда вектор электрической поляризации будет определяться формулой

$$\vec{P} = \varepsilon_0 \left(\kappa_1 \vec{E} + \tau_1 \frac{\partial^2}{\partial z^2} \vec{E} + g_1 \Delta \vec{E} \right). \quad (13)$$

После аналогичных преобразований получим формулу для вектора магнитной поляризации

$$\vec{m} = \mu_0 \left(\kappa_2 \vec{H} + \tau_2 \frac{\partial^2}{\partial z^2} \vec{H} + g_2 \Delta \vec{H} \right). \quad (14)$$

Подставляя выражения (13), (14) в уравнения (1), приходим к требуемой дифференциальной модели. ■

Аналитическое построение базисных плоских электромагнитных полей. Построим систему плоских электромагнитных полей, распространяющихся в среде с пространственной дисперсией, т. е. полей \vec{E}, \vec{H} , удовлетворяющих уравнениям (8). Плоские поля представляют собой электромагнитные поля вида

$$\begin{aligned} \vec{E} &= (A_1 \vec{e}_x + B_1 \vec{e}_y + C_1 \vec{e}_z) \Phi(x, y) e^{vz}, \\ \vec{H} &= (A_2 \vec{e}_x + B_2 \vec{e}_y + C_2 \vec{e}_z) \Phi(x, y) e^{vz}, \end{aligned} \quad (15)$$

где $\Phi(x, y) = \exp(i\alpha_1 x + i\alpha_2 y)$; $\alpha_1, \alpha_2, v, A_j, B_j, C_j$ – постоянные.

Для аналитического описания полей (15) воспользуемся волновыми полями [11, с. 96]

$$\vec{W}^{(\mp 1)}(\vec{r}; \alpha_1, \alpha_2; k) = \frac{i}{\lambda} (\alpha_2 \vec{e}_x - \alpha_1 \vec{e}_y) \Phi(x, y) \exp(\mp v z), \quad (16)$$

$$\vec{W}^{(\mp 2)}(\vec{r}; \alpha_1, \alpha_2; k) = \frac{1}{k} \left(\mp \frac{iv}{\lambda} (\alpha_1 \vec{e}_x + \alpha_2 \vec{e}_y) + \lambda \vec{e}_z \right) \Phi(x, y) \exp(\mp v z),$$

где α_1, α_2, k – произвольные комплексные постоянные,

$$\lambda = \sqrt{\alpha_1^2 + \alpha_2^2}, \quad 0 \leq \arg \lambda < \pi, \quad v = \sqrt{\lambda^2 - k^2}, \quad -\frac{\pi}{2} \leq \arg v < \frac{\pi}{2}.$$

Заметим, что для полей (16) выполнены формулы [11, с. 98]

$$\text{rot } \vec{W}^{(\mp 1)} = k \vec{W}^{(\mp 2)}, \quad \text{rot } \vec{W}^{(\mp 2)} = k \vec{W}^{(\mp 1)},$$

$$\frac{\partial^2}{\partial z^2} \vec{W}^{(\mp 1)} = v^2 \vec{W}^{(\pm 1)}, \quad \frac{\partial^2}{\partial z^2} \vec{W}^{(\mp 2)} = v^2 \vec{W}^{(\pm 2)}, \quad (17)$$

$$\text{div } \vec{W}^{(\mp 1)} = 0, \quad \text{div } \vec{W}^{(\mp 2)} = 0, \quad \Delta \vec{W}^{(\mp j)} = -k^2 \vec{W}^{(\mp j)}.$$

Решение системы уравнений (8) вида (15) представим через волновые поля (16).

Теорема 2. Электромагнитные поля, распространяющиеся в среде с пространственной дисперсией со сферoidalными частицами, определяются формулами

$$\vec{E} = E_0 \vec{W}^{(\mp 1)}(\vec{r}; \alpha_1, \alpha_2; k_+), \quad \vec{H} = E_0 \gamma_+ \vec{W}^{(\mp 2)}(\vec{r}; \alpha_1, \alpha_2; k_+); \quad (18)$$

$$\vec{E} = E_0 \vec{W}^{(\mp 2)}(\vec{r}; \alpha_1, \alpha_2; k_+), \quad \vec{H} = E_0 \gamma_+ \vec{W}^{(\mp 1)}(\vec{r}; \alpha_1, \alpha_2; k_+); \quad (19)$$

$$\vec{E} = E_0 \vec{W}^{(\mp 1)}(\vec{r}; \alpha_1, \alpha_2; k_-), \quad \vec{H} = E_0 \gamma_- \vec{W}^{(\mp 2)}(\vec{r}; \alpha_1, \alpha_2; k_-); \quad (20)$$

$$\vec{E} = E_0 \vec{W}^{(\mp 2)}(\vec{r}; \alpha_1, \alpha_2; k_-), \quad \vec{H} = E_0 \gamma_- \vec{W}^{(\mp 1)}(\vec{r}; \alpha_1, \alpha_2; k_-), \quad (21)$$

где поля (18) – (21) с верхними индексами $-1, -2$ определяют прямые волны, а с индексами $+1, +2$ – обратные волны; поля (18), (20) определяют ТЕ-поляризованные электромагнитные поля, поля (19), (21) – ТН-поляризованные электромагнитные поля; E_0 – постоянная, физическая

размерность $[E_0] = \frac{B}{M}$;

$$\alpha_1 = k_0 \sin \theta_0 \cos \varphi_0, \quad \alpha_2 = k_0 \sin \theta_0 \sin \varphi_0, \quad k_0 = \frac{\omega}{c}, \quad \lambda = k_0 \sin \theta_0, \quad 0 \leq \theta_0 < \frac{\pi}{2},$$

c – скорость света в вакууме,

$$\gamma_{\pm} = \sqrt{\frac{-B \pm C}{2m\mu_{\lambda}}}, \quad 0 \leq \arg \gamma_{\pm} < \pi,$$

$$k_{\pm} = 2i\omega\gamma_{\pm}\mu_{\lambda} \frac{(P\mu_{\lambda} - m\varepsilon_{\lambda})}{(2P\mu_{\lambda} - B \pm C)},$$

$$C = \sqrt{B^2 - 4mP\mu_{\lambda}\varepsilon_{\lambda}}, \quad 0 \leq \arg C < \pi, \quad (22)$$

$$B = m\varepsilon_{\lambda} + P\mu_{\lambda} + \omega^2(m\varepsilon_{\lambda} - P\mu_{\lambda})^2,$$

$$m = m_1 + m_2, \quad P = P_1 + P_2, \quad \varepsilon_{\lambda} = \varepsilon_{\Pi} + P_2\lambda^2, \quad \mu_{\lambda} = \mu_{\Pi} + m_2\lambda^2.$$

Доказательство. Систему уравнений (8) запишем в виде

$$\begin{aligned} \operatorname{rot} \vec{H} &= \bar{\delta}_1 \vec{E} + \bar{\delta}_2 \Delta \vec{E} + \bar{\delta}_0 \frac{\partial^2}{\partial z^2} \vec{E}, \\ \operatorname{rot} \vec{E} &= \bar{\beta}_1 \vec{H} + \bar{\beta}_2 \Delta \vec{H} + \bar{\beta}_0 \frac{\partial^2}{\partial z^2} \vec{H}, \end{aligned} \quad (23)$$

где

$$\begin{aligned} \bar{\delta}_1 &= -i\omega\varepsilon_{\Pi}, \quad \bar{\delta}_2 = -i\omega P_1, \quad \bar{\delta}_0 = -i\omega P_2, \\ \bar{\beta}_1 &= i\omega\mu_{\Pi}, \quad \bar{\beta}_2 = i\omega m_1, \quad \bar{\beta}_0 = i\omega m_2. \end{aligned} \quad (24)$$

В качестве решения уравнений (23) выберем поля (16)

$$\vec{E} = E_0 \vec{W}^{(\mp 1)}(\vec{r}; \alpha_1, \alpha_2; k), \quad \vec{H} = E_0 \gamma \vec{W}^{(\mp 2)}(\vec{r}; \alpha_1, \alpha_2; k). \quad (25)$$

Для определения неизвестных постоянных k и γ поля (25) подставим в уравнения (23). Учитывая формулы (17), для определения величин k, γ получим систему нелинейных алгебраических уравнений

$$\begin{aligned} \gamma k &= \bar{\delta}_1 - \bar{\delta}_2 k^2 + \bar{\delta}_0 (\lambda^2 - k^2), \\ k &= \gamma (\bar{\beta}_1 - \bar{\beta}_2 k^2 + \bar{\beta}_0 (\lambda^2 - k^2)). \end{aligned} \quad (26)$$

Введем обозначения

$$\delta_1 = \bar{\delta}_1 + \bar{\delta}_0 \lambda^2, \quad \delta_2 = \bar{\delta}_2 + \bar{\delta}_0, \quad \beta_1 = \bar{\beta}_1 + \bar{\beta}_0 \lambda^2, \quad \beta_2 = \bar{\beta}_2 + \bar{\beta}_0. \quad (27)$$

Тогда система уравнений примет вид

$$\delta_2 k^2 + \gamma k - \delta_1 = 0, \quad \beta_2 \gamma k^2 + k - \beta_1 \gamma = 0. \quad (28)$$

В статье [12] система (28) разрешена аналитически в радикалах и получены решения $k = k_+, \quad \gamma = \gamma_+$ и $k = k_-, \quad \gamma = \gamma_-$, которые определяются формулами

$$\begin{aligned} \gamma_{\pm} &= \sqrt{\frac{b \mp c}{2\beta_1\beta_2}}, \quad 0 \leq \arg \gamma_{\pm} < \pi, \quad k_{\pm} = \gamma_{\pm} \frac{\delta_1\beta_2 - \beta_1\delta_2}{\beta_2\gamma_{\pm}^2 - \delta_2}, \\ c &= \sqrt{b^2 - 4\delta_1\delta_2\beta_1\beta_2}, \quad 0 \leq \arg c < \pi, \\ b &= (\delta_1\beta_2 - \beta_1\delta_2)^2 + \delta_1\beta_2 + \beta_1\delta_2. \end{aligned} \quad (29)$$

Подставив выражения (27), (24) в формулы (29), получим требуемые формулы. ■

Заключение. В настоящей работе интегро-дифференциальная модель, описывающая распространение электромагнитных волн в среде с пространственной дисперсией, которая содержит вытянутые сфероидальные частицы, ориентированные вдоль оси Oz , преобразована с точностью до величин третьего порядка малости в дифференциальную математическую модель. Интегро-дифференциальные уравнения Максвелла монохроматической электродинамики представлены в виде системы дифференциальных уравнений второго порядка. Уравнения решены аналитически, и построена система четырех прямых и четырех обратных TE - и TH -поляризованных плоских электромагнитных волн, распространяющихся в однородной среде. Число построенных независимых волн вдвое больше, чем число волн, распространяющихся в обычных магнитоэлектрических средах. Поля представлены через классические базисные электромагнитные поля. Разработанная модель описывает анизотропную среду, так как волновые числа полей зависят от направления распространения волн.

Список использованных источников

1. Виноградов, А. П. Электродинамика композитных материалов / А. П. Виноградов. — М. : Эдиториал УРСС, 2001. — 206 с.
2. Костин, М. В. К теории киральной среды на основе сферических спирально проводящих частиц / М. В. Костин, В. В. Шевченко // Радиотехника и электроника. — 1998. — Т. 43, № 8. — С. 921–926.

3. Шатров, А. Д. Модель биизотропной среды из резонансных сферических частиц с идеальной смешанной проводимостью поверхности вдоль спиральных линий / А. Д. Шатров // Радиотехника и электроника. – 2000. – Т. 45, № 10. – С. 1168–1170.
4. Проникновение электромагнитных волн через композитные экраны, содержащие идеально проводящие спирали / В. Т. Ерофеенко [и др.] // Инженерно-физический журнал. – 2011. – Т. 84, № 4. – С. 740–746.
5. Балагуров, Б. Я. О влиянии формы включений на проводимость двумерных моделей композитов / Б. Я. Балагуров // Журнал технической физики. – 2011. – Т. 81, вып. 5. – С. 5–8.
6. Ерофеенко, В. Т. Электродинамическая модель расчета эффективных параметров композитов из сферических биизотропных частиц / В. Т. Ерофеенко // Информатика. – 2014. – № 1(41). – С. 45–58.
7. Ерофеенко, В. Т. Экранирование электромагнитных волн плоским однослойным экраном из материалов с пространственной дисперсией / В. Т. Ерофеенко, В. Ф. Бондаренко // Информатика. – 2017. – № 4(56). – С. 5–15.
8. Ерофеенко, В. Т. Экранирование электромагнитных полей экранами из матричных композитов, содержащих биизотропные частицы / В. Т. Ерофеенко, В. Ф. Бондаренко // Информатика. – 2014. – № 3(43). – С. 28–43.
9. Агранович, В. М. Кристаллооптика с учетом пространственной дисперсии и теория экситонов / В. М. Агранович, В. Л. Гинзбург. – М. : Наука, 1979. – 432 с.
10. Силин, Р. А. Обратные волны и пространственная дисперсия / Р. А. Силин, И. Р. Тимошина // Радиотехника и электроника. – 2012. – Т. 57, № 7. – С. 725–733.
11. Ерофеенко, В. Т. Аналитическое моделирование в электродинамике / В. Т. Ерофеенко, И. С. Козловская. – Минск : БГУ, 2010. – 304 с.
12. Ерофеенко, В. Т. Моделирование распространения электромагнитных волн в средах с пространственной дисперсией / В. Т. Ерофеенко // Информатика. – 2017. – № 3(55). – С. 5–12.
13. Иванов, Е. А. Дифракция электромагнитных волн на двух телах / Е. А. Иванов. – Минск : Наука и техника, 1968. – 584 с.

References

1. Vinogradov A. P. *Jelektrodinamika kompozitnyh materialov. Electrodynamics of Composite Materials*. Moscow, Editorial URSS Publ., 2001, 206 p. (in Russian).
2. Kostin M. V., Schevchenko V. V. *K teorii kiralnoj sredy na osnove sfericheskikh spiralno provodyaschih chastits* [The theory of the chiral medium on the spherical helically conductive particles basis]. *Radiotekhnika i elektronika [Radio Engineering and Electronics]*, 1998, vol. 43, no. 8, pp. 921–926 (in Russian).
3. Shatrov A. D. *Model biizotropnoj sredy iz rezonansnyh sfericheskikh chastits s idealnoj smeshanoj provodimostju poverhnosti vdol spiralnyh linij* [Biisotropic medium model of resonant spherical particles with ideal mixed surface conductivity along helical line]. *Radiotekhnika i elektronika [Radio Engineering and Electronics]*, 2000, vol. 45, no. 10, pp. 1168–1170 (in Russian).
4. Erofeenko V. T., Demidchik V. I., Malyi S. V., Kornev R. V. *Pronikновение jelektromagnitnyh voln cherez kompozitnye ekrany, soderzhashhie ideal'no provodjashhie spirali* [Penetration of electromagnetic waves through composite screens containing ideally conducting spirals]. *Inzhenerno-fizicheskiy zhurnal [Journal of Engineering Physics and Thermophysics]*, 2011, vol. 84, no. 4, pp. 740–746 (in Russian).
5. Balagurov B. Y. *O vliyanii formy vkluchenij na provodimost dvumernih modelej kompozitov* [The inclusion forms influence on the two-dimensional composites models conductivity]. *Zhurnal tehnichekoj fiziki [Technical Physics Journal]*, 2011, vol. 81(5), pp. 5–8 (in Russian).
6. Erofeenko V. T. *Electrodinamicheskaya model rascheta effektivnyh parametrov kompozitov iz sfericheskikh biizotropnyh chastits* [Electrodynamic model of calculation of effective parameters of composites from spherical isotropic particles]. *Informatika [Informatics]*, 2014, no. 1(41), pp. 45–58 (in Russian).
7. Erofeenko V. T., Bondarenko V. F. *Ekranirovanie elektromagnitnyh voln ploskim odnoslojnym ekranom iz materialov s prostranstvennoj dispersiej* [Shielding of electromagnetic waves by a flat single-layer screen of materials with spatial dispersion]. *Informatika [Informatics]*, 2017, vol. 4(56), pp. 5–15 (in Russian).
8. Erofeenko V. T., Bondarenko V. F. *Ekranirovanie elektromagnitnyh polej ekranami iz matrichnyh kompozitov, soderzhashhih biizotropnye chasticy* [Screening of electromagnetic fields by screens from matrix composites containing bi-isotropic particles]. *Informatika [Informatics]*, 2014, no. 3(43), pp. 28–43 (in Russian).
9. Agranovich V. M., Ginzburg V. L. *Kristallooptika s uchetom prostranstvennoj dispersii i teorija jeksitonov. Crystal Optics with Allowance for Spatial Dispersion and the Theory of Excitons*. Moscow, Nauka Publ., 1979, 432 p. (in Russian).
10. Silin R. A., Timoshina I. R. *Obratnye volny i prostranstvennaja dispersija* [Reverse waves and spatial dispersion]. *Radiotekhnika i elektronika [Radio Engineering and Electronics]*, 2012, vol. 57, no. 7, pp. 725–733 (in Russian).
11. Erofeenko V. T., Kozlovskaja I. S. *Analiticheskoe modelirovanie v jelektrodinamike. Analytical Modeling in Electrodynamics*. Minsk, BGU, 2010, 304 p. (in Russian).
12. Erofeenko V. T. *Modelirovanie rasprostraneniya jelektromagnitnyh voln v sredah s prostranstvennoj dispersiej* [Modeling the propagation of electromagnetic waves in media with spatial dispersion]. *Informatika [Informatics]*, 2017, no. 3(55), pp. 5–12 (in Russian).
13. Ivanov E. A. *Difraktsiya elektromagnitnyh voln na dvuh telah. Diffraction of Electromagnetic Waves on Two Bodies*. Minsk, Nauka i tehnika Publ., 1968, 584 p. (in Russian).

Информация об авторах

Ерофеенко Виктор Тихонович – доктор физико-математических наук, профессор, главный научный сотрудник научно-исследовательской лаборатории математических методов защиты информации, Учреждение Белорусского государственного университета «НИИ прикладных проблем математики и информатики» (пр. Независимости, 4, 220030, Минск, Республика Беларусь). E-mail: bsu_erofeenko@tut.by

Урбанович Александр Иосифович – кандидат физико-математических наук, доцент кафедры математического моделирования и управления, факультет прикладной математики и информатики, Белорусский государственный университет (пр. Независимости, 4, 220030, Минск, Республика Беларусь). E-mail: urbanovich@gmail.com

Information about the authors

Viktor T. Erofeenko – Dr. Sc. (Physics and Mathematics), Professor, Chief Research Associate of the Research Laboratory of Mathematical Methods of Information Security, Institution of the Belarusian State University “Research Institute for Applied Problems of Mathematics and Informatics” (4, Nezavisimosti Ave., 220030, Minsk, Republic of Belarus). E-mail: bsu_erofeenko@tut.by

Aleksandr I. Urbanovich – Ph. D. (Physics and Mathematics), Associate Professor at the Department of Mathematical Modeling and Control, Docent, Faculty of Applied mathematics and computer science, Belarusian State University (4, Nezavisimosti Ave., 220030, Minsk, Republic of Belarus). E-mail: urbanovich@gmail.com

ISSN 1816-0301 (print)

УЧЕНЫЕ БЕЛАРУСИ
SCIENTISTS OF BELARUS

АЛЕКСАНДР ВАСИЛЬЕВИЧ ТУЗИКОВ
(К 60-летию со дня рождения)



5 сентября 2018 г. исполнилось 60 лет генеральному директору Объединенного института проблем информатики НАН Беларуси Александру Васильевичу Тузи́кову.

Александр Васильевич известен своими работами в области информатики, обработки изображений и распознавания образов, дискретной математики, биоинформатики. Им введены и исследованы меры сходства и симметрий выпуклых объектов произвольной размерности, основанные на теории Брунна – Минковского. Доказано, что меры сходства выпуклых многоугольников и многогранников, инвариантные относительно группы аффинных преобразований, допускают эффективное вычисление. Разработан новый метод оценивания степени симметричности выпуклых многомерных множеств, бинарных и полутоновых изображений, основанный на использовании преобразований симметризации; методы синтеза и анализа инвариантных морфологических фильтров на полных решетках; параллельные алгоритмы вычисления морфологических операций. Получены аналитические формулы вычисления геометрических моментов полигональных объектов произвольной размерности и плоских объектов, ограниченных сплайновой кривой, формулы оценки объема трехмерных объектов по нескольким непараллельным сечениям. Разработаны алгоритмы взаимной обработки нескольких изображений для автоматической реконструкции трехмерных сцен по цифровым изображениям. Ряд научно-технических разработок А. В. Тузи́кова, выполненных совместно с коллегами и учениками, используется при диагностике заболеваний на основе анализа цифровых изображений различных модальностей (ультразвуковых, томографических, цитологических и др.), а также для поддержки планирования и проведения медицинских операций. Совместно с коллегами А. В. Тузи́ков занимается исследованиями в области биоинформатики, включающими изучение белок-белковых взаимодействий на основе компьютерного моделирования и анализа баз данных белковых комплексов, структурного сравнения и предсказания взаимодействия белков, анализа генетических данных с целью выявления в них зависимостей, компьютерного моделирования и поиска в базах химических соединений миметиков антител широкого нейтрализующего действия против ВИЧ-1, которые могут быть использованы для разработки перспективных лекарственных препаратов.

Александр Васильевич опубликовал свыше 350 печатных работ, в том числе пять монографий и более 50 работ на английском языке, включая статьи в престижных зарубежных журналах (<https://scholar.google.com/citations?user=EL7ilZEAAAJ&hl=ru>). Выступал с научными докладами на международных конференциях и в ведущих зарубежных научных центрах в России, США, Канаде, Великобритании, Германии, Франции, Италии, Испании, Нидерландах, Австралии, Японии, Польше, Литве и др.

Выполняет большую научно-организационную работу: является председателем научного совета и научным руководителем (совместно с С. В. Абламейко) государственной программы научных исследований «Информатика, космос и безопасность» на 2016–2020 гг., научным руководителем программ Союзного государства «Разработка космических и наземных средств

обеспечения потребителей России и Беларуси информацией дистанционного зондирования Земли (Мониторинг-СГ)» на 2013–2017 гг. и «Исследования и разработка высокопроизводительных информационно-вычислительных технологий для увеличения и эффективного использования ресурсного потенциала углеводородного сырья Союзного государства (СКИФ-Недра)» на 2015–2018 гг. Подготовил пять кандидатов наук. Входит в состав редколлегии журнала «Вести НАН Беларуси. Сер. физ.-мат. наук», является главным редактором журнала «Информатика», членом редколлегии и рецензентом нескольких международных журналов. Он является председателем совета по защите докторских диссертаций при ОИПИ НАН Беларуси, председателем ученого совета ОИПИ НАН Беларуси, членом наблюдательного совета Парка высоких технологий, представителем Беларуси в Международной ассоциации по распознаванию образов (IAPR).

А. В. Тузиков родился 5 сентября 1958 г. в Полоцке в семье военного. В 1975 г. закончил с золотой медалью среднюю школу и поступил на факультет прикладной математики Белгос-университета. Исследовательскую работу начал на пятом курсе университета под руководством профессора В. С. Танаева, тогда же получил первые научные результаты. По окончании в 1980 г. университета был направлен на работу в Институт технической кибернетики АН БССР и в том же году поступил в аспирантуру. В 1980–1983 гг. под руководством В. С. Танаева занимается исследованиями в области теории расписаний. В 1985 г. защищает кандидатскую диссертацию в области построения оптимальных расписаний. Наряду с исследовательской работой активно участвует в общественной жизни: возглавляет комсомольское бюро, затем комитет комсомола института, а с 1986 по 1989 г. руководит комитетом комсомола АН БССР, в 1987 г. принимает участие в работе XX съезда ВЛКСМ. В 1990 г. ему присваивается звание лауреата премии Ленинского комсомола Белоруссии за работу «Комбинаторные алгоритмы решения задач оптимального планирования и проектирования» (совместно с С. А. Бородичем, М. Я. Ковалевым, Н. М. Корнеенко и В. А. Струсевичем).

В 1988 г. А. В. Тузиков начинает заниматься математической морфологией – перспективным направлением на стыке цифровой обработки изображений, стереологии и интегральной геометрии, основанном французскими учеными Ж. Серра и Ж. Матероном в начале 1970-х гг. В 1991 г. по приглашению профессора Ж. Серра проходит шестимесячную стажировку в Центре математической морфологии Горной школы Парижа. В 1995–1996 гг. провел десять месяцев в Нидерландах, сначала в Центре вычислительных наук (CWI) в Амстердаме, а затем в Университете Гронингена, где совместно с докторами Х. Хеймансом и Е. Роердинком исследовал задачи сравнения геометрических объектов и оценивания их симметричности в рамках математической морфологии. По результатам исследований им опубликована монография «Анализ симметричности и сравнение объектов на основе сложения Минковского» (1998), ряд статей в международных журналах и в 1999 г. успешно защищена докторская диссертация. В 2002 г. А. В. Тузикову присуждена Государственная премия Республики Беларусь в области естественных наук за цикл исследований «Распознавание и анализ стохастических данных и цифровых изображений» (совместно с С. В. Абламейко, Ю. С. Хариным, Р. Х. Садыховым и В. В. Старовойтовым).

С 2001 г. Александр Васильевич вместе с учениками занимается разработкой алгоритмов обработки медицинских изображений различной модальности (ультразвуковых, компьютерной и магнитно-резонансной томографии, гистологических), а также систем поддержки медицинских телеконсультаций. В 2001–2002 гг. провел пять месяцев в Высшей школе телекоммуникаций Парижа, где совместно с профессором И. Блок и О. Коллиотом разработал новый алгоритм поиска плоскости отражательной симметрии для полученных с помощью ЯМР-томографии изображений мозга, который использует максимизацию меры сходства исходного и преобразованного изображений и выбор начального решения для последующей оптимизации на основе анализа эллипсоида инерции изображения. Алгоритм апробирован на тестовых и реальных изображениях головного мозга и показал устойчивые результаты при различных уровнях искажений. Под руководством А. В. Тузикова разработано программное обеспечение для поддержки проведения медицинских телеконсультаций по лечению рака щитовидной железы, алгоритмическое и программное обеспечение анализа трехмерных ультразвуковых изображений для оценки объе-

ма объектов ультразвуковых изображений на основе сплайновой интерполяции по измерениям в непараллельных сечениях (совместно с В. В. Ляховским, С. В. Рыльковым, А. Л. Богушем).

В августе 2003 г. А. В. Тузиков назначен заместителем генерального директора по научной работе ОИПИ НАН Беларуси, а с 2005 г. возглавил лабораторию математической кибернетики. С 2004 по 2007 г., будучи заместителем научного руководителя программы Союзного государства «Космос-СТ», занимается организацией и координацией исследований по космической тематике. В этот период совместно с Д. В. Жуком проводит исследования и разработку алгоритмов взаимной обработки нескольких изображений с целью автоматической реконструкции трехмерных сцен по цифровым изображениям. Одно из важнейших приложений этих результатов – реконструкция рельефа местности по спутниковым изображениям. В 2006 г. в издательстве «Белорусская наука» вышла монография А. В. Тузикова, С. А. Шейнина и Д. В. Жука «Математическая морфология, моменты, стереообработка: избранные вопросы обработки и анализа цифровых изображений». С января 2009 г. А. В. Тузиков исполняет обязанности генерального директора ОИПИ НАН Беларуси.

С 2005 г. А. В. Тузиков уделяет большое внимание еще одному перспективному направлению современной науки – биоинформатике. Созданная им научная группа занимается разработкой математических методов и алгоритмов структурного сравнения белков и предсказания интерфейсов взаимодействия белков в тесном сотрудничестве с лабораторией профессора И. Ваксера, возглавляющего Центр вычислительной биологии в Университете Канзаса (США). Новые результаты по компьютерному моделированию и поиску в базах химических соединений миметиков широкого нейтрализующего действия против ВИЧ-1, которые могут быть использованы для разработки перспективных лекарственных препаратов, получены совместно со специалистами научной группы под руководством А. М. Андрианова. В 2017 г. А. В. Тузикову присуждена премия Национальной академии наук Беларуси за цикл работ «Компьютерный дизайн потенциальных ингибиторов ВИЧ-1, перспективных для создания противовирусных препаратов нового поколения» (совместно с А. М. Андриановым и И. А. Кашиным). Разрабатываются методы и алгоритмы анализа генетических данных, в частности выявления зависимостей между мутациями в генах и медицинскими данными для различных инфекционных заболеваний (совместно с Р. С. Сергеевым).

Александр Васильевич уделяет много времени работе с молодежью, в течение ряда лет преподает на факультете прикладной математики и информатики Белгосуниверситета (с 2001 г. – профессор), постоянно руководит научной работой студентов, магистрантов и аспирантов. По инициативе А. В. Тузикова в апреле 2017 г. на факультете прикладной математики и информатики Белгосуниверситета была создана кафедра биомедицинской информатики.

В 2014 г. А. В. Тузикову присвоено звание члена-корреспондента Национальной академии наук Беларуси.

Коллеги и друзья желают Александру Васильевичу новых творческих успехов в его многогранной научной, организационной и педагогической деятельности, здоровья и благополучия ему, его семье и близким.

Основные научные труды

1. Абламейко С. В., Анищенко В. В., Лапицкий В. А., Тузиков А. В. Медицинские информационные технологии и системы. – Минск : ОИПИ НАН Беларуси, 2007. – 176 с.
2. Андрианов А., Кашин И., Тузиков А. Разработка новых ингибиторов ВИЧ методами молекулярного моделирования. – Lambert Academic Publishing, 2018. – 127 p.
3. Андрианов А. М., Корноушенко Ю. В., Тузиков А. В., Кисель М. А. In silico дизайн и синтез новых ингибиторов ВИЧ на основе гликолипидов. – Lambert Academic Publishing, 2017. – 118 p.
4. Кибернетика и информатика в Национальной академии наук Беларуси: очерки развития / Объединенный институт проблем информатики НАН Беларуси ; науч. ред. С. В. Абламейко, А. В. Тузиков, О. И. Семенков. – Минск : Технология, 2015 – 348 с.
5. Корняко С. А., Семенков О. И., Тузиков А. В. Космические средства и технологии. Основные результаты реализации белорусской части научно-технической программы Союзного государства «Космос-НТ». – Минск : ОИПИ НАН Беларуси, 2013. – 114 с.

6. Тузиков А. В. Анализ симметричности и сравнение объектов на основе сложения Минковского. – Минск : Ин-т техн. кибернетики НАН Беларуси, 1998. – 176 с.
7. Тузиков А. В., Шейнин С. А., Жук Д. В. Математическая морфология, моменты, стереообработка: избранные вопросы обработки и анализа цифровых изображений. – Минск : Белорус. наука, 2006. – 198 с.
8. Sergeev R. S., Kavaliou I. S., Sataneuski U. V., Gabrielian A., Rosenthal A., Tartakovsky M., Tuzikov A.V. Genome-wide Analysis of MDR and XDR Tuberculosis from Belarus: Machine-learning Approach. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2017, 11 p. doi: 10.1109/TCBB.2017.2720669.
9. Andrianov A. M., Kashyn I. A., Tuzikov A. V. Computational identification of novel entry inhibitor scaffolds mimicking primary receptor CD4 of HIV-1 gp120. *Journal of Molecular Modeling*, 2017, vol. 23, no. 1, pp. 1–18.
10. Kirys T., Ruvinsky A. M., Singla D., Tuzikov A. V., Kundrotas P. J., Vakser I. A. Simulated unbound structures for benchmarking in protein docking in Dockground resource. *BMC Bioinformatics*, 2015, 16:243.
11. Andrianov A. M., Kashyn I. A., Tuzikov A. V. Computational discovery of novel HIV-1 entry inhibitors based on potent and broad neutralizing antibody VRC01. *Journal of Molecular Graphics and Modelling*, 2015, no. 61, pp. 262–271.
12. Anishchenko I., Kundrotas P. J., Tuzikov A. V., Vakser I. A. Protein models: The grand challenge of protein docking. *Proteins: Structure, Function, Bioinformatics*, 2014, vol. 82, no. 2, pp. 278–287.
13. Ruvinsky A. M., Kirys T., Tuzikov A. V., Vakser I. A. Ensemble-based characterization of unbound and bound states on protein energy landscape. *Protein Science*, 2013, vol. 22, no. 6, pp. 734–744.
14. Ruvinsky A. M., Kirys T., Tuzikov A. V., Vakser I. A. Side-Chain Conformational Changes upon Protein-Protein Association. *Journal of Molecular Biology*, 2011, vol. 408, no. 2, pp. 356–365.
15. Andrianov A. M., Anishchenko I. V., Tuzikov A. V. Discovery of Novel Promising Targets for Anti-AIDS Drug Developments by Computer Modeling: Application to the HIV-1 gp120 V3 Loop. *Journal of Chemical Information and Modeling*, 2011, vol. 51(10), pp. 2760–2767.
16. Hancharenka V., Tuzikov A., Arkhipau V., Kryvanos A. Preoperative planning of pelvic and lower limbs surgery by CT image processing. *Pattern Recognition and Image Analysis*, 2009, vol. 19, no. 1, pp. 109–113.
17. Sanko D., Tuzikov A. The Description and Visualization of Vascular Tree from 3D MR Brain Angiography Images. *Pattern Recognition and Image Analysis*, 2006, vol. 16, no. 1, pp. 54–57.
18. Snezhko E. V., Tuzikov A. V. External Force Generation for Object Segmentation on 3D Ultrasound Images Using Simplex Meshes. *Pattern Recognition and Image Analysis*, 2006, vol. 16, no. 1, pp. 89–92.
19. Bogush A. L., Tuzikov A. V. 3D Object Volume Measurement Using Freehand Ultrasound, 14th Scandinavian Conference on Image Analysis, 19–22 June 2005, Joensuu, Finland. *Lecture Notes in Computer Science*, 2005, vol. 3544, pp. 1066–1075.
20. Tuzikov A. V., Sheynin S. A., Vasiliev P. V. Computation of volume and surface body moments. *Pattern Recognition*, 2003, vol. 36, no. 11, pp. 2521–2529.
21. Sheynin S. A., Tuzikov A. V. Moment computation for objects with spline curve boundary. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, vol. 25, no. 10, pp. 1317–1322.
22. Tuzikov A. V., Colliot O., Bloch I. Evaluation of the symmetry plane in 3D MR brain images. *Pattern Recognition Letters*, 2003, vol. 24, no. 14, pp. 2219–2233.
23. Tuzikov A., Roerdink J. B. T. M., Heijmans H. J. A. M. Similarity measures for convex polyhedra based on Minkowski addition. *Pattern Recognition*, 2000, vol. 33, no. 6, pp. 979–995.
24. Heijmans H. J. A. M., Tuzikov A. Similarity and symmetry measures for convex shapes using Minkowski addition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, vol. 20, no. 9, pp. 980–993.

Правила для авторов

Редакция журнала «Информатика» просит авторов руководствоваться приведенными ниже правилами:

1. Статьи принимаются в редакцию через электронную систему подачи по адресу <http://inf.grid.by> в формате файлов текстовых редакторов Microsoft Word. Основной текст статьи не должен превышать 17 стр., включая рисунки, таблицы и достаточное количество наиболее актуальных ссылок; обзорной статьи – 10 стр., включая все основные ссылки. Текст набирается с переносами, шрифт Times New Roman 11 пт, интервал между строками одинарный, абзацный отступ 1 см, поля по 2,5 см со всех сторон.

Изложенный в статье материал должен быть четко структурированным: введение, цели и задачи, методы, результаты, заключение (выводы).

2. Статьи о результатах работ, проведенных в научных учреждениях, должны иметь разрешение на публикацию (сопроводительное письмо за подписью руководителя или выписку из заседания ученого совета, отдела или кафедры, акт экспертизы).

3. Статья в обязательном порядке должна иметь следующую структуру: индекс по универсальной десятичной классификации (УДК); инициалы и фамилии всех авторов, название статьи, полное название учреждений, где работают авторы, с указанием города, страны, аннотацию (150–250 слов), подрисовочные надписи, названия таблиц и ключевые слова (7–10) на русском и английском языках, адрес электронной почты каждого автора и номер телефона.

4. Аннотация (авторское резюме) должна кратко представлять результаты работы и быть информативной, содержательной. Приветствуется структура аннотации, повторяющая структуру статьи и включающая введение, цели и задачи, методы, результаты, заключение.

5. Формулы, рисунки, таблицы в статье нумеруются в соответствии с порядком их упоминания в тексте. Ссылки на рисунки и таблицы в тексте обязательны. Рисунки должны быть выполнены с хорошим разрешением в масштабе, позволяющем четко различать надписи и обозначения. Подрисовочные подписи с расшифровкой всех позиций, представленных на рисунке, набираются шрифтом гарнитуры основного текста размером 9 пт. Цветные иллюстрации печатаются только в том случае, когда это необходимо для понимания излагаемого материала.

6. Набор формул выполняется в формульном редакторе Microsoft Equation или Math Type. Прямым шрифтом набираются: греческие и русские буквы; математические символы (\sin , \lg , ∞); символы химических элементов (C, Cl, CHCl_3); цифры (римские и арабские); векторы; индексы (верхние и нижние), являющиеся сокращениями слов. Курсивом набираются латинские буквы, символы физических величин (в том числе и в индексе).

7. Сокращения в тексте статьи (за исключением единиц измерения) могут быть использованы только после упоминания полного термина. Единицы измерения физических величин следует приводить в Международной системе единиц (СИ).

8. Цитируемые в статье фамилии авторов теорем, теорий, законов и т. д. следует приводить в скобках на языке оригинала после русского написания. Например, теорема Эйлера (Euler).

9. Список использованной литературы оформляется в соответствии с требованиями Высшей аттестационной комиссии Республики Беларусь (ГОСТ 7.5–2008). Номер литературной ссылки в тексте дается порядковым номером в квадратных скобках. Ссылаться на неопубликованные работы не допускается.

10. Отдельно приводится список цитированных источников в *романском* (латинском) алфавите со следующей структурой: авторы (транслитерация), название статьи в транслитерированном варианте [перевод названия статьи на английский язык в квадратных скобках], название русскоязычного источника (транслитерация) [перевод названия источника на английский язык – парафраз (для журналов можно не делать)], выходные данные с обозначениями на английском языке.

11. Поступившие в редакцию статьи направляются на рецензирование специалистам. Основным критерием целесообразности публикации является новизна и информативность статьи. Если по рекомендациям рецензента статья возвращается автору на доработку, а переработанная рукопись вновь рассматривается редколлегией, датой поступления считается день получения редакцией ее окончательного варианта. Статьи не по профилю журнала возвращаются авторам после заключения редколлегии.

12. Статьи, направляемые на доработку, должны быть возвращены в исправленном виде с ответами на все замечания.

13. Редакция журнала предоставляет возможность первоочередного опубликования статей, представленных лицами, которые осуществляют послевузовское обучение (аспирантура, докторантура, соискательство) в год завершения обучения.

14. Авторы несут ответственность за направление в редакцию статей, уже опубликованных ранее или принятых к публикации другими изданиями.

15. Редакция оставляет за собой право на редакционные изменения, не искажающие основное содержание статьи. Окончательное решение о публикации принимается редакционной коллегией.

Журнал «Информатика» включен Высшей аттестационной комиссией Республики Беларусь в список научных изданий для опубликования результатов диссертационных исследований.

Индексы

00827

для индивидуальных
подписчиков

008272

для предприятий и
организаций